



# An Exploration of Word-Scale Visualizations for Text Documents

Pascal Goffin

## ► To cite this version:

Pascal Goffin. An Exploration of Word-Scale Visualizations for Text Documents. Human-Computer Interaction [cs.HC]. Université Paris Saclay (COmUE), 2016. English. NNT : 2016SACLS256 . tel-01404157

**HAL Id: tel-01404157**

**<https://theses.hal.science/tel-01404157>**

Submitted on 28 Nov 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE DE DOCTORAT  
DE  
L'UNIVERSITÉ PARIS-SACLAY  
PRÉPARÉE À  
L'INRIA

ECOLE DOCTORALE N°580  
Sciences et Technologies de l'Information et de la Communication

Spécialité de doctorat INFORMATIQUE

Par

M. PASCAL GOFFIN

AN EXPLORATION OF  
WORD-SCALE VISUALIZATIONS FOR TEXT DOCUMENTS

Thèse présentée et soutenue à  
Inria Saclay - Île-de-France / Bâtiment Digiteo Moulon, Gif-sur-Yvette  
le 3 octobre 2016 :

**Directeur de thèse:**

JEAN-DANIEL FEKETE

Directeur de Recherche (Inria Saclay, France)

**Co-encadrant de thèse:**

PETRA ISENBERG

Chargé de Recherche (Inria Saclay, France)

Composition du Jury:

Président du jury:

EMMANUEL PIETRIGA

Directeur de Recherche (Inria Saclay / LRI, France)

Rapporteurs:

CHRISTOPHER COLLINS

Associate Professor (University of Ontario Institute of Technology, Canada)

CHRISTOPHE HURTER

Professor (ENAC, France)

Examineur:

UTA HINRICHS

Lecturer (University of St. Andrews, Scotland)

WESLEY WILLETT

Assistant Professor (University of Calgary, Canada)

**Titre :** Une exploration des visualisations-mots pour du texte

**Mots clés :** Visualisations d'informations, petites visualisations contextuelles, visualisations-mots, visualisation de texte, sparklines, glyphes.

**Résumé :** Ma dissertation explore comment l'intégration de petites visualisations contextuelles basées sur des données peut compléter des documents écrits. Plus spécifiquement, j'identifie et je définis des aspects importants et des directions de recherches pertinentes pour l'intégration de petites visualisations contextuelles basées sur des données textuelles. Cette intégration devra finalement devenir aussi fluide qu'écrire et aussi utile que lire un texte. Je définis les visualisations-mots (Word-Scale Visualizations) comme étant de petites visualisations contextuelles basées sur des données intégrées au texte de documents. Ces visualisations peuvent utiliser de multiples codages visuels incluant les cartes géographiques, les heatmaps, les graphes circulaires, et des visualisations plus complexes. Les visualisations-mots offrent une grande variété de dimensions toujours proches de l'échelle d'un mot, parfois plus grandes, mais toujours plus petites qu'une phrase ou un paragraphe. Les visualisations-mots peuvent venir en aide et être utilisées dans plusieurs formes de discours écrits tels les manuels, les notes, les billets de blogs, les rapports, les histoires, ou même les poèmes. En tant que complément visuel de textes, les visualisations-mots peuvent être utilisées pour accentuer certains éléments d'un document (comme un mot ou une phrase), ou pour apporter de l'information additionnelle. Par exemple, un petit diagramme de l'évolution du cours de l'action d'une entreprise peut être intégré à côté du nom de celle-ci pour apporter de l'information additionnelle sur la tendance passée du cours de l'action. Dans un autre exemple, des

statistiques de jeux peuvent être incluses à côté du nom d'équipe de football ou de joueur dans les articles concernant le championnat d'Europe de football. Ces visualisations-mots peuvent notamment aider le lecteur à faire des comparaisons entre le nombre de passes des équipes et des joueurs. Le bénéfice majeur des visualisations-mots réside dans le fait que le lecteur peut rester concentré sur le texte, vu que les visualisations sont dans le texte et non à côté.

Dans ma thèse j'apporte les contributions suivantes : j'explore pourquoi les visualisation-mots peuvent être utiles et comment promouvoir leur création. J'étudie différentes options de placement pour les visualisations-mots et je quantifie leurs effets sur la disposition du texte et sa mise en forme. Comme les visualisations-mots ont aussi des implications sur le comportement de lecture du lecteur, je propose une première étude qui étudie les différents placements de visualisations-mots sur le comportement de lecture. J'examine aussi comment combiner les visualisations-mots et l'interaction pour soutenir une lecture plus active en proposant des méthodes de collection, d'arrangement et de comparaison de visualisations-mots. Finalement, je propose des considérations de design pour la conception et la création de visualisations-mots et je conclus avec des exemples d'application. En résumé cette dissertation contribue à la compréhension de petites visualisations contextuelles basées sur des données intégrées dans le texte et à leur utilité pour la visualisation d'informations.

**Title :** An Exploration of Word-Scale Visualizations for Text Documents

**Keywords :** Information visualization, small contextual visualization, word-scale visualization, text visualization, sparklines, glyphs.

**Abstract :** This dissertation explores how embedding small data-driven contextual visualizations can complement text documents. More specifically, I identify and define important aspects and relevant research directions for the integration of small data-driven contextual visualizations into text. This integration should eventually become as fluid as writing and as usable as reading a text. I define word-scale visualizations as small data-driven contextual visualizations embedded in text documents. These visualizations can use various visual encodings including geographical maps, heat maps, pie charts, and more complex visualizations. They can appear at a range of word scales, including sizes larger than a letter, but smaller than a sentence or paragraph. Word-scale visualizations can help support and be used in many forms of written discourse such as text books, notes, blog posts, reports, stories, or poems. As graphical supplements to text, word-scale visualizations can be used to emphasize certain elements of a document (e.g. a word or a sentence), or to provide additional information. For example, a small stock chart can be embedded next to the name of a company to provide additional information about the past trends of its stocks. In another example, game statistics can be embedded next to the names of soccer teams or players in daily reports from the UEFA European Championship.

These word-scale visualizations can then for example allow readers to make comparison between number of passes of teams and players. The main benefit of word-scale visualizations is that the reader can remain focused on the text as the visualization are within the text rather than alongside it.

In the thesis, I make the following main contributions: I explore why word-scale visualizations can be useful and how to support their creation. I investigate placement options to embed word-scale visualizations and quantify their effects on the layout and flow of the text. As word-scale visualizations also have implications on the reader's reading behavior I propose a first study that investigates different word-scale visualization positions on the reading behavior. I also explore how word-scale visualizations can be combined with interaction to support a more active reading by proposing interaction methods to collect, arrange and compare word-scale visualizations. Finally, I propose design considerations for the authoring of word-scale visualizations and conclude with application examples.

In summary, this dissertation contributes to the understanding of small data-driven contextual visualizations embedded into text and their value for Information Visualization.

To my parents and my brother.



## PUBLICATIONS

---

### 0.1 JOURNAL ARTICLES

- Pascal Goffin, Jeremy Boy, Wesley Willett, and Petra Isenberg. “An Exploratory Study of Word-Scale Graphics in Data-Rich Text Documents.” In: *IEEE Transactions on Visualization and Computer Graphics* PP.99 (2016), pp. 1–1
- Pascal Goffin, Wesley Willett, Jean-Daniel Fekete, and Petra Isenberg. “Exploring the Placement and Design of Word-Scale Visualizations.” In: *IEEE Transactions on Visualization and Computer Graphics* 20.12 (2014), pp. 2291–2300

### 0.2 POSTER AND WORKSHOP ARTICLES

- Pascal Goffin, Wesley Willett, Jean-Daniel Fekete, and Petra Isenberg. “Design considerations for Enhancing Word-Scale Visualizations with Interaction.” In: *Posters of the IEEE Conference on Information Visualization (InfoVis)*. Los Alamitos, CA, USA: IEEE, 2015
- Pascal Goffin, Wesley Willett, Anastasia Bezerianos, and Petra Isenberg. “Exploring the effect of word-scale visualizations on reading behavior.” In: *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM. 2015, pp. 1827–1832
- Pascal Goffin, Wesley Willett, and Petra Isenberg. “Sharing Information from Personal Digital Notes using Word-Scale Visualizations.” In: *Personal Visualization: Exploring Data in Everyday Life*. 2015

### 0.3 PUBLICATION *in progress*

- Pascal Goffin, Wesley Willett, Jean-Daniel Fekete, and Petra Isenberg. “Interaction Techniques for Visual Exploration using Embedded Word-Scale Visualizations.” In: ()



## ACKNOWLEDGMENTS

---

I would like to use this page to thank all the people that made this research possible but also the people that helped me enjoy the many weekends, here, in Paris.

Petra Isenberg owns a big thank you for mentoring me, providing me with the opportunity to discover the fields of Information Visualization and Human-Computer Interaction, and allowing me to pursue the research I wanted to do.

Jean-Daniel Fekete owns a big thank you for letting me be part of this group. Thanks for your feedback and inputs.

Many thanks to Wesley Willett for his immense help and support, always ready to answer my questions and giving awesome feedback about my writings. Hope you have not lost to many hairs because of me. Also keep motivating your students as you do, they will appreciate it.

Many thanks to all my colleagues that accompanied my PhD at Aviz, making for a great and fun working atmosphere, namely: Anthi, Anastasia, Benjamin, Charles, Christoph, Jeremy, Katia, Lonni, Lora, Mathieu, Manolis, Nadia, Paola, Pierre, Romain (2x), Samuel, Tobias, and Yvonne.

But also many thanks to PUC Floorball for having these great training sessions on Fridays and Sundays, the weekend tournaments and allowing me to be part of the team playing in the French Championship, namely: Alban, Clément, Eric, Erwan, Fabien, Guillaume, JE, Jéréemie, JT, Juan, Julien A, Julien L, Kevin, Loik, Marie, Martino, Mathieu, Michel, Olivier, Rod, Sidney, and Vilis. If I forgot to mention somebody, I am sorry.

Finally, a big thank you to my parents and Baruki for always supporting me and giving me the opportunity to follow my path as well as my brother for constantly challenging me and being there when I needed you. I hope to have more time now to do awesome bike rides!





## CONTENTS

---

0.1	Journal Articles . . . . .	v
0.2	Poster and Workshop Articles . . . . .	v
0.3	Publication <i>in progress</i> . . . . .	v
1	INTRODUCTION . . . . .	1
1.1	Research Context . . . . .	4
1.2	Research Scope . . . . .	6
1.3	Thesis Problem and Approach . . . . .	7
1.4	Main Contributions . . . . .	9
1.5	Overview of the Dissertation . . . . .	9
2	THE SPACE OF DOCUMENT VISUALIZATION . . . . .	13
2.1	Document-Centric and Visualization-Centric Analysis . . . . .	13
2.2	Document Visualization . . . . .	15
2.2.1	Single Document Visualization . . . . .	17
2.2.1.1	Terms in Context . . . . .	17
2.2.1.2	Pattern Analysis . . . . .	18
2.2.1.3	Overview of a Document . . . . .	18
2.2.2	Document Collection Visualization . . . . .	20
2.2.2.1	Terms in Context . . . . .	21
2.2.2.2	Pattern Analysis . . . . .	21
2.2.2.3	Overview of a Collection of Document . . . . .	22
2.2.2.4	Variation of Topics . . . . .	22
2.3	Combining Text and Graphics . . . . .	22
2.4	Summary . . . . .	24
3	SMALL-SCALE AND DATA-DRIVEN GRAPHICS . . . . .	25
3.1	Small Contextual Visualizations . . . . .	25
3.1.1	Word-Scale Visualization Designs . . . . .	26
3.1.2	Non-Data-Driven Word-Scale Graphics . . . . .	29
3.2	Word-Scale Visualizations in the Wild . . . . .	30
3.3	Summary . . . . .	31
4	USE AND DESIGN OF SMALL CONTEXTUAL VISUALIZATIONS . . . . .	33
4.1	Motivation . . . . .	33
4.2	Study-Designing Word-Scale Visualizations . . . . .	34
4.2.1	Participants . . . . .	35
4.2.2	Task . . . . .	35
4.2.3	Dataset . . . . .	35
4.2.4	Study Procedure . . . . .	37
4.2.5	Set-Up and Materials . . . . .	38
4.2.6	Analysis Procedure . . . . .	38
4.3	Results . . . . .	39
4.3.1	Data vs. Non-Data-Driven Encodings . . . . .	39
4.3.1.1	Data Provenance . . . . .	40

	4.3.1.2	Types of Data . . . . .	40
	4.3.2	Design of Word-Scale Representations . . . . .	41
	4.3.2.1	Types of Visual Representation . . . . .	41
	4.3.2.2	Reusable Graphical Systems . . . . .	42
	4.3.3	Connections Between Text and Graphics . . . . .	43
	4.3.3.1	Size of the Associated Text . . . . .	43
	4.3.3.2	Type of Connection . . . . .	44
	4.3.3.3	Legends and Captions . . . . .	45
	4.3.4	Supporting Comparison and Providing Context . . . . .	46
	4.3.5	Interaction . . . . .	48
	4.3.6	Designer's Goals . . . . .	49
4.4		Discussion: Using Word-Scale Visualizations . . . . .	51
	4.4.1	Applications of Word-Scale Visualizations . . . . .	51
	4.4.1.1	Supporting the Content . . . . .	51
	4.4.1.2	Extending the Content . . . . .	52
	4.4.1.3	Going Beyond the Content . . . . .	52
	4.4.2	Defining "Word-Scale" . . . . .	52
	4.4.3	What is an Entity? . . . . .	53
	4.4.4	Challenges when Creating Word-Scale Visual- izations . . . . .	54
4.5		Limitations . . . . .	55
4.6		Summary . . . . .	56
5		PLACING WORD-SCALE VISUALIZATIONS . . . . .	59
	5.1	Motivation . . . . .	59
	5.2	Background . . . . .	60
	5.2.1	Sparklines and Word-Scale Visualizations . . . . .	60
	5.2.2	Meta-Data within Text Documents . . . . .	61
	5.2.3	Labeling . . . . .	62
	5.2.4	Readability of Text and Small-Scale Visualization . . . . .	63
	5.3	A Design Space for Word-Scale Visualizations . . . . .	64
	5.3.1	Definitions and Framing . . . . .	64
	5.3.1.1	The Size of Word-Scale Visualizations . . . . .	64
	5.3.1.2	Static vs. Dynamic Integration . . . . .	65
	5.3.1.3	Placement Context . . . . .	66
	5.3.2	Design Space Dimensions . . . . .	67
	5.4	Word-Scale Visualization Placement in Practice . . . . .	68
	5.5	Word-Scale Visualization Placement Examples . . . . .	73
	5.5.1	Word-Scale Visualizations in Scanned Documents . . . . .	73
	5.5.2	Word-Scale Visualization For Information Scent . . . . .	74
	5.5.3	Word-Scale Visualization for Detail-in-Context . . . . .	76
	5.6	Evaluation . . . . .	77
	5.6.1	Documents . . . . .	77
	5.6.2	Factors and Procedure . . . . .	78
	5.6.3	Measures . . . . .	78
	5.6.4	Results . . . . .	79

5.6.4.1	Comparing Vertical and Horizontal Space Adjustment . . . . .	80
5.7	Considerations for Word-Scale Visualization Placement . . . . .	81
5.7.1	Placement Advice . . . . .	82
5.8	Sparklificator . . . . .	82
5.9	Considerations for Future Work . . . . .	84
5.9.1	Weaker Context and Out-Of-Context Placements . . . . .	84
5.9.2	Gaining Space . . . . .	85
5.9.3	Word-Scale Visualizations and the Document Life-cycle . . . . .	86
5.9.4	Legibility of Word-Scale Visualizations . . . . .	86
5.10	Summary . . . . .	87
6	INTERACTING WITH WORD-SCALE VISUALIZATIONS . . . . .	89
6.1	Motivation . . . . .	89
6.2	Background . . . . .	90
6.3	Design Considerations for Interaction . . . . .	92
6.3.1	Where to Capture Interaction . . . . .	93
6.3.2	Interaction Intents . . . . .	93
6.3.3	Interaction Techniques . . . . .	95
6.3.4	Interaction Scope . . . . .	96
6.3.5	Summary . . . . .	96
6.4	Interaction Techniques for Visual Exploration . . . . .	97
6.4.1	Collecting Word-Scale Visualizations . . . . .	99
6.4.2	Arranging Word-Scale Visualizations . . . . .	100
6.4.2.1	Choosing a Layout . . . . .	100
6.4.2.2	Ordering . . . . .	101
6.4.3	Comparing Word-Scale Visualizations . . . . .	101
6.4.3.1	Aligning . . . . .	102
6.4.3.2	Direct Comparison . . . . .	103
6.4.4	Navigating the Document . . . . .	103
6.5	Example Use Cases . . . . .	103
6.5.1	Comparing Stocks Market Charts . . . . .	104
6.5.2	Historian Revisiting Archival Documents . . . . .	104
6.6	Discussion of Interaction Techniques . . . . .	105
6.6.1	From Document-Centric to Visualization-Centric Analysis . . . . .	105
6.7	Considerations for Future Work . . . . .	107
6.7.1	Word-Scale Visualizations as Tokens . . . . .	107
6.7.2	Multiple Word-Scale Visualization Instances . . . . .	107
6.7.3	Interacting with Word-Scale Visualizations in a Text Corpus . . . . .	108
6.7.4	Evaluation of Interaction Techniques . . . . .	108
6.8	Summary . . . . .	109
7	TOWARDS UNDERSTANDING THE EFFECTS OF WORD-SCALE VISUALIZATIONS ON THE READER . . . . .	111
7.1	Motivation . . . . .	111

7.2	Background . . . . .	112
7.3	Experiment . . . . .	113
7.3.1	Factors . . . . .	113
7.3.2	Task . . . . .	114
7.3.3	Procedure . . . . .	114
7.3.4	Measures . . . . .	117
7.3.5	Data Collection . . . . .	117
7.3.6	Participants . . . . .	117
7.3.7	Hypotheses . . . . .	118
7.3.8	Data Analysis . . . . .	118
7.4	Results . . . . .	118
7.4.1	Performance Results . . . . .	118
7.4.2	Questionnaire Results . . . . .	120
7.5	Discussion of the Study . . . . .	122
7.6	Considerations for Future Work . . . . .	123
7.6.1	Reading Behavior . . . . .	124
7.6.1.1	Suggested Factors . . . . .	124
7.6.1.2	Suggested Setup . . . . .	125
7.6.2	Memorability and Comprehension . . . . .	125
7.7	Summary . . . . .	126
8	AUTHORING WORD-SCALE VISUALIZATIONS . . . . .	127
8.1	Why Authoring Tools . . . . .	127
8.2	Background . . . . .	128
8.3	Implications for Authoring Tools . . . . .	129
8.4	Considerations for Future Work . . . . .	130
8.5	Summary . . . . .	132
9	AN APPLICATION OF WORD-SCALE VISUALIZATIONS . . . . .	133
9.1	Motivation . . . . .	133
9.2	Background . . . . .	134
9.3	Word-scale visualizations as an Approach to Present Shared Information . . . . .	135
9.3.1	Sharing metadata and snippets . . . . .	136
9.3.1.1	Technical metadata . . . . .	136
9.3.1.2	Extracted keywords . . . . .	137
9.4	Examples . . . . .	137
9.4.1	Shared recipes with Roommates . . . . .	137
9.4.1.1	Sketch . . . . .	138
9.4.2	Book Club . . . . .	139
9.4.2.1	Sketch . . . . .	139
9.5	Challenges and Implications . . . . .	140
9.5.1	Visualization . . . . .	140
9.5.2	Privacy . . . . .	141
9.5.3	Sharing useful data . . . . .	141
9.6	Summary . . . . .	141
10	SUMMARY AND PERSPECTIVES . . . . .	143
10.1	Summary . . . . .	143

10.2 Perspectives . . . . .	146
10.2.1 Integration of Word-Scale Visualizations . . . . .	146
10.2.2 Evaluation of Word-Scale Visualizations . . . . .	147
10.2.3 Transitioning from Document-Centric to Visualization-Centric Analysis using Word-Scale Visualizations	147
10.3 Conclusion . . . . .	148
<b>I APPENDICES</b>	149
A USE AND DESIGN OF SMALL CONTEXTUAL VISUALIZATIONS	151
A.1 Illustrator Articles . . . . .	151
A.2 Questionnaire . . . . .	163
A.3 Designs by the Study Participants . . . . .	164
B PLACING WORD-SCALE VISUALIZATIONS	179
B.1 Sparklifier Source Code . . . . .	179
C TOWARDS UNDERSTANDING THE EFFECTS OF WORD-SCALE VISUALIZATIONS ON THE READER	189
C.1 Questionnaire . . . . .	189
BIBLIOGRAPHY	195

## LIST OF FIGURES

---

Figure 1.2	Bandeau at the top of the page, rules (filets) in the middle surrounding ACTE PREMIER. . . . .	1
Figure 1.1	Hieroglyphs and additional visual elements. . .	2
Figure 1.3	Two pages from the “Hours of Catherine of Cleves” from 1440 depicting different scenes of Catherine of Cleves; a magnificent example of the combination of text and graphics. . . . .	3
Figure 1.4	Three examples of the combination of text and visualizations. . . . .	4
Figure 1.5	The document-centric side (left) and the visualization-centric side (right). . . . .	5
Figure 1.6	Three of the multiple views of Jigsaw: the list view, the graph view and the document view. .	6
Figure 2.1	The document-centric side (left) and the visualization-centric side (right). . . . .	14
Figure 2.2	Kucher and Kerren’s taxonomy of text visualization techniques. . . . .	15
Figure 2.3	Taxonomy of visual text analysis tasks. . . . .	16
Figure 2.4	Word Tree visualization showing the search term “love the” in context. . . . .	17
Figure 2.5	SeeSoft visualizing a document. . . . .	18
Figure 2.6	TextArc visualizing “Alice’s Adventures In Wonderland”. . . . .	19
Figure 2.7	DocuBurst rooted at the node “physical phenomenon”, and a search query for “electricity” has been performed. The nodes that match are colored in light yellow. . . . .	20
Figure 2.8	The view of a Document Card with each rectangle on the right side representing a page of the document. . . . .	21
Figure 2.9	On the left the TileBars can be seen, where as on the right the titles of the document are shown.	22
Figure 2.10	The view of a collection of structured documents represented by the Compus system. . . .	23
Figure 2.11	One-word tag cloud of a document. . . . .	23
Figure 2.12	Each topic is visualized using a different color, the wider the current the more the topic is used.	24
Figure 2.13	Three different associations of text and visualizations. . . . .	24

Figure 3.1	Win/loss sparkline generated using Schwartz's random sparkline generator [132]. All four and more game winning streaks are shown in red. . . . .	25
Figure 3.2	HorizonGraph. . . . .	28
Figure 3.3	Visual scent providing information on the widget. . . . .	28
Figure 3.4	An "escort card" that is transcribed as "May I see you home my dear?" from 1865. . . . .	29
Figure 3.5	his is an excerpt from an automobile travel guidebook from 1909-1911 by the Automobile Club of America. Here, word-scale graphics are used to provide detailed landmarks for driving orientation. . . . .	30
Figure 3.6	Sparklines function as references back to the larger interactive visualizations in the article. . . . .	30
Figure 3.7	Example of a Sparktweet created with [81] . . . . .	31
Figure 4.1	Examples of word-scale graphics collected in our study. . . . .	34
Figure 4.2	Participant details . . . . .	35
Figure 4.3	Thumbnails of the three Wikipedia articles. From left to right: "Regions of France", "Europe", and "Human Body". . . . .	36
Figure 4.4	Number of word-scale graphics per participant . . . . .	39
Figure 4.5	Examples of data-driven and non-data-driven designs. . . . .	39
Figure 4.6	Levels of types of data, their counts and examples. . . . .	40
Figure 4.7	Levels of types of graphic, their counts, and examples . . . . .	41
Figure 4.8	Example of an implementation of a reusable graphical system. The square stands for a region unit and the circle for a department. A bar chart and a trend visualization are assembled from these units. . . . .	43
Figure 4.9	Types of connections between designs and text. . . . .	43
Figure 4.10	The word-scale visualization has been positioned to the left of the paragraph, showing the proportion of countries using the Euro (18)—represented as 18 € signs—versus the number of countries actually in Europe (28)—represented using 28 icons of a flag. . . . .	45
Figure 4.11	The word-scale visualization has been added to the title of the paragraph. It is a horizontal bar chart comparing the assets under management in Europe and North America. . . . .	45

Figure 4.12	The three word-scale visualizations are placed below the entity. The first graphic represents the average adult body as a container. Then, in the second graphic the 5 liters of blood are added. And the third graphic connects “blood” to the red/orange color. . . . .	45
Figure 4.13	Several ways in which graphics and entities serve as legends or captions. . . . .	46
Figure 4.14	A nested legend and caption explains the top bar chart in (a). The right image (b) shows an example of an explicit legend created and placed in the margin by a designer. . . . .	47
Figure 4.15	Connecting multiple visualizations with a support grid. . . . .	47
Figure 4.16	Repeated grid and icons support comparison (we added the grey annotations for explanatory purposes). . . . .	48
Figure 4.17	Overlaying a trace of one visualizations over another (we added the grey annotations for explanatory purposes). . . . .	48
Figure 4.18	Sequencing of timeline word-scale visualizations.	48
Figure 4.19	Three connected word-scale visualizations. Moving the cursor in one graphic results in highlighted values in the others (we added the red dots and the cursor to explain the interaction) .	49
Figure 4.20	Using interaction to overlay the word-scale visualizations for one country on top of the graphic for another. (Sizes are approximate—as estimated by the participant. We added the cursor and line for illustration.) . . . . .	49
Figure 4.21	Word-scale visualization showing the contrast between population and surface area for the EU, Russia, and Vatican City (we added the grey dashed lines and annotations for explanatory purpose). . . . .	50
Figure 5.1	Two examples of ruby placement: on top of the kanjis (a) and to the right of the kanjis (b). . . . .	61
Figure 5.2	Important terminology used in the paper in the context of entity, word-scale visualization, and surrounding text. . . . .	65
Figure 5.3	The word-scale visualization has three different entities each using a color. The designer chose the numeral of the second entity as the main entity. . . . .	66



Figure 5.4	Visual description of the three placement positions: <i>baseline</i> position, <i>top</i> position, and <i>right</i> position. . . . .	68
Figure 5.5	Several word-scale visualization placements options and the design space decisions that produced them. A ✓ stands for “available or needed” and a – stands for “not available or not needed”. . . . .	69
Figure 5.6	Word-scale visualizations used in a scanned World War I artifact. . . . .	74
Figure 5.7	Small word-scale visualizations that provide information scent to help viewers decide whether or not they are interested in following an article linked to from a Wikipedia page. . . . .	75
Figure 5.8	Small maps embedded as word-scale visualizations provide details-on-demand for locations in a news story. . . . .	76
Figure 5.9	Parameters used to produce the seven placement example cases used in our evaluation, along with average change in document area, visualization/visualization overlap and visualization/text, and text shift statistics for each. . . . .	79
Figure 5.10	The effect of word-scale visualization placement strategy on the overall expansion of the document as more visualizations are added. . . . .	81
Figure 5.11	Sparklificator can be found on GitHub under <a href="http://inria.github.io/sparklificator/">http://inria.github.io/sparklificator/</a> . . . . .	83
Figure 5.12	Output by Sparklificator. . . . .	84
Figure 5.13	Top: user hovers over NOME,AK. The underline signals there is additional information. Bottom: inter-line space is reduced between all other lines to fit in the annotation below the underlined word. . . . .	85
Figure 6.1	Interactions can take place directly with the entire visualization (a), the related text entity (b), connected GUI controls like buttons (c), or with individual marks in the visualization itself (d). . . . .	93
Figure 6.2	Word-scale visualization showing the contrast between population and surface area for the EU, Russia, and Vatican City (we added the grey dashed lines and annotations for explanatory purpose). . . . .	95
Figure 6.3	Three connected word-scale visualizations. Moving the cursor in one graphic results in highlighted values in the others (we added the red dots and the cursor to explain the interaction) . . . . .	95

Figure 6.4	Interaction triggers the overlay of two country word-scale visualizations (we added the line between the two graphics and the cursor for illustration purpose). . . . .	96
Figure 6.5	The interaction design space, showing the interaction intents, the scope of the interaction, the location where interaction can be triggered, and how this interaction can be triggered. . . . .	97
Figure 6.6	An example of an word-scale visualization interaction (top) described using the four design dimensions (bottom). . . . .	98
Figure 6.7	The top image shows embedded word-scale visualizations in a news article. The bottom image shows the result of the gathering interaction. . . . .	99
Figure 6.8	Collecting word-scale visualizations. . . . .	100
Figure 6.9	Arranging word-scale visualizations. . . . .	101
Figure 6.10	Ordering word-scale visualizations. . . . .	102
Figure 6.11	Aligning word-scale visualizations. . . . .	102
Figure 6.12	Comparing word-scale visualizations. . . . .	103
Figure 6.13	Word-scale visualizations of timelines embedded in an historian's note. In the bottom image, word-scale visualizations and entities are aligned along their dominant axes to help compare timespans (horizontal bar). . . . .	104
Figure 6.14	From left to right: Document with embedded word-scale visualizations. Word-scale visualizations collected around an entity of interest with text visible. Small multiples without text. A single visualization. . . . .	105
Figure 6.15	Small maps embedded as word-scale visualizations provide details-on-demand for locations in a news story. . . . .	106
Figure 7.1	The four word-scale visualization positions: (a) "top", (b) "right", (c) interactive "overlay" (d) "no word-scale visualization" (no wsv). . . . .	112
Figure 7.2	An example sentence describing information that is ambiguously encoded in the sentence (trend = increasing) and in the word-scale visualization (decreasing). Colors added for illustration. . . . .	113
Figure 7.3	An example sentence for the ant colony "Druselstein" with the statistic "average class size" characterized by a <i>number</i> and a <i>trend</i> over the last 50 ant years. Colors added for illustration. . . . .	114

Figure 7.4	A trial consists of a Sentence Page, a Question Page, and a Break Page. The Space Bar is used to move from page to the next. . . . .	115
Figure 7.5	The two multiple choice questions for the sentence on the left (same sentence as in Figure 7.2).	116
Figure 7.6	95% CI for participants' time across placement positions. . . . .	119
Figure 7.7	95% CI for the accuracy for sentences with and without word-scale visualization (in the non-ambiguous cases). . . . .	119
Figure 7.8	95% CI for people's choices in the ambiguous case. . . . .	119
Figure 7.9	Summarized questionnaire data using Bertifier [116]: Grouped participant responses w.r.t. ranking for each position. . . . .	120
Figure 8.1	Visualization reference model by Card et al. . . . .	128
Figure 9.1	This sketch is an excerpt of a recipe from the flatmates' shared recipe notebook with embedded word-scale visualizations. . . . .	138
Figure 9.2	This sketch is an excerpt of one of the book club member's note about the current book. The bar chart describes the number of occurrences of the character in each book club member's current book note. . . . .	139
Figure 10.1	From left to right: Document with embedded word-scale visualizations. Word-scale visualizations collected around an entity of interest with text visible. Small multiples without text. A single visualization. . . . .	147
Figure A.1	"Regions of France" Wikipedia article (part 1). . . . .	151
Figure A.2	"Regions of France" Wikipedia article (part 2). . . . .	152
Figure A.3	"Europe" Wikipedia article (part 1). . . . .	153
Figure A.4	"Europe" Wikipedia article (part 2). . . . .	154
Figure A.5	"Europe" Wikipedia article (part 3). . . . .	155
Figure A.6	"Human Body" Wikipedia article (part 1). . . . .	156
Figure A.7	"Human Body" Wikipedia article (part 2). . . . .	157
Figure A.8	"Human Body" Wikipedia article (part 3). . . . .	158
Figure A.9	"Human Body" Wikipedia article (part 4). . . . .	159
Figure A.10	"Zurich" Wikipedia article (part 1). . . . .	160
Figure A.11	"Zurich" Wikipedia article (part 2). . . . .	161
Figure A.12	"Zurich" Wikipedia article (part 3). . . . .	162
Figure A.13	Questionnaire given to the participants at the beginning of the use and design study. . . . .	163
Figure C.1	Questionnaire effect of word-scale visualization on reader (part 1) . . . . .	189

Figure C.2	Questionnaire effect of word-scale visualization on reader (part 2) . . . . .	190
Figure C.3	Questionnaire effect of word-scale visualization on reader (part 3) . . . . .	191
Figure C.4	Questionnaire effect of word-scale visualization on reader (part 4) . . . . .	192
Figure C.5	Questionnaire effect of word-scale visualization on reader (part 5) . . . . .	193

## LIST OF SOURCE CODE

---

Source Code 5.1	HTML text with span around the entity. . . . .	83
Source Code 5.2	JavaScript to add a word-scale visualization to the entity <i>Pellentesque</i> . . . . .	83
Source Code B.1	File “jquery.sparklificator.js” from the jQuery Sparklificator library. . . . .	179
Source Code B.2	File “renderers.js” from the jQuery Sparklificator library. . . . .	183

## INTRODUCTION

---

The combination of text and various forms and sizes of graphics in different configurations has been popular for a long time. Already the ancient Egyptians included additional visual elements to their hieroglyphs for example to differentiate, or to add additional semantic information. They for example used a visual element called Cartouche around the Pharaoh's name to clearly mark that person as the Pharaoh (see Figure 1.1a). Moreover, hieroglyphs are text and illustration at the same time (see Figure 1.1b).

Writing systems evolved and their form became more and more abstract, loosing some of their illustrative side. Still, additional visual elements were added to text. For example the bandeau, a typographic ornament added to embellish a page (see Figure 1.2), or the rule (filet) which is a line of variable size and weight that emphasizes for example text parts or images (see Figure 1.2). There exists many more such additional typographic elements that can be added to text.

Before the 15th century and before Gutenberg invented the printing press, texts were handwritten works. Thus, it was simple to add new graphics. Writers could design their own graphics and include them into their texts. The simplicity to add visual elements resulted in multiple variations only limited by the creativity of the writer. Graphics were also an opportunity to make these texts unique. An example of such a handwritten text, full of additional visual elements, are illuminated manuscripts. In Figure 1.3 two example pages from such an illuminated manuscript can be seen. Even though this specific example is from the 15th century, illuminated manuscripts—or at least the ones that survived—appeared already around the 6th and 7th century. Thus, the combination of text and graphics has a long history.

However, starting with the influence of copyists and culminating with the emergence of Gutenberg's printing press, texts became more and more consistent. The invention of the printing press led to less variations in these additional graphics embedded into text. It was a time of simplification and streamlining the process of producing text. The emphasis was on the reproduction of these works and not their beautification. Nevertheless, visual elements never disappeared. Frutiger [51] provides an overview of the multitude of visual elements that designers embed into text for different reasons and purposes.

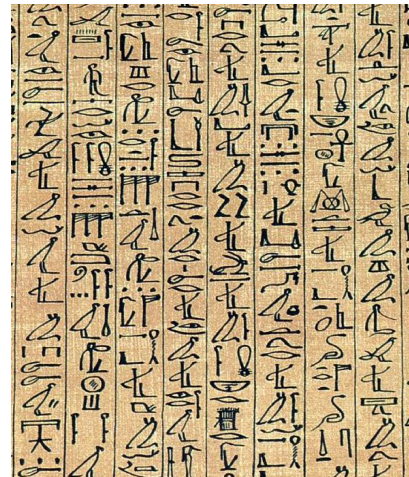
He shows that, various types of graphics of different sizes and forms exist and are used in combination with text. In this dissertation the graphics I am interested in are graphics based on or driven



Figure 1.2: Bandeau at the top of the page, rules (filets) in the middle surrounding ACTE PREMIER. Image in the public domain.



(a) The name of the Pharaoh—Thutmose III—is encircled by a cartouche. *Image in the public domain.*



(b) Section of the Papyrus of Ani showing hieroglyphs. *Image in the public domain.*

Figure 1.1: Hieroglyphs and additional visual elements.

by data—*data visualizations*. As Heer and Shneiderman [75] noted “Visualization provides a powerful means of making sense of data.”

The earliest appearance of data visualizations is dated to the 10th century as explained by Friendly [50]. However, it is difficult to identify how and when they appeared first to illustrate text with data. I found early examples of the combination of text and visualizations in the work of Leonardo da Vinci from the 16th and 17th century. Da Vinci *explains* complex situations such as the phenomenon of planet-shine in Figure 1.4a. Other examples are Copernicus who *illustrates* his theory on the Heliocentric model of the solar system by adding a graphical representation of the planets circulating around the sun as can be seen in Figure 1.4b, or Galileo who *supports* his planet theory by using sketches of his observations of Saturn through a telescope placed in-line instead of words (see Figure 1.4c). These visualizations were achieved by either providing visual representation of information from the text or by including relevant outside information as visualizations to better understand what has been written in the text.




Nowadays, researchers regularly use visualizations based on data from experiments to support their arguments in their research articles. Journalists and editors add map visualizations to illustrate the stories about the Syrian war in their newspaper articles. Chemistry teachers use visualizations of chemical formulas in chemistry textbooks to explain or teach how and why certain reactions occur.

More recently, interactive visualizations embedded in text have emerged in newspaper articles such as by the New York Times, to only mention one. Rooze [125] for example provides a curated collection of



Figure 1.3: Two pages from the “Hours of Catherine of Cleves” from 1440 depicting different scenes of Catherine of Cleves; a magnificent example of the combination of text and graphics. *Image in the public domain.*

such articles from the Guardian and the New York Times. In an effort to include data in Wikipedia pages, Wikipedians can now add interactive visualizations to Wikipedia [97] using Vega [151]—a declarative language.

This dissertation describes the integration of small, contextual and data-driven graphics into text that we call *word-scale visualizations*. Their goal is to communicate information that could not be succinctly conveyed by the text itself, close to where that data is referred to, and concurrently keeping the source text as the main point of interest for the reader. Examples include small stock charts embedded next to the name of a company such as Apple  104.39, game statistics next to the name of a soccer team such as for Belgium’s soccer team  87% pass success, or weather trends like rain averages next to the name of a city such as for Paris  26mm.

Word-scale visualizations depict data associated with the text. For example the article Europe in Wikipedia has relevant data associated with it in an “infobox” such as population number and area. Moreover, relevant data is also embedded within the text, for example the population numbers of the different member states in a table, the states which have the Euro as a currency visualized on a map, and various other interesting figures and statistics. Similarly, an article in a newspaper discussing different strategies towards immigration has associated data such as the nationalities of immigrants, their num-



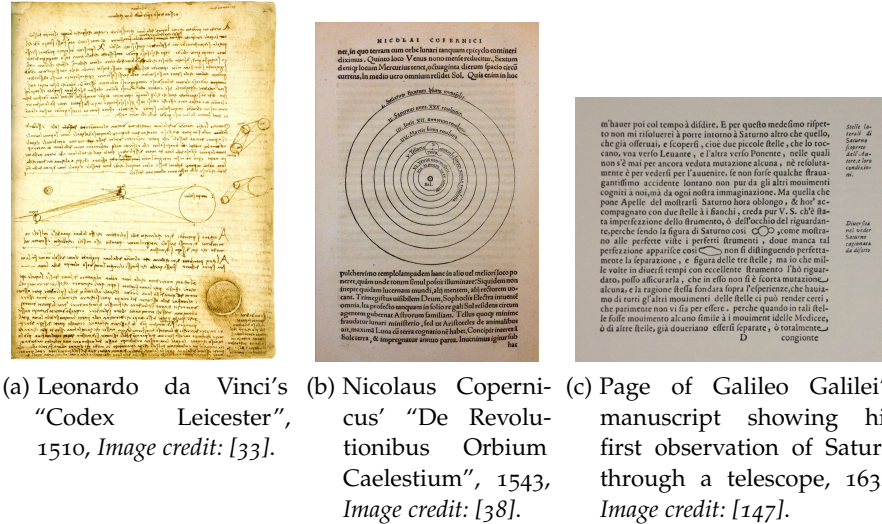


Figure 1.4: Three examples of the combination of text and visualizations.

bers, but also predictions of the outcome of these strategies on the number of immigrants. This associated data is either inherent to the text or linked from outside the text. Inherent to the text refers to the data being already in the text or embedded in text elements such as figures or tables as in the Wikipedia example before. Data linked from outside the text is data that cannot be found in the article, for example for the newspaper article example data from the immigration department that is linked (e.g. a reference) to the article describing the current immigration situation.

In the following chapters I provide the background and definition for these word-scale visualizations with respect to what they are and can be, their placement regarding the layout of the text, the interaction techniques used in connection with the visualizations and the text, a step towards understanding the effects of word-scale visualizations on readers, what should be considered when designing a word-scale visualization authoring system, and finally the illustration of an original application.

### 1.1 RESEARCH CONTEXT: A CONTINUUM FROM DOCUMENT-CENTRIC TO VISUALIZATION-CENTRIC ANALYSIS

Different fields of research use documents as a source of data in a leading role, for example in the Digital Humanities [82]. Digital Humanities scholars are interested in the analysis of related document and text snippets. In Information Visualization and Visual Analytics [86] includes the subfields of Text Visualization and Visual Text Analytics. *Text Visualization* and *Visual Text Analytics* refers to Information Visualization techniques that either focus on raw text data or on the results of text mining algorithms [91]. Similarly, *Document Vi-*



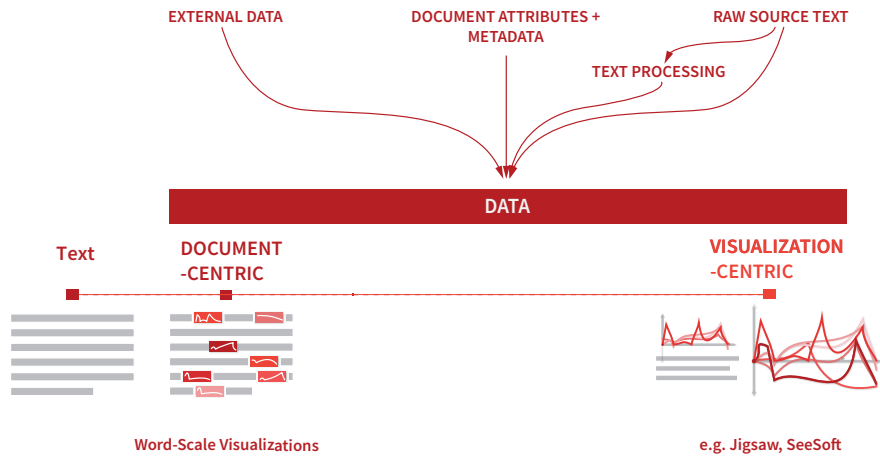


Figure 1.5: The document-centric side (left) and the visualization-centric side (right).

*sualization* focuses on documents that includes text but also attributes and metadata. An example for a document is a report. A report has attributes such as format (pdf) and typeface (Helvetica) and it has metadata such as author and last edit. In Document Visualization, researchers are either interested in representing the content of a single document, for example for the purpose of a summary, or they are interested in discovering the relationships among the documents of a corpus.

The embedding of word-scale visualizations into text is a first step towards bridging the gap between a *document-centric analysis* on one side and a *visualization-centric analysis* on the other side. These two sides are illustrated in Figure 1.5. In a document-centric analysis, the text is the main object of analysis. One representative of document-centric analysis is text as a substrate with embedded small-scale visualizations. In a visualization-centric analysis, the visualizations are separate from the text. An interesting example and counterpart of embedded small-scale visualizations is Jigsaw by Stasko et al. [65, 137]. Jigsaw visually represents documents and entities using multiple coordinated views with the aim to support sense-making tasks (Figure 1.6). With entities Stasko et al. refers to people, places, dates, organizations among others, but also includes higher level concepts like time and money. While the text is still present for example in a separate view in Jigsaw, it can still contain annotations such as highlights but usually not more evolved visualizations. Thus, the text has a secondary role. The analysis on the visualization-centric side is about the entities and how these are connected to each other.

The data used as an input to each of these visualizations or systems on both sides in Figure 1.5 is document data. Document data can be raw source text or data that originates from one or multiple text processing algorithms run on the source text, but also document

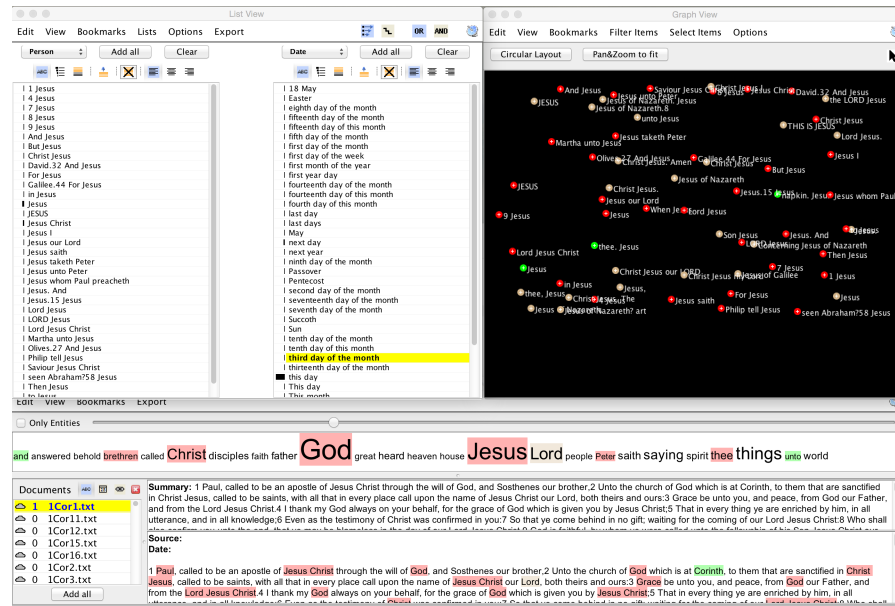


Figure 1.6: Three of the multiple views of Jigsaw: the list view, the graph view and the document view. *Image credit: [65, 137].*



attributes and document metadata such as who changed text in this document or when was the last edit.

In this dissertation I do not concentrate on one specific application domain of document data. I only explore a small part of this broad space between document-centric and visualization-centric analysis where document data plays a major role. Chapter 2 goes into more detail about where different visualizations and systems fit into this space. In Chapter 6 we suggest how to transition from a document-centric view to a visualization-centric view using interaction.

## 1.2 RESEARCH SCOPE

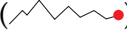


The topic of this dissertation resides in *Information Visualization* (InfoVis). InfoVis is the field of research that deals with abstract data-driven graphics or visualizations. It studies and designs interactive visual representations with the goal to amplify human cognition [21].

The space of visualization-centric analysis based on document data has already been well documented. However, the space of document-centric analysis has been relatively unexplored. In this space, I concentrate on small-scale graphics embedded in text which we call *word-scale graphics*. More specifically I am interested in the data-driven versions of these word-scale graphics—*word-scale visualizations*. We define word-scale visualizations as graphical encodings of data that are roughly word-size (ranging from the size of a single character to several words) that are placed in context with text. Word-scale visualizations can help support many forms of written discourse and can be

used in text books, manuals, notes, blog posts, reports, stories, or poems. As graphical supplements to written text, word-scale visualizations can be used to emphasize certain elements of a document (e. g. a word or a sentence), or to provide additional information; they can enable e. g. graphical comparisons, support complex relationships visually, or simply highlight quantities in the text to facilitate understanding. For example a small stock chart can be embedded next to the name of a company to provide additional information about its past performance (Apple  104.39). Hence, a reader can check the stock trends of that company. Another example is to add game statistics next to the name of the soccer teams that are qualified for the last sixteen at the Euro to allow simple comparisons between ball possession (Switzerland : Poland ) or more elaborate statistics. In contrast to larger graphics and visualizations, word-scale visualizations allow the reader to remain focused on the text, as they are set within the text, rather than alongside it. As such, word-scale visualizations have the potential to deeply change the form and meaning of a written document, and designing and integrating them requires important editorial and rhetorical decisions.

### 1.3 THESIS PROBLEM AND APPROACH

This dissertation investigates different ways on how embedding word-scale visualizations can complement written text documents. It provides the foundation and inspiration to develop ways to create illustrative, and interactive texts with the aid of small embedded visualizations. To quote Bret Victor [153] “People currently think of text as *information to be consumed*. I want text to be an *environment to think in*.” This dissertation is a step in that direction.

There have been previous examples for integrating small visualizations into text such as sparklines [148] () , Gestaltlines [16] () or SportLines [118] () , however they only scratched the surface of what could be. One possible explanation is that when Tufte coined the term sparklines, today’s ease-of-use of the internet as a source of large amounts of text and data but also as an interactive environment was not yet ready. Nowadays, almost everybody that has some programming proficiency can create interactive environments around resources provided through the internet such as text. Some of that has been made possible by for example the vast ecosystem of libraries around the language JavaScript. Moreover, the high-resolution displays currently in use were not yet available to provide the necessary screen resolution to have embedded word-scale visualizations.

The success of word-scale visualizations hinges on a mix of various potential research directions that need to be explored. These research directions are the design and the creation of small-scale visualizations (Information Visualization and Design), the study of their usability in

context with text and their readability in terms of understanding and perception (Human-Computer Interaction), but also the understanding of the effects of the integration of these visualizations on the text and on the reader, and the exploration of novel ways on how readers and analysts can use these sorts of visualizations (applications, software engineering).

If we join all these reasons with the scarce previous research on small embedded visualizations and the space constraints which calls for creative and well-thought decisions, then it is not a surprise that there are not more examples out in the wild.

In order to bring word-scale visualizations into newspaper text, blog posts or reports, I concentrated on investigating a specific set of research questions. I do not attempt to describe and solve all relevant questions around small, contextual and data-driven graphics. Instead my research strategy consisted of identifying and defining important aspects and relevant research directions towards making the integration of text and small-scale contextual visualizations as fluid as writing text and as usable as reading text.

In this dissertation each core chapter (Chapter 4 to Chapter 9) deals with one main research question (RQ1 to RQ6). The details about the study methods used are outlined in each respective chapter.

**RQ1:** We know very little about why and how a writer, designer or editor would want to integrate word-scale visualizations into his or her text. If these visualizations can be useful, how can their creation be supported and maybe even promoted? (Chapter 4)

**RQ2:** There is plenty of empty space available around and between written text. Yet, there has not been much research about what suitable placements would be good for small contextual visualizations. Where can they be placed and what are the implications of these placement options? (Chapter 5)

**RQ3:** Small contextual visualizations are usually static. What are the possibilities for using interaction and what could these interactions be used for? In addition, how could interaction support a more active reading and support an analyst's tasks such as sensemaking? (Chapter 6)

**RQ4:** What are the implications of embedding word-scale visualizations in text on reading behavior and how can they be assessed? If there are negative implications on the reading behavior, are there ways to minimize them? (Chapter 7)

**RQ5:** To make small contextual visualizations more wide spread for example designers need adapted authoring tools. What are the building blocks for such tools? (Chapter 8)

**RQ6:** What are potential novel applications for these small contextual visualizations? (Chapter 9)

## 1.4 MAIN CONTRIBUTIONS

My research builds on knowledge from two research fields: Information Visualization and Document Visualization. The main contribution of this dissertation is an expanded set of design considerations on the integration of small contextual data-driven graphics—called word-scale visualizations—into text documents. In this dissertation I contribute:

- a classification of small-scale contextual visualizations with definitions, encompassing encodings, size and connections to text (entity) (Chapter 3 and Chapter 4)
- a design space and design guidelines for the placement of small contextual visualizations illustrated with examples (Chapter 5)
- design considerations and a set of methods to interact with small contextual visualizations embedded in text (Chapter 6)
- three typical use cases of small contextual visualizations thought as a higher-level approach to generate new designs and applications and a set of main building blocks to design an authoring system for small contextual visualizations (Chapter 4, Chapter 8).

## 1.5 OVERVIEW OF THE DISSERTATION

This dissertation is framed around a set of guiding research questions (Section 1.3) about the integration and embedding of small, contextual, and data-driven graphics (word-scale visualizations) into text. I framed the beginning of this dissertation around the description of where small contextual visualizations fit into the research on Information Visualization and Document Visualization. The questions guiding my research on small contextual visualizations consisted of understanding what they are, assessing their usefulness, investigating their placement options, exploring and understanding the advantages of using interaction in the digital environment of text and visualization, exploring the implications of small contextual visualizations embedded in text for readers, investigating tools for authoring word-scale visualizations and explore potential real world applications.

In this introductory chapter, I discussed the motivation on why I explored the embedding of visualizations into text, the background and provided a high level view on what my research is all about. Moreover, I presented two analysis approaches where word-scale visualizations could be a way to bridge the gap between these two approaches. I organized the remaining chapters of the thesis as follows:

**Chapter 2:** This chapter provides background and related work from Document Visualization.

**Chapter 3:** In this chapter I revisit the different types of small-scale visualizations such as sparklines and micro visualization. Moreover, I show past examples of small-scale visualizations. I also provide examples of word-scale visualizations in the wild.

**Chapter 4:** This chapter discusses an open ended exploratory study with graphic designers focused on the design and function of small contextual visualizations embedded in text documents. We present a systematic overview of the proposed designs and examine how designers used them. Based on the results of the study we propose three typical use-cases for small contextual visualizations as high-level inspiration for authors and designers.

**Chapter 5:** This chapter presents an exploration of the placement of small contextual visualizations within text documents. We provide an illustrated design space of placement options for small contextual visualizations and identify six important variables that control the placement of the visualizations and the level of disruption of the source text. We also contribute a quantitative analysis that highlights the effect of different placements on readability and text disruption. Finally, we use this analysis to propose guidelines to support the design and placement of small contextual visualizations.

**Chapter 6:** This chapter presents a design space for interaction with small contextual visualizations. The design space covers where, when, and how interaction can be triggered for small visualizations embedded in a text document. In addition to this design space, we contribute a set of interaction techniques that allow analysts to dynamically collect, arrange, and compare small contextual visualizations while still maintaining the connection between the visualizations and the text.

**Chapter 7:** This chapter presents a study on how the integration of small contextual visualizations into a sentence affects reading speed and memorization during a brief reading task. In particular, we discuss how different placement types—with their inherent text appearance and layout changes—affect readers.

**Chapter 8:** This chapter discusses implications for the design of authoring tools for small contextual visualizations, and explores how new authoring environments could make it easier for designers to integrate them into documents.

**Chapter 9:** This chapter presents an original application of small contextual visualizations where they can be used to share information from and in personal notes. We provide two examples that illustrate the challenges and implications of using small contextual visualizations to share data in digital notes.

**Chapter 10:** This final chapter summarizes the contributions of this dissertation. In addition, I outline promising future research directions to study more closely the integration of word-scale visualizations into text, to further evaluate the effects of word-scale visualizations on reading behavior, and to investigate how to close the gap between document-centric analysis and visualization-centric analysis.





## THE SPACE OF DOCUMENT VISUALIZATION

---

In the previous chapter I introduced the topic of my dissertation and differentiated between document-centric analysis and visualization-centric analysis to put my research into context. In this chapter, I provide an overview of the research that addresses the visualization-centric analysis side. For that purpose I first provide a discussion on document-centric analysis and visualization-centric analysis and relate them to two analysis approaches from the humanities. Further, I illustrate the diversity of document visualization systems which are on the visualization-centric analysis side. Finally, I provide a high-level overview of how text and visualization can be combined.

The purpose of the following section is not an exhaustive account of visualization systems that visualize documents but an illustration of the diversity of systems visualizing one or multiple documents. Examples were selected to demonstrate this diversity and to show how users have access to the underlying document.

### 2.1 DOCUMENT-CENTRIC AND VISUALIZATION-CENTRIC ANALYSIS

In Section 1.1 I propose the distinction between a document-centric analysis and a visualization-centric analysis (see Figure 2.1). In a document-centric analysis the document is the main object of interest. The analyst needs to be able to read and immerse him or herself into the document during the analysis. This is very similar to what in the humanities is called *close reading*. Close reading [82] is a traditional approach for the analysis of the sources content. It refers to the comprehensive analysis of a text and the study of its main themes and their evolution. This includes the analysis of entities such as persons and events, the relationship between entities, the evolution of these entities, but also the analysis of word and phrase usage, and the text structure and its style.

Jänicke et al. [82] identify color, font size, connections and glyphs as techniques that support close reading. The difference between close reading and document-centric analysis is that in a document-centric analysis I suggest to also use small, contextual and data-driven graphics to provide quantitative information. I hypothesize that word-scale visualizations can be used in a document-centric setting, because these types of visualizations can contribute additional information close to the point of interest without destroying the sequence of words. Word-scale visualizations add additional information that is relevant

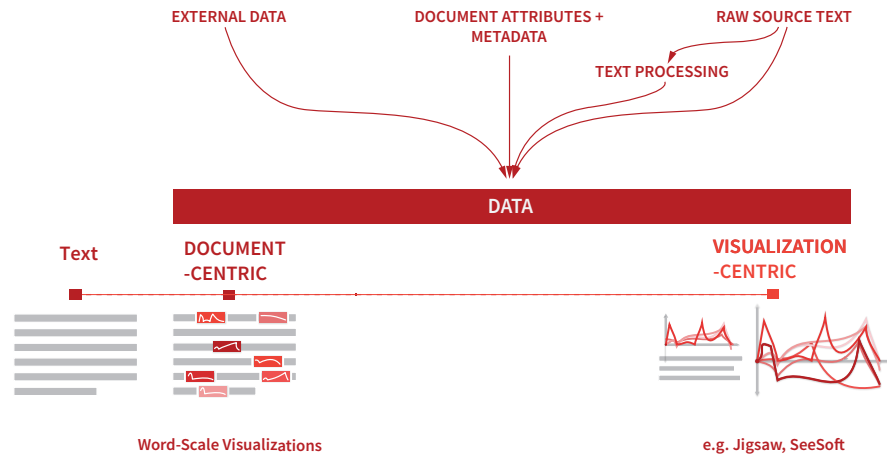


Figure 2.1: The document-centric side (left) and the visualization-centric side (right).

to the analysis of the document, as they can contribute information that is abstracted from the source document or from related documents.

In a visualization-centric analysis the text is secondary and often in a separated view, only called when needed. The visualizations used in a visualization-centric analysis are often an abstraction or summarization of one or multiple documents. These visualizations are the main point of interest for the analyst. It is still about for example analyzing entities, their relationship and their evolution as in a document-centric analysis but using abstract visualizations and interactions, away from the source text. A visualization-centric analysis is similar to what in the humanities Moretti [101] called *distant reading*. Jänicke et al. [82] identify heat maps, tag clouds, timelines, graphs, and a category called miscellaneous as techniques that support distant reading. Miscellaneous seems to include all visualizations they came across but did not mention previously. Visualization-centric analysis is also often referred to as distant reading. I use visualization-centric analysis as a term to emphasize that visualizations are the tools to analyze a single document or a collection of documents.

As argued by Jänicke et al. [82] and others, combining close reading and distant reading is important to the humanities scholar, because they often do want to go back from their visualizations to the source document, or the other way round from the source document to higher level views. Similarly, it should be possible to transition fluidly back and forth between a document-centric analysis and a visualization-centric analysis. Word-scale visualizations are already a step away from a pure document-centric analysis towards a visualization-centric analysis. Thus, I propose word-scale visualizations as a first step towards a transition to the visualization-centric side.

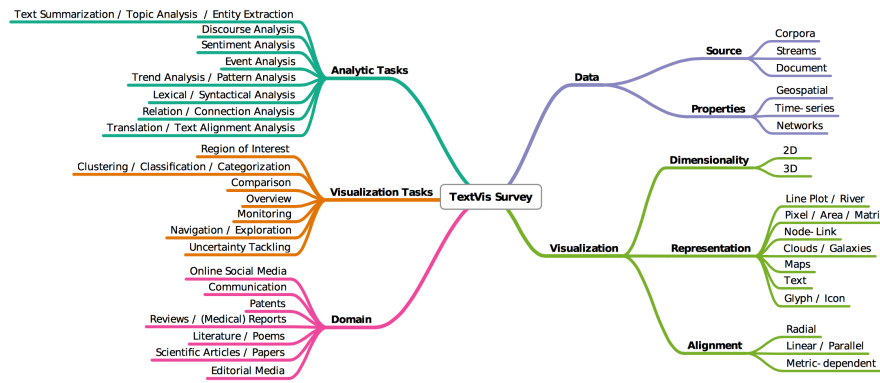


Figure 2.2: Kucher and Kerren's taxonomy of text visualization techniques.  
Image credit: [91].

## 2.2 DOCUMENT VISUALIZATION

There are several surveys in the literature that discuss document visualizations or related problems. I illustrate three proposed taxonomies that are in relation to document visualization in more details. These surveys show a diversity of possible taxonomies.

Gan et al. [55] proposed an overview of document visualizations. They classify document visualization methods into three main groups: single document visualization, document collection visualization, and extended document visualization. They further subdivide the single document visualization category into visualizations that represent: vocabulary, semantic structure and document content. The document collection visualization category is split into visualizations that show: document themes, document core content, changes between different document versions, document relationships, and document similarity. The third category contains text streaming visualizations and search visualizations.

Kucher and Kerren [91] provided an extensive and fine grained classification of text visualization techniques (see Figure 2.2). The taxonomy consists of five main categories: data, visualization, domain, analytic tasks, and visualization tasks. The first three categories are further subdivided into straight forward categories such as source for the data category and dimensionality for the visualization category. The analytic tasks category and the visualization tasks category build the core of their proposed taxonomy. Unfortunately, the categorization into analytic tasks is not well documented, as the reasoning behind this mix of analyses is not clear. The authors are mixing high-level analyses like the analysis of discourse with low-level analyses like the analysis of patterns.

The final taxonomy I describe is by Jänicke et al. [82] and introduced a categorization of text analysis tasks related to the digital humanities (Figure 2.3). They propose five main tasks: text of inter-

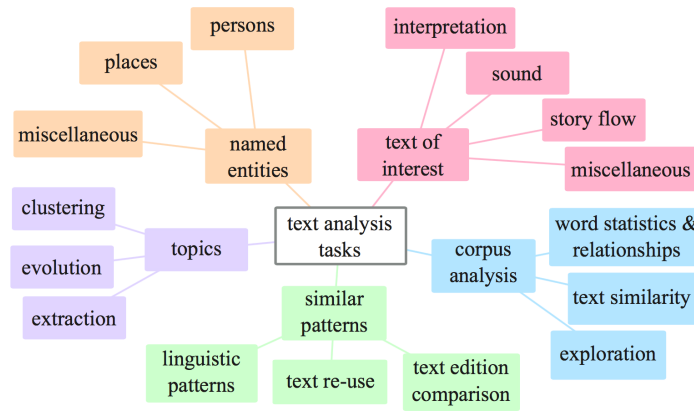


Figure 2.3: Taxonomy of visual text analysis tasks. *Image credit: [82].*

est, corpus analysis, similar patterns, topics, and named entities. The task of *text of interest* groups multiple various tasks that relate to a single literary work. Tasks in that group include the visualization of text interpretations and the analysis of the story flow. *Corpus analysis* includes the analysis of word statistics and the relationships between words. The task of *similar patterns* includes the analysis of linguistic patterns, text re-use and the comparison of text editions. The task called *topics* describes the analysis of topics for example their clustering, the analysis of their evolution, and their extraction from the text. The last task they single out is called *named entities* and includes the extraction of different types of entities.

Similarly, to the overview of document visualization by Gan et al. I separate into single document visualization and document collection visualization. Further, I divide into different low-level analysis tasks. I used the three described taxonomies to inspire my own categorization into tasks.

Before going into more details on document visualization I would like to point to the relation between the two analysis approaches mentioned in the section before—document-centric analysis and visualization-centric analysis—and document visualization. The data used for document-centric analysis stems from the techniques used in document visualization. The word-scale visualizations that form the basis for the document-centric analysis integrates that data close to where it is needed and keeps the text as the central point of interest for the analyst. For the visualization-centric analysis, techniques from the domain of document visualization are used to analyze a document or a collection of documents.



Figure 2.4: Word Tree visualization showing the search term “love the” in context. *Image credit: [160].*

### 2.2.1 Single Document Visualization

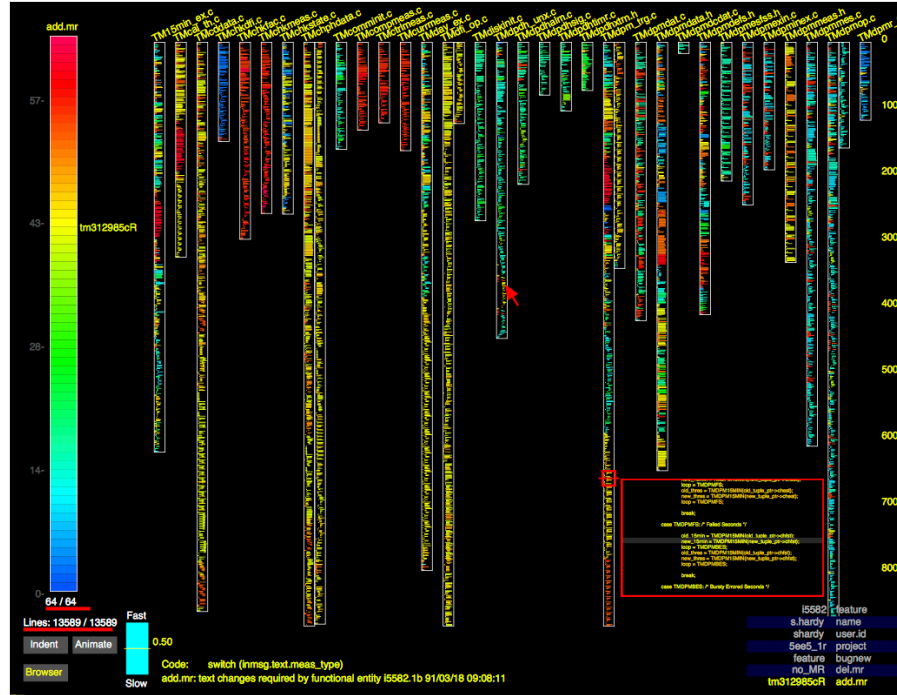
Here, I present single document visualizations that support the following tasks: analysis of words in context, analysis of patterns, and providing an overview.

### 2.2.1.1 Terms in Context

A document visualization that provides terms in context are Word Trees. The Word Tree visualization is similar to the concept of keyword-in-context (KWIC) which has been described by Hearst [72]. KWIC is used in Information Retrieval and refers to parts of sentences that show where the search terms are used within the document. Thus, KWIC does not only show in which document the search term appears but also reveals the context of where they appear.

Similarly to KWIC, the Word Tree document visualization by Wattenberg and Viégas [160] provides the visualization of search terms in context of the words that are following the search term (see Figure 2.4). Word Trees provide a simple way of exploring repetitive context as they place a tree structure over the words that follow a search term. The tree structure arranges the words following the search term spatially. Through interaction (clicking) the root node of the tree structure can be changed, thus, supporting navigation through the document. The authors also described how Word Trees have been used in Many Eyes [155]. They observed that Word Trees were also used to visualize collections of text.

Wattenberg and Viégas’s initial description of Word Trees did not mention access to the source text. Further implementations now provide access to the source text. One that provides access to the source text is the implementation of the Word Tree visualization by Davies [37].

Figure 2.5: SeeSoft visualizing a document. *Image credit: [45].*

#### 2.2.1.2 Pattern Analysis

Another type of analysis that is often supported by single document visualizations is pattern analysis. Pattern analysis refers to finding a set of data that follows a characteristic or a recognizable form.

A document visualization that supports the analysis of patterns is Eick et al.'s [45] SeeSoft. Seesoft support the quick recognition of patterns by providing an overview of line-based statistics. Initially, it was build to visualize source code. SeeSoft reduces the size of source code by transforming each line of code to a thin rectangle (see Figure 2.5). Each line can then be colored according to a statistic of interest such as location of variables or loops. Another interesting statistic that can be shown is which programmer wrote which line. A sort of brushing and linking is also implemented as clicking on a given colored line highlights all other lines with the same color. SeeSoft also provides access to the underlying source code. Clicking on a button a magnifying lens appears. By using the lens to hover over one of the thin rectangles the corresponding source code line is displayed in context of the neighboring lines. The authors also suggest that this visualization system can be used to visualize documents or a collection of documents.

#### 2.2.1.3 Overview of a Document

The summary of a document provides information about what is inside a document. I illustrate three different approaches: TextArc by Pa-

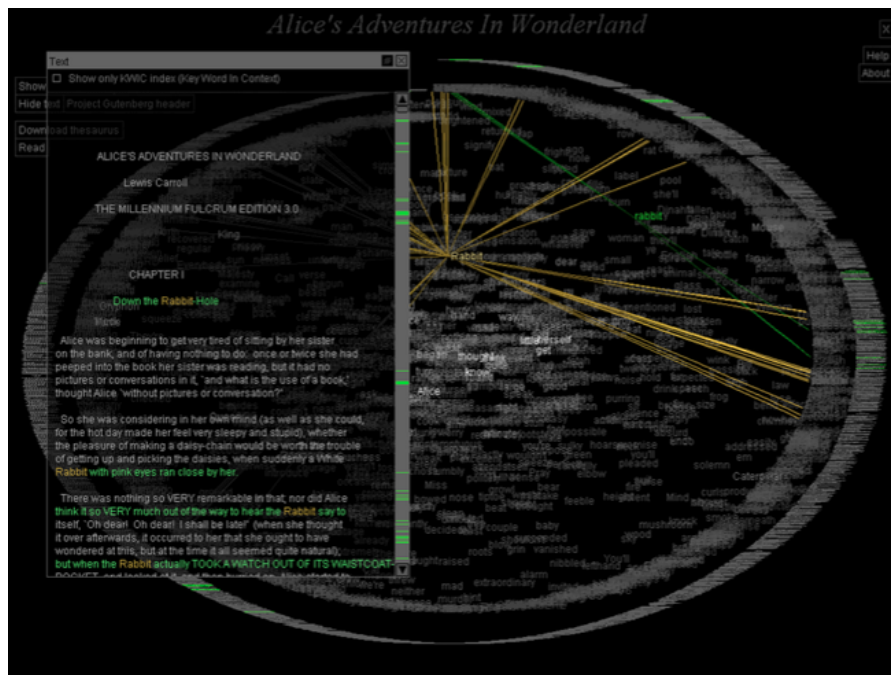


Figure 2.6: TextArc visualizing “Alice’s Adventures In Wonderland”. *Image credit: [110].*

ley which displays the content of a document through the frequency and distribution of words, DocuBurst by Collins et al. that combines the frequency of words and their semantic relation, and Document Cards by Strobel et al. which summarizes the content of a document by extracting key terms and images to provide an overview of the content.

Paley’s TextArc [110] visualizes raw text by arranging each line of the text around a large ellipsis (see Figure 2.6). The same words are then added a second time inside this large ellipsis. The positioning of the words inside the ellipsis depends on where in the text the word appears. If a word appears often in the text it appears closer to the centre of the ellipsis. TextArc is a visualization that provides information on the distribution of the words in a document. The source text can be accessed in a separate window through brushing and linking.

Collins et al.’s DocuBurst [34] visualizes an overview of the document’s content by combining word frequency and semantic relation based on a human-created lexical structure (see Figure 2.7). The source document can be accessed in a separate view linked to the visualization for details on demand. DocuBurst can also be used to compare documents. One way of reading the visualization is to follow a straight line from the node in the centre to a node on the outer circle. Each of the visited node is semantically connected to the previous node through the IS-A relation. The further away from the center the node is the more specific the classification gets. DocuBurst provides the source text in a separate window.



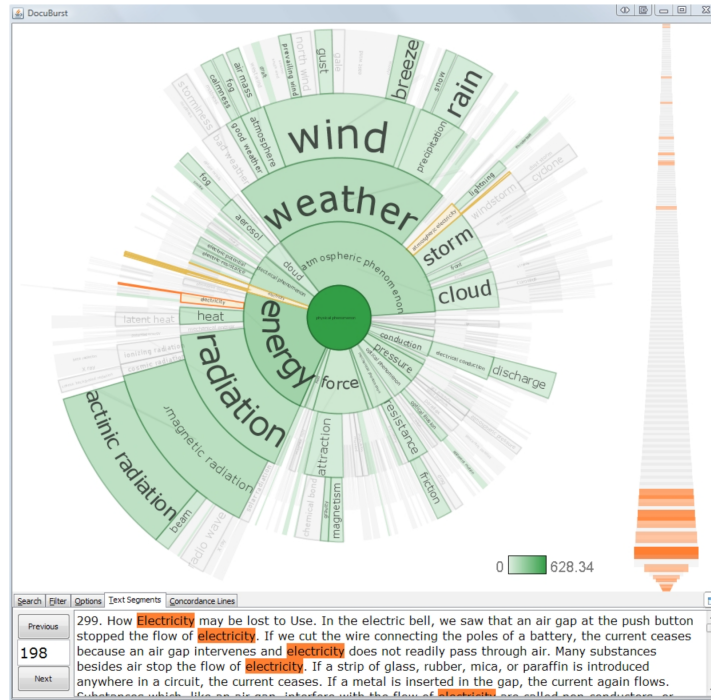


Figure 2.7: DocuBurst rooted at the node “physical phenomenon”, and a search query for “electricity” has been performed. The nodes that match are colored in light yellow. *Image credit: [34].*

Strobelt et al.’s Document Cards [141] provide a succinct summary of what is inside a document by providing a mix of images and keywords from the document (see Figure 2.8). The authors propose methods to extract key images and keywords. To extract the key images the documents are passed through an algorithm based on semi-semantic image weighting and image classification. The extraction of keywords of a document is using a text mining approach. The authors implemented a layout algorithm to group and position the images and keywords into the document card. This algorithm first iteratively places images, based on pre-calculated weights, until there is no space left to fit an image, the remaining space is filled with the keywords. On the side of each document card there are little rectangles with numbers. These numbers represent the pages of the document. Clicking on one of these numbered rectangle will open the corresponding page in the document.

### 2.2.2 Document Collection Visualization

Here, I present document collection visualizations that support the following tasks: showing terms in context, showing patterns between documents, providing an overview, and showing the variation of topics over time.





Figure 2.8: The view of a Document Card with each rectangle on the right side representing a page of the document. *Image credit: [141].*

#### 2.2.2.1 Terms in Context

Similarly, to single document visualizations, search terms can be visualized in context of multiple documents. One such visualization is the TileBars by Hearst [71]. TileBars are an iconic representation of the document (see Figure 2.9). The document is represented as a rectangle that contains little squares. Each little square represents at TextTile. A TextTile [70] is a multi-paragraph segment of a document. These segments are generated by using an algorithm called TextTiling developed by Hearst [70]. It partitions a text according to term repetition patterns. TileBars provide an indication of the length of a document, the frequency of the term in the document and its distribution within the document and across documents. The darker a TextTile, the more frequent the search term is present in that specific TextTile. The text can be accessed by selecting a document.

#### 2.2.2.2 Pattern Analysis

Similarly to SeeSoft in its visual representation is the Compus system by Fekete and Dufournaud [47]. In contrast to Seesoft, Compus visualizes entire collections of structured documents such as XML documents (see Figure 2.10). The position and size of each XML element is visualized using color and a space filling approach. Each document is represented as a set of colored segments. Each element type which for example is defined by the tag name gets its own color. The Compus system supports the comparison of document size and overall structure. It also helps the user to recognize correlations and differences between documents. It also has a very powerful sorting mechanism to sort documents according to different attributes such

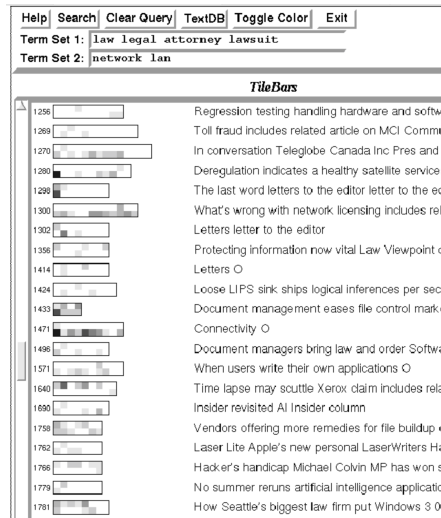


Figure 2.9: On the left the TileBars can be seen, where as on the right the titles of the document are shown. *Image credit: [71].*

as the number of colored elements. The original source text can be accessed over a tab.

#### 2.2.2.3 Overview of a Collection of Document

A visualization that provides an overview of a collection of documents but also of single documents are tag clouds (see Figure 2.11). They have their origin in an experiment by social psychologist Milgram [99] and were later popularized by Many Eyes [155] and Viégas and Wattenberg [154]. Tag clouds map word frequency to font size. Parallel tag clouds by Collins et al.'s [35] is a visualization system that uses the concept of tag clouds and parallel coordinates. It visualizes a document collection and provides access to the source document in a separate view.

#### 2.2.2.4 Variation of Topics

ThemeRiver by Havre et al. [69] is a visualization that shows variation of topics in document collections over time (see Figure 2.12). It uses the metaphor of rivers to provide information on topics at different points in time. The width of the river represents the strength or degree of use of the topic. There is no connection back to the original source document.

### 2.3 COMBINING TEXT AND GRAPHICS

Data visualizations are usually associated with text in two ways either embedded into the text as in Figure 2.13a and Figure 2.13b or as visualizations external to the text like in Figure 2.13c. External visualizations are for example placed in separate windows accessed through

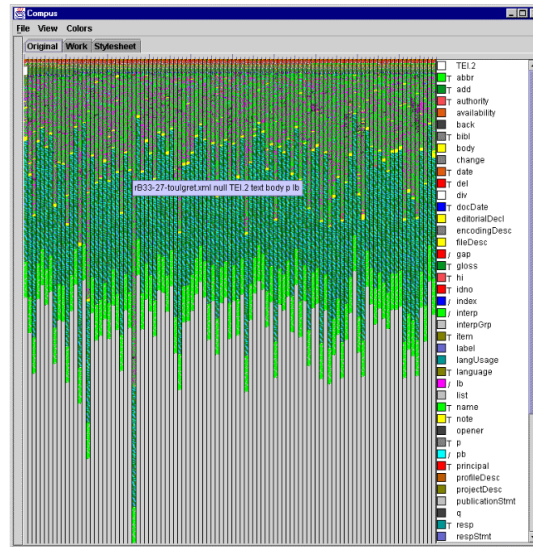


Figure 2.10: The view of a collection of structured documents represented by the Compus system. *Image credit: [47].*

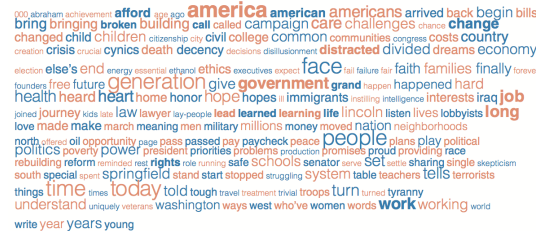


Figure 2.11: One-word tag cloud of a document. *Image credit: [154].*

a hyperlink, or details-on-demand but multiple other possibilities exist. Connecting the visualization and the part in the text that refers to it is one of the main challenges. This has also been suggested by Colin Ware [159], where he says: “If both words and images are used, methods for linking them must be selected.” One of the most often used methods to link text and visualization are text references such as “see Figure x”, which is also use in this dissertation. Other methods are to use graphical techniques such as lines (e. g. Steinberger et al. [138]’s visual links) or color, or to use interaction as suggested by Faraday and Sutcliffe [46]. Faraday and Sutcliffe argue for introducing the notion of “contact points” or coreferences between text and a representation to effectively combine two types of media. Kwon et al. [94] expressed that often there is a lack of guidance for the reader on how to interpret and manipulate embedded visualizations which sometimes are interactive. A small in-line graphic tells the reader that a guided action can be invoked by interacting with the text snippet (clicking on it). Moreover, the reader is often left alone to connect text and visualization, even for simple static bar charts as illustrated by Kong et al. [88]. Nevertheless, separate text and visualization forces

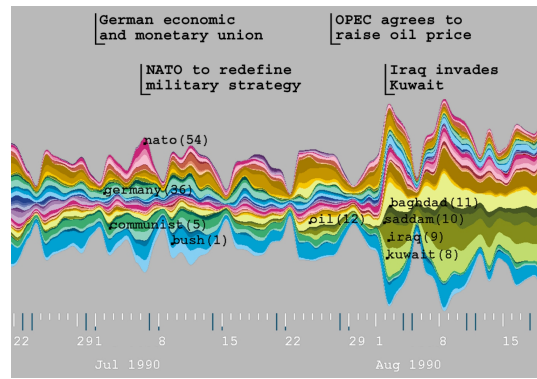


Figure 2.12: Each topic is visualized using a different color, the wider the current the more the topic is used. *Image credit: [69].*

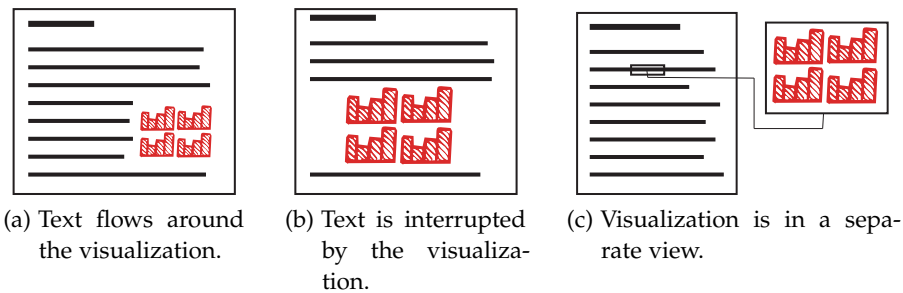


Figure 2.13: Three different associations of text and visualizations.

the reader to split attention between disparate sources of information and only their combination lead to the full picture (split-attention effect) as is explained by Chandler and Sweller [24]. Word-scale visualizations are a way of reducing the space between text and visualization, as they are placed close to the text that refers to its data.

## 2.4 SUMMARY

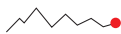
In this chapter I first described the two analysis approaches: document-centric analysis and visualization-centric analysis. Then, I related these two approaches back to what in the humanities is called close reading and distant reading. Further, I employ different examples of Document Visualization to illustrate the visualization-centric side and provide an impression of what sorts of visualizations can be used for a visualization-centric analysis. Finally, I conclude this chapter by providing related work on the combination of text and graphics.

In this chapter I did not discuss related work for visualizations that can be used on the document-centric analysis side. Therefore, in the next chapter I focus on small-scale and data-driven graphics which can be used to support a document-centric analysis.

## SMALL-SCALE AND DATA-DRIVEN GRAPHICS

In the previous chapter I provided an overview of the visualization-centric analysis side by illustrating the space of document visualizations, but I did not discuss what sorts of visualizations can be used for a document-centric analysis. Thus, in this chapter I focus on visualizations for the document-centric analysis side and provide the context for small-scale and data-driven graphics, the actual topic of my dissertation. To provide a solid foundation for my research, I report on past examples of small, contextual and data-driven graphics, highlight existing designs, and illustrate the space of non-data-driven word-scale graphics. Finally I conclude with examples of word-scale visualizations used in the wild.

## 3.1 SMALL CONTEXTUAL VISUALIZATIONS

Edward Tufte’s *sparklines* [148] are probably the best-known example of data-driven word-scale graphics. Tufte describes sparklines as *small, intense, simple, word-sized graphics with typographic resolution*. He provides several different designs, including line charts  depicting financial or health data and win/loss sparklines (see Figure 3.1) that show game result data. Tufte further suggest to employ sparklines wherever words and numbers are found, including in a sentence, a table, a headline, a map, a spreadsheet or a graphic. Small “cell-sized” sparklines modeled after Tufte’s examples have been available in Microsoft Excel since 2010 [98].

Expanding on Tufte’s definition, my co-authors and I have previously proposed the term *word-scale visualizations* [60] to define a wider variety of small embedded data-driven graphics. Typically, word-scale visualizations are open to greater variations in size than sparklines, ranging from e. g. as small as a single character to as large as several words. Word-scale visualizations are placed in context with text. In this previous work, my co-authors and I also introduced the notion of an *entity*—a specific word or sequence of words to which a word-scale visualization can be associated.






Figure 3.1: Win/loss sparkline generated using Schwartz’s random sparkline generator [132]. All four and more game winning streaks are shown in red.

Recently Parnow and Dörk introduced the term *micro visualization* [111, 112] to describe data-driven word-scale graphics that are *small in physical space and represent few data dimensions or points*. This definition reduces the concept of word-scale visualizations to those with few data dimensions or points. The authors further discuss micro visualizations in the context of typography and typographic data visualizations such as FatFonts [104]. Similar to this dissertation and previous work [60], the authors describe characteristics of micro visualizations such as their placement in or next to text and their connection to data.


While all three existing definitions vary slightly, the authors of the respective works all postulate that small data-driven graphics can “enhance” text content in particular through the addition of detail in context that could not otherwise be reported. Parnow and Dörk even postulate that micro visualizations can enhance the comprehension of text by making it more accessible, comprehensible, and memorable. However, no studies on these benefits have been proposed, yet.



### 3.1.1 Word-Scale Visualization Designs


While evaluations on the benefits of word-scale visualizations are still relatively rare, researchers have previously proposed and tested some specific word-scale visualization designs. *SportLines* [118] () , for example, show phases of player actions across a soccer field. The authors also conducted a user study where participants ranked five different sizes and two different design variations of SportLines. The first design variations consisted of a view onto the soccer field from the top and for the second design variation they used a 3D perspective view of the field. They also compared five different sizes (80×60, 65×45, 50×37, 35×26 and 20×15px) relatively as they did not control for screen resolution which resulted in different physical sizes depending on the DPI of the study participant’s screen. To compare the different sizes they did not embed the SportLines into text. The smallest size was the least preferred across all designs by the study participants.



Brandes and Nick introduced the concept of *Gestaltlines* [15] which are sparkline-sized data-intensive graphics designed to convey a general *gestalt* or overall form that resembles patterns in the data. They applied Gestaltlines to time-varying network data and showed examples of their embedding into text. However, their work mostly focused on their use in small multiples. Brandes et al. [16] re-introduced Gestaltlines as a general visualization technique to visualize multivariate data sequences. The authors tested one Gestaltline design () and found that untrained readers could quickly detect holistic patterns, outliers, and breaks in the data. The used design was based on multiple Gestalt laws to show patterns in the data. For example, the design included the law of symmetry () to show changes




in the data, or the law of similarity and the law of proximity to show local trends (.

Greenhill et al.'s *Separation Plots* [67] () can help readers assess the predictive power of models with binary outcomes. A dark red stripe stands for the actual outcome of one observation. All stripes are ordered by the fitted value the model produced and the fitted values of each observation are plotted as a black line over the plot. A perfect model would correctly predict all actual outcomes of an event and produce a plot with all dark red stripes clustered to the right side (as in the in-line example). Greenhill et al. discussed several examples and potential problems with the interpretation of their plots but did not conduct a user study. The authors also discussed the limitations of their Separation Plots, for example that if the sample size observed is too large it is impossible to distinguish between events (dark red stripes) and nonevents (light red stripes). For these cases they propose to group the probabilities of the events into discrete bands (.

While the three word-scale visualization designs discussed above are meant to be used within a text document, other designs have been proposed for non-continuous documents such as tag clouds or tables. For example, *SparkClouds* [95] () convey temporal trends for words in a tag cloud by attaching small line charts to each word. In their original paper, Lee et al. compared SparkClouds to three other trend visualizations and found that participants could read SparkClouds comparatively quickly across several tasks and preferred them over other representations.

Nguyen et al. [106] follow a similar approach and propose several tag designs that use the background of a word to convey temporal information, while placing tags on a map to further convey geospatial relationships, for example a calendar table to recognize patterns () or color-coded segments to represent the tag frequency over time (.

Beck et al. [4] proposed *word-sized eye tracking visualizations* (e.g. ) which are meant to be used alongside think-aloud transcripts in a tabular view but can also be used within text. They propose multiple designs of their word-sized eye tracking visualizations showing different data. The authors re-analyzed data from a previous small user study using the new graphics and showed how their observations could be reported using word-sized eye tracking visualizations in-line.

Another type of small data-driven graphics are data glyphs which are data-driven visual marks that “make use of different visual channels to encode multiple data dimensions” [53]. While not always the case, glyphs are often small-scale and used in small-multiple display settings. Several overview papers and chapters have been published on data glyphs [9, 31, 53, 126, 157, 158] that focus on different charac-

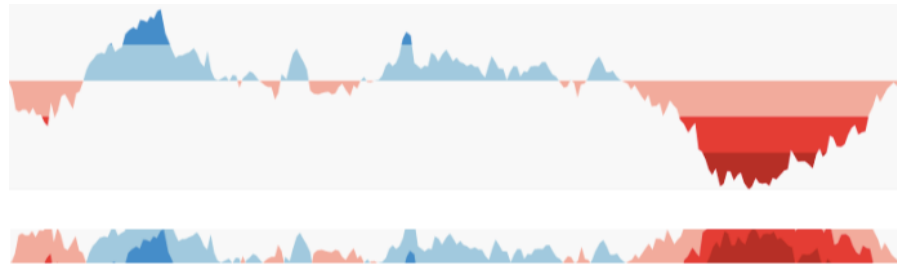


Figure 3.2: HorizonGraph [122]. Top: original chart, with colored bands. Bottom: resulting HorizonGraph. *Image credit: [122].*

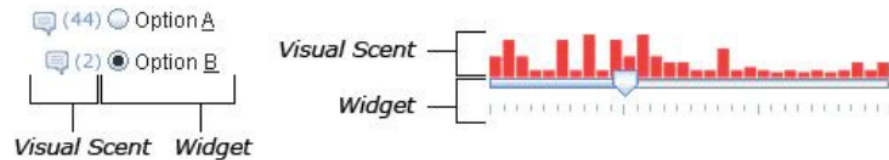






Figure 3.3: Visual scent providing information on the widget [162]. Left: visual scent provides number of comments. Right: visual scent provide data totals per tick. *Image credit: [162].*

teristics of their design space. Some well known data glyphs are the Chernoff faces [27] () , the Autoglyph [5] () , the Starglyph [134] () or the Stick Figures [119] () . The work of Fuchs et al. [53] is informative for this dissertation as it summarizes user studies on data glyphs and points out the many open questions regarding the performance of representations drawn at small sizes—that similarly apply to word-scale visualizations. Open research questions according to Fuchs et al. include the evaluation of glyphs on exploration tasks to gather insights from unknown datasets, the evaluation of glyphs using different layouts such as treemaps and not only in small multiples settings, and using quantitative studies to compare different glyph designs such as pie chart glyphs where the number of dimensions is a factor.


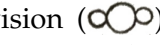

Finally, there are a number of small visualizations besides sparklines and glyphs which share similar size constraints, but have not typically been embedded with text. Horizon graphs [49, 122, 128] are one such example (see Figure 3.2). Horizon graphs use a split-space technique to represent a large number of time series in a single compact view with constrained vertical height. The height of the visualization can be kept very small by using an approach called “two-tone pseudo-coloring” [128].

Scented Widgets [162]—small data graphics embedded in GUI controls to support social navigation—showcase another opportunity for embedding small visualizations into other interfaces (see Figure 3.3). Similarly, word-scale visualizations can be used to provide information scent cues within text documents—for example by showing visitation data for hyperlinks on web pages.





### 3.1.2 Non-Data-Driven Word-Scale Graphics


Here, I illustrate five different non-data-driven word-scale graphics. The purpose of this section is not an exhaustive account of non-data-driven word-scale graphics but serves as an illustration of the diversity of existing word-scale graphics.

Small pictogrammatic, iconic, and symbolic graphics have been used alongside text for centuries. Galilei's drawings of Jupiter and its moons ( [54]) are one famous example. Equally well known is how Galilei illustrated the shape of Saturn when seen through a flawless telescope and perfect vision ( [147] in contrast to how Galilei saw it with suboptimal vision and a suboptimal telescope ( [147].

Another example is the rebus (see Figure 3.4). A rebus is a sequence of images and letters created by exchanging words or parts of words in a sentence with images. It is an allusional device that was used in the Middle Ages to denote surnames.

Emoticons (☺☺) are a more recent example of iconic representations that are incredibly common in computer mediated communication. Emoticons are usually placed in-line. Dresner and Herring [43] describe emoticons' rich set of functions within written communication that goes well beyond the expression of emotion. They categorize emoticons into three broad functional groups: emotional meanings, nonemotional meanings, and illocutionary force indicators. Emotional meanings are directly mapped to facial expressions such as happy or sad. For nonemotional meanings, the meaning is conventionally mapped onto facial expressions, such as a wink that indicates joking intent. The last category is illocutionary force indicators which have no conventional mapping to a facial expression such as a smile as a signal for downgrading a complaint to a simple assertion.

Children's books are another place where iconic graphics are sometimes used to replace words in a sentence to help beginning readers as in "Benny's Big Bubble" by J. O'Connor [108]. For example whenever the word baby would appear in the text it is replaced by an image of a baby , similarly instead of using the word hat an image representing a hat  is used.

Designers have also used symbols or icons as word-scale graphics to tag words in text to provide meta-information.  Verjat [152], for example, uses a human profile icon to make names stand out on his online CV.

Another example of extensive word-scale graphics use, is the automobile travel guidebook from 1909-1911, which was published by the Automobile Club of America [41]. In that guidebook, various word-scale graphics are used to explain how to travel from one destination to another. The word-scale graphics provide information about landmarks such as bridges, turns, and train crossings (see Figure 3.5).



Figure 3.4: An “escort card” that is transcribed as “May I see you home my dear?” from 1865. Image in the public domain.

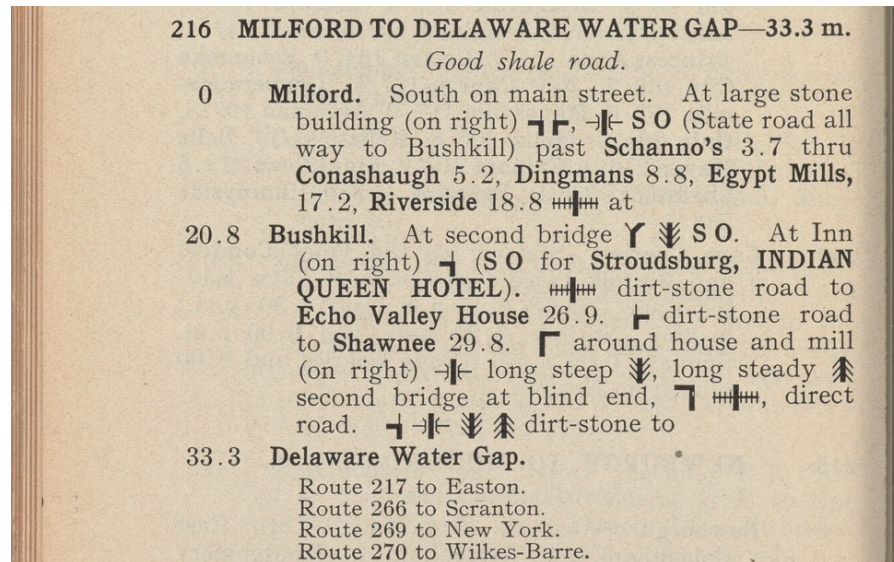


Figure 3.5: This is an excerpt from an automobile travel guidebook from 1909-1911 by the Automobile Club of America. Here, word-scale graphics are used to provide detailed landmarks for driving orientation. *Image credit: [41].*

### 3.2 WORD-SCALE VISUALIZATIONS IN THE WILD

I only found a couple of instantiations of word-scale visualization designs embedded in text. The most interesting case of embedded word-scale visualizations appeared in an article [19] by Bycoffe from FiveThirtyEight about the endorsement process for the primary in the United States. FiveThirtyEight is an online media news site specialized in data journalism. This example is interesting because it does not only show the trend of the endorsements as little word-scale visualizations in the text but the word-scale visualizations also serve as an index into the larger visualizations in the article. These larger visualization are interactive, the word-scale visualizations are not.

While each election cycle is different, nomination races since 1980 have fallen into two broad categories. In some — like the Republican contest of 1996 ↗ and both the Democratic ↘ and Republican ↗ races of 2000 — party elites rallied around a candidate well before the Iowa caucuses, and these candidates went on to win their nominations with relative ease. In others, like the Democratic contests of 1988 ↘ and 2004 ↘, there was less of an endorsement consensus, with party officials taking more of a wait-and-see approach until voters weighed in. Party elites coalesced around a candidate only after the Iowa caucuses and the New Hampshire primary.

Figure 3.6: Sparklines function as references back to the larger interactive visualizations in the article. *Image credit: [19].*



Figure 3.7: Example of a Sparktweet created with [81]

Word-scale visualizations have also been integrated into Twitter feeds. These word-scale visualizations are called Sparktweet and one such example can be seen in Figure 3.7. These Sparktweets can be created with a tool by James [81]. The original tool to create Sparktweets was created by Zach Seward, however his tool is not accessible anymore.

Another field where word-scale visualizations have appeared is in clinical research. For example Johnson et al. [84] suggest to use word-scale visualizations for comparing patterns of intracoronary pressure data. Furthermore, the authors suggest to employ them in medical records and published scientific data. Their argument is that they provide a concise representation of the data. However, the authors do not show embeddings of their word-scale visualizations in-line in text.

The current lack of integration of word-scale visualizations is probably driven in large part by the lack of appropriate tools to design and add such visualizations in-line. Current keyboard-centric authoring tools make it easy to write text, but it is hard to expressively create and integrate visualizations at the level of individual words.

### 3.3 SUMMARY


In this chapter I identified three similar definitions of data-driven word-scale graphics: sparklines, word-scale visualizations, and micro visualizations. The authors of these definitions all agree and hypothesize that data-driven word-scale graphics can complement text content with for example additional information. Furthermore, I presented a limited set of designs of data-driven word-scale graphics that have been proposed by various authors. However, the authors, did not extensively explore their integration into text.

Examples of word-scale visualizations embedded in text in the wild are rare. There has been some limited examples in data journalism such as the one by FiveThirtyEight where word-scale visualizations show the trend of endorsements for the republicans and democrats. Some examples of sparkline use have appeared in tables or as part of dashboards such as for Twitter analytics [150]. These examples usually follow the sparkline definition and their design often corresponds to a line or bar chart. However, research on the use, function,

and design of small, contextual and data-driven graphics and their application in the wild is still relatively rare. Therefore, as a first step I propose to investigate how word-scale visualizations can be used and what possible designs are.

## USE AND DESIGN OF SMALL CONTEXTUAL VISUALIZATIONS

---

In the two previous chapters I first presented the space of document visualizations followed by an overview of already existing small, contextual, and data-driven graphics that can be embedded into text such as glyphs. In this chapter, I focus on investigating how authors and designers use and design small contextual visualizations such as word-scale visualizations (e. g.  Ca).

I start by describing an open-ended exploratory study with 9 designers. The study resulted in a rich collection of different types of graphics, data provenance, and relationships between text, graphics, and data. Based on this corpus of 200 resulting graphics (see Figure 4.1 for a subset of the designs or Section A.3 for all designs), I present a systematic overview of word-scale visualization designs and examine how designers used them. I also discuss the designers' goals in creating their graphics and characterize how they used small contextual visualizations to present data, add emphasis, and create alternative narratives. From our study, I then derive three typical use-cases for word-scale visualizations, which I propose as high-level inspiration for authors and designers. I end this chapter with a reflection on small contextual visualization in light of our study results and discuss challenges that arose during the study.

The main parts of this chapter are from Goffin et al. [63] which is in submission to TVCG. Hence, the use of “we” in the remainder of this chapter refers to Pascal Goffin, Jeremy Boy, Wesley Willett, and Petra Isenberg.

### 4.1 MOTIVATION

Previous examinations of small-scale visualizations embedded in text have focused mostly on contrasting *data-driven graphics* with larger, more-fully featured data graphics [60, 112, 148, 153]. However, researchers have not explored the broader space of possible designs of word-scale visualizations or potential uses for them. The rhetorical impact of word-scale visualizations remains similarly unexplored, as is the process of authoring and integrating them with text. As a result, most guidelines on *where* and *how* to use word-scale visualizations are based only on a very limited, curated, and strictly data-driven set of examples (such as Tufte's [148]). We argue that these represent a mere fraction of the possibilities offered by word-scale visualizations, and

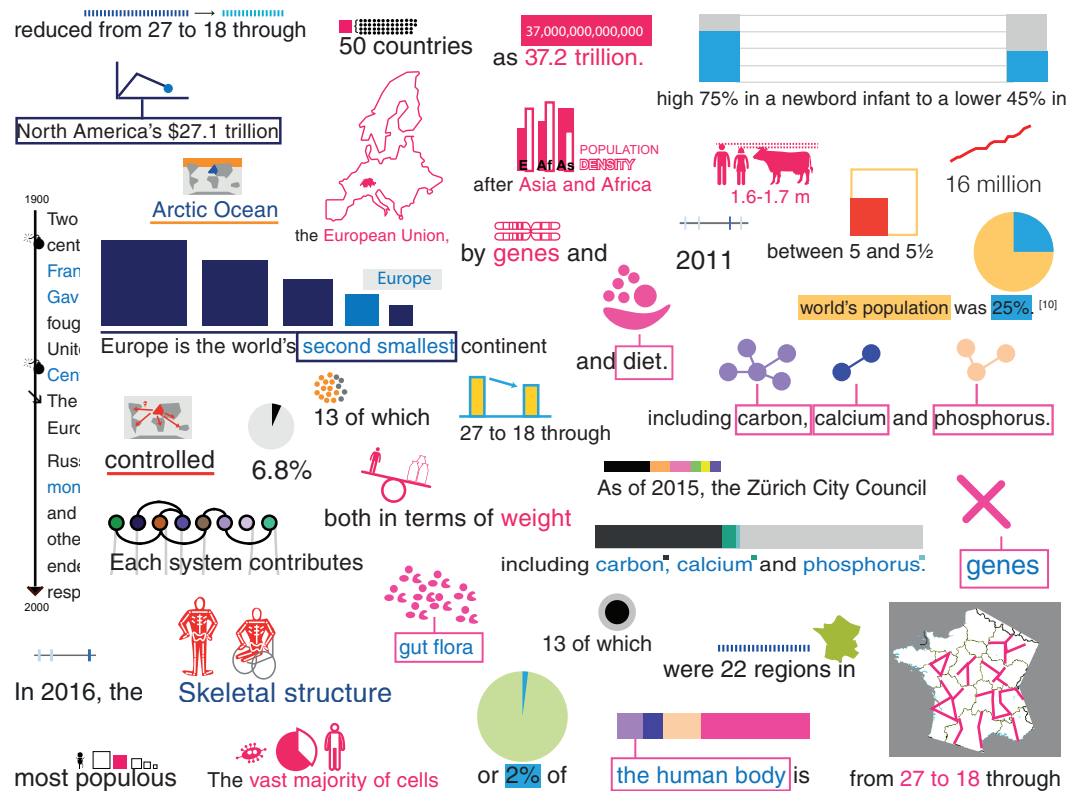


Figure 4.1: Examples of word-scale graphics collected in our study.

that many practical questions regarding their design and use remain unanswered.

To address this issue, we present the results of a study we conducted with nine graphic designers, in which we asked participants to annotate Wikipedia articles with word-scale visualizations of their own design.

## 4.2 STUDY-DESIGNING WORD-SCALE VISUALIZATIONS

To explore this diversity of possible uses and designs of word-scale visualizations, we conducted a qualitative study in which we asked nine professional graphic designers to graphically annotate existing documents. We chose Wikipedia articles as the documents to annotate because they provide a realistic scenario and environment for the integration of word-scale visualizations. We designed the study expressly to elicit new and interesting possible designs for word-scale visualizations and to explore how designers might use them to enhance existing documents and engage readers.

P	Age	Occupation
0	21-30	PhD student working on design and learning
1	21-30	Independent graphic and interaction designer
2	31-40	PhD student working on design fictions
3	21-30	(Data) graphic designer for an online news site
4	31-40	Designer and assistant professor in visualization
5	21-30	PhD student working in disability and media studies
6	21-30	Freelance graphic designer
7	21-30	Graphic designer and application/game developer
8	21-30	Teacher and practitioner in the applied arts

Figure 4.2: Participant details

#### 4.2.1 Participants

We recruited designers (4 male, 5 female) via word-of-mouth. Participants had to be expert users of Adobe Illustrator (*mean* = 8.5 years using Illustrator, *SD* = 2.60) with professional design experience through their work or field of study (*mean* = 6.6 years as a practitioner, *SD* = 3.97). We required expertise in Illustrator, as it allows flexible digital manipulation and generation of text and graphics and, compared to other vector graphics editors, is a relatively standard tool in the design community. Figure 4.2 gives details on the participants' occupations and age ranges. All were graduates or did a post-master at a prominent French design school. We compensated participants with a 20 € Amazon gift certificate.

#### 4.2.2 Task

We provided participants with a Wikipedia article in Adobe Illustrator format. Participants had to read the article carefully and envision how small visualizations embedded in the text could provide a richer reading experience. We then asked them to create small word-scale visualizations and embed them into the document using Illustrator. Participants repeated this task for three articles in total.

#### 4.2.3 Dataset

We selected subsets of three English language Wikipedia articles for participants to annotate. Figure 4.3 shows a high-level overview of the curated articles and the full documents can be found in Appendix A. We chose these articles for their high density of various data types and diversity of sub-topics. Taken together the articles contained examples of the seven data types highlighted by Shneiderman [133], in-



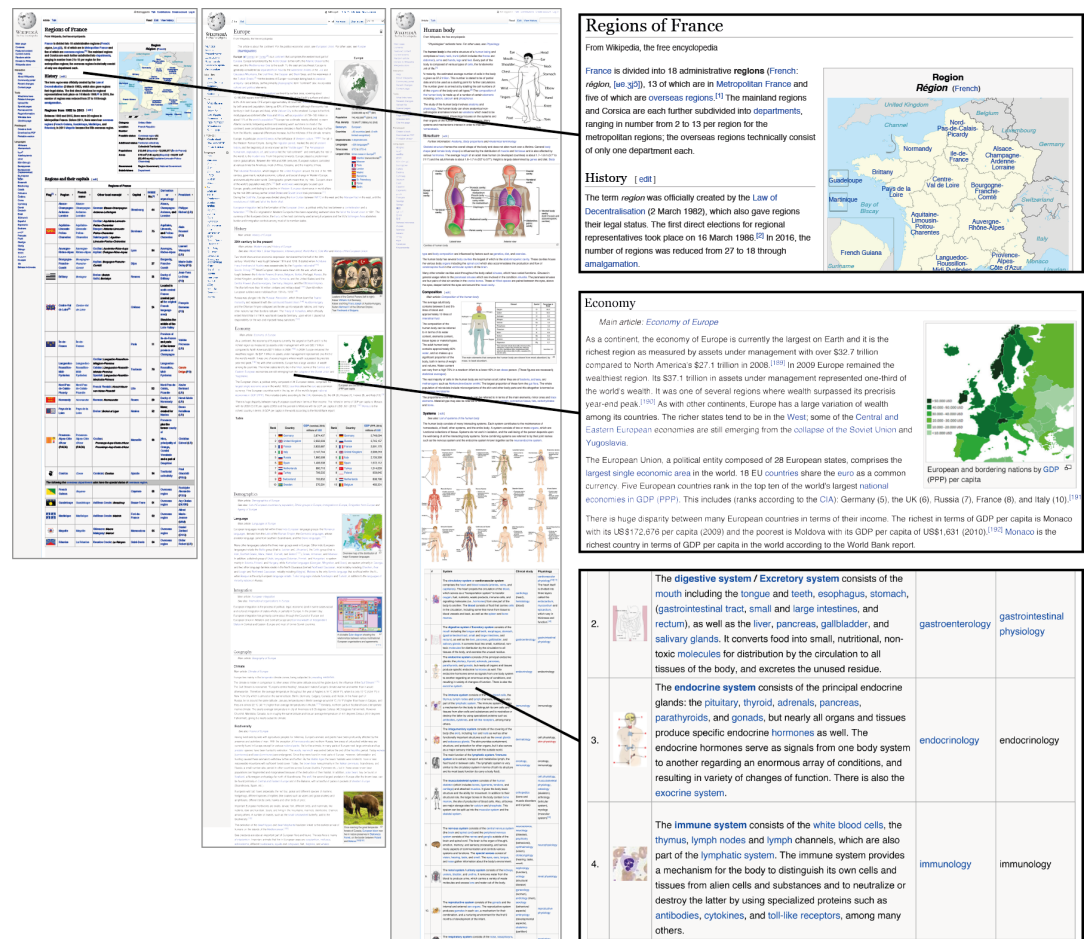


Figure 4.3: Thumbnails of the three Wikipedia articles. From left to right: “Regions of France”, “Europe”, and “Human Body”.

cluding 1D, 2D, 3D, multi-dimensional (MD), temporal, hierarchical (tree), and network data.

We used the shortest article on the “Regions of France” (see Figure A.1) as the starting article. It consisted of three short paragraphs describing the administrative regions of France and the history of their creation. At the bottom of the article, a large table provided basic information about the regions, including flags, capitals, and current presidents. To the right, an infobox included a schematic map of France and additional facts such as the populations of the largest and smallest regions. This article had only a limited set of data types: 1D, 2D, temporal, and MD.

The second text was an article on “Europe” (see Figure A.3). It contained five general paragraphs about the continent, followed by separate sections on Europe’s history, economy, languages, climate, biodiversity, and the creation of the European Union. Each of these sections included one image or data table. At the beginning of the article, an infobox provided data on various indicators (area, popula-



tion, etc.) and a map. In total, the “Europe” article contained 1D, 2D, MD, temporal, hierarchical, and network data.

The third article, the “Human Body”, contained three introductory paragraphs followed by sections about the body’s structure, its physical composition, and its anatomical systems (see Figure A.6). The article also included a large table at the bottom, which provided more detailed information about individual anatomical systems. Each section had at least one associated image. Unlike the other articles, this article contained 3D data (including diagrams of the human body). It also contained 1D, MD, hierarchical, and network data.

For each article we created an Adobe Illustrator file. To improve performance and make it easier for participants to manipulate the document, we grouped the characters in each line of text together as a single rasterized object. As a result, the smallest movable units in each document were lines of text, rather than words or characters. This allowed designers to concentrate on designing the visualizations rather than manipulating and rearranging the text. However, they were free to change the text color if needed.

Instead of the “Human Body” article, our very first participant (Po) used a Wikipedia article on the city of “Zurich” which contained no 3D or network data. We exchanged the “Zurich” article with the “Human Body” article after the first participant because we had decided in retrospect to cover all seven data types highlighted by Shneiderman with our articles. The “Human Body” article has 3D and network data which we did not have yet in our articles. To see the curated Wikipedia article on Zurich see Figure A.10. Despite this change, we chose to include Po’s data in our analysis as it did not significantly differ in quality or creativity from the other data we collected.

#### 4.2.4 Study Procedure

One experimenter was present throughout each design session. The study started with a questionnaire in which we asked about participants’ age and experience. Next, the experimenter gave a short presentation on word-scale visualizations, showed past examples of word-scale visualizations, and explained the study procedure. After this presentation, participants started the study with the article “Regions of France” (“Zurich” for Po). For 15 minutes, we allowed participants to familiarize themselves with the text of the article, generate initial ideas, and start to instantiate them with Illustrator. In order to reduce the overall study time, we allowed participants to approximate the actual data rather than trying to retrieve and reproduce exact values in their designs.

After a short break, participants then started with one of the two larger remaining Wikipedia articles. We counter-balanced the order of the larger articles between participants to mitigate fatigue effects. We

gave participants as much time as they wanted to familiarize themselves with each larger article. For the next 20 minutes participants created and integrated their data graphics into the article using Illustrator. We instructed participants to insert visualizations into the inter-line space of the main text, but not inside of tables, images, captions, infoboxes, or the navigation bar. After 20 minutes of design work, we gave participants the option to spend 5–10 additional minutes finishing their designs. All participants asked for at least 5 more minutes. We then allowed participants another short break before moving on to the second article, following the same procedure.

At the end of the study we debriefed each participant and asked them to explain each of their designs, describing their intended message and where the data originated. We also discussed the rationale behind our study and asked for any additional thoughts on the subject. Finally, we asked participants to describe how future authoring tools might make it easier for them to create and integrate graphics and text. The debrief was audiotaped and later on the responses were transcribed. On average, participants took roughly 2.5 hours to complete the entire study.

#### 4.2.5 *Set-Up and Materials*

We conducted the study in a small office with tables and chairs for up to three people. We provided blank sheets of paper for note-taking or quick idea sketches. Each participant brought their own laptop with Adobe Illustrator already installed. We asked them to do the study on their own laptop, so that they would have access to their own customized Illustrator configuration. We provided participants with a USB flash drive with the three curated Wikipedia articles, as well as a PDF copy of the slides we used to explain the study and task. All participants were conversant in English, and used English when including text in their word-scale graphics, but completed the interviews and demographic questionnaire in French.

#### 4.2.6 *Analysis Procedure*

During our analysis we first extracted all word-scale graphics produced by our participants. For each design we took a snapshot containing the entire illustration, along with any words or sentences associated with it. One of my collaborators and I then printed and clustered these snapshots using an open-coding approach, then refined the clusters through several rounds of discussion with all four collaborators. This process helped us identify a number of dimensions along which to compare and categorize word-scale graphics.

Participant	P0	P1	P2	P3	P4	P5	P6	P7	P8
# Designs	23	10	13	44	16	22	19	30	23

Figure 4.4: Number of word-scale graphics per participant

Category	Count	Examples
data-driven	159	
non-data-driven	41	

Figure 4.5: Examples of data-driven and non-data-driven designs.

### 4.3 RESULTS

We collected exactly 200 different word-scale graphics—a mean of 22 ( $SD = 10$ ) per participant (see Figure 4.4). In order to better understand the broader space of possible word-scale designs or potential uses for them, we coded all word-scale graphics according to: 1) whether or not they **encoded data**, 2) what **types of visual representation** they used, 3) how they were **connected to the text**, 4) how they **supported comparison** and **provided context**, 5) what **interaction** they suggested, and 6) the **designer's intent**. The data for the points 4, 5 and 6 was mainly collected through our interviews during which we discussed each of the designs. For the points 4 and 5 we do not provide quantitative data on how many instances of each design type per category we collected but only discuss the designs we observed. Similarly in point 6, we provide counts per study participant and not per design. Our main focus was to describe the diversity of designs, with a particular attention to the designers' editorial and rhetorical choices.

#### 4.3.1 Data vs. Non-Data-Driven Encodings

We first examined whether the graphics in our collection represented quantitative data or were more conceptual in nature. We coded graphics that used one or more visual variables (position, length, hue, value, ...) to encode data as *data-driven*. We refer to the **79.5%** (159/200) of word-scale graphics that were data-driven as *word-scale visualizations*. The remaining **20.5%** generally represented broad concepts, ideas, or messages not backed by abstract, quantitative data (see Figure 4.5).

	Level	Count	Examples
proportions, cumulative quantities		57	
absolute quantities		35	
location data		22	
changes, variations or trends		20	
temporal data		19	
rankings		15	
relational data		1	

Figure 4.6: Levels of types of data, their counts and examples.

4.3.1.1 Data Provenance

Having determined which word-scale graphics were data-driven, we examined the *provenance* of the data they used. We found that data came from one or multiple sources. Word-scale visualizations with just one data source took their data either from the associated sentence (49.7%; 79/159) or from somewhere else in the article (1.9%; 3/159). Visualizations with multiple sources either included data from the associated sentence + elsewhere in the article (11.3%; 18/159) or from the associated sentence + outside the article (37.1%; 59/159). While in theory word-scale visualizations can also visualize data exclusively from *outside the article*, we found no examples of this in our collection.

4.3.1.2 Types of Data

We then coded the *types* of data used in our word-scale visualizations. While we designed the three Wikipedia articles to contain all of Shneiderman’s data types [133], our coding deviated from his set to make more fine-grained distinctions. From Shneiderman’s set we kept *location data*, *temporal data* and *relational data*, but added *proportions & cumulative quantities*, *absolute quantities*, *rankings*, and *changes, variations or trends* (see Figure 4.6).

**Proportions, cumulative quantities:**

Visualizations in this category—such as stacked bars or pie charts—put absolute quantities in relation to other quantities. We classified 57/159 (35.8%) visualizations in this category.

**Absolute quantities:**

The 35 (22.0%) designs in this category represented single quantitative data values, often using visualizations made up of countable units—such as the black dots for countries, and blue bars for regions in the examples above.







	Level	Count	Examples
quantitative abstract visualizations		141	
geographic and schematic maps		16	
icons & symbols		36	
illustrations		7	

Figure 4.7: Levels of types of graphic, their counts, and examples

**Location data:** 

Location context or highlights were the basis of 22 (13.8%) visualizations—many of which were small maps.

**Changes, variations or trends:** 

The 20 (12.6%) visualizations in this category showed data values changing according to another data dimension such as time.

**Temporal data:** 

The designers included 19 (11.9%) visualizations depicting *temporal data*; mostly in the form of timelines and often highlighting a specific time point or period.

**Rankings:** 

*Rankings* were present in 15/159 (9.4%) designs that showed ordered sets in which typically a focus object was highlighted.

**Relational data:** 

A single design (0.1%) included *relational data* depicting connections between different systems in the human body.

#### 4.3.2 Design of Word-Scale Representations


We coded the types of visual representations used in our word-scale graphics and also noted whether or not designers followed any specific process in making their designs.

##### 4.3.2.1 Types of Visual Representation


We found examples of *quantitative abstract visualizations*, *geographic and schematic maps*, *icons & symbols*, and *illustrations* (see Figure 4.7).

**Quantitative abstract visualizations:** 





The 70.5% (141/200) of word-scale visualizations in this category used visual variable(s) like length, position, or color to encode quantities—which were typically numbers drawn from the text.

**Geographic and schematic maps:** 


8% (16/200) of the word-scale graphics were geographic and/or sche-

matic maps. We coded these separately from *quantitative abstract visualizations*, as it was often unclear whether they actually encoded quantitative data (beyond the geo-coordinates of regional boundaries). For example,  is a schematic representation of the topological surroundings of the city of Zurich—from left to right: a mountain, a lake, and a forest.

**Icons & symbols:** .

Of the word-scale graphics 18% (36/200) did not encode any quantitative data at all. Instead, they replaced a single concept or word in the text. Some *icons & symbols* directly resembled the object they referred to (such as a  standing for cranial bones), while others described an entity—a concrete piece of text (a word or sequence of words) that can have associated metadata—but needed to be learned to be understood (for example,  standing for France). Icons & symbols were also sometimes used as a part of quantitative abstract visualizations. Here, they had a secondary function to help relate the shown quantities to known ones [28], or to give contextual cues. For example in  the standard heights of men and women are related to a “known” height of a cow (according to the designer who created the graphic). In a second example  the graphic shows a ranking with a human figure on the left signalling that the graphic shows population data.

**Illustrations:** .

This group of 7 (3.5%) word-scale graphics included more complex diagrams and sketches. Often, these illustrations tried to convey a message. For example, the following small graphic , shows two human skeletons with one sitting in a wheelchair. The designer wanted to convey the message that people with disabilities are often not considered in articles on the human body.

#### 4.3.2.2 Reusable Graphical Systems

Several designers used a coherent visual theme for their word-scale graphics. They largely remained consistent in terms of coloring schemes and connection between entity and graphic (see Section 4.3.3.2). One designer (Po) went further and proposed a visual language based on the assembly of small data-driven units (see Figure 4.8). This system consisted of introducing tokens (token grammar) and then assembling them (assembly model) to express various quantities. It was similar to the construction of unit-based visualizations discussed by Huron et al. [77]. An example of such a reusable and unit-based graphical system is Isotype (“International System Of Typographic Picture Education”) coined by Neurath [105] but first described by Brinton [18] a couple of years earlier. The main idea behind Isotype is to employ countable units represented as simple pictograms that provide a semantic. These countable units are then organised in a layout that permits the comparison and countability between different cate-

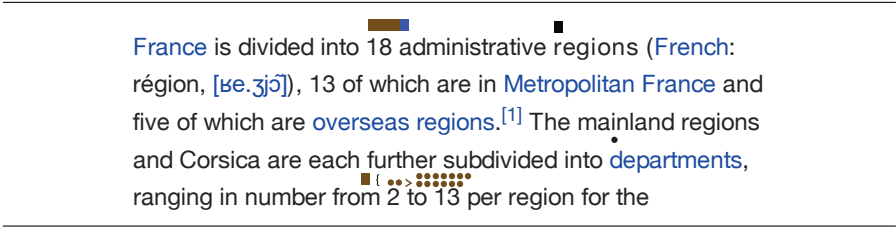


Figure 4.8: Example of an implementation of a reusable graphical system. The square stands for a region unit and the circle for a department. A bar chart and a trend visualization are assembled from these units.

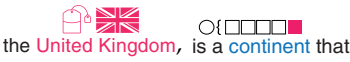
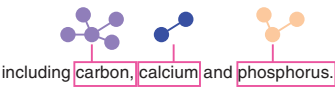


Level	Count	Examples
font color, existing hyperlink	51	
frame	36	
underlining	11	
background color	7	

Figure 4.9: Types of connections between designs and text.

gories within a visualization but also between visualizations. Isotype is an interesting type of visualization that could be transformed into word-scale visualizations.

4.3.3 Connections Between Text and Graphics

We started this study with the premise that the definition of an *entity* is: a single word or a word sequence with associated metadata that can be encoded in a word-scale visualization. However, this definition did not specify in detail how entities and their corresponding visualizations can be connected. As part of our study, we wanted to analyze these text-to-graphic connections in more detail and build a better understanding of how designers chose and communicated these relationships.

4.3.3.1 Size of the Associated Text

Designers often associated graphics with one- or multi-word entities. In the study results, we identified 139 entities, with lengths ranging from a single number or word up to the size of a sentence, paragraph, or document subsection (*median* = 2.0 words per entity, *SD* =

4.9). One-word (39%; 54/139) and two-word (25.9%; 36/139) entities were the most common. However, the relationships between entities and graphics was not always one-to-one. Although 42.5% (85/200) of graphics were connected to a single entity, 9.5% (19/200) were connected to multiple distinct entities at different locations. We also found a number of word-scale graphics (48%; 96/200) that were not clearly associated with any specific subset of the text.

#### 4.3.3.2 *Type of Connection*

The designers in our study linked 52% (104/200) of the graphics they created to one or more entities using a variety of different graphical techniques (see Figure 4.9):

**Font color / existing hyperlink:** In 49% (51/104) of the cases, a font color associated with a word served as the graphical connection between word-scale graphic and text. In 19% (20/104) of the cases the designer altered the color of the text to create the mapping—often using color present in the graphic itself. In 30% (31/104) of the cases, designers used the existing highlighted Wikipedia hyperlinks in the article as entities, sometimes even reusing the link color (dark blue) in the graphic to emphasize the connection.

**Frame:** The second most common technique (35%; 36/104) used a frame around associated words and sometimes added a reference link between frame and graphic.

**Underlining:** In 11% (11/104) of the cases, designers used a line drawn under an associated word to connect text to the graphic.

**Background color:** In 7% (7/104) of the cases, designers created a visual connection by changing the background color of the text.

Another 48% (96/200) of our word-scale graphics were not linked to text by graphical techniques but rather using spatial proximity. Effectively creating an implicit and loose connection. Thus this implicit connection creates uncertainty around what the exact boundaries of the entity are. In most cases, designers either left- or center-aligned the word-scale graphics with a related word or phrase. In some cases, designers associated graphics with whole paragraphs or subsections. For example, in Figure 4.10, P4 visually associated the word-scale visualization with a complete paragraph by placing the visualization immediately to the left of the entire text block. Similarly, in Figure 4.11 the same designer placed a word-scale visualization next to a section heading, effectively associating the visualization with the entire subsequent section. Two designers (P4, P8) added few (P4) or almost all (P8) of their designs below the entity. Several of these designs showed long vertical bars (for example in a stacked bar chart) that served as an underline in terms of its connection to the associated entity as can be seen in Figure 4.12. In total 14.5% (29/200) of the designs were po-



---

The European Union, a political entity composed of 28 European states, c  
largest single economic area in the world. 18 EU countries share the euro  
currency. Five European countries rank in the top ten of the world's largest  
economies in GDP (PPP).

---

Figure 4.10: The word-scale visualization has been positioned to the left of the paragraph, showing the proportion of countries using the Euro (18)—represented as 18 € signs—versus the number of countries actually in Europe (28)—represented using 28 icons of a flag.

---

Main article: *Economy of Europe* 

---

As a continent the economy of Europe is currently the largest on

---

Figure 4.11: The word-scale visualization has been added to the title of the paragraph. It is a horizontal bar chart comparing the assets under management in Europe and North America.

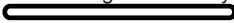


sitioned below the entity instead of above but used the same general connection types as discussed before.

#### 4.3.3.3 Legends and Captions

The design of small-scale visualizations allows little space to add captions or labels to explain a visualization. Interestingly, the designers in our study addressed this challenge in a variety of ways, summarized in Figure 4.13.

The top graphic in Figure 4.13, for example, shows how *word-scale graphics* themselves can serve *as part of a legend* to a larger graphic. Here, small icons added to the individual element names (carbon, calcium, phosphorus) label the elements of the stacked bar chart above.

---

The average adult body  
  
contains between 5 and 5½  
  
litres of blood and  
  
approximately 10 litres of  
interstitial fluid.

---

Figure 4.12: The three word-scale visualizations are placed below the entity. The first graphic represents the average adult body as a container. Then, in the second graphic the 5 liters of blood are added. And the third graphic connects “blood” to the red/orange color.




Type of Legend or Caption	Examples
word-scale graphics as legends	 including carbon, calcium and phosphorus.
entities as legends	 including carbon, calcium and phosphorus
entities as captions	 the human body is

Figure 4.13: Several ways in which graphics and entities serve as legends or captions.

In cases such as this one, the larger visualization was typically connected to multiple entities which in turn had small associated iconic representations placed in close proximity to them. For example, the designer who introduced a token-based grammar for the design of her visualizations (Section 4.3.2.2) used this mechanism to introduce the tokens of her grammar, placing individual instances of each token near the first occurrence of the corresponding term.

The second example in Figure 4.13 shows the same text as in the row before illustrated by another designer. The designer colored the element names to together serve as the legend to the associated pie chart. In general, visualizations with associated *entities as legends* included multi-entity associations in which the entities were colored according to the elements of the visualizations they were associated with.

These first two techniques used entity text to label parts of word-scale visualizations. The last example on the bottom, instead, uses an *entity as a caption* to describe the overall content of the visualizations without giving any details about the individual stacked bars.

While Figure 4.13 summarizes three common uses of legends and captions, designers also explored a range of other variations. In a few cases designers even created compound captions and legends like the one shown in Figure 4.14a. Here, the designer introduced a multi-line entity with nested caption and legend elements that help readers to interpret the larger bar chart.

Finally, one designer also created an explicit legend in the margin explaining the types of graphics introduced in her article, including representations showing geographic, temporal, and population data (see Figure 4.14b).

#### 4.3.4 Supporting Comparison and Providing Context

Designers also used a variety of strategies to support comparison between word-scale visualizations in the document:

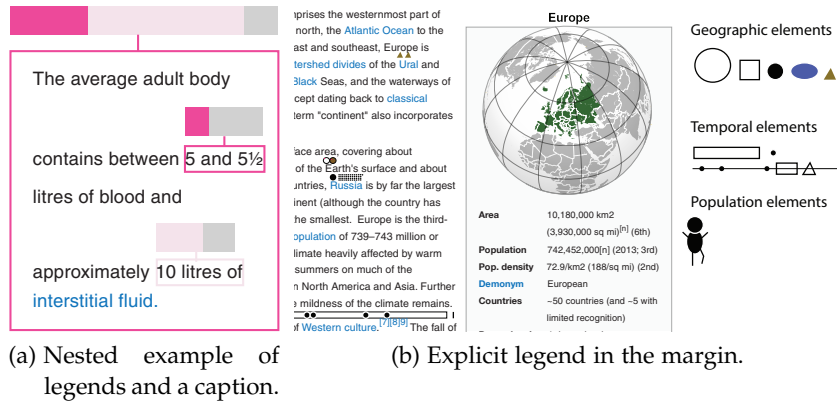


Figure 4.14: A nested legend and caption explains the top bar chart in (a). The right image (b) shows an example of an explicit legend created and placed in the margin by a designer.

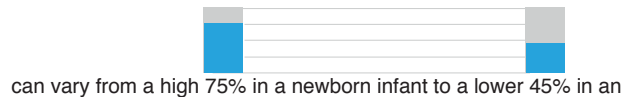


Figure 4.15: Connecting multiple visualizations with a support grid.

**Shared reference structure:** Multiple designers used shared reference structures, including reference lines and grids which spanned multiple visualizations. In one example (Figure 4.15), P1 added a grid connecting two word-scale visualizations on the same line such that the viewers could more easily compare them. In cases where visualizations were further apart in the document, some designers created duplicate reference structures to preserve comparability. In Figure 4.16, for example, P3 repeated the same reference grid and icons in two word-scale visualizations, highlighting small differences between heights more clearly.

**Traces:** Another static method we saw (P1, P4, P7) for comparing word-scale visualizations used a graphical overlay or trace of one visualization on to another (Figure 4.17).

**Sequencing:** Several designers (P3, P5, P7) created sequences of word-scale visualizations in which each visualization built upon the previous ones. For example, in Figure 4.18, P7 placed a small timeline over each date in the text. Each successive timeline also included desaturated versions of all of the previous dates, helping to place the current date in context with the earlier ones. Other examples of sequencing included sets of maps which built upon one another, each containing elements of the previous one.

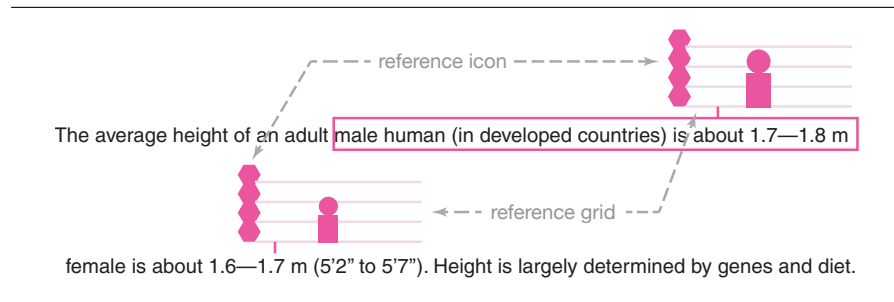


Figure 4.16: Repeated grid and icons support comparison (we added the grey annotations for explanatory purposes).

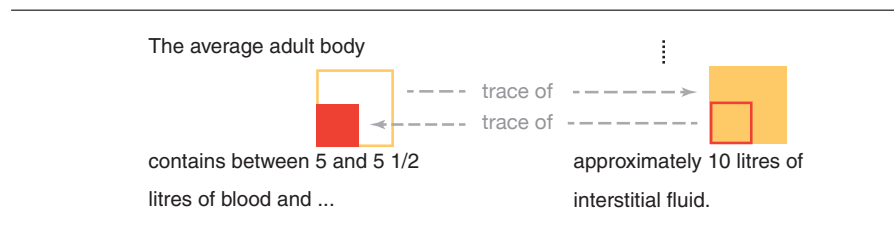


Figure 4.17: Overlaying a trace of one visualizations over another (we added the grey annotations for explanatory purposes).

#### 4.3.5 Interaction

We did not ask the designers to consider interaction when creating their word-scale visualizations. However, during the post-study debrief several participants proposed ways of augmenting their designs with interaction:

**Connecting word-scale visualizations:** P6 suggested adding interactive brushing and linking to the visualizations such as in Figure 4.19—allowing readers to interactively explore the relationship between hormonal changes and the development of fat and muscle tissue.

**Comparing word-scale visualizations:** P3 proposed several techniques for bringing multiple word-scale visualizations together to facilitate comparison. By dragging one word-scale visualization onto

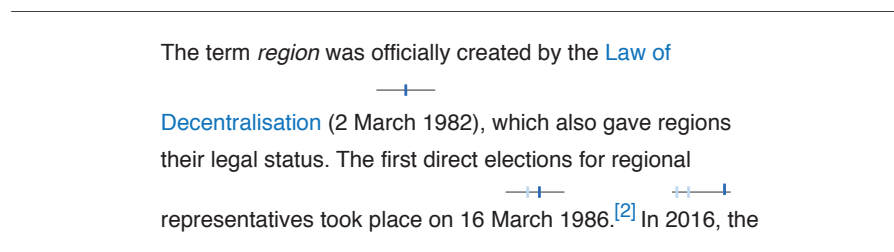


Figure 4.18: Sequencing of timeline word-scale visualizations.

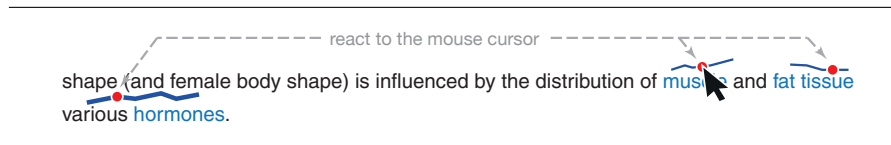


Figure 4.19: Three connected word-scale visualizations. Moving the cursor in one graphic results in highlighted values in the others (we added the red dots and the cursor to explain the interaction)

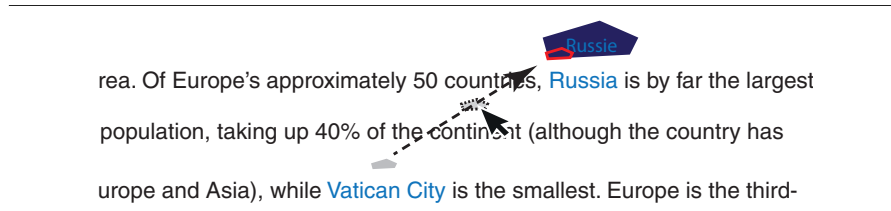


Figure 4.20: Using interaction to overlay the word-scale visualizations for one country on top of the graphic for another. (Sizes are approximate—as estimated by the participant. We added the cursor and line for illustration.)

another (as in Figure 4.20) a reader could overlay the two visualizations on top of one other. Similarly, dragging one visualizations next to another could facilitate side-by-side comparison. Both approaches would allow for dynamic comparison without the use of static support structures like those seen in the preceding section.

**Revealing word-scale visualizations:** P2 suggested using interaction to show word-scale visualizations on demand—displaying visualizations only in the paragraph that a reader is currently attending to. All word-scale visualizations outside of the paragraph could either be hidden, desaturated, or scaled down to reduce visual clutter.

**Data interactions:** Interaction could also be used to trigger filtering or highlighting on the visualizations themselves. For example, P2 proposed using interaction with labels on the right side of the word-scale visualization in Figure 4.21 to transition between alternative views. In this case, hovering over the labels for “population” (only the fillings are shown without the rectangle border) and “size” (the fillings of the rectangles are not shown) would toggle which of the two data sets the chart displayed. If there is no hovering both datasets are shown simultaneously.

#### 4.3.6 Designer’s Goals

During their debrief sessions, designers described a diverse range of goals that guided the ways they designed and integrated word-scale visualizations with the documents.

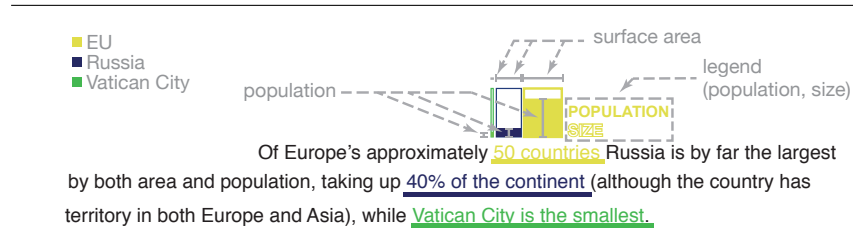



Figure 4.21: Word-scale visualization showing the contrast between population and surface area for the EU, Russia, and Vatican City (we added the grey dashed lines and annotations for explanatory purpose).


**Facilitating understanding:** The main goal all nine designers cited was *facilitating understanding*. Concretely, they mentioned using visualizations to illustrate concepts (3 out of 9 designers), make quantities explicit (6 out of 9 designers), provide scales (3 out of 9 designers), or put information into context (4 out of 9 designers). One designer also suggested using word-scale visualizations to provide more detailed information about the entity as a way to facilitate understanding.

**Summarizing the content:** Moreover, six designers expressed the use of word-scale visualizations to help *summarizing the content* of a sentence. Designers discussed using word-scale visualizations to provide a preview of the content or to expose the most important information from the text, providing readers with a clearer take-home message.

**Emphasis:** Two designers mentioned using word-scale visualizations to emphasize and draw attention to content they felt was particularly important. We were surprised that only two designers mentioned this explicitly as the presence of word-scale visualizations in text already provides a type of intrinsic emphasis effect as defined by Kyle et al. [68] where the process of mapping content to a visualizations increases its visual prominence. In addition, many of the text-to-graphic connections (font color, frames, underlining, background color) give further extrinsic emphasis [68] to the associated entities.

**Providing alternative representations:** Two designers also explicitly stated that they used word-scale visualizations to *provide alternative representations* as a supplement to the written text. They emphasized that introducing a redundant encoding of the information in the text could help make the text both more understandable and more memorable [10]. For example one of the designer suggested to add a visual representation of a cell—as known from Biology text books—to the entity “cell  ” in order to make it more memorable.

**Providing contrasting viewpoints:** Interestingly, two designers also described creating visualizations to *express a contrasting viewpoint* that differed from what was written in the text. Here, designers

mentioned hoping to encourage readers to think more deeply about the text and consider alternative interpretations of it. For example one designer proposed to add a map of the world showing where Europe was perpetrating slavery in the 16th century to provide a contrasting viewpoint on how Europe “controlled” the world at that time. Especially, pointing out that the word choice “controlled”  seems too neutral.

#### 4.4 DISCUSSION: USING WORD-SCALE VISUALIZATIONS

Having developed a fine grained understanding of the different design dimensions of word-scale visualizations, we here propose a higher-level definition of use-cases for them. We intend these to serve as inspiration and starting points for authors and designers who wish to integrate word-scale visualizations into text documents. We believe this higher-level approach is generally more accessible for generating *new* designs and applications, in contrast to the low-level characteristics from Section 4.3 which are more suitable for describing *existing* examples. Furthermore, we reflect upon our premise from before starting this study of what an entity is in light of the study results and propose an extension to the notion of associated entity. Moreover, we discuss challenges that arose during the study.

##### 4.4.1 Applications of Word-Scale Visualizations

We characterize three different ways in which designers can apply word-scale visualizations to alter the rhetorical structure of a document. Specifically, word-scale visualizations can *support the content*, *extend the content*, or *go beyond the content*.

###### 4.4.1.1 Supporting the Content

From an editorial perspective, one of the simplest ways of applying a word-scale visualization is using it to *support* its associated entities. In these cases, word-scale visualizations display data (shown above in green) that comes directly from the entity text itself.

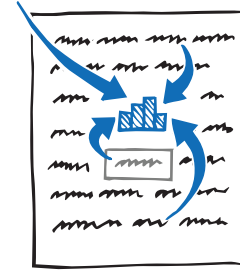


Thus, when creating word-scale visualizations that support the content, editorial choices are relatively restricted. The editor or designer needs to choose which entities to visualize, how to design the representations, and how to place and connect the visualizations to the associated text. However, they do not need to locate or prepare additional data. For example, an editor could choose a number in the text as an entity and represent that value using a chart to show the number as a part of a whole (e.g. as a pie chart). Although they

restate the existing content, these visualizations can still provide useful emphasis—facilitating better understanding of numbers or concepts [10], providing summaries, or highlighting alternative versions of the entities (see Section 4.3.6).

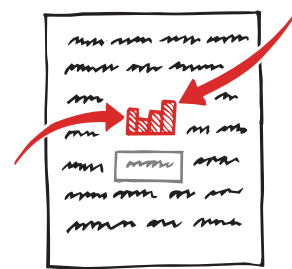
#### 4.4.1.2 *Extending the Content*

Word-scale visualizations can also be used to provide contextual data that extends the ideas put forward by their associated entities. These word-scale visualizations *extend* the content with compatible information (shown above in blue), and do so without purposefully changing the message of the sentence or proposing novel interpretations. Yet, the creation of word-scale visualizations that extend content is somewhat more complicated than for the previous group—it involves finding data from elsewhere in the text or from outside the document. For example, an author could extend a timeline connected to the date of a WWII battle by adding outside data such as the start and end dates of the war. Visualizations that extend the content can help provide context or comparisons for data already contained in the text or provide additional, supplemental material about specific entities.



#### 4.4.1.3 *Going Beyond the Content*

The previous two applications of word-scale visualizations generally complement and reinforce the content of the original document. However, authors and editors can also use word-scale visualizations that *go beyond* the content, including contrasting ideas and messages (shown above in red) which may significantly broaden, annotate, or even subvert the original meaning. By including a word-scale visualization showing external data that contradicts the existing narrative, an editor or designer can for example suggest an alternative, contrasting viewpoint (see Section 4.3.6). While these contrasting viewpoints may originate with the original author, they can also be added in response to the article at a later point.



#### 4.4.2 *Defining “Word-Scale”*

As we alluded to in Section 3.1, prior work has offered several competing terms and definitions for describing small graphics and visualizations embedded in text. Here, we revisit these descriptions in light of our study results.



Tufte’s definition of sparklines refers to them as “datawords” and describes them as “word-sized” [148]. While many of Tufte’s examples—such as his baseball win/loss charts—are much longer than typical words, most are still essentially word-height. In our study and elsewhere, we have observed many word-scale visualizations that go well beyond this size limit. As such, our definition of word-scale visualizations in Section 3.1—which allows visualizations that “are roughly word-sized, ranging from e. g. as small as a single character to as large as several words”—better captures the variety of visualizations we saw in the study.

Parnow and Dörk’s work on micro visualizations [111, 112] provides another definition for small embedded visualizations. Rather than focusing explicitly on size, their definition instead stresses simplicity—emphasizing that such visualizations contain few data dimensions or points. While we found a number of simple data-driven graphics that do fit this definition, we also found more data-intense graphics that would not be covered. As such, our definition of word-scale visualizations better captures the graphics we saw, as it does not explicitly restrict the number of data dimension or data points.

However, while our definition of word-scale visualizations seemed to be a good term to capture the data-driven graphics we saw in the study, it does not encompass the full range of non-data-driven graphics designers produced. Yet, our study clearly suggests that non-data-driven graphics at these scales can express a variety of design intents and may be just as useful a communication tool as their data-driven counterparts. A higher-level designation such as “word-scale graphics” can capture both data-driven and non-data-driven variants, and serve as a unifying term for characterizing a wide range of small graphics that can be embedded in text. Nevertheless, I would like to stress that in this dissertation I concentrate on word-scale visualizations.

#### 4.4.3 *What is an Entity?*

When first defining the term word-scale visualization, we had assumed that each visualization would explicitly be associated or linked with a single *entity* (see [60]), i. e. a single word or a sequence of words in the text. However, participants did not always follow such a close association as we had envisioned and led us to re-evaluate our initial assumptions. For example, in addition to using proximity to establish the connection between a word-scale visualization and the text, they also linked multiple disconnected pieces of text to one word-scale visualization (see Section 4.3.3). Designers also created word-scale visualizations without any explicit or implicit entity. In some cases designers added word-scale visualizations to the start of a paragraph or after a title. Designers also added word-scale visualizations verti-

cally along a paragraph (see Figure 4.10)—effectively associating the visualizations with the entire block of text.

A particularly interesting finding of our study showed that entities also served a function that went beyond the mere provision of metadata we had envisioned. Instead, they served a reference function and were used as parts of legends.

As a result, we propose an updated definition of what an associated entity is. An entity is still a concrete piece of text that can have associated metadata, however, the entity also serves as a way to link the word-scale visualization to the text, and can be part of the word-scale visualization’s legend, or of its caption. Word-scale visualizations can have one entity, multiple entities or none associated to them. If there is an entity, it is marked using for example font color or a frame. In cases where the word-scale visualization has no marked entities, it means the visualization is associated with the sentence or the paragraph, depending on its positioning.

Finally, it may be productive to more deeply consider the relationship between entities and word-scale graphics using the semiotic concepts of *referent*, *signifier*, and *signified* [39, 107, 115]. These semantic connections are particularly interesting, as entities and word-scale graphics sometimes represent the same concept, but may also introduce indirection or serve as an alternative representation of an underlying idea.

#### 4.4.4 *Challenges when Creating Word-Scale Visualizations*

One of the main challenges our participants had was to find suitable pieces of text they could complement with word-scale visualizations. The process of deciding what should and could be represented usually happened at the very beginning of the process of integrating word-scale visualizations. The designers also expressed concerns that the associated entities in the text had to be below a certain length to be representable or connectable to a word-scale visualization. We argue that our updated definition of an entity provides more freedom to support the linking of word-scale visualizations to the text.

In addition and linked to this thematic of defining associated text, our participants had difficulties finding suitable places to position their word-scale visualizations. In our instructions, we restricted the positioning of word-scale visualizations to the interline space, however sometimes designers still hesitated and sought alternative placement options—for example at the end of a title or along a paragraph (see Section 4.3.3). One participant also told us that she would have liked to rewrite the sentences to make it easier for her to integrate word-scale visualizations. She found the sentences were too convoluted, which led her to find it difficult to find suitable positions for her word-scale visualizations. Moreover, some of the information or

data that the designers wanted to integrate into the word-scale visualizations was spread over multiple lines. This made it more difficult to find good places to position data graphics. In some cases, designers also started to create a graphic first and then had a difficult time finding suitable text to associate with the word-scale visualization. As Section 4.3.3 and Section 4.3.4 show, there are multiple ways to link entities to a word-scale visualization.

#### 4.5 LIMITATIONS

The overall goal with the study was to improve our understanding of how, when, and why one might use word-scale visualizations in text documents. Yet, the results of our study and our discussion need to be interpreted in light of several specific study design decisions. In this study we focused on the design of word-scale visualizations and not on their placement. To give participants enough study time to focus on design, we only allowed them to place visualizations between lines of text. This made placement easier, since designers did not have to shift individual words or reflow text and could instead focus on creating interesting new designs. To increase opportunities for creative designs, we allowed participants to freely adjust the amount of inter-line space, allowing them to explore graphics with a greater variety of different heights, widths, and positions relative to the text. Participants were not able to edit the text itself besides changing the text color, and thus did not have full control over the phrasing, language, or organization of the content they were building on. In the post-session interview, participants had a chance to express their ideas and concerns about the placement of their word-scale visualizations. Only one of our participants mentioned that she felt her design had been restricted by the fixed text—noting that she would have liked to rewrite sentences to better fit her word-scale visualizations and use visualizations in-line as a legend. Thus, while 8/9 participants did not report design restrictions due to the fixed placement, it is likely that our results do not show the full spectrum of creativity that the participants could have applied to word-scale graphic placement.

Two further factors influencing the reported designs are the choice of text and the restricted study time. Wikipedia articles are usually much more descriptive and data-driven than articles, novels, poems, or other forms of text. Participants also did not have time to make multiple iterations over their designs, as we asked them to maximize the diversity of their designs rather than produce a few refined ones. Therefore, the designs should be seen as initial ideas, which would require iteration and further development before their effectiveness could be tested.

#### 4.6 SUMMARY

This chapter presented the results of an exploratory study on the use of word-scale visualizations in text documents. The study resulted in a rich collection of different types of graphic designs, and also helped us to identify specific design challenges unique to word-scale visualizations. Unlike larger visualizations, designers creating word-scale visualizations must consider the immediate connections between visualizations and text, and also find alternatives to common visual elements like axes, legends, and captions that work given reduced space. Designing interactions also presents a challenge, since graphics may be dispersed, small, and only loosely associated with one another. At the same time, our results also illustrate a variety of ways in which word-scale visualizations can enrich text documents. To support, inspire and serve as a starting point for the creation of word-scale visualizations, we propose three application goals for word-scale visualization: supporting the content, extending the content and going beyond the content. We see both our study results and the three application goals as starting points for how word-scale visualizations might complement a variety of types of text documents.

As a result of this study, we propose a more fine grained definition of an entity. An *entity* is a concrete piece of text (a word or sequence of words) that is associated to one or multiple word-scale visualizations and that can have associated metadata. The entity also serves as a way to link the word-scale visualization to the text, and can be part of the word-scale visualization's legend, or of its caption. Word-scale visualization can have one entity, multiple entities or none associated to them. If a word-scale visualization has an associated entity, it is marked using for example font color or a frame. In cases where the word-scale visualization has no marked entities, it means the graphic is associated with the sentence or the paragraph, depending on its positioning. Furthermore, we acknowledge that the space of small contextual graphics can also include non-data-driven graphics and proposed *word-scale graphics* as an overarching term for data-driven and non-data-driven graphics. The term word-scale graphics is based on the same definition as word-scale visualizations. Thus word-scale graphics are open to various visual encodings including geographical maps, heat maps, pie chart, and more complex visualizations on the data-driven side and includes among many others those that we observed: icons, symbols and illustrations on the non-data-driven side. Moreover, word-scale graphics can appear at a range of scales, including sizes larger than a letter, but smaller than a sentence or paragraph. However, again I would like to stress that in this dissertation we concentrate on word-scale visualizations but keep in mind that there is a whole space of non-data-driven contextual graphics.

This chapter focused on the authoring and design of word-scale visualizations. Yet more studies are required to deepen our knowledge on their integration. These studies should have a focus on the size and the embedding of the word-scale visualization into the text document taking into account readability of word-scale visualization and the text but also aesthetics of the combination of word-scale visualization and text. Furthermore, the design space of how to add captions and labels, and how they can be merged with the text is largely unexplored.

In the next chapter I begin exploring the design space of word-scale visualization further by looking at their placement in the text.



## PLACING WORD-SCALE VISUALIZATIONS

---

In the previous chapter, I discussed the use and design of small data-driven contextual visualizations. This discussion was based on the results of a study during which designers embedded small-scale visualizations into curated Wikipedia articles.

The current chapter contributes and discusses an illustrated design space of the placement options for word-scale visualizations. These placements include positioning the visualizations between lines within additional vertical and horizontal space in the document and as interactive overlays on top of the text. I identify six essential variables that control the placement of the visualizations and the level of disruption of the source text. I also contribute a quantitative analysis that highlights the effect of different placements on readability and text disruption. Based on this analysis, I propose guidelines to support the design and placement of word-scale visualizations.

The main parts of this chapter were previously published in Goffin et al. [60]. Hence, the use of “we” in the remainder of this chapter refers to Pascal Goffin, Wesley Willett, Jean-Daniel Fekete, and Petra Isenberg.

### 5.1 MOTIVATION

In our previous study about the design and use of word-scale visualizations, we restricted where designers could place them. Our objective for the restriction was to spare the participants time to look for optimal placements and to simplify the embedding of the visualizations. Furthermore, we were not interested in characterizing different types of placements. Nevertheless, participants mentioned that it was still difficult and painful to find good placement options (see Section 4.4.4). Actually, almost 50% (48%) of the designs were not associated through graphical techniques with an entity and around 15% of them were placed below the piece of text instead of above as we had instructed. The difficulty to find good placements lead the designers to explore several techniques—visual and non-visual (see Section 4.3.3.2)—to associate word-scale visualizations and text.

The most frequent method of association was to use spatial proximity. As a result of our placement restrictions to the inter-line space, we could not observe the designers’ full creativity in using proximity as a placement option. Still, we wanted to explore placement as a way to implicitly—without using visual links—associate word-scale visual-


ization and text. An additional consideration was also that placement can be dependent on the medium and the mutability of the text where a designer integrates the word-scale visualizations. Sparklines for example have typically been placed before or after the word they are related to. However, this is often not possible for different reasons—e.g. when adding information to scanned documents and other immutable texts. One way to go around this constraint is to use OCR on the scanned text which then results in a mutable text. Placing word-scale visualizations in-line with text may also be undesirable in other situations, as it requires reflowing the text and restricts the visualizations' maximum height to that of the font—making visualizations hard to read when small font sizes are chosen. In-line visualizations can also disrupt sentences, making the text more difficult to read.



The consequences of these placement decisions, in particular, are not well-understood. Thus, to better understand these consequences and to explore the large space of placement options available for integrating word-scale visualizations in text documents, we outline a design space of possible placements relative to the text.

## 5.2 BACKGROUND

The topic of this chapter closely relates to four research areas: (a) the use and placement of word-scale visualizations (b) the integration of meta-data within text documents, (c) research on labeling in visualization, and (d) the readability of texts and visualizations.

### 5.2.1 *Sparklines and Word-Scale Visualizations*

According to Tufte [148] sparklines are “small, intense, simple, word-sized graphics with typographic resolution” that can be included anywhere a word or number can be—e.g. in a sentence, table, headline, map, spreadsheet or graphic. However, Tufte provides little guidance for *placing* word-scale visualizations with respect to text, suggesting only that they be placed in a “relevant context”—usually just after the word that they complement. He presents several examples of these embeddings. One example shows sparklines embedded in-line with text in order to provide metadata for a single word, for example glucose measurements next to the word glucose (Glucose 3.4 ). In another, sparklines show temporal trends in a data table to provide information that would otherwise have required many columns.

Besides these examples, the embedding of word-scale visualizations have been discussed relatively infrequently in the research literature. Four of the examples mentioned in Section 3.1 propose more details on their placement inside the text: SparkClouds [95], Gestaltlines [16], Separation Plots [67] and SportLines [118]. SparkClouds ( ) embed sparklines below words in a tag cloud to convey



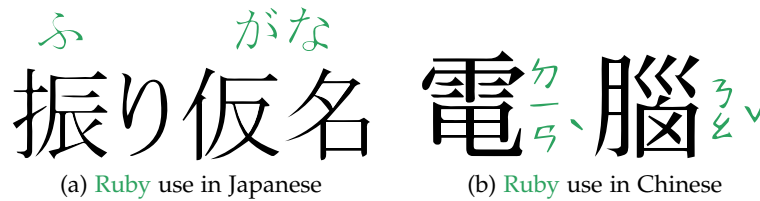


Figure 5.1: Two examples of ruby placement: on top of the kanjis (a) and to the right of the kanjis (b). 5.1a Image in the public domain, 5.1b Image credit: [79]

temporal trends. Gestaltlines (#####) by Brandes et al. show an embedding in-line with text, similar to Tufte’s glucose example. Similar embeddings are shown for Separation Plots (#####) and SportLines (#####). In this chapter, we show that other placement options exist, and can also be used to place Gestaltlines, Separation Plots, SportLines, and many other small-scale visualizations within text documents.

Glyphs are another type of small visualizations related to sparklines but are not typically placed in the context of text documents. Instead, glyphs are often used as small-multiples or replace data marks in other visualizations, e. g. on maps, scatterplots, or in graphs. Yet, glyphs are also sometimes placed inside text documents. Abdul-Rahman et al. [1], for example, used glyphs to encode and visualize a large set of poetic variables and place these glyphs in the space between lines. In the context of the glyph overview article by Ward [157], he also contributes a taxonomy of glyph placement strategies. He categorizes these strategies into data-driven and structure-driven approaches. A data-driven placement positions glyphs based on some of the dimensions of the represented data point, whereas structure-driven approaches use existing relationships between the data points (e. g. an order, hierarchy, or network links). The problem of placing a word-scale visualization next to an entity within a text document is related to the structure-driven placement option. Thus, Ward shares several placement concerns with our work, including overlap between glyphs and the addition of space to make room for glyphs.

In summary, while research has considered the problem of designing small visualizations and even embedding them in various contexts we are aware of no previous work that has explored the placement of word-scale visualizations in text documents.

### 5.2.2 Meta-Data within Text Documents

While the literature contains little advice on how to place word-scale visualizations in text, others have considered how to place general meta-data within text documents.

Annotations are a general category of meta-data placed in text documents. In East Asian scripts often parts of the “base text” is annotated with short annotations called Ruby [130]. Their purpose is usually to provide a reading on how to pronounce the base text. They are frequently used in Japan, where they are called furigana and often appear in books and magazines. They are also used in China, where they often can be found in schoolbooks. The name “ruby” originated from the 5.5pt font size used in British printing, which corresponds to half the 10pt font size used commonly in normal text. Ruby text is often placed on top as in Figure 5.1a, or on the side as in Figure 5.1b or it can also appear below the base text. Alternatively, some research has shown that the most preferred placement for larger annotations in a text document may be the margin [114]. Despite this finding, researchers have explored how to use and improve other placements. TextTearing [166], for example, allows readers to place annotations in the inter-line space via a lightweight interaction technique. Fluid Documents [26] similarly adds additional information into a text document by adjusting inter-line spacing. Fluid text includes links that expand in place to allow the text to still be read linearly while preserving the surrounding information. In an observational study Zellweger et al. [168] investigated different placement options for meta-information related to hypertext links (e. g. margins, footnotes, and in-line). They found a large variability in participants’ preferences but observed that, in general, it was desirable to keep meta-information close to the original link. Where information was placed also had a significant impact on reading speed with information close to the links being read significantly faster. An additional feature allowed meta-information to be “frozen” or kept visible in the view and participants used this feature regularly to compare different meta-information. This last feature is important for our design space and is one of the reasons we discuss both dynamic and static placements.

### 5.2.3 *Labeling*

The labeling problem is also related to our problem of placing word-scale visualizations in text documents. Bertini et al. [6] define labeling as attaching text labels to graphical marks to convey associated semantic information. Transferred to our problem, the entity is the graphical mark and the word-scale visualization is the label. Fekete and Plaisant [48] propose a taxonomy of labeling techniques with two categories: a) static labeling with the goal of finding a universal placement configuration for a whole visualization, and b) dynamic labeling where each label is treated as a dynamic object that can appear or disappear or change its location depending on the user’s interaction.

The transient visualization concept by Jakobsen and Hornbæk [80] relates to our work if we look at the labeling problem from the viewpoint of interaction and information-in-context. Their goal is to visualize information near the user’s focus of attention only when needed and without using extra screen space. This concept relates to our word-scale visualization placement problem since there are cases where a reader only wants to see the data graphics when needed. We, thus, discuss both dynamic and static placement options.

#### 5.2.4 *Readability of Text and Small-Scale Visualization*

Thus far, very little work has explored how word-scale visualizations might impact the readability of surrounding text. The most closely-related work is the previously mentioned study by Zellweger et al. [168] on Fluid Documents. The authors conducted eye tracking measurements and concluded that subjects were quite sensitive to different placement strategies but also that adjustments to document typography did not cause their point of focus to shift wildly.

Similarly, there is little existing research characterizing how small size affects the readability and perception of visualizations. For example, while several studies have compared the readability of different glyph designs (e.g. [52, 126]), we are aware of only one study (by Heer et al. [74]) that has explored the impact of size on readability. Heer et al. compared small line charts against two types of Horizon Graphs (1-band and 2-band) and investigated the impact of chart height on both designs. They found that small chart heights negatively impacted how accurately and quickly participants estimated the difference between two data points. The authors also saw that the Horizon Graph design had an influence on the amount of estimation error and that both horizon graph designs performed better than line charts. The authors conclude with the recommendation to draw line charts of 24 pixels (6.8mm) height and 2-band Horizon Graphs at 6 or 12 pixels (1.7 / 2.4 mm)—all vertical heights that fit the definition of word-sized graphics at common font sizes and display resolutions.

When designing small-scale visualizations, the choice of aspect ratio impacts the perception of trends and patterns in the data. The knowledge of how aspect ratios impact readability of different charts is relevant to our work, as some of our placement strategies are based on defining none, just one, or both of width and height of a word-scale visualization. Cleveland et al. [32] recommend banking to  $45^\circ$  to choose the aspect ratio of a line chart. This is due to the fact that an average orientation of  $45^\circ$  maximizes the discriminability of adjacent line segments. Banking to  $45^\circ$  selects an aspect ratio for a chart such that the average orientation of all line segments is  $45^\circ$ . Based on this work, Heer and Agrawala [73] presented a multi-scale banking technique which identified the different trends in the data at various

frequency scales to generate banked charts. Yet, Talbot et al. [143] showed that aspect ratio selection is not as simple as it may seem and that trade-offs between reducing errors for some tasks and increasing it for others have to be made. The authors provide a descriptive model of slope ratio estimation as a first step towards fully understanding the functions to consider when selecting an aspect ratio.

Finally, color perception is known to be impacted by changes in size [140] and, thus, is an additional important aspect to keep in mind when designing word-scale visualizations.

Next, we introduce the design space, its focus, and dimensions. Section 5.4 will detail several placement options and discusses trade-offs between word-scale visualization's placement options. In Section 5.5 we then discuss three examples that demonstrate the importance of the association between word-scale visualization and entity for the purpose of layout and interaction. Finally, in Section 5.6 we provide an in-depth analysis that examines how various placement options affect word-scale visualization placement in real documents. Based on this analysis, we provide recommendations that can help designers choose the appropriate word-scale visualization placement given their own constraints.

### 5.3 A DESIGN SPACE FOR WORD-SCALE VISUALIZATIONS







According to Schulz et al. [131] a design space consists of a finite number of design dimensions. Each of these dimensions captures one particular design decision needed to fully specify the whole. The dimensions of our design space describe constraints placed on the word-scale visualization design and text layout. Our goal is to provide a design-centered view on word-scale visualization placement that takes into account both the size of the visualization and the degree to which the visualization can disrupt the text.

#### 5.3.1 *Definitions and Framing*

We first provide a refresher on the definition of size of word-scale visualizations, introduce important terminology used in our descriptions, and situate the design space within a larger set of layout choices.

##### 5.3.1.1 *The Size of Word-Scale Visualizations*

For word-scale visualizations, we consider a wide range of visualizations that are “word-scale”—usually larger than the size of a letter, but smaller than a sentence or paragraph. For example, the height of a word-scale visualization can be restricted to font height as in most of Tufte's examples (Section 5.2.1) but can also be larger. Word-scale visualizations can also overlap or fall within the *inter-line spac-*

ing —the vertical white space between the bottom of one line of text and the top of the next—and/or the *inter-word spacing* —the horizontal white space between two consecutive words (see Figure 5.2). Allowing word-scale visualizations that are taller  and wider  than words introduces a number of alternative layout options, including using the existing inter-line and inter-word space, allowing visualizations to overlap with text , or changing the text layout to introduce additional space .

Similarly to the inter-line space, the space between two words is called inter-word space and consists of the standard spacing between words and any additional horizontal padding.

### 5.3.1.2 Static vs. Dynamic Integration

To place word-scale visualizations into text documents we have two fundamental options: static or dynamic integration. This is similar to the labeling problem as discussed in Section 5.2.3.

With *dynamic integration*, word-scale visualizations are displayed in response to user-input. In the most limited case, a user might hover the mouse cursor over an entity revealing a small visualization next to it. Alternatively, user interaction could be used to simultaneously reveal or hide all graphics in the document. The dynamic presentation of word-scale visualizations has the advantage that the text is not cluttered and can be read without the visualizations present. However, it also has several disadvantages: it requires an interactive environment and is not amenable to print media, showing visualizations individually can make comparisons more difficult, and—depending on placement options chosen—the text may have to reflow whenever visualizations appear or disappear.

*Static integration* on the other hand, means that visualizations are always present within the text. This strategy is well-suited to print media, and may also make it easier for users to compare word-scale visualizations (as noted by Zellweger et al. [168]). However, large numbers of word-scale visualizations might overwhelm or disturb reading the text.

Both static and dynamic integration entail trade-offs and it is up to the designer to make the right choice for their usage context. Our

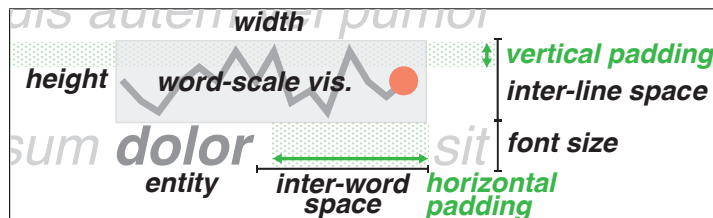


Figure 5.2: Important terminology used in the paper in the context of entity, word-scale visualization, and surrounding text.



Figure 5.3: The word-scale visualization has three different entities each using a color. The designer chose the numeral of the second entity as the main entity.

design space applies equally to static and dynamic placements but in this chapter we do not discuss interaction techniques to invoke a dynamic placement (this topic is discussed in more depth in Chapter 6).

According to the definition of word-scale visualization, they can have multiple associated entities such as in Figure 5.3. If there are multiple associated entities, either one of the entities is the “main entity” and has the most weight concerning integration or the entities have all the same weight. If there is a main entity, that entity drives the integration of the data graphic—the word-scale visualization is placed with respect to this entity. Why an entity is chosen as the main entity is highly dependent on variables such as the intended message the designer or writer wants to transmit or the given space constraints. When there is a “main entity” the same design rules as for the single entity cases applies. If all entities have the same weight then one entity can be chosen at random as the reference point for the integration. After one of the entities has been chosen the designer can proceed similarly to the single entity integration case. Other integration possibilities can exist, but the placement of a word-scale visualizations according to multiple entities is not the focus here. For the remainder of this chapter, we will focus on placement options where the word-scale visualization is associated with only one entity, irrespective of how this entity has been chosen.

### 5.3.1.3 Placement Context

We define three main placement options that relate to the contextual integration of word-scale visualizations in text. It has to be kept in mind that entities associated with word-scale visualizations are typically marked using a frame or color (see Section 4.6). According to our definition, context is based on the relationship between an entity’s bounding box and the bounding box of its associated data graphic. The three types of contextual placement are:

*Strong context:* The entity’s and visualization’s bounding boxes touch or overlap—e. g. a word-scale visualization is placed just above or next to a word.

*Weaker context:* The two bounding boxes do not touch, but the position of the visualization is still defined by the entity's position in the text—e.g. a word-scale visualization is placed in the same general text region as the entity, but the two do not touch. This placement context can be suitable for word-scale visualizations with multiple entities, but also for word-scale visualizations that are associated with a sentence (e.g. placed at the beginning of a sentence).

*Out-of-context:* Visualization position is not related to the entity position—e.g. a word-scale visualization is placed somewhere in the margin of the text.


Our design space concentrates on the strong context case where word-scale visualizations are placed in a sense “as close as possible” to the entity. We do not consider the weaker context and the out-of-context case as these introduce their own set of challenges related to linking the visualizations back to the entities. This is an interesting area of study in which there has already been some promising research (e.g. Steinberger et al.'s *context-preserving visual links* [138]).

### 5.3.2 Design Space Dimensions

The dimensions for our design space can be roughly divided into two categories: dimensions that have an impact on the word-scale visualization (1–4) and dimensions that impact the text (5 and 6):

*Control over the maximum height of the visualization:* The height of a visualization is closely related to the amount of information encoded in its vertical dimension. A designer may have limited control over the height of a visualization if it is intrinsically bound to a property of the text, such as the font height or the inter-line spacing. Full control over visualization height means that the designer can freely choose any desired height and can make height choices based on the visual encoding.

*Control over the maximum width of the visualization:* The width of a visualization is closely related to the amount of information encoded in its horizontal dimension. A designer may have limited control over the width of a visualization if it is bound to properties of the text, such as the width of the entity. Full control over visualization width means that the designer can choose the width freely.

*Word-scale visualization position:* The strong-context position of a word-scale visualization is defined by a reference point on the visualization's bounding box—we use the left bottom corner  throughout the remainder of the paper. Generally, the visualization can be placed anywhere around the entity as long as the two bounding boxes intersect. For simplicity we focus on three main positions:

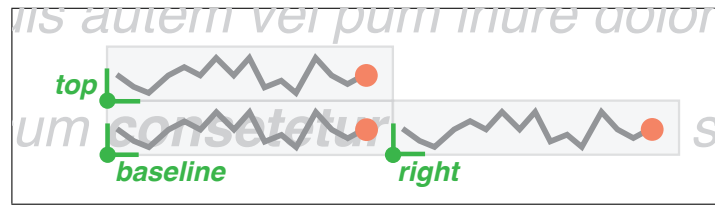


Figure 5.4: Visual description of the three placement positions: *baseline* position, *top* position, and *right* position.

- a) *baseline position*: on the left bottom corner of the entity's bounding box,
- b) *top position*: on the upper left corner of the entity's bounding box,
- c) *right position*: at the bottom right corner of the entity's bounding box (see Figure 5.4).

*Visual encoding*: The choice of visual encoding (e.g. line chart, bar chart, etc.) for the word-scale visualization has an effect on how much information can potentially be encoded within a given aspect ratio. The choice of data and visual encoding can thus affect how much control over width and height is necessary in order to ensure effective presentation of the data.

*Amount of inter-word spacing introduced*: Given a specific visualization width (freely chosen or defined through text properties), a designer can introduce additional inter-word space before or after the entity in order to control the amount of overlap between visualization and text.

*Amount of inter-line spacing introduced*: Given a specific visualization height (freely chosen or defined through text properties), a designer can decide to introduce additional inter-line space above or below the entity in order to control the amount of overlap between visualization and text. For example, additional inter-line space could be introduced to fit a visualization of given height between two lines of text without any overlap.

#### 5.4 WORD-SCALE VISUALIZATION PLACEMENT IN PRACTICE

When considering how to place word-scale visualizations, a designer may face several practical problems based on the characteristics of the document, the visualization, the data, and the usage scenario:

- **Is the text static, or can it be reflowed?** If word-scale visualizations are added to a scanned document, for example, it may not be possible to adjust the positions of words or phrases to accommodate the visualizations. However, in an electronic reading environment reflowing is easier and the inter-word spacing can be expanded to make space for word-scale visualizations.









case	control over max width of vis.	control over max height of vis.	introduce inter-word spacing	adjust inter-line spacing	overlap: vis.+ entity	overlap: vis.+ surr. text	possible collision between vis.	interaction potentially necessary	example placement
1	✓	–	✓	–	–	–	–	–	<div>uis autem vel pum iriure dolor</div> <div>m <b>consetetur</b>  sit amet,</div>
2	–	–	–	–	✓	–	–	✓	<div>is autem vel pum iriure dolor .</div> <div>m  sit amet, dolor .</div>
3	–	–	–	–	–	–	–	–	<div>is autem vel pum iriure dolor .</div> <div>m <b>consetetur</b>  sit amet, dolor .</div>
4	✓	–	✓	–	–	–	–	–	<div>uis autem vel pum iriure dolor</div> <div>m <b>consetetur</b>  sit amet,</div>
5	✓	✓	✓	✓	–	–	–	–	<div>uis autem vel pum iriure dolor</div> <div>m <b>consetetur</b>  sit amet,</div>
6	✓	–	–	–	–	–	✓	✓	<div>uis autem vel pum iriure dolor</div> <div>m <b>consetetur</b> sit <b>amet</b> color</div>
7	–	✓	–	–	✓	✓	✓	✓	<div>is autem vel pum iriure dolor .</div> <div>m  sit amet, dolor .</div>

Figure 5.5: Several word-scale visualization placements options and the design space decisions that produced them. A ✓ stands for “available or needed” and a – stands for “not available or not needed”.

- **Can the inter-line spacing be modified?** By increasing the space between lines, a designer can insert larger word-scale visualizations positioned above or below the text. However, adding inter-line spacing can also considerably increase the overall length and amount of whitespace in the document.
- **Should the visualizations be read along with the text?** If word-scale visualizations are intended to be read in the context of the sentences (as with the stock trends in Tufte’s original example [149]), the designer may wish to place them in-line with the text. If the word-scale visualizations provide supplemental information that could disrupt reading, positioning it in the inter-line space may be more appropriate.
- **How important are the visualization’s size and dimensions for readability?** The designer may need to enforce a minimum size for visualizations, e. g. for maps that contain small marks. Line charts and other slope-based visualizations may also need to be rendered using a particular aspect ratio in order to facilitate accurate reading and comparison as discussed in Section 5.2.4
- **Is there an appropriate visual encoding for the data?** Visualizations that include text labels, axes, or visually complex marks may not be effective at small scale, or may need to be re-interpreted to work within the available space. Various possible ideas for legends or captions can be found in Section 4.3.3.2

Answering these questions can be challenging, in part, because each individual choice may impact the size and layout of the document, as well as the readability of the text and word-scale visualizations. Moreover, the relationships between these decisions can be complex and their severity may depend heavily on the characteristics of the source document. For example, assume the following two designs:

**(A)** Visualizations are placed in the *baseline* position and rendered in front of their entities at 20% opacity, but are displayed at full opacity when hovered with the mouse.

**(B)** Visualizations are always shown at full opacity and placed to the *top* or *right* of their entities.

In Case A the visual encoding makes overlapping text and visualizations less problematic, but also makes comparison tasks more difficult. In Case B, comparison is easier. However, if there is not enough inter-line or inter-word space, overlap between visualizations and text *will* impede reading. A designer can avoid overlap by increasing the inter-line and/or inter-word spacing, but this can change the size and layout of the document considerably. To help illustrate the impact of and interactions between these design decisions, we discuss seven common placement options (illustrated in Figure 5.5) in greater detail.

**Case 1: Traditional In-Line Placement:** The most common placement strategy for word-scale visualizations is to increase the inter-word spacing before or after an entity and to insert the visualization in this space. This gives the designer control over the width of the visualization and avoids collisions between visualizations and text, but restricts the maximum height of the visualization to the font height plus the height of the inter-line space. Text may also need to be reflowed if adding visualizations increases line lengths past the page or column width, which may be problematic for print documents. In-line placement may also be undesirable if visualizations are hidden and revealed in response to user interaction, since hiding or revealing a visualization may change the layout of all subsequent text.

**Case 2: Overlay Word-Scale Visualizations on Entities:** Alternatively, we can choose to draw visualizations directly in front of or behind the entity. Since we introduce no additional whitespace, the layout of the text is not affected. However, the designer has restricted control over the width and height of the visualizations, which are limited by the width of the entity and font height plus inter-line spacing, respectively. This can make comparisons between visualizations difficult and can hurt readability if the entities are very short. In this placement strategy visualizations *always* collide with the entities they annotate, thus interaction may be necessary in order to disambiguate the two. For example, we can draw the visualization with a high transparency by default, but increase the opacity when a user hovers over it.

**Case 3: Using Existing Inter-Line Space:** For some documents it is possible to avoid layout changes and collisions by making visualizations the same width as the entity and positioning them within the existing inter-line space. In contrast to Case 2, this approach supports visual comparison between visualizations, but because visualization widths vary, those comparisons may still be difficult. Additionally, because the strategy constrains the visualization height to the current inter-line spacing, it is only viable for documents that are already widely-spaced, or that use very compact visual encodings.

**Case 4: Using Inter-Line & Increased Inter-Word Space:** To improve comparison and guarantee consistent visualization widths, we can combine the approaches from Case 1 and Case 3—placing fixed-width word-scale visualizations in the existing inter-line space, and increasing the inter-word spacing after the entity to avoid collisions. However, like Case 1, this approach means that text will be reflowed.

**Case 5: Increasing Inter-Line & Inter-Word Space:** We can extend the approach used in Case 4 and also insert additional inter-line space to accommodate visualizations with a chosen height. This provides the

designer with full control over width and height of the word-scale visualization, while still avoiding collisions. However, introducing padding requires text to be reflowed between lines and can increase the length of the document (especially if visualization width and height are large). Adding horizontal padding to increase the inter-line spacing means that the spacing is increased for the whole document and not for each line independently. This results in text that is potentially easier to read as the eyes always have to jump the same distance from line to line. Moreover, the composition of the text is aesthetically more consistent in contrast to the solution proposed by Chang et al. [25] where inter-line spacing is added for each line independently. When selecting word-scale visualization sizes, a designer must compromise between visualization readability and the level of text disruption. While we know of no previous work that has studied this trade-off in detail, we provide a series of disruption and readability measures in Section 5.6 that can be used to evaluate candidate designs and inform designers' decisions.

**Case 6: Allowing Visualizations To Overlap:** All of the previous cases have excluded the possibility of overlap between word-scale visualizations. However, relaxing this constraint introduces a number of new placement options. For example, by adding control over visualization width to the approach in Case 3, we can guarantee visualizations with consistent dimensions. This approach may be a good alternative if a text only has very few, widely spaced entities with visualizations. However, collisions between visualizations can appear when short entities occur in close proximity to one another (as seen in Figure 5.5). An interaction (such as hovering the mouse over one of the visualizations) may be necessary to resolve these kinds of collisions.

**Case 7: Allowing Visualizations to Overlap Surrounding Text:** Finally, while all of the previous examples have prohibited visualizations from overlapping surrounding text, this restriction can also be relaxed. Any of the prior cases can be extended by allowing the height or the width to extend beyond the padded boundaries of the current word or line. This relaxation makes it possible for some visualizations to collide with the surrounding text, as well as with other visualizations, and interaction may be required to disambiguate them. Thus, this placement option may be most suitable when larger widths and heights are necessary to make word-scale visualizations readable, and no alternative visualization technique is viable.

While illustrative, these seven cases represent only a few points in a much larger space of placement options. Each option can be further modified by additional design decisions not yet discussed here.

For example, a number of possible strategies exist for selecting maximum visualization widths. A designer might choose the word-scale visualization width based on the number of dimensions in the data, the length of the longest entity, by using perceptual guidelines for chart banking [73], or some other approach. The length could be chosen to be the same for all visualizations or chosen on an individual basis. In summary, the cases above are meant to outline the space and serve as starting points for selecting specific variations appropriate for a given setting, audience, viewing scenario, or reading task.

## 5.5 WORD-SCALE VISUALIZATION PLACEMENT EXAMPLES

In this section we describe three “real-world” examples that we created to illustrate different types of word-scale visualization embeddings. These examples are meant to demonstrate the diversity of potential applications of word-scale visualizations and spark creativity in their application to text documents. The interactions proposed in these three examples will be discussed in more detail in Chapter 6.

### 5.5.1 *Word-Scale Visualizations in Scanned Documents*

Figure 5.6 shows a clipping of a scanned transcription of a historical World War I artifact from one of the historians we collaborated with. The discussion that occurred around this project with the historians sparked some of the ideas discussed in this chapter. The text was scanned using a text recognition technique (OCR) to avoid manual transcription of the document. In this case, the historian was interested in knowing how exact the OCR technique was to potentially adjust the automatic transcription. In the specific example shown in Figure 5.6, every word became an entity as the OCR technique made a reasonable match for every word. To highlight uncertainty, we overlay rectangular bounding boxes on each entity and color-code them based on the uncertainty returned by the OCR algorithm (darker = more uncertain). When the mouse hovers over an entity, a tiny bar chart is drawn in the inter-line space, where each bar represents a possible alternative match. The height and color of each bar indicates the relative uncertainty of each of the alternative translations produced by the algorithm. Hovering over each bar shows the potential matching word and clicking a bar selects that match as the correct one.

In this example, one can consider the colored bounding box as one extremely simple word-scale visualization (a one-item heatmap) that has been placed at the baseline position and sized to the width and height of the entity. The small bar chart serves as a second word-scale visualization and is placed in the inter-line space. Together, the heatmap and bar chart illustrate two different placement strategies



## Science fiction

From Wikipedia, the free encyclopedia

For other uses, see [Science fiction \(disambiguation\)](#).

**Science fiction** is a genre of [fiction](#) dealing with imaginative content such as [futuristic](#) settings, futuristic [science](#) and [technology](#), [space travel](#), [time travel](#), [parallel universes](#), and [extraterrestrial life](#). It often explores the potential consequences of scientific and other innovations, and has been called a

Figure 5.7: Small word-scale visualizations that provide information scent to help viewers decide whether or not they are interested in following an article linked to from a Wikipedia page. Small bar charts show how frequently the page linked to has been visited in the last 30 days. The word-scale visualizations were embedded with inter-line and inter-word spacing to prevent overlaps. Detail-on-demand from each data graphic gives information about the exact visit count for each link.

to what designers proposed in Figure 4.15 and Figure 4.16 where they used a shared reference structure to support comparison and provide context. Here, we set the length of the reference line equal to the length of the bar for the most frequently visited page—in this case “futuristic” (which links to the Wikipedia page on “Future”). As in the previous example, interaction with the visualization is available. Hovering over the data graphic displays additional detail about the number of visits to the page in the last 30 days.

We set the height of the visualization equal to the height of the inter-line space and fixed the width of all visualizations at the average length of all entities in the text. In this example, we also added a small amount of vertical padding to reduce the level of visual clutter and ensure that the word-scale visualizations are associated with the entities below, rather than above them. To avoid overlap between visualizations we added inter-word spacing where the length of the visualization was longer than the length of its entity. We chose this position (Case 5 and not Case 4 due to the addition of vertical padding) because the text contains a large number of entities and the traditional placement in the inter-word space (Case 1) might disrupt reading and would cause many consecutive words to be far apart. This example further demonstrates an additional use case of vertical padding. Vertical padding is not only useful to get more vertical space for the visualization but can also help to reduce the level of visual clutter through emphasizing proximity.

## EASTERN EUROPE

## Soviet cult and pragmatism in Transnistria

Experts worry that the next  
 "Crimea" could be the  
 breakaway region of Transnistria.  
 Many locals there don't share that fear,  
 and if the last referendum holds, a large  
 majority would welcome a Russian  
 annexation.



Figure 5.8: Small maps embedded as word-scale visualizations provide details-on-demand for locations in a news story. Maps are embedded using a variant of the traditional (right) sparkline position. Interacting with a term or map expands it to show more detail.

### 5.5.3 Word-Scale Visualization for Detail-in-Context

Figure 5.8 shows a news story concerning the conflict between Russia and the Ukraine over Crimea. Within the news story the names of countries and regions are marked as entities. The word-scale visualizations for these entities are small geographical maps that are placed to the right of the word. The map visualizations provide location information that helps to put countries and regions in context. If a reader does not recognize a country from the small map, or wants more detail, they can hover over the visualization to reveal a larger version of the map with additional information. Here, we chose a variant of the Case 1 placement from Section 5.4 which introduces inter-word spacing and places each word-scale visualization slightly below the baseline position such that it is vertically centered with respect to the word. This slightly offset vertical position still avoids overlap between surrounding visualizations as long as the visualizations do not extend more than halfway into the inter-line space. In this example, adding word-scale visualizations required the text to be reflowed. However, because the number of regions mentioned in the text is limited and the map width is reasonably short, the amount of reflowing was minimal.

While it would be possible to add a larger map to the news article that shows all three regions in context, the word-scale visualizations have the benefit that they can a) show each region at reasonable size (both Crimea and Transnistria would become tiny when shown to scale next to Russia) and b) they save space in the article compared to a larger map in which all three regions would be visible. It would certainly be possible, however, to show the call-out of each map at a scale that included references to all three regions. For example, a call-



out for Russia could show small pins to indicate where Transnistria and Crimea are located—even if they are too small to see.

## 5.6 EVALUATION

Decisions about how to integrate word-scale visualizations into a text document have two main effects: a) they constrain the layout of the text and the design of the visualizations, and b) they affect the legibility and understandability of the document as a whole. While both are important areas of study, we concentrate first on exploring how different word-scale visualization placement options alter the size and layout of documents. We conducted a quantitative evaluation consisting of a simulation-based computational analysis in which we measured the impact of a number of different layout parameters, including visualization size and position, line height, and spacing. We ran the evaluation using real text documents and systematically varied the number of entities in the text in such a way that we could characterize the trade-offs between placement options for both sparse and dense distributions of word-scale visualizations.

### 5.6.1 Documents

We randomly selected a set of 15 news articles from NLTK’s ABC corpus of rural and science news [8]. These articles contained 100–246 words with average word lengths between 4.98–5.66 characters (median 5.29). We then created multiple versions of each article that contained varying numbers of entities. To create each new version, we automatically selected a number of single words in the original text to mark as entities. Even though we defined an entity as being one or multiple words the results can be generalized to multiple words. Multiple words could be seen as one long word with the same length as the single entity. We picked entities using a random log-normal distribution using a predefined average entity interval  $M$  (geometric mean) and 1.2 as the geometric standard deviation  $s$ , leading to  $\mu = \log(M)$  and  $\sigma = \log(s)$  as inputs to the log-normal random number generator. We generated tagged versions of each text using seven different entity intervals ( $M = 1, 5, 10, 15, 20, 25$ , and  $30$ ). This produced a final set of  $15 \times 7 = 105$  articles with entity intervals ranging from 1 (one word on average between entities) to 30 (thirty words on average between entities, or about one entity every two lines). We then rendered these texts as HTML documents. To simulate the layout constraints of a common news or print article, we rendered the text using 12pt Times New Roman and set the maximum width of the document to 17cm, the equivalent of a standard A4 page with 2cm margins. We chose

the font size based on work by Dyson et al. [44], who observed that 12pt type was read significantly more quickly than other type sizes.

### 5.6.2 Factors and Procedure

Using this set of HTML documents, we analyzed a series of different word-scale visualization placement options while varying five factors:

<b>word-scale vis. position</b>	<i>right / top / baseline</i>
<b>word-scale vis. width</b>	<i>entity len. / len. of shortest / len. of longest</i>
<b>word-scale vis. height</b>	<i>word height / 2.5×word height</i>
<b>line spacing</b>	<i>single-spaced / double-spaced</i>
<b>spacing adjustment</b>	<i>none / inter-line / inter-word / both</i>

The characteristic values for the line spacing, like font-size, were drawn from Dyson et al. [44]. For the visualization height we chose the word height (following Tufte's example) and also added the extreme case of  $2.5 \times$  word height to observe a condition that exhibited vertical overlap even with double-spaced text (as in Case 7, Figure 5.5). Combining these settings produced  $3 \times 3 \times 2 \times 2 \times 4 = 144$  design alternatives, including examples illustrating each of the cases given in Section 5.4. We then used a sequence of scripts to automatically apply the parameters from each design alternative to our set of sample articles and measured the changes in word-scale visualization application and document layout.

### 5.6.3 Measures

To quantify how the different placements impacted the size and readability of the document, we computed four different metrics:

*Change in Document Area:* The increase in the total area of the document following visualization insertion, padding, and reflowing (if necessary), as a percentage of the original area.

*Amount of Visualization Overlap:* The % of the surface area of the visualizations overlapped by other word-scale visualizations. As before, we count as overlapping area any place where the bounding boxes of more than one visualization overlaps.

*Amount of Text/Visualization Overlap:* The % of the surface area of the text in the document overlapped by word-scale visualizations. We count as overlapping area any place where the bounding box of one or more visualizations intersects the bounding box of a word.

*Text Shift:* The average change in the X- and Y-positions of individual words in the document following word-scale visualization insertion, padding, and reflowing, as a percentage of the original doc-

	Word-Scale Vis. Position	Word-Scale Vis. Width	Word-Scale Vis. Height	Line Spacing	Spacing Adjustment	Entity Interval	Expansion of Document Area				Vis./Vis. Overlap			Vis./Text Overlap				Text Shift X		
							0%	50%	100%	150%	0%	20%	40%	0%	20%	40%	60%	0%	10%	20%
Case 1	right	longest entity length	word height	double-spaced	inter-word only	1	125.9%											22.93%		
						5	25.1%											23.00%		
						30	4.8%											6.53%		
Case 2	baseline	entity length	word height	double-spaced	none	1								50.27%						
						5								13.40%						
						30								2.40%						
Case 3	top	entity length	word height	double-spaced	none	1	2.3%													
						5	2.3%													
						30	0.0%													
Case 4	top	longest entity length	word height	double-spaced	inter-word only	1	66.1%											23.07%		
						5	12.1%											15.60%		
						30	2.1%											2.13%		
Case 5	top	longest entity length	2.5 x word height	double-spaced	inter-line & inter-word	1	172.3%											23.07%		
						5	78.2%											15.60%		
						30	27.0%											2.13%		
Case 6	top	longest entity length	word height	double-spaced	none	1	11.1%				43.80%									
						5	4.4%				0.27%									
						30	0.3%				0.00%									
Case 7	baseline	entity length	2.5 x word height	double-spaced	none	1	3.2%				2.60%			64.47%						
						5	3.2%				0.53%			17.73%						
						30	0.0%				0.00%			3.33%						

Figure 5.9: Parameters used to produce the seven placement example cases used in our evaluation, along with average change in document area, visualization/visualization overlap and visualization/text, and text shift statistics for each. For each case, the table includes values computed for text with varying entity intervals: 1 (one word on average between visualizations), 5 (on average five words between), and 30 (on average thirty words between). The bar chart overlays are normalized by column and colored by row. Cells with only zeros have been omitted.

ument dimensions. Shifts in X occur when text is reflowed and moved within lines, while shifts in Y reflect both reflowing and changes to inter-line spacing. Since average shift in Y is linearly related to the change in document size, we do not report it independently.

#### 5.6.4 Results

As part of our quantitative evaluation we computed placement statistics for each of the 144 different design alternatives using each of the 105 source documents for a total of  $144 \times 105 = 15,120$  data points.

For the most part, the relationships between the different design dimensions are fairly predictable—as evidenced by the results in Figure 5.9, which highlights the results for each of our seven example cases. For example we see that Case 5—which introduces a considerable amount of vertical space to accommodate taller word-scale visualizations—causes the largest expansion of document area. Case 1—which adds horizontal padding and reflows the document by adding traditionally positioned visualizations, also adds space. Meanwhile, Case 4—which places word-scale visualizations in the existing inter-

line space and adds horizontal space only to prevent overlaps on longer words—adds less total area. Because Case 1, Case 4, and Case 5 all add some horizontal space, all shift the average X-position of the words considerably, especially as the density of entities increases. However, the other four cases—which do not alter the inter-word or inter-line spacing—do not cause the text to shift at all. Instead, they increase the document size only where their visualizations overrun the original margins of the document.

Similarly, the placement options in Cases 2, 6, and 7—all of which allow overlaps—result in more frequent collisions between elements. In Cases 2 and 7, which place word-scale visualizations over the entity, the overlap between text and visualization increases fairly linearly with the number of entities (see Figure 5.9, Visualization/Text Overlap). Meanwhile, Case 6—which allows overlap between adjacent charts—exhibits very few visualization collisions until the density of data graphics becomes high (Visualization/Visualization Overlap).

#### 5.6.4.1 *Comparing Vertical and Horizontal Space Adjustment*

One interesting design trade-off appears when deciding where to insert space to accommodate a word-scale visualization. One can, for example, increase the inter-line space if the word-scale visualization does not fit above the entity, thus increasing the overall length of the document. However, if instead we add *inter-word* space and insert the visualization to the right of the entity, the text that follows it must be reflowed. Because word-scale visualization lengths are usually a small fraction of the line width, it is often possible to add multiple inter-word visualizations without increasing the length of the document. However, when a new line is necessary, the document must grow by the height of an entire line plus a full inter-line space.

The relative space-efficiency of these two placement strategies depends on several factors, including the amount of existing inter-line space and the frequency with which entities occur in the text. To explore the trade-off in greater detail, we ran a second experiment where we simulated placing word-scale visualizations into texts with varying inter-line spacings and numbers of entities. We fixed the visualization height to the font height and tested three common variations for visualization width (the width of the current entity, the width of the longest word, and the width of the average word). Based on observations from our larger study we focused on inter-line spacings between 1.5 (one-and-a-half-spaced) and 2.0 (double-spaced). We also used the same set of 15 source documents but focused on cases with higher numbers of entities (entity intervals  $M = 1$  to  $M = 10$ , inclusive).

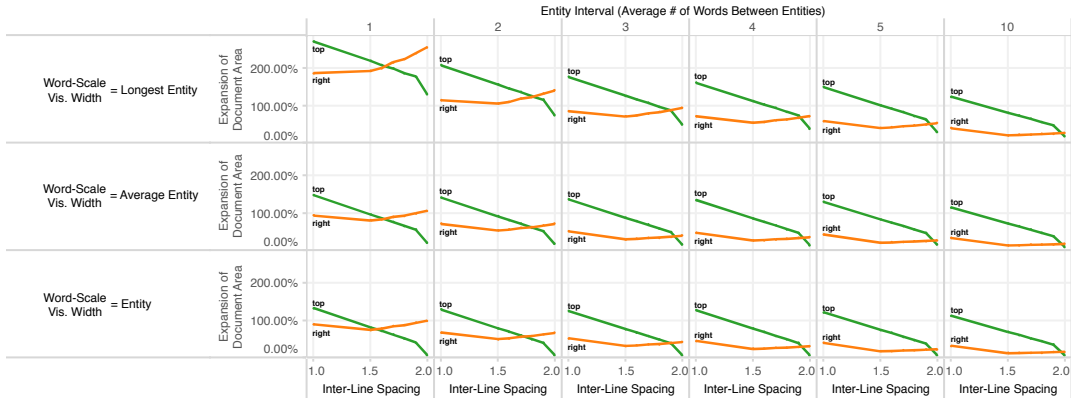


Figure 5.10: The effect of word-scale visualization placement strategy on the overall expansion of the document as more visualizations are added. Each small multiple illustrates the trade-off between inserting space for visualizations to the right of the entities they annotate (in orange) and inserting space for them above their entities (in green), for a particular combination of visualization width, inter-line spacing, and entity interval. Intersections between the lines highlight conditions where the relative space-efficiency of the two placement approaches changes.

**word-scale vis. position** *right / top*

**word-scale vis. adjustment** *inter-line / inter-word*

**word-scale vis. width** *longest entity / average entity / entity*

**line spacing** *1.0 / 1.5 / 1.6 / 1.7 / 1.8 / 1.9 / 2.0*


Our results (Figure 5.10) highlight the relationship between the two placement options, as well as inter-line space and word-scale visualization width. When the initial inter-line spacing is small, inserting visualizations to the right of entities can often considerably reduce the growth of the document—often by a factor of two or more. However, in cases where the inter-line spacing is already close to the visualization height and where the number of entities is high, positioning word-scale visualizations above the target becomes more efficient. This is especially true when longer visualizations are used, since these are more likely to trigger reflow events that add additional lines.


## 5.7 CONSIDERATIONS FOR WORD-SCALE VISUALIZATION PLACEMENT

Based on our theoretical and quantitative analyses of the design space, as well as our experience building our word-scale visualization placement test tool and examples, we discuss several considerations for placing word-scale visualizations.

### 5.7.1 Placement Advice

One of the main problems when choosing a placement for word-scale visualizations is the wealth of inter-related factors. When the placement constraints are well-defined (as in our OCR example in Section 5.5.1), the space of options becomes relatively narrow and it is easier to make an informed decision. If there are no constraints, the options may become very overwhelming. In this case, we offer two general points of advice from our quantitative analysis:

**Use inter-line space where available** . If the inter-line space is sufficiently tall to accommodate the desired visual representations, placing them above the entity in the inter-line space is often the best choice. This placement strategy reduces the need to reflow or expand the text and—unless visualizations have very long aspect ratios—it typically results in little overlap. For some small visualizations like Horizon Graphs and small line charts Heer et al. [74] give experimentally-validated minimum sizes small enough to fit between 1.5- and 2.5-spaced type. Unfortunately, similar readability or minimum size guidelines do not yet exist for most chart types, so the discrimination of whether the space is sufficient rests in the hands of the designer.

**Placing visualizations to the right (usually) saves space** . If text spacing must be modified, adjusting the inter-word spacing and placing visualizations adjacent to their entities is almost always more space-efficient than inflating the inter-line spacing to accommodate them (provided visualization heights and aspect ratios are word-like). As noted in Section 5.6.4.1, we found that adding additional inter-line space to position visualizations above their entity was more space-efficient only in a few cases—typically when the inter-line space was already very close to the desired height or when the number of entities in each line was very large. Right-side placement also has the added benefit that visualizations can typically be taller, including the height of the font plus whatever inter-line space is available.

## 5.8 SPARKLIFICATOR

As part of this research, we have built a general JQuery open-source library [64] called *Sparklificator* (see Figure 5.11) that eases the process of integrating word-scale visualizations into HTML documents, and provides a range of options for adjusting the position, size, and spacing of visualizations within the text. The library includes default visualizations, including small line and bar charts, and can also be used to integrate custom word-scale visualizations created using web-based visualization toolkits such as D3 [11]. Sparklificator implements all of

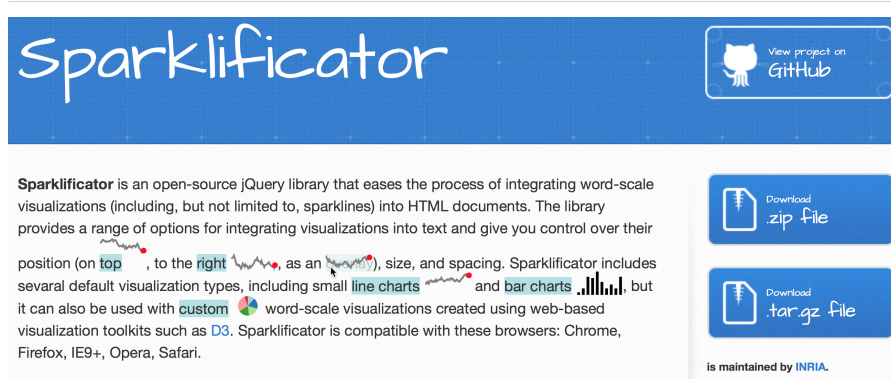


Figure 5.11: Sparklificator can be found on GitHub under <http://inria.github.io/sparklificator/>.

the seven placement cases proposed in Figure 5.5. The source code of this library can also be found in Section B.1.

Sparklificator adds word-scale visualizations to targets. Most of the time the target will be the entity or the text part the word-scale visualization is associated with. These targets can be any jQuery selection, but are most often spans containing one or more words. For example Sparklificator can add a word-scale visualization to the span around the target (word) "Pellentesque" in the sentence below.

---

```
1 Vivamus porttitor <span class='spanToSpark'>Pellentesque</span> habitant morbi tristique.
```

---

Source Code 5.1: HTML text with span around the entity.

To add a word-scale visualization of yearly rainfall to "Pellentesque" we use the following JavaScript:

---

```
1 var rainData = [20,20,25,40,45,40,46,50,60,80,85,92,99,98,101,102,92,100,105,97,95,82,73,
2               66,55,64,74,98,98,99,93,94,88,85,91,85,50,50,45,40,45,86,89,95,98,100,97,92,89];
3
4 var settings = {data: rainData, renderer: classicSparkline, position: 'top',
5               paddingHeight: true, paddingWidth: true, width: 80, height: 25};
6
7 $('spanToSpark').sparklificator();
8 $('spanToSpark').sparklificator('option', settings);
```

---

Source Code 5.2: JavaScript to add a word-scale visualization to the entity *Pellentesque*.

The *settings* object describes what settings are used within Sparklificator. The settings specify the visualization's *data* (here, the array "rainData"), *size* (here 80px by 25px), *position* ("top"), and *renderer* (here, we specify a small line chart using the "classicSparkline" ren-

derer from the `renderer.js` file). Besides a basic JavaScript type or a variable, each of the settings can also correspond to a function returning the appropriate output type for the setting. For example the width setting can be a function that calculates the width of the current entity and gives out its size. Each entity will then have a word-scale visualization that has the length of the entity. The `renderer.js` file includes all the visualization renderers provided by Sparklificator. The settings also specify whether to add *padding* around the visualization or allow the visualization to *overlap* the text (here, we add vertical and horizontal padding). The last two lines use jQuery to select the target—in this case `$(.spanToSpark')`—and adds the word-scale visualization with these settings. The result of Sparklificator can be seen in Figure 5.12.

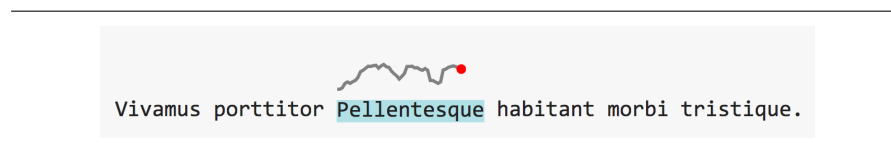


Figure 5.12: Output by Sparklificator.

## 5.9 CONSIDERATIONS FOR FUTURE WORK

There are multiple research direction to tackle for future work starting from the proposed design space of placement options for word-scale visualizations. I picked four of them that I discuss in more details: the investigation of the two not yet discussed placement contexts—*weaker context* and *out-of-context*, strategies to gain space to integrate word-scale visualizations, integrating word-scale visualizations during the document lifecycle, and the legibility of word-scale visualizations.

### 5.9.1 *Weaker Context and Out-Of-Context Placements*

In the current chapter we focused on the strong context and there are multiple cases we did not cover explicitly in our design space. As we have seen in the authoring study from Chapter 4 entities can be very diverse in size—from a short word or number (see Section 4.3.3.1) over a sentence to a whole paragraph (see Figure 4.10)—and structure such as multiple entities associated with one word-scale visualization (see Figure 5.3), entities covering multiple lines (see Figure 4.14a), or even disconnected entities (see Figure 5.3). Entities can also be sentences or paragraphs, thus, how do we position word-scale visualizations in these extreme cases? Some initial placements were already suggested in Chapter 4 like placing a word-scale visualization close



to the beginning of the sentence to emphasize that the sentence is the entity. Another case we encountered in Chapter 4 was the designer placing the word-scale visualization along a paragraph to suggest the association of the word-scale visualizations with the paragraph. We also saw a case where the title of a section was used as a placement proxy that stood for the entire section (see Figure 4.11). There are various other placement options that can be thought of when considering longer sizes for entities. It would be interesting to investigate guidelines for designers and thinking about how one could potentially automate the placement of word-scale visualizations in an authoring system.

At the beginning of this chapter we also highlighted two additional placement contexts which we did not investigate further: weaker context and out-of-context. Thus it would be an interesting avenue to investigate in which cases these placement contexts would make sense. These two contexts suggest a less strong proximal connection between the entity and the word-scale visualization. Thus, there is also interesting research to investigate how this connection between entity and word-scale visualization can be maintained with for example other gestalt laws of grouping. An overview of these gestalt laws can be found in the work of Sternberg and Sternberg [139]. In Chapter 4 some designers already suggested ideas for weaker proximal connections between entity and word-scale visualization such as frames or color.

### 5.9.2 Gaining Space

In this chapter I only addressed strategies such as adding inter-line space or inter-word space to introduce additional space to fit in word-scale visualizations. There are other options such as reducing inter-line space to accommodate word-scale visualizations in a particular text line. This is similar to the idea proposed by Chang et al. [26] where each line of text is individually moved up or down to compress inter-line space (see Figure 5.13). Such a strategy would be suitable when word-scale visualizations are shown on demand or if there are only few word-scale visualizations embedded in the text. However not enough space might be gained and additional inter-line space might have to be added. Still, less inter-line space would have to be added than with our strategy proposed in Section 5.4 – Case 5. In Case 5 we proposed to augment inter-line space for the whole document to offer a coherent view of the text. Thus, each text line has the same inter-line space. The strategy proposed by Chang is to only change inter-line space where it is needed, and where it is not needed to even reduce the inter-line space.

Another strategy to gain space is to slightly reduce or augment the letter spacing. This could be employed to either generate space be-

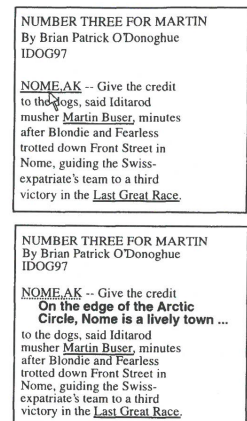


Figure 5.13: Top: user hovers over NOME,AK. The underline signals there is additional information. Bottom: inter-line space is reduced between all other lines to fit in the annotation below the underlined word. Image credit: [167].

tween words to accommodate in-line word-scale visualizations or to accommodate longer word-scale visualizations over the entity. There has been previous work on the effects of increased letter spacing (intra-word spacing) such as Paterson and Jordan's [113] that shows that increasing letter spacing has an effect on word identification and eye guidance during reading but that this effect can be modulated by inter-word spacing. Similarly, Slattery and Rayner [135] show that intra-word spacing can be reduced as long as inter-word spacing is increased without impeding reading. These two papers suggest that it is possible to do space optimization. However, there is further research needed to completely understand how much optimization can be done. For us, slightly increasing or reducing intra-word or inter-word spacing might already result in enough spaced gained to fit in the word-scale visualizations.

Other strategies could be to modify the font-size for the entity and keeping the surrounding text as it is. This could help to gain space between words and above the entity. However, these are very targeted changes and might not lead to massive gains, but it can be enough to fit a slightly higher word-scale visualization. In summary there are unexplored strategies to gain space to fit word-scale visualizations into text and which would be interesting to investigate further.

#### 5.9.3 *Word-Scale Visualizations and the Document Lifecycle*

The integration process of word-scale visualizations can happen at different stages of the lifetime of a document. The first stage in the lifecycle of a document is writing the text. If at this early stage the writers want or need to include word-scale visualizations, they often can still easily change content, form and structure of the text to optimally accommodate them. For an already written text, someone other than the writer may decide to include word-scale visualizations. This person could for example be a designer or an editor. They may not have the cooperation of or contact to the writer and thus have no control over the text content. Consequently, they cannot simply change the structure of a sentence to better accommodate the visualizations. However, they might be able to change the form of the text, for example changing font-size or inter-line spacing. Thus, there are different stages a text can be in and different constraints to take into account when integrating word-scale visualizations into text. It would be interesting to further investigate how these different stages of the lifetime of a text can influence the placement of word-scale visualizations.

#### 5.9.4 *Legibility of Word-Scale Visualizations*

The placement of a word-scale visualization depends on the size—width and height—of the word-scale visualization. The smaller the

word-scale visualization can be, the easier it is to choose a placement option that does not significantly influence the layout and flow of a text. We slightly touched on the size constraints for readable small visualization in Section 5.2.4. Similarly, to show trends in data it can be suitable to use banking techniques, that based on width and data, can deduce the relevant aspect ratio. With the aspect ratio, a suitable height can be found to show the trend that is inherent to the data of the chart. However, we did not extensively investigate these tradeoffs between sizes for readable visualizations and sizes that allow an optimal fit of the visualization into the text. It would be useful to have guidelines on size constraints for common visualizations to be still legible for designers that want to design new word-scale visualizations. However, the best would be to be able to parameterize a word-scale visualization authoring system with these empirically validated size constraints. The system would then provide multiple scales of the visualization and even suggest placement options depending on the sizes. Thus, better understanding how size and legibility of word-scale visualization is connected is an interesting avenue for future research.

## 5.10 SUMMARY

In this chapter, we described a space of options for designing and placing word-scale visualizations in text. We focused, in particular, on how word-scale visualization placements constrain the design of visualizations and the layout of documents. We illustrated the space of alternatives by discussing seven common placement choices and described how they can be varied. We also showed three real-world examples that suggest how different placement options can be applied in practice and highlight the richness and diversity of possible word-scale visualization placements, applications, and designs. In order to measure the design-related impacts of word-scale visualization placements, we conducted a quantitative evaluation consisting of a simulation-based computational analysis that measured the effect of different layout parameters, including visualization size, position, line height, and spacing on text layout. Finally, we proposed several general design considerations for placing word-scale visualizations.

We did not further investigate the weaker context and the out-of-context placement options. These two placement contexts are similar to the problem of external labeling. The main design issues with these two context placements is how to connect or visually link the word-scale visualization and its associated entity. The reader needs to know to which entity the word-scale visualization is associated to as the entity usually contains important information to understand the visualization. In Chapter 4 and more specifically Section 4.3.3 we already saw designers using weaker context placements without ex-

plicitly linking the two. Thus, investigating when and how to best use word-scale visualizations as external labels is an interesting avenue for future research.

Certain placements of word-scale visualizations can lead to situations where the reader has to interact with the text to read or the word-scale visualizations to see the data. For example, when the designer decides to use the overlay placement option, the text has still to be readable. Thus, the designer is rendering all word-scale visualizations at a low opacity to reduce visual clutter, but for the reader to be aware that there are visualizations to look at. However, when the reader wants to have a better look at the word-scale visualizations he/she needs to increase their saliency. To enable the change of opacity, interaction can be used. The next chapter discusses how interaction can be used together with word-scale visualizations.

## INTERACTING WITH WORD-SCALE VISUALIZATIONS

---

In previous chapter, I discussed the use and design of word-scale visualizations based on a study with designers. In this study in Chapter 4 some designers suggested interaction to connect multiple word-scale visualizations, to support comparison between word-scale visualizations, to support showing word-scale visualizations on demand and to interact with the word-scale visualization’s data. These considerations requires a more in-depth exploration of how interaction can improve the use of word-scale visualizations.

In this chapter I explore design consideration for using interaction with word-scale visualizations. For that purpose I categorize and organize possible interactions in a design space, discussing for example what can be affected by a given interaction—word-scale visualization or entity. Furthermore, I propose a set of interaction techniques to support the exploration and analysis of data associated with entities in text documents. This chapter ends with an examination of how interactive transitions may help bridge the gap between document-centric and visualization-centric analysis.

Parts of the first half of this chapter have been previously published in Goffin et al. [61]. Other parts come from a publication in progress by Goffin et al. [59]. Hence, the use of “we” in the remainder of this chapter refers to Pascal Goffin, Wesley Willett, Jean-Daniel Fekete, and Petra Isenberg.

### 6.1 MOTIVATION

Most past examples of sparklines and word-scale visualizations have been static. However, nowadays a lot of text can be accessed through computers, over the internet using common formats such as PDF or HTML and through dedicated reading environments like a PDF viewer, or other electronic devices such as ebooks. For example research articles are usually available through digital libraries (e. g. IEEE) in PDF format, some of them even propose research articles as HTML documents. In all these digital environments interaction is possible allowing for various manipulations for example on word-scale visualizations in the text. In addition, interaction can help to deal with the limited space available for the rendering of word-scale visualizations. For example, interaction can be used to provide more screen space by shrinking associated text or by augmenting the size of the visualizations. Further interactive operations on word-scale visualizations

include those that provide word-scale visualizations on demand and those that refine or modify what a reader can see in the word-scale visualizations. There is a multitude of ways to interact with word-scale visualizations, however past work on interactive word-scale visualizations is scarce. Before categorizing and organizing possible design considerations for word-scale visualizations, I provide relevant background on using interaction with word-scale visualizations and how other interactive analysis tools have begun to explore alternative means of integrating text and data.

## 6.2 BACKGROUND

Interaction is an important component of Information Visualization. When using visualization to find appropriate answers to formulated questions tools to drill down into the data, to refine the data for example by filtering out unwanted data points, and to be able to go back to higher level views are welcome. Moreover, interaction can be used to support information-seeking behavior. Shneiderman [133] brought these ideas together and suggested the “visual information-seeking mantra”: “Overview first, zoom and filter, then details-on-demand.” These operations can be supported through interaction.

Our work on interaction relates closely to Bret Victor’s “Explorable Explanations” [153], where he promotes the active reader, reactive documents, and the addition of contextual information to given texts. Victor based his “Explorable Explanations” on the premise that the reader is an active reader. An active reader is critical about what they read, question the proposed assumptions, ask questions when something is not clear or suspicious and consider alternatives. It is a reader that calls the content into question and wants to know more about the described data in the text. For example an active reader who reads a newspaper article about an in-depth analysis of upcoming changes in the law on income taxes wants to understand and get insights into the model behind the dependency income tax and GDP. Reactive documents are documents in which the reader can influence the rendition of embedded statistics. For example, when presenting a document showing the income tax level (e.g., 19%) of a country and its total tax revenue (e.g. 36.5% of GDP), dragging the mouse left or right over the income tax text would recalculate and update the total tax revenue displayed. To create such reactive documents Victor implemented a JavaScript library called Tangle [144]. Tangle uses a similar concept to our Sparklificator library. The developer declares the interactive variables using HTML attributes and uses JavaScript to defines how these variables are calculated. Finally Tangle connects the variables and proposes the interaction. Victor also describes how interactions could be used to elicit additional contextual information from within documents; for example, hovering over a word while

pressing a key can trigger a search on Wikipedia and show results in a box over the text.

Similar to Victor’s approach, Fluid Documents by Chang et al. [25, 167] include interactions where the text can be expanded, collapsed, and shifted to suit the current task. Our interaction techniques described later were in particular inspired by this previous work. In Fluid Documents the authors differentiate between primary information in the text and supporting information like references, derivation of formulas, and descriptions of technical terms. They add this supporting information as graphical annotations to the text. Hovering over a marked phrase causes the annotation to grow in size to improve readability. The text is reflowed to provide space for the growing annotation.

Research articles proposing interaction on word-sized data graphics are scarce. One example by Beck et al. [3] uses sparklines in source code; in their tool, hovering over the sparkline results in the visualization being enlarged and shown with more detail in a tool tip.

Perin et al.’s interactive horizon-graph [117] can be used as a word-scale visualization. It is designed to be very compact and implements several interaction techniques such as baseline panning and zoom. Baseline panning allows for interactively adjusting the horizon-graph’s baseline along the y-axis while value zooming specifies the zoom factor. Both are triggered using a continuous dragging interaction occurring over the visualization.

Although word-scale visualizations have mostly been used to add information to static text, other interactive analysis tools have begun to explore alternative means of integrating text and data. Systems like Popout Prism [142] and VisRa [109] aid document analysis by using color, size and other visual attributes to encode information about readability, sentiment, and important entities within the document text. In contrast to methods like these that encode information by modifying the text rendering itself, other work has explored encoding information outside of a document and using graphical methods to reconnect visualizations and text. For example, Steinberger et al. [138] connect entities to marks in nearby visualizations by drawing “context-preserving” graphical links which they describe as explicit edges that connect related pieces of information. These visual links however introduce visual clutter. Therefore, they developed an algorithm that determines what are salient regions and routes the links around them. Waldner et al. [156] propose a system that uses visual links to connect content from different and independent application. Geymayer et al. [56] extended this previous work to also include visual links to hidden and potentially overlooked content that is for example occluded or scrolled away. Through interactive brushing of text and highlighting of corresponding marks in a visualization, Kong et al. [88] similarly establish a text-to-graphic relationship.

Like these previous works, we embed data in text documents, yet at a smaller word-scale. We also employ interaction to connect text and data graphics but support fluid ad-hoc switching between reading text and exploring the embedded visualizations. Boy et al. [14] discuss in more detail how non-word-scale visualizations embedded in text document can suggest interactivity and data exploration. Suggesting interactivity is relevant to our work, because word-scale visualizations are supposed to be integrated into different kind of texts. Therefore, letting the reader know that there is interaction possible to explore the data in the word-scale visualizations has to be considered.

Our interaction techniques also take inspiration from a variety of other visualization tools that help analysts interactively organize and compare elements within a visualization. In particular, our approach builds on “Bring-and-Go” interactions [103], which allow analysts to explore network diagrams using their visual marks. We also take inspiration from approaches like GraphDice [7] and Small MultiPiles [2], which provide interactive tools for organizing and comparing scatterplot matrices and small multiple displays. Unlike these systems, however, we focus on organizing and manipulating word-scale visualizations while maintaining the context of a text document, rather than providing interactions in a separate visualization environment.

Finally, our work relates to the visualization literature on interaction design spaces. In that category Heer and Shneiderman [75] propose twelve task types that designers or developers can base their work on if they need to create or evaluate visual analysis tools to support information-seeking. Similarly, Yi et al. [165] propose seven general categories of interaction intents that are commonly used in Information Visualization. In contrast to our design space, however, these articles focus only on interactions in visualization-centric views, while we investigate interactions in document-centric views.

### 6.3 DESIGN CONSIDERATIONS FOR INTERACTION

In this section, we discuss where and when to trigger interactions with word-scale visualizations and which interaction techniques can be used in a document-centric space (Section 1.1). A document-centric space is the space where the document or the text is the central point of attention of the reader or analyst. We also consider how to interactively transition from a document-centric view to a visualization-centric view. In the visualization-centric space the text is not central anymore. Abstract views on the data gained from the text or document for example through text-mining algorithms is now important and the key for further analysis. Therefore, this data is visualized so the reader or analyst can better make sense of that data. We, how-



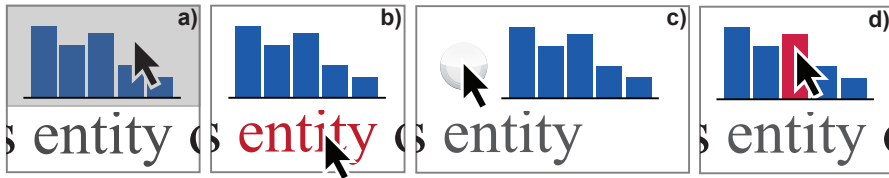


Figure 6.1: Interactions can take place directly with the entire visualization (a), the related text entity (b), connected GUI controls like buttons (c), or with individual marks in the visualization itself (d).

ever, concentrate on the document-centric space and discuss where interaction can be captured.

### 6.3.1 Where to Capture Interaction

We differentiate interaction with the entities and the word-scale visualization as a single component vs. interaction with the underlying data and its individual marks. Figure 6.1 shows four main locations that can capture interactions that affect word-scale visualizations: a) the area holding the word-scale visualizations, b) the entity, c) a button or other GUI components, and d) individual data marks of the word-scale visualization. We will not further discuss possible interactions with the underlying data marks (Figure 6.1d) as other interaction design space discussions have amply covered the topic [75, 165].

Once an interaction area has been defined, one must consider why and how an interaction will be triggered or what the result of an interaction will be.

### 6.3.2 Interaction Intents

We base our interaction intents for the document-centric space on Yi et al.’s [165] seven general categories of interaction: select, explore, reconfigure, encode, abstract/elaborate, filter and connect. All of them can be applied to the space of document-centric interactions.

*Increasing Saliency:* In order to reduce visual clutter, a designer may decide to draw all word-scale visualizations at medium or low opacity. To see a visualization more clearly, the reader can then interactively activate the visualization using an interaction techniques such as hover (see Section 6.3.3) and the reading environment will draw the visualization at full opacity.

The intent to *select* can involve choosing a particular (or multiple) word-scale visualizations for a specific purpose. In some cases, word-scale visualizations may not have to be constantly present. Using interactive controls, the reader can decide when to display them and for example hide or unhide them. This may be done at the document level, by including controls that reveal or hide all visualizations. Alter-

natively, visualizations can be revealed on a per-entity basis by clicking or hovering over the entity. A non-data-driven word-scale graphic or other visual highlight can be used to suggest this possible interaction to readers as discussed by Boy et al. [14]. This non-data-driven word-scale graphic can also serve as a way to communicate what lies in a hidden word-scale visualization. One designer in our study in Chapter 4 suggested a similar intent which was to reveal word-scale visualizations on demand (Section 4.3.5). Similarly, in order to reduce visual clutter, a designer may decide to draw all word-scale visualizations at medium or low opacity. To see a visualization more clearly, the reader can then interactively activate the visualization (select) using an interaction techniques such as hover (see Section 6.3.3) and the reading environment will draw the visualization at full opacity. Another example of where the selection intent is useful, is when resolving conflict between word-scale visualizations. In Chapter 5 we proposed to let word-scale visualizations overlap each other or the surrounding text, if adjusting inter-line spacing and/or inter-word spacing is not possible due to constraints in the document. In these cases interactions can be used to disambiguate them—for example, by bringing the one under the mouse cursor (select) to the foreground.

The desire to *explore* a word-scale visualization may in particular be a trigger for moving between document-centric and visualization-centric views to see more data in greater detail. Another possibility for exploring is to get details on demand, for example when word-scale visualizations have interactions of their own similar to larger visualizations. In Section 5.5 we showed several detail-on-demand cases where hovering the mouse over a visualization or even some of its individual components brought up additional information.

However, occasionally a word-scale visualization may provide too much detail (e. g., on a small map) in which the intent to *abstract/elaborate* may involve zooming out.

When a reader wants to change the location of the word-scale visualization, the intent is to *reconfigure* by, for example changing from a strong to a weaker context placement (see Section 5.3.1.3).

The *encode* intent involves the substitution of a word-scale visualization for another visual representation of the same data (e. g., a bar chart instead of a line chart).

When selectively showing visualizations according to their entity type (e. g. persons) then the intent is to *filter* the set of word-scale visualizations in the document. Entities can be categorized into categories (type) depending on the entity's attributes such as its grammatical association, meaning, or associated data. This categorization can be done manually by an analyst or automatically using for example text mining. One designer in our study from Chapter 4 proposed to use labels of the word-scale visualization to do filtering between alternative views of the word-scale visualization (see Figure 6.2). In

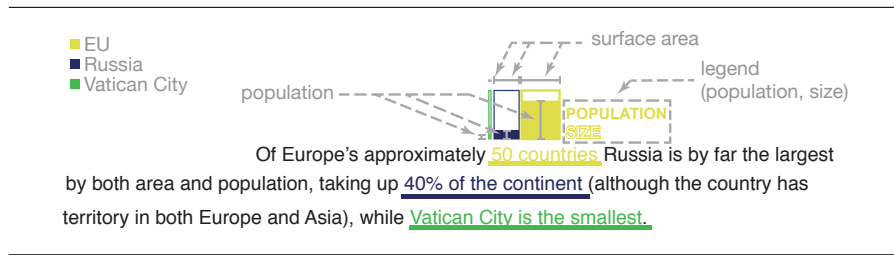


Figure 6.2: Word-scale visualization showing the contrast between population and surface area for the EU, Russia, and Vatican City (we added the grey dashed lines and annotations for explanatory purpose).

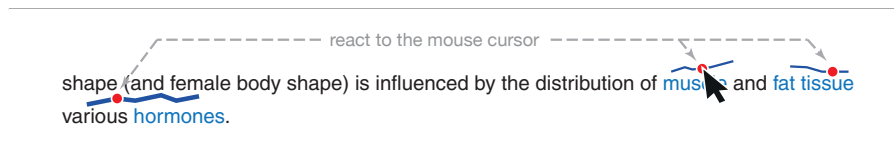


Figure 6.3: Three connected word-scale visualizations. Moving the cursor in one graphic results in highlighted values in the others (we added the red dots and the cursor to explain the interaction)

Figure 6.2 hovering over the label for “population” or “size” would toggle between the two dataset that can be displayed in the chart.

The *connect* intent comes into play when the reader wants to understand the relation between entities, for example by representing metrics for more than one entity in a single word-scale visualization. Connect could also refer to connecting or merging word-scale visualizations to generate new ones, like creating a diff of two word-scale visualization. The designers in our study in Chapter 4 also suggested connect intents, for example one designer suggested to connect word-scale visualizations using brushing and linking to see the relationship between entities (see Figure 6.3). Another designer suggested to connect two word-scale visualizations to be able to compare their data (see Figure 6.4).

### 6.3.3 Interaction Techniques

The main interaction techniques with word-scale visualizations in a desktop setting are *clicking*, *hovering*, *dragging*, *panning*, and *contextual menus*. *Clicking* is useful to select word-scale visualization or to trigger events in GUI components such as a button that turns word-scale visualizations on or off. Similarly *mouse hover* can be used to highlight or resize a word-scale visualizations to reveal more or less information as done in Fluid Documents [25, 167]. *Dragging* can also be used to resize a visualization, to zoom in/out in a map visualization, to trigger specific interactions like in the interactive horizon-graph [117],

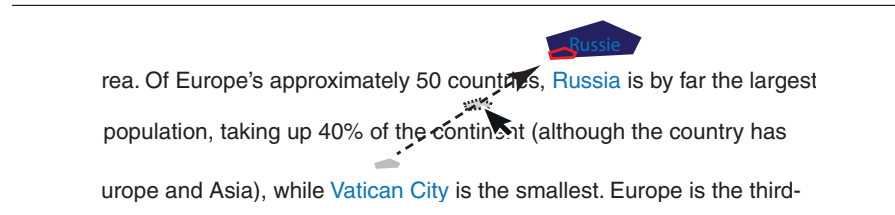


Figure 6.4: Interaction triggers the overlay of two country word-scale visualizations (we added the line between the two graphics and the cursor for illustration purpose).

or to move a word-scale visualization for example to initiate an action. An example, where moving a word-scale visualizations, has been proposed by a designer where one word-scale visualization is dragged onto another word-scale visualization to combine the two in one chart (see Figure 6.4). *Panning* can be used to change the viewpoint of a visualization like a horizontal scrolling through a timeline to see more of it. *Contextual menus* are particularly useful to present available options such as other data encodings to e. g. switch from a bar chart to a line chart. One possible type of menu that could be used as contextual menus are marking menus. If the menu items are not changing depending on some underlying data, Kurtenbach and Buxton [92] showed that marking in menus are a very efficient interaction technique. Marking refers to selecting an item by drawing a mark which depends on the user remembering where the menu item is as the menu items are not drawn.

#### 6.3.4 Interaction Scope

Any interaction in the document-centric space can be triggered for a single entity or word-scale visualization, or for multiple entities or word-scale visualizations at the same time. The number of entities or word-scale visualizations affected by an interaction matters in particular when transitioning to a visualization-centric view. If the reader is only interested in detail-on-demand for a single entity, interaction techniques and finding space to fit the information have been addressed by previous work like in Victor's "Explorable Explanations" [153], Chang et al.'s Fluid Documents [25] and Zellweger et al. [167]. However, the reader may also be interested in interacting with groups of entities and word-scale visualizations, e. g. all dates in an article or all word-scale visualizations of a single paragraph. In the latter case, a contextual menu is appropriate to assist in choosing which entity group or word-scale visualizations are selected.

#### 6.3.5 Summary

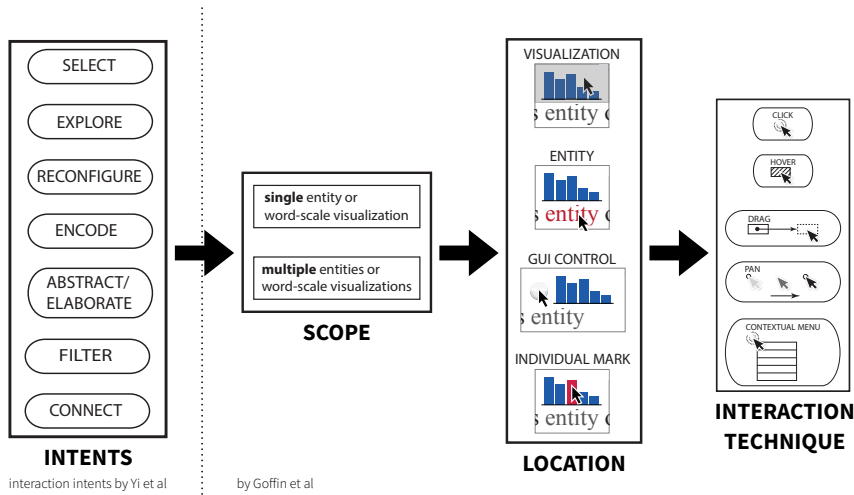


Figure 6.5: The interaction design space, showing the interaction intents, the scope of the interaction, the location where interaction can be triggered, and how this interaction can be triggered.

The design considerations described in the four previous section can be brought together and form a design space for the interaction with word-scale visualizations. The four main components are the intent of the interaction, the scope of the interaction, the area of where the interaction can be triggered, and how the interaction can be triggered. Figure 6.5 provides a summary of the main take-aways of this design space.

These four dimensions can be employed to describe available word-scale visualization interactions. For example in Figure 6.6 – top, multiple word-scale visualizations are connected through brushing and linking, interacting with one of the word-scale visualizations results in change in the other connected word-scale visualizations. Figure 6.6 – bottom, describes the intents as explore and connect and the scope of the interaction is multiple entities. The interaction is triggered on individual marks and uses the hover interaction technique.

#### 6.4 INTERACTION TECHNIQUES FOR VISUAL EXPLORATION

While word-scale visualizations can provide useful information about entities, they are less effective at supporting many other important analysis tasks such as seeing overall trends or making comparisons. One reason for this is that word-scale visualizations are usually spread across the document. Moreover, word-scale visualizations are often either statically embedded in digital documents or only provide rudimentary interactions like hovering over a mark for detail-on-demand. Therefore, we provide a set of interaction techniques that offer new possibilities for analyzing data associated with entities in text documents. Specifically, we introduce interactive techniques for collect-

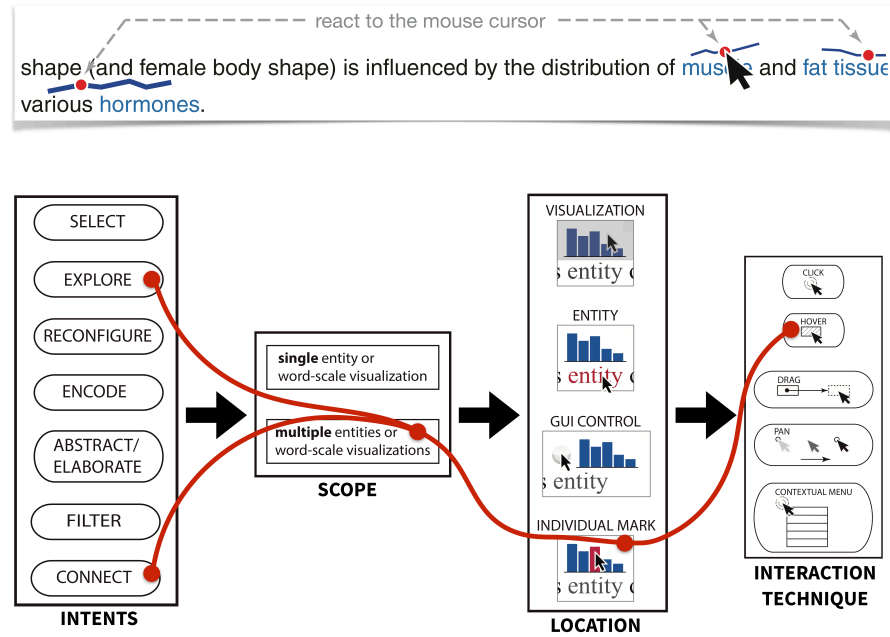


Figure 6.6: An example of an word-scale visualization interaction (top) described using the four design dimensions (bottom).

ing, arranging, and comparing word-scale visualizations embedded in documents. We also show how readers can use these techniques to analyze documents and navigate within them. We illustrate these interaction techniques using examples from economic news and historical document analysis. Finally, we discuss the potential for additional word-scale interactions and examine how our interactive transitions may help bridge the gap between document-centric and visualization-centric analysis.

To support richer active reading and data exploration with word-scale visualizations we developed interaction techniques for three common but currently unsupported analysis tasks:

- T1. Collecting word-scale visualizations.** Creating overviews by bringing together visualizations from across the document to support higher-level analysis.
- T2. Arranging word-scale visualizations.** Ordering visualizations to expose trends.
- T3. Comparing word-scale visualizations.** Providing data manipulation tools that allow analysts to make direct comparisons between visualizations.

We also demonstrate how word-scale visualizations can serve as a visual index for the document and support a fourth task:

- T4. Navigating the document.** Using word-scale visualizations as anchors to navigate to related locations in the text.

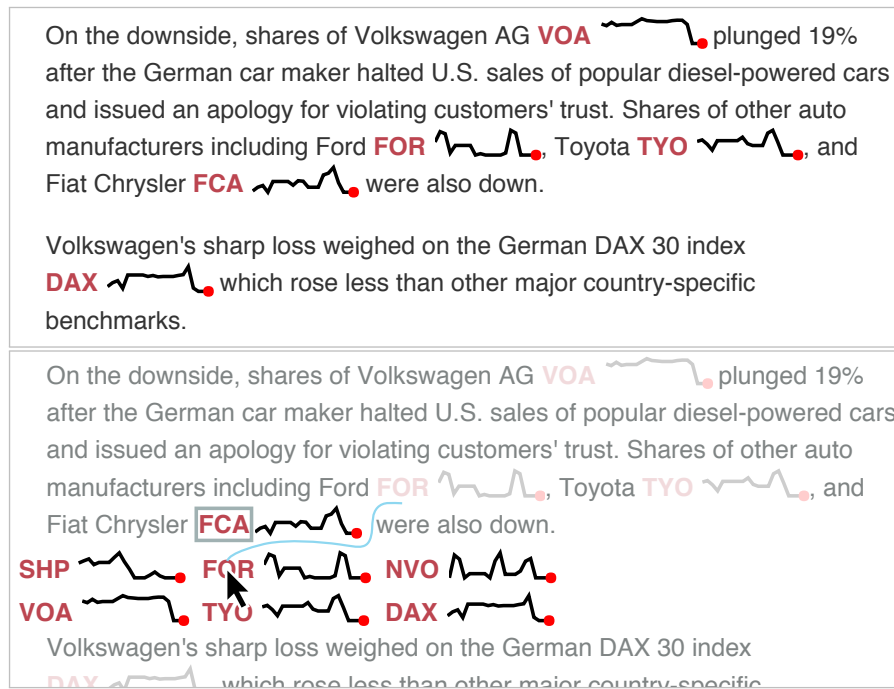


Figure 6.7: The top and bottom image show word-scale visualizations of stock market trends embedded in a news article. In the bottom image, word-scale visualizations and their entities in the document have been gathered near FCA to support comparison. Hovering over FOR reveals a trail connecting the chart to its original location.

We focus on these four tasks because, together, they form a useful set of basic interactions that support what Heer and Shneiderman refer to as “view manipulation” [75]—making it possible to select, navigate, coordinate, and organize visual elements that have, until now, been mostly static.

#### 6.4.1 Collecting Word-Scale Visualizations

When reading long documents enriched with multiple word-scale visualizations it is often difficult to get a full or partial overview of the visualizations. However, seeing an overview is often an important first step before other tasks such as arranging, comparing, sorting and navigation. Already, Shneiderman [133] suggested to first provide an overview in his famous “visual information-seeking mantra”: “Overview first, zoom and filter, then details-on-demand”.

To support full or partial overviews, we introduce two interaction techniques for collecting word-scale visualizations from across a document: *gather* and *drag*. Our *gather* interaction (Figure 6.8 – middle), inspired by “Bring & Go,” [103] allows the analyst to quickly collect a set of word-scale visualizations at one specified location in the document. The analyst can click to select multiple entities in the document and then follow with a double click to mark one as the entity

of interest (Figure 6.7 – bottom). All of the corresponding word-scale visualizations then animate from their original positions in the document to the location of the entity of interest. Alternatively, the analyst can simply double-click one entity to gather all of the visualizations in the document around it. The analyst can also gather pairs of visualizations by selecting one entity and *dragging* it to another elsewhere in the document (Figure 6.8 – right). In all cases, the original text entities stay attached to the visualizations to serve as labels.

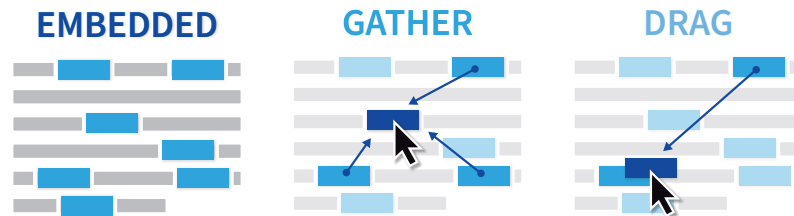


Figure 6.8: Collecting word-scale visualizations.

While the visualizations are gathered, we leave a ghosted copy of the entity and visualization at the original position in the document and de-emphasize the document text. Leaving the ghosted visualizations in place prevents the text from re-flowing and preserves the document layout. The copies also provide provenance information that helps the analyst remember the visualizations' original place in the document. For that purpose hovering over one of the gathered entities displays a trail connecting it to its original location (Figure 6.7 – bottom).

Double-clicking the entity of interest a second time or clicking elsewhere in the document releases the gathered charts, which animate back to their original positions.

#### 6.4.2 Arranging Word-Scale Visualizations

Collecting word-scale visualizations around a point of interest can help provide an overview of the visualizations but does not support more detailed comparisons. To see trends or differences more clearly, the word-scale visualizations need to be arranged into layouts that can convey an order.

##### 6.4.2.1 Choosing a Layout

We support three main small-multiples layouts [147] for arranging word-scale visualizations. We implemented these layouts so that all collected word-scale visualizations are visible on screen if the display space permits.

The *grid layout* organizes the visualizations in a grid around an entity of interest (Figure 6.9 – 1st). This layout makes good use of available screen space by expanding both vertically and horizontally but



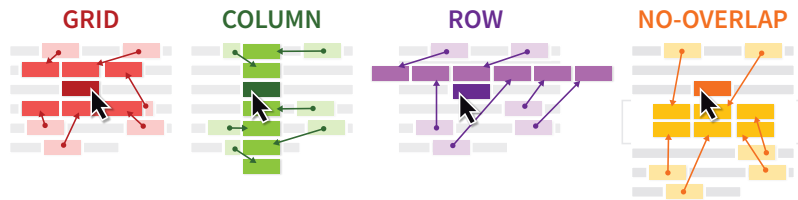


Figure 6.9: Arranging word-scale visualizations.

makes comparison across the chart axes more difficult. The *column layout* positions the selected word-scale visualizations in one column around the current entity of interest (Figure 6.9 – 2nd). For wide visualizations, this layout optimizes comparison across the principle axis, but permits fewer visualizations at a time on screen. Alternatively, the *row layout* organizes the word-scale visualizations in a row next to the entity of interest, permitting comparison on the y-axes and occluding only one line of text (Figure 6.9 – 3rd).

By default, we place these layouts in front of the document text in the area around the entity of interest. This obscures some of the text, but keeps the layout of the document stable. We also attempt to adjust the placement of the *grid* and *row* layouts such that the line of text containing the entity of interest remains visible. Alternatively, when using *row* and *grid* layouts, we can re-flow the text to create new space for the layout just below the entity of interest—similar to Yoon et al.’s approach for placing ink annotations [166] or Zellweger et al.’s [167] Fluid Links. This *no-overlap* strategy avoids collisions between the visualizations and text, but can also change the overall size of the document (Figure 6.9 – 4th and Figure 6.7 – bottom).

Other layouts, such as placing the word-scale visualizations in the page margin, are also possible. However, in these cases the layout may be compressed and the word-scale visualizations will no longer be situated near the original entity of interest or the related text.

#### 6.4.2.2 Ordering

When an analyst first gathers a set of word-scale visualizations, we place them into the chosen layout based on their original location in the document (Figure 6.10 – left). For example—space permitting—we place word-scale visualizations ordered above or left of the focus entity if they came from earlier in the document—and below or to the right of the focus entity if they appear later. However, analysts can also adjust the ordering manually (Figure 6.10 – middle) or sort the visualizations based on visual similarity (Figure 6.10 – right).

#### 6.4.3 Comparing Word-Scale Visualizations

Small multiples in an ordered or sorted layout are already much easier to compare than word-scale visualizations embedded in the

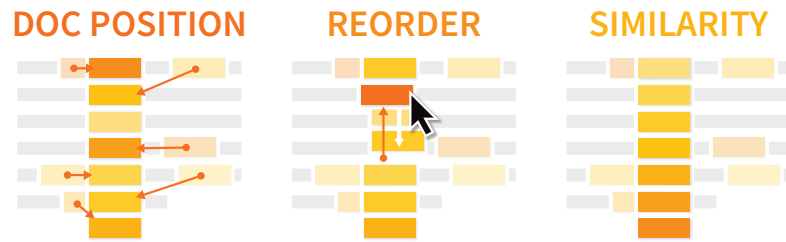


Figure 6.10: Ordering word-scale visualizations.

text. However, fine-grained comparisons, such as comparing values between charts with different axes, can still be difficult. We provide two sets of techniques for a more detailed comparison of word-scale visualizations—interactions for *aligning* charts in rows or columns and tools for *directly comparing values*.

#### 6.4.3.1 Aligning

In the row and column layouts, word-scale visualizations can be aligned in various ways. Here, we present three alignment operators that are useful for different data comparisons.



Figure 6.11: Aligning word-scale visualizations.

The first interaction left-aligns the word-scale visualizations in a column—*column-align* (Figure 6.11 – left)—and, thus, allows for vertical comparison across the x-axes of the visualizations. Yet, sometimes the x-axes of different visualizations have different scales in which case an *axis-alignment* is more useful (Figure 6.11 – middle). For example, when comparing timelines that have different start and end dates, axis-alignment can choose a common date as the alignment's center. A third alignment interaction—*pan-align* (Figure 6.11 – right)—can be applied when the word-scale visualizations are arranged in a column or row on one of the sides of the entity of interest. Panning constrained to the column or row can move the word-scale visualization of interest along the layout to simplify the side-by-side comparison of two visualizations.

#### 6.4.3.2 Direct Comparison

While side-by-side comparison can be very useful, it can be desirable to highlight differences between visualizations more directly [146]. Our *direct comparison* technique overlays two word-scale visualizations so that their axes are aligned within the same frame of reference (see Figure 6.12 – left).

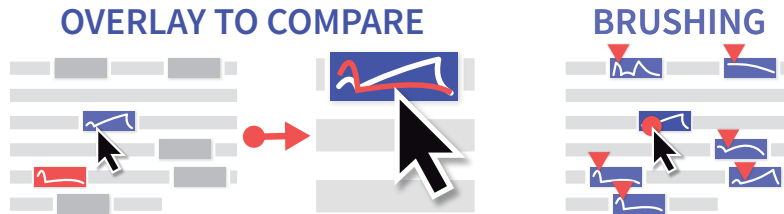


Figure 6.12: Comparing word-scale visualizations.

Another approach to highlight differences is to directly represent them [57], for example by using *brushing and linking* between word-scale visualizations (see Figure 6.12 – right). Using brushing and linking makes it possible to keep each visualization’s position stable within a small multiple layout and can even be used on the visualizations in their original positions.

#### 6.4.4 Navigating the Document

One of our goals is to support ad-hoc switches between reading a document and exploring its associated data. Therefore, we need to support quick and fluid ways of moving from a small multiple display back to the corresponding locations in the original document. We support this switch by using the word-scale visualizations as an index to the document. Double clicking on a gathered word-scale visualization releases the current layout and the view follows the clicked visualization as it animates back to its original location, automatically scrolling the window if necessary. We also add colored emphasis to the sentence surrounding the visualization to help clearly identify the entity and visualization’s original position.

### 6.5 EXAMPLE USE CASES

We briefly illustrate our techniques via two examples from (1) economic journalism and (2) historical document analysis. For that purpose we implemented the proposed interaction techniques.



Figure 6.13: Word-scale visualizations of timelines embedded in an historian's note. In the bottom image, word-scale visualizations and entities are aligned along their dominant axes to help compare timespans (horizontal bar).

#### 6.5.1 Comparing Stocks Market Charts

An analyst examines a news article to learn about recent market developments. The story includes word-scale visualizations that show the recent stock performance for the companies it mentions (Figure 6.7 – top). When examining the article, the analyst might want to compare the ups and downs of different companies mentioned in the story. However, comparing the charts is difficult because they are far apart in the document and are not aligned with one another. Using our interactive tool, the analyst can collect multiple visualizations at a single location in the document (Figure 6.7 – bottom), allowing her to more easily see relationships and make direct comparisons.

#### 6.5.2 Historian Revisiting Archival Documents

While visiting a National Archive, a historian uses a tool like the CEN-DARI Note-Taking Environment [12] to take rich notes on the documents the historian inspected. Next, the historian annotates his/her notes (Figure 6.13 – top) with small timelines showing important dates associated with people and places that appear in the text. When two timelines are embedded in the text it is difficult to determine whether their timespans overlap. Arranging them in a grid may also not help if individual timelines have varying start and end dates. However, by interactively collecting and aligning the timelines (Figure 6.13 – bottom), the analyst can quickly get a better sense of which



Figure 6.14: From left to right: Document with embedded word-scale visualizations. Word-scale visualizations collected around an entity of interest with text visible. Small multiples without text. A single visualization.

time spans overlap and draw new conclusions about the relationships between these historical figures.

## 6.6 DISCUSSION OF INTERACTION TECHNIQUES

We focus primarily on interactions that support what Heer and Shneiderman [75] call “view manipulation”. Yet, the layouts and arrangements created with our techniques can also easily serve as starting points for tasks in their “data & view specification” and “process and provenance” categories. For example, the overviews we create can simplify the task of filtering or sharing overviews of a document with embedded data.

Most importantly, however, our techniques may provide a way to bridge the gap between document-centric analysis—where the document and its sentences are the focus of attention—and visualization-centric analysis—where the focus is on supporting visual analysis of data related to the document as suggested in Section 6.6.1. Moreover, treating word-scale visualizations as manipulable tokens may allow to support an even wider range of document-related analysis tasks.

### 6.6.1 From Document-Centric to Visualization-Centric Analysis

Currently, most document analysis takes place either around the raw documents themselves, or in visualization-centric environments like Jigsaw [137] that are only loosely connected to a document. Furthermore, it is difficult to fluidly transition from document-centric views to visualization-centric ones and back during the course of an analysis. However, our word-scale interaction techniques introduce a number of interesting intermediate stages (Figure 6.14) that may support such transitions.

In many cases, visualizations may need to be enlarged or altered to support more detailed exploration. We suggest three alternative approaches for moving from a document-centric view to a visualization-centric view: 1) *In-place transition*: when the user interacts, the layout and dimensions of the text and the word-scale visualization do not change. This option is useful when the word-scale visualization al-

## EASTERN EUROPE

## Soviet cult and pragmatism in Transnistria

Experts worry that the next

"Crimea"  could be the

breakaway region of Transnistria

Many locals there don't share that fear,

and if the last referendum holds, a large

majority would welcome a Russian

annexation.



Figure 6.15: Small maps embedded as word-scale visualizations provide details-on-demand for locations in a news story. Interacting with the entity or map grows the visualization to show more detail. The text stays as it is and the visualization is overlaid over the text.

ready has the appropriate size or if the other two strategies are not applicable.

2) *Growing transition*: the size of the visualization is changed to adapt to the needs of the task at hand (see Figure 6.15). For example dragging the cursor over an entity allows the user to grow the word-scale visualization according to the distance traveled with the cursor. There are two ways in which the surrounding text can react to such an enlargement: either it can remain static and the visualization will cover and occlude the text as it grows, or the text can reflow around the visualization. In the latter case, the visualization effectively becomes an image in a document and the text flows around the image.

3) *Offsetting transition*: moving the visualization to a place where more space is available—e. g. in the margin or in a separate window—to avoid interfering too much with the readability of the document. In this case, the text does not need to be reflowed.

A document with word-scale visualizations represents one step beyond a pure document-centric view containing only text. While the text remains the focus, the embedded visualizations reveal additional data in-context. A small multiples layout like a grid exists in-between the document-centric and the visualization-centric views and may serve as a transition point between them. If the data is currently of greater interest to an analyst, he/she might hide the text completely and focus entirely on manipulating and transforming the visualizations in a more general analysis environment. However, if the text again becomes important, this process could be reversed, fluidly reintroducing the visualizations back into the document. The difficulty to connect a document-centric view and a visualization-centric view lies in finding meaningful and valid transition paths between the views. This is a promising avenue for future research.

## 6.7 CONSIDERATIONS FOR FUTURE WORK

This set of proposed interaction techniques is a first step to add meaningful interaction to a reading environment with word-scale visualizations. Multiple research directions for future work are open. I picked four promising directions and discuss them further in the following sections.

### 6.7.1 *Word-Scale Visualizations as Tokens*

In this chapter, we address several general classes of interaction tasks. However, we have not yet deeply explored operations like filtering, encoding, or abstracting/elaborating word-scale visualizations—which Yi et al. [165] also identify as important for visualization and that we mentioned in Section 6.3.2. One possible way of supporting operations like filtering is to treat word-scale visualizations as tokens that can be dynamically manipulated, sifted, and organized into groups, stacks, and piles. Using physically-inspired and force-directed interactions like those in “Kinetica” [127] or piling and sorting techniques like in “Small MultiPiles” [2] could enable more nuanced and creative ways to explore data embedded in documents or even manipulate the document itself. Investigating if more flexible layouts such as force-directed layout could be more appropriate goes into a similar direction. Therefore, investigating how well this metaphor of word-scale visualizations as tokens is useful is an interesting avenue for future research.

### 6.7.2 *Multiple Word-Scale Visualization Instances*

An interesting point that has not yet been discussed is how such interaction techniques could deal with duplicate word-scale visualizations. An interaction technique such as gathering could already handle duplicates at the stage when entities are collected and only keep one instance. However, discarding entities even if duplicate early on can lead to missing information. The word-scale visualizations can be the same, but the sentence where the word-scale visualization is embedded will definitely not be the same. Another way of dealing with duplicates is, for example, to superimpose duplicate instances like a deck of cards. The top would show the word-scale visualization but the analyst is aware that there are multiple instances of this word-scale visualization and entity. The analyst has then the possibility to explore each of the sentences associated with the word-scale visualization. Investigating how to deal with multiple instance of the same word-scale visualization entity combination is an interesting area for future work.

### 6.7.3 *Interacting with Word-Scale Visualizations in a Text Corpus*

We proposed interaction techniques that collect word-scale visualizations, arrange word-scale visualizations, compare word-scale visualizations and navigate the document. However, these techniques are currently limited to single documents. Often analyses happen over multiple documents and even text corpora. Therefore, how can we deal with a more substantial set of word-scale visualizations (the number of word-scale visualizations exceeds the available screen space). Moreover, if there are many more word-scale visualizations they will not fit all into a browser window on a regular 15" screen after a gathering interaction. The interaction system could support this by proposing selective gathering when not enough space is available to show all word-scale visualizations. Another option could be to use this idea of a continuum between document-centric and visualization-centric. When dealing with word-scale visualizations at the start of a text corpus analysis the document-centric view might not be an appropriate starting point (Figure 6.14). Therefore, starting with a visualization-centric view might be much more helpful to tweak the gathering of visualizations. When an initial selection has been done moving back to a document-centric view can be feasible again. This means that switching fluidly between document-centric and visualization-centric views is needed.

### 6.7.4 *Evaluation of Interaction Techniques*

The interaction techniques proposed in this chapter have not been evaluated yet. Therefore, investigating how well they behave is crucial. One such study could have the following form. Study participants would get two conditions. Condition 1 is a text with statically embedded word-scale visualizations. Condition 2 is a text with word-scale visualizations where the developed interaction techniques for collecting word-scale visualizations, arranging word-scale visualizations, comparison word-scale visualizations, and navigating the document through word-scale visualizations are available. The proposed tasks would be of type information extraction tasks. Examples for such information extraction tasks are: searching for an entity and a data point, doing a binary comparison between word-scale visualizations, extracting the entity with the word-scale visualization that has the lowest data point and the entity with the highest data point. The metrics we would measure are time, accuracy and the interactions they used. Moreover, it would be interesting to understand if the proposed interaction techniques suit their needs and how well the participants think the techniques helped them.



## 6.8 SUMMARY

In this chapter, I discussed design considerations for using interaction together with word-scale visualizations. I also introduced a set of new interaction techniques that enable important analysis tasks like searching for trends and comparing series which, until now, have been difficult to perform across word-scale visualizations in a document. This work is a first step since the space of interactions for small-scale and text-embedded visualizations has been largely unexplored. However, I focus on techniques that can serve as a basis for developing future tools that cover a larger and richer set of interaction tasks and intents. I investigated different aspects of word-scale visualizations, however how word-scale visualizations might change reading behavior has not yet been investigated. Embedding word-scale visualizations into text can change the appearance of the text depending on the design and placement of the word-scale visualizations. Therefore, reading behavior might change. Thus, the next chapter investigates how word-scale visualizations affects the reader.



## TOWARDS UNDERSTANDING THE EFFECTS OF WORD-SCALE VISUALIZATIONS ON THE READER

---

In the previous chapters I discussed the design and use of word-scale visualizations, their placement, and how to add interaction. However, I have not yet considered the impact of word-scale visualizations on the reader and their reading behavior. Moreover, if there is a change in reading behavior, does this change of reading behavior have implications for the reader on the comprehension and on the retention of the content of the text?

This chapter provides a first step towards the discussion of the effects of word-scale visualizations on reading behavior. I discuss an initial study that examines how different placements of word-scale visualizations affect reading speed, retention, and whether readers tend to use information from the word-scale visualizations or the text. In this study I investigate the effects on reading single sentences. After the discussion of the study results I propose considerations for future studies that investigate how multiple word-scale visualizations in a more substantial piece of text influences the reader's reading behavior.

The following evaluation of word-scale visualizations with respect to changes in reading behavior does not take into account the legibility or readability of word-scale visualization designs. These issues have been slightly touched in Chapter 5, but I deliberately do not concentrate on legibility and readability as they are highly dependent on given designs. Once certain word-scale visualizations for specific applications or more general purpose designs become established this will be the occasion to evaluate their legibility and readability.

Parts of this chapter have been previously published in Goffin et al. [62]. Hence, the use of “we” in the remainder of this chapter refers to Pascal Goffin, Wesley Willett, Anastasia Bezerianos, and Petra Isenberg.

### 7.1 MOTIVATION

In Chapter 5, I explored how different placement options like word-scale visualizations between words or between lines of text can affect the text layout. Inserting word-scale visualizations between lines (Figure 7.1a) or words (Figure 7.1b) may require to expand the inter-line space or to increase the space between words to fit them in. This often requires to reflow the text. Moreover, overlaying interactive visualizations over text such as in Figure 7.1c may also impact reading

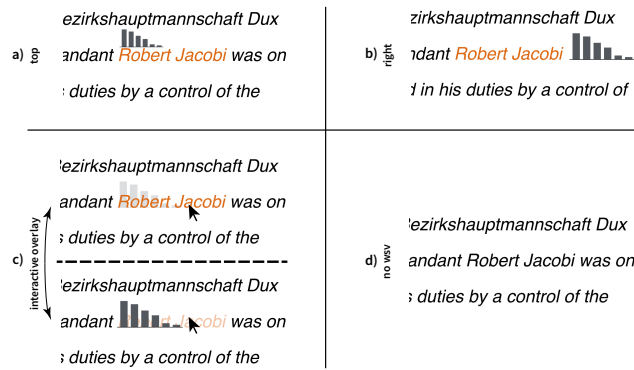




Figure 7.1: The four word-scale visualization positions: (a) “top”, (b) “right”, (c) interactive “overlay” (d) “no word-scale visualization” (no wsv).

behavior, because the text is less readable. Thus, the placement of word-scale visualizations can significantly change the layout and appearance of a text. Furthermore, the reader may have to interact with the text to see the visualizations. Therefore, we hypothesize that readers may react differently to the text depending on what placement strategy is used.

While prior work has described use cases for word-scale visualizations (see Chapter 3 and Chapter 4), it is not clear how their integration affects reading the text. The effects of their integration has not been studied in the literature. In order to address this question, we conducted an initial study that examines how placing word-scale visualizations affects reading speed, retention, and whether readers tend to use information from the word-scale visualizations or the text.

## 7.2 BACKGROUND

While example word-scale visualizations such as Gestaltlines [16]  or SportLines [118]  exist, no prior studies have explored the effects on readers when word-scale visualizations are embedded in text. Related work, however, has considered other additions to text documents and their effect on readability. For example, Diaper and Waelend [40] suggest that graphics surrounding blocks of text do not distract readers of web pages and do not hinder reader’s ability to search for information in the main part of the website. Zellweger et al. [168] conducted an observational study using an eye tracker to study the addition of textual annotations to their Fluid Documents [26]. They used several different placement conditions including the inter-line space, margins, and pop-ups. They did not find any significant differences in eye movement between conditions but subjective preferences were complex and depended on many different factors. Some participants preferred having the complete static text for the

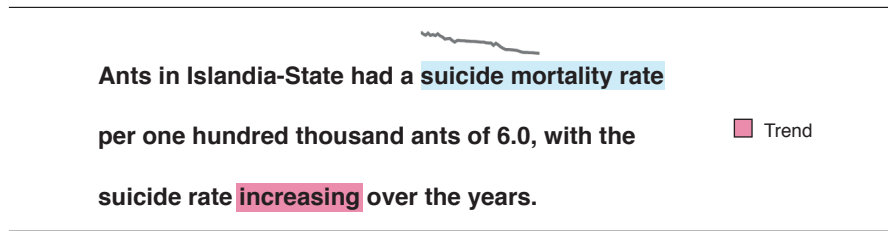


Figure 7.2: An example sentence describing information that is ambiguously encoded in the sentence (trend = increasing) and in the word-scale visualization (decreasing). Colors added for illustration.

additional information always visible. Other participants preferred the possibility to change the spatial arrangements of the primary text on demand to reveal the extra text. This study relates to ours as we similarly include a static and dynamic condition for viewing word-scale visualizations. Yet, both previous studies differ from ours as they did not study graphic visualizations that require a different type of reading compared to text or graphical icons.

Even though Zellweger et al. [168] observed different preferences by the participants, they still found evidence that placing the additional information close to the point of interest is desirable. Moreover, the placement of the information had a significant impact on reading speed and participants read the information close to the point of interest faster.

### 7.3 EXPERIMENT

Our goal was to study how the placement of word-scale visualizations in text impacts reading in a broader context. We conducted a first study that examined how word-scale visualizations are used by readers in a timed information seeking task. Even though reading is part of the task, it is basically an information seeking task because the main goal of the reader was to find specific information in the sentence or word-scale visualization to be able to answer questions. We assessed if the position of word-scale visualizations has an impact on reading and retention of information from a short sentence. Additionally, we checked whether participants preferred to retrieve information from the text or from the word-scale visualizations.

#### 7.3.1 Factors

The study factors were the four *main word-scale visualization positions* from the design space of placement options in Chapter 5: *top*, *right*, *interactive overlay*, and *no word-scale visualization* (Figure 7.1(a,b,c,d)). The “overlay” position is equivalent to the “baseline” position in Chapter 5.

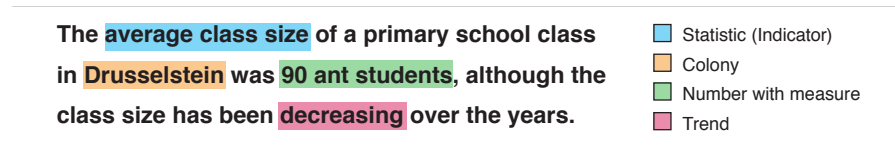


Figure 7.3: An example sentence for the ant colony “Drusselstein” with the statistic “average class size” characterized by a *number* and a *trend* over the last 50 ant years. Colors added for illustration.

Each sentence given to the participants belonged to one of five *sentence type*: The first sentence type consisted of sentences without word-scale visualizations called *no word-scale visualization sentence*. The four remaining sentence types had one word-scale visualization per sentence. Each sentence contained information that had to be used to answer questions after reading. For the cases that contained visualizations we explored four different information-related conditions—a) we encoded the information only in the sentence which we called *info in sentence*, b) we encoded the information only in the visualization which we called *info in visualization*, c) we encoded the information redundantly in both the sentence and the visualization which we called *redundant info*, and d) we encoded the information in the text and the visualization but the information contradicted each other which we called *contradicting info*. The goal of the sentences from “info in sentence” was to include questions that did not ask about the trend to force the participants to read the sentence. The goal of the sentence with “contradicting info” (Figure 7.2) was to identify from where—sentence or visualization—readers took the information to answer a post-reading question.

### 7.3.2 Task

The participant’s task was to read and remember information from a single sentence. We concentrated on single sentences as opposed to smaller texts or paragraphs to build up our knowledge on the effects of word-scale visualizations on reading from small textual building blocks. Sentences are the building blocks of text documents. Therefore, sentences were appropriate to use as a unit to measure speed of reading. Future studies—informed by this one—will include longer texts.

### 7.3.3 Procedure

We gave participants initial written instructions explaining word-scale visualizations and the study procedure. We included a short paragraph explaining the concept of word-scale visualizations, because we wanted all participants to have the same basic knowledge about

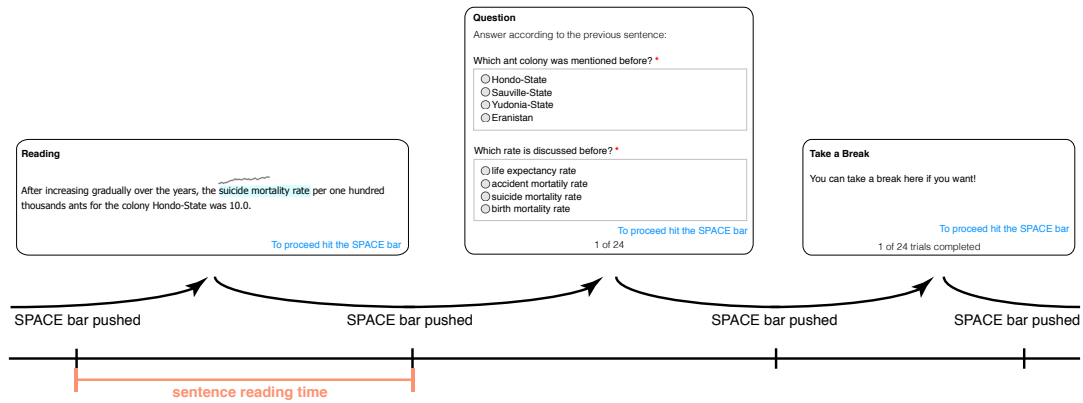


Figure 7.4: A trial starts with the Sentence Page, here with a highlighted entity and a word-scale visualization. Pushing the SPACE bar leads to the Question Page. Answering the questions and pushing the SPACE bar leads to the Break Page, and then pushing the SPACE bar leads to the next Sentence Page. This trial had the information encoded only in the sentence, therefore the first multiple choice question asked about a colony name.

word-scale visualizations. The aim of the study was not find out if the participants understood individual word-scale visualizations but to better understand how they their reading process was affected by small, contextual visualizations embedded into text. We then presented reading tasks to participants in blocks, where each participant saw all tasks for a given presentation position (*top*, *right*, *overlay* or *no word-scale visualization*) before moving on to the next placement. We used a  $4 \times 4$  Latin square design to counterbalance the blocks. At the beginning of each block we included a small training (3 sentences) to familiarize the participants with the new word-scale visualization position. The blocks for each of the position *top*, *right*, and *overlay* consisted of 24 sentences. The block for the position *no word-scale visualization* consisted of 6 sentences. In total each participant was given 90 sentences ( $3 \text{ "position"} \times 24 \text{ "sentences"} + 6 \text{ "no word-scale visualization sentences"} + 12 \text{ "training sentences"}$ ) including the three training sentences per position. The number 24 results from the four different sentence types with word-scale visualization times 6 repetitions. For the "no word-scale visualization" position we had only 6 sentences instead of 24 because only one sentence type is possible ("no word-scale visualization sentence"). In total this study resulted in 936 "trials" ( $78 \text{ "study sentences"} \times 12 \text{ "participants"}$ ). Each sentence was only displayed once to each participant.

The dataset used was about a fictional ant empire with different colonies. The sentences were inspired by the text that is found in the OECD's annual "at a Glance" series about diverse themes such as education, employment and transport. Each sentence provided a statistic (e.g. education expenses, life expectancy) about one of the

Ants in Islandia-State had a suicide mortality rate  
per one hundred thousand ants of 6.0, with the  
suicide rate increasing over the years.

What was the general trend mentioned?

- ☐ upward trend
- ☐ downward trend
- ☐ constant trend
- ☐ up-and-down trend
- ☐ ambiguous

Which rate is the sentence about?

- ☐ birth mortality rate
- ☐ accident mortality rate
- ☐ life expectancy rate
- ☐ suicide mortality rate

Figure 7.5: The two multiple choice questions for the sentence on the left (same sentence as in Figure 7.2). For the first question that asks for the trend there are five potential answers and for the second question that asks for the statistic there are four potential answers.

colonies at a specific point in time with an associated quantity and a trend description (Figure 7.3).

First the sentence (with or without word-scale visualization) was shown. When the participant was done reading, they had to press the SPACE bar to advance to the questions on the sentence. Answering them and pressing SPACE concluded a trial (screens seen in Figure 7.4). Participants could take breaks if needed between trials. At the end of the study, participants filled out a questionnaire about their reading strategy and preferences concerning word-scale visualization positioning.

We instructed participants to read the sentence and to answer the questions as quickly as possible. The questions consisted of two multiple choice questions (Figure 7.5). The first multiple choice question always asked a question appropriated to the sentence type (see Section 7.3.1). This means that for the sentence types “info in sentence”, “info in visualization”, “redundant info”, and “contradicting info” the first question always asked for the trend seen in the word-scale visualization. If the sentence type was “info in sentence” then the first question asked for some other information such as colony name, number or statistic. If the first multiple choice question was about a trend, five answers were possible (upward trend, downward trend, constant trend, up-and-down trend, and ambiguous). For the cases where the first multiple choice question was not about a trend, only four possible answers were proposed. We initially had four answers for all questions, however when we introduced the ambiguous case we needed five possible answers for the trend questions to cover all options. We decided to not change the other questions to five possible answers. The second question asked either about the colony name, the statistic, or the number present in the sentence. For the second multiple choice question there were always four potential answers. The goal of this second control question was to see how well the sentence was read and how well the relevant information was retrieved from the sentence.



#### 7.3.4 *Measures*

We used two measures of performance: sentence reading time and error rate. Sentence reading time was the interval between sentence presentation and the moment the participant pressed the SPACE bar to advance to the questions (see Figure 7.4).

We measured the error rate for the trend question for the sentences without word-scale visualizations and for the sentences with word-scale visualizations but only if the information was present in the sentence, or only in the word-scale visualization, or placed redundantly in both sentence and word-scale visualization. For these cases the answer was either right or wrong. For the case where information was ambiguously encoded we differentiated between answers that were wrong, came from the sentence, came from the word-scale visualization, or if the participant correctly reported the inconsistency by answering “ambiguous” to the trend question.

#### 7.3.5 *Data Collection*

During this study we collected quantitative data—the answers to the multiple choice questions, and the sentence reading time. This data was collected through the browser by storing all the necessary data as CSV data. Besides quantitative data, we also collected qualitative data using a questionnaire administered to the participants at the end of the study. In this questionnaire (see Section C.1) we asked multiple questions about demographics, but we also asked more specific questions concerning the participants memorising strategy for the different sentence types we presented, and if they memorised the trend of the word-scale visualization actively. They also had to assess the disturbance created by the positioning of the word-scale visualization. We also asked more specific question about the ambiguous cases. For example we wanted to know how they understood the multiple choice answer “ambiguous”, and what strategy they employed when encountering such a case. The last questions dealt with trust. For example we asked them if they could tell us which information they would trust in the following scenario: reading a newspaper article with word-scale visualizations containing ambiguous cases.

#### 7.3.6 *Participants*

We recruited twelve (8 male, 4 female) participants with their age ranging from 20–30. They were all PhD students at the Computer Science department and had expertise in Human-Computer Interaction (HCI) and Information Visualization (InfoVis). One had minor dyslexia.

### 7.3.7 Hypotheses

Our analysis was driven by three main hypotheses linked to reading time and error rate.

- (H1) Reading a sentence with a word-scale visualization increases reading time, because the reader has to change from reading text to reading a visualization [164].
  - (H1.1) Comparing individual positions, “*top*” takes more time than “*right*” since readers can no longer continue their linear reading flow.
  - (H1.2) “*Overlay*” needs longest as interaction is involved.
- (H2) Questions for unambiguous sentences with a word-scale visualization will be answered correctly more often given the memorable visual nature of the word-scale visualization.
- (H3) Readers will most often use information from the word-scale visualization to answer questions for ambiguous cases given their memorable visual nature.

### 7.3.8 Data Analysis

With respect to the concerns and recommendations regarding null-hypotheses testing [36, 42] we base our analysis and discussions on estimation, i. e., we use effect sizes with confidence intervals. All point estimates and 95% CI are based on 1000 percentile bootstrap replicates of the statistic applied to the data [20].

We log-transformed the time measures to correct for skewness [129]. All reported means are anti-logged and therefore indicate geometric means (GM) [129].

## 7.4 RESULTS

We first report on performance results and then move on to data from the questionnaire.

### 7.4.1 Performance Results

Surprisingly, we were unable to confirm H1 that embedding a word-scale visualization into a sentence had an effect on reading performance. Figure 7.6 is a breakdown of the sentence reading time per placement. The fastest placement position was “*right*” with a GM of 7s (CI [6, 8.3]). Yet, the difference to the slowest placement position “*no word-scale visualization*” (GM 7.46s; CI [6.7, 8.5]) was only 0.46s—too small to likely make a practical difference, thus we cannot corroborate H1.1. It is particularly interesting to see that interaction had no

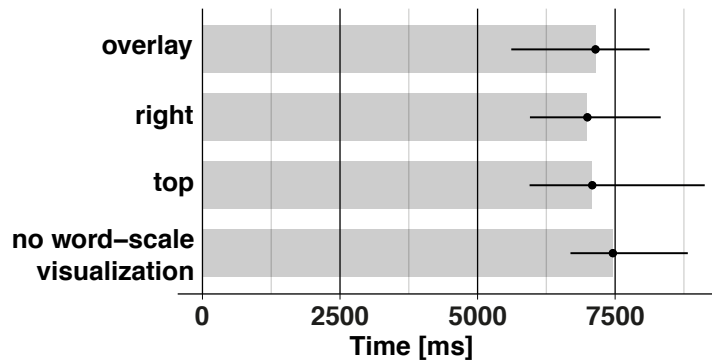


Figure 7.6: 95% CI for participants' time across placement positions.

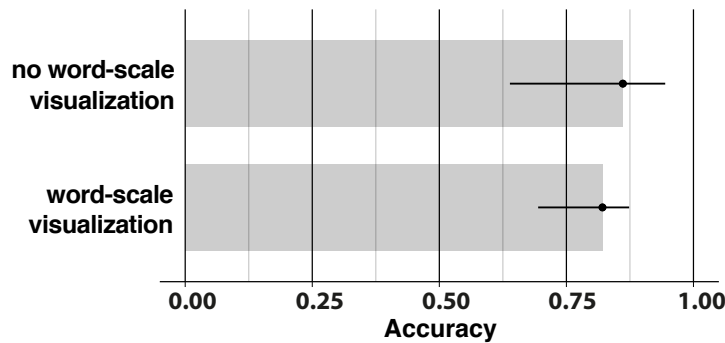


Figure 7.7: 95% CI for the accuracy for sentences with and without word-scale visualization (in the non-ambiguous cases).

practical detrimental effect with the “*overlay*” condition’s GM at 7.14s (CI [5.6, 8.1]). Therefore we are unable to confirm H1.2.

We could also not confirm H2. Figure 7.7 shows a breakdown of the accuracy rates for the two cases: “word-scale visualization” and “no word-scale visualization”. No clear difference between the two cases is evident. The difference between the case “*word-scale visualization*” (GM 0.82s; CI [0.69, 0.87]) and the case “*no word-scale visualization*” (GM 0.86s; CI [0.64, 0.94]) was only 0.04s—too small to likely make a practical difference.

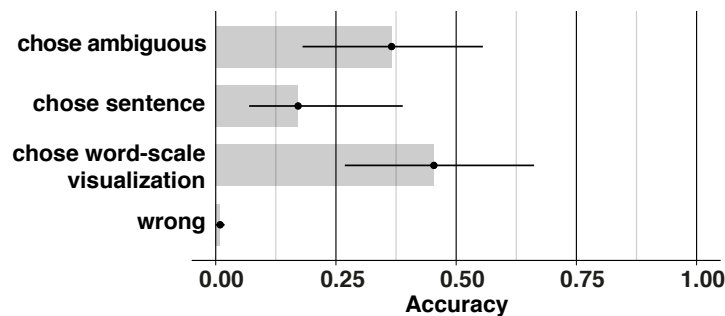


Figure 7.8: 95% CI for people’s choices in the ambiguous case.

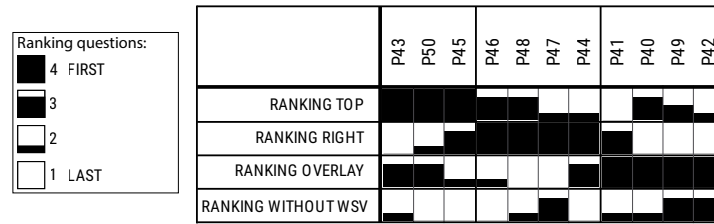


Figure 7.9: Summarized questionnaire data using Bertifier [116]: Grouped participant responses w.r.t. ranking for each position.

Figure 7.8 is a breakdown of the accuracy rate for participant’s choices with respect to the ambiguous case. Often participants recognized the ambiguity (GM 0.37; CI [0.18, 0.56]), thus they had looked at and remembered information from both the sentence and the word-scale visualization. When participants did not identify the ambiguity, they tended to pull their answer most often from the word-scale visualization (GM 0.45; CI [0.27, 0.66]) rather than the sentence (GM 0.17; CI [0.07, 0.39]). This provides some evidence in favor of H<sub>3</sub>, in that when readers did not spot the ambiguities word-scale visualizations were used more often.

#### 7.4.2 Questionnaire Results

We summarized the ranking information from the questionnaire in Figure 7.9. We removed one participant from the pool of questionnaire responses as this participant did not correctly provide a ranked answer. We did not visualize the data for the “disturbance” scores of the different positions, because they more or less corresponded to the answers from the ranking of the different positions (Questions 11 to 14 in Figure C.3).

We asked participants to rank the four word-scale visualization positions according to their personal preference (Question 15, 16 in Figure C.4). In Figure 7.9 “the more black” in the cell, the higher a position was ranked. We can distinguish three groups of participants: Three participants ranked the “*top*” position first, four ranked the “*right*” position first, and another four ranked the interactive “*overlay*” position first. Given combined ranking results, the “*top*” position (ranked median: 3) and the *overlay* position (ranked median: 3) was most preferred, while the “*right*” position was the least popular (ranked median: 3) together with the “*no word-scale visualization*” position (ranked median: 2).

No position was consistently ranked number 1, for different reasons explained by participants. P43 said in favor of “*top*”: “I prefer the “*top*” word-scale visualization over the “*interactive*” one because it provides the best of two worlds: allows for reading smoothly and allows for quick access to trend with one glance. It works for this particular example (short

text)." P44 said in favor of "right": "... right of the word was the easiest for me as I just had to follow the line to get the information ...". P42 preferred "overlay": "[The] interactive version gave me the choice to display [only when] I felt the need ...". There were also negative comments regarding different positions. E. g. P49 said: "I disliked the right position both because it forced me to stop my reading and also for aesthetics reasons". One participant also mentioned that the process of interacting with the entity distracted her from memorizing the information and therefore forgot some of the information.

In the questionnaire we asked participants about their reading and memorizing strategies for sentences without a word-scale visualization, with a word-scale visualization, and with an interactive word-scale visualization (Question 7 to 10 in Figure C.2 and Figure C.3). For the sentence without word-scale visualizations participants reported two strategies either the participants read the sentence and tried to remember the important information or they directly scanned the sentence for the four information pieces (colony, trend, number with measure, and statistic) usually asked in the following multiple choice questions. To improve memorization some of the participants used subvocalization or read the sentence out loud. However, one participant felt that speaking the keywords out loud interfered with her visual memory and as a result sometimes she did not remember some of the information in the sentence.

When the sentence had a non-interactive word-scale visualization the participants memorizing strategy changed slightly. Some of the participants started by first looking at the word-scale visualization and then read the sentence or scanned the sentence for the keywords. Others did the inverse looking at the sentence first and then looking at the word-scale visualization. Using these strategies most of the participants realized that sometimes the trend in the text did not correspond to the trend in the word-scale visualization. Therefore, they tried to validate the trend information by checking if both trends were the same. One participant mentioned that she spoke out loud the trend visible in the word-scale visualization to better remember it.

If the sentence had an interactive word-scale visualization the strategy did not change besides encouraging participants to actively hover over the entity. Some participants reported that it made them look consciously at the word-scale visualizations, because they had to interact with the entity to be able to read the word-scale visualization. We also asked if the participants actively memorized the trend in the word-scale visualization. Six of the twelve participants reported to actively memorize the trend in the word-scale visualizations, three participants reported to having it done sometimes, and three reported having not actively memorized the trend.

We also asked the participants questions about the ambiguous cases (Questions 17 to 21 in Figure C.4, and Figure C.5). All besides one reported having identified ambiguous cases. However, when asked about what ambiguous meant to them, not all responded having identified contradictions (three participants) which is what we meant with ambiguous. These three participants mostly selected the ambiguous answer when they did not remember the answer to the first multiple choice question. We also asked a question to get a better sense of how the participants reacted when they recognised a contradiction. Some only memorized the trend information from the sentence others did the opposite and tried to memorize the trend visible in the word-scale visualization. One participant even reported doing an aggregation of the two information, for example if the sentence trend was upward and the word-scale visualization trend downward the resulting answer was an up-and-down trend. For the last question on contradiction, we asked the participants to report which information they would trust more—the information in the sentence or the information in the word-scale visualization—if they encountered a case of contradiction in a newspaper article. Four out of twelve participants reported trusting the information from the word-scale visualization more, because they claimed that the visualization is based on the actual data, creating the visualization is more complex and more time consuming than writing text. Four of the participants reported trusting the information from the sentence. As reasons for trusting the sentence more they reported that visualization can deceive people who are not careful, it is easier to lie with a visualization than with text, the text contains more information on the trend. And four participants reported not being sure which information to trust, as it is easy to make mistakes in the text or the visualization, unsure as different mistakes can result in different consequences in the text but also in the visualization, no possibility to know which information is correct, and it depends on the credibility of the author.

## 7.5 DISCUSSION OF THE STUDY

When combining the data from the measurements of sentence reading time and error rate and the data from our questionnaire no single position emerges as the best. Additionally, no single position was consistently preferred by all participants. This may actually be a positive result, as it suggests that a variety of word-scale visualization positions can be experimented with in the future for different use cases. It may even be appropriate to allow readers to choose where they want the word-scale visualizations to be placed, as our participants voiced strong preferences for several different positions.

A further take-away from this study is that participants were drawn to the word-scale visualization which was also reported in the ques-

tions on reading and memorizing strategy. Moreover, they seemed to use the information in the word-scale visualization more than the information in the sentence in cases where inconsistencies in the two were not spotted. This suggests that participants did not just skip over the word-scale visualization. The interactive word-scale visualization helped some of the participants consciously look at the word-scale visualizations because they had to interact with the entity to fully see it. The overlay position supported some of the participants to actually realize that there is something to look at. Therefore, an overlay position could help to make readers aware that there is something to look at.

However, these results have to be taken with a grain of salt, as we only used one sentence instead of a more substantial text. Using only one sentence made the visualization appear less disruptive and thus had less of a visual impact on the reader. Moreover, the choice of word-scale visualization—a line without axes—increased the impression of a non-imposing and non-dense visual object.

We also narrowed our scope of types of data we collected. We collected qualitative data based on a questionnaire and we collected quantitative data on reading time and error rate. However, we did for example not collect eye-tracking data as we did not have immediate access to an eye-tracker.

All the study participants were students used to do studies and they all had a background in Human-Computer Interaction or Information Visualization. This might have led to faster sentence reading time, because they are used to working efficiently with visualizations and written text.

Our baseline was a sentence without word-scale visualizations and not a sentence with a non-word-scale visualization. We took this decision to have a comparatively small visualization reading time in comparison to the reading time of the sentence. Moreover, we wanted the visualizations to be embedded in the sentence.

## 7.6 CONSIDERATIONS FOR FUTURE WORK

The previously discussed study did not measure potential effects of word-scale visualizations on reading behavior. Therefore, future studies have to focus on better understanding reading behavior of text with embedded word-scale visualizations. Following, I discuss such a study in more detail, and propose how to better understand if word-scale visualizations have an effect on memorability and comprehension.

### 7.6.1 *Reading Behavior*

Future studies have to, in particular, change from a single sentence with a single word-scale visualization setting to a multiple sentence setting (at least one paragraph) with multiple word-scale visualizations. I now illustrate such a future study in more details, as this is the study that has to be done next. Moreover, I and the co-authors on this initial study had already some intensive discussions around this study. The results of that study would also provide much needed insights into the effects of word-scale visualizations embedded in the text on the reader.

#### 7.6.1.1 *Suggested Factors*

I suggest to compare reading behavior on a paragraph without word-scale visualizations with a paragraph having multiple embedded word-scale visualizations, with a paragraph having visualizations exclusively in the right margin, and with a paragraph having them only in the bottom margin of the paragraph. Thus, there are four main conditions. The placement option used for the word-scale visualizations embedded in the text are either the right position or the top position. I choose the right position, because it integrates the word-scale visualizations in-line which places them optimally into the reading flow of the reader. However, the reader has to change from reading text to reading a visualization and back. The top position adds an additional difficulty as the word-scale visualizations are at an unfamiliar place—outside of the reading flow—for the reader, in addition to the change from reading text to reading a visualization. The additional locations—right margin and bottom margin—are there to better understand how the reading process is affected by the distance between the point of interest and the additional information in the word-scale visualizations (distance entity to visualization). There is literature that points in favor of separating visualizations from the text—as in the bottom or right margin—such as Wright et al. [164]. Wright et al. support a larger distance between text and visualizations as they suggested that it is better to separate reading the text from viewing visualizations. However, in their experiment the visualization was on a separate page and was only accessible through clicking on a link. Text and visualization were not presented together, only the text or the visualization could be seen at once.

Similarly to the initial study, after each condition I would ask questions about the text and the data in the visualizations. These questions are there to better understand which information the participants understood and could later recall. Similarly, I would employ opposing information in text and visualization to see which information is preferred.



### 7.6.1.2 *Suggested Setup*

To get a better understanding of the reading processes I will use an eye tracker to measure eye position and eye movements. More specifically, I will gather scan path-derived metrics which includes fixations and saccades [93]. The eye tracker data is thought as a complement to reading time, accuracy rate of responses to questions, and the subjective assessment of reading performance. However, even though Carpendale [23] suggests that the advantages of the think-aloud protocol outweighs its disadvantages, I do not suggest the combination of eye tracking and think-aloud protocol. The reason for that is based on Holmes' [76] caution that the think-aloud protocol can slow down the performance of the task and can influence eye tracking data. Looking at a visualization and processing it, is much faster than articulating what can be seen in the visualization and at the same time making sense of it. The focus of this study is the reading process and the effects of the word-scale visualizations on that reading process. Thus, making the study as natural to a reading task as possible is important.

I did some initial tests with an eye tracker and observed that to be able to differentiate between text and embedded word-scale visualizations I have to use larger font sizes and line spacings. Supporting this decision is Rello and Pielot's work [123] that suggest comprehension and readability improves in online reading environments when using larger font size up to 18 points. They also observed no further effects for the objective measures for font sizes above 22 points. Objective readability is based on eye tracking data such as the duration of eye fixations. However, subjective readability decreased. Subjective readability is based on the subjective perception of how easy it was to read and understand the text. Concerning line spacing, they observed no effects for objective readability, but extreme spacings such as 1.8 had a negative effect on objective and subjective comprehension.

### 7.6.2 *Memorability and Comprehension*

Another interesting avenue for future research is to investigate whether or not word-scale visualizations can enhance memorability and text comprehension. If this is the case it would support the case for integrating word-scale visualizations into text. To investigate the effect of word-scale visualizations on memorability and text comprehension a similar setup as described in this chapter can be used. Immediately after the reading phase of each text with the different conditions I would add a recall phase. During these recall phases I would ask two to four questions about the text and the visualizations. To remove a possible learning effect, I would not provide any feedback on how they fared in answering these questions. There would be no communication about the delayed recall phase. Two weeks later the study participants would be asked to do a delayed recall phase, where they

answer the same questions they already answered in the recall phase after each text. I would then compare how much of the information they remembered correctly after the immediate recall stage and after the delayed recall stage. Moreover I would compare from where the information came to answer the questions—the text or the word-scale visualizations.

## 7.7 SUMMARY

In this chapter I discussed an initial study to better understand the effects of word-scale visualizations and their placement on reading behavior. This study used single sentences with one word-scale visualization. The results suggest that the study participants did not prefer one placement. They also seem to not neglect word-scale visualizations but use its information to answer questions even when the information in the word-scale visualization contradicts the information found in the text.

I also propose to run further studies to explore the effects of word-scale visualizations on reading behavior. I describe a potential study design in more detail where there is at least one paragraph and multiple word-scale visualizations. In addition, I suggest to use eye tracking data to get a better understanding of potential changes in reading behavior.

In the next chapter I explore what is needed for a word-scale visualization authoring environment to facilitate the process of integrating word-scale visualizations into text.

## AUTHORING WORD-SCALE VISUALIZATIONS

---

In the previous chapter, I discussed a study that investigated how different placements of word-scale visualizations affect reading behavior. In this chapter, I focus on tools that can help to lower the threshold for authoring and including word-scale visualizations into text—making integrating text and word-scale visualizations as fluid as writing text is today. More specifically I investigate design implications for future word-scale visualization authoring tools and how to make it easier for designers to integrate them into documents. One example where a word-scale visualization authoring tool could be beneficial is Wikipedia. Wikipedia already has tools to integrate interactive visualizations [97] and could be extended to also integrate word-scale visualizations. However, for that purpose appropriate authoring tools are needed.

The main part of this chapter includes work from Goffin et al. [63] which is in submission to TVCG. Hence, the use of “we” in the remainder of this chapter refers to Pascal Goffin, Jeremy Boy, Wesley Willett, and Petra Isenberg.

### 8.1 WHY AUTHORING TOOLS

Designing and integrating word-scale visualizations into text is a complex task as observed in Chapter 4’s study. In that study we asked graphic designers to read three Wikipedia articles and to conceive how word-scale visualizations could be used. After that we asked the designers to create and embed word-scale visualizations into these articles using Illustrator. During our exploratory study in Chapter 4, we purposefully recruited designers who were capable of manipulating documents to generate new word-scale visualizations from scratch. However, it is clear that not all writers, editors or authors are also competent designers or visualization specialists with the expertise to create these sorts of graphics. Currently, the amount of time necessary to manually create and integrate word-scale visualizations into a document is also very high—with individual visualizations taking orders of magnitude more time to author than raw text. However, simple authoring tools could make the process of creating and integrating word-scale visualizations considerably easier by streamlining or even automating portions of the creation and placement process.

We already created a tool called Sparklificator (see Section 5.8) that helps content creators create and embed word-scale visualizations. However this library is geared more towards programmers than for

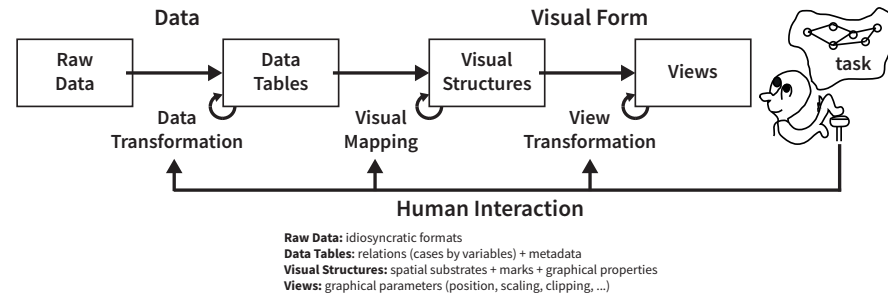


Figure 8.1: Visualization reference model by Card et al. (redrawn from Card et al. [21]).

example designers or writers that have limited programming knowledge. This library provides a lot of freedom. If some functionality is missing the developer can easily extend the library to fit their needs. However, a designer who is not a programmer will have their own designs. Thus, a more user-friendly authoring tool is needed to support this category of users.

## 8.2 BACKGROUND

The process of authoring visualizations in Information Visualization can be described through a series of transformations applied to a dataset and resulting in a visual representation of the dataset. This process is called the *visualization reference model* or *InfoVis pipeline* (Figure 8.1). The InfoVis pipeline has been described by Card et al. [21] and Chi and Riedl [29], and refined by other [22, 83, 145]. Grammel et al. [66] identified six different visualization authoring approaches: visual builder, visualization spreadsheet, textual programming, visual dataflow programming, template editor, and shelf configuration. Sparklificator for example belongs to the textual programming category, as it creates word-scale visualizations by using a regular programming language (JavaScript).

The authoring of word-scale visualizations has not been discussed in the research literature, but a few tools exist out in the wild. There are a few tools and libraries that are based on the textual programming approach, like Peity [120], sparkTable [89], or Sparklificator and that propose authoring of word-scale visualizations. There are also a few tools that use the template editor approach. For example Excel’s in-house “Sparklines – Design” ribbon and “Sparklines for Excel” by Rimlinger [124] they both propose a template-based creation of sparklines using Excel data. Another template-based tool to generate sparklines, this time for Twitter, and using Unicode is by Spencer [136]. The tool generates two different kinds of word-scale visualizations, a crude line chart using dots and a box-plot. Another way to create word-scale visualizations is to use a specific typeface to build

simple word-sized visualizations. This follows the visual builder approach because the limited set of elements of the visualization are already given. One such typeface is Kochel's [87] FF Chartwell typeface. Using a typeface greatly facilitates the integration in-line as it is like "writing" a visualization.

Heer and Agrawala [73] created a tool, also using the template editor approach, that generates word-scale visualizations using banking to  $45^\circ$  for a given dataset and typeface. This tool can for example be used to generate a word-scale visualization showing the high frequency part of the data for a detailed analysis or a word-scale visualization showing the low frequency part of the data to only provide the major trends.

There are also non-visualization authoring tools that are of interest to us due to their similarity of adding additional information close to the point of interest. Zellweger et al. [169] describes a system to author fluid annotations. Their authoring process consists of three main actions supported by the system. First the annotator selects the piece of text (anchor) to which a fluid annotation is added. Then, through the use of a contextual menu the user selects the action to write the annotation. The user writes the annotation using HTML, that permits rich formatting and the inclusion of images. The last step is to format the representation of the annotation and the piece of text to which the annotation is associated. These steps are similar to what we need to author word-scale visualizations, as we also need to first select a piece of text, add data and then create the visual representation of that data. Bouvin et al. [13] describes a prototype to author fluid annotations and integrate them into websites using different web standards. Similarly to Zellweger et al. it is based on the aforementioned three steps of selecting a piece of text, writing the annotation and styling the annotation.

### 8.3 IMPLICATIONS FOR AUTHORING TOOLS

Based on our observations in Chapter 4 as well as our own experiences designing word-scale visualizations, we identified four major decision points at which new tools could support the process of creating and integrating small-scale visualizations. In order to add a word-scale visualization to a document, designers or authors must specify the design of the visualization, as well as any entities to which it corresponds, the data it displays, and its placement relative to the text. Following are a number of important takeaways for the design of each of these operations:

**Designing the Word-Scale Visualization:** An authoring tool needs to offer some mechanism for selecting existing word-scale visualization designs or creating new ones.

In our study, a large fraction of the designs used familiar chart types and nearly a third (31%) of the visualizations were unit-based. This suggests that support for generating simple, scalable visualization designs based on basic geometric shapes could support a large number of the data-driven design alternatives that we observed.

**Selecting an Entity:** While a word-scale visualization does not always need to be associated with text, authors in our study chose to connect the majority of their visualizations to one or more entities. This suggests that it can be useful for tools to provide a mechanism for associating a word-scale visualization with one or more blocks of text. These associations can make it easier to position the visualization and can also provide context for interpreting it.

**Selecting Data:** Our study also suggests that, in many cases, the data needed to populate a design is actually contained in the associated entity (see Section 4.3.1.1). This means that data selection and entity selection can often be treated as a single operation. However, the ability to link data from outside an entity or from outside a document can support a richer and more nuanced use of word-scale visualizations (see Section 4.4.1).

**Placing the Word-Scale Visualization:** While selecting an entity specifies a relationship between that text and the associated visualization, it does not necessarily determine the visualization’s placement relative to the text. In Chapter 5 we discussed options for placing word-scale visualizations in text documents, and explored the trade-offs associated with placing visualizations immediately above an entity, in-line with the entity, or overlapping the visualization and text. Authoring systems could build on this work, showing possible placement options for each visualization and analyzing how they will impact the visibility of the visualization and the flow of the text.

## 8.4 CONSIDERATIONS FOR FUTURE WORK

More generally, tools for authoring word-scale visualizations in text could take different forms depending on the design expertise and investment of the authors. For writers with little interest in customizing designs, mechanisms for including simple word-scale visualizations could be integrated into existing text entry interfaces. For example, an interface could allow authors to create word-scale visualizations using markup embedded directly in the document text (akin to entering equations in  $\text{\LaTeX}$ ) or via in-editor search interfaces like those used to select and insert emoji in most modern operating systems. Such tools could make a number of simplifying assumptions, including supporting a limited set of chart types and reduced control over their

placement. Libraries like Peity [120], which generates simple word-scale SVG charts based on markup included in HTML documents, represent a first step towards this sort of minimalist authoring.

Alternatively, the process of adding very basic word-scale visualizations to documents could potentially be automated entirely. For example, a system could generate and place visualizations automatically next to known entities like dates or stock symbols. In this case, decisions about the design, data, and placement might be made in advance by a developer, then applied to new pages with little or no user interaction.

Another aspect that has not yet been discussed is how interaction could be integrated when authoring word-scale visualizations. Adding interactions to word-scale visualizations could be based on the four dimensions described in Section 6.3.5. These four dimensions are: what are the interaction intents, what is the scope of the interaction, where can the interaction be triggered, and how is this interaction triggered. The authoring system can go through each of the dimensions one by one. In this case, the author can first select one or more intents. This results in the authoring system proposing interactions. For example if the author selects the explore intent one of the proposed options could be “brushing and linking”. Then, follows the scope of the interaction, for some of the interactions it is already clear if the interaction affects a single word-scale visualization or multiple word-scale visualizations. This is also the step where the author selects the word-scale visualizations affected by the interaction. After the scope selection, the system asks for where the interaction can be triggered, and finally the user is asked how the interaction is triggered for example the author can select the hover interaction.

Then next step in the exploration of authoring tools for word-scale visualizations is to build such a system. An example for where such a system can be build and also be useful is Wikipedia. It could build on the already existing facility provided by Wikipedia to build interactive visualizations [97]. Such an authoring tool can be used to better understand how people from diverse backgrounds build such word-scale visualizations and for what purposes. If then these visualizations are used by common Wikipedia information seekers it could contribute an environment to better understand if the word-scale visualizations are used and how.

However, more expressive, unique, customized, or interactive designs like those described by designers in our study likely call for interfaces that provide greater control. These might include the ability to create new visual forms using image editing or coding tools, as well as greater control over the relationship between the text and graphics.

## 8.5 SUMMARY

In this short chapter I provided a set of design consideration that can be used to build word-scale visualization authoring tools. These consist of tools to design the word-scale visualizations, tools to support the selection of entities and their association to the word-scale visualizations, tools that help to select the data used in the word-scale visualizations, and finally tools that provides support and guidance when embedding the word-scale visualizations into the text.

The next chapter will discuss an original application of word-scale visualizations to support sharing important pieces of information from personal digital notes.



## AN APPLICATION OF WORD-SCALE VISUALIZATIONS

---

In the previous chapter, I provided a set of design considerations for building authoring tools for word-scale visualizations. The main motivation for this thesis lays in the Digital Humanities as exposed in the first chapters. However, multiple other domains that use text data in some way or the other could benefit from the power of word-scale visualizations. To show the potential and versatility of word-scale visualizations, I here propose an unconventional application of word-scale visualizations. I sketch how word-scale visualizations could be used in a collaborative and personal visualization setting and propose ways on how word-scale visualizations could deal with the issues around sharing and privacy. Therefore, in the following chapter, I explore the use of word-scale visualizations to enrich personal digital notes with information from collaborator's notes.

The main part of this chapter contains work from Goffin et al. [58]. Hence, the use of “we” in the remainder of this chapter refers to Pascal Goffin, Wesley Willett, and Petra Isenberg.

### 9.1 MOTIVATION

Word-scale visualizations are small and can easily embed additional information in text without being obtrusive. As a result, we can use word-scale visualizations to add additional information in the context of a note-taker's original notes. This can help note-takers better understand the overlap between their notes and their collaborators' without requiring a separate interface.

Digital notebooks can store a variety of different types of personal information, including recipes, health related information, results of web searches, shopping lists, or personal diaries. Digital notes are, thus, a unique and common type of personal data. They offer numerous opportunities for personal analysis—for example, examining the evolution of a person's culinary tastes or retracing a person's daily activity.

While note-taking is typically a solitary activity, for certain situations and tasks sharing the content of notes is important. Examples include a family planning their summer vacation or the members of a book club reading a book and preparing themselves for the next meeting. Many note-taking tools support content sharing. However, note-takers typically refrain from sharing their notes because they are concerned about how their notes and note-taking style will be per-

ceived by others [161]. As a result note-authors clean, reformat, and even rewrite their notes before sharing or simply refrain from sharing their notes at all. Yet, they seem to be comfortable with sharing some information about the notes' content—including which terms in the note are also used by their collaborators [161].

Sharing information and ideas among acquaintances and family members can be a perfect way to spark discussions, prepare events, and increase awareness of other people's thoughts. For example when planning a summer vacation, several family members may use their personal digital notebooks to collect information about potential destinations, what activities or points of interest are available at these locations, or which friends live close by and could be visited. Later, the family must collaborate and consolidate these notes in order to settle on a single vacation plan. In this vacation planning scenario, small visualizations embedded in a note about possible vacation spots could show which other family members have listed similar properties, locations, or countries and, thus, aid in the process of finding a joint destination.

We discuss how word-scale visualizations can be used to show descriptive metadata from collaborators' notes. We specifically focus on sharing only metadata extracted from notes as a simple channel for communicating shared ideas and understanding collaborators' activities. By focusing on metadata we also free collaborators from the burden of feeling that they need to reformat and clean their notes before sharing.

In the following sections, we explore and describe which types of information originating in notes could be shared between friends and family members. We discuss potential problems that can arise when sharing personal information and then elaborate on how word-scale visualizations could help to support sharing relevant pieces of information between members of a small social group. Finally, we illustrate our discussions with two examples.

## 9.2 BACKGROUND

In recent years, digital note-taking tools have become more and more available both inside and outside the workplace. Yet, general information on the characteristics of personal digital notes is relatively sparse. We base most of our knowledge on digital notes on our own previous study [161] of digital note-taking practices and their implications for visualization. We found that while notes that are digital can theoretically be easily shared and distributed, people typically felt the need to edit notes prior to sharing. Hence, we discussed how visualization can be a good way to abstract notes and prepare them for sharing. In that previous work, we discussed visualizations that can create awareness of overlapping concepts and phrases in collab-

orators' digital notes. Here, we discuss in particular how small-scale visualizations can integrate metadata from friends' and family's notes in-context of one's own notes.

Chuah and Roth [30] previously discussed how visualizations can provide common ground during collaboration. One example of the use of visualization to provide common ground was introduced by Brennan et al. [17]. In their work, private views of a graph visualization in a distributed work scenario are merged to show shared information and common ground. Keel [85] discusses a similar idea, using computational agents to identify information in private views that should be relayed to a larger group of collaborators. Similarly, we use word-scale visualizations to provide common ground for small social groups.

Other researchers have used visualization to create awareness about collaborators' search activity. Awareness of other people's data and actions is an important prerequisite for common ground formation and several tools have attempted to provide mechanisms for facilitating awareness. For example, in Cambiera [78] Isenberg et al. use small visualizations (e.g. bars) as Scented Widgets [162] to show the overlap between the documents two participants had searched for and used to provide awareness of the pair's reading histories. The authors termed this concept of explicitly visualizing collaborators' activities "collaborative brushing and linking." Similarly, Mayar and Tory [96] use partial merging to provide subtle awareness of other people's work. They term this concept of merging "linked common work" and implement it in their tool CLIP. The tool includes node-link graphs in which nodes have color coded overlapping segments to represent common and different information among collaborators. This is similar to our approach, which uses word-scale visualizations to provide subtle awareness of information that is shared between collaborators' notes. Finally, Morris and Horovitz's SearchTogether [102] creates awareness of web search histories by showing what collaborators have searched for and by explicitly sharing information via recommendations. In contrast to these previous three tools, we use small visualizations embedded directly within the text of digital notes rather than stand-alone visualization tools.

### 9.3 WORD-SCALE VISUALIZATIONS AS AN APPROACH TO PRESENT SHARED INFORMATION

Here, we discuss how word-scale visualizations embedded in notes can support shared awareness among a small social group. First, we discuss the types of data that can be shared.

### 9.3.1 *Sharing metadata and snippets*

In our previous study of digital note-taking [161], we found that note-takers were generally open to sharing their notes but were worried that unedited notes would overwhelm collaborators with unimportant information. In order to facilitate sharing notes, we therefore propose using word-scale visualizations that show abstracted metadata and snippets from notes. Using only metadata and brief snippets alleviates collaborators from the burden of having to edit their notes and, thus, simplifies sharing.

We specifically discuss descriptive metadata that is created when a note is saved or edited and snippets extracted from the notes themselves.

#### 9.3.1.1 *Technical metadata*

*Where the note was created:* Sometimes digital notebooks store GPS locations when notes are created on mobile devices with location services turned on. Seeing where notes were written could, for example, allow to see if a collaborator has visited and taken notes on a potential vacation destination. While location data can be particularly sensitive, it is also easy to abstract (e.g. reporting only the city, country, or even continent) and obfuscate [90].

*Who authored or edited a note:* In personal notebooks, notes are typically first authored and later edited by the same person. When sharing data from notes, the author/editor information might be crucial in order to judge the potential usefulness of information given existing background information in a social group. This information could also make it easier to contact the author directly and request the full content of a note.

*When was a note created or edited:* This information can be an indicator of how recently someone has engaged him or herself with the topic of a note. In addition, this information could give additional information on the note type. In our previous work [161] we found that certain types of notes, such as project logs or running lists were edited relatively frequently while notes containing information scraps or references and copied content were rarely if ever edited by note-takers. Thus, data about note creation and editing patterns—coupled with other metadata—can provide a powerful tool for deducing the importance of a note.

*Other metadata:* There is a wealth of other technical metadata available for notes that can be the basis of a shared visualization. For example, the length of a note may hint at how much information a note contains. The note content type describes if the note has been authored by the note-taker or if it only contains clippings from public sources (what Willett et al. [161] call “references and copied content”) or if it is a mix of authored and clipped content. These clippings can

be text or image clippings from websites such as Wikipedia or any other website. The name of the tags or notebooks used to file the note might also be useful when identifying related notes.

#### 9.3.1.2 *Extracted keywords*

In order to avoid sharing the raw text from notes but still provide cues on the semantic content of notes, automatically extracted key terms from notes could serve as a data source. These terms can be keyword labels applied to the note by its author or named entities like locations, dates, and surnames that are extracted directly from the text.

User-authored keywords or entities can succinctly describe the topic or ideas of a note without disclosing much unrelated information or revealing a note's disorganized structure. Entities can be found in the text through entity recognition and text processing. They can also be manually tagged by the original note-taker. For visualization purposes, entities or keywords can be used to cluster notes that are of interest to a specific topic. For example for planning a trip to New York City, notes containing entities such as "New York City", "Manhattan restaurants", or "MoMA" are relevant and may provide inspirations of what to visit and where to eat.

In addition to the mere presence of a specific entity, a diverse set of statistics can be gathered around entities. One of them is the number of occurrences of an entity in a note. Others include co-occurrence patterns with other entities or entity types and semantic relationships between entities extracted from knowledge structures such as WordNet [100, 163]. The number of occurrences, for example, can be a good indicator of what to expect from a note in terms of information relevant to a certain topic.

## 9.4 EXAMPLES

We discuss two information-sharing scenarios that illustrate how word-scale visualizations embedded in notes can show some of the data mentioned in the previous section and help support information sharing in a small social group. Additionally, we provide sketches of potential word-scale visualizations.

### 9.4.1 *Shared recipes with Roommates*

In this scenario, four flatmates live in a shared apartment. They like to cook together in the evenings and on weekends, and enjoy creating new menus. Each of the apprentice cooks has a notebook called recipes. All of them like to collect interesting recipes from their family and friends. Moreover, they enjoy browsing the Internet to find inter-



Figure 9.1: This sketch is an excerpt of a recipe from the flatmates' shared recipe notebook. Each ingredient is accompanied by a word-scale visualization depicting the number of recipes with that ingredient shared by other members of the group.

esting recipes to add to their collection. The four flatmates created a shared digital notebook where they collect and share large numbers of recipes with each other.

This shared notebook provides a mechanism for sharing specific recipes, as well as for planning future meals, identifying interesting recipe pairings, and selecting sets of recipes that share ingredients. However, if the shared folder contains a large number of recipes it may be difficult to get a sense of which recipes others in the group are interested in or which have overlapping ingredients.

#### 9.4.1.1 Sketch

Figure 9.1 shows a note containing a recipe for "Zitronencake" which has been shared by one of the flatmates. A small bar chart next to each ingredient highlights how many other recipes in the shared folder contain this ingredient as well as which roommates added them. Each of the bars corresponds to a single roommate and indicates how many recipes with that ingredient he or she has contributed. Hovering over a bar opens a small window that provides links to the three recipes with that ingredient which that roommate has most recently read or added. These visualizations make it easier for the flatmates to navigate the large collection of shared notes and identify interesting recipes added by their peers that might pair well together or which use the same ingredients.

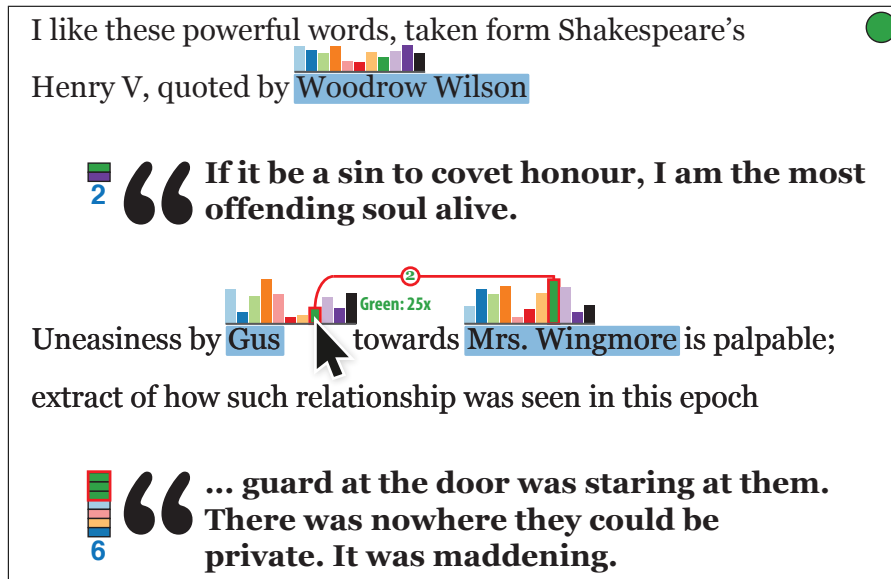


Figure 9.2: This sketch is an excerpt of one of the book club member's note about the current book. The bar chart describes the number of occurrences of the character in each book club member's current book note.

#### 9.4.2 Book Club

Our second scenario—a book club with about ten members—differs in that no notes are shared directly. Instead, each member of the club keeps their own personal notes about the current book and those notes can be annotated to provide peripheral indicators of other members' activity. Each member's individual notes on the book usually contain direct quotes from the work as well as observations about the characters, places, events, and ideas mentioned in the book, written in their own words. For example, a member may take notes discussing why they think a character behaves in one way or another or describing interesting situations that appear in the book. He or she may also discuss relationships between characters or describe feelings and ideas that occurred when reading a memorable passage.

##### 9.4.2.1 Sketch

Figure 9.2 shows how a member's personal notes about the book could be augmented to provide more information about which aspects of the book were interesting to their fellow members. Specifically, the augmented note highlights the shared use of quotes from the book as well as how often other members used the names of important characters, places, and ideas. The quotes can be seen as data from a public source—the book—and can be presented in more detail, whereas the comments about the characters and places might be seen as more private. As a result, the note can be augmented to highlight

shared mentions of those entities—letting the note-taker know which other members have something to say about these points, without directly revealing their personal thoughts on the topic. This information about shared quotes and entities can then be used to steer the conversations when the club meets in person.

In Figure 9.2, the colored circle in the upper right corner of the note indicates its owner—Mrs. Green. The bar charts over each character’s name depicts how many times that character was mentioned in each of the other members’ notes. Interaction can be used to show more details. By hovering over a bar, the note-taker can reveal the name of the book club member and the number of occurrences of the entity. Additionally, hovering highlights all bars belonging to Mrs. Green in the paragraph. Any characters that are mentioned in the same paragraph are visually connected via a line. The links between the character’s also show the strength of this relationship—number of co-occurrences of the characters in the note’s paragraphs.

Quotes drawn from the book are marked with a large quotation mark, indicating that this complete snippet of public material is present in multiple members’ notes. The bar on the left of the big quotation mark indicates which other members included the quote as well as how many times they included it (for example, Mrs. Green included the second quote three separate times, when discussing different aspects of the book). The note-taker can also click on the bar to see the names of the characters associated with the quote.

## 9.5 CHALLENGES AND IMPLICATIONS

Using word-scale visualizations to share metadata and entities drawn from personal notes requires careful consideration of the content of the visualizations and the openness of the collaborators. Moreover, designers must create visualizations that help to abstract note content but also do not unintentionally remove or create links between pieces of information.

### 9.5.1 *Visualization*

The design space for word-scale visualizations is very large. However, an interesting implication for word-scale visualizations in the context of notes is that notes are often edited. As a result, even though embedding visualizations in-line can minimize the degree to which the text is disrupted [60], it may break note-takers’ expectations about the behavior of common commands like cut, copy, paste, or backspace. For example, what happens if the cursor is to the right of a word-scale visualization? Will it be okay to delete using backspace? As a result, the best position for the word-scale visualizations may depend on whether or not the note is likely to be edited. Consider the word-scale



visualizations we placed in-line in the sketch for the recipe scenario. Since ingredients are typically given as itemized lists with plenty of empty white space to the right of each item, and recipes are less often edited than other notes [161] this placement is less likely to cause problems.

### 9.5.2 *Privacy*

The degree of privacy attached to information in a note has a wide spectrum. On one end is the raw note, that when shared with associated metadata discloses everything. The other extreme is a setting in which nothing at all is shared and all notes are kept private. In between these two endpoints we have many possibilities. Certainly there are tradeoffs between sharing and providing awareness and privacy. Word-scale visualizations can be a way to deal with this tradeoff. There is a level of flexibility available when sharing metadata and entities of a note. Not all metadata, entities, or types of entities have to be shared if not wanted. Entities can be chosen through entity recognition and text processing. For digital notebooks, however, this means that an additional infrastructure has to be available to provide the possibility to regulate which entities and metadata are shared and to whom.

### 9.5.3 *Sharing useful data*

Word-scale visualizations provide a degree of abstraction from the original note. Yet, when they are based on metadata such as entities or keywords, word-scale visualizations might provide less information than originally contained in the note. For example, relationships can disappear or be created by inferring that two characters from a book appear in the same paragraph. In fact, the two characters might appear in unrelated sentences or even in quotes from separate books, but the abstraction of the paragraph may seem to imply a relationship. The challenge of choosing the right information to share through word-scale visualizations can be complex. This has implications for the design of the word-scale visualizations. For example, a word-scale visualization that groups entities based on co-occurrence at the sentence level could help avert incorrect assumptions about the relationship between entities, but runs a greater risk of revealing the raw content.

## 9.6 SUMMARY




Word-scale visualizations provide a lightweight way of sharing important pieces of information from personal notes, and can support awareness and decision making without requiring family, friends and

collaborator to reveal their raw notes. Personal digital notes are a treasure trove for personal analytics. This investigation represents a first attempt at helping make family, friends and collaborators more aware of one another while keeping sharing light-weight, avoiding the need for note editing prior to sharing, and placing shared data in-context to already existing personal information.

This chapter provides an example of an application scenario for word-scale visualizations outside of the Digital Humanities. I provided this example because word-scale visualizations can be very versatile and could provide solutions to deal with different issues such as how to share information without compromising privacy. Different domains and issues might benefit from word-scale visualizations.

The next chapter concludes this dissertation by summarizing the content of this thesis and providing perspectives on future research directions.

## SUMMARY AND PERSPECTIVES




In this thesis I advocate the use of small, contextual and data-driven graphics embedded in text, called word-scale visualizations. Examples include small stock charts embedded next to the name of a company such as Apple  104.39, game statistics next to the name of a soccer team such as for Belgium's soccer team  87% pass success, or weather trends like rain averages next to the name of a city such as for Paris  26mm.

Word-scale visualizations are embedded in text, therefore they operate in a space where only limited room is available. The slow but inevitable emergence of high-resolution displays of different sizes that can be used to read text, leads to the necessary display resolution to have a variety of readable designs of word-scale visualizations. These word-scale visualizations can be used to emphasize certain elements of a document (e. g. a word or a sentence), or to provide additional information; they can for example enable graphical comparisons, support complex relationships visually, or simply highlight quantities in the text to facilitate understanding.

Small, contextual and data-driven graphics have received limited attention in the visualization community. This thesis identified and defined important aspects and relevant research direction towards making the integration of text and small-scale contextual visualizations as fluid as writing text and as usable as reading text. This dissertation contributes to the understanding of the value of small, contextual and data-driven graphics for Information Visualization. Following the summarization of the findings of this dissertation is a discussion of perspectives for future research.

## 10.1 SUMMARY

I first illustrated the context of where this dissertation fits in. This context consists of two different analysis approaches. Either focusing on the documents in a document-centric analysis or to abstract from the documents and get a task specific visual overview of the documents and do a visualization-centric analysis. I also suggested that for an analyst it would be very practical and useful to be able to transition fluidly from one side to the other and back. In this research context, I place word-scale visualizations which are embedded into text on the document-centric side. They are thought as a first step towards connecting document-centric analysis and visualization-centric analysis.

I then provided an overview of small-scale and data-driven graphics. There, I discussed three definitions: sparklines, word-scale visualizations and micro visualization. All three propositions agree and hypothesize that data-driven, word-scale graphics can complement the content with for example additional information. I then presented the limited set of already available designs of word-scale visualizations from the literature such as Gestaltlines () , Separation Plots () , and SportLines () . These designs are then contrasted with the limited use of word-scale visualizations in the wild.

Clearly a better understanding of word-scale visualizations was needed, after such a limited set of designs and uses. I presented an exploratory study on the use of word-scale visualizations in text document. The study culminated into a rich set of different types of graphic designs, data provenance, and connections between text, graphics, and data. It also provided insights into potential interaction uses and for what purposes word-scale visualizations can be used. A major contribution of this study was a set of three application goals for embedding word-scale visualizations: supporting the content, extending the content, and going beyond the content. We also provided a clearer definition of what a word-scale visualization is:

*Word-scale visualizations* are graphical encodings of data that are roughly word-size—ranging from the size of a single character to several words. They are placed in context with text.

The study also lead us to rethink what an entity is and its function in relation to the word-scale visualizations.

An *entity* is a concrete piece of text, that can have associated metadata. The entity serves as a way to link the word-scale visualization to the text and can serve as the visualization's legend or caption. A word-scale visualization can have multiple entities. If the word-scale visualization has an associated entity it is typically marked using font color or a frame. If the word-scale visualization has no marked entities, then the visualization is associated with the sentence or the paragraph depending on its positioning.

When word-scale visualizations have been designed and their use determined, they have to be integrated suitably into the text. The integration into the text is dependent on the media the word-scale visualizations are to be embedded in, can the text be reflowed, is it possible to add inter-line space etc. These considerations determine what the possible placement options are. In Chapter 5 I propose three placement contexts of integration: strong-context, weaker context and the out-of-context placement. I further investigated the strong context, thereby proposing a design space of placement options. The main placement options are the top position—over the entity, the right

position—to the right of the entity, and the baseline position—the word-scale visualization is placed as an overlay over the entity. I then discussed the implications for each of these placement options. Based on the results of a quantitative study consisting of a simulation-based computational analysis, I suggest to place word-scale visualizations whenever possible into the inter-line space. If this is not possible and the goal is to save space, placing word-scale visualizations to the right (usually) is the right way to go.

Past examples of word-scale visualizations are mostly static and interaction was not considered. However, nowadays a lot of text can be accessed digitally, therefore opening the door for using word-scale visualizations in combination with interaction. For example, interaction can help to provide word-scale visualizations on demand, or to help to refine or modify what can be seen in the visualizations. In Chapter 6, I proposed a design space of where interaction can be captured, why interaction is useful for word-scale visualizations, what possible interaction techniques are available, and what the possible scope of these interactions is. Moreover, I introduced a set of new interaction techniques that enable important analysis tasks like searching for trends and comparing word-scale visualizations which, until now, have been difficult to perform across word-scale visualizations in a document. These interaction techniques help collecting, arranging, and comparing word-scale visualizations embedded in documents. I also show how readers can use these techniques to analyze documents and navigate within them. Chapter 6 concludes with suggestions on how word-scale visualizations can be used to transition from a document-centric analysis to a visualization-centric analysis.

Until this point it was not clear how word-scale visualizations embedded in text would affect reading behavior. No past research has investigated if readers look at embedded word-scale visualizations, if they disturb the reading process, if the word-scale visualizations help to better understand the text, and if parts of the text are more memorable than without embedded visualizations. Therefore, I investigated how reading behavior was affected when a sentence with one word-scale visualization was presented depending on their placement. The study results suggest that the participants did not prefer one placement. They also seem to not neglect word-scale visualizations but use its information to answer questions even when the information in the word-scale visualization contradicts the information found in the text. This could suggest that participants remembered the embedded visualizations more than the text, when asked to answer questions. I also illustrate how further studies may look like to investigate more substantial text with multiple embedded word-scale visualizations. There are clearly more studies needed to better understand how word-scale visualizations affect reading behavior.

Currently, the amount of time necessary to manually create and integrate word-scale visualizations into a document is very high—with individual visualizations taking orders of magnitude more time to author than raw text. However, simple authoring tools could make the process of creating and integrating word-scale visualizations considerably easier by streamlining or even automating portions of the creation and placement process. Therefore, I proposed design consideration for the design of word-scale visualization authoring tools. A word-scale visualization authoring tool needs a mechanism to either select existing designs or to support the process of creating new designs. Then, it needs a way to select entities and to associate pieces of text to word-scale visualizations. Furthermore, the data represented in the word-scale visualization needs to be selected or linked to it. And finally after the word-scale visualization with the appropriate data has been created it has to be integrated and placed into the text. For each of these design considerations tools are necessary to provide a successful word-scale visualization authoring tool.

To provide inspiration for using word-scale visualizations I discussed and sketched an original application of word-scale visualizations. Word-scale visualizations provide a lightweight way of sharing important pieces of information from personal digital notes, and can support awareness and decision making without requiring family, friends and collaborator to reveal their raw notes.

## 10.2 PERSPECTIVES

As mentioned earlier, this thesis was also a way to identify and define relevant research directions. I now illustrate three of these research directions.

### 10.2.1 *Integration of Word-Scale Visualizations*

In Chapter 4 I reported on a study on authoring and using word-scale visualizations, however their integration was less of a goal as the study participants had clear guidelines on where to place them. As I did not emphasize the integration aspect some of the designs were not adapted to be used within dense text. Therefore, I suggest to do further studies where this time the study participants focus specifically on the integration. One approach is to use a similar design setup as for the study in Chapter 4. Thus, I propose to use again Wikipedia text, but in addition I prepare the data, the intent for the use of the word-scale visualizations and propose more diverse design examples to jump start the creation of designs. The idea is that participants will pass more time on the integration than on the design of word-scale visualizations. The study brief would clearly describe

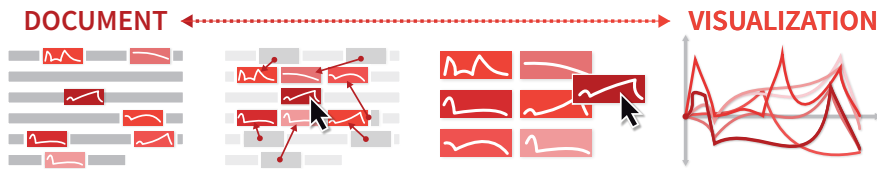


Figure 10.1: From left to right: Document with embedded word-scale visualizations. Word-scale visualizations collected around an entity of interest with text visible. Small multiples without text. A single visualization.

that the end goal of their work is to have word-scale visualizations embedded in text ready for common users to be used.

Moreover, building on the work presented in Chapter 4, I plan to study several aspects of word-scale visualization integration more closely. In particular, I plan to focus on interfaces that will make it easier to design, create, and integrate word-scale visualizations into documents both during and after authoring.

### 10.2.2 *Evaluation of Word-Scale Visualizations*

Chapter 7 clearly demonstrated that future studies are needed to better understand how word-scale visualizations are perceived by readers. Especially, empirical research on how word-scale visualizations affect reading behavior is needed. I only evaluated the setting of a single sentence with a single word-scale visualization. Having more substantial text with multiple word-scale visualizations is a must to better understand the effects of word-scale visualizations. The key to these future studies will be to use eye tracking to provide valuable insights on the reading process of texts with embedded word-scale visualizations. Moreover, better understanding how word-scale visualizations affect memorability and text comprehension could provide insights on the benefits of having text with word-scale visualizations.

### 10.2.3 *Transitioning from Document-Centric to Visualization-Centric Analysis using Word-Scale Visualizations*

At the start of this dissertation and at the end of Chapter 6 I promoted word-scale visualizations as a first step towards bridging the gap between document-centric analysis and visualization-centric analysis (see Figure 10.1). Thus, exploring different paths on how to transition from a text with word-scale visualizations to more substantial overview visualizations is another very interesting research direction. As has been documented by Jänicke et al. [82] humanities scholars are keen to be able to switch between a document-centric view on their source text to a visualization-centric view where the text is abstracted

and for example entities can be explored without the rigid frame of a text.

### 10.3 CONCLUSION

I believe that the research on small, contextual and data-driven graphics embedded in text is far from being done. The various design considerations for designs, placement, interaction, and authoring tools introduced in this thesis facilitates further investigations of these aspect of word-scale visualizations. Having appropriate word-scale visualizations authoring tools for both designers and non-designers for example for Wikipedia would certainly be useful to promote the use of word-scale visualizations more broadly.

The design space of small, contextual and data-driven graphics is so far largely unexplored. More empirical results are needed to confirm the validity and the benefits of using word-scale visualizations in text. The exploration of this design space is mainly limited by the creativity of the designers and researcher that will adventure into this direction. In summary, this dissertation contributes to the understanding of the value and usefulness of small, contextual and data-driven graphics for Information Visualization.



Part I

APPENDICES



# USE AND DESIGN OF SMALL CONTEXTUAL VISUALIZATIONS

## A.1 ILLUSTRATOR ARTICLES



**WIKIPEDIA**  
The Free Encyclopedia

Main page  
Contents  
Featured content  
Current events  
Random article  
Donate to Wikipedia  
Wikipedia store

Interaction  
Help  
About Wikipedia  
Community portal  
Recent changes  
Contact page

Tools  
What links here  
Related changes  
Upload file  
Special pages  
Permanent link  
Page information  
Wikidata item  
Cite this page

Print/export  
Create a book  
Download as PDF  
Printable version

Languages  
Afrikaans  
Alemannisch  
العربية  
Azərbaycanca  
Беларуская  
Беларуская (тарашкевіца)  
Български  
Bosanski  
Brezhoneg  
Català  
Cebuano  
Čeština  
Corsu  
Cymraeg  
Dansk  
Deutsch  
Eesti  
Ελληνικά  
Español  
Esperanto  
Euskara  
فارسی  
Français  
Frysk  
Gaeilge  
Gaelg  
Galego  
한국어  
Հայերեն  
Hrvatski  
Ido  
Bahasa Indonesia

Not logged in [Talk](#) [Contributions](#) [Create account](#) [Log in](#)

Article [Talk](#)

**Regions of France**

From Wikipedia, the free encyclopedia

France is divided into 18 administrative **regions** (French: *région*, [ʁe.ʒjɔ̃]), 13 of which are in **Metropolitan France** and five of which are **overseas regions**.<sup>[1]</sup> The mainland regions and Corsica are each further subdivided into **departments**, ranging in number from 2 to 13 per region for the metropolitan regions; the overseas regions technically consist of only one department each.

**History** [[edit](#)]

The term *region* was officially created by the [Law of Decentralisation](#) (2 March 1982), which also gave regions their legal status. The first direct elections for regional representatives took place on 16 March 1986.<sup>[2]</sup> In 2016, the number of regions was reduced from 27 to 18 through [amalgamation](#).

**Regions from 1982 to 2015** [[edit](#)]

Between 1982 and 2015, there were 22 regions in Metropolitan France. Before 2011, there were four **overseas regions** (French Guiana, Guadeloupe, Martinique, and Réunion); in 2011 Mayotte became the fifth overseas region.



**Region**  
*Région* (French)

**Category** Unitary State  
**Location** French Republic  
**Number** 18  
**Possible status** Overseas region (5)  
Région d'outre-mer  
**Additional status** Territorial collectivity  
Collectivité Territoriale  
**Populations** 212,645 (Mayotte) – 12,005,077 (Île-de-France)  
**Areas** 376 km<sup>2</sup> (145 sq mi) (Mayotte) – 84,061 km<sup>2</sup> (32,456 sq mi) (Aquitaine-Limousin-Poitou-Charentes)  
**Government** Region Government, National Government  
**Subdivisions** Department

**Regions and their capitals** [[edit](#)]

Regions of France						
Flag <sup>[3]</sup> ⇄	Region ⇄	French name ⇄	Other local name(s) ⇄	Capital ⇄	INSEE No. <sup>[8]</sup> ⇄	Derivation or etymology ⇄ President ⇄
	Alsace-Champagne-Ardenne-Lorraine	Alsace-Champagne-Ardenne-Lorraine	German: <i>Elsass-Champagne-Ardenne-Lothringen</i>	Strasbourg	44	Alsace, Champagne-Ardenne, and Lorraine Philippe Richert (LR)
	Aquitaine-Limousin-Poitou-Charentes	Aquitaine-Limousin-Poitou-Charentes	Occitan: <i>Aquitània-Lemosin-Poitau-Charantas</i> Basque: <i>Akìtania-Limousin-Poitou-Charentes</i> Saintongeais: <i>Aguilène-Limousin-Poitou-Charentes</i>	Bordeaux	75	Aquitaine, Limousin, and Poitou-Charentes Alain Rousset (PS)
	Auvergne-Rhône-Alpes	Auvergne-Rhône-Alpes	Occitan: <i>Auvèrnhe-Ròse-Aups</i> Arpitan: <i>Ôvèrgne-Rôno-Arpes</i>	Lyon	84	Auvergne, Rhône-Alpes Laurent Wauquiez (LR)
	Bourgogne-Franche-Comté	Bourgogne-Franche-Comté	Arpitan: <i>Borgogne-Franche-Comtât</i>	Dijon	27	Burgundy, Franche-Comté Marie-Guite Dufay (PS)
	Brittany	<i>Bretagne</i>	Breton: <i>Breizh</i> Gallo: <i>Bertaëyn</i>	Rennes	53	Duchy of Brittany Jean-Yves Le Drian (PS)

Figure A.1: Curated “Regions of France” Wikipedia article (part 1).












	Centre-Val de Loire <sup>[4]</sup>	<i>Centre-Val de Loire</i>		Orléans	24	Located in north-central France (central part of the original French language area) Straddles the middle of the Loire Valley	François Bonneau (PS)
	Île-de-France	Île-de-France		Paris	11	Province of Ile-de-France and parts of the former province of Champagne	Valérie Pécresse (LR)
	Languedoc-Roussillon-Midi-Pyrénées	Languedoc-Roussillon-Midi-Pyrénées	Occitan: <i>Lengadòc-Rosselhon-Miègjorn-Pirenèus</i> Catalan: <i>Llenguadoc-Rosselló-Mieidia-Pirenèus</i> Spanish: <i>Languedoc-Rosellón-Mediodía-Pirineos</i>	Toulouse	76	Languedoc-Roussillon, Midi-Pyrénées	Carole Delga (PS)
	Nord-Pas-de-Calais-Picardy	Nord-Pas-de-Calais-Picardie	French Flemish: <i>Noord-Nauw-van-Kales-Picardië</i>	Lille	32	Nord-Pas-de-Calais, Picardie	Xavier Bertrand (LR)
	Normandy	<i>Normandie</i>	Norman: <i>Normaundie</i>	Rouen	28	Duchy of Normandy	Hervé Morin (LR)
	Pays de la Loire	Pays de la Loire	Breton: <i>Broioù al Liger</i>	Nantes	52	None; created for Nantes	Bruno Retailleau (LR)
	Provence-Alpes-Côte d'Azur (PACA)	Provence-Alpes-Côte d'Azur (PACA)	Occitan: <i>Provença-Aups-Còsta d'Azur</i> <i>Prouvènço-Aup-Costo d'Azur</i>	Marseille	93	Provence plus the former county of Nice, principality of Orange, Comtat Venaissin and a part of Dauphiné	Christian Estrosi (LR)
	Corsica	<i>Corse</i>	Corsican: <i>Corsica</i>	Ajaccio	94	Territorial collectivity	Paul Giacobbi (PRG)
The following five <i>overseas departments</i> also have the special status of <i>overseas region</i> .							
	French Guiana	<i>Guyane</i>		Cayenne	03	Overseas region	Rodolphe Alexandre (PSG)
	Guadeloupe	Guadeloupe	Antillean Creole: <i>Gwadeloup</i>	Basse-Terre	01	Overseas region	Ary Chalus (GUSR)
	Martinique	Martinique	Antillean Creole: <i>Matinik</i>	Fort-de-France	02	Overseas region	Alfred Marie-Jeanne (MIM)
	Mayotte	Mayotte	Shimaore: <i>Maore</i> Malagasy: <i>Mahori</i>	Mamoudzou	05	Overseas region	Daniel Zaïdani (DVG)
	Réunion	La Réunion	Reunion Creole: <i>La Rényon</i>	Saint-Denis	04	Overseas region	Didier Robert (LR)

Figure A.2: Curated “Regions of France” Wikipedia article (part 2).



**WIKIPEDIA**  
The Free Encyclopedia

Main page  
Contents  
Featured content  
Current events  
Random article  
Donate to Wikipedia  
Wikipedia store

Interaction  
Help  
About Wikipedia  
Community portal  
Recent changes  
Contact page

Tools  
What links here  
Related changes  
Upload file  
Special pages  
Permanent link  
Page information  
Wikidata item  
Cite this page

Print/export  
Create a book  
Download as PDF  
Printable version

Languages 

Acèh  
 Адыгэбзэ  
 Afrikaans  
 Akan  
 Alemannisch  
 አማርኛ  
 Ἀγγλικά  
 Аԥсшәа  
 العربية  
 Aragonés  
 Ἀρμενացի  
 Armãneashti  
 Arpetan  
 অসমীয়া  
 Asturianu  
 Avañe'ê  
 Авар  
 Аymar aru  
 Azərbaycanca  
 تۆرکجه  
 Bamanankan  
 বাংলা  
 Bahasa Banjar  
 БӀаң-ІоӀм-гӀу  
 Basa Banyumasan  
 Башҡортса  
 Беларуская  
 Беларуская (тарашкевіца)  
 भोजपुरी  
 Bikol Central  
 Bislama  
 Български  
 Boarisch  
 བོད་སྐད་  
 Bosanski  
 Brezhoneg  
 Буряад

Not logged in [Talk](#) [Contributions](#) [Create account](#) [Log in](#)

Article [Talk](#) [Read](#) [View source](#) [View history](#)

## Europe

From Wikipedia, the free encyclopedia

*This article is about the continent. For the politico-economic union, see [European Union](#). For other uses, see [Europe \(disambiguation\)](#).*

**Europe** (<sup>i</sup>ˈjʊərəp/ or <sup>i</sup>ˈjɜːrəp/<sup>[4]</sup>) is a **continent** that comprises the westernmost part of **Eurasia**. Europe is bordered by the **Arctic Ocean** to the north, the **Atlantic Ocean** to the west, and the **Mediterranean Sea** to the south. To the east and southeast, Europe is generally considered as **separated from Asia** by the **watershed divides** of the **Ural** and **Caucasus Mountains**, the **Ural River**, the **Caspian** and **Black Seas**, and the waterways of the **Turkish Straits**.<sup>[5]</sup> Yet the borders of Europe—a concept dating back to **classical antiquity**—are arbitrary, as the primarily **physiographic** term "continent" also incorporates **cultural and political** elements.

Europe is the world's **second-smallest** continent by surface area, covering about 10,180,000 square kilometres (3,930,000 sq mi) or 2% of the Earth's surface and about 6.8% of its land area. Of Europe's approximately 50 countries, **Russia** is by far the largest by both area and population, taking up 40% of the continent (although the country has territory in both Europe and Asia), while **Vatican City** is the smallest. Europe is the third-most populous continent after **Asia** and **Africa**, with a **population** of 739–743 million or about 11% of the **world's population**.<sup>[6]</sup> Europe has a climate heavily affected by warm Atlantic currents, tempering winters and enabling warm summers on most of the continent, even on latitudes that have severe climates in North America and Asia. Further from the Atlantic, seasonal differences increase, but the mildness of the climate remains.

Europe, in particular **ancient Greece**, is the birthplace of **Western culture**.<sup>[7][8][9]</sup> The fall of the Western Roman Empire, during the **migration period**, marked the end of **ancient history** and the beginning of an era known as the "**middle ages**". The **Renaissance** **humanism**, **exploration**, **art**, and **science** led the "old continent", and eventually the rest of the world, to the **modern era**. From this period onwards, Europe played a predominant role in global affairs. Between the 16th and 20th centuries, European nations controlled at various times the Americas, most of Africa, Oceania, and the majority of Asia.

The **Industrial Revolution**, which began in the **United Kingdom** around the end of the 18th century, gave rise to radical economic, cultural, and social change in Western Europe, and eventually the wider world. Demographic growth meant that, by 1900, Europe's share of the world's population was 25%.<sup>[10]</sup> Both **world wars** were largely focused upon Europe, greatly contributing to a decline in **Western European** dominance in world affairs by the mid-20th century as the **United States** and **Soviet Union** took prominence.<sup>[11]</sup> During the **Cold War**, Europe was divided along the **Iron Curtain** between **NATO** in the west and the **Warsaw Pact** in the east, until the **revolutions of 1989** and **fall of the Berlin Wall**.

European integration led to the formation of the **European Union**, a political entity that lies between a **confederation** and a **federation**.<sup>[12]</sup> The EU originated in Western Europe but has been expanding eastward since the **fall of the Soviet Union** in 1991. The currency of the European Union, the **Euro**, is the most commonly used among Europeans and the EU's **Schengen Area** abolishes border and immigration controls among most of its member states.

### History

*Main article: [History of Europe](#)*

#### 20th century to the present

*Main articles: [Modern era](#) and [History of Europe](#)*  
*See also: [World War I](#), [Great Depression](#), [Interwar period](#), [World War II](#), [Cold War](#) and [History of the European Union](#)*

Two World Wars and an economic depression dominated the first half of the 20th century. World War I was fought between 1914 and 1918. It started when **Archduke Franz Ferdinand** of Austria was assassinated by the Yugoslav nationalist<sup>[145]</sup> **Gavrilo Princip**.<sup>[146]</sup> Most European nations were drawn into the war, which was fought between the **Entente Powers** (**France**, **Belgium**, **Serbia**, **Portugal**, **Russia**, the **United Kingdom**, and later **Italy**, **Greece**, **Romania**, and the **United States**) and the **Central Powers** (**Austria-Hungary**, **Germany**, **Bulgaria**, and the **Ottoman Empire**). The War left more than 16 million civilians and military dead.<sup>[147]</sup> Over 60 million European soldiers were mobilised from 1914 to 1918.<sup>[148]</sup>

Russia was plunged into the **Russian Revolution**, which threw down the **Tsarist monarchy** and replaced it with the **communist Soviet Union**.<sup>[149]</sup> **Austria-Hungary** and the Ottoman Empire collapsed and broke up into separate nations, and many other nations had their borders redrawn. The **Treaty of Versailles**, which officially ended World War I in 1919, was harsh towards Germany, upon whom it placed full responsibility for the war and imposed heavy sanctions.<sup>[150]</sup>

### Europe



<b>Area</b>	10,180,000 km <sup>2</sup> (3,930,000 sq mi) <sup>[n]</sup> (6th)
<b>Population</b>	742,452,000 <sup>[n]</sup> (2013; 3rd)
<b>Pop. density</b>	72.9/km <sup>2</sup> (188/sq mi) (2nd)
<b>Demonym</b>	European
<b>Countries</b>	~50 countries (and ~5 with limited recognition)
<b>Dependencies</b>	4 dependencies
<b>Languages</b>	~225 languages <sup>[1]</sup>
<b>Time zones</b>	UTC to UTC+5
<b>Largest cities</b>	Urban areas in Europe <sup>[2]</sup>

-  **Istanbul** (transcontinental)<sup>[3]</sup>
-  **Moscow**
-  **Paris**
-  **London**
-  **Madrid**
-  **Barcelona**
-  **St. Petersburg**
-  **Rome**
-  **Berlin**



Figure A.3: Curated “Europe” Wikipedia article (part 1).

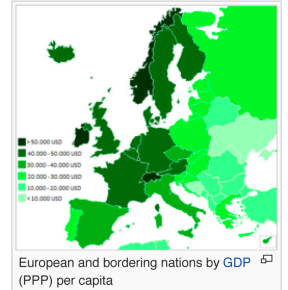
## Economy

*Main article: [Economy of Europe](#)*

As a continent, the economy of Europe is currently the largest on Earth and it is the richest region as measured by assets under management with over \$32.7 trillion compared to North America's \$27.1 trillion in 2008.<sup>[189]</sup> In 2009 Europe remained the wealthiest region. Its \$37.1 trillion in assets under management represented one-third of the world's wealth. It was one of several regions where wealth surpassed its precrisis year-end peak.<sup>[190]</sup> As with other continents, Europe has a large variation of wealth among its countries. The richer states tend to be in the [West](#); some of the [Central and Eastern European](#) economies are still emerging from the [collapse of the Soviet Union](#) and [Yugoslavia](#).

The European Union, a political entity composed of 28 European states, comprises the [largest single economic area](#) in the world. 18 EU countries share the [euro](#) as a common currency. Five European countries rank in the top ten of the world's largest [national economies in GDP \(PPP\)](#). This includes (ranks according to the [CIA](#)): Germany (5), the UK (6), Russia (7), France (8), and Italy (10).<sup>[191]</sup>

There is huge disparity between many European countries in terms of their income. The richest in terms of GDP per capita is Monaco with its US\$172,676 per capita (2009) and the poorest is Moldova with its GDP per capita of US\$1,631 (2010).<sup>[192]</sup> [Monaco](#) is the richest country in terms of GDP per capita in the world according to the World Bank report.



**Table**

Rank	Country	GDP (nominal, 2014) millions of USD	Rank	Country	GDP (PPP, 2014) millions of USD
1	<a href="#">Germany</a>	3,874,437	1	<a href="#">Germany</a>	3,748,094
2	<a href="#">United Kingdom</a>	2,950,039	2	<a href="#">Russia</a>	3,745,157
3	<a href="#">France</a>	2,833,687	3	<a href="#">France</a>	2,591,170
4	<a href="#">Italy</a>	2,147,744	4	<a href="#">United Kingdom</a>	2,569,218
5	<a href="#">Russia</a>	1,860,598	5	<a href="#">Italy</a>	2,135,359
6	<a href="#">Spain</a>	1,406,538	6	<a href="#">Spain</a>	1,572,112
7	<a href="#">Netherlands</a>	880,716	7	<a href="#">Turkey</a>	1,514,859
8	<a href="#">Turkey</a>	799,535	8	<a href="#">Poland</a>	959,845
9	<a href="#">Switzerland</a>	703,852	9	<a href="#">Netherlands</a>	808,796
10	<a href="#">Sweden</a>	570,591	10	<a href="#">Belgium</a>	483,331

## Demographics

*Main article: [Demographics of Europe](#)*

*See also: [List of European countries by population](#), [Ethnic groups in Europe](#), [Immigration to Europe](#), [Emigration from Europe](#) and [Ageing of Europe](#)*

## Language

*Main article: [Languages of Europe](#)*

European languages mostly fall within three [Indo-European](#) language groups: the [Romance languages](#), derived from the [Latin](#) of the [Roman Empire](#); the [Germanic languages](#), whose ancestor language came from southern Scandinavia; and the [Slavic languages](#).<sup>[177]</sup>

Many other languages outside the three main groups exist in Europe. Other Indo-European languages include the [Baltic](#) group (that is, [Latvian](#) and [Lithuanian](#)), the [Celtic](#) group (that is, [Irish](#), [Scottish Gaelic](#), [Manx](#), [Welsh](#), [Cornish](#), and [Breton](#)<sup>[177]</sup>), [Greek](#), [Armenian](#), and [Albanian](#).

In addition, a distinct group of [Uralic languages](#) ([Estonian](#), [Finnish](#), and [Hungarian](#)) is spoken mainly in [Estonia](#), [Finland](#), and [Hungary](#), while [Kartvelian languages](#) ([Georgian](#), [Mingrelian](#), and [Svan](#)), are spoken primarily in [Georgia](#), and two other language families reside in the North Caucasus (termed [Northeast Caucasian](#), most notably including [Chechen](#), [Avar](#) and [Lezgin](#) and [Northwest Caucasian](#), notably including [Adyghe](#)). [Maltese](#) is the only [Semitic language](#) that is official within the EU, while [Basque](#) is the only European [language isolate](#). [Turkic languages](#) include [Azerbaijani](#) and [Turkish](#), in addition to the [languages of minority nations](#) in Russia.



Figure A.4: Curated “Europe” Wikipedia article (part 2).

Integration

Main article: [European integration](#)  
See also: [International organizations in Europe](#)

European integration is the process of political, legal, economic (and in some cases social and cultural) integration of states wholly or partially in Europe. In the present day, European integration has primarily come about through the Council of Europe and European Union in Western and Central Europe and [Commonwealth of Independent States](#) in Central and Eastern Europe and most of former Soviet countries.



Geography

Main article: [Geography of Europe](#)

Climate

Main article: [Climate of Europe](#)

Europe lies mainly in the [temperate](#) climate zones, being subjected to [prevailing westerlies](#). The climate is milder in comparison to other areas of the same latitude around the globe due to the influence of the [Gulf Stream](#).<sup>[175]</sup> The Gulf Stream is nicknamed "Europe's central heating", because it makes Europe's climate warmer and wetter than it would otherwise be. Therefore, the average temperature throughout the year of Naples is 16 °C (60.8 °F), while it is only 12 °C (53.6 °F) in New York City which is almost on the same latitude. Berlin, Germany; Calgary, Canada; and Irkutsk, in the Asian part of Russia, lie on around the same latitude; January temperatures in Berlin average around 8 °C (15 °F) higher than those in Calgary, and they are almost 22 °C (40 °F) higher than average temperatures in Irkutsk.<sup>[175]</sup> Similarly, northern parts of Scotland have a temperate marine climate. The yearly average temperature in city of Inverness is 9.05 degrees Celsius (48.3 degrees Fahrenheit). However, Churchill, Manitoba, Canada, is on roughly the same latitude and has an average temperature of -6.5 degrees Celsius (20.3 degrees Fahrenheit), giving it a nearly subarctic climate.

Biodiversity

See also: [Fauna of Europe](#)

Having lived side-by-side with agricultural peoples for millennia, Europe's animals and plants have been profoundly affected by the presence and activities of man. With the exception of [Fennoscandia](#) and northern Russia, few areas of untouched wilderness are currently found in Europe, except for various [national parks](#). As for the animals, in many parts of Europe most large animals and top [predator](#) species have been hunted to extinction. The [woolly mammoth](#) was extinct before the end of the [Neolithic](#) period. Today [wolves](#) ([carnivores](#)) and [bears](#) ([omnivores](#)) are endangered. Once they were found in most parts of Europe. However, deforestation and hunting caused these animals to withdraw further and further. By the [Middle Ages](#) the bears' habitats were limited to more or less inaccessible mountains with sufficient forest cover. Today, the [brown bear](#) lives primarily in the [Balkan peninsula](#), Scandinavia, and Russia; a small number also persist in other countries across Europe (Austria, Pyrenees etc.), but in these areas brown bear populations are fragmented and marginalised because of the destruction of their habitat. In addition, [polar bears](#) may be found on [Svalbard](#), a Norwegian archipelago far north of Scandinavia. The [wolf](#), the second largest predator in Europe after the brown bear, can be found primarily in [Central and Eastern Europe](#) and in the Balkans, with a handful of packs in pockets of [Western Europe](#) (Scandinavia, Spain, etc.).

European wild cat, foxes (especially the red fox), jackal and different species of martens, hedgehogs, different species of reptiles (like snakes such as vipers and grass snakes) and amphibians, different birds (owls, hawks and other birds of prey).

Important European herbivores are snails, larvae, fish, different birds, and mammals, like rodents, deer and roe deer, boars, and living in the mountains, marmots, steinbocks, chamois among others. A number of insects, such as the [small tortoiseshell](#) butterfly, add to the biodiversity.<sup>[182]</sup>

The extinction of the [dwarf hippos](#) and [dwarf elephants](#) has been linked to the earliest arrival of humans on the islands of the [Mediterranean](#).<sup>[183]</sup>

Sea creatures are also an important part of European flora and fauna. The sea flora is mainly [phytoplankton](#). Important animals that live in European seas are [zooplankton](#), [molluscs](#), [echinoderms](#), different [crustaceans](#), [squids](#) and [octopuses](#), fish, [dolphins](#), and [whales](#).



Once roaming the great temperate forests of Eurasia, [European bison](#) now live in nature preserves in [Białowieża Forest](#), on the border between [Poland](#) and [Belarus](#).<sup>[180][181]</sup>

Figure A.5: Curated “Europe” Wikipedia article (part 3).





[Main page](#)  
[Contents](#)  
[Featured content](#)  
[Current events](#)  
[Random article](#)  
[Donate to Wikipedia](#)  
[Wikipedia store](#)

#### Interaction

[Help](#)  
[About Wikipedia](#)  
[Community portal](#)  
[Recent changes](#)  
[Contact page](#)

#### Tools

[What links here](#)  
[Related changes](#)  
[Upload file](#)  
[Special pages](#)  
[Permanent link](#)  
[Page information](#)  
[Wikidata item](#)  
[Cite this page](#)

#### Print/export

[Create a book](#)  
[Download as PDF](#)  
[Printable version](#)

#### Languages

[/English](#)  
[العربية](#)  
[অসমীয়া](#)  
[বাংলা](#)  
[Bân-lâm-gú](#)  
[Български](#)  
[Català](#)  
[Čeština](#)  
[Cymraeg](#)  
[Deutsch](#)  
[Eesti](#)  
[Español](#)  
[Esperanto](#)  
[فارسی](#)  
[Français](#)  
[Gaeilge](#)  
[Galego](#)  
[日本語](#)  
[ગુજરાતી](#)  
[한국어](#)  
[हिन्दी](#)  
[Bahasa Indonesia](#)  
[Inupiak](#)  
[Íslenska](#)  
[Italiano](#)  
[עברית](#)  
[ಕನ್ನಡ](#)  
[Қазақша](#)  
[Kinyarwanda](#)

[Not logged in](#)
[Talk](#)
[Contributions](#)
[Create account](#)
[Log in](#)

[Article](#)
[Talk](#)

[Read](#)
[Edit](#)
[View history](#)

## Human body

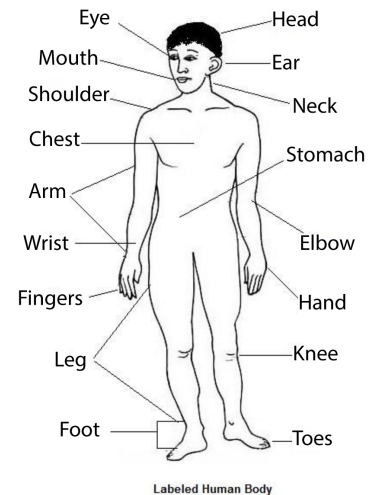
From Wikipedia, the free encyclopedia

*"Physiologies" redirects here. For other uses, see [Physiology](#).*

The **human body** is the entire structure of a **human being** and comprises a **head**, **neck**, **trunk** (which includes the **thorax** and **abdomen**), **arms** and **hands**, **legs** and **feet**. Every part of the body is composed of various types of **cells**, the fundamental unit of life.<sup>[1]</sup>

At maturity, the estimated average number of cells in the body is given as 37.2 **trillion**. This number is stated to be of partial data and to be used as a starting point for further calculations. The number given is arrived at by totalling the cell numbers of all the **organs** of the body and cell types.<sup>[2]</sup> The **composition of the human body** is made up of a number of certain **elements** including **carbon**, **calcium** and **phosphorus**.

The study of the human body involves **anatomy** and **physiology**. The human body can show anatomical non-pathological anomalies known as **variations** which need to be able to be recognised. Physiology focuses on the systems and their organs of the human body and their functions. Many systems and mechanisms interact in order to maintain **homeostasis**.

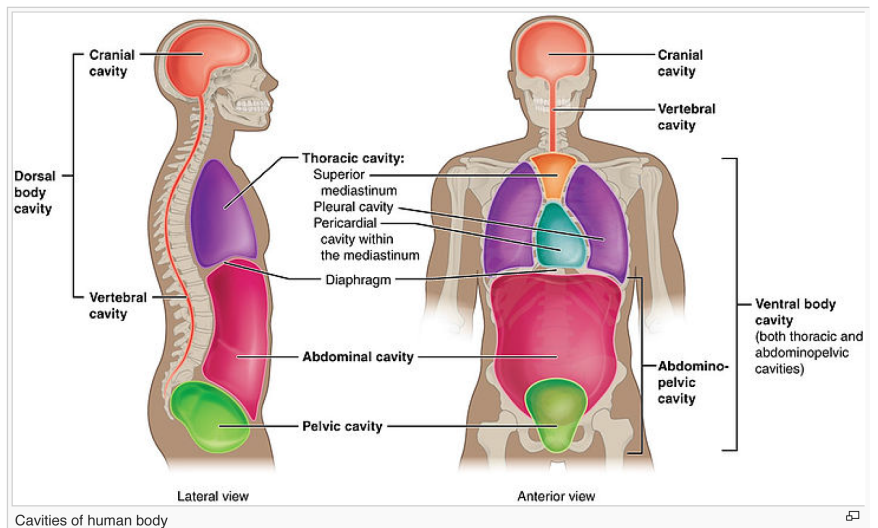


Labeled Human Body

### Structure [ edit ]

Further information: *[Anatomy](#)*, *[Body proportions](#)* and *[Anatomical terminology](#)*

**Skeletal structure** frames the overall shape of the body and does not alter much over a lifetime. General **body shape** (and **female body shape**) is influenced by the distribution of **muscle** and **fat tissue** and is also affected by various **hormones**. The average **height** of an adult male human (in developed countries) is about 1.7–1.8 m (5'7" to 5'11") and the adult female is about 1.6–1.7 m (5'2" to 5'7"). Height is largely determined by **genes** and diet. **Body**



**type** and **body composition** are influenced by factors such as **genetics**, **diet**, and **exercise**.

The human body has several **body cavities** the largest of which is the **abdominopelvic cavity**. These cavities house the various body **organs** including the **spinal cord** which also accommodates the production and flow of **cerebrospinal fluid** in the **ventricular system** of the **brain**.

Figure A.6: Curated “Human Body” Wikipedia article (part 1).



Many other smaller cavities exist throughout the body called **sinuses**, which have varied functions. *Sinuses* in general usage refers to the **paranasal sinuses** which are involved in the condition **sinusitis**. The paranasal sinuses are four pairs of vital air-cavities in the **cranial bones**. These **air-filled spaces** are paired between the eyes, above the eyes, deeper behind the eyes and around the **nasal cavity**.

**Composition** [ edit ]  
Main article: *Composition of the human body*

The average adult body contains between 5 and 5½ litres of blood and approximately 10 litres of **interstitial fluid**.

The composition of the human body can be referred to in terms of its water content, elements content, tissue types or material types. The adult human body contains approximately 60% **water**, and so makes up a significant proportion of the body, both in terms of weight and volume. Water content can vary from a high 75% in a newborn infant to a lower 45% in an **obese** person. (These figures are necessarily **statistical averages**).

The vast majority of cells in the human body are not human at all; rather they are of **bacteria**, **archaea**, and **methanogens** such as *Methanobrevibacter smithii*. The largest proportion of these form the **gut flora**. The whole population of microbiota include microorganisms of the skin and other body parts and this altogether is termed as the **human microbiome**.

The proportions of the elements of the body can be referred to in terms of the main elements, minor ones and **trace elements**. Material type may also be referred to as including water, **protein**, **connective tissue**, **fats**, **carbohydrates** and **bone**.

**Systems** [ edit ]  
See also: *List of systems of the human body*

The human body consists of many interacting systems. Each system contributes to the maintenance of homeostasis, of itself, other systems, and the entire body. A system consists of two or more **organs**, which are functional collections of tissue. Systems do not work in isolation, and the well-being of the person depends upon the well-being of all the interacting body systems. Some combining systems are referred to by their joint names such as the nervous system and the endocrine system known together as the **neuroendocrine system**.

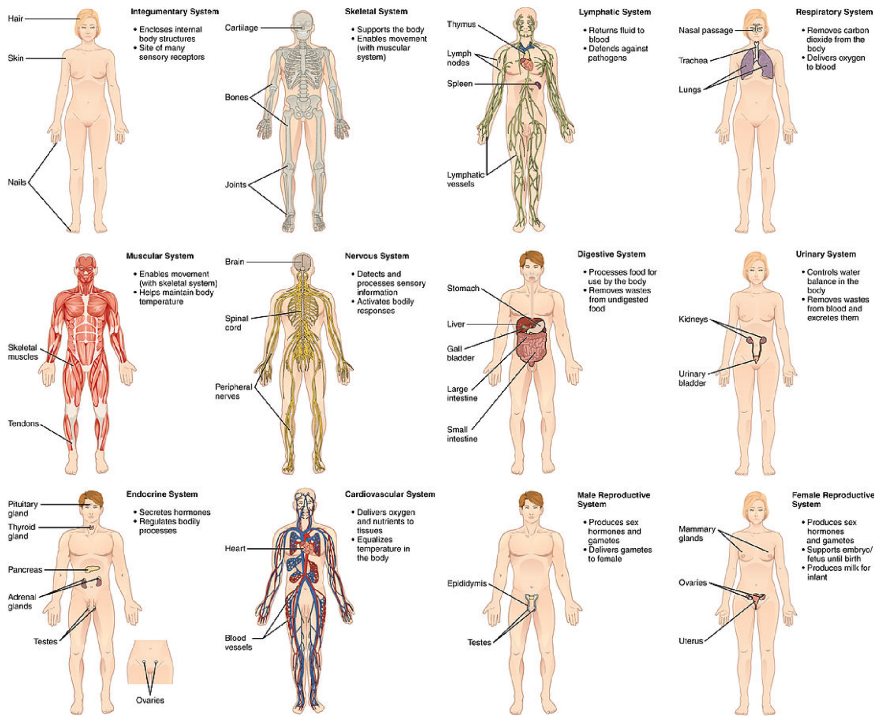


Figure A.7: Curated “Human Body” Wikipedia article (part 2).








#	System	Clinical study	Physiology
1.	 <p>The <b>circulatory system</b> or <b>cardiovascular system</b> comprises the <b>heart</b> and <b>blood vessels</b> (<b>arteries</b>, <b>veins</b>, and <b>capillaries</b>). The heart propels the circulation of the <b>blood</b>, which serves as a "transportation system" to transfer <b>oxygen</b>, fuel, nutrients, waste products, immune cells, and signalling molecules (i.e., <b>hormones</b>) from one part of the body to another. The <b>blood</b> consists of fluid that carries <b>cells</b> in the circulation, including some that move from tissue to blood vessels and back, as well as the <b>spleen</b> and <b>bone marrow</b>.</p>	<p><b>cardiology</b> (heart), <b>hematology</b> (blood)</p>	<p><b>cardiovascular physiology</b><sup>[10][11]</sup> The heart itself is divided into three layers called the <b>endocardium</b>, <b>myocardium</b> and <b>epicardium</b>, which vary in thickness and function.<sup>[12]</sup></p>
2.	 <p>The <b>digestive system</b> / <b>Excretory system</b> consists of the <b>mouth</b> including the <b>tongue</b> and <b>teeth</b>, <b>esophagus</b>, <b>stomach</b>, (<b>gastrointestinal tract</b>, <b>small</b> and <b>large intestines</b>, and <b>rectum</b>), as well as the <b>liver</b>, <b>pancreas</b>, <b>gallbladder</b>, and <b>salivary glands</b>. It converts food into small, nutritional, non-toxic <b>molecules</b> for distribution by the circulation to all tissues of the body, and excretes the unused residue.</p>	<p><b>gastroenterology</b></p>	<p><b>gastrointestinal physiology</b></p>
3.	 <p>The <b>endocrine system</b> consists of the principal endocrine glands: the <b>pituitary</b>, <b>thyroid</b>, <b>adrenals</b>, <b>pancreas</b>, <b>parathyroids</b>, and <b>gonads</b>, but nearly all organs and tissues produce specific endocrine <b>hormones</b> as well. The endocrine hormones serve as signals from one body system to another regarding an enormous array of conditions, and resulting in variety of changes of function. There is also the <b>exocrine system</b>.</p>	<p><b>endocrinology</b></p>	<p>endocrinology</p>
4.	 <p>The <b>immune system</b> consists of the <b>white blood cells</b>, the <b>thymus</b>, <b>lymph nodes</b> and <b>lymph</b> channels, which are also part of the <b>lymphatic system</b>. The immune system provides a mechanism for the body to distinguish its own cells and tissues from alien cells and substances and to neutralize or destroy the latter by using specialized proteins such as <b>antibodies</b>, <b>cytokines</b>, and <b>toll-like receptors</b>, among many others.</p>	<p><b>immunology</b></p>	<p>immunology</p>
5.	 <p>The <b>integumentary system</b> consists of the covering of the body (the <b>skin</b>), including <b>hair</b> and <b>nails</b> as well as other functionally important structures such as the <b>sweat glands</b> and <b>sebaceous glands</b>. The skin provides containment, structure, and protection for other organs, but it also serves as a major sensory interface with the outside world.</p>	<p><b>dermatology</b></p>	<p>cell physiology, <b>skin physiology</b></p>
6.	 <p>The main function of the <b>lymphatic system</b> / <b>Immune system</b> is to extract, transport and metabolize lymph, the fluid found in between cells. The lymphatic system is very similar to the circulatory system in terms of both its structure and its most basic function (to carry a body fluid).</p>	<p><b>oncology</b>, <b>immunology</b></p>	<p>oncology, immunology</p>
7.	 <p>The <b>musculoskeletal system</b> consists of the <b>human skeleton</b> (which includes <b>bones</b>, <b>ligaments</b>, <b>tendons</b>, and <b>cartilage</b>) and attached <b>muscles</b>. It gives the body basic structure and the ability for movement. In addition to their structural role, the larger bones in the body contain <b>bone marrow</b>, the site of production of blood cells. Also, all bones are major storage sites for <b>calcium</b> and <b>phosphate</b>. This system can be split up into the <b>muscular system</b> and the <b>skeletal system</b>.</p>	<p><b>orthopedics</b> (bone and muscle disorders and injuries)</p>	<p><b>cell physiology</b>, <b>musculoskeletal physiology</b>, <b>osteology</b> (skeleton), <b>arthrology</b> (articular system), <b>myology</b> (muscular system)<sup>[13]</sup></p>

Figure A.8: Curated "Human Body" Wikipedia article (part 3).






8.		The <b>nervous system</b> consists of the <b>central nervous system</b> (the <b>brain</b> and <b>spinal cord</b> ) and the <b>peripheral nervous system</b> consists of the <b>nerves</b> and <b>ganglia</b> outside of the brain and spinal cord. The brain is the organ of thought, emotion, memory, and sensory processing, and serves many aspects of communication and controls various systems and functions. The <b>special senses</b> consist of <b>vision</b> , <b>hearing</b> , <b>taste</b> , and <b>smell</b> . The <b>eyes</b> , <b>ears</b> , <b>tongue</b> , and <b>nose</b> gather information about the body's environment.	<b>neuroscience</b> , <b>neurology</b> (disease), <b>psychiatry</b> (behavioral), <b>ophthalmology</b> (vision), <b>otolaryngology</b> (hearing, taste, smell)	<b>neurophysiology</b>
9.		The <b>renal system / urinary system</b> consists of the <b>kidneys</b> , <b>ureters</b> , <b>bladder</b> , and <b>urethra</b> . It removes water from the blood to produce urine, which carries a variety of waste molecules and excess <b>ions</b> and water out of the body.	<b>nephrology</b> (function), <b>urology</b> (structural disease)	<b>renal physiology</b>
10.		The <b>reproductive system</b> consists of the <b>gonads</b> and the internal and external <b>sex organs</b> . The reproductive system produces <b>gametes</b> in each <b>sex</b> , a mechanism for their combination, and a nurturing environment for the first 9 months of development of the infant.	<b>gynecology</b> (women), <b>andrology</b> (men), <b>sexology</b> (behavioral aspects), <b>embryology</b> (developmental aspects), <b>obstetrics</b> (partition)	<b>reproductive physiology</b>
11.		The <b>respiratory system</b> consists of the <b>nose</b> , <b>nasopharynx</b> , <b>trachea</b> , and <b>lungs</b> . It brings oxygen from the air and excretes <b>carbon dioxide</b> and <b>water</b> back into the air.	<b>pulmonology</b>	<b>respiratory physiology</b>

Figure A.9: Curated “Human Body” Wikipedia article (part 4).



WIKIPEDIA  
The Free Encyclopedia

[Main page](#)  
[Contents](#)  
[Featured content](#)  
[Current events](#)  
[Random article](#)  
[Donate to Wikipedia](#)  
[Wikipedia store](#)

[Interaction](#)  
[Help](#)  
[About Wikipedia](#)  
[Community portal](#)  
[Recent changes](#)  
[Contact page](#)

[Tools](#)  
[What links here](#)  
[Related changes](#)  
[Upload file](#)  
[Special pages](#)  
[Permanent link](#)  
[Page information](#)  
[Wikidata item](#)  
[Cite this page](#)

[Print/export](#)  
[Create a book](#)  
[Download as PDF](#)  
[Printable version](#)

[Languages](#)  
★ Afrikaans  
Alemannisch  
አማርኛ  
العربية  
Aragonés  
অসমীয়া  
Arpetan  
Azərbaycanca  
বাংলা  
Bân-lâm-gú  
Беларуская  
Беларуская (тарашкевіца)  
Български  
Boarisch  
Bosanski  
Brezhoneg  
Català  
ЧӀавашла  
Čeština  
Chavacano de Zamboanga  
Cymraeg  
Dansk  
Deutsch

[Article](#) [Talk](#)

[Read](#) [Edit](#) [View history](#)

## Zürich

From Wikipedia, the free encyclopedia

Coordinates: 47°22′N 8°33′E﻿ / ﻿

*For other uses, see [Zürich \(disambiguation\)](#).*

**Zürich** or **Zurich** (/ˈzʊərɪk/, Swiss Standard German *Zürich* [ˈtsʏrɪç], German Standard German *Zürich* [ˈtsʏːʁɪç], Swiss German: *Züri* [ˈtsʏrɪ], French: *Zurich* [zyʁik], Italian: *Zurigo* [dzuˈriɡo], Romansh: *Turitg* [tuˈritɕ]) is the **largest city in Switzerland** and the capital of the **canton of Zürich**. It is located in north-central Switzerland<sup>[3]</sup> at the northwestern tip of **Lake Zürich**. The municipality has approximately 400,028<sup>[4]</sup> inhabitants, the urban agglomeration 1.315 million,<sup>[5]</sup> and the **Zurich metropolitan area** 1.83 million.<sup>[6]</sup> Zürich is a hub for railways, roads, and air traffic. Both **Zürich Airport** and **railway station** are the largest and busiest in the country.

Permanently settled for around 2000 years, Zürich has a history that goes back to its founding by the Romans, who, in 15 BC, called it *Turicum*. However, early settlements have been found dating back more than 6400 years ago.<sup>[7]</sup> During the Middle Ages Zürich gained the independent and privileged status of **imperial immediacy** and, in 1519, became a primary centre of the **Protestant Reformation** in Europe under the leadership of **Ulrich Zwingli**.<sup>[8]</sup>




The official language of Zürich is (the Swiss variety of Standard) **German**, but the main spoken language is the local variant of the **Alemannic Swiss German** dialect.

Zürich is a leading **global city** and among the world's largest financial centres despite a relatively low population.<sup>[9]</sup> The city is home to a large number of financial institutions and banking giants. Most of Switzerland's research and development centres are concentrated in Zürich and the low tax rates attract overseas companies to set up their headquarters there.


**Monocle's** 2012 "Quality of Life Survey" ranked Zürich first on a list of the top 25 cities in the world "to make a base within".<sup>[10]</sup> According to several surveys from 2006 to 2008, Zürich was named the city with the **best quality of life** in the world as well as the wealthiest city in Europe.<sup>[11][12][13]</sup> The **Economist Intelligence Unit's** Global Liveability Ranking<sup>[14]</sup> sees Zürich rank among the top ten most liveable cities in the world.

Many museums and art galleries can be found in the city, including the **Swiss National Museum** and the **Kunsthhaus**.<sup>[15]</sup> **Schauspielhaus Zürich** is one of the most important theatres in the German-speaking world.<sup>[16]</sup>


### Zürich

*Top:* View over Zürich and the lake  
*Middle:* **Fraumünster Church** on the river  
*Limmat* (left) and the **Sunrise Tower** (right)  
*Bottom:* Night view of Zürich from **Uetliberg**



Coat of arms



**Location of Zürich** [show]  
Coordinates: 47°22′N 8°33′E﻿ / ﻿

<b>Country</b>	Switzerland
<b>Canton</b>	Zürich
<b>District</b>	Zürich
<b>Government</b>	
<span> </span> • <b>Executive</b>	<i>Stadtrat</i> with 9 members
<span> </span> • <b>Mayor</b>	<i>Stadtpäsidentin</i> (list) <b>Corine Mauch SPS/PSS</b> (as of February 2014)
<span> </span> • <b>Parliament</b>	<i>Gemeinderat</i> with 125 members

Figure A.10: Curated “Zurich” Wikipedia article (part 1).



**Government** [edit]

See also: [List of mayors of Zürich](#)

The City Council (**Stadtrat**) constitutes the executive government of the City of Zürich and operates as a collegiate authority. It is composed of nine councilors, each presiding over a department. Departmental tasks, coordination measures and implementation of laws decreed by the City Parliament are carried out by the City Council. The regular election of the City Council by any inhabitant valid to vote is held every four years. The mayor ([German: \*Stadtpräsident\(in\)\*](#)) is elected as such by a public election while the heads of the other departments are assigned by the collegiate. Current mayor is [Corine Mauch](#). The executive body holds its meetings in the City Hall (*Stadthaus*), on the left bank of the Limmat. The building was built in 1883 in Renaissance style.

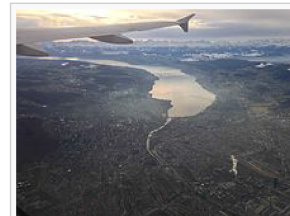
As of 2015, the Zürich City Council is made up of four representatives of the SP ([Social Democratic Party](#), one of whom is the mayor), two members of the FDP ([Free Democratic Party](#)) and one member each of [Green Party](#), CVP ([Christian Democratic Party](#)) and AL ([Alternative Left Party](#)), giving the left parties a combined six out of nine seats.<sup>[31]</sup>

**Geography** [edit]

The [Limmat](#) in Zürich

Zürich is situated at 408 m (1,339 ft) [above sea level](#) on the lower (northern) end of [Lake Zürich](#) (*Zürichsee*) about 30 kilometers (19 mi) north of the [Alps](#), nestling between the wooded hills on the west and east side. The Old Town stretches on both sides of the [Limmat](#), which flows from the lake, running northwards at first and then gradually turning into a curve to the west. The geographic (and historic) centre of the city is the [Lindenhof](#), a small natural hill on the west bank of the Limmat,

about 700 m (2,300 ft) north of where the river issues from Lake Zürich. Today the incorporated city stretches somewhat beyond the natural confines of the hills and includes some districts to the northeast in the [Glatt Valley](#) (German: *Glattal*) and to the north in the [Limmat Valley](#) (German: *Limmattal*). The boundaries of the older city are easy to recognize by the Schanzengraben canal. This artificial watercourse has been used for the construction of the [third fortress](#) in the 17th and 18th centuries.



The city stretches on both sides of the [Limmat](#), which flows out of [Lake Zürich](#). The Alps can be seen from the city center, background to the lake

**Architecture** [edit]

Compared to other cities, there are few tall buildings in Zürich. The municipal building regulations (Article 9)<sup>[61]</sup> limit the construction of high-rise buildings to areas in the west and north of the city. In the industrial district, in Altstetten and Oerlikon, buildings up to 80 meters (260 ft) in height are allowed (high-rise area I). In the adjacent high-rise areas II and III the height is limited to 40 meters (130 ft). Around the year 2000, regulations became more flexible and high-rise buildings were again planned and built. The people's initiative "40 meters is enough," which would have reduced both the maximum height and the high-rise buildings area, was clearly rejected on 29 November 2009.<sup>[62]</sup> At this time in Zürich about a dozen high-rise buildings were under construction or in planning, including the [Prime Tower](#) as the tallest skyscraper in Switzerland.



The 88-metre<sup>[60]</sup> [Sunrise Tower](#) (2005) was the first approved high-rise building in twenty years



Panoramic view of [Münsterhof](#) square with some of the [Guild houses](#)

Figure A.11: Curated “Zurich” Wikipedia article (part 2).

## Economy [ edit ]


Zürich is a leading [financial center](#) and [global city](#).<sup>[9]</sup> The [Greater Zürich Area](#) is Switzerland's economic centre and home to a vast number of international companies. By far the most important sector in the economy of Zürich is the service industry, which employs nearly four-fifths of workers. Other important industries include light industry, machine and textile industries and tourism. Most Swiss banks have their headquarters in Zürich and there are numerous foreign banks in the Greater Zürich Area. Located in Zürich, the [Swiss Stock Exchange](#) was established in 1877 and is nowadays the fourth most prominent stock exchange in the world. In addition Zürich is the world's largest gold trading centre. Ten of the country's 50 largest companies have their head offices in Zürich, among them [ABB](#), [UBS](#),<sup>[64]</sup> [Credit Suisse](#), [Swiss Re](#) and [Zürich Financial Services](#).<sup>[65]</sup>

### The Swiss stock exchange [ edit ]

The Swiss stock exchange is called [SIX Swiss Exchange](#), formerly known as SWX. The [SIX Swiss Exchange](#) is the head group of several different worldwide operative financial systems: [Eurex](#), [Eurex US](#), [EXFEED](#), [STOXX](#), and [virt-x](#). The exchange turnover generated at the SWX was in 2007 of 1,780,499.5 million [CHF](#); the number of transactions arrived in the same period at 35,339,296 and the [Swiss Performance Index](#) (SPI) arrived at a total [market capitalization](#) of 1,359,976.2 million [CHF](#).<sup>[68][69]</sup>

The [SIX Swiss Exchange](#) goes back more than 150 years. In 1996, fully [electronic trading](#) replaced the traditional floor trading system at the [stock exchanges](#) of [Geneva](#) (founded in 1850), [Zürich](#) (1873) and [Basel](#) (1876).

Since 2008, the SIX Swiss Exchange has been part of the SIX Group, as SWX Group, SIS Group and Telekurs Group merged.



The Swiss stock Exchange on Selnastrasse

Figure A.12: Curated “Zurich” Wikipedia article (part 3).

## A.2 QUESTIONNAIRE

## Questionnaire:

- What is your age?

<20      20-30      31-40      41-50      51-60      61-70      >70

- What is your gender?

Female

Male

- What is your occupation?

---



---

- What is your domain/area of expertise?

---



---



---

- What type of designers do you consider yourself to be, e.g., graphic designers, motion designers, interaction designers, etc.?

---



---



---

- How many years have you been a practitioner?

---

- How many years have you been using Illustrator?

---

- Besides illustrator what tools do you usually use?

---



---

- Have you done (or are you currently doing) layout work, either for print or the web?

---










---












Figure A.13: Questionnaire given to the participants at the beginning of the use and design study.

### A.3 DESIGNS BY THE STUDY PARTICIPANTS















Design	ID	Data-driven from SENTENCE	+ DOCUMENT from SENTENCE + OUTSIDE CUMULATIVE QUANTITIES / PROPORTIONS RANKING	CHANGE / VARIATION / TRENDS TEMPORAL ABSOLUTE QUANTITIES LOCATION RELATIONS	QUANTITATIVE ABSTRACT DATA ILLUSTRATIONS ICONS OR SYMBOLS MAPS	TAGGED ENTITY TYPE OF ENTITY TAGGING	NUMBER OF ENTITIES NUMBER OF WORDS
<p>France is divided into 18 administrative regions (French: région, [ʁẽ.ʒjɔ̃]), 13 of which are in Metropolitan France and five of which are overseas regions.<sup>[1]</sup> The mainland regions five of which are overseas regions.<sup>[1]</sup> The mainland regions and Corsica are each further subdivided into departments, ranging in number from 2 to 13 per region for the metropolitan regions; the overseas regions technically consist of only one department each.</p> <p>metropolitan regions; the overseas regions technically consist of only one department each.</p> <p>number of regions was reduced from 27 to 18 through amalgamation.</p>	01					proximity	
	02					proximity	
	03					proximity	
	04					proximity	
<p>Europe (<span>/ˈjoʊərəp/</span> or <span>/ˈjɜːrəp/</span>)<sup>[4]</sup> is a continent that compr Eurasia. Europe is bordered by the Arctic Ocean to the north, the Atlantic Ocean to the west, and the Mediterranean Sea to the south. To the east and southeast, Europe is generally considered as separated from Asia by the watershed divides of the Ural and Caucasus Mountains, the Ural River, the Caspian and Black Seas, and the waterways of</p>	05					wikipedia link	1 1
	06					wikipedia link	1 2
	07					wikipedia link	1 1
<p>Europe is the world's second-smallest continent by surface area,</p>	08					proximity	
<p>2% of the Earth's surface</p>	09					proximity	
<p>Of Europe's approximately 50 countries,</p>	10					proximity	
<p>6.8% of its land area. Of Europe's approximately 50 countries, Russia is by far the largest by both area and population, taking up 40% of the continent (although the country has territory in both Europe and Asia), while Vatican City is the smallest. Europe is the third-most populous continent after Asia and Africa, with a population of 739–743 million or</p>	11					wikipedia link	1 1
	12					proximity	
<p>Europe, in particular ancient Greece, is the birthplace of Western culture.<sup>[7][8][9]</sup> The fall of the Western Roman Empire, during the migration period, marked the end of ancient history and the beginning of an era known as the "middle ages". The Renaissance humanism, exploration, art, and science led the "old continent", and eventually the rest of the world, to the modern era. From this period onwards, Europe played a predominant role in global affairs. Between the 16th and 20th centuries, European nations controlled</p>	13					proximity	
<p>European inspiration led to the formation of the European Union, a political entity that lies between a confederation and a federation.<sup>[10]</sup> The EU originated in Western Europe but has been expanding northeast since the fall of the Soviet Union in 1991. The currency of the European Union, the Euro, is the most commonly used among Europeans and the EU's Schengen Area abolishes border and immigration controls among most of its member states.</p>	14					proximity	
<p>is the largest city in Switzerland</p>	15					wikipedia link	1 2
<p>The municipality has approximately 400,028<sup>[4]</sup> inhabitants, the urban agglomeration 1.315 million,<sup>[5]</sup> and the Zurich metropolitan area 1.83 million.<sup>[6]</sup> Zürich is</p>	16					proximity	
<p>Zürich is a leading global city and among the world's largest financial centres despite a relatively low population.<sup>[9]</sup> The city is home to a large number of</p>	17					proximity	

















Design	ID	Data-driven from SENTENCE + DOCUMENT	from SENTENCE + OUTSIDE CUMULATIVE QUANTITIES / PROPORTIONS	RANKING	CHANGE / VARIATION / TRENDS	TEMPORAL	ABSOLUTE QUANTITIES	LOCATION	RELATIONS	QUANTITATIVE ABSTRACT DATA	ILLUSTRATIONS	ICONS OR SYMBOLS	MAPS	TAGGED ENTITY	TYPE OF ENTITY	TAGGING	NUMBER OF ENTITIES	NUMBER OF WORDS
 <p>Monocle's 2012 "Quality of Life Survey" ranked Zürich first on a list of the top 25 cities in the world "to make a base within".<sup>[10]</sup> According to several surveys from 2006</p>	18	0	0	0	0	0	0	0	0	0	0	0	0	wikipedia link			1	1
 <p>The City Council (<b>Stadtrat</b>) constitutes the executive government of the City of Zürich and operates as a collegiate authority. It is composed of nine councilors, each presiding over a department. Departmental tasks, coordination measures and implementation of laws decreed by the City Parliament are carried out by the City Council. The regular election of the City Council by any inhabitant valid to vote is held every</p>	19	0	0	0	0	0	0	0	0	0	0	0	0	proximity				
<p>As of 2015, the Zürich City Council is made up of four representatives of the SP (<b>Social Democratic Party</b>), one of whom is the mayor), two members of the FDP (<b>Free Democratic Party</b>) and one member each of <b>Green Party</b>, CVP (<b>Christian Democratic Party</b>) and AL (<b>Alternative Left Party</b>), giving the left parties a combined six out of nine seats.<sup>[91]</sup></p>	20	0	0	0	0	0	0	0	0	0	0	0	0	proximity				
<p>Zürich is situated at 408 m (1,339 ft) <b>above sea level</b>.</p>	21	0	0	0	0	0	0	0	0	0	0	0	0	proximity				
<p>Zürich is situated at 408 m (1,339 ft) <b>above sea level</b> on the lower (northern) end of <b>Lake Zürich (Zürichsee)</b> about 30 kilometers (19 mi) north of the <b>Alps</b>, nestling between the wooded hills on the west and east side. The Old Town stretches</p>	22	0	0	0	0	0	0	0	0	0	0	0	0	wikipedia link			1	1
<p>Compared to other cities, there are few tall buildings in Zürich. The municipal building regulations (Article 9)<sup>[61]</sup> limit the construction of high-rise buildings to areas in the west and north of the city. In the industrial district, in Altstetten and Oerlikon, buildings up to 80 meters (260 ft) in height are allowed (high-rise area I). In the adjacent high-rise areas II and III the height is limited to 40 meters (130 ft). Around the year 2000, regulations became more flexible and high-rise buildings were again planned and built. The people's initiative "40 meters is enough," which would have reduced both the maximum height and the high-rise buildings area, was clearly rejected on 29 November 2009.<sup>[62]</sup> At</p>	23	0	0	0	0	0	0	0	0	0	0	0	0	proximity				
<p>France is divided into <b>18 administrative regions</b> (French:  <b>région</b>, [ʁe.ʒjɔ̃]), <b>13</b> of which are in <b>Metropolitan France</b> and <b>five</b> of which are <b>overseas regions</b>.<sup>[1]</sup> The mainland regions</p>	24	0	0	0	0	0	0	0	0	0	0	0	0	font color			3	9
<p>representatives took place on 16 March 1986.<sup>[2]</sup> In 2016, the</p>  <p>number of regions was reduced from 27 to 18 through</p>	25	0	0	0	0	0	0	0	0	0	0	0	0	proximity				
<p>Europe is the world's <b>second-smallest</b> continent by surface area, covering about</p>  <p>10,180,000 square kilometres (3,930,000 sq mi) or <b>2%</b> of the <b>Earth's surface</b> and about</p>	26	0	0	0	0	0	0	0	0	0	0	0	0	background 2 color			3	
<p>territory in both Europe and Asia), while <b>Vatican City</b> is the smallest. Europe is the third-most populous continent after <b>Asia</b> and <b>Africa</b>, with a <b>population</b> of 739–743 million or</p>  <p>about <b>11%</b> of the <b>world's population</b>.<sup>[3]</sup> Europe has a climate heavily affected by warm</p>	27	0	0	0	0	0	0	0	0	0	0	0	0	background 2 color			3	
<p>and eventually the wider world. Demographic growth meant that, by 1900, Europe's share</p>  <p>of the <b>world's population</b> was <b>25%</b>.<sup>[10]</sup> Both <b>world wars</b> were largely focused upon</p>	28	0	0	0	0	0	0	0	0	0	0	0	0	background 2 color			3	

Design	ID	Data-driven from SENTENCE	from SENTENCE + DOCUMENT	from SENTENCE + OUTSIDE CUMULATIVE QUANTITIES / PROPORTIONS	RANKING	CHANGE / VARIATION / TRENDS	TEMPORAL	ABSOLUTE QUANTITIES	LOCATION	RELATIONS	QUANTITATIVE ABSTRACT DATA	ILLUSTRATIONS	ICONS OR SYMBOLS	MAPS	TAGGED ENTITY	TYPE OF ENTITY TAGGING	NUMBER OF ENTITIES	NUMBER OF WORDS
<p>As a continent, the economy of Europe is currently the largest on Earth and it is the richest region as measured by assets under management with over <a href="#">\$32.7 trillion</a></p>  <p>compared to <a href="#">North America's \$27.1 trillion</a> in 2008.<sup>[189]</sup> In 2009 Europe remained the</p>  <p>including <a href="#">carbon</a>, <a href="#">calcium</a> and <a href="#">phosphorus</a>.</p> <p>The average adult body</p>  <p>contains between 5 and 5½ litres of blood and</p> <p>approximately 10 litres of <a href="#">interstitial fluid</a>.</p> <p>The average adult body</p> <p>contains between 5 and 5½ litres of blood and</p>  <p>approximately 10 litres of <a href="#">interstitial fluid</a>.</p>  <p>can vary from a high 75% in a newborn infant to a lower 45% in an <a href="#">obese</a> person.</p> <p><a href="#">France</a> is divided into 18 administrative <a href="#">regions</a> (<a href="#">French: région, [ʁe.ʒjɔ̃]</a>), 13 of which are in <a href="#">Metropolitan France</a> and</p>  <p>five of which are <a href="#">overseas regions</a>.<sup>[1]</sup> The mainland regions</p> <p>representatives took place on 16 March 1986.<sup>[2]</sup> In 2016, the</p>  <p>number of regions was reduced from 27 to 18 through</p> <p>Europe is the world's <a href="#">second-smallest</a> continent by surface area, covering about 10,180,000 square kilometres (3,930,000 sq mi) or 2% of the Earth's surface and about</p>  <p>6.8% of its land area. Of Europe's approximately <a href="#">50 countries</a>, <a href="#">Russia</a> is by far the largest by both area and population, taking up <a href="#">40% of the continent</a> (although the country has territory in both Europe and Asia), while <a href="#">Vatican City is the smallest</a>. Europe is the third-territory in both Europe and Asia), while <a href="#">Vatican City is the smallest</a>. Europe is the third-</p>  <p>most populous continent after <a href="#">Asia and Africa</a>, with a <a href="#">population</a> of 739–743 million or about 11% of the <a href="#">world's population</a>.<sup>[5]</sup> Europe has a climate heavily affected by warm</p>  <p>The <a href="#">Industrial Revolution</a>,</p>  <p> began in the <a href="#">United Kingdom</a> around the end</p>	29			0	0	0	0	0	0	0	0	0	0	0	0	background color	2	6
	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	wikipedia link	3	3
	31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	proximity		
	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	proximity		
	33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	proximity		
	34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	font color	1	2
	35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	font color	1	3
	36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	underlined and font color	3	11
	37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	font color	1	3
	38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	font color	1	2
	39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	font color	1	2

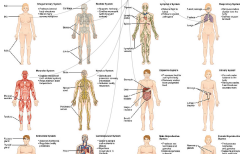
Design	ID	Data-driven from SENTENCE	from SENTENCE + DOCUMENT	from SENTENCE + OUTSIDE	CUMULATIVE QUANTITIES / PROPORTIONS	RANKING	CHANGE / VARIATION / TRENDS	TEMPORAL	ABSOLUTE QUANTITIES	LOCATION	RELATIONS	QUANTITATIVE ABSTRACT DATA	ILLUSTRATIONS	ICONS OR SYMBOLS	MAPS	TAGGED ENTITY	TYPE OF ENTITY	TAGGING	NUMBER OF ENTITIES	NUMBER OF WORDS
revolutions of 1989 and fall of the Berlin Wall.	40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	font color			1	2
European integration led to the formation of the European Union, a political entity that lies between a confederation and a	41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	font color			1	2
city, the estimated population of Paris is 37,000,000,000,000 as 37.2.trillion. This is the same as the population of the United States.	42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	font color			2	8
various hormones. The average height of an adult male human (in developed countries) is about 1.7-1.8 m (5'7" to 5'11") and the adult female is about 1.6-1.7 m (5'2" to 5'7"). Height is largely determined by genes and diet. Body	43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	font color			1	1
body by genes and environment. The average adult body contains between 5 and 5½ litres of blood and approximately 10 litres of interstitial fluid.	44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	font color			2	9
The adult human body contains approximately 60% water, and so makes up a significant proportion of the body, both in terms of weight and volume. Water content	45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	font color			1	1
The vast majority of cells in the human body are not human at all; rather they are of bacteria, archaea, and methanogens such as Methanobrevibacter smithii. The largest proportion of these form the gut flora. The whole	46	0	0	0	0	0	0	0	0	0	0	0	0	0	0	font color			1	4
From Wikipedia, the free encyclopedia	47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	proximity				
France is divided into 18 administrative regions (French: région, [ʁe.ʒjɔ̃]), 13 of which are in Metropolitan France and five of which are overseas regions.[1] The mainland regions	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	proximity				
France is divided into 18 administrative regions (French: région, [ʁe.ʒjɔ̃]), 13 of which are in Metropolitan France and five of which are overseas regions.[1] The mainland regions	49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	font color			1	14
The term region was officially created by the Law of Decentralisation (2 March 1982), which also gave regions their legal status. The first direct elections for regional	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	font color			1	13
representatives took place on 16 March 1986.[2] In 2016, the number of regions was reduced from 27 to 18 through amalgamation.	51	0	0	0	0	0	0	0	0	0	0	0	0	0	0	font color			1	14

	ID	Data-driven from SENTENCE + DOCUMENT	from SENTENCE + OUTSIDE CUMULATIVE QUANTITIES / PROPORTIONS	RANKING	CHANGE / VARIATION / TRENDS	TEMPORAL	ABSOLUTE QUANTITIES	LOCATION	RELATIONS	QUANTITATIVE ABSTRACT DATA	ILLUSTRATIONS	ICONS OR SYMBOLS	MAPS	TAGGED ENTITY	TYPE OF ENTITY TAGGING	NUMBER OF ENTITIES	NUMBER OF WORDS
<p><b>Design</b></p> <p>Between 1982 and 2015, there were 22 regions in</p>  <p>Metropolitan France. Before 2011, there were four overseas regions (French Guiana, Guadeloupe, Martinique, and</p> <p>Réunion); in 2011 Mayotte became the fifth overseas region.</p> <p>Between 1982 and 2015, there were 22 regions in</p> <p>Metropolitan France. Before 2011, there were four overseas regions (French Guiana, Guadeloupe, Martinique, and</p>  <p>Réunion); in 2011 Mayotte became the fifth overseas region.</p>	52		0	0	0	0	0	0	0	0	0	0	0	0	font color	4	5
<p>Metropolitan France. Before 2011, there were four overseas regions (French Guiana, Guadeloupe, Martinique, and</p>  <p>Réunion); in 2011 Mayotte became the fifth overseas region.</p>	53		0	0	0	0	0	0	0	0	0	0	0	0	font color	5	6
 <p>Europe is the world's <b>second-smallest</b> continent by surface area</p>	54		0	0	0		0	0	0	0	0	0	0	0	frame	1	2
 <p>2% of the Earth's surface</p>	55		0	0		0	0	0	0	0	0	0	0	0	frame	1	5
 <p>6.8% of its land area. Of t</p>	56		0	0		0	0	0	0	0	0	0	0	0	frame	1	5
 <p>Of Europe's approximately 50 countries, <b>Russia</b> is by far the largest</p> <p>Europe's approximately 50 countries, <b>Russia</b> is by far the largest</p>	57	0	0	0		0	0	0	0	0	0	0	0	0	wikipedia link	1	1
 <p>taking up <b>40% of the continent</b> (although the country has</p>	58		0	0		0	0	0	0	0	0	0	0	0	frame	1	4
<p>while <b>Vatican City</b> is the smallest.</p>	59	0	0	0		0	0	0	0	0	0	0	0	0	wikipedia link	1	2
<p>territory in both Europe and Asia), while <b>Vatican City</b> is the smallest. <b>Europe</b> is the third-</p>  <p>most populous continent after Asia and Africa, with a population of 739–743 million or</p>	60		0	0	0		0	0	0	0	0	0	0	0	frame	1	7
 <p>about 11% of the world's population.<sup>[3]</sup></p>	61		0	0	0		0	0	0	0	0	0	0	0	frame	1	6
<p>1900 2016</p>  <p>in 1914 and 1918. It t</p>	62		0	0	0	0		0	0	0	0	0	0	0	frame	1	3
 <p>fought between the <b>Entente Powers</b> (France, Belgium, Serbia, Portugal, <b>Russia</b>, the United Kingdom, and later Italy, Greece, Romania, and the United States) and the</p>	63		0	0		0	0	0	0	0	0	0	0	0	frame	1	2
 <p><b>Central Powers</b> (Austria-Hungary, Germany, Bulgaria, and the Ottoman Empire).</p>	64		0	0	0	0		0	0	0	0	0	0	0	frame	1	2
 <p>The War left more than <b>16 million</b> civilians and military dead.<sup>[14]</sup> Over 60 million European soldiers were mobilised from 1914 to 1918.<sup>[14]</sup></p>	65	0	0	0		0	0	0	0	0	0	0	0	0	frame	1	17













	ID	Data-driven from SENTENCE	from SENTENCE + DOCUMENT	from SENTENCE + OUTSIDE	CUMULATIVE QUANTITIES / PROPORTIONS RANKING	CHANGE / VARIATION / TRENDS	TEMPORAL ABSOLUTE QUANTITIES	LOCATION RELATIONS	QUANTITATIVE ABSTRACT DATA ILLUSTRATIONS	ICONS OR SYMBOLS	MAPS	TAGGED ENTITY TYPE OF ENTITY TAGGING	NUMBER OF ENTITIES	NUMBER OF WORDS
<p><b>Design</b></p> <p>As a continent, the economy of <u>Europe</u> is currently the largest on Earth and it is the</p>  <p>richest region as measured by assets under management with over <u>\$32.7 trillion</u></p> <p>compared to North America's \$27.1 trillion in 2008.<sup>[109]</sup> In 2009 Europe remained the</p> <p>As a continent, the economy of Europe is currently the largest on Earth and it is the</p>	66	0	0	0	0	0	0	0	0	0	0	frame	1	3
<p>richest region as measured by assets under management with over \$32.7 trillion</p>  <p>compared to <u>North America's \$27.1 trillion</u> in 2008.<sup>[109]</sup> In 2009 Europe remained the</p> <p>There is huge disparity between many European countries in terms of their income. The richest in terms of GDP per capita is <u>Monaco</u></p>	67	0	0	0	0	0	0	0	0	0	0	frame	1	4
 <p>with its US\$172,676 per capita (2009) and the poorest is Moldova with its GDP per capita of US\$1,331 (2010).<sup>[160]</sup> <u>Monaco</u> is the</p> <p>There is huge disparity between many European countries in terms of their income. The richest in terms of GDP per capita is Monaco</p>	68		0	0	0	0	0	0	0	0	0	frame	1	7
<p>with its US\$172,676 per capita (2009) and the poorest is <u>Moldova with its GDP per capita of US\$1,331 (2010)</u>.<sup>[160]</sup> <u>Monaco</u> is the</p> 	69		0	0	0	0	0	0	0	0	0	frame	1	9
<p>otherwise be. Therefore, the average temperature throughout the year of <u>Naples is 16 °C (60.8 °F)</u> while it is only 12 °C (53.6 °F) in</p> <p>otherwise be. Therefore, the average temperature throughout the year of Naples is 16 °C (60.8 °F), while it is only 12 °C (53.6 °F) in</p> 	70		0	0	0	0	0	0	0	0	0	frame	1	6
 <p><u>New York City</u> which is almost on the same latitude: Berlin, Germany; Calgary, Canada; and Irkutsk, in the Asian part of</p>	71		0	0	0	0	0	0	0	0	0	frame	1	3
 <p>inaccessible mountains with sufficient forest cover. Today, the <u>brown bear</u> lives primarily in the <u>Balkan peninsula</u>, Scandinavia, and <u>Russia</u> a small number also persist in other countries across Europe (Austria, Pyrenees etc.), out in these areas brown bear</p> <p>At maturity, the estimated average number of cells in the body</p>	72	0	0	0	0	0	0	0	0	0	0	frame	1	13
 <p>is given as <u>37.2 trillion</u>. This number is stated to be of partial</p> <p>all the <u>organs</u> of the body and cell types.<sup>[2]</sup> The <u>composition</u> of</p>  <p><u>the human body</u> is made up of a number of certain elements</p>	73	0	0	0	0	0	0	0	0	0	0	frame	1	2
 <p><u>the human body</u> is made up of a number of certain elements</p>	74	0	0	0	0	0	0	0	0	0	0	frame	1	3
 <p>including <u>carbon</u>, <u>calcium</u> and <u>phosphorus</u>.</p>	75	0	0	0	0	0	0	0	0	0	0	frame	1	1
 <p>including <u>carbon</u>, <u>calcium</u> and <u>phosphorus</u>.</p>	76	0	0	0	0	0	0	0	0	0	0	frame	1	1
 <p>including <u>carbon</u>, <u>calcium</u> and <u>phosphorus</u>.</p>	77	0	0	0	0	0	0	0	0	0	0	frame	1	1
<p>various <u>hormones</u>. The average height of an adult <u>male human</u> (in developed countries) is about 1.7–1.8 m (5'7" to 5'11") and the adult female is about 1.6–1.7 m (5'2" to 5'7"). Height is largely determined by <u>genes</u> and <u>diast. Body</u></p> 	78		0	0	0	0	0	0	0	0	0	frame	1	11

	ID	Data-driven from SENTENCE	DOCUMENT from SENTENCE + OUTSIDE	CUMULATIVE QUANTITIES / PROPORTIONS	RANKING	CHANGE / VARIATION / TRENDS	TEMPORAL	ABSOLUTE QUANTITIES	LOCATION	RELATIONS	QUANTITATIVE ABSTRACT DATA	ILLUSTRATIONS	ICONS OR SYMBOLS	MAPS	TAGGED ENTITY	TYPE OF ENTITY TAGGING	NUMBER OF ENTITIES	NUMBER OF WORDS
<p><b>Design</b></p> <p>Various hormones. The average height of an adult male human (in developed countries) is about 1.7–1.8 m (5'7" to 5'11") and the adult female is about 1.6–1.7 m (5'2" to 5'7"). Height is largely determined by genes and diet. Body</p> <p>Height is largely determined by genes and diet.</p> <p>Height is largely determined by genes and diet.</p> <p>The average adult body</p> <p>contains between 5 and 5% litres of blood and</p> <p>approximately 10 litres of interstitial fluid.</p> <p>The average adult body</p> <p>contains between 5 and 5% litres of blood and</p> <p>approximately 10 litres of interstitial fluid.</p> <p>The average adult body</p> <p>contains between 5 and 5% litres of blood and</p> <p>approximately 10 litres of interstitial fluid.</p> <p>The adult human body</p> <p>contains approximately 60%</p> <p>water, and so makes up a</p> <p>and volume. Water content</p> <p>bébé obèse</p> <p>can vary from a high 75% in a newborn infant to a lower 45% in an obese person.</p> <p>of bacteria.</p> <p>archaea.</p> <p>methanogens such as Methanobrevibacter smithii.</p> <p>The largest proportion of these form the gut flora</p>	79															frame	1	10
	80															frame	1	1
	80															frame	1	1
	82															frame	1	3
	83															frame	1	3
	84															frame	1	19
	85															frame	1	2
	86															frame	1	20
	87															frame	1	1
	88															frame	1	1
	89															frame	1	1
	90															frame	1	2

[illegible]

Design	ID	Data-driven from SENTENCE + DOCUMENT	from SENTENCE + OUTSIDE	CUMULATIVE QUANTITIES / PROPORTIONS	RANKING	CHANGE / VARIATION / TRENDS	TEMPORAL	ABSOLUTE QUANTITIES	LOCATION	RELATIONS	QUANTITATIVE ABSTRACT DATA	ILLUSTRATIONS	ICONS OR SYMBOLS	MAPS	TAGGED ENTITY	TYPE OF ENTITY TAGGING	NUMBER OF ENTITIES	NUMBER OF WORDS
and volume. Water content can vary from a high 75% in a newborn infant to a <b>lower 45% in an obese person.</b>	105															proximity		
	106															proximity		
<b>France</b> is divided into 18 administrative <b>regions</b>	107															proximity		
<b>France</b> is divided into 18 administrative <b>regions</b> (French: <i>région</i> , [ʁe.ʒjɔ̃]), 13 of which are in <b>Metropolitan France</b> and five of which are <b>overseas regions</b> . <sup>[1]</sup> The mainland regions	108															proximity		
<b>France</b> is divided into 18 administrative <b>regions</b> (French: <i>région</i> , [ʁe.ʒjɔ̃]), 13 of which are in <b>Metropolitan France</b> and five of which are <b>overseas regions</b> . <sup>[1]</sup> The mainland regions	109															proximity		
<b>France</b> is divided into 18 administrative <b>regions</b> (French: <i>région</i> , [ʁe.ʒjɔ̃]), 13 of which are in <b>Metropolitan France</b> and five of which are <b>overseas regions</b> . <sup>[1]</sup> The mainland regions	110															wikipedia link	1 2	
<b>France</b> is divided into 18 administrative <b>regions</b> (French: <i>région</i> , [ʁe.ʒjɔ̃]), 13 of which are in <b>Metropolitan France</b> and five of which are <b>overseas regions</b> . <sup>[1]</sup> The mainland regions	111															wikipedia link	1 2	
The term <i>region</i> was officially created by the <b>Law of Decentralisation</b> (2 March 1982), which also gave regions	112															proximity		
<b>Europe</b> ( <sup>[4]</sup> /jʊərəp/ or /jɜːrəp <sup>[44]</sup> ) is a <b>continent</b> that comprises the <b>westernmost</b> part of <b>Eurasia</b> . Europe is bordered by the <b>Arctic Ocean</b> to the north, the <b>Atlantic Ocean</b> to the	113															underlined	1 1	
<b>Eurasia</b> . Europe is bordered by the <b>Arctic Ocean</b> to the north, the <b>Atlantic Ocean</b> to the west, and the <b>Mediterranean Sea</b> to the south. To the east and southeast, Europe is	114															underlined	1 2	
<b>Eurasia</b> . Europe is bordered by the <b>Arctic Ocean</b> to the north, the <b>Atlantic Ocean</b> to the west, and the <b>Mediterranean Sea</b> to the south. To the east and southeast, Europe is	115															underlined	1 2	
<b>Eurasia</b> . Europe is bordered by the <b>Arctic Ocean</b> to the north, the <b>Atlantic Ocean</b> to the west, and the <b>Mediterranean Sea</b> to the south. To the east and southeast, Europe is	116															underlined	1 2	
<b>antiquity</b> —are <b>arbitrary</b> , as the primarily <b>physiographic</b> term "continent" also incorporates <b>cultural and political</b> elements.	117															underlined	1 1	



Design	ID	Data-driven from SENTENCE	from SENTENCE + DOCUMENT	from SENTENCE + OUTSIDE CUMULATIVE QUANTITIES / PROPORTIONS	RANKING	CHANGE / VARIATION / TRENDS	TEMPORAL	ABSOLUTE QUANTITIES	LOCATION	RELATIONS	QUANTITATIVE ABSTRACT DATA	ILLUSTRATIONS	ICONS OR SYMBOLS	MAPS	TAGGED ENTITY	TYPE OF ENTITY TAGGING	NUMBER OF ENTITIES	NUMBER OF WORDS
 or 2% of the Earth's surface	118															proximity		
 Europe, in particular <u>ancient Greece</u> ,	119															underlined	2	3
 Europe, in particular <u>ancient Greece</u> , is the birthplace of <u>Western culture</u> .	120															proximity		
 European nations <u>controlled</u> <u>abdomen</u> ), <u>arms</u> and <u>hands</u> , <u>legs</u> and <u>feet</u> . Every part of the body is composed of various types of <u>cells</u> , the fundamental unit of life. <sup>[1]</sup>	121															underlined	1	1
 all the <u>organs</u> of the body and cell types. <sup>[2]</sup> The <u>composition of</u> <u>the human body</u> is made up of a number of certain <u>elements</u> including <u>carbon</u> , <u>calcium</u> and <u>phosphorus</u> .	122															wikipedia link	1	1
 all the <u>organs</u> of the body and cell types. <sup>[2]</sup> The <u>composition of</u> <u>the human body</u> is made up of a number of certain <u>elements</u> including <u>carbon</u> , <u>calcium</u> and <u>phosphorus</u> .	123															wikipedia link	1	1
 all the <u>organs</u> of the body and cell types. <sup>[2]</sup> The <u>composition of</u> <u>the human body</u> is made up of a number of certain <u>elements</u> including <u>carbon</u> , <u>calcium</u> and <u>phosphorus</u> .	124															wikipedia link	1	1
 all the <u>organs</u> of the body and cell types. <sup>[2]</sup> The <u>composition of</u> <u>the human body</u> is made up of a number of certain <u>elements</u> including <u>carbon</u> , <u>calcium</u> and <u>phosphorus</u> .	125															wikipedia link	1	1
 all the <u>organs</u> of the body and cell types. <sup>[2]</sup> The <u>composition of</u> <u>the human body</u> is made up of a number of certain <u>elements</u> including <u>carbon</u> , <u>calcium</u> and <u>phosphorus</u> .	126															wikipedia link	1	1
 <u>homeostasis</u> .	127															wikipedia link	1	1
 <u>Skeletal structure</u> frames the overall shape of the body and does not alter much over a lifetime. representatives took place on 16 March <u>1986</u> . <sup>[2]</sup> In 2016, the number of regions was reduced from 27 to 18 through <u>amalgamation</u> .	128															proximity		
 <b>Regions from 1982 to 2015</b> [edit] Between 1982 and 2015, there were 22 regions in Metropolitan France. Before 2011, there were four <u>overseas</u>	129															proximity		
	130															proximity		

	ID	Data-driven from SENTENCE	from SENTENCE + DOCUMENT	from SENTENCE + OUTSIDE	CUMULATIVE QUANTITIES / PROPORTIONS	RANKING	CHANGE / VARIATION / TRENDS	TEMPORAL	ABSOLUTE QUANTITIES	LOCATION	RELATIONS	QUANTITATIVE ABSTRACT DATA	ILLUSTRATIONS	ICONS OR SYMBOLS	MAPS	TAGGED ENTITY	TYPE OF ENTITY TAGGING	NUMBER OF ENTITIES	NUMBER OF WORDS
<p><b>Design</b></p> 	131																wikipedia link	1	2
Europe is the world's <b>second-smallest</b> continent by surface area,	132																wikipedia link	1	1
6.8% of its land area. Of Europe's approximately 50 countries, <b>Russia</b> is by far the largest by both area and population, taking up 40% of the continent (although the country has	133																wikipedia link	1	2
while <b>Vatican City</b> is the smallest.	134																proximity		
territory in both Europe and Asia), while <b>Vatican City</b> is the smallest. Europe is the third-most populous continent after <b>Asia</b> and <b>Africa</b> , with a <b>population</b> of 739–743 million or	135																wikipedia link	1	1
currency of the European Union, the <b>Euro</b> , is the most commonly used among Europeans	136																proximity		
<b>German flowers (Roses, ranunculus, carnations, lilies, and the C</b>	137																proximity		
The War left more than 16 million civilians and military dead. <sup>[147]</sup>	138																proximity		
As a continent, the economy of Europe is currently the largest on Earth and it is the richest region as measured by assets under management with over \$32.7 trillion compared to North America's \$27.1 trillion in 2008. <sup>[189]</sup> In 2009 Europe remained the	139																wikipedia link	3	5
As a continent, the economy of Europe is currently the largest on Earth and it is the richest region as measured by assets under management with over \$32.7 trillion compared to North America's \$27.1 trillion in 2008. <sup>[189]</sup> In 2009 Europe remained the	140																proximity		
<b>Skeletal structure</b> frames the overall shape of the body and does not alter much over a lifetime. General <b>body shape</b> (and <b>female body shape</b> ) is influenced by the distribution of <b>muscle</b> and <b>fat tissue</b> and is also affected by various <b>hormones</b> . The average <b>height</b> of an adult male human (in developed countries) is about 1.7–1.8 m (5'7" to 5'11") and the adult female is about 1.6–1.7 m (5'2" to 5'7"). Height is largely determined by <b>genes</b> and diet. <b>Body</b>	141																proximity		
various <b>hormones</b> . The average <b>height</b> of an adult male human (in developed countries) is about 1.7–1.8 m (5'7" to 5'11") and the adult female is about 1.6–1.7 m (5'2" to 5'7"). Height is largely determined by <b>genes</b> and diet. <b>Body</b>	142																wikipedia link	1	1
<b>body composition</b> are influenced by factors such as <b>genetics</b>	143																wikipedia link	1	1
<b>body composition</b> are influenced by factors such as <b>genetics, diet,</b>	144																wikipedia link	1	1
<b>body composition</b> are influenced by factors such as <b>genetics, diet, and exercise.</b>	145																wikipedia link	1	2
The human body has several <b>body cavities</b> the largest of which is the <b>abdominopelvic cavity</b> .	146																wikipedia link	1	2
four pairs of vital air-cavities in the <b>cranial bones</b> .	147																proximity		
The average adult body contains between 5 and 5½ litres of blood and  approximately 10 litres of <b>interstitial fluid</b> .	148																proximity		
The term <b>region</b> was officially created by the <b>Law of</b>	149																proximity		
<b>Decentralisation</b> (2 March 1982), which also gave regions their legal status. The first direct elections for regional	150																proximity		
their legal status. The first direct elections for regional																			
representatives took place on 16 March 1986. <sup>[2]</sup> In 2016, the																			
representatives took place on 16 March 1986. <sup>[2]</sup> In 2016, the																			
number of regions was reduced from 27 to 18 through <b>amalgamation</b> .																			

[illegible]

Design	ID	Data-driven from SENTENCE + DOCUMENT from SENTENCE + OUTSIDE CUMULATIVE QUANTITIES / PROPORTIONS	RANKING	CHANGE / VARIATION / TRENDS	TEMPORAL	ABSOLUTE QUANTITIES	LOCATION	RELATIONS	QUANTITATIVE ABSTRACT DATA	ILLUSTRATIONS	ICONS OR SYMBOLS	MAPS	TAGGED ENTITY	TYPE OF ENTITY TAGGING	NUMBER OF ENTITIES	NUMBER OF WORDS
Russia, lie on around the same latitude; January temperatures in Berlin average around 8 °C (15 °F) higher than those in Calgary, and	165													proximity		
Russia, lie on around the same latitude; January temperatures in Berlin average around 8 °C (15 °F) higher than those in Calgary, and	166													proximity		
they are almost 22 °C (40 °F) higher than average temperatures in Inuitsk. <sup>[179]</sup> Similarly, northern parts of Scotland have a temperate marine climate. The yearly average temperature in city of Inverness is 9.05 degrees Celsius (48.3 degrees Fahrenheit). However,	167													proximity		
Chuvohi, Manitoba, Canada, is on roughly the same latitude and has an average temperature of -6.5 degrees Celsius (20.3 degrees Fahrenheit), giving it a nearly subarctic climate.	168													proximity		
hunting caused these animals to withdraw further and further. By the Middle Ages the bears' habitats were limited to more or less inaccessible mountains with sufficient forest cover. Today, the brown bear lives primarily in the Balkan peninsula, Scandinavia, and	169													proximity		
inaccessible mountains with sufficient forest cover. Today, the brown bear lives primarily in the Balkan peninsula, Scandinavia, and Russia; a small number also persist in other countries across Europe (Austria, Pyrenees etc.), but in these areas brown bear populations are fragmented and marginalised because of the destruction of their habitat. In addition, polar bears may be found on	170													proximity		
At maturity, the estimated average number of cells in the body	171													font color	3	3
is given as 37.2 trillion. This number is stated to be of partial	172													proximity		
all the organs of the body and cell types. <sup>[2]</sup> The composition of the human body is made up of a number of certain elements	173													proximity		
including carbon, calcium and phosphorus.	174													proximity		
various hormones. The average height of an adult male human (in developed countries) is about 1.7–1.8 m (57" to 5'11") and the adult female is about 1.6–1.7 m (52" to 57"). Height is largely determined by genes and diet. Body	175													proximity		
various hormones. The average height of an adult male human (in developed countries) is about 1.7–1.8 m (57" to 5'11") and the adult female is about 1.6–1.7 m (52" to 57"). Height is largely determined by genes and diet. Body	176													proximity		
The vast majority of cells in the human body are not human at all; rather they are of bacteria, archaea, and	177													proximity		
methanogens such as <i>Methanobrevibacter smithii</i> . The largest proportion of these form the gut flora. The whole	178													proximity		
The vast majority of cells in the human body are not human at all; rather they are of bacteria, archaea, and	179													wikipedia link	1	1
methanogens such as <i>Methanobrevibacter smithii</i> . The largest proportion of these form the gut flora. The whole	180													wikipedia link	1	1
The vast majority of cells in the human body are not human at all; rather they are of bacteria, archaea, and	181													wikipedia link	1	1
methanogens such as <i>Methanobrevibacter smithii</i> . The largest proportion of these form the gut flora. The whole	182													wikipedia link	1	1
Regions of France	183													underlined	4	58
France is divided into 18 administrative regions (French: <i>region</i> , (see 35)), 13 of which are in Metropolitan France and five of which are overseas regions. <sup>[1]</sup> The mainland regions and Corsica are each further subdivided into departments, ranging in number from 2 to 13 per region for the metropolitan regions; the overseas regions technically consist of only one department each.	184													frame	1	2
Two World Wars and an economic depression dominated the first half of the 20th century. World War I was fought between 1914 and 1918. It started when Archduke Franz Ferdinand of Austria was assassinated by the Yugoslav nationalist <sup>[149]</sup> Gavrilo Princip. <sup>[149]</sup> Most European nations were drawn into the war, which was fought between the Entente Powers (France, Belgium, Serbia, Portugal, Russia, the United Kingdom, and later Italy, Greece, Romania, and the United States) and the Central Powers (Austria-Hungary, Germany, Bulgaria, and the Ottoman Empire).	185													underlined	4	72
Two World Wars and an economic depression dominated the first half of the 20th century. World War I was fought between 1914 and 1918. It started when Archduke Franz Ferdinand of Austria was assassinated by the Yugoslav nationalist <sup>[149]</sup> Gavrilo Princip. <sup>[149]</sup> Most European nations were drawn into the war, which was fought between the Entente Powers (France, Belgium, Serbia, Portugal, Russia, the United Kingdom, and later Italy, Greece, Romania, and the United States) and the Central Powers (Austria-Hungary, Germany, Bulgaria, and the Ottoman Empire).	186													proximity		
Two World Wars and an economic depression dominated the first half of the 20th century. World War I was fought between 1914 and 1918. It started when Archduke	187													proximity		

Design	ID	Data-driven from SENTENCE	from SENTENCE + DOCUMENT	from SENTENCE + OUTSIDE	CUMULATIVE QUANTITIES / PROPORTIONS	RANKING	CHANGE / VARIATION / TRENDS	TEMPORAL	ABSOLUTE QUANTITIES	LOCATION	RELATIONS	QUANTITATIVE ABSTRACT DATA	ILLUSTRATIONS	ICONS OR SYMBOLS	MAPS	TAGGED ENTITY	TYPE OF ENTITY TAGGING	NUMBER OF ENTITIES	NUMBER OF WORDS
World War I was fought between 1914 and 1918.	183	o	o	o		o	o	o		o	o	o	o	o	o	o	proximity		
century. World War I was fought between 1914 and 1918. It started when Archduke Franz Ferdinand of Austria was assassinated by the Yugoslav nationalist <sup>[145]</sup> Gavrilo Princip. <sup>[146]</sup> Most European nations were drawn into the war, which was century. World War I was fought between 1914 and 1918. It started when Archduke Franz Ferdinand of Austria was assassinated by the Yugoslav nationalist <sup>[145]</sup> Gavrilo Princip. <sup>[146]</sup> Most European nations were drawn into the war, which was century. World War I was fought between 1914 and 1918. It started when Archduke Franz Ferdinand of Austria was assassinated by the Yugoslav nationalist <sup>[145]</sup> Gavrilo Princip. <sup>[146]</sup> Most European nations were drawn into the war, which was	184	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	proximity		
fought between the Entente Powers (France, Belgium, Serbia, Portugal, Russia, the United Kingdom, and later Italy, Greece, Romania, and the United States) and the Central Powers (Austria-Hungary, Germany, Bulgaria, and the Ottoman Empire). Gavrilo Princip. <sup>[146]</sup> Most European nations were drawn into the war, which was fought between the Entente Powers (France, Belgium, Serbia, Portugal, Russia, the United Kingdom, and later Italy, Greece, Romania, and the United States) and the Central Powers (Austria-Hungary, Germany, Bulgaria, and the Ottoman Empire). Gavrilo Princip. <sup>[146]</sup> Most European nations were drawn into the war, which was fought between the Entente Powers (France, Belgium, Serbia, Portugal, Russia, the United Kingdom, and later Italy, Greece, Romania, and the United States) and the Central Powers (Austria-Hungary, Germany, Bulgaria, and the Ottoman Empire).	185	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	proximity		
The average adult body contains between 5 and 5½ litres of blood and approximately 10 litres of interstitial fluid.	186	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	proximity		
The average adult body contains between 5 and 5½ litres of blood and approximately 10 litres of interstitial fluid.	187	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	proximity		
The average adult body contains between 5 and 5½ litres of blood and approximately 10 litres of interstitial fluid.	188	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	proximity		
The average adult body contains between 5 and 5½ litres of blood and approximately 10 litres of interstitial fluid.	189	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	proximity		
The adult human body contains approximately 60% water, and so makes up a significant proportion of the body, both in terms of weight and volume. Water content can vary from a high 75% in a newborn infant to a lower 45% in an obese person.	190	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	proximity		
The adult human body contains approximately 60% water, and so makes up a significant proportion of the body, both in terms of weight and volume. Water content can vary from a high 75% in a newborn infant to a lower 45% in an obese person.	191	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	proximity		
The adult human body contains approximately 60% water, and so makes up a significant proportion of the body, both in terms of weight and volume. Water content can vary from a high 75% in a newborn infant to a lower 45% in an obese person.	192	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	proximity		
The adult human body contains approximately 60% water, and so makes up a significant proportion of the body, both in terms of weight and volume. Water content can vary from a high 75% in a newborn infant to a lower 45% in an obese person.	193	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	proximity		
The adult human body contains approximately 60% water, and so makes up a significant proportion of the body, both in terms of weight and volume. Water content can vary from a high 75% in a newborn infant to a lower 45% in an obese person.	194	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	overlay		
The adult human body contains approximately 60% water, and so makes up a significant proportion of the body, both in terms of weight and volume. Water content can vary from a high 75% in a newborn infant to a lower 45% in an obese person.	195	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	proximity		
The adult human body contains approximately 60% water, and so makes up a significant proportion of the body, both in terms of weight and volume. Water content can vary from a high 75% in a newborn infant to a lower 45% in an obese person.	196	o	o	o	o	o	o	o	o	o	o	o	o	o	o	o	proximity		

Design	ID	Data-driven from SENTENCE + DOCUMENT from SENTENCE + OUTSIDE	CUMULATIVE QUANTITIES / PROPORTIONS	RANKING	CHANGE / VARIATION / TRENDS	TEMPORAL LOCATION	ABSOLUTE QUANTITIES	RELATIONS	QUANTITATIVE ABSTRACT DATA	ILLUSTRATIONS	ICONS OR SYMBOLS	MAPS	TAGGED ENTITY	TYPE OF ENTITY TAGGING	NUMBER OF ENTITIES	NUMBER OF WORDS
The adult human body contains approximately 60% water, and so makes up a significant proportion of the body, both in terms of weight and volume. Water content	197	0 0 0	0 0 0 0	0 0 0 0	0	0 0 0	0 0 0	0	0	0	0	0	0	proximity		
The vast majority of cells in the human body are not human at all; rather they are of bacteria, archaea, and methanogens such as <i>Methanobrevibacter smithii</i> . The largest proportion of these form the gut flora. The whole	198	0 0 0	0 0 0 0	0 0 0 0	0	0 0 0	0 0 0	0	0	0	0	0	0	proximity		
The vast majority of cells in the human body are not human at all; rather they are of bacteria, archaea, and methanogens such as <i>Methanobrevibacter smithii</i> . The largest proportion of these form the gut flora. The whole	199	0 0 0	0 0 0 0	0 0 0 0	0	0 0 0	0 0 0	0	0	0	0	0	0	proximity		
population of microbiota include microorganisms of the skin and other body parts and this altogether is termed as	200	0 0	0 0 0 0	0 0 0 0	0	0 0 0	0	0	0	0	0	0	0	proximity		

## PLACING WORD-SCALE VISUALIZATIONS

---

### B.1 SPARKLIFICATOR SOURCE CODE

---

```

1
2 (function ( $ ) {
3
4     $.widget( 'aviz.sparklificator', {
5
6         /**
7         * Default options
8         * @param {string} position - 3 available named choices: 'top', 'right', and '
          baseline' OR just {top: int, left: int}
9         * @param {int} width - the width of the word-scale visualisation
10        * @param {int} height - the height of the word-scale visualisation
11        * @param {array | function} data - the data to be visualized by the word-scale
          visualisation
12        * @param {boolean} paddingWidth - adds padding (horizontal) or allows overlap
          between word-scale visualisation and word- scale visualisation, and word-scale
          visualisation and text
13        * @param {boolean} paddingHeight - adds padding (vertical) or allows overlap between
          word-scale visualisation and word-scale visualisation, and word-scale
          visualisation and text
14        * @param {string} stackingOrder - 2 available named choices: 'front' and 'back',
          adds the possibility ('back') to render the word-scale visualization behind the
          entity (parts covered by entity) or in front ('front')
15        * @param {boolean} hoverInteraction - adds hover interaction to the entity
16        * @param {string | function} renderer - visualizes the given data for an element
17        */
18        options: {
19            position: 'top',
20            width: 100,
21            height: 50,
22            data: [],
23            paddingWidth: true,
24            paddingHeight: false,
25            stackingOrder: 'front',
26            hoverInteraction: false,
27            renderer: function() {},
28        },
29
30
31        /**
32        * Function that gets called when the plugin is called the first time, initialization
          of the plugin
33        */
34        _create: function() {
35
36            var o = this.options;
37
38            // element is the text element (entity) associated with the words-scale
          visualization
39            var element = this.element;
40            element.css('position', 'relative');
41            element.addClass('entity');
42
43            element.wrap("<span class='sparklificated'></span>");

```

```

44
45     // sparklifiedSPAN is the span which contains the word-scale visualization
      span and the entity span
46     var sparklifiedSPAN = element.parent();
47     sparklifiedSPAN.css('position', 'relative');
48     sparklifiedSPAN.css('white-space', 'nowrap');
49     sparklifiedSPAN.css('display', 'inline-block');
50
51     // sparklineSPAN is the the span around the word-scale visualization
52     var sparklineSPAN = $('<span class="sparkline"></span>');
53     sparklifiedSPAN.append(sparklineSPAN);
54
55     //set the other basic properties of the container and sparkline
56     sparklineSPAN.css('position', 'absolute');
57
58     return this;
59 },
60
61
62 /**
63  * To change one default option
64  * Example of how this function can be called: $('<div>').sparklifier('option', '
      renderer', 'classicSparkline')
65  * @param {string} option - the option name (position, width, height, data,
      paddingWidth, paddingHeight, renderer)
66  * @param {string | function} value - the value for the option
67  */
68 _setOption: function(option, value) {
69     this._super(option, value);
70     this.refresh(option);
71 },
72
73
74 /**
75  * To change multiples default options
76  * Example of how this function can be called: $('<div>').sparklifier('option', {
      renderer: 'classicSparkline', position: 'right'})
77  * @param {object} options - contains multiple options to be changed at once
78  */
79 _setOptions: function(options) {
80     this._super(options);
81 },
82
83
84 /**
85  * Called when a default option is changed
86  * @param {string} option - the option name
87  */
88 refresh: function(option) {
89     var o = this.options;
90     var element = this.element;
91
92     // sparklifiedSPAN is the span which contains the word-scale visualization
      span and the entity span
93     var sparklifiedSPAN = element.parent();
94
95     // sparklineSPAN is the the span around the word-scale visualization
96     var sparklineSPAN = element.next();
97
98     var newWidth = $.isFunction(o.width) ? o.width.call(this) : o.width;
99     var newHeight = $.isFunction(o.height) ? o.height.call(this) : o.height;
100
101     switch(option) {
102         case 'position':
103         case 'paddingWidth':
104         case 'paddingHeight':

```



```

105         this._setNewPosition(sparklineSPAN, sparklifiedSPAN, newWidth, newHeight
106             );
107         break;
108
109     case 'data':
110     case 'renderer':
111     case 'hoverInteraction':
112     case 'height':
113     case 'width':
114
115         sparklineSPAN.empty();
116
117         var data = $.isFunction(o.data) ? o.data.call(this, element[0]) : o.data;
118
119         var hoverInteraction = $.isFunction(o.hoverInteraction) ? o.
120             hoverInteraction.call(this, element[0]) : o.hoverInteraction;
121
122         o.renderer.call(this, sparklineSPAN, newWidth, newHeight, hoverInteraction,
123             data);
124
125         this._setNewPosition(sparklineSPAN, sparklifiedSPAN, newWidth, newHeight
126             );
127         break;
128     }
129 },
130
131 /**
132  * Recomputes and sets the position of the word-scale visualization using the current
133  * settings.
134  * IMPORTANT: divs/spans that are sparklified need to be visible (NOT display:
135  * none), as some of the calculations are based on the divs being connected in the
136  * DOM.
137  * @param {object} sparklineSPAN - the container in which the word-scale
138  * visualization is drawn.
139  * @param {object} sparklifiedSPAN - the outside span holding both the word-scale
140  * visualization and entity spans.
141  * @param {int} width - the width of the word-scale visualization from the current
142  * setting
143  * @param {int} height - the height of the word-scale visualization from the current
144  * setting
145  */
146 _setNewPosition: function(sparklineSPAN, sparklifiedSPAN, width, height) {
147
148     var o = this.options;
149     var elementSPAN = this.element;
150
151     // set the top position of the sparklineSPAN to 0, to guarantee a consistent
152     // calculation of the difference between sparklineSPAN and elementSPAN position
153     sparklineSPAN.css('top', 0 + 'px');
154
155     // height of the text element bounding box
156     var fontSize = elementSPAN[0].offsetHeight;
157
158     // width of the text element bounding box
159     var lengthOfEntity = elementSPAN[0].offsetWidth;
160
161     // line-height of the text element
162     var currentLineHeight = parseFloat($(elementSPAN[0]).css('line-height'));
163
164     // because the position can be a function
165     var newPosition = $.isFunction(o.position) ? o.position.call(this, elementSPAN
166         [0]) : o.position;

```

```

157         // set the width and height of the sparklineSPAN to guarantee consistent width
158         and height
159         sparklineSPAN.width(width);
160         sparklineSPAN.height(height);
161
162         var xOffset = 0;
163         var yOffset = 0; // the y position of the top edge
164         var marginTop = 0;
165         var marginRight = 0;
166         var marginLeft = 0;
167         var marginBottom = 0;
168
169         // yOffset for top position of the elementSPAN and the sparklineSPAN
170         var topPosSpanToSpark = elementSPAN[0].getBoundingClientRect().top;
171         var topPosSpark = sparklineSPAN[0].getBoundingClientRect().top;
172
173         // difference between top positions of the 2 containers
174         var diffTopPos = topPosSpanToSpark - topPosSpark;
175
176         // difference between right positions of the 2 containers
177         var diffTopPosRight = (topPosSpanToSpark + fontSize) - (topPosSpark + height);
178
179
180         // calculate length of nbsp (space between entity and sparkline)
181         var tmpNBSPNode = $('<span display=none></span>').append('&nbsp;');
182         sparklifiedSPAN.append(tmpNBSPNode);
183         var lengthOfNBSP = tmpNBSPNode[0].offsetWidth;
184         tmpNBSPNode.remove();
185
186         if (newPosition === 'top'){
187             yOffset = diffTopPos - height;
188         } else if (newPosition === 'right') {
189             xOffset = +lengthOfEntity + lengthOfNBSP;
190             yOffset = diffTopPosRight;
191         } else if (newPosition === 'baseline'){
192             yOffset = diffTopPosRight;
193         } else if (newPosition === 'left'){
194             xOffset = -width - +lengthOfNBSP;
195             yOffset = diffTopPosRight;
196         } else if (newPosition === 'below'){
197             yOffset = diffTopPos + fontSize;
198         } else if ($.isPlainObject(o.position) || $.isFunction(o.position)) {
199             xOffset = newPosition.left;
200             yOffset = newPosition.top;
201         } else {
202             console.log('PROBLEM: your option for POSITION (' + newPosition + ') is not a
203                 function or a valid default option');
204         }
205
206
207         marginTop = o.paddingHeight ? Math.max(-(yOffset-diffTopPos) - (currentLineHeight
208             - fontSize), 0) : 0;
209         marginRight = o.paddingWidth ? Math.max(xOffset + width - lengthOfEntity, 0) : 0;
210         marginLeft = o.paddingWidth ? Math.max(-xOffset, 0) : 0;
211         marginBottom = o.paddingHeight ? Math.max(yOffset - (currentLineHeight - fontSize
212             ), 0) : 0;
213
214         sparklineSPAN.css('top', yOffset + 'px');
215         sparklineSPAN.css('left', xOffset + 'px');
216         sparklifiedSPAN.css('margin-top', marginTop);
217         sparklifiedSPAN.css('margin-right', marginRight);
218         sparklifiedSPAN.css('margin-left', marginLeft);
219         sparklifiedSPAN.css('margin-bottom', marginBottom);

```

```

219
220     var currentIndex = 2;
221     this.element.css('z-index', currentIndex);
222     if (o.stackingOrder === 'front') {
223         var upOne = currentIndex + 1;
224         sparklineSPAN.css('z-index', upOne);
225     } else if (o.stackingOrder === 'back') {
226         var downOne = currentIndex - 1;
227         sparklineSPAN.css('z-index', downOne);
228     }
229 },
230
231 /**
232  * Removes all the DOM objects added by the sparklificator
233  */
234 destroy: function() {
235     var element = this.element;
236
237     element.next().next().remove();
238     element.next().remove();
239     element.unwrap();
240
241     this._destroy();
242 },
243
244 });
245
246 }( jQuery ));

```

---

Source Code B.1: File “jquery.sparklificator.js” from the jQuery Sparklificator library.

---

```

1  /**
2  * Builds a classic sparkline (word-scale visualization) without or with interaction.
3  * this.element and this.options is available in the renderer if needed
4  * @param {string} sparkSpan - the container where the word-scale visualization is placed
5  * @param {int} width - width of the word-scale visualization
6  * @param {int} height - height of the word-scale visualization
7  * @param {boolean} interaction - hover interaction for this word-scale visualization or
   not
8  * @param {array} data - word-scale visualization's array of data
9  */
10 function classicSparkline(sparkSpan, width, height, interaction, data) {
11
12     var circleRadius = 3;
13     var o = this.options;
14
15     var margin = {top: 0, right: 0, bottom: 0, left: 0},
16         widthVis = width - margin.left - margin.right,
17         heightVis = height - margin.top - margin.bottom;
18
19     var sparkContainer = d3.select(sparkSpan.get(0));
20     sparkContainer.append('svg');
21
22     var x = d3.scale.linear()
23         .domain([0, data.length-1])
24         .range([0, widthVis-circleRadius]);
25
26     var y = d3.scale.linear()
27         .domain([0, d3.max(data, function(d) { return d })])
28         .range([heightVis-circleRadius, circleRadius]);
29

```

```

30   var xAxis = d3.svg.axis()
31     .scale(x)
32     .orient('bottom');
33
34   var yAxis = d3.svg.axis()
35     .scale(y)
36     .orient('left');
37
38   var line = d3.svg.line()
39     .x(function(d, i) { return x(i); })
40     .y(function(d) { return y(d); });
41
42   var chart = sparkContainer.select('svg')
43     .style('position', 'absolute')
44     .attr('width', widthVis + 'px')
45     .attr('height', heightVis + 'px');
46
47   // select sparklifiedSPAN, as the entity might be longer than the word-scale
48   // visualization
49   var entity = $(sparkSpan).closest($('.sparklified'));
50   if (interaction) {
51     entity.on('mouseover', fade(1))
52       .on('mouseout', fade(0.1));
53   }
54
55   var gChart = chart.append('g')
56     .attr('transform', 'translate(' + margin.left + ',' + margin.top + ')');
57
58   gChart.append('g')
59     .attr('class', 'x axis')
60     .attr('transform', 'translate(0,' + heightVis + ')');
61
62   gChart.append('g')
63     .attr('class', 'y axis');
64
65   gChart.append('path')
66     .datum(data)
67     .attr('class', 'sparkline')
68     .attr('d', line)
69     .style('fill', 'none')
70     .style('stroke', 'grey')
71     .style('stroke-width', '2px');
72
73   gChart.append('circle')
74     .style('fill', 'red')
75     .attr('r', circleRadius)
76     .attr('cx', x(data.length - 1))
77     .attr('cy', y(data[data.length - 1]));
78
79   if (!interaction) {
80     gChart.style('opacity', 1.0);
81   } else {
82     gChart.style('opacity', 0.1);
83   }
84
85   // Returns an event handler for fading in the sparkline graph.
86   function fade(opacity) {
87     return function(g, i) {
88       chart.select('g')
89         .transition()
90         .style('opacity', opacity);
91
92       var textOpacity = 0.25;
93       if (opacity == 0.1) { textOpacity = 1; }
94
95       d3.select($(sparkSpan).siblings('.entity')[0]).style('opacity', textOpacity)

```

```

95     };
96   }
97 }
98
99
100 /**
101 * Builds a bar chart (word-scale visualization)
102 * this.element and this.options is available in the renderer if needed
103 * @param {string} sparkSpan - the container where the word-scale visualization will be put
104   in
105 * @param {int} width - width of the word-scale visualization
106 * @param {int} height - height of the word-scale visualization
107 * @param {boolean} interaction - hover interaction for this word-scale visualization or
108   not
109 * @param {array} data - word-scale visualization's array of data
110 */
111 function barChart(sparkSpan, width, height, interaction, data) {
112
113   var o = this.options;
114
115   var margin = {top: 0, right: 0, bottom: 0, left: 0},
116     widthVis = width - margin.left - margin.right,
117     heightVis = height - margin.top - margin.bottom;
118
119   var barWidth = 5;
120
121   // clipping data array
122   var newData = data;
123   var newLength = Math.floor(width/barWidth);
124   if (newLength < data.length) {
125     newData = data.slice(0,newLength);
126   }
127
128   var sparkContainer = d3.select(sparkSpan.get(0));
129   sparkContainer.append('svg');
130
131   var y = d3.scale.linear()
132     .domain([0, d3.max(newData, function(d) { return d })])
133     .range([0, heightVis]);
134
135   var chart = sparkContainer.select('svg')
136     .style('position', 'absolute')
137     .attr('width', widthVis)
138     .attr('height', heightVis)
139     .attr('class', 'barChart');
140
141   // select sparklificatedSPAN, as the entity might be longer than the word-scale
142   // visualization
143   var entity = $(sparkSpan).closest($('.sparklificated'));
144   if (interaction) {
145     entity.on('mouseover', fade(1))
146       .on('mouseout', fade(0.1));
147   }
148
149   var bar = chart.selectAll('g.bar')
150     .data(newData);
151
152   var gBar = bar.enter().append('g')
153     .attr('class', 'bar')
154     .attr('transform', function(d, i) { return 'translate(' + i * barWidth + ',' + (heightVis
155       - y(d)) + ')'; });
156
157   gBar.append('rect')
158     .attr('width', barWidth - 2)
159     .attr('height', function(d) { return y(d); });

```

```

157     if (!interaction) {
158         gBar.select('rect').style('opacity', 1.0);
159     } else {
160         gBar.select('rect').style('opacity', 0.1);
161     }
162
163     bar.attr('transform', function(d, i) { return 'translate(' + i * barWidth + ',' + (heightVis
164         - y(d)) + ')'; });
165     bar.select('rect').attr('height', function(d) { return y(d); });
166     bar.exit().remove();
167
168     // Returns an event handler for fading in the sparkline graph.
169     function fade(opacity) {
170         return function(g, i) {
171             chart.selectAll('rect')
172                 .transition()
173                 .style('opacity', opacity);
174
175             var textOpacity = 0.25;
176             if (opacity == 0.1) { textOpacity = 1; }
177
178             d3.select($(sparkSpan).siblings('.entity')[0]).style('opacity', textOpacity)
179         };
180     }
181 }
182
183 /**
184  * Builds a pie chart (word-scale visualization)
185  * this.element and this.options is alvailable in the renderer if needed
186  * @param {string} sparkSpan - the container where the word-scale visualization will be put
187  *    in
188  * @param {int} width - width of the word-scale visualization
189  * @param {int} height - height of the word-scale visualization
190  * @param {boolean} interaction - hover interaction for this word-scale visualization or
191  *    not
192  * @param {array} data - word-scale visualization's array of data
193  */
194 function pieChart(sparkSpan, width, height, interaction, data) {
195     var o = this.options;
196
197     var margin = {top: 0, right: 0, bottom: 0, left: 0},
198         widthVis = width - margin.left - margin.right,
199         heightVis = height - margin.top - margin.bottom;
200
201     var radius = Math.min(widthVis, heightVis) / 2;
202
203     var color = d3.scale.ordinal()
204         .range(["#a6cee3", "#1f78b4", "#b2df8a", "#33a02c", "#fb9a99", "#e31a1c"]);
205
206     // clipping data array
207     var newData = data;
208     if (data.length > 6) {
209         newData = data.slice(0,6);
210     }
211
212     var arc = d3.svg.arc()
213         .outerRadius(radius - 2)
214         .innerRadius(0);
215
216     var pie = d3.layout.pie()
217         .sort(null)
218         .value(function(d) { return d; });
219

```

```

220     var sparkContainer = d3.select(sparkSpan.get(0));
221     sparkContainer.append('svg');
222
223     var chart = sparkContainer.select('svg')
224       .style('position', 'absolute')
225       .attr('width', widthVis)
226       .attr('height', heightVis);
227
228     // select sparklifiedSPAN, as the entity might be longer than the word-scale
229     // visualization
230     var entity = $(sparkSpan).closest($('.sparklified'));
231     //var entity = $(sparkSpan).siblings($('.entity'));
232     if (interaction) {
233       entity.on('mouseover', fade(1))
234         .on('mouseout', fade(0.1));
235     }
236
237     var gChart = chart.append('g')
238       .attr('transform', 'translate(' + widthVis / 2 + ',' + heightVis / 2 + ')');
239
240     var g = gChart.selectAll('arc')
241       .data(pie(newData))
242       .enter().append('g')
243       .attr('class', 'arc');
244
245     g.append('path')
246       .attr('d', arc)
247       // .style('stroke', 'black')
248       // .style('stroke-width', '2px');
249       .style("fill", function(d, i) { return color(i); });
250
251     if (!interaction) {
252       chart.selectAll('path').style('opacity', 1.0);
253     } else {
254       chart.selectAll('path').style('opacity', 0.1);
255     }
256
257     // Returns an event handler for fading in the sparkline graph.
258     function fade(opacity) {
259       return function(g, i) {
260         chart.selectAll('path')
261           .transition()
262           .style('opacity', opacity);
263
264         var textOpacity = 0.25;
265         if (opacity == 0.1) { textOpacity = 1; }
266
267         d3.select($(sparkSpan).siblings('.entity')[0]).style('opacity', textOpacity);
268       };
269     }

```

---

Source Code B.2: File “renderers.js” from the jQuery Sparklificator library.





## TOWARDS UNDERSTANDING THE EFFECTS OF WORD-SCALE VISUALIZATIONS ON THE READER

---

### C.1 QUESTIONNAIRE

#### Word-scale Visualization Study Questionnaire

End of study questionnaire

\* Required

1. **1. Participant ID \***

.....

2. **2. What is your age?**

*Mark only one oval.*

- ☐ < 20  
☐ 20 - 30  
☐ 31 - 40  
☐ 41 - 50  
☐ 51 - 60  
☐ 61 - 70  
☐ > 70

3. **3. What is your gender?**

*Check all that apply.*

- ☐ Female  
☐ Male

4. **4. What is your native language? \***

.....

Figure C.1: Questionnaire given to the participants at the end of the study (part 1).

5. **5. What is your English fluency level according to the "European Framework of Reference for Languages" \***

(more information on the levels can be found here:

[http://en.wikipedia.org/wiki/Common\\_European\\_Framework\\_of\\_Reference\\_for\\_Languages#Common\\_reference\\_levels](http://en.wikipedia.org/wiki/Common_European_Framework_of_Reference_for_Languages#Common_reference_levels))

Mark only one oval.

- ☐ C2 (Proficient user, mastery)
- ☐ C1 (Proficient user, advanced)
- ☐ B2 (Independent user, upper intermediate)
- ☐ B1 (Independent user, intermediate)
- ☐ A2 (Basic user, elementary)
- ☐ A1 (Basic user, beginner)

6. **6. Do you have any particular problems reading or memorizing words or numbers (e.g. dyslexia)? \***

.....

7. **7. What was your reading/memorizing strategy for sentences without a word-scale visualization? \***

.....

.....

.....

.....

.....

8. **8. What was your reading/memorizing strategy for sentences with a non-interactive word-scale visualization? \***

(please explain if different for top or right word-scale visualization)

.....

.....

.....

.....

.....

9. **9. Did you actively memorize the trend of word-scale visualization? \***

Mark only one oval.

- ☐ Yes
- ☐ No
- ☐ Sometimes

Figure C.2: Questionnaire given to the participants at the end of the study (part 2).

10. **10. What was your reading/memorizing strategy for sentences with an interactive word-scale visualization? \***

.....

.....

.....

.....

.....

11. **11. Did the word-scale visualization on top (inter-line space) of the entity disturb your reading/memorizing? \***

*Mark only one oval.*

	1	2	3	4	5	
Not disturbing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very disturbing

12. **12. Did the word-scale visualization on the right of the entity disturb your reading/memorizing? \***

*Mark only one oval.*

	1	2	3	4	5	
Not disturbing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very disturbing

13. **13. Did the interactive word-scale visualization disturb your reading/memorizing? \***

*Mark only one oval.*

	1	2	3	4	5	
Not disturbing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Very disturbing

14. **14. If the interactive word-scale visualization disturbed you, could you please elaborate on why?**

.....

.....

.....

.....

.....

Figure C.3: Questionnaire given to the participants at the end of the study (part 3).

## 15. 15. Can you rank (1., 2., 3., 4.) the following according to preference? \*

where 1 = the condition of the word-scale visualization you liked the best, 2 = the one you liked second-best, and 4 = the one you liked the least.

Mark only one oval per row.

	1.	2.	3.	4.
word-scale visualization "top" position	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
word-scale visualization "right" position	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
word-scale visualization interactive version	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
sentence without word-scale visualization	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

## 16. 16. What is the reason behind your ranking in question 15. (previous question)? \*

.....

.....

.....

.....

.....

## 17. 17. Did you find any ambiguous cases? \*

Mark only one oval.

☐ Yes

☐ No

## 18. 18. What did "ambiguous" mean for you during the study? \*

(e.g. missing information, conflicting information, other)

.....

.....

.....

.....

.....

## 19. 19. What strategy did you employ when you encountered an ambiguous case? \*

.....

.....

.....

.....

.....

Figure C.4: Questionnaire given to the participants at the end of the study (part 4).

20. **20. If you encounter an ambiguous case in a newspaper which information would you trust more? \***

*Check all that apply.*

- ☐ the information in the text  
☐ the information in the word-scale visualization  
☐ not sure

21. **21. Please explain your answer from 20. (previous question)? \***

.....  
.....  
.....  
.....  
.....

22. **22. If you have any additional comments please let us know!**

.....  
.....  
.....  
.....  
.....

Figure C.5: Questionnaire given to the participants at the end of the study (part 5).



## BIBLIOGRAPHY

---

- [1] Alfie Abdul-Rahman, Julie Lein, Katharine Coles, Eamonn Maguire, Miriah Meyer, Martin Wynne, Chris R Johnson, Anne Trefethen, and Min Chen. "Rule-based Visual Mappings—with a Case Study on Poetry Visualization." In: *Computer Graphics Forum* 32.3 (2013), pp. 381–390. DOI: 10.1111/cgf.12125 (cited on p. 61).
- [2] Benjamin Bach, Nathalie Henry-Riche, Tim Dwyer, Tara Madhayastha, Jean-Daniel Fekete, and Thomas Grabowski. "Small MultiPiles: Piling Time to Explore Temporal Patterns in Dynamic Networks." In: *Computer Graphics Forum* 34.3 (2015), pp. 31–40. DOI: 10.1111/cgf.12615 (cited on pp. 92, 107).
- [3] Fabian Beck, Fabrice Hollerich, Stephan Diehl, and Daniel Weiskopf. "Visual monitoring of numeric variables embedded in source code." In: *First IEEE Working Conference on Software Visualization (VISSOFT)*. IEEE. 2013, pp. 1–4. DOI: 10.1109/VISSOFT.2013.6650545 (cited on p. 91).
- [4] Fabian Beck, Tanja Blascheck, Thomas Ertl, and Daniel Weiskopf. "Word-Sized Eye Tracking Visualizations." In: *Mathematics and Visualization Series*. Published online: April 2016, printed version to appear. Springer Verlag, 2016 (cited on p. 27).
- [5] Jeff Beddow. "Shape coding of multidimensional data on a microcomputer display." In: *Proceedings of the 1st Conference on Visualization*. IEEE Computer Society Press. 1990, pp. 238–246 (cited on p. 28).
- [6] Enrico Bertini, Maurizio Rigamonti, and Denis Lalanne. "Extended excentric labeling." In: *Computer Graphics Forum* 28.3 (2009), pp. 927–934. DOI: 10.1111/j.1467-8659.2009.01456.x (cited on p. 62).
- [7] Anastasia Bezerianos, Fanny Chevalier, Pierre Dragicevic, Niklas Elmqvist, and Jean-Daniel Fekete. "GraphDice: A System for Exploring Multivariate Social Networks." In: *Computer Graphics Forum* 29.3 (2010), pp. 863–872. DOI: 10.1111/j.1467-8659.2009.01687.x (cited on p. 92).
- [8] Steven Bird, Edward Loper, and Ewan Klein. *Natural Language Processing with Python*. O'Reilly Media Inc., 2009 (cited on p. 77).
- [9] Rita Borgo, Johannes Kehrner, David HS Chung, Eamonn Maguire, Robert S Laramee, Helwig Hauser, Matthew Ward, and Min Chen. "Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications." In: *Eurographics 2013—State*

- of the Art Reports*. The Eurographics Association. 2012, pp. 39–63 (cited on p. 27).
- [10] Michelle A Borkin, Zoya Bylinskii, Nam Wook Kim, Constance May Bainbridge, Chelsea S Yeh, Daniel Borkin, Hanspeter Pfister, and Aude Oliva. “Beyond Memorability: Visualization Recognition and Recall.” In: *IEEE Transactions on Visualization and Computer Graphics* 22.1 (2016), pp. 519–528 (cited on pp. 50, 52).
  - [11] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. “D<sup>3</sup> data-driven documents.” In: *IEEE Transactions on Visualization and Computer Graphics* 17.12 (2011), pp. 2301–2309. DOI: 10.1109/TVCG.2011.185 (cited on p. 82).
  - [12] Nadia Boukhelifa, Emmanouil Giannidakis, Evanthia Dimara, Wesley Willett, and Jean-Daniel Fekete. “Supporting Historical Research Through User-Centered Visual Analytics.” In: *Proceedings of the Sixth Workshop on Visual Analytics*. 2015. DOI: 10.2312/eurova.20151095 (cited on p. 104).
  - [13] Niels Olof Bouvin, Polle T Zellweger, Kaj Grønbaek, and Jock D Mackinlay. “Fluid annotations through open hypermedia: using and extending emerging web standards.” In: *Proceedings of the 11th International Conference on World Wide Web*. ACM. 2002, pp. 160–171. DOI: 10.1145/511446.511468 (cited on p. 129).
  - [14] Jeremy Boy, Louis Eveillard, Françoise Detienne, and Jean-Daniel Fekete. “Suggested Interactivity: Seeking Perceived Affordances for Information Visualization.” In: *IEEE Transactions on Visualization and Computer Graphics* 22.1 (Oct. 2016), p. 10. DOI: 10.1109/TVCG.2015.2467201 (cited on pp. 92, 94).
  - [15] Ulrik Brandes and Bobo Nick. “Asymmetric relations in longitudinal social networks.” In: *IEEE Transactions on Visualization and Computer Graphics* 17.12 (2011), pp. 2283–2290. DOI: 10.1109/TVCG.2011.169 (cited on p. 26).
  - [16] Ulrik Brandes, Bobo Nick, Brigitte Rockstroh, and Astrid Steffen. “Gestaltlines.” In: *Computer Graphics Forum* 32.3 (2013), pp. 171–180. DOI: 10.1111/cgf.12104 (cited on pp. 7, 26, 60, 112).
  - [17] Susan E Brennan, Klaus Mueller, Greg Zelinsky, IV Ramakrishnan, David S Warren, and Arie Kaufman. “Toward a multi-analyst, collaborative framework for visual analytics.” In: *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST)*. 2006, pp. 129–136. DOI: 10.1109/VAST.2006.261439 (cited on p. 135).
  - [18] Willard Cope Brinton. *Graphic methods for presenting facts*. Engineering Magazine Company, 1914 (cited on p. 42).



- [19] Aaron Bycoffe. *The Endorsement Primary*. <http://projects.fivethirtyeight.com/2016-endorsement-primary/>. Last read: July 2016 (cited on p. 30).
- [20] A Canty and B Ripley. *Bootstrap Functions*. 2014 (cited on p. 118).
- [21] Stuart K Card, Jock D Mackinlay, and Ben Shneiderman. *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999 (cited on pp. 6, 128).
- [22] Sheelagh Carpendale. "A framework for elastic presentation space." PhD thesis. Simon Fraser University, 1999 (cited on p. 128).
- [23] Sheelagh Carpendale. "Evaluating information visualizations." In: *Information Visualization*. Springer, 2008, pp. 19–45. DOI: 10.1007/978-3-540-70956-5\_2 (cited on p. 125).
- [24] Paul Chandler and John Sweller. "The split-attention effect as a factor in the design of instruction." In: *British Journal of Educational Psychology* 62.2 (1992), pp. 233–246. DOI: 10.1111/j.2044-8279.1992.tb01017.x (cited on p. 24).
- [25] Bay-Wei Chang, Jock Mackinlay, and Polle T Zellweger. "Fluidly revealing information in fluid documents." In: *Proceedings of Smart Graphics*. AAAI, 2000, pp. 178–181 (cited on pp. 72, 91, 95, 96).
- [26] Bay-Wei Chang, Jock D Mackinlay, Polle T Zellweger, and Takeo Igarashi. "A negotiation architecture for fluid documents." In: *Proceedings of the Conference on User Interface Software and Technology (UIST)*. ACM. 1998, pp. 123–132. DOI: 10.1145/288392.288585 (cited on pp. 62, 85, 112).
- [27] Herman Chernoff. "The use of faces to represent points in k-dimensional space graphically." In: *Journal of the American Statistical Association* 68.342 (1973), pp. 361–368. DOI: 10.1080/01621459.1973.10482434 (cited on p. 28).
- [28] Fanny Chevalier, Romain Vuillemot, and Guia Gali. "Using concrete scales: a practical framework for effective visual depiction of complex measures." In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (2013), pp. 2426–2435 (cited on p. 42).
- [29] Ed Huai-hsin Chi and John T Riedl. "An operator interaction framework for visualization systems." In: *Proceedings of IEEE Symposium on Information Visualization*. IEEE. 1998, pp. 63–70. DOI: 10.1109/INFVIS.1998.729560 (cited on p. 128).
- [30] Mei C Chuah and Steven F Roth. "Visualizing common ground." In: *Proceedings of the International Conference on Information Visualization (IV)*. IEEE. 2003, pp. 365–372. DOI: 10.1109/IV.2003.1218012 (cited on p. 135).

- [31] David HS Chung, Robert S Laramée, Johannes Kehrér, and Helwig Hauser. "Glyph-Based Multi-field Visualization." In: *Scientific Visualization*. Springer, 2014, pp. 129–137. DOI: 10.1007/978-1-4471-6497-5\_13 (cited on p. 27).
- [32] William S Cleveland, Marylyn E McGill, and Robert McGill. "The shape parameter of a two-variable graph." In: *Journal of the American Statistical Association* 83.402 (1988), pp. 289–300. DOI: 10.1080/01621459.1988.10478598 (cited on p. 63).
- [33] *Codex Leicester*. <http://www.discoveringdavinci.com>. Last read: July 2016 (cited on p. 4).
- [34] Christopher Collins, Sheelagh Carpendale, and Gerald Penn. "Docuburst: Visualizing document content using language structure." In: *Computer Graphics Forum*. Vol. 28. 3. Wiley Online Library. 2009, pp. 1039–1046. DOI: 10.1111/j.1467-8659.2009.01439.x (cited on pp. 19, 20).
- [35] Christopher Collins, Fernanda B Viegas, and Martin Wattenberg. "Parallel tag clouds to explore and analyze faceted text corpora." In: *IEEE Symposium on Visual Analytics Science and Technology (VAST)*. IEEE. 2009, pp. 91–98. DOI: 10.1109/VAST.2009.5333443 (cited on p. 22).
- [36] Geoff Cumming. "The new statistics: Why and how." In: *Psychological science* 25.1 (2014), pp. 7–29. DOI: 10.1177/0956797613504966 (cited on p. 118).
- [37] Jason Davies. *Word Tree implementation*. <https://www.jasondavies.com/wordtree/>. Last read: July 2016 (cited on p. 17).
- [38] *De Revolutionibus Orbium Caelestium*. <http://www.nypl.org/blog/2014/09/22/400-years-banned-books>. Last read: July 2016 (cited on p. 4).
- [39] Ferdinand De Saussure, Wade Baskin, and Perry Meisel. *Course in general linguistics*. Columbia University Press, 2011 (cited on p. 54).
- [40] Dan Diaper and P Waelend. "World Wide Web working whilst ignoring graphics: good news for web page designers." In: *Interacting with Computers* 13.2 (2000), pp. 163–181. DOI: 10.1016/S0953-5438(00)00036-9 (cited on p. 112).
- [41] The New York Public Library Digital Collections. (*cartographic*) *Tour book, ([1909-1911])*. The New York Public Library, Astor, Lenox, and Tilden Foundation. <http://digitalcollections.nypl.org/items/3b553d50-40e5-0131-a300-58d385a7b928>. Last read: July 2016 (cited on pp. 29, 30).
- [42] Pierre Dragicevic, Fanny Chevalier, and Stéphane Huot. "Running an HCI experiment in multiple parallel universes." In: *CHI Extended Abstracts*. ACM. 2014, pp. 607–618. DOI: 10.1145/2559206.2578881 (cited on p. 118).

- [43] Eli Dresner and Susan C Herring. "Functions of the nonverbal in CMC: Emoticons and illocutionary force." In: *Communication Theory* 20.3 (2010), pp. 249–268. DOI: 10.1111/j.1468-2885.2010.01362.x (cited on p. 29).
- [44] Mary C Dyson. "How physical text layout affects reading from screen." In: *Behaviour & Information Technology* 23.6 (2004), pp. 377–393. DOI: 10.1080/01449290410001715714 (cited on p. 78).
- [45] SC Eick, Joseph L Steffen, and Eric E Sumner. "Seesoft – a tool for visualizing line oriented software statistics." In: *IEEE Transactions on Software Engineering* 18.11 (1992), pp. 957–968. DOI: 10.1109/32.177365 (cited on p. 18).
- [46] Pete Faraday and Alistair Sutcliffe. "Authoring animated Web pages using "contact points"." In: *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. ACM. 1999, pp. 458–465. DOI: 10.1145/302979.303131 (cited on p. 23).
- [47] Jean-Daniel Fekete and Nicole Dufournaud. "Compus: visualization and analysis of structured documents for understanding social life in the 16th century." In: *Proceedings of the fifth ACM conference on Digital libraries*. ACM. 2000, pp. 47–55. DOI: 10.1145/336597.336632 (cited on pp. 21, 23).
- [48] Jean-Daniel Fekete and Catherine Plaisant. "Excentric labeling: Dynamic neighborhood labeling for data visualization." In: *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. ACM. 1999, pp. 512–519. DOI: 10.1145/302979.303148 (cited on p. 62).
- [49] Stephen Few. *Time on the Horizon*. [http://www.perceptualedge.com/articles/visual\\_business\\_intelligence/time\\_on\\_the\\_horizon.pdf](http://www.perceptualedge.com/articles/visual_business_intelligence/time_on_the_horizon.pdf). Last read: July 2016 (cited on p. 28).
- [50] Michael Friendly. "A brief history of data visualization." In: *Handbook of data visualization*. Springer, 2008, pp. 15–56. DOI: 10.1007/978-3-540-33037-0\_2 (cited on p. 2).
- [51] Adrian Frutiger. *Signs and symbols: their design and meaning*. Van Nostrand Reinhold Company, 1989 (cited on p. 1).
- [52] Johannes Fuchs, Fabian Fischer, Florian Mansmann, Enrico Bertini, and Petra Isenberg. "Evaluation of alternative glyph designs for time series data in a small multiple setting." In: *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. ACM. 2013, pp. 3237–3246. DOI: 10.1145/2470654.2466443 (cited on p. 63).
- [53] Johannes Fuchs, Petra Isenberg, Anastasia Bezerianos, and Daniel Keim. "A Systematic Review of Experimental Studies on Data Glyphs." In: *IEEE Transactions on Visualization and Computer Graphics* (2016). Published online: 31 March 2016, printed ver-

- sion to appear. ISSN: 1077-2626. DOI: 10.1109/TVCG.2016.2549018 (cited on pp. 27, 28).
- [54] Galileo Galileo. *The Galileo Manuscript*. Letter. Published online at: <http://www.lib.umich.edu/special-collections-library/galileo-manuscript>. 1609 (cited on p. 29).
  - [55] Qihong Gan, Min Zhu, Mingzhao Li, Ting Liang, Yu Cao, and Baoyao Zhou. "Document visualization: an overview of current research." In: *Wiley Interdisciplinary Reviews: Computational Statistics* 6.1 (2014), pp. 19–36. DOI: 10.1002/wics.1285 (cited on p. 15).
  - [56] Thomas Geymayer, Markus Steinberger, Alexander Lex, Marc Streit, and Dieter Schmalstieg. "Show me the invisible: visualizing hidden content." In: *Proceedings of the 32nd Annual Conference on Human Factors in Computing Systems*. ACM. 2014, pp. 3705–3714. DOI: 10.1145/2556288.2557032 (cited on p. 91).
  - [57] Michael Gleicher, Danielle Albers, Rick Walker, Ilir Jusufi, Charles D. Hansen, and Jonathan C. Roberts. "Visual Comparison for Information Visualization." In: *Information Visualization* 10.4 (Oct. 2011), pp. 289–309. DOI: 10.1177/1473871611416549 (cited on p. 103).
  - [58] Pascal Goffin, Wesley Willett, and Petra Isenberg. "Sharing Information from Personal Digital Notes using Word-Scale Visualizations." In: *Personal Visualization: Exploring Data in Everyday Life*. 2015 (cited on pp. v, 133).
  - [59] Pascal Goffin, Wesley Willett, Jean-Daniel Fekete, and Petra Isenberg. "Interaction Techniques for Visual Exploration using Embedded Word-Scale Visualizations." In: () (cited on pp. v, 89).
  - [60] Pascal Goffin, Wesley Willett, Jean-Daniel Fekete, and Petra Isenberg. "Exploring the Placement and Design of Word-Scale Visualizations." In: *IEEE Transactions on Visualization and Computer Graphics* 20.12 (2014), pp. 2291–2300 (cited on pp. v, 25, 26, 33, 53, 59, 140).
  - [61] Pascal Goffin, Wesley Willett, Jean-Daniel Fekete, and Petra Isenberg. "Design considerations for Enhancing Word-Scale Visualizations with Interaction." In: *Posters of the IEEE Conference on Information Visualization (InfoVis)*. Los Alamitos, CA, USA: IEEE, 2015 (cited on pp. v, 89).
  - [62] Pascal Goffin, Wesley Willett, Anastasia Bezerianos, and Petra Isenberg. "Exploring the effect of word-scale visualizations on reading behavior." In: *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*. ACM. 2015, pp. 1827–1832 (cited on pp. v, 111).

- [63] Pascal Goffin, Jeremy Boy, Wesley Willett, and Petra Isenberg. "An Exploratory Study of Word-Scale Graphics in Data-Rich Text Documents." In: *IEEE Transactions on Visualization and Computer Graphics* PP.99 (2016), pp. 1–1 (cited on pp. v, 33, 127).
- [64] Pascal Goffin, Wesley Willett, Jean-Daniel Fekete, and Petra Isenberg. *Sparklifier*. <http://inria.github.io/sparklifier/>. Last read: July 2016 (cited on p. 82).
- [65] Carsten Görg, Zhicheng Liu, Jaeyeon Kihm, Jaegul Choo, Hae-sun Park, and John Stasko. "Combining computational analyses and interactive visualization for document exploration and sensemaking in jigsaw." In: *IEEE Transactions on Visualization and Computer Graphics* 19.10 (2013), pp. 1646–1663. DOI: 10.1109/TVCG.2012.324 (cited on pp. 5, 6).
- [66] Lars Grammel, Chris Bennett, Melanie Tory, and Margaret-Anne Storey. "A survey of visualization construction user interfaces." In: *EuroVis-Short Papers* (2013), pp. 19–23 (cited on p. 128).
- [67] Brian Greenhill, Michael Ward, and Audrey Sacks. "The Separation Plot: A New Visual Method for Evaluating the Fit of Binary Models." In: *American Journal of Political Science* 55.4 (2011), pp. 991–1002. DOI: 10.1111/j.1540-5907.2011.00525.x (cited on pp. 27, 60).
- [68] Kyle Hall, Charles Perin, Peter G Kusalik, Carland Gutwin, and Sheelagh Carpendale. "Formalizing Emphasis in Information Visualization." In: *Eurographics 2016–State of the Art Reports*. The Eurographics Association. 2016. DOI: 10.1111/cgf.12936 (cited on p. 50).
- [69] Susan Havre, Elizabeth Hetzler, Paul Whitney, and Lucy Nowell. "ThemeRiver: Visualizing Thematic Changes in Large Document Collections." In: *IEEE Transactions on Visualization and Computer Graphics* 8.1 (2002), pp. 9–20. DOI: 10.1109/2945.981848 (cited on pp. 22, 24).
- [70] Marti A Hearst. "Multi-paragraph segmentation of expository text." In: *Proceedings of the 32nd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics. 1994, pp. 9–16. DOI: 10.3115/981732.981734 (cited on p. 21).
- [71] Marti A Hearst. "TileBars: visualization of term distribution information in full text information access." In: *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. ACM Press/Addison-Wesley Publishing Co. 1995, pp. 59–66. DOI: 10.1145/223904.223912 (cited on pp. 21, 22).
- [72] Marti A Hearst. "User interfaces and visualization." In: *Modern Information Retrieval* (1999), pp. 257–323 (cited on p. 17).

- [73] Jeffrey Heer and Maneesh Agrawala. "Multi-scale banking to 45 degrees." In: *IEEE Transactions on Visualization and Computer Graphics* 12.5 (2006), pp. 701–708. DOI: 10.1109/TVCG.2006.163 (cited on pp. 63, 73, 129).
- [74] Jeffrey Heer, Nicholas Kong, and Maneesh Agrawala. "Sizing the horizon: The effects of chart size and layering on the graphical perception of time series visualizations." In: *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. ACM. 2009, pp. 1303–1312. DOI: 10.1145/1518701.1518897 (cited on pp. 63, 82).
- [75] Jeffrey Heer and Ben Shneiderman. "Interactive dynamics for visual analysis." In: *Communications of the ACM* 55.4 (Apr. 2012), pp. 45–54. DOI: 10.1145/2133806.2133821 (cited on pp. 2, 92, 93, 99, 105).
- [76] Tim Holmes. *EYE BALLS!* <http://acuity-intelligence.com/blog/eye-balls>. Last read: July 2016 (cited on p. 125).
- [77] Samuel Huron, Sheelagh Carpendale, Alice Thudt, Anthony Tang, and Michael Mauerer. "Constructive visualization." In: *Proceedings of the Conference on Designing Interactive Systems*. ACM. 2014, pp. 433–442. DOI: 10.1145/2598510.2598566 (cited on p. 42).
- [78] Petra Isenberg and Danyel Fisher. "Collaborative Brushing and Linking for Co-located Visual Analytics of Document Collections." In: *Computer Graphics Forum* 28.3 (2009), pp. 1031–1038. DOI: 10.1111/j.1467-8659.2009.01444.x (cited on p. 135).
- [79] Richard Ishida. *Ruby example*. <https://lists.w3.org/Archives/Public/www-style/2015Aug/0241.html>. Last read: July 2016 (cited on p. 61).
- [80] Mikkel Rønne Jakobsen and Kasper Hornbæk. "Transient visualizations." In: *Proceedings of the Conference on Computer-Human Interaction (OzCHI)*. ACM. 2007, pp. 69–76. DOI: 10.1145/1324892.1324905 (cited on p. 63).
- [81] Stan James. *Sparktweet creation tool*. <http://quantifiedself.com/sparktweet-tool/>. Last read: July 2016 (cited on p. 31).
- [82] S. Jänicke, G. Franzini, M. F. Cheema, and G. Scheuermann. "Visual Text Analysis in Digital Humanities." In: *Computer Graphics Forum* (2016). DOI: 10.1111/cgf.12873 (cited on pp. 4, 13–16, 147).
- [83] Yvonne Jansen and Pierre Dragicevic. "An interaction model for visualizations beyond the desktop." In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (2013), pp. 2396–2405. DOI: 10.1109/TVCG.2013.134 (cited on p. 128).

- [84] Nils P Johnson, Daniel T Johnson, Richard L Kirkeeide, Colin Berry, Bernard De Bruyne, William F Fearon, Keith G Oldroyd, Nico HJ Pijls, and K Lance Gould. "Repeatability of fractional flow reserve despite variations in systemic and coronary hemodynamics." In: *JACC: Cardiovascular Interventions* 8.8 (2015), pp. 1018–1027. DOI: 10.1016/j.jcin.2015.01.039 (cited on p. 31).
- [85] Paul E Keel. "Collaborative visual analytics: Inferring from the spatial organization and collaborative use of information." In: *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST)*. 2006, pp. 137–144. DOI: 10.1109/VAST.2006.261415 (cited on p. 135).
- [86] Daniel Keim, Gennady Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer, and Guy Melançon. "Visual analytics: Definition, process, and challenges." In: *Information Visualization*. Springer, 2008, pp. 154–175. DOI: 10.1007/978-3-540-70956-5\_7 (cited on p. 4).
- [87] Travis Kochel. *FF Chartwell Typeface*. <https://www.fontshop.com/families/ff-chartwell>. Last read: July 2016 (cited on p. 129).
- [88] Nicholas Kong, Marti A. Hearst, and Maneesh Agrawala. "Extracting References Between Text and Charts via Crowdsourcing." In: *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. ACM, 2014, pp. 31–40. DOI: 10.1145/2556288.2557241 (cited on pp. 23, 91).
- [89] Alexander Kowarik, Bernhard Meindl, and Matthias Templ. "sparkTable: Generating Graphical Tables for Websites and Documents with R." In: *The R Journal* 7.1 (2015) (cited on p. 128).
- [90] John Krumm. "A survey of computational location privacy." In: *Personal and Ubiquitous Computing* 13.6 (2009), pp. 391–399. DOI: 10.1007/s00779-008-0212-5 (cited on p. 136).
- [91] Kostiantyn Kucher and Andreas Kerren. "Text visualization techniques: Taxonomy, visual survey, and community insights." In: *IEEE Pacific Visualization Symposium (PacificVis)*. IEEE, 2015, pp. 117–121. DOI: 10.1109/PACIFICVIS.2015.7156366 (cited on pp. 4, 15).
- [92] Gordon Kurtenbach and William Buxton. "User Learning and Performance with Marking Menus." In: *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. ACM, 1994, pp. 258–264. DOI: 10.1145/191666.191759 (cited on p. 96).
- [93] Kuno Kurzhals, Brian Fisher, Michael Burch, and Daniel Weiskopf. "Evaluating visual analytics with eye tracking." In: *Proceedings of the Fifth Workshop on Beyond Time and Errors: Novel Evaluation*

- Methods for Visualization*. ACM. 2014, pp. 61–69. DOI: 10.1145/2669557.2669560 (cited on p. 125).
- [94] Bum Chul Kwon, Florian Stoffel, Dominik Jäckle, Bongshin Lee, and Daniel Keim. “VisJockey: Enriching Data Stories through Orchestrated Interactive Visualization.” In: *Poster Compendium of the Computation + Journalism Symposium*. 2014 (cited on p. 23).
  - [95] Bongshin Lee, Nathalie Henry Riche, Amy K Karlson, and Sheelagh Carpendale. “SparkClouds: Visualizing trends in tag clouds.” In: *IEEE Transactions on Visualization and Computer Graphics* 16.6 (2010), pp. 1182–1189. DOI: 10.1109/TVCG.2010.194 (cited on pp. 27, 60).
  - [96] Narges Mahyar and Melanie Tory. “Supporting Communication and Coordination in Collaborative Sensemaking.” In: *IEEE Transactions on Visualization and Computer Graphics* 20.12 (2014), pp. 1633–1642. DOI: 10.1109/TVCG.2014.2346573 (cited on p. 135).
  - [97] *MediaWiki Extension:Graph*. <https://www.mediawiki.org/>. Last read: July 2016 (cited on pp. 3, 127, 131).
  - [98] Microsoft Corporation. *Use sparklines to show data trends*. <https://support.office.com/en-us/article/Use-sparklines-to-show-data-trends-1474e169-008c-4783-926b-5c60e620f5ca>. Last read: April 2016 (cited on p. 25).
  - [99] Stanley Milgram. “Psychological maps of Paris.” In: *Environmental Psychology: People and their Physical Settings* (1976), pp. 104–124 (cited on p. 22).
  - [100] George A Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J Miller. “Introduction to WordNet: An on-line lexical database.” In: *International Journal of Lexicography* 3.4 (1990), pp. 235–244. DOI: 10.1093/ijl/3.4.235 (cited on p. 137).
  - [101] Franco Moretti. *Distant reading*. Verso Books, 2013 (cited on p. 14).
  - [102] Meredith Ringel Morris and Eric Horvitz. “SearchTogether: An interface for collaborative web search.” In: *Proceedings of the Conference on User Interface Software and Technology (UIST)*. ACM. 2007, pp. 3–12. DOI: 10.1145/1294211.1294215 (cited on p. 135).
  - [103] Tomer Moscovich, Fanny Chevalier, Nathalie Henry, Emmanuel Pietriga, and Jean-Daniel Fekete. “Topology-Aware Navigation in Large Networks.” In: *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. ACM, 2009, pp. 2319–2328. DOI: 10.1145/1518701.1519056 (cited on pp. 92, 99).



- [104] Miguel Nacenta, Uta Hinrichs, and Sheelagh Carpendale. "Fat-Fonts: Combining the Symbolic and Visual Aspects of Numbers." In: *Proceedings of Advanced Visual Interfaces (AVI)*. ACM, 2012, pp. 407–414. DOI: 10.1145/2254556.2254636 (cited on p. 26).
- [105] Marie Neurath and Robin Kinross. *The transformer: principles of making Isotype charts*. Hyphen, 2009 (cited on p. 42).
- [106] Dinh Quyen Nguyen, Christian Tominski, Heidrun Schumann, and Tuan Anh Ta. "Visualizing tags with spatiotemporal references." In: *Proceedings of Information Visualisation (IV)*. IEEE, 2011, pp. 32–39. DOI: 10.1109/IV.2011.43 (cited on p. 27).
- [107] Winfried Nöth. *Handbook of semiotics*. Indiana University Press, 1995 (cited on p. 54).
- [108] Jane O'Connor. *Benny's Big Bubble*. Grosset & Dunlap, 1997 (cited on p. 29).
- [109] Daniela Oelke, David Spretke, Andreas Stoffel, and Daniel A. Keim. "Visual Readability Analysis: How to Make Your Writings Easier to Read." In: *IEEE Transactions on Visualization and Computer Graphics* 18.5 (2012), pp. 662–674. ISSN: 1077-2626. DOI: 10.1109/TVCG.2011.266 (cited on p. 91).
- [110] W Bradford Paley. "TextArc: Showing word frequency and distribution in text." In: *Posters of the Conference on Information Visualization (InfoVis)*. 2002 (cited on p. 19).
- [111] Jonas Parnow. "Micro Visualizations: How can Micro Visualisations enhance text comprehension, memorability, and exploitation?" <http://www.microvis.info>. Design Master Thesis. Potsdam, Germany: Potsdam University of Applied Sciences, 2015 (cited on pp. 26, 53).
- [112] Jonas Parnow and Marian Dörk. "Micro Visualizations: Data-driven typography and graphical text enhancement." In: *Posters of the Conference on Information Visualization (InfoVis)*. IEEE. 2015 (cited on pp. 26, 33, 53).
- [113] Kevin B Paterson and Timothy R Jordan. "Effects of increased letter spacing on word identification and eye guidance during reading." In: *Memory & Cognition* 38.4 (2010), pp. 502–512. DOI: 10.3758/MC.38.4.502 (cited on p. 86).
- [114] Jennifer Pearson, George Buchanan, and Harold Thimbleby. "Improving annotations in digital documents." In: *Research and Advanced Technology for Digital Libraries*. Springer, 2009, pp. 429–432. DOI: 10.1007/978-3-642-04346-8\_51 (cited on p. 62).
- [115] Charles Sanders Peirce. "Collected Writings (8 Vols.)" In: *Harvard University Press, Cambridge* 58 (1931) (cited on p. 54).

- [116] Charles Perin, Pierre Dragicevic, and Jean-Daniel Fekete. "Revisiting bertin matrices: New interactions for crafting tabular visualizations." In: *IEEE Transactions on Visualization and Computer Graphics* 20.12 (2014), pp. 2082–2091. DOI: 10.1109/TVCG.2014.2346279 (cited on p. 120).
- [117] Charles Perin, Frédéric Vernier, and Jean-Daniel Fekete. "Interactive horizon graphs: Improving the compact visualization of multiple time series." In: *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. ACM. 2013, pp. 3217–3226. DOI: 10.1145/2470654.2466441 (cited on pp. 91, 95).
- [118] Charles Perin, Romain Vuillemot, and Jean-Daniel Fekete. "SoccerStories: A Kick-off for Visual Soccer Analysis." In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (2013), pp. 2506–2515. DOI: 10.1109/TVCG.2013.192 (cited on pp. 7, 26, 60, 112).
- [119] Ronald M Pickett and Georges G Grinstein. "Iconographic Displays for Visualizing Multidimensional Data." In: *Proceedings of the International Conference on Systems, Man, and Cybernetics*. Vol. 1. IEEE, 1988, pp. 514–519 (cited on p. 28).
- [120] Ben Pickles. *Peity*. <http://benpickles.github.io/peity/>. Last read: June 2016 (cited on pp. 128, 131).
- [121] Peter Pirolle and Stuart K. Card. "Information Foraging." In: *Psychological Review* 106.4 (1999), pp. 643–675. DOI: 10.1037/0033-295X.106.4.643 (cited on p. 74).
- [122] Hannes Reijner. "The development of the horizon graph." In: *Proceeding of Workshop From Theory to Practice: Design, Vision and Visualization Extended Abstracts of IEEE VisWeek*. Citeseer, 2008 (cited on p. 28).
- [123] Luz Rello, Martin Pielot, and Mari-Carmen Marcos. "Make it big! The effect of font size and line spacing on online readability." In: *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. Vol. 16. 2015. DOI: 10.1145/2858036.2858204 (cited on p. 125).
- [124] Fabrice Rimlinger. *Sparklines for Excel®*. <http://sparklines-excel.blogspot.fr>. Last read: July 2016 (cited on p. 128).
- [125] Marije Rooze. *Collection of interactive graphic work by Guardian and the New York Times*. Website. <http://collection.marijerooze.nl>. Last read: June 2016 (cited on p. 2).
- [126] Timo Ropinski, Steffen Oeltze, and Bernhard Preim. "Survey of glyph-based visualization techniques for spatial multivariate medical data." In: *Computers & Graphics* 35.2 (2011), pp. 392–401. DOI: 10.1016/j.cag.2011.01.011 (cited on pp. 27, 63).

- [127] Jeffrey M. Rzeszotarski and Aniket Kittur. "Kinetica: naturalistic multi-touch data visualization." In: *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. ACM, 2014, pp. 897–906. DOI: 10.1145/2556288.2557231 (cited on p. 107).
- [128] Takafumi Saito, Hiroko Nakamura Miyamura, Mitsuyoshi Yamamoto, Hiroki Saito, Yuka Hoshiya, and Takumi Kaseda. "Two-tone pseudo coloring: Compact visualization for one-dimensional data." In: *Proceedings of the Conference on Information Visualization (InfoVis)*. IEEE. 2005, pp. 173–180. DOI: 10.1109/INFVIS.2005.1532144 (cited on p. 28).
- [129] Jeff Sauro and James R Lewis. "Average task times in usability tests: what to report?" In: *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. ACM. 2010, pp. 2347–2350. DOI: 10.1145/1753326.1753679 (cited on p. 118).
- [130] Marcin Sawicki, Michel Suignard, Masayasu Ishikawa, Martin Dürst, and Tex Texin. "Ruby Annotation." In: *W3C Recommendation (May 31, 2001)* (2001) (cited on p. 62).
- [131] H.-J. Schulz, T. Nocke, M. Heitzler, and H. Schumann. "A Design Space of Visualization Tasks." In: *IEEE Transactions on Visualization and Computer Graphics* 19.12 (2013), pp. 2366–2375. DOI: 10.1109/TVCG.2013.120 (cited on p. 64).
- [132] Adam Schwartz. *Randomized Sparklines*. <http://adamschwartz.co/et-notebooks/sparklines-randomized/>. Last read: July 2016 (cited on p. 25).
- [133] Ben Shneiderman. "The eyes have it: A task by data type taxonomy for information visualizations." In: *Proceedings of the IEEE Symposium on Visual Languages*. IEEE. 1996, pp. 336–343. DOI: 10.1109/VL.1996.545307 (cited on pp. 35, 40, 90, 99).
- [134] John H Siegel, Edward J Farrell, Roger M Goldwyn, and Herman P Friedman. "The surgical implications of physiologic patterns in myocardial infarction shock." In: *Surgery* 72.1 (1972), pp. 126–141 (cited on p. 28).
- [135] Timothy J Slattery and Keith Rayner. "Effects of intraword and interword spacing on eye movements during reading: Exploring the optimal use of space in a line of text." In: *Attention, Perception, & Psychophysics* 75.6 (2013), pp. 1275–1292. DOI: 10.3758/s13414-013-0463-8 (cited on p. 86).
- [136] Samuel Spencer. *Twitter Sparkline Generator using Unicode*. <http://sandbox.kidstrythisathome.com/louis/>. Last read: July 2016 (cited on p. 128).

- [137] John Stasko, Carsten Gorg, Zhicheng Liu, and Kanupriya Singhal. "Jigsaw: Supporting Investigative Analysis Through Interactive Visualization." In: *Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST)*. IEEE, 2007, pp. 131–138. DOI: 10.1109/VAST.2007.4389006 (cited on pp. 5, 6, 105).
- [138] Markus Steinberger, Manuela Waldner, Marc Streit, Alexander Lex, and Dieter Schmalstieg. "Context-Preserving Visual Links." In: *IEEE Transactions on Visualization and Computer Graphics* 17.12 (2011), pp. 2249–2258. ISSN: 1077-2626. DOI: 10.1109/TVCG.2011.183 (cited on pp. 23, 67, 91).
- [139] Robert Sternberg and Karin Sternberg. *Cognitive psychology*. 6th ed. Cengage Learning, 2011 (cited on p. 85).
- [140] Maureen Stone. "In color perception, size matters." In: *IEEE Computer Graphics and Applications* 32.2 (2012), pp. 8–13. DOI: 10.1109/MCG.2012.37 (cited on p. 64).
- [141] Hendrik Strobel, Daniela Oelke, Christian Rohrdantz, Andreas Stoffel, Daniel A Keim, and Oliver Deussen. "Document cards: A top trumps visualization for documents." In: *IEEE Transactions on Visualization and Computer Graphics* 15.6 (2009), pp. 1145–1152. DOI: 10.1109/TVCG.2009.139 (cited on pp. 20, 21).
- [142] Bongwon Suh, Allison Woodruff, Ruth Rosenholtz, and Alyssa Glass. "Popout Prism: Adding Perceptual Principles to Overview+Detail Document Interfaces." In: *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. ACM, 2002, pp. 251–258. DOI: 10.1145/503376.503422 (cited on p. 91).
- [143] Justin Talbot, John Gerth, and Pat Hanrahan. "An empirical model of slope ratio comparisons." In: *IEEE Transactions on Visualization and Computer Graphics* 18.12 (2012), pp. 2613–2620. DOI: 10.1109/TVCG.2012.196 (cited on p. 64).
- [144] *Tangled*. <http://worrydream.com/Tangle/>. Last read: July 2016 (cited on p. 90).
- [145] Matthew Tobiasz, Petra Isenberg, and Sheelagh Carpendale. "Lark: Coordinating co-located collaboration with information visualization." In: *IEEE Transactions on Visualization and Computer Graphics* 15.6 (2009), pp. 1065–1072. DOI: 10.1109/TVCG.2009.162 (cited on p. 128).
- [146] Christian Tominski, Camilla Forsell, and Jimmy Johansson. "Interaction Support for Visual Comparison Inspired by Natural Behavior." In: *IEEE Transactions on Visualization and Computer Graphics* 18.12 (2012), pp. 2719–2728. ISSN: 1077-2626. DOI: 10.1109/TVCG.2012.237 (cited on p. 103).
- [147] Edward R. Tufte. *The visual display of quantitative information*. Cheshire, CT: Graphics Press, 2001 (cited on pp. 4, 29, 100).

- [148] Edward R Tufte. *Beautiful Evidence*. Vol. 1. Cheshire, CT: Graphics Press, 2006 (cited on pp. 7, 25, 33, 53, 60).
- [149] Edward Tufte. *Envisioning Information*. Cheshire, CT: Graphics Press, 1990 (cited on p. 70).
- [150] *Twitter analytics*. <https://analytics.twitter.com/>. Last read: July 2016 (cited on p. 31).
- [151] *Vega: A Visualization Grammar*. <http://trifacta.github.io/vega>. Last read: July 2016 (cited on p. 3).
- [152] Benoit Verjat. *Semantic word-scale graphics*. <http://benoit.verjat.com>. Last read: April 2016 (cited on p. 29).
- [153] Bret Victor. *Explorable explanations*. <http://worrydream.com/ExplorableExplanations/>. Last read: June 2015 (cited on pp. 7, 33, 90, 96).
- [154] Fernanda B Viégas and Martin Wattenberg. “Timelines tag clouds and the case for vernacular visualization.” In: *Interactions* 15.4 (2008), pp. 49–52. DOI: 10.1145/1374489.1374501 (cited on pp. 22, 23).
- [155] Fernanda B Viegas, Martin Wattenberg, Frank Van Ham, Jesse Kriss, and Matt McKeon. “ManyEyes: a site for visualization at internet scale.” In: *IEEE Transactions on Visualization and Computer Graphics* 13.6 (2007), pp. 1121–1128. DOI: 10.1109/TVCG.2007.70577 (cited on pp. 17, 22).
- [156] Manuela Waldner, Werner Puff, Alexander Lex, Marc Streit, and Dieter Schmalstieg. “Visual links across applications.” In: *Proceedings of Graphics Interface*. Canadian Information Processing Society. 2010, pp. 129–136 (cited on p. 91).
- [157] Matthew O Ward. “A Taxonomy of Glyph Placement Strategies for Multidimensional Data Visualization.” In: *Information Visualization* 1.3-4 (2002), pp. 194–210. DOI: 10.1057/palgrave.ivs.9500025 (cited on pp. 27, 61).
- [158] Matthew Ward. “Multivariate Data Glyphs: Principles and Practice.” In: *Handbook of Data Visualization*. Springer, 2008, pp. 179–198. DOI: 10.1007/978-3-540-33037-0\_8 (cited on p. 27).
- [159] Colin Ware. *Information Visualization: Perception for Design*. Elsevier, 2012 (cited on p. 23).
- [160] Martin Wattenberg and Fernanda B Viégas. “The word tree, an interactive visual concordance.” In: *IEEE Transactions on Visualization and Computer Graphics* 14.6 (2008), pp. 1221–1228. DOI: 10.1109/TVCG.2008.172 (cited on p. 17).

- [161] Wesley Willett, Pascal Goffin, and Petra Isenberg. "Understanding Digital Note-Taking Practice for Visualization." In: *IEEE Computer Graphics and Applications, Special Issue on Personal Visualization and Personal Visual Analytics* 35.4 (2015), pp. 38–51. DOI: 10.1109/MCG.2015.52 (cited on pp. 134, 136, 141).
- [162] Wesley Willett, Jeffrey Heer, and Maneesh Agrawala. "Scented widgets: Improving navigation cues with embedded visualizations." In: *IEEE Transactions on Visualization and Computer Graphics* 13.6 (2007), pp. 1129–1136. DOI: 10.1109/TVCG.2007.70589 (cited on pp. 28, 135).
- [163] WordNet. <https://wordnet.princeton.edu>. Last read: July 2016 (cited on p. 137).
- [164] Patricia Wright, Robert Milroy, and Ann Lickorish. "Static and animated graphics in learning from interactive texts." In: *European Journal of Psychology of Education* 14.2 (1999), pp. 203–224. DOI: 10.1007/BF03172966 (cited on pp. 118, 124).
- [165] Ji Soo Yi, Youn ah Kang, John T Stasko, and Julie A Jacko. "Toward a deeper understanding of the role of interaction in information visualization." In: *IEEE Transactions on Visualization and Computer Graphics* 13.6 (2007), pp. 1224–1231. DOI: 10.1109/TVCG.2007.70515 (cited on pp. 92, 93, 107).
- [166] Dongwook Yoon, Nicholas Chen, and François Guimbretière. "TextTearing: Opening white space for digital ink annotation." In: *Proceedings of the Conference on User Interface Software and Technology (UIST)*. ACM. 2013, pp. 107–112. DOI: 10.1145/2501988.2502036 (cited on pp. 62, 101).
- [167] Polle T Zellweger, Bay-Wei Chang, and Jock D Mackinlay. "Fluid links for informed and incremental link transitions." In: *Proceedings of the Ninth ACM Conference on Hypertext and Hypermedia*. ACM. 1998, pp. 50–57. DOI: 10.1145/276627.276633 (cited on pp. 85, 91, 95, 96, 101).
- [168] Polle T Zellweger, Susan Harkness Regli, Jock D Mackinlay, and Bay-Wei Chang. "The impact of fluid documents on reading and browsing: An observational study." In: *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. ACM. 2000, pp. 249–256. DOI: 10.1145/332040.332440 (cited on pp. 62, 63, 65, 112, 113).
- [169] Polle T Zellweger, Niels Olof Bouvin, Henning Jehøj, and Jock D Mackinlay. "Fluid annotations in an open world." In: *Proceedings of the 12th ACM Conference on Hypertext and Hypermedia*. ACM. 2001, pp. 9–18. DOI: 10.1145/504216.504224 (cited on p. 129).