



HAL
open science

Composite visual mapping for time series visualization

Ali Jabbari

► **To cite this version:**

Ali Jabbari. Composite visual mapping for time series visualization. Computer Arithmetic. Université Grenoble Alpes, 2018. English. NNT : 2018GREAM035 . tel-01913579

HAL Id: tel-01913579

<https://theses.hal.science/tel-01913579>

Submitted on 6 Nov 2018

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

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTE UNIVERSITE GRENOBLE ALPES

Spécialité : **Informatique**

Arrêté ministériel : 25 mai 2016

Présentée par

ALI JABBARI

Thèse dirigée par **Sophie DUPUY-CHESSA** et co-encadrée par
Renaud BLANCH

préparée au sein du **Laboratoire d'Informatique de Grenoble,**
équipe Ingénierie de l'Interaction Homme-Machine
et de l'**École Doctorale Mathématiques, Sciences et**
Technologies de l'Information, Informatique

Encodage visuel composite pour les séries temporelles

Composite Visual Mapping for Time Series Visualization

Thèse soutenue publiquement le **4 juillet 2018**,
devant le jury composé de :

M. Georges-Pierre BONNEAU

Professeur à l'Université Grenoble-Alpes, Président

M. Christophe HURTER

Professeur à l'École Nationale de l'Aviation Civile, Rapporteur

Mme. Mountaz HASCOËT

Maître de Conférences à l'Université de Montpellier, Rapporteur

M. David AUBER

Maître de Conférences à l'Université de Bordeaux, Examineur

Mme. Sophie DUPUY-CHESSA

Professeur à l'Université Grenoble-Alpes, Laboratoire d'Informatique de
Grenoble, Directrice de thèse

M. Renaud BLANCH

Maître de Conférences à l'Université Grenoble-Alpes, Laboratoire
d'Informatique de Grenoble, Co-encadrant de thèse



Abstract

Time series are one of the most common types of recorded data in various scientific, industrial, and financial domains. Depending on the context, time series analysis are used for a variety of purposes: forecasting, estimation, classification, and trend and event detection. Thanks to the outstanding capabilities of human visual perception, visualization remains one of the most powerful tools for data analysis, particularly for time series. With the increase in data sets' volume and complexity, new visualization techniques are clearly needed to improve data analysis. They aim to facilitate visual analysis in specified situations, tasks, or for unguided exploratory analysis.

Visualization is based upon visual mapping, which consists in association of data values to visual channels, e.g. position, size, and color of the graphical elements. In this regard, the most familiar form of time series visualization, i.e. line charts, consists in a mapping of data values to the vertical position of the line. However, a single visual mapping is not suitable for all situations and analytical objectives. Our goal is to introduce alternatives to the conventional visual mapping and find situations in which, the new approach compensate for the simplicity and familiarity of the existing techniques. We present a review of the existing literature on time series visualization and then, we focus on the existing approaches to visual mapping.

Next, we present our contributions. Our first contribution is a systematic study of a *composite* visual mapping which consists in using combinations of visual channels to communicate different facets of a time series. By means of several user studies, we compare our new visual mappings with an existing reference technique and we measure users' speed and accuracy in different analytical tasks. Our results show that the new visual designs lead to analytical performances close to those of the existing techniques without being unnecessarily complex or requiring training. Also, some of the proposed mappings outperform the existing techniques in space constrained situations. Space efficiency is of great importance to simultaneous visualization of large volumes of data or visualization on small screens. Both scenarios are among the current challenges in information visualization.

Résumé

Les séries temporelles sont l'un des types de données les plus courants dans divers domaines scientifiques, industriels et financiers. Selon le contexte, l'analyse des séries temporelles est effectuée à diverses fins : prévision, estimation, classification et détection des tendances et des événements. Grâce aux capacités exceptionnelles de la perception visuelle humaine, la visualisation reste l'un des outils les plus puissants pour l'analyse de données, en particulier pour les données temporelles. Avec la croissance de volume et de la complexité des jeux de données, de nouvelles techniques de visualisation sont clairement nécessaires pour améliorer l'analyse des données. Elles visent à faciliter l'analyse visuelle dans le cas où des situations ou des tâches sont bien spécifiées, ou à favoriser l'analyse exploratoire non guidée.

La visualisation est basée sur *l'encodage visuel*, un processus qui consiste à associer les valeurs de données aux canaux visuels comme la position, la taille et la couleur des éléments graphiques. A cet égard, la forme la plus connue de visualisation des séries temporelles, c'est-à-dire les graphiques linéaires (*line charts* en anglais), consiste en une mise en correspondance des valeurs de données avec la position verticale de la ligne. Cependant, un seul encodage visuel ne convient pas à toutes les situations et objectifs analytiques. Notre but est d'introduire des alternatives à l'encodage visuel conventionnel et de trouver des situations dans lesquelles, la nouvelle approche compense la simplicité et la familiarité des techniques existantes. Nous présentons une revue de l'état de l'art sur la visualisation des séries temporelles, puis nous nous concentrons sur les approches existantes de l'encodage visuel.

Ensuite, nous présentons nos contributions. Notre première contribution est une étude systématique d'un *encodage visuel composite* qui consiste à utiliser des combinaisons de canaux visuels pour communiquer différentes facettes d'une série temporelle. Au moyen de plusieurs expériences avec des utilisateurs, nous comparons les nouveaux encodages visuels à une technique de référence existante et nous mesurons la vitesse et la précision des utilisateurs dans différentes tâches. Nos résultats montrent que les nouvelles conceptions visuelles conduisent à des performances proches de celles des techniques existantes sans être inutilement complexes ou nécessiter un apprentissage. De plus, certains encodages proposés surpassent les techniques existantes dans les situations de contraintes spatiales. L'efficacité spatiale est d'une grande importance pour la visualisation simultanée de grands volumes de données ou pour la visualisation sur de petits écrans. Les deux scénarios font partie des défis actuels de la visualisation d'information.

Contents

List of Figures	ix
List of Tables	xi
Acknowledgements	xiii
Introduction	1
I State of the Art	9
1 Background	11
1.1 Visualization	11
1.1.1 Visualization Reference Model	12
1.1.2 Human interaction	16
1.2 Visualization Criteria	16
1.2.1 Visualization Aspect: Goals and Tasks	17
1.2.2 Visualization Aspect: Data	18
1.2.3 Visualization Aspect: Representation	22
1.3 Temporality	23
1.3.1 Temporal Primitives: Points vs. Intervals	23
1.3.2 Temporal Structure: Linear vs. Cyclic	25
1.4 Space Efficiency Management Techniques	26
1.5 Summary	26
2 Visual Channels & Their Conjunctions	27
2.1 Marks and Visual Channels	27
2.2 List of Visual Channels	29
2.2.1 Position	29
2.2.2 Size	33
2.2.3 Hue	34
2.2.4 Chroma	34
2.2.5 Lightness	35
2.2.6 Texture and Pattern	36
2.2.7 Classification of Visual Channels	36
2.3 Visual Channels Conjunctions	38
2.3.1 Bivariate Maps	38
2.3.2 Scale Manipulation Techniques	41
2.3.3 LifeLines & MIDGAARD	41
2.3.4 Colorfield techniques	44
2.3.5 CloudLines	44
2.3.6 Order of Magnitude Markers	45
2.3.7 Slick Graphs	46
2.3.8 Horizon Graphs	47
2.4 Discussion of the Related Works	48

2.5	Summary	50
II	Contributions	53
3	Composite Visual Mapping	55
3.1	Promises	56
3.2	Decomposition Functions	56
3.2.1	Modulo Operation	57
3.2.2	Logarithmic Transformation	57
3.2.3	Frequency Separation	58
3.3	Systematic Study of Composite Visual Mapping	59
3.3.1	Implementation	62
3.3.2	Existing Composite Visual Mappings	64
3.4	Summary	64
4	Evaluation of Composite Visual Mappings	65
4.1	Challenges of Visualization Evaluation	65
4.2	Approach and Methods	67
4.2.1	Experimental Protocole	67
4.2.2	Chart Generation	70
4.3	Evaluated Composite Mappings	70
4.4	Results	71
4.4.1	Performance by Data Transformation	72
4.4.2	Performance by Visual Mapping	75
4.4.3	Performance by Number of Layers	75
4.4.4	Performance by Real Value Differences	75
4.4.5	Discussion of the Results	79
4.5	Limits and Perspectives	80
4.6	Summary	80
5	Composite Visual Mapping & Space Efficiency	83
5.1	Visual Mapping and Chart Height	84
5.2	Approach and Methods	84
5.2.1	Experimental Protocole	85
5.2.2	Chart Generation	85
5.3	Evaluated Visual Mappings	86
5.4	Pilot Studies	88
5.4.1	Pilot I: Mapping Selection	89
5.4.2	Pilot II: Color Scheme Selection	90
5.4.3	Pilot III: Texture Selection	91
5.5	Main Experiment	91
5.5.1	Results	92
5.5.2	Results from the Questionnaire	95
5.5.3	Discussion of the Results	97
5.6	Design Implications	97
5.7	Limits and Perspectives	99
5.8	Summary	99
	Conclusion	101

Appendices	105
A Data Generation	107
B Construction of an Example Composite Visual Mapping	113
Bibliography	115

List of Figures

1	Playfair’s line chart	5
1.1	Visualization Reference Model	12
1.2	an example of unsuitable visual mapping	15
1.3	Visualization Criteria	17
1.4	An example of dimension hierarchies	20
1.5	Effect of dimension ordering in a parallel coordinate plot	21
1.6	Time primitives	25
2.1	Visual channel <i>Shape</i>	28
2.2	Visual channel <i>Position</i>	30
2.3	Graphical impositions for two dimensional representation	31
2.4	Graphical impositions of diagrams	32
2.5	Linear and circular representation of same data	32
2.6	Visual channel <i>Size</i>	33
2.7	Color visual channels: hue, chroma, and lightness	35
2.8	Visual channel <i>Texture</i>	37
2.9	Vibratory effects of pattern repetition	37
2.10	Mackinlay’s ranking of visual channels by task	38
2.11	Bivariate maps	40
2.12	Dual scale charts	42
2.13	LifeLines	43
2.14	Midgaard visualization system	43
2.15	Color mapping for multi-resolution views	44
2.16	Color mapping for averaging	45
2.17	CloudLines	46
2.18	Order of Magnitude Markers	47

2.19	slick Graphs	48
2.20	Horizon Graphs	50
2.21	Qualizon Graphs	51
3.1	Exploratory table of composite visual mappings	58
3.2	Composite mapping with frequency separation	60
3.3	Configuration table of the interactive composite visual mapping table	62
3.4	Mapping of modulo sub-values to visual channels	63
4.1	Graphical interface of the experiments showing a typical trial	69
4.2	Visual mappings evaluated in the first experiment	72
4.3	Effects of transformation function: modulo vs. logarithmic	73
4.4	Comparison of the performance of the selected composite mappings	74
4.5	Effects of layering: two layers vs. three layers	76
4.6	Effects of the real value difference on accuracy	77
4.7	Effects of the real value difference on speed	78
5.1	Visual mappings examined in the second experiment	86
5.2	Absolute estimation error for line charts and horizon graphs in pilot I	89
5.3	color schemes evaluated in pilot II	90
5.4	Texture patterns evaluated in pilot III	91
5.5	Tukey comparison of discrimination time among mappings	93
5.6	Distribution of estimation error by visual mapping across all chart heights	94
5.7	Tukey comparison of estimation accuracy across chart heights	94
5.8	Tukey comparison of estimation accuracy among mappings	95
5.9	Tukey comparison of estimation time across chart heights	96
5.10	Participants reported general satisfaction	96
5.11	Linear regression lines for evaluated mappings across chart heights	98
A.1	Constrained random walk generation	110
A.2	Distribution of Real Value Differences	111
B.1	Construction of an Example Composite Visual Mapping	114

List of Tables

1.1	An example of a data table	13
1.2	General classification of visualization techniques	24
2.1	Summarized characteristics of visual channels	39
2.2	Summary of the related visualization techniques	49

Acknowledgements

This thesis becomes reality with kind help and support of many individuals.

I would like to thank Sophie and Renaud for their continuous support of my Ph.D study, for their patience, involvement, and immense knowledge. Besides my supervisors, I would like to thank all the members of my thesis committee for reviewing my work and their invaluable feedback.

My sincere thanks also go to my fellow doctoral students for their feedback, cooperation and of course friendship.

Last but not the least, I would like to thank my family: my mom who supported me throughout this enterprise and my life in general and Julie for her love and endless support specially during the difficult times.

Introduction

“A full 90 percent of all the data in the world has been generated over the last two years.”

— Åse Dragland, SINTEF

You must have heard this fun fact or similar ones lately. *Big Data*, *Data Mining*, or *Data Science* are all among the buzz words in the recent years. From industry processes to personal activities, terabytes of data is created every second around the world. However, massive collection of data is futile, unless stored data is processed and analyzed. *Data analysis* is the process of inspecting data with the goal of discovering information and constructing useful knowledge from the gathered data.

Data analysis is an essential part of all scientific research and problem solving in industry. There exist different approaches to data analysis. At the most basic level, descriptive statistics derive aggregated information such as mean and variance of data values which can help the analyst to understand the global properties of the data. More recently, data mining and machine learning methods are applied to extract categories, patterns and unusual records in data. In addition to the computational methods, visualization aims to communicate information by means of plots and graphical representations of data. Visualization is used as a tool to present discovered information, but also, as an exploratory tool to examine data and extract the information.

Among data analysis approaches, data visualization is the most reliant on human cognitive skills. The analyst does not require any sophisticated skill to be able to interpret the visual representation of data. While computational approaches need defined analytical objectives, data visualization allow for uninstructed exploration of data and unstructured insights. Such exploration of data is only limited by human visual perception, creativity, and imagination. Considering its characteristics, the field of data visualization and visual analysis is closely related to other areas such as data architecture, human-computer interaction, visual perception, and graphic design.

The great facility of visualization for presentation and exploration of data has made it a common tool in diverse areas. The applications of visualization vary from data

analysis in technical domains, to fields more accessible to general public, i.e., business analytics and data-driven journalism. Such diversity of interests and goals has led to great variety of visualization techniques.

In this thesis, we only focus on visualization of a specific type of data, i.e., time series data. First, we motivate our interest in time series visualization and we point out to the current challenges in this field.

Focus: Time Series Visualization

A time series is a set of observations (i.e. events or measured values) indexed in temporal order. Time series are widely used in science, engineering, finance, and industrial contexts. Visualization of the data collected in this form is a vital tool for experts in many domains (climatology, financial markets, production monitoring, just to name a few) for making decisions, detecting events and abnormalities, and forecasting a future event or trend.

Despite the large number of publications in recent years, visualization of time series remains an open question worth study. Most of the proposed approaches are specific to only a certain analytic problem. The existence of so many purpose-built is mainly due to the fact that it is extremely difficult to consider all task, data, and representation related aspects involved when visualizing time series. Each approach can at best enhance visual analysis with a set of specified tasks performed on specified datasets. For example, a visualization technique which is appropriate for smooth periodic time series might not be compatible with noisy time series where the undesired noise obscures the periodicity of data.

As the field of application of time series visualization grows wider, new contributions to this domain can benefit a large number of scientific and industrial domains. This promising perspective has convinced us to focus our research on improvements in time series visualization. The solutions proposed in this work aim to meet several requirements currently facing time series visualization.

Requirement: Data Volume and Complexity Management

Visualization is a means to understand the available data and to help users analyze and reason about it. The presence of the temporal dimension in data introduces further implications. Specifically, analysis of the data through time allows to understand the behavior of the source of data. This reasoning allows users to model the behavior of the data in the past and predict its evolution in the future. Such analytical process is a key part of studies in many fields from industrial processes to scientific measurements.

With the advances in sensor technology and other data gathering techniques, the need for visualization of data parameters faces several challenges. First, the number of data parameters is growing, leading to high dimensional data sets consisting of hundreds of data variables. Higher dimensionality consequently complicates simultaneous overview of all data parameters. Second, the volume of data that is recorded is ever increasing growing past billions of records per data variable. In such conditions, visualization of all data without information loss becomes practically impossible. We hence seek solutions which allow visualization of complex datasets without introducing excessive information loss.

Requirement: High Space Efficiency

The visualized data is often displayed in a visualization dashboard on a workstation screen. In many scenarios involving visualization of multidimensional data (e.g. simultaneous monitoring of tens or hundreds of production processes), the human analyst is interested in visualizing and comparing a multitude of data variables at the same time on a limited screen size. Besides stationary dashboards, handheld devices (e.g. touch screen tablets) have emerged in recent years as a new medium for visualization which offers more mobility at the expense of less exploitable screen real estate.

One common challenge at the core of both situations (stationary dashboards with multiple time series, and mobile devices) is the limit of the available screen space in terms of exploitable pixels. Screen space is thus a valuable commodity in visualization systems, since it is in proportion to the amount of data that can be visualized.

Tufte [81] introduced data-to-ink ratio as the proportion of a graphic's ink (or pixels in digital media) devoted to non-redundant visualization of data and argued that data-to-ink ratio should always be maximized within reason. In the past, various techniques have been proposed to enhance space efficiency management. We summarized these techniques in Chapter 1. These techniques either control the amount of visualized data at any given time or distort the visual configuration to represent information more efficiently.

Despite all previous efforts, the ever increasing dimensionality and volume of data sets impose space efficiency of visualization as a remaining current challenge. The compactness of our visual designs is one of the qualities that makes them appropriate for visualization of time series in current applications.

Requirement: Minimum Visual Artefacts

Practically, as the number of graphical objects (points and lines) becomes large compared to available pixels on screen, no accurate visualization of data is possible. Once

the number of graphical objects exceeds a certain level, humans perceive the resulting visualization as a shapeless blob due to occlusion [26].

The resulting visual clutter obscures data structures and makes data exploration extremely difficult. In order to address this recurring problem, several methods have been proposed in literature. Some methods reduce the number of graphical entities by sampling data points or dimensions [26], while others aim to reduce the clutter by dimension reduction or dimension reordering techniques [88]. While these methods resolve visual clutter to a certain degree, appropriate interaction techniques are required in order to preserve the integrity of data and expressiveness of the visualization.

Moreover, as the size of data increases, dynamic representation (e.g. animations) becomes confusing. In addition, 3D representation of data becomes less efficient due to the increased overlapping and occultation, and density control measures would be necessary for seeing through a 3D space [26]. Visual clutter remains an ongoing challenge in design of new visualizations. We decided to avoid such visual constraints as much as possible in design of our visualization techniques.

Possible Solutions

We can outline two approaches that address such challenges. The first approach aims to adapt complex datasets to conventional visualization techniques by volume and complexity management techniques. While this family of computational techniques are out of the scope of this thesis, we briefly overview such approach in Chapter 1. Nevertheless, these techniques can be adapted to be integrated into any visualization technique.

The second approach which concerns this thesis, consists in conceiving new visualization techniques or enhancing the existing ones in order to make them compatible with complex data sets. The most well known technique for visualization of temporal data is a *line chart*. Since its introduction in the 18th century, line charts have been used in many fields to visualize a trend in data over time – a *time series* (See Figure 1). However, as the data sets become larger and more complex and analytical objectives become more sophisticated, such simple visualization techniques need enhancements to maintain accuracy and speed.

In this regard, new visualization techniques exploit human visual capabilities in order to offer visualization techniques for complex data sets. The new techniques should maximize the amount of communicated information while taking into account the complexity and volume of data, the limits of human vision, and thrifty use of viewing space. Hence in this thesis, we look for new visual designs which improve communication of information, boost space efficiency, and profit from the less used capabilities of human vision.

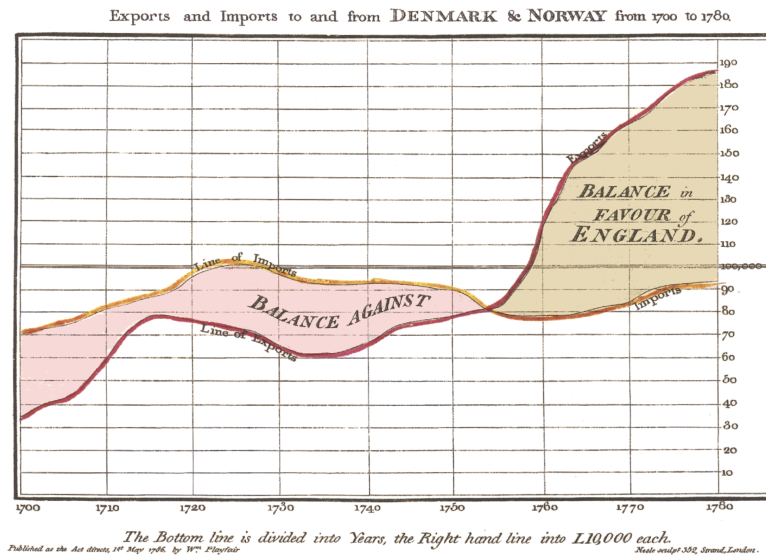


Figure 1: Playfair’s line chart [71] is considered as one of the earliest examples of the visualization technique. Here, showing the balance of trade between England, and Denmark and Norway.

We consider visual mapping as the most critical step of visualization process. We noticed in our review of the existing works that improvements in visual mapping can remarkably boost performance of the users in the evaluated analytical tasks. The most closely related existing techniques have been successfully implemented in the time series visualization systems. Their has convinced us that alternatives to conventional visual mapping are worth a systematic study.

Approach and Contributions

Our approach to visual design for time series is based on the notion of visual channels. Visual channels act as the basis for any visualization and their appropriate selection and modification can improve the overall quality of visualization. We employ combinations of visual channels in order to reinforce information communication. We make use of the extra visual channels to communicate additional information on desired properties of data. This auxiliary information is determined based on the goals of the analysis. For instance, if the analyst is interested in comparing data values at different points, such auxiliary could be the order of magnitude. Separating different facets of data in such way allows to dedicate a full range visual channel (e.g. color saturation) to the most important aspects of data and allow users to selectively attend those characteristics of data. We call this approach *Composite Visual Mapping*.

To this aim, we review the literature with a focus on the type of visual mapping. Next, we systematically study possible combinations of visual channels along with several decomposition operations. We selected some promising techniques among this design space for more in depth study. Our second contribution consists in an empirical study in which we compare our selected techniques with an existing reference technique, i.e. *horizon graphs* in some common analytical tasks. Our results show that our techniques are in general on a par with the reference technique. We hypothesize that some of our visual channels can outperform the existing techniques in specific scenarios. Our third contribution is thus an empirical study on the performance of a selection of composite mappings across different chart sizes. We show that at least some of the examined composite visual mappings perform more accurately or faster than existing techniques at small sizes.

Overall, we demonstrate that our approach to visual mapping can improve communication of target information especially in space-constrained situations. This is a recurrent situation in visualization of large amounts of data on limited screen space. At the same time, the new visual designs turned out to be easy to learn for the users and allow to benefit from underused capabilities of human vision without being unnecessarily complicated. Also, thanks to our empirical evaluations, we come up with several design guidelines for the implementation of composite mapping and combination of visual channels.

Outline of the Thesis

This thesis is structured as follows:

In **Chapter 1**, we present the theoretical background of our work. We provide the definitions and general concepts in visualization of temporal data. We also analyse the global classification of visualization techniques based on different visualization criteria.

Due to the importance of the notion of visual channels in our work, we dedicate **Chapter 2** to introducing the visual channels used through current work and the most relevant existing works. We review the literature from the perspective of visual mapping and the combinations of visual channels.

In **Chapter 3**, we introduce the notion of composite visual mappings and decompositional functions. We subsequently describe our systematic review of the entire design space using our exploratory table.

Chapter 4 describes our first evaluation of a selection of composite visual mappings. This evaluation allowed us to identify the overall strengths and limits of our approach and refine our visual designs for later studies.

Finally, **Chapter 5** presents our second evaluation on the use of composite visual

mappings and the effects of chart height. Thanks to this evaluations, we showed that at least some of our techniques outperform the existing techniques in space constraint situations. Based on our findings, we provide design guidelines on the use of composite visual mappings and combinations of visual channels.

Part I

State of the Art

Chapter 1

Background

In this first chapter, we introduce the theoretical background of our work. We provide the definitions of a few key concepts which are essential for better understanding of the work presented in this thesis. We present the general considerations in design of visualizations and overview the classification of visual representations.

1.1 Visualization

Until recently, the term *visualization* referred to forming a mental image of something in the mind. In 1987, McCormick et al. [64] introduced the notion of visualization in computer science. They defined the term visualization as follows:

“Visualization is a method of computing. It transforms the symbolic into the geometric, enabling researchers to observe their simulations and computations. Visualization offers a method for seeing the unseen. It enriches the process of scientific discovery and fosters profound and unexpected insights.”

Visualization has hence transformed from a merely internal process of constructing images in the mind to an external representation of data and ideas supporting human users in their cognitive tasks. Visualization has advantages that explain its ubiquity in many fields, from daily life to technical and scientific contexts [84]:

- Visualization harness fast and effective communication of information. The most important insights from thousands of collected measurements (e.g., maps) can be obtained by a brief glance at an image.

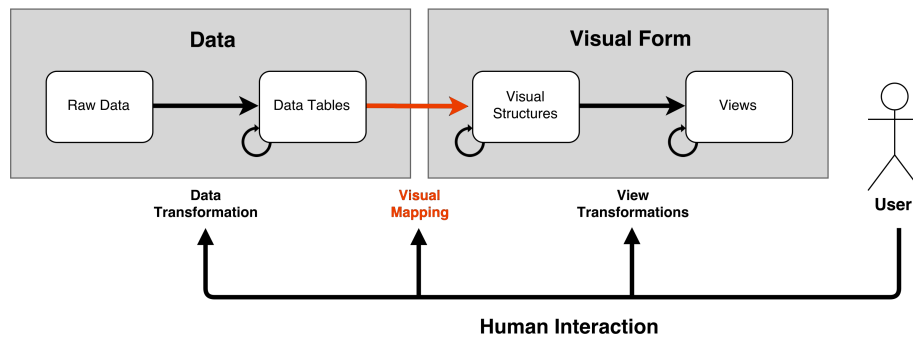


Figure 1.1: Visualization reference model as formalized by Card *et al.* [17]. The stage of visual mapping is the focus of this thesis.

- Visualization allows the perception of emergent properties that are otherwise implicit. Viewers of a visualization can detect patterns and other unanticipated properties and establish basis for further insights such as prediction, hypothesis formation, and decision making.
- Visualization can also be used as a tool for evaluating the quality of data collection by showing missing records, errors and anomalies.

In short, visualization bridges the cognitive gap between data and human user. But before proposing any new visualization technique, designers should acknowledge the fundamentals of the visualization process and its determining criteria and aspects.

1.1.1 Visualization Reference Model

We can think of visualization as a pipeline from raw data to the human viewer. The reference model proposed by Card *et al.* [17] shows the process of visualization which includes three stages and several feedback loops enabled by user interaction (see Figure 1.1). Using such a model allows the readers of this thesis to locate the focus of our efforts in the visualization process. We first explain each step in Figure 1.1.

Data Transformations

Collected data are often stored as raw data barely comprehensible to human users. The first step for visualization is thus to transform the raw data into structured sets of relations known as *data tables*. In Bertin's [14] terminology each raw data record, or *case* is called an *object*. He presents *functions* as special mathematical relationships that convert objects to variables, or *characteristics* in his terminology:

	<i>Movie_A</i>	<i>Movie_B</i>	...
<i>Title</i>	<i>Vertigo</i>	<i>Titanic</i>	...
<i>Year</i>	1958	1997	...
<i>Genre</i>	<i>Thriller</i>	<i>Romance</i>	...
...

Table 1.1: A data table containing movies data. Each movie is a case and variables are Title, Year, etc.

$$f(\text{Case}_i) = \langle \text{Variable}_{ix}, \text{Variable}_{iy}, \dots \rangle$$

Table 1.1 shows the structured data table that is obtained from a transformation function. For example in a movie data set, each *case* of the resulting data table is a movie and the variables or dimensions are the properties (e.g., genre, duration, ranking in box office) of that movie.

Most common data transformation functions separate each data record into independent variables (e.g., a location record is transformed into cartesian x , y , and z variables). The transformation of raw data into data tables frequently involves the gain or loss of information. The outcome of a data transformation is either a derived data structure (i.e., aggregation), or a derived value (e.g., mean of a series of values). These two types of transformations can be chained to form more complex data transformations e.g., derivation of the mean of a category of data (aggregated values). Bertin calls this process the *aggregation cycle*. Such cycle rearrange data tables at each level of aggregation which can be used to reveal different aspects of data. User therefore controls the kind and level of data transformation based on their interest. The choice of data transformation and the desired aspects of data lead to different choices of visual structures which is determined in *visual mapping* process.

Some other works have proposed metadata integration for more efficient visualizations. Nocke and Schumann [67] present *descriptive* (underlying properties of a data set), *derived* (extracted from a data set), and *historical* (origin of a data set) as a classification of metadata. Although integrating metadata in data transformation increases the available information, it comes with some important constraints. These constraints require the need for a priori knowledge of the domain for descriptive metadata, extra pre-processing of already too big data sets to derive metadata, and information about errors and uncertainty introduced due to operations previously applied to data for historical metadata.

Visual Mapping

Visual Mapping is the processes where focus data attributes (those attributes that are intended to be visualized) are associated to visual structures perceivable by human eye. These visual structures or *marks* (such as Points, Lines, Areas, Surfaces, or Volumes) are the building blocks of any visualization and have different *properties* (such as color and size). Visual mapping and such visual properties are at the core of our works in this thesis.

The quality of the visual mapping process is determined by the characteristic of the data and visual entities on both side of the link. The data attributes come in three basic types:

- **Nominal** (i.e., categorial, can only be equal or not equal to other attributes),
- **Ordinal** (obey a $<$ relation),
- **Quantitative** (allow numerical operations to calculate statistical attributes e.g., difference, ratio, mean).

We should note that the all quantitative data do not allow for the same numerical operations. For instance, we can calculate the difference between two timestamps (date and time) while the ratio between the two is meaningless. The visual properties also should reflect the properties of data attributes. For example, color of a visual mark does not bear any natural order and thus, is suitable for representing categorial data. Also, different shades of the same color can be appropriate for ordinal data, if the difference is perceivable without too much effort by human eye. In the Chapter 2, we will discuss such visual properties (also called *visual channels*, *visual attributes*, and *visual variables*) in more details.

Data attributes can be mapped to visual structures in multiple ways. Good and expressive mappings preserve and represent all and only the desired data attributes [62]. For instance, consider the list of the world's largest companies by revenue. The name of the companies are nominal (e.g., Toyota, Walmart), their position in the list is ordinal, and their revenue is quantitative (expressed in USD). The visual mapping in Figure 1.2, is an example of inappropriate visual mapping for such data. The height of the bars which is a quantitative visual attribute is used to map the nominal data of the country of origin. Such mapping is not expressive, as it is missing the important revenue data and implies incorrect ordinal relationship among countries.

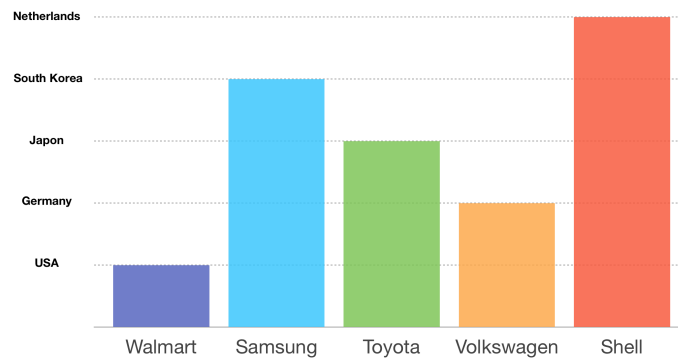


Figure 1.2: An example of unsuitable visual mapping in which, the ordinal Y axis is used to communicate nominal data attribute *Country*

View Transformations

The visual structures issued from visual mapping are barely a single view of the entire visualization where the whole researched information may not be viewable at once. View transformation exploit space and time to deliver more information and details on data. There are roughly three types of view transformations according to Card et al. [17]:

- **Location Probes** reveal supplementary information based on location. These include pop-up windows that are activated by clicking on a visual structure, or cross section planes that reveal the interiors of 3D objects.
- **Viewpoint Controls** magnify visual structures or change the viewpoint of the viewer in order to make details more visible. Sheiderman's *information seeking mantra* [79] also recommends viewpoint controls for any interactive visualization: *Overview first, zoom and filter, then details-on-demand*.
- **Distortion** deforms visual structures in order to create focus + context views. Various zoom techniques such as fish-eyes [77] and tree visualizations where nodes far from the root are proportionally shrunk [59] are examples are distortion. Distortion is effective when users can perceive the undistorted overview through the distortion [17].

Almost every such view transformations are activated by some action of the visualization's viewer. Based on their research interest, users decide to zoom or change the view to locate the information they seek. Hence, the implementation of such interactive transformations needs sufficient understanding of the users' needs and behavior.

1.1.2 Human interaction

As discussed earlier, each stage of the visualization reference model may involve some gain and loss of information. The information that user seeks may be accessible by different data and view transformations. Also, displaying the entire may not be possible or desirable due to technical and perceptual constraints. *Interaction* is the remaining part of the visualization pipeline which enables users to access and mentally combine the needed information by parametrizing each step and transformation. Interaction may stem from various intents: to mark something as interesting, to filter data and search only for particular data items, to abstract and elaborate to view less or more details, to encode and change the visual representations of data, to reconfigure and modify the views of the visualizations, and so on.

There are several means of interaction beyond standard graphical user interface controls [4]:

- **Direct Manipulation:** The main advantage of direct manipulation is that the visual feedback occurs at the imminent vicinity of the user's input. This includes point and click (or tap) selection, or zooming into detail under cursor by rotating mouse wheel.
- **Brushing and Linking:** Brushing acts like interactive queries achieved via direct manipulation. Users can highlight a selection of data by lassoing or "brushing" a subset of data items.
- **Dynamic Queries:** Dynamic queries enables users to filter out undesired data and focus on data of interest. This can be achieved by graphical widgets (e.g., sliders).

Whatever the technique of interaction, this iterative loop (See Figure 1.3) combines the computational powers of the machine and the analytical capabilities of the human user. In order to achieve a full synergy, an appropriate user interface must take human mental models into account and address computational challenges to ensure smooth execution of the interactions. A few important criteria can assess the adhesion of the visualization design to the principles of human interaction and visual expressiveness.

1.2 Visualization Criteria

In a human-centered approach, the characteristics of data and more importantly the user's goal shape the visualization design. A visualization system aims to integrate human visual perception and machine's computational power and amplify user cognition. A visualization design should thus, take a few quality criteria into account [4]:

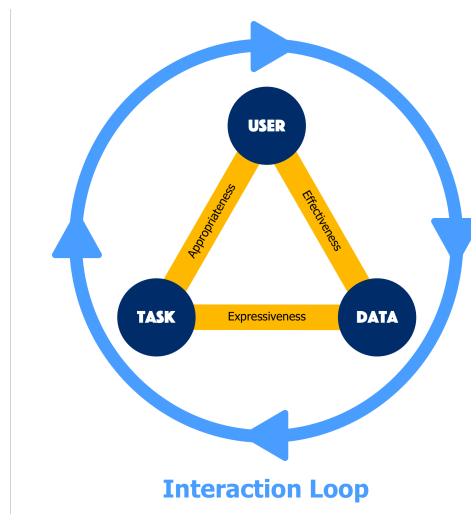


Figure 1.3: The main aspects of a visualization system and the quality criteria to consider when designing a visualization system. Interaction allows to maintain such criteria.

- *Expressiveness*: The visualization shows exactly the information contained in data
- *Effectiveness*: The visualization is compatible with human cognition, the application and the context
- *Appropriateness*: The visualization is cost effective with respect to the task and the expected goals.

Fulfilment of these quality criteria involve different underlying criteria that we discuss below. The interaction techniques that are implemented in a visualization tool also should be chosen according to these criteria. Besides the conception level, these criteria are also important for evaluation of a visualization.

Overall, a good visualization technique improves user's graphical perception of data and enhances her accuracy, performance and speed in the corresponding tasks. To ensure expressiveness, effectiveness, and appropriateness, the following factors or visualization aspects should be considered in design of visualization techniques [4].

1.2.1 Visualization Aspect: Goals and Tasks

In order to guarantee the *effectiveness* and *appropriateness* of a visualization, it is important to first acknowledge users' points of visual focus, their objectives, and the nature of the analytical tasks for achieving those goals.

Visual focus

Point vs. Local vs. Global: Human visual perception has the capacity to focus on a particular element (*point*), on groups of elements (*local*) or on an image as a whole (*global*) [74]. Point and line visualizations rely on the human Gestalt visual system [56]. Humans tend to see wholes and patterns, rather than a group of independent points. Nevertheless, in some cases user needs to extract information from a single point of data, or perform a task on a sub-group of points on a local level. For a given dataset and a specific exploring task, an effective visualization should favor the appropriate point of focus. General visualizations on the other hand, should minimize perceptual harm when the user alters her point of focus.

Objectives and Analytical Tasks

The analytical tasks that a user performs determine objectives of the visualization. A visualization's main purpose may be explorative analysis, confirmative analysis, or presentation [52].

In *exploratory analysis*, users do not assume any a priori hypothesis about data. Instead, they explore the data interactively and look for possible structures, trends, etc. Common tasks therefore involve structure change detection, cycle detection and other discovery tasks.

In contrast, in *confirmatory analysis* users assume hypothesis about data values and trends, and by performing goal-oriented tasks, they try to validate their hypotheses. Examples of such tasks include finding missing values in data, detecting if data is collected with error, comparing data point values and looking up special events.

Finally, *presentation* is the case where visualization's goal is to present already concluded facts in order to convey insights of data.

Each of these tasks engages a specific point of focus: users in tasks such as identification, i.e., looking up a data value, focus on single data points (*point focus*), while in comparison and relations seeking tasks they consider a group of data points (*local focus*). In cases where the aim is detecting overall trends and cycles in data, users consider the visualization as a whole (*global focus*).

1.2.2 Visualization Aspect: Data

The nature of the data is another aspect which crucially impacts the choice of visualization. We should distinguish two frames of reference for data: abstract and spatial data [3]. By abstract, we mean data that are not tied to a spatial context, whereas

spatial data have an inherent spatial layout. The former concerns mostly the field of information visualization (e.g., time series visualization) while the later are mostly addressed in the field of scientific visualization (i.e., volume visualization) or geographic visualization (i.e., geographic information systems - GIS). Each category requires different design approaches, since visualization of spatial data is aimed to preserve the spatial relationships between data points, while for abstract data, screen real estate can be used exclusively for conveying abstract relationships.

Visualization of the time series which are not bound to a specific spatial layout thus sits in the category of information visualization and abstract data. In the absence of any spatial mapping, the entire screen can be used to display temporal data. However, the size and ordering of data should be managed in order to obtain an efficient visualization.

Data Size Management

In any display configuration, the maximum number of data instances that could be displayed simultaneously corresponds to the number of available pixels on the screen. This means even with today's high pixel density screens, we are still constrained to a few millions of data points; and when the number of data points exceeds this limit, no visual representation can actually show all data in a single frame [49]. Even with fewer data points, visual clutter can make visualization useless, unless we control the number of displayed entities [29]. Therefore, data preprocessing and abstraction techniques are required to manage the number of displayed data points:

Data subsetting: using such techniques, we derive a subset of data from the original dataset. This can be a representative subset (by sampling) or a certain pre-fixed subset (by querying) [26].

Data segmentation: data can be divided into multiple segments based upon data attributes or data ranges [48].

Data aggregation: based upon attribute values or other properties, fewer aggregated values such as minimum, maximum, average, etc. can be constructed in order to represent the original data points. [29]

Event extraction: events are special situations in the development of time-oriented data. Events can be defined by user or be extracted by automated algorithms. In this way, information analysis is concentrated on an even higher level abstracted data which is more adequately tuned to users' goals and tasks [4].

These measures help a visualization designer to ensure the *effectiveness* of the resulting representation. Ideally, the visualization tuned regarding these criteria shows exactly the information contained in data without any missing or extra information.

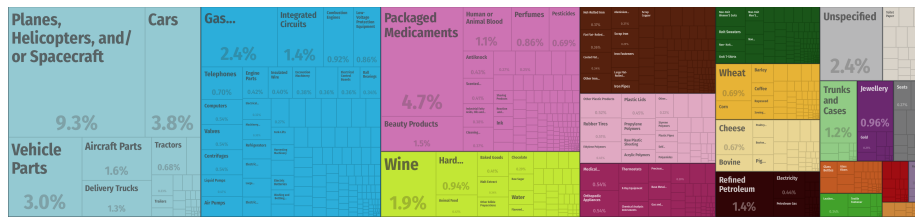


Figure 1.4: In this visualization of the exports of France, dimensions (e.g., cars, planes) are grouped by sector (e.g., transportation) to form a hierarchical view of the data. The resulting clusters are distinguishable by color. Figure adopted from *The Observatory of Economic Complexity* [80].

Dimension Management

Datasets with large numbers of dimensions (data variables) are going widespread in applications, among others, such as process monitoring, simulations and surveys. In industrial process monitoring in particular, the number of dimensions can rise to tens, even a few hundreds. We should also consider that not all dimensions share the same baseline for values or magnitude scale. Besides, each variable may be sampled at different rates and the rates can vary from milliseconds to days, depending on the characteristics of the respective sensor.

This brings a serious challenge to existing multi-dimensional visualizations: for datasets containing a few hundreds of dimensions, none of existing methods can map all dimensions at the same time without cluttering the display [88]. Dimension reduction methods (e.g. principal component analysis [7], factor analysis [37]) are aimed to find a combination of original dimensions which can explain main variations of data. Furthermore, effective dimension management techniques such as dimension ordering, spacing and filtering can ease exploration of such datasets without replacing them with non-intuitive new dimensions.

Dimension Hierarchies Ankerst et al. [8] identify the role of dimension arrangement for order-sensitive multivariate data where different arrangements of dimensions may lead to different insights of data. They define the notion of similarity between dimensions and discuss different similarity measures. They suggest rearranging dimensions by positioning those of them that exhibit similar behavior close to one another. They propose the idea of dimension hierarchies derived from similarities among dimensions. Using a dimension clustering approach, similar dimensions are tied into clusters and similar clusters into higher clusters, resulting in a hierarchical clustering of similar dimensions. In this way, a single dimension whose value is normally the average of the underlying dimensions represents the invisible dimensions of the cluster (see Figure 1.4 for an example). In order to explore all the underlying dimensions, user can traverse the resulting aggregation by interactions such as drill-down/roll-up, pan, zoom and rotation and distortion.

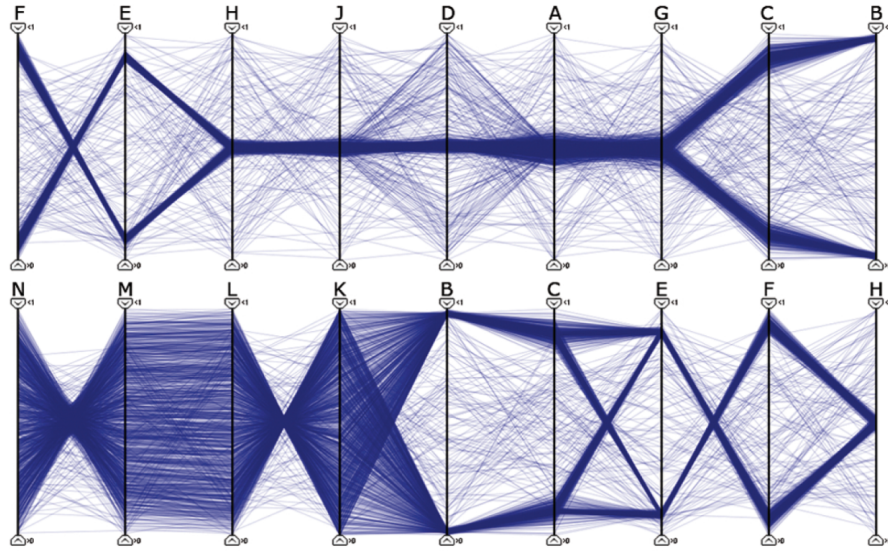


Figure 1.5: Two ordering of dimensions for the same multi-dimensional data. Ordering of dimensions in a parallel coordinate plot drastically influences cluster and correlation analysis. Figure adopted from [50]

Dimension Ordering In his book [14], Bertin illustrates the importance of dimension ordering by demonstrating how an appropriate permutation of dimensions reveal patterns and improve user’s visual understanding. In order-sensitive multivariate data visualizations, different arrangements of dimensions can reveal various aspects of the datasets. For instance, in similarity-based ordering of dimensions, those dimensions with similar patterns are positioned close to each other. This ordering enables users to find interdependencies among dimensions and reveals dominant patterns (See Figure 1.5).

Also, importance-oriented dimension ordering directs user’s attention toward more important dimensions by mapping them to more pre-attentive attributes. The importance of each dimension is decided depending user’s analytical goals [88].

However, some relationships between dimensions may stay undetected using exclusively automatic methods of dimension ordering. Given that users of a visualization often have expert knowledge of the visualized dataset, they can improve the outcomes of the automated procedure. Therefore, allowing users to interactively alter the dimension ordering is valuable. Manual rearrangement of the ordering of single dimensions or clusters of multiple similar dimensions can help users in exploratory tasks and enhance their analytic findings [68].

Dimension Spacing Dimension ordering provides users with useful information about dimension relationships. However, this information may be misleading in some cases. For example, two adjacent dimensions may be categorized in two distinct dimension clusters,

one being the ending dimension of a cluster and the other the starting dimension of the next cluster [88]. In such situation, the viewer may wrongly determine the two dimensions distant from each other. Such occurrences consequently affect users' perception of data. Dimension spacing tries to address this problem by varying the spacing between two adjacent axes or angles. This variation of spacing conveys dimension relationship information in an explicit way.

Dimension Filtering The usual dimension reduction techniques such as clustering compress the many dimensions into a few. However, the generated dimensions have little intuitive significance to users. Dimension filtering is therefore, in some way more intuitive to users as the remaining dimensions are all original dimensions of the dataset. In large dimensionality in particular, individual plots and overall trends are not visible without some kind of filtering.

Besides manual filtering, several automated filtering algorithms have been proposed [88]. However, as discussed earlier about dimension ordering, a mixture of automated algorithms and manual interactivity may be more interesting for expert users. In this manner, a dimension-filtering algorithm generates an initial filtering result, which is necessary for dealing with large dimensionality. Next, users can modify this default scheme by selecting single, multiple or clusters of dimensions to hide. This user interactivity is also necessary as they usually complement and/or improve automatic approaches.

In summary, as the number of dimensions increases, visual clutter is common and can trouble human perception of data. Several visual clutter measures have been introduced in literature [28]. Dimension reduction or management methods are then required for controlling its effects.

1.2.3 Visualization Aspect: Representation

The last aspect concerns the visual representation of time-oriented data. In this section, we discuss the fundamental sub-criteria that shape visual representation of data.

Dimensionality: 2D vs. 3D

The question of whether the use of three dimensions for visualization is useful or not has been discussed for a long time [3]. On one hand, one argument is that two dimensions are sufficient for effective data exploration. In this sense, the third dimension causes unnecessary complexity and it can also reduce analysis accuracy by occluding some data and thus, cause loss information. On the other hand, those in favor of 3D representations argue that the third dimension can encode more information into the visualization. They

try to overcome the subsequent difficulties of the third dimension by introducing novel interaction techniques.

Once again, the decision on dimensionality depends on data, user’s analytical goals, and tasks. Volume visualization for instance, requires a third dimension to reflect the 3D nature of the object, while the analysis of many abstract data do not gain significant improvement using a 3D representation.

Classification of Visualization Representations

Visualization of data which have some inherent multidimensional characteristics has been done for many years even before the introduction of computers [81]. Since the computers have been used for visualization, many existing techniques have been improved and new ones have been introduced to deal with much larger datasets and enhance user’s experience in exploring data. As specified by Keim’s [53] scheme, visualization techniques can be classified into the six categories presented in Table 1.2.

Representation aspects aim to fulfil the *effectiveness* of the visualization by guaranteeing that all information is present in data (e.g. by 3D representation of a volume). The choice of visualization technique influences the *effectiveness* of the system regarding the tasks at hand. The choice of representation technique impacts also the *appropriateness* of the visualization system which can be assessed as a cost-value ratio.

1.3 Temporality

In addition to general data-related criteria discussed earlier, temporal data possess extra characteristics that should be considered carefully in the process of visualization design. Considering time as an ordinal dimension in a 2D or 3D visualization is inadequate to express many characteristics of time-dependent information [2]. Also, there is no single correct model of time that suits all visualizations and temporal granularities [32]. The way we represent time in a visualization has important consequences on its significance and may lead to gain or loss of considerable information. Temporal primitives and temporal structure adopted from Frank’s taxonomy [32] are worth discussing in this regard.

1.3.1 Temporal Primitives: Points vs. Intervals

Data and time are related by using a set of basic elements called time primitives. These primitives can be seen as an intermediate level between time axis and data components.

Visualization Technique	Basic Idea	Notable Techniques
Geometric Projection	Visualization of geometric transformations and projections of the data	scatterplot matrices [20], landscapes [87], parallel coordinates [42]
Hierarchical	Visualization of the data using a hierarchical partitioning into subspaces	dimensional stacking [60], Treemap [78]
Icon-based	Visualization of data values as features of icons	shape-coding [11], color icons [61]
Graph-based	Visualization of large graphs using techniques to convey the meaning of the graph clearly and quickly	basic graphs (e.g. straight-line, polyline, etc.), specific graphs (e.g. symmetry-optimized and cluster-optimized graphs)
Pixel-based	each attribute value is represented by one colored pixel. For multivariate data, the attribute values for each attribute are presented in separate subwindows	recursive pattern [54], circle segment techniques [9], spiral [51]
Hybrid methods	Integrated use of multiple techniques to enhance the expressiveness of the visualization	IVEE [79] (combination of geometric, icon-based, and dynamic techniques), XmDv [75] (combination of geometric, icon-based, hierarchical and dynamic techniques)

Table 1.2: Main categories of visualization techniques according to Keim's [53] scheme with the most well known examples in bold type.

Depending on the desired application, time axis can be represented using time points or time intervals [4].

Time points are absolute time primitives, i.e. they are fixed instants along time scale. For example, data involving events during time is associated to time domain by means of time instants. On the contrary, an interval is a time primitive with an extent. It can be specified by two instants marking the beginning and the end of the interval, or by a starting point and a given duration of time. Rather than separate events in time, an interval represents the state of a variable during its length.

As shown in Figure 1.6, different relations are possible among time points and among time intervals [4]. Therefore, the choice of time primitives is important in accomplishing different analysis tasks and data exploration goals. Also, the selected time primitive affects the validity of data. For example, using time points, data values are only valid at given instants and nothing can be said about the values between two adjacent time points. Therefore, the duration for which the data values are valid should be visualized to avoid misinterpretation.

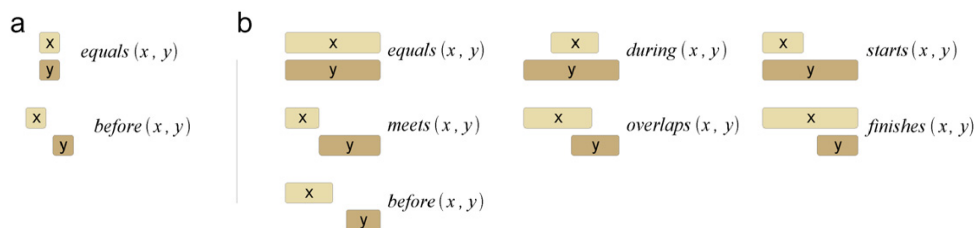


Figure 1.6: Temporal relations: (a) relations between time points; (b) relations between time intervals; respective inverse relations are possible as well. Intervals in contrast with points, can overlap to different degrees. Figure adapted from [4].

1.3.2 Temporal Structure: Linear vs. Cyclic

We naturally perceive time as proceeding linearly from past to future. A linear structure of time, thus, considers time axis as a collection of time primitives, one after the other. However, periodicity is very common in many types of data (e.g. the months of the year). On a cyclic time axis, a primitive is preceded and succeeded at the same time by any other primitive (e.g. September precedes January, but September also succeeds January). A cyclic arrangement of the time axis can incorporate temporal periodicity into the visualization and support identification of recursive structures in data.

The choice of time structure depends on the users's analytical goals. If for instance, a user aims to find a general trend in data, a linear arrangement of time axis should be sufficient. While if the user seeks to detect monthly effects on data, a cyclic temporal structure is more convenient [4].

1.4 Space Efficiency Management Techniques

Screen size is a visualization criterion that is directly correlated with the amount of data and thus, the information that can be displayed at once [86]. This makes screen size a physical limit to the amount of information that can be communicated through a visualization and marks the importance of space efficiency in visualization techniques.

Space efficiency is one of the characteristics of the visual designs studied in this thesis. We therefore, review the main approaches for improving space efficiency and reducing visual clutter. Ellis et al. [28] classify these approaches into three groups:

- **Appearance:** This group includes techniques that control how much data appear on the screen by sampling (e.g. [26]), filtering (e.g. [1]) changing marks' size or opacity (e.g. [12]), and clustering (e.g. [33]);
- **Spatial distortion:** This can be achieved by displacing marks (e.g. [55]), topological distortion (e.g. zooming and Fish-eye lenses), space-filling (e.g. Tree-Maps [78]), and pixel-oriented techniques (e.g. [53]);
- **Temporal:** This category refers to animation techniques (e.g. [25]) which display different data within a temporal interval.

In this thesis, we are interested in space efficient designs for time series visualizations with the most familiar form, i.e., horizontal charts and we focus our work on improvements on visual mapping of data by means of combinations of visual channels. In this regard, our approach is more close to space-filling and pixel-oriented techniques.

1.5 Summary

In this chapter, we presented the scientific context of this thesis. We demonstrated the role of user interaction in this process and we detailed visualization aspects related to user, data, and representation in order to show a glimpse of possible iterations in visual design. Also, in this thesis we concentrate on temporal data and so this chapter introduced the specific characteristics of the temporal axis.

We defined the most important concepts in visualization in order to better understand the theoretical and empirical work detailed in the following chapters. We detailed the visualization reference model and pointed to the process of visual mapping as the focus of the current thesis. Because of the importance of this step, we dedicate the next chapter to elaborate the visual channels and the existing approaches to visual mapping.

Chapter 2

Visual Channels & Their Conjunctions

In chapter 1, we introduced the visualization reference model. In this model, *visual mapping* relates data side and visual side of a visualization and controls how each data attribute is linked to visual structures. Visual mapping determines which properties of visual structures are used to convey information.

The bridging nature of the visual mapping process makes it sensitive both to characteristics of data and to human visual perception. It is hence important to identify the building blocks of visual mapping and the properties of the visual structures.

The notions of such visual properties, their combinations, and the resulting visual mappings are at the core of the visual designs presented in this study. Hence, we decided to dedicate this chapter to presentation of the visual channels used throughout our work, as well as the existing approaches to visual mapping in literature of time series visualization.

2.1 Marks and Visual Channels

Marks are the basic visual structures of visualizations. The elementary marks consist in *Points* (0 Dimension), *Lines* (1 Dimension), *Surfaces* (2 Dimensions), and *Volumes* (3 dimensions) [14]. Each visual mark can be displayed in different positions and various modes. The properties that describe the position and appearance of visual marks are called *visual variables*, also known as *visual channels* and *visual attributes*. For the sake of consistency, we will use the term *visual channel* in the rest of this thesis to refer to the properties of visual marks. This choice of terminology was made in analogy with



Figure 2.1: Visual channel *Shape*, for instance, is selective and associative, but it is neither qualitative, nor ordered. Theoretically, indefinite number of sign shapes are conceivable.

the concept of channels in telecommunications and computer networking. Similar to a communication channels, a visual attribute acts as a medium to convey information signal from the display to viewer's eye .

Furthermore, Bertin defines perceptual characteristics for visual channels which help to determine if a visual channel is suitable for representing a given aspect of data. The five visual channel characteristics according to Bertin are [14]:

- **Selective:** If a mark change in this channel can be perceived selectively in presence of other marks and channels,
- **Associative:** If marks can be grouped as a whole according to this channel,
- **Quantitative:** If a numerical reading of the channel is possible,
- **Order:** If changes in channel are perceived as ordered,
- **Length:** How many separate values can be perceived using this channel.

The choice of a proper visual channels for each data attribute is made based on the such criteria with regards to the characteristics of the data and the tasks and goals of visual analysis. For examples, consider the visual channel *Shape* as presented in Figure 2.1. A change in shape is associative as a mark with particular shape can mostly be perceived selectively in the presence of other shapes. It is also more or less associative as the viewer can group the same shapes together to form a "family" or cluster. Shape however is not quantitative as no numerical reading of the channel is possible. It does not assert any order neither. Marks can support numerous distinct shapes and thus the length of the variable *shape* is theoretically infinite. Hence, shape is not an appropriate visual channel for communicating quantitative values. However, it may be used for categorical data which do not inhibit a natural order.

2.2 List of Visual Channels

The first list of visual channels proposed by Bertin [14] included seven visual channels. The list of visual channels was later extended by Mackinlay [62] and Chen [36] to include 13 and more than 30 visual channels, respectively. However, a visual design space cannot be analyzed without setting its limits in advance. This study does not aim to include all visual channels. Instead, we choose Bertin’s list of visual channels with some modification as the basis of our study. Bertin’s list of visual channels includes the most basic channels and most additional channels can be considered as abstraction of Bertin’s basic channels. For instance, *angle* is not genuinely distinct from Bertin’s *position*.

As discussed in earlier, the visual channel shape is not quantitative and thus, not appropriate for quantitative data such as time series. Our focus on quantitative characteristics of data persuaded us to exclude shape and other glyph-related visual channels from the scope of our study.

In line with Bertin, visual motion (animation) is not included in our list of visual channels. Animation inherently alters the perception of time in visualization, which seems critical for visualization of time series. While dynamic representation of time (i.e. animation) may facilitate discovery of some patterns through time, it deprives users from the ability of users to read data values at more than one point in time. Such ability is essential for comparing and estimation tasks. Also, we focus on quantitative data (i.e. time series) where these tasks are quite common. For these reasons, we excluded animation from the scope of our study.

As explained in Chapter 1, in addition to the introduced occlusions and perspective distortions, three dimensional display of abstract data (i.e. time series in our application) do not lead to significant perceptual gains. We have therefore excluded three dimensional visualization from the scope of our study in line with Bertin.

The following are the visual channels most relevant to the visualization of time series.

2.2.1 Position

Planar position is the visual channels that satisfies all four perceptual characteristics (See Figure 2.2). It is selective, i.e. similar marks varying only in position on the plane can be perceived as different. It is also associative, i.e. similar marks differing in position on the plane can be perceived as similar and in the same group of marks.

The plane also allows for ordering of marks based on their position. The order however has a direction which can be defined relative to one axis (X or Y) or as rotation around a given reference point, or as marks run along a straight line. Position is also



Figure 2.2: Visual channel planar Position, is selective and associative, as each of the marks A , B , and C can be perceived selectively or together as a group of similar marks. They also inherit an order with two possible directions: A, B, C or C, B, A . It is also quantitative, numerical values can be obtained based on the coordinated of the marks (left), their lengths (middle), or their angles relative to a reference point (right).

quantitative. It permits numerical calculation based on the X and Y coordinated of a mark, as well as its length and angle (for lines) and area (for surfaces).

The two planar dimensions X and Y can be used in different ways or *impositions* in Bertin's terminology [14]. He classifies two dimensional representations in four impositions primarily based on the nature of the expressed correspondences:

- Diagrams
- Networks
- Maps
- Symbols

Diagrams allow for the correspondance between all the divisions of one component of the plane and all the divisions of another component. In Figure 2.3a, any price value can be related to any date and vice versa. Two dates or two prices however do not need to be correlated. Whereas networks allow establishing correspondances between the divisions or individuals of the same component. Maps are similar to networks while correspondances are established according to geographical order. Hence, the correspondances produce a geographical map which cannot be reordered arbitrarily. In all of the aforementioned types of impositions, any correspondence is established between the components present on the plane. However, in a symbol the correspondence is between one component on the plane and the user (exterior to the graphic) via a sign system. symbols are thus, conventional and limited to their ecosystem (e.g. washing instructions, railroad system).

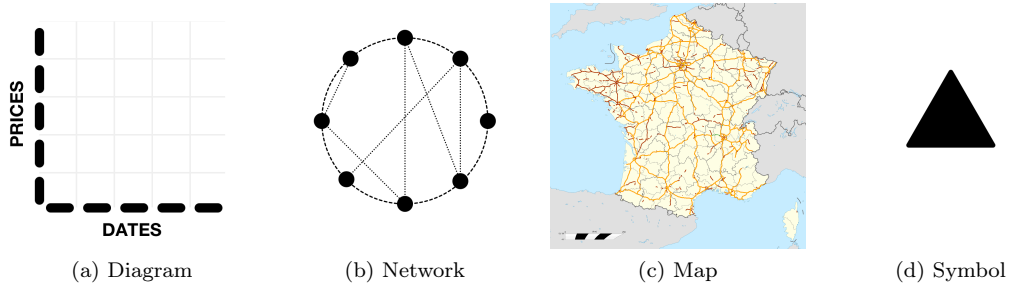


Figure 2.3: (a) Presentation of time series as diagrams allows for correlation between any timestamp (dates) and any quantity (prices) and vice versa. (b) Networks allow for displaying the correspondances between the individuals of a group (a component). (c) Maps are networks constrained by a geographical ordering. Example adapted from [22]. (d) A simple triangle symbol can be interpreted as a mountain, an increase of value, etc. based on the context and conventions.

Visualization of quantitative temporal data such as time series, which is the focus of this study, needs at least two components: one temporal component and one numerical component. Hence, networks which suit for representing the relationships between the individuals of the same component is not appropriate for our case. Also, such abstract data are not bound to any geographical and spatial arrangement and this excludes the study of maps from our scope. While other types of impositions try to reduce the representation to its essential elements, symbols like language aim to achieve immediate recognition based on a learned signification system or convention.

The imposition that corresponds best to the characteristics of time series visualization is thus, *Diagrams*. Diagrams allow for displaying both quantitative and temporal components and the correspondances between their individuals. This allows user to easily recover information about temporal instances of a given value or otherwise, read the values at a given timestamp.

With diagrams, different constructions are possible: *Rectilinear*, *Orthogonal*, and *Circular*, *Polar* (see Figure 2.4). We should note that no single representation is suitable for all analytical goals. For instance, rectilinear and orthogonal representations (e.g., simple line charts) are widely used for visualizing time series data. However, they only facilitate specific tasks such as finding the minima and maxima of data and observing of the overall behavior of data. For other tasks, e.g. detecting trends and temporal patterns, circular representations may be more helpful. With right (interactive) parametrization, circular representation reveals recurring patterns and events (See Figure 2.5).

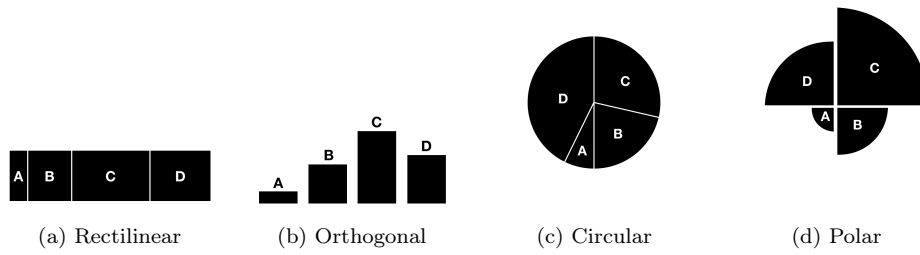


Figure 2.4: Different types of imposition for diagrams.

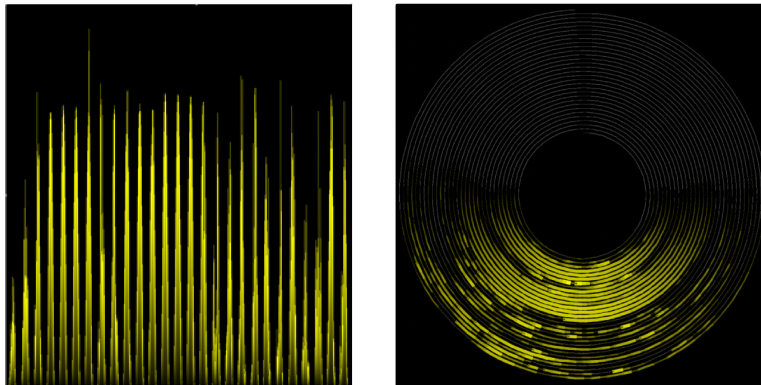


Figure 2.5: Line chart (left) and spiral (right) representations of the same sunshine intensity data. The spiral visualization is more suitable to compare days, detect cloudy periods, and spot sunrises/sunsets. Figures adapted from [85].

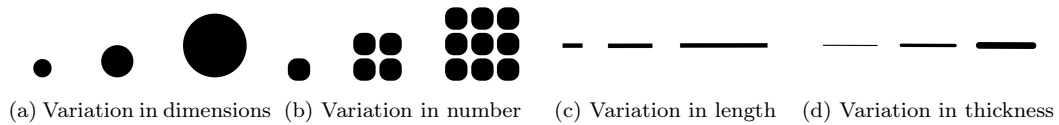


Figure 2.6: Variation in size is possible in multiple ways.

2.2.2 Size

Any visual mark can vary in size without changing in position, color, or any other visual property. A change in a mark's size is the result of the variation of its dimensions. In point representation the change in size is perceived as a change in the area occupied by that mark. It can also be achieved by multiplication of the number of the mark's identical constituent parts (see Figure 2.6b). In case of linear representation, the change in size can manifest in form of the varying thickness or length of lines. In area and volume representations, the variation in size is achieved by changing the dimensions of the surface or the volume.

The number of possible steps for size is theoretically unlimited. However, the human eye cannot distinguish more than twenty steps between two marks whose areas respectively are 1 and 10 units [14]. With selective perception, size is even more limited. Bertin advises not to use more than four or five size steps when the viewer intends to isolate marks solely based on size. Size is dissociative as the viewer cannot completely disregard it while other visual properties are present. The size of a mark can dominate other visual properties to a degree that for instance, the color of a very small mark becomes imperceptible. Size is ordinal: marks varying only in size can be ordered from small to large (or from large to small).

Color

Color is not essential for perceiving form and size of objects. However, it can be the only property to discriminate similar objects. This makes color extremely useful in data visualization. Nevertheless, its use has important limits. Besides involuntary differences in color reproduction on different screens, all humans do not perceive the colors in the same way: near 10% of the male population and 1% of the female population have some kind of color deficiency [84]. Neglecting this fact can lead to impaired perception of information for a large portion of visualization users.

Human perception of color does not correspond to the physiological properties of human eye (Three color receptors, i.e. cones). For this reason, perceptually-based color spaces reflect color vision better than color spaces such as RGB (Red-Green-Blue). A popular transformation of RGB color space is HSV (Hue-Saturation-Value). However,

HSV has a major drawback: the value (lightness) of color is not uniform across hues and saturation. Hence, HSV poorly corresponds to perceptual axes.

To overcome these perceptual drawbacks, *Commission Internationale de l'Éclairage* (CIE) introduced CIELUV color space in 1976. Polar transformation of CIELUV space allows to obtain HCL space with perceptual dimensions [41]: Hue (the dominant wavelength), Chroma (intensity of color as compared to gray), and Lightness (sensation according to which a color emits more or less light). HCL color space has been adopted by information visualization practitioners, notably ColorBrewer [38], for its accord with human perception. For the same reason, we choose HCL's hue, chroma, and lightness as color related visual channels in our study.

2.2.3 Hue

The typical human eye can perceive more than 10,000 different colors. However, cross-cultural studies strongly support the idea that certain colors, specifically, red, green, yellow, and blue, are far more valuable for information representation than others [13]. *Hue* can be interpreted as the degree to which a color can be described as similar to or different from the pure or unique colors, i.e., red, green, blue, and yellow. In Bertin's words, "hue variation is simply the perceptible difference which can be perceived between uniform uniform areas having the same value [14].

Distinct color hue may be arranged by their wavelength in the visible spectrum, otherwise, they do not inherit any natural order. Hence, color hue is selective and is more suitable for representing categories of data. Hue is also associative as the marks with the same color can be perceived as a group without major effort. Figure 2.7a shows hue variation at constant value and saturation.

2.2.4 Chroma

When asked to describe a color, people generally use adjectives such as *vivid*, *faint*, *intense*, or *colorful*. Technically speaking, colorfulness is the perceived quantity of hue content (difference from gray) in a stimulus [30]. However, colorfulness increases with luminance. In order to separate these two characteristics of color, HCL color system uses *Chroma* which, as defined by International Commission on Illumination (CIE), the colorfulness relative to the brightness of a similarly illuminated white [30].

In this way, the perceptual attributes *hue*, *chroma*, and *Lightness* are adequate for describing colors in typical environments. Other attributes, i.e. *saturation*, *brightness* can be derived from these attributes and their inclusion is thus, redundant. Figure 2.7b shows chroma variation for a certain green hue in constant lightness.



(a) Hue variation at constant chroma and lightness



(b) Chroma variation at constant hue and lightness



(c) Lightness variation at constant hue and chroma

Figure 2.7: Variation of hue, color, and lightness, respectively, while keeping other attributes constant.

2.2.5 Lightness

Lightness or *value* is defined as the brightness of a stimulus relative to the brightness of a similarly illuminated white, i.e., a reference point which appears as white in the same viewing conditions. *Brightness* in this regard, is the attribute of a visual sensation according to which an area appears to emit more or less light [30].

A color with low lightness, or value, is nearly black, and a color with high lightness, or value, appears white. The transition from black to white is independent of the color (hue). It can be a succession of grays, blues, reds, or any other color. Lightness variation is hence, ordered from black to white, or from white to black. Lightness is selective, however, Bertin advises to restrict the number of value steps to six or seven – black and white included [14]. It is dissociative, as it can impact the perception of other characteristics of the marks: for instance at high lightness, i.e., white, the size of a mark is not distinguishable on white background. Figure 2.7c shows variation of lightness at constant color hue and chroma.

2.2.6 Texture and Pattern

It is common in the literature to use the terms *texture* and *pattern* interchangeably. Although being closely related, it is important to distinguish the two concepts.

A *Pattern* is produced by orderly repetition of a mark. A variation of pattern can be obtained by changing the size, the arrangement, or the shape of the marks (see Figure 2.8). Texture variation on the other hand, is the effect created by photographic reductions or a sort of *zooming* of a pattern of marks [14]. In this way, the constituent marks of the pattern conserve their relative sizes and distances and thus, the number of marks per area remains constant. Therefore, texture variation, in contrast to pattern variation, does not vary the value of the area. In Bertin's words, *at a given value, the texture is the number of separable marks contained in a unitary area*. Figure 2.8 shows some examples of pattern from top to bottom, while showing also variation of texture from left to right, corresponding to variation from *fine* to *coarse*.

The change of value due to pattern variation is obvious in Figure 2.8. For this reason, pattern variation and value are not separable in practice and should be used together cautiously. Another point to consider when dealing with patterns and texture is the spatial frequency, i.e., the rate of repetition of the marks. Look at Figure 2.9 for instance. The high rate of repetition of black and white stripes creates an uncomfortable sensation for most of the viewers. This "vibratory" effect of texture seems to result from a some resonance at retinal level [14]. This effect therefore, should be considered when using texture and pattern in visualization.

2.2.7 Classification of Visual Channels

The visual channels position, hue, chroma, lightness, texture, and size detailed in Section 2.2 will be the channels we study throughout this study. It is important to take characteristics of each channel into account for encoding data. Table 2.1 summarizes the characteristics of the presented channels regarding to the criteria defined by Bertin.

In addition, these visual channels can be categorized by their nature into geometric channels and optical channels [19]. In this way, planar position, size (length, area, volume), and similar channels which are related to the position of marks and their occupied space on the screen are included in geometric channels. On the other hand, visual channels such as color hue, chroma, and brightness depend on the *optical* properties of the marks and hence, are inherently different from geometric channels. This categorization, however, is not strict and some visual channels, e.g. texture, locate on the border of the two categories. The variation in texture is made by changes in geometrical properties of the constructing pattern, however, it creates an optical effect.

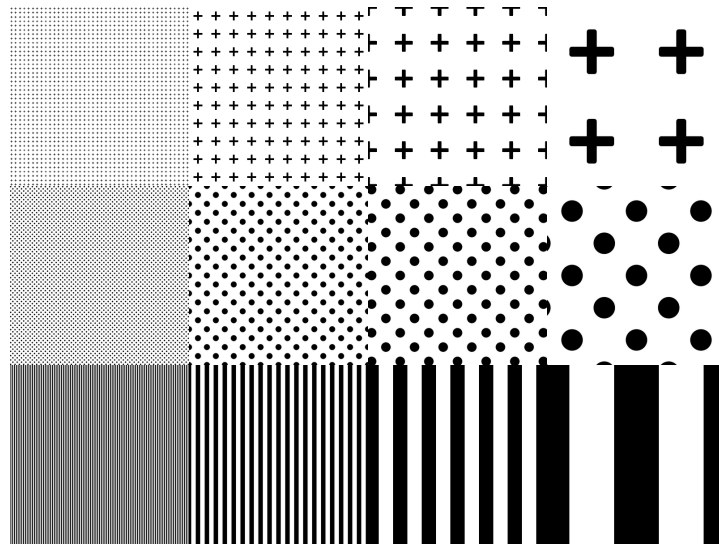


Figure 2.8: This figure shows three patterns (corresponding to the rows) and their texture variations (varying horizontally). Here, the value (the black to white ratio) increases by varying pattern (from top to bottom). By definition, it remains constant when varying texture (from left to right).

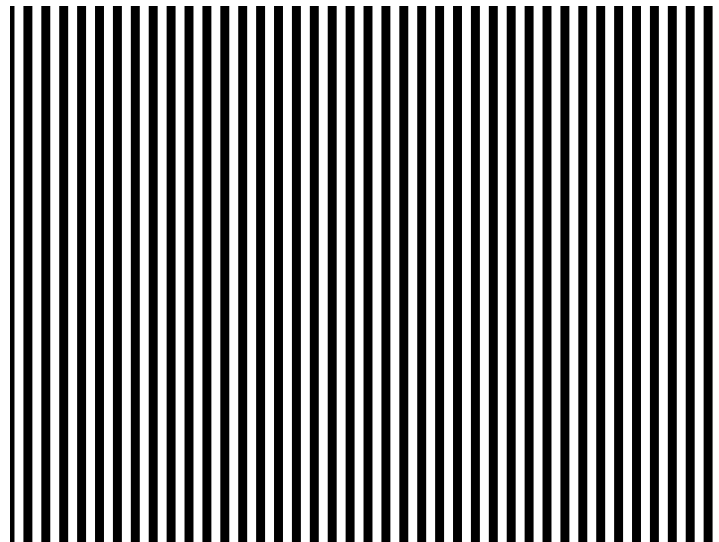


Figure 2.9: Repetition of black and white stripes can induce an uncomfortable vibratory effect.

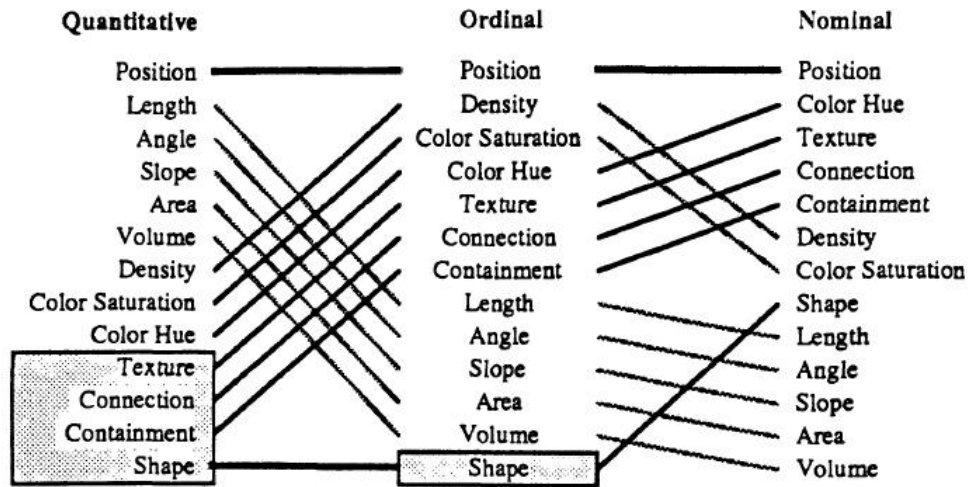


Figure 2.10: Ranking of visual channels based on type of the searched information and perceptual task. Figure adapted from [62].

Such differences in characteristics of visual channels lead to varied performance in perceptual tasks. Mackinlay [62] provides a sorting of visual channels based on the type of the perceptual task (see Figure 2.10). In this ranking, *position* is superior to all other visual channels in every perceptual type task. Interestingly, geometric channels, i.e., position (and consequently, angle and slope) and size (length, area, and volume) are the best visual channels for conveying quantitative information. Optical channels on the other hand, are better channels for communicating ordinal and nominal information in general. These disparities are of great importance when choosing a proper visual mapping and visual design.

2.3 Visual Channels Conjunctions

Visual channels described in Section 2.2 can be combined in various ways to form multi-channel or multi-variate visual mappings. In this section, we present the existing multi-variate visual mappings in order to illustrate the different approaches.

2.3.1 Bivariate Maps

Maps and spatial data visualization are situated on the borderline between the scientific visualization and information visualization. Although spatial data is not the focus of this thesis, it is worth to review bivariate maps as an example of combination of visual channels which has been widely implemented and evaluated in the literature. Bivariate

Visual Channel	Selective	Associative	Quantitative	Ordered	Length
Position	●	●	●	●	Theoretically infinite, constrained by space on the screen
Size	●	○ size dominates most other channels	●	●	Theoretically infinite, constrained by space on the screen
Hue	●	●	○	○ There is no natural ordering for color hues	Theoretically infinite, practically limited
Chroma	●	○ it can interfere with other optical properties	○	●	Theoretically infinite, practically limited
Lightness	● six or seven steps at maximum	○ Lightness dominates other color attributes	○	●	Theoretically infinite, practically limited
Texture	●	○ it can interfere with color perception	○	●	Theoretically infinite, practically limited to available resolution

Table 2.1: Recapitulation of the characteristics of the visual channels detailed in Section 2.2. Filled and half-filled circles indicate full and limited presence of the characteristic, respectively.

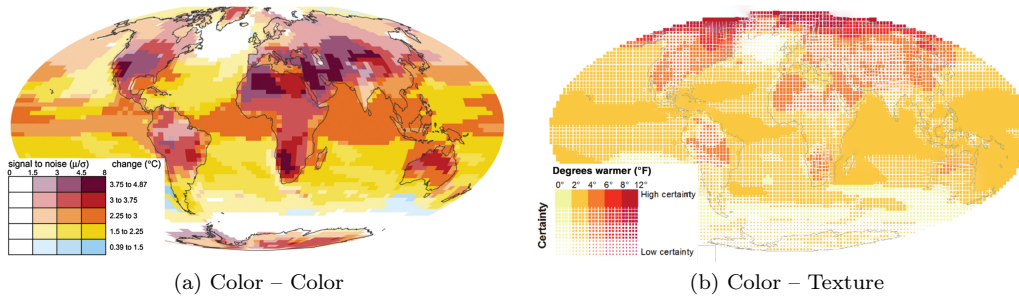


Figure 2.11: (a) using two color channels is more suitable when interaction between uncertainty and value reading is desired. (b) using texture along with color keeps values and uncertainty more separable.

maps are used to display two data variables over the same map area. In geospatial visualizations, i.e. maps, the position channel is reserved for geographical coordinates and therefore, is not available for visual mapping of other data values. This constraint requires the map designers to use one or several other visual channels in the same position. The literature in cartography has examined how the combination of color and texture in maps affect users' abilities to attend each visual channel separately or to understand the interaction between them [66]. For instance, climatological maps of observed/projected temperature changes use coincident bivariate maps to communicate information on the temperature change and its certainty via two visual channels (See Figure 2.11).

The combination of visual channels however depends on the map maker's goals: studies [73] have shown that if the goal is to easily separate the values (temperature) and uncertainty, color for the former and texture for the latter seems more suitable. Meanwhile, it may also be desirable to obscure the main value where the certainty is low. In such cases, color hue for the main value and saturation/value for certainty is an example of appropriate combination in order to introduce a fading effect.

In contrast to bivariate maps in which two visual channels are used for two data dimensions sharing the same space, other cartographers have examined the idea of redundant mapping for map symbols. Dobson [27], for instance, showed that adding color value as well as the size of a map symbol improved map interpretation. The greater the quantity represented by the symbol, the larger and the darker the symbol was. Such redundant mapping increased users' accuracy and speed compared to univariate mappings. This was a surprising finding, as selective attention studies indicate that size and value are separable visual channels.

The empirical findings in cartography and the advantages of combination of visual channels motivate us to look for their correspondences in time series visualization. In addition, the categorization of visual channels into separable, integral, and configural and the empirical evaluation of their interactions in cartography lay a foundation for our

proposed visual designs for time series.

2.3.2 Scale Manipulation Techniques

When a dataset covers a large range of magnitudes which is far greater than the the number of available pixels, quantization error is inevitable. Many quantities thus, share the same visual representation. In such cases, a single pixel can represent a large number of values with considerable quantitative differences. There exist several solutions to this problem in the literature which are mostly based on distortion of the position visual channel. The distorted position channel is sometimes accompanied with additional visual channels like color (see Figure 2.12e).

Re-expression of data

One of the most common solutions for representing numbers with varying magnitudes is to re-express them using power functions known as Tukey power ladder [82]. These functions include powers such as -1 , $-\frac{1}{2}$, \log , $+\frac{1}{2}$, $+1$ or higher. In this way, exponential changes are shown as linear when plotted on logarithmic axis. The inconvenience of such transformations is that it makes it difficult to read and compare numerical values.

Dual scale charts

In their study on dual-scale data charts [43], Isenberg et al. survey the existing dual-scale charts. They identify bifocal charts with two regions of different scales (Figure 2.12b), lens charts similar to bifocal charts with addition of a drop-off between the two scale regions (Figure 2.12c), broken charts similar to bifocal charts with addition of a visible gap between the two scale regions (Figure 2.12d), superimposed charts where focus and context charts share the same area while using two different scales on the common axis (Figure 2.12e) and finally, cut-out charts where a dedicated area displays focus the region separate from the context (Figure 2.12f).

2.3.3 LifeLines & MIDGAARD

Visualization of multi-dimensional and time-oriented data is commonplace in medical applications for monitoring different health indicators of patients during time. One of the visualization systems close to our work is *LifeLines* [70], where colored horizontal bars plotted on timelines represent an assigned symptom or action. The length of a bar is derived from the duration of the condition or event they represent. The thickness of the bar represents the significance of that action, and its color is used to visualize

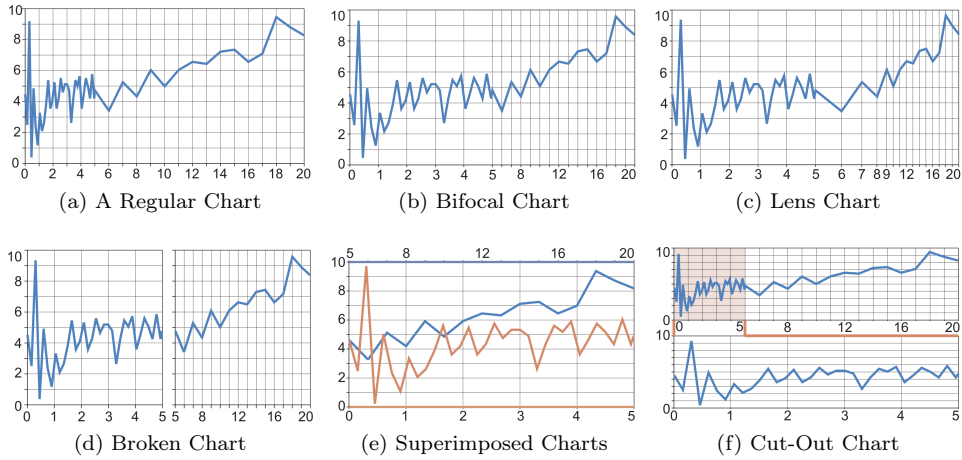


Figure 2.12: Different approaches to dual scale charts as surveyed by Isenberg et al. Figures adapted from [43].

relationships. While *LifeLines* can display presence of a symptom (e.g., fever during a time period), it is not able to visualize the quantitative variations, i.e., temperatures, during that period (see Figure 2.13).

In *Midgaard* project, Bade et al. [10] introduce a hybrid technique built on top of the qualitative approach of *LifeLines*. In addition to the qualitative approach of *LifeLines* which allows to visualize abstract attributes of data, i.e., severity of an event, *Midgaard* displays quantitative data, i.e., temperatures, in form of a line chart (see Figure 2.14a). In both techniques, however, the abstract attribute which is mapped to the additional visual channel (i.e., color) is derived based on an a priori knowledge of data, e.g., severity of a symptom, or critical levels of fever temperature, and were not directly calculated from data.

Another technique presented in *Midgaard* project, called *Extended Information Mural* aims to address the limits of data curves in visualization of high-frequency data. This technique uses slim Tukey box-plots [81] which display minima, maxima, the median, and the 20% and 75% percentile. Connecting these box-plots generates a data representation as in Figure 2.14b. The space between minima and maxima is filled by lighter color, and the space between the quartiles are filled with darker color. In this way, additional information (i.e., statistical measures) is derived directly from data and displayed using variation in color value. This representation allows for identification of the global trend of data as well as the variability of the values.

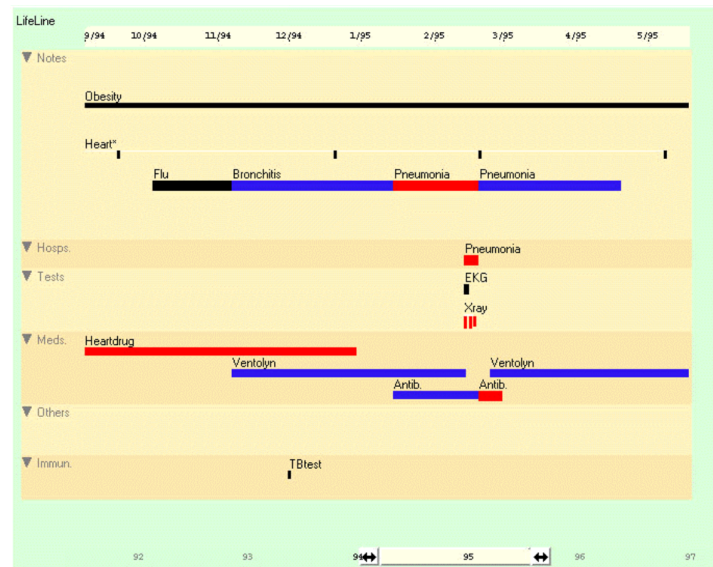


Figure 2.13: *LifeLines* [70] use additional visual channels such as the thickness of the bars and their color to display qualitative data. The mapping is determined by a priori knowledge of the context, e.g. critical levels of symptoms. Figure adapted from [70].

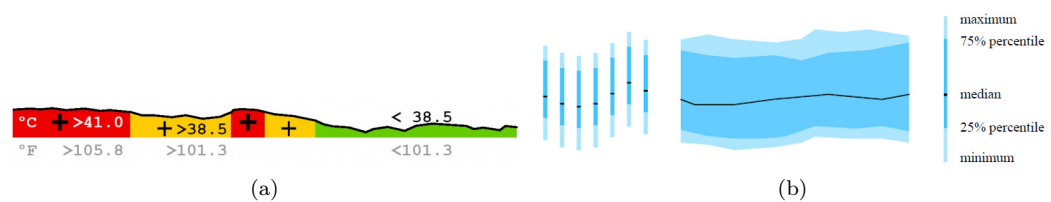


Figure 2.14: Midgaard project introduces visualization techniques which use additional visual channels to map derived data attributes: (a) the derived attribute is knowledge based and mapped to color hue, (b) the derived attributes are statistical measures directly obtained from data and mapped to color value. Figure adapted from [10].



Figure 2.15: Using color creates a “pop up” effect which attracts viewer’s attention to the details (here, local maxima). Figures adapted from [58].

2.3.4 Colorfield techniques

There exist several other time series visualization systems which benefit from conjunction of visual channels. Lam et al. [58] proposes a visualization system which incorporates multiple visual information resolutions. To achieve this aim, the system uses two visual mappings for low and high resolution views. Both interfaces use the x axis in the same way, i.e., dedicated to time dimension. The data values are mapped differently: In the low resolution view (Figure 2.15a), the values are mapped with color (saturation and brightness in HSB color space), while in the high resolution view (Figure 2.15b), they are redundantly mapped with both color and vertical position. This way of using color exploits visual attention to create a *pop out* phenomenon which allows users to quickly find details.

Similarly, Corell et al. [23] use color encoding to draw on perceptual phenomena of pre-attentive and peripheral processing. They show through empirical experiments that color mapping significantly improves participants’ perceptual averaging of the values (see Figure 2.16). However, they come short of evaluating a combination of the line chart and their colorfield.

Later, Albers et al. [5] conduct a task-driven study of several visual designs. In their evaluation, they compare two groups of visual designs: the first group includes position-based mappings such as line chart and its variations (with addition of sampled maxima and minima or overlaid average chart) or and box plot; the second group includes color-based mappings similar to the designs in [23].

The results shows that position-based mappings outperform color-based mappings in point comparison tasks, i.e., finding maxima, minima, and range of values. They also partially validated the hypothesis that color-based mappings perform better in summary comparison tasks, i.e. finding average, spread, and outliers. the authors conclude that no one mapping dominates in every task.

2.3.5 CloudLines

In *CloudLines*, Krstajic et al. [57] propose a technique for visualization of high volume event-based time series in limited space. In order to reduce the over-plotting caused by

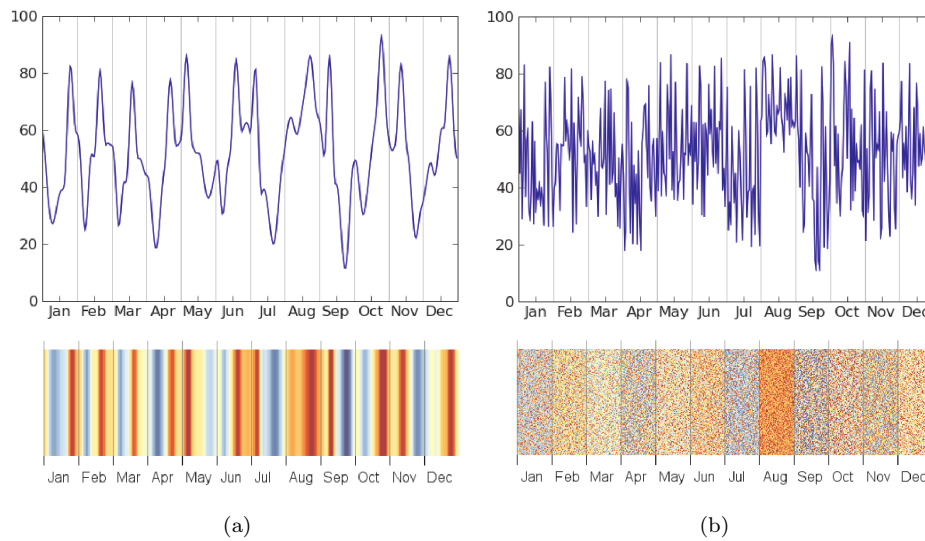


Figure 2.16: (a) Usine color mapping significantly eases mental averaging of the values. (b) The effect is more dominant in the presence of noise. In both cases, blue color corresponds to lower, and red to higher values. Figures adapted from [23].

the high number of events, visualization systems often use aggregation methods, e.g., sampling, which involves involuntary information loss. CloudLines instead, implements visual mapping enhancements to address this problem. CloudLine uses maps the *importance* of each event onto the opacity (brightness on white background) and the size of the mark representing the event (see Figure 2.17). The importance of each event is a function of its age and a density factor of the area where the event has occurred. This approach to visual mapping allows for easier identification of event sequences or episodes.

2.3.6 Order of Magnitude Markers

Order of Magnitude Markers (OOMMs) is a visualization technique for representing numerical data [15]. The motivation behind OOMMs is to present data with large dynamic range. OOMM uses logarithmic transformation:

$$v = \text{sgn}(v) \cdot (b \times 10^a)$$

to divide each data attribute into sign of the data attribute's value (negative or positive), coefficient b , and exponent a . We describe such decomposition functions in more detail in Chapter 3. Each of these components are then respectively mapped onto hue, height of the coefficient bar, and height of the exponent bar. The two bars are discriminated by their color saturation (Figure 2.18).



Figure 2.17: Step-by-step construction of a CloudLine. (*top*) An event-based time series on a linear timeline, it is practically impossible to distinguish high and low density areas. (*middle*) the calculated importance is mapped to the opacity of the marks. It is still impossible to see the details in high density regions. (*bottom*) Importance redundantly mapped to both opacity and the size of the marks which allows for identification of recent event episodes. Figure adapted from [57].

Although Order of Magnitude Markers facilitates representation of discrete numerical data in the form of bar charts, it is not suitable for visualization larger temporal data due to its modest space management.

2.3.7 Slick Graphs

Slick Graph is a smoothing technique which aims to make time series visualizations more readable without losing information [24]. In other smoothing techniques [16], high frequencies' information is lost in the process of smoothing. Slick graphs on the other hand, present the whole information to the user while preserving the benefits of the smoothing process. Compared to traditional smoothing techniques, Slick Graph has shown its superiority in various comparison tasks.

Slick graphs uses a frequency separation function to divide data attributes into high frequency and low frequency components. Low frequency components are mapped onto the vertical position (the line) and high frequency components are mapped onto the brightness (color value) of the area under the line (Figure 2.19). This separation of the frequency components allows users to view the trend in data (component of interest) while preserving noise data in the visualization. The noise component may be used by viewers to assess the certainty of their data readings.

Alternative designs for slick graph may use other ordinal retinal channels (e.g., color saturation) instead of gray scale. Such designs need further studies in order to compare their performance with grayscale.

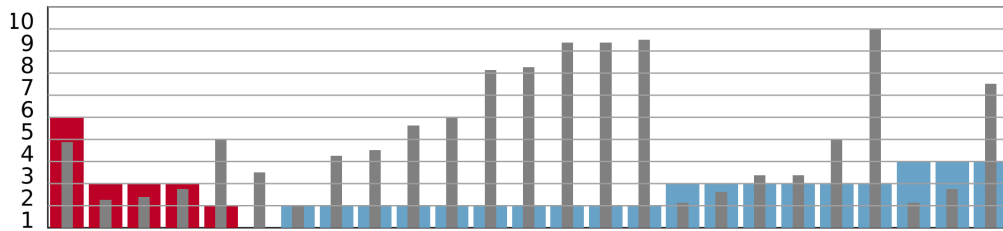


Figure 2.18: In Order of Magnitude Markers each data attribute is divided into a coefficient and a exponent of the logarithmic expression. Exponents are represented by colored bars and coefficients by grey ones. The sign of each data value is mapped onto hue (red/blue) Figure adapted from [15].

2.3.8 Horizon Graphs

Horizon Graphs [72] are one of the best known compact visualization techniques for time series. Horizon graphs combine the compactness of color mapping with spatial resolution in order to preserve details. They reduce space by dividing the chart into several bands and superposing them onto each other to create a layered form.

The horizon graphs implement a technique called *two-tone pseudo coloring* [76] in which each scalar value corresponds to two colors. This allows users to read the height of the chart with more precision. Horizon graphs implement a particular visual mapping where modulo transformation:

$$v = \text{sgn}(v).(a \times M + b)$$

decomposes each data attribute to sign of the data attribute's value (negative or positive), quotient a and remainder b . Each resulting component is then respectively mapped onto hue, saturation, and vertical position. Figure 2.20 demonstrates this process step by step.

The reduction in chart height and the resulting increase in data density enable the visualization system to display more data in a limited space. However, the deformations may obscure some patterns in data and the mental unstacking of layered charts may involve cognitive overload in some cases. Heer et al. [39] showed that horizon graphs are more accurate than line charts in discrimination and estimation tasks. Nevertheless, they demonstrated for both horizon graphs and line charts that the estimation error increases as chart height decreases. Javed et al. [49] studied horizon graphs, Stacked Graphs, and Braided Graphs with perception studies and found that none of them outperforms all other techniques, but each technique has its strengths and weaknesses depending on the task.

Later, Perin et al. [69] introduced *Interactive Horizon Graphs* in which, they im-

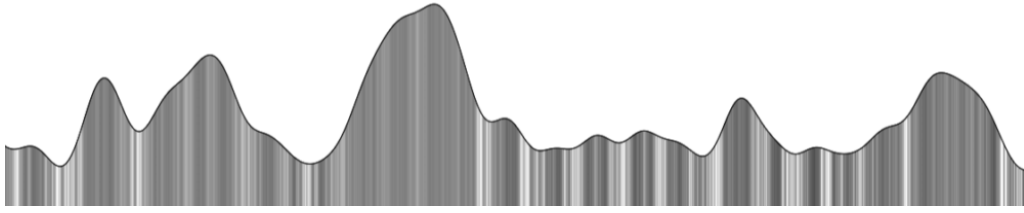


Figure 2.19: Slick Graph applies a frequency separation transformation on each data attribute. The low frequencies are mapped onto the vertical position (the smoothed line) and the high frequencies are encoded on the brightness of the area under the line. Figure adapted from [24].

plement pan and zoom interactions to achieve an interactive baseline. More recently, Federico et al. [31] introduced *Qualizon graphs*, a compact time series visualization that is based on horizon graphs. While it inherently integrates qualitative abstractions, it does not improve space efficiency further than horizon graph.

Given the strengths, limitations, and the general appeal of horizon graphs in space-efficient visualization of time series, we set horizon graphs as the comparison point and reference of our study.

2.4 Discussion of the Related Works

The conventional visual mapping consists in associating each data point to a separate visual channel. The common practice among the aforementioned techniques is the use of several visual channels for visualization of a data point. The added visual channels provide the visualization designer with additional information bandwidth or benefit from pre-attentive characteristics of human perception to guide viewers' visual attention.

However, we can note from this review that such "multi-channel" visual mappings are not achieved in a unique way. We can identify two different approaches in this regard. Some techniques map the same data value several times on different visual channels. CloudLines [57] (an event's importance on both circle's size and opacity) and redundant map symbols [27] are in this group.

Other approach is to map each data point onto several visual channels, but with this approach, the added visual channels do not redundantly show the same information. Instead, the extra visual channels represent new information extracted from the data value. Techniques such as horizon graph and Midgaard lie in this group. In horizon graph, the color brightness shows the order of magnitude calculated from each data value. Midgaard uses a secondary color hue visual channel to show abstract information (criticality of values) derived from data values based on a context-related knowledge. Table 2.2 recapitulates the multi-channel visual mappings reviewed in this chapter.


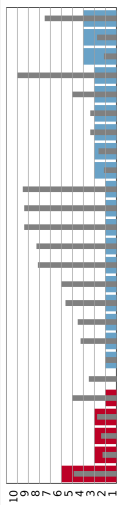

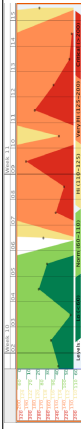

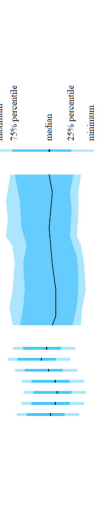


Technique	Image	Decomposition	Sub-values	Visual Channel
Slick Graph [24]		$v = f_{lo}(v) + f_{hi}(v)$	$f_{lo}(v)$ $f_{hi}(v)$	Position Brightness
OMM [15]		$v = sgn(v) \cdot (b \times 10^a)$	a b $sgn(v)$	Position Position Hue
Horizon Graph [72]		$v = sgn(v) \cdot (a \times M + b)$	a b $sgn(v)$	Brightness Position Hue
Qualizon Graph [31]		$v = sgn(v) \cdot (a \times M + b)$ M varying semantically	a b	Color Position
Midgaard [10]		Semantic abstraction	v Qualitative scale	Position Hue
Midgaard [10]		Extended Box-Plots	max min percentiles	Position Position Brightness
Multiple VIR [58]		Redundant Mapping	$value$ $value$	Position Brightness
CloudLines [57]		Redundant Mapping	$importance(event)$ $importance(event)$	Size Opacity

Table 2.2: A summary of existing time series visualization techniques which use multiple visual channels for displaying data values. Note that some techniques decompose data values to separate different information components, while others redundantly map the same values to several visual channels.

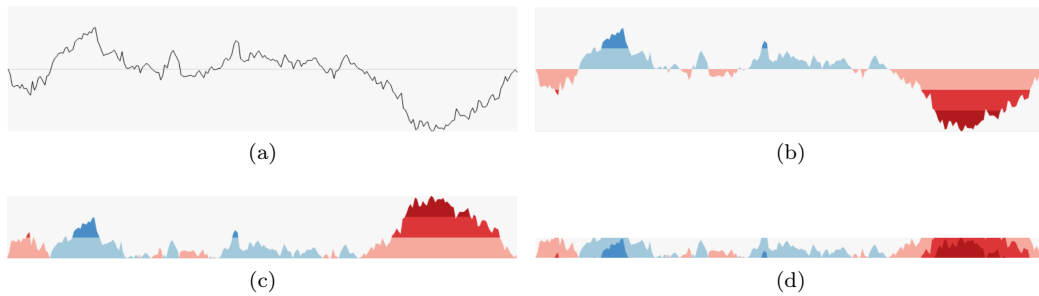


Figure 2.20: The step-by-step process of creating a three layer horizon graph for a typical time series. Figure adapted from [72].

While the first group of multi-channel mappings use the extra visual channels to attract the visual attention of the viewer, the second group aim to use the extra visual channels (and thus, the extra information bandwidth) to communicate derived information, either qualitative and abstract, or quantitative and statistical. We are particularly interested in the second group of multi-channel mappings, as they allow visual designers to display non redundant information which can be exploited by the viewers in their analytical tasks for creating new insights of data. Despite the advantages, this type of multi-channel visual mapping is less exploited in the existing visualization techniques. Our contributions in this thesis focus on this type of multi-channel visual mappings.

2.5 Summary

In this chapter, we presented the visual channels used and studied in the current thesis. We reviewed the existing visual techniques which implement some sort of conjunction of visual channels in their visual mapping of data. The works covered in this chapter are the techniques which are the most related to our work, i.e., conjunction of visual channels for time series visualization, and does not include studies on perception of visual channels or spatial visualization due to the limits of the scope of this thesis.

We categorized the reviewed techniques in two groups.

The first category includes techniques which achieve a one-to-many mapping by directly associating each data attribute to several visual channels. This type of redundant mapping makes use of the additional visual channels to attract the attention of the viewer, although the information mapped on different visual channels is identical.

The multi-channel visual mapping of the second category has a major distinction. In such techniques, data values are decomposed into sub-values through various transformation functions. The generated sub-values bear new information which was not explicit

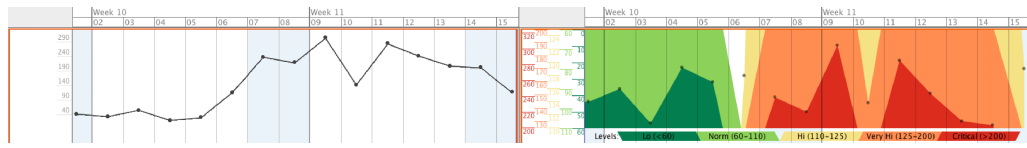


Figure 2.21: *Qualizon graphs* [31] combine horizon graph’s visual mapping technique with qualitative abstractions which need a priori knowledge of the context. left: original line chart, right: corresponding qualizon graph. Figure adapted from [31].

beforehand. Each sub-value is then mapped onto different visual channels, resulting in a one-to-many visual mapping. In contrast to the first category, the visual channels in this approach convey distinct information which was not visually separable with conventional (one-to-one) visual mapping.

The promise of increased information through such decompositional approach, motivated us to focus our efforts on the second category of multi-channel mappings. In the next chapter, we present a systematic study of this approach to visual mapping.

Part II

Contributions

Chapter 3

Composite Visual Mapping

Introducing a new visualization paradigm is not an easy endeavor, since it is very difficult to consider every aspects of data, tasks, and visual perception of users. In case of time series visualization, the time dimension has a distinct role in the analysis of data and thus, imposes additional constraints on the design of visualizations. This situation is similar to cartography where the position of visual marks is constrained by the geographical arrangement of the corresponding data. In consequence, map designers face limitations in exploiting the planar position as it is already reserved to geographical coordinates. Similarly in time series visualizations, certain visual channel (e.g. horizontal position) is always dedicated to the time dimension. Such constraints should be taken into account especially in design of alternative visual mappings.

In this work, we support the idea of a form of multi-channel visual mapping for time-series based on the findings in psychology, cartography, and existing visualization techniques. In Chapter 2, we reviewed the existing visualization which implement different kinds of multi-channel visual mapping and we categorized them in two groups:

1. Redundant mapping of each data value to several visual channels. While can be effective in drawing users' attention to that data value, no extra information is provided,
2. Division of each data attribute to several sub-values and mapping each of them to a separate visual channel. In this way, unveiled information is conveyed by each sub-value.

In this chapter, we first introduce the approach we chose, namely *composite visual mapping*. It is then necessary to define appropriate data transformations to support such mapping. These data transformations are decomposition functions which break each data to its components to reveal underlying information. To explore different possibilities of

composite visual mapping, we propose a systematic analysis and create a design space by implementing an interactive table. Finally, we point at the existing visualization techniques in the overview table.

We first introduced our systematic study of composite visual mappings during the doctoral consortium of the 28th French-speaking Conference on Human-Computer Interaction (IHM'16) [45]. The interactive tool is available online for further exploration [44].

3.1 Promises

Mapping sub-values of a single data value onto several visual channels, namely a *composite visual mapping*, can improve the visualization process for multiple reasons. Using several visual channels simultaneously, allows to distribute the information load to several channels, separate different facets of data (e.g. different orders of magnitude, different frequencies) and dedicate a full bandwidth visual channel to each of them.

The main advantage of this approach is that users can more easily focus on the desired property of data without losing the rest of information. This approach allows user to separately analyze two or more aspects of data and perform different analytical tasks using each channel (e.g. trend detection by using the channel dedicated to low frequencies, and uncertainty assessment by using the channel dedicated to the noise component). Also, by distributing the information load on several visual channels, each channel becomes less congested and thus, more readable.

This approach can be applied to different types of data. In our work we investigate its use for time series visualization. However, this approach can be put in application for other data types, i.e. geospatial data: similar to multi-variate maps, this approach can allow to display several facets of the data on the same area. For example, color value can be used to display data values and texture to display noise (as an indicator of certainty) on the same area.

3.2 Decomposition Functions

The first step in our approach to visual mapping consists in a data transformation in which, each data is *decomposed* into two or more sub-values. This process is a type data transformation (see the visualization reference model in Chapter 1) in which, interdependent sub-values are derived from the original series.

$$f(\text{value}_i) = \langle \text{subvalue}_{ix}, \text{subvalue}_{iy}, \dots \rangle \quad (3.1)$$

The decomposition function extracts and provides the data which can help the viewer with the task at hand. For instance, in a value comparison task, the order of magnitude of the values is the information which can be used to facilitate the comparison between the data points. This decomposition function, however, is not suitable for other tasks, i.e. detecting trends in data. The choice of the decomposition function, hence, depends on the analytical tasks of the viewer. Below are some of the possible decomposition functions which can be useful in composite visual mapping.

3.2.1 Modulo Operation

Modulo operation is the process of division of two numbers, which produces a quotient (a) and a remainder (b). In a division of v by M , the quotient a and b satisfy:

$$\begin{aligned} v &= \text{sgn}(v) \cdot (a \times M + b) \\ a &\in \mathbb{Z} \\ |b| &< |M| \end{aligned} \quad (3.2)$$

The resulting quotient, remainder, and the sign of the value can be further used as the sub-values for composite visual mapping. Modulo operation is the transformation function used in *horizon graphs* (See Chapter 2).

3.2.2 Logarithmic Transformation

In logarithmic transformation, each data attribute v is broken down into two sub-attributes, exponent a and coefficient b with K being a constant calculated for each dataset:

$$\begin{aligned} v &= \text{sgn}(v) \cdot (b \times K^a) \\ a &\in [0, nb) \\ b &\in [1, K) \end{aligned} \quad (3.3)$$

Similar to modulo operation, one component (a) communicates order of magnitude of data values and the other component b communicates small variations of data values. Order of Magnitude Markers (see Chapter 2) for example, uses logarithmic transformation as its decomposition function. Logarithmic transformation is specially useful for

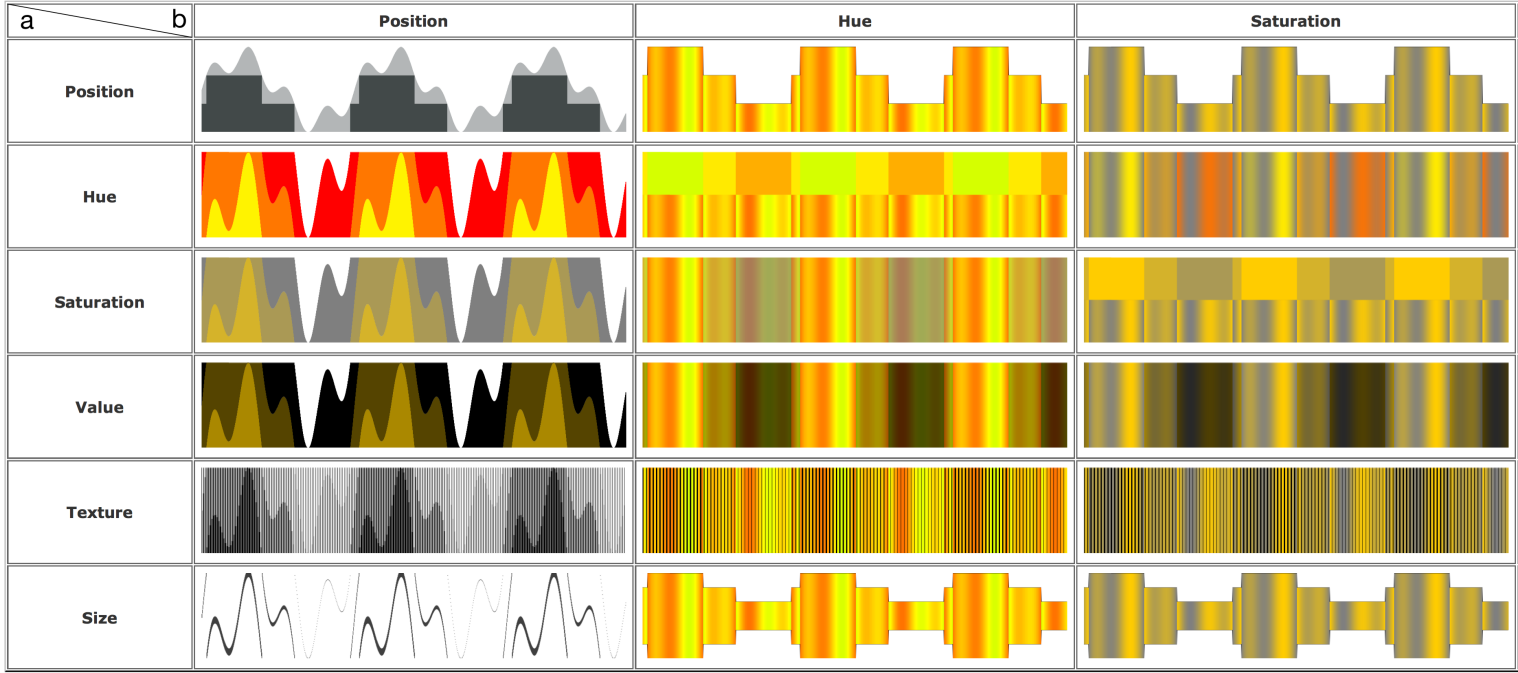


Figure 3.1: A systematic overview of possible combination of visual channels (here, using modulo operation: $v = a \times M + b$). Each sub-values a and b is mapped onto a separate corresponding visual channel. In case of combination of the same visual channels, those have been shown in two separate spaces.

datasets with magnitudes varying in large ranges.

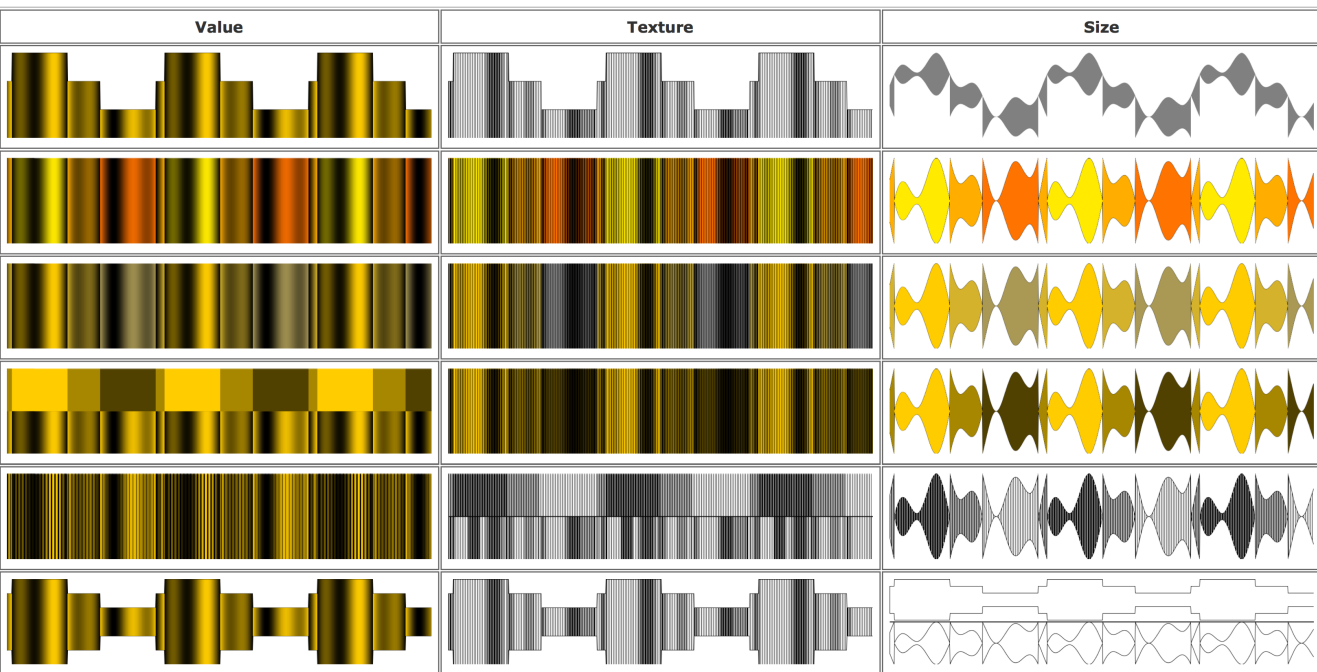
The range of data values and the selected M or K determine the number possible discrete values for the sub-value a . When modulo operation is used, the number of possible quotients of the division corresponds to number of layers or bands (as in horizon graphs).

3.2.3 Frequency Separation

Another useful decomposition method is to separate the low frequency and high frequency components of data. The low frequency component is related to the periodicity and the overall trend of values, while the high frequency component conveys the fluctuations and consequently, variability of data.

$$v = f_{lo}(v) + f_{hi}(v) \quad (3.4)$$

This operation is similar to passing a signal through a low-pass filter to obtain the low frequency signal. The high frequency signal, i.e. the noise, is then obtained by subtracting the low frequency signal from the input signal. This filtering process is achieved by different smoothing techniques. *Simple moving average* is one of the common smoothing techniques used for time series. The formula for calculating the moving average of N data points is:



$$x = \frac{1}{N} \sum_{k=0}^{N-1} x_{n-k}$$

or

$$\bar{x}_n = \bar{x}_{n-1} + \frac{x_n - x_{n-N}}{N} \quad (3.5)$$

In addition to moving average, other techniques such as kernel smoothers and exponential smoothing can be used to separate the trend and fluctuation in data. *Slick graphs* (see Chapter 2) use frequency separation as their decomposition function in their visual mapping. The resulting components from frequency separation can be mapped to different visual channels to create numerous possible composite mappings (for example, see Figure 3.2).

3.3 Systematic Study of Composite Visual Mapping

The importance of visual mapping and choice of visual channels has convinced us to concentrate our efforts on this step of the visualization reference model. Combination of visual channels promise improved visual forms that can better represent different aspects of data. However, depending on the nature of data and the analytical objectives of the user, we can imagine various combinations of visual channels. While some combinations have already been implemented and evaluated in visualization techniques, many others remain unexplored. This is the reason we think *composite visual mapping* is an approach worth a systematic study.

In order to acquire a systematic overview of the design space of composite visual mappings, we construct an exploratory table of combinations of visual channels. In this

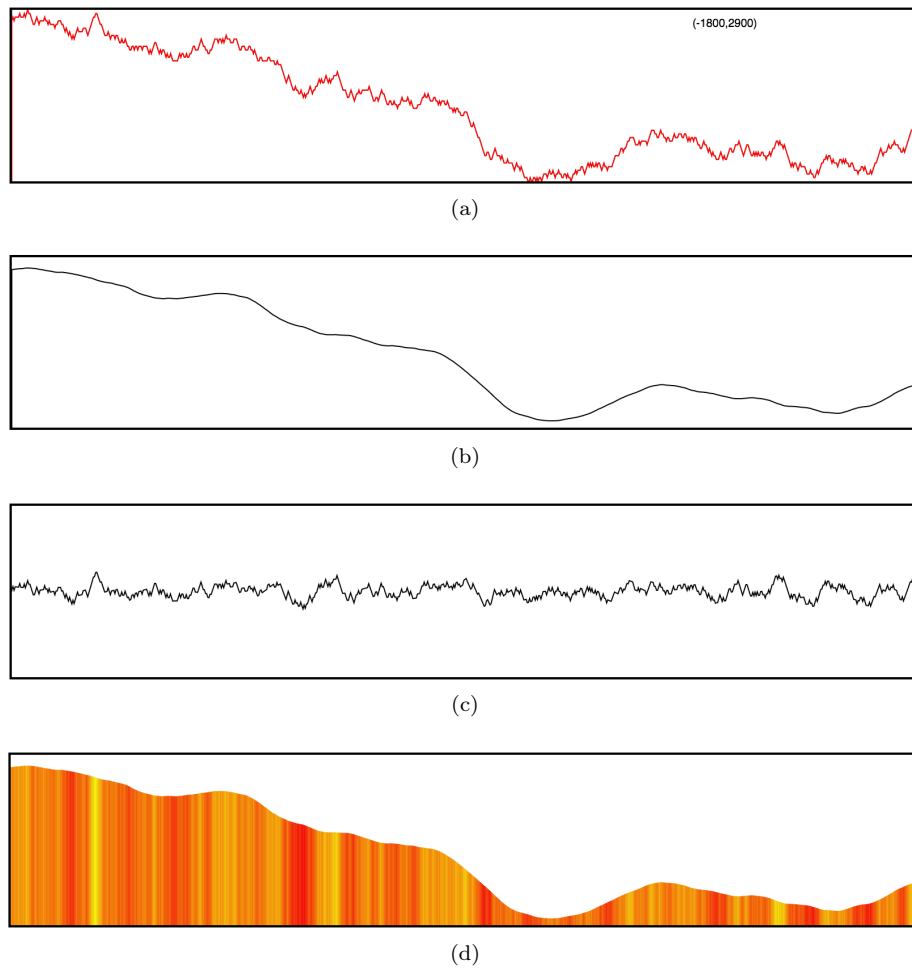


Figure 3.2: (a) Original data containing considerable amount of noise. (b) Trend component of the data obtained with moving average algorithm. (c) The noise component is obtained by subtracting the trend component from the original data. (d) A composite visual mapping using frequency separation where the trend component is mapped to the position on Y axis, and the noise component to the brightness of the color filling the curve.

current form, the table has two dimensions, allowing for composite visual mappings with two visual channels (see Figure 3). However, It is theoretically possible to generalize the table to higher dimensions to produce composite visual mappings with more than two visual channels.

Using one of the data transformation described in Section 3.2, each data value is decomposed to two sub-values a and b . Each row and column of this table corresponds to a visual channels described in Chapter 2. We attribute the sub-value a to the visual channel on each row, and the sub-value b to the visual channel of each column. In this way, each cell of the table corresponds to a visual mapping which is produced from the combination of the visual channels attributed to the cell's row and column.

In the rest of this chapter, we use the notation of C_{ij} to refer to the cell at line i and column j of the composite visual mapping table (Figure 3). For example, the first row of the table is attributed to the visual channel *position* and the second column is attributed to the visual channel *hue*. The cell at the intersection of such column and row (C_{12}) displays the resulting visual mapping from the combination of position and hue, with sub-value a mapped to position and sub-value b mapped to hue. The cell C_{21} uses the same visual channels position and hue, but with reversed mapping of sub-values: the sub-value a is mapped to hue and sub-value b is mapped to position. Although the two visual mappings are constructed from the same combination of visual channels, the results present little visual similarities.

In case of diagonal cells (C_{ii}) which are product of the identical visual channels, we mapped each of the two sub-values to the corresponding visual channel in a one-to-one way, and displayed them in two separated bands, each occupying a half of the total height of the chart. We should note that the visual representation of the combinations are not unique and other representation are possible as well. For instance, we can display most of the combinations in a spatially separate arrangement similar to those on the diagonal of the table. However, we were not quite interested by this type of representation for few reasons. First, this kind of representation is not space efficient as it uses half of the available space for data mapping, and second, this representation avoids interaction between the visual channels which is interesting for evaluation and design decisions in visualizations.

We implemented the exploratory table as an interactive tool which allows to modify data and visual parameters for visualization of selected test datasets. Such implementation would allow to investigate different configurations and the resulting visualizations for different types of data and decompositions. The interactive tool is available online [44] to explore the whole design space of our approach.

Figure 3.3: Interactive panel of the composite visual mapping table. Viewer can interactively modify the visual and data parameters of the resulting composite mappings.

3.3.1 Implementation

We implemented the table using JavaScript and HTML5’s canvas element as an exploratory tool for investigating composite visual mappings. An interactive panel allows the viewer to set the permitted range of hue values, as well as the baseline chroma and brightness (see Figure 3.3). The viewer can navigate among different test datasets and decomposition functions. As the viewer modifies the parameters, the cells of the table, i.e. the resulting combinations of the visual channels, are updated accordingly.

When modulo transformation is selected as the decomposition function, an extra parameter nb (*number of bands*) determines the number of concrete values permitted for the *quotient* (sub-value a). The divisor M and the sub-values a and b are calculated accordingly:

$$\begin{aligned}
 v &= \text{sgn}(v) \cdot (a \times M + b) \\
 M &= \frac{|Max_{data} - Min_{data}|}{nb} \\
 a &= \lfloor \frac{v}{M} \rfloor, \quad a \in [0, nb) \\
 b &= v - a \cdot M, \quad b \in [0, M)
 \end{aligned} \tag{3.6}$$

Similarly, exponent a and coefficient b are calculated for logarithmic transformation:

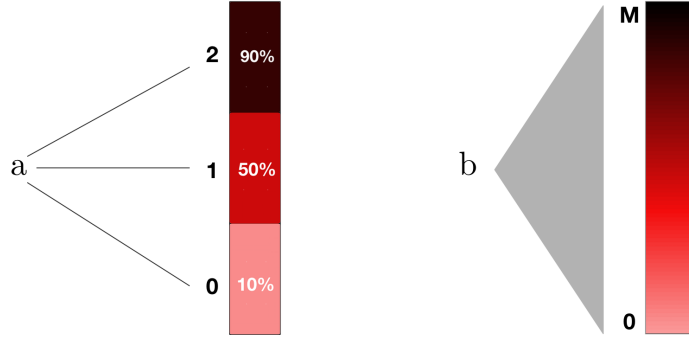


Figure 3.4: The sub-values a and b mapped onto the brightness. In modulo operation with 3 layers, a (quotient) can have three discrete values, while b is continuously mapped onto the channel variation. Mapping to other optical channels is achieved similarly.

$$\begin{aligned}
 v &= \text{sgn}(v) \cdot (b \times K^a) \\
 K &= \sqrt[nb]{\frac{Max_{data}}{Min_{data}}} \\
 a &= \lfloor \log_K v \rfloor, \quad a \in [0, nb) \\
 b &= K^{\log_K v - a}, \quad b \in [1, K)
 \end{aligned} \tag{3.7}$$

The resulting sub-values a and b are then mapped to their corresponding values of the selected visual channel (in per cent):

$$\begin{aligned}
 Channel_a &= 100 \times \frac{a+1}{nb} \\
 Channel_b &= 100 \times \frac{b}{M} \quad (\text{modulo transformation}) \\
 Channel_b &= 100 \times \frac{b}{K} \quad (\text{logarithmic transformation})
 \end{aligned} \tag{3.8}$$

Figure 3.4 shows the correspondence between the calculated sub-values and the destination visual channel (color brightness). The sub-values are similarly mapped onto color hue and chroma. We intentionally removed the two extremes of the values for brightness and chroma channels, as they correspond to nearly black, white, or gray colors which do not allow for the perception of the color's hue.

In case of texture, we chose a pattern of parallel vertical lines drawn at fixed distances. The widths of the vertical lines are proportional to the corresponding sub-value. The visual channel *size* is implemented as the thickness of the line progressing along the time axis. When size is not coupled with *position*, the line is vertically centered. You can refer to Appendix B for details on process of constructing a composite mapping.

3.3.2 Existing Composite Visual Mappings

Among the conjunctions presented in this table, only few have been implemented in the literature. Horizon graph [72] is the best known design which adopts composite visual mapping by means of modulo transformation. Depending on the design, horizon graph is the conjunction of color chroma or brightness and position in our table (C_{31} or C_{41}).

Order of Magnitude Markers (OMM) [15] is another existing technique. Using logarithmic transformation as the decomposition function, OMM redundantly maps sub-values to the position channel (C_{11} in our table). Furthermore, it uses the extra channel color hue to differentiate the sign of the quantitative values.

Slick graphs, another technique presented in Chapter 2, uses frequency separation as the the decomposition function. The resulting low frequency component (smoothed curve) is then mapped onto position and the high frequency component is mapped onto the brightness of the color filling the area below the curve (similar to C_{14} in the table). Evidently, combination of position and other retinal channels i.e. saturation and texture variation also can achieve similar effects (See Figure 3.2 for more details). Compared to Position–Brightness, however, the effectiveness of such combinations is debatable due to less visual contrast.

3.4 Summary

In this chapter, we presented the first contribution of this thesis. We presented the idea of composite visual mapping based on decomposition of single data value into several sub-values. We presented an overview of the design space using an exploratory table. The exploratory table supported by an interactive prototype allows us to systematically view the composite visual mappings that can be constructed by combination of visual channels. As mentioned earlier, the presented designs are not the only possible forms of combined visual channels and can be modified and further enhanced. From this table, we marked the existing visualization techniques, notably the well established horizon graphs.

Since the existing *horizon graphs* have been proved to be efficient for visualization of time series [39], we are interested to verify if other composite mappings benefit from the same advantages. In the next chapter, we conduct an empirical study to compare potential composite mappings with horizon graphs.

Chapter 4

Evaluation of Composite Visual Mappings

In Chapter 3, we presented the design space of composite visual mappings. We observed that among all possible combinations of visual channels, only few have been implemented in the existing visualizations. However, we are aware that not every combination of visual channels is suitable for data representation. Besides intuition, empirical work is clearly needed to evaluate the effectiveness of the new visual designs.

In this chapter, we present our evaluation of a promising subset of unexplored composite mappings. The evaluation of all combinations of visual channels is practically impossible and each combination can be achieved in several forms. Hence, we only examined a selection of composite mappings in this empirical study. We compare our selection of techniques with the existing *horizon graphs* in order to evaluate the overall relevance of the formers. We used the findings from this evaluation to improve the further iterations of our visual designs.

This chapter has drawn on our work published at PacificVis conference in 2018, entitled “Composite Visual Mapping for Time Series Visualization” [47].

4.1 Challenges of Visualization Evaluation

As the research in visualization grows, it is important that its findings are empirically validated. Many of the questions and challenges in evaluating visualizations are common to all empirical research in general, and Human-Computer Interaction (HCI), perceptual psychology, and cognitive sciences in particular. The relationship to psychological and cognitive sciences is due to the role of the human viewer and her visual perception in

effectiveness of a visualization technique. Evaluation of visualizations is also closely related to HCI in that many evaluation scenarios involve interaction tasks such as zooming, item selection, etc. [18].

Given the methodological affinities between such empirical research and visualization evaluation, we can consider the many challenges and guidelines they share with visualization evaluation. One of the shared challenges is the difficulty of obtaining a balanced sample of participants. In many cases, the participants of empirical evaluations are recruited among the domain experts who work in the same research environment, while the final users of the visualization system would be ordinary people with average knowledge of computer science. Another common challenge when evaluating a visualization system is that we can not say whether the results are due to the used visualization technique itself or the software as a whole and its usability, ergonomic aspects, etc. Using sketchy software prototypes can also impact the results from usability evaluations [34].

In addition to challenges common with HCI, there exist more specific challenges for visualization evaluation. For instance, in visualization we deal with different tasks, varying from very low level perceptual tasks (e.g. compare associate, and categorize) to high-level cognitive reasoning (e.g. understanding causal relationships, prediction, and knowledge learning). Some of these tasks, especially those high-level tasks related to decision making and interpretation, are not clearly defined and their evaluation is more challenging.

From this brief overview of challenges in evaluating visualizations, we can identify three desirable factors [65]:

- **Generalizability:** a result of the evaluation can be generalized to other people than those who participated in the experiment or other situations.
- **Precision:** measurements that were taken were precise and the control of the experimental factors was definite.
- **Realism:** a result is considered realistic if the context in which it was evaluated is similar to the final contexts in which it will be used.

Ideally, we would like to satisfy all three factors in our evaluation. However, each evaluation methodology favors one or two of these factors. For example, field studies are typically conducted in the actual context of the technique's application. The results from such studies are on one hand highly realistic, but on the other hand less precise than laboratory experiments due to lack of control in the evaluation process. The methodology should therefore be chosen wisely beforehand in order to avoid making false conclusions or exaggerating the results of a visualization evaluation.

4.2 Approach and Methods

Among the large variety of possible composite mappings, *horizon graph* is the most widely adopted and evaluated existing technique. Horizon graphs apply modulo transformation which allows to distinguish orders of magnitude based on the quotient of the division and the mapped color variation. Such choice of decomposition function, which is based on the magnitude of data values, aims to facilitate comparison of data values. Heer et al. [39] compare horizon graphs with line charts to evaluate their effectiveness in value comparison tasks. In their empirical evaluation, they measure participants' accuracy and speed in estimation of value differences and show the advantage of horizon graphs over simple line charts in terms of accuracy. Based on their findings, they recommend horizon graphs as a technique for increasing data density and perceptual optimization.

The proven effect of composite mapping in case of horizon graphs convinced us to adopt its experimental methodology as the reference of our study. The choice of an approved experimental protocol helps us to focus on the precision of the results. However, we are aware that such controlled protocol impacts realism and generalizability of our results.

We argue that the demonstrated accuracy improvements are not exclusive to horizon graphs, and at least some other composite mappings can offer similar benefits. We conducted an experiment to answer the following questions:

- H1** The new composite mappings have performances similar to those of horizon graphs in discrimination and estimation tasks
- H2** Logarithmic transformation boosts users' performance in discrimination and estimation tasks
- H3** Like with the horizon graphs, increasing the number of bands/layers will improve the accuracy and estimation speed of the new composite mappings

4.2.1 Experimental Protocole

Participants

We recruited 12 participants among people working at our laboratory. The participants, 10 male and 2 female, aged between 22 and 28 and were all computer science students or degree-holders. We acknowledge that such participant pool does not represent the general public. However, the final users of visualization systems often have above average knowledge of computers and representations of data.

Tasks

In order to be able to compare our results with the findings from earlier studies on the use of composite mapping in horizon graphs [39], we evaluated the same following analytical tasks:

Discrimination Task: Participants viewed pairs of time series. Each series of the pair was marked at a specific point. The participants were asked to determine which series had larger value at its marked point. All pairs had constant physical distance between their two discrimination points.

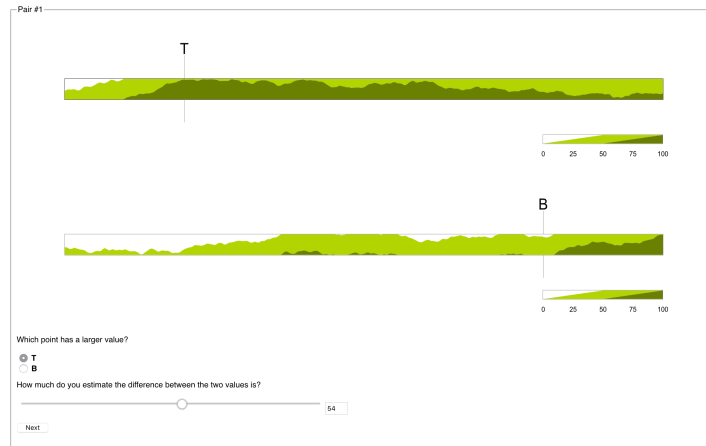
Comparison Task: Following the discrimination task, where the participants determined the point with larger value, they were asked to estimate the absolute value difference between the two reference points. All series had values varying from 0 to 100 and hence, value difference between the two points could vary from 0 to 100.

Experimental Interface

Our experimental interface was entirely created with JavaScript. In each trial, the participant is presented with a pair of two separate charts, one positioned below the other one. The two charts of each pair had identical visual configurations, while conveying two distinct time series. Each chart of a pair was marked either T or B in a fixed position through all trials (Figure 4.1).

At each trial, the participant was asked to report whether position T or B represents a greater value via radio buttons marked T and B . Participants were then asked to estimate the absolute value difference between the two positions. They responded by using a slider without tick marks in order to avoid anchoring effects [63]. The value selected via the slider was shown as numerals next to the slider. Upon clicking on "Next" button, the participant's responses, as well as the completion time for the trial was recorded and a new pair of charts replaced the former ones. The correct answer for the trial was not communicated to the participant to avoid learning effects. Once an answer was submitted, participants could not go back to review their answers.

The experiment was run in Google Chrome web browser version 59.0 on a 27-inch Apple iMac with Retina 5K display in full brightness in a room with constant ceiling lighting and no sunlight.



Which point has a larger value?

T
 B

How much do you estimate the difference between the two values is?



Next

Figure 4.1: Interface of the experiment, presenting a pair of charts with Size-Hue mapping with three-layers modulo transformation. (top) global view, (bottom) detail on the questions and answers.

Experimental Procedure

For each participant, the experiment begins with a tutorial on each of the selected mappings and the global procedure of the experiment. In this phase which lasts a couple of minutes, we introduce each visual mapping to the participant and answer to their inquiries. Participant, then, takes control of the experimental interface and starts with filling a form with information on her age, vision and education.

After this step, the first trial appears and the main experiment begins. For each trial, we measured the time a participant spent before validating her answer. During an interview at the end of the experiment, participants shared their views on the experiment and different visualizations with us.

The experiment consisted in 100 trials ($5 \text{ mappings} \times 2 \text{ transformations} \times 2 \text{ layers} \times 5 \text{ trial per condition}$) divided into 10 blocks of 10 trials. The trials in each of the 10 blocks had the same mapping and data transformation. Participants could voluntarily take breaks between blocks and resume when they were ready.

4.2.2 Chart Generation

For each chart, we generated a time series using a moving average smoothing over an unconstrained random walk. Time series' minima and maxima were then translated onto 0 and 100 respectively. This resulted in having the same value range for all charts. The charts were made on HTML 5 Canvas elements using JavaScript. All charts had a width of 2000 pixels and a height of 160 pixels. All charts measured 233×19 mm on our screen. In an attempt to preserve perceptual uniformity, we used polar CIELUV (HCL) color space. The data generation algorithm is similar to that of our other experiment and is detailed in Appendix A.

4.3 Evaluated Composite Mappings

We used our exploratory tool, the composite visual mapping table (Figure 3) to select the candidate visual mappings for evaluation. We kept high compactness, minimum visual clutter, and adaptability for quantitative data as the main requirements of our selection.

Toward this end, we consider the categorization of visual channels presented in Chapter 2: geometric channels (i.e. position and size), and optical channels (i.e. hue, chroma, lightness, and partially, texture). We tried to have a fair selection of composite mappings combining channels from each of these groups for both sub-values a and b .

Similar to horizon graphs, our selected composite mappings combine two visual channels. We had to limit the number of mappings that we test to five (horizon graph and 4 new composite mappings), otherwise evaluation of all conditions would become practically impossible. Our selection of visual mappings (see Figure 5.1) consisted of the following five composite visual mappings:

- **Line chart:** We included line chart as the reference technique for visualization of time series. Line chart's simplicity makes it by far the most familiar visualization technique for users. Line chart exclusively uses the geometric channel *position* for encoding data.
- **Horizon graph:** Among existing visualization techniques which implement composite visual mapping, horizon graph is the most extensively used technique for visualization of time series. Its similarity with a layered line chart helps to reduce the learning effort for novice users. The underlying composite visual mapping of horizon graph is a combination of color variation and position. The color variation can be achieved by either hue, chroma, or lightness variation. Similar to existing implementations of horizon graphs, we used saturation (a variation of both chroma and brightness) variation for differentiating layers of horizon graphs.

- **Saturation–Texture** In this mapping we used saturation, similar to horizon graph, for the sub-value a (quotient) and texture variation for sub-value b (remainder). The texture is produced by drawing parallel vertical lines at 5 point intervals. The thickness of each line is the average of the sub-value b in the underlying 5 points, ranging from 1 point to 5 points for the range of sub-value b .
- **Hue–Saturation:** While no natural ordering exists for hue, hue variation is commonly used in visualizations. We hence examine the use of discrete values of hue (sub-values a) with varying levels of saturation according to sub-value b .
- **Saturation–Size:** Similar to horizon graphs, in this technique saturation is used for discriminating discrete values of a . The size channel is implemented as the width of the symmetrical color-filled area for mapping sub-value b .
- **Size–Hue:** Size is implemented as the symmetrical width of the color-filled area for sub-value a . The color filling consists in the variation of color hue in a limited range (from red to green).

Here and in the rest of this thesis, the naming scheme [Visual Channel 1]–[Visual Channel 2] denotes a composite mapping consisted in sub-values a and b mapped to [Visual Channel 1] and [Visual Channel 2], respectively. Such mapping is located at row [Visual Channel 1], column [Visual Channel 2] of the composite visual mapping table.

Logarithmic transformation has been used in the literature for visualization of data with varying orders of magnitude (see Chapter 2.3). We included this transformation to evaluate its relevance with the selected composite mappings.

Another experimental variable is the number of *bands* or *layers*. In both modulo and logarithmic transformations, the sub-value a represent the order of magnitude of data. The number of envisioned orders of magnitude, i.e. the number of layers, specifies the divisor M or the base of logarithm K (see equations 3.6 and 3.7). We limited the number of layers to three because earlier studies [39] have shown that higher number of layers significantly increase the error.

4.4 Results

For the analysis, we removed an outlier trial which received wrong answer from all of the participants. This should be due to the extremely small value difference between the two markers (0.05 in a scale of 0 to 100). However, it can be a basis for further study of equality cases in order to evaluate users' estimation behavior. Otherwise, none of the participants demonstrated wildly inaccurate performance in any block of the experiment. The participants were divided into two groups: the first group started the experiment

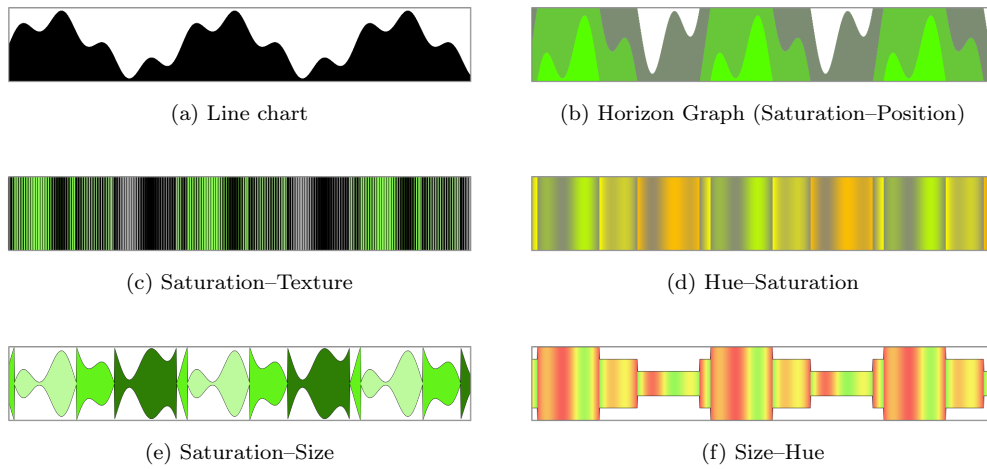


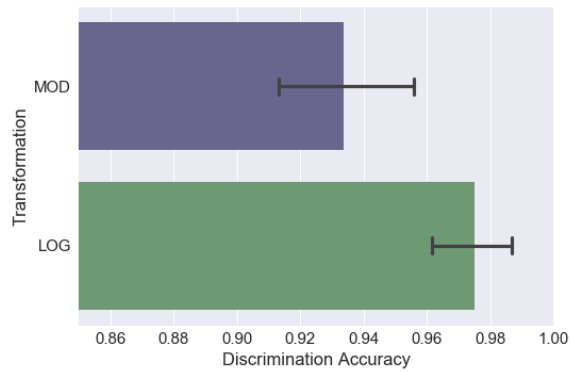
Figure 4.2: Complete selection of visual mappings examined in our empirical study displaying a dummy dataset. Except line chart which was included as the reference point, all other techniques use composite visual mapping with modulo division as the transformation function.

with modulo transformations and the second group began with logarithmic transformations. The two groups demonstrated very similar accuracy levels (closely in the same intervals of confidence) which rejects the possibility of a significant learning effect. In accordance with other studies [21], we reported logarithmic absolute error as a measure of estimation error.

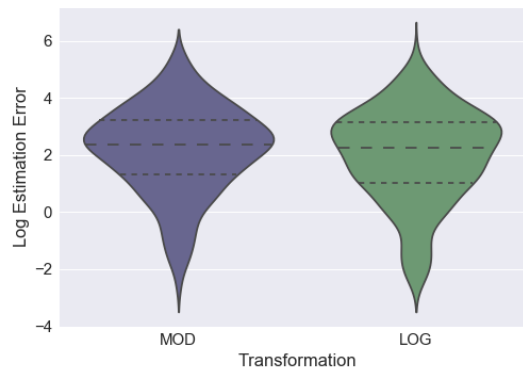
4.4.1 Performance by Data Transformation

For both modulo and logarithmic transformations, discrimination accuracy averaged above 90%. However, participants performed slightly better with logarithmic transformation in the discrimination task (Figure 4.3a). On the other hand, the estimation error does not differ between the two transformations. The violin plots [40] in Figure 4.3b allows to visually compare the distributions of logarithmic absolute error for the two transformations.

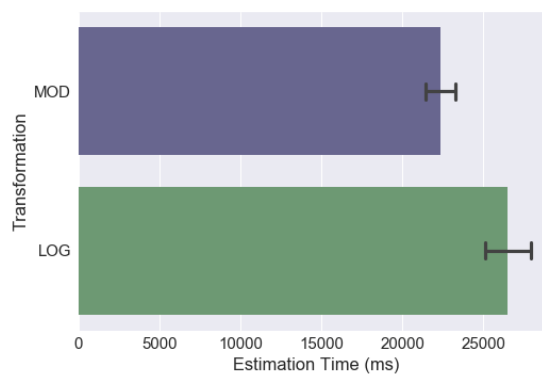
In generating the visualization pairs with logarithmic transformation, we calculated the base depending on the minimum and maximum of each dataset in order to ensure having the chosen number of “layers”. This varying base of logarithm results in varying value ranges on those charts’ legends which may require extra mental calculation to estimate value differences. Participants reported this point in the follow up interviews: they needed more mental calculations to deduce the value difference. Indeed, our results show longer completion times for logarithmic transformation (Figure 4.3c).



(a) Average discrimination accuracy for modulo and logarithmic transformations. Black brackets show the 95% confidence intervals.

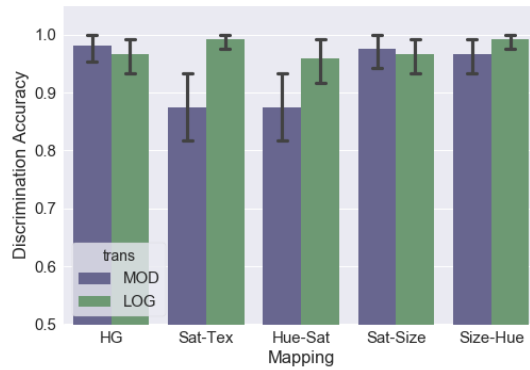


(b) The distribution of logarithmic absolute error for modulo and logarithmic transformations. They share similar distribution of error. The dashed lines represent the median, the first, and the third quartiles.

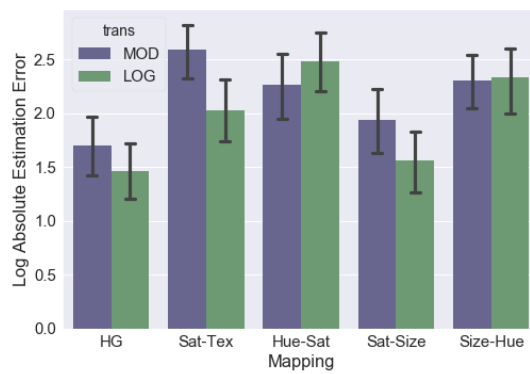


(c) Completion time for modulo and logarithmic transformations. Logarithmic transformation extends the completion time due to the irregular values it introduces in chart legends.

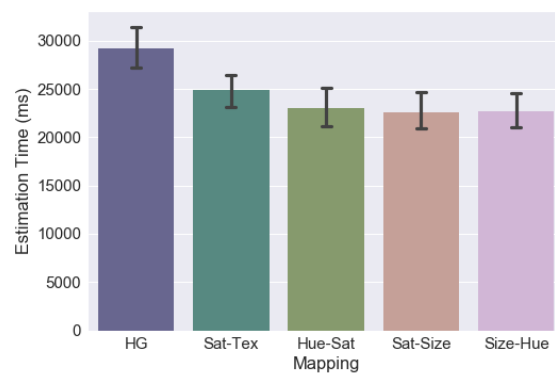
Figure 4.3: Effect of transformation function: modulo vs. logarithmic transformation.



(a) Discrimination rates for different mappings. Sat-Tex with modulo transformation performed the worst. The effect was not observed with logarithmic transformation. Black brackets show the 95% confidence intervals



(b) Lower absolute errors for Horizon Graph, while the difference (in units) is not significant in this scale.



(c) Completion time for the five visual mappings: Horizon Graph performed the slowest, while others had the same levels of completion time.

Figure 4.4: Effect of the visual mapping (visual channels combination): Horizon Graph (Saturation–Position) vs. our selection of composite mappings.

4.4.2 Performance by Visual Mapping

Figure 4.4a shows discrimination accuracy rates for the five selected visual mappings and respective rates for the two data transformations. Sat-*Tex* and Hue-*Sat* mappings with modulo transformations had the lowest average discrimination accuracy, while other conditions had similar rates all above 90%. Interestingly, in the former two mappings, charts with logarithmic transformation had much higher discrimination accuracy rates above 90%.

Absolute estimation errors (Figure 4.4b) were lower for horizon graph and Sat-*Size* mapping (at around 5 units) comparing to the others (at around 8 units). The difference, however, does not seem significant in a scale of 100 units.

The results show longer completion times for horizon graph (Figure 4.4c). The other mappings exhibit lower levels that are similar.

4.4.3 Performance by Number of Layers

The design of the experiment allowed for either two or three discrete values for the sub attribute related to the order of magnitude (i.e. quotient for modulo transformation, power coefficient for logarithmic transformation). This range of values results in two or three levels (layers or bands) of the attributed visual channels.

The results show that a three-layers design compared to a two-layers design leads to higher discrimination accuracy rates, lower estimation errors and similar completion times (Figure 4.5).

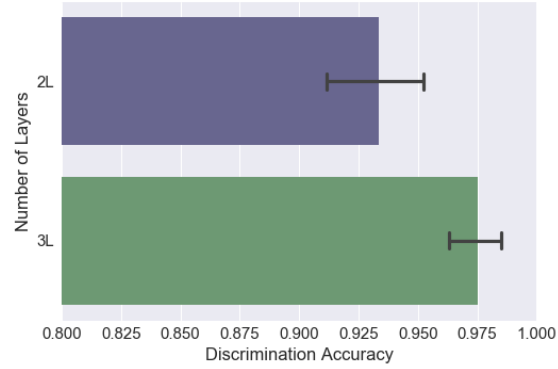
4.4.4 Performance by Real Value Differences

Figure 4.6 reports the effects of the actual value difference:

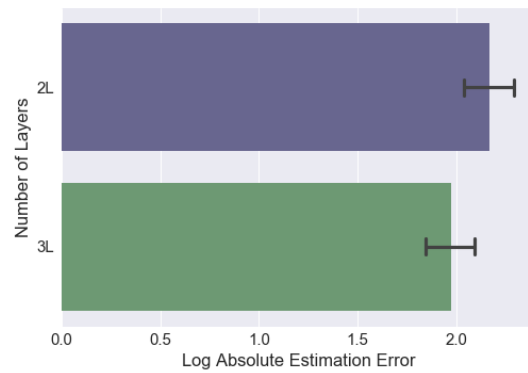
Figure 4.6a shows the effect on discrimination accuracy for the two transformations. Modulo transformation has lower scores in very small and very large actual differences. On the other hand, logarithmic transformation score retains the same levels of scores across ranges of actual difference.

Figure 4.6b reports the effect on estimation error for the five mappings. Overall, larger actual differences between the values result in higher estimations errors. However, horizon graph and Saturation-*Size* demonstrate similar behavior and are less prone to the actual value difference.

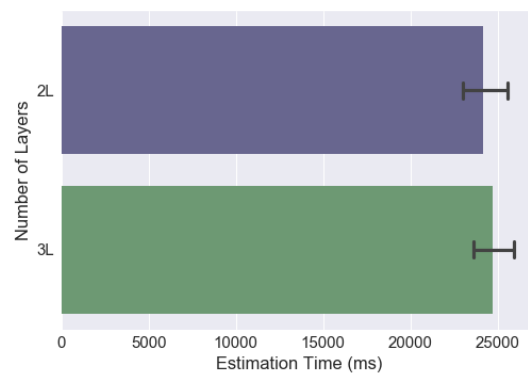
Figure 4.7 shows that the estimation time increases across the board. horizon graph



(a) Discrimination accuracy: better correct discrimination with three layer design

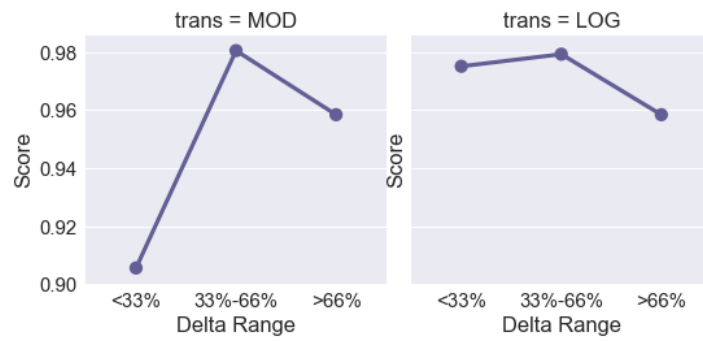


(b) Logarithmic error: lower estimation error with three layer design



(c) Completion time: no meaningful difference between the two and three layer design

Figure 4.5: Effect of layering: two layers vs. three layers design. Higher layering was excluded due to the proven lower accuracy according to earlier studies.



(a) Discrimination accuracy: Modulo vs. logarithmic transformation



(b) Logarithmic error: comparing the five mappings

Figure 4.6: Effect of the real value difference on discrimination and estimation accuracy

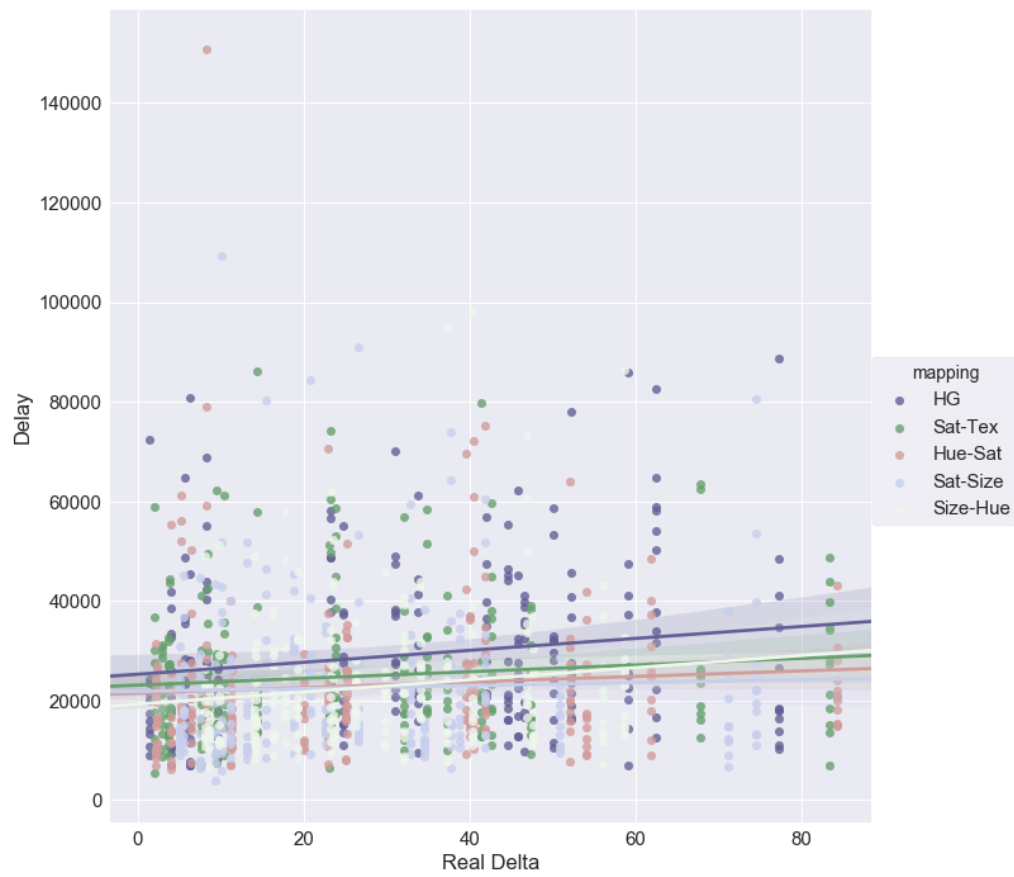


Figure 4.7: Effect of the real value difference on completion time: linear regression lines with 95% confidence intervals. horizon graph (HG) is the most vulnerable to the real value difference

is the mapping the most affected by this, while Size-Hue is the least susceptible.

4.4.5 Discussion of the Results

Our results confirm our first hypothesis **H1** that the accuracy and speed of other alternative composite mappings can attain levels similar to horizon graphs for the same discrimination and estimation task. The discrimination accuracy rates averaged above 90% for all of the mappings. Mean estimation errors were similar across different mappings, while they were slightly lower for horizon graphs and Saturation-Size. This can be explained by the fact that size, and similarly position which in case of horizon graph results in larger areas, are the most easily perceived visual channels for most users [21]. In matters of speed, horizon graphs showed to be the slowest mapping. During the interviews with the participants following the experiment, some noted that they found the mental unfolding of the horizon graphs to be a cognitively demanding task. Although this can partially explain the slower performance for horizon graphs, more training may reduce the effect. In addition, higher familiarity and confidence of users in geometric visualizations including horizon graphs might drive them to spend more time in hope of achieving more accurate estimations. With current evaluation, we cannot identify the cause of longer estimation times in horizon graphs.

Based on the interested aspects of data, many transformation functions can be imagined for dividing data attributes. In our study we focused on modulo and logarithmic transformations, two functions that are more related to the magnitude of data and that have already been used in the existing horizon graphs and order of magnitude markers. Discrimination accuracy averaged slightly higher for logarithmic transformation. This is explained by the higher discrimination accuracy averages for very small and very large differences (Figure 4.6a). This effect may be explained by the fact that with logarithmic transformation, two points both with small values are more probable to fall into the different regions of the visual channel representing the magnitude component.

The use of logarithmic transformation did not impact the estimation error, but decreased the speed of users. Overall, the effect of transformation function was limited and we failed to validate hypothesis **H2**.

Consistent with previous studies [39], using more “layers” which translates to more discrete values for the magnitude component helps users by increasing their discrimination accuracy and lowering estimation error without sacrificing the speed. However, as the previous studies have shown, more than three layers can harm the performance of the user in estimation task and thus, should be avoided. We therefore validated hypothesis **H3**.

We found that the real difference of values between the two points has an impact

on participants' estimation accuracy (Figure 4.6b). All five mappings show bigger errors with larger real value differences, but the effect is less prominent for horizon graphs and Size-Saturation. We observe that while in small real value differences all mappings show similar error levels, those two mappings show lower errors at large real value differences.

Also, the increase in real value difference has negative impact on participants' speed, with horizon graph being more prone to this effect in this regard (Figure 4.7).

4.5 Limits and Perspectives

One limitation of the present study is that we only measured the performance of users in a narrow task of discrimination and estimation. While this is a common task in visual analysis of time series, the effects of composite mapping on other tasks such as trend perception and graphical perception of rates of change remain to be investigated.

Our aim in this study was to show the general benefits of our approach to visual mapping in discrimination and estimation tasks. The empirical study allowed us to report some observed trends, yet further comprehensive evaluations are clearly needed for statistically significant results.

Another limitation of this work is that for each combination of visual channels, several visual designs are imaginable, and we only tested one particular design. Therefore, design improvements may enhance the performance of the technique.

In our evaluation, we only tested one type of data. While smoothed random walks are relevant datasets for discrimination and estimation tasks, other types of datasets (e.g. noisy data) are suitable for other analytical tasks (e.g. trend detection). Such cases require future evaluations.

In perspective, it is compelling to evaluate other potentials of these composite mappings. For example, evaluating viewer's graphical memory in dealing with geometric and optical visual channels is an interesting research avenue to explore in the future.

4.6 Summary

In this chapter, we presented the results from our first evaluation of composite visual mappings. Our experiment has shown the state-of-the-art accuracy of a selection of these mappings in a discrimination and estimation task. During interviews with participants of our evaluation, they found the new mappings straightforward to learn.

We found composite visual mapping an interesting basis for new time series visual-

izations. Future refined visual designs can be created based on this approach to visual mapping. Compactness and vast visual variability of composite mappings make them promising for many contexts and analytical tasks. It is of our great interest to further the investigation of composite visual mappings in terms of sensibility to chart size, learnability, and measures other than value estimation accuracy.

Chapter 5

Composite Visual Mapping & Space Efficiency

In Chapter 4, we described our first empirical study on composite visual mappings. In the tasks we evaluated, we observed similar performances for our selection of composite mappings and the reference horizon graph. In this chapter, we aim to study the space efficiency properties of our visual mappings.

Although horizon graphs were conceived as a means to reduce the space, earlier experiments [39] have discovered that estimation error is higher for smaller charts. We suspect that this effect is due to the dependence of horizon graphs on position channel and hence, the size of the chart. As the next step of our study, we are interested in looking at the effect of the size of the chart in estimation accuracy with different composite mappings. We suspect that the composite mappings which do not use position are less vulnerable to the effect of chart size.

The goal of this study is to evaluate composite mappings for compact visualization of time series. We expect that with some visual refinement, certain composite mappings can compete with horizon graphs in space-constrained situations. In this chapter, we describe our second empirical in which we compare a selection of composite visual mappings with the existing horizon graphs across different chart heights for discrimination and estimation tasks. We show the advantages of our designs in smaller chart heights and based on our results, we propose several recommendations on the use of composite mappings and combinations of visual channels.

This studies described in this chapter have been submitted to the 30th French-speaking Conference on Human-Computer Interaction (IHM'18) as a scientific paper [46].

5.1 Visual Mapping and Chart Height

As we stated earlier, the physical screen space is a valuable commodity which directly influences the amount of viewable information. We overviewed the existing space efficiency techniques in Chapter 1. In addition to those approaches, we hypothesize that composite visual mappings can enhance the space efficient properties of visualizations due to their varying degree of dependence on chart height. By reducing chart height, we can achieve higher space efficiency in different scenarios. For instance, we can fit larger numbers of charts in the same screen space, or adapt visualizations to smaller screens such as portable device.

Tufte [81] advises for charts with greater length than height. According to Tufte, horizontally stretched time-series are more accessible to our eye because our eye is naturally practiced in detecting deviations from horizon. It also helps elaborate the workings of the causal variable (time) in more detail [81]. Hence, it is quite common to reduce the height of a chart while preserving the same length (temporal window), in order to increase the space filling properties of the visualization and thus, fit more data (e.g. more data dimensions) in the same space.

However, it seems that vertical reduction of a chart has not the same effects on all visualization techniques. For example, in *Visualizing Data* Cleveland [20] showed how the aspect ratio of a line chart can impact user's perception of trends in data. We hypothesize the degree to which a visualization technique is impacted by reduction of its height is at least partially related to its underlying visual mapping. Previous studies on line charts and horizon graphs have shown that such visual mappings which employ geometric visual channels (e.g. size and position) are heavily impacted by the vertical size reduction [39].

5.2 Approach and Methods

We conducted a series of experiments on our proposed visualization designs with several objectives in mind. First, we wanted to reproduce one finding of prior studies on Horizon Graphs [39] which suggests that Horizon Graphs perform better than simple line charts as the most basic form of time series visualization. Furthermore, we wanted to compare our proposed visual designs with Horizon Graph in the same tasks of discrimination and estimation. We also wanted to examine the effect of chart height reduction on our propositions and Horizon Graphs.

We had a number of hypotheses in this study:

H1 Horizon graphs perform better than simple line chart in discrimination and estima-

tion task thanks to composite mapping (reproduction of previous studies).

H2 While Horizon Graph and the mappings which use geometric channels perform better in larger chart heights. However, they lose their superiority in smaller chart heights due to spatial effects (physical size, banking angle, etc.).

H3 In small chart heights, our designs may be on a level with Horizon Graph in terms of accuracy, and some can even top it in terms of speed.

5.2.1 Experimental Protocole

In a continuing effort to compare other alternative composite visual mappings with horizon graphs, we carried out the second empirical study with an experimental protocol similar to that of the first experiment. In order to refine our visual designs, we conduct several pilot tests prior to the main experiment.

Regardless of the experiment, at each trial the participant viewed a pair of charts. The two charts displayed two distinct time series using identical visual design. At fixed positions, one chart was marked with T and the other with B (see Figure 4.1). At each trial, participants first performed a discrimination task in which they were asked to report whether T or B points to a larger value. The time elapsed before a participant selects the corresponding radio button was registered as discrimination time. Participants could change their answers and the extra discrimination time was taken into account. Next, participants performed an estimation task were asked to estimate the absolute value difference between the two points. They communicated their estimations using a continuous slider without tick marks in order to avoid anchoring effects [63]. Once the participant clicked on the "Next" button, the time elapsed since the discrimination task is recorded as the estimation time. At this point, a new trial replaces the old one and participants could no more go back to review their answers.

5.2.2 Chart Generation

Each experimental trial consisted of a pair of non-identical time series generated by running a moving average smoothing algorithm over a constrained random walk. The constrained random walk was intended in order to ensure a uniform distribution of values at marker points, starting and ending points, and consequently, the real value difference between the two markers of a trial pair. The smoothing process removes high frequency fluctuations which would otherwise result in large value differences between adjacent pixels. The data generation algorithm is detailed in Appendix A.

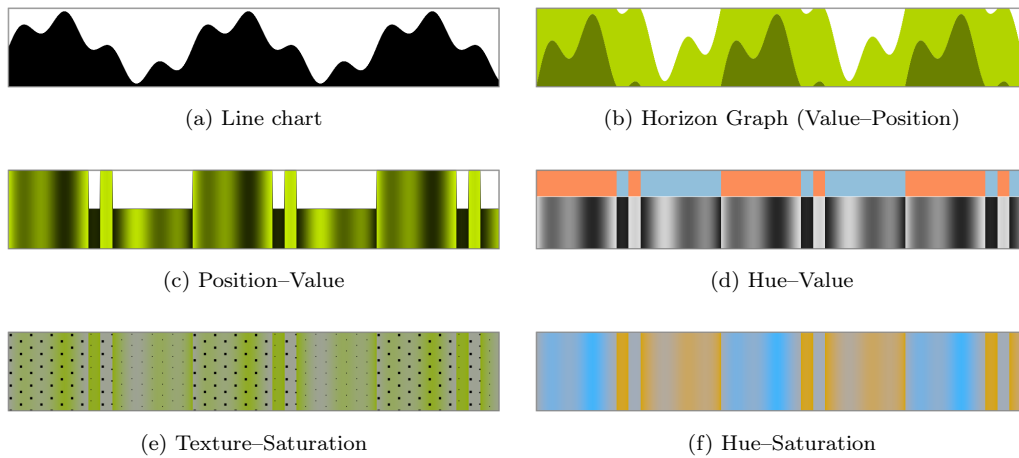


Figure 5.1: Complete selection of visual mappings examined in this study displaying a dummy dataset. Mappings (a), (b), and (c) use one geometric visual channel. Mappings (d), (e), and (f) exclusively use optical channels. Except line chart which was included as the reference point, all other techniques use composite visual mapping with modulo division as the transformation function.

We admit that this type of data encompasses only a fraction of common time series among many, for instance, time series with bursty or noisy characteristics. Having said that, we believe that smoothed random walks are a relevant type of test data for the decomposition function that we have chosen (modulo division) based on the tasks that we evaluate (discrimination and estimation). Also, previous studies [39], carried out their evaluation with this type of data, and using the same data type allows to compare the results with the existing findings. Nevertheless, other types of data (e.g. noisy data) can be used for more relevant tasks (e.g. trend detection) with different decompositions (e.g. frequency separation).

In all evaluated time series, values varied between 0 and 100 and the width of the charts corresponded to the length of the time series. The charts were generated on HTML5 Canvas elements using JavaScript. In an attempt to preserve perceptual uniformity, we chose polar CIELUV (HCL) color space for chart generation. However, we respectively refer to chroma and lightness as *saturation* and *value* due to familiarity of users and readers with these terms.

5.3 Evaluated Visual Mappings

Our initial selection of visual mappings consisted in six visualization techniques (see Figure 5.1) from which, line charts and Horizon Graphs were adopted from the

existing techniques. In the other four techniques in form of [Visual Channel 1]–[Visual Channel 2], we used composite visual mapping where quotient of the modulo division was mapped onto [Visual Channel 1], and the remainder onto [Visual Channel 2] similar to our reference point, the Horizon Graphs. In our selection of visual channels, we consider the categorization of channels into *geometric* and *optical* channels, as described in Section 2.2.7.

Line chart:

We included the line chart as the most basic and familiar time series visualization technique. As suggested by Tufte [81], a shaded, high contrast line chart might be better than a floating thin line.

Horizon Graph:

As the most relevant technique to our approach, horizon graphs are included in our list as the reference point. Many coloring schemes is conceivable for differentiating layers in Horizon Graphs. Hue variation for different layers can suffer from the lack of natural ordering of hues. We adopted the more common color value (*Lightness* in HCL color space) sequencing due to its better contrast than saturation sequencing (*Chroma* in HCL color space). In this way, horizon graph implements a composite visual mapping of Value–Position. We selected the final configuration by conducting a pilot test described in Section 5.4.

Position–Saturation:

This mapping consists of the same combination of visual channels as in horizon graphs but in the reversed order. The position visual channel was used for order of magnitude information and thus, in discrete steps. Saturation varied continuously to map the remainder of the modulo operation. This design was included to examine the effect of the order of the visual channels in a composite mapping.

Hue–Saturation:

There is no intuitive order of hues that is acceptable to everyone. Therefore, continuous variation of hue (“the rainbow”) should be avoided for quantitative data. In discrete quantities (e.g. order of magnitude) the effect is more tolerable but still is context and user dependent. For example, consider two green and red hues used in a unspecified context. While some users may associate green to higher values and red to lower values (as in

battery charge indicators), others may attribute red to values exceeding an authorized threshold, and greens to lower values below that limit. Also, users with color deficiency may misinterpret the data mapped on hue. We used a blue-orange color scheme for this design to reduce these effects.

Double-banded:

We are aware of interactions between colors and graphical elements and their effects on visual perception [6]. Hence, this design spatially separates the two visual channels into two dedicated bands. Continuous values of the remainder of modulo division naturally need more spatial resolution than discrete values of the order of magnitude. Hence, we have encoded the order of magnitude on the upper band, which has half the width of the lower band for smaller fluctuations. The upper band can be filled with color value sequencing or discrete hues with equal color value resulting in Value-Value and Hue-Value composite mappings, respectively. We chose the final design among other alternatives after running a pilot test detailed in Section 5.4.

Texture-Saturation:

We decided to include this design in our list because of the particularity of the texture channel. Although considered as an optical channel, its impact is directly related to the geometric properties of its motifs. In our opinion, this positions texture on the borderline between optical and geometric visual channels and worth studying for spatial efficiency.

Among the visual designs detailed above, line charts, Horizon Graphs, and Position-Saturation rely on geometric visual channels (i.e. position and size), Hue-Saturation and the Double-banded design rely solely on optical channels and finally, we consider Texture-Saturation on the borderline of the two categories. This varied selection of the visual designs was intended to examine the effect of the type of visual channels in space-constrained visualizations.

5.4 Pilot Studies

The selected combination of visual mappings can be implemented in different ways, using various coloring schemes and configurations. We conducted a series of pilot studies to choose our definitive visual designs for the main experiment. We are aware of the statistical weakness of the results of the evaluations with such small sample sizes. Yet, we opted for these tests as an objective aid to refine our final experimental design.

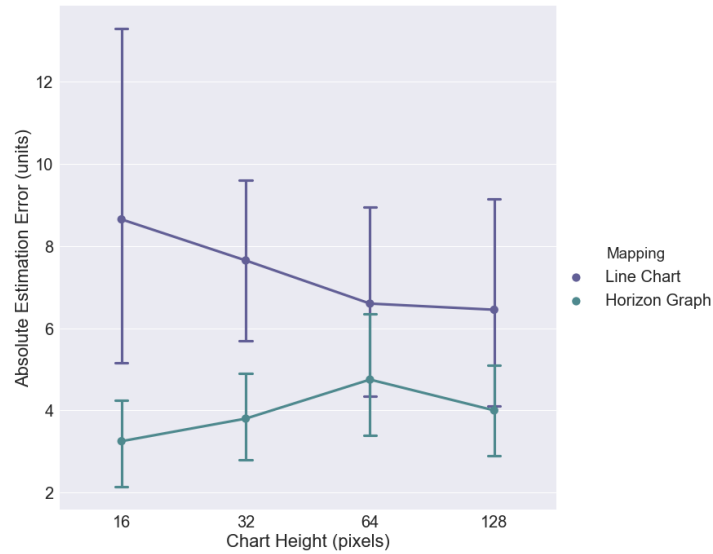


Figure 5.2: Absolute estimation error for line charts and Horizon Graphs in pilot I. Brackets show confidence intervals at 95 percent.

5.4.1 Pilot I: Mapping Selection

For statistical reasons, we had to reduce the number of variables in our experimental design. Toward this aim, in a first pilot study we tried to identify the most pertinent composite mappings among the six initial visual mappings in Figure 5.1. This experiment also allowed us to compare composite mappings with line chart as the reference visualization technique. All charts had the same length of 2000 pixels and we tested four chart heights (128, 64, 32, and 16 pixels). At 128 pixels height, the chart measured 233×15 mm on the screen.

We conducted this experiment with 5 participants (4 male and 1 female) among people working in our lab. The experiment consisted in 120 trials per subject (6 mappings × 4 sizes × 5 trial per condition), divided into 5 experimental blocks. The experimental protocol is described in Section 5.2.1.

For all visual designs, discrimination accuracy averaged above 95%, so we focused on estimation accuracy rates. Among our four candidate designs, Hue-Sat had the highest levels of estimation error, while showing no advantage in terms of speed. Based on these results (even though not statistically significant) and the discussions with participants, we decided to exclude this mapping from the main experiment.

Comparing line charts and horizon graphs (Figure 5.2), we observed one of the effects already reported in [39]. As in the preceding study, accuracy decreased at smaller chart heights for both line charts and horizon graphs. However, this effect was most prominent for line charts. Given this observation, we consider our first hypothesis (**H1**) as valid and

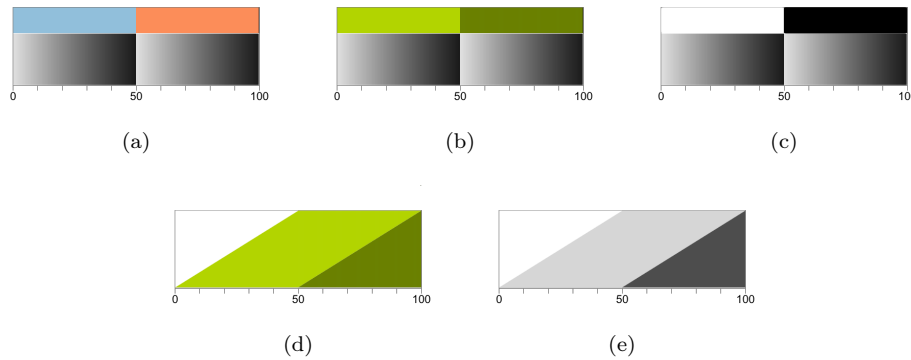


Figure 5.3: Different color schemes tested in pilot II. We finally selected (a) for the double-banded design and (d) for horizon graphs in the main experiment.

we exclude line charts from our final experiment and keep horizon graph as the reference point.

5.4.2 Pilot II: Color Scheme Selection

The choice of color can impact the performance of the visualizations and hence, we conduct the second pilot experiment to choose the color schemes for horizon graphs and the double-banded design. It has been established that the maximum spectral sensitivity of the human eye under daylight conditions is at a wavelength of 555 nm [35] (i.e. green hues). For this reason, we chose shades of green (varying the color value) as a candidate color scheme in addition to monochromic (no hue effect and maximum color contrast) for horizon graphs and the double-banded design. We also included a third color scheme with red and blue hues (with constant and equal color value) for the double-banded design (the upper band) resulting in a Hue-Value composite mapping.

We conducted a pilot test on the 5 color schemes (2 for horizon graphs and 3 for the double-banded) with 3 participants that have not had participated in the previous pilot. We tested four chart heights (1, 1/2, 1/4, and 1/8 scale factors) where the tallest chart corresponded to 128 pixels and physical size of 15 millimeters on our screen. In horizon graphs, we saw no difference between green gradients and chromatic in terms of accuracy. This can be explained by the fact that with both color schemes, users solely relied on the (equal) difference of color value and presence of hue had a minimal role. Yet, in the discussions with participants they reported their preference for colored horizon graphs over the chromatic version. Thus, we selected green color scheme for horizon graphs in the main experiment.

Regarding the double-banded design, although the accuracy was similar for all color schemes, surveyed users overwhelmingly favored red and blue scheme over other options.

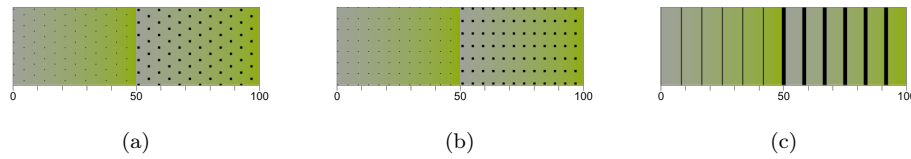


Figure 5.4: We tested three textures for the Texture–Saturation design. We finally selected (a) for the main evaluation.

In discussions, they reported that they found it extremely fast and easy to detect red zones (and therefore, higher values) without confusing the order of magnitude of the two colors. Interestingly, some participants found chromatic double-banded design confusing, as they considered darker regions corresponding to smaller values (i.e. “off” pixels) and lighter regions to larger values (i.e. “on” pixels). Therefore, we selected the double-banded design with blue/red as the final design for the main experiment.

5.4.3 Pilot III: Texture Selection

Our third pilot study aimed to finalize our Texture–Saturation design among three candidate textures (see Figure 5.4). For space efficiency reasons, we aimed to minimize the occlusion of the superimposed motifs. Therefore, we avoided complex motifs or inclined lines which take more space than straight lines or dots. We tested three chart heights (1, 1/4, and 1/8 scale factors) where the tallest chart corresponded to 64 pixels and physical size of 8 millimeters on our screen.

Six persons, 4 males and 2 females, performed the discrimination and estimation task on 45 trials (15 trial per condition). Accuracy was similar for all three cases. Nevertheless, a few observations by the participants convinced us to choose the texture 5.4a. Some participants reported experiencing uncomfortable vibratory sensations similar to those described in Section 2.2 with 5.4b and 5.4c due to the repetitive pattern. Impact of spatial frequency (e.g. “spreading” effect) on visual and color perception is well documented in the literature [83]. These effects can heavily influence the perception of colors between the bars in 5.4c. Some users also reported that in some instances, a marker was positioned exactly on a bar which naturally covered the underlying color and so, they based their estimation on the color in adjacent zones.

5.5 Main Experiment

We used the indicative findings of our pilot tests to finalize the design of the main experiment. As a result, we removed line charts (Figure 5.1a) and Hue–Saturation (Figure 5.1f)

for the reasons discussed earlier. The four remaining visual mappings, i.e. horizon graphs, Position–Value, Hue–Value, and Texture–Sat were used as they appear in Figure 5.1.

In order to maintain a reasonable number of conditions, we only tested three scale factors (1, 1/4, 1/8) where factor 1 corresponded to a height of 64 pixels and 8 millimeters on our screen. All charts measured 2000 pixels wide corresponding to 233 millimeters in physical size on our screen. We recruited twenty unpaid participants (17 males, 3 females) aged between 22 and 39 (with median age of 27.5) among people working in our laboratory. All participants were degree holders in computer science and had normal or corrected vision and no color deficiency. For all trials, participants were told to keep a fixed viewing distance of about 50 centimeters from the screen.

During the training phase preceding the main experiment, we introduced each visual design to the participants. Next, they performed the discrimination and estimation task on each training set ordered from the tallest to the shortest chart height with three trials per condition. This tutorial consisted of a total of 4 (mapping) \times 3 (size) \times 3 (trials per condition) = 36 trials per participant. During this stage which lasted about 15 minutes, participants could familiarize with all conditions and develop their strategies for the demanded tasks.

The main evaluation that followed the tutorial consisted in a total of 4 (mapping) \times 3 (size) \times 7 (trials per condition) = 84 trials per participant. Trials were divided into four experimental blocks of 21 trials and the order of the trials in each block was randomized for each participant. In order to make sure that participants make use of both visual channels, the two comparison points of each trial were intentionally located in different orders of magnitude. We also balanced the 7 trials of each condition (mapping and size, combined) for value difference between the comparison points. Participants performed the experiment with the procedure detailed in Section 5.2.1. At the end of the experiment, participants filled in a questionnaire about the evaluated techniques.

5.5.1 Results

For all conditions, discrimination scores averaged 98% or higher. Therefore, we focus on interpretation of other measurements. Linear regression lines in Figure 5.11 report overall observed trends for each visual mapping regarding different measures. For more in depth analysis of these trends, we used multi-factor analysis of variance (ANOVA) to identify statistically significant differences in performance. When significant differences were found, we used t-tests and Tukey’s honest significant difference test (HSD) to identify pair-wise significance.

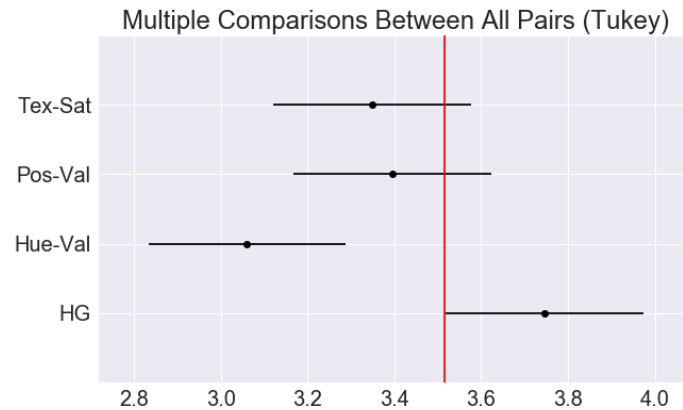


Figure 5.5: Tukey's honesty significant difference (HSD) comparisons show significant difference in discrimination time between Hue-Value and horizon graphs (pair-wise $p < 0.001$). Measures are in seconds.

Discrimination Time

ANOVAs showed a significant effect of visual mapping on discrimination speed with $F(3, 57) = 15.894, p < 0.001$. Tukey HSD comparisons in Figure 5.5 showed that Hue-Value is significantly faster than horizon graphs regarding this criterion ($t(57) < 0.001$). Furthermore, pair-wise t-test showed that at the smallest chart height (4 pixels), participants performed significantly faster with Hue-Value and Texture-Saturation than horizon graph ($t < 0.01$).

Estimation Error

Figure 5.6 shows distributions of estimation errors by each visual mapping across all chart heights. Even though they look similar, ANOVAs showed a significant effect of visual mapping ($F(3, 57) = 15.021, p < 0.001$). Post-hoc Tukey's HSD comparisons identified Texture-Saturation as the only significantly different visual mapping in terms of estimation error averaged across chart heights (see Figure 5.8).

We found a significant difference in estimation accuracy across different chart heights ($F(2, 57) = 15.033, p < 0.001$). Tukey's HSD comparisons confirm significant differences between the smallest chart height and the two taller heights (see Figure 5.7). In order to identify the exact effects of size, we calculated t-tests which confirm that even if the differences in estimation accuracy is insignificant between 1/16 and 1/4 scale factors, horizon graph has significant worse accuracy in those chart heights than the tallest chart height (with pair-wise significances $p < 0.05$).

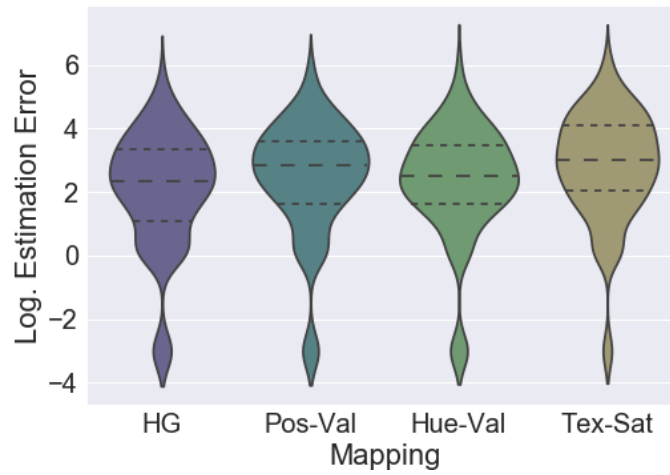


Figure 5.6: Distribution of logarithmic estimation error by visual mapping across all chart heights

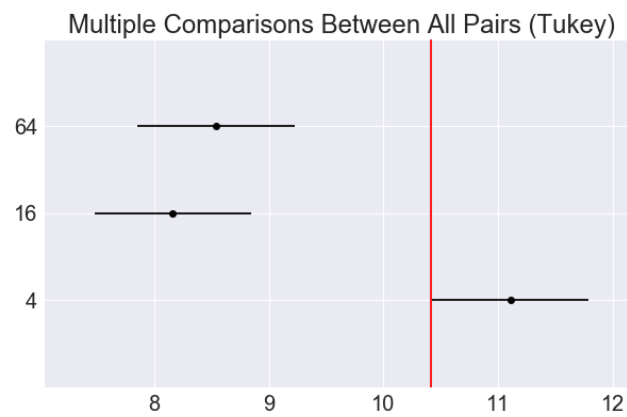


Figure 5.7: Tukey's honest significant difference (HSD) comparisons show significant higher errors in the smallest chart height. Errors are reported in units.

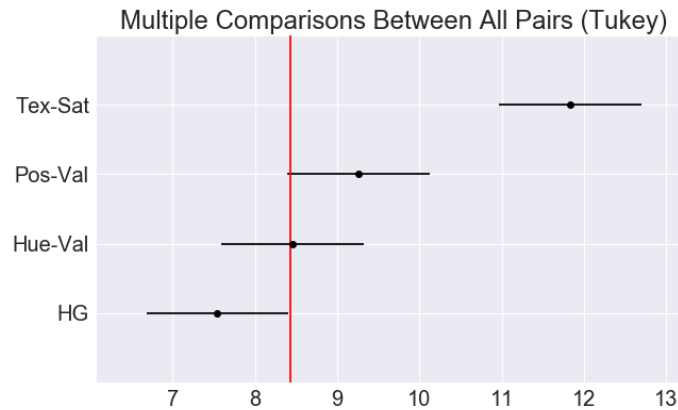


Figure 5.8: Tukey’s honesty significant difference (HSD) comparisons show no significant difference in estimation accuracy between horizon graphs and other mappings except Texture–Saturation (pair-wise $p < 0.001$). Errors are reported in units.

Estimation Time

We found a significant effect of chart height in estimation error ($F(2, 38) = 4.42, p = 0.012$). Tukey HSD comparisons reveal that estimation time decreases with chart height and the difference is significant between the smallest and tallest chart height (see Figure 5.9). We did not find any significant effect of visual mapping on estimation time ($F(3, 57) = 1.94, p = 0.14$). However, we found a weak significance of mapping effect between horizon graphs and Texture–Saturation at the tallest chart height by looking at t-test results (pair-wise $p = 0.04$).

5.5.2 Results from the Questionnaire

Following the evaluation, participants filled in a questionnaire in which, they rated and commented the four mappings. Participants were asked to give a note out of 0 to 10 (10 being the highest note) to each mapping for discrimination and estimation tasks in small and large chart heights, respectively. They also rated their overall satisfaction and their familiarity with each technique using the same scale.

Horizon graph was the most familiar visual mapping for the most of the participants, seconded by Position–Value. Other techniques were much less known to the participants, even though five of the participants had already participated in our pilot experiments. Participants reported Hue–Value as their favorite technique, followed by horizon graph (see Figure 5.10). For the discrimination task, participants in average preferred Hue–Value regardless of chart’s height. Horizon graphs was most favored for estimation task in large chart sizes and Hue–Value in small chart heights.

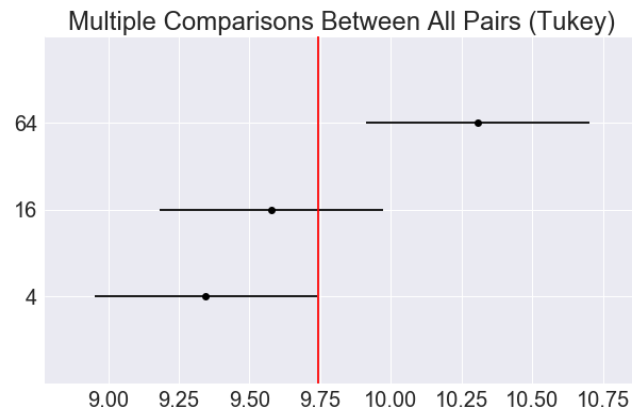


Figure 5.9: Tukey's honesty significant difference (HSD) comparisons show faster estimation at smaller chart heights.

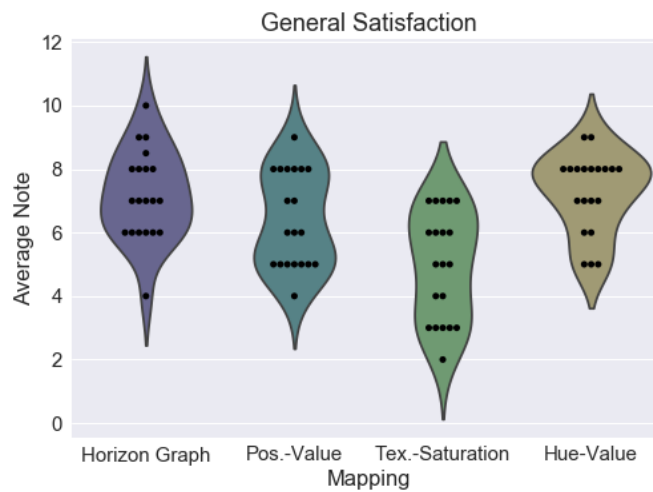


Figure 5.10: General Satisfaction for each visual design. Each black point represents one participant's answer.

5.5.3 Discussion of the Results

In the first pilot study, we witnessed partial reproduction of the earlier findings from comparison of line charts and horizon graphs. We consequently accept our first hypothesis (**H1**) as valid.

Our results confirm that at the tallest chart heights (1 scale factor), horizon graphs have the best estimate accuracy compared to other designs ($p < 0.05$ for all pair-wise comparisons). The accuracy is impacted by reduction of charts' height, yet the effect is not identical for all designs. Accuracy differences between the smallest and tallest chart heights were most significant for horizon graphs (pair-wise $p = 0.011$) and Texture-Saturation (pair-wise $p < 0.001$) and not significant for Hue-Value and Position-Value (see also Figure 5.11a).

At the smallest chart heights, horizon graph lost its absolute superiority and we showed that accuracy of Hue-Value, Position-Value, and horizon graphs converge (as seen in Figure 5.11a). This validates our second hypothesis (**H2**) that in small chart heights, our visual designs are on a level with horizon graphs.

We did not observe significant impact of chart height on discrimination time and Hue-Value was significantly faster than horizon graphs across all chart heights (see Figure 5.11c for trends). Similar to findings in previous works, we found that estimation time decreases as chart height decreases and that it converges for all mappings in the smallest chart height (see Figure 5.11b). This confirms our third hypothesis (**H3**) that in small chart heights mappings such as Hue-Value offer the same levels of accuracy as horizon graphs and yet, offer advantages in terms of speed.

5.6 Design Implications

Based on the described empirical results and discussions with participants of our study, we offer the following design recommendations for effective use of the introduced visual designs and composite visual mapping in general.

Optical visual channels gain advantage as chart size decreases

Users of visualization systems have naturally more confidence in comparing physical attributes like size and position than perceptual attributes like color. At least for the task we tested, we showed that accuracy difference is reasonable in large sizes and insignificant in small sizes. We suggest that optical visual channels can be used more widely in cases where perception of size-related attributes are impacted, i.e. when chart size decreases.

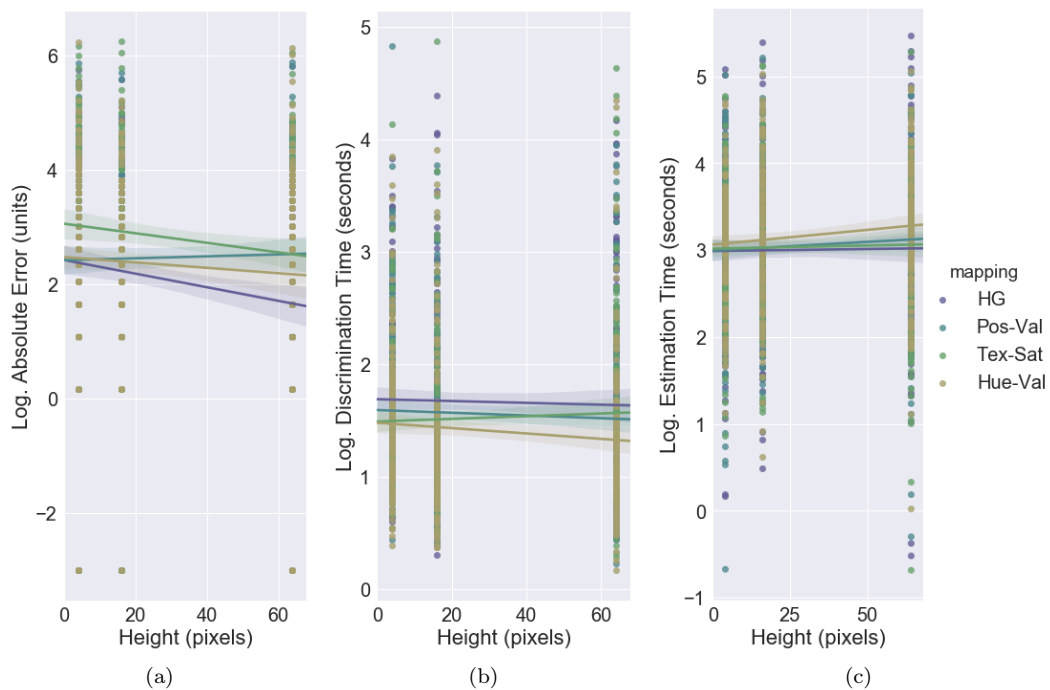


Figure 5.11: Linear regression lines with 95% confidence intervals for different mappings across the tested chart heights showing trends for (a) estimation errors, (c) discrimination time, and (b) estimation time.

Using texture in composite mapping is tricky

Blending texture and variation of other optical channels should be considered with a grain of salt. Many patterns for texture can unnecessarily occlude information, introduce visual clutter, or create undesired effects (e.g. vibratory sensations). Some other patterns may appear too minimalistic and difficult to distinguish in small sizes. Also, texture can seriously interfere with perception of the underlying color.

Hue coding is fast, but context should be considered

Mapping information on the hue channel can increase discrimination speed even in small chart sizes. However, interpretation of hues is context dependent and some hues (e.g. bright reds and yellows) are easier to spot. For example, participants in our experiment found it extremely easy to spot a red region as it rapidly raises *a red flag*. As the participants were asked to report the higher value, the selected order of hues (reds > blues) was in coherence with this comparison task.

Geometric channels can still be used in small chart heights

Horizon graph is still a powerful space-efficient technique for time series visualization and further improvements in screen resolutions can ease the use of geometric visual channels in small screen sizes. However, size-related visual channels may suit better for discrete information (e.g. order of magnitude) or data with less fluctuation (e.g. trend component of a time series) in small chart sizes.

5.7 Limits and Perspectives

Our design recommendations, nevertheless, are valid with some limitations. Due to experimental and statistical constraints, we only investigated common but narrow tasks of comparison and estimation and only examined one type of data. Visual analysis of time series involves many other tasks that have not been evaluated in this study. For future studies, modified visual designs (e.g. frequency separation instead of modulo division) can be evaluated for corresponding appropriate tasks (e.g. trend detection) with their routinely associated data (e.g. noisy data).

Another limitation of this study is that our implementation of the selected combinations of visual channels are only some among many possible visual designs. As observed in the results of our pilot studies, the performance of new mappings depend on their design and could be further enhanced.

Also, profile of the participants of the study was limited to people working in our laboratory. Half of the participants had already participated in one of our earlier studies or pilots. This could introduce an involuntary bias. However, we did not observe common behavior or better performance among this group of participants.

In our study we only examined two layered (i.e. values divided into two levels using modulo division). As in earlier studies on effects of layering in horizon graphs [39], future work can evaluate the effectiveness of our visual design in higher numbers of layering. While our study report important trends on the effects of chart height and visual mapping, more participants with more diverse backgrounds can improve the statistical robustness of our analysis.

5.8 Summary

In this chapter, we presented our second empirical study on a selection of composite visual mappings. Similar to horizon graphs, these visual designs combine two visual channels for visualizing the same data attribute, in order to separate different facets of

data and increase the information bandwidth. We pit our designs against horizon graph in discrimination and estimation tasks and observed that while horizon graph has superior performance in larger chart heights, it is significantly impacted by chart height reduction. On the other hand, our proposed visual designs that mainly use optical visual channels instead of geometric channels have proven performances similar to horizon graph in small chart heights. One of our visual designs, Hue–Value, even offers faster discrimination speed while remaining as accurate as horizon graphs in small chart heights.

We note that despite the general mistrust of optical channels for quantitative visualization due to their perceptual nature, they perform as good as geometric visualizations in space-constrained visualizations in the tasks we evaluated. We evaluated the effects of chart height and visual mapping on discrimination and estimation tasks and based on the results, we proposed several design recommendations (advantages of optical channels in low resolutions, the blending effects, reported artefacts, etc.).

We learned from our evaluations and discussions that the new visual designs are easy to learn by novice users and mostly positively received by participants. We should remember that users are more familiar with geometric representations of data and that further training may improve their performance with our proposed designs.

Conclusion

Summary of Contributions

In response to the existing challenges in visualization of large multidimensional time series, we aimed to come up with visual designs based on an alternative approach to visual mapping. We intended to verify that our approach to visual mapping facilitates visual analysis and is space efficient by design. Such visual designs would be beneficial for improving visual analysis of large amounts of data in space constraint situations.

To this aim, we reviewed the literature of time series visualization with a focus on visual mapping and the visual channels. We noticed from this review that most of the existing mappings depend mainly on geometric visual channels (e.g position and size) to communicate quantitative information. In this manner, the use of optical visual channels (e.g. color) is mostly limited to redundant encoding of the same information in order to attract the attention of the user (e.g. a graphical mark which is larger in size is also darker in color). This choice may be explained by the advantages of physical attributes such as size over more perceptual attributes like color. However, physical attributes are naturally more sensitive to the space constraints of visualization and optical channels could compensate for compromises in those situations.

Nevertheless, our goal was not to exclusively use optical channels to replace geometric channels. We got interested in combinations of visual channels which would distribute the information load. We identified few techniques which implement what we called composite visual mapping. This approach consists in decomposing each data value into several sub-values, each conveying information on specific facets of data. The choice of the decomposition operation is made based on the interest of the user and the analytical task at hand. Each resulting sub-value is then mapped onto a separate visual channel. This one-to-many approach dedicates full bandwidth visual channels to each facet of data. With the visual channels selected properly, users can exclusively attend the desired facet of data and hence, increase their precision and speed in dealing with that property of data.

A systematic study of the design space for such approach is the first contribution of this thesis. We achieved this goal thanks to an exploratory table which allowed us to identify the potentially interesting composite mappings for further evaluation.

Our second contribution is an evaluation of a selection of promising composite visual mappings where we compare them with the well established horizon graphs. Our results globally hinted that other composite mappings can produce similar accuracy to that of horizon graphs, while improving users' speed in discrimination and estimation tasks.

A second empirical study based on the findings from our first evaluation constitutes our third contribution in this thesis. In this study, we evaluate the effects of chart height across different composite visual mappings. The results demonstrated that while horizon graphs were superior to our techniques at higher chart heights, they lose their advantage in smaller heights. In addition, at least one visual design (*Hue-Value*) performed faster than horizon graph. The results demonstrated that even without any previous knowledge of the new techniques, they lead to similar, and in some aspects, better performance in discrimination and estimation tasks at small chart heights. Small chart heights is a recurrent situation when the display space is limited due to the size of the screen or the amount of data. Based on the findings of this series of evaluations, we came up with several design guidelines on combination of visual channels.

Perspectives

In our empirical studies, we reproduced the experimental protocol of the previous studies on horizon graphs [39]. We hence only evaluated the new visual mappings in limited tasks, i.e. discrimination and estimation. While the results show similar accuracy among the new techniques and horizon graphs in those tasks, further studies are necessary to evaluate the performance of the new mappings in other analytical tasks, e.g. trend detection.

Also, we exclusively used artificially generated random walks as test data for our evaluations. While this choice helped us to control experimental parameters, it obviously does not cover all types of data. Further studies should be carried out to evaluate new composite mappings for other types of data, e.g. bursty data. Also, our empirical studies were oriented toward the precision of the evaluations. Realistic evaluations should be carried out with realistic data in order to be able to comment on the effectiveness of our techniques in real applications.

In future research, it would be an asset to investigate combinations of more than two visual channels. Increasing the number of visual channels may also require new interaction techniques (e.g. data visualization in VR/AR). A generalized version of our exploratory table allows to study such higher-dimensional combinations.

Another future research avenue may investigate the use of composite visual mapping for types of data other than time series, e.g. geospatial data. Similar to composite visual mapping in time series which allows to separate evaluation of different features of data throughout time, applying the same approach to maps allows to visualize the geographical distribution of the main features, e.g. order of magnitude.

Finally, real-world time series visualization tools are promising areas for implementation of our proposed designs. Future works on visualization tools can combine composite visual mappings with the approaches that were excluded from the scope of this thesis. For instance, a user may need to overview a large number of time series in a limited space (*See the whole*). This can be accomplished with data management techniques (e.g. aggregation) and our composite mappings thanks to their space efficiency properties. Furthermore, the user can apply additional interactive space efficiency techniques (e.g. distortion lenses) to dig into more details.

Implementing these techniques allow to view and explore larger datasets. This step will allow to assess scalability, advantages, and limits of our visual designs in practice.

Appendices

Appendix A

Data Generation

For both experiments, we used random walks to generate the evaluated time series. The script below shows the computation of constrained random walks used in Chapter 5. We used Moving average to smooth the generated data. The smoothing factor remained constant for all time series.

Script

```
1 import numpy as np
2 import matplotlib.pyplot as plt
3
4 w = 2000 # length of time series
5 subt = 12 # number of unique conditions
6 k = 6 # smoothing factor
7
8 dlen = 7 # number of repetitions per condition
9 deltas = [100*i/(dlen+1) for i in range(1,dlen+1)] # bins for value differences
10         (deltas)
11 noise = np.random.uniform(-5,5,dlen) # noise added to delta bins
12
13 total = subt*dlen # total number of time series to generate
14
15 x_t = .2*w # position of marker T
16 x_b = .8*w # position of marker B
17 nu = int(x_t) + k - 1 # number of points before the first marker
18 nc = int(x_b - x_t) - 1 # number of points between the two markers
19
20 outfile = open('DATA.txt', 'w')
```

```
20
21 # Calculate possible variation of delta
22 def getMargin(d):
23     m = 20 - abs(0.5*(d-50))
24     return m
25
26 # Get values at two ends of the series
27 def getEnds():
28     nends = 2
29     ends = []
30     lows = np.random.uniform(0,50,nends)
31     highs = np.random.uniform(50,100,nends)
32     for i in range(nends):
33         one = [lows[i], highs[i]]
34         np.random.shuffle(one)
35         ends.append(one)
36     return ends
37
38 # Constrained random walk from y0 to y1 in n steps
39 def walk(y0, y1, n):
40     y = y0
41     res = [y0]*(n+1)
42     unifs = np.random.uniform(0,1,n)
43     for i in range(0,n):
44         theta = (1-(y1-y)/(n-i))/2
45         if ((unifs[i]<=theta and y>=1) or y>99): y -= 1
46         else: y += 1
47         res[i+1]=y
48     return res
49
50 # Generate a pair of random walks with real difference "delta"
51 def generatePair(delta,offset):
52
53     lo = 50 - delta/2 + offset
54     hi = 50 + delta/2 + offset
55     targets = [lo, hi]
56     target = targets[:]
57     np.random.shuffle(target) # Shuffles order of the two marked values
58
59     endPoints = getEnds()
60     C21 = walk(endPoints[0][0], target[0], nu)
61     C22 = walk(target[0], endPoints[0][1], nc+nu+1)
62     C1 = C21 + C22 # concatenate two pieces of the first series
63
64     C21 = walk(endPoints[1][0], target[1], nc+nu+1)
65     C22 = walk(target[1], endPoints[1][1], nu)
```

```
66     C2 = C21 + C22 # concatenate two pieces of the second series
67
68     return [C1, C2]
69
70 # Moving Average for smoothing the generated random walks
71 def smooth(res):
72     data = []
73     order = 2*k+1
74     dw = len(res)
75     for i in range(k,dw-k):
76         value = 0
77         for j in range(-k,k+1):
78             value += 1/order * res[i+j]
79         data.append(value)
80     return data
81
82 # Write data to file
83 def writeData(ts):
84     data = smooth(ts)
85     for x,y in enumerate(data):
86         print('% 7.4f' % y, file = outfile)
87
88 # Generate and write pair of series
89 def writePair(d,r):
90     pair = generatePair(d,r)
91     writeData(pair[0])
92     writeData(pair[1])
93
94 ds=[] # real deltas
95 for i in range(dlen):
96     d = deltas[i] + noise[i]
97     m = getMargin(d)
98     sig = np.random.uniform(-m,m,subt)
99     for j in range(subt):
100         ds.append(d)
101         r = sig[j]
102         writePair(d, r)
103
104 # Plot deltas distribution
105 plt.hist(ds, bins=100)
106 plt.title("Delta Distribution")
107 plt.show()
```

The algorithm above is illustrated in Figure A.1. Each series is composed of two constrained random walks:

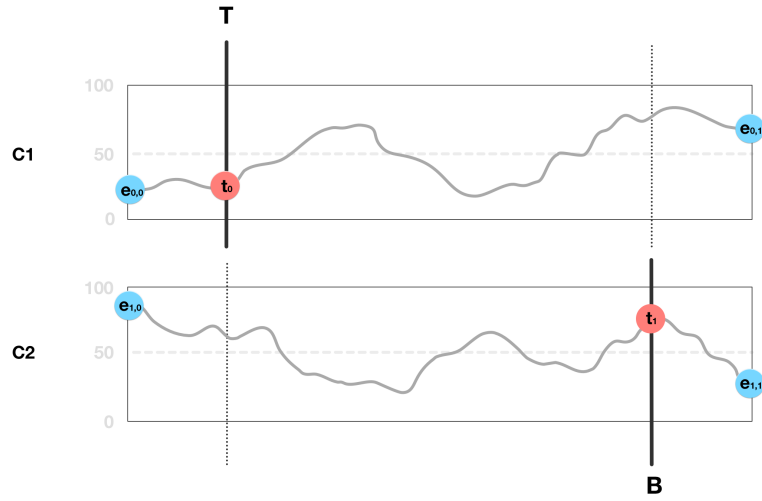


Figure A.1: Generating a pair of random walks by its constraining points.

- $C_1 = C_{11}(\text{from } e_{0,0} \text{ to } t_0) + C_{12}(\text{from } t_0 \text{ to } e_{0,1})$
- $C_2 = C_{21}(\text{from } e_{1,0} \text{ to } t_1) + C_{22}(\text{from } t_1 \text{ to } e_{1,1})$

The function `getEnds()` randomly draws the end points $e_{0,0}$ and $e_{0,1}$ (as well as $e_{1,0}$ and $e_{1,1}$) from opposite sides of the 50% bar. The target points t_0 and t_1 denote the values at marker T and B , respectively. The difference between the two values t_0 and t_1 is equal to the *delta* in the algorithm. Note that this difference of values can vary from the delta bins because of the added random noise. For each passed *delta*, the added noise is constrained (variance calculated by `getMargin()`) in a way to ensure that the two target points are always located in opposite sides of the 50% bar. The order of the end points and target points of series are randomized throughout all pairs.

Real Value Difference Distribution

Figure A.2 shows the distribution of real value differences between the two evaluation points (T and B) in the experiment described in Chapter 5. We made sure that the two reference points' values are located in opposite sides of the 50% bar, i.e. in two different layers in a two layers configuration. This choice was intentionally made to guarantee that viewers take advantage of both visual channels. As seen in the script, we added noise to the *deltas* to variate them around the initial 7 bins (see Figure A.2).

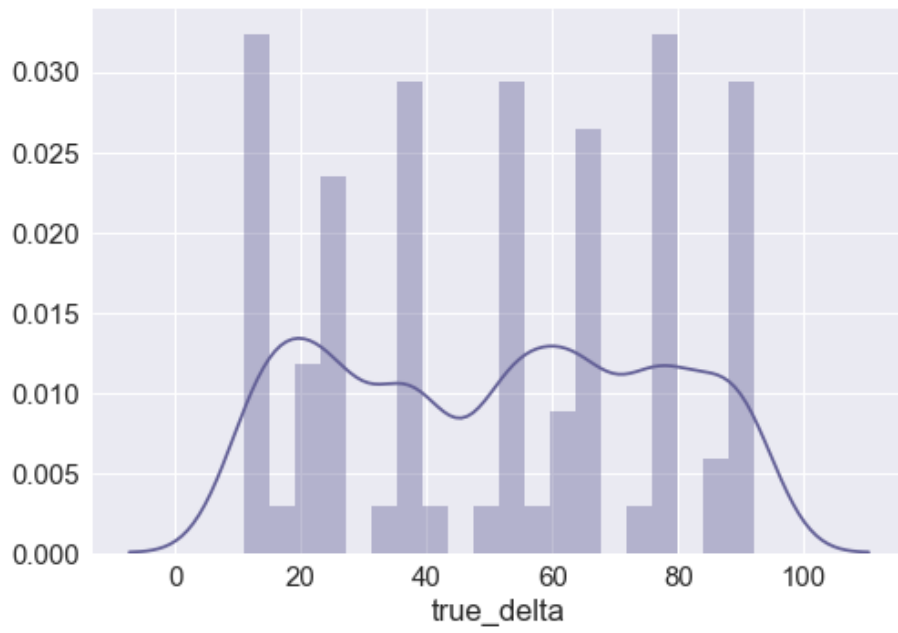


Figure A.2: Distribution of real value differences (*deltas*) in pairs of time series evaluated in Chapter 5.

Appendix B

Construction of an Example Composite Visual Mapping

Figure B.1 illustrates construction of a composite visual mapping (*Position–Saturation*) using modulo operation. First, each data value is decomposed to its components according to Equation 3.2. Next, each new sub-value is conventionally mapped onto *Position* and *Saturation* visual channels. The two representations are multiplexed to obtain the final composite mapping. However, the multiplexing step is not unique and can be achieved in multiple ways. For instance, here we could also put the two primary mappings together in two separate charts. The final choice of multiplexing was mostly made based on space efficiency. The same process is valid for construction of other composite mappings.

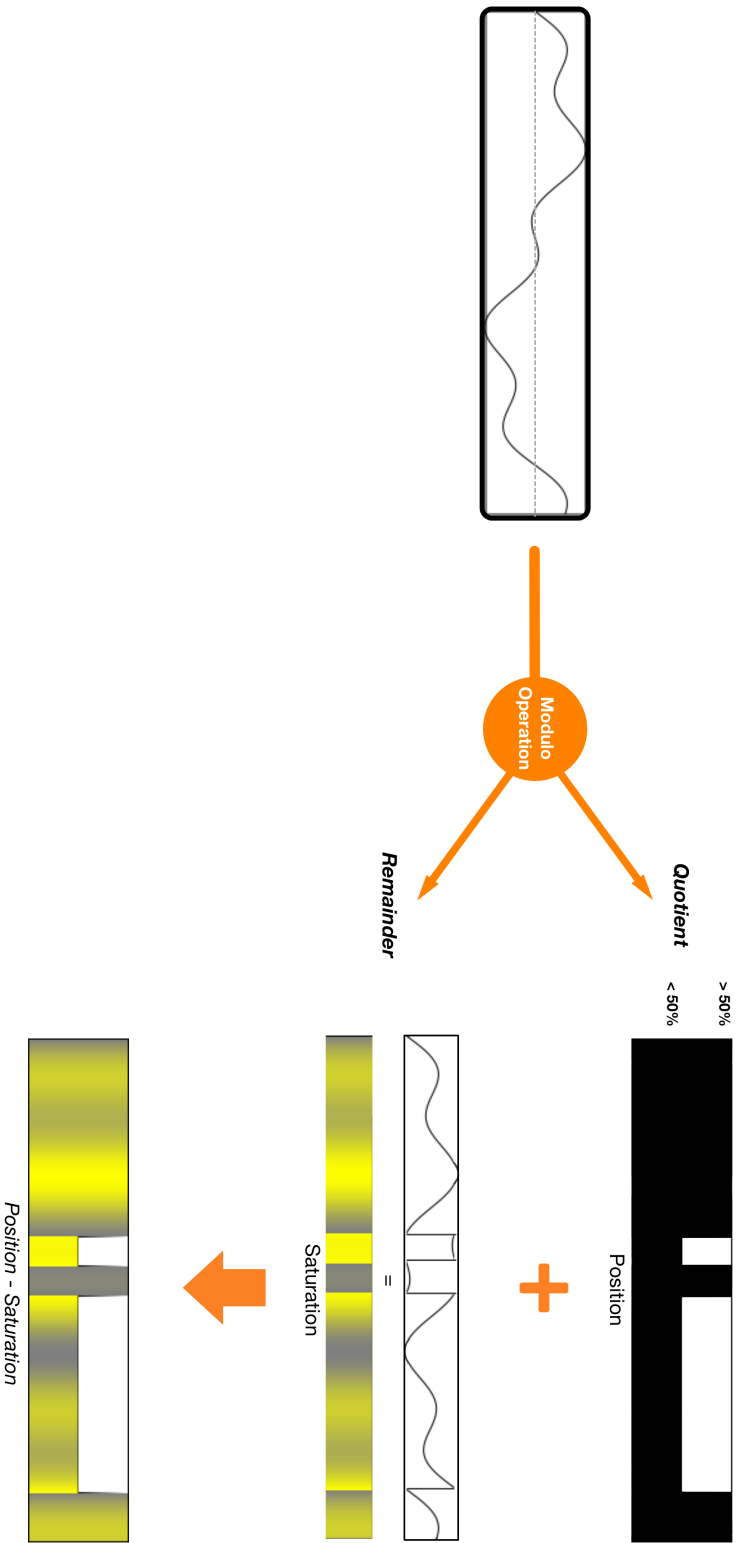


Figure B.1: Construction of the visual mapping *Position-Saturation* using modulo operation.

Bibliography

- [1] AHLBERG, C., WILLIAMSON, C., AND SHNEIDERMAN, B. Dynamic queries for information exploration: An implementation and evaluation. In *Proceedings of the SIGCHI conference on Human factors in computing systems (1992)*, CHI '92, ACM, ACM, pp. 619–626.
- [2] AIGNER, W. *Interactive Visualization and Data Analysis : Visual Analytics With a Focus on Time*, 2013.
- [3] AIGNER, W., MIKSCH, S., MÜLLER, W., SCHUMANN, H., AND TOMINSKI, C. Visualizing time-oriented data-A systematic view. *Computers and Graphics (Pergamon)* 31, 3 (2007), 401–409.
- [4] AIGNER, W., MIKSCH, S., SCHUMANN, H., AND TOMINSKI, C. *Visualization of Time-Oriented Data*, 1st ed. Springer Publishing Company, Incorporated, 2011.
- [5] ALBERS, D., CORRELL, M., AND GLEICHER, M. Task-Driven Evaluation of Aggregation in Time Series Visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (2014)*, ACM, pp. 551–560.
- [6] ALBERS, J. *Interaction of color*. Yale University Press, 2006.
- [7] ANDERSON, T. *An Introduction to Multivariate Statistical Analysis*. Wiley Series in Probability and Statistics. Wiley, 2003.
- [8] ANKERST, M., BERCHTOLD, S., AND KEIM, D. A. Similarity clustering of dimensions for an enhanced visualization of multidimensional data. In *Proceedings of the 1998 IEEE Symposium on Information Visualization (Washington, DC, USA, 1998)*, INFOVIS '98, IEEE Computer Society, pp. 52–.
- [9] ANKERST, M., KEIM, D. A., AND KRIEGEL, H.-P. Circle segments : A technique for visually exploring large multidimensional data sets. In *Visualization '96, Hot Topic Session, San Francisco, CA, November, 1996 (1996)*.
- [10] BADE, R., SCHLECHTWEG, S., AND MIKSCH, S. Connecting time-oriented data and information to a coherent interactive visualization. *Proceedings of the 2004 conference on Human factors in computing systems - CHI '04* 6, 1 (2004), 105–112.

-
- [11] BEDDOW, J. Shape coding of multidimensional data on a microcomputer display. In *Proceedings of the 1st Conference on Visualization'90* (1990), IEEE Computer Society Press, pp. 238–246.
- [12] BEDERSON, B. B., SHNEIDERMAN, B., AND WATTENBERG, M. Ordered and quantum treemaps: Making effective use of 2d space to display hierarchies. *ACM Transactions on Graphics (TOG)* 21, 4 (2002), 833–854.
- [13] BERLIN, B., AND KAY, P. *Basic color terms: Their universality and evolution*. California UP, 1969.
- [14] BERTIN, J. *Semiology of graphics: diagrams, networks, maps*. University of Wisconsin press, 1983.
- [15] BORGIO, R., DEARDEN, J., AND JONES, M. W. Order of Magnitude Markers: An Empirical Study on Large Magnitude Number Detection. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 2261–2270.
- [16] BYRON, L., AND WATTENBERG, M. Stacked graphs - Geometry & aesthetics. *IEEE Transactions on Visualization and Computer Graphics* 14, 6 (2008), 1245–1252.
- [17] CARD, S., AND MACKINLAY, J. *Information visualization*, 1999.
- [18] CARPENDALE, S. Evaluating information visualizations. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 4950 LNCS (2008), 19–45.
- [19] CHEN, M., AND FLORIDI, L. An analysis of information visualisation. *Synthese* 190, September 2012 (2013), 3421–3438.
- [20] CLEVELAND, W. S. *Visualizing data*. Hobart Press, 1993.
- [21] CLEVELAND, W. S., AND MCGILL, R. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association* 79, 387 (1984), 531–554.
- [22] CONTRIBUTORS, W. Carte du réseau autoroutier de la france. https://commons.wikimedia.org/wiki/File:France_autoroutes_map-fr.svg.
- [23] CORRELL, M., ALBERS, D., GLEICHER, M., AND FRANCONERI, S. Comparing Averages in Time Series Data. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2012).
- [24] DAUTRICHE, R., BLANCH, R., TERMIER, A., AND SANTANA, M. Traceviz: A visualization framework for interactive analysis of execution traces. In *Actes De La 28IÈMe ConfÉrence Francophone Sur L'Interaction Homme-Machine* (New York, NY, USA, 2016), IHM '16, ACM, pp. 115–125.

- [25] DE BRUIJN, O., AND SPENCE, R. Rapid serial visual presentation: A space-time trade-off in information presentation. In *Proceedings of the Working Conference on Advanced Visual Interfaces* (2000), ACM, pp. 189–192.
- [26] DIX, A., AND ELLIS, G. By Chance - Enhancing Interaction with Large Data Sets Through Statistical Sampling. In *Proceedings of the Working Conference on Advanced Visual Interfaces* (2002), AVI '02, ACM, pp. 167–176.
- [27] DOBSON, M. W. Visual information processing and cartographic communication: The utility of redundant stimulus dimensions. *Graphic communication and design in contemporary cartography*. Chichester, UK: John Wiley & Sons (1983), 149–75.
- [28] ELLIS, G., AND DIX, A. A Taxonomy of Clutter Reduction for Information Visualisation. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (2007), 1216–1223.
- [29] ELMQVIST, N., AND FEKETE, J.-D. Hierarchical Aggregation for Information Visualization : Overview , Techniques , and Design Guidelines. *IEEE Transactions on Visualization and Computer Graphics* 16, 3 (2010), 439–454.
- [30] FAIRCHILD, M. D. *Color appearance models*. John Wiley & Sons, 2013.
- [31] FEDERICO, P., HOFFMANN, S., RIND, A., AIGNER, W., AND MIKSCH, S. Qualizon Graphs: Space-efficient Time-series Visualization with Qualitative Abstractions. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces* (2014), AVI '14, ACM, ACM, pp. 273–280.
- [32] FRANK, A. U. Different types of “times” in gis. *Spatial and temporal reasoning in geographic information systems* (1998), 40–62.
- [33] FUA, Y.-H., WARD, M. O., AND RUNDENSTEINER, E. A. Hierarchical parallel coordinates for exploration of large datasets. In *Proceedings of the conference on Visualization'99: celebrating ten years* (1999), IEEE Computer Society Press, IEEE Computer Society Press, pp. 43–50.
- [34] GREENBERG, S., AND BUXTON, B. Usability evaluation considered harmful (some of the time). In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (New York, NY, USA, 2008), CHI '08, ACM, pp. 111–120.
- [35] GROSS, H., SINGER, W., AND TOTZECK, M. Handbook of optical systems. *Energy* 17 (2005), 5.
- [36] HANSEN, C. D., CHEN, M., JOHNSON, C. R., KAUFMAN, A. E., AND HAGEN, H. *Scientific visualization: uncertainty, multifield, biomedical, and scalable visualization*. Springer, 2014.
- [37] HARMAN, H. H. *Modern factor analysis*. University of Chicago Press, 1976.

- [38] HARROWER, M., AND BREWER, C. A. Colorbrewer.org: an online tool for selecting colour schemes for maps. *The Cartographic Journal* 40, 1 (2003), 27–37.
- [39] HEER, J., KONG, N., AND AGRAWALA, M. Sizing the horizon: the effects of chart size and layering on the graphical perception of time series visualizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2009), CHI '09, ACM, pp. 1303–1312.
- [40] HINTZE, J. L., AND NELSON, R. D. Violin plots: a box plot-density trace synergism. *The American Statistician* 52, 2 (1998), 181–184.
- [41] IHAKA, R. Colour for Presentation Graphics. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003) 2*, Dsc (2003).
- [42] INSELBERG, A., AND DIMSDALE, B. Parallel coordinates for visualizing multi-dimensional geometry. In *Computer Graphics 1987*. Springer, 1987, pp. 25–44.
- [43] ISENBERG, P., BEZERIANOS, A., DRAGICEVIC, P., AND FEKETE, J.-D. A Study on Dual-Scale Data Charts To cite this version : A Study on Dual-Scale Data Charts. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2469–2478.
- [44] JABBARI, A. Composite visual mapping table. <http://iihm.imag.fr/jabbari/>.
- [45] JABBARI, A. Multiple visual mapping. In *28ième conférence francophone sur l'Interaction Homme-Machine*, M. Beaudouin-Lafon and J. Vanderdonckt, Eds., Actes des Rencontres Doctorales de la 28ième conférence francophone sur l'Interaction Homme-Machine. Fribourg, Switzerland, Oct. 2016.
- [46] JABBARI, A., BLANCH, R., AND DUPUY-CHESSA, S. Beyond horizon graph: Space efficient time series visualization with composite visual mapping. (Submitted), 2018.
- [47] JABBARI, A., BLANCH, R., AND DUPUY-CHESSA, S. Composite visual mapping for time series visualization. In *2018 IEEE Pacific Visualization Symposium (PacificVis)* (2018), IEEE, pp. 116–124.
- [48] JAIN, A. K., MURTY, M. N., AND FLYNN, P. J. Data clustering: A review. *ACM Comput. Surv.* 31, 3 (Sept. 1999), 264–323.
- [49] JAVED, W., MCDONNELL, B., AND ELMQVIST, N. Graphical perception of multiple time series. *IEEE Transactions on Visualization and Computer Graphics* 16, 6 (2010), 927–934.
- [50] JOHANSSON, S., AND JOHANSSON, J. Interactive dimensionality reduction through user-defined combinations of quality metrics. *Visualization and Computer Graphics, IEEE Transactions on* 15, 6 (2009), 993–1000.
- [51] KEIM, D. *Visual support for query specification and data mining*. PhD thesis, Aachen: Shaker, 1995.

-
- [52] KEIM, D. A. Visual Techniques for Exploring Databases. *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining Tutorial Program* (1997).
- [53] KEIM, D. A. Designing pixel-oriented visualization techniques: Theory and applications. *IEEE Transactions on Visualization and Computer Graphics* 6, 1 (2002), 59–78.
- [54] KEIM, D. A., ANKERST, M., AND KRIEGEL, H.-P. Recursive Pattern: A Technique for Visualizing Very Large Amounts of Data. In *Proceedings of the 6th Conference on Visualization'95* (1995), IEEE Computer Society, p. 279.
- [55] KEIM, D. A., AND HERRMANN, A. The gridfit algorithm: An efficient and effective approach to visualizing large amounts of spatial data. In *Proceedings of the Conference on Visualization '98* (Los Alamitos, CA, USA, 1998), VIS '98, IEEE, IEEE Computer Society Press, pp. 181–188.
- [56] KOFFKA, K. *Principles of Gestalt psychology*, vol. 44. Routledge, 2013.
- [57] KRSTAJIC, M., BERTINI, E., AND KEIM, D. A. CloudLines : Compact Display of Event Episodes in Multiple. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (2011), 2432–2439.
- [58] LAM, H., MUNZNER, T., AND KINCAID, R. Overview Use in Multiple Visual Information Resolution Interfaces. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (2007), 1278–1285.
- [59] LAMPING, J., AND RAO, R. The hyperbolic browser: A focus + context technique for visualizing large hierarchies. *Journal of Visual Languages & Computing* 7, 1 (1996), 33–55.
- [60] LEBLANC, J., WARD, M. O., AND WITTELS, N. Exploring n-dimensional databases. In *Proceedings of the 1st conference on Visualization'90* (1990), IEEE Computer Society Press, pp. 230–237.
- [61] LEVKOWITZ, H. Color icons: merging color and texture perception for integrated visualization of multiple parameters. In *Proceedings of the 2nd conference on Visualization'91* (1991), IEEE Computer Society Press, pp. 164–170.
- [62] MACKINLAY, J. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics* 5, 2 (1986), 110–141.
- [63] MATEJKA, J., GLUECK, M., GROSSMAN, T., AND FITZMAURICE, G. The Effect of Visual Appearance on the Performance of Continuous Sliders and Visual Analogue Scales. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2016), 5421–5432.

- [64] MCCORMICK, B. H. Visualization in scientific computing. *Computer graphics* 21, 6 (1987).
- [65] MCGRATH, J. E. Human-computer interaction. In *Readings in Human-Computer Interaction*, R. M. Baecker, J. Grudin, W. A. S. Buxton, and S. Greenberg, Eds. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1995, ch. Methodology Matters: Doing Research in the Behavioral and Social Sciences, pp. 152–169.
- [66] NELSON, E. S. The Impact of Bivariate Symbol Design on Task Performance in a Map Setting. *Cartographia* 37, 4 (2002), 61–78.
- [67] NOCKE, T., AND SCHUMANN, H. Meta data for visual data mining. *Proceedings Computer Graphics and Imaging, CGIM 2* (2002).
- [68] PENG, W., WARD, M. O., AND RUNDENSTEINER, E. A. Clutter reduction in multi-dimensional data visualization using dimension reordering. *Proceedings - IEEE Symposium on Information Visualization, INFO VIS* (2004), 89–96.
- [69] PERIN, C. Interactive Horizon Graphs: Improving the Compact Visualization of Multiple Time Series. *CHI '13 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2013), 3217–3226.
- [70] PLAISANT, C., MILASH, B., ROSE, A., WIDOFF, S., AND SHNEIDERMAN, B. Life-Lines : Using Visualization to Enhance Navigation and Analysis of Patient Records Kaiser Permanente Colorado HCIL Technical Report No . 98-08 (October 1998). In *Proceedings of the SIGCHI conference on Human factors in computing systems* (1996), ACM, pp. 221–227.
- [71] PLAYFAIR (1786), W. *Commercial and Political Atlas and Statistical Breviary*. Cambridge University Press, 2005.
- [72] REIJNER, H. The development of the horizon graph. *IEEE Visualization Workshop: From Theory to Practice: ...* (2008), 1.
- [73] RETCHLESS, D. P., AND BREWER, C. A. Guidance for representing uncertainty on global temperature change maps. *International Journal of Climatology* (2015).
- [74] ROBERTSON, P. K. A methodology for choosing data representations. *IEEE Comput. Graph. Appl.* 11, 3 (May 1991), 56–67.
- [75] RUNDENSTEINER, E. A., WARD, M. O., YANG, J., AND DOSHI, P. R. Xmdv-tool: Visual interactive data exploration and trend discovery of high-dimensional data sets. In *Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data* (2002), SIGMOD '02, pp. 631–631.
- [76] SAITO, T., MIYAMURA, H. N., YAMAMOTO, M., SAITO, H., HOSHIYA, Y., AND KASEDA, T. Two-tone pseudo coloring: Compact visualization for one-dimensional data. *Proceedings - IEEE Symposium on Information Visualization, INFO VIS* (2005), 173–180.

-
- [77] SARKAR, M., AND BROWN, M. H. Graphical fisheye views. *Commun. ACM* 37, 12 (Dec. 1994), 73–83.
- [78] SHNEIDERMAN, B. Tree Visualization with Tree-Maps : 2-D Space-Filling Approach. *ACM Transactions on graphics (TOG)* 11, 1 (1992), 92–99.
- [79] SHNEIDERMAN, B. The eyes have it: a task by data type taxonomy for information visualizations. *Proceedings 1996 IEEE Symposium on Visual Languages* (1996), 336–343.
- [80] SIMOES, A. The observatory of economic complexity. <https://atlas.media.mit.edu>.
- [81] TUFTE, E. R., AND GRAVES-MORRIS, P. *The visual display of quantitative information*, 2 ed. Graphics press Cheshire, CT, 1983.
- [82] TUKEY, J. W. *Exploratory data analysis*. Pearson, 1977.
- [83] WANDELL, B. A. *Foundations of vision*. Sinauer Associates, 1995.
- [84] WARE, C. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004.
- [85] WEBER, M., ALEXA, M., AND MULLER, W. Visualizing time-series on spirals. *IEEE Symposium on Information Visualization, 2001. INFOVIS 2001.* (2001), 7–13.
- [86] WENZ, S. Beyond-the-desktop interactive visualizations. *Trends in Information Visualization* (2009), 54.
- [87] WRIGHT, W. Research report: Information animation applications in the capital markets. In *Information Visualization, 1995. Proceedings.* (1995), IEEE, pp. 19–25.
- [88] YANG, J., WARD, M. O., RUNDENSTEINER, E. A., AND HUANG, S. Visual hierarchical dimension reduction for exploration of high dimensional datasets. *VISSYM '03: Proceedings of the symposium on Data visualisation 2003 1* (2003), 19–28.