



**HAL**  
open science

# Contributions to geovisualization for territorial intelligence

Rosa Marina Donolo

► **To cite this version:**

Rosa Marina Donolo. Contributions to geovisualization for territorial intelligence. Emerging Technologies [cs.ET]. INSA de Lyon; Università degli studi di Roma "Tor Vergata" (1972-..), 2014. English. NNT : 2014ISAL0075 . tel-01371535

**HAL Id: tel-01371535**

**<https://theses.hal.science/tel-01371535v1>**

Submitted on 10 May 2017

**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.



**2014 ISAL 0075**

**Ciclo XXIII**

**PhD THESIS**

**Jointly awarded by**

**INSTITUT NATIONAL DES SCIENCES APPLIQUEES DE LYON  
LIRIS Laboratoire d'InfoRmatique en Image et Systèmes d'information  
Ecole doctorale Informatique et Mathématiques INFOMATHS  
and**

**UNIVERSITÁ DEGLI STUDI DI ROMA TOR VERGATA**

**Dipartimento di Informatica, Sistemi e Produzione**

**Scuola di dottorato in GEOINFORMAZIONE**

**by**

**Rosa Marina Donolo**

**Contributions to Geovisualization for Territorial Intelligence**

# **Contributions to Geovisualization for Territorial Intelligence**

## **PhD Thesis Table of Contents Rosa Marina Donolo**

### **Acknowledgments**

**Abstract**

**Résumé**

**Riassunto**

### **CHAPTER I - Introduction**

- 1.1 Motivation and co-supervision of the dissertation
- 1.2 Context of the research
- 1.3 Themes and Issues of the research
- 1.4 Goals in geovisualization
- 1.4 Objectives of the reserach
- 1.6 Thesis outline

### **CHAPTER II - State of the Art – Part I**

- 2.1 Cartography and geomatics
  - 2.1.1 Cartography
  - 2.1.2 Geomatics
  - 2.1.3 Cartography and satellite data
- 2.2 Spatial Information Systems and Geographic Information Systems
- 2.3 Taxonomy of data visualization
- 2.4 Scientific visualization
- 2.5 Information visualization
  - 2.5.1 Definition of information visualization
  - 2.5.2 Core areas of development
  - 2.5.3 Visualization of data: the process is a transformation
  - 2.5.4 Information visualization parameters
  - 2.5.5 Scales of visualization
  - 2.5.6 Graphic semiology
  - 2.5.7 Infovis research centers
- 2.6 Geovisualization and visual analytics
  - 2.6.1 Definitions of geovisualization
  - 2.6.2 Definitions of visual analytics
  - 2.6.3 Examples of spatial data representations in geovisualization
  - 2.6.4 Smart representations for smart communication
  - 2.6.5 Smart communication for smart cities
  - 2.6.6 Chorems

## **CHAPTER III: - State of the Art Part II**

- 3.1 Research literature review
- 3.2 Theoretical foundations of information visualization and geovisualization
- 3.3 Graphic semiology
  - 3.3.1 Old and new visual variables
  - 3.3.2 The use of visual variables
  - 3.3.3 Visual variables for guiding visual attention
- 3.4 Psycho-cognitive framework of visual variables
- 3.5 Conclusions and future developments in information visualization

## **CHAPTER IV: An Experimental Proposal for the Evaluation of Visual Variables - Part I –**

- 4.1 The evaluation of data visualization
- 4.2 Utility of psycho-cognitive tests
- 4.3 The experimental proposal: objectives of the prototype tests
- 4.4 Presentation of the tests
- 4.5 Structure of the tests
- 4.6 Description of the test #1 in 2D: “passenger flow in a subway system”
- 4.7 Instructions of the online test #1
- 4.8 Results of the test #1
- 4.9 Conclusions and remarks for the test #1

## **CHAPTER V: An Experimental Proposal for the Evaluation of Visual Variables - Part II –**

- 5.1 Description of the test #2 in 2.5-D “urban noise on building facades”
- 5.2 Instructions of the online test #2
- 5.3 Results of the test #2
- 5.4 Conclusions and remarks for the test #2

## **CHAPTER VI: Conclusions and future perspectives**

- 6.1 Initial purposes and objectives of the dissertation
- 6.2 Achievement of the objectives and synthesis of the results
- 6.3 Research future perspectives

### **Bibliography**

### **Webliography**

### **List of figures**

### **List of publications**

### **Appendix A: Joint PhD agreement**

# Contributions to Geovisualization for Territorial Intelligence

## PhD Thesis abstract

This PhD research work is placed in the domain of Geovisualization used to implement Territorial Intelligence and decision support systems. This research work was born through the establishment of an agreement between Tor Vergata University, Rome, and INSA (Institut National des Sciences Appliquées), Lyon. The co-supervision of this thesis was born from the necessity of a multidisciplinary approach to the research topic, taking advantage of the skills in urban planning, environment and territory modeling at the Geoinformation doctoral school of Tor Vergata University, and taking advantage of the skills in Spatial Information Systems and Geovisualization at the LIRIS Laboratory of INSA.

The motivation that led us to deal with this research topic was the perception of a lack of systematic methods and universally approved empirical experiments in data visualization domain. The experiments should consider different typologies of data, different environmental contexts, different indicators and methods of representations, etc., in order to support expert users in decision making, in the urban and territorial planning and in the implementation of environmental policies.

Geovisualization or “Geographic Visualization” concerns the visual representations of geospatial data and the use of cartographic techniques to support visual analytics. The objective of the research was to find a simple and empirical method – an online test – that could be easily reproduced, for instance in a WebGIS, to support public administration in different contexts and for different tasks. The study and analysis of visual displays could lead to improvements and to standard guidelines in Geovisualization science and visual analytic methods.

The importance of Geovisualization lays in many aspects of the management of territorial and social issues; it helps to improve territorial development and to exploit territorial resources in a proper way, but it is especially important as a decision support system as it permits to elaborate a lot of data and to extract information and knowledge; we can probably say that the most rapid way to get information is through visualization.

In modern societies, we have to deal with a great amount of data every day and Geovisualization permits the management, exploration and display of big and heterogeneous data in an interactive way that facilitates decision making processes. Geovisualization gives the opportunity to the user to change the visual appearance of the maps, to explore different layers of data and to highlight problems in some areas by the citizens. Despite these advantages, one of the most common problems in Information Visualization is to represent data in a clear and comprehensible way. Spatial data have a complex structure that includes spatial component, thematic attributes, and often the temporal component. Moreover, very often the purpose of visual design is to represent simultaneously different layers of information in different contexts and in multi-dimensional systems. Actually there are limited scientific foundations to guide researchers in visual design of spatial data, and there are limited systematic and standard methods to evaluate the effectiveness of the solutions proposed.

In this Phd research work, some contributions will be provided to the creation of a systematic assessment method to evaluate and to develop effective geovisualization displays. An empirical evaluation test is proposed to assess the effectiveness of some map displays, analyzing the use of three elements of visual design:

1. the spatial indicators to be represented and their context of visualization,
2. the physical dimensions of map displays,
3. the visual variables to represent different layers of information.

We have elaborated two tests, one for each indicator displayed (and for its context of visualization); the three indicators are: the flow of passengers in a subway map, the temperature displayed on the facades of a building. The physical dimensions chosen for the map displays of these indicators are respectively, 2D and 2.5D. This approach is mostly based on user's perception about visual variables and can be considered an application of a bottom-up visual saliency model; the input stimulus is given to the observer through the use of four visual variables: color, value, texture and size; each visual variable represents an indicator in a different moment. Visual variables are first used separately to represent the indicators at different times, then are combined increasing the difficulty of the interpretation of the maps. The results of the tests suggest which visual variable is the most efficient (fastest to understand) and which visual variable is the least efficient (less fast to understand).

This empirical experiment on visual variables is a prototype and could be extended to different contexts, scenarios and applications to provide more general indications about perception and understanding of visual variables and their combinations, but could also give indications about trends and critical points in the interpretation of different map displays.

With these empirical tests we would like to suggest an approach to individuate preferences and problems in data visualization, in order to improve the understanding of data visualization and to simplify the interpretation of data for improving private and public decision making and Territorial Intelligence.

## CHAPTER I - Introduction -

The work presented in this dissertation is placed in the domain of Information Visualization and geovisualization, that are fields of Visual Design, and concern the study of interactive methods for representing data; these fields of research are developing thanks to the new technologies of real-time graphic computers.

The purpose of Information Visualization and geovisualization, is to manage, communicate and rapidly display large amounts of data and in particular to provide graphic representation of data to display and reveal complex information in short time. Successful visualizations are interesting not only for their aesthetic design, but also for showing details, patterns and connections among data, that efficiently generate new understanding and new knowledge.

The purpose of this thesis work is to encourage and support the creation and the sharing of systematic methods to represent data, and especially to evaluate the perception and the understanding of visual information for implementing territorial intelligence.

Territorial intelligence has been defined by J. Girardot, [Girardot, 2009], “the science having for object the sustainable development of territories and having for subject the territorial community”.

To pursue this objective, in this dissertation have been elaborated and proposed two empirical tests about two different representations of spatial data. Each representation of data uses different combinations of visual variables in different temporal ranges and contexts, and in different physical dimensions.

The users of the tests should indicate if they understand the information given in the different representations. These tests can provide indications about perception and understanding of visual variables and can also provide indications about the best combination of visual variables in map displays.

With the implementation of these empirical tests we would like to suggest an approach and a method to identify preferences and problems in data visualization, in order to improve the understanding of data visualization and to simplify the interpretation of visual data, and in order to promote the development of territorial intelligence.

### 1.1 Motivation and Co-supervision of the Thesis

The motivation that led us to deal with this research topic was the perception of a lack of systematic methods and universally approved empirical experiments in data visualization domain that could consider different typologies of data, different environmental contexts, different methods of representations etc.

For this reason we would propose a simple and empirical method – a set of online tests – to analyze the use and the efficiency of visual variables. We wanted that this method could be easily reproduced for different scenarios, using different combinations of visual variables, and also we wanted that this tests could be available online, in order to increase automatically and in real time the number of tested users; finally we would like to validate and generalize the results and the rules of using visual variables for different

contexts and data. We have so implemented these tests online on visual variables and we have analyzed how tested people perceive a visual variable and the associated information.

The study of the understanding of visual representation of data, can lead to interesting improvements and to standard guidelines in geoinformation displays and visual analytics methods to support expert and non-expert users.

This research work has infact two main purposes:

1. to support expert users in the preparation of EIA<sup>1</sup>, SEA<sup>2</sup>, and local development plans, in the urban and extra-urban planning and in the implementation of environmental policies using a powerful tool as geovisualization,
2. but also to support non-expert users in the approach with Information Visualization.

This thesis work has been realized in a co-supervised PhD research program between the Geoinformation<sup>3</sup> doctoral school at the Faculty of Environmental Engineering at “Tor Vergata University” of Rome and the LIRIS<sup>4</sup> Laboratory at the InfoMaths<sup>5</sup> doctoral school at the “Institut National des Sciences Appliquées” - INSA - of Lyon.

The idea of establishing an agreement for the joint supervision of this thesis was born from the necessity of a multidisciplinary approach, taking advantage of the knowledge and the skills of experts in geomatics, urban planning, environment and territory modeling at the Geoinformation doctoral school of Rome, and in the other hand taking advantage of the knowledge and the skills in GIS<sup>6</sup>, Spatial Information Systems, and geospatial data visualization at the LIRIS Laboratory.

At the Geoinformation doctoral school, the research was more oriented to the evaluation of the quality of the urban and extra-urban environment, analyzing the different methods of representation and visualization of environmental indicators in urban and extra-urban areas. The objective was to improve the understanding and the management of spatial data related to environmental impacts especially in urban areas.

At the InfoMaths doctoral school, the research focused on three aspects: Information Visualization and geovisualization, the analysis of visual variables and the study of the human-machine interface.

## 1.2 Context of the research

The management of territorial resources and services in modern society deals with the management of a lot of data; a growing amount of data and spatial data are produced and available every day. Digital spatial data sets have grown rapidly in scope, coverage and volume [Miller & Han, 2001] and this caused a period of transition from data-

---

<sup>1</sup> EIA; Environmental Impact Assessment

<sup>2</sup> SEA; Strategical Environmental Assessment

<sup>3</sup> <http://www.disp.uniroma2.it/geoinformation/>

<sup>4</sup> <http://liris.cnrs.fr/>

<sup>5</sup> <http://infomaths.univ-lyon1.fr/>

<sup>6</sup> GIS; Geographic Information System



poverty to data-richness in Information Science. In particular we have access to vast digital data resources that include geospatially referenced data coming from:

- 1) meteorological measurements,
- 2) land use measurements,
- 3) transportation measurements,
- 4) health, property and census statistics for population,
- 5) telecommunications data,
- 6) and other topics.

This transition towards data richness has been facilitated by the diffusion of new technologies like GPS receivers<sup>7</sup>, Remote Sensing, Digital Photogrammetry, LIDAR<sup>8</sup> and Radar Interferometry. Moreover this change was accelerated by the technical advances in hardware and software products and in the increased computing power to process raw data, the lower cost of data storage, etc. The systems of representations have been implemented by commercial products, (like ESRI ArcGIS, Spotfire, IBM ILOG Visualization), or are issues of research projects (like GeoVISTA or Jigsaw).

It is also very important the development and diffusion in the last years of Internet and of web-services that permitted the accessing and the delivering of data on-line, and the sharing of open-source software products for data representation and visualization; these softwares are often produced by the Internet Community, and are very useful, even if they often need a good skill in computer programming (like the Infovis Toolkit, Prefuse, VTK, Processing 2, etc.).

In the end, web-services and web-site are very useful to the development of geovisualization as they often collect the best practices in visualization design, with several examples, like shows the screenshot of one page of the website [www.visualcomplexity.com](http://www.visualcomplexity.com), in fig.1.

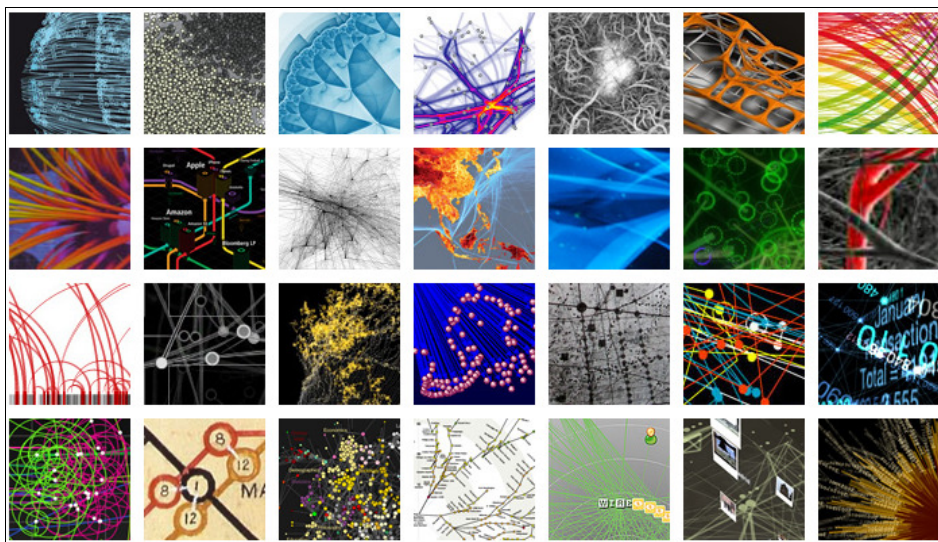


Fig.1; Examples of non-spatial information visualization. The website [www.visualcomplexity.com](http://www.visualcomplexity.com) shows a collection of visual representations of data.

<sup>7</sup> GPS: Global Positioning System

<sup>8</sup> LIDAR; Laser Imaging Detection and Ranging

The technical advances in hardware, software and data have been so important that their effect on the problems studied and the methodologies used in geovisualization have been fundamental. At the same time however, new problems have arisen in identifying relevant data and in evaluating resolution, coverage, provenance, interoperability, cost and conformance with the models to be used. Data-richness has led to new and more challenges to manage the quantity and the quality of data. For this reason the visualization design process is very complex; it is composed by a phase of processing of data, (such as filtering and selection of data), a phase of aggregation of data in index, indicators etc, a phase of analysis of the context in which data have to be represented and finally by the creation of visual representations of the data. Geospatial data in addition to the thematic component have the spatial component, and often the time component; generally attributes are represented using both traditional cartographic techniques (based on the use of visual variables like colors, textures, and using symbols, legends etc), and advanced computer techniques such as map animation and interactive 3D views.

In fact an important “revolution” in geovisualization has been the possibility of creating interactive maps, that means that user involvement is increasing in the creation and in the sharing of views; the user has become the focus of a new collaborative dimension of data visualization. This was possible thanks to the development of new technologies and software and thanks to the development of the websites called "2.0" where the users, in few clicks, can make new visualizations, as well as modify existent visualizations. A simple web browser allows the user to access all the features that previously were only available through the installation and updating of software products. Recently appeared new web-services (as ManyEyes, OpenViz, Swivel etc.) that allow the networking of the views. These web-services offer to users the possibility to transfer their datasets and to visualize them with different visual representation servers. The views are generated in few clicks and are communicated via a URL that can be posted in a public *forum* and commented. Conventional software tools, previously only available in local version of a machine, are now available in the form of websites, web-services or Java applications, (Table Public, SAP Business Objects BI, TIBCO Spotfire Analytics Server, etc.)

The development of the software and of the Internet technologies has permitted to everyone to create or to modify visualizations and to spread it thanks to the ease of use, low cost of transportation, design and consultation of data and thanks to the availability of numerous examples of visualizations.

The spread of the Internet among the general public has also good implications for the collaborative planning efforts, increasing the participation of the public while decreasing the amount of time it takes to debate about more controversial planning decisions.

geovisualization is useful in the private domain, in which people explore and manage data for personal purposes, and in the public domain, in which professionals and planners present their “visual thinking” to people and use geovisualization to explore real-world environments, models and scenarios based on spatio-temporal data. Planning relies also on collaboration between people and professionals, in fact planners use geovisualization as a tool for modeling territorial problems and policy concerns of the general public.

We can finally say that developments in hardware, software and Internet have led to the development of geovisualization. Every year at international conferences as IEEE VIS<sup>9</sup> or ACM CHI<sup>10</sup>, new visualization methods are presented to create these views and

---

<sup>9</sup> <http://ieevis.org/>

to interact with them. These methods are often specifically related to the type of data, the scope, and the user and so their generalization it is not immediate. A theoretical cognitive framework and an iterative testing approach to these methods, based on usability engineering principles, would surely permit a generalization and standardization of these methods. The final, but important aspect to consider is the context in which data have to be represented and their scale of representation; data visualizations can be displayed in urban areas or in extra-urban areas. In both cases, the visualization of geo-spatial data permits detailed analysis of different informative layers, and a graphic synthesis for territorial information, as a great amount of data can be visualized through spatial indicators. To summarize, geovisualization is an interesting and useful field of research for different reasons:

1. It can reduce the time to search information, and support decision-making,
2. It can enhance the recognition of patterns, relations, trends and critical points etc.
3. It can give a global vision of a situation, a phenomenon, etc.,
4. It enables the use of human visual memory and the capability of perceptual processing of data,
5. It permits a better interaction between user and the information system.
6. and it can possibly lead to the discovery of new bunch of knowledge.

In the following figure (fig. 2), are represented the main functions of geovisualization according to MacEarchen [MacEarchen, 2001]. The space is defined by *task*, *user*, and *interaction* enabled by the *interface*; in fact the visual design process depends on many things as the size and the nature of data, the task and the level of user experience, but also by the interface between the user and the system of data elaboration. The central diagonal of this geovisualization-use space depicts four geovisualization functions: explore, analyze, synthesize and present data.

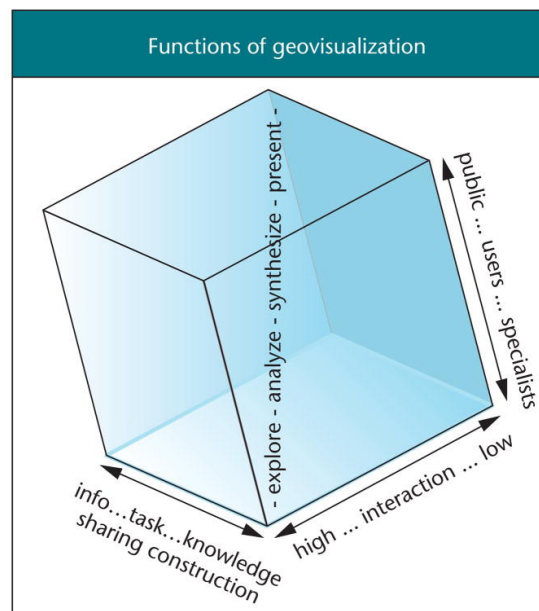


Fig. 2; Functions of geovisualization; the figure is a modified version of a MacEachren's figure.

<sup>10</sup> <http://chi2013.acm.org/>

The understanding and interpretation of data is made possible by human vision, human cognitive faculties, computational tools, domain skills and data integration, turning large heterogeneous data volumes into information and into knowledge.

In order to place our work in the period of transition and in the research domain described, we try to summarize the general context of our research:

1. Many visual data and visual representations exist in different forms (images, views, software products, web services, etc.)
2. A collective and community way of working and a social assessment is offered to users through the web services and social networks, in which users become content producers.
3. The communications network (Internet) allows the monitoring of user activity, and the interconnection of data sources, and permits to share visualizations, comments, methods, etc.
4. The users are able to compare their work and opinions in the web-sites, so they need solutions that improve information for better understanding, processing data and to improve decision making.

### **1.3 Themes and Issues of the Research**

Despite the clear advantages of geovisualization, there are still some important issues that are object of research, and that we recall below:

- 1) The complexity of visualization displays:

Cartographic representations of territorial and environmental data can be really complex for non-expert users (and sometimes also for expert users), as they often present simultaneously different levels of information and in different and multi-dimensional contexts. This is why research and experiments about understanding visualization are very important; they give indications about the preferences of the public in the comprehension of data, and they check the potential and efficiency of visual representations.

- 2) The usability of geovisualization:

Usability [Nielsen, 2000], is the capability of use and the “learnability” of a human-made object; the object of use can be a software application, website, book, tool, machine, process, or anything a human interacts with. Usability includes methods of measuring usability, such as the study of the principles behind an object’s perceived efficiency or elegance. Usability differs from user satisfaction and user experience because usability also considers usefulness. In our work we would like to analyze the usability of geovisualizations.

- 3) The adaptability of visualizations:

The adaptability of an interface (in our case the interface is the visual display of data) is the possibility of a personalized system made by a user, for example by means of a setup menu. This customization can be both depending on user preferences, capabilities and scope, or can depend on technical constraints. User preferences can be linked to certain visual impairments (blindness, age, etc.) and therefore

require the use of high-contrast colours or other visual variables, or to vary the font size, or change the background colour.

4) The flexibility of visualizations:

Flexibility is used as an attribute of various types of systems. In the field of engineering systems design, it refers to designs that can adapt when external changes occur. Flexibility has been defined differently in many fields of engineering, architecture, biology, economics, etc. In the context of engineering design we can define flexibility as the ability of a system to respond to potential internal or external changes affecting its value delivery, in a timely and cost-effective manner. In other words flexibility for an engineering system is the ease with which the system can respond to uncertainty in a manner to sustain or increase its value delivery. The flexibility also allows the evaluation of interfaces, leaving the opportunity for evaluators to vary the design alternatives and test the impact on user performance (such as the speed to perform a task or the acceptable error rate, etc.). Most current commercial visualization software tools include an adaptable and flexible visualization.

5) The interoperability of visualizations:

Interoperability permits to a user to work in the context of one application and then to use his results in a context of another application. It consists in being able to connect several different applications through the use of some formats of files similar or convertible. This issue is still difficult to solve especially because there is a problem of interoperability of visualization softwares, but also of geospatial data; a geovisualization system may access and collect data from multiple sources, and so it is difficult to combine them and to transmit them in an interoperable system.

6) The scientific foundations of geovisualization techniques;

The visual analytics community acknowledges the need for more scientific foundations to guide research in visual analytics field [Liu and Stasko, 2010]. A theoretical methodology is necessary as a reference in geovisualization; actually most of the visualization projects tend to be based on considerations on single case-studies or are guided by aesthetics instead of being based on standard foundations.

7) The standardization and generalization of geovisualization methods;

Despite the immediacy in transmitting information, one of the most common problems of geovisualization is that often representations of data are valid only for a certain place and a certain period; that is way visual characteristics need to be encoded in a standard way including all the main cases that can be found.

8) The scientific foundations for testing geovisualization techniques;

A standard methodology will not only formalize the geovisualization design but will also facilitate the establishment of the metrics to evaluate its effectiveness. Presently only limited systematic experiments are used to evaluate the effectiveness of geovisualization design. The question is to apply usability engineering tests in the traditional cartographic practice of "user testing" by evaluating visualization effectiveness throughout the lifecycle of the representation design. Applying usability engineering to geovisualization, however, may be problematic because of the novelty of

geovisualization and the associated difficulty of defining the nature of users and their tasks. It is necessary for interdisciplinary work to involve geographic information scientists, cognitive scientists, usability engineers, computer scientists, and others.

9) Cognitive issues;

As it is stated in the paper “Cognitive and usability issues in geovisualization” [Slocum et al., 2001], cognitive and usability issues should be considered and integrated in the context of six major research themes:

- geospatial virtual environments (GeoVes);
- dynamic representations (including animated and interactive maps);
- metaphors and schemata in user interface design;
- individual and group differences;
- collaborative geovisualization; and
- evaluating the effectiveness of geovisualization methods.

In particular they state that the cognitive theory may need to be developed, as the traditional cognitive theory for static two-dimensional maps may not be applicable to interactive three-dimensional immersive virtual environments and dynamic representations.

10) The information overflow and data redundancy, that should always be avoided.

## 1.4 Goals in geovisualization

The goals currently pursued in geovisualization field have been well explained by the ICA<sup>11</sup> Commission on Visualization and Virtual Environments. The initial ICA working group was extended in 1995 and become the Commission on Visualization; then in 1999 was called Commission on Visualization and Virtual Environments. Anyway the group has always emphasized the need for an effort to link cartographic research activities in visualization with those in other information science disciplines, particularly those prompted by the U.S. National Science Foundation-sponsored ViSC<sup>12</sup> report [McCormick, 1988] and has supported science and society to cope with a rapidly increasing volume of geospatial data.

We think that it is particularly useful to mention the intents of the new ICA Commission on Cognitive Issues in Geographic Information Visualization (CogVis) from 2011-2015. Their declared and attended goals are<sup>13</sup>:

- 1) Promote the awareness of cognitive issues in cartography, developing human-centred cartographic theory and practice based on empirical findings on the use of cartographic displays for spatio-temporal inference and decision-making.
- 2) Define short and medium term research goals that address key issues associated with building a theoretical base to support the construction and use of cogni-

---

<sup>11</sup> ICA: International Cartography Association

<sup>12</sup> ViSC: Visualization in Scientific Computing

<sup>13</sup> <https://www.geo.uzh.ch/microsite/icacogvis/mission.html>

tively adequate and perceptually salient visual displays of geographic information.

The specific research field include:

- 1) The research on empirical geovisualization design (2D-3D, static, animated and interactive, virtual and immersive, mobile, etc.),
- 2) the analysis of visual displays and tools to understand spatial cognition and spatial reasoning,
- 3) the study of cognitive principles to support human-visualization interaction research

It is however very important to have an interdisciplinary and international approach in this research field.

## 1.5 Objectives of the dissertation

In this research context the goals of our research work are:

- To make easier and more user-friendly the complex representations of geospatial data; and to pursue the search of guidelines to make easier and more user-friendly the complex representations of geospatial data.
- To support to the visual analytics community in the researches on geovisualization, encourage the creation and the sharing of visual representations of information.
- To underline the role of geovisualization in knowledge discovery and information communication.
- To identify more rigorous scientific methods to improve effectiveness of visual analytic tools in different contexts.
- To identify the experiences, tests and metrics to evaluate the effectiveness of visual analytics and geovisualization tools in different contexts.
- To spread the information about the available visual analytics solutions and experiments.

Our global research goal is to prove that being conscious and capable to use a “smart visualization” can be very useful for business and private purposes, but above all for scientific and public purposes. geovisualization is and will be a more and more important tool in the management of information and in the communication of data, especially for the development of smart cities and smart territories.

Our research aims at proving that the use of a more conscious and clear visualization of data, that we like to call “smart visualization”, could be very useful for personal, scientific and business purposes but overall for urban and environmental planning. geovisualization is and will be a more and more important tool in the management of information and communication especially for the development of smart cities. This is what we intend to contribute in geovisualization for territorial intelligence. More specifically our research concerns the design of a testing system for evaluating impacts and effects of different data representations on users in order to improve their capacity of decision making. The research focuses on the requirements for implementing a test on the validation of visual variables; we propose an empirical, perception-based evaluation approach

for assessing the effectiveness and efficiency of some cartographic design principles applied to 2D map displays. The approach includes a bottom-up visual saliency model that is applied to the assessment of four commonly used visual variables for designing 2D maps: size, color, intensity and texture. With this approach, we decided to propose an online test on geovisualization with the following objectives:

- 1) To evaluate which could be the most effective representation of data, changing the potentialities and the effectiveness of visual variables in different representations of data, in different contexts and dimensions.
- 2) To discover the characteristics that could help the understanding of data visualization.
- 3) To evaluate and to simplify the understanding of spatial data and interactive maps;

We can say that the test itself is a prototype of an empirical method to improve and evaluate geospatial data representations. In more general terms, we would like to make more common and useful the use of geovisualization and above all to suggest people to use it to its full potential and with more consciousness.

To reach these objectives, and to define our visual tools, we will refer to the works of J. Bertin [Bertin, 1967] on visual and spatial analysis, to present different data representations that are evaluated in some tests in order to check the potentialities and the efficiency of visual representations in different contexts. These tests can be reproduced in different contexts and the results can be extended for different data representations. With the presented approach we would like to provide cartographers, GIS scientists and visualization designers an assessment method to develop effective and efficient geovisualization displays with a correct use of visual variables.

## 1.6 Thesis outline

This dissertation is divided into six Chapters: the **Chapter I** introduces the topic of the research and the reasons for having chosen it; the motivations of the collaboration between the two Universities, INSA of Lyon and Tor Vergata University in Rome are described; then the context of the work it is presented, and the themes and issues of the research. Finally are explained the goals of the research and the preliminary objectives of the thesis.

The **Chapter II** analyzes the “State of the art” and it is divided into two sections: the state of the art about geovisualization, the state of the art of computer tools for territorial intelligence. In the first part we define Information Visualization and geovisualization and we analyze their advantages, the main areas of application, the visualization parameters; we describe the visual variables in graphic semiology, and finally we describe other aspects as the interoperability of visual variables, the scales of information visualization etc. In the second part we analyze the state of the art of the computer tools for territorial intelligence, with the definition of territorial intelligence, smart cities and the visualization for territorial intelligence.

In **Chapter III** we analyze the “State of the art” about psycho-cognitive tests in the interpretation of different representation methods, we analyze similar experiments and case-studies. We also investigate the current theoretical and cognitive foundations of



Information Visualization; some remarks are proposed about interesting discoveries and limitations concerning the state of the art. The purpose of this Chapter is to clarify the context and the approach of the dissertation.

In the **Chapter IV** it is presented the test #1 in 2-D, the design choices and the methodology used, it is presented the description of the structure of the test, the technical description, and the test instructions. At the end of the chapter the results are given and analyzed in tables and histograms.

In **Chapter V** it is presented the test #2 in 2,5-D, the design choices and the methodology used, it is presented the structure of the test, the technical description, and the test instructions. At the end of the chapter the results are given and analyzed in tables and histograms.

In **Chapter VI** we discuss the limits and potentialities of our system of tests. We discuss the achievement of the objectives of our work and we give a synthesis of the results. The conclusions of the work are illustrated and some future perspectives of the research are analyzed.

## CHAPTER II

### State of the Art – Part I

In order to set a clear framework for this dissertation, the research about the “State of the art” has been divided into two chapters. The chapter II concerns the existing literature about information visualization and geovisualization and the chapter III concerns the existing literature on psycho-cognitive foundations of information visualization.

In particular the chapter II introduces the context of the research work, by analyzing previous researches and empirical works in the fields of information visualization and geovisualization and introduces some concepts and definitions about cartography, geomatics, Geographic Information Systems, information visualization, and visual analytics. This introduction to the dissertation through an *excursus* on related sciences is important as all these disciplines make the theoretical background for advanced digital cartography and for geovisualization.

The first paragraphs of the chapter II are analyzed the basic elements of cartography, Spatial Information Systems and Geographic Information Systems. The other paragraphs are about the taxonomy of data visualization and the classification of data visualizations. Relevant space is given to the main topics, parameters, scales of information visualization and it is also given a mathematical formalization of the visualization process. Some interesting examples of geospatial representations are given order to understand the potential of geovisualization.

The last paragraph concerns “the chorems”, a technique of representation that is very interesting to display schematized representations of territories and cities and for making summaries of spatial databases.

## 2.1 Cartography and Geomatics

### 2.1.1 Cartography

Cartography is the study and practice of making maps. The word comes from ancient Greek *χάρτης*, *khartes* that means “paper”, and *γραφειν*, *graphein* that means “to write”. Cartography permits to communicate data through spatial representations. It is a science that combines different disciplines as computer science, aesthetics, topography, and geomatics. Modern cartography is closely integrated with geographic information science and constitutes the theoretical and practical foundation of Geographic Information Systems, (GIS) and of geovisualization. Traditional cartography was based largely on pen and paper; many of the conventions of manual cartography have been made in a restricted environment which imposes a limited view of reality. Computer science and digital technology were fundamental for overcoming these constraints, and led to the evolution of cartography and spatial analysis. The developments of hardware and software and research have produced the “digital cartography”. Now it is possible to find a computerized map of almost anything; there is a huge variety of different styles and types of maps. The basic elements of a map remain the same in digital cartography, but we have a more powerful way of representing data, which must be designed with accuracy, considering the audience and its needs, and considering that an interesting interactive map engages the user’s capabilities.

The cartographic design process concerns a systematic transformation of spatial data into a multi-dimensional spatial display. This process is typically performed by applying “scientific

cartographic design” methods, as well as aesthetic rules. For scientific design we intend a systematic, transparent, and reproducible design. Principles and details of the map design process can be found in many of the cartography books [Dent, 1999]; [Slocum et al., 2008]. In the following list we recall the basic elements of a map, that are the same in traditional and digital cartography:

1. The title that indicates the object of the representation,
2. The legend that explains the symbols on the map,
3. The scale that concerns the proportions of real objects described and the graphical objects represented in the map,
4. The credits (or metadata) as the data sources, the author(s), the date, etc.,
5. The map features like objects, land, water sources, towns, and other geographical features,
6. The map symbols: they have a great variety of forms and functions,
7. The labelling: it is useful to orient the reader on the map and provide important information regarding its purpose,
8. The orientation of the map (the North arrow),
9. The map projections, (It consists in the representation of the mapped objects on bidimensional support).

Also the fundamental issues and objectives of traditional cartography are the same in digital cartography:

1. The selection of the objects to be represented,
2. The editing of maps,
3. The reduction of the complexity of the map, (it consists in the elimination of the characteristics of the mapped objects that are not relevant to the map’s purpose),
4. The map design, (it consists in the management of the elements of the map to best convey its message to its audience).

Concerning the last point, if the user is unable to identify what is shown in the map in a reasonable time, the map may be regarded as useless; a correct map design is fundamental to make a meaningful map. Displaying several variables at the same time on the map allows comparison between data and show relationships between data, and adds value to the purpose of the map, but there should be no confusion concerning the purpose of the map. In order to convey the message of the map, the map-maker must design it in a way that helps the reader in the understanding of its purpose. The title of a map may provide the link necessary for communicating that message, but the overall design of the map fosters the manner in which the reader interprets it [Monmonier, 1993]. The map can play many roles: visual communication, reasoning and decision making support, data collection, data comparison.

The interesting change is that recently cartographers have not only been interested in what looks good or what visually communicates well in a map, but in how and why a particular design solution works better than others. In particular the link with the techniques of mapping and visual communication is very close.

Although some design principles have been successfully and internationally accepted as conventions, (for example the conventional meaning of colours: red usually indicates danger, green indicates safety, or red indicates hot, blue cold etc.), in the statistics community [Palsky, 1999], very few of the proposed conventions have actually been tested systematically for their effectiveness and efficiency. Users tend to extract information based on perceptual salience rather than on thematic relevance [Lowe, 2003], [Fabrikant, Goldsberry, 2005]. A system of seven visual variables was proposed by the french cartographer Jacques Bertin in his famous work “Graphic Semiology” [Bertin, 1967], and later this system was extended by various cartographers and researchers, like Morrison

[Morrison, 1974] and MacEachren [MacEachren, 1995]. More recently in the information visualization literature are important the contributions of MacKinley [Mackinley, 1999].

According to MacEachren, in all the research works there is very little empirical evidence on the effectiveness and efficiency of visual variables [MacEachren, 1995]; so, how can cartographers, GIS scientists and infographic researchers be sure that their design decisions produce effective and efficient displays?

We can answer by simply citing MacEachren, that in order to understand how and why certain displays are more successful than others for decision making, an empirical evaluation of design principles, and a systematic look into the relationships between perceptual salience and thematic relevance in visualization design is needed [MacEachren & Kraak, 2001].

## 2.1.2 Geomatics

We need to talk about geomatics because it is the science that connects different fields of our interest; we propose the definition of geomatics<sup>1</sup> given by Natural Resources Canada<sup>2</sup>:

*“Geomatics consists of products, services and tools involved in the collection, integration and management of geographic data. Geographic information can be retrieved from various sources, including earth-orbiting satellites, ground-based instruments and airborne and seaborne sensors. These data are transformed into digital maps and other usable forms with state-of-the-art information technology.”*

The new scientific term Geomatics was coined by Pollock and Wright [Pollock & Wright, 1969], combining the terms Geodesy and Geoinformatics, and was defined as *“the discipline of gathering, storing, processing, interpreting and delivering geographic information or spatially referenced information”*. Geomatics includes the tools and techniques used in land surveying, remote sensing, cartography, geographic information systems, global navigation satellite systems, photogrammetry, geography and related forms of earth mapping. The term was originally used in Canada, because it is similar in origin to both French and English, but has since been adopted by the International Organization for Standardization, the Royal Institution of Chartered Surveyors, and many other international authorities, although especially in the United States there is a preference for the term “geospatial technology”. These disciplines include surveying and mapping, geodesy; and, more recent disciplines such as remote sensing, GIS (Geographic Information Systems) and new techniques in GPS (Global Positional System). In particular GIS involves the organized integration of hardware, software, geo-referenced digital information and visualization technologies, to capture, store, up-date, manipulate, analyze, and display all forms of spatial information. These disciplines have in common the fact that they are concerned with information that has spatial properties and can then be geo-referenced, usually with global coordinates (latitude and longitude).

## 2.1.3 Types of maps

---

<sup>1</sup> The term “geomatics” is practically mostly used in Canada. In the USA and UK, the expression “geoprocessing” is more common.

<sup>2</sup> Ministry of Natural Resources Canada (NRCan); <http://www.nrcan.gc.ca/>

In cartography a map can provide a general overview for a geographical area with different levels of detail; the level and the area depend on the point of view and the purpose of the map maker.

The first kind of map that was used was the **topographic map**; a topographic map is traditionally defined as a map that shows both natural and man-made features. The Canadian Centre for Topographic Information provides this definition: “A topographic map is a detailed and accurate graphic representation of cultural and natural features on the ground”.

In particular a topographic map is characterized by large-scale detail and quantitative representation of the terrain, using contour lines that represent elevations.

A topographic map is typically published as a map series, made up of two or more map sheets that are combined to form the whole map.

The topographic maps are distinguished from smaller-scale **chorographic maps** that cover large regions, **planimetric maps** that do not show elevations, and **thematic maps** that focus on specific topics.

A thematic map is a type of map especially designed to show a particular theme connected with a specific geographic area. A thematic map is a map that focuses on a specific theme or subject area, whereas in a general map the variety of phenomena, geological, geographical, political, regularly appear together. The contrast between thematic and general maps, lies in the fact that thematic maps use the base data, such as coastlines, boundaries etc., only as points of reference for the phenomenon being mapped. General maps portray the base data, such as boundaries, lines of transportation, settlements, for their own sake. Thematic maps emphasize spatial variation of one or a small number of geographic distributions. While general maps show where something is in space, thematic maps tell a story about that place. Barbara Petchenik described the difference as “in place, about space”, [Petchenik, 1979]. Thematic maps sometimes represent spatial variations and relationships among geographical distributions.

So thematic maps have the following purposes: they provide specific information about particular locations, they provide general information about spatial patterns and they can be used to compare patterns on two or more maps.

According to the number of data sets, there are different types of thematic maps. A thematic map is *univariate* if the non-location data is all of the same kind. Population density, health issues, annual rainfall etc., are examples of univariate data. A thematic map is *bivariate* if it shows the geographical distribution of two distinct sets of data. For example, a map showing rainfall and a disease’s spread, may be used to explore a possible correlation between the two phenomena. More than two sets of data leads to a *multivariate* thematic map. For example, a single map that shows population density, annual rainfall and cancer rates. It is important to analyze the methods of thematic mapping; cartographers use several methods to create thematic maps, but four methods are more common:

1. **Choropleth maps** are thematic maps with special many similarities with digital images. The Chorological mapping methods are used today in digital images. To illustrate the concept in a simple way, we take as an example the mapping of a settlement. The population density can be mapped chorologically spreading a grid on a topographic map. For each square of the map, is counted the number of dwellings present. Choropleth mapping shows statistical data aggregated over predefined regions, such as counties or states, by coloring or shading these regions. For example, countries with higher rates of infant mortality might appear darker on a choropleth map. This technique assumes a relatively even distribution of the measured phenomenon within each region. Generally speaking, differences in color are used to indicate qualitative differences, such as land use, while differences in saturation or lightness are used to indicate quantitative differences, such as population.

2. **Proportional symbol maps** are maps that the proportional symbol technique uses symbols of different sizes to represent data associated with different areas or locations within the map. For example, a disc may be shown at the location of each city in a map, with the area of the disc being proportional to the population of the city. A dot map is a proportional symbol map and might be used to locate each occurrence of a phenomenon, with a proportional dot, as in the fig. 2.1 that represents the number of homeless people in the U.S.A. for each state.
3. A **dasymetric map** is an alternative to a choropleth map. As with a choropleth map, data are collected by enumeration units. But instead of mapping the data so that the region appears uniform, *ancillary information* is used to model internal distribution of the phenomenon. For example, population density will be much lower in forested area than urbanized area, so in a common operation, land cover data (forest, water, grassland, urbanization) may be used to model the distribution of population reported by census enumeration unit such as a tract or county.
4. **Isopleth maps** are also known as contour maps or isopleth maps depict smooth continuous phenomena such as precipitation or elevation. Each line-bounded area on this type of map represents a region with the same value. For example, on an elevation map, each elevation line indicates an area at the listed elevation. An isopleth map is a planimetric graphic representation of a 3-D surface. Isopleth mapping requires 3-D thinking for surfaces that vary spatially.

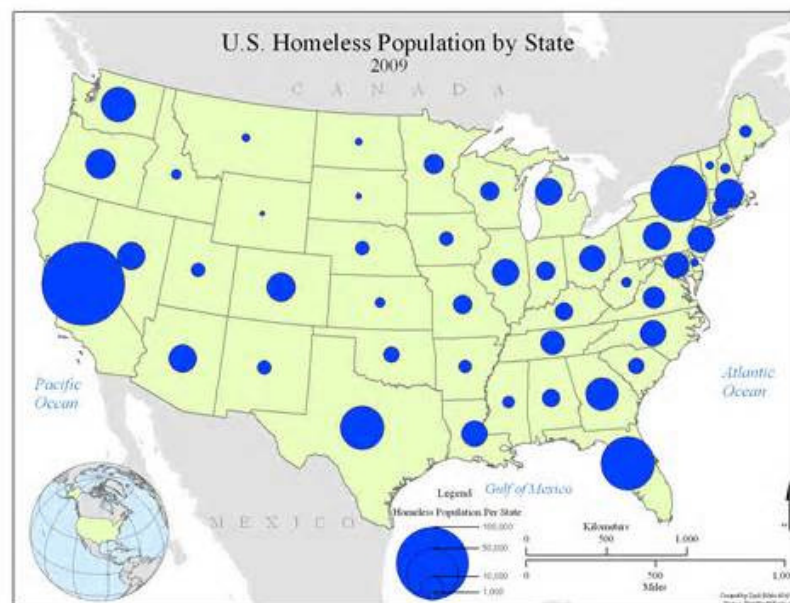


Fig 2.1 ; Example of dot map representing the homeless population by State in U.S.A.

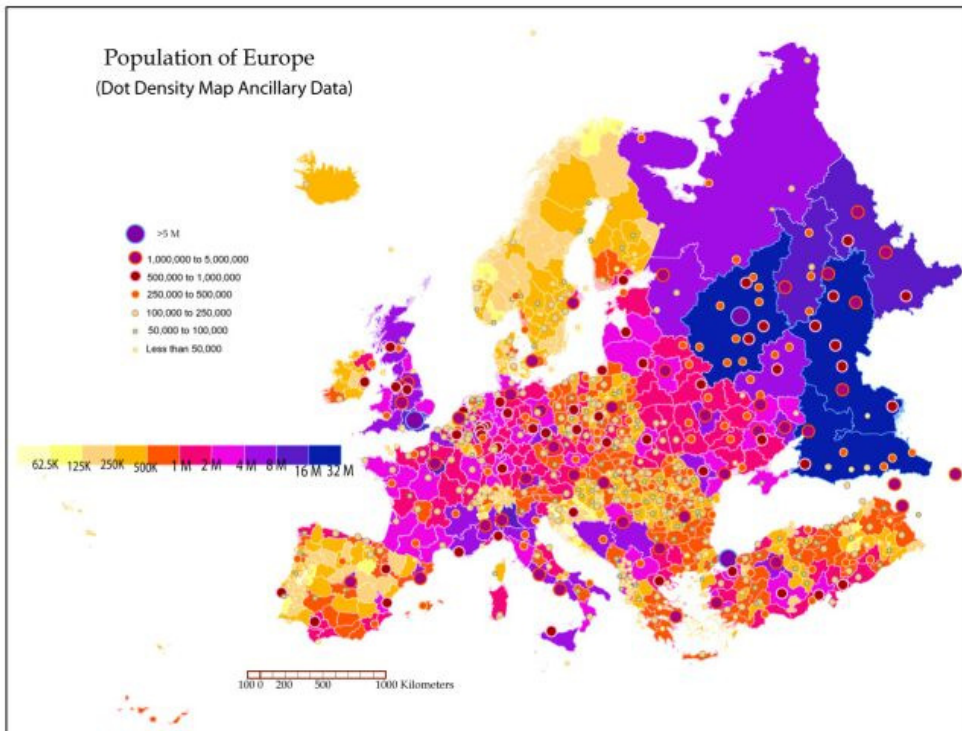


Fig 2.2 ; Example of choropleth map with overlap of a dot map, both representing the population of Europe.

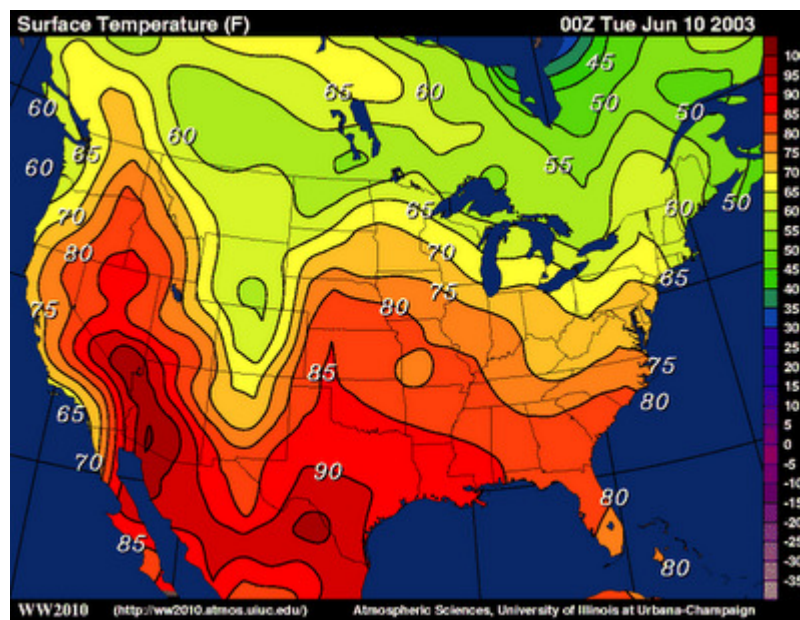


Fig 2.3 ; Example of isopleth map, representing the surface temperature.

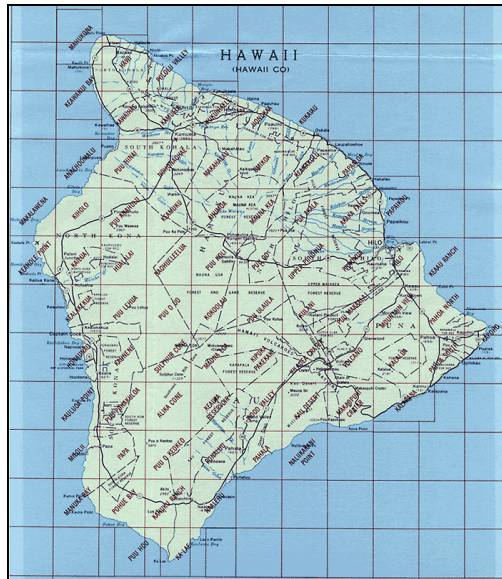


Fig. 2.4: Topographic map; a model with a grid

In fig. 2.5 there is an example of classification, using a chorological matrix and the associated histogram. A chorological matrix consists in numbers placed in a grid with a coordinate system. The geographical distribution of a settlement is converted into digits (numbers) in order to allow the computer to manage data. The histogram shows the distribution of data in the chorological matrix. On the basis of the histogram, the image data can be divided in to different classes.

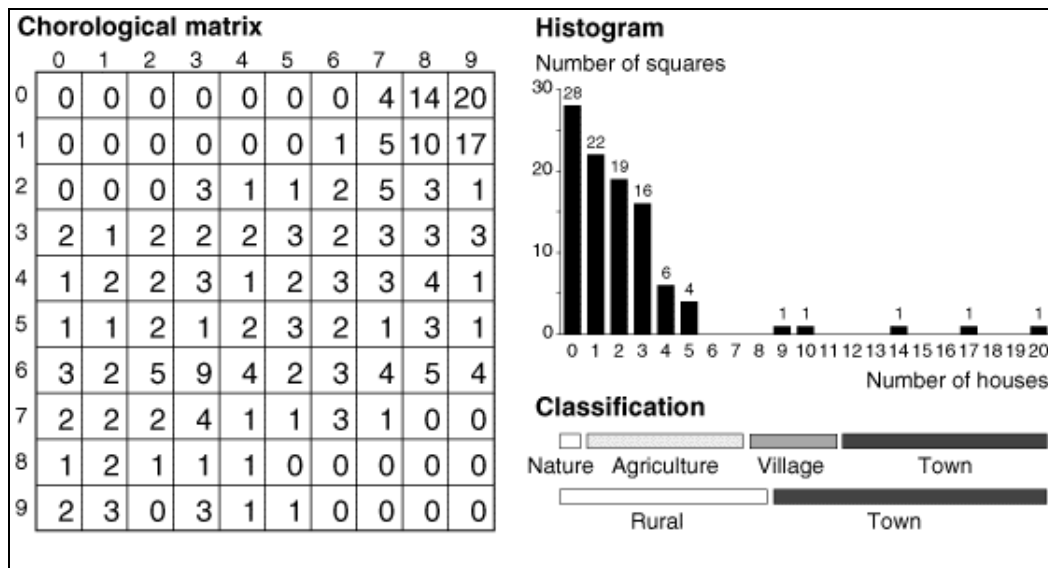


Fig. 2.5: The chorological matrix and the associated histogram. (ref.---

The illustration shows two examples of classifications based on the histogram shown: a classification is based on four classes (agriculture, towns, cities, etc.) , another classification is based on two classes (rural and city).

With a histogram, it is possible to indicate the number of dwellings in each square, for example (0,1,2, etc.). The model was overlapped on a map of the UTM grid, with the number of dwellings in each square to form the basis of the chorological matrix. Analyzing the overview given by the histogram, it is possible to find interesting classes: the squares



with 0 houses might for example be defined as forest and recreational areas, the squares from 1 to 6 houses as agricultural areas, squares from 7 to 10 houses could be defined as countries, while the squares with more than 10 homes as a city, etc.

The selected classification is placed in the histogram and each class is given shades of gray or colors. The squares in the grid (see fig. 2.6) are colored according to their classification (number of houses) and constitute a thematic map.

The problem of classification of data in categories and the problem of a correct classification depend on the purpose of the map and of the cartographer, according to the aim pursued in the map. The classifications are always a source of problems in visual design, because they depend on the arbitrary decisions of the map maker, and the same chorological matrix can often form the basis of many different maps.

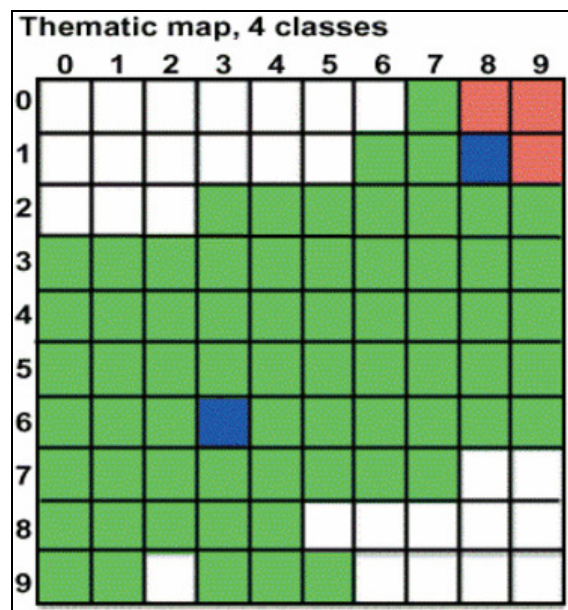


Fig. 2.6; Grid showing the four classes (and four colours) of the classification process for a thematic map.

The classifications often require compromises, for instance, if there are four classes, it could be that a suburban area is classified in the class of the country. The number of classes is important; if the classification system is based on a large number of classes there will be a good degree of detail, while if the squares are grouped in small classes (large), there will be a loss in quality of detail. Also a digital image is a chorological matrix; the size of the squares in the grid and the spatial resolution of the image depend on the system that provides the data. Moreover the number of classes is determined by the ability of the equipment to distinguish variations. The digital images often contain a division into 256 classes (each of which has a numeric value coupled with the class number), which is exactly in step with the capacity of one byte in the computer.

The individual squares in the grid are represented by a pixel, that is a picture element. The numerical value in each square of the matrix is transferred to the corresponding pixel with the same coordinates (x, y). Each pixel is colored in gray depending on the value of the pixel itself, and finally the matrix is displayed on the screen as an image or a thematic map.

The geographic coordinates of each area produce a chorological matrix, but the matrix may be subject to image processing, and the directions can be manipulated by assigning colors to the values of the pixels (classification). The map can be manipulated by adding other sources of data, multiplying, subtracting, dividing, etc. These techniques of digital image processing

are used to manage large chorological matrices and images coming from remote sensing satellites.

Remote sensing and digital image processing techniques are useful to help and update the maps. They are important tools to create real-time mapping of territories and environmental changes.

## **2.2 Spatial Information Systems and Geographic Information Systems**

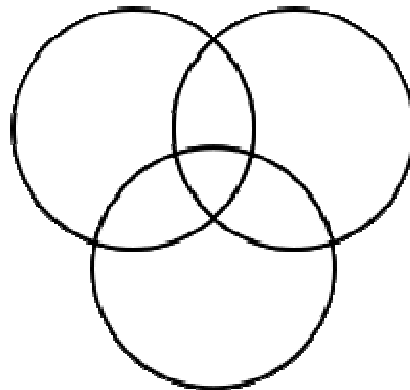
Spatial Information Systems (SIS) describe the physical location of features and the metric relationships between objects. This new information technology field concerns the collection, management and the analysis of data that have a geographic, temporal, and spatial elements. It also includes development and management of related information technology tools, such as aerial and satellite remote sensing imagery, the Global Positioning System, and computerized Geographic Information Systems (GIS). In general, GIS is a subset of a SIS. In order to better understand the difference between spatial data and geographical data, (this difference can be found also in the two spatial data types in SQL Server that distinguish geography and geometry) we propose this definition: geometric data deals with points, lines, curves, and areas on flat surfaces, while geographic data deals with the same “objects” mapped to the earth's surface. Geographic objects in a Cartesian system are defined with coordinate pairs that are simply  $X$  and  $Y$  (and eventually  $Z$ ) values on a grid, while objects in a geographic system are defined by using coordinate pairs that represent latitude and longitude, and sometimes elevation. The critical distinction is in terms of size, distance and distortion. For instance on a planar geometric coordinate system, the distance between two points can be calculated using the Pythagorean theorem and the  $x$ ,  $y$  coordinates of the involved points. That does not work for latitude/longitude pairs of coordinates in a geographic system because of the earth's curvature; what affects distance also distorts size and shape. Geographic mapping systems have grown increasingly more accurate over time, and the geography data types can be defined in terms of a number of different standards, each of which has a unique identifier. Though most uses of GIS throughout the world are almost routine, there is increasing emphasis on its application in strategic planning and decision-making [Goodchild, 2000].

## **2.3 The taxonomy of visualization techniques**

In the field of data visualization a lot of researches concern the classification of data visualization methods but also it is important the elaboration of a taxonomy of data visualizations. We recall the function of a taxonomy: in general terms it consists in the individuation of functional categories for the different visualisations. Any taxonomy or classification must involve categories which integrate their members in exclusion to those of other categories, but also should provide new insight on how the categories relate and overlap. Most of the academic researches about the definition of a taxonomy of data visualizations begin with the analysis of the structures to describe the data and the representations. The problem is that these taxonomies are structured more with the type of the data rather than the experience of the viewer, and the resulting classification systems result separated from the understanding of visualizations. We can say that actually the taxonomy of data visualizations is divided into three main categories according to the type of

data represented: Geovisualization, Information Visualization and Scientific Visualization, as shown in the fig. 2.7.

*Geovisualization*



*InfoVizualization*

*Scientific Visualization*

Fig. 2.7; The three main areas of the taxonomy of data visualizations.

The efforts are made to find a way to categorize visualizations by their type, while it should be understood that the grouping of visualizations is an intuitive process for the users. If the categories of the taxonomy originate from the viewer's experience of visualization, it is better.

In this state of the art we analysed three similar attempts to reach a common data visualization taxonomy, analysing not only the type of data, but also the perception and experience of visualization according to users; these researches have been described in the following works: "A Tour Through the Visualization Zoo" by Jeffrey Heer, Michael et al., [Heer et al., 2010], Christian Behrens's list of Design Pattern categories [Behren, 2008] and Manuel Lima's "Syntax of a New Language" in Visual Complexity [Lima, 2011]. These researchers influenced the following researches on data visualization taxonomies, especially the research based on a survey of projects described on the web-site [www.visualizing.org](http://www.visualizing.org), in which a group of researchers proposed the first version of their own taxonomy in a collaborative project. They based the categories of their taxonomy on the following principles:

1. The taxonomy is born from the analysis of the more successful types of visualisations, why we choose one over another, etc.
2. The categories are meant to describe a visualization's primary method of comparing data. For instance a map with circles scaled to show the population in major cities is comparing quantities (chart), but the primary comparison is geographic (map).
3. Some visualizations may fall into multiple categories equally.
4. The inclusion of "infographics" in their taxonomy model.

Concerning the last point and following the ideas of Tufte [Tufte, 1990], these innovative researchers argue that some visualizations are distinguished from others by the importance they dedicate to explanatory or communicative goals rather than to other characteristics as the capacity of data comparison and analysis. A network graph, for instance, dedicates most of its visual potential to representing interconnections, an infographic representation dedicates its potential to brand, visual style and illustrations, so Infographics needs a separate category to mark this emphasis on communication over comparison. The problem is that sometimes classifications mix purposes or styles, patterns, types of data structures and data structures; researchers consider that there are not only a lot of different categories, but

also complex intersections between categories: sometimes it is difficult or impossible to select a single category or to choose the proper classification for a visualization, and so some researchers say that they might define an interesting group of categories, but not a classification or a taxonomy.

## 2.4 Scientific Visualization

Scientific visualization is an interdisciplinary branch of science according to Friendly [Friendly, 2008], “primarily concerned with the visualization of three dimensional phenomena (architectural, meteorological, medical-biological, etc.), where the emphasis is on realistic renderings of volumes, surfaces, illumination sources, and so forth, perhaps with a dynamic (time) component”. It is also considered a branch of computer science that is a subset of computer graphics. The purpose of scientific visualization is to graphically illustrate scientific data to enable scientists to understand, illustrate, and catch insight from their data. Scientific data may be observations, modelling results, and have the property of having an explicit representation, often exclusive as opposed to information visualization. Scientific visualization often concerns data representations in 3D environments. Information visualization differs from scientific visualization: “it’s infovis (information visualization) when the spatial representation is chosen, and it’s scivis (scientific visualization) when the spatial representation is given”, [Muzner, 2008].

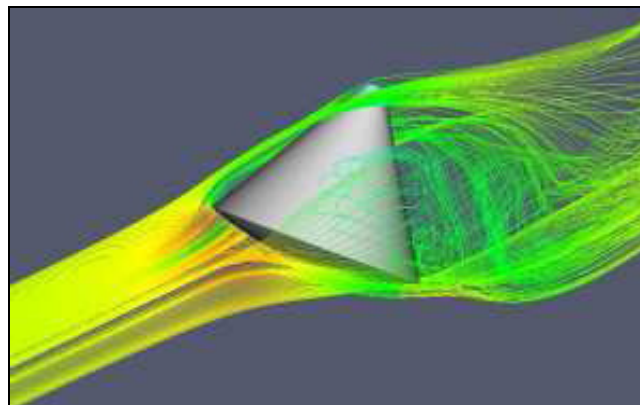


Fig. 2.8; Example of Scientific Visualization in aero-dynamics or in fluid-dynamics.

## 2.5 Information Visualization

In this section we will give some definitions about information visualization and its characteristics. Information visualization concerns the visual representation of abstract data in order to reinforce human cognition. Abstract data include both numerical and non-numerical data, such as text and geographic information.

Information visualization is the process of generating graphics from data to better use the human visual and the human cognitive system. Information visualization is a reflection process that goes beyond perception of data. Information visualization methods can be accessed on a variety of media, as paper or screen, and can be interactive, dynamic and

collaborative. The objective of the methods of information visualization are very often possible to manage and understand massive/large quantities of data. The images resulting from this process are varied, beautiful and inspiring, but viewing is however a means, not an end as said Ben Shneiderman [Shneiderman, 1999] in these words: “The purpose of visualization is insight, not pictures”.

Some examples of views of information visualization that it is possible to create, using free or commercial software, are shown in the website [www.visualcomplexity.com](http://www.visualcomplexity.com). In figure 2.3, it is shown a part of this collection of data visualizations, in particular there are examples representing non-spatial information, like social networks.

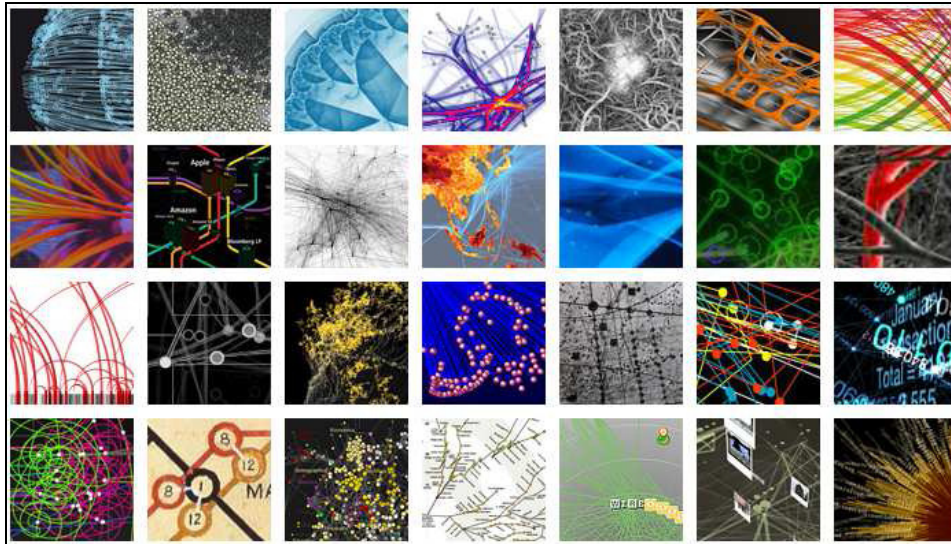


Figure 2.9.; Examples of information visualization from the site [www.visualcomplexity.com](http://www.visualcomplexity.com).

## 2.5.1 Definition of Information Visualization

As previously told, information visualization is a relatively new field in terms of information technology, especially the domain of Human - Machines Interaction (HMI). The graphical presentation devices such as screens have made important progress since the end of '80. In that period there were already some visualization instruments available but were rather intended to represent the state of the system, then the result of a complex operation. However, information visualization appeared and can be positioned in a scientific context already developed. The key areas that have contributed to technical and conceptual development of Information Visualization, are: computer graphics, geographic information systems, and scientific visualization.

Apparently, there are several definitions of Information Visualization or Visual Information; let us examine some of them. McCormick [McCormick, 1987] defines Information Visualization in this way: “*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. In many fields it is already revolutionizing the way scientists do science.*” Information Visualization is defined by Card [Card1997] as “*a process of data transformation into a visual form that permits the user to observe and no longer extract information*”.

A growing number of people would understand and use the spatial data characteristics, clusters, trends and relations without knowing difficult mathematical and statistical formulas. We do believe that visualization is a means of making sense of data for both experts and non-experts. Information visualization is the study of visual representations of abstract data to reinforce human cognition. The abstract data include both numerical and non-numerical data, such as text and geographic information. There are a growing number of studies that concern the use of different methods of spatial data representation in order to improve the interaction between the user and the information system.

The study of different aspects of information visualization needs a multidisciplinary and multidimensional analysis. The basic concepts of information visualization were given in the studies on Graphic Semiology by J. Bertin [Bertin, 1983] and in the studies on the perceptual cognitive human system [Card, 1983]. The progress in the field of information visualization is determined by the evolution of computer graphics, and also by the development of GIS as it can provide a quicker and intuitive access to information and that permits to obtain very complex and expressive representations of data. Sometimes however representations are full of visual elements that are not necessary to understand the information represented, or that distract the viewer from this information. Markings and visual elements can be called “chartjunk” if they are not part of the minimum set of visuals necessary to communicate the information understandably.

According to Tufte [Tufte, 1983], excellence in graphics consists of complex ideas communicated with clarity, precision, and efficiency. Graphical displays should induce the viewer to think about the substance, present many numbers in a small space, make large data sets coherent, encourage the eye to compare different pieces of data, reveal the data at several levels of detail, from a broad overview to the fine structure. Information workspaces are not oriented around visualizations themselves, but around tasks. An information workspace might contain several visualizations related to one or several tasks. Examples of elements which might be called “*chartjunk*” include unnecessary text or inappropriately complex font faces, ornamented chart axes and display frames, pictures or icons within data graphs, ornamental elements. Examples of this type include items depicted out of scale, noisy backgrounds making comparison between elements difficult in a map. The term “*chartjunk*” was coined by Edward Tufte in his 1983 book “The Visual Display of Quantitative Information”. Tufte wrote: “*The interior decoration of graphics generates a lot of ink that does not tell the viewer anything new. The purpose of decoration varies - to make the graphic appear more scientific and precise, to enliven the display, to give the designer an opportunity to exercise artistic skills. Regardless of its cause, it is all non-data-ink or redundant dataink, and it is often chartjunk*”.

There are a lot of advantages in data visualization, that we summarize in this list:

1. Reduction of the time of comprehension, management, elaboration and transfer of spatial information,
2. Easier identification of hidden characteristics and relations between spatial data (critical points, positive aspects or anomalies),
3. Improvement of the user-information system interaction,
4. Synthesis of the information from a big amount of available data.

It is important to analyze in detail these advantages of data visualization:

The first advantage in data visualization is the rapid communication to the observer of any critical or risk situations; this characteristic helps speed decision making both in emergency situations and in ordinary situations. Through data visualization it is possible to evaluate a situation at the first sight, even if complex (such as fire, earthquake, nuclear accident, etc.) and to take advantage of the fact that the characteristics of the phenomenon are immediately detectable in relation to the space. Friedman [Friedman2008] underlines the importance of a simple representation of spatial data: he argues that designers should not make the visual

representation of data too complex or “beautiful” in order to keep the intuitiveness and immediacy in the communication of information. Data visualization on maps gives the user a global vision of the territorial phenomena and geographic information systems permit queries into the database and permit to visualize the results of spatial queries immediately. A well designed GIS and a spatial data representation reduces the time of information elaboration because it reduces the time of information elaboration because it reduces the time of all the operations commonly made in spatial analysis: archiving, selection, integration, standardization, interrogation and elaboration of data.

The second advantage is about the identification of hidden characteristics; the representation of information highlights some characteristics of spatial data, identifies negative and positive points and highlights relations that are not obvious from digital data. This advantage comes from the fact that information visualized in the space permits the observer to shift the elaboration of data from the cognitive system to the perceptual system. This fact forces the user to visually link information with other territorial phenomena that are occurring in the same place and time. The difficulty to visualize and relate information without spatial graphic support is solved automatically thanks to the visual representations. For instance the use of colour can highlight the aggregation between data. The use of different levels of classification and different scales of representation can represent different levels of precision in the information. The use of animation permits the representation of variable parameters in space and time. In general visual representation of data helps the management of a great amount of information in a fast and stimulating way using some parameters like spatial proximity, colour, shape, dimension, classification and scale.

The third advantage concerns the user-information system interaction; Data visualization is a tool that encourages the interaction between users and informative systems, through operations like zoom, pan, animation, interrogation, integration, and it improves especially if the GIS is online. The graphic representation of data not only gives static results but also dynamic and interactive results.

The fourth advantage concerns the property of synthesis of the data; visualization techniques can be useful to investigate a big amount of data, as we can really consider visualization as a technique of visual data mining and at the same time a way to make a synthesis of data.

## 2.5.2 Information Visualization core areas of development

Information visualization actually focuses on five core areas of development:

1. Information signatures: to create mathematical signatures from text, multimedia, and sensor data, discovering new ways of summarizing key features in large, heterogeneous data sets. The focus is on performance and scalability, creating computationally-efficient representations of complex data sets.
2. Visual design: to develop new ways to transmit information with data through visual representations. Aesthetic depictions are created for complex patterns and relationships that summarize visually the output of our information signatures.
3. Analytic methods: to design software products and methods that guide users in exploring and drawing insight from visual representations. To help people use visualizations to create and to test assumptions, to communicate results, and to challenge assumptions.
4. Natural user interaction: this is to know that interaction with visual interfaces is where the insight happens. In addition to creating new interactive visual environments for the web, mobile devices, and desktop applications, we need to explore emerging techniques for gesture and touch interfaces and new display hardware that brings users closer to their data than ever before.
5. User experience: to help people work with information. We bring a user centered design approach to our work, collaborating closely with users to understand their problems, test solutions, and deliver usable and useful methodologies of visualization design.

## 2.5.3 Visualization of data: the process is a transformation

The information visualization pipeline describes the process of creating visual representations of data. This chain, called “infovis pipeline” (fig. 2.10) has steps and operations that will identify the main steps of space design.

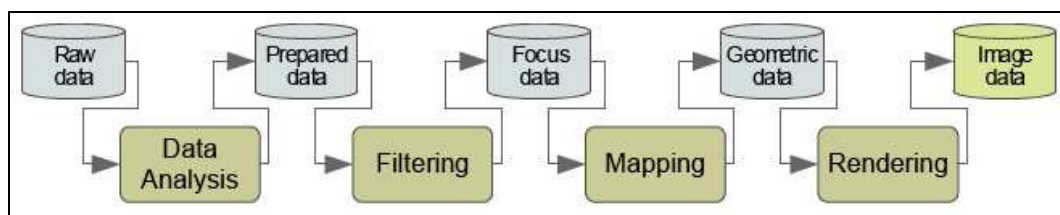


Figure 2.10; The “Infovis pipeline”.

We try to explain the four actions that produce the transformation from “raw data” to “image data”:

1. Data Analysis: data are prepared for visualization (for example by applying a smoothing filter, interpolating missing values, or correcting erroneous measurements) usually computer-centered, little or no user interaction.
2. Filtering: selection of data portions to be visualized; usually user-centred.
3. Mapping: focus data are mapped to geometric primitives (e.g., points, lines) and their attributes (e.g., color, position, size); most critical step for achieving expressiveness and effectiveness.



4. Rendering: geometric data are transformed to image data.

Another way to represent this process of visualization is explained in a really interesting paper entitled “Visual Analytics: Scope and Challenges”, by Daniel Keim [Keim et al., 1999], from the University of Konstanz, and in collaboration with Northwest National Laboratory, National Visualization and Analytics Center (NVAC), that formalized in mathematical terms the process of visualization.

According to them, the input for the data sets used in the visual analytics process are heterogeneous data sources (i.e., the internet, newspapers, books, scientific experiments, expert systems).

From these rich sources, the data sets, called “ $S$ ” =  $S_1, \dots, S_m$ , are chosen, where each  $S_i$ ,  $i$  ( $1, \dots, m$ ) consists of attributes  $A_{i1}, \dots, A_{ik}$ .

The goal of the process is the variable insight “ $I$ .” Insight is either directly obtained from the set of created visualizations “ $V$ ” or through confirmation of hypotheses “ $H$ ” as the results of automated analysis methods. Arrows represent the function “transition” from one set to another one. The formalization of the visual analytics process is the following:

The visual analytics process is a transformation “ $F$ ”,  $F: S \rightarrow I$ , where  $F$  is a concatenation of functions  $f \{DW, VX, HY, UZ\}$  defined as follows:

- $DW$  describes the basic data pre-processing functionality with  $DW: S \rightarrow S$  and  $W \{T, C, SL, I\}$  including data transformation functions  $DT$ , data cleaning functions  $DC$ , data selection functions  $DSL$  and data integration functions  $DI$  that are needed to make analysis functions applicable to the data set.
- $VW, W \{S, H\}$  symbolizes the visualization functions, which are either functions visualizing data  $VS: S \rightarrow V$  or functions visualizing hypotheses  $VH: H \rightarrow V$ .
- $HY, Y \{S, V\}$  represents the hypotheses generation process.

It is possible to distinguish between functions that generate hypotheses from data  $HS: S \rightarrow H$  and functions that generate hypotheses from visualizations  $HV: V \rightarrow H$ .

Moreover, user interactions  $UZ, Z \{V, H, CV, CH\}$  are an integral part of the visual analytics process. User interactions can either effect only visualizations  $UV: V \rightarrow V$  (i.e., selecting or zooming), or can effect only hypotheses  $UH: H \rightarrow H$  by generating a new hypotheses from given ones. Furthermore, insight can be concluded from visualizations  $UCV: V \rightarrow I$  or from hypotheses  $UCH: H \rightarrow I$ .

The typical data pre-processing applying data cleaning, data integration and data transformation functions is defined as  $DP = DT(DI(DC(S_1, \dots, S_n)))$ .

After the pre-processing step, either automated analysis methods  $HS = \{f_{sl}, \dots, f_{sq}\}$  (i.e., statistics, data mining, etc.), or visualization methods  $VS: S \rightarrow V$ ,  $VS = \{f_{vl}, \dots, f_{vs}\}$ , are applied to the data, in order to reveal patterns and relations.

## 2.5.4 Information visualization parameters

The choice of the representation methods depends on different factors that are also elements of a correct design of the maps:

- The choice of the model of spatial data,
- The placement of visual objects,
- The classification of the classes or categories of the values of an indicator,
- The use of visual variables (colors, size etc.),
- The choice of the representation scale,

- The kind and the number of geographical objects that are used to represent a spatial variable (points, lines, polygons),
- The kind of attributes that are used,
- The representation of an environmental phenomenon through an integrated indicator that summarizes different aspects of the phenomenon,
- The representation of information spatially co-existent with the overlaying of information layers and with the use of different “graphic dimensions” to help the visualization. (multidimensional approach),
- The analysis of spatial data and spatial indicators in different moments over time,
- The possibility of spatial data representations inspired by natural and biological structures,
- The attention to the specifications imposed by regulation and in particular by the INSPIRE directive<sup>3</sup> in the process of standardization of cartographic maps.
- The use of data directly measured on ground, or of data elaborated and integrated in spatial indicators.

### 2.5.5 Scales of Visualization

In order to better understand the power of representing a big amount of data through visualization, we have analysed some cases of representation of a million of items on screens. Existing information visualization techniques are usually limited to the display of a few thousand items. Researchers try to answer the question: “*To what extent can information visualization scale?*”. To achieve this goal, they are developing special techniques to:

- experiment with non standard visual attributes such as shading, transparency and stereo-vision,
- use animation to help understanding view changes,
- experiment with new interaction techniques for dynamic labeling and animated “tours” to quickly explore a data set with several different views.

We would like to report the experiments made by some researchers [Plaisant, Fekete, 2002] that describe new interactive techniques capable of handling a million items. They evaluated the use of hardware-based techniques available with newer graphics devices, new algorithms, as well as new animation techniques and non-standard graphical features such as “stereovision” and “overlap count”. These techniques have been applied to two popular information visualizations: treemaps and scatter plot diagrams, but they are generic enough to be applied to other representations as well. They are relying on hardware graphics acceleration to allow for smooth transitions between views, interpolation between layouts and synthesis of graphics attributes such as “overlaps” (among other things). Visualizing one million of items on a 1600 x 1200 screen is a challenge in term of visualization, graphics, perception and interaction. It is necessary to select appropriate visual attributes:

#### 1. Synthetic overlap attribute

If treemaps are visualization techniques where areas never overlap, scatter plots can not avoid overlap. Even with hundreds of items, the distribution tends to be unuseful with areas of high density that are hard to see. Transparency is useful when up to five items overlap, but with one million items, hundreds of overlapping items are not rare. To solve that problem, is synthesized an *overlap attribute* to show the item density.

#### 2. Transparency and stereovision

Using accelerated graphics card provides the display of graphic attributes which were not available on traditional graphics.. Transparency is required with overlapped items but is

not sufficient by itself to understand the number of overlaps. Transparency is only useful when it can be varied interactively to reveal overlaps and density of overlapping items.

### 3. Animation and interaction

Exploring a large data set requires trying several mappings from data attributes to visual attributes. A special problem arises when changing views using time multiplexing: the whole screen changes and the user has usually no clue on where the regions of the previous view have gone on the new one. Some researchers have designed techniques to manage these changes in a smoother way.

### 4. View flipping

When the positions of the data items are preserved flipping between views enables quick comparisons thanks to the retina persistency.

### 5. Interpolation of geometric attributes

When the geometry is modified between two views and the display flips from one to the other, it is hard and long to understand relationships between views, even with a few data items. Fekete and Plaisant [Fekete, Plaisant et al., 2007], implemented a set of interpolation techniques to animate the transformations from one view to the other so that the eye can follow one or a set of close items and understand their trends between both views. They also provided a slider to interactively interpolate from the current view to the previous for further control and understanding of the movements of items. The simplest technique for animation is linear interpolation. When several visual attributes change, linear interpolation is confusing and only allows users to track position changes at best.

## 2.5.6 InfoVis Research Centers

Throughout the world, there are many laboratories and real or virtual communities dedicated to information visualization. Before detailing the works and contributions, for some of them some rapid presentations will be given.

### **The Human-Computer Interaction Lab (HCIL) at the University of Maryland, USA**

The Human-Computer Interaction Lab (HCIL) at the University of Maryland, College Park, designs, implements, and evaluates new interface technologies that are universally usable, useful, efficient and appealing to a broad cross-section of people. To this end, the HCIL develops advanced user interfaces and design methodology. The HCIL is an interdisciplinary laboratory with topics of study as Information Studies, Computer Science, and Psychology. Current work includes new approaches to information visualization, interfaces for digital libraries, multimedia resources for learning communities, zooming user interface, technology design methods with and for children, mobile and pen-based computing and instruments for evaluating user interface technologies. The laboratory is currently directed by Jennifer Golbeck; its previous directors were Ben Shneiderman (1983-2000), Ben Bederson (2000-2006), and Allison Druin (2006-2011). The major work of Ben Shneiderman [Shneiderman, 1996] in recent years has been on information visualization, originating the treemap concept for hierarchical data. He also developed dynamic queries sliders with multiple coordinated displays. His work continued on visual analysis tools for time series data, TimeSearcher, high dimensional data, Hierarchical Clustering Explorer, and social network data.

Current work deals with visualization of temporal event sequences, such as found in Electronic Health Records. He also defined the research area of universal usability to encourage greater attention to diverse users, languages, cultures, screen sizes, network speeds, and technology platforms. Shneiderman has been criticized by others for over-promoting technology beyond its effective use, including by David F. Noble.

### **National Visualization and Analytics Center (NVAC) at the Pacific Northwest National Laboratory**

In 2004, the National Visualization and Analytics Center - NVAC- was created as a resource for visual analytics technology and tools. NVAC gathered experts in the field from government, industry, and academia, and wrote an agenda to guide the new field of visual analytics. Since then, NVAC has expanded the focus beyond the needs of DHS<sup>3</sup> and formed the Visual Analytics Community<sup>4</sup>. Its mission-focused program brings together innovative leadership, powerful partnerships, and advanced visual analytics research. NVAC supports DHS missions to counter current and future terrorist attacks in the United States and around the globe and to prepare for and respond to natural and manmade disasters. NVAC has helped to establish and coordinates with a family of visualization and analytics centers to bring a wide range of new technologies to bear through academic, government, and industrial partnerships.

---

3

4 <http://vacommunity.org>

## **Fraunhofer Institute for Intelligent Analysis, and Informations Systems IAIS, Germany**

Another important research center on Information Visualization is at Fraunhofer in Germany: Fraunhofer Institute for Intelligent Analysis and Informations Systems IAIS. One of the most important work is the one of Natalia and Gennady Andrienko, 2005, “Exploratory Analysis of Spatial and Temporal Data A Systematic Approach Approx. Their work refers to Exploratory data analysis (EDA)” about detecting and describing patterns, trends, and relations in data, motivated by certain purposes of investigation. They state that when something relevant is detected in data, new questions arise, causing specific parts to be viewed in more detail. So EDA has a significant appeal: it involves hypothesis generation rather than mere hypothesis testing. The authors describe in detail and systemize approaches, techniques, and methods for exploring spatial and temporal data in particular. They start by developing a general view of data structures and characteristics and then build on top of this a general task typology, distinguishing between elementary and synoptic tasks. This typology is then applied to the description of existing approaches and technologies, resulting not just in recommendations for choosing methods but in a set of generic procedures for data exploration. There are also tested solutions – illustrated in many examples – for reuse in the catalogue of techniques presented. Students and researchers will appreciate the detailed description and classification of exploration techniques, which are not limited to spatial data only. In addition, the general principles and approaches described will be useful for designers of new methods for EDA.

## **2.6 Geovisualization and Visual Analytics**

### **2.6.1 Definition of Geovisualization**

Geovisualization, is a short term for “Geographic Visualization” and refers to a set of tools and techniques to support geospatial data analysis through the use of interactive visualization. Like the related fields of scientific visualization and information visualization, geovisualization emphasizes information transmission. Geovisualization communicates geospatial information in ways that, combined with human understanding, allow data exploration and decision-making processes. Traditional, static maps have a limited exploratory capability; geovisualization allows more interactive maps, including the ability to explore different layers of the map, to zoom in or out, and to change the visual appearance of the map. Geovisualization represents a set of cartographic technologies and practices that take advantage of the ability of modern technologies to render changes to a map in real time, allowing users to adjust the mapped data. It is important to make complex data more clear and useful through a proper visual design, for instance inventing new visual metaphors, creating analysis algorithms, etc. geovisualization needs the expertise from different disciplines – statisticians, machine vision experts, modelers, and domain scientists – have to work with text, image, audio, video, and digital data, in order to help people discover patterns, trends, relationships, and events in complex data.

## 2.6.2 Definition of Visual Analytics

According to J. Thomas and K. Cook [Thomas et al., 2005] in their book *Illuminating the Path: The R&D Agenda for Visual Analytics*, visual analytics is fundamental for the fields of information visualization and scientific visualization and is "*the science of analytical reasoning facilitated by visual interactive interfaces*." "Visual analytics is developing together with science and technology developments in analytical reasoning, interaction, data transformations and representations". Visual analytics brings together several scientific and technical communities from computer science, information visualization, cognitive and perceptual sciences, interactive design, graphic design, and social sciences. Visual analytics integrates new tools with innovative interactive techniques and visual representations to enable human-information interaction. The design of the tools and techniques is based on cognitive, and perceptual principles. This science of analytical reasoning provides the framework upon which one can build both strategic and tactical visual analytics technologies for threat analysis, prevention, and response. Analytical reasoning is central to the analyst's task of applying human judgments to reach conclusions on the basis of some information. Visual analytics has some overlapping goals and techniques with information visualization and scientific visualization. There is currently no clear consensus on the boundaries between these fields, but the three areas can be distinguished as follows:

- Scientific visualization deals with data that has a natural geometric structure,
- Information visualization handles abstract data structures such as trees or graphs
- Visual analytics is especially concerned with sense making and reasoning.

Visual analytics seeks to marry techniques from information visualization with techniques from computational transformation and analysis of data. Information visualization forms part of the direct interface between user and machine, amplifying human cognitive capabilities in six basic ways:

6. by increasing cognitive resources, such as by using a visual resource to expand human working memory,
7. by reducing search, such as by representing a large amount of data in a small space,
8. by enhancing the recognition of patterns, such as information organized in space by its time relationships,
9. by supporting the perception of relationships that are otherwise more difficult to induce,
10. by perceptual monitoring of a large number of potential events,
11. by providing a medium that, unlike static diagrams, enables the exploration of a space of parameter values.

These capabilities of information visualization, combined with computational data analysis, can be applied to analytic reasoning to support the sense making process. Visual analytics is a multidisciplinary field that includes main focus areas:

- Analytical reasoning techniques that enable users to obtain deep insights that directly support assessment, planning, and decision making.
- Data representations and transformations that convert all types of conflicting and dynamic data in ways that support visualization and analysis.
- Techniques to support production, presentation, and dissemination of the results of an analysis to communicate information in the appropriate context to a variety of audiences.
- Visual representations and interaction techniques that take advantage of the human eye's capabilities to allow users to see, explore, and understand large amounts of information at once.

- Analytical reasoning techniques are the method by which users obtain deep insights that directly support situation assessment, planning, and decision making. Visual analytics facilitate high-quality human judgment with a limited investment of the analysts' time. Visual analytics tools must enable different analytical tasks such as:
- Understanding situations quickly, as well as the trends and events that have produced current conditions.
- Identifying possible alternative futures and their warning signs
- Monitoring current events for emergence of warning signs as well as unexpected events.

These tasks are generally conducted through a combination of individual and collaborative analysis. Visual analytics must enable hypothesis-based and scenario-based analytical techniques, providing support for the analyst to reason based on the available evidence. Visual representations translate data into a visible form that highlights important features, including commonalities and anomalies. These visual representations make it easy for users to perceive salient aspects of their data quickly. Augmenting the cognitive reasoning process with perceptual reasoning through visual representations permits the analytical reasoning process to become faster and more focused.

### 2.6.3 Examples of geovisualizations in urban and extra-urban areas

The management of data in urban contexts is usually complex as a great amount of different data, at different scale, has to be analyzed. Moreover almost all data that we collect in urban contexts are (or can be transformed into) spatial data. The building of an “information data infrastructure” is necessary to ensure efficiency for urban processes management. The management of spatial data has a fundamental role in urban information design that is part of the information virtuous cycle, fig. 2.11, [Hill, 2008]. Urban information design is one of the most important tools in the analysis of urban processes and in the planning of a “smart city”.

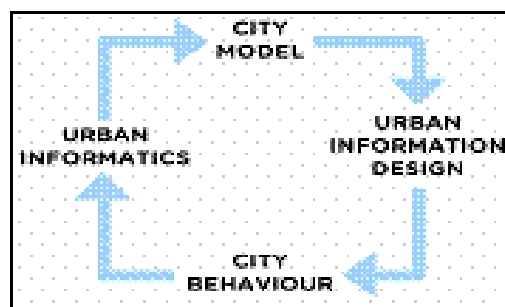


Fig. 2.11; Information virtuous cycle.

The management of spatial data can be described in the following main steps:

1. the selection of the urban processes that we want to represent,
2. the selection of the spatial data that are involved in urban processes,
3. the dimensional analysis of the physical quantities that are involved in spatial data,
4. the multiscale and multiview modeling of thematic and geometric spatial data,
5. the integration of thematic and geometric spatial data,
6. the representation of urban data and indicators.

In the last years there has been a great development of different techniques of representation of spatial data in different kinds of maps (maps of points, lines, polygons, choropleth maps, isopleth maps, cartograms etc.). Spatial indicators can be represented in different ways: in two or three dimensions, depending on the bi-dimensional or tri-dimensional character of the phenomena. If the phenomenon under analysis is mainly in two dimensions, it is common to represent it using the same number of spatial variables. But if the phenomenon to be described needs an indicator that summarizes several spatial variables, it can happen that in the representation it loses its original dimension, and the phenomenon is represented using a different dimension, or it is represented dimensionless. In the last years different methods of representation have been proposed, but not always these methods simplify the visual analysis; the simultaneous representation of different phenomena in urban areas often makes the interpretation of the map very complex (example: urban data synthesized in integrated indicators, or the representation of the same indicator in different periods in the same map, etc.). These methods introduce a more difficult level of data reading, but they have a positive effect: a better interpretation of urban phenomena. The critical points, the trends, the relations between spatial data and the spatial correlations of objects and phenomena on the territory are highlighted.

An example of the benefits obtained by developing new approaches to the mapping of data in geovisualization is illustrated below. Figure 2.6 that shows the customer sales data for a retail establishment based in the Greater Toronto Area (GTA) using a traditional GIS approach. Figure 2.6 illustrates the interpretative gain resulting from 3D mapping of the same underlying data. The adapted GIS system used in this example provides the user with the ability to quickly manipulate the level of Z-axis deformation and to explore the map from any perspective.

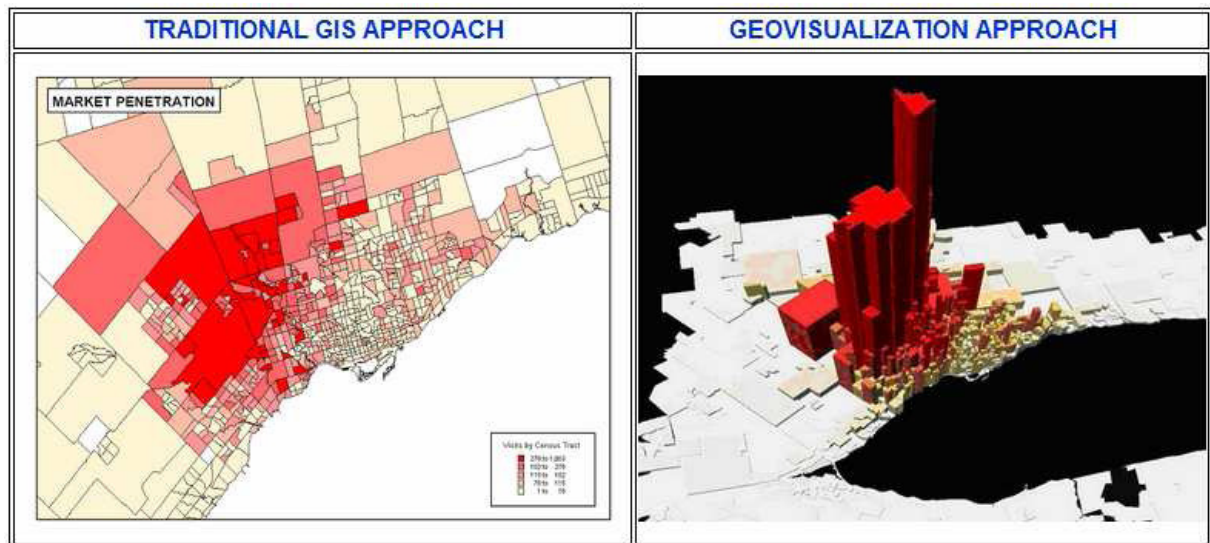


Fig. 2.12; Left side; Traditional cartographic representation; Right side; Advantages in mapping geospatial data with a geovisualization approach.

The following figures show some examples of spatial representation of data and indicators in urban and extra urban areas in a traditional GIS approach and in a more interesting geovisualization approach:



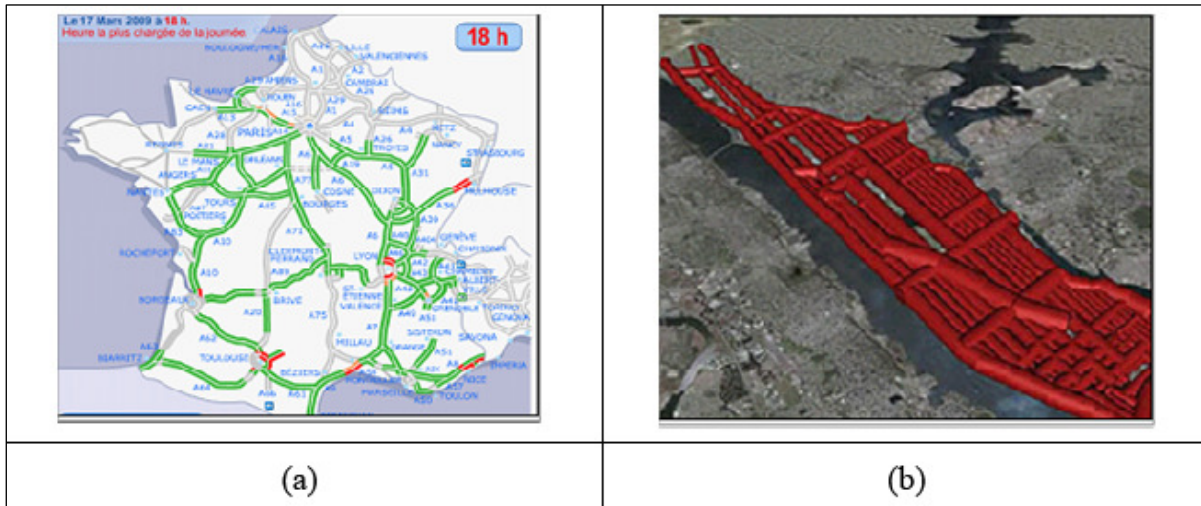


Fig. 2.13; In the left image (a), it is shown the indicator of traffic flow in two dimensions in extra-urban area, France (cfr. [www.coralys.com](http://www.coralys.com)); In the right image (b), it is shown the indicator of traffic flow in three dimensions, in urban area, New York (cfr. [www.nadiaamoroso.com](http://www.nadiaamoroso.com)).

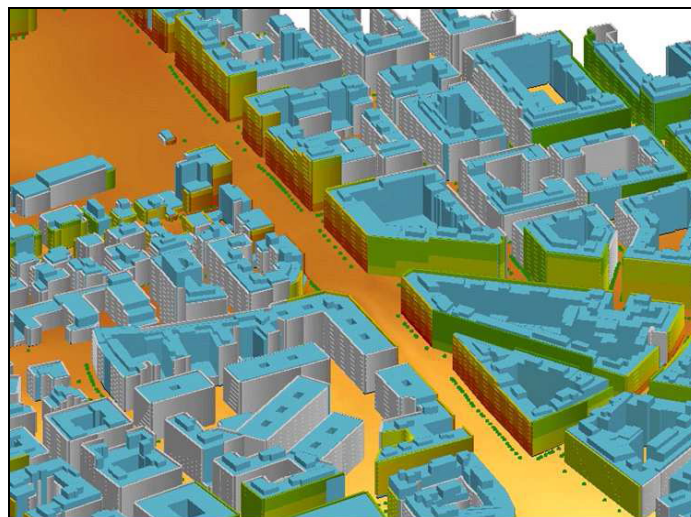


Fig. 2.14; Indicator of temperature (or noise) on the facade of buildings in urban areas [M. Ioannilli].



Fig. 2.15; Spatial indicators: intensity of the WI-FI signal of Internet – continuous indicator in three dimensions [Nadia Amoroso, 2010].



Fig. 2.16; Spatial indicators: intensity of the WI-FI signal of Internet – continuous indicator in three dimensions.

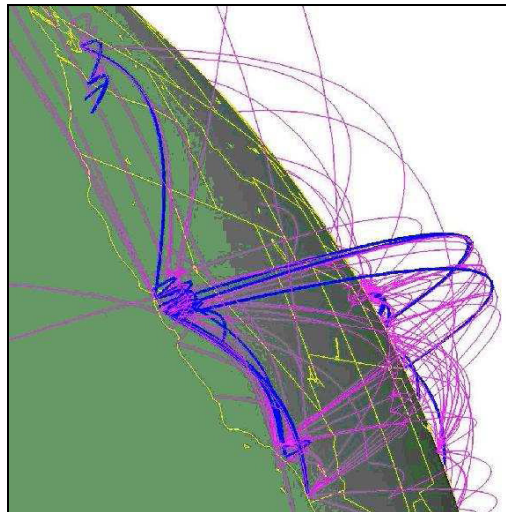


Fig. 2.17; Flow of information (or flow of road traffic) – non-continuous indicator in three dimensions.



Fig. 2.18; Flow of information (or flow of flights) – non-continuous indicator in three dimensions by TechGYD.

## 2.6.4 Smart representations for smart communication

It is important to understand that the use of smart communication is necessary to implement the smart city model; in particular the communication between local authorities and citizens not only concerns visible variables and physical infrastructures, but also invisible variables and virtual infrastructures. The techniques of a smart communication are similar for visible or invisible variables, and are based on the representation of urban and extra-urban infrastructures and processes on thematic maps.

Good communication about the topics related to urban and extra-urban life and management is surely improved and supported by the use of representation of urban and extra-urban data, phenomena and processes on maps. But a correct smart communication is not based only on a traditional representation, but on a smart representation. By smart representation we mean the exploitation of all the techniques available to improve the visualization of data on a map. First of all, the organization in synthetic indicators of data to be represented are important; then the use of the techniques of mixing these indicators, and finally it is useful to take advantage of all the visual variables, of all the spatial dimensions (even the third spatial dimension), and of all the surfaces of representation that are available in the system to be represented. As a final step the use of these so called “intelligent maps” has to be implemented online through the use of WebGIS.

Concerning the importance of the third dimension in the representation of urban and extra-urban maps, we have to underline that modern cities cannot be any more represented only in bi-dimensional maps. Old but also modern cities often have a reticular structure formed by main streets, and by their intersection that form squares and meeting points. Along the streets and in the squares are usually distributed different services as schools, hospitals, theatres, banks etc.; these points are considered the nodes of a network that have been traditionally represented in two dimensions in a map in a simplified manner such a network of a subway. Urban environment is developing in every direction; the main reasons of this fast development are the population growth, the increasing number of vehicles, and the need of new buildings. The height of the new buildings is increasingly more and more important; this is why the third dimension is so fundamental in the representation of modern and smart cities. The third dimension is important not only to represent the buildings, but also to represent the visible and invisible phenomena that happen at several meters above and under the ground, that can be represented using the surfaces of the buildings as a support. The third dimension can also be used to represent visible or invisible phenomena and variables that have no relationship with the height and that do not need to be represented on the surface of a building (like crimes). This is the case of phenomena that happen in towns but that usually are not represented in 3D maps. The advantage of the third dimension is that maps can be clearer as more variables can be represented in the third dimension. Moreover variables and indicators can be associated with other visible or invisible variables related among them, like crimes and population density, as shown in figures 2.19 and 2.20.

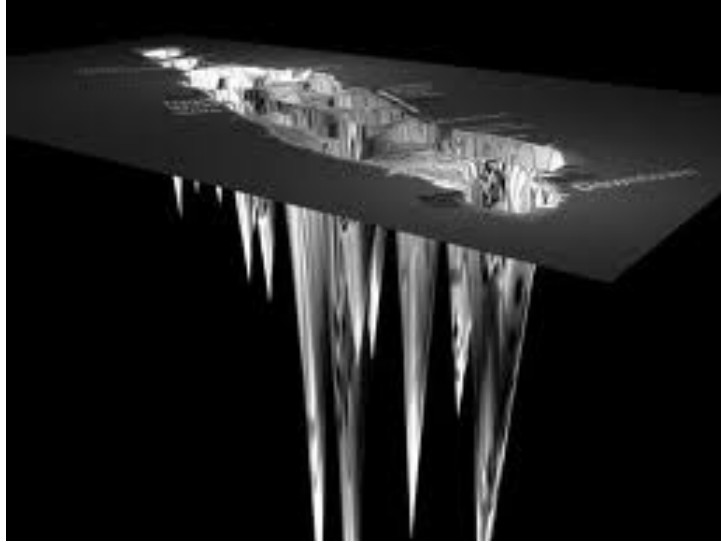


Fig. 2.19; Crimescape in New York, by Nadia Amoroso.

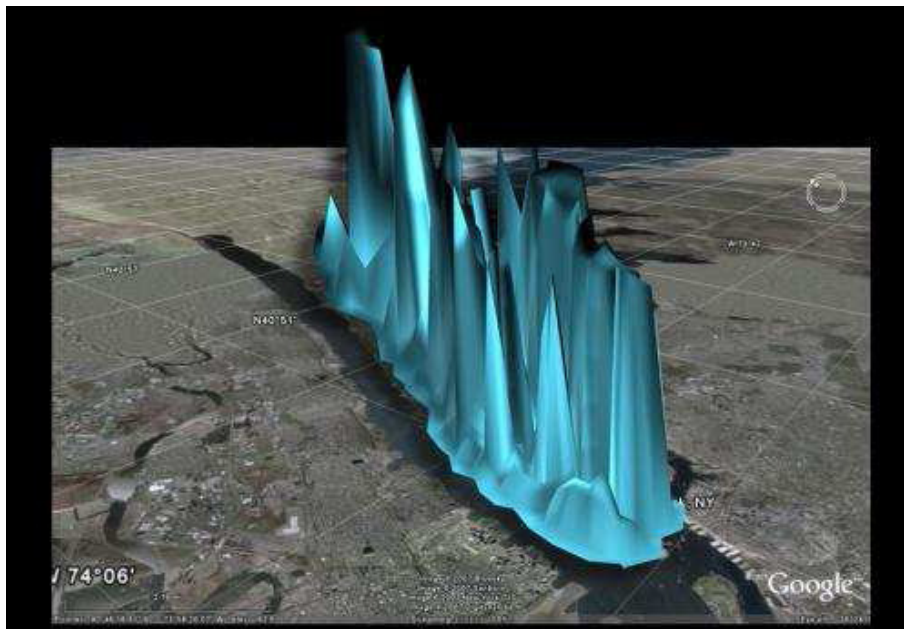


Fig. 2.20; Population density in New York, by Nadia Amoroso.

Also abstract data can be represented; invisible variables or invisible “forces” are present in a city, and especially as these variables have no spatial reference, it is more difficult to see the structure of these variables and even more difficult to see the effects of domination of some phenomena over others phenomena. Hegemonic networks can be physical and visible as technological infrastructures, (Internet or telephonic network) or virtual and invisible infrastructures as the social media networks. Dr. Paul M. Torrens [Torrens], at Arizona State University’s Department of Geography, is conducting research into the invisible geography of wireless wi-fi Internet signals at Salt Lake City, UT. He has produced an image that represents wi-fi Internet signals graphically and it is part of a work to map “urban data clouds” and to study the geography of new emerging technologies and their impact on dense urban environments.



Fig.2.21; The invisible geography of wireless wi-fi Internet signal, Salt Lake City, UT.

The same representation can be used to highlight financial infrastructures; the banks can be represented as visible nodes in the network, while the transactions and the transfer of money are invisible and can be represented as virtual lines of the network.

In fig. 2.21 the representation of an invisible financial indicator such the “market value” in New York.



Fig. 2.21; Market value in New York, by Nadia Amoroso.

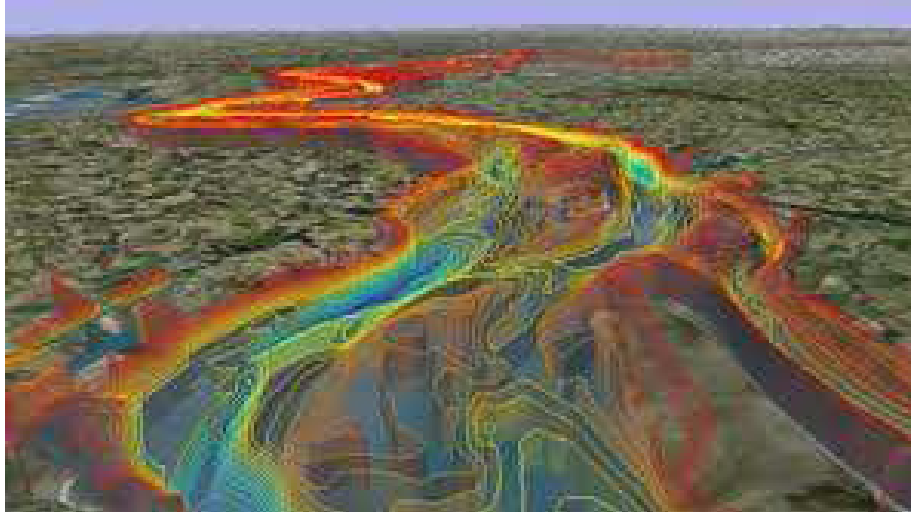


Fig. 2.22; Georeferenced image from a Google Earth application.

### 2.6.5 Smart communication for smart cities

In particular urban environment is developing in different, complex and rapid ways. The challenge of the moment is to face this issue within the model and the philosophy of the “smart city”. The biggest obstacles to the implementation of a smart city model can be probably found in the structure of urban areas and in the stratification of old urban infrastructures that make difficult the setting of new technological infrastructures. Moreover it is difficult to implement a smart city model also in new urban areas, because these areas are often the result of recent urban sprawl and shapeless urbanization, and it is difficult to set the new technological infrastructures. But in this section, we would like to highlight that in addition to these problems, there is another obstacle in the implementation of the smart city model: the difficulties encountered by the authorities in the information delivered to citizens about the problems and the resources of a city. We think that a smart information and communication system, with the support of web-GIS and smart visualization, will surely help the implementation of smart city models and smart grid models and the inclusion of citizens in the management of cities and countries. The smart city philosophy is spreading and developing in Europe and in the world. Actually the most current question and challenge for local authorities is: “which are the infrastructures required to respond to smart cities’ requirements?”. In the “smart city wheel” in fig. 2.16, Boyd Cohen [Cohen, 2005] gives a first synthetic answer; all the elements necessary to have a smart city are shown; *smart economy, smart environment, smart government, smart living, smart mobility and smart people*. In particular under the “slice” *smart government* it is written *ICT and e-government*. Taking this wheel into consideration, urban performance depends not only on the city's hard infrastructure (physical capital), but also on the availability of information and communication and of social infrastructure (intellectual capital and social capital).

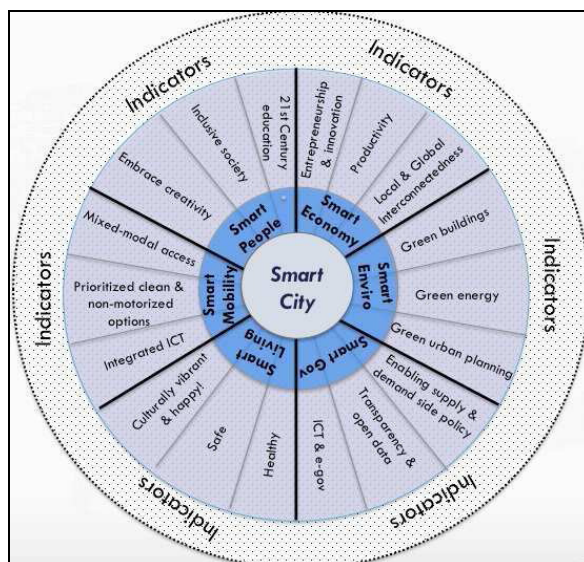


Fig. 2.23; Smart city wheel by Boyd Cohen

We would like to highlight the growing importance of Information and Communication Technologies (ICTs), social and environmental capital in developing the competitiveness of cities. Social and environmental capital are useful to distinguish smart cities and what goes under the name of digital or intelligent cities. Physical and virtual ICT infrastructures and networks are fundamental to develop smart cities and also to connect smart cities among them. The awareness and the knowledge of how to use territorial resources and energy are fundamental. Moreover the information towards and from citizens is one of the possible ways to solve the problems of social inclusion, social cohesion, and public participation of citizens in the government of the city; considering the number of people that live far from the city centre and in suburb areas, the challenge is to let the citizens feel part of a city, of a community, but also of a self-sufficient system; only in this way the citizens can be active and involved in the research of the solutions of urban and extra-urban problems.

## 2.6.6 Chorems

Chorems can be defined as a representation of elementary structure of a geographic area or as schematized representations of territories, and they generate visual summaries of spatial databases. For spatial decision-makers, it is more important to identify and map problems than facts. Until now, chorems were made manually by geographers who needed an synthetic knowledge of the territory under study, in order to decide what the salient phenomena are, and who had no problems to cartography them. In order to represent chorems in a standard way there are several publications concerning the methodologies based on spatial data mining, which provide a more rigorous approach to select the important features.

## **CHAPTER III**

### **State of the Art - part II**

#### **Theoretical foundations of Information Visualization and Geovisualization**

This Chapter analyzes some of the existing researches and methodologies about information visualization and geovisualization, and in particular analyzes some experiments on data visual displays. Information visualization and geovisualization have as a basis several science and research fields such as computer graphics, psycho-cognitive theories and human-machine interaction studies. The evaluation of data visualization needs to combine elements of formal science with aspects of empirical science for measuring the effectiveness and validity of visualizations.

In this Chapter will be analyzed the attempts to build the theoretical foundations of information visualization especially through the evaluation of data visualization.

At the end of the Chapter, some final remarks about the literature review are given.

### **3.1 Research literature review**

A literature review was conducted to investigate the available theories, approaches and experiments in the field of information visualization. Investigations were conducted in different areas such as computer graphics, psycho-cognitive science, and human-machine interaction theories; particular attention was paid to the attempts to build a theoretical foundations of information visualization, especially through the evaluation of data visualization.

The objectives of this literature review are to inform the readers about the context of the dissertation and about the potentialities of geovisualization. As said in the first Chapter, currently one of the most important issues of geovisualization and information visualization is the need for more scientific foundations. A theoretical methodology is necessary as a reference because actually most of the visualization projects tend to be based on considerations on single case-studies or are guided by a specific objective and task, instead of being based on standard foundations. The standardization and generalization of geovisualization methods are necessary as often representations of data are valid only for a certain place and a certain period. Visual characteristics need to be encoded in a standard way including all the main cases that can be found and that show similar results in the interaction between users and visual displays. The following step is to understand the reasons for these results. A standard methodology will not only formalize the geovisualization design but will also facilitate the establishment of the metrics to evaluate its effectiveness, so the need of scientific foundations should also concern the testing of geovisualization techniques, using, for instance, the usability engineering principles. Actually only limited systematic experiments are used to evaluate the effectiveness of visualization design.

Towards this issue, the analysis of the psycho-cognitive theories is important; as stated in the paper “Cognitive and usability issues in geovisualization” [Slocum et al., 2001],



cognitive and usability issues should be considered and integrated in the context of six main geovisualization research themes:

1. geospatial virtual environments;
2. dynamic representations (including animated and interactive maps);
3. metaphors and schemata in user interface design;
4. individual and group differences;
5. collaborative geovisualization;
6. evaluating the effectiveness of geovisualization methods.

The new methods for visualizing geospatial data in a 3-D environment and in dynamic contexts, will be, as well, more used if they could be developed within a theoretical cognitive framework, validated also for 2-D environment and static visual displays, tested with usability engineering principles.

## 3.2 Theoretical foundations of Information Visualization and Geovisualization

The field of information visualization is born from other disciplines such as computer science, graphics, statistical modeling, psycho-cognitive science etc., and probably for this reason, is not based on a specific theory. It involves experts like geographers, geographic information scientists, cognitive scientists, usability engineers, computer scientists; only some of them tried to build a solid common underlying theory, having the interest of building a generally valid theory for their experiments.

The information visualization practice has grown and has developed during the years thanks to the elaboration of empirical experiments and projects carried out in different countries and contexts. The purpose of a theory is to provide a framework to analyze phenomena; an information visualization theory would enable users to evaluate visualizations with a reference to an established and agreed framework. In Information visualization this framework can be used to evaluate and predict the effect of a new visualization method, the users' understanding of visualizations, and their use of it. The absence of a framework for information visualization makes the results achieved in this field difficult to describe, to validate and be reproduced.

This does not mean that a single theory would be able to represent the whole information visualization field; probably multiple theories are needed and should be integrated. In order to build the theoretical foundations of information visualization, researchers had to draw theories from multiple disciplines such as many existing cognitive and perceptual theories, statistical methods as information visualization is related to many fields and requires the use of different perspectives. Researchers draw ideas from them to formulate some assumptions and theories for information visualization.

In this literature review, we would like to report three different approaches to the building of the foundations of information visualization:

1. data-centric predictive theory,
2. information theory,
3. and scientific modeling.

First of all, we have to consider that the model that underlies all these approaches, consists in the connection between the understanding of visualizations and the understanding of languages. In order to transform the characteristics of spatial data into signs that can be perceived from a map, a language is required as for the written language theories. For this reason it is interesting to analyze the linguistic theories. In linguistic theories there are two perspectives: considering language as a representation and considering language as a process.

In the first case, the lexical signs are syntactically ordered to produce a semantic concept which the reader understands with reference to a specific code, and the concept is understood within a context.

According to the semiotic theory of de Saussure, [de Saussure, 1913] a sign is a relation between a perceptible object (signifier, referrer) and a concept (signified, referent). For example, a pair of numbers (the referrer) may mean a geographical location in one context (one possible referent), and may mean a person's age in another (a different referent). The relationship between the real object observed (referent) and its meaning (the conceptual image that is formed in the mind of the observer) is mediated by a conceptual mark in the form of sign (signifier), as shown in fig. 3.1.

The sign means both the “acoustic image” (word verbalized by phonemes), that the “written image” (word signified by letters, ideograms, or other forms of written representation).

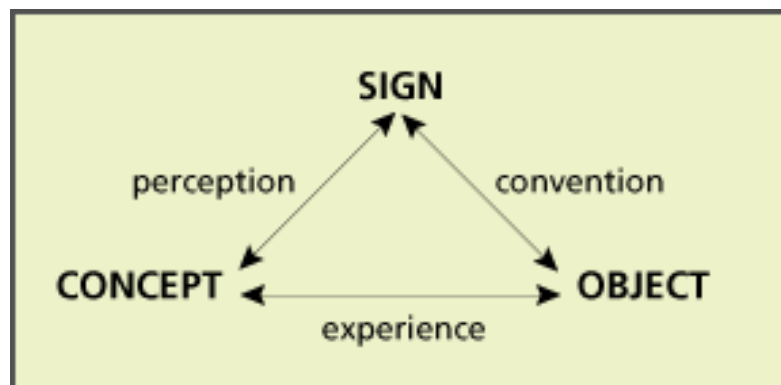


Fig. 3.1; The semiotic triangle, according to the linguistic theory of de Saussure.

In this approach, the concept of language is not accepted like a nomenclature in which there is a correspondence between sign and concept. The sign is an entity made of an arbitrary union of a concept (meaning or significance) and its acoustic image (significant); the association between meaning and significant is not linked to any natural law, as evidenced by the variety of idioms, although once established in each language, this association becomes standard and can no longer be changeable from a single speaker. It is possible to identify the “value” of an element of the language only in a “differential way”, through the relationship with the other elements of the system that allow its identification in “opposition”. The language is called structural because the determination of the value or identity of the sign, in its dual aspect, phonic and conceptual, works under the assumption that there is the totality of the linguistic system. The relationships and differences between the signs are divided into two distinct parts of the language activity: the “syntagmatic relations”, according to which the value of each sign is determined by the relationship with the sign that precedes and / or follows it, and these relations are formed by the linear sequence of words, and “associative relations” later renamed “paradigmatic” by Louis Hjelmslev [Hjelmslev, 1968]. According to him all the signs that can appear in the same context have associative (paradigmatic) relationships between them, are relations that once one relation is chosen, it excludes all other relations.

Similarly, we can extend the concept and consider the style in which language is written; the concept is understood within a style; the same response may be stimulated by a different set of signs, or the same set arranged with a different syntax, as this may produce a different emotional response. The view of language proposed by de Saussure can be considered as a static representation of meaning. In contrast, there is the view of Bakhtin [Bakhtin, 1973], who considers language a dynamic process where a text is manipulated, and its meaning constructed dynamically. Bakhtin’s theory of the dynamic interpretation of linguistic texts was based on the social context of language interpretation and the construction of ideologies within cultures and countries. Bakhtin says that this active understanding, creates new meanings: “establishes a series of complex interrelationships, consonances and dissonances, ... various different points of view, conceptual horizons come to interact with one another” [Bakhtin, 1973].

This view of “language as a process” was used as a complement to the “language as representation” perspective of de Saussure to build a framework for information visualization.

We can extend these two approaches to the formalizing of the language as well as to the formalizing of visualizations of spatial data. Data represented in visualizations have to be explored, manipulated and adapted, and both the user and the computer do perform this processing in a human-computer interaction process. This is probably why, even if visual representations of data are more intuitive and immediate to understand, the rules and theories behind visualizations are perhaps more difficult to formalize than the rules and the theories behind languages.

Each of the three sections that follow, give a different approach to suggest a theory for information visualization. These approaches were not originally developed within the linguistic model exposed in the previous paragraph, but each of them can be related to that framework:

1. Data-centric predictive theory;

This theory concerns the interpretation of a visualization through its external form (referents, and lexical, syntactic, semantic, pragmatic and stylistic structures); this is an activity typically performed by a reader;

2. information theory;

This theory concerns the exploration and manipulation of the external representation by the reader in order to discover more about the underlying model; this task is usually done through interaction facilities provided by a visualization tool;

3. scientific modeling;

This theory concerns the exploration and manipulation of the internal data model by the system in order to discover interrelationships, trends and patterns, to represent them appropriately.

Concerning geovisualization, the first approach was developed by Natalia Andrienko [Andrienko, 1997] who takes a data-centric view. She considers the dataset and the signs that describe it, and how the characteristics of the dataset and the requirements of the visualization for a specific task may be matched. She suggests the exploration of the data model to identify the best syntax to use for the given signs and to predict the patterns in datasets, and facilitate the perception of these patterns.

The second approach was developed by Matthew Ward [Ward, 2010] who supports the information theory. His view is focused on the meaning of the information in the visualizations and in the flow of information through all stages of the visualization. He analyzes how to design visualizations considering measurements of information transfer, content or loss, and looking for a means for validating visualizations. In this case, there is no internal exploration of the data in the internal model, but there is the analysis of the external representation.

The third approach concerns the development of some useful models for a scientific approach to visualization; one of them was developed by T.J. Jankun-Kelly, [Jankun-kelly et al., 2007], who called it “the visual exploration model” that describes the dynamic process of user exploration and manipulation of visualizations in order to determinate a new syntactical arrangement. Another model, called “visual transformation design”, uses transformation functions applied to the data model to provide design guidelines based on visualization parameters. This model concerns the exploration of the data model to suggest syntax to stimulate the response of the user.

### 3.3 Graphic Semiology

First of all we can define as “Graphics” the process of displaying information on a visual display in the form of diagrams, graphs, pictures, symbols and signs. Then we can define Graphic Semiology – or graphic semiotics – a field of semiotics and in particular of visual semiotics, that was formalized by Jacques Bertin [Bertin, 1967]. According to Bertin, graphic semiology is “the set of rules of a graphic sign system for the transmission of information”. Graphic semiology is a system of signs that can be used to understand maps; it is a discipline that concerns: the transcription of data into a sign in the graphics system, the processing of the data to show the information, and the design of images to provide this information. It uses the properties of the plan to show the relationships of similarity, order or proportionality between given datasets.

According to J. Bertin [Bertin, 1967], a graphic system is composed by marks, also called “signs”. Marks or signs are graphical objects that can be points, lines, areas, surfaces etc.; lines are visualized by signs of some thickness; areas have a length and width and are in a two-dimensional dimension; surfaces are areas in a three-dimensional space, but with no thickness, and volumes that have a length, a width and a depth. The described marks or signs are the basic units of graphics.

Bertin also developed some methods through which these units can be modified, including changing in position, size, shape, color etc. These modifications are called visual variables or visual attributes and each of them have certain characteristics. Crush and Pumain [Crush and Pumain, 1989] define the visual variables as “a way to vary the graphic signs”.

Bertin, defined seven visual variables (*shape*, *size*, *color*, *color value*, *color intensity*, *texture*, and *position* of the graphic sign), and different kind of signs according to their graphic sign (points, lines, areas, etc.), or to their dimensionality (0-D, 1-D, 2-D, etc.) as shown in fig. 3.2.

	<i>Points</i>	<i>Lines</i>	<i>Areas</i>
<i>Shape</i>		<i>possible, but too weird to show</i>	<i>cartogram</i>
<i>Size</i>			<i>cartogram</i>
<i>Color Hue</i>			
<i>Color Value</i>			
<i>Color Intensity</i>			
<i>Texture</i>			

Fig. 3.2; the six fundamental visual variables from J. Bertin model, in relation with the type of graphic signs(or with the dimensionality of the graphic object).

In the system of signs used, there are always at least two visual components: the two planar dimensions. The other visual variables are used to make the rendering of the signs. The representation of graphic objects or data will consist of a correspondence between components of signs and graphical representation of signs.

The complexity of a graphical representation is related to the number of components and categories in each component.

According to Bertin, there are important rules in graphic semiology, that ensure a better communication:

- **Readability:** it facilitates the understanding of the map. As noted by Bertin it is important to “detach the shape from the background”.
- **Generalization:** it reduces the level of details in order to simplify the data to adjust to a new level.
- **Identification:** for an easier reading of the map it is important to incorporate some elements: (i) the title to quickly identify the contents of the map; (ii) the caption includes all the symbols and color codes used; the scale concept of actual size; source; author: authenticate the contents of the map; (iii) the orientation.

With these rules, graphic semiology represents the mapping grammar.

### 3.3.1 Old and new visual variables

The first formalization of the graphic language was the “graphic sign language”, introduced by Bertin with his list of visual variables and his rules for graphic semiology. His work has been then adapted and extended; with the advent of new computer techniques, new visual variables were created or individuated as *sharpness*, *motion*, etc.. Thanks to computer graphics development, it was possible to add visual variables concerning color, light and optical properties, like *saturation*, *brightness* or *focus* etc., but also new variables such as *orientation* and *arrangement* of graphic objects and *perspective-height* (2.5-D), *height* (3-D), were introduced as shown in the visual variables schema in figures 3.3 and 3.4.

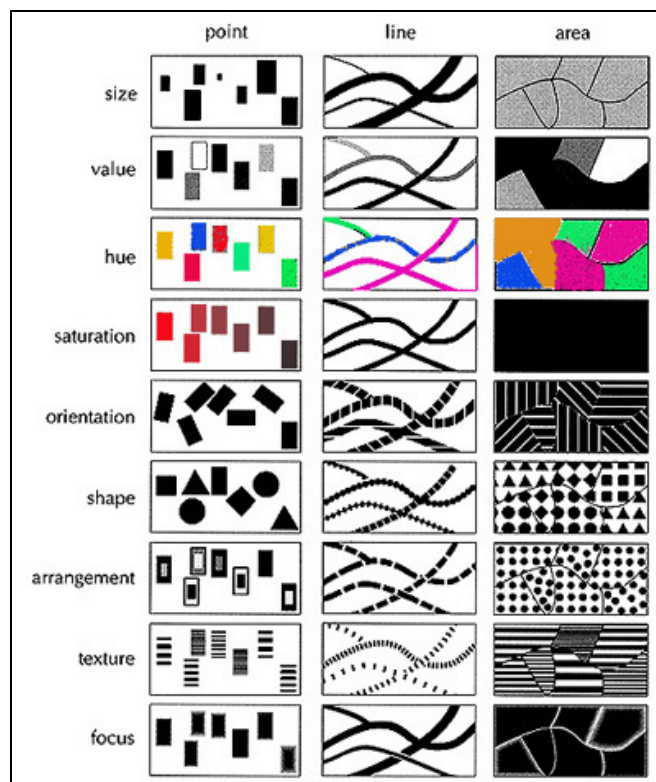


Fig. 3.3; Extended list of visual variables in relation with the type of signs in 0-D (point) ,1-D (lines), and 2-D (areas) displays.

The list of seven visual variables proposed by Bertin in 1967 was later expanded by Morrison [Morrison, 1974], MacEachren [MacEachren, 1995] and Jock D. Mackinlay [Mackinlay, 1999] who also provided a different sorting for the accuracy of visual variables, based on the task.

The list of visual variables was further expanded by several later research works. Thanks to the development of computer graphics, the visual variable *motion* and all the changes in motion such as *direction*, *speed*, *frequency*, *rhythm*, *flicker*, *trails*, and *style* were introduced into the graphic sign language. An interesting study on the visual variable *motion* was carried on in a project on visualization of mobility data described in this link: <http://casualdata.com/senseofpatterns/>, (a screen-shot is shown in fig. 3.3).

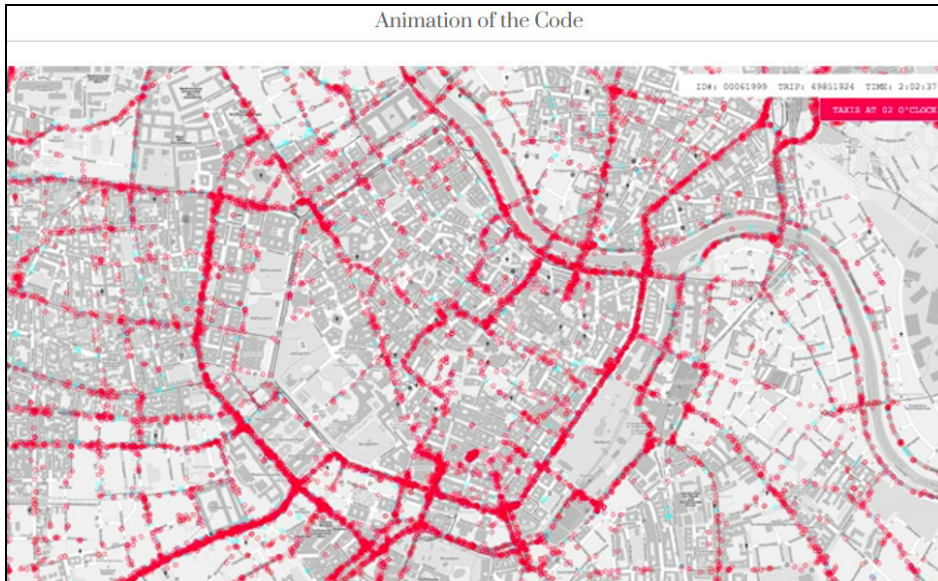


Fig. 3.3; Sense of patterns: visualizing mobility data, 2011.

Visual Variables for Quantitative Phenomena					
	Point	Linear	Areal	2½-D	True 3-D
Spacing					
Size					
Perspective Height					None Possible
Hue					
Lightness					
Saturation					

Fig. 3.4; Extended list of visual variables in relation with the kind of signs in 0-D (point) ,1-D (lines), and 2-D (areas) , and 2.5-D (volumes) displays.



### 3.3.2 The use of visual variables

The process of mapping data to visual variables is called visual mapping. It is important to know the characteristics of visual variables when creating visual data representation. The choice of different visual variables for representing different aspects of the same information influences the perception and understanding of the represented data. The choice of the visual variables that would be most appropriate to represent each aspect of information depends on its characteristics. The main characteristics of visual variables are:

1. Selective: if a mark changes in this variable and as an effect can be selected from the other marks easily the visual variable is said to be selective,
2. Associative: several marks can be grouped across changes in other visual variables,
3. Quantitative: if the difference between two marks in this variable can be interpreted numerically, the visual variable is quantitative,
4. Qualitative: if the visual variables highlights qualitative characteristics,
5. Ordered: if the variable supports ordered reading it is an ordered visual variable, This means that a change could be read as more or less (in size you can order marks according to their area),
6. Length: the length defines how many values the variable features. (For example, how many shades of grey can be recognized),
7. Categorical: if the visual variable is suitable to highlight categories.

According to visual rules and different types of data (selective, associative, quantitative, qualitative, ordered categorical etc.) it is possible to use different types of visual variable and implementations (timely, areal and linear implementations) and to determine their effectiveness in representing some properties as shown in figures 3.5 and 3.6.

	numerical	ordinal	nominal	
location	Green	Green	Green	
size	Green	Green	Green	
crispness	Red	Green	Red	
resolution	Red	Light Green	Red	
transparency	Red	Light Green	Yellow	
color value	Yellow	Light Green	Red	
color saturation	Yellow	Green	Red	
color hue	Orange	Orange	Green	
texture	Yellow	Yellow	Light Green	
orientation	Orange	Orange	Green	
arrangement	Red	Red	Orange	
shape	Red	Red	Green	

Fig. 3.5; Visual variables classified according to their characteristics and to their effectiveness.

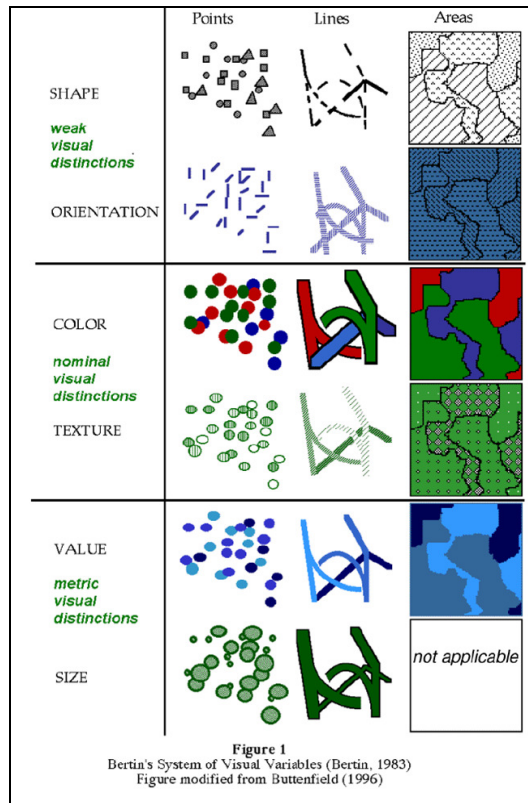


Fig. 3.6; Visual variables classified according to their characteristics and to their effectiveness.

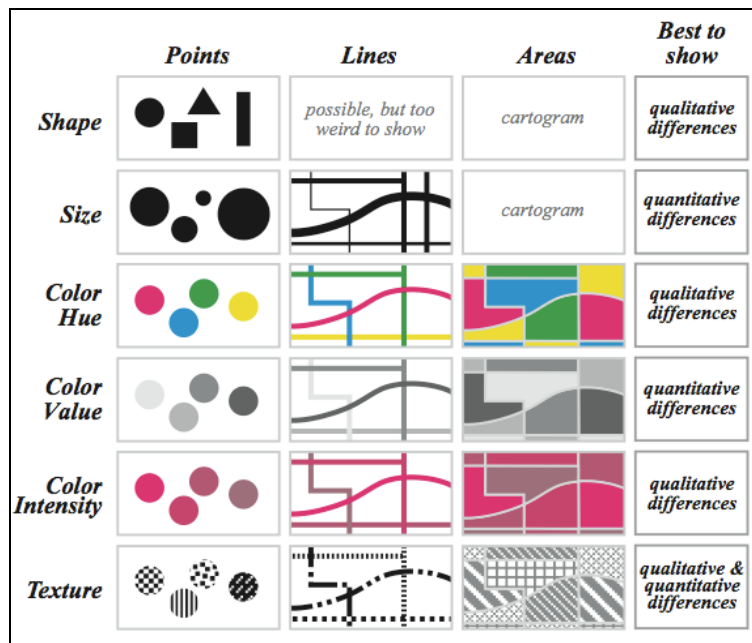


Fig. 3.7; Bertin's six fundamental visual variables, in relation with their use.

### **3.4 Towards a psycho-cognitive framework of visual variables**

Bertin's contributions can be understood within the context of the work of Gestalt psychologists such as Wertheimer and Koffka in the 1920s, reviewed by Gregory [Gregory, 1987] and Goldstein [Goldstein, 1989], who claimed that the arrangement of features in an image plane will influence the perceived thematic relations of elements. Bertin's proposals have been supported by later experimental evidence for classic visual search tasks proposed by Treisman and colleagues [Treisman & Gelade, 1980]. Wolfe & Horowitz [Wolfe & Horowitz, 2004], in a study summarizing several years of visual research in psychology and neuroscience, listed color, size and orientation as the variables to guide visual attention for static displays. The result was that visual variables were not always congruent with the organization that Bertin suggested.

### **3.5 Conclusions and future developments in Information Visualization**

In this work we argue that the framework is continuously growing under the influence of technological, conceptual and user-oriented developments. Technological developments in particular allow the representation of geo-spatial data in virtual environments, which can be experienced through multiple senses. It is interesting to note that the graphic system now includes also visual attributes to describe languages well codified for other senses, like touch and hearing, and even smell and taste are being investigated. In addition, the developments of geovisualization concern, everyday more, the user-oriented approach.

An important point that still needs to be developed through common experiments and research is the classification and the coding of visual variables of the sign language according to the use; the theoretical framework also needs further development in this direction.

To conclude, the use of visual variables in cartographic representations and especially the use of dynamic visual variables in cartographic animations should still be improved and implemented.

# CHAPTER IV

## An experimental proposal for the evaluation of data visualization

### - Test #1 -

#### 4.1 The evaluation of data visualization

The evaluation of data visualization is actually one of the most discussed topic. Not always the methods used in the theories of “human-computer interaction” are appropriate for assessing information visualization systems, moreover several visualization tools have been developed recently, but little effort has been made to evaluate the effectiveness and utility of these tools; this is the reason why it is important to investigate new evaluation procedures.

We would recall some metrics used in the evaluation of visualization systems:

1. Time required to understand the visualization system;
2. Evaluation of the error rates in the interpretation of the visualization system;
3. Retention of how to use the interface over the time.

We also recall some criteria that can be used in general in an evaluation process and in particular to evaluate visualization systems:

1. Analyzing *functionality*, that is the parameter to investigate to what extend the system provides the characteristics required by the users;
2. Analyzing *effectiveness*, that permits to understand if the visualizations provide new insight and information, and how;
3. Analyzing *efficiency*, that means to what extend the visualizations help the users in achieving a performance and pursuing a task;
4. Analyzing *usability*, that is the indicator to measure how easily the users interact with the system; this parameter provides the indication if the information is transmitted in a clear and understandable way;
5. Analyzing *usefulness*, or the indicator of how much the visualizations are useful for decision making systems.

Actually it is possible to classify the techniques of evaluation into two main categories: *Analytic evaluation* and *empiric evaluation*. Analytic evaluation, is based on formal analysis models and should be conducted by experts; It can be further divided into *heuristic evaluation* and *cognitive evaluation*: the heuristic evaluation is the procedure in which experts imagine the reactions of users in using visualization displays and describe the potential problems they foresee for such users; the cognitive evaluation, is the procedure in which experts aim to achieve a specific task using a prototype for analyzing the problems of interaction between users and visualization system.

In contrast to the analytic evaluation, there is the empirical evaluation that is realized through experiments with user testing; this is the case of our research work.

Empirical evaluation can be further distinguished in *quantitative studies* and *qualitative studies*.

Quantitative studies, consist of an analysis of determinate hypotheses tested through direct measurements;

Qualitative studies, consist of the analysis of qualitative data, which can be obtained through questionnaires, interviews and in general observing users interacting with the system, in order to understand different phenomena.

We can say that our system of tests consist in the analysis of both quantitative and qualitative data. In particular we opted for a “controlled experiment”, where the evaluator can manipulate a number of factors (in our case the visual variables) associated with the interface design and can study their effects on various aspects of user performance.

Another common method used to exchange impressions and to individuate problems in the analysis of a particular visualization interface, are *forums*, *blogs* or *internet community groups*. Often in these forums there are *administrators* and *focus group*, that consist in groups of individuals selected and assembled by researchers to discuss and animate the exchanges of ideas about the topic that is the subject of the research interviews; it is usually important the collaboration between users and administrators that ask specific questions to acquire information about users' impressions. A very important online community is the InfoVis Wiki Community<sup>1</sup> that groups the knowledge and the experiences of the most skilled researchers in the world.

---

<sup>1</sup> [http://www.infovis-wiki.net/index.php?title=Main\\_Page](http://www.infovis-wiki.net/index.php?title=Main_Page)

## 4.2 Utility of psycho-cognitive tests

Cognitive psychology is a sub-discipline of psychology that studies the internal mental processes. It concerns the study of how people perceive, remember, think, speak, and solve problems. Cognitive psychology differs from other psychological approaches in two principal ways:

It accepts the use of the scientific method, and rejects introspection as a valid method of investigation in contrast with the approaches of the “freudian psychology” that acknowledges the importance of the past experiences of the individual (especially concerning the relationships with relatives etc.) in the evolution of the person and in the present behaviour of the individual;

Cognitive psychology acknowledges the existence of internal mental states (such belief, desire, knowledge, motivation, etc.), that influence people in the behaviour, and in the ability to choose and evaluate, but also in the knowledge and communication processes.

Cognitive psychology is probably a good scientific basis to build an empirical theory on visual perception and on cognitive elaboration of spatial data. The visualization of spatial data involves a process of mental abstraction to pass from the perceptive system to the cognitive system. The process that leads from data to information is a cognitive transformation of information and is elaborated by the cognitive system, also in the case in which the process of interpretation of data comes from the interaction between the user and a graphic interface.

The mental elaboration is subjective and not standard, but presents some common elements in different people. Psycho-cognitive tests are one of the most efficient way to identify these recurrent elements that determine the preferences of people in data visualization; in this research we have proposed a test that can be defined “psycho-cognitive”.

It is clear that the experience and the culture of the user has a fundamental influence on the results of the tests. To be effective the representation has to be user-friendly and as similar as possible to the user’s mental image of the object. But it is important to understand that the mental images of the same object, can be different in different countries and cultures, and this is why it is necessary also to propose new visualization displays in order to stimulate new, but also universally valid ways of communication.

At the same time it is important to recall a universal condition of individuals that is the capacity of human memory, that doesn’t depend on the origin or on the culture of people. According to Miller [Miller, 1956], the number of objects that a common person can hold in his/her working memory, is five plus or minus two. So the immediate memory impose big limitations on the amount of information that humans are able to receive, process and remember; it is even more important and necessary to select and focus on the most important information.

## 4.3 The experimental proposal: the prototype tests

In this dissertation we elaborated an experimental proposal in which we present some prototype tests concerning the perception and understanding of some data representations in order to evaluate the role and the efficiency of visual variables in information visualization. The tested users are asked to analyze some maps with a growing level of complexity of interpretation, as there is a growing number of visual variables that represent a growing number of information layers.

We would like to recall that similar kind of tests have been used in some empirical researches, in particular there is an important experiment based on the testing of visual variables that inspired us:

“Evaluating the effectiveness and efficiency of visual variables for geographic information visualization”; a research carried out by Fabrikant and Garlandini in 2009, and presented at COSIT 2009, [Fabrikant et al., 2009].

We have also been inspired by the PhD dissertation of C. Bonhomme [C. Bonhomme, 2000] that proposes to assess a common approach to the data representation evaluation.

In this dissertation, we would like to prove that we can understand which are the most effective data visualization techniques, analyzing through some tests, the perception of different representations of data. The global objectives of our prototype tests are:

Analysis of some representations of data in order to identify critical and positive points in mapping;  
Providing support to identify standard methods for improving the visualization of data;  
Providing support to a more extensive cognitive study to identify preferences of users in visual representations.

The specific objective of our tests is to verify the efficiency of different methods of representation of spatial data using different visual variables. The first part of our test concerns the evaluation of the efficiency of visual variables in representing spatial continuous indicators as the flow of passengers in a subway system, the noise in an urban area or the data flow in the internet. In order to increase the difficulty level of interpretation of the maps, different combinations of visual variables are tested to represent the indicators simultaneously in different moments.

The main questions that we try to answer with the proposed tests are:

1. Which is the visual variable more efficient (understandable) by users in the proposed contexts? Which is the visual variable less efficient (understandable)?
2. What is the maximum number of visual variables that can be combined on the same map until the map becomes illegible?
3. Which are the best combinations of visual variables in order to represent a spatial indicator without generating errors of interpretation?

As we said previously, it is possible to find the first list and the description of visual variables in the book “Graphic Semiology” of Bertin [Bertin, 1967], that can be considered as one of the most important formal theories concerning spatial objects.

In Bertin’s theory visual variables are separated into two main groups: visual variables such as *colour*, *value*, *texture* and the geometric variables such as *form*, *dimension*, *orientation* to which we can add the variable *position* of the graphic object, that can be indicated with  $P(x, y)$  in the plan or with  $S(x, y, z)$  in the space. Visual variables used in representations are not exclusive and each one can express independent information. Sometimes some variables are proposed in a different way, especially concerning the variables *position*, it is often separated into two visual variables: the horizontal position of the object  $H(x, y)$  and vertical position of the object  $V(z)$ ; sometimes the variable *position* is not even considered as a visual variable. In general, however, the third dimension “z”, and the variable time “t”, can be considered as supplementary variables of the Bertin’s model of visual variables.

The most effective use of Bertin’s retinal variables is summarized in a visual format in the book “Making maps: a visual guide to map design for GIS” by John Krygier and Denis Wood [Krygier and Wood, 2010]. Krygier and Wood showed in the chart proposed below (fig. 4.4), how to represent the attribute expressed by the visual variable through points, lines, or areas, and also showed which visual variables are most effective for representing qualitative or quantitative differences according to Bertin.

	<i>Points</i>	<i>Lines</i>	<i>Areas</i>	<i>Best to show</i>
<i>Shape</i>		possible, but too weird to show	cartogram	qualitative differences
<i>Size</i>			cartogram	quantitative differences
<i>Color Hue</i>				qualitative differences
<i>Color Value</i>				quantitative differences
<i>Color Intensity</i>				qualitative differences
<i>Texture</i>				qualitative & quantitative differences

Fig. 4.4; Visual variables and their properties according to J. Bertin, [Krygier, Wood, 2010].

Concerning the first test, we have decided to select and analyze only four visual variables: *colour*, *intensity*, *size* and *texture*, and we have decided to exclude the following visual variables: *position*, *shape*, and *orientation*. This choice was due to some considerations:

The scenarios and the indicators represented in the tests are independent of the position and the geolocation. For instance in the first test the scenario is a “subway system”, and the indicator is the “passenger flow”; the variable *position*, relative to an origin, of the indicator is not interesting for our purposes, and has been excluded. This choice is due to the fact that in our tests we represent continuous indicators (as passenger flow, noise, or Internet data flow, etc.) that do not change their value according to the variable *position* or according to the location of the scenarios (as a subway system, the district of a city, or a telecommunication network, etc.), that we consider closed systems.

We do not take into account each point of the space, but only the single segment between each metro stop or between each segment of the building facades or of each arch of the telecommunication network.

The visual variables *shape* and *orientation* have not been taken into account because in our tests we have chosen to represent four layers of information (as we represent the passenger flow in four temporal ranges), that is a number lower than the number of the visual variables that we could use (six). In particular the variable *orientation* in our study case does not add any information to our representation of the subway system, as the subway schema is invented and not geo-referenced.

## 4.4 Presentation of the tests

In this dissertation we propose two tests concerning the representation of two different indicators displayed respectively in 2-D and in 2.5-D, at different times. Each visual variable and each combination of visual variables has been used to represent these spatial indicators in different moments.

In the first test we have opted for a representation in two dimensions (2-D), showing the indicator “passenger flow” in a metro network, using as a spatial support for the representation, the metro



linear segments between the metro stops, as shown in fig. 4.1. This indicator is represented in 4 different moments through the use and the combination of 4 visual variables.

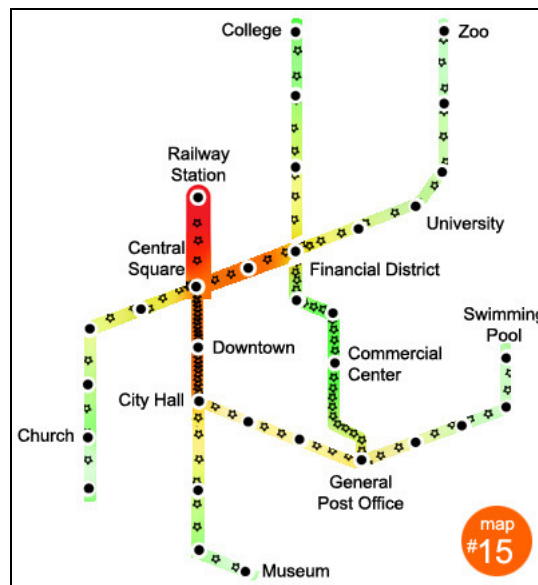


Fig. 4.1; Example of representation of passenger flow in a subway system in four different moments, as proposed in the first test.

This example was chosen because the mixing of visual variables was easy. An alternative could have been an analysis of bi-dimensional areas representing land use or urban activities.

The second test (that will be presented in chapter V) concerns the representation of the indicator “noise” in a quarter of a city (fig. 4.2), using as spatial graphic display the façades of a building; the noise is represented in four different moments through the combination of 4 visual variables, in a 2.5-D dimensionality system, but with a 3-D dimensionality effect. The analysis of 3-D objects is not so easy to deal with, due to the necessity of having 3-D visualizing devices, for that reason, in the second test we have opted for representing 2.5-D objects with 2-D layouts.

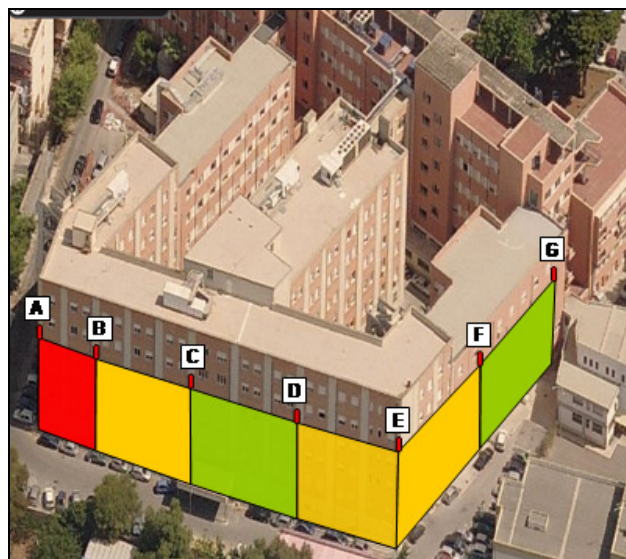


Fig. 4.2; Example of representation of noise on the façades of a building, using colours to indicate different noise levels, as proposed in the second test.

Another interesting test on visual variables, could concern the indicator “Internet data flow” using as background/scenario a map of the world, and using the arches diagrams as a graphic display to represent the flows of data, (see fig. 4.3), in a 2.5-D real dimensionally system, but with a 3-D dimensionality effect

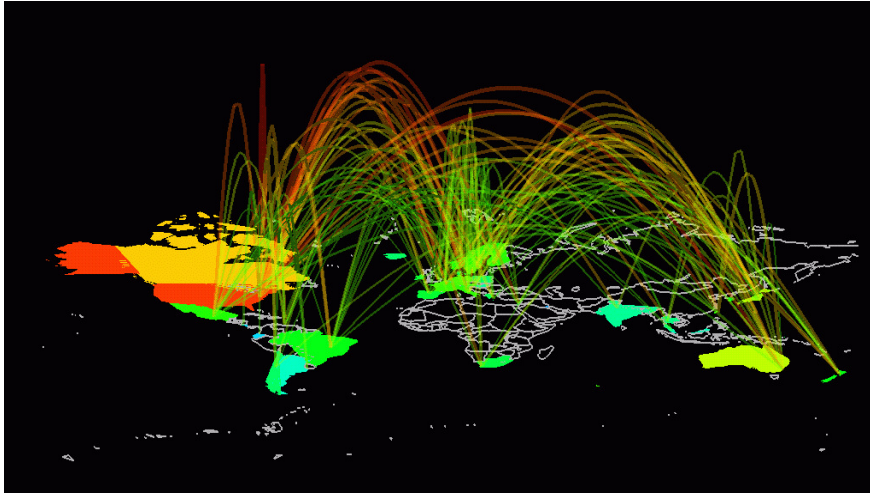


Fig. 4.3; Example of representation of “Internet data flow” in Europe in four different moments indicated with four different colours, while the intensity of the flow is given by the high of the arches.

As we can see the visual variable *color* can be used to indicate the “Internet data flow” at different moments (see fig. 4.3), and at the same time the high of the arcs can indicate different levels of Internet data flow.

## 4.5 Structure of the tests

The structure of each test is the following:

1. introduction and instructions for the execution of the test
2. training with three examples of maps
3. test execution
4. registering credentials and comments of the tested user
5. evaluation and score of the tested user.

Each test is structured in sections; for each section a group of thematic maps is presented. In the first test the maps show the passenger flow in a subway system that is represented in different moments using different combinations of visual variables. There are two types of potential users of these maps: experts in GIS and information visualization (city transport designers and stakeholders) and non-expert users (passengers, citizens). Hence we divided these potential users in two separate groups. The test conditions are very important: it was decided that the tested persons will have a very short time to answer the questions. Answers should be given in a very short time in order to catch the perception of the observer and not his reasoning. The tested users have to indicate what information is given by the map. The answers given will be compared with the correct answers and the number of wrong answers will be counted, generating the interpretation of the maps with a different combination of visual variables.

Totally in the test, are proposed 15 maps that represent all the possible combinations for the four variables previously described and selected:

1. four maps represent the four visual variables
2. six maps represent the combination of two of the four visual variables
3. four maps represent the combination of three of the four visual variables
4. one map represents the combination of all four variables simultaneously.

We have proposed the following groups of questions to the tested users:

1. one question for each of the four maps with only one visual variable represented
2. two questions for each of the six maps with two combined visual variables
3. three questions for each of the four maps with three combined visual variables
4. four questions for the map with all four visual variables presented simultaneously.

The total number of questions proposed in the test to users is 32 for each test; the maximum time available to answer is 25 seconds for each question, so 800 seconds for each test, that means 13,4 minutes of duration of each test.

Some considerations have brought to the reduction of the original number of visual variables analysed in our tests; it was observed that the use of all the variables of Bertin's model would have lead to a very high number of possible combinations of visual variables and so would have lead to a very high number of possible questions in the tests.

## 4.6 Description of the test #1 in 2D: “passenger flow in a subway system”

The first test concerns the representation of the passenger flow in a subway system in two dimensions (2D); in order to represent the passenger flow in different temporal ranges we have selected the following four visual variables *color intensity*, *texture*, *color hue* and *size* (see fig. 4.4) that we will respectively indicate with the short names: *intensity*, *texture*, *color* and *size*.

An invented subway system schema is used as a graphic support to represent the indicator of the passenger flow at different moments: 1) on working days, during rush hours, 2) on week-ends, 3) on vacation days 4) current passenger flow. The passenger flow is represented in these four different moments using the four visual variables: *intensity*, *texture*, *colour* and *size* (fig. 4.5).

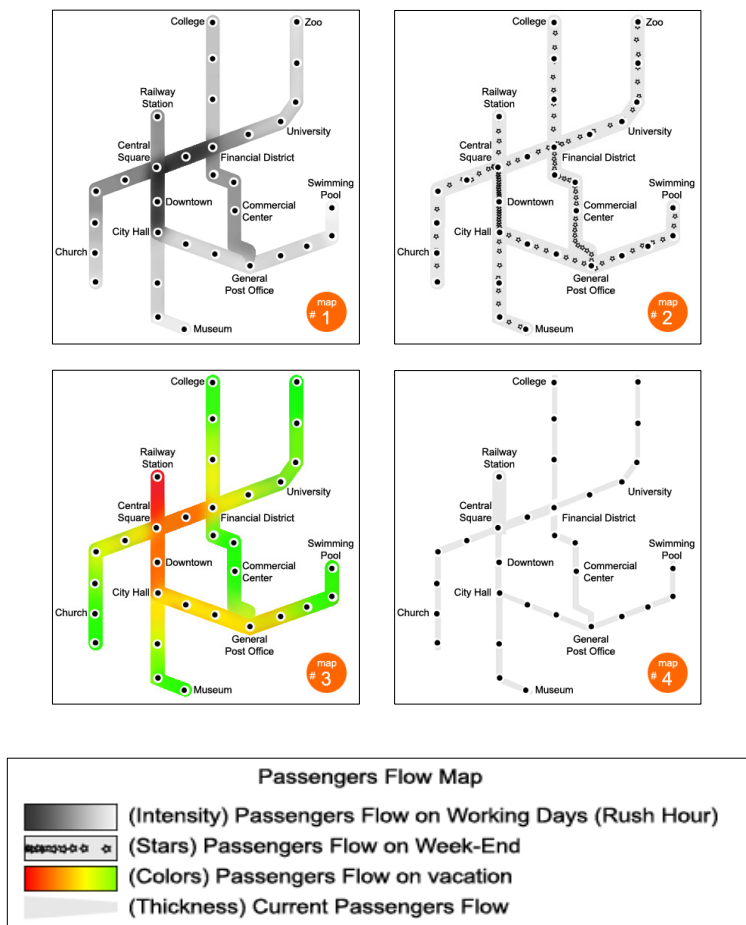


Fig. 4.5; Example of a subway map on which the indicator “flow of passengers” is represented in four different moments, each moment is represented with one of the four selected visual variables: in clockwise order; *intensity*, *texture*, *size* and *colour*, as described in the legend.

The following figure, (fig. 4.6) shows the subway schema with different combinations of visual variables.

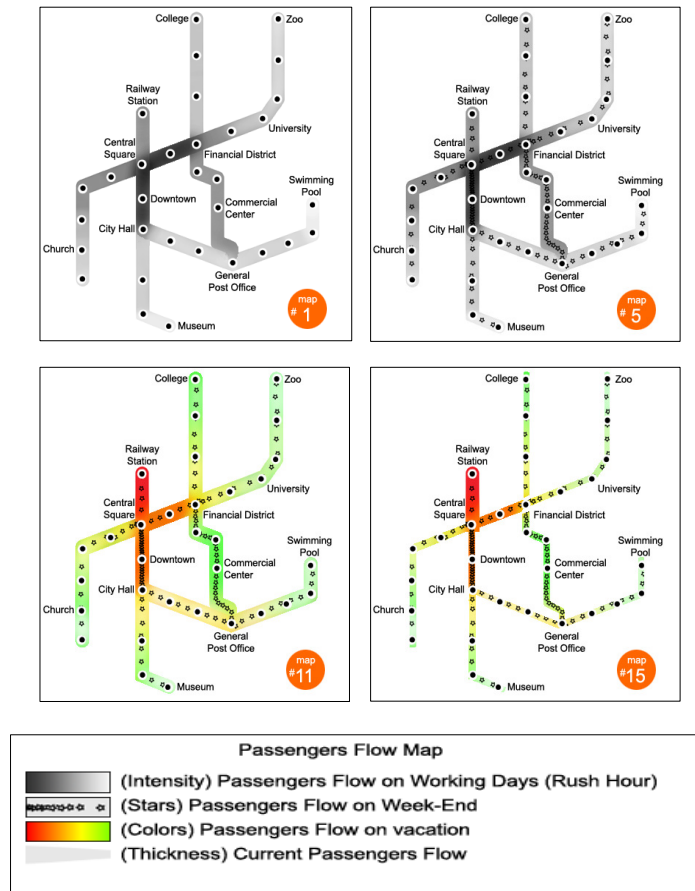


Fig 4.6; Four examples in clockwise order of the combination of respectively four visual variables used to represent the passenger flow in a subway system.

We decided to determine the efficiency of the combinations of visual variables in the transmission of information verifying whether the information given by a single visual variable in the representations is clear for the user. In order to check the efficiency the tested user was asked to interpret the information presented in different maps. For example in the fig. 4.7 (map n.15) the user should indicate which the passenger flow on week-ends between two Metro stops is; this flow is indicated in the map using the variable texture. If the answer of the user is wrong, we can conclude that the variable texture is not clear in this particular combination. Each question is presented like in the screenshot in fig. 4.7.

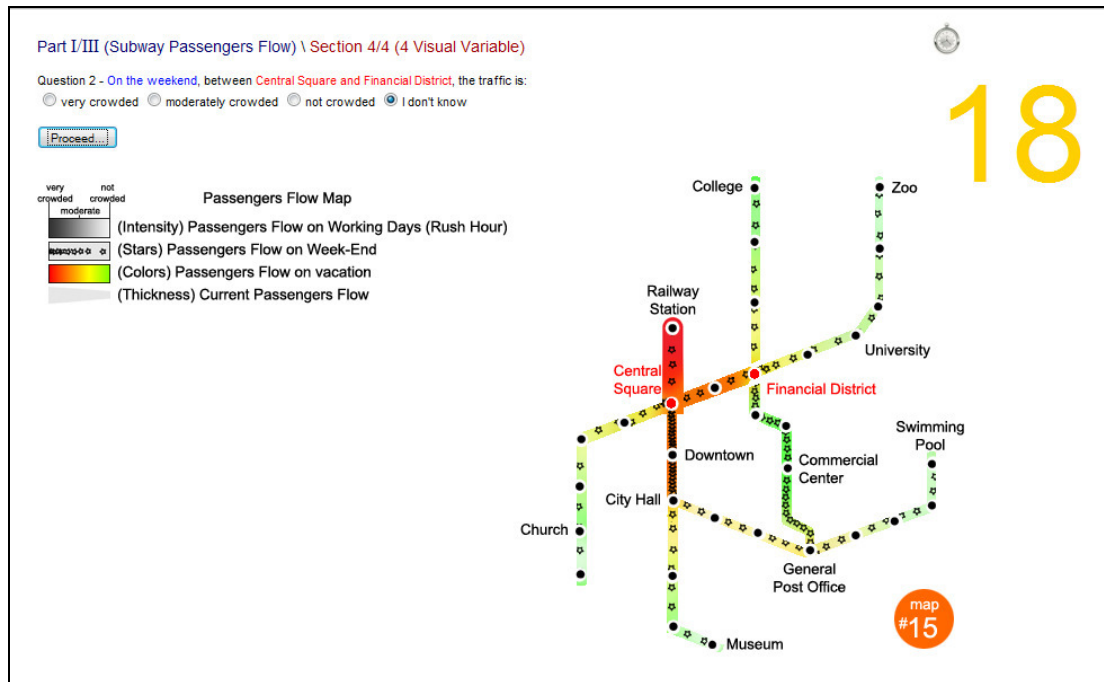


Fig 4.7: Screenshot of one of the questions proposed in the test: the tested user has to understand the passenger flow between two metro stops on the week-ends.

In the example given in fig. 4.7, the tested user has to indicate the passenger flow during week-ends between two metro stops (Central square and Financial district). The question permits to verify the efficiency and readability of the variable *texture* in co-existence with the other three visual variables, *color*, *size* and *intensity*. The time to answer the question is very short in order to avoid that the tested person gets the correct answer through a long reasoning; the time is measured by a timer in the upper right of the interface-screen of the test.

In order to make statistically valid the results of the test, we have tested more than 100 people including almost 50 expert and 50 non-expert users, but the test is still online and the users grow and the results are automatically updated.

## 4.7 Instructions of the online test #1

In order to consult all the representations proposed in the tests we invite the reader to consult the test #1 about the representation of passenger flow in a subway system, that is available online at the link: <http://www.metrorosamarina.altervista.org/index.php>.

The passwords to access the test are: for expert users “roma”, and for non-expert users “puglia”.

All the 32 maps presented in the online interface of the test are implemented in JavaScript language and in PHP.

## 4.8 Results of the test #1

We give the results in the following tables and graphs, considering the rate of correct answers for each map proposed (with single or multiple visual variables displayed) and the measurement of the medium and maximum time of answer, for each user and map. In figure 4.8, the table of the results of the first section of the test #1, “passenger flow in a subway”.

vl = variable intensity
tx = variable texture
cl = variable color
ta = variable size

Section 1, test #1				
Tested variable	vl	tx	cl	ta
Question n.	1	2	3	4
% of correct answers	82	89	52	86
# of answers	98	95	97	95
# of correct answers	80	85	50	82
# answer out of time	2	5	3	5
maximum time of answer (s)	25	25	25	25
medium time of answer (s)	15	13	12	13

Fig. 4.8; Table of the results of the first section of the test #1: “passenger flow in a subway”

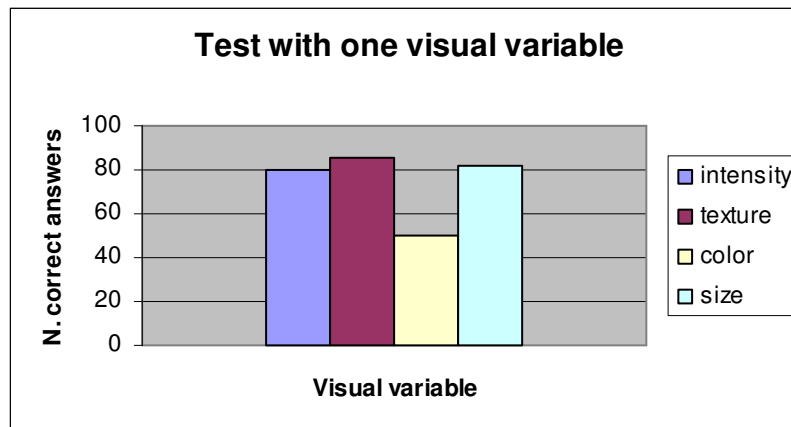


Fig. 4.9; Table of the results of the first section of the test #1; number of correct answers with only one visual variable tested and displayed for each map.

The histogram (fig. 4.9) represents the number of correct answers using only one visual variable at a time for each map to represent the flow of passengers in four different moments.

The interesting result is that the representation of the flow of passengers with the visual variable “color”, usually one of the most common visual variable, in this case has generated less correct answers than the representations that used the other visual variables. This can be due to the choice of using a continuous scale of colours (see fig. 4.10), instead of using a discrete scale of colors, but this also confirms that the visual variable “color” is more suitable to represent qualitative instead of quantitative aspects of data and indicators.



Fig. 4.10; the continuous scale of colors and the difficulties in the interpretation.

In the second section of the test #1, are presented the maps with the combination of two visual variables; in the following figure (fig. 4.11), we show the case of the combination of two visual variables (*intensity* and *texture*) used to represent simultaneously the flow of passengers in two different temporal ranges.

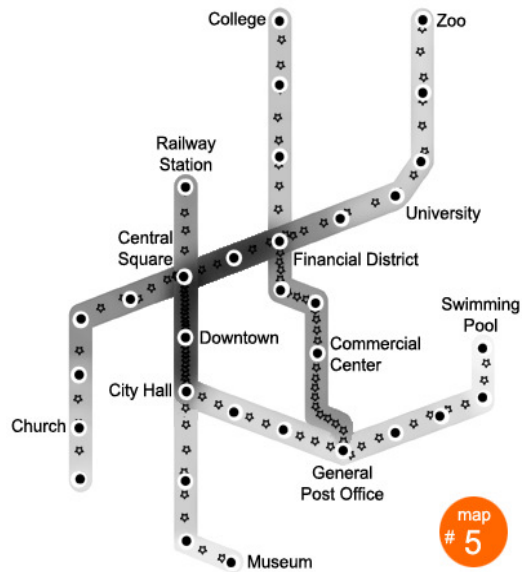


fig. 4.11; the combination of two visual variables, *intensity* and *texture* (represented by the stars).



In figure 4.12, the tables of the results of the second section of the test #1, “passenger flow in a subway”.

Section 2, part 1, test #1						
Combined variables	vl+tx	vl+tx	vl+cl	vl+cl	vl+ta	vl+ta
Tested variable	vl	tx	vl	cl	vl	ta
Question n.	1	2	3	4	5	6
% of correct answers	96	98	82	93	96	98
# of answers	96	98	88	97	95	97
# of correct answers	92	96	72	90	91	95
# answer out of time	4	2	12	3	5	3
maximum time of answer (s)	25	22	25	21	25	23
medium time of answer (s)	20	17	23	22	20	15

Section 2, part 2, test #1						
Combined variables	tx+cl	tx+cl	tx+ta	tx+ta	cl+ta	cl+ta
Tested variable	tx	cl	tx	ta	cl	ta
Question n.	7	8	9	10	11	12
% of correct answers	99	96	99	99	96	99
# of answers	99	100	98	99	100	99
# of correct answers	98	96	97	98	96	98
# answer out of time	1	0	2	1	0	1
Maximum time of answer (s)	20	20	23	20	16	18
medium time of answer (s)	15	12	13	12	10	9

Fig. 4.12; tables of the results of the second section of the test #1; number of correct answers with two visual variables tested and displayed for each map.

In figure 4.13 the histogram represents the number of correct answers in case of the combination of two visual variables representing simultaneously the flow of passengers in two different temporal ranges.

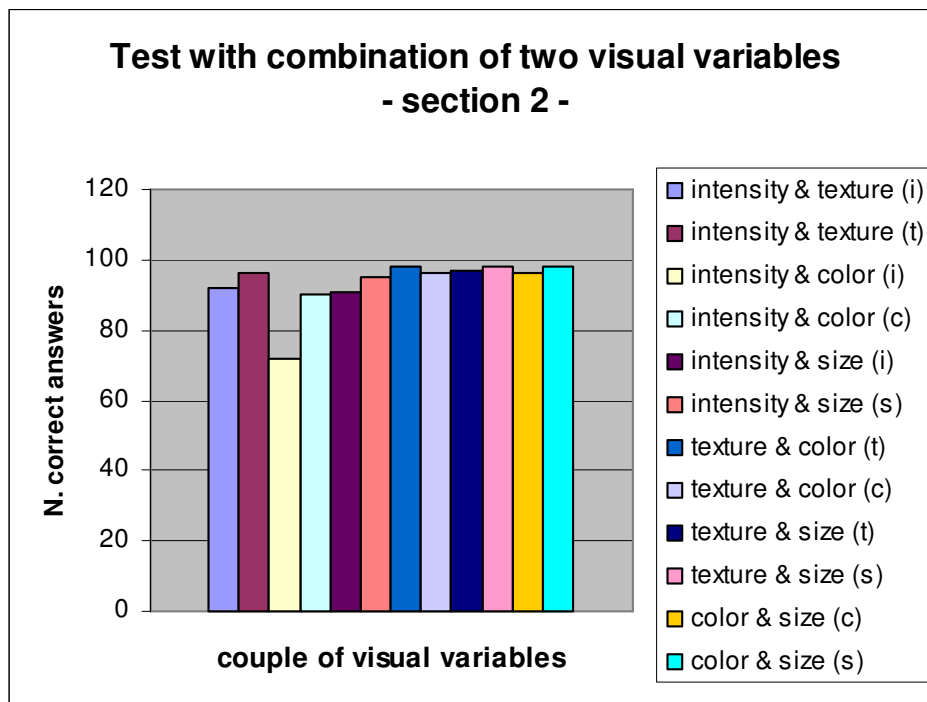


Fig. 4.13; histogram of the results of the second section of the test #1; number of correct answers with two visual variables tested and displayed for each map.

Concerning the perception and the understanding of the combinations of two visual variables, the interesting result is that the representation of the indicator *passenger flow* with the combination of the visual variables in which there is the variable “intensity” resulted less efficient (less rapid to understand) and less effective (understandable), in particular when combined with the visual variable “color”.

Also in this case we can probably attribute this kind of result to the choice of using a continuous scale of colors, and so there are too much shades of colours to indicate only three different “status” of the passenger flow in the subway (very crowded, crowded, not crowded). With a discrete scale it would be easier to understand the intensity of the passenger flow, and to simply understand that the color green means not crowded, orange means moderately crowded and red means crowded.

In figure 4.14, the tables of the results of the third section of the test #1, “passenger flow in a subway”.

Section 3, part 1, test #1						
Combined variables	vl+tx+cl	vl+tx+cl	vl+tx+cl	vl+tx+ta	vl+tx+ta	vl+tx+ta
Tested variable	vl	tx	cl	vl	tx	ta
Question n.	1	2	3	4	5	6
% of correct answers	98	96	97	98	93	97
# of answers	61	70	64	82	89	95
# of correct answers	60	67	62	80	83	92
# answer out of time	39	30	36	18	11	5
maximum time of answer (s)	25	25	25	22	19	17
medium time of answer (s)	22	20	22	19	18	15

Section 3, part 2, test #1						
Combined variables	cl+ta+vl	cl+ta+vl	cl+ta+vl	tx+cl+ta	tx+cl+ta	tx+cl+ta
Tested variable	cl	ta	vl	tx	cl	ta
Question n.	7	8	9	10	11	12
% of correct answers	97	97	97	96	98	99
# of answers	65	75	63	95	98	92
# of correct answers	63	73	61	91	96	91
# answer out of time	35	25	37	5	2	8
maximum time of answer (s)	25	25	25	21	16	20
medium time of answer (s)	20	19	20	14	10	12

Fig. 4.14; the tables of the results of the third section of the test #1, “passenger flow in a subway”.

In figure 4.15 the histogram represents the number of correct answers in the case of the combination of three visual variables for each map, representing simultaneously the flow of passengers in three different temporal ranges.

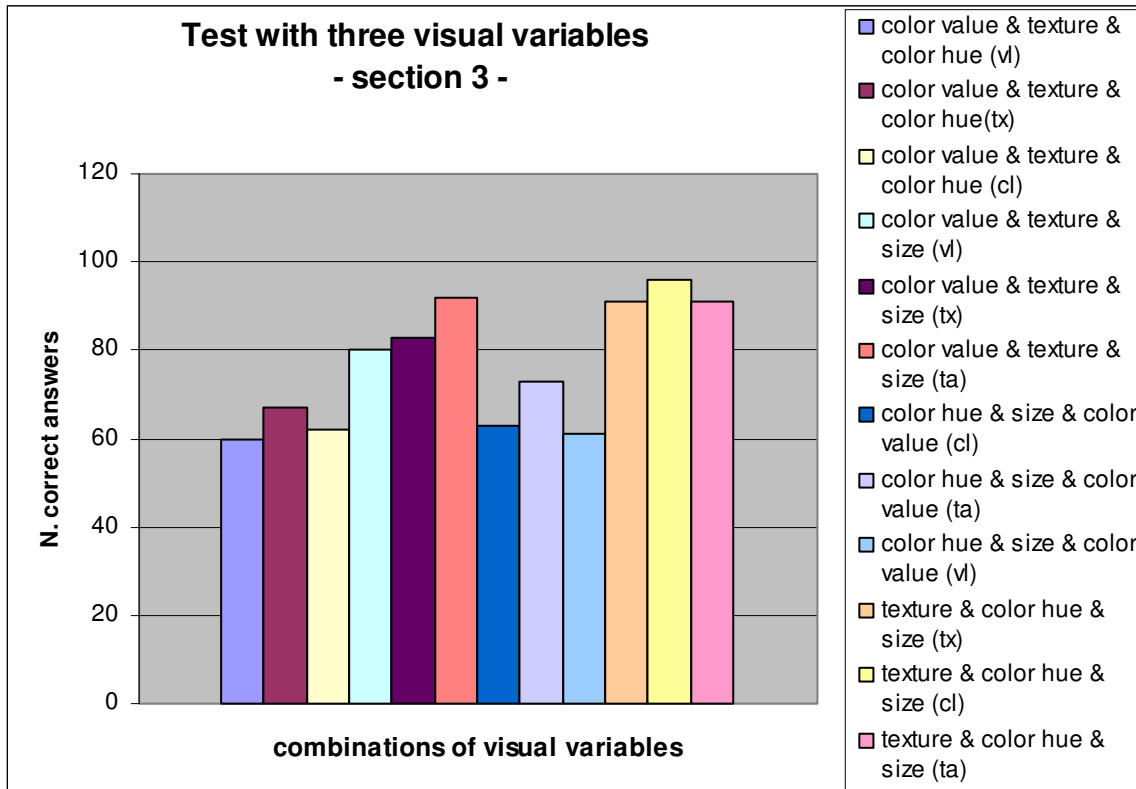


Fig. 4.15; histogram of the results of the third section of the test #1; number of correct answers with the combination of three visual variables for each map.

Concerning the perception and the understanding of the combinations of three visual variables, the interesting result is that the representation of the indicator *passenger flow* with the combination of the visual variables in which there is the variable “intensity” resulted less efficient (less rapid to understand) and less effective (understandable), in particular when combined with the visual variable “color” and “texture”.

Also in this case we can probably attribute this kind of result to the choice of using a continuous scale of colors, and so there are too much shades of colors to indicate only three different “status” of the passenger flow in the subway (very crowded, crowded, not crowded). With a discrete scale it would be easier to understand the intensity of the passenger flow, and to simply understand that the color green means not crowded, orange means moderately crowded and red means crowded.

In figure 4.16, the table of the results of the fourth section of the test #1, “passenger flow in a subway”.

Section 4, test #1				
Combined variables	vl+tx+cl+ta	vl+tx+cl+ta	vl+tx+cl+ta	vl+tx+cl+ta
Tested variable	vl	tx	cl	ta
Question n.	1	2	3	4
% of correct answers	94	87	99	99
# of answers	63	91	96	93
# of correct answers	59	79	95	92
# answer out of time	37	9	4	7
maximum time of answer (s)	25	24	19	21
medium time of answer (s)	19	16	10	11

Fig. 4.16; the table of the results of the fourth section of the test #1, “passenger flow in a subway”.

In figure 4.17 the histogram represents the number of correct answers in the case of the combination of four visual variables for each map, representing simultaneously the flow of passengers in four different temporal ranges.

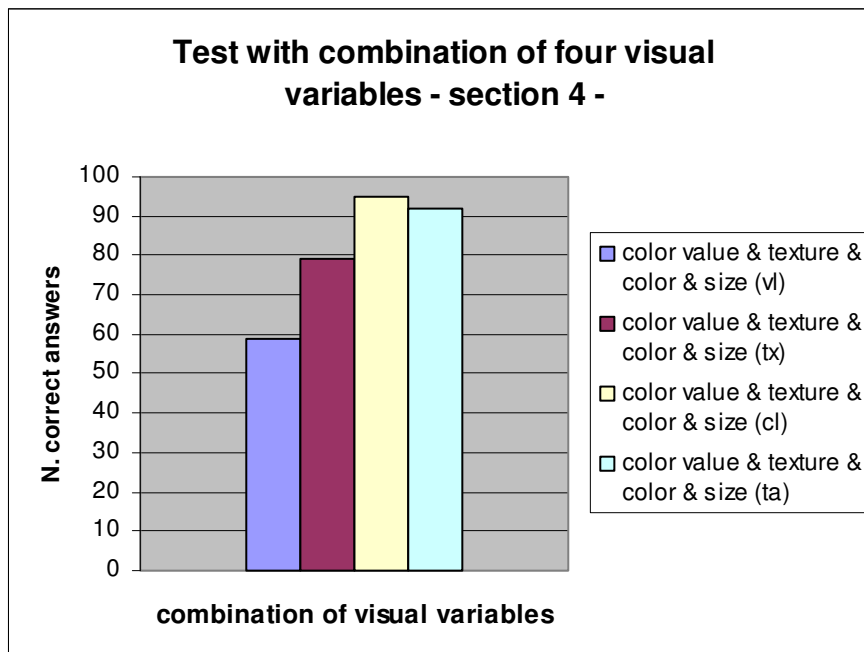


Fig. 4.17; histogram of the results of the fourth section of the test #1; number of correct answers with the combination of four visual variables for each map.

Concerning the perception and the understanding of the combinations of four visual variables, the interesting result is that the representation of the indicator *passenger flow* with the combination of the four visual variables in which there is the variable to depict is the “*intensity*”; so we can say that also in this section the variable that resulted less efficient (less rapid to understand) and less effective (understandable), is the “*intensity*”.

Also in this case we can probably attribute this kind of result to the choice of using a continuous scale of colors, and so there are too much shades of colors to indicate only three different “status” of the passenger flow in the subway (very crowded, crowded, not crowded). This can bring to misunderstand, confuse and not distinguish the visual variable “color” with the visual variable “intensity”.

#### **4.9 Conclusions and remarks for the test #1: “passenger flow in a subway system”**

From the analysis of the results of the test,#1 it is possible to answer to the questions exposed in the paragraph 4.1; the questions proposed were:

1. Which is the visual variable more easily understandable by users for the proposed context?  
And which are the less understandable visual variables?

We can answer that the most understandable variable resulted to be the “texture”, and the less understandable the “intensity”.

2. What is the maximum number of visual variables that can be combined on the same map until the map becomes illegible?

We can say that up to all four visual variables, users were able to understand the questions concerning the map understanding and the map purpose; we only noticed an exception in which in the question concerned the range of time “rush hours at working days” represented by the visual variable “intensity”.

3. Which are the best combinations of visual variables in order to represent a spatial indicator without generating errors of interpretation?

We can say that the best combinations of visual variables resulted to be: “colour” and “size”, “colour” and “texture” and also “texture” and “intensity”.

In conclusion, concerning the perception and the understanding of the combinations of visual variables in the test #1, we can say that the visual variable that, combined with other visual variables, resulted less efficient (less rapid to understand) and less effective (understandable) was the “*intensity*” in particular when combined with the visual variable “*color*”.

## **CHAPTER V**

### **An experimental proposal for the evaluation of data visualization**

#### **- Test #2 -**

## 5.1 Presentation of the Test #2; “urban noise on building facades”

In this chapter we describe the second test elaborated in this thesis work: the test is a 2.5-D representation of the indicator “urban noise”, and the context is a city-quarter in an urban area with streets and buildings.

The facades of a building are used as a graphic support to represent the indicator of the urban noise in four different moments:

1. on working days, during rush hours,
2. on week-ends,
3. on vacation days
4. currently.

As for the first test, also for the second test we have selected only four visual variables to represent the urban noise in four different temporal ranges; the selected visual variables are: *color hue*, *texture*, *color value* and *size* and represent three different levels of “urban noise”: high-medium and low, as shown in the discrete legend in figure 5.1.

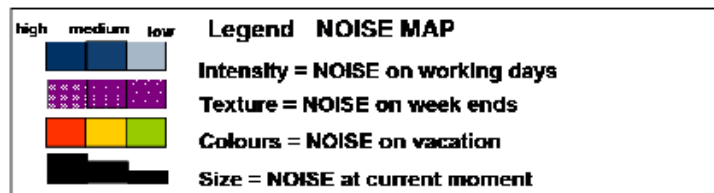


Fig. 5.1; the discrete legend that represents the urban noise in four different temporal ranges and with three different levels: high-medium-low.

Usually visual variables are divided into two groups: the retinal variables such as *color hue*, *color value*, *texture*, *shape*, *size*, and *orientation*, and the planar variables, that indicate the position of the graphic objects: in the plan  $p(x, y)$  or/and in space  $s(x, y, z)$ . In particular some retinal variables are also called geometric variables such as *shape*, *size*, and *orientation*.

For the test #2 we chose to analyze the most common visual variable (*color*) and the less common visual variable (*texture*); we also wanted that in the test should be used both qualitative oriented visual variables (*color hue*), that are usually selected to show qualitative differences of the spatial features, and quantitative oriented visual variables (*size*), that are usually selected to show quantitative differences of the spatial features (see fig. 5.2).

	Points	Lines	Areas	Best to show
Shape	Circle, Triangle, Square, Rectangle	possible, but too weird to show	cartogram	qualitative differences
Size	Small, Medium, Large Circle	Thin, Medium, Thick Line	cartogram	quantitative differences
Color Hue	Red, Green, Blue, Yellow Circle	Red, Green, Blue, Yellow Line	Red, Green, Blue, Yellow Area	qualitative differences
Color Value	Light, Medium, Dark Gray Circle	Light, Medium, Dark Gray Line	Light, Medium, Dark Gray Area	quantitative differences
Color Intensity	Light, Medium, Dark Red Circle	Light, Medium, Dark Red Line	Light, Medium, Dark Red Area	qualitative differences
Texture	Grid, Dotted, Striped, Solid	Dotted, Striped, Solid	Grid, Dotted, Striped, Solid	qualitative & quantitative differences

Fig. 5.2; six fundamental visual variables, in relation with their best use.



Finally some considerations have brought to the reduction of the original number of visual variables that could have been analyzed in our tests; it was observed that the use of all the visual variables would have lead to a very high number of possible combinations of visual variables and so would have lead to a very high number of possible questions in the tests.

The first section of the test concerns the evaluation of efficiency of each visual variable alone representing the spatial continuous indicator “urban noise”; in the second, third and fourth sections of the tests different combinations of visual variables are used to represent the same indicator in different moments, in order to increase the difficulty level of the test. In the second section of the test, there is a combination of two visual variables, in this way it is possible to understand which of the two visual variables is more efficient and rapid to understand. In the third section there is a combination of three visual variables, and in the fourth section there is a combination of four visual variables. To summarize, in the test #2 there are 15 maps that represent all the possible combinations of the four variables previously described and selected:

1. in the first section four maps represent the four visual variables alone,
2. in the second section six maps represent the combination of two of the four visual variables
3. in the third section four maps represent the combination of three of the four visual variables
4. in the fourth section one map represents the combination of all four variables simultaneously.

We have proposed the following groups of questions to the tested users:

1. in the first section one question for each of the four maps with only one visual variable represented
2. in the second section two questions for each of the six maps with two combined visual variables
3. in the third section three questions for each of the four maps with three combined visual variables
4. in the fourth section four questions for the map with all four visual variables presented simultaneously.

The total number of the questions proposed in the test is 32, as in the test #1. In fig. 5.3; it is showed the screen-shot of the first section of the test #2, “urban noise on building facades”.



Fig. 5.3; screen-shot of the first section of the test #2, “urban noise on building facades”.

## 5.2 Objectives of the Test #2; “urban noise on building facades”.

The purpose of the test #2 is to verify the efficiency of visual variables in different contexts and with different methods of representation; for these reasons we have used a different scenario for this test, but we have chosen the same number and type of visual variables of the previous test, and we have also used the same structure of the test, in order to keep a certain consistency and coherence.

The focal questions that we try to answer with the proposed test are:

1. Which is the visual variable more easily understandable (efficient) by users in the proposed contexts? And which is the visual variable less understandable (efficient)?
2. What is the maximum number of visual variables that can be combined on the same map until the map becomes illegible?
3. Which are the best combinations of visual variables in order to represent a spatial indicator without generating errors of interpretation?

## 5.3 Description of the Test #2; “urban noise on building facades”

The test #2 is divided in four sections; in the first section are presented four representations of the indicator “urban noise” in four different moments, each moment is represented with one of the four selected visual variables color value, *texture*, *size* and *color hue*, as previously described. In fig. 5.4 it is shown the representation of the urban noise through the visual variable *intensity* in different shades of blue. The urban noise is monitored and represented in different regular sections of a building façade, indicated with letters from A to G.

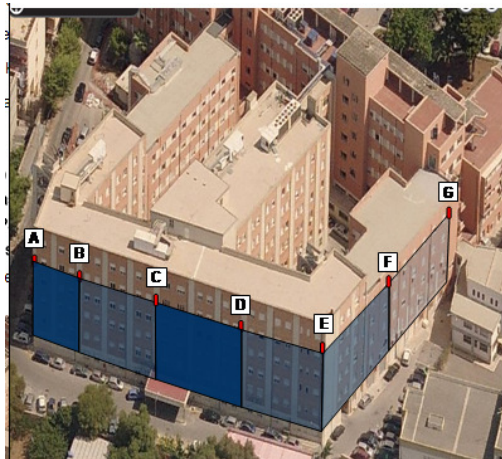


Fig. 5.4; Representation of the urban noise through the visual variable color value.

In figure 5.5 it is shown the representation of the visual variable *texture*, represented using different density of white points on a purple background.



Fig. 5.5; Representation of the urban noise through the visual variable *texture*.

In figure 5.6 it is shown the representation of the urban noise through the visual variable *color*, in a discrete scale from red to green.



Fig. 5.6; Representation of the urban noise through the visual variable *color hue*.

In figure 5.7 it is shown the visual variable *size*, represented using a different extension of the façade sections.

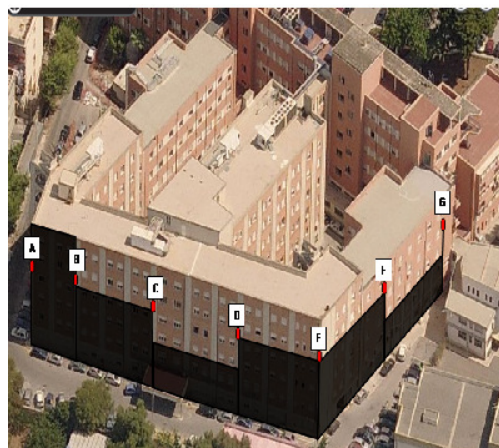
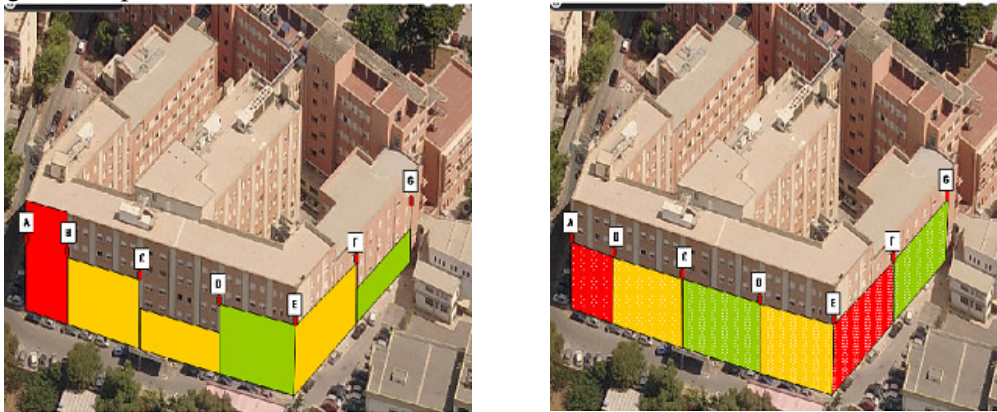


Fig. 5.7; Representation of the urban noise through the visual variable *size*.

In the second section are presented six representations of the indicator “urban noise” in different moments; each representation presents a combination of two of the four selected visual variables (*color value*, *texture*, *size* and *color hue*), and are proposed two questions for each representation in order to analyze the most efficient visual variable for each combination.

Fig. 5.8; Representation of the urban noise



through the combination of the visual variables *size* and *color hue*, and *texture* and *color hue*.

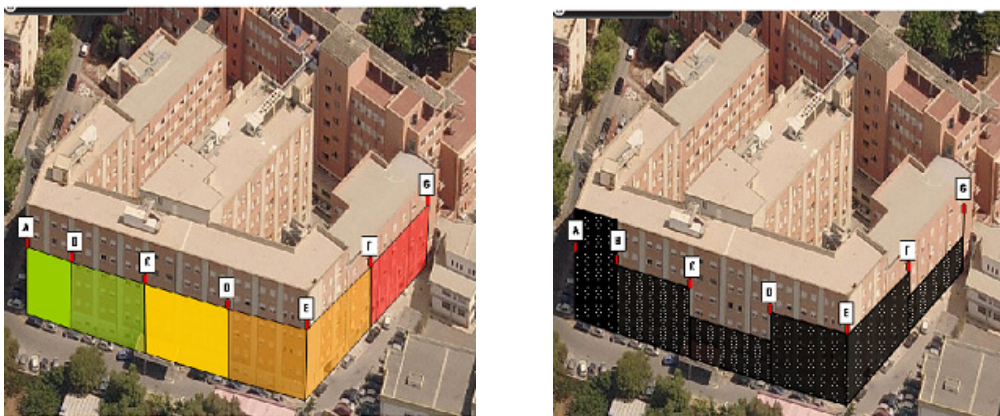


Fig. 5.9; Representation of the urban noise through the combination of the visual variables *color hue* and *color value* and through the combination of *size* and *texture*.



Fig. 5.10; Representation of the urban noise through the combination of the visual variables *size* and *color value*, and *texture* and *color value*.

In the third section are proposed to the users four maps and twelve questions (three questions for each map). In fig 5.10 there is the screen-shot of one of the questions proposed in the third section of the the 2# test;

Section 3/4 (3 Visual Variable)

Question 2 - On the weekend, between E and F, the noise is:  
 high  medium  low  I don't know

Proceed...

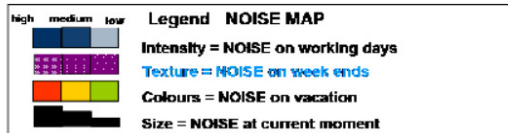


Fig 5.11: Screen-shot of one of the questions proposed in the test.

In the example given in fig. 5.11, the tested user has to indicate the urban noise during week-ends between two points in the same building façade, (the points are indicated with E and F). The question permits to verify the efficiency and readability of the variable “*texture*” in co-existence with the other two visual variables, *color hue*, and *color value*; We have proposed this example in order to show that it is not difficult to misunderstand the correct meaning of the representation, especially if the time to answer is short.

The time to answer is measured by a timer in the upper right side of the interface-screen of the test (in the fig. 5.11 the clock shows that there are still sixteen seconds remaining to answer). In the fourth section it is proposed only one map (fig. 5.12), with the combination of all the four visual variables, and are proposed four questions, one for each visual variable.



Fig. 5.12; Representation of the urban noise through the combination of four visual variables *size*, *color hue*, *texture* and *color value*.

## 5.4 Instructions for the online test #2

We invite the reader to consult the test concerning the representation of the indicator “urban noise” on the facades of a building; the test is available online at the link: <http://www.rosamarina.altervista.org/index.php> and the password to access is “spatial”.

It is also possible to consult all the maps proposed in the test.

In order to better perform the test, an important consideration about the duration of the test, is necessary; the maximum time available to answer to the questions is 25 seconds for each question, so 800 seconds for the 32 questions of the test, that means 13,4 minutes of duration of the whole test. The choice of a short time available to make the test was due to two main reasons:

1. the online tests should not be too long in order to not discourage the tested subject,
2. in order to preserve and catch the instinctive ability of intuitive perception of the tested subject.

Concerning the second point, the time to answer the question is very short in order to avoid that the tested person gets the correct answer through a long reasoning. A final consideration concerning the implementation of the test online: all the 32 maps presented in the test and the online interface are implemented in JavaScript language and PHP.

## 5.5 Results of the test #2

The results of the test #2 are shown in the following tables and histograms. In the elaboration of the results we have considered two main parameters: the rate of correct answers for each question proposed (with single or multiple visual variables displayed in the relative map) and the average and maximum answer time for each question proposed.

In order to make statistically valid the results of the test #2, we have considered the answers of more than 100 volunteer users, but the test is still available online, so the users are growing day by day and the results are automatically updated.

In the following figures 5.13 and 5.14, are reported the table and the histogram of the results of the first section of the test #2 concerning the evaluation of single visual variables.

Section 1, test #2				
Tested variable	vl=color value	tx=texture	cl=color hue	ta=size
Question n.	1	2	3	4
% of correct answers	90	84	96	97
# of answers	100	95	100	98
# of correct answers	90	80	96	95
# answer out of time	0	5	0	2
maximum time of answer (s)	22	25	16	25
medium time of answer (s)	10	12	8	10

Fig. 5.13; Table of the results of the first section of the test#2: “urban noise”  
 The histogram in fig. 5.14, represents the number of correct answers using only one visual variable for each map to represent urban noise in four different moments.

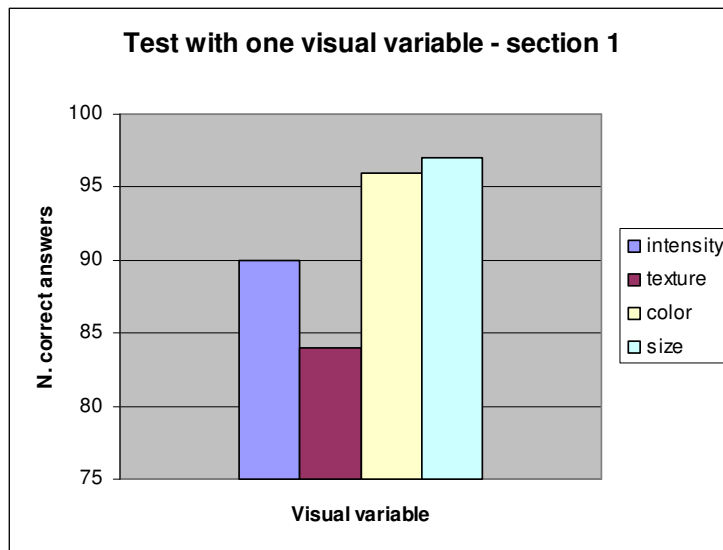


Fig. 5.14; histogram of the results of the first section of the test#2: “urban noise”; number of correct answers for each visual variable tested.

The interesting result is that the representation of the urban noise with the visual variable “*texture*”, has generated less correct answers than the representations that used the other visual variables. This can be due to the choice of using a discrete scale for the visual variable “*color hue*” (see fig. 5.2), instead of using a continuous scale, that in the previous test#1, had generated confusion in the understanding of the visual variable “*color hue*”, but this also confirms that the visual variable “*size*” is the more suitable to represent quantitative aspects of the indicators (as indicated in fig. 5.3) and quantitative data type like the urban noise level,

In the second section of the test we analyzed the representation of urban noise through the combination of two visual variables, as example see figure 5.15, in which are combined the visual variables *color hue* and *size*.



Fig. 5.15; the combination of two visual variables, *color hue* and *size*.

In the following tables (fig. 5.16 and 5.17), are reported the results of the second section of the test #2, concerning the combination of two visual variables used to represent simultaneously the urban noise in two different moments.

Section 2, part 1, test #2						
Combined variables	vl+tx	vl+tx	vl+cl	vl+cl	vl+ta	vl+ta
Tested variable	vl	tx	vl	cl	vl	ta
Question n.	1	2	3	4	5	6
% of correct answers	84	84	97	98	98	99
# of answers	95	95	98	100	96	99
# of correct answers	80	80	95	98	94	98
# answer out of time	5	5	2	0	4	1
maximum time of answer (s)	25	25	25	8	25	25
medium time of answer (s)	12	12	8	10	10	12

Fig. 5.16; Table n.1 of the results of the second section of the test#2: “urban noise”

Section 2 , part 2, test #2						
Combined variables	tx+cl	tx+cl	tx+ta	tx+ta	cl+ta	cl+ta
Tested variable	tx	cl	tx	ta	cl	ta
Question n.	7	8	9	10	11	12
% of correct answers	95	98	97	101	100	99
# of answers	95	98	93	97	100	99
# of correct answers	90	96	90	98	100	98
# answer out of time	5	2	7	3	0	1
maximum time of answer (s)	25	18	25	18	16	16
medium time of answer (s)	8	10	10	12	8	10

Fig. 5.17; Table n.2 of the results of second section of the test#2: “urban noise”.



In the figure 5.18 the histogram represents the number of correct answers in case of the combination of two visual variables.

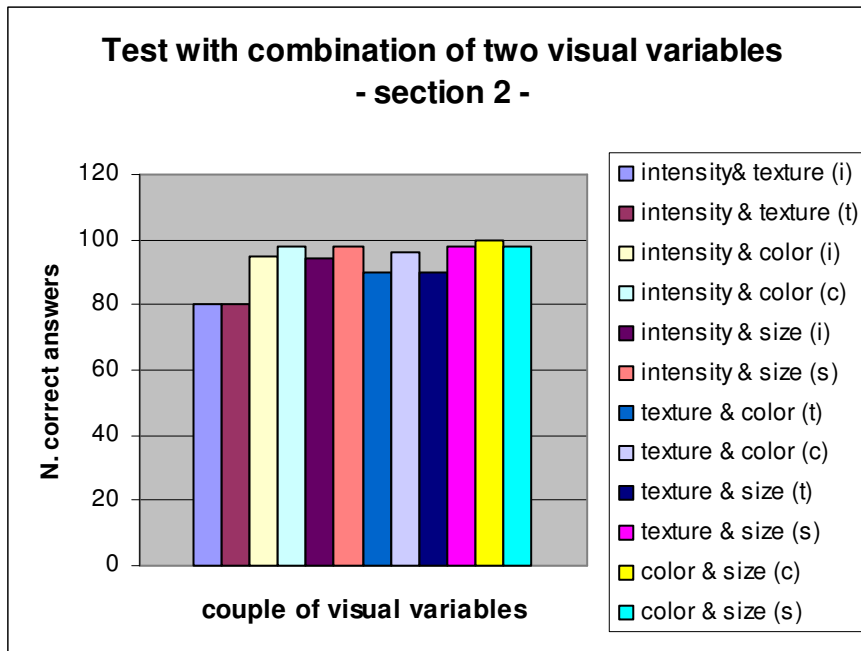


Fig. 5.18; histogram of the first results of the test; number of correct answers with up to two visual variables combined and tested for each map.

Concerning the combinations of two visual variables, the interesting result is that the representation of the indicator “urban noise” resulted less efficient (less rapid to understand) and less effective (understandable), in particular when the visual variable *color value* (or *intensity*) was combined with the visual variable *texture* and the visual variable *color hue*.

Also in this case we can probably partially attribute this kind of result to the choice of using white points to represent the texture, that do not result very clear when overlaid on a coloured background. But this also confirms the fact that the visual variable texture is weak to represent quantitative data type, and more indicated to represent categorical data type, like land cover classes, etc.

Also the visual variable color value (or intensity) is weak to represent quantitative data, and it is more indicated to represent continuous ordered data type.

In the following figures 5.19 and 5.20, are reported the tables and the histograms of the results of the third section of the test #2, concerning the combination of three visual variables used to represent simultaneously the urban noise in three different temporal ranges.

In the following tables (fig. 5.19), are reported the results of the third section of the test #2, concerning the combination of three visual variables used to represent simultaneously the urban noise in three different moments.

Section 3, part 1, test #2						
Combined variables	vl+tx+cl	vl+tx+cl	vl+tx+cl	vl+tx+ta	vl+tx+ta	vl+tx+ta
Tested variable	vl	tx	cl	vl	tx	ta
Question n.	1	2	3	4	5	6
% of correct answers	89	94	99	98	96	100
# of answers	90	95	98	93	95	100
# of correct answers	80	89	97	91	91	100
# answer out of time	10	5	2	7	5	0
maximum time of answer (s)	25	25	20	25	23	22
medium time of answer (s)	20	17	9	18	15	8

Section 3, part 2, test #2						
Combined variables	cl+ta+vl	cl+ta+vl	cl+ta+vl	tx+cl+ta	tx+cl+ta	tx+cl+ta
Tested variable	cl	ta	vl	tx	cl	ta
Question n.	7	8	9	10	11	12
% of correct answers	99	99	98	96	99	98
# of answers	98	100	89	92	100	100
# of correct answers	97	99	87	88	99	98
# answer out of time	2	0	11	8	0	0
maximum time of answer (s)	20	17	25	25	8	9
medium time of answer (s)	8	9	19	20	8	9

Fig. 5.19; tables of the results in the case of the combination of three visual variables for each map.

In the figure 5.20 the histogram represents the number of correct answers in the case of the combination of three visual variables for each map.

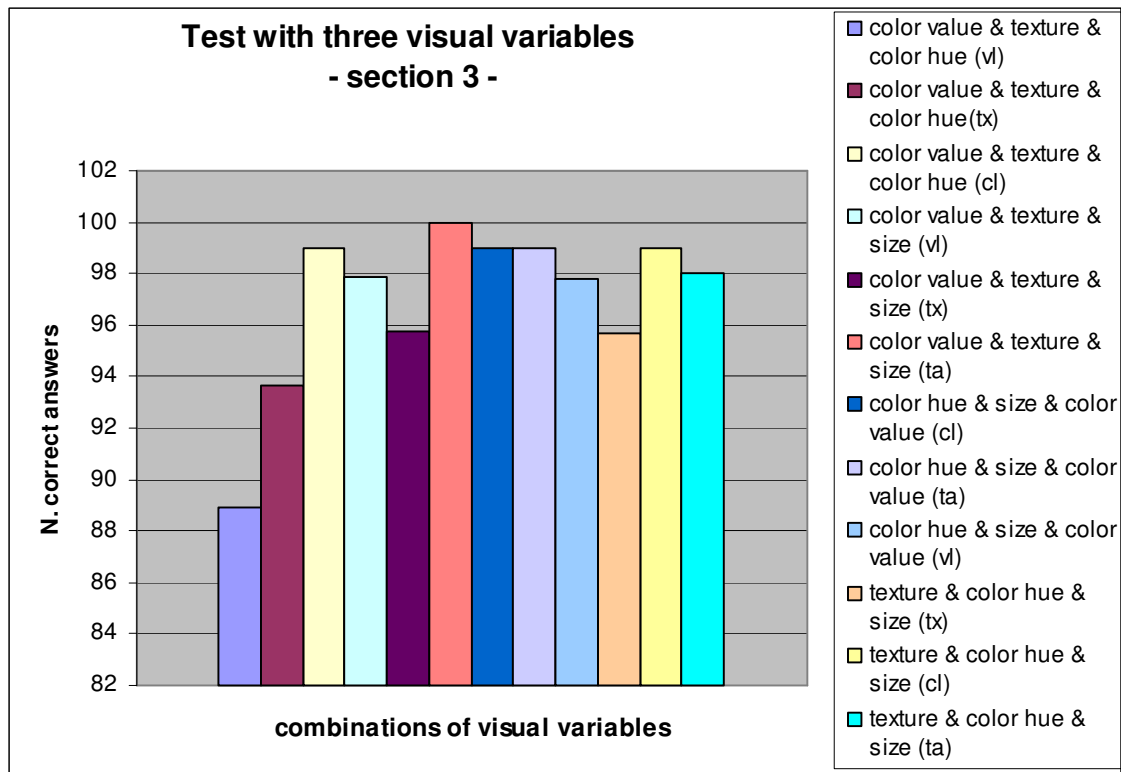


Fig. 5.20; histogram of the results of the third section of the test #2; number of correct answers with up to three visual variables combined and tested for each map.

Concerning the combinations of three visual variables, the interesting result is that the representation of the indicator “urban noise” resulted less efficient (less rapid to understand) and less effective (understandable), in particular when the visual variable *color value* (or *intensity*) was combined with the visual variable *texture* and the visual variable *color hue*, and the question asked to detect the visual variable *color value* (indicated with *vl* in the legend).

This confirms the fact that the visual variable *color value* (or *intensity*) is weak to represent quantitative data, and it is more indicated to represent continuous ordered data type, and confirms that humans are less sensible to catch this visual variable.

This also confirms the idea that the visual variable *texture* is weak to represent quantitative data type, and it is more indicated to represent categorical data type, like land cover classes, etc.

In the figure 5.21 the table of the results of the fourth section of the test #2, in the case of the combination of four visual variables for each map.

Section 4, test #2				
Combined variables	vl+tx+cl+ta	vl+tx+cl+ta	vl+tx+cl+ta	vl+tx+cl+ta
Tested variable	vl	tx	cl	ta
Question n.	1	2	3	4
% of correct answers	44	61	100	98
# of answers	45	90	100	98
# of correct answers	20	55	100	96
# answer out of time	55	10	0	2
maximum time of answer (s)	25	25	15	20
medium time of answer (s)	12	12	8	10

Fig. 5.21; table of the results of the fourth section of the test #2; number of correct answers with up to four visual variables combined and tested for each map.

In the figure 5.22 the histogram represents the results of the fourth sections of the test #2; number of correct answers in the case of the combination of four visual variables for each map.

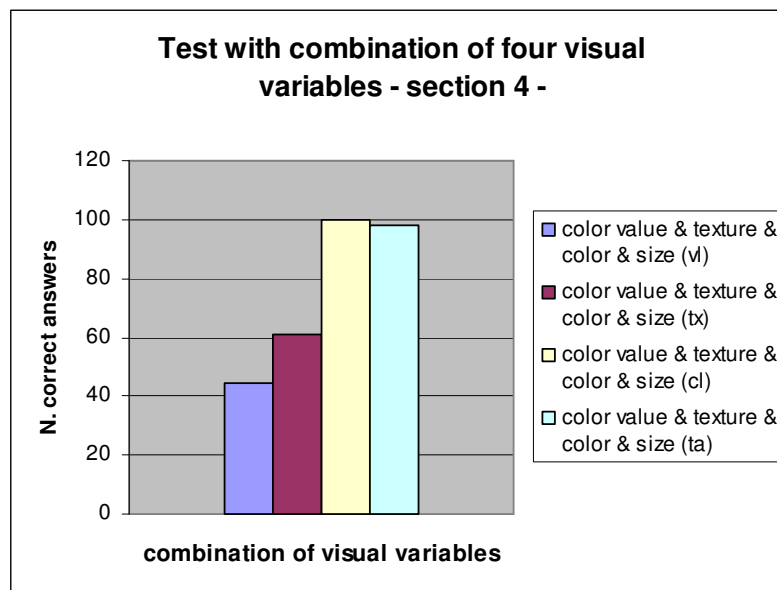


Fig. 5.23; histogram of the results of the fourth section of the test #2; number of correct answers with up to four visual variables combined and tested for each map.

Concerning the combinations of four visual variables, the interesting result is that the representation of the indicator “*urban noise*” resulted less efficient (less rapid to understand) and less effective (understandable), in particular when it was asked to detect the visual variable *color value* (or *intensity*), indicated with *vl* in the legend, and the visual variable *texture*, indicated with *tx* in the legend.

This confirms the fact that the visual variable *color value* (or *intensity*) is weak to represent quantitative data, and it is more indicated to represent continuous ordered data type, and confirms that humans are less sensible to catch this visual variable.

This also confirms the idea that the visual variable *texture* is weak to represent quantitative data type, and it is more indicated to represent categorical data type, like land cover classes, etc.

## 5.6 Conclusions and remarks for the test #2

From the analysis of the results of the test #2, it is possible to answer to the questions exposed in the paragraph 5.2:

1. Which is the visual variable more easily understandable by users for the proposed test #2? And which are the less understandable visual variables?

We can answer that the most understandable variable resulted to be the color hue, and the less understandable the “color value” in all the four sections of the test #2.

2. What is the maximum number of visual variables that can be combined on the same map until the map becomes illegible?

We can say that up to all four visual variables, users were able to answer the questions concerning the map understanding and the map purpose.

We only noticed an exception; the question concerning the visual variable “color value”, was not always caught and understood by users.

3. Which are the best combinations of visual variables in order to represent a spatial indicator without generating errors of interpretation?

We can say that the best combinations of visual variables resulted to be: color hue and size, and color hue and texture.

In conclusion, concerning the perception and the understanding of the combinations of visual variables, we can say that the visual variable that combined with other visual variables results less efficient (less rapid to understand) and less effective (understandable) is the *color value*, in particular when combined with the visual variable *color hue*.

## CHAPTER VI

### Conclusions and future perspectives

#### 6.1 Initial purposes and objectives of the dissertation

Generally there are many difficulties to represent the different type of geo-spatial data simultaneously, in order not only to communicate, but also to extract new information for both private citizens and public decision-makers; new research fields known as “information visualization” and “geovisualization” can be of great interest to address these issues, and are fundamental tools in the development and in the management of smart cities and territorial intelligence.

Information Visualization and geovisualization are now accepted and growing fields, but questions remain about the use of them and the maturity of new visualizations.

Actually there are several researches with detailed analysis of how to display data for precise, effective, quick analysis; there are several books with indications for the design of high-resolution displays, for the representation of few or big amount of data, of different kinds of data, time-series, relational graphics, and multivariate maps, etc. There are also studies about the detection of graphical deception, the analysis of design variation in relation with data variation, and on the aesthetics of data graphical displays.

The critical point is that there are few usability studies and controlled experiments that could permit a generalization and a standardization of the results in these fields. For this reason we thought that a systematic development of a set of tests would facilitate the comparison of techniques and help to identify their potentialities under different conditions. With the implementation of these empirical tests we wanted to suggest an approach and a method to identify preferences and problems in data visualization, in order to improve the understanding of data visualization and to simplify the interpretation of visual data.

We cite a phrase of MacEachren, who, in order to understand how and why certain displays are more successful than others for decision making, said that: “an empirical evaluation of design principles, and a systematic look into the relationships between perceptual salience and thematic relevance in visualization design is needed”, [MacEachren & Kraak, 2001].

For this purpose in this work we have proposed an experiment to evaluate the problems in the use of Bertin’s visual variables [Bertin, 1967], and their combinations used to represent different indicators in different temporal ranges. The experiment consisted in the evaluation of two different study-cases in which two indicators were displayed using four different visual variables. The indicators were represented in two different contexts and are displayed in different dimensional systems (2D and 2.5D); the objectives of the study have been

1. the assessment of the most efficient and effective visual variable,
2. the assessment of the most efficient and effective combination of visual variables,
3. the assessment of the maximum number of visual variables that can be combined without making the maps illegible.

## 6.2 Achievement of the objectives and synthesis of the results

The objectives listed in the previous paragraph have been achieved; the first objective was achieved ranking the most effective visual variables by the highest number of correct answers to the questions concerning the visual variables combinations (or by the number of errors detected for each visual variable analyzed).

The second objective was achieved, with the individuation of the best combinations of visual variables by the number of errors detected for each combination (or by the number of the combinations of visual variables that produced the highest number of correct answers). We also determined the maximum number of visual variables that permitted a clear representation of spatial indicators, that finally is four; so we can say that it is possible to use, at least, up to four visual variables in a spatial data representation, and not only two like it is common to see (usually the most common visual variables *color* and *size*).

We have also found that the four tested visual variables (*color*, *value*, *size*, *texture*) are all “attention guiding”, as people performed the probability of a correct interpretation of each of them significantly above the probability of chance (50%) in detecting, localizing and describing a change in the display. This is in accordance to the summary of results presented in Wolfe and Horowitz’s meta study [Wolfe and Horowitz, 2004] on attention guiding attributes. These authors listed color, motion, orientation, and size as visual variables capable to guide visual attention. However, unlike Bertin, Wolfe and Horowitz did not provide a ranking of the attributes. In our experiment, our empirical results can provide some evidence to order the Bertin’s visual variables that we have tested. Since early eye movement studies on visual displays [Buswell, 1935] and since the research described in the paper “Evaluating the effectiveness and efficiency of visual variables” [Yarbus, 1967], it has been known that people concentrate their attention on interesting and informative scene regions [Henderson & Ferreira, 2004].

In our experiment we found that the visual variable “size” was the most efficient and effective variable to guide viewer’s attention in quantitative data representations in 2D and 2.5D maps. According to Bertin, size is the only visual variable that has quantitative, ordered, selective and disassociative characteristics. In fact, Bertin attributes to the visual variable “size” the most “disassociativeness”. The visual variable *size* emphasizes a change, and it could be an aspect of interest, and a very useful quality to guide attention. Also the texture gave good results in terms of understanding. On the contrary the visual variable *intensity* appeared to be least effective and efficient of the four tested visual variables. For the visual variable *color* the result is not so clear; color seems to have a slight (but non significant) advantage towards the visual variable *intensity*, perhaps because we decided to use in the legend and in the maps a continuous scale of colors, instead of a discrete one.

In Bertin’s system, the visual variable *color* differs from the visual variable *size* only in the lack of a quantitative characteristic, thus we would have expected color to perform better than size for change detection.

These results might perhaps support Wolfe & Horowitz’s evaluation of luminance polarity (contrast in brightness or color value) as an attention-guiding attribute.

Other interesting questions were:

1. Do we achieve the same results in the two different study-cases?
2. Do the context, the display environment, the data type, and the dimensions of the display (2D, or 2.5D) influence the perception and the understanding of visual variables?

Our results permitted us to answer “yes” to the first question, and to answer “no” to the second question;

The small differences in the results, and the general preference for the metro maps and not for the facades maps, can be explained by the display of the metro map that being visually the most used and usual and the least complex (having fewer visual distracters), according to the computational

saliency model. Even if the representation of the passenger flow in the metro maps, had two disadvantages: the continuous scale of colors and the smaller size of the graphic support (the thickness of a metro line segment, instead of the areas of the facades surfaces)

Towards the issue concerning the more or less familiar graphic interfaces,, there is an interesting consideration that was done by Raskin [Raskin, 1994]. The term “intuitive” or “natural user interfaces” is often listed as an important trait in usable interfaces, but Raskin has discouraged using this term in user interface design, saying that “easy to use interfaces” are often considered easy because of the user's exposure to previous similar and conventional systems, so the term “familiar interfaces” should be preferred. This position is sometimes also illustrated with the remark “The only intuitive interface is the nipple; everything else is learned”.

Bruce Tognazzini [Tognazzini, 1992] even denies the existence of “intuitive” interfaces, since such interfaces must be able to perceive the patterns of the user's behaviour and draw inferences; instead, he advocates that “users could intuit the workings of an application by seeing it and using it”

### 6.3 Research Future Perspectives

A further development of the experiment could concern the comparison between visual variables in common interfaces and visual variables in very unusual interfaces, in order to better evaluate the intuitive capacities of the tested user. A further development of the experiment could also concern the identification of the most appreciated methods of representation of spatial data, through the comparison of different maps (and different methods of representation) representing the same indicator with the same visual variables. This will also lead to a classification of the more usual or unusual methods of representation and visualization. A further implementation of the tests could be the comparison of different indicators displayed on the maps, with the same method of representation. Moreover it will be interesting to try to determine the best visual variables to represent simultaneously different spatial indicators in different times and with different methods of representation.

In the next steps of the research it would be useful the study of how *motion*, *animation* and *interactivity* can be considered as new visual variables to add to the Bertin's model; it is surely a way to improve interactivity between pc and human-user, that can improve the customization and personalization of spatial objects and indicators in data mapping.

Finally, the continuation of this research could concern the evaluation of the efficiency of tri-dimensional, interactive and real time representations of spatial data and could analyze the interaction between the user and the system of representation, especially underlining the differences between men and women perception.

In this way it is possible to explore and improve different methods and new ways of data representation that can make urban and extra-urban data visualization and interpretation easier and more available for everybody.



## Bibliography

- [Amoroso, 2010], Amoroso N., *The Exposed City, Mapping the Urban Invisibles*, New York: Routledge, 2010, 176 p., ISBN: 978-0-415-55180-9.
- [Andrienko et al., 1996], Andrienko N., Andrienko G., *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*, Berlin, Springer, 2005.
- [Andrienko et al., 2005], Andrienko N., Andrienko G., *Visual Analytics of Movement*, Springer, 2013.
- [Andrienko et al., 2008], Andrienko N., Andrienko G., Fabrikant S. I. et al., *Geovisualization of dynamics, movement and change: key issues and developing approaches in visualization research*, *Information Visualization archive*, Volume 7, Issue 3, June 2008, p. 173-180, Palgrave Macmillan, 2008.
- [Behren, 2008] Christian Behrens's list of Design Pattern categories
- [Bertin, 1967], Bertin, J., *Semiology of Graphics: Diagrams, Networks, Maps*. Madison: University of Wisconsin Press, 1983; (French edn., 1967).
- [Brimicombe, 1999], Brimicombe A.J., *Small may be beautiful, but is simple sufficient?*, *Geographical and Environmental Modelling*, 1999, Vol.3, p. 9-33.
- [Buswell, 1935], Buswell, G.T., *How People Look at Pictures*. Chicago: University of Chicago Press, 1935.
- [Card, 1997], Card, S. K., Mackinlay, J. D., *The Structure of the Information Visualization Design Space*. *Proceedings of IEEE Symposium on Information Visualization (InfoVis 1997)*, Phoenix, Arizona, Color Plate 125, 1997, p. 92-99.
- [Card et al., 1999], Card, S. K., Mackinlay, J. D., Shneiderman, B., *Using Vision to Think. Readings in Information Visualization*, San Francisco : Morgan Kaufmann, 1999.
- [Carpendale, M. S. T, 2003], Carpendale, M. S. T., *Considering Visual Variables as a Basis for Information Visualisation*, Department of Computer Science, University of Calgary, Canada, 2003.
- [Chi et al., 1997], Chi Ed H., Barry P., Riedl J. T., Konstan J., *A Spreadsheet Approach to Information Visualization*. *Symposium on Information Visualization (InfoVis 1997)*, Phoenix, Arizona, USA, 1997, p. 17-24.
- [Chi et al., 1998], Chi Ed H., Riedl. J. T., *An Operator Interaction; Framework for Visualization Systems*. *Symposium on Information Visualization (InfoVis 1998)*, Research Triangle Park, North Carolina, USA, 1998, p. 63-70.
- [Chi, 1999], Chi. Ed H., *A Framework for Information Visualization Spreadsheets*, Ph.D. Thesis, University of Minnesota, Minnesota, USA, March 1999.

[Chi, 2000], Chi E. H., A Taxonomy of Visualization Techniques using the Data State Reference Model. Proceedings of IEEE Symposium on Information Visualization (InfoVis 2000), IEEE Computer Society Press, 2000, p. 69-75.

[Cottam J.A, 2012], Cottam J.A., Watch This: A Taxonomy for Dynamic Data Visualization, IEEE Conference on Visual Analytics Science and Technology, Seattle, WA, USA, October 14-19, 2012.

[Damy et al., 2007], Damy, Girardot J-J., Masselot C., The contribution of the computer science as well as the information and communication science approaches for the editorial function of the territorial information systems. Papers on territorial intelligence and governance, participative action-research and territorial development. Observatorio Local de Empleo, International Conference of Territorial Intelligence, Huelva, Spain, 2007, p. 24-27.

[Dent, 1999], Dent, B.D., Cartography Thematic Map Design, 5th Ed., New York: McGraw Hill, 1999.

[De Saussure, 1913]

[Dos Santo set al., 2004], Dos Santos S., Brodlie K., Gaining understanding of multivariate and multidimensional data through visualization. Computers & Graphics, 28 (3), 2004, p. 311-325.

[Fabrikant et al., 2005], Fabrikant S.I., Goldsberry K., Thematic Relevance and Perceptual Salience of Dynamic Geovisualization Displays. Proceedings, 22th ICA/ACI International Cartographic Conference, A Coruña, Spain, July 9-16, 2005.

[Fabrikant et al., 2004], Fabrikant S.I., Montello D.R., Ruocco M., Middleton R.S., The Distance-Similarity Metaphor in Network-Display Spatializations. Cartography and Geographic Information Science 31(4), 2004, p. 237-252.

[Fabrikant et al., 2006], Fabrikant S.I., Montello D.R., Mark D.M., The Distance-Similarity Metaphor in Region-Display Spatializations. IEEE Computer Graphics & Application, 2006, p. 34-44.

[Fabrikant et al., 2009], Fabrikant S. I., Garlandini S., Evaluating the Effectiveness and Efficiency of Visual Variables for Geographic Information Visualization, K. Stewart Hornsby et al. Eds., Springer-Verlag, Berlin-Heidelberg, COSIT 2009, LNCS 5756, 2009, p. 195-211.

[Fabrikant et al., 2008], Fabrikant S.I., Rebich-Hespanha S., Hegarty M., Cognitively Adequate and Perceptually Salient Graphic Displays for Efficient Spatial Inference Making. Annals of the Association of American Geographers, 100(1) 2010, p. 1-17.

[Fekete, Plaisant et al., 2007], Plaisant C., Fekete J., Grinstein G., Promoting Insight Based Evaluation of Visualizations: From Contest to Benchmark Repository, IEEE Transactions in Visualization and Computer Graphic, 2007.

[Fry, 2008], Fry B., Visualizing Data, Sebastopol: O'Reilly Media Inc., 2008.

[Gahegan M. et al., 2002], Gahegan M., Introducing GeoVISTA Studio: An Integrated Suite of Visualization and Computational Methods for Exploration and Knowledge Construction in Geography. Computers, Environment and Urban Systems, 2001, 26 (4), p. 267-292.

[Garland, 1994], Garland K., Mr Beck's Underground Map, Harrow Weald, Middlesex, UK, Capital Transport, ISBN 1854141686, 1994.

[Giovannelli, 2010], Giovannelli A., Goodman's Aesthetics, The Stanford Encyclopedia of Philosophy, ed. Edward N. Zalta, 2010.

[Girardot, 2009], Girardot J.-J., The CENTI scientific results and prospects, International Conference of Territorial Intelligence. Papers on tools and methods of Territorial Intelligence, MSHE, Besançon, 16-17 October, 2008.

[Guoping Qiu, 2007], Guoping Qiu, Advances in Visual Information Systems: 9th International Conference (VISUAL), IEEE, Inc. Staff, Visual Analytics Science and Technology (VAST), A Symposium of the IEEE, 2007.

[Goldberg, 1999], Goldberg J.H., Kotval X.P., Computer Interface Evaluation using Eye Movements: Methods and Constructs. International Journal of Industrial Ergonomics 24, 1999, p. 631–645.

[Goldstein, 1989], Goldstein E.B., Sensation & Perception, 2nd ed., Belmont: Wadsworth Publishing Company, CA, USA, 1989.

[Goodchild, 1988], Goodchild, Stepping Over The Line: Technological Constraints And the New Cartography, Cartography and Geographic Information Science, 1988. Vol. 15, n°3, pp. 311-319

[Goodman et al., 1968], Goodman, Nelson N., Languages of Art: an approach to a theory of symbols. Indianapolis: Bobbs-Merrill, 1968.

[Griffin, 2008], Griffin A.L., Visual Variables, ed. Kemp, K., Encyclopaedia of Geographic Information Science, SAGE Publication, Thousand Oaks, CA, USA, 2008, p. 506-509.

[Haber et al., 1990], Haber R. B., McNabb D. A., Visualization Idioms: A Conceptual Model for Scientific Visualization Systems. In Visualization in Scientific Computing, IEEE Computer Society Press, 1990, p. 74–93.

[Hall, 2006], Hall A. J. P., Else/Where Mapping: mapping new cartographies of networks and territories, Minneapolis, University of Minneapolis Design Institute, 2006.

[Henderson, 2004], Henderson J.M., Ferreira, F., Scene Perception for Psycholinguists, Henderson J.M., The interface of language, vision, and action: Eye movements and the visual world, New York: Psychology Press, , 2004, p. 1-58.

[Henderson, 2004], Henderson J.M., Ferreira F., The Integration of Language, Vision, and Action: Eye Movements and the Visual World, Psychology Press, New York, p. 105-134, 2004.

[Heer et al., 2010], Jeffrey Heer, Michael et al., A Tour Through the Visualization Zoo,

[Itten, 2003], Itten J., Art de la Couleur, Paris: Dessain et Tolra, 2003, ISBN 2040217886.

[Itti, 2001], Itti L., Koch C., Computational Modeling of Visual Attention, Nature Reviews Neuroscience 2(3), 2001, p. 194-203.

[Itti, 1998], Itti L., Koch C., Niebur E., A Model of Saliency-Based Visual Attention for Rapid Scene Analysis, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20 (11), 1998, p. 1254–1259.

[Ioannilli et al., 2002], Ioannilli M., Schiavoni U., *Fondamenti di sistemi informativi geografici*, *Texmat*, p. 342, 2002.

[Ioannilli, 2004], Ioannilli M., *New Information Requirements for Urban and Environmental Management: the Role of Geographic Databases*, *Proceedings of the 24th UDMS Conference*, Venice, Italy, 2004.

[James et al., 1967], James J., Thomas and Kristin A., *Illuminating the Path: The R&D Agenda for Visual Analytics* National Visualization and Analytics Center, 1967, p. 4.

[Jankun-kelly et al., 2007], Jankun-kelly et al., *A model and framework for visualization exploration*, *IEEE Transactions of visualization and computer graphics*, 2007.

[Keim et al., 2008], Keim D. A., Mansmann F., Schneidewind J., Thomas J., Ziegler H., *Visual analytics: Scope and challenges*, *Visual Data Mining*, *Lecture Notes in Computer Science* Vol. 4404, p 76-90 2008.

[Keim et al., 2008], Keim D., Andrienko G., Fekete J. D., et al., *Visual Analytics: Definition, Process, and Challenges*, 2008, *Lecture Notes in Computer Science* Vol. 4950, p. 154-175.

[Kerren et al., 2008], Kerren A., John T. Stasko, Jean-Daniel Fekete, and Chris North, *Information Visualization - Human-Centered Issues and Perspectives*, *Lecture Notes in Computer Science* 4950, p. 154-175.

[Kerren et al. 2012], Kerren A., Schreiber F., *Toward the Role of Interaction in Visual Analytics*, *Proceedings of the 2012 Winter Simulation Conference WSC 2012*, Berlin, Germany, IEEE Computer Society Press, 2012, p. 1-13.

[Kielman et al., 2009], Kielman J., Thomas J., *Foundations and Frontiers of Visual Analytics*, *Information Visualization*, Special Issue, Guest Eds., Volume 8, Number 4., 2009, p. 239-314, 2009.

[Kosara, 2008], Kosara R., *Visual Analytics*, *ITCS 4122/5122*, Fall 2007. Retrieved 28 June 2008.

[Kovalerchuk et al., 2004], Kovalerchuk B., Schwing J., *Visual and Spatial Analysis: Advances in Data Mining, Reasoning, and Problem Solving*, 2004.

[Krygier et al., 2005], Krygier J., Wood D., *Making Maps: A Visual Guide to Map Design for GIS*, New York: Guilford Press, 2005.

[Laurini, 1992], Laurini R., Thompson D., *Fundamentals of Spatial Information Systems*, London: Academic Press, p. 680, 1992.

[Laurini et al., 2008], Laurini R., Sebillio M. et al., *A chorem-based approach for visually synthesizing complex phenomena*, *Journal Information Visualization*, Vol. 7, Issue 3, Palgrave Macmillan, p. 253-264, June 2008.

- [Lima, 2011], Lima M., "Syntax of a New Language" in Visual Complexity
- [Liu, et al., 2008] Liu Z., Nersessian N., Stasko J., Distributed Cognition as a Theoretical Framework for Information Visualization, IEEE Transactions on Visualization and Computer Graphics, Vol. 14, n. 6, p. 1173-1180, Nov.-Dec. 2008, doi:10.1109/TVCG.2008.121.
- [Llyod, 1997], Lloyd R., Visual Search Processes Used in Map Reading. Cartographica 34 (1), p. 11-12, 1997.
- [Llyod et al., 2002], Lloyd R., Hodgson M.E., Visual Search for Land Use Objects in Aerial Photographs. Cartography and Geographic Information Science 29 (1), p. 3-15, 2002.
- [Lowe, 2003], Lowe R.K., Animation and Learning: Selective Processing of Information in Dynamic Graphics. Learning and Instruction (13), p. 157-176, 2003.
- [May Yuan et al., 2007], May Yuan, Hornsby K. and S, Computation and Visualization for Understanding Dynamics in Geographic Domains, Boca Raton : CRC Press 2007.
- [McCandless, 2012], McCandless D., Information is Beautiful, London: Collins, 2012.
- [McCormick, 1988], McCormick B. H., Visualization in scientific computing, ACM SIGBIO newsletter, Volume 10, Issue 1, March 1988, ACM New York, USA, 1988, p. 15 – 21.
- [MacEachren, 1994], MacEachren A.M., Some Truth with Maps: A Primer on Symbolization & Design. University Park: The Pennsylvania State University, ISBN 0892912146, 1994.
- [MacEachren, 1995], MacEachren, A.M., How Maps Work. Representation, Visualization and Design; New York, USA, Guilford Press, 1995.
- [MacEachren et al., 2001], MacEachren A.M., Kraak M.-J., Research Challenges in Geovisualization, Cartography and Geographic Information Science 28, 2001, p. 3 -12.
- [Mackinlay et al., 1986], Mackinlay, Jock D., Automated Design of Optimal Visualization", APT, 1986.
- [Mackinlay, 1986], Mackinlay, J.D., Automating the Design of Graphical Presentations of Relational Information. ACM Transactions on Graphics 5(2), 1986, p. 110-141.
- [Miller & Han, 2009], Miller H. J., Han J., Geographic Data Mining and Knowledge Discovery, Second Edition, Taylor & Francis Group, 2009.
- [Monmonier, 1991], Monmonier M., How to lie with maps, Chicago: The University of Chicago Press, , 1991.
- [Monmonier, 1993], Monmonier M., Mapping It Out. Chicago : University of Chicago Press, 1993, ISBN 0226534170.
- [Morrison, 1974], Morrison, J.L., A Theoretical Framework For Cartographic Generalization with the Emphasis on the Process of Symbolization. International Yearbook of Cartography 14, 1974, p.115-127.

[Muller, 1981], Muller, J.C.: Bertin's Theory of Graphics; A Challenge to North American Thematic Cartography. *Cartographica* 18(3), 1981, p. 1-8.

[Mullet et al., 1995], Mullet K., Sano D., *Designing Visual Interfaces: communication oriented techniques*, Englewood: Prentice-Hall, 1995.

[Murgante, 2008] Murgante B., *L'informazione geografica a supporto della pianificazione territoriale*, Milano: Franco Angeli, 2008.

[Palsky, 1999], very few of the proposed conventions have actually been tested systematically for their effectiveness and efficiency.

[Parkhurst et al., 2002], Parkhurst, D., Law, K., Niebur, E., Modeling the Role of Saliency in the Allocation of Overt Visual Attention. *Vision Research* 42, 2002, p.107-123.

[Petchenik, 1979], Petchenik B., From Place to Space: The Psychological Achievement of Thematic Mapping, *Cartography and Geographic Information Science* 6 (1), 1979, p. 5-12.

[Pollock & Wright, 1969],

[Raskin, 1994].

[Rensink et al., 1997], Rensink, R.A., O'Regan, J.K., Clark, J.J., To See or Not to See: The Need for Attention to Perceive Changes in Scenes. *Psychological Science* 8, 1997, p. 368-373.

[Rensink et al., 2002], Rensink, R.A., Change Detection. *Annual Review of Psychology* 53, 2002, p. 245-277.

[Rensink et al., 2005], Rensink, R.A., Change Blindness, Laurent Itti, Geraint Rees and John K. Tsotsos Eds., *Neurobiology of Attention*, San Diego: Elsevier, 2005, p. 76-81.

[Rosenholtz et al., 2007], Rosenholtz, R., Li, Y., Nakano, L., Measuring Visual Clutter. *Journal of Vision* 7(2), 2007, p. 1-22.

[Radach et al., 2003], Radach R., Jukka Hyona, Heiner Deubel, *The Mind's Eye: Cognitive and Applied Aspects of Eye Movement Research*, Elsevier, Amsterdam, NL, 2003, p. 573-605.

[Simons, 2000], Simons, D.J.: Current Approach to Change Blindness. *Visual Cognition* 7(1/2/3), 2000, p. 1-15.

[Slocum et al., 2008], Slocum, T.A., McMaster, R.B., Kessler, F.C., Howard, H.H.: *Thematic Cartography & Geographic Visualization*, 3rd ed., Upper Saddle River: Prentice Hall, 2008.

[Shneiderman, 1996], Shneiderman B., The eyes have it: A task by data type taxonomy of information visualizations, *Proc. IEEE Symposium on Visual Languages 1996*, IEEE, Los Alamos, California, September 1996, p. 336-343.

[Skupin., 2000], Skupin A.: From Metaphor to Method: Cartographic Perspectives on Information Visualization, *IEEE Symposium on Information Visualization*. Salt Lake City, Utah, USA, 2000.

[Stuart et al., 1999], Stuart K., Card J.D., Mackinlay, Shneiderman B, Readings in Information Visualization: Using Vision to Think, San Diego: Academic Press, 1999.

[Swienty et al., 2007], Swienty, O., Kurz, F., Reichenbacher, T.: Attention Guiding Visualization in Remote Sensing IIM Systems. Photogrammetrie, Fernerkundung, Geoinformation 4, 2007, p. 239-251.

[Tognazzini, 1992]

[Tominski, 2006] Christian Tominski, Event-Based Visualization for User-Centered Visual Analysis, PhD Thesis, Institute for Computer Science, Department of Computer Science and Electrical Engineering, University of Rostock, Germany, 2006.

[Tufte, 2005], Tufte, Edward R., Envisioning Information, Cheshire: Graphic Press LLC, 2005.

[Tufte, 2006], Tufte, Edward R., The Visual Display of Quantitative Information, Cheshire: Graphic Press LLC, 2006.

[Ware, 2000], Ware C., Information Visualization: Perception for Design, Morgan Kaufman, San Francisco, California, USA, 2000.

[Wilkinson, 1998], Wilkinson L., Grammar of Graphics, Berlin, Springer, 1998.

[Wolfe, et al. 2004], Wolfe, J.M., Horowitz, T.S., What Attributes Guide the Deployment of Visual Attention and how do they do it ? Nature Reviews Neuroscience 5(6), 2004, p. 1-7.

[Wright et al., 2008], Wright, R.D., Ward, L.M., Orienting of Attention, New York, Oxford University Press, 2008.

[Wong et al., 2004], Wong P.C., Thomas J., Visual Analytics, IEEE Computer Graphics and Applications, Volume 24, Issue 5, Sept.-Oct. 2004, p. 20-21.

[Yarbus, 1967], Yarbus, A.L., Eye Movements and Vision, New York: Plenum Press, 1967.

[Zhaoping et al., 2007], Zhaoping L., May K.A., Psychophysical Tests of the Hypothesis of a Bottom-up Saliency Map in Primary Visual Cortex, Public Library of Science, Computational Biology, 2007.

## Webliography

[www.flowingdata.com](http://www.flowingdata.com)

[www.informationisbeautiful.net](http://www.informationisbeautiful.net)

[www.nadiaamoroso.com](http://www.nadiaamoroso.com)

[www.visualizing.org](http://www.visualizing.org),

[www.visualcomplexity.com](http://www.visualcomplexity.com)

[www.visual-analytics.org](http://www.visual-analytics.org)

<http://nvac.pnl.gov>

<http://infovis.uni-konstanz.de>

<http://etlab.eng.uab.edu/research-highlights/visualization/>