Introduction to Information Visualisation

by

Riccardo Mazza

University of Lugano
Faculty of Communication Sciences

March 2004

This document was produced with LyX and LaTeX on GNU/Linux.
The versioning has been managed with CVS.

1 Introduction

There are several situations in the real world where we try to understand some phenomena, data, and events by graphics. Some aspects, such as when people need to find a route on a city map, the stock market trends during some period, the unemployment diffusion in Europe, etc. may be understood better using graphics rather than text (Ruddle et al., 2002). Let us consider for example a situation when a person needs to find a route to go from Lugano to Pisa. This information can be represented both in a pictorial way, such as a map, and in a textual way, such as the textual description of a route. See this example represented in Figure 1.

Some specific aspect that may be interesting when one is finding a route, such as the possibility for finding an alternative route, or the existence of historical places in the vicinity of the route, may be understood better from the graph rather then the textual format.

There are some specific aspects that may be understood better from graphics than from the textual format, such as the possibility for finding an alternative route, or the existence of historical places in the vicinity of the route. Graphics, if well constructed, may make the human cognitive process of constructing a mental image of the route easier (Ruddle et al., 2002).

1.1 Graphics for presentation

Graphics are a mean to display facts about the data in a way that others can see and understand the underlying structure and the hypothesis about the data (Rober, 2000). Tufte (1983, p. 13) writes, "*graphical excellence consists of complex ideas communicated with clarity, precision, and efficiency*". See for instance Figure 2. This map by Charles Joseph Minard portrays the losses suffered by Napoleon's army in the Russian campaign of 1812. Beginning at the Polish-Russian border, the thick band shows the size of the army at each position. The path of the Napoleon's retreat from Moscow in the bitterly cold winter is depicted by the dark lower band, which is tied to temperature and time scales. Edward Tufte, a highly reputed visual designer, statistician and academic, comments on this image " *It may well be the best statistical graphic ever drawn*" (Tufte, 1983, p. 40).

1.2 Graphics for explorative analysis

Graphics also are a mean for finding and identifying structures and properties in a given data set (Card et al., 1999; Tufte, 1983; Rober, 2000). The special properties of visual perception of data may facilitate the finding of relationships, trends, revealing hidden patterns, or as Bertin (1981, p. 16) says, "*it is the visual means of resolving logical problems*". To illustrate, Figure 3 represents a map of London's Soho district were an outbreak of cholera appeared in 1845. Black dots represent individual deaths from cholera, and **x** marks the position of the water pumps. This map allowed Dr. John Snow to observe that most of the deaths of the area were concentrated around the Broad

Figure 1: Driving directions from Lugano to Pisa provided both in graphical and textual format. Images from `http://www.viamichelin.com`

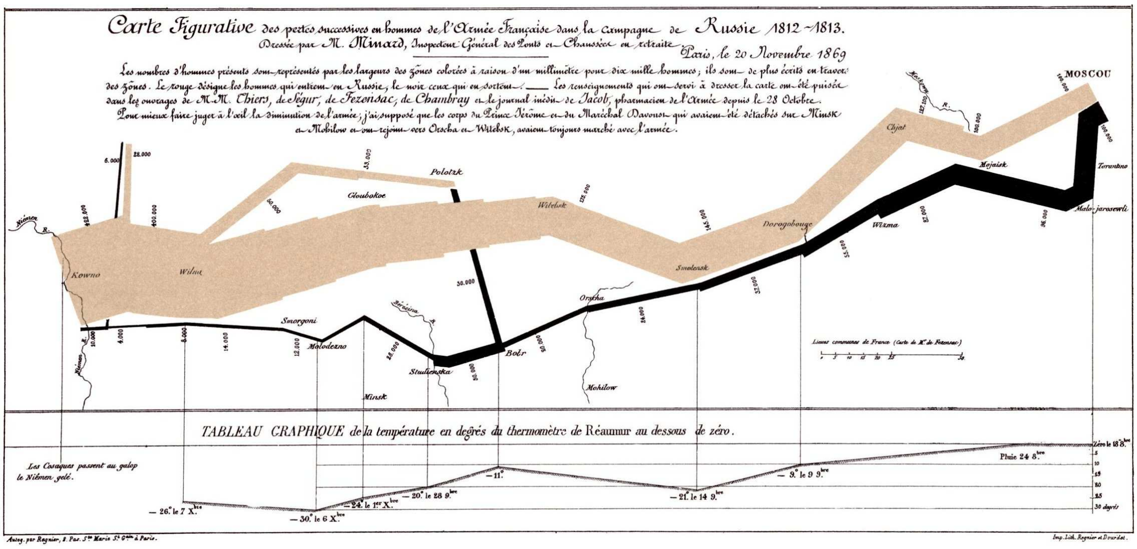

Figure 2: The Napoleon's army route in the Russian campaign of 1812 by Charles J. Minard, taken from Tufte (1983).

Street water pumps, which was discovered to be a cause of the diffusion of cholera in the area. (Tufte, 1983; Spence, 2001).

1.3 Graphics for confirmative analysis

Graphics are also the visual mean to confirm or reject some hypothesis about the data. For example, operators in the stock market exchange know that stock market indexes between different countries influence each other. This can be illustrated by the picture in Figure 4 representing the values of MibTel, the Italian stock market index and the US Down Jones market index within one year. It is easy to recognise that the increasing and decreasing values of both indexes come together. This relation, explicitly presented in the graphics, would be represented in the textual representation of the values by symbolic formulas which are less expressive and intuitive that a picture (Larkin and Simon, 1987).

2 Information Visualisation

Graphical representations are often referred by eminent authors with the term "*visualisation*" (or *visualization* in the more diffused American version of the term). For instance, Card et al. (1999, p. 6) define the term visualisation as "*the use of computer-supported, interactive, visual representations of data to amplify cognition*". It has been noted by Spence (2001) that there is a diversity of uses of the term "visualisation". For instance, in a dictionary the following definitions can be

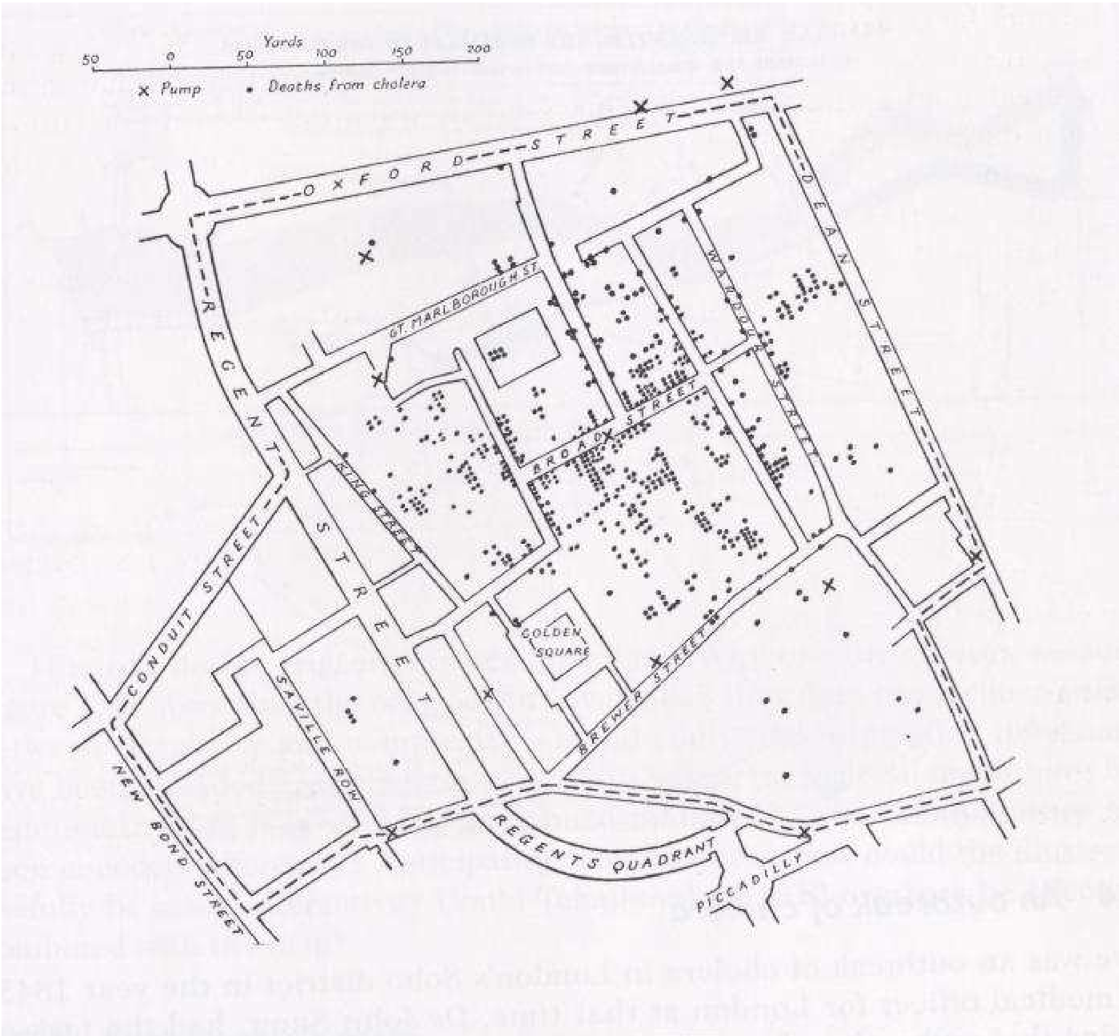

Figure 3: An 1845 map of London's Soho district plotted by Dr. John Snow showing deaths from Cholera and the location of water pumps. Dots represent cholera cases and X represent water pumps. Image from Tufte (1983).

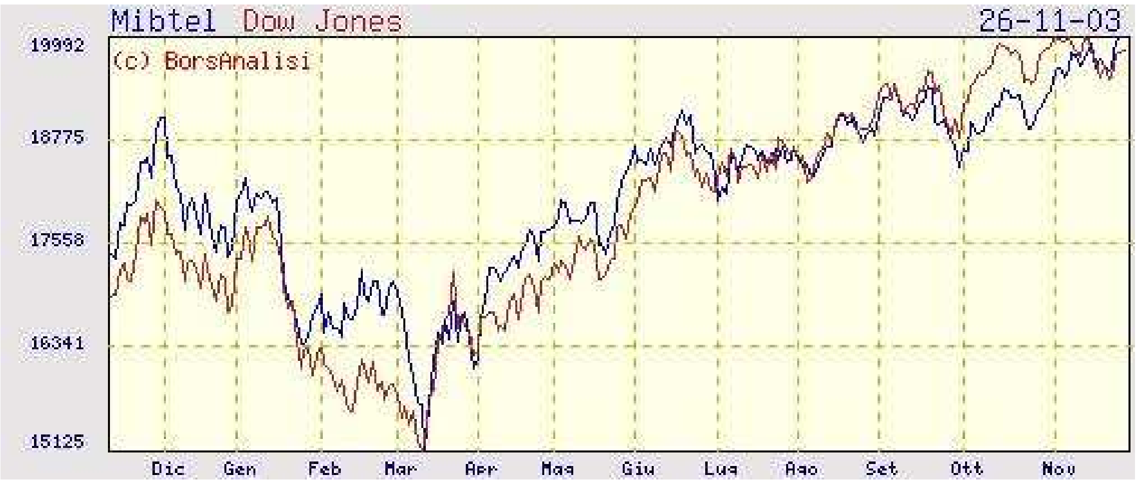

Figure 4: A comparison chart between the Italian stock market index, MibTel (in blue) and the United States' Down Jones market index (in red). Taken from `http://www.borsanalisi.com`

found:

> V*isualize: form a mental image of ..* [1].
>
> *Visualization: The display of data with the aim of maximizing comprehension rather than photographic realism*[2].
>
> *Visualization: the act or process of interpreting in visual terms or of putting into visible form*[3]

These definitions reveal that visualisation is an activity in which humans are engaged, as an internal construct of the mind (Spence, 2001; Ware, 2000). It is something that cannot be printed on a paper or displayed on a computer screen. With these considerations we can summarise that visualisation is a cognitive activity, facilitated by graphical external representations from which people construct internal mental representation of the world (Ruddle et al., 2002; Spence, 2001; Ware, 2000). Computers may facilitate the visualisation process with some visualisation tools. This is especially true in the latest years with the use of more and more powerful computers at low cost. However, the above definition is independent from computers: although computers can facilitate visualisation, it still remains an activity that happens in the mind. Some authors use the term "visualisation" to refer to both the printed graphical representation and the cognitive process

[1] *The Concise Oxford Dictionary. Ed. Judy Pearsall. Oxford University Press, 2001. Oxford Reference Online. Oxford University Press.*

[2] *A Dictionary of Computing. Oxford University Press, 1996. Oxford Reference Online. Oxford University Press.*

[3] *Merriam-Webster Online dictionary* `http://www.webster.com`

of understanding an image. In this thesis, we maintain the distinction between the creation of a pictorial representation of some data and the formation of an internal mental model of the data when interpreting the pictorial representation.

Information Visualisation is a relatively new discipline concerned with the creation of visual artefacts aimed at amplifying cognition. A number of new definitions have been produced to define the scope of this discipline. The Information Visualisation Research Group at the Institute for Software Research at University of California, Irvine cite on its Web pages:

> *Information visualization focuses on the development and empirical analysis of methods for presenting abstract information in visual form. The visual display of information allows people to become more easily aware of essential facts, to quickly see regularities and outliers in data, and therefore to develop a deeper understanding of data. Interactive visualization additionally takes advantage of people's ability to also identify interesting facts when the visual display changes, and allows them to manipulate the visualization or the underlying data to explore such changes*[4].

Similarly, the User Interface Research Group Web-site of the Palo Alto Research Centre (PARC-XEROX) defines that:

> *Information Visualization is the use of computer-supported interactive visual representations of abstract data to amplify cognition. Whereas scientific visualization usually starts with a natural physical representation, Information Visualization applies visual processing to abstract information. This area arises because of trends in technology and information scale. Technically, there has been great progress in high-performance, affordable computer graphics. At the same time, there has been a rapid expansion in on-line information, creating a need for computer-aid in finding and understanding them. Information Visualization is a form of external cognition, using resources in the world outside the mind to amplify what the mind can do*[5].

The above definition is appropriate nowadays with the computers being a constant part of our life, but as Spence (2001) and Hearst (2003) pointed out the role of computers is merely of a mean that facilitates visualisations. Hearst (2003) summarises that IV is:

> *The depiction of information using spatial or graphical representations, to facilitate comparison, pattern recognition, change detection, and other cognitive skills by making use of the visual system.*

[4] `http://www.isr.uci.edu/research-visualization.html`

[5] `http://www2.parc.com/istl/projects/uir/projects/ii.html`

IV has a long history having its origin from the historical works of J. H. Lambert [1728-1777] and William Playfair [1759-1823] who were the first to introduce graphics, in contrast with the tabular presentation of data and are considered the inventors of the modern graphics design (Tufte, 1983). Starting with Playfair, data were represented with methods of plotting. Other important contributions were made more recently by Jacques Bertin and Edward Tufte. Bertin, a French cartographer, was the first who tried to define, in 1967, a theory of IV by the identification of the basic elements of diagrams and described a framework for graphics design (Bertin, 1983). In 1983 Tufte published his theory of data graphics focused on the maximisation of the density of useful information in graphics (Tufte, 1983). Both Bertin and Tufte's theories have influenced the current development of IV.

2.1 Cognitive amplification

Graphics aid thinking and reasoning in several ways. For example, let us take a multiplication (a typical mental activity) e.g. 27×42 in our head, without having a pencil and paper. This will take usually at least five times longer than when using a pencil and paper (Card et al., 1999). The difficulty in doing this operation in the mind is holding the partial results of the multiplication in the memory until they can be used:

$$
\begin{array}{ccccc}
 & & 2 & 7 & \text{x} \\
 & & 4 & 2 & \\
\hline
 & & 5 & 4 & \\
1 & 0 & 8 & - & \\
\hline
\mathbf{1} & \mathbf{1} & \mathbf{3} & \mathbf{4} &
\end{array}
$$

This is an example which shows how visual and manipulative use of the external representations and processing amplifies cognitive performance. Graphics use the visual representations that help to amplify cognition. They convey information to our minds that allows us to search for patterns, recognise relationship between data and perform some inferences more easily. Card et al. (1999) propose six major ways in which visualisations can amplify cognition:

1. by increasing the memory and processing resources available to users;

2. by reducing the search for information;

3. by using visual representations to enhance the detection of patterns;

4. by enabling perceptual inference operations;

5. by using perceptual perception mechanisms for monitoring;

6. by encoding information in a manipulable medium.

IV is becoming an increasingly important discipline because the availability of powerful representations may facilitate the way we present and understand large complex dataset. Larkin and Simon (1987) argued in their seminal paper "*Why a diagram is (sometimes) worth ten thousand words*" that the effectiveness of graphical representations is due to their spatial clarity. Larkin and Simon compared the computational efficiency of diagrams and sentences in solving Physics problems, and concluded that diagrams helped in three basic ways:

1. **Locality** - is enabled by grouping together information that is used together. This avoids large amounts of search and allows different information closely located to be processed simultaneously. For example, Figure 4 puts together information about the history of two different stock market indexes and allows to process their evolution immediately.

2. **Minimising labelling** - is enabled by using location to group information about a single element, avoiding the need to match symbolic labels and leading to reducing the working memory load. For example, driving directions from Lugano to Pisa provided in graphical format in Figure 1 use visual entities such as lines depicted in red with a yellow stripe in the middle to denote a highway. Turning points (such as in Parma in the example) are clearly indicated by a crossing of the roads. Symbolic textual representations used in the textual format of the map are unnecessary because the connections are explicitly represented in the graphics.

3. **Perceptual enhancement** - is enabled by supporting a large number of perceptual inferences which are easy for humans to perform. For example, in Figure 3 the link between deaths from cholera and the location of a water pump responsible for the spread of cholera could be recognised immediately.

IV definitions introduce the term "abstract data", for which some clarification is needed. The data itself can have a wide variety of forms, but one can distinguish between data that have a physical correspondence and is closely related to mathematical structures and models (e.g. the airflow around the wing of an aeroplane, or the density of the Ozone layer surrounding earth) and data that is more abstract in nature (e.g. the stock market fluctuations, or the effects of temperature on the Napoleon's army movements in the Russian campaign). The former is known as *Scientific Visualisation*, and the latter as *Information Visualisation*. (Spence, 2001; Uther, 2001; Hermann et al., 2000). Scientific Visualisation was developed in response to the needs of scientists and engineers to view experimental or phenomenal data in graphical formats (examples are given in Figure 5), while Information Visualisation is dealing with unstructured data sets as a distinct flavour (Hermann et al., 2000). This thesis deals with Information Visualisation, and hence Scientific Visualisation is beyond the scope of this review.

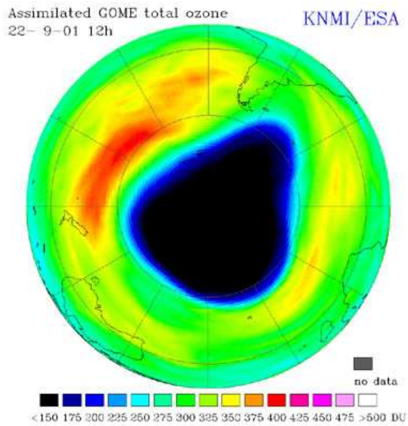

Figure 5: Two examples of scientific visualisations. On the left, representation of the laminar flows on the wing profile (image taken from `http://www.esa.int/export/esaMI/High_School/ESAF8BG18ZC_1.html`). On the right, Ozone hole over the South Pole during September 2001 (image taken from `http://www.esa.int/export/esaSA/ESALRRVTYWC_earth_1.html`)

A key question in IV is how we convert abstract data into a graphical representation, preserving the underlying meaning and, at the same time, providing new insight (Hearst, 2003). There is no “magic formula” that helps the researchers to build systematically a graphical representation starting from a raw set of data. It depends on the nature of the data, the type of information to be represented and its use, but more consistently, it depends on the creativity of the designer of the graphical representation. Some interesting ideas, even if innovative, have often failed in practice. Tufte (1983) and Bertin (1981) list a number of examples of graphics that distort the underlying data or communicate incorrect ideas. Tufte indicates some principles that should be followed to build effective well designed graphics. In particular, a graphic should:

- show the data;
- avoid distorting what the data have to say;
- present many data in a small space;
- make large data sets coherently;
- encourage inferential processes, such as comparing different pieces of data;
- give different perspectives on the data - from broad overview to the fine structure.

Graphics facilitate IV, but a number of issues must be considered (Shneiderman, 2002; Tufte, 1983; Spence, 2001):

Data Tables	Visual Structures	Views	Human interaction	Tasks	Level
Spatial (scientific) Geographic Documents Time Hierarchies Networks World Wide Web	Position Marks Proprieties: -Connection -Enclosure -Retinal -Time Axes: -Composition -Alighment -Folding -Recursion -Overloading	Brushing Zooming Overview+ detail Focus+context	Dynamic queries Direct Manipulation Magic lenses	Overview Zoom Filter Details-on-demand Browse Search Read fact Read comparison Read pattern Manipulate Create	Delete Reorder Cluster Class Promote Average Abstract Instantiate Extract Compose Organize

Table 1: Specific techniques for Information Visualisation identified by Card et al. (1999)

1. Data is nearly always multidimensional, while graphics represented on a computer screen or on a paper are presented in a 2D dimensional surface;

2. Sometimes we need to represent a huge dataset, while the number of data representable on a computer screen or on a paper is limited;

3. Data may vary during the time, while graphics are static;

4. Humans have remarkable abilities to select, manipulate and rearrange data, so the graphical representations should provide users with these features.

The above issues are considered in the IV field, and a number of methods and techniques have been proposed to meet these requirements. Card et al. (1999, p.33) give a comprehensive list of eight types of data, eleven visual structures, four views, three types of human interaction, eleven tasks and eleven levels that a user might want to accomplish with a visualisation tool (see Table 1). This is a multitude of possibilities. Next sections will describe some techniques that can be used to display pictorial representations and that could be (or have already been) considered to display students' data in educational environments.

2.2 Issues to consider in Information Visualisation

Before using one or more IV techniques we have to consider several issues (Spence, 2001; Card et al., 1999; Hearst, 2003; Reed and Heller, 1997):

1. **The problem.** This relates to what has to be presented, found, or demonstrated.

2. **The nature of the data.** Data types could be *numerical* (e.g. list of integers or reals), *ordinal* (non-numerical data having a conventional ordering, such as days of the week), and *categorical* (data with no order, such as names of persons or cities).

3. **Number of data dimensions.** Depending on the number of dimensions (also called *attributes* or *variables* (Spence, 2001)), representations are said to be handling *univariate* (one dimension), *bivariate* (two dimensions), *trivariate* (three dimensions), and *multivariate* (four or more dimensions) data. We perceive our world in three spatial dimensions, so it is easy to map and interpret up to three dimensions. However, handling more than three dimensions is very frequent in real world situations and represents one of the most challenging tasks in IV.

4. **Structure of the data**. This could be *linear* (data coded in plain data structures such as arrays, tables, alphabetical lists, sets, etc.), *temporal* (data which changes during the time), *spatial* or *geographic* (data which has a correspondence with something physical, e.g. maps, floor plans, 3D CAD; usually this is a subject of scientific visualisation and is not considered to be IV in the strict sense), *hierarchical* (data that naturally arise in taxonomies, the structures of organisations, disk space management, genealogies, etc.), *network* (data describing graph structures, i.e. nodes and links, nodes representing a data point, and a link representing a relationship between two nodes).

5. **Type of interaction**. Whether the resulting graphical representation is *static* (e.g. a print or a static image on a display screen), *transformable* (users can manipulate how the representation is rendered, such as zooming or filtering), or *manipulable* (users may control parameters during the process of image generation, i.e. restricting the view to certain data ranges)

Each one of the previous can suggest the use of one or more techniques, the most important (and used) ones are outlined next.

2.3 Techniques for representing univariate and bivariate linear data

Univariate and bivariate linear data are easily representable as a single image which relates data values with some scale in a plain. Conventional approaches are *scatterplots*, *charts*, and *histograms*. When used with bivariate data, scatterplots are helpful for encouraging the viewer to assess the possible causal relationship between the two plotted variables (Tufte, 1983). Figure 6 on the left represents a one-dimensional scatterplot illustrating the prices of some cars. The only attribute here is represented by the numeric value of a car, mapped onto the scale of the axes. This image could be helpful to easily locate the lowest, the highest values, and the general distribution of values and brunching on data (Spence, 2001). Figure 6 in the middle represents a

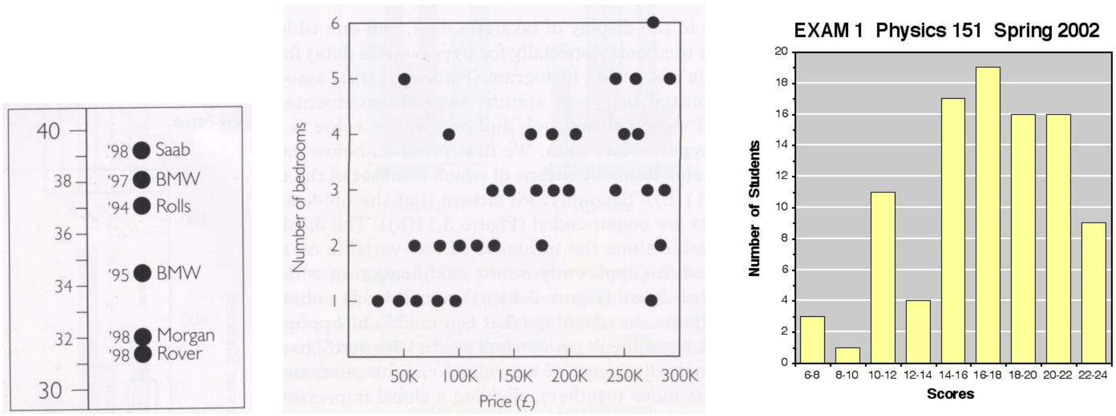

Figure 6: Samples of 1D (left), 2D (middle) scatterplots, and histogram . Images from Spence (2001) (left and middle images) and `http://www.physics.umass.edu/p151sec2s02/` (right image).

two-dimensional scatterplot which relates prices of houses with the number of bedrooms from a house-hunting archive. Users can clearly identify global trends (price of houses increases with the number of bedrooms), as well as items far from other having similar characteristics (e.g. a one-bedroom house selling for more than 250.000 GBP) (Spence, 2001). Another form of representation is the histogram (Figure 6, right). The example reports distribution of scores on a Physics exam. Histograms are suitable to facilitate comparisons between the values of a dimension, while scatterplots are helpful to convey overall impression of relationships between two variables (Hearst, 2003).

2.4 Techniques for representing trivariate linear data

Three-dimensional linear data is easily representable in a three dimensional space, as we are living in a three-dimensional world. However, the problem still arises because we basically can represent in a printed image or on a computer screen a two dimensional representation of a three-dimensional space, hence some approximations are necessary (Spence, 2001). Figure 7 left illustrates an example of a 3 dimensional scatterplot were each dimension is linearly mapped on an axis. The impassable barrier of the 2D image makes impossible to distinguish between the number of bedrooms of C and B. To overcome this problem some solutions have been proposed, such as to project items on the plans to identify the exact position (see Figure 7, centre) or to replace one dimension with a different shape of the items according to this value (see Figure 7, right).

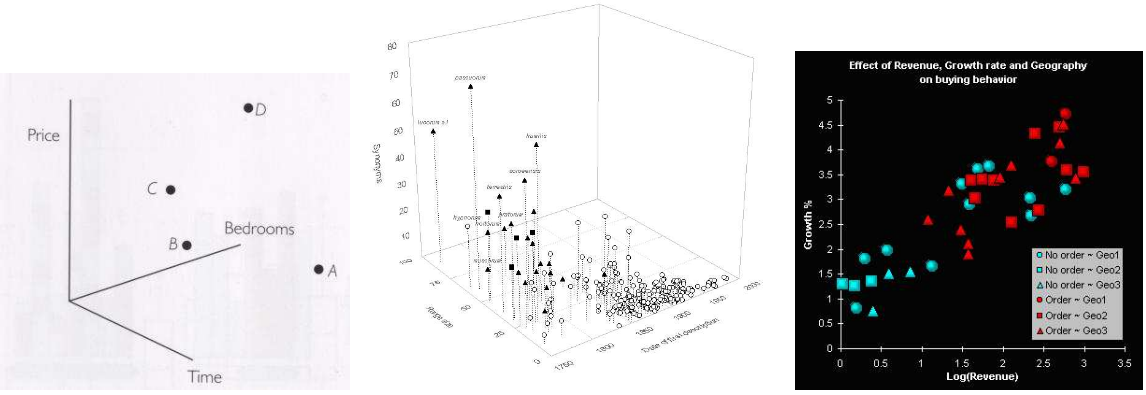

Figure 7: Some examples of 3D scatterplots. Images from Spence (2001) (left), `http://www.nhm.ac.uk/entomology/bombus/introduction.html`(centre), `http://www.kovcomp.co.uk/xlstat/tools/t01.html`(right).

2.5 Techniques for representing multivariate linear data

Very often real world cases cope with situations where relationships between more than three variables must be analysed. Let's think about an analysis of factors such as age, living place, job, and sex on the appearance of cancer on a sample of patients. This is a challenging aspect in IV, because some proprieties of images have to be explored to distinguish between several variables in a 2D drawing plane. For this purpose several methods have been proposed. Sachinopoulou (2001) suggested a classification into six groups, summarised in the following table:

Methods	**Description**	**Some known techniques**
Geometric	Transforming and projecting data in a geometric space.	Scatterplot matrix, Hyperslice, Prosection views, Surface and volume plots, Parallel coordinates, Textures and rasters.
Icon	Relies on a geometric figure (the icon) where the values of an attribute is associated with one features of this, such as the colour, a shape, the orientation.	Chernoff faces, Stick figure, Colour icon, Glyphs and Autoglyph.
Pixel	Use pixel as basic representation unit, and manipulate pixels to represent data.	Space fillings and Mosaic plots.
Hierarchical	Include trees and hierarchies and are useful when the data has some hierarchical or network structure.	Hierarchical axes, Dimension stacking, Threes, Worlds within worlds, Infocube.
Distorsion	Propose to distort the tree-dimensional space to allow more information to be visualised.	Perspective Wall, Pivot table and table lens, Fish eye view, Hyperbolic trees, Hyperbox.
Graph based	Represent data using nods and edges and is adopted when the large graphs should be represented.	Basic graph, Hyperbolic graph.

Techniques cited in the rightmost column of the table above are described in details in Sachinopoulou (2001); Card et al. (1999); Spence (2001); Chen (1999). These techniques are able to represent data in multidimensional space. A comprehensive description of multidimensional techniques is beyond the scope of this work, rather we focus on those techniques that have been used in CourseVis.

A number of additional techniques for multidimensional data representations exist. Differently from the techniques cited in the table above, they don't offer the possibility to establish relationship between each variable, but may represent several attributes allowing relationships on a subset of variables. Some of them are *composition*, *layering and separation*, *micro-macro readings*, and *small multiplies*.

Composition

The basic idea of the composition technique (Mackinlay, 1986; Card et al., 1999) is the orthogonal placement of axes that encode the same information, creating a 2D metric space of multidimensional data. An example of this technique is the diagram of the Napoleon's army route in Figure 2. This is an example of *single-axis composition* technique (Mackinlay, 1986) where attributes *army size*, *army longitude*, *army latitude*, and *temperature* have identical horizontal axis which is the *time*. Figure 8 illustrates how C. J. Minard composed these variables in the famous diagram.

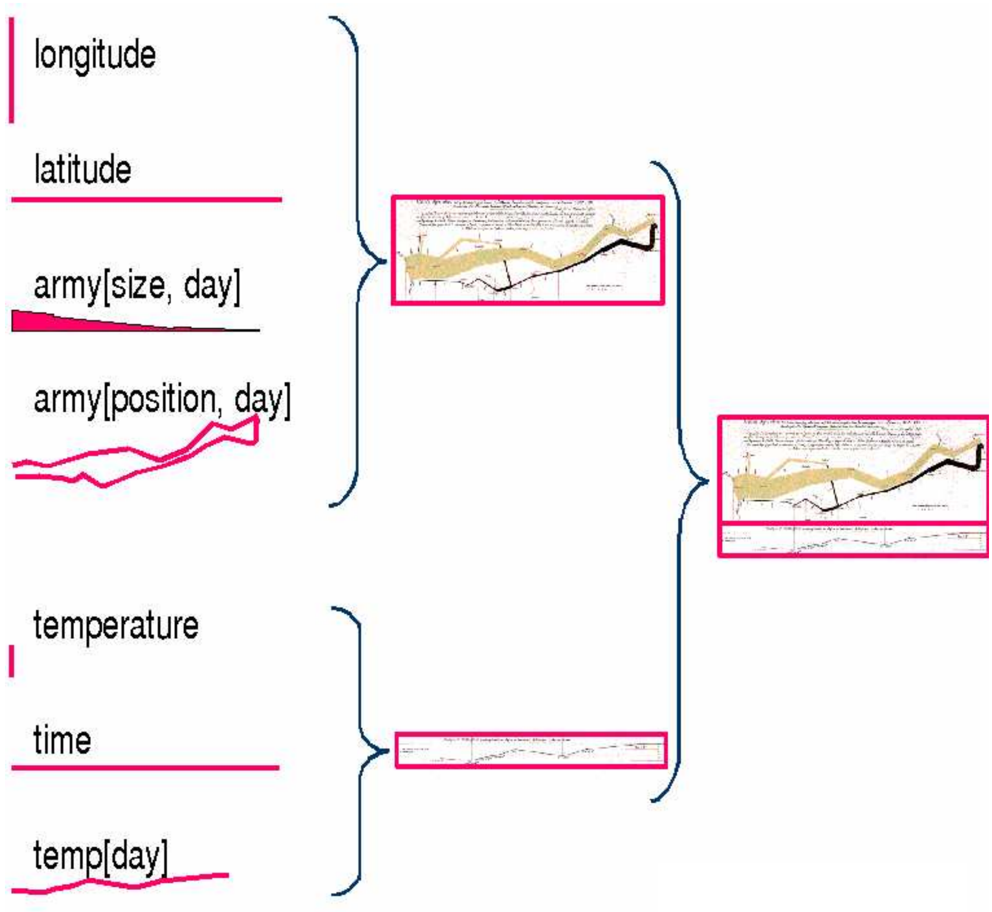

Figure 8: Single-axis composition technique used by C. J. Minard in the Napoleon's army route in the Russian campaign diagram. Image from Mackinlay (2000).

Layering and separation

Layering and separation is a technique illustrated, among others, by Tufte (1990) and concerns the visual differentiation of various aspects of the data. Tufte argues that "*confusion and clutter are failure of the design, not attributes of information ... the point is to find design strategies that reveal detail and complexity - rather than to fault the data for an excess of complication*" (Tufte, 1990, p. 53). He proposes layering and separation as one of the most powerful devices for reducing noise and enriching the content in graphics, and is achieved by distinction of colour, shape, size, addition of elements that direct the attention via visual signals, or ordering data to emphasise layer differences (Ruddle et al., 2002). An example of this technique is visible in Figure 9. It shows a city map of the centre of Florence. This picture illustrates at least 2 layers of reading: the location of historical places and the map of the streets to find a path. Colours have been used to distinguish between parks (in green), interesting places for tourists (brown), river (blue), and major streets (yellow). Shapes have been used to direct attention of tourists in historical places, and give an idea about what sort of building it is. This map can be used also by someone who is not interested in historical monuments in Florence but has to find the route to move from the *train station* to *Piazza Signoria*.

Micro-macro readings

Micro-macro reading "*is a method for presenting large quantities of data at high densities in a way that a broad overview of the data is given and yet immense amount of detail is provided*" (Ruddle et al., 2002). It encodes information at different levels of detail, such as the same image can be used to detect fine-grained level on information encoded (micro processing) as well as large-grained level of information (macro processing). An example of micro-macro reading is depicted in Figure 10. Here the picture allow the comparison of between-cycle variation (macro) and within-cycle variation (micro). Also shows an apparent growth trend in the wingspan of recent cycles (Tufte, 1990).

Small multiples

Small multiples technique consists of the same graphical design structure repeated several times (Tufte, 1983). It is used to compare at a glance series of graphics showing the same combination of variables while another variable changes. This method considers positioning similar graphical elements together in order to emphasise changes of the data (Ruddle et al., 2002). Figure 11 illustrates a small multiples design for representing the coverage of mobile phones in Switzerland for the tree main operators. Comparisons are possible to decide which operator utilises on a specific region for best signal reception.

Figure 9: A city map of the centre of Florence, streets and historical places are encoded with different shapes and colours. Image from `http://www.hotelitaliani.it/Mappa/Informazioni/PiantinaFirenze.htm`.

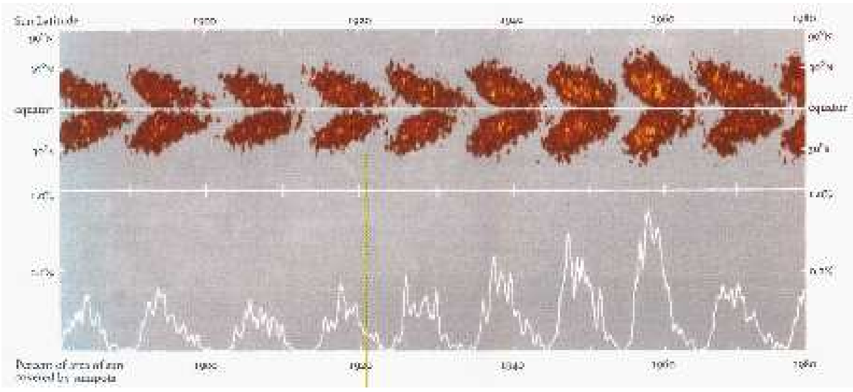

Figure 10: Diagram representing sunspots from 1880 to 1980 with the sine of the latitude marking a sunspot placement. Micro-macro readings allow combining patterns, details, average and variations with the same image. The lower time-series shows the total area of the sun's surface covered by sunspots by summing over all latitudes. Diagram by David H. Hathaway, Marshall Space Flight Centre, NASA, taken from Tufte (1990).

Figure 11: Small multiples example. The coverage for mobile telephone network in Switzerland between the tree operators: Swisscom (on the left), Orange (on the centre), and Sunrise (on right).

2.6 Techniques for representing spatial data

With the term spatial (or geographic) data we refer to representation techniques which have a correspondence with something physical. For example, data maps (Tufte, 1983) represent one or more attributes associated with areas of a map. See for instance the map in Figure 12 representing the cancer mortality in the USA at Lung, Trachea, Bronchus and Pleura for white males in years 1970 - 1994 in counties. Data maps can carry a huge volume of data in small space, but carry an intrinsic problem: they may give a wrong illusion that the importance of the information is geographic, rather than factual (Ruddle et al., 2002; Tufte, 1983). For instance, the figure doesn't allow ordering the counties according to the number of cancer deaths, or the number of people living in the county.

Space-times narrative design is another technique for representing spatial data. Its basic idea is to add spatial dimensions to the design of the graphics, so that the data are moving over the space as well as over the time (Tufte, 1983, 1990). The Napoleon's army route in Figure 2 is an excellent example. The graphics shows, among other things, how the crossing of the Berenzina river caused lots of deaths on Napoleon's troops, the latitude and longitude of movements, and the direction of the movements. It depicts the dynamics of the whole march with a static image.

2.7 Techniques for representing hierarchical and network data

There are many situations where data to be represented is linked in a graph structure, for example the organisation chart in a company, pages in a Web site, or DNA sequences. This is often referred as *network information visualisation*, and involves gaining insight into a structure that may consist of many data items (Reed and Heller, 1997). These are usually graphically represented with a drawing graph. Graphs are sometimes problematic, because a graph with few nodes is easy to draw and to comprehend visually, but real world situations often need to handle large data sets. See for example Figure 13, where two examples of network drawing are depicted. The image on the top is a representation of an intranet: the nodes are pages and links are URL connections

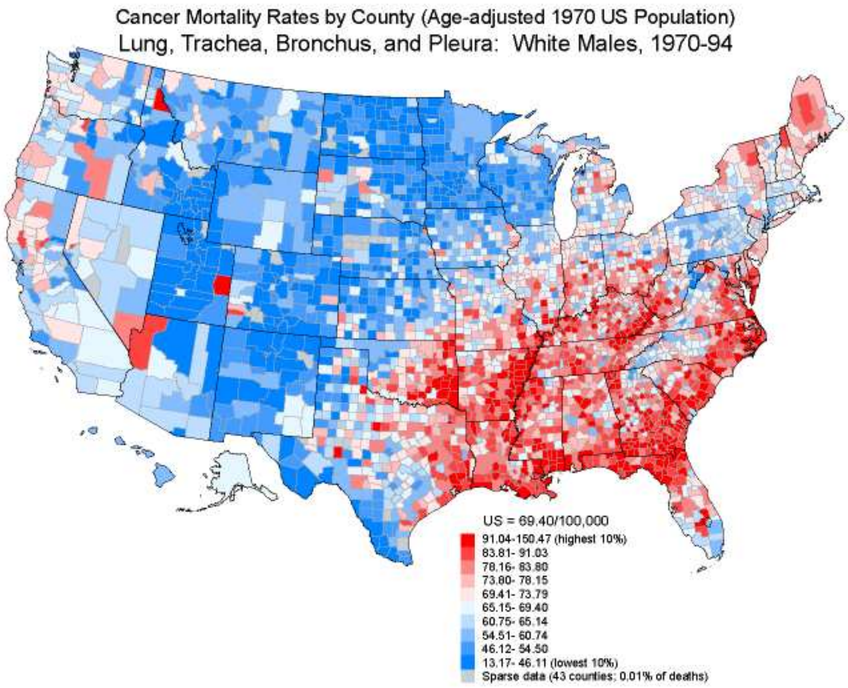

Figure 12: A data map example. Cancer mortality in the USA years 1970 - 1994 - white males by county. Image from USA National Cancer Institute `http://www3.cancer.gov/atlas/2.html`.

between pages. This network consists of 24 nodes easily represented in a 2D-picture. However, some situations have to deal with a large number of nodes. Figure 13 on the bottom represents an istance of a network with many nodes and links represented in a 3D picture. In this case, the limitation of the design pane requires rendering the network as interactive graphs. By engaging their visual image, a user is able to navigate through large networks, and to explore different ways of arranging the network components on the screen.

2.8 View transformations

A problem humans are experiencing in their everyday life is to have too many things placed in a limited space: books on shelves, addresses in agenda, windows on a computer screen, data to display in a Personal Digital Assistant (PDA). The information explosion phenomena of last years leads to the existence of more data than what can easily be displayed at once. "*Too much data, too little display area*" is a common problem in Information Visualisation (Spence, 2001). There are several techniques proposed to solve this problem, some of which are zooming, panning, scrolling, focus+context and magic lenses (Spence, 2001).

- **Zooming** is the increasing magnification of a decreasing (or increasing) fraction of a two-dimensional image.
- **Panning** is the smooth movement of a viewing frame over a two-dimensional image of greater size.
- **Scrolling** is the movement of data past a window able to contain only a part of it (such as we are doing with the scrolling of a long document in a word processing program).
- **Focus+context**'s basic idea is to illustrate at the same time the overall picture (the context) and to see details of immediate interests (the focus). This technique allows users to expand and contract selected sections of a large image, thereby displaying simultaneously the contents of individual sections of a document as well as its overall structure.
- **Magic lenses** follow the metaphor of reading a text by the means of a lens that enlarges the size of the text. In IV it can be used to place a lens upon the area of interest and receive more detailed information on the data amplified with the lens. For instance, magic lenses could be applied to Figure 9: an application could show this map to tourists and a lens placed over parts of the map could show details about the historical place selected.

An example of zooming, scrolling, and panning is illustrated in Figure 14. Focus+context technique is illustrated in an example in Figure 15. An example of magic lenses is illustrated in Figure 16

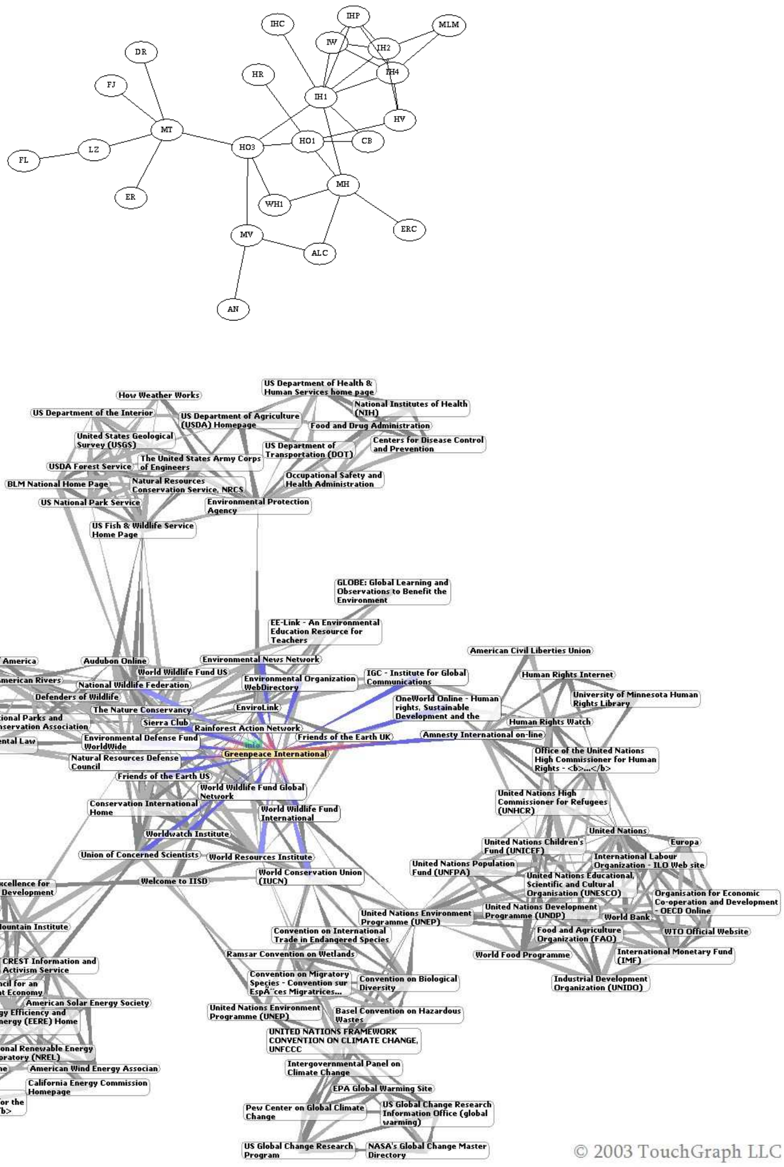

Figure 13: Two examples of network drawing. On the top, a graph of a small dimension intranet (image from Graphviz drawing software Web site: `http://www.research.att.com/sw/tools/graphviz/examples/`), on the bottom, a graph of some Internet Web sites (image from TouchGraph drawing software Web site: `http://www.touchgraph.com/bi.php?img=greenpeace_new.jpg`).

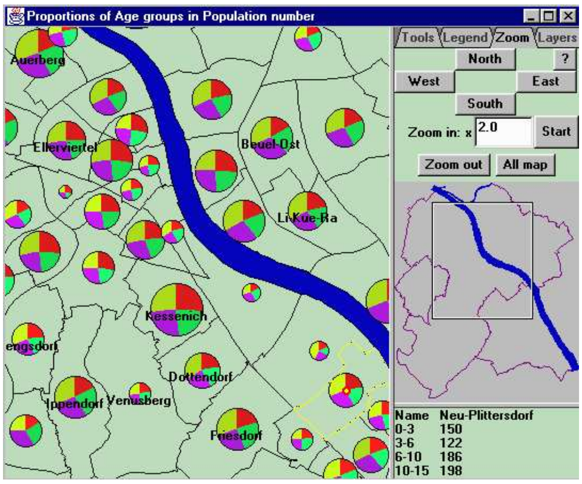

Figure 14: Zooming, scrolling, and panning operations from Descartes: a tool to support visual exploration of spatial data produced by Dialogis (`http://www.dialogis.com`). The middle of the right section of the window illustrates a schematic view of the entire territory. The frame shows the size and position of the territory fragment currently displayed in the map area relative to the whole territory. Using the buttons at the top right (north,south, east, west) users can scroll the territory shown in the map window.

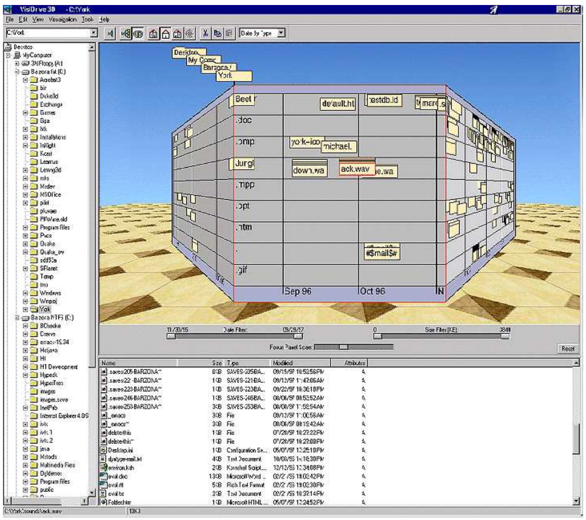

Figure 15: The perspective wall invented by Mackinlay et al. (1991). It is a 2D layout wrapped around a 3D structure, upon which label corresponds to files on the computer. Different labels denote different type of entities (folders, text files, wav, ...), located according to precise criteria (date of creation of the file in x-axis and file type on y-axis). Image from `http://www.inxight.com`.

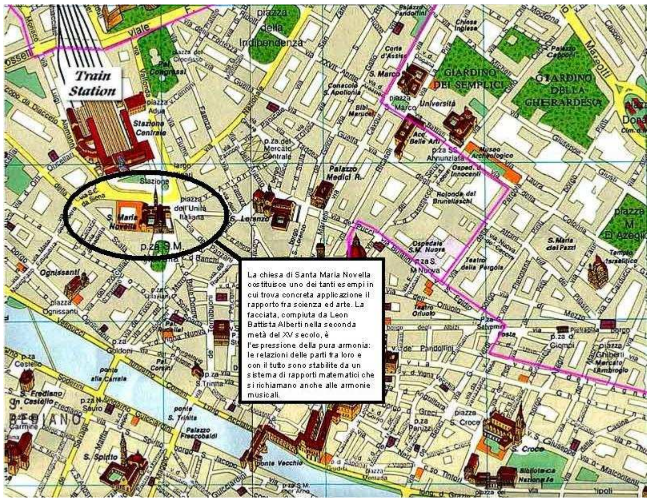

Figure 16: The magic lenses techniques applied to the historical places in a city map of the centre of Florence. Image from `http://www.hotelitaliani.it/Mappa/Informazioni/PiantinaFirenze.htm` and modified

References

Bertin, J. (1981). *Graphics and Graphic Information Processing*. Walter de Gruyter, Berlin.

Bertin, J. (1983). *Semiology of Graphics: Diagrams, Networks, Maps*. The University of Wisconsin Press. English Edition.

Card, K. S., Mackinlay, J. D., and Shneiderman, B. (1999). *Readings in Information Visualization, using vision to think*. Morgan Kaufmann, Cal. USA.

Chen, C. (1999). *Information Visualisation and Virtual Environments*. Springer.

Hearst, M. (2003). Information visualization: Principles, promise, and pragmatics. Handouts of the tutorial at CHI 2003 Conference on Human Factors in Computing Systems.

Hermann, I., Melancon, G., and Marshall, M. S. (2000). Graph visualisation and navigation in information visualisation. *IEEE Transaction on Visualization and Computer Graphics*, 1(6):24–43. `http://www.cwi.nl/InfoVisu/Survey/StarGraphVisuInInfoVis.html`.

Larkin, J. H. and Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. In Glasgow, J., Narayahan, H., and Chandrasekaram, B., editors, *Diagrammatic reasoning - cognitive and computational perspectives*, pages 69–109. California: AAAI PressThe MIT Press. 1995. Reprinted from Cognitive Science, 11, 65-100, 1987.

Mackinlay, J. D. (1986). Automating the design of graphical presentations of relational information. *ACM Transaction on graphics*, 5(2):110–141.

Mackinlay, J. D. (2000). Information visualisation. Course notes. University of Aarhus, Denmark. `http://www.daimi.au.dk/~mackinla/iv-course/index.html`.

Mackinlay, J. D., Robertson, G. G., and Card, S. K. (1991). The perspective wall: Detail and context smoothly integrated. In *ACM Conference on Human Factors in Computing Systems (CHI '91)*, pages 173–179.

Reed, M. and Heller, D. (1997). Olive, on-line library of information visualization environments. `http://otal.umd.edu/Olive`.

Rober, N. (2000). Multidimensional analysis and visualization software for dynamic spect. Master's thesis, Otto-von-Guericke-Universitat, Magdeburg.

Ruddle, R., Brodlie, K., and Dimitrova, V. (2002). Communication, visualisation and interaction. University of Leeds, School of Computing. Handbook of the course.

Sachinopoulou, A. (2001). Multidimensional visualization. Technical report, Espoo 2001. Technical Research Centre of Finland, VTT Tiedotteita, Meddelanden, Research Notes 2114. `http://www.inf.vtt.fi/pdf/tiedotteet/2001/T2114.pdf`.

Shneiderman, B. (2002). Tutorial on information visualization. INFORMS Roundtable. November 2002, San Jose, CA.

Spence, R. (2001). *Information Visualisation*. Addison-Wesley.

Tufte, E. R. (1983). *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, Connecticut, USA.

Tufte, E. R. (1990). *Envisioning Information*. Graphics Press, Cheshire, Connecticut, USA.

Uther, J. (2001). *On the visualisation of large user models in web based systems*. PhD thesis, School of Information Technologies, University of Sydney.

Ware, C. (2000). *Information Visualization, perception for design*. Morgan Kaufmann Publishers, San Francisco, USA.