

Four Applications for Data Visualization

Andrea Valente
Via Fratelli Carle,9 Torino (IT),
Laurea triennale in Ingegneria
dell'organizzazione d'impresa
Tel.3334156494-
val.andrea@hotmail.it-
s188153@studenti.polito.it

ABSTRACT

I wrote this work for the final thesis of my bachelor at Polytechnic of Turin in association with CSI Piemonte, where I have attended my stage period, during the period between June and September 2011. I want show some works that I ended during this one. I used different data visualization technologies such as Gapminder from Google Data Explorer and tools like ManyEyes, Open eXplorer and Fineo. Processed data come from the CSI Piemonte. The objective of this work is to show to the CSI Piemonte new data visualization tools, with the intent to create new prospects for the business reality in the field of treatment and exploitation of digital data.

Categories and Subject Descriptors

Statistical Sciences

Keywords

data, visualization, charts, tools, information.

1. INTRODUCTION: AIMS AND BACKGROUND

CSI Piemonte manages for its clients 1300 alphanumeric databases, 1400 spatial databases and 160 databases for decision support to Piedmont PA (public administration). The intelligent use of public property information becomes an important player for the business, for the research, for the democracy and for government. One of the requirements that have emerged in recent years, not only within the PA, is to make statistical information more understandable and accessible.

“We discover the world through our eyes” [1] says Stephan Few, one of the most important manager of visual communication. According Few forms of communication such as the graphics are still largely primitive. The charts were invented to bring to light the significant data, data that would be impossible to interpret with a simple table. But in recent time founded the discipline of Data Visualization, this takes on increasing importance in the field of Business Intelligence (BI). Data Visualization has become subject of study and research also in the academic world.

Few helps us to define the most common visualization techniques: dashboards, geo-spatial visualization and dynamic visualization.

2. DESCRIPTION

Now i want to describe the visualization tools that i have used for statistical representation.

Google public data Explorer

(<http://www.google.com/publicdata/home>)

It's a tool that facilitates the exploration, visualization and communication of large data sets of public interest. Useful to create views of public data, add links to connect to them or incorporate the data into their web pages. It currently has four display types: line chart, bar graph animated, animated map, animated bubble chart.

ManyEyes (<http://www-958.ibm.com/software/data/cognos/manyeyes/>)

It's a social network visualization designed by IBM. Allows you to create and share charts and infographics. Once you choose the data set to be transformed into visual form, you can upload and select the output type chosen. The potential load data and the immediate implementation of very high quality graphics to offer users a wide range of representation of structured data as well as unstructured information. With a few clicks you can get very pleasant bubble charts, GIS, cloud charts, bar charts and pie charts, with the ability to perform drill-down operations (segmentation and creation of hypercubes) for displaying the detail information. More than this, is possible get (and download) data sources used to easily build graphs, and use these settings directly on Many Eyes platform to build other types of views based on them. To view the different graphics is necessary to install the Java plug-in.

Open eXplorer (<http://www.ncomva.com/>)

It joins the geospatial component to the dynamic display. The main interface, which the user interacts, is divided into 4 boxes: the thematic map, scatter plot, histogram (or parallel coordinate graph) and the area of path analysis (the content browser) inside of which are written the story-telling. Another very interesting is the presence of a bar of "play", if selected, it starts time component, it is present below four distinct areas.

Fineo (<http://www.densitydesign.org/>)

It is web application that uses the patterns of Sankey diagrams, charts that represent the continuous data streams such as money, energy or materials in a system, in order to represent the relationships between the size of categorical data. Often breaking the data into categories can be crucial to interpret real-world phenomena, and this innovative app manages to make sense of a set of multidimensional data through an interactive approach. This application has been created by Density Design, a research laboratory of the Department of Design at the Polytechnic of Milan. The research objective is to exploit the potential of information visualization through the construction of various design-computer app.

3. APPLICATION AND RESULTS

The data used were provided by CSI Piemonte, concerning demographic information about Turin and the cities around it, and data concerning the use of ICT in the firms of Piedmont.

ManyEyes (Matrix Chart)

The matrix chart divides the screen into a grid. Rows represent the values in one text column (e.g., education level) and columns represent another text column (e.g., index of knowledge of English). Each cell then shows a circle or bar that represents the value for its row/column combination. The two modes of the matrix, circles and bars, are useful in different situations. Bars, whose height represents numeric values, are better for exact comparisons and allow space for more columns. Circles, which show values via area, are good for showing non-negative values that vary greatly, and allow space for more rows. You can flip between the two modes in the "Expert Options" menu to the left of the visualizations. You can display a third dimension of data using color. When you select a column to be represented by color, circles will turn into miniature pie charts and bars will be broken into differently-colored pieces. This option should be used with caution, since it's possible to create a cluttered display, but it can often be effective. A matrix chart takes a table with at least two text columns, for the x- and y-categories. If there is a third text column it can be used to determine colors. If there are no numeric columns, then the bubbles or bars simply show the count of each category combination.

Google Data Public Explorer

With the work done with the Data Explorer application I got three views: a line chart, an animated bar chart and an animated bubble chart. To upload my data was necessary to create a dataset DSPL (Dataset Publishing Language). It is a .ZIP file that contains a set of XML files and CSV files. The XML file contains metadata, the information about the data used, and CSV files are simple tables that contain data in the dataset that I wanted to propose. For the implementation of a database with Data Explorer is therefore required knowledge of XML (Google guarantees a tutorial). The interface with which the user interacts, once implemented the data, offers the possibility to observe the indicators that can be chosen for analysis. Clicking on a marker will be possible to proceed with the display. Loaded this page the user can interact with charts available choosing the menu at the top left. The line chart shows metric values over time in a traditional, line chart format. The animated bar chart shows a bar for a single metric at each point in time. The bar colors can be linked to a different metric. As time advances, the bars show the new values in an animated fashion. The animated bubble chart plots two metrics against each other, one on the x-axis, and one on the y-axis. Additional metrics can be linked to the bubble sizes and colors. As time advances, the bubble positions, sizes, and styles change according to their new values in an animated fashion.

Fineo

The flow chart that this application allows you to create, in addition to being visually effective, demonstrates a simple implementation of good data. To change the collection of information desired, the user needs to upload a file. TSV, an extension that Excel can create. In my case I wanted to show, for the two indicators 'Using ICT' and 'importance levels', information relating to ICT questionnaire given me by CSI Piemonte. The graph allows you to analyze the various branches of business, how important it is for employees to use ICT. On the left are the areas

where employees work ordered by decreasing values (from top to bottom), bringing the mouse pointer over one of these variables (eg 'Personnel Management') we note that the streams converge on the variables of the other indicator on the right (in this case 'significance levels'). Conversely, selecting one of the variables on the right side (eg 'Low Importance'), will perform an analysis regarding the level of importance for the various sectors. The range of variables is calculated according to the percentage present in the dataset. Although the magnitude of the flow is proportional to the value assigned.

Open eXplorer

The display of thematic map shows, in the upper left navigation tools, analysis tools and the legend of the chromatic scale used. The navigation tools allow you to intervene on location and zoom of the map. The Scatter Plot is a graph that allows the simultaneous display of four sizes for each country represented: the first one combined with a map (color of the bubble), the other two are represented on the axis of scatter plot (X axis, Y axis) and the last one is related to the width of the bubble. The parallel coordinate (PCP) graph allows you to simultaneously view with lines and histograms, different sizes. The PCP is a geo-visualization technique used to identify trends and data sets (clusters) to support in-depth analysis on the relationship between territories and indicators. With this tool is possible to represent also a dynamic time graph. The data file to be imported can be done using Excel and must be well designed with respect to the position of the values and required metadata. The Excel file containing the spreadsheet with the data must be saved in Unicode format .TXT.

4. CONCLUDING REMARKS

The study is part of a program of research and development compared to the intelligent use of data. As for my case, I used four application tools that provide new graphic techniques than those currently existing in the CSI Piemonte, two of which (ManyEyes and Open eXplorer) introduce the concept of story-telling.

5. REFERENCES

- [1] Visual communication., *Stephen Few, Cognos Innovation Center; 09-2006.*

