ePSIplatform

European Public Sector Information Platform

Topic Report No. 2012 / 05

Data Visualisation

Authors: Daniel Dietrich and Henri Myrttinen
Published: May 2012

1 Executive Summary

With the development of new ICT tools and methods as well as the exponential growth in the number of people accessing data through the Internet or smartphone apps, data visualisation has become a major sector of growth and development. While traditionally datasets were relatively difficult to combine and visual representation tools limited (e.g. maps, charts or graphs), the new advances allow for virtually unlimited combinations of datasets and creative visualisations. With new tools, many of which are open source, data re-users can quickly and easily trawl, gather, combine, analyse and visually represent data sets in creative ways. Re-use has already led to a range of new, user-friendly and innovative ways of presenting PSI. This has not only had direct benefits for users, customers and citizens but has also had direct benefits for PSBs who have been able to glean new insights through the combination and visualisations of datasets in new ways. Visualisation is one of the most dynamic fields of data re-use and not all forms of visualisation which are born through experimentation can be expected to have added value. The more open the re-use of PSI is, though, the more spaces are opened up for this experimentation and the greater the probability of new innovations.

2 Introduction

Data Visualisation is, in brief, the graphic representation of datasets in a way which ideally simplifies or enhances their understanding. While traditionally this has been mainly restricted to limited numbers of data sets and relatively simple forms of visualisation (e.g. graphs, charts or maps), the rapid development of software programmes now allow for a wide range of innovative and interactive visualisations which can combine multimedia with more static representations. These new programmes tend to be far easier and cheaper to use (and in some cases are freely downloadable) than previous models. At the same time, open access to datasets has increased, flanked by a marked spread of increasingly visual, mobile, affordable and powerful ways of accessing information (e.g. smartphones, tablets or netbooks). In combination, these three trends have revolutionised the ways in which data can and is analysed, combined and represented visually.

The main questions that this Topic Report seeks to answer are:

- what is the added value of PSI re-use in Data Visualisation;
- what are the potentials and
- what are the limitations of Data Visualisation?

In addition, the report will showcase some innovative examples of data visualisation.

3 Added value of PSI re-use in Data Visualisation

The greatest potential benefit of PSI re-use in terms of data visualisation is that the re-users are able to find new, unconventional ways of combining, analysing and presenting the data graphically. Datasets from different sources can be combined to develop applications which

allow different users – be they citizens, corporations, or PSBs themselves – to comprehend the information in new, easily comprehendible and often interactive ways. This can improve PSB or corporate service provision, open new insights for addressing societal challenges, assist in the development of new economic possibilities or lead to increased citizen participation.

Increased access to PSI combined with the availability of new visualisation tools significantly lowers the threshold for interested parties to trawl, combine, analyse and visually represent different datasets. The greater the room for experimentation is, the greater, at least in theory, the potential will be to develop innovative and useful data visualisation applications. Data visualisation is however not an end in itself and both relies on and leads towards certain interpretations of data.

To use a hypothetical example for data visualisation, numerous studies have established a causal link between poverty, poor school performance and environmental pollution, more precisely lead poisoning has been established in the USA[1]. Combining the four datasets of socio-economic indicators, residential districts, school grades and use of paints containing lead, the researchers were able to establish this link. Data combination, analysis and visualisation of these could arguably have speeded up this process, has the "hot spots" of poverty, poor grades and use of lead paints overlap on city maps.

The example, however, also shows the potential limitations of data visualisation; or more precisely, that visualisation needs to be based on proper data analysis. Returning to the hypothetical example, this important public health insight was not a guaranteed outcome and a different reading of the data could have been used to "prove" for example that poorer people are less intelligent and/or unscrupulous users of environmentally hazardous products – which could then also have been represented visually as "hot spots".

4 Potentials

The potential markets for PSI re-use and data visualisation are vast and growing daily as visual representations gain in popularity over other forms of data representation. Some of the main potential users for visual PSI data re-use are journalists and media representatives, NGOs and other civil society actors, citizens and consumers, research institutions, PSBs themselves as well as commercial companies.

Researchers, media and NGOs – but also PSBs – can use data visualisation to analyse and explain complex causal relations. Open access to PSI also allows citizens to combine datasets and visualise their concerns vis-à-vis decision-makers and PSBs, e.g. by visually

---

[1] More precisely, the case involved the use of lead in cheap paints in social housing projects, which severely affected the mental development of infants and young children exposed to the paints in the housing projects. It has however to be noted that the phenomenon of poor educational performance in socio-economically marginalised neighbourhoods is in no way a simple monocausal phenomenon and can not only be attributed only to the use lead paints .

representing persistent delays in service delivery. Visually presented PSI data can also activate citizen participation in political and decision-making processes. A highly successful, early example of data visualisation from the grassroots up to activate local, national and international responses to a crisis was the documentation, visualisation and mapping of human rights abuses in Kenya following the 2007-2008 elections using ushahidi.com[2], a name which means "testimony" in Swahili.

5 Drawbacks

While data visualisation has immense potential positive, innovative and creative solution-finding, it can not be assumed that all creative re-arrangements and visualisations will lead to added value for the users and re-users of PSI data. As Vitaly Friedman cautions,

> *"main goal of data visualization is to communicate information clearly and effectively through graphical means. It doesn't mean that data visualization needs to look boring to be functional or extremely sophisticated to look beautiful. To convey ideas effectively, both aesthetic form and functionality need to go hand in hand, providing insights into a rather sparse and complex data set by communicating its key-aspects in a more intuitive way. Yet designers often fail to achieve a balance between form and function, creating gorgeous data visualizations which fail to serve their main purpose — to communicate information".*

As noted earlier with the example of correlating poverty, poor educational performance and environmental pollution, the visual representation of PSI data such as statistics can easily be used to seemingly "prove" causal links which may not exist or which are only one strand of a complex relationship. As with any other form of data re-use, caution needs to be taken to avoid excessive simplifications, and as visualisations by definition aim towards a simplified presentation of complex datasets and their relations, extra care is needed.

6 Examples for new Technologies and Methods

Data visualisation has been made increasingly accessible through the development of new software and there is a growing online community sharing and discussing ways of better visualising data. Below some of the key tools and sites discussing data visualisation are showcased.

A growing range of tools, including free open source as well as proprietary and commercial ones, are available for data visualisation. These include the following:

- Infogr.am - A online programme platform for data visualisation

---

[2] http://ushahidi.com/about-us

- Gapminder is an award-winning Swedish-based non-profit venture which aims at promoting the use of data visualisation for social ends, such as improving access to and quality of education, combating the spread of HIV/Aids or attaining the UN Millennium Development Goals (MDGs). The site provides powerful open source desktop software for making semi-professional interactive animated statistics.
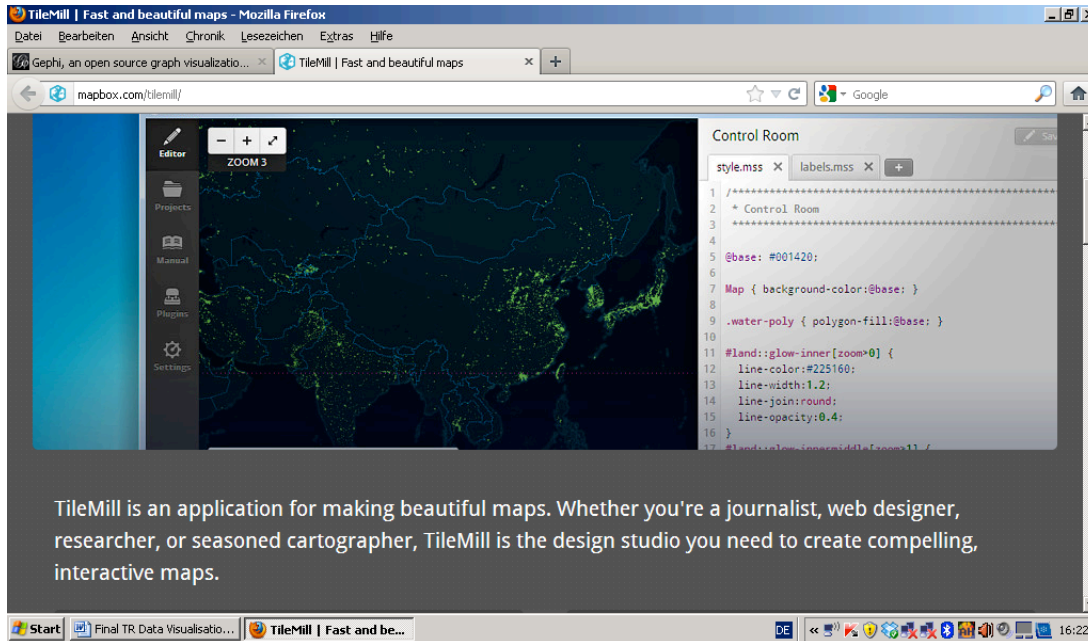

- Many Eyes is a web service for uploading data and then creating graphic representations which are then viewed and commented upon by other users of the site.

- Tableau Public is commercial desktop software for creating and sharing data visualisations on the web.

- VIDI is an online platform, which provides free downloadable software tools for making charts, maps and diagrams for websites.

- Ushahidi is a free, open source tool to collect, visualise and share information, e.g. through interactive mapping. This has been used in the past to update, track, document and communicate human rights abuses as the 2007-2008 Kenyan Election crisis unfolded.

- TileMill is an open source desktop tool for visualising data by creating interactive maps, using geodata and other datasets.

- OpenHeatMap is a web-based service for easy and quick creation of heat-maps from any data source in CSV or spreadsheet format. These can be uploaded online and are converted into heat maps in a short period of time.

- OpenLayers provides open source tools that facilitate the making of dynamic maps for web pages. It can display map tiles and markers loaded from any source.

- Gephi is open source desktop software for all kinds of networks and complex systems, dynamic and hierarchical graphs.
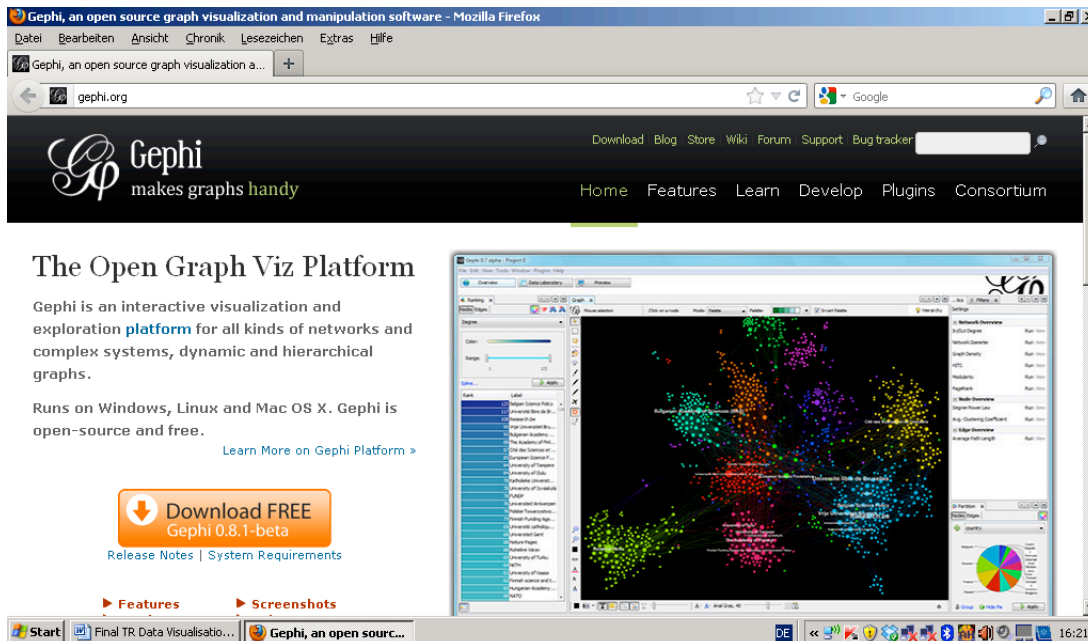
- OpenSpending is a web service to analyse and visualise budget or spending data.

Websites showcasing and discussing data visualisation

The following websites and blogs are some of the central discussion forums that provide platforms for introducing and debating new data visualisation innovations:

- Drawing By Numbers is a website is directed mainly towards the use of open source tools for visualising data for the purpose of civil society activism and advocacy. It provides how-to manuals, discussion forums and tools for opening, analysing and visualising data.

- vis4.net blog - The blog of the influential data visualisation expert Gregor Aisch discusses different forms and tools for data visualisation, with a focus on mapping

- Information is Beautiful is the blog of data journalist and information designer David McCandless. In addition to discussing data visualisation methods, the site provides some downloadable examples of data visualisations

- Flowing Data is a website that provides a discussion forum on data visualisation tools and methods as well as tutorials for users

- Information aesthetics is a further online discussion forum analysing different data visualisation tools and methods

- Visualizing.org is a discussion and information sharing site in which members of the "data visualisation community" can upload their visualisations and discuss these with other members, The site also organises data visualisation marathons.

Practical examples

Data visualisation has been put to use in practically all fields, including commercial and non-commercial uses, research, marketing, campaigning, advocacy or public service information.

Two examples showcased here with a particular significance for PSI re-use are journalism and data visualisation of PSI.

Journalism

Two websites which promote the use of data visualisation in journalism are:

- http://datadrivenjournalism.net/ and
- http://datajournalismhandbook.org/

Several leading global newspapers, such as the New York Times and the Guardian have been pioneering data-driven-journalism and have been concentrating on developing data visualisation tools for this purpose. The Guardian was awarded the first prize for "data visualisation and storytelling" in the 2012 Data Journalism Awards for its visualisation of the spread of rumours via Twitter during the 2011 riots in the United Kingdom.

Visualisation of PSI

Complex PSI data can be visualised for the purpose of public information or campaigning. Two examples of this are

- the visualisation of European states' energy use, imports, CO2 emissions and other energy-related open PSI information at http://energy.publicdata.eu/ee/vis.html
- and the visualisation of MDG-relevant data by the Swedish organisation Gapminder (see also above).
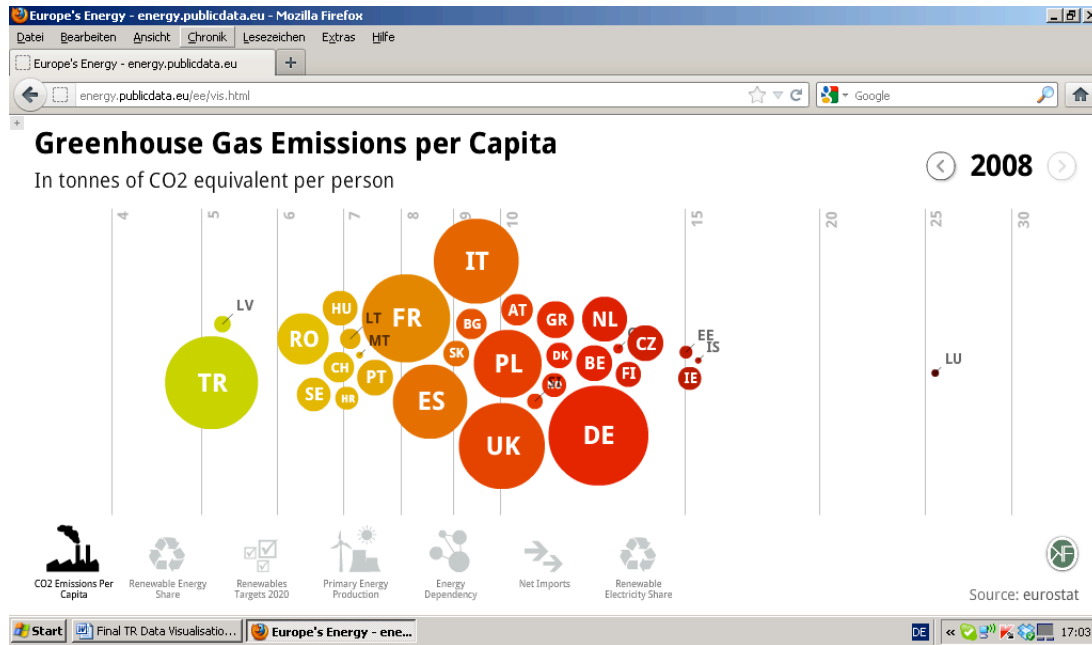
These visualisations of open PSI data allows users, be they citizens, CSOs or PSBs, to easily grasp complex data and compare various data sets with one another.

7 Conclusion

Visualisation of datasets has grown and developed rapidly over the past few years with the increased spread of new software as well as increasingly powerful, affordable, available and mobile ICT applications, such as netbook, smartphones and tablets. The fact that these applications put a high premium on the visual presentation of information has served to further increase the importance of data visualisation.

In terms of the re-use of PSI and its visualisation, there is a great potential for combining complex datasets and presenting these in a manner that makes them easier to understand and navigate for citizens, customers, researchers but also for the PSBs themselves. Apart from presenting information, however, visualisation can also be a tool that helps analysis - e.g. by making previously unknown correlations between two or more datasets visible, thus allowing for a better response to problems and improving service provision. Caution however should be taken to make sure that in the simplification of data presentation through visualisation the analysis is not also simplified and that the visualisation does not become an end in itself.
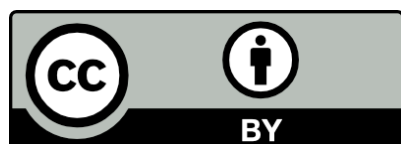
About the Authors

Daniel Dietrich was born in 1973 in Frankfurt, Germany. His academic work covers political science, computer science and communication science in Frankfurt and Berlin. He worked as Research Associate at Technical University Berlin, Department of Internet and Society until the end of 2011. He has been working for the Open Knowledge Foundation (OKFN), since 2009 and is Chairman of the German Chapter of the Open Knowledge Foundation. He is the Project Coordinator for the OKF Project Open Definition as well as the Coordinator of the Working Group on Open Government Data and the Working Group on Open Data in the EU. He is the co-founder of the Open Data Network, a non-profit advocacy organisation to promote Open Data, Open Government and Transparency in Germany, Europe and beyond. In 2011 he became Editor of the ePSI platform.

Henri Myrttinen was born in Helsinki, Finland, in 1975. He has been working academically and practically on issues related to political and social transformations with various institutions and organisations for the past 15 years. Geographically, his focus has mostly been in Central and Eastern Europe, Southeast Asia and Sub-Saharan Africa. He received his Ph.D. from the University of KwaZulu-Natal, South Africa, in 2011 and is currently working in Berlin with the Mauerpark Institute.

Copyright information