Visualizing Complexity – backdrop, repertory, objectives, principles

Michał B. Paradowski
Institute of Applied Linguistics,
University of Warsaw
ul. Browarna 8/10, PL-00-311
Warsaw, Poland
PhD
Tel. 0048225520988
michal.paradowski@uw.edu.pl

ABSTRACT

This paper examines several aspects of data visualization today. It commences with a discussion of the role of image in contemporary culture and society, and presents the different broad types and aims of data vis (presentation; verification and confirmation; analysis and exploration), with a focus on the research goals of exploratory visualization and the role and application of simulation. The historical developments which have led to the current widespread popularity of infovis and earliest canonical examples of good practice will be presented, before an array of current types of diagrams and infographics, from classic to experimental, and the factors influencing the choice of the appropriate mapping technique. Finally, a Decalogue of a data scientist is proposed, with principles which must be applied in the design of infovis interfaces, based on input from aesthetics, cognitive and mathematical sciences, studies on human perception, and ethics.

Categories and Subject Descriptors

A1 [**Introductory & Survey Literature**]; D.2.2 [**Design Tools & Techniques**]; H.5.2 [**User Interfaces**]; I.6.3 [**Simulation & Modeling Applications**]; J.m [**Computer Applications – Misc.**]

Keywords

Data visualization; image; infovis; exploratory data analysis; simulation; mapping techniques; principles.

1. THE ROLE OF IMAGE IN CONTEMPORARY CULTURE

A picture is worth a thousand words.

Most of us encounter hundreds, if not thousands of images every day. Internet services with pictures and videos, such as Flickr and Picasa or YouTube and Vimeo, register billions of visits daily, giving proof of the exceptional popularity of amateur photography and the growing exhibitionism and voyeurism of society on a nearly epidemic scale. Artists from Michael Jackson through Gotan Project to Lady Gaga have been well aware of the public's expectation of a multimedia show. More and more corporations switch to PowerPoint as the main medium of both internal and external communication, while illustrated books for adults fill shelves of every bookstore.

Thus contemporary culture manifests itself more and more "in and through that which is visual" [1]. This trend accentuating the growing omnipresence and meaning of images in contemporary culture is often considered to be a reaction to the linguistic turn and logocentrism of the 1960s and 1970s (based on Wittgenstein's philosophy of language, the structuralism of Levi-Strauss and Foucault, and the hermeneutics of Ricœur, Gadamer, Derrida, and Frye; [2]), and analogously termed as the visual/pictorial/iconic turn ([3]; [4]; [1]) or even, somewhat euphorically, as the "image revolution" [5].

While in the research practice of traditionally visually-oriented sciences (anthropology, archaeology, ethnology, folklore and art history) visualizations have been used already since the mid-19th c., they have now advanced to the status of an indispensable tool of knowledge production, retention, expression, distribution, and assimilation [6]. The proliferation of visual forms and visual-textual hybrids is much more than merely attractive diagrams and graphics; it is the new form of communication of the 21st c., so widespread that the field has been using the abbreviation "infovis".

This should not come as a surprise. From among the preferred ways of discovering the world and processing and assimilating information, in the majority of society it is visual modality that dominates, and only later verbal, kinaesthetic and tactile.

2. DATA VISUALIZATION

The greatest value of a picture is when it forces us to notice what we never expected to see.
—John Tukey, 1977

In a nutshell, data visualization means information in the form of an image, communicating its content with the help of colours, shapes, lines, layout and hierarchy, best served with a dash of creativity. A truly stunning visualization is much more than the sum of its parts. Effective visual representations of data usually fall back on the results of scientific research on visual perception and the psychology of perception in order to facilitate the comprehension of content and parallel, simultaneous processing of information and to make evident important properties and details in the data, such as relationships, principles, trends and regularities, structures, symmetries, similarities and differences and anomalies which would be hard to detect with the sole use of analytical methods.

3. THE ROLE OF SIMULATION

A special place in data visualization is played by simulation. This is by no means a novel development; already 1500 years before Confucius Chinese generals would run simulations of complicated military operations without sacrificing the life of a single soldier, which later brought them victories over not infrequently outnumbering enemy armies. Today, simulations and mathematical models are used primarily to analyse and visualize events and processes logistically difficult to carry out, whose empirical execution would be blocked by an ethics committee, laden with a high risk level, time- and money-consuming, too complex to be able to predict their results with the help of analytical methods, invisible to the human eye, taking place very fast or very slowly, or occurring on a micro or macro scale.

4. GROWTH IN THE ROLE AND MEANING OF VISUALIZATION TODAY

I keep saying the sexy job in the next ten years will be statisticians.
—Hal Varian, 2008

Several factors have contributed to the popularity and recent growth in the role and meaning of visualizations:

1. we are living in an age of numbers, in which everything undergoes measurement and counting;
2. exponentially increasing vast quantities of rich sources of data, and their widespread availability;
3. development of IT;
4. propagation of analytical techniques and interdisciplinary skills;
5. growing role of aesthetics.

5. TYPES OF DIAGRAMS AND INFOGRAPHICS

Visualizing data is like photography. Instead of starting with a blank canvas, you manipulate the lens used to present the data from a certain angle.
—Paul Butler, 14.12.2010

The choice of the appropriate mapping technique depends on:

- attributes of the data: format – measurement scale, number of observations, number of variables; distribution; statistical significance; longitudinal vs. cross-sectional observation,
- context,
- author's intentions and the question what we want to present,
- expectations of the addressee.

Deemed as best-suited for the different goals are:

- for visualizing proportions: pie chart, doughnut, word tree, treemap, tree diagram;
- for comparison of several values: bar chart, bubble chart, tag cloud;
- to track change over time: line graph, stack graph, timeline, radial diagram, time matrix, animation;
- to observe relationships between data: scatterplot, network diagram, Venn diagram;
- for geographical distribution and spatial relations: map; equal-area cartogram;

There obviously also exist combinations of the above, not infrequently coupled with interactive interfaces.

6. DECALOGUE OF DATA ANALYST AND SCIENTIST

Give me six hours to chop down a tree and I will spend the first four sharpening the axe.
—Abraham Lincoln

In the deluge of visualizations, we find a plethora of its poor instantiations. Yet, the effectiveness of conveying information may be a matter of life and death. Tufte [7] recounts how the catastrophe of the *Challenger* happened after its engineers had failed to convince the board of NASA of the danger of a lift-off in low temperatures, partly due to the form in which the message was communicated. This is why the author of a visualization ought to adhere to these elementary recommendations:

1. begin by asking a question;
2. know what you are looking for;
3. double-check your data, sources, and facts;
4. make sure you understand your data;
5. remember the purpose and target audience;
6. check geometry;
7. name the graphic and explain the coding;
8. place the data in a context understandable to the receiver;
9. avoid triumph of form over content and visual chaos;
10. avoid 3D effects, textures and gradient fill-ins and superfluous, saturated, garish and non-contrastive colours.

Acknowledgments:

The author thanks Guillaume Dumas, Sébastien Heymann, Monika Konieczny, and Taivo Lints for discussion.

7. REFERENCES

[1] Boehm, G. 1994. Die Wiederkehr der Bilder. In same (Ed.): *Was ist ein Bild?*, München: Fink, 11–38.

[2] Kawecki, W. 2010. Od kultury wizualnej do teologii wizualnej, *Kultura – media – teologia* 1 (1), 23–31.

[3] Maar, C., Burda, H. 2004. *Iconic Turn. Die neue Macht der Bilder*, Köln: DuMont.

[4] Mitchell, W.J.T. 1992. The Pictorial Turn, *ArtForum* 30, 7, 89–94.

[5] Flusser, V. 1996. *Die Revolution der Bilder. Der Flusser-Reader zu Kommunikation*, Mannheim: Bollmann.

[6] Schnettler, B. 2007. Auf dem Weg zu einer Soziologie visuellen Wissens, *Sozialer Sinn* 8, 2, 189–210.

[7] Tufte, E.R. 6.09.2005. PowerPoint does rocket science—and better techniques for technical reports, http://www.edwardtufte.com/bboard/q-and-a-fetch-msg?msg_id=0001yB&topic_id=1&topic=Ask+E%2eT%2e.

[8] Paradowski, M.B. 2011, in press. Dekalog analityka danych i infografika – *quid, cur, quomodo*. *Wiedza i Edukacja*.