Visualizing data: new pedagogical challenges

Isabel Meirelles

information visualization, systems, graphical interfaces, design education

The paper examines the burgeoning practice of visualizing data. It begins with a brief overview of this broad field and the nature of the practice throughout history. The focus is on computational interactive visualizations and the ways in which technology has given way to an expansive and expanding practice mainly centered on current issues. Information visualizations are ubiquitous and critically important to understanding several fields today, covering a wide range of content and functionality: from scientific visualizations to visual explanations of socio-political events. Technology has affected the practice in several ways, from graphical methods to the agents involved with such complex data representations: the authors and users of these applications. Selected graphical tools are examined as a means to identify recent trends. The paper concludes with questioning the ways in which we are (or not) preparing design students to tackle these new information communication challenges. The goal is to discuss—and ultimately suggest—the relevance of integrating theoretical, visual and technical aspects of structuring and representing large amounts of data into design undergraduate education.

Introduction

We live in a world that is socially and culturally media-dependent. Digital media have become a significant part of our daily interactions and means of communication. The past two decades have seen a growing need for the design of systems that work towards facilitating the way we store, access, retrieve and analyze information. The need to contextualize information so as to help us navigate the complexity of the data-rich and hyper-connected environment is ever-more present. Data visualizations are good at providing context and uncovering trends and patterns that can facilitate decision-making. New technologies have increased the possibilities of communicative expression and communication design is at the forefront of this phenomena. This paper investigates the burgeoning practice of visualizing data in the current computational domain and the implications in design education.

Definition

The visualization of data is ubiquitous and critically important to understanding several fields today. Data representations can take different forms, such as notation systems, maps, diagrams, interactive data explorations, and other graphical inventions. The practice has a long history and has been used extensively for solving problems and for communicating information by a large number of disciplines: from the sciences to engineering, from music to design. It covers a vast territory that merges different media, disciplines and techniques. In most cases, it is domain specific with specific methods and conventions for data encoding.

Information visualization depends upon cognitive processes and visual perception for both its creation (encoding) and its use (decoding). If the decoding process fails, the visualization fails. Efficiency in conveying meaning will depend on how the visual description stands for the content being depicted, whether the correspondences are well defined, reliable, readily recognizable, and easy to learn (Pinker, 1990).

Visual displays of information can be considered cognitive artefacts, in that they can complement and strengthen our mental abilities. Literature in Cognition and in Information Visualization (e.g., Norman, 1993; Card et al, 1999; Ware, 2004) suggests that the cognitive principles underlying graphic displays are: to record information; to convey meaning; to increase working memory; to facilitate search; to facilitate discovery; to support perceptual inference; to enhance detection and recognition; and to provide models of actual and theoretical worlds.
There are several terms and definitions currently in use for the various practices of visualizing data. Differences range from the medium, whether static or dynamic, to who is involved in developing them. For the sake of simplicity I will use the terms “data visualization” and “information visualization” interchangeably herein as “the use of computer-supported, interactive, visual representations of abstract data to amplify cognition” (Card et al, 1999: 7).

Digital technology

Digital technology has affected and expanded the way we visualize data: from what data we can gather, how we analyze them, to who is involved, both as makers and users. Two factors have played a major role:

- The computer as a platform for analysis and visual presentation of data;
- The network of computers as a platform for gathering and distributing visual presentation of data.

Computers as platform

If on one hand the use of computers in visualization is not recent, dating back at least 50 years; on the other, the market now offers personal computers with processing power and graphic capacity to perform complex tasks that were only executable in large and expensive mainframe stations, mostly located in research labs. Today anyone can use personal computers to interact with complex data sets in real time (unthinkable few years ago) while displaying in computer screens, also rapidly growing in resolution.

Similar trend can be found in the development of applications which until recently required sophisticated skills in programming. Programming languages have become more accessible thus broadening the range of those involved in generating data visualizations. Consider, for example, the wide adoption of Processing, an open source programming language and environment created by Ben Fry and Casey Reas in 2001, and currently used for research, pedagogical, commercial and artistic purposes (Reas & Fry, 2007). Also worth mentioning is the availability of open APIs (application programming interface) for gathering and analyzing data.

Network of computers as platform

Two recent developments have changed how we deal with the already interconnected digital environment: the Web 2.0 and the Semantic Web. In a nutshell, Web 2.0 refers to technologies that have enabled the internet as a social platform, where virtual communities and social applications are now prevalent, such as social-networking sites and tools, wikis, weblogs, video sharing, etc.

Tim Berners-Lee, who in 1989 invented and helped implement the World Wide Web (WWW) as a system for linking documents (web of documents) over the internet, has been involved with the development of the Semantic Web (web of data). The latter is conceptualized as a system for linking data from various sources such that they could be integrated and associated in different ways, and ultimately foster new knowledge. In a recent talk at the TED 2009 conference Berners-Lee urged the audience to join him in making all sorts of data available, including asking all to shout out loud: “raw data now!” To illustrate the possibilities of achieving better results when querying data rather than documents, he presented initiatives such as the “linked health care data” and how researchers have been using the resource to solve medical questions.¹

protocols as well as in making data sets available, such as the Open Data Movement. Examples of databases available online include DBpedia (access to structured information from Wikipedia), Geonames (geographical database) and MusicBrainz (community music metadatabase), to mention three. The public sector has also been active in opening their databases to the general public. A recent example of governmental data being used in information visualization is the interactive component to The New York Times story “In New York, Number of Killings Rises With Heat” that uses the New York Police Department data set to plot homicides since 2003 in the geographical space of New York City.

We have also to consider the vast amount of data that we produce in our daily interactions with digital media. Whether intentionally or not, we leave traces when twittering, talking on the cell-phone, or posting messages or photos on web-sites. The traces are in fact data that can be used for different purposes, including surveillance. Kenneth Cukier examines in The Economist Special Report on Managing Data (02.27.2010), the benefits and caveats of going from scarce to superabundant data proliferation and issues involving data security and privacy, which are currently at the center of policy discussions around the world.

A new era in information visualization?

It is possible to argue that the social-semantic computational alliance has fueled a new era in information visualization. On one hand, the internet as medium provides access to increasing volumes of data, from open data sets to data generated by our interactions with digital media. On the other, the need of cognitive artefacts to help us deal with, and, ultimately, make sense of, the information overload in which we currently live has propelled the creation of a growing number of online information visualizations. Could we consider the expansion in data visualizations in similar ways to economical models of supply-demand? Whatever the answer, the current technological environment—from the democratization of tools to the ever-more connected global computer network— is acting as a catalyst for a new generation of information visualization that needs the medium for both its production and distribution.

This is not the first time that we have experienced the desire to gather all sorts of knowledge while trying to minimize complexity by creating tools to enhance understanding while providing new models. Take for example the second half of the seventeenth century which saw the development of encyclopedias (e.g., the Encyclopédie of Diderot and D’Alembert in 1751; the Encyclopaedia Britannica in 1768) as well as museums (e.g., the British Museum, London, opened in 1759; the Hermitage, Saint Petersburg, established in 1764; the Uffizi Gallery, Florence, open in 1765; the Louvre, Paris, established in 1793). Two systems aimed at both organizing and archiving knowledge.

Perhaps it is also not coincidental that William Playfair devised the first methods for the visual representation of statistical data at the same period. The Commercial and Political Atlas, originally published in 1786, examined British commerce with other nations, and it is considered the first public document to contain charts (Playfair, 2005: 6). Some of his graphical inventions were not immediately adopted and had to wait for the next generation of visual representations of quantitative data, which happened in the second half of the nineteenth century. At this point we see an explosion of data representations and advancements in graphical methods devised by key innovators such as Charles Joseph Minard and Étienne-Jules Marey, to mention two Frenchmen. And of course, in the twentieth century the use of computers in data processing brings us closer to where we find ourselves. The new online data visualizations inscribe themselves on the history of graphical representations as much as on current developments in new media technology (e.g., Manovich, 2001, Frielie & Daniels, 2004 & 2005).

3 Story published June 18, 2009: In New York, Number of Killings Rises With Heat: <http://www.nytimes.com/2009/06/19/nyregion/19murder.html>, 06.20.09
4 Works by Minard and Marey can be found in several books including Tufte (1997), and Wainer (1997).
Data abundance

Computers have facilitated the processes of gathering and analyzing large data sets, in many cases unfeasible without computational capacity. The amount of data used in visualizations is evident when we compare recent projects to previous graphical displays. Changes in database size can be traced back to the beginning of computational data processing in visualizations, as Bertin acknowledges in his preface to the English version of his seminal book *Semiology of Graphics* (1967/1983: ix):

Thanks to the computer, information processing has developed prodigiously. We now know that “understanding” means simplifying, reducing a vast amount of “data” to the small number of categories of “information” that we are capable of taking into account in dealing with a given problem. … Our forerunners, who did not have the advantage of the computer and were generally unaware of the potential of matrix permutation, proceeded by successive simplifications. The time consumed by such process severely limited the scale and scope of research possibilities. Now, with the computer, all manner of comparisons seem within rapid reach.

The internet, for example, has provided access to a vast amount of data, that is also in constant growth. In other words, nowadays it is possible to continuously search and gather data, such that databases no longer need to be static entities. The fact that we can keep adding to a database has fostered novel methods for gathering, sorting and representing data that are in constant change. Also relevant are the kinds of content that this factor alone has opened up for examination, such as human interactions both in the physical and virtual worlds.

The accessibility to large data sets has propelled the development of tools and methods aimed at managing, manipulating and analysing structured and unstructured data. If in the past it was possible to manually structure and visualize data in the form of information graphics, nowadays computation methods are intrinsic to how we deal with large volumes of data aimed both at examination and communication purposes. The term Big Data well expresses the state of the field and the challenges ahead of us.

User-generated content

The dissemination of social networks over the internet has transformed and impacted the way we communicate and interact in the ever-more connected global environment. We have been producing a vast amount of data that travels through the internet and can be easily accessed and extracted. A number of information visualizations have focused on examining online user-generated content, data that would not exist if not for the digital environment in the first place. These projects tend to have databases of millions of data extracted from various online sources. Most projects tend to analyze the complexities of social interactions both in the virtual and physical contemporary spaces.

A well-known example is the data visualization *We Feel Fine* developed by Jonathan Harris and Sep Kamvar and initiated in August 2005. The application looks for “human feelings” in weblogs by searching occurrences of the phrases “I feel” and “I am feeling” every few minutes. The result is a database of millions of entries, that, according to the authors, increases by 15,000 – 20,000 “new feelings” per day. Images posted along with verbal information are saved as expressing the correspondent feelings. Also extracted are data related to age, gender, country, state, and city of the blog’s owner, the latter used to retrieve weather conditions together with the date of the posting. The same data is used as categories for interactive manipulation of the content. The data is computed statistically, including different ways one can examine this immense database (see figures 1-4).

---

5 [http://www.wefeelfine.org/methodology.html], 06.25.09
Figures 1–4: Screenshot of the data visualization We Feel Fine: <http://www.wefeelfine.org/>, 06.25.09

The applet presents information in a fun and effective manner. It is a good representative of several aspects propelled by recent technologies, such as the examination of online user-generated content, extraction of different classes of data, a continuously growing database, and the use of interactive statistical analysis. Finally, it is worth mentioning that they have opened the API as explained: “since we are borrowing from the feelings of thousands of people across the world to make our piece, we find it fitting for other artists to be able to borrow from our work to make theirs.”

Data-centeredness

It is possible to say that our lives have become data-centered. Not only have businesses taken advantage of analyzing and identifying new niches and economies, but as individuals we also have access and make use of several digital applets to gather, quantify and visualize our daily activities: from what we eat, where we go, to the music we listen. For example, a quick search at iPhone applications results in a profusion of tools that help track data while also charting them.

A growing number of my students are collecting data as part of their routine by using mobile or online gadgets to gather and analyze qualitative and quantitative information. A curious aspect of this trend of personal data tracking is that students are sharing charts in the same way they share photos. It is also not uncommon to have access to graph information to visually check statistical data online at various service software. Take for example the open source

<http://www.wefeelfine.org/methodology.html>, 06.25.09
blogging software WordPress which has an integrated statistical system providing up-to-the-minute charts on data such as the number of visitors, post popularity, etc. This represents a change in how we communicate and share information. It also points to the need of educating larger audiences on both how to code and decode visual information.

**Agents: authors and audience**

Data visualizations are no longer analytical tools for experts alone, rather, they range from online navigation tools to museum installation, from iPhone gadgets to social network systems. Perhaps the fact that most projects reside online might have added to the perception that we are exposed to a larger number of visualizations. On the other hand, if we examine the distribution of these projects we discover that they appear everywhere now, from news media to films (e.g., An Inconvenient Truth, 2006).

Major international art museums (not science or technology institutions) have recently commissioned and exhibited information visualizations. Among the most active institutions commissioning work are the Whitney ArtPort (Whitney Museum of American Art), and the Tate Online (Tate UK). In relation to exhibitions it is worth mentioning the Design and the Elastic Mind held at the MoMA, New York in 2008 (Antoneli, 2008). Mark Lombardi should be remembered here as a pioneer in exhibiting in art institutions his immense diagrams (mostly visualizing political scandals), which he called Narrative Structures (Hobbs, 2004).

The use of infographics in the media is not new, but has certainly contributed to the popularization of the practice. For example, the Society for News Design has been promoting the World Infographics Awards since 1992. A well-known example in the U.S. is the daily newspaper the USA Today that, since its release in 1982, has as one of its distinguishing features the large use of diagrams and visual explanations. In general, the practice has been carried over to the online version, such as the acclaimed visualizations by The New York Times. It is worth noting that Steve Duenes, Graphics Director of The New York Times, is one of the keynote speakers at SIGGRAPH 09. For the first time SIGGRAPH dedicated exhibition space for what they have called Information Aesthetics Showcase, as they explain: “in recognition of the increasingly prominent role that information visualization and data graphics are assuming in our digitally mediated culture.”

**Do-it-yourself trend**

The accessibility to open databases has fostered online services offering tools that allow anyone without programming knowledge to generate data visualizations. There are two main audiences for these services: general public and specialized enterprises. An example of the latter is the open-source project Hadoop administered by the Apache Software Foundation. The platform is directed at consolidating, combining and understanding data and has been widely used by dominant companies in technology, media and finance.

In the first set we find applets that allow the general public to use open data sets or upload their own data in order to generate visualizations using methods provided by experts. One of the most successful services is the application ManyEyes developed by Fernanda Viegas and Martin Waterberg at the IBM’s Visual Communication Lab in Cambridge, MA. Their goal and motivation is “to democratize visualization and to enable a new social kind of data analysis.”

The New York Times Visualization Lab is a version of ManyEyes that allows online readers

---

7 [http://en.wordpress.com/features], 03.22.10

8 Links to listed projects:<http://artport.whitney.org/>; <http://www.tate.org.uk/intermediaart/>, 06.20.09

9 [http://www.siggraph.org/s2009/galleries_experiences/information_aesthetics/index.php], 06.20.09

10 [http://hadoop.apache.org/], 03.22.10

to visualize data generated by the newspaper’s editors. The rational for the tool is quite enlightening: “Just as readers’ comments on articles and blogs enhance our journalism, these visualizations – and the sparks they generate – can take on new value in a social setting and become a catalyst for discussion.” 12

Another application in this area is Wolfram|Alpha devised by Stephen Wolfram and released in May 2009. The application uses Wolfram’s analytical and mathematical language Mathematica to compute, analyze and visualize data sets. The project’s long-term goal is quite grandiose and proposes “to build on the achievements of science and other systematizations of knowledge to provide a single source that can be relied on by everyone for definitive answers to factual queries.” 13

Google has also devised a set of gadgets for displaying data, which includes “motion chart,” the animated graphical method devised by Swedish Professor Hans Rosling. The tool was first presented at the Ted 2006 conference, where Rosling revealed through animated statistical data how much we can learn by looking at the rates of change and data patterns over time. 14

Similar to publishing tools that enable anyone to write and publish blogs without the need to learn programming, these tools are helping to educate, spread and consolidate the value of data visualizations.

Visualization as navigation

Another recent trend is the use of visualization methods employed as navigation tools. A well-known example is the “tag cloud” or “word cloud” popularized by Flickr, the image and video sharing online community. It has become a common tool to navigate content by hierarchical ranking and employed in many web-sites. The same method has been extensively used also for data visualization purposes, such as the project “Inaugural Words: 1789 to the Present” by The New York Times (January 17, 2009), which analyzes the inaugural addresses of all Presidents in American history. 15

Launched in March 2009, the Flickr Clock is a new application developed by Stamen Design to serve as a browser to watch videos. Videos are organized by the time and in the order in which they were uploaded to the site. The application reminds me of a bookshelf, an endless bookshelf of moving images (see figure 5).

Figure 5: Screenshot of Flickr Clock <http://www.flickr.com/explore/clock>, 06.26.09

13 <http://www.wolframalpha.com/about.html>, 06.18.09
A much simpler navigation tool, but worth mentioning, is the flow diagram used by the MoMA to present the exhibition calendar online (see figure 6).

Figure 6: Screenshot of MoMA’s calendar <http://moma.org/visit/calendar/exhibitions>, 06.28.09

Serving a more specialized audience is the outstanding project *Well-formed Eigenfactor Visualizations* by Moritz Stefaner, which explores emerging patterns in citation networks. A series of data visualizations present the information flow in science based on the Eigenfactor metrics and hierarchical clustering. The database is comprised of around 60,000,000 citations from more than 7000 journals, originated from Thomson Reuters’ *Journal Citation Reports* 1997–2005. The project is a compound of four interactive displays that provide: an overview of the whole citation network in a circular schema; a visualization of changes in the journals’ influence and clustering over time in the form of flow diagram; a hierarchical clustering in the form of a treemap; and a spatial map with journals represented as circular nodes positioned in the plane according to clustering, and node size given by the citation score. The four different data visualizations present novel possibilities for navigating content while providing context that support insights (see figures 7–10).

Figures 7–10: Screenshot of *Well-formed Eigenfactor Visualizations* <http://well-formed.eigenfactor.org>, 06.22.09
New pedagogical challenges

The central question is how to prepare not only future generations, but ourselves included, to deal with data proliferation: from learning how to structure and analyze data to developing skills and methods for effectively visualizing information. It is critical to foster understanding of relationships between visual thinking, visual representation, and visual communication. How can we provide informed criteria to support the design process of structuring, representing, and communicating information in static and dynamic media?

I am not alone in advocating for the urgent need for visual literacy in all fields of knowledge (e.g., Horn, 1998). More specifically in what concerns the demands of current visualization practices, I would like to suggest that learning computational analytical methods are crucial to structuring large data-sets, which would be unfeasible with manual manipulation. Programming languages have the potential of expanding cognitive, and creative abilities and skills used in the process of solving design problems, more specifically in dealing with complex systems.

It is possible to argue that the learning experience of solving communication design problems in a dynamic environment can create new opportunities for achieving effective communication solutions in any medium. Furthermore, the introduction of programming languages at an early stage in the education can inform students toward experimenting with and deciding which medium and format is most appropriate and effective for a given design problem. The premise is an education that fosters the understanding of human-centered and context-based information communication, rather than methods centered on object or product development. Ultimately, a pedagogical model integrating multi-dimensional and interdisciplinary ways of thinking and exploring design problems. Ben Fry offers a holistic solution of how we might move forward (2008: 5):

Given the complexity of data, using it to provide a meaningful solution requires insights from diverse fields: statistics, data mining, graphic design, and information visualization. However, each field has evolved in isolation from the others. Thus, visual design—the field of mapping data to a visual form—typically does not address how to handle thousands or tens of thousands of items of data. Data mining techniques have such capabilities, but they are disconnected from the means to interact with the data. … We must reconcile these fields as parts of a single process. Graphic designers can learn the computer science necessary for visualization, and statisticians can communicate their data more effectively by understanding the visual design principles behind data representation. The methods themselves are not new, but their isolation within individual fields has prevented them from being used together.

Conclusions

It is unquestionable that there have been drastic changes in how we create and consume information. The interconnected digital world has affected the storage, retrieval and analysis of data. Data visualizations currently play a major role in helping us navigate and make sense of the information overload and the complex data-rich environment we experience daily. It is a two-way road: new technologies have fostered the development of novel methods for visualizing data, at the same time that there is a need for cognitive artefacts that can provide theoretical models for dealing with the ever-more connected global computer network.

As described above, the result is a growing number of data visualizations developed by a wide range of people, from programmers to designers, from sociologists to architects, and in most cases by interdisciplinary teams. However, the wide spread of these applications does not guarantee their quality. A fundamental question remains: What are the implications to the design community, more specifically to design education? Are we preparing students to contribute to this burgeoning effort in data visualization? How can we advance the study and development of information visualization practice?

Design practice, criticism and education today face new challenges due not only to innovations in technology—affecting both how we produce and how we communicate—but also to new paradigms in media communication. There is a need to acknowledge the interplay of technology and analytical tools in the design pedagogy. Programming has become a necessary knowledge in modern information visualization practices. There is also the need to integrate other fields of knowledge into the design process of structuring and representing information. At
least two areas are fundamental in my view: visual perception and programming languages. The new designer will need to be media and visual literate.

References

About the author
Isabel Meirelles
Associate Professor, Department of Art + Design, Northeastern University, Boston, MA.

Meirelles holds a B.Arch from Febasp, Sao Paulo, Brazil; an M.Arch in History and Theory of Architecture from the Architectural Association School of Architecture, London; and an MFA in Graphic Design from Massachusetts College of Art, Boston. Professional experience includes work as architect and urban designer, head of museum departments, and art director in publication and interactive design. Meirelles past professional practices inform and shape her current research in design. Her scholarly work focuses on the theoretical and experimental examination of the fundamentals underlying how information is structured, represented, and communicated in different media. Her research has been disseminated in scholarly design journals such as *Visible Language*, and international conferences such as SIGGRAPH.

Email: m.meirelles@neu.edu