

Panel: The Impact of Social Data Visualization

Robert Kosara*
UNC Charlotte

Ola Rosling†
Google

Warren Sack‡
University of California, Santa Cruz

Fernanda B. Viégas§
IBM Visual Communication Lab

Sara Wood¶
Swivel

1 INTRODUCTION: SETTING INFORMATION FREE BY VISUALIZING IT (ROBERT KOSARA)

In his famous talk at TED, Hans Rosling [7] used animated and interactive graphics to get his point across. Rosling’s energetic style and interesting message certainly helped, but the means he used were an incredible demonstration on the power of visual communication. From the perspective of current visualization research, though, the displays used were simple, and one has to wonder if they would have been accepted for publication at the InfoVis conference. This raises the question of whether our criteria for evaluating visualizations are really the right ones, and what other criteria there might be.

Information – some argue – wants to be free. Certainly data that was collected with the support of public funds should be freely available, and a lot of it is, at least in principle. But access can mean many things, from easily usable tables of numbers to obscure data formats and querying interfaces to printed tables in locked PDFs.

Availability of data does not equate meaningful access. Examples like the 2000 US Census [8] show that just making the data available is not enough if the format is obscure and the interface for accessing it is overly bureaucratic. Even simple visualizations like gCensus [4] and a project by Juice Analytics [5], which both create layers for Google Earth [3], finally provide access to that data years after it has been published. In a similar way, TheyRule [6] shows data that is available in principle, but that needed to be collected in one place and shown visually to make connections visible that were hidden in plain sight.

An interesting problem that arises from the growing availability of useful data visualization tools for the general public is that while making data available so far did not necessarily lead to that data being actually used for investigation, be it by researchers, journalists, or just interested individuals. Will less data become available for fear of data mining by the general public? Will the possibility of connecting many different data sources make breaches of privacy possible that will lead to tighter restrictions on what data can be published at all?

Of course, all of the above is more a question of general data processing than visualization. Freeing data from an obscure format does not require or even involve visualization, and neither do many kinds of analysis. What makes visualization so powerful is its compelling visual nature. Seeing a graph display numbers like unemployment or, recently, climate change, is much more interesting to look at and more impressive than reading a table with numbers.

Visualization, therefore, is an agent of change, a powerful tool that needs to be understood not just in terms of mathematics or perception, but in terms of the impact it can have on the views and opinions of people.

*e-mail: rkosara@unc.edu

†e-mail: olar@google.com

‡e-mail: wsack@ucsc.edu

§e-mail: viegasf@us.ibm.com

¶e-mail: sara@swivel.com

2 DATA ANIMATION (OLA ROSLING)

(Written by Robert Kosara on behalf of Ola Rosling, due to time constraints)

Gapminder was developed to support the presentation of health data for all countries in the world over more than forty years. Its target audience is an informed and interested general public, rather than visualization researchers. The graphical representation was therefore chosen to be visually compelling yet simple enough to be understood by a lot of people.

The strength of Gapminder is not just its effective visualization, but also its narrative structure and the use of animation. Unlike most general visualization tools, gapminder is very specific to Hans Rosling’s presentation, and thus able to perfectly support it. All transitions are animated, to make it easy for the audience to follow what is going on, and also to see developments that happen over time. In addition, many animated objects can leave traces, to produce static images that still show developments for a few select objects.

3 OVERCOMING COGNITIVE BIAS (WARREN SACK)

Data concerning groups or networks of people, i.e., “social data,” are prone to a number of misunderstandings frequently referred to as “social biases” or “attributional biases,” but more generally termed “cognitive biases” in the fields of cognitive science, social psychology, and elsewhere. For example, the false consensus effect is well known in public opinion polling: people tend to overestimate the number of people who agree with their own opinion. Pluralistic ignorance is another example of social bias: it occurs when almost everyone rejects a group norm, yet believes that most other group members accept it. False consensus and pluralistic ignorance are symptomatic of the fact that people have a hard time estimating the distribution of others opinions about some given topic. Other kinds of cognitive bias affect related areas of cognition and perception. For example, a number of biases make people prone to errors in estimating probabilities. Another set of cognitive biases influence memory and, unfortunately, make eye witness testimony in a court of law error prone. Some of these forms of cognitive bias can be explained by noting that people tend to employ heuristics rather than algorithms or scientific methods to make decisions. Conversely, we can understand various forms of scientific methodology as procedures for overcoming these cognitive biases (e.g., the employment of random statistical sampling to overcome ones naturally tendency to rely too heavily on personal experience and anecdotal evidence).

Following in a longer tradition of technological development (e.g., Douglas Engelbarts project of “Augmenting Human Intellect”) many in the field of information visualization have sought to design and evaluate visualization technologies according to their ability to amplify or augment our cognitive abilities [1]. The general form of the argument is that computers can be used to build upon and strengthen our cognitive and perceptual abilities. Unfortunately, this is an unworkable approach in certain domains such as within the realm of social data because our cognitive “abilities” are liabilities prone to error and bias. Or, in other words, while seeing is believing, we should not believe everything we see because our eyes can deceive us. Consequently, I propose a complementary methodology of design and evaluation in which we judge social data visualization technologies according to their powers to help us

overcome our cognitive biases. I will present results from an NSF-funded project in which my students and I are working to design an interface and search engine appropriate for deliberative democracy in which the cognitive bias of false consensus is minimized.

4 MANY EYES: DEMOCRATIZING VISUALIZATION (FERNANDA VIÉGAS)

Historically visualization technology has been accessible only to the elite in academia, business, and government. But recent years have witnessed internet-based visualizations ranging from political art projects (e.g. Theyrule [6]) to New York Times stories (Faces of the Dead [2]). These displays are seen by thousands and have brought visualization to a new, large audience. Unfortunately, the revolution is not yet complete: while lay users can view many sophisticated visualizations, they have few ways to create them.

In order to “democratize” visualization, we have built Many Eyes, a web site where users may upload data, create interactive visualizations, and carry on discussions. The goal is to support collaboration around visualizations at a large scale by fostering a social style of data analysis in which visualizations not only serve as a discovery tool for individuals but also as a medium to spur discussion among users.

To support this goal, the site includes novel mechanisms for end-user creation of visualizations and asynchronous collaboration around those visualizations. Traditionally, visualization builders have focused on systems that support a single user or small groups, while seeking ways to present huge graphs, very high dimensional data sets, and tables with billions of rows. Many Eyes embodies the reverse perspective: Instead of scaling the size of the data, we scale the size of the audience. What happens when we design for massive public visualization?

Since the site launched earlier this year, users have uploaded data and created graphics on everything from DNA microarray data, to co-occurrences of names in the New Testament, to Senate testimony of a top White House official. Apart from data analysis, our preliminary results show that Many Eyes is used for goals ranging from journalism and advocacy to personal expression and social interaction. We contend that these findings suggest a growing role for visualization as an expressive medium—and the power of putting visualization into the hands of the people.

5 EMPOWERING THE PUBLIC (SARA WOOD)

Numbers affect our every day life. They are in newspaper articles, they serve as justification for policies, they help guide our decisions and support (or refute) our opinions. But finding them is no small task. The data behind the quotes seem to be ephemeral. Despite the huge leap in access to information enabled by the web, the focus has been primarily on narrative text. Without access to those data, consumers of information are left with only half of the picture. We are forced to take claims at face value rather than develop alternative hypothesis of our own. This leads to a society where often facts are neither understood fully nor considered critically. This is a problem.

A solution to this problem is that any individual whether they are in Bangladesh or in Boston would have equal access to all relevant data both for their own decision-making and for research purposes. To make this widely accessible, implementation would need to utilize the technologies and methods of the social web. Swivels online community and suite of features enables users to compare and contrast indicators from vastly different data sets on different topics for a variety of purposes – which is what is needed as a start to get people to explore data and consider its implications on their own lives.

The ability to visualize data is an essential piece to the puzzle of data exploration and use, and can no longer be reserved for academics. The fear data providers have of having their data trivial-

ized, misunderstood or misrepresented also apply to the process of making visualizations, even simple ones. We believe that a strong community supported with data and tools will help advance the role of both data and visualizations in positive directions.

Visualizing data changes how data are understood and increases interest in data generally, which will encourage more and better data to be developed. Making data more visual exposes patterns in data not otherwise apparent. Putting visualizations on the web further encourages insight by allowing a broad audience to access, comment on and discuss what they see.

6 BIOGRAPHIES

Robert Kosara is an Assistant Professor of Computer Science at the University of North Carolina at Charlotte (UNCC). He received his M.Sc. and Ph.D. degrees in Computer Science from Vienna University of Technology in Austria. He considers the visualization of data not only an interesting topic in itself, but wants to see visualization widely adopted in practice.

Ola Rosling is one of the co-founders of gapminder, and the main developer behind the system. Ola is in the process of moving from Sweden to California, where he is joining Google.

Warren Sack is a media theorist and software designer. He has exhibited work at the ZKM—Center for Art and Media, Karlsruhe, Germany; the Walker Art Center in Minneapolis; the New Museum for Contemporary Art in New York City; and, on the Artport website of the Whitney Museum of American Art. Warren earned his B.A. from Yale College and his Ph.D. from the MIT Media Laboratory. He currently teaches in the Digital Arts & New Media M.F.A. program and in the Film & Digital Media Department at the University of California, Santa Cruz.

Fernanda B. Viégas is a research scientist in IBMs Visual Communication Lab. Together with Martin Wattenberg, she created Many Eyes. Her work addresses the social and collaborative aspects of data visualization, focusing on representations of online communities to support identity, collective memory, and story-telling. Her visualization-based artwork has been exhibited in galleries in New York, Los Angeles, and Boston.

Sara Wood is the Chief Data Officer for Swivel. Sara has spent the better part of the last decade at institutions working with some of the world’s most important data: the World Health Organization, Harvard School of Public Health, the UN and UNDP. Previous to that she worked for a number of technology companies and research organizations where she helped to solve emerging issues of content and data management on the web.

REFERENCES

- [1] S. K. Card, J. D. MacKinlay, and B. Shneiderman, editors. *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann Publishers, 1999.
- [2] G. Dance, A. Pilhofer, A. Lehren, and J. Damens. Faces of the dead in Iraq. http://www.nytimes.com/ref/us/20061228_3000FACES.TAB1.html.
- [3] Google Earth. <http://earth.google.com/>.
- [4] I. Haque. gCensus. <http://gocensus.stanford.edu/gocensus/>.
- [5] Juice Analytics. Census data visualization. <http://www.juiceanalytics.com/weblog/?p=119>.
- [6] J. On. They rule. <http://theyrule.net/>.
- [7] H. Rosling. Talk at TED. <http://video.google.com/videoplay?docid=2670820702819322251,2006>.
- [8] U.S. Census Bureau. <http://www.census.gov/>.