

# Panel: Changing the World with Visualization

Robert Kosara, Sarah Cohen, Jérôme Cukier, Martin Wattenberg

---

## 1 INTRODUCTION

Data collected by government agencies, or by other parties paid for with taxpayer money, have been available in the past. Most of those data were difficult to get access to, and even harder to understand. Recent efforts like data.gov are making vast amounts of government and other data easily accessible. While the raw data is now available, for the vast majority of people, it is still of little use. The data formats are often not compatible with spreadsheet software, and even if: who is going to sift through thousands – or even millions – of numbers to try to understand them?

Visualization is what makes data accessible, intelligible, and interesting. Even a few simple charts can shed light on a large dataset, and will be seen by many more people than the actual numbers.

## 2 EXPLORATION TO PUBLICATION: INFOVIS IN INVESTIGATIVE AND PUBLIC AFFAIRS JOURNALISM (SARAH COHEN)

News organizations have long harnessed interactive visualizations to let their readers and viewers explore the data that was collected for stories. A decade ago, for example, the Washington Post allowed readers to explore unsolved murders in Washington, D.C. in a crude but effective online map that was among the first of its kind in the mainstream media. Now it is common.

At the New York Times, real-time maps let readers explore the 2008 election results as they were released [3]. At the Washington Post, each of the Obama administrations appointees are coded and presented by their connections, their agencies and their stage in the process [6].

However, this work often comes too late for the journalists researching a story. The graphics are produced by artists at the end of the process for presentation rather than at the beginning for discovery. Those deep and insightful interactives that are presented to readers cannot help shape the story that late. Instead, reporters often have to find these trends on their own and sometimes just hope that their interviews and reviews of documents and databases will reveal the same results that their readers will take away from exploring interactives online.

Of course, reporters have long had their own versions of early visualizations: maps tacked up to the wall and highlighted in different colors to show local property owners or network diagrams sketched on a whiteboard showing the interconnected local contractors and politicians. Analysis of public records databases has always involved crude static graphics. Now some newsrooms are becoming a little more sophisticated. In particular, whenever time and place are important in a story, interactive visualizations can help us discover broader truths and lead us to the most illustrative examples. Many are never published, and some are used for only one day.

For a 2008 story on deaths among detainees awaiting deportation [4], we used an interactive map with frequency charts that showed each death by age, by year and by cause. We used it for 10 minutes to see very quickly what may have taken us much longer to find out otherwise: young detainees had died more frequently in the West from

accidental and other sudden causes, while those in the East were sicker and older. It told us where we had to go to find the stories.

The same year, we used the same basic template to identify the best city to focus on in Iraq reconstruction [7], and the best neighborhood for recent abuses among landlords [5]. We have used small multiples to quickly identify agencies and programs; sparklines to highlight searches agricultural subsidy spending; and timelines to isolate key components of the presidents first 100 days.

Most of these visualizations will never see the light of day, though they are often used as a start for the published versions. The reason is that these are unconfirmed early documents that would require detailed checking to make them publishable. Instead, we let them lead us to the story that we can confirm.

Unfortunately, our data sources are almost never on the Web or otherwise generally available, except to the extent we are using them to match or supplement information we have gotten in other ways. On detainees and another story on lead contamination in water pipes, we had documents leaked by a source. We spent six months trying to get records for *Forced Out* – what few records there were on the city's data feed were overly censored. On Iraq, it was a Freedom of Information Act request that took six months to get. This is still the key to investigative or accountability journalism: uncovering something that has never been uncovered before.

## 3 CAN DATA VISUALIZATION IMPROVE KNOWLEDGE AND DECISION-MAKING? (JÉRÔME CUKIER)

Coming from a governmental organization (Organisation for Economic Co-operation and Development, OECD), the most obvious way I see in which data visualization can change the world is through transmission of knowledge from experts to the people. If all citizens had perfect information each time they made a decision, our economy and society would be stronger. People could have access to the same quality of information that their elected officials use when making policies.

The idea that government agencies should not only collect data and turn that information into a public good, but also present it visually for maximum impact is not new. Almost 100 years ago, Brinton and Neurath described the time saved and the added accuracy of information gleaned from visual representations [1].

However, in the second half of the past century, we have mostly tried to achieve this by pushing abstract indicators such as GDP and, as far as the general public is concerned, with very limited success. In 2007, we conducted a study [2] with the Institute for Studies and Economic Analyses (ISAE) where we asked people to tell us what they thought was the GDP growth rate of their country. Only 3% were able to make an educated guess. However, this is one of the numbers which is the most represented visually by the mainstream media.

So we come to question our approach. Do we really have to claim people's mind share with data that do not necessarily relate to them in order to help them? On one hand, we are trying to come up with different indicators, different ways to measure progress. On the other, we are trying to open up and improve access to our data, so that others can create visual tools that can help solve specific problems. There have been several initiatives to publish large data sets, such as data.gov or Amazon's public data sets. Increasingly, organizations are willing to open their data to others that might turn it into something useful.

The situation where the ones who represent data are not the ones that collect it raises a new set of questions. First our personal bias as data providers: the fear that our data would be misinterpreted. We have

- 
- Robert Kosara is with UNC Charlotte, E-mail: rkosara@unc.edu.
  - Sarah Cohen is with Duke University, E-mail: sarah.cohen@duke.edu
  - Jérôme Cukier is with the OECD, E-mail: jeromecukier@gmail.com.
  - Martin Wattenberg is with IBM Research, E-Mail: mwatten@us.ibm.com.

Manuscript received 31 March 2009; accepted 27 July 2009; posted online 11 October 2009; mailed on 5 October 2009.

For information on obtaining reprints of this article, please send email to: tvcg@computer.org.

difficulties to let go of data, to allow people to play with our stuff for fear they do not take into account all their subtleties and limitations. But even if we overestimate the risk of errors, will they ask the best questions? Will they use the right series? There is also no link between the visual quality of a representation and the reliability and verifiability of data. Visualizations add a new layer of abstraction on top of another one: observation of facts. Will a given visualization help a viewer understand a message better? Will it be truer than bare data? Just because a visualization is well-made, does it make it right?

Finally, how can the right data visualizations reach their audience? It is difficult enough to find data and make it discoverable. How to do it with visualizations which, by nature, come in all shapes and sizes?

#### 4 WHAT HAPPENS AFTER “AHA!” (MARTIN WATTENBERG)

The world will not change itself. To have an effect a tool needs to make people think, “I have to take matters into my own hands.”

What does this mean for visualization? At the least, a change in perspective. Our field grew out of a scientific tradition of dispassionate analysis. Yet passion and emotion are forces that can change the world. A diagram that yields an “Aha!” will not lead to change unless the viewer then says, “Oh no!” or “But that’s my neighborhood!” or “So that’s how I can help!”

Of course, we have to be careful in making an emotional appeal with information graphics. Advocacy for a good cause does not justify a deceptive diagram. With visualization we have the opportunity to ground debates in honest, detailed discussion of the facts. And making a visualization emotionally engaging does not mean giving up on “aha” moments. On the contrary, a moment of insight, in which people see facts and patterns for themselves, can be rhetorically powerful.

One key to emotionally engaging visualization is storytelling. The graphic itself may tell the whole story, or it may be part of a larger story that an advocate is telling. (A good analogy may be photography and photojournalism.) Figuring out how to balance a desire for a point of view in a visualization, without distorting the data, is a broad and critical design challenge.

A second key to action is to show people data that is relevant to them. Such data could be specific to a person – perhaps customized to their neighborhood – or specific to a news event. That in itself leads to a host of technological questions – how can we make data easily searchable, or crowdsource the gathering of data? It also means we need to spend more effort on visualization of documents and unstructured text, which are increasingly available in electronic form.

To enhance the power of visualization, we need to focus on what comes after the “aha” moment. We need to study the emotional power of different designs, in addition to their clarity. And we need to invent new ways to find and display the data that matters to our users.

#### 5 CHANGING THE WORLD WITHOUT RISKING TENURE (ROBERT KOSARA)

Over the last few years, *Visualization for the Masses* has become an active area of research in visualization. Despite the existing work, there are still issues that make this kind of work problematic, though: How can we evaluate a visualization for a broad audience? How do we know it is of any use? How do we compare different ones?

Visualization has hardly invented broad dissemination of information to a mass audience (in fact, it has hardly even discovered it), so we need to look at other fields: communication, film, advertising, journalism, etc. There is an abundance of work on how to study audiences, in many different ways, for many different purposes. Many of these will certainly need to be adapted for visualization, but there is plenty of work that has already been done for us.

One way of evaluating visualizations are field studies. These cannot be limited to online studies, but need to involve all available media. How do readers of a newspaper (on paper) read a visualization differently from online readers? Does screen size make a difference? What about age, sex, native tongue, education, cultural background, etc.? Visualization papers tend to assume a particular type of user who is hard to find in the real world: the domain expert. We need to find a way to talk to non-expert, interested people with average intelligence

and computer skills. What do they want to learn? Where do they get their news? And how can we reach them?

A metric that some may consider unscientific is uptake by the popular media (including blogs). A visualization that gets linked to, shown on TV, reprinted in newspapers, etc., undoubtedly has made a bigger difference than one that gets ignored. This requires us to pick up current topics and react quickly – within a day, rather than within six to nine months. It also requires us to leave the familiar world of academia, and go out into the (potentially hostile) real world. The impact of such work will be much larger than what we do today.

Until recently, finding real, interesting data to visualize was a challenge. Now that such data are easily and abundantly available, we have an obligation to put our knowledge and tools to use to make something out of that data – and in doing so, to make a difference in the world.

#### 6 BIOGRAPHIES

**Robert Kosara** is Assistant Professor of Computer Science at University of North Carolina, Charlotte (UNCC). His research focuses on two seemingly opposite directions: theoretical foundations of visualization, and using visualization for communication. His credentials include the website EagerEyes.org and the open source Parallel Sets program. Robert received his M.Sc. and Ph.D. degrees in Computer Science from Vienna University of Technology (Vienna, Austria).

**Sarah Cohen** became the Knight Professor of the Practice of Journalism and Public Policy at Duke University in July. Prior to the appointment, she worked as a database editor for The Washington Post for 11 years, focusing on national and local investigative reporting. She has shared in most national prizes for her work there, including the Pulitzer Prize in Investigative Reporting, the Goldsmith Award for investigative reporting, the Robert F. Kennedy journalism prize for public service, and the Selden Ring Award in investigative reporting. She also was a nominated finalist for the Pulitzer Prize in public service. Before the Post, Cohen worked as the training director for Investigative Reporters and Editors, a grass-roots training organization that works to improve in-depth journalism. She has a masters’ degree from the University of Maryland in public affairs reporting and a BA from the University of North Carolina – Chapel Hill in economics.

**Jérôme Cukier** is a data editor at the Organisation for Economic Co-operation and Development (OECD). His job is to increase the quality, usefulness and usage of OECD statistics. He holds a master’s degree from the EM Lyon (Lyon, France) and an MBA from the University of Texas at Austin.

**Martin Wattenberg** is a computer scientist in IBM’s Visual Communication Lab, which researches new forms of visualization and how they can enable better collaboration and storytelling. Prior to joining IBM, Wattenberg was the Director of Research and Development at SmartMoney.com, a joint venture of Dow Jones and Hearst. He is also known for his visualization-based artwork, which has been exhibited internationally. Wattenberg holds a Ph.D. in mathematics from U.C. Berkeley.

#### REFERENCES

- [1] W. Brinton. *Graphic Methods for Presenting Facts*. The Engineering Magazine Company, 1914.
- [2] F. Fullone, M. Gamba, E. Giovannini, and M. Malgarini. *What Do Citizens Know About Statistics? The Results of an OECD/ ISAE Survey on Italian Consumers*. Technical Report, OECD, 2007.
- [3] The New York Times. *President Map*. <http://elections.nytimes.com/2008/results/president/map.html>.
- [4] The Washington Post. *Careless Detention*. <http://www.washingtonpost.com/wp-srv/nation/specials/immigration/index.html>.
- [5] The Washington Post. *Forced Out*. <http://www.washingtonpost.com/forcedout>
- [6] The Washington Post. *Head Count*. <http://www.washingtonpost.com/headcount>.
- [7] The Washington Post. *Money as a Weapon*. <http://www.washingtonpost.com/cerp>.