

Storytelling: The Next Step for Visualization

Robert Kosara and Jock Mackinlay, *Tableau Software, Seattle*

Presentation—specifically, its use of elements from storytelling—is the next logical step in visualization research and should be a focus of at least equal importance with exploration and analysis.

Much of the early visualization research focused on novel techniques, leading to questions about which one to use and for what task. This gave rise to evaluation papers that compared techniques and tried to ascertain the perceptual mechanisms behind them.

Today, researchers have a good understanding of the visualization design space, to the point where they can find suitable techniques for most datasets and tasks. Although more research is clearly needed in this area, the body of existing work provides useful ways of working with data. However, solid knowledge about the best ways to present and communicate data is still lacking.

Humans have always tied facts together into stories, effectively presenting information and making a point in a memorable way. In addition to offering an incredibly popular way of conserving data and passing it on, stories also provide the connective tissue between facts that make them memorable.¹ The use of elements from storytelling is therefore the next logical step in visualization research, specifically because storytelling can offer an effective way to present data.

STORY DEFINITION AND MODELS

At its essence, a story is an ordered sequence of steps, each of which can contain words, images, visualizations, video, or any combination thereof. Here, we focus on stories that primarily consist of visualization steps, which can include text and images but essentially are based on data.

In traditional stories, order roughly corresponds with time, which is crucial for understanding causality: events that happen earlier can influence later events, but not the other way around.¹ Stories often are not told in a linear fashion, but rather use flashbacks and other literary devices. However, within each segment, the order must be consistent—and the order of different segments clear—for the story to be comprehensible.

This working model of story construction is the basis of the way journalists work. They collect information through research and interviews, which gives them the key facts, and then they tie those facts together into a story. Because the goals and tasks during the research phase differ from those in the writing phase, so do the tools. For example, if the journalist works on a TV show, some of the material from the research phase, such as pieces of video, might end up in the final story, but most of the source material only serves as raw background information.

Data analysts can use visualization for exploration, analysis, or presentation. However, the way they use the technology can differ, so the choice of technique will differ, as will how much and which data is shown. Although visualization researchers often tacitly assume that the tools they use for analysis are suitable for presentation

as well, this is a very limiting assumption. The goals and approaches in analysis are different from the ones in presentation, where the main objective is to get a point across or explain a finding.

HISTORY OF STORYTELLING IN VISUALIZATION

Some of visualization's earliest examples aimed to show and explain, not analyze. For example, Florence Nightingale used charts not to analyze data about the causes of death in the Crimean War, but to emphasize the size of the problem to numerically illiterate politicians. Similarly, John Snow's map of cholera cases in 1850s London was not intended to find the water pump that was spreading the disease, but to present Snow's evidence after he identified it using other means.² Likewise, Charles Minard's map of Napoleon's march on Moscow was primarily a means of telling the story of the soldiers' plight and less an analytic tool for understanding complex related data.

In modern visualization literature, the earliest work on storytelling that we could find describes the use of storytelling techniques to show the development of a hostage situation.³ But while the article makes interesting points about the power of storytelling, it did not describe actual visualization, perhaps because it was primarily based on map views without numerical data. It did, however, describe the need to communicate the key information about a situation clearly and concisely, and argued that stories are a good vehicle for this purpose.

More recently, Edward Segel and Jeffrey Heer classified the patterns and techniques that news media outlets use to tell stories visually.⁴ They identified commonly used approaches, including different layouts and semantic story structures. One of the more interesting structures is what they called the martini glass, which starts with a broad introduction, then narrows to make a particular point, and then opens up interaction and exploration to the viewer.

Jessica Hullman and Nick Diakopoulos explored the role of rhetoric in narrative visualization and how it frames the presented data.⁵ Specifically, they identified approaches for communicating authority and data completeness, showing how these cues can help to prioritize particular interpretations.

On the evaluation side, George Robertson and colleagues investigated the effectiveness of animation in presentation and analysis—in particular, the type used by Gapminder—and found that had limited effectiveness.⁶ Still, Gapminder demonstrates that animated transitions are useful for explaining visualizations and getting people interested in data.

Given that a presentation's goal is generally to get a point across and to have the audience remember it, visualization's effect on memory is important. Just like

stories, embellishments add context to the presented information. In a study of the effects of embellishments on memory, embellished charts—similar to infographics—were easier to remember.⁷

Data-based infographics are often considered bad visualization—and, when done for marketing purposes, they often are. However, visual journalism has a history of incorporating informative and well-designed graphics that attract readers' attention without distracting from the data. Although little work has been done on understanding these techniques from an academic perspective, interest in the area is growing.⁸

Researchers can use stories not only to support discussion and decision making but also for process analysis.

But storytelling is not limited to information visualization. Research into the role that storytelling should play in scientific visualization reveals that it is valuable when illustrating findings gathered with complex tools.⁹ In this case, storytelling features often include different views of the same data to make it easier to understand but are less concerned with overall story structure.

Telling stories about data is a natural outcome when visualization is used in collaborative settings. Systems such as Many Eyes and Tableau Public have long been used as vehicles for telling stories about the data being visualized.¹⁰ In a more structured context, researchers can use stories not only to support discussion and decision making but also for process analysis.¹¹ Stories can thus serve as part of a finding's provenance, similar to an event's narrated history.

TYPICAL STORYTELLING SCENARIOS

Visualization researchers present information in a variety of settings in front of many diverse audiences. Each of the following scenarios has requirements related to the techniques used, the presentation structure, the amount of anticipated interaction, and so on.

Self-running presentations for a large audience

Most news media presentations follow a similar style: a presentation is created once and then viewed by multiple people independently, without interaction between the presenter and the audience. Some of these stories are entirely self-running—such as in a film clip—while others require the user to click through a slideshow, but neither involves any interaction beyond simple timeline control.

The goal of such stories is similar to that of a written feature: to get a point across and explain it in sufficient

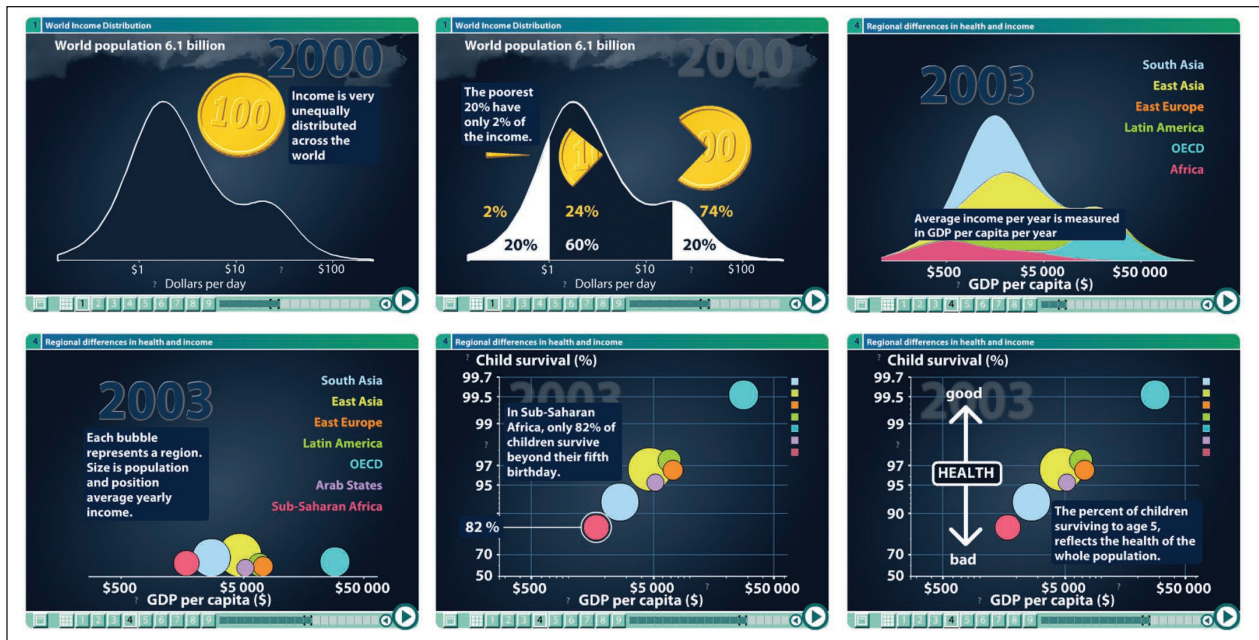


Figure 1. Gapminder. This sequence transitions data from a stacked area chart to a scatterplot, explaining what to look for in the visualization.

detail for the viewer to both understand it and trust that it is based on real facts and data.

A key concern with self-running presentations for large audiences is how to engage readers or viewers. To address this, many presentations offer a static view with a teaser and an information bite that does not require interaction. This is similar to including a catchy title and lede in a printed feature, both of which are meant to pique the readers' curiosity and entice them to read the rest of the article.

In addition to informing about an issue, a story often tries to raise the readers' or viewers' awareness or generate interest in a topic. To provide a deeper connection, the story aims to get the reader or viewer closer to the data, or to at least find out how it relates to them—one example is to provide a map that connects the viewer to the immediate area.

Live presentations

Most business presentations employ live speakers in front of live audiences. A good example of this model is Hans Rosling's Gapminder presentation at the 2006 TED conference (<http://on.ted.com/s8BV>). The main difference compared to self-running presentations is that presenters can respond to questions.

A presentation based on a live visualization lets the presenter pause the story and interact in response to questions. The presentation can even adapt to changes made at one point that carry forward through subsequent presentation steps. In addition to the usual kinds of interactions used in visualization, the additional layer of annotation, highlighting, and so on can be especially useful in this setting.

Individual or small-group presentations

Although scenarios in which individuals or small groups need to present their research results might not seem different from the previous scenario, they do potentially involve more interaction between the presenter and the audience. This requires the presentation tool to be more flexible than a simple slideshow so that the presenter can answer the multitude of questions that might come up during the presentation. For example, in a discussion of quarterly results, questions about specific sales or marketing measures might be asked that were not part of the story but are of interest to people in the audience.

In addition to being able to ask and answer questions, it is often useful to record the kinds of questions asked so they can be reviewed later. A well-presented story is likely to lead to new questions for consideration when creating a revision or that the presenter should follow up on. In this way, the presentation becomes a vehicle not only for dissemination of information but also for collecting and condensing additional knowledge.

CONCRETE EXAMPLES

Both self-running presentations and live presentations in front of large audiences require making design choices when preparing stories that use visualization.

Gapminder

Gapminder (<http://gapminder.org>) is a type of animated presentation that uses bubble charts, scatterplots with point size representing a value (such as population) to show transitions over time. Although animated transitions

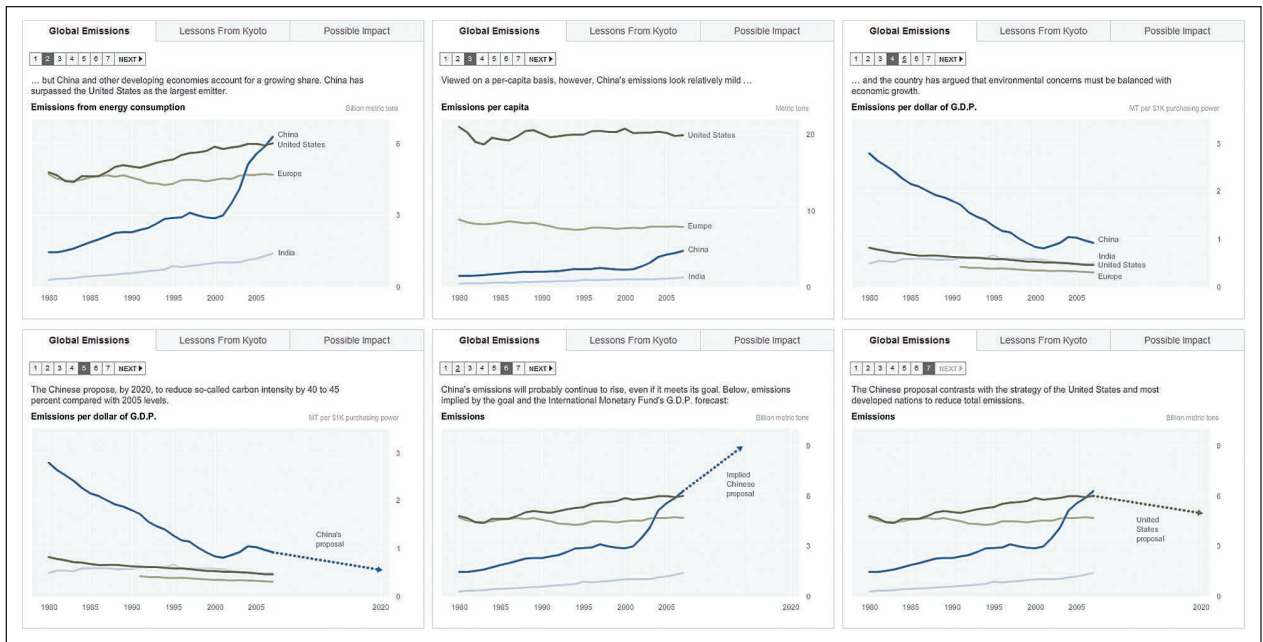


Figure 2. Excerpts from *The New York Times* slideshow “Copenhagen: Emissions, Treaties, and Impacts.” The controls let the user move back and forth between steps, with content structured like a dialogue.

have a slight detrimental effect on people’s ability to follow trends,⁹ they are both entertaining and captivating, and lend themselves well to live presentations in front of an audience.

In addition to transitions, Gapminder also demonstrates the effectiveness of building views up gradually so the audience can follow along, even when the visualization is relatively complex. Figure 1 shows an example of the use of Gapminder, with several steps first explaining a distribution and percentiles, a stacked area chart, and a scatterplot. The entire transition is quite complex, but by breaking it up into small steps and using entertaining but apt imagery, the audience can follow it with relative ease.

In the transition from the stacked area chart to the scatterplot and bubble chart, the differently colored layers—in this scenario, representing continents—turn into small circles or bubbles. The bubbles then slightly rearrange to split off the Arabic countries from Africa and Asia. An *x*-axis that represents income determines their horizontal position, and the area chart also uses the same axis. Once the bubbles are explained, a vertical axis is unrolled, which the bubbles stick to. This animation very simply but clearly explains the idea that bubble location is determined not by just one value, but two—the second one being a measure for health.

A more thorough understanding of all the different aspects of this presentation would be extremely useful to guide further development of presentation tools using visualization. In particular, does engagement help people understand data or does it get in the way? Which kinds of

animation are helpful, and which only distract? What is the tradeoff between distraction and engagement?

Slideshows

Although the slideshow metaphor is simple, it facilitates telling almost any story. For example, in a story about the 2009 Copenhagen climate conference, *The New York Times* used slideshow controls to provide a way of visualizing data (<http://nyti.ms/sFYztk>).

As Figure 2 shows, the content’s structure is compelling because it describes a relatively complex subject, with different stakeholders who have different goals and ideas about what should be done about climate change. The presentation walks the reader through those differences and also shows the results of implementing the Kyoto protocol in countries that have already agreed to it.

The story is also of interest because its comparison of the metrics that lead to different interpretations mirrors a common use case in business data. There are many ways to measure things that lead to clashing views of the same process. Understanding these differences and creating a common view is a task that a well-constructed story can support.

Using a visualization such as a line chart also facilitates simple interaction. For example, while the interaction in Figure 2 is limited to just highlighting grayed-out data values in some views, it is focused and keeps the user from straying too far from the point of the story. This makes it easy for the user to pick up the thread after any interactions, allowing the insertion of interaction points in multiple places without making the story overly complex.

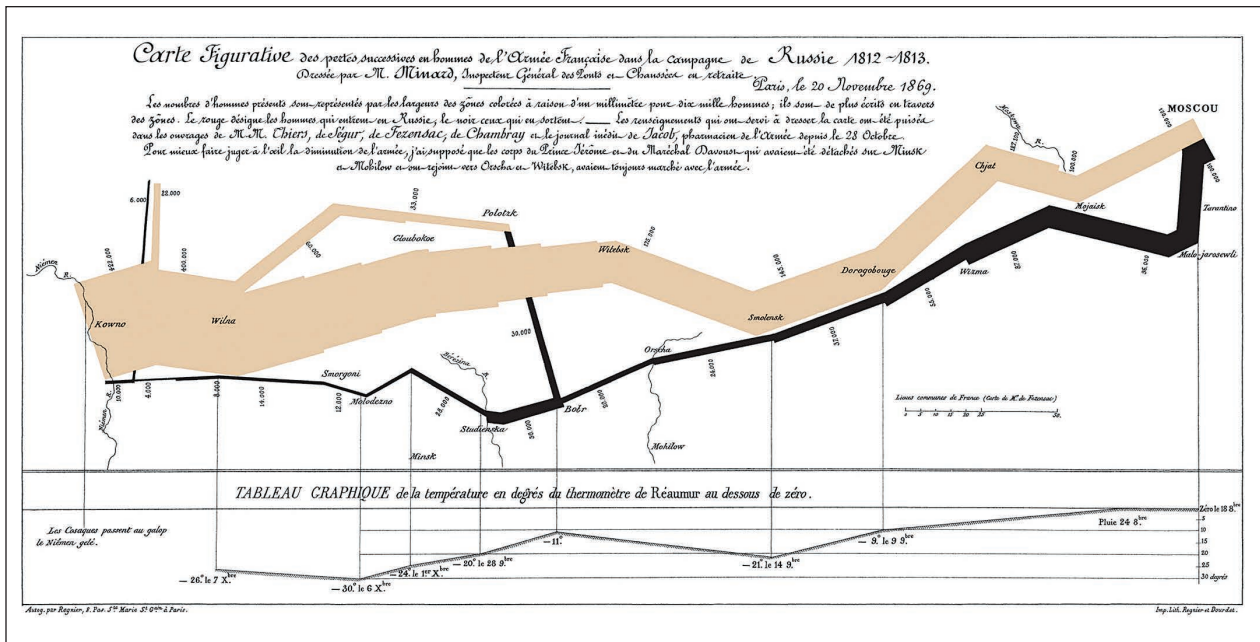


Figure 3. Napoleon's Russian campaign. Although Minard's map is often cited as an example of visual storytelling, it does not include the typical elements of storytelling, such as progression through time.

Other design choices exist, of course, but this focused example represents a pragmatic yet interesting take on the visual storytelling design space.

RESEARCH DIRECTIONS

Storytelling research in visualization straddles the boundaries of several fields, including traditional computer graphics and visualization, cognitive psychology, and the social sciences.¹²

Storytelling approaches and affordances

While Segel and Heer identified several genres and strategies, their sample was limited to newspaper stories and a particular presentation scenario.⁴ Developing a deeper understanding of storytelling strategies in visualization and examining a much broader sample of presentations will provide a richer library of approaches. It will also require a critical evaluation of the effectiveness of each of these stories.

As a starting point, we propose the concept of storytelling affordances—features of a visualization that provide a narrative structure and guide the reader through a story. A fundamental feature of stories is that they provide a temporal structure, even if it is not necessarily linear. Time is closely related to causality, so providing the causal relationships between facts and events ties the individual parts together in a cohesive structure.

Figure 3 depicts Minard's famous graphic tracking the number of men Napoleon Bonaparte lost during his ill-fated march on Moscow, which offers a particularly interesting example to study. Specifically, the graphic

depicts the size of Napoleon's army at different stages during the campaign as the width of the tan and black line. Drawn on top of a minimally styled map, the line provides both temporal and spatial information. The left-to-right direction is a natural one, making it easy for people to follow who are used to that reading direction. The connection with the temperature chart at the bottom also provides a hint regarding the primary cause of the soldiers' deaths.

In slideshows, affordances are typically obvious, but there are cases where they are less so or even missing. Understanding these affordances will make it possible to create more effective stories that can be read effortlessly while providing ample information.

Evaluation

While there are undoubtedly many interesting stories to be found in the news media, no clearly defined metrics or evaluation methods are available to measure their effectiveness. Developing these methods will require the definition of—and agreement on—goals, such as what we expect stories to achieve and how to measure that.

Currently, many researchers evaluate visualizations based on the time it takes to complete a task and the response accuracy. However these metrics are not relevant for understanding stories—meaningful story metrics include engagement and interest, ability to remember key points, information provided to make more informed decisions, and so on.

Controlled studies are often done in the lab, typically within a relatively short time frame. Story evaluation will

require a very different approach to account for various scenarios and to reflect real-world uses. The use of crowdsourcing platforms such as Amazon's Mechanical Turk, for example, makes it possible for visualization studies to reach a wider and more diverse audience than the usual student population used in lab studies.

Memory, context, and embellishments

The effects of visualization on memory have not been studied in much depth to date. Although this is understandable when it comes to analysis, effective presentations must create memories. Visualization tends to be generic and minimalist: analysts prefer techniques that work with a wide range of datasets, and adding embellishments—so-called chart junk such as images or unusual color schemes—to visualizations is generally verboten.

However, the features that set a visualization apart are exactly the ones that make it memorable. Scott Bateman and colleagues' study⁷ provided the first glimpse into this topic, but the study's design had flaws, the kinds of visualizations and infographics it used were limited, and it lacked interactivity.

The space of possible questions to ask and configurations to test is huge, so the results of studies in this area will be of immediate use to many of the people working with data.

Interaction

Interaction is one of the most important aspects of visualization. The ability not only to see the data but also to quickly change the view, add different data, and so on makes analyzing it much faster and more effective. Stories are traditionally told without interaction, and unlike analysis processes, are predefined and meant to be delivered in their entirety.

Among the attempts at interactive storytelling, computer games are perhaps the most interesting and certainly the most popular. However, it is debatable whether games are stories and not just worlds that the player can explore, similar to traditional data exploration or analysis.

Clearly, there are uses for interaction in visualization stories that do not interfere with the story arc. At the very least, opening up a visualization for interaction at the story's end provides a convenient starting point for exploration and goes beyond a simple slideshow. Pausing the story for interaction is another easy-to-imagine scenario, particularly in response to questions from the audience when using visualization in a live presentation.

There is a tradeoff between interaction and focus: the former tends to distract from the story. Stories that respond to and change based on interaction, such as by selecting a particular part of the data or asking questions that the user is interested in, are conceivable, but it is unclear how

to do this without the interaction causing some form of interference.

Annotations and highlights

Visualization is a powerful tool, but telling a story might require augmentation through some other means of communication, such as written text, audio, video, or links to more information. Moreover, guiding the user through the story might require highlighting, arrows, or other tools.

Most related research considers visualization to be entirely self-contained and independent of its use and surroundings. To make it part of a story, the visualization must fit in with a presentation's other elements. The balance between text and visualization becomes an issue when too much text takes away from the data but too little text leaves the viewer confused and unable to see the connections. Where a visualization story might be placed and how it is tied into a publication or website will influence the design decisions that go into building it.

Opening up a visualization for interaction at the story's end provides a convenient starting point for exploration and goes beyond a simple slideshow.

Learning from other disciplines

Storytelling is practiced in many disciplines, offering a huge opportunity to learn from other fields. Of particular interest are ideas from the performing arts and film, especially those subdisciplines concerned with telling stories: screenwriting, choreography, directing, and so on. Another relevant discipline is journalism, which continues to increase its focus on integrated stories that contain text, images, audio, and video.

In addition to providing access to a vast collection of knowledge, interacting with other disciplines also opens up considerable opportunities for collaboration with artists, designers, filmmakers, and journalists.

Techniques specific to storytelling

Any visualization can be used as part of a story, but some techniques lend themselves to storytelling better than others. For example, techniques such as the connected scatterplot, in which the points are connected with a line in some order, or the slope graph, essentially a single-axis pair from parallel coordinates, generally are not very useful for analysis. However, they can be quite effective as storytelling devices for specific data. How well they work depends not only on the data structure but also on actual values. If the connected scatterplot results in a large number of tangled lines, it does not provide any value, but if the values change relatively smoothly and

in slightly unexpected ways, the scatterplot provides the starting point for a story.

We are not aware of a systematic study of visualization techniques for their effectiveness as storytelling devices. Such studies could give a better understanding of the design space for visualization techniques that are useful in storytelling.

Stories and collaboration

Storytelling is an inherently collaborative activity: there is no point in creating a story if there is no audience. Little research has been done on collaboration in visualization, and storytelling could provide an interesting starting point for more work in this area. As the small-group presentation scenario demonstrates, stories naturally lead to questions, which lead to discussions, which lead to deeper analysis.

In addition to being a good way to present data, stories also offer an effective means of packaging information and knowledge in a way that is easy for another person to understand. Perhaps visual data stories will become the way to preserve information about complex data and processes and pass on knowledge in the future—not unlike they once were, before reading and writing became common skills.



IEEE Open Access

Unrestricted access to today's groundbreaking research
via the IEEE Xplore® digital library

IEEE offers a variety of open access (OA) publications:

- Hybrid journals known for their established impact factors
- New fully open access journals in many technical areas
- A multidisciplinary open access mega journal spanning all IEEE fields of interest

▶ Discover top-quality articles, chosen by the IEEE peer-review standard of excellence.

Learn more about IEEE Open Access
www.ieee.org/open-access



Storytelling promises to open entirely new avenues of visualization research. Moving from exploration to analysis to presentation is a natural progression, mirrored by the research effort focused on these steps over time. As the field becomes more mature, researchers must focus on presentation. This will prove even more crucial as visualization is increasingly used for decision making. □

References

1. M. Austin, *Useful Fictions: Evolution, Anxiety, and the Origins of Literature*, Univ. Nebraska Press, 2011.
2. S. Johnson, *The Ghost Map*, Riverhead Trade, 2007.
3. N. Gershon and W. Page, "What Storytelling Can Do for Information Visualization," *Comm. ACM*, vol. 44, no. 8, 2001, pp. 31-37.
4. E. Segel and J. Heer, "Narrative Visualization: Telling Stories with Data," *Trans. Visualization and Computer Graphics*, vol. 16, no. 6, 2010, pp. 1139-1148.
5. J. Hullman and N. Diakopoulos, "Visualization Rhetoric: Framing Effects in Narrative Visualization," *Trans. Visualization and Computer Graphics*, vol. 17, no. 12, 2011, pp. 2231-2240.
6. G. Robertson et al., "Effectiveness of Animation in Trend Visualization," *Trans. Visualization and Computer Graphics*, vol. 14, no. 6, 2008, pp. 1325-1332.
7. S. Bateman et al., "Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts," *Proc. ACM Conf. Human Factors in Computing Systems (CHI 10)*, ACM, 2010, pp. 2573-2582.
8. A. Cairo, *The Functional Art*, New Riders Press, 2012.
9. K.-L. Ma et al., "Scientific Storytelling Using Visualization," *IEEE Computer Graphics and Applications*, vol. 32, no. 1, 2012, pp. 12-19.
10. C.M. Danis et al., "Your Place or Mine? Visualization as a Community Component," *Proc. ACM Conf. Human Factors in Computing Systems (CHI 08)*, ACM, 2008, pp. 275-284.
11. W. Willett et al., "CommentSpace: Structured Support for Collaborative Visual Analysis," *Proc. ACM Conf. Human Factors in Computing Systems (CHI 11)*, ACM, 2011, pp. 3131-3140.
12. L. Zunshine, *Why We Read Fiction: Theory of Mind and the Novel*, Ohio State Univ. Press, 2012.

Robert Kosara is a visual analytics researcher at Tableau Software, Seattle, with a special interest in the communication of, and storytelling with, data. Kosara received a PhD in computer science from the Vienna University of Technology. Contact him at rkosara@tableausoftware.com.

Jock Mackinlay is vice president of visual analysis at Tableau Software, Seattle. His research interests include information visualization and human-computer interaction. Mackinlay received a PhD in computer science from Stanford University. He is a member of the IEEE Computer Society and ACM. Contact him at jmackinlay@tableausoftware.com.



Selected CS articles and columns are available for free at <http://ComputingNow.computer.org>.