Presentation-Oriented Visualization Techniques

Robert Kosara
Tableau Research

Data visualization research focuses on data exploration and analysis, yet the vast majority of visualizations people see were created for a different purpose: presentation. Whether we are talking about charts showing data to help make a presenter's point, data visuals created to accompany a news story, or the ubiquitous infographics, many more people consume charts than make them.

The techniques used to present data are mostly those used in analysis: bar charts, line charts, and so on. Although we understand them well, that understanding is based on their role in analysis. What if the presentation goals were different? How would that impact how the techniques we used? And are there techniques uniquely suited to data presentation but not necessarily as ideal for exploration and analysis?

Meet the class of visualization techniques I call *presentation-oriented techniques*. Let's turn around the usual view of visualization that treats presentation as an afterthought and instead focus on this use. These techniques are not restricted to presentation, but I consider their usefulness for presentation first and any usefulness for analysis as merely a nice bonus.

Criteria and Goals

First, we need to look at what data presentation and communication mean in terms of what we expect from the techniques. In analysis, we tend to use the same techniques everywhere because we know they almost always work. Bar charts, scatterplots, and such are great tools for that purpose, but presenting different data with the same techniques will not help make a lasting impression. Did that last bar chart you saw show the number of cell phones used or life expectancy? Which of the 15 scatterplots was the one showing correlation between income and fertility rate?

Two criteria are specific to presentation techniques, but not helpful (or even counterproductive) for analysis: *memorability* and *engagement*.

Memory is not a goal in analysis, where the user wants to switch between many different views and ask many different questions (see Figure 1). Unique, memorable charts would be more confusing and distracting than useful in that context. But presentation is all about getting a point across and making it stick. To achieve that goal, there need to be hooks your audience's memories can latch onto. The slick, clean, but forgettable standard chart types don't help with that. More unusual techniques that create memorable shapes, or that allow the use of icons to represent what the data is about, will be more memorable.

To get people to actually pay attention to a presentation, they need to find the views interesting and engaging. This is not only important in a news media setting, where a story needs to stand out among many distractions, but also when trying to keep an audience's attention during a long presentation. Learnability is part of this, especially for unusual techniques. If the way the visualization works cannot be grasped quickly, potential viewers will get frustrated and move on.

The second pair of criteria goes against the ideas much visualization research is based on. I argue that presentation-oriented techniques need to be *specific rather than general* and *compact rather than scalable*.

Of course it's important to have analysis techniques that don't turn into an unreadable mess when you apply a filter or change the mapping on one of your axes—that would disrupt the flow of the analysis. But that is not a concern in presentations. A presentation is much more deliberate and curated, so it can use niche techniques that only work in a small number of cases. For example, news media like *The New York Times* and *The Washing-*

 Published by the IEEE Computer Society

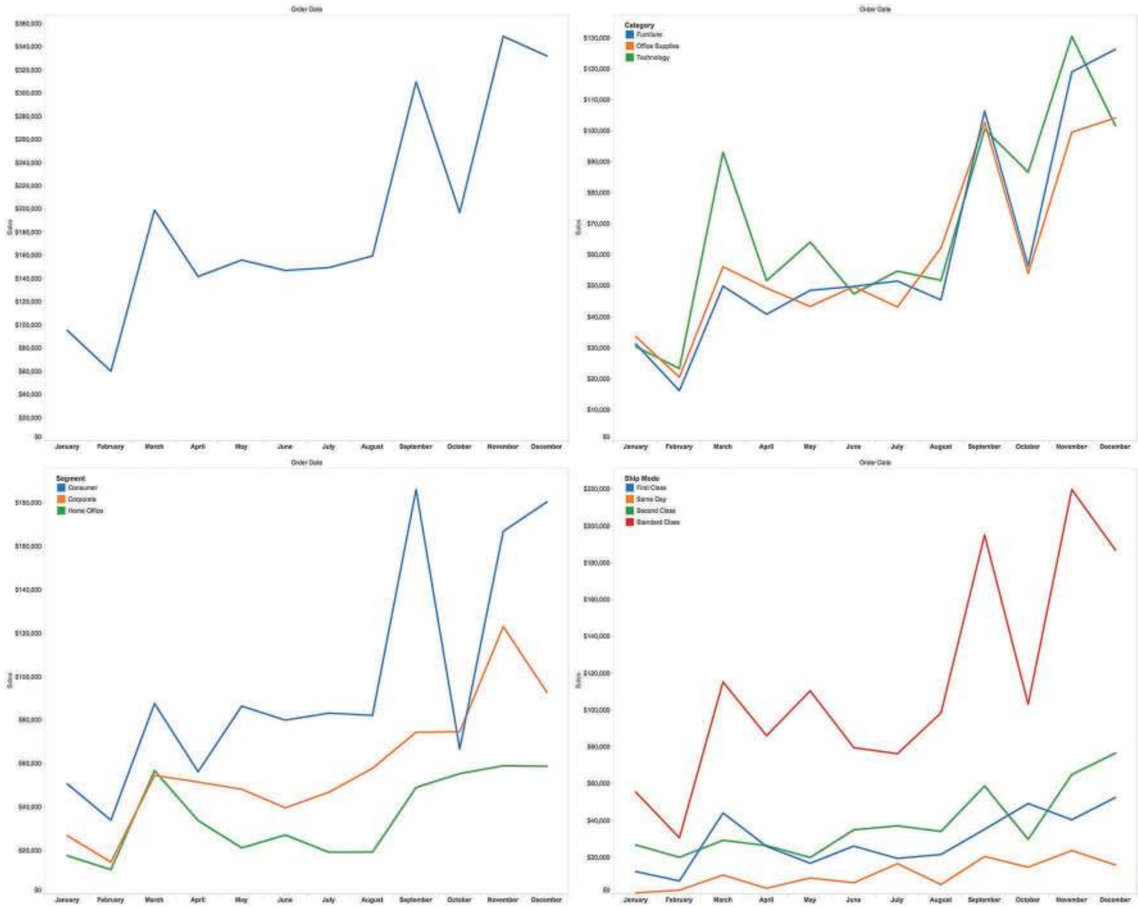


Figure 1. Example of a typical analysis session showing business data. Such line charts are effective but forgettable, which is fine in an analysis setting. The requirements are different in presentation, however.

ton Post can explore unusual and niche techniques: they want something that works for a particular dataset without worrying about generality.

Finally, presentation techniques do not need to scale as much as ones used for analysis. Although the data amounts used during analysis can be huge, results are usually presented at a much coarser level or with a much smaller subset of the data. Scaling to large data is a general problem in visualization, but most techniques can handle at least thousands of data points. Most presentation techniques do not, but that's okay; they don't usually need to. Their ability to present a few dozen or maybe hundreds of data points well is all that's needed for most use cases.

Just as certain general analysis techniques work well for certain types of questions—scatterplots for finding correlations, bar charts for ranking and comparison, and so on—presentation-oriented techniques have particular strengths that make them well suited for certain uses and datasets. Understanding these is important to turning presentation into a first-class citizen in visualization.

There is a key difference, however, between techniques that may be limited but still applicable to different data and custom-built information graphics (such as the ones Nigel Holmes created[1]). Custom designs are clearly outside the scope of visualization, whereas techniques—however limited—fit.

ISOTYPE

It's not always clear why techniques were developed, and sometimes the initial idea was really analysis rather than presentation. This is not the case with the *International System of Typographic Picture Education* (ISOTYPE), however. Otto and Marie Neurath developed it in the 1930s to communicate knowledge through data, rather than for analysis.[2]

ISOTYPE is a broad system, but its most interesting feature from a visualization point of view are the unit charts, which stack objects on top of or next to each other to represent quantities (see Figure 2). Each little symbol represents a multiple, such as 10,000 workers or 5 million heads of cattle. On a higher level, these charts can still be read as bar charts by comparing the length of the resulting bars. They are also quite universal because the objects are recognizable across language barriers

Figure 2. ISOTYPE charts stack objects on top of or next to each other to represent quantities. Our study found that the technique does not hurt reading, but does help memory and produces charts that are more engaging.[3]

and do not require viewers to read a title or legend.

In a recent paper, Steve Haroz, Steven Franconeri, and I showed that ISOTYPE charts are equal in terms of reading speed and accuracy to bar charts, and the added visual information helps people remember what they saw.[3] We did not find a single instance where an ISOTYPE chart performed worse than a bar chart.

Clearly, ISOTYPE charts are not likely to be useful in analysis. Finding good shapes to represent the different categories is a nontrivial problem. The numbers each object represents would also keep changing throughout an analysis session as the user filters the data, which would require readjusting that number. And the memory benefits are of little use when digging through data, where remembering individual charts is simply not important.

However, none of this takes away from the usefulness of ISOTYPE as a visually compelling, readable, and friendly way of representing information. It is an effective presentation-oriented technique.

The Connected Scatterplot

A technique that has seen some use in the news media in recent years, but hasn't gotten any attention from the research community so far, is the connected scatterplot. It's a simple technique at its core but can be visually interesting and effective. It also fails spectacularly for many datasets.

As the name suggests, the technique consists of a scatterplot with points that are connected in a sequence, which is usually temporal (see Figure 3).[5] This means that the two axes really represent two time series, with points that coincide in time. The resulting chart shows time along the line but does not show distance in time (only in the values), which restricts the use to cases where the samples are equidistant. This is commonly the case for data that is reported on a schedule (such as monthly, quarterly, or yearly), however, which is of interest for news pieces.

The great strength of the connected scatterplot is the interesting shape it can create, which invites closer study. It also lends itself to annotation, both of points and stretches of the line with particular shapes or directions (which all encode certain patterns between the two time series). We recently studied the technique[5] and found that users were able to read and understand the charts. They also found them more engaging than dual-axis line charts of the same data.

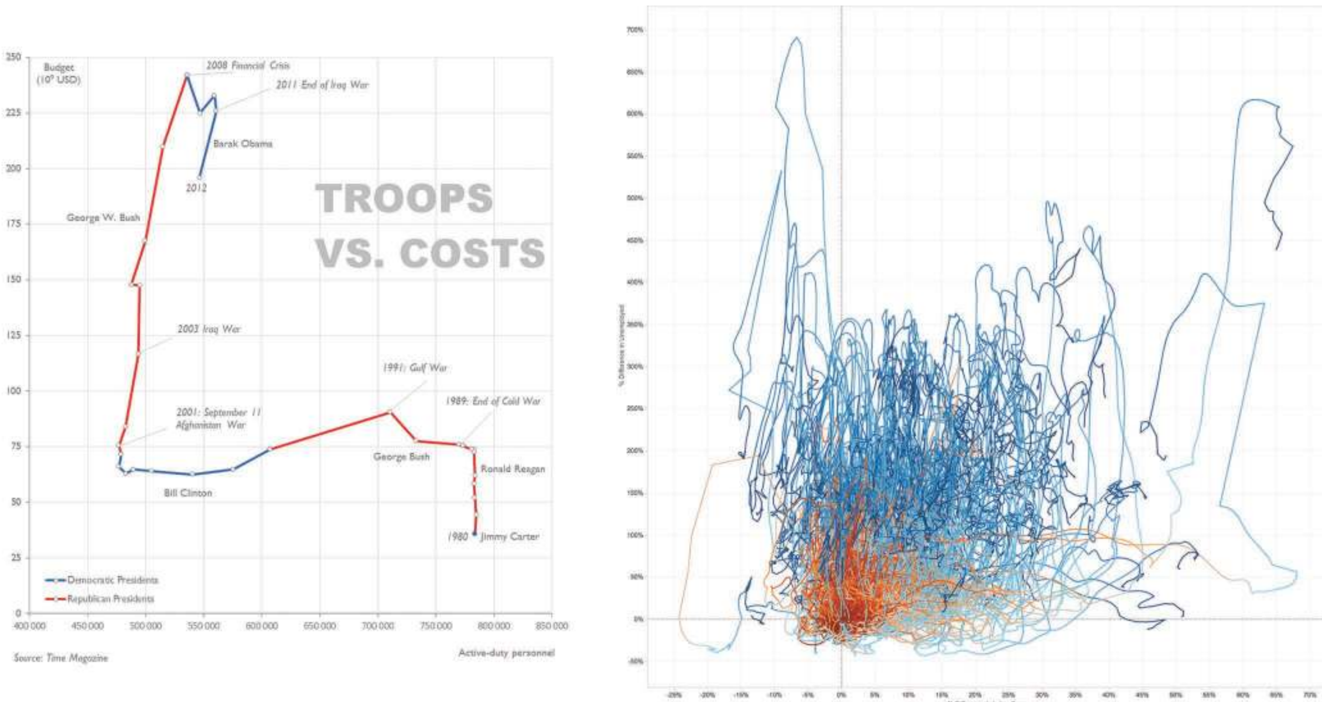
That said, the technique often fails. Tangled lines that are otherwise the domain of node-link diagrams are common, as are jaggy shapes that are hard to read and comprehend. The pay-off in the cases where it works is an engaging and interesting technique that provides for much more interesting reading than dual-axis line charts or small multiples.

ThemeRiver and Streamgraph

The ThemeRiver was published in 2001 without much fanfare.[6] Several years later, the rather similar Streamgraph caused quite a stir at InfoVis 2008 (see Figure 4).[7] Streamgraph was used in a *New York Times* piece about blockbuster movies[8] that won some of the most prestigious awards in news graphics, including *Best of Show* at Malofiej 2009, the Pulitzer Prize for News Graphics.

At the InfoVis conference, however, there was much criticism of the Streamgraph: it's a stacked area chart with many irregularly shaped items stacked on top of each other, making precise reading of any one of them impossible. But it worked to show the common shapes of the key blockbusters: a quick increase and a smoother but relatively quick drop back to almost zero. It also made for a much more interesting and exciting piece than a proven, general, but boring chart like so many area or bar charts in small multiples would have.

ThemeRiver and Streamgraph are not general techniques. I've seen implementations of the ThemeRiver that tried to deal with spiky data or many streams, and they were a disaster. The technique does not scale beyond 20 or so streams, and it does not work well when the width of the stream changes dramatically between time steps. It also isn't appropriate for streams that appear and disappear suddenly. But it does work well for some uses, where the point is more the overall shape and relative width of a small number of streams,



(a) (b)

Figure 3. The connected scatterplot technique. This example plots two time series against each other, one on each axis. (a) This can work well to show data and allow for annotations, (b) but it can also devolve into an incomprehensible hairball. Understanding the limitations of the technique helps make the right choice. (Left image courtesy of Jorge Camoes.[4])

or where the shapes of most of the streams are similar and only large differences in size are of interest. Those are perfectly valid presentation uses for this technique.

Napoleon's and Hannibal's Troop Movements

Charles Minard's depiction of the number of soldiers in Napoleon's army during the 1812/1813 Russian campaign is perhaps the most famous example of data visualization (see Figure 5, top). Often referred to as Napoleon's March, it uses a technique that is sometimes called a flow map, sometimes compared with a Sankey diagram, and some might argue that it is really a connected scatterplot.

Whatever the case may be, it is not a commonly useful technique. While the variations on the famous chart are legion (and mostly terrible[9]), the technique is not one you actually find used in many other cases. There is no question that it works well for this particular data and use, but it does not lend itself well to other datasets, and certainly not analysis.

It is so limited, in fact, that even Minard struggled to use it for another, similar map. On the same sheet as the map of Napoleon's troop movements (also dated 20 November 1869), he drew a map of Hannibal's troops crossing the Alps in 218 BCE (see Figure 5, bottom). Unfortunately, Hannibal did not conveniently travel from west to east for a neat left-to-right movement, but at an annoying

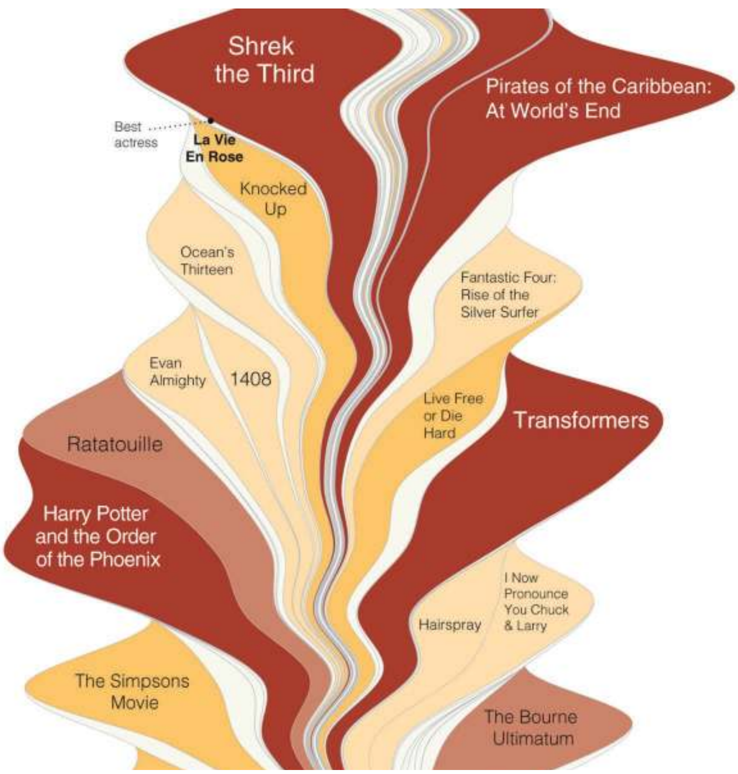


Figure 4. A Streamgraph example. The similar ThemeRiver was introduced as an analytic tool, but both have limitations that make them mostly useful when the number of streams and their shapes can be precisely controlled. This Streamgraph shows the ebb and flow of weekly movie box office revenues over time. (Courtesy of Lee Byron and Martin Wattenberg.[7])

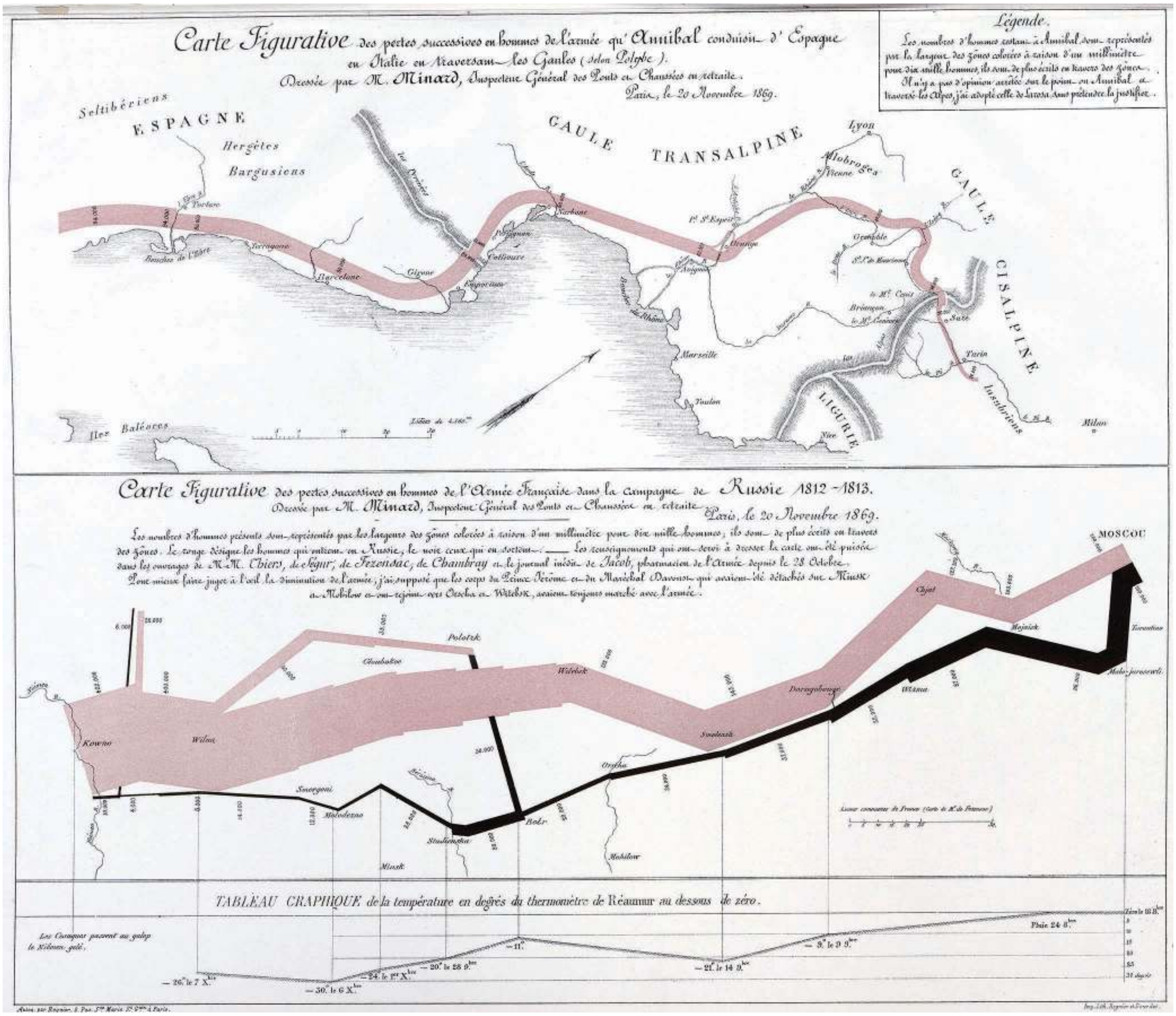


Figure 5. Minard march visualizations. (bottom) Minard's famous visualization of the troop strength in Napoleon's army is a decidedly single-use technique that does not generalize well and depends on the shape of the particular dataset shown. (top) Even Minard recognized this when creating this chart showing Hannibal's troops crossing the Alps, which required rotating the underlying map. (Courtesy of École nationale des ponts et chaussées.)

angle along the Spanish, French, and Italian coasts. Minard's solution was to rotate the map clockwise 45 degrees to create a more readable version than the more common north-is-up orientation.

Minard's map is certainly impressive, but it needs to be seen for what it is: a tool for data presentation, not analysis, of a very particular dataset (and also a historical document). It does not generalize well to other data, in particular when the direction of travel does not coincide with our common Western, left-to-right reading direction.

Presentation Contexts

Many techniques are used in presentation contexts, regardless of whether they are primarily analysis or presentation techniques. These include graphics that accompany news stories, information graphics, and informational pamphlets.

Pie charts are often used to present, rather than to analyze, data. That needs to be taken into account when evaluating their usefulness. Perhaps the pie chart's attractiveness and familiarity outweigh its low precision when a number of conditions hold, such as the number of items being low or the differences between values being large. It might get the attention that a bar chart won't.

The Voronoi treemap[10] is an interesting exercise and visually attractive, but it is much less readable than the more traditional rectangular treemap. It has been used in a number of information graphics and news stories to show part-whole data. While clearly more visually appealing and engaging than the rectangular treemap, it is also more limited in the number of data items it can show. Compared with a pie chart, it may be more readable when the range between small and large values is large because thin pie slices become even harder to judge. It is certainly much more interesting to look at and more unusual, though.

Presentation is a concern that can go beyond the pure representation of data. Periscopic's gun deaths piece adds to the actual data of homicide victims by extrapolating their potential lifespans.[11] Whatever the merits of doing this, the way the data is depicted is unique and does not fit any existing category. It does not appear to be terribly effective for analysis, but it is impressive (and memorable) as a presentation piece.

There are more techniques that are more commonly used in presentation than analysis, including Sankey diagrams, unit charts (like ISOTYPE), bubble charts, and word clouds. A technique does not exist in isolation, and it can be more or less effective depending on where and how it is used.

News graphics are fascinating because journalists seem much more willing to experiment beyond currently established ideas than most academics. The results don't always work, but the experiment is still valuable. We can learn much more from failure than from not trying.

Even the resulting ideas that do work are often limited. Rather than lamenting those limitations and ignoring the techniques, academic research in visualization would do well to study and understand them. There are undoubtedly interesting and important lessons to be learned. Limitations can also be valuable when they demonstrate issues or mechanisms that underlie visualization.

I believe that it is paramount for the academic visualization field to start thinking about presentation as an equal part of data visualization, together with exploration and analysis. Ignoring the cutting edge of visualization work that is being done outside of academia would be a mistake, and much of this work is focused on presentation.

We need to develop the tools and criteria to understand those techniques and their use cases, so we can understand how and when they work, and when they do not. Then we can come up with new ways of representing data that are effective under these criteria and guide people who want to use the best means for presenting their data and insights. We have a good body of existing work for data analysis in visualization. It's time we started to build up the same for data presentation.

References

1. S. Bateman et al., "Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts," *Proc. SIGCHI Conf. Human Factors in Computing Systems* (CHI), 2010, pp. 2573–2582.
2. O. Neurath, *From Hieroglyphs to Isotype: A Visual Autobiography*, Hyphen Press, 2010.
3. S. Haroz, R. Kosara, and S. L. Franconeri, "ISOTYPE Visualization – Working Memory, Performance, and Engagement with Pictographs," *Proc. SIGCHI Conf. Human Factors in Computing Systems* (CHI), 2015, pp. 1191–1200.
4. J. Camoes, "Chart Redraw: Troops vs. Cost (Time Magazine)," 2013; www.excelcharts.com/blog/redraw-troops-vs-cost-time-magazine/.
5. S. Haroz, R. Kosara, and S.L. Franconeri, "The Connected Scatterplot for Presenting Paired Time Series," *IEEE Trans. Visualization and Computer Graphics* (TVCG), preprint, 20 Nov. 2015, doi:10.1109/TVCG.2015.2502587.
6. S. Havre, B. Hetzler, and L. Nowell, "ThemeRiver: Visualizing Theme Changes over Time," *Proc. IEEE Symp. Information Visualization*, 2000, pp. 115–123.
7. L. Byron and M. Wattenberg, "Stacked Graphs: Geometry & Aesthetics," *IEEE Trans. Visualization and Computer Graphics*, vol. 14, no. 6, 2008, pp. 1245–1252.
8. M. Bloch et al., "The Ebb and Flow of Movies: Box Office Receipts 1986–2008," *New York Times*, 23 Feb. 2008; www.nytimes.com/interactive/2008/02/23/movies/20080223_REVENUE_GRAPHIC.html?_r=0.
9. M.-J. Kraak, *Mapping Time: Illustrated by Minard's Map of Napoleon's Russian Campaign of 1812*, Esri Press, 2014.
10. M. Balzer and O. Deussen, "Voronoi Treemaps," *Proc. IEEE Symp. Information Visualization*, 2005, pp. 49–56.
11. Periscopic, "U.S. Gun Deaths in 2013," http://guns.periscopic.com.

Robert Kosara *is research scientist at Tableau Research. Contact him at rkosara@tableau.com.*

Contact department editor Theresa-Marie Rhyne at theresamarierhyne@gmail.com.