

Cartographic Malpractice

A Review of Bis² Super Graphics

Stephen Few, Perceptual Edge
Visual Business Intelligence Newsletter
May/June 2009

We learn as much from our failures as from our successes; probably more, because there are far more of them. When I come across failures in data visualization, I try to sift and share lessons from them so you can avoid the same mistakes. In this article, I'll describe one such failure—one that is fresh.

Many of the world's greatest innovations arise from the intersection of ideas, perspectives, and disciplines. Two people can approach a problem from different perspectives and their collaboration might bring connections and possible solutions to light that were previously unknown. An architect, physicist, or electrical engineer might get involved in data visualization and imagine designs that had never occurred to those who've been working in the field for years. For this reason, last October I became excited about the innovations that might emerge from the efforts of the new *Vizbybis*² division of *Bis*². I was approached by Andrew Cardno of *Bis*², the person in charge of the company's new *Vizbybis*² business unit, with a request for my services. Andrew is a cartographer. In an email to me, Andrew wrote: "I have been following your publications on the info viz industry and your position gives me hope that, even in today's market, we can create a product that is both commercially viable and follows good info viz practices." Andrew explained that he was developing a new set of data visualizations, called *Super Graphics*, which were heavily influenced by cartographic display techniques. Cartography is the oldest form of data visualization. It has developed time-tested techniques over centuries. Possible adaptations of these techniques to other forms of data visualization seemed promising, so I agreed to review each of Andrew's new visualizations as they were developed and recommend ways to improve them.

At the outset, Andrew decided that he would produce 10 new visualizations in total. As I learned more about his plan, I became concerned that it wasn't based on the existence of 10 viable extensions of cartographic techniques. It's dangerous to base a product on an approach that works in some situations—in this case contoured heatmaps, which work for particular types of geo-spatial displays—assuming that the approach will address other problems as well. Solutions begin with a thorough understanding of a real problem and only then proceed to design and development, allowing the nature of the problem to determine the approach that is used to solve it. According to the *Bis*²'s website, Andrew was given a "mandate to create completely new ideas, new technology and new ways of doing things." Completely new ideas—those that are effective—don't emerge in response to mandates.

The more that I learned about *Super Graphics*, the more concerned I became that Andrew was trying to force a predetermined solution on a set of data analysis requirements that were already being addressed quite well by existing visualizations. Every example of a planned *Super Graphic* that I was shown used the same contoured heatmap approach. They appeared to differ only in the shape of the plot area (a spiral, rectangle, square, etc.) and in the nature of the variables that they addressed. Andrew had previously used contoured heatmaps successfully when he developed visualizations for a company called *Compudigm*, which produced software for monitoring gambling activity in casinos (*Compudigm* has since been purchased by Bally's, a gambling interest). *Compudigm*'s application was spatial, designed to show the location of activities on the casino floor. Rather than focusing on space, however, *Vizbybiz*²'s efforts were venturing into dimensions such as time, companies, and products that are less familiar to cartographers.

Contour lines work well for displaying contained regions of like values. They're commonly used on maps to display regions of like elevation, illustrated in the example below.

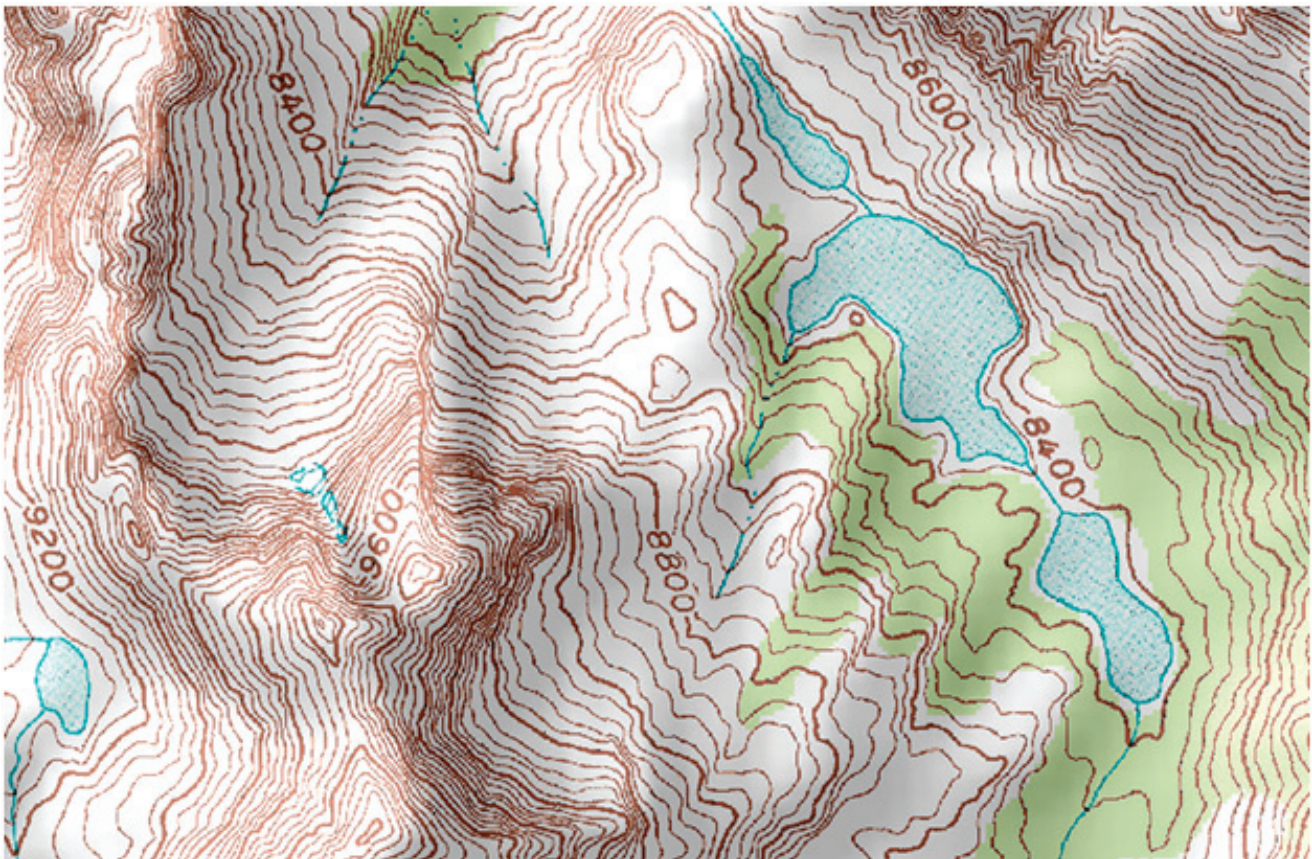


Figure 1

With relatively little effort we can spot the highest peak, which exceeds an altitude of 9,600 feet. Each region that's outlined with a contour on maps like this is either higher or lower than the region that surrounds it. The meaning of contours that define spatial regions is easy to understand, but can they also be understandably and meaningfully used to display regions of time or other categorical dimensions, such as companies or products?

Let's begin to pursue this question by considering time. Like space, time is continuous. One moment flows into the next, much as one location blends into the next, without discrete boundaries. Perhaps, just as one point in space can be combined with adjacent points by a contour to form a region of similar elevation on a map, adjacent points in time that share a common value (for example, revenues within a specified range) can be meaningfully contained within a contour as well. So far, so good, but there is a difference between space and time that must be taken into account: space is continuous in all directions, but time as we perceive it is continuous in one direction only, flowing from the past into the present on its way to the future in a straight line.

We can display monthly revenues as a linear path from left to right as shown below.



Figure 2

We can display these changing revenues graphically, using a line that moves up and down as it proceeds from left to right.

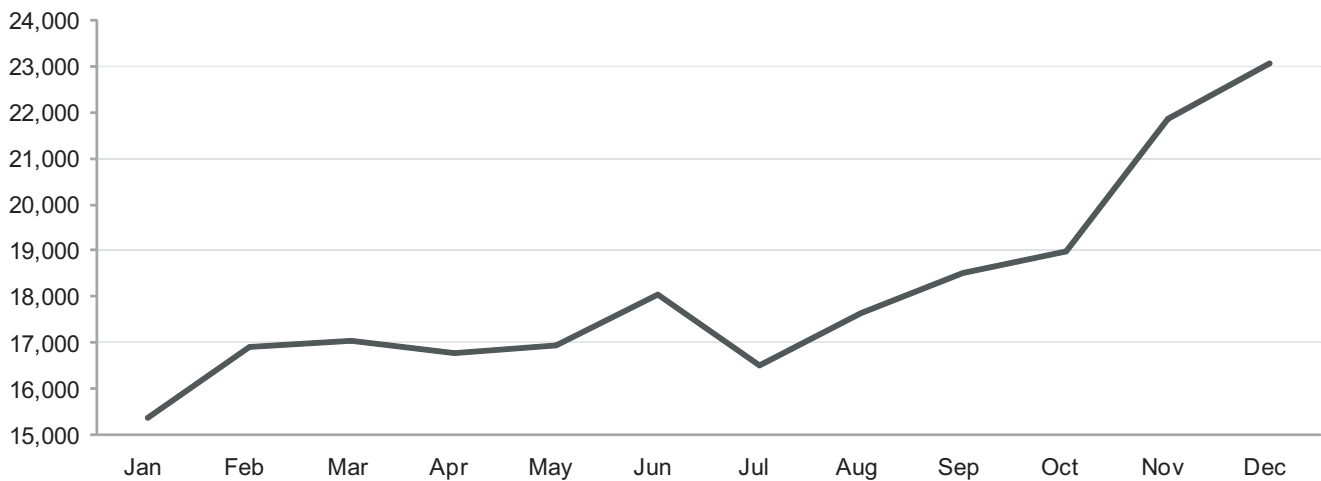


Figure 3

Can contours be used to group ranges of like revenues along this path? If so, how should they be drawn? Imagine that we want to use contours to mark monthly revenues that fall within the same \$2,000 interval (greater than \$15,000 and less than or equal to \$17,000, greater than \$17,000 and less than or equal to \$19,000, etc.). Using the linear display shown in Figure 2, the contours could be drawn as follows:

Revenue	15,384	16,934	17,038	16,774	16,953	18,051	16,502	17,655	18,525	18,977	21,854	23,052
Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec

Figure 4

Now let's color-code the ranges of like values within the contours in the form of a heatmap using the following sequence of light to dark colors:

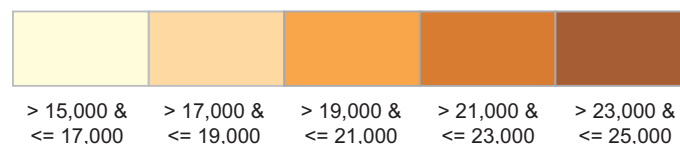


Figure 5

Our timeline now looks like this:

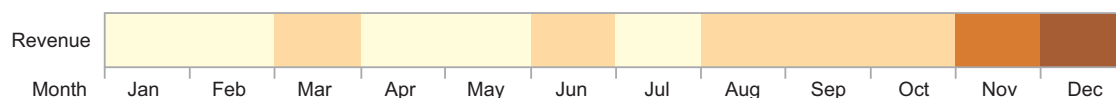


Figure 6

Now that we're using colors to represent revenue ranges, we no longer need the vertical gray contour lines, because the left and right edges of a particular contour are clearly delineated by the heatmap colors alone. Our contours have straight vertical edges that always mark the precise beginning of a particular month and the end of another. One month does not blend into another because this would indicate a smooth increase or decrease in revenue that began during the last few days of one month and continued into the first few months of the next, which is not necessarily what happened. Now, rather than a year's worth of revenues only, let's use the heatmap approach to display 25 years worth of revenues, one year per row.

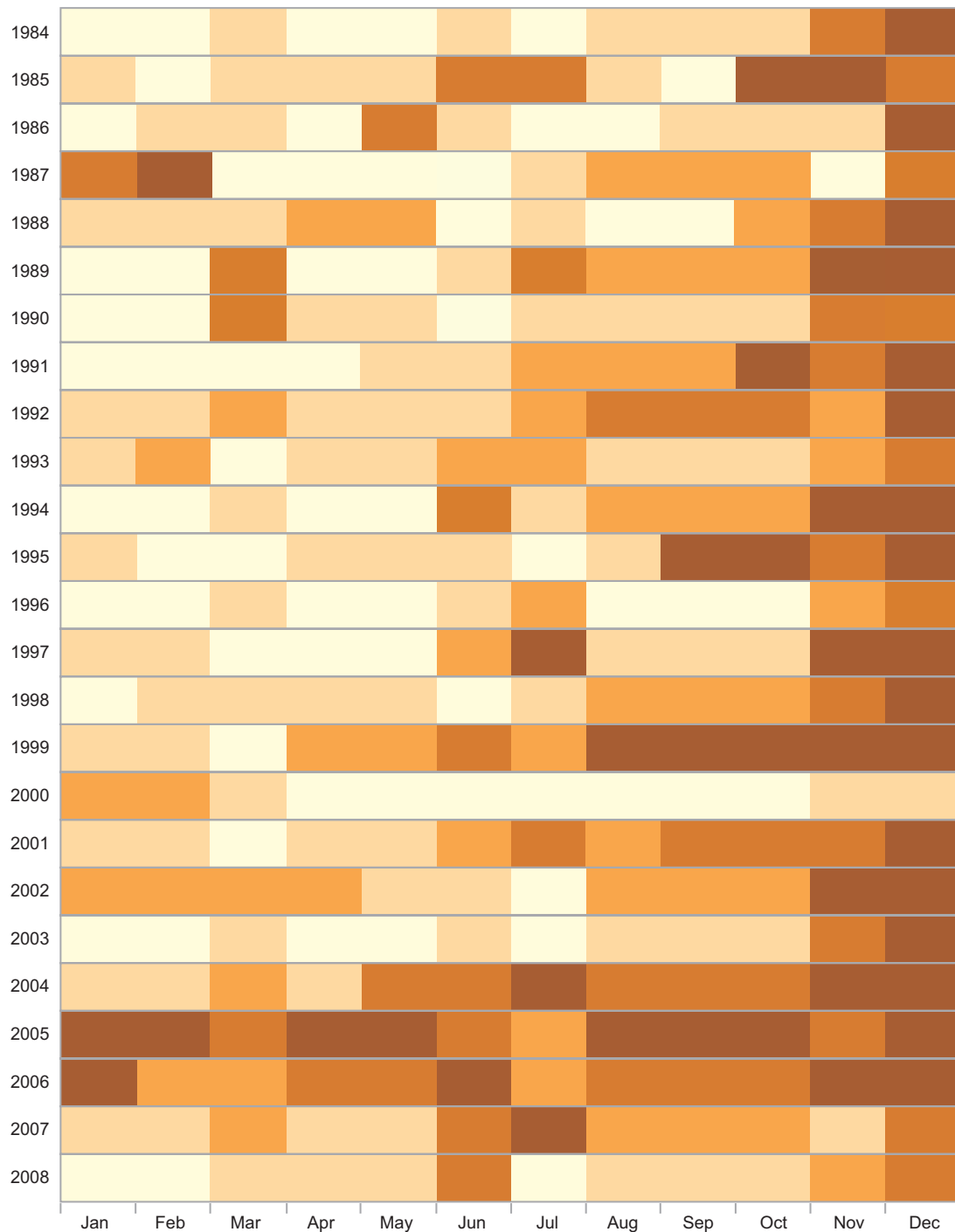


Figure 7

We're being careful to separate one year from the next using a thin gray line because revenues don't flow between years, except linearly from December of one year to January of the next. For example, it wouldn't be appropriate to blend values in May, 2007 with those of May, 2008, because time doesn't flow directly between them.

Now imagine that, rather than different years, we want to display different products, one per row. This time we'll use the following continuous range of color to encode sales revenues from \$15,000 to \$25,000, rather than five distinct colors for \$2,000 intervals.



In this display, it's especially important to delineate the rows, because products are discrete—revenues definitely don't flow from one product to another.

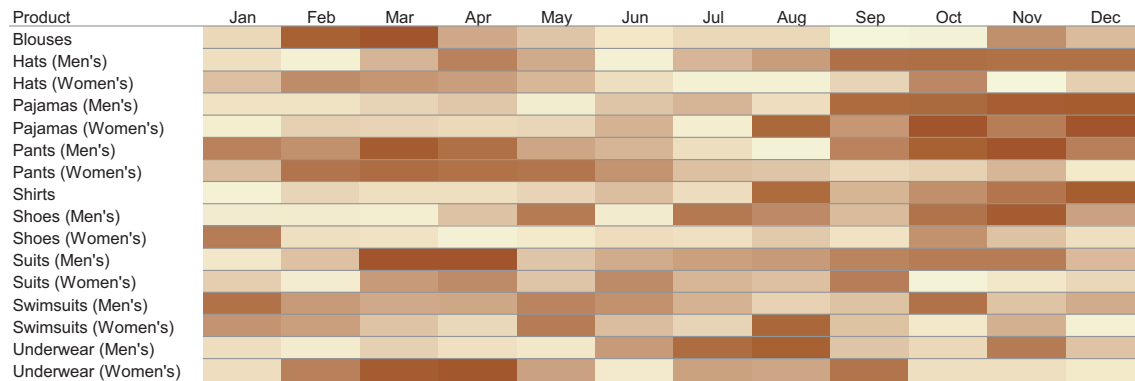


Figure 8

Because there is no particular sequence in which these products should be displayed, the overall picture of product revenues can be significantly altered by sorting the products differently from the alphabetized arrangement above, such as into separate groups of men's and women's products (see below)...

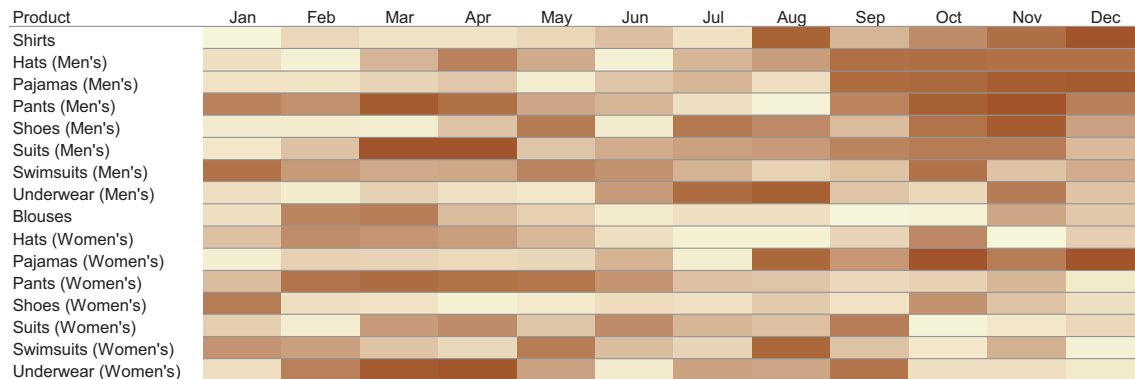


Figure 9

...or from products with the lowest to the highest revenues in the month of December (see below).

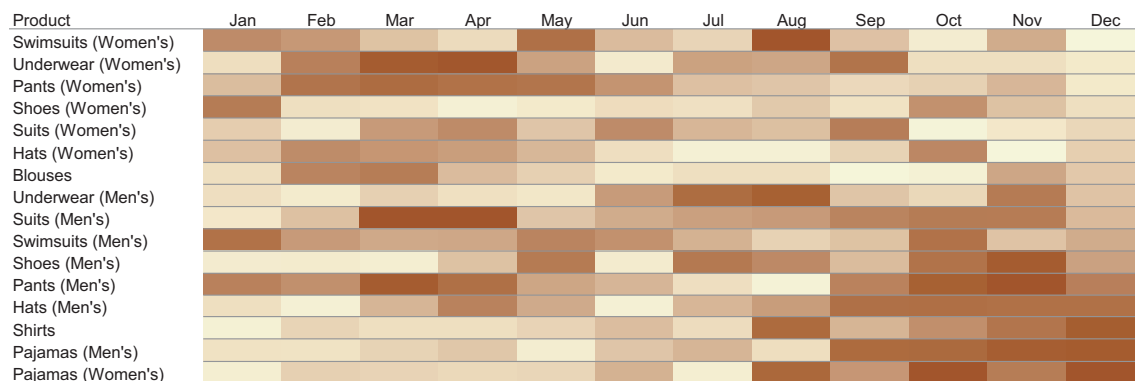


Figure 10

There is nothing new about these visualizations. They are simple heatmaps arranged as tables (a.k.a. matrices) of columns (months) and rows (products). This is easy to do in Excel, which is how I created the last three examples (figures 8 through 10) using Excel's conditional formatting to translate the values into a range of colors. When we want to examine a large set of values at once—more than we could effectively display in one or more line graphs—a heatmap matrix is an effective way to provide an overview. It doesn't support

precise comparisons between values, because we can't perceive differences in color intensity precisely, but it does provide an effective means to spot extremes (the highest and lowest values) and predominant patterns (for example, the fact that most men's products sold best in the second half of the year).

While working to produce his Super Graphics, Andrew could hardly adhere to the design decisions I've illustrated above, for that wouldn't produce a new form of visualization. Andrew took a different route. Let's take a look now at what he's designed and assess whether it's super. Here's one of the four new visualizations that have been released so far, called the *Pivotal Super Graphic*, which I borrowed from Bis2's website. It attempts to display data similar to the product revenue examples above. Each row represents a discrete item (for example, a product), time flows horizontally from left to right (in this cases labeled as weeks 1, 2, 3, etc., except that the 12th week is missing), and each of the four sections represents a larger interval of time (in this case, quarters).

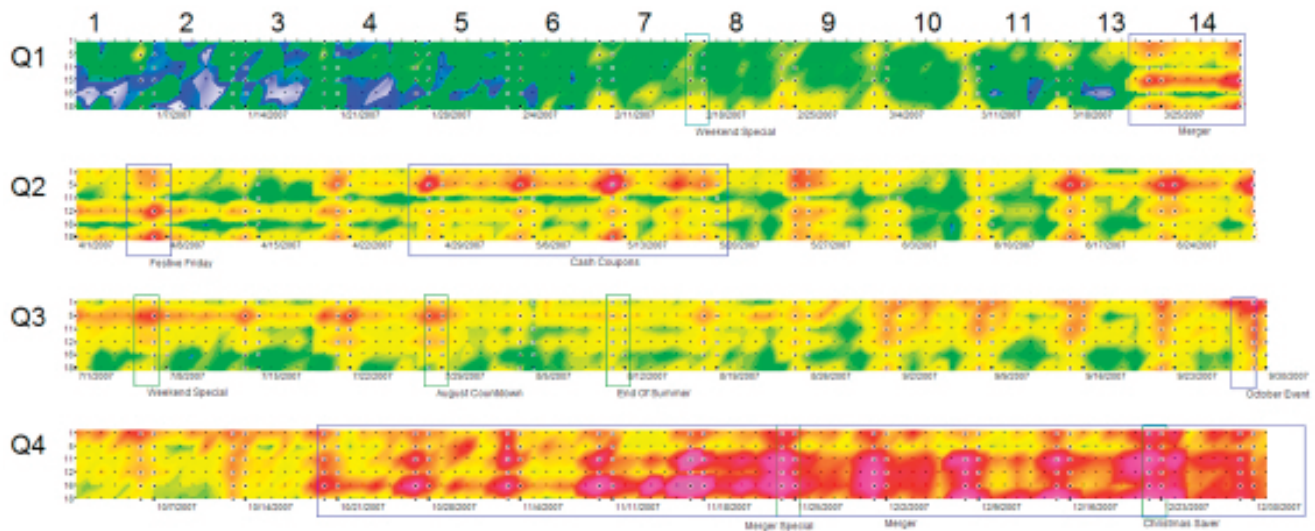


Figure 11

Colors encode the values, but a legend has not been provided (Super Graphics supposedly don't need them). Notice how the colors form irregular shapes that cut across columns and rows. I'll cut to the chase and let you know that after reviewing this particular visualization for a couple of hours per Andrew's request, I found that I couldn't proceed, so I wrote the following email.

Andrew,

I'm concerned that this visualization is flawed in such a fundamental way, it wouldn't be appropriate for me to spend an entire day critiquing it, because the changes that are required to correct its problems would produce an entirely different display that already exists: a heatmap matrix. I'm concerned that you are trying to apply visualization techniques that work for spatial data (i.e., a contoured heatmap), which do not work for many other types of data, such as time series associated with a categorical variable...In your primer you write: "The structure and nature of data may be utilized to determine how to best present information in a data visualization." I wholehearted agree with this statement, but believe that the structure and nature of the data in many significant ways did not determine how you are presenting the data in this visualization. Rather, I believe that you started with a bias for contoured heatmaps, because of your cartographic orientation, and then forced this approach on the data.

The rows in this visualization consist of categorical items, which are discrete and should be displayed as such. Your visualization, however, merges and blends values between these categorical items. This misrepresents the data...A reordering of the categorical items would result in a completely different visualization and would lead to a different interpretation. Even the blending of color horizontally through time, although more appropriate because of the gradual and continuous nature of change through time, fails to accurately display abrupt changes in value. A standard heatmap matrix would work much more effectively than the contoured heatmap approach.

I could proceed with a detailed review, but I want to give you the option of calling it off and owing me nothing for the work that I've done so far, given my concerns that a detailed critique and recommendations wouldn't be worthwhile because the visualization fails to work in a fundamental way, which can't be corrected by minor revisions. I think that I might be most helpful to you at this point by taking some time to quickly review each of your super graphics—not in detail, but at a very high level—to see if others might also have fundamental problems that should be considered before proceeding with them. Do all of your planned super graphics make use of contoured heatmaps? If so, did you determine this approach because the data demand it or because this is your default approach? I believe that contoured heatmaps are quite powerful for particular types of data and for featuring particular relationships within the data, but their application is limited.

Please let me know how you'd like to proceed.

Take care,

Steve

The Pivotal Super Graphic misrepresents the data by merging and blending values associated with discrete items that should have been kept separate. If the rows were sorted differently, the apparent patterns and their meanings would change. “Enabling human capability” is one of the mottos that Bis² is using to promote Super Graphics, but these visualizations are in fact enabling the dark side of human perception—our tendency to see patterns and find meanings that aren't actually there.

Without a legend, it isn't clear how to read the color scale. Which color is greatest and which is least? Is blue in the middle; is it high, is it low? What does light gray represent and where does it fit in relation to the other colors? Even with a legend, this color scale would not intuitively represent a sequential range of quantities. Below, I've created a separate square for each of the colors that appear in the Pivotal Super Graphic. If you were asked to put them in order from the lowest value to the highest, would the right way of sequencing them be obvious?

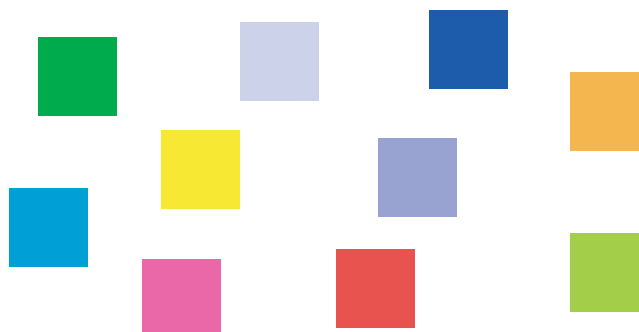


Figure 12

Intuitive representations of quantitative ranges can be created by following the rules of sequential or diverging color scales taught by another cartographer, Cynthia Brewer of Penn State University. Here are two examples of seven-color scales from her [ColorBrewer](#) application that would do the job nicely: a sequential scale above and a diverging scale below.

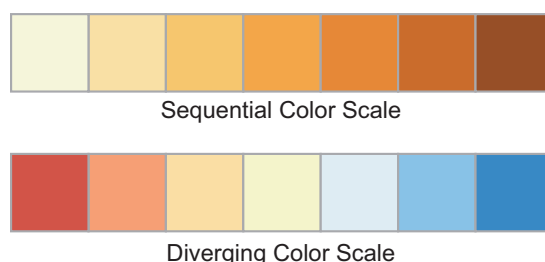


Figure 13

The diverging scale would work best if you wish to express values as below and above average: the red range for below and the blue range for above. Andrew's use of color is surprising given his cartographic training. Cynthia Brewer demonstrates what cartographers and data visualizers of all stripes should know about color.

Below is my attempt to reproduce as a line graph the data in Q1 of the Pivotal Super Graphic to more truthfully, clearly, and precisely reveal patterns of change.

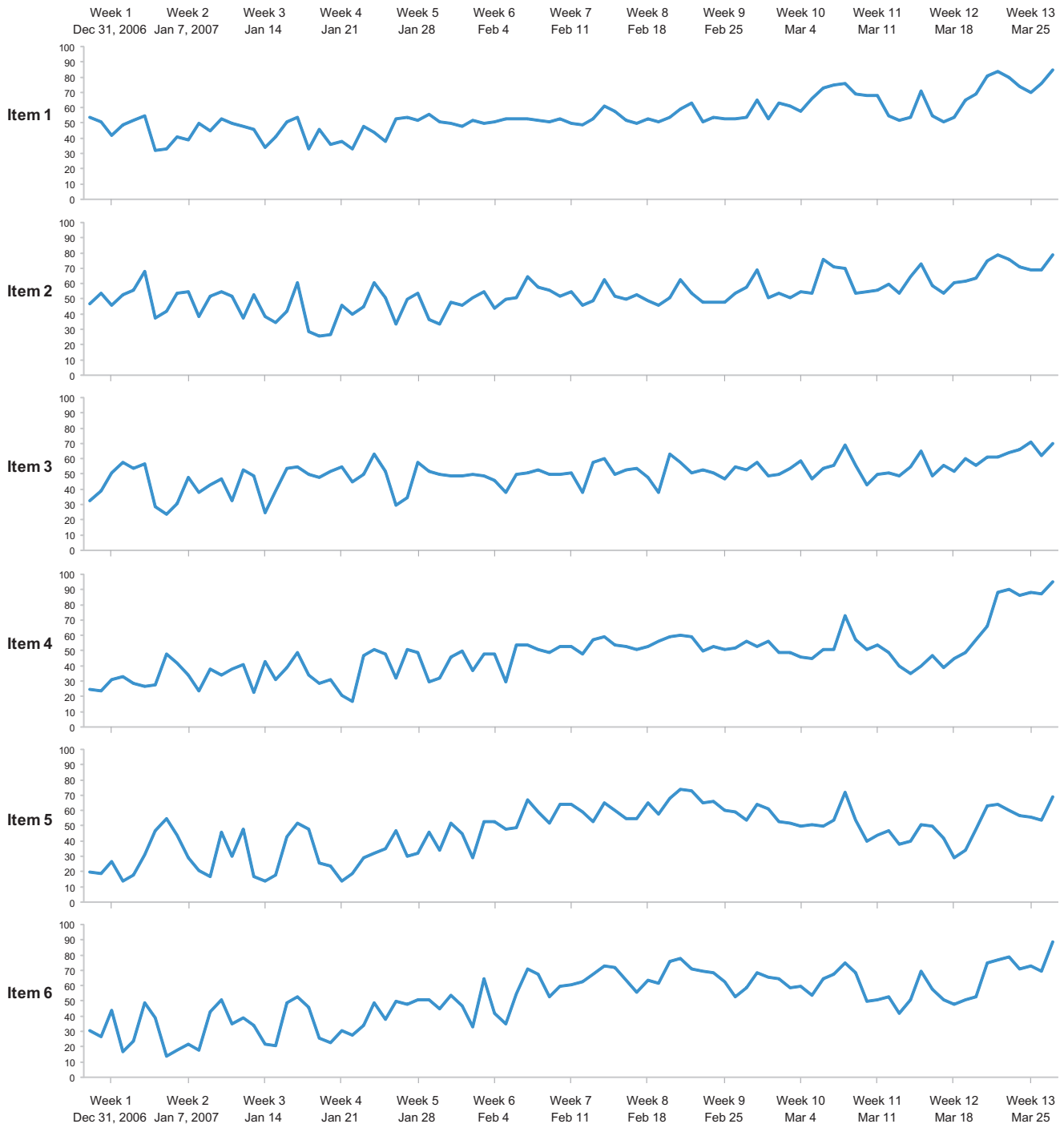


Figure 14

On the next page, I've enhanced this series of line graphs by highlighting the lowest values (those below 30) to make them easier to see and compare across multiple graphs. These are the values that appear as white

and gray in Andrew's Pivotal Super Graphic, which I assume represent the low end of the color scale. Well-designed interactive visualizations should allow us to highlight any range of values that we wish to focus on at the moment, without distraction from other aspects of the data.

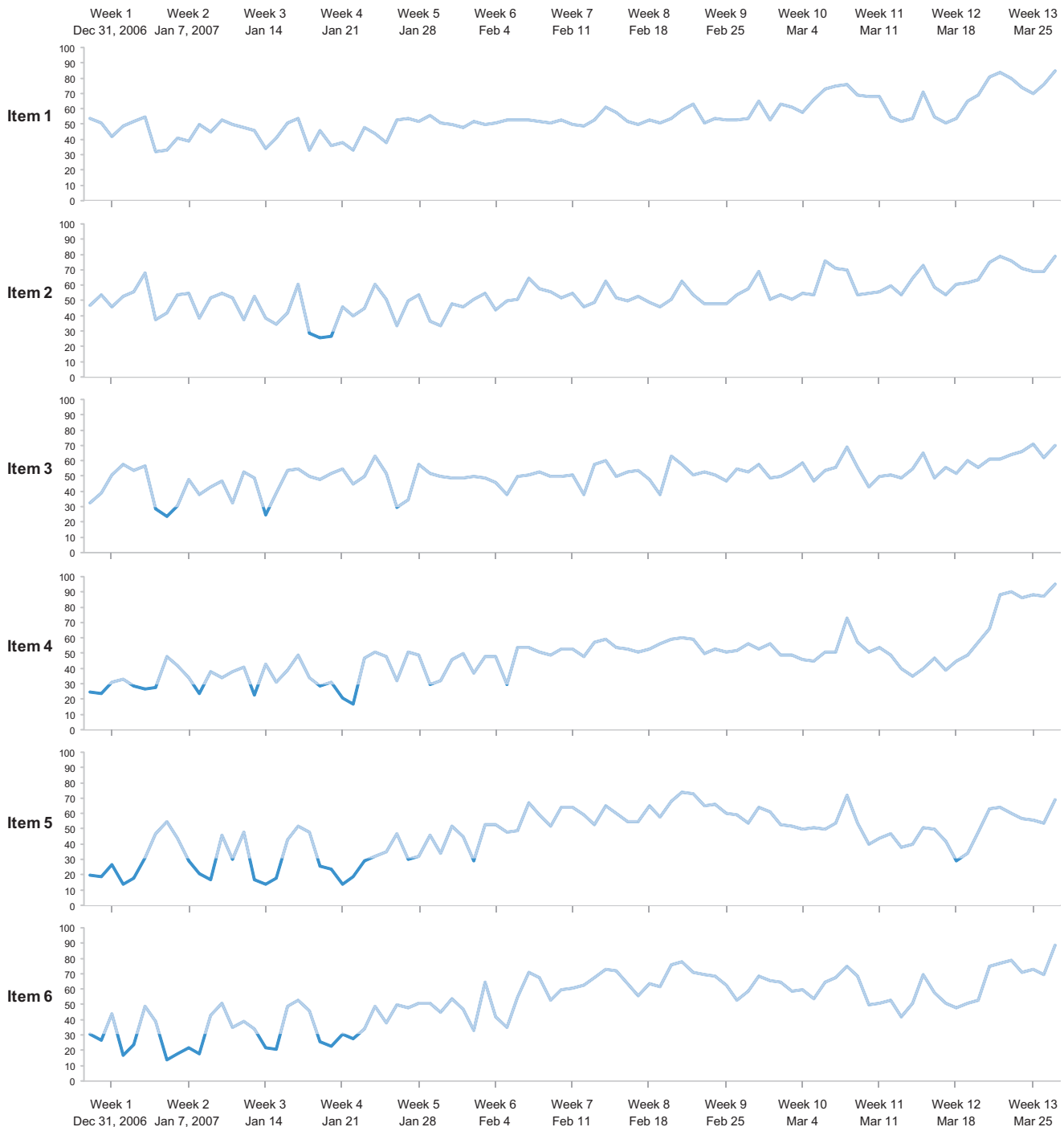


Figure 15

Only so many line graphs such as these can fit on a single screen, however, and there are times when we want to see more data. On such occasions, we could opt for a plain old heatmap matrix like the one in Figure 10, or if we have a copy of Panopticon's software, we could switch to a [Horizon Graph](#), which would allow us to squeeze a great deal of time-series information onto the screen without sacrificing readability, accuracy, and meaning.

The Pivotal Super Graphic was the second of Andrew's innovations that I reviewed. The first was called the *Temporal Super Graphic*. Here's an example from Bis2's website:

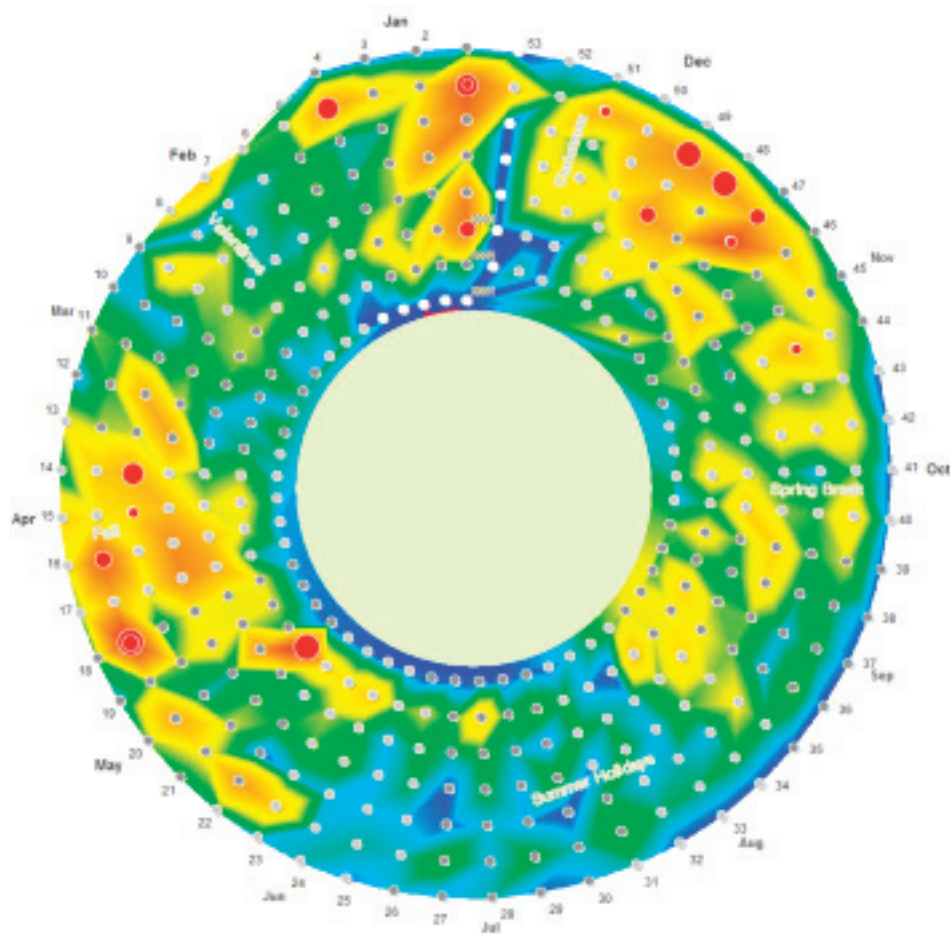


Figure 16

In addition to the problems that I cited above for the Pivotal Super Graphic, this multi-year time-series display suffers from another problem. The spiral arrangement of the display, by its very nature, causes successive cycles of time (in this case, successive years) to be displayed differently. Moving one cycle at a time from the innermost to the outermost, each is physically longer than the one preceding it. This difference affects our perception of the cycles. If the earliest and the latest years have precisely the same values, they will not look precisely the same. Larger areas of concentrated color (e.g., a region of red) in the outer section of the spiral would look greater than the same values in the inner section of the spiral, because less space would be used to encode the same values in the inner section.

Despite my critique of these visualizations, I like Andrew Cardno. Quite a lot, actually, but I can't ignore the fact that his Super Graphics are less effective by far than other graphics that already exist. Valuable new ways of visualizing data will probably emerge as extensions of cartographic techniques, but these Super Graphics don't qualify—at least not in their current form. I have little doubt that Andrew has many useful innovations within him waiting to be born, but when they are, they will arise in response to a careful study of real data visualization needs, not from an arbitrary mandate to create 10 new Super Graphics and a fixation on contoured heatmaps.

Be wary of new business intelligence (BI) solutions that clothe themselves in pseudo-scientific descriptions and eye-catching garb to convince you that they're credible. While being interviewed for a podcast by Ron Powell of the *B-eye-Network*, Andrew made a series of seemingly scientific claims about Super Graphics that are not at all scientific or true, such as "A Super Graphic has 100 times more information than a regular data visualization." Anyone familiar with data visualization will recognize this statement for what it is—a marketing fabrication.

If more BI vendors actually did real research before creating new products, their products would practically market themselves. As it is, most BI marketing departments exploit buyers' lack of expertise by entertaining them with false claims and flashy effects rather than real solutions to real problems. As recognition of data visualization's value has grown among BI vendors, its commercialization has rapidly expanded in quantity but seldom in quality. This will only change as we who use these products become educated—as we learn what works and what doesn't—and stop issuing purchase orders for the latter.

About the Author

Stephen Few has worked for over 25 years as an IT innovator, consultant, and teacher. Today, as Principal of the consultancy Perceptual Edge, Stephen focuses on data visualization for analyzing and communicating quantitative business information. He provides training and consulting services, writes the bi-monthly *Visual Business Intelligence Newsletter*, speaks frequently at conferences, and teaches in the MBA program at the University of California, Berkeley. He is the author of three books: *Show Me the Numbers: Designing Tables and Graphs to Enlighten*, *Information Dashboard Design: The Effective Visual Communication of Data*, and *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. You can learn more about Stephen's work and access an entire library of articles at www.perceptualedge.com. Between articles, you can read Stephen's thoughts on the industry in his blog.