Data Visualizations that Set the Bar Where It Ought to Be

Stephen Few
Visual Business Intelligence Newsletter
December 2006

This is the fifth and final article in a series that features the winning solutions to the 2006 Data Visualization Competition, which I judged for the Business Intelligence Network (www.B-EYE-NETWORK.com). In the fifth scenario of the competition, participants were invited to submit any original visual presentation of quantitative data to showcase their skills in visual communication.

Here's the scenario as it was described to participants:

> This is your chance to show your stuff, with few restrictions. You may submit any data visualization that you believe displays quantitative business data in a manner that effectively supports people in their efforts to understand the data in meaningful ways. Please describe in writing what the visualization means and how people would interact with it to gain understanding.

While reviewing the solutions that were submitted, in addition to clarity of communication and ease of use, I looked for those that demonstrated a high level of data visualization expertise by successfully tackling a complex problem, by exhibiting a great deal of innovation, or both.

Two solutions that were submitted for this final scenario tied for first place: one from Xan Gregg of SAS Institute, who also placed first in the overall competition, and the other from Jock Mackinlay of Tableau Software, who placed a close second in the overall competition. A third solution, submitted by Dylan Cotter of Spotfire who placed third in the overall competition, also deserves an honorable mention. Each of these solutions is quite different and together they exhibit a nice spectrum that includes many of the best designs and uses of data visualization. It also happens to be true that the three products that were used to create these solutions are some of the best products currently available for visual analysis.

The Winning Solutions

Xan Gregg of SAS Institute used his company's product JMP to create a visualization that could be placed on the Web site of a town's water authority for use by residents to look up the local reservoir's current water level and to make sense of it in light of comparisons to

"conservation levels" (explained below), 2005 levels, and levels in 2002, which was a year of drought (Figure 1).

Reservoir Status, February 28, 2006

Current Level: 92% of capacity
Status: Normal

Graph shows levels for 2002, 2005, and 2006 over shaded areas representing conservation stages based on 77-year rainfall history.

Legend:
- Normal
- Advisory
- Stage 1
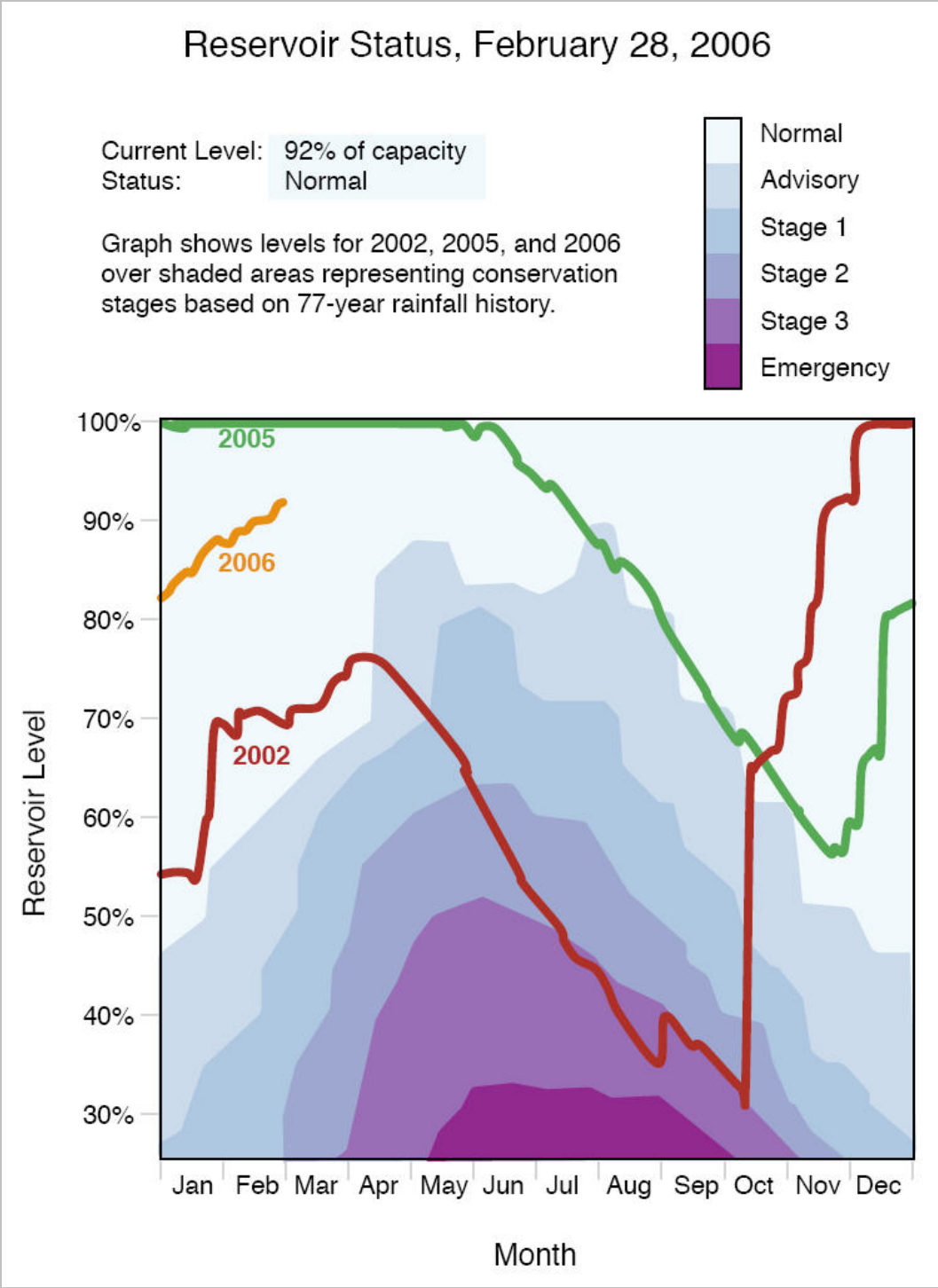- Stage 2
- Stage 3
- Emergency

Figure 1

Xan's inspiration for this visualization came from an existing display, which you can see on the Web site of the Orange Water and Sewer Authority located in Carrboro, North Carolina

(http://www.owasa.org/pages/supplydemandgraph.asp). I believe that Xan was successful in simplifying this water authority's basic message, making it much more approachable and easy to understand.

This graph uses lines in the form of a line graph to display the daily storage levels separately for the current year, prior year, and recent drought year of 2002. These lines are displayed on top of contour lines—filled areas of color, ranging from light blue to dark purple—to show increasing ranges of water supply criticality based on historical levels. Here's how Xan explains the contour lines:

> *For each combination of month and relative reservoir storage (in 5% increments), we have the number of times in 77 years the reservoir would have fallen to a critical level based on current water usage and historical rainfall data. The conservation stages are defined in terms of the probability of going critical (critical divided by 77):*
>
> | | |
> |---|---|
> | *0 - 1%* | *Normal* |
> | *1 - 3%* | *Advisory* |
> | *3 - 8%* | *Stage 1* |
> | *8 - 21%* | *Stage 2* |
> | *21 - 47%* | *Stage 3* |
> | *47% +* | *Emergency* |

Measures, such as these reservoir levels, are meaningless by themselves. They need to be seen in context, usually in the form of comparisons. Xan sifted through a great deal of data to find what he deemed to be the most meaningful context, giving residents of the town who view the graph enough to understand 2006 reservoir levels, but not so much that they would be overwhelmed or intimidated by a complicated display. As of February 28, 2006, when Xan created this graph with up-to-date data, residents could determine with relative ease that, even though levels were in the normal range so far, they were not as good as the previous year, and even though levels were heading up, they headed up during the same months in the drought year of 2002 before severely declining to critical levels. Much more than this can also be determined from this simple picture, such as the fact that in October 2002 the drought was broken with an astounding amount of rainfall in a single day, jumping all the way from a Stage 3 state of near emergency up into the normal range in a single bound.

One of the mottos of my work is "elegance through simplicity," which I believe Xan has demonstrated beautifully.

Jock Mackinley of Tableau Software used his company's product, also called Tableau, to analyze a company's weekend sales performance (Figure 2).
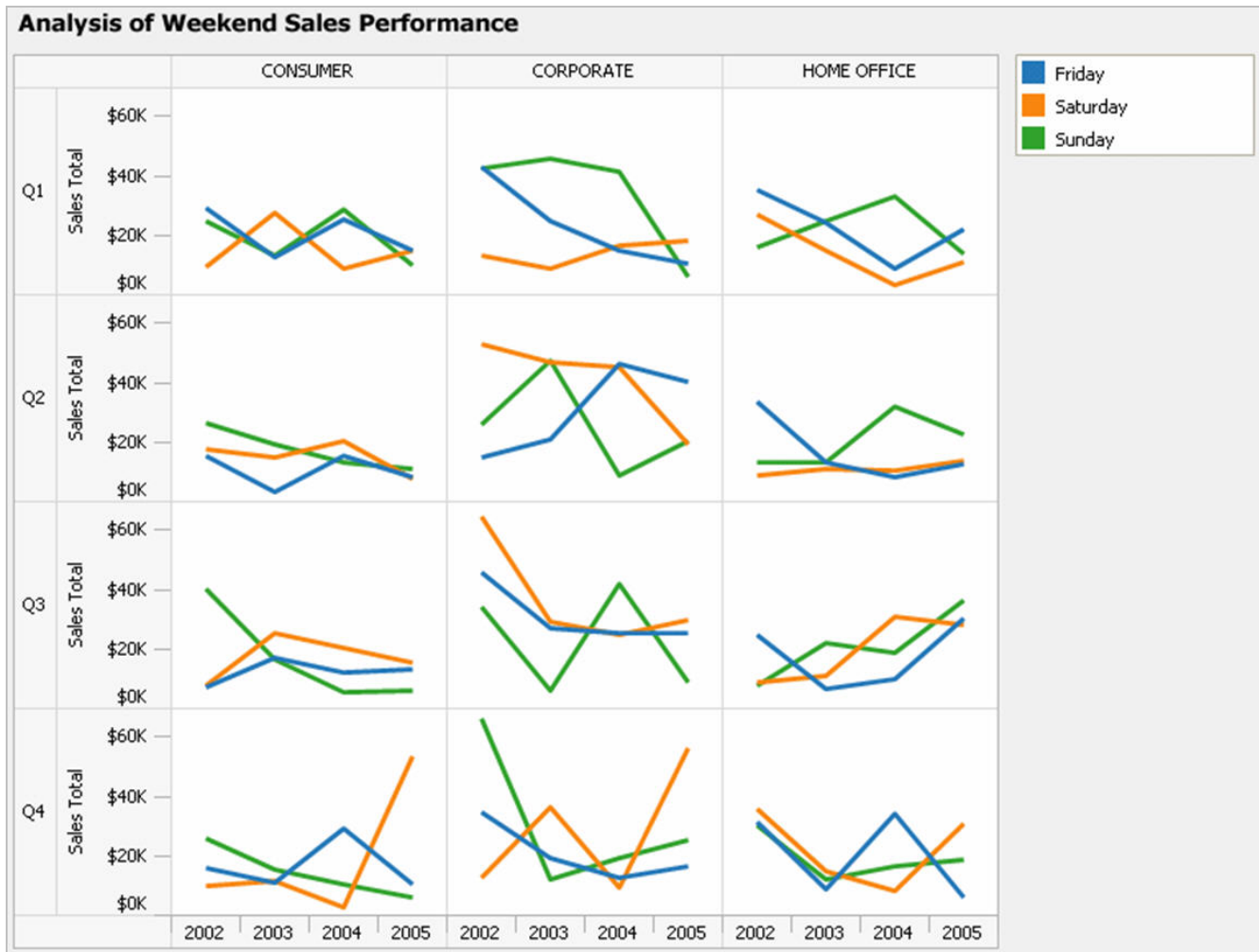
Figure 2

Here's what Jock had to say about his solution:

> *Good visualizations often display everyday data in an insightful way rather than unusual data in an elaborate way. This entry is based on everyday data (sales) and a proven visualization technique (trend lines). It's motivated by the following question: Which days of the week in which markets witnessed large 2004-2005 sales improvements?*
>
> *An important characteristic is its effectiveness at revealing and comparing trends. It summarizes sale dates at three levels of detail simultaneously, each as a separate visual attribute of a table: columns (year), rows (quarter) and colors (day). Without viewing the data in this manner, could you have noticed that Saturday sales skyrocketed in two markets* [Consumer and Corporate] *from 2004 to 2005? Could you have discovered that this "Saturday effect" was unique to Q4? Could you have compared the magnitude of these changes to other quarters and other areas of the business?*

The specific discoveries that Jock points out are only a few of the most dramatic trends that can be seen in this multidimensional display. Others can also be spotted with ease. For

instance, notice that with few exceptions, Friday sales declined from 2004 to 2005, undermining in many cases the increase in Saturday sales.

One of the important benefits derived from using software such as Tableau is not only the rich multidimensional display that can be constructed, but also the ease with which it can be constructed and then altered with each new question that comes to mind. I haven't bothered to confirm my suspicion, but I'd be willing to bet that Jock's display did not spring fully formed from his imagination before approaching the data, but was the happy product of lengthy experimentation. This is as it should be. Software that makes it easy to pursue one question after another without delay encourages us to explore.

Honorable Mention

One more solution deserves to be featured in this article. Dylan Cotter used Spotfire's new product Spotfire DXP to build a multi-view display to analyze a company's sales performance. It consists of three graphs and a table arranged on a single screen for easy comparison (Figure 3).
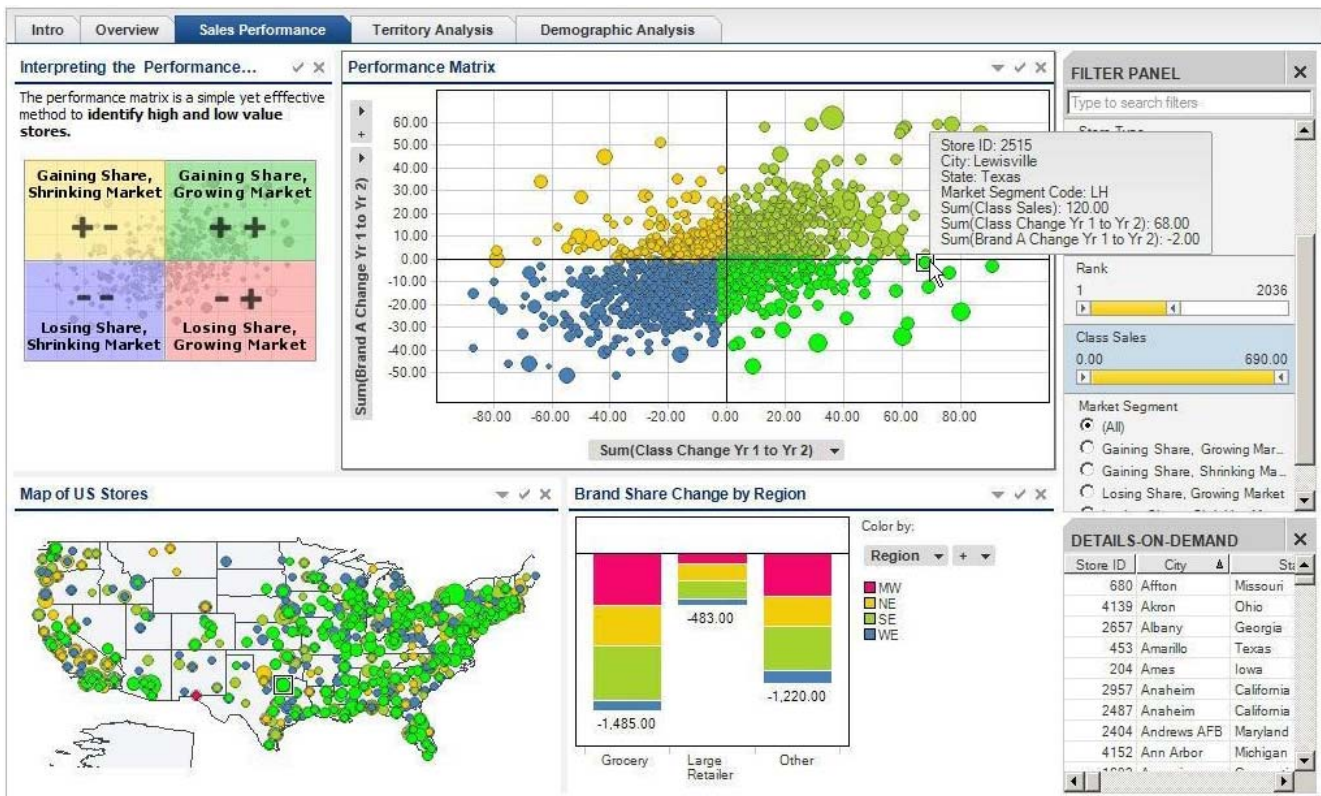

Figure 3

The screen print that Dylan supplied gives us a glimpse into functionality that is built into the software for interaction with the data, such as the pop-up details-on-demand window that must have appeared when Dylan pointed with his mouse to one of the data points in the scatterplot (upper right), and also the controls (sliders, radio buttons, and check boxes) on the right that allowed him to dynamically filter the data on the fly without taking his eyes off the graphs where the results appeared immediately.

Each of the three graphs provides a different view of the data, They work together to bring relationships to light that would be impossible to see if you could only look at one of these graphs at a time, independent of the others. For instance, notice that by dividing the stores into four categories—Gaining Share/Shrinking Market, Gaining Share/Growing Market, Losing Share/Shrinking Market, and Losing Share/Growing Market—which was easily achieved by using a scatterplot, and then by assigning distinct colors to each quadrant and using those same colors to display the stores on a United States map, it became easy to see important features, such as the predominance of stores that are losing share in a growing market (the green circles).

You might be wondering why Dylan didn't assign the color red rather than bright green to the data points that appear in the lower right quadrant of the scatterplot (Losing Share/Growing Market) to match the key to the left of the scatterplot, which labels the quadrants. Actually, he did. Although the passive nature of the screenshot does not clearly reveal the fact that a set of data points in the scatterplot were actively selected, the bright green color that we see in the lower right quadrant was assigned when Dylan drew a rectangle around those points to select and highlight them. Also, when Dylan highlighted those points in the scatterplot, corresponding points were automatically highlighted with the same bright green color on the map as well. The bar graph to the right of the map only displays data for the stores that were selected in the scatterplot—those losing share in a growing market—which allows us to easily see how declining sales in those stores are distributed into categories (grocery, larger retailer, and other) and into regions.

Spotfire DXP is one of only a few products that allow you to examine a data set from multiple perspectives simultaneously and to interact with it in ways that makes relationships between various aspects of the data visible on a single screen, thereby working around the limitations of short-term memory.

Xan, Jock, and Dylan form quite a trio of expert data visualizers. With their uncommon names, they could go on tour as the business intelligence equivalent of a rock band. Perhaps I'll ask them to open for me during my 2007 teaching tour. I would be happy to share the stage with them anytime.

_____

About the Author

Stephen Few has worked for over 20 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 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 two books: _Show Me the Numbers: Designing Tables and Graphs to Enlighten_ and _Information Dashboard Design: The Effective Visual Communication of Data_. 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.