

## The Surest Path to Visual Discovery

Stephen Few  
April 11, 2006

We use visual representations of data primarily for two purposes: 1) to search for meaningful patterns and then examine them once they're found to gain understanding; 2) to communicate meaningful findings to others. To date, business intelligence has focused on the latter, the communication of data, and is only now beginning to recognize the rich opportunities of visual discovery and analysis. Way back in 1965, the person most responsible for the emergence of visual data analysis, John Tukey, made the following prescient declaration:

*As yet I know of no person or group that is taking nearly adequate advantage of the graphical potentialities of the computer...In exploration they are going to be the data analyst's greatest single resource. (Source: John Tukey, "The Technical Tools of Statistics," *American Statistician*, 19, 1965.)*

The greatest benefits of data visualization can be found in its ability to bring important findings to light and to help us think productively and swiftly, leading to the poignant experience of "Aha, I see!"

There are many possible paths to discovery, but some are surer and faster than others. When skilled seekers venture into the world of data exploration, they tend to follow a particular path that Ben Shneiderman of the [University of Maryland](#) recognized and expressed in the form of a mantra:

*Overview first, zoom and filter, then details-on-demand.*

Shneiderman got it right. This is a sure path to discovery!

### Overview First

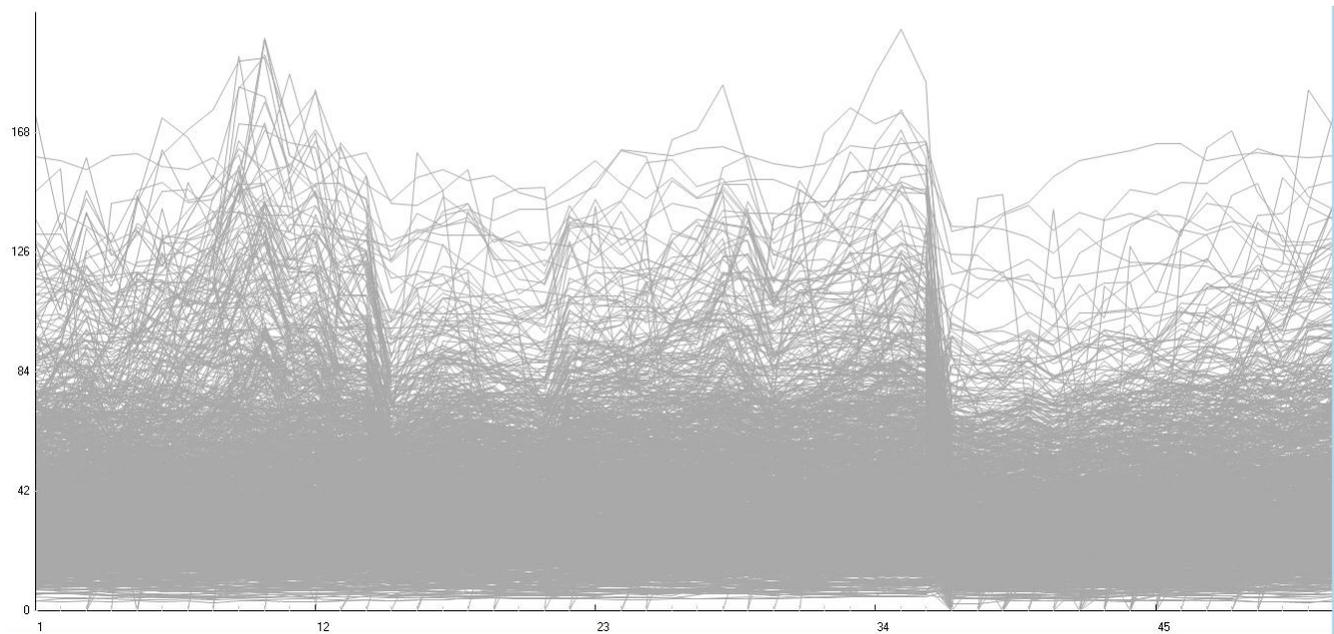
About a year ago, I was watching a BBC television program about the training of intelligence agents (a.k.a., spies). From this program, I learned that spies are explicitly taught to get a quick overview of a situation and then to focus on anything that doesn't look quite right, anything in the picture that bothers them. Scanning the big picture for points of interest is also a good first step when seeking business intelligence.

*Having an overview is very important. It reduces search, allows the detection of overall patterns, and aids the user in choosing the next move. A general*

*heuristic of visualization...is to start with an overview. (Source: Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman, *Readings in Information**

*Visualization: Using Vision to Think*, Academic Press, San Diego, California, 1999.)

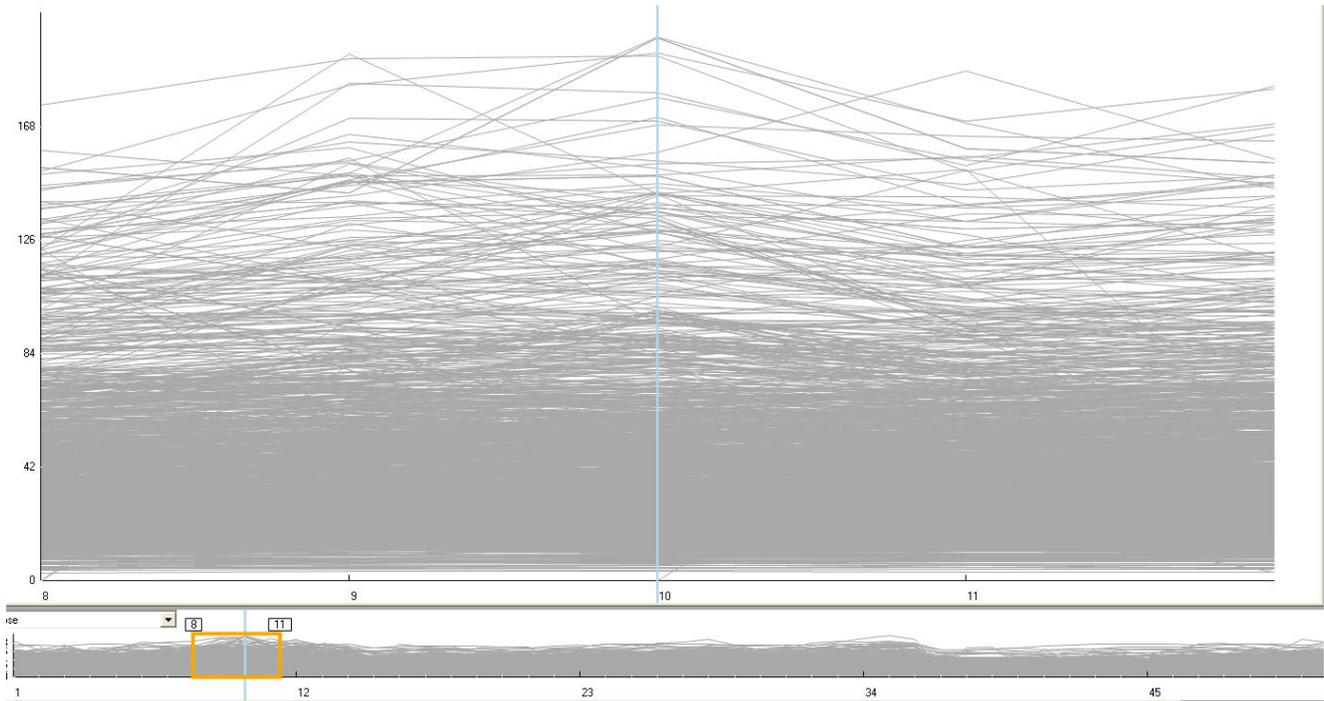
Figure 1 shows a graph in *TimeSearcher 2*, software developed at the University of Maryland, which I'll use to illustrate Shneiderman's approach. This is probably the most cluttered line graph you've ever seen. It displays the closing prices of 1,430 individual stocks across 52 weeks of time. We certainly wouldn't use a line graph with this much data to compare the performance of individual stocks, but notice how a display such as this allows you to discern what appears to constitute normal activity (closing prices roughly between \$7 and \$80) as well as exceptions to the norm. For instance, notice the interesting peak of activity around week 10.



**Figure 1:** This graph contains a line for each of 1,430 stocks. Each line visually encodes 52 weeks of closing prices.

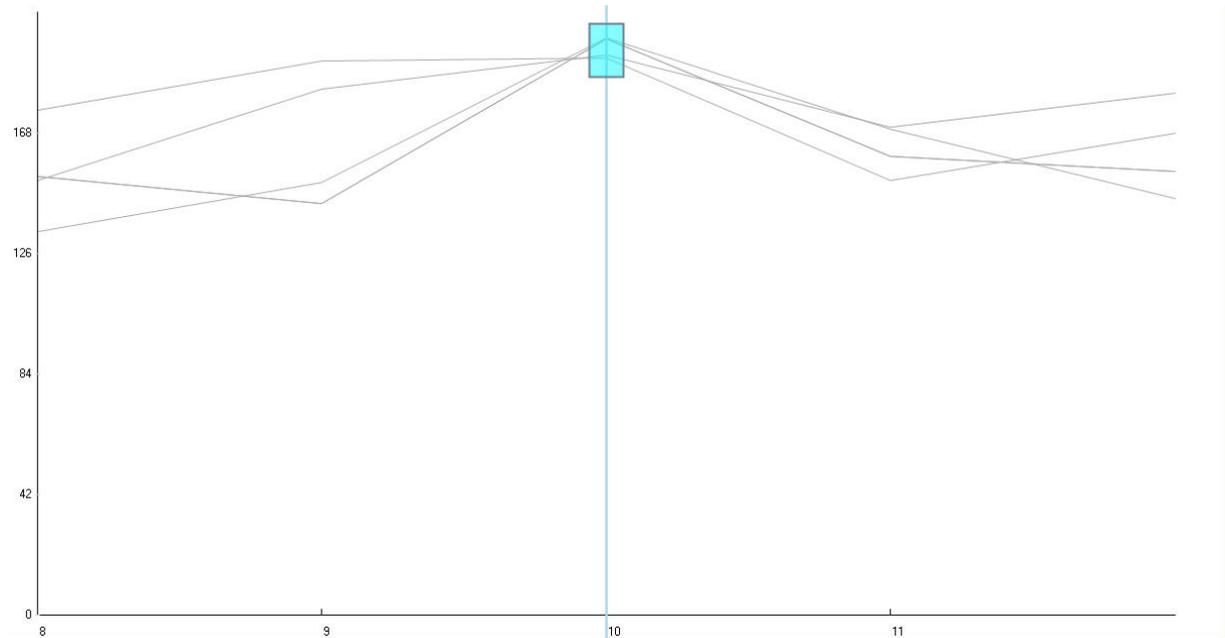
## Zoom and Filter

Once we spot interesting patterns in the overview, the next logical step is to focus on one in particular and examine it more closely. Let's zoom in on those high closing prices around week 10 and take a closer look. In Figure 2, I've narrowed the range of time from the full 52 weeks to weeks 8 through 12. Notice the orange rectangle in the smaller line graph at the bottom, which is the mechanism that I used to narrow the period of time displayed in the large graph. This method of adjusting the range of time is not only simple to use, but it also provides the added benefit of keeping us constantly reminded of where the four-week period that we're currently focusing on fits into the larger 52-week whole - concurrent focus plus context.



**Figure 2:** The larger graph above focuses on the four weeks of time that were selected using the gold rectangle in the smaller graph below.

Within this narrowed range of time, we can now explore the peak of activity around week 10 in greater detail, but we could see it even better if all but those stocks with the high closing prices were removed from view. They're distracting. Figure 3 shows the result of this filtering activity. The small blue rectangle at week 10 is called a *timebox* in *TimeSearcher 2*, which I drew around the highest closing prices to select them and filter all else from view.



**Figure 3:** The blue rectangle, a *timebox*, was used to select only those stocks with high closing prices during week 10.

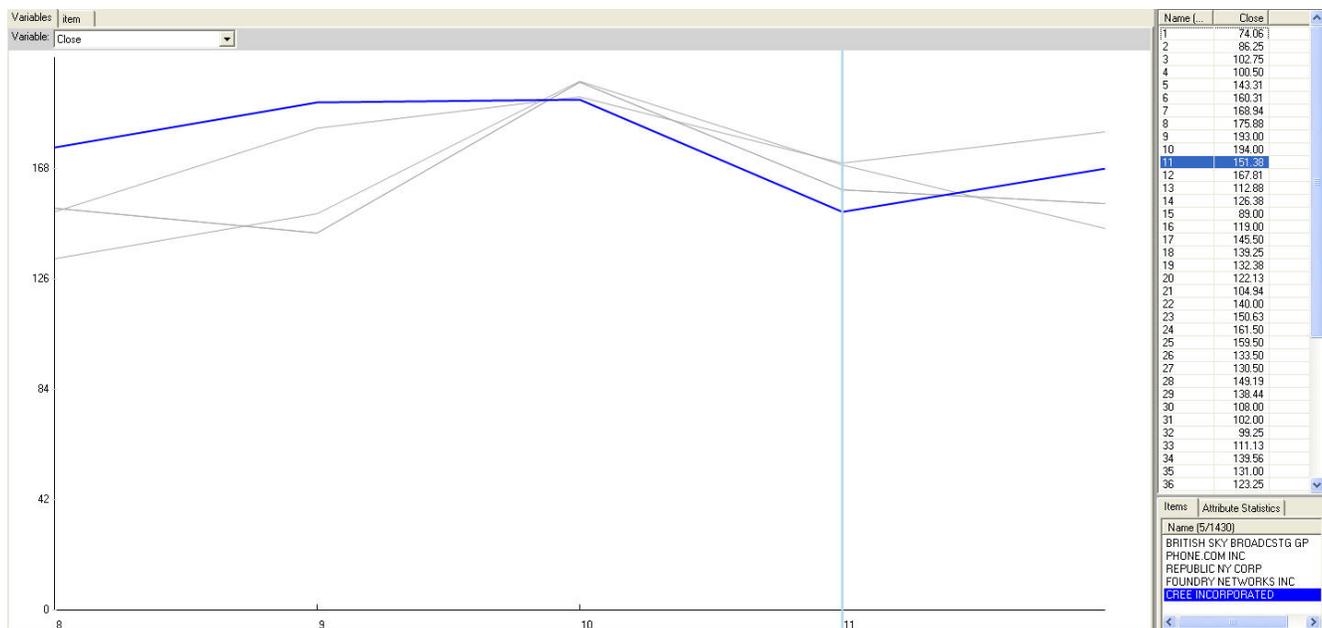
Filtering data is a critical step in the process of data analysis.

*Users often try to make a "good" choice by deciding first what they do not want, i.e., they first try to reduce the data set to a smaller, more manageable size. After some iterations, it is easier to make the final selection(s) from the reduced data set. This iterative refinement or progressive querying of data sets is sometimes known as hierarchical decision-making. (Source: Stuart K. Card, Jock D. Mackinlay, and Ben Shneiderman, Readings in Information Visualization: Using Vision to Think, Academic Press, San Diego, California, 1999, quoting a research paper by Kumar, Plaisant, and Shneiderman.)*

With only four stocks in the graph, we can now easily compare them, noticing how they behaved similarly in some cases and quite differently in other cases.

## Details-On-Demand

It is while examining the data closely that we often need to know additional details, such as the names of these stocks and their precise closing prices. Having easy access to these details only when we need them without having them in the way (interfering with the picture until we do) is valuable functionality. In Figure 4, I've expanded the screenshot to show the details-on-demand panels at the right of the graph, which lists the names of the stocks below and the precise closing prices for each week above. Because I've selected a particular line in the graph, which is highlighted in blue, and a particular point in time (week 11), those details are highlighted in the detail panels.



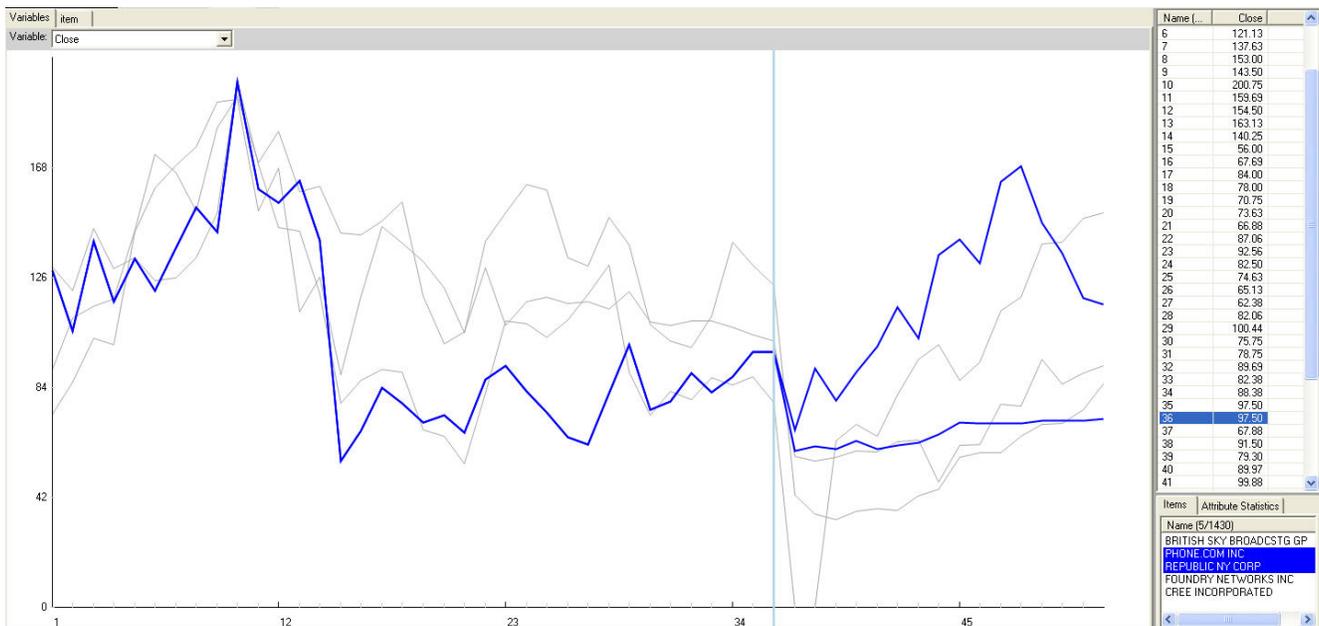
**Figure 4:** The panels on the right display details as text associated with the line highlighted as a blue line in the graph during week 11.

Using separate display panels for the text details that are always available is one of the ways that details-on-demand can be provided. Another common means that also works well is a pop-up window that appears when we click on or hover over particular items and locations in the graph, such as a particular point on one of the lines.

## Experiencing Analytical Flow

You might have noticed that five stocks are listed in the details panel, but only four lines are visible in the graph. This is odd. By selecting each stock in turn in the details panel, we discover that two of them are associated with what appears to be a single line in the graph. This invites further investigation.

Analytical navigation is often cyclical in nature. We don't just start with the overview, zoom and filter, then access details-on-demand, and end the process there. An examination of the details often prompts us to expand the picture to an overview once again. Being able to easily and quickly follow the path wherever it leads is critical to productive data analysis. In Figure 5, we can see that when time has been expanded back to the full 52 weeks, it is revealed that two stocks - Phone.com Incorporated and Republic NY Corporation - closed at precisely the same weekly prices up until week 36, where they finally diverged. We would have to investigate further to find out why these two stocks behaved identically up until week 36, but I suspect this is due to a data entry error. The person keying the data probably skipped to the wrong row of numbers for a brief period while entering the closing prices for one of these two stocks.



**Figure 5:** This graph of the entire 52 weeks shows that two stocks had precisely the same weekly closing prices through week 36 when they finally diverged.

When we explore and try to make sense of data, the software should enable us to follow the path of "overview first, zoom and filter, then details-on-demand," bouncing back and forth, here and there, with ease and without interrupting our train of thought.

*In a data exploration interface, it is important that the mapping between the data and its visual representation be fluid and dynamic. Certain kinds of interactive techniques promote an experience of being in direct contact with the data. Rutkowski (1982) calls it the principle of transparency; when transparency is*

*achieved, "the user is able to apply intellect directly to the task; the tool itself seems to disappear."* (Source: Colin Ware, *Information Visualization: Perception for Design*, Second Edition, Morgan Kaufmann Publishers, 2004.)

When software gets out of the way, invisibly and seamlessly augmenting cognition, insights can rapidly emerge. This experience of peak performance in any task is sometimes called "flow." This term was coined by the psychologist Mihaly Csikszentmihalyi (pronounced "chick-sent-me-high"), whose work I admire greatly. Shneiderman and his colleague Ben Bederson at the University of Maryland argue that "People are primarily interested in focusing on their tasks and not on operating the interface - and yet so much of a user's experience with a computer is manipulating widgets, resizing windows, and selecting from menus." With good visualization software, or any software, for that matter:

*The computer becomes a "tool" in the best sense of the word—an extension of the user's body. Time passes quickly, and the users develop a sense of control and confidence while making progress toward their goals...[Well designed interfaces] enable users to forget they are using a computer and think only of the important work they are accomplishing.* (Source: Ben Bederson and Ben Shneiderman, *The Craft of Information Visualization*, Morgan Kaufmann Publishers, San Francisco, CA, 2003.)

This is the goal. Software that enables data exploration and analysis in this manner is the key to productively mining our massive resources of data. Some software is already making it possible to closely approach this ideal. Tools such as this will produce the next gold rush in the business intelligence industry, shifting the norm from a few measly nuggets sifted from the stream here and there to entire trainloads of gleaming insight.

---

## 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](http://www.perceptualedge.com). Between articles, you can read Stephen's thoughts on the industry in his [blog](#).