

Information technology has not delivered on its promises. Yes, we live in the *information age*, and yes, much has changed—but to what end? Do you know more today than before? Are you smarter? Do you make better decisions? Too often we still make the same bad decisions, but now we make them faster than before, thanks to technology’s questionable gift of “more and faster.” This is hardly the better world that we imagined.

It is the mission of business intelligence (BI) to help us harness the power of information to work smarter. Intelligence—“*the faculty of understanding*” (according to the *Oxford English Dictionary*)—is the solid ground on which businesses must build to succeed. Information is the stuff with which intelligence works to produce the *understanding* needed to effect change, but more data delivered faster can actually lead to less understanding and even bad decisions if we lack the skills and tools needed to tame and make sense of it.

Gartner coined the term *business intelligence* in the mid-1990s and defined it as follows:

An interactive process for exploring and analyzing structured and domain-specific information to discern trends or patterns, thereby deriving insights and drawing conclusions. The business intelligence process includes communicating findings and effecting change.

The BI industry often loses sight of this clear vision. The “I” of BI—*intelligence*—can only be achieved by fully engaging the half of human-computer interaction that possesses intelligence: the human half. Business intelligence is only as effective as its ability to support human intelligence. This requires software that seamlessly interacts with our brains to support and extend our cognitive abilities. Unfortunately, BI software too often gets in the way, interrupting and undermining the thinking process rather than complementing and extending it. When BI software does its job, however, you find yourself submerged in thoughts about the data, not about the software and the hoops you must jump through to reach insight.

As a business person trying to get the job done, your only concern when dealing with data is “What does it mean?” so you can move on the whole point of the matter, which is to decide “What should I do about it?” That is it. Did you know that there are experts who spend all their time determining the most effective ways to make sense of information to produce good decisions? It is true. While you are sitting in your office or cubicle doing your job, they are sitting in research laboratories, most at universities and a few at commercial enterprises, doing their best to help you. Their findings get published, but rarely anywhere that you would ever see them.

I will do my best in this article to introduce you—people who must make sense of information to solve real business problems—to a few simple data analysis techniques for discovering meaning in your data that might otherwise remain invisible using more traditional approaches. I will use *Spotfire DecisionSite*, a popular visual analytics software package, to illustrate these techniques. All of these techniques leverage the strength of our most powerful sense: *vision*.

I See

Because vision is our dominant sense, the acts of seeing and thinking are intimately connected. It is not an accident that we use expressions such as “I see” to describe the experience of understanding. In fact, almost every word we use to describe understanding, including *insight*, *illumination*, and *enlightenment*, is a visual metaphor. Of all the sense receptors in the human body, 70% are located in our eyes. Not only does vision offer a richer, more nuanced perception of the world around us than our other senses, but it does so through a significantly broader bandwidth at much higher speeds of delivery. Researchers focus primarily on developing *visual* methods of exploring and representing information because this is the channel that can deliver the richest perception possible, resulting, when done effectively, in the richest understanding. To fulfill its promise, BI software must incorporate visual analysis methods—not just any visual analysis methods, but those that actually work.

Information visualization—technologies and techniques that support the analysis and communication of data using visual media—is not separate from business intelligence. As the Gartner’s definition of business intelligence made clear, when visualizations are used to support an “interactive process for exploring and analyzing structured and domain-specific information to discern trends or patterns”, they are doing precisely what business intelligence is meant to do. Information visualization is a powerful approach to business intelligence. When used effectively, visualization software extends the reach of traditional business intelligence to new realms of understanding—not as one means among many, but often as the only effective means available. I believe that information visualization will enable the next significant leap in BI’s evolution.

All BI software vendors have recognized the appeal to their customers of graphical data displays, but few understand their value; few understand what works and what does not, let alone why. Nothing illustrates this more vividly than the current popularity of dashboards. Like a feeding frenzy among sharks, vendors sensed the potential market of dashboards and surged in to devour as much as possible, as quickly as possible, without taking the time to understand it. Rather than stepping back and asking why these single-screen consolidations of information for rapid monitoring were so appealing, they assessed the situation superficially and began to compete with one another in an absurd display of flash and dazzle. This will change in time as the market matures and a critical mass of unhappy customers rebels against these daft dashboard displays that fail to communicate. Dashboards, like other visual media for the analysis and communication of information, have tremendous potential, but only when properly designed to connect and interact with our eyes (how we see) and our brains (how we think).

To harness the power of visual analysis and communication, BI vendors must take the time to do two things that they often overlook:

1. Focus on the useful human needs that their technologies should support.
2. Become familiar with the findings of information visualization research, which will teach them how to address these human needs in ways that actually work.

Given the central goal of BI to help business people make sense of information to enable smart decisions, BI vendors must become experts in the visual techniques that alone can enable much of the data discovery and analysis that businesses require.

Even the federal government has recognized the potential of visual analytics. A new government program is establishing regional research centers at universities throughout the country to develop visual techniques and tools for exploring and making sense of data (see National Visualization and Analytics Center). Funded primarily by the Department of Homeland Security (DHS), but available for participation by all government agencies, these research centers are bringing together academic, commercial, and government talent to leverage the power of visual analytics for greater insight by those who make the decisions that affect our national interests. Their motto is “detecting the expected...discovering the unexpected.”

Simple Information Visualization Techniques for Business Intelligence

Information visualization research has produced many useful techniques for discovery and analysis. I will introduce three of these techniques in this article using *Spotfire DecisionSite* to demonstrate them:

1. Filtering data directly and instantly
2. Viewing data from multiple coordinated perspectives simultaneously
3. Highlighting subsets of data simultaneously and automatically in multiple views

Filtering Data Directly and Instantly

One of the most common steps that you take repeatedly when examining and trying to make sense of data is *filtering*. Whenever you see something interesting, you need to take a closer look at it without the distraction of other data, so you filter out the information that is not relevant to your immediate interest. This is part and parcel of the data analysis process—nothing new—but the way that filtering is done with most traditional BI software is clumsy and tends to interrupt the free flow of analysis.

Using traditional software, the process typically is as follows:

1. You examine your data in a graph and notice something interesting.
2. You turn away from the data to construct a query to filter the data and then you run the query.
3. You wait for the query to return the new results.
4. You construct a new graph to view the filtered data.
5. You return to the data to examine it in its new form.

This fragmented process introduces an annoying stutter into the thinking process and can act as a disincentive to asking new questions of the data. It slows you down and interrupts your train of thought.

Wouldn't it be helpful if you could filter the data in a way that did not require you to take your eyes off of it, thereby allowing you to see the effects of the filtering process as it is happening, without delay? Information visualization researchers recognized this need several years ago and have worked hard to provide solutions, which they call *dynamic queries*.

Let us look at this in the context of a real-world analysis of business data. Imagine that we are responsible for making sense of our company's sales. We sell five different wines to retailers and have a particular interest in the sales of Merlot. We have the data to examine sales of Merlot across two years compared to overall wine sales, distributed across four regions (WE, MW, NE, and SW), thirty territories (1-30), and three customer types (large, medium, and small), to name a few of the available items. Now, let us focus on Merlot's percentage of overall sales in each of the two years across the four regions by displaying it as a simple bar graph (see Figure 1).

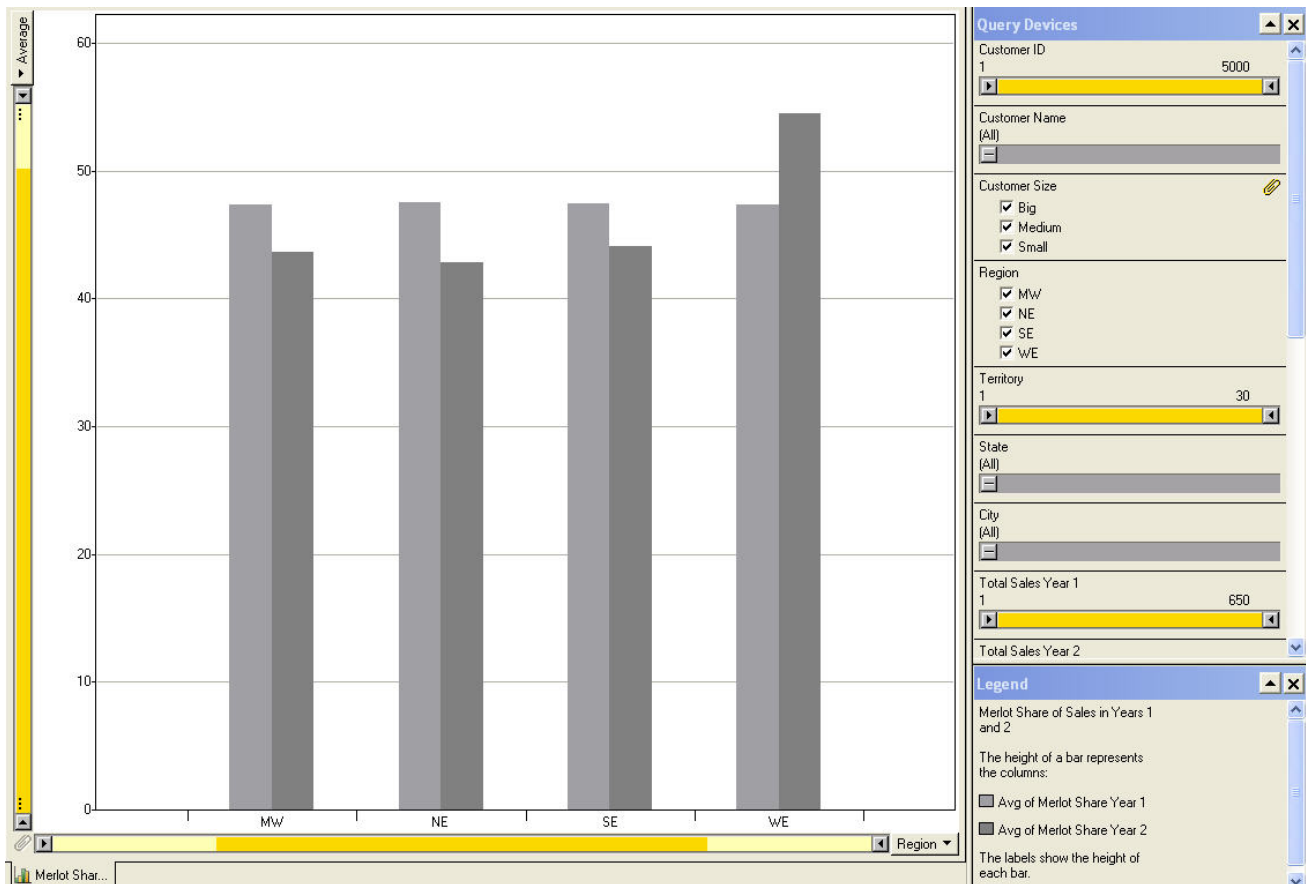


Figure 1

It only takes a second to notice that Merlot's share of overall sales has decreased from year 1 to 2 in all but region WE (west). This leads us to wonder if this pattern holds true for all types of customers—big, medium, and small. Notice that to the right of the graph there is a panel labeled "Query Devices." This panel includes simple mechanisms for filtering the data, one for each column that is available (Customer ID, Customer Name, Customer Size, etc.). Let us

focus for now on the query device for the *Customer Size* column. Because Customer Size is a categorical variable—one that consists of a limited set of discrete rather than continuous values—the query device is in the form of a check box for each of the values. Right now, each value is checked; but if we want to quickly see Merlot’s share of sales for big customers only, we can easily uncheck medium and small to view the result (see Figure 2).

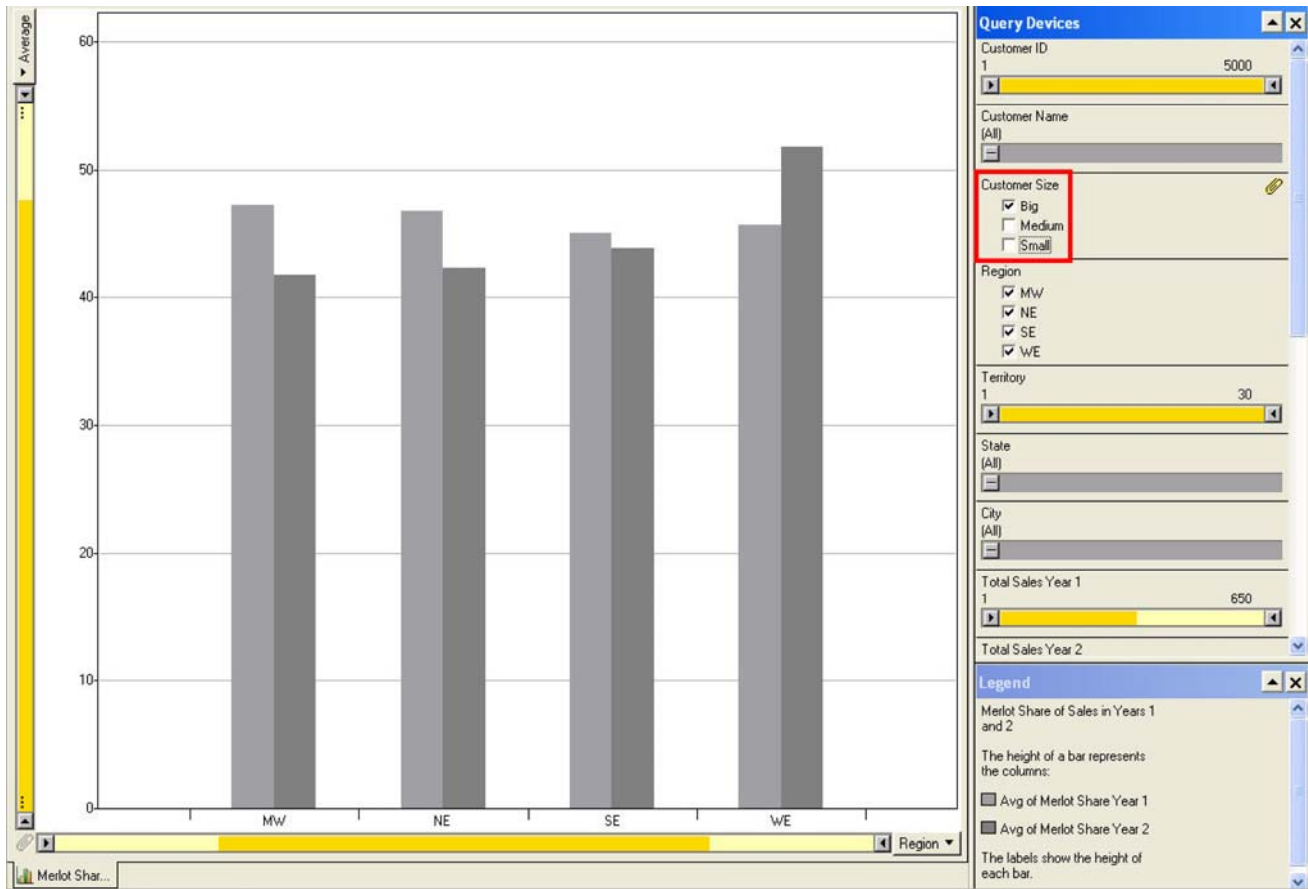


Figure 2

The general pattern holds true for big customers, but the decrease of Merlot’s share in the SE (southeast) region is not as pronounced. What you could not see, because you are reading my explanation of this process rather than doing it yourself and seeing what happens, is that there was no disconnection between wondering what the pattern might look like for big customers alone, filtering the data, and seeing the result. It involved a fluid interaction of thought and action.

Let us pursue this line of investigation further by filtering out all but the west region to see if all states in this region display a similar pattern. Because we are focusing on the change in Merlot’s share from year 1 to year 2, we will change the graph to show a single bar per state that measures the amount of change, rather than a separate bar for each year. Figure 3 shows the result of those changes.

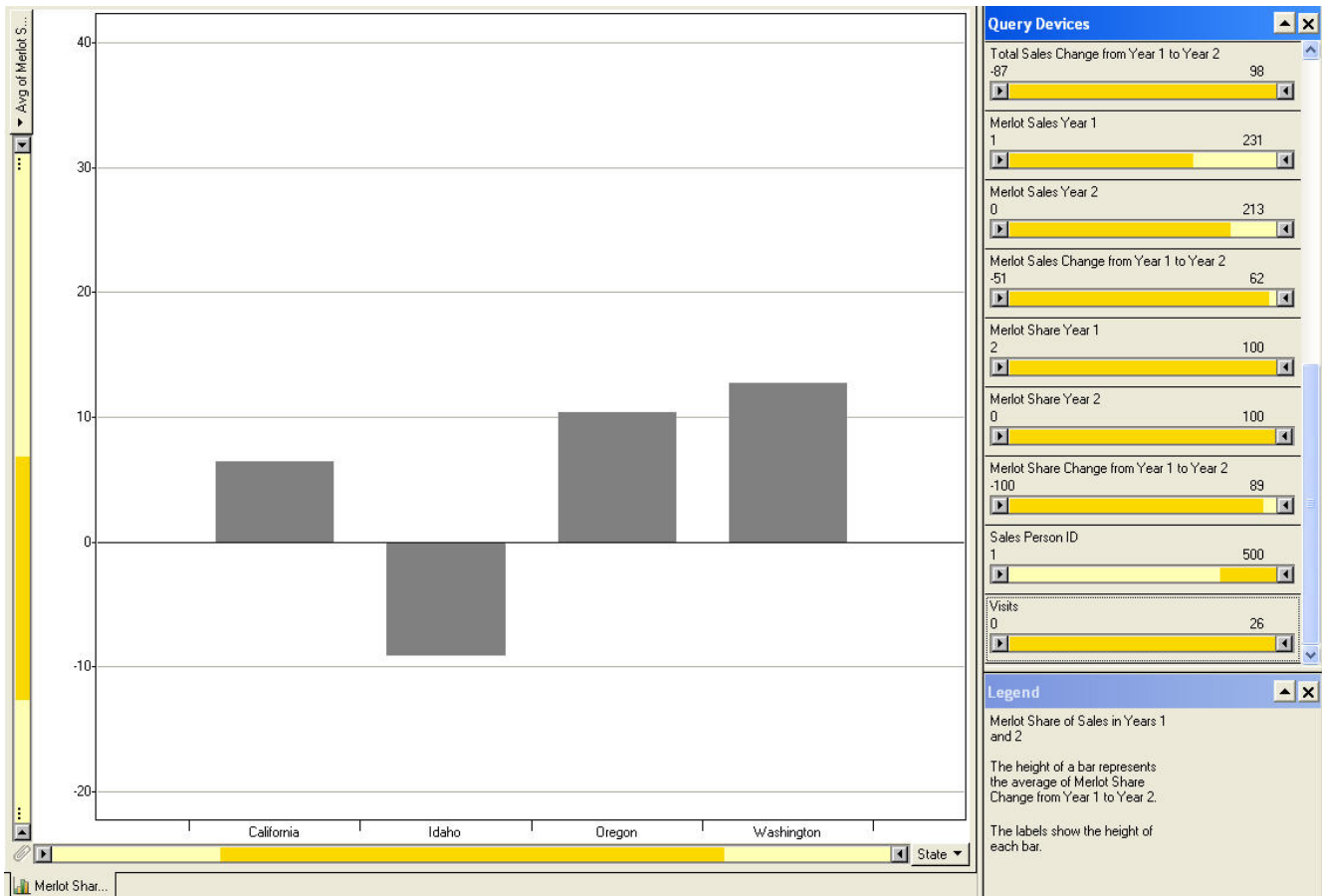


Figure 3

Of the four states in the west region, Idaho stands out as an exception to the pattern of an increase in Merlot's share. Now let us use the query devices to do something interesting. One of the measures in our data is the number of visits that were made by our sales team to each of the customers in the second year. Let us see if there is a relationship between the number of visits and the fact that Merlot's share decreased in Idaho. The query device for the number of visits is displayed in the form of the slider, shown in Figure 4.



Figure 4

This type of slider, with a control at each end, allows you to easily set the range of values to any range you want. Notice that the low and high ends of the slider are labeled with the values 0 and 26, which indicates the entire range of values in the number of visits measure. To filter the data based on the number of visits, in most BI software you would have to define the range, submit the query, and wait for the results; but with this slider, we can adjust the range dynamically and see the results change in the graph as we do so. For instance, we can adjust the slider to include only those customers who received one or more visits, then two or more visits, and so on, seeing the results of each adjustment in the graph as we proceed. Nothing much happens when we adjust the slider from one or more visits through three or

more visits, but look at Figure 5 to see the change that occurs from left to right when we adjust the slider from four or more visits to six or more visits.

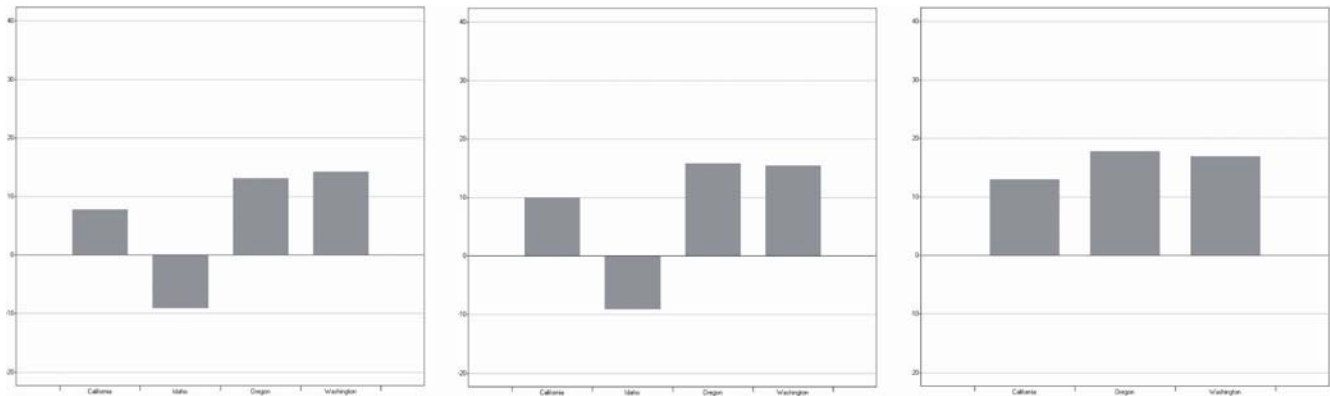


Figure 5

With five or more visits (middle graph), the change in Merlot's share increased a little in California and Oregon, but with six or more visits (right graph), Idaho disappeared altogether. In other words, none of the customers in Idaho received more than five visits from the sales team out of maximum of 26 visits that some customers received. Being able to see the data change in the graph as we are filtering it caused this meaningful aspect of the data to leap out.

Dynamic queries, enabled through simple filtering controls such as sliders, offer an enormous advantage when exploring data. These query devices represent a big leap in usability, although I suspect that we will see continued evolution of such user interfaces to the point that we do not even notice them, but interact with them entirely intuitively.

Viewing Data from Multiple Coordinated Perspectives Simultaneously

Multiple views of the same data, each displaying the data from a different perspective, can be quite revealing when seen together. I call these "faceted analytical displays." One of our most fundamental limitations when analyzing data resides in short-term memory. This is that part of memory that is active when we are consciously thinking about something. No more than four to eight chunks of information can be held in short-term memory at any one time. Traditional BI methods of analysis are often so fragmented that significant observations have already been forgotten by the time we make our next observations, which means that we cannot compare them to one another. Most of the insights that we gain through analysis come in the form of comparisons. Consequently, good data analysis software needs to compensate for the limits of short-term memory, primarily in two ways:

1. By organizing meaningful data into larger chunks so that the four to eight chunks of information that we can hold in short-term memory at any one time are as rich as possible.
2. By placing as much of the data that we need to compare within eye span so we do not need to keep it all in short-term memory to make the comparisons.

Take a look at Figure 6. In this table, each of the 24 numbers represents a chunk of information when stored in short-term memory, which means that only a handful of them can be considered at any one time.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Domestic	1,983	2,343	2,593	2,283	2,574	2,838	2,382	2,634	2,938	2,739	2,983	3,493
International	574	636	673	593	644	679	593	139	599	583	602	690

Figure 6

The graph of the same data in Figure 7, however, represents the values for each category (domestic and international sales) as a single line in the graph—a visual object that we can hold in short-term memory as a single chunk of information. This is one of the reasons that graphs are such a powerful medium for the analysis of data.

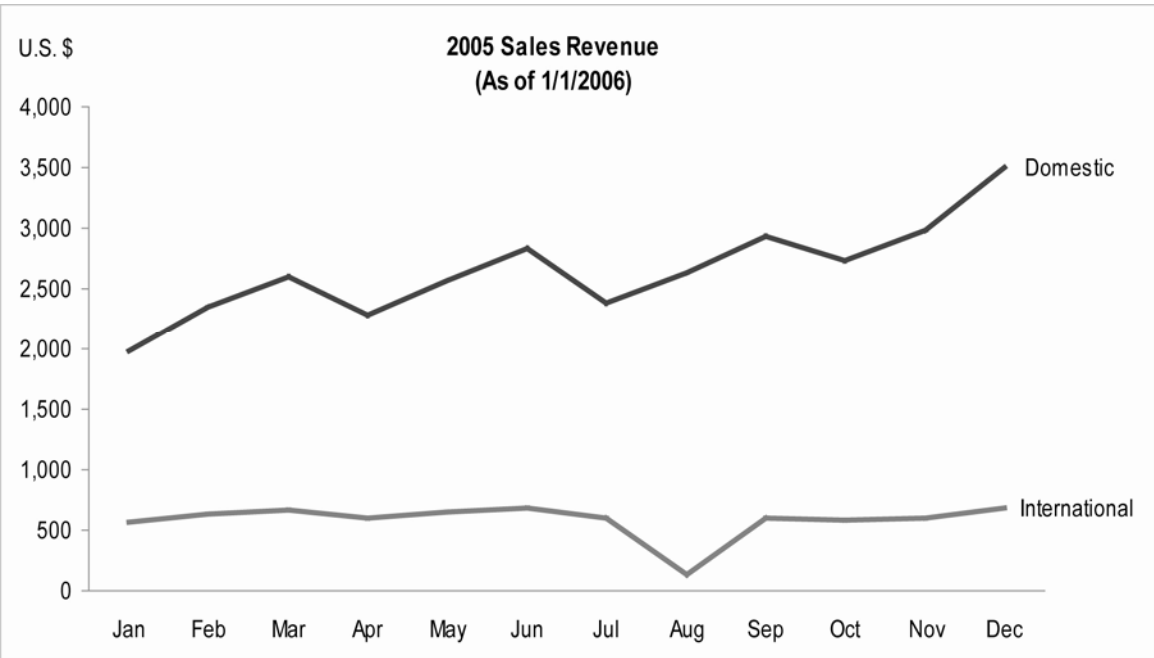


Figure 7

When comparisons need to be made between different values, sets of values, or views of the data, they need to be available to our eyes at the same time. Nothing needs to be held for long in short-term memory when it is all there before our eyes and we can quickly swap the chunks that we need in and out of memory as we are making connections and considering what they mean. Software can support this need by allowing us to place several different data displays on a single screen. Let us look at a rich assortment of perspectives on the same set of sales data that we have been exploring.

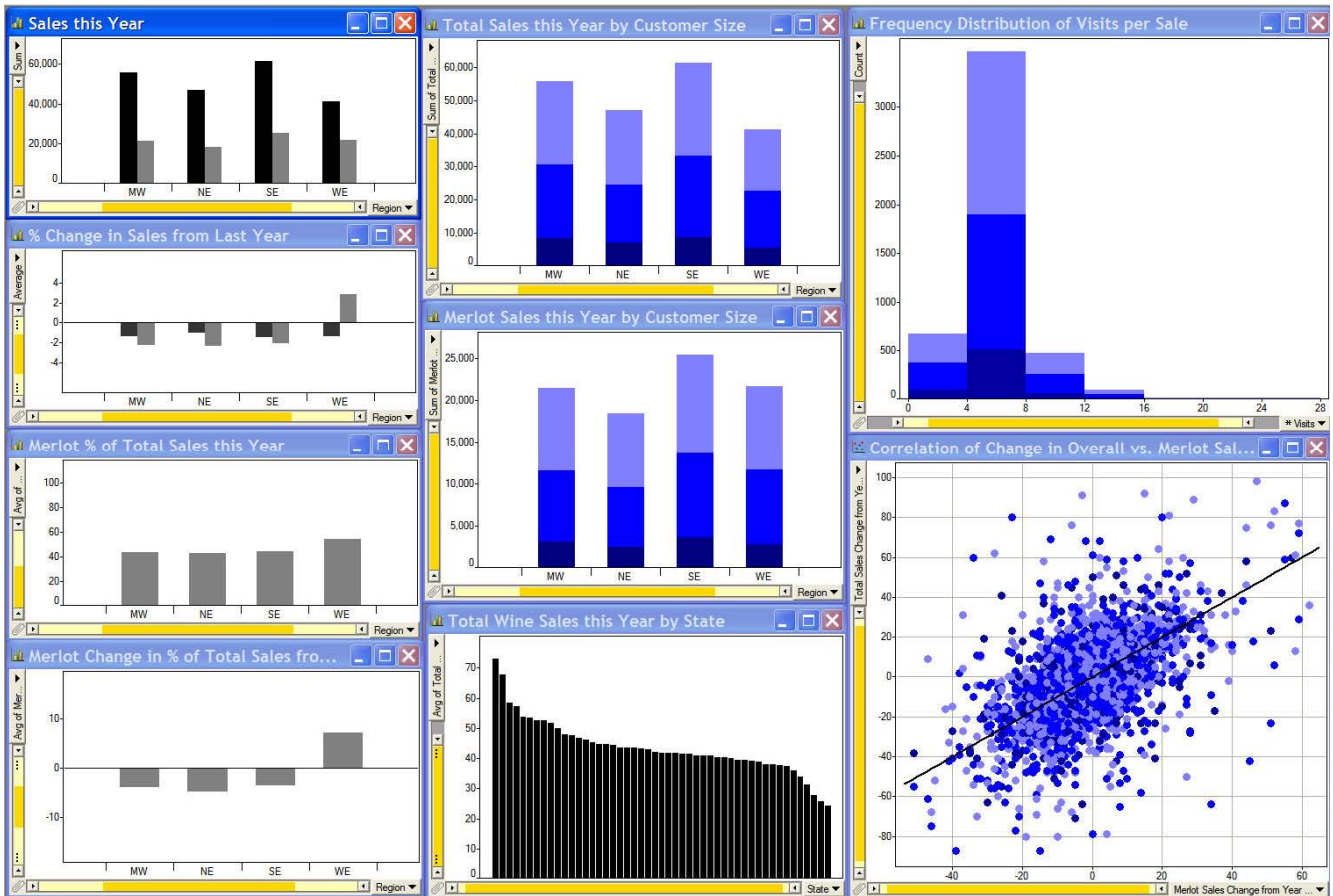


Figure 8: Example of a faceted analytical display created using Spotfire DecisionSite.

Starting with the graph in the upper left-hand corner of Figure 8 and working down each column, here is what we see:

1. Total wine sales (black bars) compared to Merlot sales (gray bars) in U.S. dollars for this year per region.
2. Percentage change in sales between last year and this year—total and Merlot—per region.
3. Merlot’s percentage of total wine sales per region.
4. Change in Merlot’s percentage of total sales from last year to this year per region.
5. Total wines sales this year per customer size (darkest gray represents big customers, medium gray represents medium-sized customers, and light gray represents small customers) by region.
6. Merlot sales this year per customer size by region.
7. Merlot share of total sales this year per customer size by region.
8. Frequency distribution of sales based on the number of sales visits that led to each sale (0-3, 4-7, 8-11, etc.).

9. Scatterplot showing the correlation of the percentage change in total wine sales from last year to this year (vertical axis) to the percentage change in Merlot wine sales from last year to this year (horizontal axis).

Each of these graphs provides a different perspective on our sales. Seen together, they help us see connections (interesting interactions) between various aspects of sales. Study these graphs for a few minutes to see what worthwhile stories they tell you about sales.

Here are a few features that caught my attention while examining these graphs:

1. Merlot sales in the west region represent approximately 55% of total sales, the greatest share of any region, but this is mostly due to a dramatic increase in Merlot sales in the west. Despite this strong showing in the west, Merlot's share of total wine sales actually went down a little in the midwest and northeast and increased only slightly in the southeast. Merlot's high share of total sales in the rest is strongest among small customers.
2. Even though the west barely beat the northeast in overall wine sales this year (third place of the four regions), they came in a strong second in Merlot sales, nearly catching up with southeast. In fact, Merlot sales to medium-sized customers in the west roughly match those in the southeast.
3. Most sales resulted from between three and seven visits to a customer, and the number of visits that seems to be required for sales is fairly consistent between big, medium, and small customers.
4. While there is a positive correlation between changes in total wines sales vs. Merlot sales from last year to this year (that is, as total wines sales increased or decreased, Merlot sales tended to do the same), there are a number of exceptions to this pattern, especially in the lower right corner where a disproportionate number of Merlot sales increases appear relative to decreases in overall sales.

Many of these discoveries, as well as others you perhaps made while examining this collection of complementary graphs, might have remained invisible using more traditional BI methods of analysis, and certainly would have been much more difficult and time-consuming to ferret out.

Highlighting Subsets of Data Simultaneously and Automatically in Multiple Views

This next technique builds on the one we just examined. The revelatory power of multiple perspectives on the data seen together can be extended through a technique that information visualization researchers call *brushing*. Let us say that we are looking at one of the graphs in Figure 8 and we become interested in a particular subset of the data, such as those Merlot sales that decreased since last year even though overall wines sales increased, as shown in the upper left quadrant of the scatterplot. Now imagine that we have a brush that we can use to paint a rectangle around these particular data points in the scatterplot to highlight them. Figure 9 shows the result of this action.

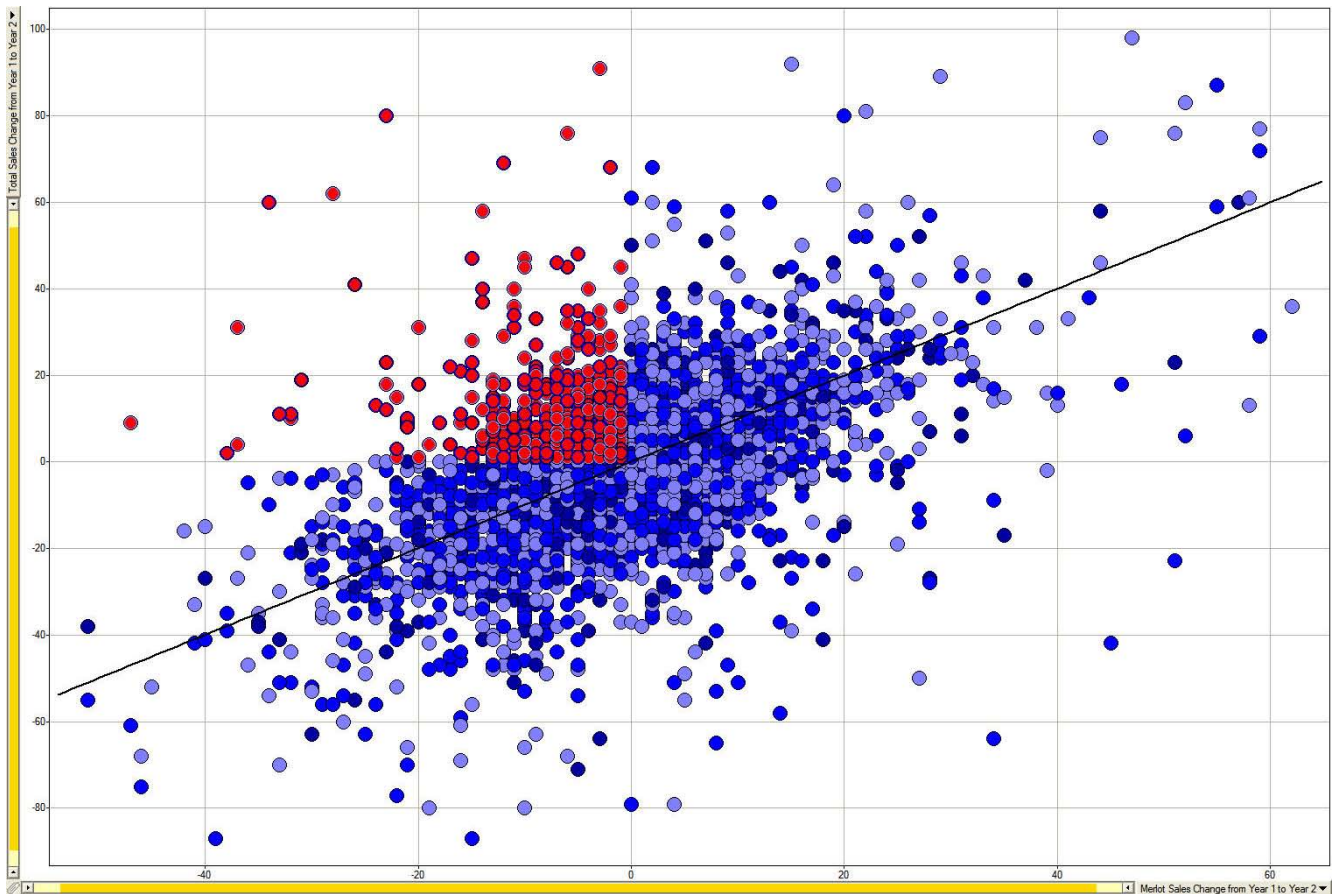


Figure 9: All values above zero on the Y-axis and below zero on the X-axis have been highlighted.

Now all the values of overall wine sales that are greater than zero corresponding to Merlot sales that are less than zero are highlighted in red. Simply making them stand out in the scatterplot, however, is not the point of this exercise. What we really want to see is how this particular subset of data behaves in all of the other graphs on the screen. That is exactly what brushing does for us, and it does it automatically. Take a look now at the full screen shown in Figure 10 and see if it leads you to any interesting insights.

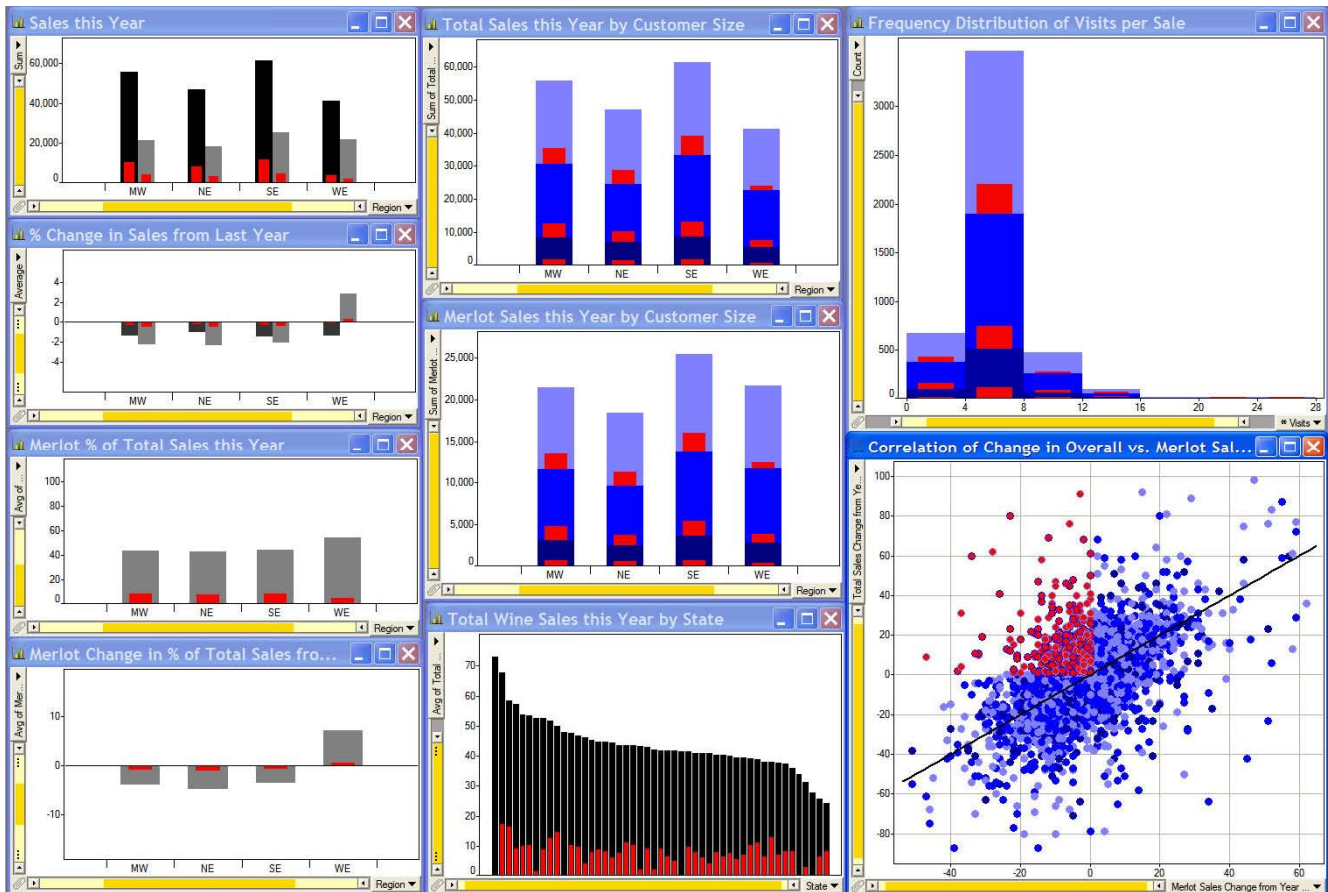


Figure 10: Brushing caused a subset of data that was highlighted in one chart to automatically appear highlighted in all charts. This example was created using Spotfire DecisionSite.

One of the first things that I noticed is that the decreases in Merlot sales that are out of sync with corresponding increases in overall wine sales occurred less often in the west (especially seen in upper graph in the left column and the middle graph in the center column). Relative to customer size, this pattern also seems to be disproportionately strong in the southeast (see the top and middle graphs in the center column).

Now let us say that we want to see if sales in the west fall disproportionately in any particular area of the scatterplot. To see this, we can brush the west region in any of the graphs in the left or center columns and see the results highlighted in every graph, including the scatterplot shown in Figure 11.

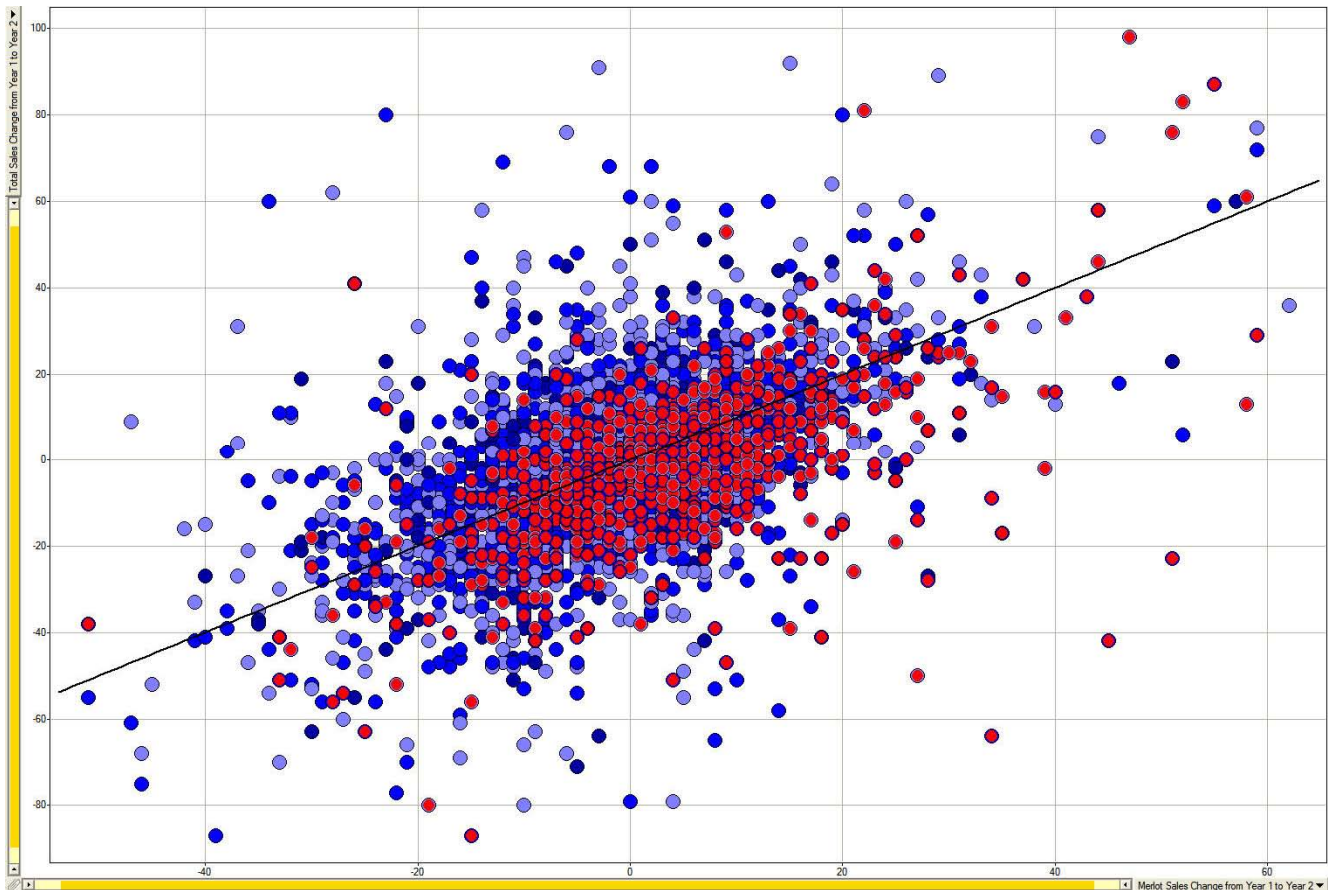


Figure 11

Notice that a disproportionate number of sales in the west (highlighted in red) appear below the trend line and in the right half (increases in Merlot sales) of the graph. This reaffirms our previous observation that the west has contributed more than other regions to increases in Merlot sales, especially when overall wine sales changed less than the average degree.

In the BI Zone

Even with this simple data set, we could spend hours pursuing lines of investigation as quickly as they come to mind. When the kinds of analytical functionality that we have just examined are enabled by software in such simple and efficient ways, the step-by-step process—that begins with observation, then raises a question, and is followed by manipulation of the data to pursue that question, resulting in new observations and insights—becomes fluid, without interruption. When this happens, we can achieve a state of awareness and insight that psychologist Mihaly Csikszentmihalyi (pronounced “chick-sent-me-high”) calls “flow” and others more colloquially call being “in the zone.” This can only happen when our awareness of the software we are using recedes into the background and we become fully immersed in the data and its rich story.

With the right tools, you can clear the fog and learn to analyze data at the speed of thought. Take advantage of the techniques that information visualization researchers have developed to help you work smarter. Achieve the promise of business intelligence today.

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.