improve your vision and expand your mind with visual analytics

AUTHOR:
Stephen Few

DATE:
August 1, 2007

INTRODUCTION

More and more, the quality of our lives, the vitality of our businesses, the effectiveness of our governments, and the well being of our world relies on our ability to find, examine, and make sense of information. Decisions ranging from where to go for dinner to whether or not to militarily invade a country are based on access to relevant and accurate information, and our ability to understand it. As we know from recent world events, the consequences of inaccurate, incomplete, or misunderstood information can be dire. Although lives might not be at stake when most business decisions are made, similar scenarios involving information that result in the wrong decisions are played out in the course of business every day. Humans will always have the capacity for bad decisions — it's all too natural — but we can dramatically reduce the number that result from misinformation or misunderstanding by bringing good analytical skills to the process, supported by good analytical tools. No tool can enable enlightening analysis better than good visual analysis software.

Data analysis is something that is done by humans, not computers. At best, computers can assist the process, which they can do quite powerfully with well designed software. Software is well designed when those who build it not only understand what people really need to accomplish, but also understand the strengths and limitations of the people who will use it. Too much software exists that doesn't understand the humans who use it — how we perceive and how we think. As a result, we limp along with software tools that barely do the job, steeped in the clumsy mechanics of using the software as we work around its foibles, rather than immersed in the process the software supposedly supports. Good software fades into the background when you use it, allowing you to see what you need to see and do what you must without distraction. For the process of data analysis, we should demand tools that help us discover, think about, interact with, and understand information, with our attention fully on the data, lost in thought and primed for illumination.

USING THE STRENGTHS AND SUPPLEMENTING THE LIMITATIONS OF THE HUMAN BRAIN

The human brain is highly evolved and unique in its ability to reason abstractly, which is required to analyze data. Nevertheless, it has its limitations. One of its primary limitations, which affects our ability to analyze data, involves short-term memory (a.k.a. working memory).

You have probably heard about "the magical number seven, plus or minus two," which Harvard psychologist George Miller proposed as the limit of short-term memory back in the 1950s. Actually, research that has been done since Miller's original work indicates that the limits of visual memory are actually more restrictive — closer to four chunks of information. If we can only hold onto this many chunks of information at a time while thinking about something, it is helpful when software assists us in two ways:

- It packages information into big chunks to take the best advantage of the limited storage space in short-term memory
- It allow us to view a great deal of information simultaneously so we don't need to rely as heavily on memory to keep track of things

It is partly its ability to work around the limitations of short-term memory that makes visual analysis software so useful.

Sight is our most powerful sense. Any way you do the math, what we perceive through our eyes surpasses our other senses, both in terms of pure volume of information and speed of input. Seventy percent of the sense receptors in our bodies reside in our eyes. In terms of bandwidth of input, sight is ten times greater than touch, which comes in second. We also see greater subtlety and nuance than is available through any of our other senses. For instance, we are capable of discriminating extremely slight differences in color. This powerful sense makes us especially talented and agile visual pattern detectors. It is not accidental that the expression "I see" is an acceptable stand in for "I understand."

Tables of information encoded as text — the traditional business intelligence display — are almost useless for spotting patterns in quantitative information. But present that same information visually, in the form of a graph, and a picture suddenly make visible what would have otherwise remained hidden in the numbers. Joseph Berkson once referred to the effect of a good data visualization as "interocular traumatic impact" — a conclusion that hits us between the eyes. Well put. This is the experience we seek when analyzing data.

"When designers of data analysis software decide to include visualization functionality in their products, they are making a design choice that understands and respects the workings of the human brain."

When you need to spot and examine trends, patterns, and exceptions in data, visual representations surpass all other representations by far. It is easy to spot and estimate a downwards trend in sales when it is displayed in a simple line graph. It is easy to see a cluster of products with an abnormal relationship between marketing expenses and sales revenues when they show up as an isolated group of dots in a scatterplot. It is easy to detect exceptional behavior when a single bar in a graph towers over its comrades.

Visual representations of data work in a direct way to help us overcome the limits of short-term memory. Information is stored in memory as chunks. A single number in a table can be stored as a single chunk in memory. If you study a table of monthly revenue figures and then look away, only up to four or so months of revenue values at most will be retained. The same 12 months of data, however, when displayed as a line in a graph, forms a single picture of the entire year's sales, which can be stored as a single chunk. In other words, by properly encoding data visually, we are able to dramatically increase the total amount of information that we can simultaneously store in memory. When designers of data analysis software decide to include visualization functionality in their products, they are making a design choice that understands and respects the workings of the human brain.

Another way to work around the limitations of short-term memory involves placing all the information an analyst must consider at any one time simultaneously on the screen. Only then can the necessary comparisons be made. If you must switch back and forth between separate screens or scroll around the screen to view everything, you will never remember what you were looking at a moment ago when you wish to compare it to what you're looking at now.

Good data analysis software displays information visually and provides ways to place a great deal of information on the screen at one time. One powerful way that this can be done is through a device I call a visual crosstab.

VISUAL CROSSTABS

A crosstab, short for cross tabulation, is a table that breaks data down into parts, such as individual departments or months, and does so both down the rows and across the columns. The values that are displayed in a crosstab represent the intersection of those parts. Here's a simple example of a crosstab:

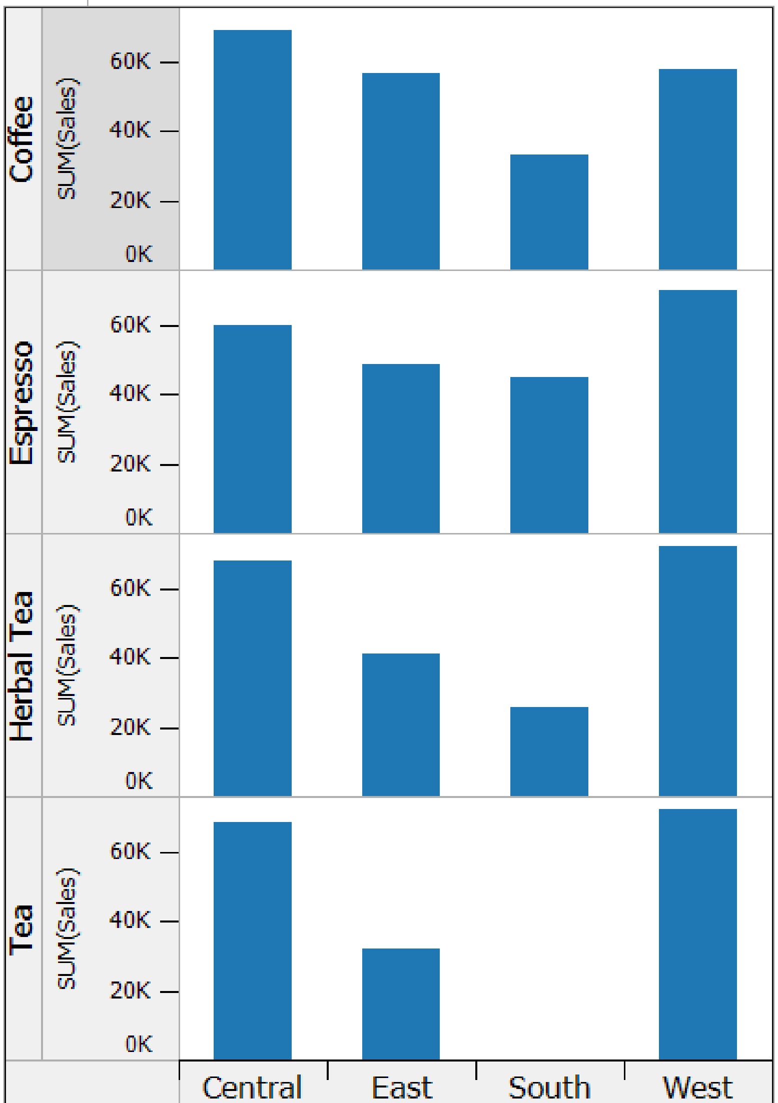
FIGURE 1: A simple crosstab display of quantitative data

	Central	East	South	West
Coffee	69,080	56,640	33,256	57,856
Espresso	59,703	48,405	44,989	69,911
Herbal Tea	67,885	41,362	25,683	72,285
Tea	68,380	32,172		72,220

Product types (coffee, etc.) are displayed down the rows and regions (central, etc.) across the columns. Each cell in the body of the table (the white section in Figure 1) is located at the intersection of a particular product type in a particular region, and expresses the quantitative value of that intersection. Typical of crosstabs, the quantitative values — sales revenues in this case — are displayed as text. Anyone who has ever used spreadsheet software, such as Microsoft Excel, is familiar with this type of display.

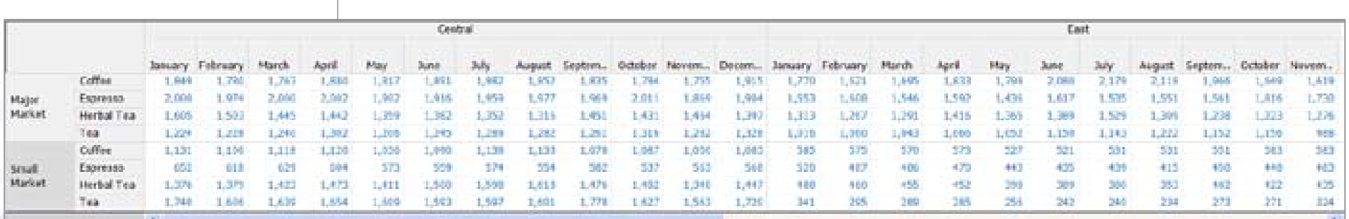
Now, let's take advantage of the familiar crosstab arrangement of data, but extend it by displaying the values visually. Here in Figure 2 are the same sales values, this time encoded graphically as bars:

FIGURE II: A visual crosstab display of quantitative data.



Because this display only includes 16 values, it isn't very interesting, but you can nevertheless already see how much easier it is to compare the values to one another. Let's add more values to the display by breaking it down into more parts, to more fully appreciate the advantages of visual displays for analysis. In Figure 3, you can see that months and market types have now been added:

FIGURE III: This table displays sales revenues, with market types and product types arranged down the rows, and regions and months arranged across the columns.

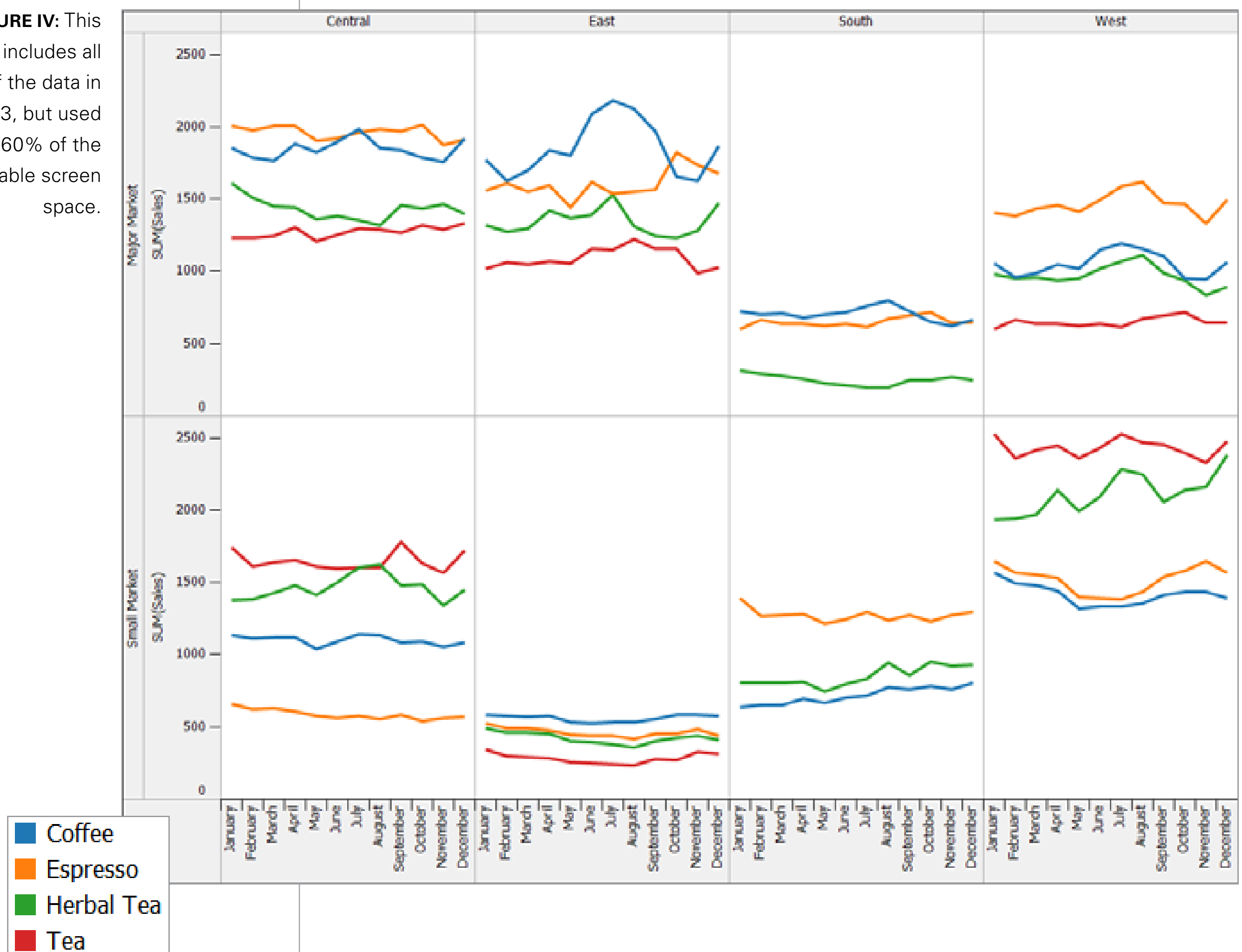


		Central											
		January	February	March	April	May	June	July	August	Septem...	October	Novem...	Decem...
Major Market	Coffee	1,644	1,790	1,767	1,880	1,317	1,851	1,882	1,857	1,835	1,794	1,755	1,915
	Espresso	2,008	1,974	2,000	2,082	1,902	1,916	1,953	1,977	1,964	2,011	1,869	1,984
	Herbal Tea	1,605	1,503	1,445	1,442	1,359	1,362	1,352	1,315	1,451	1,431	1,464	1,397
	Tea	1,229	1,228	1,240	1,302	1,200	1,245	1,289	1,282	1,281	1,319	1,212	1,328
Small Market	Coffee	1,131	1,156	1,118	1,120	1,030	1,090	1,138	1,133	1,078	1,087	1,056	1,083
	Espresso	652	618	629	604	573	559	574	554	582	537	553	568
	Herbal Tea	1,376	1,375	1,423	1,473	1,411	1,500	1,598	1,613	1,476	1,482	1,348	1,447
	Tea	1,748	1,606	1,635	1,654	1,469	1,563	1,597	1,601	1,778	1,627	1,543	1,726

		East										
		January	February	March	April	May	June	July	August	Septem...	October	Novem...
Major Market	Coffee	1,770	1,621	1,695	1,633	1,798	2,088	2,179	2,118	1,966	1,649	1,419
	Espresso	1,553	1,608	1,546	1,592	1,436	1,617	1,535	1,551	1,561	1,416	1,730
	Herbal Tea	1,313	1,287	1,291	1,416	1,365	1,389	1,529	1,306	1,238	1,223	1,276
	Tea	1,010	1,000	1,043	1,000	1,052	1,150	1,143	1,222	1,152	1,150	988
Small Market	Coffee	385	375	370	373	327	321	331	331	331	383	383
	Espresso	320	407	406	470	443	435	439	415	450	448	463
	Herbal Tea	400	460	455	452	359	389	306	353	462	422	435
	Tea	341	305	289	285	254	242	240	234	273	271	324

As you can tell by the presence of the horizontal scroll bar at the bottom, only a little more than half of the data is currently visible. Despite this problem, test yourself to see how easily you can find which product type in which region and market type sold the most or sold the least during the year. Also, try to get a sense of the pattern of Espresso sales across the year in small markets in the East. Text-based tables don't support these typical operations of analysis. Look now at the same data displayed in a visual crosstab:

FIGURE IV: This display includes all of the data in Figure 3, but used only 60% of the available screen space.



Notice how easily you can now see and compare the sales trends of the product types, regions, and market types. Now try to spot which product type in which region and market type had the highest sales and which had the lowest? Several other features in the data are now clearly visible as well, such as those product types that had the most volatility versus stability across the year in particular regions and market types.

The value of visual displays such as the one above, which arrange a series of small graphs horizontally, vertically, or in a matrix, has been known for many years. In 1983, Edward Tufte dubbed them small multiples in his book The Visual Display of Quantitative Information. Others sometimes call them trellis displays. What Tufte described, however, involved a series of graphs that varied along a single variable, such as regions or products, but not along more than one at the same time. Tableau Software introduced displays designed like the one above as a logical extension of Tufte's idea to increase the power of the crosstab display that was already familiar to most business people. The traditional crosstabs and pivot tables, which we once welcomed with great fanfare because they enabled us to explore and analyze data multidimensionally, have now been surpassed in analytical possibility by new visual displays that extend this familiar paradigm.

Don't let the simplicity of the visual crosstab beguile you into the mistaken belief that it is commonplace or unworthy of your attention (to say nothing of heartfelt gratitude). I'm familiar with this brand of response, because much of what I teach about analyzing and presenting data is so easy to learn and simple to practice, people often respond with a casual "well…duh" at the end of class. Truth be told, however, they didn't know what I taught them before the class and had regularly turned out ineffective graphs or stumbled haphazardly through the analysis process as a result. Sometimes the best innovations are those that seem the simplest and most obvious upon reflection, but remained unknown and unavailable until somebody came up with the idea and generously shared it. We should all be grateful to the folks at Tableau Software for their simple but elegant and powerful means of displaying and interacting with data.

SIMPLE IS OFTEN BEST

Despite the many flashy visual displays that many software vendors advertise today, the best are usually the simplest. The flashy stuff usually gets in the way of the data, and once you dig past you, you often discover that there is nothing meaningful there. Familiar 2-D graphs with an X and a Y axis, which encode quantitative values as points, lines, and bars almost always work best for analyzing business information.

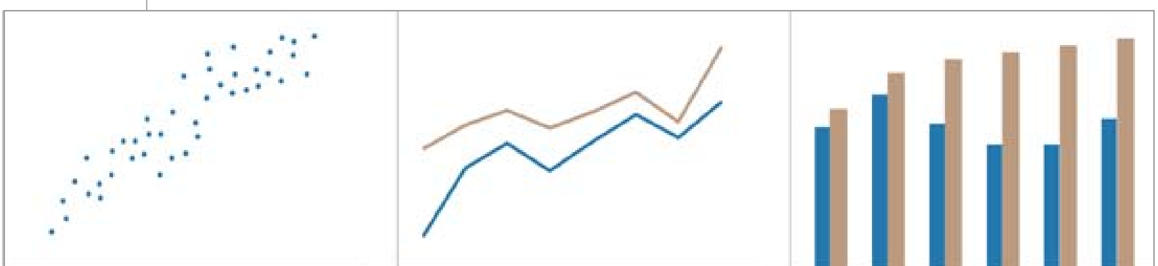


FIGURE V: Points, lines, and bars are usually the best ways to encode quantitative values.

Points, lines, and bars in a simple graph work well when properly designed, because they encode data in ways that we can easily perceive. Visual perception excels in spotting differences in the 2-D location of objects (such as points in a scatter plot or the up and down patterns of a line) and in the length of objects (such as bars that are arranged next to one another along a common baseline). These are the two most powerful ways to encode quantitative values.

Sometimes, however, when we need to add more dimensions to the data we're examining, we're forced to rely on other visual means to encode quantitative values, because 2-D location and length have already been used. Good visual analysis software recognizes that the next best ways to encode values, which we perceive less precisely, but can be used in a pinch, are color intensity (for example, from light to dark expressions of a single hue) and size (such as dots in a scatter plot that vary in size). Figure 6 illustrates this practice.

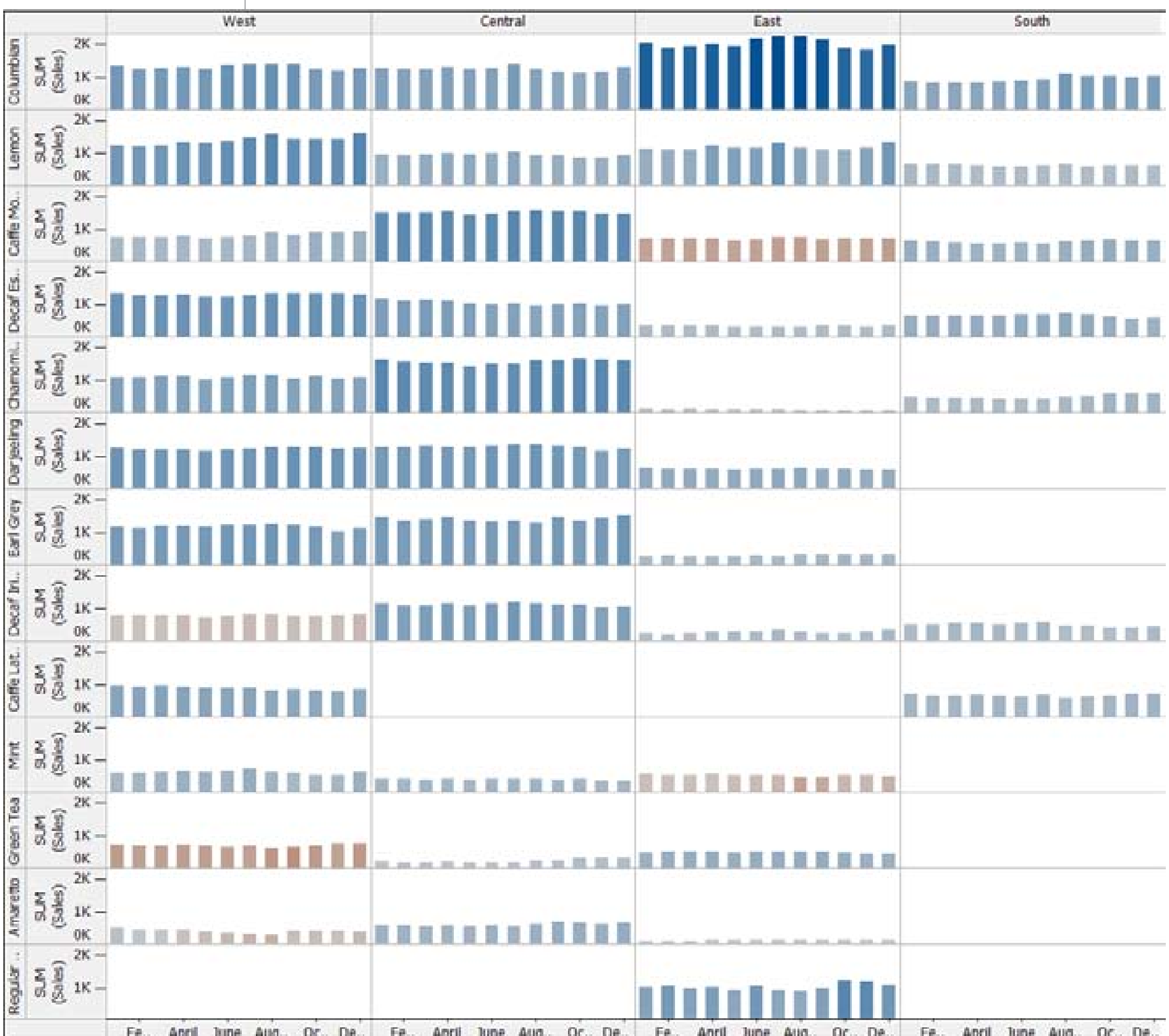


FIGURE VI: This display allows us to view and compare sales (the heights of the bars) and profits (the colors of the bars) simultaneously.

“When we need to add more dimensions to the data we’re examining, we’re forced to rely on other visual means to encode quantitative values, because 2-D location and length have already been used. Good visual analysis software recognizes that the next best ways to encode values, which we perceive less precisely, are color intensity and size.”

Here we see monthly sales (one month per bar) of individual products by region, but color is now being used to display another quantitative variable. The regions and products have also been sorted in descending order of sales to make comparisons easier. The heights of the bars represent sales revenues and their colors (ranging across increasing intensities of blue for positive values, increasing intensities of red for negative values, and gray for values near zero) represent profits (see the color scale in Figure 7).

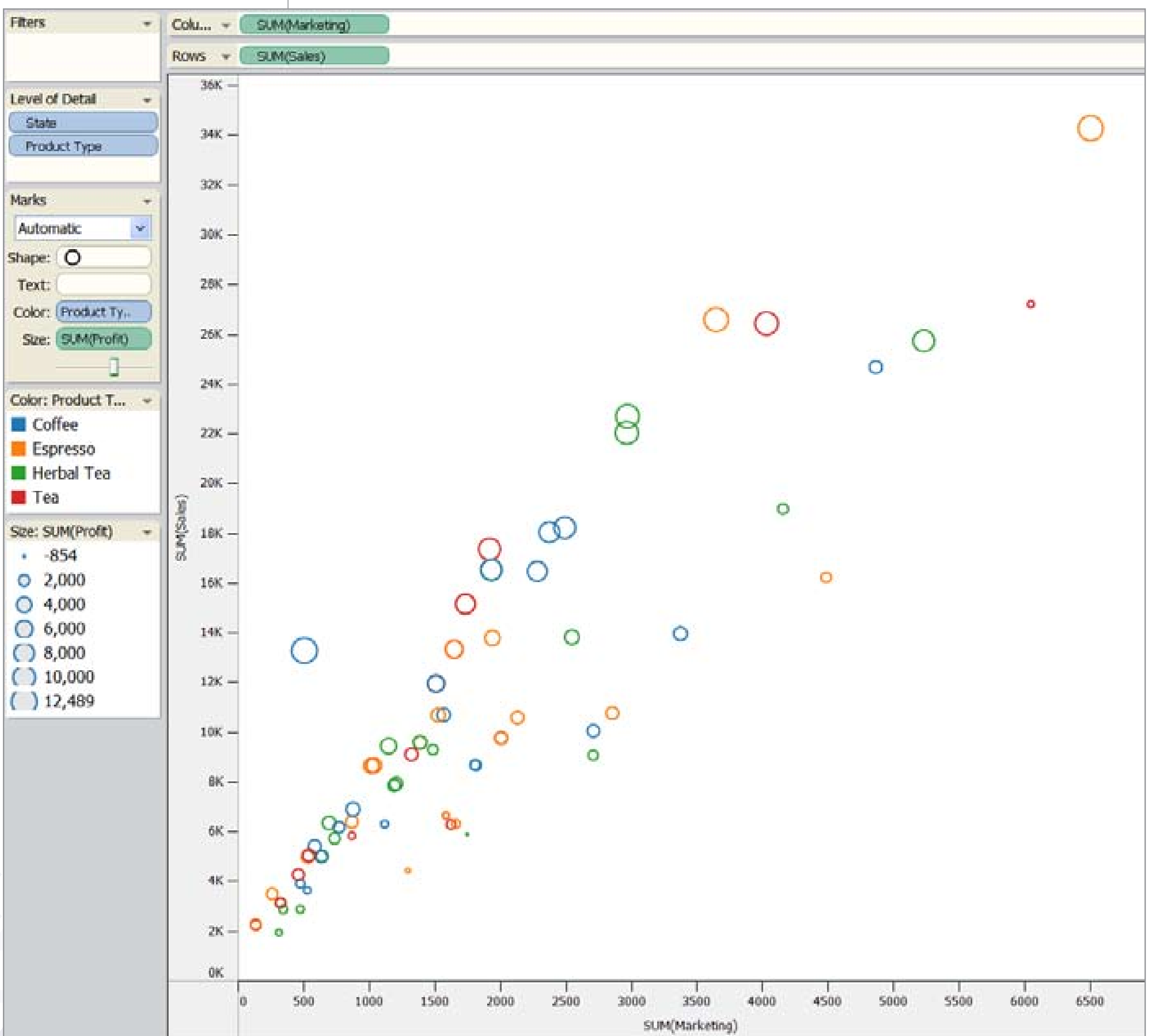
FIGURE VII: This continuous color scale varies on the right from light to dark blue for profits, and on the left from light to dark red for losses, with gray in the middle for profits and losses near zero.

Even though we are not able to interpret slight differences in color or the precise profits represented by the colors, we don’t need to in this case. We’re using color intensity to spot extremes and basic patterns, such as the fact that the greatest losses occurred for Caffe Mocha in the East and Green Tea in the West, consistently throughout the year. We can also see that Mint tea suffered losses in the East, mostly in the hot months of August and September. Because revenues and profits can be examined together in this display, we quickly spot the fact that the greatest revenues and profits are both earned by Columbian Coffee in the East.

Besides color intensity, the other visual attribute that can be used to encode quantitative values when 2-D location or length cannot b e used, is size. Figure 8 illustrates this practice in a scatterplot.

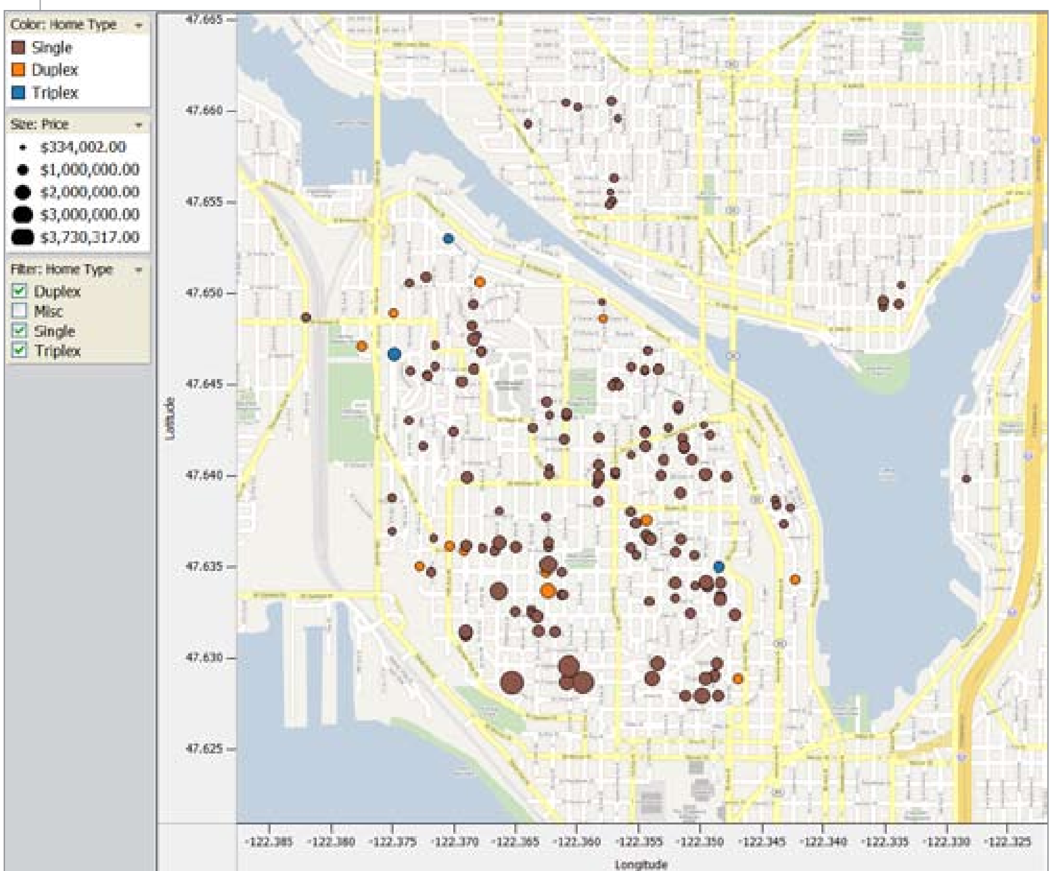
Circles are used to encode three quantitative variables: marketing expenses as their position along the X-axis, revenues as their position along the Y-axis, and profits as their size. Each circle represents a different product type sold in a particular state. It wouldn’t be possible to precisely interpret and compare the profits based on the sizes of these small circles, but in this case that’s not necessary. Right now we’re looking for patterns. In general, there is a correlation between marketing and revenues, seen in the fact that the circles are arranged in a roughly linear fashion, beginning at the bottom left corner and moving upwards to the upper right corner. This is what we would hope for: that the money we spend on marketing products usually results in increased sales. One exception shows itself as the large blue circle (a coffee, which we know from the color legend on the left) that stands out on the left near 14K on the sales revenue scale did well despite very little marketing investment. If you look at the upper right section of the scatterplot, you'll notice a small red circle that ranks second in marketing expenses and revenues, but the small size of the circle tells us that it earned very little profit. It is difficult spot these kinds of relationships in the data when you cannot examine all of these variables at once.

Figure VIII: This scatterplot uses the size of the dots to encode a third quantitative variable.



The visualizations that are used in the analysis of quantitative data usually represent abstract information. This requires a translation of abstract concepts, such as rising costs, into visual objects that represent these concepts clearly. A rising line in a line graph works wonderfully for showing how costs are changing through time, but there are frequent occasions when the information we analyze includes a spatial component (where things are located) that cannot be shown in X and Y axis graphs. We have a great deal of experience graphing when things happen (time) but visualizing where things happen (space), is less familiar in business analysis. Spatial information, especially geographical information, however, can be displayed quite naturally in the form of a map — a visualization we've been using for millennia. When abstract information such as home sales must be seen in the context of where they occurred to fully understand what's going on, map-based displays are the obvious solution (see Figure 9).

FIGURE IX: When seeing where things are located is helpful to make sense of the data, well-designed map-based displays are ideal.



For instance, if you want to find out where the highest concentration of sales occurred or where few if any occurred, the geographical display in Figure 9 makes this clear in a way that any other form of display could not. Map-based displays are not the only visualizations other than X and Y axis graphs that come in handy, but they are certainly the type that you'll need most often when analyzing business data.

"It isn't enough to make these operations available; software should allow you to perform them effortlessly, without interrupting your train of thought or view of the data."

ANALYTICAL INTERACTION AND NAVIGATION

Good visual analysis software takes advantage of the simple, familiar, and effective graphs, and brings them to life by making them interactive. The process of finding what's meaningful in your data and then making sense of it (a.k.a. analysis) requires interaction. Some of the most common and useful interactions in data analysis include:

- Filtering out what's not relevant
- Sorting the data to see it in order of magnitude
- Moving between high-level (the big picture) and low-level (the details) views of the data
- Drilling up and down through levels in hierarchically structured data
- Changing your view of the data, such as by switching to a different type of graph, to view it from a different perspective

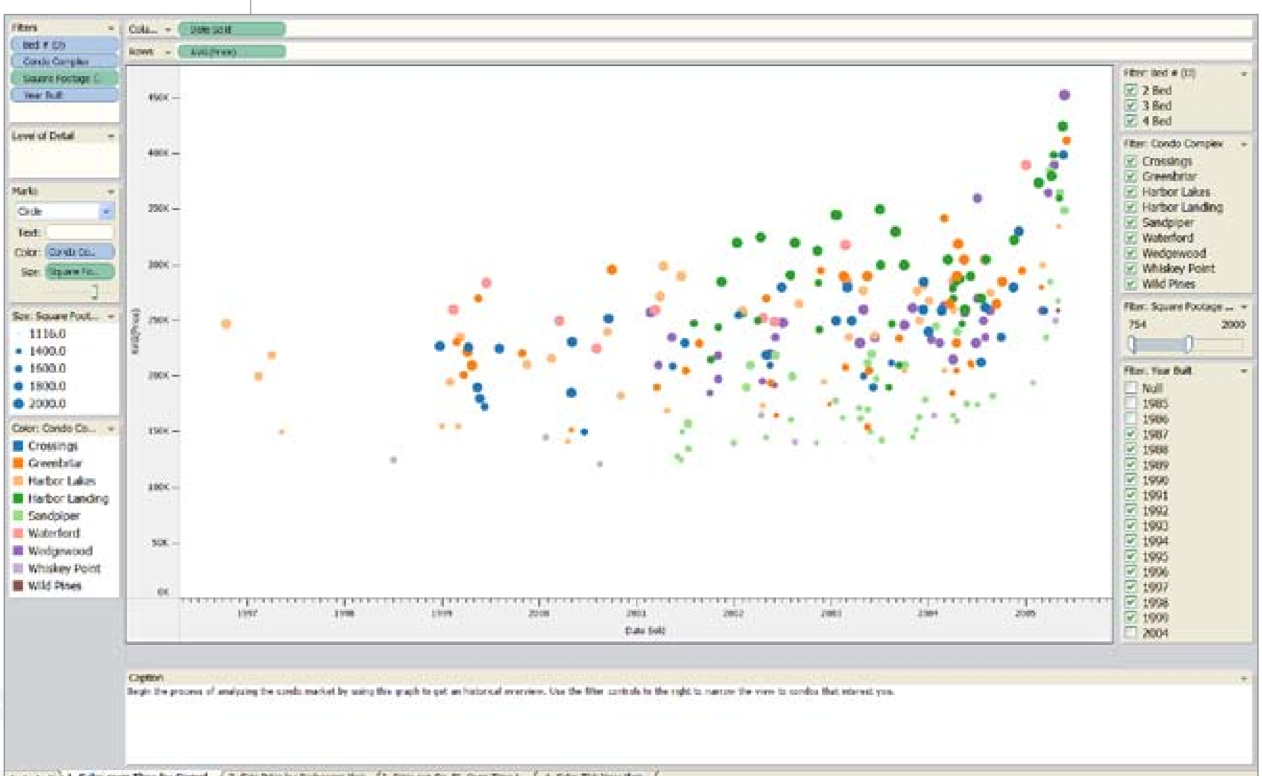
It isn't enough to make these operations available; software should allow you to perform them effortlessly, without interrupting your train of thought or view of the data.

The eminent psychologist Mihaly Csikszentmihalyi (pronounced, chick-sent-me-high) coined the term "flow" for the experience of total engagement in an activity at a level of peak performance. Anyone can experience flow, losing touch with space and time while absorbed in an activity, sailing along in a state of mastery. Top athletes experience this often, as do writers, scientists, doctors, artists, and even data analysts. The experience of analytical flow is far less common than it could be, however, because most of the software tools that analysts use get in the way. The mechanics of using the software are often cumbersome, which fragments our attention and distracts us from becoming fully immersed in the data. Imagine being in a state of analytical flow, seamlessly moving between the detection of interesting discoveries, to closer examination of them, to the removal of data that's irrelevant, to shifts in the display to see the data from different perspectives, to new discoveries, to further filtering, and on and on as you interact with the data, just as fast as you can think of things you wish to see. Well crafted visual analysis software is bringing us closer and closer to this ideal.

I use the term analytical navigation to describe the process of moving step by step, from one activity or view to another, when analyzing data. Some types of analysis require a form of navigation that is much like "stream of consciousness" writing. We need to remain completely flexible and let our perceptions and thoughts take us wherever they will in rapid succession. On the other hand, many of the analyses that we do in the course of business are more predictable in nature, best accomplished by performing a prescribed series of steps in a particular sequence. An expert must determine the steps and sequence, but once they're determined, the process ought to be automated as much as possible to guide people through the steps with minimum effort and delay.

Capturing the best practices of experts in the form of guided analytics, a path that others can navigate on auto-pilot, is a feature that contributes enormously to the analytical productivity of an organization. By "navigate on auto-pilot," I'm not suggesting the people don't need to think — they absolutely do — but simply that they don't need to invent from scratch over and over the process that's required to navigate through a routine analytical process. Figure 10 illustrates a simple approach to guided analytics, using predefined displays and filter controls separated into multiple tabs — one for each step in the process — along with instructions to guide people through the process (shown in the caption at the bottom).

FIGURE X: The tabs at the bottom allow people to step through the analytical process and the instructions in the caption tell them what to do.



"If you are accustomed to using text-based tables to analyze data, the thought of switching to a visual form of analysis might seem incomplete — where are the details, the precise numbers that you're accustomed to seeing? Good visual analysis software hides these details until we need them, and makes them extremely easy to see when we do."

EASY ACCESS TO ALL THE DATA, BUT ONLY WHEN YOU NEED IT

If you are accustomed to using text-based tables to analyze data, the thought of switching to a visual form of analysis might seem incomplete — where are the details, the precise numbers that you're accustomed to seeing? The answer is that they are always waiting just out in the wings for the moment that you need them, being careful to not get in your way until you do. The truth is, when we're engaged in analyzing data, most of our time should be spent on searching for and examining trends, patterns, and exceptions — precise values are not only unnecessary, they clutter the screen and slow us down. Good visual analysis software hides these details until we need them, and makes them extremely easy to see when we do.

This ability to easily access the details, but only when you need them, is called details-on-demand. Good visual analysis software allows you to hover with your mouse pointer over any object in a chart (for instance, a data point or bar) and gives you the details behind that object in a small pop-up window (see Figure 11).

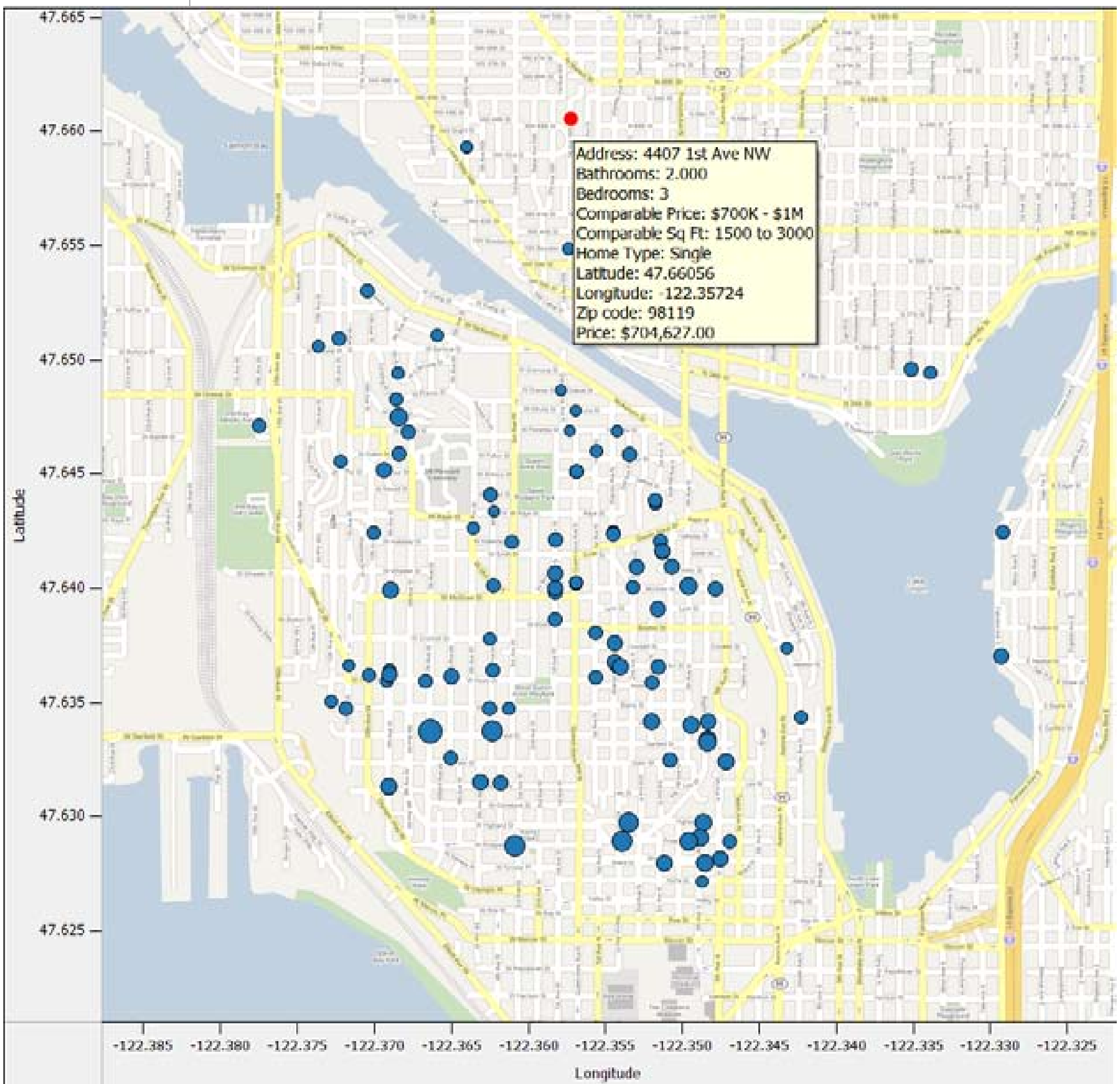
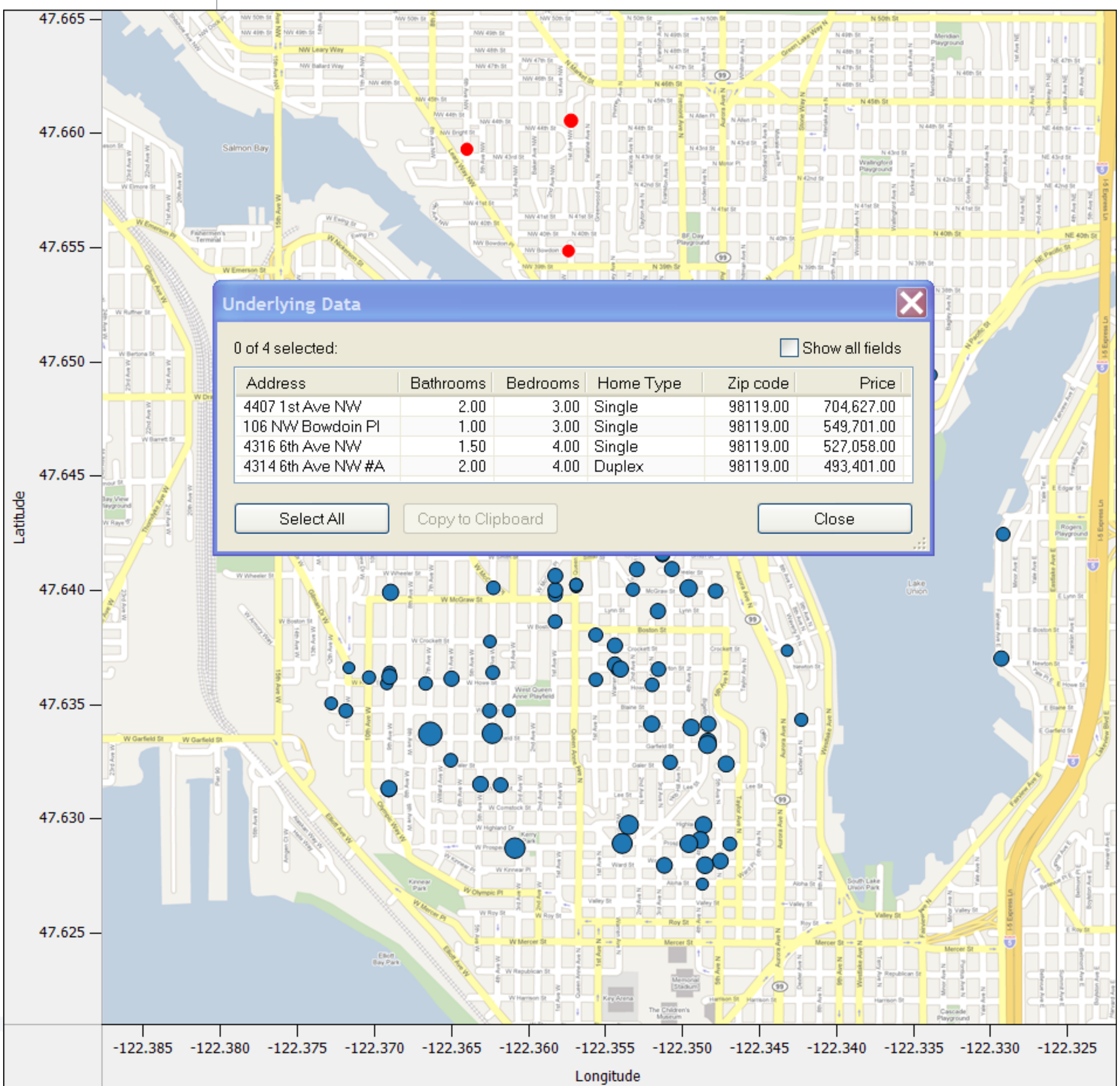


Figure XI: Details have been accessed by hovering over an item.

Figure XII: Details have been accessed for a selected group of items.

Figure 12 illustrates another means that's available in Tableau, which allows you to select an entire collection of objects and then open a window to see details for them all. In this example, I selected several house sales (the upper dots highlighted in red) and easily accessed the underlying data for the set.



Sometimes the supplementary information that you need doesn't reside in the data you're analyzing, but can be accessed from elsewhere. Details-on-demand functionality should ideally extend to any data that you need, making it simple and efficient to access it when it's useful. If you're examining house sales in the map-based display in Figure 12 and decide to look closer at a particular house, you might want to see an actual photo of the neighborhood and compare its sales price to the values of its neighbors. The ability to select that house and then easily hyperlink to a Web-based service that provides this information is a natural extension of the details-on-demand concept. Figure 13 shows the results of this simple process, which I was able to view and then easily switch back to the original map without missing a beat.

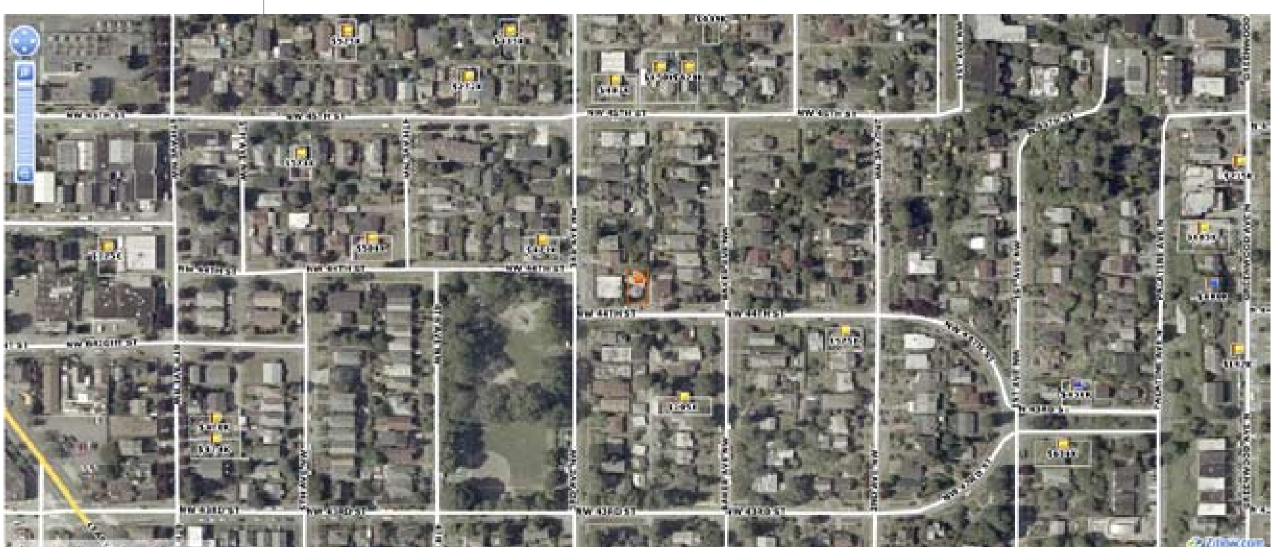


FIGURE XII: Additional information that resides elsewhere can be linked to with ease, such as this satellite photo with home prices from www.zillow.com.

FINAL WORD

Visual analysis software can speed up the process of analysis and open your eyes to discoveries and insights you might not ever experience otherwise, but only if it is well designed. I've presented several characteristics of well designed visual analysis software in this paper, which all fit into one of three categories:

- Effective visualizations
- Effective interactions
- Effective navigation

To take full advantage of your powerful eyes and pattern-seeking brain, good visual analysis software uses visual representations that present what's meaningful in the data in ways that your eyes can see and brain can understand. It should also design those visualizations in ways that keep you focused on what's important and never distracted by visual content that doesn't add to your understanding of the data. Good visual analysis software provides every means to interact with the data that you frequently need (filtering, sorting, etc.), and allows you to perform each action conveniently, without losing sight of the data. And finally, good visual analysis software makes it easy to navigate through the analytical process, from view to view, interaction to interaction, overview to detail, riding the wave of thought smoothly throughout the process. Don't settle for anything less.

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 consulting and training 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*.

ACKNOWLEDGEMENT

Images used in this paper are courtesy of Tableau Software, Inc.

Tableau is Business Intelligence software that helps people visually explore and investigate the information in databases. We call this process "visual analysis." It enables people to transform raw data into smart decisions using a drag and drop process that creates vivid, interactive visualizations.