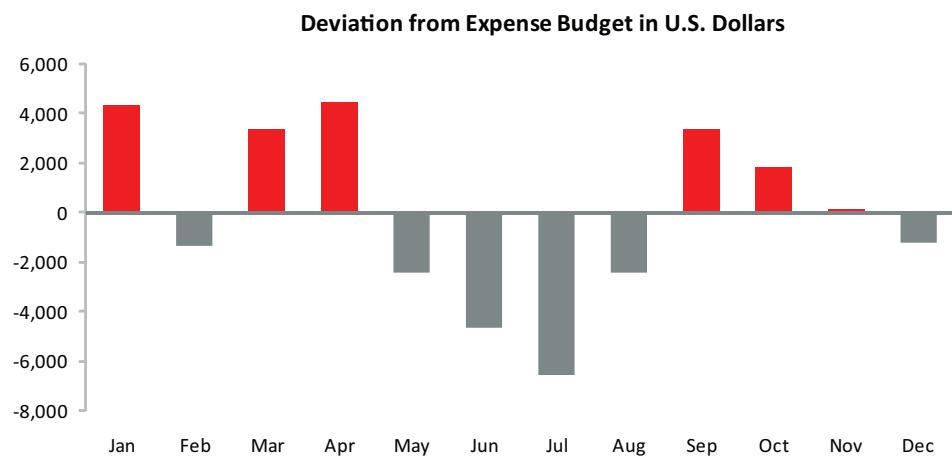
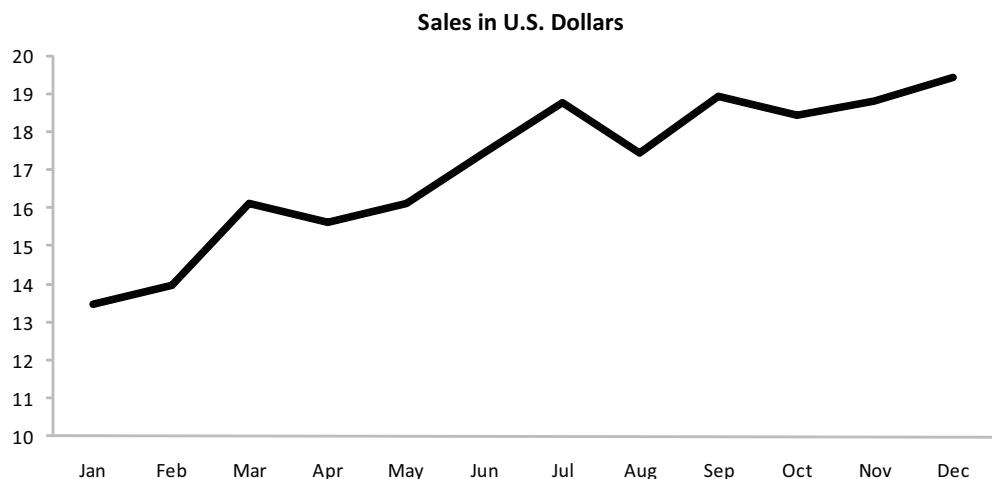


Introduction to Geographical Data Visualization

Stephen Few, Perceptual Edge
Visual Business Intelligence Newsletter
 March/April 2009

The important stories that numbers have to tell often involve location—where things are or where they've occurred. When we display quantitative information on a map, we combine visual displays of both abstract and physical data. Quantitative information is abstract—it doesn't have physical form. Whenever we represent quantitative data visually, whether on a map or otherwise, we must come up with visual objects that represent abstract concepts in a clear and understandable manner, such as "sales are going up," represented by a line, or "expenses have deviated from the budget in both directions during the course of the year," represented by bars extending up or down from a baseline of zero.

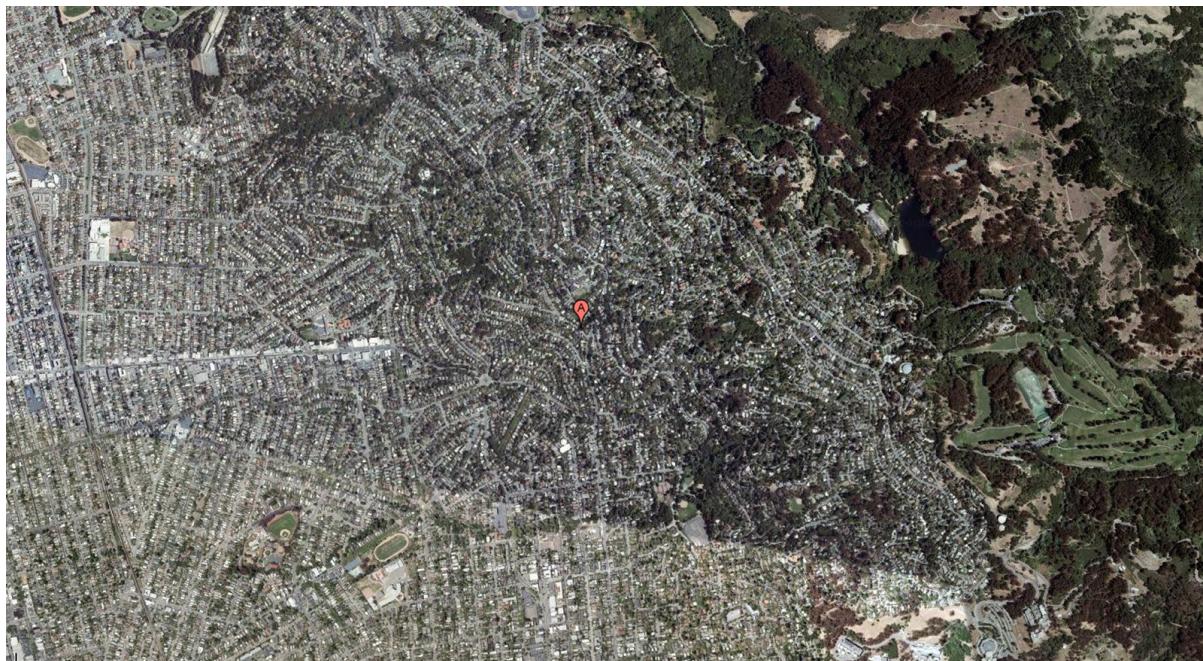


Geographical information on the other hand is physical. When we display it, we do our best to represent those physical characteristics of land masses, bodies of water, terrain, roads, and so on, that concern us. On the following map, the land masses and bodies of water represented are physical, but the political boundaries and the red circles, which represent Internet usage in 2005, are abstract.



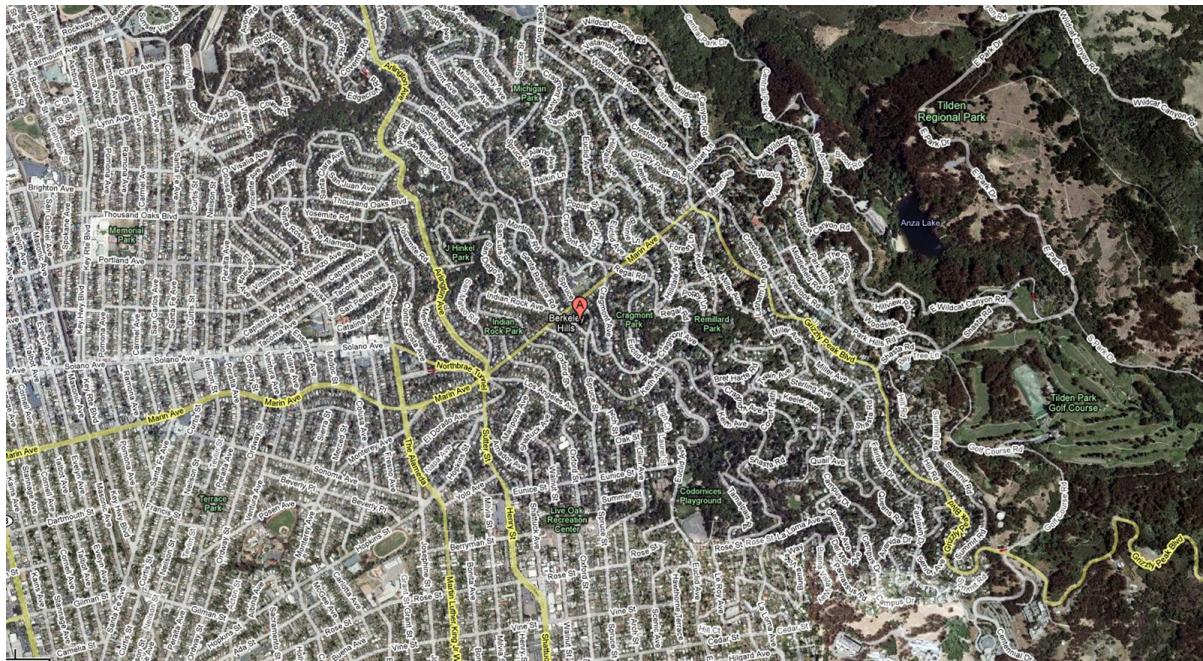
We represent physical reality on a map in ways that leave out those aspects that aren't pertinent to the task at hand. Otherwise, rather than a map, we would rely entirely on satellite photographs, which provide an accurate 2-D representation of geography.

The red icon labeled "A" below marks the location of my house in Berkeley, California, as photographed from a satellite and displayed by Google Maps.



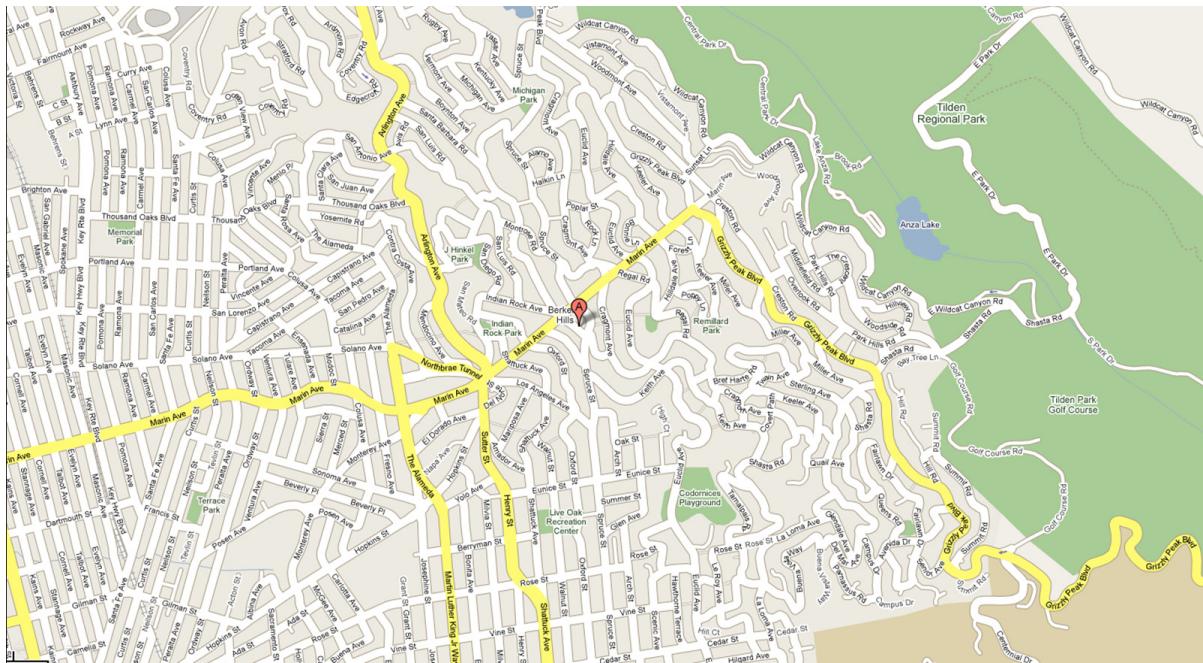
Aerial photographs of a geographical region work well for some purposes but poorly for others. While viewing my community in this way, my eyes are drawn to the expanse of open space on the right, which is Tilden Regional Park where I often hike. My eyes are also drawn to the two athletic fields in the lower left. Looking further, I notice that there appear to be more trees in the hills where I live than in the flatlands a short walk down the road. What I can't get from this view, however, is a clear sense of where I live in relation to familiar streets.

In the view below, streets and their names have been overlaid on top of the photo to make them easier to see and identify.



With this view on my large computer screen, I can now get a better sense of where I live in relation to other street locations that are familiar, but I have to work fairly hard, because there are too many details in the photo that distract me from the information that I currently need.

In the next view, this problem has been corrected by switching to Google's map view.

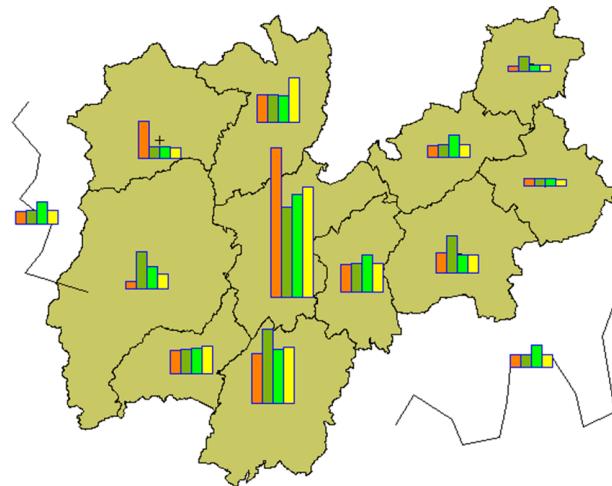


Now details that don't concern me have been removed and the features that I care about—mostly the streets—have been abstracted from their true physical appearance and simplified in a way that allows me to see only what's useful, without distraction from anything that isn't.

Different representations of geography are required for different tasks. Cartographers spend years learning how to design maps to specifically and effectively support their many uses. When we add quantitative information like sales or levels of Internet usage to a map, we must take care just as cartographers do to design an effective display for the task at hand.

Methods for Encoding Quantitative Data on a Map

Many of the visual objects that represent data well in graphs, such as bars and lines, do not work well on maps. In the following example, it isn't easy to compare the lengths of bars except for those that are immediately next to one another.

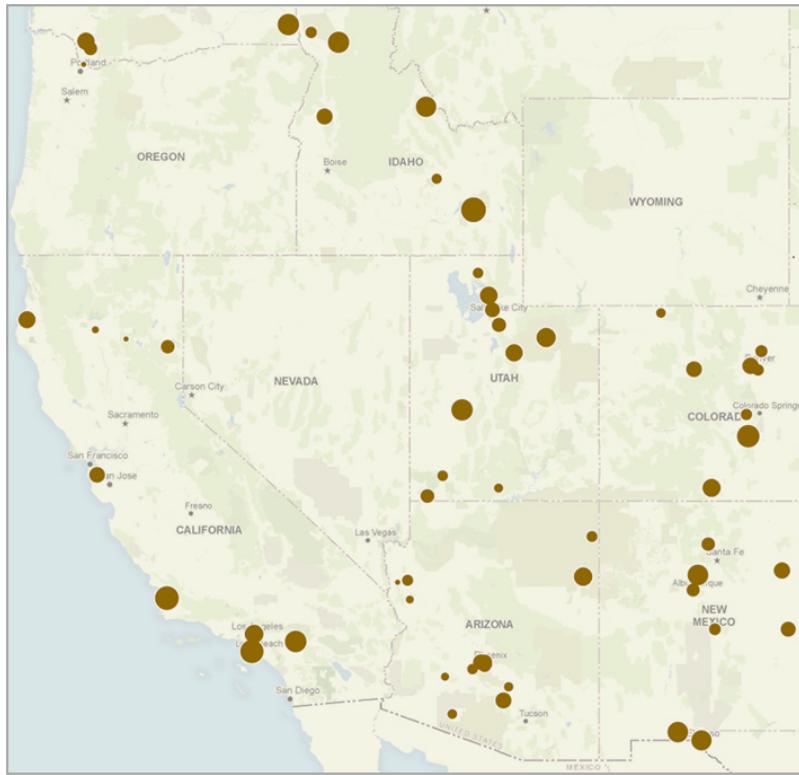


Bar graphs rely on our natural ability to compare the lengths of objects such as bars by arranging them side by side along a common baseline. Because the sets of bars that appear on this map are positioned in various geographical locations and therefore don't share a common baseline, our ability to compare values is impaired. Besides this problem, on many maps there simply wouldn't be enough room to place bar charts everywhere they're needed without overlapping them, which would make many unreadable due to occlusion.

Two approaches to displaying quantitative information on maps usually work best: variations in color intensity, in size, or both. The map below, which I borrowed from Gretchen Peterson's excellent new book *G/S Cartography: A Guide to Effective Map Design*, illustrates the use of color intensity for displaying quantities. Imagine that this is a country divided into provinces, and that various intensities of the color orange are being used to encode average household income—the darker the color the greater the income. This approach displays an aggregate measure for each province rather than a measure for each household. Geographical displays of this type are called choropleth maps.



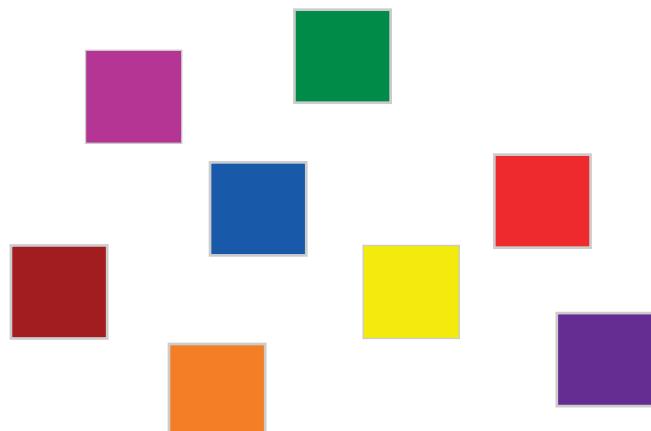
The next map below uses circles that vary in size to encode differences in value—the larger the greater. Imagine that each circle represents a retail store and that their sizes indicate the amount of sales at each.



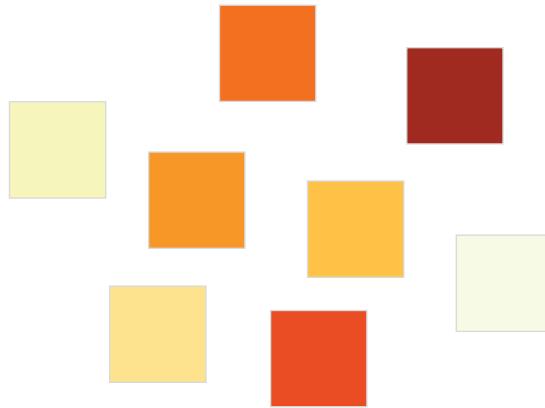
In this case, rather than displaying an aggregated value per state, each circle marks the location of an individual store. This design allows us to see a level of detail that would be lost had we color-encoded entire regions as done on the choropleth map. Both are valid approaches; they simply serve different needs.

On this map, if we wanted to see sales aggregated to the state level instead of individual stores, we could color code each state or we could instead display circles (or some other simple symbol) of various sizes, one per state. This is often the better method because on choropleth maps, because large areas of color stand out more than small areas of the same exact color, we tend to notice them more, even though both have the same value.

Before we depart from this topic, let me mention another guideline to keep in mind when encoding quantitative values as color. Avoid the use of rainbow colors—several distinct hues—when displaying a quantitative range. Looking at the eight colors below, assuming that they represent different quantitative values, try to put them in order from least to greatest.



We don't perceive distinct hues as ordered—they're simply different. Hues like those above work great for separating items on a map or graph into different groups, such as blue for Democrats and red for Republicans, but they don't work for the expression of quantitative differences. The following colors, alternatively, do the job nicely.



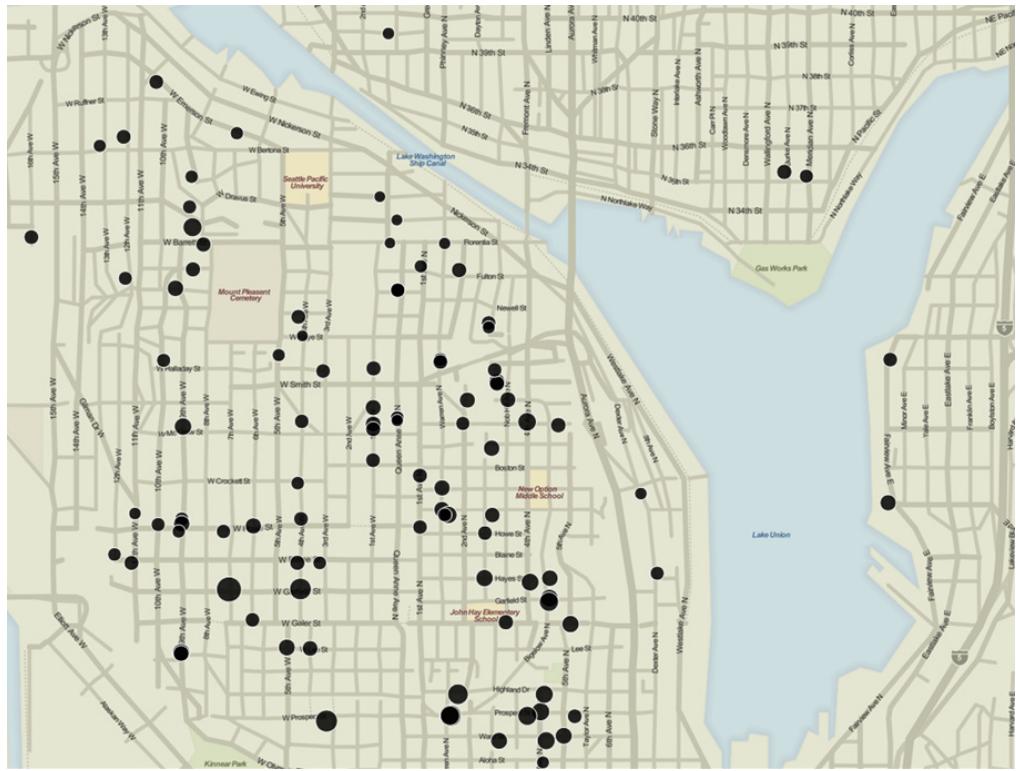
These colors work because they vary not by hue but by intensity from light and lowly saturated to dark and fully saturated, which we intuitively perceive as ordered.

Methods for Featuring the Data, Not the Map

When we display quantitative information on a map, the map should only include those geographical features that are needed to provide meaningful context for the data. Maps that we use for driving directions, thanks to the free services of Google, MapQuest, Yahoo, Microsoft, and others, are not ideally designed for most data displays. The information that's useful for driving is rarely useful for displaying quantitative data. If your only means of displaying data geographically involves mashups on a map that was designed for driving, you won't feature the data effectively, as you can see in the example on the next page.

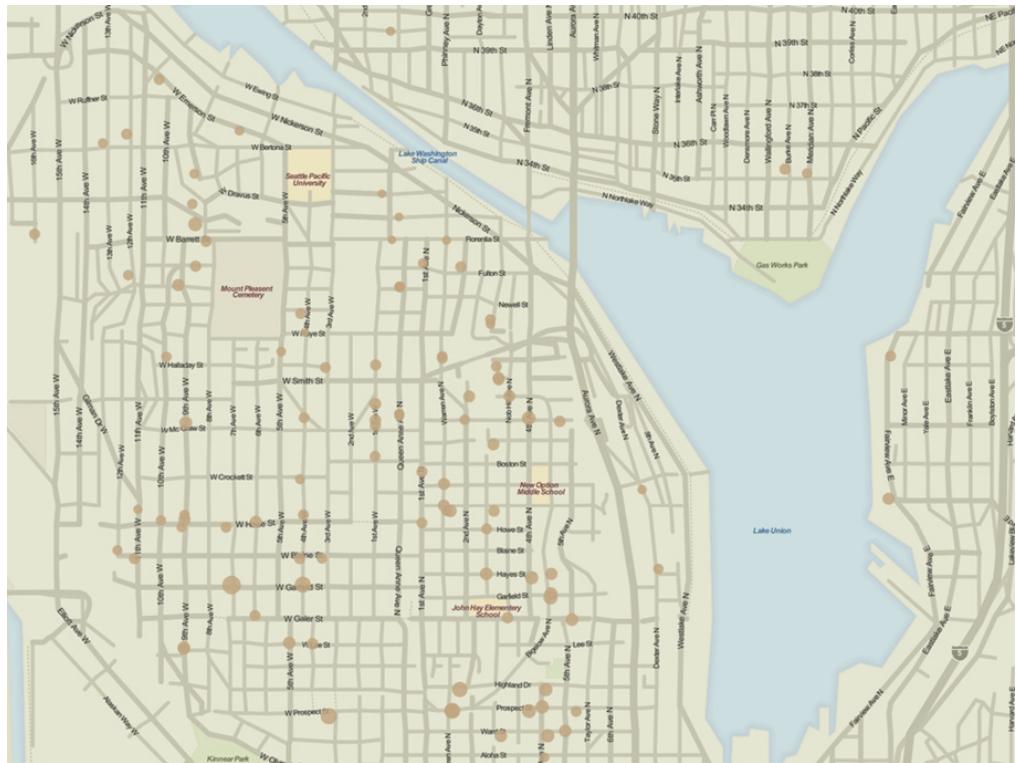


The example on the next page, which I created using Tableau Software, is a good example of a map that was thoughtfully designed for displaying data. Notice how the geographical features have been pared down and visually subdued to allow the data to stand out.

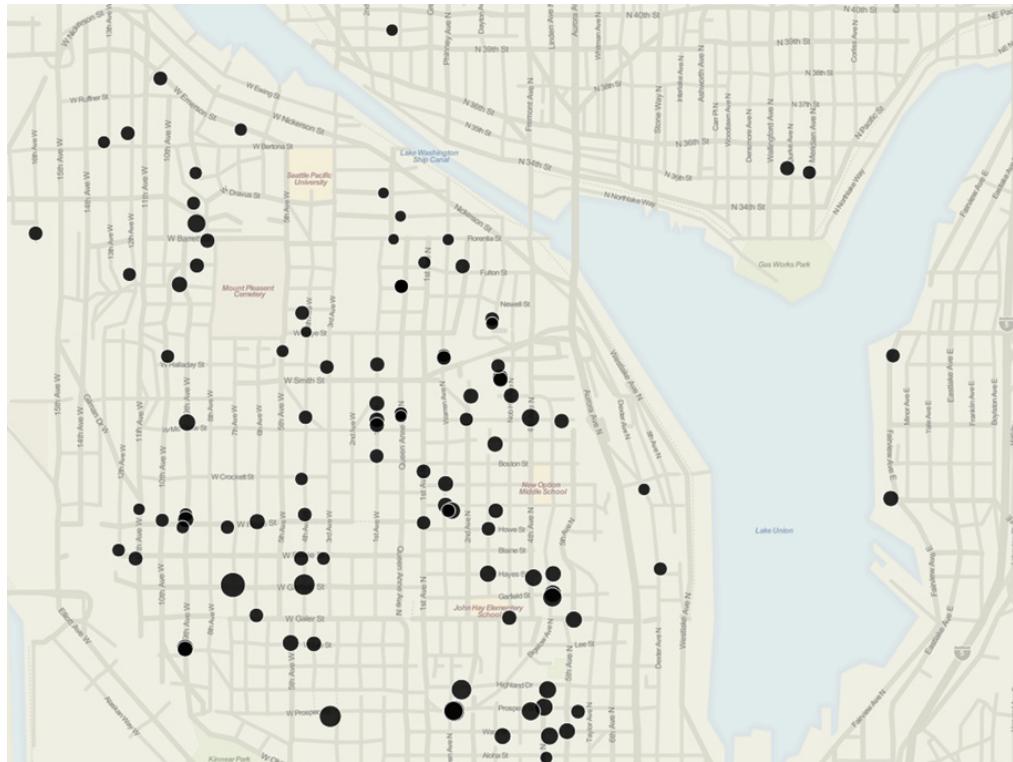


If you use software that provide maps that have been designed for this purpose, along with control over the inclusion or omission of specific features such as zip code or area code boundaries, you have the means to design a good geographical data display.

Even with a good data visualization tool you can still create an ineffective display by making bad design choices. In the following example, which displays the same information as above, the data fails to stand out because the data points are too light and their color is too similar to the geographical features of the map.

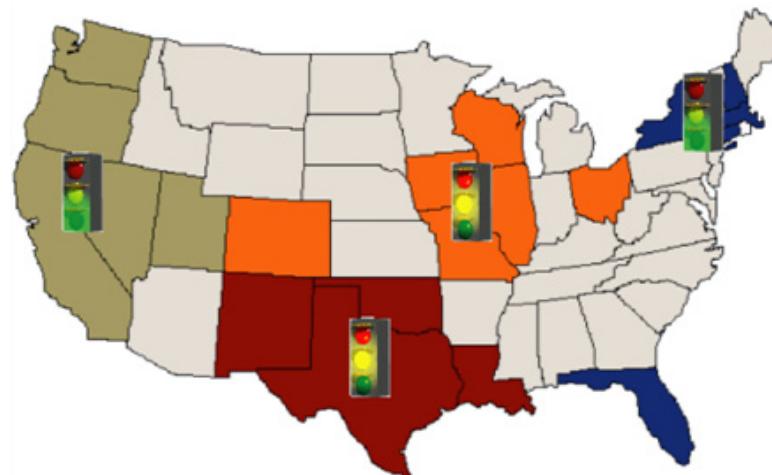


By choosing a lighter color—one that's similar to the colors of the map—I've caused the data values to fade somewhat into the background, making them difficult to see. Besides choosing colors for data that are darker or brighter than the other colors on the map, we can also cause the data to stand out more by pushing the map further into the background. In the following example, I've used a feature available in Tableau to "washout" the map, causing it to become less salient, but not so much less that it doesn't remain plenty visible to do its job.

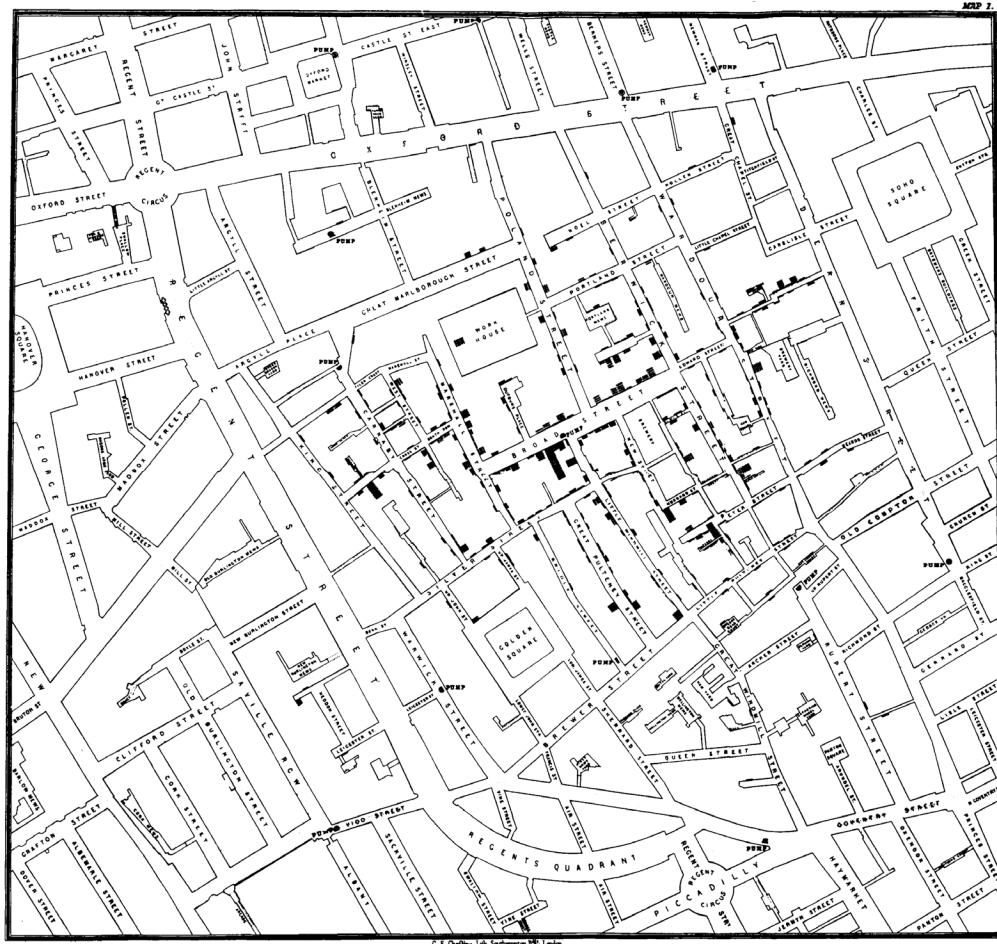


Good Uses of Geographical Data Displays

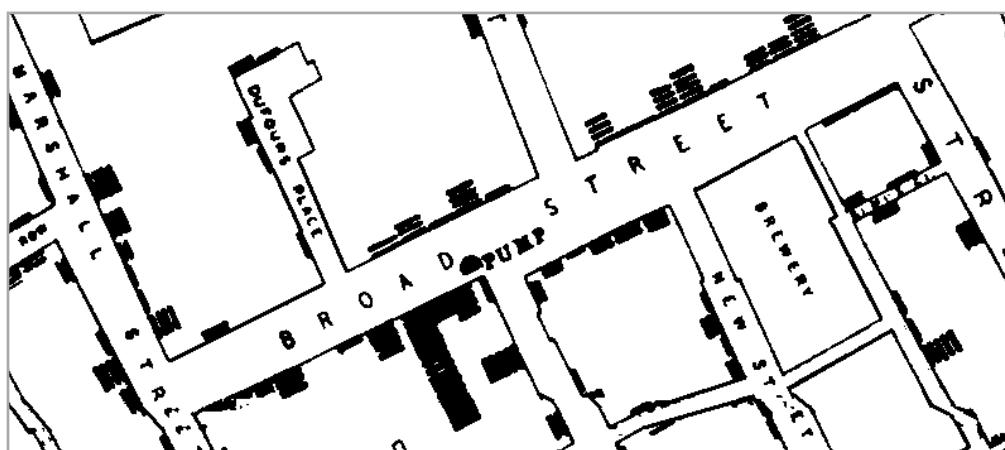
Just because you can display data on a map doesn't mean you should. With the increased availability of affordable mapping software today, it is tempting to throw everything onto a map, but it's only useful when location is an important part of the meaning you're trying to discover or the story you're trying to tell. The following map displays sales information for four regions much less effectively than a simple table or bar graph with regional labels such as "West," "North Central," "South Central," and "East."



In contrast, the following example was incredibly useful when it was drawn back in the 1860s by John Snow, a medical doctor who worked quickly and diligently to discover the source of a deadly cholera epidemic in his London neighborhood.



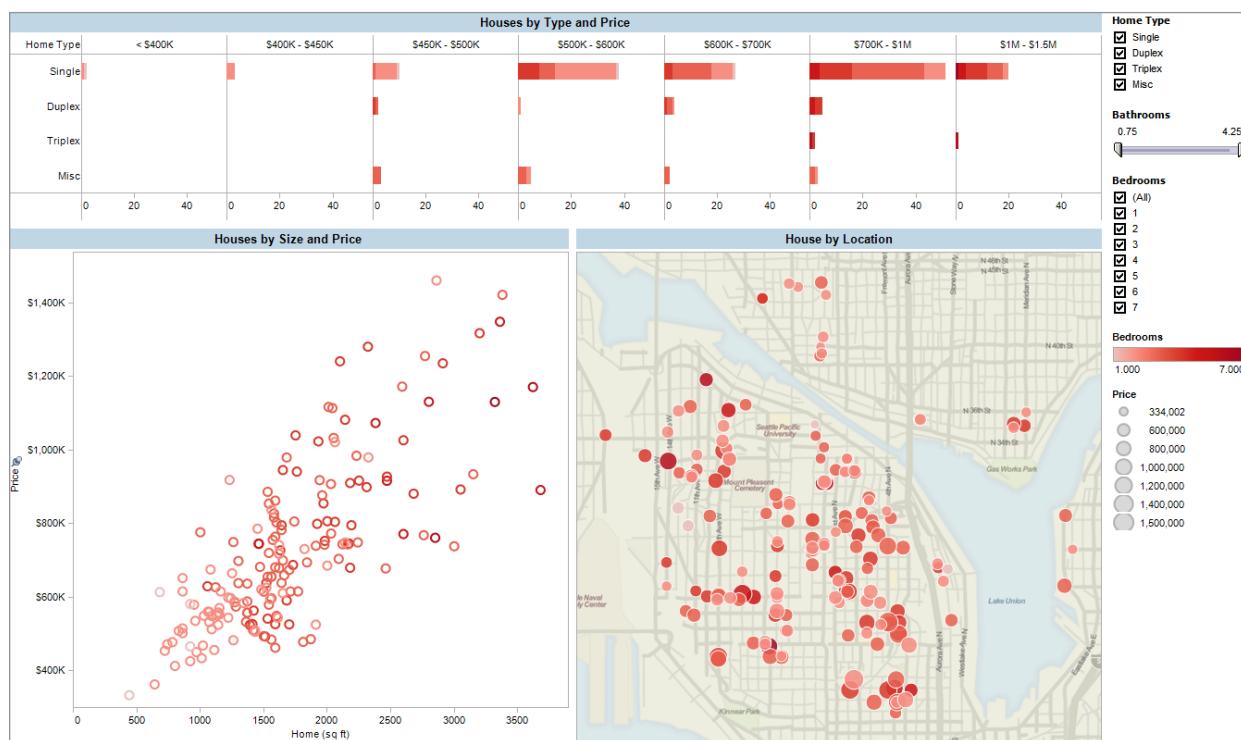
Snow suspected that, contrary to the “miasma theory” of cholera transmission, which assumed that it spreads through the air, the disease was actually transmitted through the water supply. Like any good analyst, he decided to test his hypothesis, and did so by going door to door throughout the neighborhood to record every location where someone had died from cholera while the epidemic was still occurring. He recorded the number of deaths that occurred at specific locations using bars. He also recorded the location of every well where people drew their supplies of water, which he marked with the word “PUMP.”



By displaying this information on a map, he was able to easily see that most deaths occurred close to the Broad Street pump. Not every death, however, so he investigated the outliers to determine if they contradicted the likelihood that the Broad Street well was the epidemic's source. While doing so he heard accounts such as the story of an old woman who had previously lived near the Broad Street well, but after moving closer to another, didn't like the taste of its water as well, so she had her son's make frequent visits with buckets of water from the Broad Street well to maintain her drinking supply. Based on the evidence that he gathered, Snow convinced the authorities to remove the pump handle from the Broad Street well, and within a short time the epidemic abated. His geographical data display literally made a life or death difference for many who might have otherwise succumbed to the disease. No one would argue that this wasn't a meaningful use of a geographical display.

Powerful Collaborations between Maps and Other Visual Displays

Even though geographical displays like Snow's cholera map can be quite powerful by themselves, their analytical use is magnified when used in collaboration with other quantitative displays. Many data discoveries can only be made when you view data from several perspectives, not one at a time, but simultaneously. In the following example, three distinct views of house sales have been arranged on the screen together for easy comparison.



Each circle in both the scatterplot and on the map represents a house that sold during a particular span of time. In all three views, the varying intensity of red represents the number of bedrooms from one (light red) to seven (dark red). On the map, the varying sizes of the circles represents sales price. These three views allow us to see several dimensions of house sales at once: sales price, comparable price categories, house size in square feet, number of bedrooms, geographical location, and home type (single, duplex, triplex, and others). Each view is designed to feature particular dimensions in a manner that is simple and easy to read (for example, size and price in the scatterplot). Because these views are not just arranged simultaneously on the screen, but also linked for coordinated interaction, we could use them to discover relationships between these dimensions by interacting with the data, such as by filtering out or highlighting the same set of houses in all three views through a single action. The integration of well designed geographical data displays with graphs in a coordinated manner adds a critical perspective to our analyses when geographical location is a meaningful part of the story. As data analysis products make it easier and easier to do this using thoughtfully designed maps and powerful data interactions, we'll be able to add a clear view of *where* to the *what and when*, which we can already explore relatively well.

Further Reading

Brewer, Cynthia A., *Designing Better Maps: A Guide for GIS Users*, ESRI Press, 2005.

Peterson, Gretchen N., *GIS Cartography: A Guide to Effective Map Design*, CRC Press, 2009.

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 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.