

Practical Rules for Using Color in Charts

Stephen Few, Perceptual Edge
Visual Business Intelligence Newsletter

February 2008

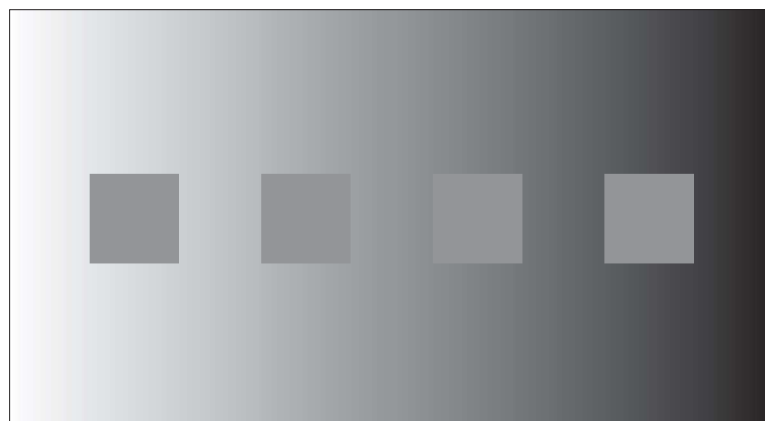
We can use color in powerful ways to enhance the meaning and clarity of data displays, but only when we understand how it works, what it does well, and how to avoid problems that often arise when it's used improperly. My expertise in color, which focuses almost exclusively on its proper use in data displays, is practical in nature. My grasp of color theory—a fairly complex field of study—does not run deep. This article is intended for people like me who must understand color insofar as it applies to quantitative data displays. We are concerned with the practical use of color, not its use for artistic expression. Becoming familiar with a few simple rules about color and its use will serve us well.

Color in Context

Let's begin with an important fact about color perception that must be understood to use color effectively. Like all aspects of visual perception, we do not perceive color in an absolute manner. Our perception of an object is influenced by the context that surrounds it. In other words, although our eyes sense absolute wavelengths of light, which we call color, we perceive an object's color in contrast to the one or more colors that surround it. Visual perception is relative, not absolute. I'll illustrate this using the small gray square below:



I will now place four copies of this square at various locations in a large rectangle, which varies in color from white on the left along a progressive gradient of gray until it reaches pure black on the right. Notice how different the small square looks at each location.



No, I am not playing tricks. I have done nothing to alter the color of the small rectangle. We perceive them as different from one another, because each is immediately surrounded by

a different shade of gray. When surrounded by light gray on the left, the small rectangle appears darker than it does when surrounded by dark gray on the right.

So, what does this mean in practical terms? Two practical rules emerge from this observation:

- Rule #1** If you want different objects of the same color in a table or graph to look the same, make sure that the background—the color that surrounds them—is consistent.
- Rule #2** If you want objects in a table or graph to be easily seen, use a background color that contrasts sufficiently with the object.

One straightforward application of the Rule #1 to graphs is to avoid using gradients of color in the background or varying the background color in any other way. Don't give into the temptation to decorate a graph in a way that undermines its ability to present data clearly.

Rule #2 cautions us to choose colors carefully, always making sure that they are easy to see and that they effectively serve the purpose for which we are using them. I'll illustrate this point using a display that is becoming increasingly familiar, but is seldom done well. With *Microsoft Excel* and several other software products, you can display quantitative data in the form of a *heatmap*. A heatmap is a visual display that encodes quantitative values as color. We are all familiar with weather maps, which use colors to represent varying amounts of rainfall or degrees of temperature. Heatmaps need not be arranged geographically; they can also be structured as a matrix of cells, such as a tabular arrangement of values in a spreadsheet. Here's a typical example of what I've seen created with Excel:

	Laptops	Desktop PCs	Harddisks	Flash Memory	Screens	Keyboards	Printers	Scanners	PDAs	Projectors	Cameras
California	-2,216	4,497	884	3,252	8,564	3,418	6,582	-3,891	2,333	1,356	5,450
Colorado	3,410	0	2,338	2,676	1,567	367	1,361	3,249	828	1,272	-141
Connecticut	0	0	4	0	2,998	1,219	673	0	0	610	989
Florida	0	0	583	765	2,305	940	1,737	2,727	0	1,344	497
Illinois	0	0	9,384	3,405	4,362	2,331	4,495	3,464	1,366	0	1,569
Iowa	377	0	412	6,577	210	4,487	872	189	5,622	0	3,466
Louisiana	0	2	1,455	1,501	1,138	0	920	1,349	0	0	991
Massachusetts	0	0	-297	0	12,489	707	0	0	0	492	367
Missouri	0	0	1,224	-169	986	367	1,011	1,125	-210	-45	-685
Nevada	0	875	374	884	-45	11,934	410	210	4,497	-10,978	4,355
New Hampshire	1,009	0	-166	0	897	376	0	0	0	882	-43
New Mexico	0	-208	-161	377	902	0	369	-683	0	0	202
New York	0	0	-17,034	0	8,564	3,254	0	0	3,405	2,326	7,589
Ohio	820	0	2,671	495	827	2,305	355	-134	2,724	0	705
Oklahoma	0	2,722	492	495	1,274	0	2,305	700	-15,034	0	571
Oregon	-139	357	818	708	395	487	2,671	834	2,310	2,727	1,269
Texas	0	1,356	3,415	808	5,452	0	2,336	1,567	0	0	829
Utah	1,130	1,006	1,501	1,009	678	-156	1,222	979	367	-208	904
Washington	0	768	488	2,998	1,661	1,452	1,418	564	918	-5	1,141
Wisconsin	497	0	1,418	1,450	576	916	767	1,740	2	0	1,337

In this example, traffic light colors of green, yellow and red are being used to encode high profits (green), low profits or losses (yellow), and high losses (red) across several product types and states. It is probably true that the values that are of greatest concern to the person viewing them are those in dark red and dark green, but they are the hardest values to read, because there is not enough contrast between black text and dark background colors for the numbers to stand out. I'll come back to heatmaps a bit later. For now, let's move on to other rules for using color in data displays.

Use Color Meaningfully and with Restraint

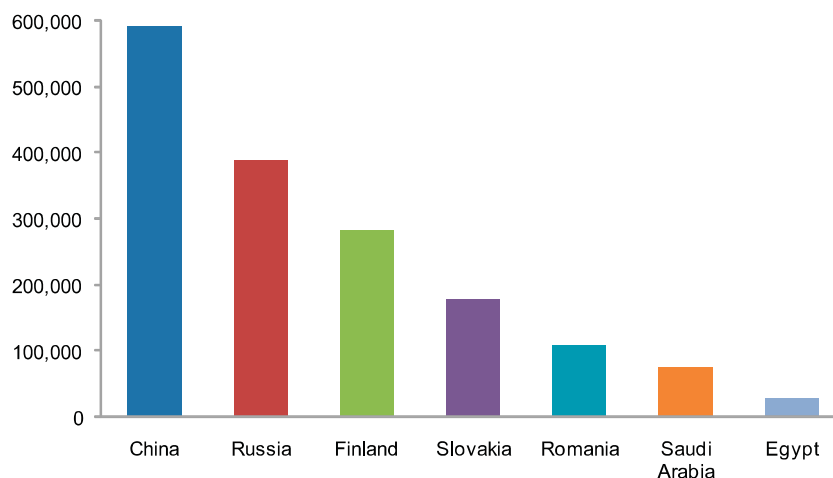
I have a friend, Maureen Stone, who is an expert in color. She was once kind enough to write a [guest article](#) for my newsletter to introduce basic color theory and how it applied to data displays. Once, while chatting casually at a conference, she and I were bemoaning the fact that color is often applied haphazardly to data displays, tragically undermining their value. Regarding color, people tend to think that more and brighter is better. To illustrate the point that the opposite is true, Maureen mentioned that conferences dedicated to color and its use for information displays, exhibit great restraint in their use of color during presentations. For example, their PowerPoint slides tend to include grays far more often than you'll typically see elsewhere. These experts realize that color should be used meaningfully, not arbitrarily or gratuitously. Here's our next rule:

Rule #3 Use color only when needed to serve a particular communication goal.

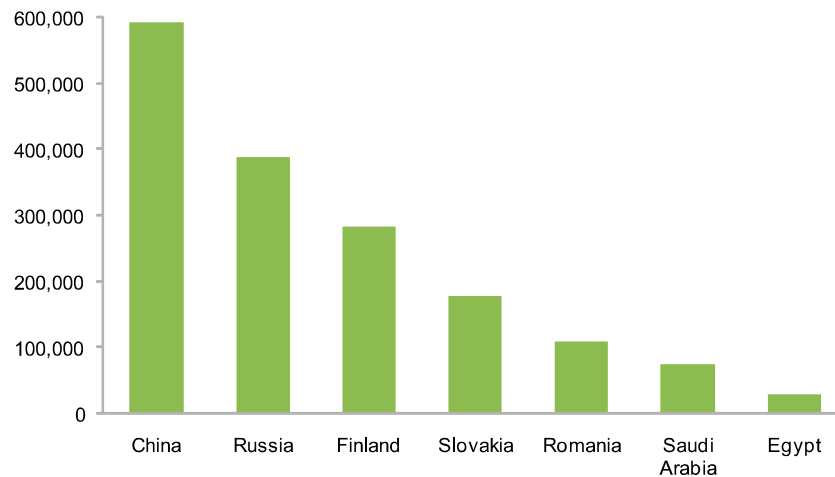
We should only add color to an information display to achieve something in particular—something that serves the goal of communication. Don't use color to decorate the display. Dressing up a graph might serve a purpose in advertising, but it only distracts people from what's important—the data—in an information display.

Rule #4 Use different colors only when they correspond to differences of meaning in the data.

The following graph illustrates one version of what we should avoid:



What do the different bar colors in this graph mean? Not a thing. The labels along the X-axis tell us what the bars represent. The colors add no meaning or value, but their presence suggests that they do. Consciously or not, when people look at a data display and see visual differences, they try to determine the meaning to those differences. Suggesting meanings, which aren't there, wastes people's time, prompting them to expend mental resources in a fruitless pursuit. This graph is supposed to help people compare sales amounts associated with various countries. However, notice how much more your eyes are encouraged to compare the bars when they look alike, as shown on the following page, than when they look different, as shown above.



Whenever you're tempted to add color to a data display, ask yourself these questions: "What purpose will this color serve?" and "Will it serve this purpose effectively?" If the answer is "It serves no useful purpose" or "It serves a purpose, but something other than color or this particular color would do the job better," avoid using it.

Define Standard Palettes of Colors for Particular Purposes

Nature uses color more wisely than most people do in tables and graphs. In nature, what purpose do the bright colors of flowers serve? They attract pollinators—birds and bees—to the flowers so they'll spread the pollen and thereby propagate the species. This same strategy applies to data displays as well. When properly used, color can powerfully draw attention to particular data.

We can use color effectively for three fundamental purposes in a data display to promote communication:

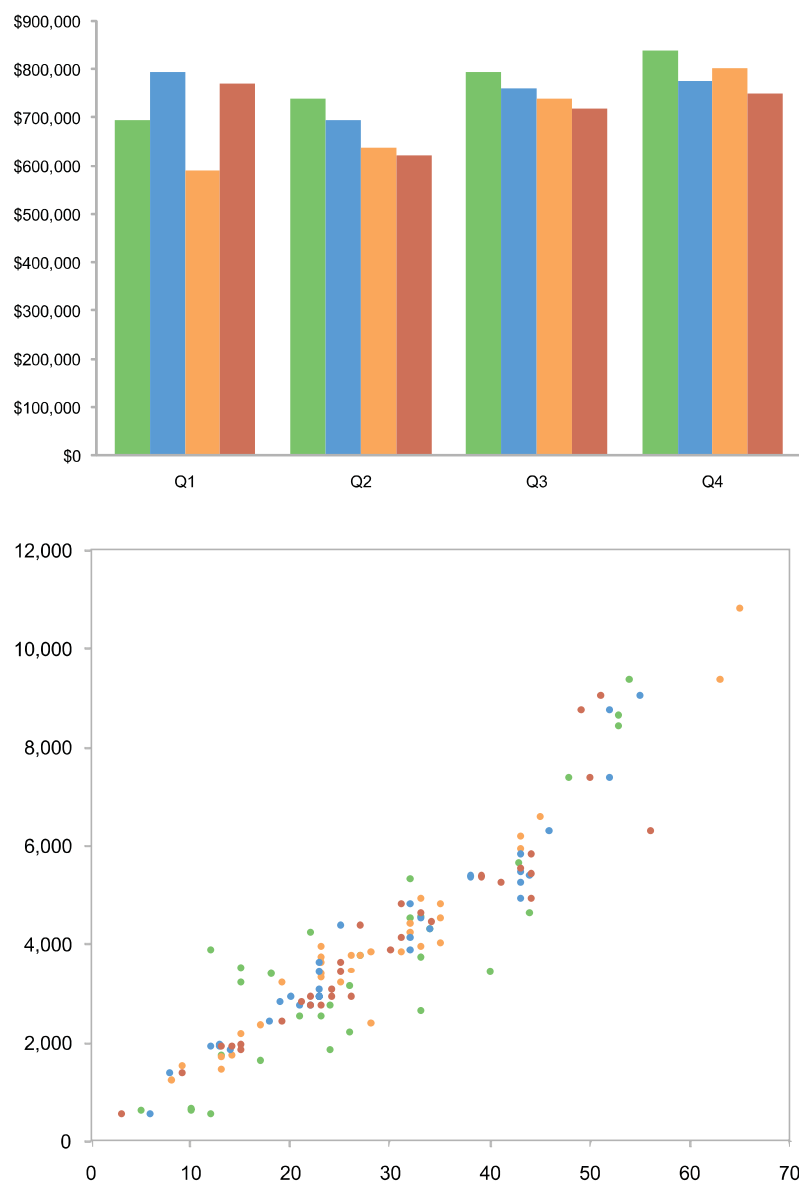
- To highlight particular data
- To group items
- To encode quantitative values

We're now ready for the next rule:

Rule #5 Use soft, natural colors to display most information and bright and/or dark colors to highlight information that requires greater attention.

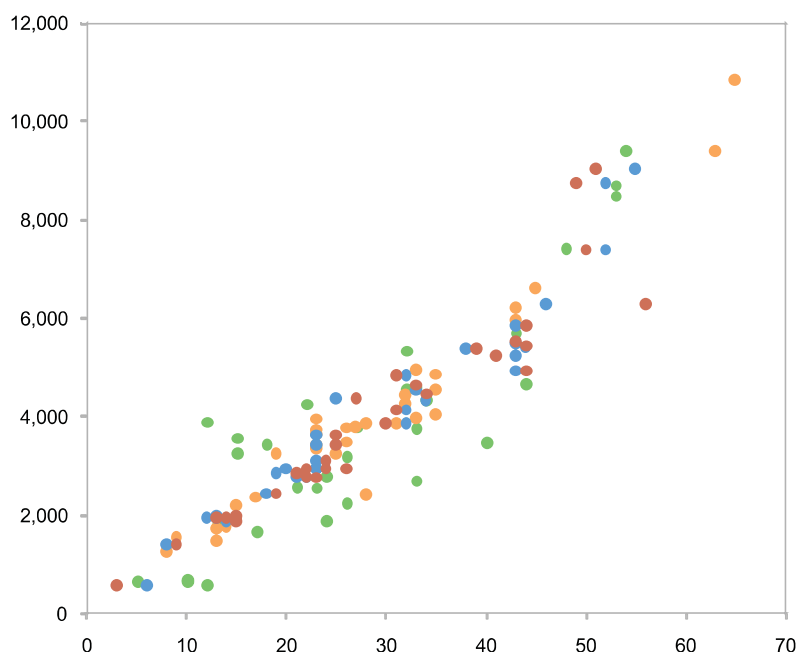
Rather than taking time to select a color from a huge list of possibilities each time we need one, time will be saved and the effectiveness of our choices will be improved by standardizing on a few good colors. I recommend that you standardize on a few color palettes, each designed for particular purposes. I maintain one palette of bright, dark colors, another of medium shades that are easy on the eyes, and a final set of light, pale colors. The different levels of visual salience that these palettes represent serve particular purposes.

Bright, dark colors are primarily useful for highlighting data in graphical displays, such as a particular set of bars, and they also come in handy for text. Colors are harder to see and tell apart when objects are small or thin, such as text, thin lines, and small data points. Notice in the two graphs below, that medium shades of color, which work well for bars, are harder to see and distinguish when used for small data points.



When objects are small or thin, colors that we use to encode them must sometimes be brighter and/or darker than otherwise necessary. For this reason, I sometimes choose from my palette of bright/dark colors for encoding thin lines or small data points.

Another solution that can often allow us to stick with softer colors for lines and data points is to simply thicken the lines or enlarge the data points, as shown here:



I use the palette of light, pale colors for those parts of tables and graphs that don't display data, such as the axes, and for data that I want to de-emphasize. Not all information is created equal. Sometimes information must be included that plays only a minor role in the message you're trying to communicate.

When color is used to organize data into separate groups (for example, a different color per sales region), we want those colors to look very different from one another to clearly delineate the groups. We want them to be distinct from one another, but not in a way that makes some stand out more than others. They should look different without varying in perceived intensity. For this purpose, we need a collection of hues that are as distant from one another as possible along the color spectrum, but similar in intensity (that is, from light to dark and pale to bright). Here's a palette of eight sample hues that fall in the medium range of intensity:



Notice that these colors are not only easy to distinguish, they are also well balanced. No one color pops out more than the others. Here's a bright, dark palette of the same hues (technically, black is not a hue, but the complete absence of color), which could be used to highlight data:



Finally, here's a light, pale palette that is once again based on the same hues.



If we choose colors for our standard palettes wisely, there is rarely a need to look further. Choosing wisely is the key, and you don't need to be an expert in color or take the time to do extensive usability testing to achieve this end. Instead, you can rely on experts who provide predefined palettes via the Internet for free. Cynthia Brewer of Penn State University is one such expert. She provides many well-designed color palettes through a simple Web-based application called [Color Brewer](#).

If you take a look at Color Brewer, you'll notice that Cynthia divides her palettes into three types:

- Categorical
- Sequential
- Diverging

Categorical palettes are the ones that are used to separate items into distinct groups. Sequential and diverging palettes are primarily used to encode quantitative differences. Quantitative values are arranged sequentially, from low to high or high to low; they vary in a continuous manner, increasing or decreasing from one end to the other in perceptually equal increments. Understanding this leads us to our next rule:

Rule #6 When using color to encode a sequential range of quantitative values, stick with a single hue (or a small set of closely related hues) and vary intensity from pale colors for low values to increasingly darker and brighter colors for high values.

We can intuitively perceive colors of increasing intensities as numbers of increasing values. We don't, however, perceive differences of hue in this manner. To see for yourself, try putting the following hues in order from low to high.

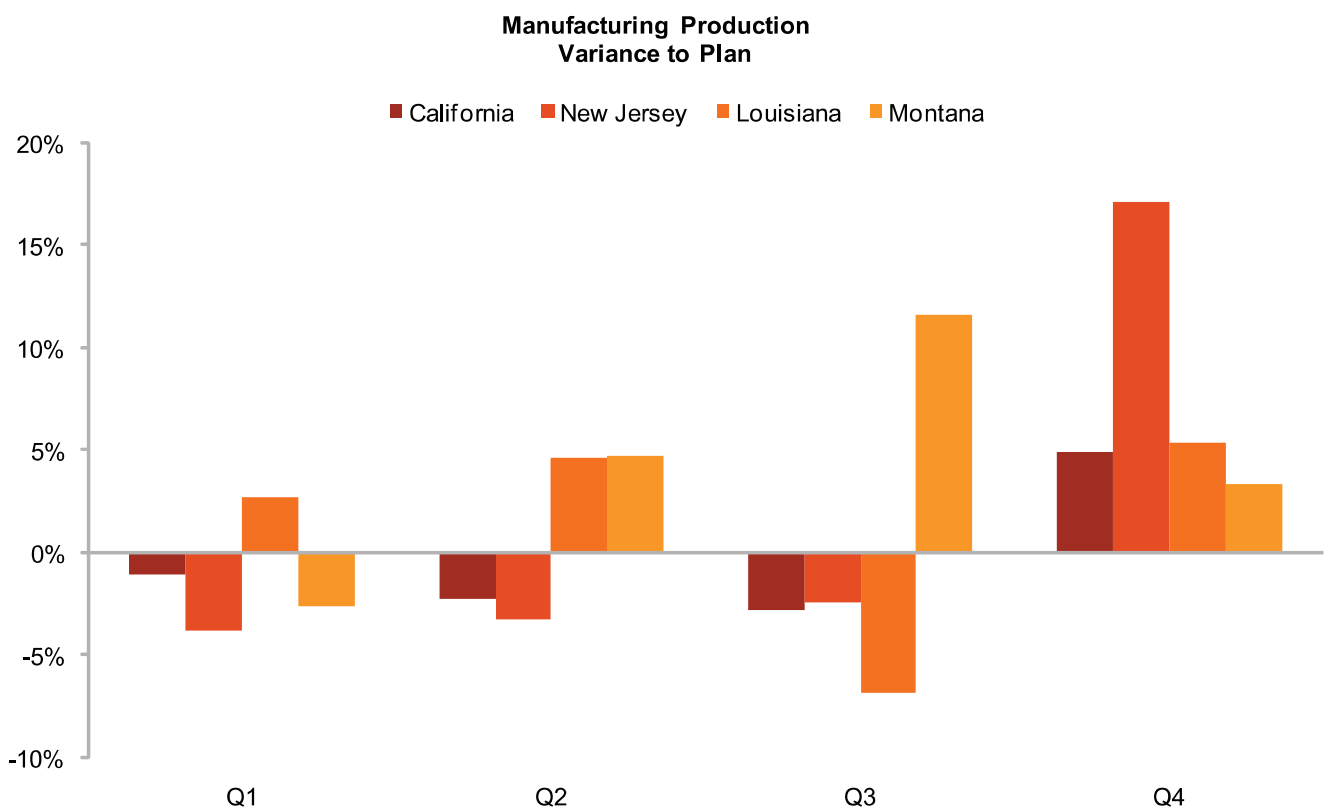


If you've studied color theory and taken the time to memorize the order of hues along the spectrum of light, you might have been able to order the hues above by wave frequency, but

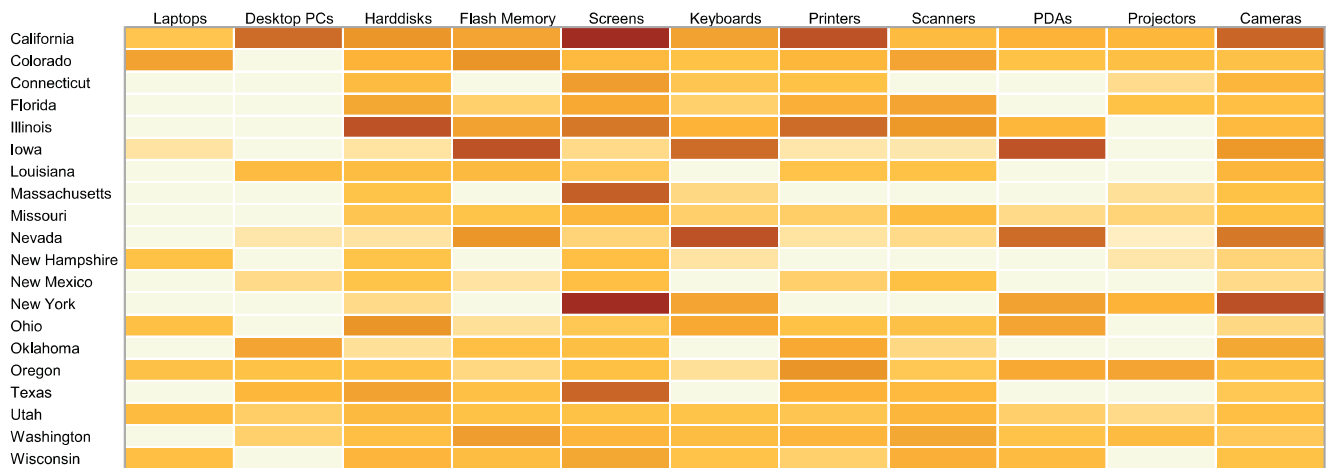
not because you intuitively perceived them in this order. Now, if you try to place the following colors in order from low to high, you will have no trouble nor require any training.



We could meaningfully assign these colors to the four sets of bars in the following graph to encode the order in which four manufacturing facilities performed relative to the production plan during the course of the year as a whole, from the worst (California, with the darkest set of bars) to the best (Montana, with the lightest).



We could also use a set of ordered colors in the form of a heatmap to encode a large number of sales values in a way that could be used to find out which products in which states sold the best, which sold the worst, and which states or products did the best or worst overall, as seen on the following page.



Notice that I did not display the text values in the cells of this heatmap. I did this, not only because some of the text would not stand out well in contrast to the heatmap colors, but also because when scanning colors to spot patterns in the data, we don't want to be distracted and slowed down by reading text. Keeping the numbers turned off (invisible) while exploring the data works best. If we want to look up precise values at times, we could keep a second copy of these cells with the text values visible and no background fill colors nearby in the spreadsheet for this purpose. If we only need a precise value rarely, and only one at a time, we wouldn't need the text values to be visible at all, but could simply click on the heatmap cell that interests us and read the value in Excel's formula bar.

It's useful to maintain several standard palettes of ordered colors, such as the sample below:



A single set of ordered colors, which progresses in perceptually equal intervals, such as the one above, is what Cynthia Brewer calls a sequential palette. But what does she mean by a *diverging* palette? The answer isn't complicated. A diverging palette is one that consists of two sets of sequential colors—one that increases in intensity from the middle of the palette going upwards and one that increases from the middle going downwards. Here's an example:



Diverging palettes, which I prefer to call *dual-ordered palettes*, encode a range of quantitative values that are above or below some logical breakpoint in the middle. For example, a company's profits can be displayed using a heatmap with a dual-ordered palette of colors: one set for profits and another for losses, with a zero midpoint. Another example that could be displayed as a heatmap using a dual-ordered palette involves comparisons to an average, such as the percentage difference for each U.S. state between actual incidences of violent crime and the national average. Essentially, this is just another case of

positive and negative values, with zero in the middle as the national average.

Because dual-ordered palettes consist of two ordered palettes, we don't need to separately maintain standard palettes of dual-ordered colors in addition to those that are ordered. We can easily construct a dual-ordered series of colors from two ordered palettes when one's needed.

Choose the Best Colors for Each Component

Tables and graphs are composed of some objects that represent actual data and some that don't. For example, most graphs use bars, lines, or data points to display values. These data objects should dominate our attention when we examine graphs. Other objects also encode information that is less central to the task and more supportive in nature, such as text (for example, a graph's title, axis titles, and labels along the axes, which provide quantitative scales and names for the things being measured) and tick marks along an axis, which tell us precisely where specific values are located along the scale.

In graphs, the most common non-data objects are axis lines (excluding tick marks and labels, which do encode data), background fill colors, and borders (for example, around the plot area or legend). Particular components of a graph, whether they encode data or not, usually work best when displayed in particular ways. Here's the next rule, which addresses this matter:

Rule #7 Non-data components of tables and graphs should be displayed just visibly enough to perform their role, but no more so, for excessive salience could cause them to distract attention from the data.

The colors we use to display non-data components should usually be light, so they carry little visual weight (that is, salience). Here are a few best practices to keep in mind:

Component Default Color

Axis lines	Use thin gray lines of medium intensity.
Borders	Whether around the graph as a whole, its plot area, or a legend, when borders are needed (and usually they are not), use thin gray lines of medium intensity.
Background	Use white (or in Excel select "None" for color).

We want data components to dominate the landscape and be easily seen. Here are a few suggestions for a good set of default colors:

Component Default Color

Bars	Use a distinct hue of medium intensity for each data series.
Lines	For thin lines, use a distinct hue of fairly high intensity for each; otherwise, use distinct hues of medium intensity.
Data Points	For small points, use a distinct hue of fairly high intensity for each; otherwise, use distinct hues of medium intensity.

A Few Remaining Rules

Simple rules about the following two additional topics remain:

- Deficiencies in color perception
- Effects of light and shadow

Deficiencies in Color Perception

Some people cannot distinguish particular colors that most people distinguish quite easily. It might surprise you to learn that approximately 10% of males and 1% of females suffer from abnormal color perception commonly known as *colorblindness*. The vast majority of these folks cannot distinguish red and green. This means that when we follow the popular convention today of color coding “poor” values as red and “good” values as green, we are displaying this important distinction in a way that many people cannot discern. Although it won’t solve the problem for everyone, the following rule does the job for most:

Rule #8 To guarantee that most people who are colorblind can distinguish groups of data that are color coded, avoid using a combination of red and green in the same display.

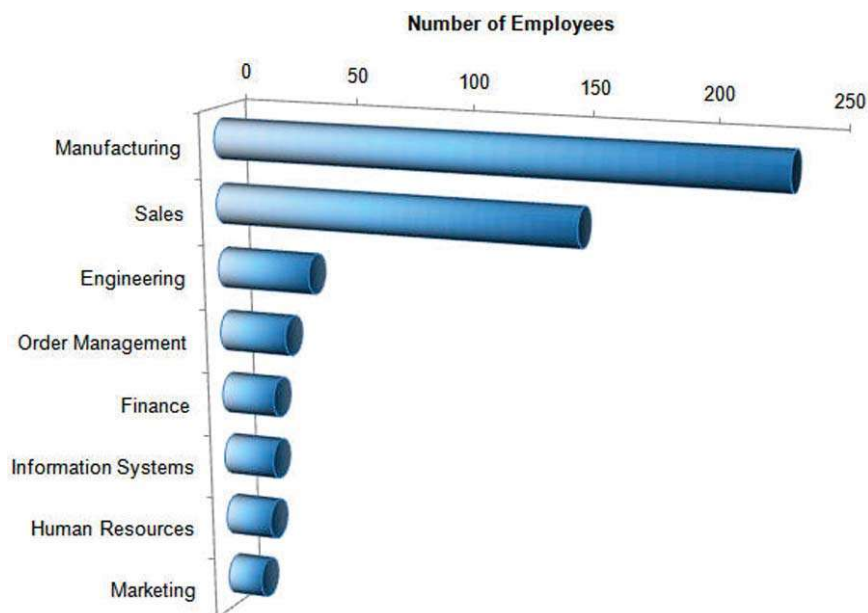
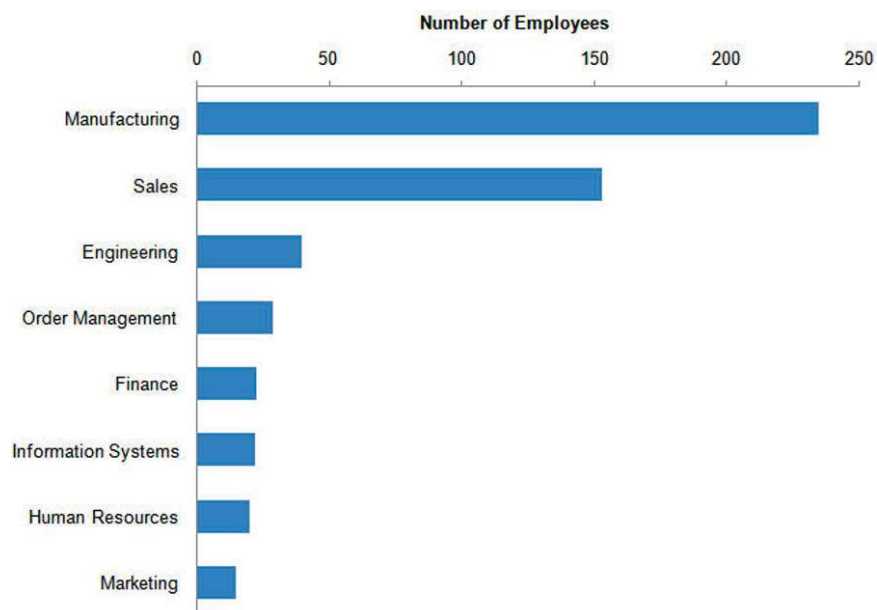
A good substitute in a heatmap for distinguishing positive and negative values is blue and red, rather than green and red.

Effects of Light and Shadow

Most software products that are used to produce graphs today, including Excel, provide an array of visual effects to dress up graphs. Effects of light and shadow, especially those that are designed to make two-dimensional objects such as bars, lines, and data points, appear three dimensional, are all the rage. Unfortunately, these effects introduce problems that can undermine the clear communication that we aim for in our graphs. They add only superficial appeal, but rarely any real value. Here’s the last of the rules:

Rule #9 Avoid using visual effects in graphs.

Which of the following two graphs is easier to read?



In a graph, there is no need to make things look like objects in the real world. We want to display data as simply and clearly as possible.

Follow this simple set of nine rules and you'll improve the quality of your tables and graphs without increasing your workload. If the information is worth displaying, it's worth displaying well.

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.