Elegance Through Simplicity
Stephen Few
October 16, 2004

Experts in any domain tend to lose awareness of the individual steps that they go through to produce their work. As their expertise grows and deepens over time, it becomes second nature, intuitive, and automatic. Ask them to break the process down for you and they might just stare blankly, genuinely befuddled. They know how to do it, but that knowledge has become so intimately integrated into their minds and bodies, they can no longer articulate the process. In this article, I'll try to save you some time and trouble by revealing the steps I take when designing data presentations.

This series of articles has examined the poor state of data presentation today, common mistakes you should avoid, how visual perception works and how you can apply that knowledge to the visual presentation of data, and the seven quantitative relationships that regularly need to be presented via tables and graphs (see Resources). Now let's examine the steps involved in designing effective data presentations. You'll be amazed at how simple they are; in fact, *simplicity* is the guiding principle of effective data presentation.

The Fundamental Challenges

The process of creating an effective data presentation begins with a clear understanding of the data and a firm grasp of your message. With this knowledge in hand, you must then face two fundamental challenges in your effort to present the data:

1. Determine the best medium of display. (Sentence, table, or graph? If a table, which kind? If a graph, which kind?)
2. Design the components of the medium you have chosen to display the data and your message clearly, without distraction.

The previous article in this series, "Eenie, Meenie, Minie, Moe" already addressed the first step in this process. Now we can proceed to the second step: what you should do to make each component of the table or graph that you've selected clearly encode and support the data, with the most important information shining brightest of all.

The Data-Ink Ratio

Edward Tufte, the expert whose work in the field of data presentation I cite and praise more often than any other, laid the conceptual foundation for visual clarity years ago. In his 1983 book, *The Visual Display of Quantitative Data*, he succinctly stated the goal: "Above all else show the data." He then proceeded to show us how. Tufte introduced a concept that he coined the "data-ink ratio":

A large share of ink on a graphic should present data-information, the ink changing as the data change. Data-ink is the non-erasable core of a graphic, the non-redundant ink arranged in response to variation in the numbers represented. Then,

Data-ink ratio = *data-ink ÷ total ink used to print the graphic*

= *proportion of a graphic's ink devoted to the non-redundant display of data-information*

= *1.0 - proportion of a graphic that can be erased without loss of data-information.*

The data-ink ratio is the proportion of ink (or pixels, when displaying information on a screen) that's used to present actual data, without redundancy, compared to the total amount of ink (or pixels) used in the entire display, such as in a table or graph. The goal is to design a display that has the highest possible data-ink ratio (that is, as close to the total of 1.0 or 100% as possible), without eliminating something that is necessary for effective communication.
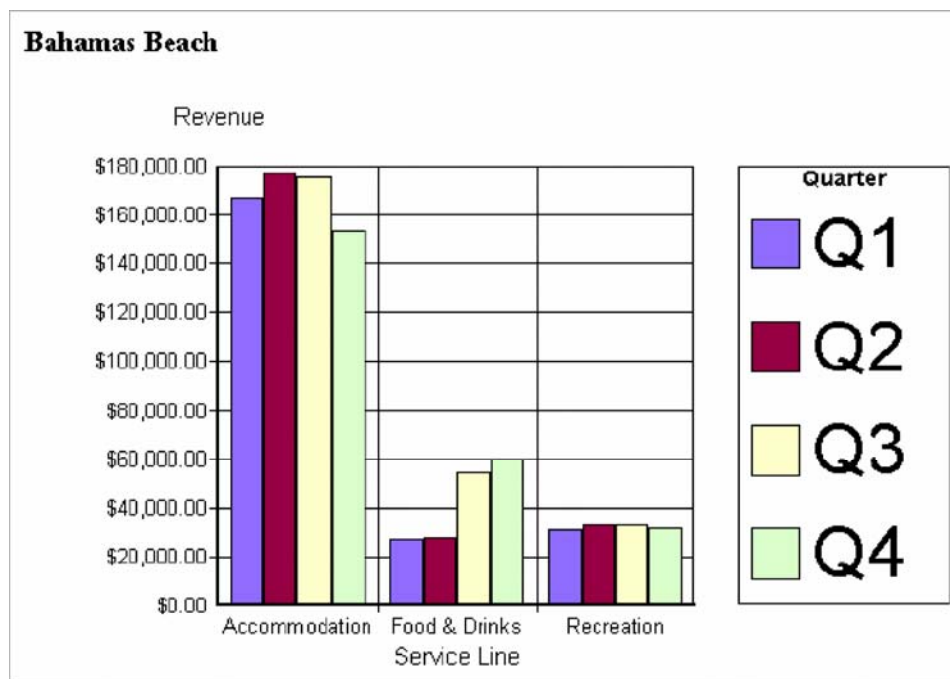
Figure 1: Example of a graph (taken from Business Objects' user documentation) with a low data-ink ratio.

To begin to understand and appreciate this concept, consider Figure 1. Everything that communicates actual data is data ink. In this example, the bars, the tick marks and numeric values along the vertical axis, the labels along the horizontal axis, the colors and labels that identify the four quarters, and the titles are all data. Despite the many data components, however, this graph falls far short of an optimal data-ink ratio. Several examples of non-data ink stand out as completely unnecessary. Can you identify them? In my opinion, the border

around the legend, the border around the plot area, and the grid lines are all unnecessary data ink.

Based on the concept of the data-ink ratio, Tufte describes the goal of effective data display:

> *Maximize the data-ink ratio, within reason. Every bit of ink on a graphic requires a reason. And nearly always that reason should be that the ink presents new information.*

Now, let's learn how to achieve this goal.

Designing Clear Data Displays

The process of maximizing the data-ink ratio consists of two major steps, each of which can be further broken down into two minor steps:

1. Reduce the non-data ink.
    a. Remove unnecessary non-data ink.
    b. De-emphasize and regularize the remaining non-data ink.
2. Enhance the data ink.
    a. Remove unnecessary data ink.
    b. Emphasize the most important data ink.

You begin by identifying all of the non-data ink components of your display and asking the following question of each: "Will any data be lost or diminished if this component is removed?" For example, if the border around the legend is removed, will the data and its message suffer? If the answer is "No," then the component should go. Once you've completed this step, the only non-data ink remaining is that which somehow supports the data and its message in a necessary way. The writer Antoine de St. Exupery insightfully expresses a fundamental principle of elegant communication: "In anything at all, perfection is finally attained not when there is no longer anything to add, but when there is no longer anything to take away."

The most common examples of useful non-data ink in a business graph are the axes. Rather than encoding data, the axes define the space in which the data is displayed. In doing so, they play a useful support role. As such, they should be visually muted to the point where they're just visible enough to do their job, no longer competing with the data for attention. Also, if you always design non-data ink components to look the same wherever and whenever they appear, thus regularizing them, they'll never jump out demanding undue attention.

Now let's enhance the data ink, first by identifying and eliminating any that isn't necessary, which will automatically make what remains more accessible. Looking once again at Figure 1, do you see any data ink that isn't necessary? Four items jump out to me:

1. The cents portion of the revenue values along the vertical scale is useless and suggests a level of numeric precision that the graph doesn't support.
2. The repetition of the dollar sign for each revenue value is redundant.

3. The tick marks along the horizontal scale aren't needed to mark the positions of the three service lines (Accommodation, Food and Drinks, and Recreation).
4. The black borders around the bars add no value.

The cents, tick marks, and bar borders may be eliminated altogether, and the dollar signs may be replaced by simply adding the label "U.S. $" to the vertical axis.

Assuming that no one service line is more important than the others and that the four quarters are of equal significance to your message, none of these should stand out above the rest. If "Food and Drinks" needed to be the main focus of the graph, however, you could make it visually stand out above the other two service lines, such as by giving it a brighter color, but more on this a little later.
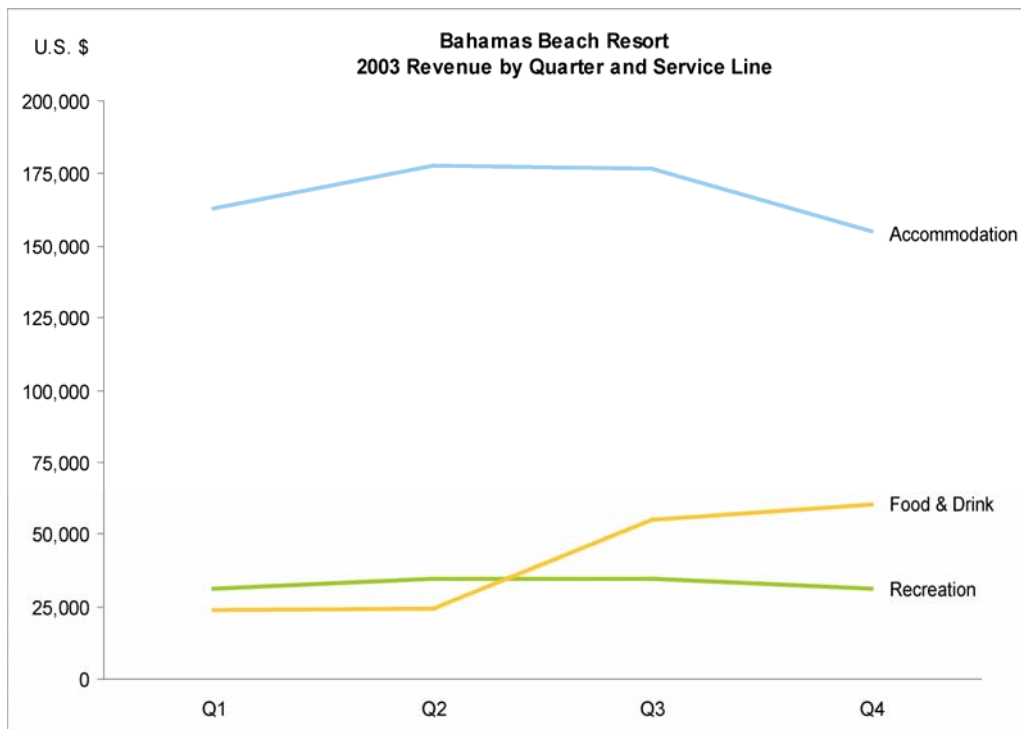
Figure 2: Example of a graph with a high data-ink ratio.

Given the equality of the three service lines and four quarters, the data that ought to be highlighted is probably the pattern of each service line's revenue as it moves through the four quarters and the comparative performance of the three service lines in any one quarter. Although bars can be used to accomplish this, I think that these features of the data can be highlighted best by encoding the values as lines of different colors, one for each service line, rather than as bars. Take a look at Figure 2 to see how well this works.

The three prominently displayed lines, without anything else to distract from them in the plot area of the graph, makes the key features of the data stand out clearly from the text data that surrounds them. Notice that the graph has been further simplified by eliminating the legend and labeling each service line immediately to the right of the line that represents it. Notice also that the weights of the three lines are the same and that no one line color stands out more prominently than the others, thus supporting the equal importance of the three data sets.

Visual Means to Highlight Data

In the original graph shown in Figure 1, the differing fonts and their sizes suggest distinctions in the data that don't really exist. For example, the huge font used for labeling the four quarters in the legend suggests that the quarters are somehow more important than the other data, but this clearly isn't the case. Making something bigger than the rest is a useful way to make it stand out as more important, so you must be careful to keep the sizes of similar components the same unless you really do want to emphasize them.

There are two fundamental means to visually highlight data:

1. Use visual attributes that are perceived as greater.
2. Use contrasting visual attributes.

In Part 3 of this series, I described the visual attributes (size, color, and so on) that are most useful in data presentation, including a distinction between those that are perceived quantitatively (for example, the longer the line the greater its value) and those that are useful only for distinguishing separate sets of data. These visual attributes apply directly to the topic at hand. The following attributes are perceived quantitatively to the extent that some variations are clearly perceived as greater than others:

- Bigger (longer, wider, or both)
- Higher (located closer to the top)
- Darker
- Brighter

Bigger objects, such as bigger fonts, stand out more than their smaller cousins. Objects that are located closer to the top stand out more than those that are lower on the page or screen. Darker or brighter colors usually stand out more than lighter, less saturated colors.

You don't need to rely only on these attributes that are perceived quantitatively to highlight items. A contrasting version of any visual attribute, rendered differently from the norm— whatever the norm—demands attention. For example, if all the text in a graph is black, except for the text that labels one of the bars, your attention will be drawn to that bar. Our visual sense is highly attuned to anything that differs from its surroundings. Experience through our long evolutionary journey has taught us that contrasts from the norm often signal danger. Anything different catches our eyes and begs us to assign meaning, to understand the significance of that difference.

In addition to these visual attributes, the following non-quantitatively perceived attributes are also useful for highlighting data through visual contrast:

- Different orientation (such as an italicized font)
- Different shape (such as a different font or a differently shaped point
- Different hue
- The addition of something that doesn't appear elsewhere, such as a border, fill color, text label, or pointer (for example, an arrow).
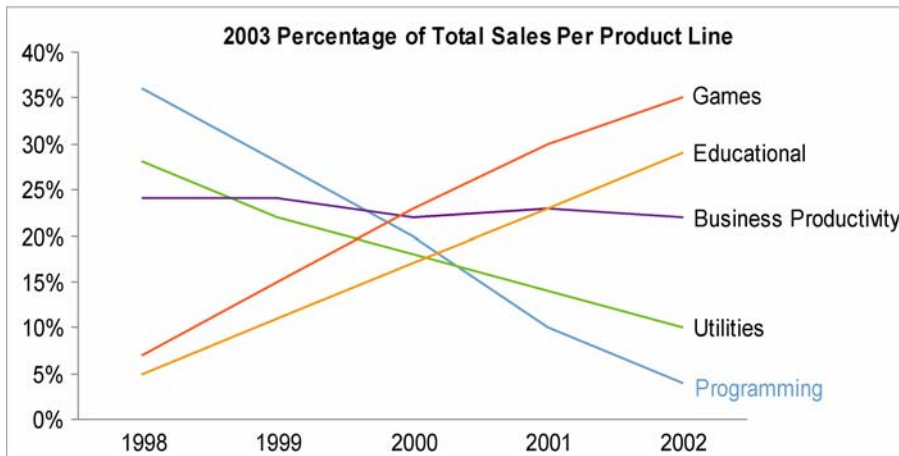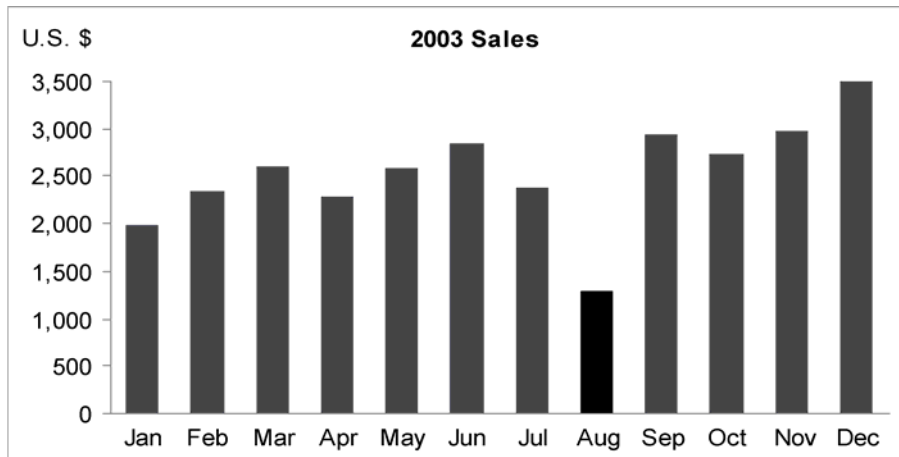
Figure 3: Highlighting data using visual tributes that are either greater (top) or contrasting (bottom).

Figure 3 shows examples of using visual attributes that are greater and contrasting to highlight data. Bear in mind that contrast only exists when there is a norm against which to create a distinction. If you have four sets of bars in a graph, each rendered in a different pastel hue, the one that happens to be soft pink isn't going to stand out as obviously important in contrast to the baby blue, sea-foam green, and terracotta orange. If, however, the other sets of bars are each a different shade of gray, the soft pink is going to clearly stand out. If you make everything bright, not only will nothing stand out, you run the risk of assaulting your readers' senses. A page or screen full of bright colors might entertain preschoolers watching the *TeleTubbies* on TV but it doesn't work for adults. You would do well to replace the bright colors that your charting software likely provides as the defaults for the bars, lines, and data points of graphs with softer earth tones, reserving the bright and dark colors for those items that need to clearly shout "Look at me!"

Strive to maintain consistency of appearance in a data presentation, except when the data itself differs in meaning or importance and visual differences can be used to communicate differences in the data. One common mistake is the use of color variations for bars that are

part of the same data set, which is entirely unnecessary. The left-hand graph in Figure 4 encodes each bar as a different color, but we already know that each bar represents a different country by its position along the horizontal axis and the corresponding labels. Different colors are not only unnecessary in this graph, but they also make it more difficult to see the six bars as a single related data set and discern the relationships among the bars as they range from the largest value to the smallest.
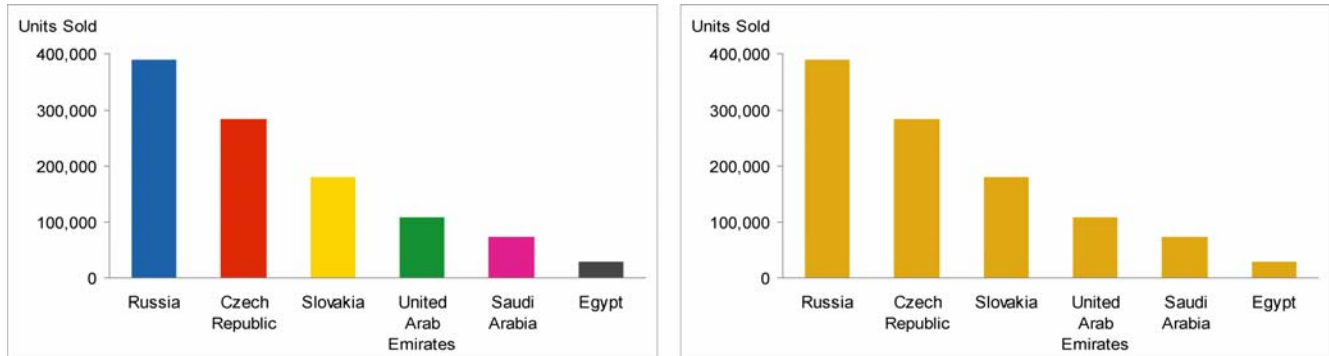
Figure 4: The graph on the left misuses distinctions in color in a meaningless and distracting way.

Effective data presentations are rooted in a constant awareness of the goal, which is to *communicate*—not to entertain or impress. Keep this goal in mind whenever you design data presentations and pursue it using the principles and practices that I've introduced in this series of articles. Always ask yourself, "Does this design present the data and its message clearly?" If the information is worth presenting, it is worth presenting well.

References

- Edward R. Tufte, *The Visual Display of Quantitative Data,* Graphics Press, 1983.
- Kevin Mullet and Darrel Sano, *Designing Visual Interfaces,* Sun Microsystems, 1995.

(This article was originally published in *Intelligent Enterprise*.)

_____

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.