Data visualization is part art and part science. The challenge is to get the art right without getting the science wrong, and vice versa. A data visualization first and foremost has to accurately convey the data. It must not mislead or distort. If one number is twice as large as another, but in the visualization they look to be about the same, then the visualization is wrong. At the same time, a data visualization should be aesthetically pleasing. Good visual presentations tend to enhance the message of the visualization. If a figure contains jarring colors, imbalanced visual elements, or other features that distract, then the viewer will find it harder to inspect the figure and interpret it correctly.

In my experience, scientists frequently (though not always!) know how to visualize data without being grossly misleading. However, they may not have a well-developed sense of visual aesthetics, and they may inadvertently make visual choices that detract from their desired message. Designers, on the other hand, may prepare visualizations that look beautiful but play fast and loose with the data. It is my goal to provide useful information to both groups.

Ugly, Bad, and Wrong Figures

Throughout this book, I frequently show different versions of the same figures, some as examples of how to make a good visualization and some as examples of how not to. To provide a simple visual guideline of which examples should be emulated and which should be avoided, I am labeling problematic figures as "ugly," "bad," or "wrong" (Figure 1-1):

Ugly
: A figure that has aesthetic problems but otherwise is clear and informative

Bad
: A figure that has problems related to perception; it may be unclear, confusing, overly complicated, or deceiving

Wrong
: A figure that has problems related to mathematics; it is objectively incorrect

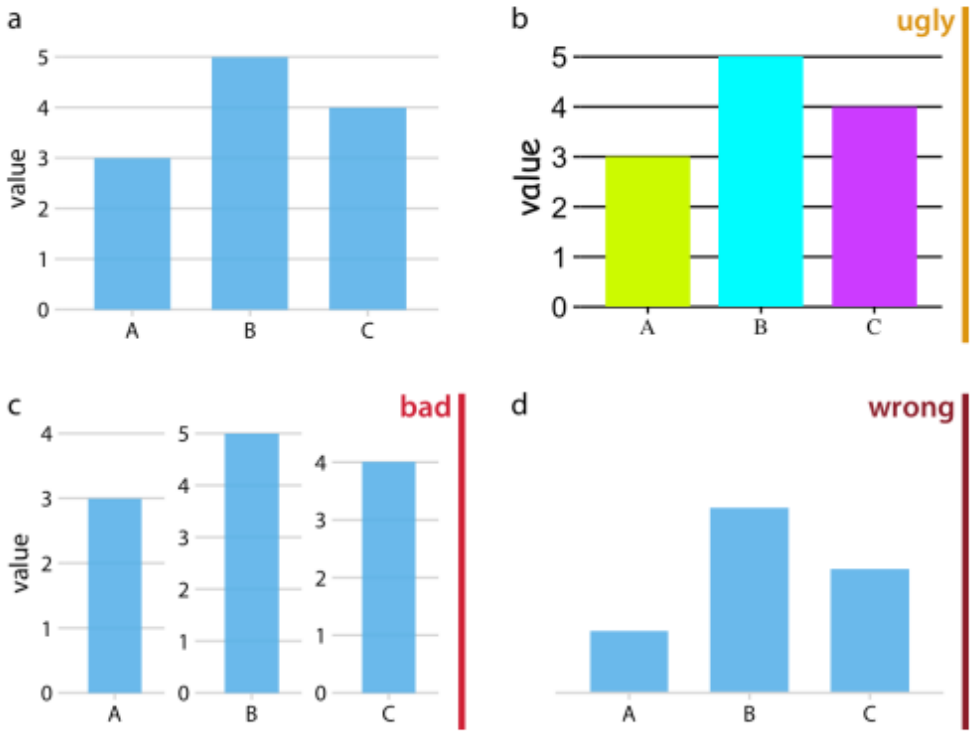

Figure 1-1. Examples of ugly, bad, and wrong figures. (a) A bar plot showing three values (A = 3, B = 5, and C = 4). This is a reasonable visualization with no major flaws. (b) An ugly version of part (a). While the plot is technically correct, it is not aesthetically pleasing. The colors are too bright and not useful. The background grid is too prominent. The text is displayed using three different fonts in three different sizes. (c) A bad version of part (a). Each bar is shown with its own y *axis scale. Because the scales don't align, this makes the figure misleading. One can easily get the impression that the three values are closer together than they actually are. (d) A wrong version of part (a). Without an explicit* y *axis scale, the numbers represented by the bars cannot be ascertained. The bars appear to be of lengths 1, 3, and 2, even though the values displayed are meant to be 3, 5, and 4.*

I am not explicitly labeling good figures. Any figure that isn't labeled as flawed should be assumed to be at least acceptable. It is a figure that is informative, looks appealing, and could be printed as is. Note that among the good figures, there will still be differences in quality, and some good figures will be better than others.

I generally provide my rationale for specific ratings, but some are a matter of taste. In general, the "ugly" rating is more subjective than the "bad" or "wrong" rating. Moreover, the boundary between "ugly" and "bad" is somewhat fluid. Sometimes poor design choices can interfere with human perception to the point where a "bad" rating is more appropriate than an "ugly" rating. In any case, I encourage you to develop your own eye and to critically evaluate my choices.