How to Avoid Some Common Graphical Mistakes

Naomi B. Robbins

NBR, 11 Christine Court, Wayne, New Jersey 07470.

E-mail: naomi@nbr-graphs.com

Introduction

A major obstacle to graphical literacy is the widespread availability of software that pays more attention to decorating graphs and to wowing the audience than to communicating clearly and accurately. People who would never write a business report using the fanciest typeface of their word processor think nothing of decorating their graphs with the fanciest options available in their graphing software. This results in figures that mislead, confuse, and cause poor decisions to be made based on a misunderstanding of the data. This paper discusses several common graphical mistakes so that the audience will avoid these mistakes once they are aware of them.
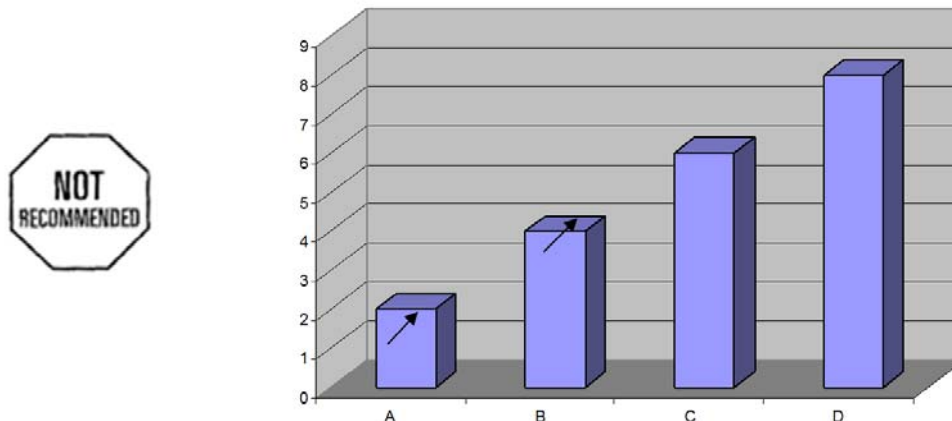
Using Unnecessary Dimensions

Figure 1: Unnecessary dimensions are used to decorate this bar chart.

A very common way to decorate graphs is to include unnecessary dimensions such as in Figure 1.

There are a number of problems with this approach. First, readers do know how to read the graph. Do you read from the front of the bar as the arrow on the A bar indicates, or the back of the bar as the arrow on the B bar indicates? Figure 2 shows the same data without the extra dimension, indicating that the correct way to read this pseudo-three-dimensional figure drawn using the default options of Excel 2003 is not either of the

choices listed above. Notice that the bars in Figure 1 do not touch the back wall of the chart. The designers tried to compensate for this but almost everyone reads it incorrectly and underestimates the values of the bars.
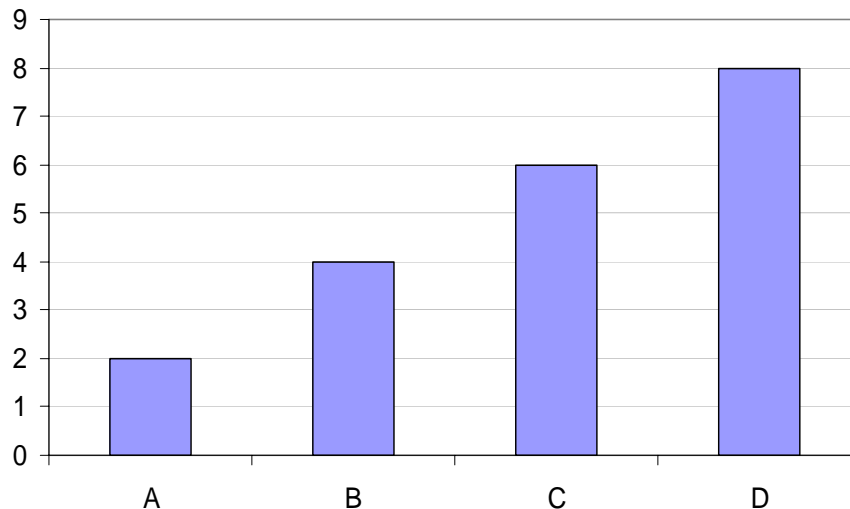

Figure 2: Shows the same data as Figure 1 drawn in a two-dimensional bar chart.

A second problem with pseudo 3-D charts is that different software packages use different default options or different algorithms to draw these charts. For example, if you plot the same values using Microsoft Graph, the software for Word and PowerPoint, you will see that the default there does touch the back wall. For some

other software packages you do read the values from the front as the arrow on the A bar of Figure 1 suggests. It is totally unacceptable that the way to read a graph depends on the software used to draw the graph.

A third problem with pseudo three-dimensional charts is that they suggest we know more about the data than we really do. A two-dimensional figure displays two variables. In the case of Figure 2 we know if the data belongs to group A, B, C or D and we also know the value of the data as shown in the height of the bar. Three-dimensional figures should correspond to three variables, but we have no additional information for Figure 1.

Take some simple data and draw a 2-D and pseudo 3-D pie chart. You will notice that the extra ridge on the front of the pie emphasizes some wedges at the expense of others. Adding unnecessary dimensions or adding perspective to the pies distorts the data.

Showing Changes in One Dimension with Area or Volume
We just observed that changing the length of an object displays changes in one dimension. Yet many graph designers use area or volume to show one-dimensional changes. This exaggerates the change. Figure 3 shows two boxes; each side of the larger one is twice that of the smaller one. The visual impression is that the larger box is more like eight times that of the smaller one. I say more like since the study of human perception

shows that we underestimate volumes and we perceive the larger box to be a little less than eight times the size of the smaller one.
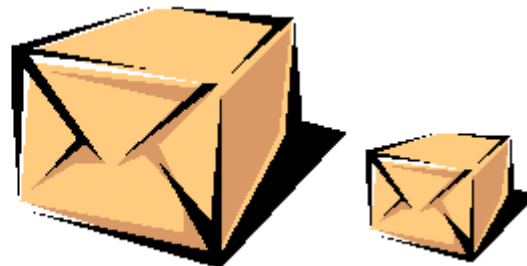

Figure 3: Shows two boxes with each side of the larger one twice the length of the smaller one.

The height of the largest person at the right of Figure 4 is 1.82 times the height of the smallest person at the left of the figure, but visually it appears that the person at the right is larger than that since the width was varied in addition to the height. A more accurate visual interpretation of these numbers would hold the width constant and only vary one dimension. A similar problem occurs when pictograms use differently sized figures for their categories; the result is that smaller numbers may be visually bigger than larger numbers.
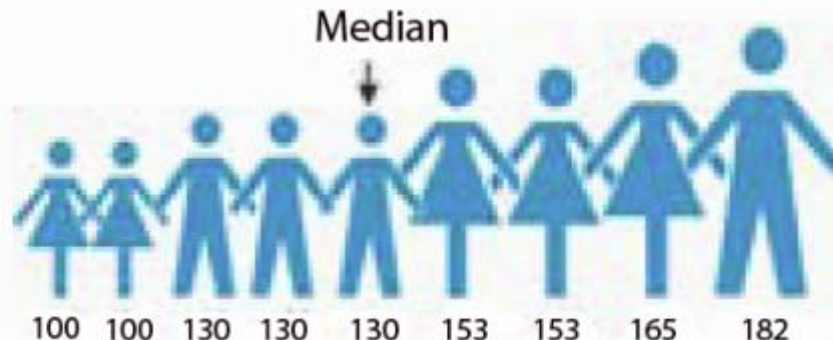
Figure 4: Illustrates using area for a one-dimensional change using an example included in a curricula to help users understand official statistics.

Not Making Data Stand Out

The data and model(s) you are graphing are the most important elements in a graph; therefore, the data should be the first thing you notice when you look at a graph. Often when we look at a graph the elements that stand out the most are grid lines, long slanted labels, or other non-data elements. Should this occur, the figure would benefit from a redesign. The changes needed are often quite simple such as making grid lines a light shade of gray or not labeling every tick mark.

Making Mistakes with Scales

Mistakes with scales are among the most common graphical mistakes. A few examples include having equally spaced tick marks for uneven intervals, using too many or too few tick marks and labels, and having a bar chart without a zero baseline.

The times at which data are available are often not equally spaced. For example, in Figure 5, life expectancies in the United States are shown for 10 year intervals from 1900 to 1990 and then for 1992 and 1998. However, the spacing of the tick marks do not reflect this change so that the trend at the right side of the chart is misleading. This is a very common mistake, especially with Excel users who do not understand the assumptions Excel makes with its charting capabilities. Excel assumes the horizontal axis of a line chart displays a categorical variable; i.e., a variable that contains categories or labels such as the name of countries with no intrinsic ordering. Figure 6 spaces its intervals correctly.
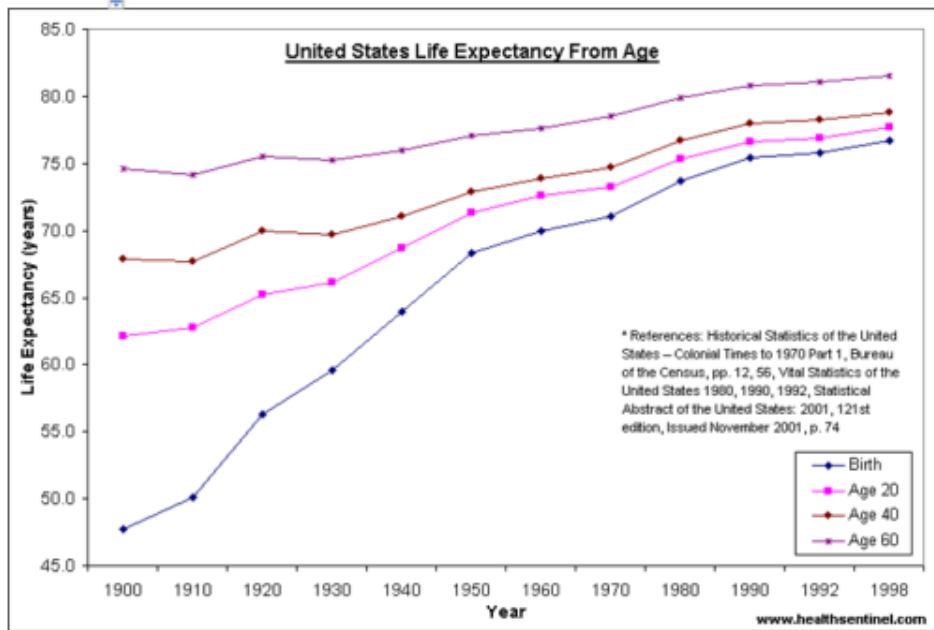
Figure 5: Has equally spaced intervals representing uneven periods of time.
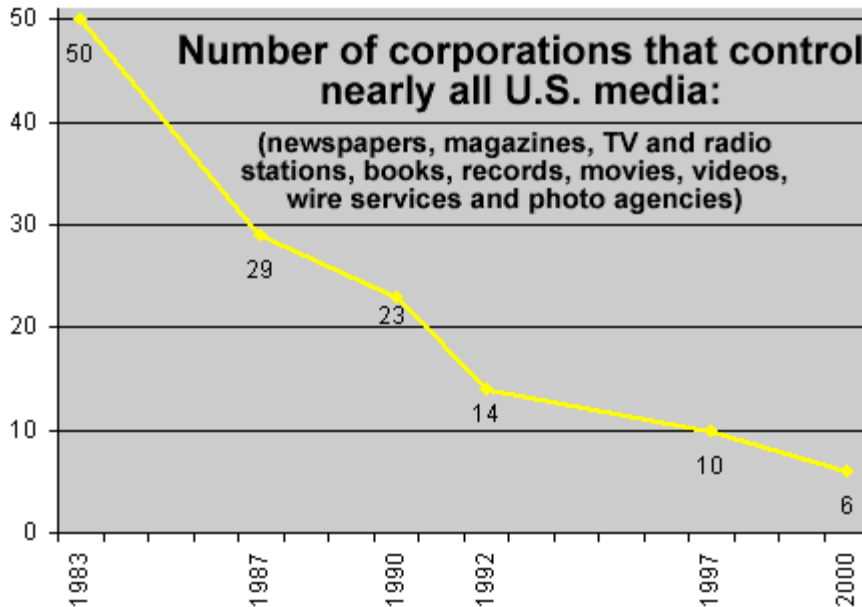
Figure 6: Spaces the data points properly.

Other common problems with tick marks and their labels include using too many or too few. Too many tick
mark labels often cause run-on labels that are difficult to read or having to slant the labels, taking up a good

deal of the available real estate of the figure. Too many tick mark labels can cause the labels to be more prominent than the data. Too few tick marks can make it difficult to estimate the values of the data points, as in Figure 7. Placing tick marks and the maximum value of a scale at round numbers rather than at a value such as 47, 153 would make it easier for readers to interpolate values.
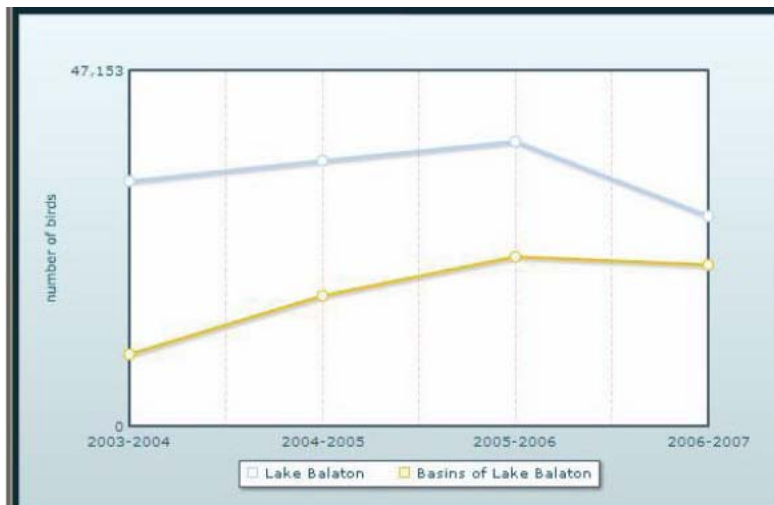

Figure 7: Has no tick marks on the vertical scale making it difficult to estimate values.

Whether or not al scales require zero it a controversial issue, but most experts agree that all bar charts require

a zero baseline. We judge values of the data by the length of the bar in a bar chart, and judgments of length require a zero baseline.
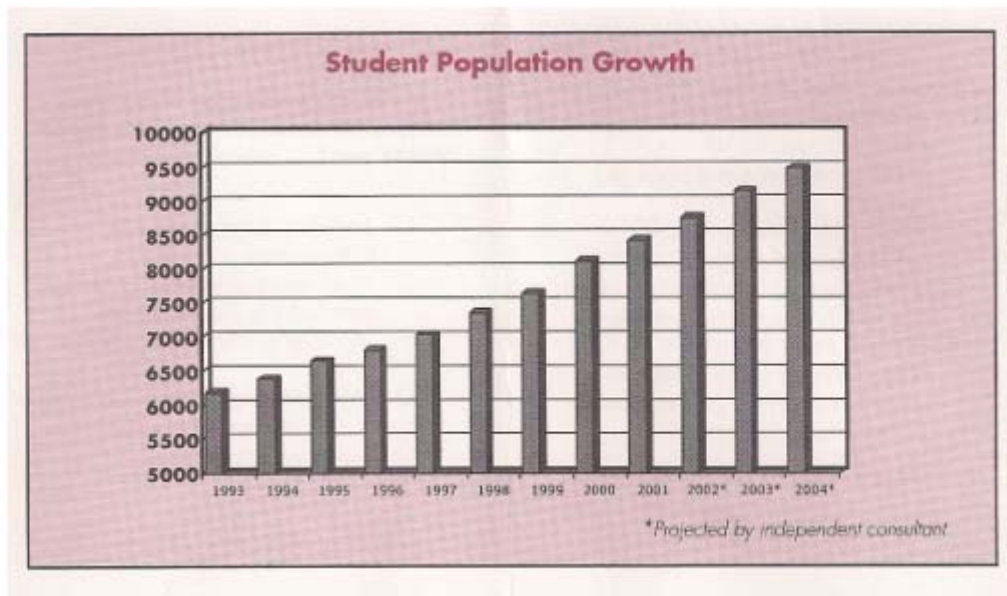

Figure 8: Starts the bars at 5000, distorting the visual comparisons of their lengths.

Improving the Design of Tables

Understanding the data in tables is also enhanced by proper design and avoiding common mistakes. Numbers should always be right-adjusted to ease comparisons. Far too many tables ignore this rule and center the numbers. It is very difficult to add numbers or to compare numbers that are not right-adjusted. Another common mistake with tables is to include too many decimal places. Printing out as many places as computers provide draws attention to the insignificant digits and away from the digits that matter. Never include more decimal places than are needed to understand the data.

Rules and gridlines break up the data and inhibit useful comparisons. White space is the preferred separator between rows and columns. Rules are useful for separating summary statistics such as averages or totals from the main data or separating titles from the data. Comparisons are easier down columns than across rows so care must be given to the design of the table.

Conclusion

Tables and graphs help readers to understand data and turn information into knowledge. However, poorly designed tables and graphs can confuse, mislead, and deceive. Understanding some common graphical mistakes can go a long way to help designers of graphs inform rather than mislead and to prevent readers of graphs from being misled. Many more examples are included in Chapter 2, 6, and 7 of Robbins [1]. The

frequency with which these mistakes occur, even in statistical journals, make this paper necessary to include although its contents are quite elementary.

Source of Figures

Fig. 1 Robbins, Naomi B. (2005). *Creating More Effective Graphs*. John Wiley, Hoboken, NJ.

Fig. 2 ibid.

Fig. 3 Robbins, Naomi. B.(2009)   All Rights Reserved.

Fig. 4 http://www.stat.ucla.edu/~jsanchez/books/ASA-2008.pdf, accessed Apr. 10, 2009

Fig. 5 http://www.healthsentinal.com/Vaccines/DiseaseAndRelatedData.htm, accessed Oct. 5, 2008.

Fig. 6 http://www.prorev.com/stats.htm, accessed Apr. 10, 2009.

Fig. 7 http://www.oecd.org/dataoecd/47/22/40010783.pdf?contentId=40010787, accessed Feb. 14, 2009.

Fig. 8 Wayne, NJ Public Schools Community Survey, 2001.

REFERENCES (RÉFERENCES)

[1]   Robbins, Naomi B. (2005). *Creating More Effective Graphs*. John Wiley, Hoboken, NJ.