Visualizing Data: Challenges to Presentation of Quality Graphics—and Solutions

Naomi Robbins

Naomi Robbins, a consultant and seminar leader who specializes in the graphical display of data, offers training on the effective presentation of data. She also reviews documents and presentations for clients. She is the author of *Creating More Effective Graphs*.

Three common challenges statisticians and others face when preparing data for presentation include poor options and defaults in many software packages used for creating graphs, managers and colleagues who are socialized to expect figures that attract attention, and poor instructions from conference organizers. This article addresses each of these challenges and offers some tips for dealing with them.

Poor Options and Defaults in Many Software Packages

Many software programs for drawing charts and graphs offer defaults and options that are full of fancy embellishments that detract from the clear and accurate communication of data. Some software vendors think graphs that wow the audience with the complexity of their artwork will produce more sales. Therefore, they include unnecessary dimensions, use confusing ribbons in place of lines, and offer graph forms that do not communicate well. Unfortunately, these frills and decorations may distort the data, make it more difficult to understand, and may lead to poor decisions being made based on the data.

Figure 1 shows the results of an Internet/mail survey of ASA members with six to 15 years of membership. The figure appears in the October 2005 issue of *Amstat News*. Members were asked if they agreed that their primary position was professionally challenging; they also were asked about the importance of increasing professional recognition. In addition, a number of demographic variables were included.

A major problem with pseudo-three-dimensional bar charts like this one is that almost no one reads them correctly. Note that the bar for "Agree/

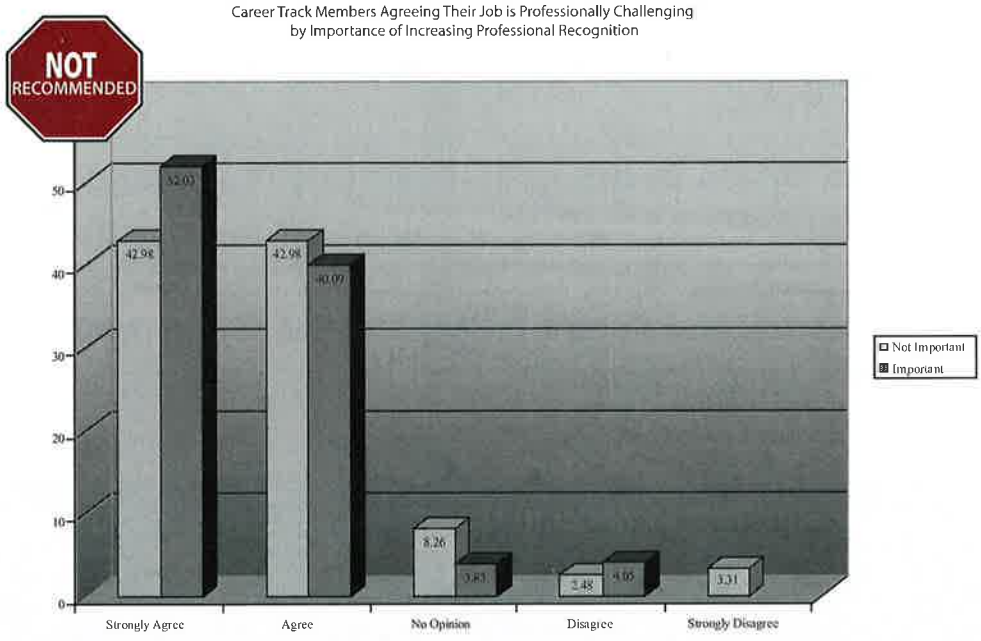

Figure 1. The use of unnecessary dimensions in bars. All titles and labels are from the original figure.

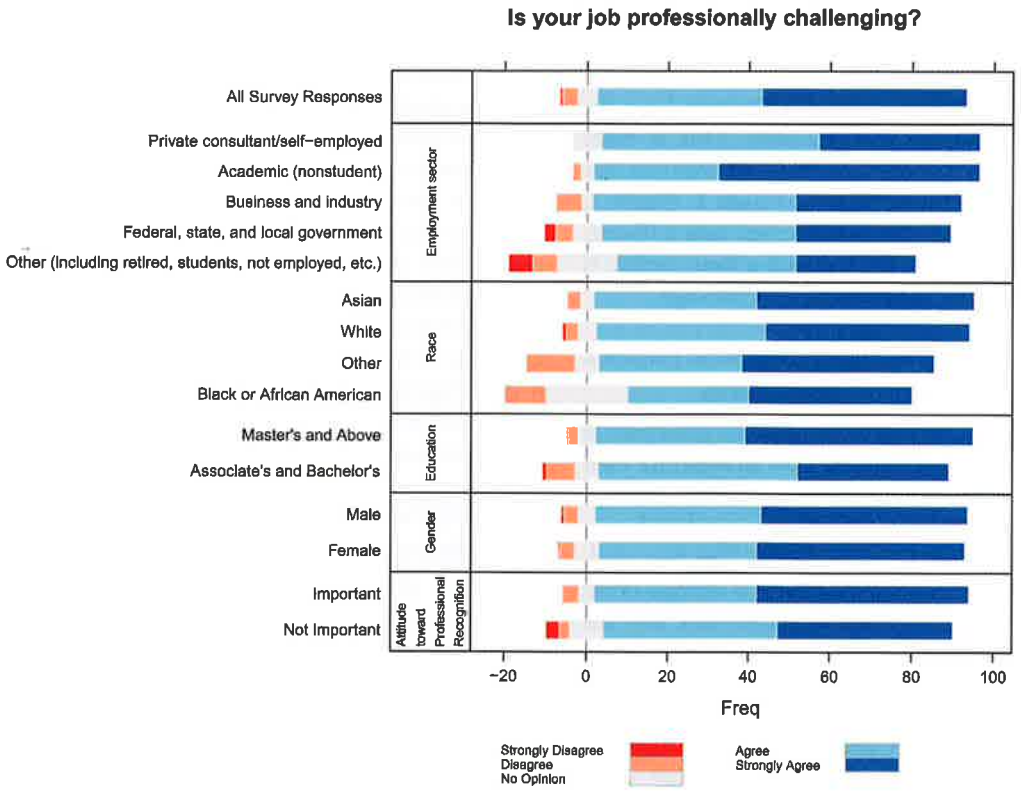

Figure 2. A two-directional stacked bar chart showing the percentage of various demographic groups that agreed or disagreed their positions were professionally challenging

Important" is labeled 40.09; however, all points on the top of the bar lie below the grid line labeled 40. This happens because the bar does not touch the back wall.

I assume it was designed so a plane tangential to the top of the bar would look as if it were the correct height. It doesn't work for me. This figure looks as if it was drawn using a version of Microsoft Excel prior to 2007 with the gap depth (i.e., the distance from the back of the bar to the back wall) set as the default. Note that gap depth is an option that can be changed in Excel. It confuses the audience when the labels do not match the visual representation.

The pseudo three-dimensional problem is easily solved by using a two-dimensional bar chart. In Excel, that means sticking with "2-D Column" or "2-D Bar" charts and never using what I call pseudo-three-dimensional charts: the "3-D Column," "3-D Bar," "Cylinder," "Cone," and "Pyramid" options. These are "pseudo" three-dimensional since they only display two dimensions, despite their 3D appearance. Data that are truly three-dimensional with three variables cannot be displayed with these charts. I often recommend Trellis displays, described in *Creating More Effective Graphs,* for plotting three variables.

There are numerous other problems with this figure. The small font size of the title, labels, and legend make them difficult to read. The grid lines are too prominent, distracting attention from the more important elements of the graph. The variably shaded gray background also takes attention away from the data. Some readers have difficulty interpreting the title when the graph is separated from the article in which it was originally published, since the full text of the article clarifies the title.

The diverging stacked bar chart of Figure 2 shows a much improved way of conveying all the information available in Figure 1, with the addition of a breakdown by employment sector, race, education, and gender. Yet, the figure takes about the same space as Figure 1. Adding unnecessary dimensions is just one example of a poor option that occurs in many software programs.

While standard stacked bar charts are difficult to read, centering the bars around zero ("No Opinion" in this case) makes it easy to check whether the majority in any subgroup find their job professionally challenging or not. The length of the bar to the right of zero shows the percent who agree that their positions are professionally challenging, while the length to the left of zero shows those who

disagree. People with no opinion are split down the middle. The shading shows whether the agreement was strong.

In a glance, we can see which groups have the strongest agreement/disagreement. This is more difficult to notice from a table or separate bar graphs for each category, as in Figure 1. Figure 2, programmed by Richard Heiberger, uses a diverging stacked bar chart created with a forthcoming R function to be included in the HH package.

Managers and Colleagues Who Expect Figures That Attract Attention

Statisticians who are well versed in the principles of effective graphs often ask me how to convince their managers or colleagues that the figures requested are misleading or inappropriate. They say their bosses want figures with a "wow" factor. The first challenge and this one together form a vicious cycle: Managers like the decorated graphs that they see software vendors providing, and software vendors believe the managers prefer and demand them.

Almost no one would write a business report in a font that attracts attention, such as Algerian. Business writers reserve display fonts for single words or phrases in advertisements or for invitations to a child's birthday party. There are a number of analogies about using display fonts and graphs that attract attention for the wrong reasons: They both emphasize the design, rather than the message the words or graphs are meant to communicate. They both show off the designer's skills with technology. There may have been a day when people were impressed that you could produce these fancy graphs, but today it is no more impressive than knowing how to change fonts. I often find that making analogies with words helps others see the parallels and encourages them to show the same respect for numbers.

Another effective technique I use to convince others that their favorite graphs do not communicate well is to ask questions about the data that are difficult to answer from their preferred graph. As an exercise at meetings or seminars, participants answer questions about fancy graphs they did not draw. The people who proposed the graphs can readily see that colleagues misinterpret the data. For example, you could show Figure 1 without the data labels and ask how high the "Agree/Important" bar is. When most of the others in the room underestimate the value, the person who designed or requested it will likely realize its limitations.

A number of years ago, there was a discussion on S-news, a support group for the S-Plus software, about the use of pie charts. I still remember a message from Eric Gibson, who said that when he was asked to draw a graph he thought did not communicate well, he did not lecture the requester or play better than thou. He prepared what he was asked to do, but also what he thought should be done. Then, he delivered them, saying, "I always like to give my clients a choice." Many times, the client would see the superiority of his method and use his figure.

A number of statisticians have told me they like to give a book emphasizing principles of effective graphs to their management or clients who request graphs they dislike. Another option is to arrange for a seminar or short course for the department or organization about communicating data clearly. Offering training courses is often the solution when it is management that appreciates communicating data clearly and staff who include unnecessary decorations in their charts.

Poor Instructions from Conference Organizers

A number of conferences advise their speakers to use yellow text on a dark blue background. They claim these are the easiest to read. I've seen articles about effective presentations that recommend light on dark and others that recommend dark on light. There is a problem when light on dark is used and handouts are made from the slide decks, since the colors are inverted for the handouts so the text shows up. The problem is that the original graphs usually have a light background with dark data markers and text. Then, when the colors are inverted, the graphs are illegible.

Also, the handouts are often black and white, even if the original slides were in full color. Any colors used to distinguish points or lines are lost. Handouts are often referred to years after a presentation, so intelligible handouts are essential. I have seen many prominent statisticians with useless handouts since they followed the directions of the conference organizers. The solution: Just say no. I have refused to use yellow on navy, but explained my reasons. The conference organizers replied that they wished other speakers gave as much thought to their handouts.

Summary

Even statisticians well acquainted with the principles of effective graphs face challenges when trying to visualize data. These challenges may be caused by the software the graph designer is required to use, the instructions given by management, or the instructions given by conference organizers. Solutions include recognizing these problems so you choose software options and software carefully, selecting a method for communicating with management that you are comfortable with and is appropriate for the situation, and considering the consequences of following instructions when preparing slides for presentations and speaking up if necessary. ■