

Quantitative Literacy Across the Curriculum

Improving Graphs in College Textbooks

Naomi B. Robbins, Ph.D. and Joyce Robbins, Ph.D.

Visual Business Intelligence Newsletter

February 2010

Introduction

This newsletter is well known for its criticism of the graphs produced by many Business Intelligence (BI) software vendors. This criticism is well-deserved. In this issue we examine an equally disturbing source of confusing and misleading graphs: those found in introductory college social science textbooks. Given the importance of statistics to the social sciences and the prevalence of social science requirements, it's fair to assume that introductory courses significantly influence students' understanding of data presentation. We would expect that the wealth of graphs in social science textbooks would conform to the highest standards of data presentation principles.

What one finds in these textbooks, however, is often disappointing. While the quality does vary among textbooks, a surprising number include graphs which contain all kinds of distortions due to poor graphing techniques. Furthermore, we have found that the same bad graphs appear in edition after edition of these books; the popular social science textbooks today tend to come out with a new edition on a very frequent basis. Not only do students miss substantively what the graph attempts to communicate, but they learn inappropriate methods for displaying data visually. In this article, we provide examples and describe in detail some errors found using pie, bar and line charts as well as tables. It is our hope that by drawing attention to this situation we can begin the process of correcting the problem. Although there are problems with the graphs in many texts, we will single out one textbook that contains more than its share of distorted graphs: *Introduction to Sociology* by Henry Tischler.

Graph forms covered: a short disclaimer

Pie charts, bar charts, line graphs, and tables are among the most common forms of communicating numbers, and not surprisingly, they are the most common forms found in social science textbooks. This in itself is problematic, since many data presentation experts recommend against overuse of pie charts and bar charts. In our view, pie charts do not communicate data well and should be used rarely. (For more information on problems with pie charts, see Becker and Cleveland [1], Tufte [11], Robbins [5], and Few [4].) Bar charts are more controversial. Cleveland [2] recommends dot plots instead of bar charts. Tufte [11] criticizes bar charts since they have high redundancy with a low data-ink ratio. Following Cleveland, we do not recommend using bar charts with error bars, logarithmic scales, or when zero is not included in the scale. We find that bar charts get cluttered quickly while dot plots do not. We are more open to the use of bar charts for very small data sets. Properly drawn line graphs tend to communicate numerical data well. Tables are extremely useful when precision is needed but the design of most tables could be improved.

That said, for the purpose of this critique, we take the position that whatever form is chosen, it should be executed properly. Therefore, rather than take aim at forms that we think are less effective and promote those that are more effective, we point out problems within the chosen graph genre.

In short, whether we like it or not, these graph forms are widely known and used and are likely to continue to be used frequently. Therefore, they should be designed to communicate numerical data as clearly as possible. Line graphs are less likely to have errors than pie and bar charts, yet we find many design flaws with this form as well. In the following sections we point out errors that distort the data in a number of graphs.

Pie Charts

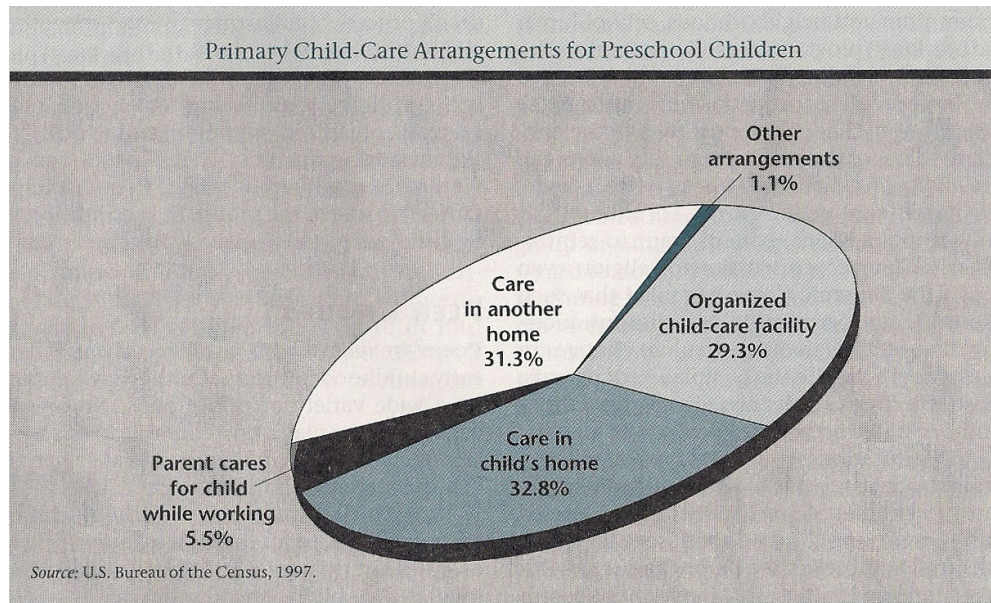


Figure 1. Pie charts do not communicate information well. The added pseudo-third dimension and tilted circle shown above make the problem worse.

Adding a pseudo-third dimension distorts the data, emphasizing some wedges at the expense of others. Adding perspective or tilting the bar distorts the data even more.

A graph is a pictorial representation of numbers. At the least, the elements of the graph representing the numbers should be proportional to the numbers themselves. The percentages in a pie chart are represented by the angles of the wedges, the areas of the wedges, or the arc lengths of the wedges. To our eyes, the angle of the "Organized child-care facility" wedge looks much smaller than the "Care in another home" wedge, even though the values in the labels are close (29.3% vs. 31.3%). Tilting and elongating the pie is poor graphical practice and does not belong in an introductory text. Interestingly enough, the previous edition of this text used a simple two-dimensional pie chart for similar data. Although software packages increasingly offer elaborate design options, in most cases, for communicating data, less is more.

Bar Charts

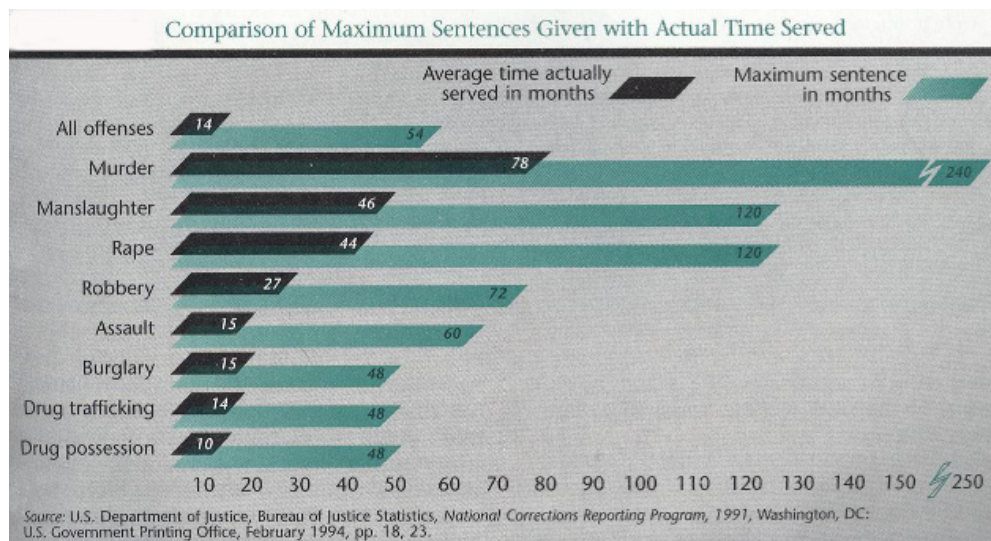


Figure 2. Most readers will not know how to read these charts; the slanted ends are confusing.

The numbers in a bar chart are represented by the length of the bar or the position of the end of the bar. Neither is clear with the bars of Figure 2 because the ends of the bars slant.

Notice that the “average time actually served” for drug possession is ten months and ten is one of the labels on the horizontal scale. However, since there are no tick marks, we cannot tell exactly where ten falls on the scale. Both the top and bottom of the end of that bar appear larger than the midpoint of the label “10”. Graphs should communicate quantitative information rather than confuse readers and make them do detective work to decipher the information.

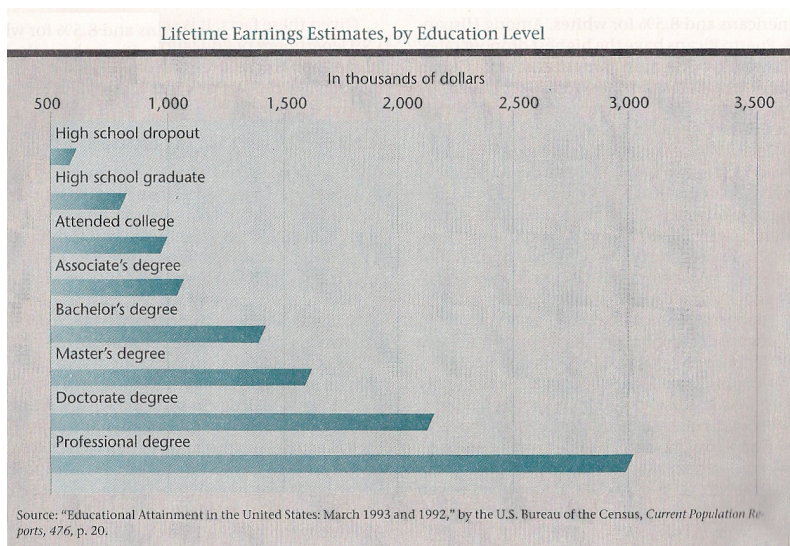


Figure 3. This figure also suffers from slanted bar ends and is deceptive since the judgment of lengths requires a zero base.

The bar chart in Figure 3 is more confusing since, in addition to the slanted ends of the bars, the chart is missing a zero baseline. This is crucial since we judge the values of bars by their lengths and lengths require a zero base. Without zero the comparison is deceptive. Even if you think you are comparing the bars by the position of their endpoints, you cannot help but notice the lengths of the bars as well. For example, the length of the bar for “high school graduate” looks about triple that for “high school dropout” but based on the tick marks, both are between 500 and 1000.

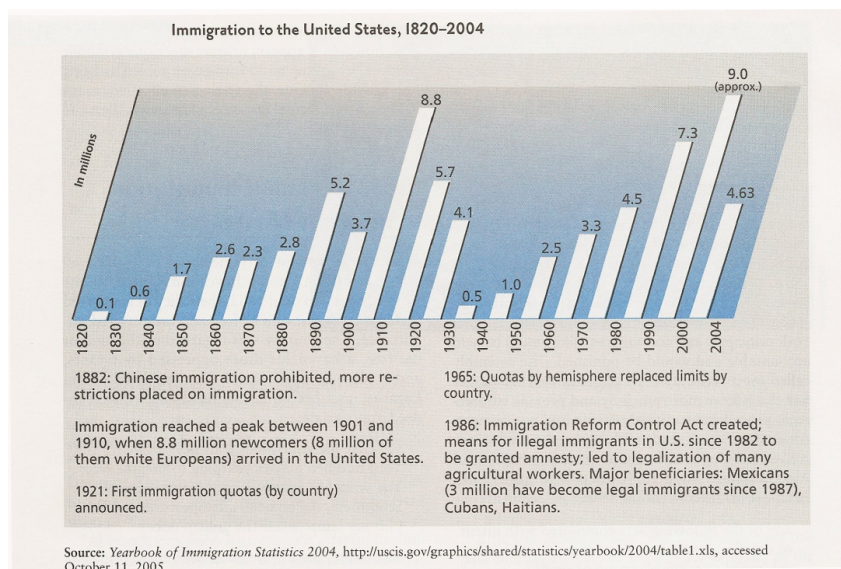


Figure 4. The slanting bars and frame for this graph only serve to confuse. Furthermore, the reader can't tell whether each bar is associated with a decade or year, and if the latter, whether the label is to the right or left of the bar.

The ends of the bars slanted in Figures 2 and 3; here the whole bar slants. This serves no purpose other than to confuse the reader. Comparing the heights of the first two bars shows that the figure is not drawn to scale: the height of the “0.6” bar should be six times the height of the “0.1” bar which it clearly is not.

In addition, it is not clear if the values of the bars represent immigration for the years labeled or for decades. If it is years, it is not clear which label is associated with each bar since there are 20 labels and 19 bars. (It is easy to determine that there are more labels than bars by noting that there is a label to the left of the first bar and also one to the right of the last bar.) However, we know from other sources that the data correspond to immigration by decade. (See Dinnerstein et al. [3] for an excellent graph of the same data.) This is impossible to determine from this figure.

Another problem with Figure 4 is that it suggests that immigration decreased after 2000. However, the final bar represents four years while the others represent ten years, and therefore the short bar is deceptive.

Line Graphs

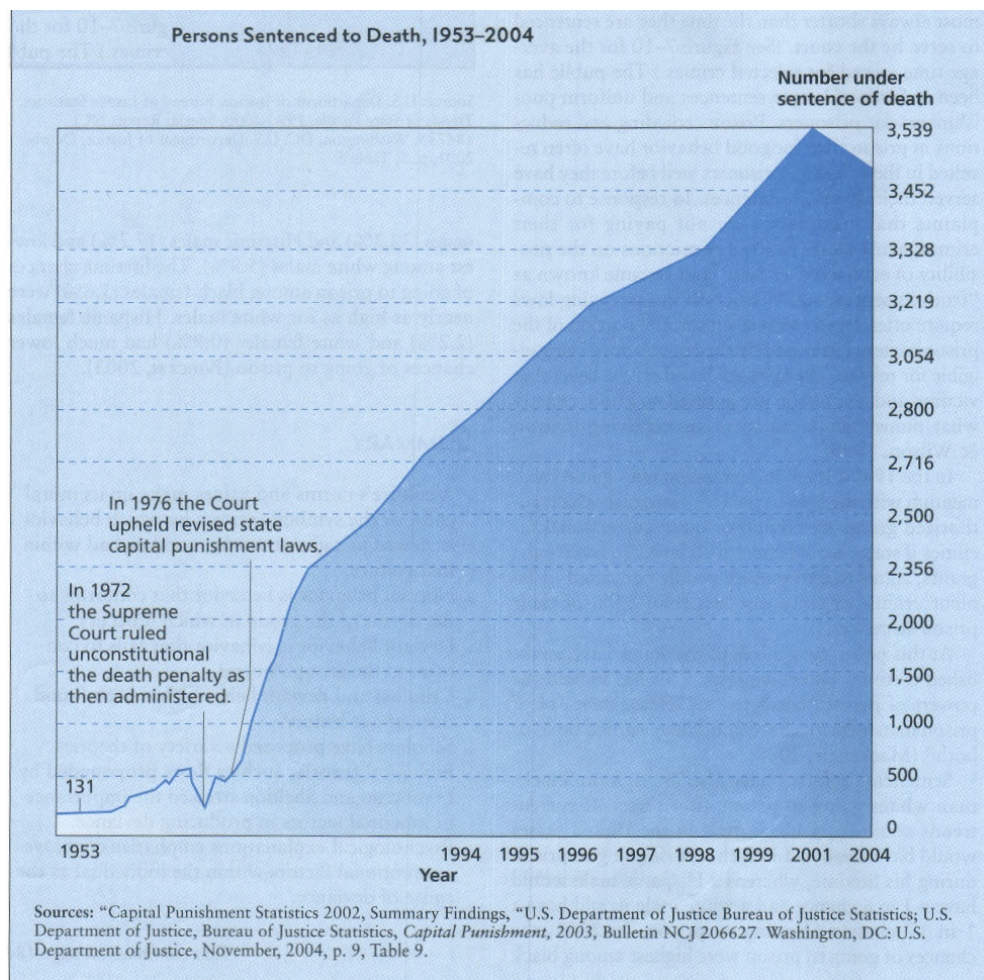


Figure 5. This figure distorts the data by using evenly spaced tick mark labels for unequal intervals.

This figure shows the number of prisoners sentenced to death from 1953 through 2004, a period of just over 50 years. The data from 1953 to 1994 (over a forty year span) takes up less than half of the horizontal axis while the remaining eleven years takes over half. The labels from 1994 to 2004 are evenly spaced but do not all represent the same intervals of time: from 1994 to 1999 the interval represents one year, then the next interval represents two years and the final one three years. On the vertical axis we have evenly spaced gridlines that begin at intervals of 500 prisoners. We see 0, 500, 1000, 1500, 2000, and then 2356. Evenly spaced tick marks should represent the same intervals when using a linear scale.

The footnote in the figure shows that the source of the data is the US Department of Justice. We were curious to see how they presented these data so we went to their Web site and found the following figure.

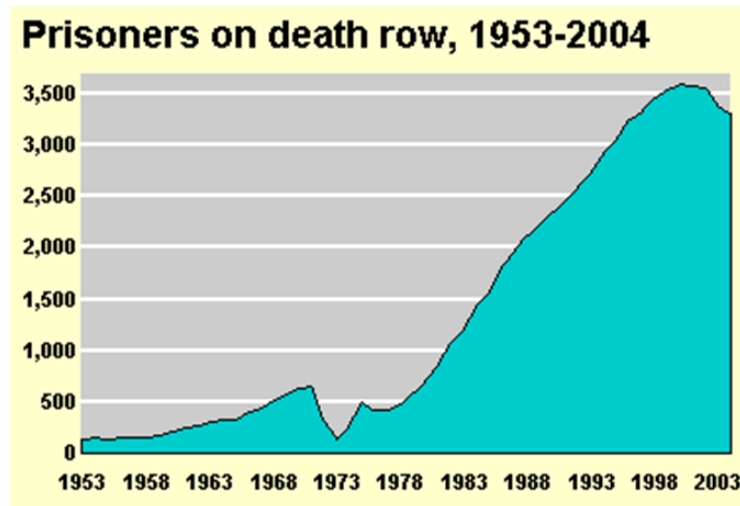


Figure 6. This figure, showing the same data as Figure 5, appears on the Department of Justice Web page.

Notice that the horizontal axis is labeled at five year intervals and the vertical axis at 500 prisoners, so that this figure is drawn to scale. We'd prefer that the horizontal axis had tick marks so we could tell more precisely where each year fell, but this is very minor next to the distortions of Figure 5. If we compare the two figures we notice that the percentage of the time axis with a relatively low number of prisoners on death row is much larger in Figure 6 than in Figure 5. We also notice the differences of the rate of increase. The big bulge in Figure 5 does not appear in Figure 6.

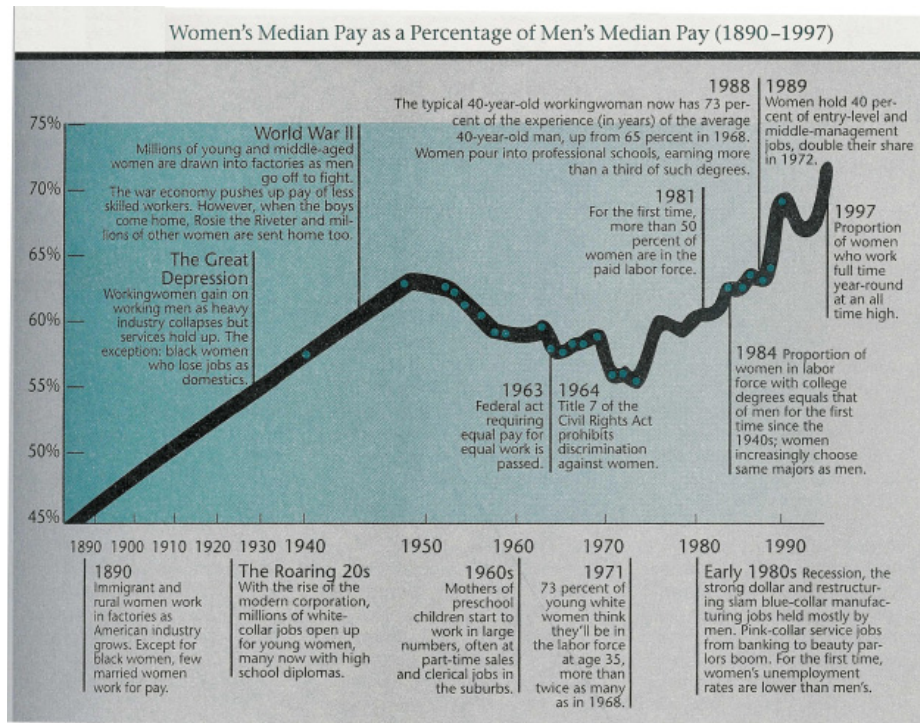


Figure 7. Here all the labels are ten years apart but the tick marks are not equally spaced.

In Figure 5 the labels did not represent even intervals while in Figure 7 the labels do represent even intervals but their distribution over the axis is not even. The end result is the same: a distortion of the data.

Tables

| Worldwide Health Comparisons | | | |
|------------------------------|------------------------------------|----------------------------------|--------------------------|
| | Life Expectancy at Birth (1992) | Per Capita Health Expenditure | Population per Doctor |
| Japan | 78.6 | \$1,771 | 610 |
| Sweden | 77.7 | 2,372 | 370 |
| Spain | 77.4 | 877 | 280 |
| Greece | 77.3 | 274 | 580 |
| Canada | 77.2 | 1,847 | 450 |
| Netherlands | 77.2 | 1,664 | 410 |
| Australia | 76.7 | 1,466 | 440 |
| France | 76.6 | 1,912 | 350 |
| Israel | 76.2 | 509 | 350 |
| United Kingdom | 75.8 | 1,003 | 710 |
| Germany | 75.6 | 1,782 | 370 |
| United States | 75.6 | 2,932 | 420 |
| Ireland | 75.0 | 886 | 681 |

Sources: U.N. Development Program, 1994 Human Development Report; Organization for Economic Cooperation and Development.

Figure 8. The numbers in the middle column are not right adjusted.

Right adjusting numbers makes comparisons much easier. It is also important to right adjust numbers when addition will be performed (not the case here.) The right-most digits (ones' column), the second right-most digits (tens' column), etc. should all line up in the "Per Capita Health Expenditure" column. This table does have some positive features. The data is ordered by life expectancy at birth, which makes comparisons easier. The dollar sign only appears once in the first row. Too many dollar signs clutter the table and make the numbers more difficult to read.

Conclusion

We stress that we are criticizing the illustrations in these books, not the content. The distortions of these graphs should not serve as a role model of how to present quantitative data to college students being exposed to quantitative methods for the first time. We contacted editors of the publisher and were impressed with how seriously they reacted to our comments. Bad graphs are not confined to the social sciences, however; they appear in books and journals in all fields including science, medicine, and business. Therefore, whatever your field, we encourage you to check the graphs for common errors, and take the trouble to contact the publisher if you find misleading or deceptive graphs. Sample text for an email appears in the next paragraph. Editor emails are often included with publishing information in the text itself or can be found on publisher web sites. Please send a note to naomi at nbr-graphs.com if you have other ideas on how to improve the quality of graphs in textbooks. By working together we can raise awareness and set higher standards for communicating quantitative information.

Sample Text (Choose appropriate phrase from bracket): As a {member of the (statistical graphics, data visualization, sociology, etc.) community or as a parent of a college student}, I object to the distortion in the graphs displayed in some of your text books. In particular, the graphs in {name the text you are referring to} should not be used in a curriculum designed to introduce young minds to quantitative ideas. Graphs in textbooks should serve as examples of excellence.

References

1. Becker, Richard and William S. Cleveland. 1996. *S-Plus Trellis Graphics User's Manual*. Mathsoft, Inc., Seattle and Bell Labs, Murray Hill, New Jersey.
2. Cleveland, William S. 1994. *The Elements of Graphing Data*. Hobart Press, Summit, NJ.
3. Dinnerstein, Leonard, Roger Nichols and David Reimers. 2003. *Natives and Strangers: A Multicultural History of Americans. 4th Edition*. Oxford University Press, New York.
4. Few, Stephen. 2004. *Show Me the Numbers: Designing Tables and Graphs to Enlighten*. Analytics Press, Oakland, CA
5. Robbins, Naomi B. 2005. *Creating More Effective Graphs*. Wiley, Hoboken, NJ.
6. Robbins, Naomi B. 2006. "Dot Plots: A Useful Alternative to Bar Charts," http://www.perceptualedge.com/articles/b-eye/dot_plots.pdf
7. Tischler, Henry. 2007. *Introduction to Sociology, 9th Edition*. Wadsworth Publishing.
8. Tischler, Henry. 2002. *Introduction to Sociology, 7th Edition*. Wadsworth Publishing.
9. Tischler, Henry. 1998. *Introduction to Sociology, 6th Edition*. Harcourt College Pub.
10. Tischler, Henry. 1996. *Introduction to Sociology, 5th Edition*. Harcourt College Pub.
11. Tufte, Edward. 1983. *The Visual Display of Quantitative Information, First Edition*. Graphics Press, Cheshire, Connecticut. (second edition 2001).

Note: The tenth edition of Tischler's textbook was published in 2010; we obtained a copy after this article was written. We are happy to say that the bars in the figure corresponding to our Figure 4 are no longer slanted. The other problems with that figure remain. The scales are still distorted on the figure corresponding to our Figure 5 and a number of bar graphs with no zero remain.

About the Authors

Joyce Robbins, Ph.D., is Assistant Professor of Sociology at Touro College in New York. Her research in political sociology combines quantitative and qualitative methods. She received her Ph.D. in sociology from Columbia University, M.A. in sociology and anthropology from Tel Aviv University and B.S.E. in civil engineering and operations research from Princeton University.

Naomi B. Robbins, Ph.D., is a consultant and seminar leader who specializes in the graphical display of data. She trains employees of corporations and organizations on the effective presentation of data with customized programs. She also reviews documents and presentations for clients, suggesting improvements or alternative presentations of the data as appropriate. She is the author of *Creating More Effective Graphs*, published by John Wiley (2005). Dr. N. B. Robbins received her Ph.D. in mathematical statistics from Columbia University, M.A. from Cornell University, and A.B. from Bryn Mawr College. She had a long career at Bell Laboratories before forming NBR, her consulting practice.

Discuss this Article

Share your thoughts about this article by visiting the [Quantitative Literacy Across the Curriculum](#) thread in our discussion forum.

This was published as a guest article in Stephen Few's monthly Visual Business Intelligence Newsletter. A complete library of Stephen Few's articles, as well as other guest articles, is available at www.perceptualedge.com/library.php.