4-2005

Graphic Violence

Dale K. Hathaway
Olivet Nazarene University, hathaway@olivet.edu

"The information conveyed by the graph will often be remembered long after the rest of the article is forgotten."

Graphic Violence

Dale Hathaway
Olivet Nazarene University

You can say anything with statistics. When it comes to statistical graphs, that statement may in fact be true. Graphs and displays are one of the most common places for abuses in the field of statistics. As students of mathematics you have insight into quantitative data lacking in much of the general public. You need to be aware of the common abuses that often take place in graphical displays.

To the general public a display may be more important than the article it accompanies. The display is a visual stimulus that will be seen by those who do not take the time to read the article but do take a quick glance at the graph. The information conveyed by the graph will often be remembered long after the rest of the article is forgotten. Yet all too often the statistical graphs accompanying articles in newspapers and magazines are ambiguous. Below are a few of the ones I've discovered reading my morning paper, the *Kankakee Daily Journal*, over a bowl of Cheerios. The folks at the *Kankakee Daily Journal* were good sports in letting me reproduce some of their graphs; other publications' blunders have been recreated to protect the guilty.

Missing Zero

Probably the most common graphical mistake is the missing zero. Consider the graph of Kankakee Community College enrollment in Figure 1. If we looked at the picture, not the numbers, it would appear that KCC enrollment has decreased by over 50%, but the actual decrease, from 3,555 students to 2,992 students, is around 16%! This misinterpretation of the picture is due to the graph not having a true zero on the vertical scale. The bottom of the graph is actually around 2500 with each increment adding an additional 500 students. While we math types do tend to look at the numbers and would understand the decrease, the general population often does not. The picture is the image that they are left with, and the picture can be deceiving.

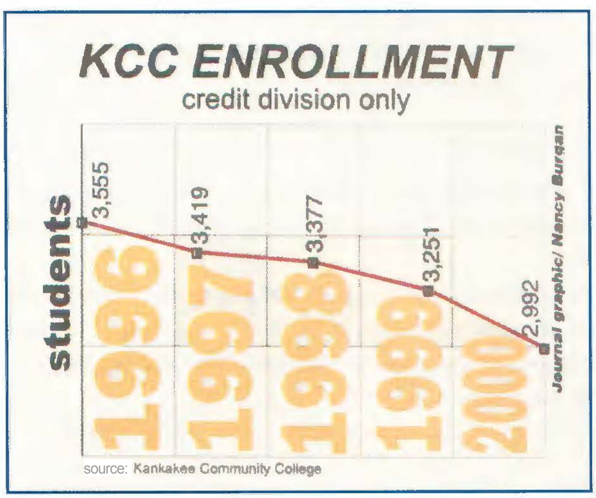

Figure 1. Kankakee Community College enrollment with a missing baseline zero.

Squaring the Effect (or even Cubing)

Another misrepresentation that often occurs is what is known as squaring the effect. A two-dimensional figure is sometimes used to represent the linear data, as in the houses in Figure 2. The problem is when houses of different sizes are drawn, they are drawn to scale, meaning that both dimensions are changed. By changing both dimensions, the linear change is squared, which in turn impacts the interpretation of the graph since we tend to examine change by comparing the area

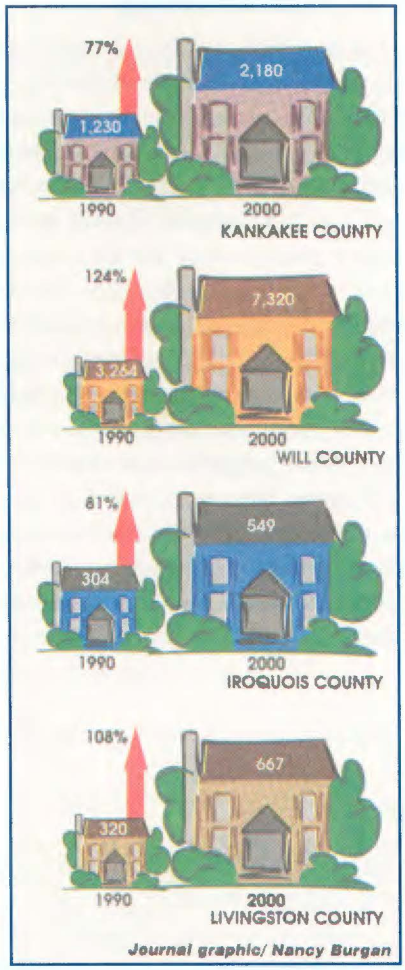

Figure 2. Squaring the effect with houses.

of the two-dimensional figures rather than that of a single linear dimension. For example, the larger house for Livingston county seems about four times the size of the smaller house, but it is only representing a 108% increase, meaning it should be just over twice the size. The actual numbers are included in the houses, and the arrow may be intended to force the visual examination to concentrate on the vertical change, but the end result is that those large houses look much bigger than they should. The person who looks at the picture and not the numbers gets the wrong impression.

This effect can be magnified even further by using a three-dimensional item to represent the data. Consider Figure 3 where a book (all three dimensions are visible) is used to represent the percentages. It's hard to imagine how many of the smallest book would be needed to equal the larger book when you take into account the smaller size on all three dimensions. Yet the smallest book is supposed to represent over one fifth of the largest book. We call this type of graphic abuse 'cubing the effect.'

Hidden Bias

Figure 4 shows hidden bias. The graph, done by a bookstore, shows a typical breakdown of each dollar spent on new books. It seems to show that the publisher gets by far the largest share and that what the bookstore receives is similar to what the author receives. But look closely at the last four categories: freight, school, bookstore salaries, and bookstore earnings and expenses. First of all, it's not clear what the difference is between bookstore salaries and bookstore earnings. Second, the broken-down expenses are paid for by money the bookstore receives. Didn't the publisher have expenses? Didn't the author have expenses? Why are these expenses not broken out the way the bookstore expenses were? By only breaking down their expenses, the bookstore is presenting a biased view of where each dollar goes. Figure 5 shows what the graph would look like if the total percentage each party received were shown. The publisher may get 66% of the dollar, but it has by far the largest expenses in publishing the book. The bookstore gets about 24% of the dollar and the author gets around 10%. Beware of a possible hidden bias any time the creator of the graph has a vested interest in the graph.

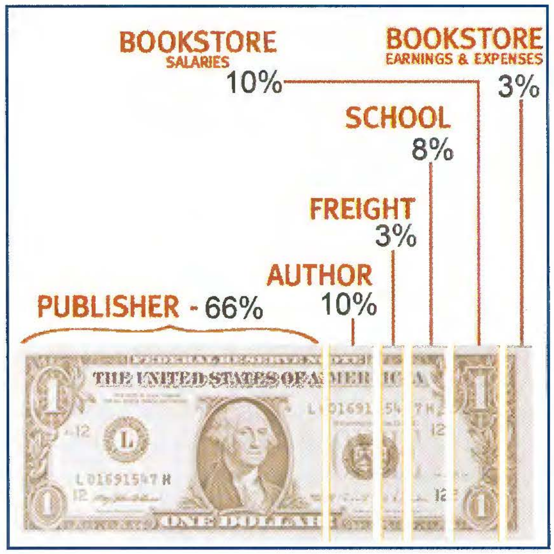

Figure 4. Where your textbook dollar goes, with bias.

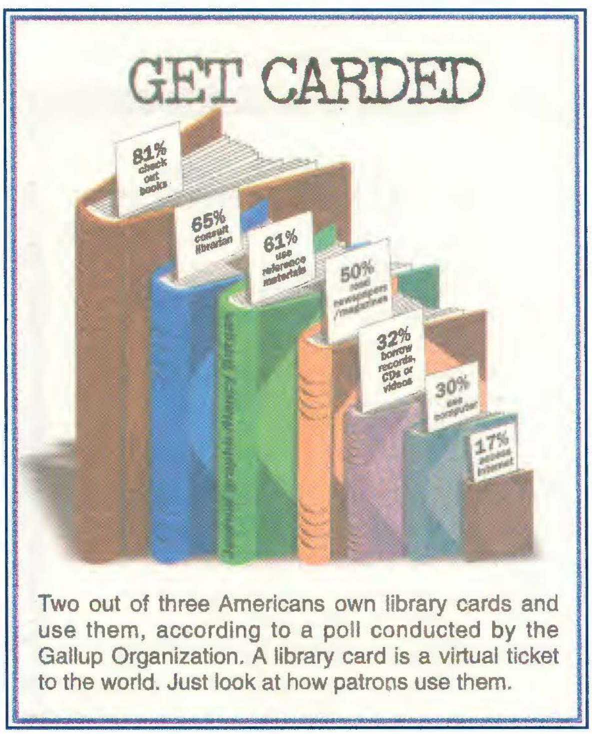

Figure 3. Cubing the effect with books.

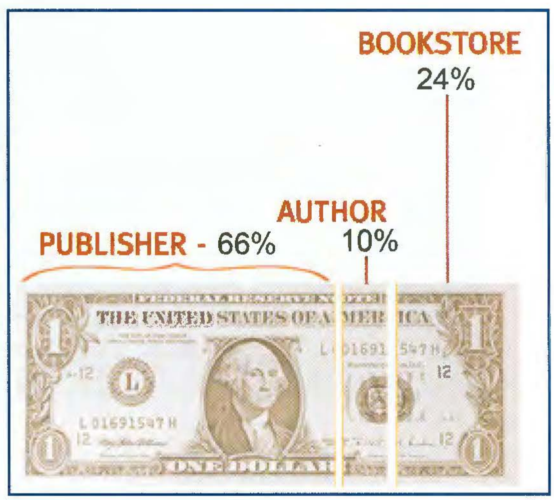

Figure 5. Where your textbook dollar goes, without bias.

Wrong Graph Type

Different graphical displays serve different purposes. Pie charts show a percentage of the whole, line graphs show changes over time, and bar graphs and histograms are used to show comparisons among various items. Each type of graphical display has a specific purpose. Because of this, sometimes the type of statistical display used for a graph is inappropriate. The graph to the right in Figure 6 is presented as a pie chart but should be displayed using a bar graph. Pie charts are used to show percentages of the whole, but here the data was collected in such a way that multiple responses were possible and the percentages total 137%. There is nothing wrong with multiple responses on a question, but the data should then be displayed in a manner appropriate for those responses. Using a pie chart in this fashion only confuses the reader.

Figure 7 is my personal candidate for the worst graph ever created because of the number of mistakes. See how many you can find. The answers are given on page 30.

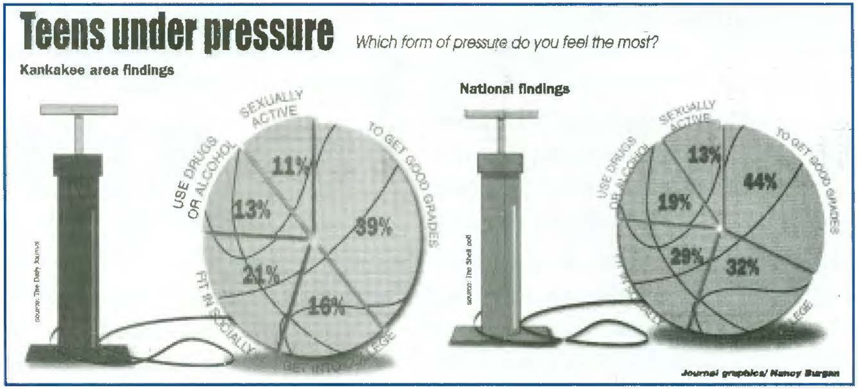

Figure 6. An inflated pie chart.

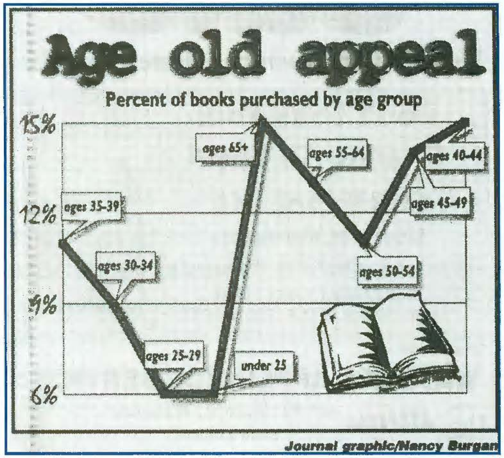

Figure 7. How many mistakes can you find? Answer on page 30.

Answers from page 16

First of all, the ordering of the age categories makes no sense. The ages start in the late 30s, then decrease to those under 25, then jump up to the over-65 category, and finally decreases from there back to the 40s. Youngest to oldest would seem to make more sense since that is the natural progression of ages. Second, open-ended categories like 'under 25' or 'over 65' should be avoided if possible because they are hard to compare to other intervals of a fixed size. Speaking of fixed size, all the middle categories are five years in length except for the interval from 55–64. Since this interval is twice as long as most of the others it cannot really be compared to the others. Another problem is that the graph has a missing zero; the lower end of the vertical scale is 6%. The final problem is that the wrong display type is used. A line graph should be used to show changes over time; the data are percentages of the whole (they do total to 100%). ■