Q: What is the best way to represent data?

By Bill Robertson

BRIAN DISKIN

"Very funny, Art-Boy. Just make sure I get the cartoon on time."

A: To answer that question, let's look at various ways to represent data. Below are several situations along with graphs or charts that help visualize them.

Galen, Martha, Sabrina, and Wally work on their homework different amounts of time each week. Galen averages 10 hours per week, Martha averages 5 hours per week, Sabrina averages 3 hours per week, and Wally averages 1 hour per week (Figure 1).

A car begins at rest and speeds up as it moves down the road. Its speed at different times is given in this table, and represented in the graph in Figure 2.

Time (seconds)	Speed (miles/hour)
1	5
2	10
3	17
4	27
5	41

A study was done comparing the average annual income of 50-year-old adults and the number of years of their formal education. The results are shown in Figure 3.

By the time you've gotten this far, I hope you realize that these three different graphs or charts (Figures 1–3) are really bad ways to represent the data in the three different situations. In Figure 1, the graph indicates that there's some kind of meaning to that line that's drawn between the data points. Is that line supposed to indicate that there's a trend of doing less homework each week the farther down the alphabet you are in the first letter of

Figure 1.

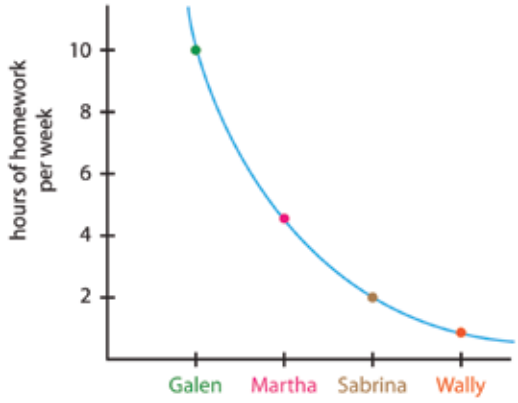
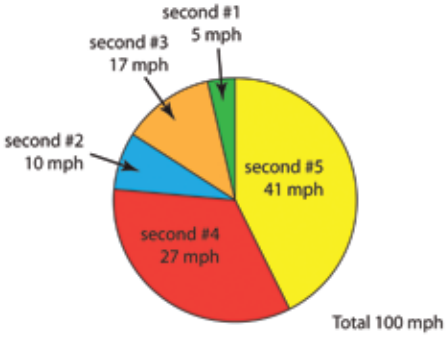


Figure 2.



your name (in case you didn't notice, the names are in alphabetical order)? Nonsense. That line is meaningless. Yet many students get the idea that whenever they graph data, they're supposed to make a line graph as in Figure 1.

Figure 2 is a pie chart, but does it tell us anything more about Table 1 than we already know from the numbers in the table? No, and in fact it's relatively confusing to put these numbers in a pie chart. But pie charts are attractive, aren't they? Why not use them?

Figure 3 is probably the best representation of the data of the three examples, but there's a much better way of showing what's going on. Speaking of better ways, Figures 4a, 4b, and 4c show graphs or charts that are relatively useful for showing the situations depicted in Figures 1–3.

If you just want to show that different numbers are associated with different people or places or groups of people, a bar graph (the graph of students and homework hours) shows this just fine. There is no overall trend. But the graph in Figure 1 implies that there is some kind of trend. Students love to connect dots, and you don't want them doing this when it doesn't mean anything. Similar to the bar graph is the histogram shown in Figure 3. The main difference between a histogram and a bar graph is that a histogram does show a definite trend. Figure 3 shows that there is an upward trend in annual income the more years of education you have. But Figure 3 is limited in what it shows. The graph of annual income and education shown in Figure 4c is much more revealing, in that it shows all the data instead of just categories, and you can judge just how well the two numbers (income and education) are correlated. That graph is known as a scatterplot, and it reveals that while there is a correlation between annual income and education, it's not a strong correlation. (This is

Figure 3.

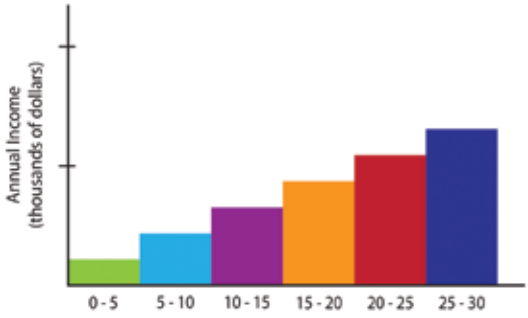


Figure 4a.

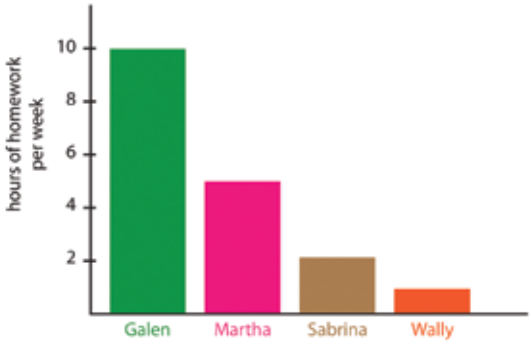


Figure 4b.

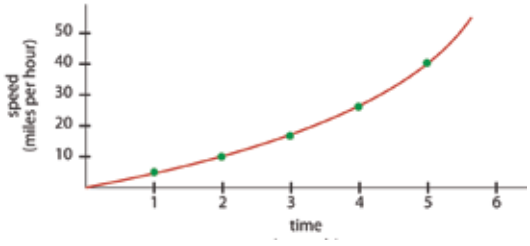
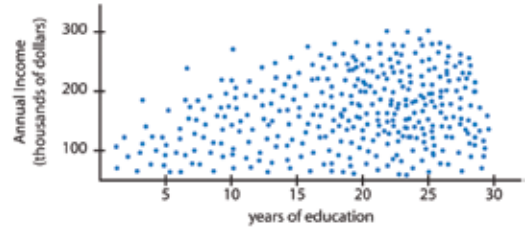


Figure 4c.



made-up data, by the way.) Finally, the graph of speed versus time in Figure 4b shows a definite trend in what the car is doing as time moves along—it's speeding up. That isn't at all clear in Figure 2, which is a pie chart.

General Guidelines

So, there are general guidelines for the best way to show data. If you have a clear relationship between two variables, especially if one variable is changing with time, then a line graph is a great idea. Here, though, you have to be careful about just connecting the dots. That's often not what you should do. If you suspect a correlation between two variables, as with income and education, then you can usually best tell the story by using a scatterplot. If all you want to do is show that different numbers are associated with different people or categories (as with the kids and the homework), then bar charts are great. Pie charts are similar to bar charts, and relative amounts are sometimes easier to see in a pie chart. For example, compare the graphs in Figures 5a and b. They show the same data, but the pie chart demonstrates relative amounts better.

Of course, maybe you don't want to reveal everything in a graph. In that case, there are many ways to hide things. Figure 3 is an example of this. That histogram seems to show a clear trend, even though the scatterplot data indicate that the correlation isn't all that strong between income and education. Figure 6 shows voting data in two different ways. One hides the fact that different regions in a state voted differently, and one makes that apparent.

And if you want to exaggerate your point, you can always cut off part of your graph, as demonstrated in Figure 7. Showing only part of the data can make it seem like an effect is much larger than it really is.

So, there isn't any one best way to represent information, but some methods can be misleading or can even border on being fraudulent. One thing that's for sure is that plotting data points on a set of axes and connecting them with a line won't work for everything, even though that might seem like the most scientific way of doing things. ■

Bill Robertson (wrobert9@ix.netcom.com) *is the author of the NSTA Press book series,* Stop Faking It! Finally Understanding Science So You Can Teach It.

Figure 5a.

Figure 5b.

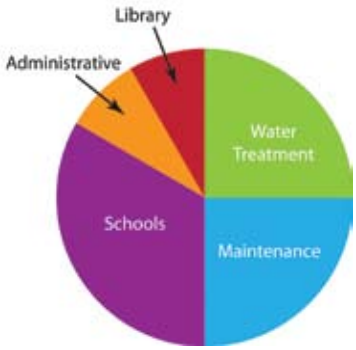


Figure 6.

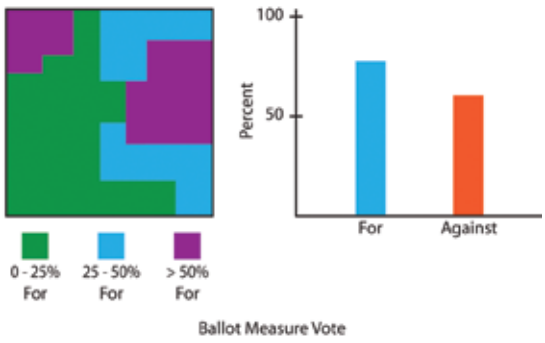


Figure 7.