

This tutorial is going to show you some misleading graphics. The point is that certain times people create graphics that make you want to think a certain way. They're trying to sway you to believe something. And they'll distort or mislead you using graphics to try and get you to believe a certain way.

Let's take a look. We'll start with one that is a pretty obvious visual distortion. So suppose we had three buddies, Paul, Hector, and Juan that were looking at the number baseball cards they had. This was probably drawn by Juan. And we'll get to why we think so in a second.

But you'll notice that the y-axis has unequal scaling. This apparently represents 50, whereas this apparently represents 70. But they're the same size gap. And even more ludicrous, this apparently represents 5. But it's the same gap. So some graphs are drawn with unequal scales so as to represent things disproportionately.

There are some that are more tricky. So let's take a look at the preferred brand of dish soap. So let's say that 15 people selected Brand A. And 8 people selected B. And 20 people selected C.

Now, both of these graphs can show us that. But which one is more accurate? As it turns out, the one on the left is more accurate.

The one on the right clearly was used to exaggerate the difference between B and C. Notice how much taller C is than B in the graph in the right as opposed to the graph on the left. It's only about twice as tall as B here, whereas it's about five times taller than B here.

If you use a graph like that, a bar graph or a histogram, it's a good idea to start the vertical axis at zero, unless there's a good reason not to start there. For instance, if you were tracking the change in home prices starting from zero and going all the way up to say 300,000 won't show a big difference between 300,000 and 280,000.

However, to the homeowner, that drop in \$20,000 is significant.

Graphs beginning anywhere besides zero have a tendency to exaggerate differences. But graphs starting at zero sometimes can minimize very real differences. So it's important to know what you're trying to emphasize.

Let's take another look. Suppose a class of 18 students was asked their favorite sport and some kids said soccer and some kids said baseball. Three said soccer. Five said baseball. And the remaining 10 said basketball.

Suppose a student drew this graph. Now, what's wrong with this graph? Well, twice as many students chose basketball. And look. It goes up to 10. And five students chose baseball. And look. It goes up to 5.

So what's the problem? Well, the problem is that while the height of the basketball is twice the height of the

baseball, it's also twice the width. And so something that's twice the height and twice the width, if you compare the areas taken up by the basketball and the baseball, it's about four times as much area taken up by the basketball. In fact, you can even see that more clearly by putting the box that represented the baseball inside the box that represented the basketball. And it's clearly only about 1/4 the size.

To make matters worse, technology has introduced us to lots of different misleading graphs. This is kind of a ridiculous graph. It's meant to show across four different cities-- I don't know what it's supposed to be measuring, frankly. There's no label.

It's supposed to show across time how these different markets are doing. Maybe this is a store. I can't even tell what this is supposed to be measuring or what these numbers zero through 100 actually mean.

So the additional problem comes from what are we comparing here? Are we comparing the height of these things, the volume of these things. For instance, the cone to the cylinder, a cone only has about 1/3 the volume of a cylinder with the same base and height. And so it doesn't really make sense to be comparing cones and pyramids to cylinders and boxes. It just makes no sense.

Also, because this graph is three-dimensional, there's no way to easily compare heights. Is this cone supposed to be taller, shorter, or the same height as the cylinder behind it? I can't tell. And it's going to be very hard for anyone to tell. This is an incredibly misleading graphic.

To make matters worse, technology, like certain spreadsheet programs, have allowed you to look at all of these different graphs. Now, cones, cylinders, and all of these being three-dimensional, all that these do is distort your data. The best choice if you're going to use bar graphs would be the simple ones, the simple two-dimensional ones here at the top.

So taking a look, this is not a good choice. The better choice, being that you're trying to compare across time, would be something like a time series. January, February, March, April, May, June, July. Again, I'm not even 100% sure that these are where they're supposed to be, because I couldn't compare the heights easily. I had to guess at what these values were going to be.

But this is a lot more useful to anyone reading it than this would be. This is flashier. But the information gets hidden.

And so to recap, graphical displays can be manipulated in many different ways. If you use an inappropriate scale, you can exaggerate the differences. Or you can use areas to make differences seem larger than they actually are. Or you can use three-dimensional displays that aren't really clear at all.

As statisticians, our goal is to make the complicated simple, to make the data easy to understand. We're trying to clean up a messy world. Our goal is clarity. And all these misleading graphics don't do that.

So the terms that we used were misleading graphics; perceptual distortion, which was the area issue that we had with the baseball and basketball example; and scales. Where do you start your graph? Do you start it at zero? Or do you start it somewhere else?

Good luck. And we'll see you next time.