Teacher Tune-up

Quick Content Refresher for Busy Professionals

Graphing Scientific Results

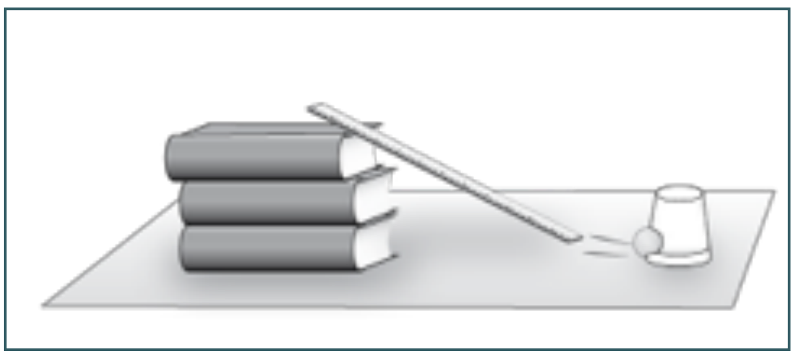
How do you graph the results of a scientific investigation? The best answer is: any way that tells the story. Your students may have many ways of displaying their results. Celebrate them all, but ask the class to find graphs they think are good at showing what's important, and discuss what features make them good.

Some aspects of graphs are worth pointing out and insisting on early: Neatness. Labeled axes. Clear labels and scales for the axes. Units for quantities.

There are, however, additional conventions about graphs that you should know about. Students don't need to attend to these at first (or maybe ever), but it's reasonable in a class, when you have enough experience with "invented graphing," to nudge students towards orthodoxy, all other things being equal.

Independent or Dependent?

Scientists distinguish between independent and dependent variables. In an experiment, the independent variable is the aspect that you choose to change or manipulate in some way. The dependent variable is the aspect that you measure to understand its response to the independent variable. So in an experiment where students roll a ball down a ruler propped up by books and see how far the ball pushes a paper cup it collides with at the bottom, the number of books used to tilt the ramp is the independent variable and the distance the cup moves is the dependent variable.



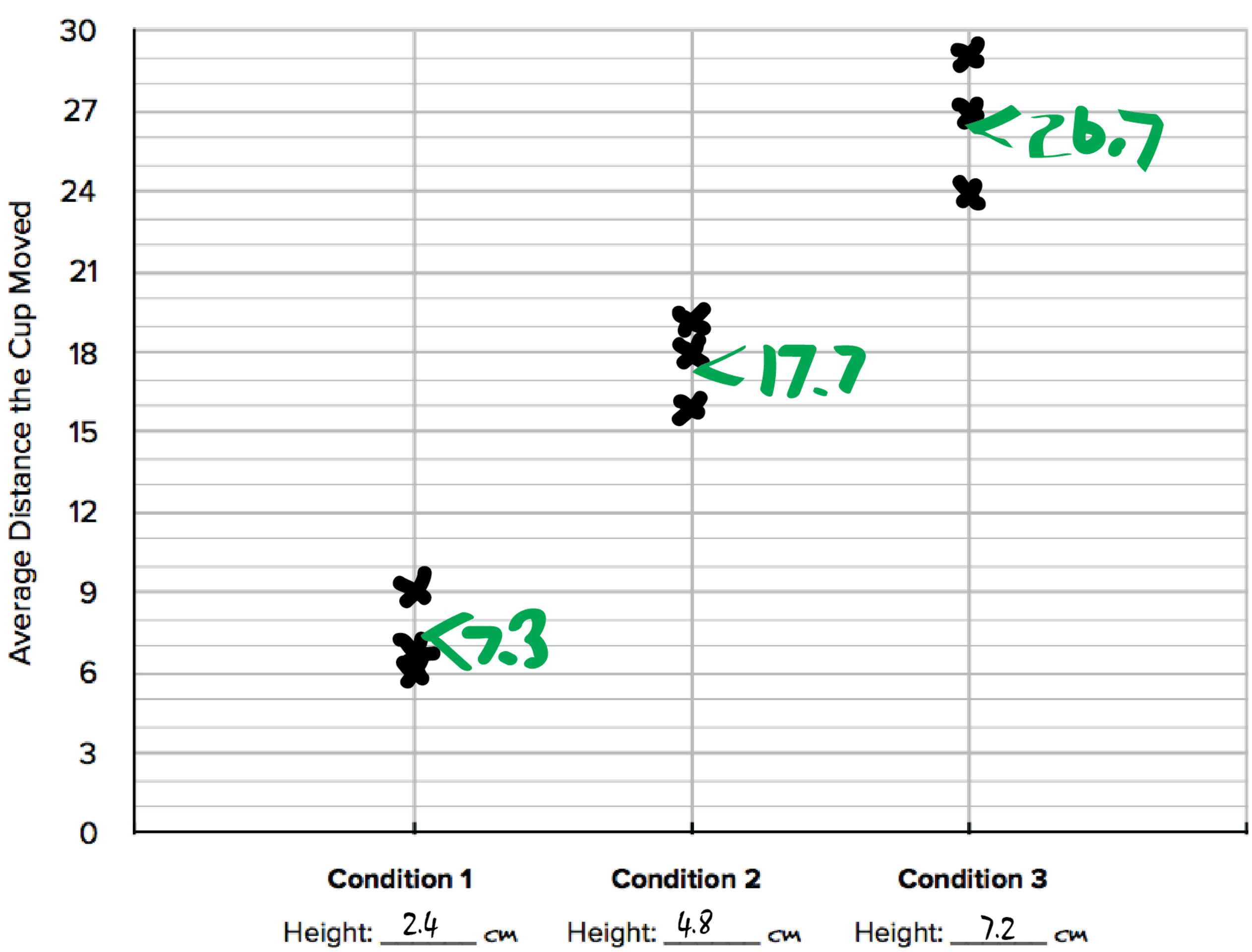
This terminology is confusing to many students (and adults), perhaps because the words are so close. There are alternatives. Some teachers simply call them x and y variables. Some scientists use the terms treatment (for independent) and response (for dependent), which, for experiments, makes a lot of sense.

Different names for the two variables

Independent	***Dependent***
X	*Y*
treatment	*response*
This is the variable that you manipulate directly.	*This is the variable you measure to collect and analyze data.*

What Goes on Which Axis

In general—but not always—the independent (treatment, x) variable goes on the horizontal axis. This placement holds true whether that variable is numerical or not. The dependent (response, y) variable goes on the vertical. For example, in the aforementioned activity where students roll balls down a ruler, the treatment—the height at the elevated end of the ruler—goes on the horizontal axis. (Note in the following image that students conducted three trials for each height, with distances marked on the vertical axis and an average result calculated for each condition.)



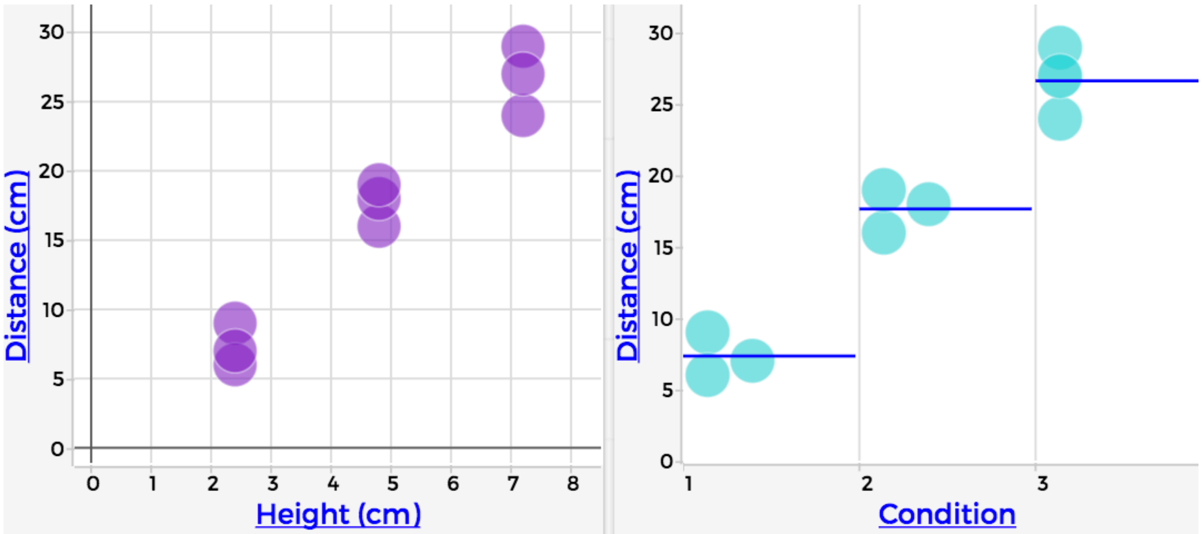
(From Unit T3: Claims and Evidence)

It's not clear why this axis placement of variables is the convention. Perhaps it's because it matches the mathematical convention for a function, where x is the "input" and y is the "output." If you enjoy a good mnemonic, here's a way to remember the convention: DRY MIX.

D = dependent variable
R = responding variable
Y = graph information on the vertical or y-axis

M = manipulated variable
I = independent variable
X = graph information on the horizontal or x-axis

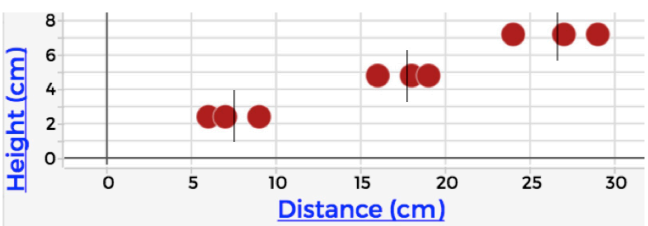
Here are two computer versions of the same graph. Notice the difference in the graphs between how Height and Condition are treated.



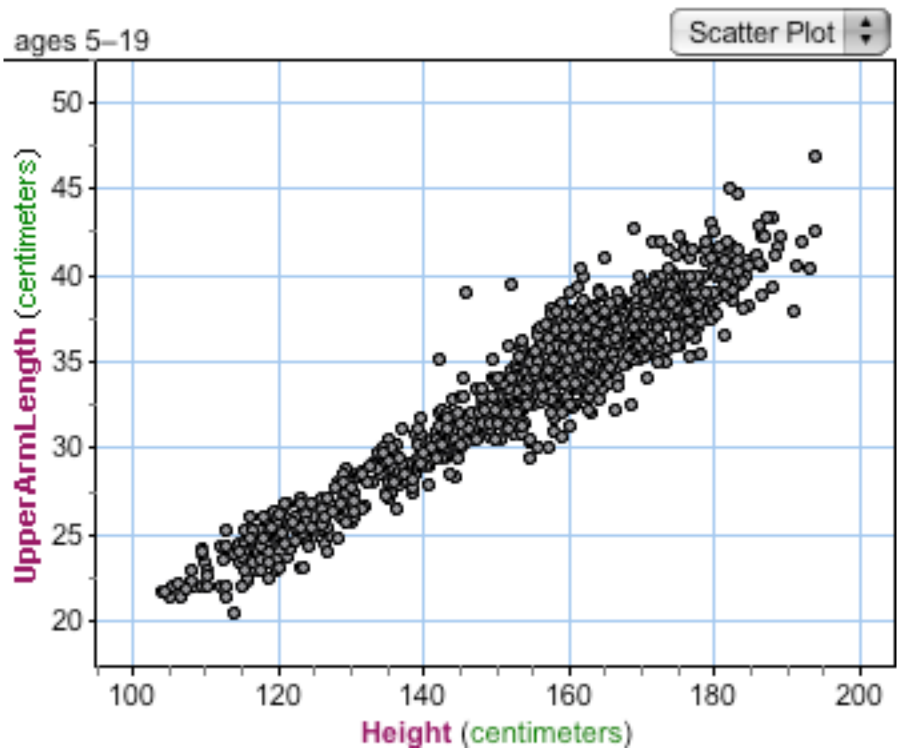
The graph on the left is a scatter plot that shows two numerical or continuous variables—the height and the distance. The one on the right, which shows the same data, is three parallel dot plots, with a separate dot plot for each condition. In this case, each condition is a "categorical" variable. Frequently, categorical variables have values that are words such as "male" and "female."

We use categorical variables to define groups in experiments. The ball-rolling activity actually involved three separate conditions, and things were supposed to be identical for all three rolls within each condition. So there are really three separate groups of data, and our right-hand graph would be a sensible alternative to the scatter plot: we want to see how each condition affects the numerical distance. Also, it makes sense to compute an average for each group, given that three trials were conducted under each condition.

Some variables have a natural direction, however, and this can trump the conventions. Height is naturally vertical. Width is naturally horizontal. In the ramp activity, you can make the case for putting the distance the cup moves—which is horizontal—on the horizontal axis. So even though we determined the height (so height is independent), a graph like this makes a picture that more resembles the actual situation, in that a steeper ruler results in a greater distance rolled.

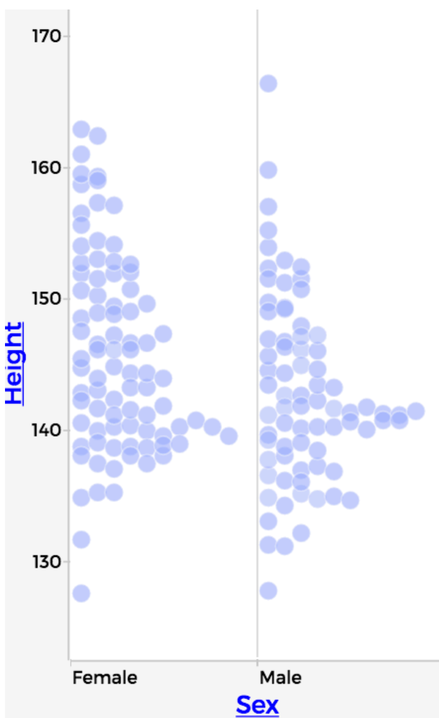


The distinction between independent and dependent variables doesn't apply in some non-experimental situations, such as in observational studies. For example, suppose you collect data on students' height and upper arm lengths. Neither is a treatment or response; so you can put the labels wherever you like.



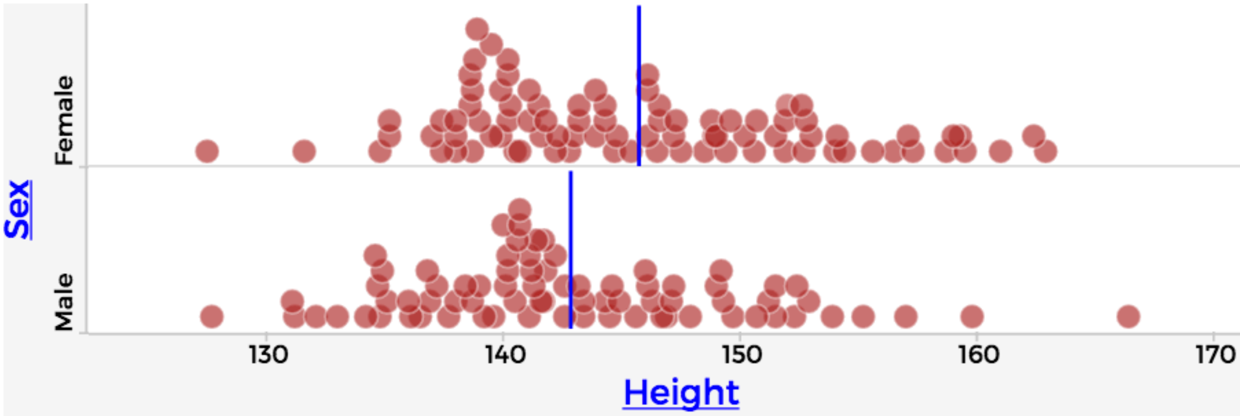
(1000 children and teenagers. 2003 data from NHANES, a national health survey.)

In other situations, you might have a variable that represents the grouping you're interested in; then it takes on the "independent" role. For example, suppose you're studying the heights of 10-year olds, comparing males and females. Sex is not a treatment; nevertheless, it's acting like an independent variable, and you can argue that you're studying whether height "depends" on sex.



(163 10-year-olds in the United States. 2003 data from NHANES, a national health survey.)

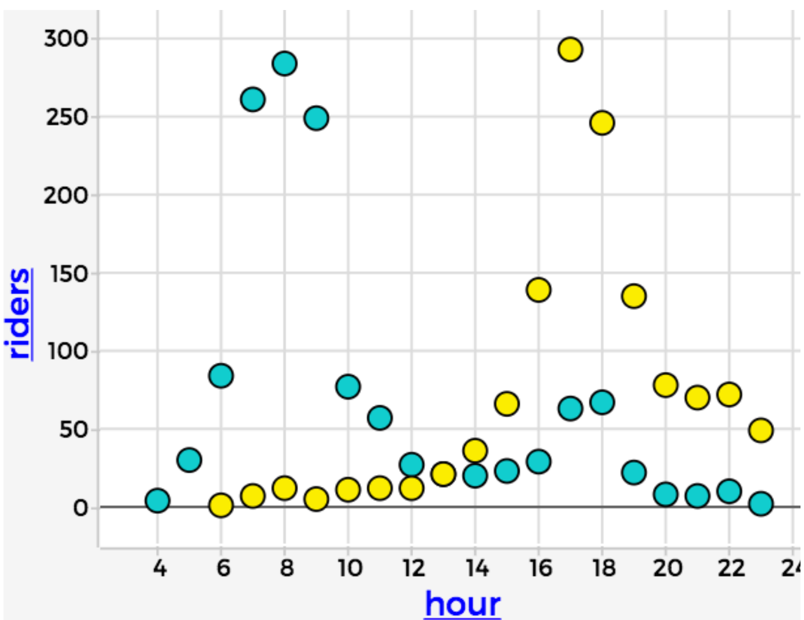
Other considerations as prosaic as the shape of the graph can change the choice of axes. If you have more room for a horizontal graph, you might make that height graph like this. You can also add a line for the mean.



(same NHANES data on 10-year-olds)

Plotting Change over Time

Change over time is an important special case. If you're graphing something that changes over time, the time variable goes on the horizontal axis. This placement is consistent with the dependent/independent idea, since we can never control time. For example, this graph shows the number of passengers each hour, riding the train from an outlying station into the city (blue) and back (yellow) over the course of a day. A graph like this, with data at a sequence of times, is called a time series.



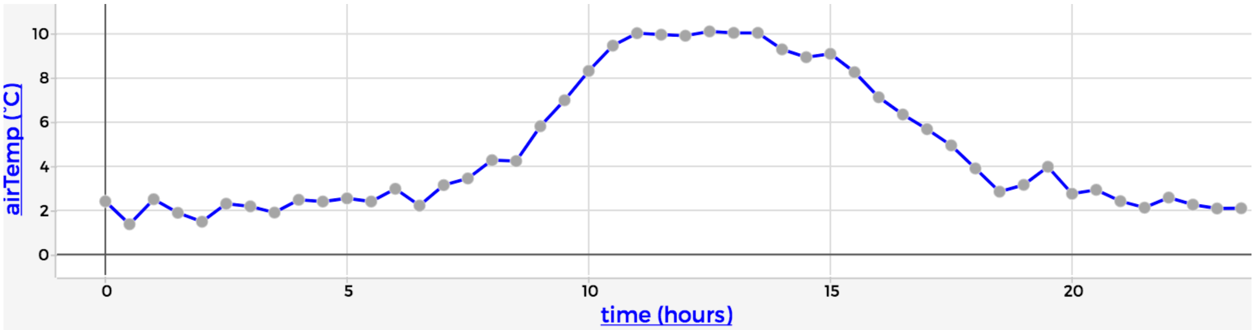
(Bay Area Rapid Transit data, April 15 2015, between Rockridge and Embarcadero stations)

In this graph, notice how we use color to distinguish the two things we're comparing—the inbound and outbound patterns. (Notice, too, that you can detect a slight "reverse commute": people who travel from the big city to work in the outlying area, then return to the city against the predominant flow of passenger traffic.)

By the way: this rule about "time" applies to the time at which something happens, not to a duration. Suppose you have an experiment about how long you can hold your breath. That "time" could go on the vertical axis.

Connecting Lines

In general, do not connect the dots on your graphs! You can make an exception when the quantity is changing continuously, so that it makes sense to imagine values between the dots. Temperature is a good example.



(January 8 2000: air temperature at a station in Blodgett forest, California)

When you can, plot all the data!

In our graph from the ball-rolling investigation earlier, students plotted the individual data values as well as the averages. Why is this approach a good idea?

- It lets you see the relationship between an average and its values.
- If the student makes a mistake in computing the average, you can see it on the graph.
- You can see the variability in the original data.