"Wait, but what am I supposed to graph?" – A guide to pulling unknowns out of graphs *or* 'How I got LoggerPro to tell me what I want'

When completing a lab that is investigating *what a given relationship is* (e.g. the relationship between the force stretching a spring and how far it stretches), it is straightforward to determine what to graph. The research question tells you: if you want the relationship between blangos and blorts, you should graph blangos on one axis and blorts on the other. *We don't know the relationship, so we want to graph the data that most straightforwardly shows us that relationship*. Traditionally, we plot the independent variable from our experiment on the x axis and the dependent on the y, but that is just a convention.

Now, when completing a lab that is trying to *find or determine some quantity*, it is a different story. There in fact are many approaches to graphing in cases like these, depending on what the research question is asking. Described here is one approach that is effective with the use of curve fitting software. (there are other tricks when we only have pen and paper)

Let's take a silly example, abstracted so you can not get tied down with the context. My research question is "What is the value of D for this system?" I know that D is an important quantity when thinking about how my system works, but it is something that isn't directly measurable. I can't walk up to my system and ask, "Hey what's your D?". But, I can measure and calculate a bunch of other stuff: *B, C, n, Z,* and even *U*. I'm even clever enough to know that there is a relationship between all of these things, and I can write down an equation that I learned in class:

$$4BZ = \frac{1}{4}Cn^2 + \frac{1}{3}\frac{Dn^2}{U^2}$$

Knowing this equation, and knowing I can measure all those variables, I get really excited because, hey, I can just solve for D after I do my measurements! I do some algebra, and I get to:

$$D = \left(\frac{48BZ}{n^2} - 3C\right)\frac{U^2}{4}$$

If I do one measurement of all of my parameters, and plug things in, I indeed can calculate a value for D. If I'm a bit more careful, I realize that I should at least take multiple measurements and do an average. Sometimes, this is sufficient to answer the question. However, this is prone to a major pitfall: *what if my original equation was in some way wrong or missing a component?* If that were the case, all of my calculations would be incorrect, because they were built from an originally incorrect equation. It would be very difficult for me to identify *in what way I was wrong*, because all I have is the final output. Typically, *solving the equation for the variable you are interested in can be a weak means of actually determining it. Don't do this for your moment of inertia lab!*

What are you supposed to do? I can measure all of these different things for this system, what do I graph?

First, ask: what is straightforward to manipulate in my system? Depending on what you're actually working with, it may be easier or harder to vary one of these many variables in your equation. You should choose ones that you can experimentally vary and measure most easily. **In my example let's say Z and n are easy to find.** If you have a lot of options, choosing one pair is a good place to start.

Next, determine which of your two variables is the simplest to solve for in your equation. In my example, it is easy to write a function that says Z = _____, and somewhat more complex to solve for n = _____. **As much as possible, we want to minimize algebra steps** so that we can more directly see the ways in which our function doesn't work well for our data. In the example, solving for Z gives:

$$Z = \frac{1}{4B}\left(\frac{1}{4}Cn^2 + \frac{1}{3}\frac{Dn^2}{U^2}\right)$$

(Don't lose sight, we are trying to determine D!)

Looking at this, I know that when *n* changes, *Z* changes, as long as everything else is constant. In fact, I know that if I plot *n* on the x-axis, and *Z* on the y-axis, there SHOULD be a *quadratic* relationship (as long as I'm right about my original theory). This gets a bit easier to see when I factor out the *n* squared:

$Z = \frac{1}{4B}\left(\frac{1}{4}C + \frac{1}{3}\frac{D}{U^2}\right)n^2$ or in other words $Z = (SOMESILLYNUMBER)n^2$. That's definitely a simple quadratic relationship. D should somehow be incorporated in the value for my SOMESILLYNUMBER when I do a curve fit.

At the start of your lab, you should explain what your goal is to graph, and explain the original theory that helped. For example: "I will be plotting Z on my y-axis and n on my x-axis. I expect that there should be a quadratic relationship between the two because I know that the _____ of the system should obey the relationship $4BZ = \frac{1}{4}Cn^2 + \frac{1}{3}\frac{Dn^2}{U^2}$. This reflects how the ____ becomes both ___ and ___, which makes sense because ________. By varying Z and n, and measuring all the other parameters, my quadratic curve fit should depend on D. By this method I will answer the research question."

Concerned student says: "Ok so if I use Z on my y-axis and n on my x-axis, it should be quadratic. But Z isn't my dependent variable… :'(" Don't forget that **the convention of plotting the independent on the x-axis is just that, a convention. Do what works best for your analysis!**

Now we go do the lab. Since you are able to directly measure everything (Z, B, C, U, n) except the thing you want to find (D), you are in good shape. With the two variables you selected as the variables plotted on your x and y axes (a scatterplot), we can use our measured values of the other parameters to have LoggerPro find D. To do this, I click on "Define function…" in the curve fit window, and carefully type in the right hand side of my formula: $Z = \frac{1}{4B}\left(\frac{1}{4}C + \frac{1}{3}\frac{D}{U^2}\right)n^2$, except I plug in my measurements for my constant *B, C,* and *U* values. Write a letter to represent the thing you don't know (Sometimes you can have multiple!!!), in this case D. The end point would be (lets say B was 10, C was 8 and U was 2): $Z = \frac{1}{4(10)}\left(\frac{1}{4}8 + \frac{1}{3}\frac{D}{2^2}\right)n^2$.

It is often a good idea to NOT multiply this stuff out by hand so that (1) you can easily adjust values in case you realize U was actually 2.3 or something, and (2) you don't have to round, as LoggerPro will do the calculations. Really, writing $Z = \frac{1}{40}\left(2 + \frac{D}{12}\right)n^2$ or even more simplified versions is OK.

When I do a curve fit with this function, LoggerPro will determine which value of D best aligns with your data. Your curve fit has answered your research question! Wooooo… (Now you have to go figure out if your answer is reasonable)

Why is this process better than averaging and just solving for D from the start:
-You can see the ways your theory was wrong as you attempt to use it to find a value. If you undershoot your points, you may not have actually measured your y variable correctly. If you have an unexpected intercept, it means that there was an additional term in your original theory.
-The fit does its best to hit all of your data points at once, and you can easily tell which of your data points are less trustworthy than others on the graph. When averaging, the less accurate points and the more accurate points all get lumped together.

Note: solving for D just using your equation and a single point is a reasonable way to get a sense of whether your procedure is on the right track before you spend a lot of time doing measurements. **Do not expect individual values to be reliable when calculated one-by-one.** If they all gave the same answer, that would mean your points perfectly fell on the same line, and we all know that is rarely the case.