

Using Microsoft Excel for Constructing Graphs

file: d:\B162-2014\graphing_using_Excel.wpd
date: September 14, 2014

Introduction

Previously we used Microsoft Excel in its most basic mode: as a record keeping device and for doing some simple calculations.

Another very useful feature of spreadsheet programs is their ability to help you construct graphs quickly and easily. In days past, a scientist might construct a graph in a spreadsheet for a quick look at the data and then produce the final graph (for a presentation or publication) using a much more sophisticated (and expensive) graphics package such as SigmaPlot. As Excel has evolved, so too has its graphing capabilities and the current version can do a respectable job of producing almost-publication ready graphics. There are still some annoyances and limitations but you can use Excel for most of your graphing needs.

Graphing has another use besides presenting nice-looking images. Graphing can powerfully portray vast quantities of information and it can be used (and misused) to influence interpretation of information.

Probably the finest examination of this topic is found in Edward Tufte's books, starting with

Tufte, E.R. (1983) *The Visual Display of Quantitative Information*. Graphics Press, Cheshire, Connecticut.

and Tufte's subsequent books.

Tufte argues that graphs are seldom used to their fullest potential, and in fact, are quite often misused. Some of the topics he covers include:

- Graphic ducks (where the presentation assumes a life of its own; quite common with computer generated graphics)
- Scaling and miscalling, particularly in reference to area (i.e. 2D graphics)
- Misuse of ink (use ink to tell the story, not for its own sake)
- Placement of information on a graph (show how adding labels can make a regression line look even better)

Be sure to take a good look at this book. Look at Minard's masterpiece graph depicting Napoleon's march on Moscow. This graph tells the fate of 420,000 men, all but 10,000 perishing on the march to, and particularly from, Moscow.

Strive for such elegance in your own graphs. Never turn in a graphic duck such as an "exploded 3D pie chart".

It is quite clear that the designers of Microsoft Excel did not read Tufte's book because of some of the default settings they built into graphs; however, fortunately, these can be overcome with just a few keystrokes.

Things to consider when making graphs:

Scaling: Use the full page for the data, i.e., choose a scale for the axes that plots the data and not a bunch of white space (unless that is your intention)

Background shading: For this course, **always remove background shading** for any printed assignment. Background shading is only of value for visual presentations, e.g., slide shows or PowerPoint presentations. It is never used in printed publications because it detracts from the image and doesn't reproduce correctly anyway

Legends: Legends can be useful, but position them with care, so as not to use up tons of space that should be devoted to information. Also, never put a legend on a graph that has only one symbol (which in many cases means that you will have to delete the silly legend that Excel puts on every graph by default).

Titles: Graphs in journals do NOT have titles, whereas those in PowerPoint presentations often do. In printed publications, the title goes in the Figure caption, below the Figure, which is produced entirely independently of the figure.

Why all this concern about using space wisely? The bottom line is that space costs money. Space on a printed page costs a lot of money. Two examples will illustrate this.

Scientific publications, such as journals, are enormously costly to produce because they are printed on high quality paper and have small distributions (a few thousand copies). Editors and printers regard every square centimeter of page as sacred. Authors frequently have to pay actual hard, real dollars to use that space (~\$100/page is typical). As a result, you will seldom find a full-page graph in a scientific journal.

Consultant reports are exactly the opposite. Typically a consultant is being paid to produce a report on some topic, e.g., the consequences of releasing water at the Nimbus Dam on salmon spawning in the American River. Although few will admit such a thing outright, both contractor and contractee seem to prefer a nice fat report. A fat report with tons of graphs surely indicates that a lot of work was done and much useful information was found. This is crap. Most such reports are submitted and dutifully filled where they belong: on a shelf, only to be tossed out a few years later. If you look at many of these reports you will see why. They are often loaded with Appendices of charts and graphs. Often you will see multiple, similar looking graphs, plotting the same information but with one variable changed, e.g., Graph 1: the number of sockeye salmon; Graph 2: the number of coho salmon; Graph 3: the number of chinook salmon, etc.... In many cases, were someone forced to pay per graph, the authors would find a much more succinct way of presenting the information. This would have two advantages: it would use less paper, and surprisingly, one well-done graph may be far BETTER than ten poorly done graphs at conveying the important data.

Some terms

Independent variable: this is the variable that is free to vary; typically plotted on the x-axis (the horizontal axis)

Dependent variable: this variable depends on the independent variable; typically plotted on the y-axis (the vertical axis)

Let us say that you are plotting a graph of your grades versus the number of hours put into assignments. In this case, the number of hours invested is the independent variable. You choose how many hours to put in. The grade that you get depends on the number of hours that you put in. Therefore the grade is the dependent variable.

Not every graph has true dependent and independent variables. For example, if we plotted a graph of Standard Length versus Age for a bunch of fish, which variable is dependent and which is independent? The truth is that neither variable is independent. Nonetheless in those cases we may often refer to the variable plotted on the x-axis as the independent variable.

When we say "plot **something** versus **another-thing**", we mean to plot "**something**" on the y-axis and "**another-thing**" on the x-axis by convention. In other words, we plot "y versus x" and NOT "x versus y". It is very important that you understand and remember this for the rest of your life. Getting this simple convention backwards will cost you huge numbers of points on exams and make your life miserable.

Making a simple XY graph

We are going to make an XY graph (often called a Scatterplot) showing the relationship between Total Length (TL) and Standard Length (SL) of a sample of fish.

To do so, start a new spreadsheet.

In the first row, enter your name and a title, e.g., "John Smith – Simple XY Graph"

In the second row, I usually put the filename, e.g., "File: simple_graph.xlsx" This helps you to find the file later on, when you are looking for it.

In the third row, enter the date, e.g., "Date: September 14, 2014" [Note: it is bad practice in science to write dates as 02/10/14. The reason is that the ordering used in the United States, i.e., month/day/year is exactly backwards from that used in most of the rest of the world]

Get used to always putting this kind of information at the top of your spreadsheets.

In row 6, enter the following titles in cells A6, B6 and C6 respectively: Fish, SL(mm), TL(mm)

Starting in A8, enter the following information, under the appropriate title.

<u>Fish</u>	<u>SL (mm)</u>	<u>TL (mm)</u>
1	135	150
2	182	200
3	203	220
4	220	245
5	234	247
6	256	280
7	266	279
8	277	290
9	287	310
10	300	340

We will now use the Chart Wizard to help us make a graph. There are various ways to use this wizard; I strongly suggest the following. It is not the quickest, but it does ensure that you graph what YOU want, not what the designers of Excel want to graph.

Put your cursor at cell E5. Select Insert | Scatter. Select "Scatter with only Markers" (the upper left of the five choices).

REALLY IMPORTANT: Several things happened when you created a chart/graph and some of them are not immediately obvious. Look at the main menu along the top of Excel. When you created a new chart, that menu added THREE items to itself. These are collectively known as the Chart Tools. They include "Design", "Layout" and "Format" tabs. You are automatically put into the "Design" tab to select data series, but you will soon need other functions that are not there... such as the ability to add Axis labels. Where are they? They are under the "Layout" tab.

Also, notice that if you click outside of the graph, e.g., by accident, those tabs go away! Click in the graph and they reappear.

First, position the graph just to the right of the data columns and shrink it a little so that it will all print on one page.

Next, let us add some data. Click "Select Data". The "Select Data Source" dialog box will appear.

Click on the "Add" button. Up will pop the "Edit Series" dialog box that you can fill in to add a series to the graph. [This is 100% preferable to using the box at the top of the Select Data dialog which says "Chart Data Range". I strongly discourage you from using that box as it will frequently give you the wrong answer]

To work with the "Edit Series" dialog box, do the following:

Type a name in the Series name box, e.g., Total length vs Standard length

At the right end of the "Series X values" box, there is a little spreadsheet icon. Click on it. A small dialog box will appear, but most importantly, now you can use the mouse to select a range of data. In this case, the X values are the SL values. So, select cells B8..B17. Hit the Enter key.

Do the same thing for the "Series Y values" box, but this time select C8..C17 and hit the Enter key.

Select the OK button and you will leave the "Edit Series" dialog.

Select OK to leave the “Select Data Source” dialog box.

You should notice that Excel has done a bunch of work for you and created a graph of the Total length data versus the Standard length data. However, Excel has also done a bunch of stuff that you didn’t want it to do and now you have to correct that.

First, notice that Excel has connected the points with lines. This is completely incorrect. Anytime a line connects two points, that means that the intermediate values are valid, i.e., any point on the line is meaningful. This is not true in this case.

So, we need to get rid of those lines. Using the mouse, on the graph, right click on one of the data points or on one of the line segments. A menu will pop up. Select “Format Data Series”. Click on “Line Color” (not Marker Line Color – that is something else) and select “No line”.

Select “Close” and click anywhere on the graph, but not on a datapoint, gridline or legend. You should now see the data points properly displayed.

Delete the useless legend. There is no point in having a legend for only one data series. To do so, right click on the legend and “Delete” it.

Remove the horizontal grid lines. To do so, right click on one of the lines and “Delete”.

Now it is time to add Axis titles. Recall the **Important Notice** above and click on the “Layout” tab. Now you have access to all sorts of things, like adding title, axes, etc. Complete the graph by adding an appropriate X axis and Y axis.

Notice that Excel has used the data series name as the graph title. This is okay. You can change it if you want to. Once an element is on the graph, such as an axis title, you can adjust it by right-clicking on it.

You might be tempted to further modify the graph. Keep this in mind: **JUST BECAUSE YOU CAN, DOES NOT MEAN THAT YOU SHOULD**. Right click in the central portion of the graph. Select “Format Plot Area”. Select “Picture or Texture Fill”. Notice that someone at Microsoft had a fondness for fishes – one of the available textures is a pattern of fossil fish. Select this. While it looks cool for a few minutes, it actually makes it harder to interpret the graph, so sadly we will not keep it.

Change the plot area format back to plain old boring white.

You will typically want to print your graph and/or spreadsheet. Which of these gets printed depends on where the mouse was the last time you clicked it. For instance, if you were working in the graph, and the graph is “selected”, hitting Print causes only the graph to print. However, if you were working on the main spreadsheet, once you hit Print, it will print the whole spreadsheet (including the graph). Understand the distinction. I strongly suggest that you always “Preview” before printing to make sure of what you are getting and how it will line up on the paper.

It is also important to save your work on a Jump Drive. You will likely need it later and it will NOT be saved on the lab laptops (they are set to periodically erase everything that you type)..

Assignment:

Produce the graph as per the above instructions. Print the spreadsheet showing your numbers and the graph.

HINTS:

Did you just ignore the instructions and select the data and hit “chart” and up popped a chart? If you did, go directly to Jail and do not pass “Go”. Do not collect \$200. You will also not collect a good grade. READ THE INSTRUCTIONS. This is a simple graph. Later on you will be some much more elaborate things with Excel. If you don’t follow the instructions, you won’t have a chance of success in this course.

-- END