# **Plotting Data with Microsoft Excel**

Here is an example of an attempt to plot parametric data in a scientifically meaningful way, using Microsoft Excel. This example describes an experience using the Office X version for Macintosh. The details may change with different versions, but the principle stays the same – the author must control the appearance of all aspects of the figure. We cannot count on programmers to have correctly guessed our intended use of the application.

If you find that it is just too awkward to prepare a proper figure using a computer program, then do it by hand. If you use a hand drawn graph do keep in mind its purpose. If it is to be a working document, not to be presented or published, then you may want to make it as large as you can, ignoring margins, for maximum accuracy. For a paper or presentation, just keep the figure well within the margins of your sheet of graph paper, since the purpose is to illustrate the result, not to use the graph as a working document. You don't need to fill up all available space. If you submit a manuscript for publication or conduct a formal presentation, then you will probably have to prepare a computer generated figure.

#### **PROGRAM DEFAULT**

Table 1 represents measurements of heights of tree seedlings as they were allowed to grow. If you haven't much experience plotting with Excel you might want to open a new spreadsheet, enter the data, and follow along with the example. The data were entered into an Excel spreadsheet in columns in the same order as they appear in the table, with time (the independent variable) listed in the first column.

Table	1.	Time	course	of g	rowth	of	seedlings	of	three	com	mon
North	Am	erican	tree spe	ecies.	Meas	sure	ements are	e he	ight of	f seed	lling
in cent	time	eters.									

Time (days)	Acer palmatum	Quercus rubra	Morus alba
0	1	1.5	1
2	1.5	2	2.2
4	2.2	2.7	3.7
6	3.2	3.2	5.4
8	4.3	3.5	7
12	5.2	3.7	8.7
16	5.6	3.8	10.3
20	6.3	4.8	16

After selecting all of the data, including the headings, punch the plot icon and then just hit the NEXT buttons and then the finish buttons. Figure 1 shows what might come of it. Obviously, you must make some choices right from the beginning.

A lot of things are wrong with figure 1. The most grievous error is that the independent variable was not recognized as such. All data were plotted versus row number, which is the program default for column, bar, and line charts. The plot type is not appropriate either. Column type graphs are best used for independent variables in the

form of categories. To show a relationship between continuous variables such as height versus time, one would typically use a line graph or a scatter plot with trend lines.

If you want to see just how inappropriate (even bizarre) some plot types are for these kinds of data, try making a "Doughnut," "Radar," or "Surface" chart with the same data.



Fig. 1. First attempt at plotting a time course of seedling growth using Microsoft Excel.

## SELECT A GRAPH TYPE

Now try selecting a line graph, then go to the finish without choosing other options.



Fig. 2. Attempt at using a line graph.

Obviously, the phrase "line graph" means something different, to the Microsoft Excel programmers, than it does to most of the rest of us. When you choose line graph in Excel, the program <u>still treats the independent variable as a set of categories</u>. Time is a continuous variable, so the x axis should go from zero to twenty days. In fact, time itself is treated as a dependent variable.

### Try a scatter plot

Select the category "XY (Scatter)" and go with the defaults.



*Fig. 3.* Table 1 data as a scatter plot, default settings.

Aha! It seems that if you want the x axis to represent a continuous variable, you must select a scatter plot. Notice that this plot type recognizes the first column of data as the dependent variable.

Under Chart/Source data.../Series you can delete or add dependent variables and/or change the independent variable. The program lets you make the selection directly on the chart, by highlighting the desired column(s) of data. You might want to practice a bit.

#### ADD THE ESSENTIALS

You don't need to add much to your scatter plot to make it acceptable. The axes should be labeled, of course. There should also be some kind of trend line for each data set. Under Chart Options.../Titles you have the opportunity to enter axis labels. Remember that in a typical publication the figure caption goes beneath the graph. We don't give a graph a title, which Excel places above the figure.

For figure 4, smooth trend lines were added from the Chart Type... menu by selecting a different option for the type of scatter plot. You must let the data "tell you" what type of lines are appropriate. See the documents on error analysis and curvefitting and the summary, below.

(http://www.owlnet.rice.edu/~labgroup/pdf/Errorsandfits.htm or http://www.owlnet.rice.edu/~labgroup/pdf/errorsandfits.pdf



Fig. 4. Scatter plot with axis labels and smoothed trend lines.

## About curve fitting

The lines in the graph of figure 4 are what the program calls "smoothed lines." Another choice is to use straight lines connecting the data points, a method also called "interpolation." In either case, the lines are strictly there to help guide the eye. Such lines are not theoretical curve fits. True curve fitting is based upon mathematical methods such as the method of least squares, and frequently requires the use of replicate samples, means, and standard errors.



Fig. 5. Scatter plots with inappropriate trend lines.

A true curve fit often does not intersect each and every data point. Figure 5 shows examples of inappropriate choices of trend lines. The data are obviously scattered due to experimental error, but the student used interpolation anyway. In figure 6 the same data are visualized using a user-defined trend line.



Fig. 6. Scattered data visualized using a user-defined trend line

See our document on error representation and curvefitting for introductory material on experimental error, its representation on graphs, and for basic material on curve fitting.

#### **REMOVE NON-ESSENTIALS**

Now that the basic plot is ready (back to figure 4), let's make it acceptable for publication by getting rid of the "computer junk," and shaping it up.

Under Chart Options... the gridlines tab allows you to extend the axis tic marks across the plot area, a feature that is completely non-essential to the effectiveness of this figure. Do note, however, that some publications allow tic marks to enter the graph space, and a grid may be positively useful if you intend to read numbers from the graph.

Eliminate the gridlines from the major y axis units. You don't need the gray fill or the borders either. Double click in the graph area and you can choose "none" for border and area. Double click the border around the entire figure, and you can choose "none" for the larger border around the whole thing. The result should resemble the graph in figure 7.



Fig. 7. Gridlines and borders removed.

For a presentation, you may wish to include a legend in the plot area, extended tic marks, color, borders, and perhaps other "bells and whistles." What you choose to do depends on the purpose of the figure. Just make sure that it you are the one making the decisions.

Background shading reduces contrast, borders serve no useful purpose, color is seldom used for a published graph, and conventions often require the legend to be included in the figure caption, not in the plot area itself. Many journals ignore former conventions and permit all sorts of trend lines and legends, as long as they are useful, appropriate, and understandable.

In print journals we seldom use color. It adds to publication costs, and consider that there are color blind people. Getting rid of color can be a little trickier. Select one of the data points to highlight the whole set. Go to Format/Selected Data Series... and make everything black, both for lines and data symbols. Play with symbol size and style, trying different options until you see what you want. An editor may require specific plot symbols, by the way, further stifling your creative independence.



*Fig.* 8. More formatting changes - larger symbols, different line styles, all black and white.

Oops. So your figure isn't so colorful anymore. So what? No one will be upset if you figure has all of the essentials and none of the non-essentials. Now we'll do a couple of more things, and we're ready.

#### **PROPORTION AND ORIENTATION**

Note that the plot area of figure 8 is a bit wider than it is tall. What is most important - the time axis, the height axis, or the relationship between the two? To show the relationship most effectively, square up the plot area by selecting the graph, grabbing an anchor point with the mouse, and changing the axis proportions.

If you were plotting a single function, then creating a square plot area, and having at least one data point near the end of each axis should display the slope (in this case rate of growth) most effectively. To best show the differences among the three rates of growth, you might adjust proportions so that the middle set (data for *Acer palmatum*) produces a slope near forty-five degrees. The y axis would have to be taller than the x axis is long. Such adjustments are a matter of personal preference and perhaps trial and error to see what looks best.

The y axis label is easier to read in a horizontal orientation. You can double click on the label, select the alignment tab in the dialog box, and change the orientation to horizontal. You may have to re-do the proportions. The legend might work better placed in the field itself (area within the ranges of the axes. You can simply move it to a good place.

To produce figure 9, make the recommended changes to figure 8 then select the chart, go to Format/Selected Chart Area..., select the Font tab, and make all of your titles/labels the same font and size. Changing proportions and copying/pasting apparently affects type proportions and everything. By the way, the choice of font style is yours to make, but fancy, elaborate styles can be hard to read and distracting. We went with Times, 18 point in the Excel document before pasting it in.

Add an informative caption, and you're set.



*Fig. 9.* Differential growth rates of seedlings from three commonly planted landscape trees.

Your simple black-and-white figure isn't as dramatic as a full color version with all of the "eye candy," but this way the reader is not distracted from the message that seedlings of *Morus alba* grew much faster than those of the other two species. Figure 9 shows the graph as it might appear in a journal. Note that by convention a figure caption is placed below the corresponding figure. A journal editor might put the caption to the left, but rarely to the right or above. It's easier for a reader to find the information if he/she knows where to look, especially in a complicated results section.

Compare figure 9 with the first figure you generated (figure 1). Quite a difference, isn't there? The lesson here is, decide what <u>you</u> want your figure to look like. Don't let some anonymous programmer do it for you.