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 98 version. The details may change with newer 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.

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>
<thead>
<tr>
<th>Time (days)</th>
<th>Acer palmatum</th>
<th>Quercus rubra</th>
<th>Morus alba</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1.5</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1.5</td>
<td>2</td>
<td>2.2</td>
</tr>
<tr>
<td>4</td>
<td>2.2</td>
<td>2.7</td>
<td>3.7</td>
</tr>
<tr>
<td>6</td>
<td>3.2</td>
<td>3.2</td>
<td>5.4</td>
</tr>
<tr>
<td>8</td>
<td>4.3</td>
<td>3.5</td>
<td>7</td>
</tr>
<tr>
<td>10</td>
<td>5.2</td>
<td>3.7</td>
<td>8.7</td>
</tr>
<tr>
<td>12</td>
<td>5.6</td>
<td>3.8</td>
<td>10.3</td>
</tr>
</tbody>
</table>

After selecting all of the data, including the headings, punch the plot icon and let the program plot the data, using the built in defaults. Figure 1 shows what might come of it.

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 such as height versus time, one would typically use a line graph.
If you want to see just how inappropriate (and bizarre) some plot types are for these kinds of data, try making a "Doughnut," "Radar," or "Surface" chart with the same data.

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

**MAJOR CORRECTIONS**

Under Chart/Source Data select the series tab and delete Time (days) from the list of dependent variables. Before closing the box by selecting OK, click in the area called Category (X) axis labels, select the time data on the spreadsheet (excluding the heading). Figure 2 shows a new graph with time as the independent variable.

Figure 2. Figure 1 modified so that time is the independent variable.
Since a line graph is a more appropriate way of presenting the data, try creating figure 3 from figure 2 by choosing Chart/Chart Type... and selecting a line graph. If you turned it a paper with figure 3 you would receive partial credit (not much, but some at least).

If you wish, you might go back to the original data - select all, punch the plot icon, and select XY (Scatter) as the plot type, choosing a subtype with data symbols plus lines. The result should look a lot like that of figure 3. For scientific data involving continuous relationships, a scatterplot produces an appropriate graph most often and with less fine tuning than the other choices.

**ABOUT LINE GRAPHS AND CURVE FITTING**

The lines in the graph of figure 3 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. A true curve fit often does not intersect each and every data point.

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.

**ADD ESSENTIALS, REMOVE NON-ESSENTIALS**

Now that the basic plot is ready, let's make it acceptable for publication. We'll start by labeling the axes.
For publication, axes must be labeled. Under Chart/Chart Options... a large dialog box pops up. The Titles tab lets you type in labels for each axis. The figure title typically goes into a caption that is placed beneath the graph. Captions are often submitted separately from the figures, so we won't worry about a figure title right now.

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.

Label both axes and eliminate the gridlines from the major y axis units. The result should resemble the graph in figure 4.

![Figure 4](image)

Figure 4. Figure 3 with gridlines removed, titles added to the axes, and font style changed.

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.

Create figure 5 by double clicking in the plot area and selecting "none" under Area, in the dialog box for formatting plot area. While there, select "none" under Border. Double click on the border of the entire chart, and get rid of the patterns under "Format chart area" as well. Under Chart/Chart Options... select the Legends tab and deselect "show legend." We're on a roll - don't stop there...

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 and white, 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.
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.

**PROPORTION AND ORIENTATION**

Two more changes to make, and the figure should be acceptable to the most finicky of publishers. Note that the plot area of figure 5 is 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. Double click on the label, select the alignment tab in the dialog box, and change the orientation to horizontal.

To produce figure 6, make the recommended changes to figure 5 then select the chart, go to Format/Selected Chart Area..., select the Font tab, and make all of your titles/labels 12 point, Times, regular type. Changing proportions 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.
Figure 6. Result of adjustments to plot area proportions. Note that the axis scales were not altered, just axis proportions.

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. Finally, figure 7 shows the graph as it might appear in a journal.

Figure 1. Differential growth rates of seedlings from three commonly planted landscape trees.

- *Acer palmatum*
- *Morus alba*
- *Quercus rubra*

Figure 7. Publishable figure with complete title and legend. In addition to providing a complete legend, the species names were italicized, as per convention.