

Almost Everything You Wanted to Know About Making Tables and Figures

Once your statistical analyses are complete, you will need to summarize the data and results for presentation to your readers. Data summaries may take one of 3 forms: text, Tables and Figures.

Text: contrary to what you may have heard, not all analyses or results warrant a Table or Figure. Some simple results are best stated in a single sentence, with data summarized parenthetically:

Seed production was higher for plants in the full-sun treatment (52.3 +/-6.8 seeds) than for those receiving filtered light (14.7+/- 3.2 seeds, $t=11.8$, $df=55$, $p<0.001$.)

Tables: Tables present lists of numbers or text in columns, each column having a title or label. Do not use a table when you wish to show a trend or a pattern of relationship between sets of values - these are better presented in a Figure. For instance, if you needed to present population sizes and sex ratios for your study organism at a series of sites, and you planned to focus on the differences among individual sites according to (say) habitat type, you would use a table. However, if you wanted to show us that sex ratio was related to population size, you would use a Figure.

Figures: Figures are visual presentations of results, including graphs, diagrams, photos, drawings, schematics, maps, etc. Graphs are the most common type of figure and will be discussed in detail; examples of other types of figures are included at the end of this section. Graphs show trends or patterns of relationship.

Organizing your presentation: Once you have done your analyses and decided how best to present each one, think about how you will arrange them. Your analyses should tell a "story" which leads the reader through the steps needed to logically answer the question(s) you posed in your Introduction. The order in which you present your results can be as important in convincing your readers as what you actually say in the text.

How to refer to Tables and Figures from the text: *Every* Figure and Table included in the paper MUST be referred to from the text. Use sentences that draw the reader's attention to the relationship or trend you wish to highlight, referring to the appropriate Figure or Table only parenthetically:

Germination rates were significantly higher after 24 h in running water than in controls (Fig. 4).

DNA sequence homologies for the *purple* gene from the four congeners (Table 1) show high similarity, differing by at most 4 base pairs.

Avoid sentences that give no information other than directing the reader to the Figure or Table:

Table 1 shows the summary results for male and female heights at Bates College.

Abbreviation of the word "Figure": When referring to a Figure in the text, the word "Figure" is abbreviated as "Fig.", while "Table" is not abbreviated. Both words are spelled out completely in descriptive legends.

How to number Tables and Figures: Figures and Tables are numbered independently, in the sequence in which you refer to them in the text, starting with Figure 1 and Table 1.

Placement of Figures and Tables within the Paper: In consideration of your readers, place each Table or Figure as near as possible to the place where you first refer to it (e.g., the next page.) For manuscripts (e.g. lab papers), Tables and Figures are usually put on separate pages from text material.

The "Acid Test" for Tables and Figures: Any Table or Figure you present must be sufficiently clear, well-labeled, and described by its legend to be understood by your intended audience without reading the results section, i.e., it must be able to stand alone and be interpretable. Overly complicated Figures or Tables may be difficult to understand in or out of context, so strive for simplicity whenever possible. If you are unsure whether your tables or figures meet these criteria, give them to a fellow biology major (not in your course) and ask them to interpret your results.

Descriptive Legends or Captions: To pass the "acid test" above, a clear and complete legend (sometimes called a caption) is essential. Like the title of the paper itself, each legend should convey as much information as possible about what the Table or Figure tells the reader: the subjects of the experiment, the treatment applied or the relationship displayed, location (if a field experiment), and sample sizes and statistical tests if they are not displayed elsewhere. Do not simply restate the axis labels with a "versus" written in between.

Example:

Figure 1. Height frequency (%) of White Pines (*Pinus strobus*) at Thorncrag Bird Sanctuary, Lewiston, Maine, before and after the Ice Storm of '98. Before, n=137, after, n=133. Approximately 16% of canopy pines were topped by heavy ice loads, averaging 3.6 m stem length decrease. Four trees fell during the storm and were excluded from the post-storm survey.

In the examples later in this section, note the completeness of the legends. When you are starting out, you can use one of these examples (or an appropriate example from a published paper) as a model to follow in constructing your own legends.

Where do you place the legend?

- **Table legends** go above the body of the Table and are left justified; Tables are read from the top down.
- **Figure legends** go below the graph; graphs and other types of Figures are usually read from the bottom up.

The Anatomy of a Table

Table 4 below shows the typical layout of a table in three sections demarcated by lines. Tables are most easily constructed using your word processor's table function or a spread sheet such as Excel. Gridlines or boxes, commonly invoked by word processors, are optional for our purposes, but unlikely to be permitted in a journal.

Table 4. Population variation in hatch success (mean percent) of unfertilized eggs for females from populations sampled in 1997. N = number of females tested.

Population	mean (%)	Standard deviation	Range	N
Beaver Creek ^T	7.31	13.95	0-53.16	15
Honey Creek ^T	4.33	7.83	0-25.47	11
Rock Bridge Gans Creek ^T	5.66	13.93	0-77.86	38
Cedar Creek ^P	6.56	9.64	0-46.52	64
Grindstone Creek ^P	8.56	14.77	0-57.32	19
Jacks Fork River ^P	5.28	8.28	0-30.96	28
Meramec River ^P	5.49	10.25	0-45.76	45
Little Dixie Lake ^L	7.96	14.54	0-67.66	71
Little Prairie Lake ^L	6.86	7.84	0-32.40	36
Rocky Forks Lake ^L	3.31	4.12	0-16.14	43
Winegar Lake ^L	10.73	17.58	0-41.64	5
Whetstone Lake ^L	7.36	12.93	0-63.38	57

^T = temporary stream, ^P = permanent streams, ^L = lakes. **---footnotes**

---Table legend

---Column titles

---Table body (data)

---Lines demarcating the different parts of the table

Table 2. Log-likelihood tests of deviation from 1:1 sex ratios for nymphs collected from each population in 1997 and 1998. Values are ratios of female:male; sample sizes are in parentheses. Bonferroni corrected probabilities are shown with an asterisks.

Population	Year	
	1997	1998
Beaver Creek ^L	9.00:1(20)***	2.67:1 (22)*
Honey Creek ^T	9.00:1(56)***	2.27:1 (98)***
Rock Bridge ^T	3.33:1(26)**	2.09:1 (68)**
Cedar Creek ^P	2.05:1(119)***	1.87:1 (198)***
Grindstone Creek ^P	-	2.26:1 (140)***
Jacks Fork River ^P	2.89:1(35)**	5.17:1 (37)***
Meramec River ^P	2.80 1(38)**	2.41:1 (58)**
Little Dixie Lake ^L	2.45:1(494)***	2.46:1 (384)***
Little Prairie Lake ^L	2.38:1 (71)***	2.08:1 (157)***
Rocky Forks Lake ^L	2.55:1 (213)***	2.93:1 (299)***
Winegar Lake ^L	3.41:1 (207)***	2.34:1 (204)***
Whetstone Lake ^L	2.69:1 (381)***	2.01:1 (268)***

* significant at $p < 0.05$; ** significant at $p < 0.005$; *** significant at $p < 0.001$.

^T = temporary stream, ^P = permanent streams, ^L = lakes.

Table 2. Planting date, mean planting density, and total number of seed clams planted in plots at Filucy Bay and Wescott Bay in 1979.

Location	Plot code	Planting date	Mean planting density in no. clams/m ² ± 1 st. dev. (N)	Total no. clams planted
Filucy Bay	F10 x 30	5-16-79	994 ± 39(5)	298200
	F3 x 10	5-24-79	994 ± 39(5)	29820
Wescott Bay	W10 x 25	5-16-79	994 ± 39(5)	248500
	W3 x 10	6-2-79	895 ± 35(5) ^a	26850

^aCalculated after clams were planted based on estimated 11% mortality of seed clams between 5-24 and 6-2-79.

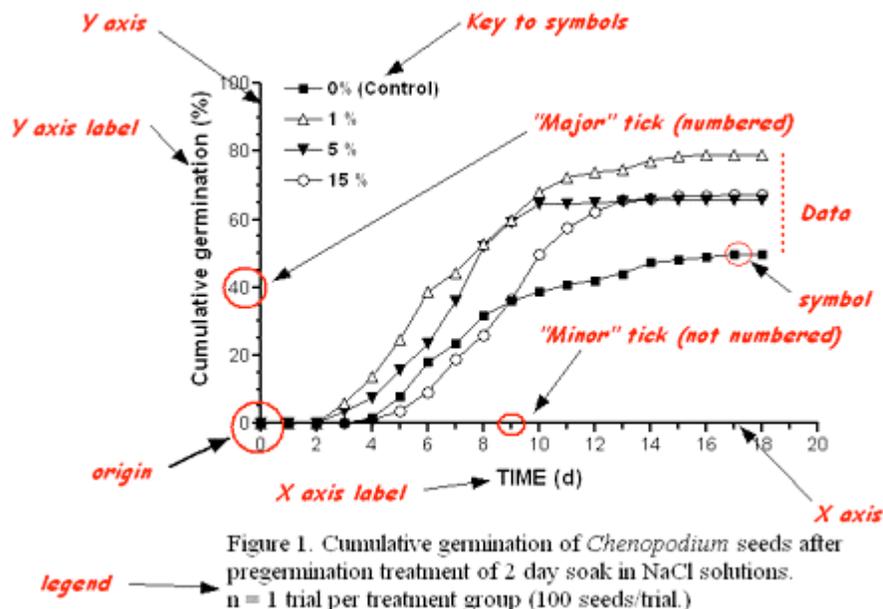
In these examples notice several things:

- the presence of a period *after* "Table #";
- the legend goes *above* the Table;
- *units* are specified in column headings wherever appropriate;
- lines of demarcation are used to set legend, headers, data, and footnotes apart from one another.
- *footnotes* are used to clarify points in the table, or to convey repetitive information about entries;
- footnotes may also be used to denote statistical differences among groups.

The Anatomy of a Figure

The sections below show when and how to use the four most common Figure types (bar graph, frequency histogram, XY scatterplot, XY line graph.) The final section gives examples of other, less common, types of Figures.

Parts of a Graph: Below are example figures (typical line and bar graphs) with the various component parts labeled in red. Refer back to these examples if you encounter an unfamiliar term as you read the following sections.



Some general considerations about Figures:

- **Big or little?** For course-related papers, a good rule of thumb is to size your figures to fill about one-half of a page. Readers should not have to reach for a magnifying glass to make out the details.
- **Color or no color?** Most often black and white is preferred. The rationale is that if you need to photocopy or fax your paper, any information conveyed by colors will be lost to the reader. However, for a poster presentation or a talk with projected images, color can be helpful in distinguishing different data sets. Every aspect of your Figure should convey information; *never use color simply because it is pretty.*
- **Title or no title?** *Never use a title for Figures included in a paper;* the legend conveys all the necessary information and the title just takes up extra space. However, *for posters or projected images,* where people may have a harder time reading the small print of a legend, a larger font title is very helpful.
- **Offset axes or not?** Elect to offset the axes only when data points will be obscured by being printed over the Y axis.
- **Error bars or not?** Always include error bars (SD or SEM) when plotting means. In some courses you may be asked to plot other measures associated with the mean, such as confidence intervals.

Four Common Figure Types

Bar Graph

Bar graphs are used when you wish to compare the value of a single variable (usually a summary value such as a mean) among several groups. For example, a bar graph is appropriate to show the mean sizes of plants harvested from plots that received 4 different fertilizer treatments. (Note that although a bar graph might be used to show differences between only 2 groups, especially for class purposes, editors of many journals would prefer that you save space by presenting such information in the text.)

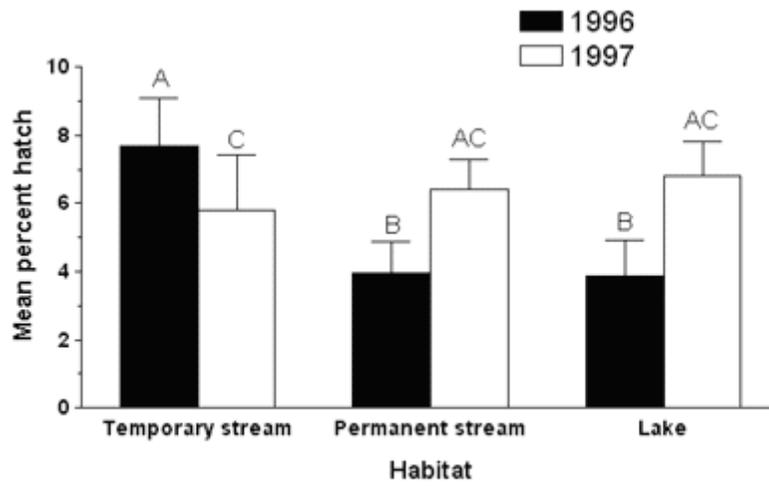


Figure 3. Effects of habitat and year on tytoparthenogenetic capacity (mean % hatching success \pm 1 SD of unfertilized eggs) in mayflies. Means with different letters are significantly different (Tukey's HSD, $p < 0.05$).

In this example notice that:

- legend goes *below* the figure;
- a period follows "Figure 1" and the legend itself; "Figure" is not abbreviated ;
- the *measured* variable is labeled on the Y axis. In most cases units are given here as well (see next example);
- the *categorical* variable (habitat) is labeled on the X axis, and each category is designated;
- a *second* categorical variable (year) within habitat has been designated by *different bar fill patterns*. The patterns *must* be defined in a *key*, located wherever there is a convenient space within the graph.
- error bars are included, extending +1 SD or SEM above the mean.
- statistical differences may be indicated by a system of letters above the bars, with an accompanying note in the caption indicating the test and the significance level used.

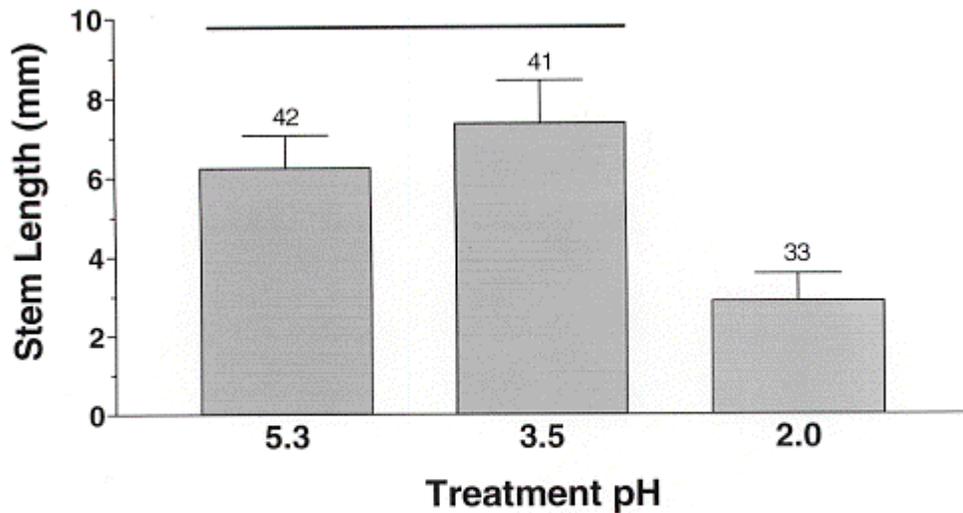


Figure 1. Mean stem length (± 1 SD) of seedling clover watered to soil saturation daily for 2.5 weeks with simulated acid rain of varying pH. The control (pH 5.3) was normal city tapwater. The pH 3.5 and 2.0 water was acidified with 2 M sulfuric/ 1 M nitric acid solution. Line over bars indicates groups which were not significantly different (Kruskal-Wallis Test and Dunn's Multiple Comparison's Tests). Number over bar indicates sample size.

Notice here:

- the completeness of the legend, which in this case requires over 3 lines just to describe the treatments used and variable measured.
- axis labels, with units;
- treatment group (pH) levels specified on X axis;
- error bars and group sample sizes accompany each bar, and each of these is well-defined in legend;
- statistical differences in this case are indicated by lines drawn over the bars, and the statistical test and significance level are identified in the legend.

Frequency Histogram

Frequency histograms (also called frequency distributions) are bar-type graphs that show how the measured individuals are distributed along an axis of the measured variable. Frequency (the Y axis) can be *absolute* (i.e. number of counts) or *relative* (i.e. percent or proportion of the sample.) A familiar example would be a histogram of exam scores, showing the number of students who achieved each possible score. Frequency histograms are important in describing populations, e.g. size and age distributions.

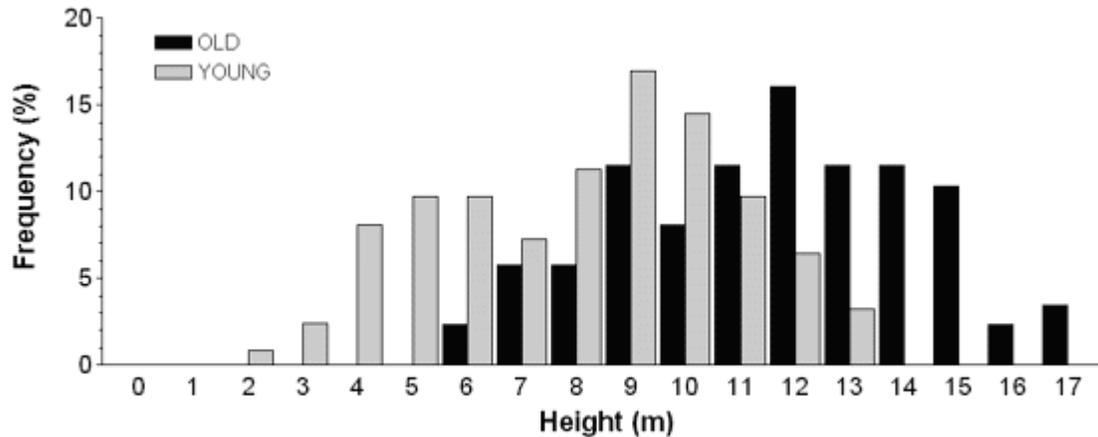


Figure 4. Height distribution in two recruitment cohorts of eastern white pine (*Pinus strobus*) near the eastern shore of Lake Auburn, Maine, in January 2001. N = 88 OLD and 123 YOUNG trees.

Notice several things about this example:

- the Y axis includes a clear indication ("%") that relative frequencies are used. (Some examples of an absolute frequencies: "Number of stems", "Number of birds observed")
- the measured variable (X axis) has been divided into categories ("bins") of appropriate width to visualize the population distribution. In this case, bins of 0.2 cm broke the population into 7 columns of varying heights. Setting the bin size at 0.5 cm would have yielded only 3 columns, not enough to visualize a pattern. Conversely, setting the bin size too small (0.05 cm) would have yielded very short columns scattered along a long axis, again obscuring the pattern.
- the values labeled on the X axis are the bin *centers*;
- sample size is clearly indicated, either in the legend or (in this case) the graph itself;
- the Y axis includes numbered and minor ticks to allow easy determination of bar values.

X,Y Scatterplot

These are plots of X,Y coordinates showing each individual's or sample's score on *two* variables. When plotting data this way we are usually interested in knowing whether the two variables show a "relationship", i.e. do they change in value together in a consistent way?

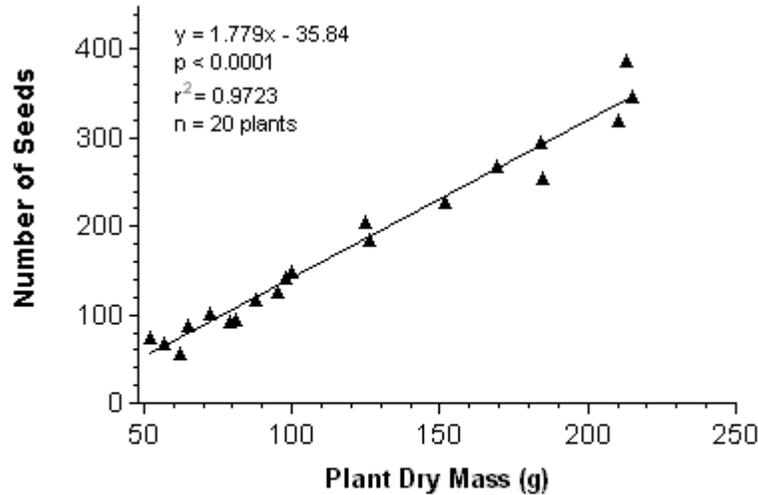


Figure 3. Seed production as a function of plant biomass in waterlilies (*Nuphar luteum*) harvested from Great Works Pond in Northern Maine in August, 2001.

Note in this example that:

- each axis is labeled (including units where appropriate) and includes numbered and minor ticks to allow easy determination of the values of plotted points;
- sample size is included in the legend or the body of the graph;
- if the data have been analyzed statistically and a relationship between the variables exists, it may be indicated by plotting the regression line on the graph, and by giving the equation of the regression and its statistical significance in the legend or body of the figure;
- the range of each axis has been carefully selected to maximize the spread of the points and to minimize wasted blank space where no points fall. For instance, the X axis is truncated below 50 g because no plants smaller than 52 g were measured. The ranges selected also result in labeled ticks that are easy to read (50, 100, 150..., rather than 48, 96, 144...)

Which variable goes on the X axis? When one variable is clearly *dependent* upon another (e.g. height depends on age, but it is hard to imagine age depending on height), the convention is to plot the **dependent variable on the Y axis and the independent variable on the X axis**. Sometimes there is no clear independent variable (e.g. length vs. width of leaves: does width depend on length, or vice-versa?) In these cases it makes no difference which variable is on which axis; the variables are inter-dependent, and an X,Y plot of these shows the *relationship* BETWEEN them (rather than the effect of one upon the other.)

In the example plotted above, we can imagine that seed production *might* depend on plant biomass, but it is hard to see how biomass could depend directly on seed production, so we choose biomass as the X axis. Alternatively, the relationship might be indirect: *both* seed production *and* plant biomass might depend on some other, unmeasured variable. Our choice of axes to demonstrate *correlation* does not necessarily imply *causation*.

X,Y Line Graph

Line graphs plot a series of related values that depict a change in Y as a function of X. Two common examples are a growth curve for an individual or population over time, and a dose-response curve showing effects of increasing doses of a drug or treatment.

When to connect the dots? If each point in the series is obtained from the same source and is dependent on the previous values (e.g. a plot of a baby's weight over the course of a year, or of muscle strength on successive contractions as a muscle fatigues), then the points should be connected by a line in a dot-to-dot fashion. If, however, the series represents independent measurements of a variable to show a trend (e.g. mean price of computer memory over time; a standard curve of optical density vs. solute concentration), then the trend or relationship can be modeled by calculating the best-fit line or curve by regression analysis. *Do not connect the dots* when the measurements were made independently.

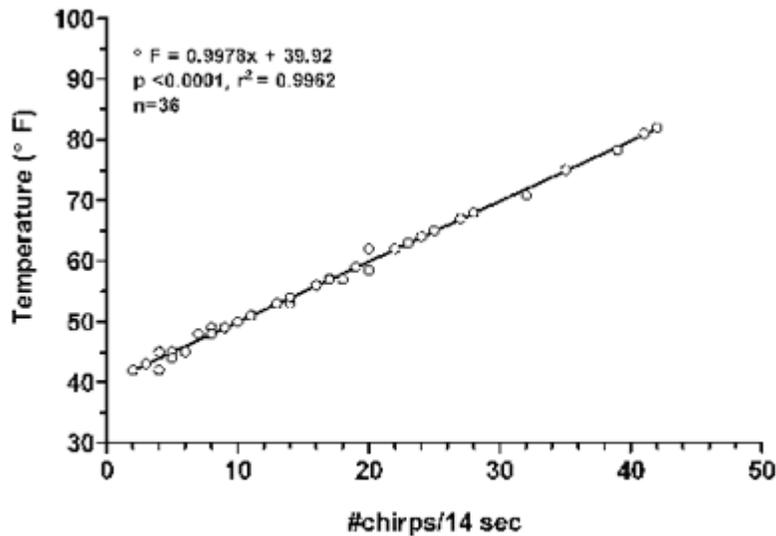


Figure 7. Temperature-dependence of cricket chirp frequency in south central Maine. Temperature (F) = #chirps in 14 sec + 40. n = 36 cricket chirp bouts.

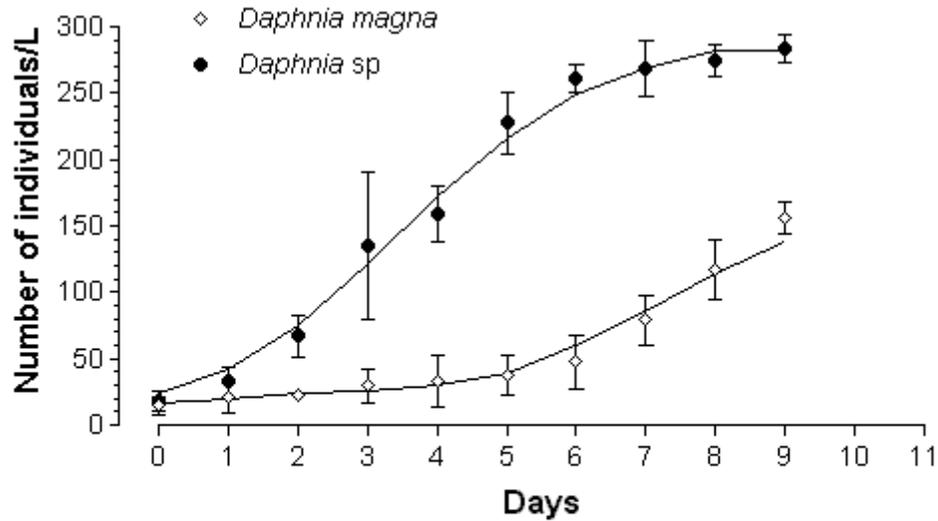


Figure 2. Mean population density (± 1 standard deviation) of two species of *Daphnia* following artificial eutrophication of a small farm pond by application of organic fertilizer. Six replicate 1 L water samples were drawn from 50 cm depth at 1100 hr each day.

In this example notice:

- a different symbol is used for each group (species), and the key to the symbols is placed in the body of the graph where space permits. Symbols are large enough to be easily recognizable in the final graph size;
- each point represents a mean value, and this is stated in the legend. Error bars are therefore plotted for each point and defined in the legend as well.
- because measurements were taken on independent groups for each species, the points are NOT connected dot-to-dot; instead a curve is fitted to the data to show the trend.

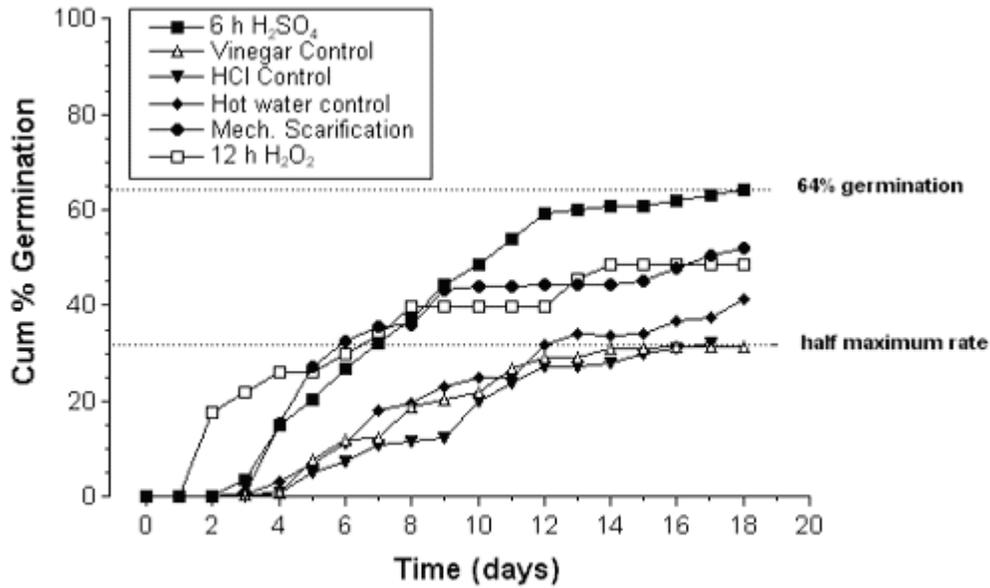


Figure 2. Cumulative germination of gourd seeds following various pregermination treatments. n = 100 seeds per trial.

Notice here that:

- this time the dots ARE connected dot-to-dot within each treatment, because cumulative percent germination was measured within the same set of seeds each day, and thus is *dependent* on the measurements of the prior days;
- a different symbol is used for each treatment, and symbols are large enough (and connecting lines fine enough) so that all can be easily read at the final graph size;
- in addition to the key to symbols, two other kinds of helpful information are supplied in the body of the figure: the values of the highest and lowest final cumulative percents, and a dashed line (baseline) showing the lowest cumulative % germination achieved. This baseline is defined in the legend.

Modified 1-9-03 [gja](#)

Copyright 2002 [Department of Biology, Bates College](#), Lewiston, ME 04240