

## **DATA REPRESENTATION AND GRAPHICAL TECHNIQUES**

### **Introduction**

Data obtained by biologists from experiments, measurements or observations are eventually presented in the results section (sometimes in the discussion) of a scientific report or manuscript. An important aspect of a results section is the representation of the data. In many cases the investigator will have gathered the results at great financial and personal (in terms of man hours) cost. The scientist must present, or represent, these results in the best possible way - clearly and unambiguously. This ensures that they are perfectly understandable to other workers who will eventually read the report. Data presentation is therefore a valuable skill that must be acquired.

Scientific data (results) can be presented in the form of tables or as graphs. It is the graphical representation of data which is the focus of this section. Good and appropriate graphs allow rapid comparison of the effects of different treatments, the affect of parameters on different species, or assessment of trends, be they temporal or relationships between variables.

There are numerous ways of representing data graphically, some of which are more appropriate to certain types of data than others. Selecting the appropriate type of graph that best shows the important features of your data can sometimes be confusing. There are also several components that go into constructing a graph, and care when considering these can go a long way towards making the graph easy for readers (and yourself) to interpret. Almost as important as the graph itself is the caption, or legend, describing what the graph represents.

### **Visual Presentation of Results**

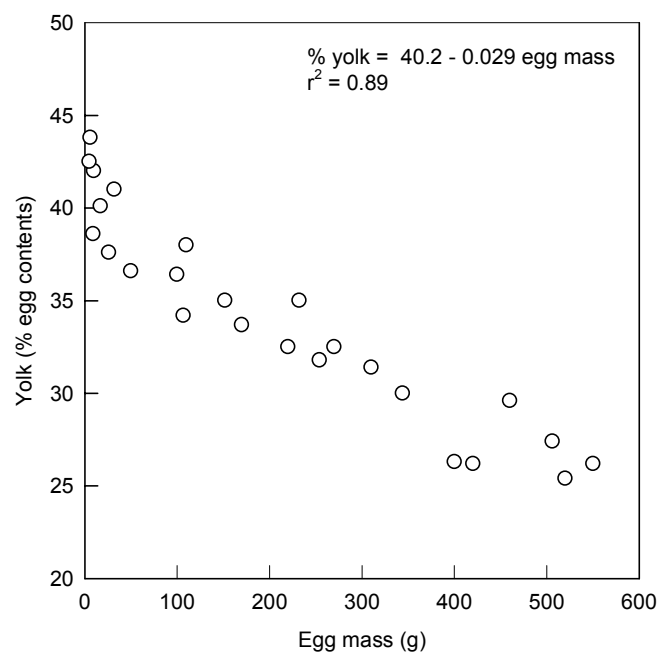
As mentioned above, results may be presented in two fundamental ways:

1. Tables
2. Graphs

There are many cases where data can be presented either way, and whether to use a table or a graph is not always obvious. The advantage of a table is that it allows other readers to see precise numbers that may be difficult to determine from a graph (e.g.  $22.367 \pm 3.678$ ). Thus,

if precision is important for complete understanding of the findings, then the data are best presented as tables. A list of species on which various measurements have been taken (e.g. length, width, height etc.) is also often best represented in a table.

The advantage of a graph is that it has more visual impact which makes it easier to see trends and patterns in some types of data. The data represented as a scatter plot in Figure 1 immediately reveals that there is a relationship between yolk content and egg mass. If the data were presented in a table, this relationship would be much harder to see.



**Figure 1. Yolk content (% of egg content) in relation to egg mass for 32 species of Procellariiformes**

There are a number of ways of plotting data as graphs and some data can be represented by more than one graph type. Often, however, a specific type of graph is more appropriate to certain types of data. The first thing to be decided, therefore, is the type of graph to use.

### **Types of Graph - Which One To Use?**

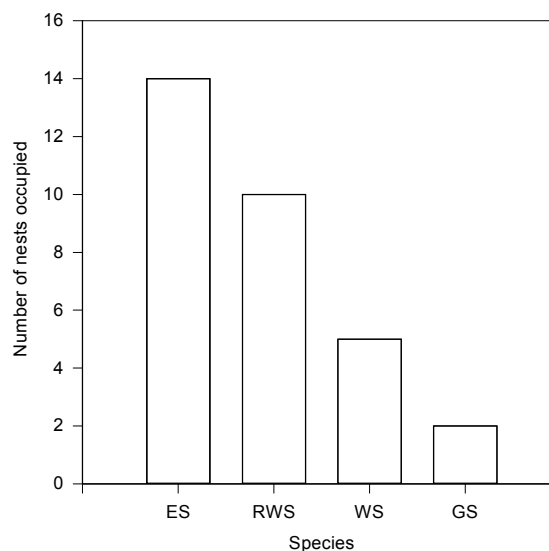
There are five basic types of 2-dimensional graphs. Note that on occasions data must be plotted as 3-dimensional graphs, but these are rather specialised and the majority of scientific reports will not require them. This section concentrates on fundamental 2-D plots.

The five types are:-

3. Bar graphs or histograms (vertical or horizontal).
4. Point graphs (with or without lines joining the points).
5. Scatter plots (which may have a line showing a trend).
6. Pie diagrams.
7. Kite diagrams.

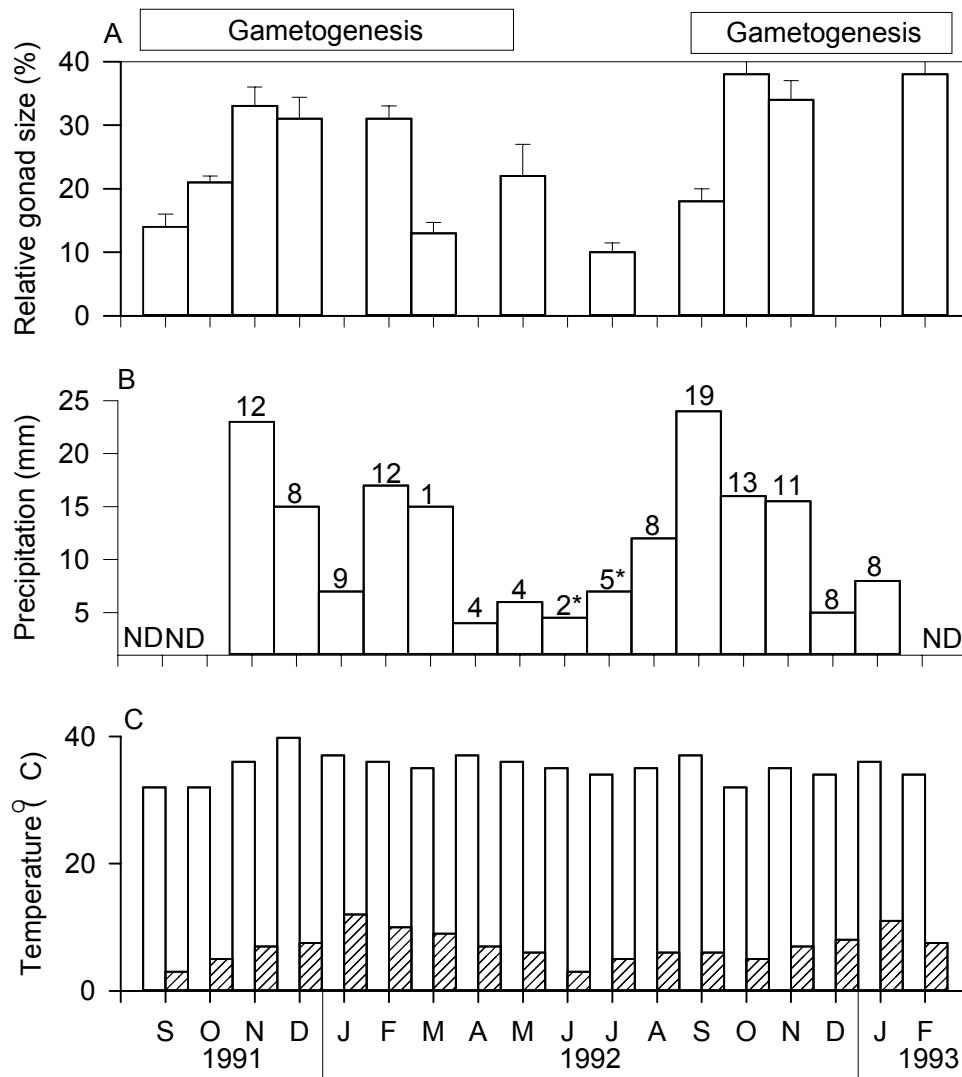
### *Bar Graphs and Histograms*

Bar graphs or histograms are useful for data involving frequencies. For example: numbers (or percentage) of animals in different size categories; number (or percentage) of animals per month or per site; rainfall over specific time periods. The category (size, number, month) is plotted on the horizontal or X-axis, and the number (number of observations or percentage) in each category is plotted on the vertical or Y-axis. The difference between a bar graph and a histogram is that bar graphs are used when dealing with data gathered from discrete variables - the bars on the graph do not touch each other. Histograms are used to represent continuous data and the bars should be continuous (compare Figure 2 with Figure 3 upper and middle histograms), although this convention is often not enforced anymore.



**Figure 2. Number of nest boxes occupied by European starlings (ES), Redwinged starlings (RWS), Wattled starlings (WS) and Glossy starlings (GS).**

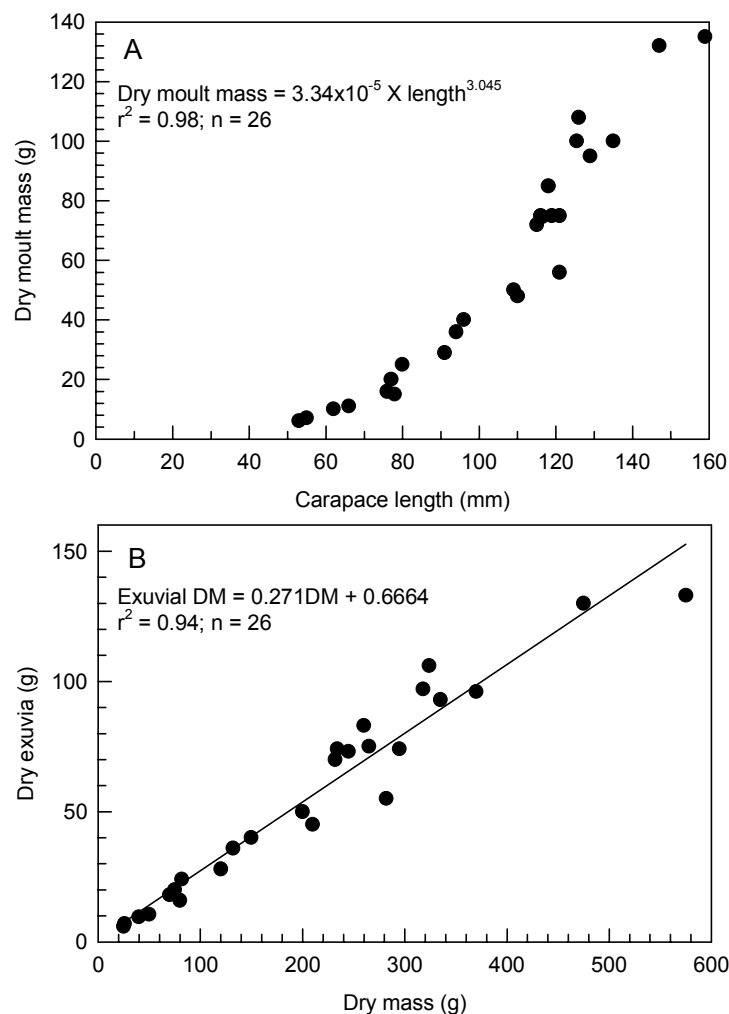
Bar graphs and histograms are not restricted to data involving total numbers or percentages, but means and variability in the form of standard deviations or standard error can also be represented in this way (Figure 3 upper histogram).



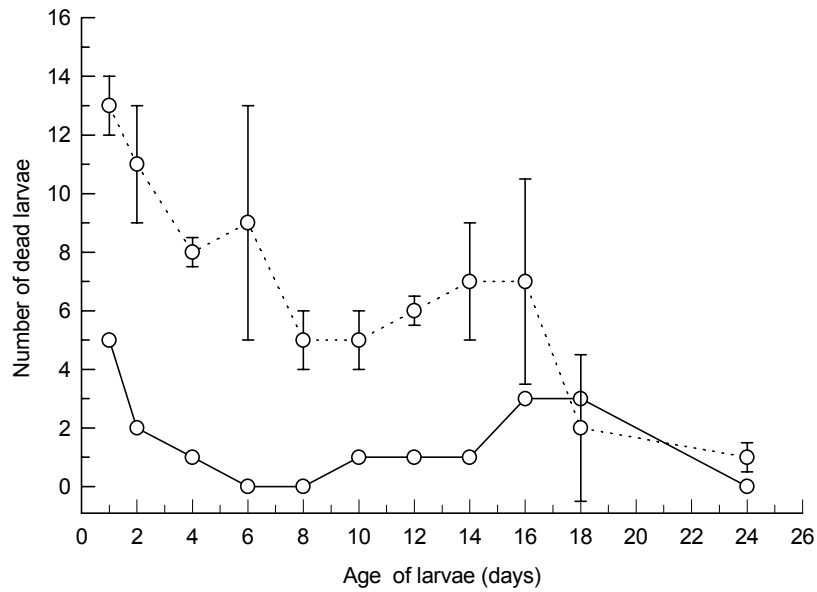
**Figure 3.** A, monthly change in size of the ovotestis of *Xerocerastis minutus*, expressed as percentage of the cross-sectional area of the visceral hump ( $\bar{x} \pm S.D.$ ). Horizontal blocks represent months in which gametogenesis was occurring. B, total monthly precipitation from fog at the Hamilton range weather station; numbers above histograms are total number of fog days. Asterisk indicates incomplete months, data from which may therefore be an underestimate. ND = no data. C, maximum and minimum (stippled) air temperatures recorded in each month.

### Line or Point Graphs

Line graphs are often used when plotting data of two variables e.g. oxygen consumption at different temperatures; blood osmolarity in relation to urine osmolarity. Data can be plotted as individual points (Figure 4) or as means with the variability (Figure 5). If the data are continuous, then the means can be joined by lines (as shown in Figure 5). When all data points are plotted as a scattergram (e.g. length width relationship) a line, based on the appropriate statistics (e.g. a regression or correlation) can then be plotted through these points to describe the relationship between the variables mathematically (Figure 4). This line should never be fitted by eye as you would be simply guessing.



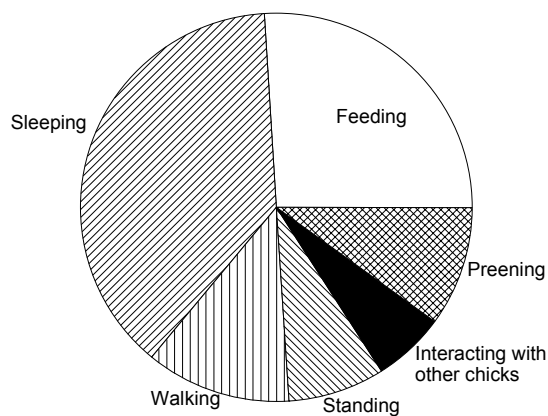
**Figure 4.** Exuvial mass of different sized *Jasus lalandii* (A = carapace length; B = dry mass).



**Figure 5. Comparison of the number of dead ostracods ( $\bar{x} \pm \text{S.D.}$ ) during a 26 day experiment in which the crustaceans were either infected or uninfected by trematode parasites.**

### *Pie Diagrams*

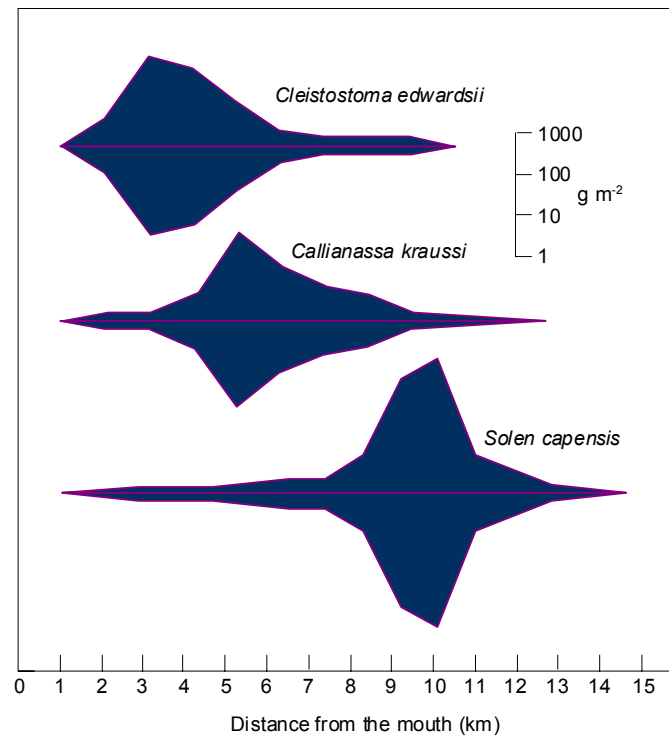
Pie graphs are often used to show ratios or percentages and an example of a pie diagram is presented below.



**Figure 6. Time-activity budget of ostrich chicks. Data are expressed as percentages over 24h.**

## Kite Diagrams

Kite diagrams are sometimes used in ecology to show density and distribution of species along a transect line or other environmental gradient.



**Figure 7. Biomass and distribution of dominant macrobenthic invertebrates in the Kariaga river estuary.**

There are other types of graph, but the above, particularly the histogram and point graph, are the most frequently used.

## How Many Graphs? - Comparative Graphs

When there is only one data set a single graph is all that can be used. Often several different data sets will have been collected or separate experiments carried out. For example, data might have been collected monthly on temperature, rainfall and reproductive condition of an animal over a 12 month period. Rather than plotting these parameters as individual graphs it might make more sense to combine the data into one graph so that reproductive condition can be pictured in relation to temperature and rainfall (e.g. Figure 3). Another example of plotting

all the data on one graph could be data (e.g. oxygen consumption) collected over the same time period but for different species or for one species at different temperatures or weights. Such graphical representation would allow immediate visual comparison of similarities or differences.

### **How Should The Data Be Plotted? - Independent and Dependent Variables**

When plotting a graph it is important to ensure that the right data are plotted on the right axis. In most cases there is a convention for deciding which is plotted on the X and which on the Y axis and is based on which of the variables is dependent on the other. For example, the rate of oxygen consumption of an animal may be dependent on ambient temperature, its body size or sex, but not the other way around i.e. the ambient temperature, body size or sex are not dependent on the animal's oxygen consumption. In the example just given, therefore, oxygen consumption is the dependent variable and this always goes on the Y-axis. Temperature, body size or sex are independent variables and always go on the X-axis.

Sometimes it is not possible to distinguish the dependent and independent variables and in these cases the choice of axes is arbitrary. For example there may be a relationship between two size parameters of an animal, both increasing as the animals grows. The two variables may be related, but it is not possible to say that one is dependent on the other.

### **Scale and Labelling**

Once the type of graph has been selected and the independent and dependent variables identified, the graph can be drawn. It is important to choose a sensible scale for the axes so that the graph does not look ridiculous (e.g. too compacted so that trends cannot be seen or too expanded so that trends are exaggerated). The axes must be labelled correctly and neatly by printing. Make sure that the labels are large enough to be easily read. Data points on point graphs should be represented by neat symbols and on comparative graphs different symbols must be used for different sets of data. If the bars of bar graphs and histograms are to have filling, ensure that this is done neatly. If points can be joined or variability is shown, make sure that a ruler is used to join the points. **Never draw free-hand.** When drawing the graph ensure that there is sufficient room (about 3 cm) on the paper to enable labels for the axes to be placed to the left side of the y-axis and beneath the x-axis.



## **The Legend**

The legend should contain enough text, along with the information on the graph itself, to enable the reader to understand the graph completely. If this is not possible, then the legend is inadequate. At the same time, the legend should be succinct. Avoid legends that begin “A graph showing....” or “A histogram showing....”.

## **DANGER: Plotting Graphs Using Computers!**

There are numerous computer software spreadsheets and graphical packages that are capable of producing publication quality graphs. Many of these are, however, designed for businesses who want eye-catching, colourful and unusual graphs to impress clients at presentations or in glossy reports. These features are undesirable for scientific graphs and many features required for good scientific graphs are often lacking. There are good scientific graphic packages available but the appearance of the final product still rests with the user. Even with the best graphic software, inappropriate selection of the type of graph and poor graph design results only in publication quality garbage. Producing graphs on computer can sometimes take a long time to get right and there are times when a neat graph produced with a pen and a ruler is still the best. The reports of undergraduate students are not penalised if graphs are drawn in pen and ink, as long as they are neat and meet the criteria outlined in this section. Nevertheless the ability to use a graph plotting package is a very desirable skill, and one which should be acquired.