

3.6 THREE-DIMENSIONAL (MULTIVARIATE) AND TWO-DIMENSIONAL-PLUS GRAPHS

3.6.1 Three-dimensional (multivariate) graph

Multivariate graphs have high information content, which can be good, but is sometimes more than can be shown readily. The main problem with these graphs arises when they become too complex to be easily interpreted. Exercise caution. There are various ways to make such graphs more effective.

x - y - z graphs use a ‘mock 3D’ representation on the flat (2D) page. It is usually hard to represent three dimensions accurately on a flat sheet, and only some types of datasets will lend themselves to this treatment. The wire-frame style (i.e. a 3D line graph) is generally best, but its effectiveness depends on the exact shape of the data (Fig. 21); some data may be hard to see, with points hidden behind other points. Various approaches to graphing a 3D dataset are shown in Fig. 22. The points-on-a-stick graph (a 3D-scatter: Fig. 22D) can be difficult to make sense of. Never draw such points without the sticks as this makes it impossible to interpret them! For on-screen analysis or presentations, rotating graphs can be very useful and effective, but are unsuitable for publication.

3D histograms and 3D bars (Fig. 23A) are generally poor graphs: at best they are hard to interpret; at worst they create unnecessary distraction from the data and qualify as ‘chartjunk’ (see section 4.11). It is generally better to present these data as 2D multivariate bar graphs beside each other, or as x - y graphs (Fig. 23B) if appropriate (see Box 2 for an outline of methods to convert 3D bar charts to x - y line graphs using Excel).

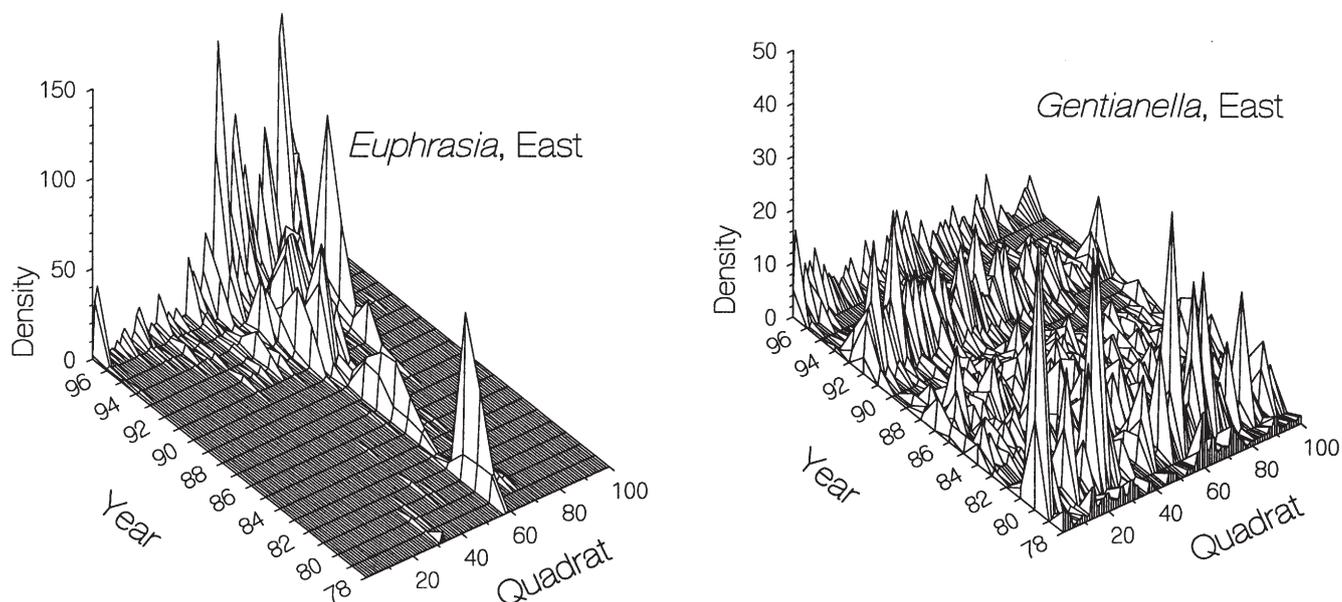
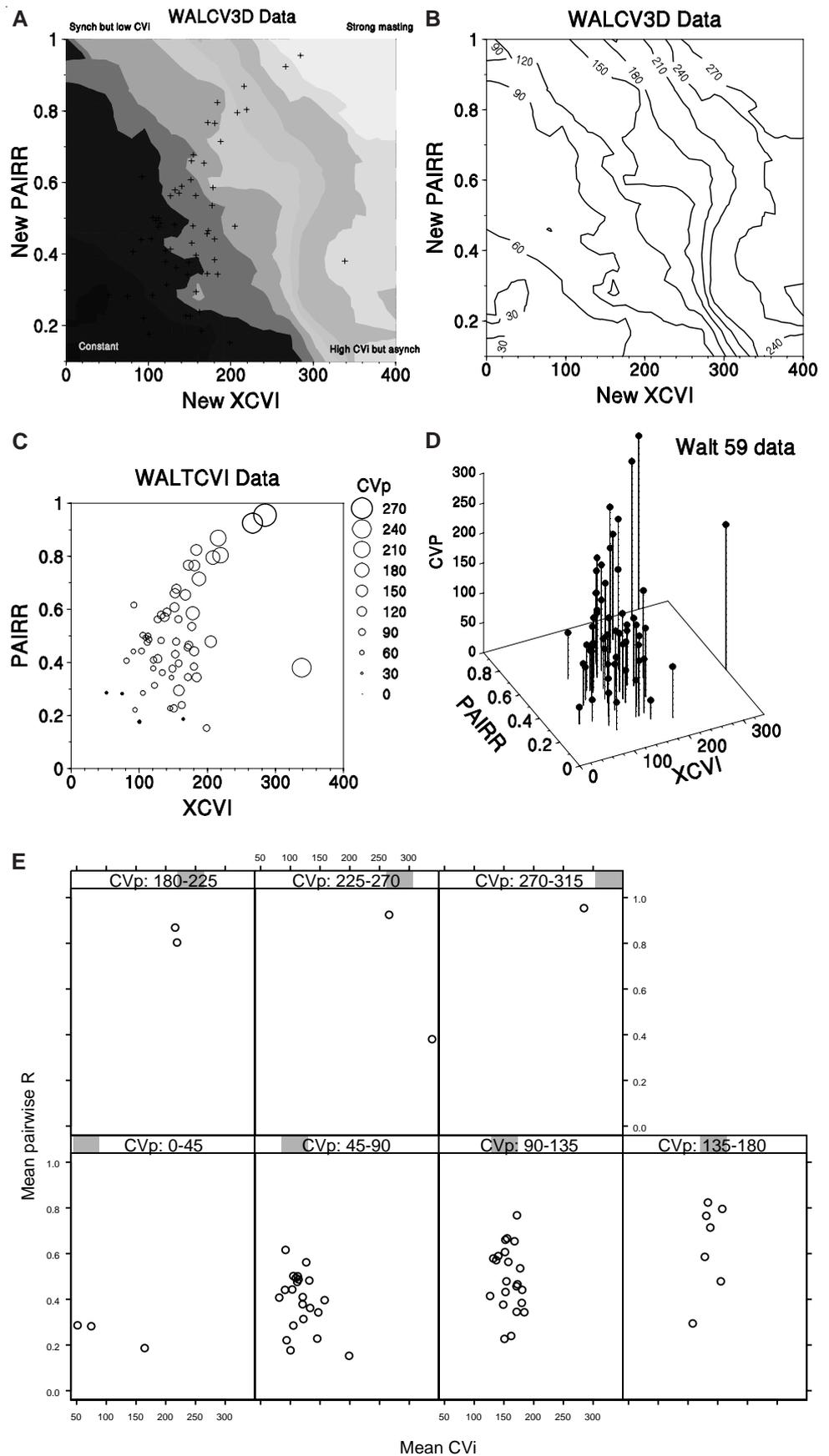


Figure 21. 3D graphs showing wire-frame representations, i.e. data are regularly spaced on a grid and each point is joined by lines. Note that this is much more successful where there is a clear, simple pattern (e.g. *Euphrasia*), preferably with a lot of short-scale autocorrelation (i.e. adjacent points tend to have similar values) than when there is a lot of scatter (e.g. *Gentianella*). This type of graph may be hard to draw if the original data are not collected on a regular x , y grid; interpolation of the values for the regular grid points may be necessary, which is undesirable and arguably misleading.

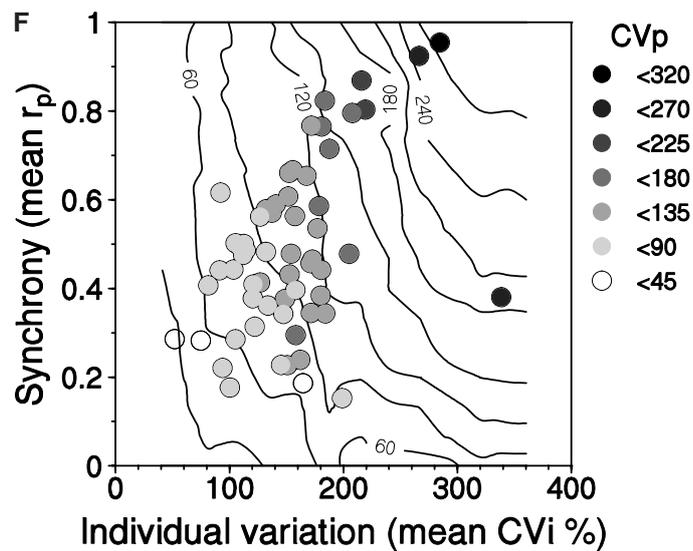
Original caption: Distribution of flowering individuals of four short-lived plants in the East transect, Castle Hill N.N.R., 1978–97. The total numbers of flowering plants seen in each 0.5 x 0.5 m quadrat in each year is shown. The species were *Gentianella amarella*, [*Rhinanthus minor*, *Medicago lupulina*] and *Euphrasia nemorosa*.

Figure 22. Six ways of graphing a 3D dataset to show the relationships among CVi (mean individual plant variation in seed crops), R (mean pairwise R—synchrony among plants), and CVp (population variation in seed crops). Four 3D representations of the same data (A–D) were tried before settling on the final published graph (F; overleaf); version E is an alternative approach, developed by IW. Version A shows the contours more clearly using colour and the location (but not the value) of the data points. However, as can be seen, this does not reproduce well in black and white. Version B shows only the contours, in monochrome. Version C shows the value of each data point (by the size of the circle) but not the contours. Version D shows the value of the 59 data points using a wire stick model, but not the contours. Version E shows the same 59 data points (but not the simulation contours), with nine panels for different levels of CVp, shown graphically in the strips at the top of each panel. The multipanel approach allows a clearer view of the relationship of CVi and R at the given levels of R, although it requires more space. Panel F (overleaf) was chosen because it shows both the actual values for each of the 59 real data points (by shading of the circles), and the contour lines for the simulations, all in monochrome as required by the journal. (The contour lines are more strongly smoothed in the final version (F) than in earlier versions (A, B), but should perhaps be smoothed even more to avoid giving an unwarranted impression of precision in the detailed contour patterns.)



Continued on next page

Figure 22—continued



Original caption: Interrelationships among mean r_p , CV and mean CVi based on randomized permutations. For each of the 59 datasets used in the analyses, annual data for individual plants were reshuffled to vary synchrony from high to low while holding mean CVi constant, and calculating the resulting CVp. Plotted are CVp contours resulting from 270 different combinations of each dataset, giving a total of 16,000 reshuffled datasets. Circles are the actual CVp values for the 59 datasets.

Box 2: Conversion of 3D bar charts (Fig. 23A) to x–y line graphs (Fig. 23B) using Excel

1. The key step is to **change the chart type**: select the Chart Area, click on 'Chart' on the menu bar, select 'Chart type', choose 'Line', and select the line with markers on it (usually the default).

2. **Maximise the size of the graph** within the overall chart window. Select the graph, by clicking in or near the graph itself until a dotted grey border appears. The graph can now be extended to take up as much as possible of the chart window by dragging on the black squares at the corners and in the middle of the sides. The default grey background is readily deleted by pressing 'Delete' on the keyboard while the graph area is selected.

3. **Labels for the axes** are essential, and are entered in the dialogue box from the 'Chart Options' option of the Chart menu, selecting the 'Titles' page. By default, the labels are in bold, so click on each label and change the bold format to regular. For publication, it is desirable to delete the overall title, except possibly for an on-screen presentation.

4. **Remove the horizontal gridlines** cluttering up the graph by clicking on one of them to select them, and again pressing 'Delete' on the keyboard.

5. **Remove the border on the chart window and the legend** by double-clicking on each in turn and, in the Patterns tab, setting 'Border' to none.

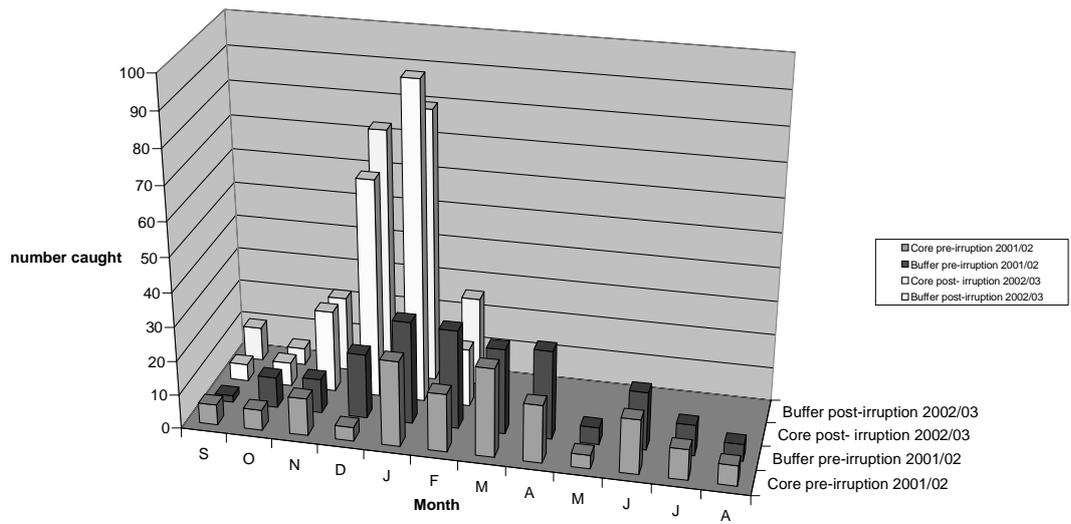
6. **Multiple lines need to be formatted** so they can be clearly identified, even when printed in black and white. For example, the pre-irruptions lines can be formatted as dotted lines with circles as markers, and the post-irruption can have solid lines with triangles. The core can be indicated with solid markers, while the buffer has hollow markers. To implement this, double-click on the line to get 'Format Data Series' dialogue box. Select 'Patterns', and under 'Lines' choose the line type and colour wanted. At this point, it is also possible to change the marker type, and foreground and background colours—to black, or no colour for example.

7. The **overall shape of the graph can be adjusted** if need be (provided it is embedded in a worksheet rather than in its own window) by selecting the whole graph window and changing it in the normal Windows manner.

8. **Final tidying up** may involve adjusting font size (in the tool bar), after selecting the whole graph or legend or an area of text. Then the graph area and legend may need to be resized, and the legend moved.

A

Okarito Rowi Sanctuary Stoat Captures



B

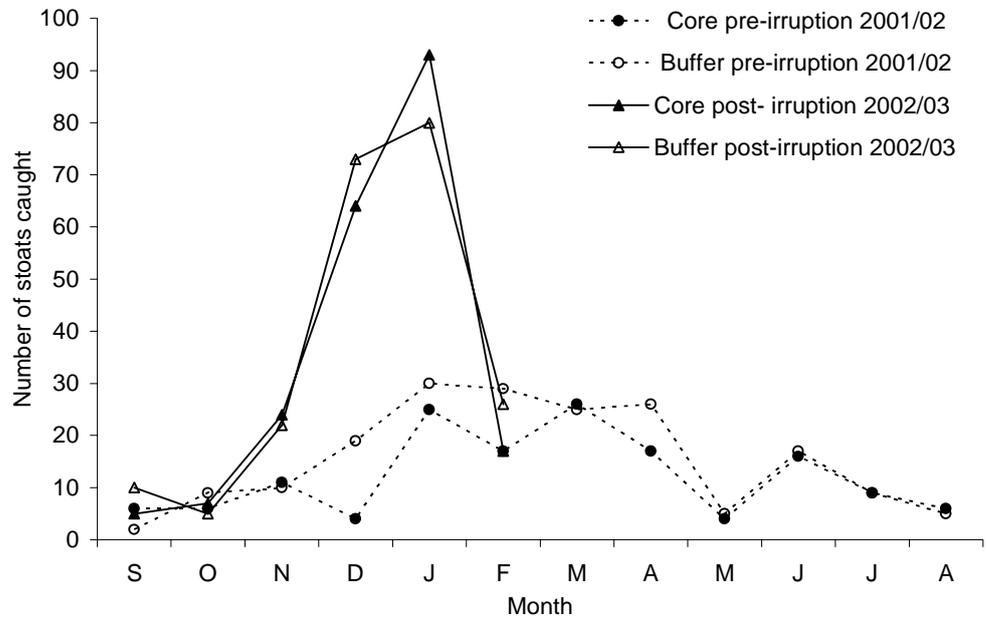


Figure 23. Stoat data from Okarito used in graph workshop. A. 3D bar chart intended to be published in colour. B. The same dataset presented as a monochrome x - y line graph (see Box 2 for methods on how to convert A to B).

3.6.2 Two-dimensional (2D)-plus graph

This is an x - y graph with various other ways of representing the value on the third axis. These usually work better than ‘mock 3D’ graphs. This type includes data maps, where the x - and y -axes show the physical location on the ground and the third dimension shows the value of some variable of interest at that spot. The third dimension can be shown using contours as in maps (this works only where the change in the third dimension is gradual); bubbles of varying sizes; shading; small pies or bars; or trajectories. An example using bars is shown in Fig. 24. Some of these methods work with particular data—for example, trajectories can be useful for showing a sequence of points through time; points are marked and successive points linked by lines, with labels or arrows to show the start and end.

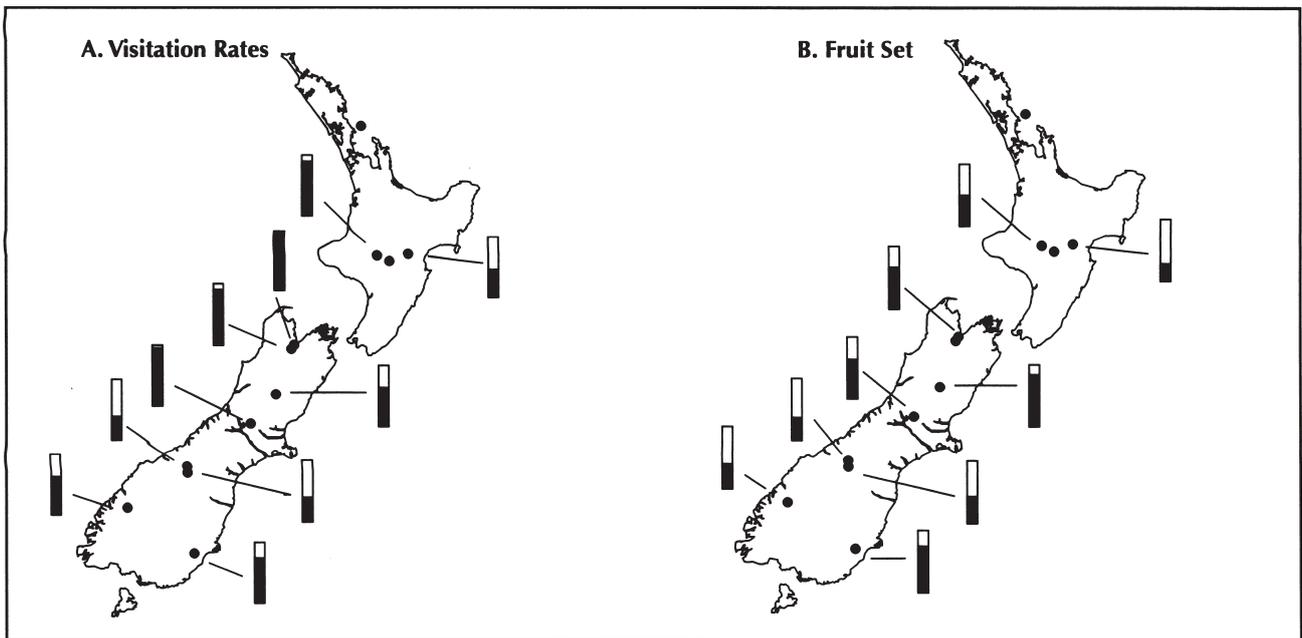


Figure 24. Two data maps, showing the value of a third variable (left, visitation rate by birds to mistletoe flowers; right, fruit set by mistletoe plants) at points located by geographic position (latitude, longitude). The third variable is shown by the fraction of a bar which is filled; this is better than simply having dark bars of variable length. Note that if the key message was the relationship between visitation rate and fruit set at each site, a plain x - y graph plotting the two directly against each other would be superior. The data could alternatively have been presented with the visitation rates (and possibly site names) on the map instead of the bars.

Original caption: Visitation rates for all mistletoe study sites during the 1997–98 season are shown in figure 9a. Visitation was quite low at Lake Ohau, the Temple and in the Kawekas. This signifies a lack of pollination in these areas. Fruit-set for all mistletoe study sites during the 1997–98 season are shown in figure 9b. Fruit-set generally correlated with visitation rates.

3.7 MULTIPANEL GRAPH

Multiple related graphs have become increasingly easy to prepare with the ready availability of high-powered computers. If you have more than two or three variables, this is really the only failsafe method to represent data: with care it can work for almost any dataset, as opposed to the tricks like those discussed above that work only for particular types of data. The use of multiple graphs (e.g. Fig. 10) has been systematised and extended by Cleveland (1993) to support graphs such as Figs 12B, 22E and 25. The simple but powerful underlying concept is to break the limitations of the two dimensions that are represented on two axes by using multiple panels of such graphs. Conventional graphs, typically displaying two to three variables, are systematically arranged in a series of panels that allow users to see a number of variables and their interactions simultaneously. With careful allocation of variables to axes, it becomes possible to view and even analyse complex interactions graphically, without needing to standardise, or to fit models that usually require strong assumptions to be made. Relationships within the data, obscured by the limitations of standard two- or three-variable graphs, are illuminated once other influential variables are controlled for within the multipanel graph structure. The power of this approach is shown in Fig. 25.

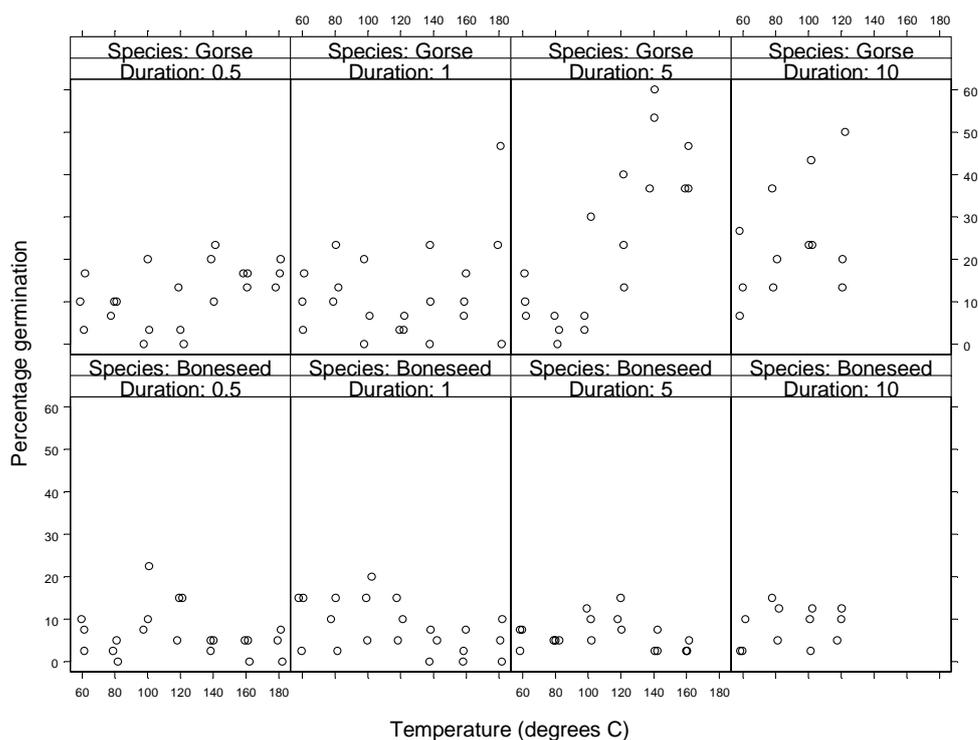


Figure 25. This set of graphs from S-PLUS shows, within each panel, germination rates (on the *y*-axis) for seeds exposed to different temperatures (on the *x*-axis). However, in addition to these two variables, an additional two variables are shown using panels. The upper row of panels is for gorse, and the lower for boneseed; the horizontal panels represent increasing exposure times to the raised temperatures. Thus four variables are presented. One is categorical—gorse / boneseed; one is continuous—percentage germination; and two represent discrete levels of potentially continuous variables—temperature (eight levels, but note that the temperature data has been jittered slightly to avoid points being hidden under each other) and duration of exposure (in minutes; four levels). This graph allows an almost immediate answer to a key research question for the experiment: ‘Does gorse or boneseed germination respond to heat’, and allows rapid exploration of further questions, like the effects of increasing temperature and duration of exposure to heat.

Continue to next file: docts32c.pdf