

TOPIC 4

GRAPHICAL PRESENTATION

Public agencies are very keen on amassing statistics – they collect them, raise them to the n th power, take the cube root, and prepare wonderful diagrams. But what you must remember is that every one of the numbers comes in the first instance from the village watchman, who puts down what he damn well pleases.

Sir Josiah Stamp

Introduction

diagrams are easy to understand

We have so far confined ourselves to presenting data in tabular form. The presentation of data in a diagram is often easier for a user to understand. But care needs to be taken when making graphs, because it is easy to mis-represent statistical data. In this topic the main types of graphical presentation are presented along with the types of data the graphs are most suitable for depicting statistical data.

General rules when preparing graphs

people prefer graphs

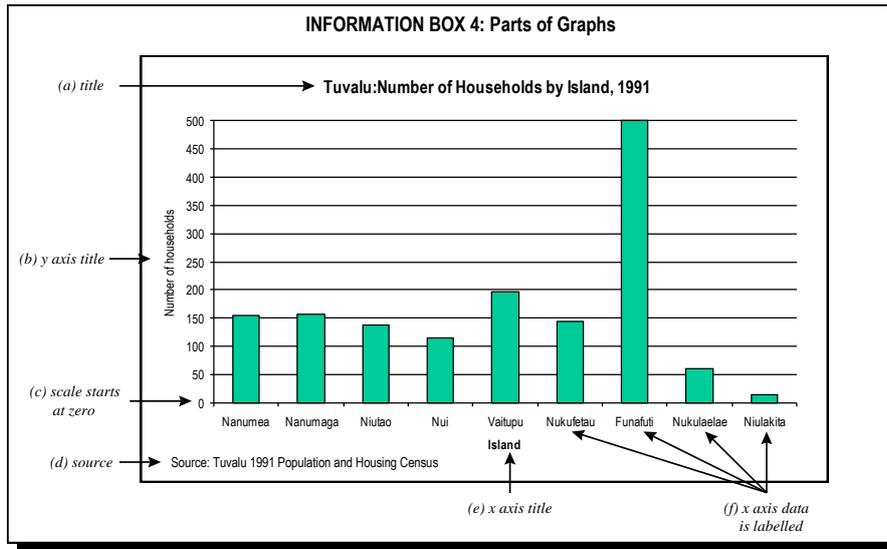
Many people prefer to look at a graph, rather than examine a table of numbers or read a page of writing. A graphical presentation is often more easily understood than a numeric one – a picture is worth a thousand words, although in this case numbers.

general rules

We must always be aware that different graphs show different aspects of the same data. As we prepare graphs there are a number of general guidelines to follow to ensure that the graph can be understood and interpreted.

- ☆ Graphs should have a clear, self-explanatory title.
- ☆ The units of measurement should be stated.
- ☆ Graphs should be simple and not too cluttered.
- ☆ All axes should be carefully labelled.
- ☆ Include the source of the data.
- ☆ The scale on each axis should not distort or hide any information. That is, the graph should show the data without changing the message of the data.

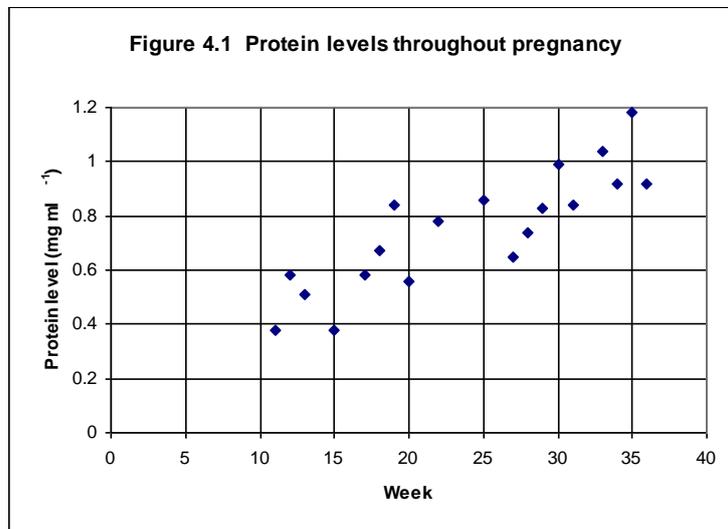
- ☆ Graphs should clearly show any trend or differences between the data.
- ☆ Use 'scale breaks' for false origins – where the scale on the axis does not start at zero to avoid misinterpretation. A scale break is indicated by a 'wiggle' on the axis.
- ☆ The type of presentation needs to be chosen carefully. This includes the size and shape of the plotting symbol (e.g. asterisks, dots or squares) and the method, if any, of connecting points (straight line, curve, dotted line, etc).
- ☆ Graphs should be accurate in a visual sense – if one value on the chart is 15 and another 30, then the second value should appear to be twice the size of the first.
- ☆ A trial and error approach can be very helpful in improving graphs. It is unlikely that you will find the right graph the first time.



Scatter diagrams

simple to see relationships clearly

One of the simplest forms of graphical presentation is the **scatter diagram**. Scatter diagrams are usually used to show the relationship between two variables (bivariate data), to detect any outliers or to detect clusters of observations.



Source: Illustrative data only.

outliers – the unexpected

Figure 4.1 depicts quantitative bivariate data, where each observation consists of two measurements – time in weeks and a protein level. It shows the change during pregnancy in the protein levels. We can see the **outliers** – that is, the observations which are higher or lower than the general pattern of the other observations.

Bar charts**qualitative and quantitative data**

Bar charts are used to describe the distribution of **qualitative and quantitative data** where the data is grouped into **equal sized class intervals**. *The frequency of the variable is represented by the height of the bar.*

comparisons

Bar charts are good to use when you want to make a comparison between two variables – such as differences between males and females. With nominal data, the order of the categories on the axis is arbitrary, although the variables are often ordered from largest to smallest in frequency.

vertical or horizontal

Bar charts can be vertical or horizontal, although they are usually vertical with the variable categories on the x axis and their frequencies on the y axis.

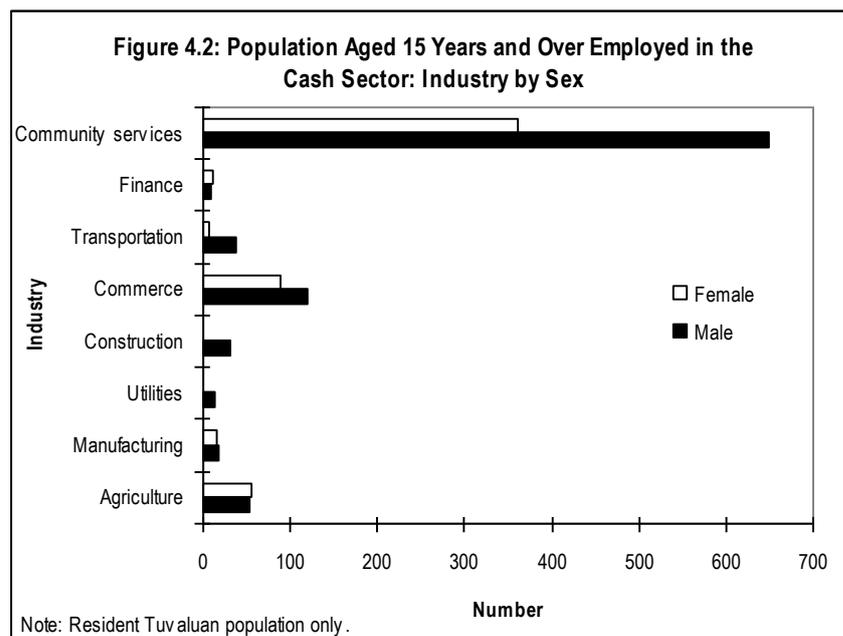
TIP

Vertical bar charts are better to show the frequency distribution across the variable categories.

Horizontal bar charts are used when the labels for the variable categories are too long to fit neatly on the x axis. It is also easier to read the scale on horizontal bar charts.

example

The chart used in Information Box 4 is a bar chart (vertical). Figure 4.2 is a horizontal bar chart. Imagine how small the labels would have to be to fit neatly on the horizontal axis!



Source: Tuvalu 1991 Population and Housing Census.

Pie charts

slice of the pie

Pie charts are one of the most commonly used graphs for **qualitative** data. Each “slice” of the pie represents the relative class frequency of that category. They are simple to construct, but are most effective when you have **five or fewer classes** in your data.

small number of classes

Generally we use pie charts when there is a small number of classes and we are interested in the relative frequency of *each category* (rather than comparisons with other variables).

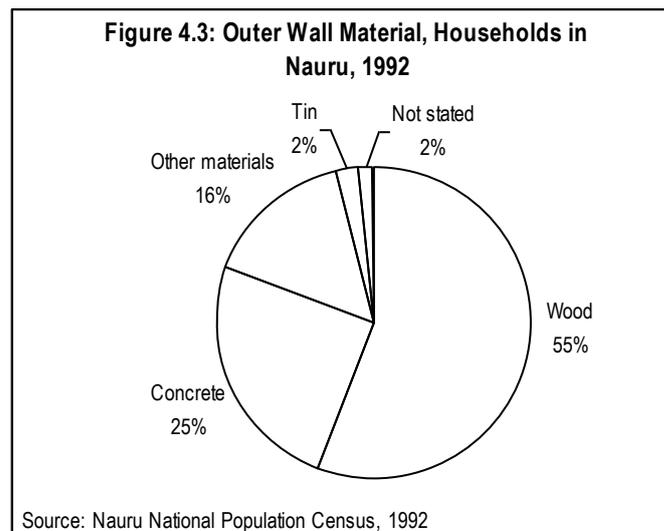
TIP



An important guideline when constructing pie charts is that the segments should be **ordered by size**. This makes it easier to read and to understand the information.

example

Figure 4.3 below contains data from the Nauru 1992 Census. In this chart the percentage of each variable is displayed. This is a useful option when creating pie charts since there is no scale to read from as in bar charts.



Line graphs

independent and dependent variables

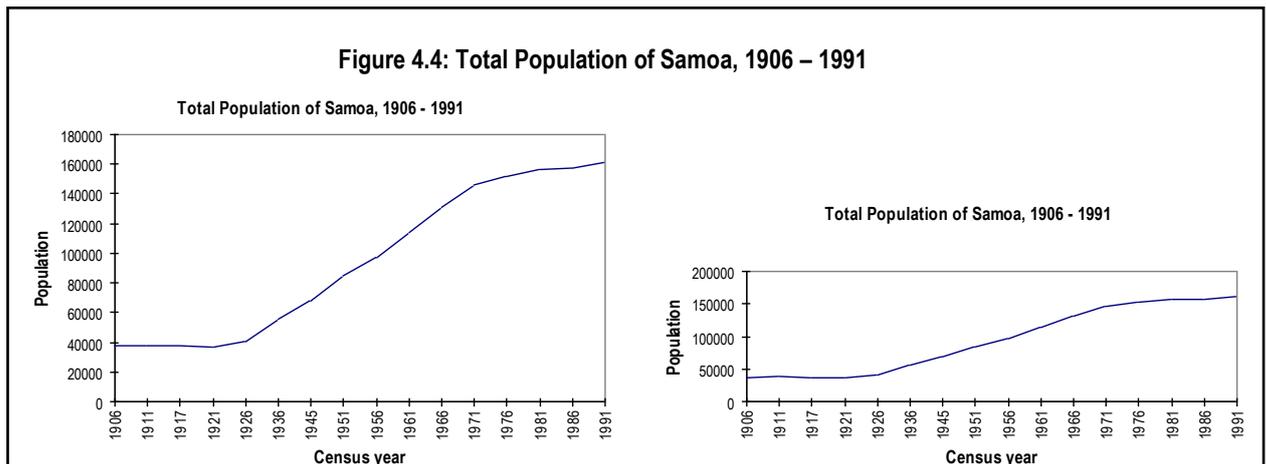
Line graphs are used for bivariate data. They have two axes – the horizontal (x) axis on which the scale for values of the **independent variable** is recorded, and the vertical (y) axis on which the scale for the **dependent variable** is recorded. The independent variable is the variable on which predictions are based, while the dependent variable changes in relation to the independent variable. Time is an example of an independent variable – it depends on nothing. For example, when we look at population counts from successive censuses, we could use a line graph. Census year would be the independent variable, and the population count would be the dependent variable.

common usage

Line graphs are commonly used to display data *over a time period* and when you want to show *trends* in data. Care must be taken when preparing line graphs that the scale on the vertical (*y*) axis does not give the wrong impression about the data.

example

Figure 4.4 contains two line graphs with the same data, but different scales are used on the *y* axis. The second graph gives a much different impression than the first. Looking at the first graph you would assume that the population increased quite radically from the 1930s – 1960s, but the second graph does not give the same impression. It is particularly important when comparing different populations to keep the same scale.



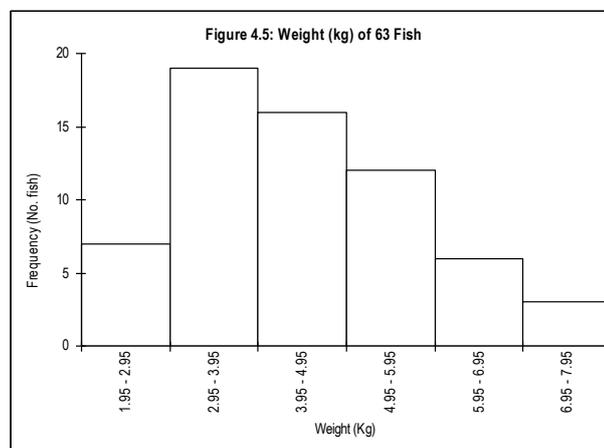
Source: Census of Population and Housing, 1991, Western Samoa

Histograms**area of each bar is proportional to frequency**

The most common method of representing a frequency distribution for **continuous data** (or data that is assumed to be continuous) is by using a **histogram**. A histogram is a vertical bar chart with the *area of the bars (not necessarily the heights of the bars) proportional to the class frequencies* (whereas for bar charts both the height and area of the bars is proportional to the class frequency).

different from bar charts

Histograms differ from bar charts because the vertical bars are always side by side, without gaps. This is to reflect the continuous nature of the data. Generally a histogram has equal width bars, although when class intervals are not equal the width of the bars has to be adjusted.

example

Source: Table 3.1

area is important

You must remember when plotting a continuous distribution, it is the **area of each rectangle** (and not the height) that is **proportional to the frequency**. It is only in the case of equal class intervals (as in Figure 4.5) that the frequency is proportional to the height of the rectangle.

unequal classes

Let us look at another example of a histogram. The histogram in Figure 4.6 (from Table 4.1) shows how we should deal with a distribution with unequal class intervals and with open-end class intervals. Remember it is the **area of the bars** which is important in a histogram. This means when we have unequal classes we have to make sure the area of each bar accurately represents its frequency. The class 100 – 199 is twice the size of the 50 – 99 class, so in the histogram the width of the 100 – 199 rectangle has to be twice the width of the 50 – 99 rectangle.

height = frequency density

To ensure the area of the rectangle reflects the class frequency, the height of the rectangle must then be the frequency density.

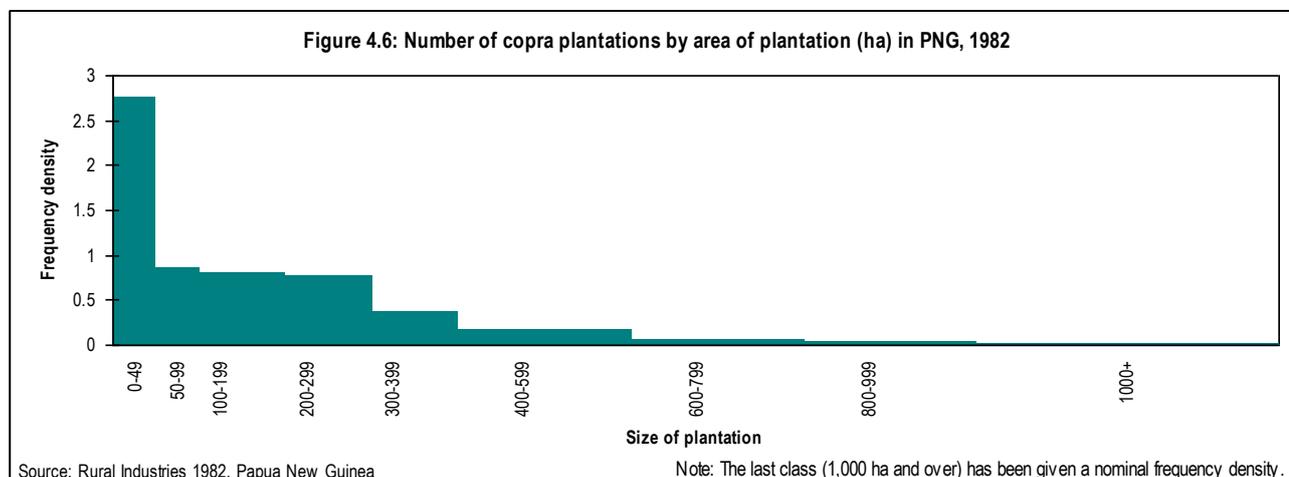
Note that the method for dealing with open-end class intervals has been footnoted, to ensure the reader is informed of the method used.

Table 4.1 Number of Copra plantations by area of plantation in PNG, 1982

Class Group (ha)	Class Interval	Class Frequency	Frequency Density (per ha)
0 - 49	50	138	2.76
50 - 99	50	43	0.86
100 - 199	100	81	0.81
200 - 299	100	77	0.77
300 - 399	100	38	0.38
400 - 599	200	34	0.17
600 - 799	200	13	0.07
800 - 999	200	6	0.03
1,000 and over		8	–
Total		438	

Source: Rural Industries 1982, Papua New Guinea.

Figure 4.6: Number of copra plantations by area of plantation (ha) in PNG, 1982



Source: Rural Industries 1982, Papua New Guinea

Note: The last class (1,000 ha and over) has been given a nominal frequency density.

open ended classes

In Figure 4.6 there is also a problem in deciding how to deal with open-ended classes, in this case the '1,000 and over' class. We do not know the width or the interval of this class and we have no way of making even a reasonable guess. In some situations it is possible to assume an upper class limit – for example if we were dealing with data on age at marriage we could assume an upper limit of 60, since it is unlikely that people over 60 would marry.

In Figure 4.6 we have no way of knowing the upper limit of the 1,000 and over class so we cannot calculate the height of the rectangle to represent this part of the total frequency. It would be wrong to leave it out altogether, so we have to decide what to do with this class.

two ways of dealing with open ended classes

The problem can be dealt with in one of the two ways. Firstly, we can assume an upper limit for the distribution and draw the rectangle accordingly (such as with the age at marriage illustration). The second alternative is to draw the rectangle with a nominal height but leave it open and footnote the method applied, as in Figure 4.6 above. This would then indicate an open-ended interval. This second method, however, only works when the frequency in the open-ended class interval is small (as is the case in Figure 4.6), but this should be the case (large frequency open ended classes should always be avoided).

discrete data

When we draw a histogram for **discrete data**, we have two possible approaches. Let us look at Table 4.2 given below:

Table 4.2 Employment size group by number of paid employees, 1996

Employment size group	Wage and salary earners	Earners per enterprise = freq * class mid point
1–4	5,053	=5,053 ÷ 4 = 1,263.25
5–19	15,260	=15,260 ÷ 15 = 1,017.33
20–49	15,312	=15,312 ÷ 30 = 510.40
50–99	11,714	=11,714 ÷ 50 = 234.28
100 and over ¹	62,742	= 62,742 ÷ 400 = 156.85
Total	110,081	

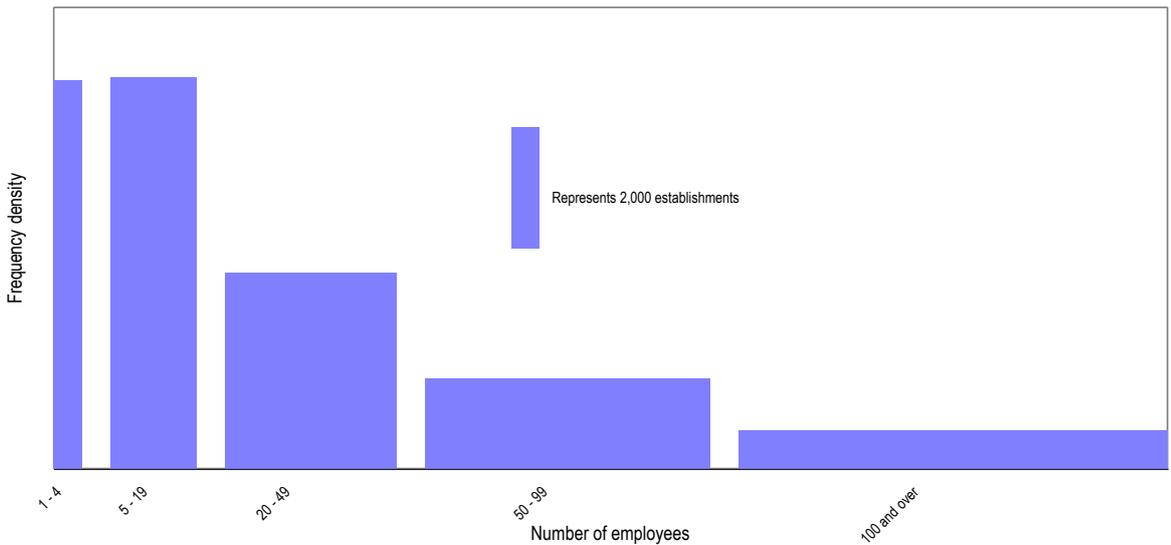
Assumed upper limit of 500.

Source: 1996 Annual Employment Survey, Fiji Bureau of Statistics, Table 2, p. 15.

two alternatives

The first approach is to draw a histogram with the rectangles separated by spaces. This signifies to the reader that the data is discrete. This approach is shown in Figure 4.7.

Figure 4.7: Employment size group by number of paid employees, 1996



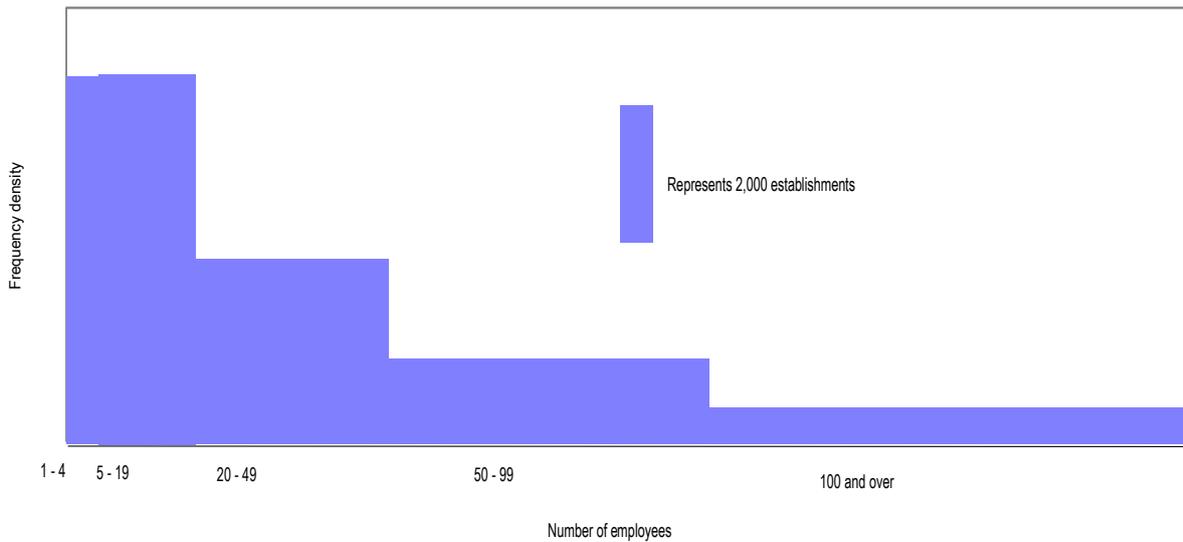
Source: 1996 Annual Employment Survey, Fiji Bureau of Statistics

Note: The 100 and over class has been given a nominal frequency density

second alternative

The second approach is to proceed as if the data were, in fact, continuous. If we have a distribution similar to that given in Table 4.2, then the unit of measurement (in this case single persons) is small compared with the range of values we have observed. In such a case, we can draw an approximate histogram by proceeding as if the data were continuous. The histogram is drawn in Figure 4.8 and looks somewhat like Figure 4.7 since the horizontal scale is divided into unequal intervals. You must realise that *Figure 4.7 is only an approximation* because the data are discrete and so we cannot, strictly speaking, represent them by a continuous rectangles. For the purpose of representing the general shape of the distribution, however, the diagram is quite adequate.

Figure 4.8: Employment size group by number of paid employees, 1996



Source: 1996 Annual Employment Survey, Fiji Bureau of Statistics

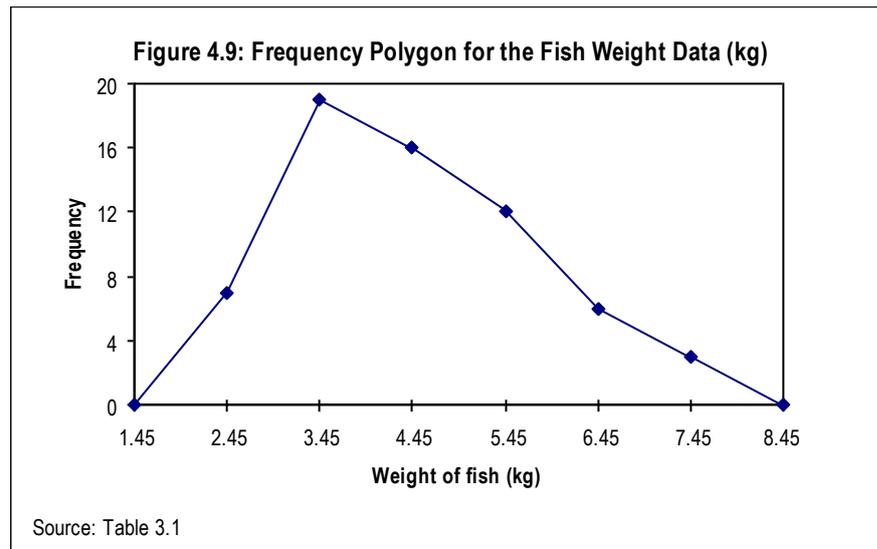
Note: The 100 and over class has been given a nominal frequency density

Frequency polygons

frequencies plotted at class midpoint

An alternative type of graph, suitable for **continuous data**, or for *discrete data which can be considered to be approximately continuous*, is a **frequency polygon**. In this graph the frequencies of each class are plotted at the class midpoint, and successive points joined up by straight lines.

An example of a frequency polygon with equal class intervals is given in Figure 4.9 below. The data used is that of fish weight data from Table 3.1.



unequal class width

When the class intervals are unequal lengths then, as in the case of a histogram, the frequency densities or relative frequency densities are to be used.

must start and end at the x axis

The beginning and end of the polygon should be extended to the horizontal axis, to the mid-points of the classes below and above those covered by the distribution.

For a discussion of the graphical presentation of cumulative frequency distributions, refer to the following section 'More on Frequency Distributions'.

... Exercises ...

1. (a) Using Excel construct a pie chart for the following data:

Table 1. Country of origin of visitors to Port Vila, 1997
Random sample of 35 visitors from June to November

Country	Number of Visitors
Australia	14
Japan	3
New Caledonia	5
New Zealand	6
USA	5
Total	33

Source: 1997 Vanuatu Visitors Survey.

- (b) Briefly describe the distribution of the data in the pie chart, and assume your reader is not familiar with pie charts.

2. Construct a histogram in Excel for the following data:

Table 2: Shoppers at the Vegetable Market over 30 Days

Number of shoppers	Frequency
25 - 34	3
35 - 44	5
45 - 54	6
55 - 64	5
65 - 74	4
75 - 84	3
85 - 94	4
Total	30

More on Graphical Presentation

Cumulative frequency distributions

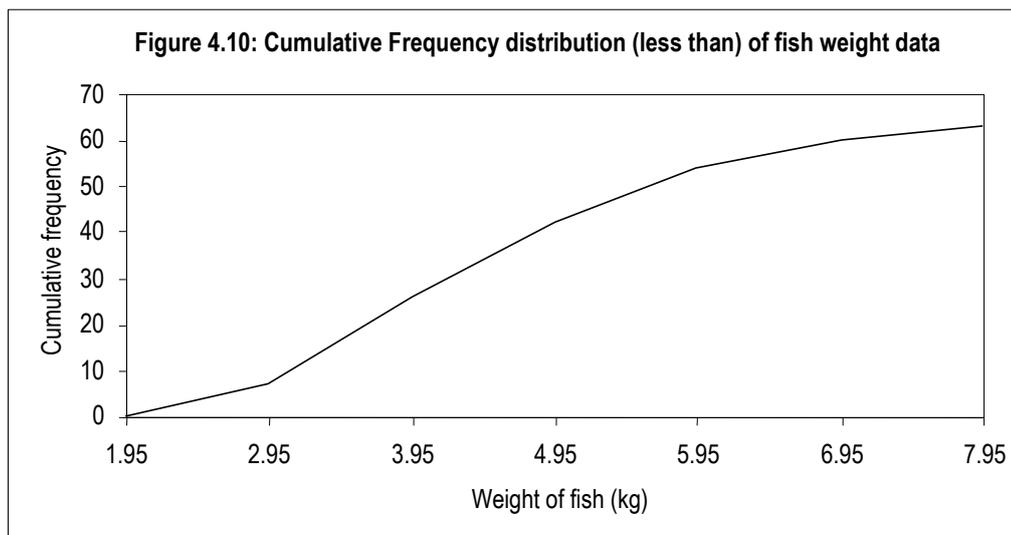
ogives

Just as for some purposes it is better to construct a cumulative frequency distribution, so we find it useful to represent this graphically. A cumulative frequency distribution of the fish data is given in Table 4.3. A graphical representation of the cumulative frequency distribution (either less than or greater than) is called the **ogive** and the less than ogive of the Table 4.3 data is given in Figure 4.10.

Table 4.3 Cumulative frequency distribution of fish weight data (kg)

Class Group	Class frequency	Cumulative Frequency (less than)	Cumulative Frequency (greater than)
2.0 - 2.9	7	7	63
3.0 - 3.9	19	26	56
4.0 - 4.9	16	42	37
5.0 - 5.9	12	54	21
6.0 - 6.9	6	60	9
7.0 - 7.9	3	63	3
Total	63		

Source: Table 3.1



Source: Fish Data from Table 4.3

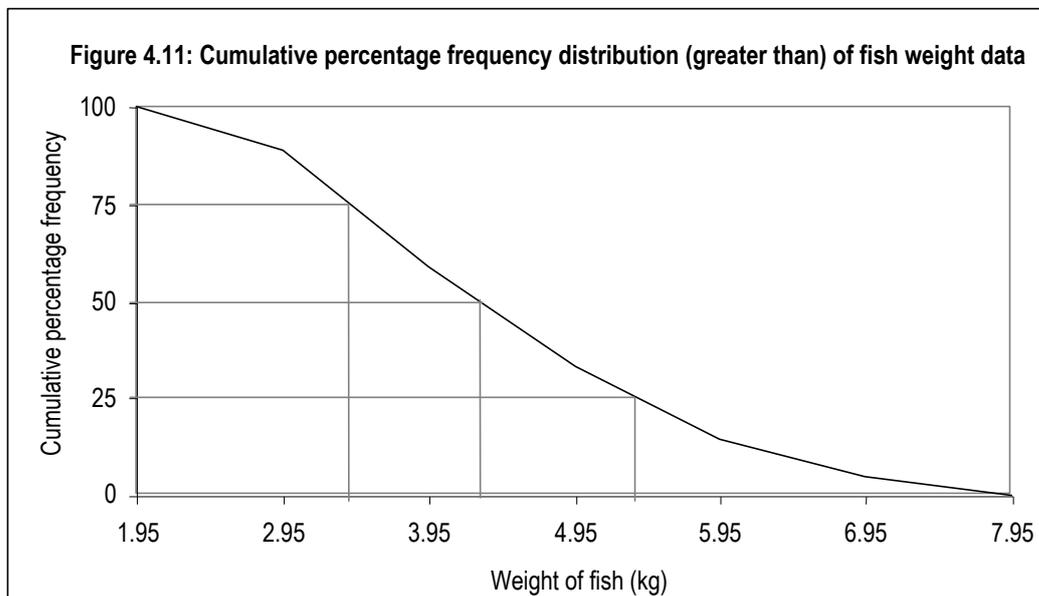
greater than ogive

The line in Figure 4.10 represents the cumulative frequency distribution (less than). We can also draw a similar diagram for the cumulative frequency distribution (greater than). This is shown in Figure 4.11. The data used is that from Table 4.4 over.

Table 4.4 Cumulative frequency distribution (greater than) of fish weight data (kg)

Class Group	Class frequency	Cumulative Percentage frequency (greater than)
2.0–2.9	7	100.0
3.0–3.9	19	88.9
4.0–4.9	16	58.7
5.0–5.9	12	33.3
6.0–6.9	6	14.3
7.0–7.9	3	4.8
Total	63	

Source: Table 3.1



Source: Table 4.4

cumulative frequencies show distribution of the data

In Figure 4.11, instead of plotting the actual frequencies on the vertical axis, we have plotted the cumulative percentage frequencies. We can clearly see from this diagram that 100 per cent of the observations were greater than 1.95 kilograms, 75 per cent are estimated to be greater than 3.4 kilograms, 50 per cent are estimated to be greater than 4.4 kilograms and 25 per cent are estimated to be greater than 5.4 kilograms. This is the graphical representation of quartiles, will be discussed in detail in the section “More on measures of location”.

two distributions

Finally, let us look at the ogive of the age at marriage of brides and grooms data from Table 3.14.



Source: Table 3.14

comparisons between variables and observations

Because the two distributions have the same base, it is helpful to show their cumulative frequencies on the same diagram. In Figure 4.12, we can clearly see the 'lag' effect, that is, the difference of age at marriage of brides and grooms. Grooms tend to be older than brides as can be seen from the ogive above.

Excel 97 – Creating charts

It is easy to create and format charts in Excel using the Chart Wizard (). Charts can be *embedded* on the worksheet next to your data, or they can appear as a *chart sheet*. A chart sheet is a sheet in a workbook that contains only a chart. The position of the chart depends on your preference.

WARNING

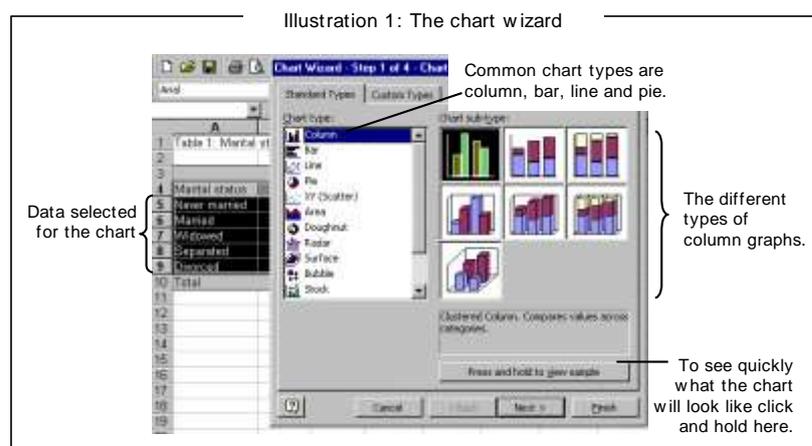


Charts are easy to create in Excel, and the best way to create them is as a chart sheet. Chart sheets do not have to be formatted as much for printing or copying as an embedded chart. As you will see from the following exercises, embedded charts need a lot of formatting before they can be copied or printed.

Create a chart

The main thing to be aware of when creating a chart in Excel is to select the correct *x* and *y* variables. For example, when basing a chart on a table containing Male, Female and Total variables, you wouldn't usually chart the 'Total' column. Generally you do not chart the total column if you are charting the parts which make up the total. If you include the total it will affect the scale of the chart, and change the emphasis of data items.

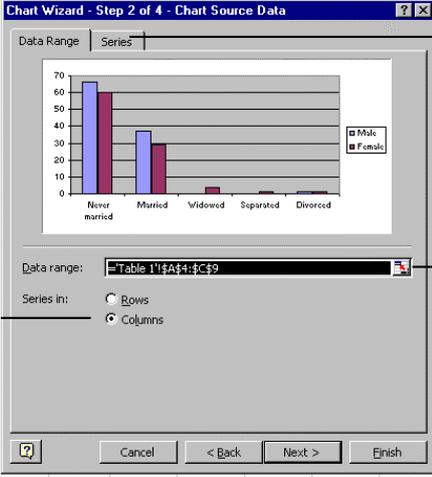
1. Select the cells you want to chart.
2. Click on the Chart Wizard button (). The Chart Wizard appears.
3. In the Step 1 Dialog box you select the type of chart you want to make. Select the correct type of chart for your data.



4. In the Step 2 Dialog box, the chart is previewed. If it doesn't look the way you want it to you can change the data and labels selected for the chart. This is quite difficult to use if you aren't familiar with

specifying cell ranges in Excel. Some people find it easier to cancel out of the chart and make the data selection again, and start the chart over. Click .

Illustration 2: The chart wizard - Step 2



If the graph is not showing the correct **labels** select them in the Series tab.

If the graph is not showing the correct **data** click here to change the data range.

Usually your data will be in column format.

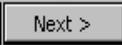
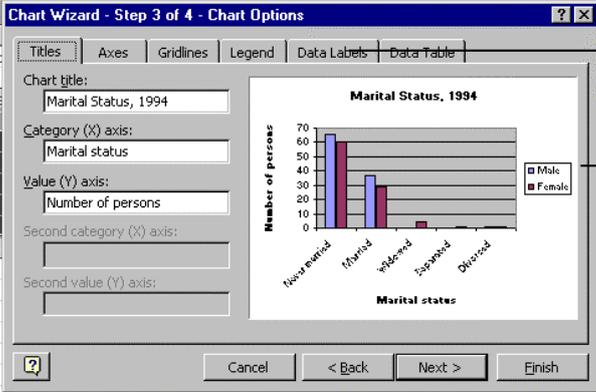
5. In the Step 3 Dialog box, on the Titles tab, type in the chart title, and labels/units for the x and y axes. You can click on the different tabs to change the legend, data labels and data table. Displaying the data table can be useful for presentations. Click .

Illustration 3: The chart wizard - Step 3

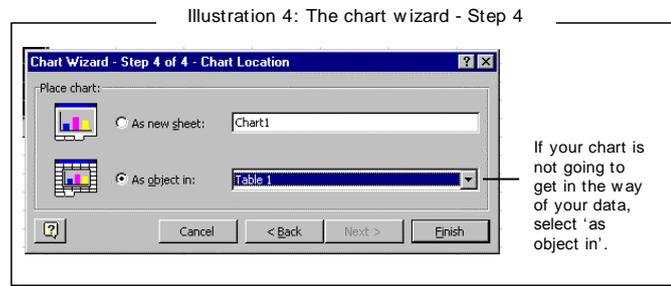


If you want the data values displayed use the **Data Labels** tab.

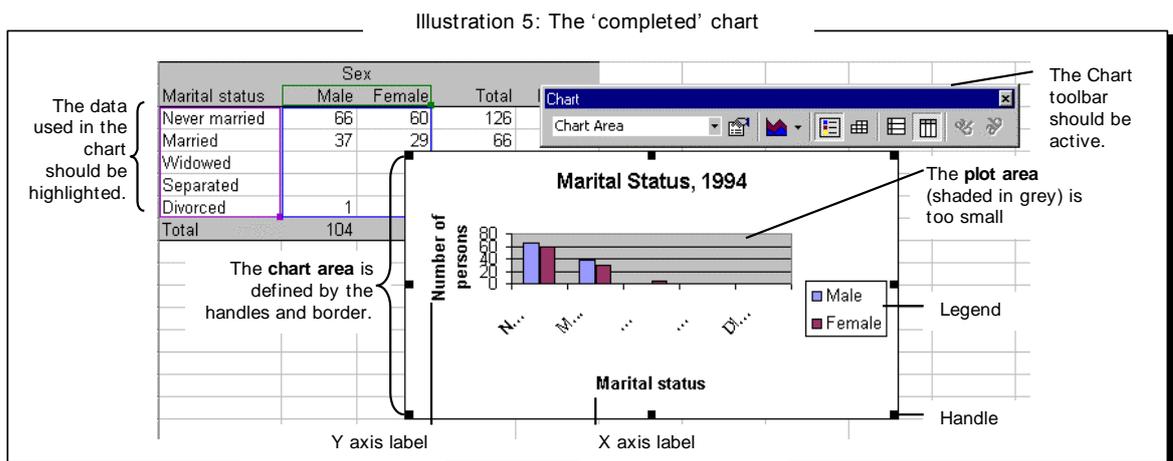
Use the **Legend** tab if you do not want a legend.

The title and labels are displayed in the preview.

6. In the Step 4 Dialog box, select the location of the Chart. The formatting of the 'As Object In' option is not as good as the 'As New Sheet' option. When working with a large amount of data, you would select the 'as new sheet' option. In this example. The 'as object in' option has been used.



7. Click .
8. The new chart is created in the middle of the active worksheet. Your worksheet could look similar to Illustration 4. The chart looks bad and needs formatting!! Excel is not the best software for creating charts!



Format a chart for printing

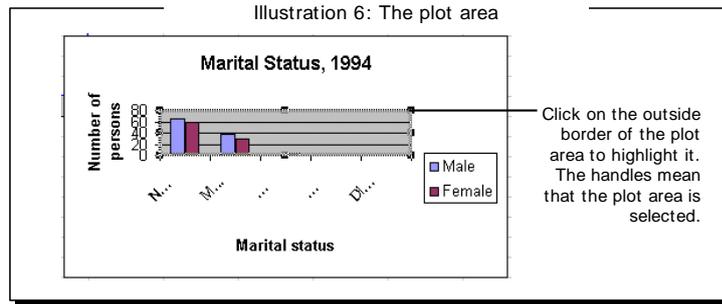
TIP



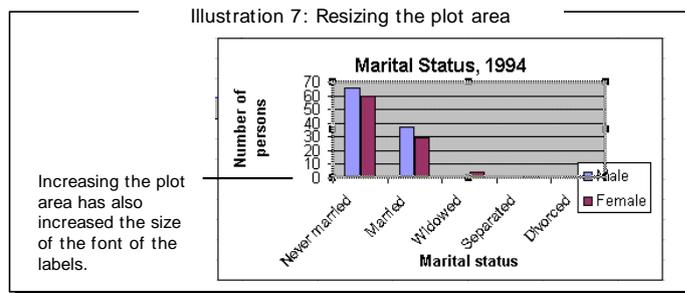
As a general rule, if you want to change the look of something on your chart select the object and double click on it. This will open a dialog box that should contain options for formatting.

Now that you have created an embedded chart, you need to format it for printing – the first thing to fix is the Chart Area which is too small.

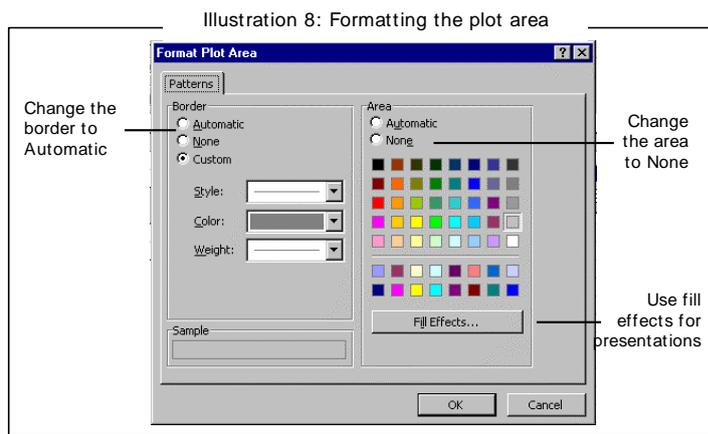
1. Click on the border of the grey Plot area, and handles should appear.



2. Hold the mouse pointer over the handle in the top centre of the page, and it will change to an \cdot . You can now drag on the handle to make the Plot area larger. Do the same with the middle handle at the bottom of the Plot area.
3. Your chart should now look Illustration 7. Notice how the scale of the chart changed when the size of the plot area was increased.

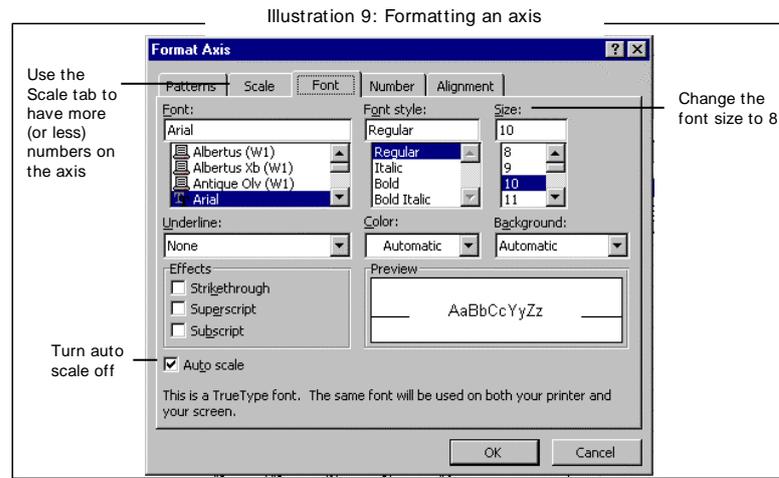


4. The grey plot area is not very good for printing and photocopying (unless you have a colour printer). You can easily change it to clear (white) by holding the mouse over the Plot area until you see the text "Plot area" (this is called a chart tip label). Double click and the Format Plot Area dialog box appears.

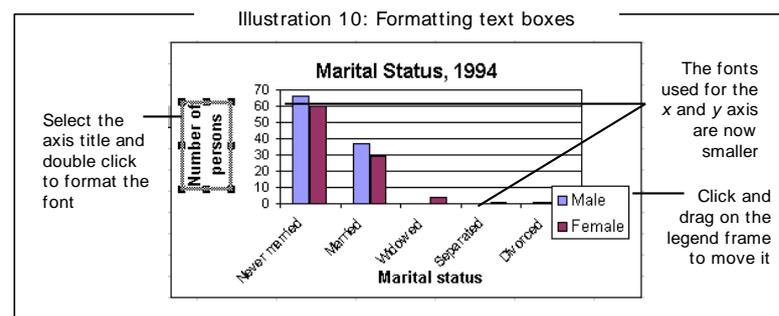


5. Under Border, change the option from Custom to Automatic. Under Area, change the option from grey to None, and press e to close the dialog box. The plot area should now be white.
6. Another puzzling feature is the automatic font size of the title and axis labels. By default Excel 'auto scales' the font – that means if you make the chart area bigger, Excel will adjust the size of the font automatically. This is not a good feature of Excel! Double click on the y axis to open the Format Axis

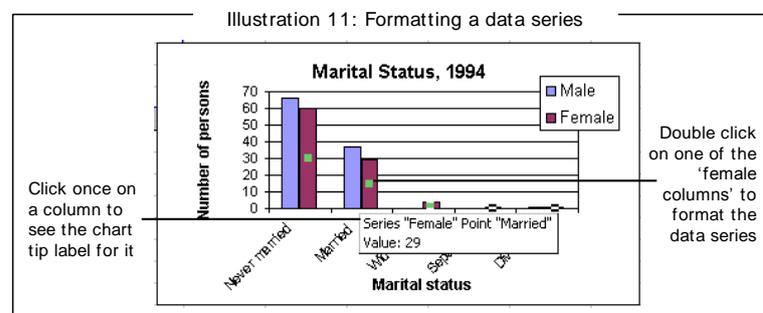
dialog box (hold the mouse over the y axis until the chart tip label says 'value axis'). On the Font tab change the font size to 8 and turn Auto scale off.



7. You have to do this for the title, the legend text and the x axis label as well. Your chart should now look like Illustration 10.

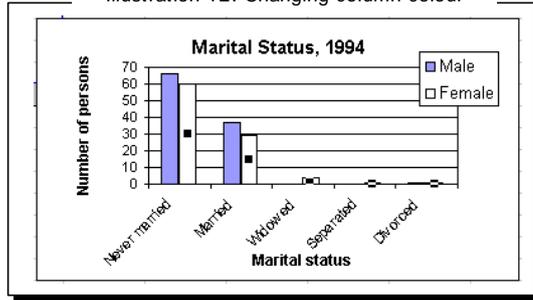


8. If you have a colour printer you won't have problems with the colours of the bars. But if you don't, and you will be photocopying your report, it is best to change the colours to black and white. To change the colour of one of the 'series', double click on one of the bars. If you click once on a data series, small boxes should appear on it, as shown in Illustration 11. Double click on the data series to open the Format Data Series dialog box.



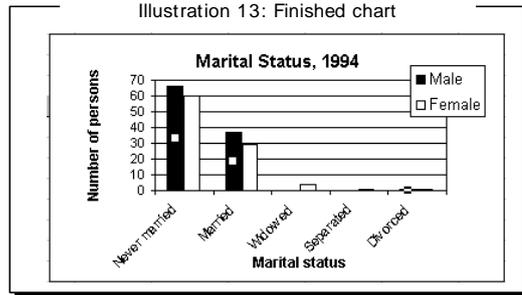
9. If you are not in the Patterns option of the Format Data Series dialog box, click on the Patterns tab. Under Area, click on the **white** colour and press to close the dialog box. Your chart should now look like Illustration 12.

Illustration 12: Changing column colour



10. Change the colour and/or pattern of the other data series to black and white options. Your chart should now look something like Illustration 13.

Illustration 13: Finished chart



One step further ...

Creating pictograms

When you are preparing charts for presentations, sometimes you want to add some 'wow' factors. One effective way of doing this is to create pictograms, where the data series is a clip art image rather than shading.

WARNING



Pictograms are one of the most common ways of using statistics to mislead. Be careful not to distort the area of the image – technically the area of the image represents its frequency. If you create a pictogram use the method here.

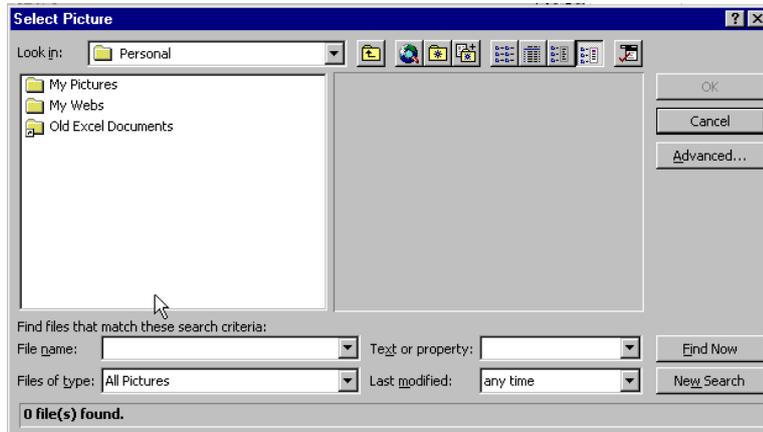
Putting an image in the data series

You decide that you will do a presentation on migration and will use the chart on Nationality and Status, but you want to make it more eye-catching.

1. Double click on the Temporary data series. Click on the Fill Effects button to open the Fill Effects dialog box. Click on the Texture tab to look at some of the other formatting you can apply to the data series.



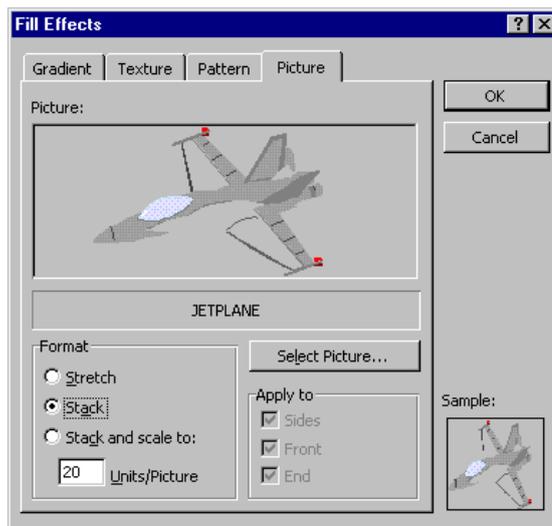
- Click on the Picture tab, and click on the  button. Excel now asks you where the picture is. A dialog box like this will open (the directory you are in will be different from the example here).



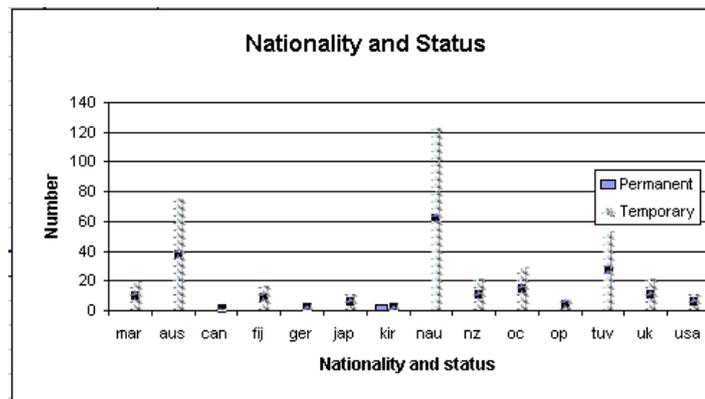
- You have to find your clip art directory. In Office 97 this will be:



- In the Clipart directory you will have a few directories to choose from. Double click on the directory called Popular. When selecting the image, make sure you get a SIMPLE one – if the picture has too much detail the chart will look silly. Also you select a picture which is RELEVANT to your data – baby bottles for fertility, boats for international trade, people for population. Here you would choose a suitcase or aeroplane or something with a similar travel theme. Here select the Jetplane and click OK (try to use a bright colourful picture – this one is a bit dull and grey).



5. In the Fill Effects dialog box make sure the Stack option is selected. Then press OK. Click OK in the Format Data Series dialog box. Your chart should now look like this:



6. The grey in the jetplane means that the picture is too weak. You would change the picture to a brighter one.

