This document contains supplementary information on using `Minitab` to produce graphical and numerical summaries from data. Also included is some guidance on using `Minitab` to calculate probabilities from the normal distribution.

1 Starting Minitab

`Minitab` is a computer package available on most university computers that allows you to analyse data both graphically and numerically.

`Minitab` is run by clicking on

`Start > All Programs > Minitab > Minitab 16 Statistical Software`

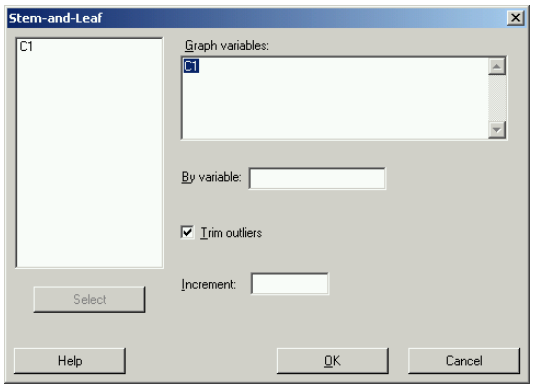
You will see two windows: a session window and a worksheet. Data are entered into columns labelled `C1`, `C2`, `C3`, etc in the worksheet.

2 Stem and leaf plots

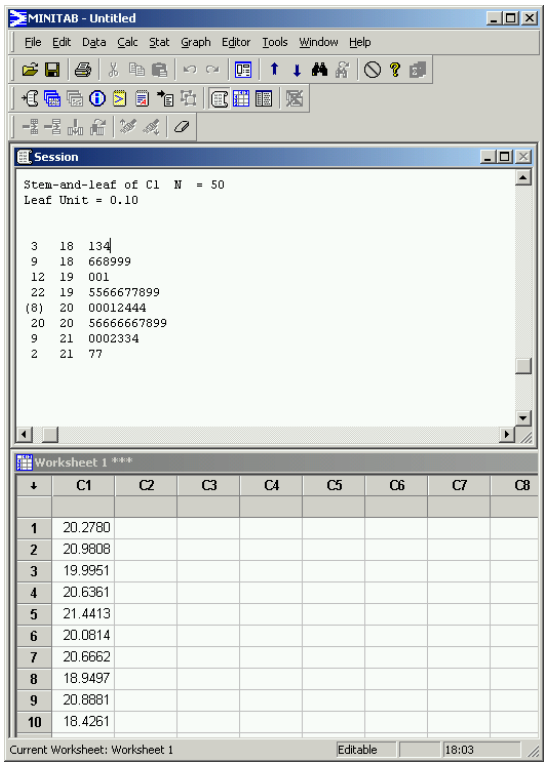
Suppose `C1` contains some data. To obtain a stem and leaf plot of these data you would need to do the following:

`Graph > Stem-and-Leaf...`

This brings up the window below. You need to type in `C1` under `Graph variables` and click `OK`. If you want you can choose the stem unit by entering a value in `Increment` first, otherwise the programme selects this for you.



This creates a stem and leaf plot in the session window:



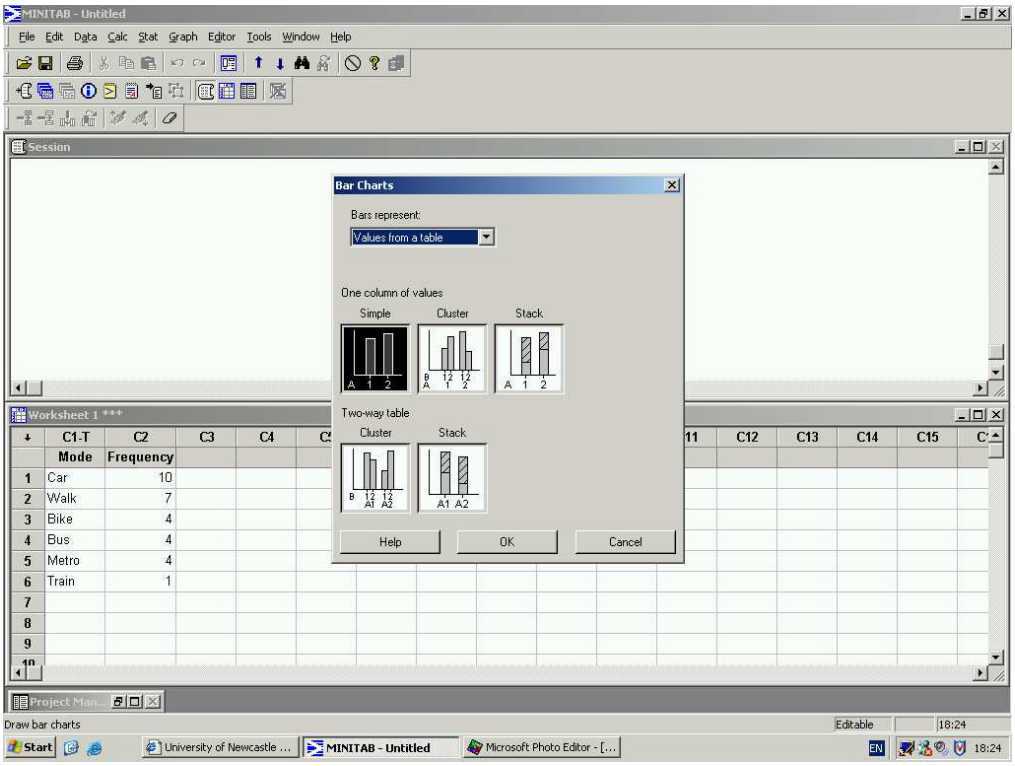
It is easy to see some of the advantages of graphically presenting data. For example, here you can clearly see that the data are centred around a value in the low 20's and fall away on either side. From stem and leaf plots we can quickly and easily tell if the distribution of the data is symmetric or asymmetric. We can see whether there are any **outliers**, that is, observations which are either much larger or much smaller than is typical of the data. We could perhaps even tell whether the data are **multi–modal**, that is to say, whether there are two or more peaks on the graph with a gap between them. If so, this could suggest that the sample contains data from two or more groups.

3 Bar Charts

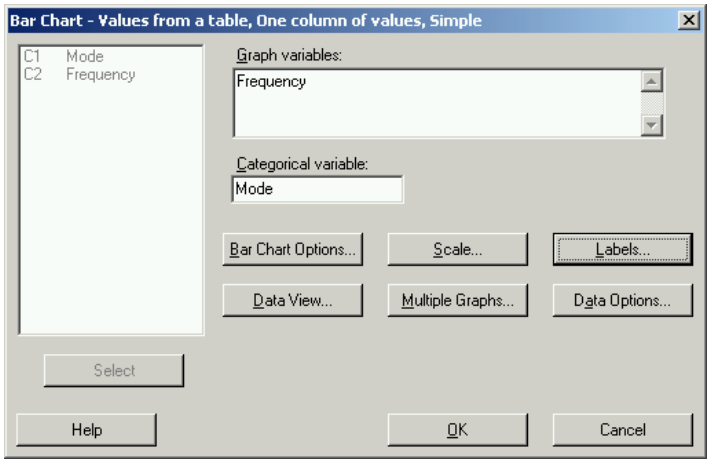
Bar charts are a commonly–used and clear way of presenting categorical data or any ungrouped discrete frequency observations. As with stem and leaf plots, various computer packages allow you to produce these with relative ease.

Bar charts are easily drawn using `Minitab`:

1. First enter the data in the worksheet, either in summary format or as raw data, with column `C1` containing the categories and the (raw or frequency) counts in column `C2`.
2. `Graph > Bar Chart...`

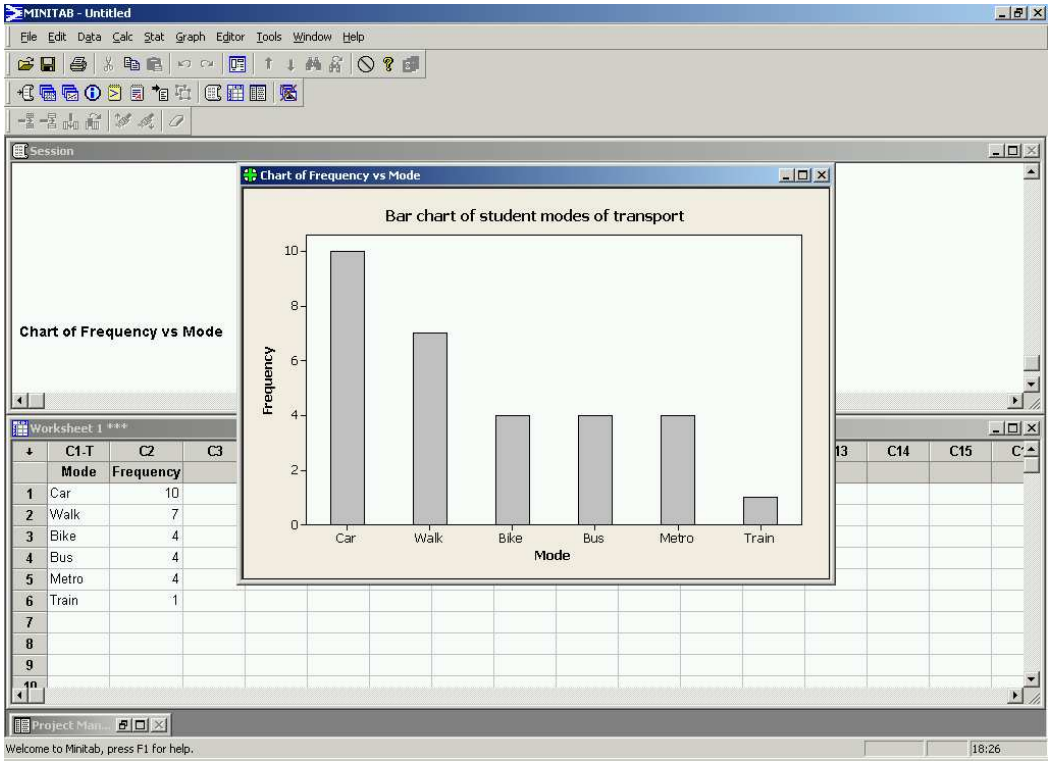


3. Select the appropriate data format (raw data or tabulated data), the columns containing the data, and the graph format.



4. When ready click OK.

This procedure produces the chart



This bar chart clearly shows that the most popular mode of transport is the car and that the metro, walking and cycling are all equally popular (in our small sample). Bar charts provide a simple method of quickly spotting simple patterns of popularity within a discrete data set.

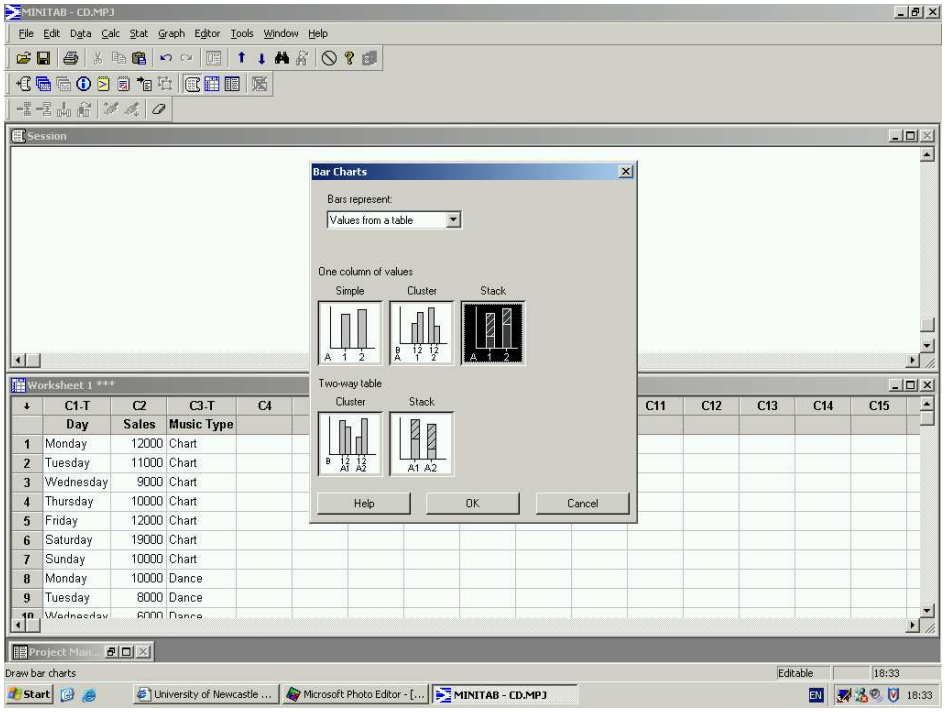
3.1 Multiple Bar Charts

Consider the data below showing daily sales of CDs by music type at an independent retailer:

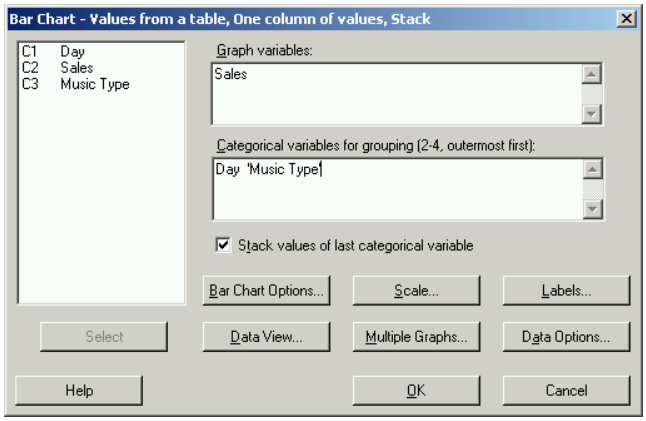
Day	Chart	Dance	Rest	Total
Monday	12000	10000	2700	24700
Tuesday	11000	8000	3000	22000
Wednesday	9000	6000	2000	17000
Thursday	10000	5000	2500	17500
Friday	12000	11000	3000	26000
Saturday	19000	12000	4000	35000
Sunday	10000	8000	2000	20000
Total	83000	60000	19200	162200

Multiple bar charts can be produced in `Minitab` as follows:

1. Enter the data into the worksheet, the types of music in columns and the days as rows.
2. `Graph > Bar Chart...`
3. Select the appropriate data format and the `Stack` graph format.

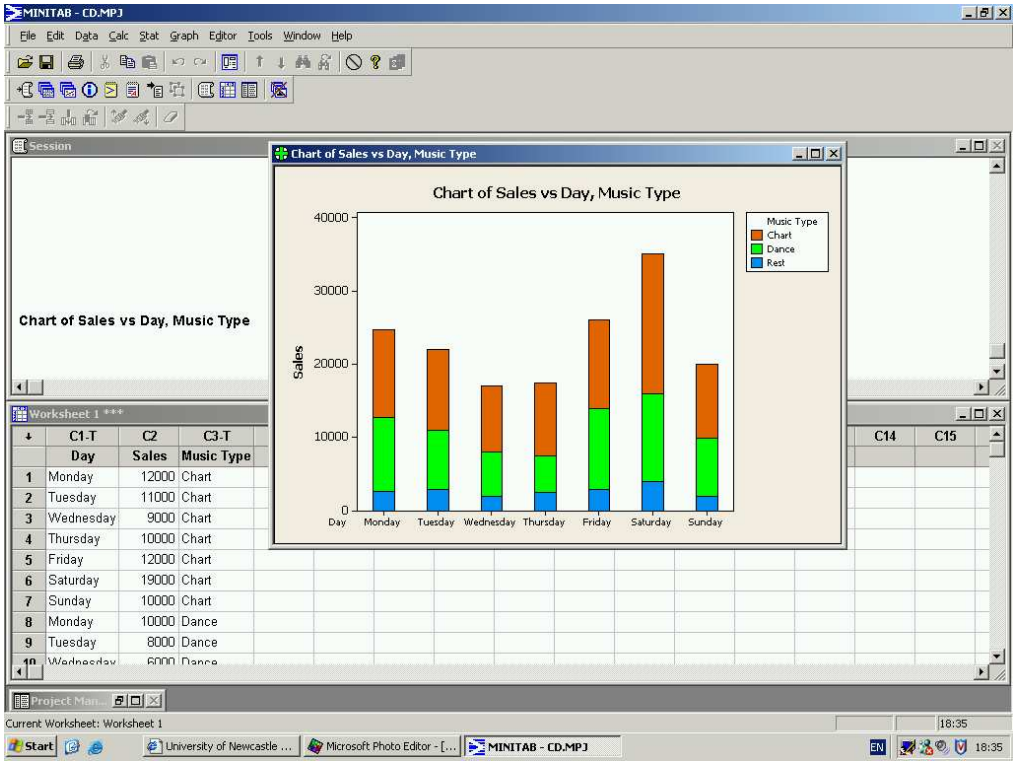


4. Click `OK`.
5. Enter the column containing the `Sales` data under `Graph variables` and the `Day` and `Music Type` in the grouping dialogue box.



6. Click `OK`.

The `Minitab` worksheet and chart this produces are as follows:

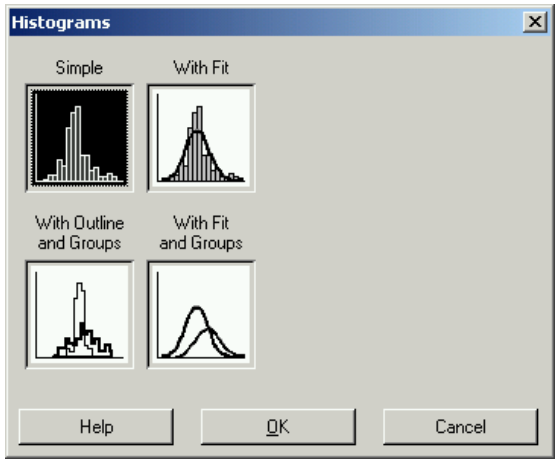


These types of charts are particularly good for presenting such financial information or illustrating any breakdown of data over time – for example, the number of new cars sold by month and model.

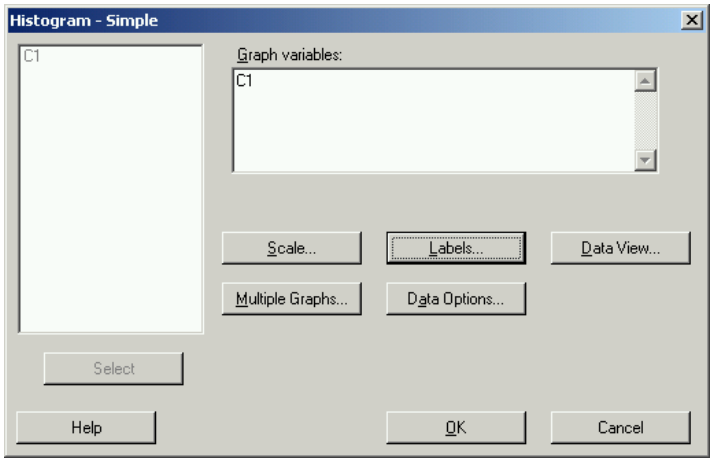
4 Histograms

Normally, as with stem and leaf plots and bar charts, we would get `Minitab` to produce histograms for us, rather than draw them by hand.

1. Enter the data in column `C1` of the worksheet. For illustrative purposes I have randomly generated 500 observations in this column.
2. `Graph > Histogram...`
3. Select the `Simple` graph format.



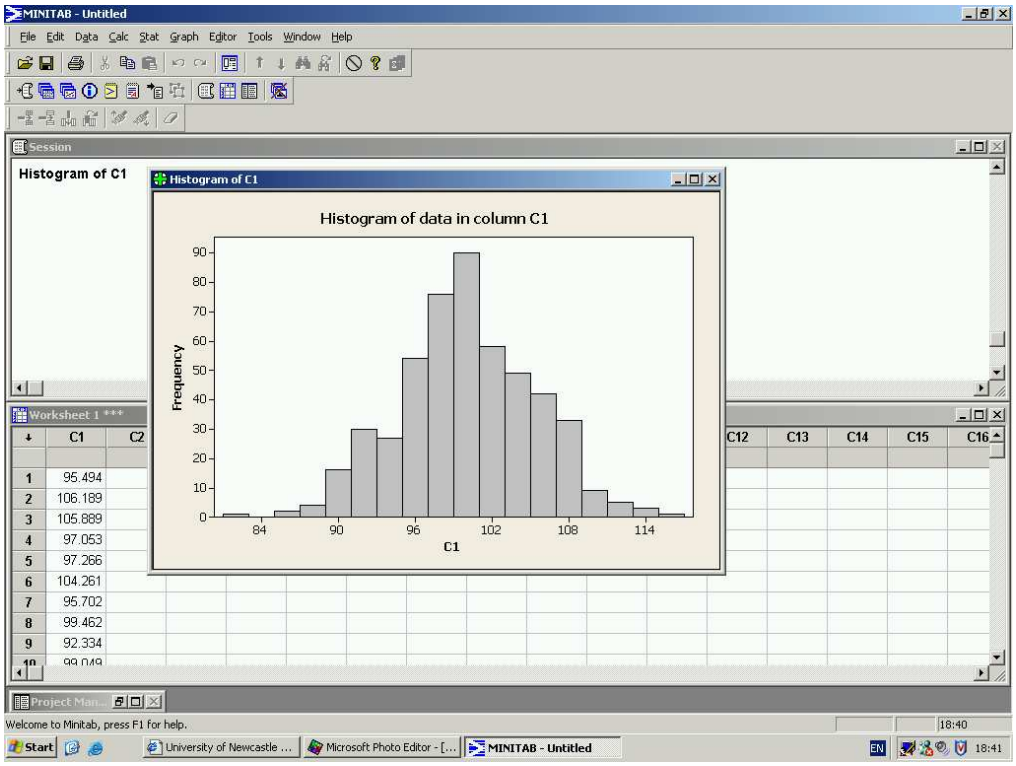
4. Select `C1` under `Graph variables`.



   Note: various advanced options are available e.g. a title can be added by clicking `Labels`

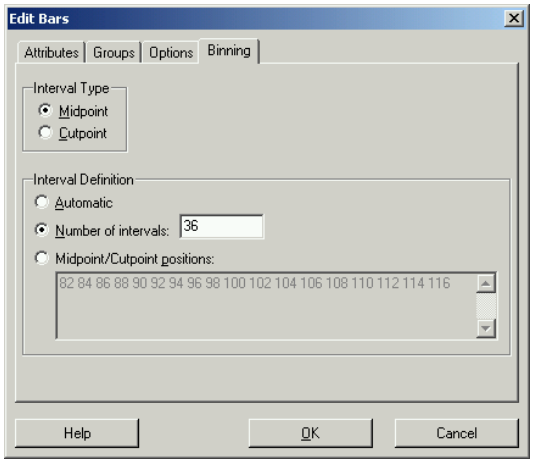
5. When happy with your choices click `OK`.

These instructions produce the following histogram:

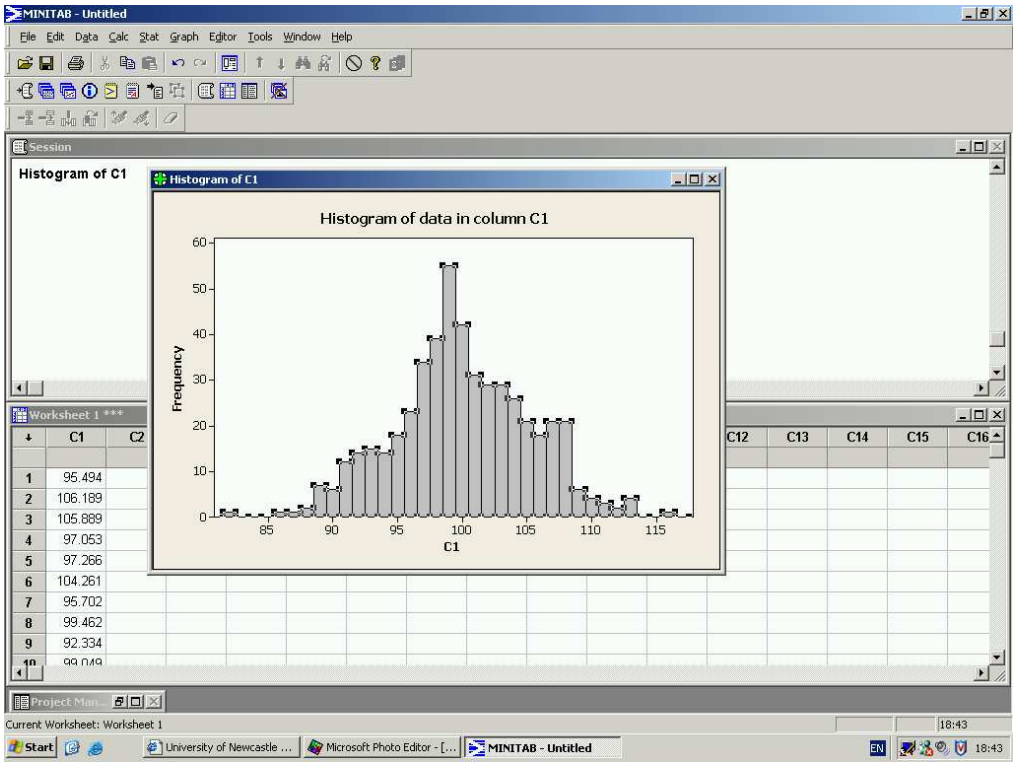


The histogram produced can be amended by `right-clicking` on the graph. For example, the intervals used in the histogram can be changed or, more simply, the number of intervals using `Edit bars > Binning`.

We can double the number of intervals (from 18 to 36 intervals) using the `Binning` dialogue box



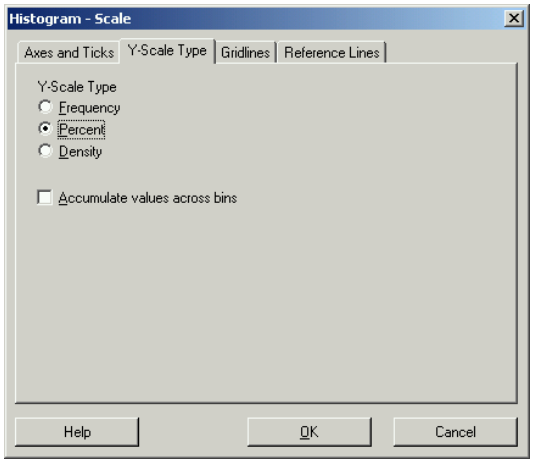
This changes the histogram to



4.1 Percentage Relative Frequency Histograms

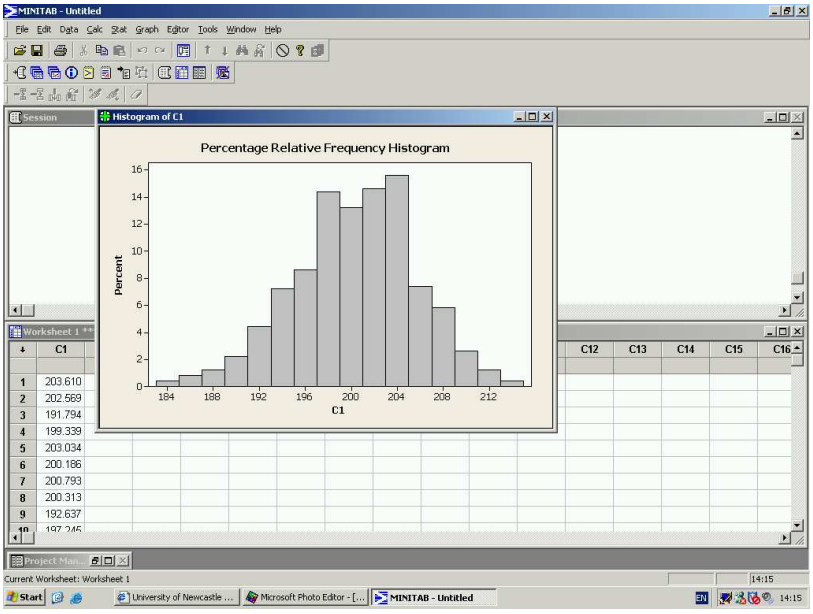
Percentage relative frequency histograms can be produced in `Minitab` as follows:

1. Place the data to be graphed in a column of the worksheet. For illustrative purposes 500 observations have been generated in column `C1`.
2. `Graph > Histogram`
3. As with ordinary histograms, select the `Simple` graph format, click on `OK`, and select column `C1` under `Graph variables`.
4. Select `Scale...` then `Y-Scale Type` and tick the `Percent` button



5. Click on `OK` and again on `OK`.

This produces the following histogram:

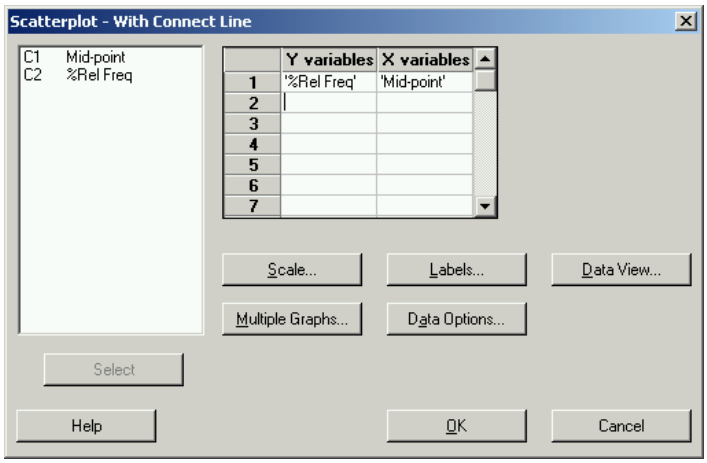


Note that the y-axis now contains the relative percentages rather

5 Relative Frequency Polygons

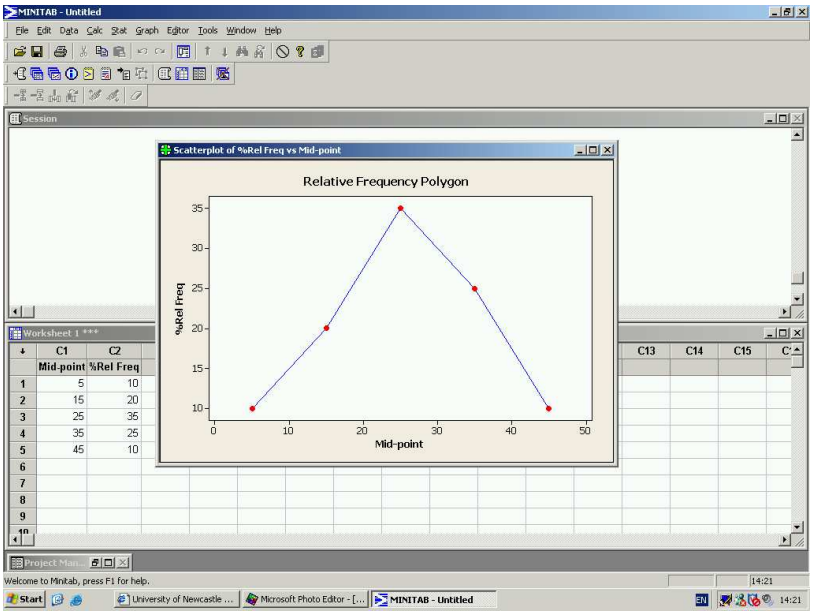
These can be produced in `Minitab` as follows.

1. Place the data in the worksheet using column `C1` for the mid–points and column `C2` for the percentage relative frequencies.
2. `Graph > Scatterplot...`
3. Select the `With Connect Line` option and click on `OK`.
4. Enter the column with the percentage frequencies (`C2`) under `Y variables` and the column with the midpoints (`C1`) under `X variables`



5. Add a title by clicking on `Labels...` etc.
6. Click on `OK`.

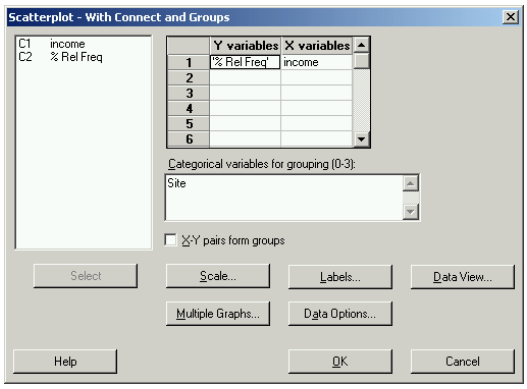
These instructions produce the graph:



These percentage relative frequency polygons are very useful for comparing two or more samples – we can easily "overlay" many relative frequency polygons, but overlaying the corresponding histograms could get really messy! Consider the data on gross weekly income (in £) collected from two sites in Newcastle; see page 74 of the notes.

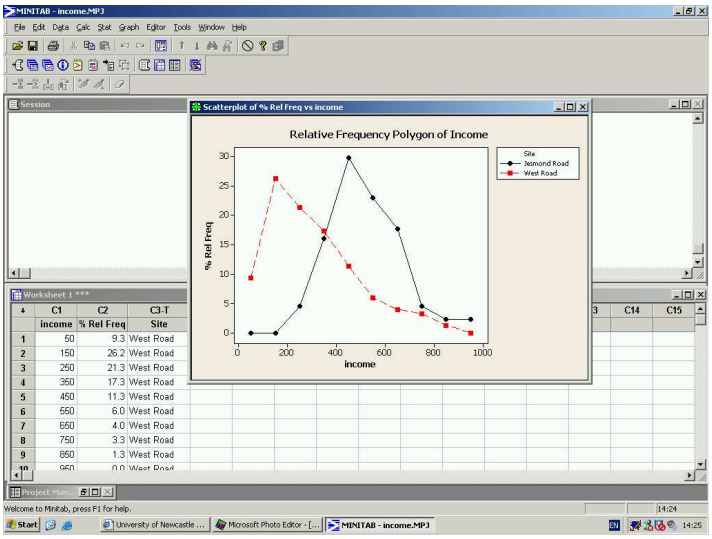
We can produce a graph containing polygons for both locations using `Minitab` instructions very similar to those above:

1. Place the data in the worksheet using column `C1` for the mid–points, column `C2` for the percentage relative frequencies and column `C3` for the site where the data were taken.
2. `Graph > Scatterplot...`
3. Select the `With Connect and Groups` option and click on `OK`.
4. Enter the column with the percentage frequencies (`C2`) under `Y variables` and the column with the midpoints (`C1`) under `X variables`. Also enter the `Site` column (C3) in the box for `Categorical variables for grouping`.



5. Add a title by clicking on `Labels...` etc.
6. Click on `OK`.

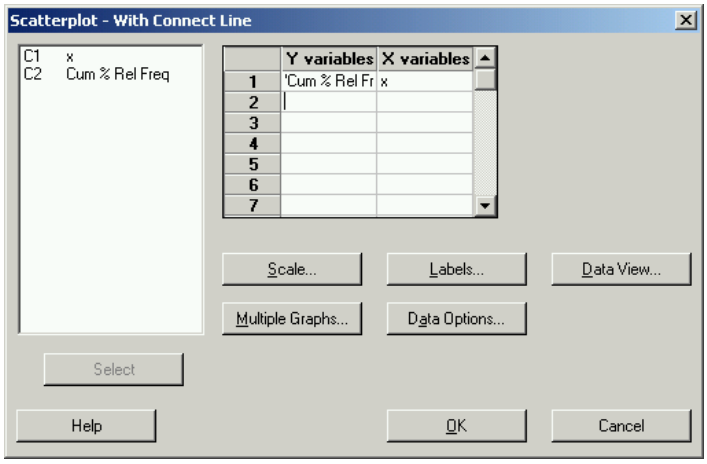
The polygon produced looks like



6 Cumulative Frequency Polygons (Ogive)

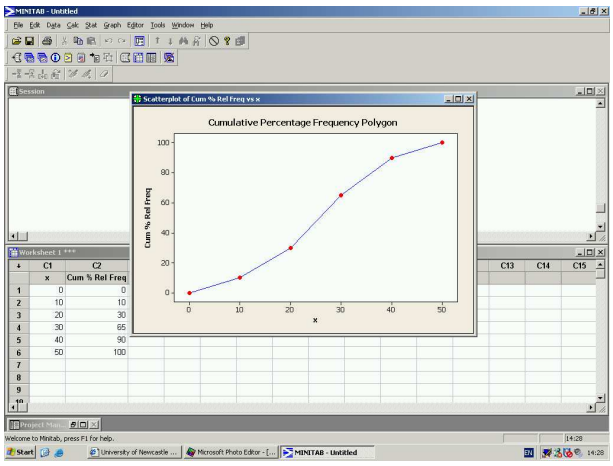
This graph can be produced using the following `Minitab` instructions:

1. In column `C1`, enter the end points of the class intervals, as well as the starting point of the smallest class.
2. In column `C2`, enter 0 against the starting point and the cumulative percentage relative frequencies against the relevant end point.
3. `Graph > Scatterplot...`
4. Select the `With Connect Line` option and click on `OK`
5. Enter the column with the percentage frequencies (C2) under `Y variables` and the column with the midpoints (C1) under `X variables`

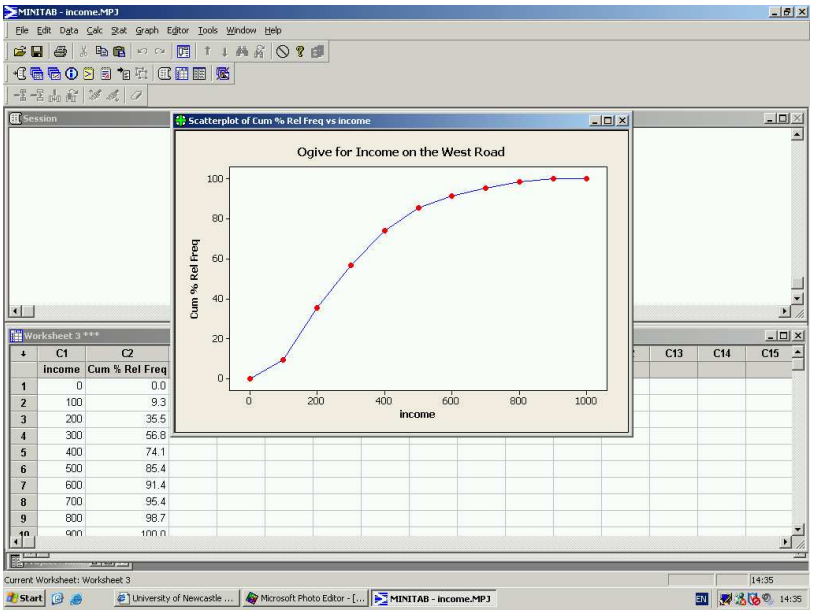


6. Add a title by clicking on `Labels...` etc.
7. Click on `OK`.

This produces the following graph:



Applying this procedure to the income data from the West Road survey gives the ogive:

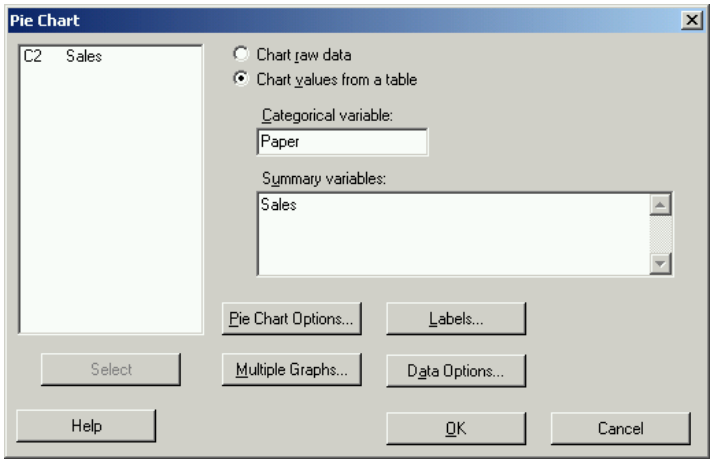


7 Pie Charts

Consider the data on newspaper sales to 650 students that were presented in question 6, Section 3.6 of the notes.

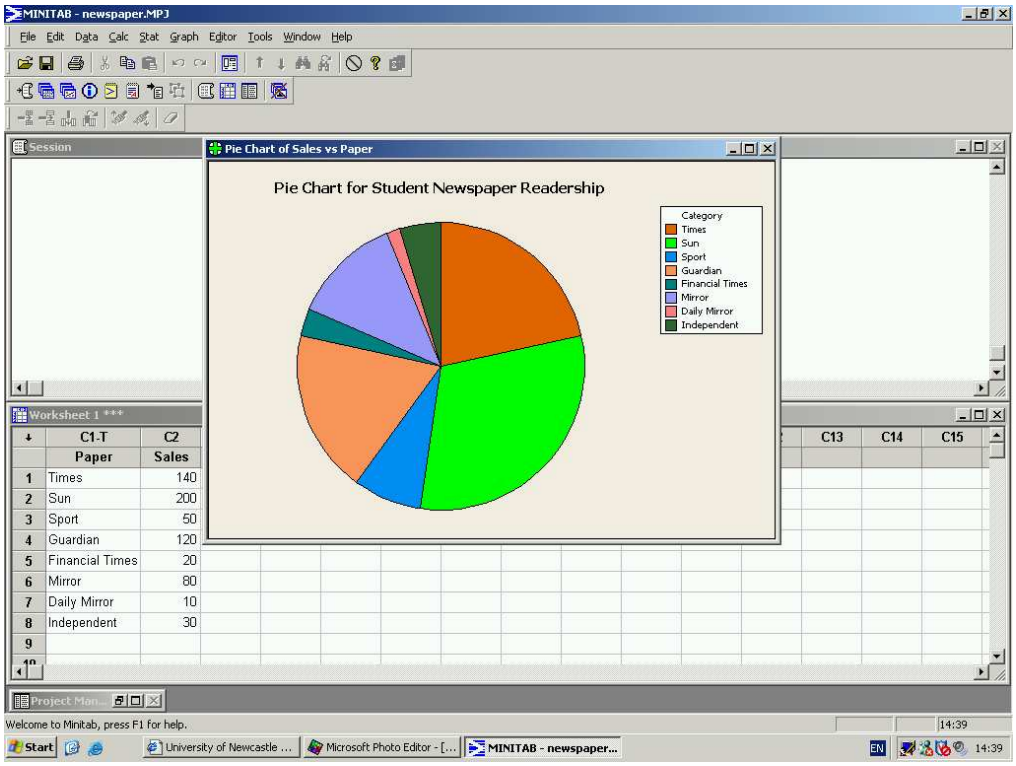
In `Minitab`, a pie chart for these data would be obtained as follows:

1. Enter the data into a worksheet, with category name in column `C1` and frequencies in column `C2`.
2. `Graph > Pie Chart...`
3. Tick the button for `Chart values from a table`
4. Enter the `Category` column under `Categorical variable:` and the `Frequency` column under `Summary variables:`



5. Add a title and click OK.

This produces the following pie chart:



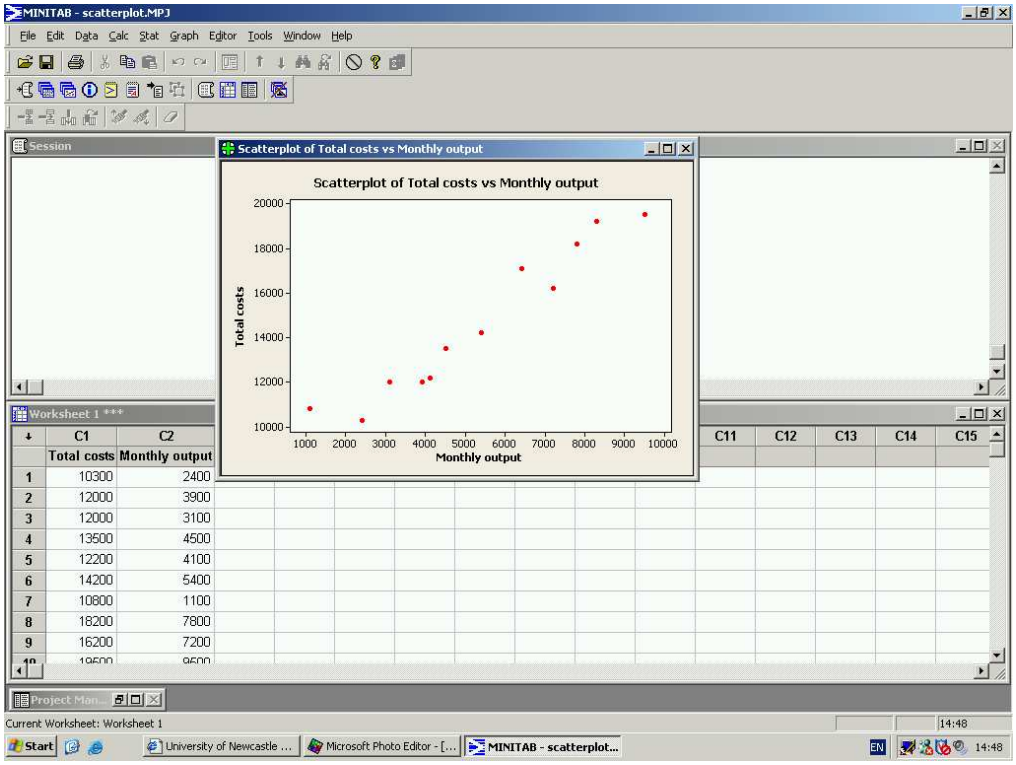
It shows that The Sun, The Times and The Guardian are the most popular papers.

Note that the pie chart is a simple circle. Some computer software will draw “perspective” pie charts, pie charts with slices detached etc. It is best to avoid such gimmicks which merely obscure the information contained in the chart.

8 Scatter Plots

Consider the data for monthly output and total costs at a factory that were given in Section 3.4.8 of the notes.

If you were interested in the relationship between the cost of production and the number of units then a scatter plot can be produced using `Minitab` (Select `Graph` then `Scatterplot` then `Simple` and insert the required variables).

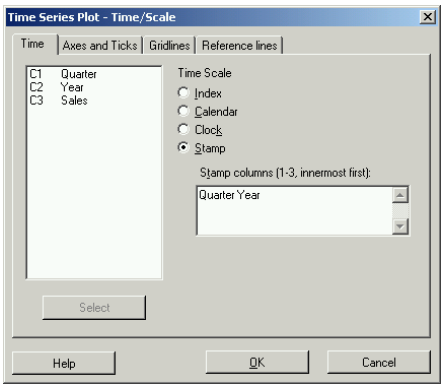


9 Time Series Plots

Consider the data on the number of computers sold (in thousands) by quarter (January-March, April-June, July-September, October-December) at a large warehouse outlet that were given in Section 3.4.9 of the notes.

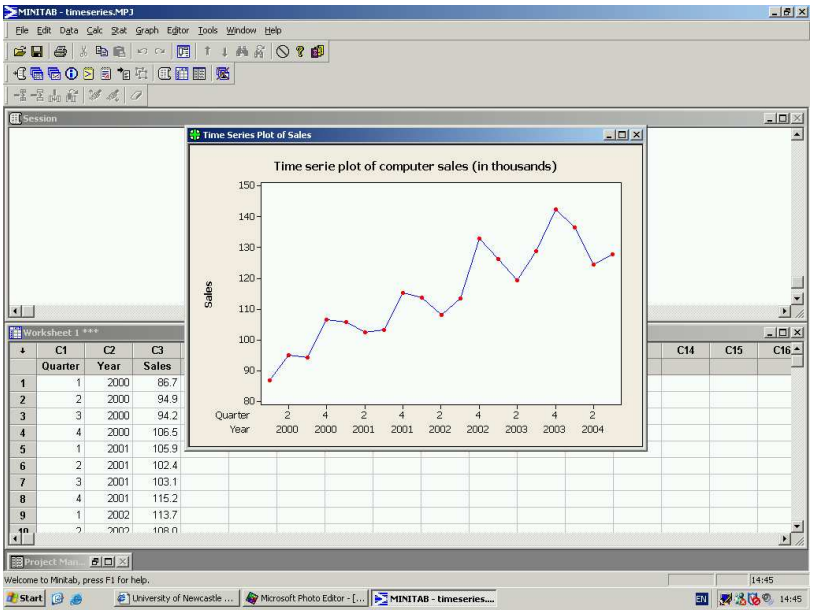
In Minitab a time series plot can be obtained using:

1. Enter the data into a worksheet, with the Quarter, Year and Sales in columns C1, C2 and C3.
2. Click on Graph and select Time Series Plot...
3. Select the Simple graph format and click on OK.
4. Enter the Sales column in the Series: box.
5. Now click on Time/Scale..., check the Stamp button and enter the Quarter and Year columns under Stamp columns



6. Click OK.
7. Add a title etc.
8. Click OK.

The time series plot is:

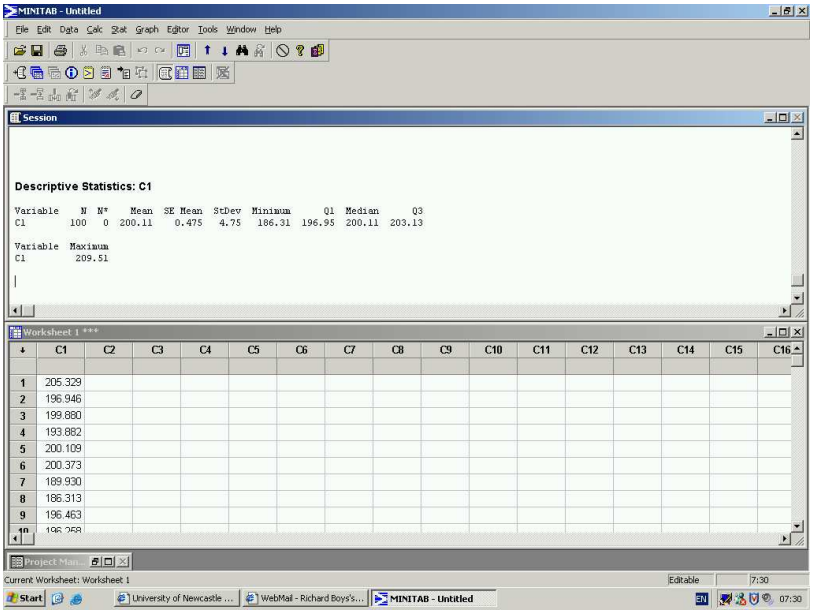


10 Summary statistics in Minitab

Minitab can be used to calculate all of the basic numerical summary statistics covered in Chapter 3. These summaries for data in a selected column can be obtained using the commands

```
Stats > Basic Statistics > Display Descriptive Statistics
```

The results are output in the session window as follows:

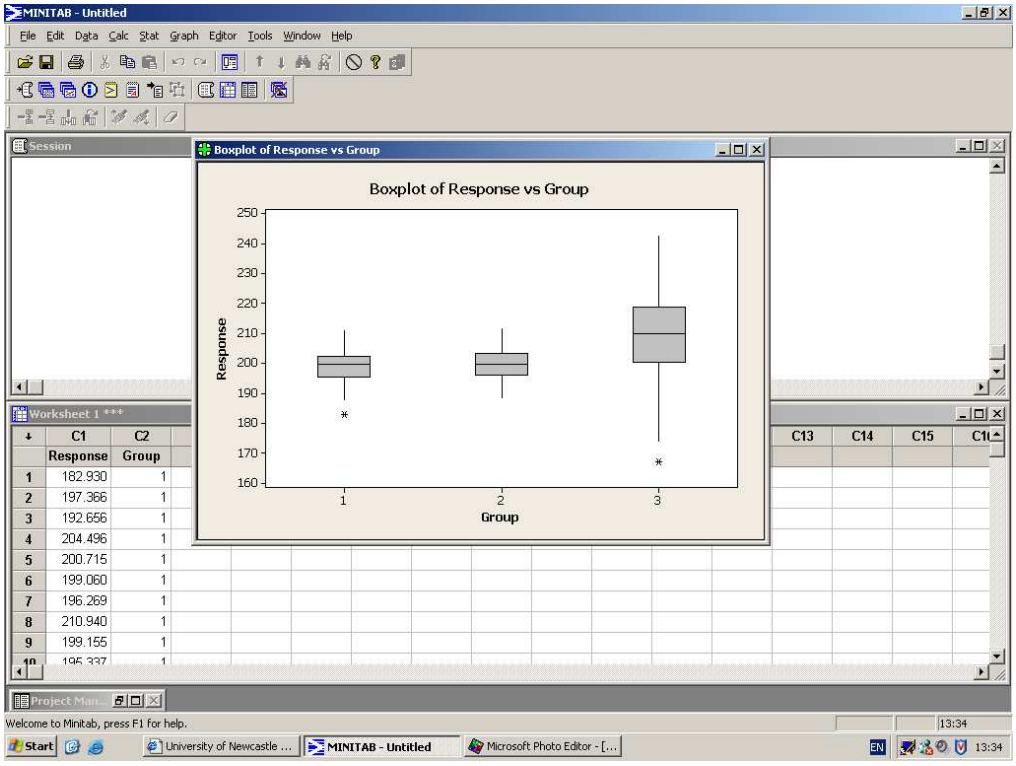


11 Box plots

`Minitab` will produce box plots using the following commands.

1. Enter the data into the worksheet, say column C1
2. `Graph > Boxplot...` and select the `Simple` graph format
3. Next enter the column containing the data under `Graph variables:`
4. Add a title using `Labels...`
5. Click on `OK`.

If the data have subgroups, such as results from three different surveys, then box plots of the sample data can be plotted by group by first entering the group variable into the worksheet, say as column `C2`, and then selecting the `With Groups` graph format. The group variable is then entered into the subsequent dialogue box under `Categorical variables for grouping`. Displaying group structure is one of the main uses of box plots. For example:



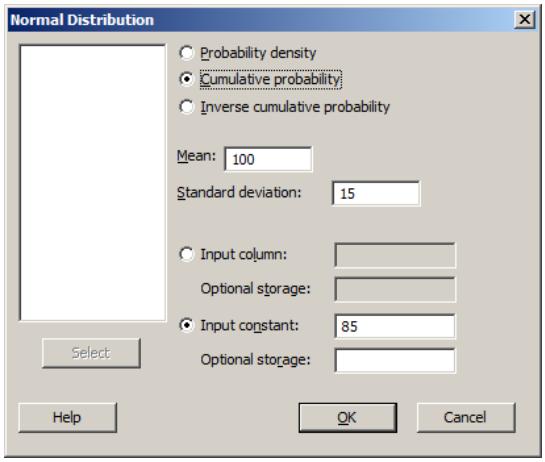
clearly shows that although there is overlap between the three sets of data, the first and second datasets contain roughly similar responses and that these are quite different from those in the third set. Note that the asterisks (*) at the ends of the whiskers is the way `Minitab` highlights outlying values.

12 Calculating probabilities from the normal distribution

`Minitab` can be used to calculate normal probabilities. The following commands will calculate probabilities $P(X < x)$ and also values of x that satisfy $P(X < x) = p$:

1. `Calc > Probability Distributions > Normal`

   opens up dialogue box



2. Select `Cumulative probability` for $P(X \leq x)$ or `Inverse cumulative probability` for the value of x satisfying $P(X \leq x) = p$
3. Enter the `Mean` (μ) and the `Standard Deviation` (σ)
4. Select `Input Constant` and enter the value for x or p (as appropriate)
5. Click `OK`
6. The answer is displayed in the Session Window: