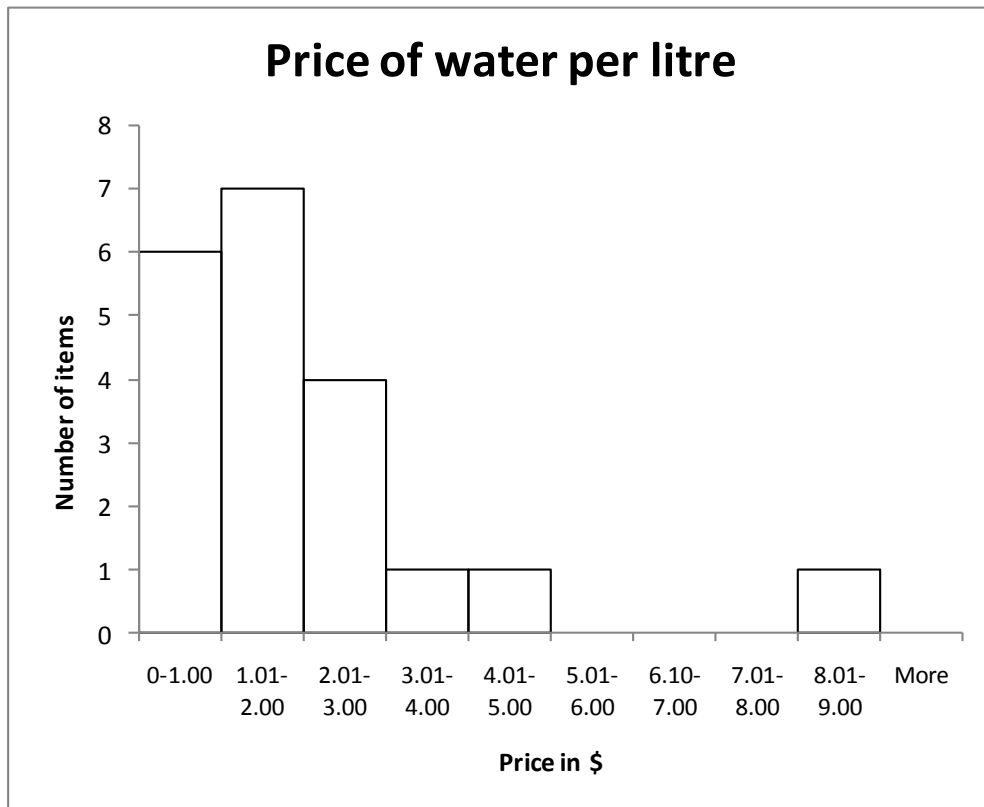


Using Excel to create a histogram

Open up the Excel workbook containing the data. We will use water.xls for this example. It gives the price per litre for different brands of water tested by Consumer in Nov 2005. You want to produce a histogram that looks like this:

Figure 1: The intended product

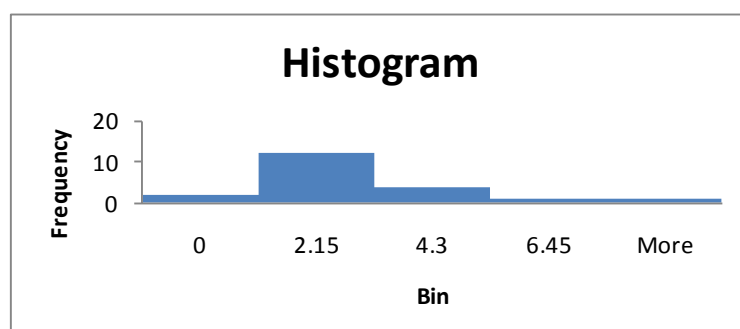


Features of a good histogram

- The bars are touching.
- There are between 5 and 15 bars.
- The ranges that the bars include are clear (eg 1.01-2.00)
- The heading is informative and the axis titles are also correct
- There is no shading.

This is much more difficult than it sounds, as Excel, left to default values, will create an incorrect histogram.

Figure 2: What Excel will produce if left to its own devices! (This is the same data as in Figure 1.)



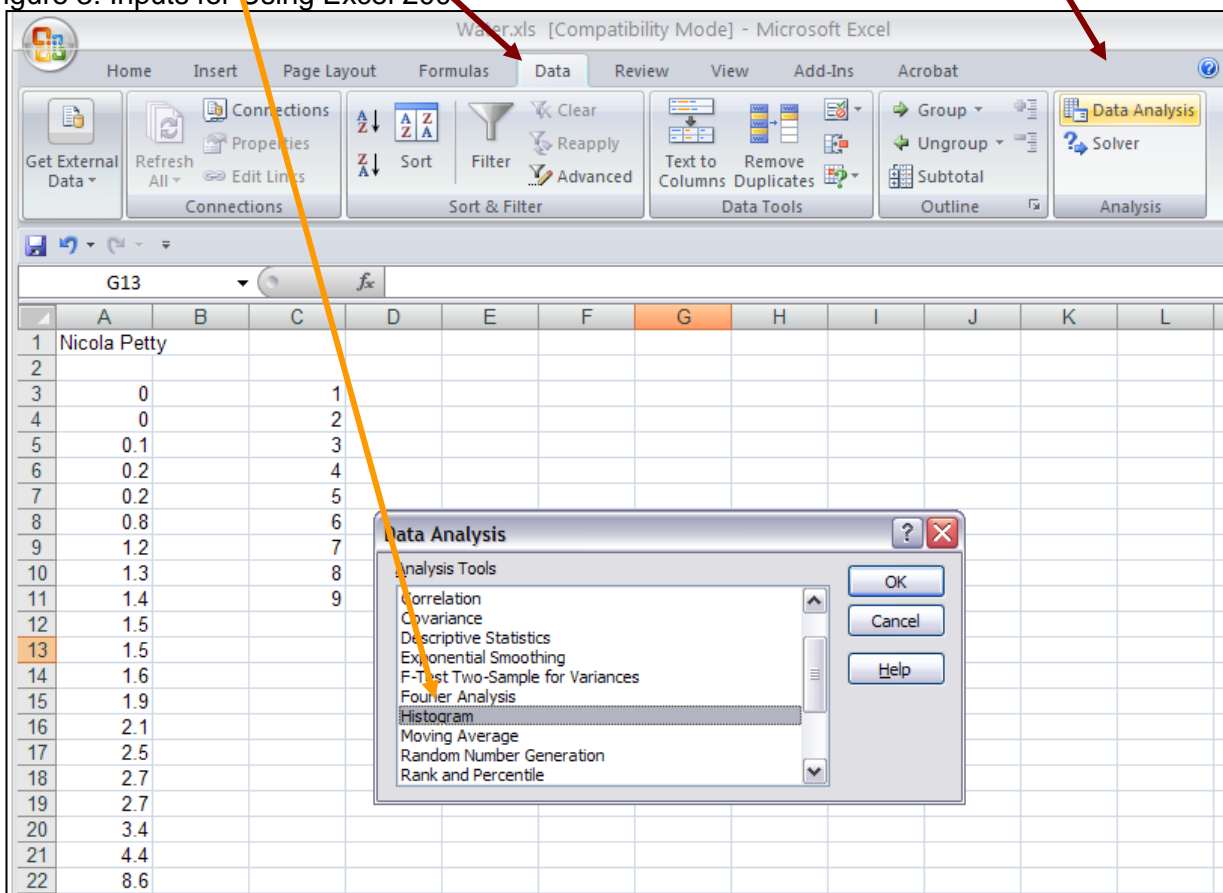
Step 1 – Identifying the ‘Bin Range’ – (see Figure 7 to see the effect of differing bin sizes.)

- The aim is to have between **5 and 15** equal sized intervals, generally of integer width. (Eg 0-5, 6-10, 11-15 etc, not 0-2.33333, 2.33334 – 4.6666 etc). Fewer observations will require fewer bins.
It is common to decide to change bin ranges after the first attempt.
- Highlight all the data (in column 1)
- Use ‘Sort & Filter’ in the Editing section of the Home ribbon to sort column 1 into ascending order to get an idea of the data.
- By doing this you can see the smallest value is 0 and the biggest is 8.6, so using an interval width of 1 puts all the data into 9 intervals of equal width.
- The numbers you give Excel are the inclusive upper bound. Thus, next to the number 3 will be a count of all the values that are less than or equal to 3 but more than 2.
- I chose intervals of width 1, so enter 1,2,3...9 in column C. (You can enter the first two values and use the fill tool to make it quicker.) These are the Bin Ranges.

Step 2 – Using the Excel Tool-Pak to create Histogram

1. Choose the Data ribbon, and at the right hand end should be Data Analysis.
If Data Analysis is not listed there, you will need to add it using the Windows button, Excel Options, Add-ins, Manage Excel Add-ins, then check Analysis ToolPak.
2. Select ‘Histogram’ and click ok.

Figure 3: Inputs for Using Excel 2007



3. Go to Input Range and click in the box next to it.
4. Highlight all your data, but not the label (if there is one).
Shortcut: Select the top value in your data (probably A3) and hold down Ctrl,Shift, then press ↓(Down arrow)
5. Do the same with 'Bin Range'
6. Click in the Output Range input box and select a cell in a clear part of the spreadsheet.
7. Make sure Chart Output is selected with a tick.
8. Click on 'ok'
9. You should see a frequency table and a chart output. (See Figure 5.)

Figure 4: Inputs for a Histogram.

| | A | B | C | D | E | F | G | H | I | J |
|----|--------------|---|---|---|---|---|---|---|---|---|
| 1 | Nicola Petty | | | | | | | | | |
| 2 | | | | | | | | | | |
| 3 | 0 | | 1 | | | | | | | |
| 4 | 0 | | 2 | | | | | | | |
| 5 | 0.1 | | 3 | | | | | | | |
| 6 | 0.2 | | 4 | | | | | | | |
| 7 | 0.2 | | 5 | | | | | | | |
| 8 | 0.8 | | 6 | | | | | | | |
| 9 | 1.2 | | 7 | | | | | | | |
| 10 | 1.3 | | 8 | | | | | | | |
| 11 | 1.4 | | 9 | | | | | | | |
| 12 | 1.5 | | | | | | | | | |
| 13 | 1.5 | | | | | | | | | |
| 14 | 1.6 | | | | | | | | | |
| 15 | 1.9 | | | | | | | | | |
| 16 | 2.1 | | | | | | | | | |
| 17 | 2.5 | | | | | | | | | |
| 18 | 2.7 | | | | | | | | | |
| 19 | 2.7 | | | | | | | | | |
| 20 | 3.4 | | | | | | | | | |
| 21 | 4.4 | | | | | | | | | |
| 22 | 8.6 | | | | | | | | | |
| 23 | | | | | | | | | | |

Figure 5: Raw output from Excel



Step 3 – Fixing up the output.

You need to make the columns touch, remove legend, add titles and correct the labels on the horizontal axis.

- **Size:** Make the chart the size you wish it to finish up at by dragging the handles on the edges of the frame.
- **Legend:** Remove the 'Legend' by clicking on it and then pressing Delete.
- **Titles:** Use the Layout ribbon to add meaningful and correct Chart and Axis titles.
- **Column width and colour:** Right click in a column to open the Format Data Series dialog. (Or go to the Format ribbon, and select Series Frequency, then Format selection).

In the Series Options Tab, make the Gap width 0%.

In the Fill Tab select No fill.

In the Border Colour Tab, select Solid line (and Black).

Close.

- **Correct labels:** Change the text labels in the table under "Bins". In this example the value 2 will be replaced by 1.01 – 2.00, as the value given is the inclusive upper bound. This is time-consuming, but **VERY important** to make the graph correctly interpretable. If you find a date appearing, you need to change the format to Text then try again.
- **Format Axis labels:** If the labels you have just added are a bit long, you may need to adjust them. On the graph, right click on one of the axis labels and select Format Axis.

In the Axis Options Tab, Specify interval unit: 1. This will make the text wrap if needed.

In the Alignment Tab, make the Custom angle 0.

Change the size of the font, by right clicking an axis label and selecting Font.

You may need to do this to the vertical axis as well to make it look good.

- **You have now created a good quality histogram in Excel.**

| Bin | Frequency |
|-------------|-----------|
| 0-1.00 | 6 |
| 1.01 - 2.00 | 7 |
| 2.01 - 3.00 | 4 |
| 3.01 - 4.00 | 1 |
| 4.01 - 5.00 | 1 |
| 5.01 - 6.00 | 0 |
| 6.01 - 7.00 | 0 |
| 7.01 - 8.00 | 0 |
| 8.01 - 9.00 | 1 |
| More | 0 |

Figure 6: Good Histogram produced using Excel 2007.

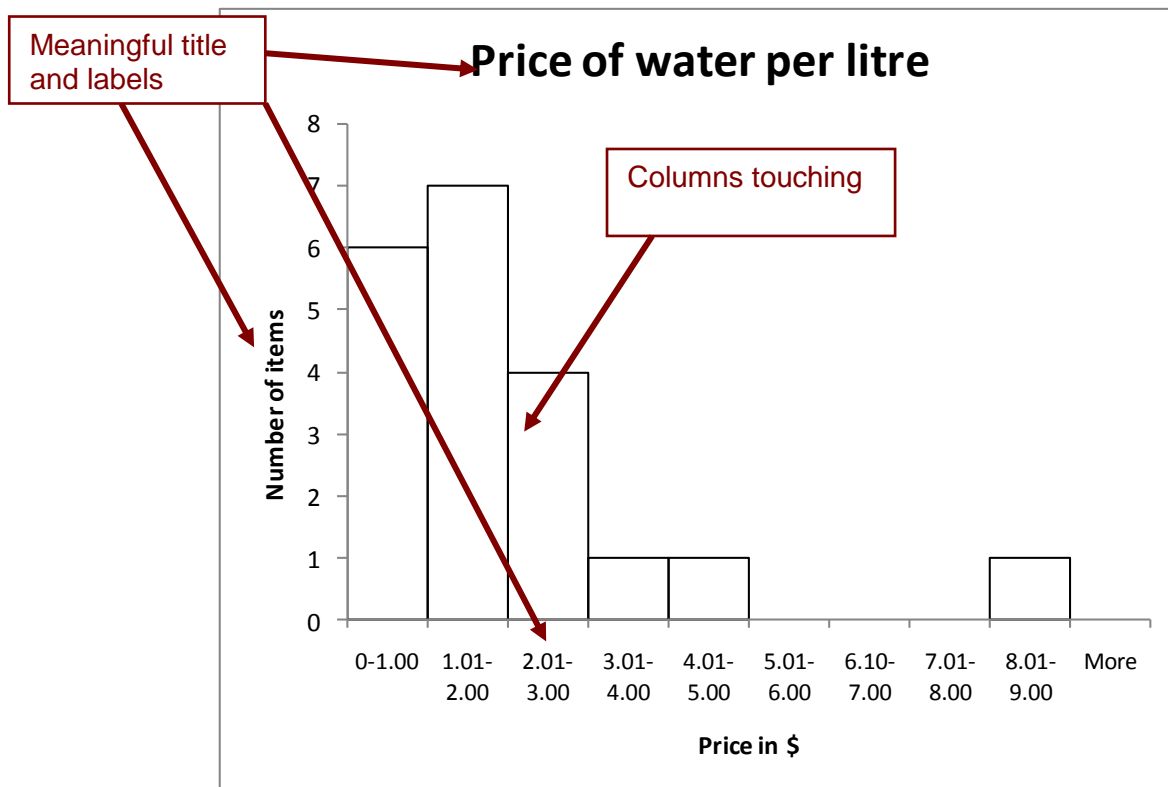
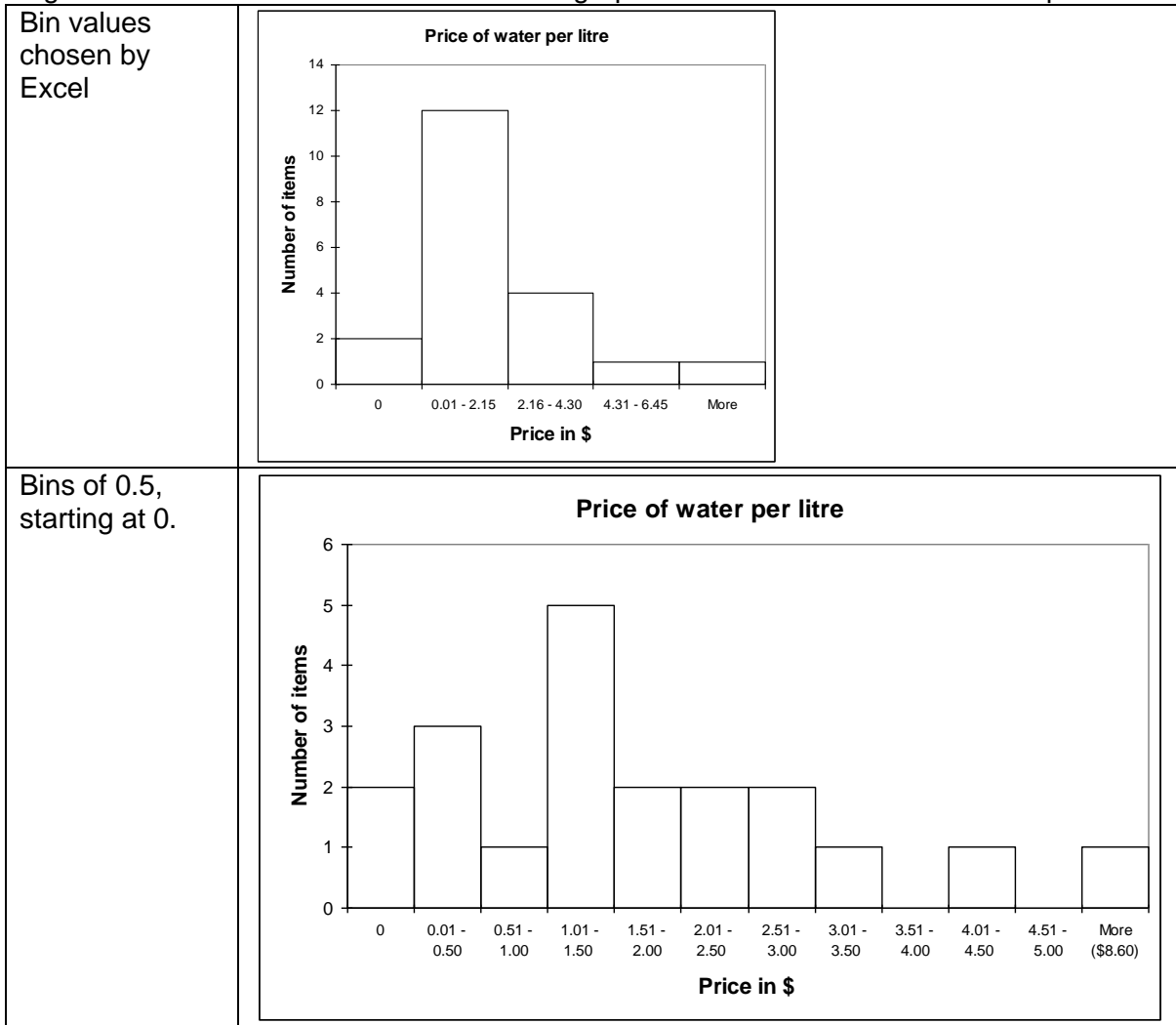


Figure 7: Illustrates the effect of bin size on graphs of the same data as the example.



Nicola Petty
 Statistics Learning Centre 2010
www.statslc.com

Data from
<http://www.consumer.org.nz/topic.asp?docid=2422&category=Food&subcategory=Drink&topic=Bottled%20water&title=Products%20compared&contenttype=general> accessed Jan 09.