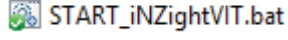
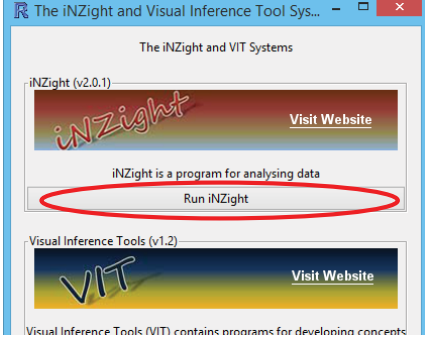
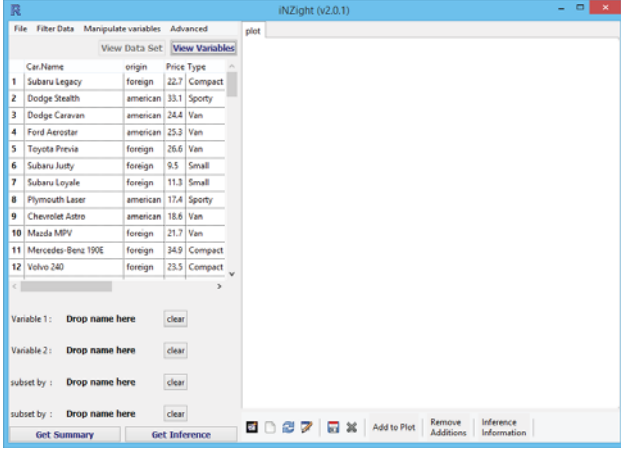
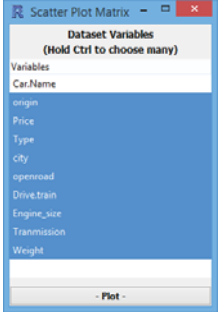
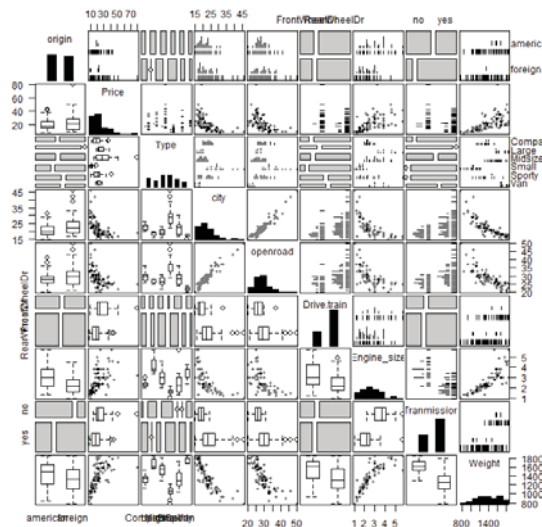


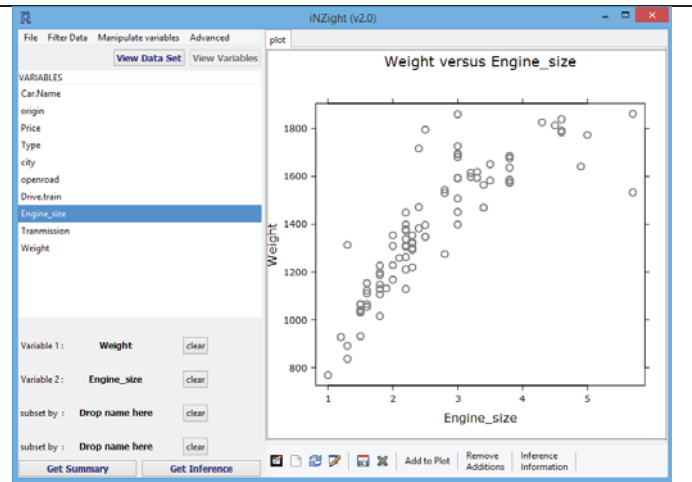
Data – Using iNZight

The next section that we need to do is the data section. This is reproducing the graphs on Page 2 using iNZight, as well as a few other things. The example below will go through using the Rugby dataset for weight by position.

<p>First up we need to start iNZight by clicking on the shortcut in the iNZight folder that looks like this.</p>	
<p>After it has had some time to think it will open up a window that looks like this.</p> <p>To start off with we need to open the iNZight window by clicking on the 'Run iNZight' button (circled)</p>	
<p>This will bring up the main iNZight window that looks like this. We then want to import the data, by clicking on 'File' and 'Import Data'. Browse for the right file and follow the prompts, and once imported it should look like this.</p>	
<p>Often it is useful to get an overview of the whole dataset. To do this go to 'Advanced' and 'Quick Explore' and 'Pairs'.</p> <p>This will give the window shown on the right where I have selected all of the variables apart from the name of the car.</p>	
<p>Pressing plot will give us a graph of all of the possible combinations of variables like the one show on the right.</p> <p>From this we can work out what there is a relationship between, and what variables might be appropriate to use.</p>	

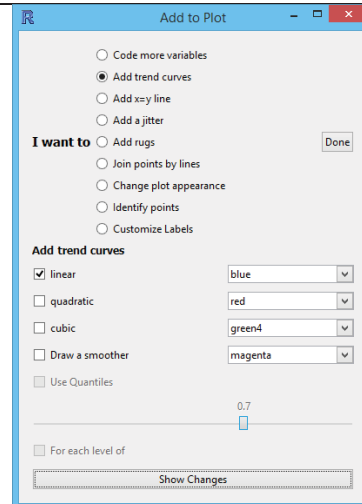
Once we have determined the variables we would like to investigate you drag the title of each variable onto 'Variable 1' and 'Variable 2'. Your dependent variable goes onto variable 1 and your independent variable goes onto variable 2.

Note: you can make it easier to see all the variables by clicking on the 'View Variables' button.



You also need to add a trend curve to the plot (we normally use a linear trend). You do this by clicking add to plot, and selecting add trend curve. Tick the box next to linear as shown to the right and click the 'Show Changes' button at the bottom. You can now close the window.

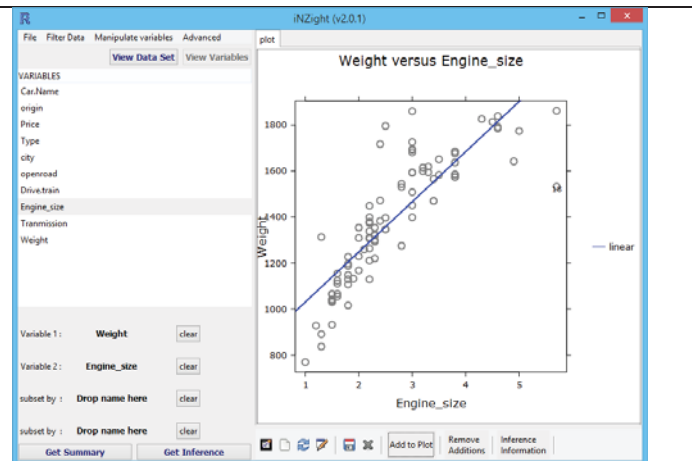
Note: You can now easily identify outliers... if you click on 'identify points' then 'locate' if you click on the point it will display the information next to that point on the graph... by default it displays the id which you can use to look up the data point in the list, but you can also get it to show any of the other variables... see example on the graph in the next row where point 18 is identified. You can still click on the point even if the mouse cursor is still 'spinning'.



The final thing you need to do is get the summary, you get this by pressing the 'Get Summary' button.

Note: you can make it easier to copy the graph by clicking on the 'New' button under the plot which will open the graph in a new window which can be right clicked on to copy. Bitmaps normally work best when copying into a word document, as metafiles sometimes will lose some of the plotted points. (Right clicking on the main window and pressing copy doesn't work)

If you want to change the title of the graph or the axis labels this can be done using the 'Add to Plot' button and the 'Customise Labels' option. It is always a good idea to change the labels to be a good description and include the units.



$$\text{Weight} = 217.97 * \text{Engine_size} + 812.32$$

$$\text{Correlation} = 0.8451$$

Now it is your turn. For each dataset you need to produce:

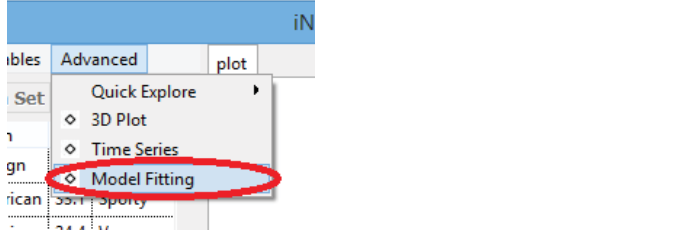
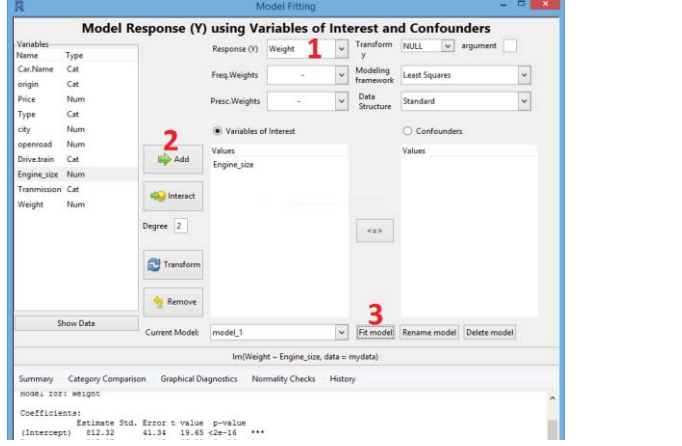
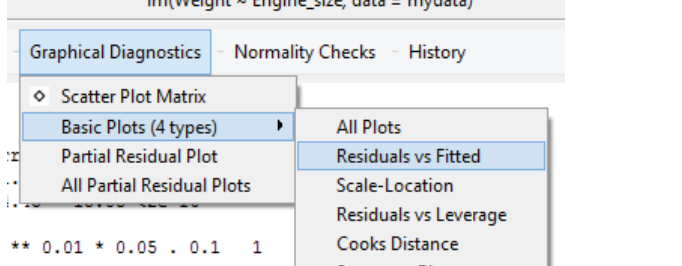
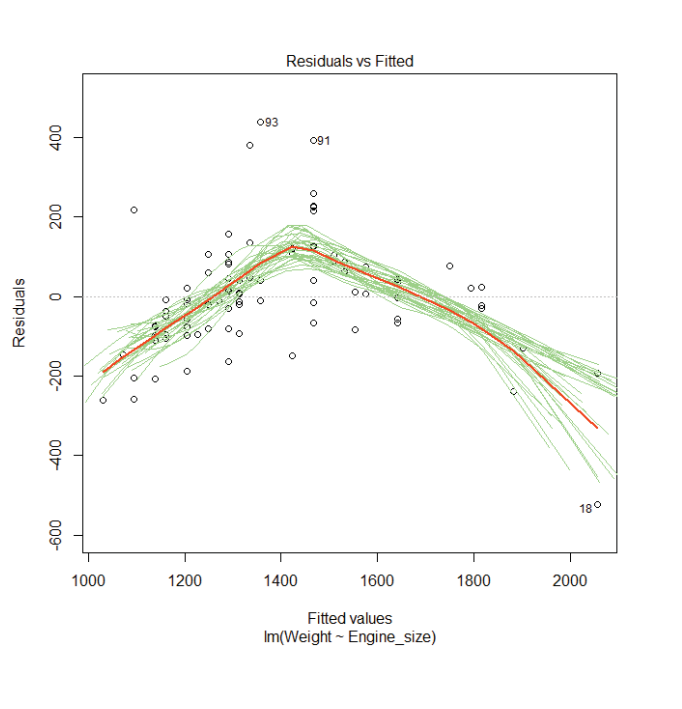
- The scatter plot.
- The equation of the trend (summary).

The scatter plots are at the front of the booklet, and the dataset summaries are included as an appendix so you can check your answers

Note: The axis titles given on iNZight are not sufficient, so once inserted into your document you will need to put a text box over the top of them to include a proper title and axis labels (including units)

Residuals – Merit Level Skill

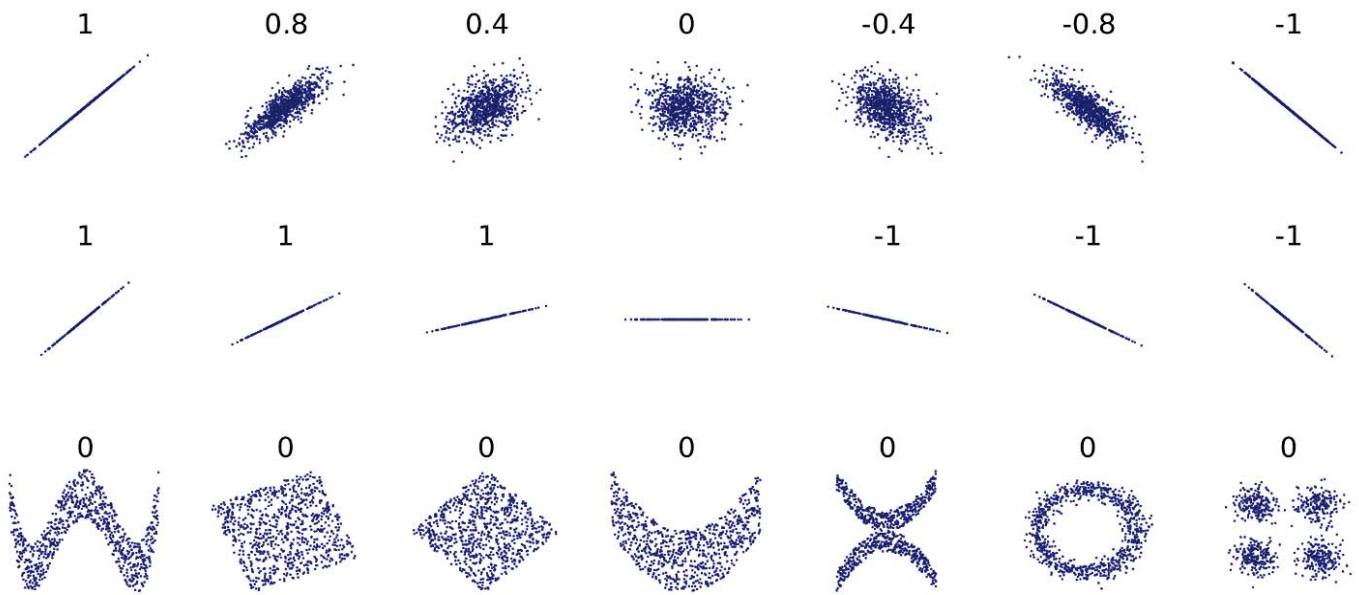
One of the ways that we can analyse how well our model fits the data and therefore how accurate our predictions are is by looking at the residuals. We can create a plot of the residuals using iNZight.

<p>Once the data has been loaded in you need to go to the 'Model Fitting' option under the 'Advanced' menu.</p>	
<p>You need to add the dependent variable (or the one you are predicting) into the response variable (Y) selector box. (1)</p> <p>Under variables you need to select the independent variable (or the variable you are using to predict) press the 'Add Variable' button (2).</p> <p>Then press the 'Fit Model' button. (3)</p> <p>This will display an output similar to the one shown on the right, and will fit the same linear model that was fitted earlier.</p>	
<p>You then need to go to 'Graphical Diagnostics' → 'Basic Plots' and click on 'Residuals vs. Fitted'</p>	
<p>This will give the output shown to the right.</p> <p>You need to insert this into your report and you can use this to analyse how accurate your predictions are.</p> <p>The flat dotted line is where your predictions will be, and the red (dark centre line) and green lines (light lines either side) indicate the range of values that you would expect the trend to take (using a weighted average and bootstrapping).</p> <p><i>For excellence you can also think about if another model might fit the data better or if there is any outliers that are adversely affecting the model and remove them (as long as you can justify why you should remove them).</i></p>	

Appendices

Examples of Correlation Coefficient (r)

Below are graphs showing the different correlation coefficients (r values) above each one.



Source: http://upload.wikimedia.org/wikipedia/commons/d/d4/Correlation_examples2.svg

The important thing to note here is that correlation coefficients, or r-values, only tell you about **linear** relationships. If the relationship is non-linear then it is not useful (see the graphs at the bottom).

Separating Variables

For Merit and Excellence one of the things you can do is separate out the data by a categorical variable to see if this is affecting the trend. You do this by dropping a variable onto the 'Subset By' section on iNZight as shown below (in this case origin of the car). This gives the output on the right.

