Level 3 – AS91581 – 4 Credits – Internal

Investigate Bivariate Measurement Data

Written by Jake Wills – MathsNZ – jwills@mathsnz.com

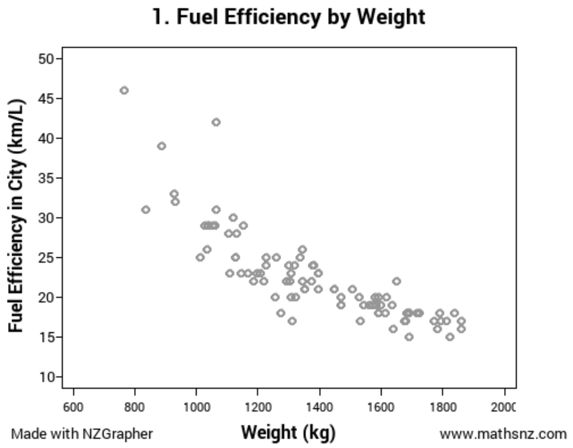
Achievement	Achievement with Merit	Achievement with Excellence
Investigate bivariate measurement data.	Investigate bivariate measurement data, with justification.	Investigate bivariate measurement data, with statistical insight.

Part 1: Problem

For each of the graphs below write a good relationship question. A question should have:

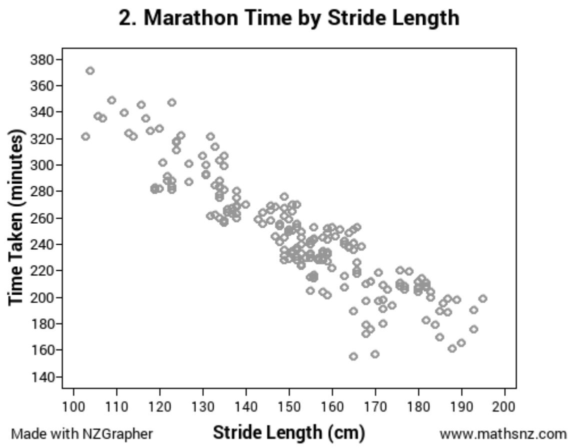
- Why you are interested in looking at this relationship (i.e. the context).
- What you are trying to find a relationship between.
- What you are planning on predicting (what is on the y-axis).

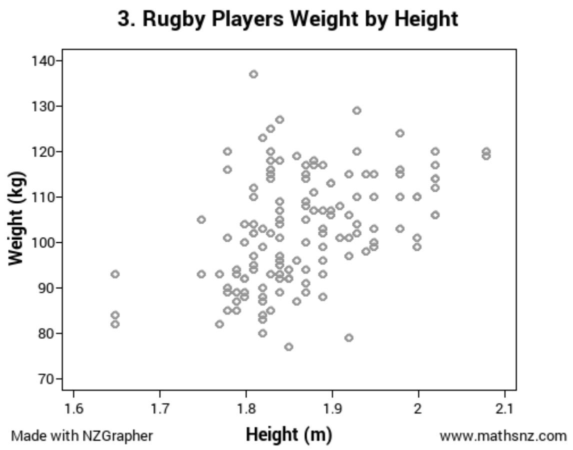
You should be referencing your context (i.e. research – it doesn't matter how you reference, just that you do… using footnotes is normally easiest). The first one has been done for you.

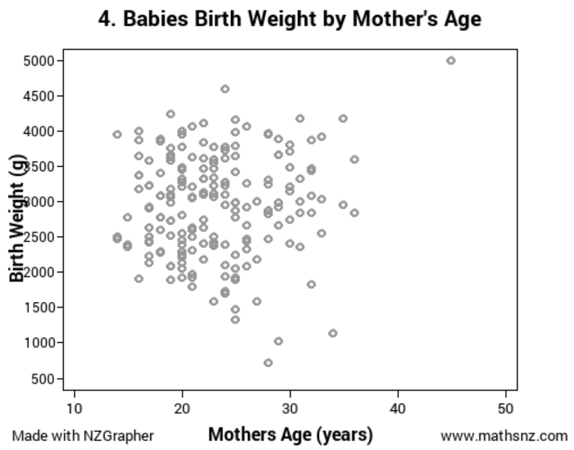


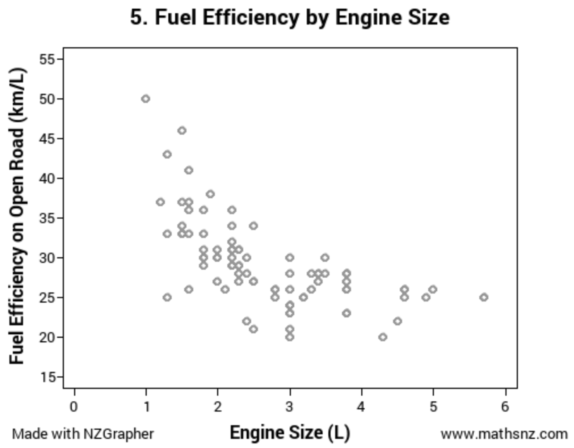
I am looking at buying a car and I have heard that heavy cars use more petrol[1]. Therefore I wonder if there is a relationship between the weight of a car and the car's fuel efficiency for the purpose of predicting the fuel efficiency.

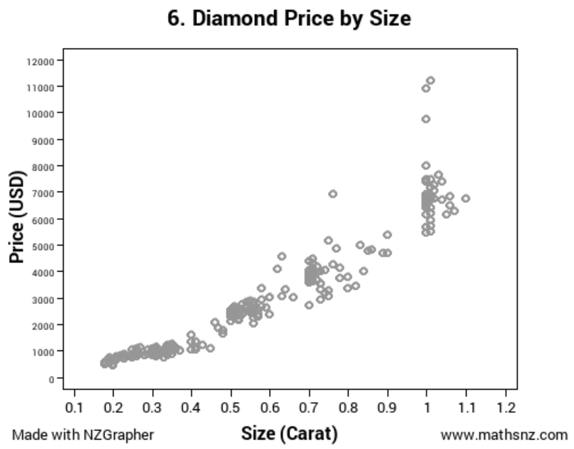
1. http://www.theaa.com/motoring_advice/car-buyers-guide/cbg_emissions.html






4. Babies Birth Weight by Mother's Age
Birth Weight (g)
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
10
20
30
40
50
Mothers Age (years)
Made with NZGrapher
www.mathsnz.com

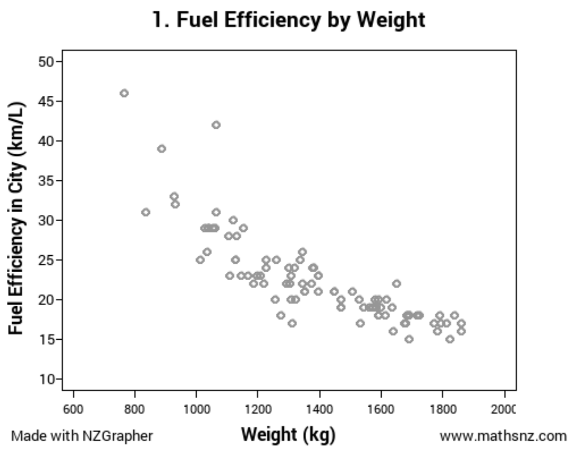
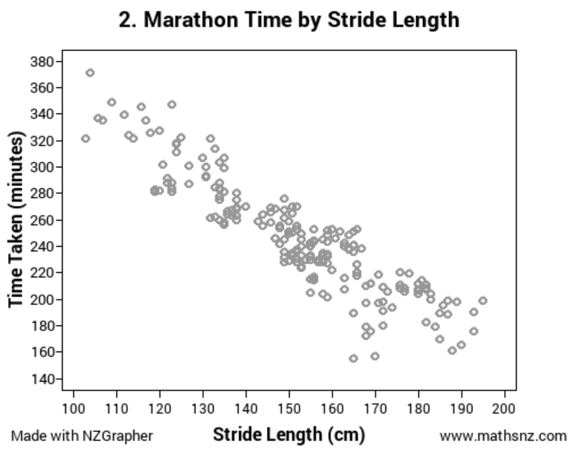
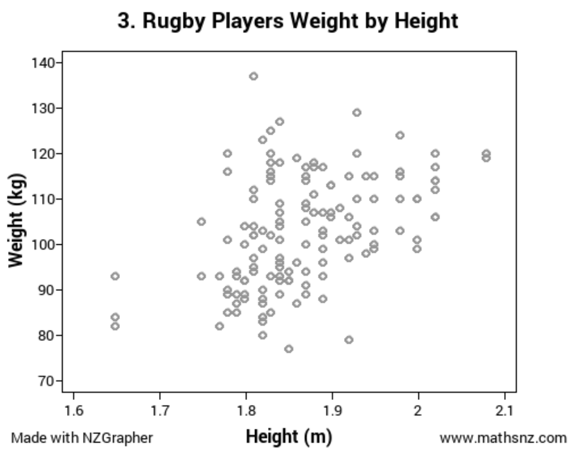

5. Fuel Efficiency by Engine Size
Fuel Efficiency on Open Road (km/L)
55
50
45
40
35
30
25
20
15
0
1
2
3
4
5
6
Engine Size (L)
Made with NZGrapher
www.mathsnz.com

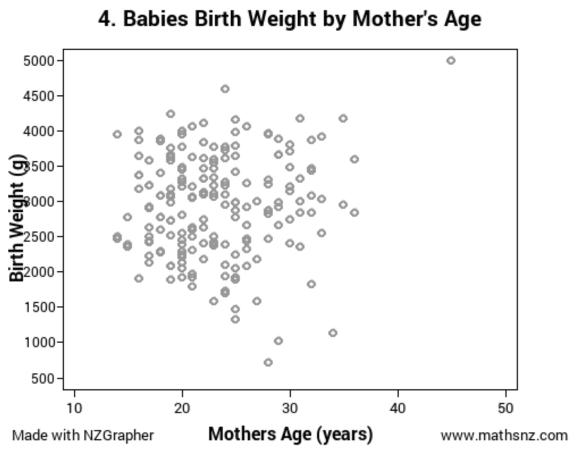

6. Diamond Price by Size
Price (USD)
12000
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1.1
1.2
Size (Carat)
Made with NZGrapher
www.mathsnz.com

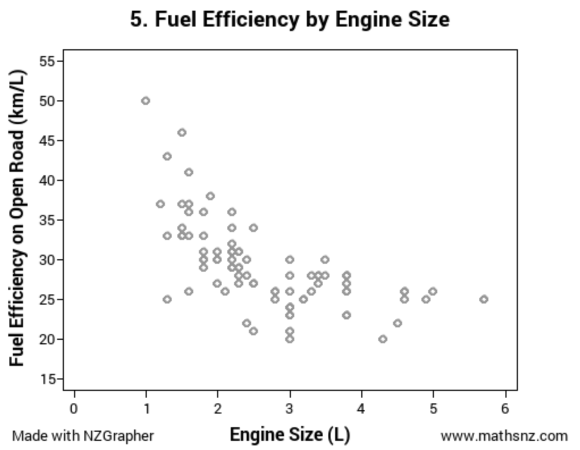
Part 2: Plan

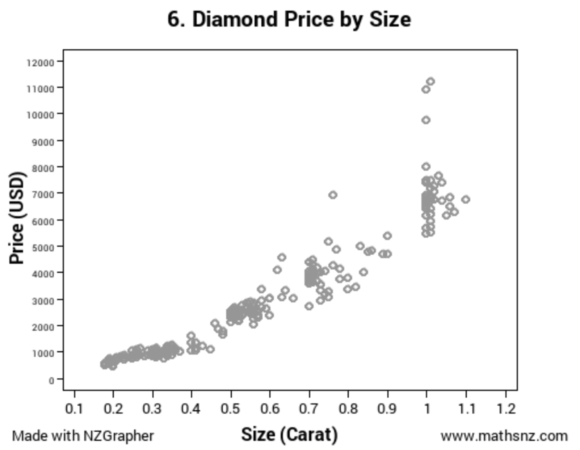
Part 2.1: Identifying the Variables

The next thing that we need to do is identify our variables and say what units are being used. The independent variable is the variable on the x-axis, and the dependent variable is the variable on the y-axis that we are wanting to predict.
Identify the variables for each of the graphs below. The first one has been done as an example for you.

	The independent variable is the weight of the car, which is the mass of the car in kilograms. The dependent variable is the fuel efficiency in cities and on motorways, measured in kilometres per litre.
	
	


4. Babies Birth Weight by Mother's Age
Birth Weight (g)
Mothers Age (years)
Made with NZGrapher
www.mathsnz.com

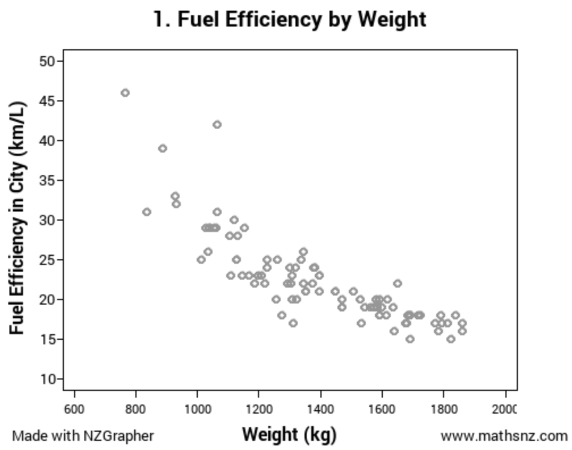

5. Fuel Efficiency by Engine Size
Fuel Efficiency on Open Road (km/L)
Engine Size (L)
Made with NZGrapher
www.mathsnz.com


6. Diamond Price by Size
Price (USD)
Size (Carat)
Made with NZGrapher
www.mathsnz.com

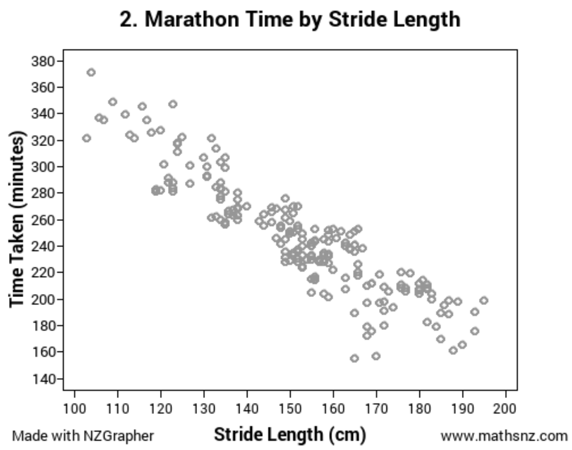
Part 2.2: Naming the Source

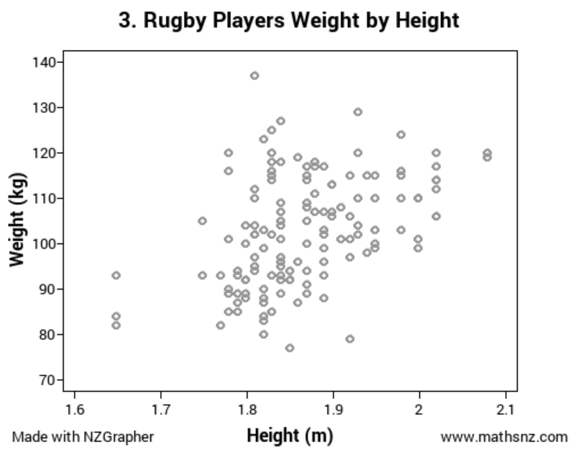
In order for our report to have validity, we need to state where the data has come from. Name the source for each of the graphs below. The first one has been done as an example for you.

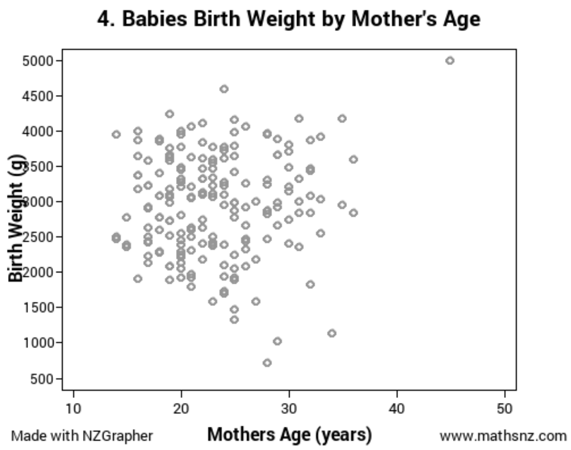
Hint: Use the Information of all of the data sources.

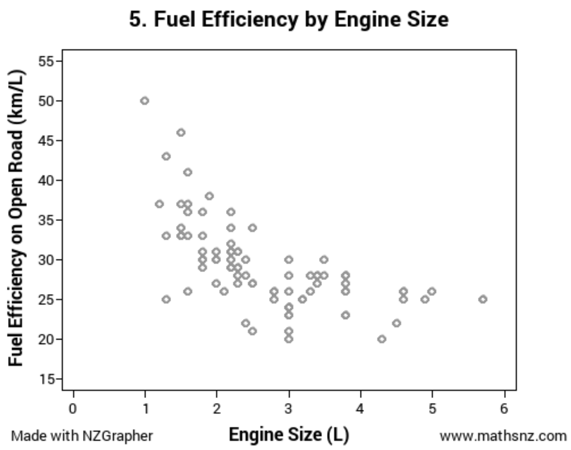


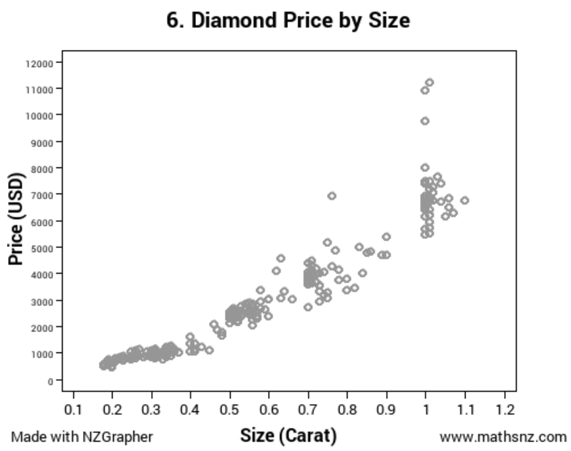
The data used in this investigation is from new vehicles sold in America in 1993.





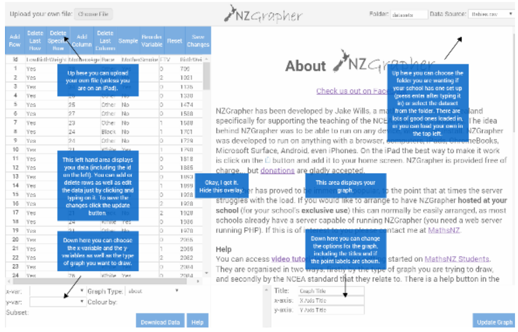
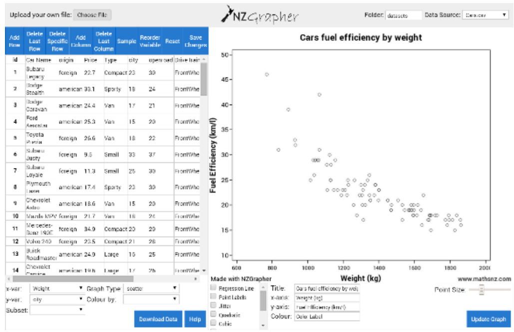
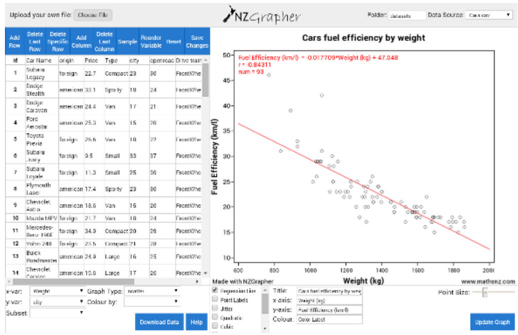

4. Babies Birth Weight by Mother's Age
Birth Weight (g)
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
10
20
30
40
50
Made with NZGrapher
Mothers Age (years)
www.mathsnz.com


5. Fuel Efficiency by Engine Size
Fuel Efficiency on Open Road (km/L)
55
50
45
40
35
30
25
20
15
0
1
2
3
4
5
6
Made with NZGrapher
Engine Size (L)
www.mathsnz.com


6. Diamond Price by Size
Price (USD)
12000
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1.1
1.2
Made with NZGrapher
Size (Carat)
www.mathsnz.com

Part 3: Data – Using NZGrapher

The next section that we need to do is the data section. This is reproducing the graphs on Page 2 using NZGrapher. The example below will go through using the cars dataset for weight by engine size.
NZGrapher runs on anything with a browser… Macs, PCs, iPad, Android, ChromeBooks and more.

First up we need to start NZGrapher by going to the link in the box to the right.	www.jake4maths.com/grapher
The first time you load NZGrapher it will display an overlay with descriptions as to what all the different areas do as shown to the right. To load your data in either select it from the dropdown in the top right, or upload it in the top left corner and press go.	
To draw a scatter plot there are just three things you need to do. 1. Select the x-variable… this is your independent variable that will be on the x-axis, in this case it's engine size. 2. Select the y-variable… this is your response or dependent variable, in this case it's weight. 3. Select the graph type… for this we want the scatter. This will give a graph with just the points. You need to check the graph title and axis labels to make sure they are appropriate (include units where necessary) and press update graph. To save or copy the graph just right click on it and press 'Copy Image' or 'Save Image As' or whatever your device says that is similar. 4. Once you have the graph without any regression line you should add in the regression line by pressing the 'Regression Line' check box. *Note 1: The summary statistics are automatically overlaid in red, if you want to remove them just un-tick the summary statistics box.* *Note 2: If you want to identify the outliers, if you click the 'Point Labels' checkbox this will add little numbers next to the points that correspond with the point id.*	 

Now it is your turn. For each dataset you need to produce the scatter plot for each dataset. Don't forget to add appropriate titles and units to your graph and axis.

Part 4: Analysis

We now start on the Analysis section of our report. The acronym we use for this section is TARSOG. The most important thing that you need to remember in this section is that what you can see with your eyes in the most important, not just the numbers, and your comments should be linked to the context.

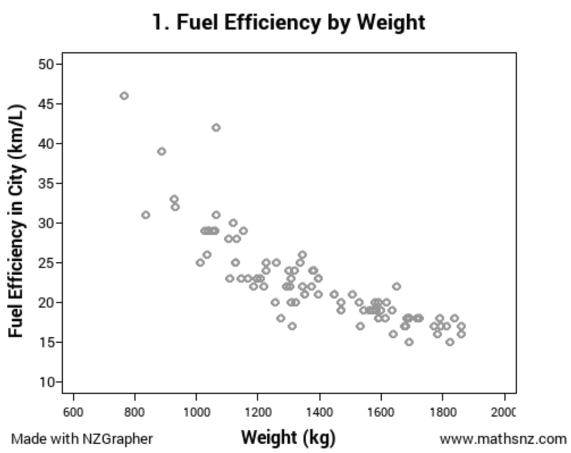
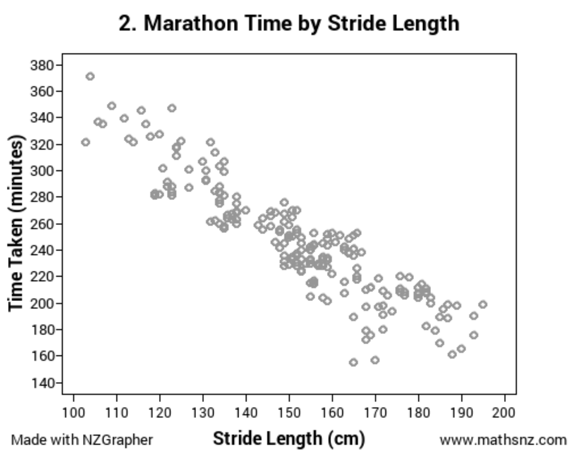
Note: to be going for a **Merit** or **Excellence** Grade in this TARSOG section you need to be justifying these features in context and using research to back up your statements.

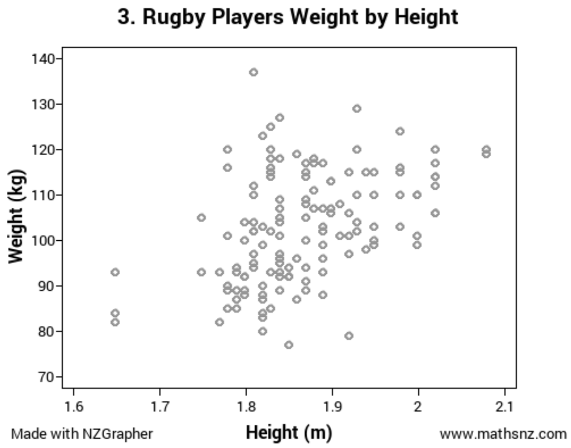
Part 4.1: Trend

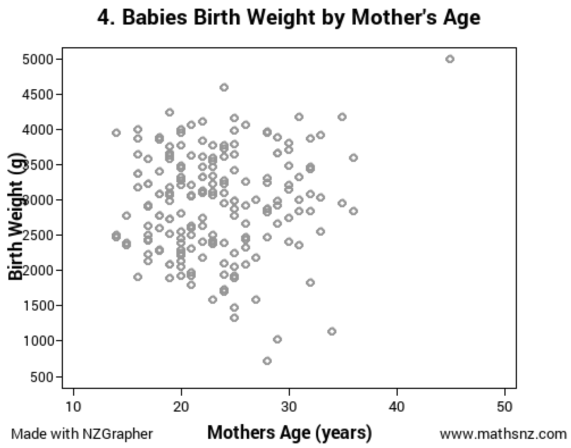
The first comment we need to make is about the trend. There are three statements we need to make about the trend.

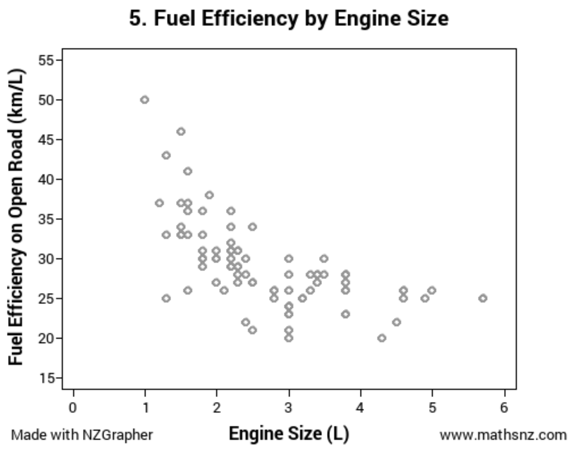
- How strong the trend is: weak, moderate, strong (or somewhere in between),
- If the trend is positive or negative (does it go up or down) and
- Is the trend linear (most circumstances we look at – forming a straight line) or non-linear.

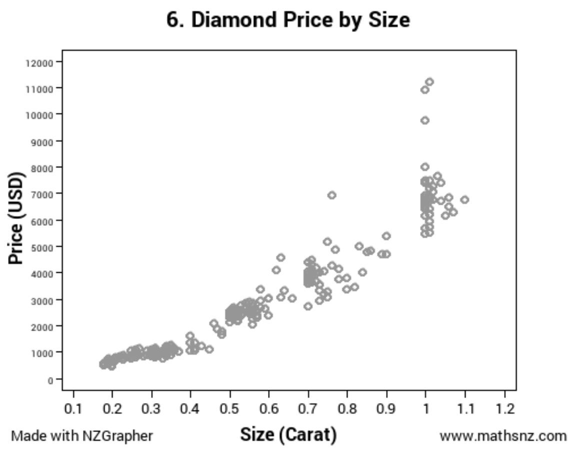
Write a trend statement for each of the datasets. Again the first one has been done for you.

<table>
<tr><td>
</td><td>From the graph I can see a strong negative linear relationship between the weight of a car and the fuel efficiency of the car.</td></tr>
<tr><td>
</td><td></td></tr>
</table>


3. Rugby Players Weight by Height
Weight (kg)
Height (m)
Made with NZGrapher
www.mathsnz.com


4. Babies Birth Weight by Mother's Age
Birth Weight (g)
Mothers Age (years)
Made with NZGrapher
www.mathsnz.com

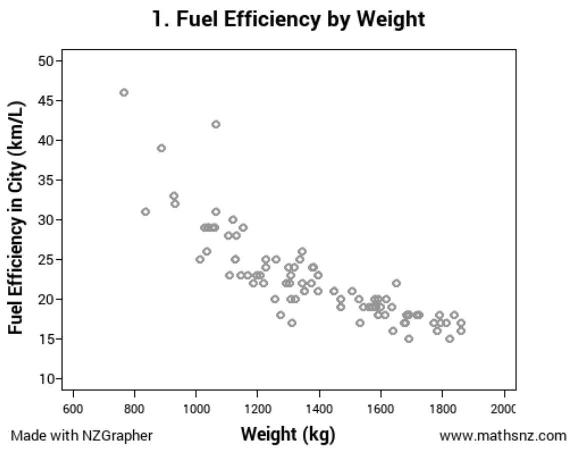
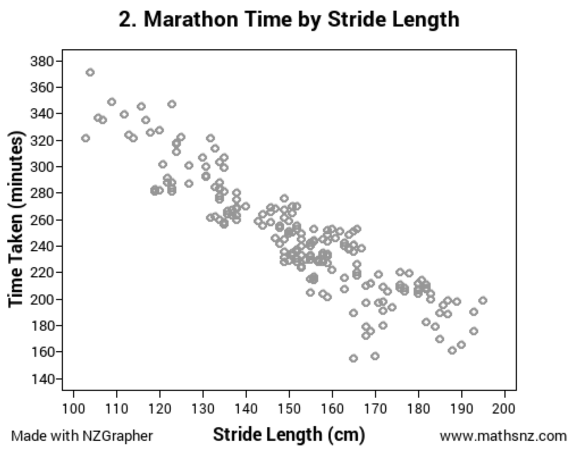
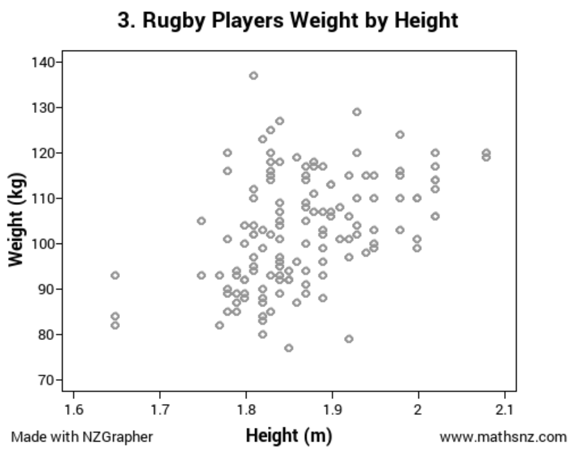

5. Fuel Efficiency by Engine Size
Fuel Efficiency on Open Road (km/L)
Engine Size (L)
Made with NZGrapher
www.mathsnz.com

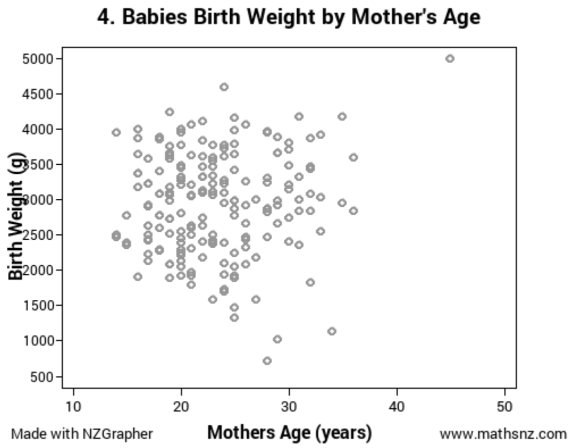

6. Diamond Price by Size
Price (USD)
Size (Carat)
Made with NZGrapher
www.mathsnz.com

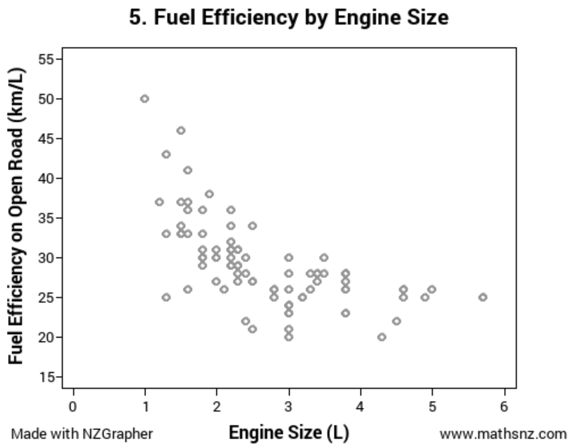
Part 4.2: Association

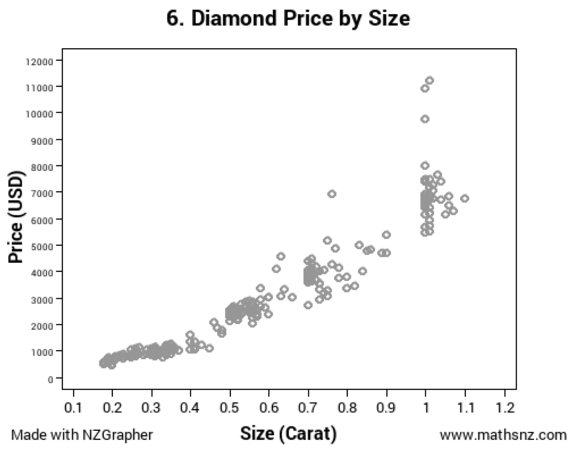
Association is about explaining why the relationship is either positive or negative, and it is important to link back to the context.

Discuss the association for each of the sets of data, the first one has been done for you

	I can see that the association is negative because as the weight of the car increases, the fuel efficiency of the car decreases.
	
	


4. Babies Birth Weight by Mother's Age
Birth Weight (g)
Mothers Age (years)
Made with NZGrapher
www.mathsnz.com


5. Fuel Efficiency by Engine Size
Fuel Efficiency on Open Road (km/L)
Engine Size (L)
Made with NZGrapher
www.mathsnz.com


6. Diamond Price by Size
Price (USD)
Size (Carat)
Made with NZGrapher
www.mathsnz.com

Part 4.3: Relationship

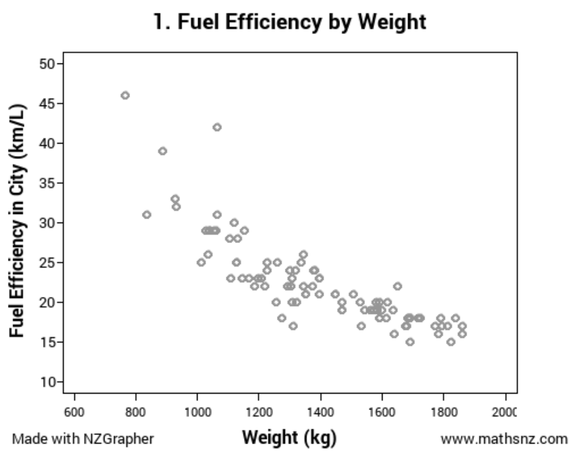
The statement about the relationship is about justifying the strength of the trend that **you can see** on your graph. It is important that you are commenting on what **you can see**. You can use the correlation coefficient (r-value) to **back up** your strength statement, but it should only be used as a backup… **what you can see** is the most important.

The r-value is a number between -1 and 1 indicating how strong the relationship is. The closer it is to 1 or -1 the stronger the relationship is, and the closer it is to zero the weaker the relationship is. A positive r-value indicates that the trend is positive, a negative r-value indicated the trend is negative.
The number line below is **just a guide**, remember what you can see with your eyes is most important.

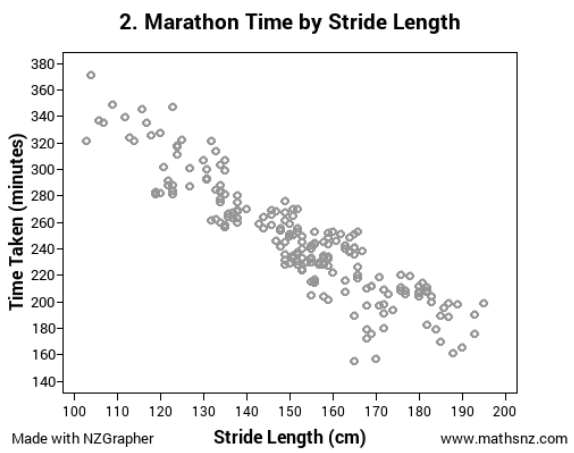
Strong	Moderate	Weak	No Relationship	Weak	Moderate	Strong

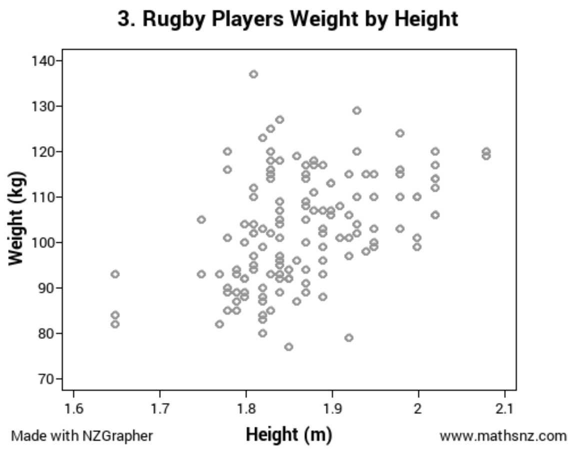
-1 | -0.8 | -0.6 | -0.4 | -0.2 | 0 | 0.2 | 0.4 | 0.6 | 0.8 | 1

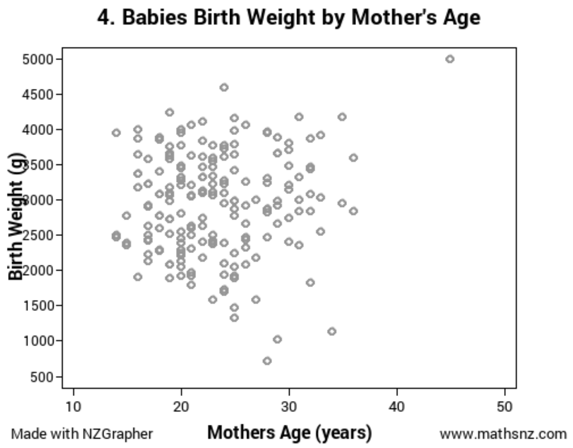
Discuss the relationship for each of the sets of data, the first one has been done for you.

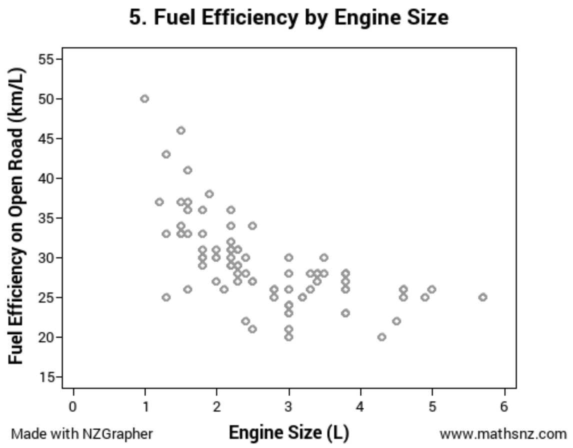


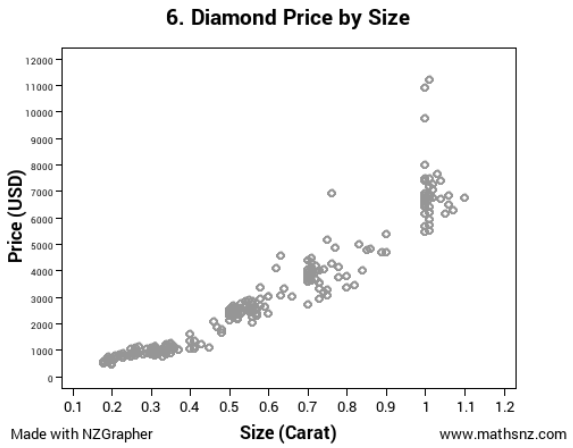
The relationship is strong and linear as I can see most the points form a fairly consistent pattern. This is confirmed by the correlation coefficient of -0.8431, indicating that the linear relationship is quite strong as r is between -0.75 and -1.




3. Rugby Players Weight by Height
Weight (kg)
Height (m)
Made with NZGrapher
www.mathsnz.com


4. Babies Birth Weight by Mother's Age
Birth Weight (g)
Mothers Age (years)
Made with NZGrapher
www.mathsnz.com

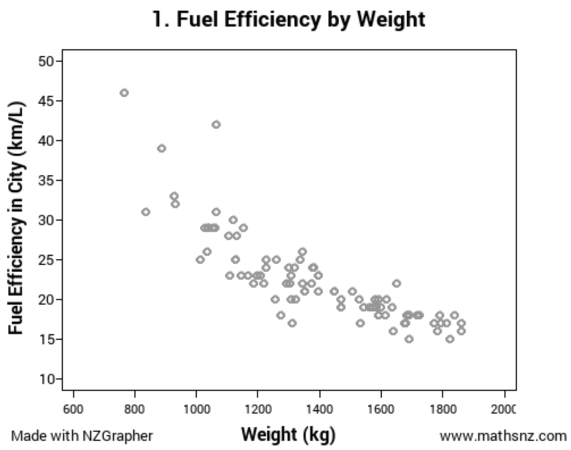

5. Fuel Efficiency by Engine Size
Fuel Efficiency on Open Road (km/L)
Engine Size (L)
Made with NZGrapher
www.mathsnz.com


6. Diamond Price by Size
Price (USD)
Size (Carat)
Made with NZGrapher
www.mathsnz.com

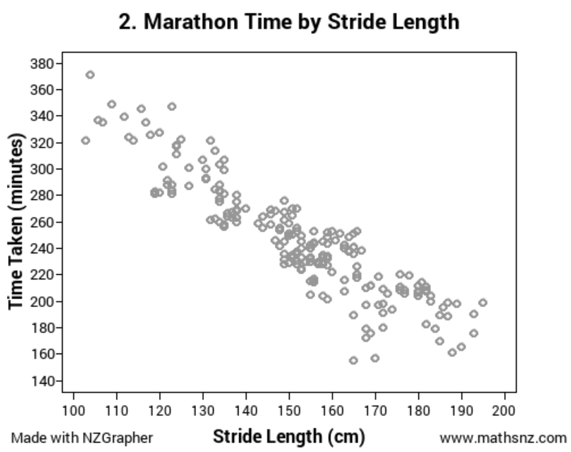
Part 4.4: Scatter

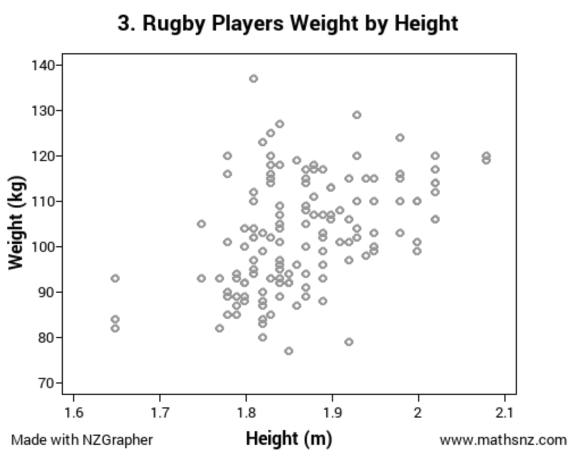
In the scatter section you need to look and see how consistent the scatter is. Are there any areas that are denser or sparser than others?

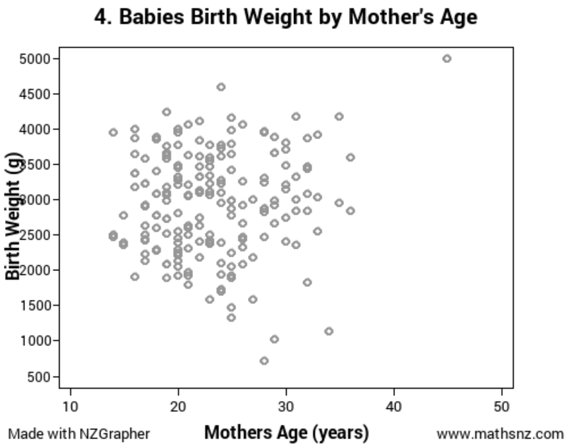
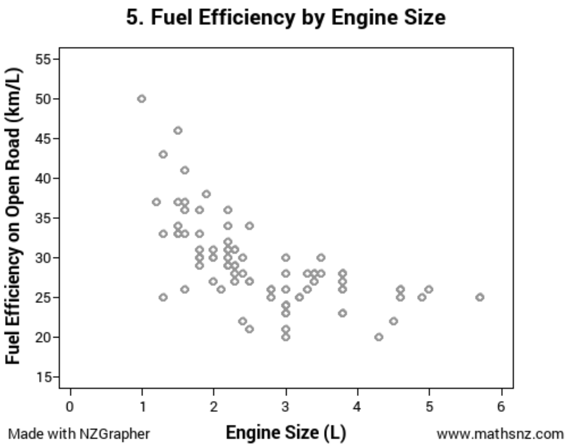
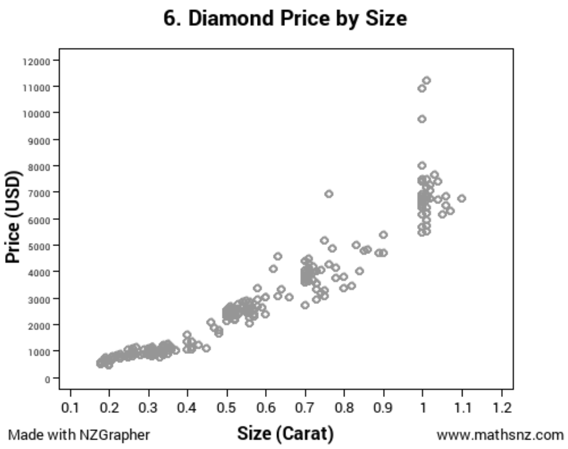
Discuss the scatter for each of the sets of data, the first one has been done for you.



The scatter appears to be reasonably consistent for car weights above about 1000 kg but below this there does appear to be fewer cars, probably due to not many small cars being manufactured.





4. Babies Birth Weight by Mother's Age
Birth Weight (g)
Mothers Age (years)
Made with NZGrapher
www.mathsnz.com

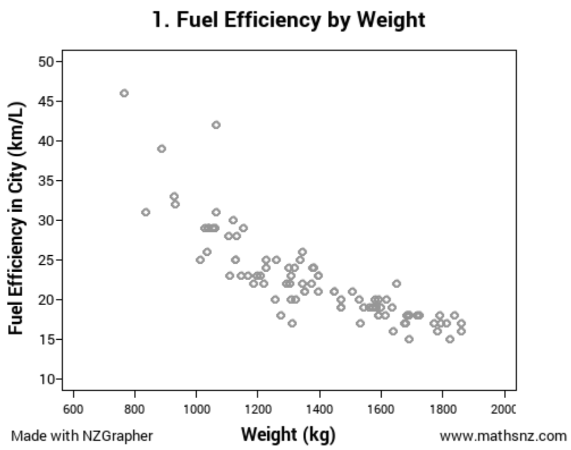
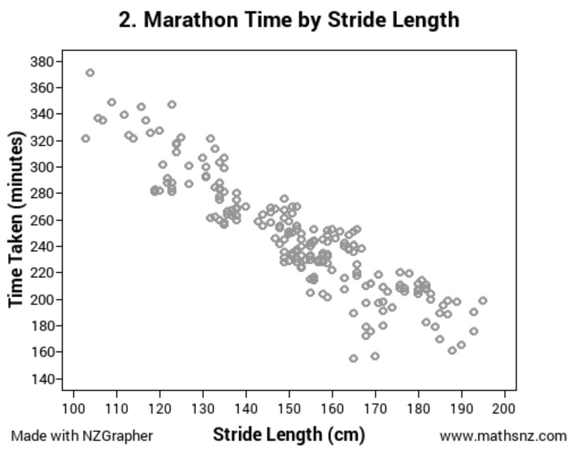
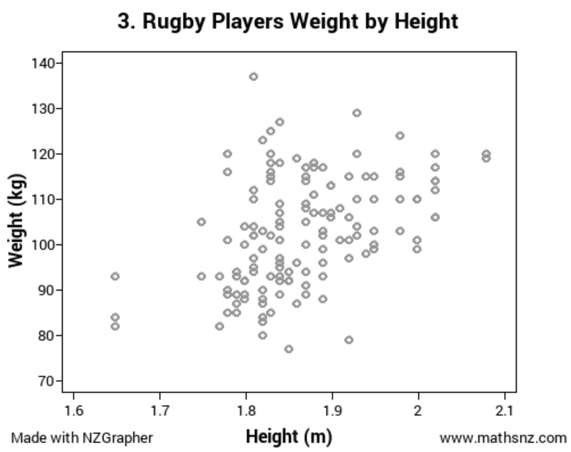
5. Fuel Efficiency by Engine Size
Fuel Efficiency on Open Road (km/L)
Engine Size (L)
Made with NZGrapher
www.mathsnz.com

6. Diamond Price by Size
Price (USD)
Size (Carat)
Made with NZGrapher
www.mathsnz.com

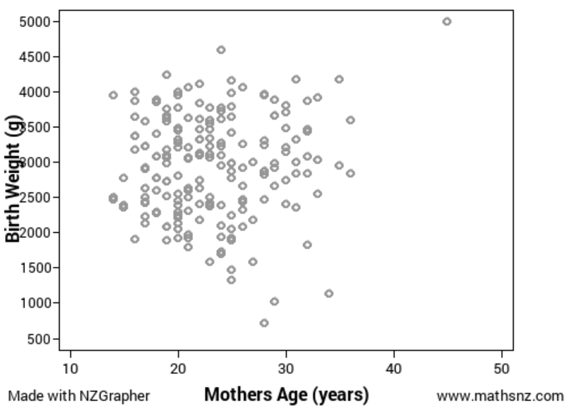
Part 4.5: Unusual Values (Outliers)

In a large number of graphs there will be points (1 or 2) that do not follow the trend. These are called unusual values or outliers. When you identify an outlier you need to find it on the data list and find out as much information about it as you can in order to explain why it might be an outlier.
You could be thinking about the impact of these outliers on your model.

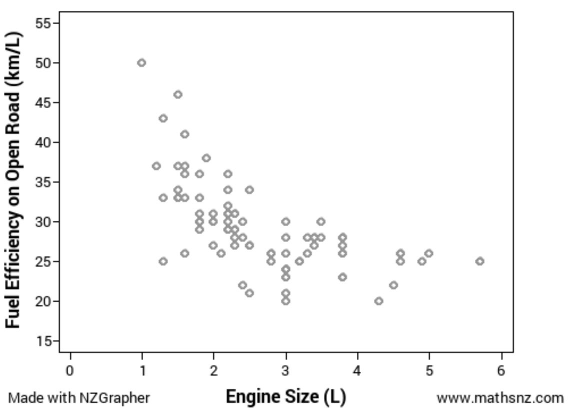
Discuss the outliers for each of the sets of data, the first one has been done for you.

	There are two cars that have higher fuel efficiency rates than expected. The first is a Geo Metro with a weight of 769 kg and a fuel efficiency of 46 km/l in the city. The second is a Honda Civic with a weight of 1066 kg and a fuel efficiency of 42 km/l. Both of these cars have very small engines so I expect this will have increased their fuel efficiency.
	
	

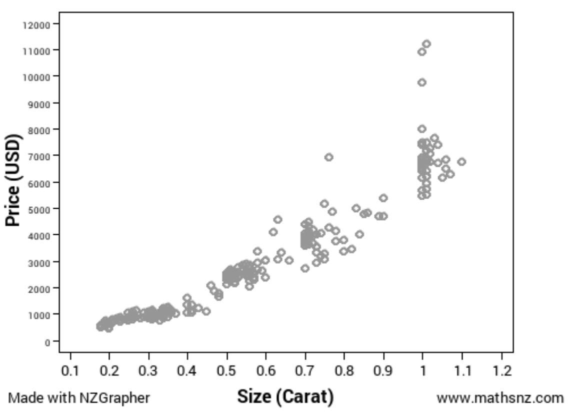
4. Babies Birth Weight by Mother's Age



5. Fuel Efficiency by Engine Size



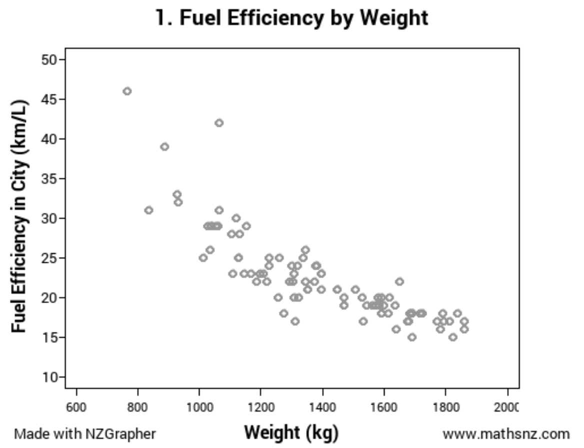
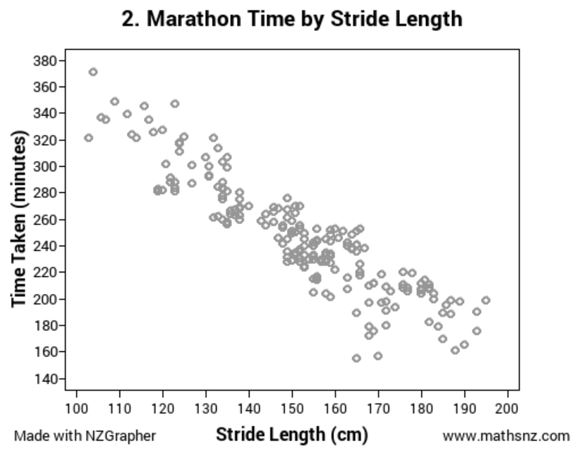
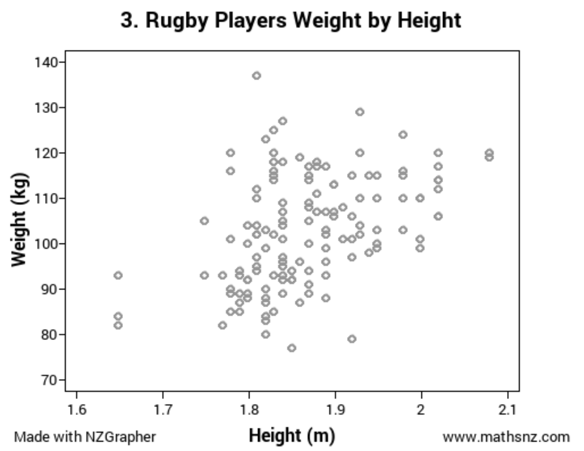
6. Diamond Price by Size

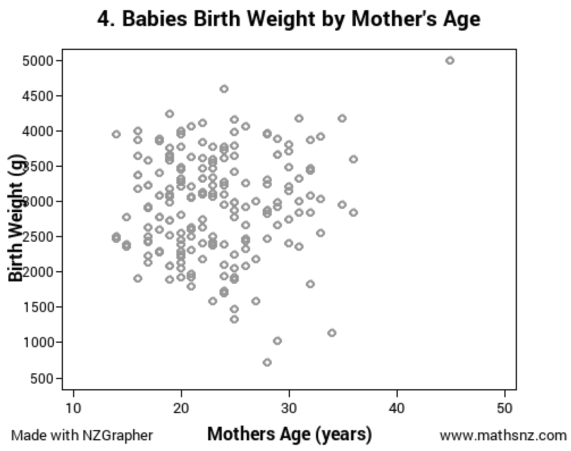


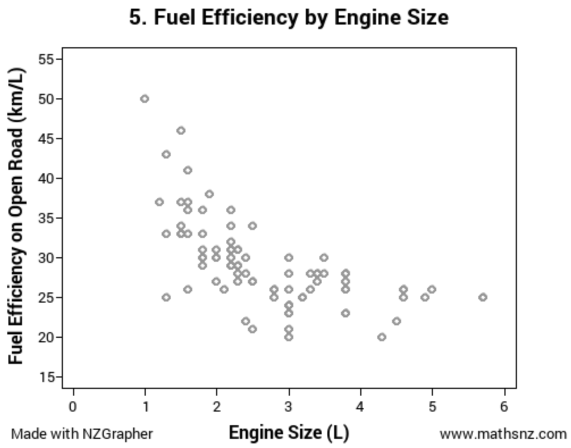
Part 4.6: Grouping

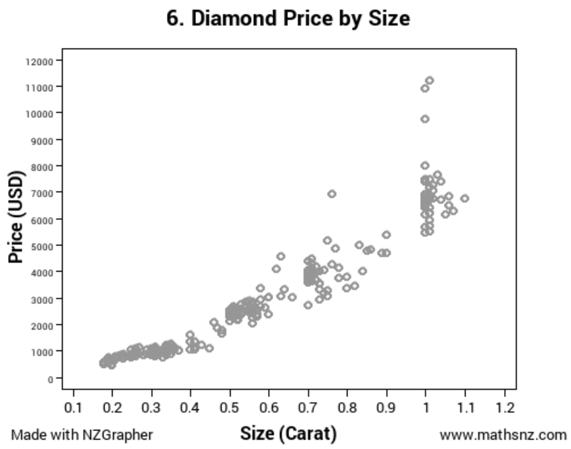
Sometimes in graphs you can end up with two groups (or clusters) of data. If this happens you need to comment on it and what might be causing it, otherwise you can comment that there is not any obvious grouping. Again link it to what you can see.

Discuss the grouping for each of the sets of data, the first one has been done for you.

	Looking at the graph I cannot see any obvious groupings. This is what I would expect as there are not really two different sizes of cars, they are all on a continuous range.
	
	


4. Babies Birth Weight by Mother's Age
Birth Weight (g)
5000
4500
4000
3500
3000
2500
2000
1500
1000
500
10
20
30
40
50
Made with NZGrapher
Mothers Age (years)
www.mathsnz.com


5. Fuel Efficiency by Engine Size
Fuel Efficiency on Open Road (km/L)
55
50
45
40
35
30
25
20
15
0
1
2
3
4
5
6
Made with NZGrapher
Engine Size (L)
www.mathsnz.com

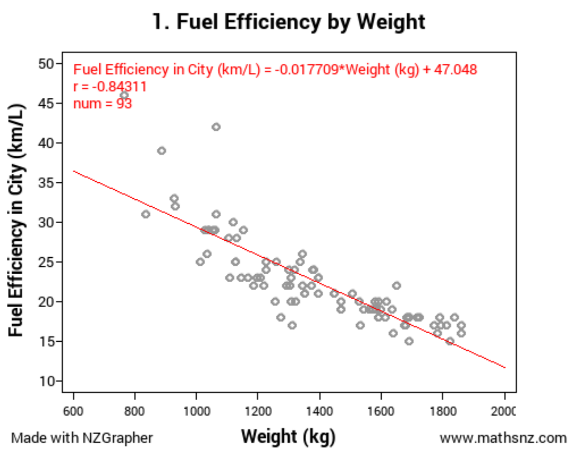

6. Diamond Price by Size
Price (USD)
12000
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1.1
1.2
Made with NZGrapher
Size (Carat)
www.mathsnz.com

Part 4.7: Interpretation of Regression Line

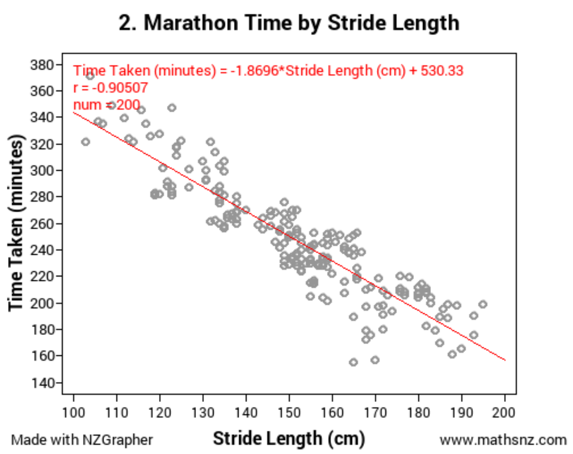
Teachers note: this is not a requirement of the standard, but it does round out the discussion nicely.

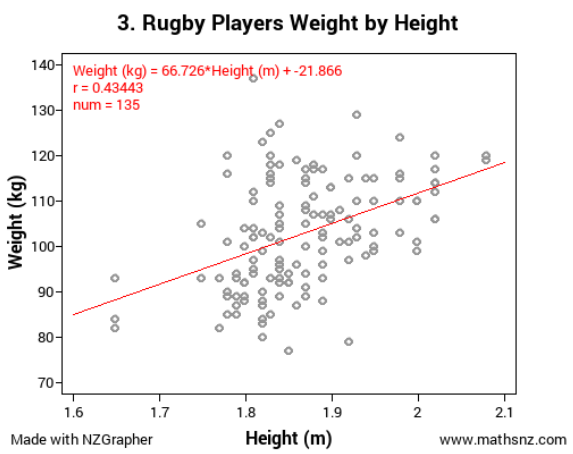
One of the key bits of information that we get given from NZGrapher is the equation of the regression line. Interpreting the gradient of this regression line is an important comment to make. It is vital that you realise that this is only giving the **average** increase over the whole graph, and not a fixed amount for every unit.

Interpret the regression line for each of the sets of data, the first one has been done for you.

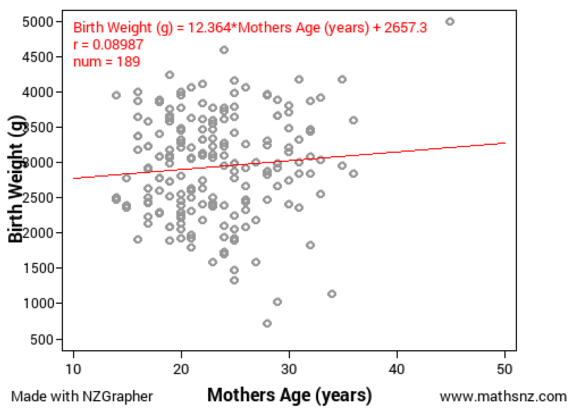


The regression line of
Fuel Efficiency = -0.017709 × Weight + 47.048 means that for every one kilogram increase in the car's weight, the fuel efficiency decreases by 0.017709 kilometres per litre on average.

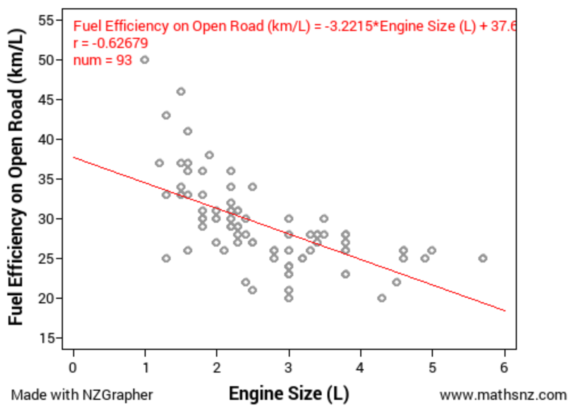




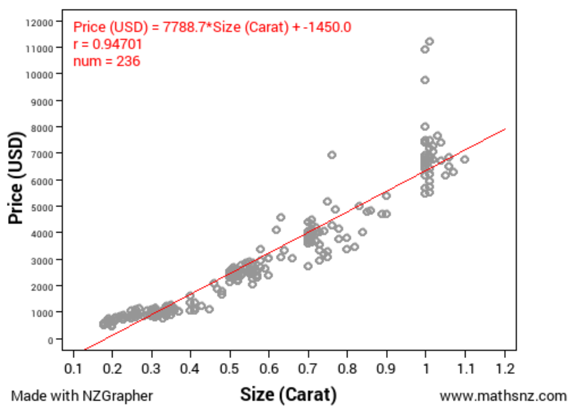
4. Babies Birth Weight by Mother's Age



5. Fuel Efficiency by Engine Size



6. Diamond Price by Size



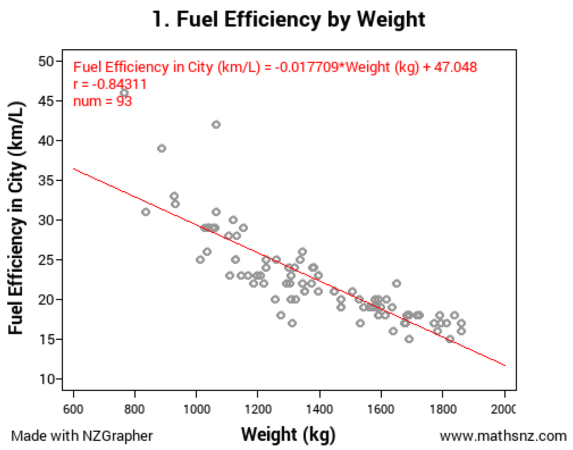
Part 4.8: Predictions

As well as interpreting the regression line we need to use this line to make **at least two** predictions and comment on how reliable we think the predictions are based on the strength of the relationship and the scatter on the graph close to the point we are predicting.

In order to do this you need to substitute two different values into the equation. With bivariate data we can only safely make predictions inside our data range, so your predictions should be able to be plotted on your graph. It is also vital that you **round the prediction sensibly** (usually the same as the original data for that variable was rounded to).

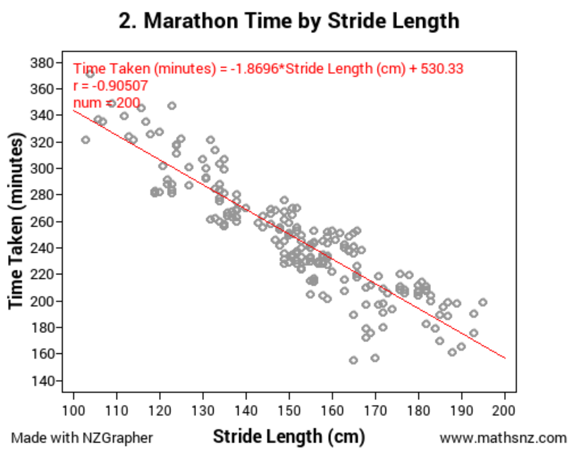
You could expand on this further by discussing the confidence in the predictions in depth or linking this to the residuals (see section later on). You could also reflect on how relevant these predictions are.

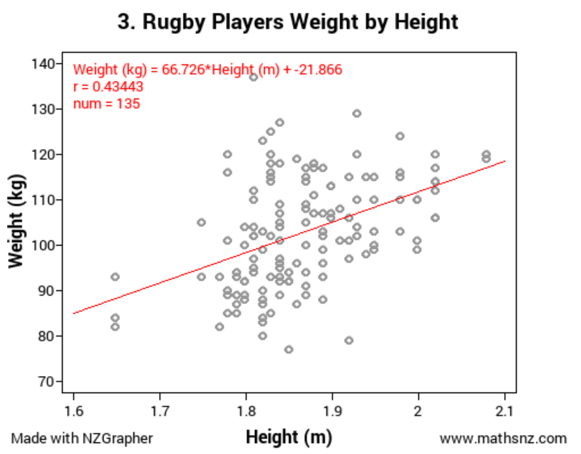
Make two predictions inside the data range for each of the sets of data, the first one has been done for you.

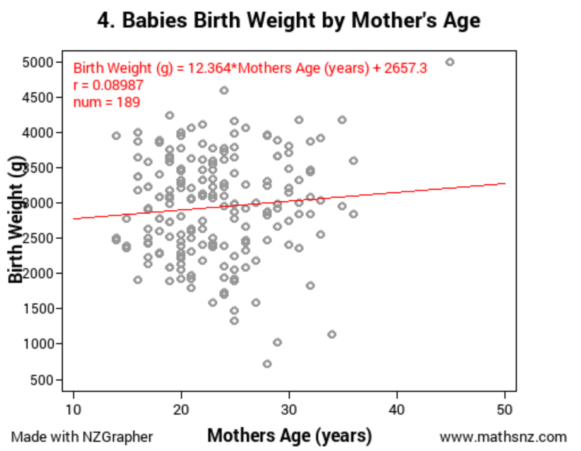


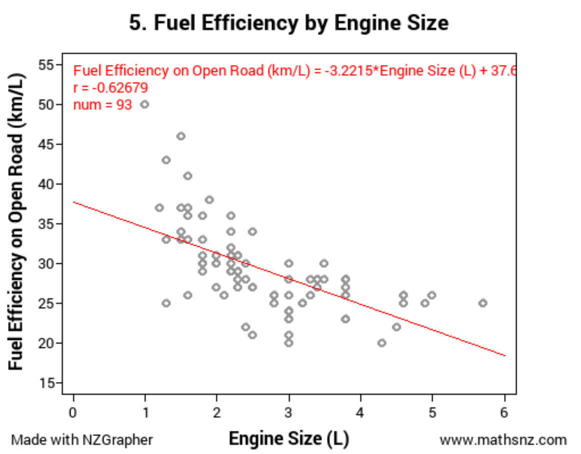
-0.017709 × 1200 + 47.048 = 26.7972
Based on my regression line I would predict that a car that weighs 1200 kg would have a fuel efficiency of approximately 27 kilometres per litre. I am / am not confident in this prediction because...

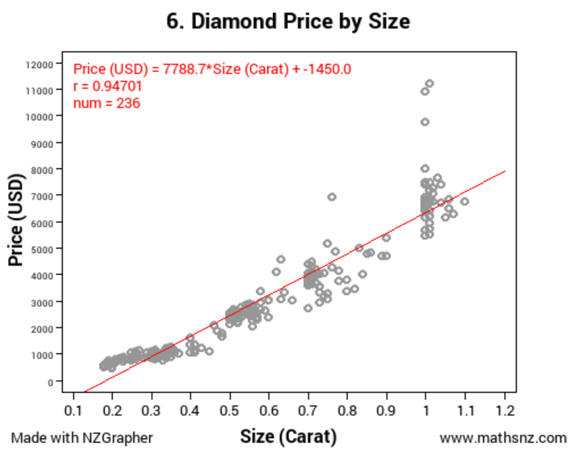
-0.017709 × 1600 + 47.048 = 18.7136
Based on my regression line I would predict that a car that weighs 1600 kg would have a fuel efficiency of approximately 19 kilometres per litre. I am / am not confident in this prediction because...




3. Rugby Players Weight by Height
Weight (kg) = 66.726*Height (m) + -21.866
r = 0.43443
num = 135
Weight (kg)
Height (m)
Made with NZGrapher
www.mathsnz.com

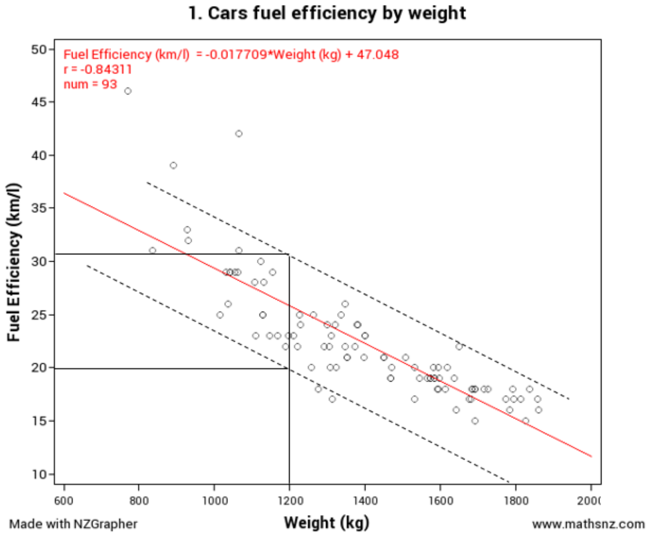

4. Babies Birth Weight by Mother's Age
Birth Weight (g) = 12.364*Mothers Age (years) + 2657.3
r = 0.08987
num = 189
Birth Weight (g)
Mothers Age (years)
Made with NZGrapher
www.mathsnz.com


5. Fuel Efficiency by Engine Size
Fuel Efficiency on Open Road (km/L) = -3.2215*Engine Size (L) + 37.6
r = -0.62679
num = 93
Fuel Efficiency on Open Road (km/L)
Engine Size (L)
Made with NZGrapher
www.mathsnz.com


6. Diamond Price by Size
Price (USD) = 7788.7*Size (Carat) + -1450.0
r = 0.94701
num = 236
Price (USD)
Size (Carat)
Made with NZGrapher
www.mathsnz.com

Part 4.9: Using the Graph for Confidence in Predictions

Looking at graph and how the scatter is can give us a good indication as to how reliable our predictions are likely to be, and therefore create an interval that we think our predictions might be in.



If we look at this graph here we can see that most of the points are between the dotted lines (you put these in manually by eye).
If we go back to the prediction that I made earlier... we can't be completely certain a car that weighs 1200kg will have a fuel efficiency of 27 km/l, but we can be reasonably confident that the fuel efficiency for that car will be somewhere between 20 and 31 km/l.

Part 4.10: Cause and Effect and Correlation

Teachers note: this is not a requirement of the standard, but it does round out the discussion nicely.
While there is a relationship between two variables, there are two reasons why we cannot make causal statements:
- We don't know the direction of the cause – Does X cause Y or does Y cause X?
- A third variable may be involved that is responsive for the covariance between X and Y, we call this a lurking variable.

Causal relationships can only be determined by controlled experiments, which we look at in a different standard.

Part 4.11: Residuals

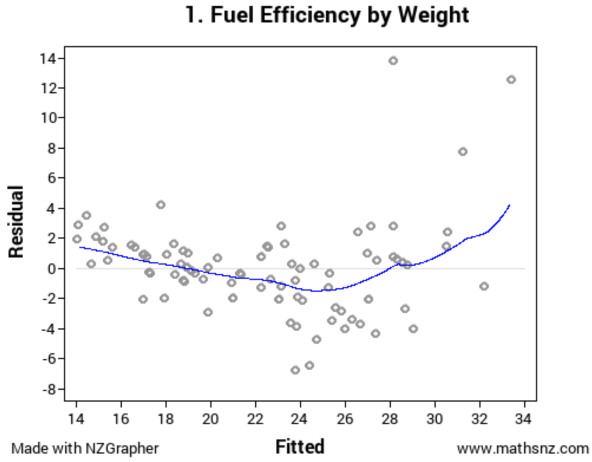
Teachers note: this is not a requirement of the standard, but it does round out the discussion nicely.

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

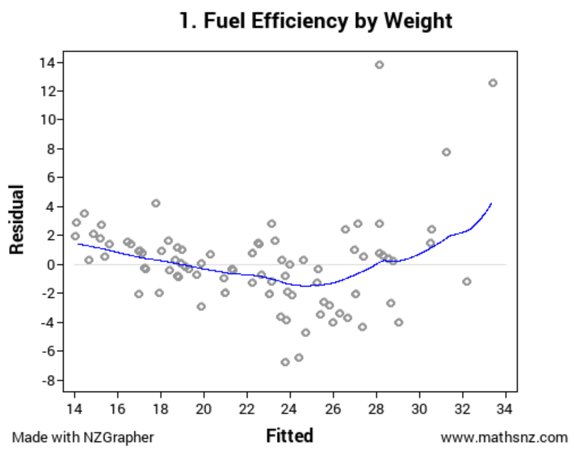
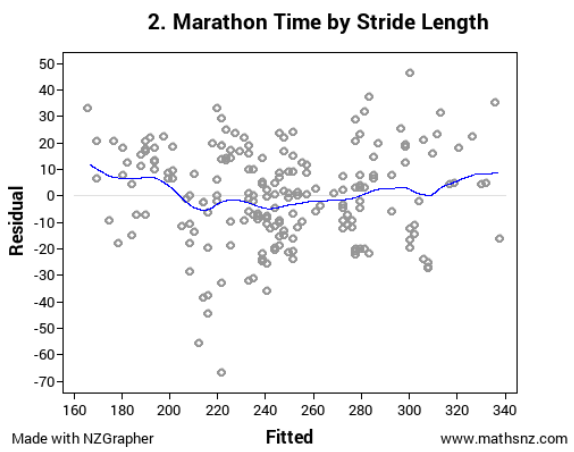
This part is really easy… all you need to do is change the graph type from the graph that we did earlier to 'residuals'.

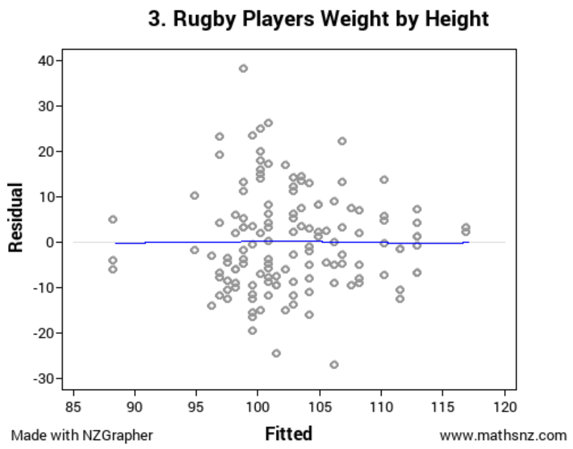
This gives the output shown to the right, which shows the expected (or fitted) values on the x-axis and the difference between the fitted and the actual (the residual) on the y-axis.

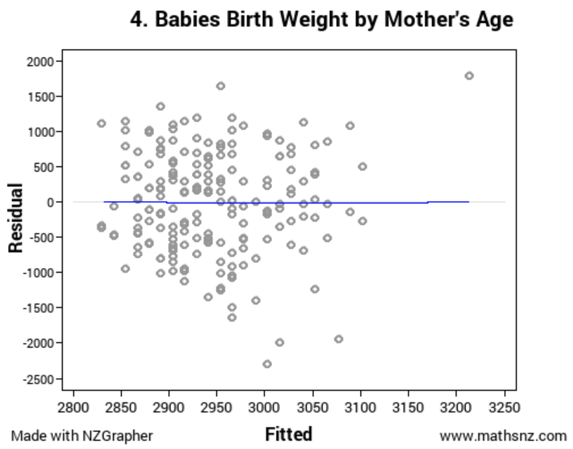
The line that is put in is a weighted average curve that shows the overall trend of the data.

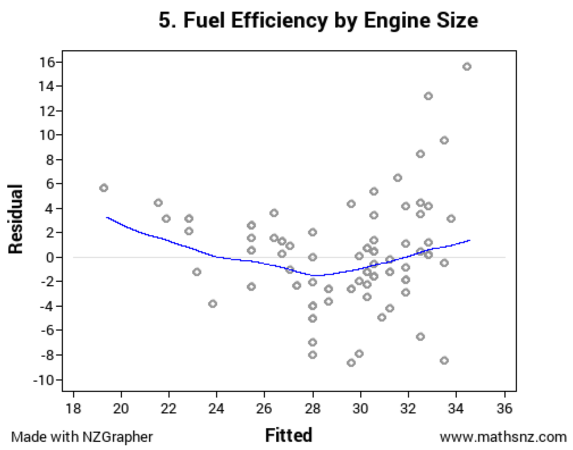


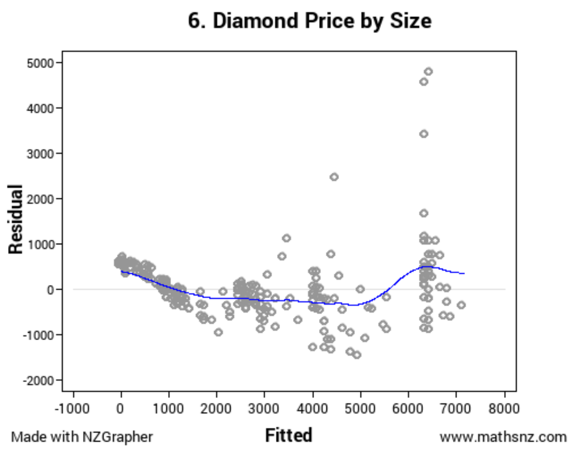
For each of the sets of data, generate the residuals plot and use this to justify how accurate you think your predictions are, the first one has been done for you.

<table>
<tr><td>
</td><td>Looking at my residuals plot I think that the first car I predicted to have a fuel efficiency of 27 km/l might actually be slightly less than this due to most of the values being below the predicted line in the middle of the range. Based on looking at the residuals the car that I predicted to have a fuel efficiency of 19 km/l probably will be very close to this as all of the points round 19 km/l are very close to the predicted line.</td></tr>
<tr><td>
</td><td></td></tr>
</table>


3. Rugby Players Weight by Height
Residual
Fitted
Made with NZGrapher
www.mathsnz.com


4. Babies Birth Weight by Mother's Age
Residual
Fitted
Made with NZGrapher
www.mathsnz.com

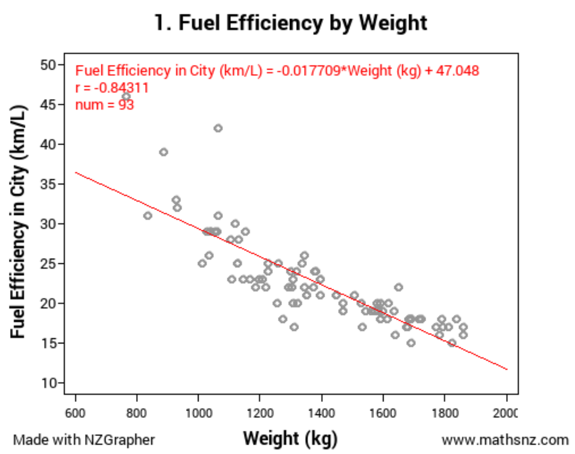

5. Fuel Efficiency by Engine Size
Residual
Fitted
Made with NZGrapher
www.mathsnz.com


6. Diamond Price by Size
Residual
Fitted
Made with NZGrapher
www.mathsnz.com

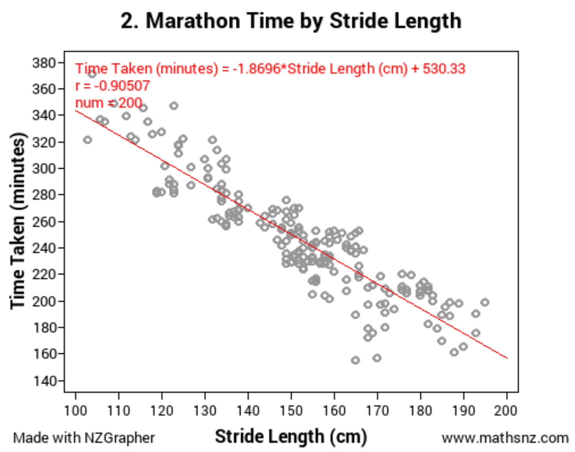
Part 5: Conclusion

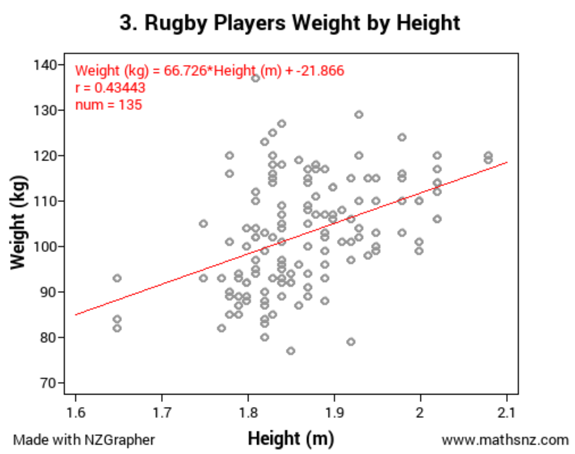
We now need to make a concluding statement to summarise our report. You need to include a statement around the relationship, and it needs to be linked back to what you are investigating.

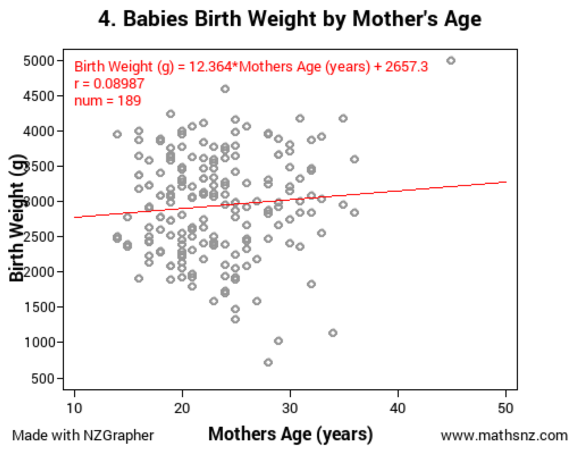
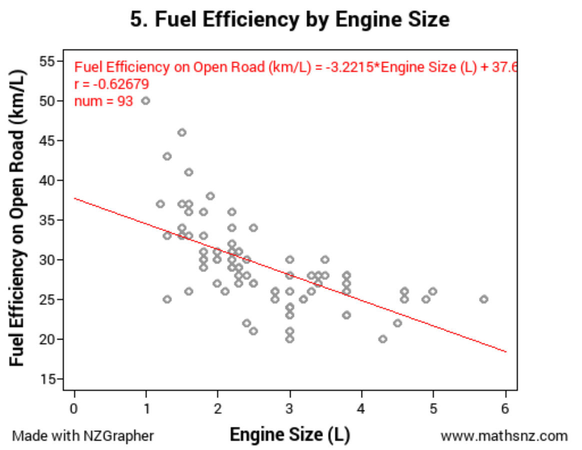
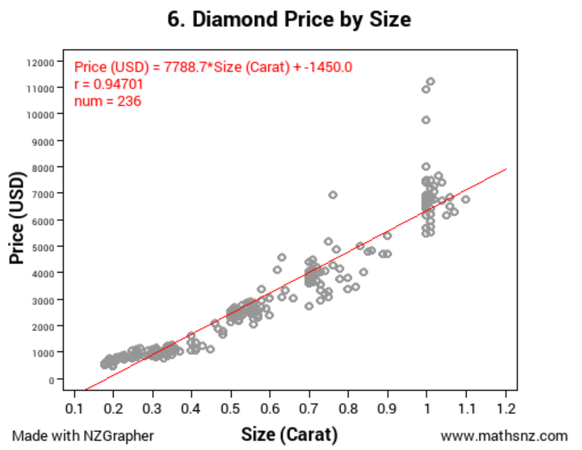
Make a conclusion for each of the sets of data, the first one has been done for you.



In conclusion I think there is a strong negative relationship between the weight of cars and the fuel efficiency – the heaver the car, the more fuel it will use, therefore if we know the weight of a car we should be able to predict the fuel efficiency. This is useful for me to know because if I want a car that will use less petrol I know I should buy a lighter car.





Congratulations, you now have written up a report for 5 different sets of data.

Part 6a: Writing Your Own Internal 1

Using the framework below write a report on the kiwi data.

Kiwi Birds

Problem

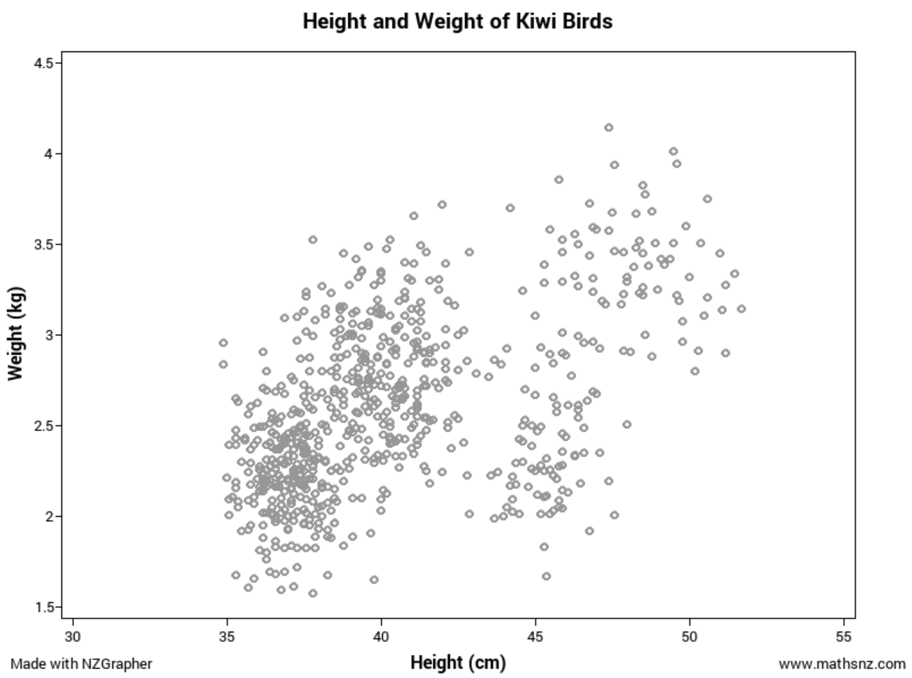
Question and Purpose linked to context, reason for investigation given.

Plan

Variables Identified

Source Named

Data



Graph without Regression Line Given

Analysis

Trend

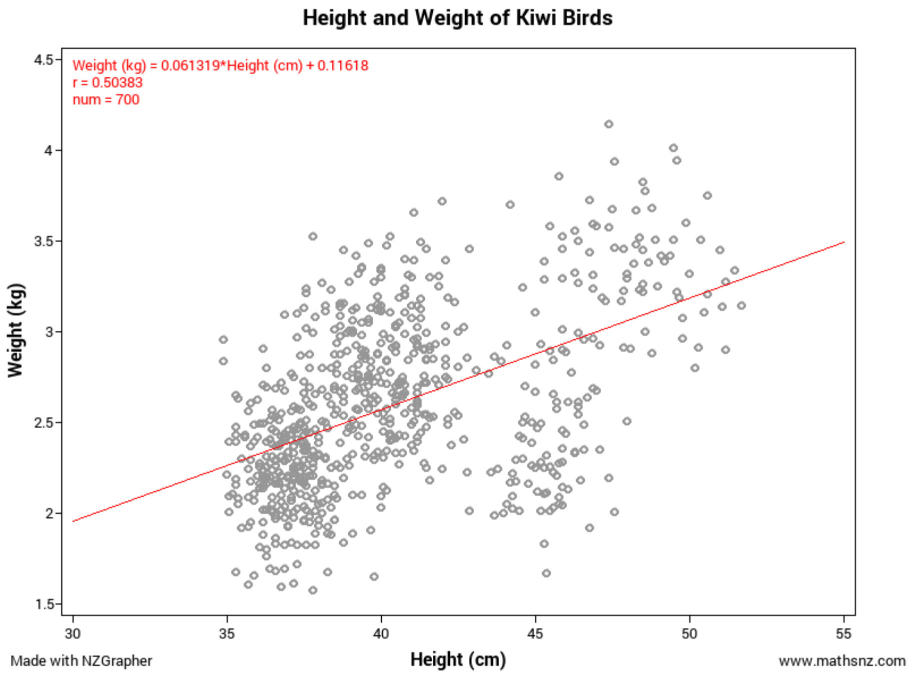
Association

Relationship

Scatter

Outliers

Grouping



Graph with Regression line given

Interpretation of Regression Line

Predictions

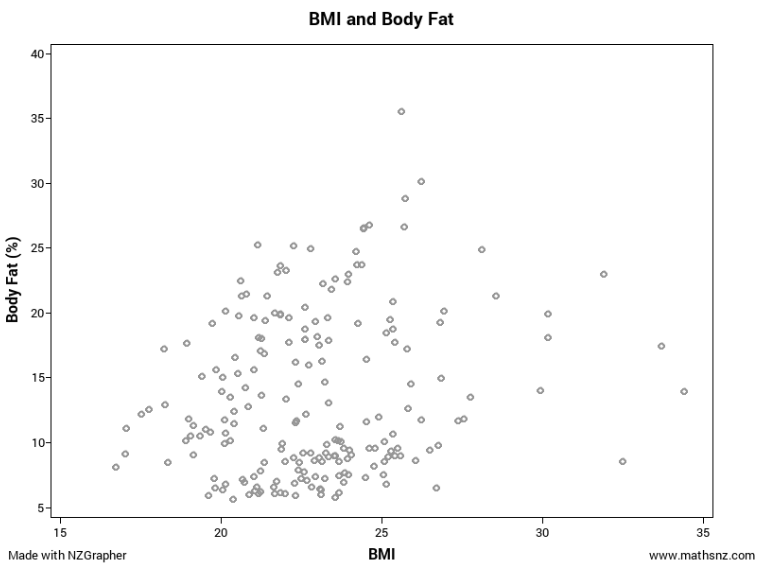
Conclusion

Summarise and link back to the purpose

Part 6b: Writing Your Own Internal 1

This time you have just been provided with a title and graphs. Using these write your own internal. This is using the Sports Science dataset.

BMI and Body Fat




BMI and Body Fat
Body Fat (%) = 0.40537*BMI + 4.2018
r = 0.18756
num = 202
Body Fat (%)
BMI
Made with NZGrapher
www.mathsnz.com

Data Set Information

Babies

The data on 189 births were collected at Baystate Medical Center, Springfield, Mass. during 1986. The goal of this study was to identify risk factors associated with giving birth to a low birth weight baby (weighing less than 2500 grams). Data was collected on 189 women, 59 of which had low birth weight babies and 130 of which had normal birth weight babies.

Variable	Description
LowBirthWeight	No = Birth Weight >= 2500g Yes = Birth Weight < 2500g
MothersAge	Age of the Mother in Years
Race	Race of the mother
MotherSmoke	Smoking Status During Pregnancy
FTV	Number of Physician Visits During the First Trimester
BirthWeight	Birth Weight in Grams

Cars

With rising costs of owning and running a car, and environmental awareness, buyers are becoming more conscious of the features when purchasing new cars. The data supplied is for new vehicles sold in America in 1993.

Variable	Description
Vehicle Name	
Origin	Country of manufacture • America Foreign
Price	US $1000
Type	Small, midsize, large, compact, sporty, van
City	City MPG (miles per gallon by EPA rating)
OpenRoad	Highway MPG
Drive Train	Front Wheel Drive Rear Wheel Drive
Engine Size	Size in litres
Manual Transmission	Yes, No
Weight	Weight of car in Kg

Diamonds

Every diamond is unique, and there are a variety of factors which affect the price of a diamond. Insurance companies in particular are concerned that stones are valued correctly.
Data on 308 round diamond stones was collected from a Singapore based retailer of diamond jewellery, who had the stones valued.

Variable	Description
Carat	Weight of diamond stones in carat units 1 carat = 0.2 grams
Colour	Numerical value given for quality of colour ranging from 1=colourless to 6=near colourless
Clarity	Average = score 1, 2 or 3 Above average = score 4, 5 or 6
Lab	Laboratory that tested & valued the diamond, 1 = laboratory 1, 2 = laboratory 2
Price	Price in US dollars

Kiwi
A sample of kiwi birds around New Zealand was collected in order to help with conservation efforts. The original data is from: http://www.kiwisforkiwi.org/ and was sourced from the secondary school guides (*http://seniorsecondary.tki.org.nz/Mathematics-and-statistics/Achievement-objectives/AOs-by-level/AO-S7-1*)

Variable	Description	
Species	GS-Great Spotted NIBr-NorthIsland Brown Tok-Southern Tokoeka	
Gender	M-Male F-Female	
Weight(kg)	The weight of the kiwi bird in kg	
Height(cm)	The height of the kiwi bird in cm	
Location	NWN-North West Nelson CW-Central Westland EC-Eastern Canterbury StI-Stewart Island NF-North Fiordland	SF-South Fiordland N-Northland E-East North Island W-West North Island

Teachers note: this is a synthesised dataset based on real data. At the time of creating the data set there were around 25,000 brown, 17,000 great spotted and 34,500 southern tokoeka. These numbers formed the basis of the data set, but instead of being out of around 76,000 the data set contains around 700 birds. The data was generated using the population parameters, including gender, location, height, weight and species in Fathom. The size of the population was so that it was too big to use all the data (when doing by hand) but not too big that it couldn't be created for students to use as a "population" to sample from.

Marathon
The data is a sample taken from marathons in NZ.
It is a simple random sample of 200 athletes.

Variable	Description
Minutes	How many minutes they completed the marathon in
Gender	Male (M) or Female (F)
AgeGroup	Younger (under 40) or older (over 40)
StridelengthCM	The persons average stride length over the marathon in cm.

Rugby
The data is real data and comes from http://www.rugby-sidestep-central.com/

Variable	Description
Country	New Zealand or South Africa
Position	Forward or Back
Weight	The weight of the player in kilograms (kg)
Height	The height of the player in metres (m)

Assessment Guidelines – 91581 – Investigate Bivariate Measurement Data

	Achieved (all compulsory)	Merit	Excellence
Problem	Identify a purpose and pose a relationship question which is informed by contextual knowledge.	The question is justified in context and linked to research.	The choice of variables is reflected on and linked to the context and research.
Plan	Data source is identified. The explanatory and response variables are clear.		
Data	Scatter plot(s) is produced with tile and labelled axis, units should be evident		
Analysis	Students need to use a visual inspection to describe features in the data, **before** fitting a model. Features need to include the strength and direction of the relationship, and could include whether a linear model is appropriate, clusters and unusual values. All comments should be in context.	Students need to take care to justify the existence of any unusual value/outlier with reference to the data set and the context. Evidence from the display is used to justify the features discussed	Contextual evidence and research is integrated to support discussion about the features of the data.
Model Fitting	An appropriate model is fitted and its equation is expressed in context.	Students could interpret the gradient in context and/or comment on the strength of the fit using the r value. Students discuss how well the model fits the data across the x value range in context. An analysis of the residuals may be used.	The fit of the model could be discussed by use of a residual plot with insight. The adequacy and strength of the model is reflected upon.
Predictions	A prediction is made for the response variable, using the model equation, in context, for at least one value of the explanatory variable. Prediction should be rounded appropriately.	The precision of the prediction could be discussed by reviewing the strength of the relationship and the scatter on the graph close to the relevant explanatory data value. Causation may be discussed in context.	The precision of the prediction could be discussed by use of error lines and/or a residual plot with insight. The relevance of predictions could be reflected upon with insight. Causation discussion linked to research with insight. Reflect on the model and consider other regression models based on evidence from the display or residual plot and justify and evaluate such models with insight including reflecting on prediction values. Improvements to the model could be considered by considering other variables e.g. separating the variable into relevant subsets or looking at another related variable with insight
Conclusion	Students need to answer their investigative question and state the nature of the relationship. The conclusion links to the purpose of the investigation.	The conclusion is linked to the question with contextual and evidential support.	Reflect on the investigation. Discuss limitations of and improvements to the investigation with insight.

Final grades will be decided using professional judgement based on a holistic examination of the evidence provided against the criteria in the Achievement Standard. Evidence for the award of a grade may be found anywhere in the report.