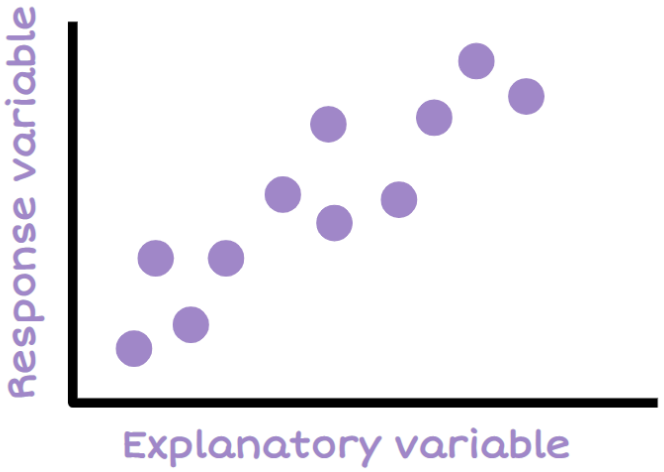
Level 1 Relationship Investigations Workbook

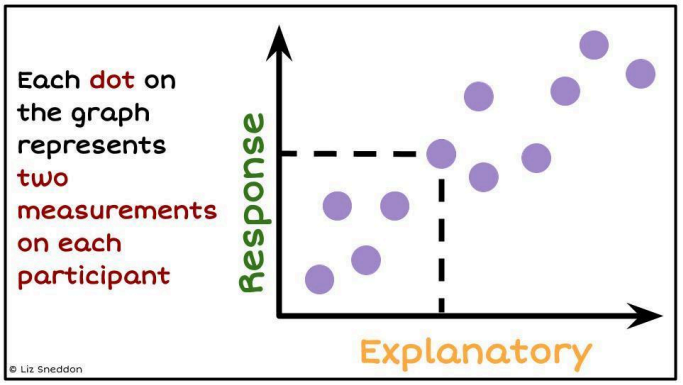


Name:

By Liz Sneddon

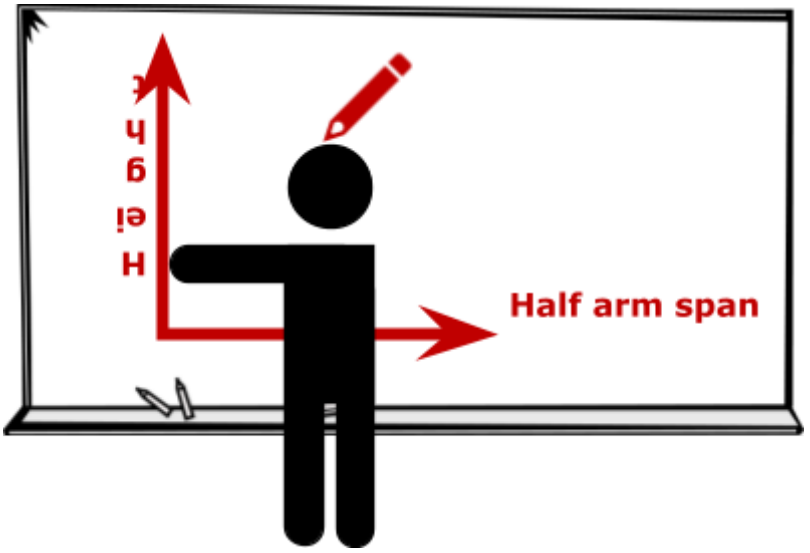
Scatter graphs

Before we get into the PPDAC cycle for collecting data and analysing the features of graphs, we are first going to have a look at scatter graphs.



Class Activity

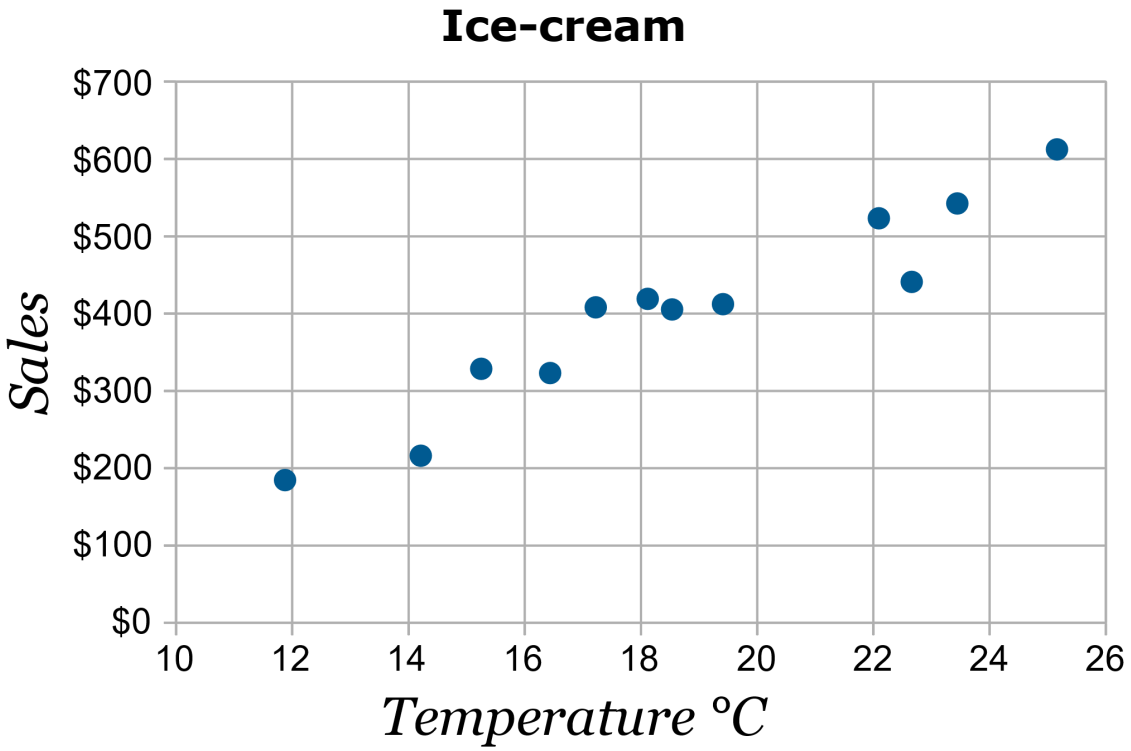
We are going to collect data and draw a scatter graph all at the same time. The two variables we are going to measure the height of ākonga and the length of (half) arm span.



Instructions:

1) Teachers have drawn the axes on the whiteboard (height versus half arm span).
2) Ākonga will come up to the whiteboard one at a time standing facing the whiteboard.
3) Each ākonga will hold out their left arm straight out, with the tip of their finger touching the y axis.
4) The teacher, or another ākonga, will draw a cross where the top of the ākonga head is.
5) Repeat this with all ākonga in the class.

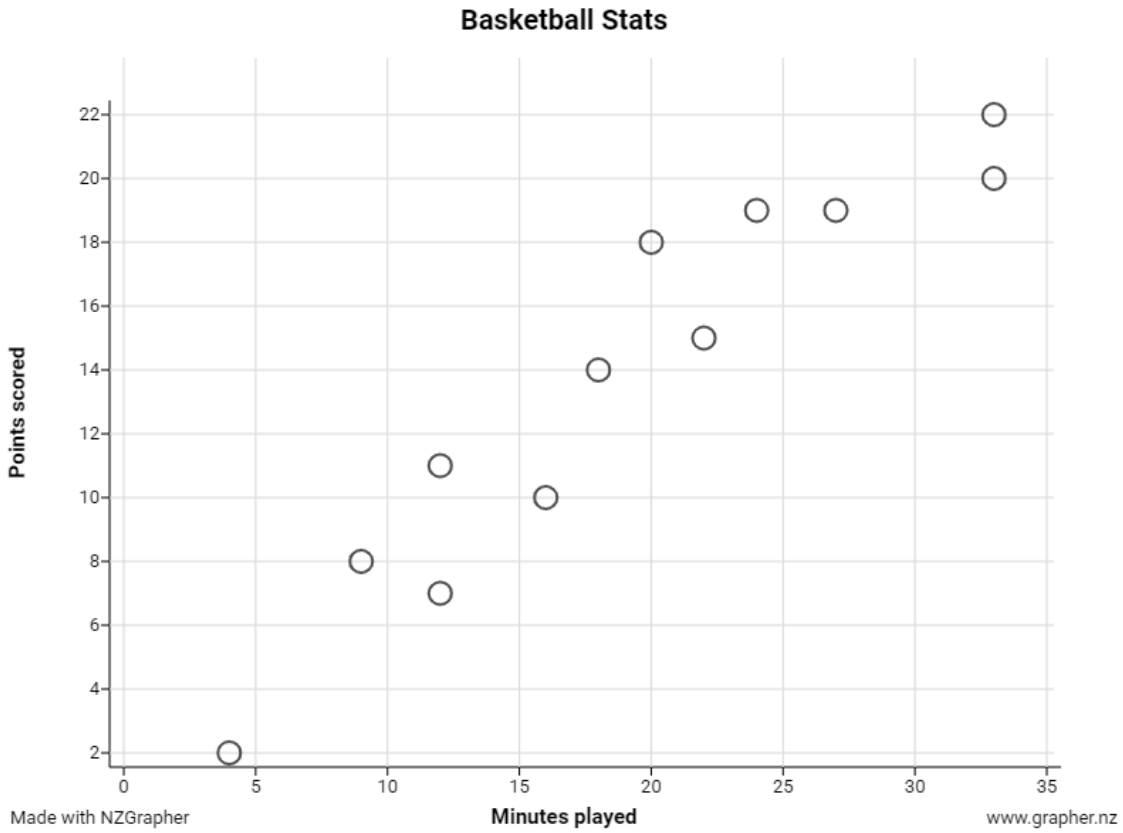
Example:



a)	What are the variables in this graph?	Temperature (°C) Sales ($)
b)	How much money did they make on a day when the temperature was approximately 22°C?	Around $520
c)	What was the temperature when they had approximately $180 of sales?	Around 11.8°C
d)	What do you notice the sales of ice cream when the temperatures are low?	The sales are lower.
e)	What do you notice about the sales of ice cream when the temperatures are high?	The sales are higher.

Exercise 1:

1)



a) What are the variables in this graph?

_____

_____

b) How many points were scored for the basketball player who played for 16 minutes?

_____

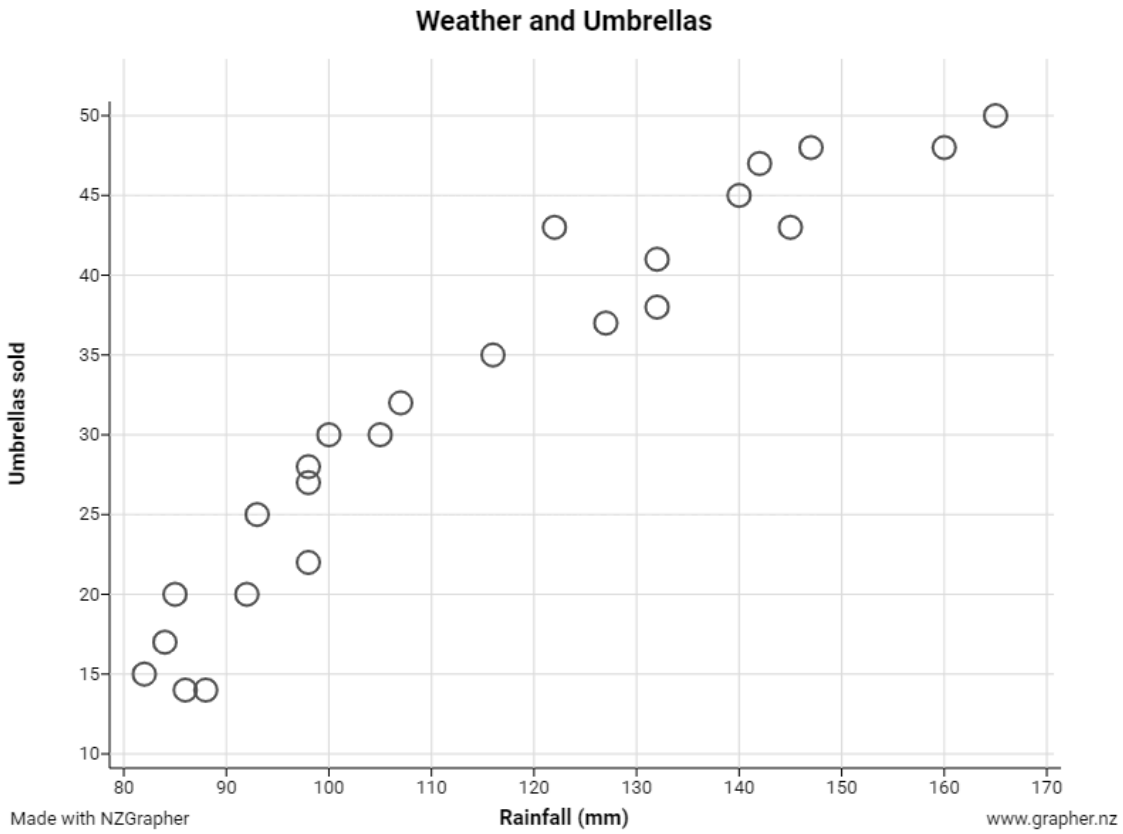
c) How much time did the basketball player who got 18 points spend on court?

_____

d) Looking at the players who spend only a **short** time on the court, the points they score are **low / high.** (circle correct answer)

e) Looking at the players who spend only a **lot of** time on the court, the points they score are **low / high.** (circle correct answer)

2)



a) What are the variables in this graph?

_____

_____

b) How many umbrellas were sold on the day that had 100mm of rain?

_____

c) How much rainfall was there on the day where 35 umbrellas were sold?

_____

d) Looking at days with a **small** amount of rainfall, the number of umbrellas sold are **low / high**. (circle correct answer)

e) Looking at days with a **high** amount of rainfall, the number of umbrellas sold are **low / high**. (circle correct answer)

3)

a) What are the variables in this graph?

_____

_____

b) What is the price of the car for the person who has a salary of $100,000?

_____

c) What is the salary for the person who has a car worth $30,000?

_____

d) Looking at the people who have a **low** salary, their car prices are

low / high. (circle correct answer)

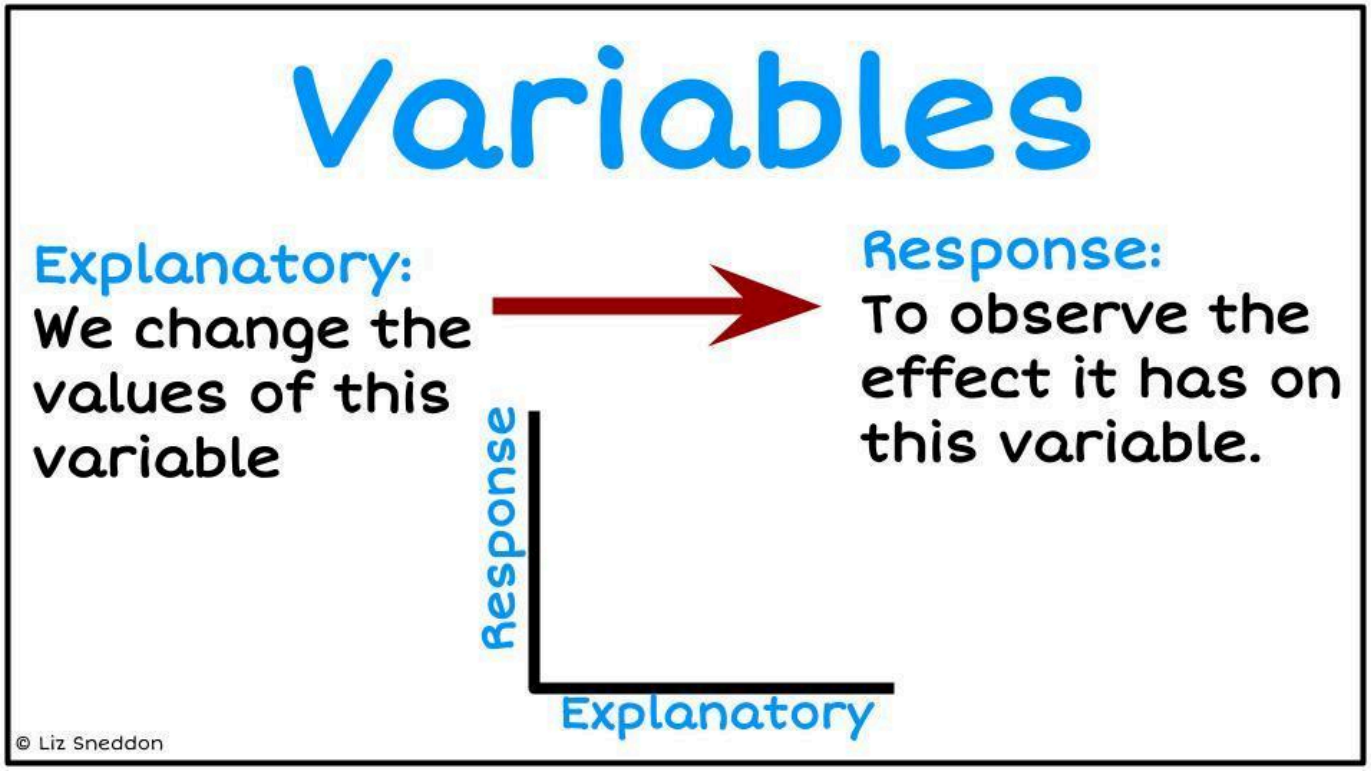
e) Looking at the people who have a **high** salary, their car prices are

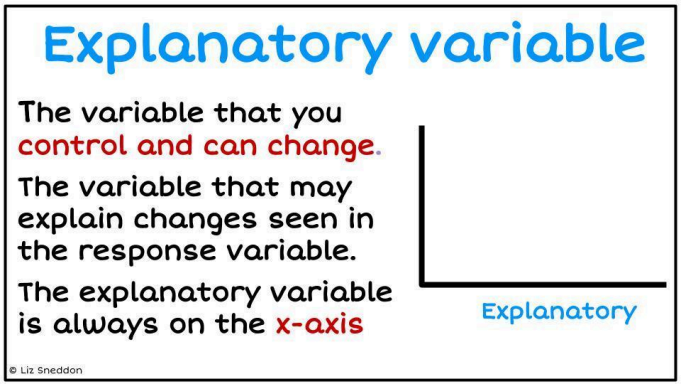
low / high. (circle correct answer)

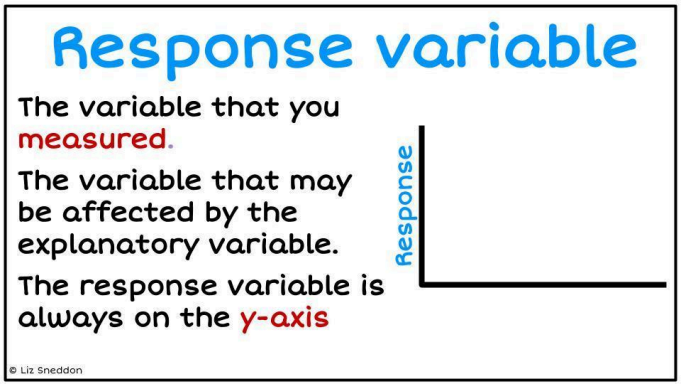
Explanatory & Response Variables

Given two numeric variables, we need to work out which one is the explanatory

variable, and which is the response variable.







Example:

Work out which is the explanatory variable and response variable for the following numeric variables:

- Amount of water a plant is given,
- The height of the plant.

Ask yourself, which way around makes sense

1) Does the amount of water given to a plant explain the height of the plant?
   OR
2) Does the height of the plant explain the amount of water given?

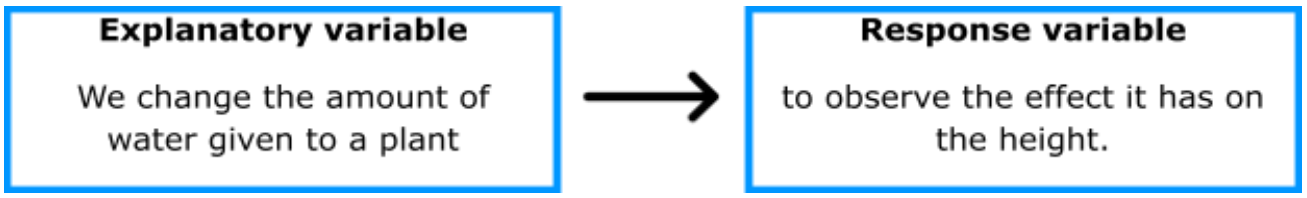
The first option is correct – we would **control or change** how much water a plant is given, and then we can **measure** the height.

This means:

Explanatory variable = amount of water given to a plant

Response variable = height of the plant.

Another way to think about it is:



Exercise 2:

Decide which variable is the explanatory and response variables for the following situations. Explain your reasoning.

1) The amount of sunlight that a plant gets and the height of the plant.

Explanatory Variable	
Response Variable	
Explanation why.	

2) The age of a person and their shoe size.

Explanatory Variable	
Response Variable	
Explanation why.	

3) The height of a roller-coaster and the speed it goes.

Explanatory Variable	
Response Variable	
Explanation why.	

4) The distance to drive to a location and the time it takes to drive there.

Explanatory Variable	
Response Variable	
Explanation why.	

5) The speed an object falls at when it is dropped and its weight.

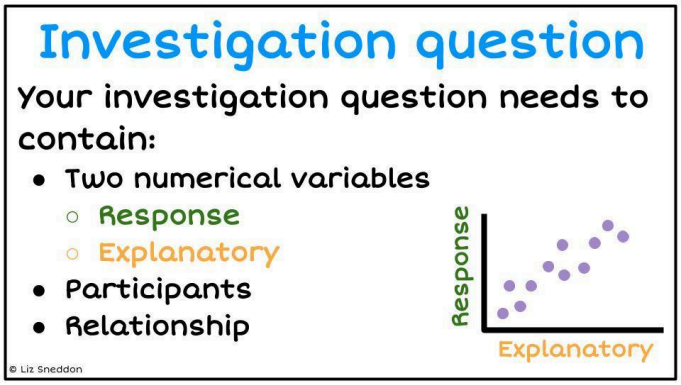
Explanatory Variable	
Response Variable	
Explanation why.	

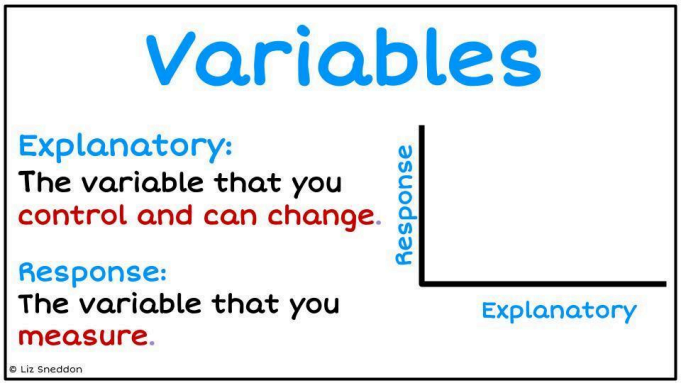
6) The daily temperature and the number of ice-cream cones sold.

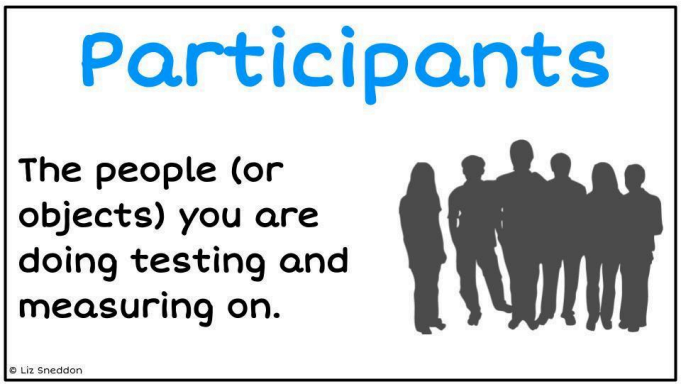
Explanatory Variable	
Response Variable	
Explanation why.	

Problem

The investigation question is to investigate **if there is a relationship** between two numeric variables. Sometimes we use the word **correlation** when describing a relationship.







Example:

I wonder if there is a relationship between a person's foot length and hand span for ākonga at your school?

Exercise 3:

Write investigation questions for the following datasets.

1) **Participants:** a random sample of top rugby players on the website: 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 meters (m)

There is one possible relationship question that you can write from the dataset above. Write this below.

Numerical variable 1 (and units)	
Numerical variable 2 (and units)	
Investigation question:	

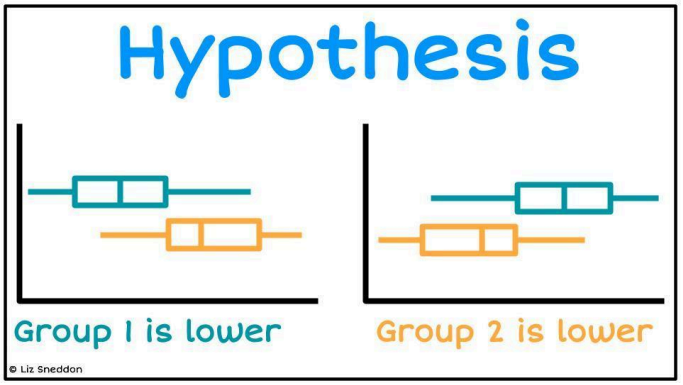
2) **Participants:** the dataset is a sample of new vehicles sold in America in 1993.

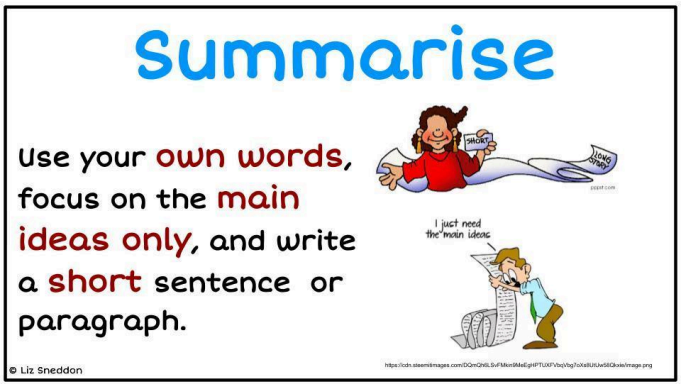
Variable	Description
Origin	America or Foreign
Price	The price of the car in US$ (thousands of dollars)
City	Fuel efficiency of the car in the city (miles per gallon)
OpenRoad	Fuel efficiency of the car on highways (miles per gallon)
Drive Train	Front Wheel Drive or Rear Wheel Drive
Engine Size	The size of the engine (litres)
Manual Transmission	Yes (manual) or No (Automatic)
Weight	The weight of the car (kg)

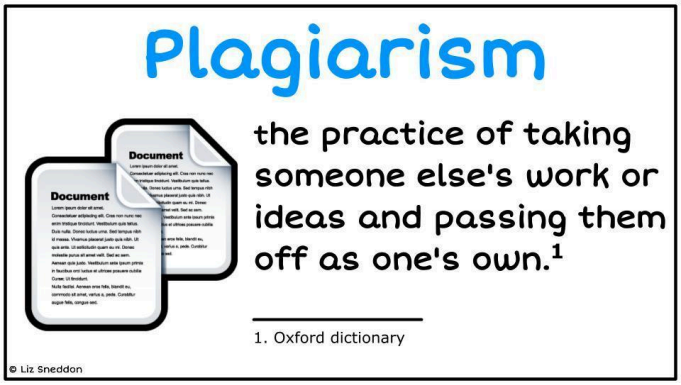
Write at least 3 different investigation questions using the variables above.

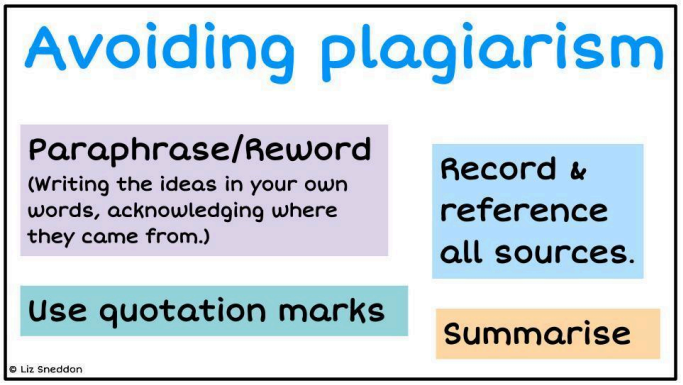
Hypothesis

Once you have identified what type of investigation you can do, the next step is to form a hypothesis, using contextual knowledge and a little research.

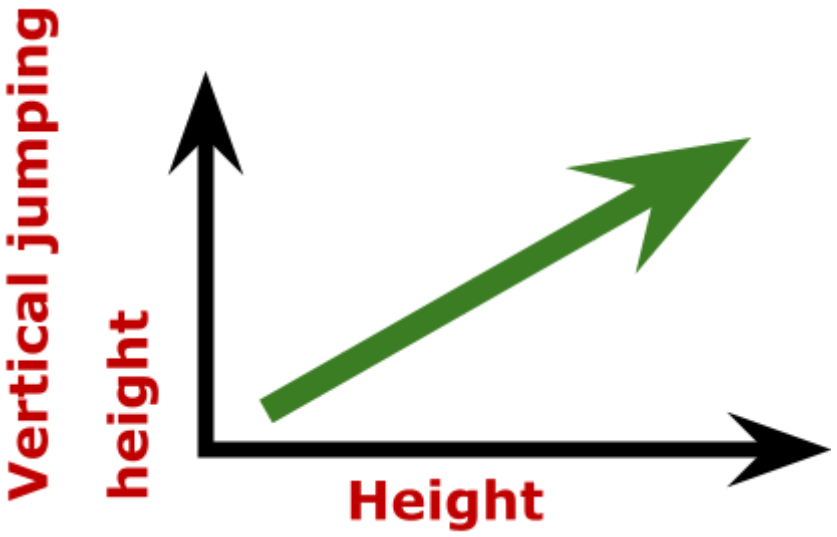








Example:

Investigation question:	I wonder if there is a relationship between the vertical height that ākonga can jump and their height, for Year 11 ākonga in our maths class.
Identify Variable 1:	Vertical height that ākonga can jump
Identify Variable 2:	Height of ākonga
Draw a graph of your Hypothesis: (Include axis labels)	
Hypothesis:	I hypothesise that as ākonga grow taller, that they will be able to jump vertically higher. Because taller ākonga have longer legs, and therefore they have a higher centre of gravity, and with longer legs I expect that they will be able to get more momentum to jump higher.

Exercise 4:

1)

Investigation question:	I wonder if there is a relationship between the height of ākonga and the percentage of shots that go into the basketball hoop, for ākonga in our class.
Identify Variable 1:	
Identify Variable 2:	
Draw a graph of your Hypothesis: (Include axis labels)	

Hypothesis:

2)

Investigation	I wonder if there is a relationship between the height of bottles and the volume of liquid that the bottle

question:	holds, for the bottles in the class collection.
Identify Variable 1:	
Identify Variable 2:	
Draw a graph of your Hypothesis: (Include axis labels)	

Hypothesis:

Plan / Whakamahere

In the first workbook, Exploring data, we

Organising data

	Explanatory variable	Response variable

covered a lot of information about the different data collection methods, including primary and secondary data, sources of variation and types of studies.

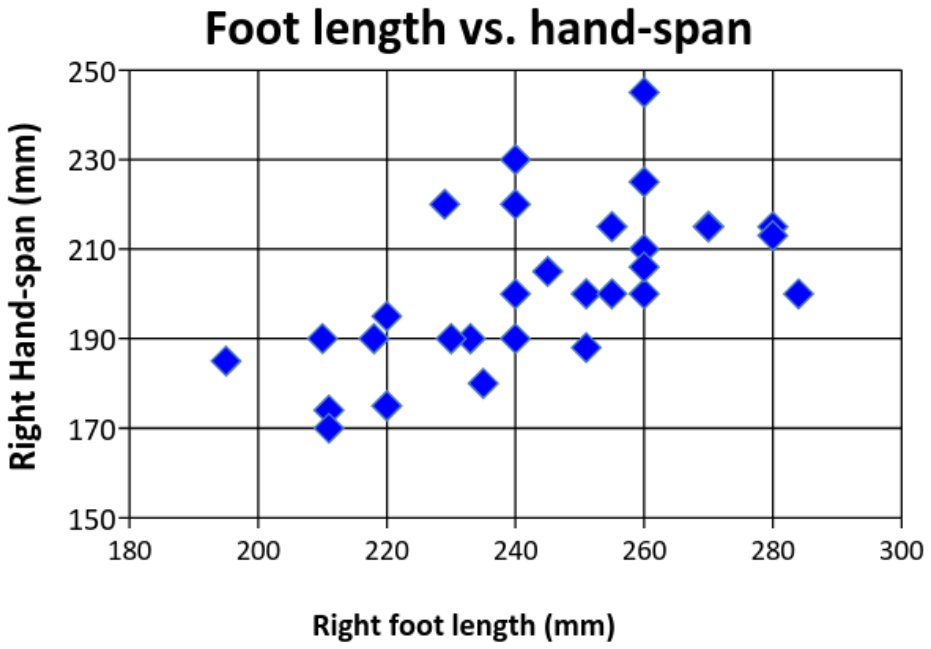
For this assessment, the requirement is to have at least 30 pairs of data.

Data / Raraunga

As you collect your data, don't forget to record any observations that might affect the accuracy and consistency of the data.

Example:

Data was collected from a sample of Year 11 ākonga at your school. These 30 ākonga measurements are shown on the graph below. An analysis and conclusion are given below.



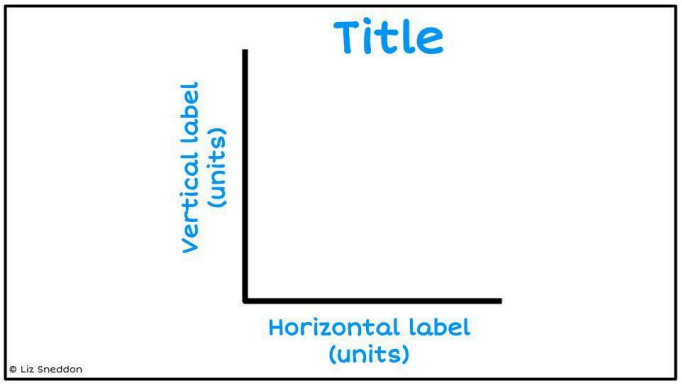
Improvements:

I could improve my investigation by getting ākonga to remove not just their shoe, but also their sock. This is because I noticed that some ākonga had quite thick socks, while some girls had very thin panty hose. So, it is possible that the measurements for foot length are not as accurate as they could be.

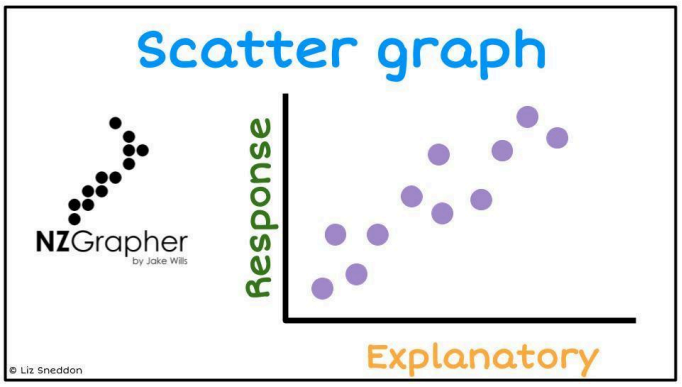
Drawing Scatter Graphs

Once the data is entered into NZGrapher, your next step is to draw a scatter graph, remembering to add:

- axis labels and units, and
- title.



You also need to make sure that you put the explanatory variable on the horizontal axis, and the response variable on the vertical axis.



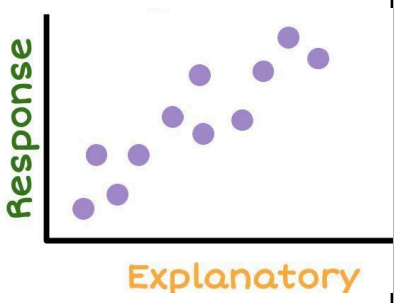
Drawing scatter graphs in NZGrapher

Step 1:

Go to the bottom of the screen, click on the **Graph Type** menu, and select **Scatter Graph**.

Step 2:

Select the **variables** for your graph. Check that you have them the correct way around.



(Variable 1 = Explanatory variable)

(Variable 2 = Response variable)

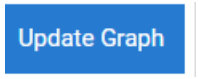
Step 3:

Add axes labels, units, and a title to the graph.

Step 4:

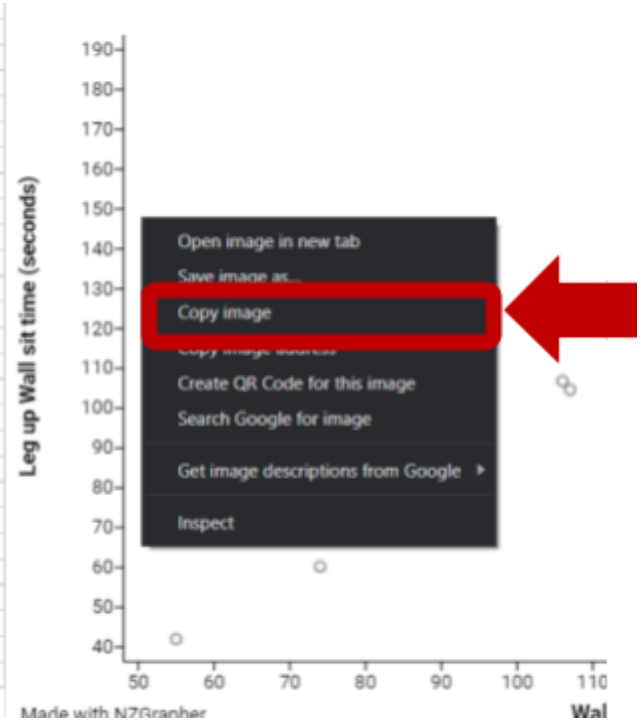
The last thing you need to do is to copy the graph and paste it into your

Once you have entered the labels and title, then click the **Update Graph** button (it's on the right-hand side at the bottom of the page).



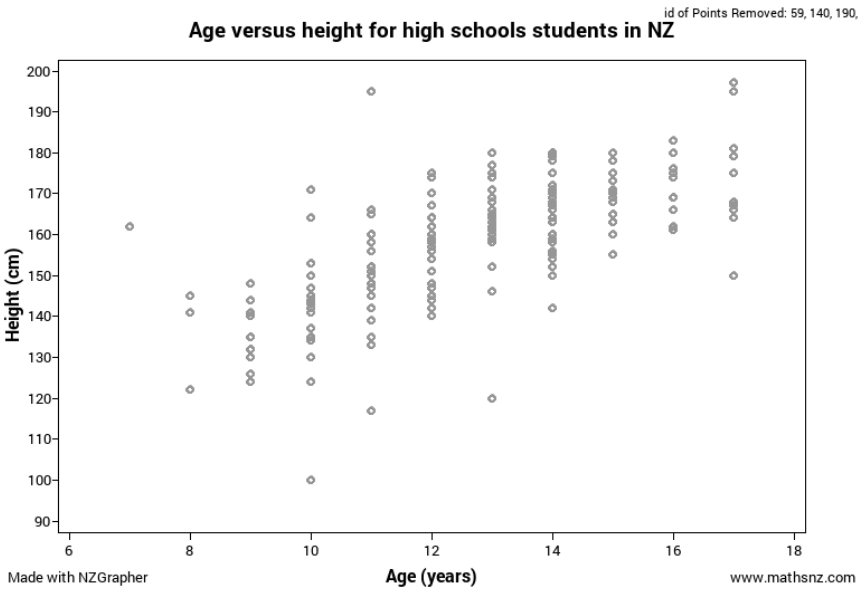
document.

To do this, move the mouse pointer so that it is over the graph. Then click on the **right mouse button** and select "**Copy image**"



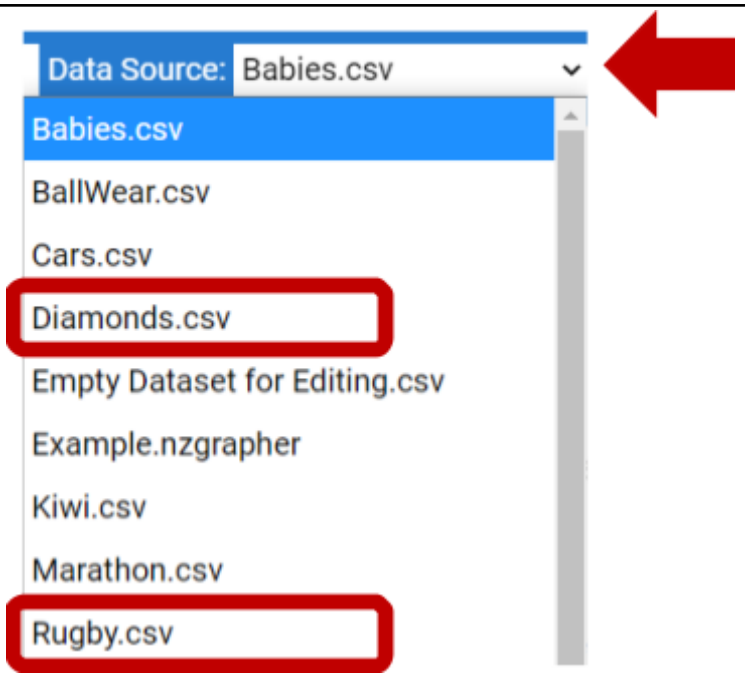
Then go to your document and paste the image.

Example:



Exercise 5:

Go to NZGrapher and explore one of the following two datasets. Then choose two numeric variables and draw a scatter graph. Include the axes labels and units, and title for the graph.

To find the datasets, go to the top right-hand side, click on the Data Source, and select the matching dataset below.	

1) Diamonds.csv

Explanatory Variable	
Response Variable	

2) Cars.csv

Explanatory Variable	
Response Variable	

3) Go back to your data from the helicopter. Enter this data into a spreadsheet, clean the data, copy the data into NZGrapher and draw a scatter graph.

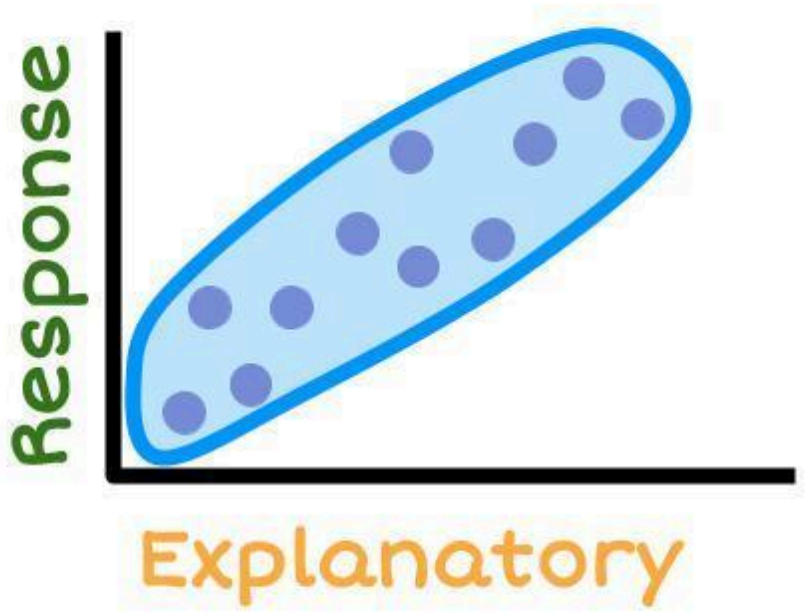
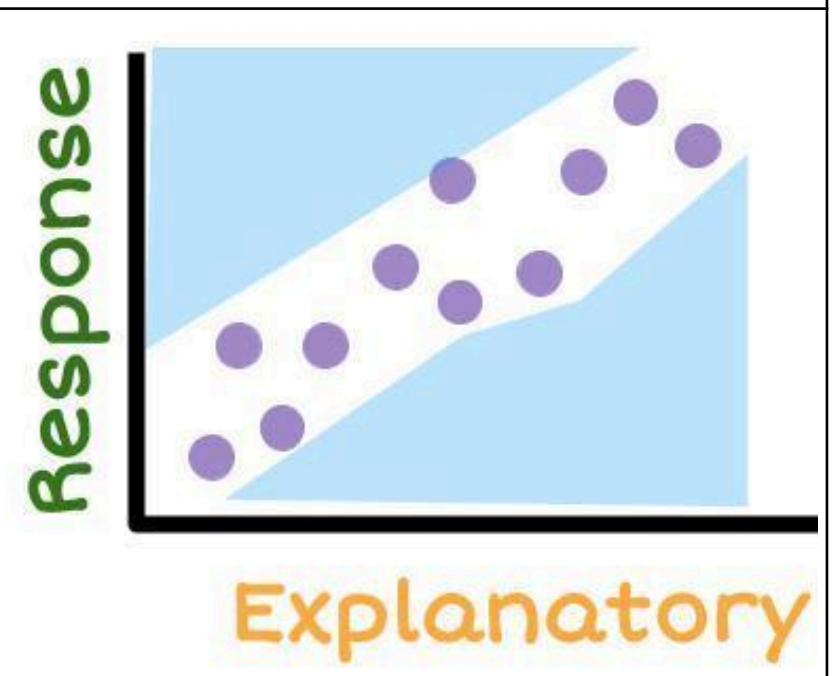
Explanatory Variable	
Response Variable	

Analysis & Conclusion

Once we have the scatter graph, the next step is to analyse any patterns and features we can see in the graph. To help us identify any patterns and features, we want to visually examine the graph. To do this we need to focus on the data.

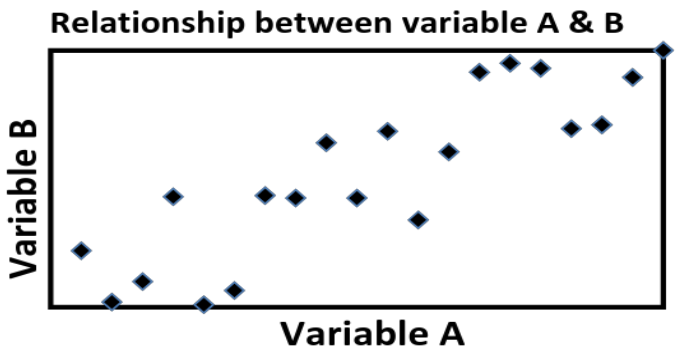
There are 2 methods that I'm going to suggest that you could use to help visually

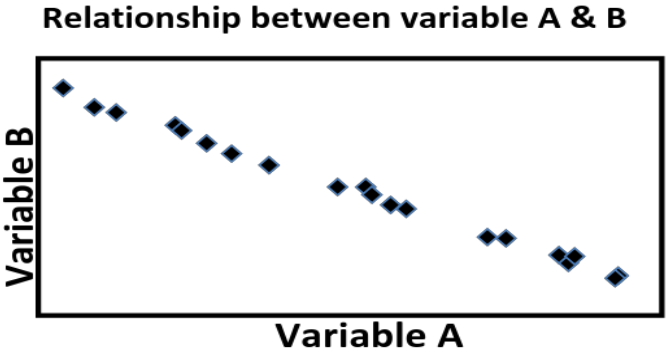
focus on the data. You can choose to use either of these.

Method 1: Shading in. Draw a blob around the data (you can ignore obvious outliers), the shade the area **inside** the blob. Or you can draw a line above and a line below (like train tracks) and shade the area **between** the lines.	
Method 2: Shading out. Shade the areas where there is **not** any data.	

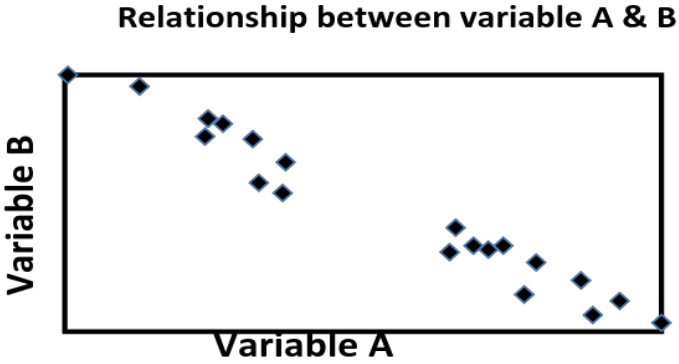
Exercise 6:

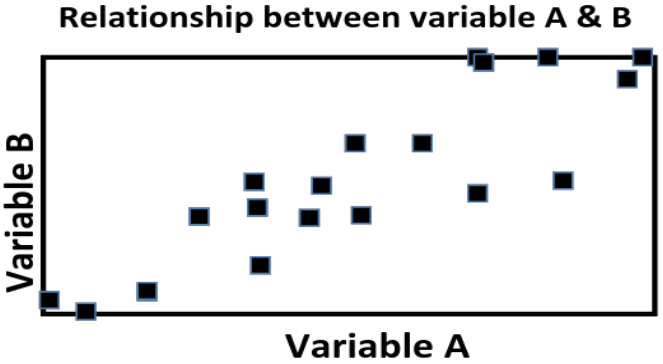
1) Using a highlighter, shade **IN** the following graphs.





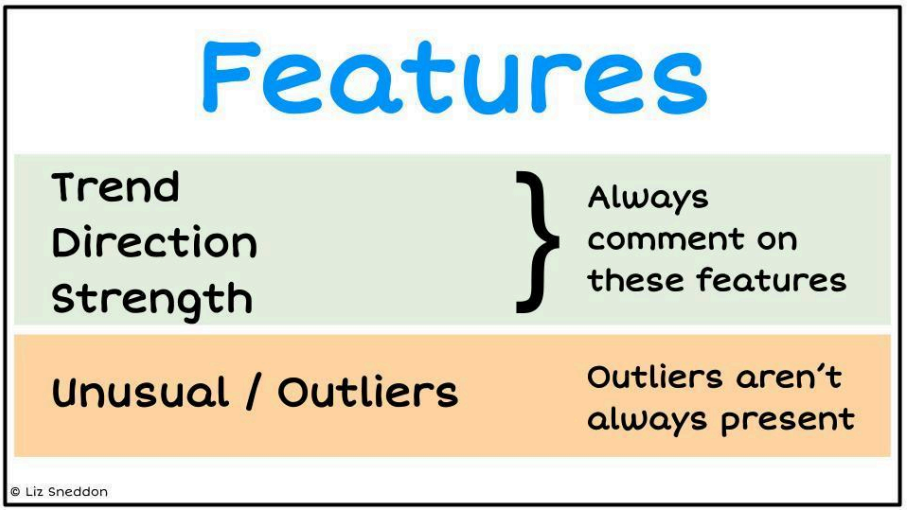
2) Using a highlighter, shade **OUT** the following graphs.





Features

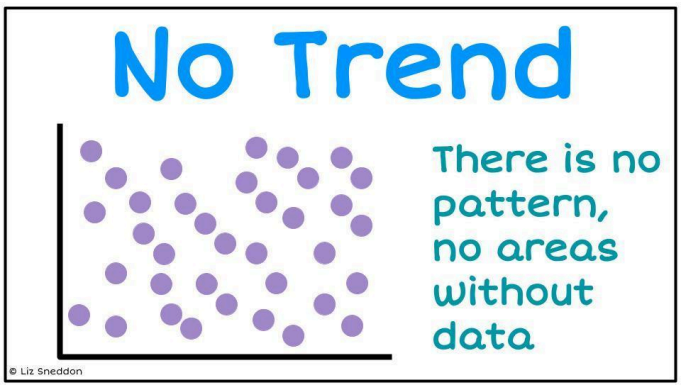
Once we have a visual picture of the data, we want to identify different features.



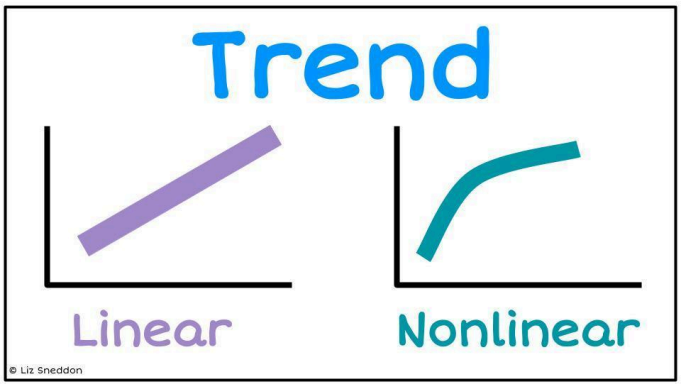
Let's look at these features one at a time, and then we can put it altogether

Trend

First, we need to decide whether there is a trend pattern.

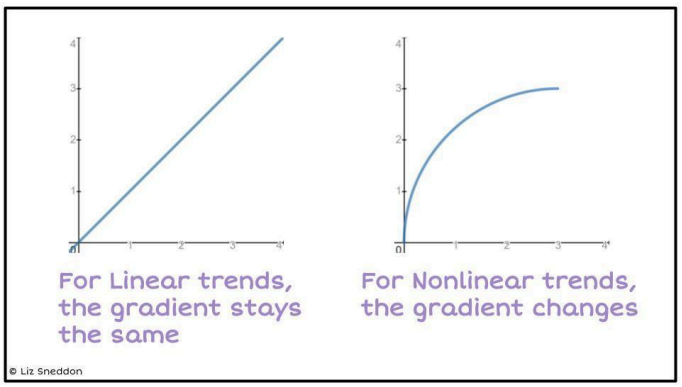


If there is a pattern we need to decide if it looks more like a linear or non-linear pattern (curved).

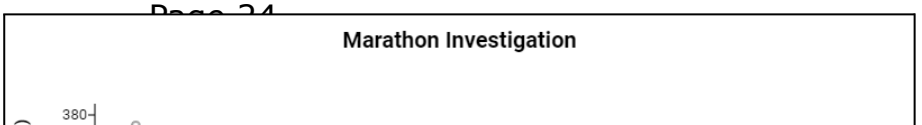


To justify or explain why we make this decision, we need to think about whether the gradient / slope is steady / constant, or whether the gradient is changing.

(The gradient measures how steep something is.)

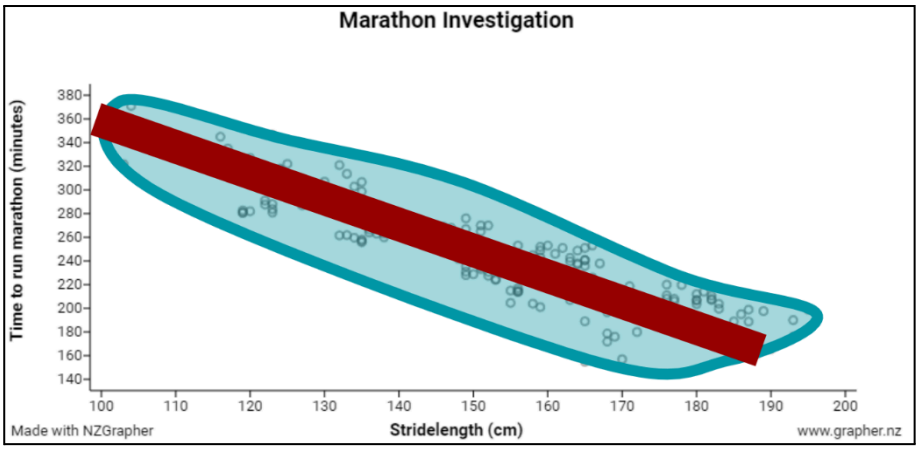


Example 1:



The graph shows athletes in NZ who ran a marathon, and their stride length and the length of time it took them to complete the marathon.

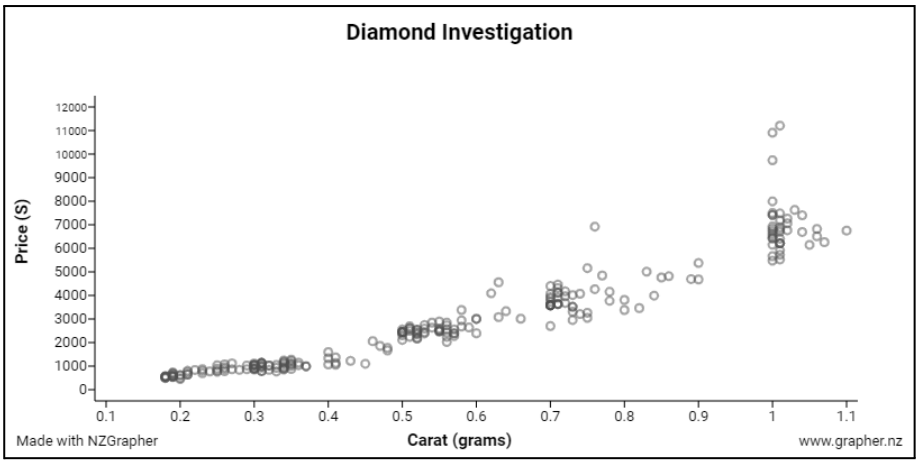
To identify if the pattern is linear or nonlinear, draw a **blob** around the data (excluding any outliers), and draw a **line through the middle of the blob pattern** to see whether that line is linear or nonlinear.



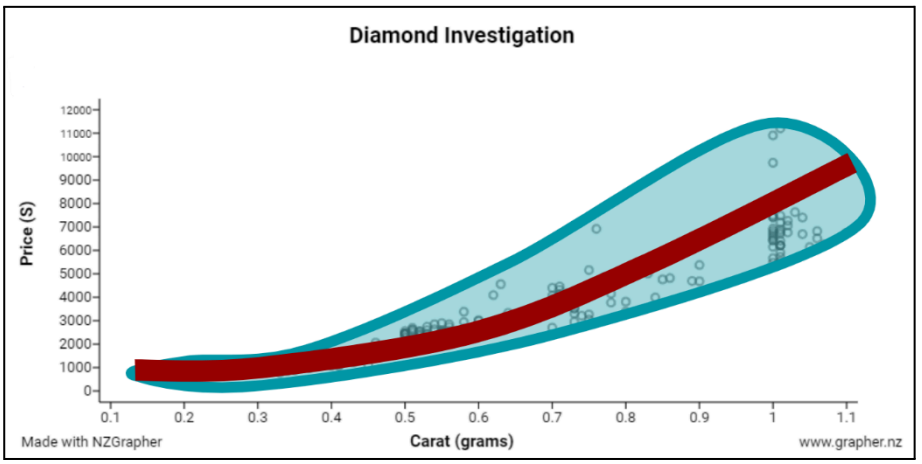
There is a linear relationship between the stride length and the time it takes athletes to complete a marathon. It is linear, because as the stride length increases, the marathon time decreases at a steady rate (e.g. constant gradient).

Example 2:

The graph shows the weight (number of carats) of diamonds and their price from a Singapore based retailer.

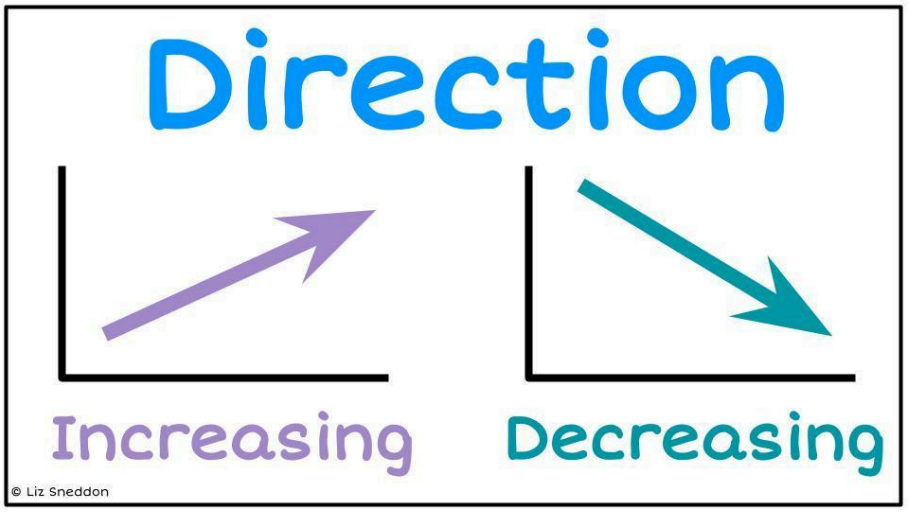


There is a nonlinear relationship between the weight (carat) of diamonds and the price of diamonds. It is nonlinear, because as the number of carats increases, the price increases at an increasing rate (e.g. increasing gradient).

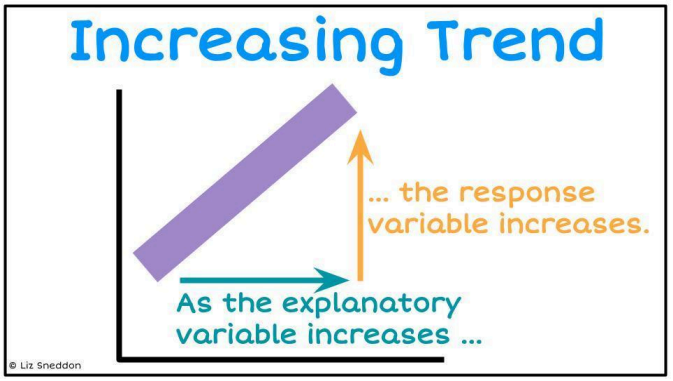


Direction

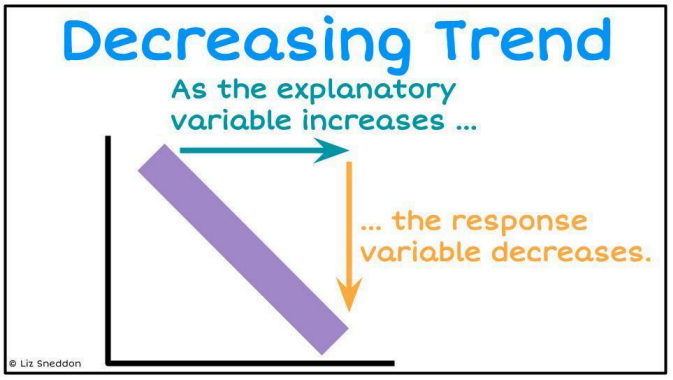
Once we know that there is a trend, the next step is to identify whether the trend is going up (positive or increasing) or going down (negative or decreasing).



To justify or explain an increasing trend, we always discuss (in context) how as the **explanatory variable** (horizontal variable) **increases**, the **response variable** (vertical variable) **increases**.

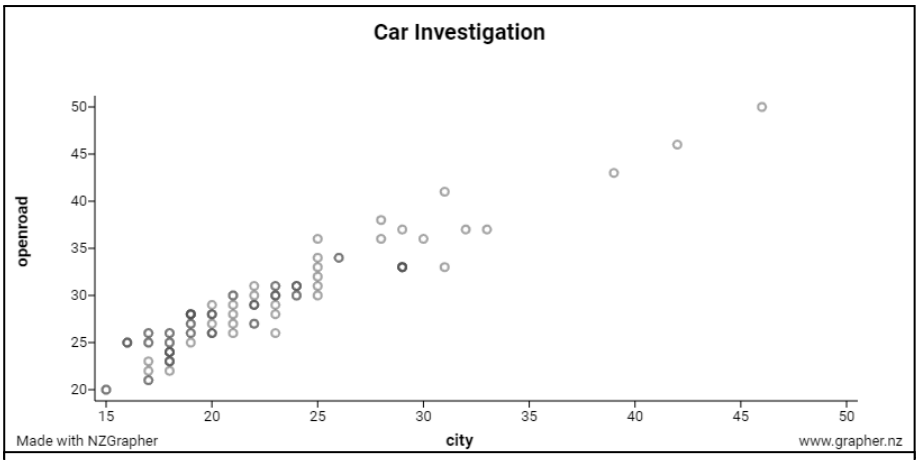


To justify or explain a decreasing trend, we always discuss (in context) how as the **explanatory variable** (horizontal variable) **increases**, the **response variable** (vertical variable) **decreases**.

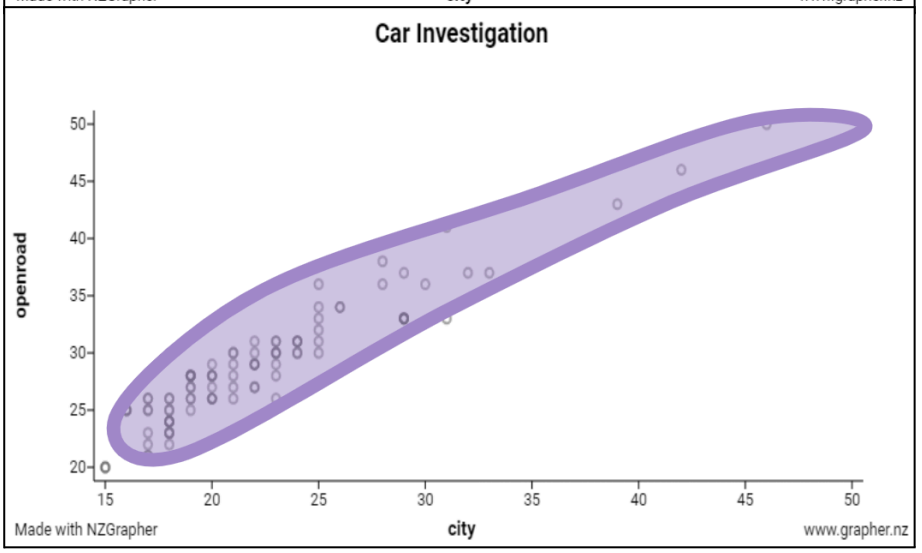


Example 1:

The graph shows new cars sold in 1993 in America, and the fuel efficiency (miles per gallon) when driving in the city, and the fuel efficiency when driving on open roads.



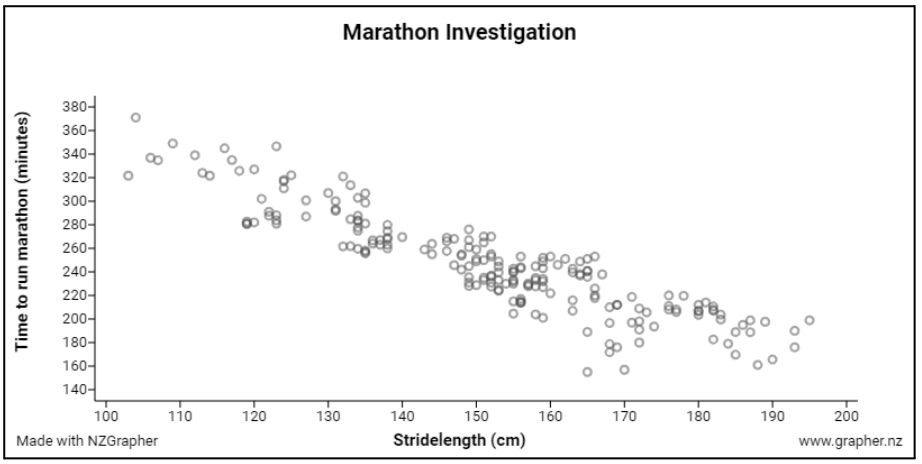
To identify the direction of the pattern, draw a **blob** around the data (excluding any outliers), and see whether the blob pattern is increasing or decreasing.



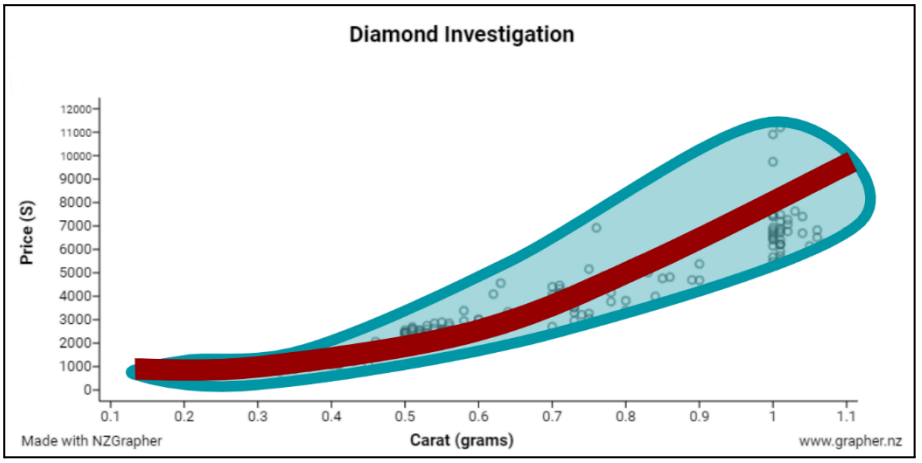
There is an increasing relationship between the fuel efficiency of cars driving in the city and fuel efficiency driving on open roads. As the fuel efficiency in the city increases, the fuel efficiency in open roads increases.

Example 2:

The graph shows athletes in NZ who ran a marathon, and their stride length and the length of time it took them to complete the marathon.

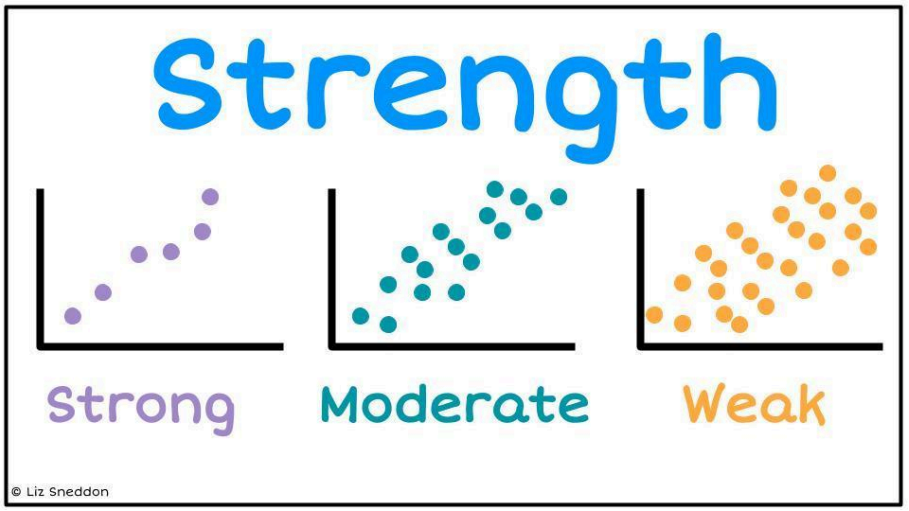


There is a decreasing relationship between the stride length and the time it takes athletes to complete a marathon. As the stride length increases, the time it takes to run a marathon decreases.



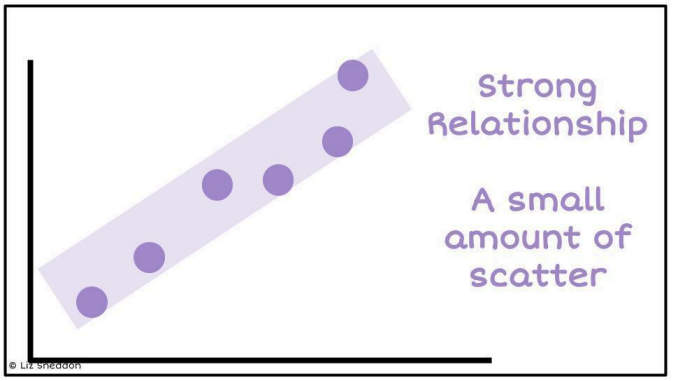
Strength

Next, we look at how tightly packed (or not) the data points are to the trend.



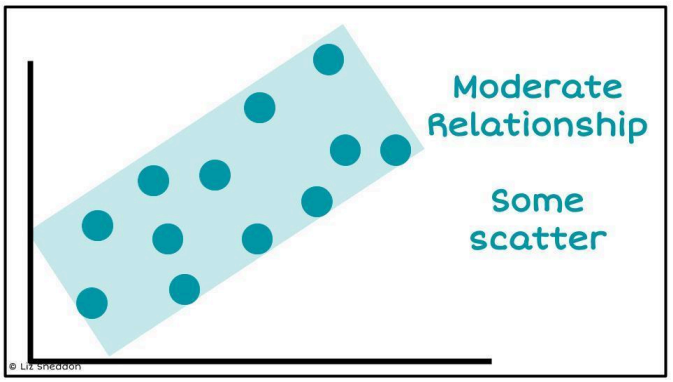
To justify or explain a **strong** relationship, we look at **how much scatter or variation** there is.

It's like using a thin paintbrush to paint the trend.



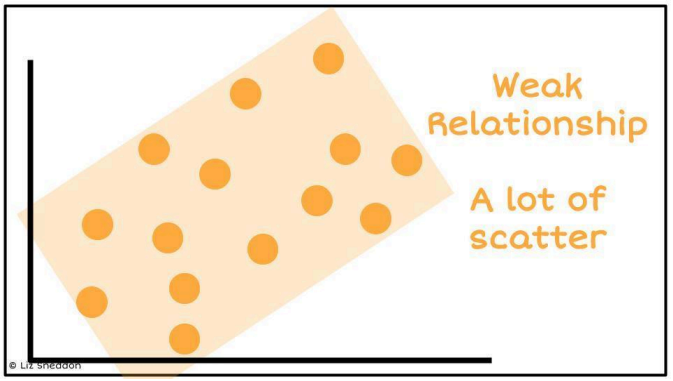
To justify or explain a **moderate** relationship, we look at **how much scatter or variation** there is.

It's like using a medium width paintbrush to paint the trend.



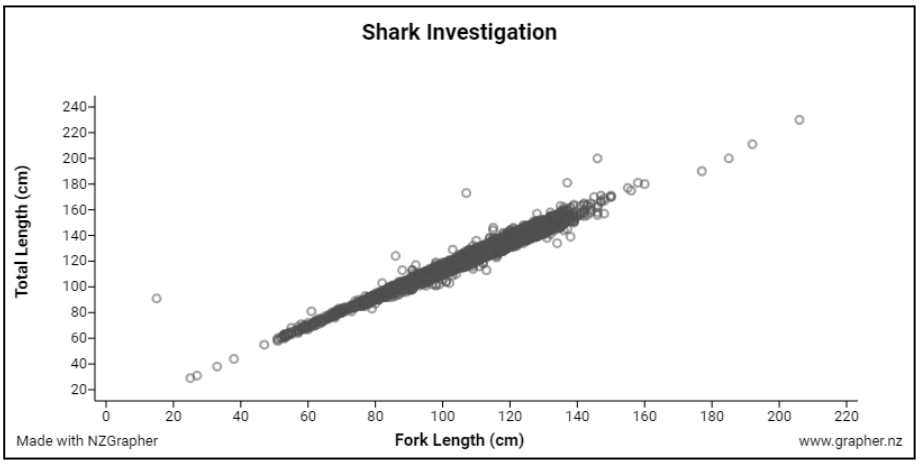
To justify or explain a **strong** relationship, we look at **how much scatter or variation** there is.

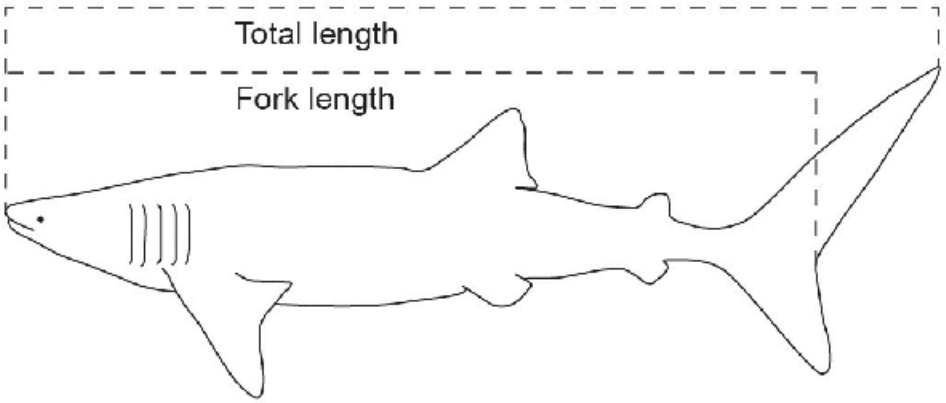
It's like using a wide paintbrush to paint the trend.



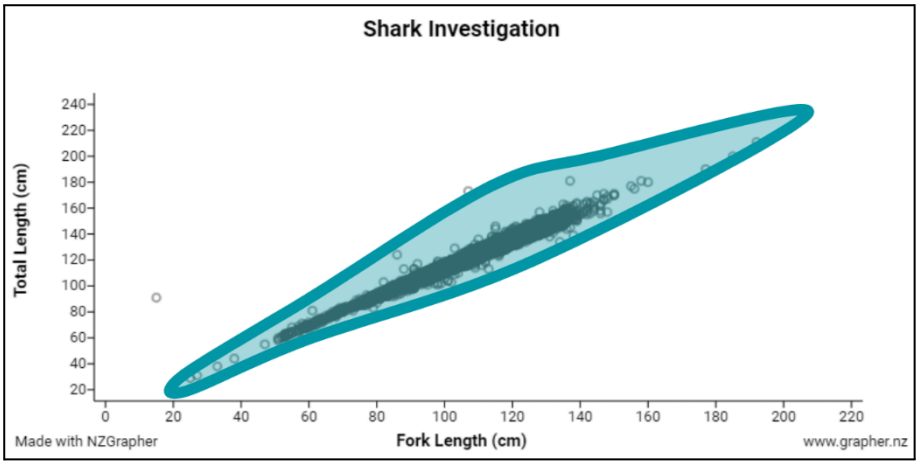
Example 1:

The graph shows a sample of sharks and the measurements of their fork length (distance from their head to where their tail splits into fins) and their total length.





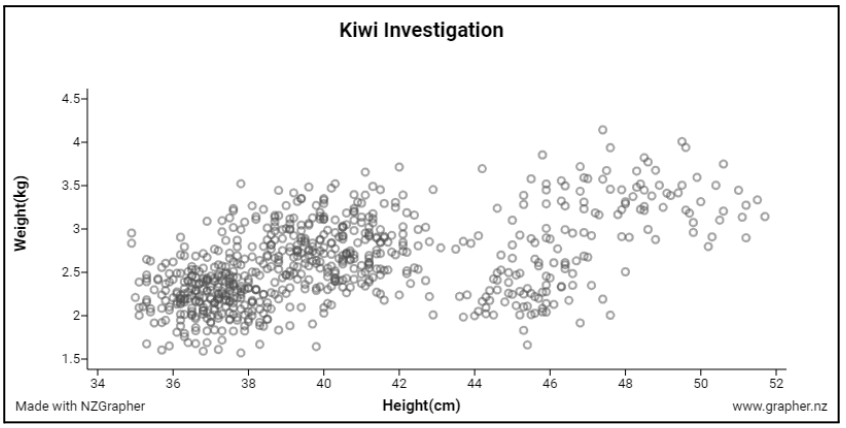
To identify the direction of the pattern, draw a **blob** around the data (excluding any outliers), and see whether the blob pattern is wide or narrow.



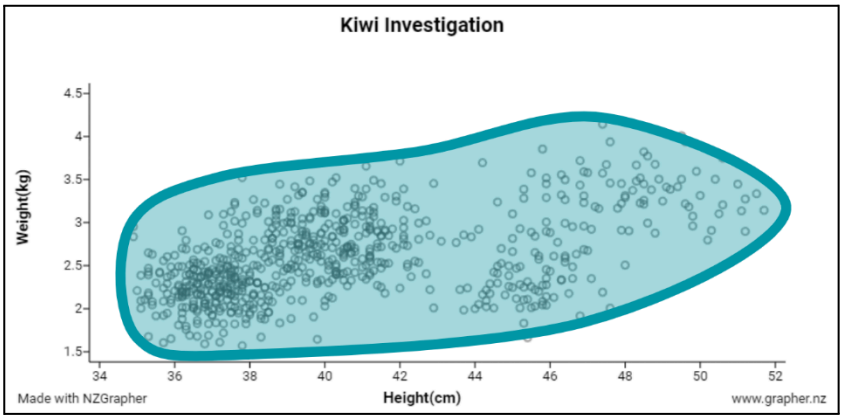
The relationship between the fork length and the total length of sharks is strong. The relationship is strong because there is only a small amount of scatter, and the data points are packed closely together.

Example 2:

The graph shows a sample of kiwi birds and their height and weight measurements.

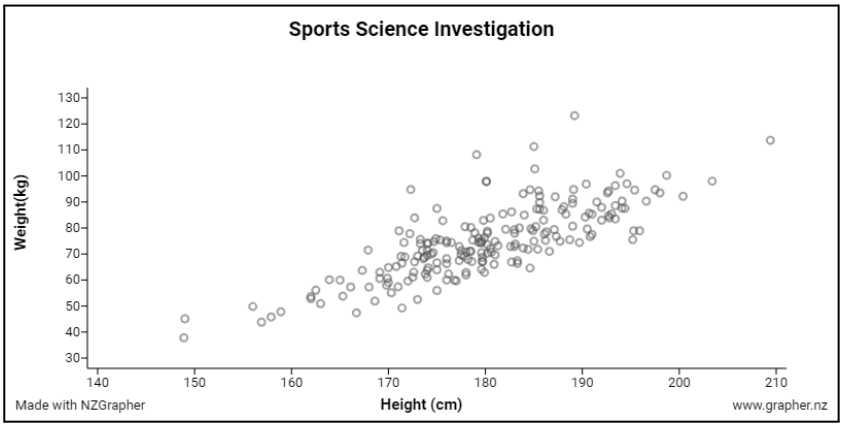


The relationship between the height and weight of kiwi birds is weak. The relationship is weak because there is a lot of scatter between the data points.

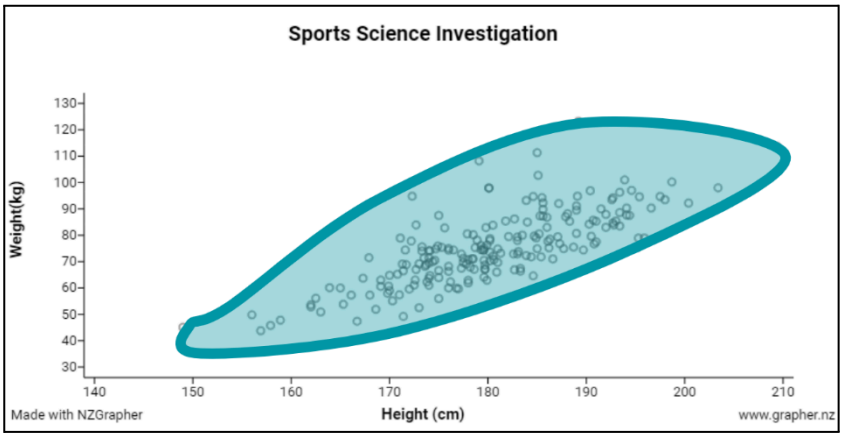


Example 3:

The graph shows a sample of athletes from the Australian Institute of Sport, and the height and weight measurements of these athletes.

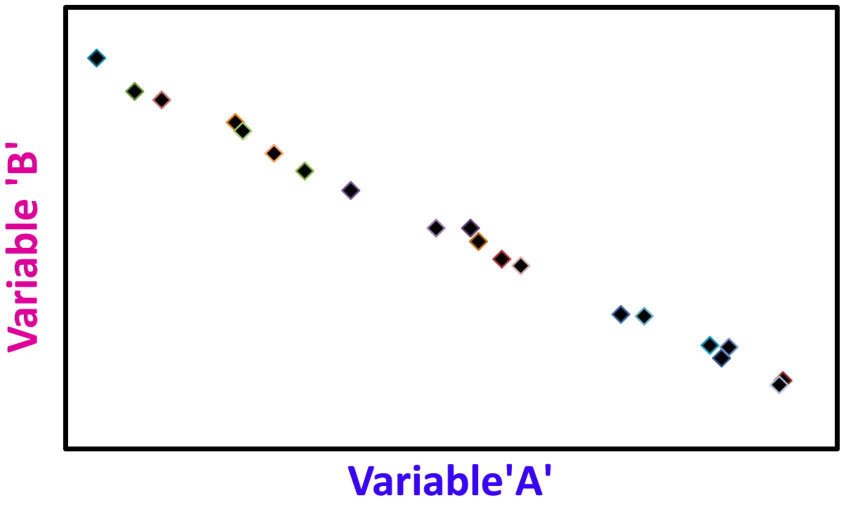


The relationship between the height and weight of athletes is moderate. The relationship is moderate because there is a moderate amount of scatter between the data points.



Exercise 7:

1) Shade the data. Then decide whether there is a relationship between variable A and B. If there is a relationship, decide the trend, direction, and strength.



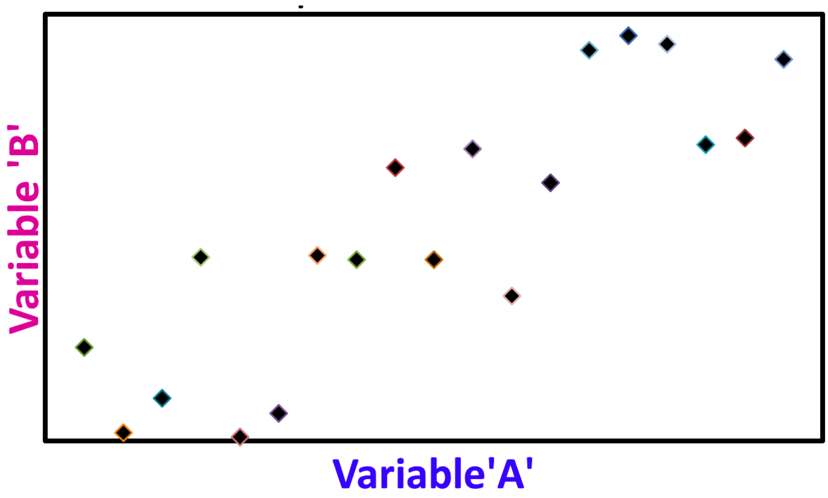
Relationship:
Yes / No

Trend:
Linear / Non-linear

Direction:
Positive / Negative

Strength:
Strong / Moderate / Weak

Outliers: (circle any points)
Yes / No



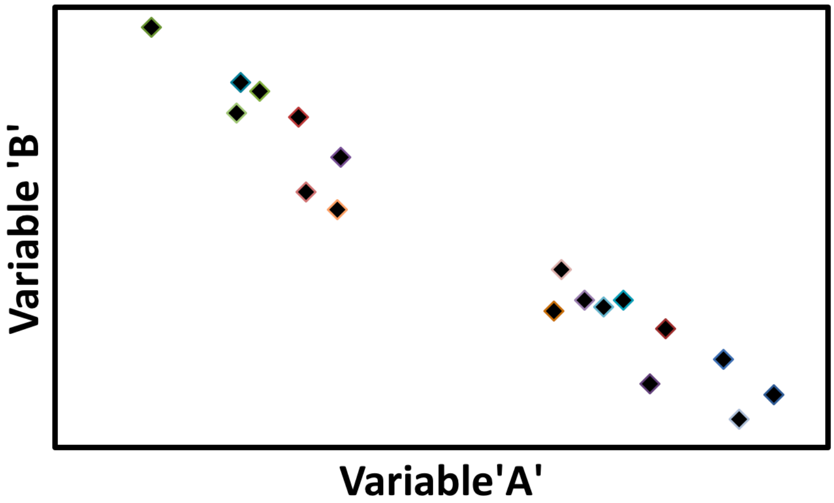
Relationship:
Yes / No

Trend:
Linear / Non-linear

Direction:
Positive / Negative

Strength:
Strong / Moderate / Weak

Outliers: (circle any points)
Yes / No



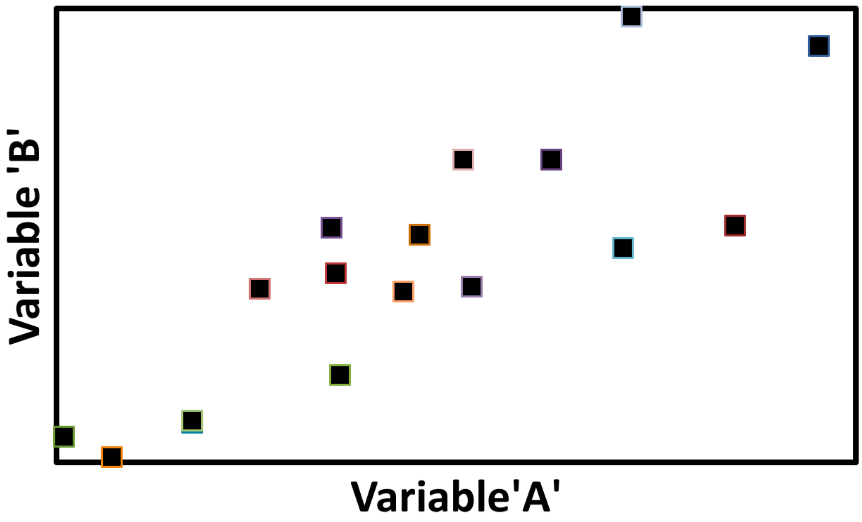
Relationship:
Yes / No

Trend:
Linear / Non-linear

Direction:
Positive / Negative

Strength:
Strong / Moderate / Weak

Outliers: (circle any points)
Yes / No



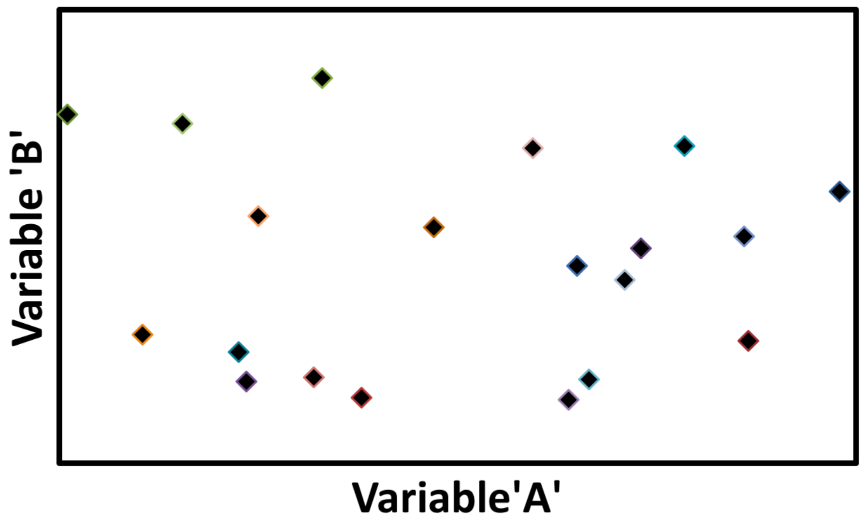
Relationship:
Yes / No

Trend:
Linear / Non-linear

Direction:
Positive / Negative

Strength:
Strong / Moderate / Weak

Outliers: (circle any points)
Yes / No



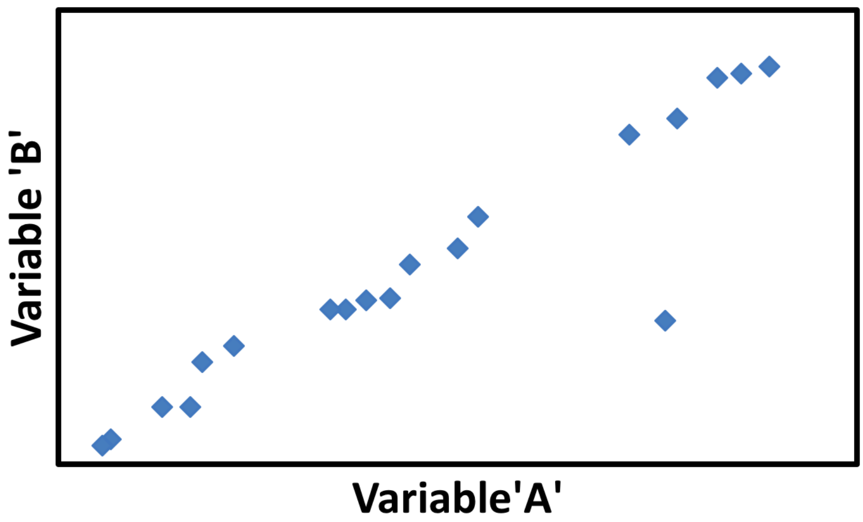
Relationship:
Yes / No

Trend:
Linear / Non-linear

Direction:
Positive / Negative

Strength:
Strong / Moderate / Weak

Outliers: (circle any points)
Yes / No



Relationship:
Yes / No

Trend:
Linear / Non-linear

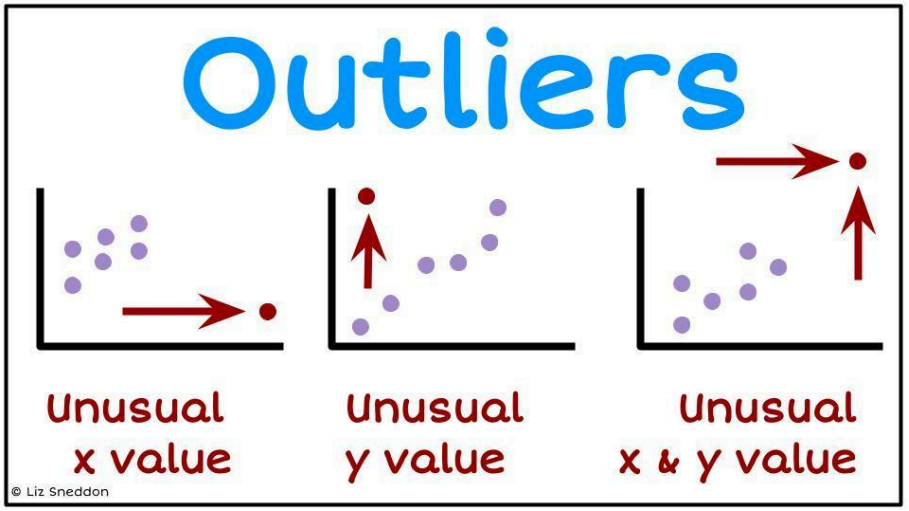
Direction:
Positive / Negative

Strength:
Strong / Moderate / Weak

Outliers: (circle any points)
Yes / No

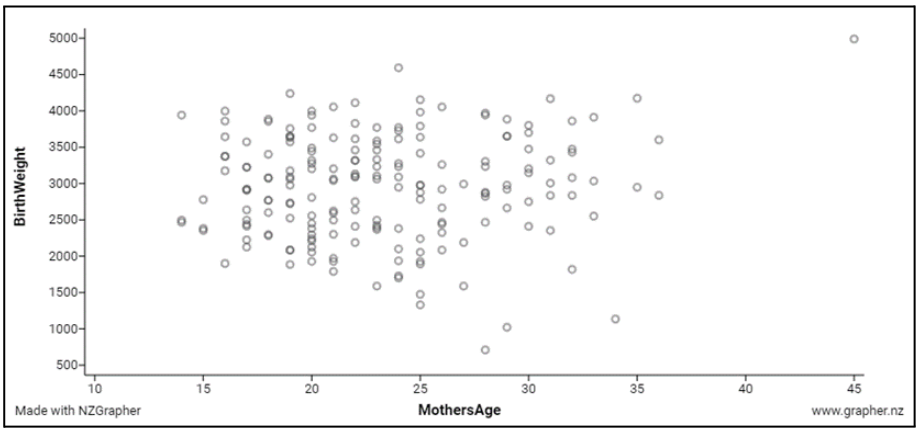
Outliers

An outlier is a point that is **A LONG WAY AWAY** from the trend pattern. Be careful when identifying these points, and make sure you state the coordinates.



Example 1:

The graph shows a sample of mothers and babies from Baystate Medical Centre, Springfield, Mass. during 1986. Data on the age of the mother and the birth weight of their baby was recorded.

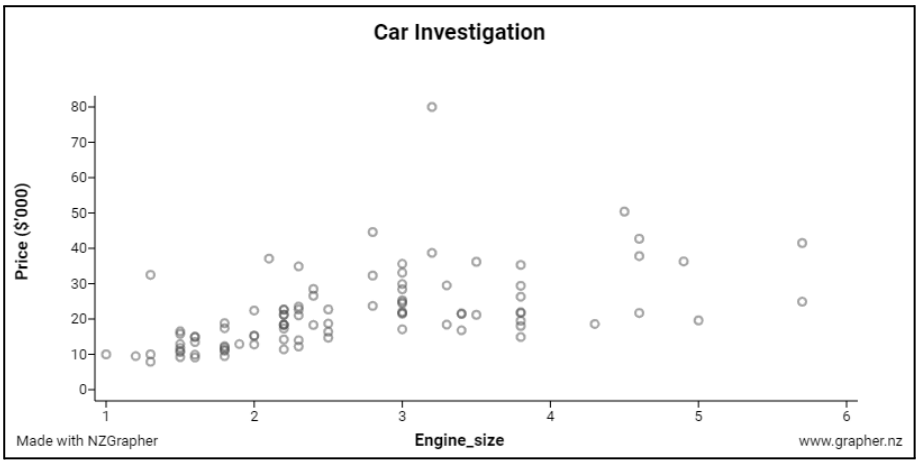


To identify any outliers, draw a **blob** around the data (excluding any points that are a long way away from the pattern), and identify the values of the outliers.

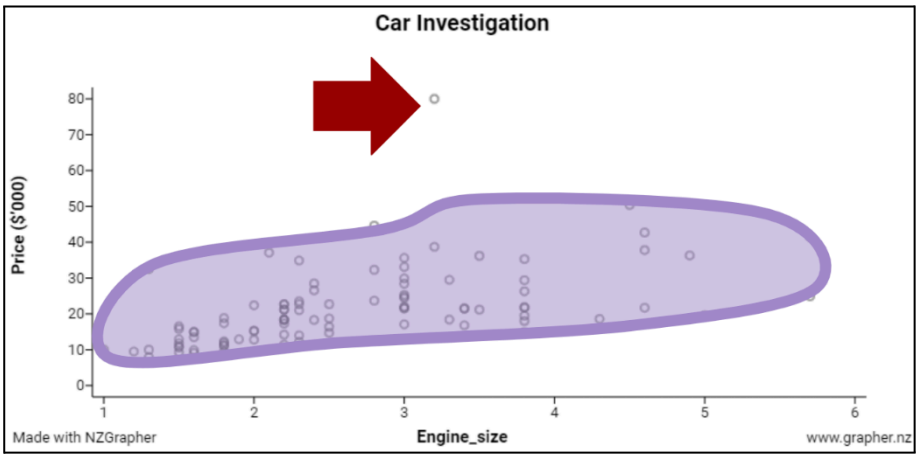
There is one outlier, a mother who is 45 years old with a baby whose birth weight is 5kg. This is unusual as the mother is older than expected, and the baby heavier than expected.

Example 2:

The graph shows new cars sold in 1993 in America, and data on the engine size and price was collected.



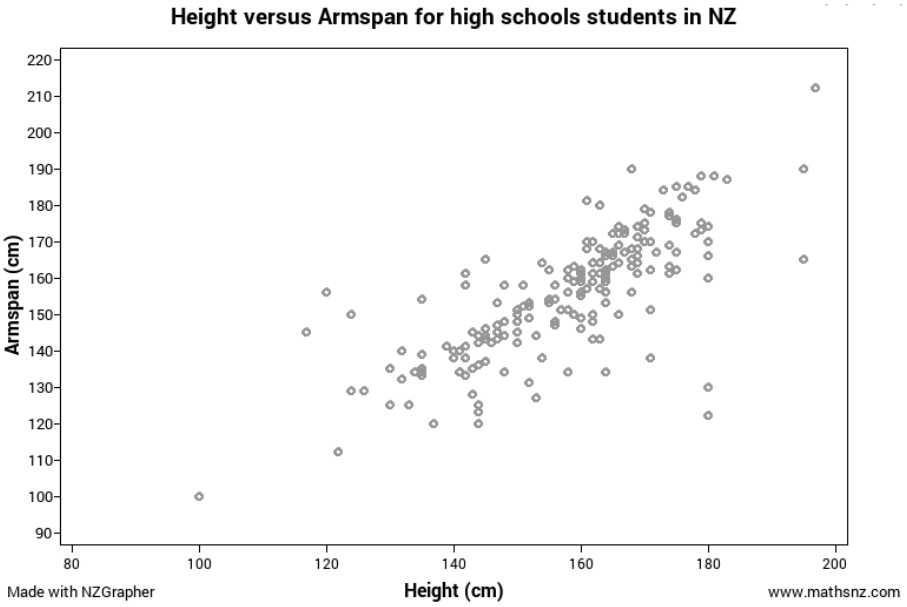
There is one outlier, a car with an engine size of just over 3L and a price of around $75,000. This is unusual as the more expensive than expected.



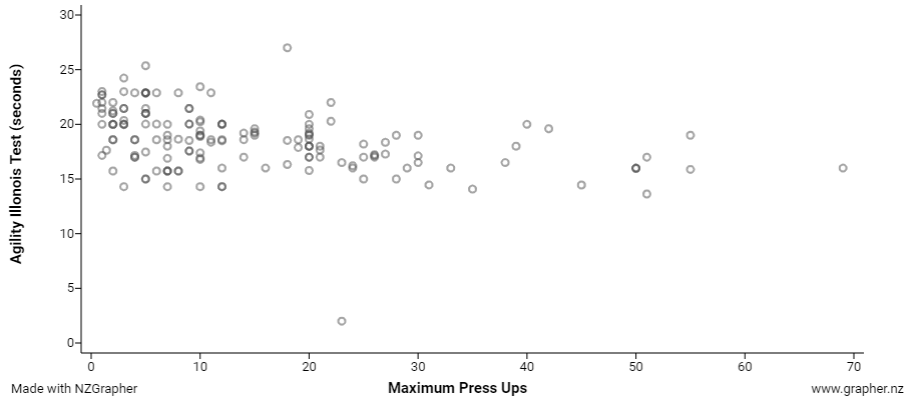
Exercise 8:

Identify any outliers in the following graphs. Highlight them on the graph, and then write a sentence estimating their coordinate points.

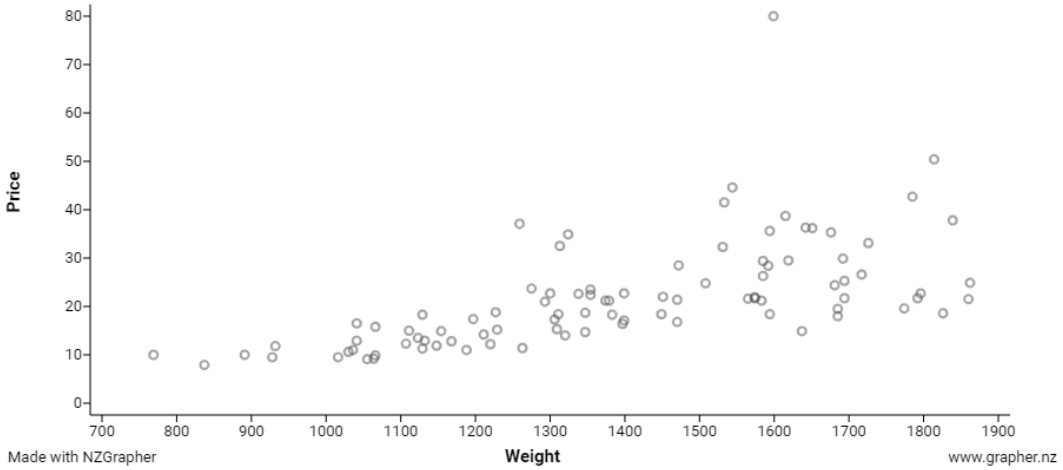
1)



2)

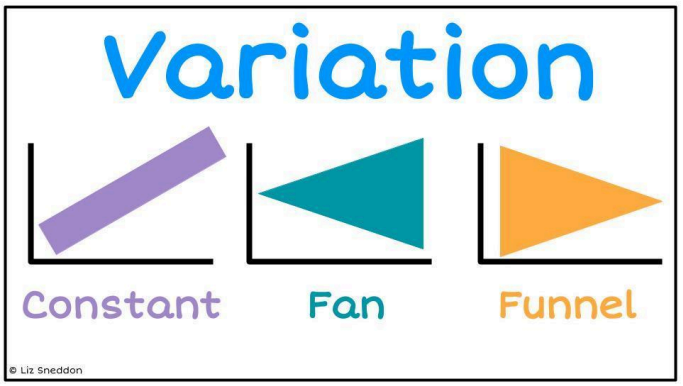


3)


Price
80
70
60
50
40
30
20
10
0
700
800
900
1000
1100
1200
1300
1400
1500
1600
1700
1800
1900
Weight
Made with NZGrapher
www.grapher.nz

Extra feature – Variation

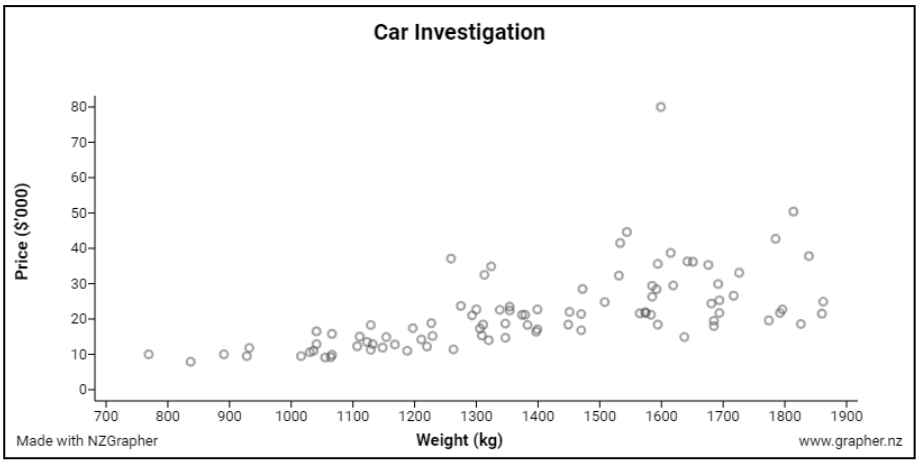
There is one additional feature that may be present in some scatter graphs, and that is looking at the pattern of whether the variation in the response variable is staying constant, increasing, or decreasing.



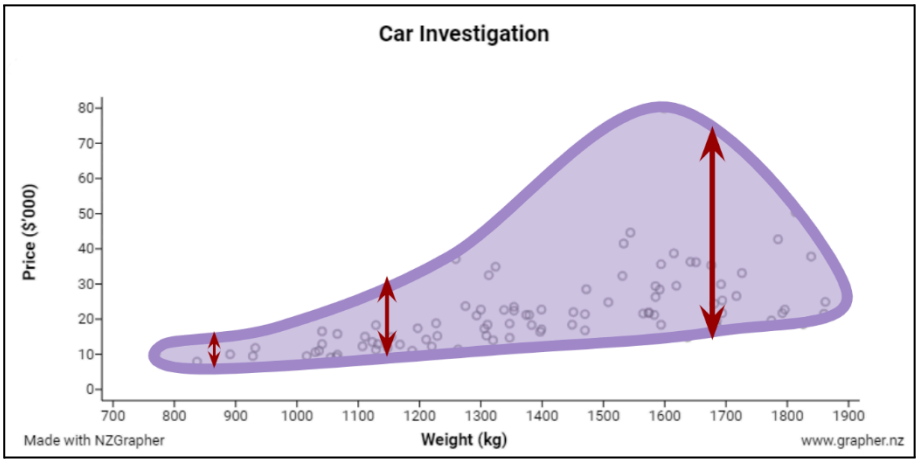
When looking at whether the variation has changed significantly, check whether it has doubled, tripled, etc. (Small changes are just due to random variation).

Example 1:

The graph shows new cars sold in 1993 in America, and the weight of the car and its price.



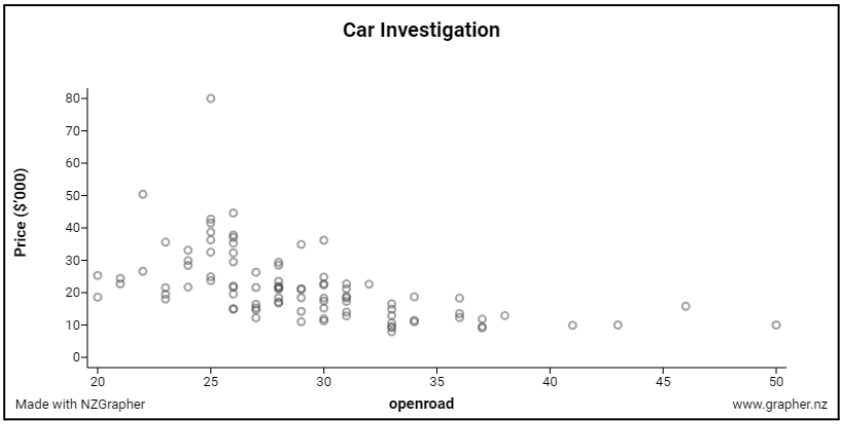
To identify the direction of the pattern, draw a **blob** around the data, and see whether the width of the blob pattern is roughly constant, increasing (fan effect) or decreasing (funnel effect).



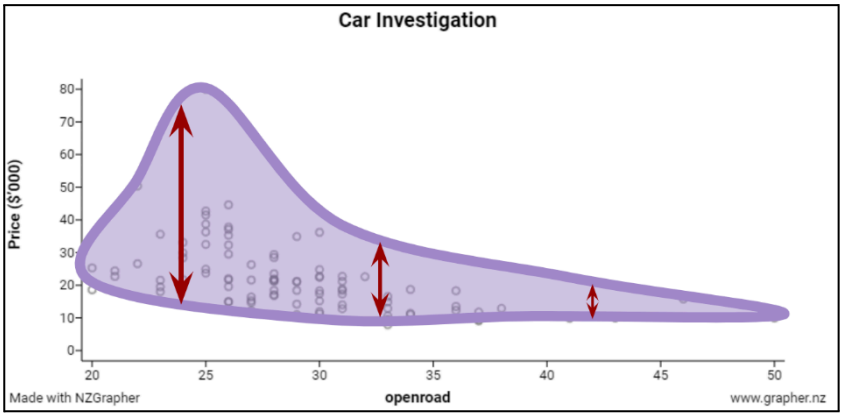
As the weight of the car increases, the variation in the price is increasing too.

Example 2:

The graph shows new cars sold in 1993 in America, and the fuel efficiency (miles per gallon) when driving in the city, and the fuel efficiency when driving on open roads.

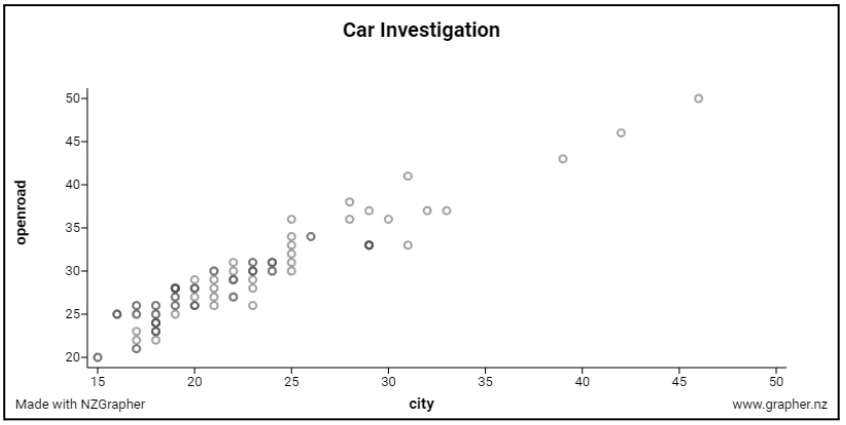


As the fuel efficiency of the car when driving on open roads increases, the variation in the price is decreasing.

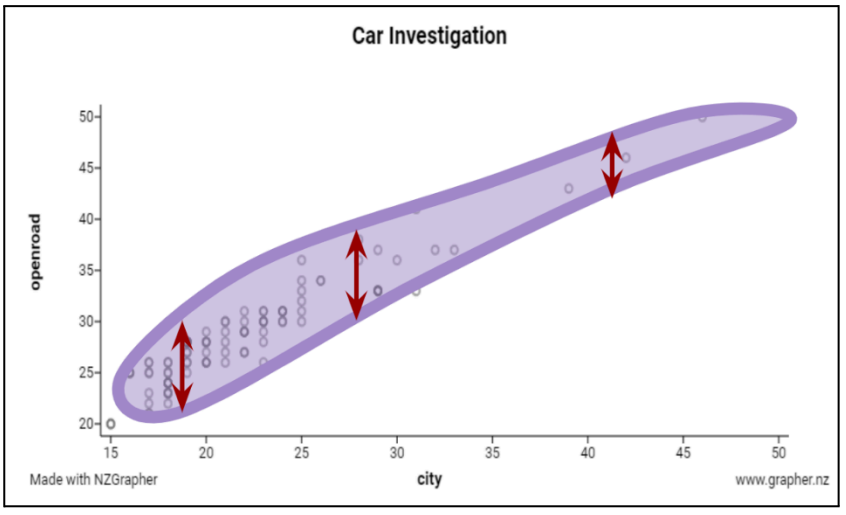


Example 3:

The graph shows new cars sold in 1993 in America, and the fuel efficiency (miles per gallon) when driving in the city, and the fuel efficiency when driving on open roads.

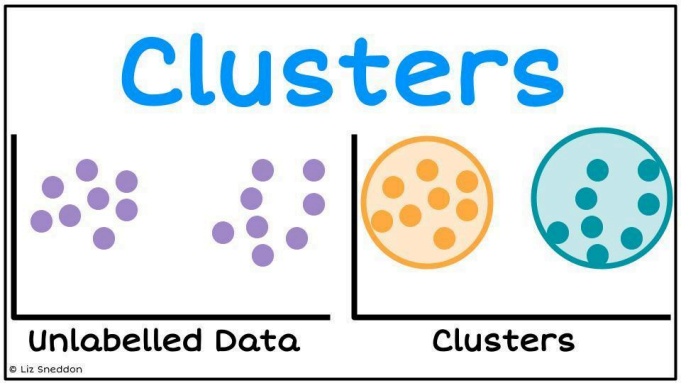


As the fuel efficiency of cars driving in the city increases, the variation in the fuel efficiency of cars driving on open roads stays roughly constant.



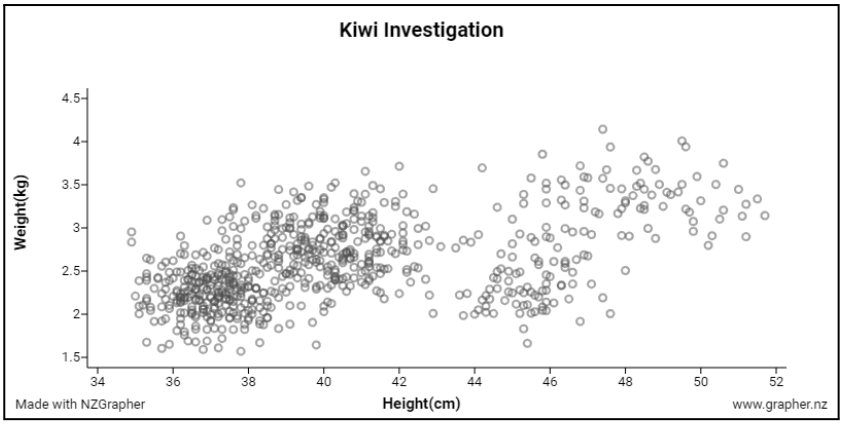
Extra feature – Clusters

You need to be careful when trying to identify clusters. Sometimes there will be gaps in the data just due to random variation.
Usually clustering occurs when there is an underlying categorical variable (such as sex (male/female), ethnicity, etc.)

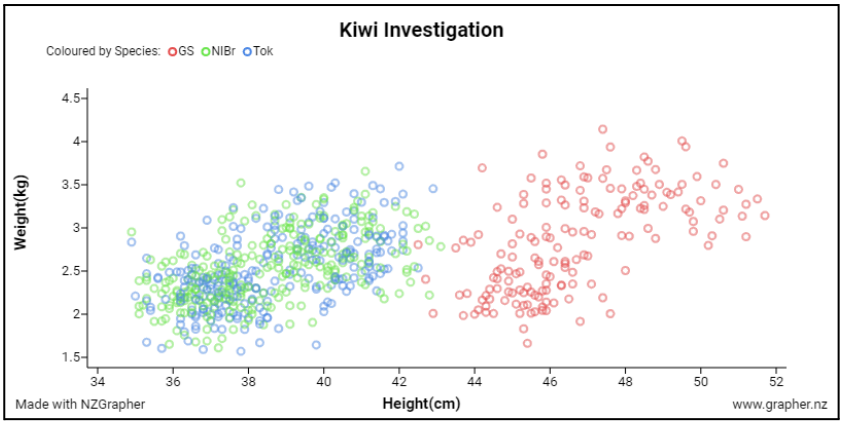


Example 1:

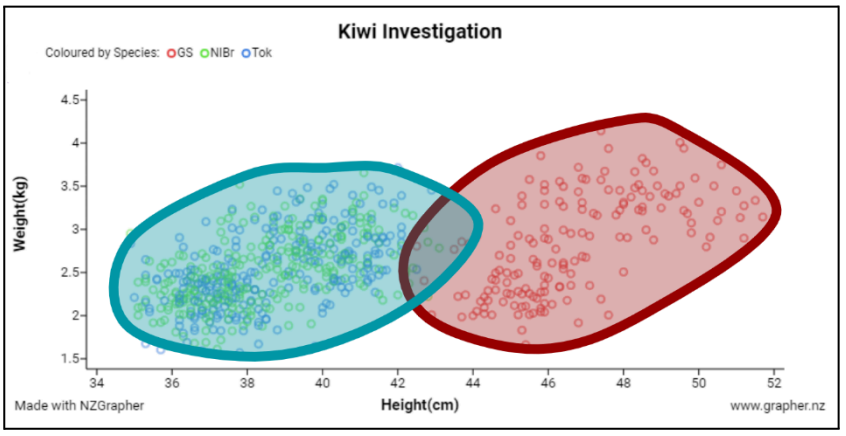
The graph shows a sample of kiwi birds and their height and weight measurements.



To identify the direction of the pattern, colour the categorical variable to notice any cluster patterns. Then draw a **blob** around the clusters.

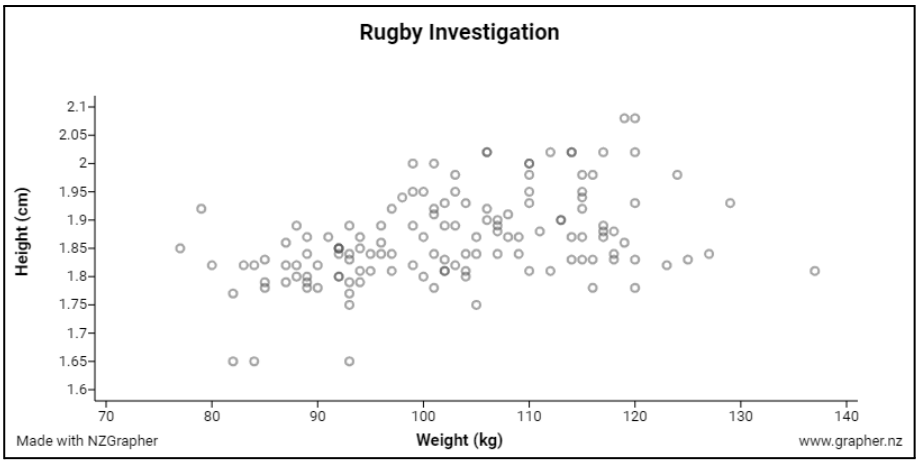


There are two clusters on the graph showing the relationship between height and weight of kiwi birds. One cluster is the Great Spotted kiwis, who are much taller but have similar weights to the other two species, the North Island Brown and Southern Tokoeka who form the second cluster having very similar weights and heights.

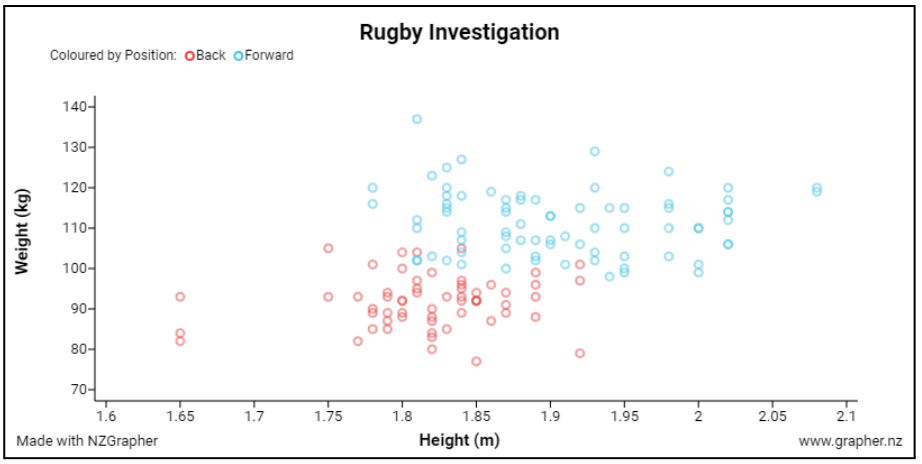


Example 2:

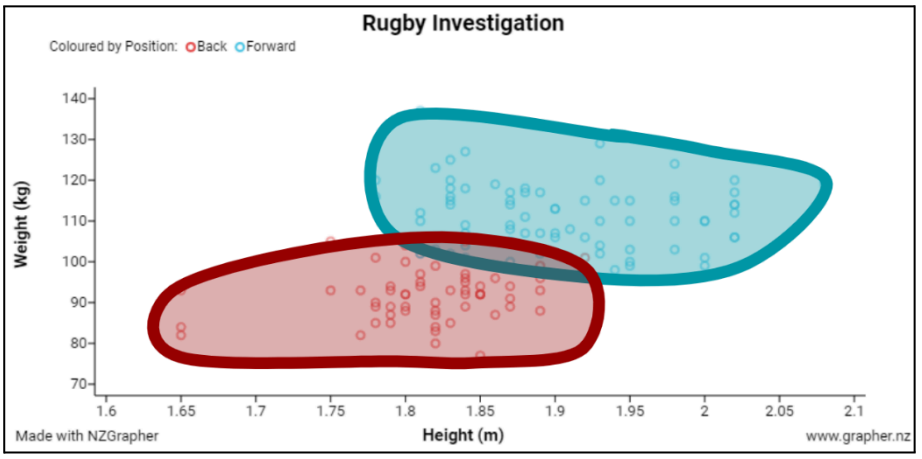
The graph shows a sample of top rugby players and their heights and weights.



To identify the direction of the pattern, colour the categorical variable to notice any cluster patterns. Then draw a **blob** around the clusters.



There are two clusters on the graph showing the relationship between height and weight of rugby players. One cluster is the Forwards, who tend to be taller and heavier. The second cluster is the Backs, who tend to be shorter and lighter.



Putting it altogether

Once you have identified the trend, direction, and strength of the relationship between the explanatory and response variables, we then need to write this up **in context**. Writing it in context means that you need to refer to both the **explanatory variable** and **response variable**.

Example:

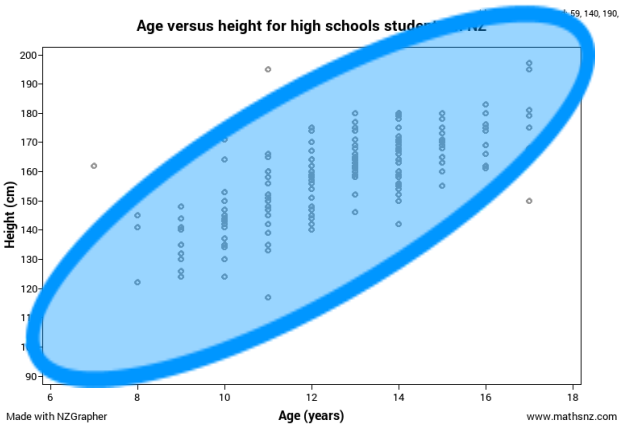
Trend:

Identify:

I notice that there is a **linear** trend between the age of high school ākonga and their height.

Justify:

It is a linear trend because there is a **steady slope**.



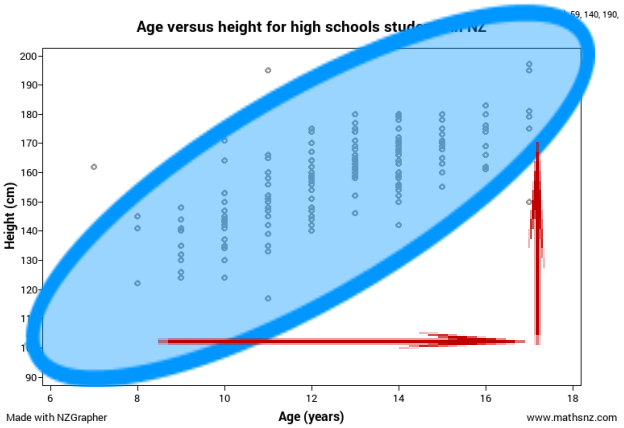
Direction:

Identify:

I notice that there is an **increasing** trend between the age and height of high school ākonga.

Justify:

It is an increasing trend because as the **age** of high school ākonga **increases**, the **height increases**.



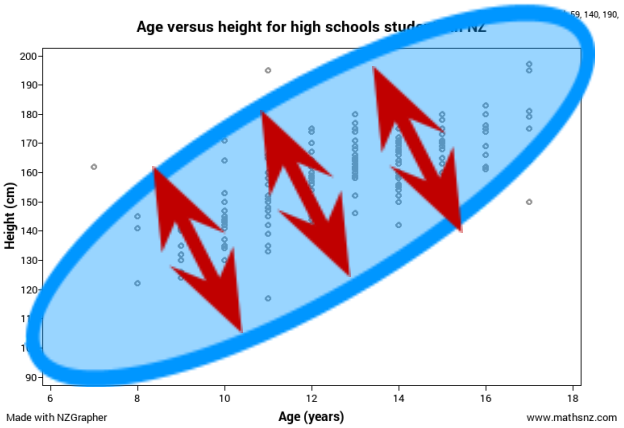
Strength:

Identify:

I notice that there is **weak** relationship between the age and height of high school ākonga.

Justify:

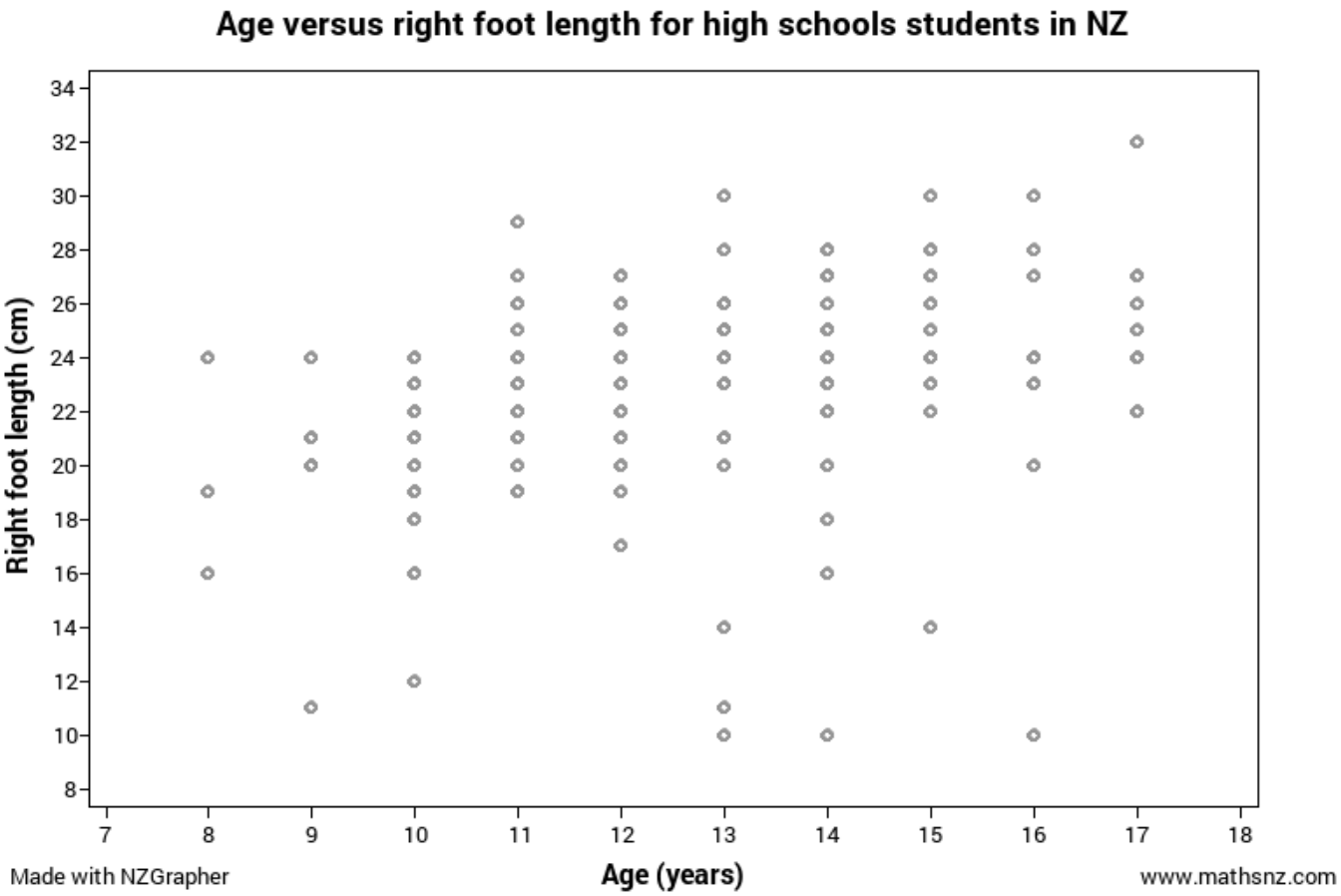
It is weak because the data points are widely scattered.



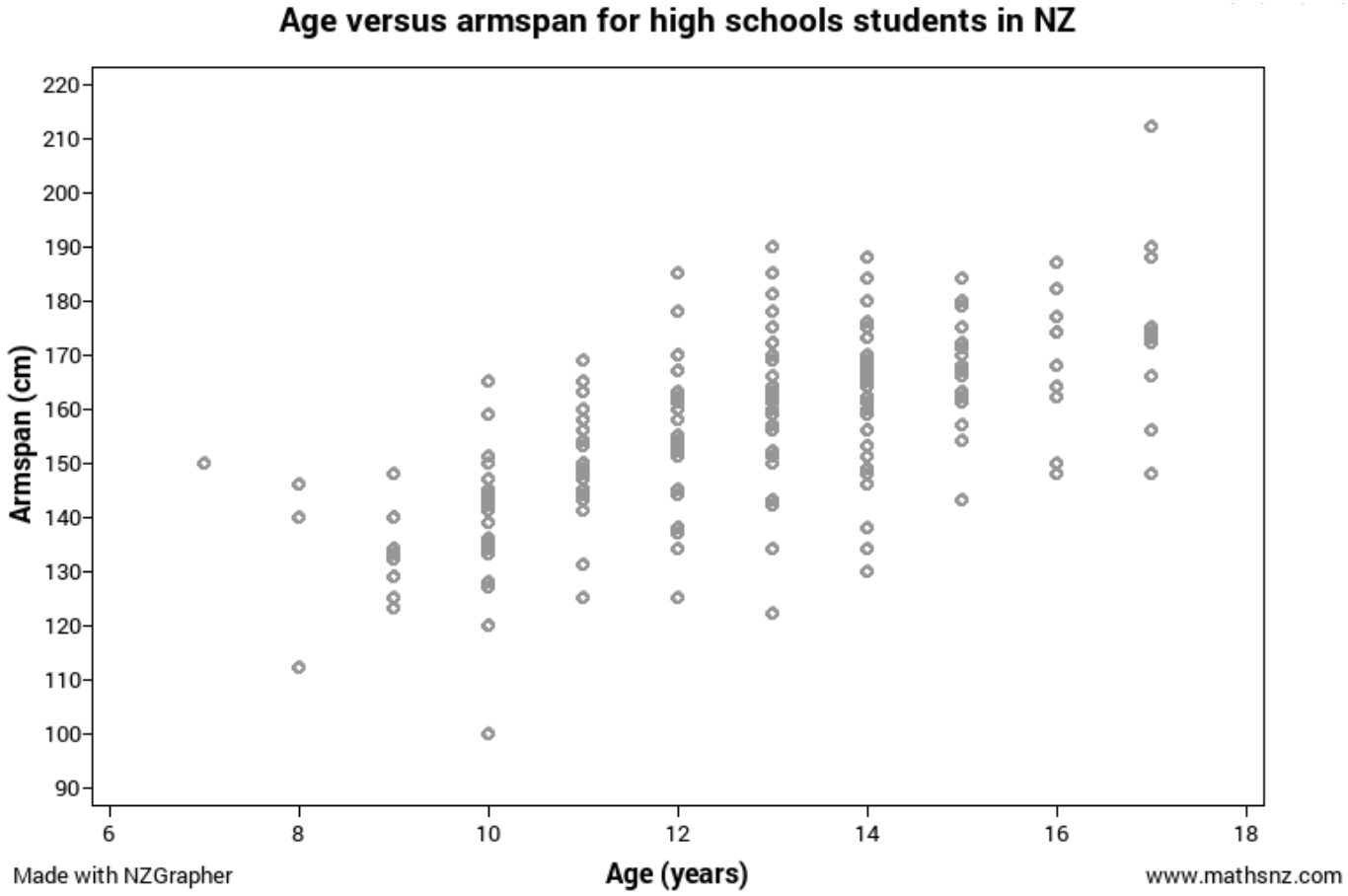
Exercise 9:

For the following graphs, shade the data, then describe the **features** (trend, direction, and strength) in **context**.

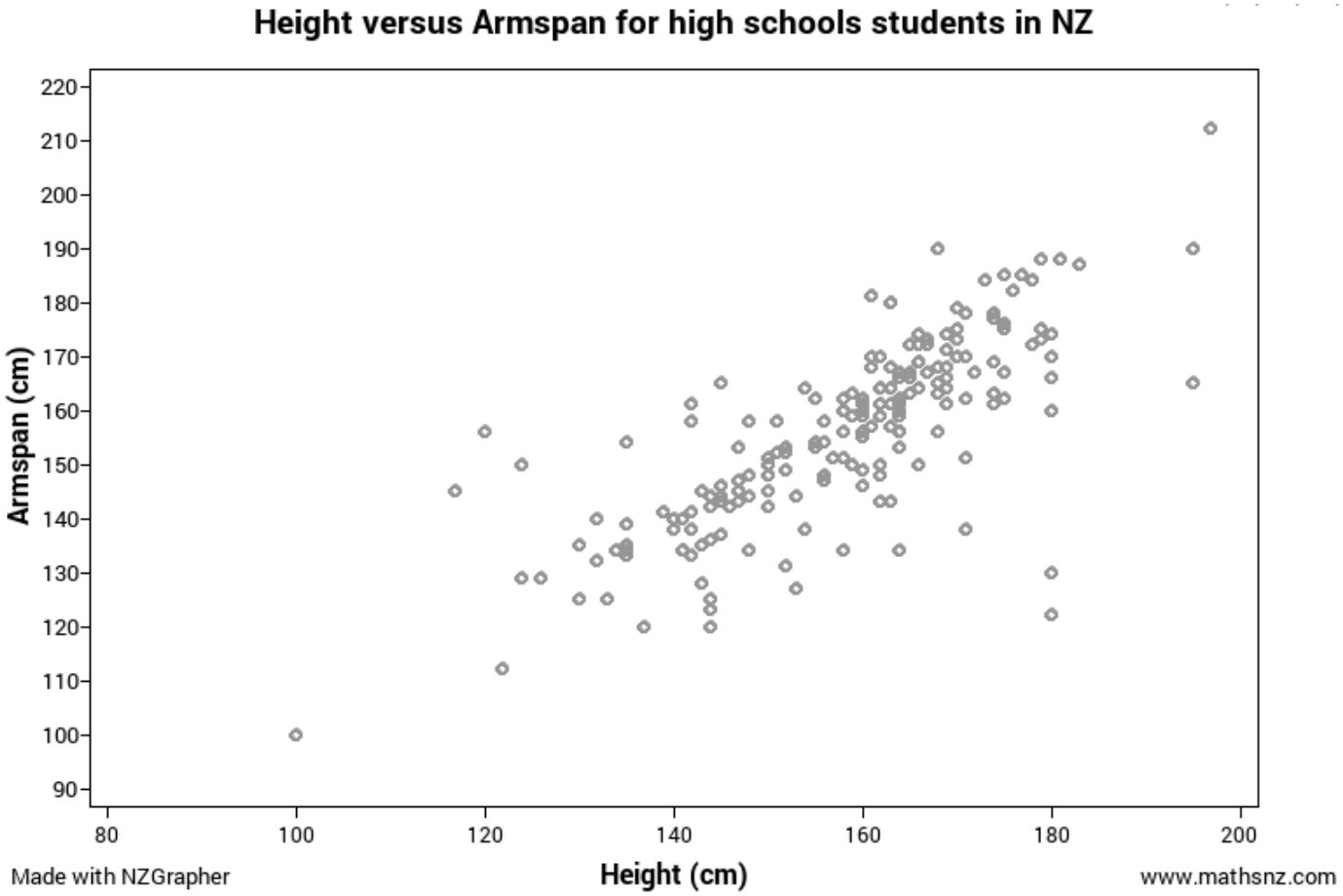
1)


Age versus right foot length for high schools students in NZ
Right foot length (cm)
34
32
30
28
26
24
22
20
18
16
14
12
10
8
7
8
9
10
11
12
13
14
15
16
17
18
Age (years)
Made with NZGrapher
www.mathsnz.com

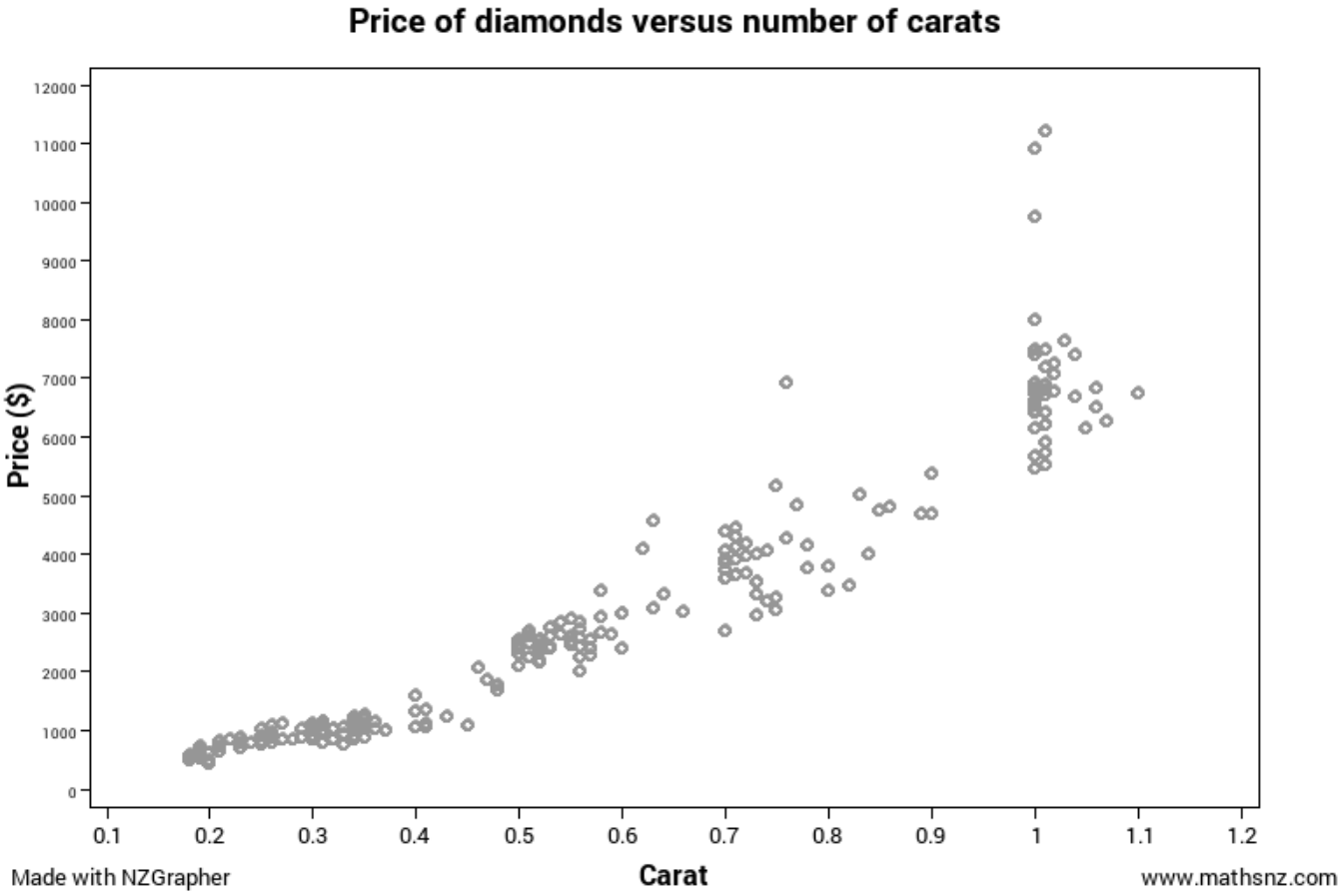
2)


Age versus armspan for high schools students in NZ
Armspan (cm)
220
210
200
190
180
170
160
150
140
130
120
110
100
90
6
8
10
12
14
16
18
Age (years)
Made with NZGrapher
www.mathsnz.com

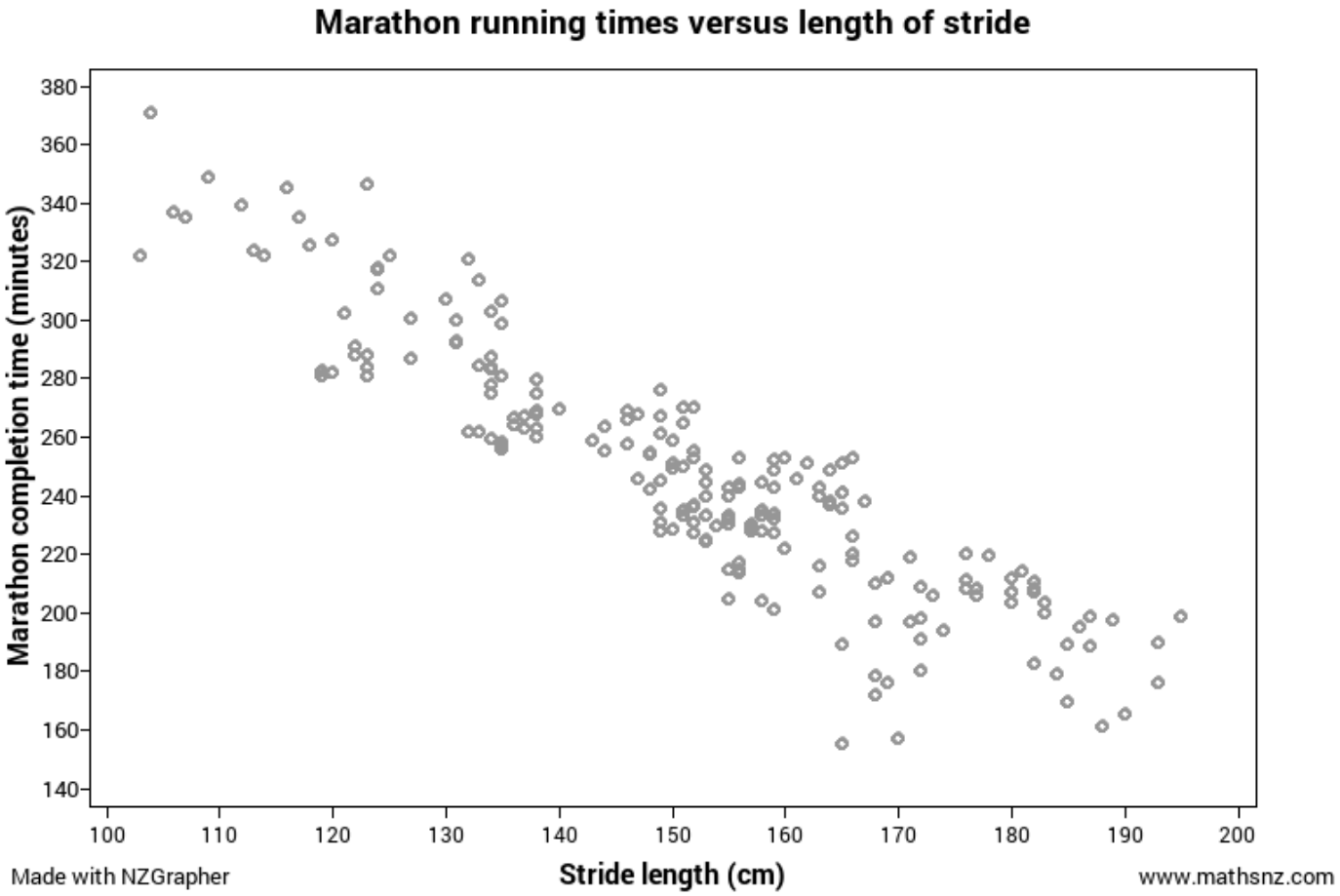
3)


Height versus Armspan for high schools students in NZ
Armspan (cm)
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80
100
120
140
160
180
200
Height (cm)
Made with NZGrapher
www.mathsnz.com

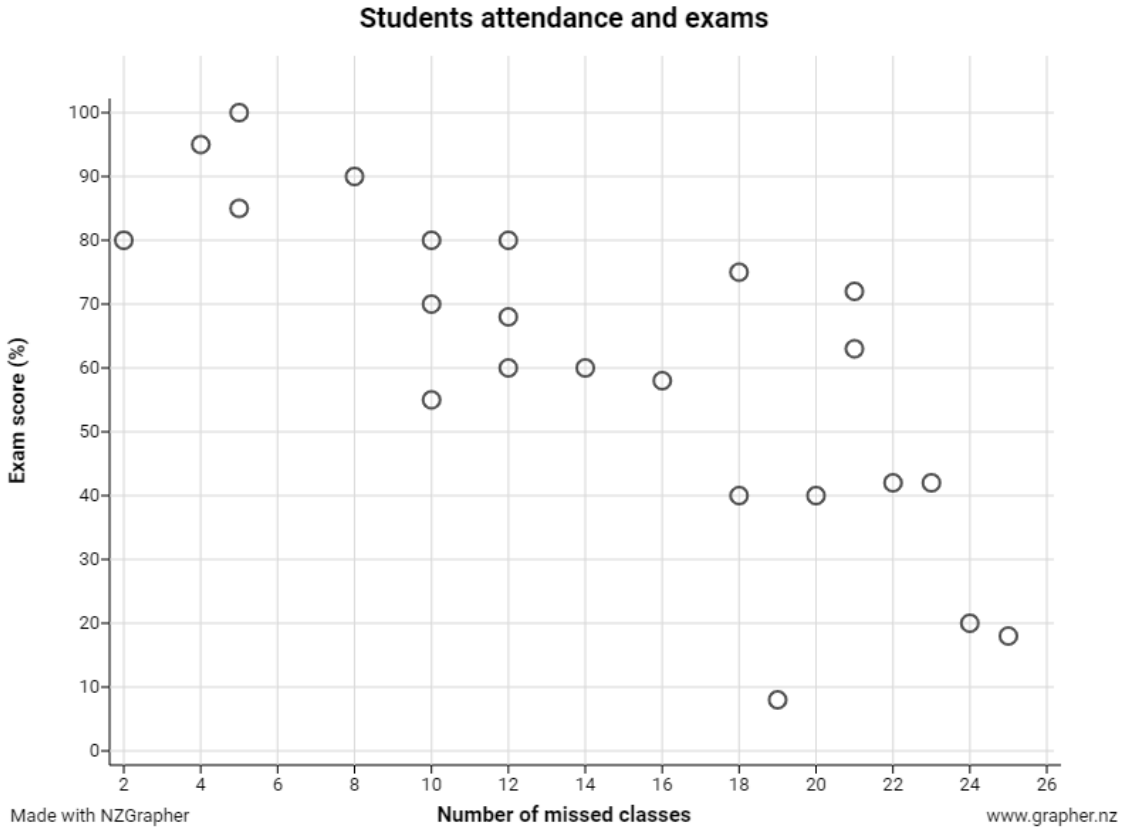
4)


Price of diamonds versus number of carats
Price ($)
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
Carat
Made with NZGrapher
www.mathsnz.com

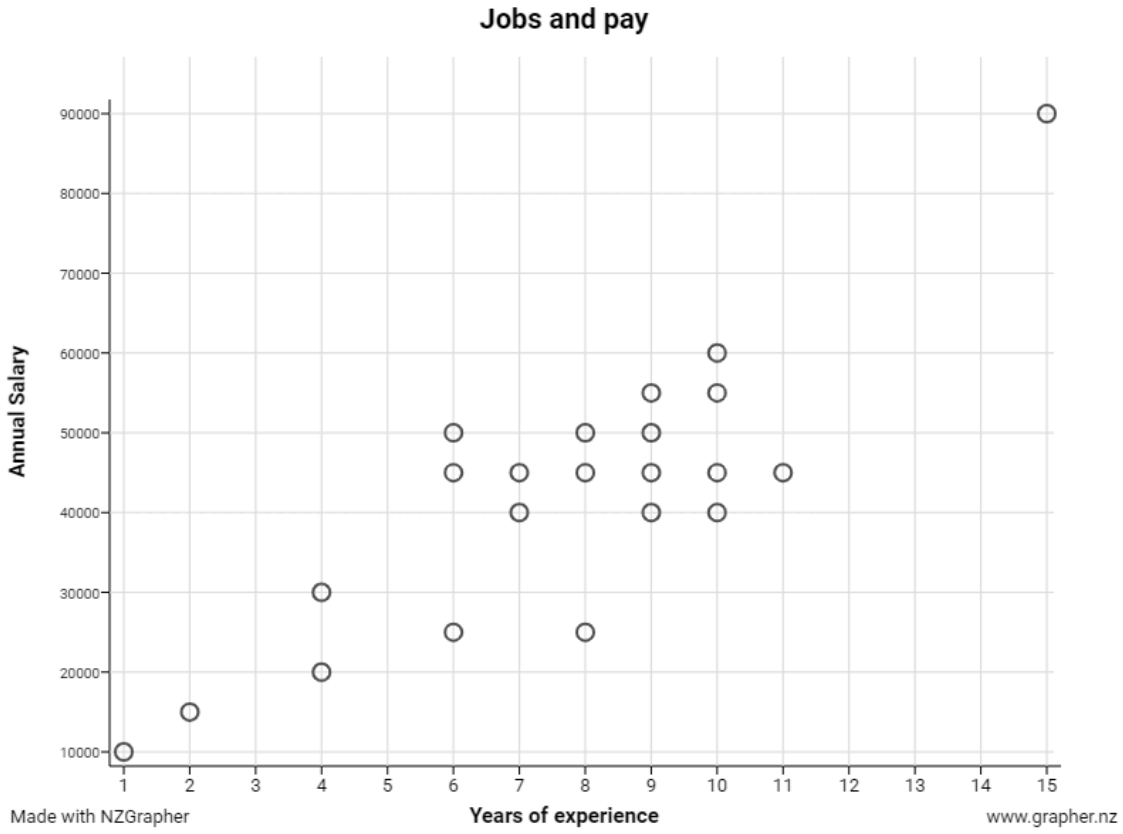
5)


Marathon running times versus length of stride
Marathon completion time (minutes)
380
360
340
320
300
280
260
240
220
200
180
160
140
100
110
120
130
140
150
160
170
180
190
200
Stride length (cm)
Made with NZGrapher
www.mathsnz.com

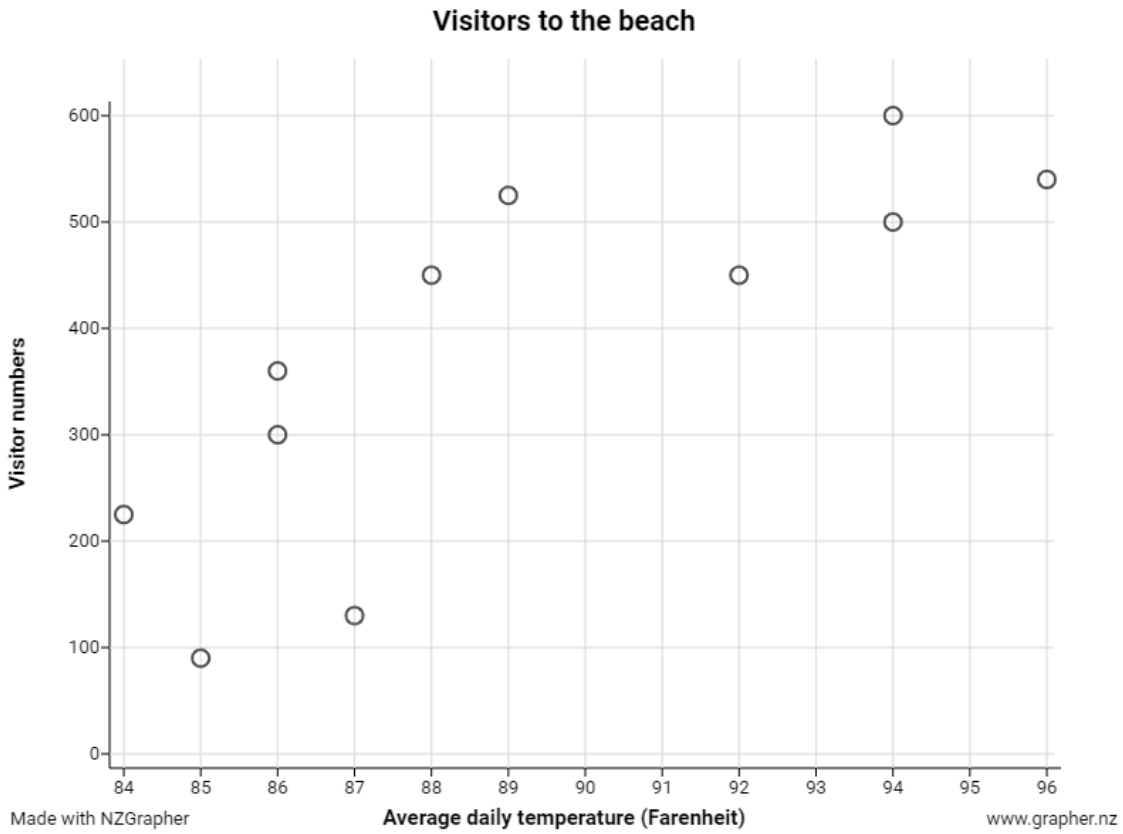
6)


Students attendance and exams
Exam score (%)
100
90
80
70
60
50
40
30
20
10
0
2
4
6
8
10
12
14
16
18
20
22
24
26
Number of missed classes
Made with NZGrapher
www.grapher.nz

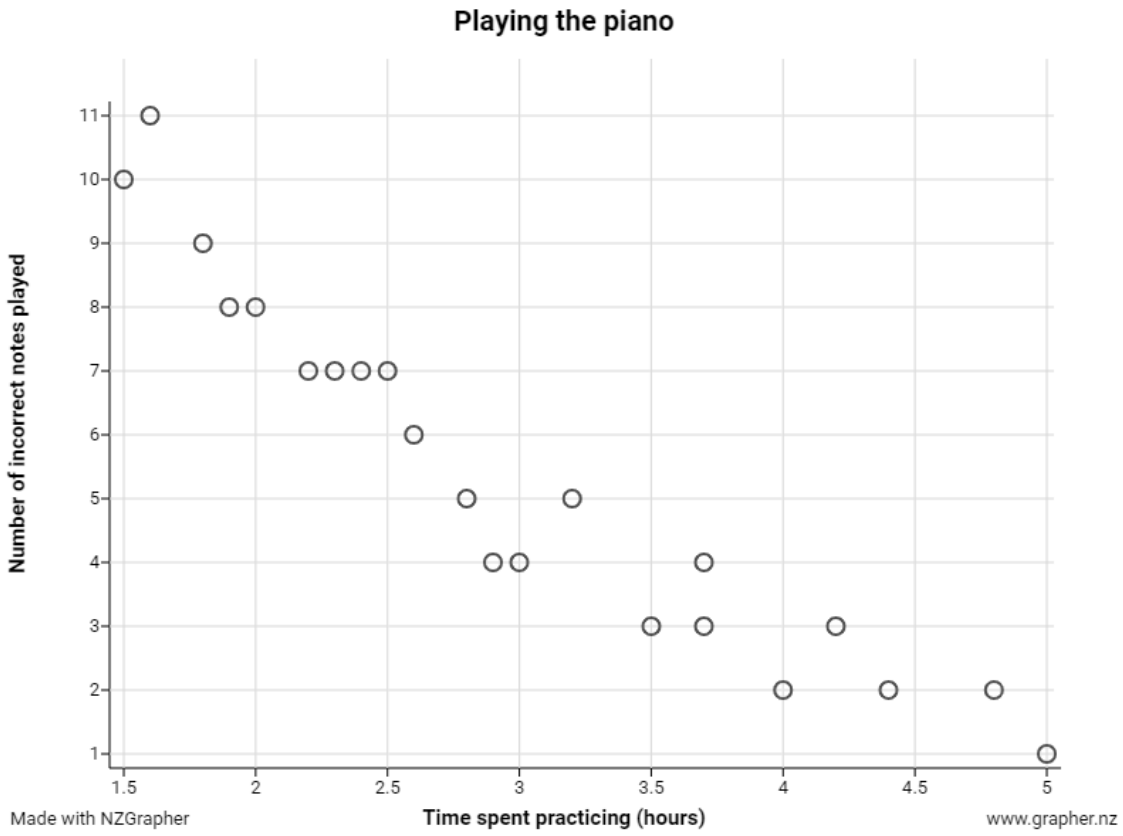
7)


Jobs and pay
Annual Salary
90000
80000
70000
60000
50000
40000
30000
20000
10000
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Years of experience
Made with NZGrapher
www.grapher.nz

8)


Visitors to the beach
Visitor numbers
600
500
400
300
200
100
0
84
85
86
87
88
89
90
91
92
93
94
95
96
Made with NZGrapher
Average daily temperature (Farenheit)
www.grapher.nz

9)

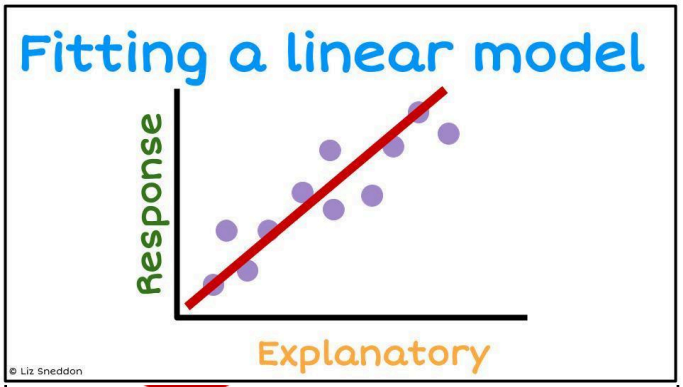

Playing the piano
Number of incorrect notes played
11
10
9
8
7
6
5
4
3
2
1
1.5
2
2.5
3
3.5
4
4.5
5
Time spent practicing (hours)
Made with NZGrapher
www.grapher.nz

Conclusion

Once we have identified the patterns in the data, then we want to fit a model to the data so that we can make predictions.

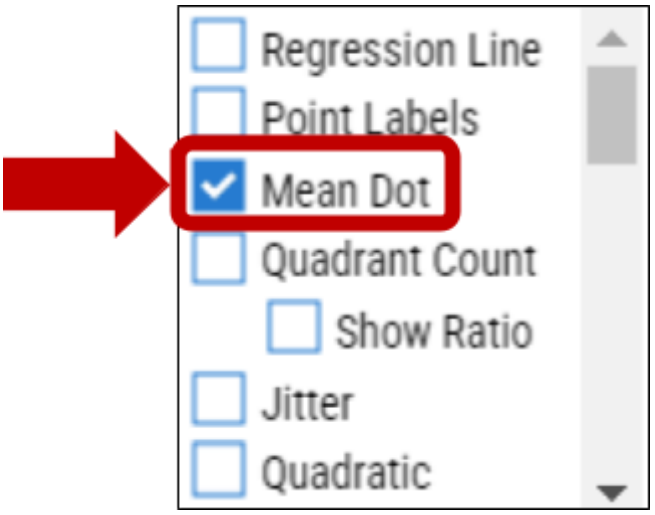
Fitting a model in NZGrapher

For NCEA Level 1, we want you to visually find a model to fit through the data. We will focus only on linear models. We will use NZGrapher to fit the model.

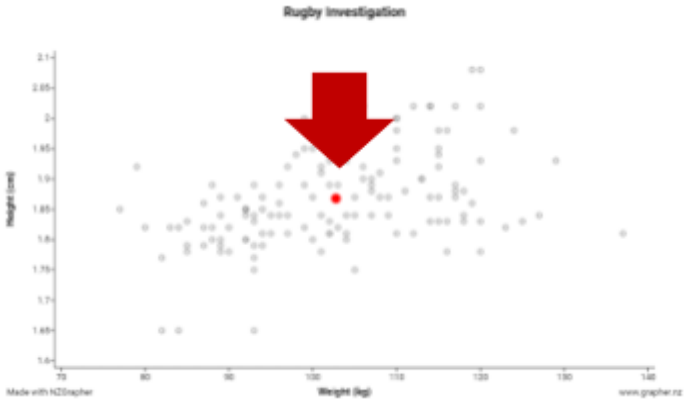


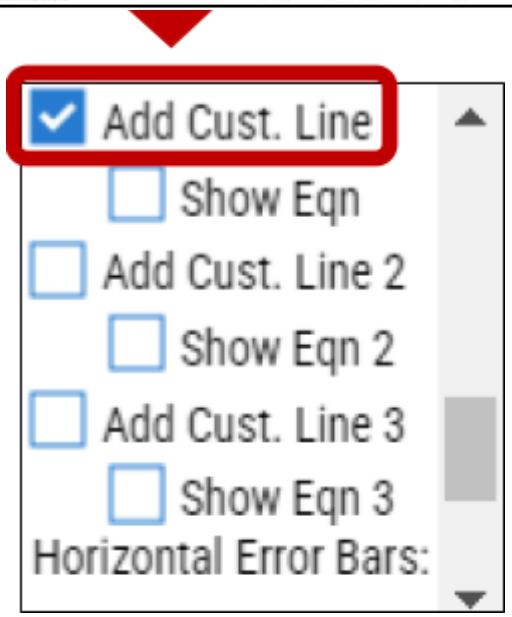
Step 1:

Make a scatter graph using the instructions earlier in this workbook. Tick and select the **Mean dot**.

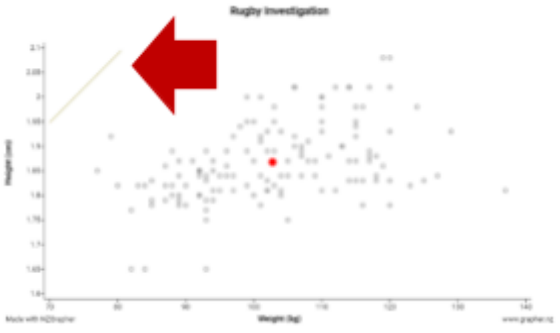


This puts a red dot which is the centre of both variables.





This will add two yellow dots with a straight line between them on the top right of your scatter graph.



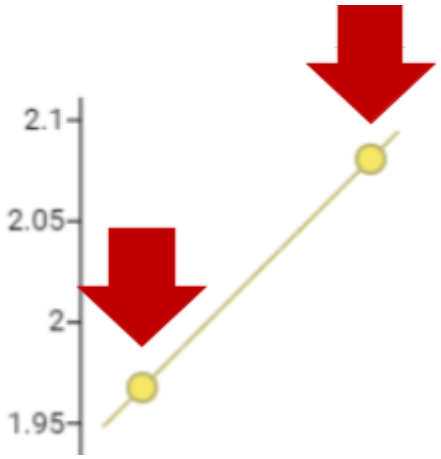
Step 3:

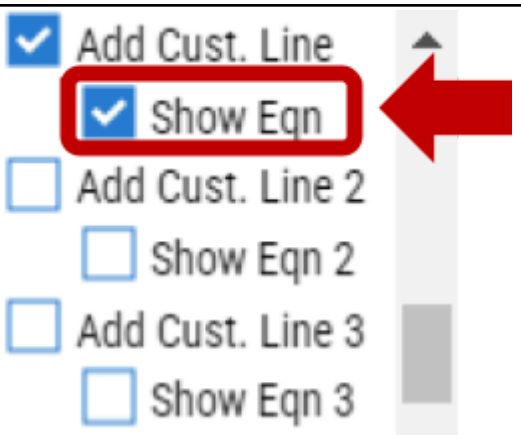
Click on each of the yellow dots and move the line around until it goes through the red mean dot and follows the pattern (it should go through the middle of the

Step 4:

Lastly, tick and select the **Show Eqn**. This will add the equation of your line to your graph (we will use this equation for predictions).

pattern).

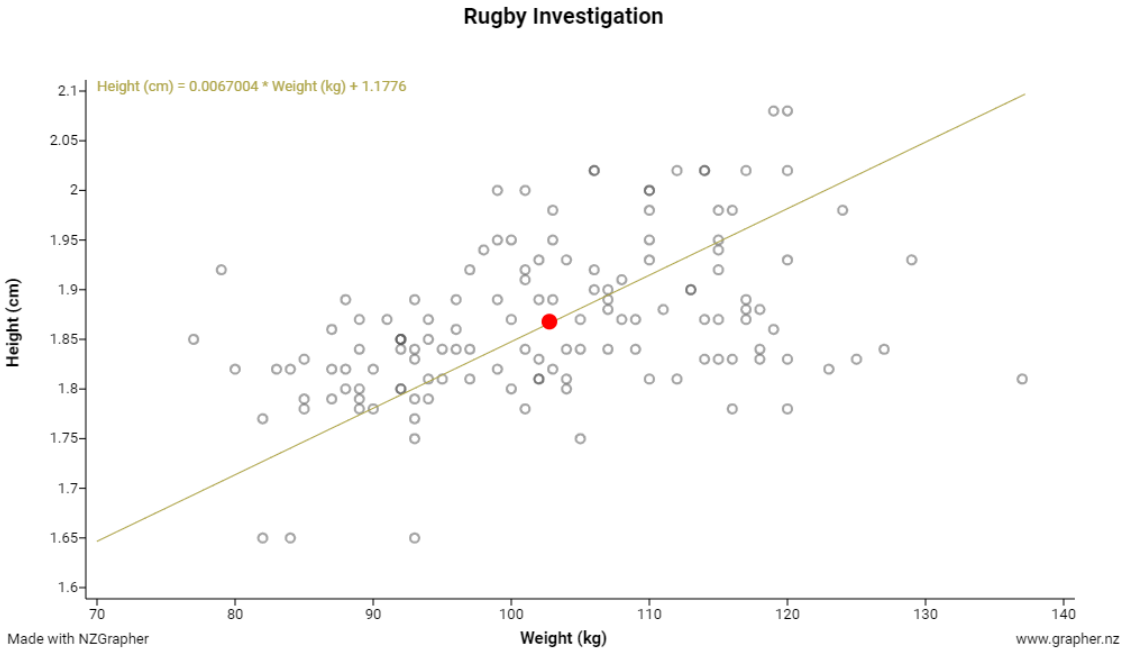




Then copy and paste your graph into a document.

Example:

For the rugby dataset in NZGrapher, I have fitted a model to a scatter graph of weight and height.



The equation of the linear model I fitted (as shown on the top left corner of the graph) is:

Height = 0.0067004 * Weight + 1.1776

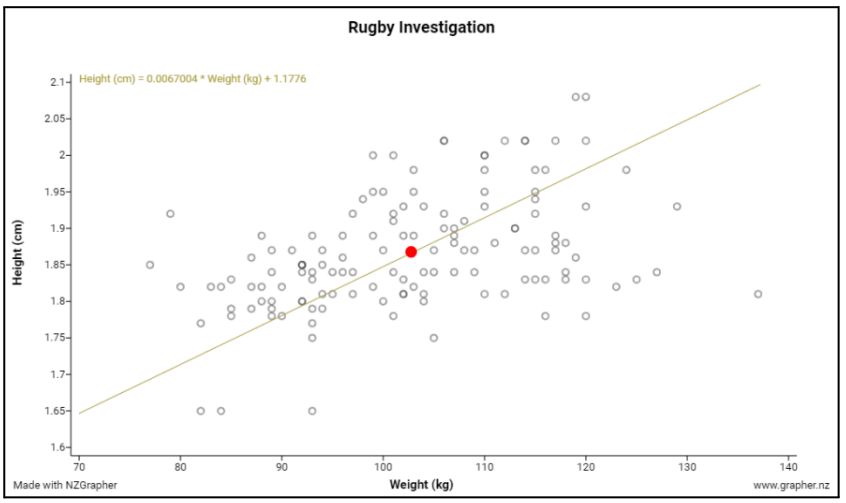
Making Predictions

Once you have come up with a suitable model for your data, then you need to use your equation to make a prediction. Choose a value for the explanatory variable (horizontal axis) that is within the same range of data on the graph, then substitute this into the equation, and write a sentence interpreting the prediction in context.

Example:

For the rugby dataset in NZGrapher, the model I fitted is:

Height = 0.0067004 * Weight + 1.1776



Looking at the explanatory variable (on the horizontal axis), Weight, I can see that the weights have values between approximately 75kg to 140kg.

I need to choose a value between these points, so I have randomly chosen 100kg.

Next, I substitute this weight into the equation, calculate it, and round the answer appropriately:

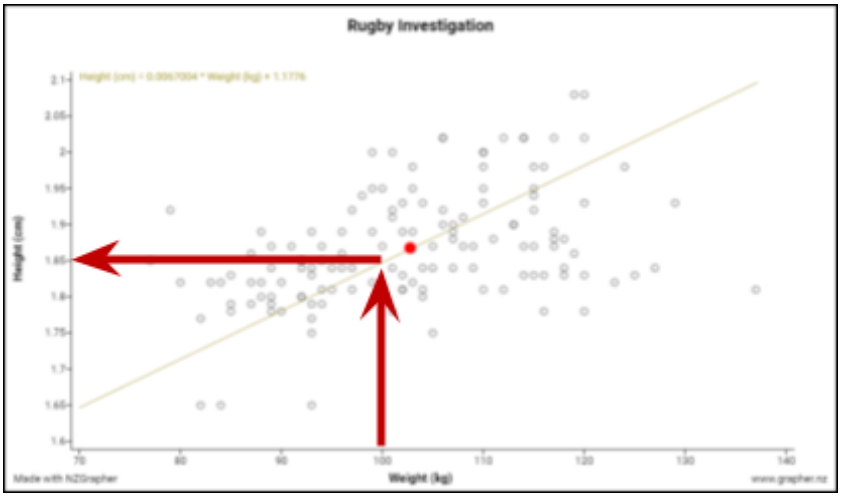
Height = 0.0067004 * Weight + 1.1776

Height = 1.84764 m

Height = 1.85 m

Lastly, I need to interpret this in context:

> I predict that a top rugby player in the sample who weighs 100kg, will be approximately 1.85 metres.



(In terms of what this means for the graph, start on the explanatory variable at the value of 100kg, then draw a vertical line that goes up to the fitted line, and then draw a horizontal line that goes across to the response variable, which we can see is approximately 1.85 metres.)

Exercise 10:

Go to NZGrapher, select the dataset specified below, draw a scatter graph, and add an appropriate model to your graph. Save the graph and write the equation you have found below. Lastly, make a prediction, showing your working, answer, and units, and write this in a sentence in context.

1)

Dataset	Cars.csv
Explanatory Variable	Weight (kg)
Response Variable	Price (US thousands of dollars)
Prediction	

2)

Dataset	Cars.csv
Explanatory Variable	Engine size (Litres)
Response Variable	Price (US thousands of dollars)
Prediction	

3)

Dataset	Marathon.csv
Explanatory Variable	Stridelength (cm)
Response Variable	Minutes
Prediction	

4)

Dataset	Cars.csv
Explanatory Variable	Fuel efficiency on the open road (miles per gallon)
Response Variable	Fuel efficiency on highways (miles per gallon)
Prediction	

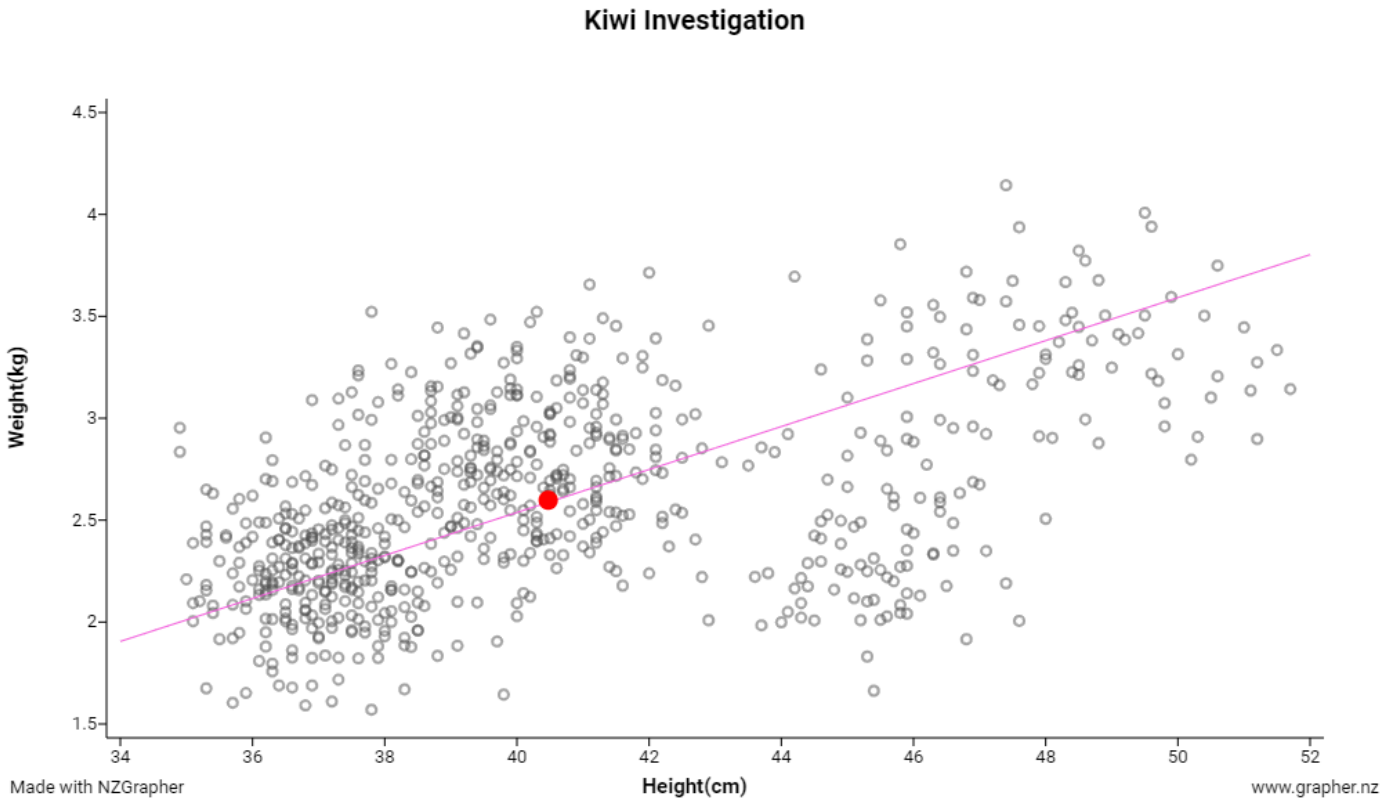
Linking findings to Purpose

Example:

Purpose:

I was interested in exploring data on kiwi birds as I am interested in the conservation of the species as they are being driven towards extinction due to three main threats: predators, lost habitat, and fragmented kiwi populations.

Findings:



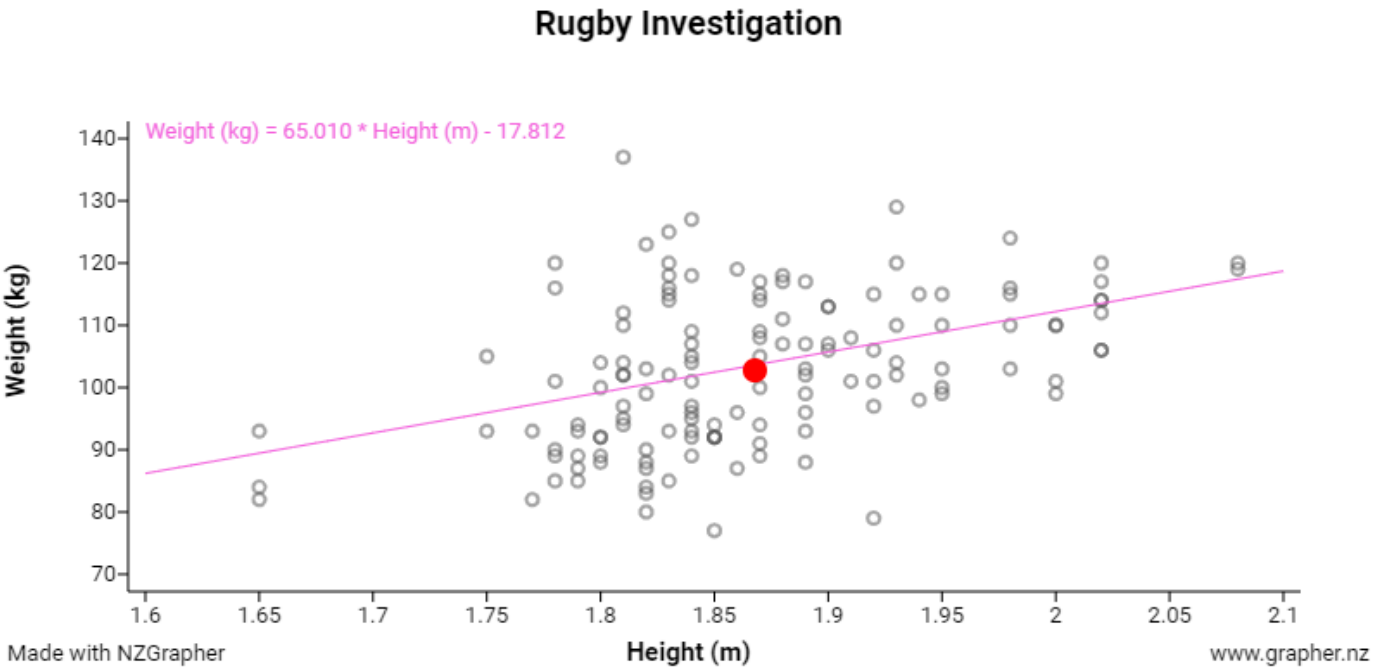
Conclusion:

My investigation into the weights and heights of kiwi birds found that there is a positive, linear, weak relationship. This is useful information for anyone involved in the monitoring of the kiwi populations in NZ, as we expect that if they find a short kiwi bird, that its weight would be lower, and if they find a taller kiwi bird, that its weight would be higher. So, when monitoring the health of kiwis, this information will be helpful in identifying kiwis that are at risk or not healthy.

Exercise 11:

For each of the questions below, connect the results to the given purpose.

1)

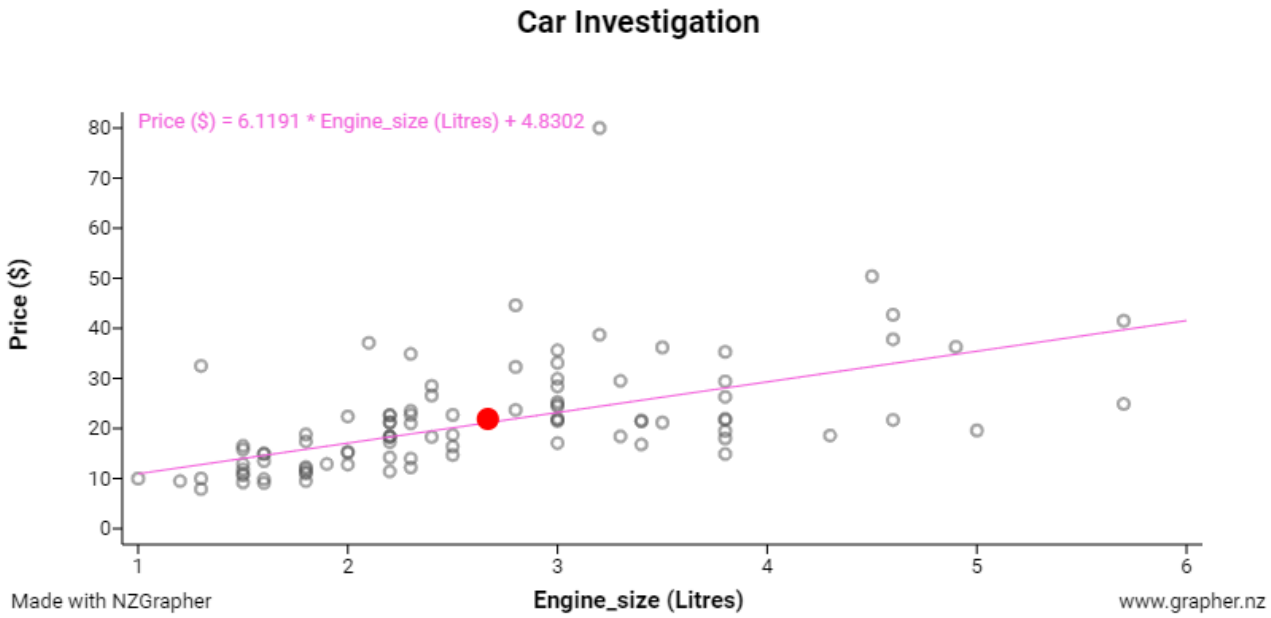


Purpose:

Fitness trainers are interested in the typical weights for top rugby players dependant on their height because that can guide them in developing appropriate training programmes.

Conclusion:

2)



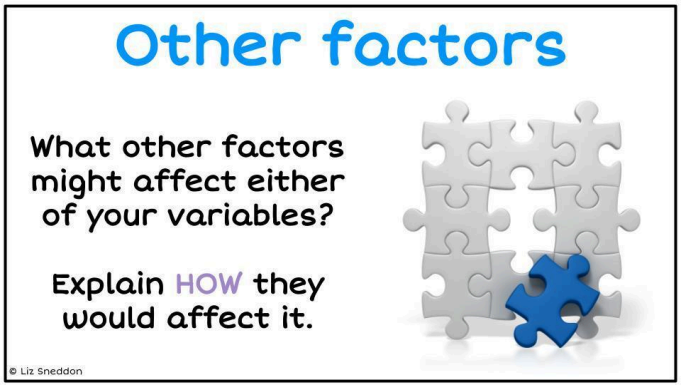
Purpose:

Ākonga in Year 13 are considering buying a car, which is why they are exploring data on the engine size and price of cars.

Conclusion:

Other factors

Think about making a puzzle. If you have one piece of the puzzle, you can learn a lot - the colours, the shapes, and guess where it might go in the big picture. But you know that this is only one piece of the puzzle, and that there are lots of other pieces that when put together make a whole different picture.



Data is like a puzzle. You might investigate one particular piece, one particular comparison, but we always need to be aware that there are many other pieces in this puzzle. Before we can paint the whole picture, we need to think about what the other pieces might be.

This is the idea of other factors. You need to consider what other factors might also affect the numerical variables that you are investigating and explain **WHY** this factor might be affecting either of the variables (response or explanatory).

Example:

Problem:
I wonder if there is a relationship between the length of a person's right foot and their hand span, for ākonga in Mrs Sneddon's Year 9 maths class.

Other variables or factors:
There are a number of different factors that might affect the length of a person's feet. For example, their genetics may influence this. If their parents have small feet, then that would make the child more likely to have small feet too. Equally with hand span - if either of the parents had large or small hand spans, their children are more likely to have the same as the parents.

Another factor that could affect both the hand span and foot length is how tall a person is. If a person is taller, it is likely that both their hand span and foot length are longer than a shorter child (e.g. a 5-year-old).

Exercise 12:

Explain some other factors that may affect either or both variables.

1) **Problem:**
I wonder if there is a relationship between the length of the rotor blades and the time it takes for the helicopter to drop to the ground.

Other factors:
What other factors do you think would affect the time it takes for the helicopter to drop to the ground?

2) **Problem:**

I wonder if there is a relationship between a person's age and the amount of time that they spend on their cell phone each day, for teenagers at Saint Kentigern College.

Other factors:

What other factors do you think would affect the time that teenagers in NZ spend on cell phones each day or the age of a teenager?

3) **Problem:**

I wonder if there is a relationship between the number of hours someone works and the amount of money, they get paid, for adults in NZ.

Other factors:

What other factors do you think would affect the amount of money that adults in NZ earn per year, or the number of hours someone works?

4) **Problem:**

I wonder if there is a relationship between the amount of time teenagers' study for each week and the number of NCEA credits they get, for all teenagers in NZ.

Other factors:

What other factors can you identify?