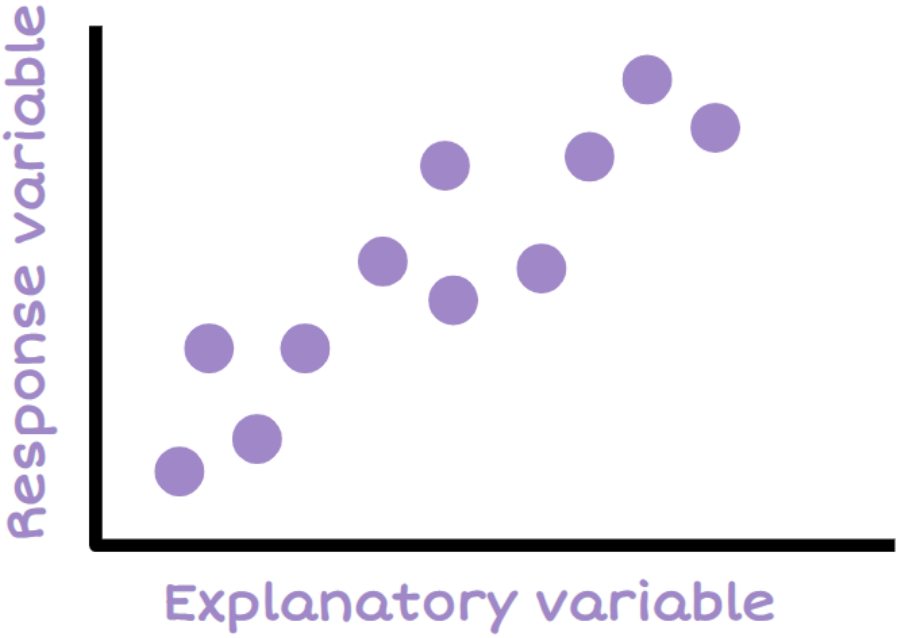
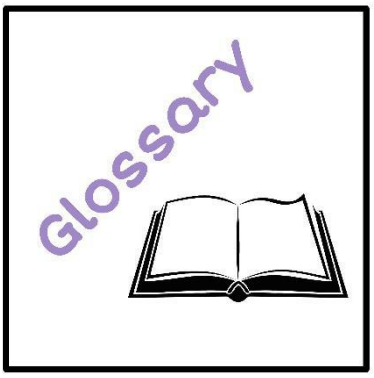
Level 1 Bivariate Workbook



Name:

By Liz Sneddon



Scan the QR code and it will take you to the Glossary on Mrs Sneddon's website.
Write down the definitions in the table.

Terminology	Definition
Data Types	(copy diagram)
Numeric Data	
Discrete Data	
Continuous Data	

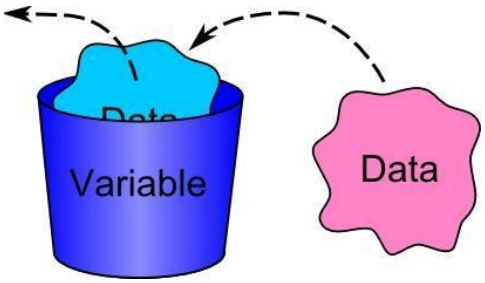
Categorical Data	
Variable	
Bivariate	
Explanatory Variable	
Response Variable	

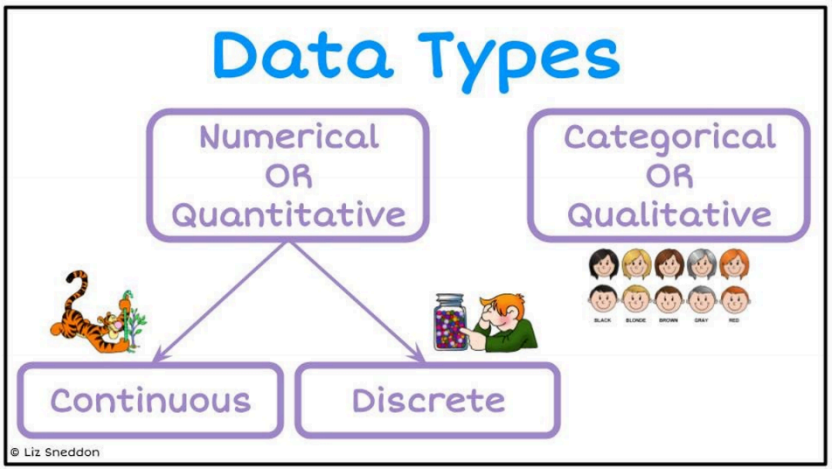
Participants	

Background Information

For this standard we want to explore possible relationships between **two numeric variables**. Let's have a little explore of some of the background information.

Variables, Data & Data Types





Example:

Here is a spreadsheet:

	A	B	C	D	E	F
1	Age	Gender	PE?	Wall sit time (seconds)	Leg up Wall sit time (seconds)	Height (cm)
2	15	F	no	185	15	168
3	15	M	no	70	31	187
4	15	M	no	71	40	170
5	15	F	no	91.76	12.18	156.5
6	15	F	no	74	21	147

Each **row** is a set of **data** belonging to a student.

Each column is a variable.

Variable	Data Type

Age	Numeric (Discrete)
Gender	Categorical
Whether they take PE this year	Categorical
Wall sit time	Numeric (Continuous)
Leg up wall sit time	Numeric (Continuous)
Height	Numeric (Continuous)

Exercise:

For each of the datasets below, name the numeric and categorical variables.

1) The Bungee spreadsheet:

	A	B	C	D	E	F
1	**Elastic type**	**Length (cm)**	**Number of marbles**	**Stretched length (cm)**	**Weight (grams)**	**Did it hit obstacle(s)?**
8	Narrow	50	20	95	112	No
9	Narrow	50	18	79	112	No
10	Wide	27	20	30	107	No
11	Wide	39	20	60	119	No
12	Wide	45	19	70	106.4	Yes

Numeric Variables:

Categorical Variables:

2) The Marathon dataset:

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)
Stridelength	The persons average stride length over the marathon in cm.

Numeric Variables:
Categorical Variables:

Scatter graphs

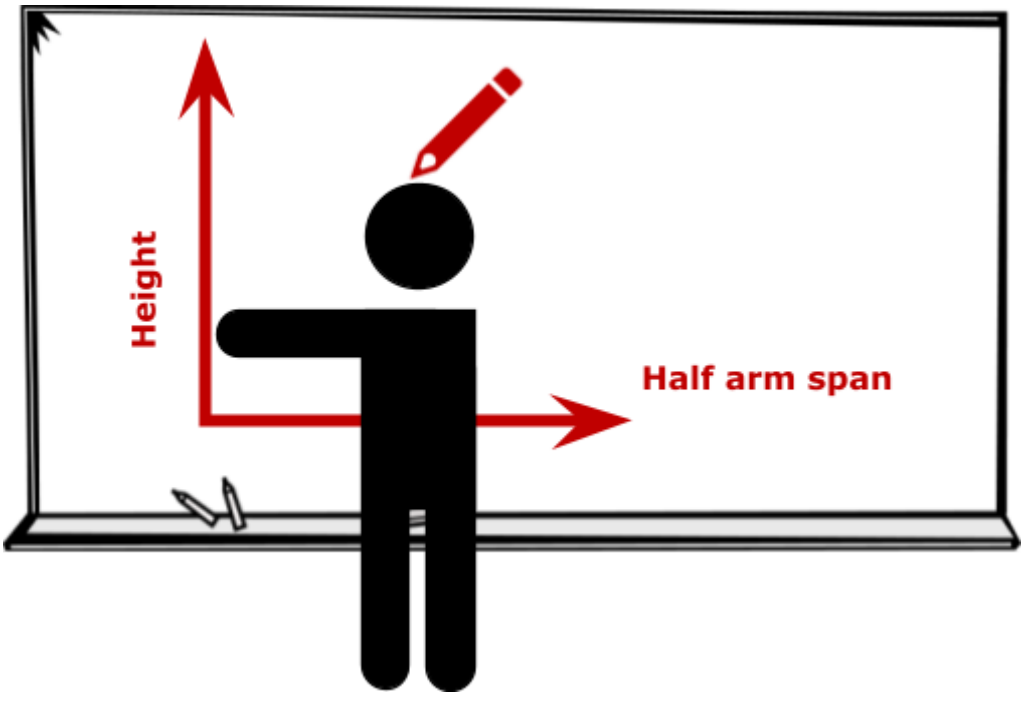
Before we get into the PPDAC cycle for collecting data and analysing the features of graphs, we are first going to have a look at scatter graphs.

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 students and the length of (half) arm span.

Instructions:



1) Teachers have drawn the axes on the whiteboard (height versus half arm span).

2) Students will come up to the whiteboard one at a time standing facing the whiteboard.

3) Each student will hold out their left arm straight out, with the tip of their finger touching the y axis.

4) The teacher, or another student, will draw a cross where the top of the student's head is.

5) Repeat this with all students in the class.

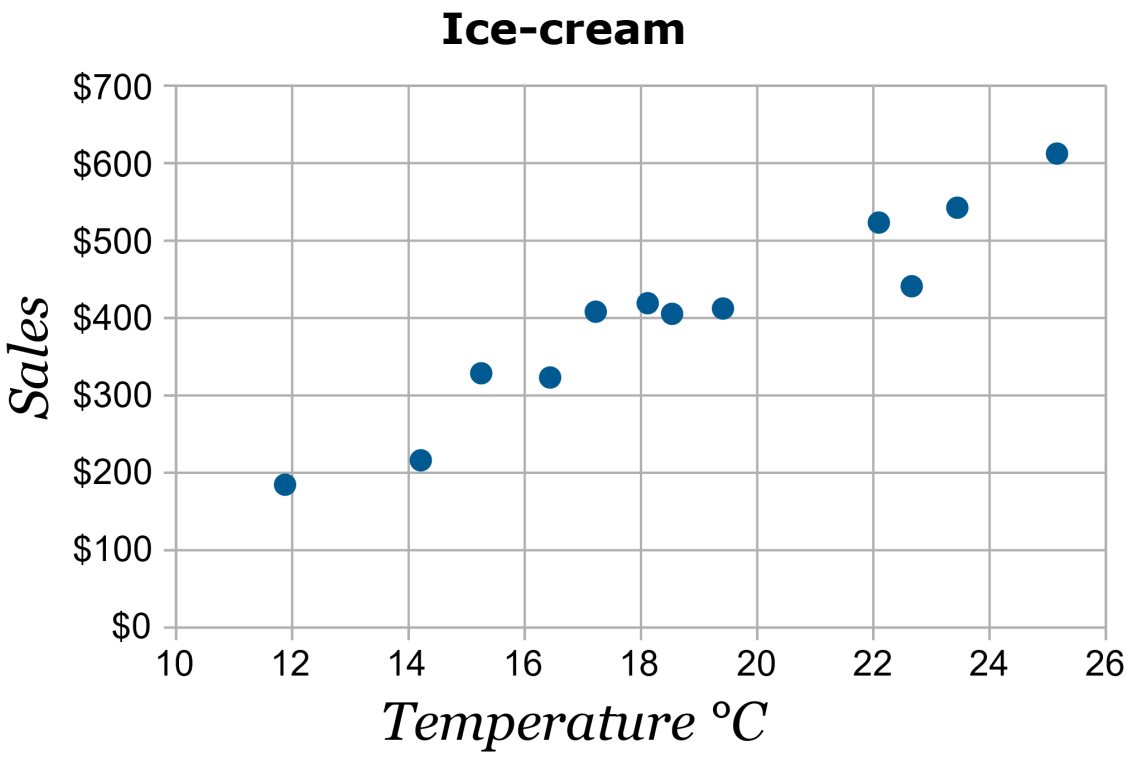
Discussion:

What do you notice on the graph of the class?

What improvements can you suggest?

Think about how we can improve the data and the graph.

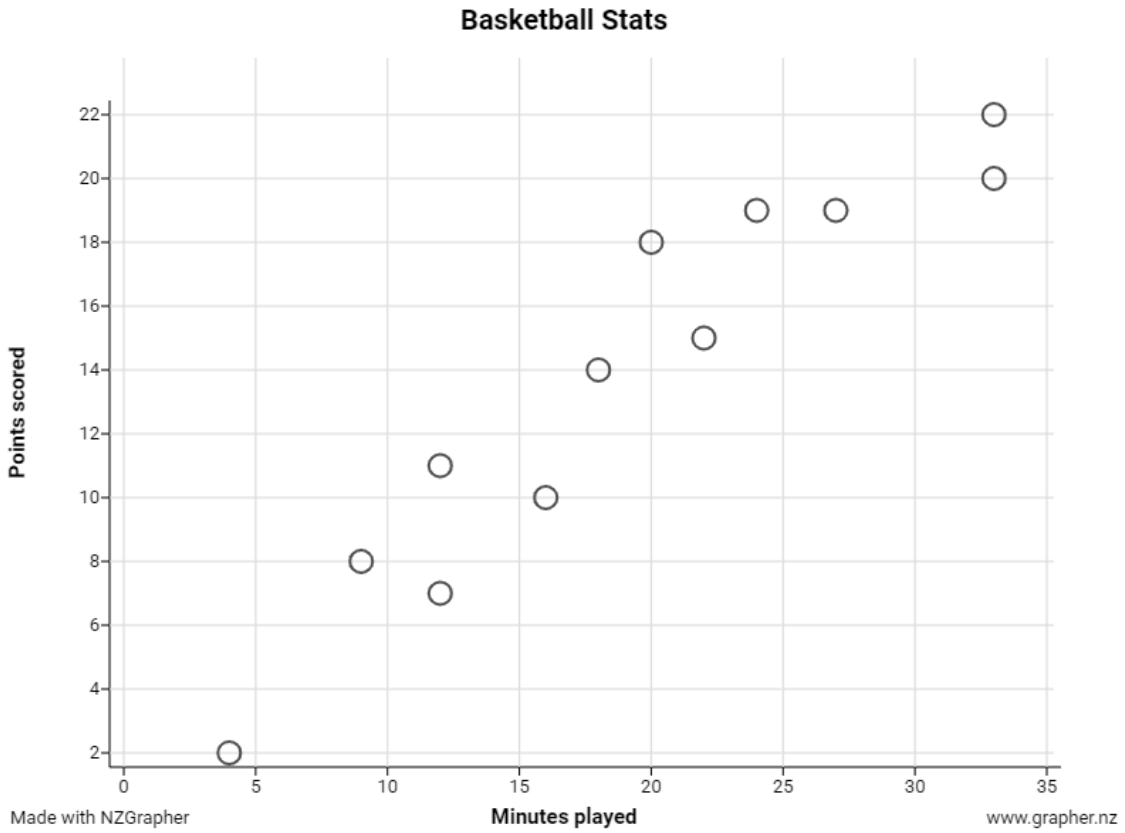
Example:



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.

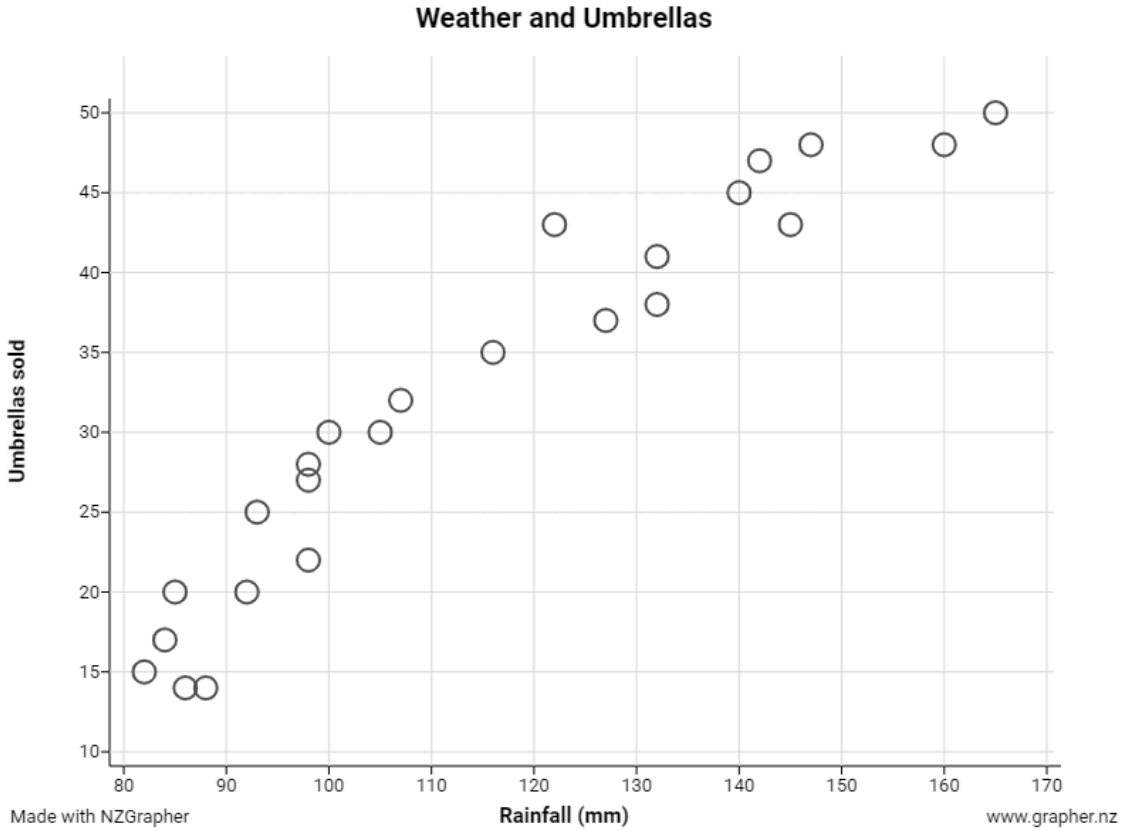
Exercise:

1)



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?

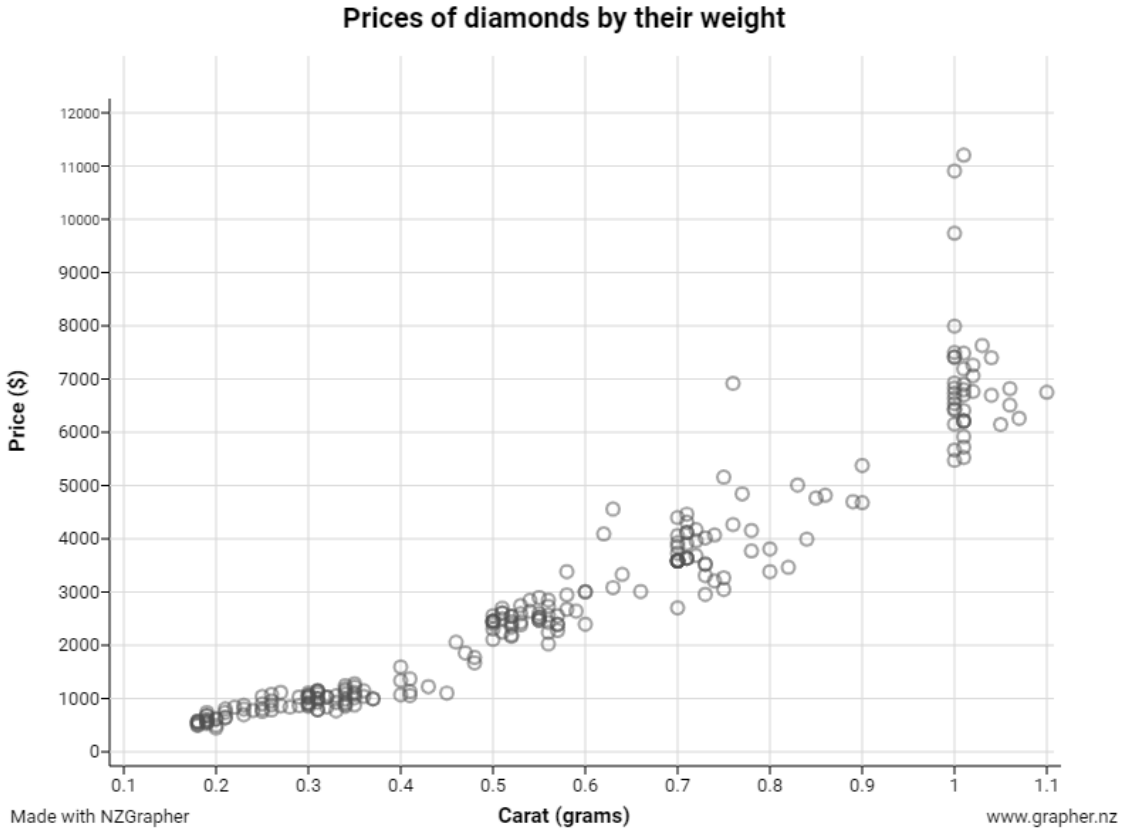
d) Looking at days with a **small** amount of rainfall, the number of umbrellas sold are **low / high.** (circle correct answer)

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

3)

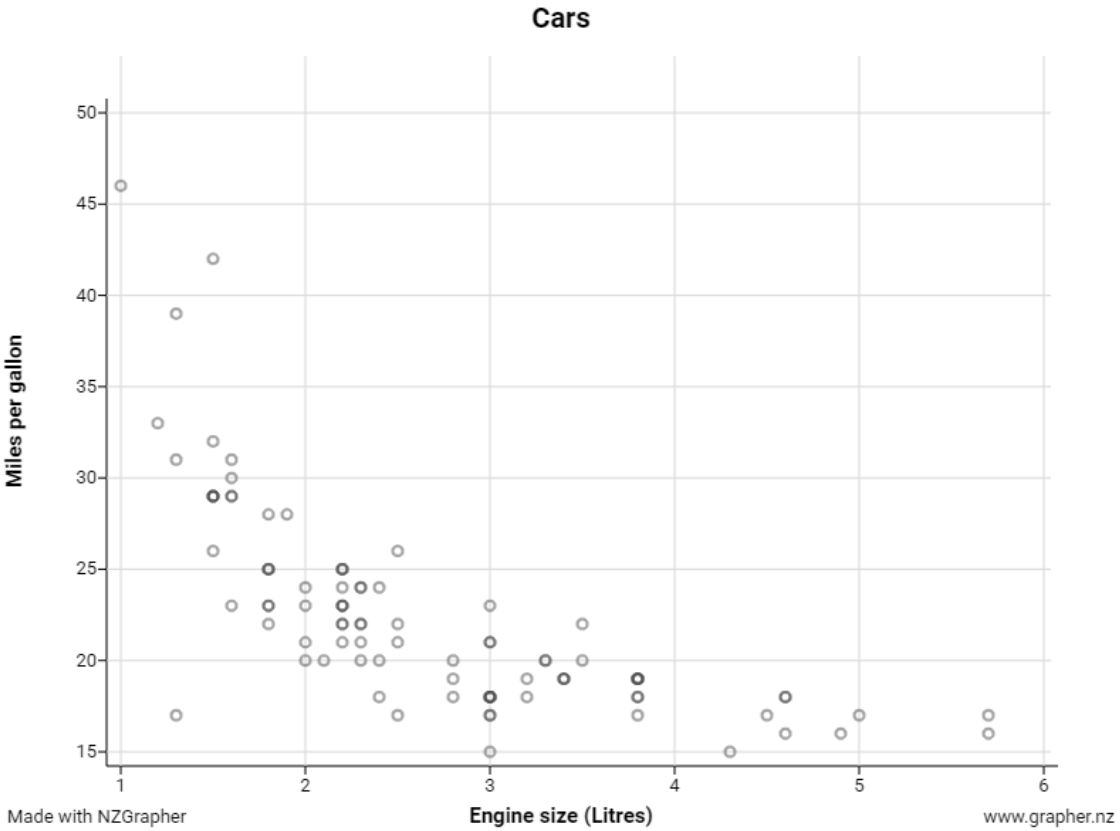
a)	What are the variables in this graph?	
b)	What is the price of the car for the person who has a salary of $100,000?	
c)	What is the salary for the person who has a car worth $30,000?	
d)	Looking at the people who have a **low** salary, their car prices are **low / high.** (circle correct answer)	
e)	Looking at the people who have a **high** salary, their car prices are **low / high.** (circle correct answer)	

4)



a)	What are the variables in this graph?	
b)	How many diamonds cost more than $8000?	
c)	How many diamonds weigh between 0.8 grams and 0.9 grams?	
d)	Looking at diamonds that are **small (low carat)**, the prices are **low / high.** (circle correct answer)	
e)	Looking at diamonds that are **large (big carat)**, the prices are **low / high.** (circle correct answer)	

5)



a) What are the variables in this graph?

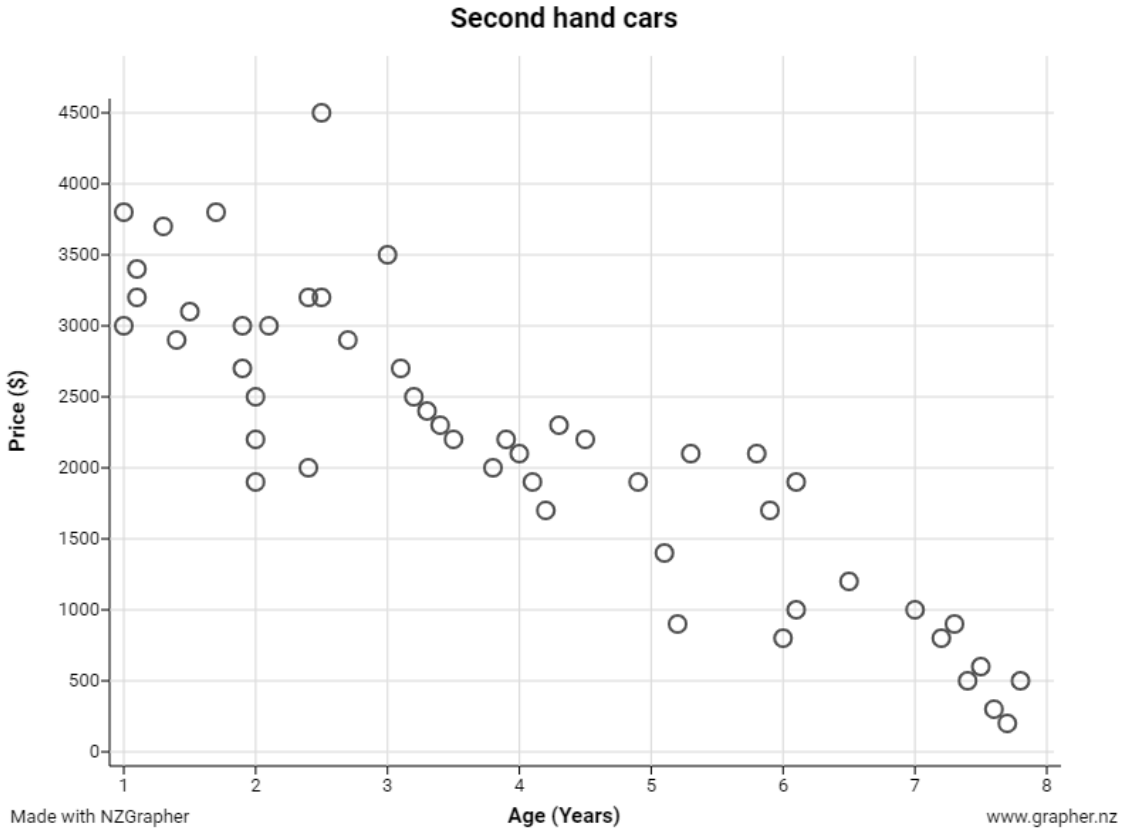
b) How many cars do more than 35 miles per gallon?

c) What is the largest and smallest engine size?

d) Looking at cars with **small** engine sizes, the fuel efficiency (miles per gallon) is **low / high.** (circle correct answer)

e) Looking at cars with **big** engines, the fuel efficiency (miles per gallon) is **low / high.** (circle correct answer)

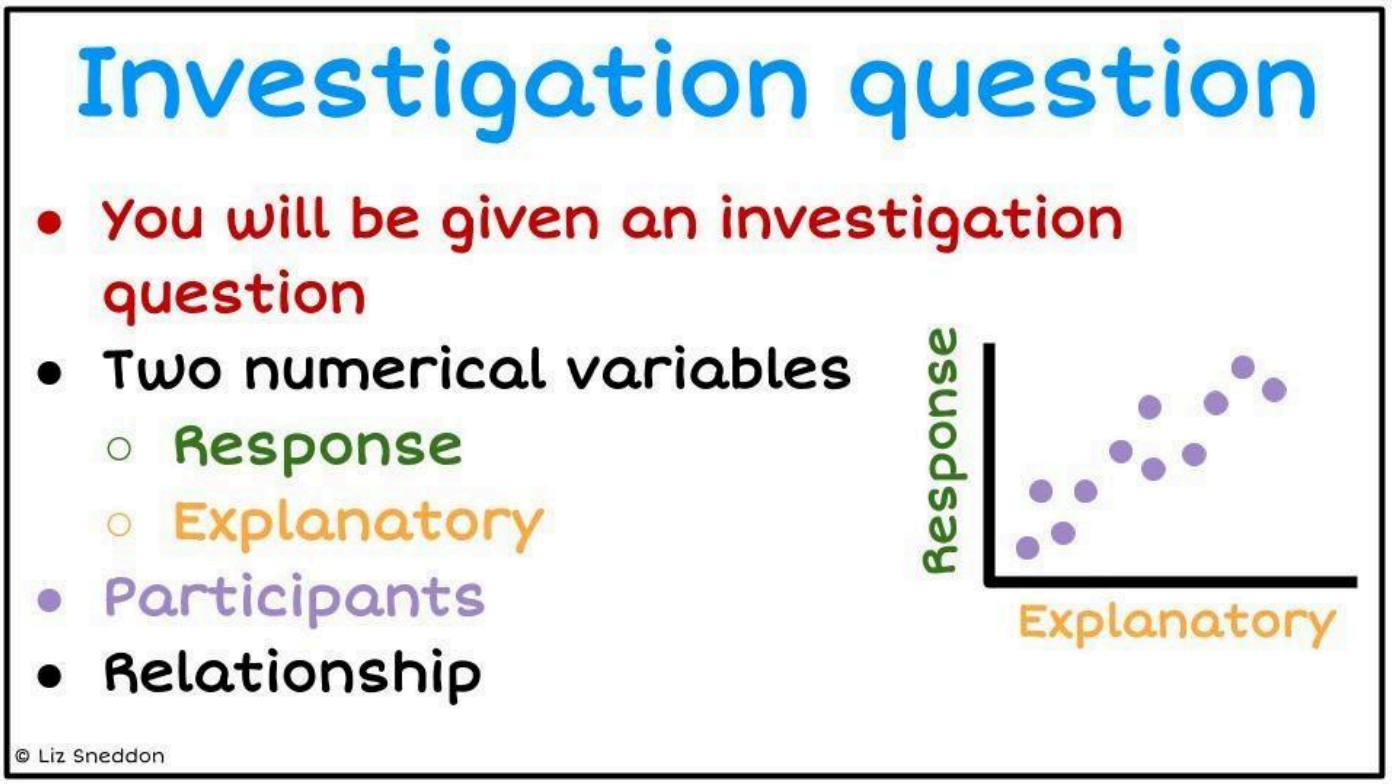
6)



a)	What are the variables in this graph?	
b)	How many cars for sale are between 3 and 5 years old?	
c)	How many cars are over $2000?	
d)	Looking at secondhand cars which are **relatively new (age is small)**, the price of cars is **low / high.** (circle correct answer)	
e)	Looking at secondhand cars which are **older**, the price of cars is **low / high.** (circle correct answer)	

Problem

You will be given an investigation question, which is to investigate **if there is a relationship** between two numeric variables.



Sometimes we use the word **correlation** when describing a relationship.

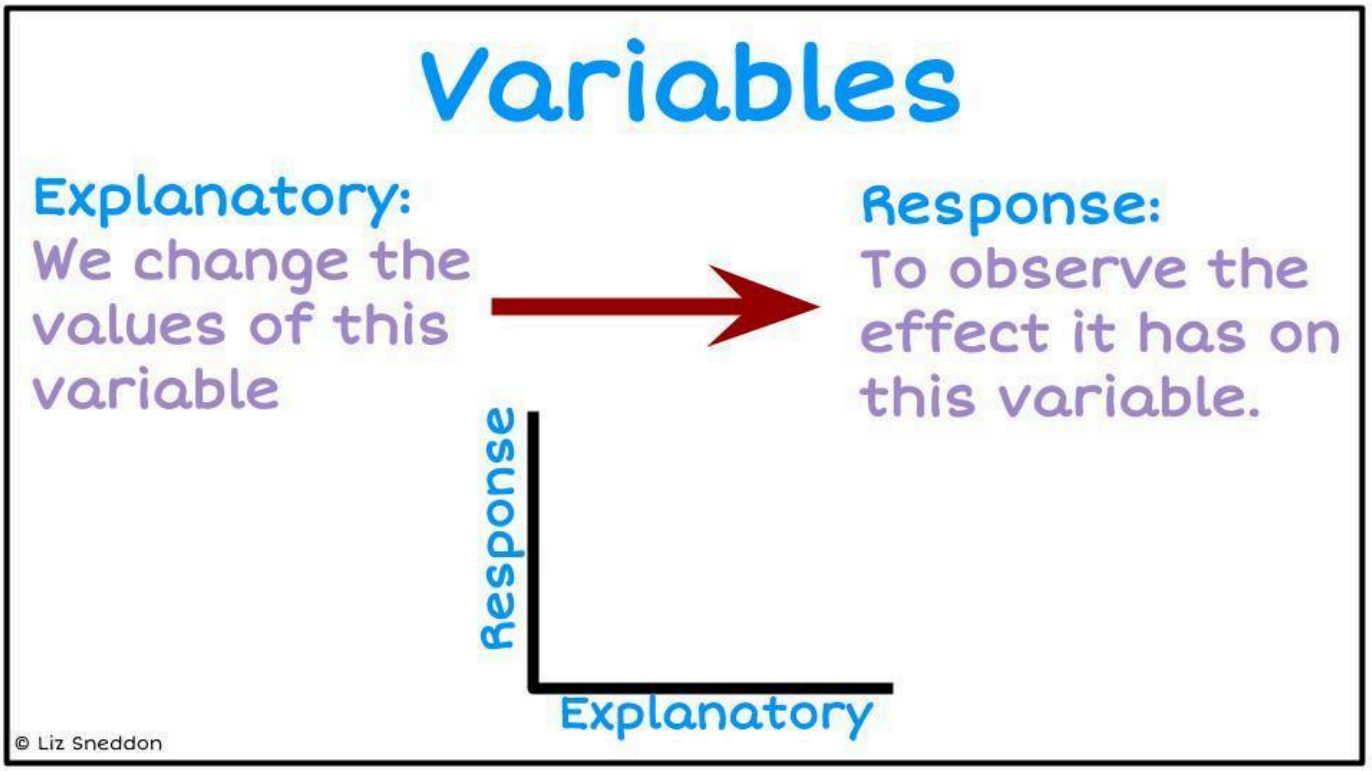
Example:

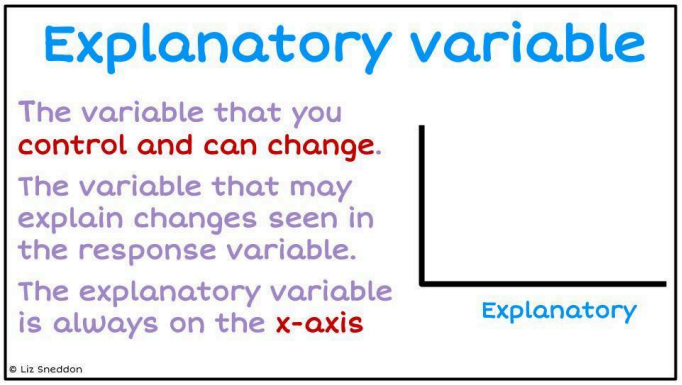
Problem

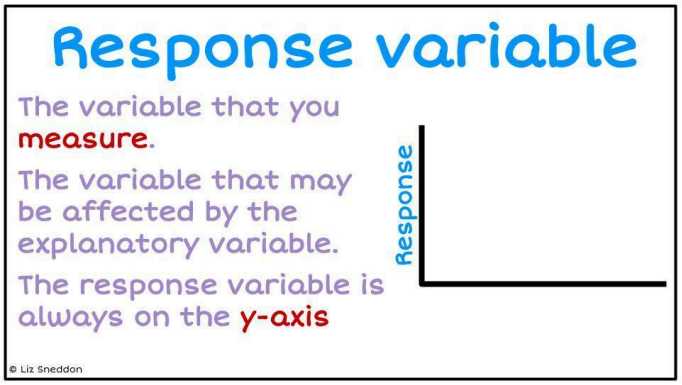
I wonder if there is a relationship between a person's foot length and hand span for students at your school?

Explanatory versus 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?

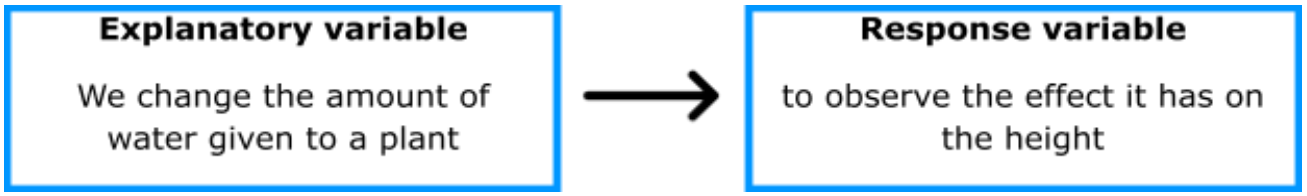
The first option is correct – we would **control or change** how much water a plant is given, and then we can **measure** the height.

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:

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.

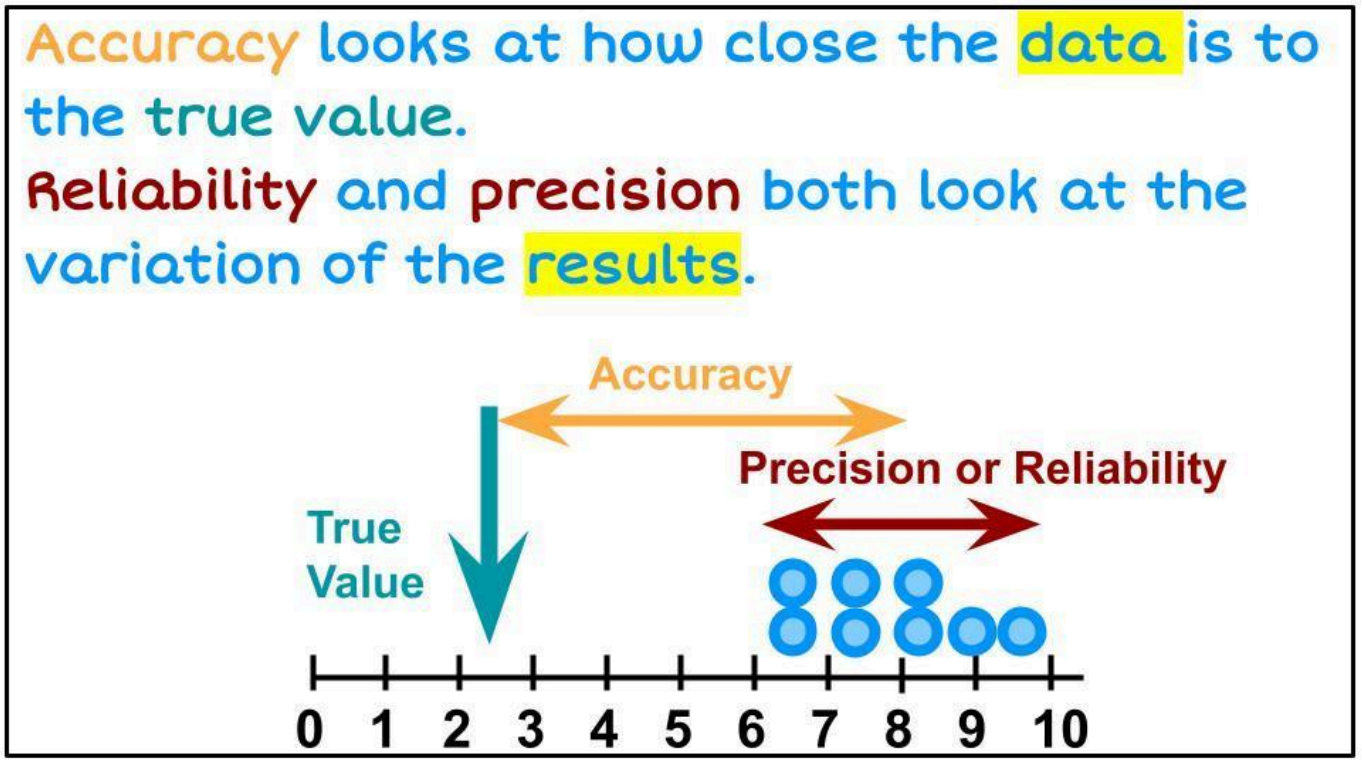
Explanatory Variable	
Response Variable	
Explanation why.	

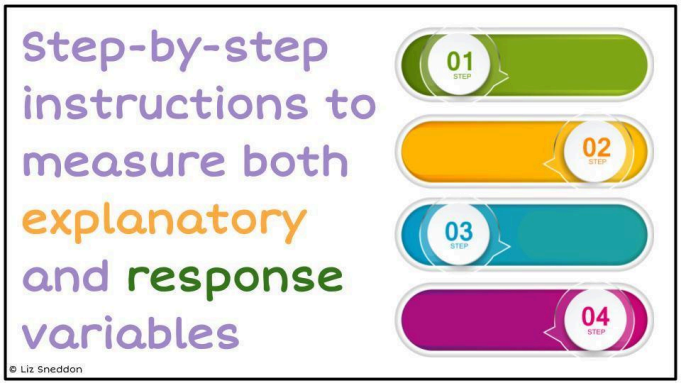
6) The daily temperature and the number of ice-cream cones sold.

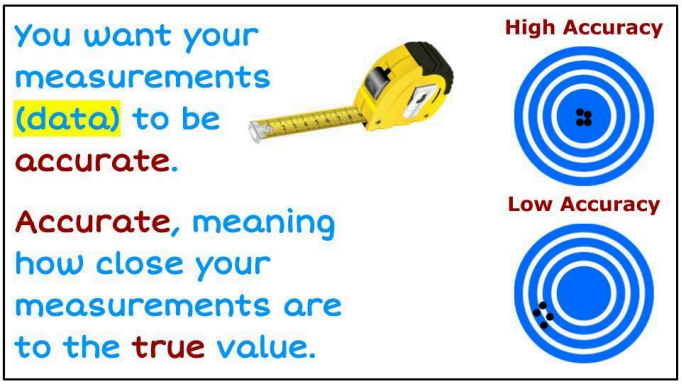
Explanatory Variable	
Response Variable	
Explanation why.	

Plan

In order to collect good quality data, there are a few things we want to think about. The main goal is to collect data that is accurate, leading to results that are reliable and precise.







Example:

Problem

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

Plan

For each student in our class (aiming for around 30 students), measure both the foot length and hand span following the instructions below.

For measuring foot length:

1) Ask the person to remove their right shoe.
2) Ask the person to place their right foot against a wall, facing outwards.
3) Make sure that their heel is touching the wall.
4) Using a tape measure, measure the distance (in mm) from the wall to the longest toe.
5) Record this measurement on a data table.

For measuring hand span:

1) Ask the person to place their right hand flat on a piece of paper on a desk, palm down.
2) Ask the person to spread their fingers as wide as they can.
3) Using a pen, mark the edge of the persons' smallest finger and thumb.
4) The person can now remove their hand.
5) Using a tape measure, measure the distance (in mm) between the two marks.
6) Record this measurement on a data table.

Exercises:

Look at the example plan and answer these questions.

1) Explain why we ask the student to remove their shoes.

2) What equipment would you suggest using to measure the following?

Time to run 100meters	
Weight of an apple	
Capacity of liquid in a bottle	
Volume of music on the radio	
Angle of a wheelchair ramp	

3) Circle the words that complete the sentences below.

Smaller sample sizes take a ___shorter / longer_____ time to collect data but give results that are _____more / less_____ reliable.

Larger sample sizes take a _____shorter / longer_____ time to collect data but give results that are _____more / less_____ reliable.

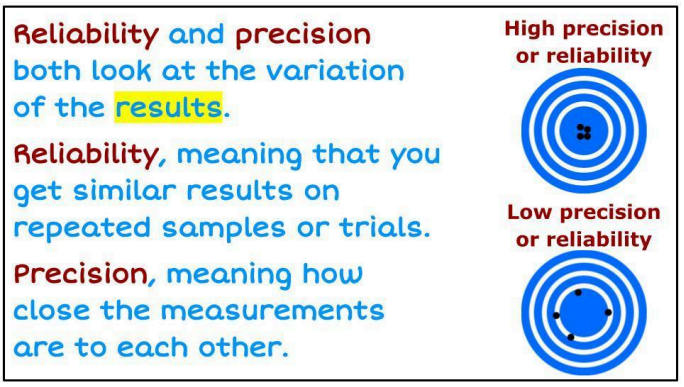
4) Working in groups of 2 or 3, write a set of instructions to walk from the doorway of the classroom to the opposite corner (without walking into the furniture, tripping over or walking into other students).

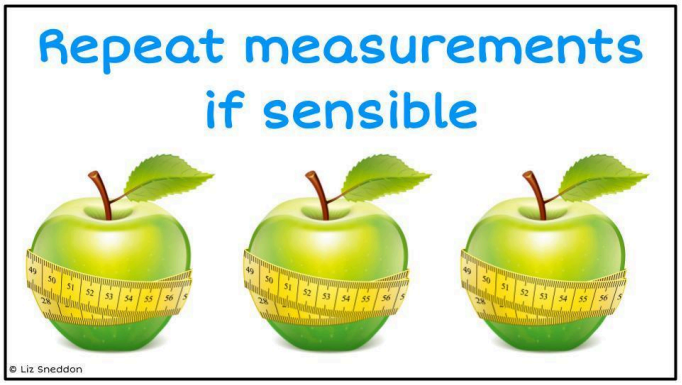
5) Give your instructions to your teacher to test them out.
As your teacher tests them, think about what improvements could you make and write these down.

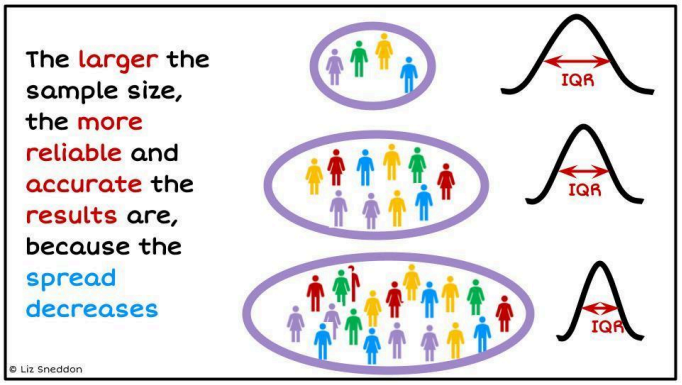
Controlling sources of variation

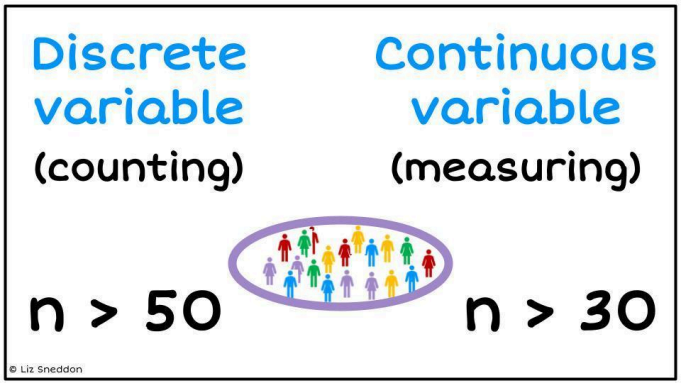
When coming up with your plan, you need to think about how you can minimise the amount of variation - making sure that all the measurements are done in the same way.

Here are some things to think about:









Example:

When measuring foot length and handspan, some of the factors I will control are:

- Using the same measuring tape, so that all the measurements are consistent.
- Getting students to put their hand down on a piece of paper, so that their hand is as flat as possible. This will make the measurements accurate.
- Get students to take their shoe off when I measure the length of their foot, because the different shoes people wear could have a different end (e.g. pointed, flat, curved) which would change the measurements and not be an accurate measurement of the length of their foot.

For Excellence, you want to reflect on what would happen if the source of variation had been ignored.

Exercises

Problem 1

I wonder if there is a relationship between a person's **height** and **weight** for students in your class?

Plan

Do a brainstorm below to think about how you can measure a person's height and weight accurately. What are some sources of variation that need to be controlled?

Write a detailed plan of how you are going to measure a person's height and weight, making sure you include controlling the sources of variation you identified.

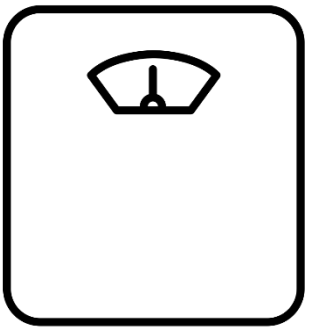
Explanatory variable: Height

Equipment:

Step by step instructions:

Response variable: Weight

Equipment:



Step by step instructions:

Problem 2

I wonder if there is a relationship between the length of the rotor blade and the time it takes the helicopter to hit the ground?

Information:

- Collect a template from your teacher. Notice that the helicopters have marks on the rotors every 0.5 cm, and you need to put a paper clip at the bottom.
- Make sure you drop the helicopters with a variety of sized rotor blades at least 30 different times. You can do several measurements with each rotor blade length.
- If made correctly, the helicopters will spin as they fall to the ground.
- We will drop the helicopters from the bridge by the Science Centre.

Plan

Do a brainstorm below to think about how you can measure the rotor length and the time it takes to drop to the ground accurately. What are some sources of variation that need to be managed / controlled?

Write a detailed plan of how you are going to measure the length of the rotor blade and the time it takes to reach the ground, making sure you include controlling the sources of variation you identified.

Things to consider…

1. *How will you measure each variable?*
2. *How much data will you be collecting?*
3. *Who will be doing what?*
4. *What will you do with the data?*

Explanatory Variable	
Response Variable	

Equipment:

Step by step instructions:

Instructions cont.

Data

You will need to measure and record the data for your investigation. You will need to set up a table similar to the one below.

Sample	Measurement 1	Measurement 2
1		
2		
3		
...		
30		

Exercise:

Collect your data with your helicopter, recording the data in the table below.

Trial Number	Rotor blade length (cm)	Time to hit the ground (seconds)
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

Trial Number	Rotor blade length (cm)	Time to hit the ground (seconds)
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		
26		
27		
28		
29		
30		

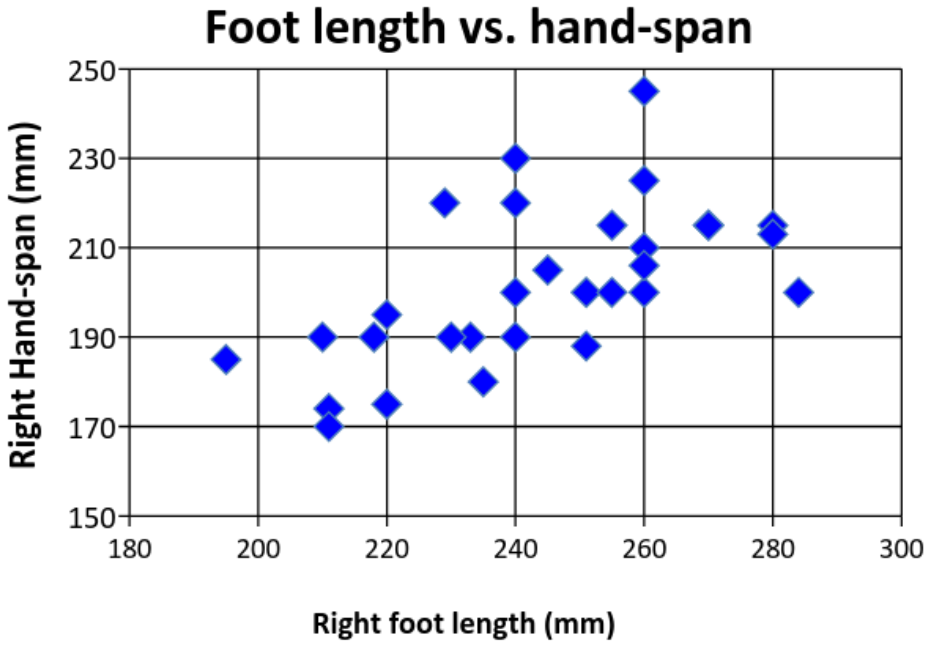
Insight for Excellence

There are a number of ways that you can show statistical and contextual insight.

You want to reflect on what improvements you could suggest if you collected another set of data. How could you better control any sources of variation that you didn't identify when you first wrote your plan?

Example:

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



Improvements:

I could improve my investigation by getting students to remove not just their shoe, but also their sock. This is because I noticed that some students 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.

Exercise:

Think back to the plan you made to investigate the relationship between the rotor length of the helicopter and the time it took to drop. What improvements can you suggest if you were to do this again?

Cleaning data

Look for the following issues:

- Data entry mistakes
- Incorrect units
- Missing data

But you **CANNOT** change/delete data unless you **KNOW** that it is a mistake.

If you are **CERTAIN** the data is wrong, then either correct the value or make the cell blank (or enter a 0).

Exercise:

1) Find any data that doesn't make sense and highlight the values.

Elastic type	Length (cm)	Number of marbles	Stretched length (cm)	Weight (grams)	Obstacles
WIDE	69 cm	6	890 mm	34.8gm	Yes
widE	690	8	92	46.4	no
	0.69 meters	10	95	58	YES
cord		12	98 cm	69.6 grams	NO
CORd		14	99	81.2	No
CORD		16	1020 mm	92.8	Yes
Cord	690	18	106	104.4 gm	yes
string	690	20	1080 mm	116	
Narrow	68.03 cm	-10	1.16meters	-58	yes

Write the corrected values in the table below.

A	B	C	D	E	F
Elastic type	**Length (cm)**	**Number of marbles**	**Stretched length (cm)**	**Weight (grams)**	**Obstacles**

Entering Data into NZGrapher

You will be using NZGrapher to draw a graph of your data.

You will need to enter your data into a spreadsheet first (either Excel or Google Sheets), including the headings (names of the variables).

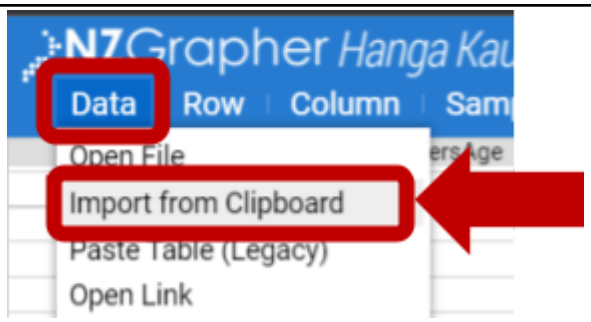
Example:

Here are two variables from the Fitness dataset:

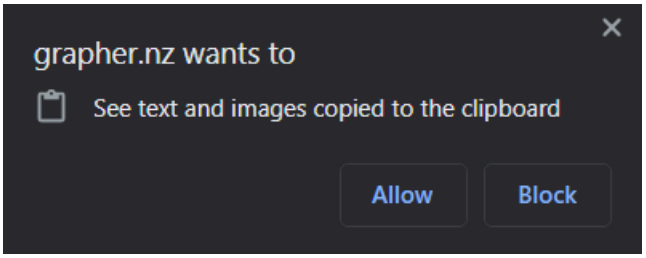
Leg up Wall sit time (seconds)	Height (cm)
15	168
31	187
40	170
12.18	156.5
21	147

Steps for entering data into NZGrapher

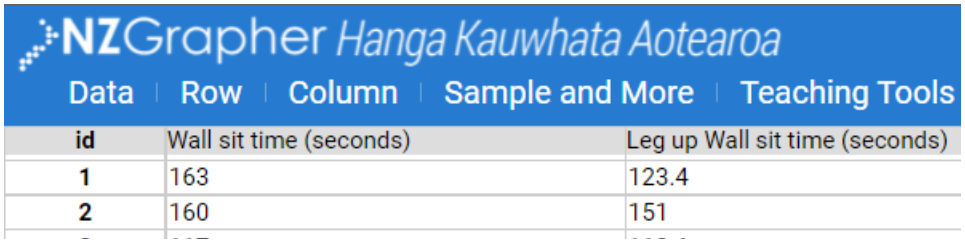
Step 1: The quickest way to get data into NZGrapher is to **select all** of the data in your spreadsheet and **copy** this.	**Step 2:** Next, go to the NZGrapher website.
Step 3: Click on the **Data** menu and select "**Import from Clipboard**".	**Step 4:** NZGrapher will ask you to give your permission to copy information from the clipboard.



Click "**Allow**".



This then imports your data into NZGrapher:

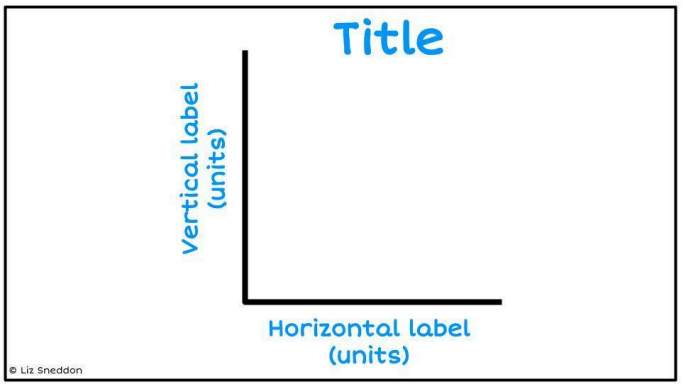


id	Wall sit time (seconds)	Leg up Wall sit time (seconds)
1	163	123.4
2	160	151

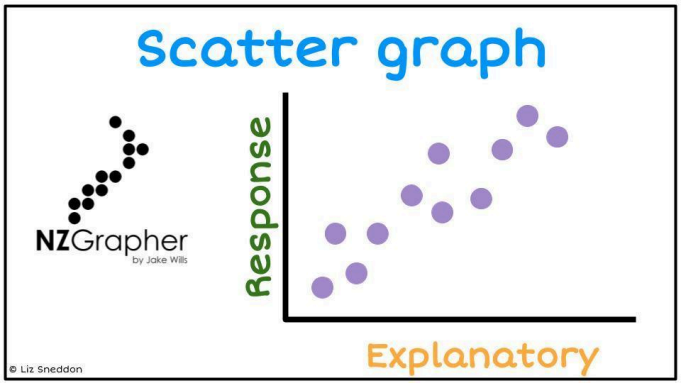
Drawing Scatter Graphs

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.



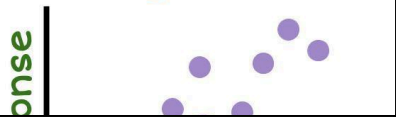
Steps to draw a scatter graph in NZGrapher

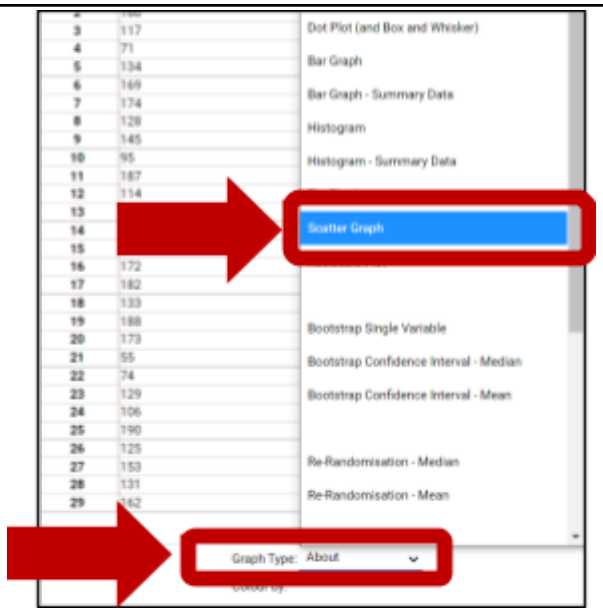
Step 1:

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

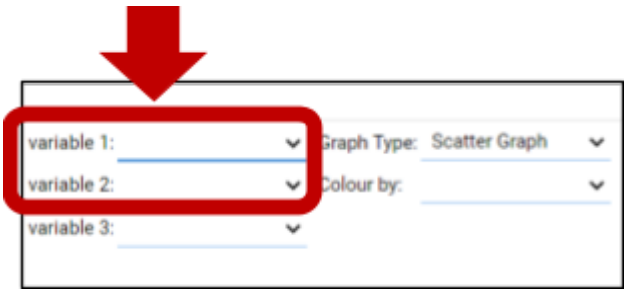




them the correct way around.

(Variable 1 = Explanatory variable)

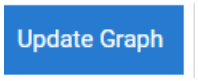
(Variable 2 = Response variable)



Step 3:

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

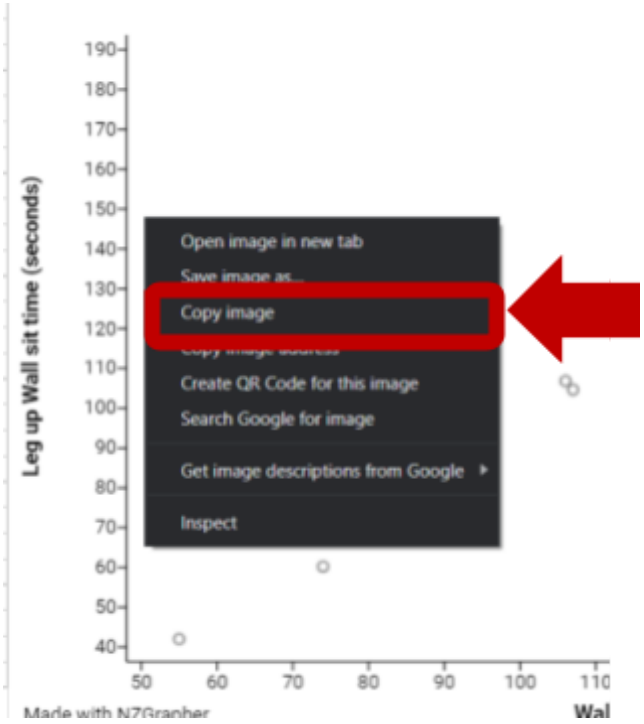
Once you have entered the labels and title, then click the **Update Graph** button (it's on the right-hand side at the bottom of the page).



Step 4:

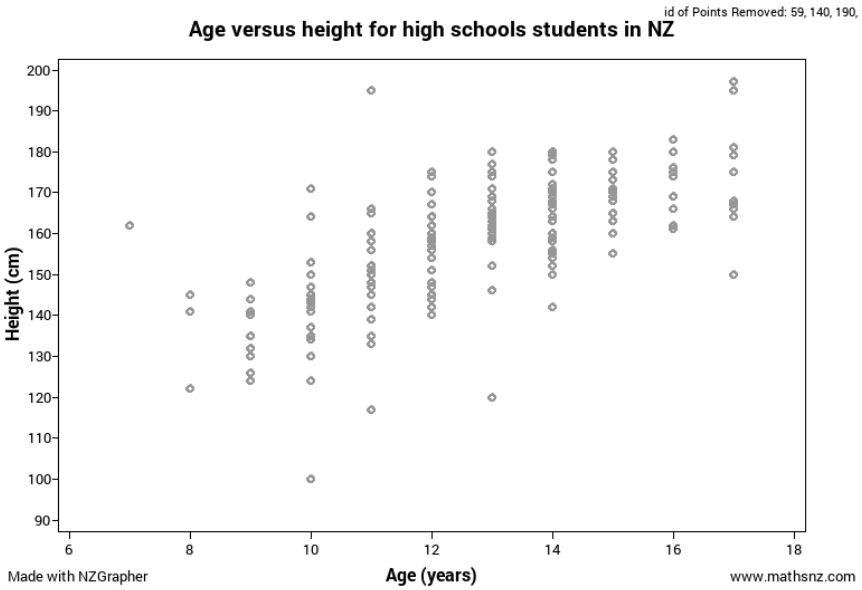
The last thing you need to do is to copy the graph and paste it into your document.

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**"



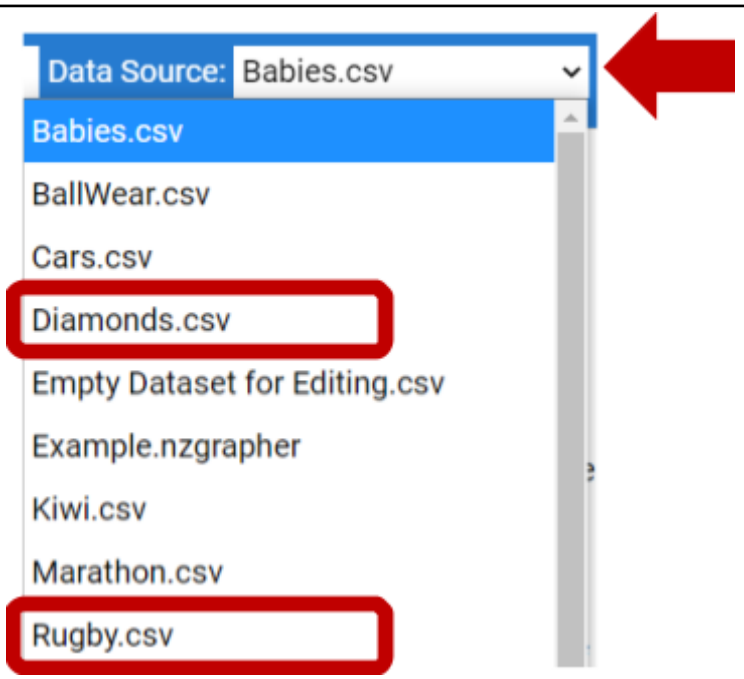
Then go to your document and paste the image.

Example:



Exercise:

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) Rugby.csv

Explanatory Variable	
Response Variable	

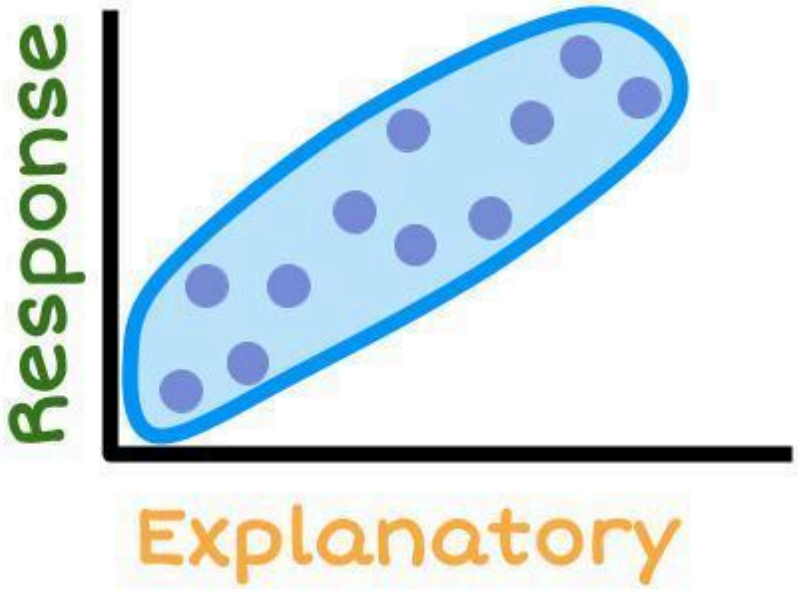
3) Go back to your data from the helicopter. Enter this data into a spreadsheet, clean the data, copy the data into NZGrapher and draw a scatter graph.

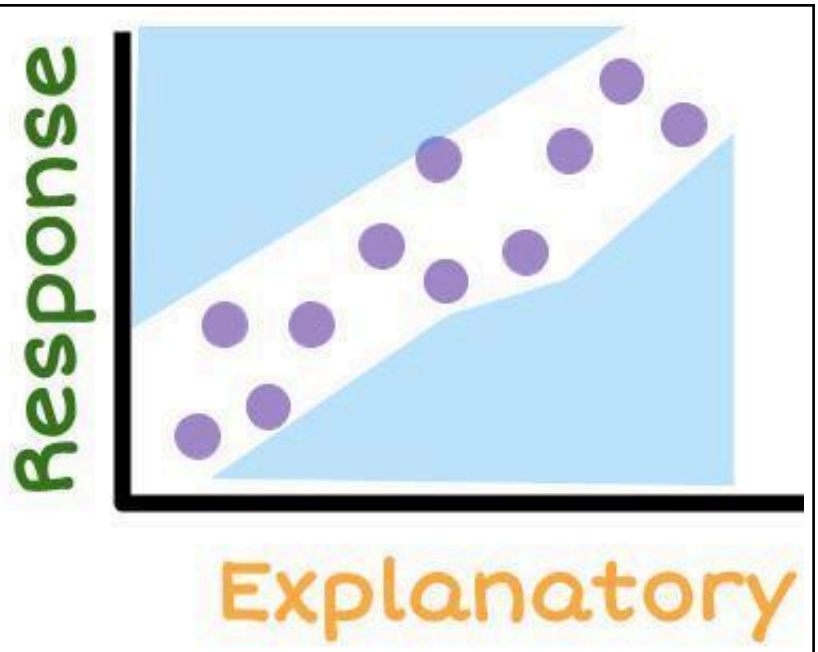
Explanatory Variable	
Response Variable	

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.

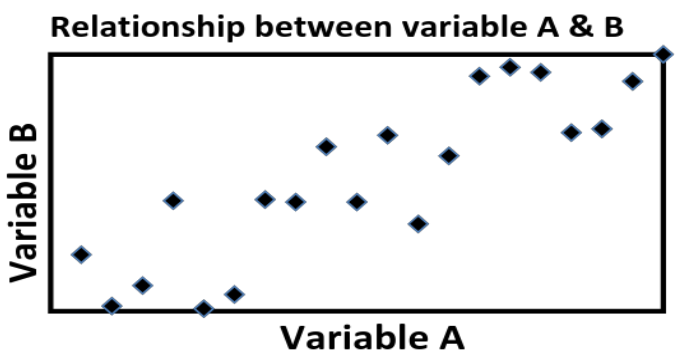
There are 2 methods that I'm going to suggest that you could use to help visually focus on the data. You can choose to use any of these.

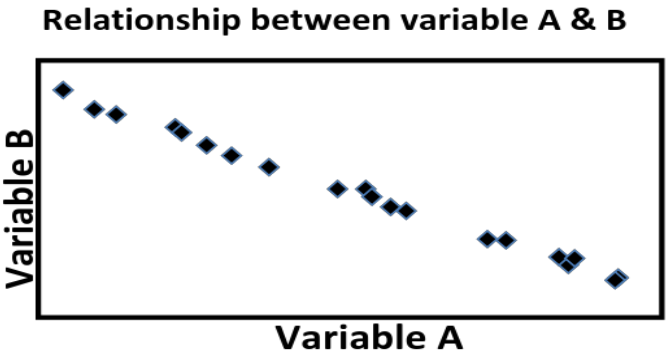
Method 1: Shading in. Draw a line above and below the data, the shade the area **between** these lines.	

Method 2: Shading out. Shade the areas where there is **not** any data.	

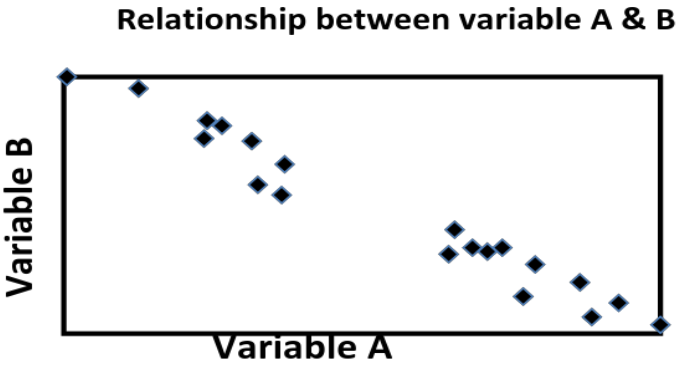
Exercise:

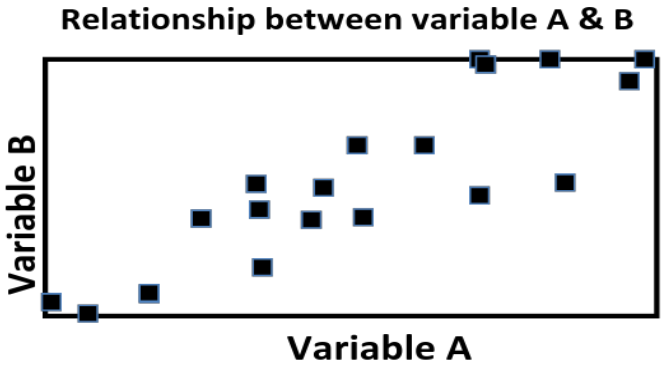
1) Using a highlighter, shade **IN** the following graphs.





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

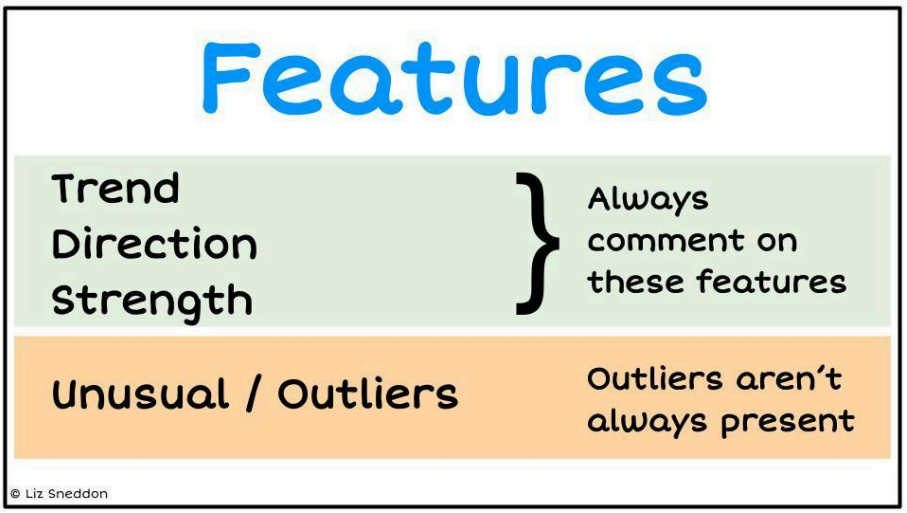




Features

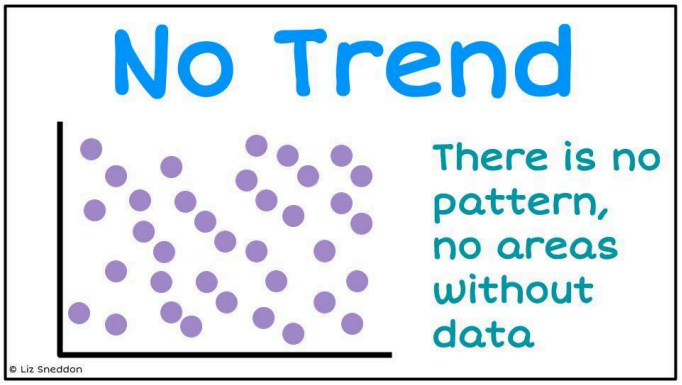
Once we have a visual picture of the data, we want to identify different features.

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

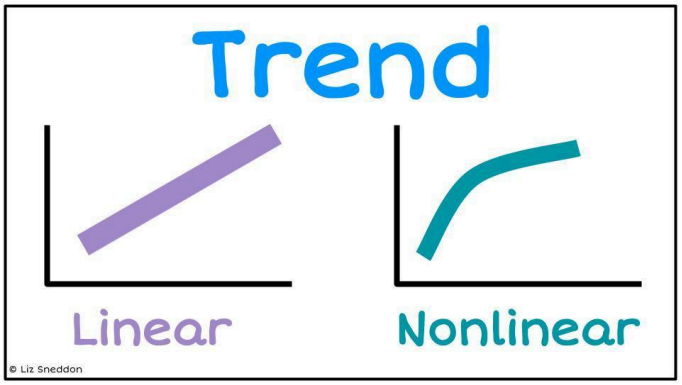


Trend

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

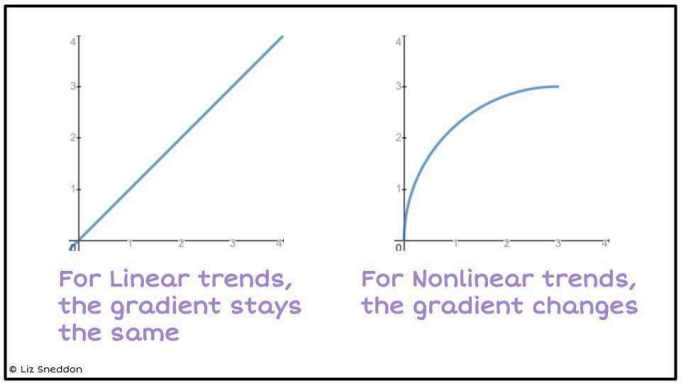


If there is a pattern we need to decide if it looks more like a linear or non-linear pattern (curved).



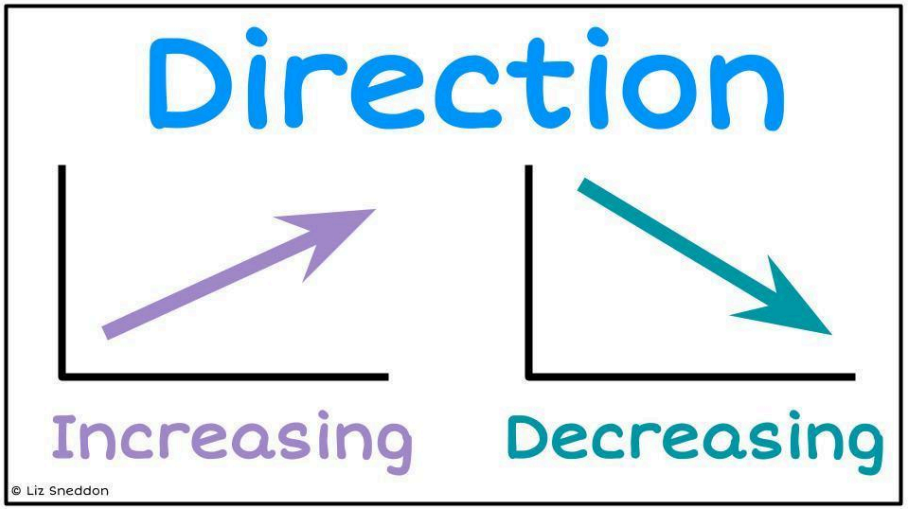
To justify or explain why we make this decision, we need to think about whether the gradient / slope is steady / constant, or whether the gradient is changing.

(The gradient measures how steep something is.)

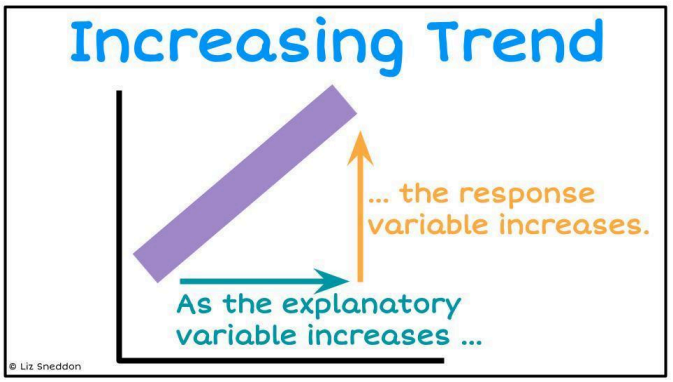


Direction

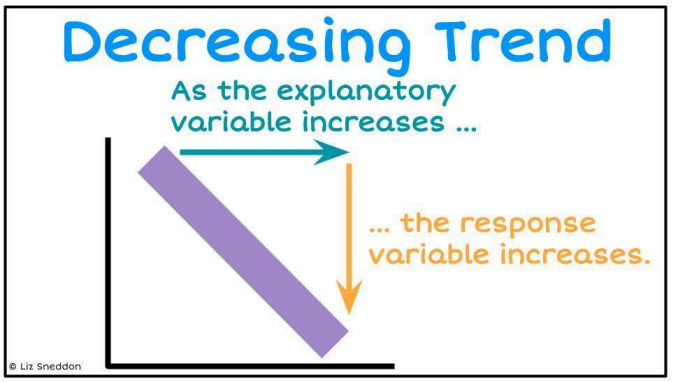
Once we know that there is a trend, the next step is to identify whether the trend is going up (positive or increasing) or going down (negative or decreasing).



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**.

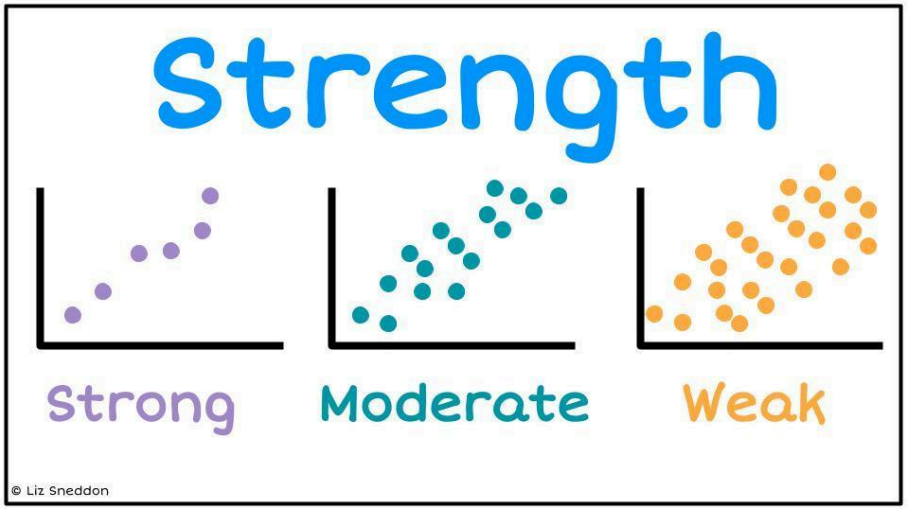


To justify or explain a decreasing trend, we always discuss (in context) how as the **explanatory variable** (horizontal variable) **increases**, the **response variable** (vertical variable) **decreases**.



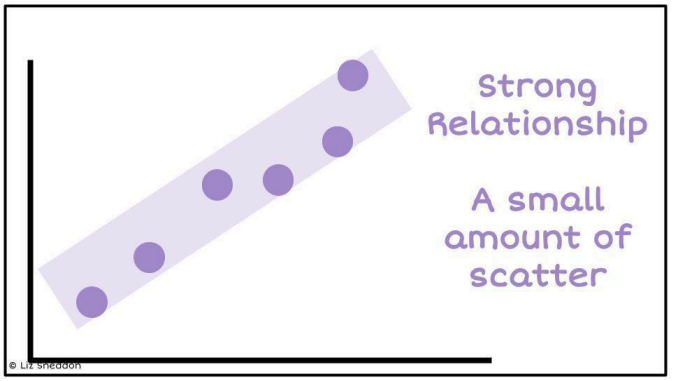
Strength

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



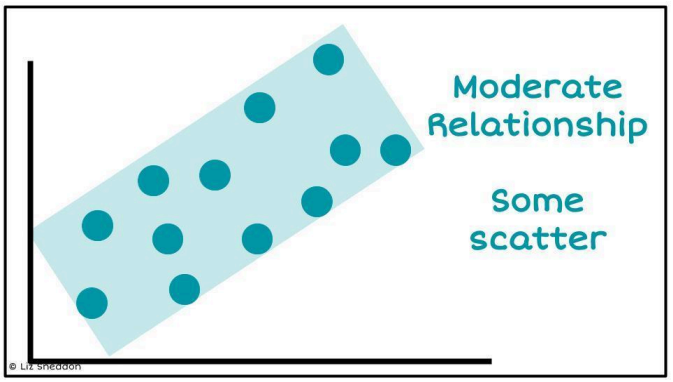
To justify or explain a **strong** relationship, we look at **how much scatter or variation** there is.

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



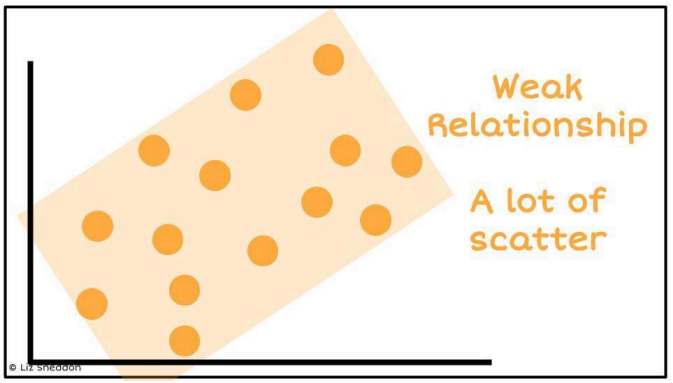
To justify or explain a **moderate** relationship, we look at **how much scatter or variation** there is.

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



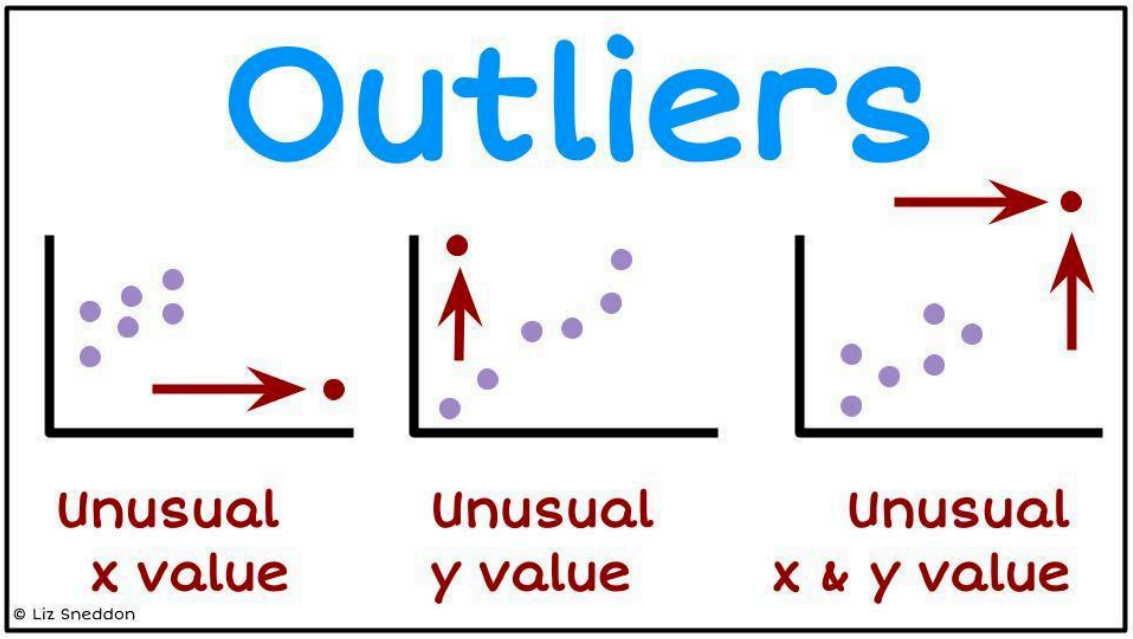
To justify or explain a **strong** relationship, we look at **how much scatter or variation** there is.

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

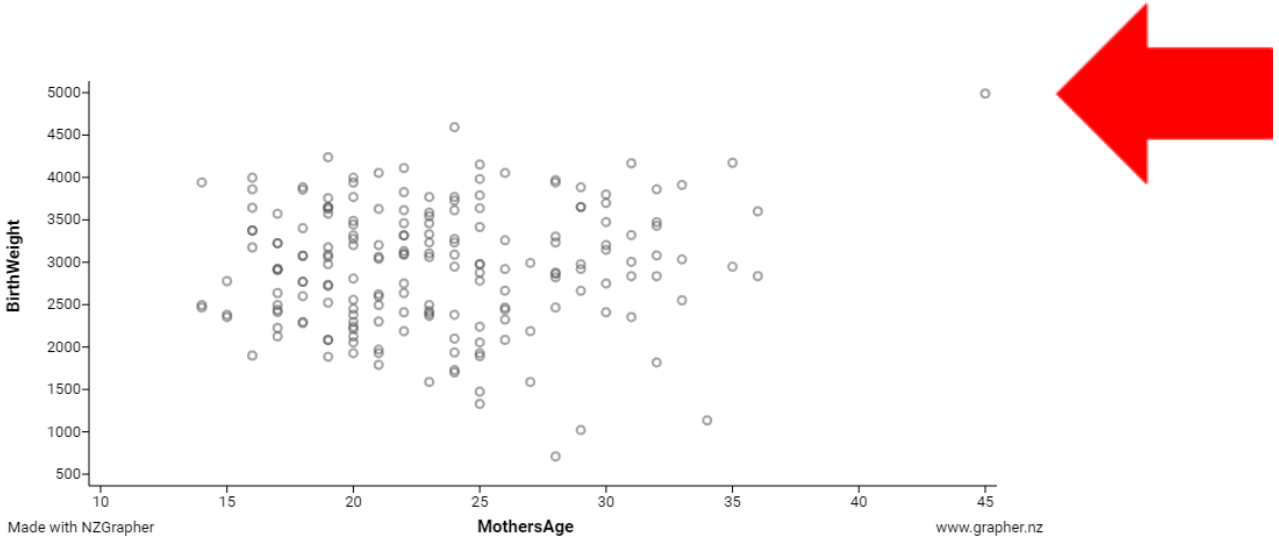


Outliers

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:

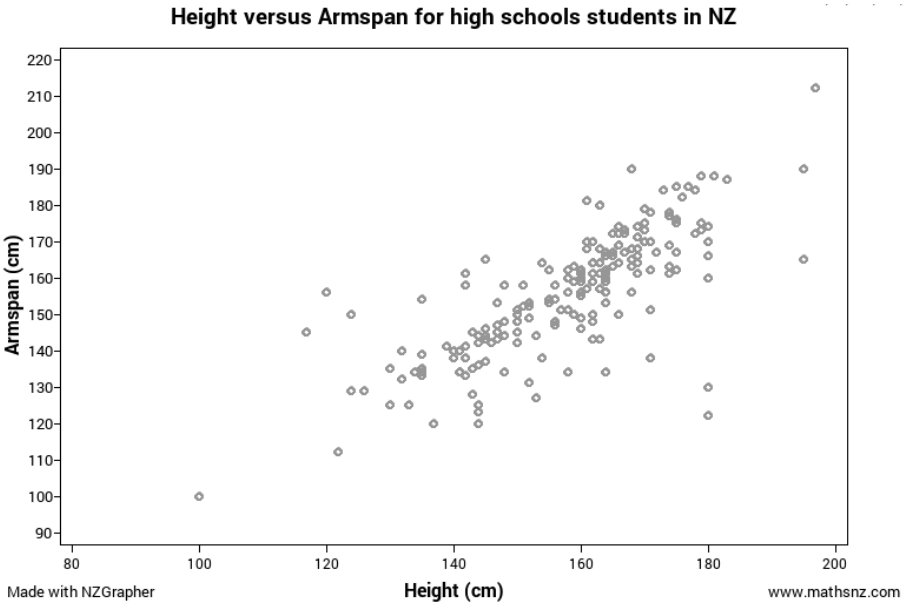


Looking at the graph above of the mothers age and birthweight of her baby, we can see that there is an outlier - this is a mother who is 45 years of age and the baby weighs around 5000grams. The mother's age and the baby's weight are both unusually large.

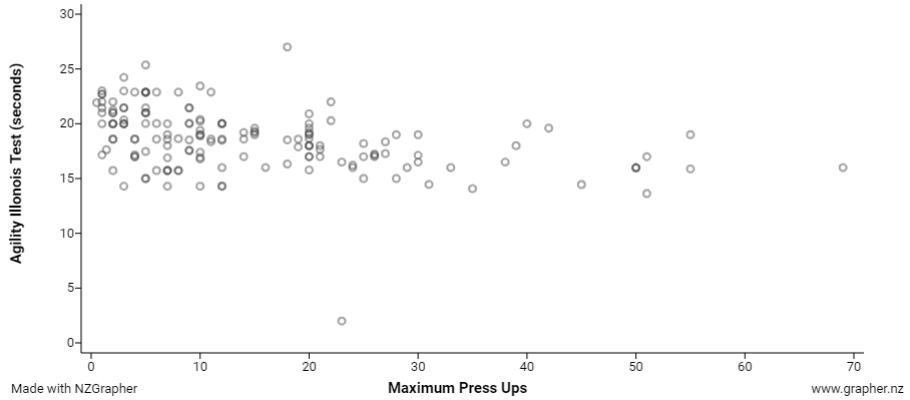
Exercise:

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

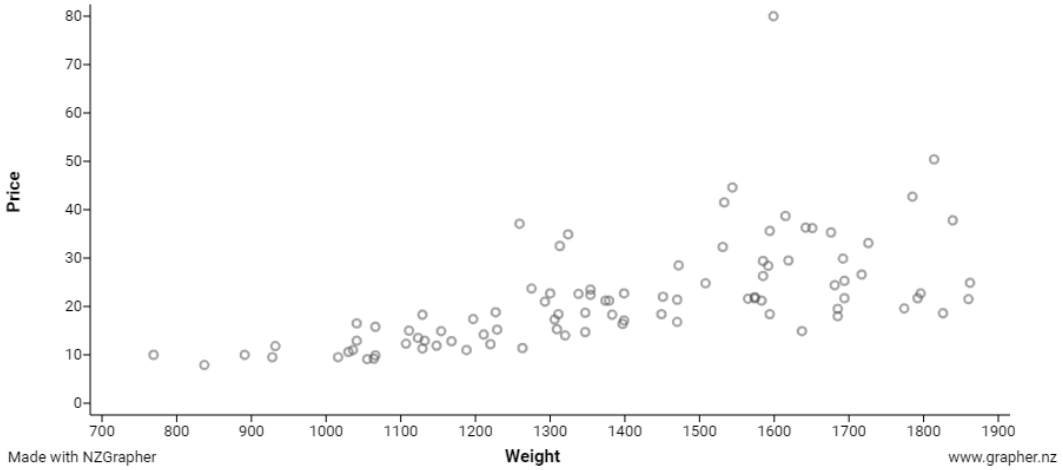
1)



2)

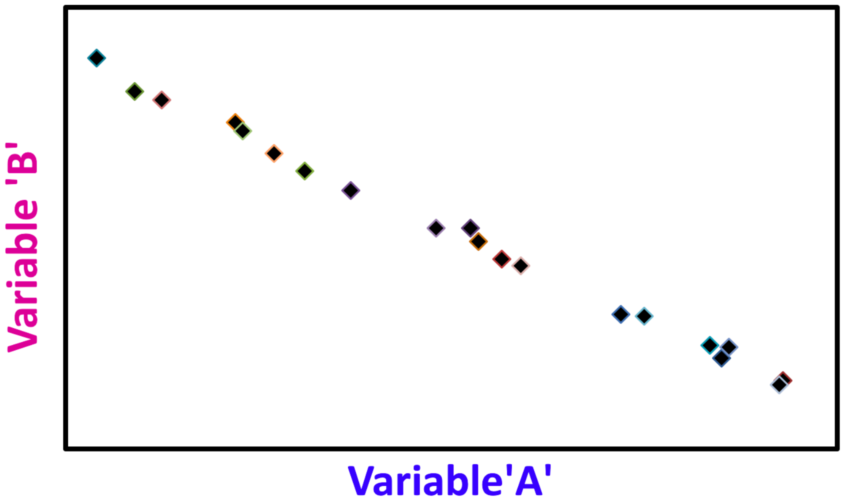


3)



Exercise:

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.



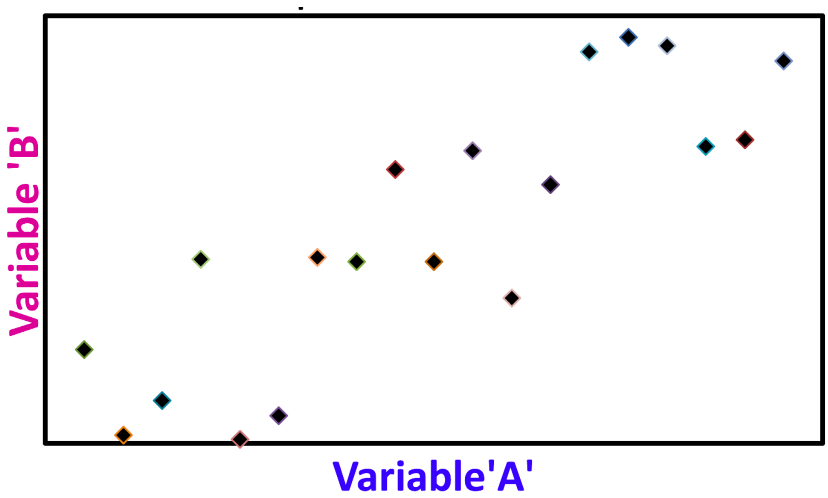
Relationship:
Yes / No

Trend:
Linear / Non-linear

Direction:
Positive / Negative

Strength:
Strong / Moderate / Weak

Outliers: (circle any points)
Yes / No



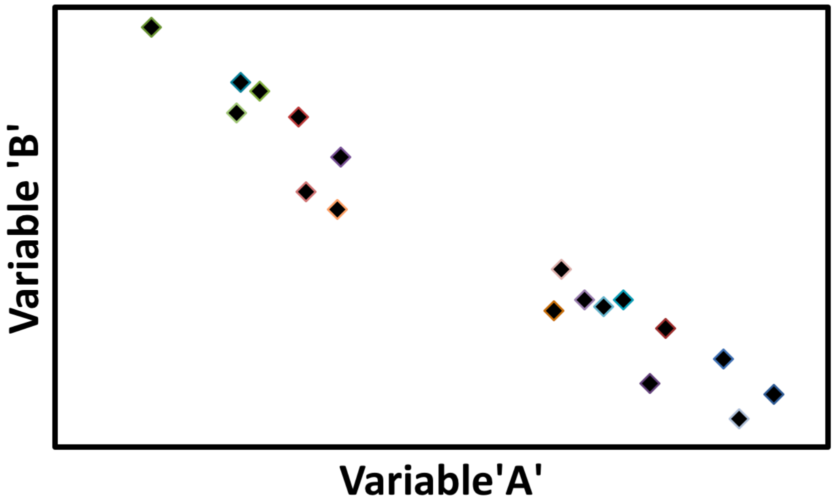
Relationship:
Yes / No

Trend:
Linear / Non-linear

Direction:
Positive / Negative

Strength:
Strong / Moderate / Weak

Outliers: (circle any points)
Yes / No



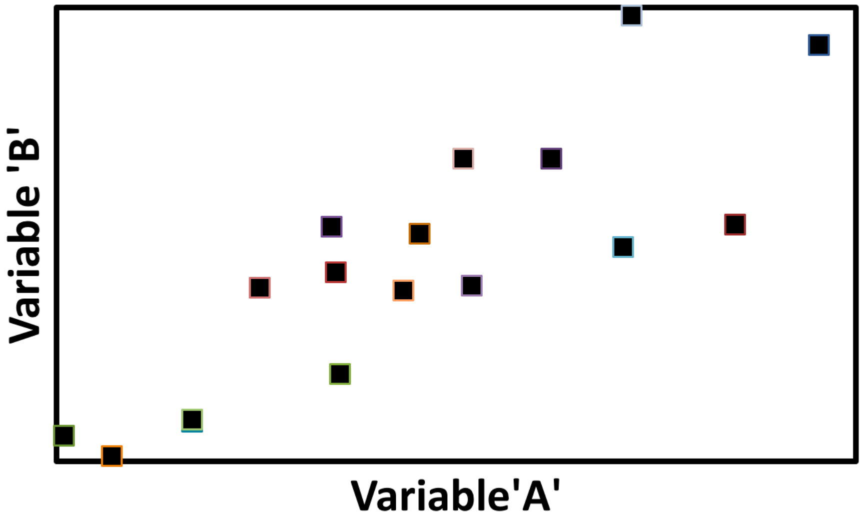
Relationship:
Yes / No

Trend:
Linear / Non-linear

Direction:
Positive / Negative

Strength:
Strong / Moderate / Weak

Outliers: (circle any points)
Yes / No



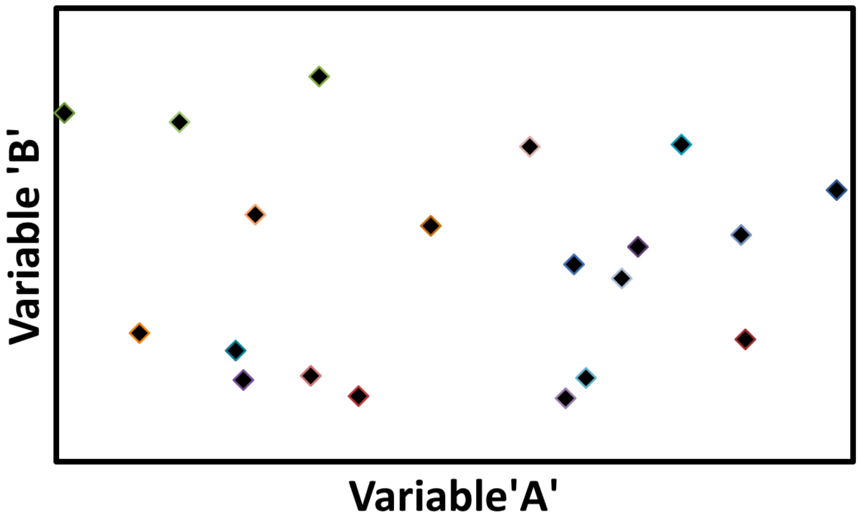
Relationship:
Yes / No

Trend:
Linear / Non-linear

Direction:
Positive / Negative

Strength:
Strong / Moderate / Weak

Outliers: (circle any points)
Yes / No



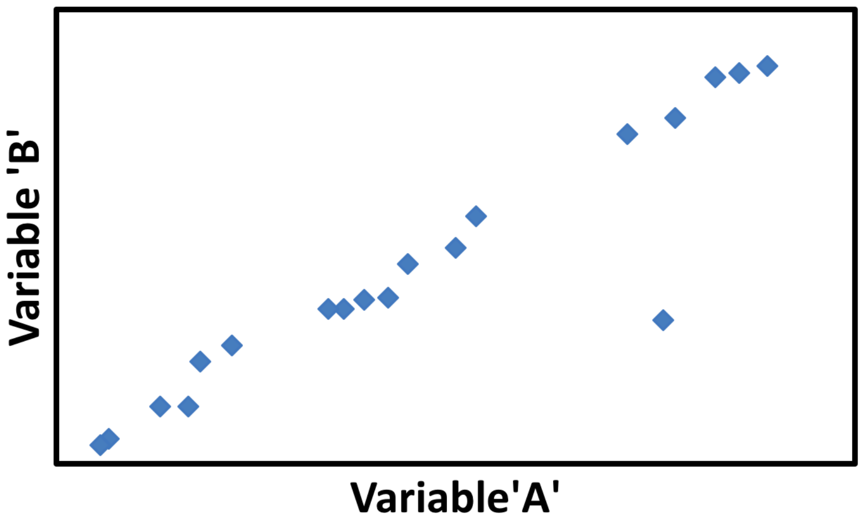
Relationship:
Yes / No

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

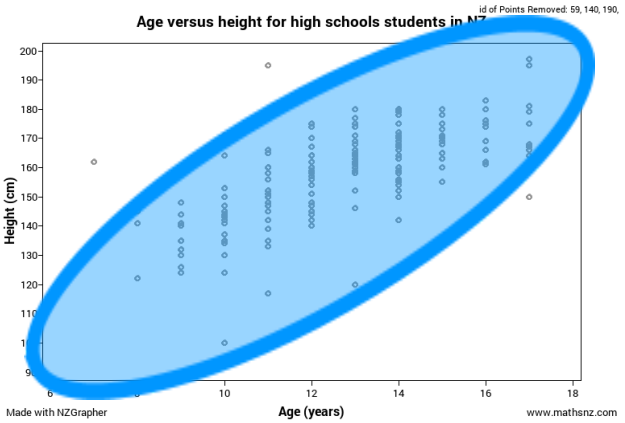
Outliers: (circle any points)
Yes / No

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:



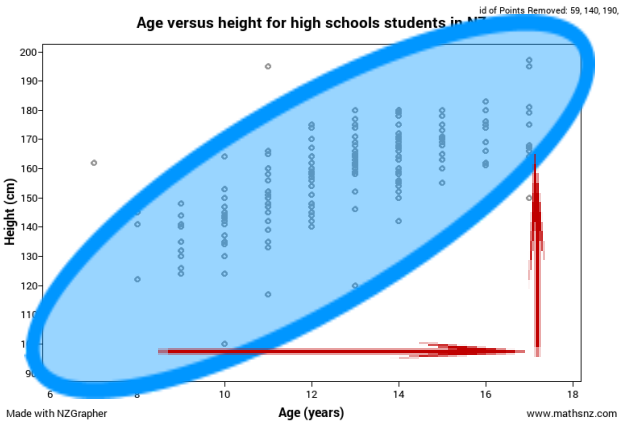
Identify:

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

Justify:

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

Direction:



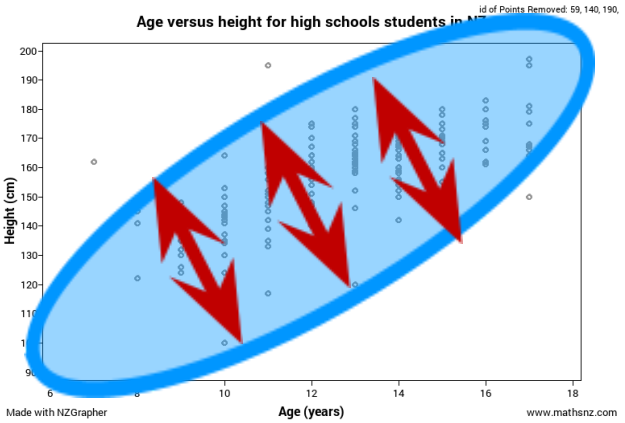
Identify:

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

Justify:

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

Strength:



Identify:

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

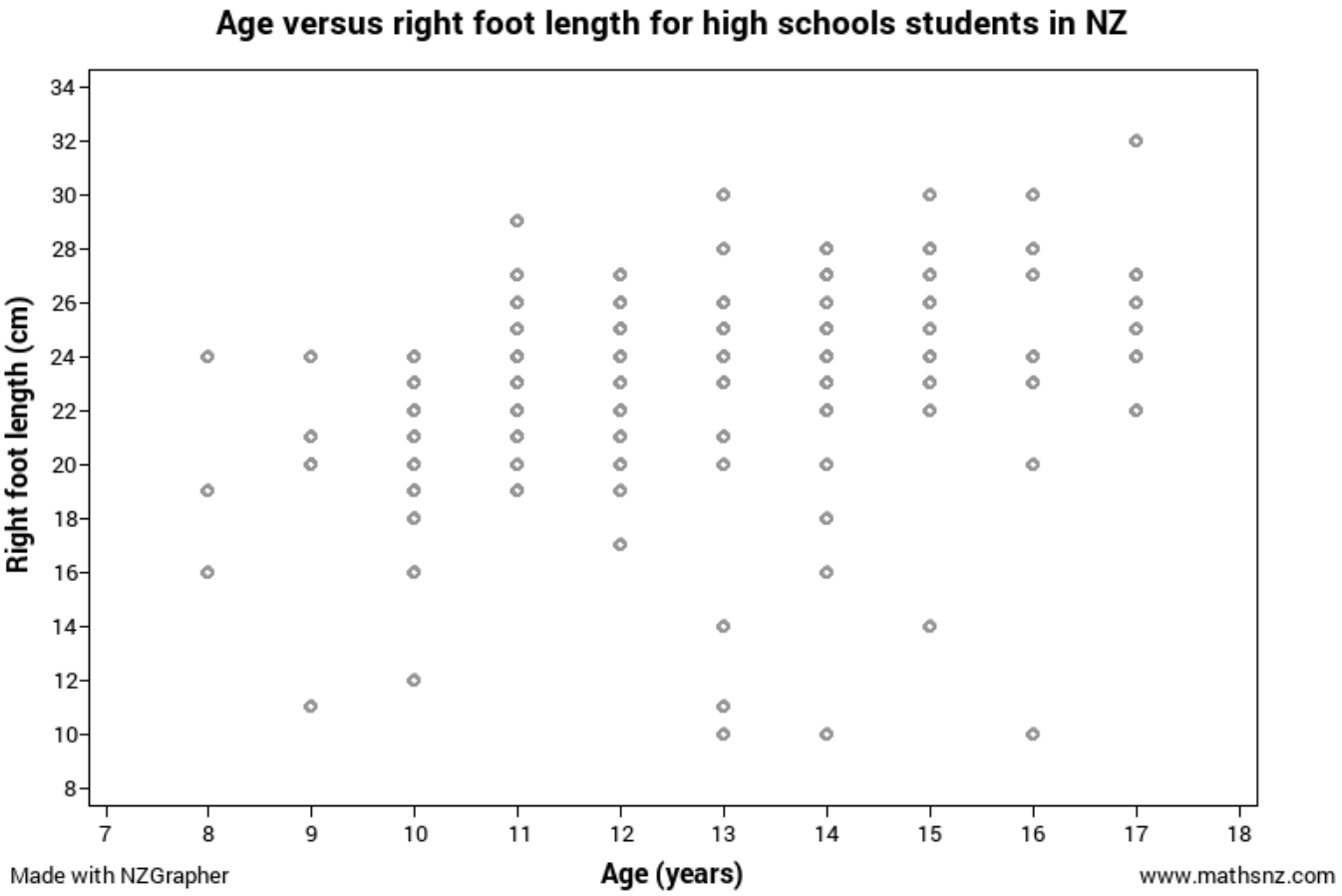
Justify:

It is weak because the data points are widely scattered.

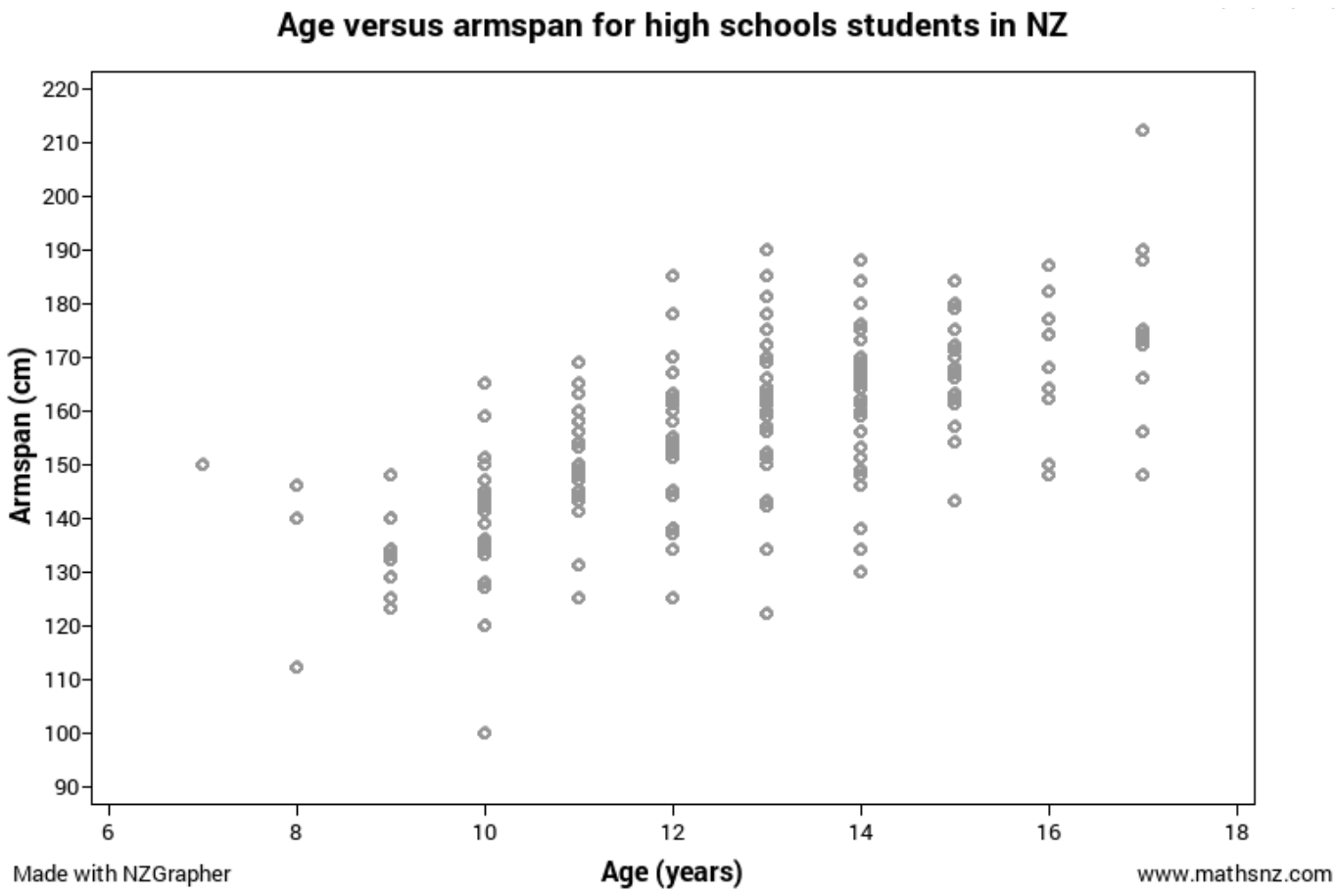
Exercise:

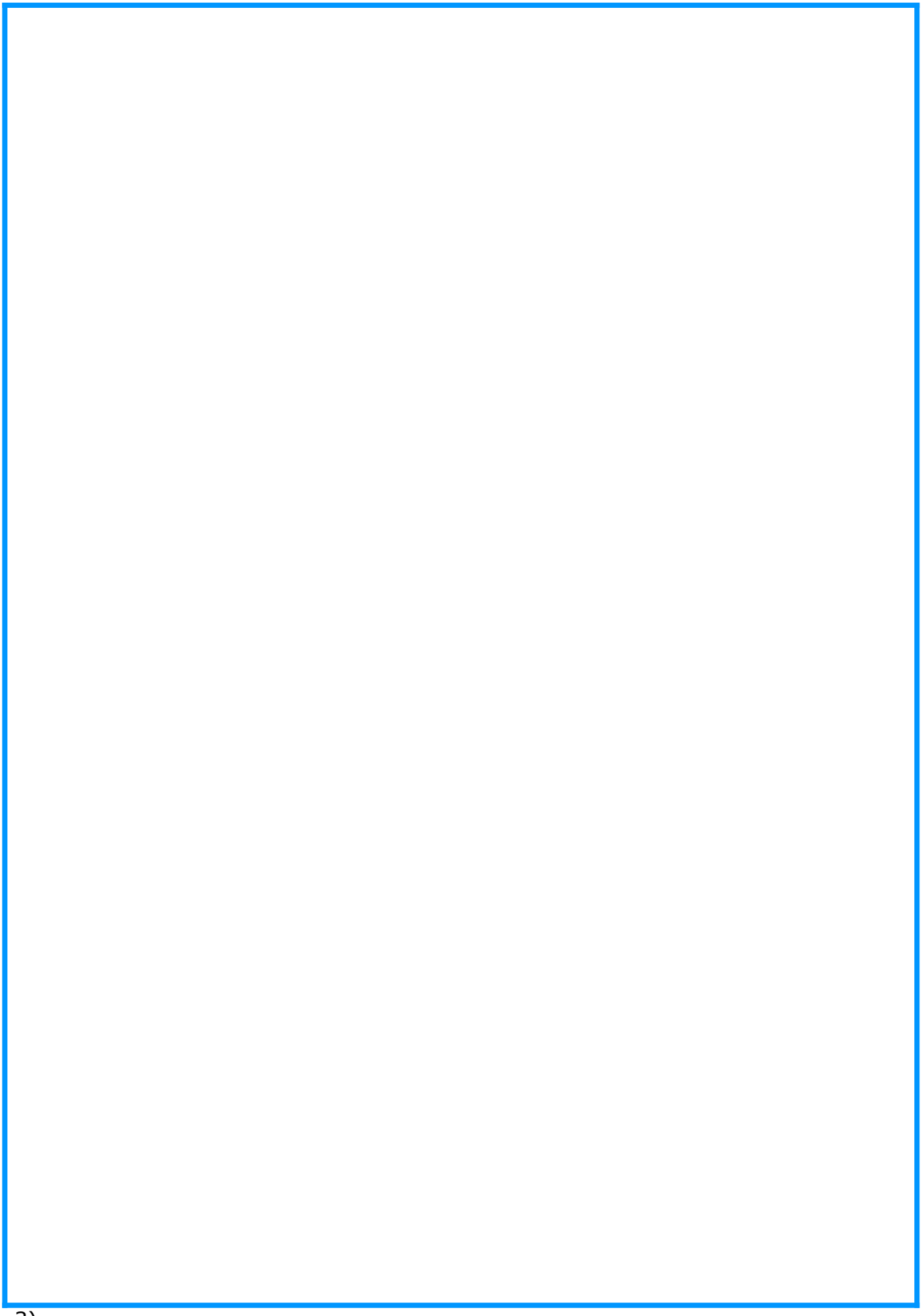
For the following graphs, shade the data, then describe the **features** (trend, direction and strength) in **context**.

1)

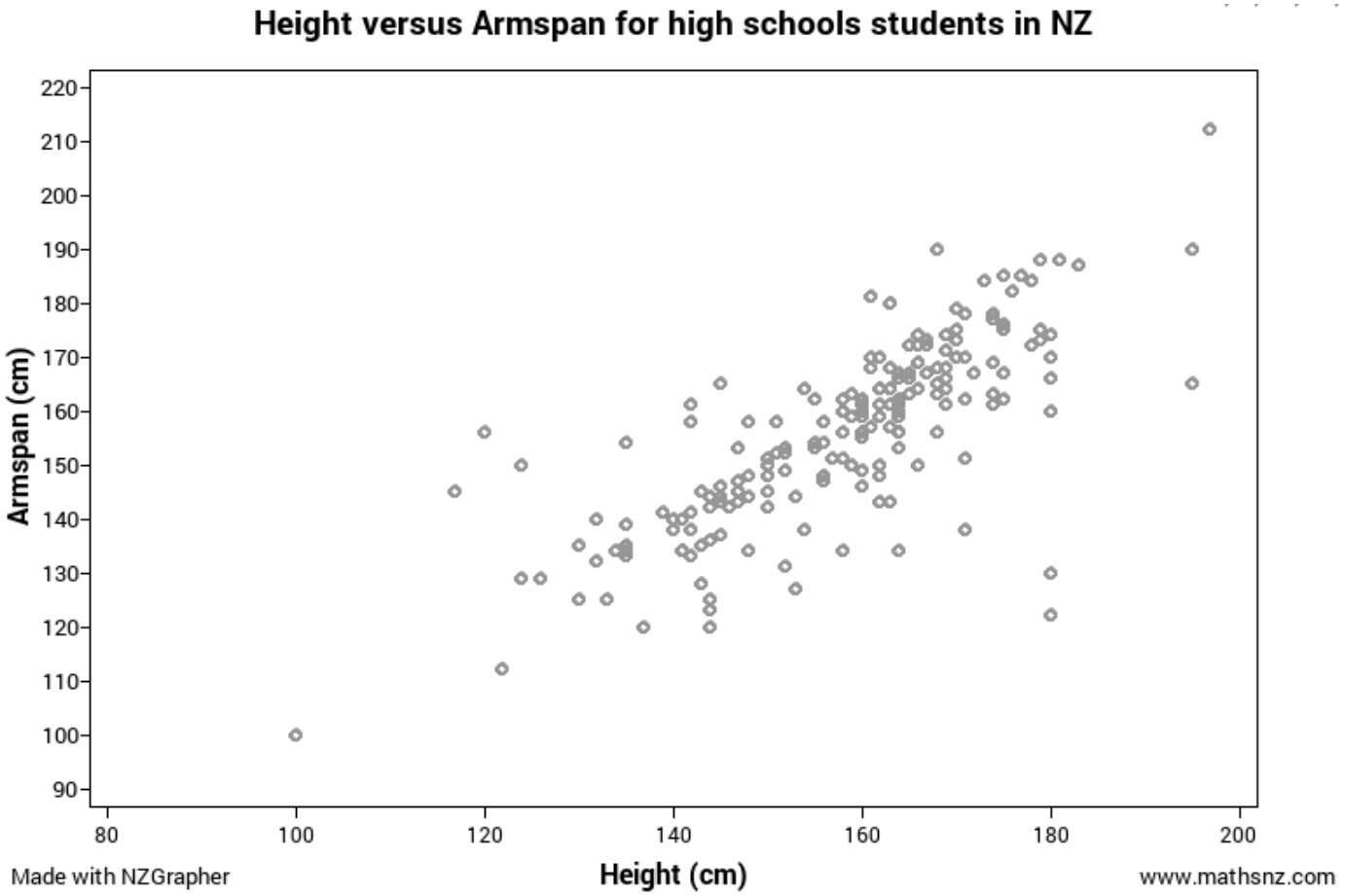


2)

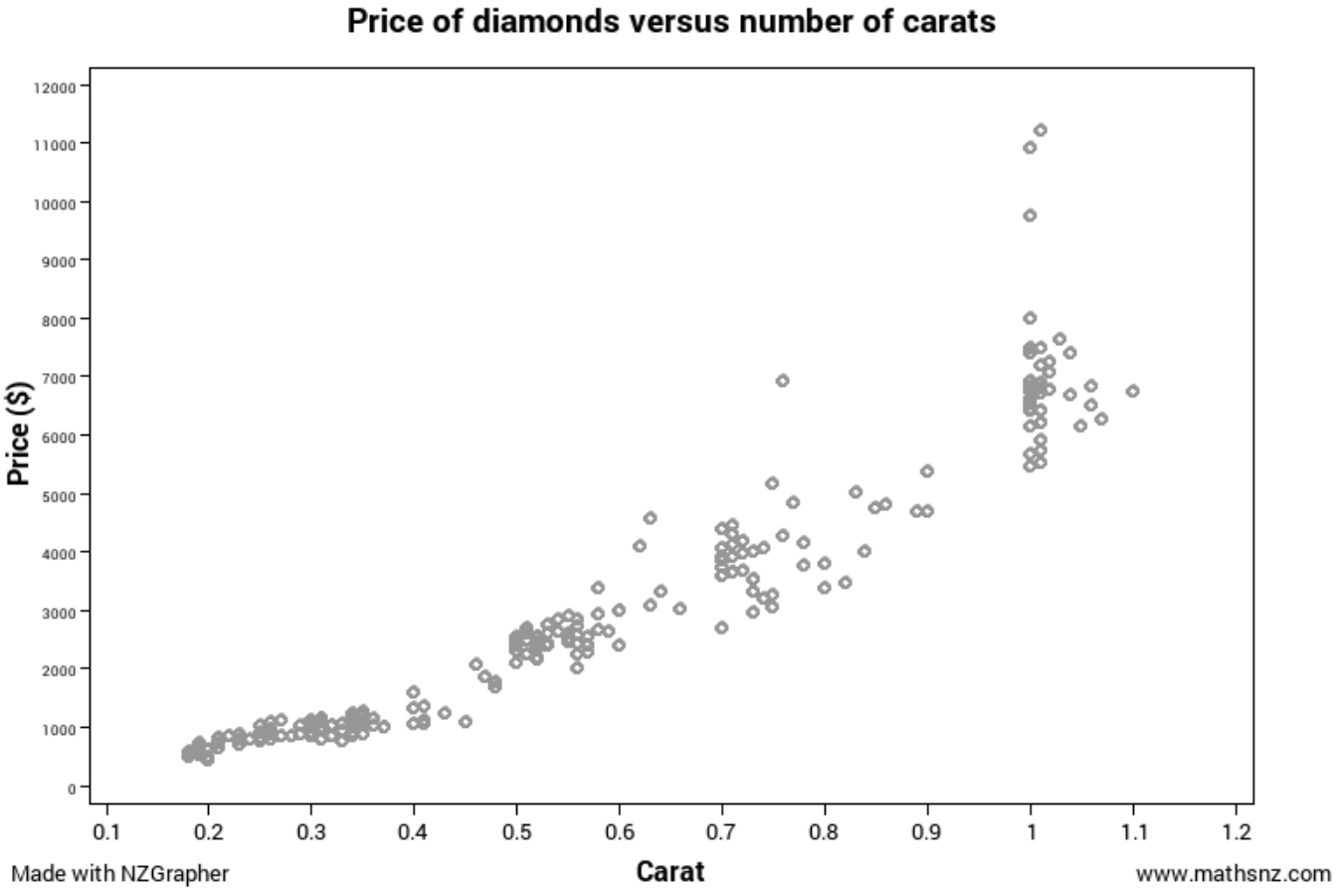

Age versus armspan for high schools students in NZ
Armspan (cm)
220
210
200
190
180
170
160
150
140
130
120
110
100
90
6
8
10
12
14
16
18
Age (years)
Made with NZGrapher
www.mathsnz.com



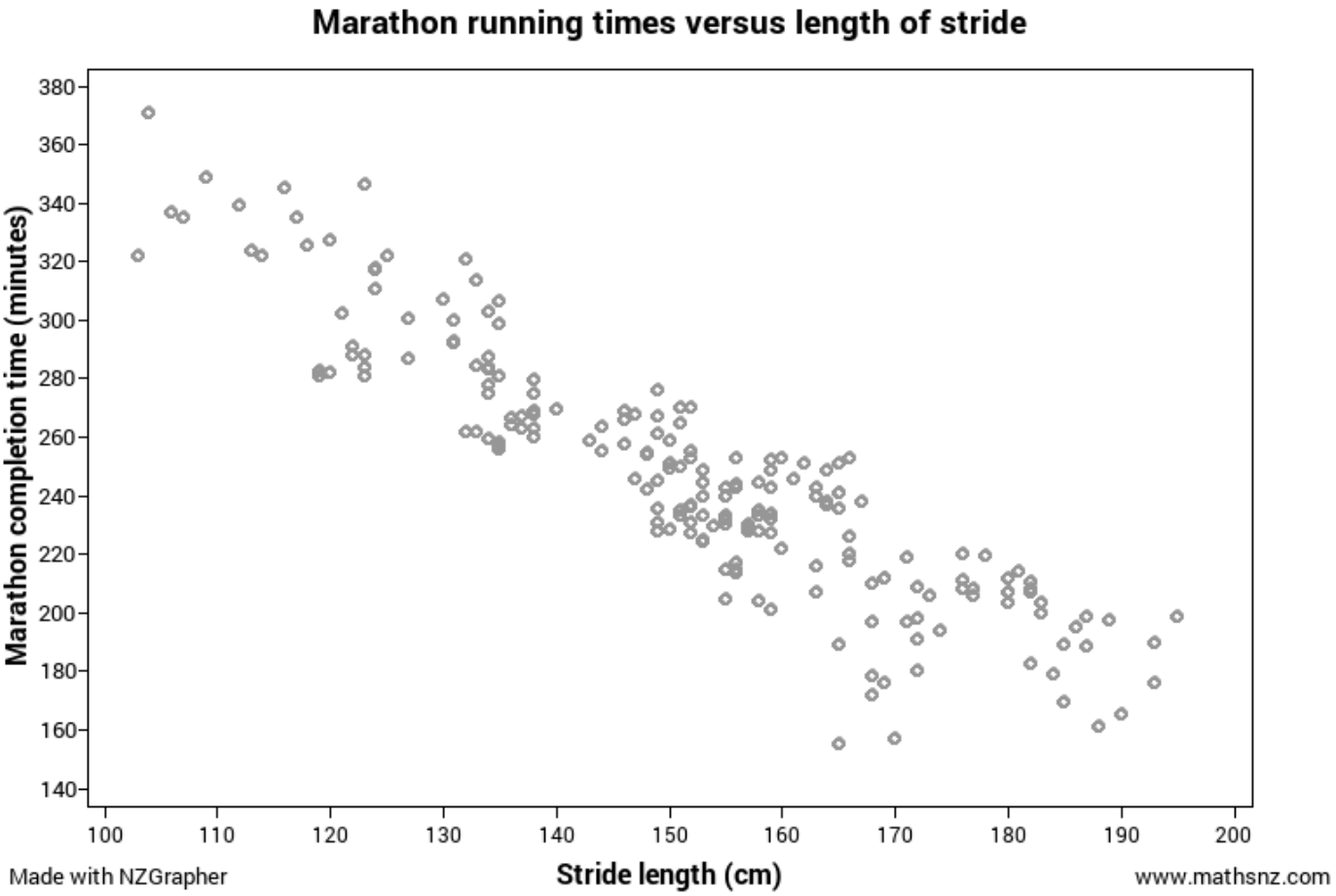
3)

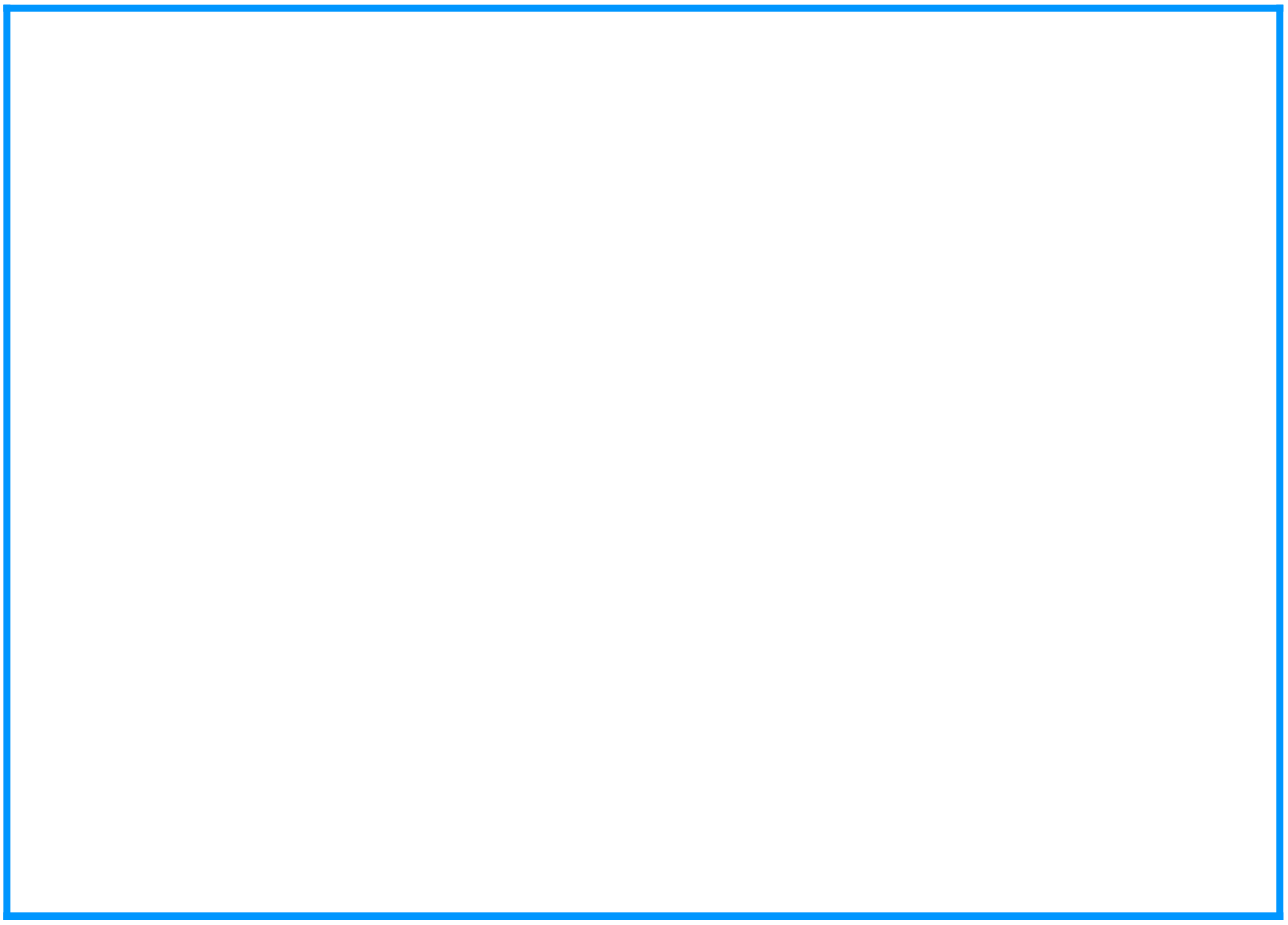

Height versus Armspan for high schools students in NZ
Armspan (cm)
220
210
200
190
180
170
160
150
140
130
120
110
100
90
80
100
120
140
160
180
200
Height (cm)
Made with NZGrapher
www.mathsnz.com

4)

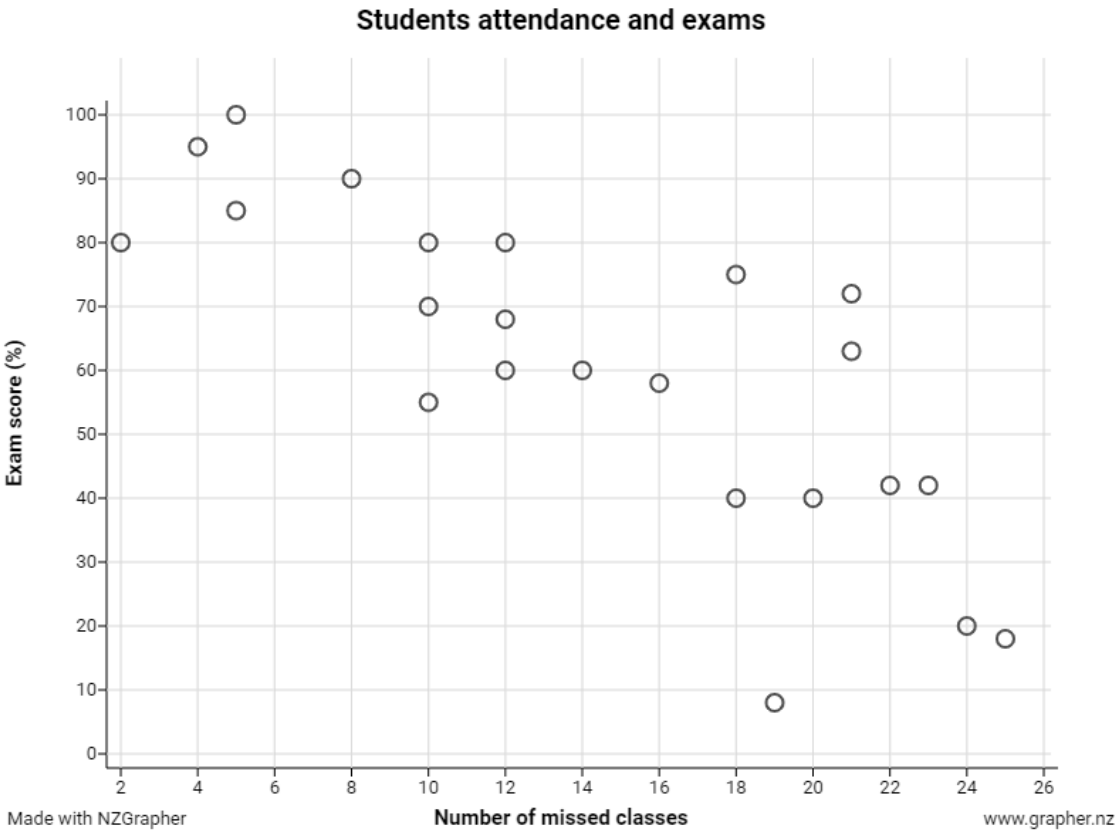

Price of diamonds versus number of carats
Price ($)
12000
11000
10000
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1.1
1.2
Carat
Made with NZGrapher
www.mathsnz.com

5)

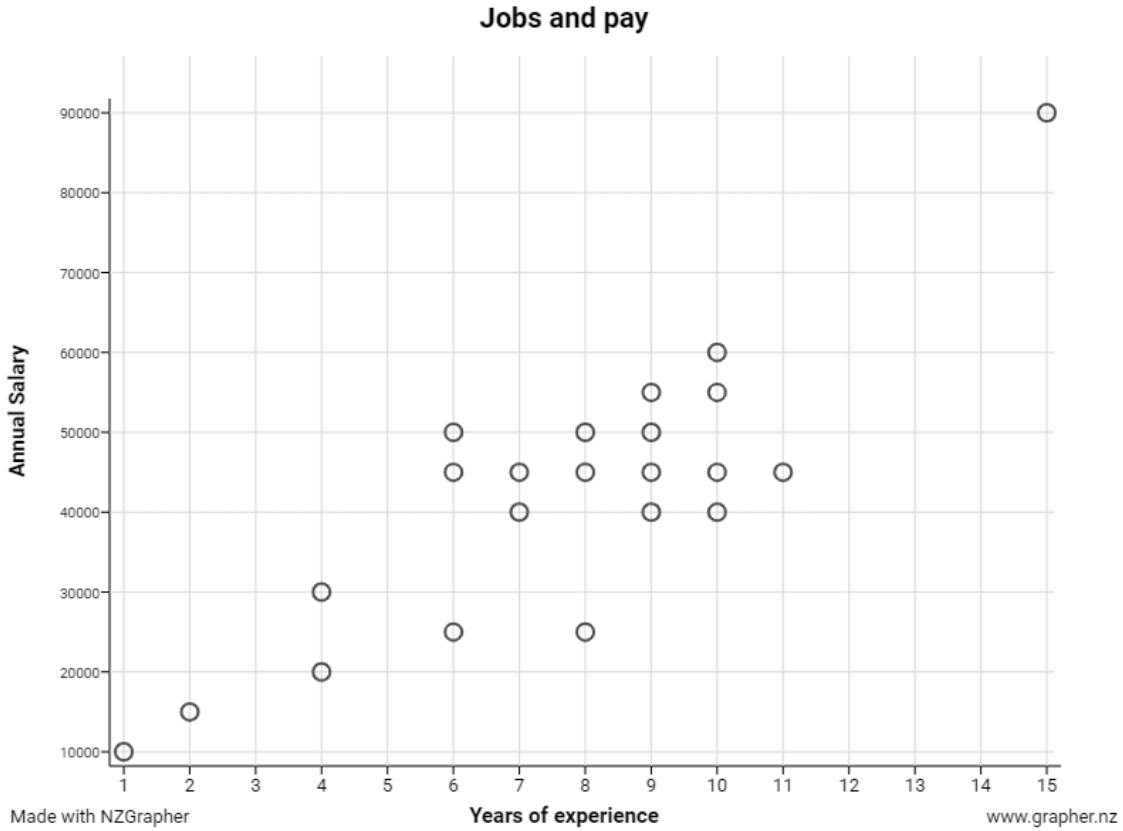

Marathon running times versus length of stride
Marathon completion time (minutes)
380
360
340
320
300
280
260
240
220
200
180
160
140
100
110
120
130
140
150
160
170
180
190
200
Stride length (cm)
Made with NZGrapher
www.mathsnz.com

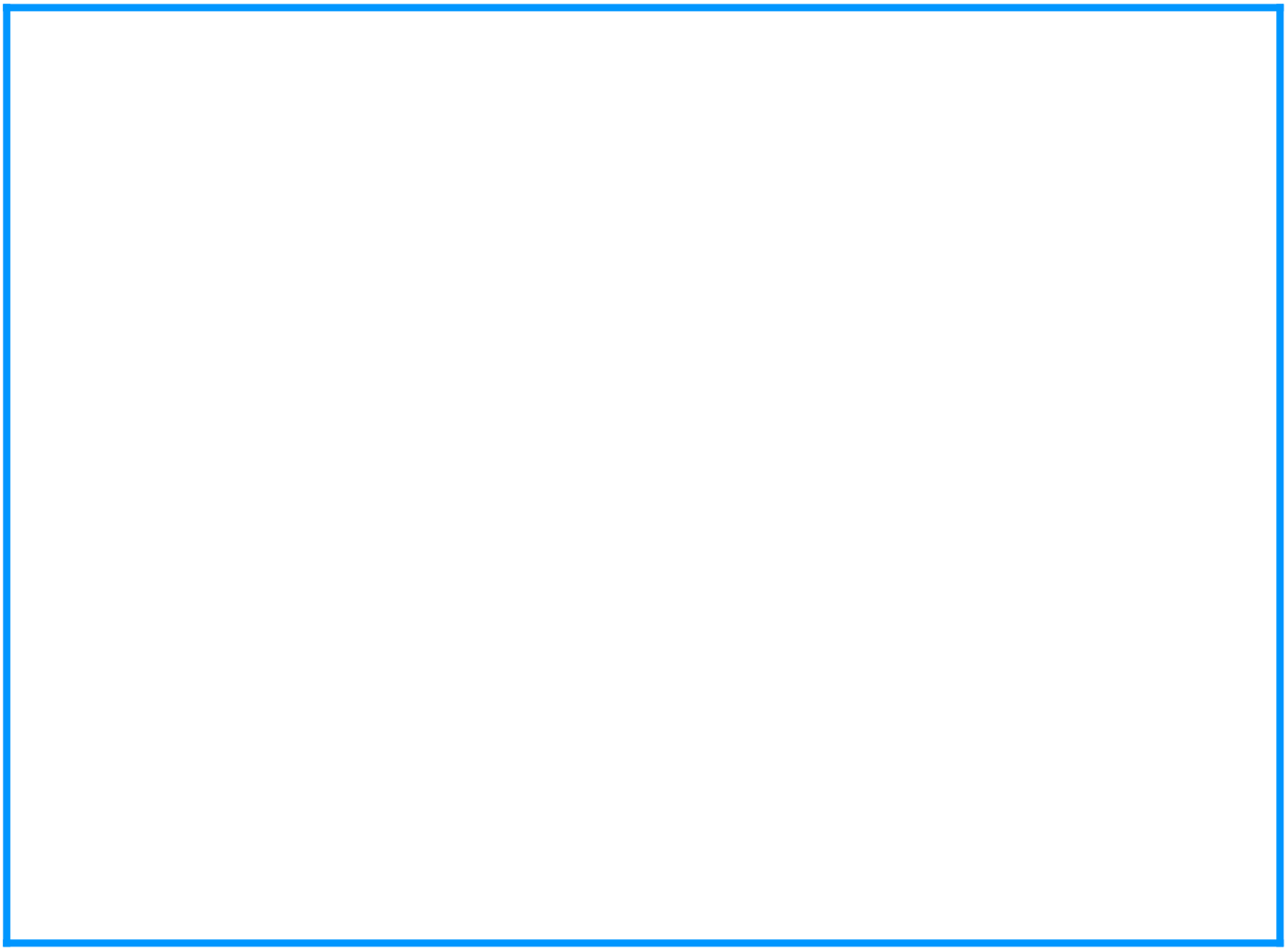


6)

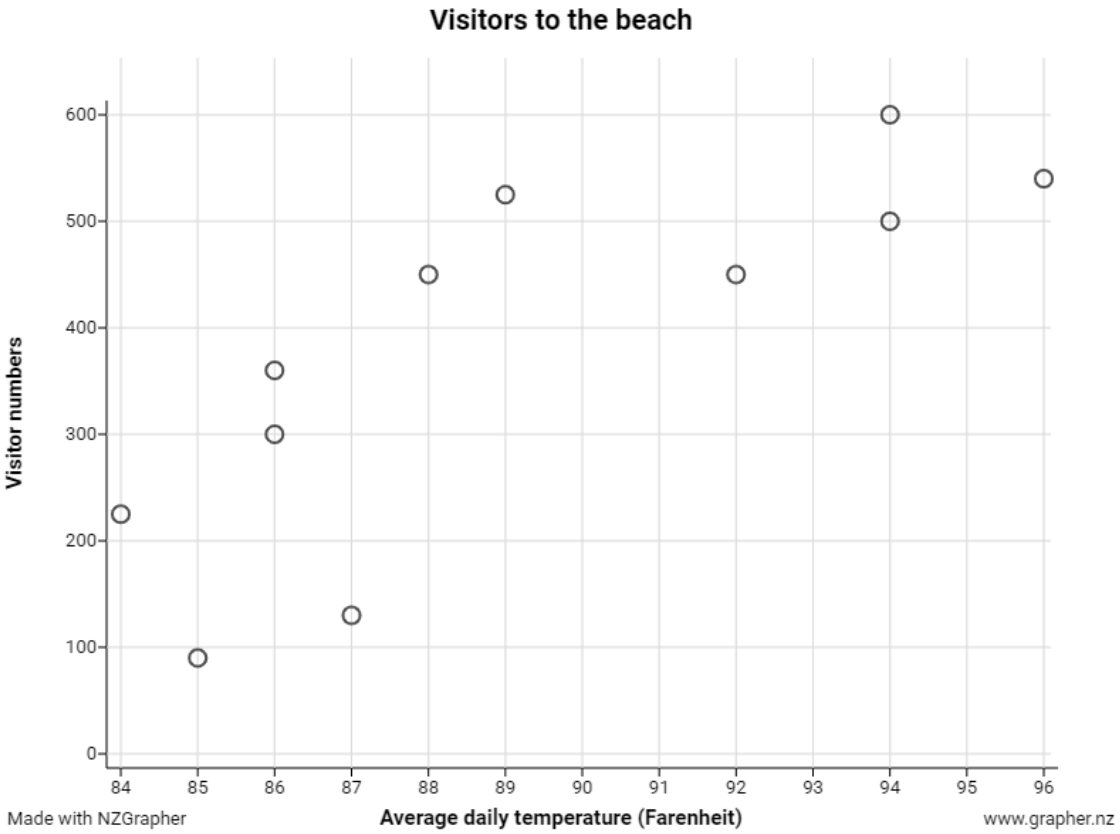

Students attendance and exams
Exam score (%)
100
90
80
70
60
50
40
30
20
10
0
2
4
6
8
10
12
14
16
18
20
22
24
26
Number of missed classes
Made with NZGrapher
www.grapher.nz

7)

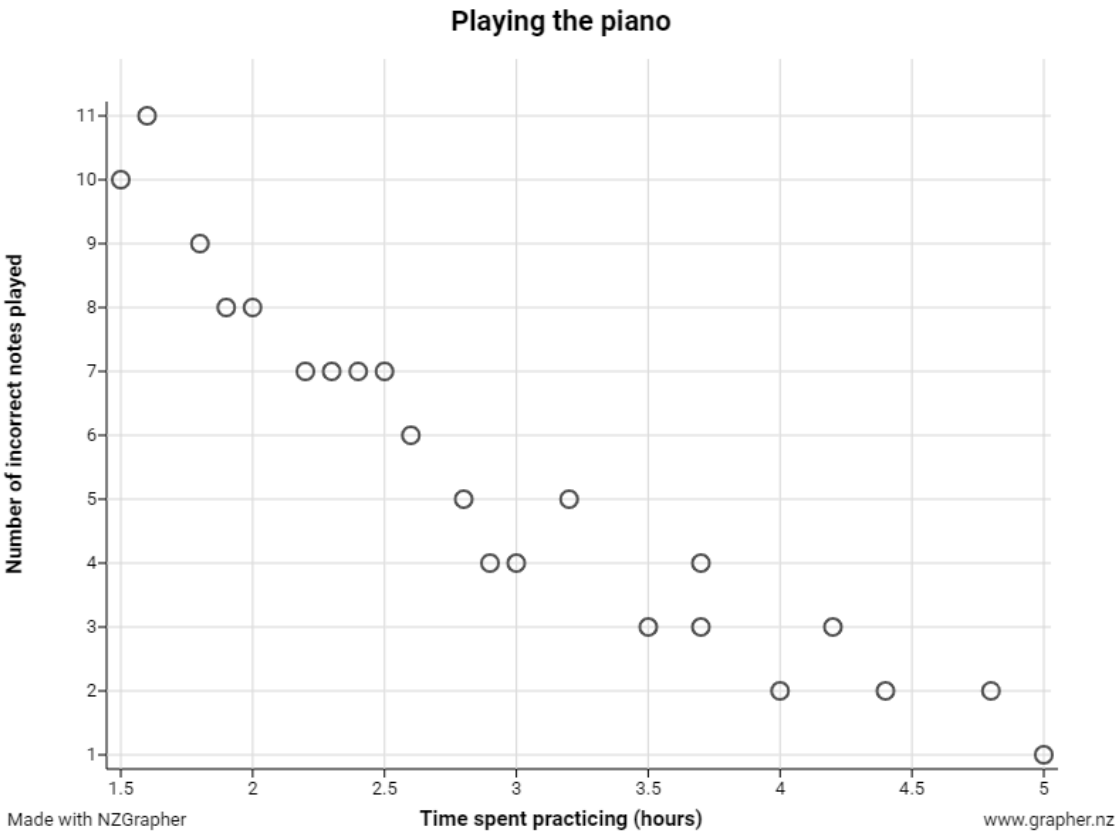

Jobs and pay
Annual Salary
90000
80000
70000
60000
50000
40000
30000
20000
10000
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Years of experience
Made with NZGrapher
www.grapher.nz



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
Average daily temperature (Farenheit)
Made with NZGrapher
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

Insight for Excellence

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 students 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:

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?