Year 9 Statistics (NZC: S4.1, S5.1)

**9.10 Investigate a given data set using the statistical enquiry cycle**

Written by Jake Wills – MathsNZ – [jwills@mathsnz.com](mailto:jwills@mathsnz.com)

|  |  |  |
| --- | --- | --- |
| **Achievement** | **Achievement with Merit** | **Achievement with Excellence** |
| Investigate a given data set using the statistical enquiry cycle. | Investigate a given data set using the statistical enquiry cycle, with justification. | Investigate a given data set using the statistical enquiry cycle, with statistical insight. |

[Introduction to the PPDAC Cycle 2](#_Toc475864528)

[PPDAC Cycle 1 – Eye Colour 3](#_Toc475864529)

[Writing a good paragraph 4](#_Toc475864530)

[Achieved, Merit and Excellence Paragraphs 5](#_Toc475864531)

[Some good sentence starters 6](#_Toc475864532)

[Extra questions to be thinking about at each stage of PPDAC 6](#_Toc475864533)

[Terminology 7](#_Toc475864534)

[Identifying Investigation Type 8](#_Toc475864535)

[Census at School 9](#_Toc475864536)

[PPDAC Cycle 2 – Height and Arm Span 10](#_Toc475864537)

[Extra Bivariate Investigation Ideas 12](#_Toc475864538)

[Mean, Median and Mode 14](#_Toc475864539)

[Making a Call 15](#_Toc475864540)

[PPDAC Cycle 3 – Boys’ and Girls’ Heights 16](#_Toc475864541)

[Extra Comparison Investigation Ideas 17](#_Toc475864542)

[PPDAC Cycle 4 – Students Attendance 19](#_Toc475864543)

[Extra Time Series Ideas 21](#_Toc475864544)

[Applying our Findings to Other Situations 23](#_Toc475864545)

[Teaching Plan 25](#_Toc475864546)

[Formatting Data for Time Series on NZGrapher 26](#_Toc475864547)

# Introduction to the PPDAC Cycle

# PPDAC Cycle 1 – Eye Colour

**Problem:**

I wonder if students in our class are more likely to have blue or brown eyes?

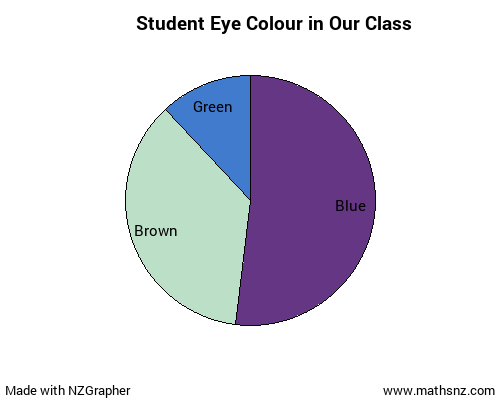
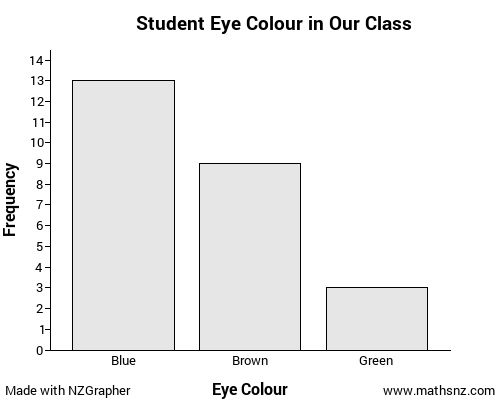
**Plan:**

One student is going to go around the class and record what colour each student’s eyes are in a tally chart.

**Data:**

|  |  |  |
| --- | --- | --- |
| **Blue** | **Brown** | **Green** |
| **~~||||~~ ~~||||~~ |||** | **~~||||~~ ||||** | **|||** |
| **13** | **9** | **3** |

**Analysis:**



More students have blue eyes. We know this as the blue bar is tallest. The blue sector is also biggest in the pie chart.

**Conclusion:**

We found that our class is more likely to have blue eyes (13 out of 25) than to have brown eyes (9 out of 25). There were also 3 students in our class with green eyes.

# Writing a good paragraph

You may have come across an acronym like SEXY, SEED, PEC, OSEM or many more for writing a good paragraph, possibly in English or Social Studies, and the same idea applies here… in this example we will be using the SEXY example (but all the others work fine too).

The paragraphs should follow the format:

**S** - Statement/Topic Sentence: WHAT are you trying to argue?

**EX** - Explanation / Example: HOW does your evidence support your argument?

WHAT evidence is there? (the E and X can be swapped)

**Y** - Why is this important, why do we care? This is also a chance to think about the context.

It is a good idea to think through the whole paragraph and what you want to say before you start writing it. This will help you form much better paragraphs.

**Teachers Note:** we have used SEXY here as that is what our English department uses for their essay structures. There are lots of other acronyms that you can use like OSEM, PEC, PEEL, PEAL, MEAL etc. but we went with this so the kids get used to the same format across the school… why re-invent the wheel when we don’t need to.

# Achieved, Merit and Excellence Paragraphs

Achieved is making a point and explain how it links to what is in the graph.

Merit is either making a good point and linking it to the graph, or backing up with what it means.

Excellence is making a good point and linking it to the graph and the wider context.

An **Achieved** paragraph might look like:

From my graphs, I can see more students in our class have blue eyes as the blue bar is the tallest.

A **Merit** paragraph might look like:

From my graphs, I can see more students in our class have blue eyes. I know this as the blue bar is tallest on the bar graph. The blue sector is also biggest in the pie chart.

An **Excellence** paragraph might look like:

From my graphs, I can see more students in our class have blue eyes. I know this as the blue bar is tallest on the bar graph. The blue sector is also biggest in the pie chart. This would mean if a student came up to me from my class they would be more likely to have blue eyes than any other colour.

# Some good sentence starters

|  |  |
| --- | --- |
| **S – Statement / Topic Sentence**  From my graph…  We can see that…  I can say…  I notice… | **E – Explanation**  Therefore…  As a result…  Because…  Alternatively… |
| **X – Example**  I know this because…  Firstly… Secondly…  However…  A clear example of this is… | **Y – Why?**  We can therefore say…  This would mean…  This matters because… |

# Extra questions to be thinking about at each stage of PPDAC

**Problem**

* Why is this a problem?
* Who cares about this problem?

**Plan**

* Where did the data come from and how was it collected (if I’m not collecting it myself)?
* What other variables might I need to be thinking about?

**Data**

* Were there any issues collecting the data?
* Did I get a good sample?

**Analysis**

* What is going on in the data?
* What might be causing this?

**Conclusion**

* Are there any alternative explanations?
* What does this mean in real life?
* Are there any other questions I should now be asking?

# Terminology

**Sample** - A group of objects, individuals, or values selected from a population. The idea is for this sample to provide estimates for the population.

**Sample Size** - The number of objects, individuals, or values in a sample.

**Survey** - A systematic collection of data taken by questioning a sample of people.

**Census** - A study that attempts to measure every unit in a population.

**Experiment** - A process or study that results in the collection of data, the outcome of which is unknown.

**Measurement (Continuous Numerical) Data** - Data that can take any value in an interval of numbers. Eg: Height, Weight.

**Counting (Discrete Numerical) Data** - Data that can take only distinct values, usually whole numbers. Eg: number of left handed people in the class.

**Group or Category (Categorical) Data** - Values can be organised into groups. These groups (or categories) must be chosen so that they do not overlap and that every value belongs to one and only one group, and there should be no doubt as to which one. Eg: Eye colour

**Time Series Data** - A data set gathered over time. For one object, such as climate in Auckland, the values of a variable (or several variables) are obtained over time. Usually there are equal intervals between the times. Eg: Temperature, attendance in class.

**Symmetry** – When numerical data is distributed the same way either side of the centre.

**Skew** - A lack of symmetry in a distribution of a numerical distribution in which the values are stretched out in one direction.

**Average** - A number that is representative or typical of the centre of a set of numerical values. In this sense, the number used could be the mean or the median or mode.

**Mean** - Calculated by adding all the values together and dividing by the number of values.

**Median** - The middle number when a set of data is ordered numerically.

**Mode** - A value in a set of numerical data that occurs most often.

**Lower Quartile** - A number that is a quarter of the way through the data when it is ordered, from the lower end.

**Upper Quartile** - A number that is a quarter of the way through the data when it is ordered, from the upper end.

**Minimum** - The smallest number when a set of data is ordered numerically.

**Maximum** - The largest number when a set of data is ordered numerically.

**Outlier** - A data point whose values for the variable we are looking at are much bigger or smaller than most of the other data points. This can be either on a dot plot or a scatter graph

# Identifying Investigation Type

Summary (Univariate data) – will often lead to bar graphs or pie charts

Relationship (Bivariate data) – will often lead to scatter graphs

Comparison (Multivariate data) – will often lead to dot plots and box and whisker graphs

Time series (data collected over time) – is looking at data over time

# Census at School

These lessons use a dataset from Census at school… the best way to understand where this data has come from is to take part in the census at school project. Taking part is free and helps create an awesome resource for schools around New Zealand and around the world to use.

Once a class completes the census at school survey the teacher is also given a link to download their class’ dataset, so that you can use it to produce your own graphs and investigation.

Full details on taking part in the census at school project are here:

<http://new.censusatschool.org.nz/take-part/>

# PPDAC Cycle 2 – Height and Arm Span

**Problem**

I wonder if there is a relationship between the height and arm span of students in my class

**Plan**

We are going to record the height of students in my class by getting them to take their shoes off, get them to stand against a wall with a tape measure on it and have someone else put a book on their head and record their height in centimetres.

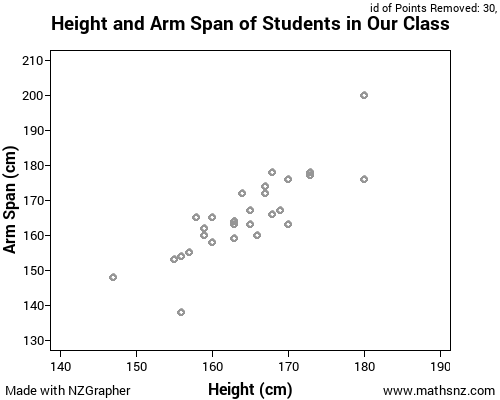
We are going to record the arm span of students in my class by standing facing a whiteboard and raising both arms until they are stretched out as far as we can, one of them touching the left of the board and getting a partner to read off how far I am reaching on the measuring tape.

I will record this data in a table.

**Data**

|  |  |
| --- | --- |
| **Height (cm)** | **Arm span (cm)** |
| 159 | 160 |
| 163 | 159 |
| 163 | 163 |
| 173 | 177 |
| 165 | 167 |
| 155 | 153 |
| 170 | 163 |
| 167 | 172 |
| 156 | 138 |
| 169 | 167 |
| 166 | 160 |
| 163 | 164 |
| 170 | 176 |
| 167 | 174 |
| 168 | 166 |
| 156 | 154 |
| 165 | 163 |
| 160 | 158 |
| 180 | 200 |
| 163 | 164 |
| 160 | 165 |
| 168 | 178 |
| 173 | 178 |
| 164 | 172 |
| 157 | 155 |
| 158 | 165 |
| 159 | 162 |
| 147 | 148 |
| 180 | 176 |
| 168 | - |

**Analysis**



I notice there was one student who did not record their arm span. As a result, we only have 29 points on our graph, not 30.

From my graph, I can see there is a positive trend. I know this because the points seem get higher as they go further right and the ellipse slopes upwards. Therefore, I can say as height increases, so does the arm span.

We can also see that the trend is also reasonably strong. I know this because the ellipse is quite skinny.

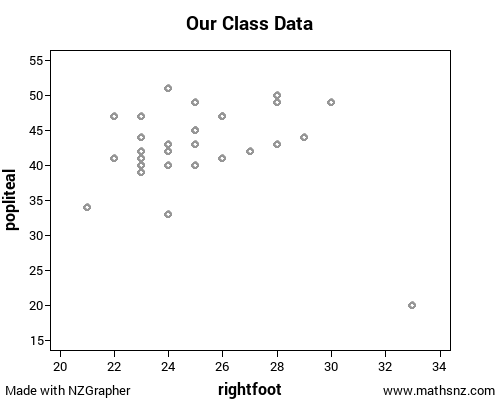
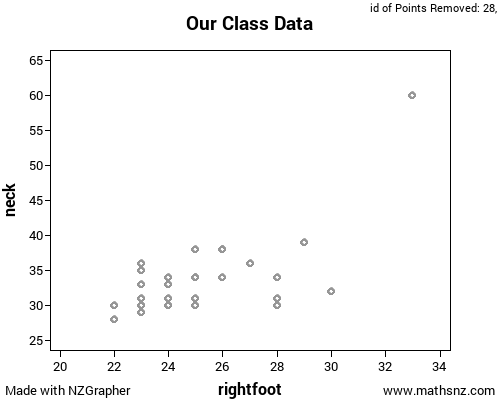
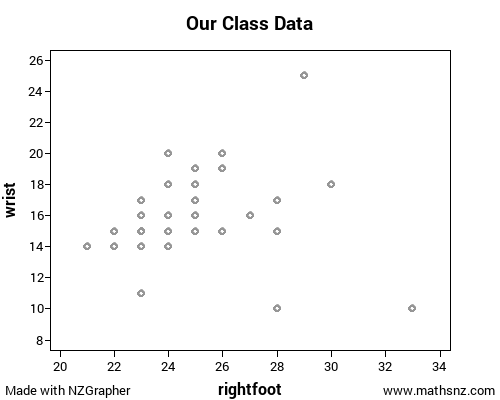
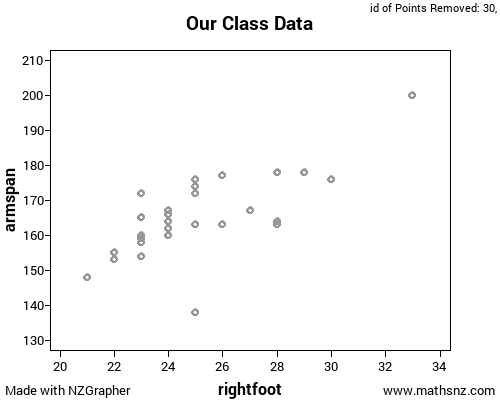
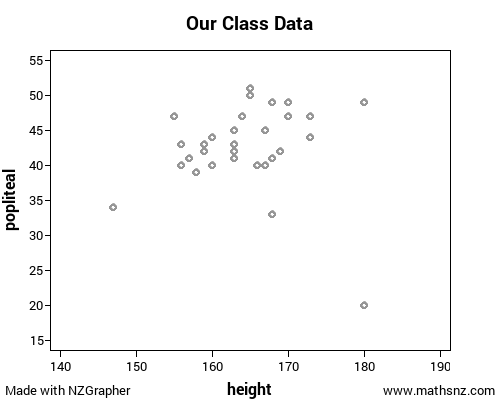
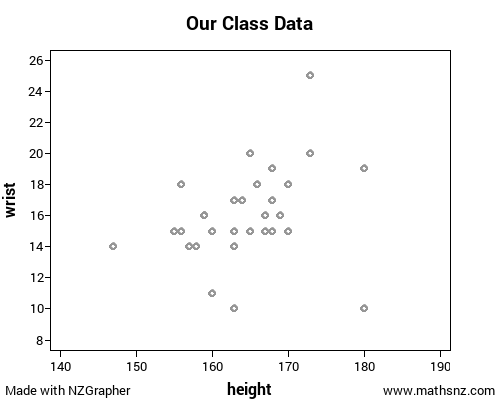
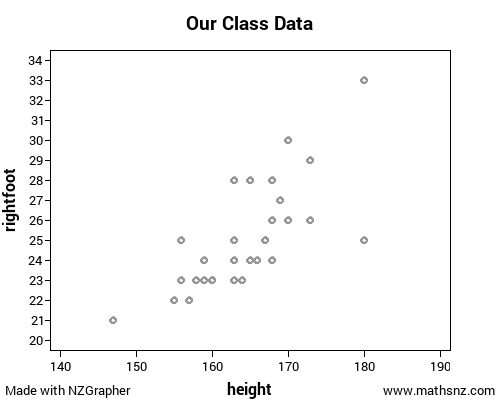
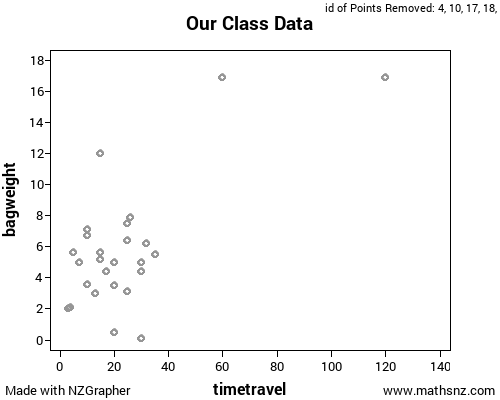
I can also see there is one student who has a much longer arm span than the others. They have a height of 180cm and an arm span of 200cm. This may be because they didn’t measure it properly.

**Conclusion**

This means if a student in our class is tall I would expect them to have a big arm span, and if they are short I would expect them to have a small arm span.

# Extra Bivariate Investigation Ideas

Below are some ideas of some bivariate relationships that you could investigate if you have finished the previous write up. You should write one of these up on your own doc and then share it with your teacher.



# Mean, Median and Mode

Let’s look at these 5 numbers: 2, 2, 2, 3, 6

To find the **mean** we add them together and divide by how many there are.

2 + 2 + 2 + 3 + 6 = 15

15 ÷5 = 3

So the **mean** is 3.

To find the **median** we put them in order and find the middle number

~~2~~, ~~2~~, 2, ~~3~~, ~~6~~

So the **median** is 2

The **mode** is the number that occurs most often.

In this case the **mode** is 2.

Let’s look at these numbers instead: 2, 3, 4, 4, 4, 5, 6, 7, 7, 8

Again, to find the **mean**, we add them together and divide by how many there are.

2 + 3 + 4 + 4 + 4 + 5 + 6 + 7 + 7 + 8 = 50

50 ÷ 10 = 5

So the **mean** is 5.

To find the **median** we put them in order and find the middle number

~~2, 3, 4, 4~~, 4, 5, ~~6, 7, 7, 8~~

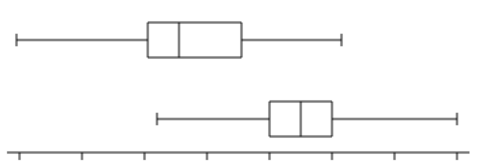
This time there are 2 numbers in the middle, so we the **median** is half way between them, so 4.5

The **mode** is the most common number, so in this case 4.

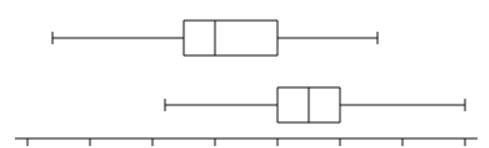
# Making a Call

The first thing you should check is “do the boxes overlap?” The boxes are the square bits in the middle of the graph that go from the lower quartile to the upper quartile.

If they **don’t overlap**, like this, then you can say in the population there is a difference.

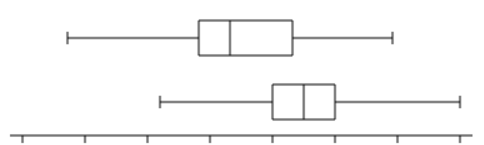


If they are just touching, like this, then you can also say in the population there is a difference, as they still **don’t overlap**.

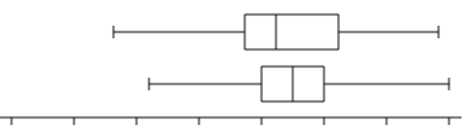


If they **do overlap** the next thing we are looking for is **medians outside the other box**, like this, the median for the top box is below the left of the bottom box, and the median of the bottom box is higher than the right of the top box. If this is the case we can still say there is a difference in the population.

*(Teachers Note: this should be used only if there is between 20 and 40 in each group)*



**Otherwise**, we cannot say there is a difference in the population. This doesn’t mean they are the same, it just means we don’t have enough information to make a call.



# PPDAC Cycle 3 – Boys’ and Girls’ Heights

*Teachers note: you could do this with students aged 12/13/14 instead of boys / girls if that works better in your teaching environment.*

**Problem**

I wonder if Year 9 boys or Year 9 girls tend to be taller? I am going to use the students in our class to try and work this out.

**Plan**

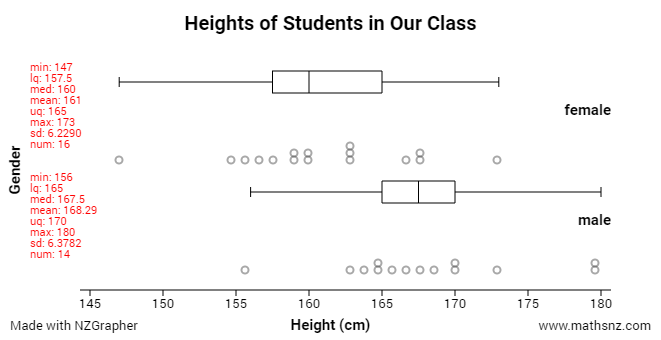
We are going to record the height of students in my class by getting them to take their shoes off, get them to stand against a wall with a tape measure on it and have someone else put a book on their head and record their height in centimetres.

**Data**

Boys Heights (cm): 165, 170, 156, 169, 166, 163, 170, 167, 165, 180, 168, 173, 164, 180

Girls Heights (cm): 159, 163, 163, 173, 155, 167, 168, 156, 160, 163, 160, 157, 158, 159, 147, 168

**Analysis**



I notice the males’ median height is 167.5cm whereas the females’ median height is only 160cm. This means the males median height is 7.5cm bigger than the females median height.

I notice the females’ heights are more spread out than the males’ heights. I know this because box is much wider for the females than for the males. This means the middle 50% is more spread out for the females than for the males.

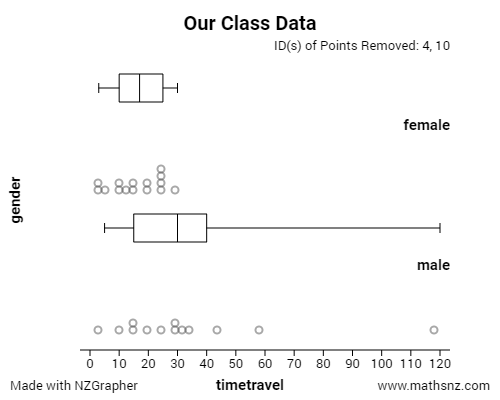
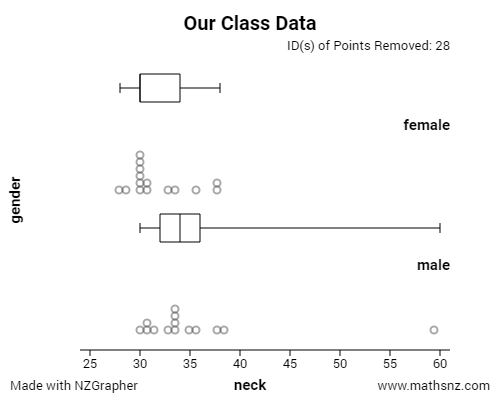
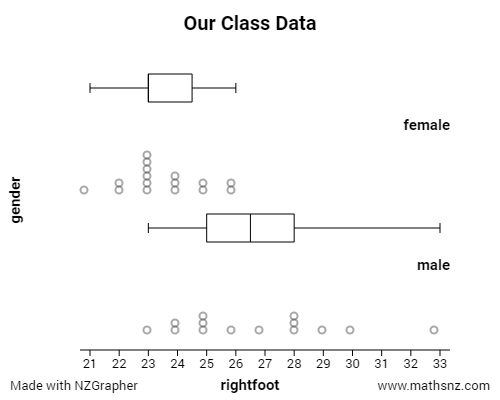
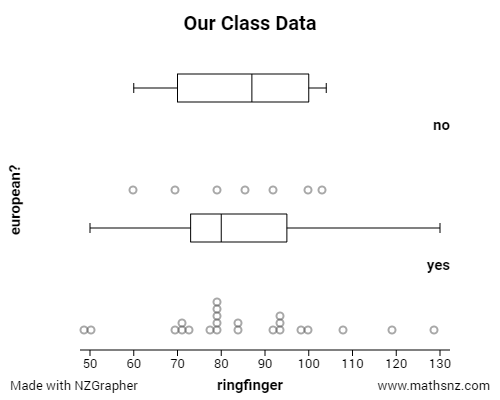
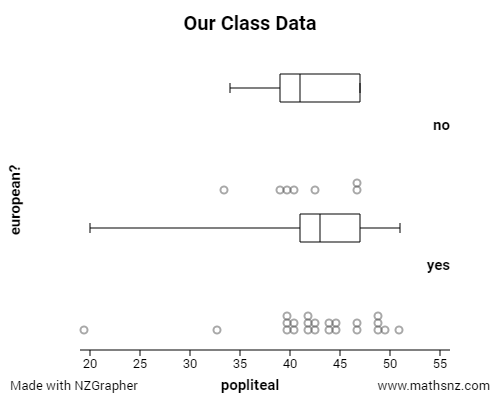
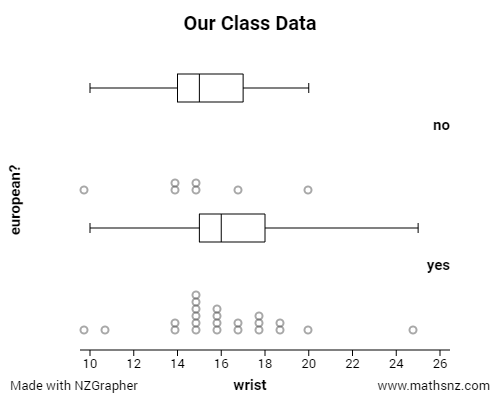
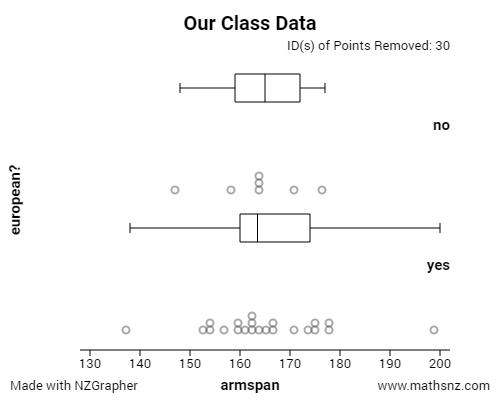
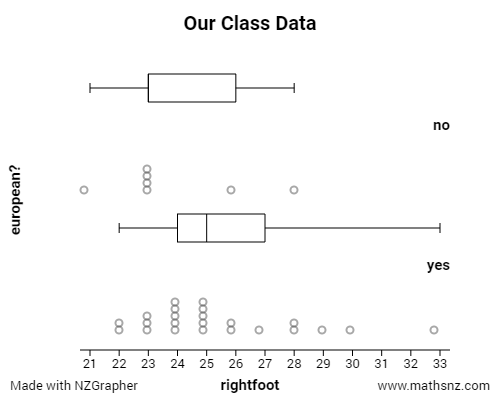
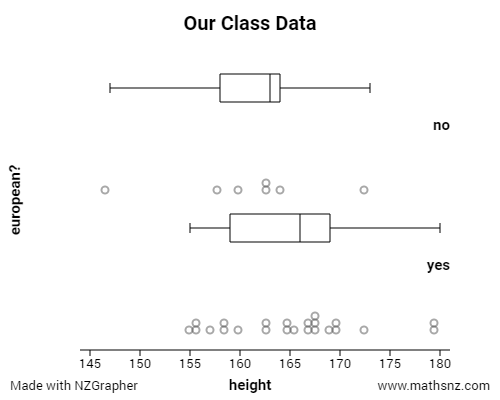
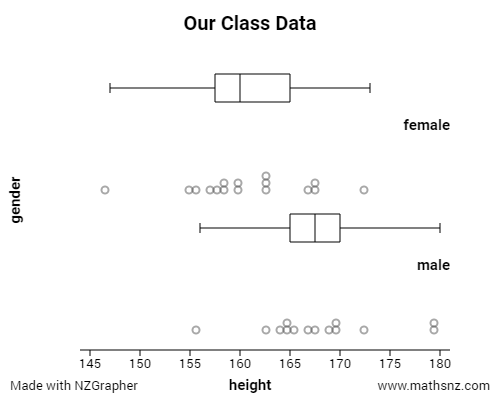
I notice there is one female that is much shorter than the other females. I can see this as there is one point further to the left than the others. Looking at the data, she is only 12 years old, so it may be because she hasn’t grown as much yet.

**Conclusion**

Overall I can say Year 9 males tend to be taller than Year 9 females as the boxes do not overlap.

# Extra Comparison Investigation Ideas

Below are some ideas of some comparisons that you could investigate if you have finished the previous write up. You should write one of these up on your own doc and then share it with your teacher.



# PPDAC Cycle 4 – Students Attendance

**Problem**

I wonder if there is any pattern to the attendance of Year 9 students at our school?

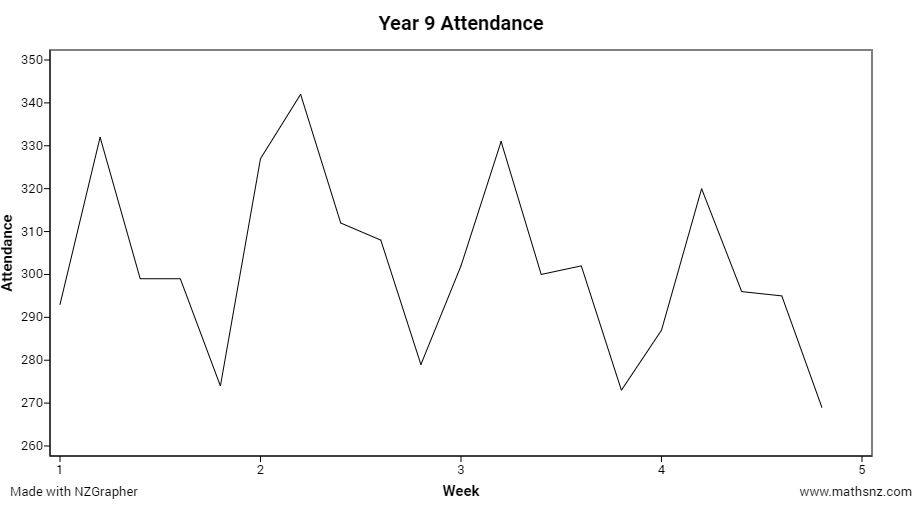
**Plan**

We will collect data for each day, over the first 4 weeks of Term 2. We will record how many Year 9 students were at school each day.

**Data**

|  |  |  |  |
| --- | --- | --- | --- |
| **Week** | **Day** | **NZGrapher Format** | **Attendance** |
| 1 | 1 | 1W1 | 293 |
| 1 | 2 | 1W2 | 332 |
| 1 | 3 | 1W3 | 299 |
| 1 | 4 | 1W4 | 299 |
| 1 | 5 | 1W5 | 274 |
| 2 | 1 | 2W1 | 327 |
| 2 | 2 | 2W2 | 342 |
| 2 | 3 | 2W3 | 312 |
| 2 | 4 | 2W4 | 308 |
| 2 | 5 | 2W5 | 279 |
| 3 | 1 | 3W1 | 302 |
| 3 | 2 | 3W2 | 331 |
| 3 | 3 | 3W3 | 300 |
| 3 | 4 | 3W4 | 302 |
| 3 | 5 | 3W5 | 273 |
| 4 | 1 | 4W1 | 287 |
| 4 | 2 | 4W2 | 320 |
| 4 | 3 | 4W3 | 296 |
| 4 | 4 | 4W4 | 295 |
| 4 | 5 | 4W5 | 269 |

**Analysis**



**Overall Trend**

Overall there appears to be a slight decrease in the number of students attending. I can see this as the highest points in week 2, 3 and 4 keep dropping, as do the lowest points. The highest point in Week 2 is 342 whereas the highest point in Week 4 is only 320. The lowest point in Week 2 is 279, and the lowest point in Week 4 is 269. This means as the term goes on less Year 9 students are attending school.

**Seasonality**

I can see that the highest attendance each week is always on a Tuesday. I know this because the second point in each week is always the furthest up the graph. Tuesdays are normally about 30 students higher than any other day in the week. This means more students are at school on a Tuesday than any other day.

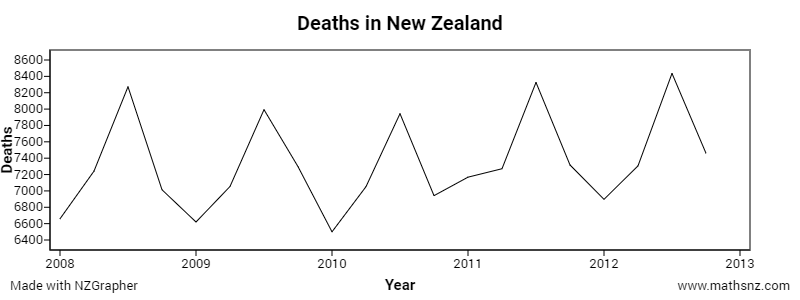
I notice that the lowest attendance each week is always on a Friday. I know this as the last point in each week is always the lowest down on the graph. Fridays are normally about 30 students lower than any other day in the week. This means less students are at school on a Friday than any other day. I think this might be because people are more tired by the end of the week so more likely to be sick, and more likely to go away on a holiday.

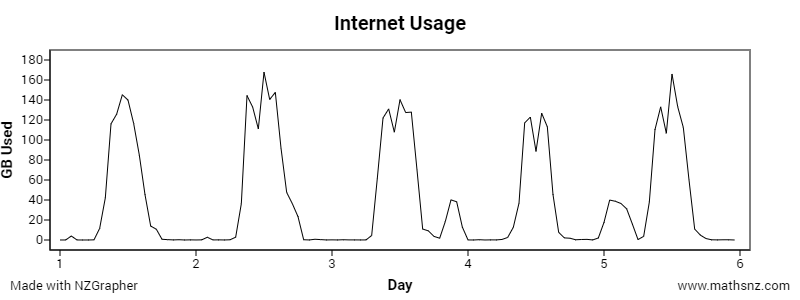
**Conclusion**

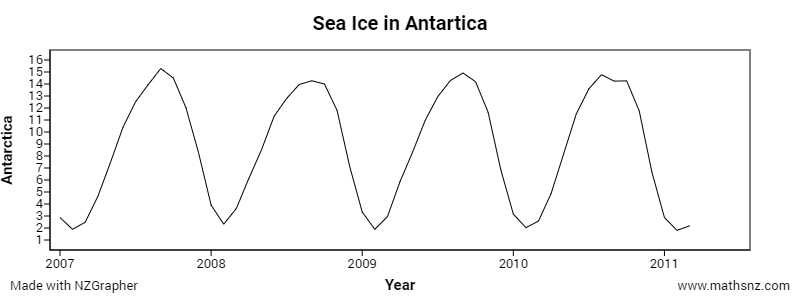
Yes, there is a pattern to the attendance of Year 9 students. Tuesdays is always high and Fridays is always low, and overall there is a slow drop in the number of Year 9 students attending.

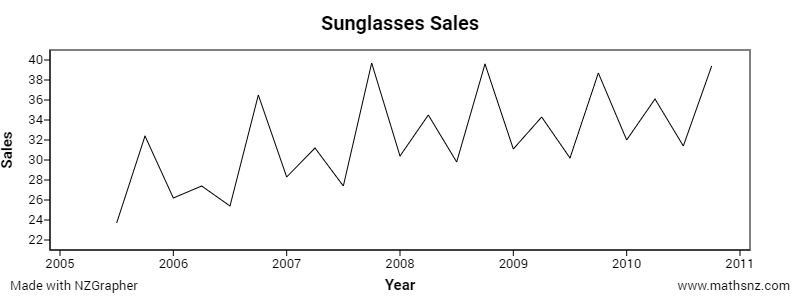
# Extra Time Series Ideas

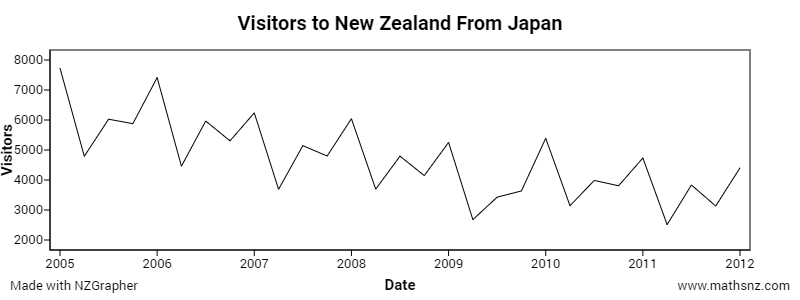
Below are some ideas of some time series that you could investigate if you have finished the previous write up. You should write one of these up on your own doc and then share it with your teacher.









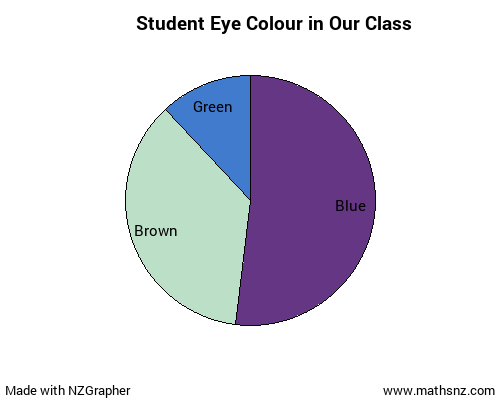
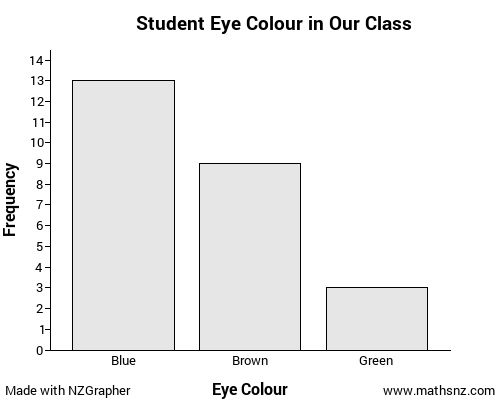


# Applying our Findings to Other Situations

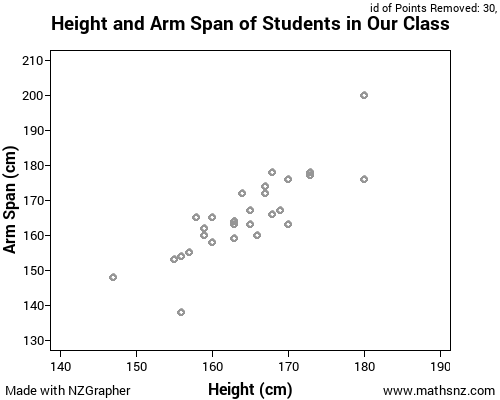
*Teachers Note: This is ideally done as a classroom discussion around this, and then looking at how the findings of the investigations you have done can be applied and what situations are appropriate and not appropriate.*

So far, we have only been looking at analysing our data, but not really making too many links to other situations. One of the big parts of statistics is making links to other situations and thinking about weather what you have found out can apply to other situations.

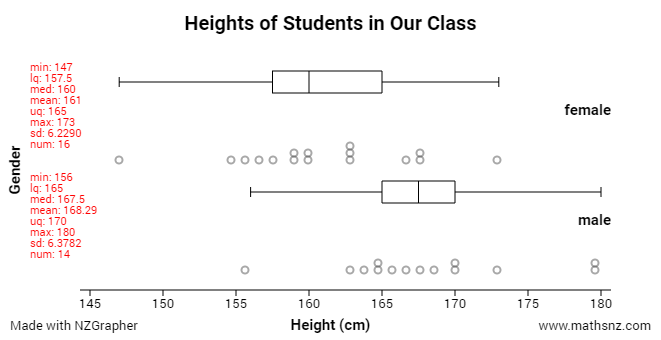
Let’s re-look at the investigations we’ve looked at during these lessons.



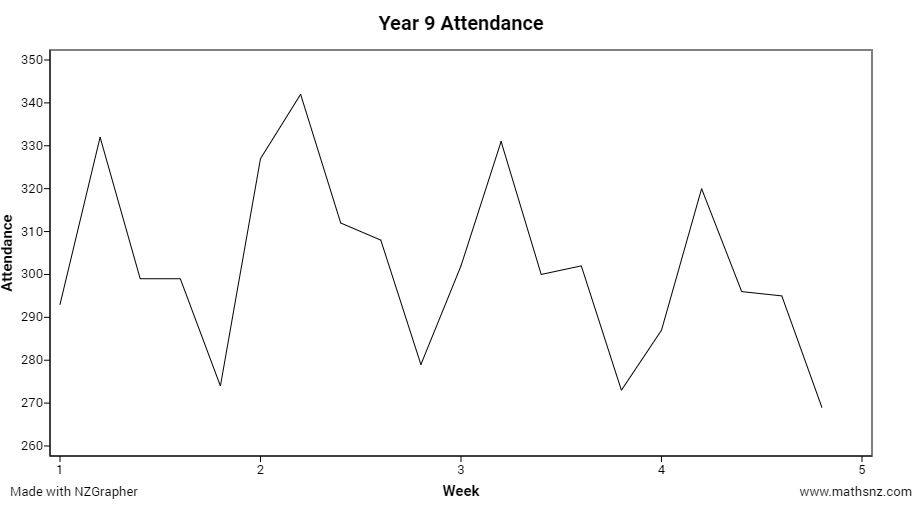
In this investigation, we looked at eye colours of students in our class. We found that more students had blue eyes than brown eyes. This was only for a small number of students in our class though. If we were to look at more people we might find something quite different.



In this investigation, we found that the taller you are the longer your arm span. This is something that is likely to hold true for all people, regardless of their age or ethnicity. So in this case the findings can probably be extended quite well.



In this investigation, we found that the males were taller than the females. This was for a group of Year 9 students, so it probably would be okay to generalise to Year 9 students, but may not be appropriate to extend it to students in other year levels, as males and females grow at different times and different rates.



In this investigation, we found that at this school in Term 2, the attendance was highest on Tuesdays and lowest on Fridays, and was dropping slightly as the term went on. There are several things that could alter this. What happens if there was a public holiday? A school Trip? Do all schools have the same pattern? We need to think about these things before we can apply these findings to the whole year at the school, let alone other schools.

# Teaching Plan

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Week** | **Lesson 1** | **Lesson 2** | **Lesson 3** | **Lesson 4** |
| **1** |  | Welcome and introductions  Intro to PPDAC | PPDAC Cycle 1 – Eye Colour | Good Paragraphs |
| **2** | Waitangi Day | Terminology | Identifying Investigation Type | [Census at School](http://new.censusatschool.org.nz/take-part/) data collection |
| **3** | [Census at School](http://new.censusatschool.org.nz/take-part/) online entry | PPDAC Cycle 2 – problem, plan, and data | PPDAC Cycle 2 – analysis and conclusion | Averages |
| **4** | PPDAC Cycle 3 – example | PPDAC Cycle 3 – extra investigation | PPDAC Cycle 4 – example | PPDAC Cycle 4 – extra investigation |
| **5** | Applying our Findings to Other Situations | Catch-up Day | Class choice investigation – problem, plan, and data | Class choice investigation – analysis and conclusion |
| **6** | Give out assessment | Working on assessment | Working on assessment | Working on assessment |
| **7** | Assessment due in – start next topic |  |  |  |

# Formatting Data for Time Series on NZGrapher

NZGrapher has these 6 formats for time series data:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Format** | **Seasons** | **Example** | **Notes** | **Think of as** |
| Yearly | 1 | 2015, 2016, 2017 | Just numbers by themselves | YEAR |
| Quarterly | 4 | 2015Q1, 2015Q2, 2015Q3, 2015Q4 then 2016Q1, 2016Q2 | must be 1 digit after the Q | YEAR "Q" QUARTER |
| Monthly | 12 | 2015M01 - 2015M12, 2016M01, 2016M02 | must be 2 digits after the M | YEAR "M" MONTH |
| Daily  (7 Days) | 7 | 2001D1 - 2001D7, 2002D1, 2002D2 | must be 1 digit after the D | WEEKNUMBER "D" DAY |
| Daily  (5 Days) | 5 | 2001W1 - 2001W5, 2002W1, 2002W2 | must be 1 digit after the W | WEEKNUMBER "W" DAY |
| Hourly | 24 | 2001H01 - 2001H24, 2002H01, 2002H02 | must be 2 digits after the H | DAYNUMBER "H" HOUR |

There is one more thing to note... the first two graphs in NZGrapher (time series, and recomposition) I have re-written the back end for to make them faster and more versatile, so they aren't quite as picky with the format of the numbers before the letters... they can be any length as long as they are consistent (ie all 1, 2, 3, 4 or 5 digits).

The last two (seasonal effects and forecasts) I haven't yet re-written, so they require 4 digits before the letter.

For Year 9 you probably don’t need the last two so any number of leading digits is fine, I normally try go for 1 or 2 for the daily or hourly data, but makes sense to have 4 for the monthly or quarterly data. For Year 13 it’s important to have the 4 digits.