

Level 3 - AS91580 - 4 Credits - Internal

Investigate Time Series Data

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

Achievement	Achievement with Merit	Achievement with Excellence
Investigate time series data.	Investigate time series data,	Investigate time series data,
	with justification.	with statistical insight.

Part 1: Problem	2
Part 2: Plan	4
Part 2.1: Explain the Variables	4
Part 2.2: Naming the Source	6
Part 3: Data – Using NZGrapher	8
Part 4: Analysis	10
Part 4.1: Long Term Trend	10
Part 4.2: Seasonal Pattern	12
Part 4.3: Outliers	14
Part 4.4: Appropriateness of the Model	16
Part 4.5: Predictions	18
Part 4.6: Reasons for Confidence	20
Part 5: Conclusion	22
Part 6a: Writing Your Own Internal 1	24
Part 6b: Writing Your Own Internal 2	26
Sample Internal (at Achieved Level)	
Data Set Information	
Assessment Guidelines – 91580 – Investigate Time Series Data	32

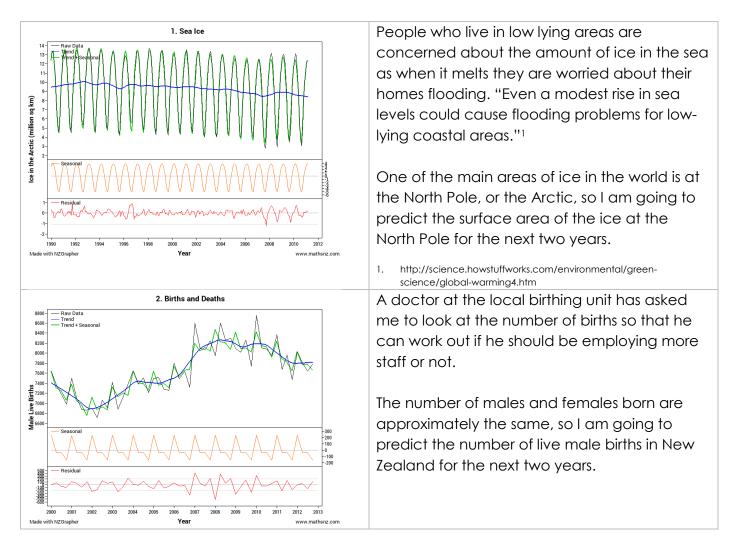


Part 1: Problem

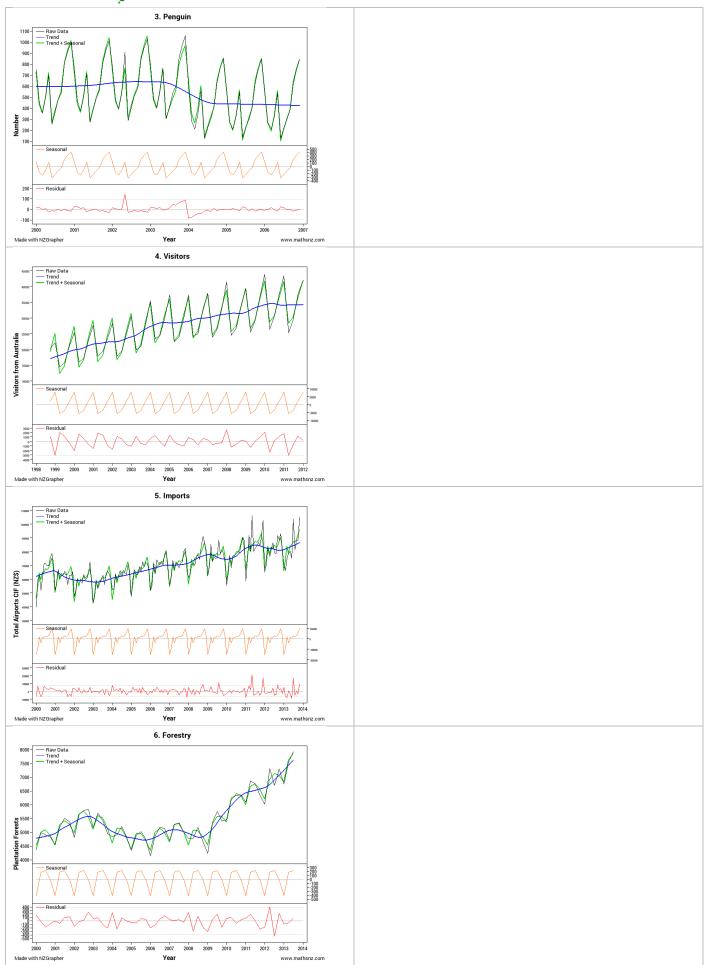
For each of the graphs write a good purpose. A purpose should give a reason as to why you would want to look at this dataset. You should be trying to link this to research. You then need to state what you are going to be predicting, and link this to your purpose.

The first two have been done for you.

Note: NZGrapher will always predict for the next two years





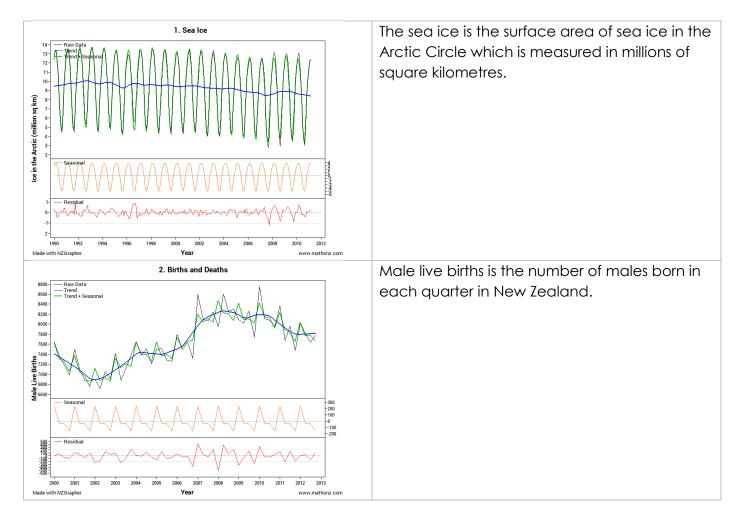




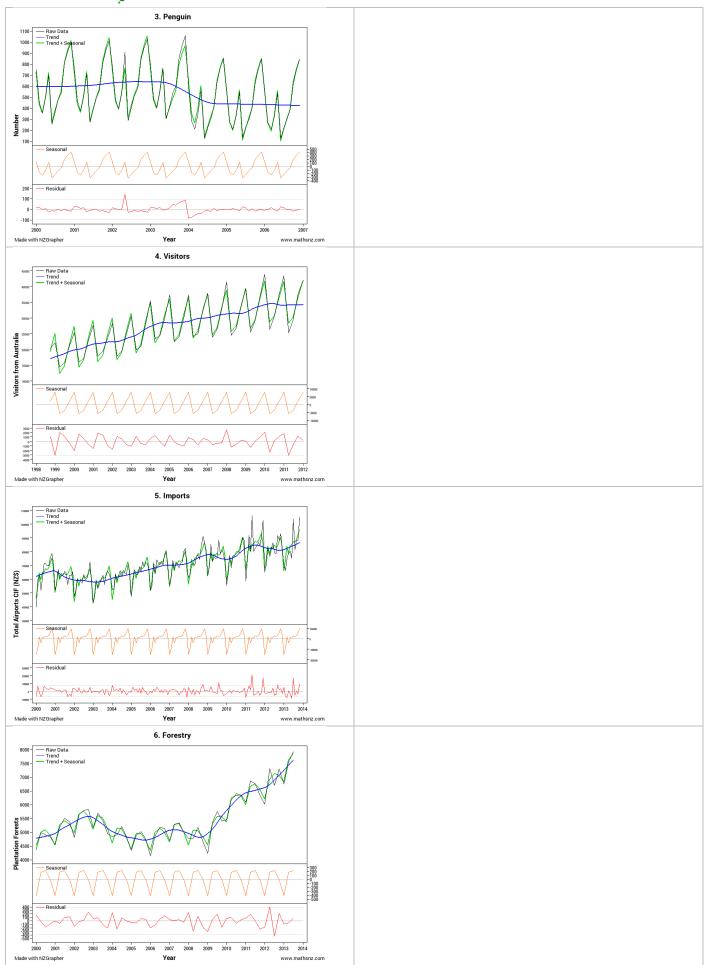
Part 2: Plan Part 2.1: Explain the Variables

The next thing that we need to do is explain our variable(s) and say what units are being used. Identify the variables for each of the graphs.

The first two have been done as examples for you.



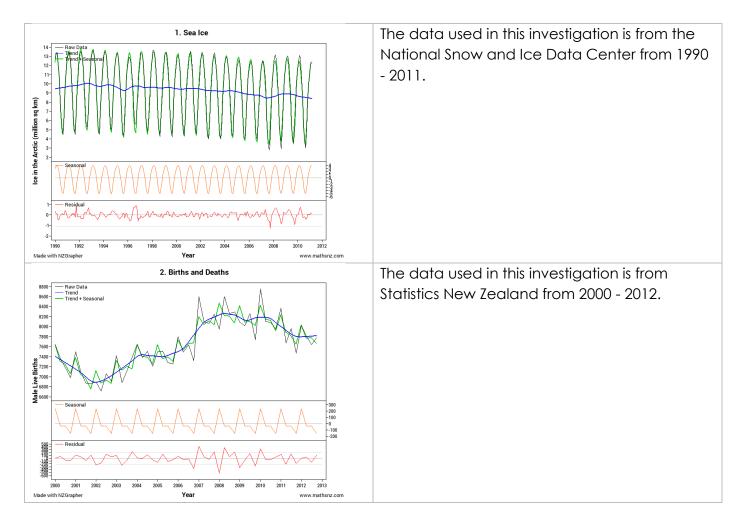




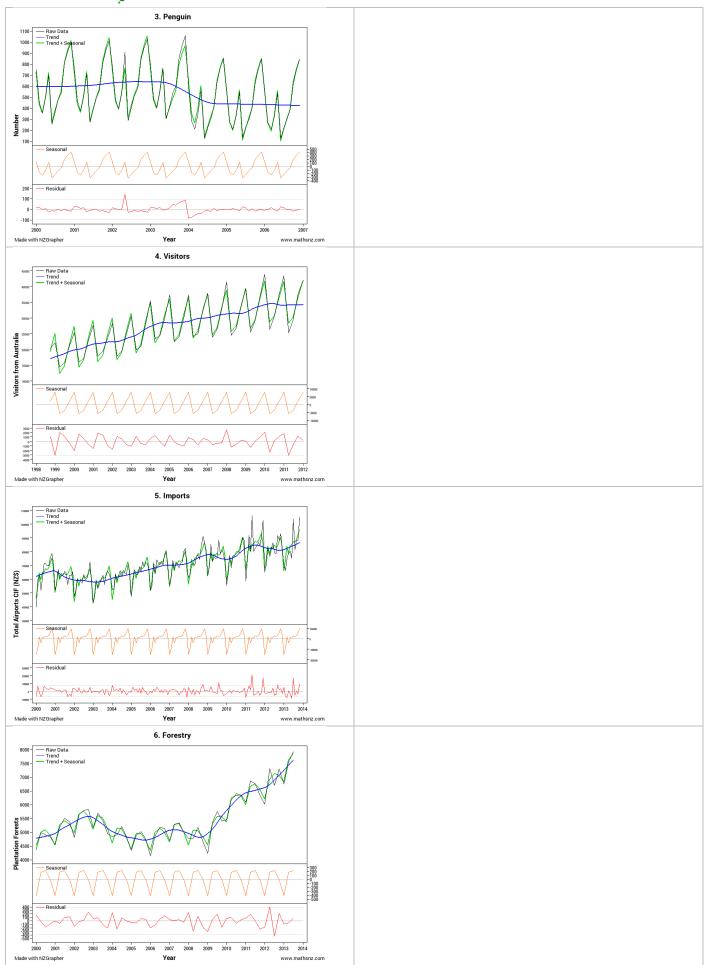


Part 2.2: Naming the Source

In order for our report to have validity, we need to state where the data has come from. Name the source for each of the graphs. The first two have been done as examples for you. Hint: Information of all of the data sources is in the appendix.







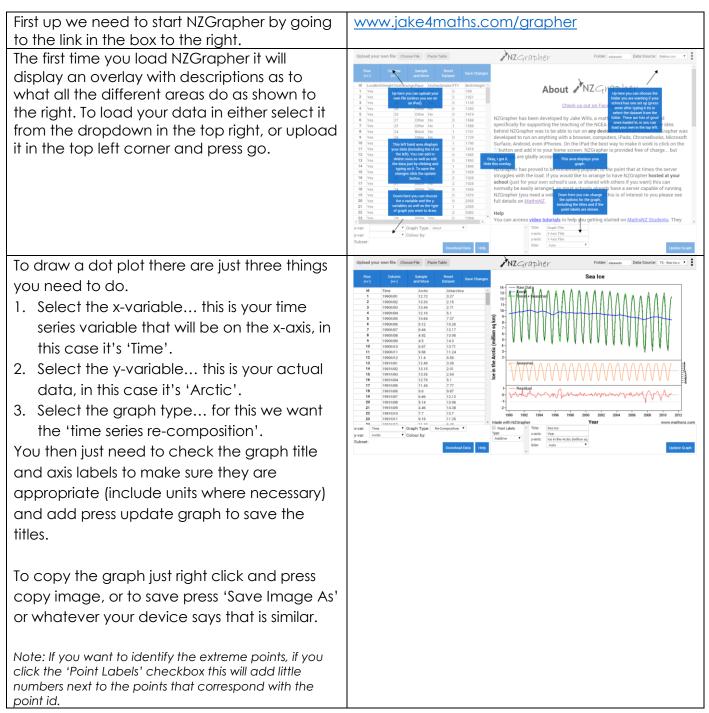


Part 3: Data – Using NZGrapher

The next section that we need to do is the data section. This is reproducing the graphs on Page 2 using NZGrapher. The example below will go through using the Rugby dataset for weight by position.

NZGrapher runs on anything with a browser... Macs, PCs, iPad, Android, ChromeBooks and more.

The example below uses the sea ice dataset.





To get the segrend effect graph just change	Upload your own file: Choose File Paste Table	NZGrapher Folder: datasets Data Source: TS-Sea tocc .
To get the seasonal effect graph just change	Row Column Sample Reset Save Chapper	Sealce
the graph type to 'time series seasonal effects'		Star Data 13 12 12 12 12 12 12 12 12 12 12
Note: you will need to change the x-axis title to reflect if it is monthly or quarterly (or something else).	7 1950407 6.44 12.17 6 1950407 6.2 12.55 4 1950409 4.5 14.3 9 1950409 4.5 14.3 9 1950409 4.5 14.3 10 1950407 1.4 5.6 12 1950417 1.4 5.6 13 1951407 1.2.4 2.9 14 1951407 1.3.5 2.4 15 1951407 1.3.6 2.4 16 1951407 1.4.4 7.7 17 1951407 6.6 1.2.12 19 1951407 6.6 1.2.12 19 1951407 6.6 1.2.12 20 1951407 6.6 1.2.12 21 1951407 6.6 1.2.12 21 1951407 6.6 1.2.12 21 1951407 4.4 1.2.36	More that lates the second sec
The final graph we need is the forecast	Upload your own file: Choose File Paste Table	NZGrapher Folder: datasets Data Source: TS- Sea tocc •
graph.	Row Column Sample Reset Save Changes (+/-) (+/-) and More Dataset Save Changes	Sea Ice
To get the visual output you just change the graph type to 'time series forecasts'	Id Time Actic Amarcica 1 1906001 12.72 3.27 2 1906002 13.33 2.16 3 1906002 13.44 2.71 4 1906003 13.44 2.71 5 1906004 12.16 6.1 6 1906005 0.14 7.27 6 1906006 0.12 10.26 7 1906007 6.44 12.17 8 1906006 4.5 14.3	10
You will also need the text output of the predictions which you can get by clicking on the tick box below the graph.	10 1500110 8.07 13.71 11 1500112 1.8 13.21 12 1500112 1.4 6.6 13 101101 2.40 3.59 14 101102 1.16 5.01 15 101101 1.16 2.01 15 101102 1.15 2.01 16 101102 1.15 2.01 17 101100 1.15 2.44 19 101100 0.60 9.77 19 101100 6.66 1.12.2 20 101100 6.66 1.12.2 21 101100 6.66 1.12.2 21 101100 6.46 1.13.2 21 101100 6.66 1.12.2 21 101100 1.14 1.2 22 101100 7.13.2 1.14 24 101100 7.13.2 1.16 24 101100 7.13.2 1.16	B P 0 P

Now it's your turn. For each dataset produce all the outputs shown above.

Teachers: for tips and tricks on how to format data for time series click on the time series link on: <u>http://www.mathsnz.com/inzight-tips/</u>



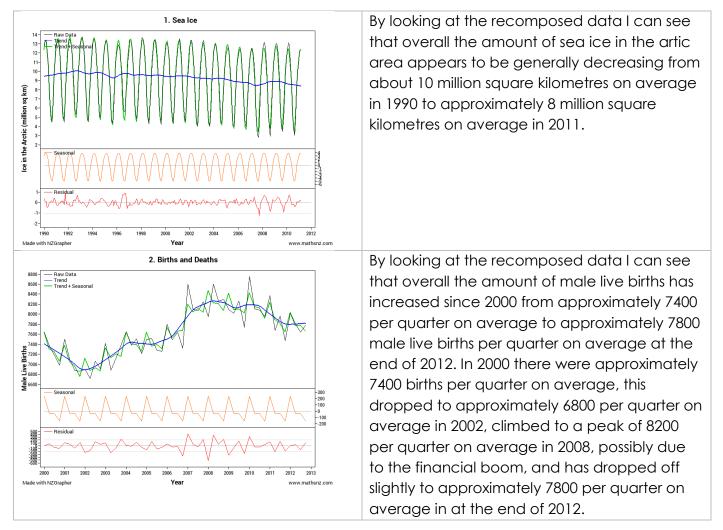
Part 4: Analysis

We now start on the analysis section of our report. In all of these sections you should be attempting to explain the cause of what you can see, ideally with references to back up your statements.

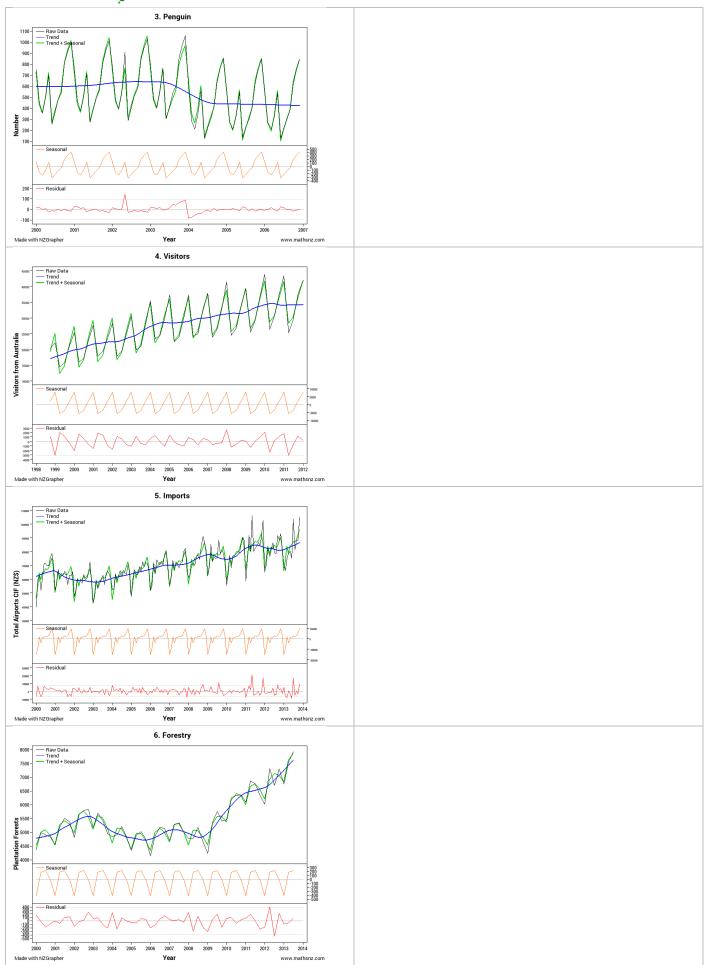
Part 4.1: Long Term Trend

For the long term trend we want to be discussing the overall trend. This is what is happening to the blue line on our recomposed data graphs. Is it increasing, decreasing or doing something else? It is good to give numbers for different dates to back up your comments. For higher grades you should be giving possible reasons and link to your purpose. You should always start by stating the start and finish values, and then elaborate on what is happening in between. You could also think about how much of the variation in the raw data is due to the change in this trend.

For each of the graphs write a comment about the long term trend. The first two have been done for you.





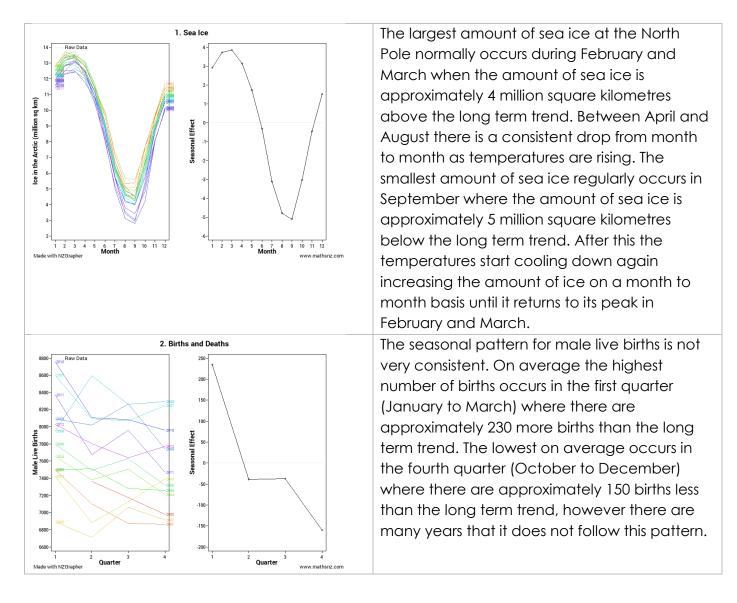




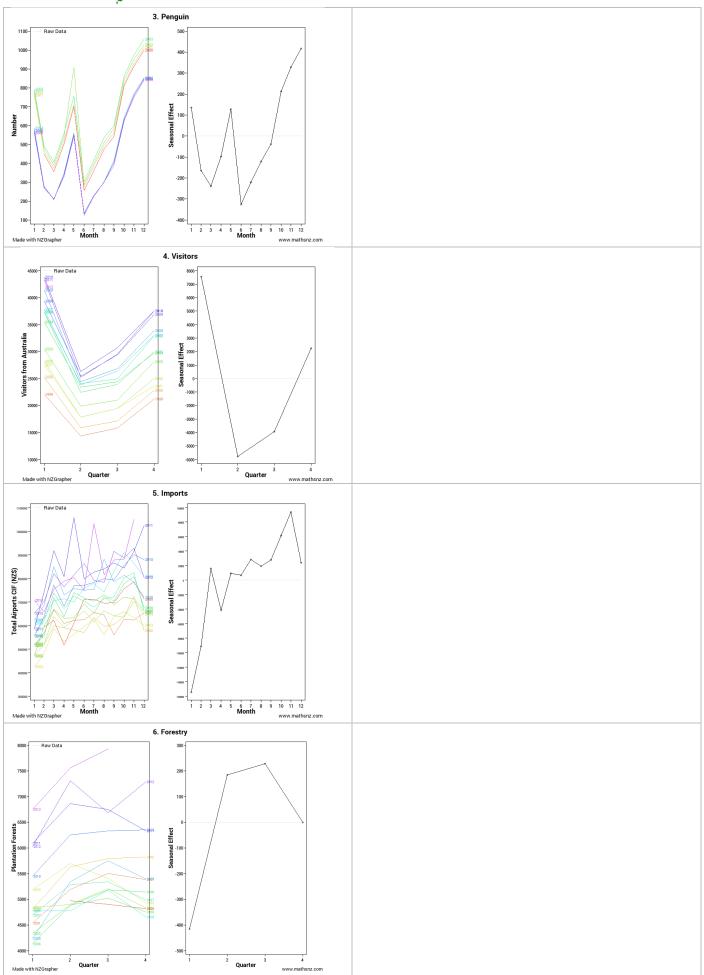
Part 4.2: Seasonal Pattern

With the seasonal pattern we are looking at what is happening in each month or each quarter. You want to be commenting on any highs or lows and also the change between them. For higher grades you should be giving possible reasons and link to your purpose and also think about the relative size of the season pattern to the overall variation.

For each of the graphs write a comment about the long term trend. The first two have been done for you.









Part 4.3: Unusual Points

The next step involves a little bit of mathematical calculations. To work out if a point is an outlier tend to use the rule of thumb that the residual is more than 10% of the overall spread. NZGrapher automatically draws these in as light grey lines for you in the residuals section. To easily identify points if you tick the 'Point Labels' button it puts the id of the row next to each point.

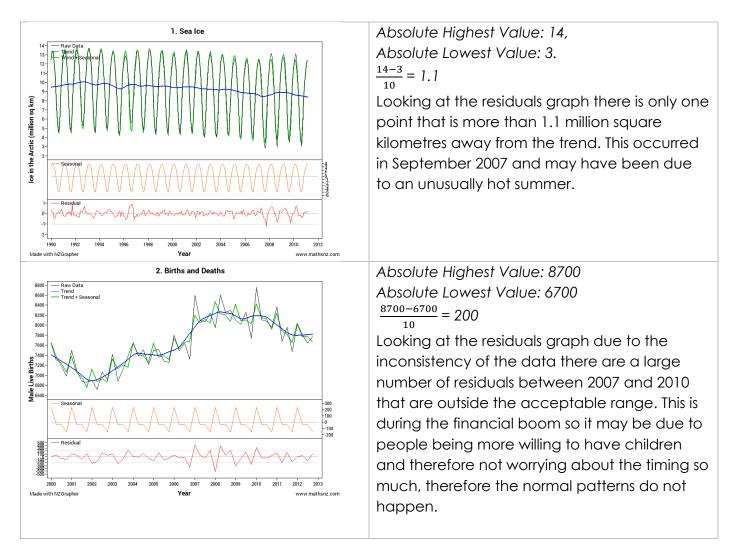
To calculate this we use the following calculation:

Absolute Highest Value – Absolute Lowest Value

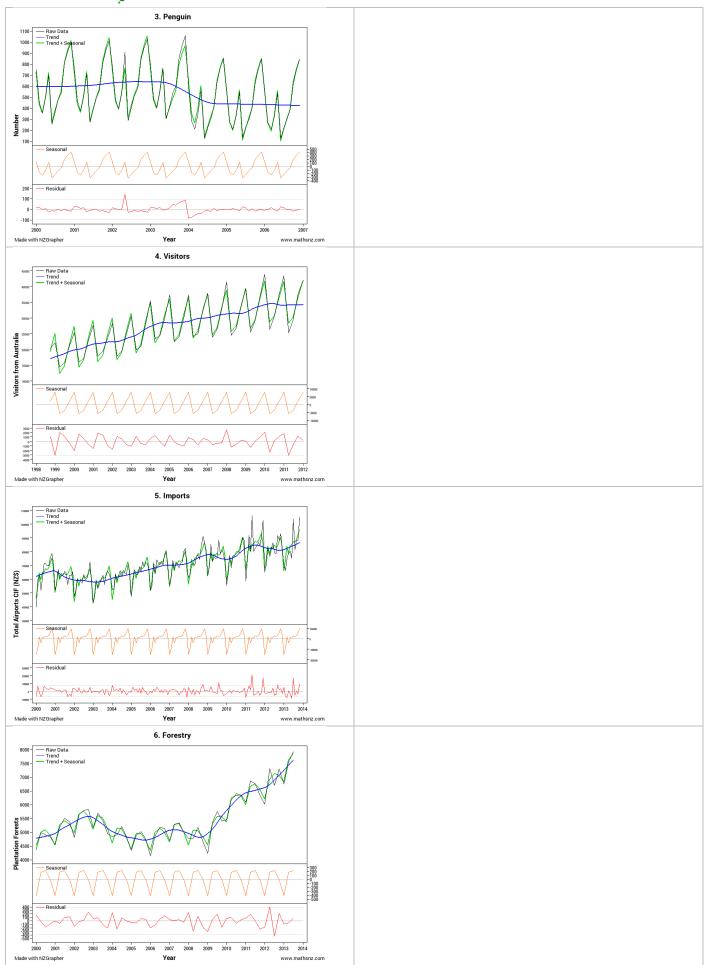
10

If any of the residuals are either larger or smaller than this value we need to comment on them and what might be causing them. You could also be thinking about how big the variation is of the residuals is as a component of the overall variation.

For each of the graphs calculate if there are any outliers and comment on any unusual features. The first two have been done for you.







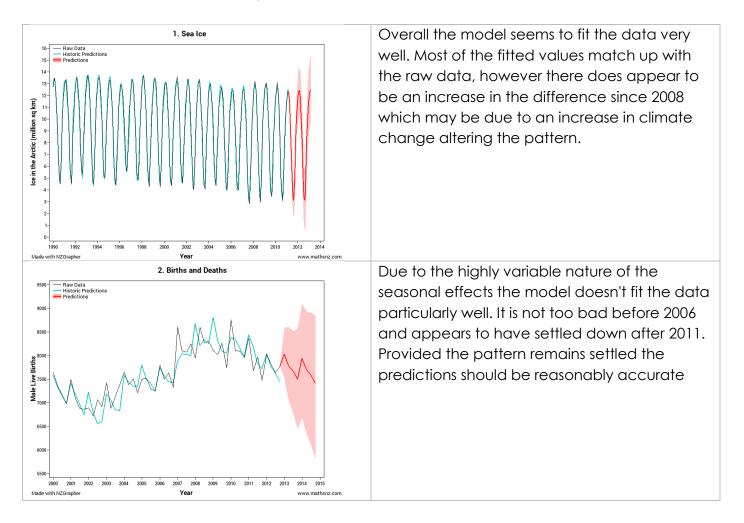


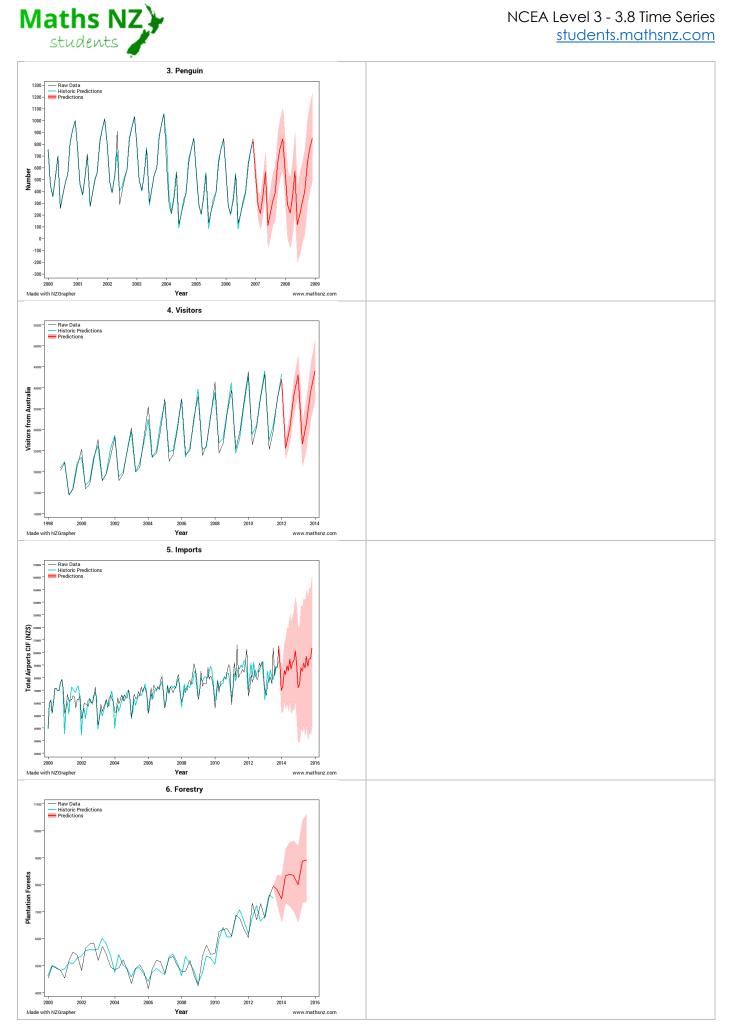
Part 4.4: Appropriateness of the Model

At this point in your report you should include the forecast graph we produced earlier.

You need to make a comment around how well the forecast model fits the data by looking at how the historic predictions fits with the raw data, particularly commenting on any time periods where it does not fit particularly well. Try and link these to how accurate you think the predictions will be. You can also discuss how well the recomposed data we looked at earlier fits the data.

For each of the graphs comment on how well the model fits the data. The first two have been done for you.







Part 4.5: Predictions

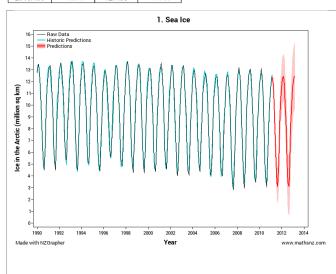
At this point in your report you should include the forecast output we produced earlier eg:

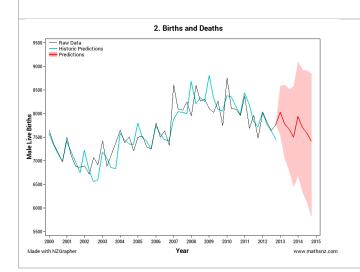
Time	Min	Prediction	Max
2011M04	11.117	11.690	12.274
2011M05	9.3005	10.102	10.888
2011M06	6.9482	7.8910	8.8742
2011M07	3.9685	5.1036	6.2087
2011M08	2.0775	3.3529	4.6057
2011M09	1.5600	3.0741	4.5038
2011M10	3.5709	5.1246	6.6439
2011M11	6.5583	8.1878	9.9006
2011M12	8.3936	10.095	11.894
2012M01	9.6028	11.406	13.346
2012M02	10.218	12.227	14.232
2012M03	10.369	12.420	14.533
2012M04	9.5365	11.706	13.927
2012M05	7.8806	10.118	12.481
2012M06	5.5889	7.9068	10.341
2012M07	2.7299	5.1194	7.5392
2012M08	0.90775	3.3687	5.7740
2012M09	0.48142	3.0900	5.5093
2012M10	2.3890	5.1405	7.6785
2012M11	5.4112	8.2036	10.868
2012M12	7.3447	10.111	12.798
2013M01	8.7271	11.422	14.285
2013M02	9.3710	12.242	15.102
2013M03	9.5273	12.435	15.353

We then need to write up **two** of these predictions in sentences. In these sentences it is VITAL you include the units, and round the numbers appropriately, and not only give the exact number as well as a range that it is likely to be between.

For each of the graphs make two predictions, the first two have been done for you.

Note: every time the predictions get produced in NZGrapher the min and max will be slightly different as they are produced using a bootstrapping process

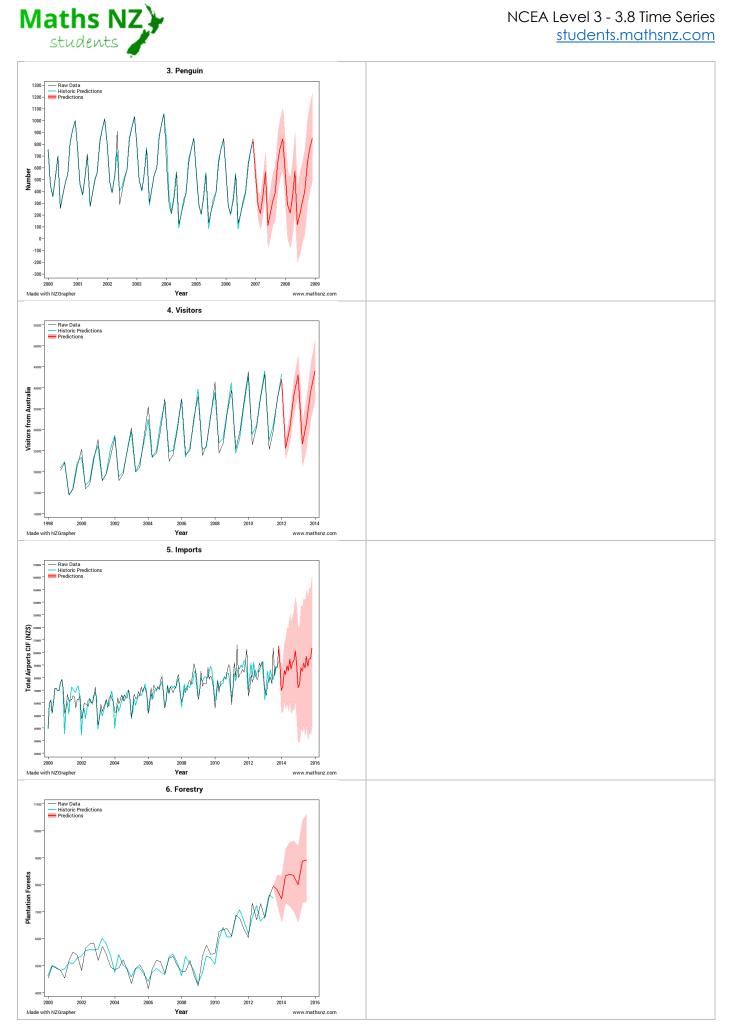




In April 2011 I would expect the amount of sea ice for the Artic to be 11.69 million square kilometres. However I cannot be completely certain about my prediction, but I would expect the area of ice will between 11.11 million square kilometres and 12.27 million square kilometres. In March 2013 I would expect the amount of sea ice for the Artic to be 12.44 million square kilometres. However I cannot be completely certain about my prediction, but I would expect the area of ice will between 9.53

million square kilometres and 15.35 million square kilometres.

In January to March 2013 I would expect there to be approximately 8030 male live births. However I cannot be completely confident in my prediction, but I would expect the number of male live births for this quarter to be somewhere between 7540 and 8590. In October to December 2014 I would expect there to be approximately 7410 male live births. However I cannot be completely confident in my prediction, but I would expect the number of male live births for this quarter to be somewhere between 5890 and 9050.



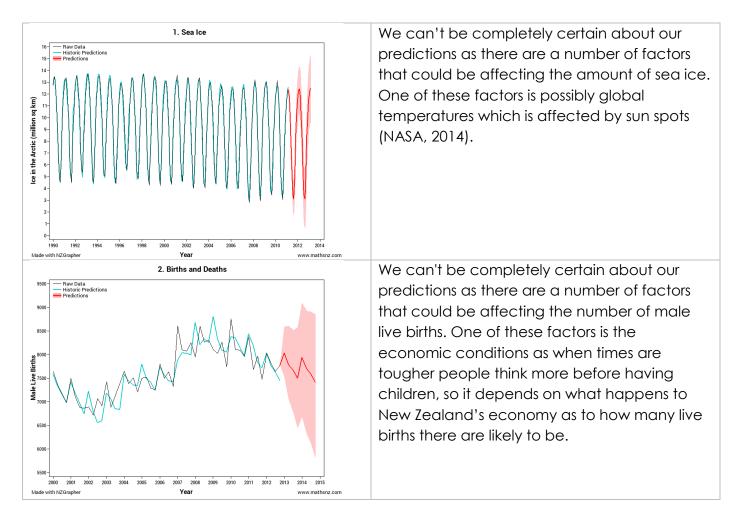


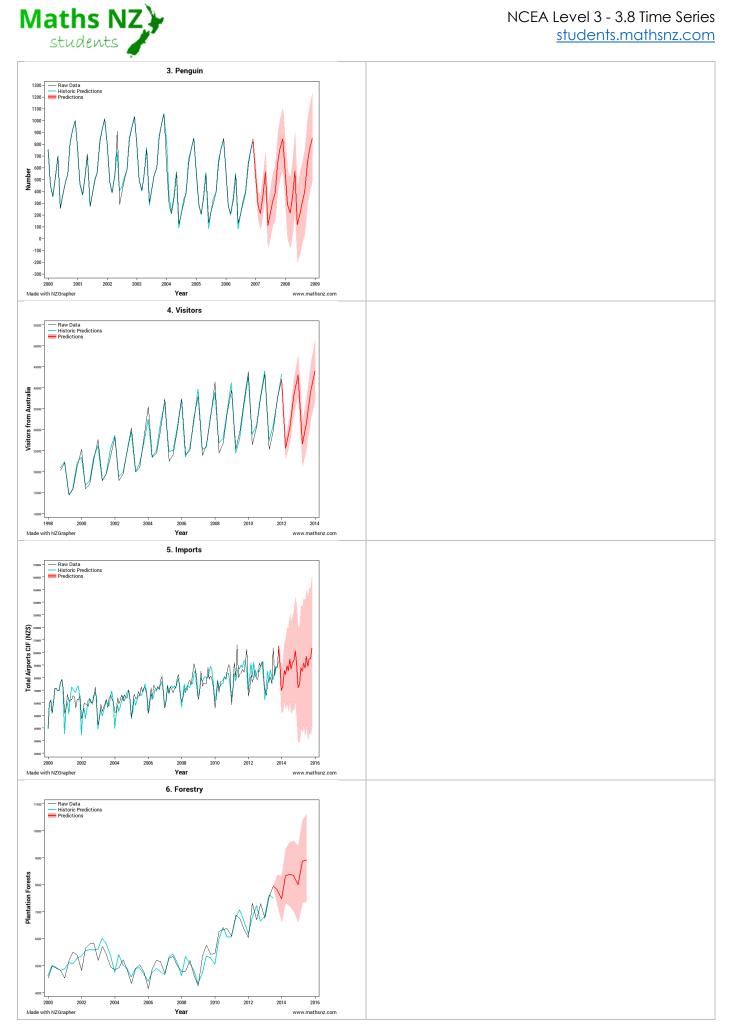
Part 4.6: Reasons for Confidence

With our predictions we normally cannot be completely confident in our predictions. It is useful to discuss other factors that may be influencing the predictions and how much effect these will have on your predictions.

This is a great time to include some extra research.

For each of the graphs make comments about your predictions. The first two have been done for you.





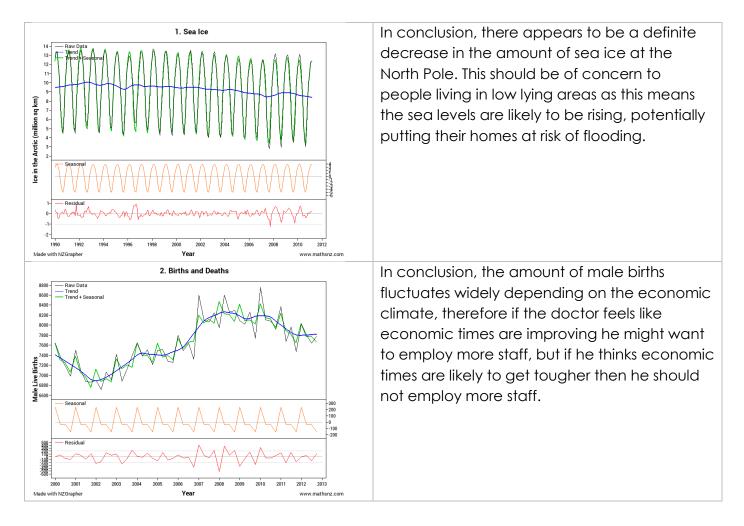


Part 5: Conclusion

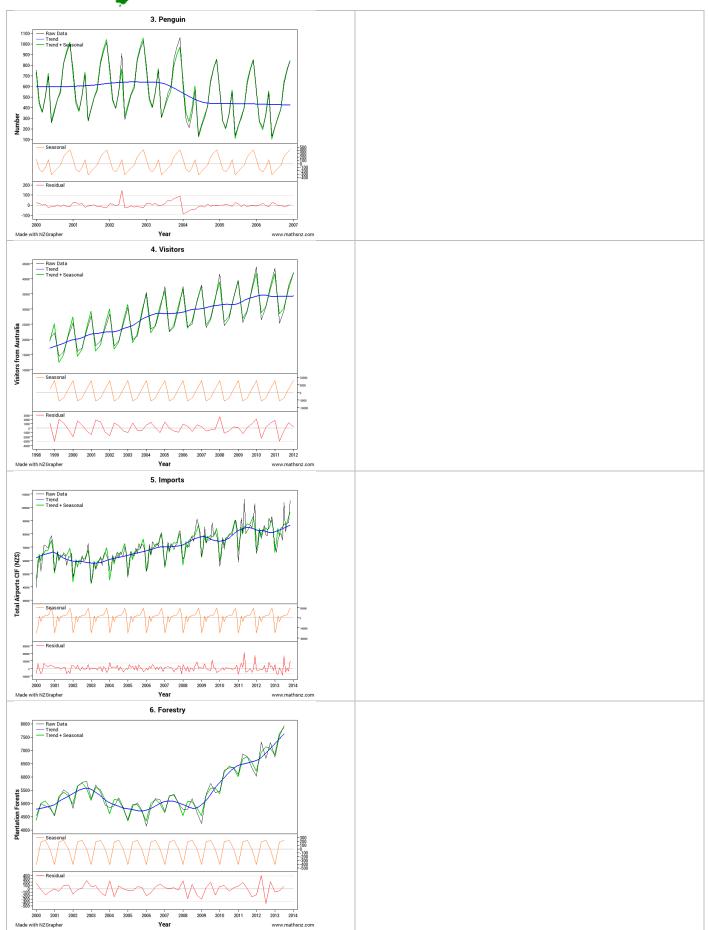
In the conclusion we need to summarise and link back to the purpose. It is important to:

- Give a brief summary
- Link back to the research... do our findings match up?
- What is the impact of these findings?

For each of the graphs write up a conclusion, the first two have been done for you.





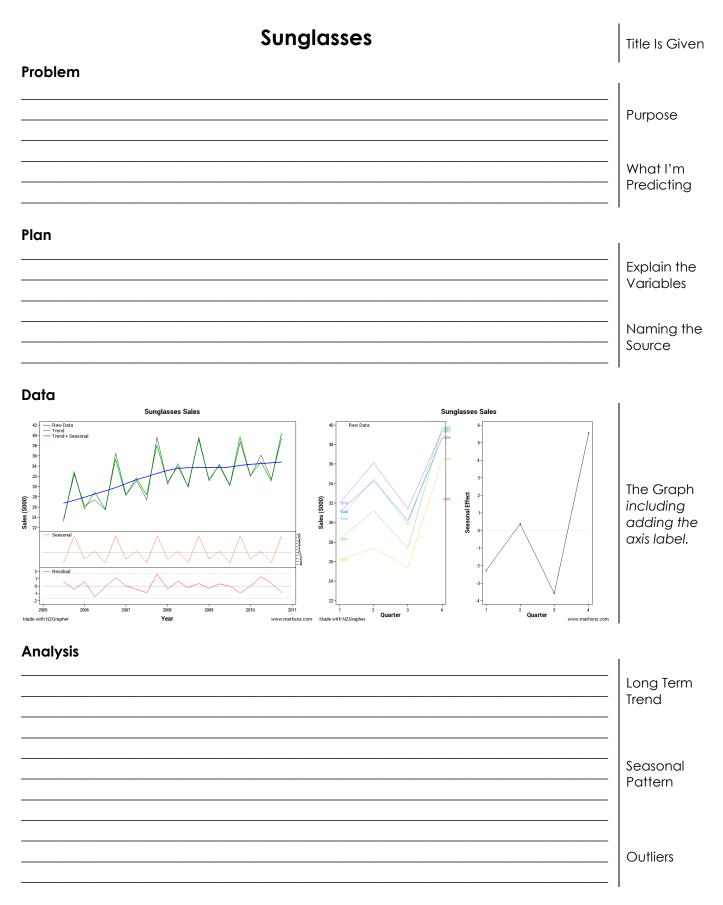


Congratulations, you have now written up a report for 4 different sets of data, so you now should be able to write up your own internal. Don't forget to give your report a title.

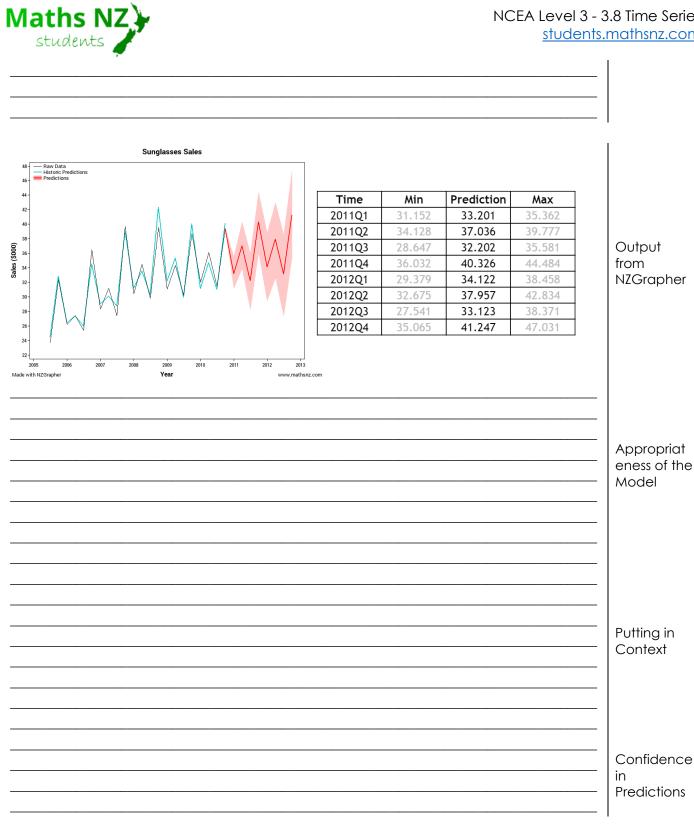


Part 6a: Writing Your Own Internal 1

Using the framework below write a report on the sunglasses sales data.





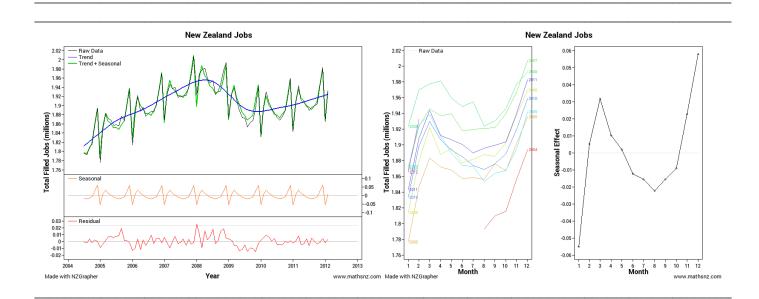


Conclusion

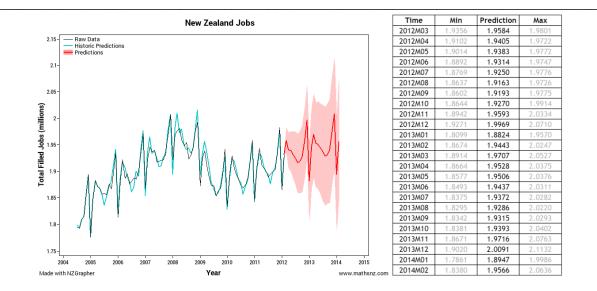


Part 6b: Writing Your Own Internal 2

This time you have just been provided with graphs. Using these write your own internal.









Ice at the North Pole	Title Is Given
Problem People who live in low lying areas are concerned about the amount of ice in the sea as when it melts they are worried about their homes flooding. "Even a modest rise in sea levels could cause flooding problems for low-lying coastal areas." ¹	Purpose
One of the main areas of ice in the world is at the North Pole, or the Arctic, so I am going to predict the surface area of the ice at the North Pole for the next two years.	What I'm Predicting
Plan The sea ice is the surface area of sea ice in the Arctic Circle which is measured in millions of square kilometres.	Explain the Variables
The data used in this investigation is from the National Snow and Ice Data Center from 1990 - 2011.	Naming the Source
<figure><figure></figure></figure>	The Graph including adding the axis label.
Analysis By looking at the recomposed data I can see that overall the amount of sea ice in the artic area appears to be generally decreasing from about 10 million square kilometres on average in 1990 to approximately 8 million square kilometres on average in 2011.	Long Term Trend
The largest amount of sea ice at the North Pole normally occurs during February and March when the amount of sea ice is approximately 4 million square kilometres above the long term trend. Between April and August there is a consistent drop from month to month as temperatures are rising. The smallest amount of sea ice regularly occurs in September where the amount of sea ice is approximately 5 million square kilometres below the long term trend. After this the temperatures start cooling down again increasing the amount of ice on a month to month basis until it returns to its peak in February and March.	Seasonal Pattern

¹ <u>http://science.howstuffworks.com/environmental/green-science/global-warming4.htm</u>



Absolute Highest Value: 14, Absolute Lowest Value: 3. $\frac{14-3}{10} = 1.1$ Looking at the residuals graph there in only one point that is more than 1.1 million square kilometres away from the trend. This occurred in September 2007 and may have been due to an unusually hot summer.	Outliers	
I. Sea leaInternet definition of predictionsInternet definition of predictions <td colspa<="" td=""><td>Output from NZGrapher</td></td>	<td>Output from NZGrapher</td>	Output from NZGrapher
Overall the model seems to fit the data very well. Most of the fitted values match up with the raw data, however there does appear to be an increase in the difference since 2008 which may be due to an increase in climate change altering the pattern.	Appropriateness of the Model	
In April 2011 I would expect the amount of sea ice for the Artic to be 11.69 million square kilometres. However I cannot be completely certain about my prediction, but I would expect the area of ice will between 11.11 million square kilometres and 12.27 million square kilometres. In March 2013 I would expect the amount of sea ice for the Artic to be 12.44 million square kilometres. However I cannot be completely certain about my prediction, but I would expect the area of ice will between 9.53 million square kilometres and 15.35 million square kilometres.	Putting in Context	
We can't be completely certain about our predictions as there are a number of factors that could be affecting the amount of sea ice. One of these factors is possibly global temperatures which is affected by sun spots.	Confidence in Predictions	
Conclusion In conclusion, there appears to be a definite decrease in the amount of sea ice at the North Pole. This should be of concern to people living in low lying areas as this means the sea ice levels are likely to be rising, potentially putting their homes at risk of flooding.	Summarise and Link Back to the Purpose	



Data Set Information

Births and Deaths

Data on the number of births and deaths in New Zealand. The data is sourced from Statistics New Zealand.

Variable	Description
Quarter	Quarterly
Male Live Births	Number of males born during the quarter
Female Live Births	Number of females born during the quarter
Male Deaths	Number of male deaths during the quarter
Female Deaths	Number of female deaths during the quarter

Forestry

The volume of wood removed from different types of forests in New Zealand. The data is sourced from the Ministry for Primary Industries.

Variable	Description
Quarter	Quarterly
Natural Forests	The volume of wood removed from Natural Forests in millions of m ³
Plantation Forests	The volume of wood removed from Plantation Forests in millions of m ³

Imports

Information on imports to and from New Zealand. The data is sourced from Statistics New Zealand.

Variable	Description
Month	Monthly
TotalAirportsCIF	Cost, insurance and freight of imported goods in NZ\$(000)
TotalParcelPostCIF	Cost, insurance and freight of imported goods in NZ\$(000)
TotalSeaportsCIF	Cost, insurance and freight of imported goods in NZ\$(000)
TotalAirportsWeight	Weight of imported goods in tonnes
TotalParcelPostWeight	Weight of imported goods in tonnes
TotalSeaportsWeight	Weight of imported goods in tonnes

Jobs

The number of people in employment in New Zealand. The data is sourced from Statistics New Zealand.

Variable	Description
Month	Monthly
Total Filled Jobs	The number of jobs that are filled



Penguin

Data on the number of penguins at the Phillip Island Penguin Parade in Australia.

Variable	Description	
Month	Monthly	
Number	The number of penguins in the colony	

Sea Ice

The data is the surface area of sea ice in millions of square kilometres. The data is sourced from the National Snow and Ice Data Center.

Variable	Description	
Time	Monthly	
Arctic	Million Square Kilometres of Ice in the Arctic	
Antarctica	Million Square Kilometres of Ice in Antarctica	

Sunglasses

Data on the value of sunglasses sold.

Variable	Description	
Quarter	Quarterly	
Sales	Amount of sales in thousands of dollars	

Visitors

The visitors' dataset is the number of people entering New Zealand on a Visitor Visa from Australia, China, Japan and the UK.

The data is sourced from Statistics New Zealand.

Variable	Description	
Date	Quarterly	
Australia	Number of visitors in the quarter from Australia	
China, People's Republic of	Number of visitors in the quarter from China	
Japan	Number of visitors in the quarter from Japan	
United Kingdom	Number of visitors in the quarter from the UK	



Assessment Guidelines – 91580 – Investigate Time Series Data

	Achieved (all compulsory)	Merit Achieved PLUS	Excellence Merit PLUS
Problem	Identify a purpose for the investigation	Purpose is clear (compulsory)	Research is used to develop purpose (compulsory)
Plan	Select a variable to investigate that links to the purpose	Context is researched	
Data	Graph the raw and the smoothed data Appropriate model is fitted and can be given as an equation, a fitted line or a graph of the smoothed data Must have the following: • Recomposition graph • Seasonal Plot and estimated seasonal effects • Raw data plus predictions plus prediction intervals Remember to include the units on the axis if appropriate.	Graph must have a title, correctly labelled axis and any series shown on the graph(s) are clearly identified	Other relevant variables are discussed. This could include creating a new variable from the variables given
Analysis	Give quantitative description of the trend and linked to context including stating the start and end points. (long term trend) Seasonal pattern described and linked to context (seasonal effects) Other relevant features are identified (residuals) Forecasts are made with correct units	The appropriateness of the model is justified throughout the entire range of x-values Other relevant features must be explained in context and comments made must be supported with statistical evidence Forecasts are given in context and rounded correctly Discussion given on how precise the predictions might be based on reliability of the trend or seasonal components Understanding shown that forecasts are estimates	Provide possible explanations for the features of the graph A comparison of predicted values for the most recent data values of the model could be made Comparison between two data sets could be made
Conclusion	Conclusion is consistent with the purpose of the investigation	Conclusion must be given in context (compulsory)	The purpose of the investigation is addressed (compulsory) There is a reflection on the analysis with respect to the background research undertaken Impact of the findings is discussed

Final grades will be decided using professional judgement based on a holistic examination of the evidence provided against the criteria in the Achievement Standard.