

Level 3 - AS91580 - 4 Credits - Internal

Investigate Time Series Data

Written by J Wills – MathsNZ – <u>jwills@mathsnz.com</u>

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	
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)	28
Data Set Information	30
Assessment Guidelines – 91580 – Investigate Time Series Data	32

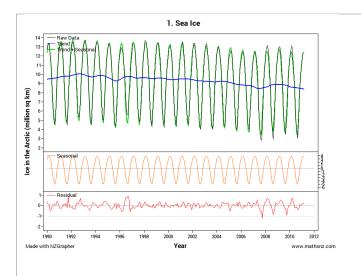


Part 1: Problem

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.

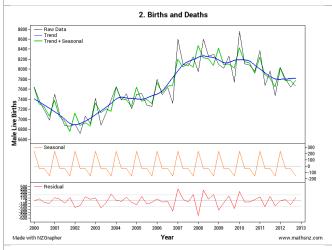
Note: NZGrapher will always predict for the next two years



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."

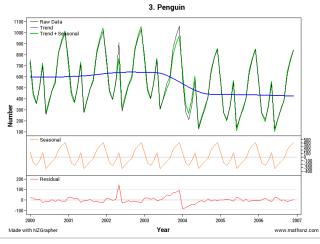
One of the main areas of ice in the world is at the North Pole, or the Arctic, so I am going to predict the surface area of the ice at the North Pole for the next two years.

 http://science.howstuffworks.com/environmental/greenscience/global-warming4.htm

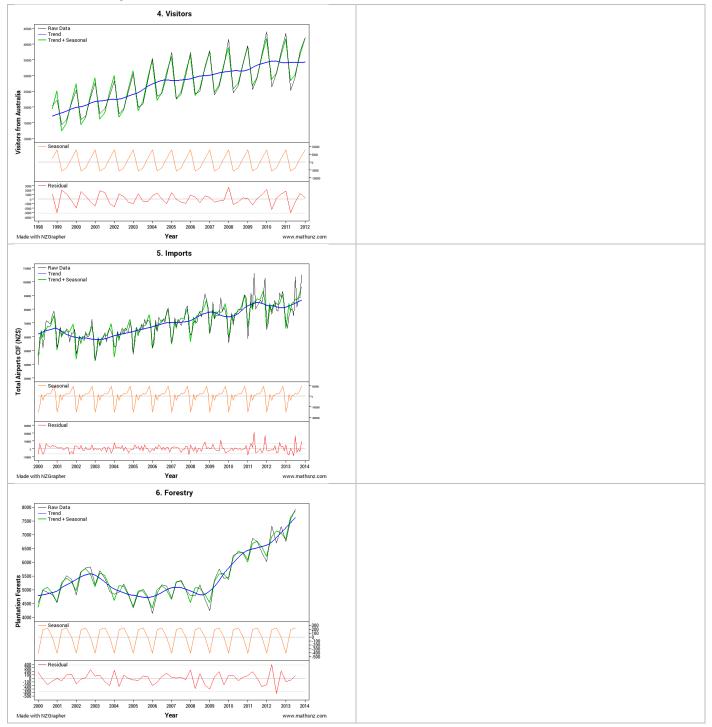


A doctor at the local birthing unit has asked me to look at the number of births so that he can work out if he should be employing more staff or not.

The number of males and females born are approximately the same, so I am going to predict the number of live male births in New Zealand for the next two years.







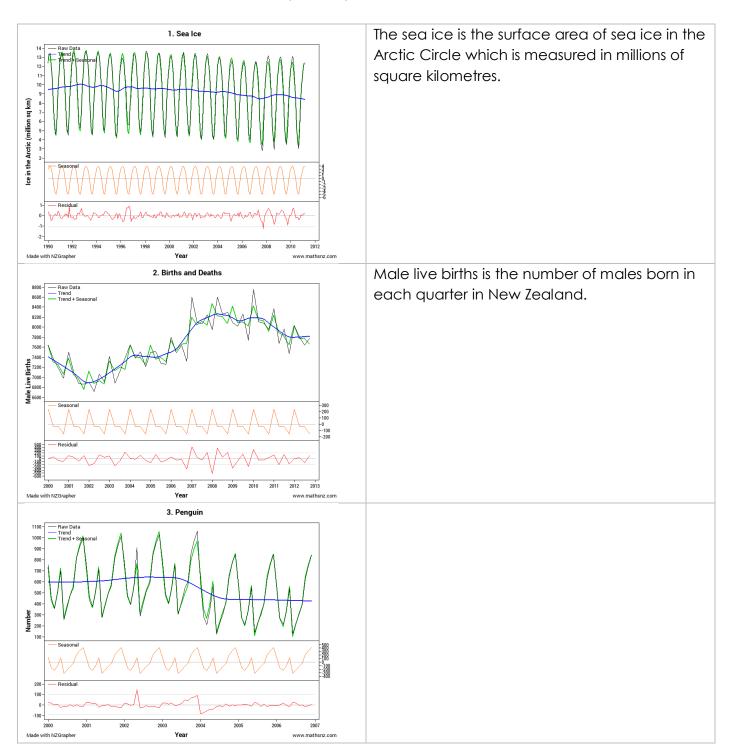


Part 2: Plan

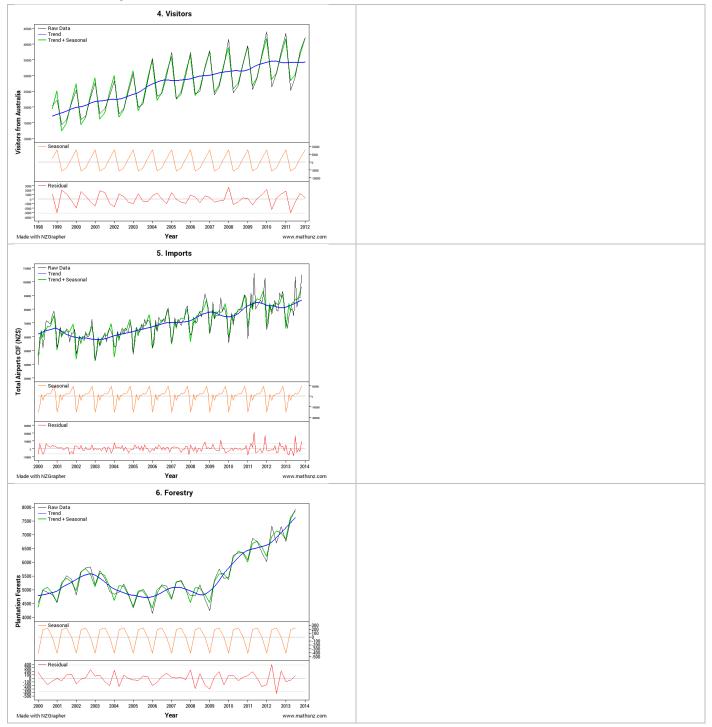
Part 2.1: Explain the Variables

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.



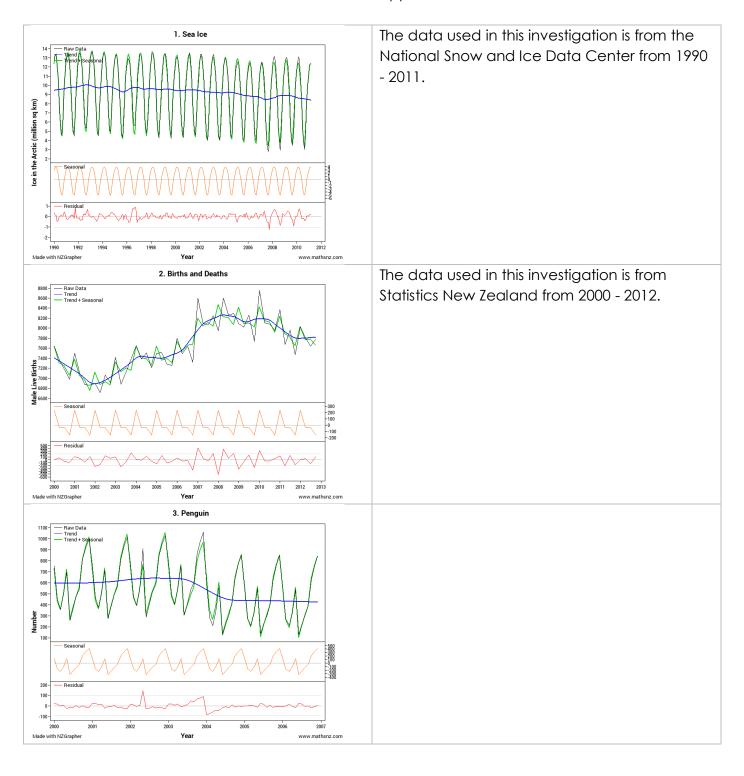




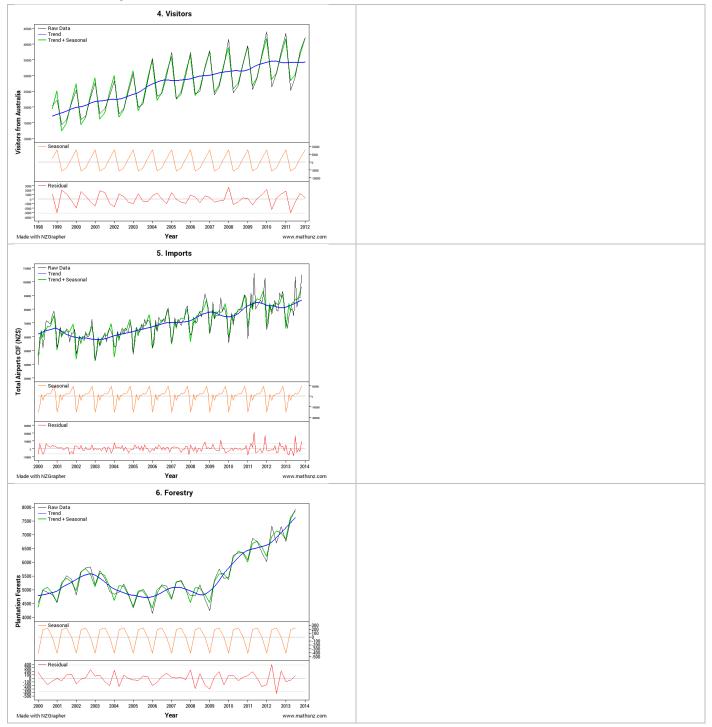


Part 2.2: Naming the Source

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.

First up we need to start NZGrapher by going to the link in the box to the right.

The first time you load NZGrapher it will display an overlay with descriptions as to what all the different areas do as shown to the right. To load your data in either select it from the dropdown in the top right, or upload it in the top left corner and press go.

www.jake4maths.com/grapher



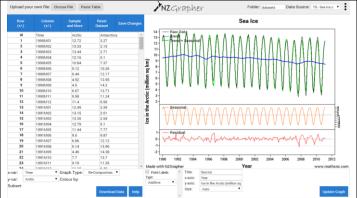
To draw a dot plot there are just three things you need to do.

- 1. Select the x-variable... this is your time series variable that will be on the x-axis, in this case it's 'Time'.
- 2. Select the y-variable... this is your actual data, in this case it's 'Arctic'.
- 3. Select the graph type... for this we want the 'time series re-composition'.

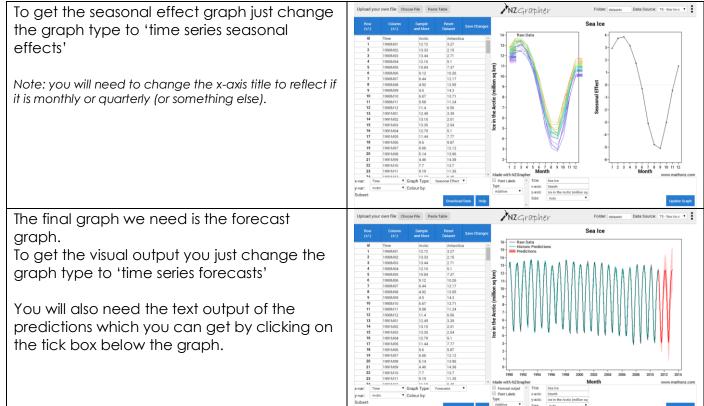
You then just need to check the graph title and axis labels to make sure they are appropriate (include units where necessary) and add press update graph to save the titles.

To copy the graph just right click and press copy image, or to save press 'Save Image As' or whatever your device says that is similar.

Note: If you want to identify the extreme points, if you click the 'Point Labels' checkbox this will add little numbers next to the points that correspond with the point id.







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: http://www.mathsnz.com/inzight-tips/



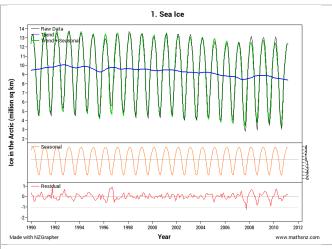
Part 4: Analysis

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.

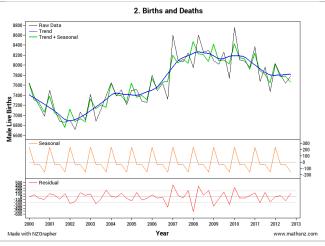
Part 4.1: Long Term Trend

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.

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

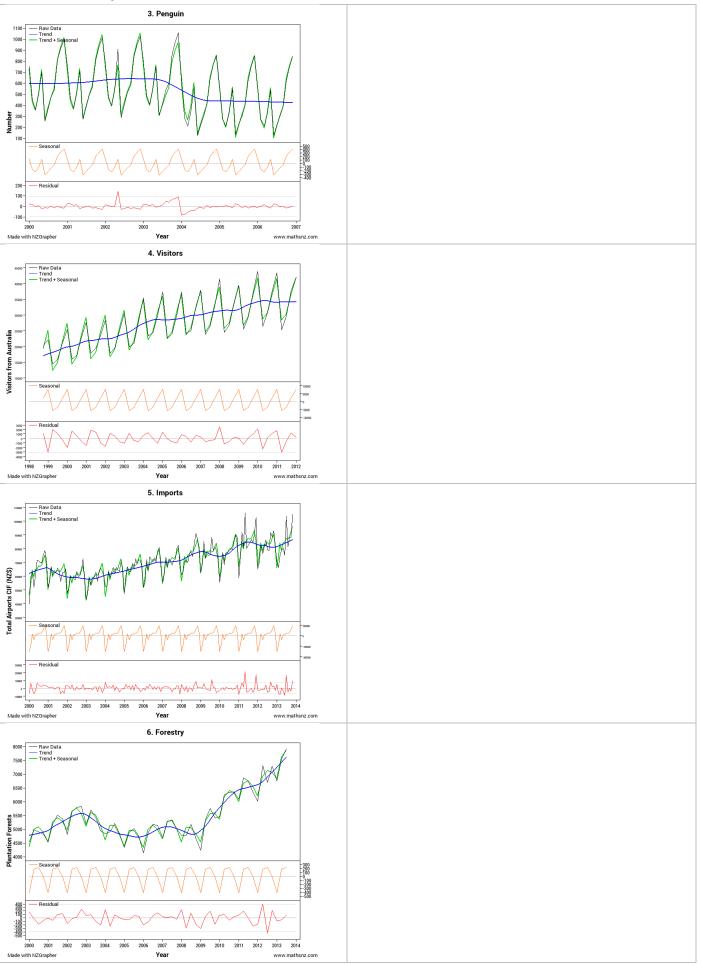


By looking at the recomposed data I can see that overall the amount of sea ice in the artic area appears to be generally decreasing from about 10 million square kilometres on average in 1990 to approximately 8 million square kilometres on average in 2011.



By looking at the recomposed data I can see that overall the amount of male live births has increased since 2000. In 2000 there were approximately 7400 births per quarter on average, this dropped to approximately 6800 per quarter on average in 2002, climbed to a peak of 8200 per quarter on average in 2008, possibly due to the financial boom, and has dropped off slightly to approximately 7800 per quarter on average in 2013.



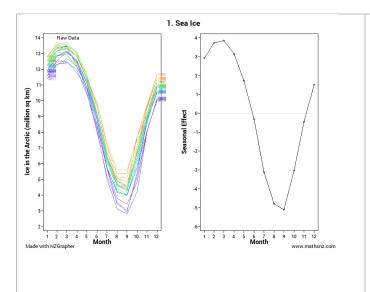




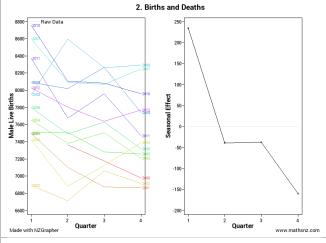
Part 4.2: Seasonal Pattern

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.

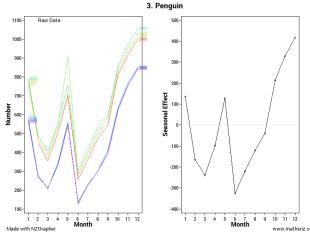
For each of the graphs write a comment about the long term trend. The first two have been done for you.



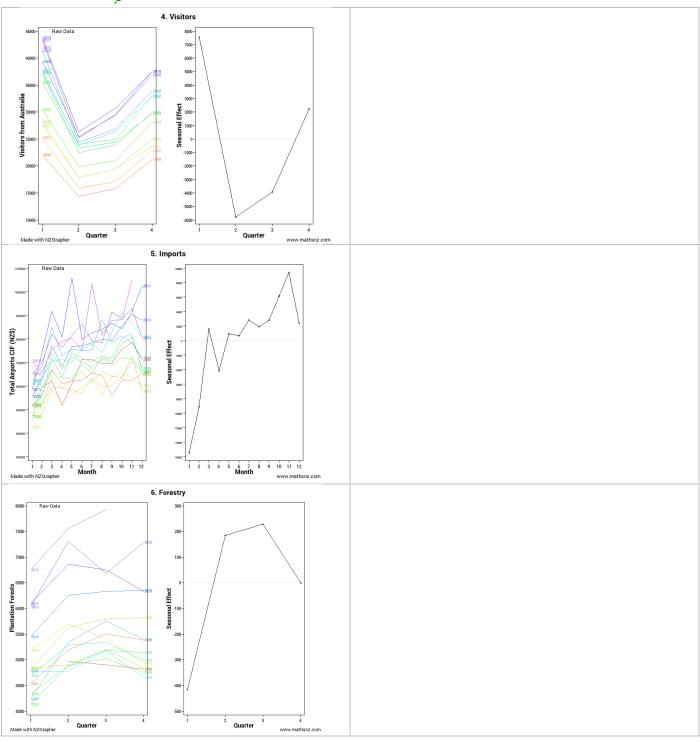
The largest amount of sea ice at the North Pole normally occurs during February and March when the amount of sea ice is approximately 4 million square kilometres above the long term trend. Between April and August there is a consistent drop from month to month as temperatures are rising. The smallest 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.



The seasonal pattern for male live births is not very consistent. On average the highest number of births occurs in the first quarter (January to March) where there are approximately 150 more births than the long term trend. The lowest on average occurs in the fourth quarter (October to December) where there are approximately 200 births less than the long term trend, however there are many years that it does not follow this pattern.









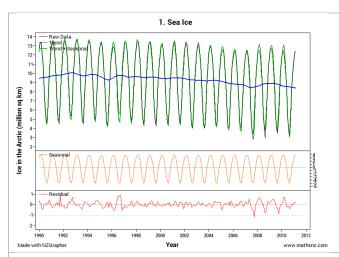
Part 4.3: Outliers

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:

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.

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.

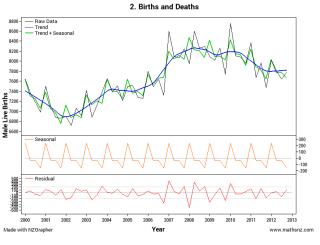


Absolute Highest Value: 14,

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.

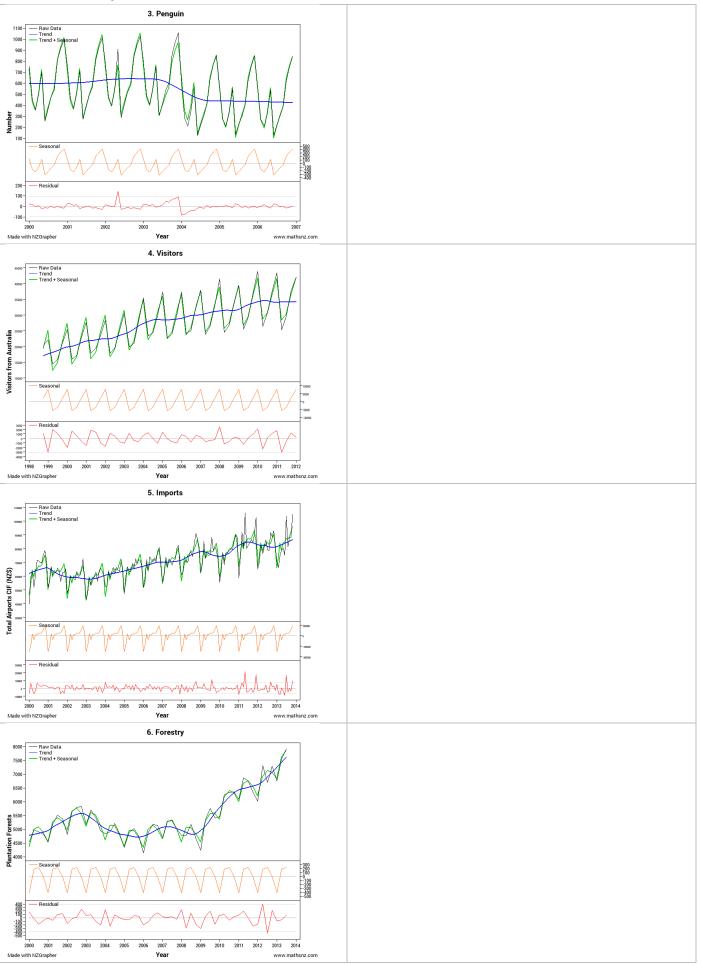


Absolute Highest Value: 8700 Absolute Lowest Value: 6700

$$\frac{8700 - 6700}{10} = 200$$

Looking at the residuals graph due to the inconsistency of the data there are a large number of residuals between 2007 and 2010 that are outside the acceptable range. This is during the financial boom so it may be due to people being more willing to have children and therefore not worrying about the timing so much, therefore the normal patterns do not happen.





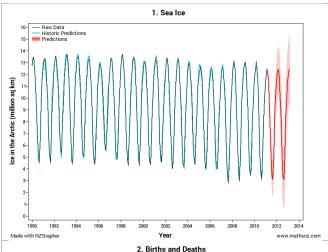


Part 4.4: Appropriateness of the Model

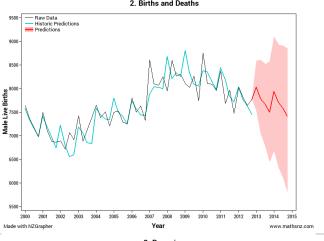
At this point in your report you should include the forecast graph we produced earlier.

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.

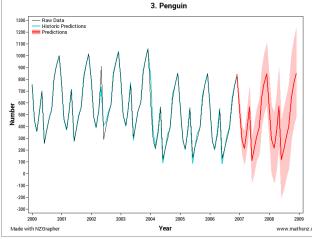
For each of the graphs comment on how well the model fits the data. The first two have been done for you.



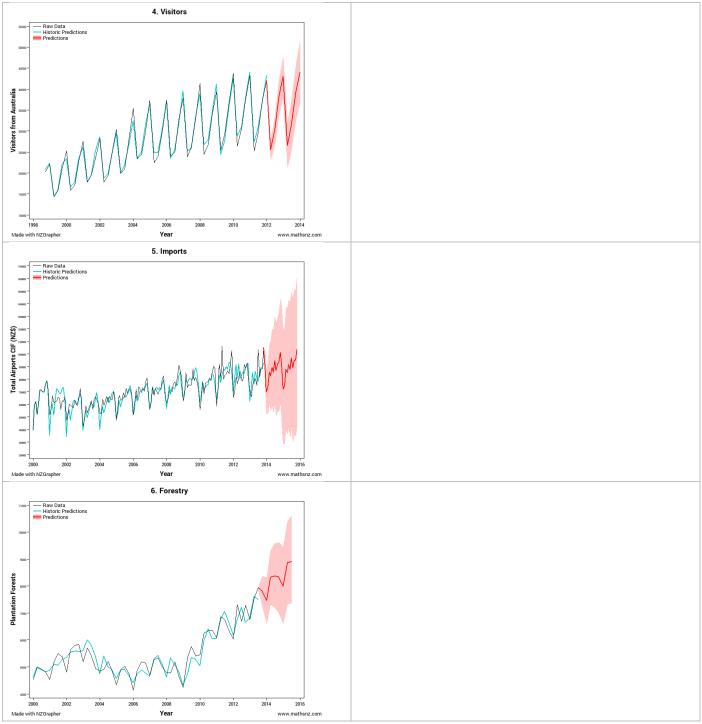
Overall the model seems to fit the data very well. Most of the fitted values match up with the raw data, however there does appear to be an increase in the difference since 2008 which may be due to an increase in climate change altering the pattern.



Due to the highly variable nature of the seasonal effects the model doesn't fit the data particularly well. It is not too bad before 2006 and appears to have settled down after 2011. Provided the pattern remains settled the predictions should be reasonably accurate









Part 4.5: Predictions

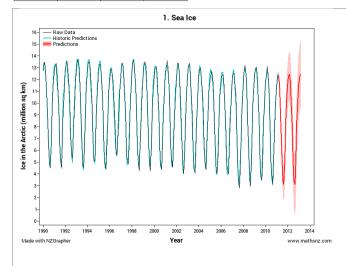
At this point in your report you should include the forecast output we produced earlier eg:

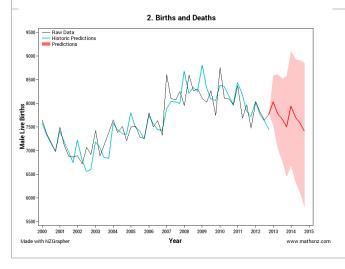
Time	Min	Prediction	Max
2011M04	11.117	11.690	12.274
2011M05	9.3005	10.102	10.888
2011M06	6.9482	7.8910	8.8742
2011M07	3.9685	5.1036	6.2087
2011M08	2.0775	3.3529	4.6057
2011M09	1.5600	3.0741	4.5038
2011M10	3.5709	5.1246	6.6439
2011M11	6.5583	8.1878	9.9006
2011M12	8.3936	10.095	11.894
2012M01	9.6028	11.406	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.

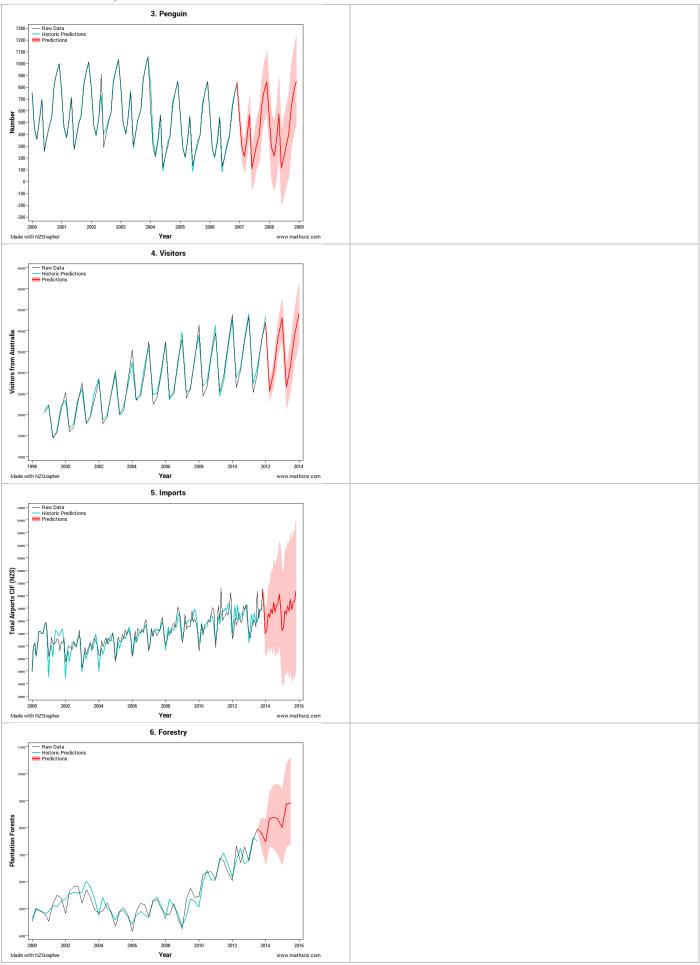
In March 2013 I would expect the amount of sea ice for the Artic to be 12.44 million square kilometres. However I cannot be completely certain about my prediction, but I would expect the area of ice will between 9.53 million square kilometres and 15.35 million square kilometres.

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.





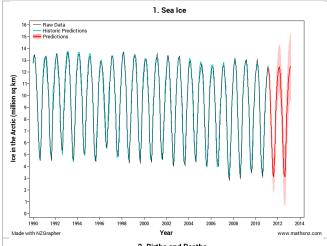


Part 4.6: Reasons for Confidence

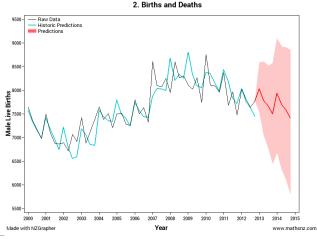
With our predictions we normally cannot be completely confident in our predictions. It is useful to discuss other factors that may be influencing the predictions and how much effect these will have on your predictions.

This is a great time to include some extra research.

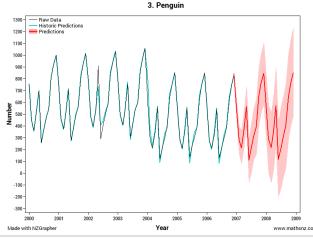
For each of the graphs make comments about your predictions. The first two have been done for you.



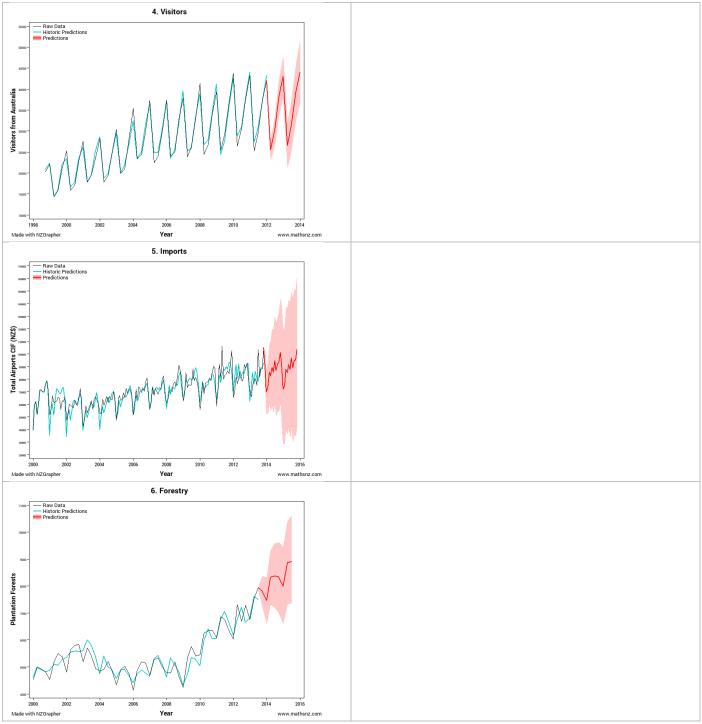
We can't be completely certain about our predictions as there are a number of factors that could be affecting the amount of sea ice. One of these factors is possibly global temperatures which is affected by sun spots (NASA, 2014).



We can't be completely certain about our predictions as there are a number of factors that could be affecting the number of male live births. One of these factors is the economic conditions as when times are tougher people think more before having children, so it depends on what happens to New Zealand's economy as to how many live births there are likely to be.







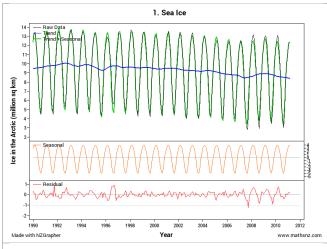


Part 5: Conclusion

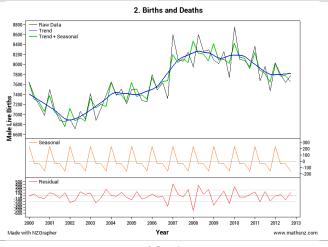
In the conclusion we need to summarise and link back to the purpose. It is important to:

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

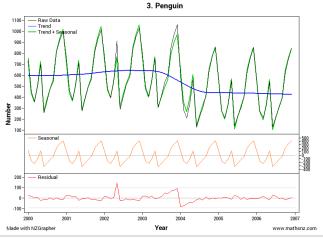
For each of the graphs write up a conclusion, the first two have been done for you.



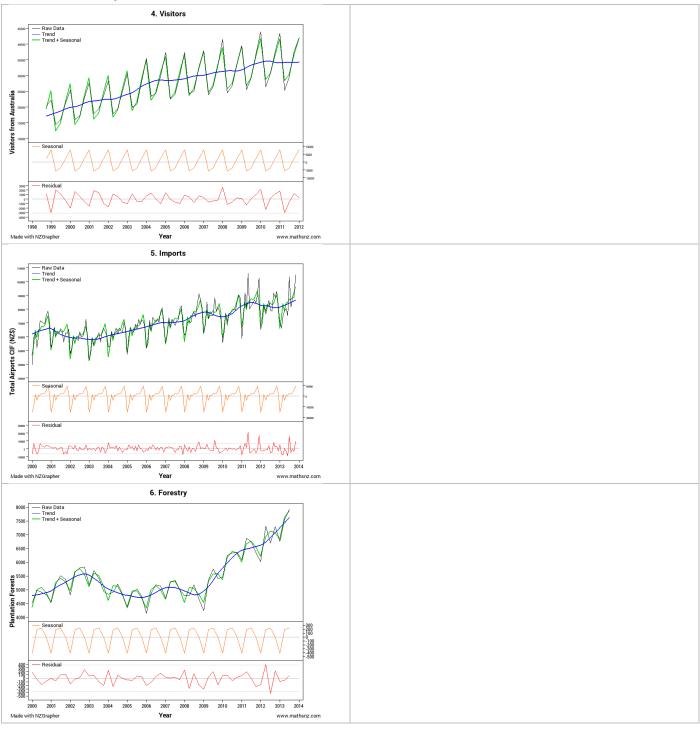
In conclusion, there appears to be a definite decrease in the amount of sea ice at the North Pole. This should be of concern to people living in low lying areas as this means the sea levels are likely to be rising, potentially putting their homes at risk of flooding.



In conclusion, the amount of male births fluctuates widely depending on the economic climate, therefore if the doctor feels like economic times are improving he might want to employ more staff, but if he thinks economic times are likely to get tougher then he should not employ more staff.







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.

Sunglasses	Title Is Given
Problem	,
	Purpose
	What I'm Predicting
Plan	ı
	Explain the Variables
	Naming the Source
Data	
Sunglasses Sales Sunglasses Sales Sunglasses Sales Sunglasses Sales Sunglasses Sales Sunglasses Sales Made with NZGrapher Sunglasses Sales Sunglasses Sales	The Graph including adding the axis label.
Analysis	I
	Long Term Trend
	Seasonal Pattern
	Outliers

Purpose



					_
Sunglasses Sales					— I
48- — Raw Data — Historic Predictions					
46 Predictions			I		
42-	Time	Min	Prediction	Max	
40-	2011Q1	31.152	33.201	35.362	
38-	2011Q2	34.128	37.036	39.777	Outout
Sales (6000) 36-	2011Q3	28.647	32.202	35.581	Output
<u>8</u> 34-	2011Q4 2012Q1	36.032 29.379	40.326 34.122	44.484 38.458	from
32-	2012Q1 2012Q2	32.675	37.957	42.834	NZGrapher
30-	2012Q2 2012Q3	27.541	33.123	38.371	
26-	2012Q3 2012Q4	35.065	41.247	47.031	
24-	201201	33.003	111217	17.1031	
22-					
					_
					_
					Appropriat
					eness of the
					— Model
					_
					_
					—
					_
					_
					Putting in
					Context
					_
					_
					_
					_
					- Confidence
					Confidence
					Predictions
					— Fredictions
Conclusion					ı
					_
					— — Summarise
					_ and Link
					Back to the



Part 6b: Writing Your Own Internal 2

This t	ne you have just been provided with graphs. Using these write your own internal.
2.02- 2- 1.86- 1.96- 1.94- 1.92- 1.9- (\$1.88- 1.81- 1.84- 1.87- 1.76- 1.76- 0.03- 0.02- 0.01- 0.01- 0.02- 0.01- 0.02- 0.02-	New Zealand Jobs Rew Data Trend Tre



	New Zealand Jobs	
2.15-	Raw Data Historic Predictions Predictions	
2.1-		
2.05- Su		
oillim) sı	1 . AA A	
Total Filled Jobs (millions)		
Total I.9-		
1.85		
1.8-	J	
1.75	-	
	2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 vith NZGrapher Year www.mathsr	

Time	Min	Prediction	Max
2012M03	1.9356	1.9584	1.9801
2012M04	1.9102	1.9405	1.9722
2012M05	1.9014	1.9383	1.9772
2012M06	1.8892	1.9314	1.9747
2012M07	1.8769	1.9250	1.9776
2012M08	1.8637	1.9163	1.9726
2012M09	1.8602	1.9193	1.9775
2012M10	1.8644	1.9270	1.9914
2012M11	1.8942	1.9593	2.0334
2012M12	1.9271	1.9969	2.0710
2013M01	1.8099	1.8824	1.9570
2013M02	1.8674	1.9443	2.0247
2013M03	1.8914	1.9707	2.0527
2013M04	1.8664	1.9528	2.0375
2013M05	1.8577	1.9506	2.0376
2013M06	1.8493	1.9437	2.0311
2013M07	1.8375	1.9372	2.0282
2013M08	1.8295	1.9286	2.0220
2013M09	1.8342	1.9315	2.0293
2013M10	1.8381	1.9393	2.0402
2013M11	1.8671	1.9716	2.0763
2013M12	1.9020	2.0091	2.1132
2014M01	1.7861	1.8947	1.9986
2014M02	1.8380	1.9566	2.0636

· · · · · · · · · · · · · · · · · · ·	

Sample Internal (at Achieved Level)

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."

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

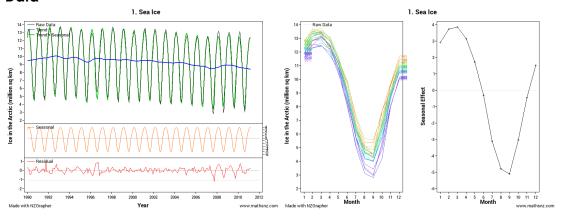
The sea ice is the surface area of sea ice in the Arctic Circle which is measured in millions of square kilometres.

Explain the Variables

The data used in this investigation is from the National Snow and Ice Data Center from 1990 - 2011.

Naming the Source

Data



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.

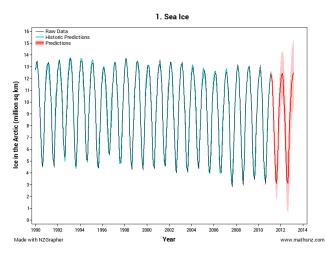
Seasonal Pattern

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



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



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

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 axes 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 (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.