Climate Change and Global Warming

Annie Rana
16CSU049
The NorthCap University
ranannie97@gmail.com

Anshum Gill
16CSU052
The NorthCap University
anshumgill@gmail.com

Abstract

Present paper focuses on the climatic changes along the equator covering countries falling at 15°N to 15°S and countries falling between 60°N-90°N and 60°S-90°S because the equatorial region receives abundant sunlight and an analysis of the poles represents the adversity of climate change. There has been a concern about whether a changing climate might significantly alter current weather patterns in such a way that the life form would be affected drastically. R programming made it possible to study and analyse the changing temperature patterns along the equator and the poles. Using Linear Regression Model, it was possible to predict the average temperatures of the future. It helped us to answer certain questions like, are heat waves becoming stronger and more intense with time and what effect will it have on the global warming phenomena.

1.Introduction

The planet's temperature is rising. The trend is clear and unmistakable. Global warming is known as a persistent threat to the environment, the livelihoods of people and their health. The sea levels are rising; number of wildfire occurrences is growing; dangerous heat waves are becoming more common. Extreme climatic conditions like floods, cyclones, thunderstorms occur more frequently than ever. Droughts are starting to develop a perennial nature.

This compels us to scientifically find the trends in temperatures across the globe so as to realise the damage the mankind has caused to the environment over the past few decades.

This paper focuses on land temperature changes along the equator (15°N-15°S) and the North and South poles (60°N-90°N and 60°S-90°S). The Equatorial Region was considered as it is the closest to the sun and experiences more direct sunlight than any other region on the planet, it is also the warmest region on the planet hence the outliers would represent a drastic change. The Polar Regions are considered as they offer a visual representation of global warming. The increase in temperature is causing glaciers/ice caps situated in the poles to melt thereby increasing the sea levels across the globe which causes numerous disasters threatening our very existence. [1]

Using the data from 1950-2013 for the country of India, the average temperature patterns for the year 2950-3013 were predicted to understand how the current temperature trend would impact on the future.

The common belief around global warming is that it represents the increase in temperatures but actually, Global Warming is a scientific study of atmospheric sciences which enables the researchers to analyse and predict the various parameters surrounding climate like, the thermal energy accessible to a region, the likelihood of snow and hail reaching the surface. [2]

2.Dataset

Climate Change: Earth Surface Temperature Data

The dataset was obtained from Kaggle.com. The early data was recorded using mercury thermometers where any variation in the visit time of data collection impacted the measurements. Given the complexity of the dataset there are a range of organizations which record the said data. The three most cited Land and Ocean temperature datasets are NOAA's MLOST, NASA's GISTEMP and the UK's HadCrut. [3]

The data was repackaged from a newer compilation done by the Berkeley Earth, affiliated with Lawrence Berkeley National Laboratory. The study combined 1.6 billion temperature reports from 16 pre-existing archives. They applied methods that allowed frequent weather observations to be included, which meant that fewer observations had to be discarded. [3]

The data consists of average temperatures of cities across the world recorded over a month from 1750-2010. Total of 8,599,212 million records exist in the dataset.

The dataset includes the following columns:

- dt: The date of observation in yyyy-mm-dd format.
- AverageTemperature: The average temperature recorded over a month.
- AverageTemperatureUncertainty: The 95% confidence interval around the average.
- City: List of all the cities across the world.
- Country: List of countries across the globe.
- Latitude: Latitude location of the cities (N or S).
- Longitude: Longitude location of the cities (W or E).

dt	AverageTemperature	AverageTemperatureUncertainty	City	Country	Latitude	Longitude
1743-11-01	6.068	1.737	Ã...rhus	Denmark	57.05N	10.33E
1744-04-01	5.788	3.624	Ã...rhus	Denmark	57.05N	10.33E
1744-05-01	10.644	1.283	Ã...rhus	Denmark	57.05N	10.33E
1744-06-01	14.051	1.347	Ã...rhus	Denmark	57.05N	10.33E
1744-07-01	16.082	1.396	Ã...rhus	Denmark	57.05N	10.33E
1744-09-01	12.781	1.454	Ã...rhus	Denmark	57.05N	10.33E
1744-10-01	7.950	1.630	Ã...rhus	Denmark	57.05N	10.33E
1744-11-01	4.639	1.302	Ã...rhus	Denmark	57.05N	10.33E
1744-12-01	0.122	1.756	Ã...rhus	Denmark	57.05N	10.33E
1745-01-01	-1.333	1.642	Ã...rhus	Denmark	57.05N	10.33E
1745-02-01	-2.732	1.358	Ã...rhus	Denmark	57.05N	10.33E
1745-03-01	0.129	1.088	Ã...rhus	Denmark	57.05N	10.33E

Table 1 Raw Dataset

3.Data cleansing

As the dataset consists of 8,599,212 rows of data, there were numerous NA or NULL values included in the dataset too. To be able to carry out the research it was necessary to remove these values first. As a result, a total of 364,130 records were removed from the dataset.

The dataset also consisted of only character or string datatype, hence a type conversion was required on certain columns.

In R [4], the dt column was converted into Date Type format from which Years and Months were recorded into separate new columns.

Both the latitude and longitude values consisted of either of N/S or E/W at the end of the string, example: 77N latitude and 20E longitude. In order to separate the data into Equatorial and Polar Regions a conversion from string type to integer type was required. As integer types can only consist of numerical values, the N/S or E/W delegation had to be omitted. This was done by taking the length of the substring i.e. the latitude or the longitude and the last character was removed after which it was converted into an integer datatype.

The City and AverageTemperatureUncertinity columns were removed as they were not required for the research.

Then a subset of the data was created to include records only after the year 1950 as to get a reasonably new observation on Climate Change.

AverageTemperature	Country	Latitude	Longitude	Year	Month
26.437	Nigeria	5.63	8.07	1950	1
27.105	Nigeria	5.63	8.07	1950	2
28.610	Nigeria	5.63	8.07	1950	3
28.176	Nigeria	5.63	8.07	1950	4
27.358	Nigeria	5.63	8.07	1950	5
26.700	Nigeria	5.63	8.07	1950	6
25.432	Nigeria	5.63	8.07	1950	7
25.490	Nigeria	5.63	8.07	1950	8
25.651	Nigeria	5.63	8.07	1950	9
26.025	Nigeria	5.63	8.07	1950	10
26.643	Nigeria	5.63	8.07	1950	11
26.800	Nigeria	5.63	8.07	1950	12

Table 2 Dataset after Cleansing

4.Linear Regression and R programming

For the purpose of data manipulation, R language was chosen for this research as it offers a variety of options to analyse and predict data.

In order to predict the temperature of any country over the coming years or any year in particular, a Linear Regression model was established.

Linear Regression is a modelling technique where a linear relationship is found between a target or the variable to be predicted and one or more predictors. The said relationship is statistical in nature and not deterministic. [5]

A relation between two or more variables is called deterministic if one variable can be accurately expressed by another variable or a combination of other variables for example, temperature can be expressed in either Celsius or Fahrenheits where we can obtain Fahrenheit if we have Celsius and vice-a-versa. Hence there is an exact relation between Celsius and Fahrenheit, this type of relation is known as deterministic relation. [6]

Whereas in Linear Regression we have a statistical relation between the target and predictors i.e. there isn't an exact relation but we obtain a best fit. A best fit is when the total prediction error for all the data points is as little as possible.

Given a dataset $\{y_i, x_{i1}, x_{i2}, x_{i3} \dots \dots .. x_{ip}\}\binom{n}{i=1}$ of n statistical units, linear regression assumes a linear relation between y_i and *X*, where $X = \{x_{i1}, x_{i2} \dots \dots . x_{ip}\}\binom{n}{i=1}$.

Thus, the model takes the form of $y_i = b + w_{i1}x_{i1} + w_{i2}x_{i2} \dots \dots$, where b represents a bias and $\{w_{i1}, w_{i2}, w_{i3} \dots .\}$ represent the weight of corresponding *x* values.

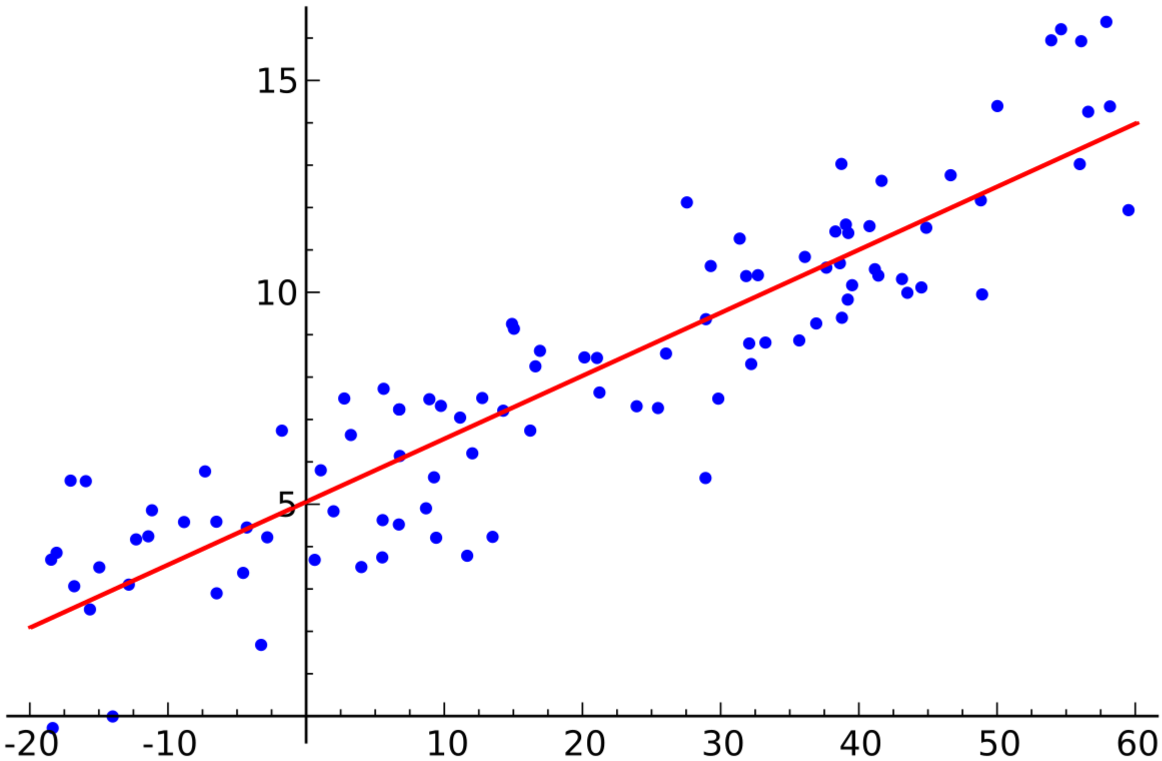

Figure 1 Random data points and their linear regression [7]

As a general practice it is recommended to divide the dataset into training and testing datasets. After the data cleansing 585,238 observations were left of which 75% were taken as training dataset i.e. 438,928 observations and 25% or 146,310 observations were taken as testing dataset.

In our model AverageTemperature was predicted using Country, Latitude, Longitude, Year and Month as predictors. A simple model was created using the *lm()* function in R which takes the AverageTemperature as the first argument and ~. as the next argument representing all of the remaining columns.

In R a *prediction()* function exists which can predict values, in this case AverageTemperature, based on all the other variables provided.

The summary of the model consists of various sections which help us identify and determine the relation. These values include-

1. Residuals – Residuals are essentially the difference between the actual records and the values that the model predicted. The Residual section is divided into 5 parts, Min or Minimum value, 1Q or the first quantile, Median Value, 3Q or the third quantile and Max or Maximum value. [8]
2. Coefficients – The coefficients are divided into two types, one is the intercept or the *b* value, and other are the terms which help us in determining the slope of the regression line. Coefficient is further expressed in four columns. [9]
    1. Estimate – It is the value of the slope calculated by the regression model.
    2. Standard Error – It is the measure of the variability of the estimate of the coefficient. Here lower is better (relative to the estimate value) and Std. Error is generally an order of magnitude less than the estimate.
    3. t-value – It measures whether the corresponding variable is important to the model or not.
    4. Pr(>|t|) – It is known as the variable p-value. It represents the probability of whether the corresponding variable is NOT relevant to the model.
3. Significance Stars – The asterisks following the Coefficient values are known as the significance stars which imply how significant the corresponding variable is to the model. *** represents high significance, * represents low significance and nothing would represent that the variable is not significant to the model. [9]
4. Residual Standard Error – It is the standard deviation of the Residuals.
5. R-Squared – Measure of how well the model fits with 1 being the best.
6. F-statistic – It performs F-Test on the model i.e. it takes a model with fewer parameters than the current model and checks whether it performs better or worse. Generally, a model with more parameters should perform better. [9]

5.Conclusion

Using basic plotting techniques available in R we can conclude that there has been high rise in temperature on the Earth's surface.

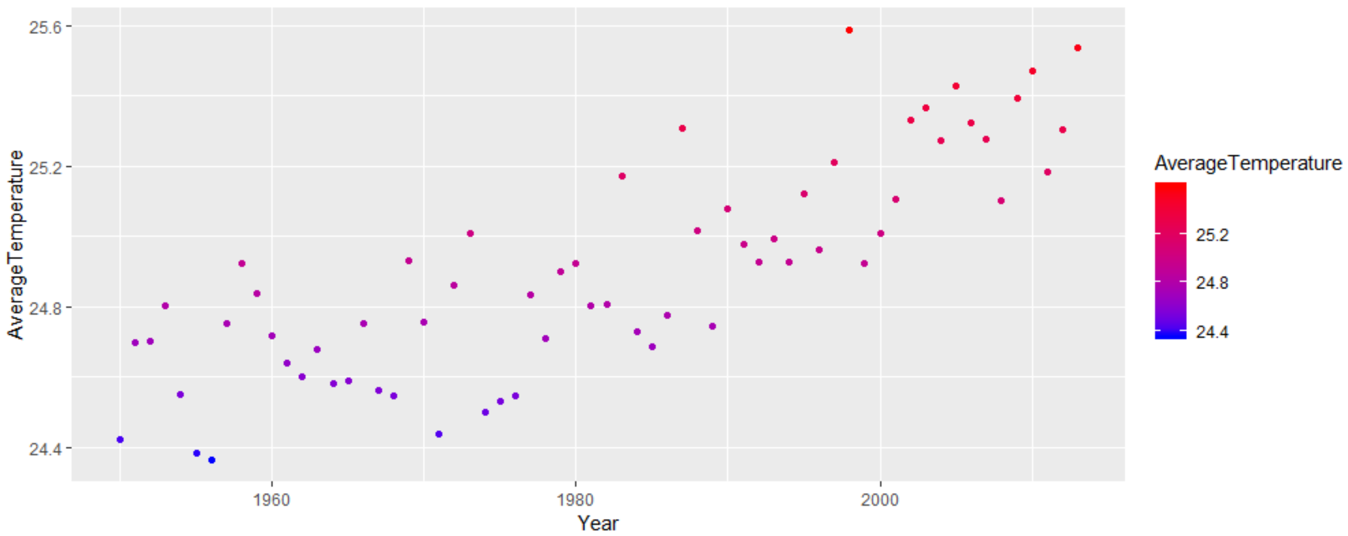

Figure 2 Scatter Plot of Mean Avg. Temp vs. Year

A scatter plot of Mean of Average Temperatures calculated over the year of the equitorial as well as polar regions can be observed here which clearly shows a sharp rise in temperature.

The following graph represents the Mean Average Temperatures calculated over the year of only the Polar regions and there can be observed a rise of 1°C-2°C which might not seem drastic but in reality, is. According to a report by the US Army Corps of Engineering [10], the thickness of ice can be reduced by 3.7mm per warming degree (Celsius). Therefore, a temperature of 1°C sustained for 24 hours would reduce the thickness of ice by 3.7mm, which when considered for a time on the scale of years would drastically reduce the ice present in the Polar regions.

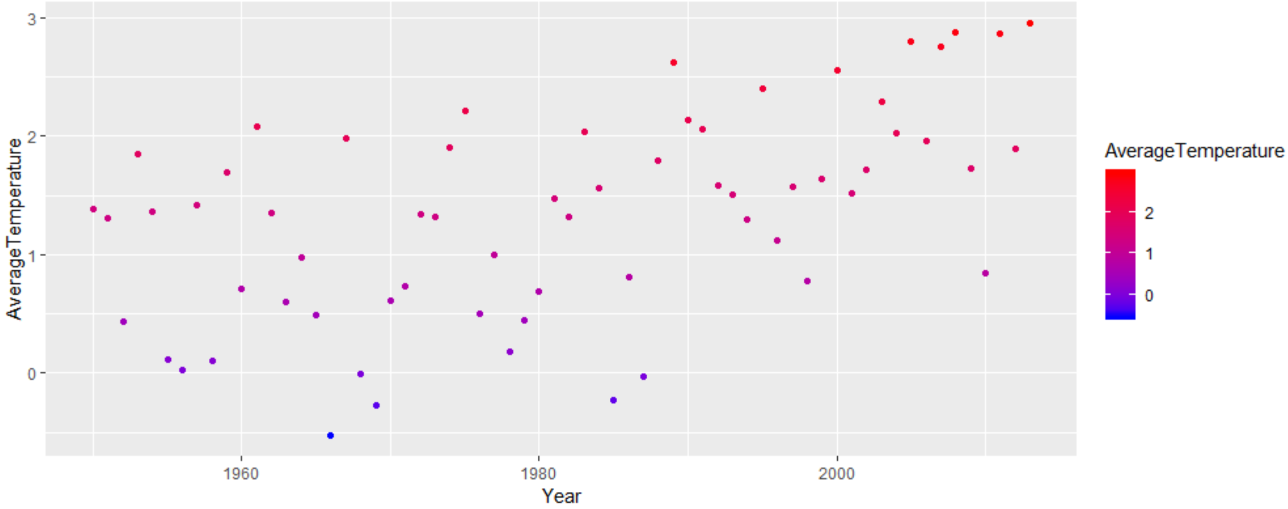

Figure 3 Average Temperature vs Year for Polar Regions

6.Result

1. Residuals

Min	1Q	Median	3Q	Max
-37.753	-1.117	0.023	1.005	22.742

Table 3 Residual Table

2. Residual Standard Error, R-Squared and F-statistic

```
Residual standard error: 2.766 on 438860 degrees of freedom
Multiple R-squared:  0.7518,    Adjusted R-squared:  0.7518
F-statistic: 1.984e+04 on 67 and 438860 DF,  p-value: < 2.2e-16
```

Figure 4 Residual Standard Error and F-Statistic

3. Coefficients

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	-3.210e+00	4.540e-01	-7.070	1.56e-12	***
CountryBenin	4.729e+00	7.098e-02	66.616	< 2e-16	***
CountryBrazil	4.659e+00	5.975e-02	77.964	< 2e-16	***
CountryBurkina Faso	5.721e+00	9.751e-02	58.664	< 2e-16	***
CountryBurma	7.046e+00	1.223e-01	57.629	< 2e-16	***
CountryBurundi	-9.089e-01	1.289e-01	-7.052	1.77e-12	***
CountryCÃ´te D'Ivoire	3.595e+00	6.171e-02	58.254	< 2e-16	***
CountryCambodia	7.777e+00	1.496e-01	51.999	< 2e-16	***
CountryCameroon	2.518e+00	6.009e-02	41.901	< 2e-16	***
CountryCentral African Republic	2.914e+00	1.297e-01	22.464	< 2e-16	***
CountryChad	5.289e+00	9.653e-02	54.786	< 2e-16	***
CountryColombia	2.416e+00	7.675e-02	31.473	< 2e-16	***
CountryCongo	1.745e+00	1.269e-01	13.748	< 2e-16	***
CountryCongo (Democratic Republic Of The)	1.377e+00	5.855e-02	23.517	< 2e-16	***
CountryCosta Rica	5.017e+00	1.413e-01	35.507	< 2e-16	***
CountryDjibouti	8.240e+00	1.288e-01	63.951	< 2e-16	***
CountryEcuador	2.213e-01	8.483e-02	2.609	0.00907	**
CountryEl Salvador	5.773e+00	9.635e-02	59.917	< 2e-16	***
CountryEquatorial Guinea	2.894e+00	9.839e-02	29.414	< 2e-16	***
CountryEthiopia	-6.487e-01	6.768e-02	-9.584	< 2e-16	***
CountryFinland	-1.560e+01	1.188e-01	-131.315	< 2e-16	***
CountryGabon	3.491e+00	9.879e-02	35.334	< 2e-16	***
CountryGambia	3.620e+00	9.683e-02	37.387	< 2e-16	***
CountryGhana	3.937e+00	6.613e-02	59.539	< 2e-16	***
CountryGuinea	3.553e+00	7.736e-02	45.924	< 2e-16	***
CountryGuinea Bissau	5.358e+00	1.265e-01	42.364	< 2e-16	***
CountryGuyana	5.917e+00	1.316e-01	44.970	< 2e-16	***
CountryHonduras	4.339e+00	1.415e-01	30.674	< 2e-16	***
CountryIceland	-1.772e+01	1.604e-01	-110.526	< 2e-16	***
CountryIndia	6.667e+00	7.780e-02	85.684	< 2e-16	***
CountryIndonesia	6.716e+00	9.891e-02	67.898	< 2e-16	***
CountryKenya	-1.393e+00	7.467e-02	-18.662	< 2e-16	***
CountryLiberia	3.681e+00	1.277e-01	28.839	< 2e-16	***
CountryMalawi	-1.234e+00	1.262e-01	-9.779	< 2e-16	***
CountryMalaysia	6.961e+00	9.799e-02	71.043	< 2e-16	***

Table 4 Coefficient Table (1 of 2)

CountryMali	5.063e+00	9.717e-02	52.100	< 2e-16	***
CountryMozambique	2.362e+00	9.894e-02	23.871	< 2e-16	***
CountryNicaragua	7.170e+00	9.988e-02	71.778	< 2e-16	***
CountryNiger	6.140e+00	8.421e-02	72.906	< 2e-16	***
CountryNigeria	4.410e+00	5.351e-02	82.422	< 2e-16	***
CountryNorway	-1.691e+01	1.253e-01	-135.037	< 2e-16	***
CountryPanama	6.767e+00	1.119e-01	60.457	< 2e-16	***
CountryPapua New Guinea	6.617e+00	1.724e-01	38.376	< 2e-16	***
CountryPeru	-2.437e+00	8.028e-02	-30.351	< 2e-16	***
CountryPhilippines	7.943e+00	1.103e-01	71.983	< 2e-16	***
CountryRussia	-1.954e+01	1.194e-01	-163.566	< 2e-16	***
CountryRwanda	-2.699e+00	1.293e-01	-20.879	< 2e-16	***
CountrySenegal	4.598e+00	8.433e-02	54.520	< 2e-16	***
CountrySierra Leone	3.753e+00	9.687e-02	38.747	< 2e-16	***
CountrySingapore	6.907e+00	1.481e-01	46.629	< 2e-16	***
CountrySomalia	5.153e+00	7.811e-02	65.975	< 2e-16	***
CountrySri Lanka	6.728e+00	8.692e-02	77.407	< 2e-16	***
CountrySudan	6.276e+00	6.917e-02	90.729	< 2e-16	***
CountrySuriname	5.698e+00	1.327e-01	42.935	< 2e-16	***
CountrySweden	-1.534e+01	1.332e-01	-115.153	< 2e-16	***
CountryTanzania	1.171e+00	6.246e-02	18.747	< 2e-16	***
CountryThailand	7.847e+00	9.670e-02	81.146	< 2e-16	***
CountryTogo	4.373e+00	9.712e-02	45.025	< 2e-16	***
CountryUganda	2.300e+00	8.783e-02	26.194	< 2e-16	***
CountryUnited States	-1.825e+01	2.002e-01	-91.182	< 2e-16	***
CountryVenezuela	5.231e+00	7.166e-02	73.006	< 2e-16	***
CountryVietnam	7.126e+00	1.002e-01	71.113	< 2e-16	***
CountryYemen	4.839e+00	1.007e-01	48.041	< 2e-16	***
CountryZambia	-6.506e-01	7.076e-02	-9.194	< 2e-16	***
Latitude	-4.032e-02	1.873e-03	-21.529	< 2e-16	***
Longitude	-2.797e-02	8.839e-04	-31.643	< 2e-16	***
Year	1.336e-02	2.272e-04	58.792	< 2e-16	***
Month	-4.125e-02	1.210e-03	-34.089	< 2e-16	***

Table 5 Coefficient Table (2 of 2)

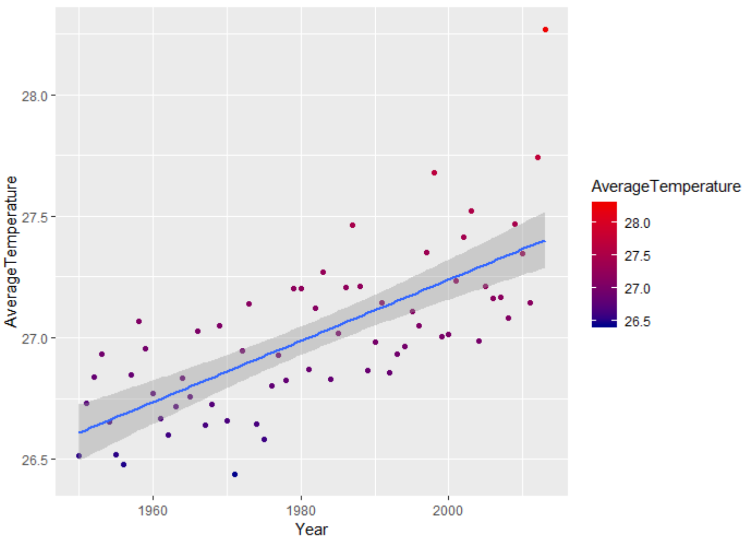

Figure 5a Plot of Average Temperature vs. Year for India (1950-2013)

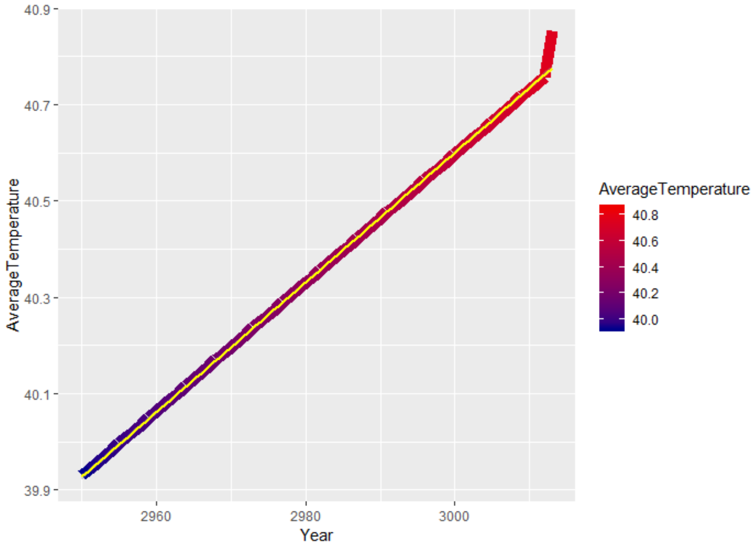

Figure 5b Plot of Average Temperature vs. Year for India (2950-3013)

Using the above created model plots were generated for the country of India. The figure on the left shows the temperature from years 1950-2013 with a regression line along with a 95% Confidence Region. The figure on the right uses the model to predict temperatures for the year 2950-3013 for the country of India with a regression line in yellow. It can be observed that the minimum temperature on the left figure was observed as approximately 26.5°C whereas with the figure on the right it can be observed that the minimum temperature is approximately 39.9°C.

7.References

[1] Caryl-Sue, “Latitude,” National Geographic, 6 November 2012. [Online]. Available: https://www.nationalgeographic.org/encyclopedia/latitude/.

[2] “What is Global Warming?,” National Geographic, [Online]. Available: https://www.nationalgeographic.com/environment/global-warming/global-warming-overview/.

[3] BerkleyEarth, “Climate Change: Earth Surface Temperature Data,” 2016. [Online]. Available: https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data/home.

[4] J. Chambers, “What is R?,” The R Foundation, [Online]. Available: https://www.r-project.org/about.html.

[5] D. M. Lane, “Introduction to Linear Regression,” Onlinestatbook, [Online]. Available: http://onlinestatbook.com/2/regression/intro.html.

[6] Stephanie, “Types of Relationships STAT 414/415,” Statistics How to, 21 October 2016. [Online]. Available: https://www.statisticshowto.datasciencecentral.com/deterministic/.

[7] Sewaqu, “Linear Regression,” Wikipedia, 5 November 2012. [Online]. Available: https://en.wikipedia.org/wiki/Linear_regression#/media/File:Linear_regression.svg.

[8] "QUICK GUIDE: INTERPRETING SIMPLE LINEAR MODEL OUTPUT IN R," Felipe Rego, 23 October 2015. [Online]. Available: https://feliperego.github.io/blog/2015/10/23/Interpreting-Model-Output-In-R.

[9] yhat, "Fitting & Interpreting Linear Models in R," 18 May 2013. [Online]. Available: http://blog.yhat.com/posts/r-lm-summary.html.

[10] "How Ice Melts," thebowens, [Online]. Available: http://thebowens.name/HowIceMelts.php.