INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES TOULOUSE

ISS Big Data Analytics Report

The purpose of this report is to show the skills that I have acquired and the concepts those I understood as a result of the courses and the TP sessions. The report is of 5 pages.

Last-Name 1: Xiong

First-Name 1: Xuantang

Last-Name 2: SELVARAJ

First-Name 2: MOHAN PRABHU

Date

18-Jan-2019

Introduction

Effects of climate change has already started leaving impact on the environment. All the predictions that scientists came up in the past has started occurring in recent decades among which the rise in temperature is the key issue leading to various other effects such as loss of sea ice, rise in sea level and more intense heat waves.

The major aspect of long-term rise in the average temperature of the earth's climate system is termed as Global warming. Thanks to Kaggle and Berkeley Earth for the data collection of global temperatures since 1750. The dataset ***"Climate Change: Earth Surface Temperature Data"*** consists of Global Temperatures together with the Global Land Temperatures by geographical divisions such as City, State and Country.

Hypothesis

In this report, we explore the dataset, analyze and attempt to confirm that the key factor for increase in temperature across the globe is due to global warming & climatic changes across the globe.

Analysis Steps

1) **Data cleaning and preprocessing**

First, we selected the major countries, which includes "China", "United States", "France", "Japan","Germany", "India","United Kingdom", "Italy", "Brazil", "Canada". And then we delete the missing data. As shown in the figure below, each set of data includes the time when the data was recorded, and the country in which the data is collected and the average temperature.

year <int>	month <int>	Country <chr>	date <date>	avg_Temp <dbl>
1743	11	France	1743-11-01	10.203
1743	11	Germany	1743-11-01	5.468
1743	11	Italy	1743-11-01	9.538
1743	11	United Kingdom	1743-11-01	7.086
1744	4	France	1744-04-01	13.190
1744	4	Germany	1744-04-01	8.438
1744	4	Italy	1744-04-01	12.759
1744	4	United Kingdom	1744-04-01	7.619
1744	5	France	1744-05-01	14.133
1744	5	Germany	1744-05-01	11.498

2) Data observation

How is the temperature variation of the selected countries during a year?

First, we can look at the distribution of the data to determine if there is a problem with the data. We set the abscissa to the month and the ordinate to the average temperature to observe how the average temperature of different countries changes with the month during the year.

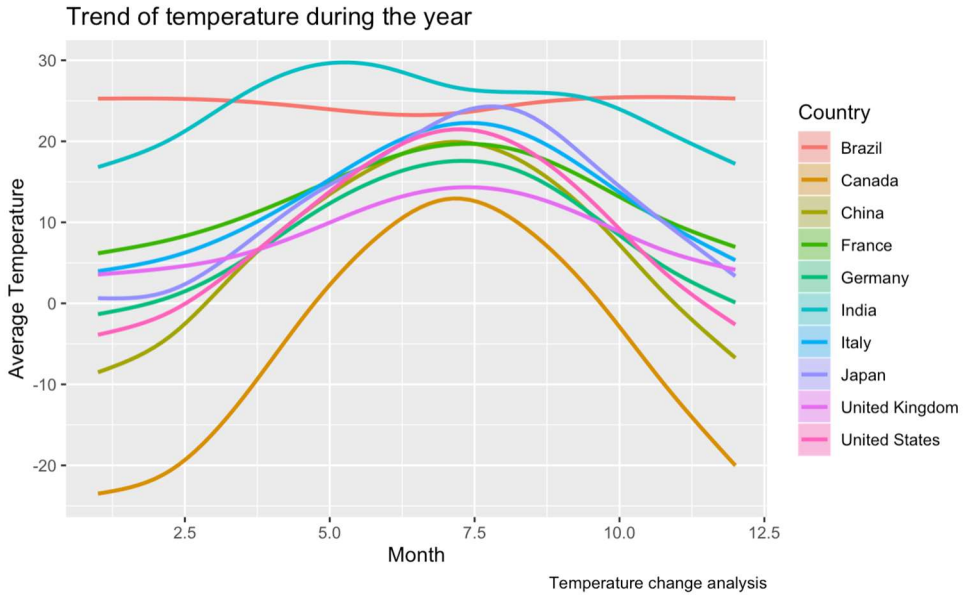


Figure-1

The result is shown in the **figure-1**, which fits our expectations well. As you can see, Brazil is in the southern hemisphere and 80% of the land is located in the tropics, so the annual temperature is greater than 25 degrees and the temperature is lowest in July. For France, Japan, and the United States, whose main land areas are located in the northern temperate zone, January has the lowest temperature, and July has the highest temperature, which varies widely. As for Canada, where part of the land area is located in the northern cold zone, the average temperature is less than zero at half the time of the year.

Temperature Trend Analysis

How is the change in temperature of the selected countries during the last 300 years?

Now we change the horizontal axis to year, and the vertical axis is still the average temperature to observe the changes in the average temperature of different countries in the past two hundred years**.**

As shown in the **figure-2**, the solid line represents the average temperature, and the transparent part around the solid line represents the degree of dispersion of the data. Similar to the above analysis, the countries in the tropics have high average temperatures and small temperature changes. As the latitude increases, the temperature decreases and the range of change becomes larger.

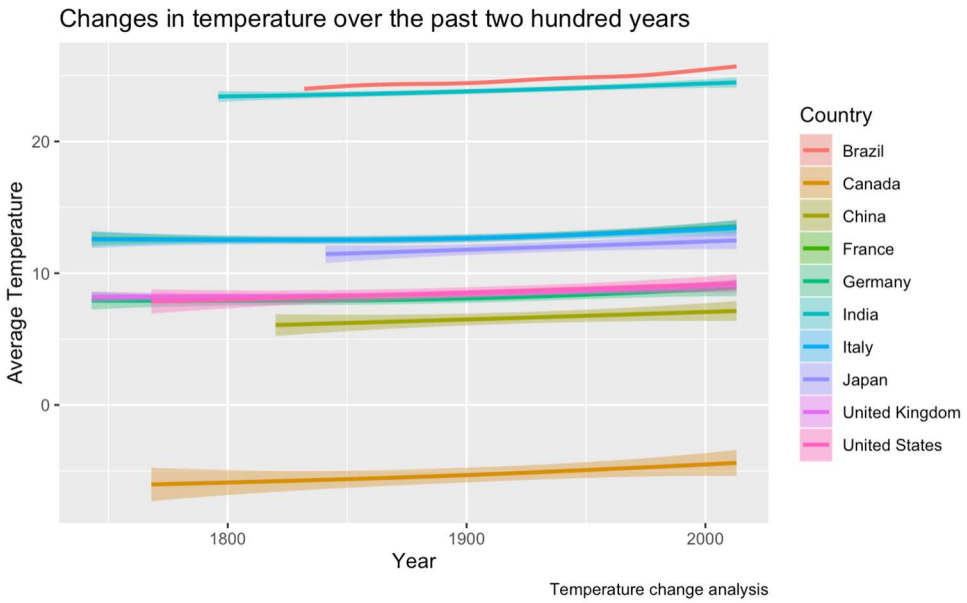


Figure-2

Although not obvious in the graph above, it can still be seen that the average temperature of each country has increased over the past two hundred years. To make the observations clearer, from **Figure-3**, we observe the average temperature change in just one country - Japan.

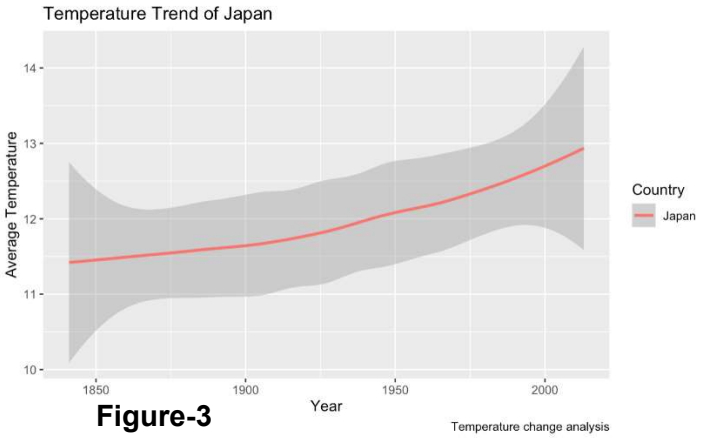
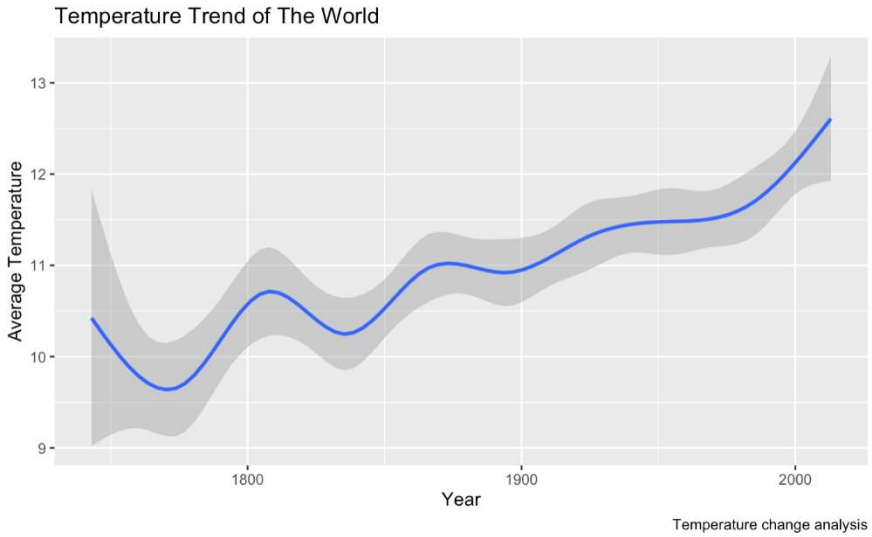


Figure-3

We can see that the average temperature in Japan rose from 11.5 to 13 degrees. Especially after 1900, the rate of temperature rise accelerated significantly.

How can we relate the trend in temperature change of those selected countries to the global temperature change?

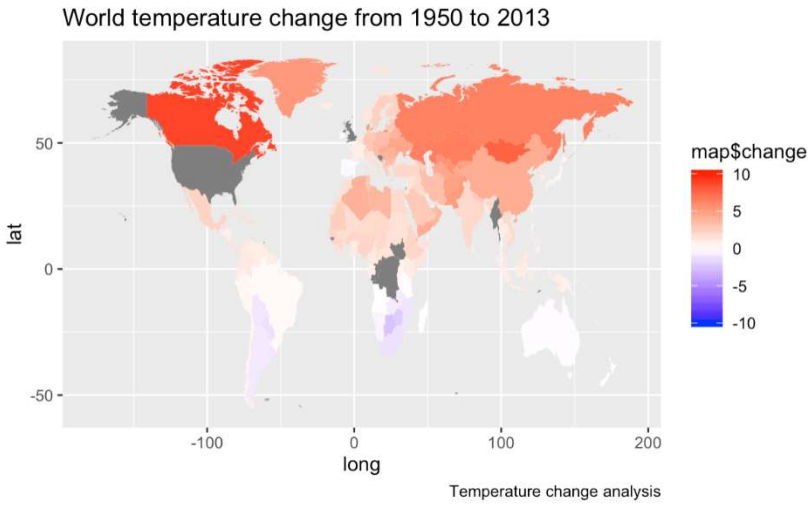


Next, we observe the global temperature change. We fit the temperature data of the above countries into a curve to get the **figure-4.**

Figure-4

From the **figure-4**, we can see that the temperature changes with the increase of time, but the overall trend is rising, which has increased by about 2 degrees in the past two hundred years.

What is the rate of change in global temperature based on data from the last few decades?



Finally, we calculated the difference between the average temperature of 2013 and 1850 in each country in the world, and made the **figure-5** based on the change in temperature. The darker the color, the more the temperature rises. The color changes towards dark in most of the countries which is evidence of global warming.

Figure-5

To add more evidence, we have also used 'GlobalTemperatures' from the dataset as found from **figure-6** to see how the overall land temperature was increasing over the years. By this way, we are also making sure of the accuracy of the data.

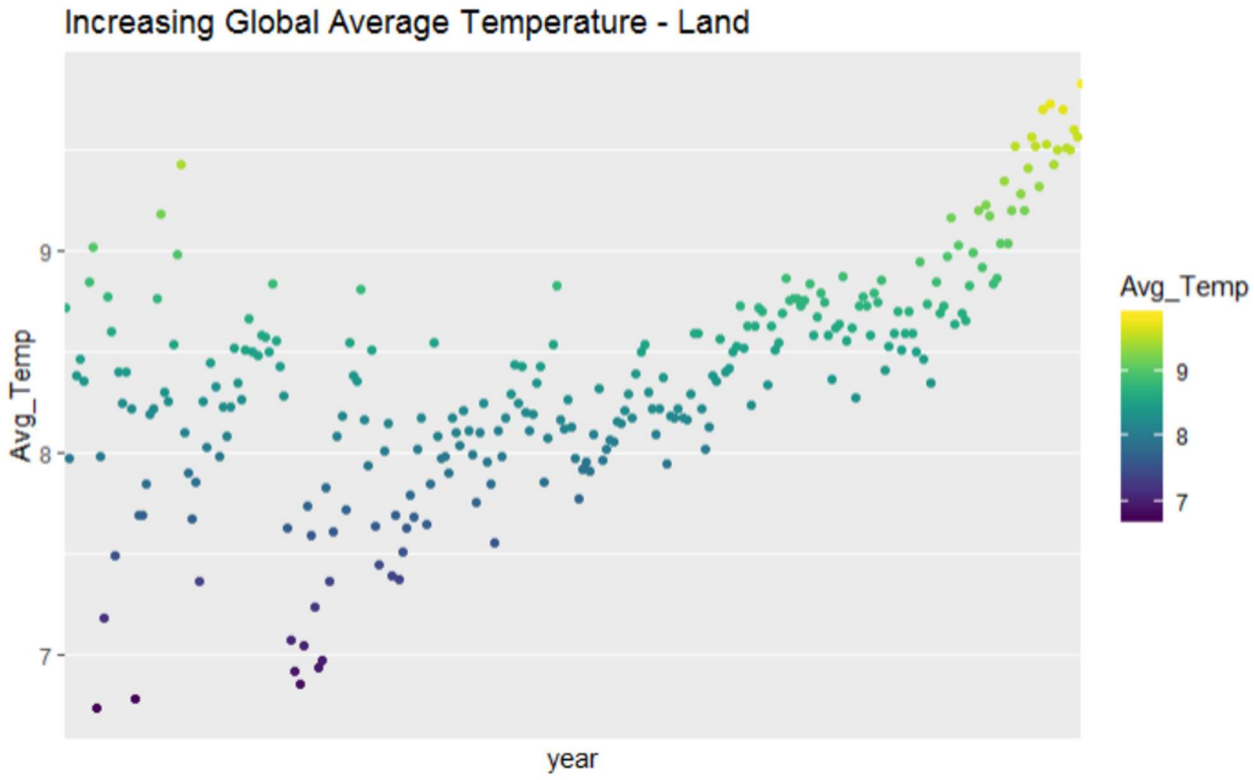


Figure-6

To be more precise about the increasing temperature, we have filtered the average temperatures that are above 6.5. As we see from **figure-6**, the land temperature is clearly steadily increasing over the years. This shows to be the effect of the global warming.

Conclusion

In our exploration of the dataset to validate our hypothesis that the temperature rise is due to global warming, we have used 2 datasets to improve our accuracy.

1) GlobalLandTemperatureByCountry
2) GlobalTemperatures

We have used the inferences/observations made from analysis of data on 10 geographically dispersed countries to improve our accuracy of predictions.
We started with exploring the data of the selected countries to find out how is the temperature variation over a period of a year. Followed by, we learnt the temperature change trend for past 300 years. From the trend analysis results, we tried to relate it at a global scale. We inferred that the temperature has risen by 2 degree Celsius during the last 200 years. Further, the difference between the latest temperature average as of 2013 and temperature average as of 1850 clearly show the temperature rise. The land temperature results does reinforce the evidences.

Therefore, as per the inferences made from data analysis results, we believe that the acceleration of temperature is due to the global warming in our planet and the climatic changes.