Udacity Data Analysis Degree Project 01: Weather Trends

Simon Thornewill von Essen

***Abstract*—Structured Query Language (SQL) was used to download Comma Separated Values (CSV) files of global yearly average temperatures and those from Berlin. The downloaded data was then analyzed using Python. It was found that global temperatures are rising exponentially over recent years and are unlikely to be anomalous. It has been hypothesised that the trend in rising temperature is due to increased carbon dioxide in the atmosphere since this rise in temperature takes place roughly at the same time as the industrial revolution.**

I. Introduction

THE discussion on Global warming has been a topic of fierce debate in politics in recent years. The debate is largely centered around whether or not global temperatures are spiking, what mechanisms drive this change and what is to be done about it. In this publication, global weather data has been analyzed to determine whether or not temperatures are climbing in Europe compared to the global average. The mechanisms behind global warming and what can be done to address it will be outside of the discussion scope.

II. Methods

The data which is used in this analysis was collected from a database using Structures Query Language (SQL) and was subsequently processed using various libraries in Python. (i.e. Pandas, Numpy, Matplotlib, etc.) The scripts that contain step by step information about how it was written and can be found at the following url [1]. After a simple visualisation was created, the data was transformed using the Numpy library to fit a seven year moving average in order to make the data less noisy.

A. SQL

The data was exported using the following code;

```
SELECT *
FROM city_data
WHERE city = 'Berlin';

Select *
FROM global_data;
```

The resulting CSV files were saved to the github repository.

B. Python

The CSV files were imported into Python using Pandas. Upon initial investigation of the data it was found that there were missing values before the year 1950 for the Berlin dataset in addition to missing years past 2013 for the global dataset. These years were dropped in order to give the data the same shape for analysis. Cleaning the data in this way meant that no imputation was necessary.

Afterwards, the data was transformed into a 7 year moving average with the following Python algorithm where the variable "moving" is the amount of years used for the moving average:

```
moving = 7
average = []
TempMovingAverageDF = AvgTempDF

for i in range(len(AvgTempDF)):
    average.append(
        np.average(
        AvgTempDF.iloc[(i-moving):i, 1]))

for i in range(len(AvgTempDF)):
        TempMovingAverageDF.iloc[i, 1] = average[i]
```

It should be noted that this is an abstracted algorithm to bring the point of the code across to the reader and is not an exact copy of the code, for more information see the github repository. After the data was cleaned and transformed it was plotted using the Matplotlib library.

III. Results and Discussion

It was found that the average global temperature is increasing over time as can be seen in figure 1 where the average global temperature, shown in orange, has high fluctuations in the mid 1700s which slowly stabilizes into an upward trend. This upward trend coincides with the start of the industrial revolution where humans began to harness fossil fuels as a form of energy causing emissions of greenhouse gases.

The variance in Berlin's temperature changes widely relative to the average global temperature. This difference in variance is due to the fact that the global temperatures cover a much wider than a single city and includes various biomes with differing temperatures due to change in atmospheric composition, the type of precipitation, latitude/longitude, altitude and sunlight exposure. Covering such a large and varied area means that it is very difficult to change the average values over time. The average temperature is also driven downward relative to Berlin for the same reason as Berlin does not have icy tundras and arctic areas. Berlin also does not contain scorching deserts ans so the conclusion is that on average berlin is warmer than most places on the planet which might not feel intuitive until one accounts for the antarctic,

[1] https://github.com/SThornewillvE/Udacity-Project—Exploring-Weather-Trends

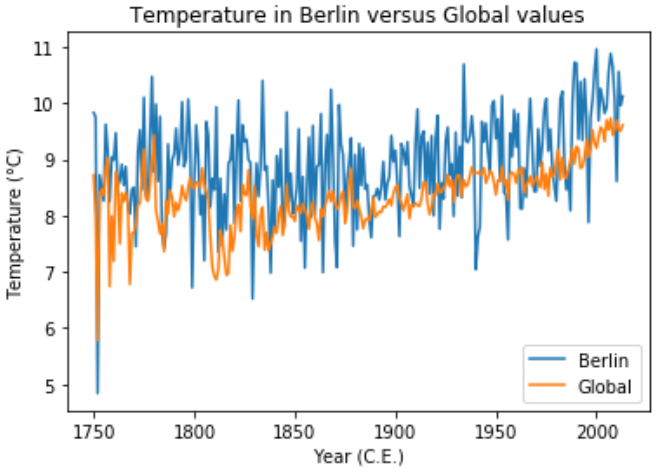

Fig. 1. Average temperature in Celsius over time in years before data transformations with Berlin shown in blue and Global values shown in orange.

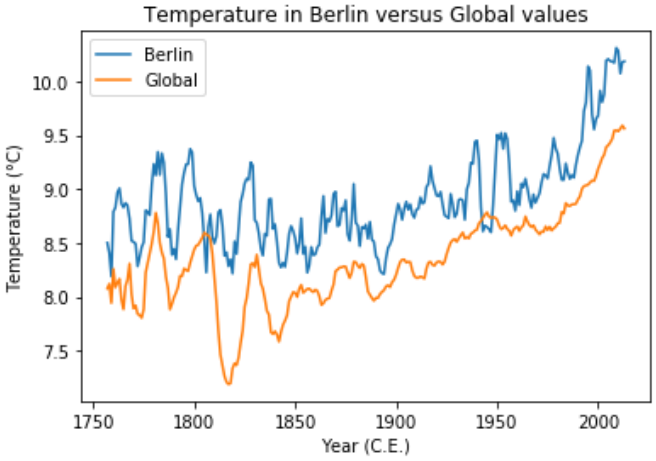

Fig. 2. Seven year moving average of the temperature in Berlin (shown in blue) against the corresponding Global values.

oceans and other such regions which contribute a large area to the world's surface area relative to places hotter than Berlin.

When the data was transformed to a 7 year moving average, as can be seen by figure 2, the trends become more obvious because they are less obscured by noise and become more consistent as they become less prone to outliers. As seen with the consistent global temperatures, taking an average of sections of the data mean that outliers are washed away by more normal values. However, if temperatures rise uniformly then this will result in a much higher average.

It is well known to statisticians that mean is a metric that is prone to outliers so the data can likely be made much less noisy by using a moving mode instead of a moving average because the mode is a metric of central tendency which is known to be less prone to outliers making likely to be a better candidate than the mean.

The same trend of Berlin having a higher temperature and the presence of large fluctuations in temperatures during the late 1700s is still evident. However, the exponential rise of temperature over time becomes much easier to spot as the temperature consistently rises as the years progress. This suggests that the forces which are driving climbing temperatures are compounding themselves. This could coincide with the rising carbon dioxide emissions per capita as the global population increases exponentially for instance. However, the direct cause cannot easily be inferred from this data alone. Until the process causally driving this change is determined, no recommendation can be made.

This evidence is cause for alarm because there are many systems on the planet which are dependent on temperature from atmospheric chemistry which determines the quality of air to dissolved carbon dioxide in the water, in addition to the suspicion that this is driven by carbon dioxide, which can decrease the acidity of various ocean bodies which can have knock on effects for aquatic life. Organisms are part of a large network and so the effects of these changes could become very widespread. Hence, if the rising temperature should be a problem then it should also be able to see these effects across various scientific disciplines. It is worth noting that the rising temperature of the globe by roughly 1.5C is not trivial and requires a large amount of energy to do as discussed earlier in this document. Hence, whatever is causing this ought to be a proportionally large process.

IV. Conclusion

In conclusion, there is evidence to suggest that the temperature of the globe has been rising over the past century at a rate which seems to be exponential. This could have a profound effect on global ecosystems as many of them are dependent on temperature in order to function naturally. More research is required in order to determine where this change in temperature comes from and what should be done about it if at all and what this could mean for the world.