Project 1

Explore Weather Trends Analysis

Submitted By Jacky Wang

Submitted to Udacity reviewer

In partial fulfillment of the requirements for
Data Analyst Nanodegree, Udacity

April 20, 2021

Introduction

The purpose of this report is to analyze the temperature data from both a global and a local scale and to create a visualization to present the findings. The global temperature change has been a hot topic. As people debate about its validity, and scientists try to understand what causes the change, analysing the temperature data around the world has always been an important step.

The steps taken in the analysis from how the data is prepared to how the visulization is created will be explained. Several charts are created during the analysis. The trends, similarities and differences between the world's average and the city's average will also be analyzed based on the charts, including the calculation of the correlation coefficient. A linear equation is made to estimate the local average temperature based on the global average temperature. The local city has to be in the closest big city to where I live, therefore the chosen city in this report is Ottawa, Canada.

Analysis Procedure

Prepare the data

The data is taken from SQL database Schema. The language used is SQL script. There are three tables in the database, city_list, city_data, and global_data. First, I try to locate if the Ottawa is in the city_list, and save it as a CSV file for further analysis. The steps to extract the data from the database are the following.

Use this script to get the table structure:

```
SELECT * FROM city_list;
```

Then use this script to see if Ottawa is in the table:

```
SELECT * FROM city_list

WHERE city = 'Ottawa';
```

Use this script to get the structure of table city_data:

```
SELECT * FROM city_data;
```

Then use this script to get the data from city_data:

```
SELECT year, avg_temp

FROM city_data

WHERE city = 'Ottawa' and country = 'Canada';
```

Download and save it as a CSV file.

And similarly, use the following script to get information from global_data and save it as a CSV file.

```
SELECT * FROM global_data;
```

Data analysis

The tool used in this analysis is Microsoft Excel 2016. The steps taken for the analysis are outlined here:

Create a .xlsx file for analysis

Open both csv files in Excel and copy the content of one csv file to another csv file. In particur, I copied Ottawa's data (just the temperature data) into Global's data side by side and Save As a xlsx file in order to save the analysis.

Calculation of moving average for both data

First of all, in order to compare the data precisely, I chose the same year range, namely 1750-2013, from both data. So I truncated the data out of this range, relabelled the columns Global and Ottawa, and created two new columns, MV-Global and MV-Ottawa. Then I calculated the moving average for each of column Global and Ottawa and put in MV-Global and MV-Ottawa respectively. I experimented with different year range, 5, 10, and 20 years, to see how they go, and ended up choosing 10 years. The reason was that the 5-year moving average did not smooth out the line quite well, whereas the 20-year moing average smoothed out the line too well. As the result, both data starts from year 1759 to 2013.

Analysis Process and Visualizations

Figure 1 shows the result of analysis for temperature trends globally and in Ottawa. This chart is a line chart created by taking data from both MV-Global and MV-Ottawa. The green line is the global temeperature trend, and the brown line is the Ottawa temperature trend. The x-axis is the year, and the y-axis is the moving average temperature.

In order to understand more about this relation, I also calculated the correlation coefficient. However, the correlation coefficient was calcuated from the original average temperature rather than from moving average, using built-in correl() function in Excel. The correlation coefficient is 0.718.

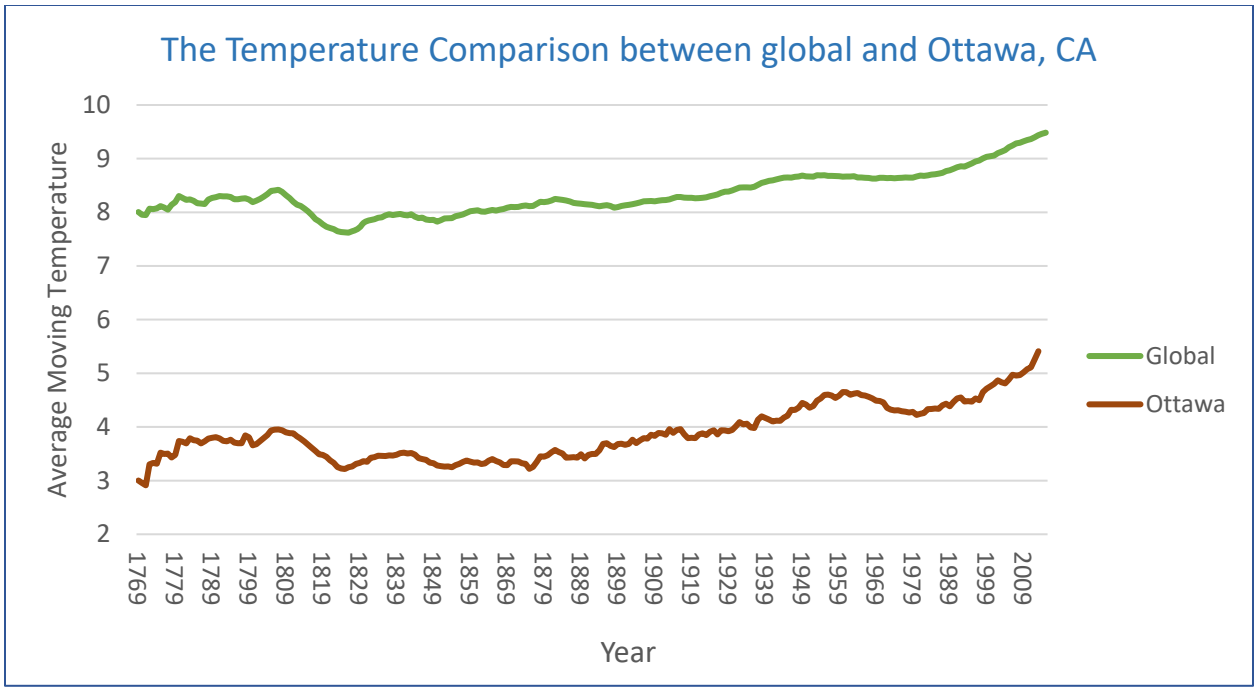

Figure 1: Temperature Global versus Ottawa. The temperature trends for both global and city Ottawa (CA) were displayed

To visually see this trend, I also created two scatter plots from the data, one from moving averages, the other from original temperature data, along with their linear equations and R-squared measures as shown in Figure 2 and 3. Figure 2 was obtained from calculated moving average temperature, and Figure 3 was obtained from raw original temperatures. They were created to understand if they differ too much.

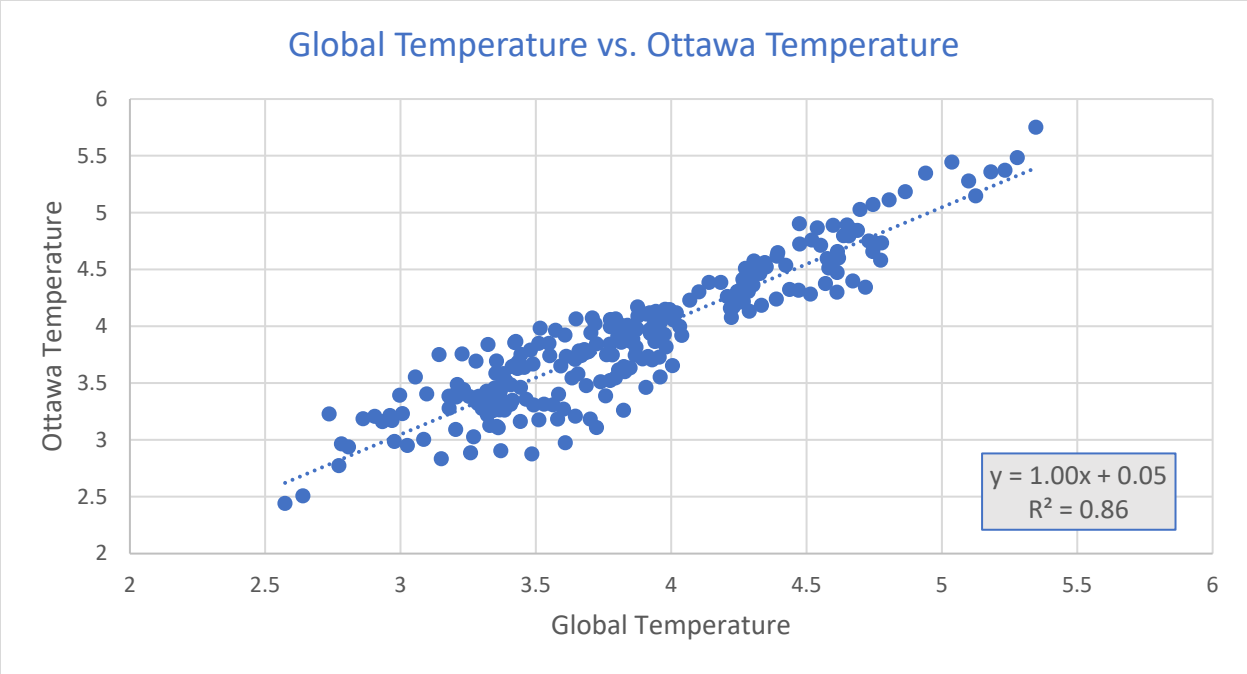

Figure 2: Linear Relation. The trend between two temperatures shows there is a linear relationship between each other (using moving average)

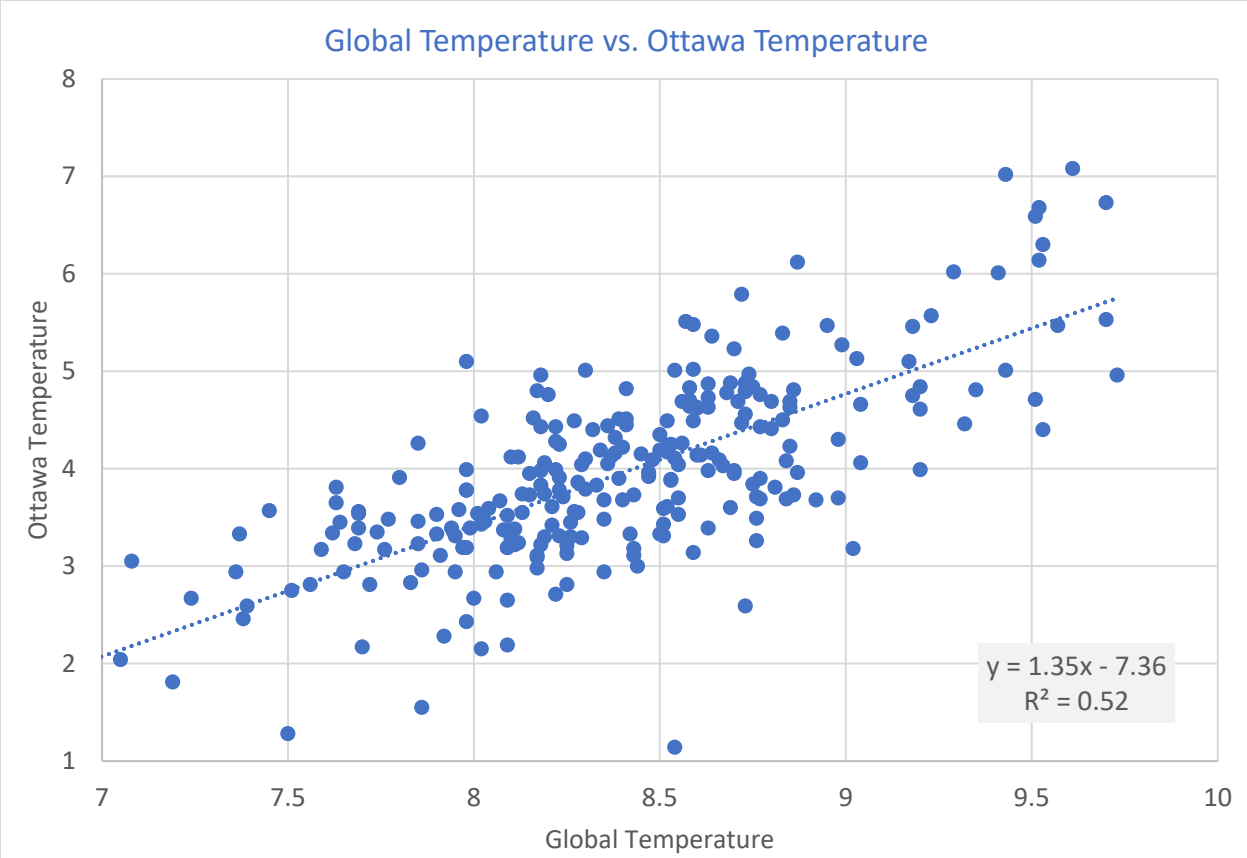

Figure 3: Linear Relation. The trend between two temperatures shows there is a linear relationship between each other (using original temperature average)

Observations

1. According to the calculated correlation coefficient, 0.718, the data is positive and highly correlatated, which explains the strong similarities in the trends to a consistent higher temperature over the years either globally or locally in Ottawa.
2. According to Figure 3, the calculated R-Squared value is 0.52, which indicates that the linear equation obtained does not explain very well the relationship between global temperature and local temperature.
3. According to Figure 1, The temperature in Ottawa has been lower compared to global temperature due to the primary reason that Ottawa is located in high longitude.
4. The big discrepency between Figure 2 and Figure 3 may be due to a greater range of moving average chosen.
5. The Ottawa's temperature can be roughly estimated with the linear equation in Figure 3.
6. Overall, based on these Figures, the world is getting hotter and the change is consistent over the last few centuries.

Conclusion

The world's temperature is getting hotter globally or locally in Ottawa, and the change is consistent over the last centuries. The temperature changes are highly positive correlated. We can obtain the estimated Ottawa temperature from global temperature using the linear equation. The temperature is lower compared to global temperature primarily due to its location.