Kathleen M. Flaherty

25 May 2018

Data Analyst Nanodegree Program – Term 1

Project: Explore Weather Trends

First, I formulated two separate SQL queries; one to access data regarding average yearly temperatures in New York City, and one to access data regarding average global temperatures.

Following are the two SQL statements which ran without error and pulled the intended data:

```
SELECT year, avg_temp

FROM city_data

WHERE city LIKE 'New York'
```

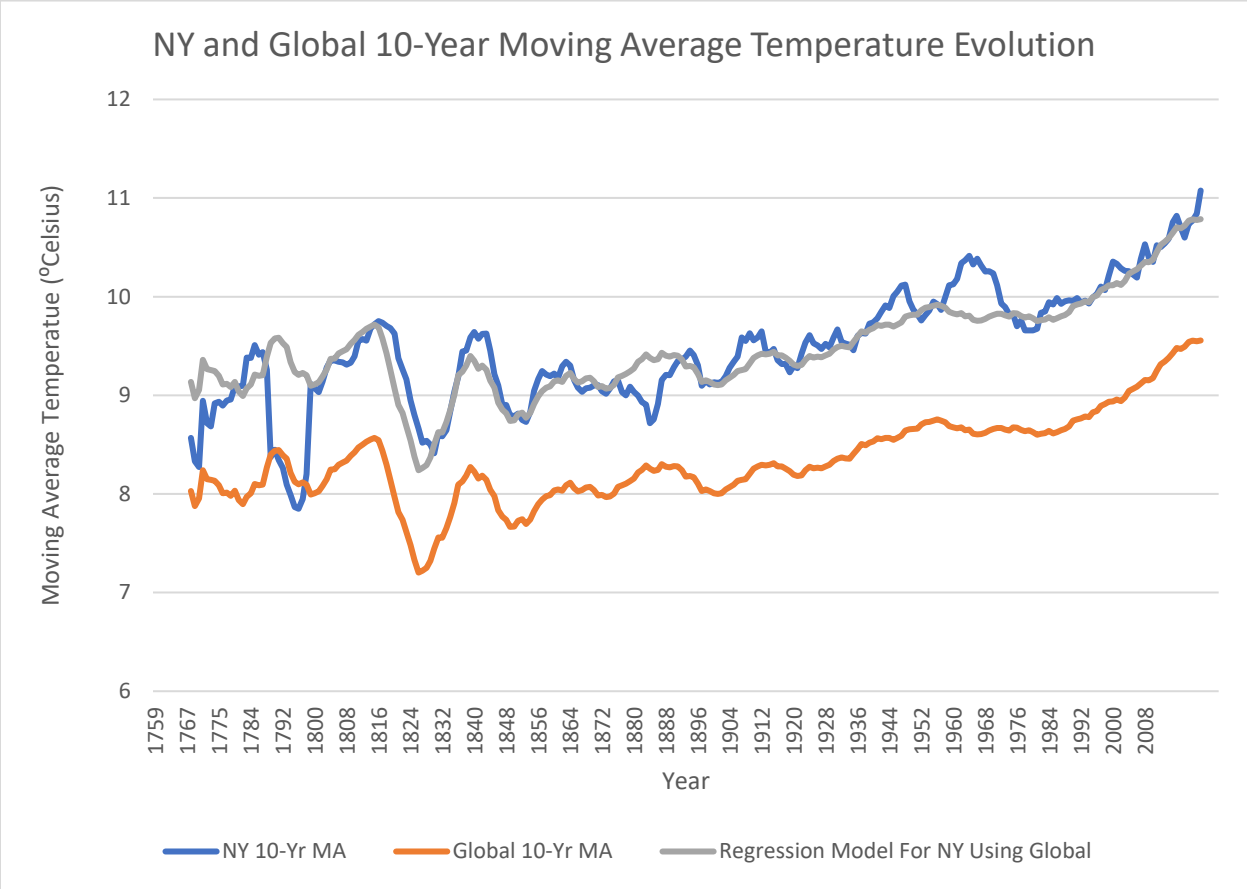
```
SELECT year, avg_temp

FROM global_data
```

When my queries proved successful and I was able to download both corresponding CSV files, I merged the two temperature columns in order to efficiently compare New York City vs. global yearly temperatures. When I merged all the required columns, I realized that data regarding the average annual temperature in New York City started in the year 1743, whereas data regarding the average annual global temperature did not start until the year 1750. Also, between 1746 and 1749, there was a discrepancy in temperature data in the New York City – the temperature data was missing.

Therefore, since global and New York City data did not overlap until 1750, I decided to eliminate all rows prior to 1750. I also removed the row pertaining to 1780 because data (average temperature) was missing for New York City. In order to organize the data, I also eliminated the two extra rows that expressed global temperature recordings for the years 2014 (9.57 °C) and 2015 (9.83 °C) because there was no corresponding data for those years in New York City. I decided that leaving this in the spreadsheet could skew the line graph.

Once I cleaned the data, I calculated the moving averages for lookback of 5 years, 10 years and 20 years and then visualized the line plots. To calculate NY 10 MA, I first used the average function on the B2 to B11 cells (=AVERAGE(B2:B11)), and then used Copy+Paste to fill the remaining rows. I performed the same operation for Global 10 MA and used C2 to C11 cells (=AVERAGE(C2:C11)), and finally I used Copy+Paste to fill the remaining rows. The 10-year moving average gave me optimal insights. Hence, I have included results for 10-year moving average in my spreadsheet. Please refer to the line chart below (the chart also includes the regression model line for NY using Global data):




Here are my observations (Please see spreadsheet for the calculations):

1. Temperatures have been increasing over time, both in New York and on a global scale. In New York, the increase has been 0.01 °C/year. Globally, the increase has been 0.006 °C/year.

2. That being said, the temperature in New York City was higher than the global temperature by 1.13 °C.

3. From 1975, the rate of increase in temperature was higher - in New York, the value being 0.03 °C/year (vs. 0.01 °C/year over the entire period since 1750). Globally, the rate of increase was 0.024 °C/year (vs. 0.006 degrees over the entire period since 1750).

4. Both global and New York temperatures exhibited a decline in the period between 1808 and 1817. In New York, the temperature decreased by 0.11 °C/year. Globally, the temperature decreased by 0.12 °C/year.

5. Between 1953 and 1966, the temperature in New York decreased by 0.04 °C/year, while the global temperature plateaued (~0 °C/year).

Correlation Coefficient

I calculated the correlation coefficient between New York average temperatures and global temperatures. The value is: 0.5634. Between New York and global temperatures, there is a positive correlation.

When the New York temperature goes up, generally so does global. They generally move in the same direction.

Regression Model: Estimate New York Temperature Using Global Temperature

I used a regression model to estimate the New York temperature based on average global temperature. I used Data Analysis Tools in of Excel to do these calculations.

NY Temperature = 1.08191305094123 x Global Temperature + 0.447148920947518

The Adjusted R Square Value is 0.314819298152399, so I can be confident about this model.

Favorite City Visualization

I also looked up the temperature data for my favorite city, Istanbul. This is the SQL query I used to access the information:

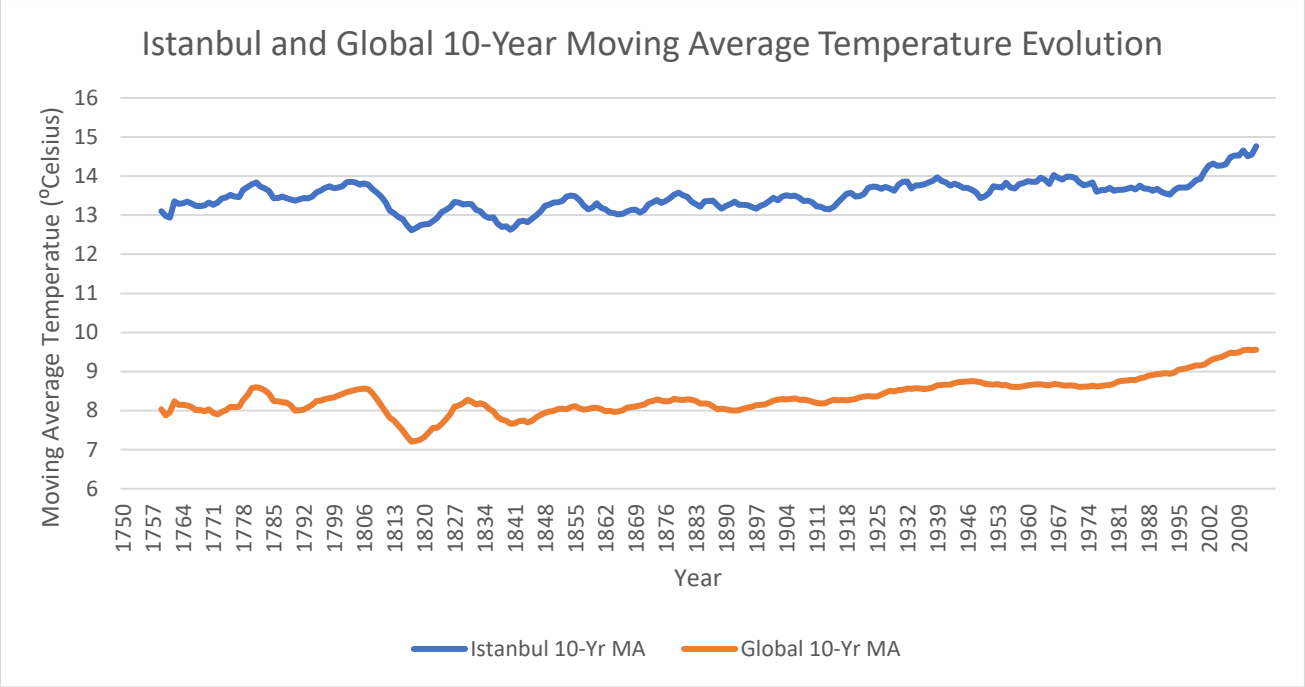
```
SELECT year, avg_temp

FROM city_data

WHERE city LIKE 'Istanbul'
```

First, I cleaned the data. Please note that in this dataset, the year 1780 was not omitted like it was in the New York dataset. Therefore, global moving average values are slightly different.

Next, I proceeded to calculate the 10-Year moving average and then generated the line chart:




Here are my observations for Istanbul (Please see spreadsheet for the calculations):

1. Temperatures have been increasing over time, both in Istanbul and on a global scale. In Istanbul, the increase has been 0.007 °C/year. Globally, the increase has been 0.006 °C/year.

2. That being said, the temperature in Istanbul was higher than the global temperature by 5.15 °C.

3. From 1975, the rate of increase in temperature was higher - in Istanbul, the value being 0.024 °C/year (vs. 0.007 °C/year over the entire period since 1750). Globally, the rate of increase was 0.024 °C/year (vs. 0.006 degrees over the entire period since 1750).

4. Both global and Istanbul temperatures exhibited a decline in the period between 1808 and 1817. In Istanbul, the temperature decreased by 0.11 °C/year. Globally, the temperature decreased by 0.12 °C/year.

## Correlation Coefficient

I calculated the correlation coefficient between Istanbul average temperatures and global temperatures. The value is: 0.7186. Between Istanbul and global temperatures, there is a positive correlation.

When the Istanbul temperature goes up, generally so does global. They generally move in the same direction.

## Hypotheses

Temperatures have been increasing steadily since 1840, both in New York and on a global scale, perhaps due to the dawn of the Industrial Revolution: I think that the Industrial Revolution resulted in burning of fossil fuels on commercial scale, which caused the release of extra greenhouse gases such as carbon dioxide into the atmosphere. This accumulation of greenhouse gases has trapped heat on the Earth's surface; since the heat is not able to escape into space, it has led to an increase in temperature, both globally and in New York.

There was a significant decrease in temperature at the beginning of the 19th century (more, specifically, starting in the year 1808). I researched possible causes, finally to discover that between 1808 and 1815 there was a series of large volcanic eruptions worldwide, the most significant being Mount Tambora in Sumbawa, Indonesia. As a result of these eruptions, there was an accumulation of atmospheric dust which prevented the sunlight from reaching Earth. Therefore, these massive eruptions caused the temperature to drop during this period.

I think that the global average is less volatile than the New York average because the global average is calculated as an average temperature of cities across the world. Therefore, this automatically results in a smoothing effect on the global temperature data. For global temperature, the fluctuations in other cities cancel each other out – in other words, there's less noise in global temperature data. On the other hand, New York temperature has its own fluctuations which cannot be counteracted by fluctuations in other cities.

## Works Consulted

"Industrial Revolution." *Encyclopaedia Britannica,* https://www.britannica.com/event/Industrial-Revolution. Accessed 24 May 2018.

"9 questions about climate change you were too embarrassed to ask." *Vox,* https://www.vox.com/science-and-health/2017/6/1/15724164/9-questions-climate-change-too-embarrassed-to-ask. Accessed 24 May 2018.

"Year Without a Summer." *Wikipedia,* https://en.m.wikipedia.org/wiki/Year_Without_a_Summer. Accessed 24 May 2018.