Project – Explore Weather Trends

Udacity Data Analyst Nanodegree

Matthew Incledon

September 14, 2018

Outline

To generate analysis of global weather trends versus the closest city to me, New York, the first step was to understand how the data is structured in the database and what data is available to me. The following queries provided me this insight:

```
select * from city_data limit 100;

select * from global_data limit 100;

select * from city_list limit 100;
```

Once I knew the data I needed existed in the city_data and global_data tables, I was able to write queries to extract a csv of the data sets I needed:

```
select * from city_data where city = 'New York';

select * from global_data;
```

With this data, I performed the following steps:

1. Imported csv's into separate tabs in Excel
2. Converted temperatures from C to F to sanity check the values
3. Created columns for moving averages with a formula so I could update the average size as best-fit for the analysis
4. Created a third tab with combined data and set a defined range of years based on quality of data. I ended up using 1760 – 2013
5. Populated NY and Global temperatures in Combined tab using index/match from raw/calculated data tabs

I generated a line chart with these 3 columns and formatted the chart for aesthetics and trimmed anything not needed.

Moving Average

To calculate the moving average, I started with the previous 5 years. This generated a line much smoother than just the yearly average though it was still quite choppy. I increased the average set to 10 years and this smoothed out the data in a way where the data trends are still visible.

The formula I used for average was the built-in AVERAGE() function in excel:

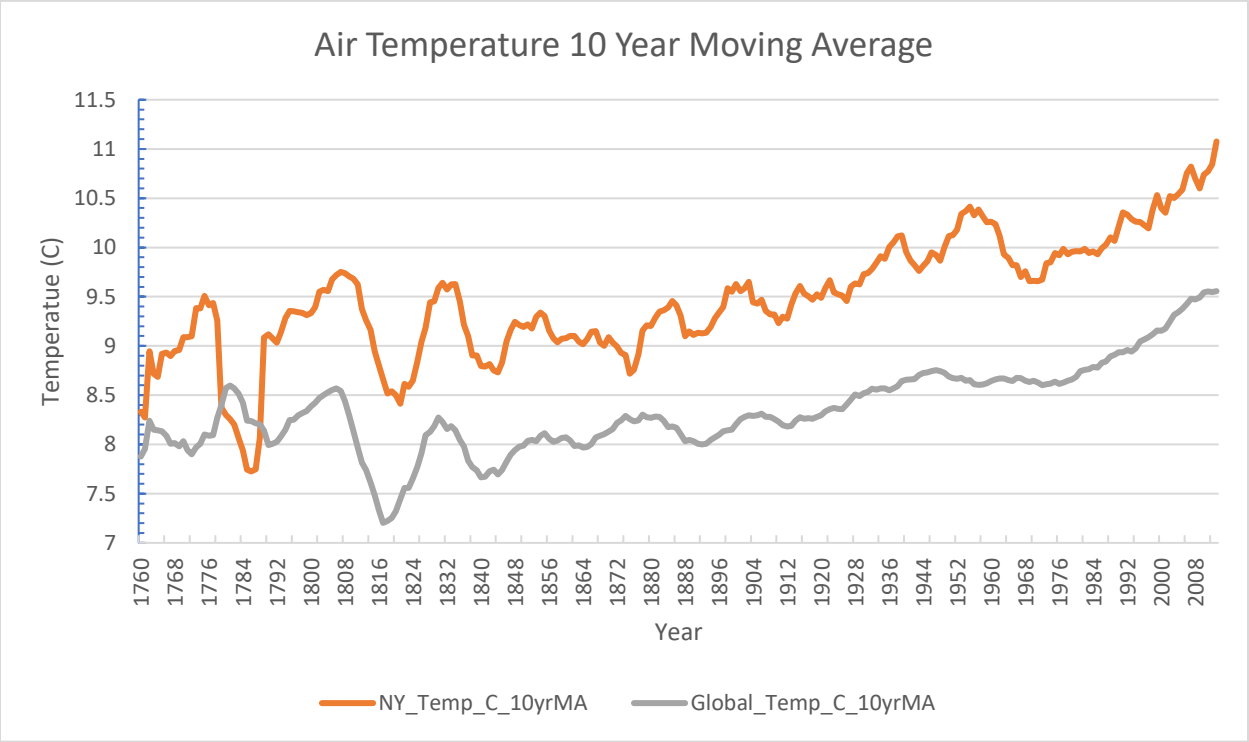
year	city	country	avg_temp	avg_temp_F	10yr_MA
1743	New York	United States	3.26	37.868	
1744	New York	United States	11.66	52.988	
1745	New York	United States	1.13	34.034	
1746	New York	United States			
1747	New York	United States			5.35
1748	New York	United States			6.395
1749	New York	United States			1.13
1750	New York	United States	10.07	50.126	10.07
1751	New York	United States	10.79	51.422	10.43
1752	New York	United States	2.81	37.058	7.89
1753	New York	United States	9.52	49.136	8.2975
1754	New York	United States	9.88	49.784	8.614
1755	New York	United States	6.61	43.898	7.922
1756	New York	United States	9.94	49.892	7.752
1757	New York	United States	8.89	48.002	8.968
1758	New York	United States	8.15	46.67	8.694
1759	New York	United States	9.01	48.218	8.567
1760	New York	United States	7.73	45.914	=AVERAGE(D10:D19)

I shifted the 10 year average by a year for each year, to create a true moving average:

19	**1760**	New York	United States	7.73	45.914	=AVERAGE(D10:D19)
20	1761	New York	United States	10.18	50.324	=AVERAGE(D11:D20)
21	1762	New York	United States	9.55	49.19	=AVERAGE(D12:D21)

Line Chart

The following is the line chart created to show the temperature trends between NY and globally since 1760, a 10 year moving average is used:



Observations

1: Average temperature is increasing over time

It is clear from the data that the global and NY average annual temperature increases over time. Globally, annual average temperate has increased 1.5C since 1850, from 8C to 9.5C.

2: New York is consistently 1C above the global average

New York is consistently ~1C warmer than the average global temperature. Since 1990 this divergence seems to be growing slightly but more data will need to be collected to see if this is trending greater or if it is just a fluctuation.

Given New York's climate experiences all of the seasons, this fact surprises me as I'd expect the number of warm climate areas and cold climate areas globally would average out the global temperature to most closely match New York's temperature.

3: Global temperature started increasing in 1850

There is an inflection point around 1850 where the temperature was steady and then started to increase annually. This coincides with the end of the industrial revolution. Whether this is correlation or coincidence cannot simply be determined with this set of data.

4: Weather goes through peaks and troughs

Average temperature does not follow a linear trend but rather peaks and troughs over time that average out to a trend. Though fluctuations seem to be less than half of a degree C, these are still significant fluctuations that should be taken into account when considering weather patterns historically and in predictions.

Individual cities also follow similar trends but translated either warmer or cooler than the global average, based on city climate. The city trends are also more volatile than global, likely due to a more concentrated set of data and globally being more muted due to a larger data set.