Benjamin Knight
Udacity Data Analysis
Explore Weather Trends

NYC vs. Global Weather Trends (1750-2013)

My goals going into this analysis were to:

- **Collect** the data
- **Calculate** the necessary information
- **Arrange** the data into a clear and accurate representation
- **Articulate** clearly what the data can tell us about both the past, and its implications for the future.

Below I'll go through my process step-by-step with both words and images, and finally the final representations of the data I was given to work with.

Collecting the Data:

- I retrieved the provided data using some basic SQL commands and downloaded the .txt files

```
SELECT *
FROM global_data
```

(I repeated this for both city_list and city_data.)

- Once there I uploaded the .txt files to both Sheets (since it was recommended), and my Jupyter Notebook (since my focus for this course is to improve my programming skills.):

(Importing into Sheets)

```
In [263]: import numpy as np
          import pandas as pd
          import matplotlib as plt
```

```
In [264]: cdf = pd.read_csv('csv files\city_data.txt')
          gdf = pd.read_csv('csv files\global_data.txt')
```

(Pulling the csv files into my notebook.)

- Once that was accomplished I further trimmed the data (specifically city_data) to only contain the information I was looking for. This involved limiting my data to only New York City Temperatures, and also starting the data at 1750. I chose 1750 as the start date for two reasons:
    - Firstly the global data doesn't even begin until 1750, and so there was nothing earlier to compare the city data to.
    - I could have used the city data to map trends and estimate the missing global data (we'll get to that later), however NYC was missing several years of data between 1743 and 1750 which would have skewed any rolling averages. Using the gross average of NYC temperatures to fill the gaps would have made the global temperature prediction for those years inaccurate and would have defeated the purpose of estimating in the first place.
    - There was one other year of data missing for NYC (1780). For such a small slice of data I was comfortable filling in the missing data with the gross average.

```
In [265]: #Reduce the data to just NYC
          cdf = cdf[cdf['city']=='New York']

In [266]: #Global data starts at 1750
          #There are also several years missing from pre-1790 in NYC
          #so we'll start NYC data at 1750 for a cleaner output

          cdfnew = cdf[cdf['year']>1749]

In [267]: #Because otherwise Python yells at me
          cdfnew = cdfnew.copy()

In [268]: #There's no data for 1780 in the NYC df
          #So I've just put in the overall average temperature in its place

          mean = cdfnew['avg_temp'].mean()

          cdfnew.fillna(mean, inplace = True)
```

(Using Python to trim the NYC data.)

Year	City	Country	NYC Avg_Temp
1750	New York	United States	10.07
1751	New York	United States	10.79
1752	New York	United States	2.81
1753	New York	United States	9.52
1754	New York	United States	9.88
1755	New York	United States	6.61
1756	New York	United States	9.94

(The Sheets data was trimmed as well.)

Calculating The Data:

- Once I had all the relevant data it was time to crank out the quicker part of the project: turning that data into something relevant and useable. Again I'll show my process through both Python, and Sheets.
  - Using the data given I put together the moving average calculations taught in the previous lesson (modified to a 10-year moving average). I felt the 10-year moving average made the most sense given that the data encompasses nearly 300 years.

=AVERAGE(B2:B11)

A	B	C
year	Global avg_temp	Global 10-Year M
1750	8.72	
1751	7.98	
1752	5.78	
1753	8.39	
1754	8.47	
1755	8.36	
1756	8.85	
1757	9.02	
1758	6.74	
1759	7.99	8.03

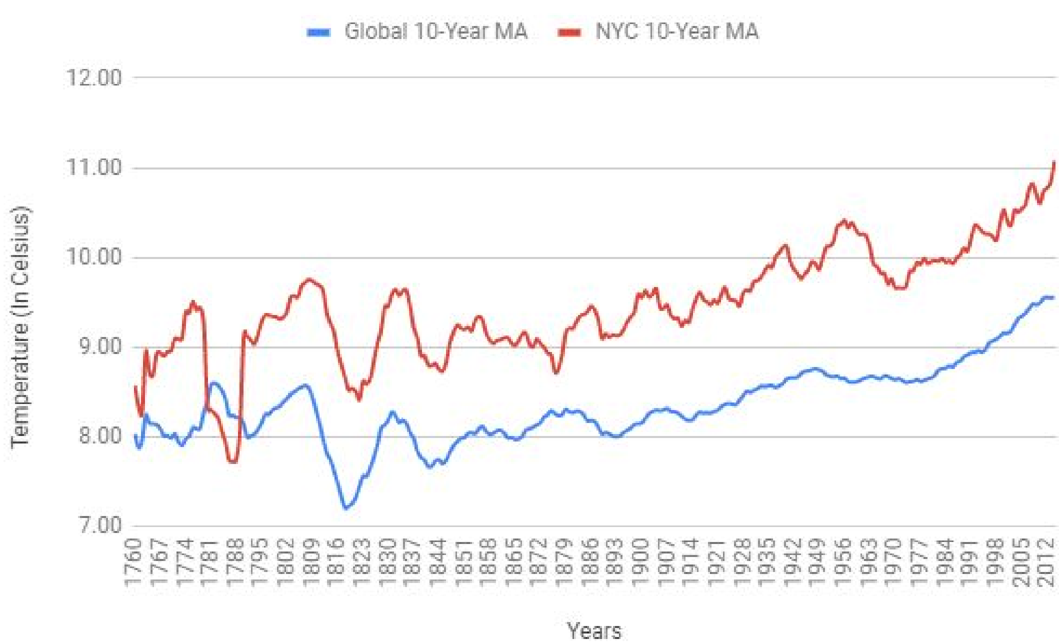
(10-year moving Average in Sheets…)

```
In [270]: #Time to get those rolling averages.
          #I felt a decade window was appropriate for a span of ~300 years

          cdfnew['CMA 10'] = cdfnew['avg_temp'].rolling(window = 10).mean()

In [271]: gdf['GMA 10'] = gdf['avg_temp'].rolling(window = 10).mean()
```

(... and in Python!)

Arranging the Data:

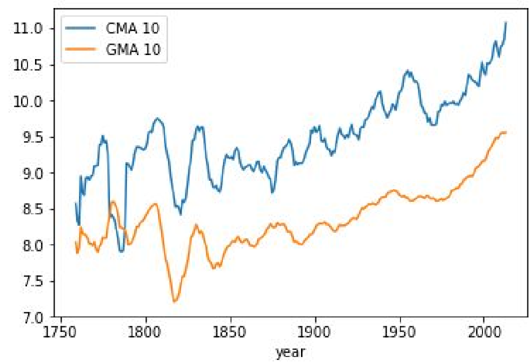
- This proved to be a bit trickier, and more time consuming as I'm less familiar with both Sheets' graphing abilities and Python's matplotlib.plot.line(). However after (much) trial and error I came up with some results I'm proud to share:
  - For Sheets I found the graphing function wholly unintuitive until I was able to experiment heavily with it. After a few stumbles I was able to wrap my head around it:



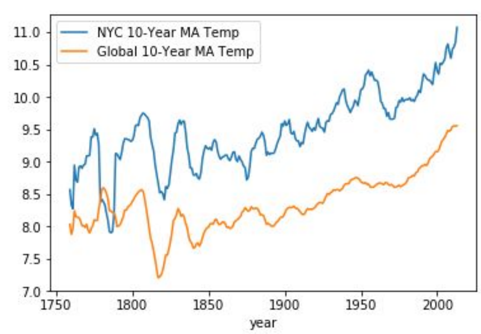
(Line-Graphed Data in Sheets)

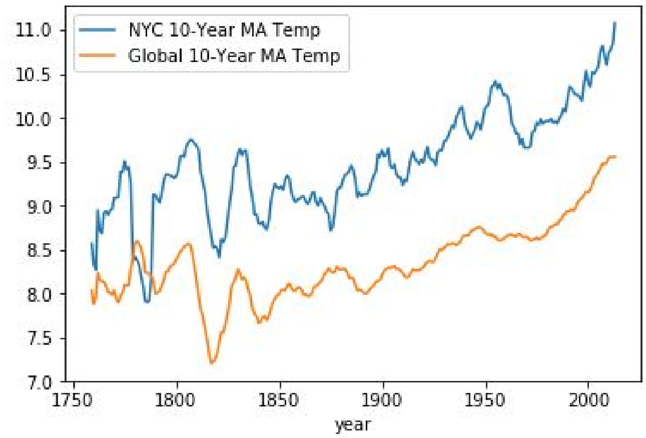
  - The graphing for Python was a bit different and definitely a bigger struggle, but it was easily the better learning experience. Learning how to plot, merge (properly), plot again, and then some was a great way to brush up on my Pandas skill (images below):

```
In [331]: #This was after many failed experiments on how to join and/or merge
          #used suffixes here to differentiate the 'avg_temp' columns
          mergedf = cdfnew.merge(gdf, on = 'year', suffixes = ('_city','_global'))

In [332]: #Here it is, but it could be clearer.
          mergedf.plot.line(x='year', y=['CMA 10','GMA 10'])

Out[332]: <matplotlib.axes._subplots.AxesSubplot at 0x1327f2470f0>
```



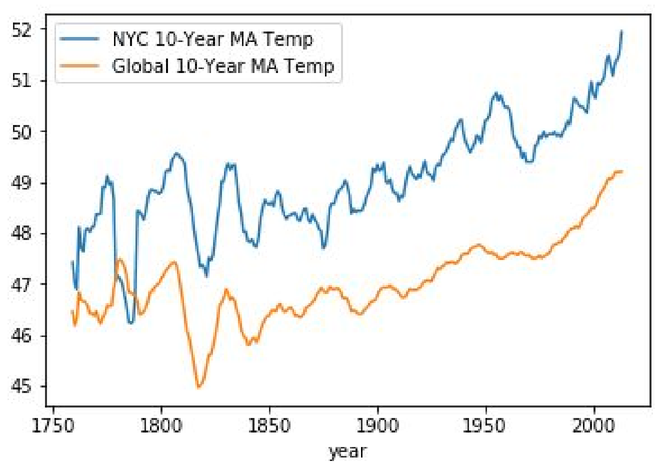

```
In [276]: #So I made my column labels more descriptive for the graph

          mergedf.rename(columns = {'CMA 10':'NYC 10-Year MA Temp', 'GMA 10':'Global 10-Year MA Temp'}, inplace = True)

In [290]: mergedf.plot.line(x='year', y=['NYC 10-Year MA Temp','Global 10-Year MA Temp'])

Out[290]: <matplotlib.axes._subplots.AxesSubplot at 0x1327efdaf60>
```

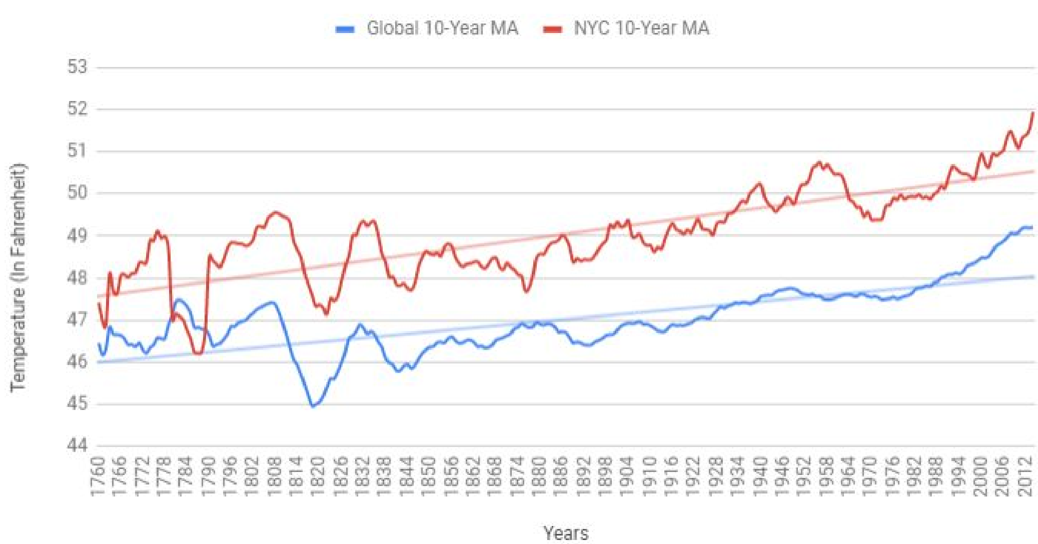




(My start to finish process for creating a line graph in Pandas, could not for the life of me get a y-axis label to print though.)

- As I final bonus I translated them both into Fahrenheit for better comprehension from those unfamiliar with Celsius:





Analyzing the Data:

- Now that I had the data all laid out I was able to make some really fascinating observations, historical correlations, and some less than savory predictions.
  - I think, to get it out of the way, we'll address the clear trend: the temperature is rising at an increasing rate. It's also not hard to see that the beginning of this steady rise is the early 20th century (the beginning of the second industrial revolution.) These spikes are even more noticeable in New York City (having been the birthplace for many industries). What we will do with this information is anyone's guess, but hopefully it's to begin repairing the damage done before it's too late.
  - On a lighter note there are some really great observations to be made here. One of my favorites is you can clearly see when the famous "year without a summer" happened in 1816/1817. A volcanic eruption caused severe weather disturbances and cold snaps worldwide and it looks like NYC was no exception to the rule. *Fun Fact:* It was this unpredictable weather that would trap Mary Shelly indoors with Lord Byron, and (eventual husband) Percy Bysshe Shelly which lead to a game of who could conceive the best ghost story. Shelly's won, and would go on to become the first science fiction novel: Frankenstein.
  - After viewing some research based on the data, it turns out that New York City was subjected to some particularly bad (and isolated) cold snaps in 1780, leading to the only time in almost 300 years where the average temperature of the city dips below the global average. It actually got so cold that one could walk from Manhattan to Staten Island across the frozen Hudson!
  - As the data shows, New York City follows many of the same temperature trends as the global average (which tracks since it's on the globe). However, New York City is notably warmer, which can be explained by both the global average temperature having to account for the Arctic and Antarctic circles, but also the Urban Heat Island Effect. This is where highly urbanized cities like NYC have a higher localized temperature than the surrounding areas. Large contributing factors to this phenomenon are the abundance of concrete and asphalt, which absorb solar radiation, the lack of foliage, and buildings, which block wind and

reflect heat to the surface. This is also why pre-industrial NYC tends to be closer to the global average, as this effect was less severe.

  - Lastly, another sign of man-made climate change is evident in the reduction of extreme seasonal patterns. You'll notice that before the 20th century there are periods of extreme cooling followed by a warmer waves. However, after the turn of the century (again, second industrial revolution) those patterns begin to become less extreme, and in fact after the mid-1970's the pattern almost completely disappears. It is replaced instead by an ever-ominous increase in yearly average temperature with little fluctuation.

Sources:

Urban Heat Island:

https://www.nationalgeographic.org/encyclopedia/urban-heat-island/

Industrial Revolution and Climate Change:

https://climate.nasa.gov/causes/

"Carbon dioxide (CO2). A minor but very important component of the atmosphere, carbon dioxide is released through natural processes such as respiration and volcano eruptions and through human activities such as deforestation, land use changes, and burning fossil fuels. Humans have increased atmospheric CO2 concentration by more than a third since the Industrial Revolution began. This is the most important long-lived "forcing" of climate change."

1816: The Year Without a Summer:

https://scied.ucar.edu/shortcontent/mount-tambora-and-year-without-summer

The Winter of 1780:

https://nj1015.com/the-winter-of-1780-so-cold-you-could-walk-across-the-hudson/