Shaneaka Robinson
Explore Weather Trends – Project 1
Data Analyst Nanodegree

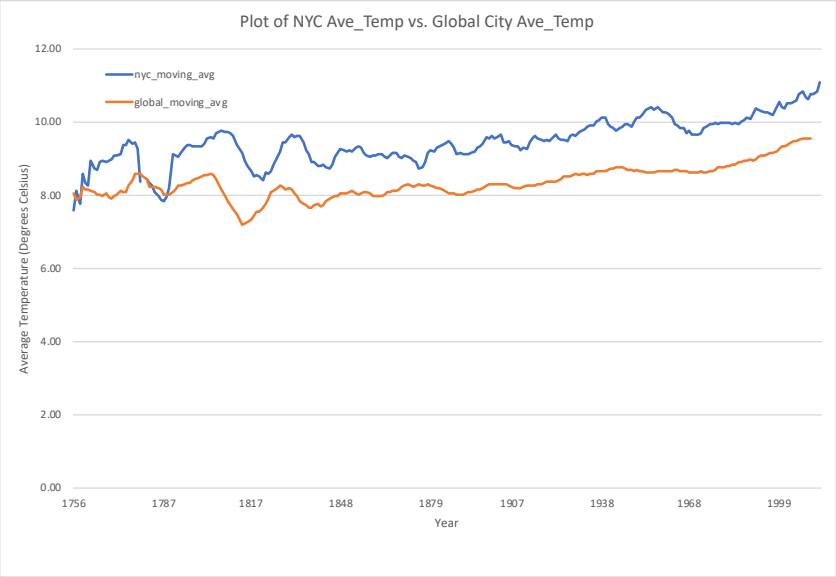
Introduction:
For this project I chose to manually generate the plots using Microsoft Excel and write a script in Python to get the experience using this syntax. I have broken down the steps taken for each method below.

Initial Setup:
Using the following SQL syntax each table was downloaded to a .CSV file and stored on in folder on local desktop.

```
SELECT *
FROM city_data

SELECT *
FROM city_list

SELECT *
FROM global_data
```

Microsoft Excel:

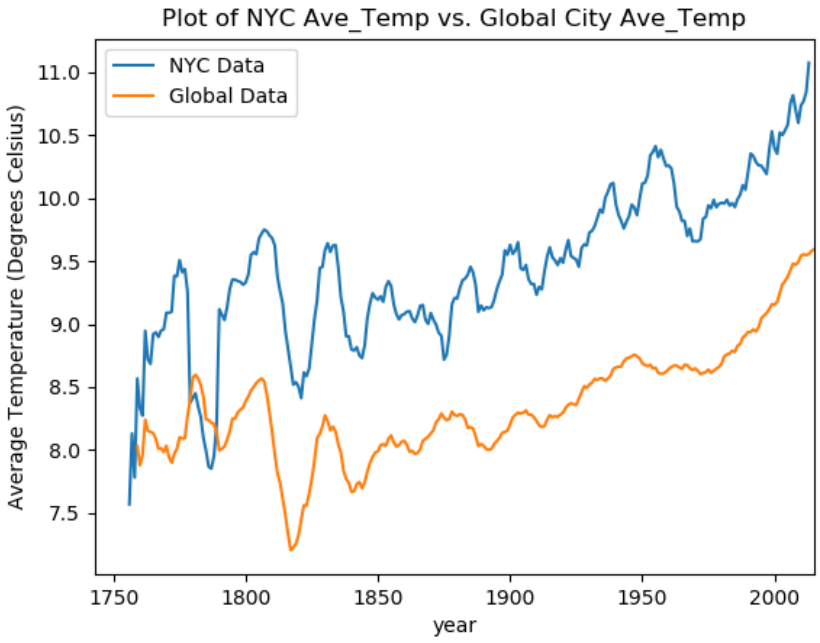
The .CSV city and global data were combined into an .XSLX workbook for analysis.
The city data was then filtered to NYC (nearest local city).
Any rows with missing data were then removed to ensure the moving average was consistently calculated.
The moving average was then calculated using the built in excel AVERAGE() function. The average was taken over a 10 year period for both the local (NYC) data and the Global data.
The data was then plotted on a line graph shown below.



Python Script:

The Python code shown below has been commented to show the steps taken to generate the line chart and make the final observations.

```
import pandas as pd
import matplotlib.pyplot as plt

#Read city_data file
city_data = pd.read_csv("city_data.csv")
#Read city_list file
city_list = pd.read_csv("city_list.csv")
#Read global_data file
global_data = pd.read_csv("global_data.csv")

#Extract my nearest city (New York City) into nyc_data
nyc_data = city_data[city_data.city == "New York"]
#Elimainate cells with no data
nyc_data_complete = nyc_data.dropna()

#Create new column within each data set and
#calculate the moving average for both NYC and Global data over a 10 year period
nyc_data_complete["nyc_moving_avg"] = nyc_data_complete["avg_temp"].rolling(window=10).mean()
global_data["global_moving_avg"] = global_data["avg_temp"].rolling(window=10).mean()

#Plot both NYC and Global data on the same plot
ax = nyc_data_complete.plot(x="year", y="nyc_moving_avg", label ="NYC Data")
global_data.plot(x="year", y="global_moving_avg", label ="Global Data", ax=ax)

#y-axis label
ax.set_ylabel("Average Temperature (Degrees Celsius)")
#Plot Title
plt.title("Plot of NYC Ave_Temp vs. Global City Ave_Temp")
#Generate Chart
plt.show()

# Difference in mean between NYC and Gloabl Temps
nyc_temps = nyc_data_complete["nyc_moving_avg"]
global_temps = global_data["global_moving_avg"]
mean_diff = nyc_temps.mean() - global_temps.mean()
print(nyc_temps.mean())
print(global_temps.mean())
print(mean_diff)
```



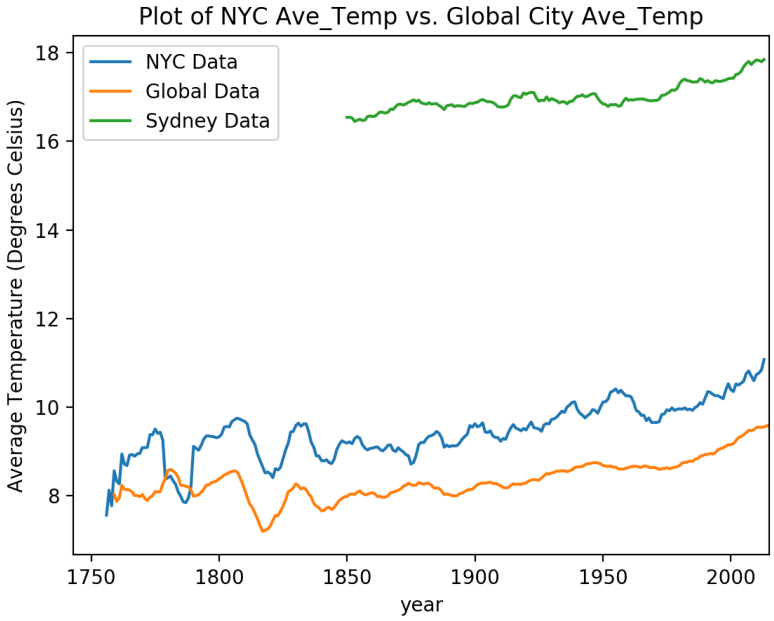
Overall Observations:

- Global temperatures are on average ~1.09°C less than local (NYC) temperatures.
  - Mean NYC temperature = ~9.45
  - Mean Global temperature = ~8.35
- Temperatures on both a local (NYC) scale and global scale are gradually getting warmer.
- While local (NYC) temperatures are always higher than global temperatures both sets of data follow a similar pattern.
- Since about the 1960 temperatures on both a local (NYC) and global level have been increasing on a linear scale

Additional Observations:

Just out of curiosity I wanted to compare data to a city almost 10K miles away.
Here is what I learned:

- Sydney temperatures are on average ~7.54°C more than local (NYC) temperatures.
  - Mean NYC temperature = ~9.45
  - Mean Sydney temperature = ~16.99
- Sydney temperatures are on average ~8.63°C more than Global temperatures.
  - Mean Global temperature = ~8.35
- Temperatures in Sydney are increasing over time but it doesn't appear to be the same linear path as NYC and Global temperatures


Plot of NYC Ave_Temp vs. Global City Ave_Temp
NYC Data
Global Data
Sydney Data
Average Temperature (Degrees Celsius)
18
16
14
12
10
8
1750
1800
1850
1900
1950
2000
year