OUTLINE PROJECT 1

STEPS

1. **Extracted the data to CSV's using the SQL workspace provided.**
    a. Started by using **SELECT * FROM city_list** so I could see what the table looked like.[1]
    b. After seeing the city_list table, I then used **SELECT * FROM city_list WHERE country = 'United States'**. This gave me all U.S. cities which had data available for me to use for this project.
    c. I scrolled down to look for Los Angeles, assuming that would be the closest city with data available; however, I ended up seeing Long Beach, which is actually closer to where I live than Los Angeles. I decided to use Long Beach instead.
    d. Used **SELECT * FROM city_data WHERE city = 'Long Beach'** to get all of the temperature data available for Long Beach.
    e. Downloaded/extracted it to a CSV file called results.csv
    f. Queried the database again, using **SELECT * FROM global_data** this time, which gave me the global average temperature data.
    g. Extracted this to a different CSV file named results(1).csv.
2. **Used Microsoft Excel to get all of the data into one file.**
    a. Copy/pasted the data from results(1).csv (the global average temperatures) to a column in results.csv, ensuring that the corresponding years matched up.

[1] This command caused some lag. In hindsight, it probably would have been better to add LIMIT 5 at the end so it wouldn't have slowed down so much.

b. Noted the columns that existed in my csv file: year, city, country, avg_temp, global_temp. avg_temp has Long Beach's average temperature data.
c. Deleted the cells where there was data available for the global temperature, but not for Long Beach. [2]
d. Saved my changes.

3. Calculated 10 year moving averages for each data set.[3]

a. Labeled two new columns in results.csv, 10_yr_MA_lb and 10_yr_MA_global.
b. Located the cell in 10_yr_MA_lb column corresponding to the year 1858, as this was the first year it would be appropriate to calculate the 10 year moving average, given the available data.
c. In that cell, typed =AVERAGE(
d. Then highlighted the cells from the avg_temp column corresponding to the years 1849 to 1858.
e. Pressed enter, got the 10 year moving average for 1858.
f. Used the method in the moving average lesson to calculate the rest of the 10 year moving averages for Long Beach.[4]
g. Repeated using the global_temp column to find the 10 year moving averages for average global temperatures.
h. Saved changes.

---

[2] There was corresponding data for Long Beach average temperature and global average temperature from 1849 to 2013

[3] I wasn't sure if using a decade would be appropriate for the moving average, but it seemed like a nice round number.

[4] CLICK AND HOLD ON THE CELL, DRAG DOWN MOUSE AND HIGHLIGHT CELLS UNTIL THE END OF THE COLUMN, CTRL+D

4. **Coded a way to graph the data using Python.**

CODE AND COMMENTS:

#Imported necessary packages

```
import pandas as pd

import matplotlib.pyplot as plt
```

#Read the file into python as a pandas DataFrame, df

```
df = pd.read_csv('C:/Users/Andrew/Downloads/results.csv')
```

 # Extracted the years for which 10 year moving averages exist, i.e. 1858 and later. To make sure I only got the years and not the indexes.[5] Stored the years in variable, x.

```
x=df.loc[df['year']>=1858,'year'].values
```

#Extracted the 10 year moving average for Long Beach temperature from 1858 onwards. Stored in variable, y1

```
y1= df.loc[df['year']>=1858,'10_yr_MA_lb'].values
```

[5] It may have been a little better to set the actual index of the dataframe to be the years, and then proceed from there.

#Extracted the 10 year moving averages for global temperature from 1858 onwards. Stored in variable, y2

```
y2= df.loc[df['year']>=1858,'10_yr_MA_global'].values
```

#Plotted the years vs. the moving averages for Long Beach in red, and labeled

```
plt.plot(x,y1,c='r',label='Long Beach')
```

#Plotted years vs. moving averages for global in black, and labeled

```
plt.plot(x,y2,c='k',label='Global')
```

#Labeled x-axis

```
plt.xlabel('Year')
```

#Labeled y-axis

```
plt.ylabel('10 Year Moving Average Temp (°C)')
```

#Created legend, set it in 'best' location, and set labels

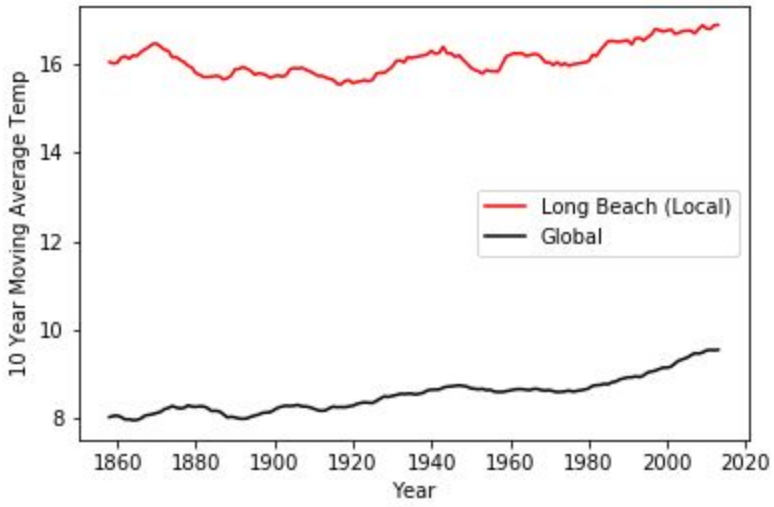
```
plt.legend(loc='best',labels = ['Long Beach','Global'])
```

#Saved figure as 'project1graph.png'[6]

[6] Originally used plt.show() before saving the figure, but this left me with a blank image.

plt.savefig(‘project1graph.png’)

GRAPH:



KEY CONSIDERATIONS:

To make the graph simple to understand, I assigned colors to each line corresponding to each set of data. I made the Long Beach line red, and the global line black. I also labeled the axis so the meaning to the graph would be understood. The x-axis is labeled ‘Year’, and the y-axis is labelled ‘10 Year Moving Average Temp’. Lastly, I created a legend so it is easily understood that the red line represents Long

Beach's data, and the black represents the global data. Something I felt could have been more fine tuned if I had more knowledge about moving averages was the choice of 10 years as my moving average. I wasn't completely sure what number of years to use, and simply chose 10 because it was a nice round number.

OBSERVATIONS:

Overall, the 10 year moving average temperature for Long Beach has always been a good amount higher than the global 10 year moving average temperature. The difference has been fairly consistent over time.

Changes in the 10 year moving average temperature for Long Beach seem to be more abrupt than changes for the global 10 year moving average temperature, signified by the steeper slopes for periods of time on the Long Beach Line.

The overall trend for the 10 year moving average temperature for both Long Beach and the entire world is increasing temperatures.

Of course, over the last 100 years, there has been short periods where the temperature has decreased, signified by the presence of negative slopes for sections of the lines for both Long Beach and the globe. Because of this, people can technically argue that the temperature is just in an uptrend for right now, and it will drop again. However, if you look at the overall trend, from 1860 to 2013, the temperature has clearly increased; the 10 year moving average is currently the highest it has ever been based on our data.