Outline of steps to prepare the data visualization

Step 1. Obtain city_list data and find the local city---SQL, python

Code:

```
SELECT *
FROM city_list;
```

After downloading the “city_list” from SQL as csv file, I opened the file in Jupyter notebook with python and Pandas

```
Code:
df.loc[df.country == 'Canada']
```

I filtered the country and found that “Ottawa” is in the list

Step 2. Obtain the city_data and global_data---SQL

Code:

```
SQL query
SELECT c.year,
        c.avg_temp AS ottawa_temp,
        g.avg_temp AS global_temp,
        c.city
FROM city_data c
JOIN global_data g
ON c.year = g.year
WHERE c.city = 'Ottawa';
```

Use SQL again to get local temperature data in Ottawa and global temperature data as one single CSV file.

Missing value created during joining the two tables are dropped directly with 264 rows left

Step3. Prepare the data for visualization---Python, Pandas

Code:

```
temp = pd.read_csv("/content/drive/My Drive/Data analyst project/
concat table.csv", parse_dates=["year"], index_col= "year")
temp.drop('city', axis =1, inplace = True)
```

Import the data and convert the “Year” column to datetime stamp and set it as index.

Drop the “city” column with only two columns left(Ottawa temperature and global temperature)

Step 4. Calculate moving average—Python, Pandas

Code:

```
temp['Ottawa'] = temp.iloc[:,0].rolling(window = 7).mean()
temp['Global'] = temp.iloc[:,1].rolling(window = 7).mean()
temp.drop(labels=['ottawa_temp', 'global_temp'], axis =1, inplace
=True)
```

Create two columns for moving average in Ottawa and around the globe respectively. Drop the original temperature columns with only the moving average columns left.

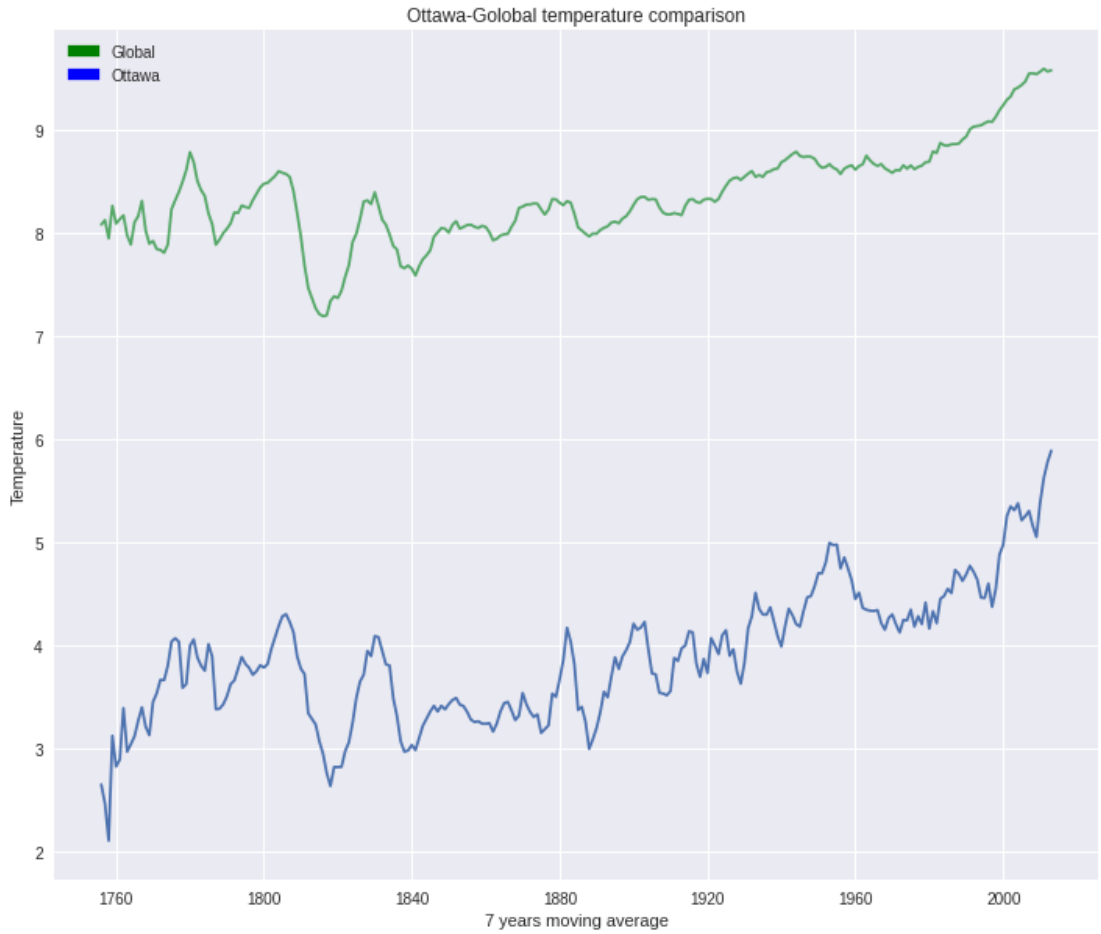
Step 5. Plot the data---Python, Matplotlib

Code:

```
fig, axes = plt.subplots(figsize = (12,10))
axes.plot(temp)
axes.set_xlabel('7 years moving average')
axes.set_ylabel('Temperature')
axes.set_title('Ottawa-Golobal temperature comparison')
green_patch = mpatches.Patch(label = 'Global',color = 'green' )
blue_patch = mpatches.Patch(label = 'Ottawa', color = 'blue')
plt.legend(handles = [green_patch, blue_patch])
plt.show()
```

I tried several moving averages (3 years, 5 years, 7 years, and 10 years). It seems that using 7 years moving average can preserve the fluctuation in ottawa and at the same time see the general trends of the temperature easily.

Line Chart



Observations

Similarity:

- The temperature rises gradually over the past 250 years.
- There was a dramatic dip in averaged temperature in around 1820 in both data.

Difference:

- The average global temperature is obviously higher than that in Ottawa at any time in the history
- Temperature in Ottawa has more fluctuation than that in global data.
- Temperature in Ottawa has a steeper rise in temperature after 2000