Name: **Parth Alpeshkumar Patel**
Date: 31st March 2020
Course: Udacity Data Analyst Nanodegree Project

Explore Weather Trend

Summary

In this project, you will analyze local and global temperature data and compare the temperature trends where you live to overall global temperature trends.

Instructions

Your goal will be to create a visualization and prepare a write up describing the similarities and differences between global temperature trends and temperature trends in the closest big city to where you live. To do this, you'll follow the steps below:

Extract the data from the database. There's a workspace in the previous section that is connected to a database. You'll need to export the temperature data for the world as well as for the closest big city to where you live. You can find a list of cities and countries in the city_list table. To interact with the database, you'll need to write a SQL query.

- Write a SQL query to extract the city level data. Export to CSV.
- Write a SQL query to extract the global data. Export to CSV.

Solution:

- I started by displaying and exploring all the tables by using the below given command:
  Select * from city_data;
  Select * from city_list;
  Select * from global_data;

- I then looked for the New York city, to which I am closet too, in the city_list table.
  select * from city_list
   where country like 'United States' and city like 'New York';

- I realized that there are columns named avg_temp which can cause confusion when we are looking to create a master table putting city average temperature and global average temperature side-by-side. So I ran this query:

  ALTER TABLE GLOBAL_DATA RENAME COLUMN AVG_TEMP TO GLOBAL_AVG_TEMP;

  ALTER TABLE CITY_DATA RENAME COLUMN AVG_TEMP TO CITY_AVG_TEMP;

- Once I managed the naming of columns appropriately, I used the below given query to get a master table.

```sql
SELECT GLOBAL_DATA.YEAR, CITY_DATA.CITY, GLOBAL_DATA.GLOBAL_AVG_TEMP,
CITY_DATA.CITY_AVG_TEMP
FROM GLOBAL_DATA
INNER JOIN CITY_DATA
ON GLOBAL_DATA.YEAR = CITY_DATA.YEAR;
```

At this point, I am mentioning the query used by me before downloading the data into a csv file, but there is also an alternate approach wherein we can only extract data of the New York City from the master table to give us the focus point of our project i.e. choosing a city (nearest to you) and comparing it's average temperature with the global average temperature.

```sql
SELECT GLOBAL_DATA.YEAR, CITY_DATA.CITY, GLOBAL_DATA.GLOBAL_AVG_TEMP,
CITY_DATA.CITY_AVG_TEMP
FROM GLOBAL_DATA
INNER JOIN CITY_DATA
ON GLOBAL_DATA.YEAR = CITY_DATA.YEAR
WHERE CITY_DATA.CITY LIKE 'NEW YORK';
```

- With our approach we have a master table consisting of 4 columns namely year, city, city_avg_temp, global_avg_temp in a csv file which we import in a ipython notebook using Jupyter console.
- The codes used in Jupyter can be found in the Jupyter notebook itself attached in this project or a pdf file named as 'report.pdf'.
- My logic while using python to plot the comparison is listed below:
  - Importing necessary packages for data manipulation.
  - Reading the excel or csv file into a data container.
  - Convert that data container into a pandas data frame.
  - Filtering the data to only have New York city's data.
  - Import matplotlib.pyplot package to plot comparison.
  - Calculate the moving average for both the temperatures.
  - Use the pyplot commands to actually plot the comparison as a graph

- Considerations for plotting the averages:
  - The Y-axis will host the range of average temperatures.
  - The X-axis will host the timeline of years from 1750s to 2013.
  - The legend describes the color of the curves.

Conclusion:

Temperature comparison <of New York and Global average>
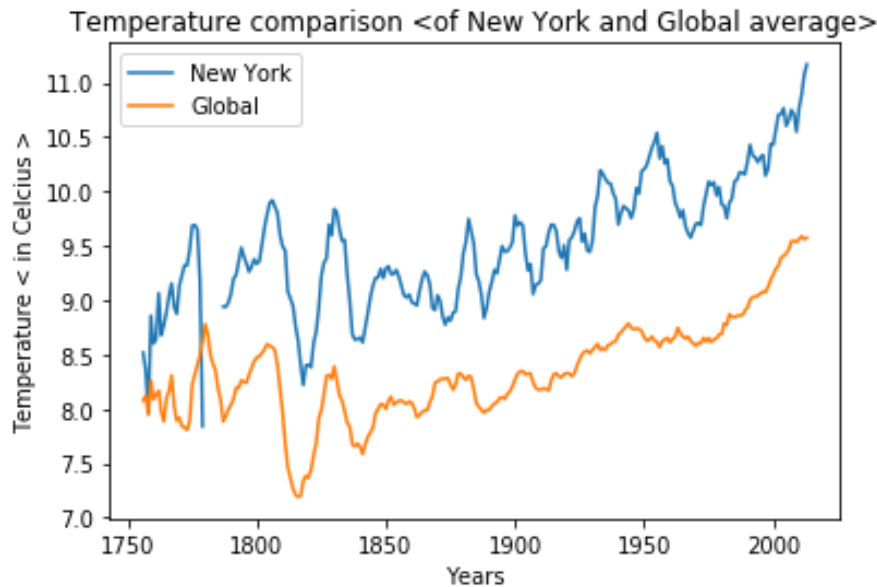
- When we experiment by increasing the rolling parameter i.e. taking moving averages like 10, 50, 150, the temperature curves tend to distinctly separate out from each other and creating more smoother version of respective moving averages.
- At the same instance, how will we compensate in terms of loosing the insights about fluctuations. The trade-off raises some tricky situation.
- The similarity we can see at first glance of the plot is that both the average temperatures have a positive slope and coming to second half of 18th century, the slope becomes steeper, which can be attributed to development and urbanization taking place in both NYC and the world as well.
- The difference between both the averages tend to stay nearly constant at about 1-1.5 °C.
- On average the NYC average temperature has always been higher than the global averages. The reason could be the fact about metropolitan area having more pollution from higher energy consumption and industries and offices it accommodates.
- At various points (years in our case) we see that the fall or rise in global average temperatures is reflected in NYC's average temperature too. We can deduce that those years were either hotter or colder for the world or earth as a whole.
- Finally, the overall trend looks like the world is getting hotter year by year with minor exceptions. The trend has been consistent over last few hundred years because of modernization and its polluting effect on the atmosphere.