Explore Weather Trends

Data Extraction

1. The following 2 SQL queries were used to extract city level and global data. My city is Indianapolis and hence I included a WHERE condition for city = 'Indianapolis' in my query.

   SELECT * FROM city_data WHERE city = 'Indianapolis';
   This query fetches all data for Indianapolis from the city_data table.

   SELECT * FROM global_data; This query fetches all data from the global_data table.

2. In this project, I used Python to read the data from csv files, clean it and visualized the data using the matplotlib library. Scripts were run from Jupyter Notebook.
3. Data was stored in cityTempDF and globalTempDF data frames respectively using the Pandas library in Python.
4. cityTempDF data frame had average temperature values by year for Indianapolis ranging from 1743 to 2013.
5. globalTempDF data frame had average global temperatures by year from 1750 to 2015.

Data Cleaning

1. Before calculating the moving average for the temperature in Indianapolis and for the global data, I analyzed the data to check for null values. The 'avg_temp' column in my local city "Indianapolis" had 5 null values out of the total 270 records. No avg_temp value was recorded for the years 1746, 1747, 1748, 1749 and 1780 in Indianapolis.
2. Handling missing data is important. Even though the number of records with null values for avg_temp in cityTempDF is just 5, the period from 1746 – 1749 had null values consecutively.
3. I used mean imputation to handle missing data for the 5 years. Mean average temperature for Indianapolis is 11.14 and was used to replace the 5 null values in the cityTempDF data frame.
4. The globalTempDF had no null values.

Calculating Moving Average

I used the matplotlib library to plot a line chart for both average temperature by year for Indianapolis and the globe with the X axis representing values from 1740 to 2015 at intervals of 10 years. With too many spikes and the line charts overlapping each other, I decided to look at the moving average for 5 years.

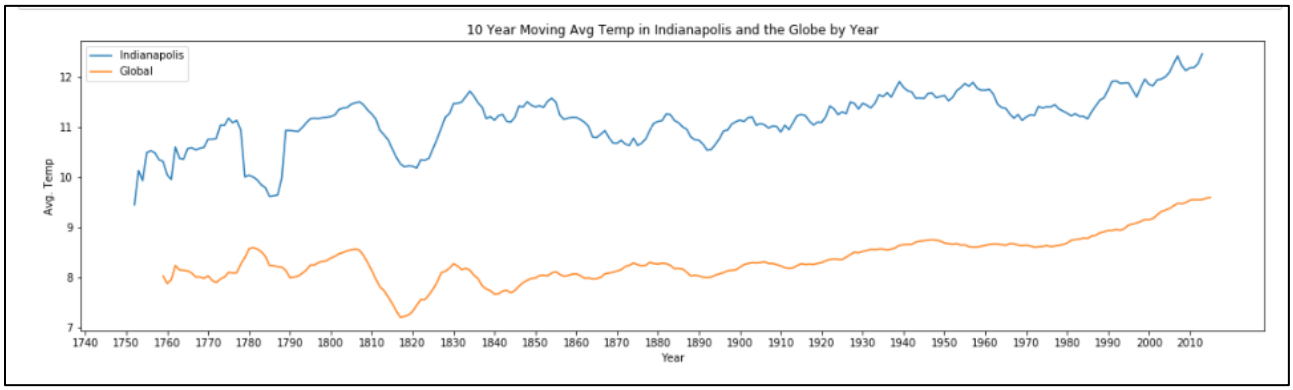
For moving average, I calculated 5-year, 7-year and 10-year moving average temperature for both Indianapolis and global temperature. The following function was used in Python to calculate 5 year moving average for Indianapolis.

```
cityTempDF['5yr_moving_avg_temp'] = cityTempDF['avg_temp'].rolling(window=5).mean()
```

The 5-year moving average still had a lot of spikes and decided to visualize 7-year moving average. The 7-year moving average for global temperatures were smooth however, the 7-year moving average for Indianapolis had a lot of spikes and wasn't clear to derive insights.

I feel the 10-year moving average is much clear to understand with more smoother curves in the line chart and hence I decided to analyze the 10-year moving average line chart.

The following is a line chart for 10-year moving average temperature for Indianapolis and the globe.



Observations

1. In the above line chart, it can be observed that from the year 1900 the 10-year moving average temp for the globe is increasing consistently. The 10-year moving average temperature during the period from 1940 to 1980 is constant and is subsequently increasing again from 1980. The world is getting hotter from 1980.
2. Overall there is a trend that shows that the world is consistently getting hotter since the past 120 years. (since 1900)
3. The 10-year moving average temperature in my local city "Indianapolis" is comparatively hotter than the global average. This can be seen by the gaps between the blue and orange lines.
4. Overall the 10-year moving average for Indianapolis follows a similar trend to global temperatures maintaining a consistent difference over time.