Data Analyst Nanodegree

Suelen Fenali

Exploring Weather Trends

Preparing the data

- The data was extracted using SQL. It was extracted two data sets:
    - one containing the closest city to São Paulo, which is Guarulhos.

```
SELECT *
FROM city_data
WHERE
        country = 'Brazil' AND city = 'Guarulhos';
```

    - second dataset containing the global temperatures averages

```
SELECT *
FROM global_data;
```

Calculating the moving averages

- In order to calculate the moving averages it was used Python and pandas library. The figure 1 shows the code to calculate the moving averages.

- The bigger is the window to calculate the moving averages, the less you see small variations. Since, the objective is to identify a trend and not anomalies a moving window of 50 years looks good to observe how the weather is behaving. Figure 2 illustrates four different moving window size being applied, and how the trends become more evident.

Loading CSV file with the city and global data

```
In [1]: import pandas as pd
        cities_data = pd.read_csv("cities_data.csv")
        global_data = pd.read_csv("global_data.csv")
```

Creating a function to calculate the moving average

```
In [2]: def calculate_moving_averages(city_df, window_size):
            moving_av = pd.DataFrame(city_df['avg_temp'].rolling(window=window_size).mean())
            moving_av.columns = ['moving_averages']
            return moving_av
```

Calculate moving averages and saving in the Dataframes "guarulhos" and "global_data"

```
In [3]: guarulhos = cities_data.loc[cities_data['city'] == 'Guarulhos']

        guarulhos = guarulhos.join(calculate_moving_averages(guarulhos, 50))
        global_data = global_data.join(calculate_moving_averages(global_data, 50))
```

Figure 1 - Code to calculate the moving averages

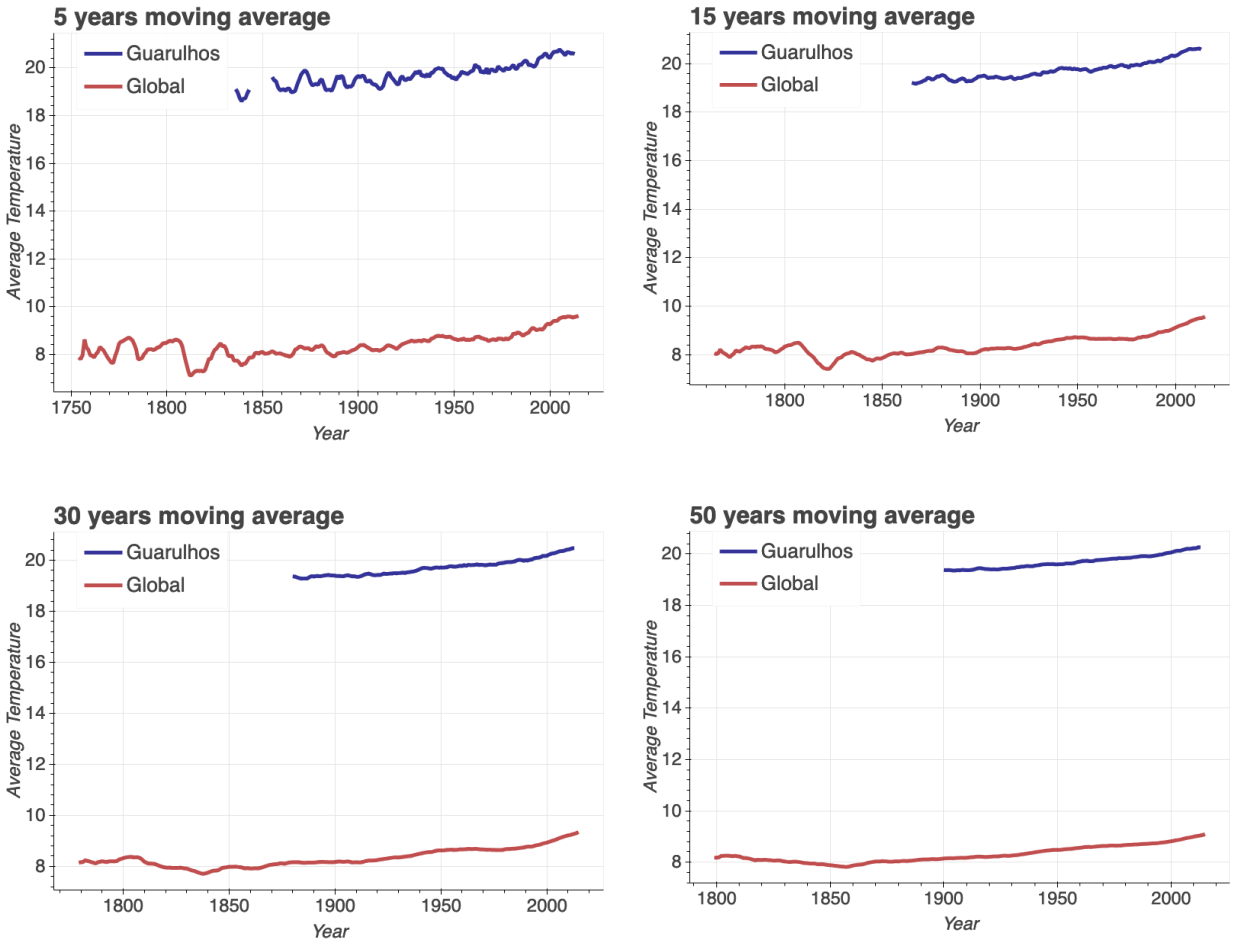

Figure 2 - Comparison among moving averages for 5, 15, 30 and 50 years.

Visualizing and comparing the trends

The chart below shows the weather trends for Guarulhos and Global.

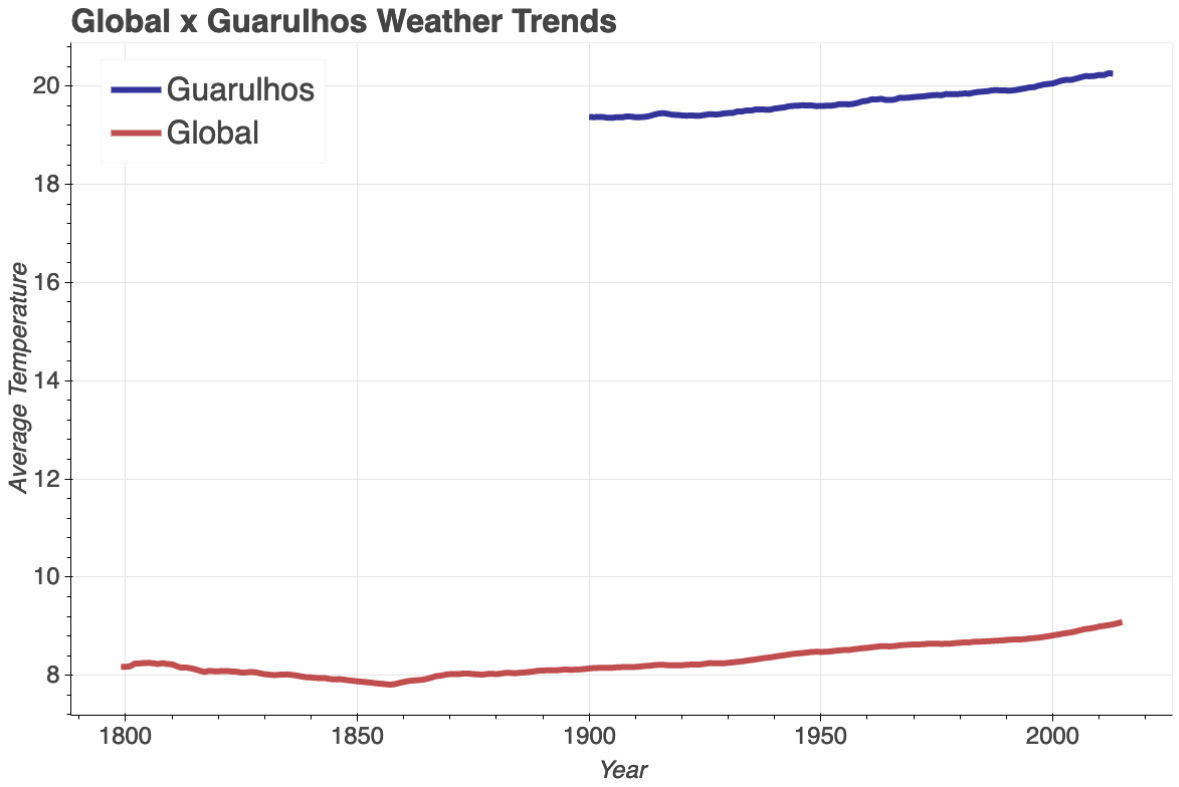

Figure 3 - Chart comparing global and Guarulhos trends

Observations

1. The global average temperature, and Guarulhos temperature are increasing along the years. Both are rising in a similar trend.
2. The average temperature in Guarulhos is approximately 11 degrees Celsius higher than the average global temperature.
3. Guarulhos city is getting hotter since when we have data to analyse. The world started getting hotter around 1860, before that, the temperature was decreasing in the previous 50 years.
4. In the last hundred years (1913 to 2013) the temperature increased 1.07 degrees Celsius for Guarulhos and 1.1°C for the global trend.
5. The maximum average temperature registered for Guarulhos was 21.3 degrees in 2002, while the highest global average temperature was 9.83 degrees in 2015.
6. The minimum average temperature registered for Guarulhos was 18.28 degrees in 2002, while the lowest global average temperature was 5.78 degrees in 2015.

Extra

This section presents some extra insights about the data.

Correlation coefficient

The correlation coefficient (using pearson correlation) between the global average temperatures and Guarulhos temperatures is: 0.794454. It was calculated using the following code lines .

```
In [17]: data_to_correlate = temperature_data[['global', 'guarulhos']]
         data_to_correlate = data_to_correlate.dropna()
         data_to_correlate.corr(method='pearson')
```

Out[17]:

	global	guarulhos
global	1.000000	0.794454
guarulhos	0.794454	1.000000

Figure 4 - Code to calculate the correlation between Guarulhos and global temperatures.

This is a strong correlation, so for future years we could predict Guarulhos temperature based on the global trends.

Multiple cities

Other three cities were added to the chart line, so, it was possible to have a better understanding of the weather trends for other regions of the world.
To extract the information the SQL below was used:

```
SELECT *
    FROM city_data
    WHERE
        (country = 'Brazil' AND city = 'Guarulhos') OR
        (country = 'Australia' AND city = 'Brisbane') OR
        (country = 'Germany' AND city = 'Berlin') OR
        (country = 'Canada' AND city = 'Toronto'');
```

The moving average of 50 years was calculated for the other cities using the same code shown previously. The chart line follows in Figure 5.

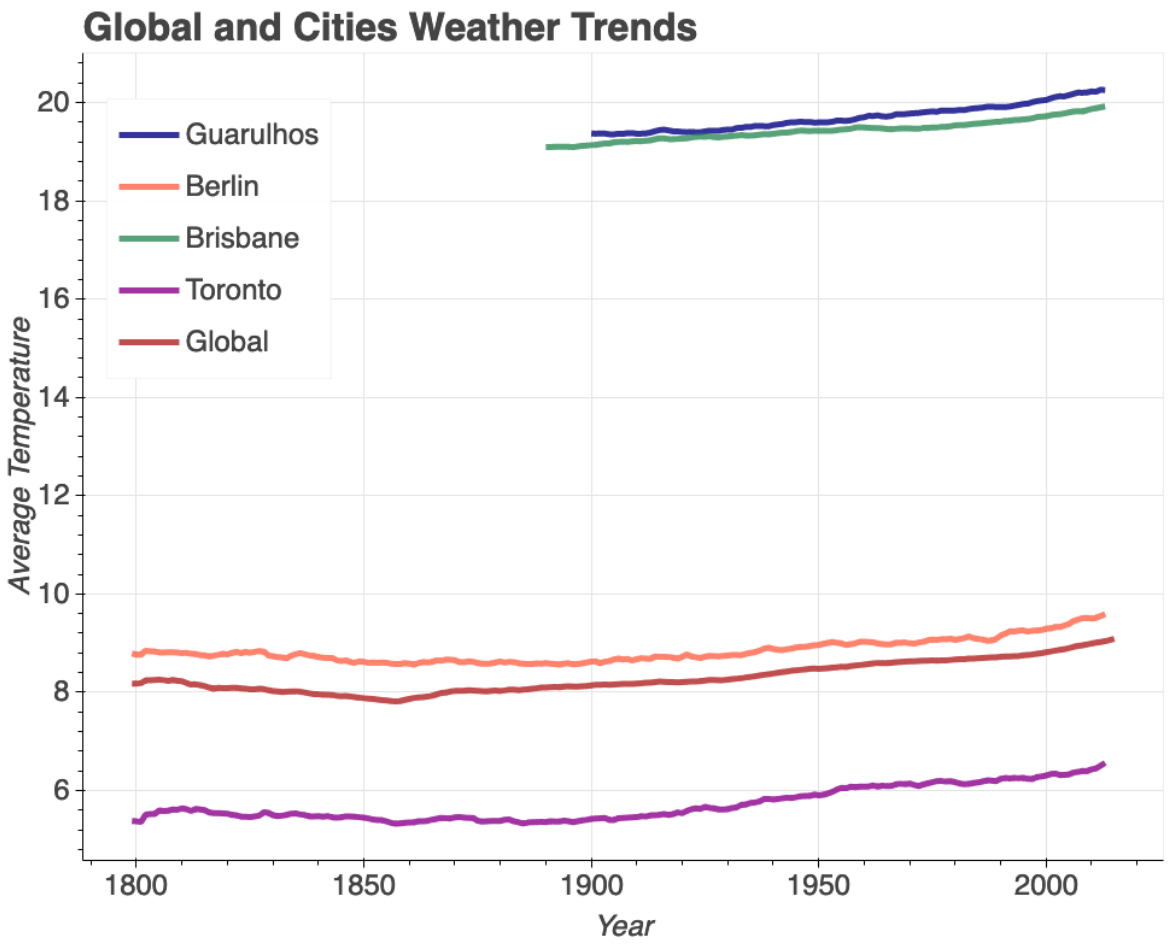

Figure 5 - Chart comparing global and cities trends

1. All the cities follow the increasing trend in the last hundred years.
2. Berlin and Toronto, which have data from the beginning, follow the same trend as the global one.
3. In Berlin and Toronto there is a slightly decrease of the temperature between 1800 and 1850. However the average starts increasing after then, just as the global trend.
4. Toronto is the city with the lowest temperature among the cities observed.
5. Brisbane has a weather trend very similar to Guarulhos
6. Berlin is the closest city to the world weather trend.