

Exploring Weather Trends in San Francisco, United States with Global Data
Frances Chan
Data Analyst Nanodegree
7/10/20

Introduction:

This project focused on analyzing local and global temperature data and comparing the temperature trends in San Francisco, United States to overall global temperature trends. I am required to create a visualization and interpret the temperature trends in a write up. The data is extracted from a database provided in the workspace by using SQL queries. Next, I am required to calculate the moving averages and create a line chart. Lastly, I need to make at least four observations.

Steps and data extraction:

1. In order to find the city nearest to me, I inputted:

```
SELECT *  
FROM city_list;
```

in which is San Francisco.

2. To look at the average temperatures for San Francisco by year (°C), I put in:

```
SELECT *  
FROM city_data;
```

3. To look over the average global temperatures by year (°C), wrote:

```
SELECT *  
FROM global_data;
```

4. In order to exact both city_data and global_data specifically just for San Francisco and the global data, I inserted:

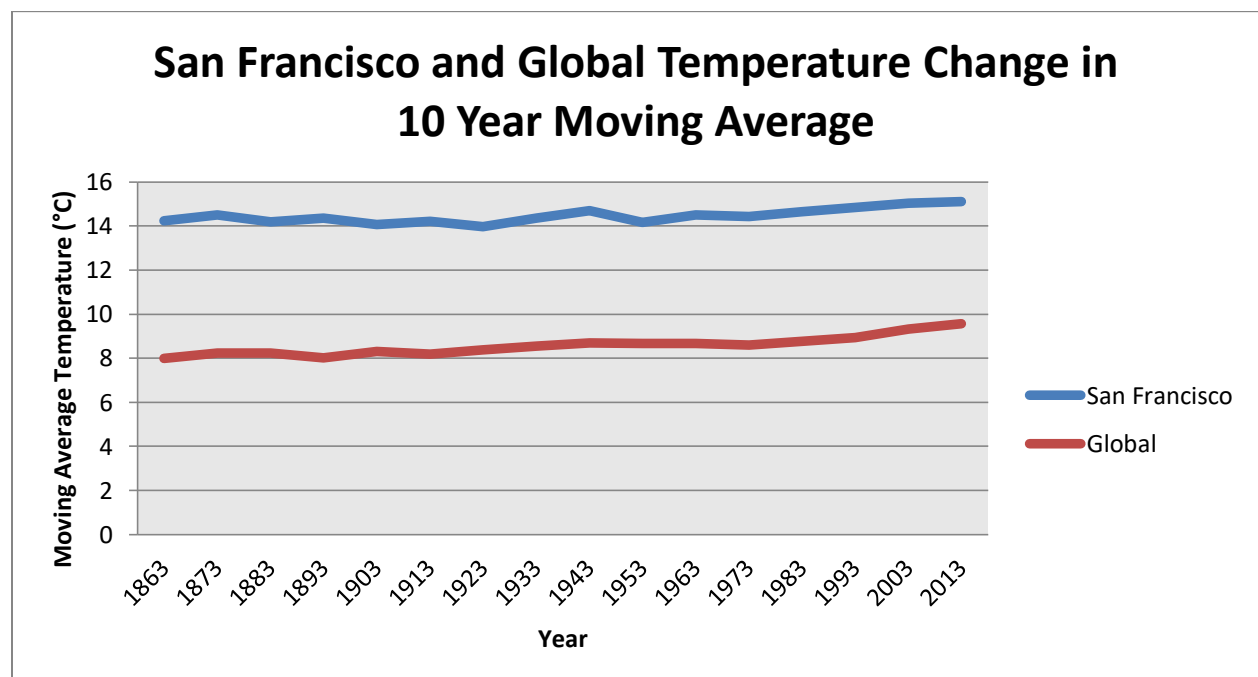
```
SELECT city_data.year, city_data.city,  
city_data.country, city_data.avg_temp,  
global_data.year, global_data.avg_temp  
global_avg_temp  
FROM global_data FULL OUTER JOIN city_data ON  
global_data.year = city_data.year  
WHERE city = 'San Francisco';
```

The output is a total of 165 results with year from 1849 to 2013.

5. After exporting the city_data (San Francisco) and global_data to CSV, I used Microsoft Excel, specifically the AVERAGE() function to calculate the moving average. Note: I am aware that I can calculate the moving average using Python or R. However, I chose not to because I haven't install NumPy nor Pandas. I have learned Python using elsewhere software libraries. The interval of moving average I have decided to use is 10 year moving average. As a result, I started from the year 1853 to 1863 to 1873, and so on forth to 2013 (10 year moving average). Either way some years are missing due to whichever moving average interval I choose.

Key considerations:

- The chart has a title.
- The y-axis is Moving Average Temperature (°C).
- The x-axis is Year.
- The chart includes a legend.
- Used different line colors for San Francisco and the global moving average.



Observations:

1. San Francisco's moving average is warmer than the global moving average. Over time the difference is pretty consistent as about 6 °C difference between the both.
2. Over time, both San Francisco's moving average and the global moving average have been increasing slowly only about 2 decimal points, not even 1 °C.
3. The overall trend is that the world is getting warmer and seems to be pretty consistent over the last few hundred years as there is only about 2 °C change.
4. In San Francisco, the moving average ranges from 13.97 to 15.1 °C from year 1853 to 2013 whereas the global moving average ranges from 7.99 to 9.56 °C from 1853 to 2013.
5. The highest moving average year is 2013 (2003-2013 decade) for both San Francisco and the global data. 1863 (1853-1863 decade) was the year with the lowest moving average (7.99 °C) for the global data and 1923 (1913-1923 decade) was the year with the lowest moving average (13.97 °C) for San Francisco.
6. There is definitely a positive correlation for both, the correlation coefficient is 0.82 via calculated on Microsoft Excel.