Exploring Weather Trends
OCTOBER 2018

OVERVIEW

I analyzed San Francisco and Global temperature data and compared trends. Afterward, I created visualizations to aid understanding of the data.

ANALYSIS

1. **What tools did you use for each step? (Python, SQL, Excel, etc)**

    a. I used SQL to extract the temperature data for *Global* and for *Local* (for San Francisco, CA) and exported the data to a CSV file. I used the following queries for each:

        i. Global Data:

        ```
        SELECT *
        FROM global_data;
        ```

        ii. Local Data:

        ```
        SELECT year, city, avg_temp
        FROM city_data
        WHERE city = "San Francisco";
        ```

    b. I then uploaded the CSV file into Google Drive and then used Google Sheets to calculate the moving average and to create line charts.

2. **How did you calculate the moving average?**

    a. To calculate the moving average, I created a series of different subsets consisting of 5-year increments starting with a 10-year moving average,10, 15, and 20-year periods for both Global and Local Data. I used the **AVERAGE** function in Google Sheets to calculate each of the three different moving average time periods that I tested.

```
=AVERAGE(C2:C11)
```

	A	B	C	D	E	F
	year	city	avg_temp	10yr_mov_avg	15yr_mov_avg	20yr_mov_avg
	1849	San Francisco	14.12			
	1850	San Francisco	13.80			
	1851	San Francisco	14.39			
	1852	San Francisco	13.81			
	1853	San Francisco	14.40			
	1854	San Francisco	13.98			
	1855	San Francisco	14.20			
	1856	San Francisco	14.10			
	1857	San Francisco	14.78			
	1858	San Francisco	14.19	14.18		
	1859	San Francisco	13.71	14.14		

Image 1: Moving average in 5-year increments (10, 15, and 20-year moving averages) for Local Temperature Data.

3. **What were your key considerations when deciding how to visualize the trends?**

   a. The time period of the data captured between the Global Temperature Data and the Local Temperature Data only overlapped partially. The Global Temperature Data had data from 1750–2015, while the Local Temperature Data had data from 1849–2013. Thus, I selected to focus on the time frame in which there was data for both data sets (1849–2013).

   b. To determine the starting date for each of the moving-average formulas (**=AVERAGE()**), I needed to add +n (n= 10, 15, and 20 (year periods)) to the 1849 date.

   c. Since I was comparing data sets for temperatures, I used colors in my line chart to indicate cool (in the title), warm (Global), and hot (Local).

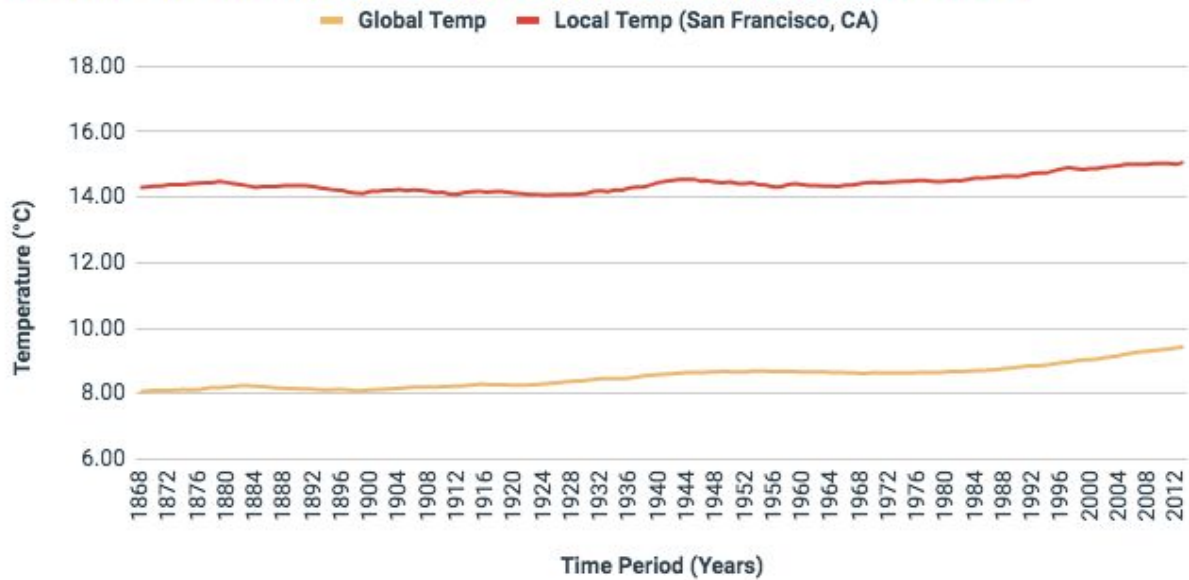Image 2: 20-year moving average comparing Global and Local temperatures

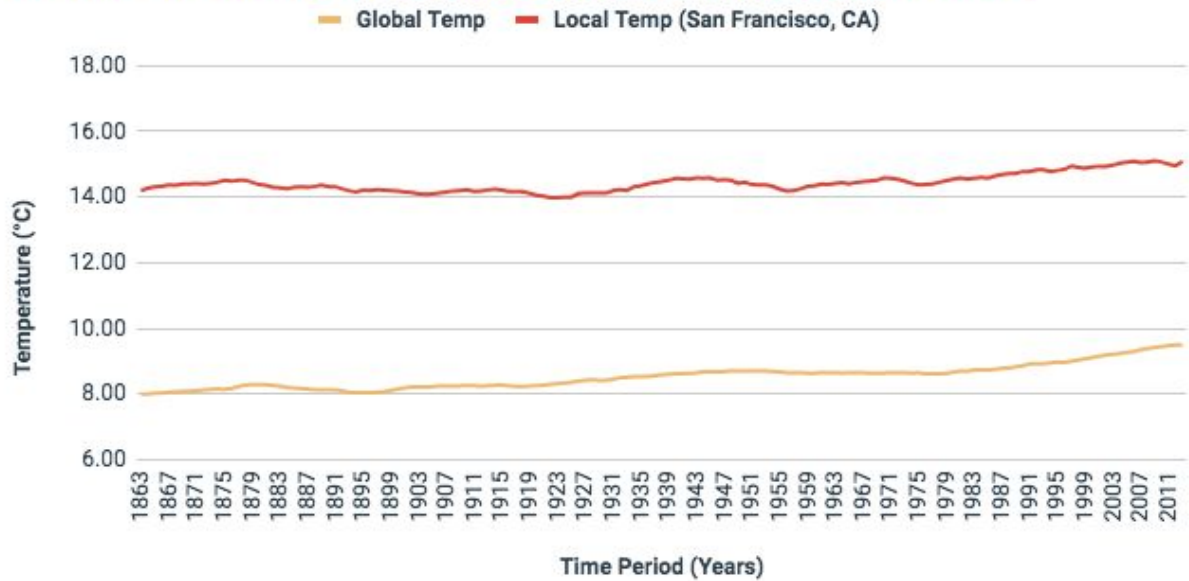Image 3: 15-year moving average comparing Global and Local temperatures

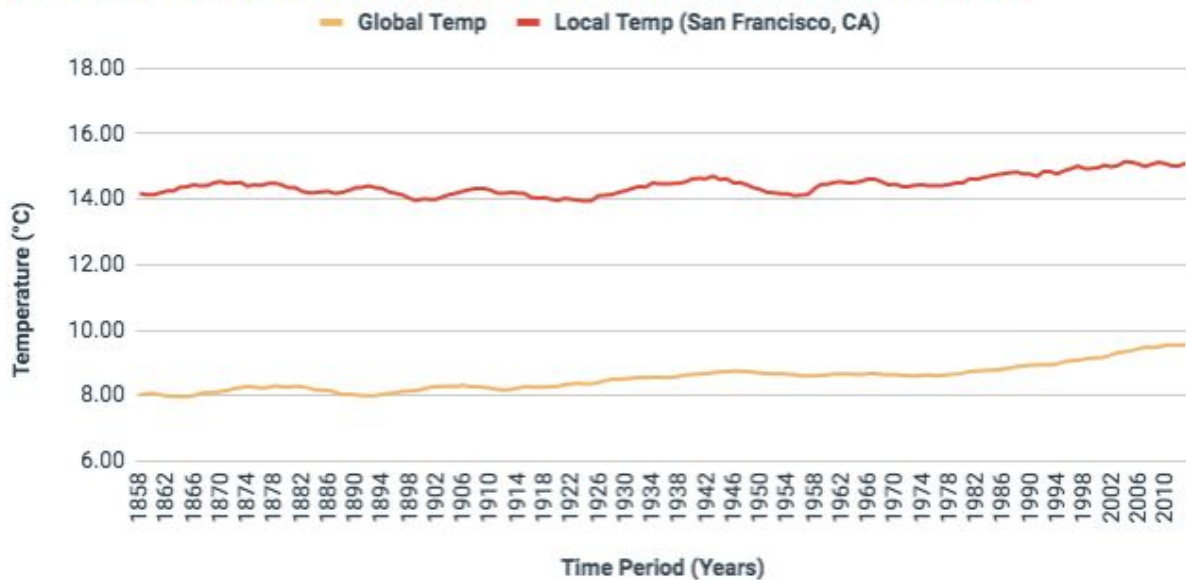Image 4: 10-year moving average comparing Global and Local temperatures

	MIN	MAX	AVG	STDEV
Local Temp (San Francisco, CA)	13.22	16.23	14.45	0.505
Global Temp	7.56	9.73	8.55	0.460

Image 5: Some descriptive statistics for Global and Local temperatures from 1849–2013

OBSERVATIONS

- The Global and Local temperatures are both increasing overtime from 1849–2013.
- On average, the Local temperature is double the Global temperature, with the average temperature being 14.45°C for Local and 8.55°C for Global.
- The Local temperature is increasing faster than the Global temperature, with the standard deviation being 0.505 for Local and 0.460 for Global.
- The hottest years on record for both Local and Global temperature have happened during the last decade, with Local occurring in 2013 (16.23°C), and Global occurring in 2007 with (9.73°C).