Exploring Weather Trends:
Understanding the correlation between local and global average temperatures

In this weather trends project, I explore the correlation between local and global average temperatures by comparing weather trends in San Francisco against global average temperatures. In this analysis, I am looking to answer the following question: **how indicative global estimates are of local temperature averages?** I am solely focusing on weather in San Francisco for this exercise. I realize that including data points from other parts of the world would significantly help answer the question above, but for the sake of time, I decided to limit my universe of elements to a specific city.

I first started by joining the city_data and global_data schemas to view San Francisco and global average temperatures side-by-side.

```
SELECT c.*, g.avg_temp AS global_temp
FROM city_data c
JOIN global_data g
     ON c.year = g.year
WHERE c.city = 'San Francisco'
```

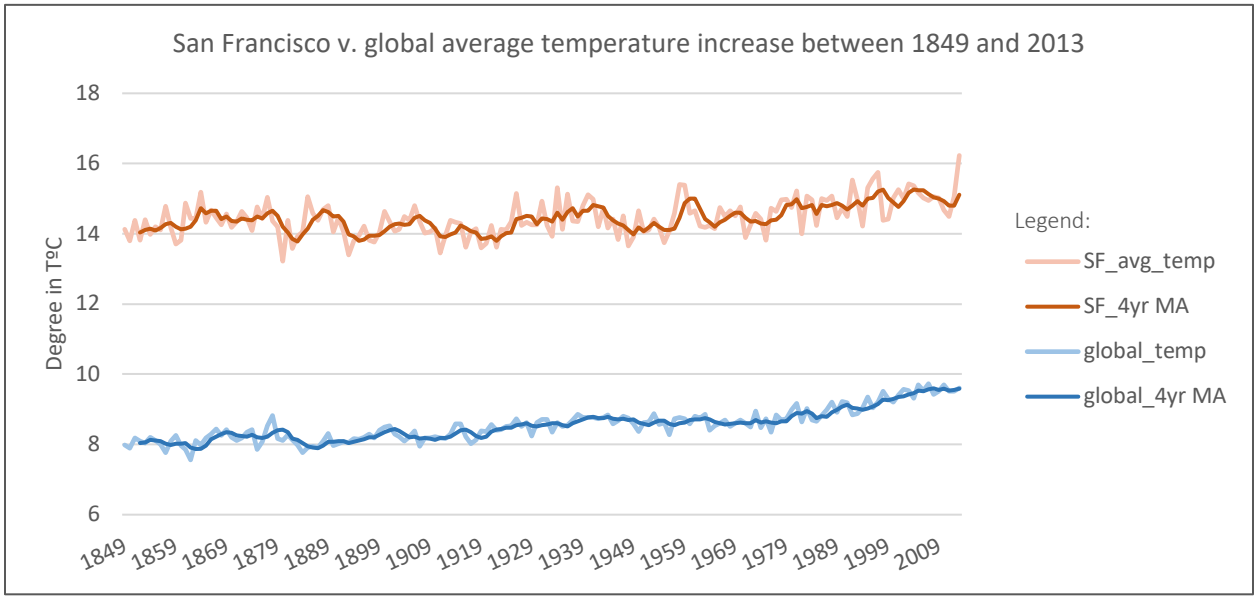
The query returned 165 rows, of which only the top ones are shown below:

year	city	country	avg_temp	global_temp
1849	San Francisco	United States	14.12	7.98
1850	San Francisco	United States	13.80	7.90
1851	San Francisco	United States	14.39	8.18
1852	San Francisco	United States	13.81	8.10
1853	San Francisco	United States	14.40	8.04
1854	San Francisco	United States	13.98	8.21
1855	San Francisco	United States	14.20	8.11

In the downloaded and saved .csv file, I computed a 4-year moving average (MA) of San Francisco and global average temperatures in Excel. I chose a relatively short time range to compute the moving average to avoid smoothing the curve out too much and, as a result, allow for the observation of variations in weather trends.

year	city	country	avg_temp	global_temp	MA_avg_temp	MA_global_temp
1849	San Francisco	United States	14.12	7.98		
1850	San Francisco	United States	13.8	7.9		
1851	San Francisco	United States	14.39	8.18		
1852	San Francisco	United States	13.81	8.1	=AVERAGE(D2:D5)	

I then created a Pivot Table to visualize the weather trends using a line chart. I plotted both the average temperatures and the moving averages of San Francisco and global temperatures.



As you can see, **the temperatures in San Francisco are consistently higher that the global estimates**, about 69% higher on average. That observation itself might lead us to believe that the weather in San Francisco is exceedingly hotter compared to the global norm. We mustn't forget, however, that there are cities with weathers much more blistering than San Francisco, like Bangkok, Timbuktu, Ahvaz, Kuwait City, or Phoenix.[1] Thus, the moving average of temperatures in San Francisco itself isn't statistically significant or relevant enough here to understand how global average temperatures are calculated.

A big difference we also note is the fact that local and global temperatures increase at a different rate. While temperatures in San Francisco, measured by the moving average, only increase by approximately 8% between 1849 and 2013, global temperatures increase much more rapidly, at 19% for the same time period. Hence, we need to look elsewhere to see the relationship between locally measured average temperatures of San Francisco and global average temperature estimates.

While we might not yet think that global estimates are an accurate representation of the reality, we notice from the graph that both lines shift upward. In order to see the relationship between the

[1] Based on a blog post by Jess MacDonald published on June 2019 entitled "The World's Hottest Cities." Link: https://www.tripsavvy.com/the-worlds-hottest-cities-4070053

two lines, we need to take a look at the correlation coefficient, an indicator that helps us measure the linear correlation between two variables. In this case, the coefficient of San Francisco and global temperature moving averages is approximately 0.71.[2] **The coefficient seems to indicate here a relatively strong positive correlation**, which is made evident by the **occasional synchronic movement of both lines**.

For instance, temperatures in San Francisco dropped by 0.73 ºC between 1879 and 1883 while global temperatures dropped by 0.28 ºC in the same timeframe. Between 1916 and 1919, San Francisco average temperatures dropped by approximately 0.3 ºC while global average temperatures dropped by ~0.24 ºC. If the downward trend is reflected in the moving average of global temperatures, it is fair to assume that San Francisco is not the only city with similar weather trends. Cities in other parts of the world must have also undergone relatively colder years. If we look at two cities with contrasting weathers, Khartoum, Sudan, and Omsk, Russia, we will observe the same downward and upward trends during certain time periods reflected in the moving average of global temperatures.

Conclusion

While we were able to point to differences and similarities in the data between San Francisco and global average temperatures, it is important to keep in mind that it is only one data point among 70792 included in the full dataset. We should not base our entire analysis on a single data point to retrieve statistically relevant insights. Instead, a multivariate analysis would be a better approach here. Calculating the global average temperature is a complex task, involving measurements at various geographic levels, with the possibility that there might not even be data due to the absence of weather stations in certain locations. In another analysis, it would be interesting to explore differences in weather trends between geographic areas containing clusters of cities to see how they reflect in the global temperature benchmark.

[2] The Pearson product-moment correlation coefficient is measured by dividing the covariance by the product of the two variables' standard deviations: https://www.investopedia.com/terms/c/correlationcoefficient.asp