Project Report: Explore Weather Trends

Steps

SQL Query to extract global temperature data:	SELECT * FROM global_data;

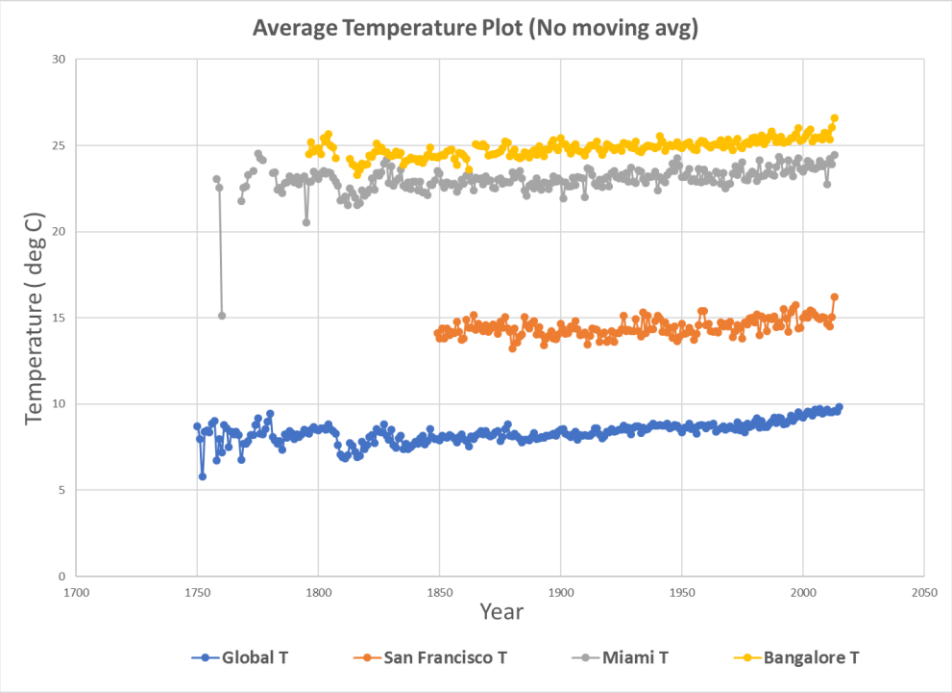
SQL Query to pull out cities in the US:	SELECT * FROM city_list ORDER BY country;

From the list of US cities, I picked San Francisco (biggest city closest to where I live) and Miami (because I'm interested in the analyzing the weather in Miami), and also picked Bangalore, a city in India.

SQL queries to extract city data:		
SELECT * FROM city_data WHERE city='San Francisco';	SELECT * FROM city_data WHERE city='Miami';	SELECT * FROM city_data WHERE city='Bangalore';

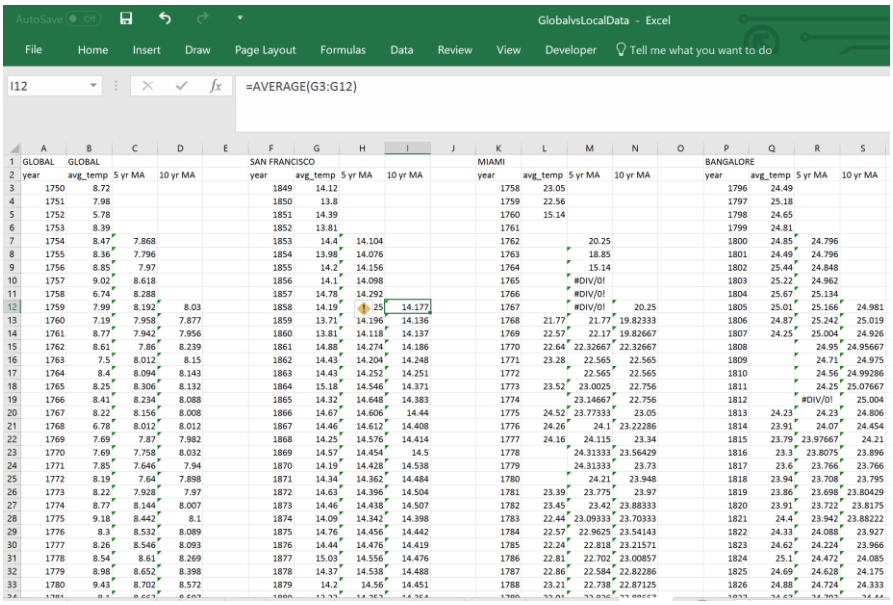
The data extracted using the queries was exported to Excel files.

Analysis in Excel: A graph of avg temp vs year was plotted to get an idea about the data:

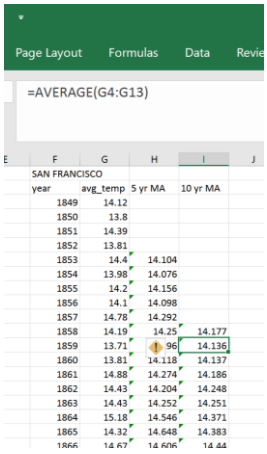


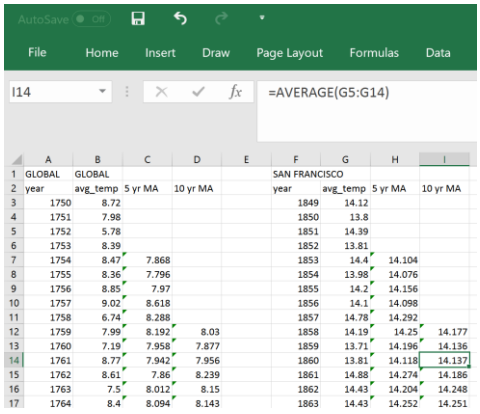
This graph shows that the temperature data for some years is missing, and calculating the moving average may help handle missing data and help observe the trend better

Moving average for 5 year and 10 year periods were calculated. I picked the 10 year moving average for the comparison plot since the 5 year moving average had points where division by zero was encountered due to missing data. It is assumed that the 10 year moving average is reasonable imputation to fill in the gaps in data.

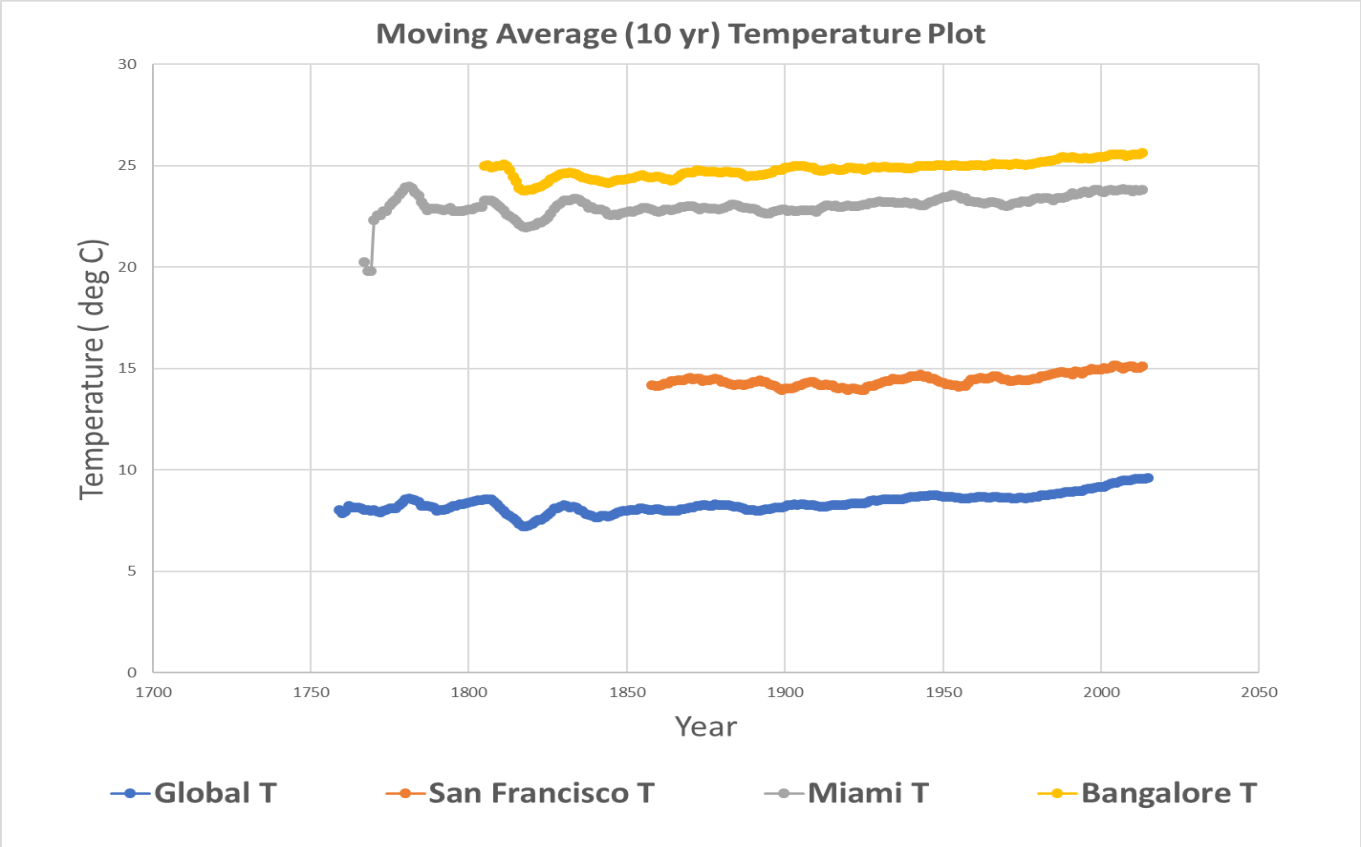


A few screenshots showing the calculation of the moving average from the spreadsheet:





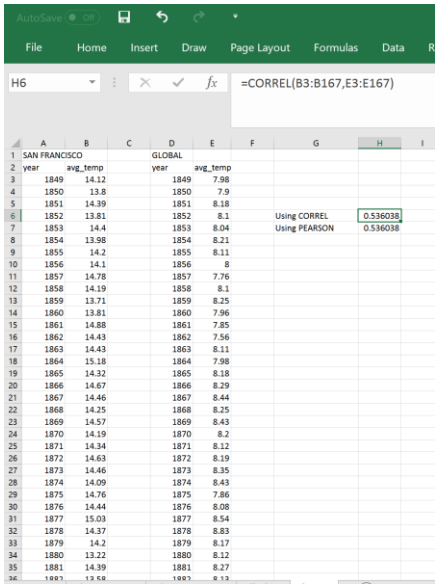
A line plot showing the 10 yr moving average temperatures vs year



Observations:

- The chosen cities (San Francisco, Miami and Bangalore) are hotter compared to the global average. In order of decreasing heat, Bangalore > Miami > San Francisco> global temperature.
- Bangalore and Miami are pretty close to each other with Bangalore being only slightly hotter than Miami, consistently over the years. SanFrancisco is significantly cooler than Bangalore and Miami.
- Over time, the average local temperatures in San Francisco , Miami and Bangalore have gradually increased and so has the global average temperature showing that the world has been getting hotter over the years.
- This trend has been consistent over the last few hundred years and there are no steep increases. Before 1850, the fluctuations in global temperature have been more. Around 2000, the global temperature increase is more evident while 1850 – 2000 has not seen too much variation.

Correlation Coefficient

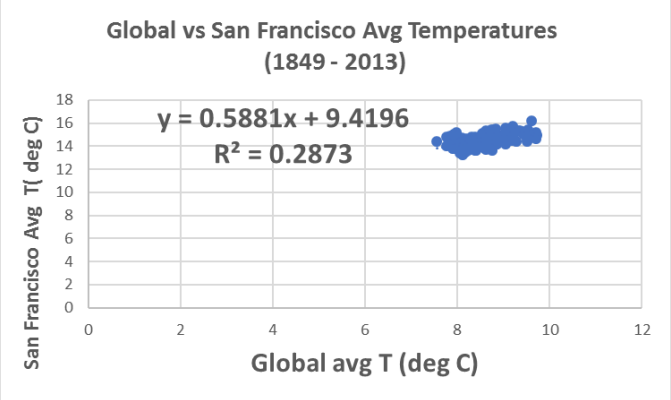


	A	B	D	E	G	H
1	SAN FRANCISCO		GLOBAL			
2	year	avg_temp	year	avg_temp		
3	1849	14.12	1849	7.98		
4	1850	13.8	1850	7.9		
5	1851	14.39	1851	8.18		
6	1852	13.81	1852	8.1	Using CORREL	0.536038
7	1853	14.4	1853	8.04	Using PEARSON	0.536038
8	1854	13.98	1854	8.21		
9	1855	14.2	1855	8.11		
10	1856	14.1	1856	8		
11	1857	14.78	1857	7.76		
12	1858	14.19	1858	8.1		
13	1859	13.71	1859	8.25		
14	1860	13.81	1860	7.96		
15	1861	14.88	1861	7.85		
16	1862	14.43	1862	7.56		
17	1863	14.43	1863	8.11		
18	1864	15.18	1864	7.98		
19	1865	14.32	1865	8.18		
20	1866	14.67	1866	8.29		
21	1867	14.46	1867	8.44		
22	1868	14.25	1868	8.25		
23	1869	14.57	1869	8.43		
24	1870	14.19	1870	8.2		
25	1871	14.34	1871	8.12		
26	1872	14.63	1872	8.19		
27	1873	14.46	1873	8.35		
28	1874	14.09	1874	8.43		
29	1875	14.76	1875	7.86		
30	1876	14.44	1876	8.08		
31	1877	15.03	1877	8.54		
32	1878	14.37	1878	8.83		
33	1879	14.2	1879	8.17		
34	1880	13.22	1880	8.12		
35	1881	14.39	1881	8.27		

Taking the avg temperature data for SanFrancisco , between 1849 – 2013, we have global and local data available for those years.

Using the CORREL () and PEARSON() functions in Excel, the correlation coefficient was found to be **0.53.**
The two variables were: Average global temperature and Average local temperature in San Francisco.

Predicting City temperature based on Global temperature



Using this equation where y is the local avg temperature, and x is the global avg temperature, the city temperature can be predicted based on the global temperature.

y = 0.5881x + 9.4196

The low R^2 value suggests that the fit is not great. A linear relationship does not relate the data very well.