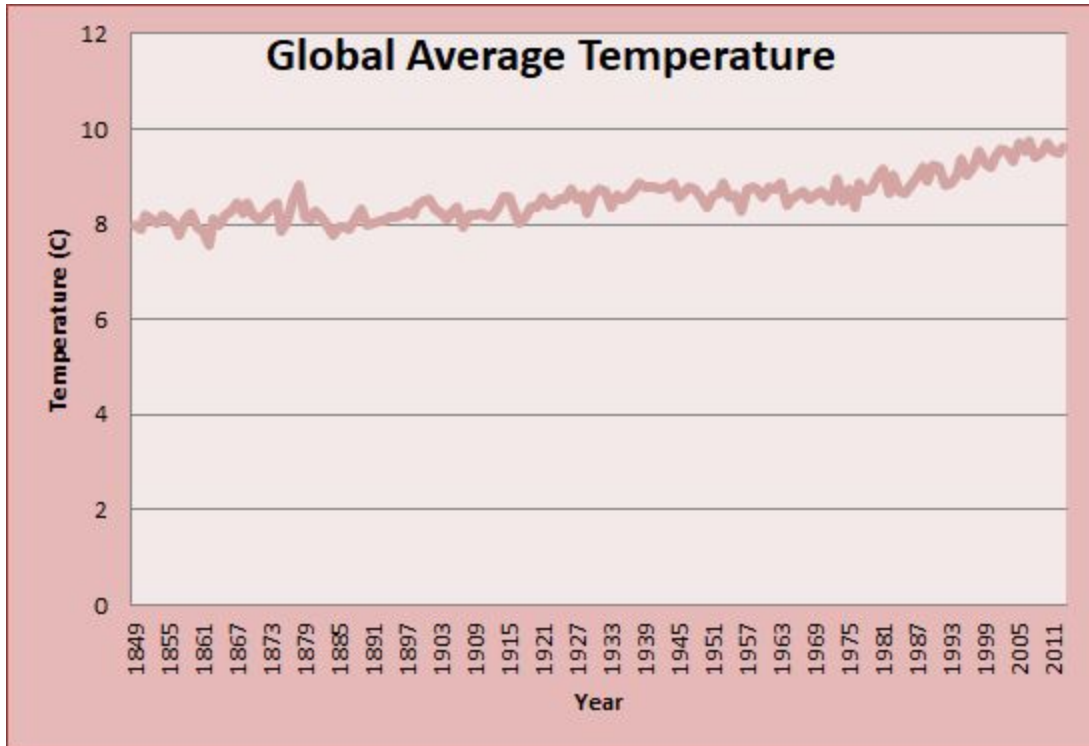# Explore Weather Trends

Project 1

Imesha Kuruppu

04.22.2020

Udacity-Data Analyst nano degree

Introduction

In this project, I have analyzed local temperatures of San Jose(USA) and global temperature data and compared the temperature trends in San Jose to global temperature trends.

Procedure

1. Extract data from the database and export it to a CSV file.
2. Analyze the data in the CSV file and calculate the moving average.
3. Create a line chart visualization based on data.
4. Derive conclusions based on observations.

Tools used

1. SQL - To extract data from the database.
2. Excel - Calculate the moving average, correlation coefficient, and plotting the line chart.
3. Google documents - Writing the project report.

Extract data from the database

I have used the following SQL query to extract relevant temperature data for San Jose and Global in respective years.

SELECT c.year,

c.avg_temp AS city_avg_temp,

g.avg_temp AS global_avg_temp

FROM city_data c

JOIN global_data g

ON c.year = g.year

WHERE c.city = 'San Jose'

Results were saved into a CSV file. When we open the CSV file in Excel it looks like:

A	B	E
year	city_avg_temp	global_avg_temp
1849	14.12	7.98
1850	13.8	7.9
1851	14.39	8.18
1852	13.81	8.1
1853	14.4	8.04
1854	13.98	8.21
1855	14.2	8.11
1856	14.1	8
1857	14.78	7.76

Figure 1 - Only the first few lines are displayed in the above table.

Calculate moving average

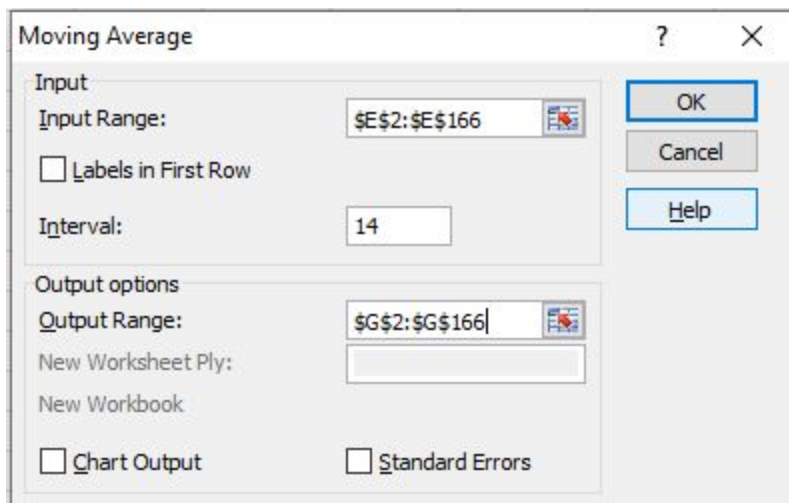I used Excel to calculate the 14-year moving average as below.

Figure 2 - Calculating moving average

After calculating the moving average, the results look as below.

	A	B	D	E	G
year		city_avg_temp	San Jose	global_avg_temp Global	
	1849	14.12	#N/A	7.98	#N/A
	1850	13.8	#N/A	7.9	#N/A
	1851	14.39	#N/A	8.18	#N/A
	1852	13.81	#N/A	8.1	#N/A
	1853	14.4	#N/A	8.04	#N/A
	1854	13.98	#N/A	8.21	#N/A
	1855	14.2	#N/A	8.11	#N/A
	1856	14.1	#N/A	8	#N/A
	1857	14.78	#N/A	7.76	#N/A
	1858	14.19	#N/A	8.1	#N/A
	1859	13.71	#N/A	8.25	#N/A
	1860	13.81	#N/A	7.96	#N/A
	1861	14.88	#N/A	7.85	#N/A
	1862	14.43	14.18571429	7.56	8
	1863	14.43	14.20785714	8.11	8.009285714
	1864	15.18	14.30642857	7.98	8.015
	1865	14.32	14.30142857	8.18	8.015
	1866	14.67	14.36285714	8.29	8.028571429
	1867	14.46	14.36714286	8.44	8.057142857

Figure 3 - only the first few lines are displaying in the above table.

Calculate the correlation coefficient

I used excel to calculate the correlation coefficient. The correlation coefficient indicates how strongly two variables are related to each other.
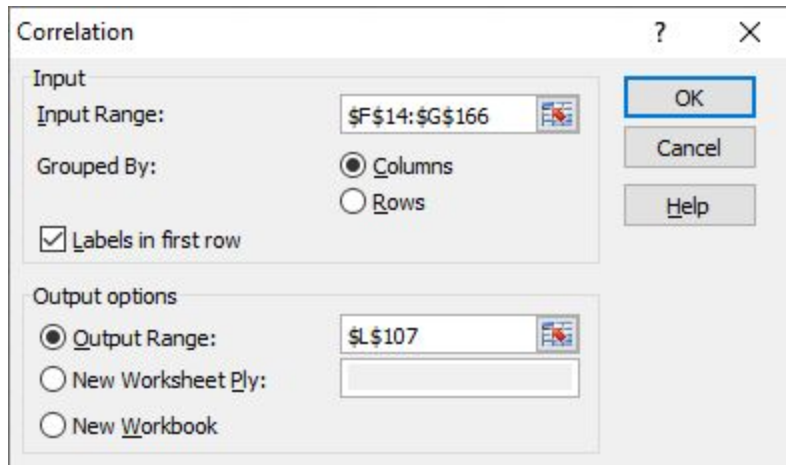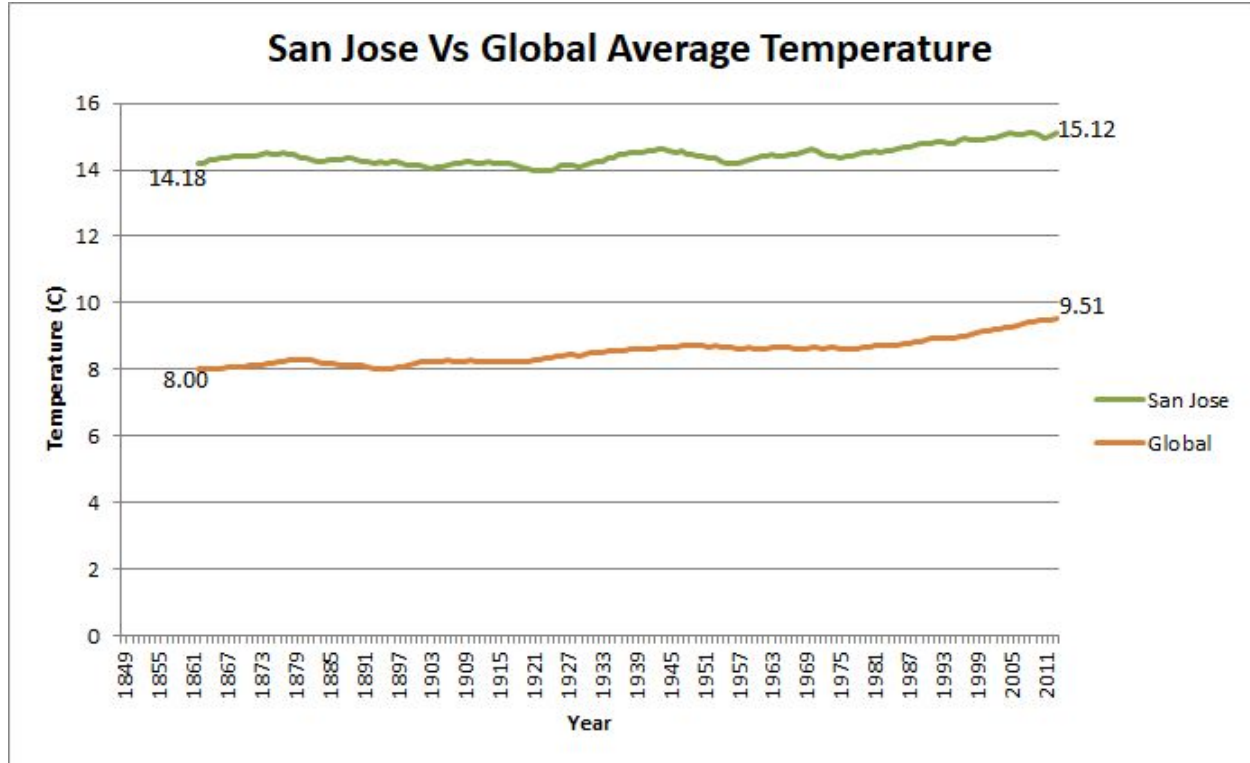

Figure 4 - Calculating correlation

Result :

	Global	San Jose
Global	1	
San Jose	0.837248	1

The correlation coefficient of San Jose and global temperatures is 0.8

Line Chart

The line chart was plotted using excel.

Observations

Observations from the above chart.

1. San Jose local temperatures are **always** higher than global average temperatures. This indicates San Jose city is hotter when compared to the global average. This is true from 1849 to 2013.
2. Both San Jose city temperatures and average global temperatures have increased over time.
3. The correlation coefficient of San Jose and Global average temperatures is 0.8. This value indicates a positive correlation. That means San Jose city temperatures increase when the global average temperature increases and vice versa.
4. San Jose city temperature increase ⇒ 15.12 - 14.18 = 0.98 C, Global average temperature increase ⇒ 9.51 - 8.00 = 1.51 C. According to the result, Global temperature increase is higher than the San Jose city temperature increase.

5. According to the line chart, the world is getting hotter and hotter. This trend has been consistent since 1849.

References

1. Correlation - https://www.excel-easy.com/examples/correlation.html
2. Moving Average - https://www.excel-easy.com/examples/moving-average.html
3. Line Chart - https://www.excel-easy.com/examples/line-chart.html