**Explore Weather Trends Project**

**Amelia Ho**

**Extracting Data:**

For the first step of the assignment I used various SQL queries in order to extract relevant data and download them into CSV files.

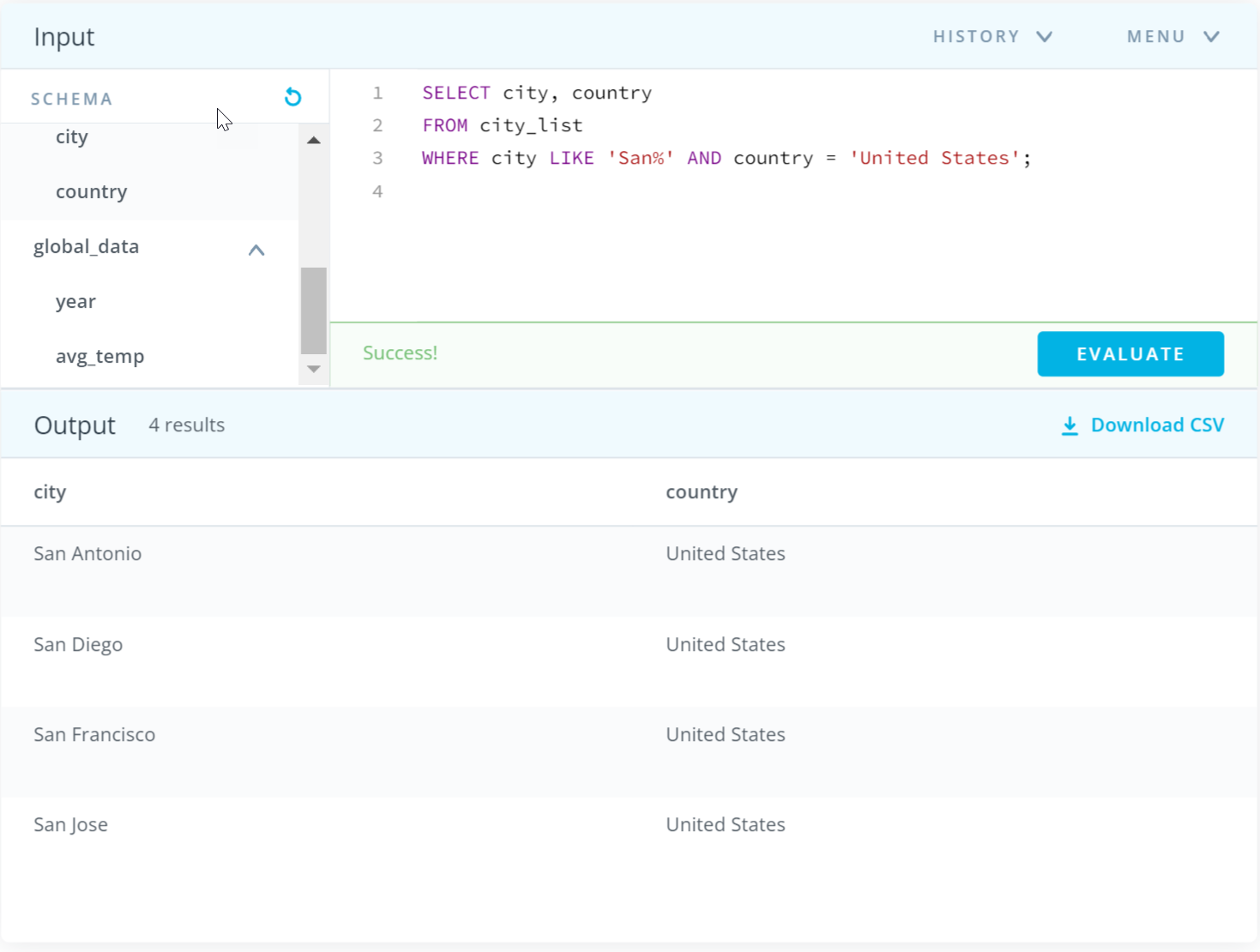
I needed to find the largest city closest to me. Since I knew that would be either San Francisco or San Jose (sometimes San Jose isn't well known enough to be considered a big city), I used this query to view cities that started with "San" in the United States. The results showed that there was San Antonio, San Diego, San Francisco, and San Jose. Since San Jose is closer to me, I picked San Jose as my city. I also wanted to make sure there was only one city that was named 'San Jose' and not another one in another country, so I ran the second query to double check that when I pulled data from the table city\_data, I would be pulling from the city I wanted San Jose, United States.

**Query 1:**

SELECT city, country

FROM city\_list

WHERE city LIKE 'San%' AND country = 'United States';



**Query 2:**

SELECT city, country

FROM city\_list

WHERE city = 'San Jose';



Since I already know that San Jose is in the United States, I decided not to extract the country column and only extracted the columns for city, year, and average temperature. Because the year and average temperature are in a different table (city\_data) then city\_list, I used a join statement to connect the 2 tables. I then downloaded the data from this query into a CSV.

**Query 3:**

SELECT cl.city, cd.year, cd.avg\_temp

FROM city\_list cl

JOIN city\_data cd

ON cl.city = cd.city

WHERE cl.city = 'San Jose';

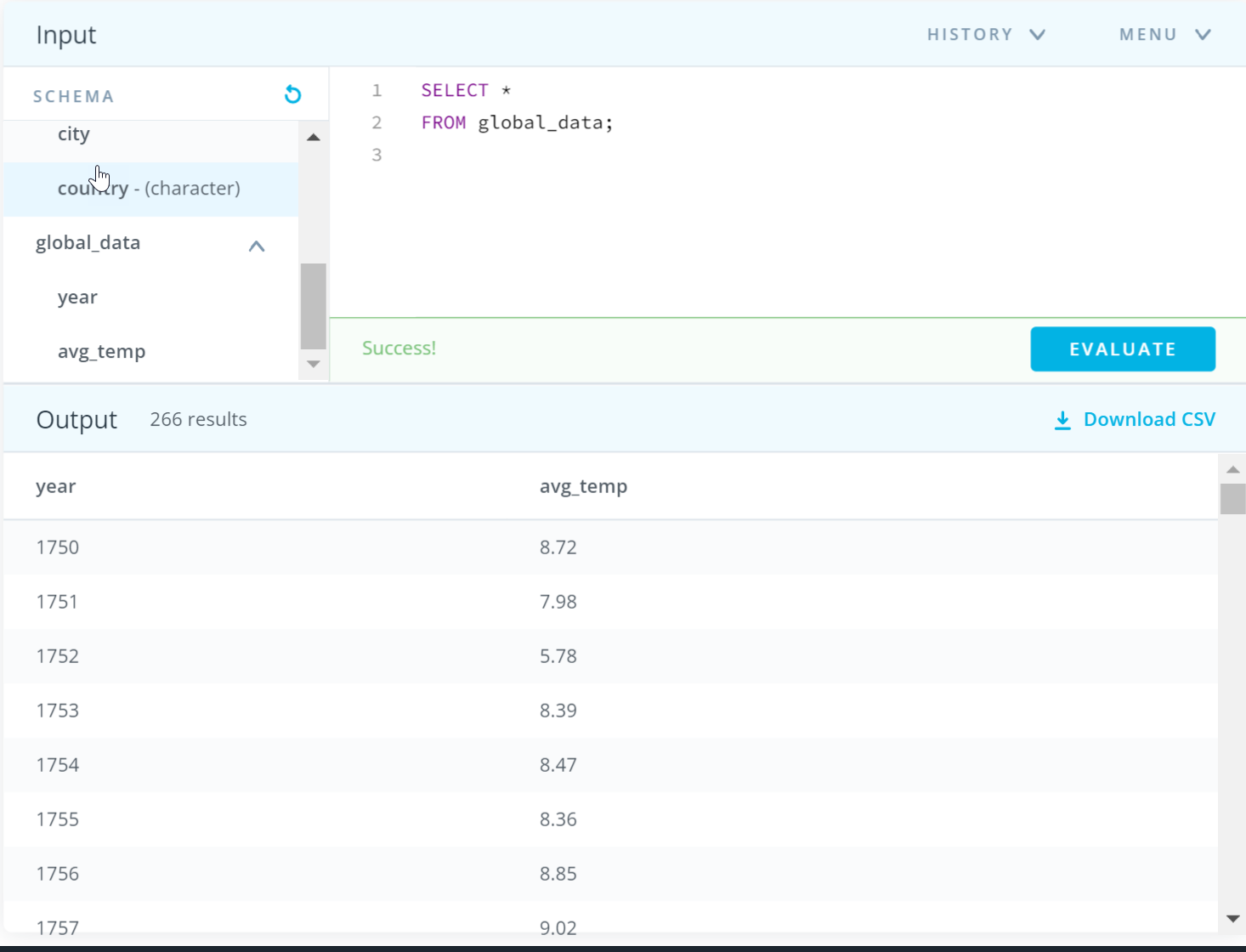


To extract the global data, I simply selected all the columns in the global\_data table and downloaded it as a csv using this query:

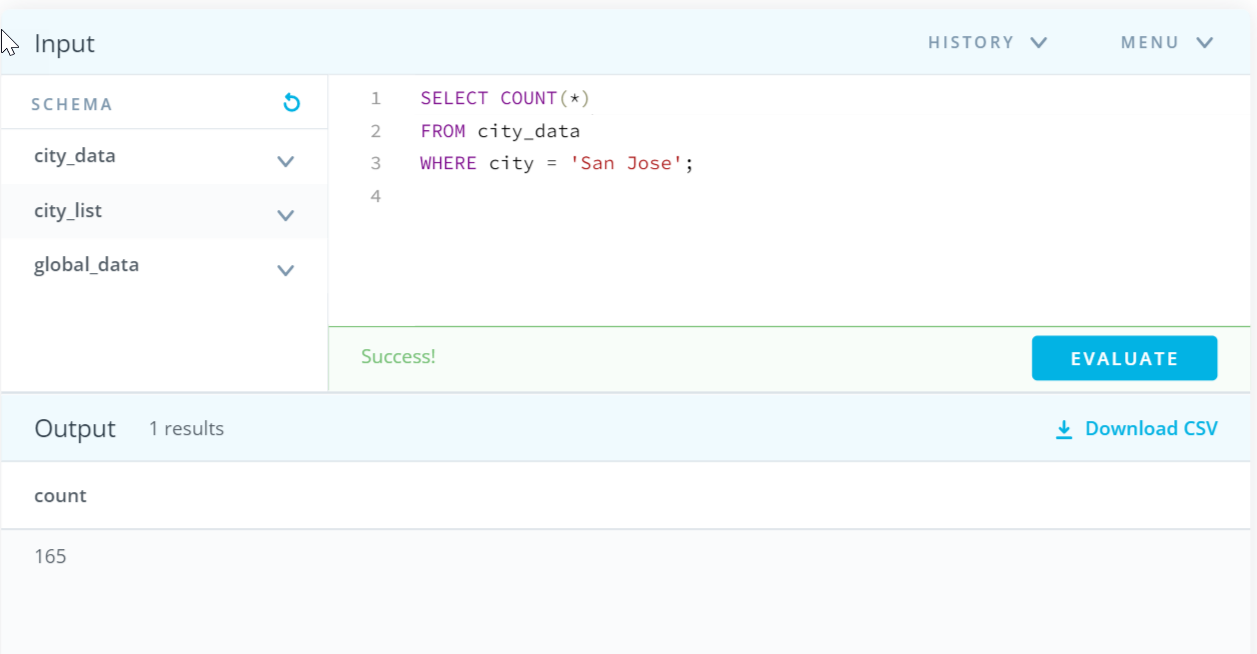
**Query 4:**

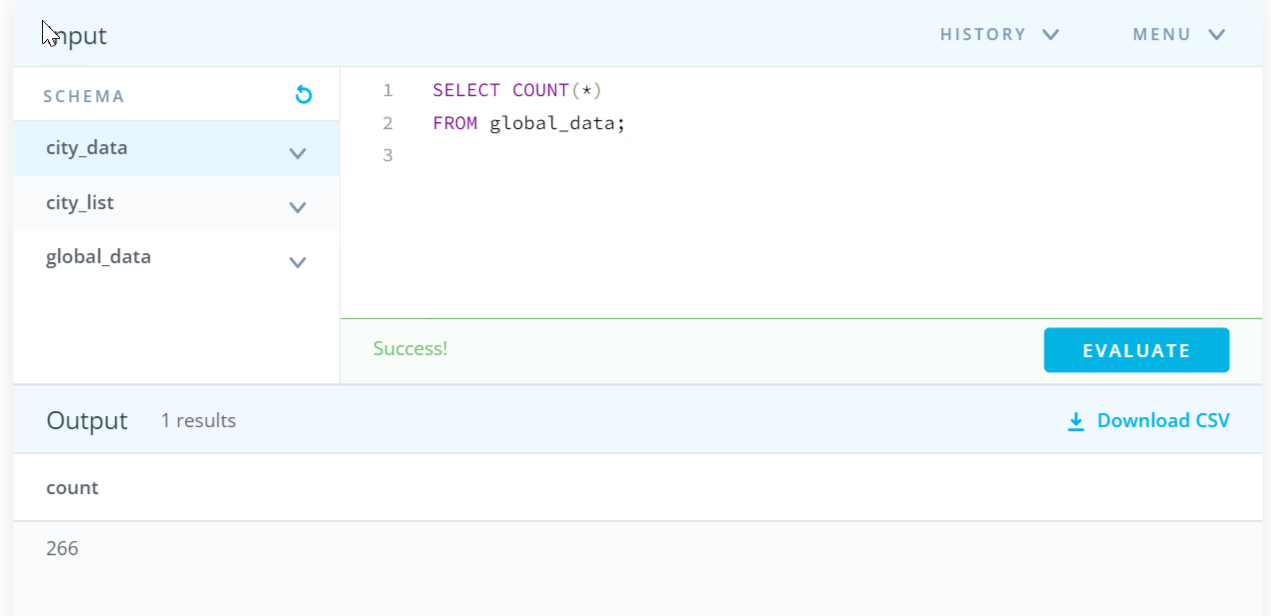
SELECT \*

FROM global\_data;



I converted them into Excel files so I could work on the second part of the assignment, creating a line chart based off of the global and city data averages. After looking at the data, I noticed that the tables didn’t have the same number of rows. To check that I didn’t miss any rows from the table when downloading to CSV, I did a quick COUNT query for both the city and global data.



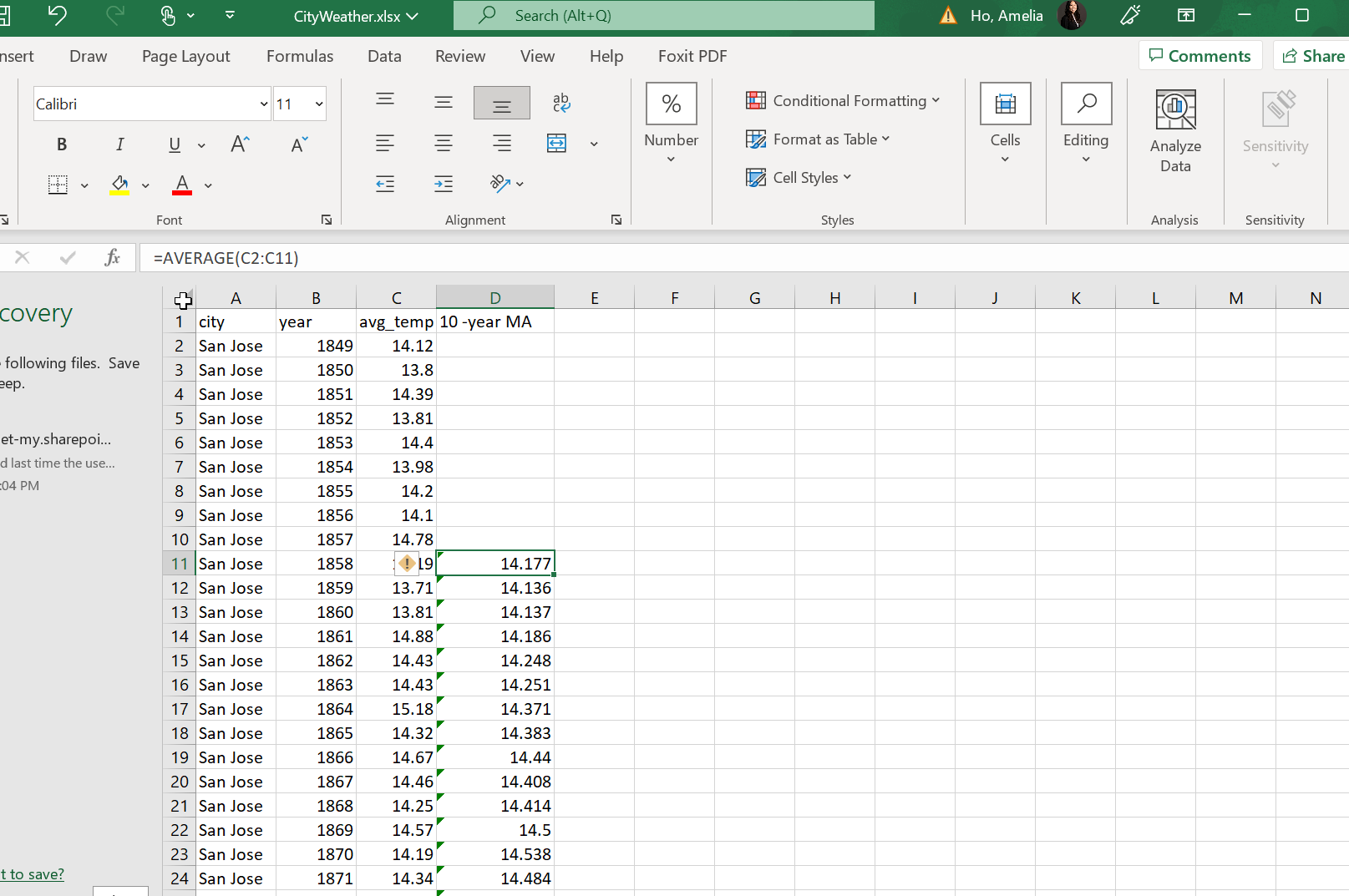


**Calculating the Moving Average:**

Using what I learned in the previous lesson on Moving Averages, I used the Average() function to find the 10-year moving averages for both city and global data. I chose a 10-year moving average specifically because that is the length of a decade, which I thought was a significant period of time.

**City Data Excel File**

Formula for Moving Average -> = AVERAGE(C2:C11)

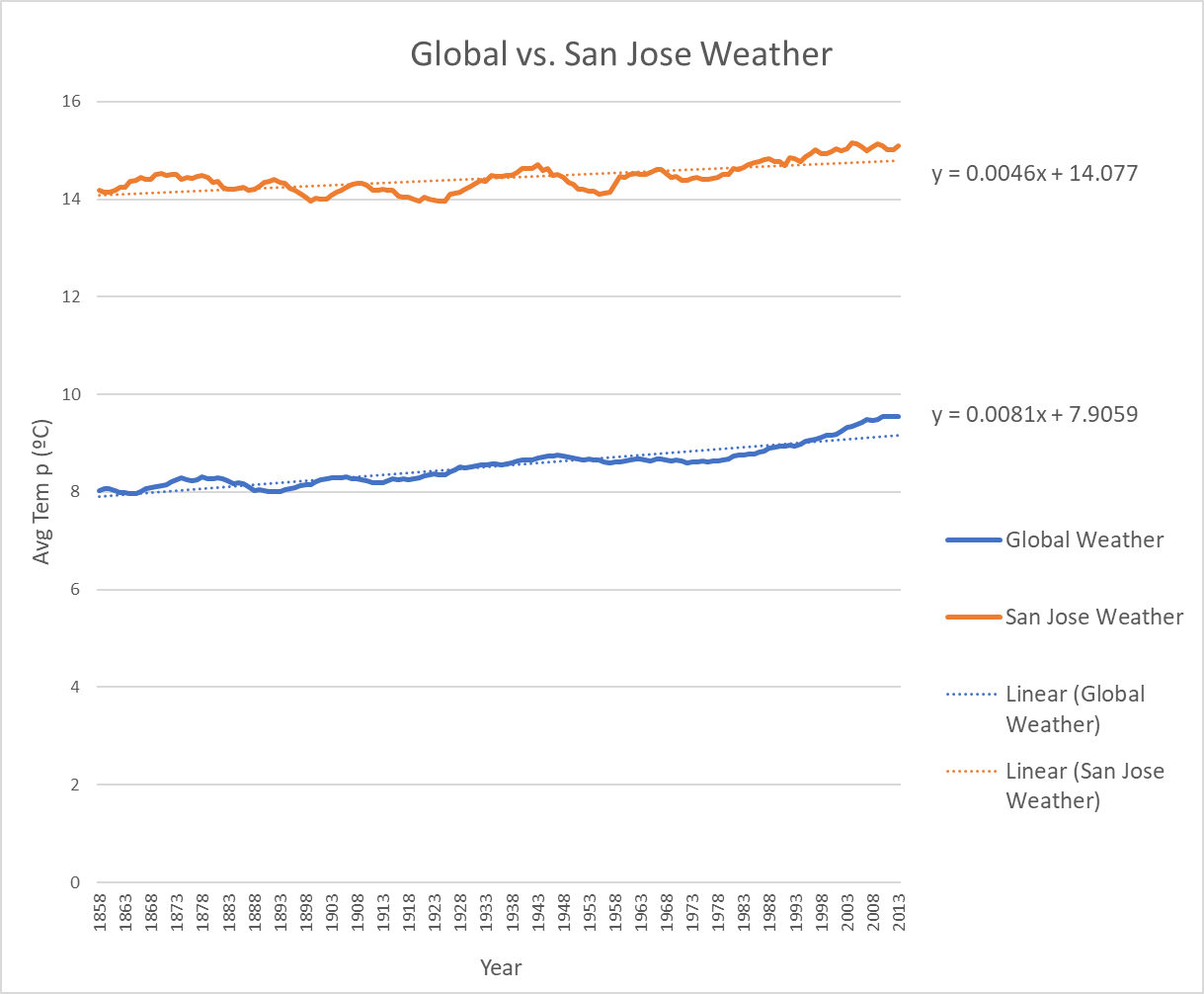
****

**Global Data Excel File**

Formula for Moving Average -> = AVERAGE(B2:B11)

**Line Chart and other considerations:**

Some of the key considerations I had regarding visualizing the data came from dataset size, readability, and purpose. Because the data for San Jose had less rows than the global data, I created a line chart based on how many years of data San Jose had (the date range was from 1858 - 2013). Even though the 10-year moving average for global data starts in 1759, the 10-year moving average for San Jose begins in 1858, so I started the line chart in 1858 and ended it in 2013 (even though global data had data up to 2015). To make it obvious the lines were from two different datasets, I made them contrasting colors of orange and blue. I added axis titles, a legend, and a chart title for clarity. I also added the temperature units on the axis so the audience knew it was Celsius and not Fahrenheit. Additionally, I plotted a linear trendline for both data sets so I could compare how much each dataset increased in temperature. I chose a linear trendline so I would be able to make an approximate prediction for weather in further years.



**Observations:**

Since data was first collected, the general trend for the weather has been positive, meaning that the world is getting hotter over time. This trend has been consistent for the last hundred years or so. This could be attributed to a number of factors such as global warming, build up of greenhouse gasses, and deforestation. San Jose, United States has been consistently hotter than the global temperature by around 6 degrees Celsius. Based on the trendline correlation coefficient for global and San Jose weather, global weather has been increasing faster than San Jose weather (0.0081 versus 0.0046). This means the world is heating up quicker compared to how much San Jose has been heating up. Looking at the line chart, the moving averages of San Jose have more dips in the line chart compared to globally, so the weather varies more in San Jose generally than in the world historically.

**Reviewer suggested links:**

[**Calculating Moving Average in Excel [Simple, Weighted, & Exponential] (trumpexcel.com)**](https://trumpexcel.com/moving-average-excel/)

[**How to Create a Chart From Start to Finish (visme.co)**](https://visme.co/blog/create-a-chart/)