

Extracting Data from SQL database

The first step is to list all the cities available in the city_list table and to pick the city that is closest, to use for local temperature trend.

I ran the following query to look at cities in United States, and picked San Jose as the city of reference, for local temperature after scanning through the results.

```
SELECT *  
FROM city_list  
WHERE country = 'United States';
```

Having picked San Jose, we now move to the city_data table and use the below query to extract all available temperature data for this city.

```
SELECT *  
FROM city_data  
WHERE city = 'San Jose';
```

We notice that this city has data from 1849 and 2013 whereas SELECT * FROM global_data has data from 1750 to 2015. So, we also dump out another CSV using the BETWEEN clause to extract global_data for the range for which we have the local temperature data, to be able to inspect that range better.

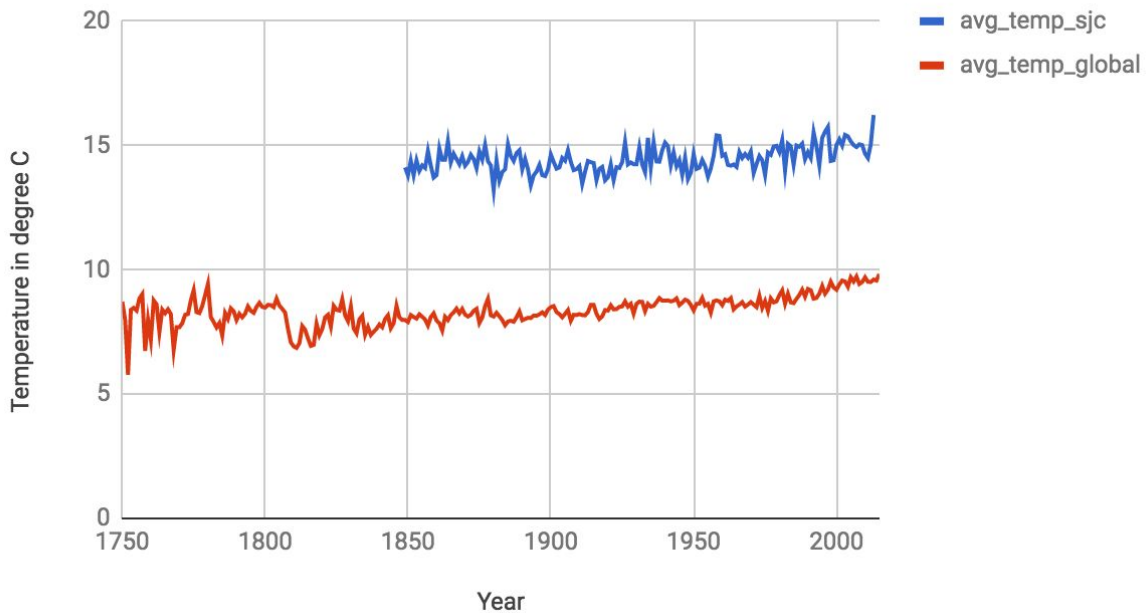
```
SELECT *  
FROM global_data  
WHERE year BETWEEN 1849 AND 2013;
```

Manipulating data and creating visualization

We open global temperature csv file in google spreadsheet and then import the second csv file (sjc temperature) by pointing to the row corresponding to 1849 of next available column using the 'replace at cell' option. This allows us to inspect both series in the same spreadsheet.

Initially we plot a line graph of the global temperature vs SJC temperature to see how noisy the data is and over what range we need to do the moving average, to achieve sufficient smoothing.

San Jose vs Global temperature trends



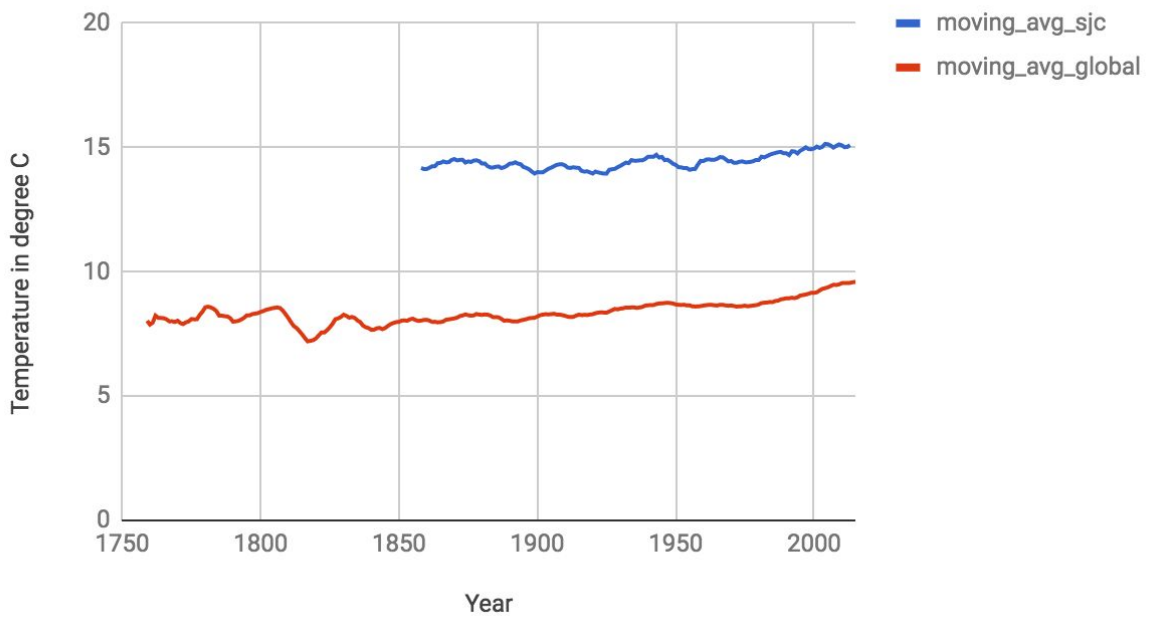
We now start out creating the moving average over two cells using the AVERAGE function, and keep increasing the moving window size, while inspecting the line graph visually, until we see that sufficient smoothing has been achieved.

For our data, this window size is 10 years.

This is the formula in the initially cell, which is applied to other rows by dragging down.

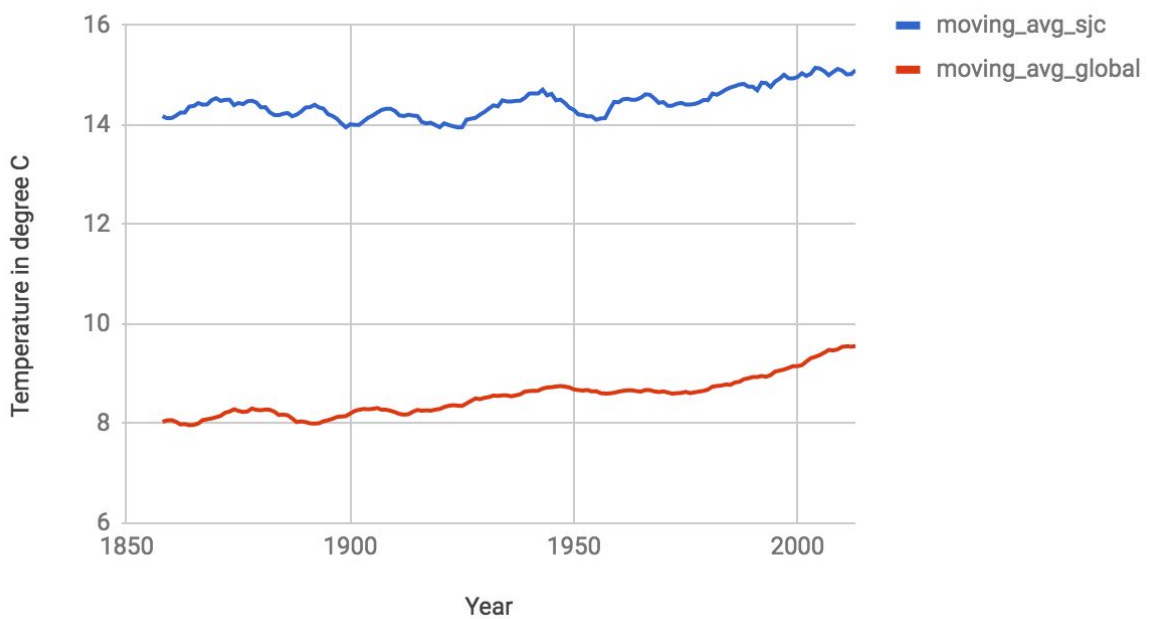
=AVERAGE(D2:D11)

San Jose vs Global temperature 10 yr moving avg trend



Zoomed view for the years 1849 and 2013.

San Jose vs Global temperature 10 yr moving avg trend



Interpreting the data visualization

- 1) On average, the temperature in San Jose is about 6 C higher than the global average temperature. The gap seems to be more or less consistent, the shrinkage in the gap over the past 100 years is barely noticeable.
- 2) Overall, the temperatures seem to be rising, having gone up almost 2C, from 1850 to 2015. The rise in San Jose temperature is slightly slower, but it also goes up overall.
- 3) The correlation coefficient is pretty high (~ 0.8) between the two trends, however San Jose seems to be coping slightly better than other places in the world, which might be responsible for fact that San Jose temperatures haven't risen as fast as the global temperatures of late.
- 4) The fact that window size of 10 years worked well, also shows how noisy the temperature trends can be in the short term, and this could be vital to understanding how people experience temperature trends, and how the noise might mislead and misinform any policy that is based on the unsmoothed noisy raw data.