Project 1: Explore Weather Trends

by Patricia Chandía

For this project the following databases were available: city_list, city_data and global_data; each one of them includes a few columns with some variables that I used.

First, using SQL queries I looked for the data of Chileans cities:

```
SELECT *
FROM city_list
WHERE country = 'Chile';
```

The city nearest to my address was Santiago de Chile, but before collect all the data associate to Santiago in the city_data database, I realized that there are some other cities called Santiago in other countries like Dominican Republic and Phillippines, so I added a second condition to my selection:

```
SELECT *
FROM city_data
WHERE city = 'Santiago' and country = 'Chile';
```

Then, I downloaded the .csv file and I got 159 results with the average temp from year 1855 to 2013.

Now, for the global temperatures, I selected the data from global_data database:

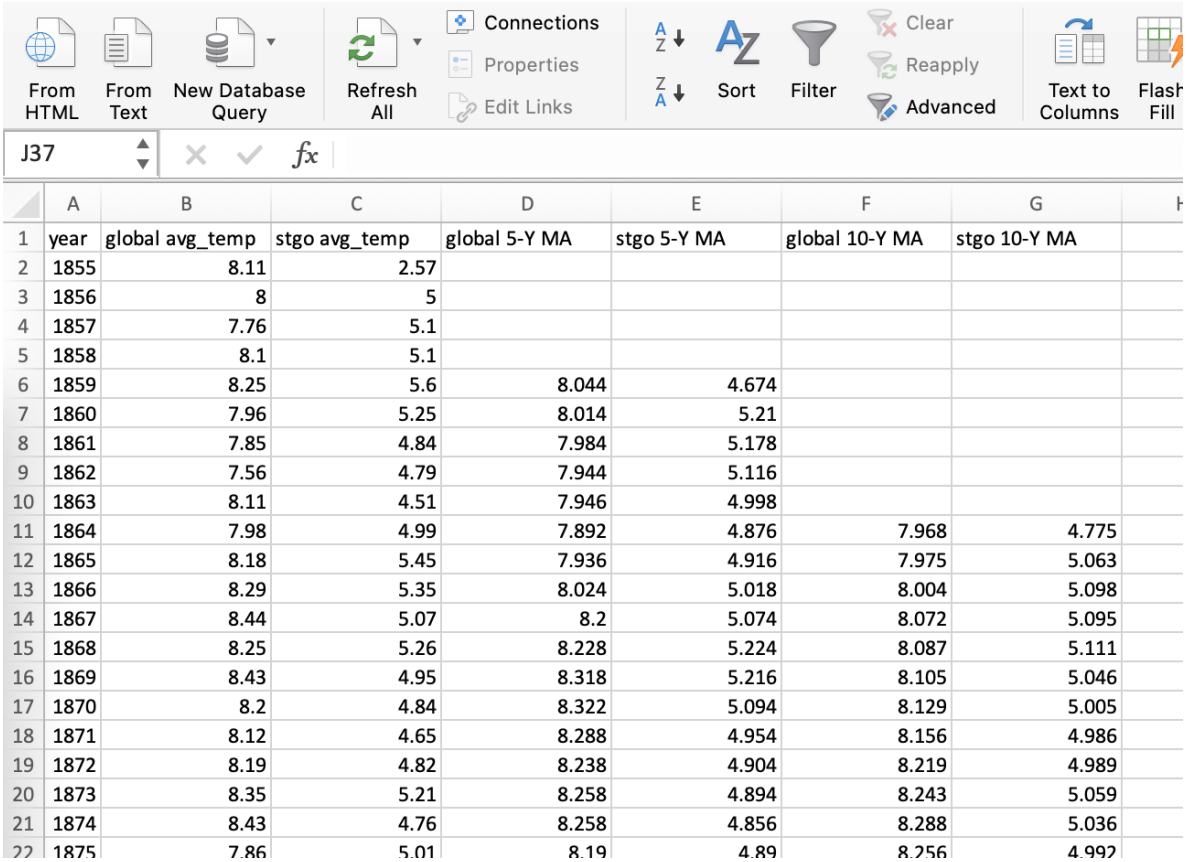
Then I select
```
SELECT *
FROM global_data;
```

But the global data starts in 1750 while Santiago data is from 1855 to 2013, so I apply a filter as well:

```
SELECT *
FROM global_data
WHERE (year >= 1855 and year <= 2013);
```

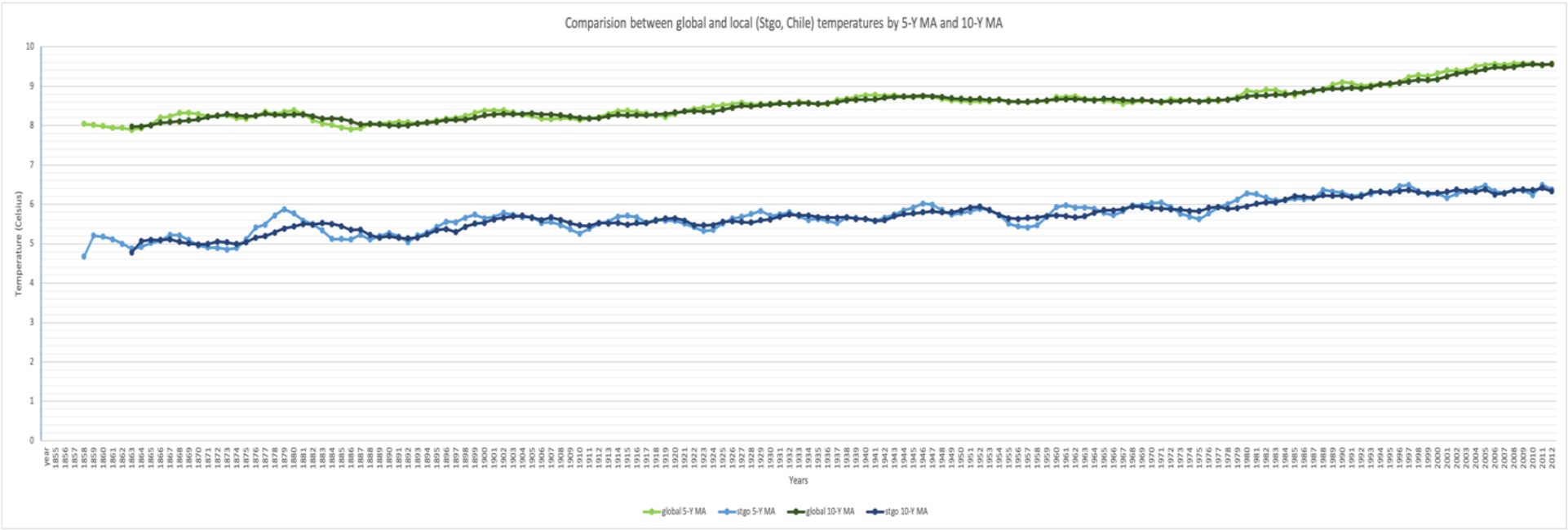
And finally, I obtained the same years to compare and I downloaded this second .csv file.

In order to continue with the analysis, I used Excel, so I opened both .csv files from data extract From Text tool in Excel and rename the columns B and C as global avg_temp and stgo avg_temp respectively and then I added new columns to calculate the moving averages (AVERAGE function) for global and local temps in range of 5 years and 10 years in columns D and E:



	A	B	C	D	E	F	G
1	year	global avg_temp	stgo avg_temp	global 5-Y MA	stgo 5-Y MA	global 10-Y MA	stgo 10-Y MA
2	1855	8.11	2.57				
3	1856	8	5				
4	1857	7.76	5.1				
5	1858	8.1	5.1				
6	1859	8.25	5.6	8.044	4.674		
7	1860	7.96	5.25	8.014	5.21		
8	1861	7.85	4.84	7.984	5.178		
9	1862	7.56	4.79	7.944	5.116		
10	1863	8.11	4.51	7.946	4.998		
11	1864	7.98	4.99	7.892	4.876	7.968	4.775
12	1865	8.18	5.45	7.936	4.916	7.975	5.063
13	1866	8.29	5.35	8.024	5.018	8.004	5.098
14	1867	8.44	5.07	8.2	5.074	8.072	5.095
15	1868	8.25	5.26	8.228	5.224	8.087	5.111
16	1869	8.43	4.95	8.318	5.216	8.105	5.046
17	1870	8.2	4.84	8.322	5.094	8.129	5.005
18	1871	8.12	4.65	8.288	4.954	8.156	4.986
19	1872	8.19	4.82	8.238	4.904	8.219	4.989
20	1873	8.35	5.21	8.258	4.894	8.243	5.059
21	1874	8.43	4.76	8.258	4.856	8.288	5.036
22	1875	7.86	5.01	8.19	4.89	8.256	4.992

The medium averages were calculated in range of 5 and 10 years so the curve could be smoothed and then graphed into a line chart:



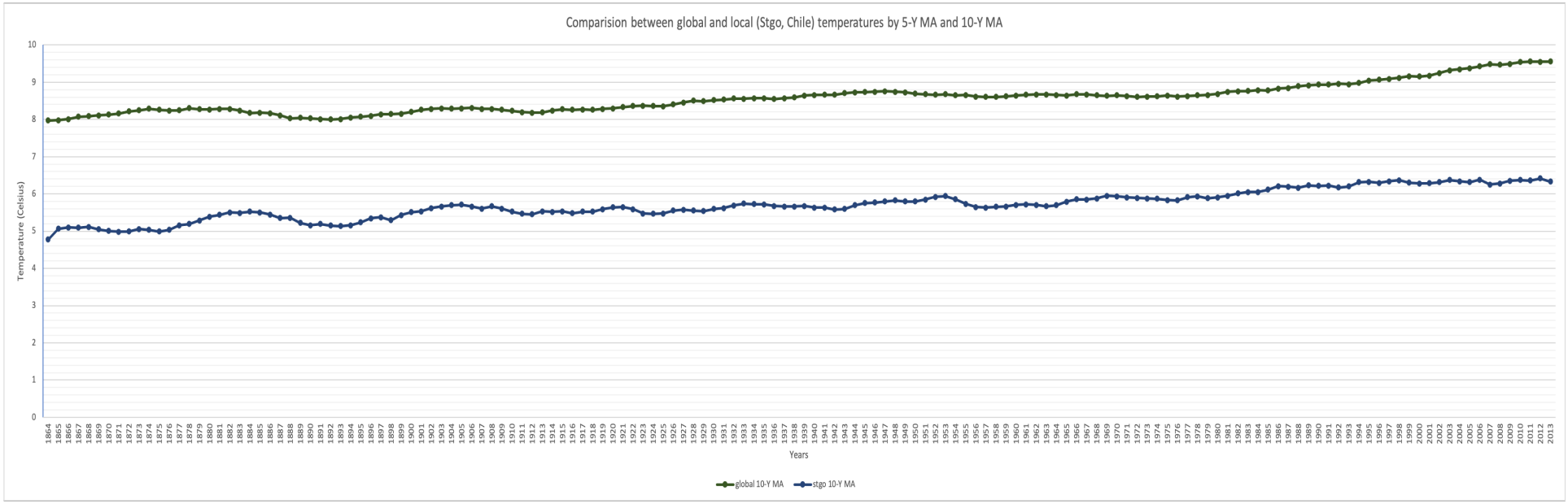
The darker lines show the 10 Years Moving Average for Global and Local (Stgo) temperatures so it is so much smoother than the 5 Years Moving Average and it allows to visualize the trends better.

In the same way, I did a brief academic review and found that there are some articles on climatology that use the 10 Years moving average for temperature or precipitation variables to study climate change related topics[1] and the World Meteorological Organization uses de decade criteria to compare global temperatures[2].

[1] Pulido-Velásquez, M. et al. (2015). Integrated assessment of the impact of climate and land use changes on groundwater quantity and quality in the Mancha Oriental system (Spain).

[2] https://public.wmo.int/en/media/press-release/2019-concludes-decade-of-exceptional-global-heat-and-high-impact-weather

For this analysis I selected the 10 Years Moving Average and here are the conclusions:



Observations:

1. Consistently, the temperature using the 10 years moving average for Santiago de Chile is always lower about 3 Celsius degrees over time than the global 10 years moving average.
2. Anyway, the trends of both, global and local, data shows an increase in the average temperature and it shows clearly in big periods of time.
3. Another aspect to consider is that the increase seems to be faster in the last decades, as is showed if we look at the behavior of the slope.
4. The average for global temperature always will be smoother because it takes data from all over the world and the local temperature shows the effect of climate anomalies as “Fenómeno de la Niña” and “Fenómeno del Niño”.
5. It seems to be a correlation between the both, global and local, temperature, but we cannot forget that we are working with moving averages.

Extra:

Correlation

I used the CORREL function in Excel to calculate if there is some correlation between global and local temperatures and maybe think about a predictive model:

	L	M
	0.69215272	YEARLY
	0.87396175	5-Y MA
	0.91047068	10-Y MA

I obtained those results and is obvious that the correlation using the 10 years moving average is super high (0.9) because it's already normalized by the moving average procedure. Anyway, if we use the yearly data, the correlation goes down to 0.69 so it makes me think that it would be a low satisfactory predictive model.