Exploring weather trends project

1) Objective of the project:

The project is a case study that gives data about annual averages temperature from different locations or cities all over the world, and global yearly average data temperature. I have to choose the nearest city to my location and try first to analyst its temperature trends over the years, and then compare it to the global temperature trends.

Since I'm from morocco, I looked over city list Table, and I found the two major cities in my country, Rabat the capital and Casablanca. Rabat is an administrative city; it's big but not as big as Casablanca. The last one, we call it the economic capital of morocco it has a population of about 4 million and lot of industrial zones, so it has a big impact on climate change in comparison to Rabat. For that, my study will be about Casablanca.

2) Extract the data:

2.1) City data:

I ran a simple query to check all columns in the city data table, as shown in fig1:

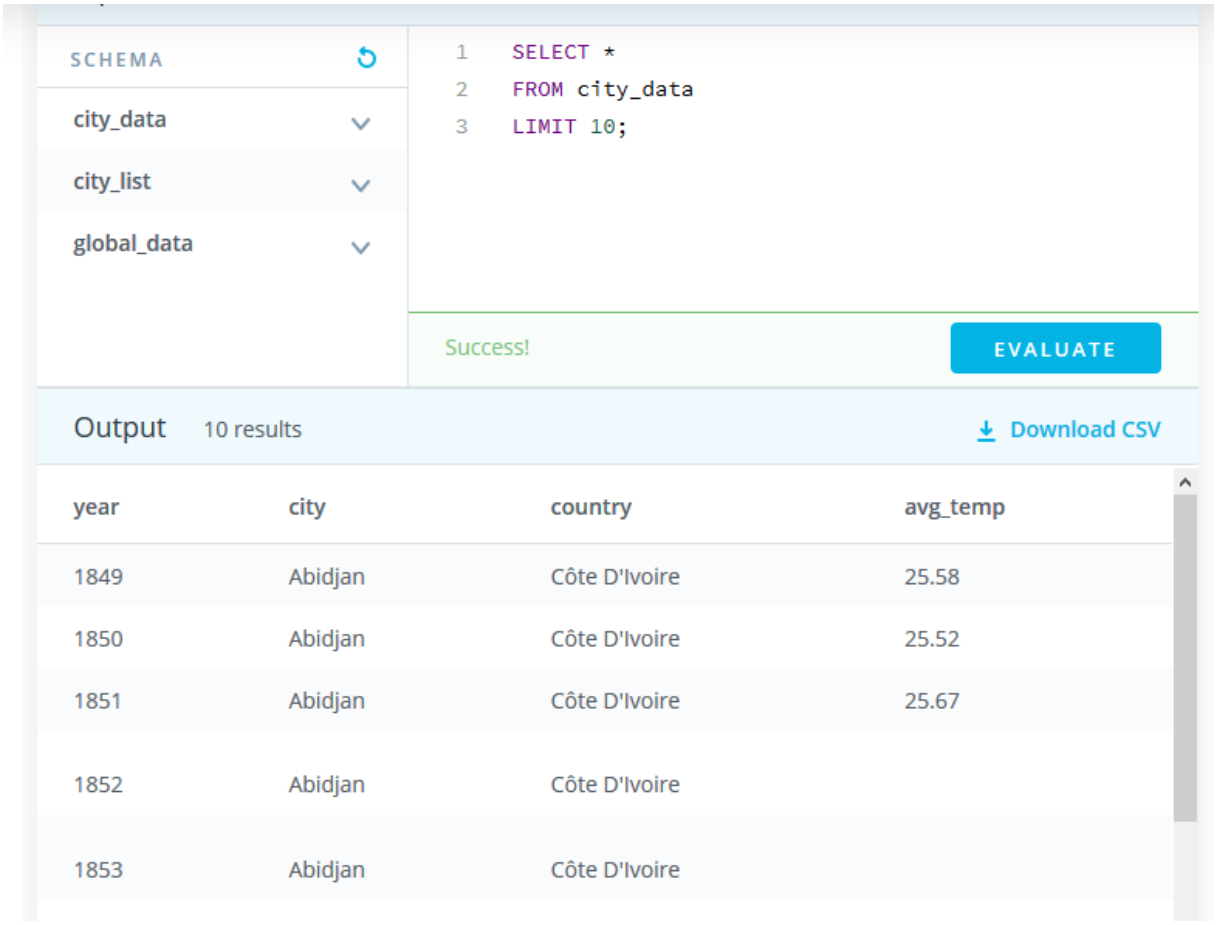


Fig1: Checking columns in city data table

The first note, there is NULLs in the result table, so I have to deal with missing information. For that I ran another query to see the size of missing data regarding Casablanca, as shown in fig2.

Input HISTORY MENU

SCHEMA

city_data

city_list

global_data

```
1  SELECT *
2  FROM city_data
3  WHERE country = 'Morocco' AND city = 'Casablanca'
4  ORDER BY year;
```

Success! EVALUATE

Output 251 results Download CSV

year	city	country	avg_temp
1763	Casablanca	Morocco	16.36
1764	Casablanca	Morocco	18.95
1765	Casablanca	Morocco	17.27
1766	Casablanca	Morocco	17.34

Fig2: Checking missing data for Casablanca

Returned table had missing data for 1775, 1781 and 1782 only, which, I think, is not a big deal, so I decided to omit those lines.

The final query is shown in Fig3, I ran it and I got my CSV file.

Input HISTORY MENU

SCHEMA

city_data

city_list

global_data

```
1  SELECT *
2  FROM city_data
3  WHERE country = 'Morocco' AND city = 'Casablanca'
   AND avg_temp IS NOT NULL
4  ORDER BY year;
```

Success! EVALUATE

Output 248 results Download CSV

1769	Casablanca	Morocco	16.96
1770	Casablanca	Morocco	16.82
1771	Casablanca	Morocco	17.33
1772	Casablanca	Morocco	17.72
1773	Casablanca	Morocco	17.07
1774	Casablanca	Morocco	17.72

Fig3: Getting yearly averages temperature of Casablanca

2.2) Global data:

I ran another query to get global temperature data from global data table as shown in Fig4.

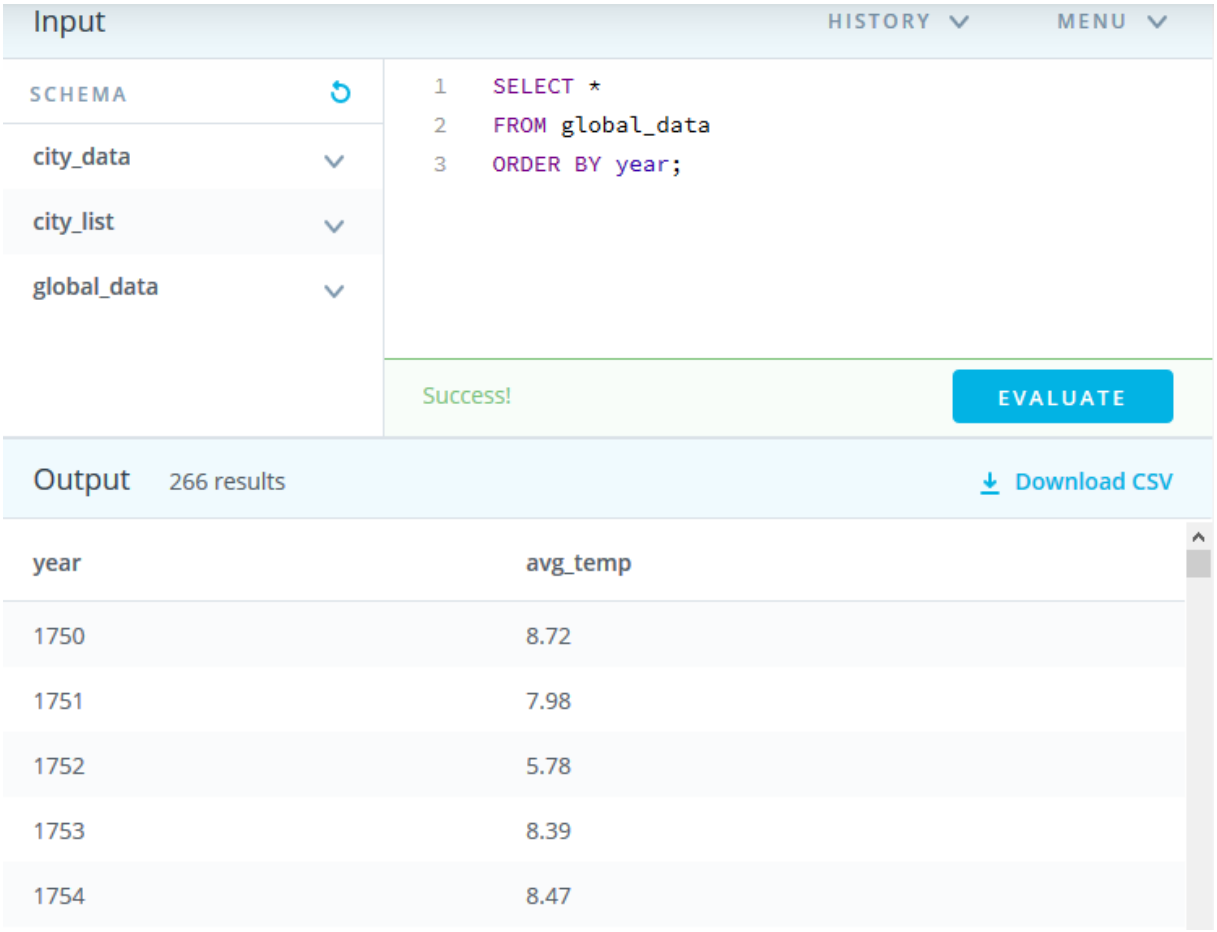


Fig4: Getting data from global data table

I checked the returned table and I found no missing data (fortunately).

So I exported the global temperature data as a CSV file.

I needed another table that combines the average global temperature and Casablanca's average temperature in order to plot and compare the data from the two sources. So I ran the query shown in fig5.

Input HISTORY MENU

SCHEMA

city_data

city_list

global_data

```
1  SELECT cd.year, cd.city, cd.avg_temp
   city_avg_temp, gd.avg_temp global_avg_temp
2  FROM city_data cd
3  JOIN global_data gd
4  ON gd.year = cd.year AND cd.country = 'Morocco'
   AND cd.city = 'Casablanca'
5  WHERE cd.avg_temp IS NOT NULL
6  ORDER BY cd.year;
```

EVALUATE

Output 248 results Download CSV

year	city	city_avg_temp	global_avg_temp
1763	Casablanca	16.36	7.50
1764	Casablanca	18.95	8.40
1765	Casablanca	17.27	8.25
1766	Casablanca	17.34	8.41
1767	Casablanca	17.28	8.22

Fig5: Joining the two tables (city data and global data)

3) Create a line chart:

3.1) City chart:

I used EXCEL to export the CSV files and to plot the data. I started by plotting the yearly averages temperature of Casablanca, using the yearly averages data, and I got the figure below (fig6):

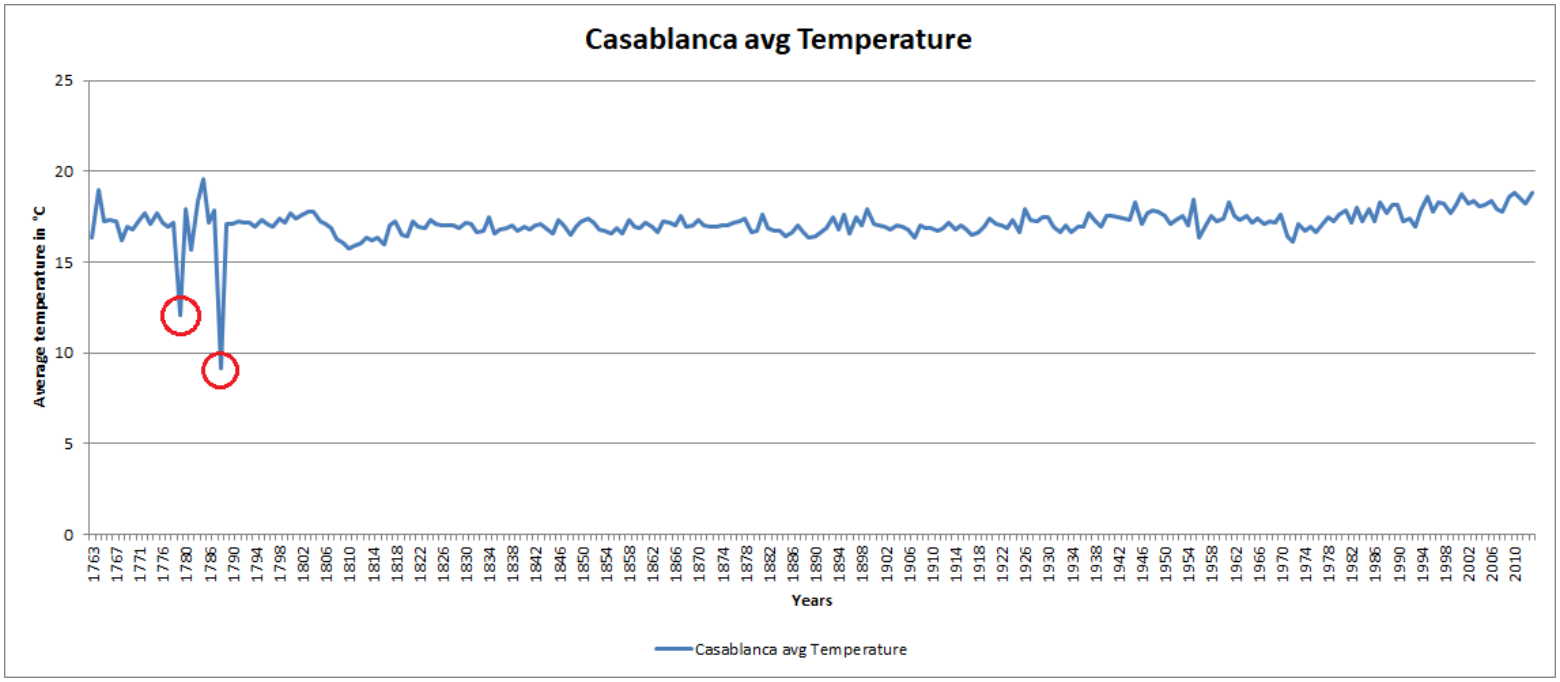


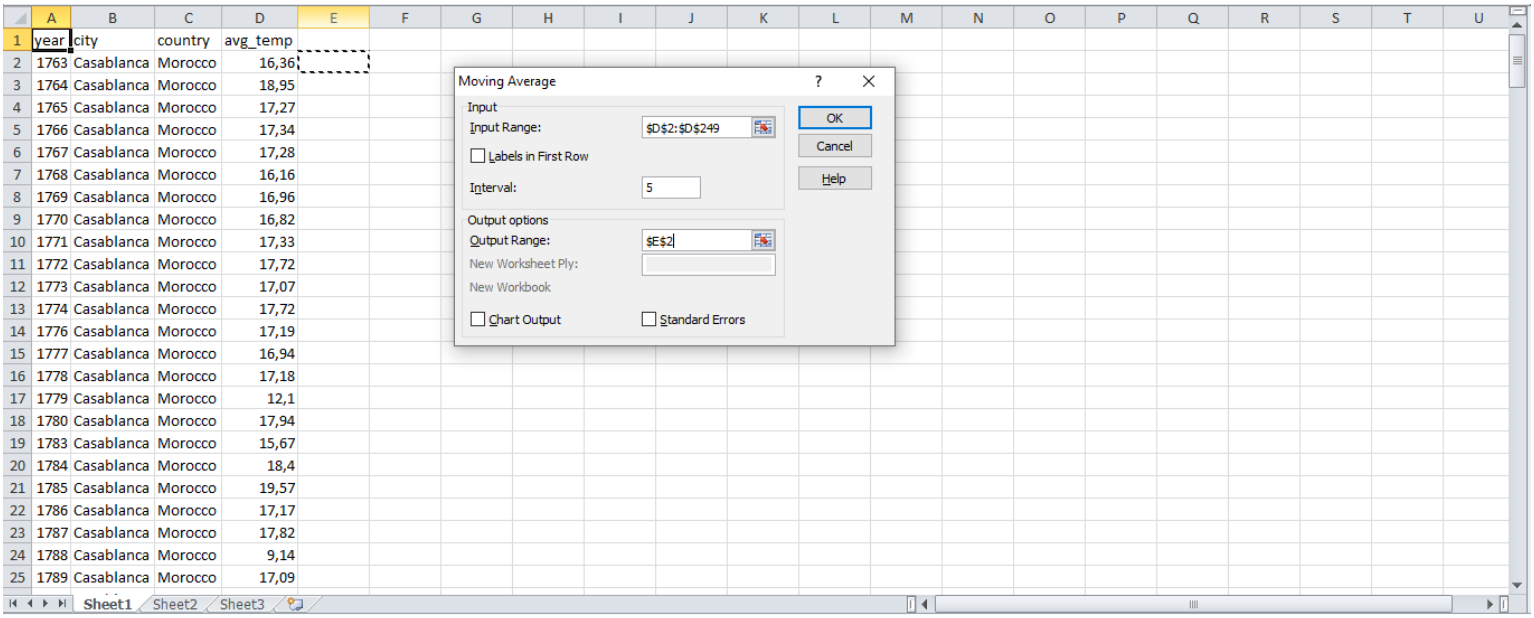
Fig6: plotting the yearly averages temperature of Casablanca

Note: the plotting is too spiky and gives a lot of details, so it's hard to interpret.

I noticed also two surprising minimums (red circles in fig6), I tried to find more details about those years in order to understand the reason of those drastic drops but I didn't find anything.

As mentioned in the project instructions, I needed to smooth out the line using moving average to make trends more observable.

So I used a tool called data analysis in EXCEL to calculate the moving average over my data (fig7)



year	city	country	avg_temp
1763	Casablanca	Morocco	16,36
1764	Casablanca	Morocco	18,95
1765	Casablanca	Morocco	17,27
1766	Casablanca	Morocco	17,34
1767	Casablanca	Morocco	17,28
1768	Casablanca	Morocco	16,16
1769	Casablanca	Morocco	16,96
1770	Casablanca	Morocco	16,82
1771	Casablanca	Morocco	17,33
1772	Casablanca	Morocco	17,72
1773	Casablanca	Morocco	17,07
1774	Casablanca	Morocco	17,72
1776	Casablanca	Morocco	17,19
1777	Casablanca	Morocco	16,94
1778	Casablanca	Morocco	17,18
1779	Casablanca	Morocco	12,1
1780	Casablanca	Morocco	17,94
1783	Casablanca	Morocco	15,67
1784	Casablanca	Morocco	18,4
1785	Casablanca	Morocco	19,57
1786	Casablanca	Morocco	17,17
1787	Casablanca	Morocco	17,82
1788	Casablanca	Morocco	9,14
1789	Casablanca	Morocco	17,09

Fig7: Calculating 5 years Moving average

I had also to choose between multiple window sizes to get the best amount of smoothing and keep enough data in the line.

5 years Moving average:

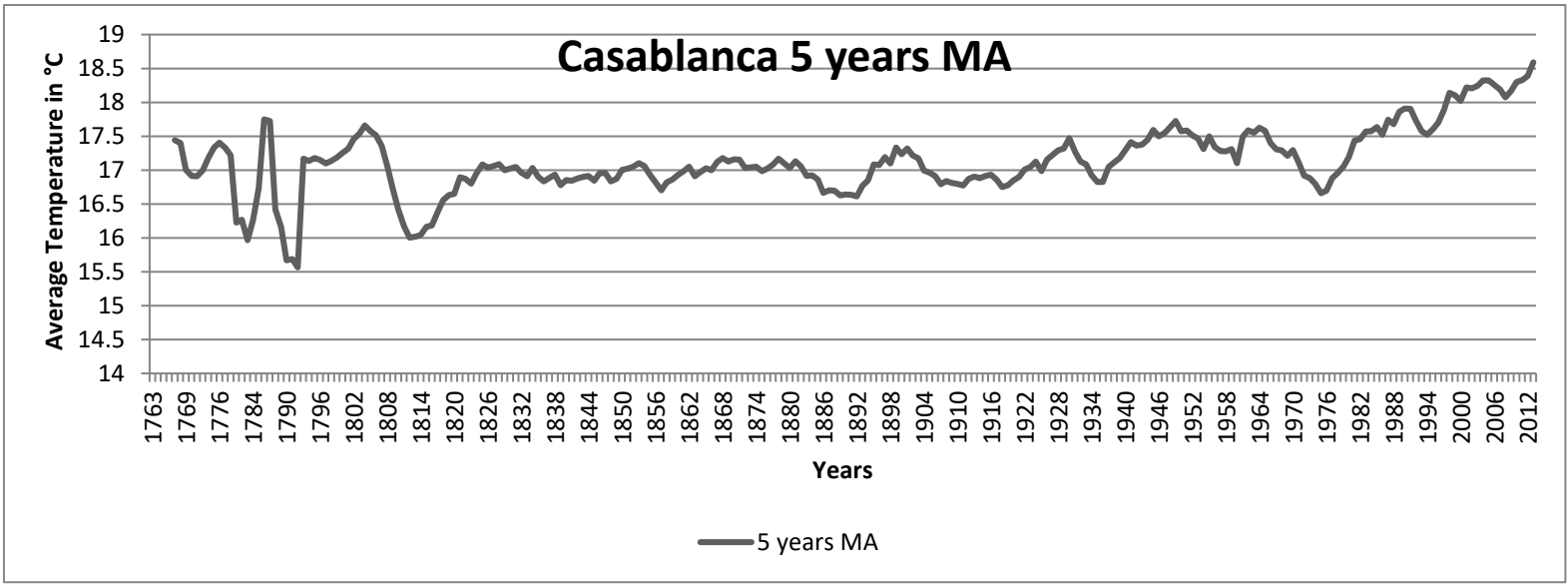


Fig8: Casablanca's 5 years moving average line chart

The plotting is less noisy than the original one (see fig8), it keeps more data but it doesn't show clearly the trends. For example, in the time window between 1892 and 1901, I can see clearly there is a local uptrend, but the moving average line keeps more fluctuation in this period of time, the same thing can be said about the period between 1976 and 2013. In summary, the five years moving average line is smooth but not smooth enough to make trends more clear.

10 years moving average:

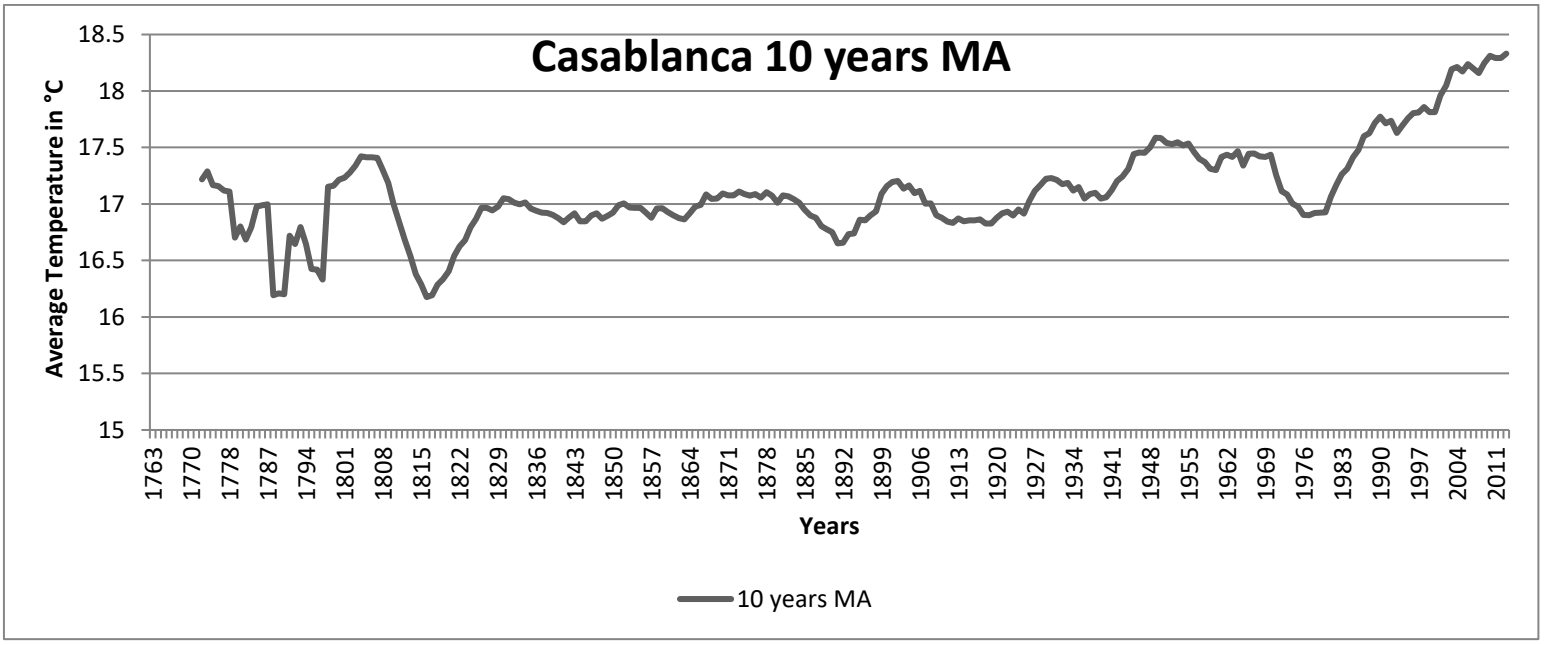


Fig9: Casablanca's 10 years moving average line chart

The 10 years moving average shows less fluctuation than the previous one, especially in the period of time before 1804. We can also observe clearly local trends:

- a downtrend between 1804 and 1815
- an uptrend between 1891 and 1900
- an uptrend since 1977
- And no trend between 1820 and 1970

15 years moving average:

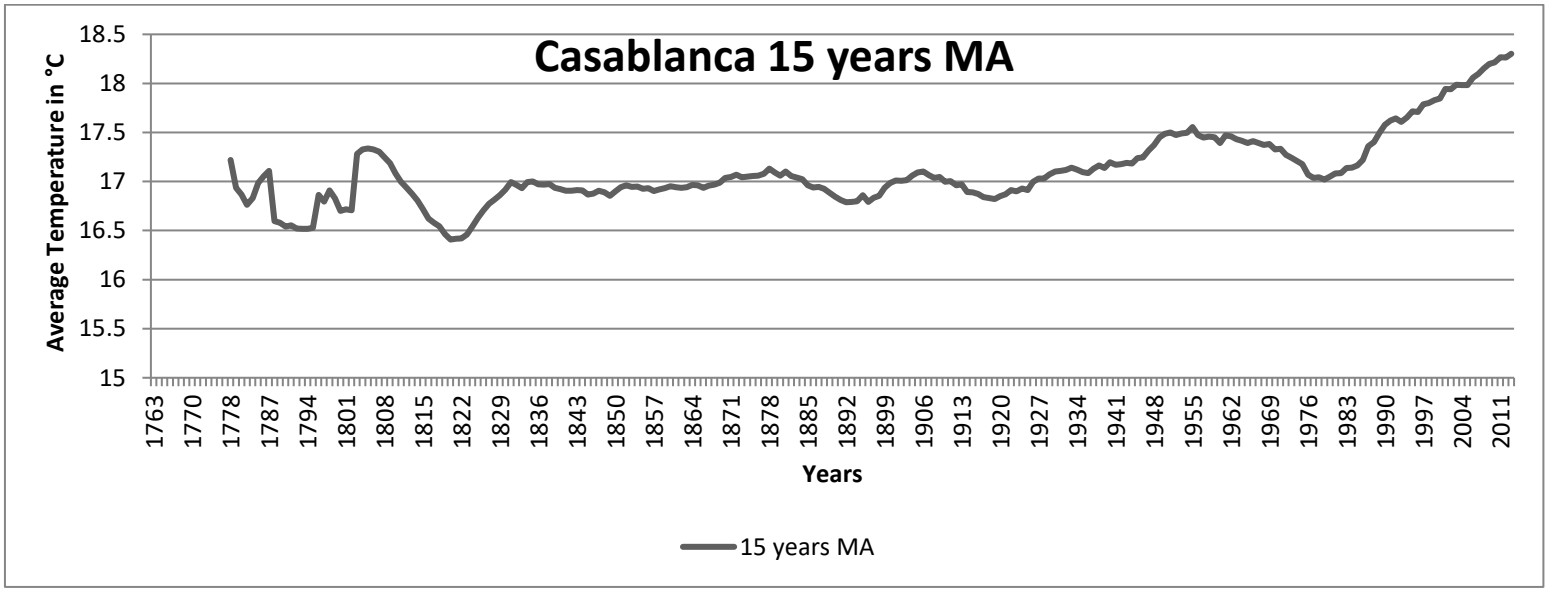


Fig10: Casablanca's 15 years moving average line chart

This line is too smooth and too flat to show local trends, it omits a lot of details and the only trend I see here is the one between 1977 and 2013.

In summary, the 5 years moving average omit a lot of fluctuation from the original data, but it's not clear enough to show all trends in the data. The 10 years moving average is less noisy than the previous one and it show more data regarding local trends. The final one is too smooth and the line is

too flat to disclose all regions of interests, but it is useful to see the global picture. So finally I decided to use the 10 years moving average to analyze the data

3.2) Global chart:

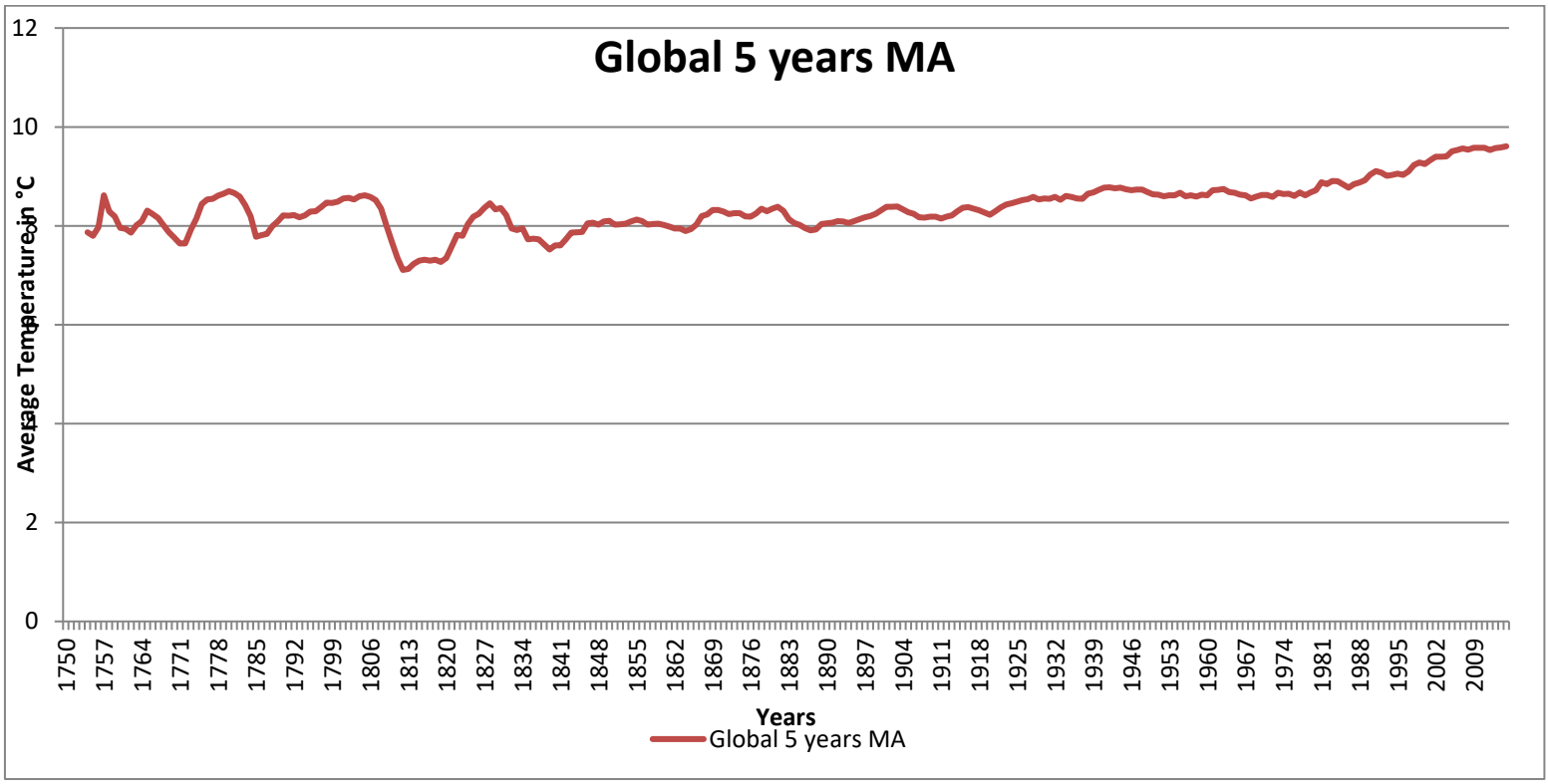


Fig11: Global 5 years moving average line chart

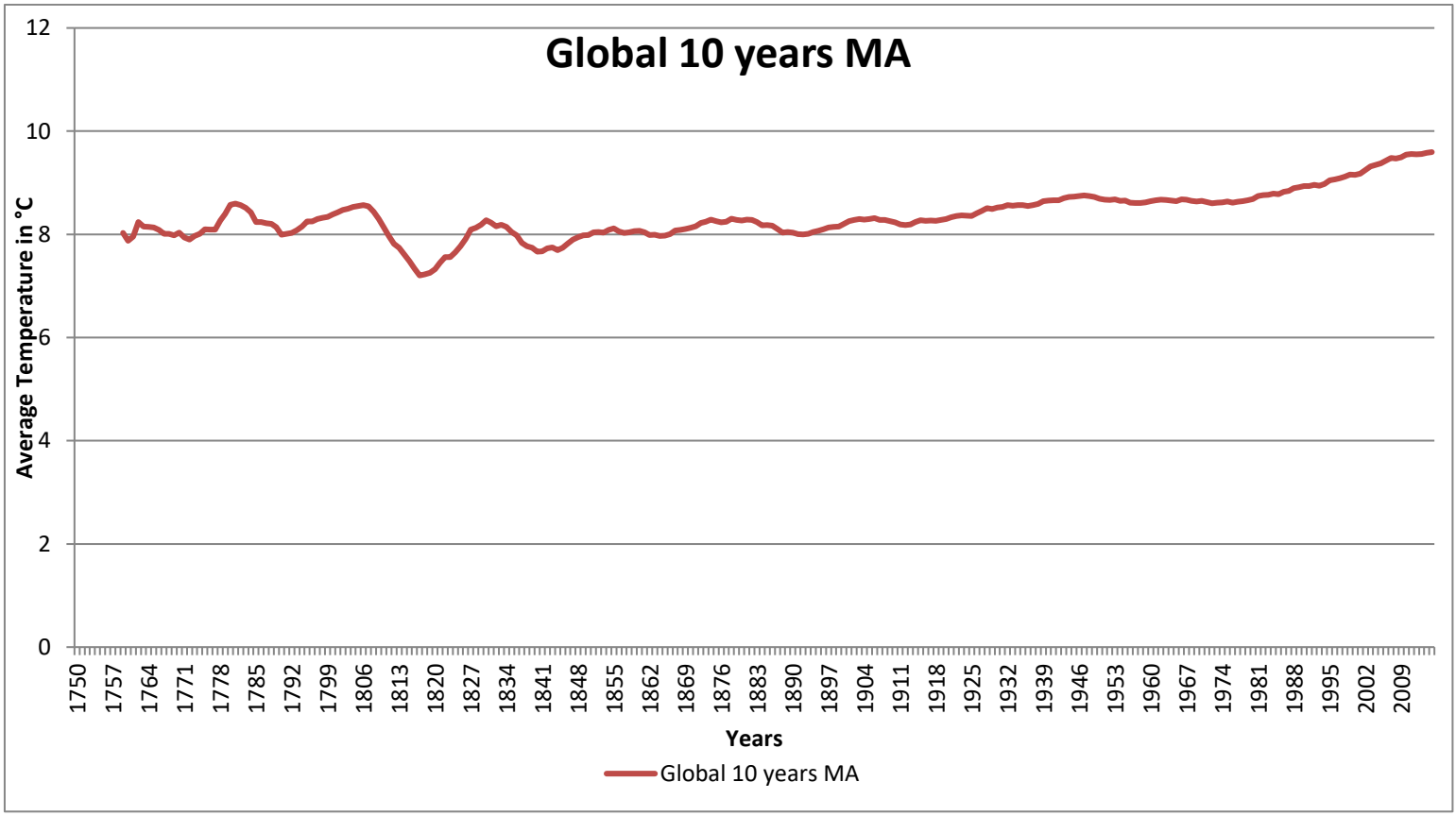


Fig12: Global 10 years moving average line chart

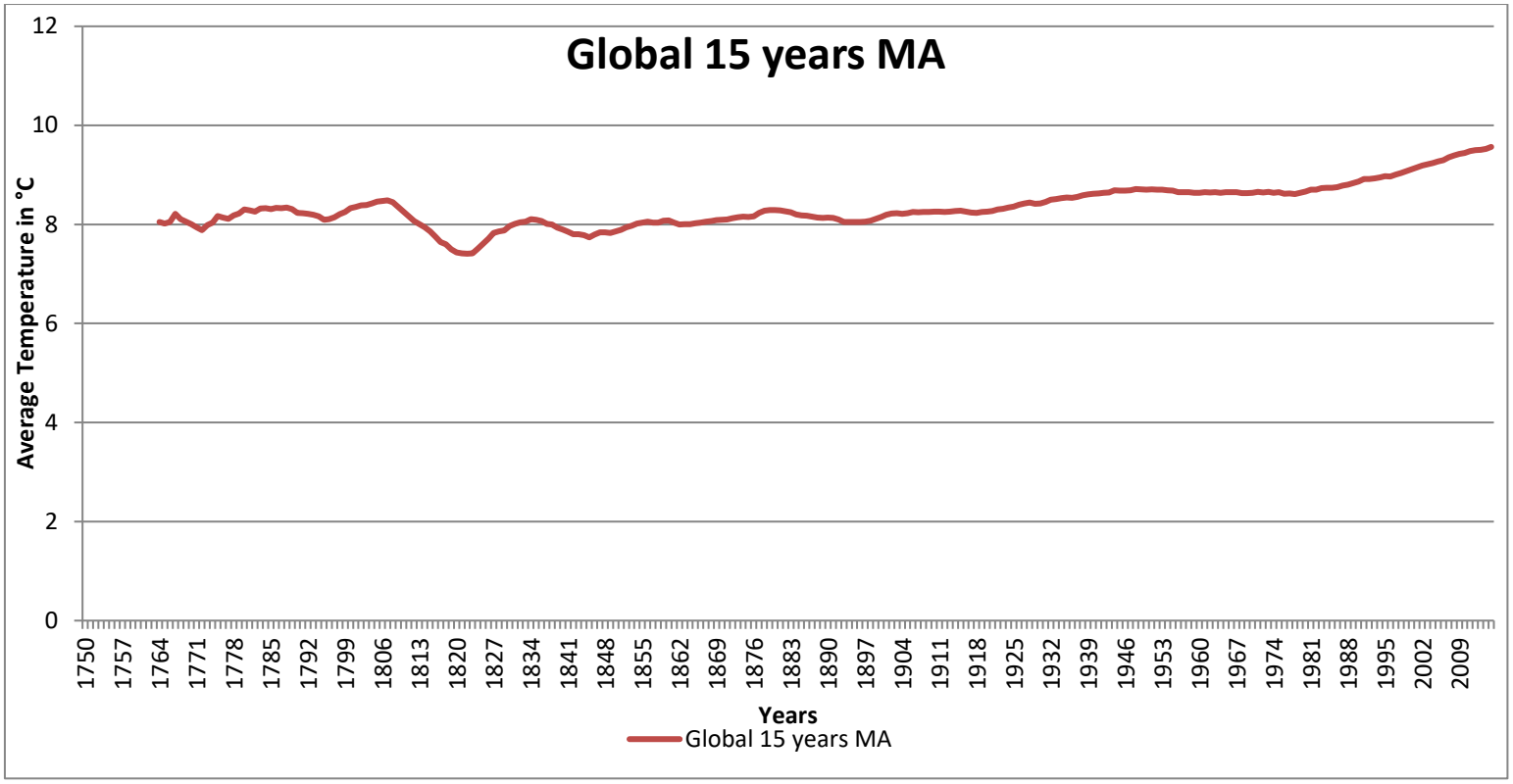


Fig13: Global 15 years moving average line chart

Note: In this case, the 5 years moving average is too noisy to reflect clearly all the trends within the plotting data; the 10 years moving average is too good to show local and global trends, for example I can see a clear:

- Uptrend between 1790 and 1804,
- Down trend between 1804 and 1817
- Uptrend between 1818 and 1829
- And a global uptrend from 1977 to the end of the line

The 15 years moving average is beautifully smooth so it can show all global trends and some local trends.

Finally, I decided to use the 10 years moving average to analyze local data and global data.

3.3) Global and local line chart:

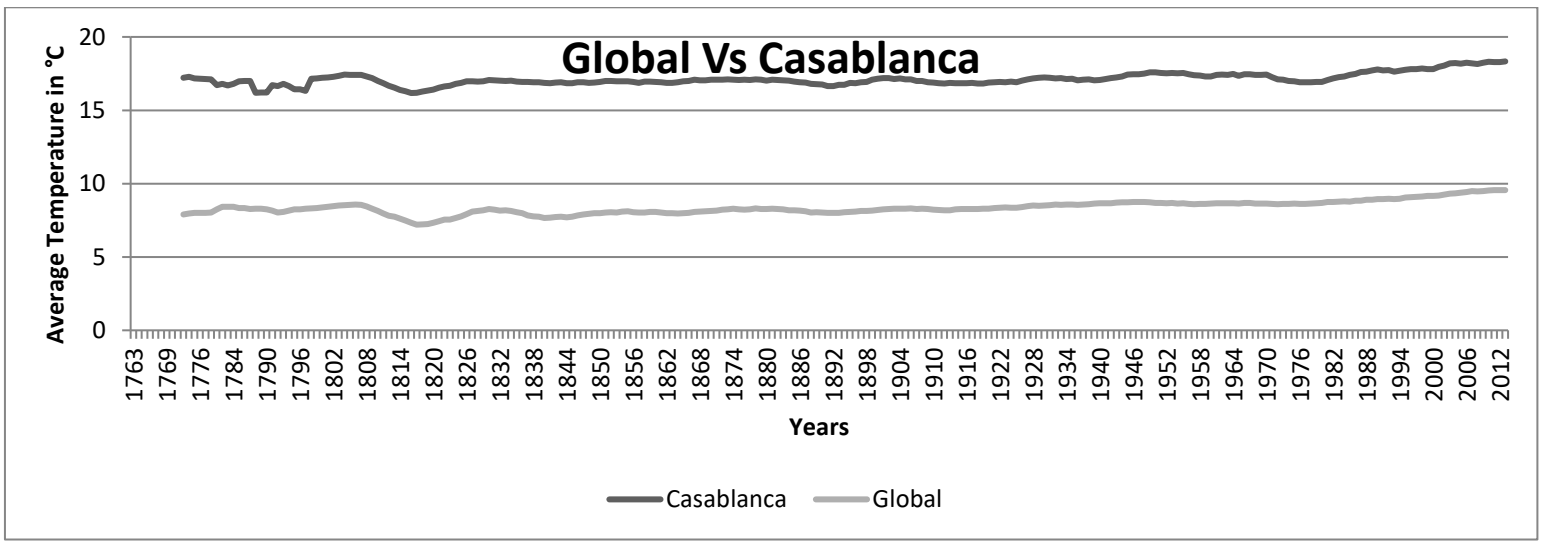


Fig14: Casablanca's temperature Vs Global Temperature line chart

4) Observations:

- Is Casablanca cooler or hotter on average compared to the global average?

As we can see and without using any calculation formula, the moving average representing Casablanca is way above the global moving average, so we can say that Casablanca is a hotter city on average compared to the global average.

- How do the changes in Casablanca's temperatures over time compare to the changes in the global average?

If we divide the time axis to different window times we can see there are a lot of similarities regarding trends, for example we can see in the two lines:

- ✓ No trend from the beginning of the line until 1807
- ✓ A downtrend between 1807 and 1817
- ✓ An uptrend between 1817 and 1830
- ✓ No trend between 1830 and 1976
- ✓ A clear uptrend from 1977 to the end of the line.

So in summary, I see a lot of similarities between the changes in Casablanca's temperature over time and global average regarding trends.

- What does the overall trend look like? Is the world getting hotter or cooler? Has the trend been consistent over the last few hundred years?

As I mentioned in the previous point, we can see ups and downs over time; but if we see the global picture we see no trend until the 70's of the 20th century, and then the temperature increases and the world is getting hotter.

- What are the minimum and the maximum average temperature in Casablanca?

Analyzing the 10 years moving average for Casablanca I found a minimum point in 1816 of 16.176°C and a maximum point in 2013 of 18.331°C