Project: Explore Weather Trends

Author: Ana Chan
E-mail: ana2004@gmail.com

In this project, I have analysed the data and trends from global temperatures and Lisbon temperatures (my home town) as well as from other cities with specific weather characteristics, locations and contrasts: Sydney (Southern Hemisphere), Cairo (tipically with high temperatures) and Novosibirsk (in Russia with low temperatures).

1. Extraction of data from the database

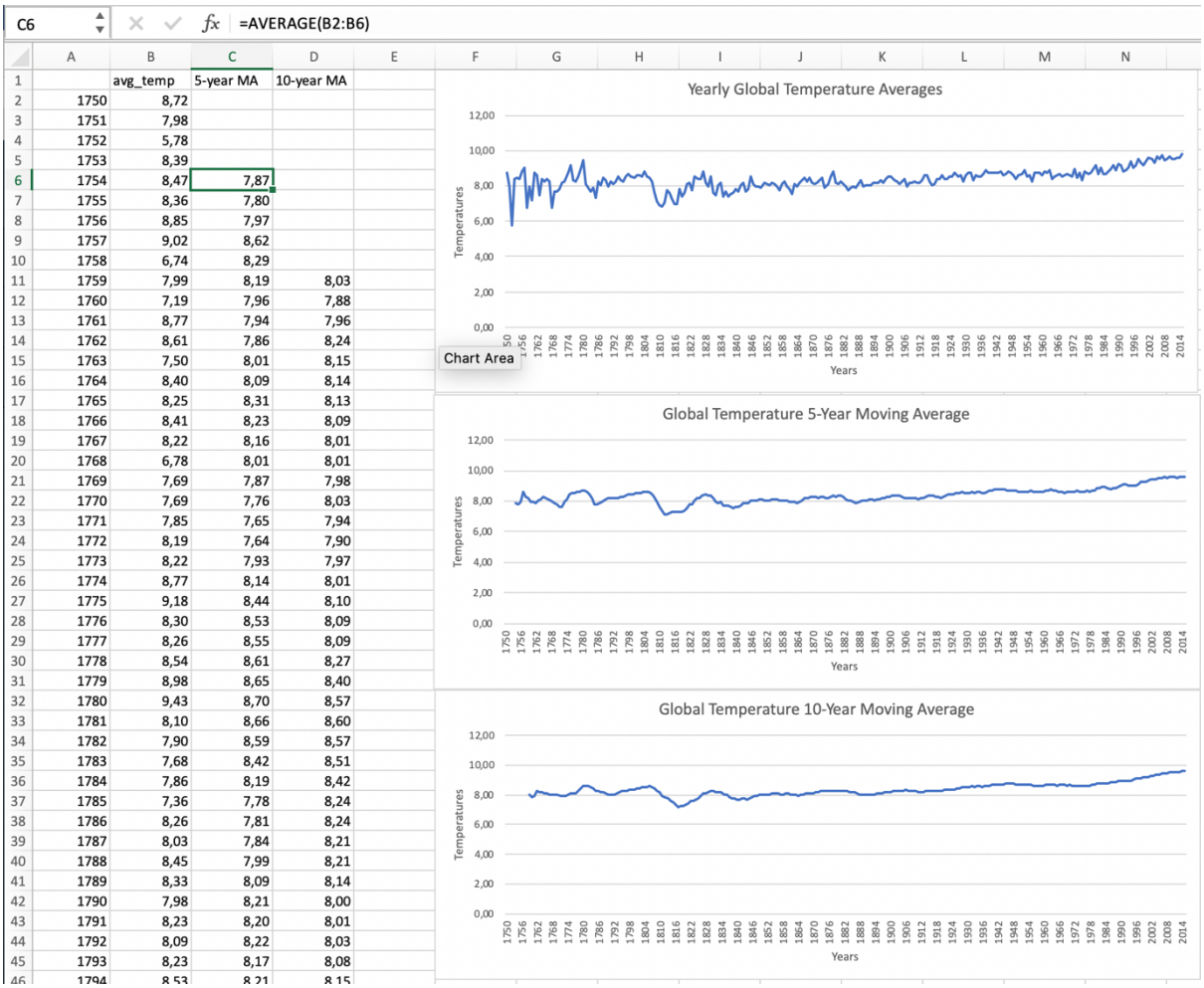
The data was extracted from the database using SQL and the queries results were downloaded to CSV files. These were the SQL queries used to extract the global data and the data for the different cities:

- Global temperatures:
  *SELECT **
  FROM global_data;
- Lisbon:
  *SELECT **
  FROM city_data
  WHERE city='Lisbon';
- Sydney:
  *SELECT **
  FROM city_data
  WHERE city='Sydney';
- Cairo:
  *SELECT **
  FROM city_data
  WHERE city='Cairo';
- Novosibirsk:
  *SELECT **
  FROM city_data
  WHERE city='Novosibirsk';

2. Moving Averages calculation

To work with the data contained in the CSV files, I have used Excel. The data was imported from the CSV files into Excel as data from text. In order to prepare the data for the line charts and improve the visibility of the trends, moving averages were calculated.

Concerning the selected year range for the calculation of the moving average, I have considered 2 options: 5 years and 10 years.



	A	B	C	D
1		avg_temp	5-year MA	10-year MA
2	1750	8,72		
3	1751	7,98		
4	1752	5,78		
5	1753	8,39		
6	1754	8,47	7,87	
7	1755	8,36	7,80	
8	1756	8,85	7,97	
9	1757	9,02	8,62	
10	1758	6,74	8,29	
11	1759	7,99	8,19	8,03
12	1760	7,19	7,96	7,88
13	1761	8,77	7,94	7,96
14	1762	8,61	7,86	8,24
15	1763	7,50	8,01	8,15
16	1764	8,40	8,09	8,14
17	1765	8,25	8,31	8,13
18	1766	8,41	8,23	8,09
19	1767	8,22	8,16	8,01
20	1768	6,78	8,01	8,01
21	1769	7,69	7,87	7,98
22	1770	7,69	7,76	8,03
23	1771	7,85	7,65	7,94
24	1772	8,19	7,64	7,90
25	1773	8,22	7,93	7,97
26	1774	8,77	8,14	8,01
27	1775	9,18	8,44	8,10
28	1776	8,30	8,53	8,09
29	1777	8,26	8,55	8,09
30	1778	8,54	8,61	8,27
31	1779	8,98	8,65	8,40
32	1780	9,43	8,70	8,57
33	1781	8,10	8,66	8,60
34	1782	7,90	8,59	8,57
35	1783	7,68	8,42	8,51
36	1784	7,86	8,19	8,42
37	1785	7,36	7,78	8,24
38	1786	8,26	7,81	8,24
39	1787	8,03	7,84	8,21
40	1788	8,45	7,99	8,21
41	1789	8,33	8,09	8,14
42	1790	7,98	8,21	8,00
43	1791	8,23	8,20	8,01
44	1792	8,09	8,22	8,03
45	1793	8,23	8,17	8,08
46	1794	8,53	8,21	8,15

The moving average calculations were made as follows:

- 5-year moving average:
  - Create a new column “5-year MA”
  - in the 5th cell of this column, enter the AVERAGE() function with the first 5 years data as argument: “=AVERAGE(B2:B6)”
  - Copy this formula to all cells in this column in order to calculate the moving average for all years

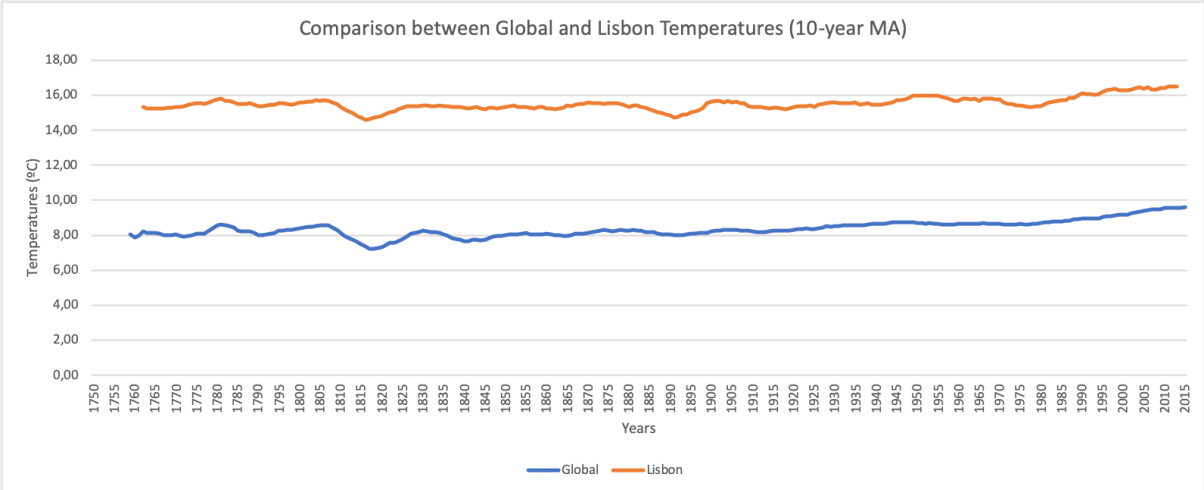
- 10-year moving average:
  - Create a new column “10-year MA”
  - in the 10th cell of this column, enter the AVERAGE() function with the first 10 years data as argument: “=AVERAGE(B2:B11)”
  - Copy this formula to all cells in this column in order to calculate the moving average for all years

Comparing the line charts for both ranges, I have decided to work with the 10-year range as it allows to visualize better the trends and plots smoother lines.

Besides of the 10-year moving average for the global temperatures, a “10-year moving average” column was calculated for each city in this project: Lisbon, Sydney, Cairo and Novosibirsk.

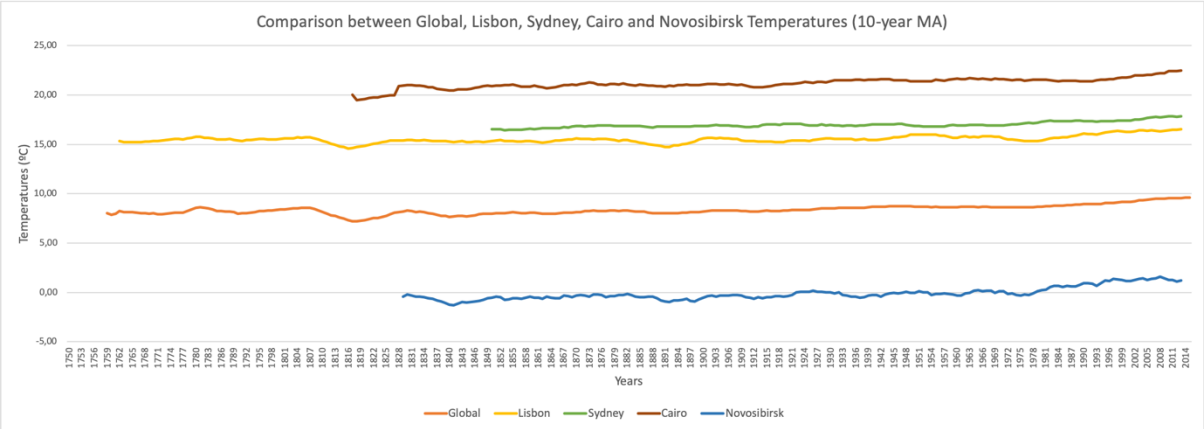
3. Generation of the Line Charts

Line Chart: Comparison between Global and Lisbon Temperatures (10-year moving average)



This line chart was generated with the data of 3 columns in the Excel spreadsheet: the years, the 10-year moving average (MA) of the global temperatures and the 10-year moving average (MA) of Lisbon temperatures.

Line Chart: Comparison between Global, Lisbon, Sydney, Cairo and Novosibirsk Temperatures (10-year moving average)

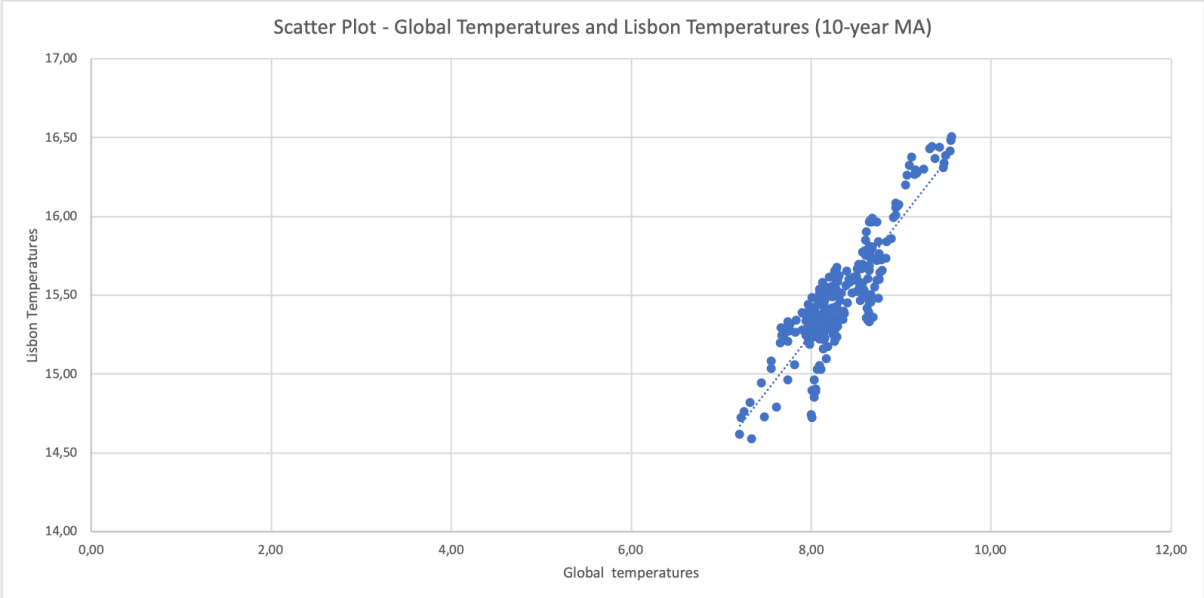


This line chart was generated with the data of 6 columns in the Excel spreadsheet: the years, the 10-year moving average (MA) of the global temperature and the 4 columns containing the 10-year MA of the temperatures for each of the following cities: Lisbon, Sydney, Cairo and Novosibirsk.

4. Calculation of the Correlation Coefficient for Global and Lisbon 10-year MA temperatures

In order to measure the closeness of the linear relationship between the Global temperatures and Lisbon temperatures, the correlation coefficient was calculated between the 10-year moving averages of these variables. This was done using Excel.

The first step was to generate the scatter plot for the data of these 2 columns and add the corresponding trend line.



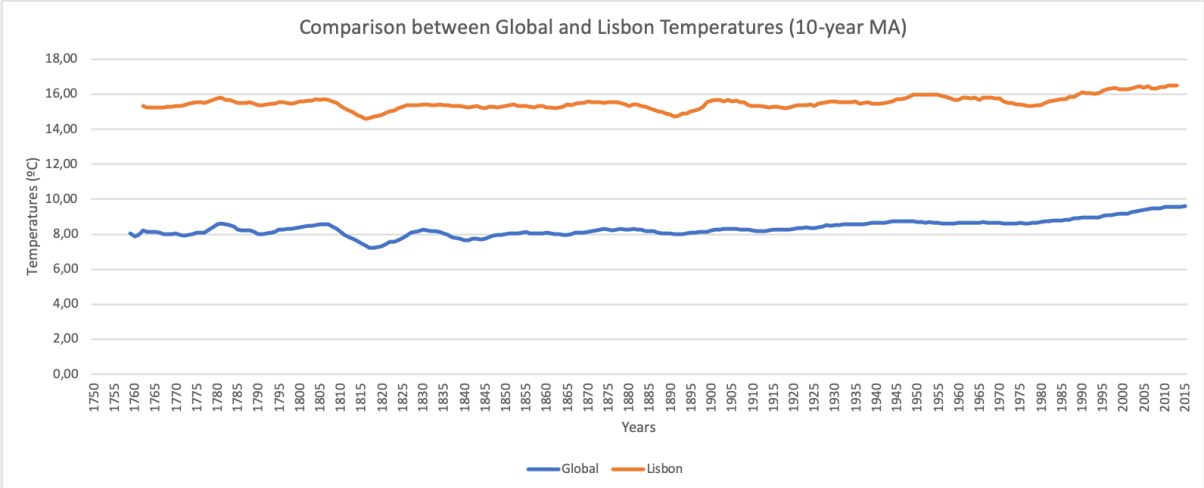
The correlation coefficient was calculated using the function CORREL(array1; array2), where the argument “array1” was selected from the column cells containing the 10-year MA of the global temperatures and the argument “array2” was selected from the column cells containing the 10-year MA of Lisbon temperatures.

The value of the correlation coefficient between the average global temperatures and Lisbon temperatures is 0,8897.

Having a correlation coefficient close to 1 (0,8897) indicates a strong positive linear relationship between the global temperatures and Lisbon temperatures: as global temperature values increase, the values of Lisbon temperatures also increase in a linear way.

5. Observations of Temperature Trends

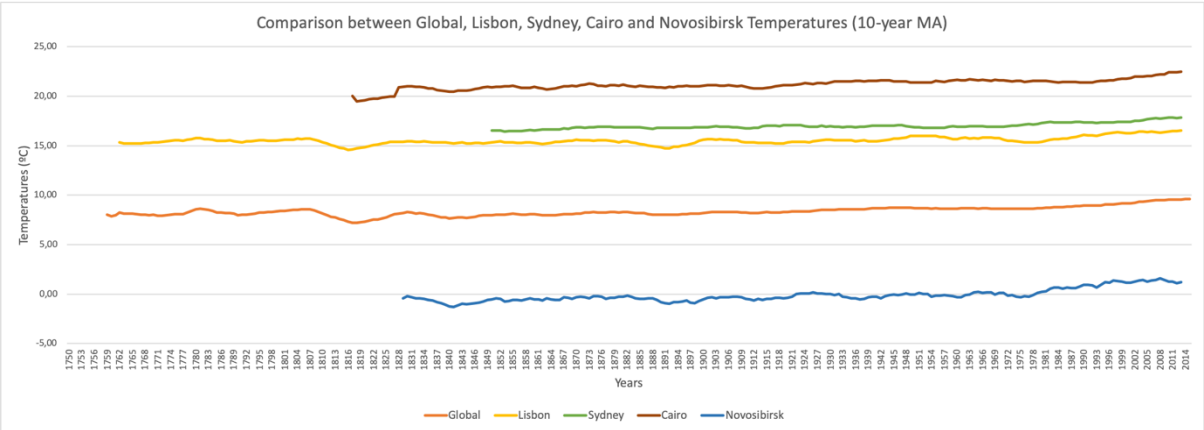
Comparison between Global and Lisbon Temperatures (10-year moving average)



The comparison between the average global temperatures and Lisbon temperatures in this chart leads to the following observations:

1. Lisbon average temperatures are approximately 7ºC higher than the global average temperatures and this difference is consistent over the years.
2. The year with maximum values for the average temperature in both cases is the last year that we have data for: 2013 for Lisbon and 2015 for global temperature.
3. Global and Lisbon average temperatures have a similar evolution. Both recorded increasing or decreasing values simultaneously: in both cases the temperature curves have increased and have reached high peaks in the years 1781, 1808 and 1830, and have also both decreased with a low peak in 1816 and 1817. Both curves are consistently increasing over the years since 1978.
4. Overall, the average values of the global and Lisbon temperatures have been oscillating inside a constant interval until 1910 and since that year there is a clear trend in both curves that indicates a consistent rise of the temperature, in spite of some oscillations between 1910 and 1978.
5. Concerning the evolution trends for the future years, it is possible to forecast a continuous increase for both global and Lisbon average temperatures based on the consistent rise of the curves since 1910, on the fact that both temperatures reach a new maximum value every year since 1978, and also on the positive linear relationship between these data confirmed by a correlation coefficient of 0,8897.

Comparison between Global, Lisbon, Sydney, Cairo and Novosibirsk Temperatures (10-year moving average)



This chart shows the comparison between the average global temperature and the average temperatures in 4 cities that represent weather and location diversity in the globe: Lisbon (Northern Hemisphere), Sydney (Southern Hemisphere), Cairo (high temperatures) and Novosibirsk (low temperatures).
Here are some observations:

1. Cairo has the highest average temperatures, approximately 13ºC above the global temperatures. Sydney average temperatures are also above the global temperatures with a difference of roughly 8ºC. Lisbon average temperatures, as seen before, are approximately 7ºC higher than the global temperatures. Novosibirsk has the lowest average temperatures, approximately 8,5ºC bellow the global temperatures.
2. The year with maximum values for the average temperature in all these cities, except Novosibirsk, is the last year that we have data for: 2013 for Lisbon, Sydney and Cairo and 2015 for global temperature. In the case of Novosibirsk, the maximum value was recorded on 2008.
3. Overall, the average temperature values in all cases have been increasing since 1917, with a consistent rise of the temperature since 1978 for all curves, except the last few years in the case of Novosibirsk.
4. Based on the persistent rise of all curves in the last years, the evolution trend for these cities indicates a continuous increasing of the average temperatures in the next years.

6. Conclusions

Comparison between Global and Lisbon Temperatures

The comparison between the global and Lisbon average temperatures indicates that in spite of having a constant difference of approximately 7ºC, their evolution over the years is very similar with a strong positive linear relationship. Regarding the future trends, it is expected a continuous rise of the temperatures as, in both cases, there was a consistent increasing of the temperature and every year has registered the new maximum temperature since 1978.

Comparison between Global, Lisbon, Sydney, Cairo and Novosibirsk Temperatures

Concerning the comparison between the global temperatures and the 4 cities representing the weather diversity, it is possible to observe that their evolution is overall similar with an increasing trend since 1978, in spite of all the differences between the weather characteristics like Cairo with high temperatures and Novosibirsk with low temperatures, and the location like Lisbon in the Northern Hemisphere and Sydney in the Southern Hemisphere. In this comparison, it is also expected a continuous increasing of the average temperatures for the next years.