Maximilian Braeuninger

Nanodegree ‘Data Analyst’ - Term 1

Project: Explore Weather Trends

1) Extract data
2) Manipulate data
3) Visualize data
4) Interpret data

1)

This part will explain which SQL queries I used to get the necessary data from the database. For the global data I used the following query:

*SELECT * FROM global_data*

Since I live in Berlin, Germany I then tried to find out which German cities are listed in the data:

*SELECT * FROM city_list WHERE country = ‘Germany’*

I found out Berlin is in there, so I went with it:

*SELECT * FROM city_data WHERE city = ‘Berlin’*

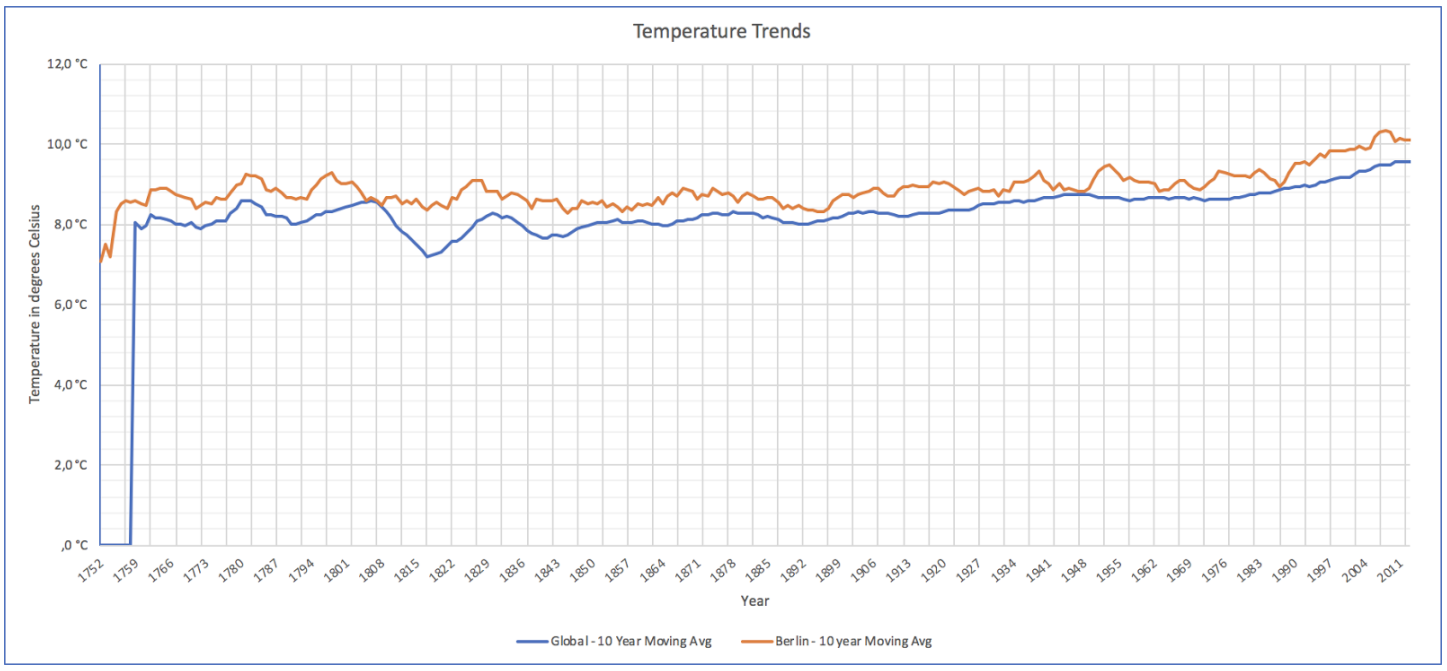
2)

I put the resulting csvs global_data and berlin_data in two seperate sheets in one excel file. To compare the global trend against the local trend I decided to use a 10-year moving average. It is calculated as the average of the current year and the previous 9 years. Global temperature data is available from 1750 on, Berlin data from 1743 on (with no data available from 1746 - 1749). That means my first valid data point is from 1752 (Berlin). Note that the first global datapoint is available from 1759 on only.

After calculating the moving averages for both datasets I put all years from 1752 onwards in a new spreadsheet and pulled the moving average for global and for berlin data through an INDEX(MATCH())-formula into the new table.

Additional calculations revealed that the correlation coefficient between both datasets is 0.566, which indicates a positive relationship between both curves.

3)



Since my goal is to compare the global trend with the local trend in Berlin I decided to go for a simple line chart. My main reasons to do so were that you can see right away what the annual difference is, and that the slope of each curve makes it easy to see bigger trends.

4)

The chart clearly shows that the average temperature in Berlin is above the global average, indicating that it is generally warmer. In most years it seems to be at least 0.5 °C warmer.

The average difference between Berlin and the globe seems to get smaller over time, especially in the 18th and the beginning of the 19th century the difference is often bigger than 1 °C.

It happened more than one time that the average temperature in Berlin dropped contrary to the global trend. This can be observed in the beginning of the 1800s, in the 1940s and in the 1980s. At these timeframes it happened that both moving averages were almost at the same level.

The pace at which the average temperatures of both Berlin and the globe rise increases over time, especially since the mid of the 20th century. This observation is supported by taking into account the correlation coefficient of 0.566, which indicates a positive correlation between both datasets.

However this relationship cannot exactly be called strong, therefore temporary differences in the curves are possible (see observation 3).