Udacity - Data Analyst Nanodegree

Project: Explore weather trends

The goal of this project is to analyze local and global temperature data and compare the temperature trends of the city closest to where I live, in this case Hamburg (Germany), to overall global temperature trends.

This report is organized as follows:

- First, the SQL query to extract the desired temperature data from the database provided by Udacity is presented.
- Then, the moving averages of the global and city temperature data are calculated, so as to visualize them in a line chart.
- Finally, some observations are made about the temperature trend information derived from the graph.

1. Extracting the data

In order to extract the temperature data, SQL commands were used in the platform provided by Udacity.

To select the city closest to where I live, I explored which cities of my country (Germany) exist in the database by performing the following query:

```
SELECT city
FROM city_list
WHERE country = 'Germany'
```

This produced the following results:

```
Berlin
Hamburg
Munich
```

I finally chose Hamburg since it's the one closest to where I live. In order to extract the global and Hamburg temperature information, I performed the following query:

```
SELECT global_data.year as year, global_data.avg_temp as global_avg_temp,
city_data.avg_temp as
            hamburg_avg_temp
FROM global_data, city_data
WHERE city_data.city = 'Hamburg' AND global_data.year = city_data.year AND
          (global_data.avg_temp IS NOT NULL AND city_data.avg_temp IS NOT
      NULL)
```

The query retrieves the **avg_temp** column from both the **global_data** and **city_data** tables, as well as the year column from the **global_data** table. The query takes into consideration that both global and Hamburg temperature values are not empty, and that the years of both sets match.

This search produces an output of 264 entries between the years 1750 to 2013. The following table shows the first six entries.

year	global_avg_temp	hamburg_avg_temp
1750	8.72	9.31
1751	7.98	8.94
1752	5.78	4.65
1753	8.39	8.12
1754	8.47	7.88
1755	8.36	7.69

Table 1. First six entries of the SQL query result

2. Calculate and visualize moving averages

In order to calculate and visualize the moving averages, the search results were downloaded as a .csv file and then imported into a Google Sheets spreadsheet. Two columns were added to store the moving average values for the Hamburg and global temperatures: **global_mv_avg** and **hamburg_mv_avg.** It was decided to consider a 10-year window for the moving average calculation so as to observe relatively smooth trends without too many abrupt fluctuations.

To calculate the 10-year moving average, the AVERAGE function of Google Sheets was used. In the case of the global temperatures, the formula **=AVERAGE(B2:B11)** was entered In the tenth row of the **global_mv_avg**, since cells B2 to B11 contain the global temperature values corresponding to the first 10 entries of **global_avg_temp**. Then the same formula was dragged down to fill in the rest of the **global_mv_avg** values. The same steps were performed to calculate **hamburg_mv_avg.**

The first entries of the table with the 10-year moving averages are shown below.

year	global_avg_temp	global_mv_avg	hamburg_avg_temp	hamburg_mv_avg
1750	8.72		9.31	
1751	7.98		8.94	
1752	5.78		4.65	
1753	8.39		8.12	
1754	8.47		7.88	
1755	8.36		7.69	

1756	8.85		8.64	
1757	9.02		8.48	
1758	6.74		7.76	
1759	7.99	8.03	8.69	8.016
1760	7.19	7.877	8.36	7.921
1761	8.77	7.956	8.86	7.913
1762	8.61	8.239	7.81	8.229
1763	7.5	8.15	7.8	8.197
1764	8.4	8.143	8.25	8.234

Table 2. First results of the 10-year moving average calculation

In order to visualize the moving averages over time, the line chart shown below was created, using the **`year`** column as the horizontal axis and the **`global_mv_avg`** and **`hamburg_mv_avg`** columns as the vertical axis. The values on the vertical axis are limited between 6 and 10 °C for a better visualization, since the temperature values do not exceed this range.

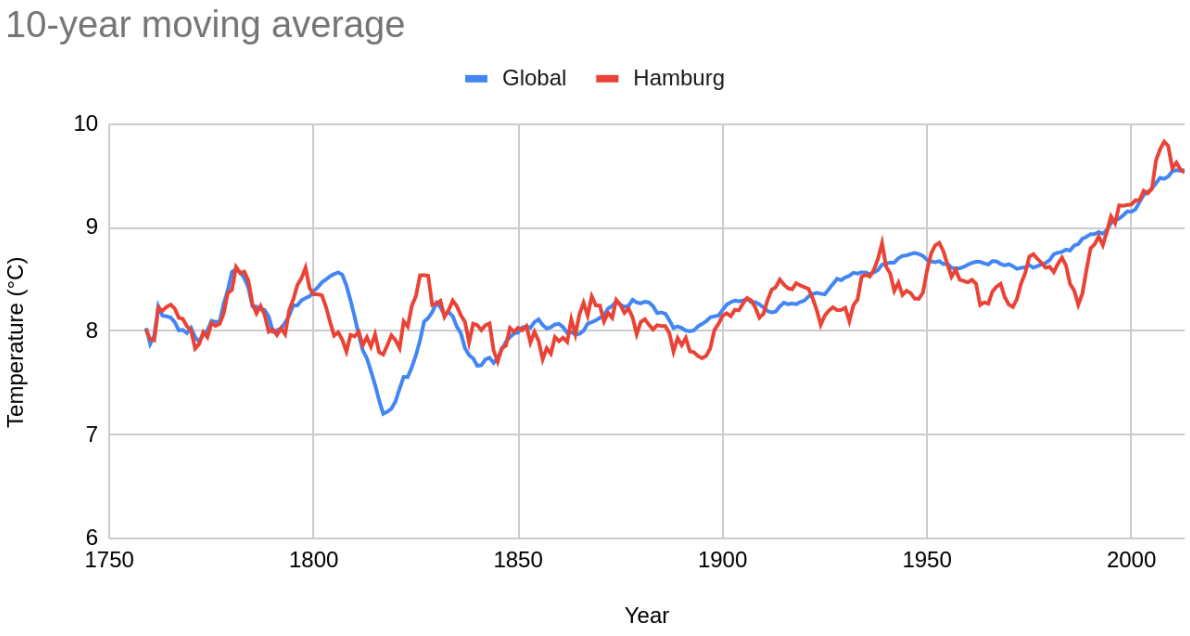

Figure 1. Line chart with moving averages corresponding to Hamburg (red) and global values (blue)

3. Observations

- It can be observed in the graph that both Hamburg's and the global temperature trend follow a similar evolution in time. In fact, the difference between the two curves remains less than 1°C throughout the observed time range.

- The largest difference between the two curves takes place between 1810 and 1820, where a large decrease in global temperature can be observed, while the temperature in Hamburg remains almost constant.
- It can be seen that from 1850 onwards the temperature in Hamburg is on average slightly lower than the global temperature.
- In the century between 1750 and 1850 both the global temperature and Hamburg temperatures have fluctuated, but from 1850 onwards the trend is increasing. Therefore it is clear that the world in general is getting warmer.
- Furthermore, it can be observed that in the century between 1850 and 1950 the temperature increase in both cases is slightly less than 1°C. However, from 1950 to 2000 that same increase occurs in half the time (approximately 50 years). Therefore it can be concluded that the temperature increase in the world is getting faster.

Conclusion

In this project, temperature trends of the world and of one city in particular, Hamburg, were analyzed.

For this purpose, a SQL query was performed to extract the annual temperature data from a database. The data was then analyzed in a Google Sheets spreadsheet. The 10-year moving average was calculated and a line chart was generated to compare the two trends.

Finally, the most relevant observations derived from this information were mentioned. The most important may be that for more than one and a half century the world's temperature has been rising, and that this increase is getting faster.