Data Analyst Project: Explore Weather Trends

Overview
In this project, I will analyze local and global temperature data and compare the temperature trends where I live to overall global temperature trends.
My goal is to create a visualization and prepare a write up describing the similarities and differences between global temperature trends and temperature trends in the closest big city to where I live.

Step 1: Data Extraction
In this step we'll extract the data from the database hosted on the Udacity site. Our goal is to export the temperature data for the world as well as for the closest big city.

The city_list table contains the names of some large cities. I've written a short query to find the cities in Spain, which are Barcelona, Madrid, and Valencia:

SELECT * FROM city_list WHERE country LIKE 'Spain';

Altering column names:
By examining the database SCHEMA through the Udacity interface, I discovered that both city_data and global_data contain a column named 'avg_temp'. I have changed the column names so that I'll be able to extract a joined table:

ALTER TABLE city_data RENAME COLUMN avg_temp to city_avg_temp;

ALTER TABLE global_data RENAME COLUMN avg_temp to global_avg_temp;

Extract the data:
Now I am ready to extract the data, in this case I want to get the available data for Madrid:

SELECT global_data.year, global_data.global_avg_temp, city_data.city_avg_temp

FROM global_data JOIN city_data
ON global_data.year = city_data.year
WHERE city LIKE 'Madrid';

The result is a downloadable CSV file with the data for year, global_avg_temp, and city_avg_temp. I joined the city_data table stipulating that I wanted only that data from the global_data table where the years matched up, so my result has already been truncated.

Step 2: Open the CSV file

I chose Google Docs, and simply imported the CSV file from my downloads folder. On first glance, the data for Madrid spans the years from 1750 to 2013 and there isn't any missing data.

Step 3: Make a line chart

In the line chart below, I've plotted the 50-year moving average for global temperatures in blue (50 YR Global MA), as well as the 50-year moving average for my chosen city, Madrid, in red (50 YR Madrid MA). I tried shorter and longer year intervals for the moving average, and feel that the 50 year interval is a good range to see the general trends, providing fairly smooth lines without losing detail. The calculation was made for the average of each 50 year range within Google Docs and plotted using the handy "Insert chart" from the toolbar, and then editing the data range as well as the titles.

Temperature Change Over Time
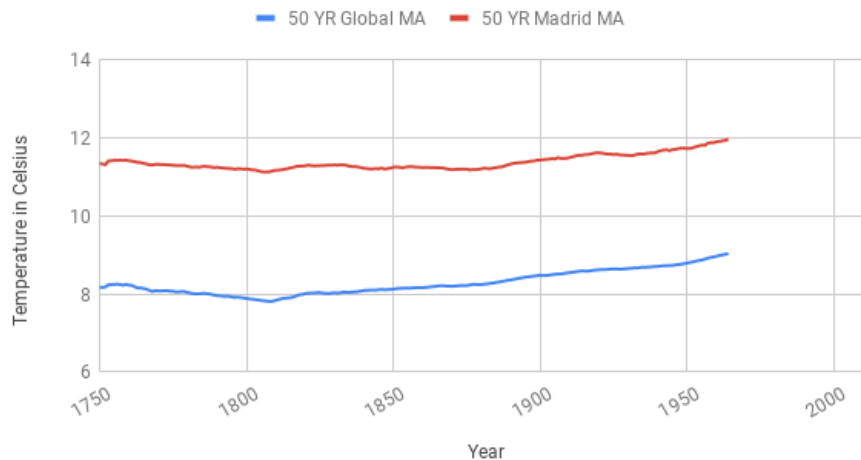
Step 4: Observations

1. The temperature in Madrid is on average higher than the global average.
The temperature in Madrid on average is approximately 3°C higher than the average global temperature, for the given dataset, and remains about the same over time although with some deviation. For the years 1750 to 2013 the **overall** average was approximately 8.36°C globally, while in Madrid it was around 11.45°C.

2. The general trend is increasing temperatures both globally and locally.
The general trend particularly since the early 1800's is an increase in temperature. This increase becomes more pronounced globally starting in 1852, I can hypothesize that this is was perhaps caused in part by an increase in industrialization. We might also be curious about the changes in data collection, for example

3. There are some periodic deviations between the temperature trend globally versus Madrid.
During this same period of global growth in the mid 1850's, however, the data for Madrid stays fairly flat. There is actually a slight decrease in the period from 1850 to 1861 which has a direct effect on the moving average. In the late 1800's we can visualize a jump to again closely match the global trend, with a similar 3°C gap between the global and Madrid temperature.
There is also a small dip downwards in the data from Madrid in the years 1926-1928, and again from 1971 to 1983. A more in-depth analysis of the reasons for these temperature deviations might include looking at trends in other nearby cities, both in Spain and in Europe, as well as examining historical events. In any case I would need to find further data to substantiate the findings.

4. The rate of change is slightly lower for Madrid.
On average, for every 1°C that the temperature increases globally, the temperature in Madrid increases approximately 0.591°C. The temperature in Madrid therefore appears to be increasing at a slightly slower rate respective to the global temperature. I could postulate that perhaps this is due to many factors. Some factors could include the development rate, and we might want to look at how data was collected (i.e., was it done in a manner consistent with the global data collection, was the data collected in the same spot, etc.). I would need further data to back up other claims.

References

SQL commands: https://www.w3schools.com/sql/default.asp

Google Docs Spreadsheet: https://docs.google.com/spreadsheets/d/1qaWstdO4NoKfXxAk-XuIcYbBeVSnwbTiPelao0-Gp7c/edit?usp=sharing

Otto Roberson
otto.roberson@gmail.com