



Explore Weather Trends

Abdulrahman Mohammed Alobaidy

Cohort 9

Data Analyst Nanodegree Program

Email: AbdulrahmanAlobaidy2001@gmail.com

Steps

This section outlines the process of extracting the information from the database, cleaning it, calculating the moving average then finally exporting and visualizing it.

1. Data Extraction

To start we had to extract the data from the database.

In order to determine what are the cities that we can choose from we have to take a look at the `city_list` table and filter them to the country of “Iraq” since this is the country that I live in:

```
SELECT *  
FROM city_list  
WHERE country = 'Iraq'
```

the data that we are interested in is for the city of “Irbil” in “Northern Iraq”.

Now that we have identified our target city, we proceed to examine the city data:

```
SELECT *  
FROM city_data  
WHERE city = 'Irbil'
```

Here we notice that the `avg_temp` for the year **1816** is **NULL**, so we have to take this into consideration.

The data begins from the year **1808** to the year **2013**

Now we extract this data from the `city_data` table and **JOIN** it with the global data from the `global_data` table:

```
WITH c as (  
  SELECT *  
  FROM city_data  
  WHERE city = 'Irbil' AND avg_temp IS NOT NULL  
)  
  
SELECT g.year, g.avg_temp Global_AVG_Temp, c.avg_temp local_AVG_TEMP  
FROM global_data g  
JOIN c  
ON c.year = g.year
```

This joining operation will not include the row for the year **1816** from both the global and the local data sets, so that it doesn't affect the calculations for the moving average as we can't consider NULL as 0.

Now we download the data in **CSV** format.

2. Data Processing

Now we process the csv file with **Python** to calculate the moving average for both the global and local average temperature for each year.

The code for this was written specifically for this project and is on Github, you can find it [here](#).

I preferred this code so that I could specify the time block that I will be calculating the moving average for, I ended up producing two data sets, the first one with 10 years block, and the second one for 25 years block.

Then we export the data sets to separate csv files.

Finally, I performed analysis on the the initial data in order to extract specific numbers, these are mentioned in the conclusions section, the code for this can be found [here](#).

3.Data Visualization

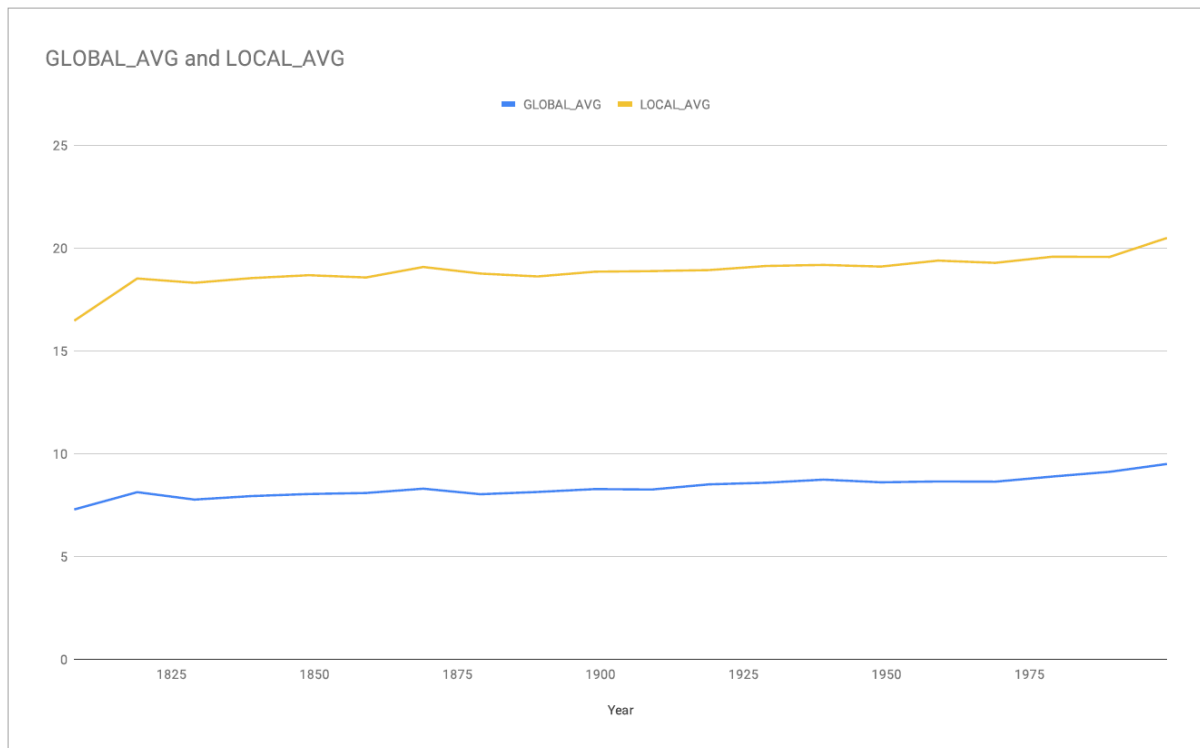
The aim now is to visualize the two data sets that we ended up with from the last step.

At this point we need to determine what kind of chart to use in order to best visualize the data, since we have a quantitative field that we want to visualize over a period of time in order to observe trends, we prefer to go with a **Linear plot**.

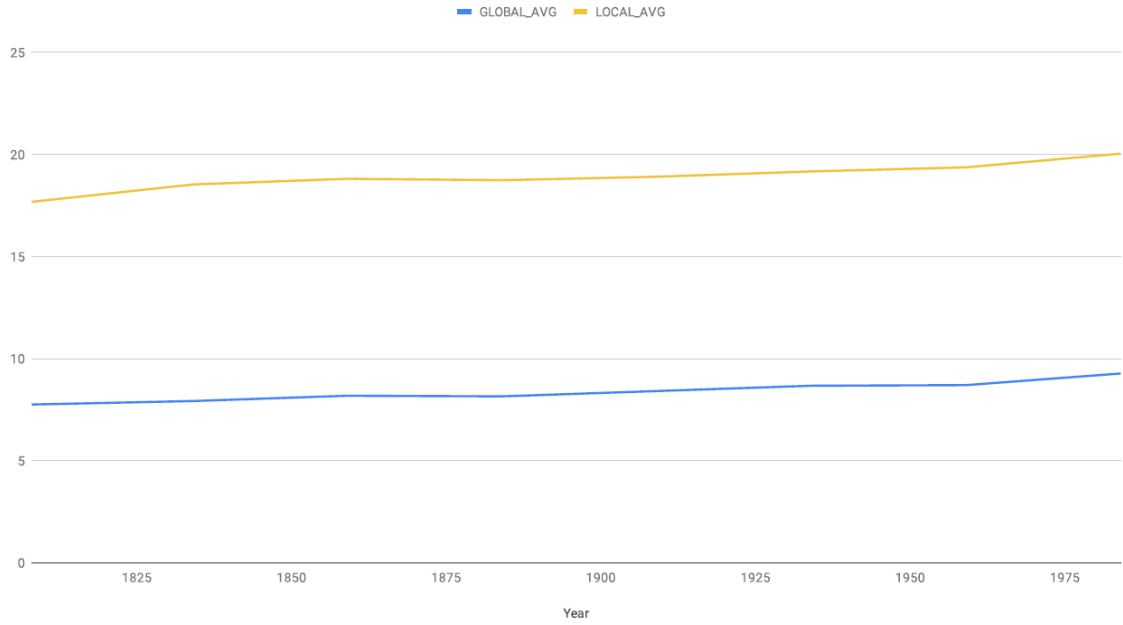
To make this possible, I used **Google Sheets**.

The following are two charts representing:

- 10-year Block
- 25-year Block respectively



GLOBAL_AVG and LOCAL_AVG



Conclusions

1. Statistics

Here are a few stats resulted after analyzing the initial data set

Property	Degrees Celsius	Year(s)
MIN_GLOBAL	6.86	1811
MAX_GLOBAL	9.73	2007
MIN_LOCAL	5.07	1818
MAX_LOCAL	21.97	2010
MIN_LOCAL_GLOBAL_DIFFERENCE	2.76	1818
MAX_LOCAL_GLOBAL_DIFFERENCE	14.40	2013
MIN_GLOBAL_GLOBAL_DIFFERENCE	-0.88	1830-1831
MAX_GLOBAL_GLOBAL_DIFFERENCE	0.85	1817-1818
MIN_LOCAL_LOCAL_DIFFERENCE	-16.31	1817-1818
MAX_LOCAL_LOCAL_DIFFERENCE	11.69	1818-1819
GLOBAL_DIFFERENCE	1.98	-
LOCAL_DIFFERENCE	9.61	-
AVERAGE_GLOBAL	8.4	-
AVERAGE_LOCAL	18.93	-
AVERAGE_DIFFERENCE	10.56	-
AVERAGE_GLOBAL_GLOBAL_DIFFERENCE	0.01	-
AVERAGE_LOCAL_LOCAL_DIFFERENCE	0.05	-

Ddddd

2.Observations

The following are some observations regarding the aforementioned stats:

- Notice that **AVERAGE_LOCAL** is higher than the **AVERAGE_GLOBAL** by about 10 °C.
- Notice that the year 1818 was the coldest year in Irbil with 13.86 °C below the **AVERAGE_LOCAL**, the **MIN_LOCAL** being at 5.07 °C, while the **AVERAGE_LOCAL** rests at 18.93 °C, this is the only year in the data set range where Irbil is lower than the global temperature, which is apparent when comparing the **MIN_GLOBAL** and the **MIN_LOCAL**.
- Notice that the lowest global temperature was in 1811 with the temperature being 6.86 °C, while the highest global temperature was in 2007 rising to 9.73, this rising is indicated by the **AVERAGE_GLOBAL_GLOBAL_DIFFERENCE** which indicated that the average increase of temperature from year to year is 0.01 °C, which signifies the the upward trend in the global temperature that is obvious by the two plots we saw earlier.
- The average difference between global temperature and the local temperature is 10.56 °C, with the local being the higher, which is supported by both plots.
- The lowest difference between the two fields was in 1818, again the coldest year in Irbil in almost 200 years, with only 2.76 °C between the local and global temperatures.

- The highest difference between the two fields was in 2013 by 14.40 °C, with the local being at 21.77 °C and the global at 9.61 °C.
- The largest drop in global temperature in almost 200 years was from 1830 to 1831, dropping by 0.88 °C.
- The largest increase in global temperature was from 1817 to 1818, increasing by 0.85 °C.
- The largest drop in local temperature was from 1817 to 1818, dropping by 16.31 °C.
- The largest increase in local temperature was from 1818 to 1819, increasing by 11.69 °C, recovering from the coldest year.
- The difference of global temperature between the beginning of the data set and its end is 1.98 °C, while that of the local temperature is 9.61 °C, which is significantly higher than that of global temperature.
- The year-to-year global difference in temperature is 0.01 °C, while that of the local temperature is 0.05 °C, which is five times the global.

Final Thoughts

Here are some thoughts on the overall analysis of the data set and to summarize the findings:

- Both the global and local temperatures are following an upward trend with the local being significantly higher, both of which are evident by looking at both the plot and data.
- The only outlier in the Irbil city data aside from the NULL 1816 was the year 1818 being significantly lower than average local and a little lower than the global minimum.

Resources

Here is a the [Github repository](#) created for the sole purpose of this project, it contains the original data sets, the extracted data set, the SQL instructions, the codes written to analyze the data and generate the results along with this report itself.