Project: Explore Weather Trends

By: Fatima M Atieh

Overview

In this project, I analyzed local and global temperature data and compared the temperature trends of the closest city to where I live (since the city I live in doesn't exist in the database) to overall global temperature trends.

Goal

The goal of this project 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.

Steps

The steps that were taken to reach the goal are:

1- Data Extraction form database

2- Data Exploration and Cleaning

3- Data Visualization

4- Observations

STEP 1: Data Extraction form database

Tools used: **SQL**

1. In this step I need to find the city I live in or the closest to where I live in the database. I live in **Jordan** so I used the below query to check if city I live in exists in the database

```
select *

from city_list

where country = 'Jordan' and city = 'Amman'
```

And I got 0 results

So I wrote the same query for the biggest closest city, I chose **Cairo** that is in **Egypt**

```
select *

from city_list

where country = 'Egypt' and city = 'Cairo'
```

And I got a result

2. After I could retrieve the city, I wrote another query to extract the data I need to calculate the **moving average** which are the **year** and **average temperature** by checking the columns name and data in the **city_data** table

```
select year, avg_temp

from city_data

where city = 'Cairo'
```

And that returned the average temperature from **years 1808 to 2013**
Then I downloaded the results as a **CSV** file and imported them in an **Excel Sheet** to explore them in the next step

3. Similarly, I extracted the **global** data

```
select year, avg_temp

from global_data

where year >= 1808 and year <= 2013
```

And had to add the last condition of the year because the data in **global** data starts from **1750 to 2015** whereas in Cairo it starts from 1808 to 2013 (Did it after the exploration of data)

STEP 2: Data Cleaning and Exploration

Tools: **Excel Sheets**

As I mentioned earlier, the **results** I got from the last queries were **exported as a CSV file** and then I **opened** them as **Excel Sheets** to explore them and see what cleaning steps the data might need.

2.1 Data Cleaning

After I opened them in Excel Sheets I noticed that the data of **global** starts from **1750 to 2015** whereas in **Cairo** it starts from **1808 to 2013**, so there is missing data in Cairo's data from years that are before 1808 and years after 2013 therefore I made the **comparison on** the data in common years (from **1808 to 2013**) and excluded the rest.

2.2 Data Exploration

After I chose the range of years for both Global and Cairo's data to be the same, it's time to calculate the **moving average** for them.

On the Excel Sheet I used **=AVERAGE(B2:B8)** formula to calculate the **moving average for 7 years** for Global and Cairo average temperature, then clicked and dragged the formula down to the next cells to get the average for the rest of data.

The step was performed on Global and Cairo's data similarly.

D8 | fx | =AVERAGE(B2:B8)

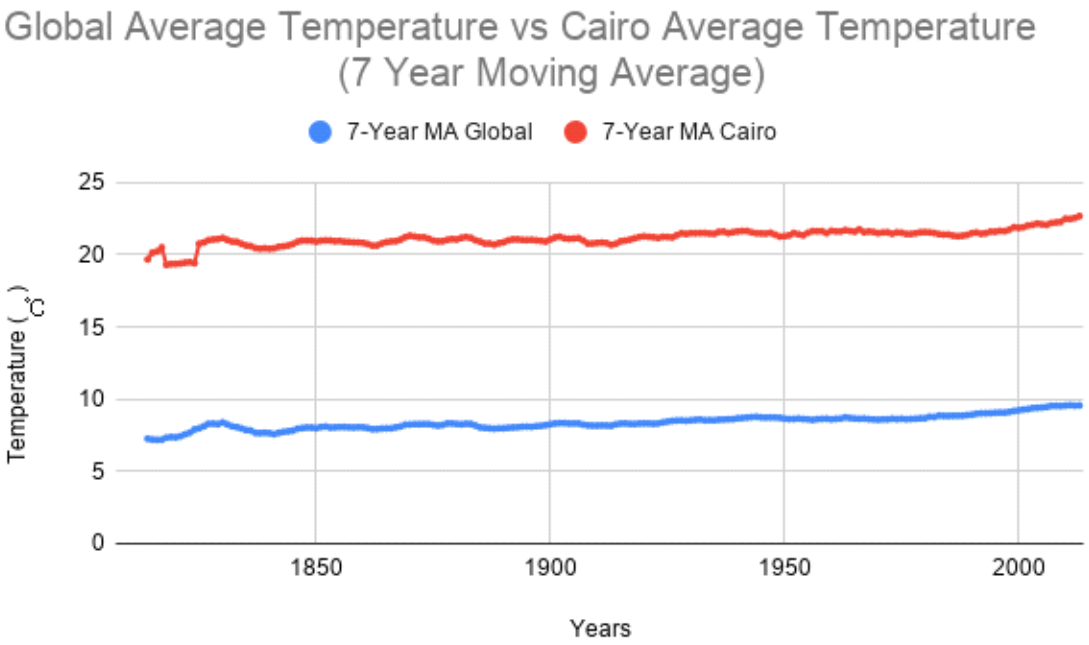
	A	B	C	D	E
1	year	Global avg_temp	Cairo avg_temp	7-Year MA Globa	7-Year MA Cairo
2	1808	7.63	17.11		
3	1809	7.08	19.87		
4	1810	6.92	19.93		
5	1811	6.86	20		
6	1812	7.05	19.93		
7	1813	7.74	20.51	7.267142857 ×	
8	1814	7.59	20.43	=AVERAGE(B2:B8)	19.68285714
9	1815	7.24	20.3	7.211428571	20.13857143
10	1816	6.94	20.51	7.191428571	20.23
11	1817	6.98	21.88	7.2	20.50857143
12	1818	7.83	11.6	7.338571429	19.30857143
13	1819	7.37	20.31	7.384285714	19.36285714
14	1820	7.62	20.58	7.367142857	19.37285714
15	1821	8.09	20.63	7.438571429	19.40142857
16	1822	8.19	20.72	7.574285714	19.46142857

STEP 3: **Data Visualization**

Tools: Excel Sheets

After calculating MA and obtaining the data needed to do the visualization, I used a **line chart** to plot the data so I can be able to do observations in the next step.

I plotted the years on the x-axis and the MA on the y-axis so it's easier to compare and notice the differences.



STEP 4: Observations

From the above visualizations I could observe that:

1. By comparing the Global average temperature and Cairo average temperature , we can see the big difference in the average temperatures as **Global** temperature ranges between **5-10** degrees while **Cairo** slightly **below 20** for almost the first 2 decades and **then above 20** degrees.

2. Cairo is much **hotter** than the Global average temperature.

3. Both of Global and Cairo's temperature kept slightly increasing over years after 1850

4. An observable increase in the average temperature for Cairo in the last 2 decades

5. An observable decrease in Cairo's temperature trends in the 2nd decade then an observable increase afterwards

6. An observable increase in Global temperature trends in the 3rd decade then a slight decrease afterwards

7. A **higher variance** in **Cairo's** temperature trends over years comparing to the Global temperature trends

8. Trends for **Global** were slightly **consistent** from **1850 to 1900** with small increase

9. The **world is getting hotter**, and it remained slightly consistent for some years in the last 10 decades

Correlation Coefficient

Tools: Excel Sheets

I calculated the Correlation Coefficient as another way to have insights other than line chart using this formula **=CORREL(D8:D207, E8:E207)**

Correlation Coefficient for **Global** trends compared to **Cairo's** trends is **0.9129388815** which means that there's a strong relationship among both of them.

Adding another city

***All previous steps were done before adding the new city, so you may see a small difference in the columns header inside the formulas than those on the sheet**

Tools: **SQL**, **Excel Sheets**

I did all the previous steps on another third city I chose which is **Washington**.

STEP 1:

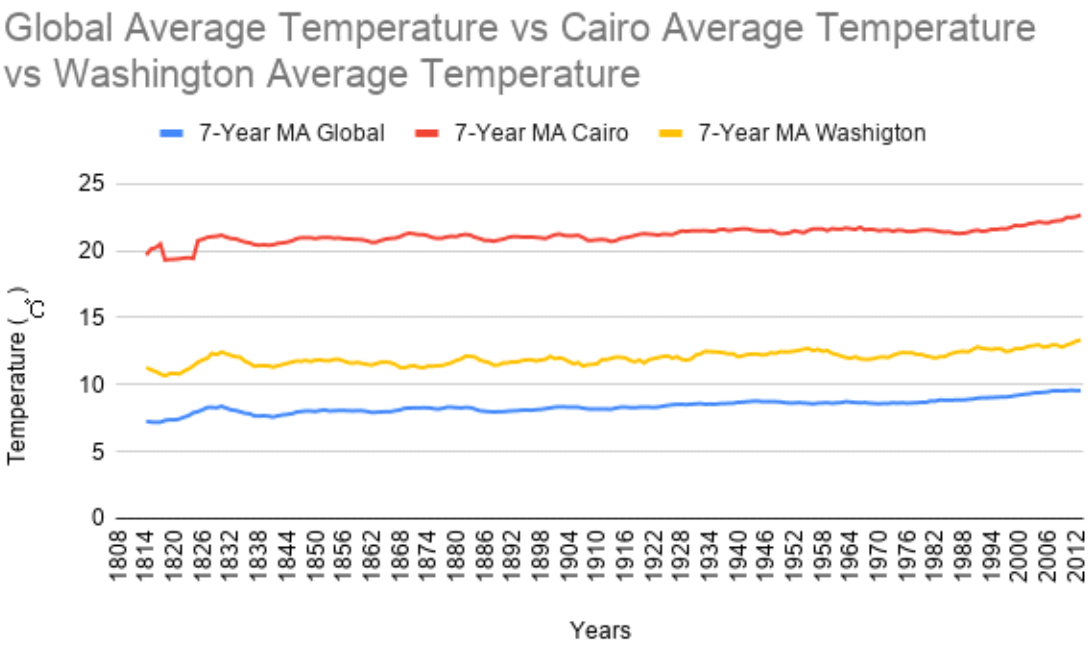
```
select year, avg_temp

from city_data

where city = 'Washington' and year >= 1808 and year <= 2013
```

STEP 2:

2.1 I made the comparison on the data in common years (from 1808 to 2013) and excluded the rest.

2.1 **=AVERAGE(D2:D8)**

STEP 3:



STEP 4:

1. Washington has more **variance** in temperature trends than Global

2. Washington's temperatures range between **10 and 15** degrees

3. An observable increase in Washington's trends starting from 1825 tp 1832

4. Washington is **hotter** than the Global but cooler than Cairo

Correlation Coefficient

I calculated the Correlation Coefficient using this formula **=CORREL(E8:E207, G8:G207)**

Correlation Coefficient for Global trends comparing to Washington's trends is **0.8930217116** which means that there's a strong relationship among both of them but *less than the relationship between Global and Cairo*.