

# EXPLORE WEATHER TRENDS

---

Project- 1

June 1, 2020

Udacity- Data Analyst Nonodegree

**Elham Daha**

## Overview

In this project, I will analyze local temperature and the global temperature data. The comparison is between the average temperature trends in Nashville, Tennessee in USA, where I live, with the overall global temperature trends.

## Instruction

- **Extract the data:** I extracted data from the database using SQL and exported to CSV file.
- **Open up the CSV:** I used Excel Spreadsheet to open the CSV and calculate moving average of city vs global temperature.
- **Create a line chart:** I plot a line chart visualizing in excel to compare my city's temperature with the global temperature.
- **Make observations:** my conclusion about the similarities and differences between the world averages and my city's averages, as well as overall trends.

## Extracting data from the database using SQL

**Step 1:** to extract the global data- there are 266 results available to download to CSV

The screenshot shows a web-based SQL interface. On the left, under the 'Input' tab, there is a 'SCHEMA' section with a refresh icon and a list of tables: 'city\_data', 'city\_list', and 'global\_data', each with a dropdown arrow. The main query area contains a single SQL statement: `1 select * from global_data`. Below the query area, a green bar indicates 'Success!'. To the right of this bar is a blue button labeled 'EVALUATE'. At the bottom, the 'Output' section shows '266 results' and a blue link with a download icon labeled 'Download CSV'.

**Step 2:** to see the available cities for country USA

The screenshot shows the same SQL interface. The 'SCHEMA' section now includes an additional table, 'city', and the 'city\_list' table has an upward arrow. The query area contains two SQL statements: `1 select * from city_list` and `2 where country= 'United States'`. A green bar indicates 'Success!'. The 'EVALUATE' button is present. The 'Output' section shows '52 results' and a 'Download CSV' link. Below this, a table of results is visible with two columns: city names and 'United States'. The visible rows are: Kansas City, Las Vegas, and Long Beach.

I live in Nashville and I will work on this city. There are 52 results for Nashville. As long as city\_data and global\_data has a column with a same name “avg\_temp” which that one contains the global temperature and the other one presents the city temperature, I renamed the column “avg\_temp” in “global\_avg\_temp” and “city\_avg\_temp” by using Alter Table in SQL.

The screenshot shows a SQL interface with an 'Input' section. On the left, there is a 'SCHEMA' dropdown menu with a refresh icon. Below it, a list of tables is shown: 'city\_data', 'city\_list', 'city', 'country', and 'global\_data'. The 'city\_data' and 'global\_data' tables are selected. To the right of the schema list, there are two SQL queries:

- 1 `Alter Table city_data rename column avg_temp to city_avg_temp;`
- 2 `Alter Table global_data rename column avg_temp to global_avg_temp;`

Below the queries, there is a green 'Success!' message and a blue 'EVALUATE' button. At the bottom, the 'Output' section shows 'No data to download'.

### Step 3: Joining two tables

The screenshot shows a SQL interface with an 'Input' section. On the left, the same schema list is visible. To the right, there are five SQL queries:

- 1 `Select g.year, g.global_avg_temp, c.city_avg_temp`
- 2 `from global_data as g`
- 3 `join city_data as c`
- 4 `on g.year=c.year`
- 5 `where city='Nashville'`

Below the queries, there is a green 'Success!' message and a blue 'EVALUATE' button. The 'Output' section shows '264 results' and a 'Download CSV' button. Below this, a table of results is displayed:

year	global_avg_temp	city_avg_temp
1750	8.72	14.92
1751	7.98	15.72
1752	5.78	8.54
1753	8.39	14.27
1754	8.47	14.32
1755	8.36	11.90
1756	8.85	14.44
1757	9.02	13.92

At the bottom of the table, there are 'Menu' and 'Expand' buttons.

I got the required data with 264 results. In order to complete the extraction part, I downloaded the file as CSV format as “results.csv”.

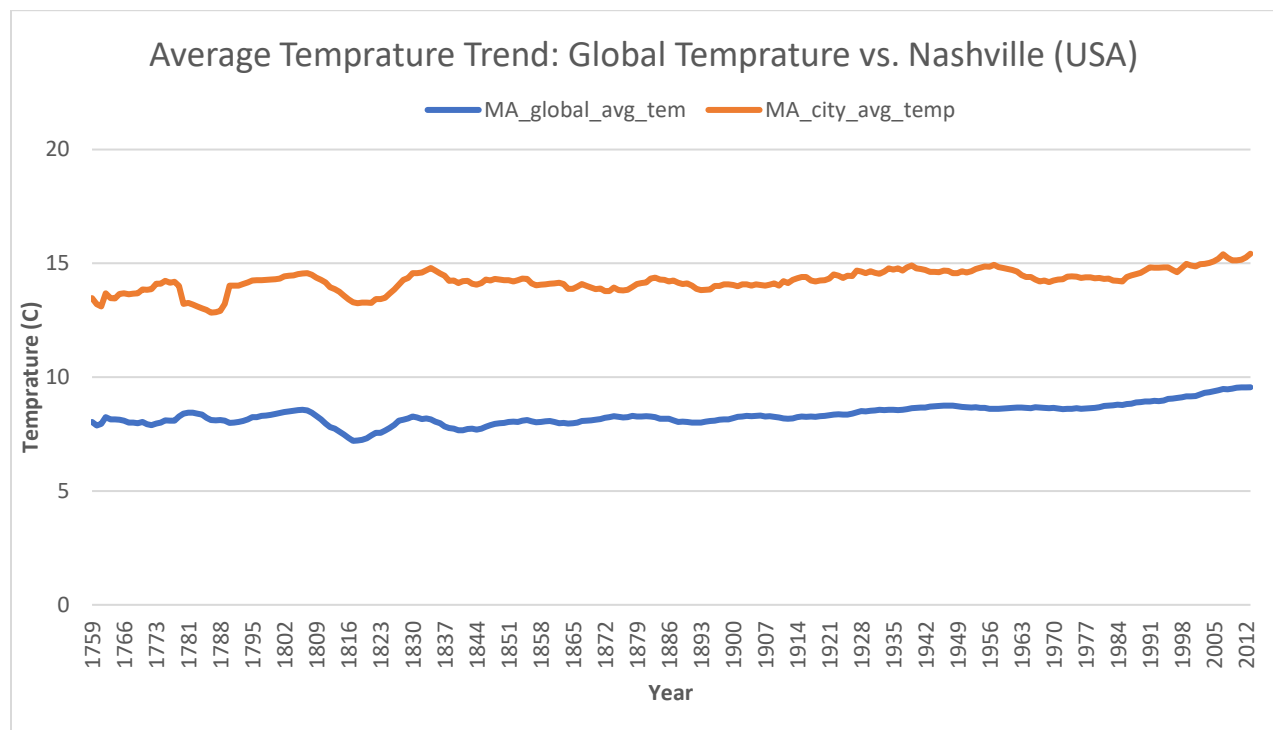
## Moving Average & Trend Visualization

**Step 1:** CSV file opened in excel spreadsheet (or Google Sheets) and I deleted the missing year data (used if function to find the null columns). Only one row was deleted due to null value in city\_avg\_temp.

**Step 2:** Calculate the moving average for the first 10 years to smooth out the lines. (In my example, 10 years moving average line chart is smoother than 5 years)

year	global_avg_temp	MA_global_avg_tem	city_avg_temp	MA_city_avg_tem
1750	8.72		14.92	
1751	7.98		15.72	
1752	5.78		8.54	
1753	8.39		14.27	
1754	8.47		14.32	
1755	8.26		13.9	
1756	8.85		14.44	
1757	9.02		13.92	
1758	6.74		12.81	
1759	7.99	8.03	13.83	AVERAGE(D7:D16)
1760	7.25	7.877	12.28	AVERAGE(D8:D17)
1761	8.17	7.956	14.77	13.108
1762	8.81	8.239	14.27	13.081
1763	7.5	8.35	12.2	13.474
1764	8.4	8.145	14.18	13.46
1765	8.25	8.332	13.85	13.456
1766	8.43	8.008	14.7	13.082
1767	8.32	8.008	13.57	13.64
1768	6.78	8.012	13.09	13.468
1769	7.68	7.862	14.02	13.487
1770	7.69	8.032	13.88	13.847
1771	7.85	7.94	14.68	13.838
1772	8.18	7.898	14.52	13.863
1773	8.25	7.77	14.01	14.104
1774	8.77	8.007	14.17	14.103
1775	9.18	8.1	15.15	14.232
1776	8.5	8.085	13.89	14.153
1777	8.26	8.093	13.89	14.19
1778	8.54	8.265	11.23	14.064
1779	8.06	8.308	6.17	13.225
1781	8.1	8.439	14.37	13.208

**Step 3:** created a line chart for global average temperature and the City of Nashville average temperature to compare their trends. In order to visualize the trends, I considered to show a smooth trend rather than a fluctuated one. So, I worked with the 10-year moving average (tested two ranges). I tried to change column year type to date in order to format the unit and bound of “year” axis. But data only contains “yyyy” and I was not able to convert it to date in SQL. However, this problem can be solved in Tableau.



## Observations

- Nashville's temperature is higher than the overall global temperature.
- In 18<sup>th</sup> century, the global temperature trend was steady, but Nashville temperature was increasing. Overall, there is a correlation between global and local temperature. (checked the correlation and it is 0.7)
- Overall trend is incremental. Even though it is not increasing rapidly but the city's and global temperature is increasing in the entire time frame which can be due to increase Global Warming, specifically in 20<sup>th</sup> century. In 18<sup>th</sup> and 19 centuries, the global temperature was on constant rise.
- In 1780s, Nashville's temperature fell significantly while the global temperature was almost steady. The beginning of 19<sup>th</sup> century, both global and local temperature had a significant decrease that did not last long and after few years increased to its peak. Still, today's temperature is higher than 19<sup>th</sup> century peak.
- According to the graph, average global temperature had only increased 1.5°C throughout 260 years while Nashville average temperature had an increase of 2°C. It means that Nashville is going to be hotter than the overall global temperature.

## PS:

I googled to find a convert code in case if I can change the year type.

I checked GitHub of some students.