Project: Exploring Weather Trends

Dec 31. 2021

Introduction

In this project, I analyzed the local and global temperature data, and compared the temperature trends where I live to overall global temperature trends. Based on provided data, Raleigh is the nearest city to me and it has available temperature data. I visualized and summarized the temperature trends following these steps: a) Extract data; b) Open up the CSV file and process the data; c) Create a line chart; d) Make observations. The following sections will represent each step.

Step 1: Extract Data

With the provided workspace in Udacity, I exported temperature data from the world as well as for Raleigh. Fig. 1 is the screenshot of the SQL query that I wrote to extract the global data, and I saved the data in "global_data.csv". Then I worked on the local data. I first checked whether Raleigh exists in the provided database, as shown in Fig. 2. Since Raleigh does included in the database, I wrote another SQL query to extract its local temperature data, as given in Fig. 3. I saved the local data in "city_data.csv".

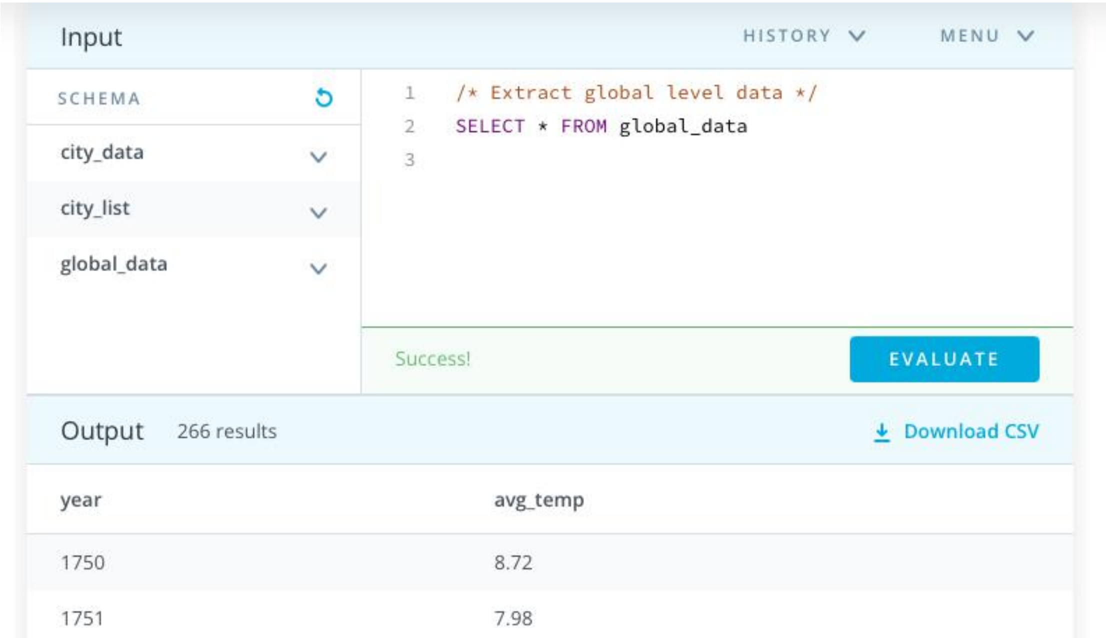


Figure 1. The screenshot of the SQL query to extract global data

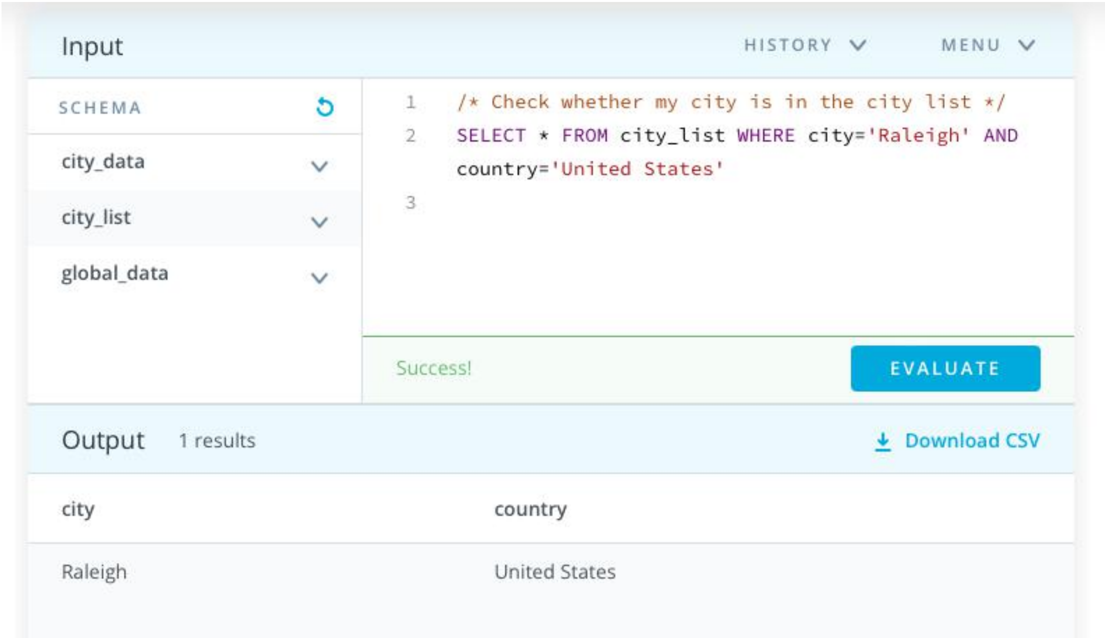


Figure 2. The screenshot of the SQL query to check whether Raleigh exists

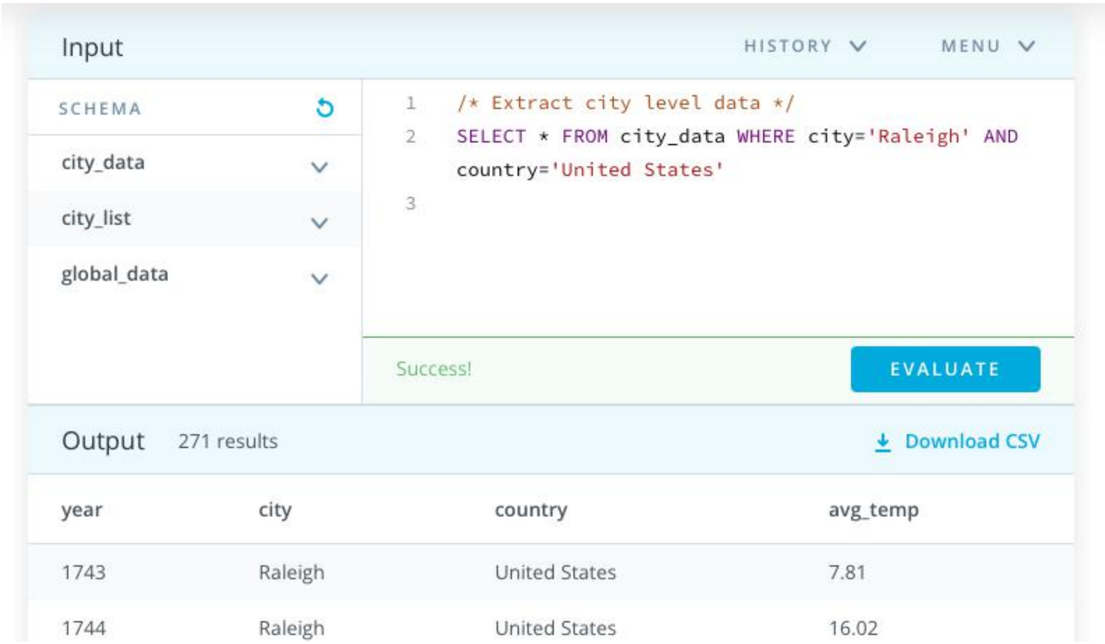
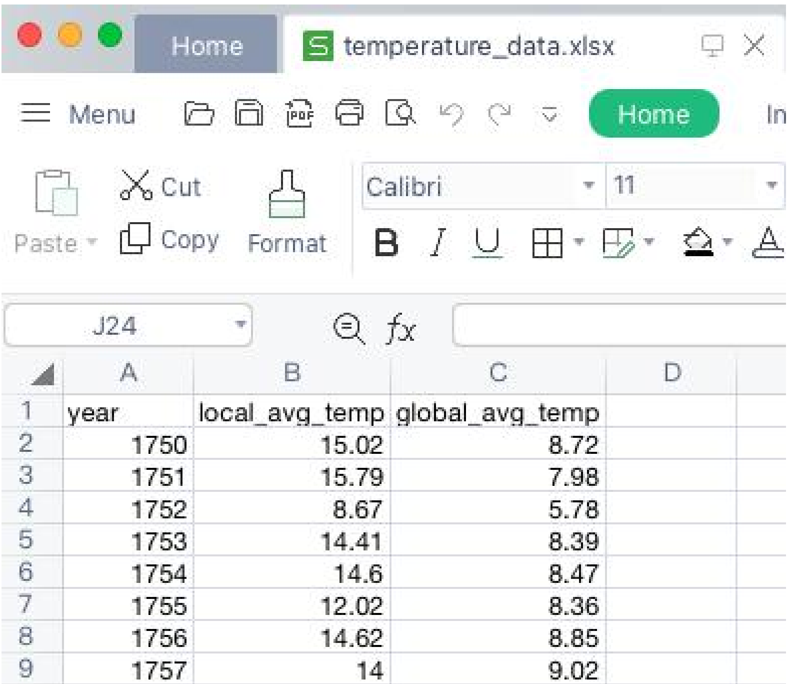


Figure 3. The screenshot of the SQL query to extract Raleigh's data

Step 2: Open up the CSV and process the data

A) Open up the CSV file and organize the data

As it is recommended in the description of this project, I used Excel to open, process and visualize the data. Since Raleigh's temperature data is from 1743 to 2013, while the world temperature record is from 1750 to 2015, I merged "global_data.csv" and "local_data.csv" to a new file, selected data from 1750 to 2013, and saved the data in a new file named as "temperature_data.xlsx", as shown in Fig. 4.

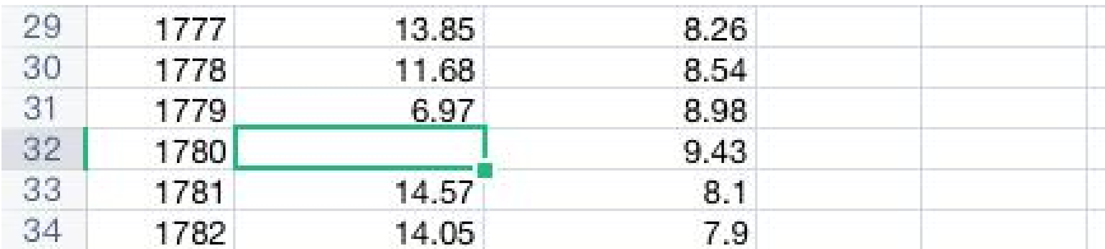


	A	B	C	D
1	year	local_avg_temp	global_avg_temp	
2	1750	15.02	8.72	
3	1751	15.79	7.98	
4	1752	8.67	5.78	
5	1753	14.41	8.39	
6	1754	14.6	8.47	
7	1755	12.02	8.36	
8	1756	14.62	8.85	
9	1757	14	9.02	

Figure 4. The screenshot of the opened and processed data

B) Deal with the missing data

Since there is one year - 1780, that is not provided with local average temperature, I need to treat this missing data, see Fig. 5. Considering it is only 1 year's data among hundreds of years' data, the missing data's influence to the final result could be ignored, I used the average temperature of 1779 and 1781 to fill the missing data in 1780. Then the filled data for the year of 1780 would be (6.97+14.57)/2=10.76 (approximate to the second decimal). Note that if there are too many years of data missing, the treatment would be different.



29	1777	13.85	8.26		
30	1778	11.68	8.54		
31	1779	6.97	8.98		
32	1780		9.43		
33	1781	14.57	8.1		
34	1782	14.05	7.9		

Figure 5. The screenshot of the year that lacks of local average temperature

Step 3: Create a line chart

Instead of directly obtaining the figure of moving averages, I started with yearly temperature at first, as shown in Fig. 6. It is clear that Raleigh's temperature is always higher than global temperature despite one year between 1750 to 1800. The chart also looks too volatile and the moving average strategy would be helpful to smooth the chart.

Considering temperature would not have large change within 5 years, I first selected the moving average window to be 5 years. The formula that I used to is shown in the red circle in Fig. 7. This is the formula to calculate Raleigh's 5-year moving average of year 1750 to year 1754. For the rest of this column, I just cluck and drag the formula down to the last row of this column. For the global 5-year moving average, I just just same strategy but in different column. The formula to calculate global 5-year moving average of year 1750 to year 1754 that I input was "=AVERAGE(C2:C6)".

The 5-year moving average result is given in Fig. 8. Comparing to the curve of yearly temperature, the curve of 5-year moving averages looks smoother. To further display the effect of moving averages, I used 20 year's average temperature (with input "=AVERAGE(B2:B21)" and "=AVERAGE(C2:C21)"), which is shown in Fig. 9. Comparing Fig. 7, 8,and 9, it is clear that the more years that moving window includes, the smoother the curve of the average temperature would be.

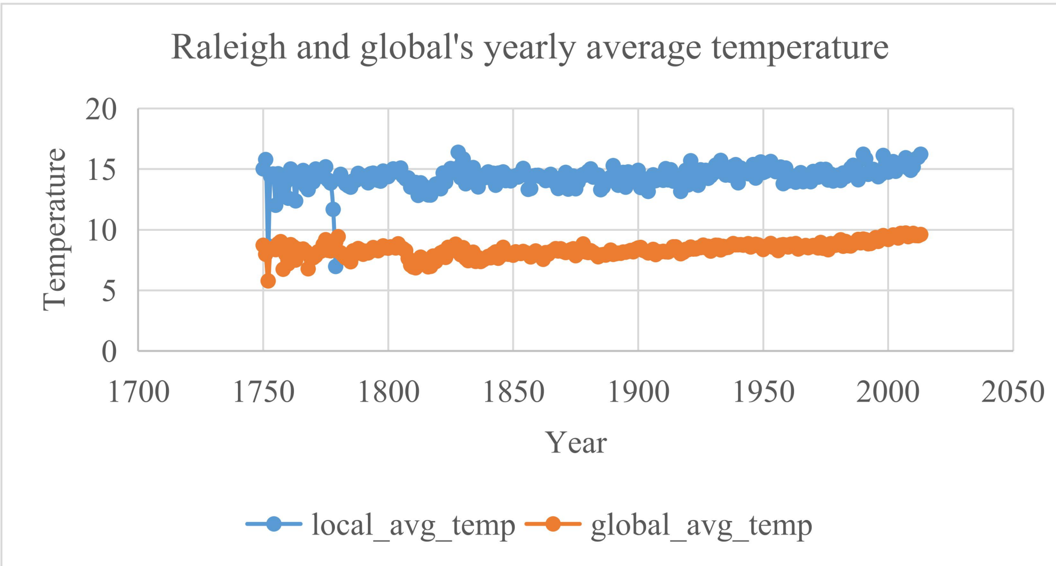
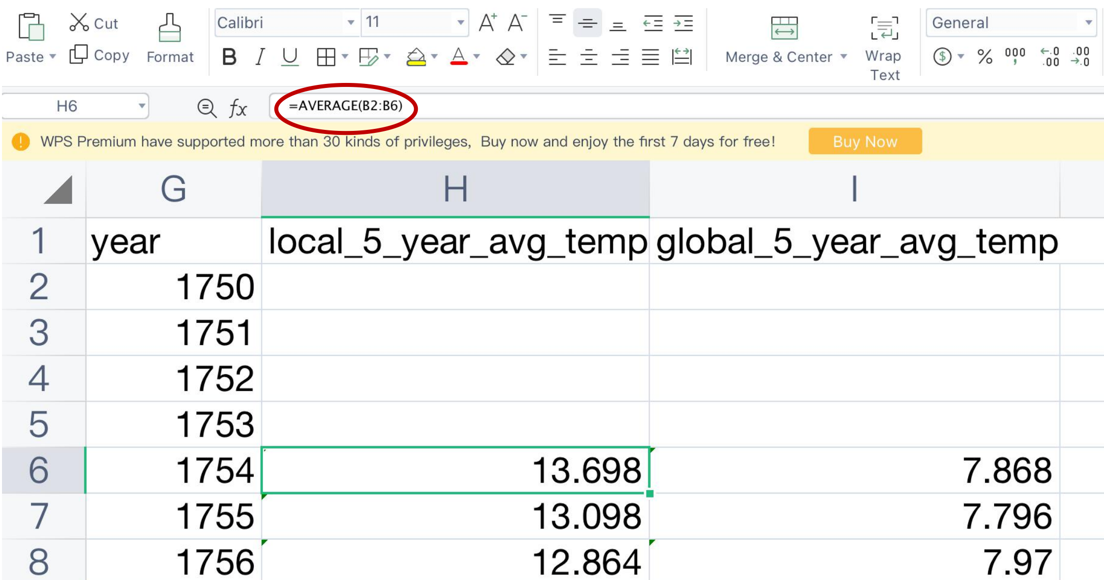


Figure 6. The chart of Raleigh and global's yearly average temperature



	G	H	I
1	year	local_5_year_avg_temp	global_5_year_avg_temp
2	1750		
3	1751		
4	1752		
5	1753		
6	1754	13.698	7.868
7	1755	13.098	7.796
8	1756	12.864	7.97

Figure 7. The screenshot of formula to calculate 5-year moving average temperature

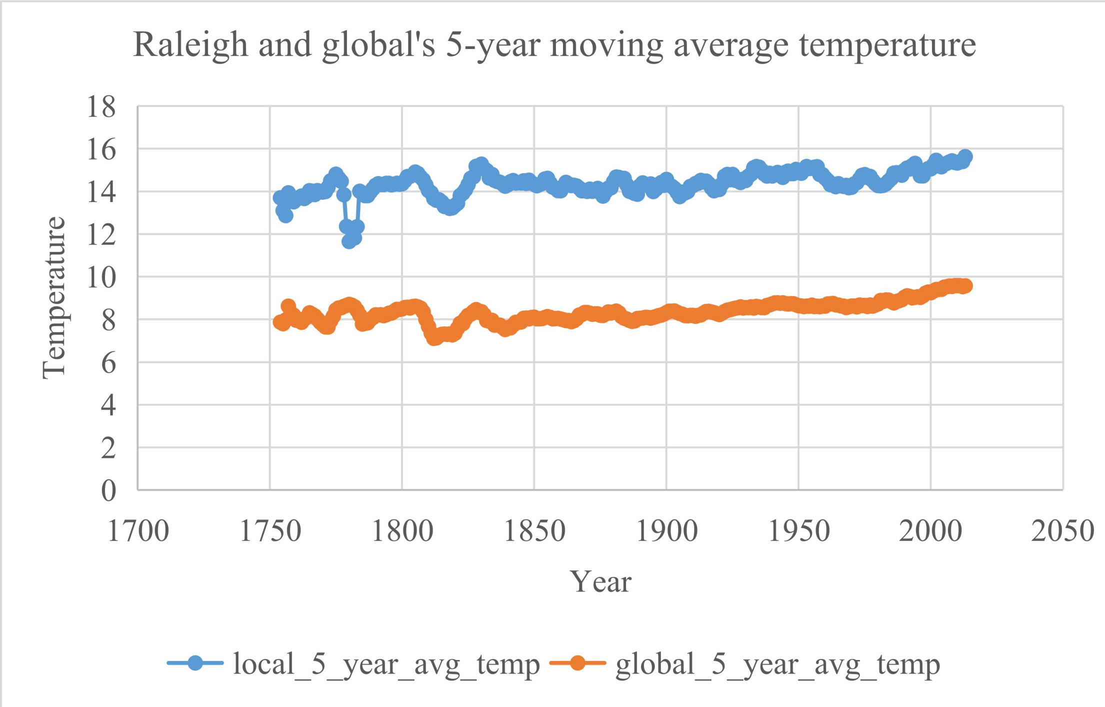


Figure 8. The chart of Raleigh and global's 5-year moving average temperature

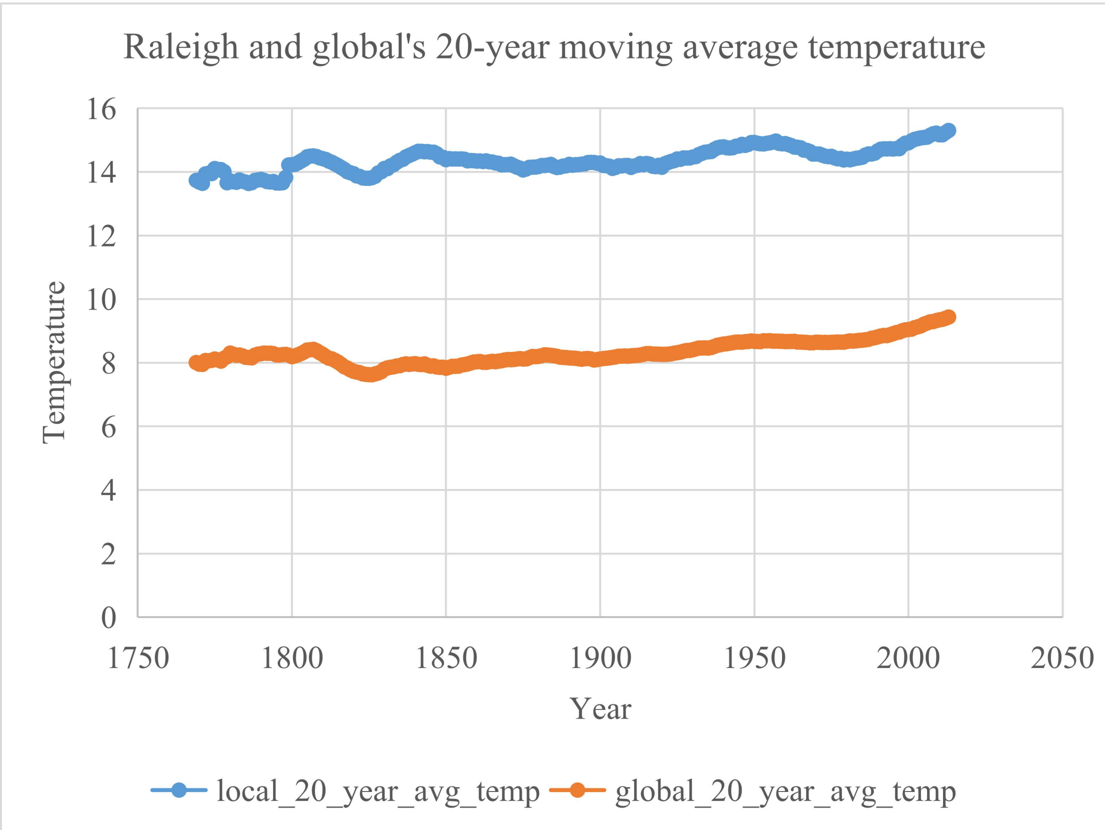


Figure 9. The chart of Raleigh and global's 20-year moving average temperature

Step 4: Make observations

Based on the visualized data, here are my observations:

- Despite one year (between 1750 and 1800), Raleigh is hotter comparing to the global average.
- Based on the chart, the difference of Raleigh's temperature and global temperature has been consistent over time.
- Overall speaking, Raleigh's local temperature changes along with global temperature: when the global temperature is higher, Raleigh's local temperature typically also gets higher; when the global temperature is lower, Raleigh's local temperature typically also gets lower.
- Overall speaking, both global and Raleigh's local temperature show a tendency of increasing over years.