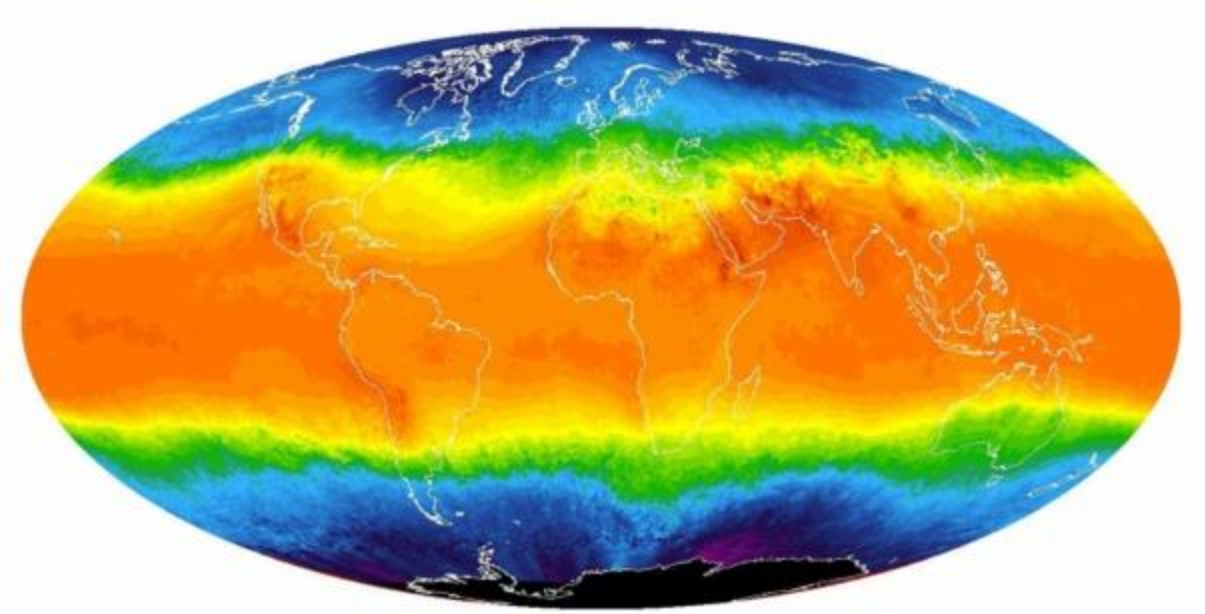
1 Project One



EXPLORING WEATHER TRENDS

by

Chun Shen

Data Analyst Nanodegree - Udacity

May, 2019

1. Summary

This technical report is in response to the project requirement from Data Analyst Nanodegree by Udacity starting with May 14, 2019. The report analyzes local (Houston) and global temperature data and compares local (Houston) temperature trends to overall global temperature trends.

To analysis temperature trends, SQL queries were written to extract the data from a database, moving averages calculations were performed to smooth data, data visualization were created using Excel. Finally, the report provides the similarities and differences between local (Houston) and global temperatures.

Overall conclusions for the temperature data analysis are:

- Both global and Houston are getting warmer.
- In the past 30 years, temperature changes significantly globally.
- Houston is much hotter in the world.
- During early time, Houston temperature is relatively stable.

2. Objective

In the project, we will get familiar with SQL, and practice to download data from a database. Then we will create a visualization using Excel and analyze local and global temperature data, and accurately make four observations.

3. Technical Work

Houston and global average temperature data analysis were performed to show the similarities and differences between them.

3.1 Data Extraction

Four SQL queries were written on the provided workspace which was connected to a database. There are three tables in the database:

- city_list - This contains a list of cities and countries in the database.
- city_data - This contains the average temperatures for each city by year (ºC).
- global_data - This contains the average global temperatures by year (ºC).

3.1.1 SQL query to survey city_list

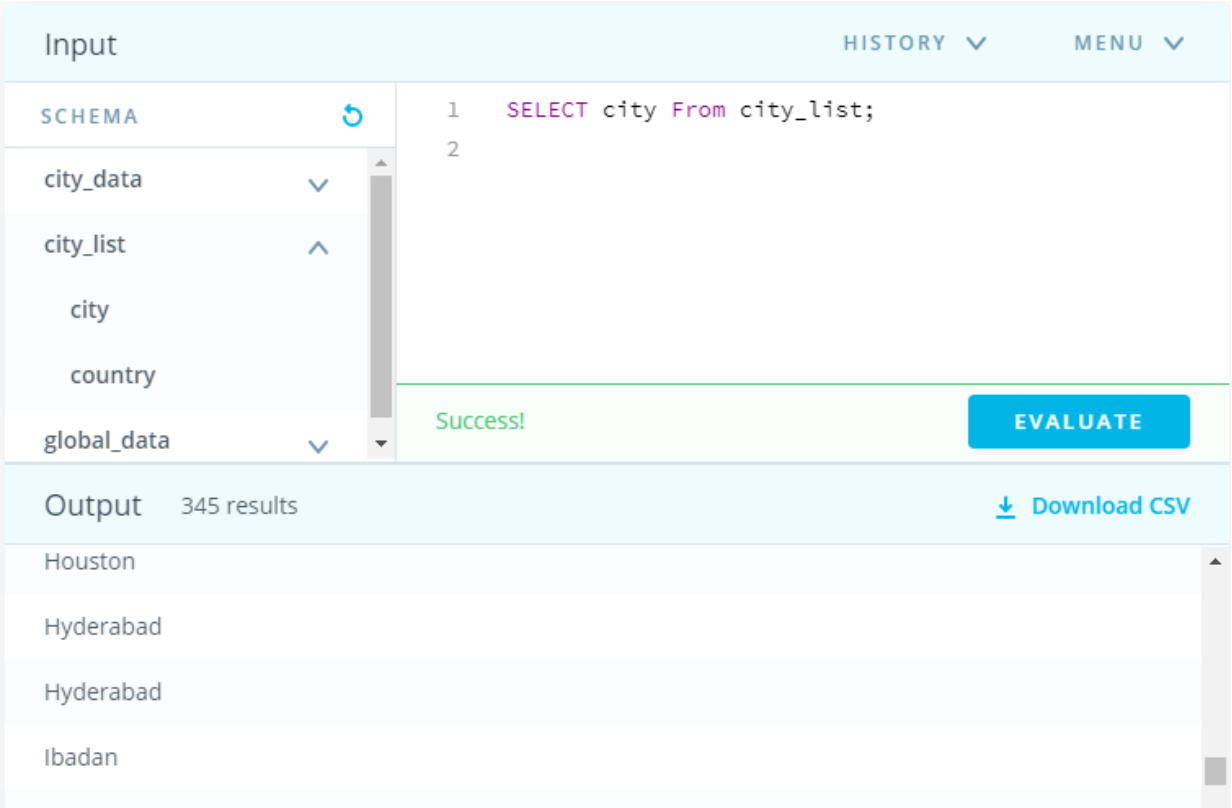


Fig. 1 SQL query to view the details of city_list

The table includes total 332 cities from 137 countries. The city I live in was not on the list, so I select Houston which is nearest to me.

3.1.2 SQL query to extract the city level data

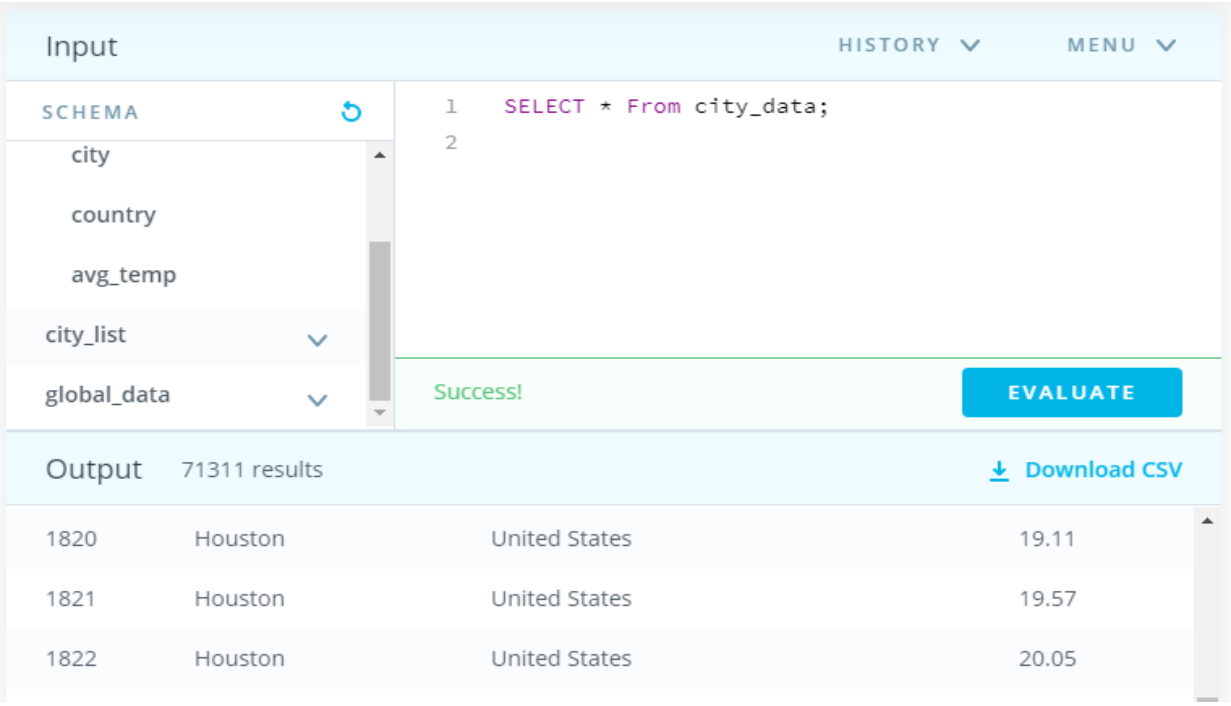


Fig. 2 SQL query to view the details of city_data

This table records yearly average temperatures data. There are totally 194 continuous data points from 1820-2013 for Houston. Then the Houston data was exported to CSV.

	A	B
1	year	avg_temp
2	1820	19.11
3	1821	19.57
4	1822	20.05
5	1823	19.62
6	1824	20.19
7	1825	20.44
8	1826	20.17
9	1827	20.83
10	1828	20.41
11	1829	20.00

Fig. 3 Export Houston data to CSV

3.1.3 SQL query to extract the global data

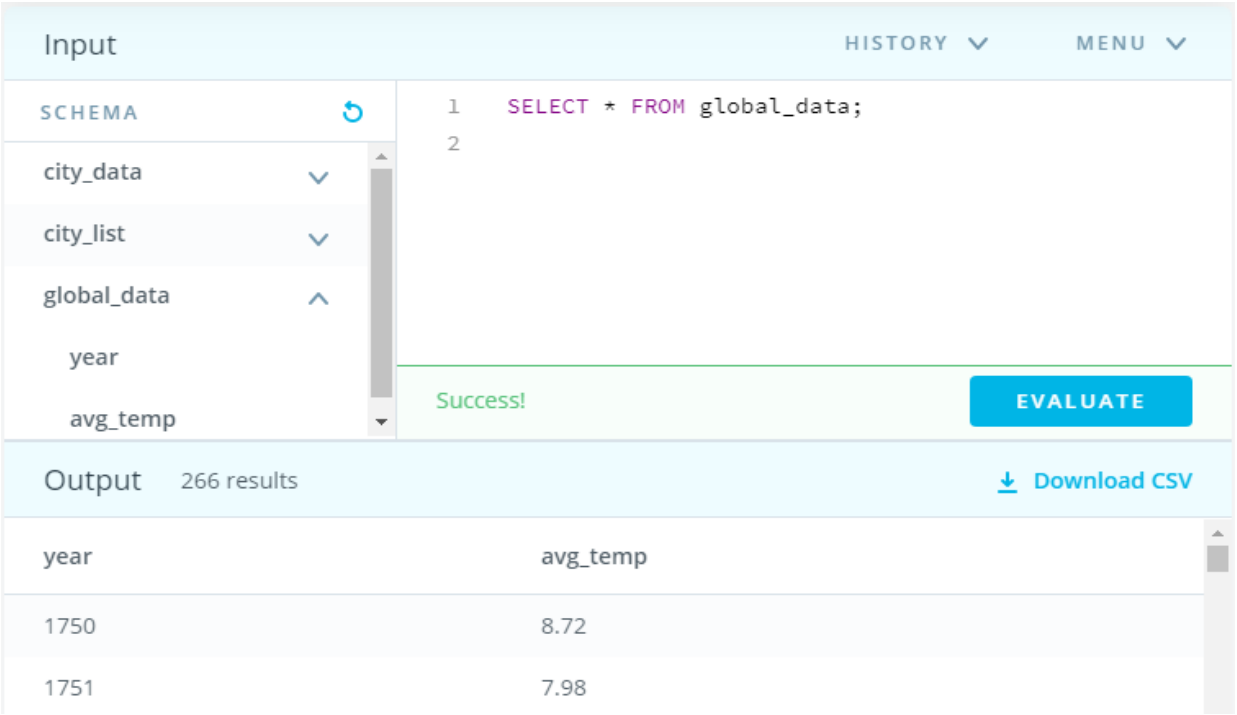


Fig. 4 SQL query to view the details of global data

Global temperature data starts with 1750, and has the records for 266 years up to 2015. The data was successfully exported to CSV.

	A	B
1	year	avg_temp
2	1750	8.72
3	1751	7.98
4	1752	5.78
5	1753	8.39
6	1754	8.47
7	1755	8.36
8	1756	8.85
9	1757	9.02
10	1758	6.74
11	1759	7.99
12	1760	7.19

Fig.5 Export global data to CSV

3.1.4 A SQL query to join city_data and global_data

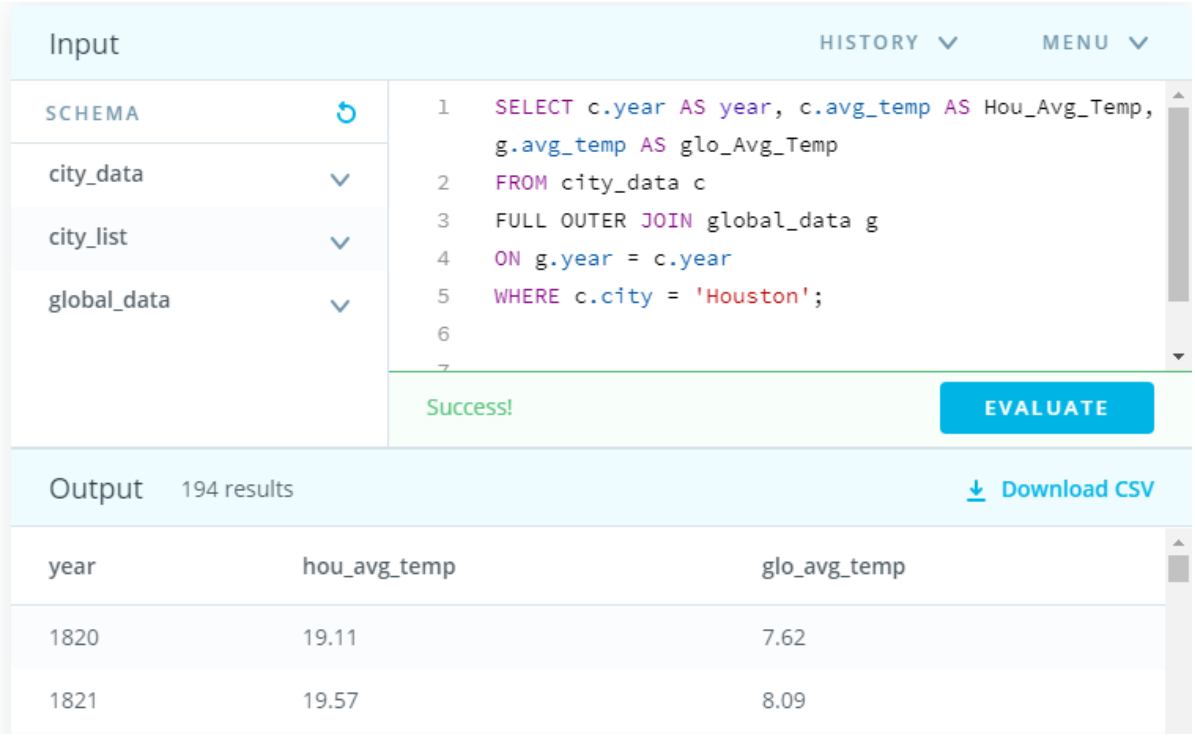


Fig. 6 SQL query to combine city_data and global_data

FULL OUTER JOIN was used to join city and global tables, so that we can export needed information from both tables at same time, and get alerts if there are any mismatched years. Finally, data was sucessfully exported to CSV.

	A	B	C
1	year	hou_avg_temp	glo_avg_temp
2	1820	19.11	7.62
3	1821	19.57	8.09
4	1822	20.05	8.19
5	1823	19.62	7.72
6	1824	20.19	8.55
7	1825	20.44	8.39
8	1826	20.17	8.36
9	1827	20.83	8.81

Fig. 7 Export local and global data to CSV

3.2 Data Visualization

In this project, key consideration of data visualization is to create clear charts so that we can easily observe the long term trends of Houston and global temperatures.

3.2.1 Global Temperature Trend

Open up global CSV file, and start with original yearly average data. The chart has a lot of fluctuations as Fig. 8 shows. To better interpret the data and make it easier to observe long term trends, moving average calculation was applied to smooth out data.

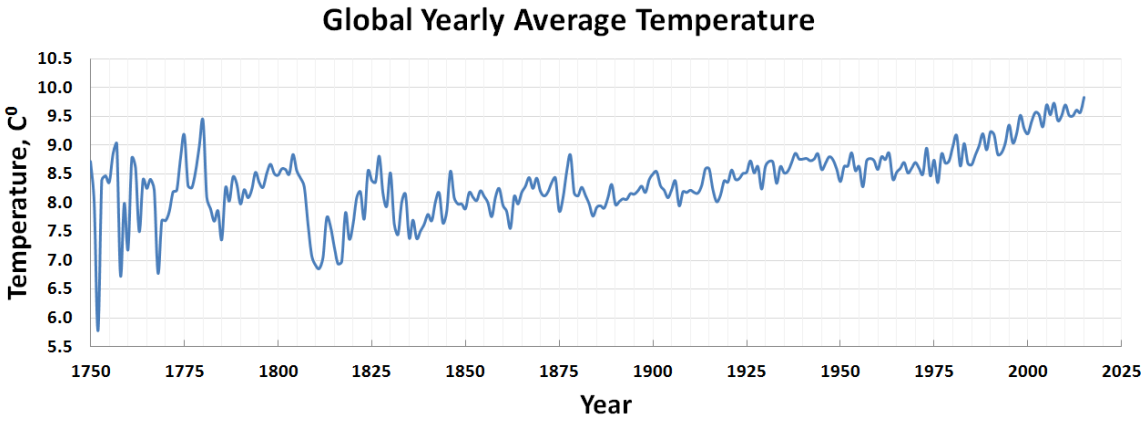


Fig. 8 Global yearly average temperature

Moving Average

We calculate moving averages in the spreadsheet. We select 10-years moving averages since 10-years seems a good time span, which captures a clear long-term trend. To start, create a second column called 10-year avg_temp, which is where the moving average field will be stored. Go down to the tenth year (1975) and use the AVERAGE() function to calculate the average temperatures for the first ten years of avg_temp data, as seen in Fig. 9.

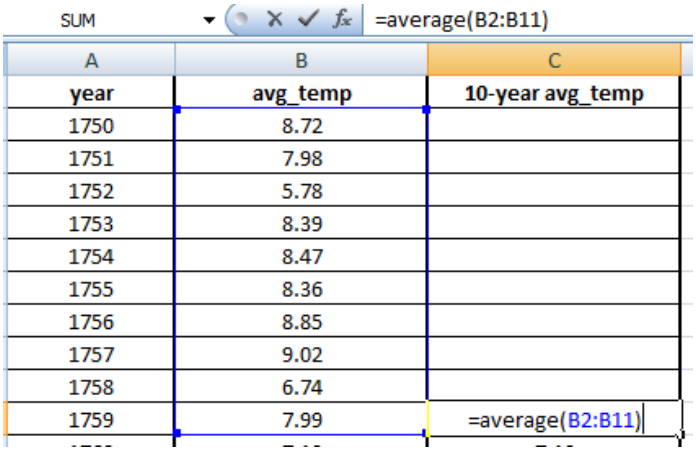


Fig. 9 Moving average calculation example

Then click and drag the formula all the way down to the end of the data.

10-Year Moving Average Temp

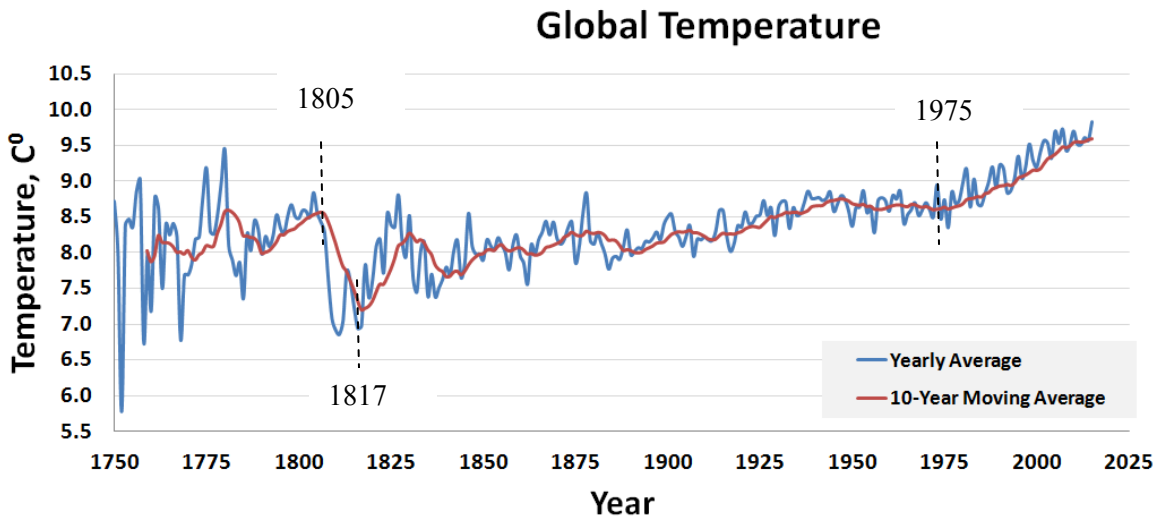


Fig. 10 Global yearly and 10-year average temperatures

The red line in Fig. 10 shows global temperature trend based on10-year moving averages, which makes it much easier to observe changes in the certain periods.

Observations

- In general trend, global is getting hotter. Especially, since 1975, global temperature has been steadily increasing, which may explain why glaciers are melting at a faster rate and sea levels are rising now.
- Notice there is big drop in global temperature between 1805 and 1817, and the total change in temperature is more than 1^{o}.

3.2.2 Houston Temperature Trend

Again, we started with original yearly average data. Since the chart also shows a lot of fluctuations, 10-years moving average was applied to smooth our data as we did for global data.

Fig. 11 below shows Houston temperature trends from 1820 to 2013 based on yearly and 10-year moving averages. Clearly, 10-year moving average line (red one) is more easily to observe a long term trend.

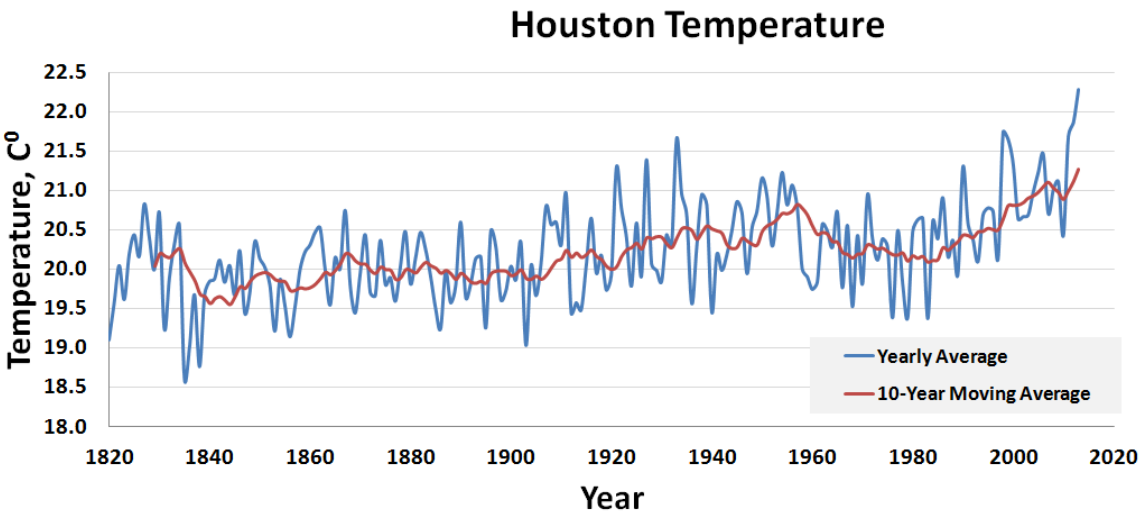


Fig. 11 Houston yearly and 10-year average temperatures

Observations

- In general trend, Houston is getting warmer, following the global trend.
- Before 1900, there wasn't significant temperature change in Houston. The upward trend becomes clear during 1900-1957 and 1983 to 2013 which is end data recored in the table.

3.2.3 Comparison of Houston trend with global trend

Fig. 12 compares temperature trends for Houston and Gobal with same horizontal and vertical scales during 1820 - 2013 based on yearly and 10-year moving averages.

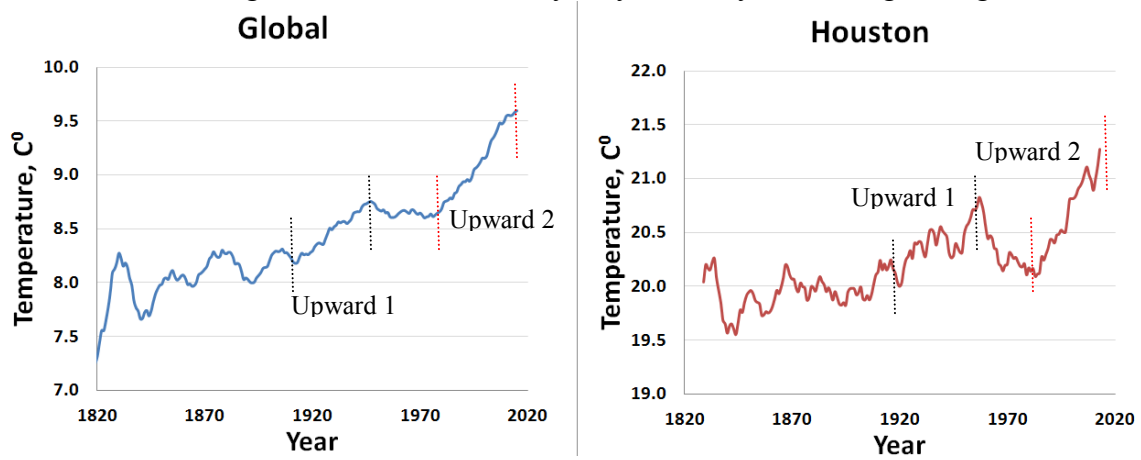


Fig. 12 10-year Moving Average Temperature Comparison

Observations

- Before 1920, Houston temperature change is relatively small compared to global temperature.
- Both shows two clear upward trends after 1920, and temperature rises very quickly in the second upward period.
- Between two rising periods, global temperature is relatively stable while Houston became cooler. Even that, Houston temperature is much higher than globals all the time.

4. Conclusions

Based on data analysis on Houston and global data, we have conclusions:

- Both global and Houston are getting warmer.
- During the recent 30 years, temperature changes significantly in both Houston and global (roughly 1 C^o since 1970 as shown in the Upward 2 in Fig. 12)
- Houston is much hotter in the world.
- Before first clear rising trend, Houston temperature is relatively stable.