Explore Weather Trend

Exploration Steps:

1. Select the right city to investigate for this project.

Use a SQL query to find all available cities in the United States since I currently live in US. The code is shown as below:

```
SELECT country, city
FROM city_list
WHERE country = 'United States'
ORDER BY 1
```

After running the query, I checked all available cities in the results and noticed that New York City (NYC) is the nearest big city. Thus, my project will be focusing on the similarities and differences between NYC and global temperature data.

2. Select and download NYC and global temperature data.

I ran two queries separately to select NYC and global temperature data and downloaded them as CSV files. The two queries codes are shown as below:

```
/* NYC data*/
SELECT *
FROM city_data
WHERE city = 'New York'
/*Global data*/
SELECT *
FROM global_data
```

3. Check and clean data.

I opened the two CSV files to go through the data first. A big difference between NYC data and global data is that NYC data is between 1743 and 2013 while global data is between 1750 and 2015. Also, for NYC data, 5 data points were missing: from 1746 to 1749 and 1780.

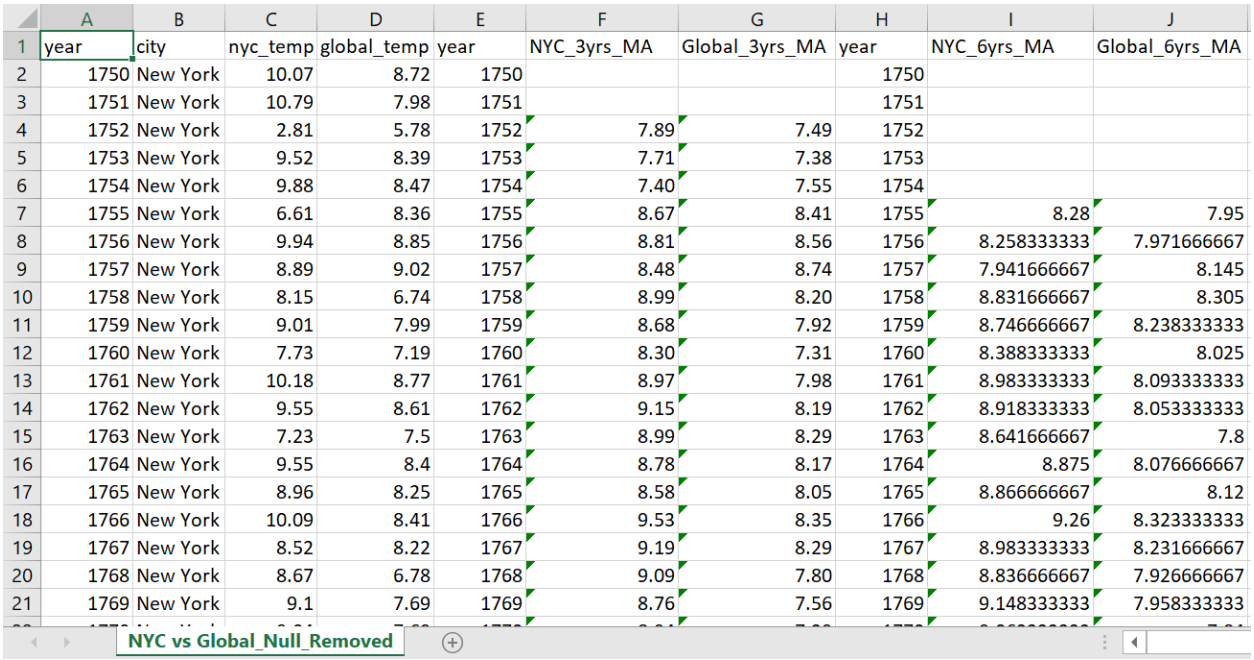
For comparison, it'd be better to put NYC data and global data in the same CSV file and only include the yearly temperature data that exist in both datasets. To achieve this, I use a SQL query with inner join and where clause to clean up the data. The code is as below:

```
SELECT c.year, c.city,
       c.avg_temp nyc_temp,
       g.avg_temp global_temp
FROM city_data c
JOIN global_data g
ON g.year = c.year
WHERE c.city = 'New York' AND c.avg_temp IS NOT NULL AND g.avg_temp IS NOT
NULL
```

After running this query, I downloaded the data as a CSV file. It contains data from both NYC and global datasets and removes all unmatched or Null data.

4. Calculating moving averages.

I opened the CSV file in step 3 in Excel. As instructed in the Udacity online course, I calculated 3yrs MA and 6yrs MA for both NYC and global data. Below is a screenshot of the Excel file interface.



	A	B	C	D	E	F	G	H	I	J
1	year	city	nyc_temp	global_temp	year	NYC_3yrs_MA	Global_3yrs_MA	year	NYC_6yrs_MA	Global_6yrs_MA
2	1750	New York	10.07	8.72	1750			1750		
3	1751	New York	10.79	7.98	1751			1751		
4	1752	New York	2.81	5.78	1752	7.89	7.49	1752		
5	1753	New York	9.52	8.39	1753	7.71	7.38	1753		
6	1754	New York	9.88	8.47	1754	7.40	7.55	1754		
7	1755	New York	6.61	8.36	1755	8.67	8.41	1755	8.28	7.95
8	1756	New York	9.94	8.85	1756	8.81	8.56	1756	8.258333333	7.971666667
9	1757	New York	8.89	9.02	1757	8.48	8.74	1757	7.941666667	8.145
10	1758	New York	8.15	6.74	1758	8.99	8.20	1758	8.831666667	8.305
11	1759	New York	9.01	7.99	1759	8.68	7.92	1759	8.746666667	8.238333333
12	1760	New York	7.73	7.19	1760	8.30	7.31	1760	8.388333333	8.025
13	1761	New York	10.18	8.77	1761	8.97	7.98	1761	8.983333333	8.093333333
14	1762	New York	9.55	8.61	1762	9.15	8.19	1762	8.918333333	8.053333333
15	1763	New York	7.23	7.5	1763	8.99	8.29	1763	8.641666667	7.8
16	1764	New York	9.55	8.4	1764	8.78	8.17	1764	8.875	8.076666667
17	1765	New York	8.96	8.25	1765	8.58	8.05	1765	8.866666667	8.12
18	1766	New York	10.09	8.41	1766	9.53	8.35	1766	9.26	8.323333333
19	1767	New York	8.52	8.22	1767	9.19	8.29	1767	8.983333333	8.231666667
20	1768	New York	8.67	6.78	1768	9.09	7.80	1768	8.836666667	7.926666667
21	1769	New York	9.1	7.69	1769	8.76	7.56	1769	9.148333333	7.958333333

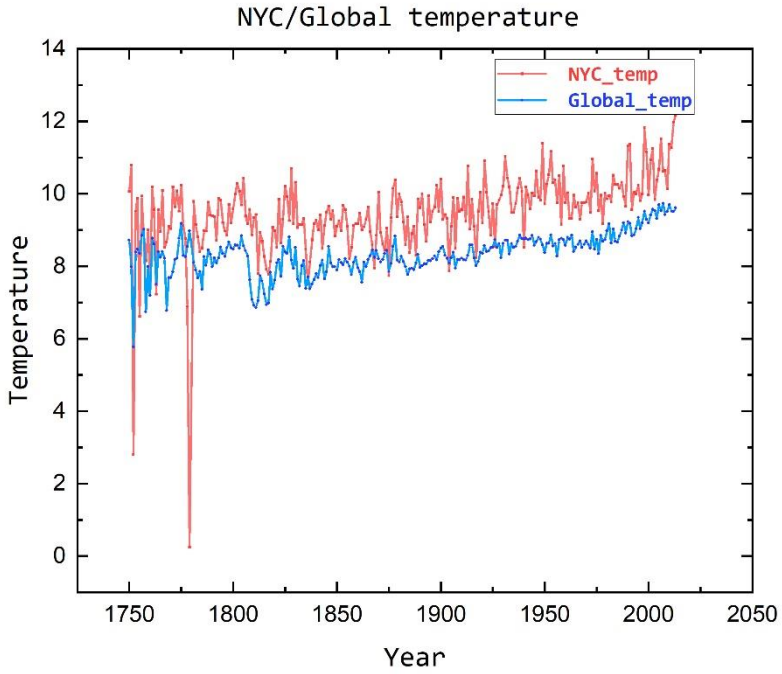
NYC vs Global_Null_Removed

5. Plot line charts using the processed data.

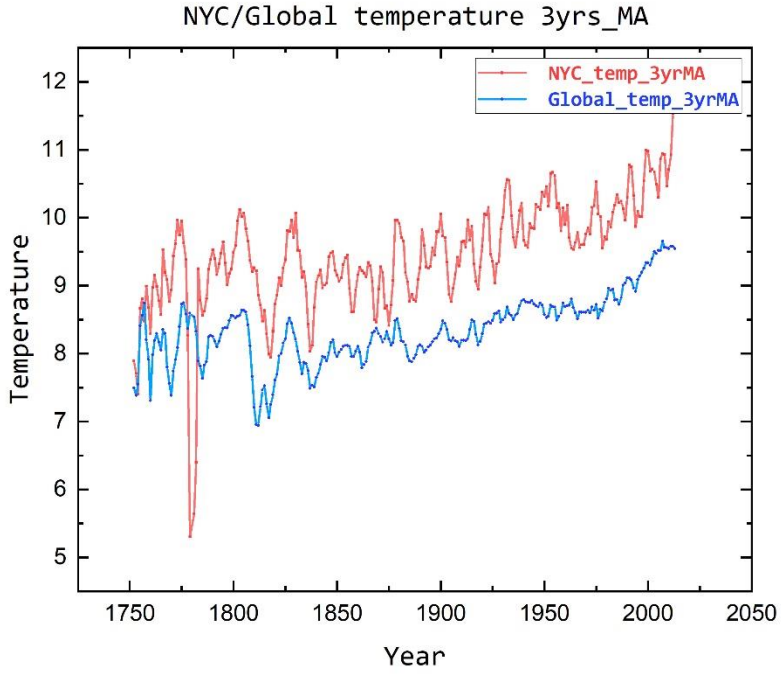
I use the software Origin 2020 to plot line charts. The key reason I choose Origin is that Origin is highly functional in scientific graphing and provides many customization options. For comparison, I plotted 3 plots: NYC vs Global temperature raw data, NYC vs Global temperature 3yrs_MA, NYC vs Global temperature 6yrs_MA.

Line Charts:

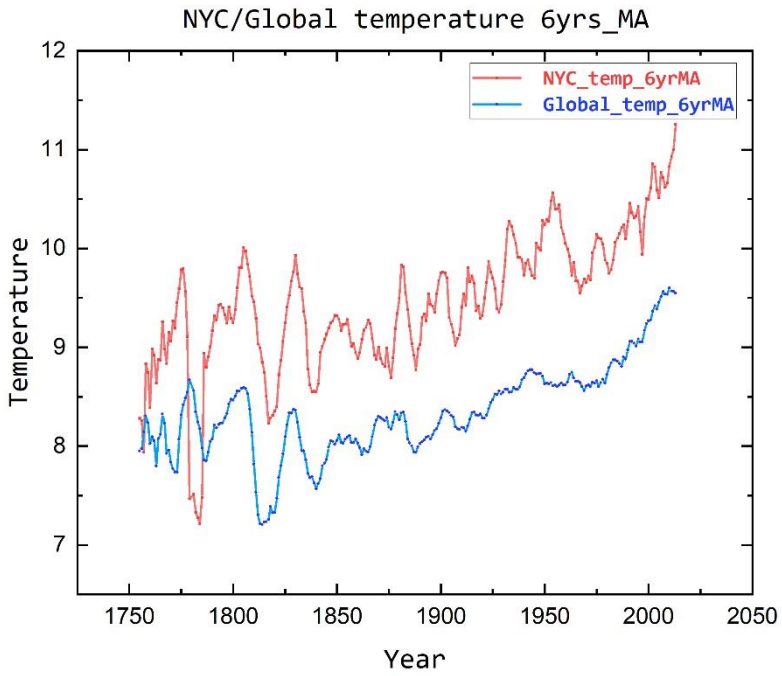
1. NYC vs Global yearly temperature raw data



2. NYC vs Global 3yrs MA



3. NYC vs Global 6yrs MA



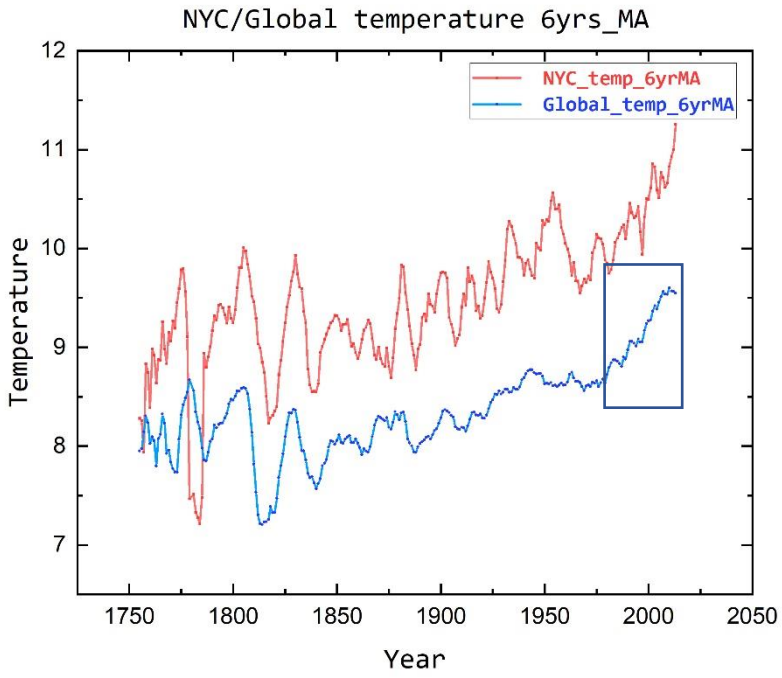
Observations:

From these line charts, the smoothing effect of the moving average plots is obvious. Though some yearly details are omitted, the overall trend of how the temperature changes through these years is presented in a better way.

1. NYC is **hotter** on average compared to the global average except for a short period between 1778 and 1782. It is interesting to investigate what really happened during 1778-1782. After searching it online, I found that many articles online have mentioned that the winter of 1779-1780 has been called the harshest in the 18th century[1-3]. Over the course of the winter, New Jersey had twenty-six snowstorms and six among them were blizzards[2]. Combined with all the later temperature records of NYC, the winter of 1779-1780 could be called the harshest in recent three centuries.

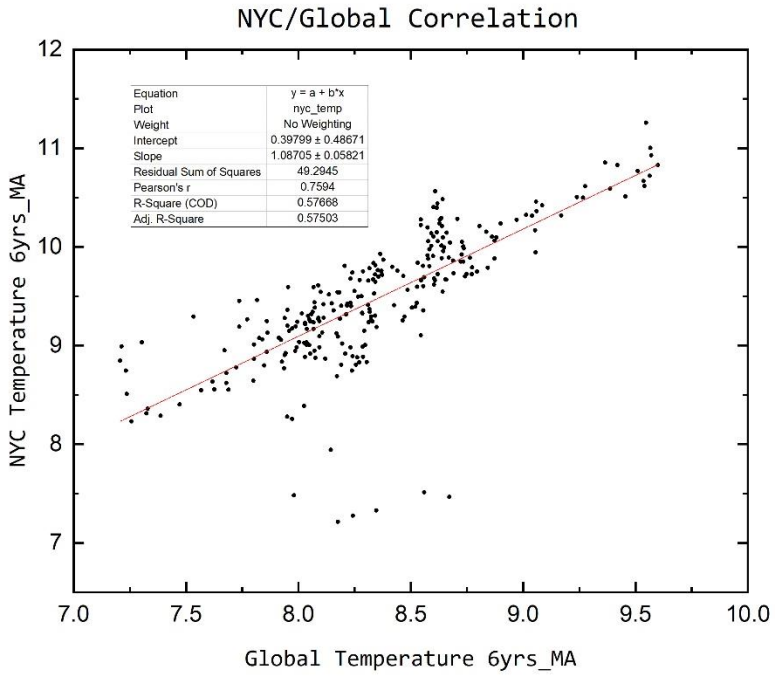
2. From all three line charts, it can be seen that the temperature fluctuation in NYC is **bigger** than that of the global average. One of the reasons I can think of is the extensive human activities and large population density in NYC.

3. According to the line charts, the world is getting **hotter** over the last few hundred years. This trend is especially obvious over the last 100 years. I also notice that the temperature increase seems to accelerate in the recent 50 years by looking at the slope change of global 6yrs_MA plot.



4. Both NYC and global plots show a very sharp temperature drop in the from late 18th century to the first twenty years in 19th century. This period is known as the Dalton Minimum[4]. It is intriguing to see how the weather might have an impact on the history. Napoleon was defeated in Russia in 1812 because of a bitterly cold winter and widespread famines spread across Europe[4].

5. I also used Origin to get a plot showing the correlation between the 6yrs_MA data of NYC and globe. The plot is shown in the next page. According to the plot, there is a certain positive linear correlation between NYC temperature data and global temperature data. The Pearson's correlation coefficient is 0.7594.



Reference:

[1] https://www.encyclopedia.com/history/encyclopedias-almanacs-transcripts-and-maps/winter-1779-1780

[2] https://weatherworksinc.com/news/winter-1779-80-revolution

[3] https://emergingrevolutionarywar.org/2016/01/23/the-hard-winter-of-1779-1780/

[4] https://www.sciencedirect.com/topics/earth-and-planetary-sciences/medieval-warm-period