

Explore Weather Trends

YI WANG

Data Preparation

- Data exploration step, I went through those 3 datasets to check the data structure. Since I'm living at a city in NJ, I choose the temperature of New York as my local city.
- I first check the city names in the data set city_list. From which I've found that the name of NYC in the data is 'New York'. With such information, I use following code to extract and merge the data. The general idea is that I need to create a master data set including Year, Average City Temperature, and Average Global Temperature. As the information are stored in different datasets, I also need to do an inner join to combine them.

Figure 1: Data Processing Code

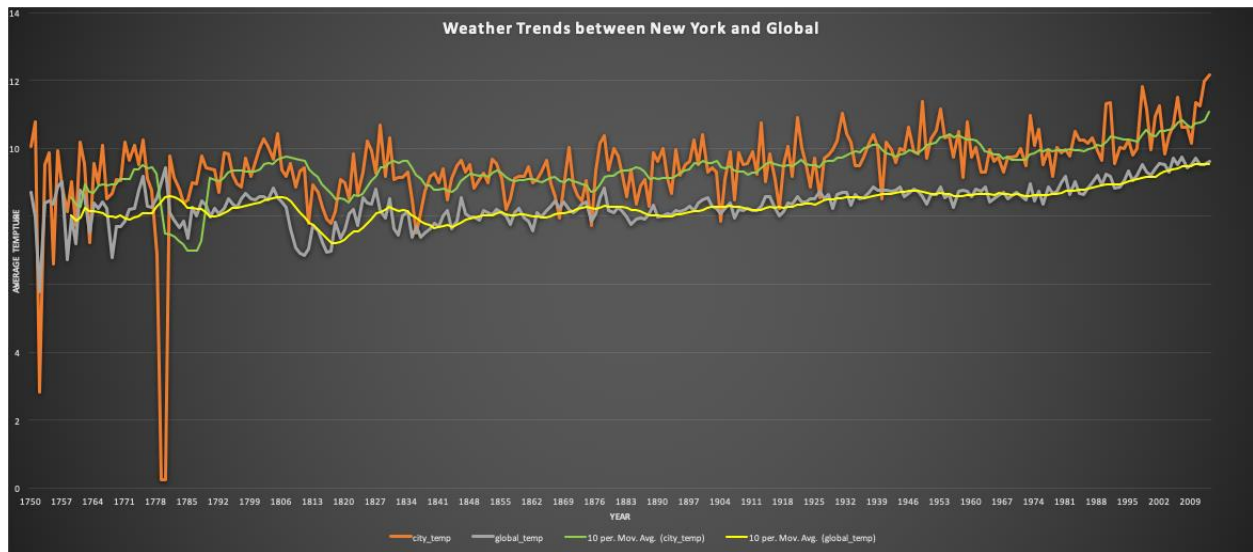
```
1 select city_data.year, city_data.avg_temp as  
   city_temp, global_data.avg_temp as global_temp  
2 from city_data  
3 inner join global_data  
4 on city_data.year = global_data.year  
5 where city = 'New York';
```

- Exported to CSV, I found there was a missing value of city_temp in year 1780. After some further research, I've found that this is actually a famous period of "hard winter"¹ in history. During which time, the temperature is extremely low. Thus, instead of taking an average of 1779 and 1781, I just used the temperature of 1779 to fill this missing value.
- I'm using 10 years as the window to calculate moving average. The screen short below showed a part of the result of the moving average, and I also marked the moving average in the line chart after.

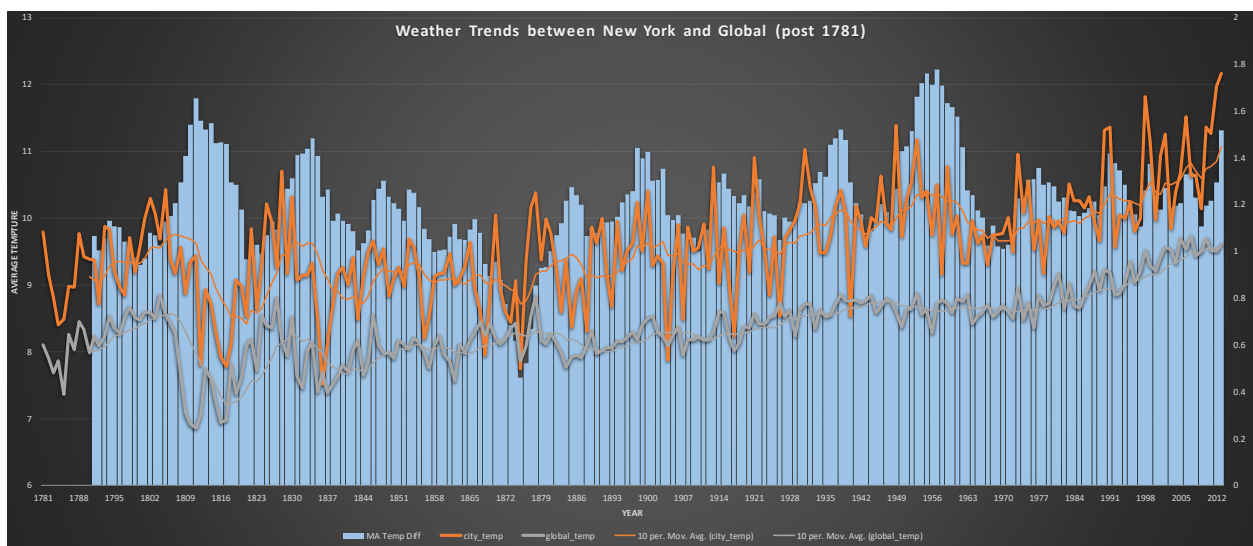
	year	city_temp	global_temp	MA_CityTemp	MA_GlobalTemp	correlation
2	1750	10.07	8.72			44.403%
3	1751	10.79	7.98			
4	1752	2.81	5.78			
5	1753	9.52	8.39			
6	1754	9.88	8.47			
7	1755	6.61	8.36			
8	1756	9.94	8.85			
9	1757	8.89	9.02			
10	1758	8.15	6.74			
11	1759	9.01	7.99	8.567	8.03	
12	1760	7.73	7.19	8.333	7.877	
13	1761	10.18	8.77	8.272	7.956	
14	1762	9.55	8.61	8.946	8.239	
15	1763	7.23	7.5	8.717	8.15	
16	1764	9.55	8.4	8.684	8.143	

¹ <https://emergingrevolutionarywar.org/2016/01/23/the-hard-winter-of-1779-1780/>

Line Chart



After removing the outlier of 1779~1780, I could actually zoom in to better visualize the correlation between New York and Global temperature. In addition, I've also added the moving average difference as bar chart.



Observations

- My city is hotter on average compare to the global average. Overall, the temperature difference is around 1. The temperature difference has been consistent from 1785 to 1830, and the years after 1940. The trough around 1872 suggests that the temperature of the New York is decreasing while the global temperature stay stable, while the peak around 1956 suggests the opposite story.
- The world is getting hotter because the overall trend of the average temperature is a little bit upwards, and this trend has been consistent since 1834.

- As described earlier, the local temperature during year 1779 to 1780 was much lower than the global, I found that the winter of that year was really bad, over the course of the winter New Jersey had 26 snowstorms and 6 of those were blizzards. And that winter was called “hard winter”.
- The correlation between local and global temperature is around 44.4%, which means there is a close relationship between them.