

# Exploring Weather Trends

Kaiwen Yao

Dec.8th, 2020

## Table of Contents

<b>1. Outline .....</b>	<b>- 1 -</b>
1.1.What tools did you use for each step?.....	- 1 -
1.2. How did you calculate the moving average?.....	- 1 -
1.3. How to calculate correlation coefficient? .....	- 2 -
1.4. What were your key considerations when deciding how to visualize the trends? .....	- 2 -
<b>2. Line Chart .....</b>	<b>- 3 -</b>
<b>3. Four Observations.....</b>	<b>- 3 -</b>

# 1. Outline

## 1.1. What tools did you use for each step?

First of all, I wrote SQL queries to extract data from the temperatures.

Queries 1 : Select \* from city\_data where country = 'San Francisco';

Queries 2 : Select \* from city\_data where country = 'New York';

Queries 3 : Select \* from city\_data where country = 'Tianjin';

Queries 4 : Select \* from global\_data;

Then I combined these CSV by using Python. Here's the code.

```
import pandas as pd
```

```
SFO = pd.read_csv("./SFO.csv",parse_dates=True)
```

```
NYC = pd.read_csv("./NYC.csv",parse_dates=True)
```

```
Tianjin = pd.read_csv("./Tianjin.csv",parse_dates=True)
```

```
city_list = pd.concat([SFO, NYC, Tianjin], axis=0)
```

Thirdly, graphs and numerical values were determined using EXCEL. Here's the charts.

Notice: There're some missing value in the database. If the temperature data are available for the years before and after, I replace the missing value with their average value. If not, I delete it.

For NYC, I deleted Year 1746, 1747, 1748, 1749s' data and replaced Year 1780's temperature by the average value of Year 1779 and Year 1781. For Tianjin, I replaced Year 1838's temperature by the average value of Year 1837 and Year 1839.

## 1.2. How did you calculate the moving average?

A	B	C	D
1	Year	Avg temp	SFO Moving avg temp
2	1849	14.12	
3	1850	13.8	
4	1851	14.39	
5	1852	13.81	
6	1853	14.4	
7	1854	13.98	
8	1855	14.2	
9	1856	14.1	
10	1857	14.78	
11	1858	14.19	14.177
12	1859	13.71	14.136
13	1860	13.81	14.137
14	1861	14.88	14.186
15	1862	14.43	14.248

I take San Francisco's temperature for example:

10 Years Moving Average

Temperature (Year 1858) = Average (B2:B11)

I continue this process all the way to the end of the dataset. Copy the cell and paste it all the way down to the end of the data (through row 166).

### 1.3. How to calculate correlation coefficient?

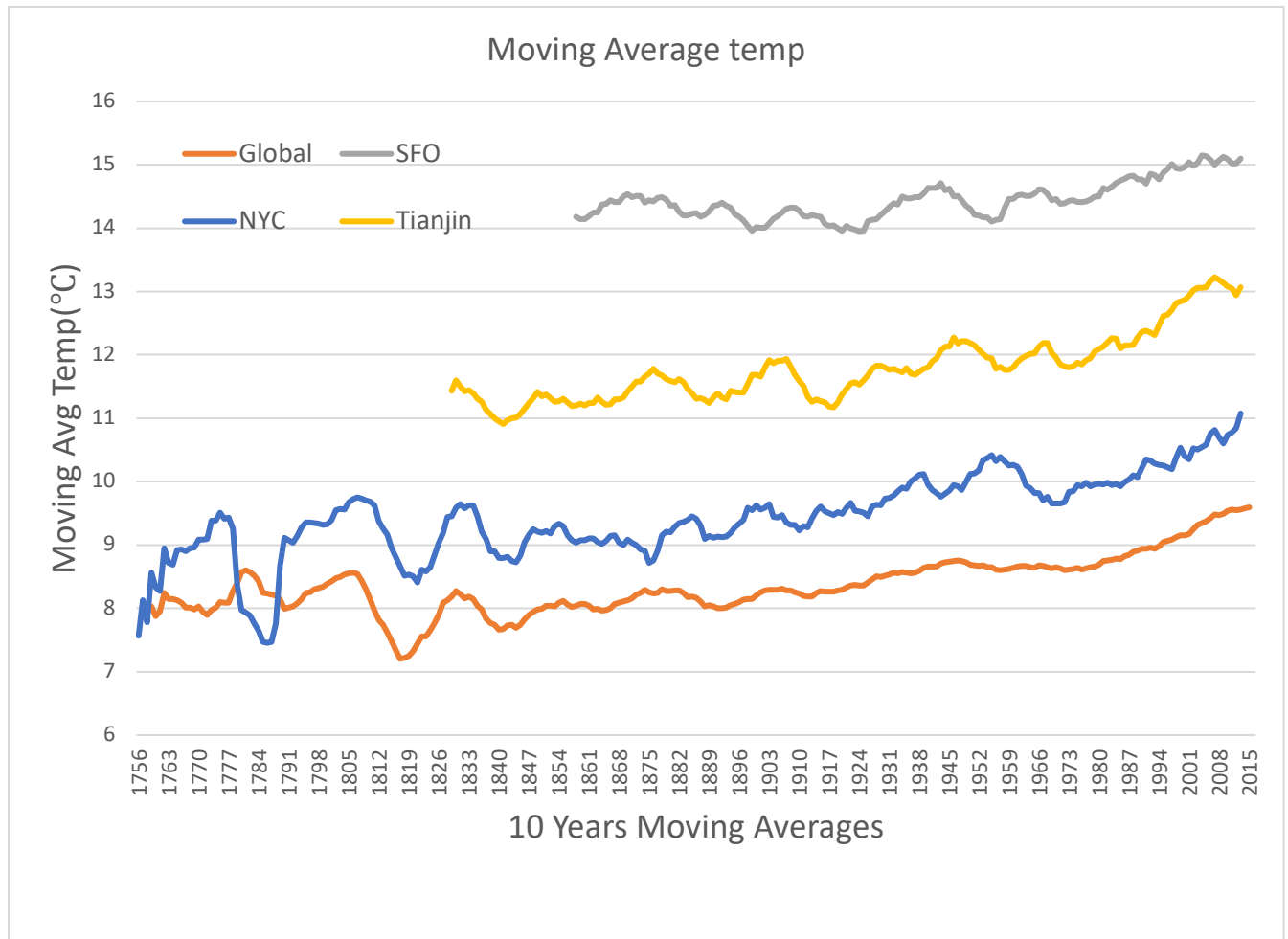
I used Excel function CORREL to calculate correlation coefficient between each city and Global. According to the following table, correlation coefficient between Tianjin and Global is the highest, SFO rank second, NYC is the lowest.

correlation coefficient	SFO	NYC	Tianjin
Global	0.81601817	0.75246068	0.96375541

### 1.4. What were your key considerations when deciding how to visualize the trends?

- Purpose:  
Analyze local and global temperature data and compare the temperature trends where you live to overall global temperature trends.
- Chart type:  
Since we need to compare multiple cities' temperature with global temperature, Line graph is the most appropriate visualization for the data. We can easily know how temperature of each city change over time.
- Color:  
I changed colors for the different data sets. [local and global]

## 2. Line Chart



## 3. Four Observations

- SFO & Global
  - Similarities:
    - Over time, although both types of temperature change up and down, they are basically show an upward trend, which means the world is getting hotter.
    - According to the chart, we can easily know that the upwards trend has been consistent over the last hundred years and probably will keep increase in the future.
  - Difference:
    - The San Francisco city's temperature has been consistently higher than the global average over time.
    - From 1859 to 2013, San Francisco's temperature has changed 1.179°C, which is lower than global temperature changed. Global temperature has changed 1.619°C.
- NYC & Global

- Similarities:
  - After 1789, Over time, although both types of temperature change up and down, they are basically show an upward trend, which means the world is getting hotter.
  - According to the chart, we can easily know that the upwards trend has been consistent over the last hundred years and probably will keep increase in the future.
- Difference:
  - The New York city's temperature has been consistently higher than the global average after 1789.
  - From 1756 to 2013, New York's temperature has changed 3.601°C, which is higher than global temperature changed. Global temperature has changed 2.391°C.
- Tianjin & Global
  - Similarities:
    - Over time, although both types of temperature change up and down, they are basically show an upward trend, which means the world is getting hotter.
    - According to the chart, we can easily know that the upwards trend has been consistent over the last hundred years and probably will keep increase in the future.
  - Difference:
    - The Tianjin city's temperature has been consistently higher than the global average over time.
    - From 1829 to 2013, San Francisco's temperature has changed 2.273°C, which is higher than global temperature changed. Global temperature has changed 1.923°C.