The Boston city and global weather trends during 1750 to 2013

In this project, I choose Boston (US) city, which is my nearest big city to analyze and compare with global weather trends. The outline of the project steps is as follows:

Data Extraction: SQL is used to extract the data from database. The data results are downloaded in csv format and then copied to Excel for analysis.

Trend Analysis: Line chart created by Excel is the visualization tool to analyze. 5 and 10 years (a decade) moving average are calculated and visualized to smooth the fluctuations over time.

The Association Analysis: Scatter plot and correlation coefficient are used to study the relationship of the time series.

Forecast: Linear regression and exponential smoothing are used to forecast the temperature of Year 2014.

Data Extraction

Find the available data: I tried to find Worcester city, which I lived in using SQL codes:

```
select * from city_list where city = 'Worcester' limit 5
```

The output shows 0 results, so I loaded all cities available in the U.S. and choose Boston as the target city. Then the temperature data are extracted from both *city_data* and *global_data*. The SQL codes are:

```
select * from city_list where country = 'United States'
select * from city_data where city = 'Boston'
select * from global_data
```

The results are downloaded as csv files, and data are copied to excel sheet for study.

The original data shows Boston has temperature records from year 1743 to 2013, with year 1780 data missing, and the global data is from year 1750 to 2013. In the following study, the data before 1750 is ignored. Vlookup function is used to match the year for Boston and global temperature.

The data can also be extracted using SQL codes:

```
select c.year as Year, c.avg_temp as boston_data, g.avg_temp as global_data
from city_data c left join global_data g on (c.year = g.year) where c.city =
'Boston'
```

or

```
select c.year as Year, c.avg_temp as boston_data, g.avg_temp as global_data
from city_data c, global_data g where (c.year = g.year) and c.city = 'Boston'
```

Moving Average Calculation and Visualization

5-year and 10-year moving average are both calculated using AVERAGE function. The screenshot Fig.1 demonstrates the calculation process. Fig. 2 and Fig.3 display the line charts respectively. Clearly, 10-year moving average has more clear trends compared to 5-year moving average, so it is chosen for further study.

Fig.1 screenshot of five-year moving average calculation

K13 =AVERAGE(D9:D13)

	A	D	E	H	I	K	L
1	year	Boston	Global	Boston_10	Global_10	Boston_5	Global_5
9	1750	7.88	8.72				
10	1751	8.6	7.98				
11	1752	0.36	5.78				
12	1753	7.35	8.39				
13	1754	7.75	8.47			6.39	7.87
14	1755	4.28	8.36			5.67	7.80
15	1756	7.76	8.85			5.50	7.97
16	1757	6.65	9.02			6.76	8.62
17	1758	6.09	6.74			6.51	8.29
18	1759	6.8	7.99	6.35	8.03	6.32	8.19
19	1760	5.53	7.19	6.12	7.95	6.57	7.96
20	1761	8.05	8.77	6.06	7.87	6.62	7.94

Fig.2 Temperature Line Chart of Boston and Global

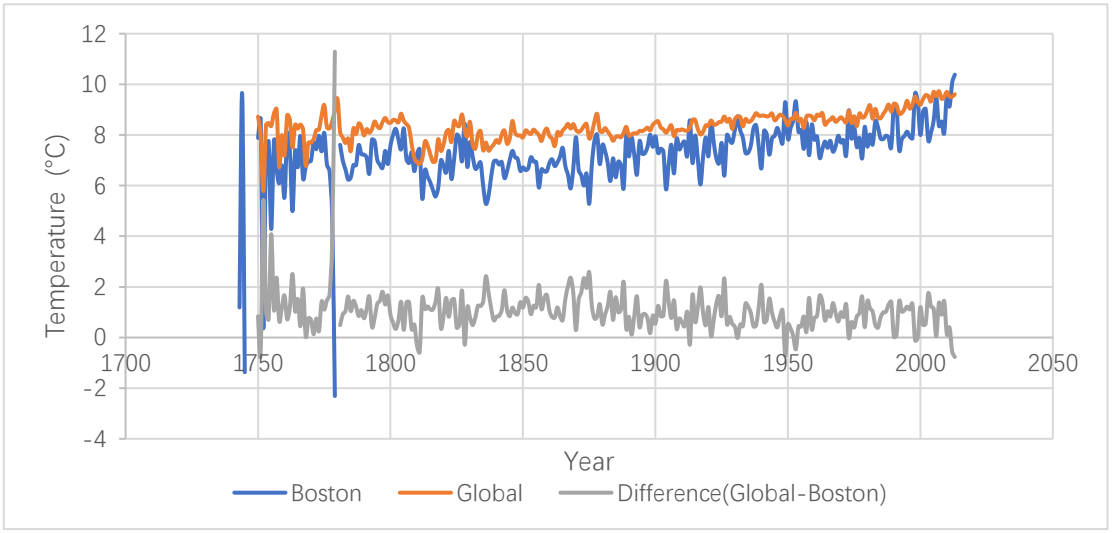

Fig.3 Five-Year Moving Average Line Chart of Temperature

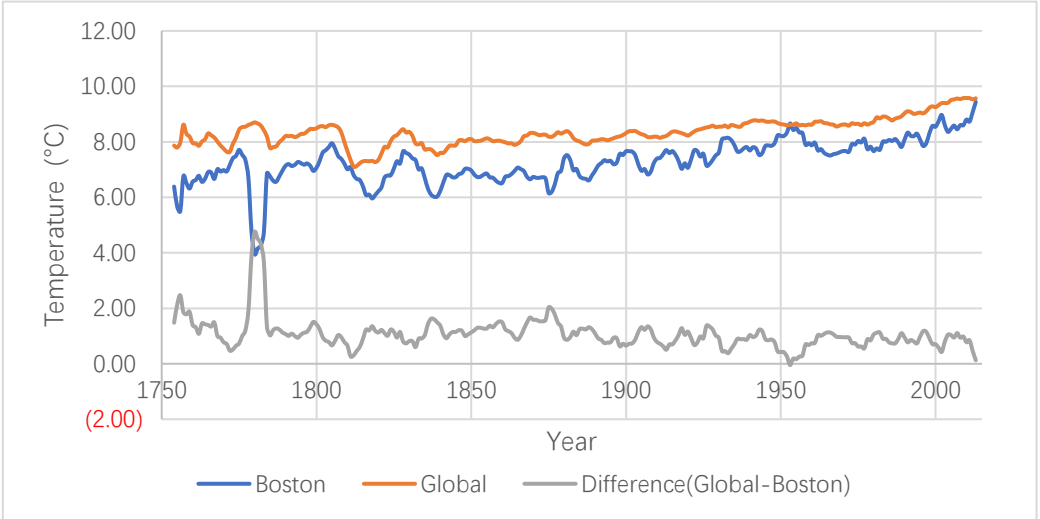

Fig.4 Ten-Year Moving Average Line Chart of Temperature

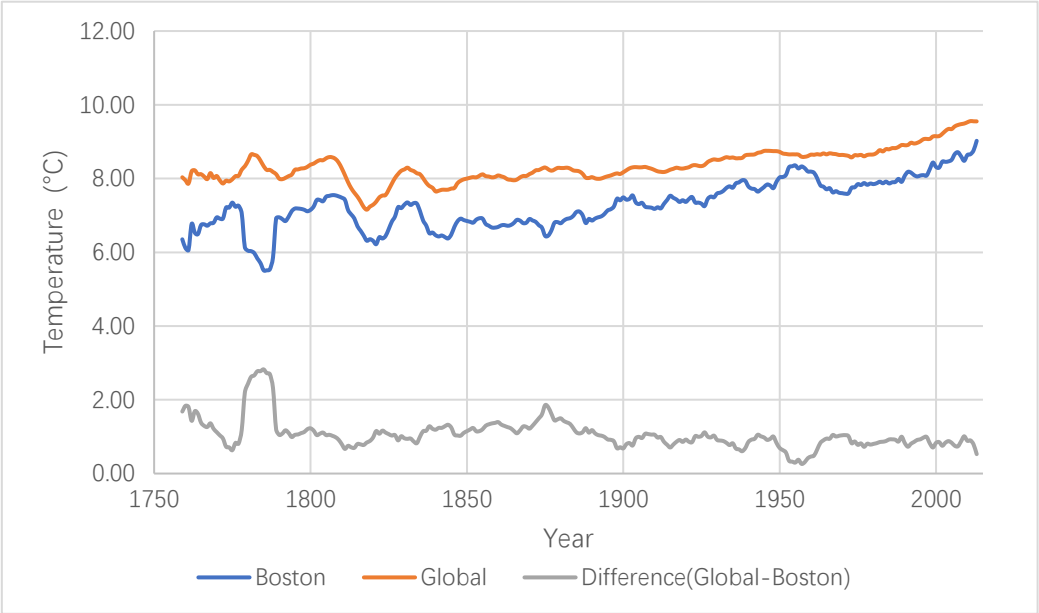

Table 1 Mean Temperature

Year	Boston (°C)	Global (°C)
1750-1850	6.778	8.032
1750-1900	6.834	8.068

Observations

1. The temperature of the Boston city is much more fluctuating than global. It is obvious that global average temperature consists weather information of many cities, thus the data is “smoother” than a single city.

2. Before around year 1900, time series shows the weather for both Boston and global are horizonal patterns, which mean the data fluctuate around a constant mean, which are list in Table 1. After 1900, although the Fig.2-4 show up and down movement over the hundreds of years, we can easily observe the time series for Boston and global both demonstrate trend patterns.

3. Apparently, the world is getting warmer over the several hundred of years. The trend is accelerated from last century. The trend of the Boston city is more obvious than the global data.

4. Most of time in history, Boston is cooler compared to the global average. The grey lines in Fig. 2-4 illustrate the difference is consistent over time after year 1800.

Association between the Boston city and the global temperature

Fig.5 Scatter Chart of temperature pairs

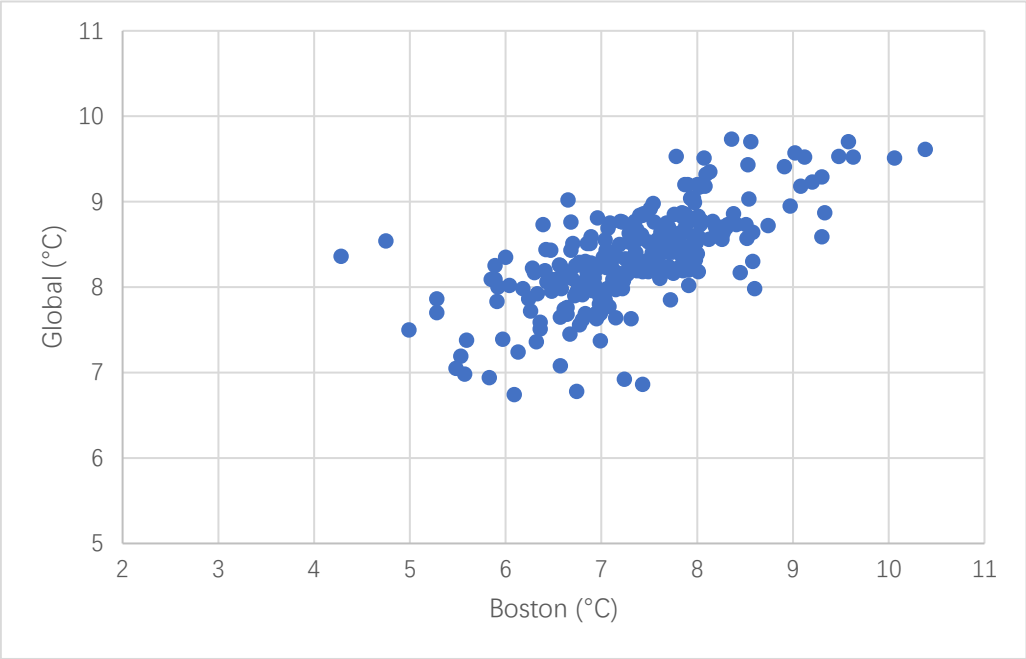

Table 2 Correlation coefficient of the original data

	Boston	Global
Boston	1	
Global	**0.5779958**	1

Table 3 Correlation coefficient of the 10-year moving average data

	Boston_Avg	Global_Avg
Boston_Avg	1	
Global_Avg	**0.796413175**	1

The scatter plot Fig.3 displays Boston and global temperature in pairs. Apparently, there is a positive correlation between them. The calculation of correlation coefficient also indicates that when the global weather become warmer over time, Boston has the similar trend.

Forecast

Linear regression is first applied to fit and forecast Boston and global temperature. Temperature from 1900 to 2013 is select as the input data because of the observation 2. Excel Data Analysis – Regression tool was used, and the concise results are list in Table 4. FORECAST.LINEAR function can also be applied to get the same forecast results. Since the Adjusted r square is not quite large, especially for Boston data, the forecast results may not be accurate.

Table 4. Linear Regression Forecast

	Boston	Global
Forecast for 2014	8.649	9.343

Adjusted r square	0.2896	0.7036

Exponential smoothing is also applied to forecast with FORECAST.ETS function. The results in Table 5 show that exponential forecast predicts higher temperature than linear forecast. Intuitively, they will be more reliable results since the global warming trend is obvious according to the historical data.

Table 5. Exponential Forecast

	Boston	Global
Forecast for 2014	9.4982	9.620