Project 1: Explore Weather Trends (Washington, U.S)

OUTLINE:

1. <u>What tools did you use for each step? (Python, SQL, Excel, etc)</u>
   Firstly, I wrote SQL query to extract data to CSV file from database. Since Washington DC is the nearest city from my city, I selected its temperature data as my exploratory data. Second, I checked missing value and delete it since it will affect the result. Then, I calculated the moving average in EXCEL and created the line chart of it. Finally, I had my result and gave the report.

2. <u>How did you calculate the moving average?</u>
   In EXCEL, using '=AVERAGE(xx:xx)' to calculate the moving average where 'xx' is the cell you want to analyze.

3. <u>What were your key considerations when deciding how to visualize the trends?</u>
   - Change horizontal axis interval unit
   - Legend name
   - Plot name
   - Select data

STEPS:

1. <u>Extract data</u>
   Write a SQL query to extract the city level data and the global data. Export to CSV.

```
1    SELECT *
2    FROM city_list
3    WHERE country = 'United States';
```

```sql
1    SELECT d.year, d.avg_temp as city_avg_tmp,
     g.avg_temp as global_avg_tmp
2    FROM city_data as d
3    JOIN global_data as g
4    ON d.year = g.year
5    WHERE d.city = 'Washington'
6    ORDER BY d.year;
```

2. <u>LINE CHART (Moving Average)</u>

At the beginning of analyze data, I checked about missing values and omit it since it will affect the result. Here is the code in R:

1. Read the data

```r
library(readxl)
data = read_excel("TEMP_results.xls")
```
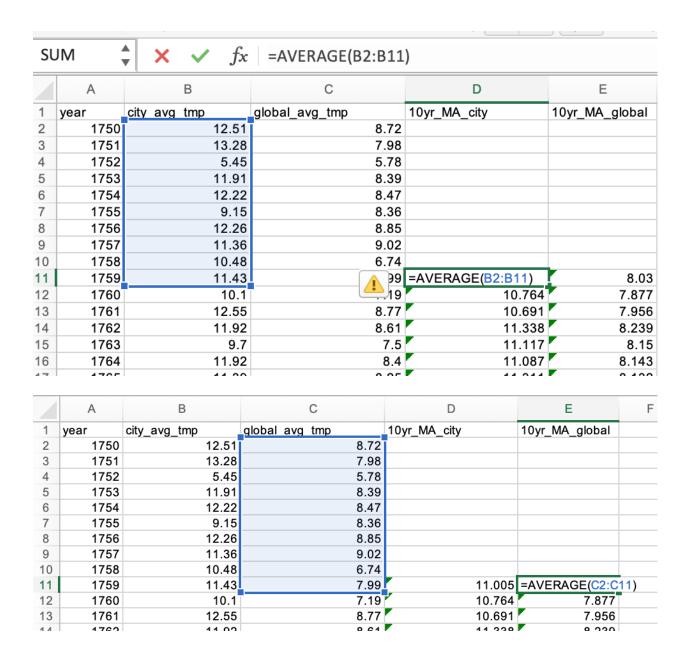
2. Check missing data

```{r}
is.na(data)
```

3. Omit missing data

```{r}
data = na.omit(data)
```

4. Write the final data into excel
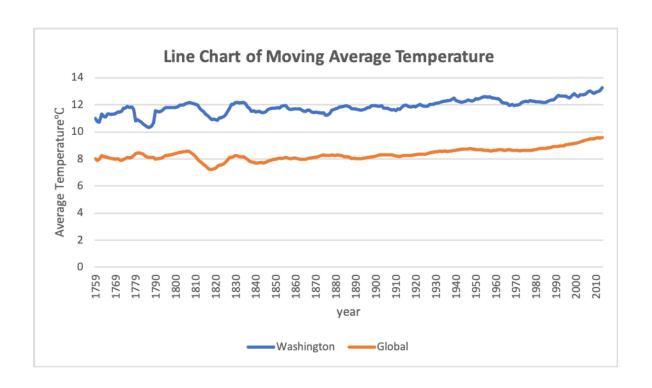
```{r}
library(WriteXLS)
WriteXLS(data, "temp_without_NA.xls")
```

After that, I did 10 year Moving Average of both city level data and global level data as following. Firstly, I used '=AVERAGE(B2:B11)' to calculate the first 10 years moving average value and drag down to apply to whole city level data. Then I did the same thing on global level data by using '=AVERAGE(C2:C11)'.

	A	B	C	D	E
1	year	city_avg_tmp	global_avg_tmp	10yr_MA_city	10yr_MA_global
2	1750	12.51	8.72		
3	1751	13.28	7.98		
4	1752	5.45	5.78		
5	1753	11.91	8.39		
6	1754	12.22	8.47		
7	1755	9.15	8.36		
8	1756	12.26	8.85		
9	1757	11.36	9.02		
10	1758	10.48	6.74		
11	1759	11.43	⚠ 99	=AVERAGE(B2:B11)	8.03
12	1760	10.1	.19	10.764	7.877
13	1761	12.55	8.77	10.691	7.956
14	1762	11.92	8.61	11.338	8.239
15	1763	9.7	7.5	11.117	8.15
16	1764	11.92	8.4	11.087	8.143

	A	B	C	D	E	F
1	year	city_avg_tmp	global_avg_tmp	10yr_MA_city	10yr_MA_global	
2	1750	12.51	8.72			
3	1751	13.28	7.98			
4	1752	5.45	5.78			
5	1753	11.91	8.39			
6	1754	12.22	8.47			
7	1755	9.15	8.36			
8	1756	12.26	8.85			
9	1757	11.36	9.02			
10	1758	10.48	6.74			
11	1759	11.43	7.99		11.005	=AVERAGE(C2:C11)
12	1760	10.1	7.19		10.764	7.877
13	1761	12.55	8.77		10.691	7.956
14	1762	11.92	8.61		11.338	8.239

After I got the full moving average result. I created the line chart for these two column data.

Line Chart of Moving Average Temperature

OBSERVATIONS

1. After applying 10 years Moving Average Method, the global average temperature is around 8°C (7°C − 9.5°C). And the city level average temperature is around 12°C (10.2°C − 13.1°C)
2. The overall trend for city level data and global level data are both increasing slowly.
3. Washington is hotter compared to the global average from the above line chart. And the difference had been nearly consistent over time. However, we can see that the difference suddenly reduced between the year 1779 and 1789, which was caused by the sudden decline of city level data and rise of global level data.
4. It is shown that there is a drop between 1812 and 1825 on both city level data and global level data. After that, there is a rise between 1825 and 1839 on both data. Therefore, it seems like that the changes maintain a synchronized increase in Washington's temperatures over time compare to the changes in the global average.