Project 1 - Exploring Weather Trends

2nd March 2020, Udacity Data Analyst Nanodegree

1. Data Extraction

The SQL code used to extract data list dataset is shown in the box below, this is to extract all the variables in the city_list database.

```
Select *
from city_list;
```

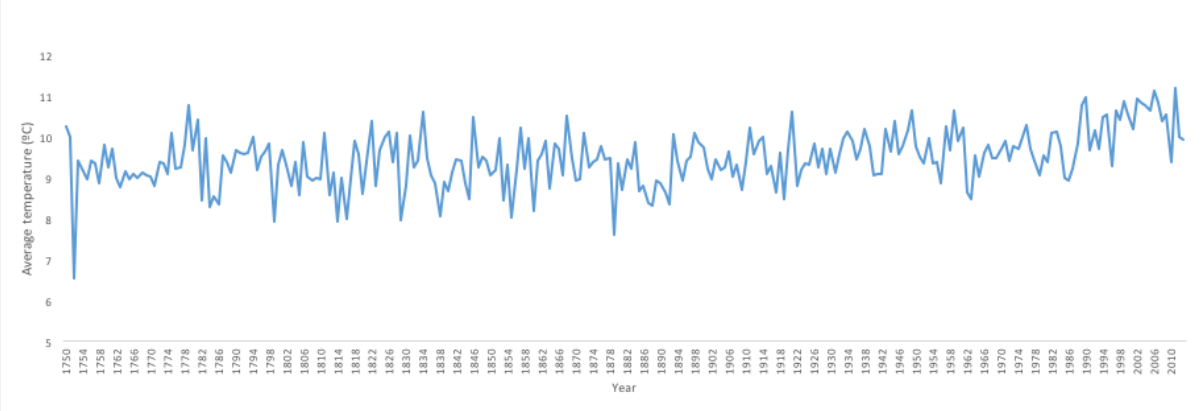
After having looked at the city list data, I decided to choose city London in United Kingdom to analyze. The below SQL code is used to extract the UK London city temperature data.

```
Select *
from city_data
WHERE city='London' AND country='United Kingdom';
```

2. Data manipulation

I have used Excel to do all the analysis. (Note: as the global temperature data starts from year 1750, all my analysis is done for period 1750 to year 2013.)

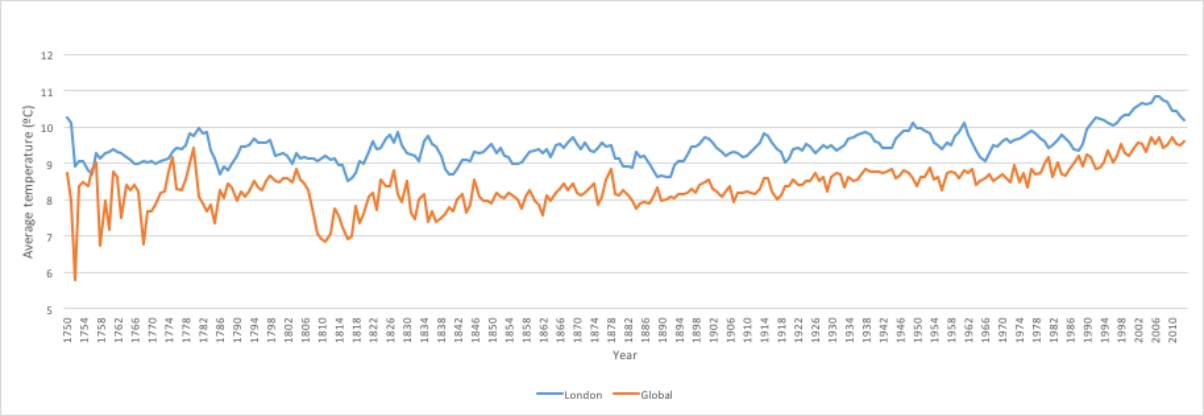
Chart 1: Average weather temperature in London UK from year 1750 to 2013



The chart above was created using the raw data and it shows a lot fluctuation over the years, so I have computed the 5-year moving average to smooth the data. As an example, the smoothed value for year 1990 is calculated as the average of temperatures 1986-1990. The global temperature was matched to London

temperature dataset using VLOOKUPs as both datasets contain the same variable 'year'. The chart below shows the global average temperature over the period as well as the smoothed London temperature over the same the period.

Chart 2: London vs. Global average temperature, Year 1750 -2013

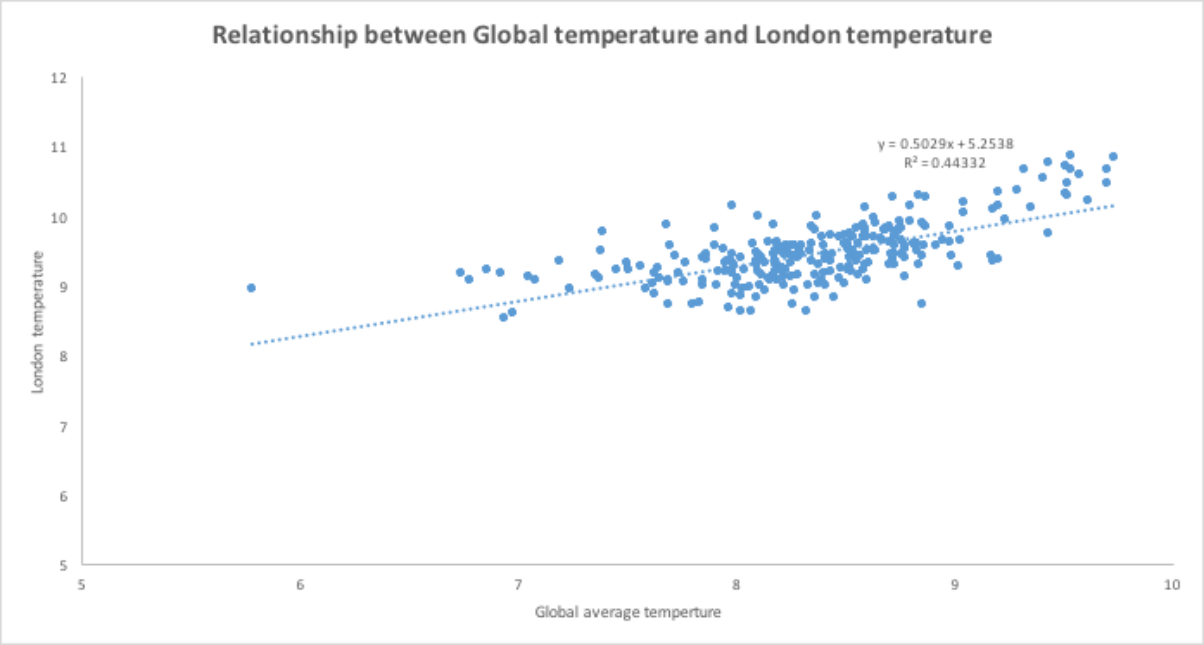


Some observations:

- It is noticed that trend for the temperature in London is very similar to the global average temperature trend, but has a slightly higher temperature in general compare to the global average.
- Both trends have fluctuations, but in general, they are in upward trend.
- The lowest average temperature recorded in the datasets appeared to be in year 1752, with 6.54 degrees recorded for London and 5.78 degrees for global average.
- After approximately 1886, the global temperature started increasing at a higher rate, this may be due to the global warming effect came to play. In particular, temperature in the past 50 years increased at a much significant rate.

3. Further analysis

To further analyze the relationship between the two datasets, I have computed the correlation between London city temperature and global temperature. The correlation coefficient is 0.67, and obtained using function CORREL in excel. This suggests that there is a fairly strong positive correlation between the two datasets.



In order to determine the relationship between the two datasets, I have fitted a liner model with equation

$$y = 0.5029x + 5.2538$$

Where
y = London temperature
x = global average temperature

This equation can be used to predict the temperature based on a given global temperature.