# Project: Weather Trends

## Extracting Data

I initially investigated the data provided and determined what information I would need to explore how global and local temperatures have changed. Initially, using SQL I used the following query to determine if my local city (London) was in the list:

```
select *
from city_list
where city = 'London'
```

From this query, there were two results: London, Canada and London, United Kingdom. From this I realised I would need to specify London, United Kingdom in my main query. To extract the data using SQL, I used the following query:

```
SELECT city_data.year, city_data.city,
city_data.avg_temp as city_temp, global_data.avg_temp as global_temp
from city_data, global_data
where city_data.year = global_data.year
and city_data.avg_temp IS NOT NULL
and city_data.city='London'
and city_data.country= 'United Kingdom'
```

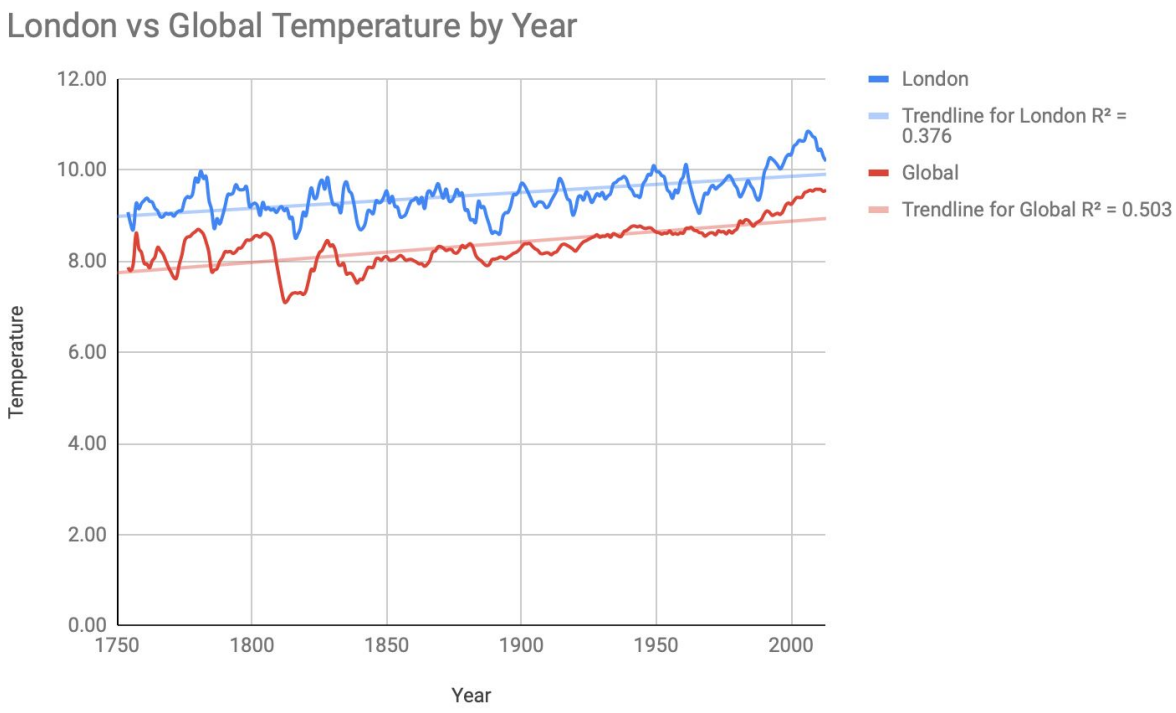
## Moving Averages

To smooth out the data and to avoid daily fluctuations, I used moving averages to observe the long term trend. In Google Sheets, I used a 5-year moving average. Starting at 1754, I took the average of the year and its prior 4 years:

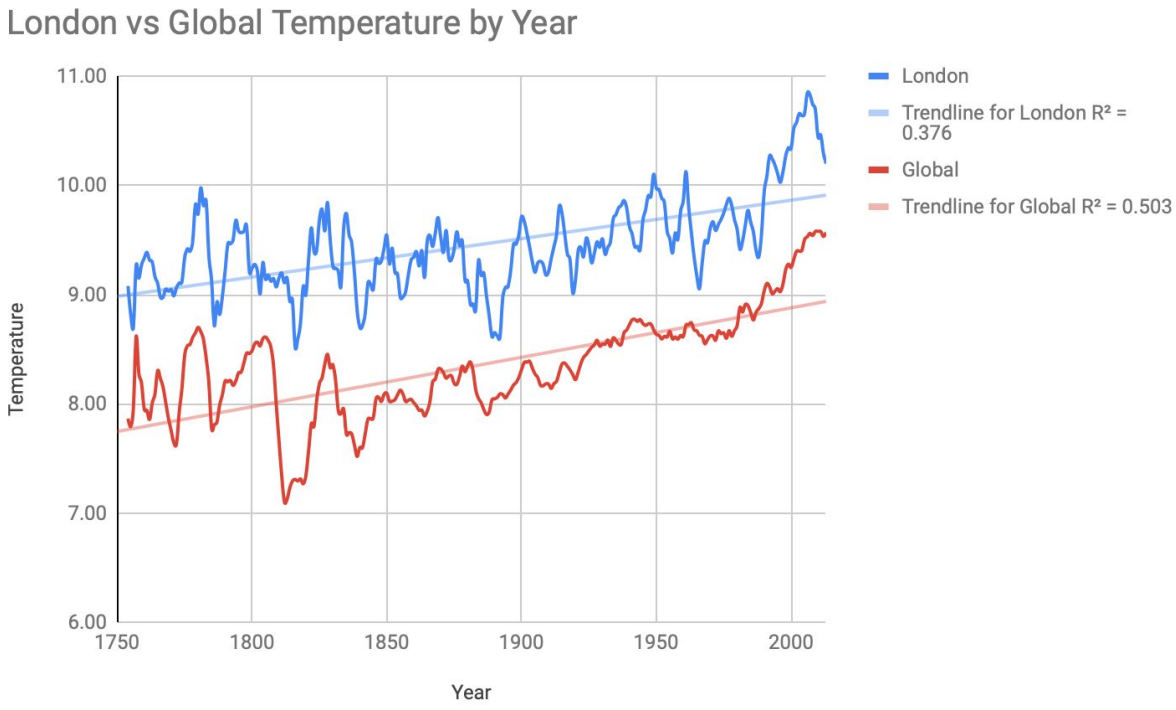
| year | city | city_temp | global_temp | London | G |
|---|---|---|---|---|---|
| 1750 | London | 10.25 | 8.72 | | |
| 1751 | London | 9.99 | 7.98 | | |
| 1752 | London | 6.54 | 5.78 | | |
| 1753 | London | 9.42 | 8.39 | | |
| 1754 | London | 9.2 | 8.47 | =AVERAGE(C2:C6) | |

I repeated the formula for the rest of the data and for global temperatures. I then used these Moving Averages to plot a Line Graph of temperature against time, in this case year. This was my initial graph:

Data Analyst Nanodegree: Alessia Tidman




I noticed that the line graph was condensed to the top which made it difficult to see the fluctuations, so I amended the vertical axis to show values of 6 to 12:



## Observations

With help from the trend line, we can see that both London and Global temperature on average has increased over the 250 year period. However, fluctuations have been inconsistent and irregular.

The shape of London is somewhat similar to the shape of Global temperature, i.e typically when the Global temperature drops/increases so does London. However, more recently the average temperature for London has been more erratic compared to average Global temperature - London average temperature rises and falls more drastically whereas Global temperature does this on a much smaller scale.

From the graph, we can also see that on average London is much hotter than the Global average. The largest Global Average is 9.58 compared to London which is 10.84.

As seen on the graph legend, we have $R^2$ values which is the percentage of variability around the mean that is explained by the model. If we take the root of these values, we obtain the Pearson Correlation Coefficient (r) which measures the correlation between two variables.

| | $R^2$ | r | Trendline Equation |
|---|---|---|---|
| London | 0.376 | ±0.61 | Temp = 3.52E-03*year + 2.83 |
| Global | 0.503 | ±0.70 | Temp = 4.52E-03*year -0.166 |

Ideally, a good model would explain 70% or more of the variability around the mean. For London, the $R^2$ is 37.6% whereas the Global $R^2$ is 50.3%. Neither would be considered a good model based on a 70% threshold, however the Global model explains for more variability compared to the London model.

Looking at r value, we can see that Global Temperature has a stronger positive correlation than London. We can see this slight difference on the graph itself.

Using the trendline equation, I can predict the temperature for 2020:

```
London : Temp = 3.52E-03*2020 + 2.83 = 9.94
Global: Temp = 4.52E-03*2020 -0.166 = 8.96
```

In 2020, we can expect the Temperature in London to be warmer than the Global average which is in line with my previous observation that London is warmer than the Global average.