Exploring Weather Trends

Purpose

The purpose of this project is to review temperature data and analyse and compare local temperature trends at where I live and the overall global temperature trends.

Workflow

1. Extract the data
2. Open up the CSV
3. Create a line chart
4. Make observations

Step 1: Extract the data

The following temperature data sets were obtained from the database using SQL:

- London, United Kingdom
- Global

First of all, the *city_list* table was used to verify the existence of temperature data for London, United Kingdom. I used the following SQL query:

```
SELECT *
FROM city_list
WHERE city = 'London'
```

The database outputted two results: 'London, Canada' and 'London, United Kingdom'. Therefore, temperature data set for London, United Kingdom exists in the database.

Second of all, the *city_data* table was used to obtain temperature data set for London, United Kingdom with the following SQL query:

```
SELECT *
FROM city_data
WHERE city = 'London' AND country = 'United Kingdom'
```

The resulted data set was then downloaded onto my computer in the form of a CSV file.

Finally, the *global_data* table was used to obtain the global temperature data set with the following SQL query:

```
SELECT *
FROM global_data
```

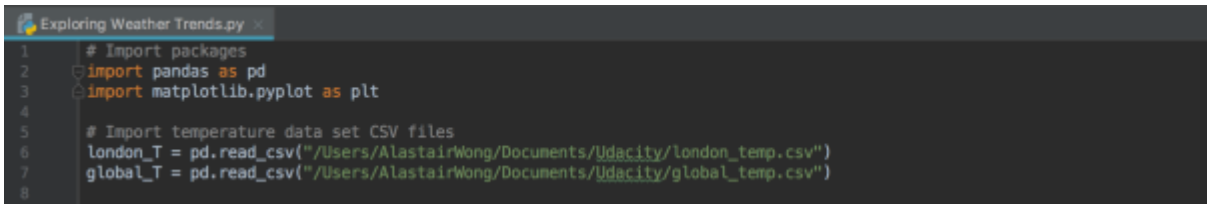
The resulted data set was then downloaded onto my computer in the form of a CSV file.

Step 2: Open up the CSV

I used Python to open up the temperature data set CSVs. Please note that the data sets were renamed as follows:

- London temperature data set: *london_temp.csv*
- Global temperature data set: *global_temp.csv*

I stored the London and global temperature data sets in variables *london_T* and *global_T*, respectively, with the pandas read_csv function. The Python syntax is shown in Figure 1.


Exploring Weather Trends.py

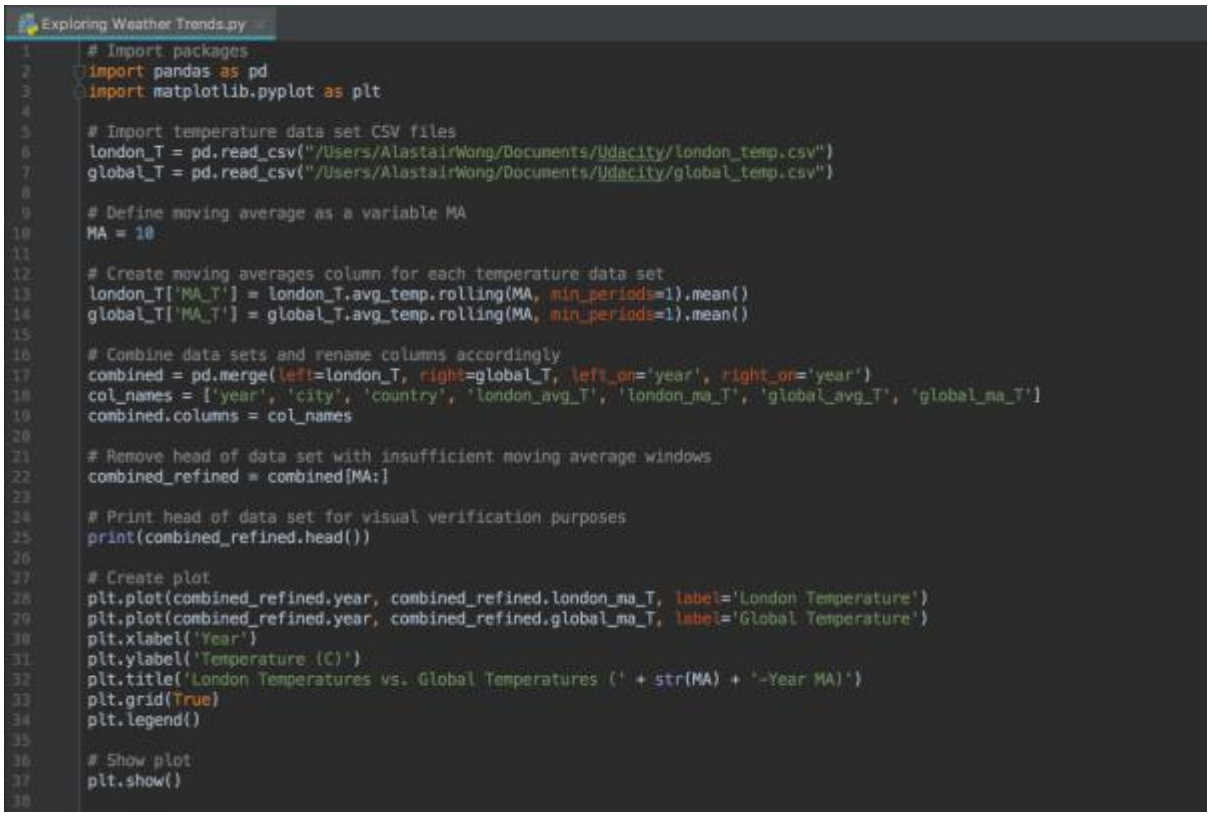
```
# Import packages
import pandas as pd
import matplotlib.pyplot as plt

# Import temperature data set CSV files
london_T = pd.read_csv("/Users/AlastairWong/Documents/Udacity/london_temp.csv")
global_T = pd.read_csv("/Users/AlastairWong/Documents/Udacity/global_temp.csv")
```

Figure 1: Opening up the CSV with pandas package

Step 3: Create a line chart

In order to create the line chart to compare the two temperature trends, I had to go through the following steps:

1. To add moving average column to each temperature data set
2. To plot the temperature trends onto a graph


Exploring Weather Trends.py

```
# Import packages
import pandas as pd
import matplotlib.pyplot as plt

# Import temperature data set CSV files
london_T = pd.read_csv("/Users/AlastairWong/Documents/Udacity/london_temp.csv")
global_T = pd.read_csv("/Users/AlastairWong/Documents/Udacity/global_temp.csv")

# Define moving average as a variable MA
MA = 10

# Create moving averages column for each temperature data set
london_T['MA_T'] = london_T.avg_temp.rolling(MA, min_periods=1).mean()
global_T['MA_T'] = global_T.avg_temp.rolling(MA, min_periods=1).mean()

# Combine data sets and rename columns accordingly
combined = pd.merge(left=london_T, right=global_T, left_on='year', right_on='year')
col_names = ['year', 'city', 'country', 'london_avg_T', 'london_ma_T', 'global_avg_T', 'global_ma_T']
combined.columns = col_names

# Remove head of data set with insufficient moving average windows
combined_refined = combined[MA:]

# Print head of data set for visual verification purposes
print(combined_refined.head())

# Create plot
plt.plot(combined_refined.year, combined_refined.london_ma_T, label='London Temperature')
plt.plot(combined_refined.year, combined_refined.global_ma_T, label='Global Temperature')
plt.xlabel('Year')
plt.ylabel('Temperature (C)')
plt.title('London Temperatures vs. Global Temperatures (' + str(MA) + '-Year MA)')
plt.grid(True)
plt.legend()

# Show plot
plt.show()
```

Figure 2: Creating a moving average temperature line chart with pandas data frame manipulation and matplotlib.pyplot

Key considerations when visualising the temperature trends include the following:

- The moving average window used which is most appropriate for analysis
- Proportion of the data used for the purpose of the analysis

At the beginning, moving average window (in years) was defined as a variable, *MA*, so that I would be able to experiment with different moving average windows. In Figure 2, *MA* was set to 10 years as it was found to be the lowest moving average window which had sufficiently cut out noises for my analysis. The reader is referred to the Appendix section to compare the differences between the resulting line charts generated using 1-year, 5-year, 10-year, and 25-year moving averages.

Then, moving averages column was created for each temperature data set using the *.rolling(MA, min_periods=1)* method along with *.mean()* at the tail. This would, for each row in each data set, compute moving average of current temperature data along with the past 9 data points (total of 10 data points). The resulting calculated moving average value would be stored in a new column *MA_T.*

The two data sets were then combined into one using the pandas *merge()* function. This would ensure that the same time frame from both data sets were used for analysis.

The head of the combined data set was removed as the first (MA – 1) rows of the data set would have been based on insufficient moving average window.

Using *matplotlib.pyplot* package, the following were done in order to plot moving average temperature data for London and global:

- *plt.plot* was used to plot the two temperature trends, with 'year' on the x-axis and 'temperature' on the y-axis. Appropriate labels were also given to each plot.
- X-axis and y-axis labels were added using *plt.xlabel()* and *plt.ylabel()* functions.
- Title was added using *plt.title()* function.
- Grid was added using *plt.grid()* function in order to increase the ease of reading the charts.
- Legend was added using *plt.legend()*

Finally, the chart was generated using *plt.show()* function. Figure 3 shows the resulting line chart.

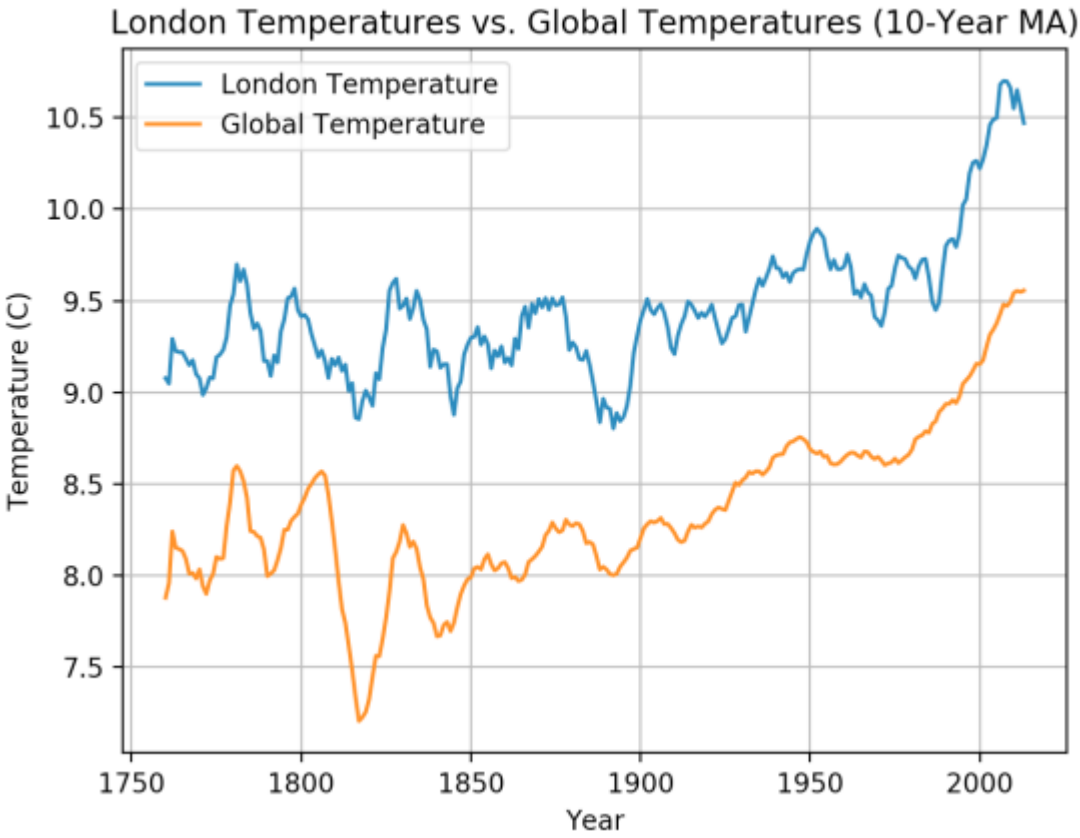


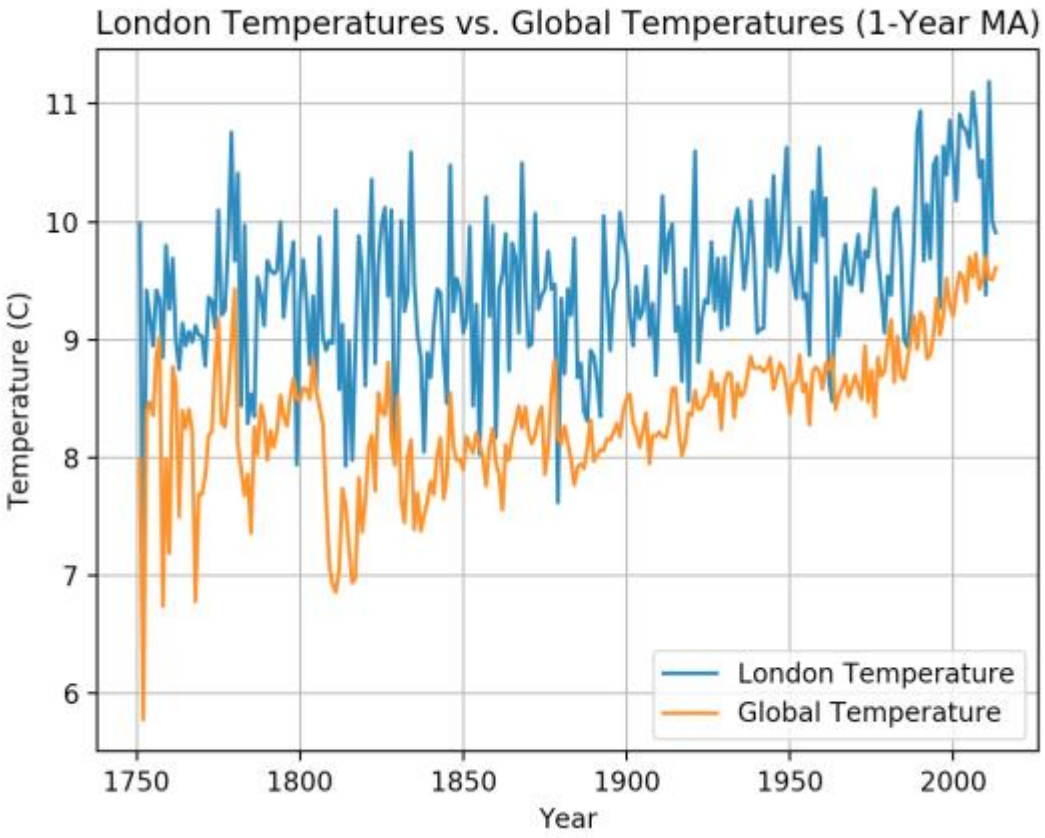
Figure 3: London temperatures vs. global temperatures 10-year moving average trends.

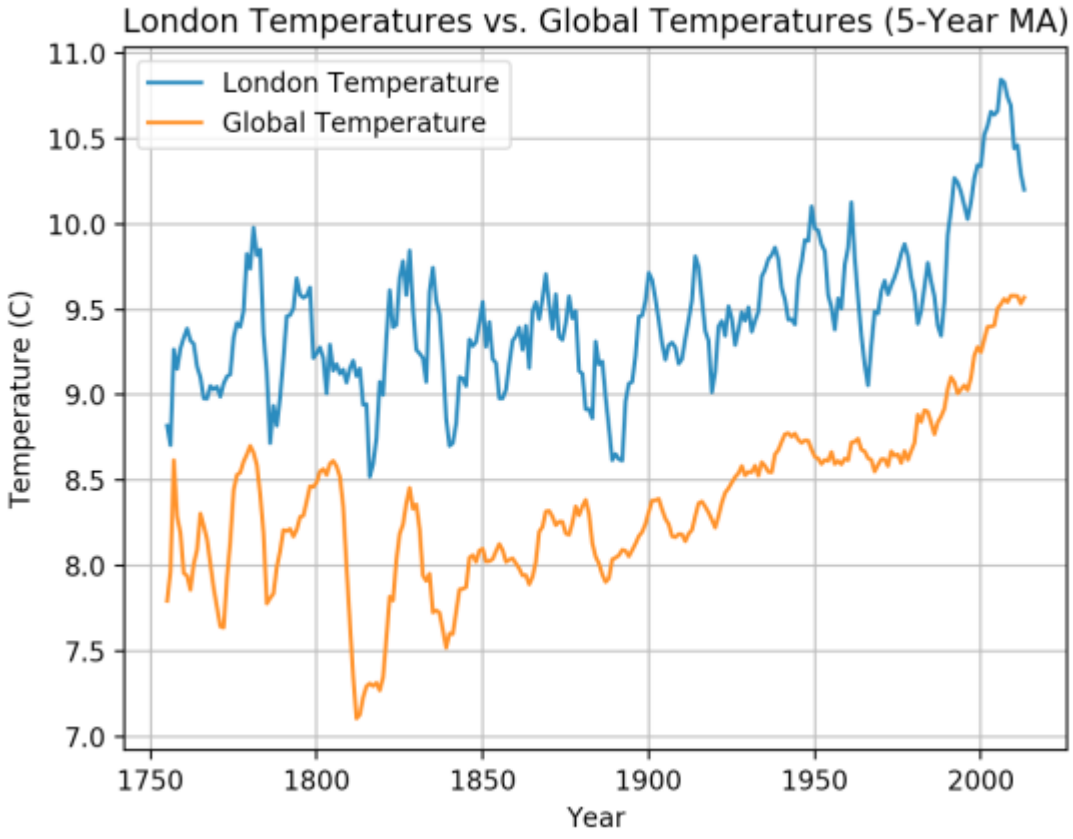
Step 4: Make observations

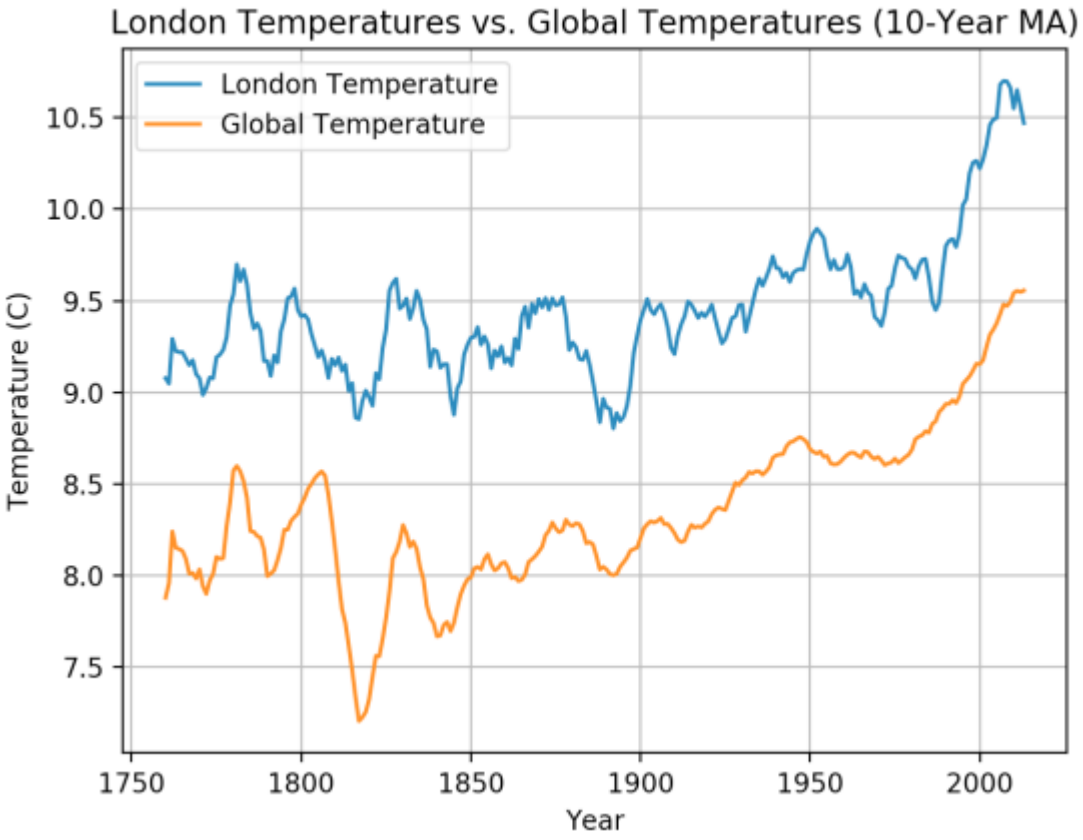
Using Figure 3, it can be seen that London temperature has been consistently hotter compared to the global overall temperature over the past circa. 250 years. On average, London temperature has been around 1C hotter than global average temperature. In addition, it was noticed that the global average trend was much smoother compared to the London trend; this is probably an effect of averaging over all cities/countries around the globe. However, it can also be seen that the London and global temperature trends both have peaks and troughs roughly around the same time periods.

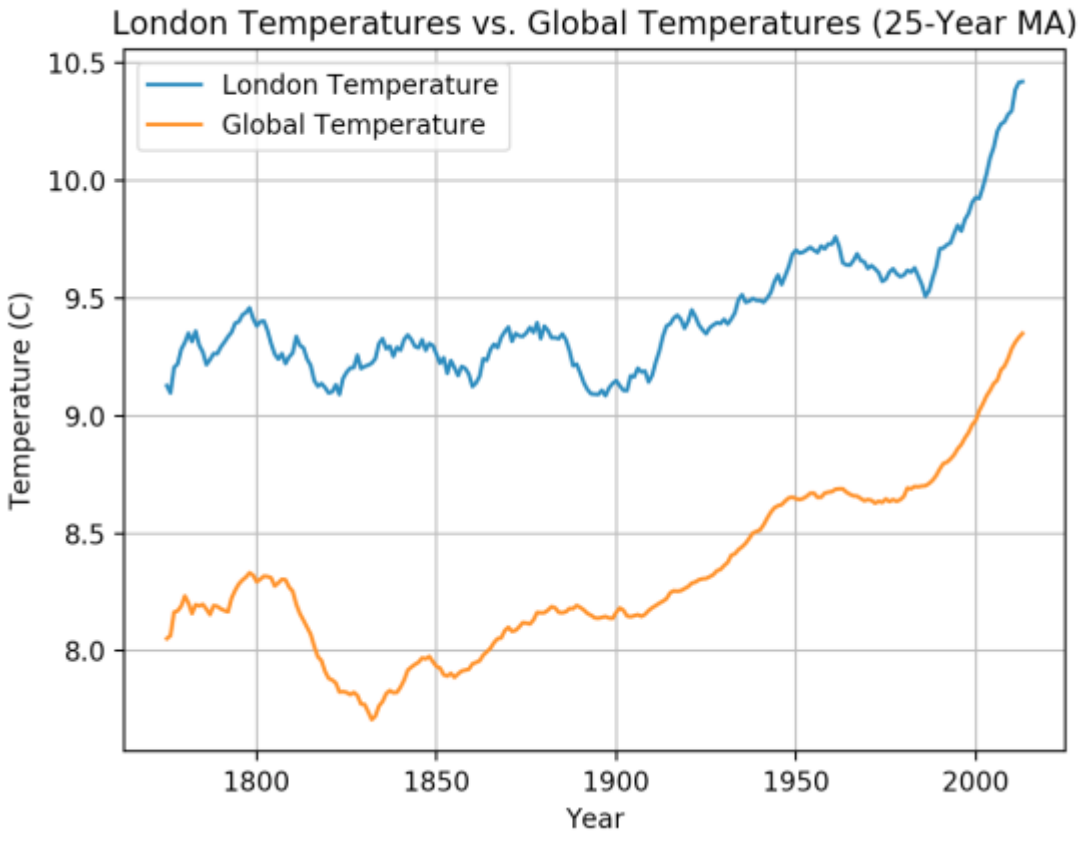
Overall, both temperature trends have been increasing overtime, with particularly more significant increases beyond 1900. Both temperature trends increased by roughly the same amount. This increase could be attributed to the increased usage in carbon-based fuels in the 20th and 21st centuries.

Appendix


London Temperatures vs. Global Temperatures (1-Year MA)
Temperature (C)
Year
London Temperature
Global Temperature
11
10
9
8
7
6
1750
1800
1850
1900
1950
2000


London Temperatures vs. Global Temperatures (5-Year MA)
Temperature (C)
Year
London Temperature
Global Temperature
11.0
10.5
10.0
9.5
9.0
8.5
8.0
7.5
7.0
1750
1800
1850
1900
1950
2000


London Temperatures vs. Global Temperatures (10-Year MA)
London Temperature
Global Temperature
Temperature (C)
10.5
10.0
9.5
9.0
8.5
8.0
7.5
1750
1800
1850
1900
1950
2000
Year


London Temperatures vs. Global Temperatures (25-Year MA)
London Temperature
Global Temperature
Temperature (C)
10.5
10.0
9.5
9.0
8.5
8.0
1800
1850
1900
1950
2000
Year