Pierre Hentges
Udacity Data Analysis Nanodegree – term 1, project 1

Project Submission: Exploring Weather Trends

1. Extracting data from the database

To extract the global temperature data, I used the following query, and exported to CSV:

```
SELECT *
FROM global_data;
```

In order to obtain local temperature data, I first needed to determine the closest city for which information is available in the database:

```
SELECT *
FROM city_list
WHERE country = 'United Kingdom';
```

This query gave 5 results, of which London is the closest city. Query to extract the local temperature data:

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

Using the AND operator to limit results to the UK is important because the database contains data from two cities called London, one in the UK and another one in Canada.

The query results were exported to CSV.

2. Opening the CSV data

I chose Gnumeric (http://www.gnumeric.org) to import and process data, and to make the line chart. Gnumeric is recognised among other spreadsheet software for its accuracy esp. in statistical analysis.

The CSV data could be imported without problems into Gnumeric. However there is a discrepancy in the years for which temperature records are available : global data are available from 1750, whereas the local temperature data start in 1743, without data for the years 1746-1749.

At first, the a single column was created for the year, starting with 1743, and the global and local temperature data were pasted to start from the appropriate cell.

In the final chart, the London temperature data for 1743-1745 were not used in later calculations for the moving averages, for three reasons: (1) there is a suspicious amount of variation, far larger than for other London temperature data, which could be due to imprecision in early measurements, and which is visually distracting ; (2) the interruption of available data for 1746-1749 ; (3) data from before 1750 are not useful for comparisons with global data, which start in 1750.

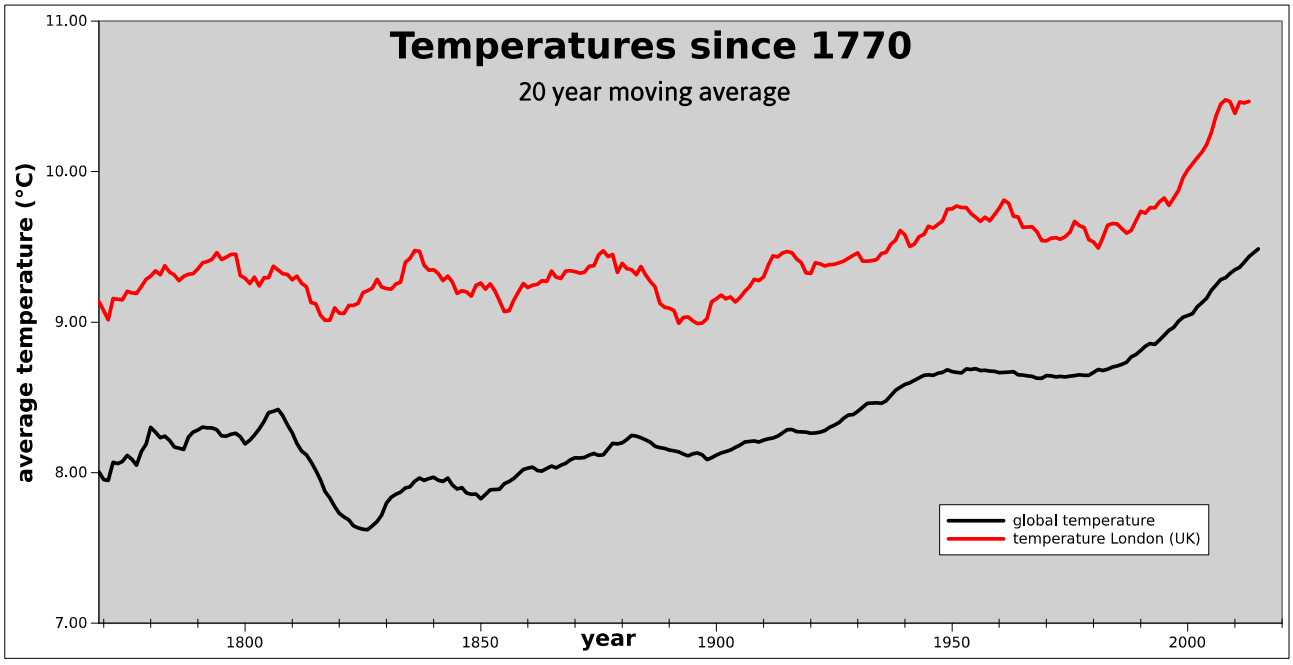
3. Create a line chart

Moving averages were calculated using Gnumeric's *=average()* function using cell data in the preceding rows. To determine the best period over which moving averages should be calculated, I tried out 5 year m.a. (5 years is the average periodicity of the El Niño Southern Oscillation), 10 year m.a., and 20 year m.a.. The moving averages were then used to create line charts.

While the line for the 5 year and 10 year made longer-term trends sufficiently observable for the global temperature data, the resulting line was inadequate for the the local temperature data, which had a noisier/turbulent appearance.

However moving averages calculated over 20 year period produced line charts in which the longer-term temperature trends are well visible for both local and global temperatures, while still preserving sufficient detail. A 20 year moving average is used in the line chart below.

A disadvantage of using a longer period is that it can only be calculated for a shorter timespan – for example, with a 20 year moving average, the earliest year for which temperatures can be shown is 1770, even though temperature records are available from 1750.



4. Observations

London is warmer than global average temperatures, by 1 to 1.5℃ .

London and global temperatures have both increased over the past two centuries, by slightly more than 1℃. For both, the warming trend can be discerned from the mid-19th century, though it has accelerated since, particularly since the 1980s.

The warming trend is interrupted by periods of decreasing or stagnant temperatures. For the global temperatures, this is most prominent in the large decrease of 1806-1825 , the 1885-1900 decrease, the stagnation 1950s-1980.

However, the trough of each decrease is higher than the trough of the preceding decrease, and the peak of each increase is higher than the preceding peack, supporting the notion that there is a robust warming trend.

The lines representing local and global temperatures have a strong tendency to move in the same direction, i.e. each global temperature decrease is also observed in London, as is each increase. The two lines largely have a parallel appearance, though the parallelism is not perfect.

The space between the London and global temperature lines has been slightly narrowing since the mid-19th century, and especially since 2000, though it is unkown if this trend will continue. This implies that the temperature difference between London and global temperatures is decreasing, and that London is warming more slowly than global temperatures.

As pointed out above, the London line has a noisier/more turbulent appearance. This can be explained by the idea that local data reflect measurements made in a single geographical point, while the global data are averages from a large number of locations, leading to a geographical smoothing effect with the global temperature date, in addition to the temporal smoothing of using moving averages for both data sets.