Udacity Data Analyst Nanodegree

Project One: Exploring Weather Trends

Student Name: Sharath V

Project Submitted on 08-June-2021

Table of Contents

Steps to Prepare Data for Visualization and Analysis

The following steps were taken to prepare data and analyze the weather trends:

1. Selecting the closest big city from the 'city_list' database.

2. Querying the temperature date for the world and the closest big city (which is Amsterdam in this case) using SQL Workspace in the learning platform.

3. Downloading and exporting the data to CSV.

4. Using Excel to analyze the CSV and preparing the line chart for moving average of global and local temperature trends.

SQL query used

```
select
cd.year,
cd.city,
cd.country,
cd.avg_temp city_avg,
gd.avg_temp global_avg
from city_data cd
join global_data gd on cd.year = gd.year
where cd.city = 'Amsterdam'
order by cd.year
```

Deciding the period for Moving Average (MA)

The next step involved deciding the period for the moving averages. Based on the size of the downloaded dataset, which was 264 rows, both 5-year and 10-year Moving Average (MA) line charts were prepared.

Key considerations for deciding the Moving Average

1. To make sure the MA period chosen smoothens out the data to observe the long-term trends while not losing out the details that is needed to analyze.

Calculation of Moving Average (MA)

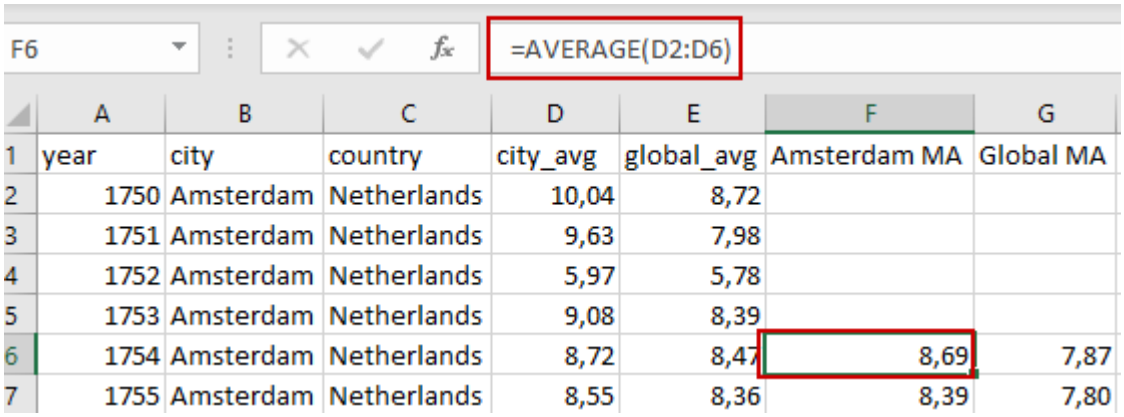
MA (for 5-year) was calculated using the following steps:

1. Two additional columns were created (refer fig.1). One for local (column F) and the other one for global MA (column G).

2. For 5-year MA, AVERAGE () function was used in cell F6, which referred to the 5th year (year 1754). This function returned the average of the local temperature for the first five years.

3. The formula was dragged and used in column G6 to calculate the global MA.

4. Next, selecting both Mas (cell F6 and G6), the formula was dragged to the cells below till the end of the dataset (till cell F265, G265).



	A	B	C	D	E	F	G
1	year	city	country	city_avg	global_avg	Amsterdam MA	Global MA
2	1750	Amsterdam	Netherlands	10,04	8,72		
3	1751	Amsterdam	Netherlands	9,63	7,98		
4	1752	Amsterdam	Netherlands	5,97	5,78		
5	1753	Amsterdam	Netherlands	9,08	8,39		
6	1754	Amsterdam	Netherlands	8,72	8,47	8,69	7,87
7	1755	Amsterdam	Netherlands	8,55	8,36	8,39	7,80

Key considerations on how to visualize the trend

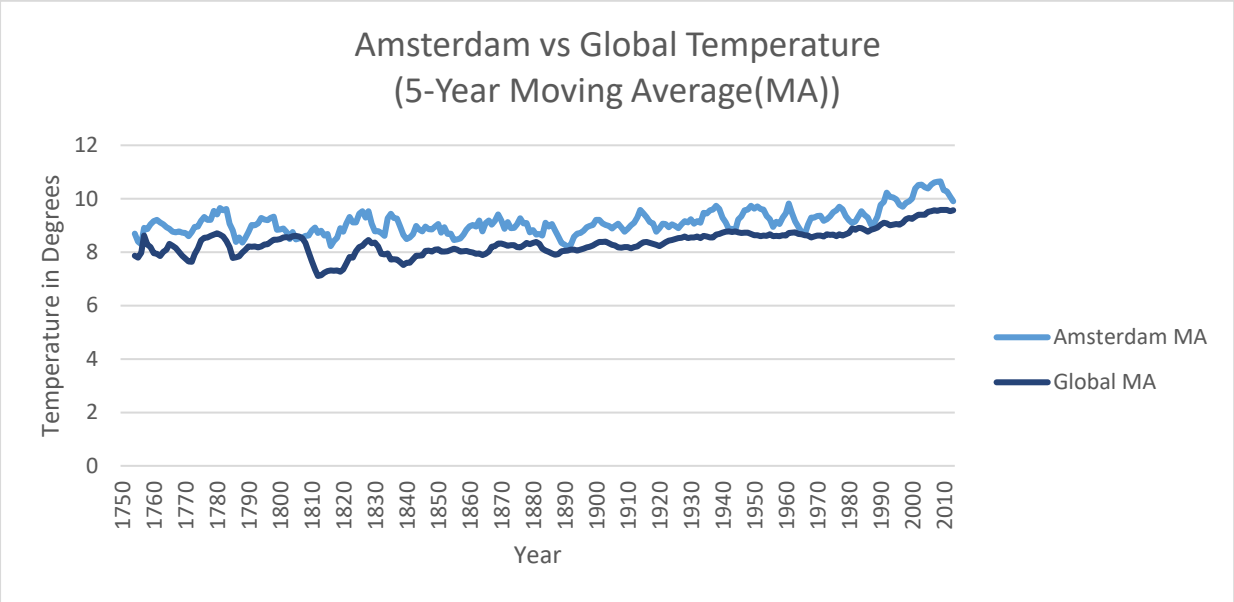
1. Clear chart titles, axis labels and legends to help the user / viewer to clearly understand the graph.

2. Use of neutral coloring to avoid association of colors to subjective or subconscious judgment of the user.

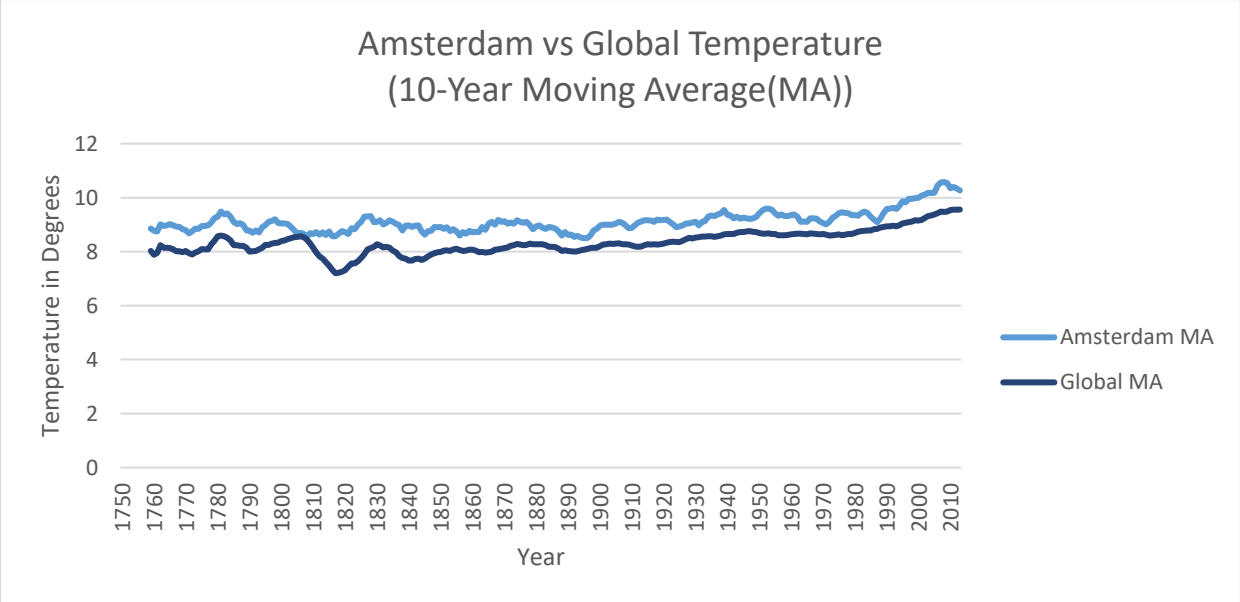
3. To clearly mention the units of measurement in the axis.

Moving Average Charts

5-year MA Chart



10-Year MA Chart



Comparing the line charts for both the MA periods, the 5-year MA provided both the long-term trend and the details required for trend analysis. The details of outliers are not easy to observe in the 10-Year MA chart. **Hence, a 5-year Moving Average was chosen for trend analysis.**

Observations about the weather trends

1. Both the local and global weather trends show that the average temperatures have increased consistently from early 1900s till 2013. However, the increase has been steady in the case of global trend while the local trend has been varying over the years.

2. The data shows that the local MA is closer to the global MA but always slightly higher than the global MA. In other words, temperatures in Amsterdam are usually slightly higher than the global average. The average difference in the MA temperatures of local and global is less than 1 degree (~ 0.77 degrees more in Amsterdam than global average).

3. The local and global MA are having a positive co-efficient correlation of 0.75 for the whole period in the graph. A positive co-efficient correlation of 0.75 means that as global MA increases, local MA increases and when global MA decreases, local MA decreases. It also indicates a moderate-to-good correlation between the two MAs.

4. The standard deviation for the average local and global temperatures is 0.77 and 0.58 respectively indicating that the local temperatures spread out more from their averages (respective averages of the entire dataset) than the global average figures.

5. An interesting period is from 1939-1944, where the local MA dropped consistently while the global MA was quite stable. This is one of the outliers' details we observe when we consider a smaller period as our sample.

6. During 1809-1812, when the global MA kept decreasing, the local MA increased steadily leading to the difference between the global and local to a maximum of 1.6 degrees.