SQL Query

Find which cities are avalable in the data

```
select * from city_list where country = 'Turkey'
```

Get the “Istanbul” data

```
select * from city_data where city = 'Istanbul'
```

Get the global data

```
select * from global_data
```

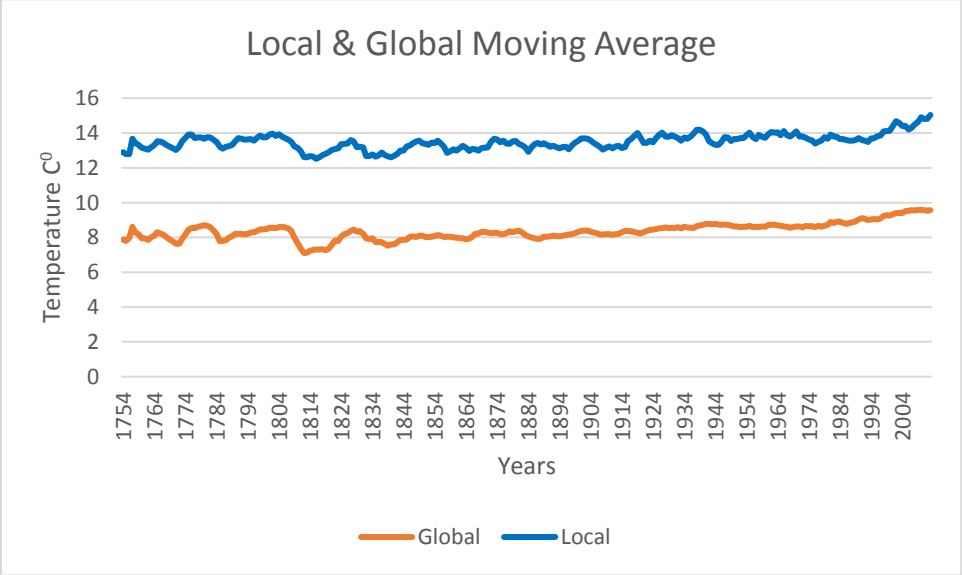
To get the yearly basis global and local data

```
SELECT a.Year, a.avg_temp global_temp, b.avg_temp local_temp

FROM global_data a INNER JOIN city_data b ON a.year = b.year

where b.city = 'Istanbul'
```

Outline

- First, I wrote an SQL Query to match the corresponding records of the same year,
- Then copy the result to an Excel spreadsheet
- Calculate the moving averages for 5 years period
- I wanted to minimize the sharp fluctuation affect of the yearly trend, instead see the average trend of the weather

Moving Average (With 5 years period)



Global Data Basic Stats

MinOfavg_temp	AvgOfavg_temp	MaxOfavg_temp	StDevOfavg_temp
5.78	8.369473684	9.83	0.58474741

Local Data Basic Stats (Istanbul)

MinOfavg_temp	AvgOfavg_temp	MaxOfavg_temp	StDevOfavg_temp
10.37	13.47677903	9.48	0.8544374

- Overall according to the basic statistics the Local weather is more hot then the global weather.
- Over the years the weather gets hot according to the previous years.
- Both global and local weather gives the same trends at some points that is they both decrease or increase.
- Global weather has more smoother increase after the year 1844 but the local weather has more fluctuation

Correlation Coefficient

I used the pearson' s correlation coefficient formula

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

$$n = 264$$
$$\sum x = 2206$$
$$\sum y = 3566$$
$$\sum xy = 29891$$
$$\sum x^2 = 18535$$
$$\sum y^2 = 48315$$

I have got the result by this calculation

Correlation Coefficient = 0.718

Correlation coefficient formulas are used to find how strong a relationship is between data. The formulas return a value between -1 and 1, where:

- 1 indicates a strong positive relationship.
- -1 indicates a strong negative relationship.
- A result of zero indicates no relationship at all.

According to the definition we can say that there is almost a strong correlation between the local and global weather.

Point Estimation

Yes we can estimate the local temperature by mean with unknown standard deviation with T-table
We can accept the confidence interval %95 of two side;

$$x - T_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < x + T_{\alpha/2}\frac{s}{\sqrt{n}}$$

x : mean of sample : 8.36

s : unknown standart deviation : 0.58

T : distribution : T(%95) = 1.984

α/2 = 0.025

n = sample size : 266 -1 = 265

⇨ By using the formula the result is

$$8.36 - 1.984\,\frac{0.58}{\sqrt{265}} < \mu < 8.36 + 1.984\,\frac{0.58}{\sqrt{265}}$$

$$8.28 < \mu < 8.36$$