

## Project : Explore Weather Trends

Nanodegree Udacity Program: Data Analyst

Student: Abderrahman BENYAHYA

Morocco, Rabat

17 May 2019

1. An outline of steps taken to prepare the data to be visualized in the chart, such as: What tools did you use for each step? (Python, SQL, Excel, etc)

During the project, I had first to extract the data from the Database in the classroom using the SQL commands. The first difficulty was to identify which city should be considered as mine.

I entered first the SQL command to verify if there are and how many cities from Morocco using the code:

```
SELECT country, city, count(city)
FROM city_data
WHERE country = 'Morocco'
GROUP By 1,2
```

Input		HISTORY ▾	MENU ▾
SCHEMA ↻		1 SELECT country, city, count(city)	
city_data ▾		2 FROM city_data	
city_list ▾		3 WHERE country = 'Morocco'	
global_data ▾		4 GROUP By 1,2	
		5	
		6	
		Success!	EVALUATE
Output		2 results	Download CSV
country	city	count	
Morocco	Casablanca	251	
Morocco	Rabat	261	

It returns the results that there are two cities Casablanca and Rabat. So I write the second code:

```
SELECT year, city, avg_temp
FROM city_data
WHERE city = 'Rabat'
```

I uploaded the result in csv and created the chart using Excel 2019 professional but also Open Office to have the csv converted properly to odt and xlsx.

## Project : Explore Weather Trends

I did the same with the global data.

```
SELECT year, avg_temp  
FROM global_data
```

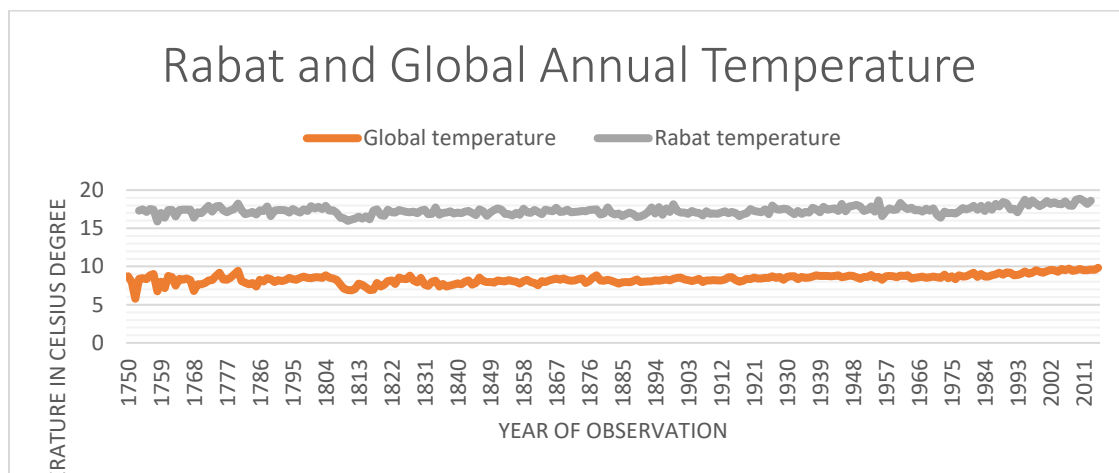
Then I created the chart by calculating the average for ten years as indicated in the classroom. But I found out that it would be more efficient to have the difference between each temperature and the average of temperature for the entire time. I calculated the average for the all period and then calculate the difference between the temperature of each year and the average.

The most effective visualization in line diagram as it should reflect the fluctuation by year. The color of the two elements (Rabat and Global) should be distinctive.

I used R and R Studio to import the csv then to calculate the Pearson correlation between the two data (Rabat and Global)

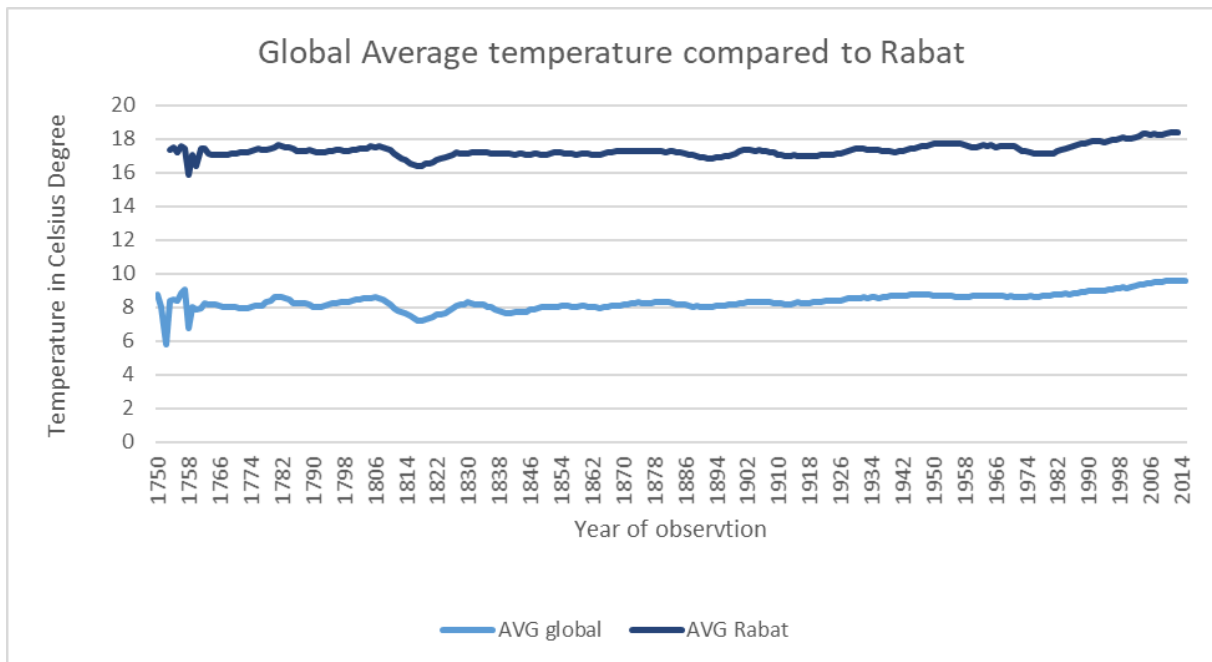
```
>smp = read.csv2(file = "compare Rabat Global.csv", dec=',')  
>cor.test(smp$AVG.global,smp$AVG.Rabat)
```

### 2. Line chart with local and global temperature trends

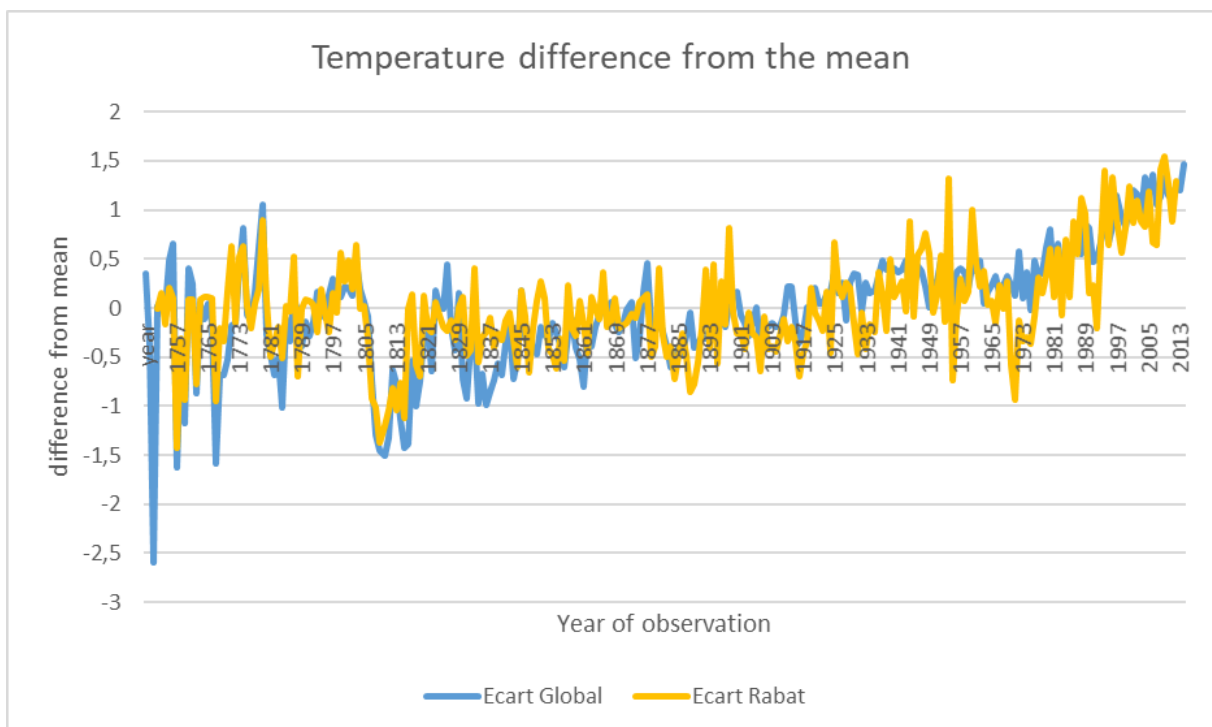


**Fig1. Annual temperature for Rabat and Global**

## Project : Explore Weather Trends



**Fig2. Average temperature Rabat and Global ten years.**



**Fig3. Distance to the mean temperature for Rabat and Global**

## Project : Explore Weather Trends

3. At least four observations about the similarities and/or differences in the trends
  - The two first line charts are very similar even if the average tends to erase the discontinuity
  - We can see that Rabat is always warmer than the global average temperature
  - The two lines are following almost the same trend both in up and down. The person's product correlation calculated with R is 0.8971999, very strong correlation to conclude that Rabat is following the global trends.
  - Using the two first chart does not give a really good picture of the climate change over the years. This is why it is preferred to calculate the average of temperature for the all period and see for each year the distance to mean.
  - The last chart shows clearly that there is a correlation between the two line (Rabat and Global). The global warming is very clear and shown by the increase of temperate since almost 1909. It seems that the distance was never so high in history as it is since 1989.