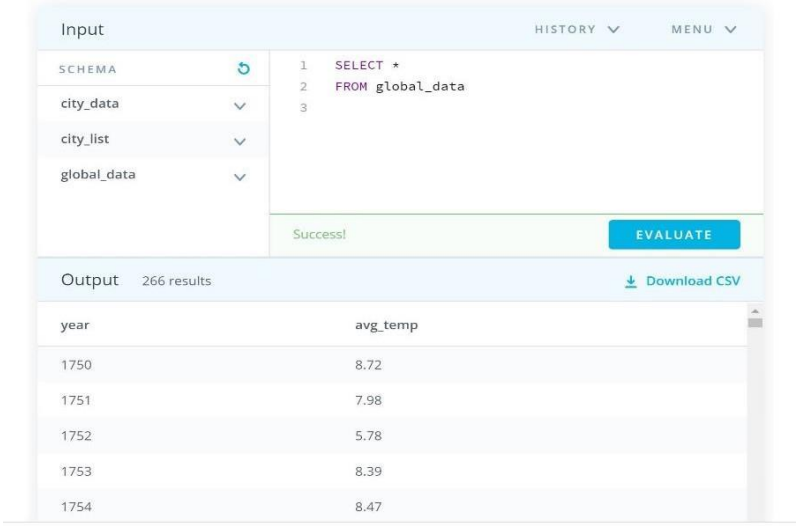
Data Analyst Course

Exploring Weather Trends project

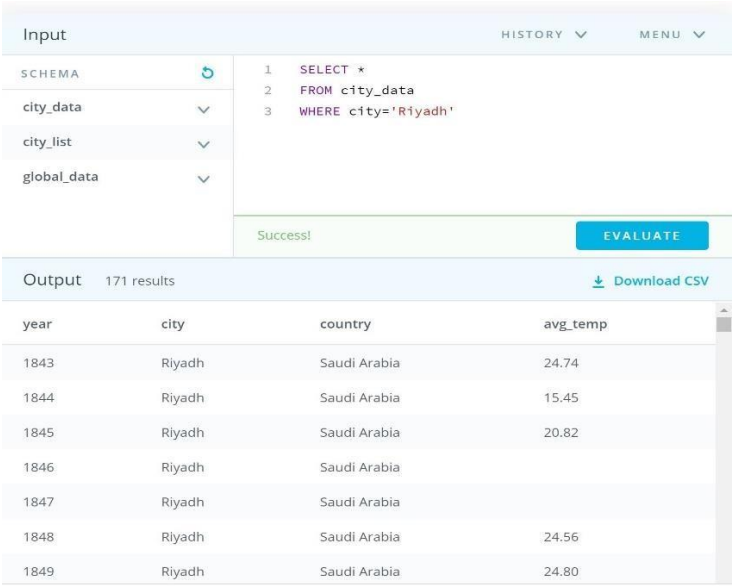
Work by :

Nada Almutairi

- **Extract the data using SQL**
  - the Global Data query



  - The Riyadh data query

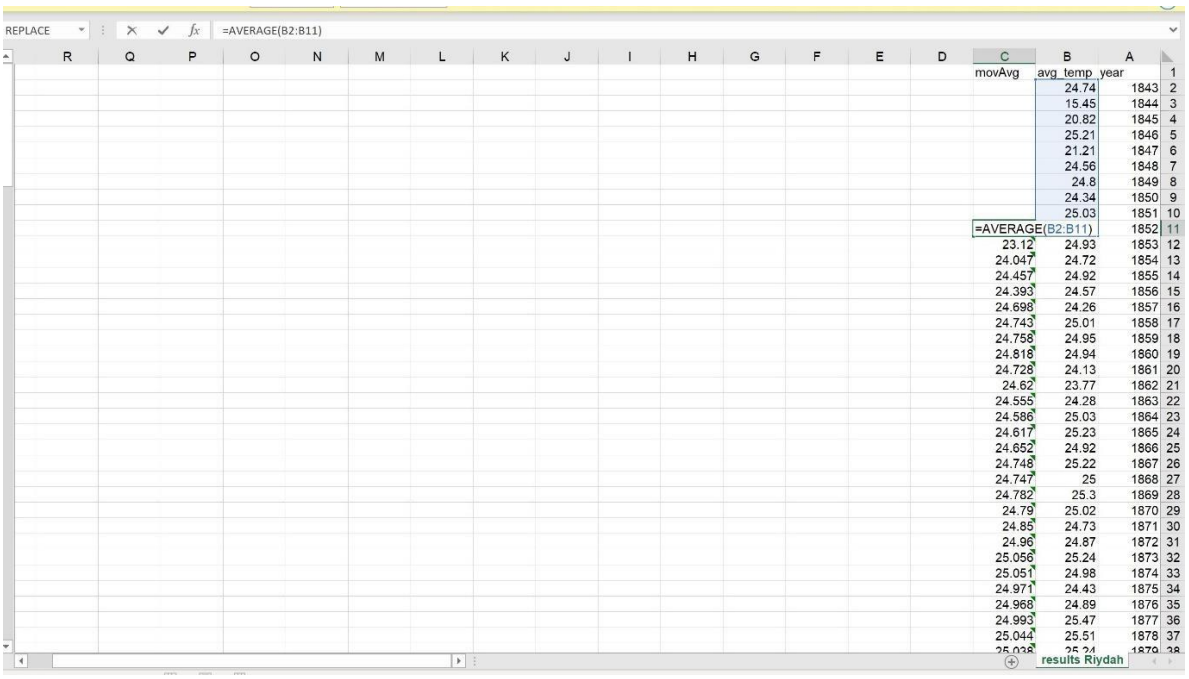


- **Preprocessing the data**
    - used Excel to Prepare the two datasets

First, Riyadh Data has two NA value. Because the data is already small , I see the good way is replacement NA's with average .
Also, I deleted City and country columns. They are useless for the comparison.

Moreover, the number of years in global data is larger than Riyadh data, it should be compatible to compare them. So I had to choose the same period which is from 1843 to 2013 .

I find ten years is the best number to calculate the moving average with. Since I try the data distribution with five years, there was too much noise data. And, with 15 years the data distribution was totally different from original data. it may loss of details



- **Visualize the data with R**
  - I choose it to visualize the Data with R .Since, I most familiar with it than other tools.

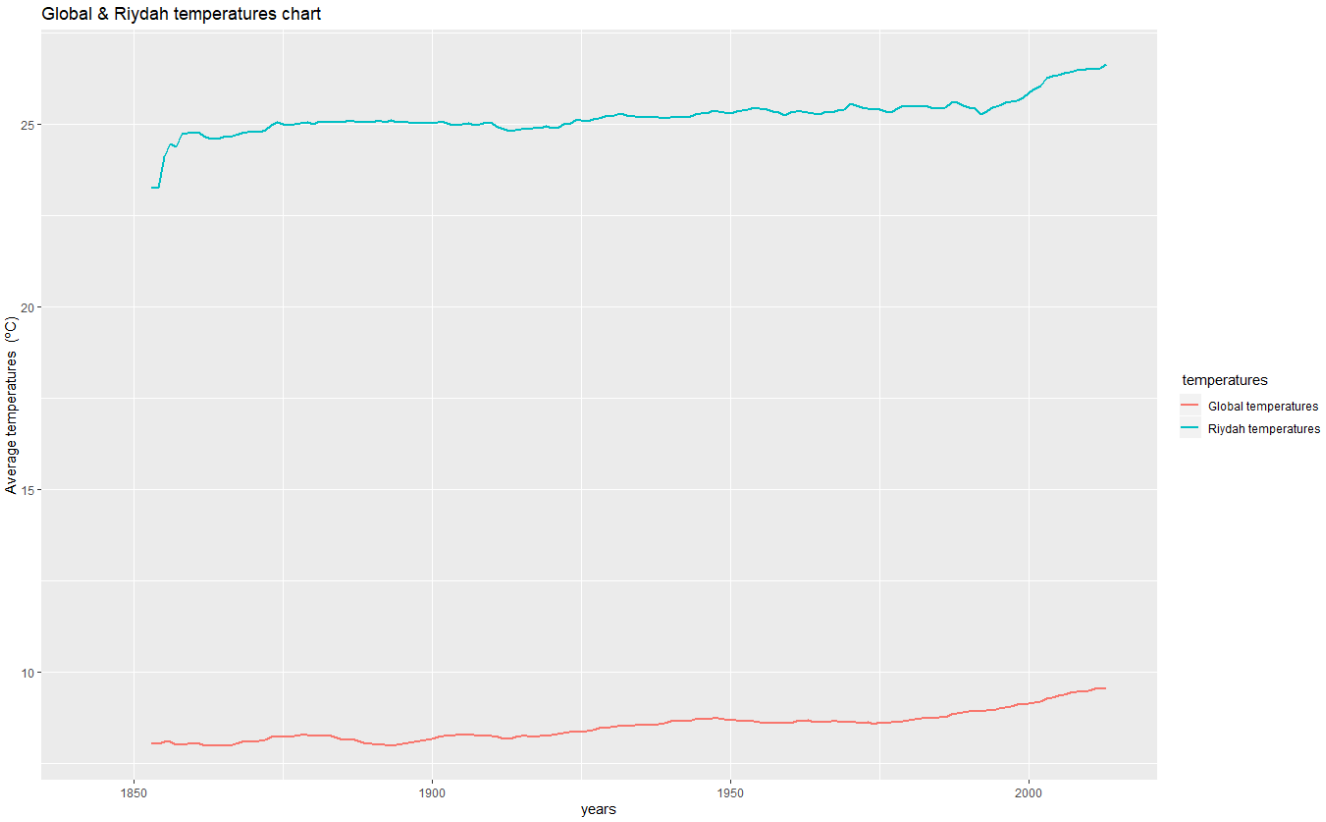
```r
#read the datasets
riyData = read.csv("results Riydah.csv")
globData = read.csv("results global.csv")



#Moving Avg. from 1843 to 1852 is alredy calculated in Excel
#To avoid Warning messages from R I choose this subset
riyData = riyData [riyData$year >=1852,]
globData = globData [globData$year >=1852,]


# requier to visualize the datasets
library(ggplot2)


ggplot() +
  geom_line(data=riyData, aes(year, movAvg , colour="red"), size=1 ) +
  geom_line(data=globData,  aes(year, movAvg , colour="blue"), size=1 ) +
  labs(x='years' , y='Average temperatures  (°C)', title='Global & Riydah temperatures chart') + #for title and axis names
  scale_color_discrete(name=" temperatures" , labels = c("Global temperatures","Riydah temperatures")) #for lagned
```



- **Findings**
  - Riyadh city temperature average is 25.19. While global average is 8.53. The difference between them is 16.66 degrees, it is clear that Riyadh is hotter than global.
  - Riyadh Temperature from 1850 to almost 1875 has risen 2.5 degrees while global temperatures remain stable.
  - In 2000 , the Riyadh has increased by two degrees then it gone up . While, The Global temperature increased by one degree.
  - Also, In 1925, the Riyadh temperature has increased as global temperature is also increased. Which means there is a positives relationship between them. To make sure I find the Correlation coefficient between the average for both global & Riyadh, and the result is 0.77 . For moving average is 0.87 . Since it is rather close to 1, we can conclude that the global temperatures and Riyadh temperature are positively related.
    Here is the code :

References

http://www.rdocumentation.org/