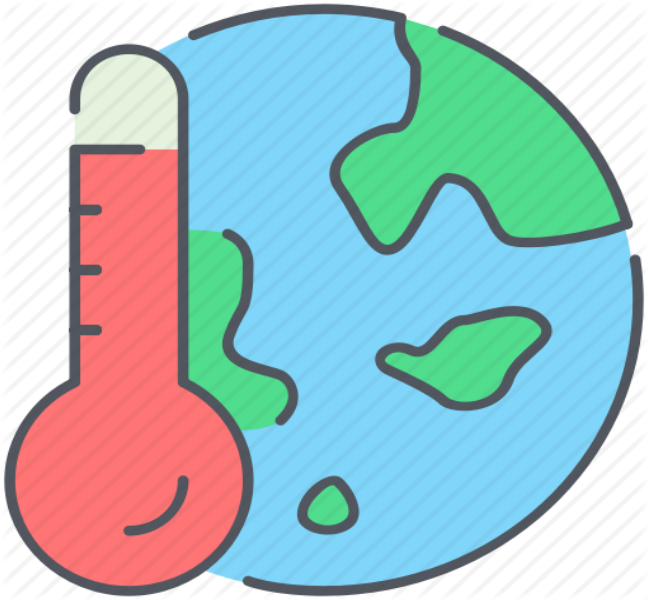
UDACITY

DATA ANALYST NANO DEGREE
2018



Explore Weather Trends

Vidhi Thakkar | 21st August, 2018

Introduction

The world is getting warmer. Whether the cause is human activity or natural variability and the preponderance of evidence says its humans. According to an ongoing temperature analysis conducted by scientists at NASA, the average global temperature on Earth has increased by about 0.8° Celsius (1.4° Fahrenheit) since 1880. Two-thirds of the warming has occurred since 1975, at a rate of roughly 0.15-0.20°C per decade.

Overview

In this project, I have been provided with a database on Udacity portal from where I have extracted, manipulated and visualized the data. I have analysed the local temperature of two cities individually, Ahmadabad (India) and Dallas (United States), in accordance with the global temperature data. I will compare and see how each city has been impacted by the climate change over the years.

Tools Used

1. SQL: To extract the data from the database
2. R: For calculating moving average and plotting line chart

Steps

1. Extraction of data from the database and export to CSV file

These are the SQL codes that I used to retrieve the data from the Udacity database.

```
SELECT global_data.year AS "Year", global_data.avg_temp AS "Global Temp",
city_data.avg_temp AS "Ahmadabad Temp"
FROM global_data
INNER JOIN city_data
ON global_data.year = city_data.year
WHERE city = 'Ahmadabad';
```

```
SELECT global_data.year AS "Year", global_data.avg_temp AS "Global Temp",
city_data.avg_temp AS "Dallas Temp"
FROM global_data
INNER JOIN city_data
ON global_data.year = city_data.year
WHERE city = 'Dallas';
```

2. Make a chart visualization based on extracted data

I used R programming language to write codes for finding out the Moving Averages of the local and global temperature and to visualise them using a Line chart.
(For simplicity, only the code written for Ahmadabad is shown)

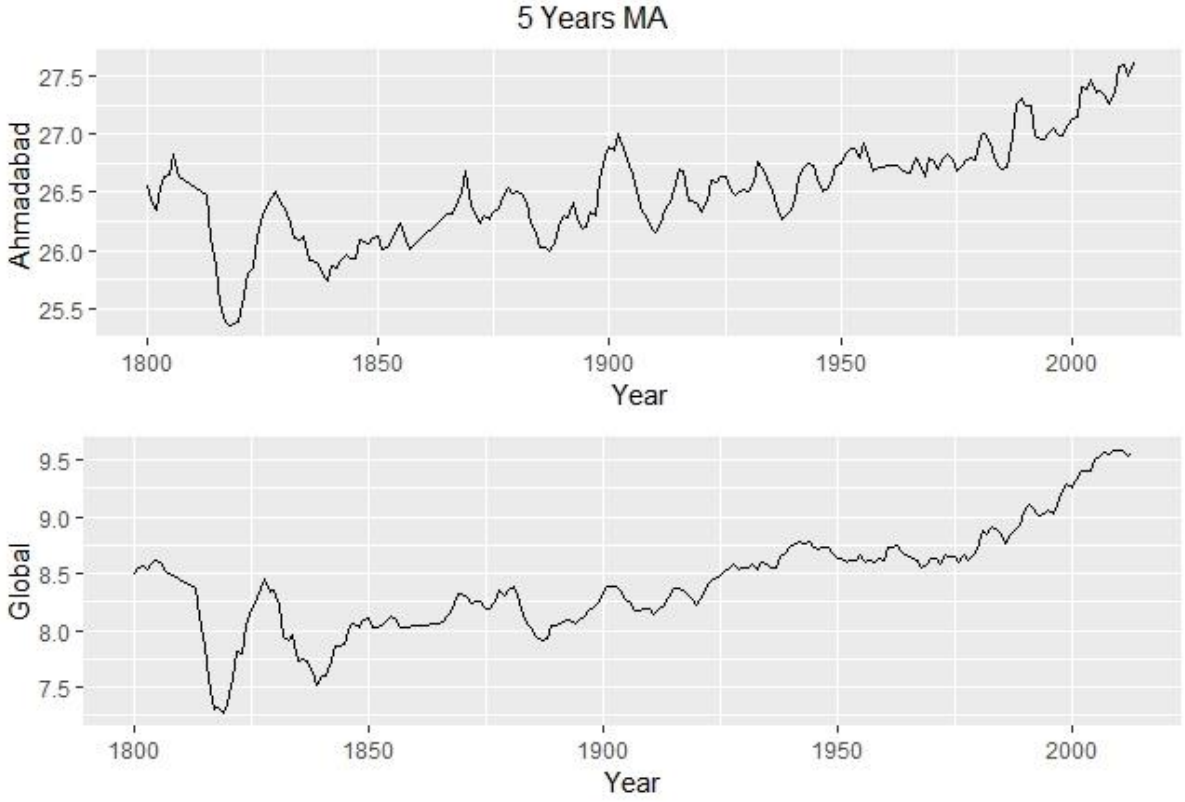
R Code

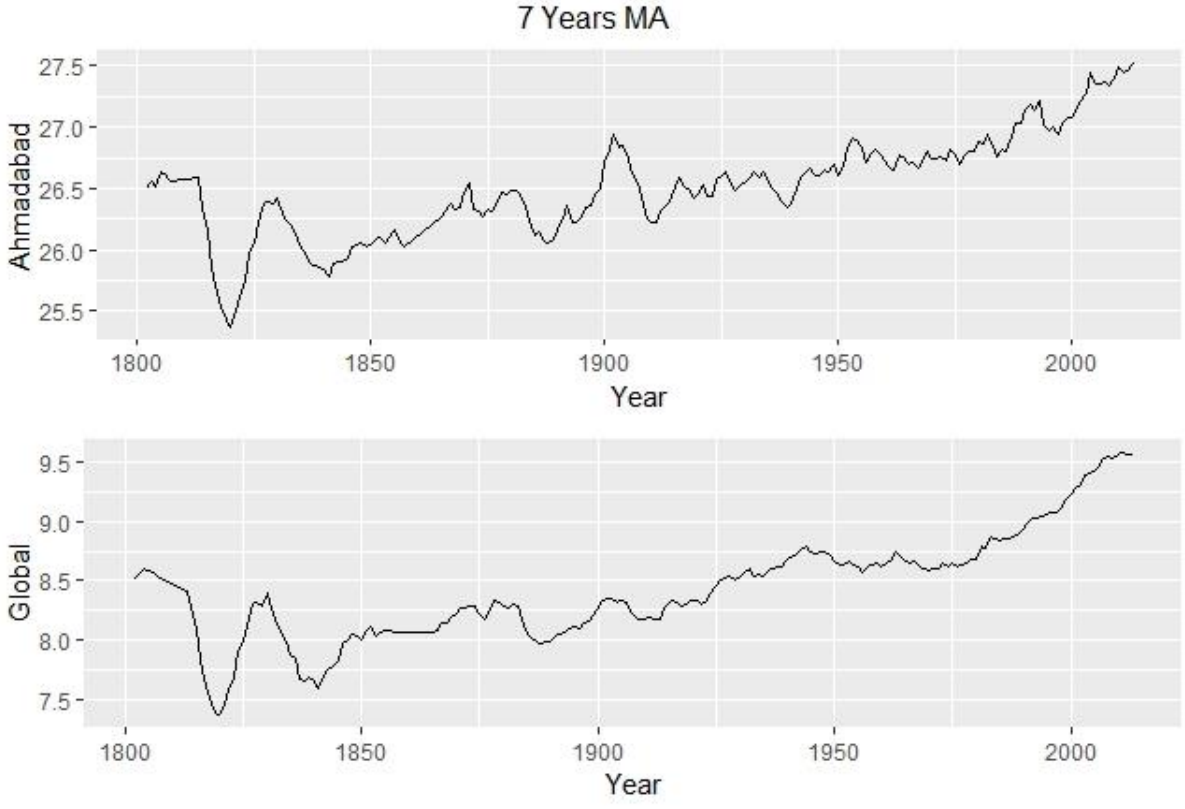
```
# a) Remove the previous data (if any)
rm(list=ls())

# b) Load the Packages
library(TTR)                  # for calculating Moving Average
library(ggplot2)              # for ggplot

library(grid)
library(gridExtra)
library(ggpubr)

# c) Set working directory
setwd("C:/Users/thakk/Desktop/Udacity Project 1")

# d) Load the files into data frames
Ahmadabad <-  read.csv("GA.csv", stringsAsFactors = FALSE)

# e) Check and Remove the rows with NA vlaues of avg temperatures
length(which(is.na(Ahmadabad)))

## [1] 12

# Ahmadabad has NA values
Ahmadabad  <- na.omit(Ahmadabad)

# a) Creating time  series  dataframes to calculate moving average
Ahmadabad_timeseries <-  ts(Ahmadabad$Ahmadabad.Temp)
Global_timeseries    <-  ts(Ahmadabad$Global.Temp)

# b) Moving  Averages  5  Years
Ahmadabad_SMA_5  <-  data.frame(SMA(Ahmadabad_timeseries,n=5))
Global_SMA_5     <-  data.frame(SMA(Global_timeseries,n=5))

# c) Moving  Averages  7  Years
Ahmadabad_SMA_7  <-  data.frame(SMA(Ahmadabad_timeseries,n=7))
Global_SMA_7     <-  data.frame(SMA(Global_timeseries,n=7))

# d) Moving  Averages  10  Years
Ahmadabad_SMA_10  <-  data.frame(SMA(Ahmadabad_timeseries,n=10))
Global_SMA_10     <-  data.frame(SMA(Global_timeseries,n=10))
```

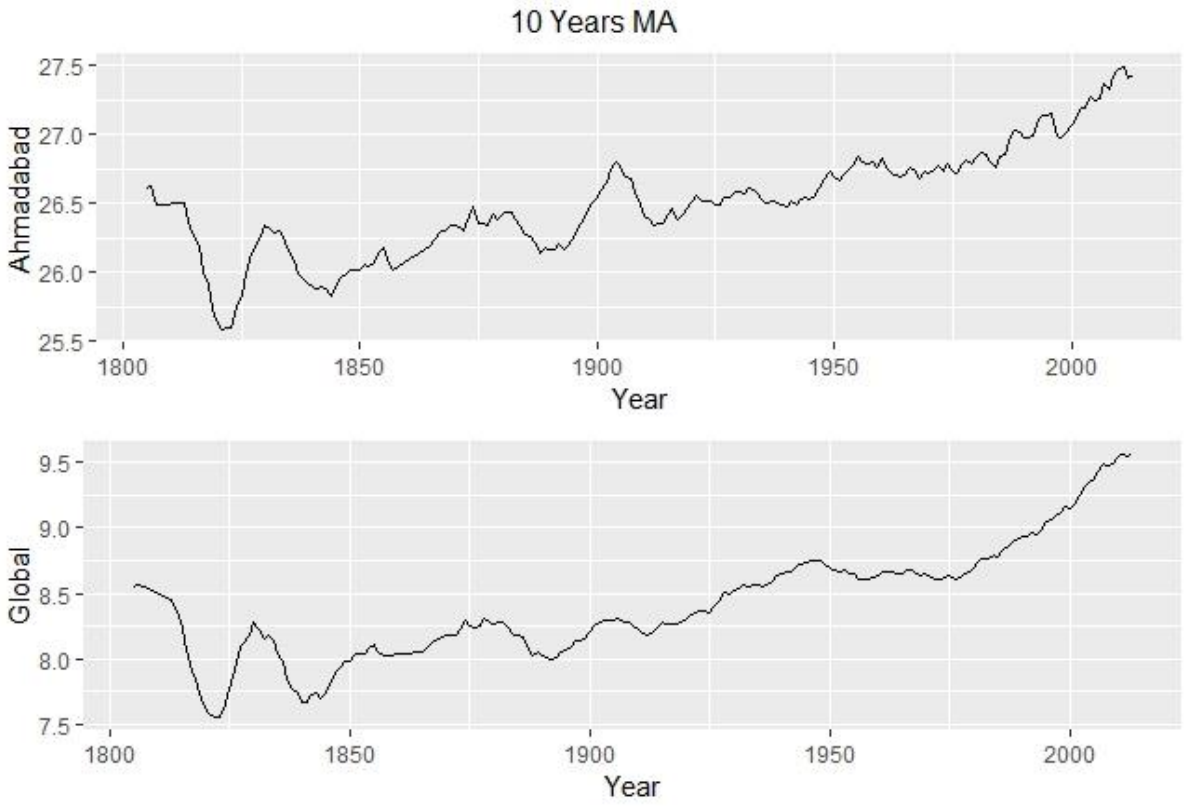
```r
# e) Combining Moving Average dataframe and year from the original data
frame
Ahmadabad_5 <- cbind(Ahmadabad_SMA_5,Ahmadabad$Year)
Global_5    <- cbind(Global_SMA_5,Ahmadabad$Year)

Ahmadabad_7 <- cbind(Ahmadabad_SMA_7,Ahmadabad$Year)
Global_7    <- cbind(Global_SMA_7,Ahmadabad$Year)

Ahmadabad_10 <- cbind(Ahmadabad_SMA_10,Ahmadabad$Year)
Global_10    <- cbind(Global_SMA_10,Ahmadabad$Year)

# f) Changing column names for all the data frames

colnames(Ahmadabad_5)[1]  <- "Ahmadabad"
colnames(Ahmadabad_7)[1] <- "Ahmadabad"
colnames(Ahmadabad_10)[1]  <- "Ahmadabad"
colnames(Ahmadabad_5)[2]  <- "Year"
colnames(Ahmadabad_7)[2] <- "Year"
colnames(Ahmadabad_10)[2] <- "Year"


colnames(Global_5)[1]  <- "Global"
colnames(Global_7)[1] <- "Global"
colnames(Global_10)[1]  <- "Global"
colnames(Global_5)[2]  <- "Year"
colnames(Global_7)[2] <- "Year"
colnames(Global_10)[2] <- "Year"

# g) Omiting NA values
Ahmadabad_5   <- na.omit(Ahmadabad_5)
Global_5      <- na.omit(Global_5)

Ahmadabad_7  <- na.omit(Ahmadabad_7)
Global_7     <- na.omit(Global_7)

Ahmadabad_10  <- na.omit(Ahmadabad_10)
Global_10     <- na.omit(Global_10)

# h) Arrange multiple plots on a page using grid.arrange

# 5, 7 and 10 Years Moving Average plots for Ahmadabad
AhmadabadPlot_1  <- ggplot(Ahmadabad_5, aes(x=Year,y=Ahmadabad))+
geom_line()
AhmadabadPlot_2  <- ggplot(Ahmadabad_7, aes(x=Year,y=Ahmadabad))+
geom_line()
AhmadabadPlot_3  <- ggplot(Ahmadabad_10, aes(x=Year,y=Ahmadabad))+
geom_line()

grid.arrange(AhmadabadPlot_1,AhmadabadPlot_2,AhmadabadPlot_3,nrow =
3, top="5, 7 & 10 Years Moving Averages(Ahmadabad)")
```

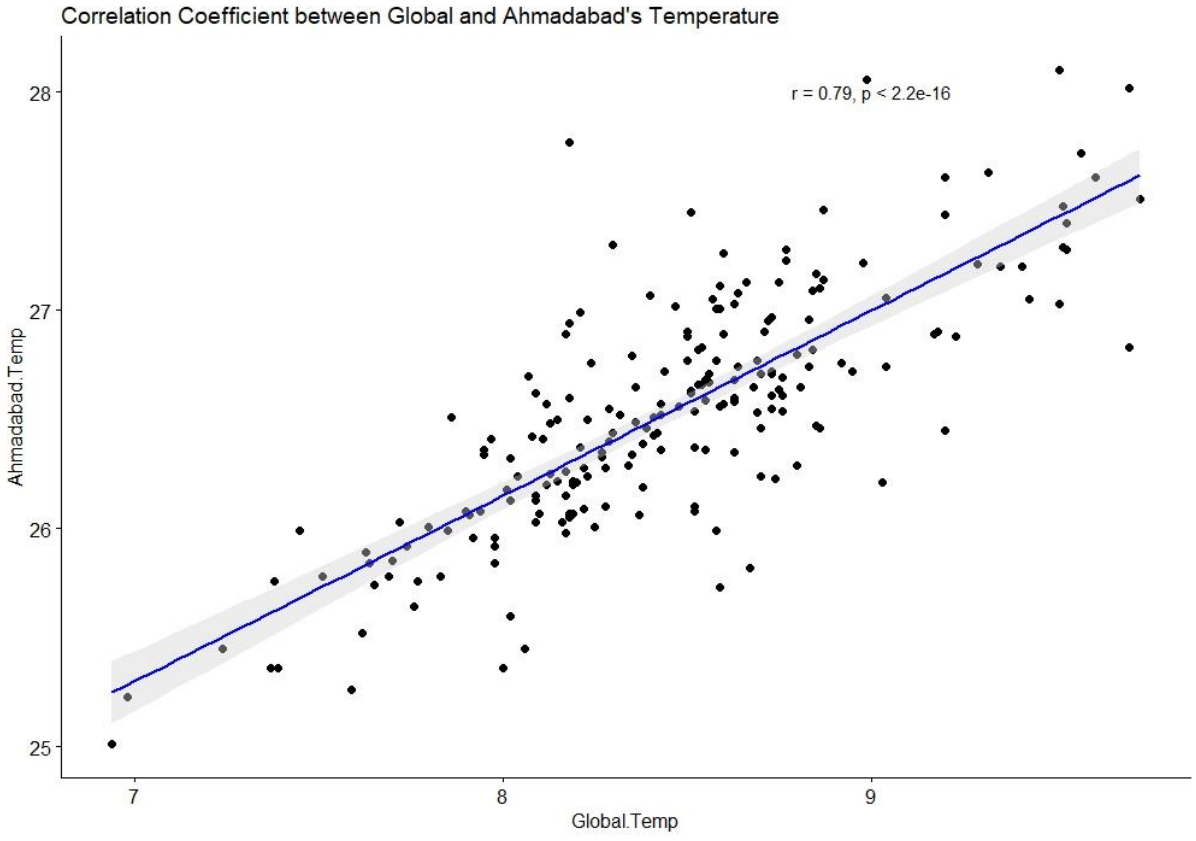
```r
#  5, 7 and 10 Years Moving Average plots for the Globe
GlobalPlot_1    <- ggplot(Global_5, aes(x=Year,y=Global))+ geom_line()
GlobalPlot_2    <- ggplot(Global_7, aes(x=Year,y=Global))+ geom_line()
GlobalPlot_3    <- ggplot(Global_10, aes(x=Year,y=Global))+ geom_line()

grid.arrange(GlobalPlot_1,GlobalPlot_2,GlobalPlot_3,nrow = 3, top="5, 7 &
10 Years Moving Averages(Global)")


# Final Plots
grid.arrange(AhmadabadPlot_1,GlobalPlot_1,top="5 Years MA")

grid.arrange(AhmadabadPlot_2,GlobalPlot_2,top="7 Years MA")

grid.arrange(AhmadabadPlot_3,GlobalPlot_3,top="10 Years MA")


# i) Scatter plot with correlation coefficient
sp <- ggscatter(Ahmadabad, x = "Global.Temp", y = "Ahmadabad.Temp",
                add = "reg.line", add.params = list(color = "blue", fill =
 "lightgray"),
                conf.int = TRUE, title = "Correlation Coefficient between
Global and Ahmadabad's Temperature" )
# Add correlation coefficient
sp + stat_cor(method = "pearson", label.x = 9, label.y = 28)

sp
```

AHMADABAD

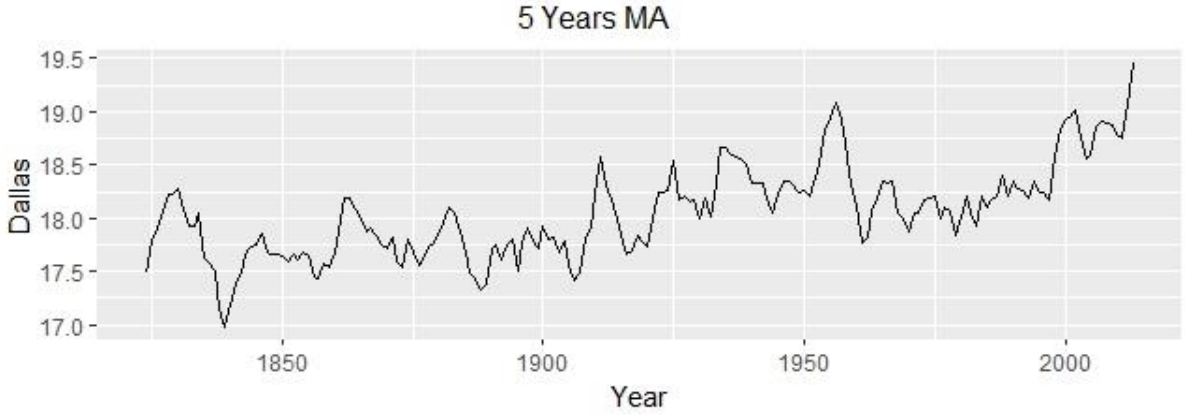

5 Years MA
Ahmadabad
27.5
27.0
26.5
26.0
25.5
1800
1850
1900
1950
2000
Year
Global
9.5
9.0
8.5
8.0
7.5
1800
1850
1900
1950
2000
Year

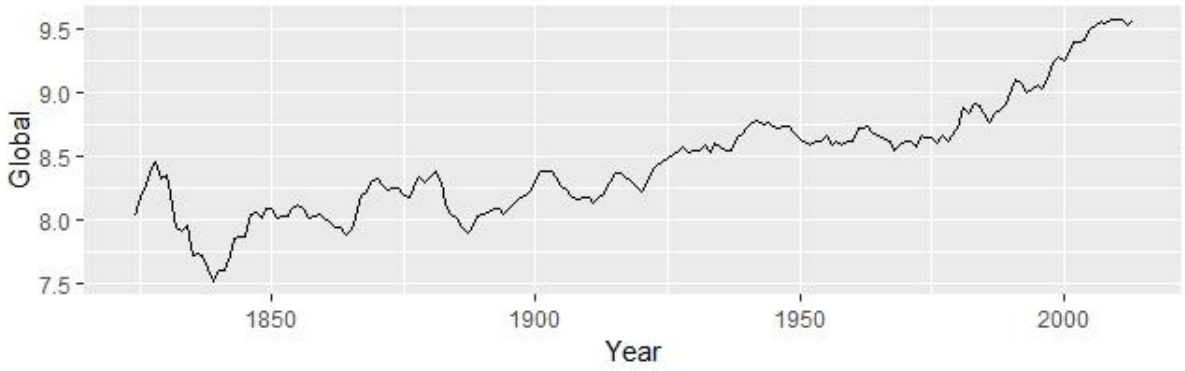

7 Years MA
Ahmadabad
27.5
27.0
26.5
26.0
25.5
1800
1850
1900
1950
2000
Year
Global
9.5
9.0
8.5
8.0
7.5
1800
1850
1900
1950
2000
Year


10 Years MA
Ahmadabad
27.5
27.0
26.5
26.0
25.5
1800
1850
1900
1950
2000
Year
Global
9.5
9.0
8.5
8.0
7.5
1800
1850
1900
1950
2000
Year

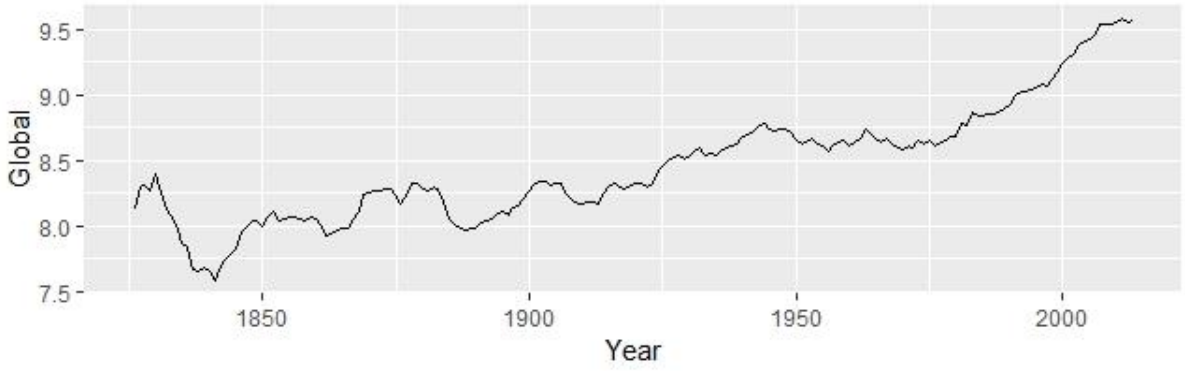

Correlation Coefficient between Global and Ahmadabad's Temperature
r = 0.79, p < 2.2e-16
Ahmadabad.Temp
28
27
26
25
7
8
9
Global.Temp

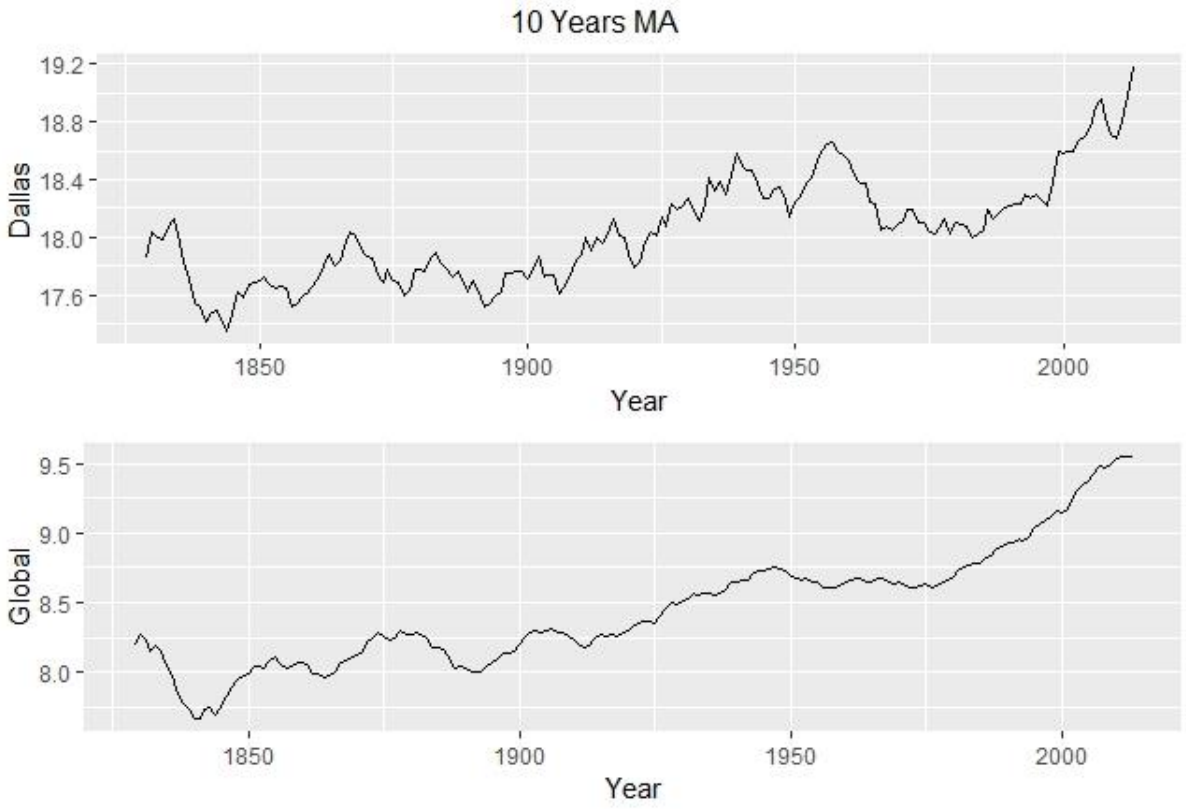
DALLAS

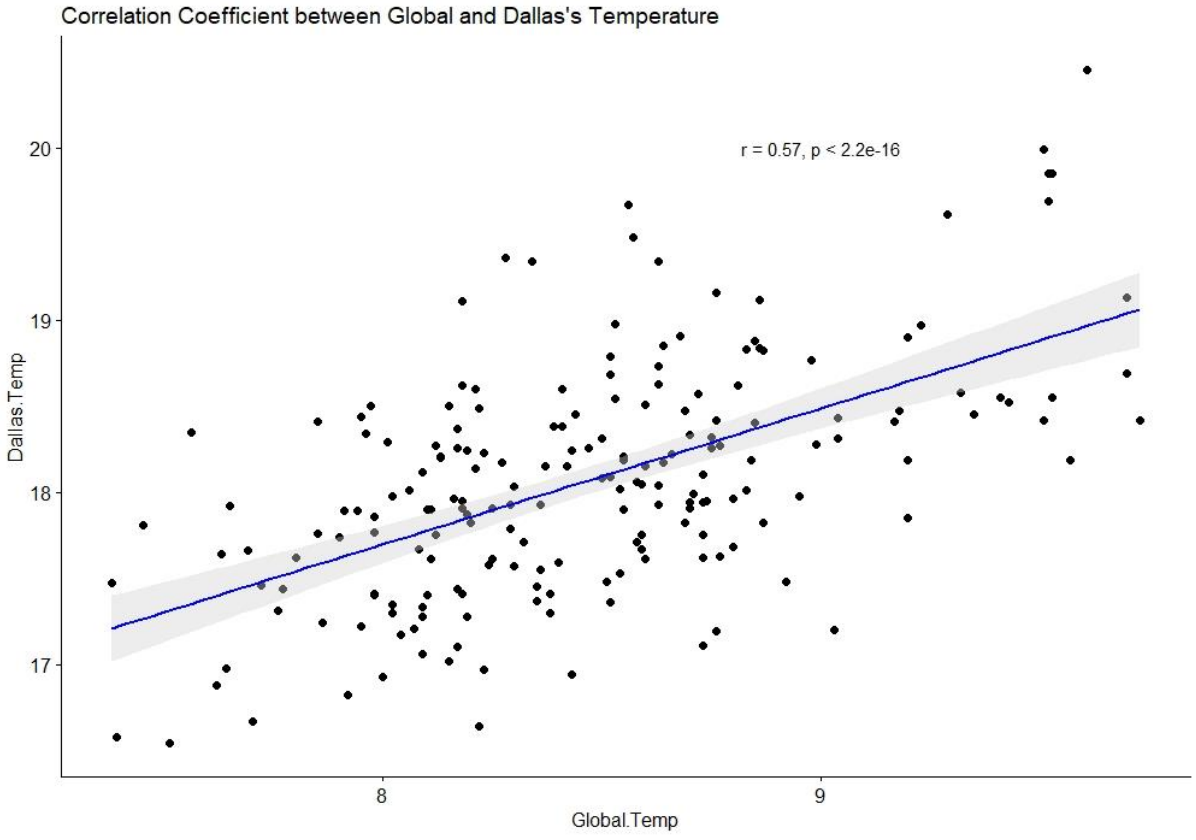

5 Years MA
Dallas
19.5
19.0
18.5
18.0
17.5
17.0
1850
1900
1950
2000
Year


Global
9.5
9.0
8.5
8.0
7.5
1850
1900
1950
2000
Year


7 Years MA
Dallas
19.0
18.5
18.0
17.5
1850
1900
1950
2000
Year


Global
9.5
9.0
8.5
8.0
7.5
1850
1900
1950
2000
Year


10 Years MA
Dallas
19.2
18.8
18.4
18.0
17.6
1850
1900
1950
2000
Year
Global
9.5
9.0
8.5
8.0
1850
1900
1950
2000
Year


Correlation Coefficient between Global and Dallas's Temperature
r = 0.57, p < 2.2e-16
Dallas.Temp
20
19
18
17
8
9
Global.Temp

Key Considerations

1. All the unrecorded years were removed from the data analysis for consistency.
2. For consistency and simplicity, the local and global temperature are taken from the same year and ending on the same year.
3. I have calculated 5,7 & 10 years Moving Average to get proper observations
4. I used the Pearson's Correlation Coefficient to measure the linear association between the Global temperature and the City Temperature.

3. Observations, Similarities and/or Differences

Ahmadabad

1. Ahmadabad's average temperature is greater than the global average. Since Ahmadabad lies near to the tropic of cancer, it is observed that all the cities of the world that lie between Tropic of Cancer and the Tropic of Capricorn have higher temperatures as compared to the global temperature.
2. There is a positive correlation (0.79) between the average temperature of the Earth and that of Ahmadabad.
3. The first observation noted was a significant drip in temperature that was seen in the year 1814 in Ahmadabad and globally. It happened due to a stratospheric volcanic eruption in the Philippines.
4. Again, a significant dip in the temperature is seen in the year 1883 when eruption of **Krakatoa Volcano** in Indonesia occurred. Every barograph around the world recorded the passage of the air wave, up to seven times. Volcanic ash was carried around the globe in thirteen days producing blue and green suns in the tropics, and then vivid red sunsets in higher latitudes. The temperature of the earth was lowered one degree for the next two years, finally recovering to normal by 1888.
5. The second observation is a cooling trend in the mid-20th century occurring between 1940 and 1975. After 1975 global warming continues.

Dallas

1. Dallas is comparatively cooler than Ahmadabad since it lies away from the tropic of cancer and the equator.
2. There is a positive correlation (0.57) between the average temperature of the Earth and that of Dallas.
3. The first observation noted was a significant drip in temperature that was seen in the year 1835 in Dallas and globally. It was due to a severe frost in America in that year.
4. Some effect of **Krakatoa Volcano** eruption on the temperature of Dallas was also observed.
5. While there is a cooling trend, some drastic changes in temperature as observed in Dallas between the year 1950 to 1975.

Conclusion

The global temperature record represents an average over the entire surface of the planet. The temperatures we experience locally can fluctuate significantly due to predictable cyclical events (night and day, summer and winter) and hard-to-predict wind and precipitation patterns. But the global temperature mainly depends on how much energy the planet receives from the Sun and how much it radiates back into space which depends significantly on the chemical composition of the atmosphere. In the global maps, the years from 1885 to 1945 tend to appear cooler, growing less cool as we move toward the 1950s. Hence, Global temperature is constantly rising and as time progresses the Earth gets warmer.

References

https://earthobservatory.nasa.gov/Features/WorldOfChange/decadaltemp.php

http://www.sthda.com/english/rpkgs/ggpubr/reference/stat_cor.html

http://www.breadandbutterscience.com/climatehistory.pdf