A COMPARISON OF WEATHER TRENDS IN BANGALORE AND THE GLOBE

S.KASHYAP SUBRAMANIAN

Project Goal: To study Temperature change in Bangalore and its relationship with Global Temperature change.

Project Data: The data was provided by **Udacity** for project purpose.

Tools Used: SQL, Excel, R (Programming Language)

Project Execution: Project execution involves the following steps.

1. Extraction
2. Data Wrangling
3. Visualization
4. Observation

I. EXTRACTION

Extraction involves extracting data from Database using SQL. The following QUERIES were used for data extraction.

1. CITY DATA:

```sql
SELECT *
FROM city_list;

```

Then from the list of cities ‘Bangalore’ was selected for project purposes. ‘Bangalore’ data was extracted using the below query.

```sql
SELECT *
FROM city_data
WHERE city='Bangalore'
ORDER BY year;

```

2. GLOBAL DATA:

```sql
SELECT *
FROM global_data
ORDER BY year;

```

Using the above queries, data was exported to CSV files.

II. DATA WRANGLING

The exported data contains empty/NUL L values for certain years in both global and city data. Those values are eliminated in order to properly analyze the dataset.

```
library(zoo)

city_data<-read.csv('bangalore_data.csv',header=TRUE)
global_data<-read.csv('global_data.csv',header=TRUE)

fcity_data<-city_data[complete.cases(city_data),]
fglobal_data<-global_data[complete.cases(global_data),]

fcity_data$ma_temp<-rep(0,211)
fcity_data$ma_temp[-1:-4]<-rollmean(fcity_data$avg_temp,k=5,align='right')

fglobal_data$ma_temp<-rep(0,266)
fglobal_data$ma_temp[-1:-4]<-rollmean(fglobal_data$avg_temp,k=5,align='right')
```

The above code is in R language. It imports data from CSV files. The entries that do not contain value or contain *NA* are eliminated by *complete.cases* function.

Moving Average: It is calculated using *rollmean* function with *k* value as 5 years. Column *'ma_temp'* holds the 5 year moving average. Shown below - CITY data, GLOBAL data

	year	city	country	avg_temp	ma_temp
1	1796	Bangalore	India	24.49	0.000
2	1797	Bangalore	India	25.18	0.000
3	1798	Bangalore	India	24.65	0.000
4	1799	Bangalore	India	24.81	0.000
5	1800	Bangalore	India	24.85	24.796
6	1801	Bangalore	India	24.49	24.796
7	1802	Bangalore	India	25.44	24.848
8	1803	Bangalore	India	25.22	24.962
9	1804	Bangalore	India	25.67	25.134
10	1805	Bangalore	India	25.01	25.166
11	1806	Bangalore	India	24.87	25.242
12	1807	Bangalore	India	24.25	25.004
18	1813	Bangalore	India	24.23	24.806
19	1814	Bangalore	India	23.91	24.454
20	1815	Bangalore	India	23.79	24.210
21	1816	Bangalore	India	23.30	23.896
22	1817	Bangalore	India	23.60	23.766
23	1818	Bangalore	India	23.94	23.708
24	1819	Bangalore	India	23.86	23.698
25	1820	Bangalore	India	23.91	23.722

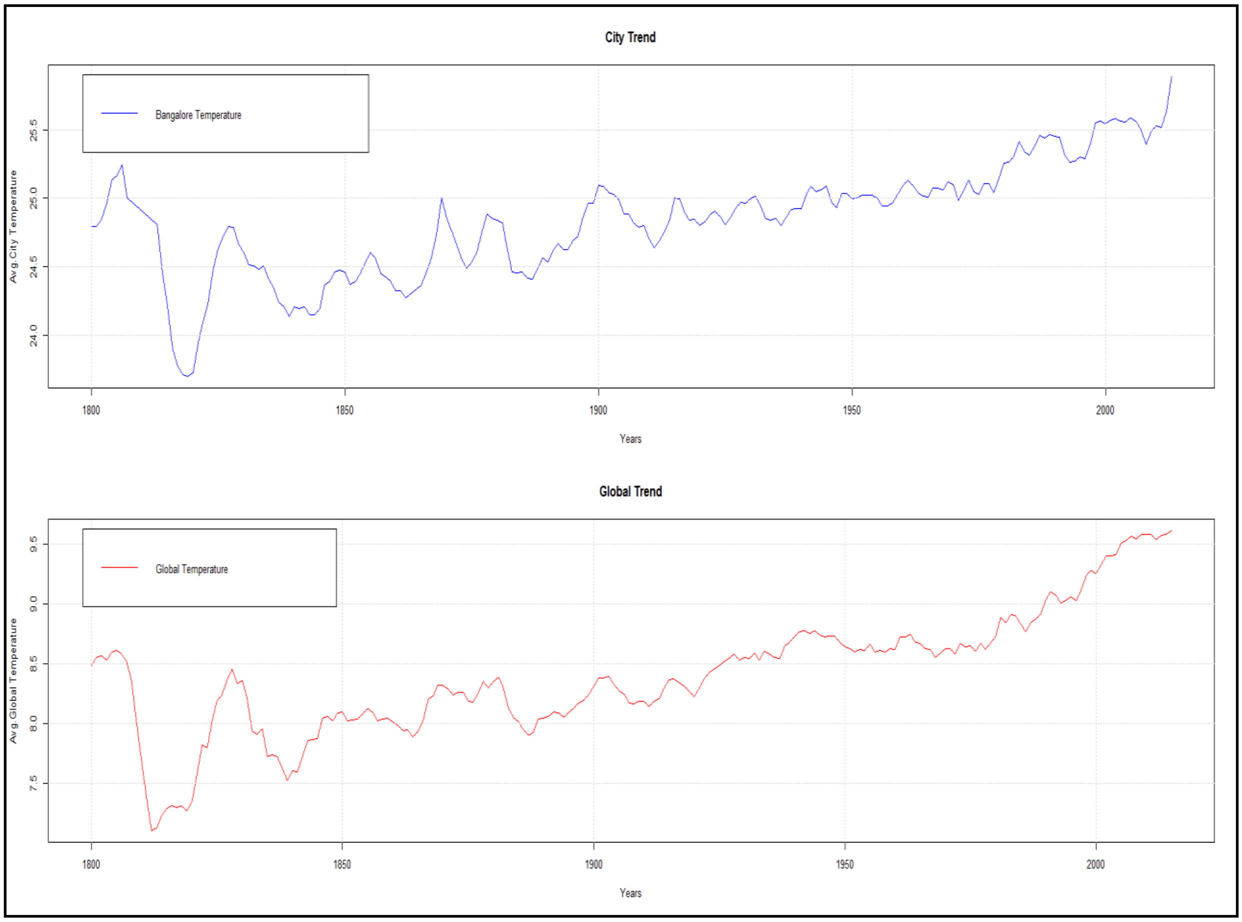
	year	avg_temp	ma_temp
1	1750	8.72	0.000
2	1751	7.98	0.000
3	1752	5.78	0.000
4	1753	8.39	0.000
5	1754	8.47	7.868
6	1755	8.36	7.796
7	1756	8.85	7.970
8	1757	9.02	8.618
9	1758	6.74	8.288
10	1759	7.99	8.192
11	1760	7.19	7.958
12	1761	8.77	7.942
13	1762	8.61	7.860
14	1763	7.50	8.012
15	1764	8.40	8.094
16	1765	8.25	8.306
17	1766	8.41	8.234
18	1767	8.22	8.156
19	1768	6.78	8.012
20	1769	7.69	7.870

III. VISUALIZATION

Now that data is ready, we can visualize data. For data comparison we consider data from the year 1980.

```
grid<-matrix(c(1,1,2,2),nrow=2,byrow=TRUE)
layout(grid)

# City Trend
plot(fcity_data$year[fcity_data$year>1799],fcity_data$ma_temp[fcity_data$year>1799],type='l',col='blue',
     main='City Trend',xlab='Years',ylab='Avg.City Temperature')
grid()
legend('topleft', inset=0.03, legend="Bangalore Temperature", col="blue", lty=1)

# Global Trend
plot(fglobal_data$year[fglobal_data$year>1799],fglobal_data$ma_temp[fglobal_data$year>1799],type='l',col='red',
main='Global Trend',xlab='Years',ylab='Avg.Global Temperature')
grid()
legend('topleft', inset=0.03, legend="Global Temperature", col="red", lty=1)
```

The above code plots 'Line Chart' for both Global and City data using 'Year' in x-axis and 'Moving Average Temperature' in y-axis. The plot looks like



From the graphs we identify that both City and Global temperatures are in the rising trend

Both the graphs show Non-Stationary behavior i.e. Mean, Variance & Covariance are not constant

In order to compare both Global and City in one graph, we can scale up Global data by a factor of 3(Approximation) and plot both data in one graph. The code is given below.

```
# Global vs City Trend

plot(fglobal_data$year[fglobal_data$year>1799],fglobal_data$ma_temp[fglobal_data$year>1799]*3,type='l',col='red',
main='Global vs Local Trend',xlab='Years',ylab='Avg.Temperature')
lines(fcity_data$year[fcity_data$year>1799],fcity_data$ma_temp[fcity_data$year>1799],col='blue')
grid()
legend('topleft', inset=0.03, legend=c("Global Temperature(scaled by factor of 3)", "Bangalore Temperature"),
       col=c("red", "blue"),lty=1:1)
```

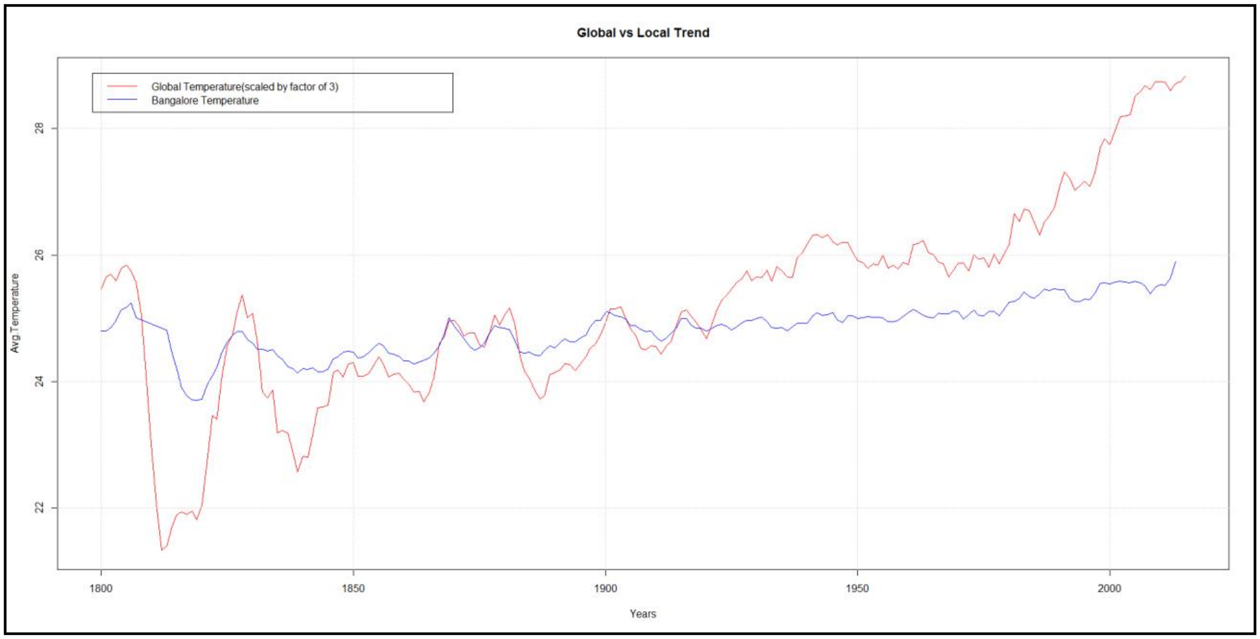
The function *plot* will plot the graph.

The function *lines* is used to plot another graph in the same plot.

The function *legend* adds the legend data in the plot.

The function *grid* will populate the grid in the plot. The required labels and colors are specified.

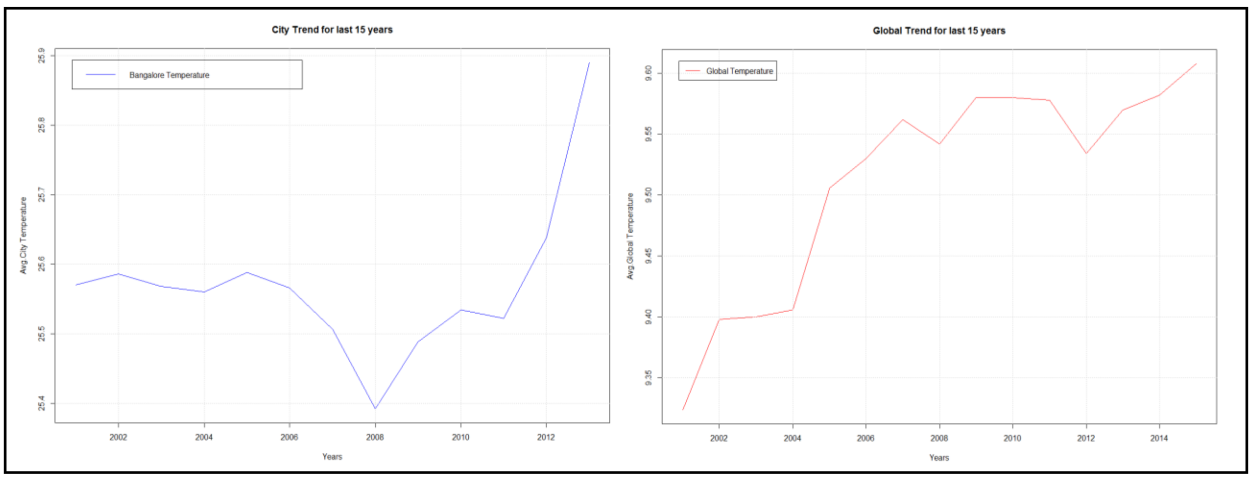
The plotted graph looks like below



The temperature data for last 15 years is plotted using the below code. We can trace how the temperature has changed for the last 15 years.

```
# Last 15 years Trend

grid<-matrix(c(1,2,1,2),nrow=2,byrow=TRUE)
layout(grid)

# City Trend - 15
plot(fcity_data$year[fcity_data$year>2000],fcity_data$ma_temp[fcity_data$year>2000],type='l',col='blue',
     main='City Trend for last 15 years',xlab='Years',ylab='Avg.City Temperature')
grid()
legend('topleft', inset=0.03, legend="Bangalore Temperature", col="blue", lty=1)

# Global Trend - 15
plot(fglobal_data$year[fglobal_data$year>2000],fglobal_data$ma_temp[fglobal_data$year>2000],type='l',col='red',
main='Global Trend for last 15 years',xlab='Years',ylab='Avg.Global Temperature')
grid()
legend('topleft', inset=0.03, legend="Global Temperature", col="red", lty=1)
```

The graph plotted is as below



There is a dip in temperature in 2008 in both the graphs. Also temperatures are increasing steeply after 2012 in both the graphs. The slopes of the graphs show steep increase in temperature. From the above two graphs we can state that

City temperature mirrors Global temperature

The temperature change for every 5 years is calculated i.e. temperature difference between current_year and current_year-5. This will give us an idea about how the temperature has changed over the years. The following code will calculate temperature difference.

```
# Temperature Difference

yrs<-seq(1800,2015,5)
temp<-0

for (yr in yrs){
current_temp<-fcity_data$avg_temp[fcity_data$year==yr]
pre_temp<-fcity_data$avg_temp[fcity_data$year==(yr-5)]
temp<-c(temp,(current_temp-pre_temp))
}
temp<-temp[-1]

city_temp_diff<-data.frame(Years=tail(yrs,40),Temp_diff=temp)


temp<-0

for (yr in yrs){
current_temp<-fglobal_data$avg_temp[fglobal_data$year==yr]
pre_temp<-fglobal_data$avg_temp[fglobal_data$year==(yr-5)]
temp<-c(temp,(current_temp-pre_temp))
}
temp<-temp[-1]

global_temp_diff<-data.frame(Years=tail(yrs,44),Temp_diff=temp)
```

The data in data frame looks like below – CITY data, GLOBAL data

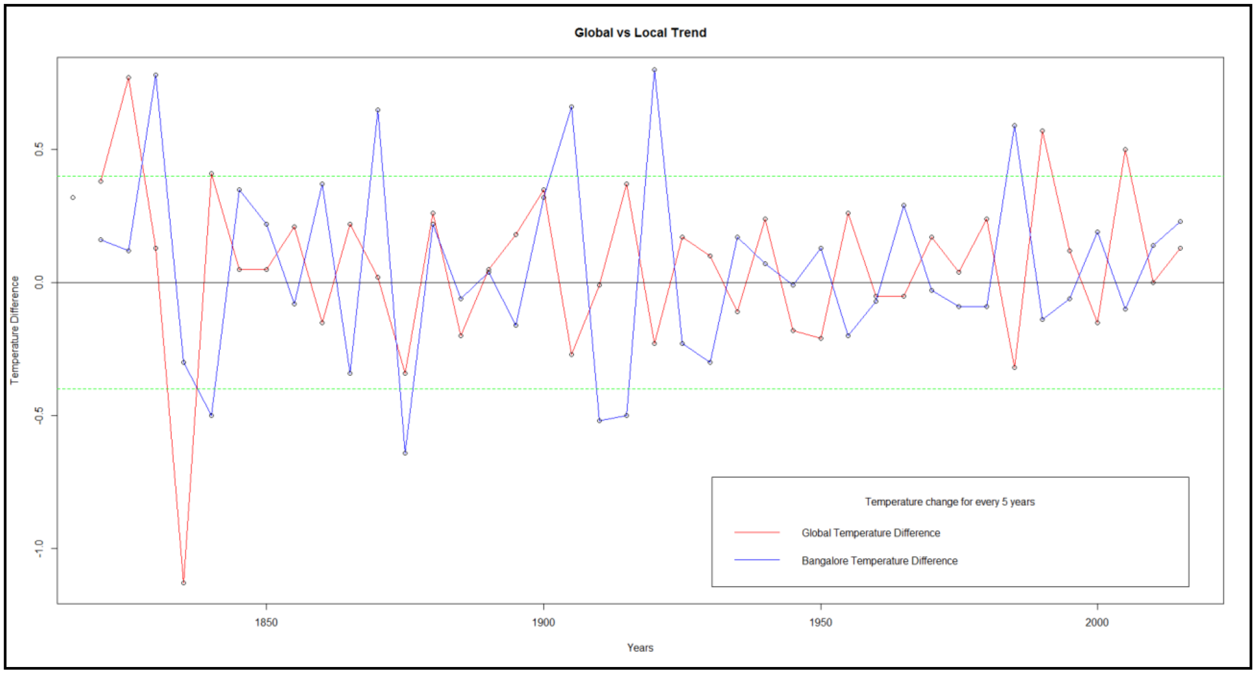
Years	Temp_diff
1820	0.16
1825	0.12
1830	0.78
1835	-0.30
1840	-0.50
1845	0.35
1850	0.22
1855	-0.08
1860	0.37
1865	-0.34

Years	Temp_diff
1800	0.13
1805	0.08
1810	-1.64
1815	0.32
1820	0.38
1825	0.77
1830	0.13
1835	-1.13
1840	0.41
1845	0.05
1850	0.05
1855	0.21
1860	-0.15
1865	0.22
1870	0.02

To plot this data we use the following code

```
plot(global_temp_diff[global_temp_diff$Years>1819,],type='l',col='red',
main='Global vs Local Trend',xlab='Years',ylab='Temperature Difference')
lines(global_temp_diff,type='p')
lines(city_temp_diff,type='l',col='blue')
lines(city_temp_diff,type='p')
abline(h=0)

legend('bottomright',inset=0.03,title="Temperature change for every 5 years", legend=c("Global Temperature Difference",
,col=c("red", "blue"),lty=1:1)

abline(h=0.5,lty=2,col='green')
abline(h=-0.5,lty=2,col='green')
```

The graph for the above code



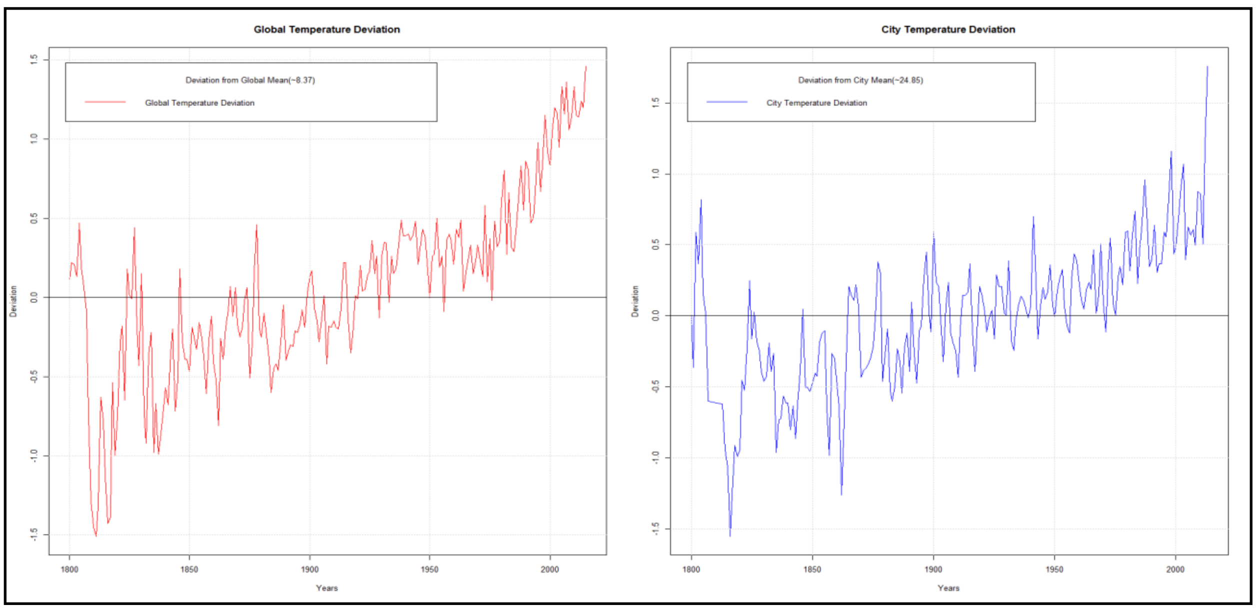
The variation in temperature for a 5-year period is within the range of (-0.4,0.4)

On observing the graph we can identify that the graph tends to move more towards the positive direction and the negative movement decreases after 1920. Also from 1800 till 1925, the magnitude of temperature difference in every five years is high compared the magnitude of temperature difference after 1925.

The magnitude of Temperature variation is less after 1925

The average mean temperature for Global data is ~8.37 and for Bangalore it is ~24.85. We can visualize how the temperatures have deviated from mean and plot them. The code and output is given below.

```
#Plot Deviation

grid<-matrix(c(1,2,1,2),nrow=2,byrow=TRUE)
layout(grid)

plot(global_deviation[global_deviation$Years>1799,],type='l',col='red',
     main='Global Temperature Deviation')
grid()
abline(h=0)
legend('topleft',inset=0.03,title="Deviation from Global Mean(~8.37)",
       legend="Global Temperature Deviation",col="red",lty=1)

plot(city_deviation[city_deviation$Years>1799,],type='l',col='blue',
     main='City Temperature Deviation')
grid()
abline(h=0)
legend('topleft',inset=0.03,title="Deviation from City Mean(~24.85)",
       legend="City Temperature Deviation",col="blue",lty=1)
```



From the temperature difference, we can calculate the total temperature change over the years

```
> global_data$avg_temp[global_data$year==2013] - global_data$avg_temp[global_data$year==1800]
[1] 1.13
>
> city_data$avg_temp[city_data$year==2013] - city_data$avg_temp[city_data$year==1800]
[1] 1.76
> global_data$avg_temp[global_data$year==2010] - global_data$avg_temp[global_data$year==1800]
[1] 1.22
>
> city_data$avg_temp[city_data$year==2010] - city_data$avg_temp[city_data$year==1800]
[1] 0.86
> city_data$avg_temp[city_data$year==2013] - city_data$avg_temp[city_data$year==2010]
[1] 0.9
```

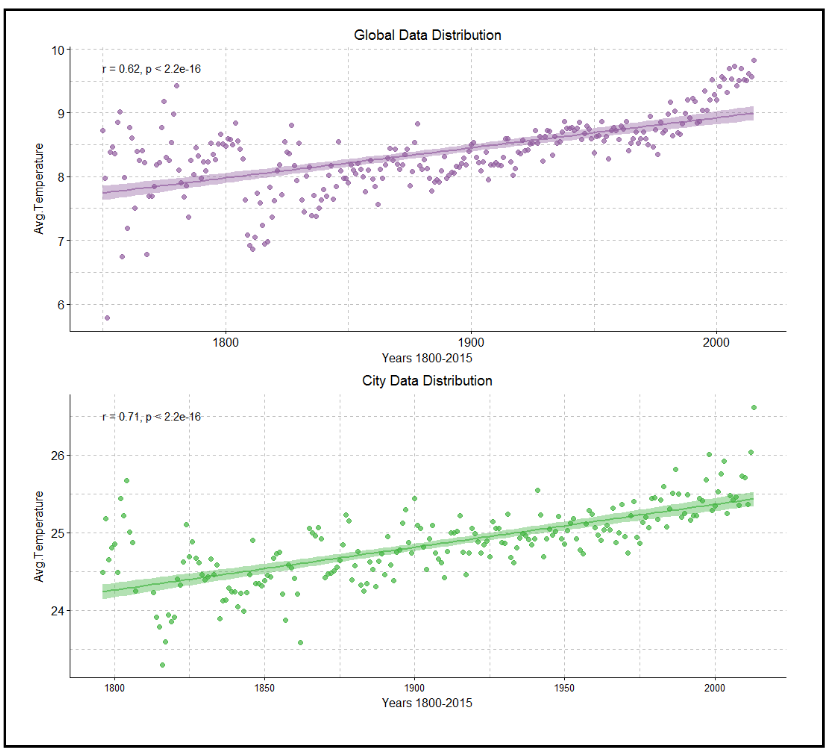
The global temperature has changed about 1.13 units and Bangalore temperature has changed about 1.76 units from 1800 - 2013. The city temperature has changed more after 2010.

The data distribution and correlation can also be identified by using scatter plot. *ggpubr* package is used for plotting correlation. The correlation is between *year* and *avg_temp* columns with data frame.

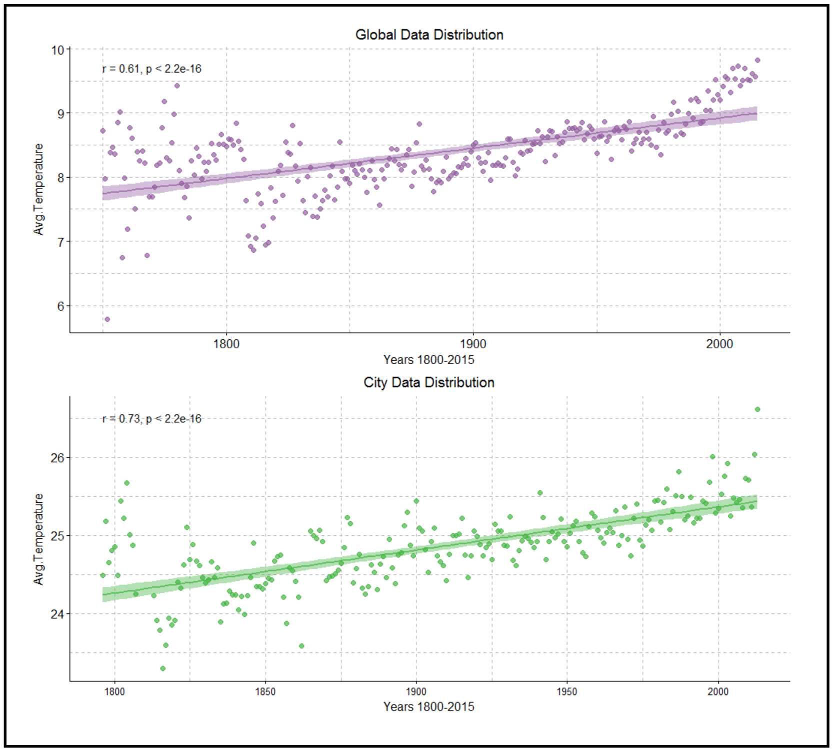
```
#------------------------------------------------------------------#
library(ggpubr)

mycol<-pal_lancet('lanonc',alpha=0.7)(9)
show_col(mycol)

gd<-ggscatter(fglobal_data, x = "year", y = "avg_temp",
      xlab='Years 1800-2015',ylab='Avg.Temperature',
      color="#925E9FB2",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "kendall")

gds<-gd+grids(color='grey',linetype='dashed')+
ggtitle("Global Data Distribution")+theme(plot.title = element_text(hjust = 0.5))


cd<-ggscatter(fcity_data, x = "year", y = "avg_temp",
      xlab='Years 1800-2015',ylab='Avg.Temperature',
      color="#42B540B2",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "kendall")

cds<-cd+grids(color='grey',linetype='dashed')+
ggtitle("City Data Distribution")+theme(plot.title = element_text(hjust = 0.5))

ggarrange(gds, cds + font("x.text", size = 10),
                    ncol = 1, nrow = 2)
```

By changing the *cor.method* attribute in ggscatter, we can plot Pearson, Spearman, Kendall correlation coefficients. The plots for the above 3 methods are below.

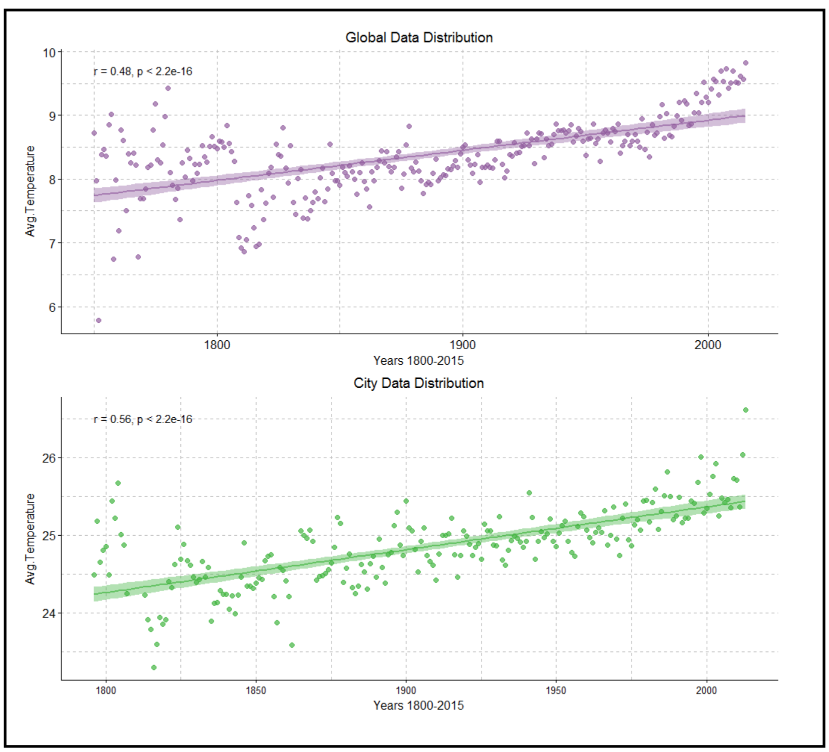
Pearson:



Spearman:



Kendall:



It is also interesting to find the correlation between global and local temperatures. The 3 methods of correlation are calculated and the outputs are as below. The data is considered from the year 1800 - 2013

Pearson:

```
> cor.test(global_data$avg_temp[(global_data$year>1799)&(global_data$year<2014)],
+          city_data$avg_temp[city_data$year>1799],
+          method = "pearson",use="complete.obs")

	Pearson's product-moment correlation

data:  global_data$avg_temp[(global_data$year > 1799) & (global_data$year <  and city_data$avg_temp[city_data$year > 1799]
014)] and city_data$avg_temp[city_data$year > 1799]
t = 24.885, df = 205, p-value < 2.2e-16
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.8283143 0.8971058
sample estimates:
      cor
0.8667757
```

Spearman:

```
> cor.test(global_data$avg_temp[(global_data$year>1799)&(global_data$year<2014)],
+          city_data$avg_temp[city_data$year>1799],
+          method = "spearman",use="complete.obs")

	Spearman's rank correlation rho

data:  global_data$avg_temp[(global_data$year > 1799) & (global_data$year <  and city_data$avg_temp[city_data$year > 1799]
014)] and city_data$avg_temp[city_data$year > 1799]
S = 217810, p-value < 2.2e-16
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.8526581
```

Kendall:

```
> cor.test(global_data$avg_temp[(global_data$year>1799)&(global_data$year<2014)],
+          city_data$avg_temp[city_data$year>1799],
+          method = "kendall",use="complete.obs")

	Kendall's rank correlation tau

data:  global_data$avg_temp[(global_data$year > 1799) & (global_data$year <  and city_data$avg_temp[city_data$year > 1799]
014)] and city_data$avg_temp[city_data$year > 1799]
z = 14.281, p-value < 2.2e-16
alternative hypothesis: true tau is not equal to 0
sample estimates:
      tau
0.6712594
```

From the above we can state that-

There is strong positive correlation between Global Temperature and Bangalore Temperature

IV. OBSERVATION

1. RISING TREND

   From the graphs we can easily identify that both global temperature and city temperature are increasing as time moves forward.

2. NON-STATIONARY

   Both Global and Local Temperature data are Non-Stationary i.e. Mean, Variance & Covariance are not constant over time.

3. MIRROR

   City Temperature change mirrors Global temperature change at many points. i.e. if there is rise in global temperature, city temperature tends to rise and if there is a fall in global temperature, city temperature decreases.

4. RANGE

   The temperature variation between 5 year periods lay in the range of -0.4 to 0.4 i.e. if temperature is recorded every five years, there is 88% probability for global temperature and 79% probability for city temperature that the rise or fall of temperature will lie within the range of -0.4 to 0.4.

5. LESS MAGNITUDE

   After 1925, the magnitude of temperature difference is less when compared to the century (1825-1924) before it. This means that temperature variation was drastic before 1925. But after 1925, the variation is comparatively less.

6. POSITIVE CORRELATION

   From the Pearson, Spearman and Kendall methods of calculation correlation, we identify that there is strong positive correlation between Global and City temperature i.e. when Global temperature increases, there is also an increase in city temperature and when global temperature decreases, there is also a decrease in city temperature. Correlation coefficient is ~0.85 between Global temperature and Bangalore temperature.