Name : Tania Moulik, Date : 9/18/17

PROJECT 1 : Explore Weather Trends

1. *Data extraction query using SQL*:

   SQL Query for city data extraction: (Local City : Chicago)
   SELECT year,avg_temp FROM city_data WHERE city='Chicago' AND country='United States'

   SQL Query for global data extraction
   SELECT year, avg_temp FROM global_data

   Then I clicked on Download csv to save it in a data file.

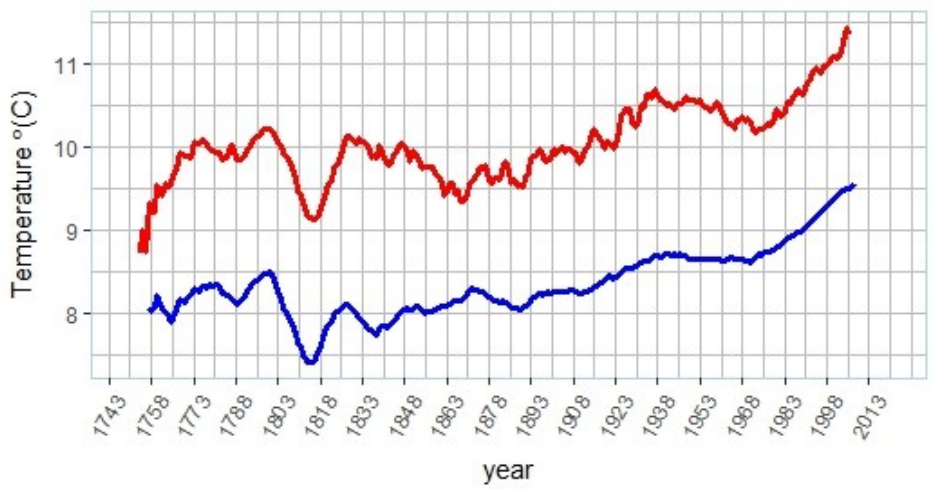
2. *Analysis* :

   I have used the R package for analyzing and visualizing the data. Some of the years had no temperature data, and these were dropped. The visualization is done using the package ggplot2.

   a. Line Chart of the moving Average.

      The moving mean is calculated as the average of temperatures over N consecutive years. The first data point takes a mean of the first N years, and the subsequent data points take the mean of the next consecutive years minus the first year, and so on. So the first data point would include 1 to N values, the 2nd, 2 to N+1, the 3rd : 3 to N+2 values and so on. I have used the **"rollmean"** function in the **"zoo"** package of R. I have also written my own function **"myrollmean"** and verified that I obtain the same results (A snippet of my function is included at the end).

      The red line shows the local city temperature, which in my case is Chicago, while the black line is the plot of the global city temperatures. The Y axis label shows the **median value** of the year interval over which the rolling mean is taken.



Local Temperature

Global Temperature

Fig. 1 : Line chart for the moving average of city and global temperatures in a 15 year window.

Some of the considerations in analyzing the trend is to see if there is a change of the temperatures over time i.e. are the temperatures rising, falling or constant? Are they trending or varying with time in the same way? Are there periods of stability where the temperatures are mostly constant? Are there some periods of extreme temperatures where the temperature fell or rose considerably with respect to the average temperature in that period?

Similarities:

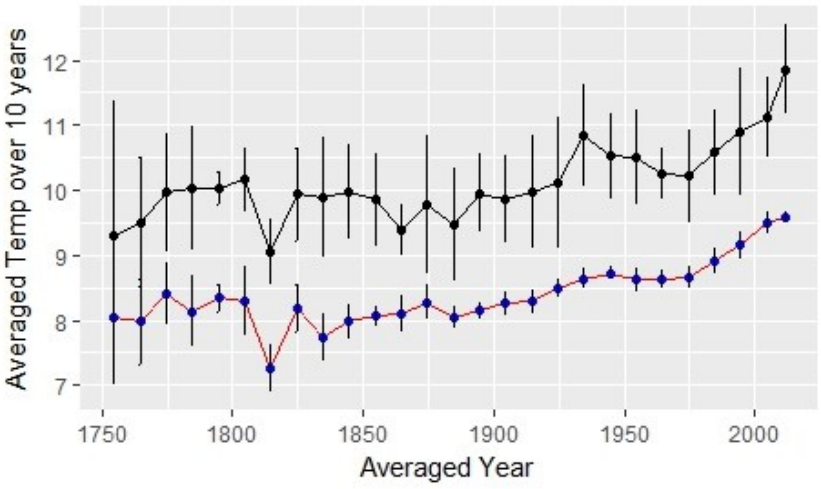
a) Both the global and city temperatures show a gradual increase after about the year 1890 with a very similar trend in the rise and fall of temperatures. The world is getting hotter!
b) Both show a similar general variation in terms of the rise and fall of the temperatures for the overall period plotted.
c) Both show a clear dip in temperature when the temperature fell considerably around 1800 and rose again around 1820. The minimum was in 1816 with a temperature of 9.1 ºC. The same year gives the minimum for the global data as well.

Differences:

d) The average city temperature is 10.0+-1.1 ºC. The average global temperature is 8.37+-0.58 ºC. The city temperatures are higher than the global temperatures by 1.67+-0.07, using error on mean as standard error/sqrt(N).
e) As can be seen from the standard error in Fig 2. the fluctuations in the city temperatures are higher than in the global temperatures.

Other Plots and Features:

1. I also did a simple average over 10-year periods. This is different from the rolling mean in that these are just averages over subsequent 10-year periods. In order to compare the global and the local temperature data, I included only the years which are common to both and removed some of the years which are not common in both data. This would allow me to histogram the difference in the temperatures of the two. The line chart along with error bars can be seen below:



City Temperature

Global Temperature

Fig. 2 : City temperatures and Global Temperatures vs Year

The histogram showing the difference in temperature can be seen below:

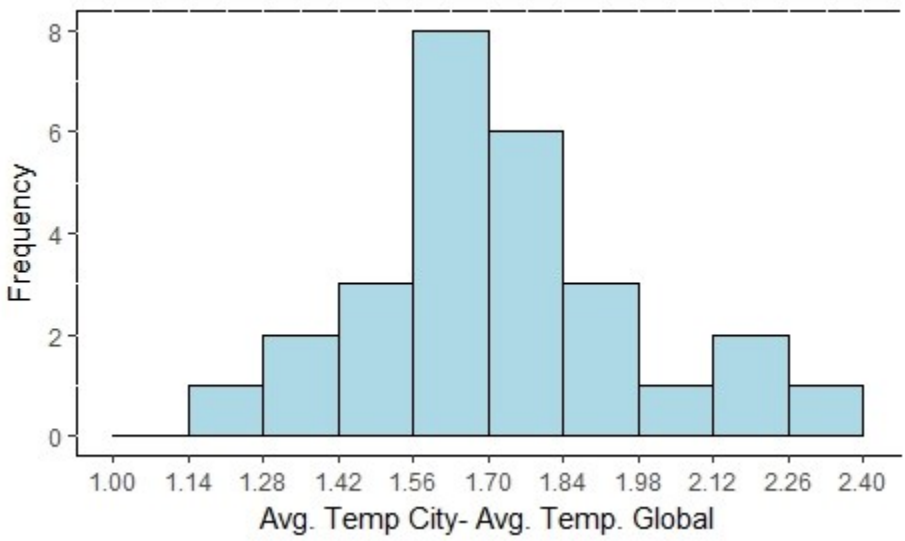
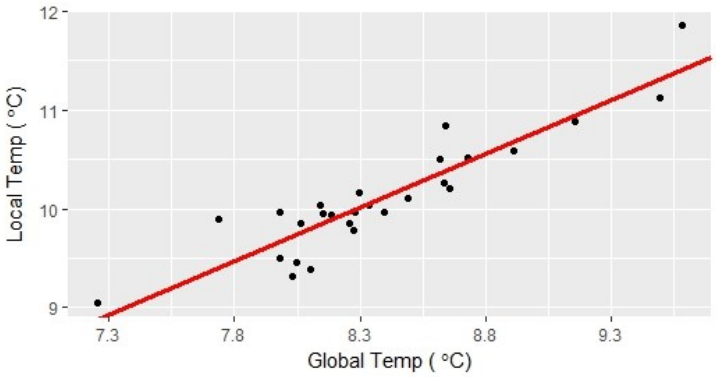
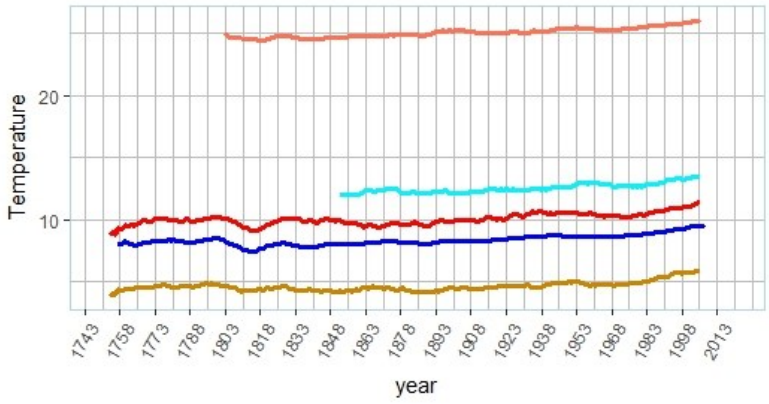
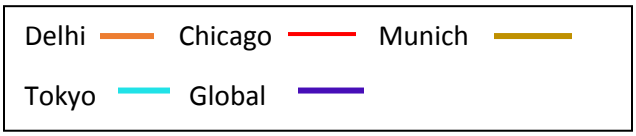


Fig 3. Difference of the 10-year averaged city and temperatures.

The scatter plot between the two temperatures can be seen below. As can be seen the temperatures are quite strongly correlated. The correlation coefficient is about 91%. One could do a linear fit to this data in order to obtain the average local temperature given the average global temperature. The fit yields the equation: y = 1.21*x-0.107, and from this plot one can "estimate" for example, around 8.5ºC of global temperature the local temperature is about 10.2 ºC.







Multiple Cities (Seen Above): I also plotted the rolling means (15-year periods) for some other cities such as, New Delhi, Munich and Tokyo. The data clearly shows that the world is getting hotter and there is a sharper temperature increase around 1960 which continues till date.

```
myrollmean <- function(v1,v2,n) {
#  v1 : Vector of avg_temp
#  v2 : Vector of year
#  n : Rolling mean window size
# The function return a data frame (d) containing the median year value and the
#rolling mean.

  start<-1
  max1=length(v1)
  d=NULL
  while (start<=max1-(n-1)) {
    mav <- mean(v1[start:(start+n-1)])
    yea <- median(v2[start:(start+n-1)])
    start <- start+1
    d = rbind(d,data.frame(yea,mav))
  }
  return(d)
}
```