Explore Weather Trends

Fienny Angelina

1 December 2018

- Initial Setup
    - Extract data from database
    - Install R packages needed in the project.
    - Load both the global and city data (Singapore data)
- Read the statistics from both plot
- Initial line plot
- Plot the moving average
    - Clean the data
- Is the increase the same for city and global?
- Does global temperature increase faster than city temperature?

Initial Setup

Extract data from database

Download the global data as csv from the course site, using the command below.

```
SELECT * from global_data;
```

Do the same thing for city data, in this case, I will be using Singapore Data

```
SELECT * from city_data where city_name='Singapore';
```

Install R packages needed in the project.

Install and load the essential R packages for Time Series analysis – `tseries` and `forecast`.

```
# install.packages("tseries")
# install.packages("forecast")
library(tseries)
library(forecast)
```

Load both the global and city data (Singapore data)

```
cityData <- read.csv("singapore_data.csv", header = TRUE)
globalData <- read.csv("global_data.csv", header = TRUE)
var <- c("year", "avg_temp")
cityData <- cityData[var]
```

The last 2 commands above are used to clean the city data by removing the column that are not needed (the city name and country name). Actually we can do it when we extract the data from the database.

Read the statistics from both plot

```
summary(cityData)
```

```
##       year         avg_temp
##  Min.   :1825   Min.   :25.47
##  1st Qu.:1872   1st Qu.:26.15
##  Median :1919   Median :26.55
##  Mean   :1919   Mean   :26.52
##  3rd Qu.:1966   3rd Qu.:26.88
##  Max.   :2013   Max.   :27.78
##                 NA's   :17
```

```
summary(globalData)
```

```
##       year         avg_temp
##  Min.   :1750   Min.   :5.780
##  1st Qu.:1816   1st Qu.:8.082
##  Median :1882   Median :8.375
##  Mean   :1882   Mean   :8.369
##  3rd Qu.:1949   3rd Qu.:8.707
##  Max.   :2015   Max.   :9.830
```

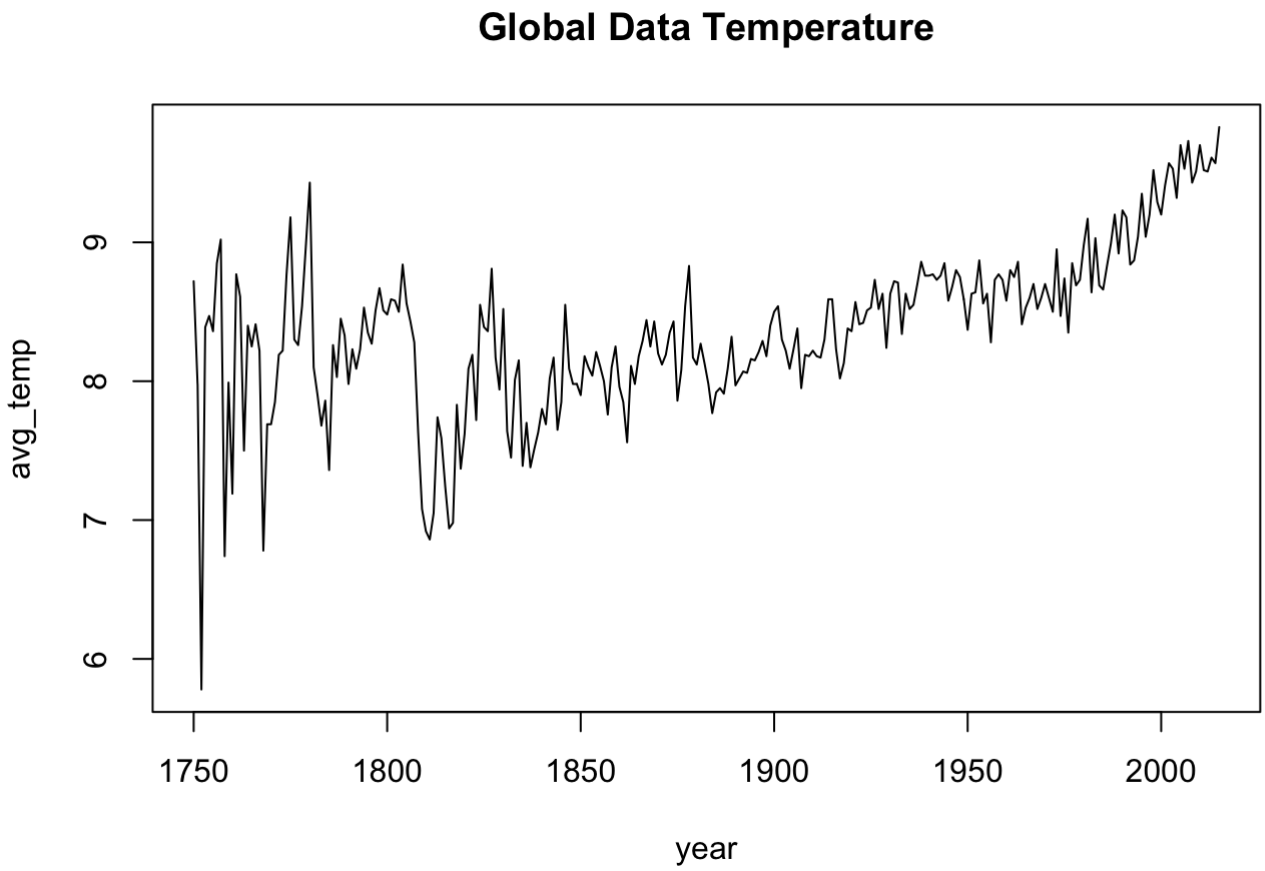
The first summary shows the city data while the second one shows the global data.

> From this point we can see that Singapore average temperature is always higher than global temperature, since the maximum global temperature has not reached Singapore temperature.

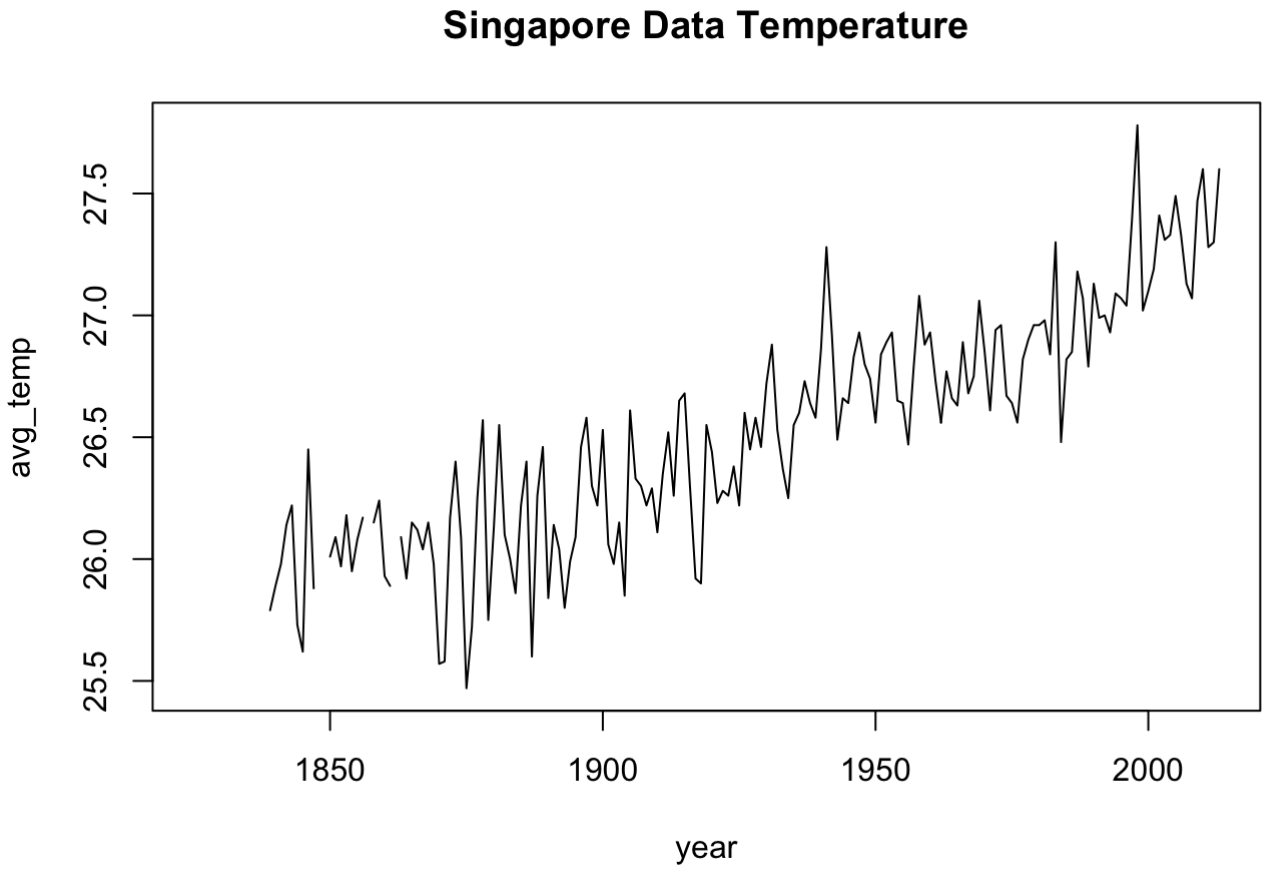
> We can also see that global average temperature has higher range than Singapore's temperature.

Initial line plot

```
plot(globalData, type="l", main="Global Data Temperature")
```




```
plot(cityData, type="l", main="Singapore Data Temperature")
```



From observing the above graph, there are a few interpretations I made: > In both data there seems to be an increase of the temperature throughout the years, especially for the city data.

> The maximum temperature of Singapore happens somewhere near 2000, yet the global temperature seems to happen at the end of the plot, in the year of 2015.

Plot the moving average

Clean the data

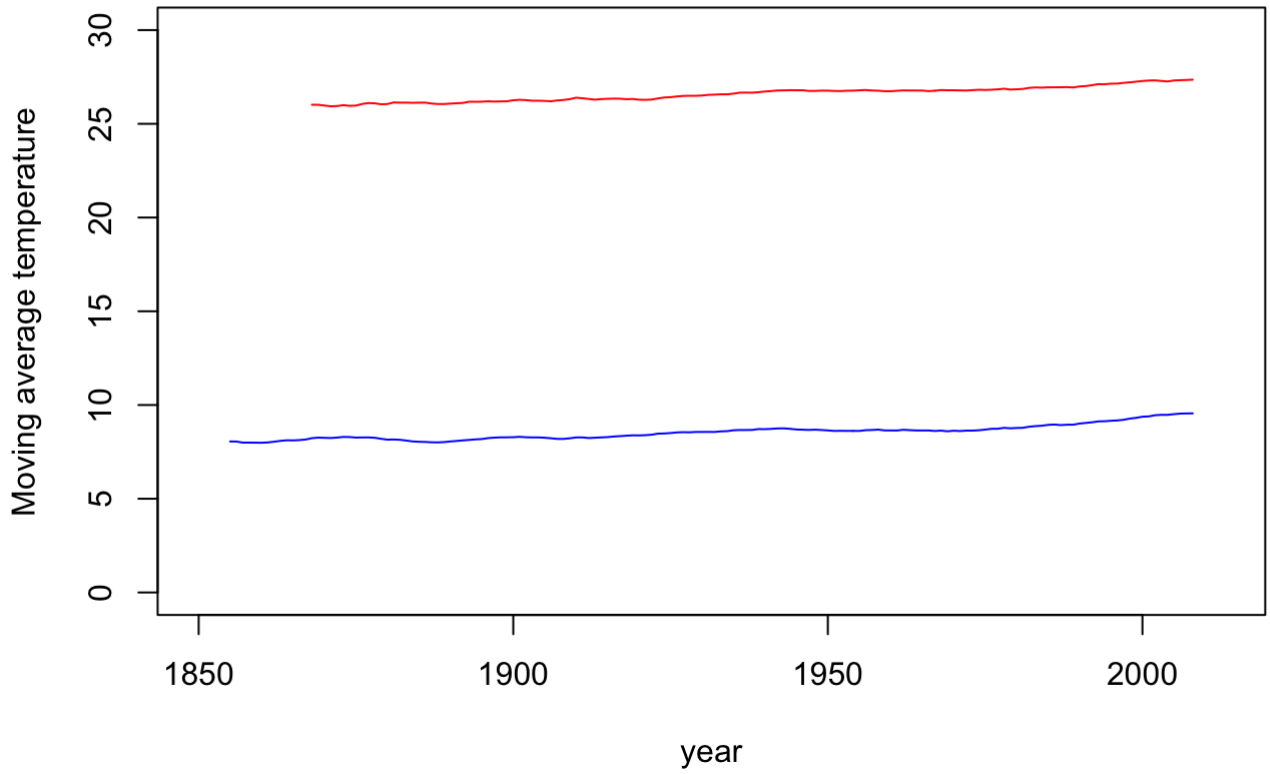
Before that, it is easy to see that Singapore Data only exists around 1850, to better measure both data equally, let's remove all data that occurs before 1850 from both dataset, also, we note that the last global data is dated 2013, yet the last city data is dated 2015, so let's remove everything after 2013.

```
globalData <- globalData[(globalData$year>=1850) & (globalData$year<=2013),]
cityData <- cityData[(cityData$year>=1850),]
```

Now try to create a moving average plot of both data with size of 11 (5 data on the left, the year's data, 5 data on the right)

```
f11 <- rep(1/11,11)
globalData$moving_avg_temp <- filter(globalData$avg_temp, f11, sides=2)
plot(globalData$year, globalData$moving_avg_temp, type="l", col="blue", xlab="y
ear", ylab="Moving average temperature", ylim=c(0, 30))

cityData$moving_avg_temp <- filter(cityData$avg_temp, f11, sides=2)
lines(cityData$year, cityData$moving_avg_temp, col="red", type="l")
```



As we can see from above, both city and global temperature has tendency to increase throughout the year.

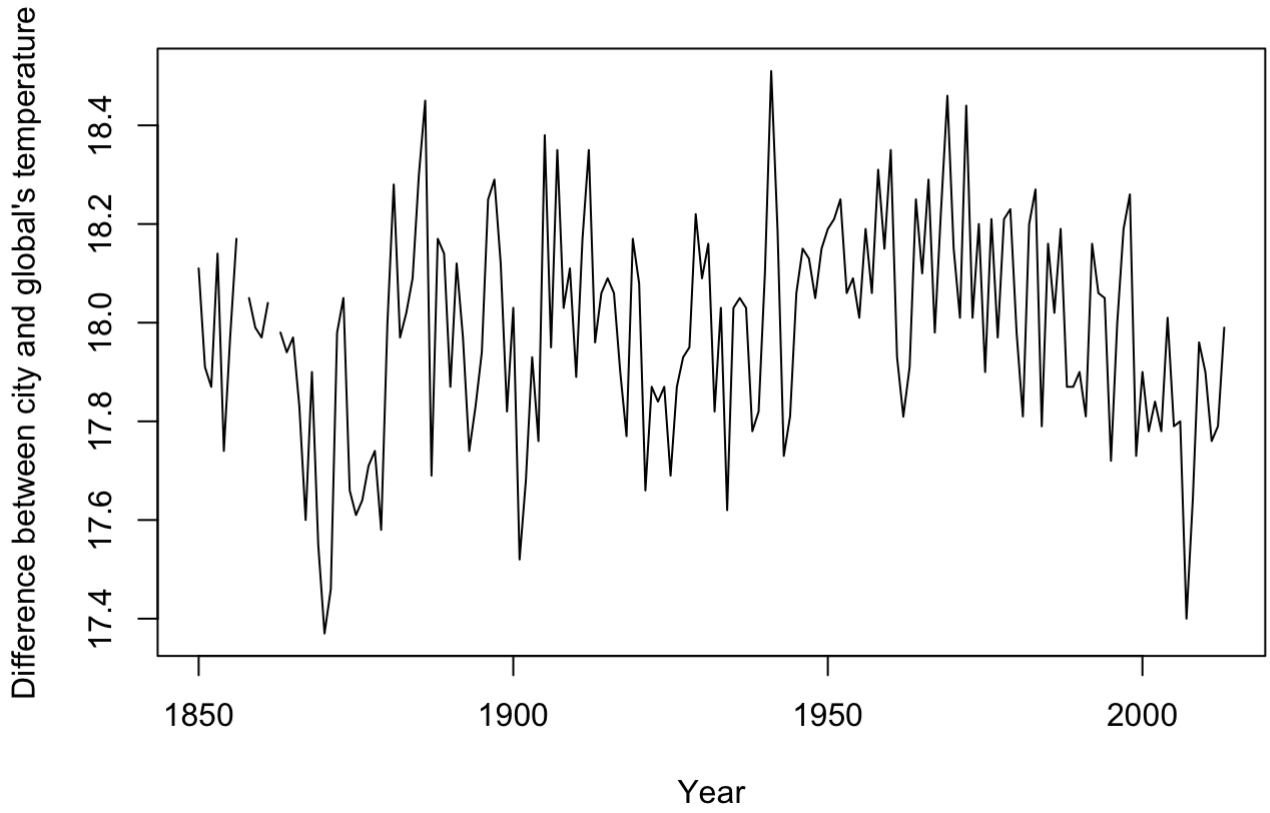
Is the increase the same for city and global?

Lets calculate the difference between cityData's temperature and globalData's.

```
diff <- cityData$avg_temp - globalData$avg_temp
```

Try to plot the data, it should be a nearly straight line if the increase is the same.

```
plot(globalData$year, diff, type="l", xlab="Year", ylab="Difference between cit
y and global's temperature")
```



From the plot above, it is hard to tell whether the data is actually increasing or decreasing. However we can see that the difference lies around 18 degrees each year. Let's try to get the summary of the difference to find its statistics

```
summary(diff)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   17.37   17.83   18.00   17.99   18.15   18.51       2
```

Let's try to smoothen the data using moving average to visualize the differences better.

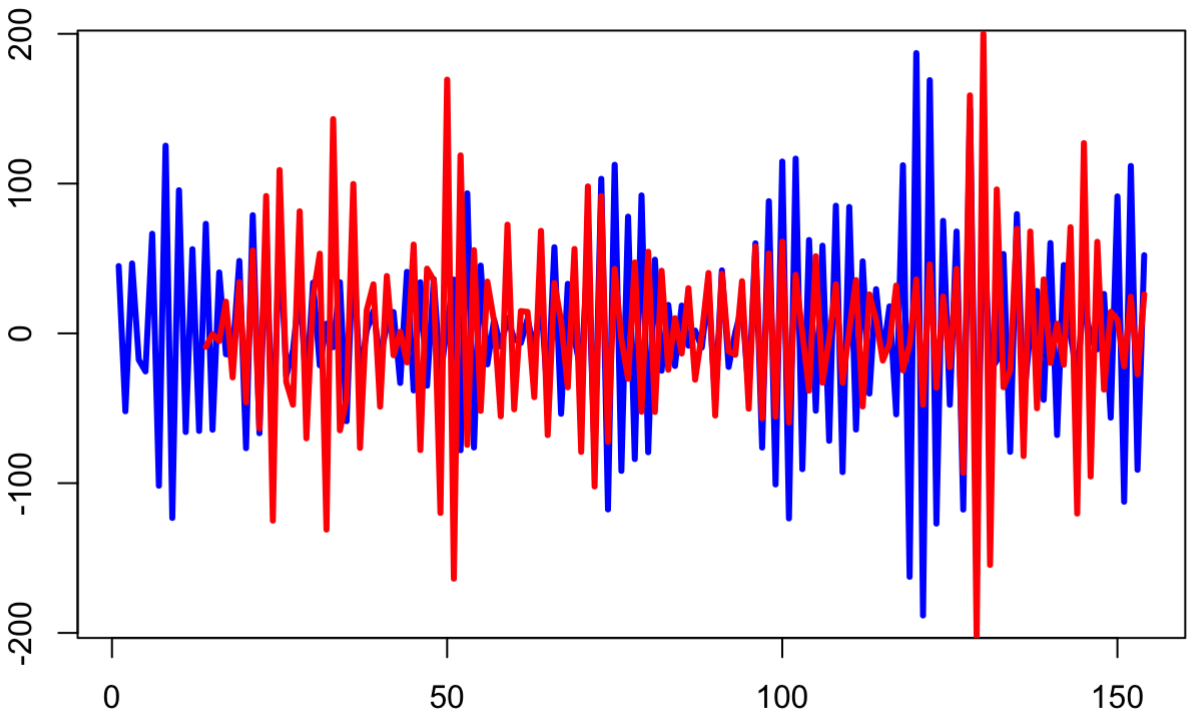
```
diff_moving_avg_temp <- filter(diff, f11, sides=2)
plot(globalData$year, diff_moving_avg_temp, type="l", xlab="Year", ylab="Differ
ence between city and global's temperature")
```

> From here, we can see that in the last few decades, the difference has tendency to decrease. However, it seems like both the city and global temperature has increased.

Does that mean global temperature has increased by more than city's since the difference is becoming smaller?

Does global temperature increase faster than city temperature?

```
plot(diff(globalData$avg_temp, differences= 10), type="l", col = "blue", lwd =
3, xlab = "", ylab = "")
lines(diff(cityData$avg_temp, differences=10), col="red", lwd = 3)
```



The red plot marks the city temperature difference, which is smaller than the global temperature. Thus, we can roughly conclude that global temperature increase faster than city temperature.