Exploring Weather Trends

Indira Yarullina

For the graphical representation of results I use R.

Upload the datasets

Setting working directory, required libraries, uploading required csv.

```
setwd("~/Desktop/Udacity/Project_1/")
library(readr)
city_data <- read_csv("city_data.csv")
```

```
## Parsed with column specification:
## cols(
##   year = col_integer(),
##   city = col_character(),
##   country = col_character(),
##   avg_temp = col_double()
## )
```

```
city_list<-read_csv("city_list.csv")
```

```
## Parsed with column specification:
## cols(
##   city = col_character(),
##   country = col_character()
## )
```

```
global_data<-read_csv("global_data.csv")
```

```
## Parsed with column specification:
## cols(
##   year = col_integer(),
##   avg_temp = col_double()
## )
```

Moving average and graphical representation.

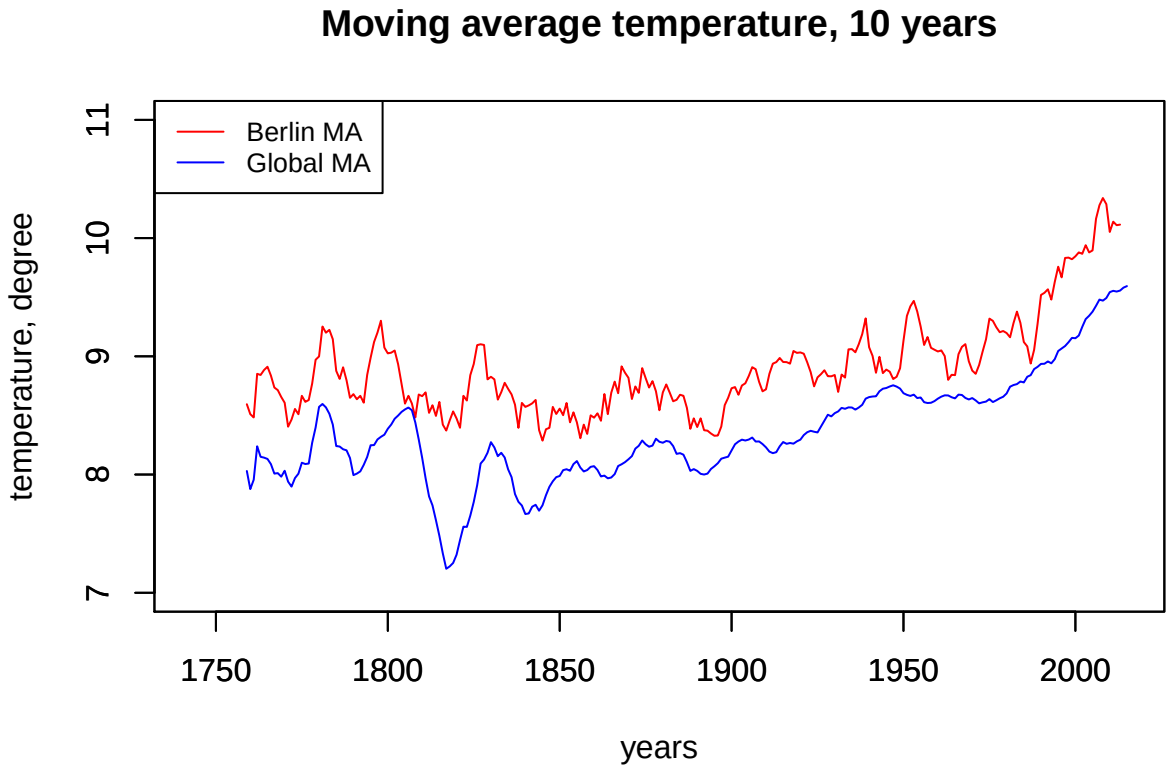
Since I am from Germany, and city_list contains only three German cities: Berlin, Hamburg and Munich, the closest one to me is Berlin.

```
city_list[city_list$country=="Germany",]
```

```
## # A tibble: 3 x 2
##      city country
##     <chr>   <chr>
## 1  Berlin Germany
## 2 Hamburg Germany
## 3  Munich Germany
```

```
#I make subset of city_data only for the Berlin
Berlin_data<-subset(city_data, city_data$city=="Berlin")
```

```r
#calculating manually the mooving average (MA) for Berlin and for the Global data
Berlin_data$MA<-NA

#creating function for mooving average
mav <- function(x,n){filter(x,rep(1/n,n), sides=1)}
Berlin_data$MA<-mav(Berlin_data$avg_temp, n=10)# Berlin, 10 years MA

global_data$MA_glob<-mav(global_data$avg_temp, n=10)

#I want to combime both mooving aaverage in order to have correct dates,
#therefore I will add to the Berlin_data information from global_data

Ber_glob<-merge(Berlin_data, global_data, by.x = "year", by.y = "year",
                all = TRUE)[,c("year","MA", "MA_glob")]

#plot the results. We plot both variables together in order to see similarities
#and differences for berlin and global together. For distinguish we use different
#colors of lines, that also will be explained in legend.

plot(Ber_glob$year,Ber_glob$MA,xlab="",ylab="",pch=0,col="red", type = "l", ylim = c(7,11))
par(new=TRUE)
plot(Ber_glob$year,Ber_glob$MA_glob,xlab="years",ylab="temperature, degree",
     main = "Moving average temperature, 10 years", col="blue", type = "l", ylim = c(7,11))
# Add a legend
legend("topleft",legend=c("Berlin MA", "Global MA"),
       col=c("red", "blue"), lty=1, cex=0.8)
```



The graph represent the combained moving averages for Berlin and global temperature changes.

-The average temperature in Berlin is higher compare the global average (approximately less than one degree

difference) and it is consistent over time.

-Moving average for Berlin more or less repeat the global trend: same fluctuating trend in short periods and increasing for the whole period.

-The overall trend for both is the upward slope - the average temperature increased in both cases approximately by 1,5 degree.

Additional analysis

There are bunch of additional analitical tools that we can apply for our data.

First of all, visual representation tells us about the possible correlation between the Berlin and Global temperatures. We can calculate the correlation coefficient.

```
cor(Ber_glob$MA, Ber_glob$MA_glob, use = "complete.obs")
```

```
## [1] 0.8576165
```

```
# I added use = "complete.obs" because in our data exist some missing variables.
```

As we see from calculation, there exist strong correlation between two variables. The results are not surprising, since Berlin create part of the global temperature.