Project N°1:
Exploring Weather Trends

Jaime Auger Esterio

Index

1. Data Extraction 3

SQL extraction 3

Reading of the CSV file 3

2. Data Visualization 4

Moving averages 4

Line chart 4

3. Observations 5

4. Bonus – Mobile Averages Variations 6

5. References 7

1. Data Extraction

SQL extraction

Two queries were made in SQL format. The first with the aim of knowing the closest cities registered in the database provided. The cities from Chile registered were Santiago, the capital of Chile, and Los Ángeles; between these, the nearest city was Santiago. Instead, the second query aimed to extract the information on global and local temperatures.

```
1   SELECT * FROM city_list WHERE city_list.country='Chile';
2
3   SELECT city_data.year as year, city_data.avg_temp as
    santiago_avg_temp, global_data.avg_temp as global_avg_temp FROM
    (city_data JOIN global_data ON city_data.year=global_data.year) WHERE
    city_data.city='Santiago';
4
```

Reading of the CSV file

The chosen tool to read and explore the data was the IDE Rstudio[1]. In first instance the data received was not the same quantity between the global and the local data because there were some maximum temperatures registered. So, to leave only the annual averages, it was used the “filter” function from the library “dplyr”. Then it was applied the “summary” function to know if there is any missing value but was not the case, there wasn´t any null value.

[1] It was chosen because it is an individual analysis of the data. Instead, if is necessary to use any other web applications or must upload the data to a server, it is more suitable to use Python because of its integrability advantages. (Rochina, 2016)

2. Data Visualization

Moving averages

There were calculated two mobile averages for each temperature, one of 5 years, and another of 10 years.

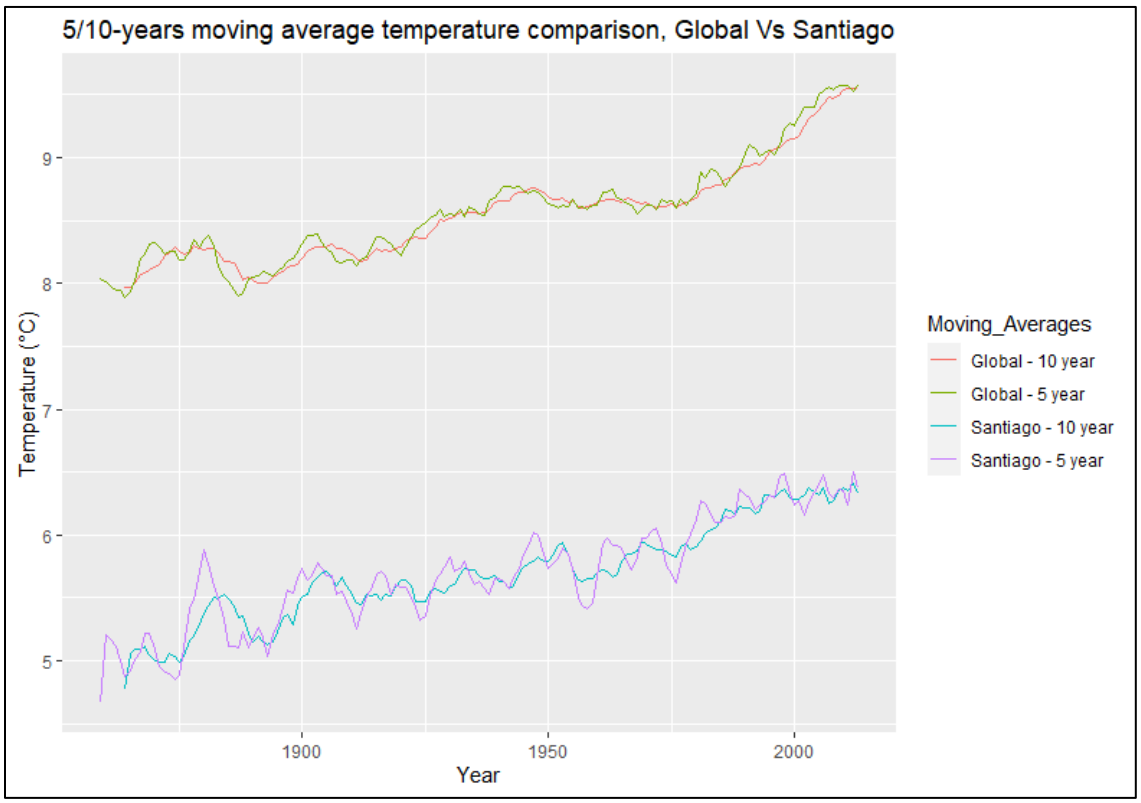
```
x<- rep(0,length(data_set$santiago_avg_temp))
y<- rep(0,length(data_set$santiago_avg_temp))
w<- rep(0,length(data_set$global_avg_temp))
z<- rep(0,length(data_set$global_avg_temp))

for (r in 1:length(data_set$santiago_avg_temp)){
  if (r>=5){
    x[r] <- (data_set$santiago_avg_temp[r] + data_set$santiago_avg_temp[r-1] + data_set$santiago_avg_temp[r-2]
             + data_set$santiago_avg_temp[r-3] + data_set$santiago_avg_temp[r-4])/5
    w[r] <- (data_set$global_avg_temp[r] + data_set$global_avg_temp[r-1] + data_set$global_avg_temp[r-2]
             + data_set$global_avg_temp[r-3] + data_set$global_avg_temp[r-4])/5
  }
  if (r>=10){
    y[r] <- (data_set$santiago_avg_temp[r] + data_set$santiago_avg_temp[r-1] + data_set$santiago_avg_temp[r-2]
             + data_set$santiago_avg_temp[r-3] + data_set$santiago_avg_temp[r-4] + data_set$santiago_avg_temp[r-5]
             + data_set$santiago_avg_temp[r-6] + data_set$santiago_avg_temp[r-7] + data_set$santiago_avg_temp[r-8]
             + data_set$santiago_avg_temp[r-9])/10
    z[r] <- (data_set$global_avg_temp[r] + data_set$global_avg_temp[r-1] + data_set$global_avg_temp[r-2]
             + data_set$global_avg_temp[r-3] + data_set$global_avg_temp[r-4] + data_set$global_avg_temp[r-5]
             + data_set$global_avg_temp[r-6] + data_set$global_avg_temp[r-7] + data_set$global_avg_temp[r-8]
             + data_set$global_avg_temp[r-9])/10
  }
}

data_set <- mutate(data_set, ma5_santiago = x)
data_set <- mutate(data_set, ma10_santiago = y)
data_set <- mutate(data_set, ma5_global = w)
data_set <- mutate(data_set, ma10_global = z)
```

Line chart

The line chart was made using the “ggplot2” library with the help provided by Statistic Globe blog (Schork, s.f.). The filter function was used again to remove all the first values lost in each mobile average because there is no need in their visualization.



3. Observations

- From the line graph, we can conclude that the temperature of the city of Santiago de Chile is, on average, 2,86 degrees cooler than the global temperature. It´s minor difference shows in the period 1900-1910 with a variance between 2,59 and 2,72.

- Temperature variations in the city of Santiago reach greater values than in global ones throughout the displayed time interval. The maximum variation in the 10-year mobile average for the local temperature was 0,29 degrees, instead of 0,07 degrees for the global one.

- Both Santiago de Chile and the world show, according to their 10-year moving averages, an increasing trend in their temperatures. Although there have been short periods of slight decreases in temperature (1886 - 1892) or stability (1900 - 1925 and 1950 - 1963), viewing the entire interval can be seen that the general trend tends to be positive or at higher temperatures.

- Between 1975 and 2000, there have been, on average, the greatest increases in temperatures both in Santiago (5,83°C – 6,28°C) and in the rest of the world (8,64°C – 9,16).

4. Bonus – Mobile Averages Variations

Like an extra analysis, there were calculated the variations of each mobile average and know how many of these variations are in the same way. The results were that the 61,39% of the variations in the 5-year mobile average were in the same way in the temperatures of Santiago and in the rest of the world. In the case of the 10-year mobile average, the 58,86% of the variations were in the same way in both locations.

```
data_set_diff <- data.frame(data_set$year[2:159],
                            diff_ma5_stgo=diff(data_set$ma5_santiago),
                            diff_ma10_stgo=diff(data_set$ma10_santiago),
                            diff_ma5_global=diff(data_set$ma5_global),
                            diff_ma10_global=diff(data_set$ma10_global))

m<- rep(0,length(nrow(data_set_diff)))
n<- rep(0,length(nrow(data_set_diff)))
for (i in 1:length(data_set_diff$diff_ma5_global)){
  if (data_set_diff$diff_ma5_stgo[i]*data_set_diff$diff_ma5_global[i] > 0){
    m[i] <- TRUE
  }
  if (data_set_diff$diff_ma5_stgo[i]*data_set_diff$diff_ma5_global[i] <= 0){
    m[i] <- FALSE
  }
  if (data_set_diff$diff_ma10_stgo[i]*data_set_diff$diff_ma10_global[i] > 0){
    n[i] <- TRUE
  }
  if (data_set_diff$diff_ma10_stgo[i]*data_set_diff$diff_ma10_global[i] <= 0){
    n[i] <- FALSE
  }
}
data_set_diff <- mutate(data_set_diff, diff_5 = m)
data_set_diff <- mutate(data_set_diff, diff_10 = n)
summary(data_set_diff)
```

```
    diff_5          diff_10
Min.   :0.0000   Min.   :0.0000
1st Qu.:0.0000   1st Qu.:0.0000
Median :1.0000   Median :1.0000
Mean   :0.6139   Mean   :0.5886
3rd Qu.:1.0000   3rd Qu.:1.0000
Max.   :1.0000   Max.   :1.0000
```

5. References

Rochina, P., 2016. *Revista Digital INESEM.* [En línea] Available at: https://revistadigital.inesem.es/informatica-y-tics/python-r-analisis-datos/ [Último acceso: 25 October 2020].

Schork, J., s.f. *Statistics Globe.* [En línea] Available at: https://statisticsglobe.com/plot-all-columns-of-data-frame-in-r [Último acceso: 24 October 2020].