

Exploring Weather Trends in Cairo

9/6/2020

Abstract

In this summary report we will explore the weather trend locally, which is Cairo in this case, and globally. We will use a moving average to compare the two trends, and will analyze the different observations after visualization in a line chart.

Getting the data:

Data were extracted from the temperatures database using Udacity's workspace and downloaded as a csv file. The following SQL queries were used:

```
-- Exploring the tables in the database
SELECT * FROM city_data
LIMIT 10;
SELECT * FROM city_list
LIMIT 10;
SELECT * FROM global_data
LIMIT 10;

-- Extracting data about Cairo and global temperatures
SELECT c.year, c.city, c.avg_temp AS avg_temp_city, g.avg_temp AS
avg_temp_global
FROM city_data c
JOIN global_data g
ON c.year = g.year
WHERE c.city = 'Cairo';
```

Analyzing the data:

Analysis of data was done in R and was performed in RStudio:

- The csv file was imported using the `read.csv()` function provided by the `readr` package.

- 5, 10, and 20 years moving averages were calculated for both Cairo and global average temperatures using the `rollmean()` function provided by the `zoo` package.

```
library(readr)
library(tidyr)
library(dplyr)
library(ggplot2)
library(zoo)

#importing the data set
cairo <- read.csv("cairo_results.csv")
glimpse(cairo)

## Rows: 206
## Columns: 4
## $ year      <int> 1808, 1809, 1810, 1811, 1812, 1813, 1814, 1815, 181...
## $ city      <chr> "Cairo", "Cairo", "Cairo", "Cairo", "Cairo", "Cairo..."
## $ avg_temp_city <dbl> 17.11, 19.87, 19.93, 20.00, 19.93, 20.51, 20.43, 20...
## $ avg_temp_global <dbl> 7.63, 7.08, 6.92, 6.86, 7.05, 7.74, 7.59, 7.24, 6.9...

#calculating 5,10, and 20 year moving averages using the rollmean function
#provided by zoo package
cairo_ma <- cairo %>%
  mutate(cairo_05 = rollmean(avg_temp_city, 5, fill = NA ),
         cairo_10 = rollmean(avg_temp_city, 10, fill = NA ),
         cairo_20 = rollmean(avg_temp_city, 20, fill = NA ),
         global_05 = rollmean(avg_temp_global, 5, fill = NA ),
         global_10 = rollmean(avg_temp_global, 10, fill = NA ),
         global_20 = rollmean(avg_temp_global, 20, fill = NA ))
glimpse(cairo_ma)

## Rows: 206
## Columns: 10
## $ year      <int> 1808, 1809, 1810, 1811, 1812, 1813, 1814, 1815, 181...
## $ city      <chr> "Cairo", "Cairo", "Cairo", "Cairo", "Cairo", "Cairo..."
## $ avg_temp_city <dbl> 17.11, 19.87, 19.93, 20.00, 19.93, 20.51, 20.43, 20...
## $ avg_temp_global <dbl> 7.63, 7.08, 6.92, 6.86, 7.05, 7.74, 7.59, 7.24, 6.9...
## $ cairo_05      <dbl> NA, NA, 19.368, 20.048, 20.160, 20.234, 20.336, 20....
## $ cairo_10      <dbl> NA, NA, NA, NA, 20.047, 19.496, 19.540, 19.605, 19....
## $ cairo_20      <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, 20.0015, 20.195...
## $ global_05     <dbl> NA, NA, 7.108, 7.130, 7.232, 7.296, 7.312, 7.298, 7...
## $ global_10     <dbl> NA, NA, NA, NA, 7.203, 7.223, 7.252, 7.322, 7.445, ...
## $ global_20     <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, 7.6480, 7.6750,...

# subsetting and reshaping each moving average for comparison
cairo_ma05 <- cairo_ma %>%
  select( year, cairo_05, global_05) %>%
```

```

gather( key= "level", value="Temperature", cairo_05, global_05)
cairo_ma10 <- cairo_ma %>%
  select( year, cairo_10, global_10) %>%
  gather( key= "level", value="Temperature", cairo_10, global_10)
cairo_ma20 <- cairo_ma %>%
  select( year, cairo_20, global_20) %>%
  gather( key= "level", value="Temperature", cairo_20, global_20)

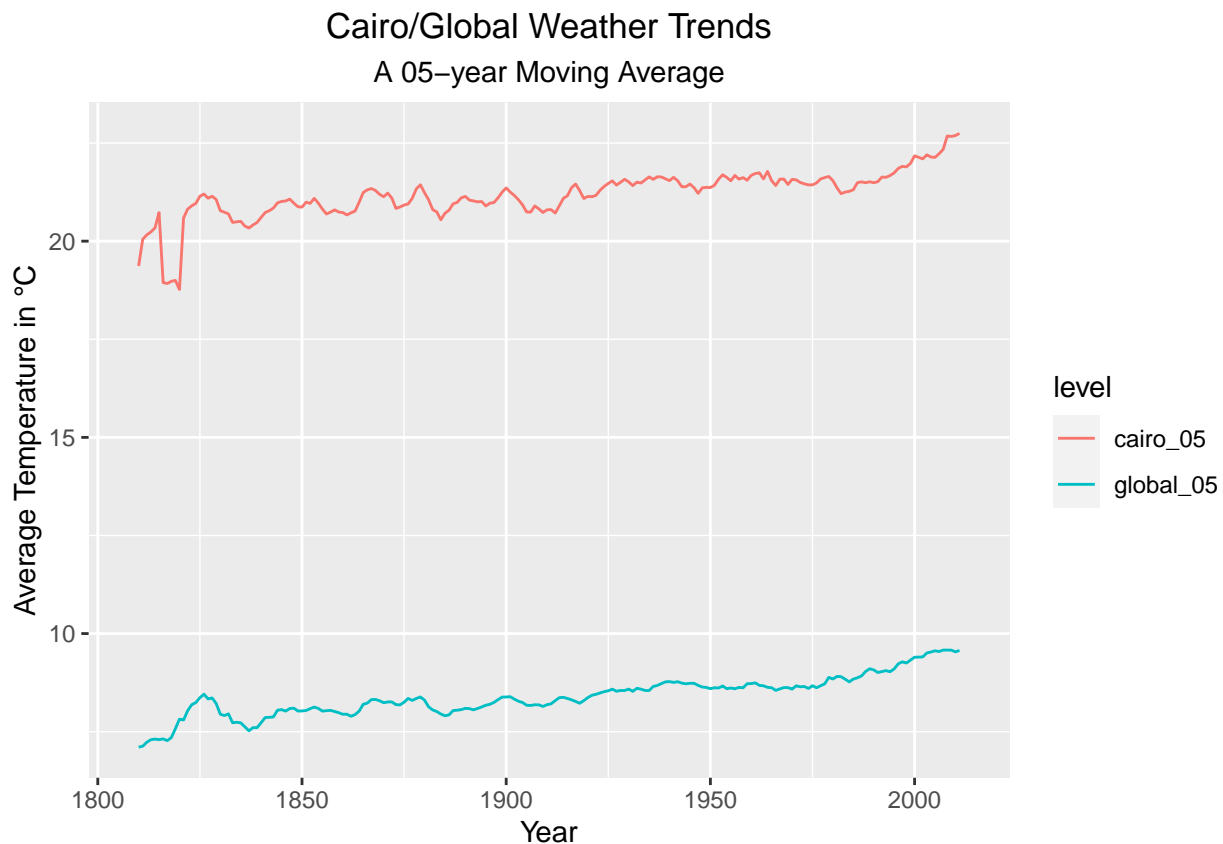
```

- Exploratory visualization for the three moving averages was performed for comparison using the `ggplot2` visualization package.

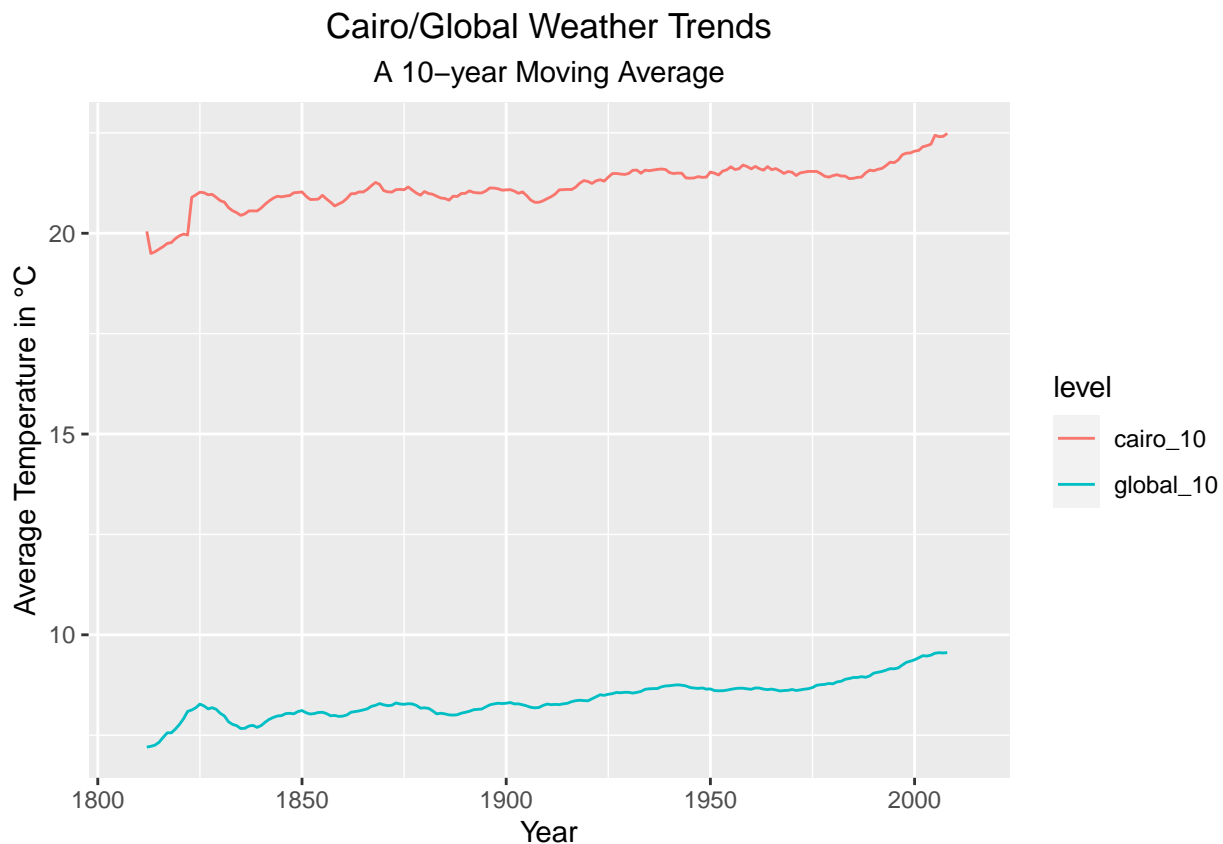
```

# visualizing and comparing the 05-year moving average of cairo and global
# temperature
ggplot(cairo_ma05 , aes(x= year, y= Temperature , color = level))+
  geom_line()+
  labs(
    x="Year",
    y="Average Temperature in °C",
    title="Cairo/Global Weather Trends",
    subtitle = "A 05-year Moving Average")+
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))

```

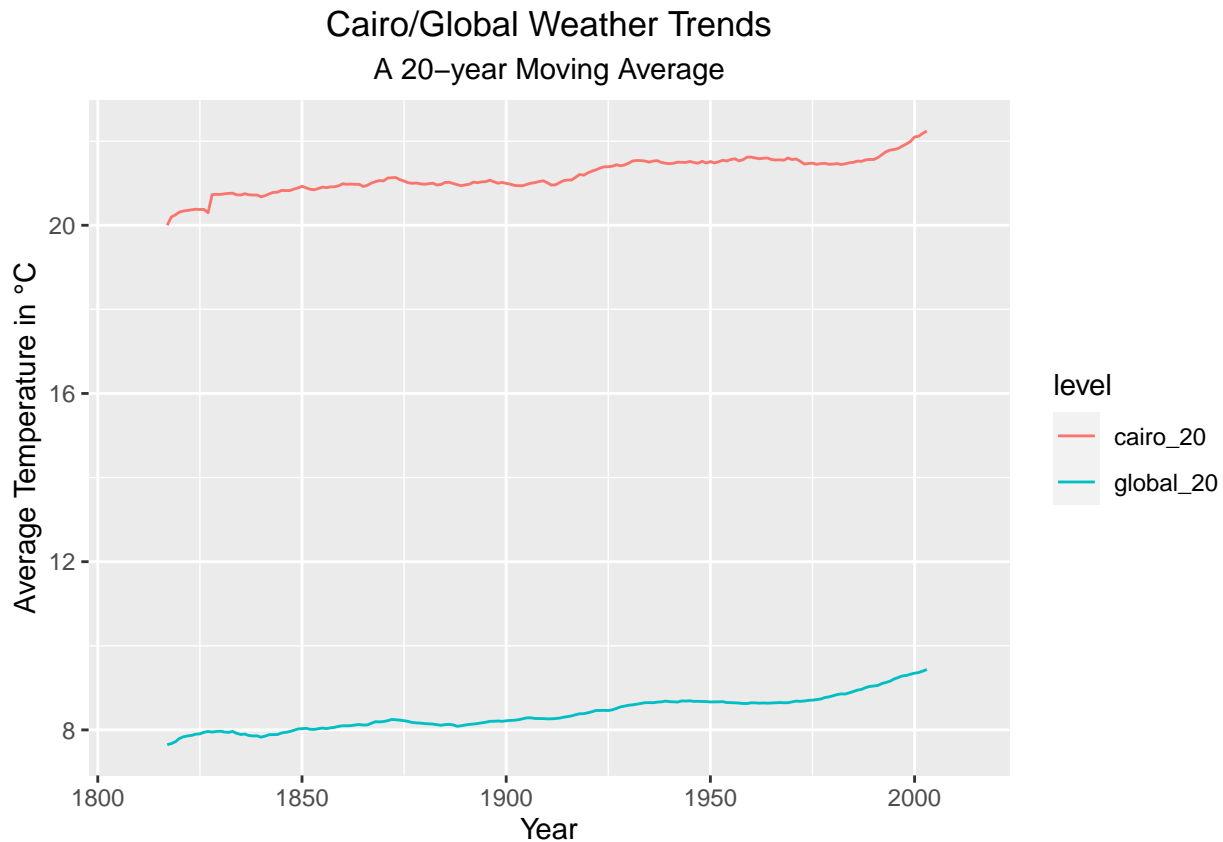


```
# visualizing and comparing the 10-year moving average of cairo and global
# temperature
ggplot(cairo_ma10 , aes(x= year, y= Temperature , color = level))+
  geom_line()+
  labs(
    x="Year",
    y="Average Temperature in °C",
    title="Cairo/Global Weather Trends",
    subtitle = "A 10-year Moving Average")+
  theme(plot.title = element_text(hjust = 0.5),
    plot.subtitle = element_text(hjust = 0.5))
```



```
# visualizing and comparing the 20-year moving average of cairo and global
# temperature
ggplot(cairo_ma20 , aes(x= year, y= Temperature , color = level))+
  geom_line()+
  labs(
    x="Year",
    y="Average Temperature in °C",
    title="Cairo/Global Weather Trends",
    subtitle = "A 20-year Moving Average")+
  theme(plot.title = element_text(hjust = 0.5),
```

```
plot.subtitle = element_text(hjust = 0.5))
```



- From the above visualizations we can see that in the case of a 20-year moving average, a much more smoother lines are obtained in which trends can be easily observed but some details could be lost, such as the early peak at the beginning.
- In case of a 10-year moving average, we have a clear trend line while still preserving details we could use.

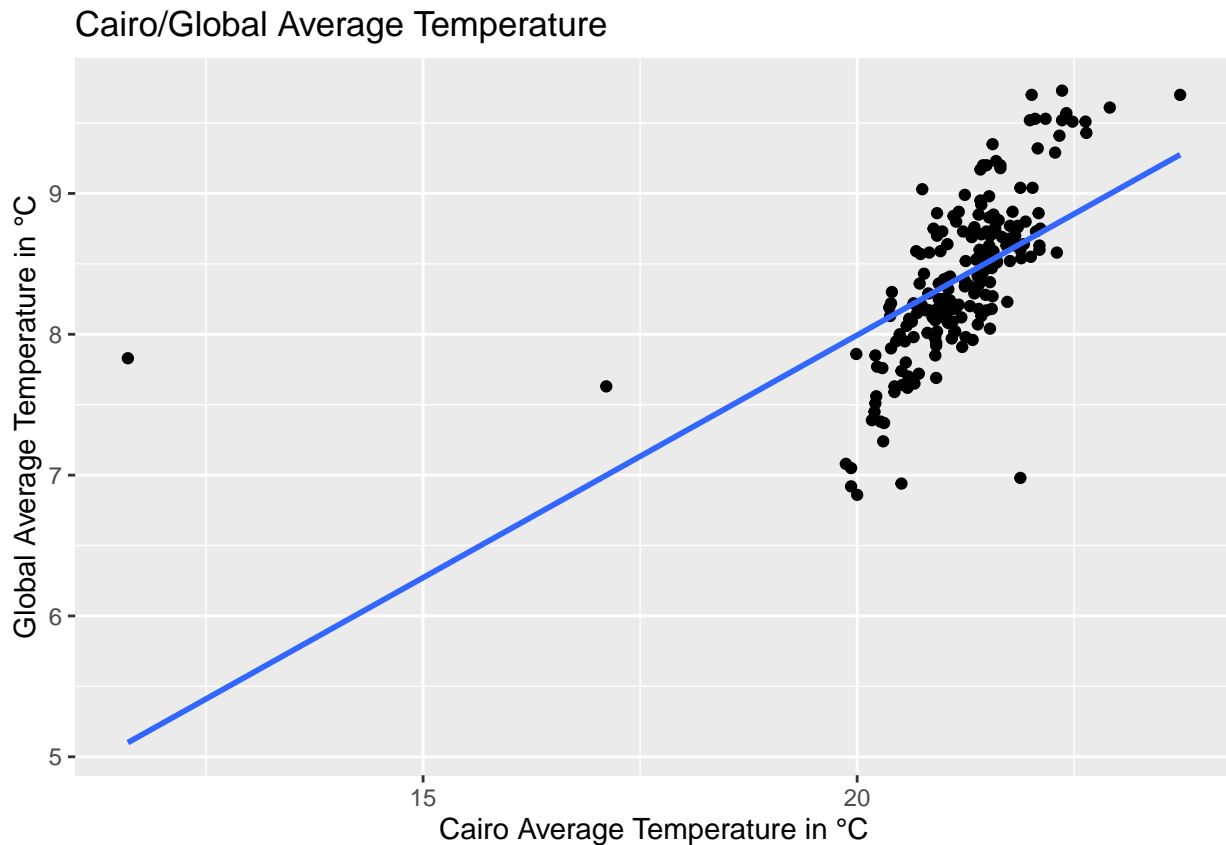
Calculating and comparing the correlation coefficients:

- Correlation coefficients were calculated for average and 10-year moving average between Cairo and global temperatures.
- Pearson correlation coefficient was used, and it was calculated by the `cor()` function provided by the base R `stats` package.

```
# correlation coefficient between the average for cairo and global
# temperature
pearson <- cor(cairo_ma$avg_temp_city, cairo_ma$avg_temp_global,
               method= "pearson")
pearson
```

```
## [1] 0.5826697
```

```
# plotting using ggplot
ggplot(cairo_ma, aes(x= avg_temp_city, y= avg_temp_global)) +
  geom_point() +
  geom_smooth(method = lm,
              se = FALSE)+
  labs(x= "Cairo Average Temperature in °C",
       y= "Global Average Temperature in °C",
       title = "Cairo/Global Average Temperature")
```

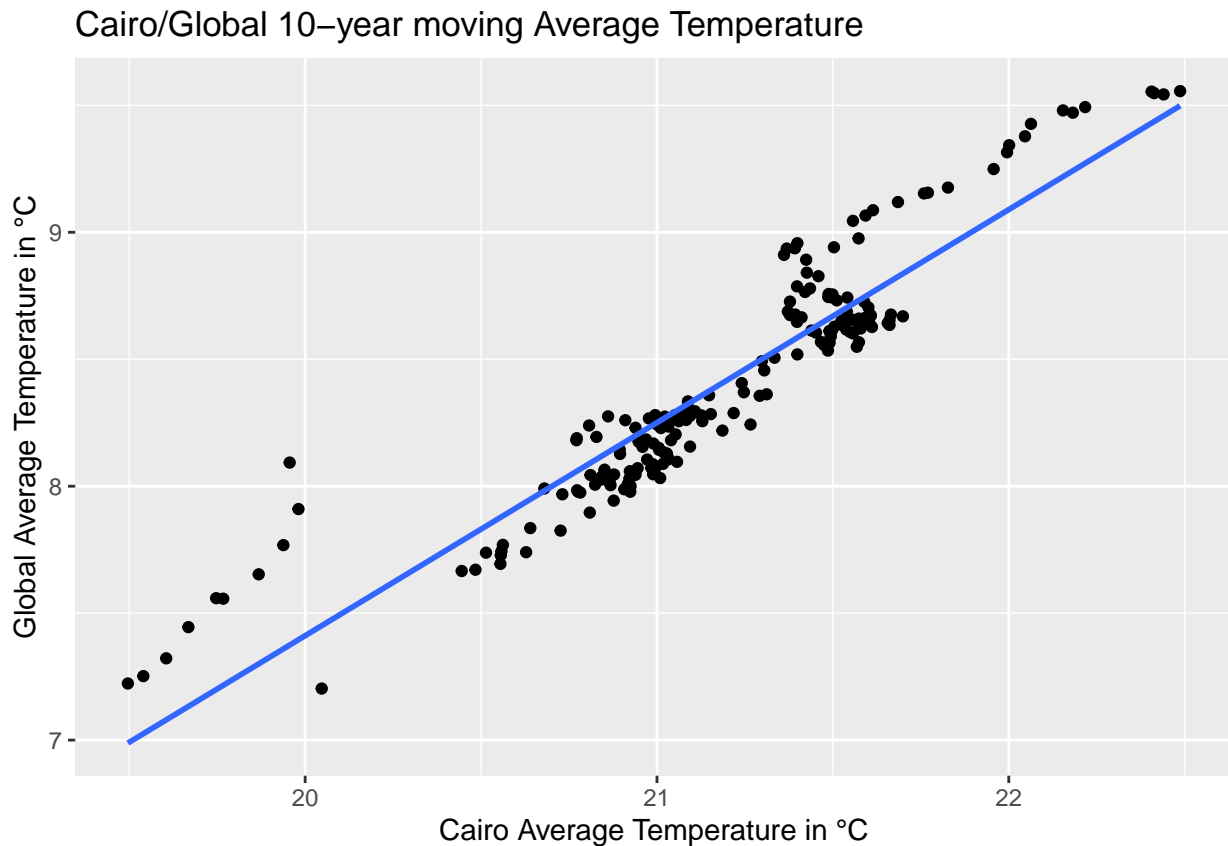


```
# correlation coefficient between the 10-year moving average for cairo and
# global temperature
pearson_ma <- cor(cairo_ma$cairo_10, cairo_ma$global_10,
                  method= "pearson", use= "complete.obs" )
pearson_ma
```

```
## [1] 0.9325437
```

```
# plotting using ggplot
ggplot(cairo_ma, aes(x= cairo_10, y= global_10)) +
  geom_point() +
  geom_smooth(method = lm,
              se = FALSE)+
  labs(x= "Cairo Average Temperature in °C",
```

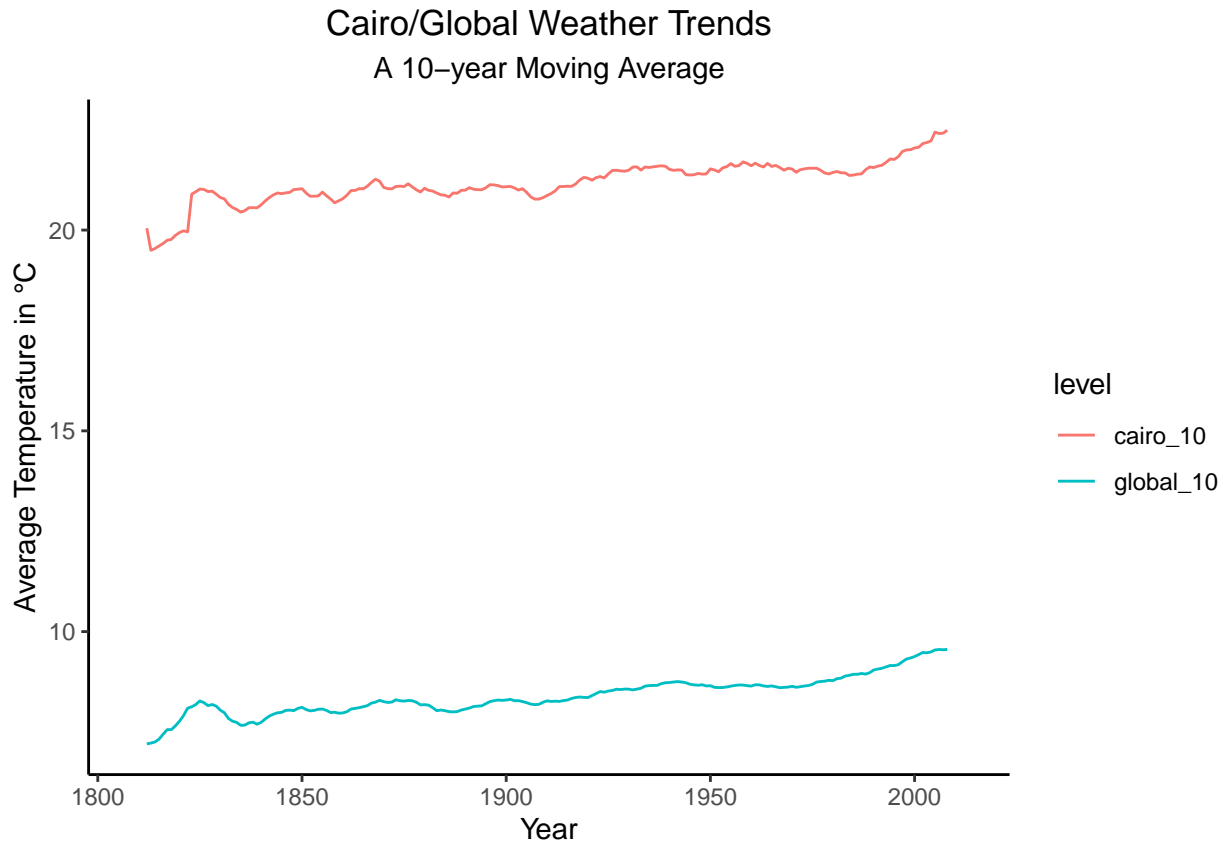
```
y= "Global Average Temperature in °C",
title = "Cairo/Global 10-year moving Average Temperature")
```



Observations:

The provided observations were made from the 10-year moving average line chart.

```
ggplot(cairo_ma10 , aes(x= year, y= Temperature , color = level))+
  geom_line()+
  labs(
    x="Year",
    y="Average Temperature in °C",
    title="Cairo/Global Weather Trends",
    subtitle = "A 10-year Moving Average")+
  theme_classic()+
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```



- From the above line chart we can see that the temperature in Cairo is on average hotter than the average of the global temperature.
- The difference in average between Cairo and global temperature is consistent throughout the chart.
- There is a sharp peak at the beginning of both Cairo and global temperature indicating a sudden increase in temperature that then began to stabilize again.
- Unlike the sharp peak that is present in both line, there is a valley of a sudden drop in temperature in the Cairo line. both the peak and the valley(in case of Cairo), could be due to some measurement errors in the past (both are in the early 1800s) or could be not, but the peak is present in both lines.
- The increase in the average temperature is steady and on the rise in both cases, indicated by the rising lines and the strong positive correlation coefficient [0.9325437] between the moving average of Cairo and global temperature.
- This trend shows an overall increase in temperature that has been consistent over the last few hundred years, from which we can expect even more increase in temperature in the coming decades (unless something is done about climate change, of course).

Comparing temperature trend in Cairo and other cities:

Here we plot and visualize the weather trends in Cairo and five other cities.

```
# importing the data set
cities <- read.csv("cities_results.csv")
glimpse(cities)

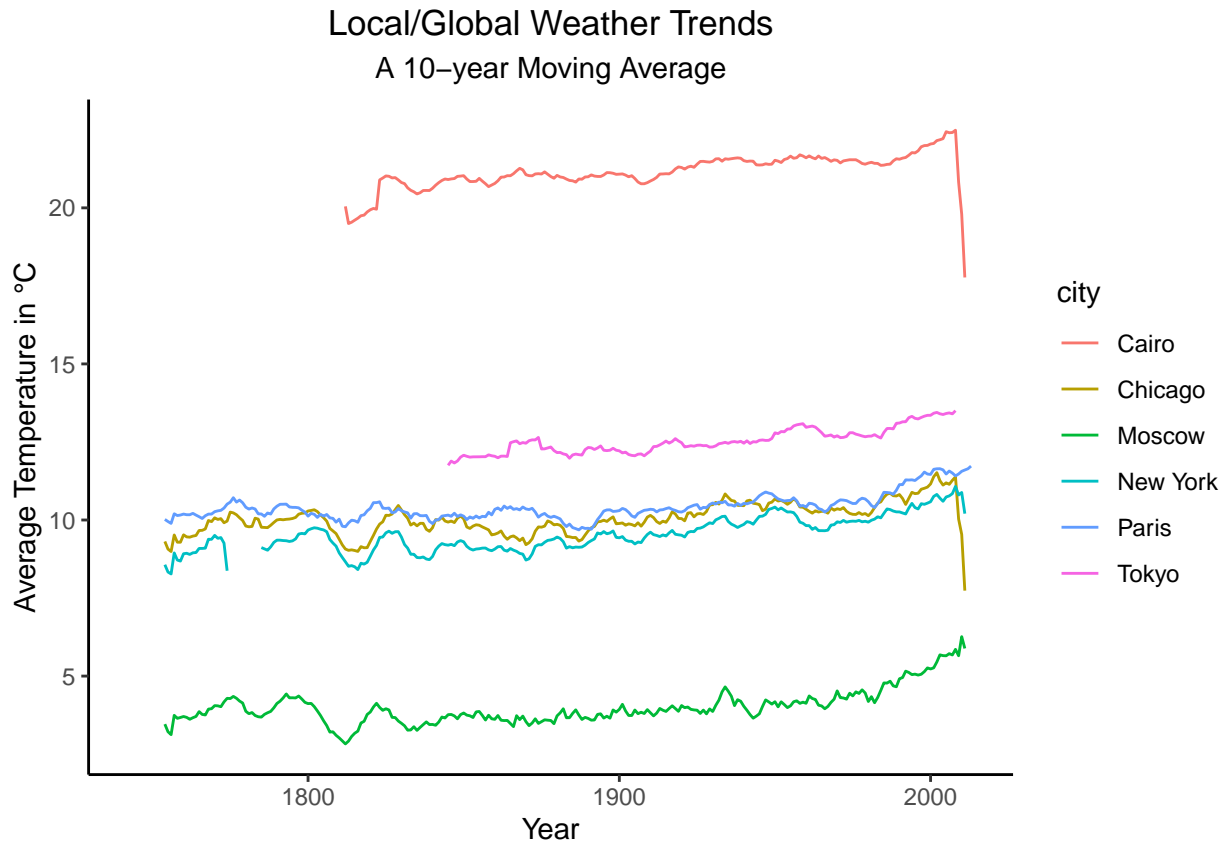
## Rows: 1,459
## Columns: 3
## $ year      <int> 1808, 1809, 1810, 1811, 1812, 1813, 1814, 1815, 1816,...
## $ city      <chr> "Cairo", "Cairo", "Cairo", "Cairo", "Cairo", "Cairo",...
## $ avg_temp_city <dbl> 17.11, 19.87, 19.93, 20.00, 19.93, 20.51, 20.43, 20.3...

# calculating 10-year moving averages using the rollmean function
# provided by zoo package
cities_ma <- cities %>%
  mutate(city_10 = rollmean(avg_temp_city, 10, fill = NA )) %>%
  select( year, city, city_10)

glimpse(cities_ma)

## Rows: 1,459
## Columns: 3
## $ year      <int> 1808, 1809, 1810, 1811, 1812, 1813, 1814, 1815, 1816, 1817,...
## $ city      <chr> "Cairo", "Cairo", "Cairo", "Cairo", "Cairo", "Cairo", "Cair...
## $ city_10    <dbl> NA, NA, NA, NA, 20.047, 19.496, 19.540, 19.605, 19.668, 19....

# visualizing the 10-year moving average of different cities
ggplot(cities_ma , aes(x= year, y= city_10, color= city))+
  geom_line()+
  labs(
    x="Year",
    y="Average Temperature in °C",
    title="Local/Global Weather Trends",
    subtitle = "A 10-year Moving Average")+
  theme_classic()+
  theme(plot.title = element_text(hjust = 0.5),
        plot.subtitle = element_text(hjust = 0.5))
```



Some strange observations are more clear in this graph:

- There is a steep decrease in the average temperature in case of Cairo and Chicago during the past few years (was not as clear in case of Cairo when we were comparing with global temperature). There is a big chance of this being an error.
- Some lines are more smoother than others, for example Tokyo, which show a continuous increase in average temperature through the past years.
- Of the six cities, still Cairo is the one with hottest temperature on average compared with the others while Moscow is the coldest.
- The peak at the early 1800s is found in all the cities except Tokyo, which has no temperature information for this period.
- Overall, a consistent trend of increase in the average temperature is found in all the cities, some more prominent than others.