Assignment 3

Nataly Aranda

05/04/2022

Analysis Report of Average Temperature Data Series - America

by Nataly Aranda

Dataset Source: https://www.kaggle.com/datasets/sudalairajkumar/daily-temperature-of-major-cities (https://www.kaggle.com/datasets/sudalairajkumar/daily-temperature-of-major-cities) , Data description: https://academic.udayton.edu/kissock/http/Weather/source.htm (https://academic.udayton.edu/kissock/http/Weather/source.htm)

The library tidyverse will be used in this analysis. Tidyvers includes packages as ggplot used for graphs and visualizations, dplyr used for data manipulation, readr used for data import and more. The first step is then call the tidyverse library as follow:

```
library(tidyverse)
```

The data set selected for this analysis is city_temperatures.csv. This is a data set with average temperatures for the major cities of the world, this data set includes 8 fields: Region, Country, State, City, month, day, year, average daily temperature (F). In total, we have 2906327 records for the 8 fields.

```
temperatures=read.csv('city_temperature.csv',TRUE,',')
str(temperatures)
```

```
## 'data.frame':    2906327 obs. of  8 variables:
##  $ Region        : chr  "Africa" "Africa" "Africa" "Africa" ...
##  $ Country       : chr  "Algeria" "Algeria" "Algeria" "Algeria" ...
##  $ State         : chr  "" "" "" "" ...
##  $ City          : chr  "Algiers" "Algiers" "Algiers" "Algiers" ...
##  $ Month         : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ Day           : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ Year          : int  1995 1995 1995 1995 1995 1995 1995 1995 1995 1995 ...
##  $ AvgTemperature: num  64.2 49.4 48.8 46.4 47.9 48.7 48.9 49.1 49 51.9 ...
```

```
map_chr(temperatures,typeof)
```

```
##         Region        Country          State           City          Month
##    "character"    "character"    "character"    "character"      "integer"
##            Day           Year AvgTemperature
##      "integer"      "integer"       "double"
```

```
head(temperatures)
```

```
##   Region Country State    City Month Day Year AvgTemperature
## 1 Africa Algeria       Algiers     1   1 1995           64.2
## 2 Africa Algeria       Algiers     1   2 1995           49.4
## 3 Africa Algeria       Algiers     1   3 1995           48.8
## 4 Africa Algeria       Algiers     1   4 1995           46.4
## 5 Africa Algeria       Algiers     1   5 1995           47.9
## 6 Africa Algeria       Algiers     1   6 1995           48.7
```

Creating a subset for analyzing only the temparatures for countries in America and Caribbean

For analyzing only temperatures in countries of America, the first step is to create a subset for data of temperatures for countries located South America, Central America and North America. This new data set will be called America_Temperatures.

```
America_Temperatures <- subset(temperatures,Region=="South/Central America & Carribean" | Region
=="North America", select=Region:AvgTemperature)
head(America_Temperatures)
```

```
##                Region Country State    City Month Day Year AvgTemperature
## 1130117 North America  Canada       Calgary     1   1 1995           12.6
## 1130118 North America  Canada       Calgary     1   2 1995            4.5
## 1130119 North America  Canada       Calgary     1   3 1995            2.5
## 1130120 North America  Canada       Calgary     1   4 1995           11.4
## 1130121 North America  Canada       Calgary     1   5 1995           11.3
## 1130122 North America  Canada       Calgary     1   6 1995            4.0
```

Cleaning the data

The next step is to clean the data:

1. Delete values without values for temperatures
2. Convert the temperature from Fahrenheit to Celsius/
3. Group the data by country and year and calculate maximum and minimum temperatures for each year in each country.

Step by step

1. Deleting the data with values of temperature of -99 (values with no temperature recorded for that day).

```
America_Temperatures <- America_Temperatures %>% filter(AvgTemperature > -99)
```

2. As the original data is in Fahrenheit, we will create a new column converting the Avg Temperature to degree Celsius.

```
America_Temperatures['AvgTemperatureC']= (America_Temperatures['AvgTemperature']-32)/1.8
```

3. Finally, we will group the data by country and year and calculate the maximum and minimum temperature of the data grouped.

```
AmericaMaxMinTemp <- America_Temperatures %>% group_by(Region,Country, Year) %>% summarise(MaxTe
mp=max(AvgTemperatureC),MinTemp=min(AvgTemperatureC))
```

```
## `summarise()` has grouped output by 'Region', 'Country'. You can override using
## the `.groups` argument.
```

```
head(AmericaMaxMinTemp)
```

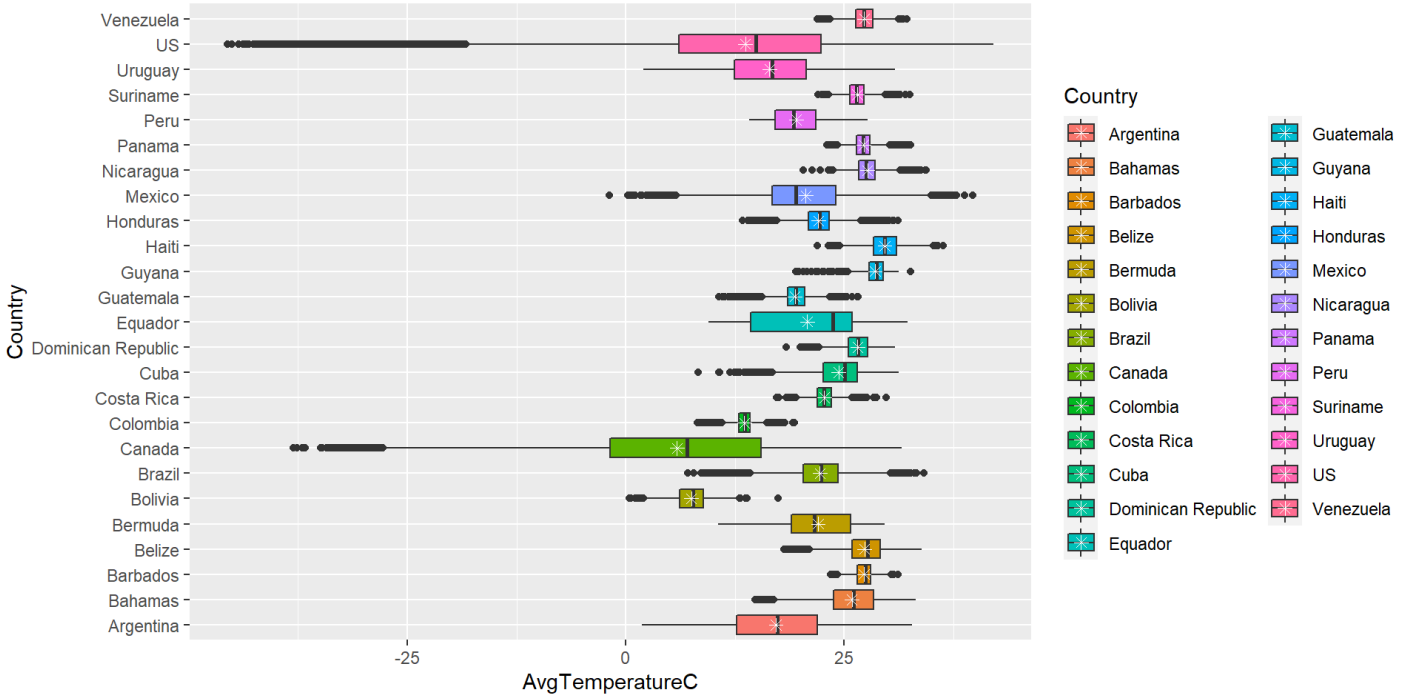
```
## # A tibble: 6 x 5
## # Groups:   Region, Country [1]
##   Region        Country  Year MaxTemp MinTemp
##   <chr>         <chr>   <int>   <dbl>   <dbl>
## 1 North America Canada   1995    30.3   -32.6
## 2 North America Canada   1996    28.4   -37.6
## 3 North America Canada   1997    27.2   -34.2
## 4 North America Canada   1998    27.1   -34.1
## 5 North America Canada   1999    30.2   -32.1
## 6 North America Canada   2000    25.4   -31.3
```

Now our data is ready to be analyzed.

Data Analysis

First, I wanted to see the variation of temperature for each country in America. For doing that, I used the America_Temperatures data set, the following graphs using box whiskers show that variations:

```
ggplot(America_Temperatures, aes(y = Country, x = AvgTemperatureC, fill = Country)) +
  geom_boxplot() + stat_summary(fun = "mean", geom = "point", shape = 8, size = 2, color = "whit
e")
```



We observed that US and Canada are the countries with most variation in temperatures. Also it is possible to see that the inter tropic countries do not present higher variations in temperatures, as was expected because they do not have stations.

How have the maximum and minimum temperatures changed in America?

To analyze how the temperature has been changed in time, I plotted the maximum and minimum temperatures by year and country. I grouped the countries in regions as follow:

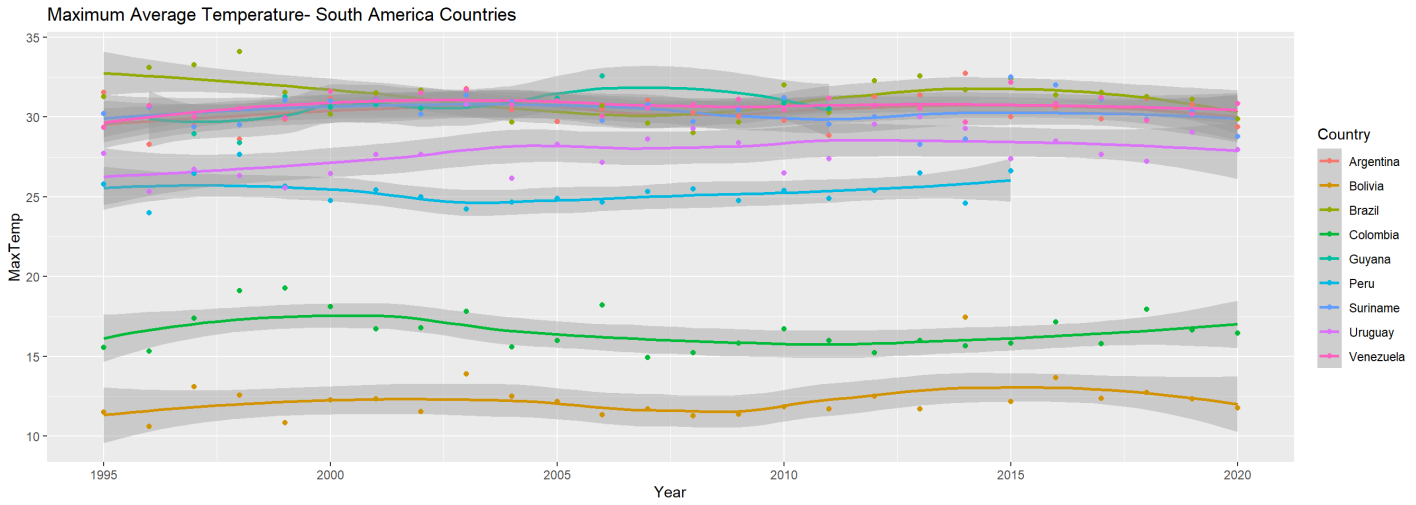
- South America: Venezuela, Colombia, Brazil, Ecuador, Peru, Bolivia, Argentina, Uruguay, Suriname, Guyana
- Central America: Panama, Costa Rica, Nicaragua, Honduras, Guatemala, Belize
- Caribbean: Bahamas, Barbados, Bermuda, Cuba, Dominican Republic, Haiti
- North America: Canada, US, Mexico

South America

The changes in time for maximum and minimum average temperatures for countries in South America are shown in the next two graphs, plotted using ggplot.

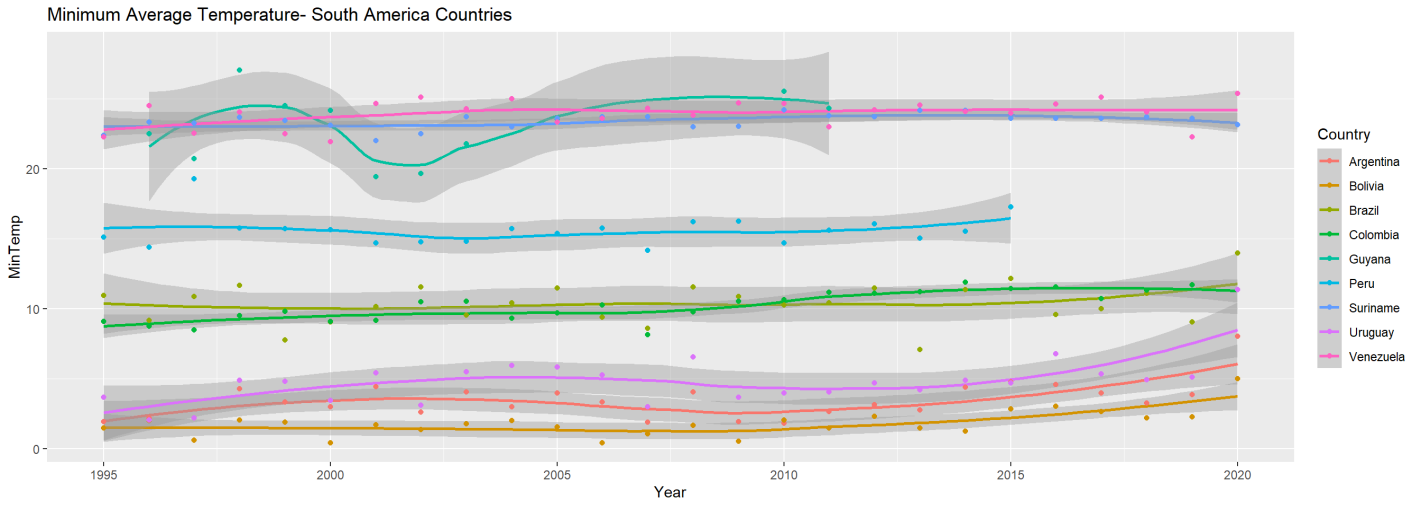
```
ggplot(data = filter(AmericaMaxMinTemp,Country == "Colombia"| Country=="Venezuela"| Country=="Ec
uador" | Country=="Peru"| Country=="Brazil" | Country=="Bolivia"| Country=="Argentina"|Country==
"Uruguay" |Country=="Suriname" |Country=="Guyana")) +  geom_smooth(mapping = aes(x = Year, y = M
axTemp, color=Country))+ geom_point(mapping = aes(x = Year, y = MaxTemp, color=Country))+ ggtitl
e("Maximum Average Temperature- South America Countries")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
ggplot(data = filter(AmericaMaxMinTemp,Country == "Colombia"| Country=="Venezuela"| Country=="Ecuador" | Country=="Peru"| Country=="Brazil" | Country=="Bolivia"| Country=="Argentina"|Country=="Uruguay"|Country=="Suriname" |Country=="Guyana" )) +  geom_smooth(mapping = aes(x = Year, y = MinTemp, color=Country))+ geom_point(mapping = aes(x = Year, y = MinTemp, color=Country))+ ggtitle("Minimum Average Temperature- South America Countries")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

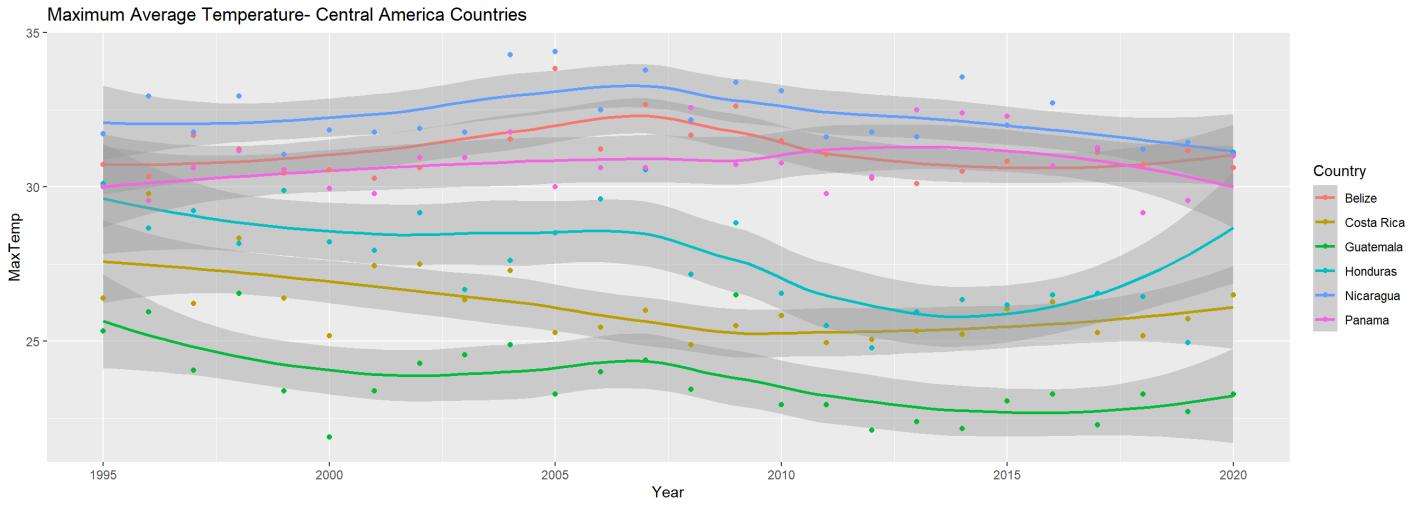


Central America

The changes in time for maximum and minimum average temperatures for countries in Central America are shown in the next two graphs, plotted using ggplot.

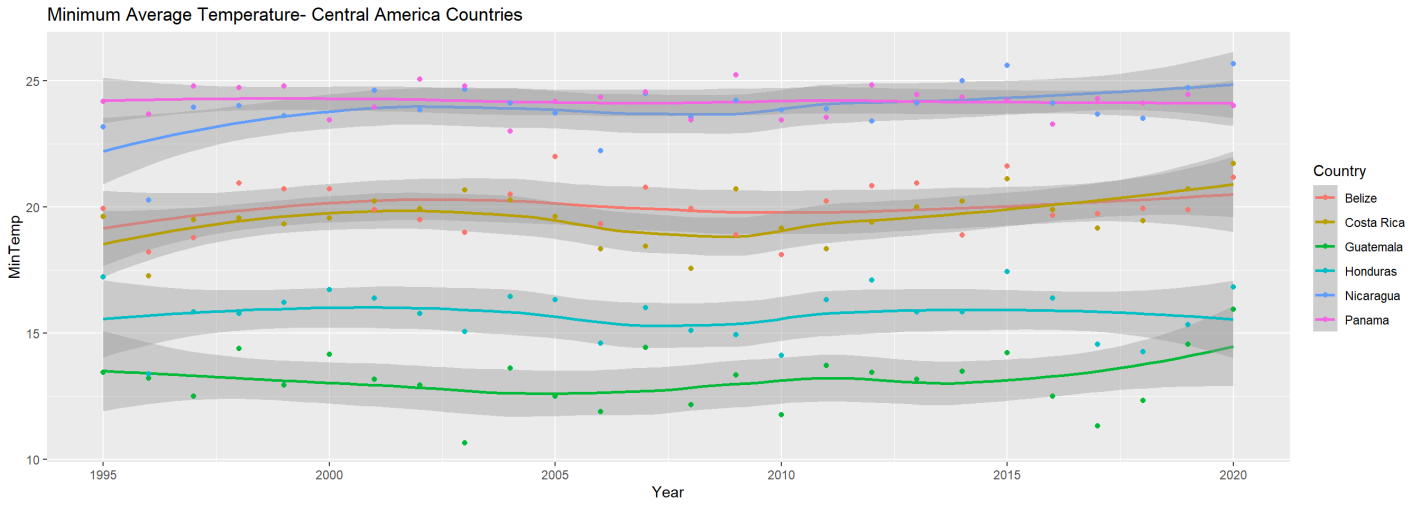
```
ggplot(data = filter(AmericaMaxMinTemp,Country == "Panama"| Country=="Costa Rica"| Country=="Nicaragua" | Country=="Honduras"| Country=="Guatemala" | Country=="Belize")) +  geom_smooth(mapping = aes(x = Year, y = MaxTemp, color=Country))+ geom_point(mapping = aes(x = Year, y = MaxTemp, color=Country))+ ggtitle("Maximum Average Temperature- Central America Countries")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
ggplot(data = filter(AmericaMaxMinTemp,Country == "Panama"| Country=="Costa Rica"| Country=="Nic
aragua" | Country=="Honduras"| Country=="Guatemala" | Country=="Belize")) +  geom_smooth(mapping
= aes(x = Year, y = MinTemp, color=Country))+ geom_point(mapping = aes(x = Year, y = MinTemp, co
lor=Country))+ ggtitle("Minimum Average Temperature- Central America Countries")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```

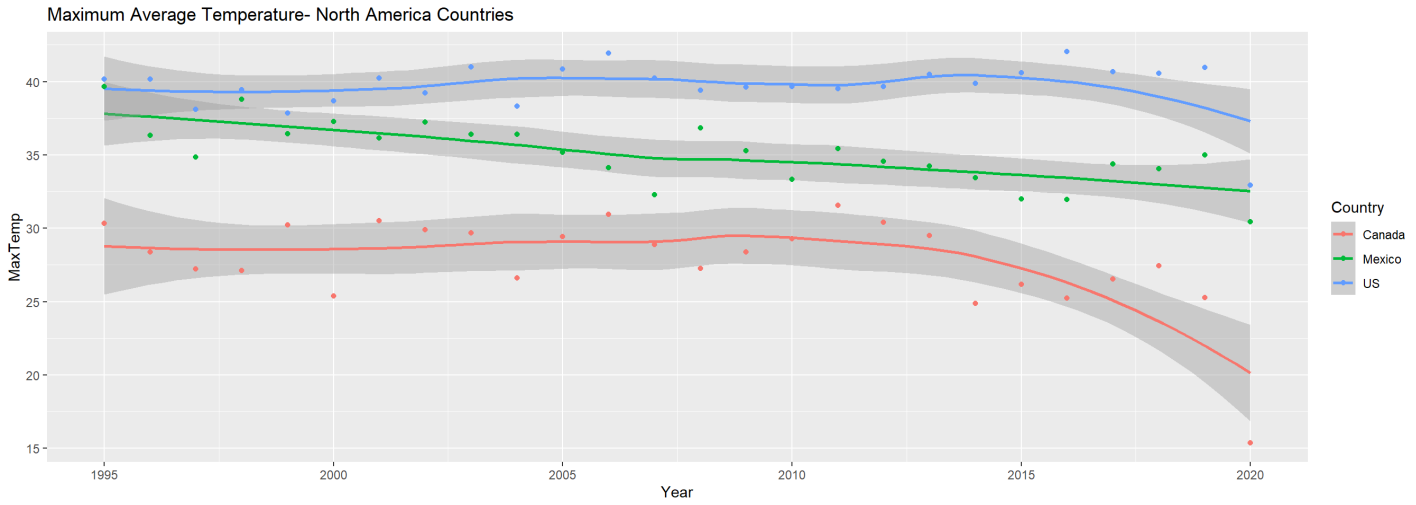


Norht America

The changes in time for maximum and minimum average temperatures for countries in North America are shown in the next two graphs, plotted using ggplot.

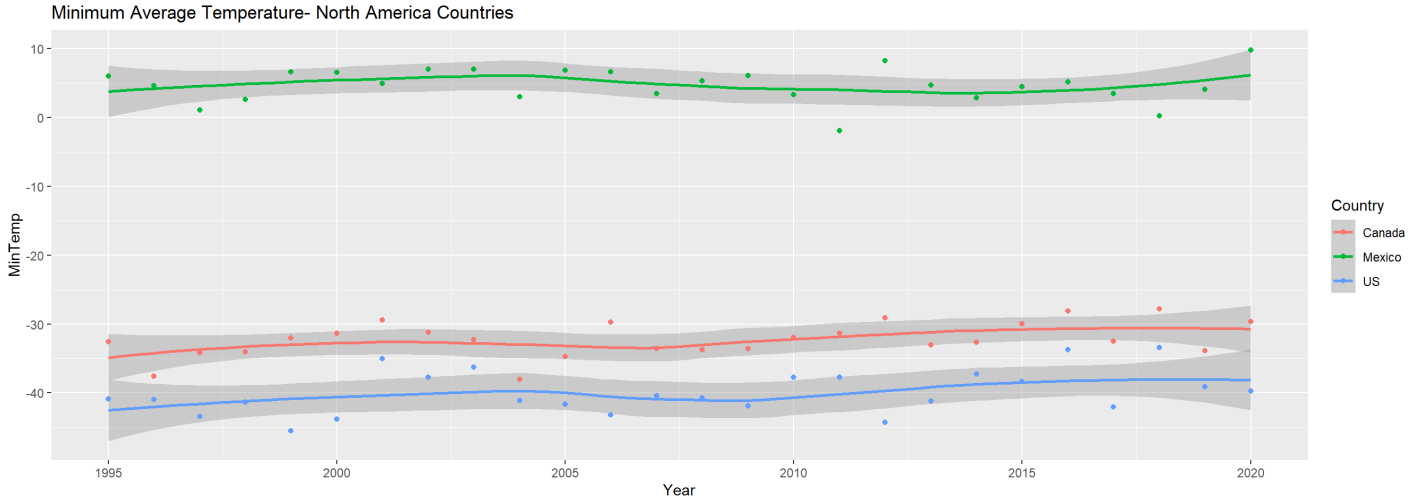
```
ggplot(data = filter(AmericaMaxMinTemp, Country=="Canada"|Country=="US" |Country=="Mexico")) +
 geom_smooth(mapping = aes(x = Year, y = MaxTemp, color=Country))+ geom_point(mapping = aes(x =
 Year, y = MaxTemp, color=Country)) + ggtitle("Maximum Average Temperature- North America Countr
ies")
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```




```
ggplot(data = filter(AmericaMaxMinTemp, Country=="Canada"|Country=="US" |Country=="Mexico" )) + geom_smooth(mapping = aes(x = Year, y = MinTemp, color=Country))+ geom_point(mapping = aes(x = Year, y = MinTemp, color=Country))+ ggtitle("Minimum Average Temperature- North America Countries")
```

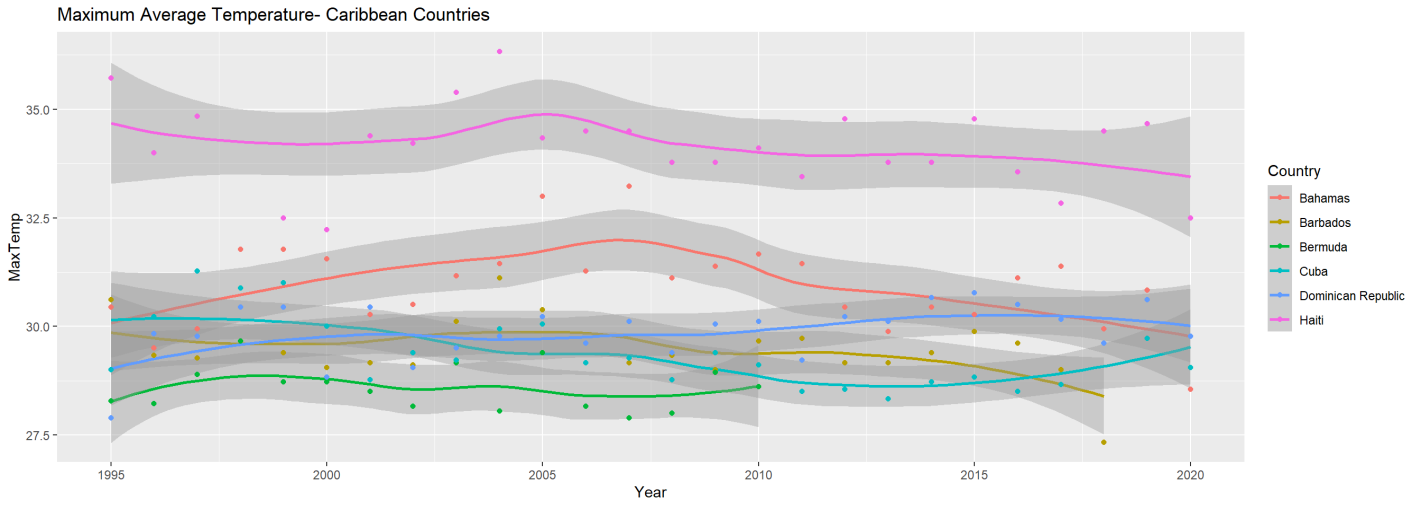
```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



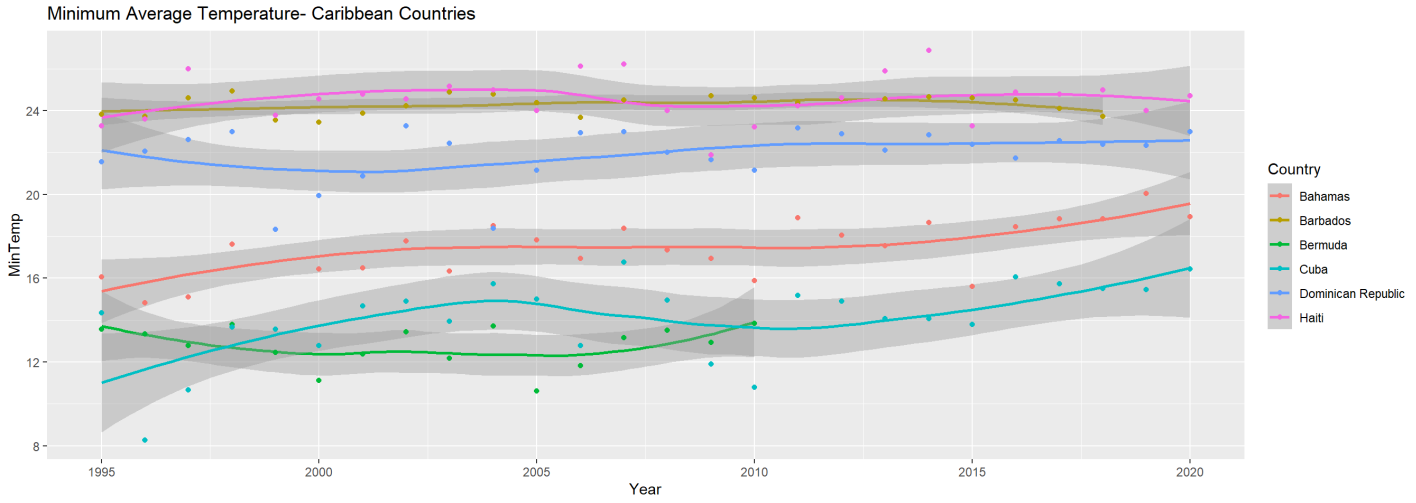
Caribbean

The changes in time for maximum and minimum average temperatures for countries in Caribbean are shown in the next two graphs, plotted using ggplot.

```
ggplot(data = filter(AmericaMaxMinTemp,Country == "Bahamas"| Country=="Barbados"| Country=="Bermuda" | Country=="Cuba"| Country=="Dominican Republic" | Country=="Haiti")) +  geom_smooth(mapping = aes(x = Year, y = MaxTemp, color=Country))+ geom_point(mapping = aes(x = Year, y = MaxTemp, color=Country)) + ggtitle("Maximum Average Temperature- Caribbean Countries")
```




```
ggplot(data = filter(AmericaMaxMinTemp,Country == "Bahamas"| Country=="Barbados"| Country=="Berm
uda" | Country=="Cuba"| Country=="Dominican Republic" | Country=="Haiti")) +  geom_smooth(mappin
g = aes(x = Year, y = MinTemp, color=Country))+ geom_point(mapping = aes(x = Year, y = MinTemp,
 color=Country))+ ggtitle("Minimum Average Temperature- Caribbean Countries")
```



Trends observed

We could observed in the plots of maximum and minimum temperatures along the time, that in general there is a trend of increase in temperature for the countries in America. The countries located in the Caribbean are the ones that showed higher increments in minimum temperature along the time. Argentina, Bolivia and Uruguay also showed a high increase of minimum temperature along the time.

On the other hand, some countries in Central America showed a decrease of maximum temperature along the time, those countries were: Honduras, Guatemala and Costa Rica. However, this behavior was not observed for the minimum temperature that showed and increment with time.

Conclusion

Using R we could analyzed the change in temperature along the time for some countries in America. We could see that in general the temperatures have increased with time for the majority of countries in America.