PSTAT174-Project

Jon Carlos (3872520)

2023-03-13

```
library(tidyverse)
library(ggplot2)
library(ggfortify)
library(ggthemes)
library(zoo)
library(forecast)
library(Mcomp)
library(MASS)
library(fabletools)
library(reshape2)
```

Within a century, the world has seen great changes in its climate.[] Countries around the world have now been experiencing harsher weather patterns with extreme weather events becoming the norm. Among these changes is a global change in temperature. In order to find the significance of the climate change, data about the temperature over tens of years in different places around the world would be necessary. Thankfully, there exists such data and this will be an analysis focusing on climate change within Europe.

FIX THIS^^^^^ Its too elementary

Analyzing Europe would be beneficial because of the different climates throughout the continent while still being relatively closer together. This makes it so that there could be an analysis around location and temperature change. Because the location and the environment could have an effect on the climate, it is necessary to look that these variables and see they make any significant difference to the temperature.

The data originates from the website Data Commons and they had acquired their data from noaa.gov also known as, **National Ocean and Atmospheric Administration**.

Being able to tell whether there is a significant change in the climate would make understanding just how problematic global warming really is. It would allow people to be able to get a digestible number that can anyone can understand. It can make people really understand how the world is changing without having to fully explain everything that is really going on.

This is a very common thing to study because of how important understanding climate change its recent effects. This means that there are definitely plenty of researchers also studying this data.

Just how much has the temperature increased? Are there differences in the amount of temperature

increase and the location? If location matters, is it an effect of the environment or is it an effect of solely the location?

```
EU_Max_Temperatures <- read.csv("Temperatures_In_Europe.csv")
names(EU_Max_Temperatures) <- c("date", "Berlin", "Paris", "Copenhagen", "Stockholm", "Rome")


head(EU_Max_Temperatures, 5)
```

```
##      date Berlin Paris Copenhagen Stockholm Rome
## 1 1900-01    8.4  12.7        6.2       5.0   NA
## 2 1900-02   17.4  16.9        6.2       4.5   NA
## 3 1900-03   15.0  16.9        7.6       5.5   NA
## 4 1900-04   24.1  25.8       19.2      14.0   NA
## 5 1900-05   29.4  28.9       23.4      21.0   NA
```

```
summary(EU_Max_Temperatures)
```

```
##      date               Berlin          Paris         Copenhagen
##  Length:1473        Min.   : 1.50   Min.   : 6.10   Min.   : 1.50
##  Class :character   1st Qu.:13.90   1st Qu.:16.00   1st Qu.:10.80
##  Mode  :character   Median :22.70   Median :23.72   Median :18.50
##                     Mean   :21.92   Mean   :23.29   Mean   :18.35
##                     3rd Qu.:29.90   3rd Qu.:29.80   3rd Qu.:25.90
##                     Max.   :38.90   Max.   :42.61   Max.   :39.10
##
##    Stockholm          Rome
##  Min.   : 0.00   Min.   :14.60
##  1st Qu.: 9.00   1st Qu.:21.00
##  Median :16.75   Median :27.00
##  Mean   :16.95   Mean   :27.23
##  3rd Qu.:25.00   3rd Qu.:33.50
##  Max.   :41.00   Max.   :42.00
##  NA's   :3       NA's   :576
```

However, there is some missing data within the column for Rome as seen above so we must remove them.

Because the data set is rather large to begin with starting all the way from the year 1900, we can get rid of data from each other column until Rome finally has some. Looking at the data set, Rome begins to have some data when `date = 1944-06`, but is still patchy with this information until around `1951-01`. Therefore, all data from before this will be removed and only from 1951 and onward will be analyzed.

```
EU_Max_Temperatures <- EU_Max_Temperatures[EU_Max_Temperatures$date >= '1951-01',]
head(EU_Max_Temperatures, 5)
```

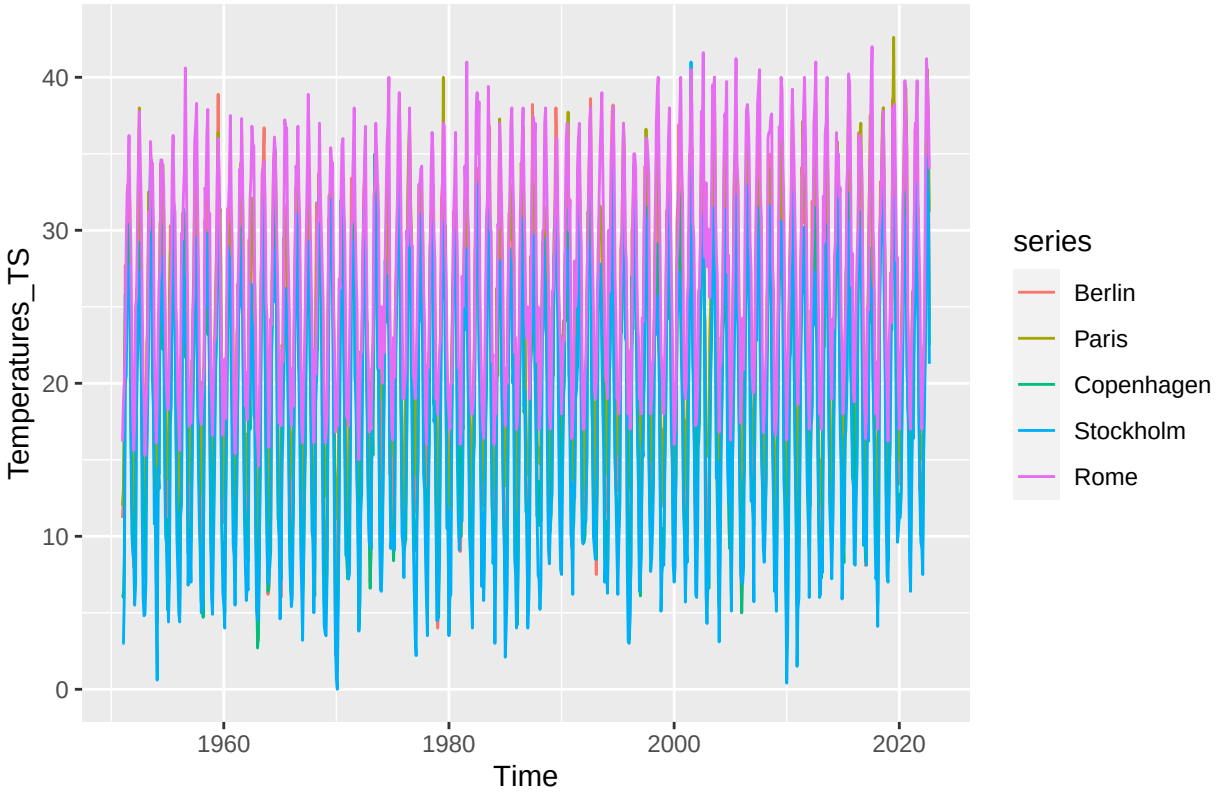
```
##        date Berlin Paris Copenhagen Stockholm Rome
## 613 1951-01   11.2  12.0      6.200       3.0 16.2
## 614 1951-02   12.5  13.1      6.000       3.0 18.7
## 615 1951-03   14.8  19.6      8.889       6.2 20.2
## 616 1951-04   27.7  24.2     25.800      20.0 22.8
## 617 1951-05   24.4  25.8     25.400      20.5 29.0
```

```
# try to get a plot of EU_Max_Temperatures in here if not whatever
```

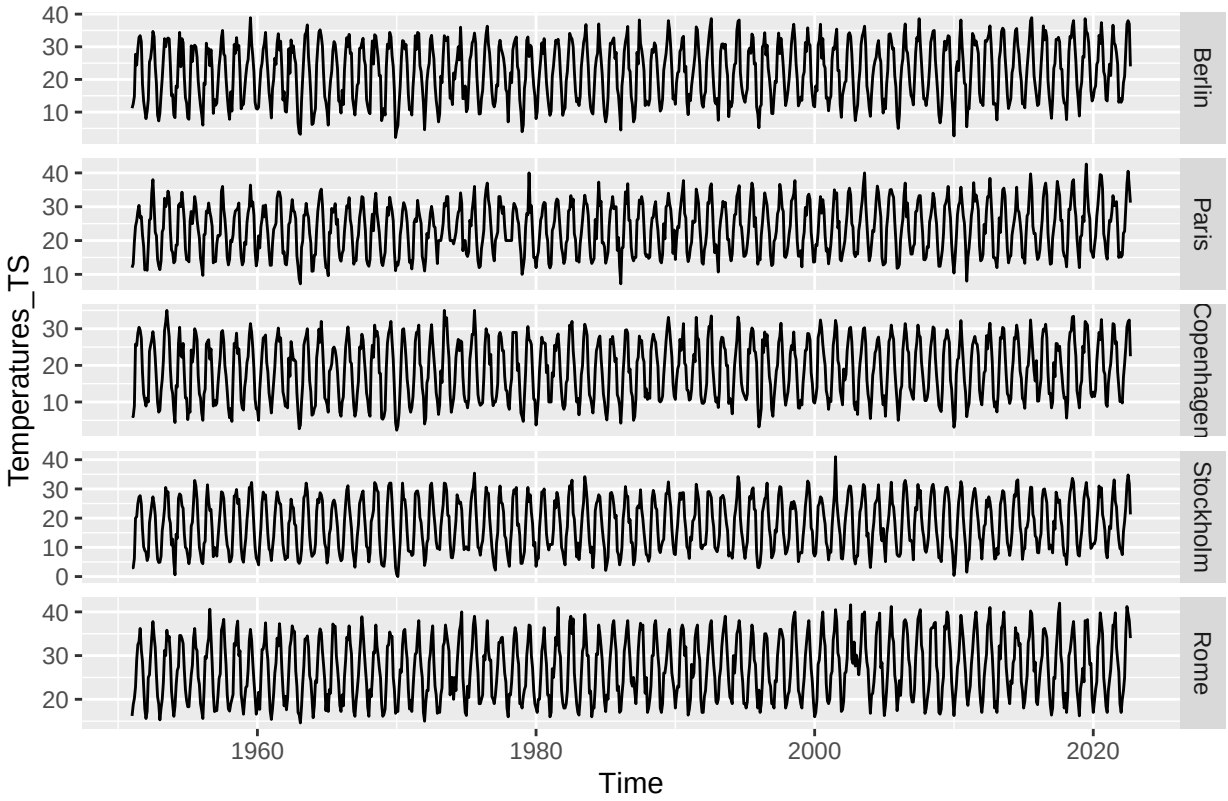
```
Temperatures_TS <- ts(EU_Max_Temperatures[,c(2:6)], start = c(1951), frequency = 12)
head(Temperatures_TS)
```

```
##          Berlin Paris Copenhagen Stockholm Rome
## Jan 1951   11.2  12.0      6.200       3.0 16.2
## Feb 1951   12.5  13.1      6.000       3.0 18.7
## Mar 1951   14.8  19.6      8.889       6.2 20.2
## Apr 1951   27.7  24.2     25.800      20.0 22.8
## May 1951   24.4  25.8     25.400      20.5 29.0
## Jun 1951   29.4  28.6     28.800      25.0 32.5
```

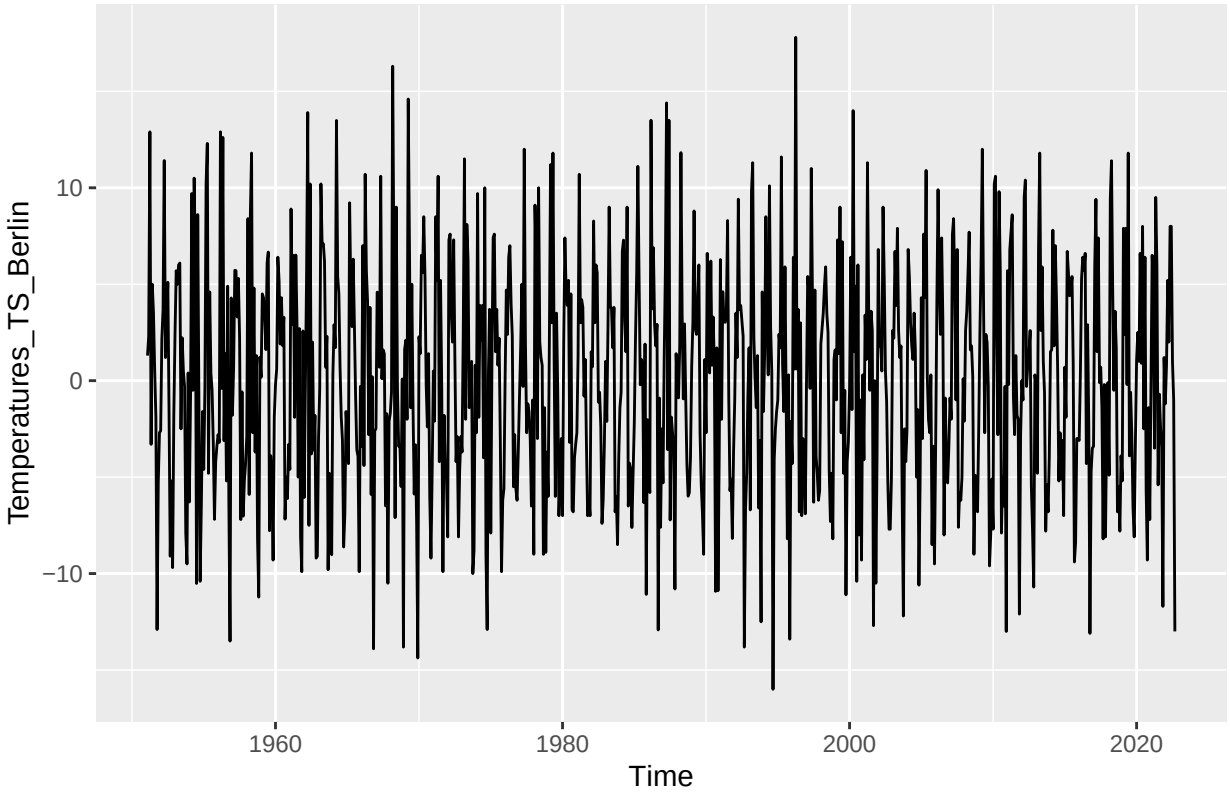
```
autoplot(Temperatures_TS)
```



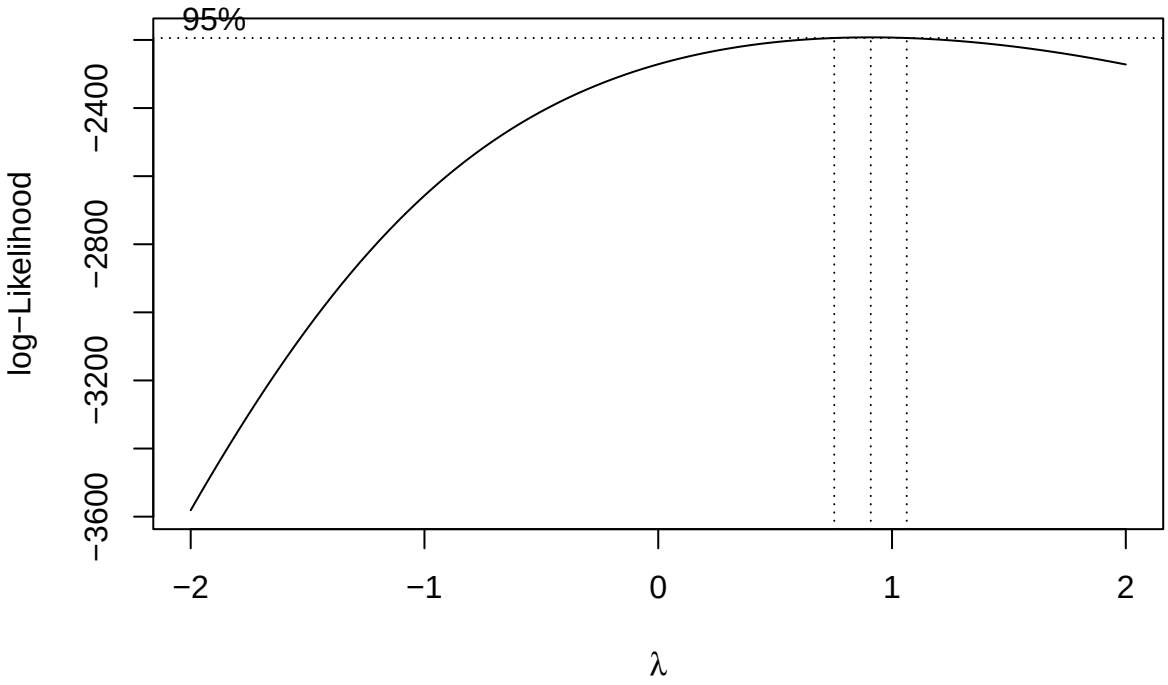

```
autoplot(Temperatures_TS, facets = TRUE)
```




```
Temperatures_TS_Berlin <- diff(Temperatures_TS[,1])

Temperatures_TS_Paris <- diff(Temperatures_TS[,2])

Temperatures_TS_Copenhagen <- diff(Temperatures_TS[,3])

Temperatures_TS_stockholm <- diff(Temperatures_TS[,4])

Temperatures_TS_Rome <- diff(Temperatures_TS[,5])


par(mfrow = c(2,3))

autoplot(Temperatures_TS_Berlin)
```

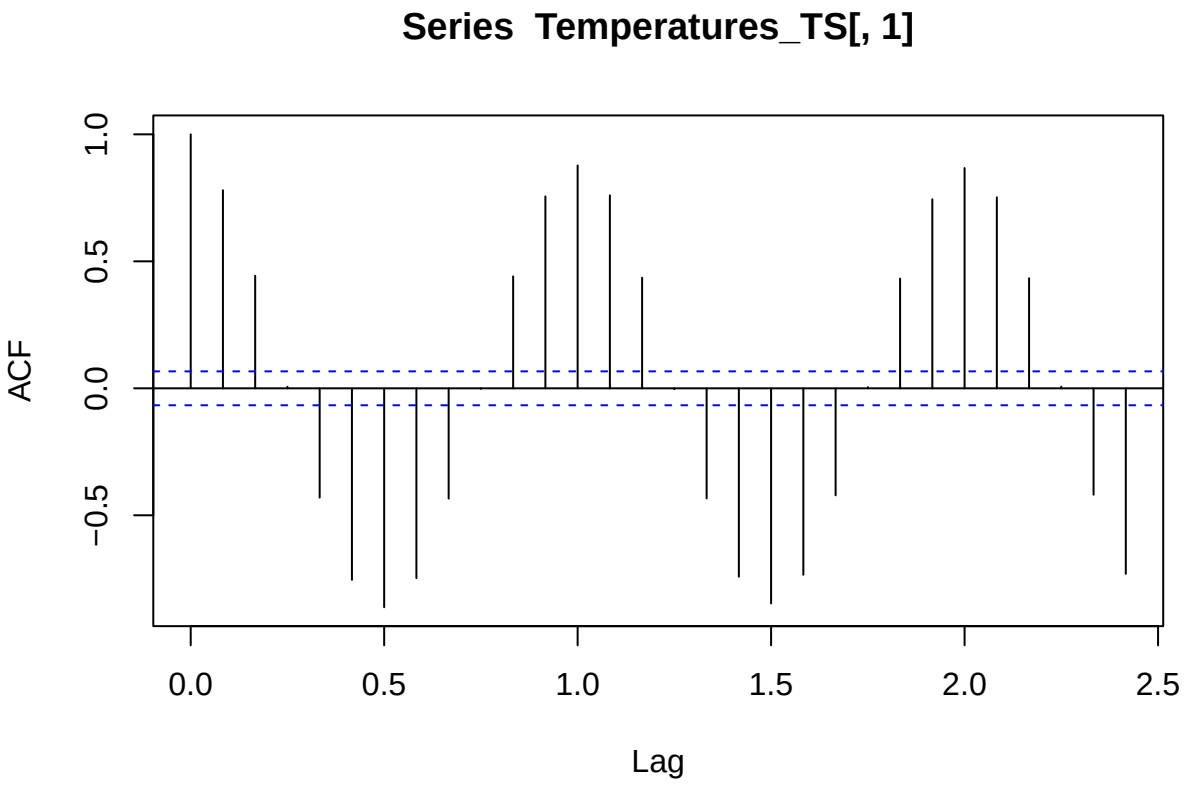



```
boxcox(lm(Temperatures_TS[,1] ~ 1))
```



When rounding to the closest *lambda* value, $lambda = 1$. Therefore, we do not need to make any BoxCox transformations to the data.

```
acf(Temperatures_TS[,1])
```


Series Temperatures_TS[, 1]
ACF
1.0
0.5
0.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
Lag