Assignment 2

Rohit Jyotiba Chougule & 19200240

04/11/2019

Question 1:

Complete your assignment using R Markdown, check that all the output and code are correctly shown in your final document. Knit your document frequently to fix errors. Once completed, submit the Rmd file and the resulting pdf or word document which shows all your code.

```
# No code here
```

Question 2:

Load in the data as an object called DublinAirport. Notice that you have to skip the first 19 lines contained in the file before beginning to read data. Line 20 contains the column names. Display the structure of the dataset.

```
DublinAirport <- read.csv(file = "mly532.csv", skip = 19)
#head(DublinAirport)
str(DublinAirport, strict.width="wrap")
```

```
## 'data.frame':    931 obs. of  12 variables:
## $ year : int 1941 1941 1942 1942 1942 1942 1942 1942 1942 1942 ...
## $ month: int 11 12 1 2 3 4 5 6 7 8 ...
## $ meant: num 6.9 6.5 4.3 2.9 6.3 8.4 10.4 13.1 14.6 14.9 ...
## $ maxtp: num 14 12.7 11.9 11.6 16.2 16.2 20.9 24.1 22.2 22.3 ...
## $ mintp: num -3.1 -3.6 -3.1 -4.3 -6.1 0.8 1.8 1.4 7.2 6.7 ...
## $ mnmax: num 9.9 9.1 6.9 5.8 9.4 11.9 14.4 18 18.9 18.4 ...
## $ mnmin: num 3.9 3.9 1.7 0 3.2 4.9 6.3 8.2 10.4 11.4 ...
## $ rain : num 67.2 41.7 91.9 25.8 76.4 ...
## $ gmin : num -5.7 -7.6 -9.5 -10.7 -8.3 -0.4 -0.7 -0.9 2.4 4.6 ...
## $ wdsp : num 12 12.5 13.1 9 10.7 15.1 12 9.4 13.4 10.8 ...
## $ maxgt: int NA NA NA NA NA NA NA NA NA NA ...
## $ sun : num 56.1 46.1 72.8 51.4 73.9 ...
```

```
head(DublinAirport$month)
```

```
## [1] 11 12  1  2  3  4
```

Question 3:

Transform the column months to a factor with labels given by the month names. [Two hints: (1) look at the arguments levels and labels of the function factor. (2) you may want to make use of the built-in constant month.name]

```
#?factor
Month_Name <- factor(DublinAirport$month, labels = month.name)
DublinAirport$month <- Month_Name
head(DublinAirport)
```

```
##   year    month meant maxtp mintp mnmax mnmin rain  gmin wdsp maxgt   sun
## 1 1941 November   6.9  14.0  -3.1   9.9   3.9 67.2  -5.7 12.0    NA  56.1
## 2 1941 December   6.5  12.7  -3.6   9.1   3.9 41.7  -7.6 12.5    NA  46.1
## 3 1942  January   4.3  11.9  -3.1   6.9   1.7 91.9  -9.5 13.1    NA  72.8
## 4 1942 February   2.9  11.6  -4.3   5.8   0.0 25.8 -10.7  9.0    NA  51.4
## 5 1942    March   6.3  16.2  -6.1   9.4   3.2 76.4  -8.3 10.7    NA  73.9
## 6 1942    April   8.4  16.2   0.8  11.9   4.9 36.9  -0.4 15.1    NA 185.4
```

```
#head(DublinAirport$month)
```

Question 4:

Use the aggregate function to compute which month has on average the highest and the lowest Precipitation Amount.

```
head(DublinAirport)
```

```
##   year    month meant maxtp mintp mnmax mnmin rain  gmin wdsp maxgt   sun
## 1 1941 November   6.9  14.0  -3.1   9.9   3.9 67.2  -5.7 12.0    NA  56.1
## 2 1941 December   6.5  12.7  -3.6   9.1   3.9 41.7  -7.6 12.5    NA  46.1
## 3 1942  January   4.3  11.9  -3.1   6.9   1.7 91.9  -9.5 13.1    NA  72.8
## 4 1942 February   2.9  11.6  -4.3   5.8   0.0 25.8 -10.7  9.0    NA  51.4
## 5 1942    March   6.3  16.2  -6.1   9.4   3.2 76.4  -8.3 10.7    NA  73.9
## 6 1942    April   8.4  16.2   0.8  11.9   4.9 36.9  -0.4 15.1    NA 185.4
```

```
Avg_Precipitation <- aggregate(DublinAirport$rain, list(DublinAirport$month), mean, na.rm=TRUE)
Avg_Precipitation <- order(Avg_Precipitation[,2])

print(paste0(month.name[Avg_Precipitation[length(Avg_Precipitation)]], " is the month with average highe
```

```
## [1] "December is the month with average highest precipitation amount"
```

```
print(paste0(month.name[Avg_Precipitation[1]], " is the month with average lowest precipitation amount"
```

```
## [1] "February is the month with average lowest precipitation amount"
```

Question 5

Create a new column which contains a factor indicating the season:

- Winter: December, January, February,
- Spring: March, April, May,

- Summer: June, July, August,
- Autumn: September, October, November

[Hint: again, look at the arguments levels and labels of the function factor.]

```
#install.packages("zoo")
library(zoo)
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
Year_qtr <- as.yearqtr(as.numeric(DublinAirport$month))/12
DublinAirport$season <- factor(format(Year_qtr, "%q"), levels = 1:4, labels = c("Winter", "Spring", "Su
head(DublinAirport$season)
```

```
## [1] Autumn Winter Winter Winter Spring Spring
## Levels: Winter Spring Summer Autumn
```

```
#head(DublinAirport)
```

Question 6

```
class(DublinAirport) <- c('WeatherData', 'data.frame')

str(DublinAirport)
```

```
## Classes 'WeatherData' and 'data.frame':  931 obs. of  13 variables:
##  $ year  : int  1941 1941 1942 1942 1942 1942 1942 1942 1942 1942 ...
##  $ month : Factor w/ 12 levels "January","February",..: 11 12 1 2 3 4 5 6 7 8 ...
##  $ meant : num  6.9 6.5 4.3 2.9 6.3 8.4 10.4 13.1 14.6 14.9 ...
##  $ maxtp : num  14 12.7 11.9 11.6 16.2 16.2 20.9 24.1 22.2 22.3 ...
##  $ mintp : num  -3.1 -3.6 -3.1 -4.3 -6.1 0.8 1.8 1.4 7.2 6.7 ...
##  $ mnmax : num  9.9 9.1 6.9 5.8 9.4 11.9 14.4 18 18.9 18.4 ...
##  $ mnmin : num  3.9 3.9 1.7 0 3.2 4.9 6.3 8.2 10.4 11.4 ...
##  $ rain  : num  67.2 41.7 91.9 25.8 76.4 ...
##  $ gmin  : num  -5.7 -7.6 -9.5 -10.7 -8.3 -0.4 -0.7 -0.9 2.4 4.6 ...
##  $ wdsp  : num  12 12.5 13.1 9 10.7 15.1 12 9.4 13.4 10.8 ...
##  $ maxgt : int  NA NA NA NA NA NA NA NA NA NA ...
##  $ sun   : num  56.1 46.1 72.8 51.4 73.9 ...
##  $ season: Factor w/ 4 levels "Winter","Spring",..: 4 1 1 1 2 2 2 3 3 3 ...
```

Question 7

Write an S3 summary method for an object of class WeatherData which produces the following statistical summaries the rain, maxtp, mintp, maxgt variables split by season: mean, standard deviation, minimum, maximum. Ignore the missing values in the calculations.

Test your function on the DublinAirport data set and comment your findings.

```
#?data.frame
summary.WeatherData <- function(Object)
{
  Season <- c('winter', 'spring', 'summer', 'autumn')

  temp_df <- data.frame(Season)

  rain <- split(Object$rain, Object$season)
  maxtp <- split(Object$maxtp, Object$season)
  mintp <- split(Object$mintp, Object$season)
  maxgt <- split(Object$maxgt, Object$season)

  summary_func <- c('mean', 'sd', 'max', 'min')

  for(i in 1:length(summary_func)){
    temp_df[paste('rain.',summary_func[i])] <- data.frame(sapply(rain, summary_func[i], na.rm=TRUE))
    temp_df[paste('maxtp.',summary_func[i])] <- data.frame(sapply(maxtp, summary_func[i], na.rm=TRUE))
    temp_df[paste('mintp.',summary_func[i])] <- data.frame(sapply(mintp, summary_func[i], na.rm=TRUE))
    temp_df[paste('maxgt.',summary_func[i])] <- data.frame(sapply(maxgt, summary_func[i], na.rm=TRUE))
  }
  print(temp_df, row.names = FALSE)
}

summary.WeatherData(DublinAirport)
```

```
 Season rain. mean maxtp. mean mintp. mean maxgt. mean rain. sd maxtp. sd
 winter   63.89017    12.73761  -3.1931624    53.04933 32.87200  1.638288
 spring   53.54915    17.16239  -0.5871795    45.57333 27.30880  2.798734
 summer   63.92165    23.02511   5.9303030    39.49333 35.43135  1.787048
 autumn   70.20948    18.02198   1.2551724    47.23661 38.11011  3.038054
 mintp. sd maxgt. sd rain. max maxtp. max mintp. max maxgt. max rain. min
  2.369322  8.767892     217.0       17.1        1.8         80       4.7
  2.401950  7.426136     151.8       23.5        6.9         66       3.6
  1.735196  6.263642     189.9       28.7       10.0         56       4.0
  2.799624  7.875278     185.8       25.1        7.5         73       3.6
 maxtp. min mintp. min maxgt. min
        4.8      -12.2         35
        9.9       -7.9         28
       18.4        0.7         27
       12.1       -8.4         27
```

Comments from the summarized data:

From the summary of the data for all these years, it seems that the highest rainfall in Dublin is during the with winters, where as the mean rainfall during the overall seasons seem to be same for winters and summers.

However, its during the autumn that the overall season mean is the highest. The mean temperature in Dublin ranges from -3 to 23 during overall seasons. The lowest ofcourse during winters, mean being around -3. The maximum temperature during summers goes up to 23. The mean windspeed seems to be almost the same across all seasons, however, summer being the lowest and winter being the highest.

Question 8

Create an S3 plot method for the class WeatherData that produces the following plots.

(1) Plot of the monthly Air Temperature (C) (maxtp, mintp).

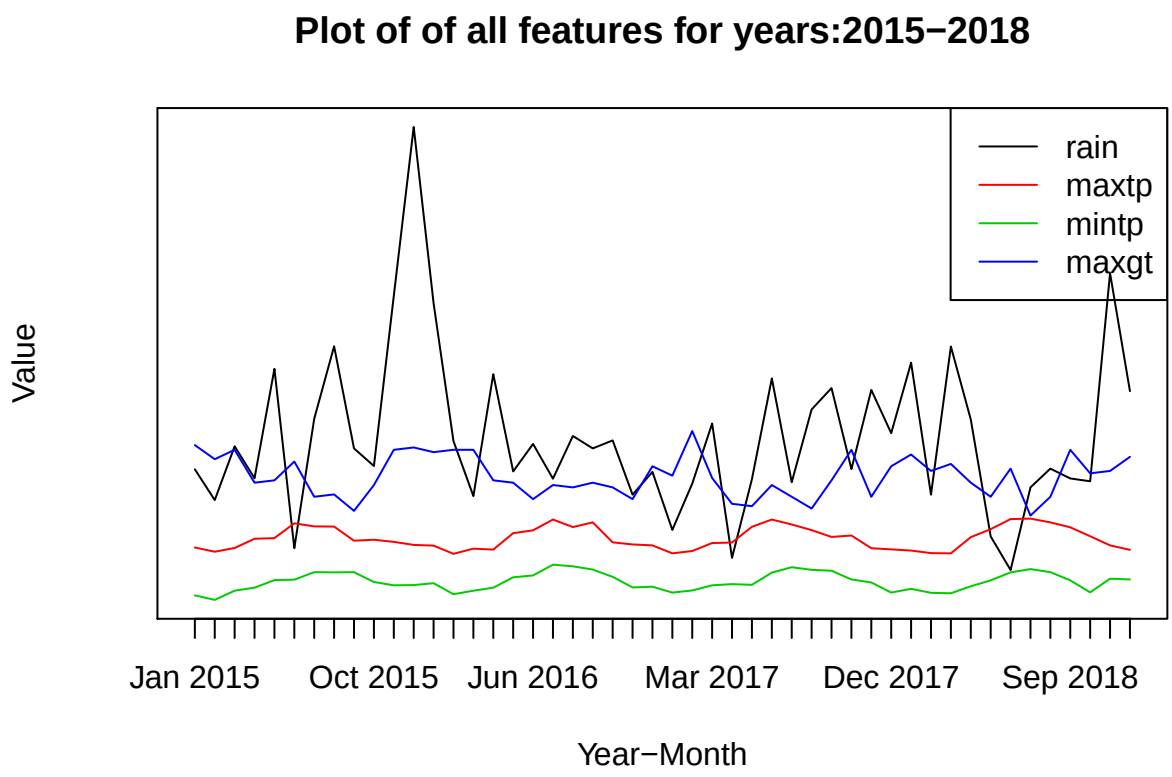
(2) Plot of the Precipitation Amount (mm) (rain).

(3) Plot of the Highest Gust (knot) (maxgt).

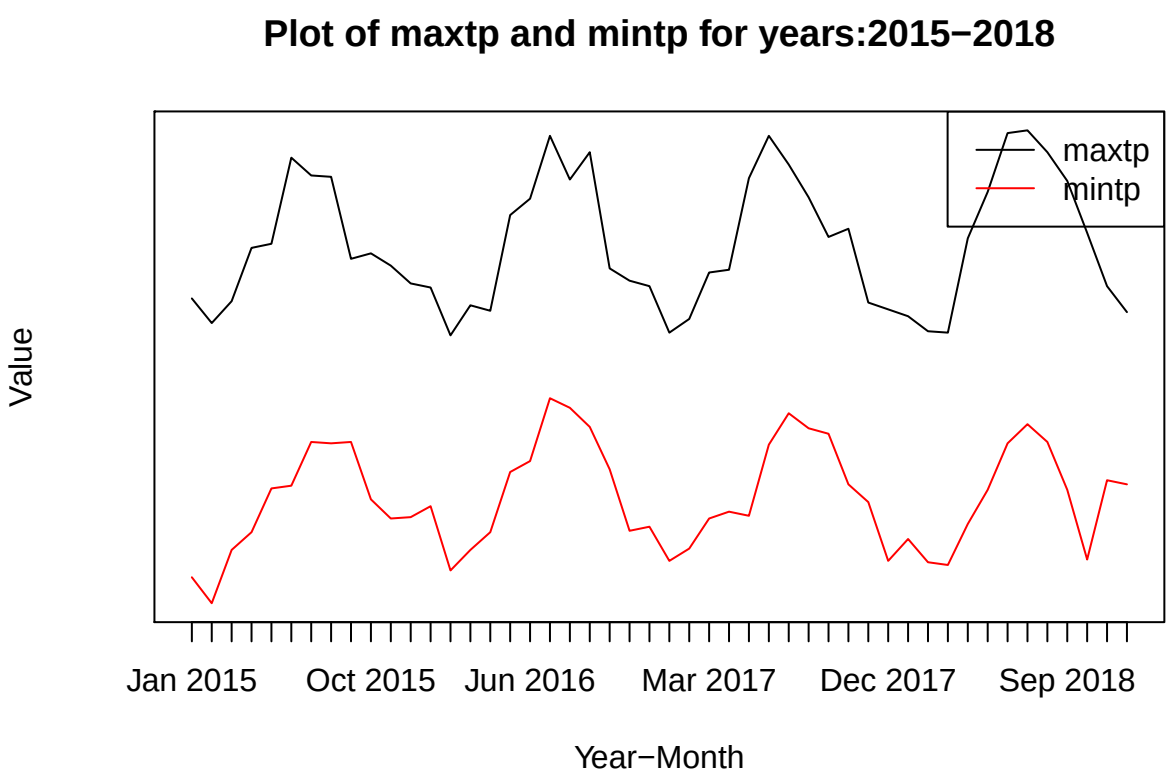
- The user must be able to decide which years to plot. By default it will use the data from 2015 until 2018.
- The user must be able to decide which plot to draw (i.e, only one of the three, two of the three, or all three plots). By default the function will create all three plots.
- The plots must be on a single panel.
- The plots must have meaningful labels and/or titles, and a legend if needed.
- Test your function on the DublinAirport data set.

```
plot.WeatherData <- function(my_Object, year_start = 2015, year_end=2018, plot_data = 0){

temp_data <- my_Object[which(my_Object$year >= year_start & my_Object$year <= year_end),]

title <- ''
xlabel <- 'Year-Month'
ylabel <- 'Value'

if (plot_data == 0){
temp_data <- temp_data[,c('rain', 'maxtp', 'mintp', 'maxgt')]
title <- paste0('Plot of of all features for years:',
year_start,'-',year_end)
}

if (plot_data == 1){
temp_data <- temp_data[,c('maxtp', 'mintp')]
title <- paste0('Plot of maxtp and mintp for years:',
year_start,'-',year_end)
}

if (plot_data == 2){
temp_data <- temp_data[,c('rain'), drop=FALSE]
title <- paste0('Plot of rain between for years:',
year_start,'-',year_end)
}

if (plot_data == 3){
```

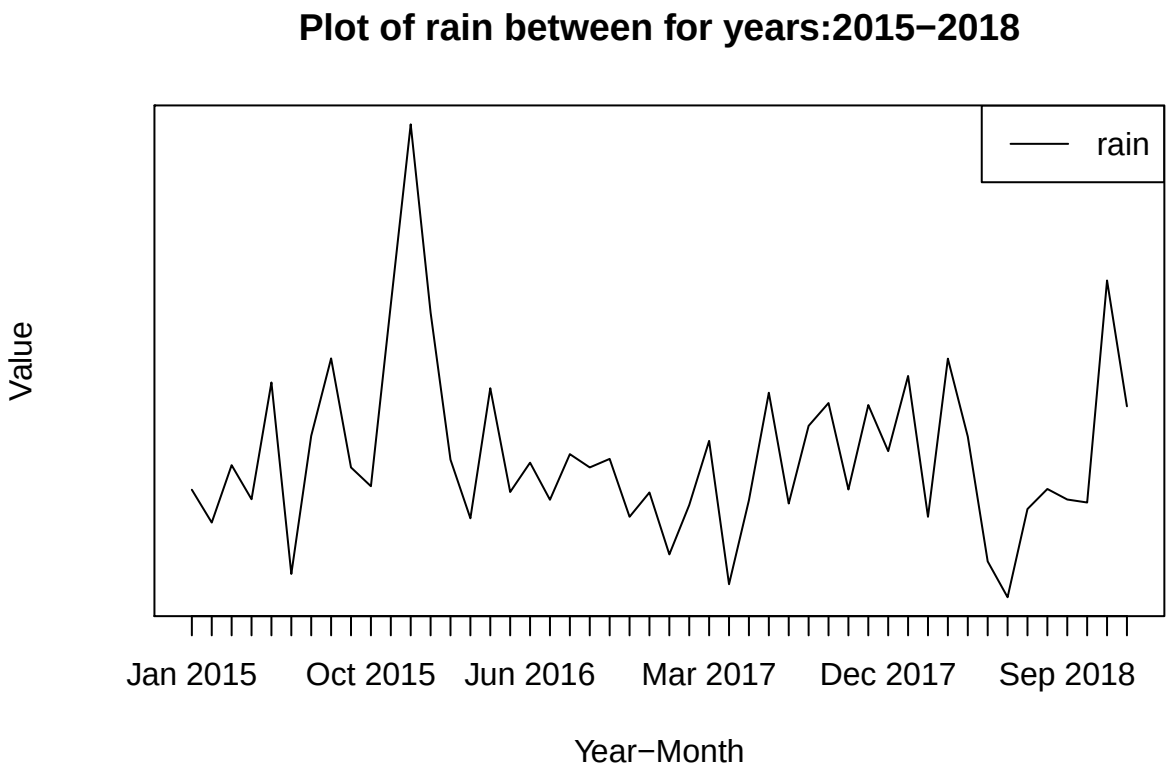
```
temp_data <- temp_data[,c('maxgt'), drop=FALSE]
title <- paste0('Plot of maxgt for years:',
year_start,'-',year_end)
}

DataSet.TS <- ts(temp_data,start=c(year_start,1),
end = c(year_end,12),frequency=12)

plot(DataSet.TS, plot.type="single", col = 1:ncol(DataSet.TS), main = title, xlab = xlabel,
ylab = ylabel , at = as.yearmon(time(DataSet.TS)), labels =as.yearmon(time(DataSet.TS)))
legend("topright", colnames(DataSet.TS), col=1:ncol(temp_data), lty=1, cex=1)
}

# When plot_data = 0, the default condition, all graphs are printed
#When the user sets plot_data = 1 it plots monthly data for air temperature
#When the user sets plot_data = 2 it plots monthly data for rain (percipitation) amount in mm
#When the user sets plot_data = 3 it plots monthly data for highest gust in knots

plot.WeatherData(DublinAirport, 2015, 2018, 0)
```



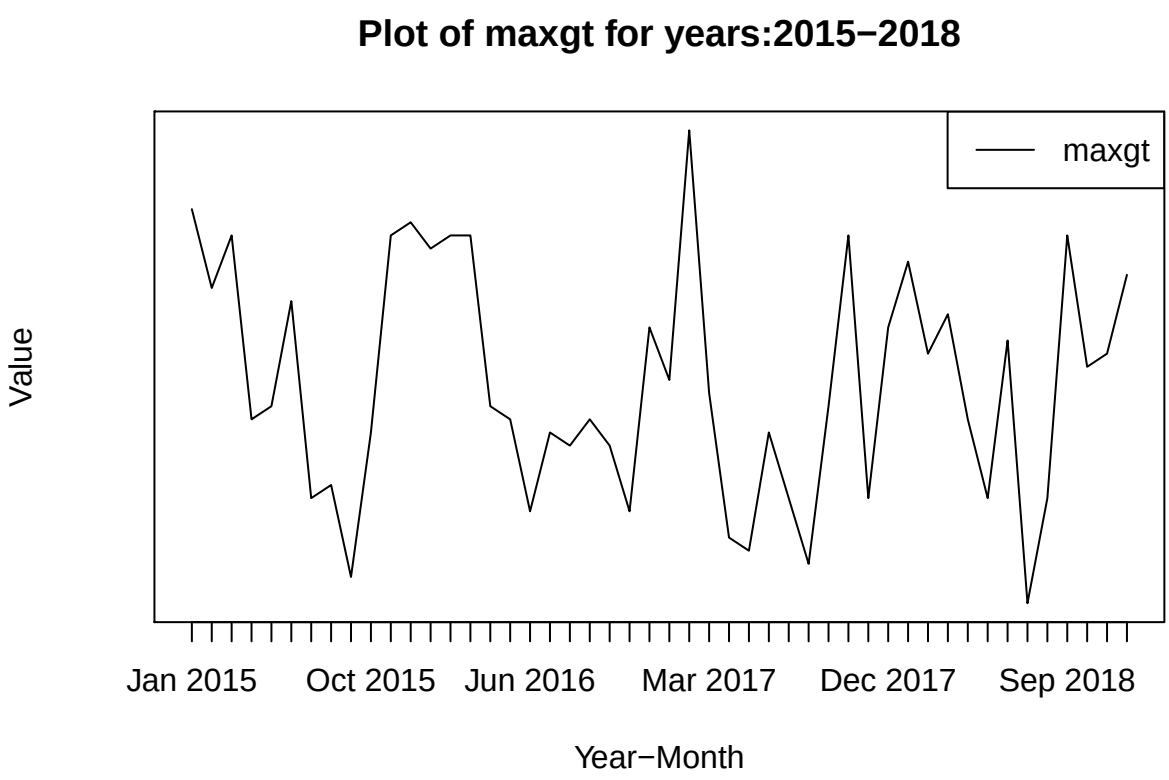

```
plot.WeatherData(DublinAirport, 2015, 2018, 1)
```




```
plot.WeatherData(DublinAirport, 2015, 2018, 2)
```




```
plot.WeatherData(DublinAirport, 2015, 2018, 3)
```


Plot of maxgt for years:2015−2018
maxgt
Value
Jan 2015
Oct 2015
Jun 2016
Mar 2017
Dec 2017
Sep 2018
Year−Month