Assignment 2

Ranjith_Dhanaraj_19200041

11/13/2019

Loading the dataset while skipping the first 19 lines.

```
DublinAirport = read.csv('mly532.csv',skip=19)
str(DublinAirport)
```

```
## '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 ...
```

Transform the column ‘months’ to lables with column names.

```
DublinAirport$month <- factor(DublinAirport$month, levels = c(1:12), labels = c(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
```

Month with the highest and the lowest percipiation amount.

```
aggregate_rain = aggregate(DublinAirport$rain, by= list(DublinAirport$month), FUN = "mean")
rownames(aggregate_rain)<-aggregate_rain$Group.1
aggregate_rain$Group.1 <- 0
highest_percipitation_month <- row.names(aggregate_rain[which.max(aggregate_rain$x),])
sprintf("The month highest precipitation is %s", highest_percipitation_month)
```

```
## [1] "The month highest precipitation is December"
```

```
lowest_percipitation_month <- row.names(aggregate_rain[which.min(aggregate_rain$x),])
sprintf("The month lowest precipitation is %s", lowest_percipitation_month)
```

```
## [1] "The month lowest precipitation is February"
```

Creating a new column 'season' based on the month.

Importing data manipulation library

```r
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

Changing the 'DublinAirport' dataframe by adding season as an additional column

```r
DublinAirport <- mutate(DublinAirport,
                        Season = case_when(
                          month %in% c("December","January","February") ~ "Winter",
                          month %in% c("March","April","May") ~ "Spring",
                          month %in% c("June","July","August") ~ "Summer",
                          month %in% c("September","October","November") ~ "Autumn",
                          TRUE ~ "NA"))

DublinAirport$Season<-factor(DublinAirport$Season,
                             levels=c("Winter","Spring","Summer","Autumn"))

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
##   Season
## 1 Autumn
## 2 Winter
## 3 Winter
## 4 Winter
## 5 Spring
## 6 Spring
```

Custom S3 Class creation

Assigning the classes 'WeatherData' and 'data.frame' to the dataframe 'DublinAirport'.

```
class(DublinAirport) <- c('WeatherData', class(DublinAirport))
class(DublinAirport)
```

```
## [1] "WeatherData" "data.frame"
```

Creating a S3 summary method for an object in WeatherData class.

```
summary.WeatherData <- function(object){
  object %>%
  group_by(Season) %>%
  select(rain, maxtp, mintp, maxgt) %>%
  summarise_all(funs(mean, max, sd), na.rm = TRUE)
}
summary.WeatherData(DublinAirport)
```

```
## Adding missing grouping variables: `Season`
```

```
## Warning: funs() is soft deprecated as of dplyr 0.8.0
## Please use a list of either functions or lambdas:
##
##   # Simple named list:
##   list(mean = mean, median = median)
##
##   # Auto named with `tibble::lst()`:
##   tibble::lst(mean, median)
##
##   # Using lambdas
##   list(~ mean(., trim = .2), ~ median(., na.rm = TRUE))
## This warning is displayed once per session.
```

```
## # A tibble: 4 x 13
##   Season rain_mean maxtp_mean mintp_mean maxgt_mean rain_max maxtp_max
##   <fct>      <dbl>      <dbl>      <dbl>      <dbl>    <dbl>     <dbl>
## 1 Winter      63.9       12.7     -3.19        53.0     217       17.1
## 2 Spring      53.5       17.2     -0.587       45.6     152.      23.5
## 3 Summer      63.9       23.0      5.93        39.5     190.      28.7
## 4 Autumn      70.2       18.0      1.26        47.2     186.      25.1
## # ... with 6 more variables: mintp_max <dbl>, maxgt_max <int>,
## #   rain_sd <dbl>, maxtp_sd <dbl>, mintp_sd <dbl>, maxgt_sd <dbl>
```

A custom summary function for objects belonging to WeatherData class has been created. We can observe that the highest mean amount of rainfall occurs in Autumn. When we take temperature into consideration, the hottest season tends to be the summer even though it has a good amount of rainfall, whereas, Spring has the lowest rainfall yet the mean temperature is less than I expected it to be. Winter has the highest gust speed on an average while Autumn comes in second. Summer and Autmn seem to have the highest range of temperature swings. Highest amount of rainfall has occured during the winter season even though the mean rainfall is less when compared to Autumn. Winter and Autumn are the seasons whcih rank first and second respectively when we take the highest gust speed. This corresponds well with the fact that they are the windiest on an average as well. Mean temperatures are of the same ranking as that of the Max and min temperature and there is nothing so special about it. Standard deviation of rain, max temp, min temp, max gust speed and min gust speed, all correlate to their mean, max and min counterparts.

Plotting graphs for 'WeatherData' Class.

Importing the library that helps us arrange graphs

```
library(gridExtra)
```

```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##     combine
```

Now, we will import the library that will help plot the graphs.

```
library(ggplot2)
```

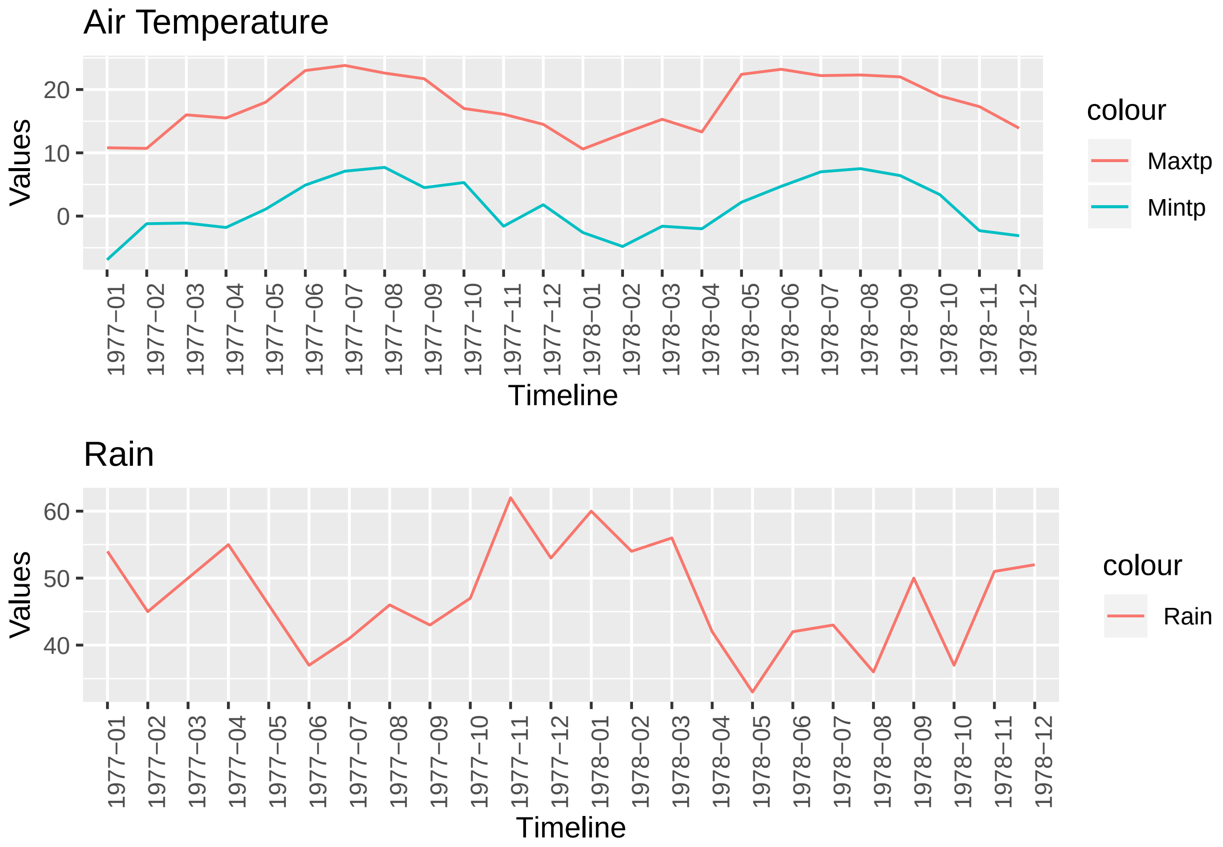
Plotting begins in the following code -

```
plot.WeatherData <- function(df, start_year = 2015, end_year = 2018,
                             plot1 = TRUE, plot2 = TRUE, plot3 = TRUE ) {
  df_sub <- subset(df, df$year >= start_year & df$year <= end_year)
  data <- data.frame(Timeline = with(df_sub,

  sprintf("%d-%02d", df_sub$year, df_sub$month)),

  mintp = df_sub$mintp,
  maxtp = df_sub$maxtp,
  rain = df_sub$rain,
  maxgt = df_sub$maxgt)
  q <- ggplot(data, aes(x= Timeline)) +
  theme(axis.text.x = element_text(angle = 90, hjust = 1),
        legend.position="right") + labs(y = "Values")
  if(plot1 == TRUE & plot2 == FALSE & plot3 == FALSE ){
    q <- q +
    geom_line(aes(y = mintp, group = 1, color = "Mintp")) +
    geom_line(aes(y = maxtp, group = 1, color = "Maxtp")) +
    labs( title = 'Air Temperature')
    q
  }
  else if(plot1 == FALSE & plot2 == TRUE & plot3 == FALSE ){
    q <- q +
    geom_line(aes(y = rain, group = 1, color = "Rain")) +
    labs( title = 'Rain')
    q
  }
  else if(plot1 == FALSE & plot2 == FALSE & plot3 == TRUE ){
    q <- q +
    geom_line(aes(y = maxgt, group = 1, color = "Gust")) +
    labs( title = 'Gust')
    q
  }
  else if(plot1 == TRUE & plot2 == TRUE & plot3 == FALSE ){
    q1 <- q +
    geom_line(aes(y = mintp, group = 1, color = "Mintp")) +
    geom_line(aes(y = maxtp, group = 1, color = "Maxtp")) +
    labs( title = 'Air Temperature')
    q2 <- q +
    geom_line(aes(y = maxgt, group = 1, color = "Rain")) +
    labs( title = 'Rain')
    grid.arrange(q1,q2)
  }
  else if(plot1 == TRUE & plot2 == FALSE & plot3 == TRUE ){
    q1 <- q +
    geom_line(aes(y = mintp, group = 1, color = "Mintp")) +
    geom_line(aes(y = maxtp, group = 1, color = "Maxtp")) +
    labs( title = 'Air Temperature')
    q2 <- q +
    geom_line(aes(y = maxgt, group = 1, color = "Gust")) +
    labs( title = 'Gust')
    grid.arrange(q1,q2)
  }
  else if(plot1 == FALSE & plot2 == TRUE & plot3 == TRUE ){
    q1 <- q +
    geom_line(aes(y = rain, group = 1, color = "rain")) +
    labs( title = 'Rain')
    q2 <- q +
    geom_line(aes(y = maxgt, group = 1, color = "Gust")) +
    labs( title = 'Gust')
    grid.arrange(q1,q2)
  }
  else{
    q1 <- q +
    geom_line(aes(y = mintp, group = 1, color = "Mintp")) +
    geom_line(aes(y = maxtp, group = 1, color = "Maxtp")) +
    labs( title = 'Air Temperature')
    q2 <- q +
    geom_line(aes(y = rain, group = 1, color = "Rain")) +
    labs( title = 'Rain')
    q3 <- q +
    geom_line(aes(y = maxgt, group = 1, color = "Gust")) +
    labs( title = 'Gust')
    grid.arrange(q1,q2,q3)
  }
}
```

We have now defined the plotting function for the ‘WeatherData’ class.

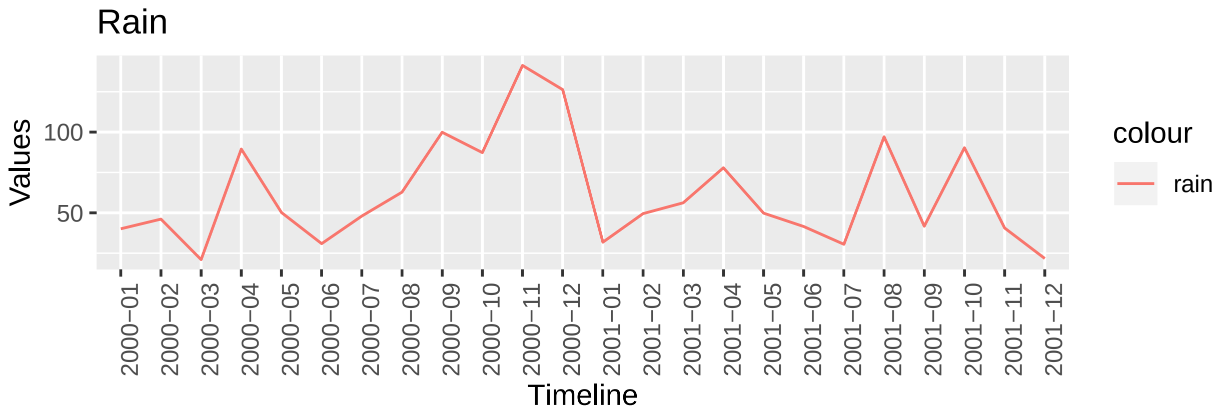
Plot 1

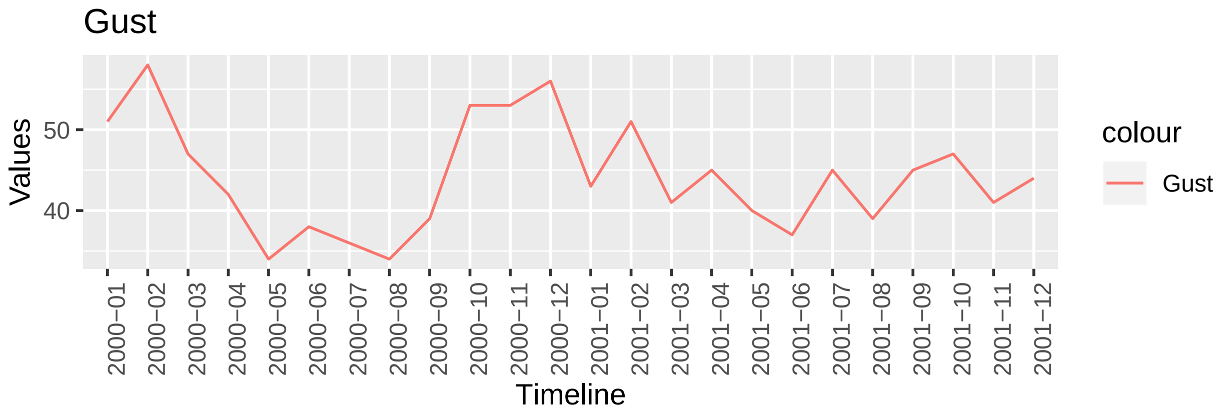
```
plot.WeatherData(DublinAirport, 1977, 1978, TRUE, TRUE, FALSE)
```



Plot 2

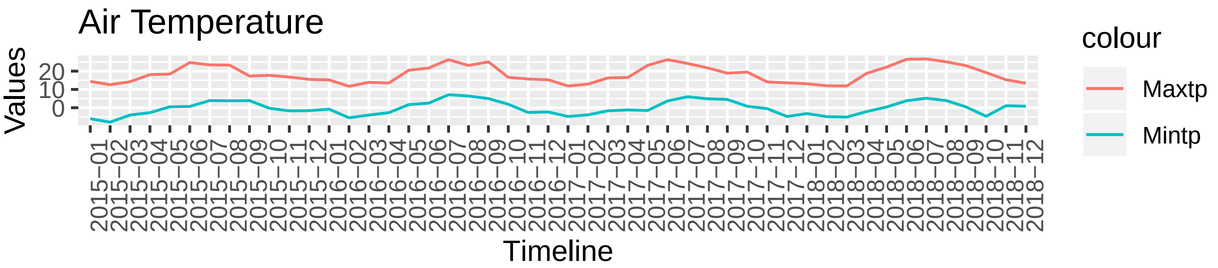
```
plot.WeatherData(DublinAirport, 2000, 2001, FALSE, TRUE, TRUE)
```

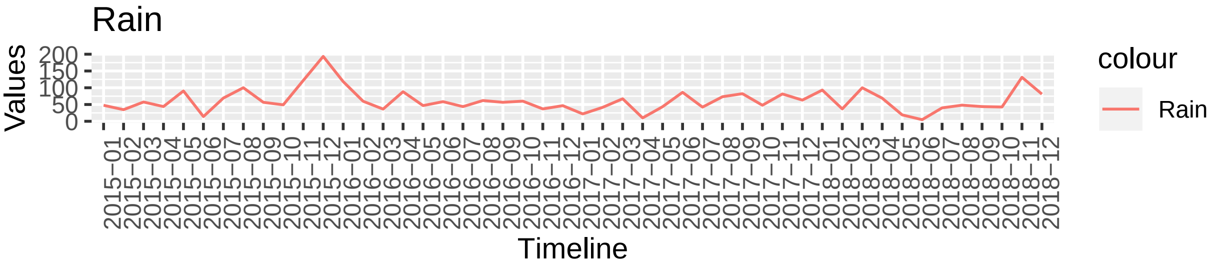


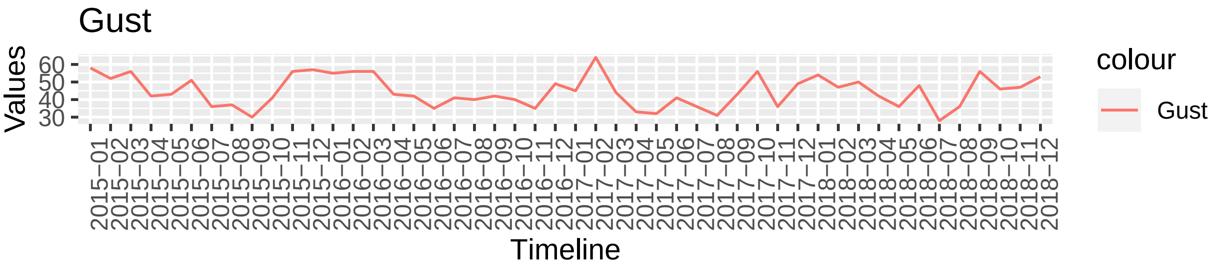


PLot 3

```r
plot.WeatherData(DublinAirport)
```


Air Temperature
Values
20
10
0
2015−01
2015−02
2015−03
2015−04
2015−05
2015−06
2015−07
2015−08
2015−09
2015−10
2015−11
2015−12
2016−01
2016−02
2016−03
2016−04
2016−05
2016−06
2016−07
2016−08
2016−09
2016−10
2016−11
2016−12
2017−01
2017−02
2017−03
2017−04
2017−05
2017−06
2017−07
2017−08
2017−09
2017−10
2017−11
2017−12
2018−01
2018−02
2018−03
2018−04
2018−05
2018−06
2018−07
2018−08
2018−09
2018−10
2018−11
2018−12
Timeline
colour
Maxtp
Mintp


Rain
Values
200
150
100
50
0
2015−01
2015−02
2015−03
2015−04
2015−05
2015−06
2015−07
2015−08
2015−09
2015−10
2015−11
2015−12
2016−01
2016−02
2016−03
2016−04
2016−05
2016−06
2016−07
2016−08
2016−09
2016−10
2016−11
2016−12
2017−01
2017−02
2017−03
2017−04
2017−05
2017−06
2017−07
2017−08
2017−09
2017−10
2017−11
2017−12
2018−01
2018−02
2018−03
2018−04
2018−05
2018−06
2018−07
2018−08
2018−09
2018−10
2018−11
2018−12
Timeline
colour
Rain


Gust
Values
60
50
40
30
2015−01
2015−02
2015−03
2015−04
2015−05
2015−06
2015−07
2015−08
2015−09
2015−10
2015−11
2015−12
2016−01
2016−02
2016−03
2016−04
2016−05
2016−06
2016−07
2016−08
2016−09
2016−10
2016−11
2016−12
2017−01
2017−02
2017−03
2017−04
2017−05
2017−06
2017−07
2017−08
2017−09
2017−10
2017−11
2017−12
2018−01
2018−02
2018−03
2018−04
2018−05
2018−06
2018−07
2018−08
2018−09
2018−10
2018−11
2018−12
Timeline
colour
Gust