Assignmnet 2

Raj Shekhar - 18200277

08/11/2019

Dublin Airport Weather Data Analysis

```
#(2):Loading in the data as an object called DublinAirport and skippping the first 19 lines cont
ained in the file.

DublinAirport <- read.csv("mly532.csv",header = TRUE,row.names = NULL,skip = 19)

#Displaying the structure of the dataset.

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 ...
```

```
#(3):Transform the column months to a factor with labels given by the month names.

DublinAirport$Date <-paste(DublinAirport$month,"-",DublinAirport$year)
DublinAirport$month <- factor(DublinAirport$month, levels = c("1", "2" , "3","4","5", "6" , "7",
"8","9", "10" , "11","12"),labels = month.name)

#Displaying the updated months colmun:
unique(DublinAirport$month)
```

```
##  [1] November  December  January   February  March     April     May
##  [8] June      July      August    September October
## 12 Levels: January February March April May June July August ... December
```

```
#(4):Computing which month has on average the highest and the lowest Precipitation Amount.

rainavg <- aggregate(DublinAirport$rain,by=list(DublinAirport$month),mean)

# Maximum Precipitation: December- 75.37564

subset(rainavg,rainavg$x == max(rainavg$x))
```

```
##     Group.1        x
## 12 December 75.37564
```

```
# Minimum Precipitation: February- 49.70256

subset(rainavg,rainavg$x == min(rainavg$x))
```

```
##    Group.1        x
## 2 February 49.70256
```

```
#(5) Creating 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

DublinAirport$season <- factor(DublinAirport$month,
       levels=c("December", "January", "February","March", "April", "May", "June","July", "Augus
t","September", "October", "November"),
       labels=c("Winter", "Winter", "Winter","Spring", "Spring", "Spring", "Summer","Summer", "S
ummer","Autumn", "Autumn", "Autumn"))

# Displaying the seasons:
unique(DublinAirport$season)
```

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

```
#(6). Assigning to the DublinAiport object the classes WeatherData and data.frame.
class(DublinAirport) <- c('WeatherData', 'data.frame')
class(DublinAirport)
```

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

```
#(7). S3 summary method for an object of class WeatherData which produces the statistical summar
ies: rain, maxtp, mintp, maxgt split by season

library(dplyr)
```

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

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

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

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

```
summary.WeatherData <- function(object){
    statssummary <- data.frame(object$rain,object$maxtp,object$mintp,object$maxgt)
    statssummary %>%
    group_by(object$season) %>%
    summarise_all(list(~ mean(., na.rm = TRUE), ~ sd(., na.rm = TRUE),~ max(., na.rm = TRUE), ~
 min(., na.rm = TRUE))) %>%
    rmarkdown::paged_table()
    }

# Applying "summary.WeatherData" function on the DublinAirport data set

summary.WeatherData(DublinAirport)
```

object$season <fctr>	object.rain_mean <dbl>	object.maxtp_mean <dbl>	object.mintp_mean <dbl>	object.maxgt_mea <db
Winter	63.89017	12.73761	-3.1931624	53.049
Spring	53.54915	17.16239	-0.5871795	45.573
Summer	63.92165	23.02511	5.9303030	39.493
Autumn	70.20948	18.02198	1.2551724	47.236

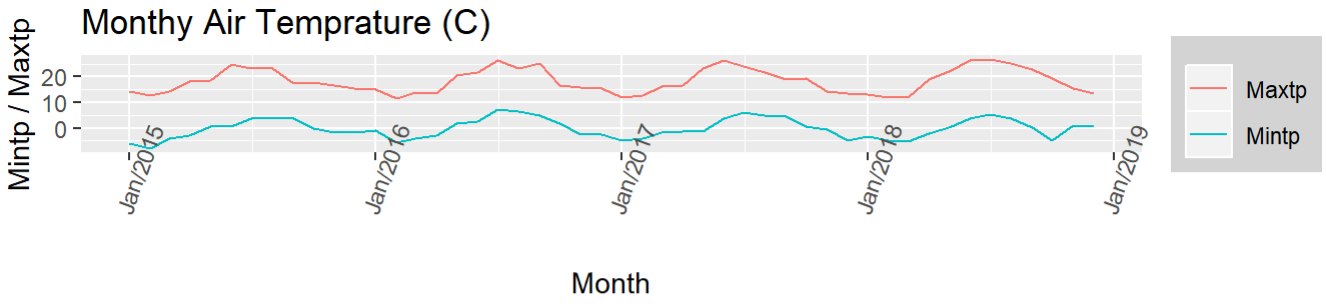
4 rows | 1-5 of 17 columns

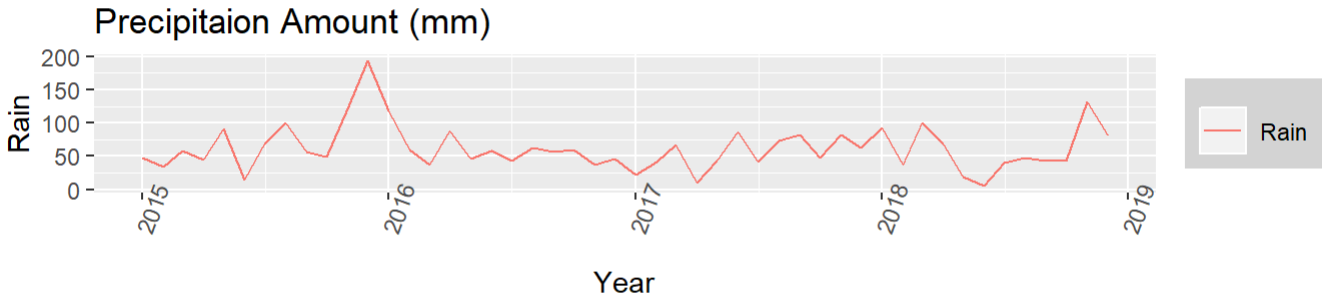
```
#Comments:
#(1) Winter: has lowest temprature among all season(as mintp_mean is -3.1931624) and has maximum
Gust and highest Precipitation Amount of 217
#(2) Autumn: has maximum rain(rain_mean is 70.20948)
#(3) Summer: has maximum temprature (maxtp mean is 23.02511) with highest air temrature of 28.7
 and low Highest Gust
#(4) Spring: has low value recorderd on an average for Rain and Maxgt in comparision to other se
asons.
```

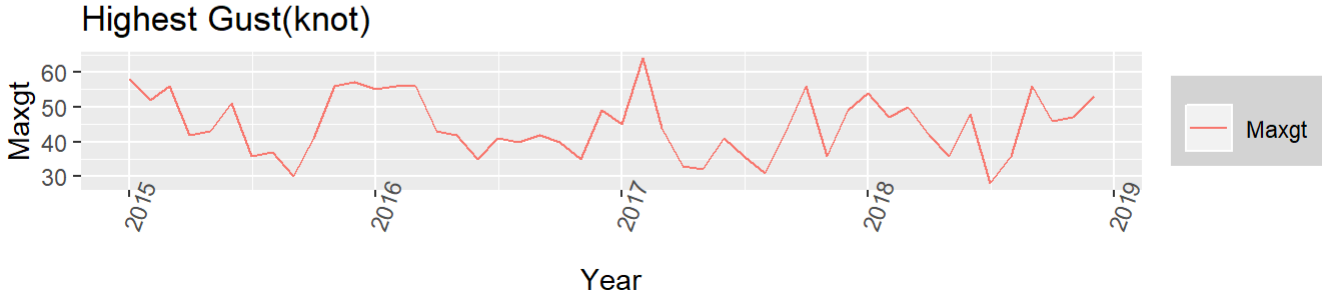
```
#S3 plot method for the class WeatherData that produces the 3 plots:

#DublinAirport$maxgt[is.na(x = DublinAirport$maxgt)] <- 0
plot.WeatherData <- function(startyear,endyear,plota,plotb,plotc) {
#By default it will use the data from 2015 until 2018 if no year is mentioned in user input
if(missing(startyear)|missing(endyear)){
 startyear=2015
 endyear=2018
}

#As per user input of statyear & endyear,subset of data will be created from orignal "DublinAirp
ort" data set and stored in "plotyears"
plotyears  <- data.frame(DublinAirport[DublinAirport$year >= startyear & DublinAirport$year <= e
ndyear,])

#Converting the subset plotyears date data in proper date fromat
plotyears$date <- as.Date(paste(plotyears$year,plotyears$month,"01",sep="-"),format = ("%Y-%b-%
d"))

#(1) Plot of the monthly Air Temperature (C) (maxtp, mintp).
plot1 <-ggplot(data=plotyears,aes(x=date,group=1))+
        geom_line(aes(y=plotyears$maxtp,color="Maxtp"))+
        geom_line(aes(y=plotyears$mintp,color="Mintp"))+
        scale_x_date(date_labels = "%b/%Y")+
        labs(x="Month",y="Mintp / Maxtp",title="Monthy Air Temprature (C)")+
        theme(axis.text.x = element_text(angle = 70),legend.background =element_rect(fill="light
grey",size=0.5, linetype="solid"),legend.title = element_blank())

#(2) Plot of the Precipitation Amount (mm) (rain).
plot2 <- ggplot(data=plotyears,aes(x=date,group=1))+
          geom_line(aes(y=plotyears$rain,color="Rain"))+
          scale_x_date(date_labels = "%Y")+
          theme(axis.text.x = element_text(angle = 70),legend.background =element_rect(fill="lig
htgrey",size=0.5, linetype="solid"),legend.title = element_blank())+
          labs(x="Year",y="Rain",title="Precipitaion Amount (mm)")

#(3) Plot of the Highest Gust (knot) (maxgt).
plot3 <- ggplot(data=plotyears,aes(x=date))+
         geom_line(aes(y=plotyears$maxgt,color="Maxgt"))+
         scale_x_date(date_labels = "%Y")+
         theme(axis.text.x = element_text(angle = 70),legend.background =element_rect(fill="ligh
tgrey",size=0.5, linetype="solid"),legend.title = element_blank())+
         labs(x="Year",y="Maxgt",title="Highest Gust(knot)")

#By default the function will create all three plots if no plot is mentioned in user input.

if(missing(plota) & missing(plotb) & missing(plotc)){
    plota=0
    plotb=0
    plotc=0
    par(mfrow=c(1,3))
 grid.arrange(plot1,
              plot2,
```

```
              plot3)
  }
#If user inputs plota=TRUE then plot1 will be shown
if(plota == TRUE){
  print(plot1)
}

#If user inputs plotb=TRUE then plot2 will be shown
if(plotb == TRUE){
   print(plot2)
}

#If user inputs plotc=TRUE then plot3 will be shown
if(plotc == TRUE){
  print(plot3)
}

}
```

```
#By default function will use the data from 2015 until 2018 end and will create all three plots
 if no input provided
plot.WeatherData()
```

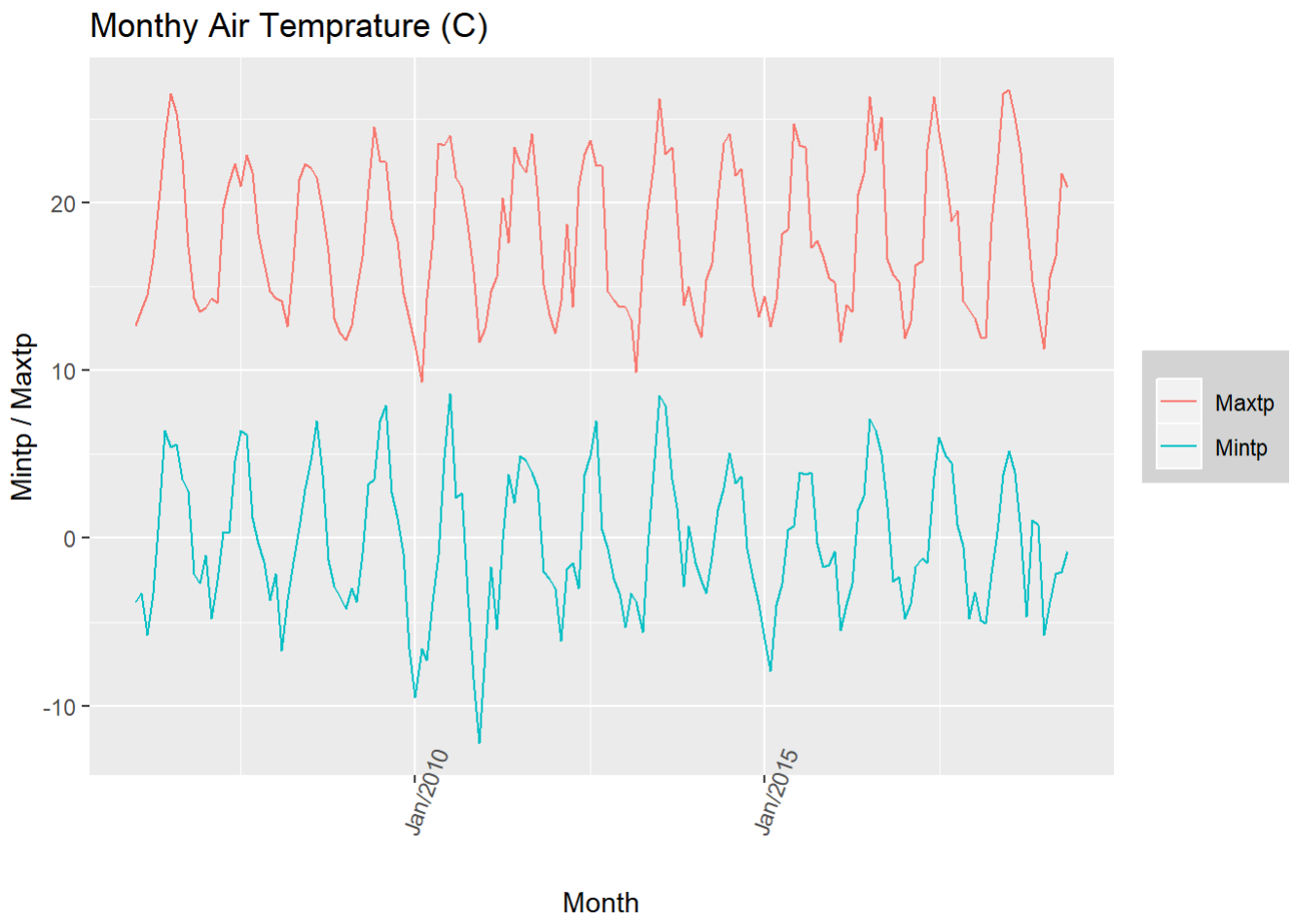






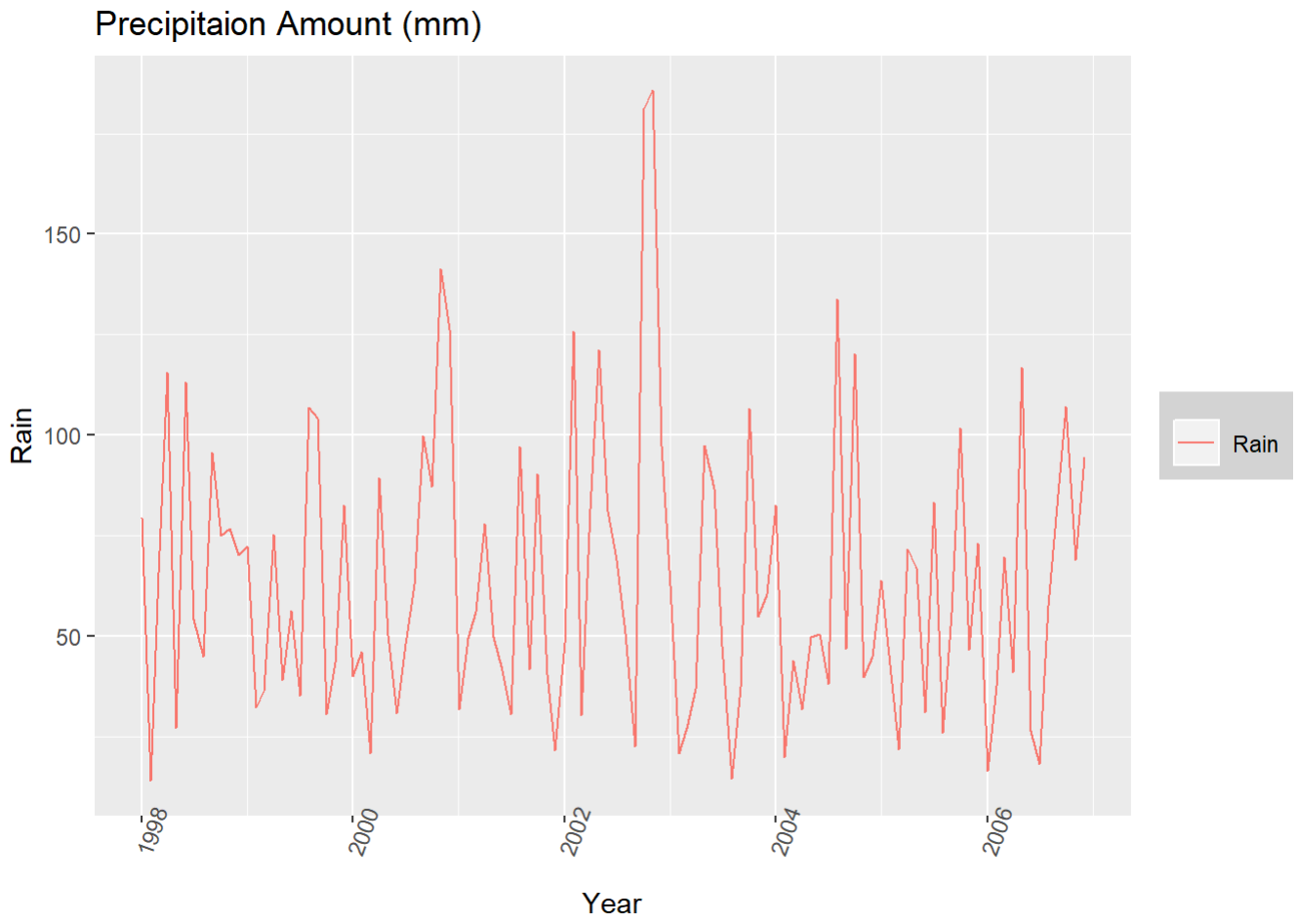
```
# Plot 1: Monthly Air temprature,the difference between maximum and minimum remains approximatle
y same as almost same distance can be seen between theor respective lines in the plot.
# Plot 2:Rain spiked around 2015 year end and is generally not varying much over the years
# Plot 3 : There is allot of variation over the year in Gust
```

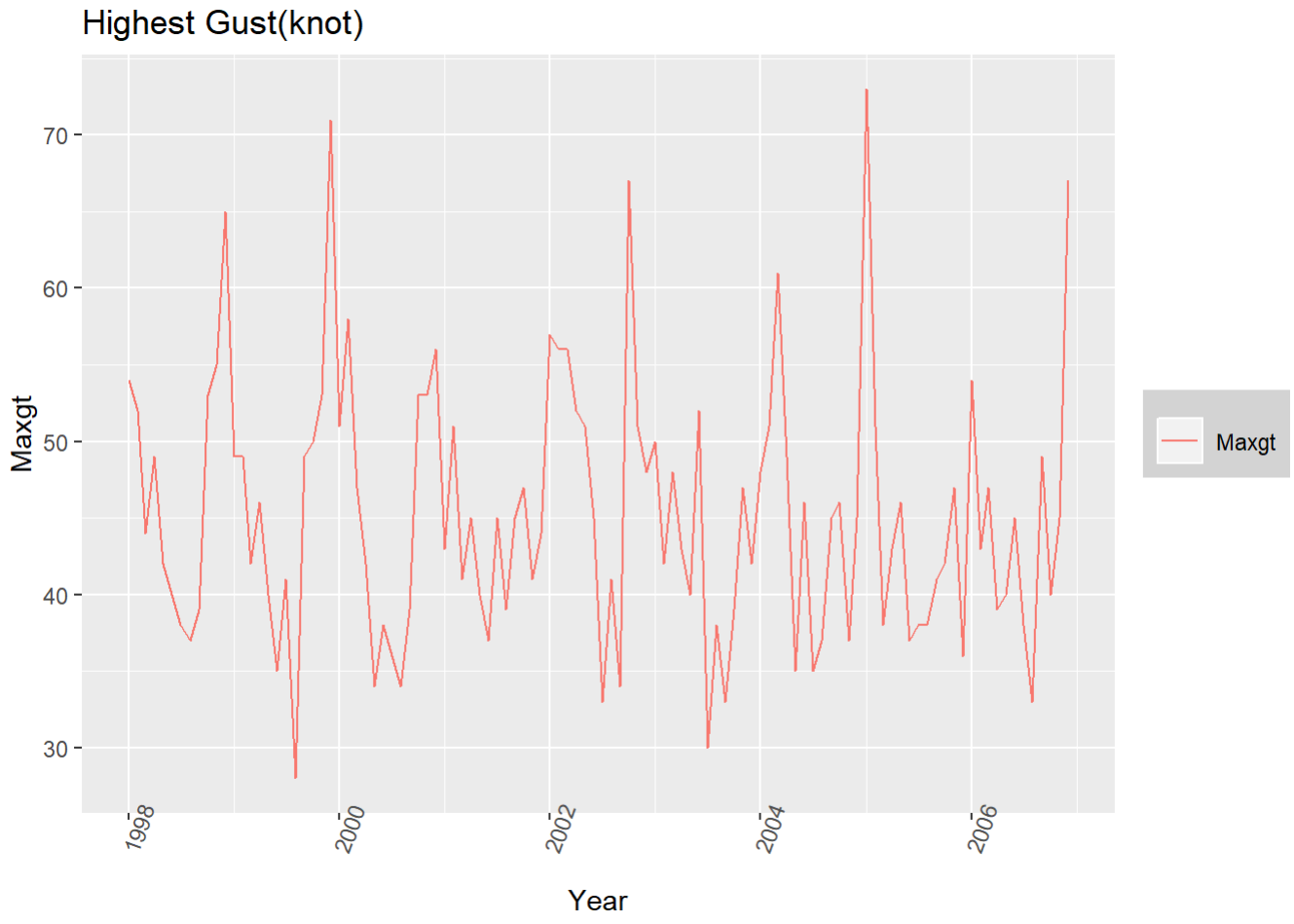
```
#User inputs which plot to draw (i.e, only one of the three,two of the three, or all three plot
s) along with specific time period.
#(1) 1st plot of three plot (Passing TRUE to plot variable for the type of plot )
plot.WeatherData(2006,20018,plota=TRUE,plotb=FALSE,plotc=FALSE)
```




```
#Comment: Maxtp is shown by red line and Mintp by the blue line
```

```
#(1) 2nd & 3rd plot among three plot
plot.WeatherData(1998,2006,plota=FALSE,plotb=TRUE,plotc=TRUE)
```


Precipitaion Amount (mm)
Rain
150
100
50
1998
2000
2002
2004
2006
Year
Rain


Highest Gust(knot)
Maxgt
70
60
50
40
30
1998
2000
2002
2004
2006
Year
Maxgt

```
#User inputs only which years to plot.As no plot type is provided by default all three plots wil
l be shown.
plot.WeatherData(1999,2016)
```