CODE:

```
library(rvest)

library(dplyr)



link <- "https://docs.google.com/spreadsheets/d/e/2PACX-
1vQA9Atb2iCAbExo6bGyneP0yYIwaHn3xtaPSOATzziQqELAbBy0hp-bWjC2Y2HWbxAbHsZMr7NmF14s0S-pcEQ/pubhtml"

rainfallURL<- read_html(link)


rainfallURL


allTables<- rainfallURL %>% html_table(fill = TRUE)

View(allTables)



#extract the data available


#JUNE

jun<- data.frame(allTables[[6]])

jun<-(jun[-1,-1]) #cleaning of data

head(jun)

View(jun)

#sum

juneTotalcol<-as.numeric(unlist(jun %>% select(Var.34)))

juneTotalRainfall<-sum(na.omit(juneTotalcol))

juneTotalRainfall


#JULY


july<-data.frame(allTables[[7]])

july<-(july[-1,-1]) #cleaning of data

head(july)
```

```
View(july)


#sum
julyTotalcol<-as.numeric(unlist(july %>% select(Var.35)))
julyTotalRainfall<- sum(na.omit(julyTotalcol))
julyTotalRainfall



#Create a matrix which contains the change of percentage (+ve / -ve) between the month of June 2022 and July 2022


length(juneTotalcol)<-length(julyTotalcol) # same lengths
changeJuneJuly<-cbind.data.frame( juneTotalcol,julyTotalcol )
changeJuneJuly<- changeJuneJuly[-1,] #cleaning of data
head(changeJuneJuly)
changeJuneJuly<-(mutate(changeJuneJuly,Change=julyTotalcol-juneTotalcol))
changeJuneJuly<-mutate(changeJuneJuly,PercentageChange=(Change/juneTotalcol)*100 )
head(changeJuneJuly)
#Total findings added as last row
changeJuneJuly<- changeJuneJuly %>% add_row
            (juneTotalcol=juneTotalRainfall,
              julyTotalcol=julyTotalRainfall,
              Change=julyTotalcol- juneTotalcol,
              PercentageChange= (Change/juneTotalcol)*100 )
#convert to matrix
changeMatrix<-(as.matrix.data.frame(changeJuneJuly))
is.matrix(changeMatrix)
tail(changeMatrix)
```

```
#(Extra Credit) State the date of the month which have
#maximum rainfall accumulated over all ten months.
jan<- data.frame(allTables[[1]])
jan<-cleanData(feb)
jan<-within(jan,rm('Var.35'))
jan <- mutate_all(jan, function(x) as.numeric(as.character(x)))
jan <- jan %>% replace(is.na(.), 0)
maxjan=(sort(colSums(jan[,1:length(jan)]),decreasing=TRUE)[1])
maxjan

feb<- data.frame(allTables[[2]])
feb<-cleanData(feb)
feb<-within(feb,rm('Var.33'))
feb <- mutate_all(feb, function(x) as.numeric(as.character(x)))
feb <- feb %>% replace(is.na(.), 0)
maxfeb=(sort(colSums(feb[,1:length(feb)]),decreasing=TRUE)[1])
maxfeb

maxjan<-getMonthMax(1)
maxMar<-getMonthMax(3)
maxApr<-getMonthMax2(4)
maxMay<-getMonthMax(5)
maxJun<-getMonthMax2(6)
maxJuly<-getMonthMax(7)
maxAug<-getMonthMax(8)
maxSep<-getMonthMax2(9)
maxOct<-getMonthMax2(10)



c('jan','feb','mar','apr','may','june','july',
  'aug','sep','oct')[which.max(c(maxjan,maxfeb,maxMar,
  maxApr,maxMay,maxJun,maxJuly,maxAug,maxSep,maxOct))]
```

```
str(maxJuly) # date of maximum rainfall in 10 months



#====================FUNCTIONS================================

cleanData <- function(myMonth)
{
  myMonth<-(myMonth[-1,-1])
  myMonth<-(myMonth[-1,-1])
  myMonth<-(myMonth[,-1])
  myMonth <- myMonth %>% replace(is.na(.), 0)
}

getMonthMax <- function(loc)
{
  df<- data.frame(allTables[[loc]])
  df<-cleanData(df)
  df<-within(df,rm('Var.35'))
  df <- mutate_all(df, function(x) as.numeric(as.character(x)))
  df <- df %>% replace(is.na(.), 0)
  maxx=(sort(colSums(df[,1:length(df)]),decreasing=TRUE)[1])
  return(maxx)

}
getMonthMax2 <- function(loc)
{
  df<- data.frame(allTables[[loc]])
  df<-cleanData(df)
  df<-within(df,rm('Var.34'))
  df <- df %>% replace(is.na(.), 0)
  df <- mutate_all(df, function(x) as.numeric(as.character(x)))
```

```
  df <- df %>% replace(is.na(.), 0);3
  maxx=(sort(colSums(df[,1:length(df)]),decreasing=TRUE)[1])
  return(maxx)

}
```

SCREENSHOTS

helloword.R | allTables

Show Attributes

Name	Type	Value
allTables	list [12]	List of length 12
[[1]]	list [154 x 35] (S3: tbl_df, tbl, data	A tibble with 154 rows and 35 columns
[[2]]	list [154 x 54] (S3: tbl_df, tbl, data	A tibble with 154 rows and 54 columns
[[3]]	list [154 x 35] (S3: tbl_df, tbl, data	A tibble with 154 rows and 35 columns
[[4]]	list [154 x 34] (S3: tbl_df, tbl, data	A tibble with 154 rows and 34 columns
[[5]]	list [152 x 35] (S3: tbl_df, tbl, data	A tibble with 152 rows and 35 columns
[[6]]	list [153 x 34] (S3: tbl_df, tbl, data	A tibble with 153 rows and 34 columns
[[7]]	list [152 x 35] (S3: tbl_df, tbl, data	A tibble with 152 rows and 35 columns
[[8]]	list [152 x 35] (S3: tbl_df, tbl, data	A tibble with 152 rows and 35 columns
[[9]]	list [152 x 34] (S3: tbl_df, tbl, data	A tibble with 152 rows and 34 columns
[[10]]	list [152 x 35] (S3: tbl_df, tbl, data	A tibble with 152 rows and 35 columns
[[11]]	list [152 x 34] (S3: tbl_df, tbl, data	A tibble with 152 rows and 34 columns
[[12]]	list [153 x 35] (S3: tbl_df, tbl, data	A tibble with 153 rows and 35 columns

(No selection)

Console | Terminal | Background Jobs

R 4.2.1 · ~/

```
> library(rvest)
> library(dplyr)
> link <- "https://docs.google.com/spreadsheets/d/e/2PACX-1vQA9Atb2iCAbEx
Mr7NmF14s0S-pcEQ/pubhtml"
> rainfallURL<- read_html(link)
> rainfallURL
{html_document}
<html>
[1] <head>\n<meta http-equiv="Content-Type" content="text/html; charset=U
[2] <script nonce="jYTUFikBog9u2F8iMdbS6w">document.addEventListener('DOM
[3] <body class="docs-gm">\n<div id="top-bar">\n<div id="doc-title"><span
> allTables<- rainfallURL %>% html_table(fill = TRUE)
> View(allTables)
```

```
14 #extract the data available
15
16 #JUNE
17 jun<- data.frame(allTables[[6]])
18 jun<-(jun[-1,-1]) #cleaning of data
19 head(jun)
20 View(jun)
21 #sum
22 juneTotalcol<-as.numeric(unlist(jun %>% select(Var.34)))
23 juneTotalRainfall<-sum(na.omit(juneTotalcol))
24 juneTotalRainfall
```

```
> #JUNE
> jun<- data.frame(allTables[[6]])
> jun<-(jun[-1,-1]) #cleaning of data
> head(jun)
       Var.2    Var.3 Var.4 Var.5 Var.6 Var.7 Var.8 Var.9 Var.10 Var.11 Var.12 Var.13 Var.14 Var.15 Var.16 Var.17 Var.18
2    STATIONS DISTRICT     1     2     3     4     5     6      7      8      9     10     11     12     13     14     15
3      ATTOCK   Attock   4.0   0.0   0.0   0.0   0.0   0.0    0.0    0.0    0.0    0.0    0.0    0.0    2.0    0.0    0.0
4 PINDI GHAIB   Attock   0.0   0.0   0.0   0.0   0.0   0.0    0.0    0.0    0.0    0.0    0.0    0.0   12.0    0.0    0.0
5       JHAND   Attock   0.0   0.0   0.0   0.0   0.0   0.0    0.0    0.0    0.0    0.0    0.0    0.0    8.0    0.0    0.0
6 HASAN ABDAL   Attock   9.0   0.0   0.0   0.0   0.0   0.0    0.0    0.0    0.0    0.0    0.0    0.0    3.0    0.0    0.0
7  FATEH JANG   Attock   0.0   0.0   0.0   0.0   0.0   0.0    0.0    0.0    0.0    0.0    0.0    0.0    2.0    0.0    0.0
  Var.19 Var.20 Var.21 Var.22 Var.23 Var.24 Var.25 Var.26 Var.27 Var.28 Var.29 Var.30 Var.31 Var.32 Var.33 Var.34
2     16     17     18     19     20     21     22     23     24     25     26     27     28     29     30  Total
3    0.0    2.0    0.0   13.0    8.0    3.0   45.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0   77.0
4    0.0    5.0    0.0    2.0    0.0    2.0   20.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0   41.0
5    0.0    0.0    0.0    9.0    0.0    0.0   18.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0   35.0
6    0.0    8.0    0.0    2.0    6.0   33.0   52.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0  113.0
7    0.0    0.0    0.0    1.0   13.0    2.0   20.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0    0.0   38.0
>
```

```
> juneTotalRainfall<-sum(na.omit(juneTotalcol))
> juneTotalRainfall
[1] 6994.8
```

```
26  #JULY
27
28  july<-data.frame(allTables[[7]])
29  july<-(july[-1,-1]) #cleaning of data
30  head(july)
31  View(july)
32
33  #sum
34  julyTotalcol<-as.numeric(unlist(july %>% select(Var.35)))
35  julyTotalRainfall<- sum(na.omit(julyTotalcol))
36  julyTotalRainfall
```

```
> july<-data.frame(allTables[[7]])
> july<-(july[-1,-1]) #cleaning of data
> head(july)
       Var.2    Var.3 Var.4 Var.5 Var.6 Var.7 Var.8 Var.9 Var.10 Var.11 Var.12 Var.13 Var.14 Var.15 Var.16 Var.17 Var.18
2    STATIONS DISTRICT     1     2     3     4     5     6      7      8      9     10     11     12     13     14     15
3      ATTOCK   Attock   0.0   0.0   0.0   5.0   0.0   0.0    6.0    0.0    0.0    0.0   20.0   65.0   12.0   84.0    0.0
4 PINDI GHAIB   Attock   0.0   0.0   0.0   0.0   0.0   0.0    0.0    0.0    0.0    0.0    0.0    7.0   18.0   15.0    0.0
5       JHAND   Attock   0.0   0.0   0.0   0.0   0.0   0.0    0.0    0.0    0.0    0.0    0.0   17.0   13.0   22.0    0.0
6 HASAN ABDAL   Attock   1.0   0.0   0.0   0.0 113.0   0.0    1.0    0.0    0.0    0.0   30.0  115.0   87.0   26.0    0.0
7  FATEH JANG   Attock   0.0   0.0   0.0   0.0   0.0   0.0    0.0    0.0    0.0    0.0    0.0   25.0   17.0   15.0    0.0
  Var.19 Var.20 Var.21 Var.22 Var.23 Var.24 Var.25 Var.26 Var.27 Var.28 Var.29 Var.30 Var.31 Var.32 Var.33 Var.34 Var.35
2     16     17     18     19     20     21     22     23     24     25     26     27     28     29     30     31  Total
3    0.0    0.0    0.0    0.0    0.0    0.0    3.0    0.0    0.0    0.0    1.0    0.0    0.0    6.0   58.0    0.0  260.0
4    0.0    0.0    0.0    0.0    0.0    0.0    4.0    0.0    0.0    4.0    0.0    0.0    0.0    0.0    0.0    2.0   50.0
5    0.0    0.0    0.0    0.0    0.0    0.0    9.0    0.0   12.0    9.0    0.0    6.0    0.0    0.0    2.0    0.0   90.0
6    0.0    0.0    0.0    0.0    0.0   28.0   16.0    0.0   38.0   17.0    4.0    0.0    0.0   39.0    4.0   47.0  566.0
7    0.0    0.0    0.0    0.0    0.0    5.0    8.0    0.0   13.0    5.0    0.0    0.0    0.0    4.0   13.0   23.0  128.0
>
```

```
> julyTotalRainfall<- sum(na.omit(julyTotalcol))
> julyTotalRainfall
[1] 28690.1
```

```r
#Create a matrix which contains the change of
#percentage (+ve / -ve) between the month of June 2022 and July 2022

length(juneTotalcol)<-length(julyTotalcol) # same lengths
changeJuneJuly<-cbind.data.frame( juneTotalcol,julyTotalcol )
changeJuneJuly<- changeJuneJuly[-1,] #cleaning of data
head(changeJuneJuly)
changeJuneJuly<-(mutate(changeJuneJuly,Change=julyTotalcol-juneTotalcol))
changeJuneJuly<-mutate(changeJuneJuly,PercentageChange=(Change/juneTotalcol)*100 )
head(changeJuneJuly)
#Total findings added as last row
changeJuneJuly<- changeJuneJuly %>% add_row
                 (juneTotalcol=juneTotalRainfall,
                   julyTotalcol=julyTotalRainfall,
                   Change=julyTotalcol- juneTotalcol,
                   PercentageChange= (Change/juneTotalcol)*100 )
#convert to matrix
changeMatrix<-(as.matrix.data.frame(changeJuneJuly))
is.matrix(changeMatrix)
tail(changeMatrix)
```

```
> length(juneTotalcol)<-length(julyTotalcol) # same lengths
> changeJuneJuly<-cbind.data.frame( juneTotalcol,julyTotalcol )
> changeJuneJuly<- changeJuneJuly[-1,] #cleaning of data
> head(changeJuneJuly)
  juneTotalcol julyTotalcol
2           77          260
3           41           50
4           35           90
5          113          566
6           38          128
7           93          513
> changeJuneJuly<-(mutate(changeJuneJuly,Change=julyTotalcol-juneTotalcol))
> changeJuneJuly<-mutate(changeJuneJuly,PercentageChange=(Change/juneTotalcol)*100 )
> head(changeJuneJuly)
  juneTotalcol julyTotalcol Change PercentageChange
2           77          260    183        237.66234
3           41           50      9         21.95122
4           35           90     55        157.14286
5          113          566    453        400.88496
6           38          128     90        236.84211
7           93          513    420        451.61290
> #Total findings added as last row
> changeJuneJuly<- changeJuneJuly %>% add_row
> #convert to matrix
> changeMatrix<-(as.matrix.data.frame(changeJuneJuly))
> is.matrix(changeMatrix)
[1] TRUE
> tail(changeMatrix)
       juneTotalcol julyTotalcol Change PercentageChange
[146,]           40          157    117        292.50000
[147,]           75          116     41         54.66667
[148,]           26           97     71        273.07692
[149,]           10          107     97        970.00000
[150,]           11           87     76        690.90909
[151,]           NA           NA     NA               NA
>
```

```r
#(Extra Credit) State the date of the month which have
#maximum rainfall accumulated over all ten months.
jan<- data.frame(allTables[[1]])
jan<-cleanData(feb)
jan<-within(jan,rm('Var.35'))
jan <- mutate_all(jan, function(x) as.numeric(as.character(x)))
jan <- jan %>% replace(is.na(.), 0)
maxjan=(sort(colSums(jan[,1:length(jan)]),decreasing=TRUE)[1])
maxjan

feb<- data.frame(allTables[[2]])
feb<-cleanData(feb)
feb<-within(feb,rm('Var.33'))
feb <- mutate_all(feb, function(x) as.numeric(as.character(x)))
feb <- feb %>% replace(is.na(.), 0)
maxfeb=(sort(colSums(feb[,1:length(feb)]),decreasing=TRUE)[1])
maxfeb

maxjan<-getMonthMax(1)
maxMar<-getMonthMax(3)
maxApr<-getMonthMax2(4)
maxMay<-getMonthMax(5)
maxJun<-getMonthMax2(6)
maxJuly<-getMonthMax(7)
maxAug<-getMonthMax(8)
maxSep<-getMonthMax2(9)
maxOct<-getMonthMax2(10)
```

```r
c('jan','feb','mar','apr','may','june','july',
  'aug','sep','oct')[which.max(c(maxjan,maxfeb,maxMar,
  maxApr,maxMay,maxJun,maxJuly,maxAug,maxSep,maxOct))]

str(maxJuly) # date of maximum rainfall in 10 months
```

DATE : 25TH July

```
> c('jan','feb','mar','apr','may','june','july','aug','se
p,maxOct))]
[1] "jan"
> str(maxJuly) # date of maximum rainfall in 10 months
 Named num 3410
 - attr(*, "names")= chr "Var.25"
>
```

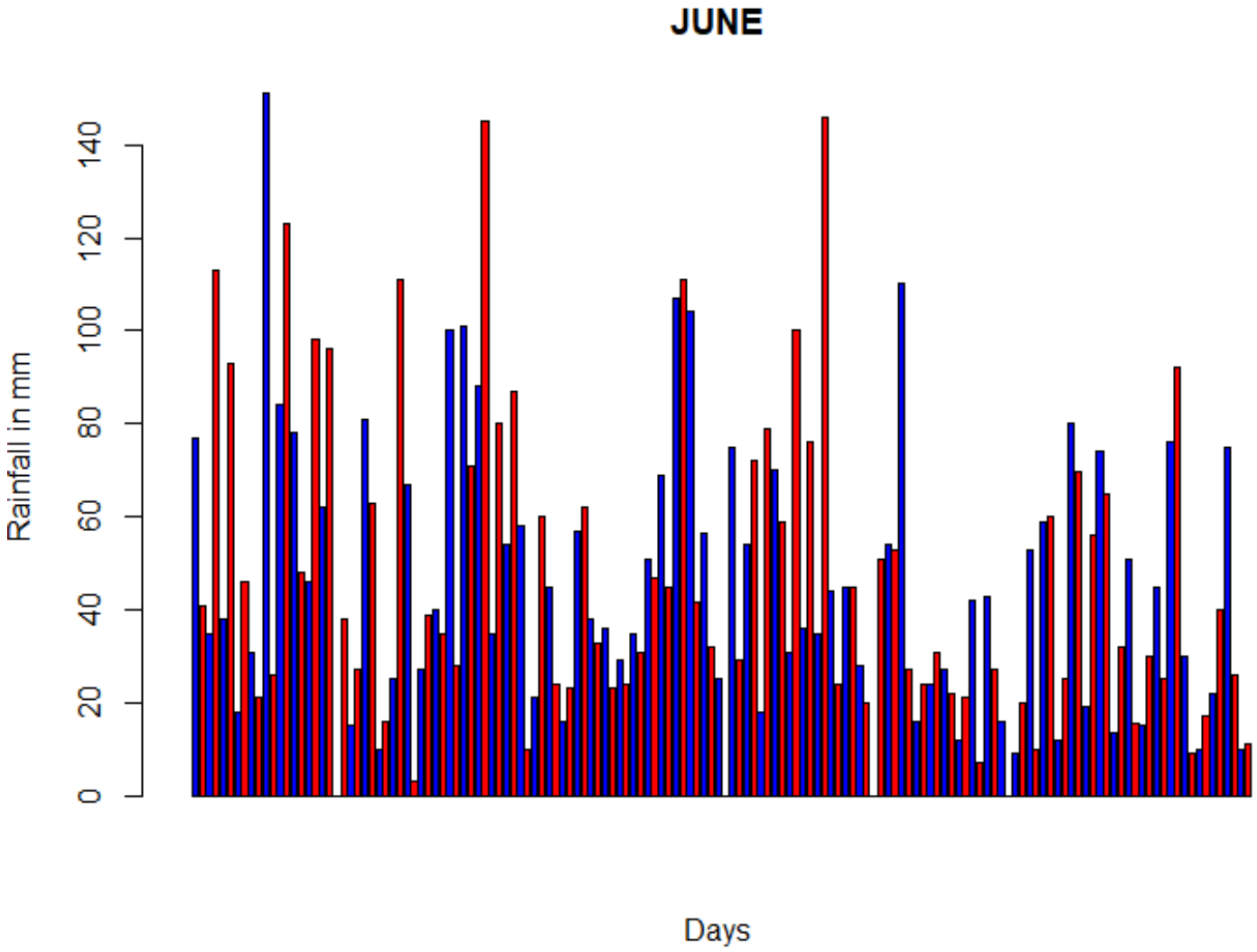
FUNCTIONS

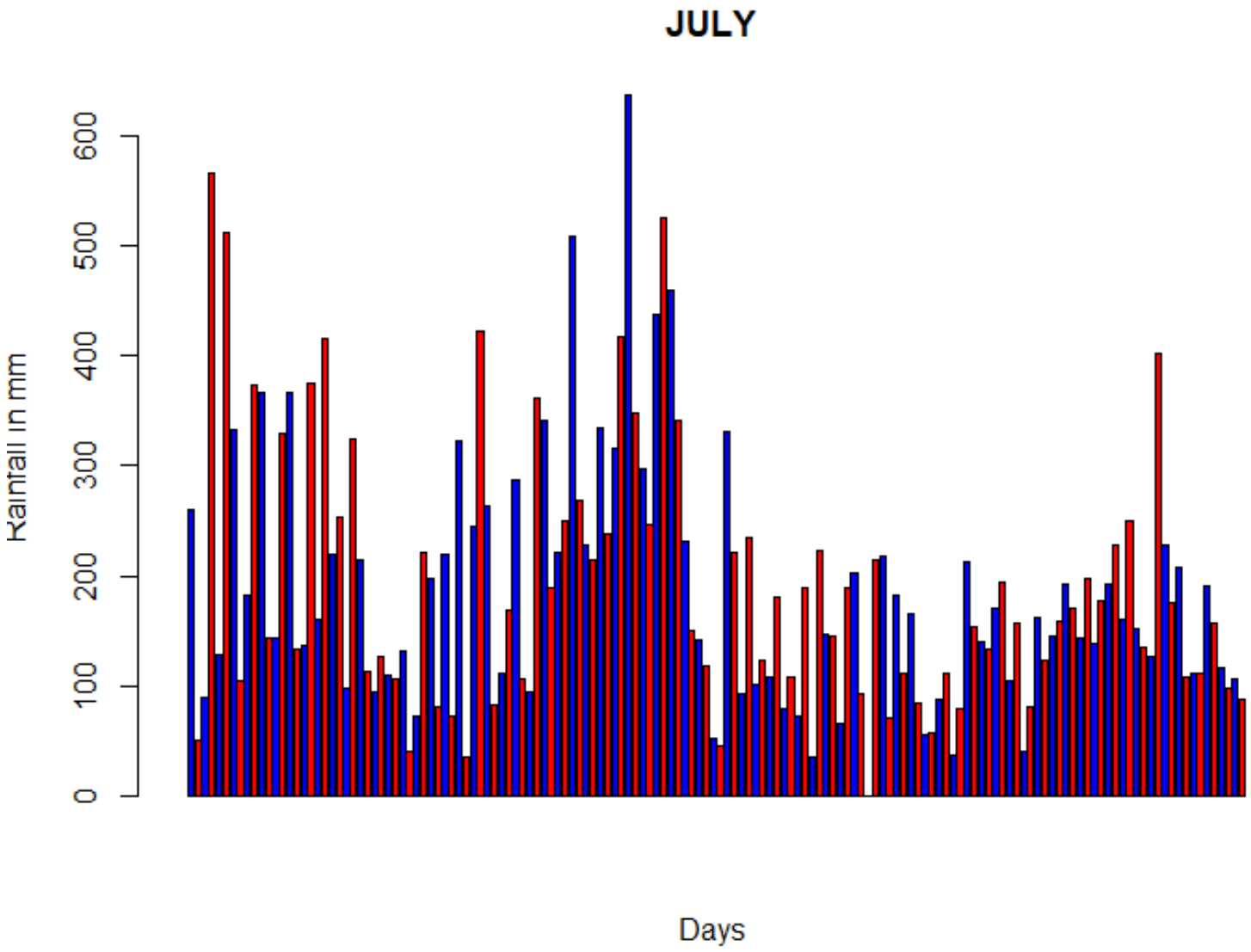
```
#=====================FUNCTIONS=================================
cleanData <- function(myMonth) {
  myMonth<-(myMonth[-1,-1])
  myMonth<-(myMonth[-1,-1])
  myMonth<-(myMonth[,-1])
  myMonth <- myMonth %>% replace(is.na(.), 0)
}
getMonthMax <- function(loc){
  df<- data.frame(allTables[[loc]])
  df<-cleanData(df)
  df<-within(df,rm('Var.35'))
  df <- mutate_all(df, function(x) as.numeric(as.character(x)))
  df <- df %>% replace(is.na(.), 0)
  maxx=(sort(colSums(df[,1:length(df)]),decreasing=TRUE)[1])
  return(maxx)
}
getMonthMax2 <- function(loc){
  df<- data.frame(allTables[[loc]])
  df<-cleanData(df)
  df<-with                                    4'))
  df <- df %>% replace(is.na(.), 0)
  df <- mutate_all(df, function(x) as.numeric(as.character(x)))
  df <- df %>% replace(is.na(.), 0);3
  maxx=(sort(colSums(df[,1:length(df)]),decreasing=TRUE)[1])
  return(maxx)
```

mutate_all(.tbl, .funs, ...)

PLOTS

```
128
129  #------------PLOTS
130  colours = c("red","blue")
131  barplot(as.numeric(juneTotalcol),main='Rainfall By Day',
132          ylab='Rainfall in mm', xlab='Days',beside = TRUE, col=colours)
133  box()
134
135  barplot(as.numeric(julyTotalcol),main='Rainfall By Day',
136          ylab='Rainfall in mm', xlab='Days',beside = TRUE, col=colours)
137  box()
138
```



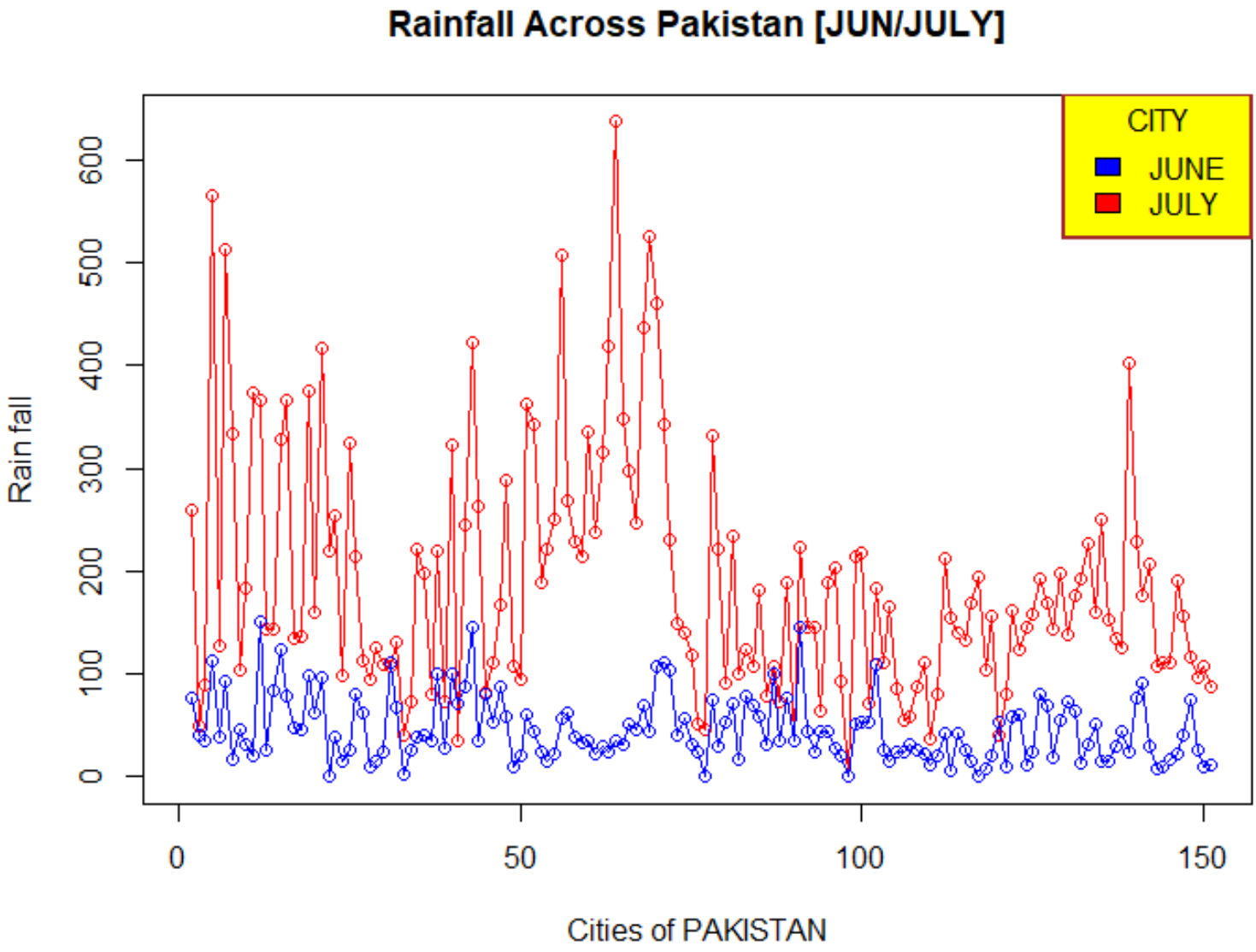



```
#------------LINE PLOT
lin1<-as.vector(juneTotalcol)
lin2<-as.vector(julyTotalcol)

plot(lin2,type = "o",col = "red", xlab = "Cities of PAKISTAN", ylab = "Rain fall",
     main = "Rainfall Across Pakistan [JUN/JULY]")

lines(lin1, type = "o", col = "blue")


legend(x = "topright", box.col = "brown",
       bg ="yellow", box.lwd = 2 , title="CITY",
       legend=c("JUNE", "JULY"),
       fill = c("blue","red"))
```


Rainfall Across Pakistan [JUN/JULY]
CITY
JUNE
JULY
Rain fall
0
100
200
300
400
500
600
0
50
100
150
Cities of PAKISTAN