Assignment_1

Morgan McCullough

9/10/2021

library('tidyverse') library("ggplot2") library("dplyr") libray("tidyr")

###inserting College.csv into a dataframe

```
collegeDf <-data.frame(read.csv("C:/Users/morga/Desktop/CPTS475/College.csv"))
###collegeDf <- collegeDf %>% drop_na()
```

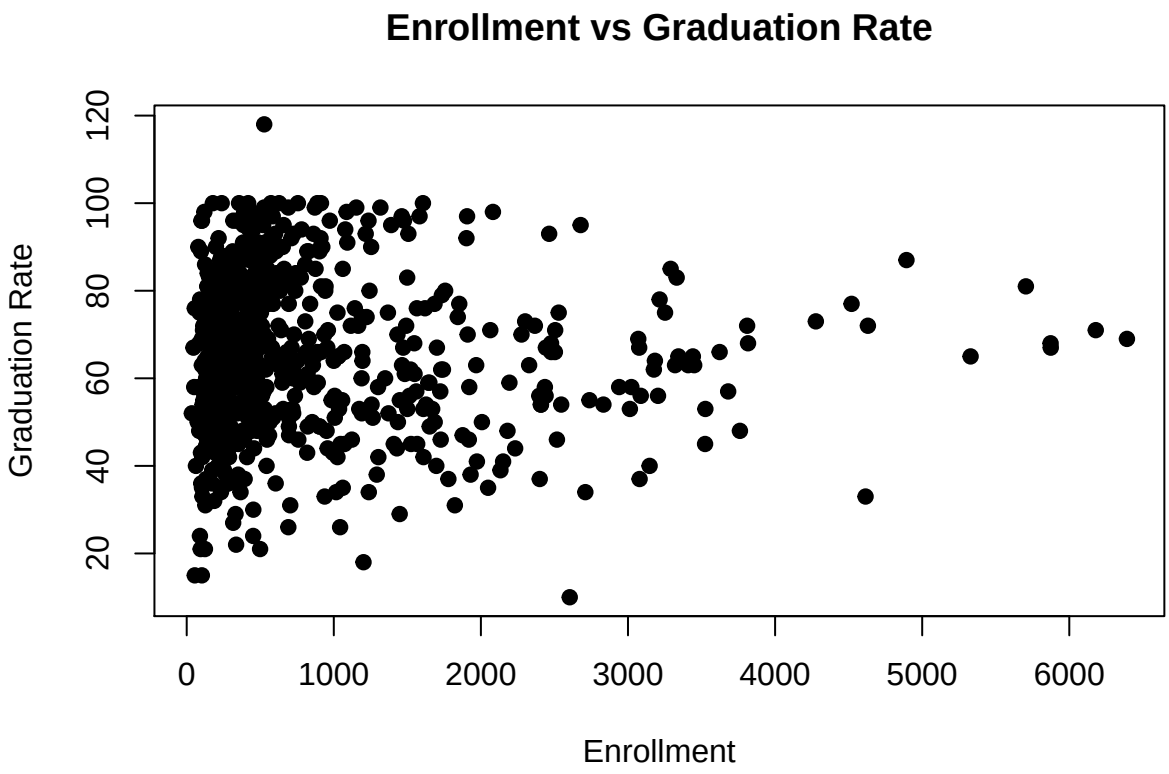
###finding median price of room and board

```
medianPriceRoom <- median(collegeDf$Room.Board)
medianPriceRoom
```

```
## [1] 4200
```

###scatterplot

```
attach(collegeDf)
plot(collegeDf$Enroll, collegeDf$Grad.Rate,
     main = "Enrollment vs Graduation Rate",
     xlab="Enrollment", ylab="Graduation Rate", pch=19)
```



D and E

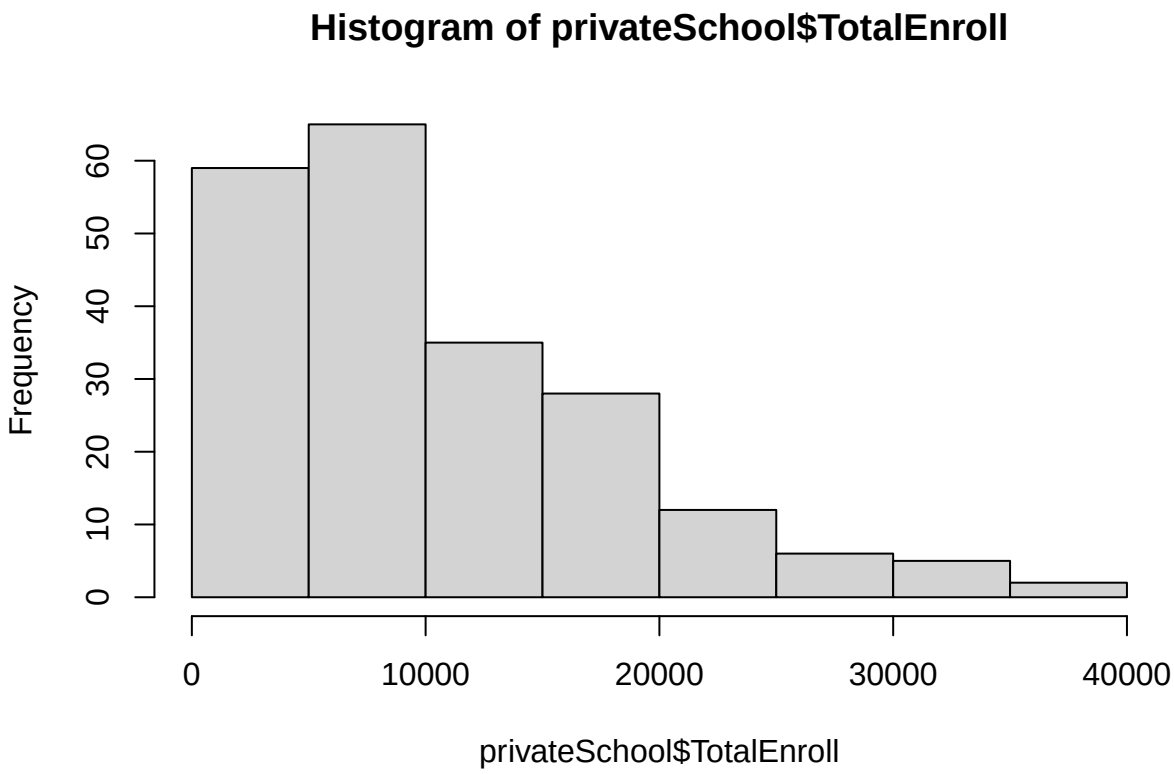
```
###D
attach(collegeDf)
```

```
## The following objects are masked from collegeDf (pos = 3):
##
##     Accept, Apps, Books, Enroll, Expend, F.Undergrad, Grad.Rate,
##     Outstate, P.Undergrad, perc.alumni, Personal, PhD, Private,
##     Room.Board, S.F.Ratio, Terminal, Top10perc, Top25perc, X
```

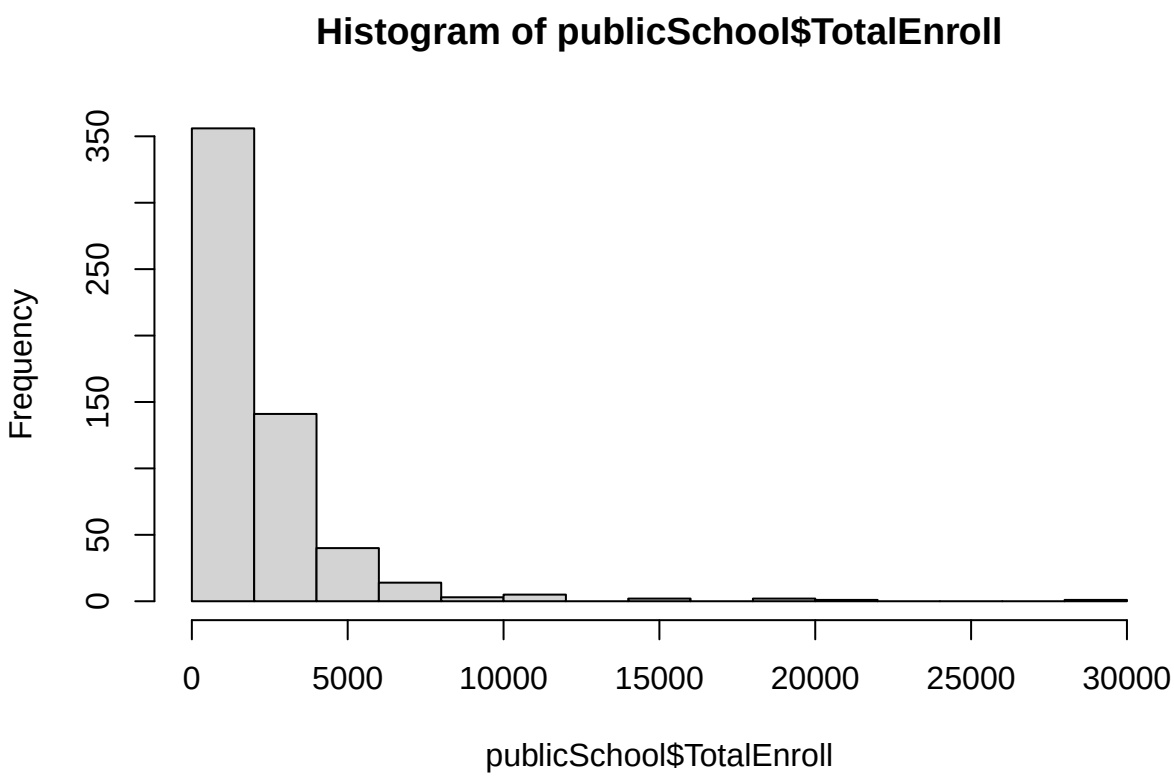
```
collegeDf$PrivateBinary[collegeDf$Private == "Yes"] <- 1
collegeDf$PrivateBinary[collegeDf$Private == "No"] <- 0

publicSchool <- collegeDf[collegeDf$PrivateBinary == 1,]
privateSchool <- collegeDf[collegeDf$PrivateBinary == 0,]
privateSchool$TotalEnroll <- (privateSchool$F.Undergrad +
                                privateSchool$P.Undergrad)

hist(privateSchool$TotalEnroll)
```




```
publicSchool$TotalEnroll <- (publicSchool$F.Undergrad +
                               publicSchool$P.Undergrad)
hist(publicSchool$TotalEnroll)
```




```
###E
collegeDf$Top[collegeDf$Top10perc <= 75] <- 0
collegeDf$Top[collegeDf$Top10perc > 75] <- 1

# boxplot(collegeDf$Top ~ collegeDf$AcceptanceRate)


detach(collegeDf)
```

###F

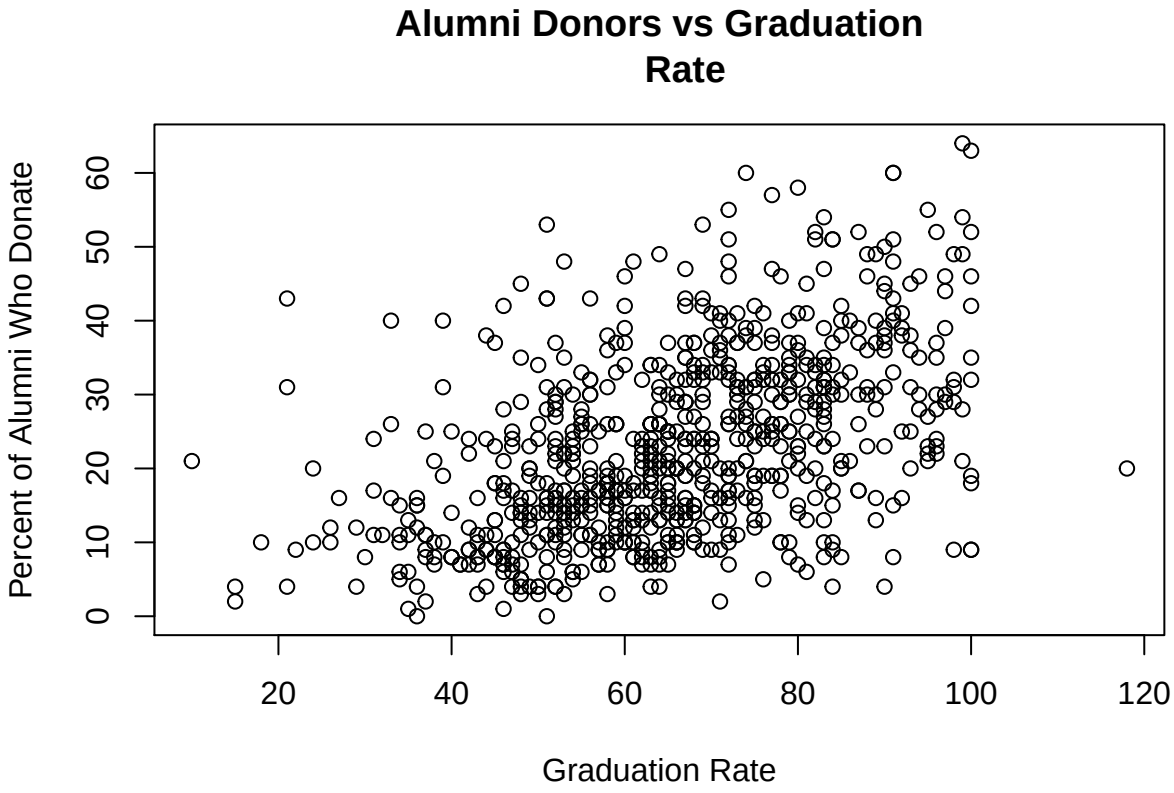
```
###Hypothesis: There is a correlation between graduation rate and percent of
###alumni who donate. My idea is that the higher the graduation
###rate, the more alumni are to donate to the college.

attach(collegeDf)
```

```
## The following objects are masked from collegeDf (pos = 3):
##
##     Accept, Apps, Books, Enroll, Expend, F.Undergrad, Grad.Rate,
##     Outstate, P.Undergrad, perc.alumni, Personal, PhD, Private,
##     Room.Board, S.F.Ratio, Terminal, Top10perc, Top25perc, X
```

```
newData <- collegeDf[order(perc.alumni, Grad.Rate ),]

plot(newData$perc.alumni ~ newData$Grad.Rate, main="Alumni Donors vs Graduation
     Rate", xlab="Graduation Rate", ylab="Percent of Alumni Who Donate")
```




```
###There are some outliers that do not fit within the trend. My prediction was
###correct.
```

###2

```
forestFireData <- read.csv("C:/Users/morga/Desktop/CPTS475/forestfires.csv")
```

###A ###Month and Day are represented as qualitative variable, both in abbreviated ### version of themselves, ###The rest of the data is quantitative and represented in numerical values.

###b

```
FFMCrange <- range(forestFireData$FFMC)
FFMCmedian <- median(forestFireData$FFMC)
FFMCstandDev <- sd(forestFireData$FFMC)
FFMCstats <- data.frame(FFMCrange, FFMCmedian, FFMCstandDev)

DMCrange <- range(forestFireData$DMC)
DMCmedian <- median(forestFireData$DMC)
DMCstandDev <- sd(forestFireData$DMC)
DMCstats <- data.frame(DMCrange, DMCmedian, DMCstandDev)

DCrange <- range(forestFireData$DC)
DCmedian <- median(forestFireData$DC)
DCstandDev <- sd(forestFireData$DC)
DCstats <- data.frame(DCrange, DCmedian, DCstandDev)

ISIrange <- range(forestFireData$ISI)
```

```
ISImedian <- median(forestFireData$ISI)
ISIstandDev <- sd(forestFireData$ISI)
ISIstats <- data.frame(ISIrange, ISImedian, ISIstandDev)

tempRange <- range(forestFireData$temp)
tempMedian <- median(forestFireData$temp)
tempStandDev <- sd(forestFireData$temp)
tempStats <- data.frame(tempRange, tempMedian, tempStandDev)

RHrange <- range(forestFireData$RH)
RHmedian <- median(forestFireData$RH)
RHstandDev <- sd(forestFireData$RH)
RHstats <- data.frame(RHrange, RHmedian, RHstandDev)

windRange <- range(forestFireData$wind)
windMedian <- median(forestFireData$wind)
windStandDev <- sd(forestFireData$wind)
windStats <- data.frame(windRange,windMedian, windStandDev)

rainRange <- range(forestFireData$rain)
rainMedian <- median(forestFireData$rain)
rainStandDev <- sd(forestFireData$rain)
rainStats <- data.frame(rainRange, rainMedian, rainStandDev)

areaRange <- range(forestFireData$area)
areaMedian <- median(forestFireData$area)
areaStandDev <- sd(forestFireData$area)
areaStats <- data.frame(areaRange, areaMedian, areaStandDev)
```

###C

```
library("dplyr")
```

```
## Warning: package 'dplyr' was built under R version 4.0.5
```

```
##
## 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
```

```
forestFireDataNew <- data.frame(forestFireData)
forestFireDataNew <- anti_join(forestFireDataNew, forestFireDataNew[20:70,])
```

```
## Joining, by = c("month", "day", "FFMC", "DMC", "DC", "ISI", "temp", "RH", "wind", "rain", "area")
```

```
FFMCrangeNew <- range(forestFireDataNew$FFMC)
FFMCmedianNew <- median(forestFireDataNew$FFMC)
FFMCstandDevNew <- sd(forestFireDataNew$FFMC)
FFMCstatsNew <- data.frame(FFMCrange, FFMCmedian, FFMCstandDev)
FFMCstatsNew
```

```
##   FFMCrange FFMCmedian FFMCstandDev
## 1      18.7       91.6     5.520111
## 2      96.2       91.6     5.520111
```

```
DMCrangeNew <- range(forestFireDataNew$DMC)
DMCmedianNew <- median(forestFireDataNew$DMC)
DMCstandDevNew <- sd(forestFireDataNew$DMC)
DMCstatsNew <- data.frame(DMCrange, DMCmedian, DMCstandDev)
DMCstatsNew
```

```
##   DMCrange DMCmedian DMCstandDev
## 1      1.1     108.3    64.04648
## 2    291.3     108.3    64.04648
```

```
DCrangeNew <- range(forestFireDataNew$DC)
DCmedianNew <- median(forestFireDataNew$DC)
DCstandDevNew <- sd(forestFireDataNew$DC)
DCstatsNew <- data.frame(DCrange, DCmedian, DCstandDev)
DCstatsNew
```

```
##   DCrange DCmedian DCstandDev
## 1     7.9    664.2   248.0662
## 2   860.6    664.2   248.0662
```

```
ISIrangeNew <- range(forestFireDataNew$ISI)
ISImedianNew <- median(forestFireDataNew$ISI)
ISIstandDevNew <- sd(forestFireDataNew$ISI)
ISIstatsNew <- data.frame(ISIrange, ISImedian, ISIstandDev)
ISIstatsNew
```

```
##   ISIrange ISImedian ISIstandDev
## 1      0.0       8.4    4.559477
## 2     56.1       8.4    4.559477
```

```
tempRangeNew <- range(forestFireDataNew$temp)
tempMedianNew <- median(forestFireDataNew$temp)
tempStandDevNew <- sd(forestFireDataNew$temp)
tempStatsNew <- data.frame(tempRange, tempMedian, tempStandDev)
tempStatsNew
```

```
##   tempRange tempMedian tempStandDev
## 1       2.2       19.3     5.806625
## 2      33.3       19.3     5.806625
```

```
RHrangeNew <- range(forestFireDataNew$RH)
RHmedianNew <- median(forestFireDataNew$RH)
RHstandDevNew <- sd(forestFireDataNew$RH)
RHstatsNew <- data.frame(RHrange, RHmedian, RHstandDev)
RHstatsNew
```

```
##   RHrange RHmedian RHstandDev
## 1      15       42   16.31747
## 2     100       42   16.31747
```

```
windRangeNew <- range(forestFireDataNew$wind)
windMedianNew <- median(forestFireDataNew$wind)
windStandDevNew <- sd(forestFireDataNew$wind)
windStatsNew <- data.frame(windRange,windMedian, windStandDev)
windStatsNew
```

```
##   windRange windMedian windStandDev
## 1       0.4          4     1.791653
## 2       9.4          4     1.791653
```

```
rainRangeNew <- range(forestFireDataNew$rain)
rainMedianNew <- median(forestFireDataNew$rain)
rainStandDevNew <- sd(forestFireDataNew$rain)
rainStatsNew <- data.frame(rainRange, rainMedian, rainStandDev)
rainStatsNew
```

```
##   rainRange rainMedian rainStandDev
## 1       0.0          0    0.2959591
## 2       6.4          0    0.2959591
```

```
areaRangeNew <- range(forestFireDataNew$area)
areaMedianNew <- median(forestFireDataNew$area)
areaStandDevNew <- sd(forestFireDataNew$area)
areaStatsNew <- data.frame(areaRange, areaMedian, areaStandDev)
areaStatsNew
```

```
##   areaRange areaMedian areaStandDev
## 1      0.00       0.52     63.65582
## 2   1090.84       0.52     63.65582
```

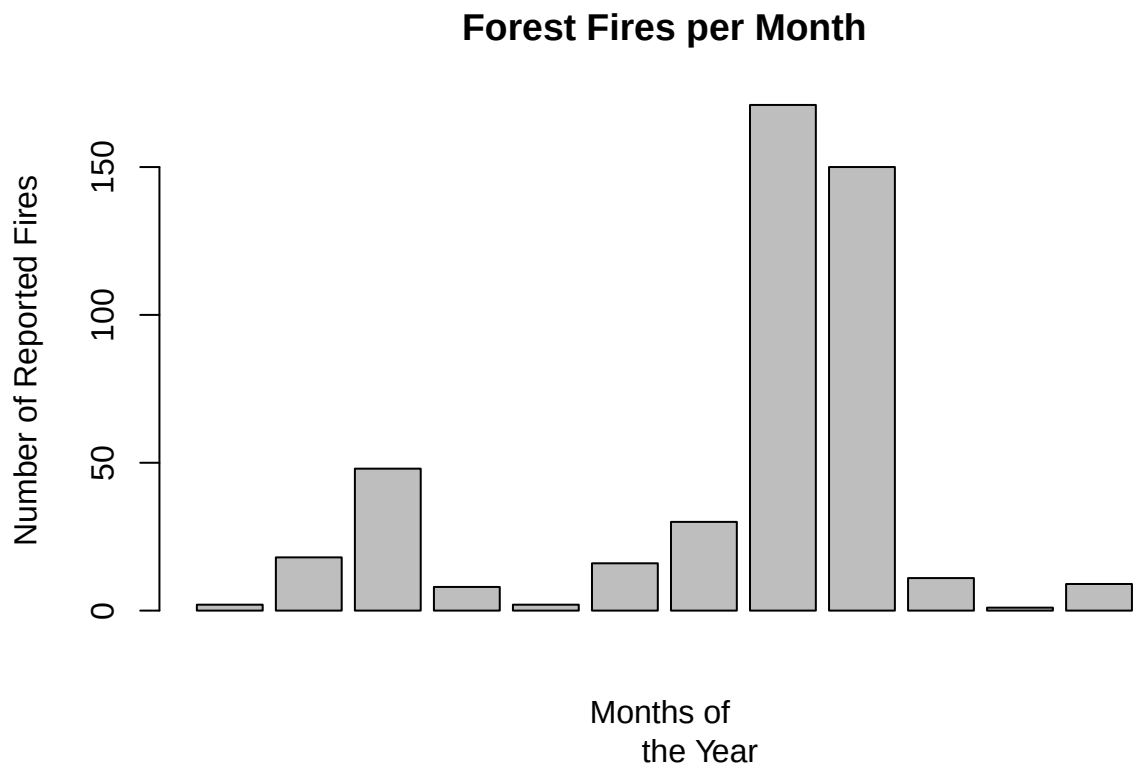
###D

```
attach(forestFireDataNew)
forestFireDataNew$MonthBinary[forestFireDataNew$month == "jan"] <- 1
forestFireDataNew$MonthBinary[forestFireDataNew$month == "feb"] <- 2
forestFireDataNew$MonthBinary[forestFireDataNew$month == "mar"] <- 3
forestFireDataNew$MonthBinary[forestFireDataNew$month == "apr"] <- 4
forestFireDataNew$MonthBinary[forestFireDataNew$month == "may"] <- 5
forestFireDataNew$MonthBinary[forestFireDataNew$month == "jun"] <- 6
forestFireDataNew$MonthBinary[forestFireDataNew$month == "jul"] <- 7
forestFireDataNew$MonthBinary[forestFireDataNew$month == "aug"] <- 8
forestFireDataNew$MonthBinary[forestFireDataNew$month == "sep"] <- 9
forestFireDataNew$MonthBinary[forestFireDataNew$month == "oct"] <- 10
forestFireDataNew$MonthBinary[forestFireDataNew$month == "nov"] <- 11
forestFireDataNew$MonthBinary[forestFireDataNew$month == "dec"] <- 12

janFires <- sum(forestFireDataNew$MonthBinary == 1)
febFires <- sum(forestFireDataNew$MonthBinary == 2)
marFires <- sum(forestFireDataNew$MonthBinary == 3)
aprFires <- sum(forestFireDataNew$MonthBinary == 4)
mayFires <- sum(forestFireDataNew$MonthBinary == 5)
junFires <- sum(forestFireDataNew$MonthBinary == 6)
julFires <- sum(forestFireDataNew$MonthBinary == 7)
augFires <- sum(forestFireDataNew$MonthBinary == 8)
sepFires <- sum(forestFireDataNew$MonthBinary == 9)
octFires <- sum(forestFireDataNew$MonthBinary == 10)
novFires <- sum(forestFireDataNew$MonthBinary == 11)
decFires <- sum(forestFireDataNew$MonthBinary == 12)

monthlyFires <- list(janFires, febFires, marFires, aprFires, mayFires,
                     junFires, julFires, augFires,
                           sepFires, octFires, novFires, decFires)
barplot(unlist(monthlyFires), main="Forest Fires per Month", xlab = "Months of
```

```
        the Year", ylab="Number of Reported Fires")
```



###E

```
library("Hmisc")
```

```
## Warning: package 'Hmisc' was built under R version 4.0.5
## Loading required package: lattice
## Loading required package: survival
## Loading required package: Formula
## Loading required package: ggplot2
## Warning: package 'ggplot2' was built under R version 4.0.5
##
## Attaching package: 'Hmisc'
## The following objects are masked from 'package:dplyr':
##
##     src, summarize
## The following objects are masked from 'package:base':
##
##     format.pval, units
```

```
attach(forestFireData)
```

```
## The following objects are masked from forestFireDataNew:
```

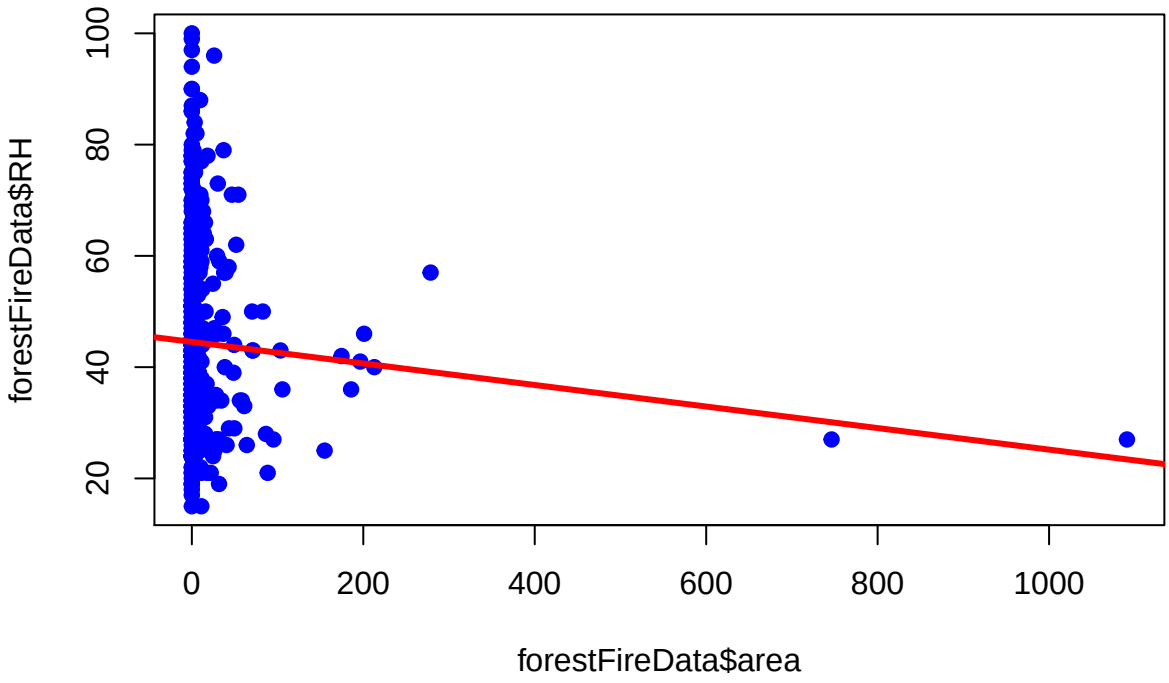
```
##
##     area, day, DC, DMC, FFMC, ISI, month, rain, RH, temp, wind
```

```
forestFireData$month[forestFireData$month == "jan"] <- 1
forestFireData$month[forestFireData$month == "feb"] <- 2
forestFireData$month[forestFireData$month == "mar"] <- 3
forestFireData$month[forestFireData$month == "apr"] <- 4
forestFireData$month[forestFireData$month == "may"] <- 5
forestFireData$month[forestFireData$month == "jun"] <- 6
forestFireData$month[forestFireData$month == "jul"] <- 7
forestFireData$month[forestFireData$month == "aug"] <- 8
forestFireData$month[forestFireData$month == "sep"] <- 9
forestFireData$month[forestFireData$month == "oct"] <- 10
forestFireData$month[forestFireData$month == "nov"] <- 11
forestFireData$month[forestFireData$month == "dec"] <- 12

forestFireData$day[forestFireData$day == "sun"] <- 1
forestFireData$day[forestFireData$day == "mon"] <- 2
forestFireData$day[forestFireData$day == "tue"] <- 3
forestFireData$day[forestFireData$day == "wed"] <- 4
forestFireData$day[forestFireData$day == "thu"] <- 5
forestFireData$day[forestFireData$day == "fri"] <- 6
forestFireData$day[forestFireData$day == "sat"] <- 7

colnames(forestFireData)[which(names(forestFireData) == "day")] <- "day_of_week"

plot(forestFireData$area, forestFireData$RH, pch=19, col="blue")
abline(lm(forestFireData$RH ~ forestFireData$area), col ="red", lwd=3)
```



###F

```
##To predict the area burned by the forest fire on the basis of other variables
## such as wind speed and initial spread index. A large factor that comes with
##grossly raging forest fires are wind speeds that help
##propel the blaze across a land. The initial spread index is a generalized
##expectation of fire spread rate.
```