UCS410

LAB AND PROJECT WORK SUBMISSION

SUBMITTED BY: MANPREET KAUR

ENROLLMENT NUMBER: 101816013

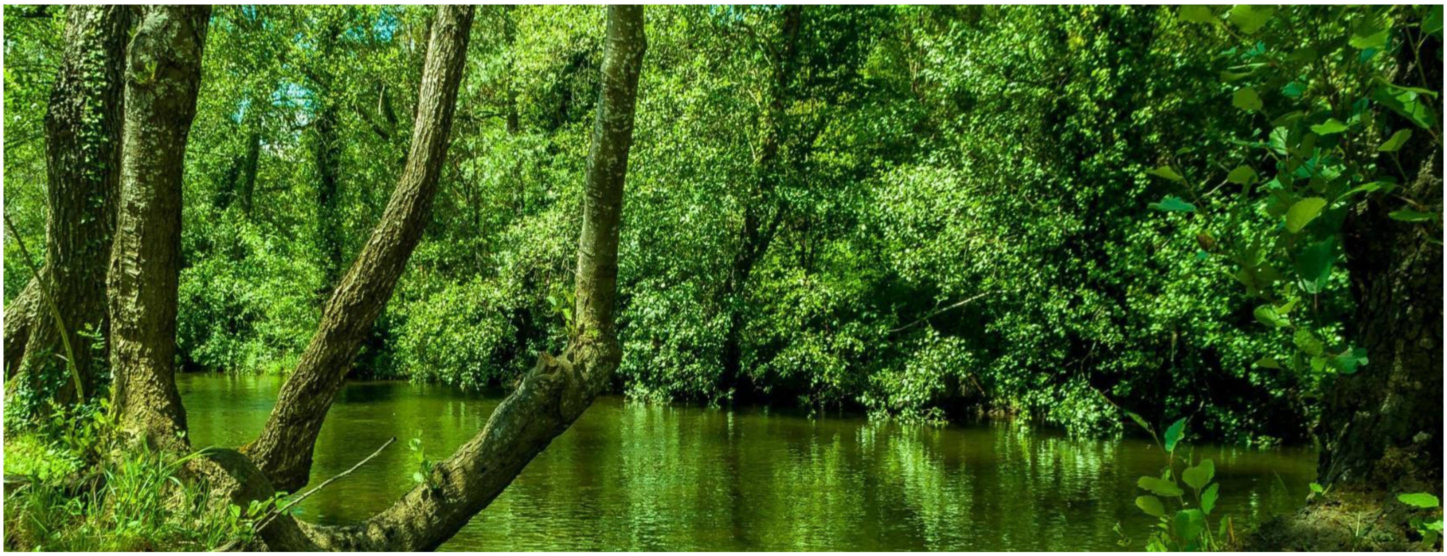
CONTENTS:

PART 1: ANALYSIS OF MONTESINHO FOREST FIRES USING R PROGRAMMING pages 1-34

PART2: LAB ASSIGNMENTS pages 34 -43

ANALYSIS OF MONTESINHO FOREST FIRES USING R

Montesinho is a beautiful protected area located in the municipalities of Vinhais and Braganza, northeastern Portugal. Sections of the southern slopes of the Serra da Coroa (Sierra de la Culebra) fall within the park.



It is home to many different kinds of animals. Its biodiversity includes the Iberian wolf, roe deer, wild boar, Iberian lynx, common genet, red fox and European otter.

The forest due to its **nature** and various other **geographical reasons** is constantly at risk of inevitable forest fires.

Parameters:

Here we are analyzing the Montesinho forest fires data set consisting of 13 variables (1 dependent variable, 4 discrete attributes and 8 continuous attributes) and 1021 entries of data collected over 3 years.

R command used to analyze the variables:

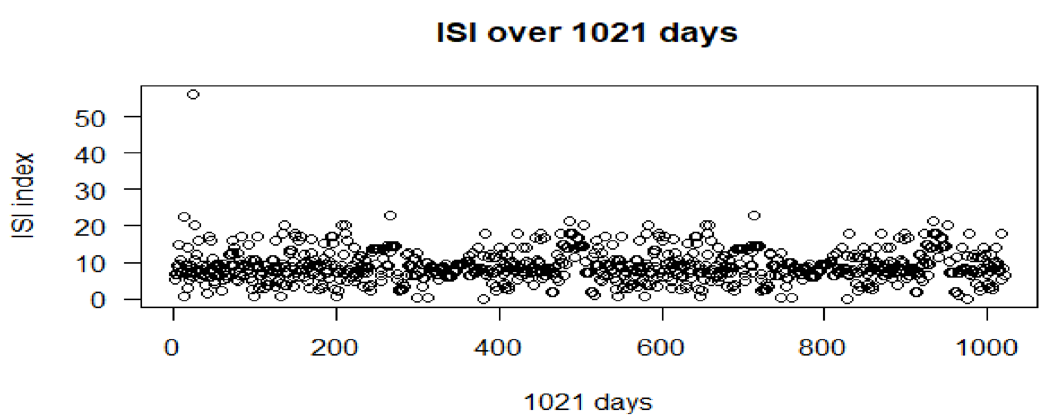
>summary (Parameter)

- Response: area: Burned area of a forest fire in hectares (ha)
- Predictors:
- X: x-axis coordinate of the Montesinho park map: 1 to 9;spatial coordinates
- Y: u-axis coordinate of the Montesinho park map: 2 to 9;spatial coordinates
- **FFMC: Fine fuel Moisture code:** A numerical rating of the moisture content of litter and other cured fine fuels: 18.7 to 96.2
- **DMC: Duff moisture code:** A numerical rating of the average moisture content of loosely compacted organic layers and medium-size woody material: 1.1 to 291.3
- **DC: Drought Code:** A numerical rating of the average moisture content of deep, compact, organic layers: 7.9 to 860.6
- **ISI: Initial Spread index :** A numerical rating of the expected rate of fire spread: 0.0 to 56.10
- **Month:** month of the year: 1 to 12
- **Day:** day of the week: 1 to 7
- **Temp:** temperature in Celsius degrees: 2.2 to 33.30
- **RH:** relative humidity in %: 15.0 to 100
- **Wind:** wind speed in km/h: 0.40 to 9.40
- **Rain:** outside rain in mm/m2: 0.0 to 6.4

Exploration of DATASET

Q1 how can forest fire be calibrated?

Ans. The forest fire is measured by the parameter/independent variable ISI. It stands for Initial spread index. This factor describes the quickness of spread of fire.



\

R commands:

>plot(ISI,main="ISI over 1021 days",xlab="1021 days",ylab="ISI index",las=1)

>summary(ISI)

Min. 1st Qu. Median Mean 3rd Qu. Max.

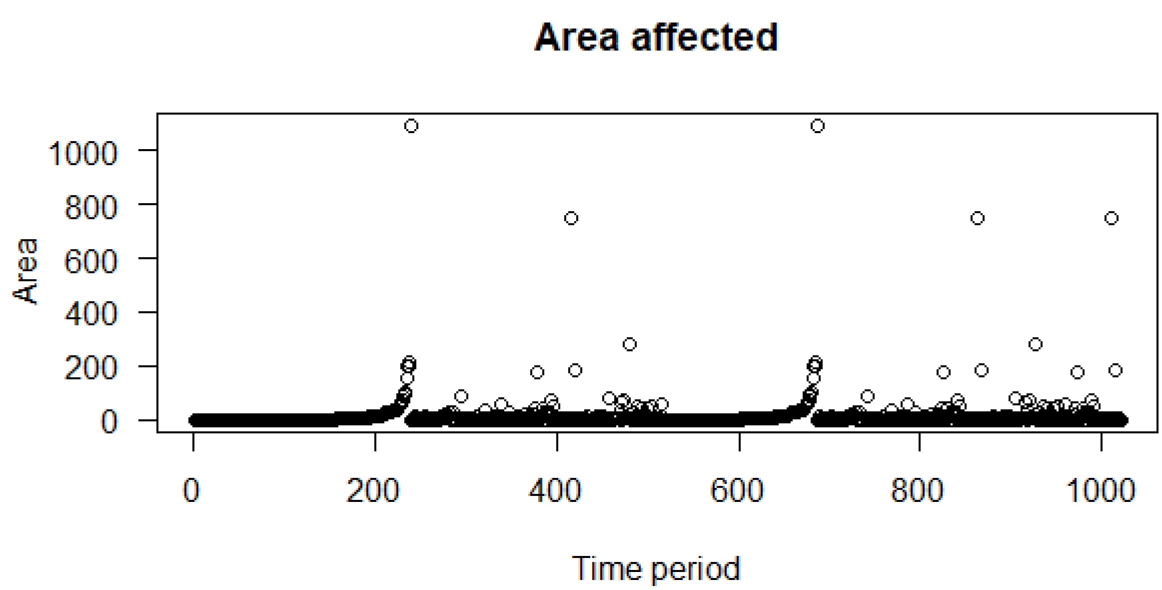
0.000 6.500 8.400 8.982 11.000 56.100

From the above plot it's clear that we have positive values and continuous values of fire spread over time.

From summarization it can be denoted that we have positive real values of quartiles with minimum value of 0,which means no fire, and 56.100 as maximum fire which means greater affected area of forest fires.

Q2. How large the forest fire can be?

The devastating effect of forest fire can be calculated by the area impacted. The area impacted is amounted by the spatial coordinates.



On analysis and summarization of the "area" dependent variable we find that most of the plot density is over the area "0" which means that most of the forest fire is localized and is not spread beyond a unit coordinate. However, this gives us the idea of fire coverage of the fire but does not completely measures the severity of it.

summary(area)

Min. 1st Qu. Median Mean 3rd Qu. Max.

0.09 2.14 6.37 24.60 15.42 1091.00

Thus with the summarization we can divide the range of areas in small median and large extent.

Categorize area by the above information:

(0,2.14) size of area = ‘small’

(2.14,15.42) size of area = ‘median’

(15.42,1091.00) size of area = ‘large’

R commands:

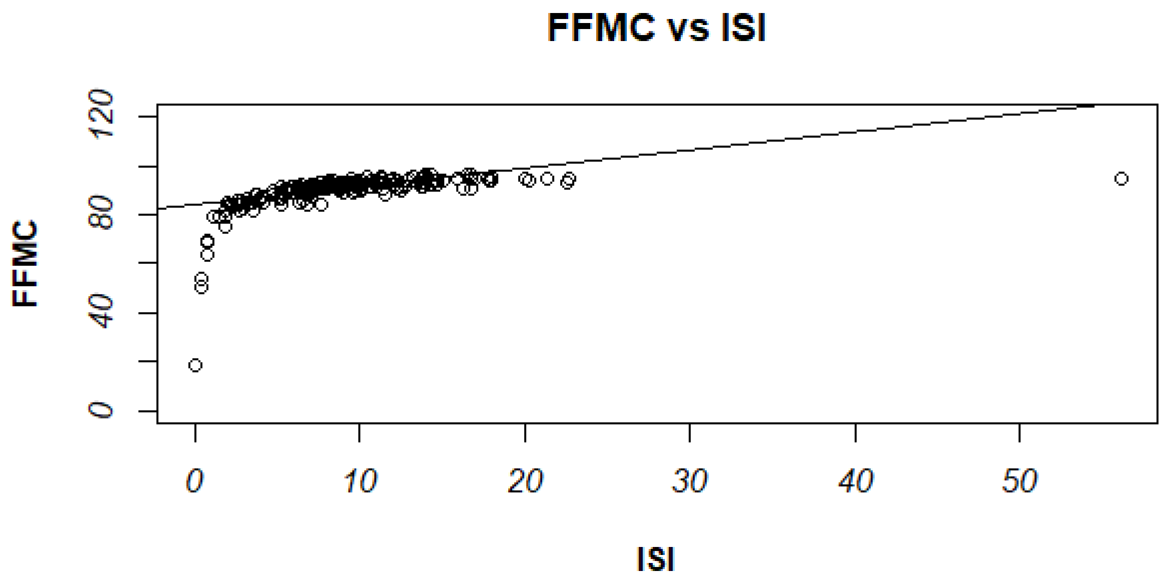
```
>summary(area)

>plot(area, main=”area affected”,xlab=”time period”,ylab=”Area”,las=1)
```

Q3. Forests are natural sources of many types of fuels. How does this affect fire?

The major fuel index in this dataset is the FFMC. It stands for fine fuel moisture code. This statistic represents the fuels in liquid form specific to this forest.

Correlation with the ISI index will amount to the fires which had fuels from the forest play a role in it.

We had covariance factor of 0.5393 which indicates the directional relationship of these variables.



R commands:

>plot(ISI,FFMC, main="FFMC vs ISI",ylim=c(0,120), font.lab=2,font.axis=3)

> abline(lm(F cor(ISI,FFMC)

>cor(FMC~ISI))

0.5393548

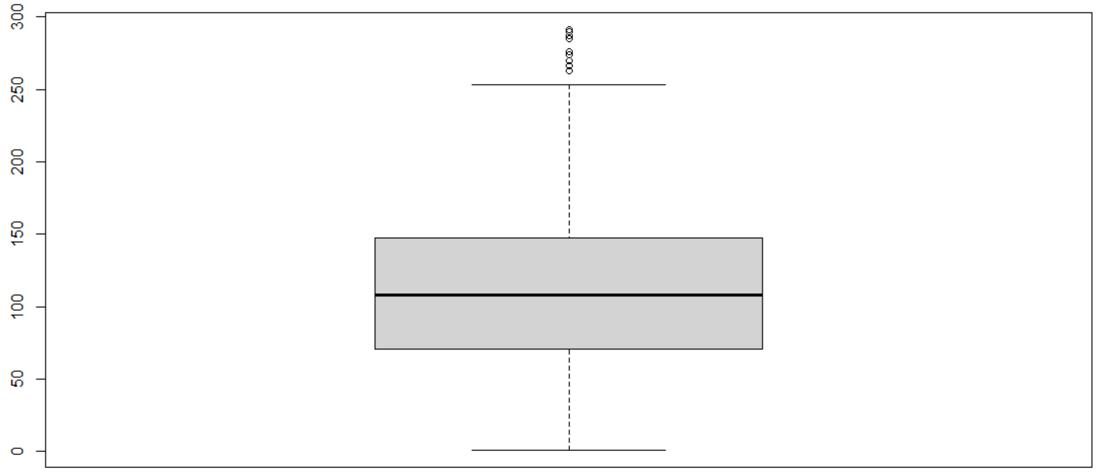
Q4. Can the moisture of forests have an impact on the fire?

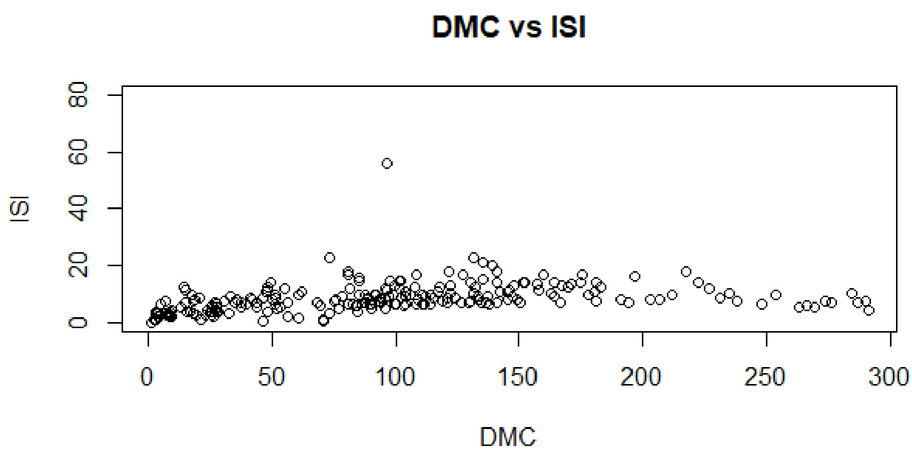
The **DMC** (DUFF MOISTURE CODE) indicates the moisture content of loosely-compacted organic layers of moderate depth. It is representative of the duff layer that is 2-4 inches (5-10 cm) deep, and has a fuel loading of about 22 tons per acre (50 t/ha).

As DMC fuels are slower in burning than the FFMC predicting the probability of **fire** ignition by lightning.

It turns out DMC can be a factor for forest fire but its covariance is less than that of FFMC, which means its less productive than the FFMC

The following boxplot represents the DMC as parameter effective of forest fire





We observe that DMC is a factor in forest fire

>boxplot(DMC)

>plot(DMC,ISI,main="DMC vs ISI",ylim=c(0,80))

>cor(ISI,DMC)

Q5. Is DC a index of potential as compare to FFMC, DMC?

This is the third moisture **index** reflects the moisture content of compact organic layers, 10-20 cm deep. **DC is** indicative of long-term moisture conditions and deep burning **fires**, being related to mop-up and patrol difficulties.

It has lesser covariance than DMC and FFMC therefore, the factor is less promising indicator of fire.

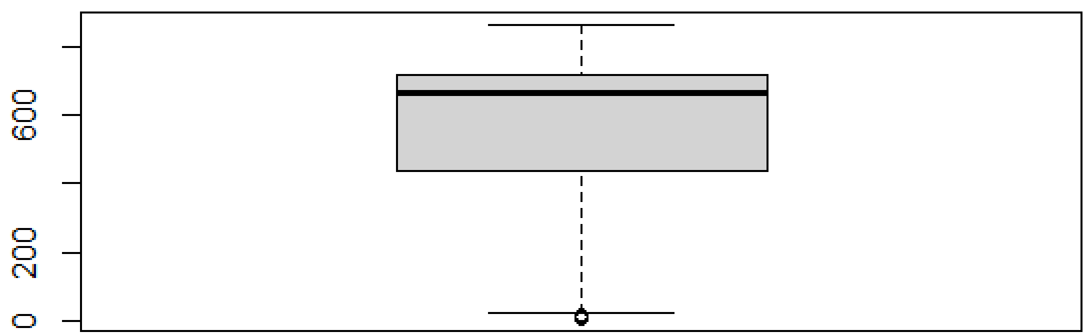
R commands:

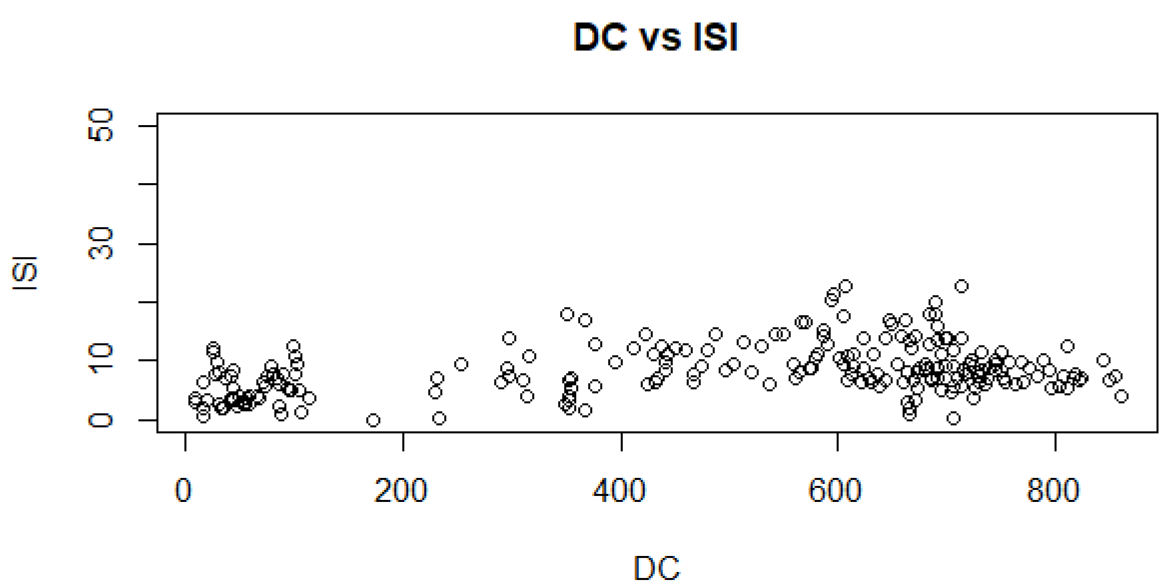
>boxplot(DC)?check on the factor

>cor(ISI,DC)

[1] 0.2565189

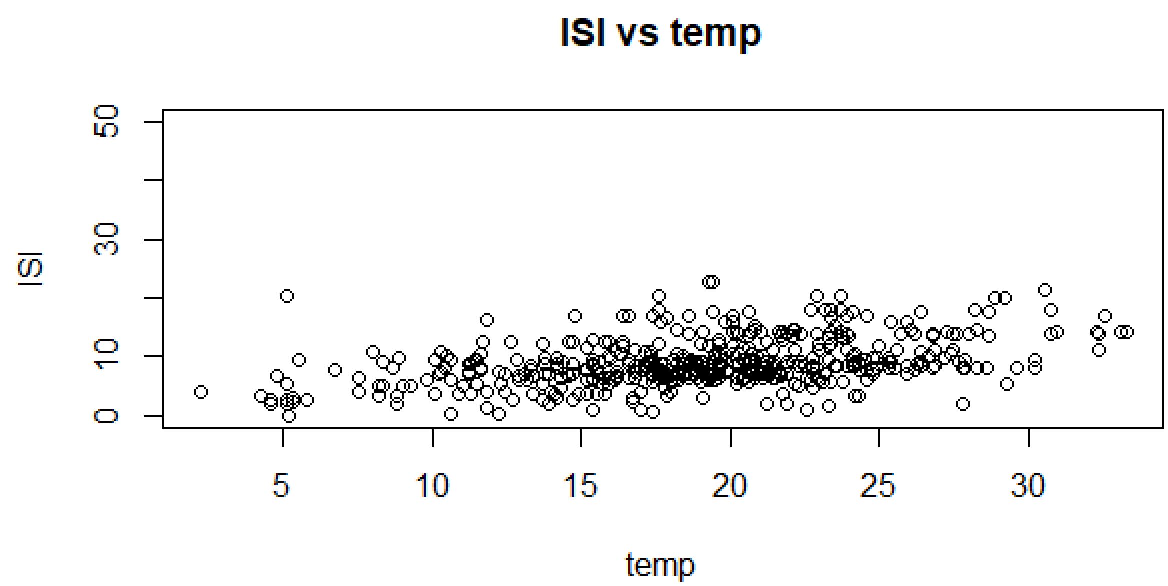
> plot(DC,ISI,main="DC vs ISI",ylim=c(0,50))





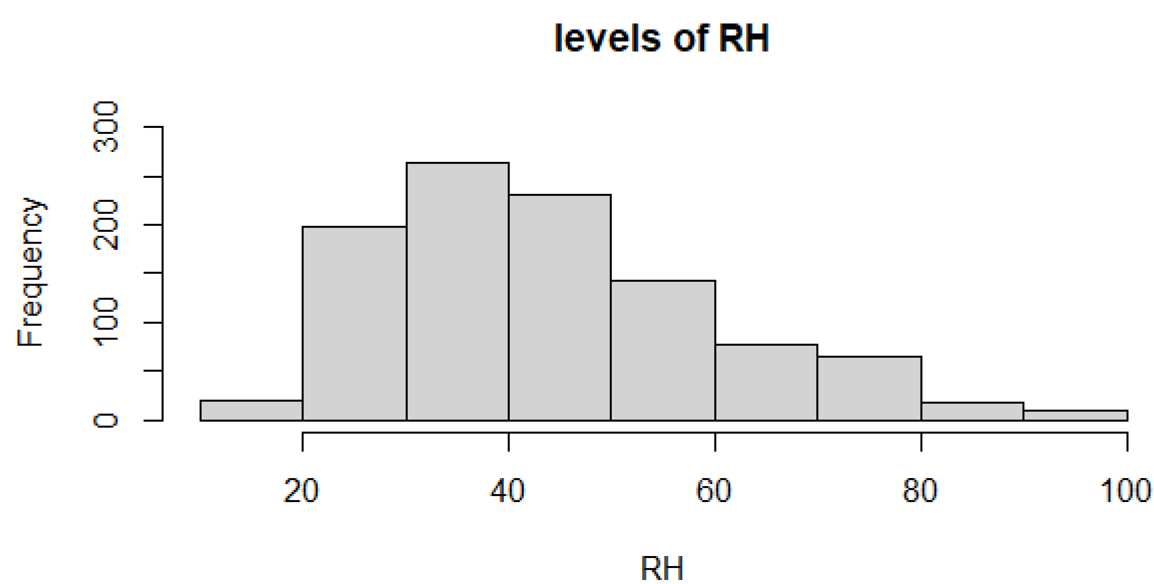
Q6. Will temperature be causative of forest fire?

By the analysis of the plot in the R it clear that the temperature in the range of 15 and 30, supports the ISI index. Hence, the optimal temperature provides the optimal thermal conditions for the fire.




```
>plot(temp,ISI,main="ISI vs temp",ylim=c(0,50))
>summary(temp)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   2.2    15.5    19.4    19.0    22.9    33.3
```

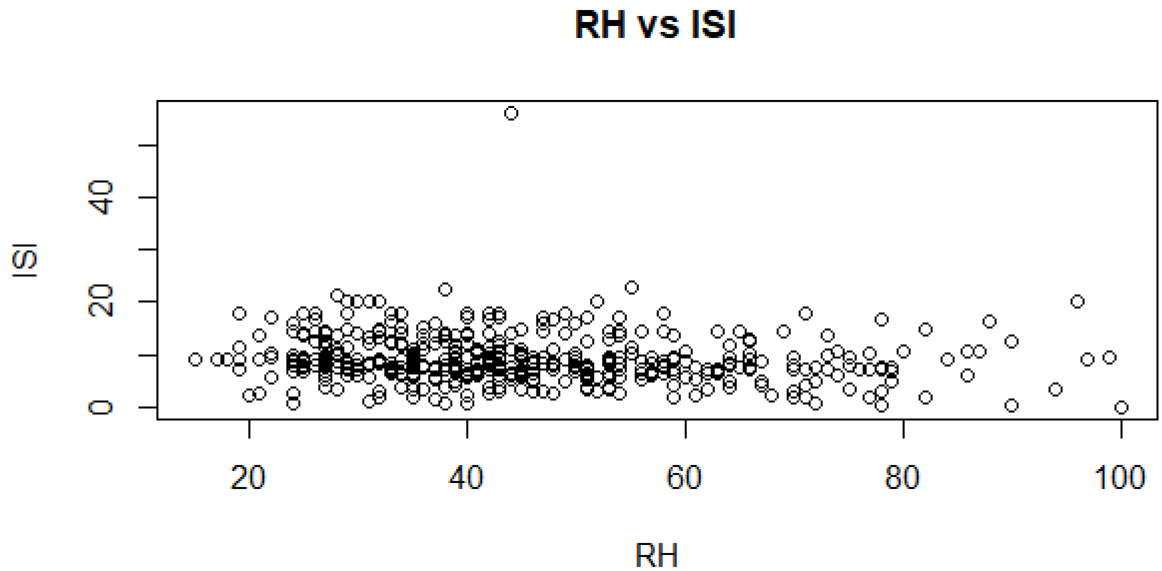
Q7. High humidity in atmosphere can lower the ignition, how does it work in case of forest fires?



The major effect is seen in the frequency of forest fires. The 20-40 bracket depicts dense region of forest fires whereas the 80-100 bracket depicts minimal forest fires.

cor(ISI,RH)

[1] -0.1598274h



However its covariance states that it might not be a complete factor towards determining the forest fire.

R commands:

> hist(RH,main="levels of RH",ylim=c(0,300))

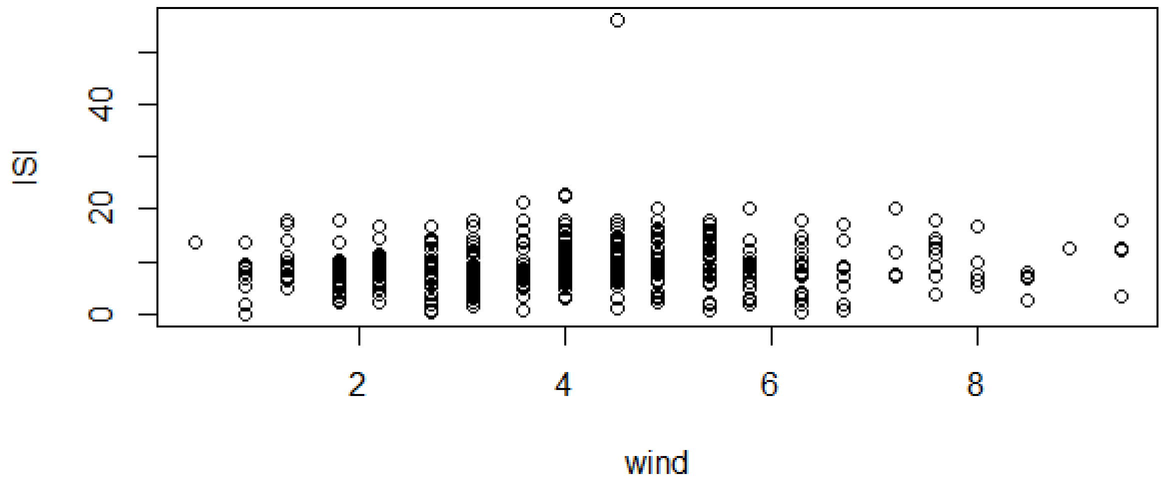
Q8. Wind at certain angle to fire may support the fire. The phenomenon is called Fanning effect.How certain is the wind parameter for the phenomenon?

>plot(wind,ISI)

> cor(ISI,wind)

The phenomenon is confirmed to happen as the forest has a regular trend with the wind and has a positive corelation index,which means that the fanning effect is a significant causative of forest fire in Montesinho park.

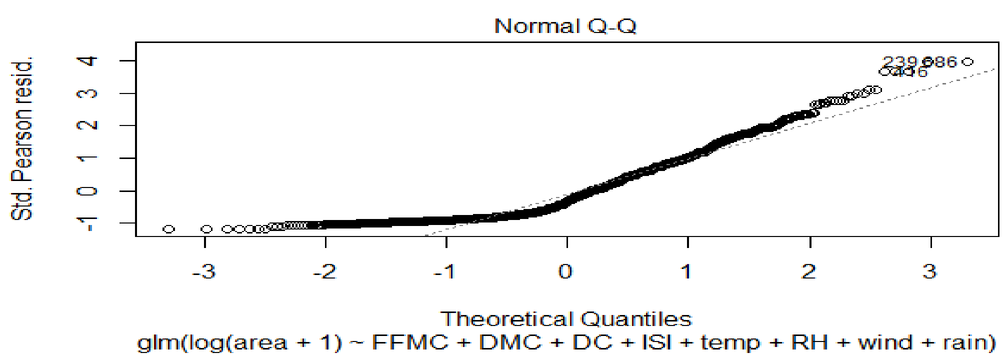
[1] 0.1094243



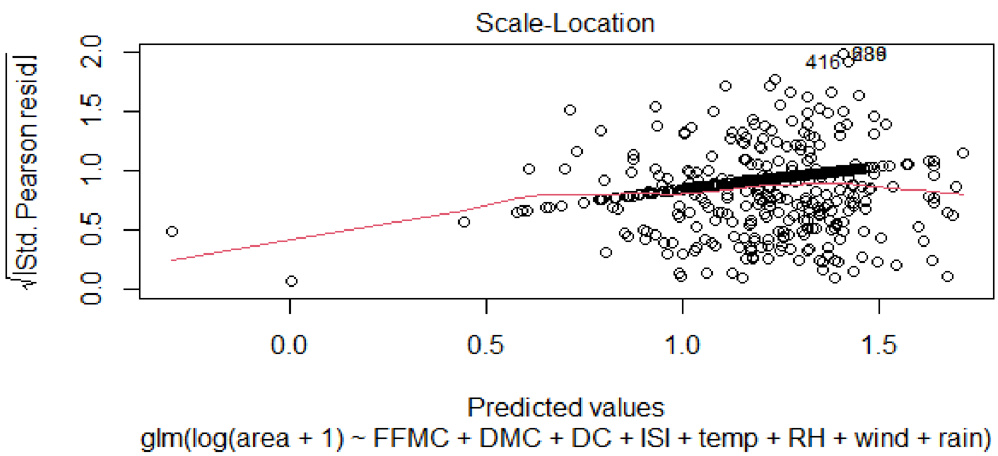
Q9.What variables are most significant?

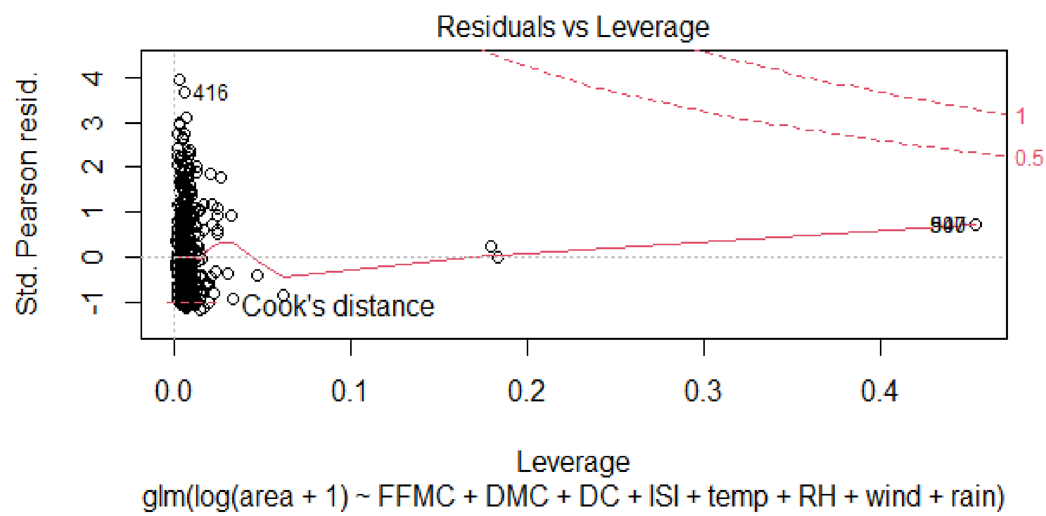
The most significant variables are the ones which are potential of indicating or conclusive about the forest fires.

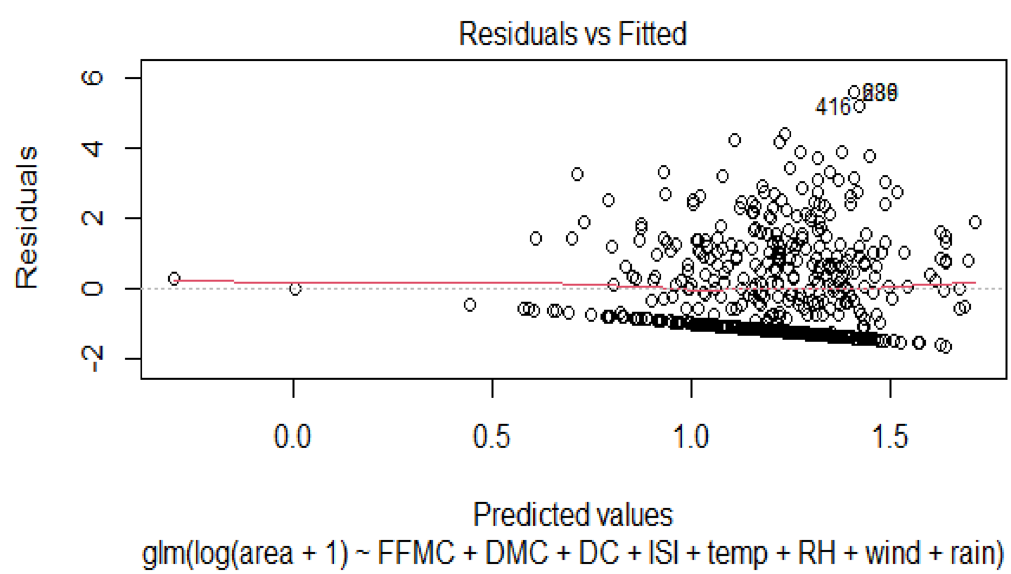
This analyis can be done with the help building of generalized linear regression model.This model has helped with its evaluation of plots such as Normal Q-Q,Location Scaling,Residual vs fitted, and the residual plots helped to develop a certain idea on the fact that what factors might be contributing the most to analyze the forest fires.



\


Scale-Location
√|Std. Pearson resid|
0.0
0.5
1.0
1.5
2.0
416
Predicted values
glm(log(area + 1) ~ FFMC + DMC + DC + ISI + temp + RH + wind + rain)


Residuals vs Leverage
Std. Pearson resid.
416
1
0.5
Cook's distance
0.0
0.1
0.2
0.3
0.4
Leverage
glm(log(area + 1) ~ FFMC + DMC + DC + ISI + temp + RH + wind + rain)


Residuals vs Fitted
Residuals
416
0.0
0.5
1.0
1.5
Predicted values
glm(log(area + 1) ~ FFMC + DMC + DC + ISI + temp + RH + wind + rain)

R commands:

Regressional model 1

```
>fit = glm(log(area+1)~FFMC+DMC+DC+ISI+temp+RH+wind+rain)

>plot(fit)

> summary(fit)


Call:

glm(formula = log(area + 1) ~ FFMC + DMC + DC + ISI + temp +

    RH + wind + rain)


Deviance Residuals:

    Min       1Q   Median       3Q      Max

-1.6394  -1.1774  -0.4596   0.8941   5.5884


Coefficients:

             Estimate Std.   Error t value  Pr(>|t|)

(Intercept)  0.1329334    0.9268847   0.143  0.88599

FFMC         0.0089377   0.0097849   0.913  0.36124

DMC          0.0005631   0.0010100   0.558  0.57728

DC           0.0005079     0.0002619   1.939  0.05274 .

ISI         -0.0297833       0.0131476   -2.265  0.02371  *

temp         0.0034369      0.0124561   0.276  0.78266

RH          -0.0047651      0.0037992  -1.254  0.21005

wind         0.0838409      0.0265166   3.162  0.00161  **

rain         0.0754290        0.1525582   0.494  0.62111

---
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 2.013344)

Null deviance: 2086.1 on 1020 degrees of freedom

Residual deviance: 2037.5 on 1012 degrees of freedom

AIC: 3622.9

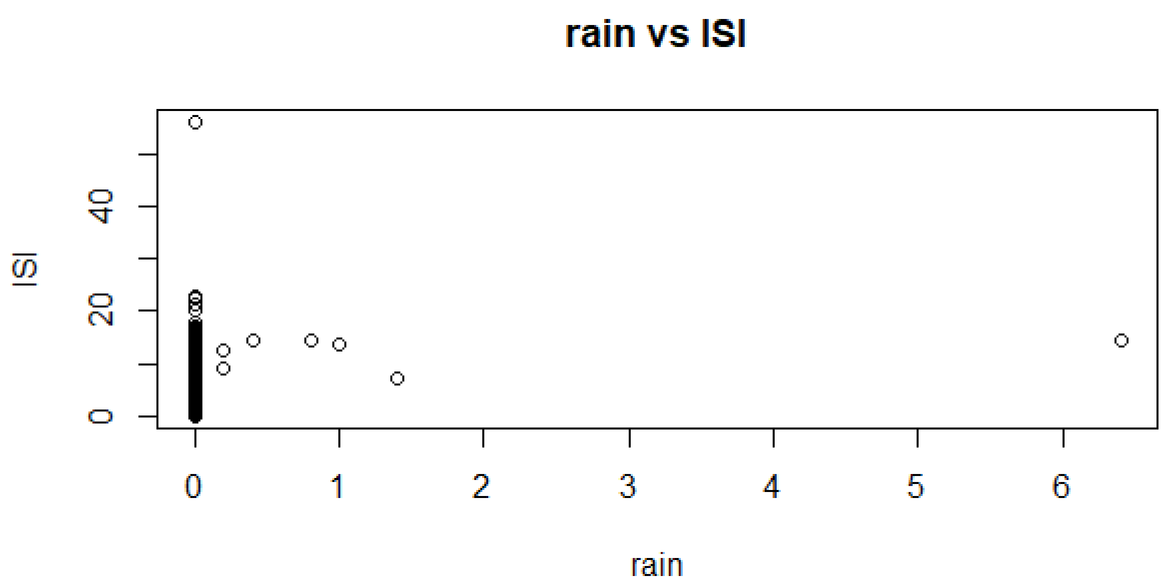
Number of Fisher Scoring iterations: 2

Thus upon the analysis of the estimated errors and possible 2 fisher iterations it can be concluded that lower error values can be the most accurate with combined with relative high estimate standard.

The FFMC has the highest estimate Std. with a considerable lower error therefore it's the most productive factor.

The DMC and RH are also productive factors with sufficient estimation value and lower err0r.

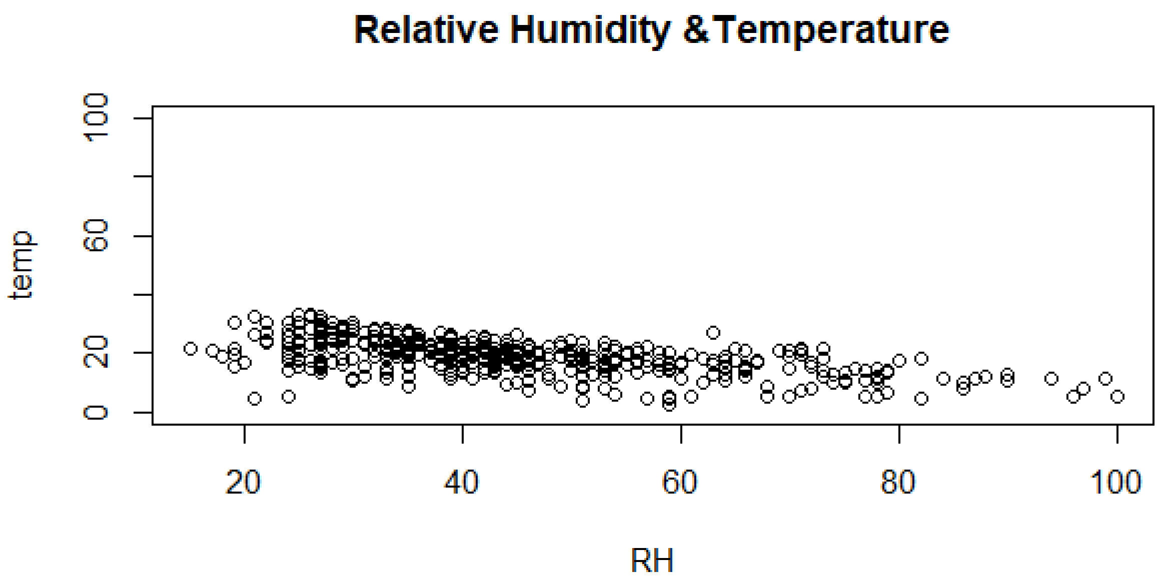
Q10. Does precipitation acts as fire extinguisher?



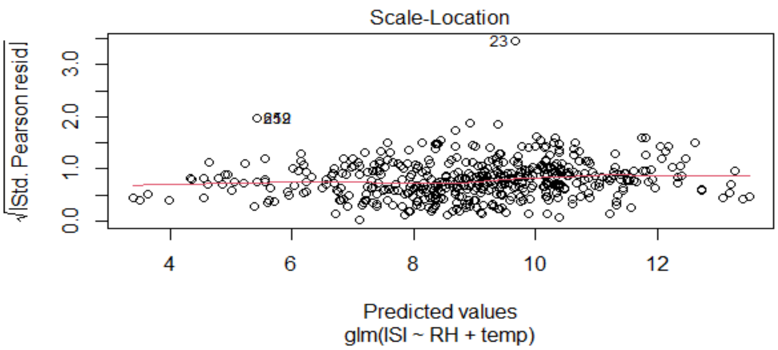
Yes, when there is no rain the trends of the forest fire are the highest. There is scarce observation of the precipitation supporting fire.

Q11. With higher temperature and lower humidity the atmosphere will facilitate, is the notion true?

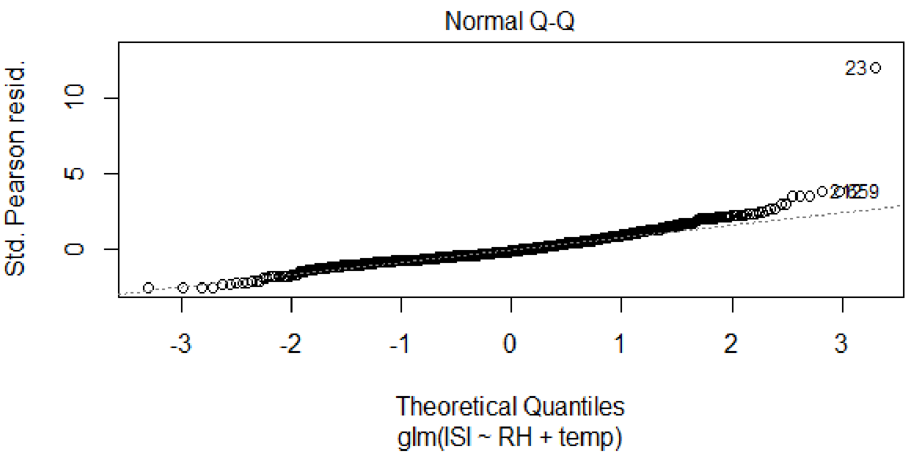
We have a min (RH) of 15 and max (temp) of 33.3.



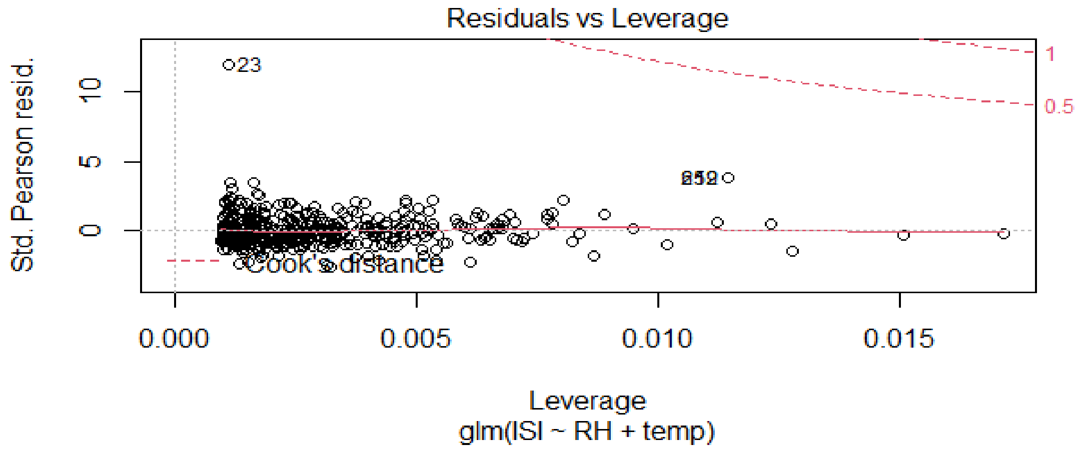
We see that a certain lower amount of RH near about 25 has the maximum higher temperature density therefore is sound conditions to the forest fires. Thus a regression analysis with the RH and temp we can ascertain the notion.

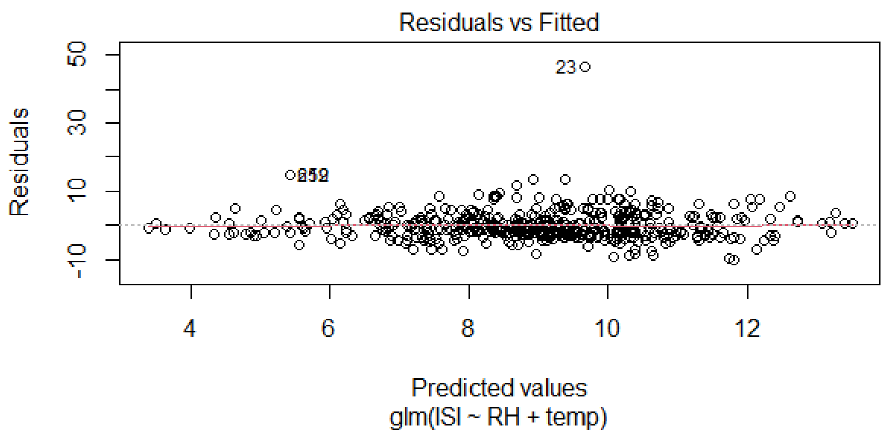


Scale-Location
√|Std. Pearson resid|
Predicted values
glm(ISI ~ RH + temp)



Normal Q-Q
Std. Pearson resid.
Theoretical Quantiles
glm(ISI ~ RH + temp)



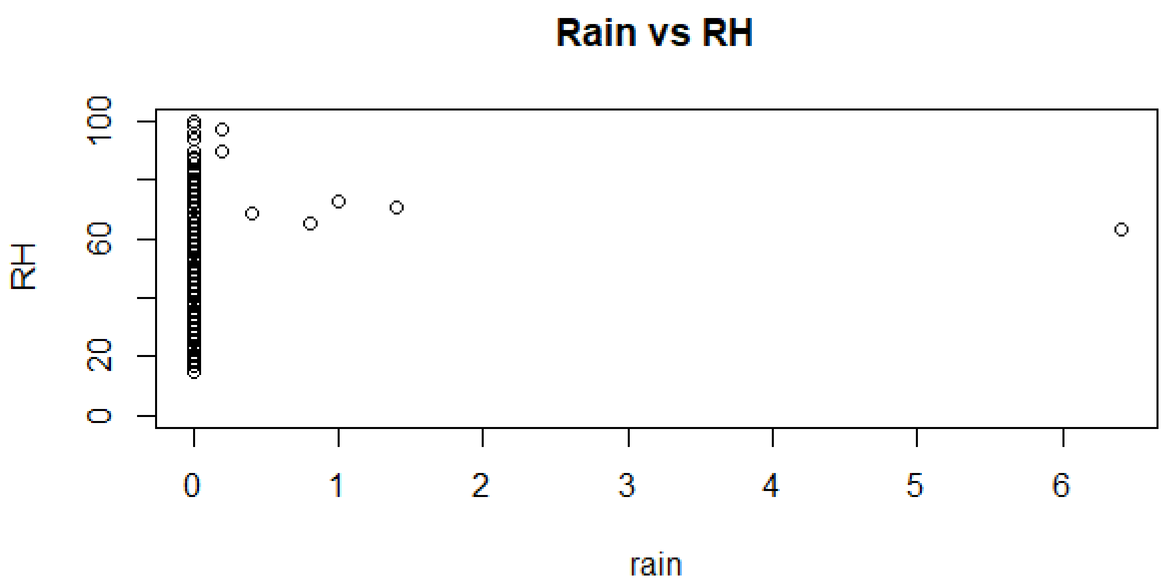


The high density of the regression models ascertain that there is an intersection in the region of humidity and RH. With a sufficiently negative covariance value it is indicated of the inverse relationship between the RH and temperature. Also, the density intersection plots are of the lower humidity and higher temperature in regards to the ISI which is an index to fire.

R commands:

```
>plot(RH,temp,type="p",main="Relative Humidity &Temperature",ylim=c(0,100))

>mod1<-glm(ISI~RH+temp)

>plot(mod1)

>cor(RH,temp)

 [1] -0.5327149
```

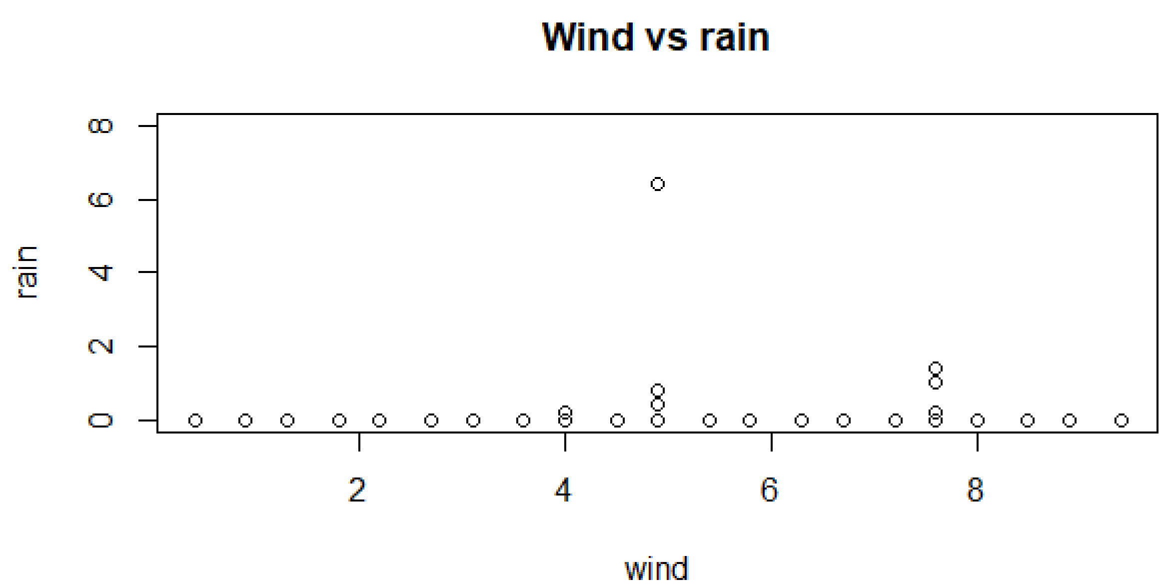
Q12. Is precipitation the source of humidity in the Montesinho forest fires?



The plot shows that rain is not the principle source of humidity to the Montesinho forest park.

Although, scarcely due its nature it does effect the humidity.

Q13. How other climatic factors such as wind and rain affect the Montesinho forest fire region?



Upon analysis it turns out that wind and rain are not very correlational to each other .Winds does support fire but rain contains water which extinguishes fire ,therefore , these climatic factors do affect the forest fires individually but combined they are oppositely affect the forest fires.

Q.14 What can we say about the weather/climatic conditions of Montesinho forest park?

With previously plotted statistics it can be said that the forest park lies on the rain shadow region. As in terms of meteorology the rain fall is scarce and temperatures are 2.2 and 33.3.It is a windy region with constant amount of wind exposure. It is naturally humid area (as precipitation is scarce)with abundance of natural forest fuels.

Q15. What is the amount of forest fire in which essential fuel resources are burnt?

On estimation of total amount of forest fire index to the total fuel resources of the soil.

It turns out about 1.18 % of the total forest essential fuels are burnt.

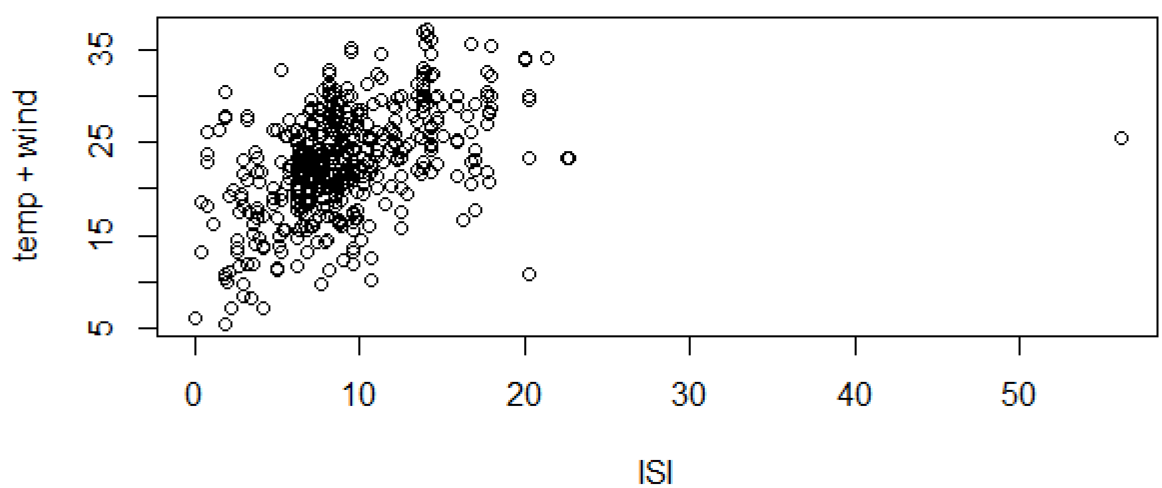
R commands:

>sum (FFMC+DMC+DC)

>sum (ISI)

>sum(ISI)/sum(FFMC+DC+DMC)

Q16. How can higher temperature and higher wind blows be a favorable condition for forest fire, both being favorable independently?



It is seen that highly dense plots of temperature and wind have decent ISI, therefore, temperature and wind are both favorable factors.

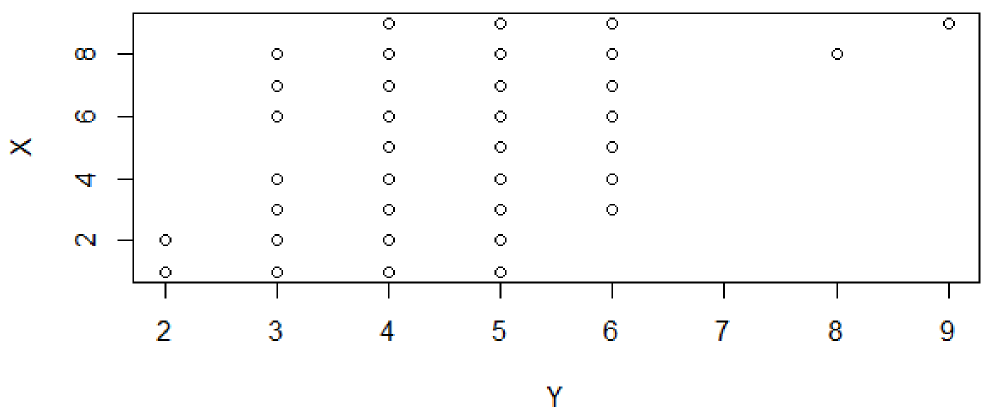
R command:

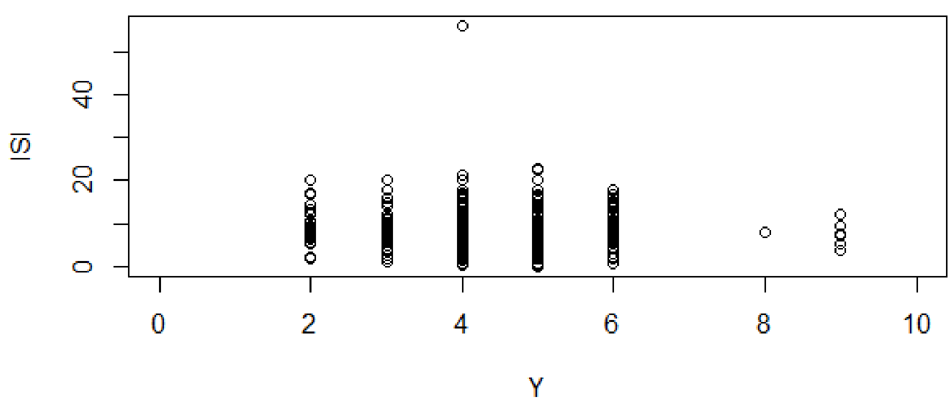
plot(ISI,temp+wind)

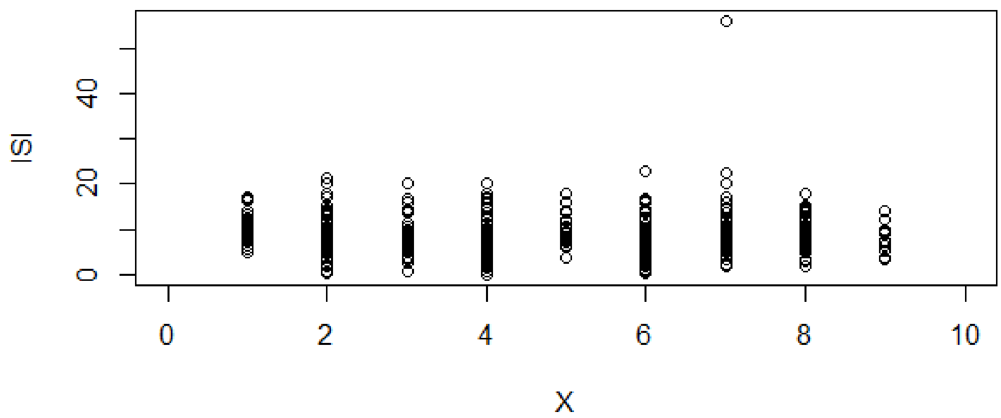
Q.17 What are the most and least vulnerable areas of the forest?

The most vulnerable regions of forest fire can be calculated with the high **ISI** among the spatial coordinates.

X˜Y+ISI







The spatial coordinates of Y with 2, 3,4,5,6 have high forest fires.

The spatial coordinates of X with 1, 2, 3, 4, 6, 8 have high forest fires.

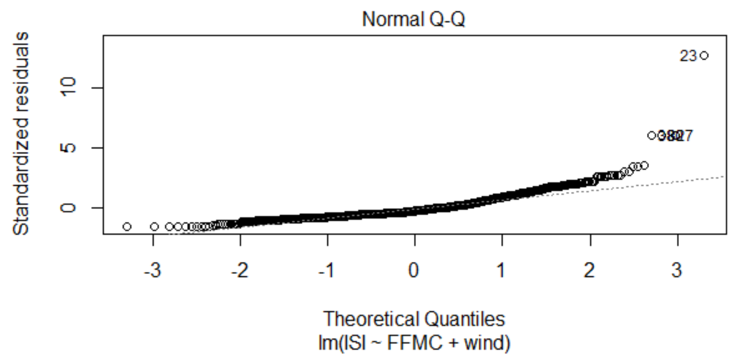
Thus (2,2),(3,3)(4,4),(5,5) are the most vulnerable regions of forest fire.

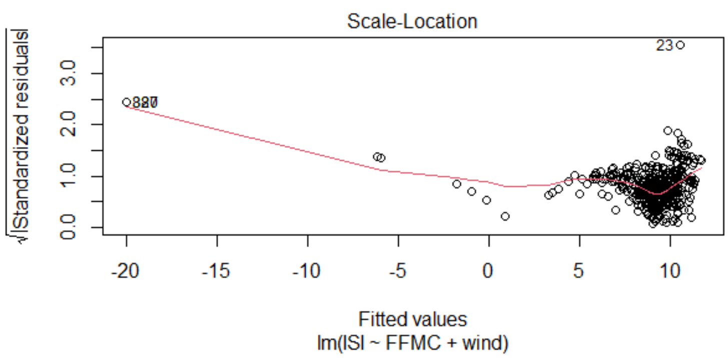
The spatial coordinates of (9, 9) are the least vulnerable regions of forest fires.

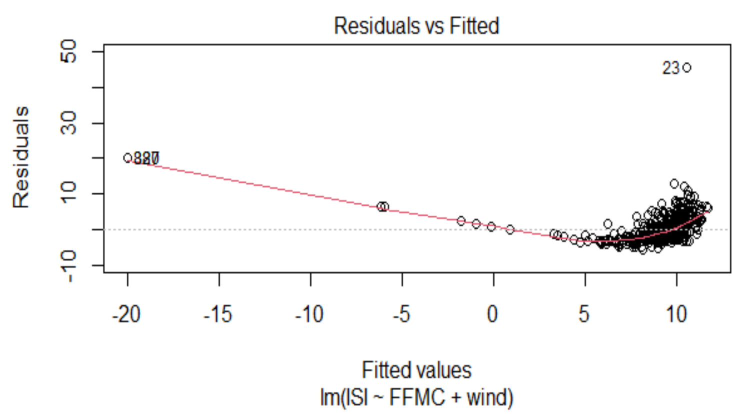
Q18. How can we calibrate the fire spread in the forest region ?

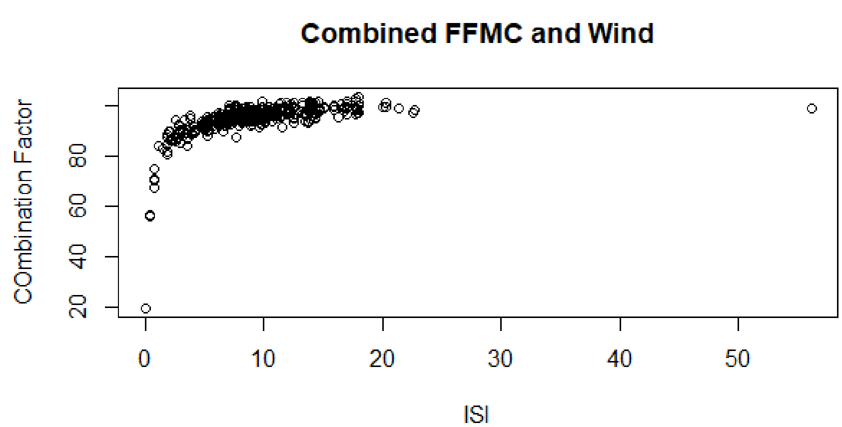
ISI - This index combines FFMC and wind speed, being a good indicator for fire spread.

```
>m1<-lm(ISI~FFMC+wind)

> plot(m1)
```









This is a precise estimation with highly dense and calculative plots.

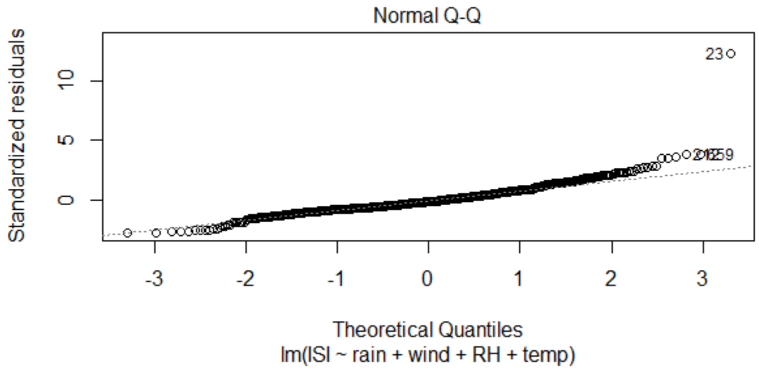
Thus ISI combined with FFMC and Wind is a good factor for determination of forest fires in the region of Fast wind blows and high FFMC.

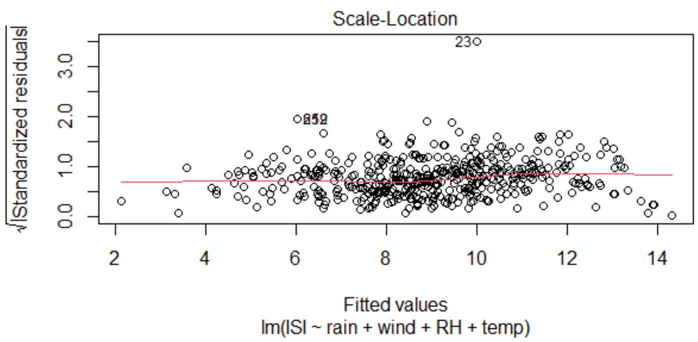
Q19. Can a better calibration be obtain ISI and total fuel index with wind and RH, all being the favorable factors?

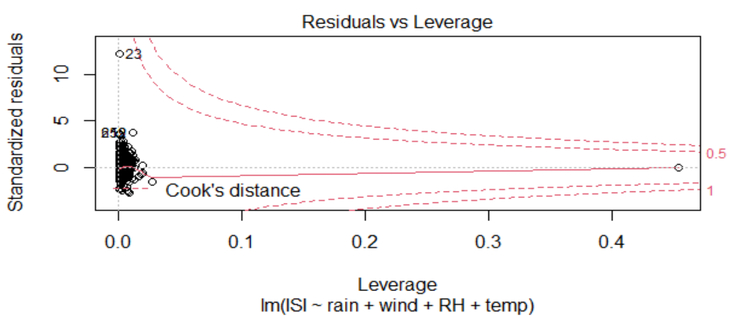
```
> mod4<-lm(ISI~rain+wind+RH+temp)

> summary(mod4)


Call:

lm(formula = ISI ~ rain + wind + RH + temp)


Residuals:

    Min      1Q  Median      3Q     Max

-10.124  -2.333  -0.417   1.769  46.093


Coefficients:

            Estimate Std. Error t value Pr(>|t|)

(Intercept) -1.896328   0.857387  -2.212  0.02721 *

rain         0.165362   0.404945   0.408  0.68310

wind         0.532464   0.067925   7.839 1.14e-14 ***

RH           0.029487   0.008721   3.381  0.00075 ***

temp         0.390960   0.024679  15.842  < 2e-16 ***

---

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1


Residual standard error: 3.775 on 1016 degrees of freedom

Multiple R-squared:  0.2353, Adjusted R-squared:  0.2322
```

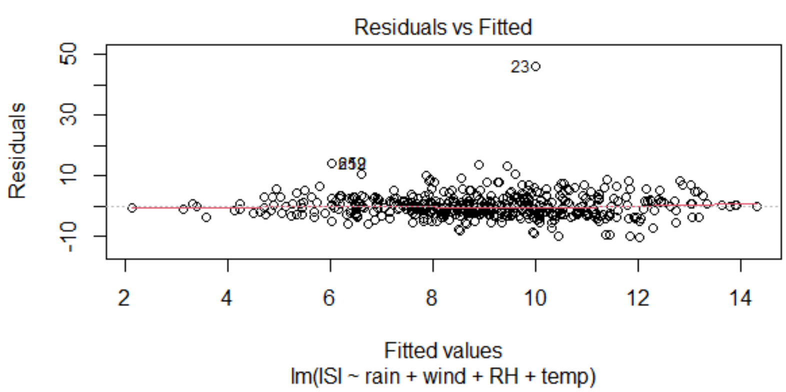
F-statistic: 78.14 on 4 and 1016 DF, p-value: < 2.2e-16

>plot(mod4)





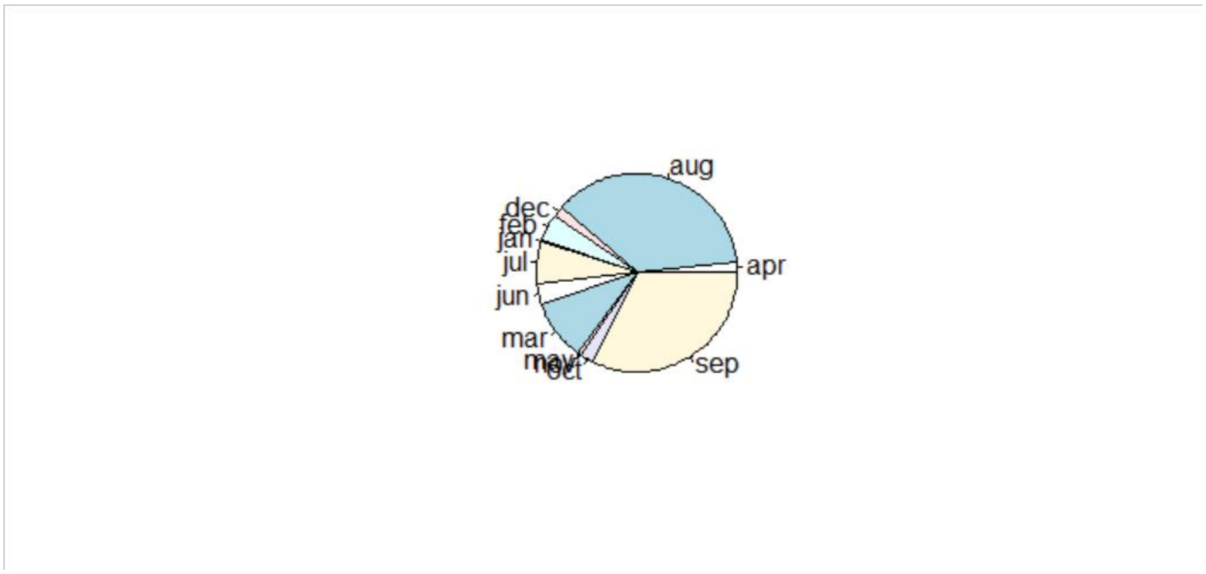




We see that combined favorable factors on an improved regression model give us even better estimation of the of the forest fires.

We obtain dense standardized plots, which indicate that fuels, wind, high temperature and low relative humidity are the certain factors facilitating forest fires.

Q 20. Montesinho Forest park is a tourist attraction, what can we suggest tourists in terms of safety?



On Plotting the forest fires with their month of occurrence frequency it is observed that August, September and March are the months of ravaging forest fires with maximum occurrence

Whereas the safest months of forest fires are January, May, October, November, and December. So, in order to ensure maximum safety we need to open the park for visits only in the recommended months for high tourists intake.

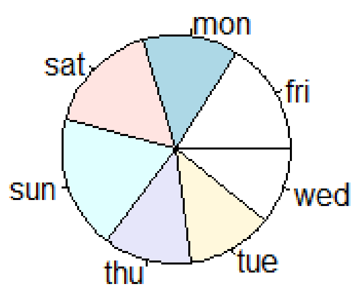
```
>mon<-c(month)

> monFac<-as.factor(mon)

> hist(table(monFac), freq=TRUE, xlab = "Month", ylab = "Frequencies")

> monFac
```

Q21. What factors can account to advise forest departments, for better management of forest fires?

The most vulnerable coordinates and months accountable for high forest fires can provide significant information for forest department to patrol the forests and use the suitable methods to extinguish fire. The forest fuels burning ignite and swell the fires. Control and timely extraction of resources can benefit the forest as well the humans. Also, the forest is prone to fires naturally as determined by climatic factors and fuel indexes, therefore, better fire emergency evacuation methods should be adopted.

Q 22. Are we in a position to predict when the next forest fire can happen/are forest fires predictable?

We take the hypothesis that forest fires are predictable: and test it with t.test()



>summary(dFac)

Fri Mon Sat Sun Thu Tue wed

165 138 165 193 126 123 111

However the days seem equally likely to have a forest fire, but weekend day Sunday (193,the most) and Saturday and following day Monday have comparatively higher chance of the fire.

The days mentioned above in the months of August September and March have higher chances of fire, as compare to other days.

To determine it with the best fitted fuel index (high covariance), and its fire index.

>x1<-rnorm(ISI)

>x2<-rnorm(FFMC)

>t.test(x1,x2)

t.test(x1,x2)

Welch Two Sample t-test

data: x1 and x2

t = -1.6242, df = 2037.7, p-value = 0.1045

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-0.16272585 0.01529227

sample estimates:

mean of x   mean of y

-0.006204541 0.067512251

With confidence interval of 95% the forest fires can be ascertained with the given parameters.

A **negative t-value** of -1.62 indicates a reversal in the directionality of the effect, which has no bearing on the significance of the difference between groups. Therefore indicates the independence of the variables.

With the summarization of the best fitted model we have r squared value :

Residual standard error: 3.775 on 1016 degrees of freedom

Multiple R-squared: 0.2353, Adjusted R-squared: 0.2322

This indicates the model is fairly good, accounting the fact natural disasters are very uncertain,we can predict with an accuracy of 23%

```
> mod8<-lm(ISI~FFMC*DMC*DC)

> summary(mod8)


Call:

lm(formula = ISI ~ FFMC * DMC * DC)


Residuals:

   Min      1Q  Median      3Q     Max
```

-4.007 -2.234 -0.798 1.168 42.601

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -5.121e+00 3.065e+00 -1.671 0.09511 .

FFMC   1.102e-01 3.564e-02   3.093 0.00204 **

DMC   -6.334e-01 1.081e-01 -5.861 6.21e-09 ***

DC   -1.178e-03 1.055e-02 -0.112 0.91116

FFMC:DMC   7.752e-03 1.191e-03   6.507 1.20e-10 ***

FFMC:DC   6.631e-05 1.182e-04   0.561 0.57506

DMC:DC   3.531e-04 1.805e-04   1.957 0.05067 .

FFMC:DMC:DC -4.951e-06 2.003e-06 -2.472 0.01360 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.15 on 1013 degrees of freedom

Multiple R-squared: 0.4689, Adjusted R-squared: 0.4653

F-statistic: 127.8 on 7 and 1013 DF, p-value: < 2.2e-16

We have about 46% accuracy

Now our finally tested mo > summary(mod9)

Call:

lm(formula = ISI ~ FFMC * DMC * DC * wind)

Residuals:

```
   Min     1Q Median     3Q    Max
-7.789 -1.776 -0.578  1.104 40.659

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        1.884e+01  7.005e+00   2.690  0.00727 **
FFMC              -1.971e-01  8.174e-02  -2.411  0.01607 *
DMC                2.013e+00  2.928e-01   6.876 1.08e-11 ***
DC                -1.066e-01  3.646e-02  -2.923  0.00355 **
wind              -7.039e+00  1.698e+00  -4.144 3.69e-05 ***
FFMC:DMC          -2.110e-02  3.191e-03  -6.613 6.11e-11 ***
FFMC:DC            1.253e-03  4.066e-04   3.081  0.00212 **
DMC:DC            -3.194e-03  5.329e-04  -5.994 2.86e-09 ***
FFMC:wind          8.864e-02  1.947e-02   4.552 5.96e-06 ***
DMC:wind          -7.947e-01  8.999e-02  -8.831  < 2e-16 ***
DC:wind            2.217e-02  6.913e-03   3.207  0.00138 **
FFMC:DMC:DC        3.376e-05  5.866e-06   5.755 1.15e-08 ***
FFMC:DMC:wind      8.640e-03  9.774e-04   8.840  < 2e-16 ***
FFMC:DC:wind      -2.488e-04  7.715e-05  -3.225  0.00130 **
DMC:DC:wind        1.100e-03  1.455e-04   7.562 8.92e-14 ***
FFMC:DMC:DC:wind -1.199e-05  1.592e-06  -7.530 1.13e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.879 on 1005 degrees of freedom
```

Multiple R-squared: 0.56,   Adjusted R-squared: 0.5534

F-statistic: 85.28 on 15 and 1005 DF, p-value: < 2.2e-16 del has an accuracy of about:

We have accuracy of about 56% to predict the forest fires based on the factors provided

R code overall for fitting, training and testing models

```
#import the Montesinho_park Dataset from the folder via .csv file ie comma seperated value

Montesinhopark_forestfires <- read.csv("D:/Academic/Montesinhopark_forestfires.csv")

View(Montesinhopark_forestfires)

hist(Montesinhopark_forestfires$X)

summary(Montesinhopark_forestfires)

names(Montesinhopark_forestfires)

attach(Montesinhopark_forestfires)

mean(X)

#another approach when data is detached and and we use a dollar sign

detach(Montesinhopark_forestfires)

mean(Montesinhopark_forestfires$FFMC)

class(FFMC)

summary(Montesinhopark_forestfires)

dim(Montesinhopark_forestfires)

Montesinhopark_forestfires$FFMC[5:20]#gives the value of FFMC from 5 to 20

Montesinhopark_forestfires$FFMC[5:20,]

mean(FFMC[day=="mon"])#value of FFMC of the day mon 's mean

x1<-Montesinhopark_forestfires[month=="mar",]#subset the data to avoid error //y can extract more of the code

x2<-Montesinhopark_forestfires[month=="feb",]

x3<-Montesinhopark_forestfires[FFMC>50,]

rain[100:200]

area[400:800]

temp<-rain>5#will return the rain units greater than 5
```

```
temp2<-as.numeric(rain>0)

r<-rain>0 & month=="mar"

md<-cbind(Montesinhopark_forestfires,temp2)

#barplot visual frequency table

 ?barplot()

table(FFMC)

count<-table(FFMC)

count

#for percentages divide the number with the total number of columns

 barplot(count)

 count<-table(FFMC)

  count

barplot(count)

percent<-table(FFMC)/1021

barplot(FFMC)

barplot(FFMC,xlab="FFMC",ylab = "%",las=1)#las means orientation of the tick mark variables

boxplot(FFMC)

boxplot(temp)

boxplot(ISI)

boxplot(DC)

boxplot(DMC)

quantile(FFMC,probs=c(0,0.25,0.50,0.75,1))

quantile(ISI,probs=c(0,0.25,0.50,0.75,1))

boxplot(ISI,main="BOXPLOT",ylab="ISI",las=1)

boxplot(FFMC~ISI)

boxplot(DC~ISI)

boxplot(DMC~ISI)

boxplot(ISI[area=="0"],ISI[rain=="0"])
```

```
boxplot(ISI~DMC,main="Litter vs Forest burnt",ylab="Expected forest burn(ISI)",las=1)

stem(ISI)

#examine the relationship between ISI & DMC/FFMC/DC

#produce a contingency table

table(ISI,FFMC)

barplot(table(ISI))

#relationships between ISI and FFMC

barplot(table(ISI,FFMC))

mosaic(table(ISI,FFMC))#not much of help

summary(ISI)

#scatterplot to examine the linear relationship

plot(ISI)

plot(ISI~DMC)

plot(ISI,FFMC,main="Scatterplot",ylab="FFMC",las=1)

#correlation

cor(ISI,FFMC)

cor(ISI,DC)

cor(ISI,DMC)

abline(ISI~DMC)

lines(smooth.spline(ISI,FFMC))

min(ISI)

max(ISI)

plot(ISI,FFMC, main="Scatterplot", cex=0.50)

plot(ISI,FFMC, main="Scatterplot", cex=0.50,cex.main=2,cex.lab=1.5)

plot(ISI,FFMC, main="Scatterplot", font.main=3)#italic

plot(ISI,FFMC, main="Scatterplot", font.main=4)#bold

plot(ISI,FFMC, main="Scatterplot", font.lab=2)

plot(ISI,FFMC, main="Scatterplot", font.lab=2,font.axis=3)
```

```
plot(ISI,FFMC, main="Scatterplot", pch=2)#triangular notations for the values

abline(lm(FFMC~ISI))#predict FFMC using ISI

#t-test analysis

t.test(ISI,mu=8.98,alternaties="less",level=0.95)

t.test(ISI,FFMC)

t.test(ISI,DMC)

var(ISI)

var(FFMC)

TAB=table(ISI,FFMC)

barplot(TAB,beside=T,legend=T)

chisq.test(TAB)

mod1<-lm(ISI~FFMC)

summary(mod1)

plot(mod1)#see all the plots you get and analyze them correctly

abline(mod1)

attributes(mod1)

coefficients(mod1)

confint(mod1)

anova(mod1)

cor(ISI,FFMC,method="pearson")

plot(cor(ISI,FFMC,method="pearson"),ylab="cor value")

#multiple linear regression in R

mod2<-lm(ISI~FFMC+DMC)

summary(mod2)

mod3<-lm(ISI~FFMC+DMC+DC)

summary(mod3)

mod4<-lm(ISI~rain+wind+RH+temp)

summary(mod4)
```

```
plot(mod3)

plot(mod4)

#interaction or effect modification

mod6<-lm(ISI~FFMC*DMC)

#how are the forest fires are distributed in the park

fit=glm(log(area+1)~FFMC+DMC+DC+ISI+temp+RH+wind+rain)

plot(fit)

summary(fit)

#what suggestions can we give on the basis of the analyzation of the relationships

plot(RH)

barplot(RH)

plot(RH,ISI)

barplot(ISI,RH)

#thus people should avoid visiting the park when relative humidity is lower

#Analyze the fanning effect

#1.work on the intervals

#draw the histograms

plot(wind,ISI)

#yes wind helps the spread of fire

plot(rain,ISI)

#when the rain is zero the fire has maximum spreading trends

plot(temp,ISI)

#when the temperature is maximum[10-30 degree celsius] the fire is maximum

plot(ISI,area)

#trends are observed with the initial spread of fire are out of nowhere

#perform the T-test in R

x1<-rnorm(ISI)

x2<-rnorm(FFMC)
```

t.test(x1,x2)

#used to determine whether the means of two groups are equal to each other. The assumption for the test is that both groups are sampled from normal distributions with equal variances. The null hypothesis is that the two means are equal, and the alternative is that they are not

plot(mean(x1),mean(x2))

Assignment -1

runif(1000000)

runif(10,min=1,max=10000)

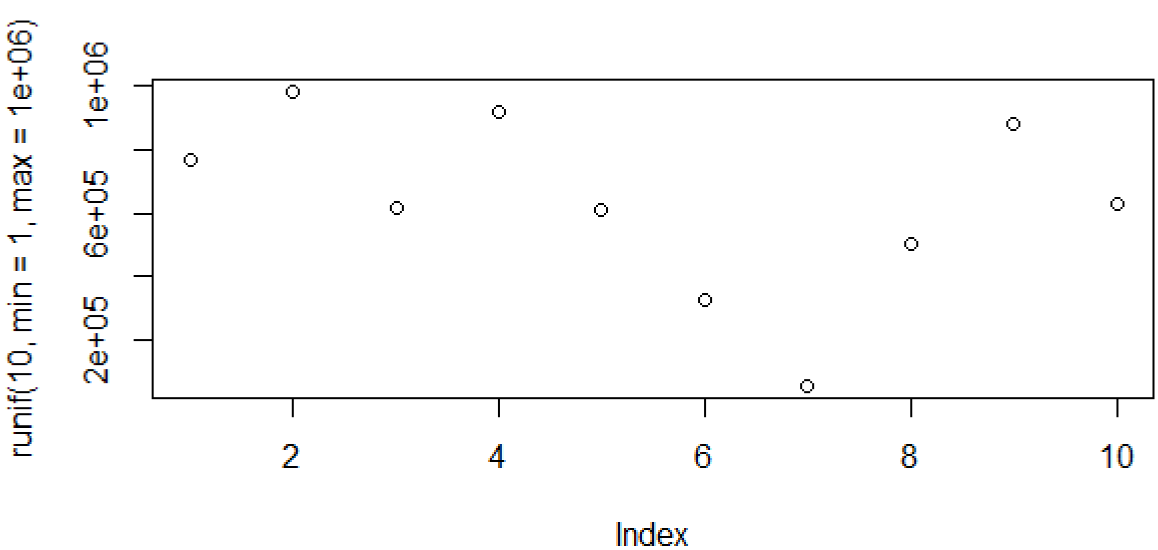
data<-c(seq(1,100000,1))

chisq.test(data)runif(10)

plot(runif(10,min=1,max=1000))

data<-c(seq(1,100000,1))

chisq.test(data)

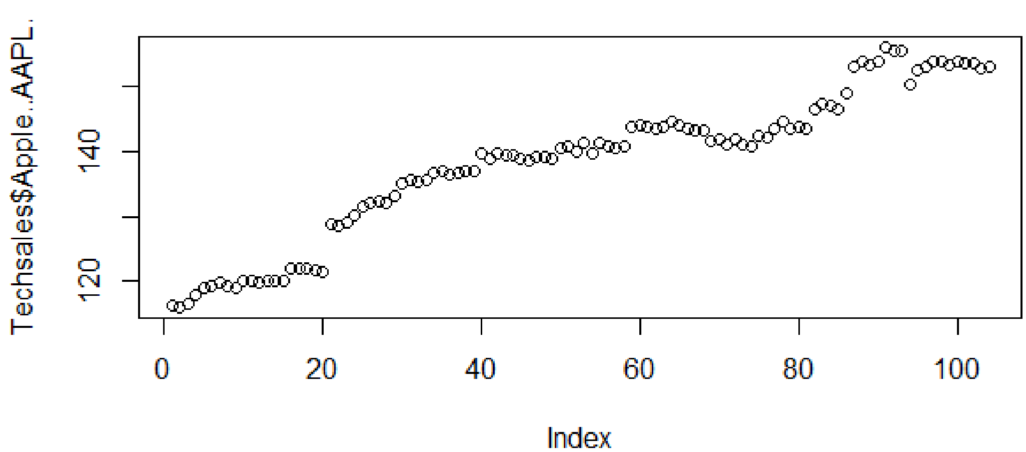


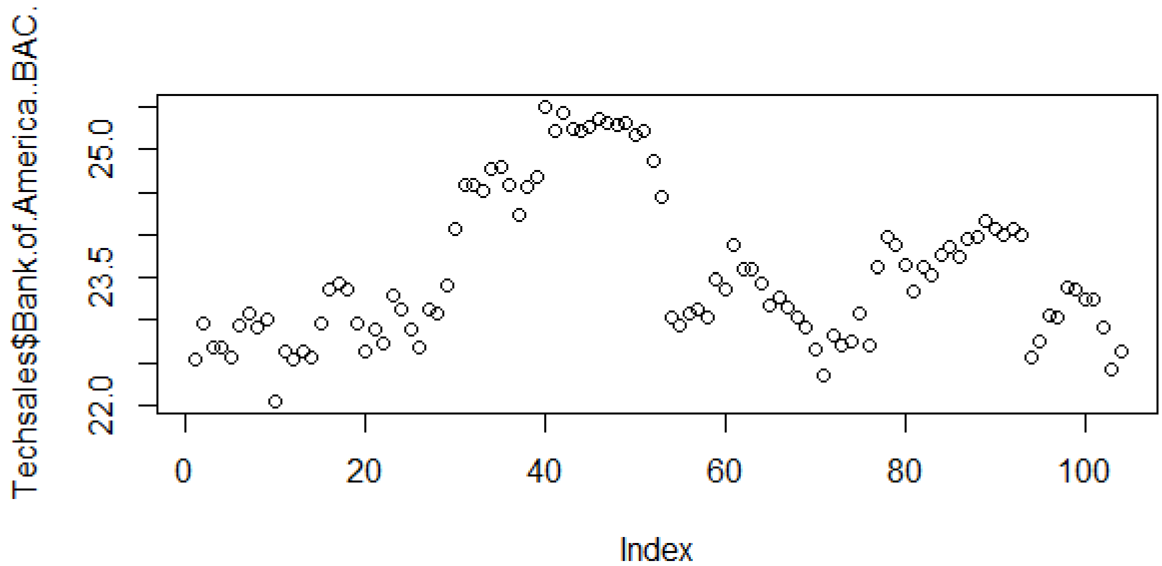
Assignment-2

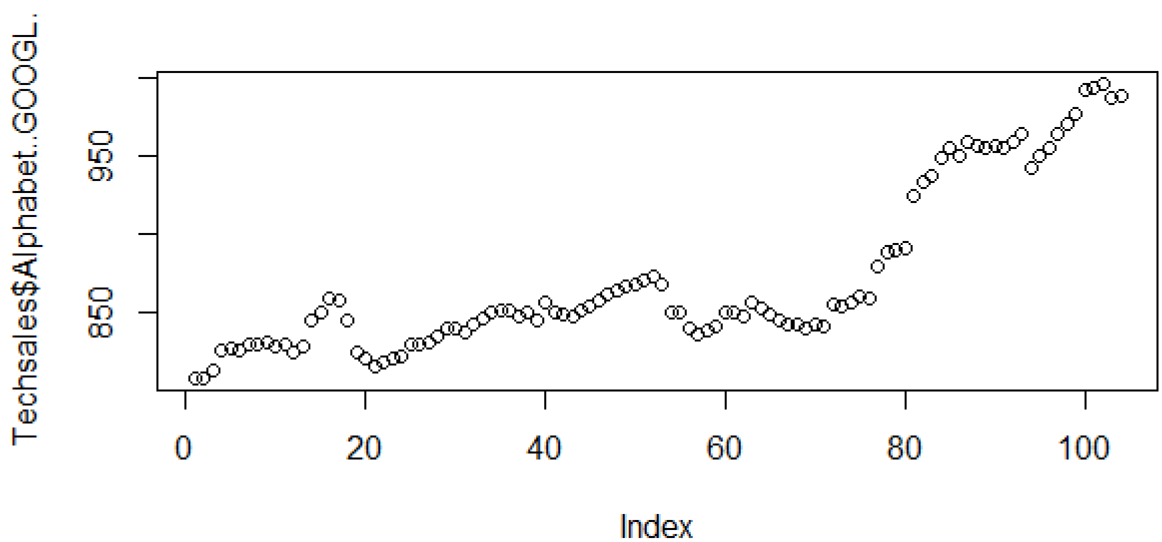
dataset<-c(12327,17129,19923)

```
barplot(dataset)

column_names<-c("New York","los Angeles","San Francisco")

colnames(Table)=column_names

Table<-matrix(cells,nrow = 1, ncol=3)

Table

NY<-c(12237)

LA<-c(17129)

SF<-c(19923)

pie(dataset)

slices<-c(12327,17129,19923)

lbls<-c("New York","los Angeles","San Francisco")

pie(slices,labels = lbls,main = "Ice cream sales")
```

Assignment-3



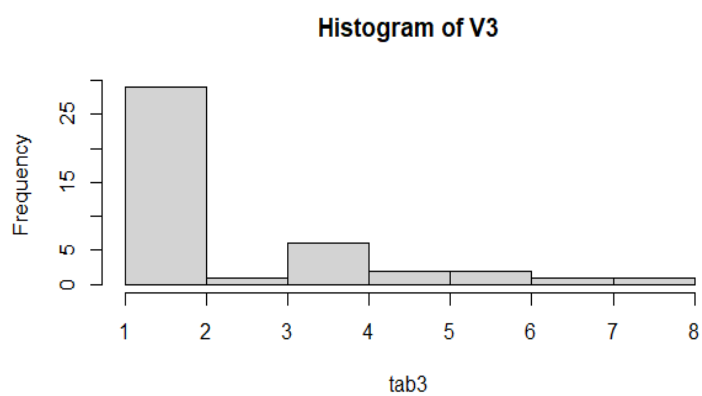


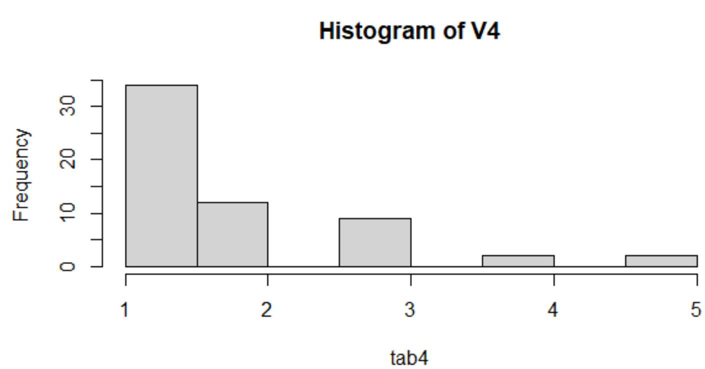


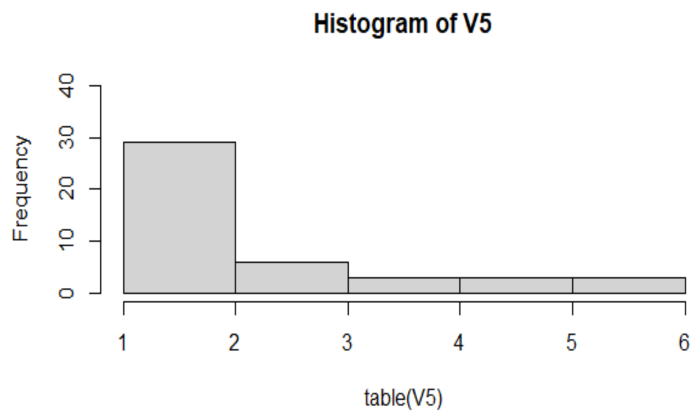
Assignment 4:

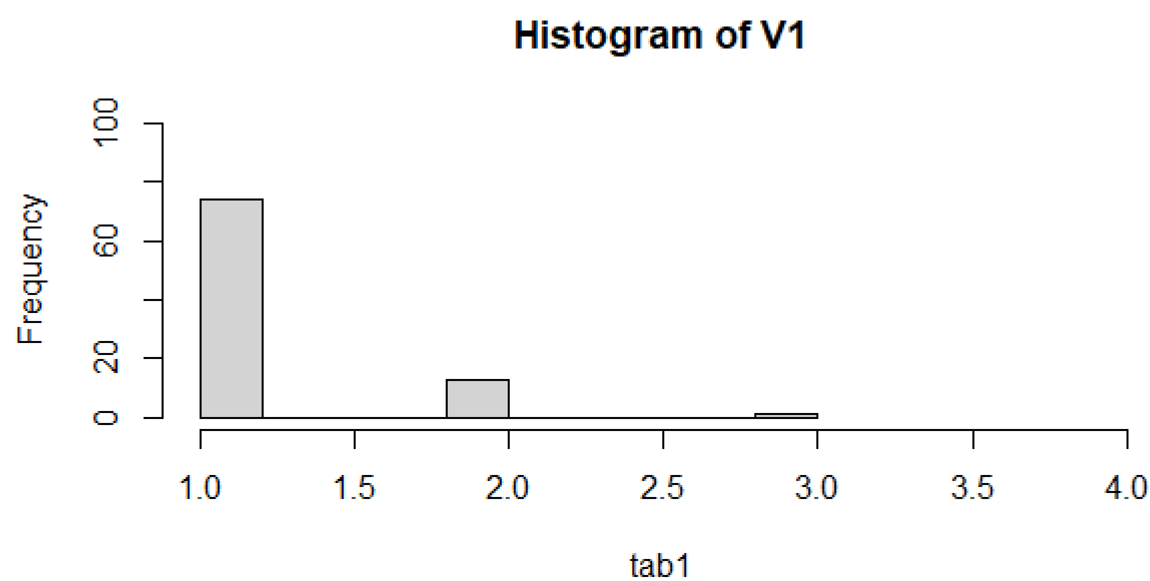
```
tab1<-c(table(V1))

> hist(tab1)

> hist(tab1,xlim=c(1,4),ylim=c(0,100))

> hist(tab1,xlim=c(1,4),ylim=c(0,100),main="Histogram of V1")

> tab2<-c(table(V2))
```

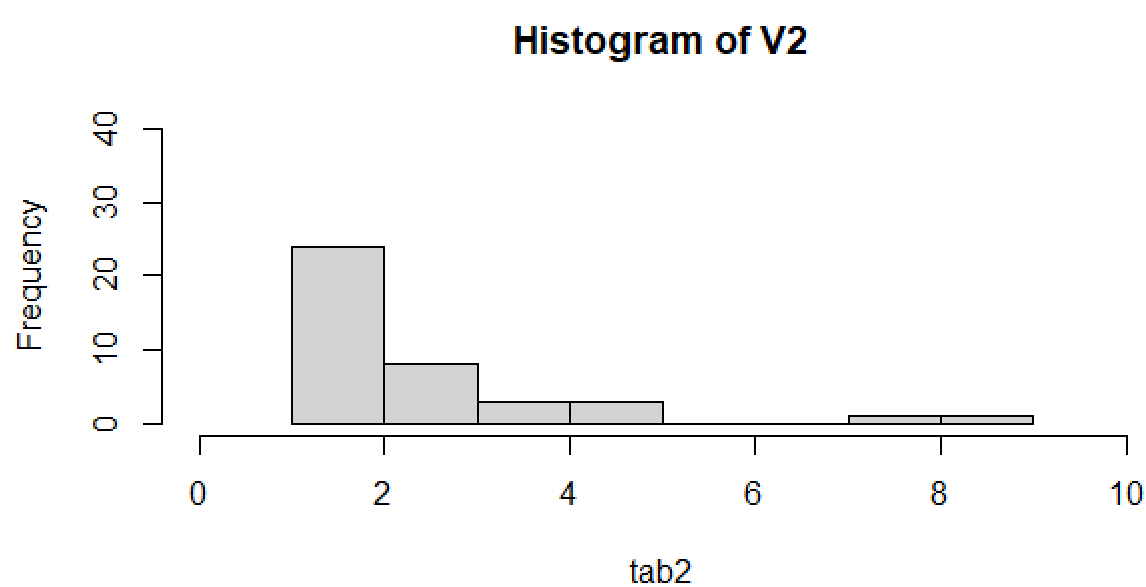
```
> hist(tab2)

> hist(tab2,main="Histogram of V2",xlim=c(0,10),ylim=c(0,40))

> tab3<-c(table(V3))

> hist(tab3)

> hist(tab3,main="Histogram of V3")

> tab4<-c(table(V4))

> hist(tab4)

> hist(tab4,main="Histogram of V4")

> hist(table(V5))

> hist(table(V5),main="Histogram of V5")

> hist(table(V5),main="Histogram of V5",ylim=c(0,40))
```






Histogram of V5
Frequency
0
10
20
30
40
1
2
3
4
5
6
table(V5)


Histogram of V1
Frequency
0
20
60
100
1.0
1.5
2.0
2.5
3.0
3.5
4.0
tab1


Histogram of V2
Frequency
0
10
20
30
40
0
2
4
6
8
10
tab2

30 Sept. Assignment

Q1.

```
x<-rpois(100, 50)

y<-rpois(100, 100)

z<-rpois(100, 150)

#inbuit model

fit=glm(z~x*y)

summary(fit)

Call:

glm(formula = z ~ x * y)


Deviance Residuals:

    Min       1Q    Median      3Q      Max

-22.1584  -7.0242   0.7954   5.6098  25.6292


Coefficients:

            Estimate Std. Error t value Pr(>|t|)

(Intercept) 84.20097   73.97613   1.138    0.258

x            1.63241    1.46442   1.115    0.268

y            0.57985    0.72327   0.802    0.425

x:y         -0.01451    0.01429  -1.015    0.313
```

(Dispersion parameter for gaussian family taken to be 102.7121)

    Null deviance: 10306.2 on 99 degrees of freedom

Residual deviance: 9860.4 on 96 degrees of freedom

AIC: 752.9

Number of Fisher Scoring iterations: 2

```
#normalize and find the values
xmat = cbind(x= x, y=y, z=z)
 ones = rep(1,nrow(xmat))
 xmat = cbind(intercept = ones,xmat)
 beta = solve(t(xmat) %*% xmat)%*% t(xmat) %*% y
 beta
[,1]
intercept -2.578382e-12
x          1.793010e-14
y          1.000000e+00
z          5.447032e-16
 lm(z~x+y)
Call:
```

```
lm(formula = z ~ x + y)


Coefficients:

(Intercept)           x           y

   158.3194      0.1533     -0.1478
```

Q2.

```
x<-rpois(100, 50)

y<-rpois(100, 100)

z<-rpois(100, 150)

#answer to question 1

summary(fit1<-lm(z~x))

summary(fit2<-lm(z~x+y+z))

summary(fit3<-lm(z~x*y))

#answer to part 2

#Model1

xmat = cbind(x= x)

ones = rep(1,nrow(xmat))

xmat = cbind(intercept = ones,xmat)

beta = solve(t(xmat) %*% xmat)%*% t(xmat) %*% x

beta

#model2
```

```
xmat = cbind(x= x,y=y,z=z)

ones = rep(1,nrow(xmat))

xmat = cbind(intercept = ones,xmat)

beta = solve(t(xmat) %*% xmat)%*% t(xmat) %*% y

beta

[,1]

intercept -2.578382e-12

x          1.793010e-14

y          1.000000e+00

z          5.447032e-16

# model 3

xmat = cbind(x= x,y=y)

ones = rep(1,nrow(xmat))

xmat = cbind(intercept = ones,xmat)

beta = solve(t(xmat) %*% xmat)%*% t(xmat) %*% x

beta

[,1]

intercept -3.685940e-14

x          1.000000e+00

y         -1.100509e-14

> summary(fit1)$r.squared
```

```
[1] 0.009520894

> summary(fit2)$r.squared

[1] 0.03298831

> summary(fit3)$r.squared

[1] 0.9916279
```