Forest Fire Project Part 2

Shi Fan Jin

11/7/2018

```
library(MASS)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
## The following object is masked from 'package:MASS':
##
##     select
## The following objects are masked from 'package:stats':
##
##     filter, lag
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
set.seed(1)
```

```
forest = read.csv("forestfires.csv", stringsAsFactors = FALSE)
```

Transform the data

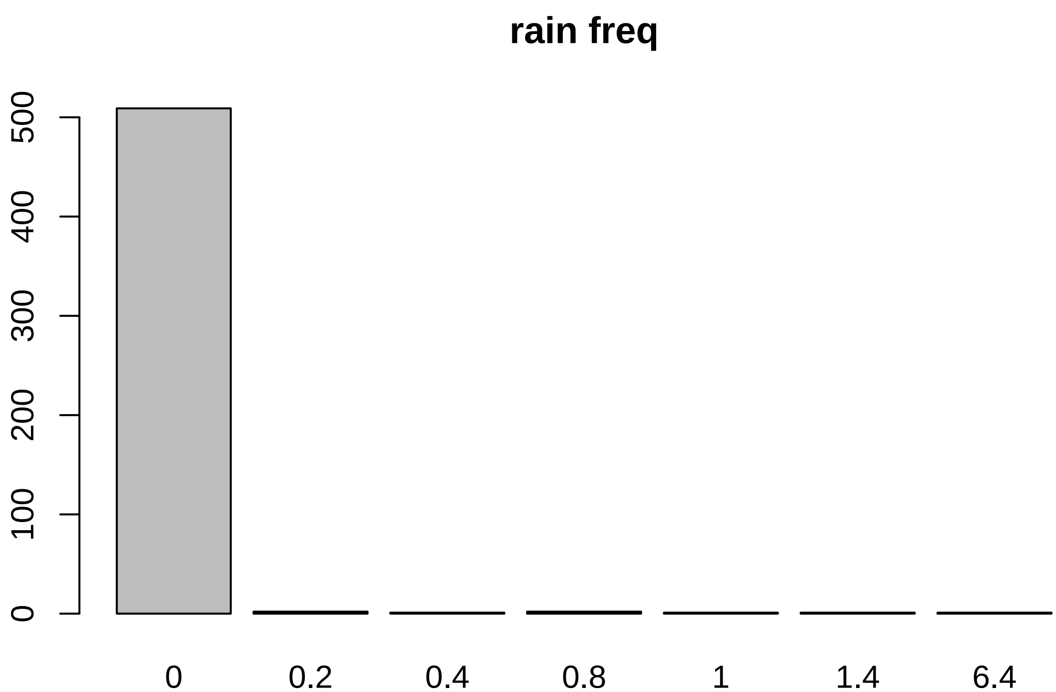
Transform the "area" using log(x + 1), "FFMC" with log(x), and remove rain.

In order to make it easier to see (spread it out)

```
table(forest$rain)
```

```
##
##   0 0.2 0.4 0.8   1 1.4 6.4
## 509   2   1   2   1   1   1
```

```
barplot(table(forest$rain), main = "rain freq")
```




```
## Most of the fire happens when there's no rain
## In this case, Let's remove "rain" for this practice
forest = forest[, -12]

# Transform the data
forest$area = log(forest$area + 1)
forest$FFMC = log(forest$FFMC)
#forest$ISI = log(forest$ISI)
```

Transform the "month" and "day" into numeric values

```
########## CHANGING month and day into numeric values
for (i in 1:517) {

  if(forest$month[i] == "jan") {
    forest$month[i] = 1
  }else if (forest$month[i] == "feb") {
    forest$month[i] = 2
  }else if (forest$month[i] == "mar") {
    forest$month[i] = 3
  }else if (forest$month[i] == "apr") {
    forest$month[i] = 4
  }else if (forest$month[i] == "may") {
    forest$month[i] = 5
  }else if (forest$month[i] == "jun") {
    forest$month[i] = 6
  }else if (forest$month[i] == "jul") {
    forest$month[i] = 7
  }else if (forest$month[i] == "aug") {
    forest$month[i] = 8
  }else if (forest$month[i] == "sep") {
```

```
    forest$month[i] = 9
  }else if (forest$month[i] == "oct") {
    forest$month[i] = 10
  }else if (forest$month[i] == "nov") {
    forest$month[i] = 11
  }else if (forest$month[i] == "dec") {
    forest$month[i] = 12
  }

  if (forest$day[i] == "mon") {
    forest$day[i] = 1
  } else if (forest$day[i] == "tue") {
    forest$day[i] = 2
  } else if (forest$day[i] == "wed") {
    forest$day[i] = 3
  } else if (forest$day[i] == "thu") {
    forest$day[i] = 4
  } else if (forest$day[i] == "fri") {
    forest$day[i] = 5
  } else if (forest$day[i] == "sat") {
    forest$day[i] = 6
  } else if (forest$day[i] == "sun") {
    forest$day[i] = 7
  }
}

class(forest$month[157])
```

```
## [1] "character"
```

```
class(forest$day[157])
```

```
## [1] "character"
```

```
forest$month = as.numeric(forest$month)
forest$day = as.numeric(forest$day)

cor(forest)
```

```
##                  X            Y       month           day         FFMC
## X      1.000000000  0.539548171 -0.06500303 -0.0249218945 -0.006798943
## Y      0.539548171  1.000000000 -0.06629179 -0.0054533368 -0.043383489
## month -0.065003032 -0.066291786  1.00000000 -0.0508365920  0.250941204
## day   -0.024921895 -0.005453337 -0.05083659  1.0000000000 -0.055045824
## FFMC  -0.006798943 -0.043383489  0.25094120 -0.0550458235  1.000000000
## DMC   -0.048384178  0.007781561  0.46664525  0.0628703973  0.301026670
## DC    -0.085916123 -0.101177767  0.86869776  0.0001049027  0.257457541
## ISI    0.006209941 -0.024487992  0.18659697  0.0329092595  0.414650759
## temp  -0.051258262 -0.024103084  0.36884151  0.0521903410  0.342932405
## RH     0.085223194  0.062220731 -0.09528038  0.0921514374 -0.277653414
## wind   0.018797818 -0.020340852 -0.08636797  0.0324781638  0.005226916
## area   0.061994908  0.038838213  0.11428008  0.0002081962  0.049879323
##                DMC            DC          ISI        temp          RH
## X     -0.048384178 -0.0859161229  0.006209941 -0.05125826  0.08522319
## Y      0.007781561 -0.1011777674 -0.024487992 -0.02410308  0.06222073
```

```
## month  0.466645252  0.8686977586  0.186596974  0.36884151 -0.09528038
## day    0.062870397  0.0001049027  0.032909260  0.05219034  0.09215144
## FFMC   0.301026670  0.2574575405  0.414650759  0.34293241 -0.27765341
## DMC    1.000000000  0.6821916120  0.305127835  0.46959384  0.07379494
## DC     0.682191612  1.0000000000  0.229154169  0.49620805 -0.03919165
## ISI    0.305127835  0.2291541691  1.000000000  0.39428710 -0.13251718
## temp   0.469593844  0.4962080531  0.394287104  1.00000000 -0.52739034
## RH     0.073794941 -0.0391916472 -0.132517177 -0.52739034  1.00000000
## wind  -0.105342253 -0.2034656909  0.106825888 -0.22711622  0.06941007
## area   0.067152740  0.0663597560 -0.010346879  0.05348655 -0.05366216
##               wind          area
## X      0.018797818  0.0619949083
## Y     -0.020340852  0.0388382135
## month -0.086367965  0.1142800820
## day    0.032478164  0.0002081962
## FFMC   0.005226916  0.0498793226
## DMC   -0.105342253  0.0671527398
## DC    -0.203465691  0.0663597560
## ISI    0.106825888 -0.0103468787
## temp  -0.227116220  0.0534865490
## RH     0.069410067 -0.0536621583
## wind   1.000000000  0.0669734893
## area   0.066973489  1.0000000000
```

```
# FOR PREDICTING "AREA"
#'temp' has the highest correlation with the area of forest fire(which is a positive correlation), foll
```

Topic 1. Model Selection

Split data into 2/3 as training set and 1/3 as the test set

```
train_size = floor((length(forest$X)/3) * 2)
train = sample(517, train_size)
test_size = length(forest$X) - train

trainset = forest[train, ]
testset = forest[-train, ]
```

Forward and Backward Selection (NEW)

```
glm.fits_new1 = glm(area~ X + month + day + FFMC + DMC + DC + ISI + temp + RH + wind, data = trainset)
```

Forward and Backward Selection (OLD)

```
# Let's make a model
#lm.fit = lm(area~ X + Y + month + day + FFMC + DMC + DC + ISI + temp + RH + wind, data = forest)
#glm.fits = glm(area~ ., data = trainset)
glm.fits = glm(area~ X + Y + month + day + FFMC + DMC + DC + ISI + temp + RH + wind, data = trainset)
# Let's try forward and backward selection to figure out which variable have more significant effect on
```

```r
##################Backward
backwards=step(glm.fits)
```

```
## Start:  AIC=1222.26
## area ~ X + Y + month + day + FFMC + DMC + DC + ISI + temp + RH +
##     wind
##
##         Df Deviance    AIC
## - FFMC   1   652.18 1220.3
## - RH     1   652.22 1220.3
## - Y      1   652.39 1220.4
## - day    1   653.11 1220.8
## - X      1   654.13 1221.3
## - temp   1   654.44 1221.5
## - wind   1   654.88 1221.7
## <none>       652.14 1222.3
## - ISI    1   657.08 1222.8
## - DMC    1   658.85 1223.8
## - DC     1   658.88 1223.8
## - month  1   664.53 1226.7
##
## Step:  AIC=1220.28
## area ~ X + Y + month + day + DMC + DC + ISI + temp + RH + wind
##
##         Df Deviance    AIC
## - RH     1   652.24 1218.3
## - Y      1   652.44 1218.4
## - day    1   653.16 1218.8
## - X      1   654.17 1219.3
## - temp   1   654.51 1219.5
## - wind   1   654.93 1219.7
## <none>       652.18 1220.3
## - ISI    1   658.69 1221.7
## - DMC    1   658.87 1221.8
## - DC     1   659.17 1222.0
## - month  1   664.74 1224.8
##
## Step:  AIC=1218.31
## area ~ X + Y + month + day + DMC + DC + ISI + temp + wind
##
##         Df Deviance    AIC
## - Y      1   652.51 1216.5
## - day    1   653.30 1216.9
## - X      1   654.19 1217.3
## - wind   1   655.01 1217.8
## <none>       652.24 1218.3
## - temp   1   657.25 1218.9
## - ISI    1   658.73 1219.7
## - DMC    1   659.16 1219.9
## - DC     1   659.52 1220.1
## - month  1   665.12 1223.0
##
## Step:  AIC=1216.46
```

```
## area ~ X + month + day + DMC + DC + ISI + temp + wind
##
##         Df Deviance    AIC
## - day    1   653.55 1215.0
## - wind   1   655.18 1215.9
## <none>       652.51 1216.5
## - X      1   656.34 1216.5
## - temp   1   657.80 1217.2
## - ISI    1   659.17 1217.9
## - DMC    1   659.94 1218.3
## - DC     1   660.82 1218.8
## - month  1   666.46 1221.7
##
## Step:  AIC=1215
## area ~ X + month + DMC + DC + ISI + temp + wind
##
##         Df Deviance    AIC
## - wind   1   656.06 1214.3
## <none>       653.55 1215.0
## - X      1   657.57 1215.1
## - temp   1   658.68 1215.7
## - ISI    1   660.22 1216.5
## - DMC    1   660.48 1216.6
## - DC     1   661.60 1217.2
## - month  1   667.37 1220.2
##
## Step:  AIC=1214.32
## area ~ X + month + DMC + DC + ISI + temp
##
##         Df Deviance    AIC
## <none>       656.06 1214.3
## - temp   1   659.93 1214.3
## - X      1   659.94 1214.3
## - ISI    1   661.38 1215.1
## - DMC    1   664.15 1216.5
## - DC     1   666.20 1217.6
## - month  1   671.84 1220.5
```

```r
# OUTPUT: area ~ X + month + DMC + DC + ISI + temp
# with the lowest AIC = 1214.32
summary(backwards)
```

```
##
## Call:
## glm(formula = area ~ X + month + DMC + DC + ISI + temp, data = trainset)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7036  -1.1277  -0.5112   0.8347   5.5171
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.2221133  0.4160590  -0.534  0.59380
## X            0.0468402  0.0331616   1.412  0.15873
## month        0.2232205  0.0783951   2.847  0.00468 **
```

```
## DMC          0.0036225  0.0017776   2.038  0.04234 *
## DC          -0.0019889  0.0008713  -2.283  0.02307 *
## ISI         -0.0292818  0.0177113  -1.653  0.09920 .
## temp         0.0225111  0.0159612   1.410  0.15936
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 1.946768)
##
##     Null deviance: 688.00  on 343  degrees of freedom
## Residual deviance: 656.06  on 337  degrees of freedom
## AIC: 1214.3
##
## Number of Fisher Scoring iterations: 2
```

```
###################Forward
nothing = glm(area~ 1, data = trainset)
forwards = step(nothing,
                scope = list(lower = formula(nothing),
                             upper = formula(glm.fits)), direction = "forward")
```

```
## Start:  AIC=1218.67
## area ~ 1
##
##         Df Deviance    AIC
## + month  1   675.76 1214.5
## + DMC    1   680.66 1217.0
## + DC     1   682.10 1217.7
## + temp   1   683.84 1218.6
## <none>       688.00 1218.7
## + Y      1   684.80 1219.1
## + X      1   685.29 1219.3
## + wind   1   686.87 1220.1
## + RH     1   687.17 1220.3
## + ISI    1   687.48 1220.4
## + day    1   687.51 1220.4
## + FFMC   1   687.84 1220.6
##
## Step:  AIC=1214.5
## area ~ month
##
##        Df Deviance    AIC
## + Y     1   671.73 1214.4
## <none>      675.76 1214.5
## + X     1   672.10 1214.6
## + DC    1   673.61 1215.4
## + wind  1   673.64 1215.4
## + ISI   1   673.86 1215.5
## + DMC   1   674.44 1215.8
## + temp  1   675.00 1216.1
## + day   1   675.21 1216.2
## + RH    1   675.22 1216.2
## + FFMC  1   675.38 1216.3
##
## Step:  AIC=1214.44
```

```
## area ~ month + Y
##
##        Df Deviance    AIC
## <none>      671.73 1214.4
## + wind  1   669.38 1215.2
## + ISI   1   669.92 1215.5
## + DC    1   670.32 1215.7
## + DMC   1   670.35 1215.7
## + X     1   670.67 1215.9
## + temp  1   671.04 1216.1
## + day   1   671.17 1216.2
## + RH    1   671.33 1216.2
## + FFMC  1   671.46 1216.3
```

```
# OUTPUT: area ~ month + Y
# with the lowest AIC = 1214.44
summary(forwards)
```

```
##
## Call:
## glm(formula = area ~ month + Y, data = trainset)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.4255  -1.1570  -0.5776   0.8698   5.6599
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.10798    0.38574   0.280   0.7797
## month        0.08653    0.03359   2.576   0.0104 *
## Y            0.08980    0.06273   1.431   0.1532
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 1.969874)
##
##     Null deviance: 688.00  on 343  degrees of freedom
## Residual deviance: 671.73  on 341  degrees of freedom
## AIC: 1214.4
##
## Number of Fisher Scoring iterations: 2
```

```
###################Use CV to see which model is better
###################Applied the trained model (both backward and forward) to the test-data to compare th
pred_back = predict(backwards, testset)
mean((pred_back - testset$area)^2)
```

```
## [1] 1.886026
```

```
pred_forward = predict(forwards, testset)
mean((pred_forward - testset$area)^2)
```

```
## [1] 1.872752
```

```
###Looks like forward selection model is better, since the MSE is lower (1.872752 < 1.886026)
```

```r
####################################TRANSFORMING >>> "fire"
# Since the 13th column indicates the burned area, if there isn't fire, the area = 0. In this case, we
fire = rep(0, length(forest$X))
forest = cbind(forest, fire)
for (i in 1:length(forest$area)) {
  if (forest$area[i] > 0) {
    forest$fire[i] = 1
  }
}
# In this case, if fire = 0, there is no fire, if fire = 1, there is a fire.


####################################FINDING MODEL
# Let's make a model (excluding month, day, since they are not numeric):
glm.fits = glm(fire~ X + Y + FFMC + DMC + DC + ISI + temp + RH + wind + area, data = forest)
# Let's try forward and backward selection to figure out which variable have more significant effect on
##################Backward
backwards=step(glm.fits)
```

```
## Start:  AIC=317.46
## fire ~ X + Y + FFMC + DMC + DC + ISI + temp + RH + wind + area
##
##         Df Deviance    AIC
## - X      1   53.401 315.48
## - wind   1   53.428 315.75
## - ISI    1   53.461 316.07
## - FFMC   1   53.479 316.24
## - temp   1   53.492 316.36
## - Y      1   53.530 316.73
## - RH     1   53.540 316.83
## <none>       53.399 317.46
## - DC     1   53.717 318.53
## - DMC    1   53.727 318.62
## - area   1  125.573 757.55
##
## Step:  AIC=315.48
## fire ~ Y + FFMC + DMC + DC + ISI + temp + RH + wind + area
##
##         Df Deviance    AIC
## - wind   1   53.430 313.76
## - ISI    1   53.463 314.08
## - FFMC   1   53.480 314.25
## - temp   1   53.493 314.37
## - RH     1   53.540 314.83
## - Y      1   53.562 315.04
## <none>       53.401 315.48
## - DC     1   53.718 316.55
## - DMC    1   53.727 316.62
## - area   1  125.766 756.34
##
## Step:  AIC=313.76
## fire ~ Y + FFMC + DMC + DC + ISI + temp + RH + area
##
##         Df Deviance    AIC
```

```
## - temp  1   53.505 312.49
## - FFMC  1   53.510 312.54
## - ISI   1   53.515 312.58
## - RH    1   53.563 313.05
## - Y     1   53.586 313.27
## <none>      53.430 313.76
## - DC    1   53.729 314.64
## - DMC   1   53.745 314.80
## - area  1  126.723 758.26
##
## Step:  AIC=312.49
## fire ~ Y + FFMC + DMC + DC + ISI + RH + area
##
##        Df Deviance    AIC
## - RH    1   53.567 311.08
## - FFMC  1   53.579 311.20
## - ISI   1   53.646 311.85
## - Y     1   53.673 312.11
## <none>      53.505 312.49
## - DMC   1   53.757 312.91
## - DC    1   53.918 314.46
## - area  1  126.749 756.37
##
## Step:  AIC=311.08
## fire ~ Y + FFMC + DMC + DC + ISI + area
##
##        Df Deviance    AIC
## - FFMC  1   53.612 309.52
## - ISI   1   53.698 310.35
## - Y     1   53.743 310.78
## <none>      53.567 311.08
## - DMC   1   53.779 311.13
## - DC    1   53.953 312.80
## - area  1  126.794 754.55
##
## Step:  AIC=309.52
## fire ~ Y + DMC + DC + ISI + area
##
##        Df Deviance    AIC
## - Y     1   53.782 309.16
## - DMC   1   53.806 309.39
## <none>      53.612 309.52
## - ISI   1   53.831 309.63
## - DC    1   54.016 311.40
## - area  1  127.154 754.02
##
## Step:  AIC=309.16
## fire ~ DMC + DC + ISI + area
##
##        Df Deviance    AIC
## - DMC   1   53.941 308.68
## <none>      53.782 309.16
## - ISI   1   53.993 309.18
## - DC    1   54.120 310.40
```

```
## - area  1  127.755 754.45
##
## Step:  AIC=308.68
## fire ~ DC + ISI + area
##
##         Df Deviance    AIC
## - ISI   1   54.088 308.09
## - DC    1   54.121 308.40
## <none>      53.941 308.68
## - area  1  127.763 752.48
##
## Step:  AIC=308.09
## fire ~ DC + area
##
##         Df Deviance    AIC
## <none>      54.088 308.09
## - DC    1   54.365 308.73
## - area  1  127.787 750.58
```

```
# OUTPUT: fire ~ Y + DC + wind + area
# with the lowest AIC = 732.78
###################Forward
nothing = glm(fire~ 1, data = forest)
forwards = step(nothing,
                scope = list(lower = formula(nothing),
                             upper = formula(glm.fits)), direction = "forward")
```

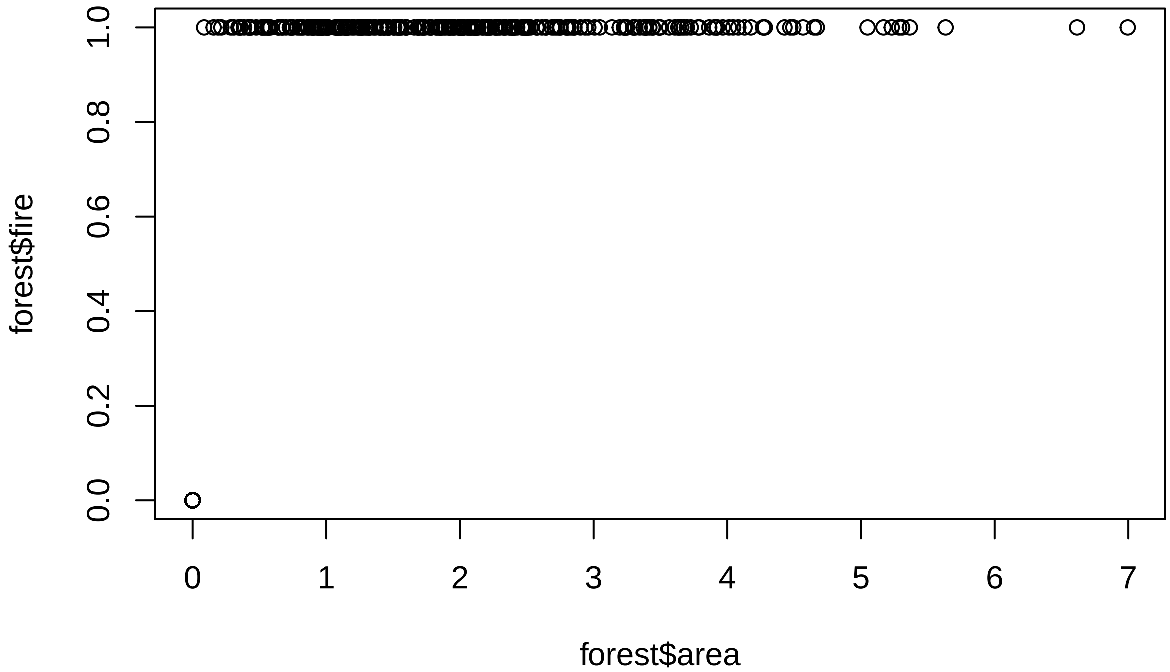
```
## Start:  AIC=753.44
## fire ~ 1
##
##         Df Deviance    AIC
## + area  1   54.365 308.73
## + DC    1  127.787 750.58
## + temp  1  128.248 752.45
## + FFMC  1  128.282 752.58
## + DMC   1  128.488 753.41
## + X     1  128.490 753.42
## <none>     128.994 753.44
## + Y     1  128.577 753.77
## + wind  1  128.594 753.84
## + ISI   1  128.830 754.79
## + RH    1  128.831 754.79
##
## Step:  AIC=308.73
## fire ~ area
##
##         Df Deviance    AIC
## + DC    1   54.088 308.09
## + ISI   1   54.121 308.40
## <none>      54.365 308.73
## + FFMC  1   54.194 309.11
## + temp  1   54.203 309.19
## + Y     1   54.269 309.81
## + X     1   54.335 310.44
## + DMC   1   54.348 310.57
```

```
## + RH     1   54.362 310.70
## + wind   1   54.362 310.71
##
## Step:  AIC=308.09
## fire ~ area + DC
##
##        Df Deviance    AIC
## <none>      54.088 308.09
## + ISI   1   53.941 308.68
## + Y     1   53.953 308.80
## + DMC   1   53.993 309.18
## + FFMC  1   54.005 309.29
## + X     1   54.038 309.62
## + wind  1   54.060 309.82
## + temp  1   54.061 309.84
## + RH    1   54.082 310.03
```

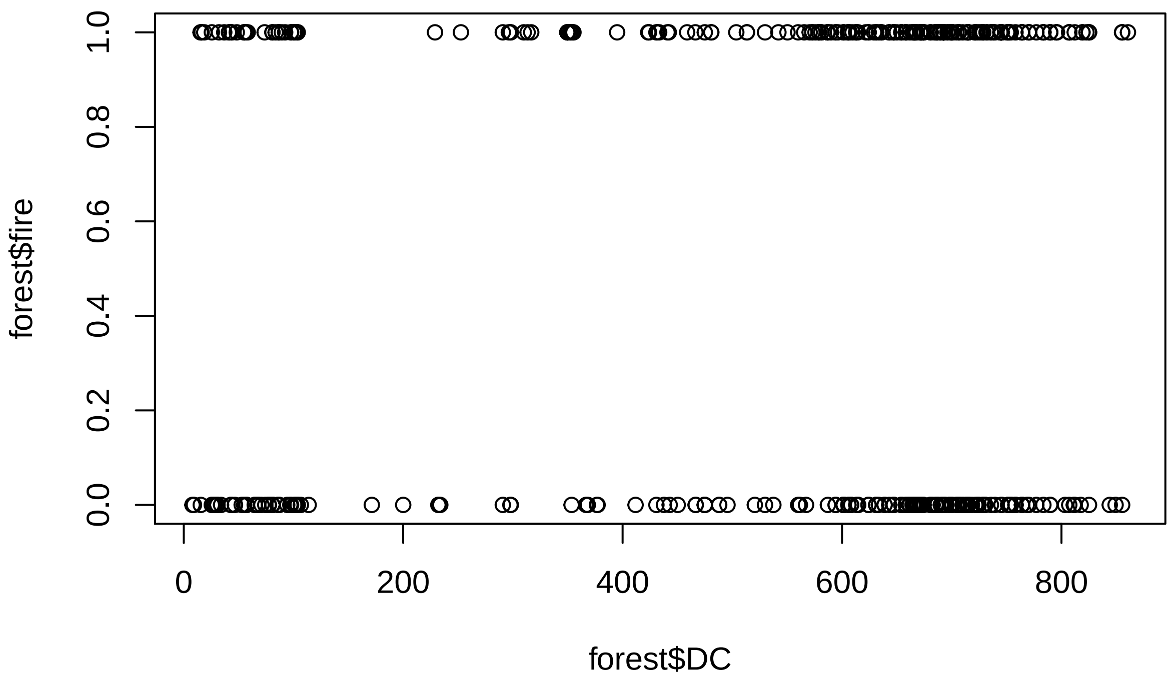
```r
# OUTPUT: fire ~ area + DC + wind + Y
# with the lowest AIC = 732.78
######### lOOKS LIKE BOTH BACKWARD AND FORWARD SUGGESTS THE SAME MODEL: fire ~ Y + DC + wind + area
glm.selected = glm(fire ~ area + DC + wind + Y, data = forest)
summary(glm.selected)
```

```
##
## Call:
## glm(formula = fire ~ area + DC + wind + Y, data = forest)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.1330  -0.2307  -0.1767   0.2568   0.7355
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 8.649e-02  7.609e-02   1.137   0.2562
## area        2.698e-01  1.029e-02  26.232   <2e-16 ***
## DC          1.077e-04  5.939e-05   1.813   0.0704 .
## wind        4.662e-03  8.180e-03   0.570   0.5690
## Y           1.355e-02  1.170e-02   1.158   0.2474
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.10531)
##
##     Null deviance: 128.994  on 516  degrees of freedom
## Residual deviance:  53.919  on 512  degrees of freedom
## AIC: 310.47
##
## Number of Fisher Scoring iterations: 2
```

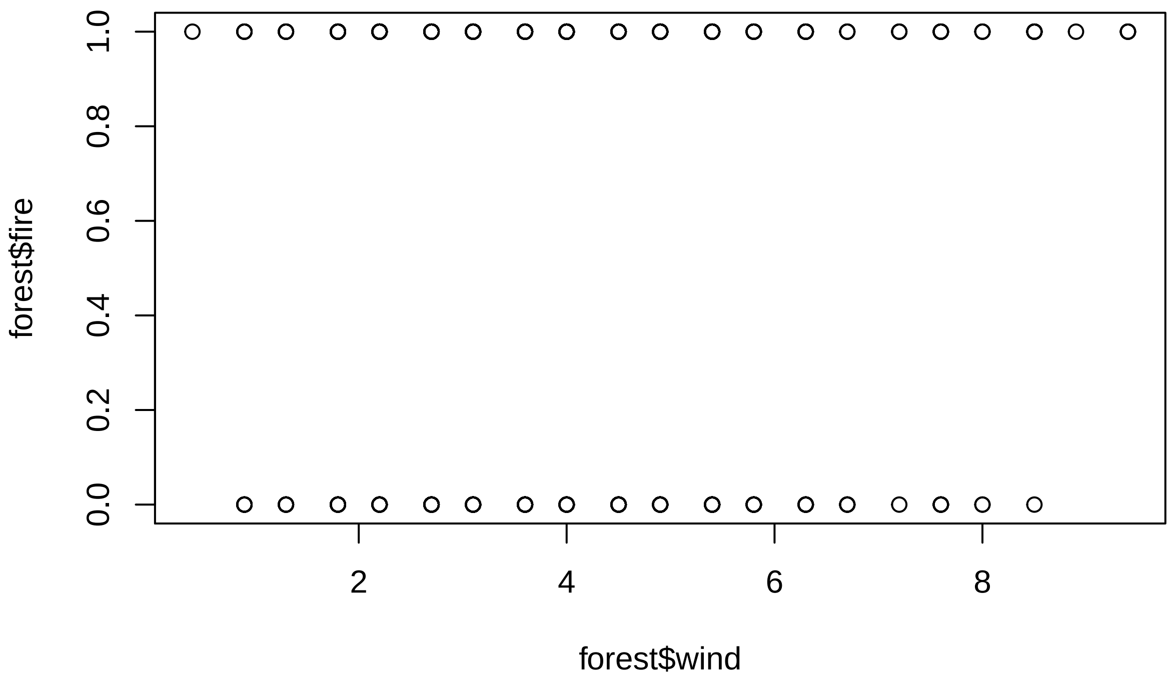
```r
plot(forest$area, forest$fire)
```



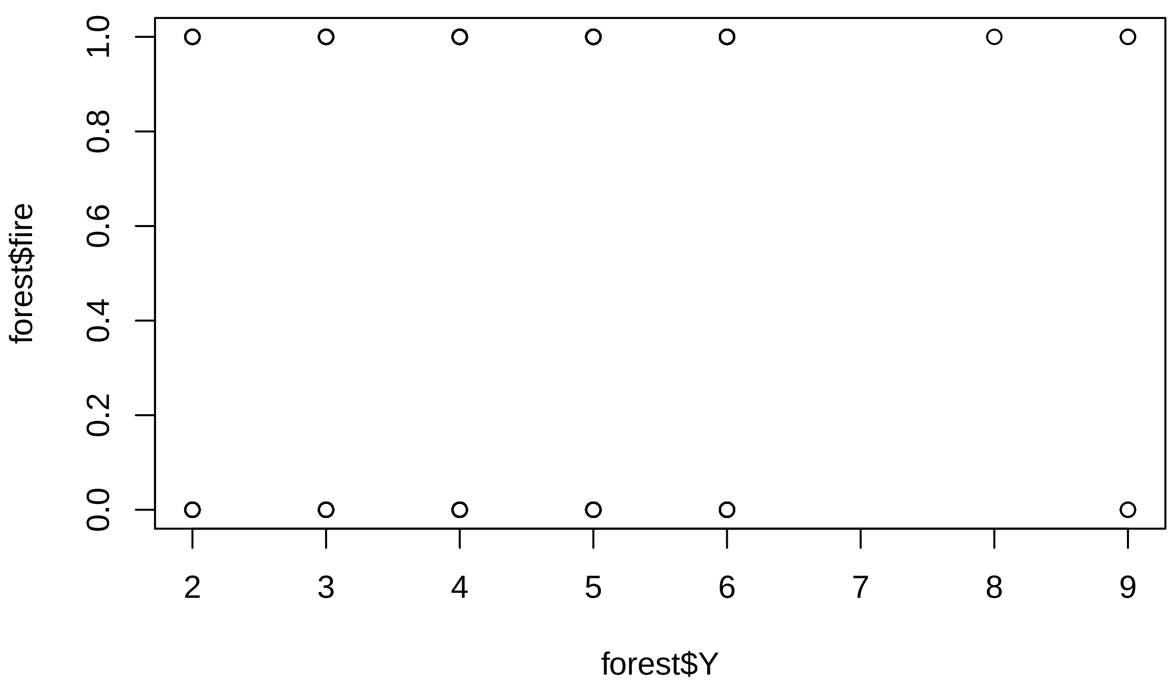

```
plot(forest$DC, forest$fire)
```




```
plot(forest$wind, forest$fire)
```




```r
plot(forest$Y, forest$fire)
```




```
#################USE linear discriminant analysis
#######################USE SVM???
##########################################USE TREE TO SEE HOW TO CLASSIFY AS SMALL FIRE, MEDIUM FIRE, BIG
```

```r
#randomly divide the dataset to make training set and testing set to see if our model is good or not
train = sample(length(forest$X), size = 450)
trainset = forest[train,]
testset = forest[-train,]
dim(trainset)
```

```
## [1] 450  13
```

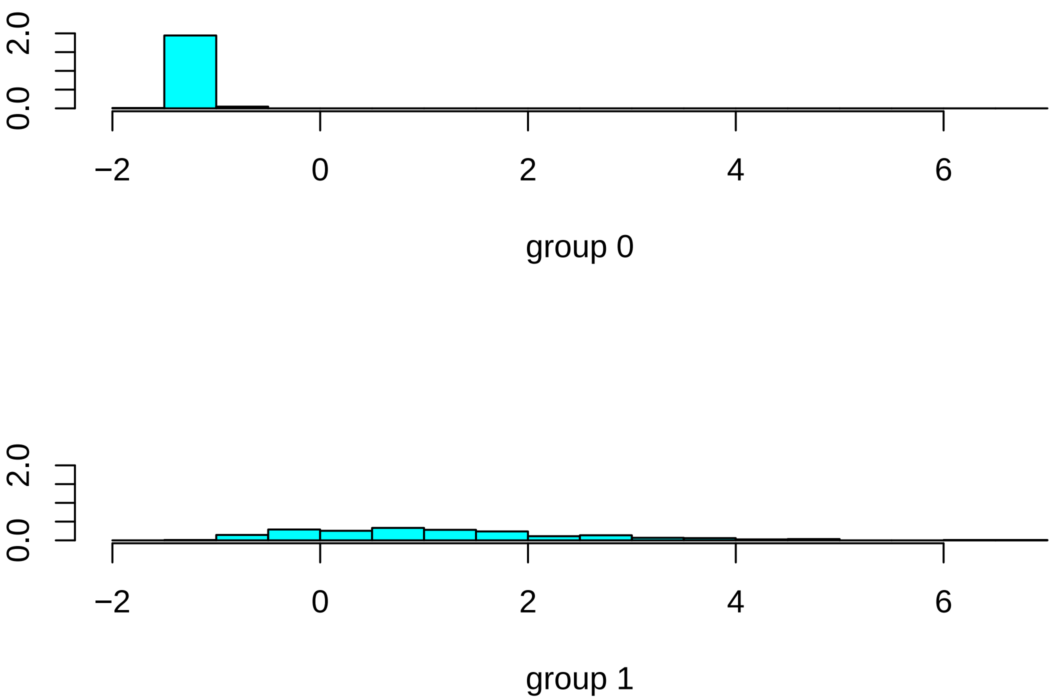
```r
dim(testset)
```

```
## [1] 67 13
```

```r
lda.fit = lda(fire~ area + DC + wind + Y, data = trainset)
lda.fit
```

```
## Call:
## lda(fire ~ area + DC + wind + Y, data = trainset)
##
## Prior probabilities of groups:
##         0         1
## 0.4777778 0.5222222
##
## Group means:
##      area       DC     wind        Y
## 0 0.00000 532.6791 3.873488 4.246512
## 1 2.13005 569.8494 4.107234 4.382979
##
## Coefficients of linear discriminants:
##               LD1
## area 1.0891759902
## DC   0.0002565042
## wind 0.0240864261
## Y    0.0620459278
```

```r
plot(lda.fit)
```



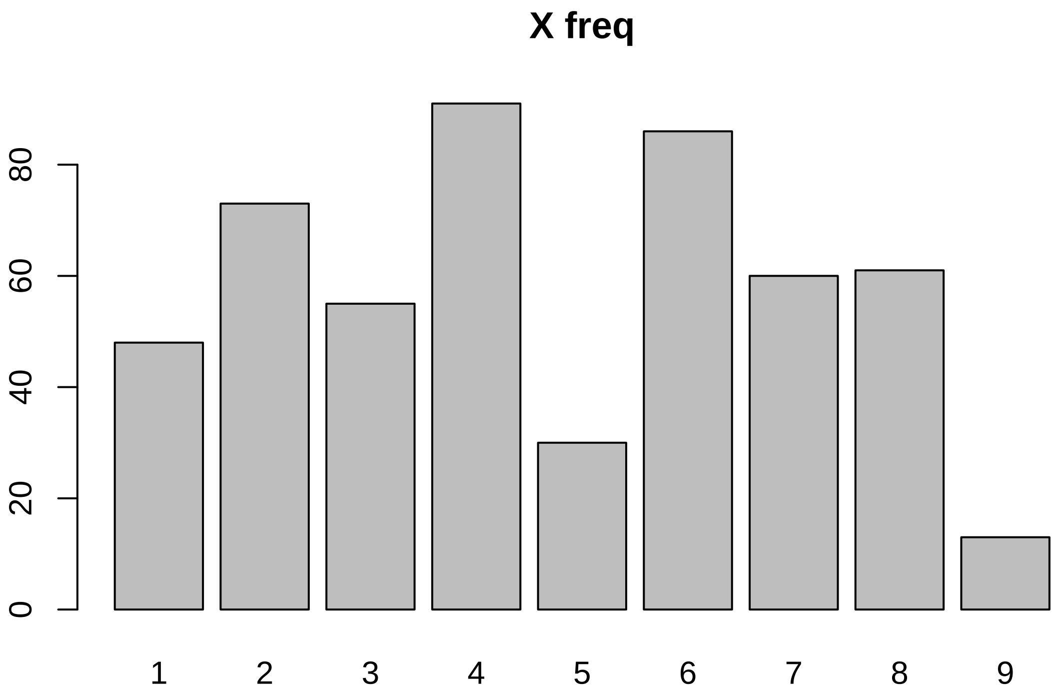
CLOSER LOOK AT EACH VARIABLES

```
##########PICK THE FIRE ONES OUT FIRST

table(forest$X)
```

```
##
##  1  2  3  4  5  6  7  8  9
## 48 73 55 91 30 86 60 61 13
```

```
barplot(table(forest$X), main = "X freq")
```

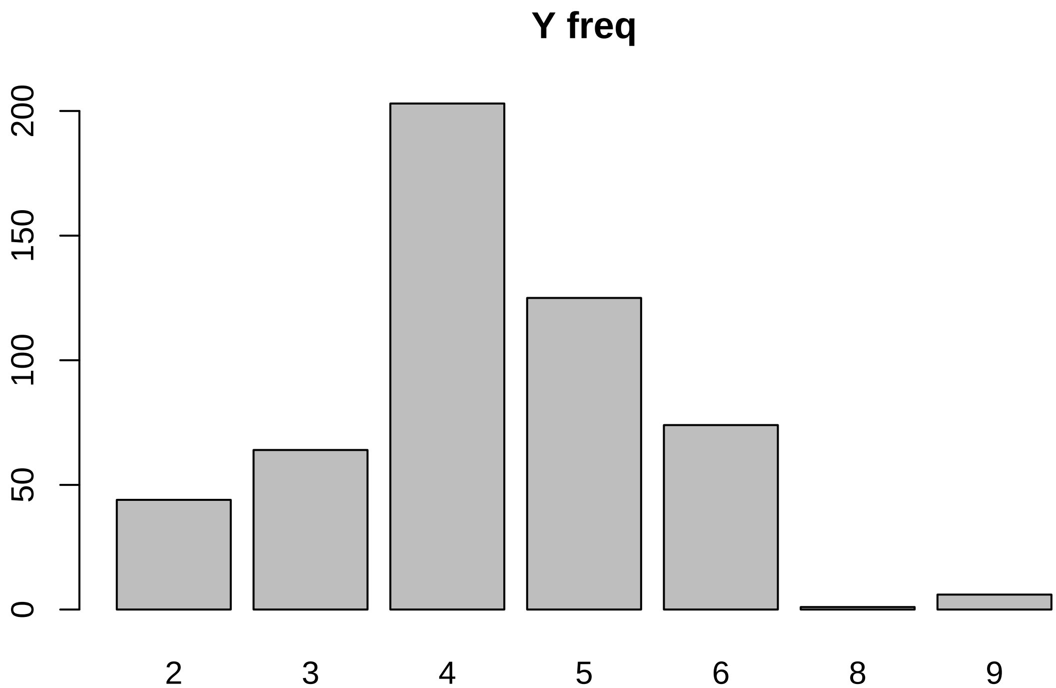



```
## Even number? ## No clear differentiation

table(forest$Y)
```

```
##
##   2   3   4   5   6   8   9
##  44  64 203 125  74   1   6
```

```
barplot(table(forest$Y), main = "Y freq")
```

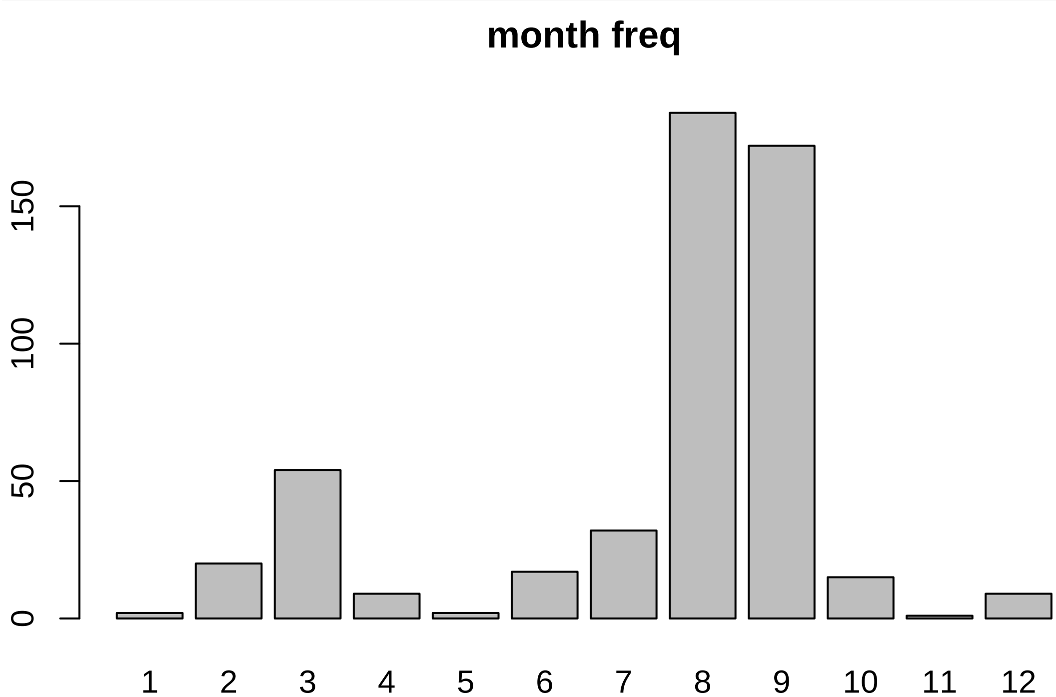



```
## Seems like most of the fire occurs in the location where Y = 4 (5)

table(forest$month)
```

```
##
##   1   2   3   4   5   6   7   8   9  10  11  12
##   2  20  54   9   2  17  32 184 172  15   1   9
```

```
barplot(table(forest$month), main = "month freq")
```

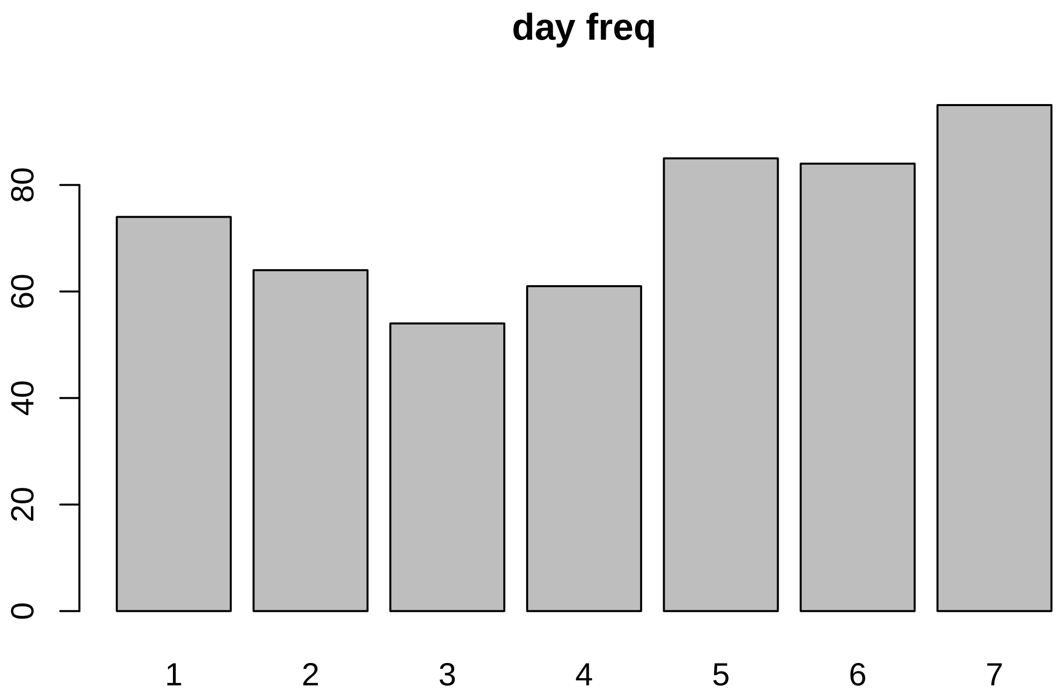


```
## Most of the fire happens in August and September

table(forest$day)
```

```
##
##  1  2  3  4  5  6  7
## 74 64 54 61 85 84 95
```

```
barplot(table(forest$day), main = "day freq")
```



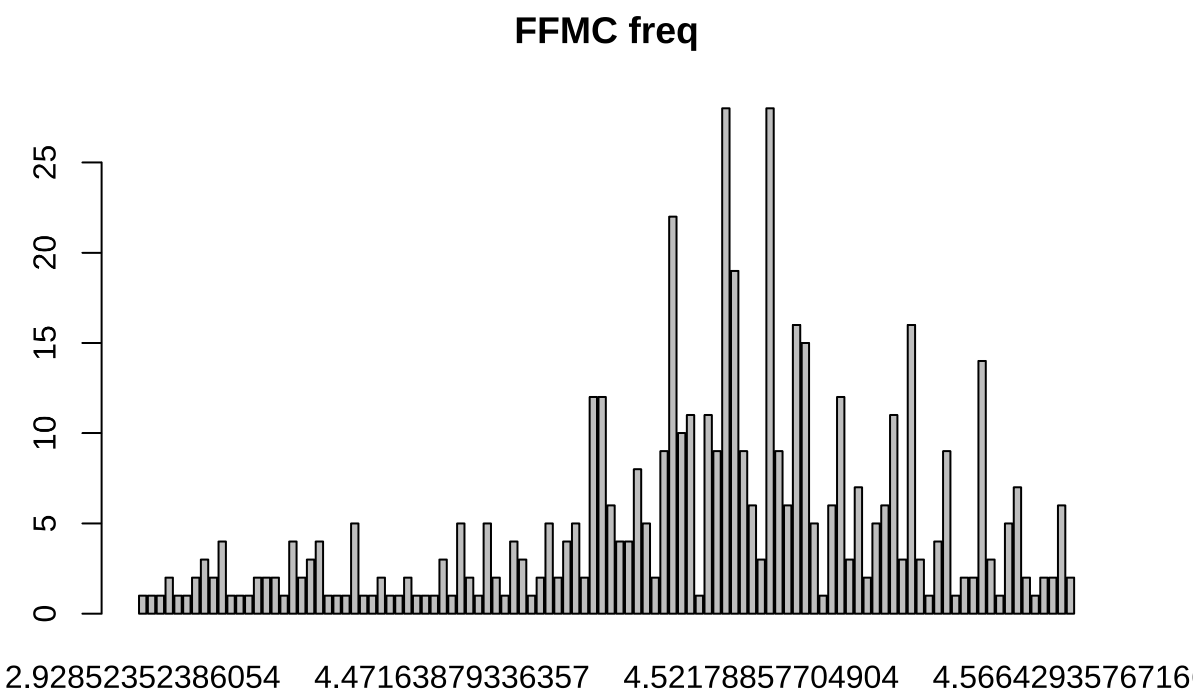

```
## No clear differentiation

table(forest$FFMC)
```

```
##
## 2.92852352386054 3.91999117507732 3.97781074596615 4.15103990589865
##                1                1                1                2
## 4.22244456484942 4.23410650459726 4.31882055877009 4.37575702166029
##                1                1                2                3
## 4.40060302024682 4.40182926197006 4.40549899085902 4.40793801645838
##                2                4                1                1
##  4.4188406077966 4.42962561347316 4.43081679884331  4.4320065669789
##                1                2                2                2
## 4.43319492124828 4.43556740160191 4.43793426661218 4.43911560165801
##                1                4                2                3
##  4.4414740933173 4.44265125649032 4.44382703557933 4.44500143383527
##                4                1                1                1
## 4.44734610079452  4.4496852831477 4.45201900649392 4.45318382899021
##                5                1                1                2
## 4.45667017766965 4.45782959808938 4.46129981556839  4.4636066216663
##                1                1                2                1
## 4.46475803227135 4.46705688385846 4.46820433091493 4.47163879336357
##                1                1                3                1
```

```
## 4.47278099794235 4.47619980469113 4.47733681447821 4.47847253294213
##                5                2                1                5
## 4.47960696301275 4.48074010760991 4.48413185761104 4.48638664999812
##                2                1                4                3
## 4.48751214251986 4.49088103958596 4.49200148788245 4.49312068217947
##                1                2                5                2
## 4.49535531998088 4.49647076906475 4.49980967033027 4.50092016461429
##                4                5                2               12
## 4.50202942706858 4.50313746042294 4.50424426739813 4.50534985070588
##               12                6                4                4
## 4.50645421304893 4.50755735712109 4.50865928560725 4.50976000118343
##                8                5                2                9
## 4.51085950651685 4.51195780426591 4.51305489708029 4.51415078760092
##               22               10               11                1
##  4.5152454784601 4.51633897228148 4.51743127168008 4.51852237926242
##               11                9               28               19
## 4.51961229762644 4.52070102936164 4.52178857704904 4.52287494326126
##                9                6                3               28
## 4.52396013056255 4.52504414150881 4.52612697864764 4.52720864451838
##                9                6               16               15
## 4.52828914165213 4.52936847257181 4.53044663979215 4.53152364581979
##                5                1                6               12
## 4.53259949315326 4.53367418428302 4.53474772169155  4.5358201078533
##                3                7                2                5
##  4.5368913452348 4.53796143629464 4.53903038348355 4.54009818924438
##                6               11                3               16
## 4.54223038621422    4.54329478227 4.54542018158232 4.54648118963941
##                3                1                4                9
## 4.54754107315146  4.5485998344997 4.54965747605783 4.55176940926098
##                1                2                2               14
## 4.55282370561588 4.55387689160054 4.55492896955134 4.55597994179732
##                3                1                5                7
## 4.55912624748668 4.56226268497681 4.56330598188939 4.56434819146784
##                2                1                2                2
## 4.56538931597625 4.56642935767166
##                6                2
```

```r
barplot(table(forest$FFMC), main = "FFMC freq")
```



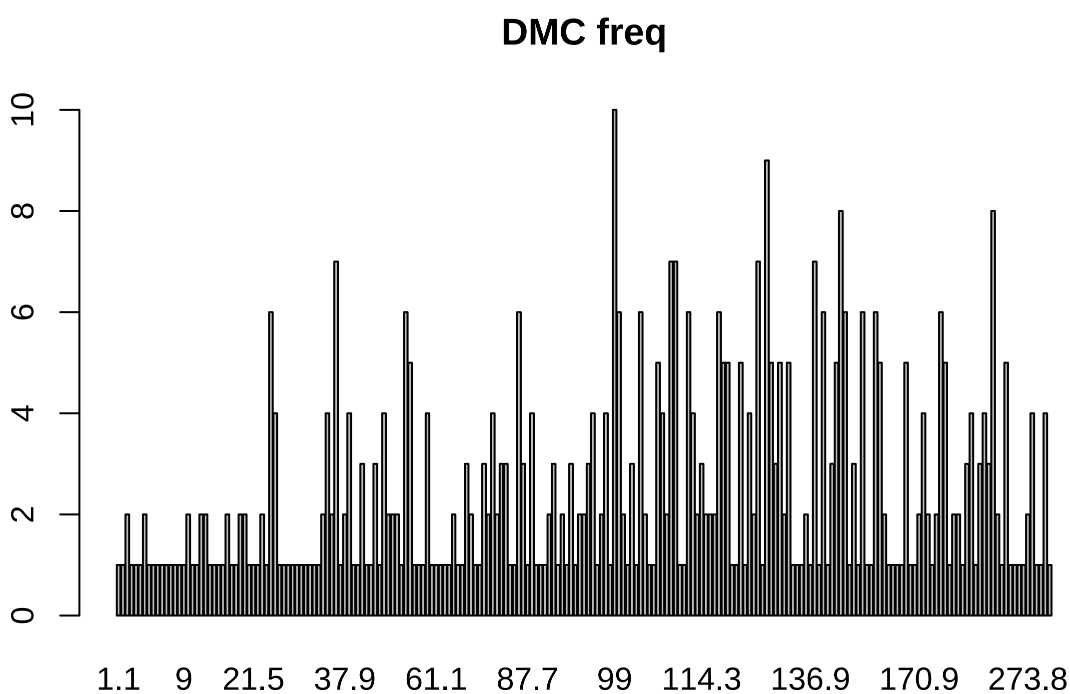

```
## Left Skewed histogram

table(forest$DMC)
```

```
##
##   1.1   2.4     3   3.2   3.6   3.7   4.4   4.6   4.9   6.6   6.8   7.3
##     1     1     2     1     1     1     2     1     1     1     1     1
##     8   8.2   8.7     9   9.1   9.3   9.5  13.2  14.6    15  15.1  15.6
##     1     1     1     1     2     1     1     2     2     1     1     1
##  17.2  17.3  18.2  18.5  18.9  19.5  20.6  21.5  23.3  23.9  24.9  25.4
##     1     2     1     1     2     2     1     1     1     2     1     6
##  25.7  26.2  26.4  26.7  27.2  27.4  27.5  27.8  27.9    28  30.7  32.8
##     4     1     1     1     1     1     1     1     1     1     1     2
##  33.3  35.4  35.8  37.6  37.9  39.7  41.5  41.9  43.7    44  46.2  46.5
##     4     2     7     1     2     4     1     1     3     1     1     3
##  47.9  48.3  48.5  49.5  50.1  51.2  51.3  52.2  53.3  55.2  56.4  56.7
##     1     4     2     2     2     1     6     5     1     1     1     4
##  60.6  61.1  62.3  68.6  69.7  70.8    71  73.2  73.4  75.3  75.6    77
##     1     1     1     1     1     2     1     1     3     2     1     1
##  78.5  80.7  80.9  81.8  82.9  84.1  84.7  84.8  85.1  85.3  87.7    88
##     3     2     4     2     3     3     1     1     6     3     1     4
##  88.2  88.8  88.9  89.5    90  90.4  91.3  91.6  91.8  92.1  93.3  94.1
##     1     1     1     2     3     1     2     1     3     1     2     2
##  94.3  96.2  96.3  96.7  96.9  97.9    99  99.6  99.9 100.2 101.3 102.2
##     3     4     1     2     4     1    10     6     2     1     3     1
## 102.3 103.2 103.8 103.9 104.2 105.8   108 108.3 108.4 109.2 110.9 111.2
##     6     2     1     1     5     4     2     7     7     1     1     6
## 111.7 112.4 114.3 114.4 115.4 117.2 117.9   119 121.1 121.2 121.7   122
##     4     2     3     2     2     2     6     5     5     1     1     5
## 122.3 124.1 124.4 126.5 127.1 129.5 130.1 130.3 131.7 132.3 133.3 133.6
##     1     4     2     7     1     9     5     3     5     2     5     1
## 134.7 135.5 135.7 136.9   137 138.1 139.4 141.1 141.2 141.3 142.4 145.4
```

```
##     1     1     2     1     7     1     6     1     3     5     8     6
##   146 147.3 147.8 149.3 150.3   152 152.6 157.3   158   160 163.2   164
##     1     3     1     6     1     1     6     5     2     1     1     1
## 164.1 166.9 167.6 169.7 170.9 175.1 175.5   178 180.4 181.1 181.3 183.1
##     1     5     1     1     2     4     2     1     2     6     5     1
## 191.4 194.1 196.8 203.2   207 212.1 217.7 222.4   227 231.1 235.1 238.2
##     2     2     1     3     4     1     3     4     3     8     2     1
## 248.4 253.6 263.1 266.2 269.8 273.8 276.3 284.9 287.2   290 291.3
##     5     1     1     1     1     2     4     1     1     4     1
```

```r
barplot(table(forest$DMC), main = "DMC freq")
```



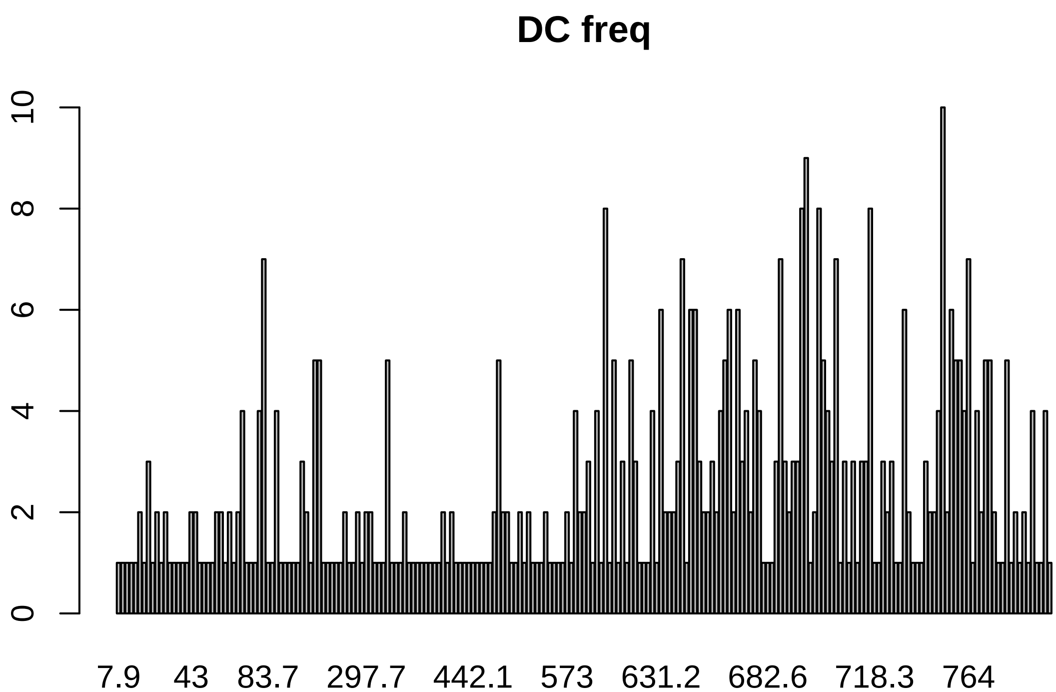

```
## No clear differentiation
```

```r
table(forest$DC)
```

```
##
##   7.9   9.3  15.3  15.5  15.8  16.2  18.7  25.6  26.6  28.3  30.2  30.6
##     1     1     1     1     1     2     1     3     1     2     1     2
##  32.1    34  36.9  41.1  41.6    43  43.5  43.6  46.7  48.3  52.8    55
##     1     1     1     1     1     2     2     1     1     1     1     2
##  55.2  56.9  57.3  58.3  64.7  67.6  70.8  73.7  74.3  77.5  80.8  83.7
##     2     1     2     1     2     4     1     1     1     4     7     1
##  85.3  86.6  87.2  89.4  92.4  94.3  97.1  97.8 100.4 100.7 102.2 103.8
##     1     4     1     1     1     1     1     3     2     1     5     5
## 106.7 113.8 171.4   200   229 232.1 233.8 252.6 290.8 296.3 297.7 298.1
##     1     1     1     1     1     2     1     1     2     1     2     2
## 309.9 313.4 316.7 349.7 350.2   352 352.6 353.5 354.6 355.2 366.7 368.3
##     1     1     1     5     1     1     1     2     1     1     1     1
## 376.6 377.2   395 411.8 423.4 424.1 430.8 431.6 433.3 437.7 440.9 442.1
##     1     1     1     1     2     1     2     1     1     1     1     1
## 442.9 450.2 458.8 466.3 466.6 474.9 480.8   488 495.6 503.6 513.3 520.5
##     1     1     1     1     2     5     2     2     1     1     2     1
```

```
## 529.8 537.4   542 550.3   560 561.6 565.5 567.2 570.5   573 575.8 578.8
##     2     1     1     1     2     1     1     1     1     2     1     4
## 581.1 586.7 587.1 589.9 594.2 596.3 601.4 605.3 605.8 607.1 608.2 609.6
##     2     2     3     1     4     1     8     1     5     1     3     1
##   613 614.5 614.7 621.7 624.1 624.2 629.1 631.2 633.6 635.9 638.8   643
##     5     3     1     1     1     4     1     6     2     2     2     3
## 647.1 649.9 654.1 658.2 661.3 661.8 664.2 664.5 665.3 665.6 666.7   668
##     7     1     6     6     3     2     2     3     2     4     5     6
## 669.1 671.2 671.9 672.6 673.8 674.4 680.7 680.9 682.6 684.4 685.2 686.5
##     2     6     3     4     2     5     4     1     1     1     3     7
## 686.9 689.1   690 691.8 692.3 692.6 694.8 696.1 698.6 699.6 700.7 704.4
##     3     2     3     3     8     9     1     2     8     5     4     3
## 706.4 706.6 706.7 706.8 709.9   713 713.9 714.3 715.1 718.3 721.1 721.4
##     7     1     3     1     3     1     3     3     8     1     1     3
## 723.1 724.3 725.1 726.9 728.6 730.2 730.6 731.7 732.3 735.7 738.1 739.4
##     2     3     1     1     6     2     1     1     1     3     2     2
## 744.4 745.3 750.5 751.5 752.6 753.8 758.1   764 768.4 770.3 777.1 783.5
##     4    10     2     6     5     5     4     7     1     4     2     5
## 789.7 795.3 795.9 803.3 807.1 811.2 812.1 817.5 819.1 822.8 825.1   844
##     5     2     1     1     5     1     2     1     2     1     4     1
## 849.3 855.3 860.6
##     1     4     1
```

```
barplot(table(forest$DC), main = "DC freq")
```



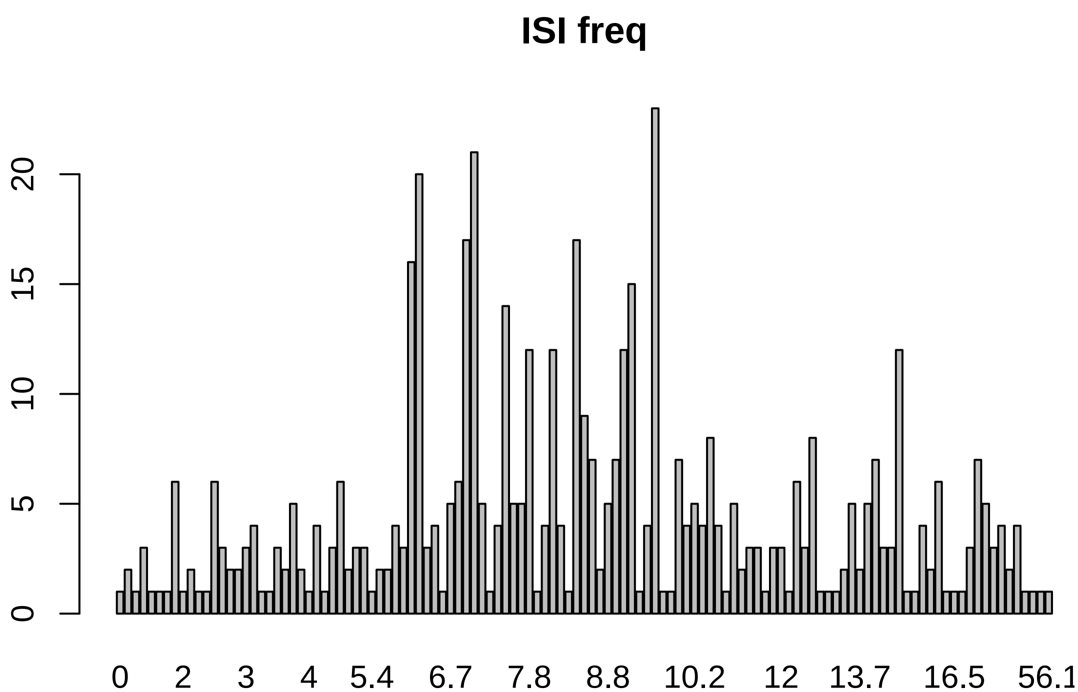

```
## No clear differentiation

table(forest$ISI)
```

```
##
##    0 0.4 0.7 0.8 1.1 1.5 1.8 1.9   2 2.1 2.2 2.3 2.6 2.7 2.8
##    1   2   1   3   1   1   1   6   1   2   1   1   6   3   2
##  2.9   3 3.2 3.3 3.4 3.5 3.7 3.8 3.9   4 4.1 4.7 4.8   5 5.1
```

```
##    2    3    4    1    1    3    2    5    2    1    4    1    3    6    2
##  5.2  5.3  5.4  5.5  5.6  5.7  5.8  6.2  6.3  6.4  6.5  6.6  6.7  6.8    7
##    3    3    1    2    2    4    3   16   20    3    4    1    5    6   17
##  7.1  7.2  7.3  7.4  7.5  7.6  7.7  7.8  7.9    8  8.1  8.2  8.3  8.4  8.5
##   21    5    1    4   14    5    5   12    1    4   12    4    1   17    9
##  8.6  8.7  8.8  8.9    9  9.2  9.4  9.5  9.6  9.7  9.8  9.9 10.1 10.2 10.4
##    7    2    5    7   12   15    1    4   23    1    1    7    4    5    4
## 10.6 10.7 10.8   11 11.1 11.3 11.4 11.6 11.9   12 12.1 12.2 12.3 12.5 12.7
##    8    4    1    5    2    3    3    1    3    3    1    6    3    8    1
## 12.9   13 13.2 13.5 13.7 13.8 13.9   14 14.1 14.3 14.4 14.6 14.7 15.1 15.9
##    1    1    2    5    2    5    7    3    3   12    1    1    4    2    6
## 16.3 16.5 16.7 16.8   17 17.7 17.9   18   20 20.3 21.3 22.6 22.7 56.1
##    1    1    1    3    7    5    3    4    2    4    1    1    1    1
```

```
barplot(table(forest$ISI), main = "ISI freq")
```



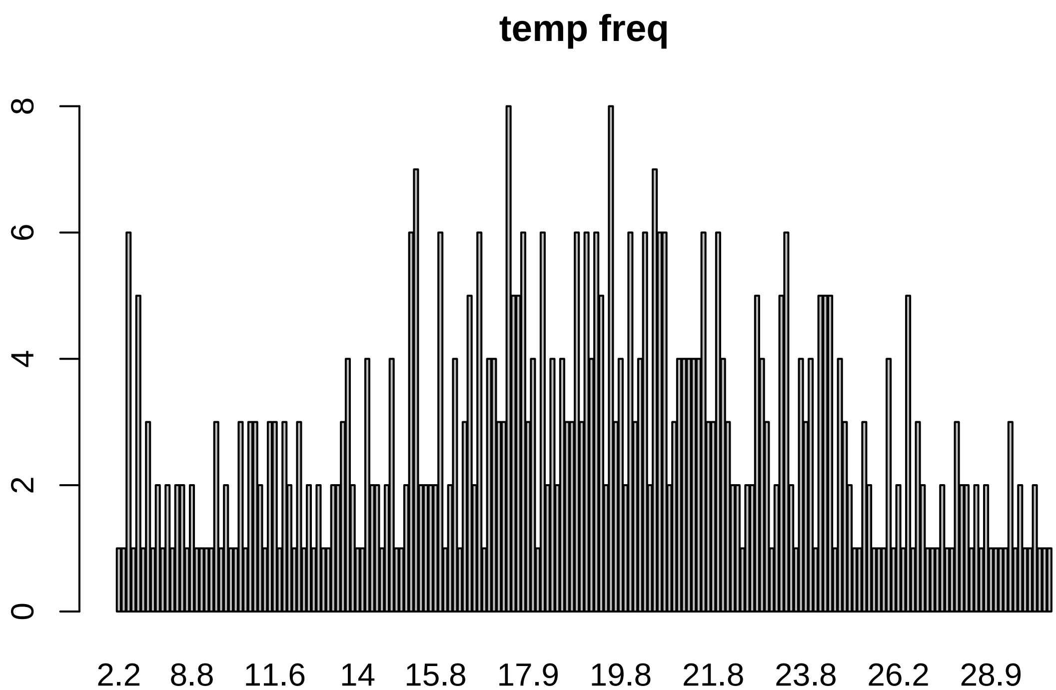

```
## No clear differentiation

table(forest$temp)
```

```
##
##  2.2  4.2  4.6  4.8  5.1  5.2  5.3  5.5  5.8  6.7  7.5    8  8.2  8.3  8.7
##    1    1    6    1    5    1    3    1    2    1    2    1    2    2    1
##  8.8  8.9    9  9.3  9.8 10.1 10.2 10.3 10.4 10.5 10.6 10.9   11 11.2 11.3
##    2    1    1    1    1    3    1    2    1    1    3    1    3    3    2
## 11.4 11.5 11.6 11.7 11.8 12.2 12.3 12.4 12.6 12.7 12.8 12.9 13.1 13.2 13.3
##    1    3    3    1    3    2    1    3    1    2    1    2    1    1    2
## 13.4 13.7 13.8 13.9   14 14.1 14.2 14.3 14.4 14.5 14.6 14.7 14.8 14.9 15.1
##    2    3    4    2    1    1    4    2    2    1    2    4    1    1    2
## 15.2 15.4 15.5 15.6 15.7 15.8 15.9   16 16.1 16.2 16.3 16.4 16.6 16.7 16.8
##    6    7    2    2    2    2    6    1    2    4    1    3    5    2    6
## 16.9   17 17.1 17.2 17.3 17.4 17.6 17.7 17.8 17.9   18 18.1 18.2 18.3 18.4
##    1    4    4    3    3    8    5    5    6    3    4    1    6    2    4
```

```
## 18.5 18.6 18.7 18.8 18.9   19 19.1 19.2 19.3 19.4 19.5 19.6 19.7 19.8 19.9
##    2    4    3    3    6    3    6    4    6    5    2    8    3    4    2
## 20.1 20.2 20.3 20.4 20.5 20.6 20.7 20.8 20.9   21 21.1 21.2 21.3 21.4 21.5
##    6    3    4    6    2    7    6    6    2    3    4    4    4    4    4
## 21.6 21.7 21.8 21.9 22.1 22.2 22.3 22.4 22.5 22.6 22.7 22.8 22.9   23 23.1
##    6    3    3    6    4    3    2    2    1    2    2    5    4    3    1
## 23.2 23.3 23.4 23.5 23.6 23.7 23.8 23.9   24 24.1 24.2 24.3 24.5 24.6 24.8
##    2    5    6    2    1    4    3    4    1    5    5    5    1    4    3
## 24.9   25 25.1 25.3 25.4 25.5 25.6 25.7 25.9 26.1 26.2 26.3 26.4 26.7 26.8
##    2    1    1    3    2    1    1    1    4    1    2    1    5    1    3
## 26.9 27.2 27.3 27.4 27.5 27.6 27.7 27.8 27.9   28 28.2 28.3 28.6 28.7 28.9
##    2    1    1    1    2    1    1    3    2    2    1    2    1    2    1
## 29.2 29.3 29.6 30.2 30.6 30.8   31 32.3 32.4 32.6 33.1 33.3
##    1    1    1    3    1    2    1    1    2    1    1    1
```

```
barplot(table(forest$temp), main = "temp freq")
```

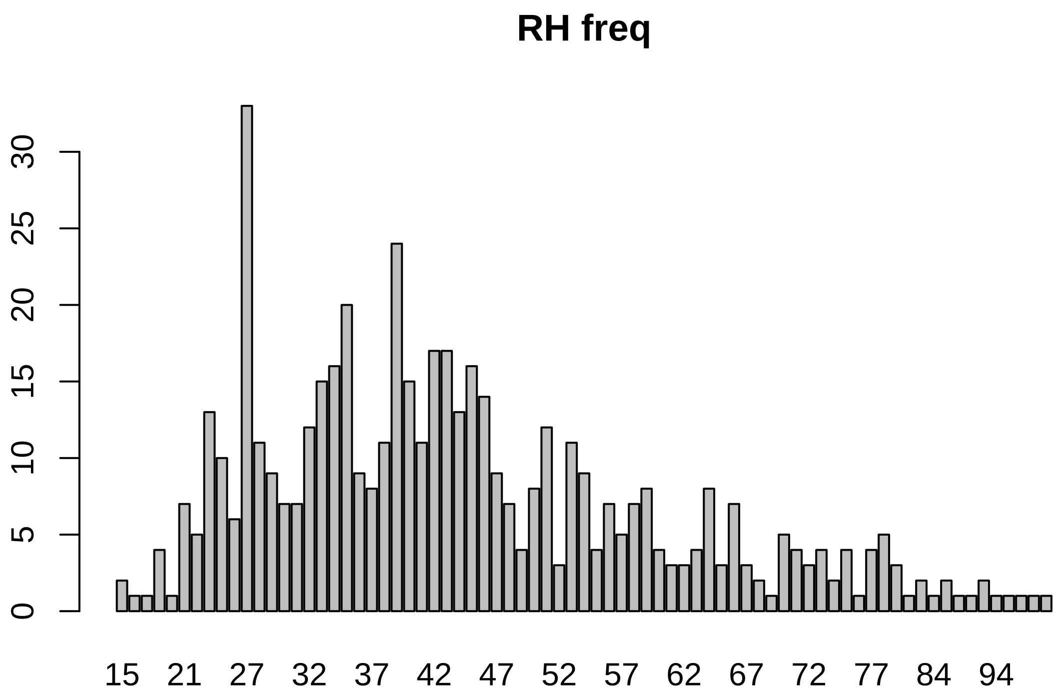



```
## No clear differentiation

table(forest$RH)
```

```
##
##  15  17  18  19  20  21  22  24  25  26  27  28  29  30  31  32  33  34
##   2   1   1   4   1   7   5  13  10   6  33  11   9   7   7  12  15  16
##  35  36  37  38  39  40  41  42  43  44  45  46  47  48  49  50  51  52
##  20   9   8  11  24  15  11  17  17  13  16  14   9   7   4   8  12   3
##  53  54  55  56  57  58  59  60  61  62  63  64  65  66  67  68  69  70
##  11   9   4   7   5   7   8   4   3   3   4   8   3   7   3   2   1   5
##  71  72  73  74  75  76  77  78  79  80  82  84  86  87  88  90  94  96
##   4   3   4   2   4   1   4   5   3   1   2   1   2   1   1   2   1   1
##  97  99 100
##   1   1   1
```

```r
barplot(table(forest$RH), main = "RH freq")
```

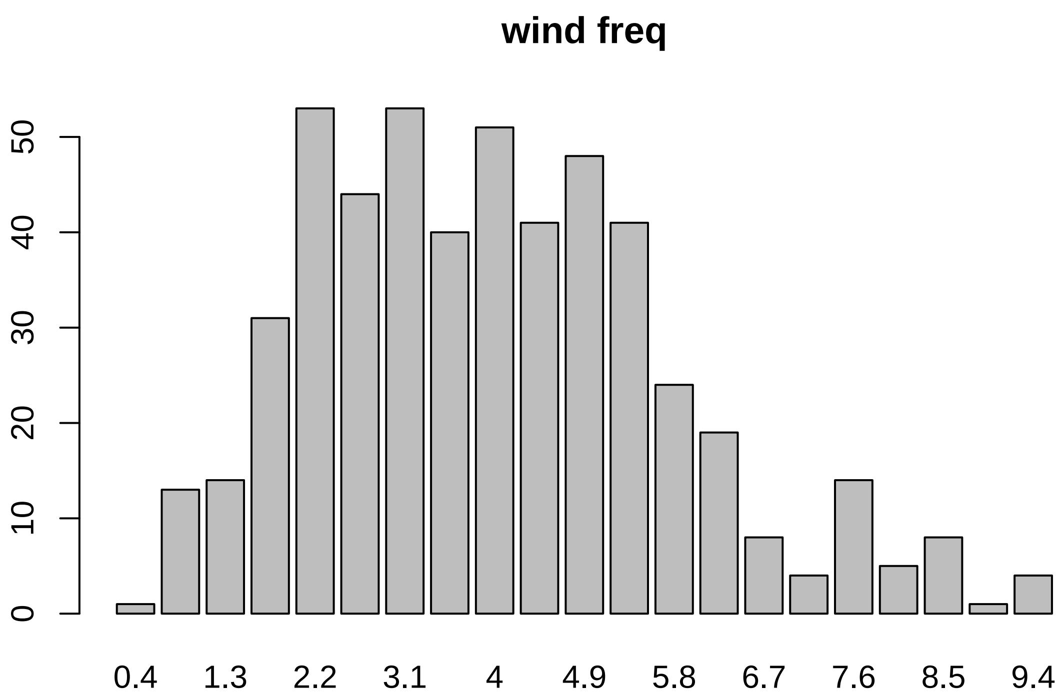



```
## Right Skewed histogram
```

```r
table(forest$wind)
```

```
##
## 0.4 0.9 1.3 1.8 2.2 2.7 3.1 3.6   4 4.5 4.9 5.4 5.8 6.3 6.7 7.2 7.6   8
##   1  13  14  31  53  44  53  40  51  41  48  41  24  19   8   4  14   5
## 8.5 8.9 9.4
##   8   1   4
```

```r
barplot(table(forest$wind), main = "wind freq")
```




```r
## Right Skewed histogram

table(forest$rain)
```

```
## < table of extent 0 >
```

```r
# barplot(table(forest$rain), main = "rain freq")
## Most of the fire happens when there's no rain

table(forest$area)
```

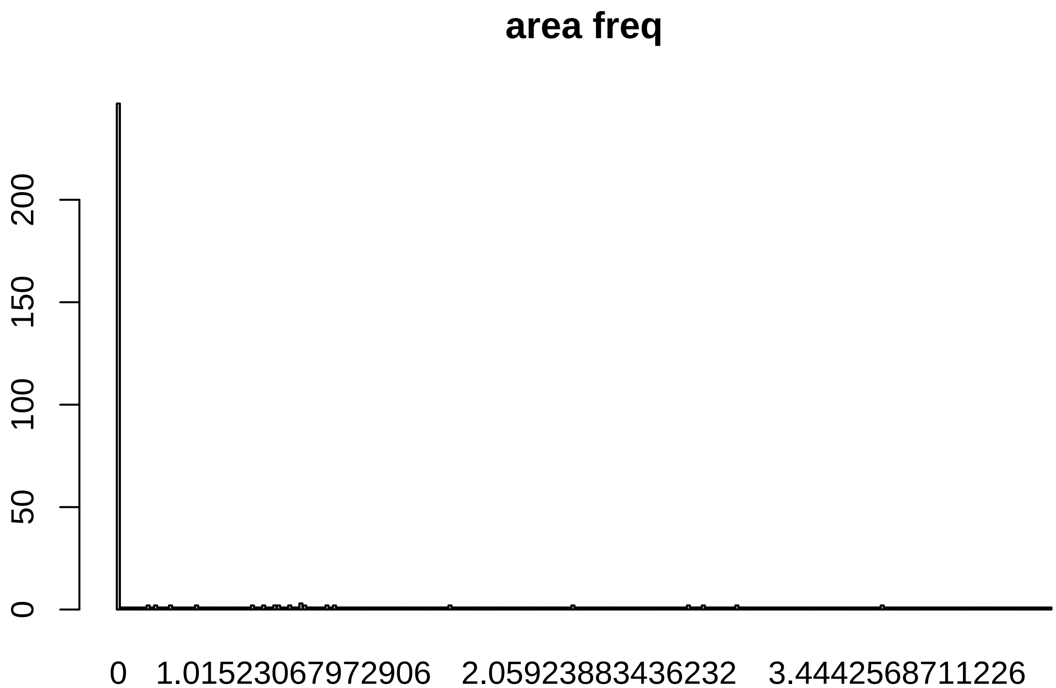
```
##
##                  0 0.0861776962410524  0.157003748809665
##                247                  1                  1
##   0.19062035960865  0.215111379616945  0.285178942233662
##                  1                  1                  1
##  0.307484699747961  0.343589704390077  0.357674444271816
##                  1                  1                  2
##  0.385262400790645  0.418710334858185  0.431782416425538
##                  1                  2                  1
##  0.438254930931155  0.476234178996372  0.518793793415168
##                  1                  1                  2
##  0.536493370514568  0.542324290825362  0.559615787935423
##                  1                  1                  1
##   0.56531380905006  0.570979546585738  0.582215619852664
##                  1                  1                  1
##  0.641853886172395  0.667829372575655  0.672944473242426
##                  2                  1                  1
##  0.698134722070984  0.727548607277278   0.73716406597672
##                  1                  1                  1
##  0.741937344729377  0.751416088683921  0.783901543828409
##                  1                  1                  1
##  0.802001585472027  0.815364813284194  0.828551817566148
```

```
##                1                1                1
## 0.858661619037519 0.867100487683383 0.887891257352457
##                1                1                1
## 0.900161349944271 0.904218150639886 0.924258901523332
##                2                1                1
## 0.940007258491471 0.947789398933526 0.959350221334602
##                2                1                1
## 0.966983846189673 0.970778917158225 0.989541193613748
##                2                2                1
##  1.00063188030791  1.01160091167848  1.01523067972906
##                1                2                1
##  1.06471073699243  1.07840958135059  1.08180517035173
##                1                3                2
##  1.09861228866811  1.10194007876078  1.10856261952128
##                1                1                1
##  1.12167756159911  1.14103300455206  1.14422279992016
##                1                1                2
##  1.15373158788919  1.15688119679209  1.16627093714192
##                1                2                1
##  1.19088756477728  1.20896034583698  1.23547147138531
##                1                1                1
##  1.24415459395877  1.25561603747777  1.26129787094521
##                1                1                1
##  1.26694760348732  1.27256559579155  1.29198368164865
##                1                1                1
##  1.30562645805244  1.31908561142644  1.32175583998232
##                1                1                1
##  1.32707500145992  1.35325450704169  1.36863942588117
##                1                1                1
##  1.39871688111845   1.4036429994545  1.40854497005471
##                1                1                1
##  1.43031124653666  1.43270073393405  1.43508452528932
##                1                1                1
##  1.45861502269952  1.46325540225602   1.4655675420144
##                1                1                1
##  1.47017584510059  1.50407739677627  1.50851199384414
##                1                1                1
##  1.53255686809814  1.53471436623816  1.54968790802833
##                1                1                2
##  1.56444054650336   1.5953389880546  1.59736533119983
##                1                1                1
##   1.5993875765806  1.65822807660353  1.68639895357023
##                1                1                1
##  1.68824909285839  1.69009581545155  1.71018781553424
##                1                1                1
##  1.72455071953461   1.7263316639056  1.73871024813824
##                1                1                1
##  1.77155676191054  1.78339121955754  1.78507048107726
##                1                1                1
##   1.8213182714696  1.82937633279936  1.84530023615608
##                1                1                1
##   1.8531680973567  1.85473426838944  1.86252854011626
##                1                1                1
##  1.87946504964716  1.89461685466776  1.91692261218206
```

```
##                1                1                1
##  1.9213246735827 1.92570744173779 1.94161522477243
##                1                1                1
## 1.95160817016995 1.96009478404727 1.98787434815435
##                1                1                1
## 1.99605993274078 1.99877363861238 2.00552585872967
##                1                1                2
##   2.006870848845 2.02022218201986 2.02419306744936
##                1                1                1
## 2.02551319965428 2.02946317187359 2.05796251000271
##                1                1                1
## 2.05923883436232 2.07442899985629 2.08193842187842
##                1                1                1
## 2.08442908319087 2.10291389786498 2.10535292346434
##                1                1                1
## 2.11625551480255 2.11745960886736 2.12345842709661
##                1                1                1
## 2.12823170584927 2.13771044980381 2.16676536985151
##                1                1                1
## 2.17475172148416 2.19722457733622 2.19944433407453
##                1                1                1
## 2.21046980408624 2.21484617868604 2.22354188565359
##                1                1                1
## 2.23108909128898 2.26072088889535 2.27006190128849
##                1                1                1
## 2.27315628230323   2.287471455184 2.30058309032337
##                1                1                1
## 2.32922702394047 2.34276688262688 2.37117788445966
##                2                1                1
##  2.3767644911683 2.39425228151987 2.39880395073459
##                1                2                1
## 2.39971180372477 2.40514168131914 2.40964416528745
##                1                1                1
## 2.42833629829961 2.46214966266538 2.46979301197795
##                1                1                1
## 2.47905623610982 2.48989419129904 2.49815187653802
##                1                2                1
## 2.50061594349318 2.50307395374345 2.50470927708418
##                1                1                1
## 2.51122395810537 2.52812576890798 2.57261223020711
##                1                1                1
## 2.57870052907436 2.61300665241532 2.64262239577975
##                1                1                1
## 2.64333388638252 2.68784749378469 2.70738331211451
##                1                1                1
## 2.72719901994097 2.74534598584591 2.75238601492226
##                1                1                1
## 2.79361608943186 2.80032547721138 2.81180943539306
##                1                1                1
## 2.83321334405622 2.85243910372751 2.85647020622048
##                1                1                1
## 2.90142159408275 2.93651291389402 2.96010509591084
##                1                1                1
## 3.00716665117965 3.04594998971461 3.13679771383259
```

```
##                1                1                1
##  3.19499288440487 3.22803376265297 3.22843003767301
##                1                1                1
##   3.2422016501717 3.24921102466427 3.29583686600433
##                1                1                1
##  3.30064012667084 3.31163730494951 3.34462703017376
##                1                1                1
##  3.37382618486602 3.38979933670979 3.39249294103201
##                1                2                1
##  3.41707073081845 3.43977686362963  3.4442568711226
##                1                1                1
##  3.48798651173455 3.49225611260912  3.4986265269937
##                1                1                1
##  3.56558123776944 3.60766939868839 3.63363097988346
##                1                1                1
##  3.63811233706028 3.65609796458956 3.67579421456528
##                1                1                1
##  3.69759139471596 3.72665681844797 3.78123071517812
##                1                1                1
##  3.79143604243903  3.8649313978943   3.902982260776
##                1                1                1
##  3.91939575975756 3.92375392830384 3.96613233107518
##                1                1                1
##  4.01259206034984 4.04375277610604 4.08260930600368
##                1                1                1
##   4.1292289640756 4.17592454921452 4.26717679299494
##                1                1                1
##  4.27332721775054 4.28082412916472 4.42783617070518
##                1                1                1
##  4.47106720146461 4.49412688719477 4.56622143584952
##                1                1                1
##  4.64813388542021 4.66964620517246 5.04908648047051
##                1                1                1
##  5.16837950943364 5.22982437010274 5.28563731339066
##                1                1                1
##  5.30797062357617 5.36541511008053 5.63310962136115
##                1                1                1
##  6.61643994756459 6.99561962542321
##                1                1
```

```r
barplot(table(forest$area), main = "area freq")
```




```
## No clear differentiation
### THIS IS A RESULT OF FIRE
```