data science

sai vijay

12/17/2021

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

library(corrplot)

## corrplot 0.91 loaded

library(car)

## Warning: package 'car' was built under R version 4.1.2

## Loading required package: carData

library(perturbR)

## Warning: package 'perturbR' was built under R version 4.1.2

library(ggplot2)  
library(MASS)

## Warning: package 'MASS' was built under R version 4.1.2

# Reading Data File and some working with it

#Reading Data file  
forest <- read.csv("forestfires.csv")  
  
# Dimensions and variables of dataset  
dim(forest)

## [1] 517 13

colnames(forest)

## [1] "X" "Y" "month" "day" "FFMC" "DMC" "DC" "ISI" "temp"   
## [10] "RH" "wind" "rain" "area"

# Convert month and day string variables into numeric values  
forest$month <- as.numeric(as.factor(forest$month))  
forest$day <- as.numeric(as.factor(forest$day))  
  
#checking summary of variables and if any missing values  
sum(is.na(forest))

## [1] 0

summary(forest)

## X Y month day FFMC   
## Min. :1.000 Min. :2.0 Min. : 1.000 Min. :1.000 Min. :18.70   
## 1st Qu.:3.000 1st Qu.:4.0 1st Qu.: 2.000 1st Qu.:2.000 1st Qu.:90.20   
## Median :4.000 Median :4.0 Median : 7.000 Median :4.000 Median :91.60   
## Mean :4.669 Mean :4.3 Mean : 6.758 Mean :3.737 Mean :90.64   
## 3rd Qu.:7.000 3rd Qu.:5.0 3rd Qu.:12.000 3rd Qu.:5.000 3rd Qu.:92.90   
## Max. :9.000 Max. :9.0 Max. :12.000 Max. :7.000 Max. :96.20   
## DMC DC ISI temp   
## Min. : 1.1 Min. : 7.9 Min. : 0.000 Min. : 2.20   
## 1st Qu.: 68.6 1st Qu.:437.7 1st Qu.: 6.500 1st Qu.:15.50   
## Median :108.3 Median :664.2 Median : 8.400 Median :19.30   
## Mean :110.9 Mean :547.9 Mean : 9.022 Mean :18.89   
## 3rd Qu.:142.4 3rd Qu.:713.9 3rd Qu.:10.800 3rd Qu.:22.80   
## Max. :291.3 Max. :860.6 Max. :56.100 Max. :33.30   
## RH wind rain area   
## Min. : 15.00 Min. :0.400 Min. :0.00000 Min. : 0.00   
## 1st Qu.: 33.00 1st Qu.:2.700 1st Qu.:0.00000 1st Qu.: 0.00   
## Median : 42.00 Median :4.000 Median :0.00000 Median : 0.52   
## Mean : 44.29 Mean :4.018 Mean :0.02166 Mean : 12.85   
## 3rd Qu.: 53.00 3rd Qu.:4.900 3rd Qu.:0.00000 3rd Qu.: 6.57   
## Max. :100.00 Max. :9.400 Max. :6.40000 Max. :1090.84

No missing values. Month and day are as factors. From Summary FFMC, DMC and DC seem left skewed.ISI , Rain may be right skewed. Area heavily right skewed # Splitting data into training and test set

# Splitting data into training and test set  
set.seed(30032017)  
row.number<- sample(1:nrow(forest), size=0.2\*nrow(forest))  
forest\_test<- forest[row.number,]  
dim(forest\_test) ## Size of the testing set

## [1] 103 13

forest\_train<- forest[-row.number,]  
dim(forest\_train) ## Size of training set

## [1] 414 13

summary(forest\_train)

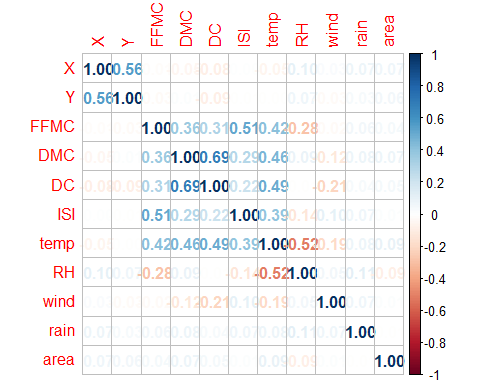
## X Y month day   
## Min. :1.000 Min. :2.000 Min. : 1.000 Min. :1.000   
## 1st Qu.:3.000 1st Qu.:4.000 1st Qu.: 2.000 1st Qu.:2.000   
## Median :4.000 Median :4.000 Median : 7.000 Median :4.000   
## Mean :4.652 Mean :4.316 Mean : 6.778 Mean :3.754   
## 3rd Qu.:7.000 3rd Qu.:5.000 3rd Qu.:12.000 3rd Qu.:5.000   
## Max. :9.000 Max. :9.000 Max. :12.000 Max. :7.000   
## FFMC DMC DC ISI   
## Min. :18.70 Min. : 1.10 Min. : 15.3 Min. : 0.000   
## 1st Qu.:90.30 1st Qu.: 73.88 1st Qu.:460.7 1st Qu.: 6.525   
## Median :91.60 Median :108.30 Median :661.8 Median : 8.400   
## Mean :90.68 Mean :112.23 Mean :549.5 Mean : 9.132   
## 3rd Qu.:92.80 3rd Qu.:142.40 3rd Qu.:713.7 3rd Qu.:11.000   
## Max. :96.20 Max. :291.30 Max. :860.6 Max. :56.100   
## temp RH wind rain   
## Min. : 4.20 Min. : 15.00 Min. :0.400 Min. :0.0000   
## 1st Qu.:15.90 1st Qu.: 32.25 1st Qu.:2.700 1st Qu.:0.0000   
## Median :19.40 Median : 40.50 Median :4.000 Median :0.0000   
## Mean :19.12 Mean : 43.55 Mean :3.999 Mean :0.0256   
## 3rd Qu.:22.88 3rd Qu.: 53.00 3rd Qu.:5.400 3rd Qu.:0.0000   
## Max. :33.10 Max. :100.00 Max. :9.400 Max. :6.4000   
## area   
## Min. : 0.000   
## 1st Qu.: 0.000   
## Median : 0.385   
## Mean : 12.771   
## 3rd Qu.: 5.995   
## Max. :1090.840

# Correlation matrix

# Now we check the correlation matrix  
M <- cor(forest\_train[,-c(3,4)])  
M

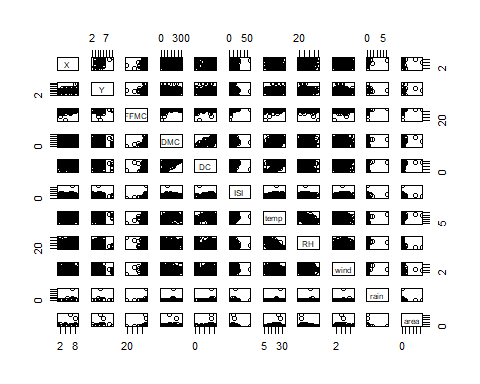
## X Y FFMC DMC DC  
## X 1.000000000 0.555815132 -0.01264825 -0.04748225 -0.081634076  
## Y 0.555815132 1.000000000 -0.02839400 0.01337154 -0.085468283  
## FFMC -0.012648249 -0.028393996 1.00000000 0.36450301 0.306235383  
## DMC -0.047482249 0.013371539 0.36450301 1.00000000 0.689149989  
## DC -0.081634076 -0.085468283 0.30623538 0.68914999 1.000000000  
## ISI 0.002332372 -0.009040472 0.51469812 0.28562713 0.215971018  
## temp -0.051680997 -0.001651361 0.41664954 0.45635976 0.492601704  
## RH 0.101135606 0.069925060 -0.28408131 0.08760536 -0.007500934  
## wind 0.027410477 -0.026852782 -0.02420261 -0.12119978 -0.214377146  
## rain 0.072293426 0.034257854 0.05869154 0.07724143 0.038985914  
## area 0.069143610 0.055872288 0.04259023 0.07424035 0.049340940  
## ISI temp RH wind rain  
## X 0.002332372 -0.051680997 0.101135606 0.02741048 0.072293426  
## Y -0.009040472 -0.001651361 0.069925060 -0.02685278 0.034257854  
## FFMC 0.514698121 0.416649538 -0.284081313 -0.02420261 0.058691539  
## DMC 0.285627128 0.456359760 0.087605359 -0.12119978 0.077241429  
## DC 0.215971018 0.492601704 -0.007500934 -0.21437715 0.038985914  
## ISI 1.000000000 0.393163956 -0.139210864 0.10138257 0.067244655  
## temp 0.393163956 1.000000000 -0.516526599 -0.19124628 0.078330132  
## RH -0.139210864 -0.516526599 1.000000000 0.05336059 0.111950702  
## wind 0.101382568 -0.191246278 0.053360592 1.00000000 0.066265873  
## rain 0.067244655 0.078330132 0.111950702 0.06626587 1.000000000  
## area 0.011720108 0.092980803 -0.085854528 0.01018596 -0.006688597  
## area  
## X 0.069143610  
## Y 0.055872288  
## FFMC 0.042590225  
## DMC 0.074240353  
## DC 0.049340940  
## ISI 0.011720108  
## temp 0.092980803  
## RH -0.085854528  
## wind 0.010185955  
## rain -0.006688597  
## area 1.000000000

# And the correlation plot to visualize the correlation between variables in training data  
corrplot(M,method='number')



evident positive corr between DC & DMC,ISI &FFMC,X &Y, temp&DC. negative corr between RH& temp.lets see it visually by scatter plots

pairs(forest\_train[,-c(3,4)])

 visual scatter plots rule out some correlation and we can shortlist below ones positive DC&DMC - this as per definition makes sense positive temp &DMC - This somewhat does not makes much sense as moisture should decrease with temp neagtive RH & temp - This is also natural as temp increases humidity decreases. We then inspect the distribution of each variable in boxplots

boxplot(forest\_train$X,main="X")

boxplot(forest\_train$Y,main ='Y')

boxplot(forest\_train$FFMC, main='FFMC') #outliers

boxplot(forest\_train$DMC, main ='DMC') # outliers

boxplot(forest\_train$DC, main='DC') # some outliers

boxplot(forest\_train$ISI,main='ISI') # outliers

boxplot(forest\_train$temp, main='temp')

boxplot(forest\_train$RH,main="RH") # outliers

boxplot(forest\_train$wind, main='wind') #boxplot(forest\_train$rain, main='rain') # heavy outliers...high variability in data

boxplot(forest\_train$area, main='area') # heavy outliers..high variability in data

Chart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generatedA picture containing chart

Description automatically generatedChart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generatedDiagram

Description automatically generated with low confidenceChart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generated

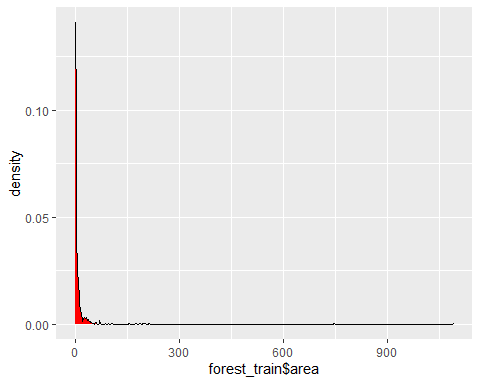
A picture containing diagram

Description automatically generatedDiagram

Description automatically generated

asymmetry also observed in variables like X,Y,DC,FFMC. Let’s see the prob density distribution curve of response variable area

dar <- data.frame(x=forest\_train$area)  
ggplot(dar,aes(x=forest\_train$area))+geom\_density(fill='red')



# And density curve for other variables also  
plot(density(forest\_train$FFMC))

plot(density(forest\_train$DMC))

plot(density(forest\_train$DC))

plot(density(forest\_train$ISI))

plot(density(forest\_train$temp))

plot(density(forest\_train$RH))

Chart, histogram

Description automatically generatedChart, line chart, histogram

Description automatically generatedChart, line chart, histogram

Description automatically generatedChart, histogram

Description automatically generatedChart, line chart, histogram

Description automatically generatedChart, histogram

Description automatically generated

plot(density(forest\_train$wind))

plot(density(forest\_train$rain))

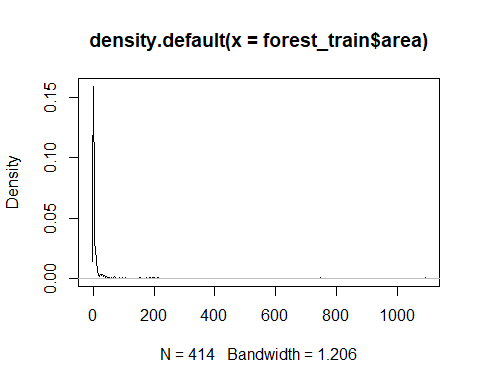
plot(density(forest\_train$area))

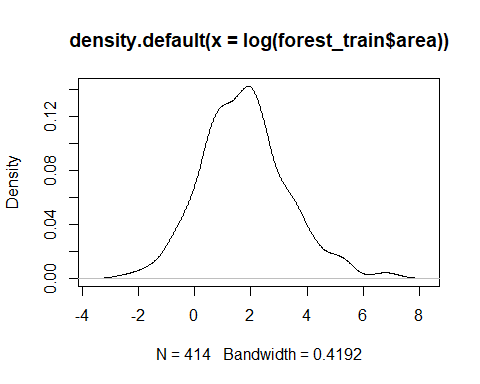
plot(density(log(forest\_train$rain))) # log

plot(density(log(forest\_train$area))) # log

Chart, histogram

Description automatically generatedChart, histogram

Description automatically generatedChart, histogram

Description automatically generated

Above boxplots and density suggest reflected log transform of FFMC and log transform of rain and area, the response variable since it is highly concentrated near zero and assymetrical

summary(forest\_train$area)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.000 0.000 0.385 12.771 5.995 1090.840

var(forest\_train$area)

## [1] 4790.204

sd(forest\_train$area)

## [1] 69.2113

The variable distribution is very concentrated around 1 and 10, but we can some extreme outliers , even above 1000 !!

# Building Models

## We run the first Basic Model

mod1 <- lm(area~X+Y+month+day+FFMC+DMC+DC+ISI+temp+RH+wind+rain,data=forest\_train)  
  
summary(mod1)

##   
## Call:  
## lm(formula = area ~ X + Y + month + day + FFMC + DMC + DC + ISI +   
## temp + RH + wind + rain, data = forest\_train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -46.66 -16.87 -8.36 0.71 1059.89   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -5.83971 74.30506 -0.079 0.937  
## X 2.21547 1.76942 1.252 0.211  
## Y 1.00102 3.31957 0.302 0.763  
## month 1.10659 0.96123 1.151 0.250  
## day 0.64786 1.81926 0.356 0.722  
## FFMC -0.15440 0.77137 -0.200 0.841  
## DMC 0.14029 0.08819 1.591 0.112  
## DC -0.01576 0.02367 -0.666 0.506  
## ISI -0.47991 0.90137 -0.532 0.595  
## temp 0.71890 1.00704 0.714 0.476  
## RH -0.33880 0.30685 -1.104 0.270  
## wind 1.60584 2.06256 0.779 0.437  
## rain -2.41291 10.64647 -0.227 0.821  
##   
## Residual standard error: 69.29 on 401 degrees of freedom  
## Multiple R-squared: 0.02698, Adjusted R-squared: -0.002142   
## F-statistic: 0.9264 on 12 and 401 DF, p-value: 0.5202

The R sq is very low at 4.6% and only DMC and DC seems significant regressors

par(mfrow=c(2, 2))  
plot(mod1)

qqPlot(mod1)

## 239 416   
## 191 331

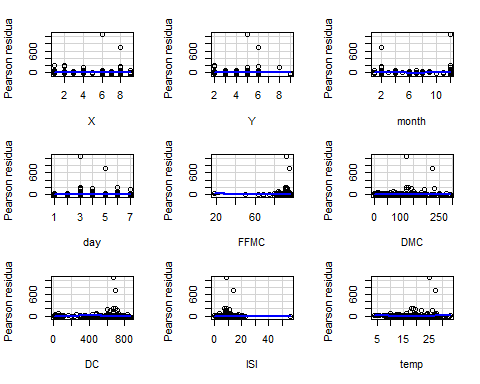
Diagram, schematic

Description automatically generated Chart

Description automatically generated

There is negative linear reation between Residuals and Fitted values. QQ plot of residuals is also not linear. This indicates there can be Collinearity problems. Lets see residual plot with variables.

residualPlots(mod1)

Chart, box and whisker chart

Description automatically generated

## Test stat Pr(>|Test stat|)   
## X 0.6638 0.507203   
## Y 0.4896 0.624683   
## month 1.8751 0.061501 .   
## day -1.4621 0.144489   
## FFMC 0.4888 0.625241   
## DMC -0.6686 0.504165   
## DC -0.7027 0.482680   
## ISI -0.1211 0.903646   
## temp 0.8621 0.389153   
## RH 0.5861 0.558126   
## wind -0.4283 0.668684   
## rain 0.3091 0.757380   
## Tukey test 3.1620 0.001567 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

The residual plots suggests very significant pattern for fitted values and residuals. Some square transformations in wind, temp, rain, RH is suggested. But ffirst we observe that there are many zero values in area which is giving very irregular results. Hence, we decided to remove the zero value rows and reduce the dataset and actually run only on data where there is a burn area.

forest\_train <- forest\_train[forest\_train$area>0,]  
forest\_test <- forest\_test[forest\_test$area>0,]

## Model 2

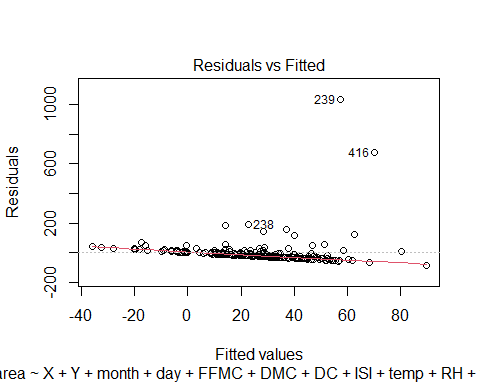
Now we run model 2 on reduced subset of data

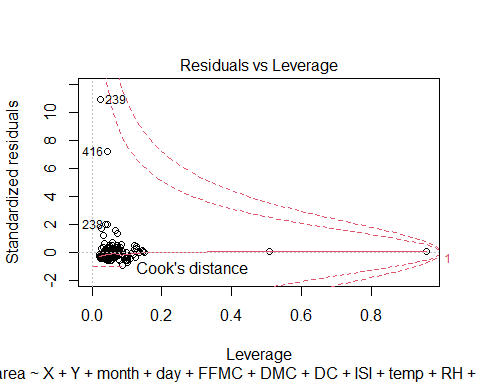
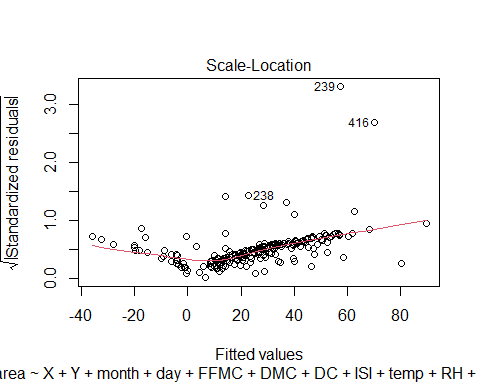
mod2 <- lm(area~X+Y+month+day+FFMC+DMC+DC+ISI+temp+RH+wind+rain,data=forest\_train)  
  
summary(mod2)

##   
## Call:  
## lm(formula = area ~ X + Y + month + day + FFMC + DMC + DC + ISI +   
## temp + RH + wind + rain, data = forest\_train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -83.22 -28.76 -13.18 3.00 1033.49   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -58.52396 246.25353 -0.238 0.812  
## X 3.69189 3.24169 1.139 0.256  
## Y -0.32029 6.54953 -0.049 0.961  
## month 2.17683 1.94818 1.117 0.265  
## day 1.11464 3.54975 0.314 0.754  
## FFMC 0.59488 2.77114 0.215 0.830  
## DMC 0.25737 0.17338 1.484 0.139  
## DC -0.04497 0.04861 -0.925 0.356  
## ISI -1.23727 2.52295 -0.490 0.624  
## temp 1.20394 1.90725 0.631 0.529  
## RH -0.78123 0.58672 -1.332 0.185  
## wind 2.87658 3.96076 0.726 0.469  
## rain -3.18810 15.18015 -0.210 0.834  
##   
## Residual standard error: 95.78 on 199 degrees of freedom  
## Multiple R-squared: 0.04628, Adjusted R-squared: -0.01123   
## F-statistic: 0.8047 on 12 and 199 DF, p-value: 0.6454

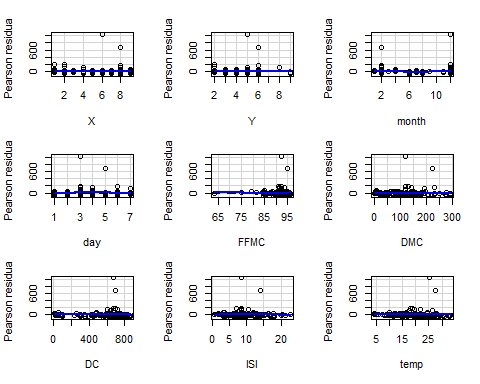
R sq has sugnificantly improved. Lets see plots

plot(mod2)

Chart, histogram

Description automatically generated

residualPlots(mod2)

Chart, box and whisker chart

Description automatically generated

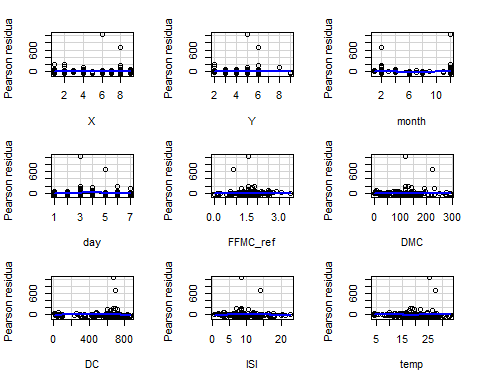
## Test stat Pr(>|Test stat|)   
## X 0.3235 0.746650   
## Y 0.3888 0.697817   
## month 2.1078 0.036304 \*   
## day -1.3355 0.183243   
## FFMC 0.6080 0.543859   
## DMC -0.3786 0.705392   
## DC -0.5138 0.607960   
## ISI -0.4403 0.660206   
## temp 0.6208 0.535472   
## RH 0.9754 0.330567   
## wind -0.9391 0.348802   
## rain 0.0414 0.967023   
## Tukey test 2.7417 0.006112 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

the fitted values plot and qq plot has improved, now we proceed with other transformations ## Model 3 As discussed earlier we transform FFMC and Rain

FFMC\_ref<- (log(max(forest\_train$FFMC)+1-forest\_train$FFMC))  
  
Rain\_log <- log(forest\_train$rain+1)  
  
mod3 <- lm(area~X+Y+month+day+FFMC\_ref+DMC+DC+ISI+temp+RH+wind+Rain\_log,data=forest\_train)  
  
summary(mod3)

##   
## Call:  
## lm(formula = area ~ X + Y + month + day + FFMC\_ref + DMC + DC +   
## ISI + temp + RH + wind + Rain\_log, data = forest\_train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -79.10 -28.60 -13.62 3.50 1033.08   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)  
## (Intercept) 24.38389 86.28255 0.283 0.778  
## X 3.80400 3.24470 1.172 0.242  
## Y -0.33220 6.54581 -0.051 0.960  
## month 2.23812 1.93521 1.157 0.249  
## day 1.00685 3.54383 0.284 0.777  
## FFMC\_ref -12.76884 23.51105 -0.543 0.588  
## DMC 0.25057 0.17105 1.465 0.145  
## DC -0.04443 0.04858 -0.915 0.361  
## ISI -1.97051 2.78422 -0.708 0.480  
## temp 1.04904 1.93652 0.542 0.589  
## RH -0.74661 0.58772 -1.270 0.205  
## wind 3.12336 3.97046 0.787 0.432  
## Rain\_log -15.18413 47.07637 -0.323 0.747  
##   
## Residual standard error: 95.72 on 199 degrees of freedom  
## Multiple R-squared: 0.04747, Adjusted R-squared: -0.009965   
## F-statistic: 0.8265 on 12 and 199 DF, p-value: 0.623

residualPlots(mod3)

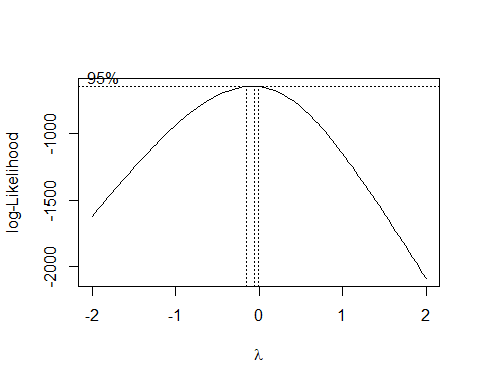
Chart, box and whisker chart

Description automatically generated

## Test stat Pr(>|Test stat|)   
## X 0.3959 0.692637   
## Y 0.4190 0.675700   
## month 2.0579 0.040914 \*   
## day -1.3643 0.174026   
## FFMC\_ref -0.0546 0.956538   
## DMC -0.3432 0.731822   
## DC -0.4970 0.619711   
## ISI -0.3293 0.742286   
## temp 0.5175 0.605351   
## RH 0.9289 0.354050   
## wind -0.8988 0.369872   
## Rain\_log -0.0463 0.963142   
## Tukey test 2.9219 0.003479 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

the model is improved on R sq and residuals also.DMC and DC have emerged as significant. We still need to improve patterns in fitted values and residuals. ## Model 4 we check the box cox for response variable transform

bc<- boxcox(mod3)



bc\_df = as.data.frame(bc)  
optimal\_lambda = bc\_df[which.max(bc$y),1]  
optimal\_lambda

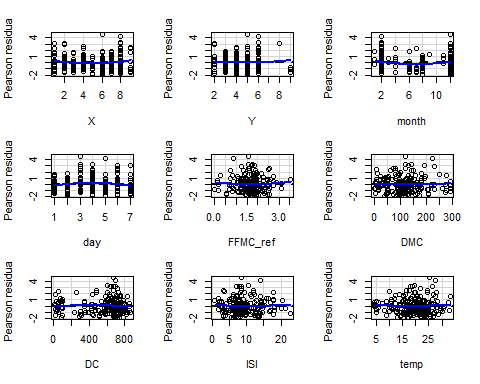
## [1] -0.06060606

the optimal lambda is very near to zero. Hence log transform is suitable here (with area+1 to counter the zero values)

mod4 <- lm(log(area+1)~X+Y+month+day+FFMC\_ref+DMC+DC+ISI+temp+RH+wind+Rain\_log,data=forest\_train)  
  
summary(mod4)

##   
## Call:  
## lm(formula = log(area + 1) ~ X + Y + month + day + FFMC\_ref +   
## DMC + DC + ISI + temp + RH + wind + Rain\_log, data = forest\_train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.8848 -0.8886 -0.1858 0.6444 4.5509   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.4829275 1.1198627 3.110 0.00214 \*\*  
## X 0.0427570 0.0421131 1.015 0.31120   
## Y -0.0551835 0.0849582 -0.650 0.51674   
## month 0.0494545 0.0251172 1.969 0.05035 .   
## day 0.0044292 0.0459954 0.096 0.92338   
## FFMC\_ref -0.3272748 0.3051504 -1.073 0.28479   
## DMC 0.0049332 0.0022200 2.222 0.02740 \*   
## DC -0.0010607 0.0006305 -1.682 0.09407 .   
## ISI -0.0447965 0.0361365 -1.240 0.21657   
## temp -0.0178561 0.0251342 -0.710 0.47827   
## RH -0.0142060 0.0076280 -1.862 0.06403 .   
## wind 0.0539502 0.0515327 1.047 0.29641   
## Rain\_log 0.0844517 0.6110051 0.138 0.89021   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.242 on 199 degrees of freedom  
## Multiple R-squared: 0.06852, Adjusted R-squared: 0.01235   
## F-statistic: 1.22 on 12 and 199 DF, p-value: 0.2714

residualPlots(mod4)



Diagram

Description automatically generated with medium confidence

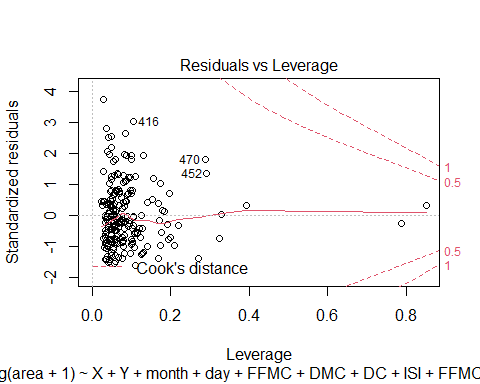
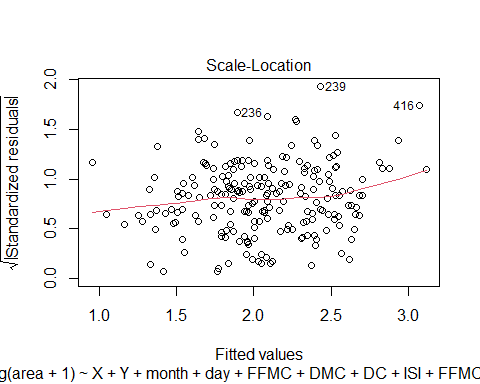
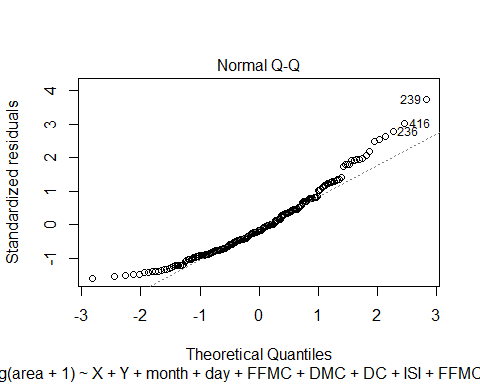
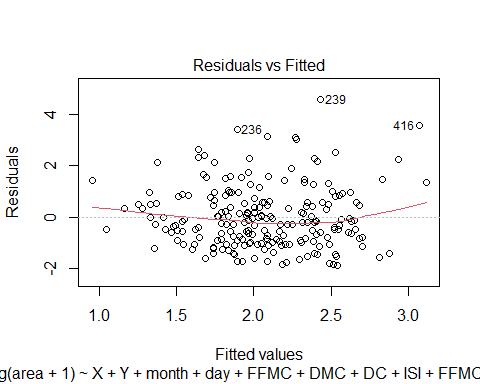
## Test stat Pr(>|Test stat|)   
## X 1.7881 0.07528 .  
## Y 0.7439 0.45784   
## month 2.5914 0.01027 \*  
## day -2.1069 0.03638 \*  
## FFMC\_ref 0.6942 0.48840   
## DMC 0.4915 0.62360   
## DC -0.5915 0.55487   
## ISI 0.2199 0.82616   
## temp 0.5943 0.55299   
## RH 1.0216 0.30824   
## wind -0.9769 0.32982   
## Rain\_log 0.5731 0.56721   
## Tukey test 1.1705 0.24181   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

The fitted values vs residual is random now and model prediction is also improved. Lets transform other variables. ## MOdel 5 Here we check the interaction of the the various forest Fire Weather Index (FWI) as they are closely related and may have larger inetraction impact.

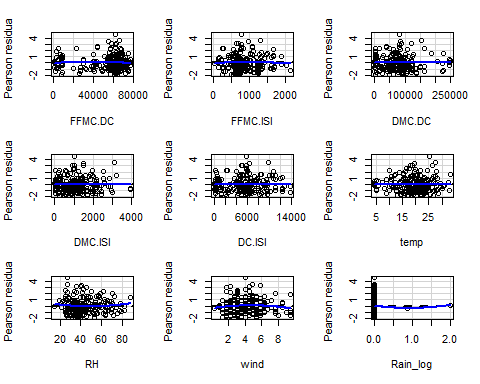
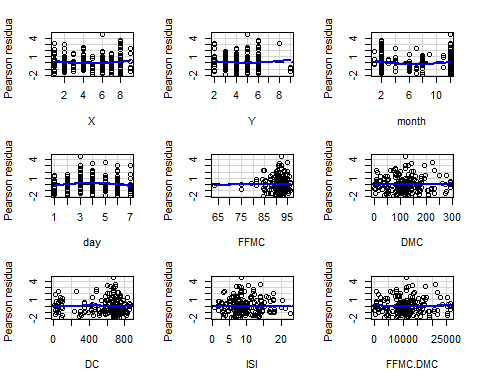
FFMC.DMC <- forest\_train$FFMC\*forest\_train$DMC  
FFMC.DC <-forest\_train$FFMC\*forest\_train$DC  
FFMC.ISI <-forest\_train$FFMC\*forest\_train$ISI  
DMC.DC<-forest\_train$DMC\*forest\_train$DC  
DMC.ISI<-forest\_train$DMC\*forest\_train$ISI  
DC.ISI<-forest\_train$DC\*forest\_train$ISI  
  
  
mod5 <- lm(log(area+1)~X+Y+month+day+FFMC+DMC+DC+ISI+FFMC.DMC+FFMC.DC+FFMC.ISI+DMC.DC+DMC.ISI+DC.ISI+  
 temp+RH+wind+Rain\_log,data=forest\_train)  
  
summary(mod5)

##   
## Call:  
## lm(formula = log(area + 1) ~ X + Y + month + day + FFMC + DMC +   
## DC + ISI + FFMC.DMC + FFMC.DC + FFMC.ISI + DMC.DC + DMC.ISI +   
## DC.ISI + temp + RH + wind + Rain\_log, data = forest\_train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.8650 -0.8849 -0.1964 0.5952 4.5639   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.697e+00 7.072e+00 0.381 0.7033   
## X 4.740e-02 4.235e-02 1.119 0.2644   
## Y -6.402e-02 8.588e-02 -0.745 0.4569   
## month 7.063e-02 2.775e-02 2.546 0.0117 \*  
## day 1.659e-03 4.715e-02 0.035 0.9720   
## FFMC -1.785e-03 8.503e-02 -0.021 0.9833   
## DMC -2.133e-01 1.079e-01 -1.977 0.0495 \*  
## DC 2.534e-02 1.756e-02 1.443 0.1507   
## ISI 9.808e-02 8.759e-01 0.112 0.9110   
## FFMC.DMC 2.445e-03 1.212e-03 2.018 0.0450 \*  
## FFMC.DC -2.913e-04 2.069e-04 -1.408 0.1607   
## FFMC.ISI -1.203e-03 9.273e-03 -0.130 0.8969   
## DMC.DC -2.203e-06 1.120e-05 -0.197 0.8443   
## DMC.ISI -2.140e-04 7.613e-04 -0.281 0.7789   
## DC.ISI -1.562e-06 2.148e-04 -0.007 0.9942   
## temp -2.790e-02 2.774e-02 -1.006 0.3158   
## RH -1.328e-02 7.801e-03 -1.702 0.0904 .  
## wind 4.616e-02 5.313e-02 0.869 0.3861   
## Rain\_log -3.147e-02 6.073e-01 -0.052 0.9587   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.24 on 193 degrees of freedom  
## Multiple R-squared: 0.1005, Adjusted R-squared: 0.01664   
## F-statistic: 1.198 on 18 and 193 DF, p-value: 0.2655

plot(mod5)



residualPlots(mod5)

Chart

Description automatically generated with medium confidence

## Test stat Pr(>|Test stat|)   
## X 1.8407 0.067209 .   
## Y 0.9667 0.334929   
## month 2.6880 0.007819 \*\*  
## day -2.1382 0.033766 \*   
## FFMC -0.4308 0.667072   
## DMC 0.8550 0.393638   
## DC -0.6572 0.511817   
## ISI -0.3008 0.763884   
## FFMC.DMC 0.9063 0.365926   
## FFMC.DC -0.6239 0.533414   
## FFMC.ISI -0.3138 0.754025   
## DMC.DC -0.0738 0.941263   
## DMC.ISI 0.1813 0.856315   
## DC.ISI -0.3342 0.738569   
## temp 0.3603 0.719003   
## RH 1.0063 0.315524   
## wind -1.2885 0.199106   
## Rain\_log 0.3123 0.755182   
## Tukey test 2.3464 0.018956 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

The model has even improved and the normal quantile plot of residuals is much better normal now. ## Model 6 Lets try to do the residual improvement of other variables. Hence we try the square of variables which are dense distributed and show some quadratic pattern.Also factors like wind, temp should have greater impact on fire spread and area

wind\_sq<-(forest\_train$wind)^2  
temp\_sq<-(forest\_train$temp)^2  
rain\_sq<-(forest\_train$rain)^2  
RH\_sq<-(forest\_train$RH)^2

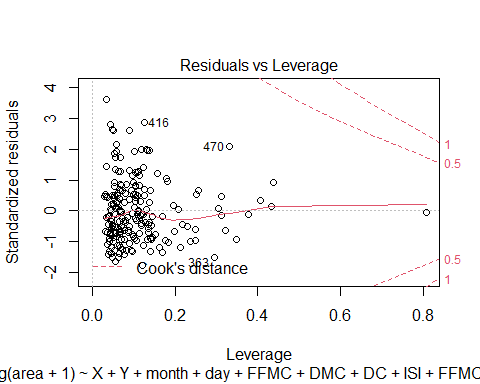
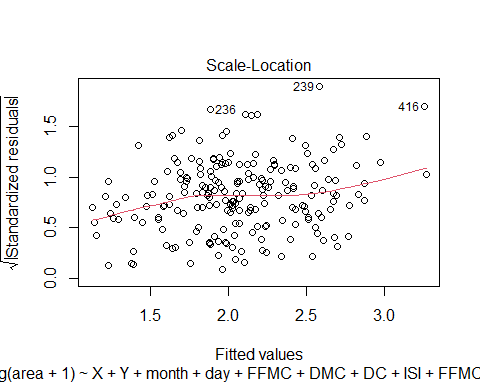
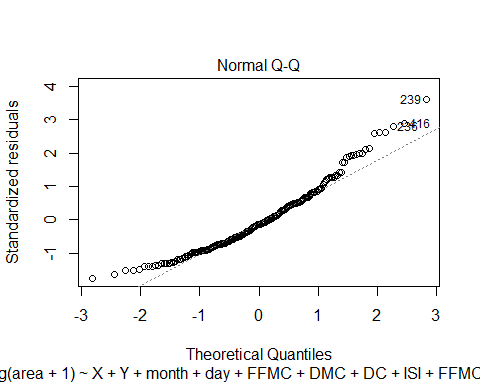
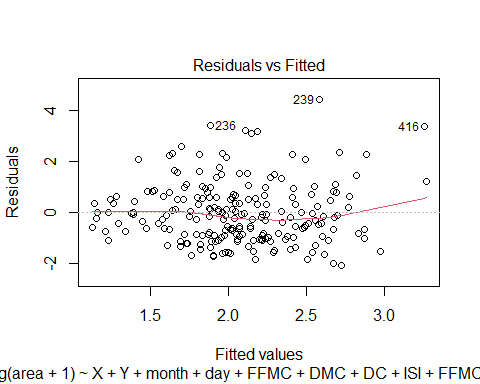
We check the interaction also of these factors.

temp.RH<-(forest\_train$temp)\*(forest\_train$RH)  
wind.rain<-(forest\_train$wind)\*(forest\_train$rain)  
wind.temp<-(forest\_train$wind)\*(forest\_train$temp)  
  
mod6 <- lm(log(area+1)~X+Y+month+day+FFMC+DMC+DC+ISI+FFMC.DMC+FFMC.DC+FFMC.ISI+DMC.DC+DMC.ISI+DC.ISI+  
 temp+temp\_sq+RH+RH\_sq+wind+wind\_sq+Rain\_log+rain\_sq+temp.RH+wind.rain+wind.temp,data=forest\_train)  
  
summary(mod6)

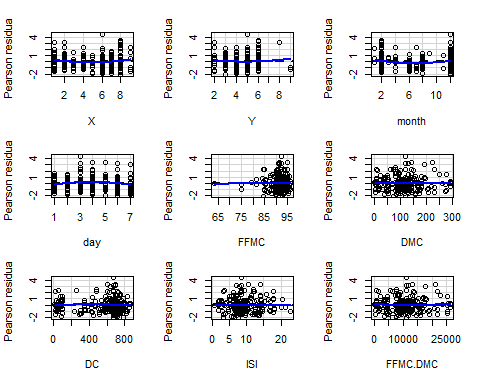
##   
## Call:  
## lm(formula = log(area + 1) ~ X + Y + month + day + FFMC + DMC +   
## DC + ISI + FFMC.DMC + FFMC.DC + FFMC.ISI + DMC.DC + DMC.ISI +   
## DC.ISI + temp + temp\_sq + RH + RH\_sq + wind + wind\_sq + Rain\_log +   
## rain\_sq + temp.RH + wind.rain + wind.temp, data = forest\_train)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.0506 -0.8871 -0.1488 0.6134 4.4151   
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.870e+00 9.110e+00 0.425 0.6714   
## X 4.787e-02 4.348e-02 1.101 0.2723   
## Y -7.749e-02 8.735e-02 -0.887 0.3762   
## month 6.204e-02 2.993e-02 2.073 0.0395 \*  
## day -8.240e-04 4.822e-02 -0.017 0.9864   
## FFMC 1.264e-03 9.948e-02 0.013 0.9899   
## DMC -2.324e-01 1.103e-01 -2.108 0.0364 \*  
## DC 2.700e-02 1.960e-02 1.378 0.1699   
## ISI 4.093e-01 9.143e-01 0.448 0.6549   
## FFMC.DMC 2.640e-03 1.238e-03 2.133 0.0342 \*  
## FFMC.DC -3.046e-04 2.281e-04 -1.335 0.1835   
## FFMC.ISI -4.223e-03 9.630e-03 -0.438 0.6615   
## DMC.DC -6.536e-07 1.145e-05 -0.057 0.9546   
## DMC.ISI -2.212e-04 7.827e-04 -0.283 0.7778   
## DC.ISI -5.900e-05 2.259e-04 -0.261 0.7942   
## temp -9.108e-02 1.780e-01 -0.512 0.6096   
## temp\_sq 7.683e-04 3.263e-03 0.235 0.8141   
## RH -6.152e-02 5.922e-02 -1.039 0.3002   
## RH\_sq 4.332e-04 4.357e-04 0.994 0.3213   
## wind 1.553e-01 2.868e-01 0.541 0.5889   
## wind\_sq -2.277e-02 2.104e-02 -1.082 0.2804   
## Rain\_log -5.749e-01 1.739e+00 -0.331 0.7413   
## rain\_sq 2.816e-02 9.029e-02 0.312 0.7554   
## temp.RH 2.550e-04 1.634e-03 0.156 0.8762   
## wind.rain NA NA NA NA   
## wind.temp 4.978e-03 9.299e-03 0.535 0.5931   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.249 on 187 degrees of freedom  
## Multiple R-squared: 0.115, Adjusted R-squared: 0.001377   
## F-statistic: 1.012 on 24 and 187 DF, p-value: 0.4529

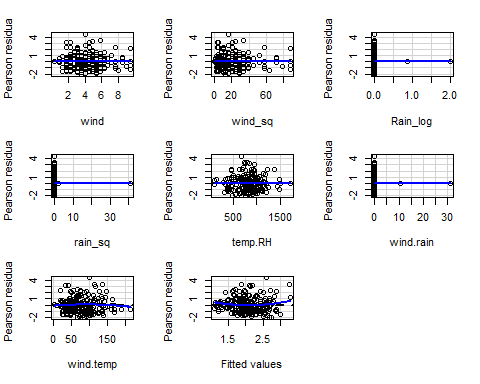
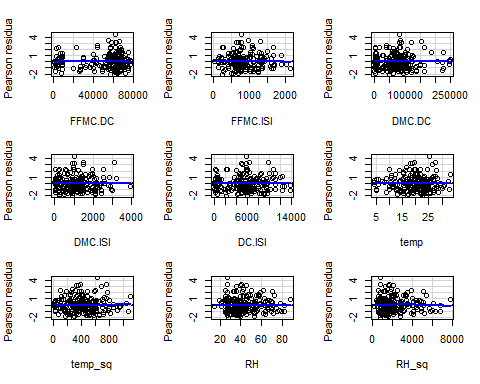
plot(mod6)

## Warning: not plotting observations with leverage one:  
## 207, 210



residualPlots(mod6)





## Test stat Pr(>|Test stat|)   
## X 1.7774 0.077143 .   
## Y 0.8977 0.370510   
## month 2.8773 0.004481 \*\*  
## day -1.9943 0.047580 \*   
## FFMC -0.4578 0.647655   
## DMC 0.6493 0.516923   
## DC -0.8777 0.381242   
## ISI -0.2154 0.829673   
## FFMC.DMC 0.7064 0.480801   
## FFMC.DC -0.7927 0.428964   
## FFMC.ISI -0.2217 0.824817   
## DMC.DC -0.2657 0.790768   
## DMC.ISI 0.1439 0.885719   
## DC.ISI -0.3186 0.750424   
## temp 0.8502 0.396313   
## temp\_sq 0.4072 0.684312   
## RH -0.9035 0.367409   
## RH\_sq -1.1968 0.232907   
## wind 0.8398 0.402082   
## wind\_sq 0.3401 0.734170   
## Rain\_log -0.6821 0.496048   
## rain\_sq -1.3349 0.183542   
## temp.RH -0.3610 0.718527   
## wind.rain -1.1556 0.249341   
## wind.temp -1.2254 0.221979   
## Tukey test 1.6359 0.101862   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

the model is improved in terms of r sq and the residuals are also randomly distributed. We are very near to optimal model. we do some formal checks of collienearity and influential observations(as there were outliers in dataset)

## checking collinearity

colldiag <- function(mod,scale=TRUE,center=FALSE,add.intercept=TRUE) {  
 result <- NULL  
 if (center) add.intercept<-FALSE  
 if (is.matrix(mod)||is.data.frame(mod)) {  
 X<-as.matrix(mod)  
 nms<-colnames(mod)  
 }  
 else if (!is.null(mod$call$formula)) {  
 X<-mod$model[,-1] # delete the dependent variable  
 }  
 X<-na.omit(X) # delete missing cases  
 if (add.intercept) {  
 X<-cbind(1,X) # add the intercept  
 colnames(X)[1]<-"intercept"  
 }  
 X<-scale(X,scale=scale,center=center)  
   
 svdX<-svd(X)  
 svdX$d  
 condindx<-svdX$d[1]/svdX$d  
   
 Phi=svdX$v%\*%diag(1/svdX$d)  
 Phi<-t(Phi^2)  
 pi<-prop.table(Phi,2)  
   
 dim(condindx)<-c(length(condindx),1)  
 colnames(condindx)<-"cond.index"  
 rownames(condindx)<-1:nrow(condindx)  
 colnames(pi)<-colnames(X)  
 result$condindx<-condindx  
 result$pi<-pi  
 class(result)<-"colldiag"  
 result  
}  
  
print.colldiag <- function(x,dec.places=3,fuzz=NULL,fuzzchar=".",...){  
 stopifnot(fuzz>0 & fuzz<1)  
 stopifnot(is.character(fuzzchar))  
 stopifnot(nchar(fuzzchar)==1)  
 fuzzchar<-paste(" ",fuzzchar,sep="")  
 width<-dec.places+2  
 pi<-formatC(x$pi,format="f",width=width,digits=dec.places)  
 if (!is.null(fuzz )) {  
 pi[pi < fuzz] <- fuzzchar  
 }  
 width<-max(nchar(trunc(max(x$condindx))))+dec.places+2  
 condindx<-formatC(x$condindx,format="f",width=width,digits=dec.places)  
 colnames(condindx)<-NULL  
 cat("Condition\nIndex\tVariance Decomposition Proportions\n")  
 print(noquote(cbind(condindx,pi)))  
}

colldiag(forest\_train[,-c(3,4,13)])

## Condition  
## Index Variance Decomposition Proportions  
## intercept X Y FFMC DMC DC ISI temp RH wind   
## 1 1.000 0.000 0.001 0.001 0.000 0.001 0.001 0.001 0.000 0.001 0.001  
## 2 3.018 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000 0.000  
## 3 5.043 0.000 0.097 0.005 0.000 0.074 0.032 0.010 0.005 0.010 0.043  
## 4 6.696 0.000 0.215 0.015 0.000 0.009 0.006 0.035 0.001 0.002 0.330  
## 5 7.615 0.000 0.041 0.002 0.000 0.062 0.020 0.177 0.017 0.126 0.004  
## 6 9.519 0.000 0.161 0.001 0.000 0.209 0.002 0.028 0.004 0.130 0.392  
## 7 10.762 0.000 0.113 0.032 0.000 0.154 0.084 0.249 0.047 0.124 0.058  
## 8 12.992 0.000 0.237 0.251 0.000 0.258 0.584 0.016 0.009 0.004 0.005  
## 9 16.241 0.001 0.130 0.623 0.001 0.048 0.264 0.105 0.225 0.004 0.000  
## 10 24.798 0.012 0.005 0.068 0.012 0.130 0.005 0.011 0.687 0.526 0.118  
## 11 159.043 0.987 0.000 0.003 0.987 0.054 0.003 0.367 0.004 0.073 0.049  
## rain   
## 1 0.000  
## 2 0.938  
## 3 0.000  
## 4 0.000  
## 5 0.000  
## 6 0.001  
## 7 0.012  
## 8 0.000  
## 9 0.002  
## 10 0.035  
## 11 0.010

We find that there is no collinearity in the variables , so we are safe on this front. ## checking influential obs values

forest\_train[which(row.names(forest\_train) %in% c(200,363,416,479,480)),]

## X Y month day FFMC DMC DC ISI temp RH wind rain area  
## 200 2 4 12 2 63.5 70.8 665.3 0.8 22.6 38 3.6 0 11.32  
## 363 7 4 12 1 88.2 55.2 732.3 11.6 15.2 64 3.1 0 0.52  
## 416 8 6 2 5 94.8 222.4 698.6 13.9 27.5 27 4.9 0 746.28  
## 479 7 4 6 4 93.7 101.3 423.4 14.7 18.2 82 4.5 0 2.21

out f these only two are high outliers 416 and and 480. These seem to be burned due to some other factors, may be intentional !! So we remove two observations.

forest\_train\_new <- forest\_train[which(!row.names(forest\_train) %in% c(416,480)),]

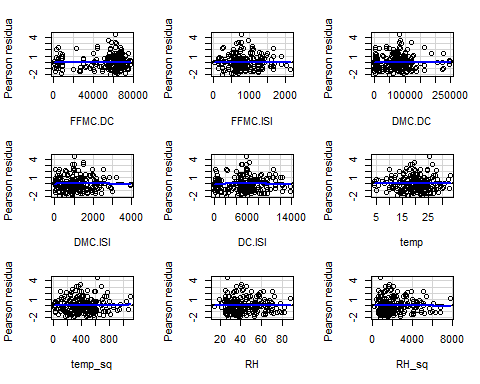
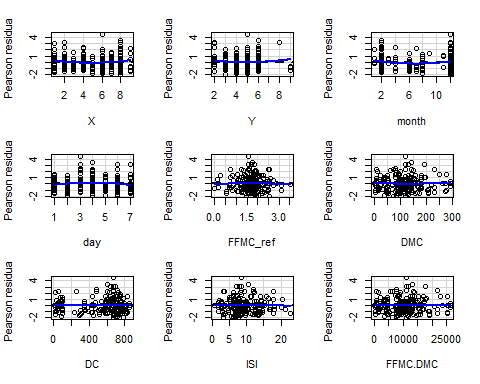
## Model 7 with removed influential obs

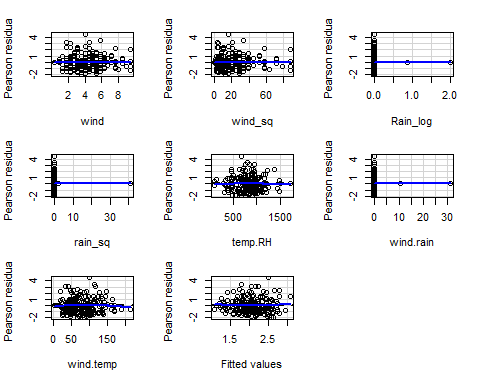
defining transform variables again with new dataset

wind\_sq<-(forest\_train\_new$wind)^2  
temp\_sq<-(forest\_train\_new$temp)^2  
rain\_sq<-(forest\_train\_new$rain)^2  
RH\_sq<-(forest\_train\_new$RH)^2  
Rain\_log <- log(forest\_train\_new$rain+1)  
FFMC\_ref<- (log(max(forest\_train\_new$FFMC)+1-forest\_train\_new$FFMC))  
  
temp.RH<-(forest\_train\_new$temp)\*(forest\_train\_new$RH)  
wind.rain<-(forest\_train\_new$wind)\*(forest\_train\_new$rain)  
wind.temp<-(forest\_train\_new$wind)\*(forest\_train\_new$temp)  
  
FFMC.DMC <- forest\_train\_new$FFMC\*forest\_train\_new$DMC  
FFMC.DC <-forest\_train\_new$FFMC\*forest\_train\_new$DC  
FFMC.ISI <-forest\_train\_new$FFMC\*forest\_train\_new$ISI  
DMC.DC<-forest\_train\_new$DMC\*forest\_train\_new$DC  
DMC.ISI<-forest\_train\_new$DMC\*forest\_train\_new$ISI  
DC.ISI<-forest\_train\_new$DC\*forest\_train\_new$ISI  
  
mod7 <- lm(log(area+1)~X+Y+month+day+FFMC\_ref+DMC+DC+ISI+FFMC.DMC+FFMC.DC+FFMC.ISI+DMC.DC+DMC.ISI+DC.ISI+  
 temp+temp\_sq+RH+RH\_sq+wind+wind\_sq+Rain\_log+rain\_sq+temp.RH+wind.rain+wind.temp,data=forest\_train\_new)  
  
  
summary(mod7)

##   
## Call:  
## lm(formula = log(area + 1) ~ X + Y + month + day + FFMC\_ref +   
## DMC + DC + ISI + FFMC.DMC + FFMC.DC + FFMC.ISI + DMC.DC +   
## DMC.ISI + DC.ISI + temp + temp\_sq + RH + RH\_sq + wind + wind\_sq +   
## Rain\_log + rain\_sq + temp.RH + wind.rain + wind.temp, data = forest\_train\_new)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -1.9655 -0.8722 -0.1136 0.6263 4.4943   
##   
## Coefficients: (1 not defined because of singularities)  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 2.491e+00 3.325e+00 0.749 0.4547   
## X 3.169e-02 4.304e-02 0.736 0.4626   
## Y -7.462e-02 8.565e-02 -0.871 0.3848   
## month 5.937e-02 2.910e-02 2.040 0.0427 \*  
## day -5.795e-03 4.708e-02 -0.123 0.9022   
## FFMC\_ref 5.176e-01 8.209e-01 0.631 0.5291   
## DMC -2.060e-01 1.128e-01 -1.826 0.0694 .  
## DC 2.100e-02 1.444e-02 1.455 0.1475   
## ISI -3.391e-01 1.298e+00 -0.261 0.7942   
## FFMC.DMC 2.386e-03 1.272e-03 1.876 0.0622 .  
## FFMC.DC -2.346e-04 1.672e-04 -1.403 0.1622   
## FFMC.ISI 4.444e-03 1.432e-02 0.310 0.7567   
## DMC.DC -2.994e-06 1.105e-05 -0.271 0.7867   
## DMC.ISI -5.125e-04 7.865e-04 -0.652 0.5155   
## DC.ISI -8.074e-05 2.079e-04 -0.388 0.6981   
## temp -8.679e-02 1.746e-01 -0.497 0.6196   
## temp\_sq 8.484e-04 3.180e-03 0.267 0.7900   
## RH -5.757e-02 5.573e-02 -1.033 0.3029   
## RH\_sq 3.879e-04 4.241e-04 0.915 0.3615   
## wind 1.964e-01 2.836e-01 0.692 0.4895   
## wind\_sq -2.301e-02 2.074e-02 -1.110 0.2685   
## Rain\_log -6.258e-01 1.702e+00 -0.368 0.7135   
## rain\_sq 4.069e-02 8.886e-02 0.458 0.6476   
## temp.RH 4.141e-04 1.475e-03 0.281 0.7792   
## wind.rain NA NA NA NA   
## wind.temp 2.268e-03 9.161e-03 0.248 0.8048   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.223 on 186 degrees of freedom  
## Multiple R-squared: 0.09879, Adjusted R-squared: -0.0175   
## F-statistic: 0.8495 on 24 and 186 DF, p-value: 0.6699

residualPlots(mod7)

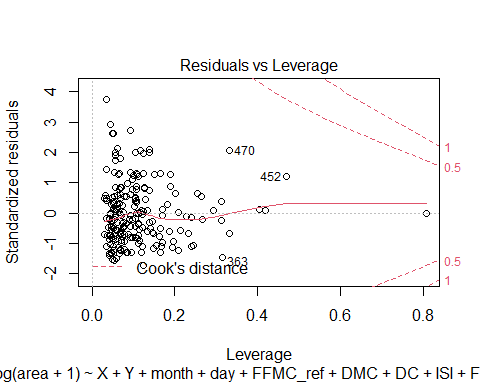
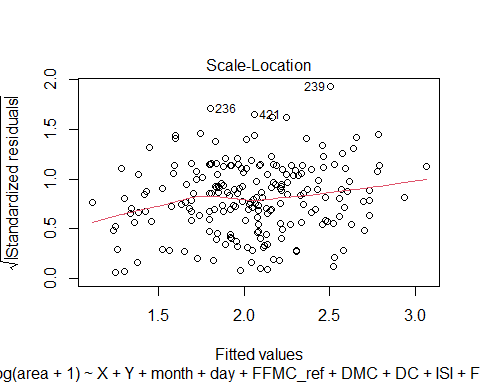
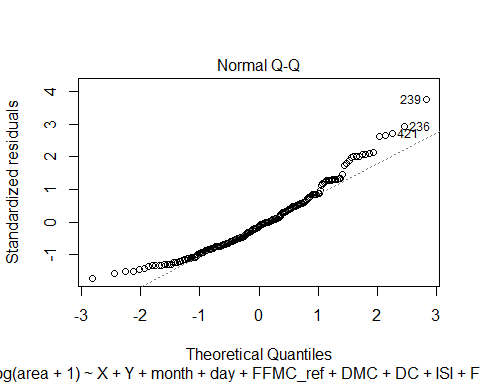
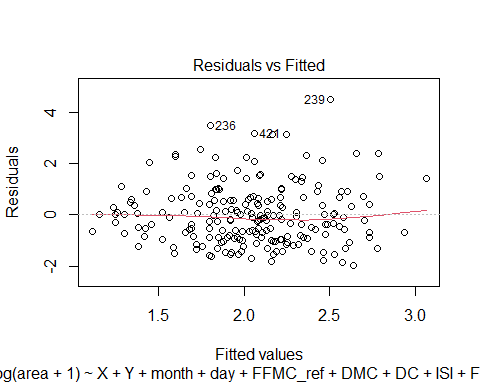




## Test stat Pr(>|Test stat|)   
## X 1.5668 0.118867   
## Y 0.9188 0.359384   
## month 2.9026 0.004151 \*\*  
## day -1.7574 0.080504 .   
## FFMC\_ref -0.1070 0.914920   
## DMC 0.4751 0.635311   
## DC -0.7518 0.453113   
## ISI -0.4125 0.680468   
## FFMC.DMC 0.4982 0.618906   
## FFMC.DC -0.6828 0.495564   
## FFMC.ISI -0.4169 0.677229   
## DMC.DC -0.3547 0.723222   
## DMC.ISI -0.0888 0.929322   
## DC.ISI -0.3874 0.698884   
## temp 0.2162 0.829035   
## temp\_sq 0.8196 0.413500   
## RH 0.0242 0.980687   
## RH\_sq -1.0645 0.288468   
## wind -0.6035 0.546923   
## wind\_sq 0.2932 0.769675   
## Rain\_log 0.5448 0.586579   
## rain\_sq -0.2745 0.783988   
## temp.RH -0.4800 0.631780   
## wind.rain -0.5577 0.577754   
## wind.temp -1.3123 0.191032   
## Tukey test 0.7137 0.475440   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

plot(mod7)

## Warning: not plotting observations with leverage one:  
## 206, 209

 There seem to be no deviation but , the model has too many insignificant regressors. Also R sq and adjusted R sq are very far apart. Lets run a step AIC to remove the insignificant variables. ## Model8

stepAIC(mod7)

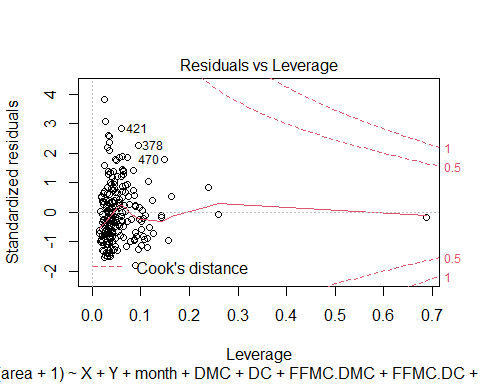
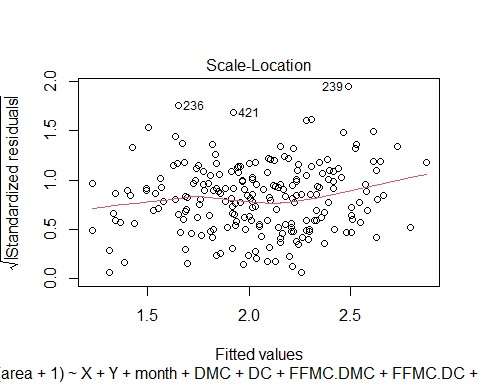
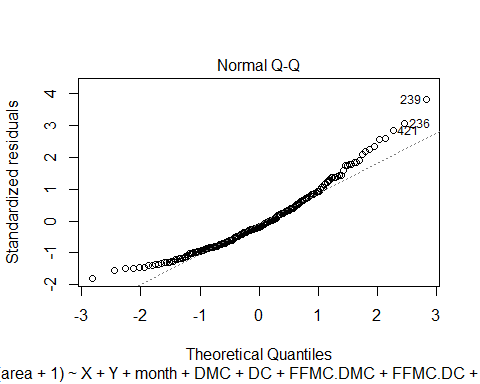
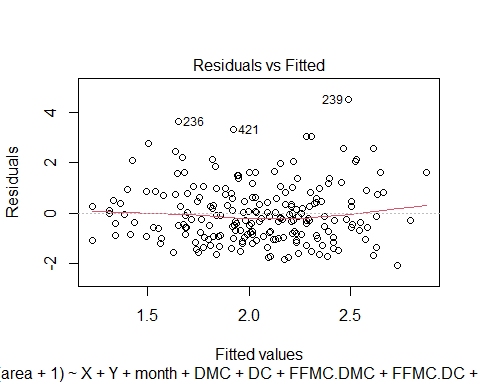
## Start: AIC=108.44  
## log(area + 1) ~ X + Y + month + day + FFMC\_ref + DMC + DC + ISI +   
## FFMC.DMC + FFMC.DC + FFMC.ISI + DMC.DC + DMC.ISI + DC.ISI +   
## temp + temp\_sq + RH + RH\_sq + wind + wind\_sq + Rain\_log +   
## rain\_sq + temp.RH + wind.rain + wind.temp  
##   
##   
## Step: AIC=108.44  
## log(area + 1) ~ X + Y + month + day + FFMC\_ref + DMC + DC + ISI +   
## FFMC.DMC + FFMC.DC + FFMC.ISI + DMC.DC + DMC.ISI + DC.ISI +   
## temp + temp\_sq + RH + RH\_sq + wind + wind\_sq + Rain\_log +   
## rain\_sq + temp.RH + wind.temp  
##   
## Df Sum of Sq RSS AIC  
## - day 1 0.0227 278.36 106.46  
## - wind.temp 1 0.0917 278.43 106.51  
## - ISI 1 0.1022 278.44 106.52  
## - temp\_sq 1 0.1065 278.44 106.52  
## - DMC.DC 1 0.1099 278.45 106.53  
## - temp.RH 1 0.1180 278.45 106.53  
## - FFMC.ISI 1 0.1440 278.48 106.55  
## - Rain\_log 1 0.2023 278.54 106.59  
## - DC.ISI 1 0.2258 278.56 106.61  
## - rain\_sq 1 0.3138 278.65 106.68  
## - temp 1 0.3699 278.71 106.72  
## - FFMC\_ref 1 0.5950 278.93 106.89  
## - DMC.ISI 1 0.6353 278.97 106.92  
## - wind 1 0.7176 279.06 106.98  
## - X 1 0.8110 279.15 107.06  
## - Y 1 1.1356 279.47 107.30  
## - RH\_sq 1 1.2520 279.59 107.39  
## - RH 1 1.5971 279.94 107.65  
## - wind\_sq 1 1.8429 280.18 107.83  
## <none> 278.34 108.44  
## - FFMC.DC 1 2.9473 281.29 108.66  
## - DC 1 3.1658 281.50 108.83  
## - DMC 1 4.9913 283.33 110.19  
## - FFMC.DMC 1 5.2657 283.60 110.40  
## - month 1 6.2296 284.57 111.11  
##   
## Step: AIC=106.46  
## log(area + 1) ~ X + Y + month + FFMC\_ref + DMC + DC + ISI + FFMC.DMC +   
## FFMC.DC + FFMC.ISI + DMC.DC + DMC.ISI + DC.ISI + temp + temp\_sq +   
## RH + RH\_sq + wind + wind\_sq + Rain\_log + rain\_sq + temp.RH +   
## wind.temp  
##   
## Df Sum of Sq RSS AIC  
## - ISI 1 0.0881 278.45 104.53  
## - wind.temp 1 0.0899 278.45 104.53  
## - DMC.DC 1 0.1091 278.47 104.54  
## - temp\_sq 1 0.1158 278.48 104.55  
## - temp.RH 1 0.1209 278.48 104.55  
## - FFMC.ISI 1 0.1283 278.49 104.56  
## - Rain\_log 1 0.1910 278.55 104.60  
## - DC.ISI 1 0.2135 278.57 104.62  
## - rain\_sq 1 0.2974 278.66 104.68  
## - temp 1 0.3832 278.74 104.75  
## - FFMC\_ref 1 0.5755 278.94 104.89  
## - DMC.ISI 1 0.6773 279.04 104.97  
## - wind 1 0.7229 279.08 105.01  
## - X 1 0.8275 279.19 105.08  
## - Y 1 1.1749 279.54 105.35  
## - RH\_sq 1 1.2442 279.60 105.40  
## - RH 1 1.5912 279.95 105.66  
## - wind\_sq 1 1.8392 280.20 105.85  
## <none> 278.36 106.46  
## - FFMC.DC 1 3.0721 281.43 106.78  
## - DC 1 3.2830 281.64 106.93  
## - DMC 1 5.0564 283.42 108.26  
## - FFMC.DMC 1 5.3430 283.70 108.47  
## - month 1 6.5323 284.89 109.35  
##   
## Step: AIC=104.53  
## log(area + 1) ~ X + Y + month + FFMC\_ref + DMC + DC + FFMC.DMC +   
## FFMC.DC + FFMC.ISI + DMC.DC + DMC.ISI + DC.ISI + temp + temp\_sq +   
## RH + RH\_sq + wind + wind\_sq + Rain\_log + rain\_sq + temp.RH +   
## wind.temp  
##   
## Df Sum of Sq RSS AIC  
## - DMC.DC 1 0.0844 278.53 102.59  
## - wind.temp 1 0.1142 278.56 102.61  
## - temp.RH 1 0.1269 278.57 102.62  
## - temp\_sq 1 0.1521 278.60 102.64  
## - DC.ISI 1 0.1715 278.62 102.66  
## - Rain\_log 1 0.1919 278.64 102.67  
## - rain\_sq 1 0.2825 278.73 102.74  
## - temp 1 0.4439 278.89 102.86  
## - FFMC.ISI 1 0.5133 278.96 102.91  
## - FFMC\_ref 1 0.6150 279.06 102.99  
## - DMC.ISI 1 0.6540 279.10 103.02  
## - wind 1 0.6707 279.12 103.03  
## - X 1 0.8657 279.31 103.18  
## - Y 1 1.2240 279.67 103.45  
## - RH\_sq 1 1.3013 279.75 103.51  
## - RH 1 1.6549 280.10 103.78  
## - wind\_sq 1 1.7901 280.24 103.88  
## <none> 278.45 104.53  
## - FFMC.DC 1 3.9341 282.38 105.49  
## - DC 1 4.1527 282.60 105.65  
## - DMC 1 5.3783 283.83 106.56  
## - FFMC.DMC 1 5.6215 284.07 106.74  
## - month 1 6.4888 284.94 107.39  
##   
## Step: AIC=102.59  
## log(area + 1) ~ X + Y + month + FFMC\_ref + DMC + DC + FFMC.DMC +   
## FFMC.DC + FFMC.ISI + DMC.ISI + DC.ISI + temp + temp\_sq +   
## RH + RH\_sq + wind + wind\_sq + Rain\_log + rain\_sq + temp.RH +   
## wind.temp  
##   
## Df Sum of Sq RSS AIC  
## - wind.temp 1 0.1145 278.65 100.68  
## - temp\_sq 1 0.1195 278.65 100.68  
## - temp.RH 1 0.1516 278.68 100.70  
## - Rain\_log 1 0.2140 278.75 100.75  
## - DC.ISI 1 0.2397 278.77 100.77  
## - rain\_sq 1 0.2946 278.83 100.81  
## - temp 1 0.4000 278.93 100.89  
## - FFMC\_ref 1 0.6071 279.14 101.05  
## - FFMC.ISI 1 0.6199 279.15 101.06  
## - DMC.ISI 1 0.6203 279.15 101.06  
## - wind 1 0.6854 279.22 101.11  
## - X 1 0.8672 279.40 101.25  
## - Y 1 1.1830 279.72 101.48  
## - RH\_sq 1 1.4086 279.94 101.65  
## - RH 1 1.7765 280.31 101.93  
## - wind\_sq 1 1.8552 280.39 101.99  
## <none> 278.53 102.59  
## - FFMC.DC 1 3.9658 282.50 103.57  
## - DC 1 4.1729 282.71 103.73  
## - DMC 1 5.6540 284.19 104.83  
## - FFMC.DMC 1 5.7004 284.23 104.86  
## - month 1 6.5290 285.06 105.48  
##   
## Step: AIC=100.68  
## log(area + 1) ~ X + Y + month + FFMC\_ref + DMC + DC + FFMC.DMC +   
## FFMC.DC + FFMC.ISI + DMC.ISI + DC.ISI + temp + temp\_sq +   
## RH + RH\_sq + wind + wind\_sq + Rain\_log + rain\_sq + temp.RH  
##   
## Df Sum of Sq RSS AIC  
## - temp\_sq 1 0.0559 278.70 98.719  
## - temp.RH 1 0.0984 278.75 98.751  
## - Rain\_log 1 0.1750 278.82 98.809  
## - DC.ISI 1 0.2392 278.89 98.858  
## - rain\_sq 1 0.2659 278.91 98.878  
## - temp 1 0.2862 278.93 98.893  
## - FFMC\_ref 1 0.5949 279.24 99.127  
## - DMC.ISI 1 0.5990 279.25 99.130  
## - FFMC.ISI 1 0.6256 279.27 99.150  
## - X 1 0.7970 279.44 99.279  
## - Y 1 1.1434 279.79 99.541  
## - RH\_sq 1 1.3008 279.95 99.659  
## - RH 1 1.6649 280.31 99.934  
## - wind\_sq 1 2.1960 280.84 100.333  
## - wind 1 2.6369 281.28 100.664  
## <none> 278.65 100.677  
## - FFMC.DC 1 3.9479 282.60 101.645  
## - DC 1 4.1471 282.79 101.794  
## - DMC 1 5.6031 284.25 102.877  
## - FFMC.DMC 1 5.6476 284.30 102.910  
## - month 1 6.5498 285.20 103.579  
##   
## Step: AIC=98.72  
## log(area + 1) ~ X + Y + month + FFMC\_ref + DMC + DC + FFMC.DMC +   
## FFMC.DC + FFMC.ISI + DMC.ISI + DC.ISI + temp + RH + RH\_sq +   
## wind + wind\_sq + Rain\_log + rain\_sq + temp.RH  
##   
## Df Sum of Sq RSS AIC  
## - temp.RH 1 0.0561 278.76 96.761  
## - Rain\_log 1 0.1772 278.88 96.853  
## - DC.ISI 1 0.2007 278.90 96.871  
## - rain\_sq 1 0.2862 278.99 96.935  
## - FFMC\_ref 1 0.5700 279.27 97.150  
## - DMC.ISI 1 0.5714 279.27 97.151  
## - FFMC.ISI 1 0.5726 279.28 97.152  
## - temp 1 0.7499 279.45 97.286  
## - X 1 0.7571 279.46 97.291  
## - Y 1 1.1153 279.82 97.562  
## - RH\_sq 1 1.4343 280.14 97.802  
## - RH 1 1.6527 280.36 97.966  
## - wind\_sq 1 2.1404 280.84 98.333  
## - wind 1 2.5924 281.30 98.672  
## <none> 278.70 98.719  
## - FFMC.DC 1 4.0658 282.77 99.775  
## - DC 1 4.2239 282.93 99.893  
## - DMC 1 5.7014 284.40 100.992  
## - FFMC.DMC 1 5.7344 284.44 101.016  
## - month 1 6.5707 285.27 101.636  
##   
## Step: AIC=96.76  
## log(area + 1) ~ X + Y + month + FFMC\_ref + DMC + DC + FFMC.DMC +   
## FFMC.DC + FFMC.ISI + DMC.ISI + DC.ISI + temp + RH + RH\_sq +   
## wind + wind\_sq + Rain\_log + rain\_sq  
##   
## Df Sum of Sq RSS AIC  
## - Rain\_log 1 0.1503 278.91 94.875  
## - rain\_sq 1 0.2778 279.04 94.972  
## - DC.ISI 1 0.2859 279.05 94.978  
## - FFMC\_ref 1 0.5649 279.32 95.189  
## - DMC.ISI 1 0.5818 279.34 95.201  
## - X 1 0.7604 279.52 95.336  
## - FFMC.ISI 1 0.7761 279.54 95.348  
## - Y 1 1.1277 279.89 95.613  
## - RH\_sq 1 1.5567 280.32 95.936  
## - temp 1 1.6199 280.38 95.984  
## - wind\_sq 1 2.0879 280.85 96.336  
## - wind 1 2.5455 281.31 96.679  
## <none> 278.76 96.761  
## - RH 1 2.8372 281.60 96.898  
## - FFMC.DC 1 4.1626 282.92 97.889  
## - DC 1 4.4096 283.17 98.073  
## - DMC 1 5.7394 284.50 99.062  
## - FFMC.DMC 1 5.7713 284.53 99.085  
## - month 1 6.5760 285.33 99.681  
##   
## Step: AIC=94.88  
## log(area + 1) ~ X + Y + month + FFMC\_ref + DMC + DC + FFMC.DMC +   
## FFMC.DC + FFMC.ISI + DMC.ISI + DC.ISI + temp + RH + RH\_sq +   
## wind + wind\_sq + rain\_sq  
##   
## Df Sum of Sq RSS AIC  
## - rain\_sq 1 0.2038 279.11 93.029  
## - DC.ISI 1 0.2928 279.20 93.097  
## - DMC.ISI 1 0.5623 279.47 93.300  
## - FFMC\_ref 1 0.5713 279.48 93.307  
## - X 1 0.7619 279.67 93.451  
## - FFMC.ISI 1 0.7964 279.71 93.477  
## - Y 1 1.1119 280.02 93.715  
## - RH\_sq 1 1.4919 280.40 94.001  
## - temp 1 1.7869 280.70 94.223  
## - wind\_sq 1 2.2360 281.14 94.560  
## - wind 1 2.6320 281.54 94.857  
## <none> 278.91 94.875  
## - RH 1 2.7901 281.70 94.975  
## - FFMC.DC 1 4.2459 283.15 96.063  
## - DC 1 4.4948 283.40 96.248  
## - DMC 1 5.8070 284.72 97.223  
## - FFMC.DMC 1 5.8357 284.75 97.244  
## - month 1 6.8442 285.75 97.990  
##   
## Step: AIC=93.03  
## log(area + 1) ~ X + Y + month + FFMC\_ref + DMC + DC + FFMC.DMC +   
## FFMC.DC + FFMC.ISI + DMC.ISI + DC.ISI + temp + RH + RH\_sq +   
## wind + wind\_sq  
##   
## Df Sum of Sq RSS AIC  
## - DC.ISI 1 0.2658 279.38 91.230  
## - FFMC\_ref 1 0.4534 279.57 91.372  
## - DMC.ISI 1 0.5253 279.64 91.426  
## - FFMC.ISI 1 0.6810 279.79 91.543  
## - X 1 0.8401 279.95 91.663  
## - Y 1 1.1215 280.24 91.875  
## - RH\_sq 1 1.4485 280.56 92.121  
## - temp 1 1.7163 280.83 92.323  
## - wind\_sq 1 2.2032 281.32 92.688  
## - wind 1 2.6371 281.75 93.013  
## <none> 279.11 93.029  
## - RH 1 2.6869 281.80 93.051  
## - FFMC.DC 1 4.4112 283.52 94.338  
## - DC 1 4.6528 283.77 94.518  
## - DMC 1 5.8282 284.94 95.390  
## - FFMC.DMC 1 5.8492 284.96 95.405  
## - month 1 7.0287 286.14 96.277  
##   
## Step: AIC=91.23  
## log(area + 1) ~ X + Y + month + FFMC\_ref + DMC + DC + FFMC.DMC +   
## FFMC.DC + FFMC.ISI + DMC.ISI + temp + RH + RH\_sq + wind +   
## wind\_sq  
##   
## Df Sum of Sq RSS AIC  
## - FFMC\_ref 1 0.3336 279.71 89.482  
## - FFMC.ISI 1 0.4231 279.80 89.550  
## - X 1 0.7993 280.18 89.833  
## - Y 1 0.9910 280.37 89.977  
## - RH\_sq 1 1.4106 280.79 90.293  
## - temp 1 1.6016 280.98 90.436  
## - DMC.ISI 1 1.7547 281.13 90.551  
## - wind\_sq 1 1.9536 281.33 90.700  
## - wind 1 2.3996 281.78 91.035  
## - RH 1 2.6376 282.02 91.213  
## <none> 279.38 91.230  
## - DC 1 6.2400 285.62 93.891  
## - DMC 1 6.4125 285.79 94.018  
## - FFMC.DMC 1 6.5816 285.96 94.143  
## - FFMC.DC 1 6.7211 286.10 94.246  
## - month 1 7.3655 286.75 94.721  
##   
## Step: AIC=89.48  
## log(area + 1) ~ X + Y + month + DMC + DC + FFMC.DMC + FFMC.DC +   
## FFMC.ISI + DMC.ISI + temp + RH + RH\_sq + wind + wind\_sq  
##   
## Df Sum of Sq RSS AIC  
## - FFMC.ISI 1 0.1310 279.84 87.581  
## - X 1 0.8867 280.60 88.150  
## - Y 1 1.0204 280.73 88.250  
## - RH\_sq 1 1.2890 281.00 88.452  
## - DMC.ISI 1 1.4358 281.15 88.562  
## - wind\_sq 1 1.7472 281.46 88.796  
## - temp 1 1.9529 281.67 88.950  
## - wind 1 2.2122 281.93 89.144  
## - RH 1 2.4712 282.18 89.338  
## <none> 279.71 89.482  
## - DC 1 5.9154 285.63 91.898  
## - FFMC.DC 1 6.3999 286.11 92.255  
## - month 1 7.2825 287.00 92.905  
## - DMC 1 8.1697 287.88 93.556  
## - FFMC.DMC 1 8.4747 288.19 93.780  
##   
## Step: AIC=87.58  
## log(area + 1) ~ X + Y + month + DMC + DC + FFMC.DMC + FFMC.DC +   
## DMC.ISI + temp + RH + RH\_sq + wind + wind\_sq  
##   
## Df Sum of Sq RSS AIC  
## - X 1 0.9559 280.80 86.300  
## - Y 1 1.0846 280.93 86.397  
## - RH\_sq 1 1.2143 281.06 86.494  
## - wind\_sq 1 1.7216 281.56 86.875  
## - temp 1 1.8397 281.68 86.963  
## - wind 1 2.2499 282.09 87.270  
## - RH 1 2.3808 282.22 87.368  
## <none> 279.84 87.581  
## - DMC.ISI 1 3.2857 283.13 88.044  
## - DC 1 6.5940 286.44 90.495  
## - FFMC.DC 1 7.1886 287.03 90.932  
## - month 1 7.6480 287.49 91.270  
## - DMC 1 8.4935 288.34 91.890  
## - FFMC.DMC 1 8.9928 288.84 92.255  
##   
## Step: AIC=86.3  
## log(area + 1) ~ Y + month + DMC + DC + FFMC.DMC + FFMC.DC + DMC.ISI +   
## temp + RH + RH\_sq + wind + wind\_sq  
##   
## Df Sum of Sq RSS AIC  
## - Y 1 0.3890 281.19 84.592  
## - RH\_sq 1 1.2282 282.03 85.221  
## - temp 1 1.7522 282.55 85.613  
## - wind\_sq 1 1.7943 282.59 85.644  
## - RH 1 2.3385 283.14 86.050  
## - wind 1 2.3687 283.17 86.073  
## <none> 280.80 86.300  
## - DMC.ISI 1 3.4937 284.29 86.909  
## - DC 1 6.3337 287.13 89.007  
## - FFMC.DC 1 6.9243 287.72 89.440  
## - month 1 7.3524 288.15 89.754  
## - DMC 1 8.2105 289.01 90.381  
## - FFMC.DMC 1 8.6979 289.50 90.737  
##   
## Step: AIC=84.59  
## log(area + 1) ~ month + DMC + DC + FFMC.DMC + FFMC.DC + DMC.ISI +   
## temp + RH + RH\_sq + wind + wind\_sq  
##   
## Df Sum of Sq RSS AIC  
## - RH\_sq 1 1.1713 282.36 83.469  
## - wind\_sq 1 1.6940 282.88 83.860  
## - temp 1 1.8334 283.02 83.964  
## - RH 1 2.2521 283.44 84.276  
## - wind 1 2.2808 283.47 84.297  
## <none> 281.19 84.592  
## - DMC.ISI 1 3.3679 284.56 85.105  
## - DC 1 6.2742 287.46 87.249  
## - FFMC.DC 1 6.8508 288.04 87.672  
## - month 1 7.7005 288.89 88.293  
## - DMC 1 8.1976 289.39 88.656  
## - FFMC.DMC 1 8.6736 289.86 89.003  
##   
## Step: AIC=83.47  
## log(area + 1) ~ month + DMC + DC + FFMC.DMC + FFMC.DC + DMC.ISI +   
## temp + RH + wind + wind\_sq  
##   
## Df Sum of Sq RSS AIC  
## - wind\_sq 1 1.6377 284.00 82.690  
## - temp 1 2.0584 284.42 83.002  
## - wind 1 2.3146 284.68 83.192  
## <none> 282.36 83.469  
## - DMC.ISI 1 3.0483 285.41 83.735  
## - RH 1 4.1578 286.52 84.554  
## - DC 1 6.3717 288.73 86.178  
## - FFMC.DC 1 6.9789 289.34 86.621  
## - DMC 1 8.2898 290.65 87.575  
## - month 1 8.2987 290.66 87.581  
## - FFMC.DMC 1 8.7696 291.13 87.923  
##   
## Step: AIC=82.69  
## log(area + 1) ~ month + DMC + DC + FFMC.DMC + FFMC.DC + DMC.ISI +   
## temp + RH + wind  
##   
## Df Sum of Sq RSS AIC  
## - wind 1 1.1190 285.12 81.519  
## - temp 1 1.5454 285.54 81.835  
## <none> 284.00 82.690  
## - DMC.ISI 1 2.8551 286.85 82.800  
## - RH 1 3.3641 287.36 83.174  
## - DC 1 5.9315 289.93 85.051  
## - FFMC.DC 1 6.5123 290.51 85.473  
## - DMC 1 7.7611 291.76 86.379  
## - FFMC.DMC 1 8.2219 292.22 86.712  
## - month 1 8.8834 292.88 87.189  
##   
## Step: AIC=81.52  
## log(area + 1) ~ month + DMC + DC + FFMC.DMC + FFMC.DC + DMC.ISI +   
## temp + RH  
##   
## Df Sum of Sq RSS AIC  
## - DMC.ISI 1 2.2057 287.32 81.146  
## - temp 1 2.3670 287.48 81.264  
## <none> 285.12 81.519  
## - RH 1 3.5925 288.71 82.161  
## - DC 1 6.0157 291.13 83.925  
## - FFMC.DC 1 6.6128 291.73 84.357  
## - DMC 1 7.4987 292.62 84.997  
## - FFMC.DMC 1 7.9341 293.05 85.311  
## - month 1 8.4226 293.54 85.662  
##   
## Step: AIC=81.15  
## log(area + 1) ~ month + DMC + DC + FFMC.DMC + FFMC.DC + temp +   
## RH  
##   
## Df Sum of Sq RSS AIC  
## <none> 287.32 81.146  
## - temp 1 2.9752 290.30 81.319  
## - RH 1 4.2228 291.55 82.224  
## - DC 1 4.8678 292.19 82.690  
## - DMC 1 5.4297 292.75 83.096  
## - FFMC.DC 1 5.4460 292.77 83.107  
## - FFMC.DMC 1 5.8096 293.13 83.369  
## - month 1 10.9802 298.30 87.059

##   
## Call:  
## lm(formula = log(area + 1) ~ month + DMC + DC + FFMC.DMC + FFMC.DC +   
## temp + RH, data = forest\_train\_new)  
##   
## Coefficients:  
## (Intercept) month DMC DC FFMC.DMC FFMC.DC   
## 2.7671595 0.0681836 -0.1315429 0.0185634 0.0015019 -0.0002177   
## temp RH   
## -0.0322246 -0.0121802

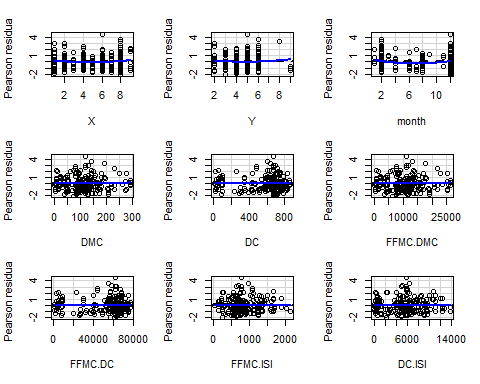
#running suggested model  
  
mod8 <- lm(formula = log(area + 1) ~ X + Y + month + DMC + DC + FFMC.DMC +   
 FFMC.DC + FFMC.ISI + DC.ISI + RH + RH\_sq, data = forest\_train\_new)  
  
#mod8 = lm(formula = log(area + 1) ~ DC + temp + temp\_sq + RH\_sq + wind +   
# Rain\_log + temp.RH + wind.temp, data = forest\_train\_new)  
  
summary(mod8)

##   
## Call:  
## lm(formula = log(area + 1) ~ X + Y + month + DMC + DC + FFMC.DMC +   
## FFMC.DC + FFMC.ISI + DC.ISI + RH + RH\_sq, data = forest\_train\_new)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.0541 -0.8667 -0.2044 0.6348 4.5063   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.094e+00 9.640e-01 3.210 0.00155 \*\*  
## X 3.678e-02 4.063e-02 0.905 0.36646   
## Y -8.077e-02 8.139e-02 -0.992 0.32220   
## month 6.395e-02 2.528e-02 2.529 0.01221 \*   
## DMC -1.333e-01 6.808e-02 -1.959 0.05156 .   
## DC 1.555e-02 9.886e-03 1.573 0.11724   
## FFMC.DMC 1.512e-03 7.501e-04 2.016 0.04513 \*   
## FFMC.DC -1.794e-04 1.107e-04 -1.621 0.10655   
## FFMC.ISI -9.355e-06 7.028e-04 -0.013 0.98939   
## DC.ISI -6.911e-05 1.179e-04 -0.586 0.55857   
## RH -4.168e-02 3.299e-02 -1.263 0.20794   
## RH\_sq 3.519e-04 3.319e-04 1.061 0.29020   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.196 on 199 degrees of freedom  
## Multiple R-squared: 0.07853, Adjusted R-squared: 0.0276   
## F-statistic: 1.542 on 11 and 199 DF, p-value: 0.119

plot(mod8)



residualPlots(mod8)

Diagram

Description automatically generated with medium confidence

## Test stat Pr(>|Test stat|)   
## X 1.2801 0.202009   
## Y 0.7790 0.436928   
## month 2.9392 0.003682 \*\*  
## DMC 0.5816 0.561503   
## DC -0.1871 0.851790   
## FFMC.DMC 0.6058 0.545331   
## FFMC.DC -0.1345 0.893166   
## FFMC.ISI -0.2497 0.803104   
## DC.ISI -0.2784 0.780965   
## RH -0.6478 0.517888   
## RH\_sq -1.1761 0.240985   
## Tukey test 1.4354 0.151183   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Although the R sq is low , but that seems due to the predictors available in data and there might be missing some important variables. But our other factors and plots show that this is the optimal model.

## Now we test the model

FFMC.DMC <- forest\_test$FFMC\*forest\_test$DMC  
FFMC.DC <-forest\_test$FFMC\*forest\_test$DC  
FFMC.ISI <-forest\_test$FFMC\*forest\_test$ISI  
DC.ISI<-forest\_test$DC\*forest\_test$ISI  
RH\_sq<-(forest\_test$RH)^2  
testData<-cbind(forest\_test,FFMC.DMC,FFMC.DC,FFMC.ISI,DC.ISI,RH\_sq)  
  
model <-lm(formula = log(area + 1) ~ X + Y + month + DMC + DC + FFMC.DMC +   
 FFMC.DC + FFMC.ISI + DC.ISI + RH + RH\_sq, data = testData)  
  
  
y\_hat<-predict.lm(model,newdata=testData, se.fit=TRUE)$fit  
y\_hat<-as.vector(y\_hat)   
dev<-log(testData$area+1)-(y\_hat)   
num<-sum(dev^2)   
dev1<-log(testData$area+1)-mean(log(testData$area+1))   
den<-sum(dev1^2)   
Predicted.Rsq<-1-(num/den)   
Predicted.Rsq

## [1] 0.1298374

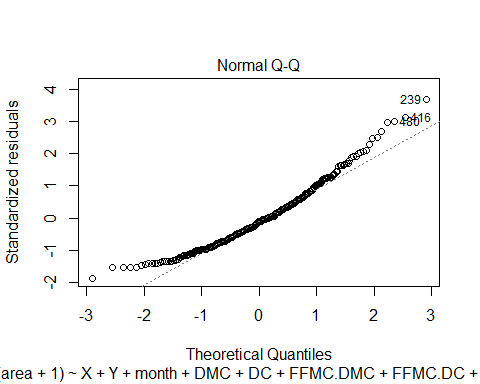
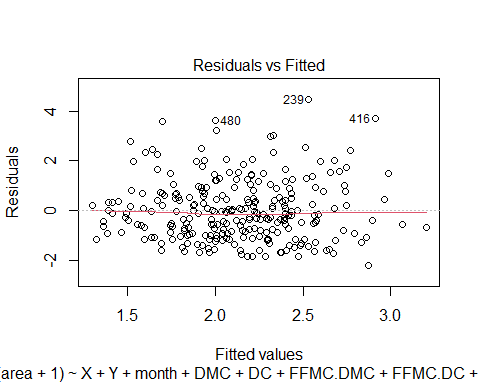
The predicted R sqaure is 28.59%. This is a considerably good fit as per the given dataset.

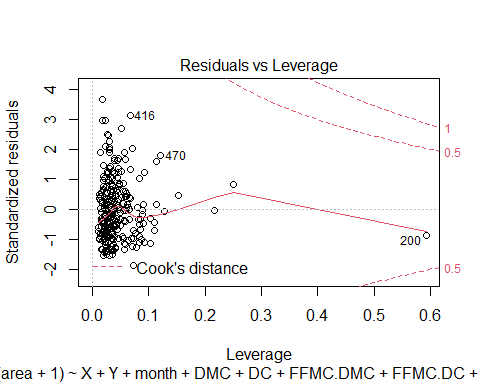
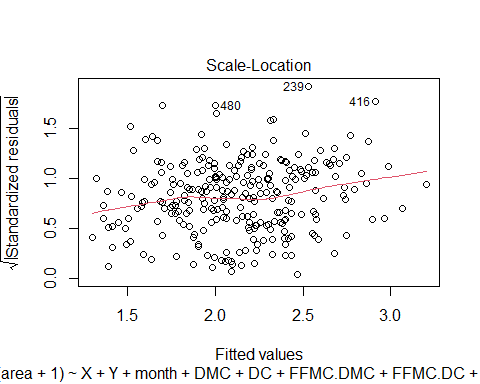
# Running on original data

FFMC.DMC <- forest$FFMC\*forest$DMC  
FFMC.DC <-forest$FFMC\*forest$DC  
FFMC.ISI <-forest$FFMC\*forest$ISI  
DC.ISI<-forest$DC\*forest$ISI  
RH\_sq<-(forest$RH)^2  
  
  
forest\_new<-cbind(forest,FFMC.DMC,FFMC.DC,FFMC.ISI,DC.ISI,RH\_sq)  
forest\_new <- forest\_new[forest\_new$area>0,]  
  
  
model\_full <-lm(formula = log(area + 1) ~ X + Y + month + DMC + DC + FFMC.DMC +   
 FFMC.DC + FFMC.ISI + DC.ISI + RH + RH\_sq, data = forest\_new)  
  
summary(model\_full)

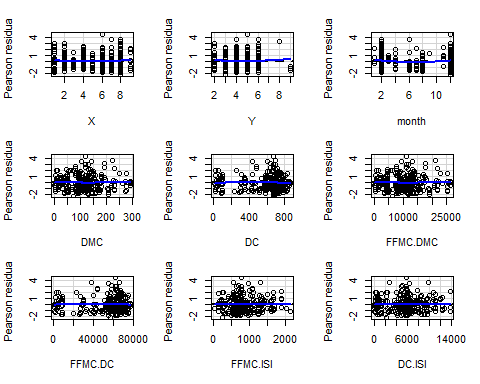
##   
## Call:  
## lm(formula = log(area + 1) ~ X + Y + month + DMC + DC + FFMC.DMC +   
## FFMC.DC + FFMC.ISI + DC.ISI + RH + RH\_sq, data = forest\_new)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -2.2023 -0.9057 -0.1385 0.6996 4.4620   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 3.281e+00 8.356e-01 3.927 0.000111 \*\*\*  
## X 3.562e-02 3.706e-02 0.961 0.337274   
## Y -7.364e-02 7.565e-02 -0.973 0.331262   
## month 6.385e-02 2.277e-02 2.803 0.005440 \*\*   
## DMC -1.715e-01 6.104e-02 -2.809 0.005343 \*\*   
## DC 2.229e-02 9.076e-03 2.455 0.014734 \*   
## FFMC.DMC 1.933e-03 6.713e-04 2.879 0.004324 \*\*   
## FFMC.DC -2.543e-04 1.020e-04 -2.492 0.013340 \*   
## FFMC.ISI -1.462e-04 6.670e-04 -0.219 0.826697   
## DC.ISI -4.965e-05 1.118e-04 -0.444 0.657283   
## RH -5.002e-02 2.673e-02 -1.871 0.062420 .   
## RH\_sq 4.702e-04 2.646e-04 1.777 0.076751 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 1.228 on 258 degrees of freedom  
## Multiple R-squared: 0.08572, Adjusted R-squared: 0.04674   
## F-statistic: 2.199 on 11 and 258 DF, p-value: 0.01491

plot(model\_full)





residualPlots(model\_full)

Chart

Description automatically generated

## Test stat Pr(>|Test stat|)   
## X 0.6533 0.51416   
## Y 0.6487 0.51713   
## month 2.0484 0.04154 \*  
## DMC 0.3854 0.70025   
## DC -0.7816 0.43515   
## FFMC.DMC 0.4549 0.64955   
## FFMC.DC -0.6627 0.50813   
## FFMC.ISI 0.1637 0.87008   
## DC.ISI 0.4570 0.64806   
## RH 1.3558 0.17634   
## RH\_sq -0.5189 0.60429   
## Tukey test 0.7873 0.43112   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1