Forest Fire Project Part 3

Changning Liu, Shi Fan Jin, Esther Law

April 30, 2018

Introduction

In this final project, we are required to analyze the Forest Fires dataset, which contains information on forest fire data from the Montesinho natural park in northeast Portugal. The data were collected from January 2000 to December 2003 and there are 517 observations of 13 variables, listed below

- Spatial and temporal variables
  (a) X - x-axis spatial coordinate within the Montesinho park map: 1 to 9
  (b) Y - y-axis spatial coordinate within the Montesinho park map: 2 to 9
  (c) month - month of the year: 'jan' to 'dec'
  (d) day - day of the week: 'mon' to 'sun'
- FWI component variables: The forest Fire Weather Index is the Canadian system for rating fire danger
  (a) FFMC - FFMC index denotes the moisture content surface litter and influences ignition and fire spread: 18.7 to 96.20
  (b) DMC - DMC index represent the moisture content of shallow organic layers: 1.1 to 291.3
  (c) DC - DC index represent the moisture content of deep organic layers: 7.9 to 860.6
  (d) ISI - ISI index is a score that correlates with fire velocity spread: 0.0 to 56.10
- Weather variables
  (a) temp - temperature in Celsius degrees: 2.2 to 33.30
  (b) RH - relative humidity in %: 15.0 to 100
  (c) wind - wind speed in km/h: 0.40 to 9.40
  (d) rain: outside rain (in mm=m2); denotes the accumulated precipitation within the previous 30 minutes.
- Area: total burned area (in hectares). Note that all the entries denote fire occurrences and a value of zero indicates that the burned area was less than 0.01 hectares or 100m2.

My goal is to analyze the relationship between the size of the burned area of forest fires and the meteorological and other variables given and I want to find out the variables that influence the total burned area in forest fires.

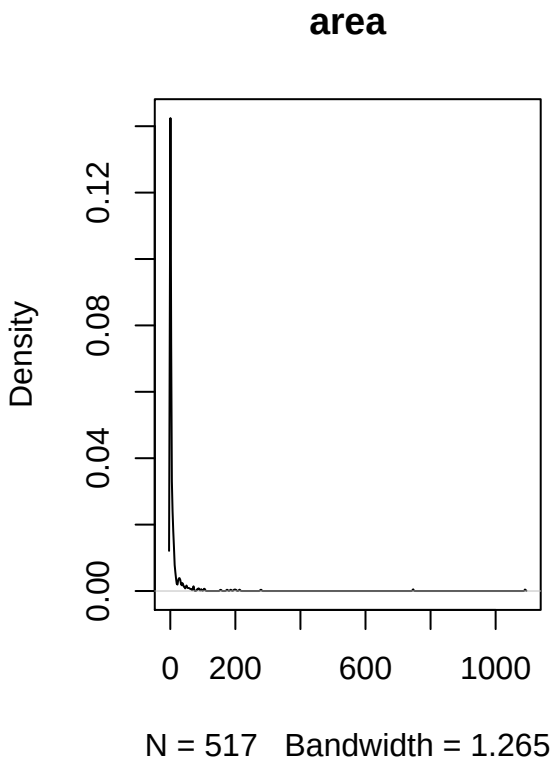
Data exploration

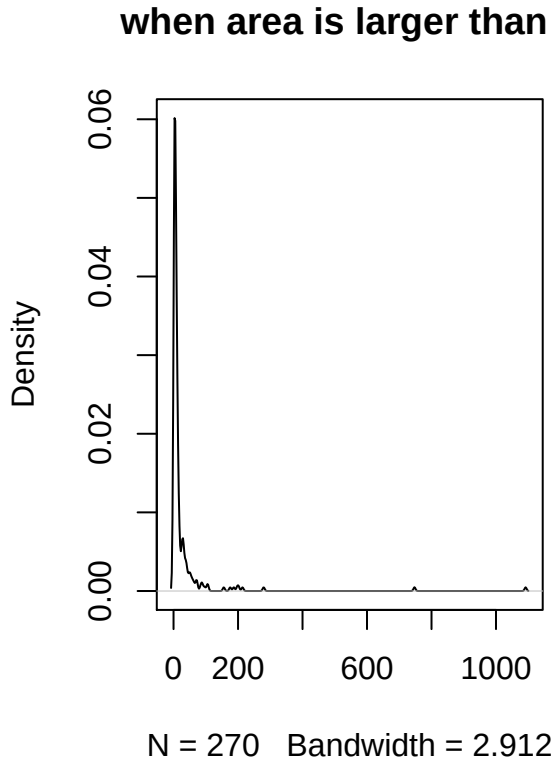
I first check the data structure to get a preliminary idea about the data including the typeand quantile of each variable,etc. Also, I observed the distribution of our response variable, area. It appears that our response variable is highly right skewed which implies that a log transformation may be helpful when conducting regression analysis. And since our purpose is to predict the size of burned area and find the variables that may influence the size of fire, I only focus on the data that have positive burned area.

```
## 'data.frame':    517 obs. of  13 variables:
##  $ X    : int  7 7 7 8 8 8 8 8 8 7 ...
##  $ Y    : int  5 4 4 6 6 6 6 6 6 5 ...
##  $ month: Factor w/ 12 levels "apr","aug","dec",..: 8 11 11 8 8 2 2 2 12 12 ...
##  $ day  : Factor w/ 7 levels "fri","mon","sat",..: 1 6 3 1 4 4 2 2 6 3 ...
##  $ FFMC : num  86.2 90.6 90.6 91.7 89.3 92.3 92.3 91.5 91 92.5 ...
##  $ DMC  : num  26.2 35.4 43.7 33.3 51.3 ...
##  $ DC   : num  94.3 669.1 686.9 77.5 102.2 ...
##  $ ISI  : num  5.1 6.7 6.7 9 9.6 14.7 8.5 10.7 7 7.1 ...
```

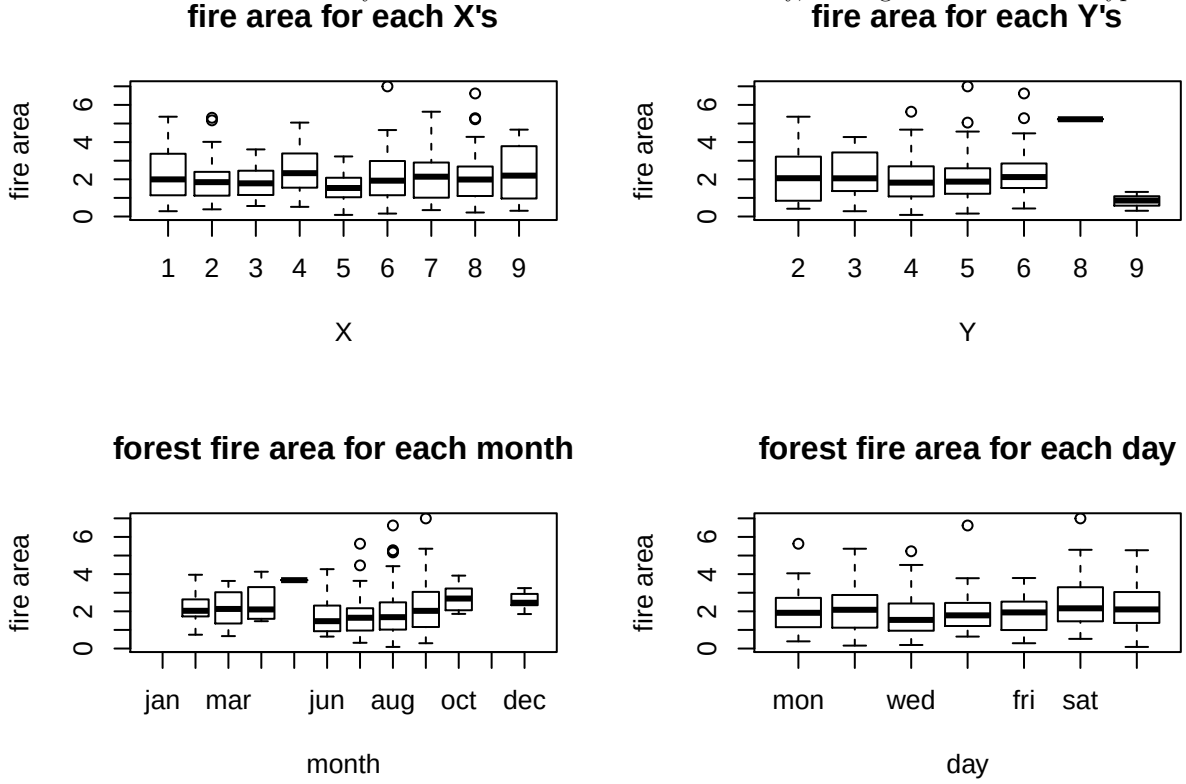
```
##  $ temp : num  8.2 18 14.6 8.3 11.4 22.2 24.1 8 13.1 22.8 ...
##  $ RH   : int  51 33 33 97 99 29 27 86 63 40 ...
##  $ wind : num  6.7 0.9 1.3 4 1.8 5.4 3.1 2.2 5.4 4 ...
##  $ rain : num  0 0 0 0.2 0 0 0 0 0 0 ...
##  $ area : num  0 0 0 0 0 0 0 0 0 0 ...
```

```
##        X               Y           month      day           FFMC
##  Min.   :1.000   Min.   :2.0   aug    :184   fri:85   Min.   :18.70
##  1st Qu.:3.000   1st Qu.:4.0   sep    :172   mon:74   1st Qu.:90.20
##  Median :4.000   Median :4.0   mar    : 54   sat:84   Median :91.60
##  Mean   :4.669   Mean   :4.3   jul    : 32   sun:95   Mean   :90.64
##  3rd Qu.:7.000   3rd Qu.:5.0   feb    : 20   thu:61   3rd Qu.:92.90
##  Max.   :9.000   Max.   :9.0   jun    : 17   tue:64   Max.   :96.20
##                                (Other): 38   wed:54
##       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
##
```

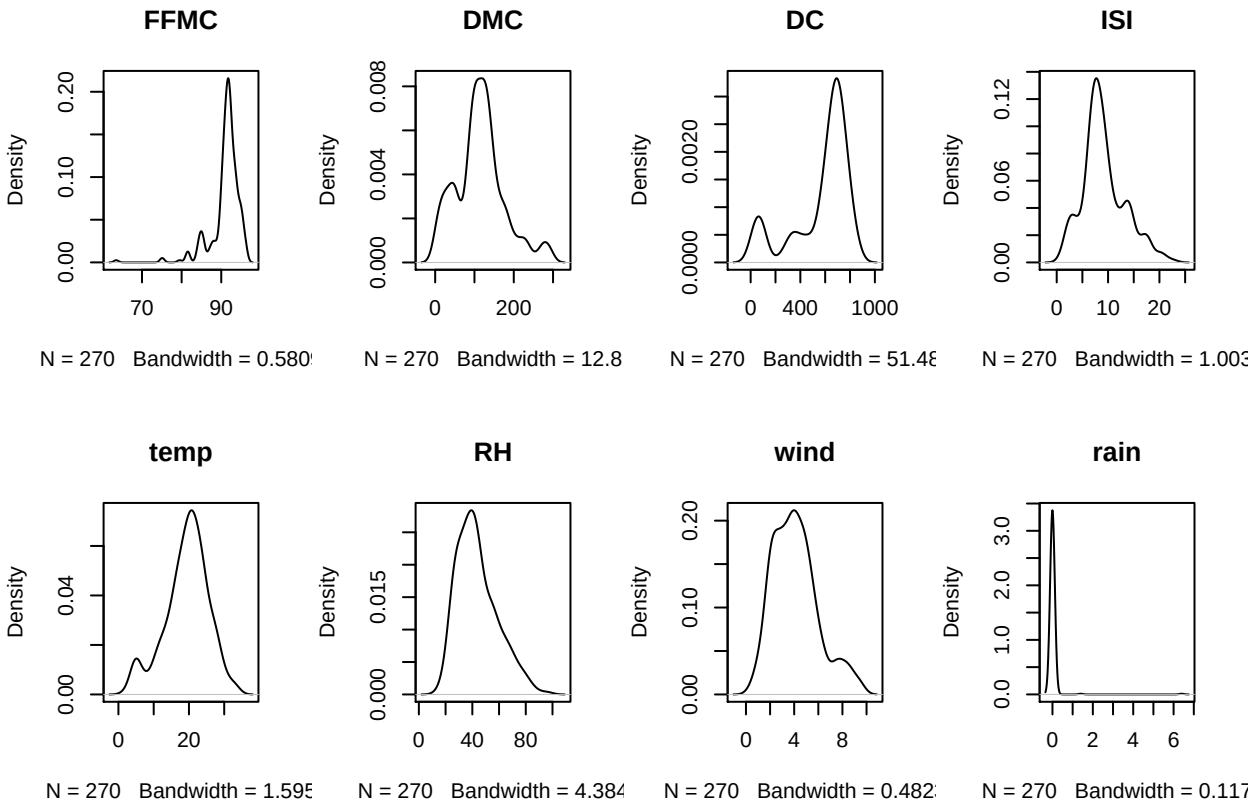




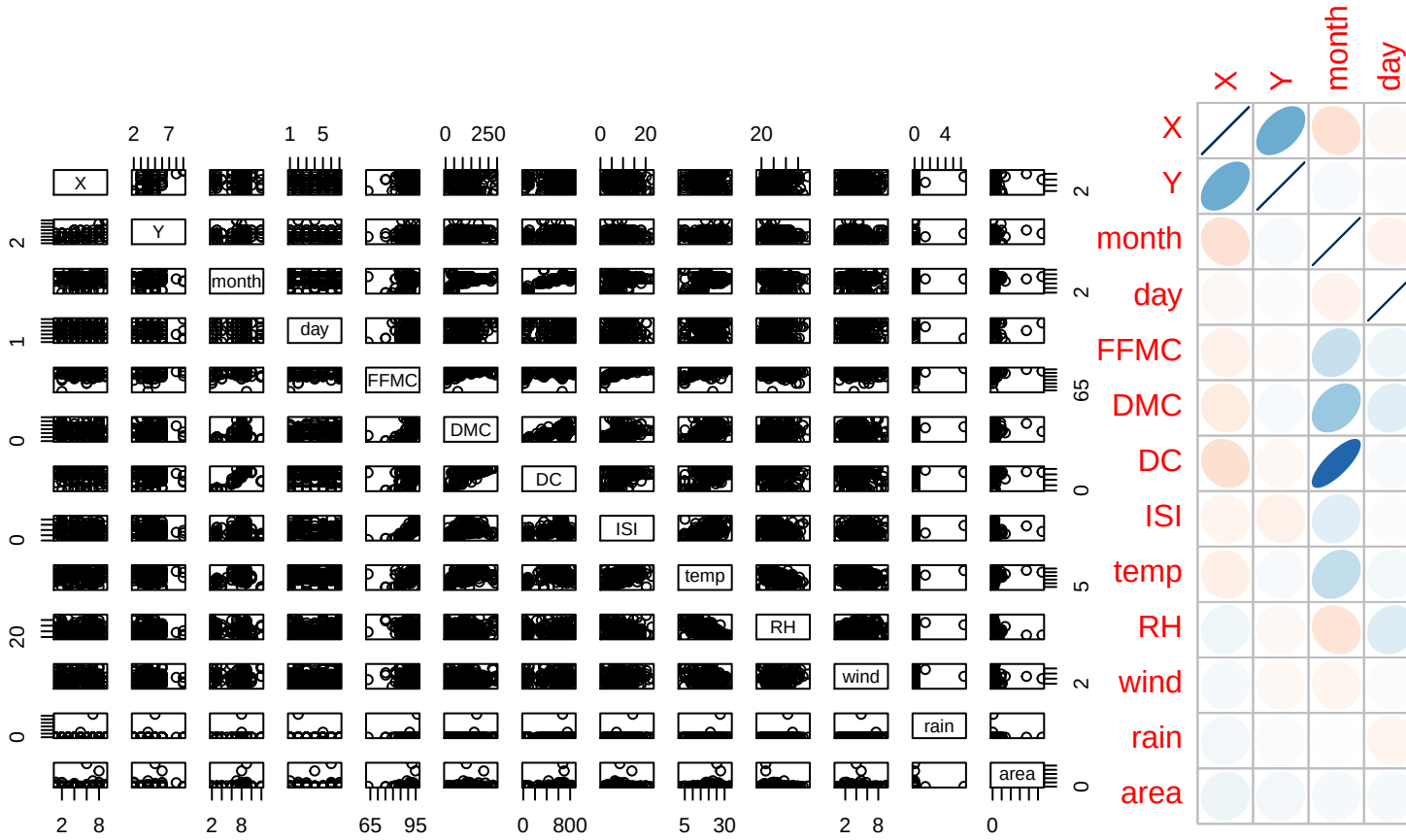
then try to explore the explanatory variables. First, I plot the boxplots for the Spatial and temporal variables. I also re-order the factors month and day so that the order now are from jan to dec and Monday to Sunday. It appears that for different months the number of fires differ significantly. For example there are no fire obesrved in jan and nov while a great amount of fires was found in aug and sep. It may implies that month is an influential variable. Also in order to better analysis the factor variables month and day, I changed these data into type integer.



I then try to explore the distribution of the explanatory variables expecially the distributions of FWI component variables and weather variables. It appears that FFMC and rain are heavily skewed which suggest transformation maybe helpful.



Another thing I want to know is the correlation among the explainatory variables especially the correlation among FWI component variables and weather variables.I plot both the scattermatrix plot and corrplot.

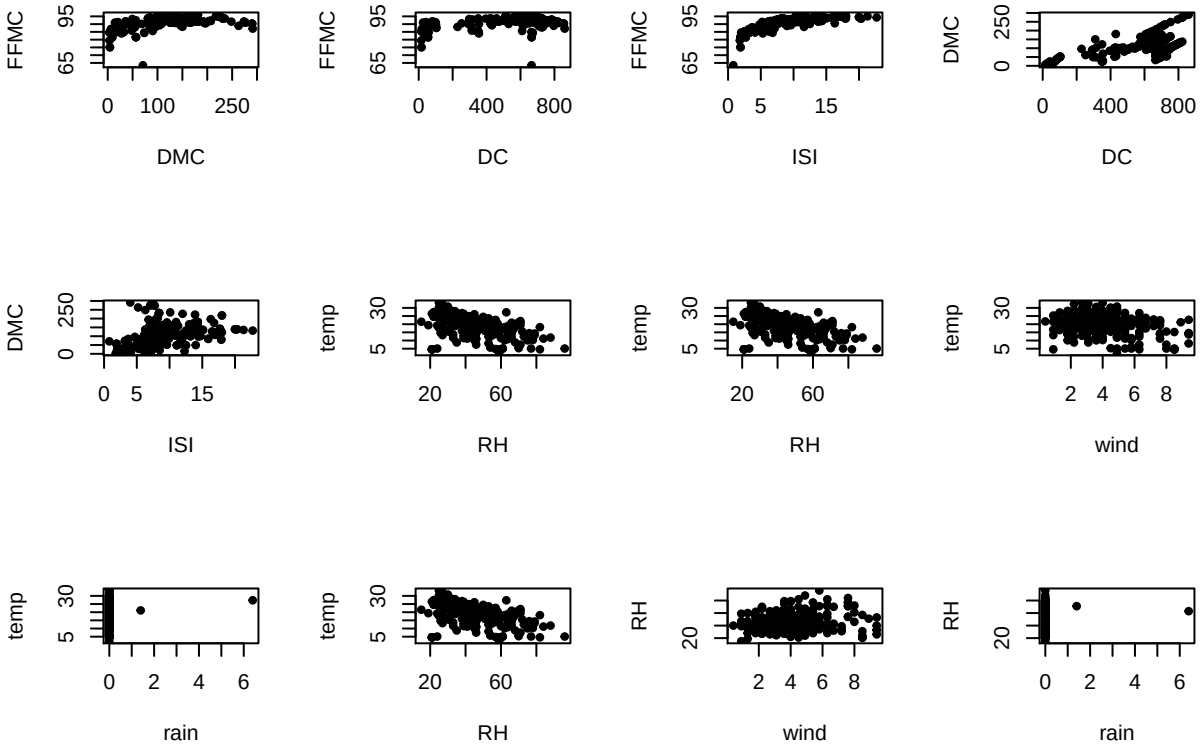



```
##                  X           Y       month         day        FFMC
## X       1.00000000  0.49703882 -0.15377607 -0.03085277 -0.07217614
## Y       0.49703882  1.00000000  0.03578095 -0.01723025 -0.02112302
## month  -0.15377607  0.03578095  1.00000000 -0.06680611  0.22969667
## day    -0.03085277 -0.01723025 -0.06680611  1.00000000  0.07509169
## FFMC   -0.07217614 -0.02112302  0.22969667  0.07509169  1.00000000
## DMC    -0.10461401  0.03753295  0.37187129  0.12998221  0.48025000
##                 DMC
## X      -0.10461401
## Y       0.03753295
## month   0.37187129
## day     0.12998221
## FFMC    0.48025000
## DMC     1.00000000
```

```
##                 DC          ISI         temp          RH         wind
## DC     1.00000000  0.256826065  0.49570270 -0.08221711 -0.237593410
## ISI    0.25682607  1.000000000  0.46602711 -0.14880362  0.072651791
## temp   0.49570270  0.466027115  1.00000000 -0.49754787 -0.320563247
## RH    -0.08221711 -0.148803620 -0.49754787  1.00000000  0.138489957
## wind  -0.23759341  0.072651791 -0.32056325  0.13848996  1.000000000
## rain   0.03637664  0.067225499  0.08098833  0.09992007  0.049111940
## area   0.04673457  0.002121065  0.11029294 -0.10484626  0.002085841
##              rain         area
## DC     0.03637664  0.046734570
## ISI    0.06722550  0.002121065
## temp   0.08098833  0.110292941
```

```
## RH    0.09992007 -0.104846261
## wind  0.04911194  0.002085841
## rain  1.00000000 -0.012901119
## area -0.01290112  1.000000000
```

To view the correlation more clearly, I compared the explainatory variables(FWI and weather variables) one-by-one. I found that there exist a moderate to strong positive correlation between FFMC and ISI,DMC and DC,while at the same time there is a moderate to strong negative correlation between temp and RH,etc.

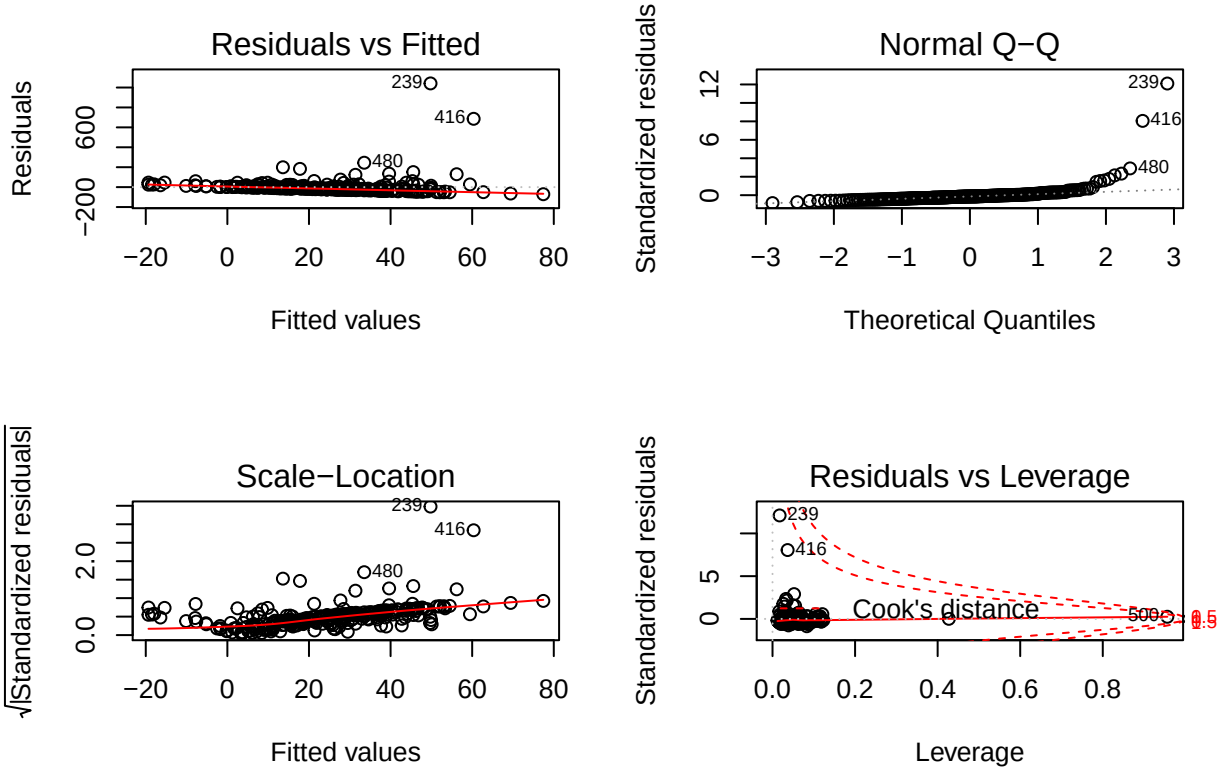


Regression analysis

Now let'S do the regression, I started with the easiest full model with out interaction term. But it seems this model fit terriable. There is an obvious pattern among the residuals and fitted value, and only about 3.8% of the variation can be explained by the model.

```
##
## Call:
## lm(formula = data1$area ~ ., data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
##  -71.13  -26.43  -11.93    2.48 1041.05
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -57.14665  211.22520  -0.271    0.787
## X             3.56649    2.64106   1.350    0.178
## Y            -1.56635    5.41352  -0.289    0.773
## month         4.22121    5.05635   0.835    0.405
## day           1.92123    2.67121   0.719    0.473
```

```
## FFMC          0.46215    2.36044   0.196    0.845
## DMC           0.17705    0.13652   1.297    0.196
## DC           -0.05242    0.05928  -0.884    0.377
## ISI          -1.83270    1.97426  -0.928    0.354
## temp          1.42478    1.48729   0.958    0.339
## RH           -0.45318    0.48253  -0.939    0.349
## wind          1.81215    3.27672   0.553    0.581
## rain         -4.00644   13.70705  -0.292    0.770
##
## Residual standard error: 86.79 on 257 degrees of freedom
## Multiple R-squared:  0.03828,    Adjusted R-squared:  -0.00663
## F-statistic: 0.8524 on 12 and 257 DF,  p-value: 0.5964
```



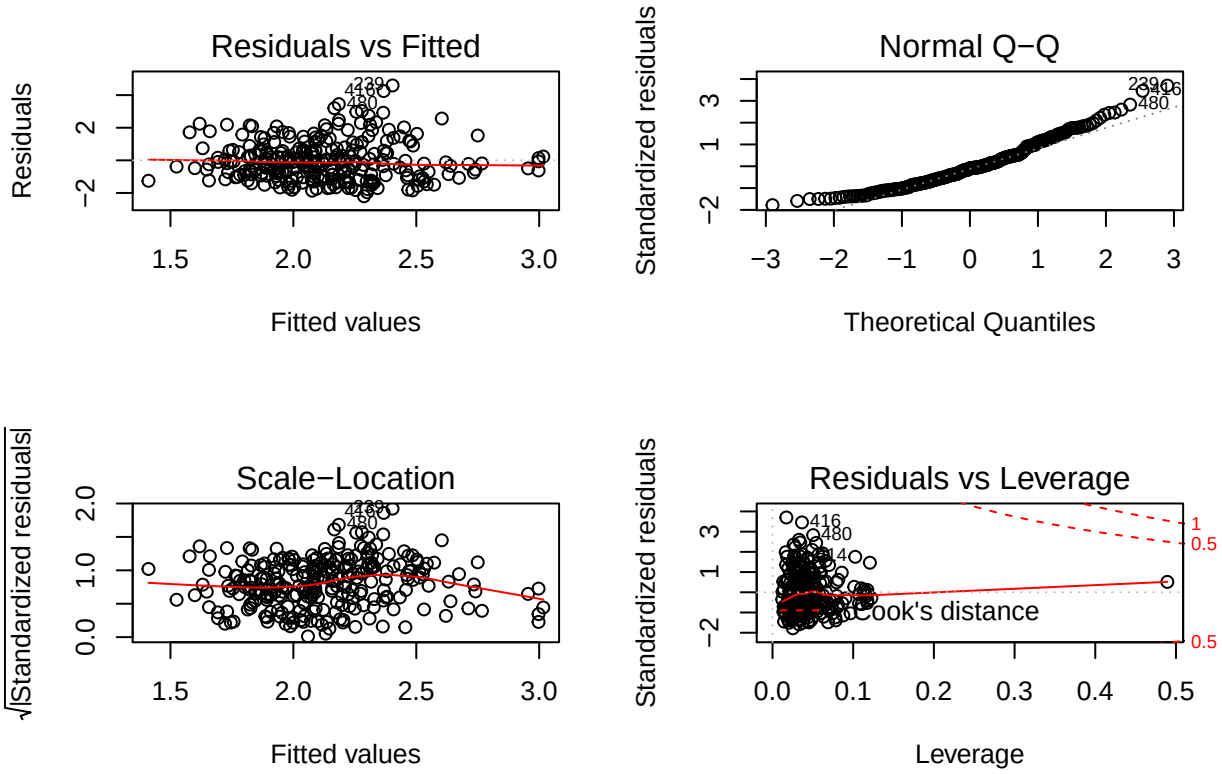
As mentioned earlier, I try to log transform the area variable and FFMC. Also I dropped the variable rain since only 2 out of 270 observations that have nonegative value and from it's p-value of the full model implies that it is not statistically significant.

* modified1.lm <-lm(log(data1$area+1)~X+Y+month+day+FFMC_LOG+DMC+DC+ISI+temp+RH+wind,data = data1)

This time it seems that the modified model is better than the previous one, at least the residuals vs fitted value seems to be random.

```
##
## Call:
## lm(formula = log(data1$area + 1) ~ X + Y + month + day + FFMC_LOG +
##     DMC + DC + ISI + temp + RH + wind, data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.2011 -0.9210 -0.1176  0.6110  4.5921
##
```

```
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.7453281 12.2987240   0.142   0.8873
## X            0.0426007  0.0380735   1.119   0.2642
## Y           -0.0909403  0.0780945  -1.164   0.2453
## month        0.1187862  0.0729191   1.629   0.1045
## day          0.0506013  0.0383380   1.320   0.1880
## FFMC_LOG     0.0988263  2.7522609   0.036   0.9714
## DMC          0.0040659  0.0019652   2.069   0.0395 *
## DC          -0.0014796  0.0008544  -1.732   0.0845 .
## ISI         -0.0404861  0.0276380  -1.465   0.1442
## temp         0.0006711  0.0211967   0.032   0.9748
## RH          -0.0086298  0.0068223  -1.265   0.2070
## wind         0.0267494  0.0469957   0.569   0.5697
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.252 on 258 degrees of freedom
## Multiple R-squared:  0.04912,    Adjusted R-squared:  0.008583
## F-statistic: 1.212 on 11 and 258 DF,  p-value: 0.2791
```

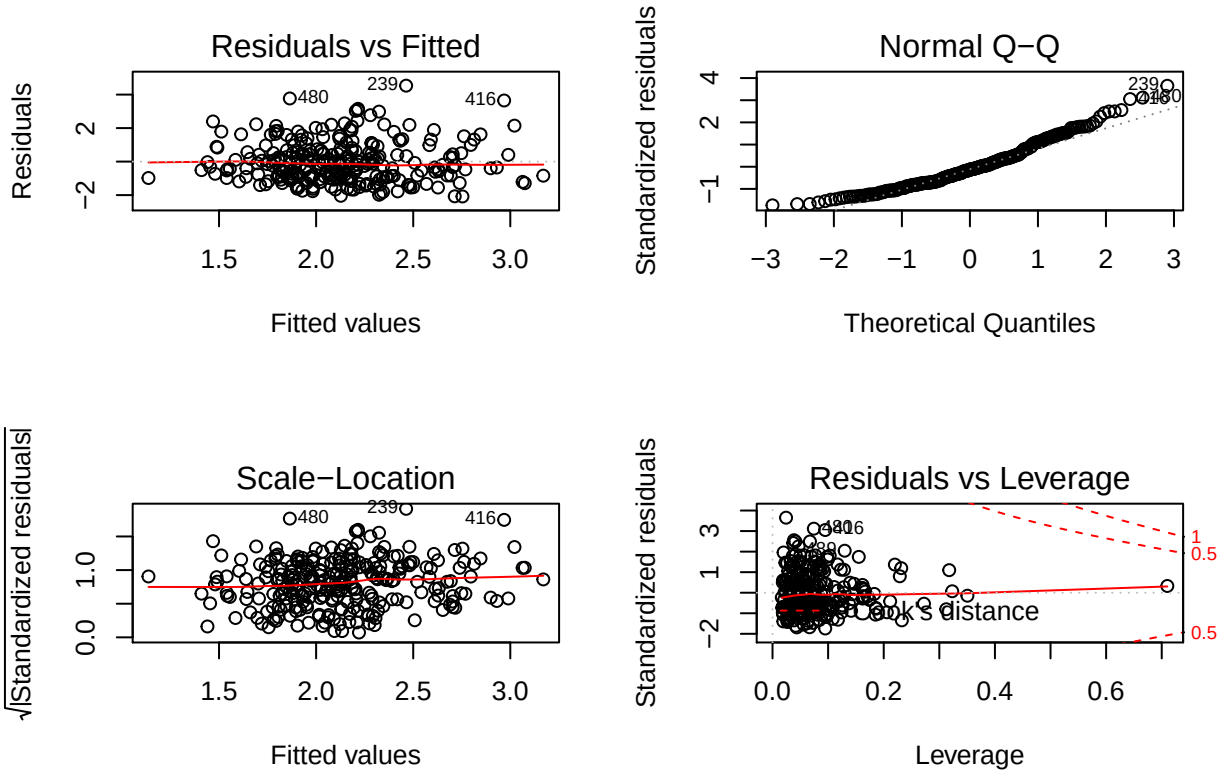


Now I try to add the interaction terms.
* modified2.lm<-lm(log(data1$area+1)~X+Y+month+day+FFMC_LOG+DMC+DC+ISI+temp+RH+wind+ FFMC_DMC+FFMC_DC+FFMC_ISI +DMC_DC+DMC_ISI+DC_ISI+temp_RH+temp_wind+RH_wind, data = data1)

The model now may able to explain 7.5% of the variation and the residuals seems to be random.

```
##
## Call:
## lm(formula = log(data1$area + 1) ~ X + Y + month + day + FFMC_LOG +
```

```
##     DMC + DC + ISI + temp + RH + wind + FFMC_DMC + FFMC_DC +
##     FFMC_ISI + DMC_DC + DMC_ISI + DC_ISI + temp_RH + temp_wind +
##     RH_wind, data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0807 -0.8953 -0.1466  0.6274  4.5320
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.389e+00  3.125e+01  -0.044    0.965
## X            5.105e-02  3.894e-02   1.311    0.191
## Y           -1.267e-01  8.105e-02  -1.563    0.119
## month        1.496e-01  9.819e-02   1.524    0.129
## day          3.521e-02  4.098e-02   0.859    0.391
## FFMC_LOG     6.513e-01  6.991e+00   0.093    0.926
## DMC         -1.474e-01  9.563e-02  -1.542    0.124
## DC           1.612e-02  1.759e-02   0.917    0.360
## ISI          3.166e-01  8.090e-01   0.391    0.696
## temp         8.030e-03  5.777e-02   0.139    0.890
## RH           2.009e-02  2.738e-02   0.734    0.464
## wind        -7.780e-02  2.397e-01  -0.325    0.746
## FFMC_DMC     1.633e-03  1.077e-03   1.516    0.131
## FFMC_DC     -1.874e-04  2.039e-04  -0.919    0.359
## FFMC_ISI    -3.151e-03  8.522e-03  -0.370    0.712
## DMC_DC       4.813e-06  1.043e-05   0.461    0.645
## DMC_ISI     -8.105e-05  7.001e-04  -0.116    0.908
## DC_ISI      -1.284e-04  2.024e-04  -0.634    0.526
## temp_RH     -1.391e-03  9.413e-04  -1.478    0.141
## temp_wind    9.237e-03  8.205e-03   1.126    0.261
## RH_wind     -1.340e-03  3.276e-03  -0.409    0.683
##
## Residual standard error: 1.257 on 249 degrees of freedom
## Multiple R-squared:  0.07564,    Adjusted R-squared:  0.001398
## F-statistic: 1.019 on 20 and 249 DF,  p-value: 0.4403
```

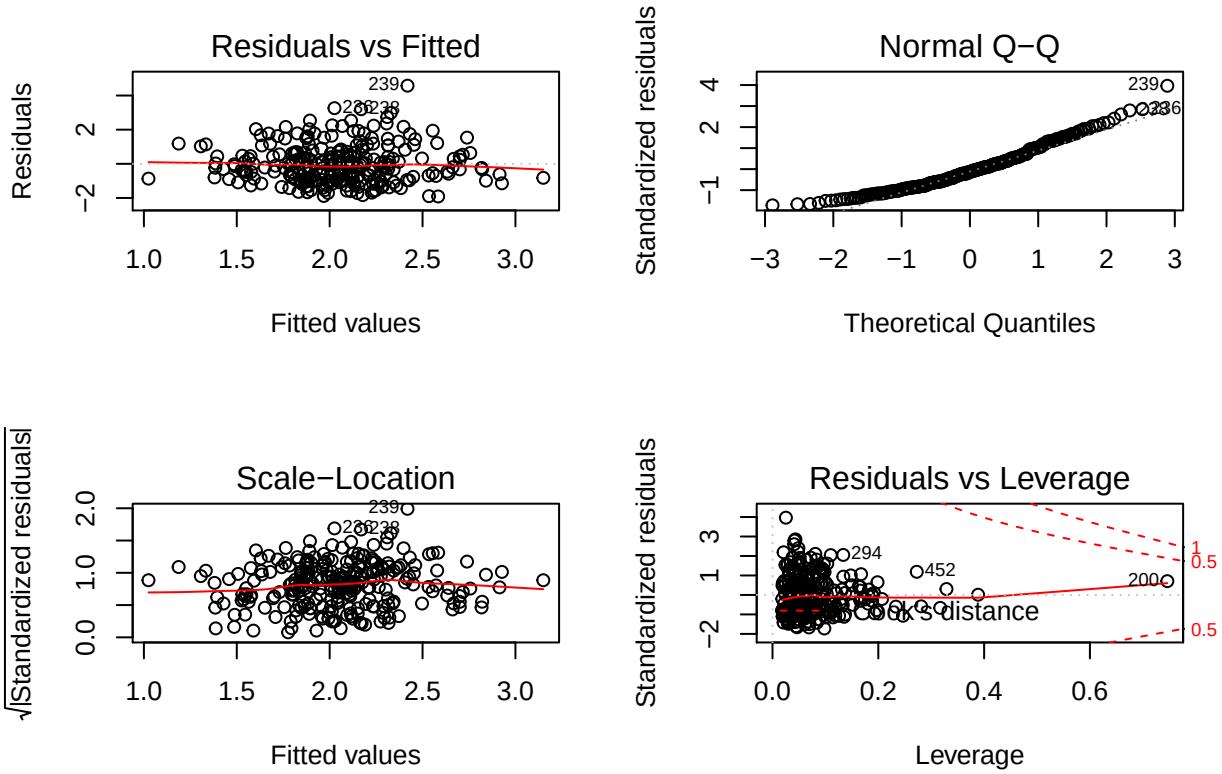


From the boxplot and density plot of the explainatory varable I made previously, I find some outliers, so I tried to find the influential points according to cook distance. I detected ten points that have high cook distance. Let's discard these observations to seem how much our model will improve without these data.

- modified3.lm <-lm(log(data3$area+1)~X+Y+month+day+FFMC_LOG3+DMC+DC+ISI+temp +RH+wind+FFMC_DMC3+FFMC_DC3+FFMC_ISI3 +DMC_DC3+DMC_ISI3+DC_ISI3+temp_RH3+temp_wind3 +RH_wind3,data = data3)

```
##     X Y month day FFMC   DMC    DC  ISI temp RH wind rain   area
## 212 7 4     8   6 93.5 139.4 594.2 20.3  5.1 96  5.8    0  26.00
## 224 2 2     7   5 88.3 150.3 309.9  6.8 13.4 79  3.6    0  37.02
## 363 7 4     9   5 88.2  55.2 732.3 11.6 15.2 64  3.1    0   0.52
## 378 2 2     8   6 93.7 231.1 715.1  8.4 21.9 42  2.2    0 174.63
## 416 8 6     8   4 94.8 222.4 698.6 13.9 27.5 27  4.9    0 746.28
## 421 8 8     8   3 91.7 191.4 635.9  7.8 26.2 36  4.5    0 185.76
## 470 6 3     4   7 91.0  14.6  25.6 12.3 13.7 33  9.4    0  61.13
## 480 7 4     7   1 89.2 103.9 431.6  6.4 22.6 57  4.9    0 278.53
## 489 4 4     8   2 95.1 141.3 605.8 17.7 19.4 71  7.6    0  46.70
## 514 2 4     8   7 81.6  56.7 665.6  1.9 21.9 71  5.8    0  54.29
##           cook
## 212 0.02674659
## 224 0.02516539
## 363 0.02613714
## 378 0.01776262
## 416 0.04641338
## 421 0.02136949
## 470 0.02046808
## 480 0.03715004
## 489 0.02968276
## 514 0.02661409
```

```
##
## Call:
## lm(formula = log(data3$area + 1) ~ X + Y + month + day + FFMC_LOG3 +
##     DMC + DC + ISI + temp + RH + wind + FFMC_DMC3 + FFMC_DC3 +
##     FFMC_ISI3 + DMC_DC3 + DMC_ISI3 + DC_ISI3 + temp_RH3 + temp_wind3 +
##     RH_wind3, data = data3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9111 -0.8754 -0.1432  0.6101  4.5765
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -8.727e+00  3.006e+01  -0.290   0.7718
## X            3.879e-02  3.675e-02   1.056   0.2922
## Y           -1.096e-01  7.855e-02  -1.395   0.1642
## month        1.246e-01  9.274e-02   1.343   0.1804
## day          4.587e-02  3.892e-02   1.179   0.2397
## FFMC_LOG3    2.229e+00  6.708e+00   0.332   0.7400
## DMC         -1.792e-01  1.021e-01  -1.755   0.0805 .
## DC           2.200e-02  1.811e-02   1.215   0.2255
## ISI          6.946e-01  8.313e-01   0.836   0.4042
## temp         2.377e-02  5.684e-02   0.418   0.6762
## RH           2.526e-02  2.775e-02   0.910   0.3637
## wind         9.555e-03  2.305e-01   0.041   0.9670
## FFMC_DMC3    1.951e-03  1.142e-03   1.708   0.0889 .
## FFMC_DC3    -2.580e-04  2.090e-04  -1.235   0.2182
## FFMC_ISI3   -7.054e-03  8.738e-03  -0.807   0.4203
## DMC_DC3      1.141e-05  1.071e-05   1.065   0.2879
## DMC_ISI3    -5.306e-04  6.996e-04  -0.758   0.4489
## DC_ISI3     -2.919e-05  2.018e-04  -0.145   0.8851
## temp_RH3    -1.587e-03  9.931e-04  -1.598   0.1114
## temp_wind3   5.501e-03  7.805e-03   0.705   0.4817
## RH_wind3    -2.648e-03  3.226e-03  -0.821   0.4127
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.169 on 239 degrees of freedom
## Multiple R-squared:  0.08054,    Adjusted R-squared:  0.003603
## F-statistic: 1.047 on 20 and 239 DF,  p-value: 0.4081
```



It seems that after discarding the influential points, my model can explain more of the variation. However I am not quite sure whether it is legitmate to drop this points because I can not say these data are wrong. I did this just due to the assumption that a linear model is appropriate and my model is correct.

My final attempt is to add some quadratic form of the FWI and weather variables into the model. I am not sure if these are helpful but if they are they may improve my model and if not they will be dropped when I shrink our model use AIC method.

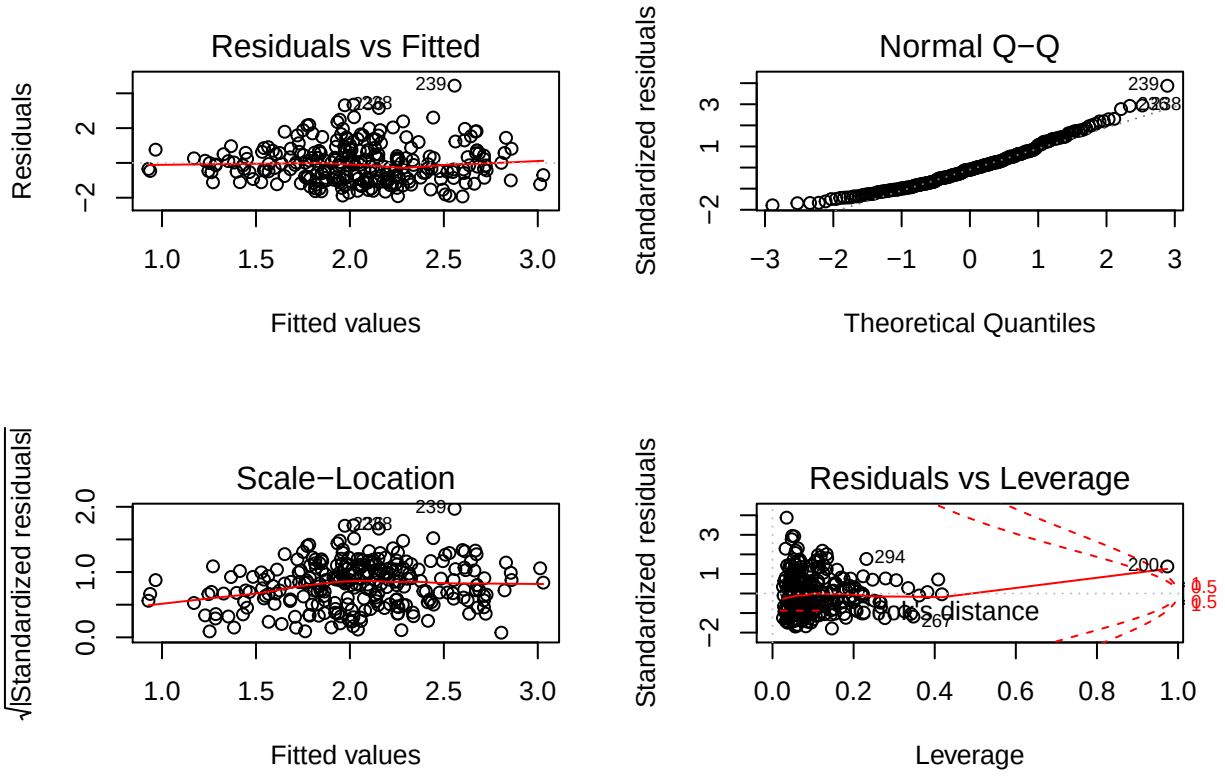
- fullmodel2.lm <- lm(log(data3$area+1)~X+Y+month+day+FFMC_LOG3+DMC+DC+ISI +temp+RH+wind+FFMC_DMC3+FFMC_DC3+FFMC_ISI3 +DMC_DC3+DMC_ISI3 +DC_ISI3+temp_RH3+temp_wind3 +RH_wind3+FFMC_SQ+DMC_SQ+DC_SQ+ISI_SQ +temp_SQ +wind_SQ+RH_SQ,data=data3)

```
##
## Call:
## lm(formula = log(data3$area + 1) ~ X + Y + month + day + FFMC_LOG3 +
##     DMC + DC + ISI + temp + RH + wind + FFMC_DMC3 + FFMC_DC3 +
##     FFMC_ISI3 + DMC_DC3 + DMC_ISI3 + DC_ISI3 + temp_RH3 + temp_wind3 +
##     RH_wind3 + FFMC_SQ + DMC_SQ + DC_SQ + ISI_SQ + temp_SQ +
##     wind_SQ + RH_SQ, data = data3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.9266 -0.8579 -0.1080  0.6205  4.4385
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  6.471e+00  1.348e+02   0.048   0.9617
## X            3.399e-02  3.844e-02   0.884   0.3774
## Y           -9.609e-02  8.012e-02  -1.199   0.2316
## month        2.792e-01  1.361e-01   2.052   0.0413 *
```

```
## day          6.576e-02  3.973e-02   1.655   0.0992 .
## FFMC_LOG3   -1.343e+00  3.471e+01  -0.039   0.9692
## DMC         -1.708e-01  1.167e-01  -1.463   0.1448
## DC           1.013e-02  2.000e-02   0.507   0.6130
## ISI          8.226e-01  1.511e+00   0.544   0.5867
## temp         7.396e-02  1.767e-01   0.419   0.6759
## RH          -1.860e-02  5.091e-02  -0.365   0.7152
## wind         3.842e-01  3.242e-01   1.185   0.2373
## FFMC_DMC3    2.041e-03  1.312e-03   1.556   0.1211
## FFMC_DC3    -2.050e-04  2.275e-04  -0.901   0.3684
## FFMC_ISI3   -8.432e-03  1.766e-02  -0.478   0.6334
## DMC_DC3     -1.347e-05  2.536e-05  -0.531   0.5958
## DMC_ISI3    -5.158e-04  8.158e-04  -0.632   0.5278
## DC_ISI3     -6.666e-05  2.075e-04  -0.321   0.7483
## temp_RH3    -8.928e-04  1.322e-03  -0.676   0.5000
## temp_wind3   2.035e-04  8.900e-03   0.023   0.9818
## RH_wind3    -2.516e-03  3.429e-03  -0.734   0.4639
## FFMC_SQ      5.193e-06  2.836e-03   0.002   0.9985
## DMC_SQ       4.674e-06  3.956e-05   0.118   0.9061
## DC_SQ        8.126e-06  4.626e-06   1.757   0.0803 .
## ISI_SQ       1.502e-03  7.620e-03   0.197   0.8439
## temp_SQ     -1.134e-03  3.056e-03  -0.371   0.7110
## wind_SQ     -3.291e-02  1.884e-02  -1.747   0.0820 .
## RH_SQ        3.517e-04  3.439e-04   1.023   0.3076
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.166 on 232 degrees of freedom
## Multiple R-squared:  0.1125, Adjusted R-squared:  0.009187
## F-statistic: 1.089 on 27 and 232 DF,  p-value: 0.3539

## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced

## Warning in sqrt(crit * p * (1 - hh)/hh): NaNs produced
```



Finally, let's run the step function to see what the final model would be.

```
## Start:  AIC=106.08
## log(data3$area + 1) ~ X + Y + month + day + FFMC_LOG3 + DMC +
##     DC + ISI + temp + RH + wind + FFMC_DMC3 + FFMC_DC3 + FFMC_ISI3 +
##     DMC_DC3 + DMC_ISI3 + DC_ISI3 + temp_RH3 + temp_wind3 + RH_wind3 +
##     FFMC_SQ + DMC_SQ + DC_SQ + ISI_SQ + temp_SQ + wind_SQ + RH_SQ
##
##               Df Sum of Sq    RSS    AIC
## - FFMC_SQ      1    0.0000 315.23 104.08
## - temp_wind3   1    0.0007 315.23 104.08
## - FFMC_LOG3    1    0.0020 315.23 104.08
## - DMC_SQ       1    0.0190 315.25 104.09
## - ISI_SQ       1    0.0528 315.28 104.12
## - DC_ISI3      1    0.1402 315.37 104.19
## - RH           1    0.1813 315.41 104.23
## - temp_SQ      1    0.1869 315.41 104.23
## - temp         1    0.2382 315.47 104.28
## - FFMC_ISI3    1    0.3098 315.54 104.33
## - DC           1    0.3486 315.58 104.37
## - DMC_DC3      1    0.3833 315.61 104.39
## - ISI          1    0.4028 315.63 104.41
## - DMC_ISI3     1    0.5431 315.77 104.53
## - temp_RH3     1    0.6201 315.85 104.59
## - RH_wind3     1    0.7314 315.96 104.68
## - X            1    1.0625 316.29 104.95
## - FFMC_DC3     1    1.1035 316.33 104.99
## - RH_SQ        1    1.4209 316.65 105.25
## - wind         1    1.9075 317.13 105.65
## - Y            1    1.9546 317.18 105.69
```

```
## <none>                    315.23 106.08
## - DMC          1    2.9090 318.14 106.47
## - FFMC_DMC3    1    3.2892 318.52 106.78
## - day          1    3.7226 318.95 107.13
## - wind_SQ      1    4.1465 319.37 107.48
## - DC_SQ        1    4.1934 319.42 107.52
## - month        1    5.7219 320.95 108.76
##
## Step:  AIC=104.08
## log(data3$area + 1) ~ X + Y + month + day + FFMC_LOG3 + DMC +
##     DC + ISI + temp + RH + wind + FFMC_DMC3 + FFMC_DC3 + FFMC_ISI3 +
##     DMC_DC3 + DMC_ISI3 + DC_ISI3 + temp_RH3 + temp_wind3 + RH_wind3 +
##     DMC_SQ + DC_SQ + ISI_SQ + temp_SQ + wind_SQ + RH_SQ
##
##               Df Sum of Sq    RSS    AIC
## - temp_wind3   1    0.0007 315.23 102.08
## - DMC_SQ       1    0.0200 315.25 102.10
## - FFMC_LOG3    1    0.0448 315.27 102.12
## - ISI_SQ       1    0.0872 315.31 102.15
## - DC_ISI3      1    0.1404 315.37 102.19
## - RH           1    0.1815 315.41 102.23
## - temp_SQ      1    0.2047 315.43 102.25
## - temp         1    0.2556 315.48 102.29
## - DC           1    0.3615 315.59 102.38
## - DMC_DC3      1    0.3834 315.61 102.39
## - DMC_ISI3     1    0.5700 315.80 102.55
## - temp_RH3     1    0.6304 315.86 102.60
## - RH_wind3     1    0.7409 315.97 102.69
## - FFMC_ISI3    1    0.8388 316.07 102.77
## - ISI          1    0.9825 316.21 102.89
## - X            1    1.0973 316.32 102.98
## - FFMC_DC3     1    1.1700 316.40 103.04
## - RH_SQ        1    1.4281 316.66 103.25
## - wind         1    1.9275 317.15 103.66
## - Y            1    2.0032 317.23 103.73
## <none>                     315.23 104.08
## - DMC          1    2.9528 318.18 104.50
## - FFMC_DMC3    1    3.3727 318.60 104.85
## - day          1    3.7228 318.95 105.13
## - wind_SQ      1    4.1465 319.37 105.48
## - DC_SQ        1    4.3553 319.58 105.65
## - month        1    5.8107 321.04 106.83
##
## Step:  AIC=102.08
## log(data3$area + 1) ~ X + Y + month + day + FFMC_LOG3 + DMC +
##     DC + ISI + temp + RH + wind + FFMC_DMC3 + FFMC_DC3 + FFMC_ISI3 +
##     DMC_DC3 + DMC_ISI3 + DC_ISI3 + temp_RH3 + RH_wind3 + DMC_SQ +
##     DC_SQ + ISI_SQ + temp_SQ + wind_SQ + RH_SQ
##
##              Df Sum of Sq    RSS    AIC
## - DMC_SQ      1    0.0201 315.25 100.10
## - FFMC_LOG3   1    0.0445 315.27 100.12
## - ISI_SQ      1    0.0889 315.32 100.15
## - DC_ISI3     1    0.1405 315.37 100.20
```

```
## - RH         1    0.1927 315.42 100.24
## - temp_SQ    1    0.2423 315.47 100.28
## - DC         1    0.3610 315.59 100.38
## - temp       1    0.3680 315.60 100.38
## - DMC_DC3    1    0.3878 315.62 100.40
## - DMC_ISI3   1    0.5696 315.80 100.55
## - temp_RH3   1    0.6750 315.90 100.64
## - RH_wind3   1    0.8212 316.05 100.76
## - FFMC_ISI3  1    0.8472 316.08 100.78
## - ISI        1    0.9973 316.23 100.90
## - X          1    1.1122 316.34 101.00
## - FFMC_DC3   1    1.1693 316.40 101.04
## - RH_SQ      1    1.4529 316.68 101.28
## - Y          1    2.0223 317.25 101.74
## <none>                   315.23 102.08
## - DMC        1    2.9666 318.19 102.52
## - FFMC_DMC3  1    3.3863 318.61 102.86
## - day        1    3.9324 319.16 103.30
## - DC_SQ      1    4.3795 319.61 103.67
## - wind_SQ    1    4.4513 319.68 103.73
## - wind       1    4.7954 320.02 104.00
## - month      1    5.9490 321.18 104.94
##
## Step:  AIC=100.1
## log(data3$area + 1) ~ X + Y + month + day + FFMC_LOG3 + DMC +
##     DC + ISI + temp + RH + wind + FFMC_DMC3 + FFMC_DC3 + FFMC_ISI3 +
##     DMC_DC3 + DMC_ISI3 + DC_ISI3 + temp_RH3 + RH_wind3 + DC_SQ +
##     ISI_SQ + temp_SQ + wind_SQ + RH_SQ
##
##             Df Sum of Sq    RSS     AIC
## - FFMC_LOG3  1    0.0372 315.29  98.127
## - ISI_SQ     1    0.1131 315.36  98.190
## - DC_ISI3    1    0.1468 315.40  98.218
## - RH         1    0.1869 315.44  98.251
## - temp_SQ    1    0.2574 315.51  98.309
## - temp       1    0.3904 315.64  98.418
## - DC         1    0.4144 315.66  98.438
## - DMC_DC3    1    0.5884 315.84  98.581
## - DMC_ISI3   1    0.6005 315.85  98.591
## - temp_RH3   1    0.6973 315.95  98.671
## - RH_wind3   1    0.8504 316.10  98.797
## - FFMC_ISI3  1    0.8893 316.14  98.829
## - ISI        1    1.0377 316.29  98.951
## - X          1    1.2258 316.47  99.106
## - FFMC_DC3   1    1.3169 316.57  99.180
## - RH_SQ      1    1.4697 316.72  99.306
## - Y          1    2.0505 317.30  99.782
## <none>                   315.25 100.097
## - DMC        1    3.1919 318.44 100.716
## - FFMC_DMC3  1    3.6216 318.87 101.066
## - day        1    3.9548 319.20 101.338
## - DC_SQ      1    4.4789 319.73 101.764
## - wind_SQ    1    4.5416 319.79 101.815
## - wind       1    4.9946 320.24 102.184
```

```
## - month      1    6.2534 321.50 103.203
##
## Step:  AIC=98.13
## log(data3$area + 1) ~ X + Y + month + day + DMC + DC + ISI +
##     temp + RH + wind + FFMC_DMC3 + FFMC_DC3 + FFMC_ISI3 + DMC_DC3 +
##     DMC_ISI3 + DC_ISI3 + temp_RH3 + RH_wind3 + DC_SQ + ISI_SQ +
##     temp_SQ + wind_SQ + RH_SQ
##
##             Df Sum of Sq    RSS     AIC
## - DC_ISI3    1    0.1098 315.40  96.218
## - ISI_SQ     1    0.1432 315.43  96.245
## - RH         1    0.1637 315.45  96.262
## - temp_SQ    1    0.2361 315.52  96.322
## - temp       1    0.3851 315.67  96.445
## - DMC_DC3    1    0.5609 315.85  96.589
## - DMC_ISI3   1    0.6391 315.92  96.654
## - temp_RH3   1    0.8366 316.12  96.816
## - RH_wind3   1    0.8690 316.15  96.843
## - FFMC_ISI3  1    0.9691 316.25  96.925
## - ISI        1    1.0752 316.36  97.012
## - DC         1    1.1249 316.41  97.053
## - X          1    1.2097 316.50  97.123
## - RH_SQ      1    1.4485 316.73  97.319
## - Y          1    2.0377 317.32  97.802
## <none>                   315.29  98.127
## - FFMC_DC3   1    3.1206 318.41  98.688
## - DMC        1    3.5025 318.79  99.000
## - day        1    3.9177 319.20  99.338
## - FFMC_DMC3  1    3.9278 319.21  99.346
## - DC_SQ      1    4.4416 319.73  99.764
## - wind_SQ    1    4.5072 319.79  99.818
## - wind       1    4.9713 320.26 100.195
## - month      1    6.3310 321.62 101.296
##
## Step:  AIC=96.22
## log(data3$area + 1) ~ X + Y + month + day + DMC + DC + ISI +
##     temp + RH + wind + FFMC_DMC3 + FFMC_DC3 + FFMC_ISI3 + DMC_DC3 +
##     DMC_ISI3 + temp_RH3 + RH_wind3 + DC_SQ + ISI_SQ + temp_SQ +
##     wind_SQ + RH_SQ
##
##             Df Sum of Sq    RSS    AIC
## - ISI_SQ     1    0.1299 315.53 94.325
## - RH         1    0.1966 315.59 94.380
## - temp_SQ    1    0.2695 315.66 94.440
## - temp       1    0.4112 315.81 94.557
## - DMC_DC3    1    0.7139 316.11 94.806
## - temp_RH3   1    0.7897 316.19 94.868
## - RH_wind3   1    0.8945 316.29 94.954
## - FFMC_ISI3  1    1.0284 316.42 95.064
## - ISI        1    1.1002 316.50 95.123
## - X          1    1.1557 316.55 95.169
## - DMC_ISI3   1    1.4961 316.89 95.448
## - DC         1    1.5276 316.92 95.474
## - RH_SQ      1    1.5709 316.97 95.510
```

```
## - Y          1    1.9414 317.34 95.813
## <none>                   315.40 96.218
## - day        1    3.9075 319.30 97.419
## - DMC        1    4.0471 319.44 97.533
## - wind_SQ    1    4.3999 319.80 97.820
## - FFMC_DC3   1    4.5903 319.99 97.975
## - DC_SQ      1    4.6674 320.06 98.037
## - FFMC_DMC3  1    4.7389 320.13 98.095
## - wind       1    4.8726 320.27 98.204
## - month      1    6.4427 321.84 99.475
##
## Step:  AIC=94.32
## log(data3$area + 1) ~ X + Y + month + day + DMC + DC + ISI +
##     temp + RH + wind + FFMC_DMC3 + FFMC_DC3 + FFMC_ISI3 + DMC_DC3 +
##     DMC_ISI3 + temp_RH3 + RH_wind3 + DC_SQ + temp_SQ + wind_SQ +
##     RH_SQ
##
##             Df Sum of Sq    RSS    AIC
## - RH         1    0.1735 315.70 92.468
## - temp_SQ    1    0.2378 315.76 92.521
## - temp       1    0.3912 315.92 92.647
## - DMC_DC3    1    0.6335 316.16 92.846
## - temp_RH3   1    0.8266 316.35 93.005
## - FFMC_ISI3  1    0.9420 316.47 93.100
## - RH_wind3   1    0.9505 316.48 93.107
## - ISI        1    0.9853 316.51 93.136
## - X          1    1.2210 316.75 93.329
## - DC         1    1.4567 316.98 93.522
## - DMC_ISI3   1    1.4946 317.02 93.554
## - RH_SQ      1    1.5445 317.07 93.594
## - Y          1    1.9549 317.48 93.931
## <none>                   315.53 94.325
## - day        1    3.8046 319.33 95.441
## - DMC        1    4.0986 319.62 95.680
## - wind_SQ    1    4.2957 319.82 95.841
## - FFMC_DC3   1    4.4815 320.01 95.992
## - DC_SQ      1    4.6074 320.13 96.094
## - wind       1    4.8687 320.39 96.306
## - FFMC_DMC3  1    4.9758 320.50 96.393
## - month      1    6.4815 322.01 97.612
##
## Step:  AIC=92.47
## log(data3$area + 1) ~ X + Y + month + day + DMC + DC + ISI +
##     temp + wind + FFMC_DMC3 + FFMC_DC3 + FFMC_ISI3 + DMC_DC3 +
##     DMC_ISI3 + temp_RH3 + RH_wind3 + DC_SQ + temp_SQ + wind_SQ +
##     RH_SQ
##
##             Df Sum of Sq    RSS    AIC
## - temp_SQ    1    0.4150 316.11 90.809
## - DMC_DC3    1    0.6142 316.31 90.973
## - FFMC_ISI3  1    0.8874 316.59 91.198
## - temp       1    0.9188 316.62 91.223
## - ISI        1    0.9283 316.63 91.231
## - X          1    1.2415 316.94 91.488
```

```
## - DC         1    1.4621 317.16 91.669
## - DMC_ISI3   1    1.4973 317.20 91.698
## - RH_wind3   1    1.6081 317.31 91.789
## - Y          1    2.0157 317.71 92.123
## - RH_SQ      1    2.0351 317.73 92.138
## <none>                   315.70 92.468
## - temp_RH3   1    3.4063 319.11 93.258
## - day        1    3.6701 319.37 93.473
## - wind_SQ    1    4.1274 319.83 93.845
## - DMC        1    4.1654 319.86 93.876
## - FFMC_DC3   1    4.5711 320.27 94.205
## - DC_SQ      1    4.7077 320.41 94.316
## - FFMC_DMC3  1    5.0513 320.75 94.595
## - wind       1    5.3513 321.05 94.838
## - month      1    7.3510 323.05 96.452
##
## Step:  AIC=90.81
## log(data3$area + 1) ~ X + Y + month + day + DMC + DC + ISI +
##     temp + wind + FFMC_DMC3 + FFMC_DC3 + FFMC_ISI3 + DMC_DC3 +
##     DMC_ISI3 + temp_RH3 + RH_wind3 + DC_SQ + wind_SQ + RH_SQ
##
##             Df Sum of Sq    RSS    AIC
## - FFMC_ISI3  1    0.8634 316.98 89.518
## - ISI        1    0.9305 317.04 89.574
## - DMC_DC3    1    1.0392 317.15 89.663
## - X          1    1.3345 317.45 89.905
## - RH_wind3   1    1.4393 317.55 89.990
## - DC         1    1.4485 317.56 89.998
## - temp       1    1.5679 317.68 90.096
## - RH_SQ      1    1.6205 317.73 90.139
## - DMC_ISI3   1    2.0783 318.19 90.513
## - Y          1    2.0904 318.20 90.523
## <none>                   316.11 90.809
## - temp_RH3   1    3.2268 319.34 91.450
## - day        1    3.8117 319.93 91.926
## - DMC        1    3.8624 319.98 91.967
## - FFMC_DC3   1    4.4111 320.52 92.412
## - wind_SQ    1    4.4124 320.53 92.413
## - FFMC_DMC3  1    4.7930 320.91 92.722
## - DC_SQ      1    4.7981 320.91 92.726
## - wind       1    5.2621 321.38 93.102
## - month      1    7.1343 323.25 94.612
##
## Step:  AIC=89.52
## log(data3$area + 1) ~ X + Y + month + day + DMC + DC + ISI +
##     temp + wind + FFMC_DMC3 + FFMC_DC3 + DMC_DC3 + DMC_ISI3 +
##     temp_RH3 + RH_wind3 + DC_SQ + wind_SQ + RH_SQ
##
##             Df Sum of Sq    RSS    AIC
## - ISI        1    0.2517 317.23 87.725
## - DC         1    0.8414 317.82 88.208
## - RH_wind3   1    0.9799 317.96 88.321
## - temp       1    1.1440 318.12 88.455
## - RH_SQ      1    1.1654 318.14 88.473
```

```
## - X          1    1.3110 318.29 88.592
## - DMC_ISI3   1    1.5640 318.54 88.798
## - DMC_DC3    1    1.5846 318.56 88.815
## - Y          1    1.7824 318.76 88.976
## <none>                   316.98 89.518
## - temp_RH3   1    2.9236 319.90 89.906
## - DMC        1    3.0910 320.07 90.042
## - FFMC_DC3   1    3.5763 320.55 90.436
## - wind_SQ    1    4.0412 321.02 90.812
## - FFMC_DMC3  1    4.1175 321.09 90.874
## - wind       1    4.5781 321.56 91.247
## - day        1    4.8334 321.81 91.453
## - DC_SQ      1    5.5126 322.49 92.001
## - month      1    7.5015 324.48 93.600
##
## Step:  AIC=87.72
## log(data3$area + 1) ~ X + Y + month + day + DMC + DC + temp +
##     wind + FFMC_DMC3 + FFMC_DC3 + DMC_DC3 + DMC_ISI3 + temp_RH3 +
##     RH_wind3 + DC_SQ + wind_SQ + RH_SQ
##
##             Df Sum of Sq    RSS    AIC
## - DC         1    0.5898 317.82 86.208
## - RH_wind3   1    0.9048 318.13 86.465
## - RH_SQ      1    1.0112 318.24 86.552
## - temp       1    1.1410 318.37 86.658
## - X          1    1.3877 318.62 86.860
## - Y          1    1.8439 319.07 87.232
## - DMC_DC3    1    1.8705 319.10 87.253
## <none>                   317.23 87.725
## - temp_RH3   1    2.7374 319.97 87.959
## - DMC        1    3.0692 320.30 88.228
## - DMC_ISI3   1    3.2267 320.46 88.356
## - FFMC_DC3   1    3.9754 321.20 88.963
## - wind_SQ    1    4.2292 321.46 89.168
## - FFMC_DMC3  1    4.3524 321.58 89.268
## - wind       1    4.7308 321.96 89.574
## - day        1    5.1430 322.37 89.906
## - DC_SQ      1    5.4265 322.66 90.135
## - month      1    7.3791 324.61 91.703
##
## Step:  AIC=86.21
## log(data3$area + 1) ~ X + Y + month + day + DMC + temp + wind +
##     FFMC_DMC3 + FFMC_DC3 + DMC_DC3 + DMC_ISI3 + temp_RH3 + RH_wind3 +
##     DC_SQ + wind_SQ + RH_SQ
##
##             Df Sum of Sq    RSS    AIC
## - RH_wind3   1    0.7814 318.60 84.846
## - RH_SQ      1    0.9581 318.78 84.990
## - X          1    1.3710 319.19 85.327
## - temp       1    1.6290 319.45 85.537
## - Y          1    1.7305 319.55 85.620
## <none>                   317.82 86.208
## - DMC_DC3    1    2.4814 320.30 86.230
## - temp_RH3   1    2.5602 320.38 86.294
```

```
## - DMC_ISI3   1    2.7555 320.57 86.452
## - wind_SQ    1    4.0299 321.85 87.484
## - DMC        1    4.1730 321.99 87.599
## - wind       1    4.4129 322.23 87.793
## - day        1    5.0551 322.87 88.311
## - FFMC_DMC3  1    6.4356 324.25 89.420
## - DC_SQ      1    7.7861 325.60 90.501
## - FFMC_DC3   1   10.3101 328.13 92.508
## - month      1   12.7053 330.52 94.399
##
## Step:  AIC=84.85
## log(data3$area + 1) ~ X + Y + month + day + DMC + temp + wind +
##     FFMC_DMC3 + FFMC_DC3 + DMC_DC3 + DMC_ISI3 + temp_RH3 + DC_SQ +
##     wind_SQ + RH_SQ
##
##             Df Sum of Sq    RSS    AIC
## - RH_SQ      1    0.2130 318.81 83.020
## - temp       1    1.2858 319.89 83.893
## - X          1    1.4579 320.06 84.033
## - Y          1    1.9881 320.59 84.464
## - temp_RH3   1    2.4353 321.04 84.826
## <none>                   318.60 84.846
## - DMC_ISI3   1    2.5104 321.11 84.887
## - DMC_DC3    1    2.6518 321.25 85.001
## - wind       1    3.7628 322.36 85.899
## - DMC        1    4.1465 322.75 86.208
## - wind_SQ    1    4.3843 322.98 86.400
## - day        1    5.2798 323.88 87.120
## - FFMC_DMC3  1    6.4678 325.07 88.072
## - DC_SQ      1    7.8221 326.42 89.153
## - FFMC_DC3   1   10.5367 329.14 91.306
## - month      1   13.8309 332.43 93.895
##
## Step:  AIC=83.02
## log(data3$area + 1) ~ X + Y + month + day + DMC + temp + wind +
##     FFMC_DMC3 + FFMC_DC3 + DMC_DC3 + DMC_ISI3 + temp_RH3 + DC_SQ +
##     wind_SQ
##
##             Df Sum of Sq    RSS    AIC
## - X          1    1.4376 320.25 82.190
## - temp       1    1.6202 320.43 82.338
## - Y          1    1.9009 320.71 82.566
## - DMC_ISI3   1    2.3943 321.21 82.965
## <none>                   318.81 83.020
## - DMC_DC3    1    2.4838 321.30 83.038
## - temp_RH3   1    3.8493 322.66 84.140
## - wind       1    4.1007 322.91 84.343
## - DMC        1    4.1138 322.93 84.354
## - wind_SQ    1    4.7791 323.59 84.889
## - day        1    5.3040 324.12 85.310
## - FFMC_DMC3  1    6.3530 325.17 86.150
## - DC_SQ      1    7.6349 326.45 87.173
## - FFMC_DC3   1   10.3514 329.16 89.328
## - month      1   13.6213 332.43 91.898
```

```
##
## Step:  AIC=82.19
## log(data3$area + 1) ~ Y + month + day + DMC + temp + wind + FFMC_DMC3 +
##     FFMC_DC3 + DMC_DC3 + DMC_ISI3 + temp_RH3 + DC_SQ + wind_SQ
##
##             Df Sum of Sq    RSS    AIC
## - Y          1    0.8044 321.06 80.842
## - temp       1    1.4845 321.74 81.392
## - DMC_DC3    1    2.4439 322.69 82.166
## <none>                   320.25 82.190
## - DMC_ISI3   1    2.5510 322.80 82.253
## - temp_RH3   1    3.5744 323.83 83.076
## - DMC        1    3.9182 324.17 83.351
## - wind       1    4.2985 324.55 83.656
## - wind_SQ    1    4.8813 325.13 84.123
## - day        1    5.0270 325.28 84.239
## - FFMC_DMC3  1    6.0889 326.34 85.087
## - DC_SQ      1    7.3549 327.61 86.093
## - FFMC_DC3   1    9.9043 330.15 88.109
## - month      1   12.6748 332.93 90.282
##
## Step:  AIC=80.84
## log(data3$area + 1) ~ month + day + DMC + temp + wind + FFMC_DMC3 +
##     FFMC_DC3 + DMC_DC3 + DMC_ISI3 + temp_RH3 + DC_SQ + wind_SQ
##
##             Df Sum of Sq    RSS    AIC
## - temp       1    1.4526 322.51 80.016
## - DMC_ISI3   1    2.3222 323.38 80.716
## <none>                   321.06 80.842
## - DMC_DC3    1    2.5620 323.62 80.909
## - temp_RH3   1    3.4624 324.52 81.631
## - DMC        1    3.9049 324.96 81.985
## - wind       1    4.1530 325.21 82.184
## - wind_SQ    1    4.6487 325.70 82.580
## - day        1    5.0085 326.06 82.867
## - FFMC_DMC3  1    6.0700 327.13 83.712
## - DC_SQ      1    7.7857 328.84 85.072
## - FFMC_DC3   1   10.0919 331.15 86.889
## - month      1   12.2798 333.33 88.601
##
## Step:  AIC=80.02
## log(data3$area + 1) ~ month + day + DMC + wind + FFMC_DMC3 +
##     FFMC_DC3 + DMC_DC3 + DMC_ISI3 + temp_RH3 + DC_SQ + wind_SQ
##
##             Df Sum of Sq    RSS    AIC
## - DMC_ISI3   1    2.0365 324.54 79.652
## <none>                   322.51 80.016
## - temp_RH3   1    3.3045 325.81 80.666
## - DMC_DC3    1    3.4017 325.91 80.744
## - DMC        1    3.6484 326.16 80.941
## - wind       1    4.2952 326.80 81.456
## - day        1    5.1475 327.66 82.133
## - wind_SQ    1    5.1608 327.67 82.143
## - FFMC_DMC3  1    6.0410 328.55 82.841
```

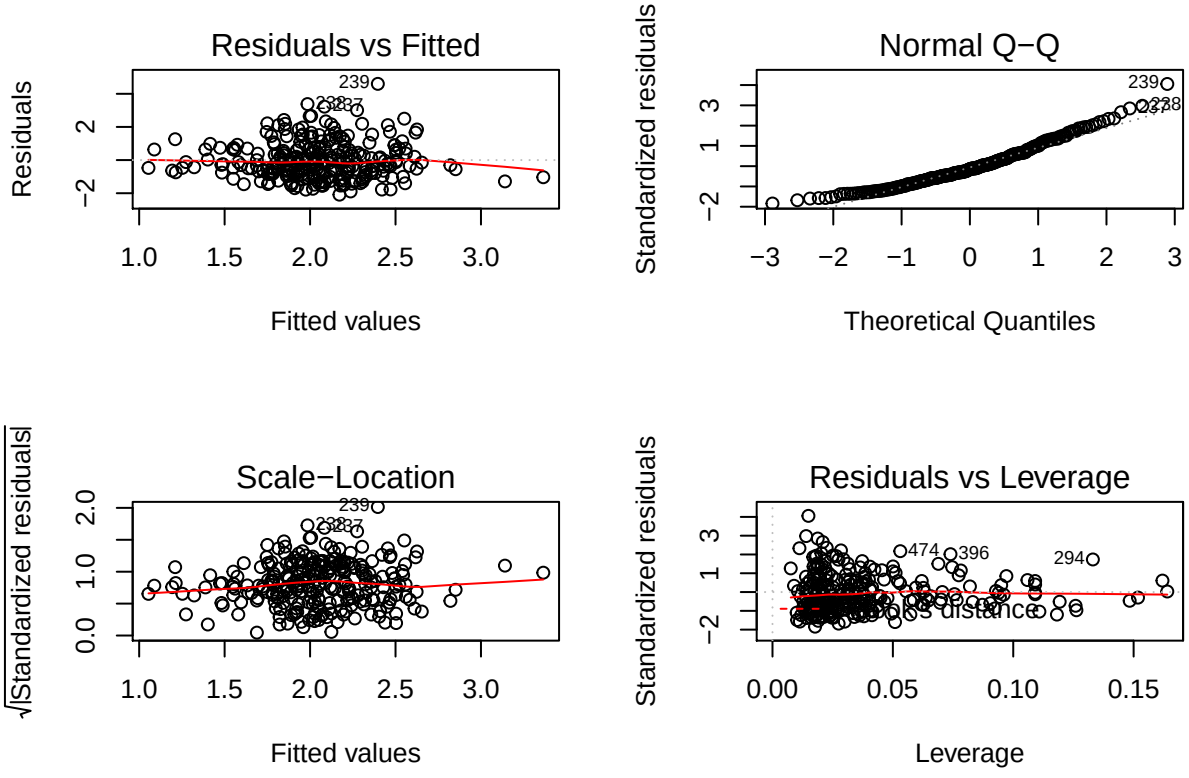
```
## - DC_SQ      1    7.0711 329.58 83.655
## - FFMC_DC3   1    8.8630 331.37 85.065
## - month      1   10.9532 333.46 86.699
##
## Step:  AIC=79.65
## log(data3$area + 1) ~ month + day + DMC + wind + FFMC_DMC3 +
##     FFMC_DC3 + DMC_DC3 + temp_RH3 + DC_SQ + wind_SQ
##
##             Df Sum of Sq    RSS    AIC
## - DMC        1    2.0583 326.60 79.296
## <none>                   324.54 79.652
## - DMC_DC3    1    3.2751 327.82 80.263
## - temp_RH3   1    3.7302 328.27 80.624
## - wind       1    3.7575 328.30 80.645
## - FFMC_DMC3  1    4.1147 328.66 80.928
## - wind_SQ    1    4.9417 329.49 81.581
## - day        1    6.2137 330.76 82.583
## - DC_SQ      1    7.9526 332.50 83.947
## - FFMC_DC3   1   10.1123 334.66 85.630
## - month      1   12.2074 336.75 87.253
##
## Step:  AIC=79.3
## log(data3$area + 1) ~ month + day + wind + FFMC_DMC3 + FFMC_DC3 +
##     DMC_DC3 + temp_RH3 + DC_SQ + wind_SQ
##
##             Df Sum of Sq    RSS    AIC
## <none>                   326.60 79.296
## - DMC_DC3    1    3.3975 330.00 79.987
## - wind       1    3.8196 330.42 80.319
## - FFMC_DMC3  1    4.5722 331.17 80.911
## - wind_SQ    1    4.8670 331.47 81.142
## - day        1    5.9836 332.59 82.016
## - DC_SQ      1    6.0790 332.68 82.091
## - temp_RH3   1    6.5348 333.14 82.447
## - FFMC_DC3   1    8.0546 334.66 83.630
## - month      1   10.1747 336.78 85.272
##
## Call:
## lm(formula = log(data3$area + 1) ~ month + day + wind + FFMC_DMC3 +
##     FFMC_DC3 + DMC_DC3 + temp_RH3 + DC_SQ + wind_SQ, data = data3)
##
## Coefficients:
## (Intercept)        month          day         wind    FFMC_DMC3
##   1.134e+00    2.268e-01    7.659e-02    2.643e-01    2.124e-04
##    FFMC_DC3      DMC_DC3     temp_RH3        DC_SQ      wind_SQ
##  -8.510e-05   -2.231e-05   -8.190e-04    7.723e-06   -3.131e-02
```

Now let's check our final model.
* finalmodel.lm <-lm(log(data3$area + 1) ~ month + day + wind + FFMC_DMC3 + FFMC_DC3 + DMC_DC3+ temp_RH3 + DC_SQ + wind_SQ, data = data3)

Although the r-squared seems to be low,it is still much better than my origional model.

```
##
## Call:
```

```
## lm(formula = log(data3$area + 1) ~ month + day + wind + FFMC_DMC3 +
##     FFMC_DC3 + DMC_DC3 + temp_RH3 + DC_SQ + wind_SQ, data = data3)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.0859 -0.7961 -0.1639  0.6493  4.5991
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.134e+00  5.409e-01   2.097  0.03696 *
## month        2.268e-01  8.126e-02   2.791  0.00566 **
## day          7.659e-02  3.579e-02   2.140  0.03331 *
## wind         2.643e-01  1.546e-01   1.710  0.08853 .
## FFMC_DMC3    2.124e-04  1.135e-04   1.871  0.06254 .
## FFMC_DC3    -8.510e-05  3.427e-05  -2.483  0.01368 *
## DMC_DC3     -2.231e-05  1.384e-05  -1.613  0.10808
## temp_RH3    -8.190e-04  3.662e-04  -2.237  0.02620 *
## DC_SQ        7.723e-06  3.580e-06   2.157  0.03195 *
## wind_SQ     -3.131e-02  1.622e-02  -1.930  0.05472 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.143 on 250 degrees of freedom
## Multiple R-squared:  0.08045,    Adjusted R-squared:  0.04735
## F-statistic:  2.43 on 9 and 250 DF,  p-value: 0.01155
```



Conclusion

unfortunately, my model can only explain a tiny proportation of the total variation. The possible explaination are either the data itself can only explained that much information or a linear model is not sufficient to do the work. So, maybe I shall turn to nonlinear regression to solve the problem. Anyway, based on my model, month, day,DMC,DC,FFMC,RH wind are the influential variables that may affect the size of a forest fires.