

Project Performance

Alison Jing Huang

4/21/2018

Load the data

```
##   X Y month day FFMC  DMC    DC  ISI temp RH wind rain area
## 1 7 5     8   1 86.2 26.2  94.3  5.1  8.2 51  6.7  0.0   0
## 2 7 4    11   6 90.6 35.4 669.1  6.7 18.0 33  0.9  0.0   0
## 3 7 4    11   3 90.6 43.7 686.9  6.7 14.6 33  1.3  0.0   0
## 4 8 6     8   1 91.7 33.3  77.5  9.0  8.3 97  4.0  0.2   0
## 5 8 6     8   4 89.3 51.3 102.2  9.6 11.4 99  1.8  0.0   0
## 6 8 6     2   4 92.3 85.3 488.0 14.7 22.2 29  5.4  0.0   0
```

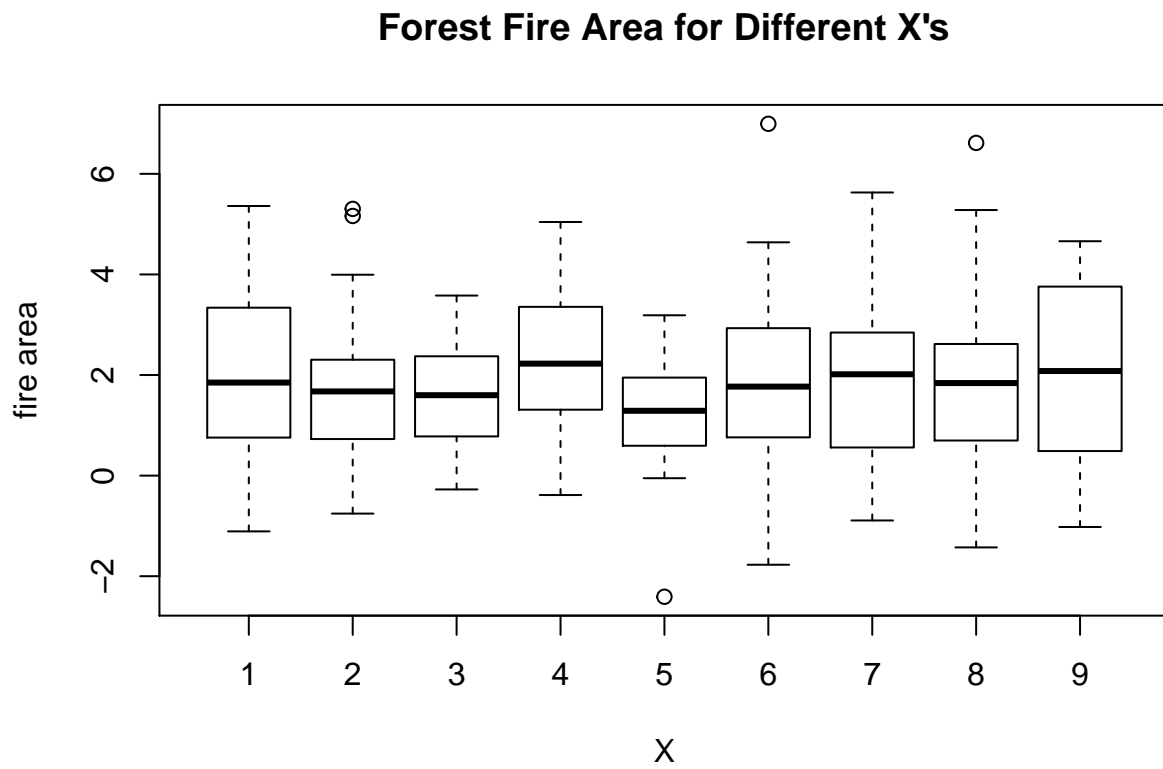
check the likelihood of Fire Observations

```
##
## FALSE  TRUE
##  0.52  0.48
```

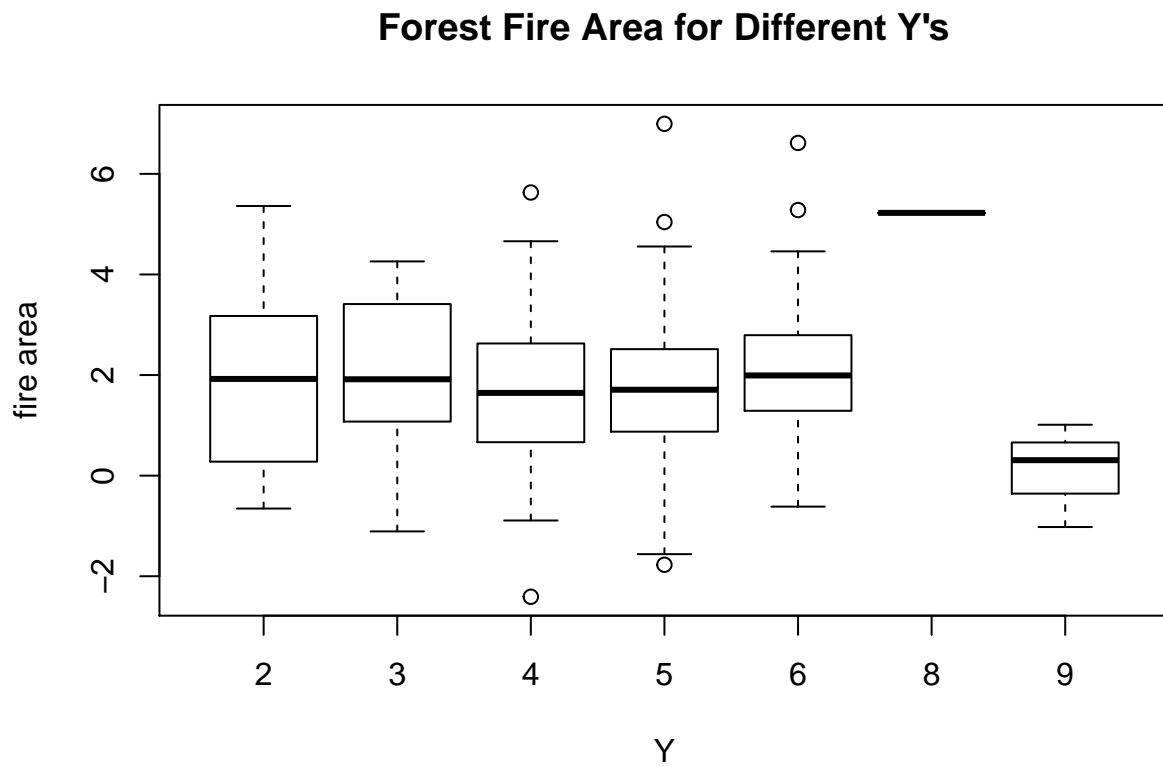
The result shows that there is 48% percent of the time there is no observation of a forest fire.

Linear Regression Models

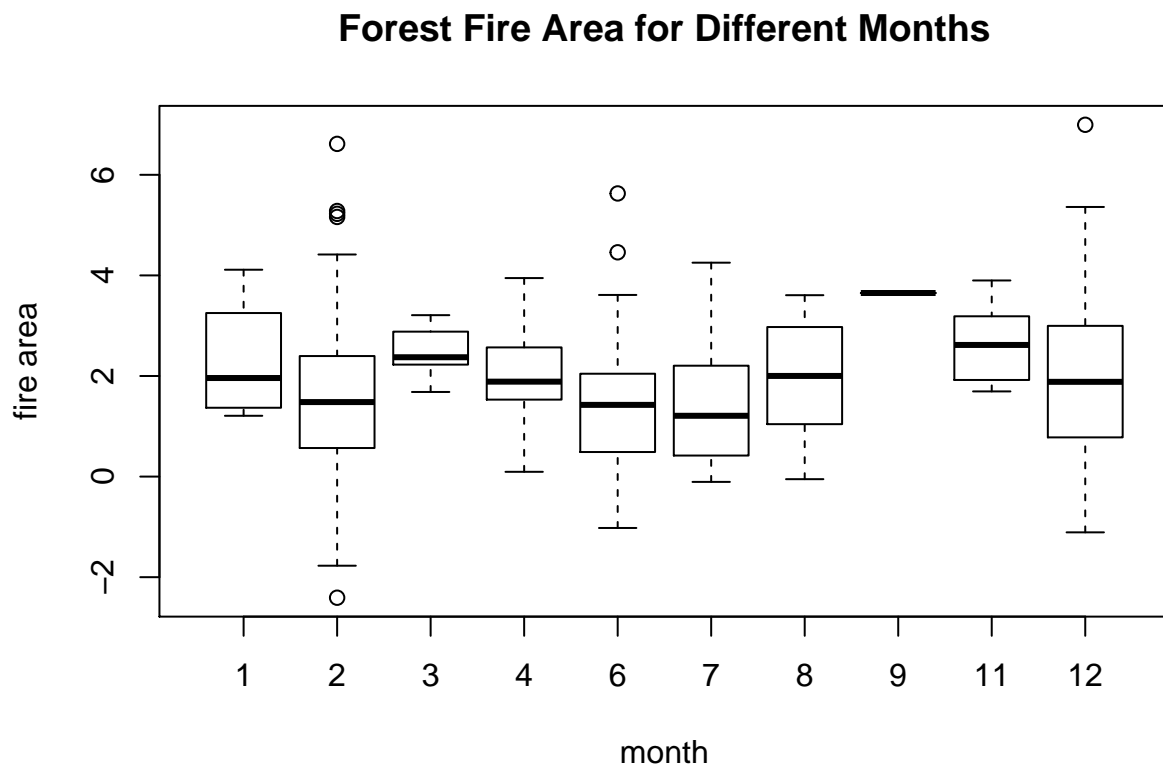
Boxplot of forest fire area for different X's



Boxplot of forest fire area for different Y's



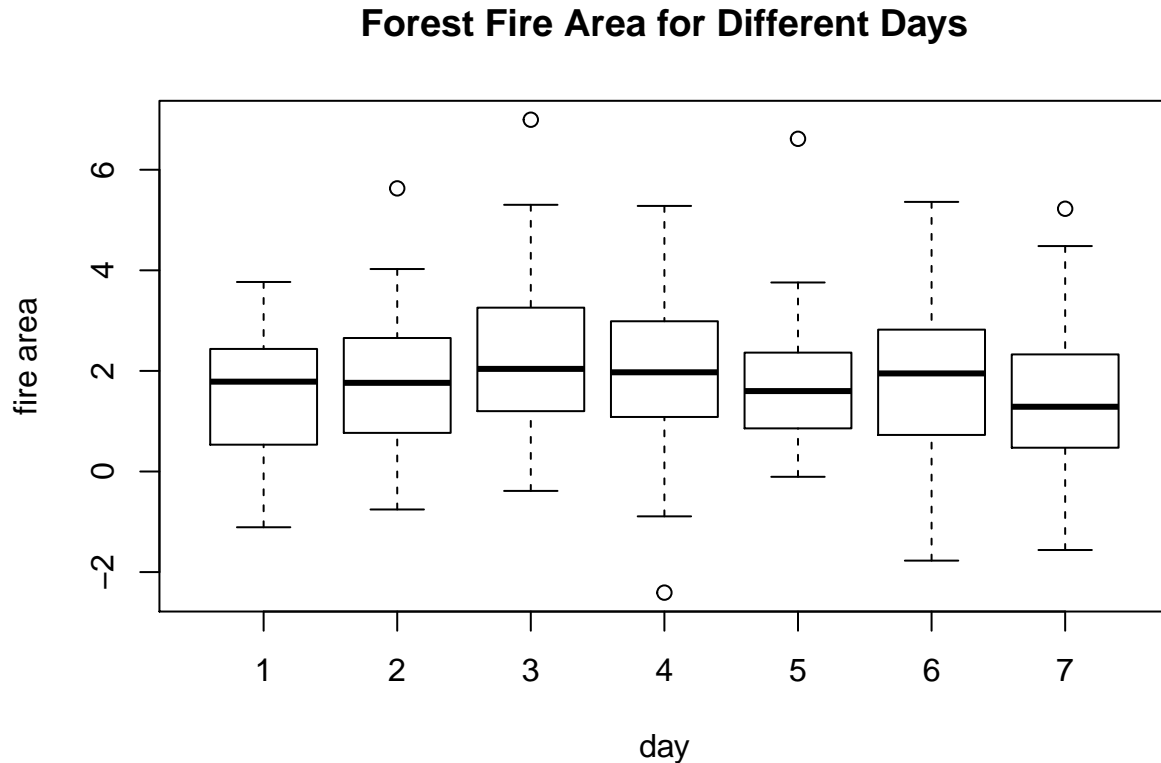
Boxplot of forest fire area for different Months



Here based on a total of 517 fire observations and distributions across different months, we can deduce that there are more fire observations in August and September.

Boxplot of forest fire area for different days

```
boxplot(log(area)~day, data=df1, xlab="day", ylab="fire area", main="Forest Fire Area for Different Days")
```



```
reg_day <-lm(log(area)~ day, data= df1)
summary(reg_day)
```

```
##
## Call:
## lm(formula = log(area) ~ day, data = df1)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.2495 -1.1092  0.0004  0.8867  5.1355
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   1.91216    0.20352   9.395  <2e-16 ***
## day          -0.01765    0.04745  -0.372    0.71
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.529 on 268 degrees of freedom
## Multiple R-squared:  0.0005161, Adjusted R-squared:  -0.003213
## F-statistic: 0.1384 on 1 and 268 DF, p-value: 0.7102
```

Scatterplots of all different predictor variables.

