linear Model Performance on Forest Fire Dataset

Alison Jing Huang

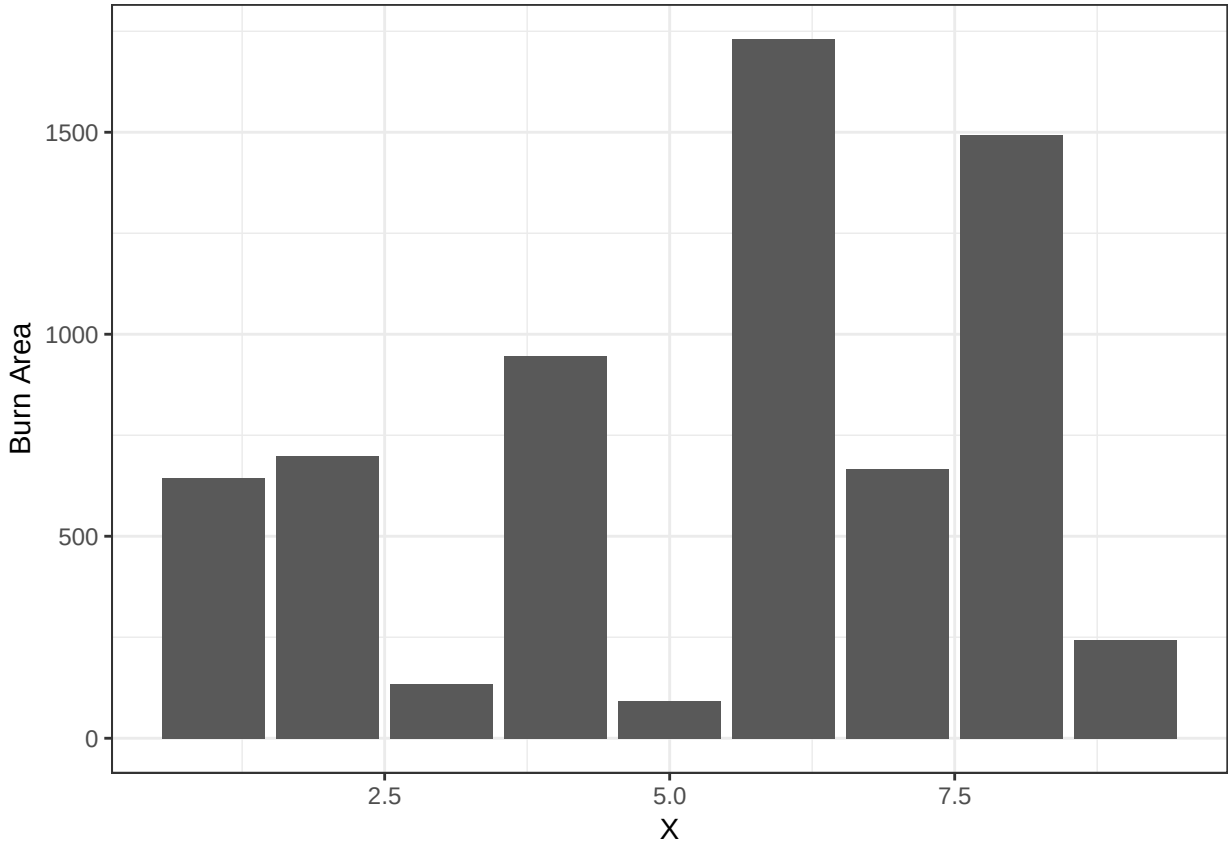
4/15/2018

Load the data

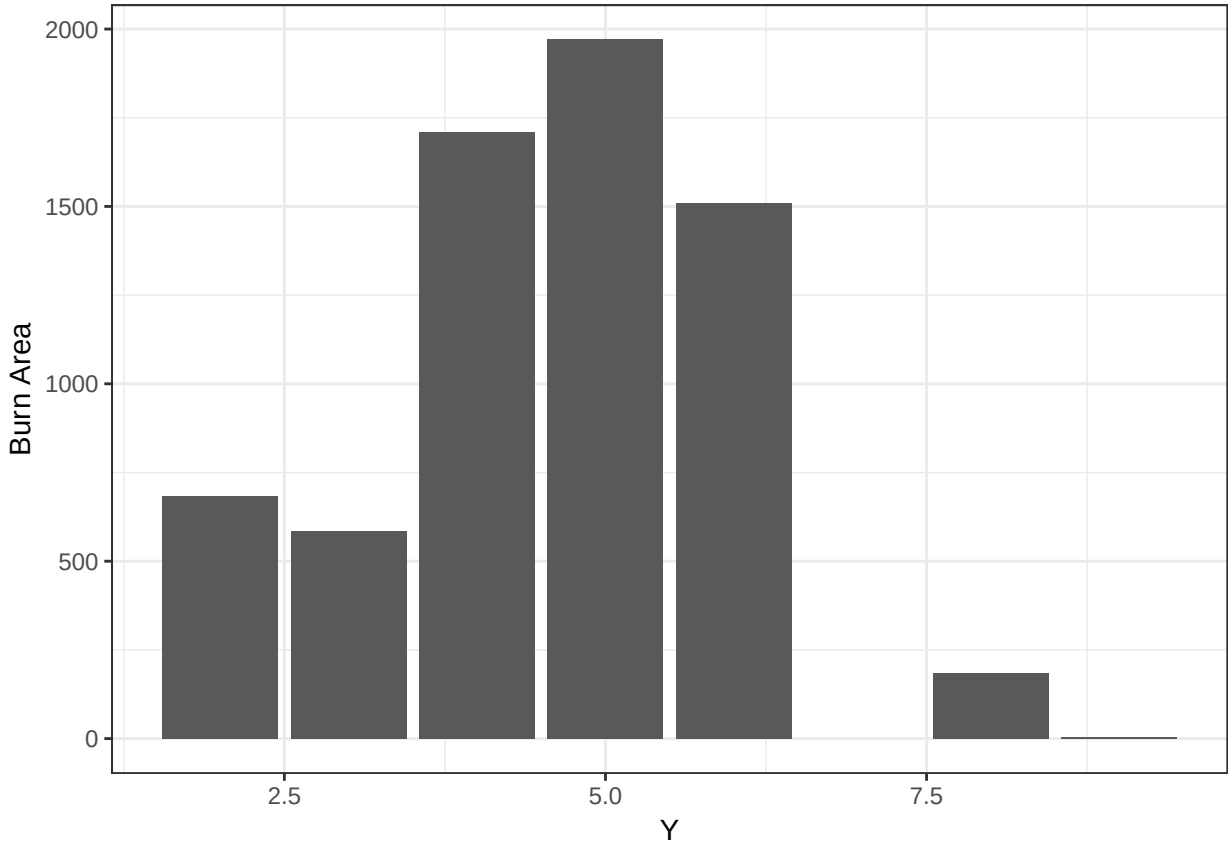
```
##   X Y month day FFMC  DMC    DC  ISI temp RH wind rain area
## 1 7 5     8   1 86.2 26.2  94.3  5.1  8.2 51  6.7  0.0    0
## 2 7 4    11   6 90.6 35.4 669.1  6.7 18.0 33  0.9  0.0    0
## 3 7 4    11   3 90.6 43.7 686.9  6.7 14.6 33  1.3  0.0    0
## 4 8 6     8   1 91.7 33.3  77.5  9.0  8.3 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
```

After we conducted initial analysis on the dataset, the next step is to create a linear model on the first fires data. Recall earlier we have transform the raw datset to all numerical variables, and renamed it as `fires`, next use ggplot2 to examine different variable with respect to the response variable - **"AREA"**.

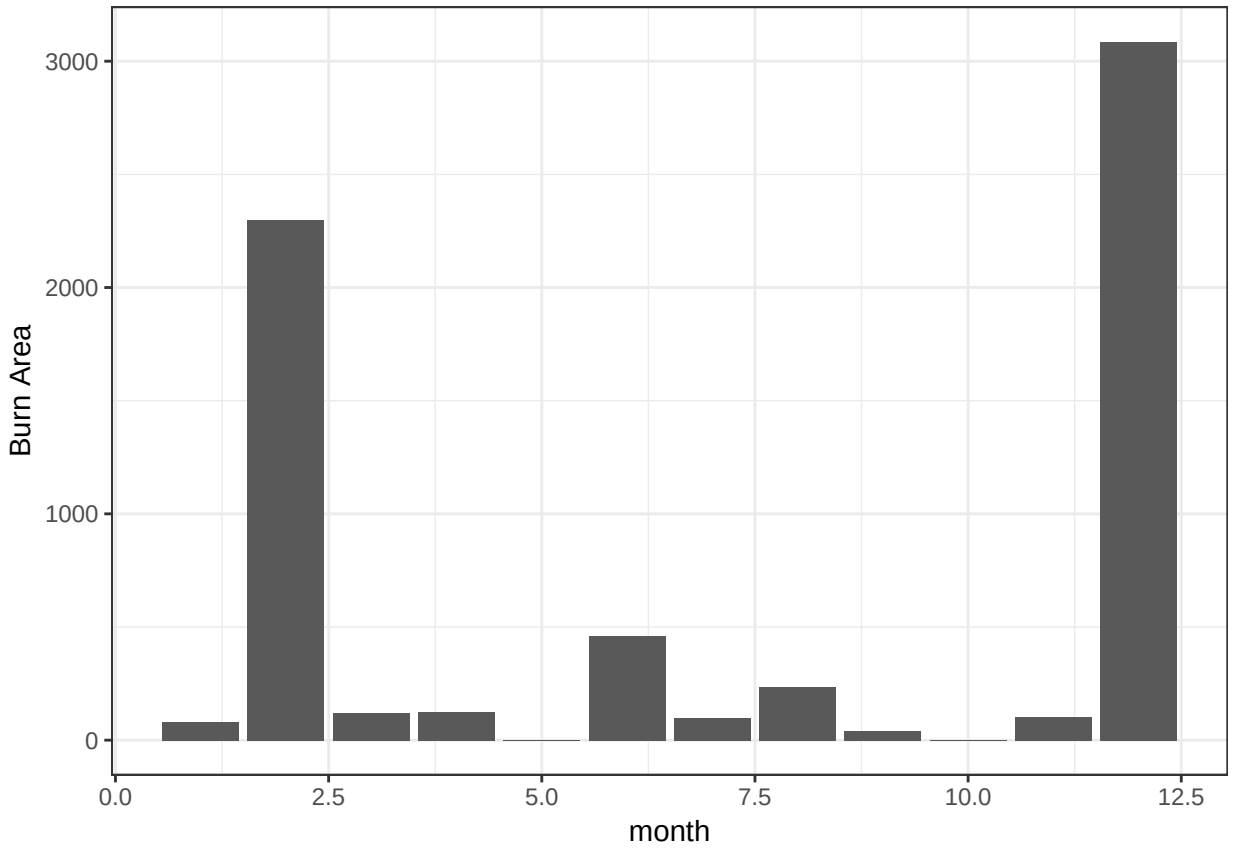
Relationship between X and AREA



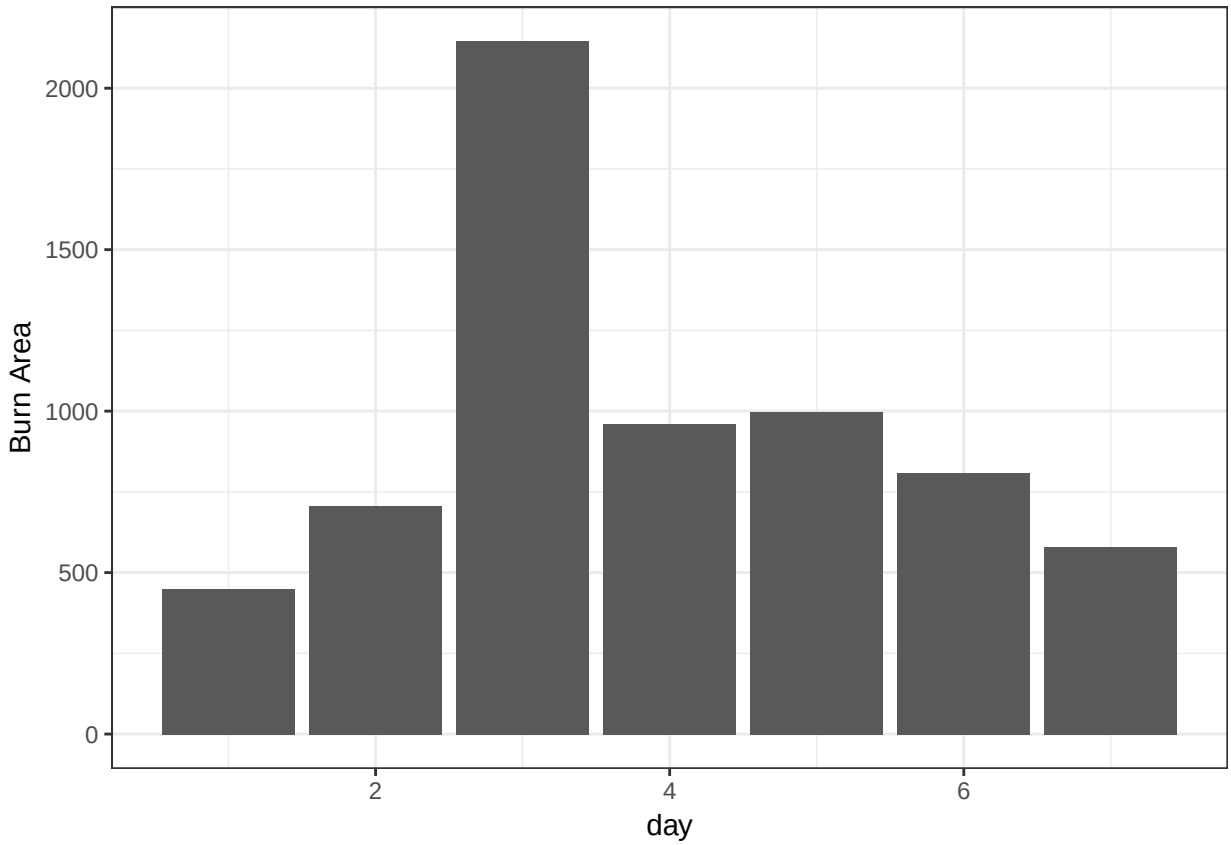
Relationship between Y and AREA



Relationship between MONTH and AREA



Relationship between DAY and AREA



Relationship between FFMC and AREA