Forest Fires

Sarvesh Kaushik, Jay Gupta, Colin Brennan

4/27/2022

Input Region CSV

```
forestfireregion1 <- read.csv("/Users/jaygupta/Desktop/ITMD 514/Final Project/ForestFire
Region1.csv", sep = ",")
```

Data Preparation Region

```
names(forestfireregion1)[names(forestfireregion1)=="ï..day"] <- "day"
dim(forestfireregion1)
```

```
## [1] 122  14
```

```
str(forestfireregion1)
```

```
## 'data.frame':    122 obs. of  14 variables:
##  $ day        : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ month      : int  6 6 6 6 6 6 6 6 6 6 ...
##  $ year       : int  2012 2012 2012 2012 2012 2012 2012 2012 2012 2012 ...
##  $ Temperature: int  29 29 26 25 27 31 33 30 25 28 ...
##  $ RH         : int  57 61 82 89 77 67 54 73 88 79 ...
##  $ Ws         : int  18 13 22 13 16 14 13 15 13 12 ...
##  $ Rain       : num  0 1.3 13.1 2.5 0 0 0 0 0.2 0 ...
##  $ FFMC       : num  65.7 64.4 47.1 28.6 64.8 82.6 88.2 86.6 52.9 73.2 ...
##  $ DMC        : num  3.4 4.1 2.5 1.3 3 5.8 9.9 12.1 7.9 9.5 ...
##  $ DC         : num  7.6 7.6 7.1 6.9 14.2 22.2 30.5 38.3 38.8 46.3 ...
##  $ ISI        : num  1.3 1 0.3 0 1.2 3.1 6.4 5.6 0.4 1.3 ...
##  $ BUI        : num  3.4 3.9 2.7 1.7 3.9 7 10.9 13.5 10.5 12.6 ...
##  $ FWI        : num  0.5 0.4 0.1 0 0.5 2.5 7.2 7.1 0.3 0.9 ...
##  $ Classes    : chr  "not fire" "not fire" "not fire" "not fire" ...
```

```
summary(forestfireregion1)
```

```
##       day            month          year       Temperature          RH
##  Min.   : 1.00   Min.   :6.0   Min.   :2012   Min.   :22.00   Min.   :45.00
##  1st Qu.: 8.00   1st Qu.:7.0   1st Qu.:2012   1st Qu.:29.00   1st Qu.:60.00
##  Median :16.00   Median :7.5   Median :2012   Median :31.00   Median :68.00
##  Mean   :15.75   Mean   :7.5   Mean   :2012   Mean   :31.18   Mean   :67.98
##  3rd Qu.:23.00   3rd Qu.:8.0   3rd Qu.:2012   3rd Qu.:34.00   3rd Qu.:77.75
##  Max.   :31.00   Max.   :9.0   Max.   :2012   Max.   :37.00   Max.   :89.00
##        Ws           Rain              FFMC            DMC
##  Min.   :11   Min.   : 0.0000   Min.   :28.60   Min.   : 0.700
##  1st Qu.:14   1st Qu.: 0.0000   1st Qu.:65.92   1st Qu.: 3.725
##  Median :16   Median : 0.0000   Median :80.90   Median : 9.450
##  Mean   :16   Mean   : 0.8426   Mean   :74.67   Mean   :12.315
##  3rd Qu.:18   3rd Qu.: 0.5000   3rd Qu.:86.78   3rd Qu.:16.300
##  Max.   :26   Max.   :16.8000   Max.   :90.30   Max.   :54.200
##        DC              ISI              BUI             FWI
##  Min.   :  6.90   Min.   : 0.000   Min.   : 1.10   Min.   : 0.000
##  1st Qu.: 10.05   1st Qu.: 1.125   1st Qu.: 5.10   1st Qu.: 0.500
##  Median : 35.55   Median : 2.650   Median :11.20   Median : 3.000
##  Mean   : 53.16   Mean   : 3.656   Mean   :15.43   Mean   : 5.578
##  3rd Qu.: 79.03   3rd Qu.: 5.600   3rd Qu.:21.68   3rd Qu.: 8.700
##  Max.   :220.40   Max.   :12.500   Max.   :67.40   Max.   :30.200
##    Classes
##  Length:122
##  Class :character
##  Mode  :character
##
##
##
```

```
head(forestfireregion1)
```

```
##   day month year Temperature RH Ws Rain FFMC DMC   DC ISI BUI FWI  Classes
## 1   1     6 2012          29 57 18  0.0 65.7 3.4  7.6 1.3 3.4 0.5 not fire
## 2   2     6 2012          29 61 13  1.3 64.4 4.1  7.6 1.0 3.9 0.4 not fire
## 3   3     6 2012          26 82 22 13.1 47.1 2.5  7.1 0.3 2.7 0.1 not fire
## 4   4     6 2012          25 89 13  2.5 28.6 1.3  6.9 0.0 1.7 0.0 not fire
## 5   5     6 2012          27 77 16  0.0 64.8 3.0 14.2 1.2 3.9 0.5 not fire
## 6   6     6 2012          31 67 14  0.0 82.6 5.8 22.2 3.1 7.0 2.5     fire
```

```
sum(is.na(forestfireregion1))
```

```
## [1] 0
```

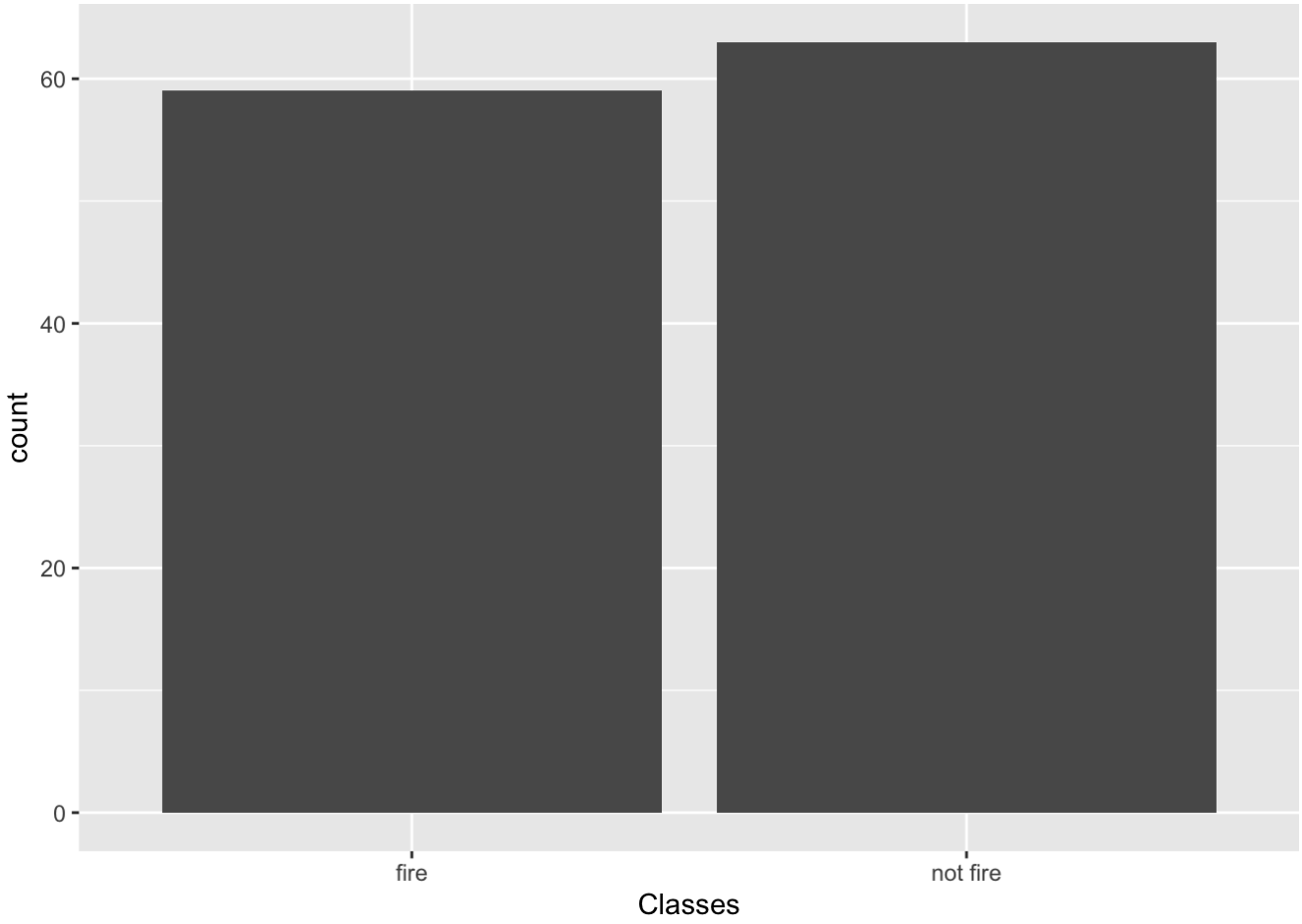
```
sum(duplicated(forestfireregion1))
```

```
## [1] 0
```

Region 1 is a data frame that contains 122 observations and 14 features. There are no duplicate or missing values.

Bar Plots Region

```
# Region 1 Classes
ggplot(data = forestfireregion1) + geom_bar(mapping = aes(x = Classes))
```



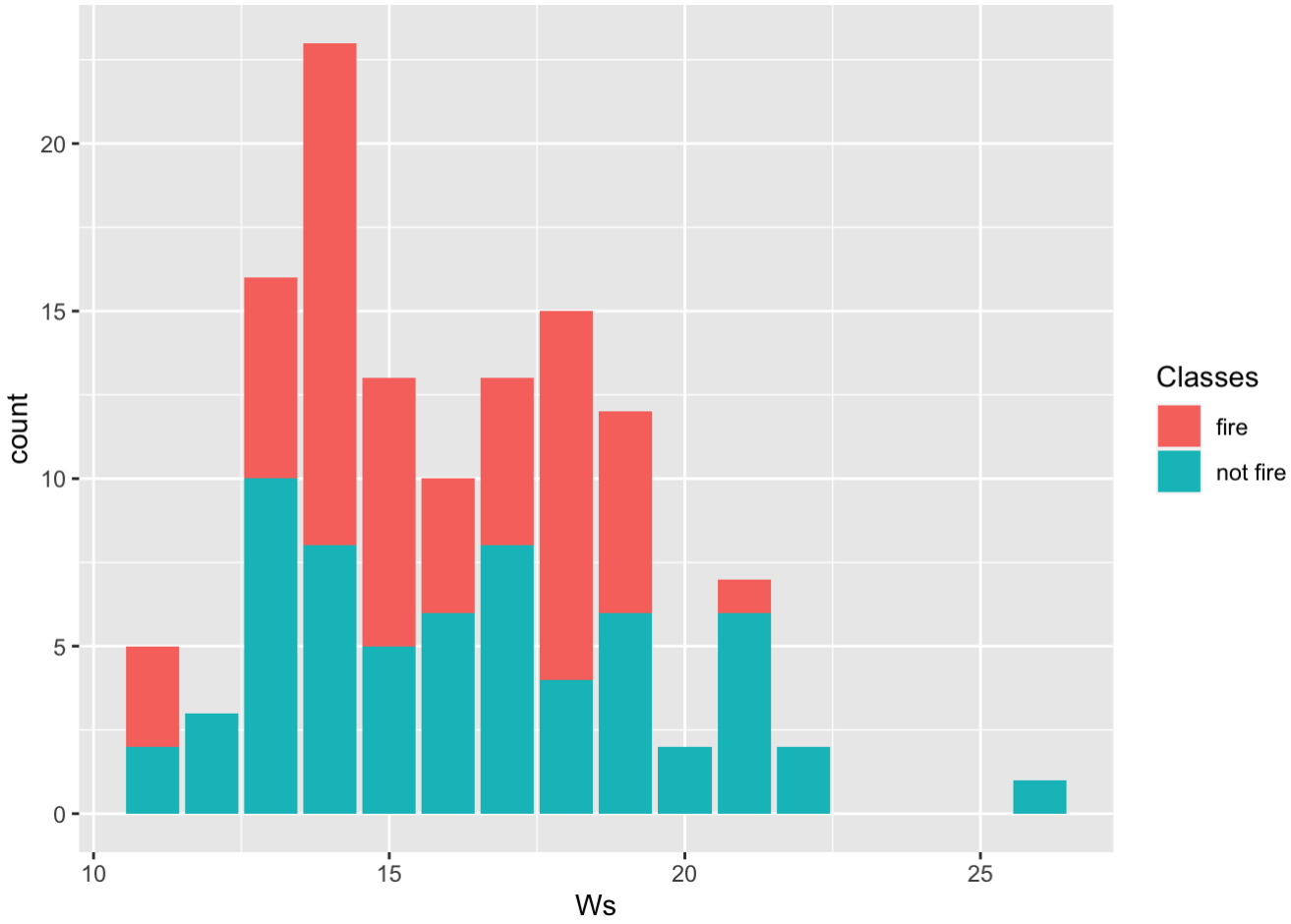

```
# Frequency of rain and temperature to understand them better
as.data.frame(table(forestfireregion1$Temperature))
```

```
##    Var1 Freq
## 1    22    2
## 2    24    1
## 3    25    6
## 4    26    4
## 5    27    4
## 6    28    7
## 7    29   11
## 8    30   10
## 9    31   18
## 10   32   14
## 11   33   13
## 12   34    7
## 13   35   15
## 14   36    9
## 15   37    1
```

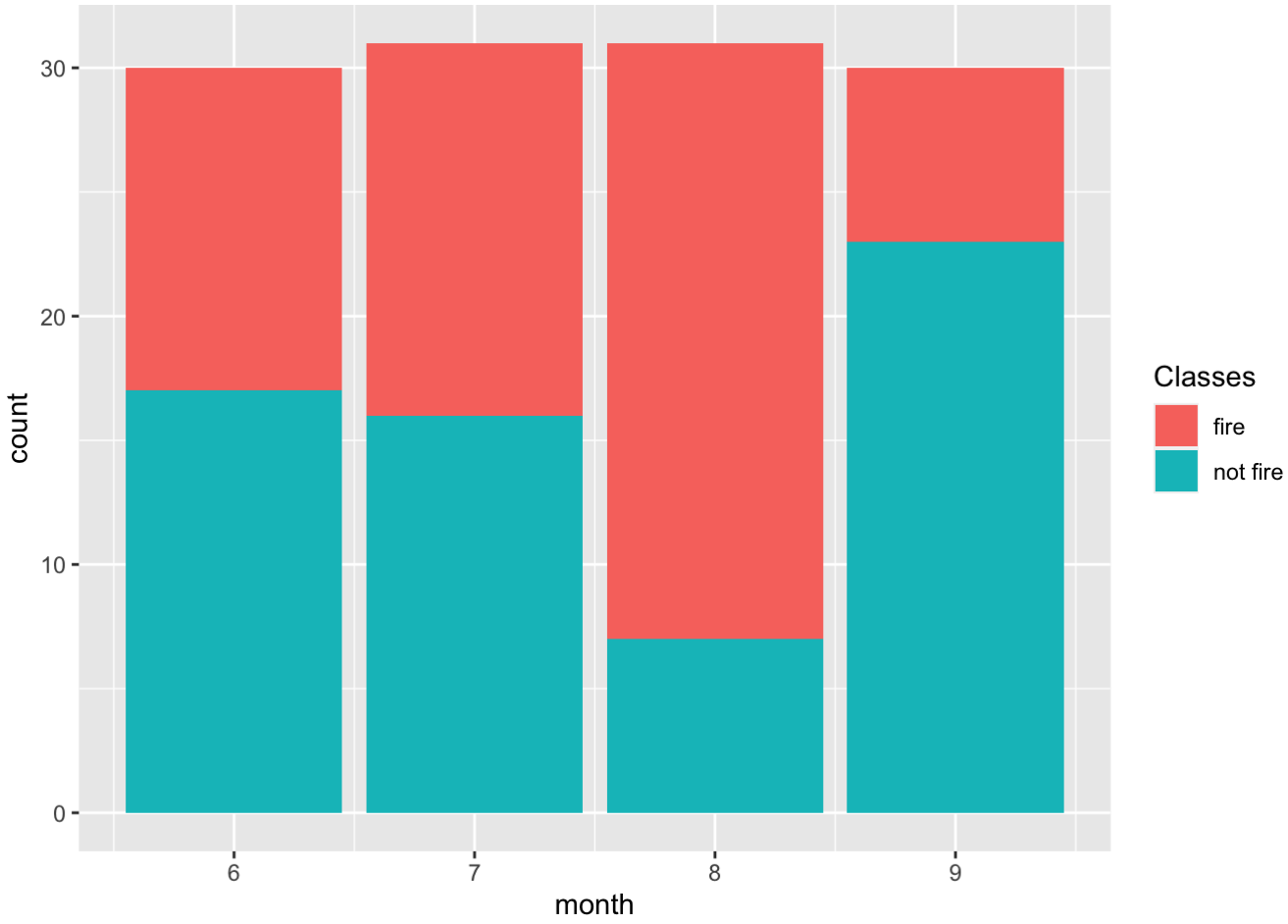
```
as.data.frame(table(forestfireregion1$Rain))
```

```
##    Var1 Freq
## 1     0   66
## 2   0.1    9
## 3   0.2    2
## 4   0.3    8
## 5   0.4    5
## 6   0.5    2
## 7   0.6    4
## 8   0.7    4
## 9   0.8    1
## 10  0.9    1
## 11    1    2
## 12  1.2    2
## 13  1.3    1
## 14  1.4    2
## 15  1.8    3
## 16  2.5    1
## 17  3.1    1
## 18  3.8    1
## 19  4.6    1
## 20  5.8    1
## 21  7.2    1
## 22  8.3    1
## 23 10.1    1
## 24 13.1    1
## 25 16.8    1
```

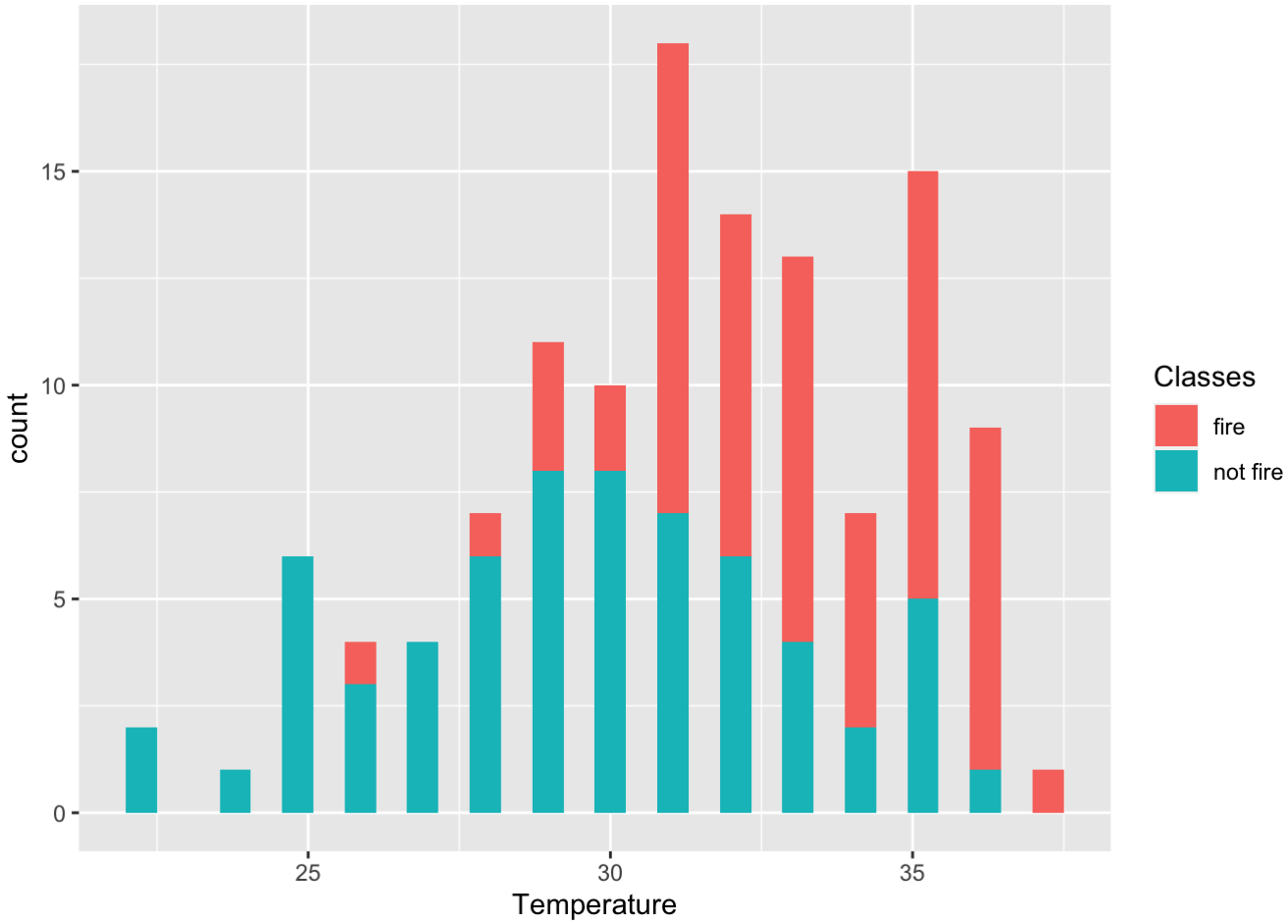
```
# Temperature has most values concentrated in categories 31, 32, 33 and 35. Only a few o
bservations made for categories 22, 24 and 37.

# Rain has most values concentrated in categories 0, 0.1 and 0.3. There are quite a lot
 of observations with less frequencies.

# Classes compared with Wind Speed
ggplot(data = forestfireregion1, aes(x = Ws)) +
  geom_bar(aes(fill = Classes))
```



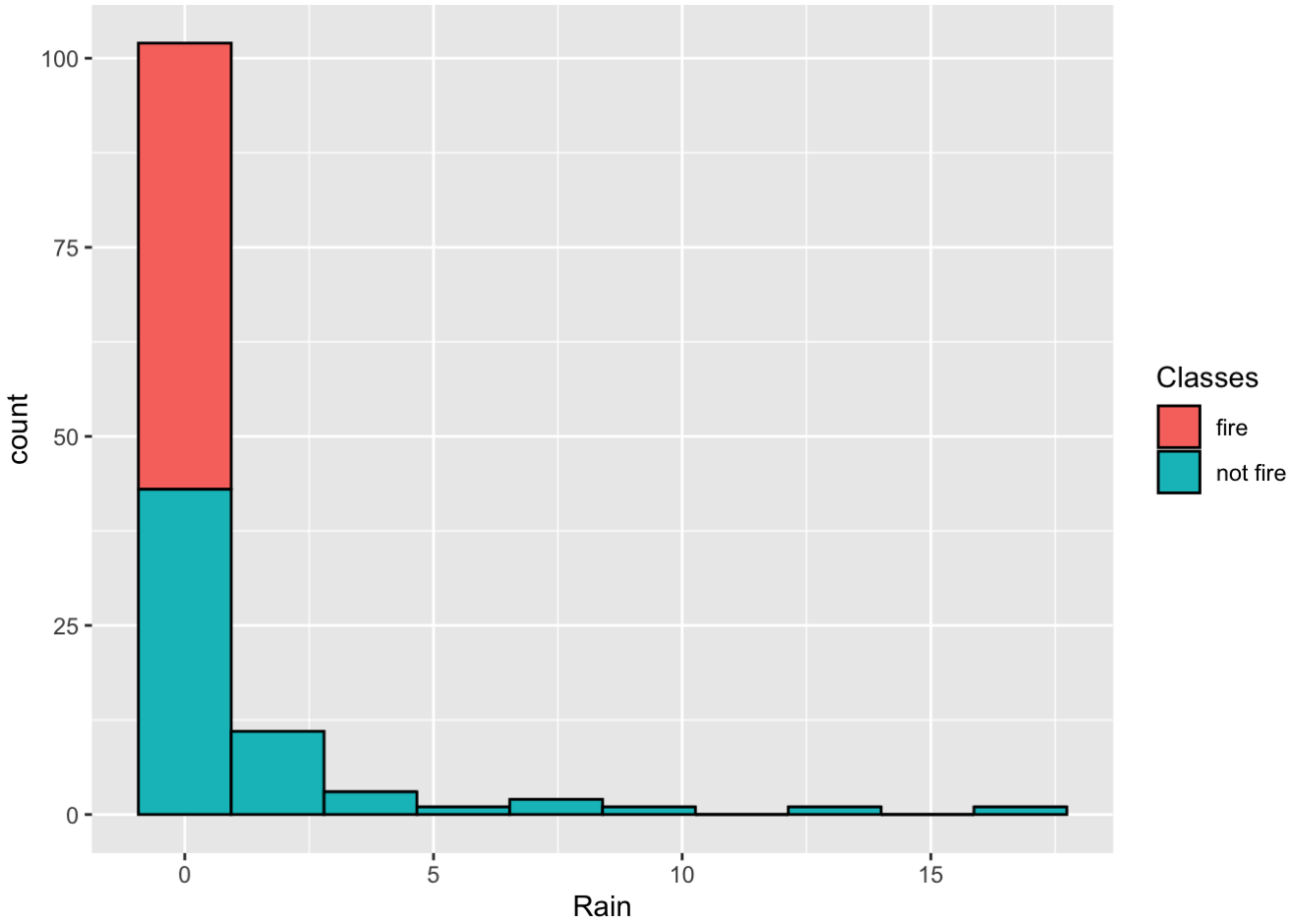

```
# Classes per Month
ggplot(data = forestfireregion1, aes(x = month)) +
  geom_bar(aes(fill = Classes))
```




```
# Temperature vs Classes
ggplot(data = forestfireregion1, aes(x = Temperature)) +geom_histogram(aes(fill = Classe
s),bins = 30)
```



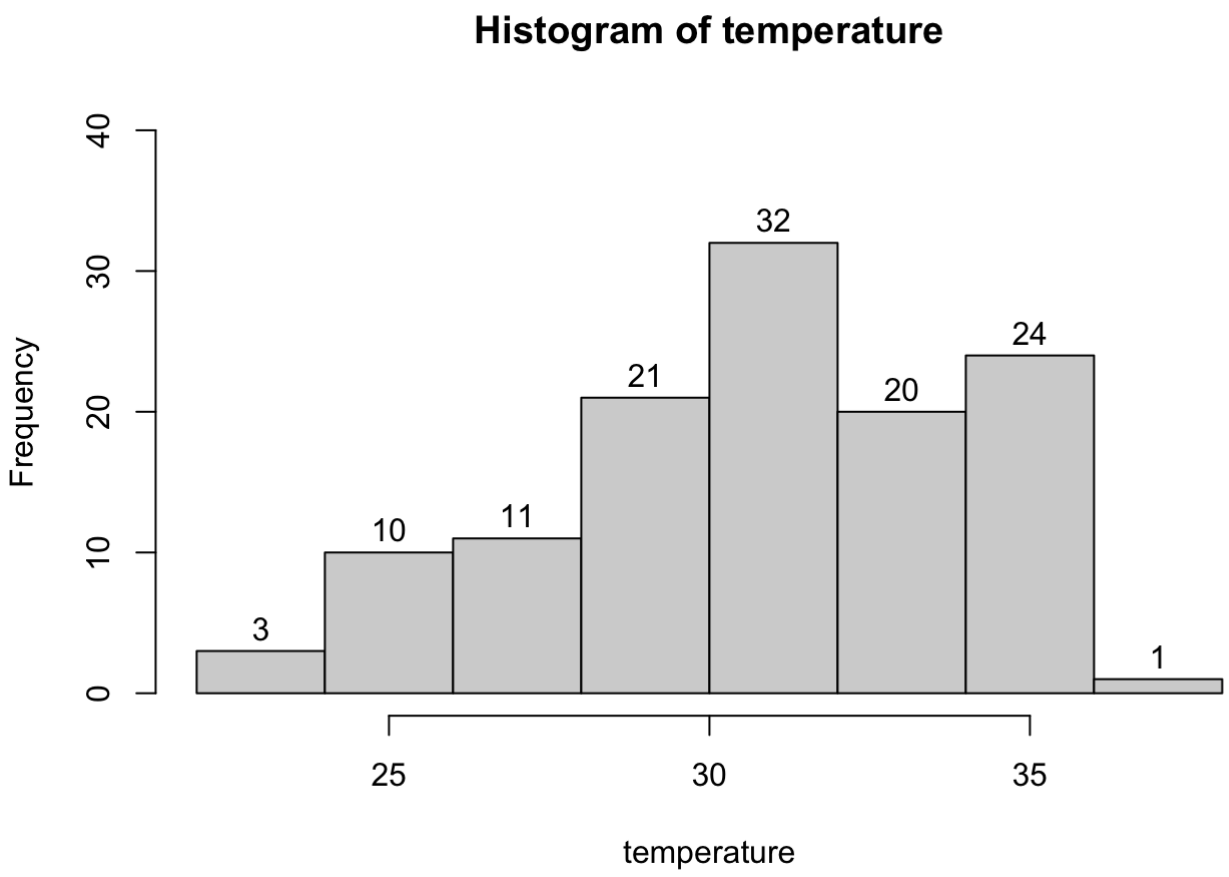

```
# Rain vs Fire
ggplot(data = forestfireregion1, aes(x = Rain)) + geom_histogram(aes(fill = Classes), bi
ns =10, color = "black")
```



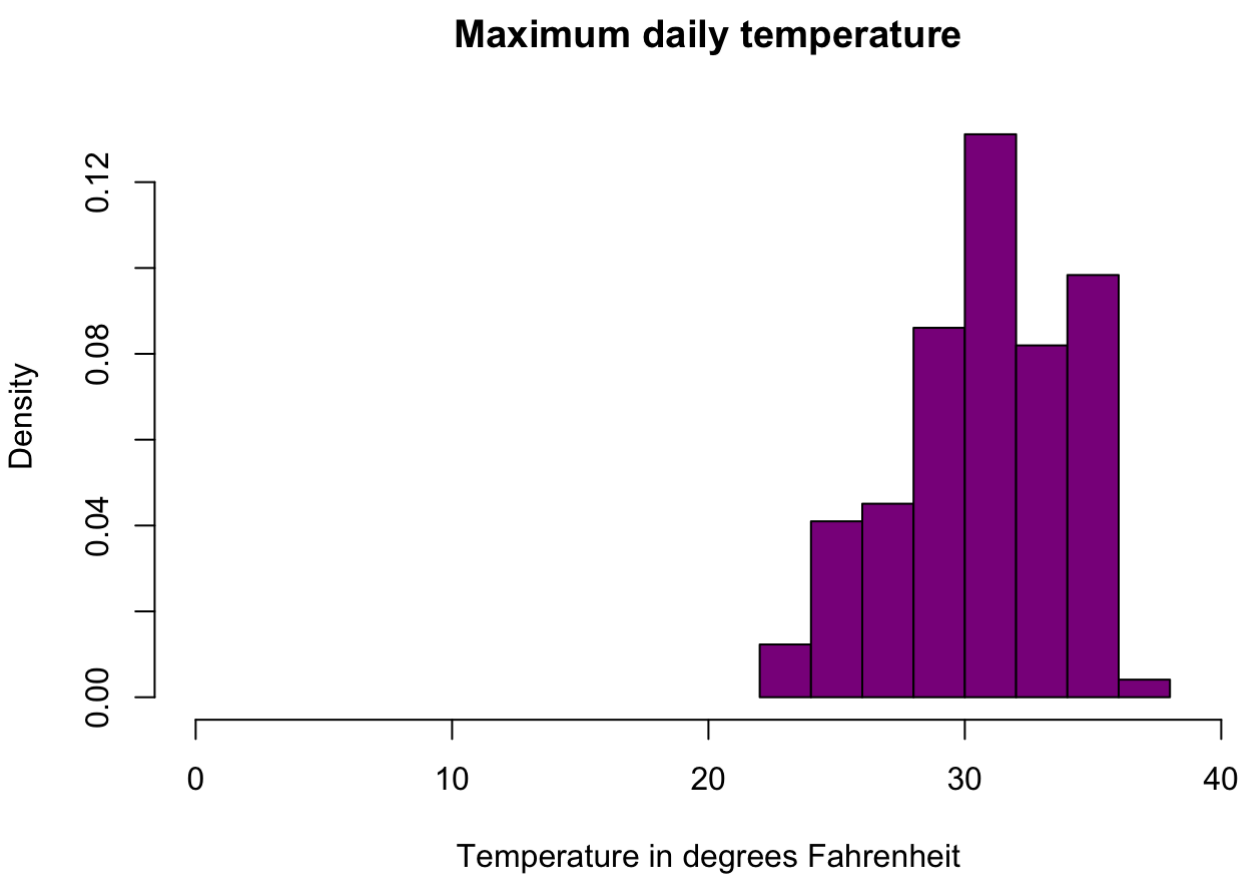
The Chances of fire are greater when less Rain Precipitation

EDA For Temperature

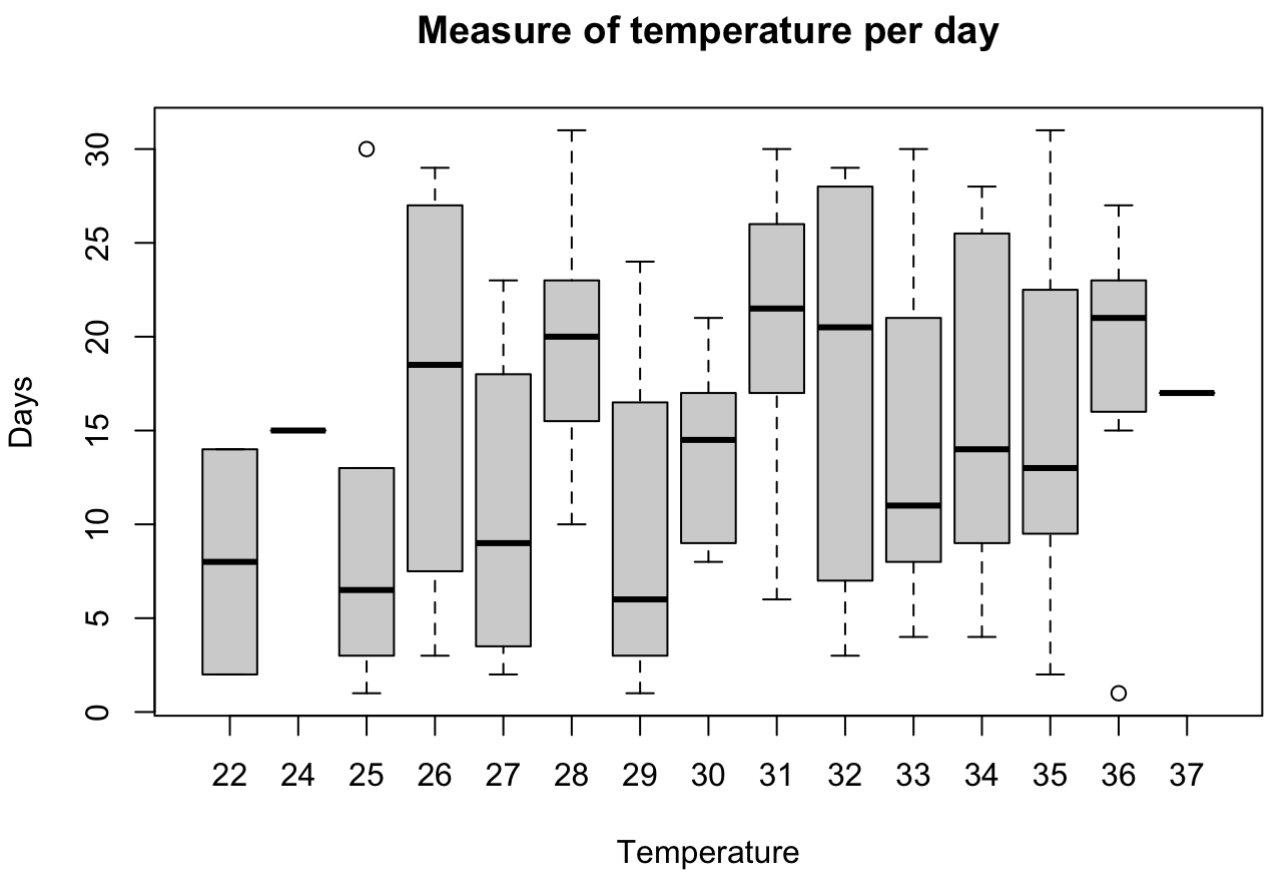
```
temperature=forestfireregion1$Temperature
h1=hist(temperature,ylim=c(0,40))
text(h1$mids,h1$counts,labels=h1$counts, adj=c(0.5, -0.5))
```



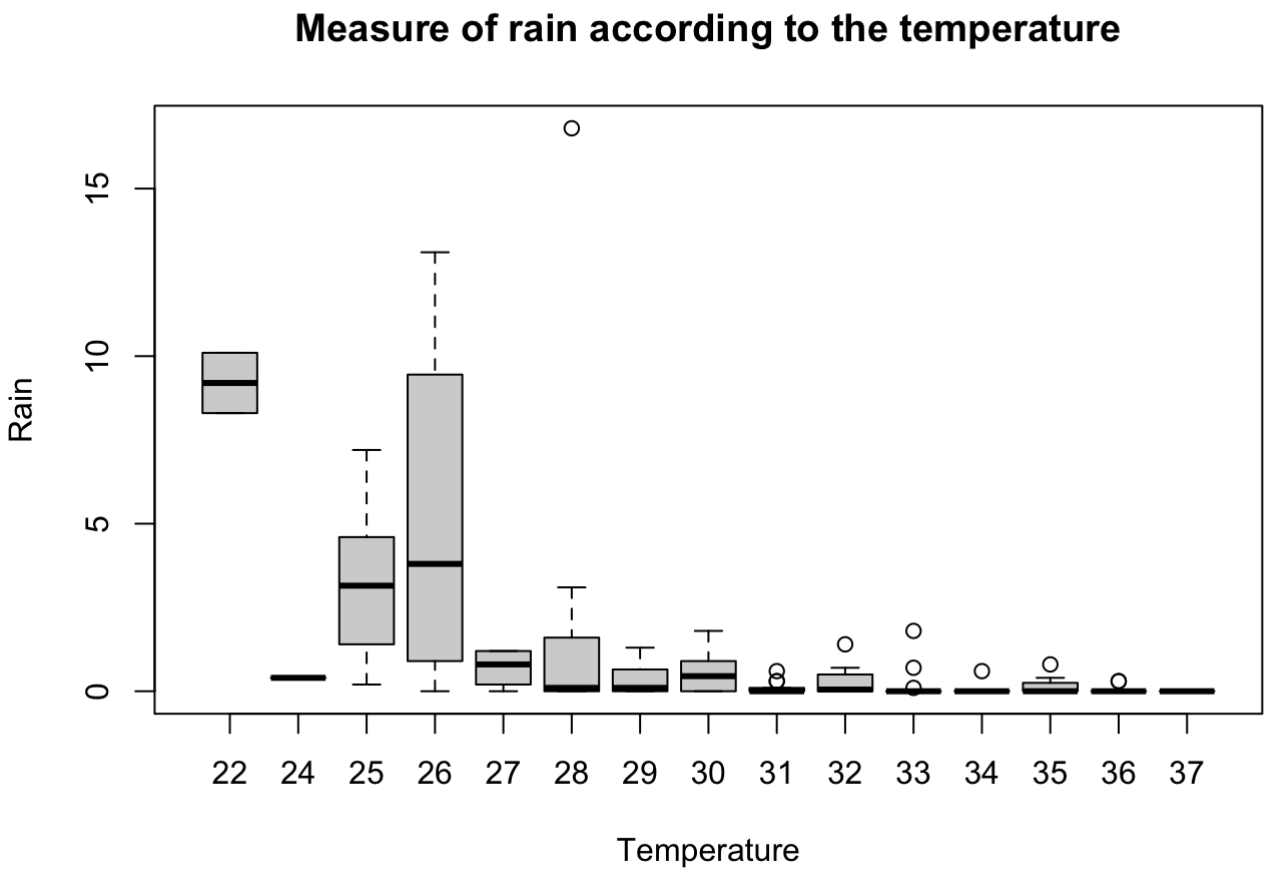

```
hist(temperature,main="Maximum daily temperature",xlab="Temperature in degrees Fahrenhei
t",xlim=c(0,40),col="darkmagenta",freq=FALSE)
```



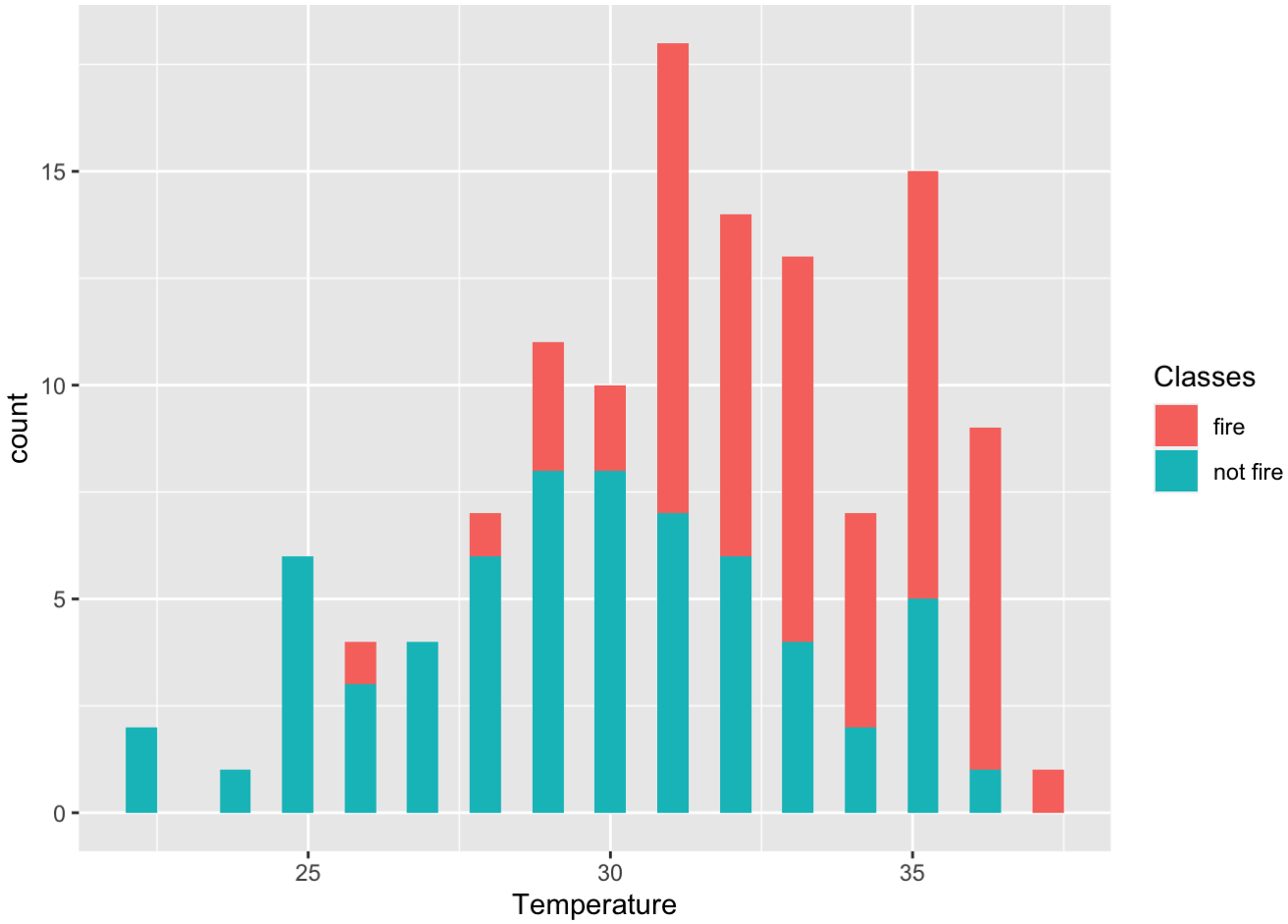

```
boxplot(day~Temperature,data=forestfireregion1,main="Measure of temperature per day",xla
b="Temperature",ylab="Days")
```




```
boxplot(Rain~Temperature,data=forestfireregion1,main="Measure of rain according to the t
emperature",xlab="Temperature",ylab="Rain")
```

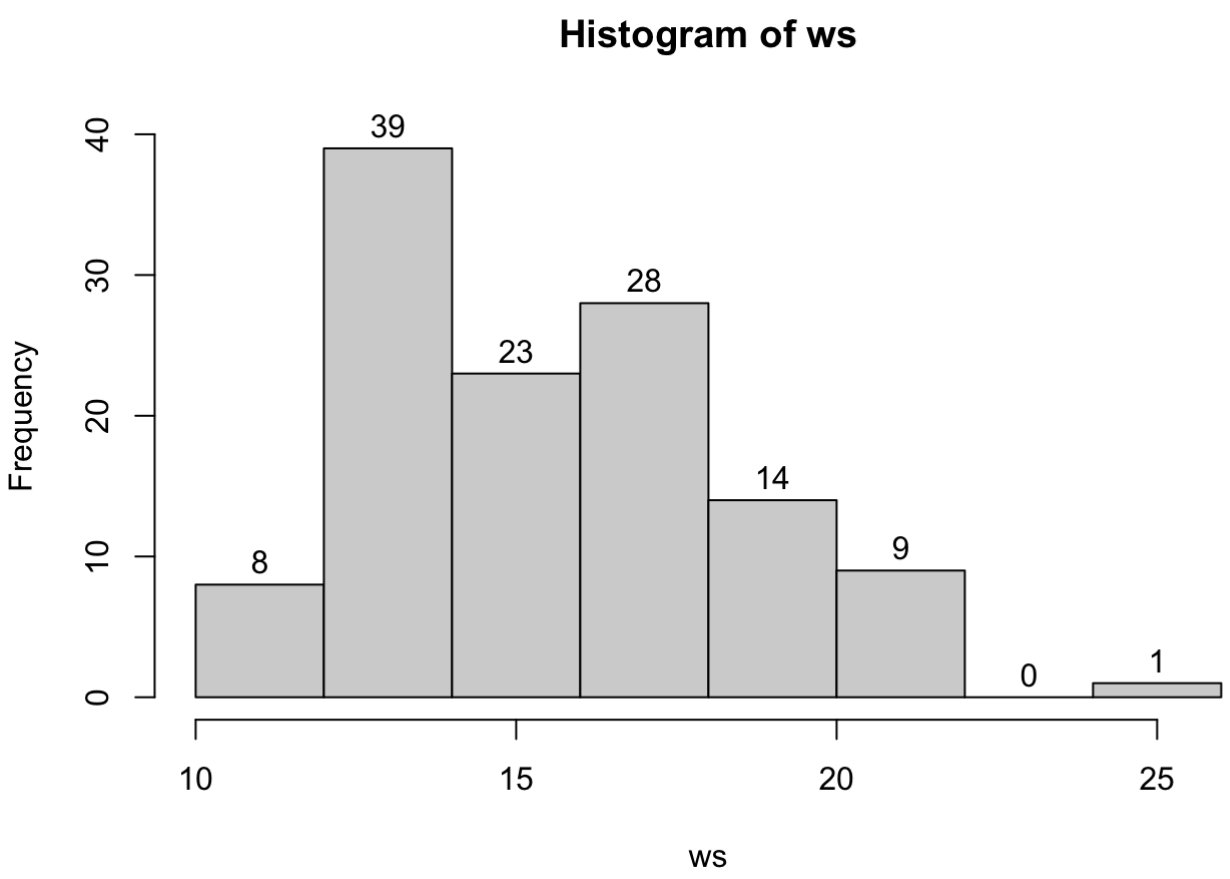



```
ggplot(data = forestfireregion1, aes(x = Temperature)) +geom_histogram(aes(fill = Classe
s),bins = 30)
```

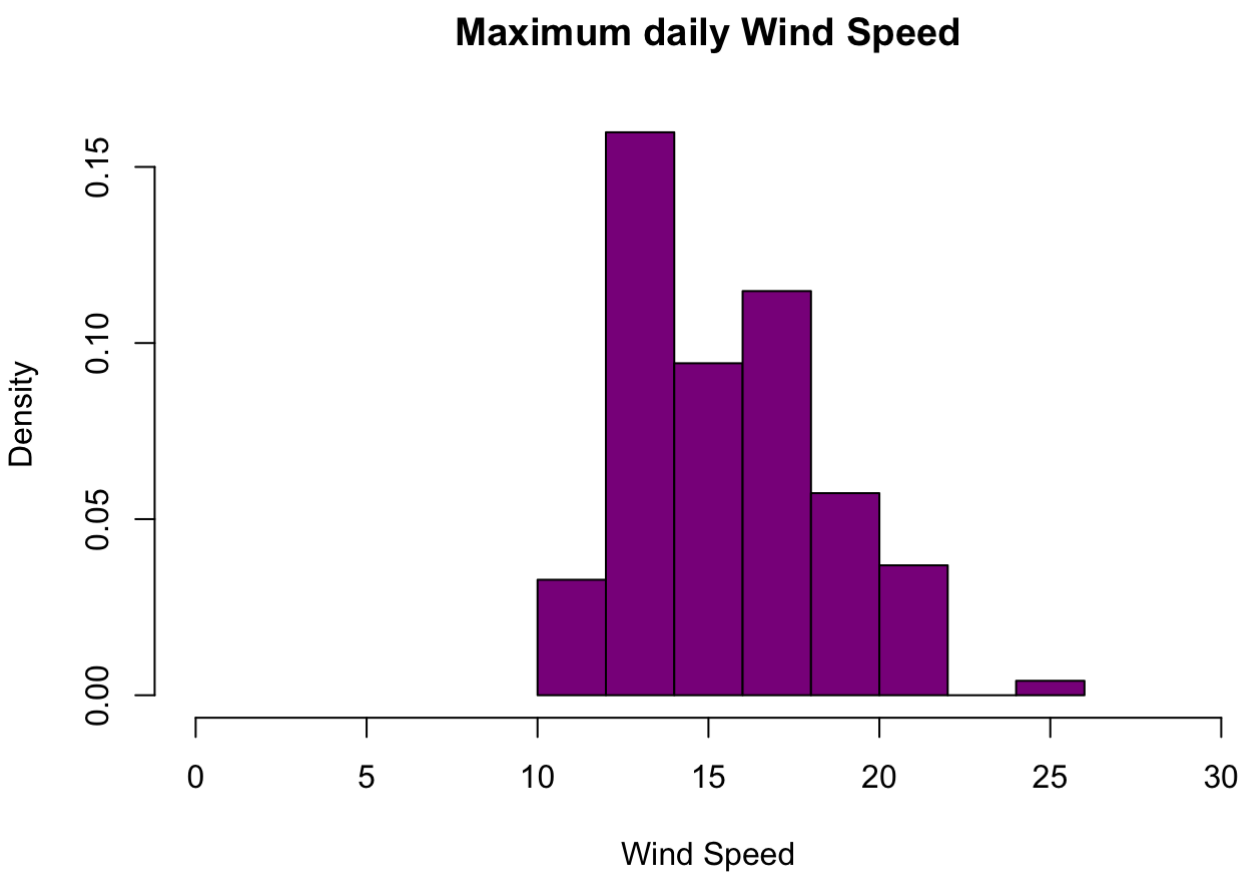


EDA For Wind Speed

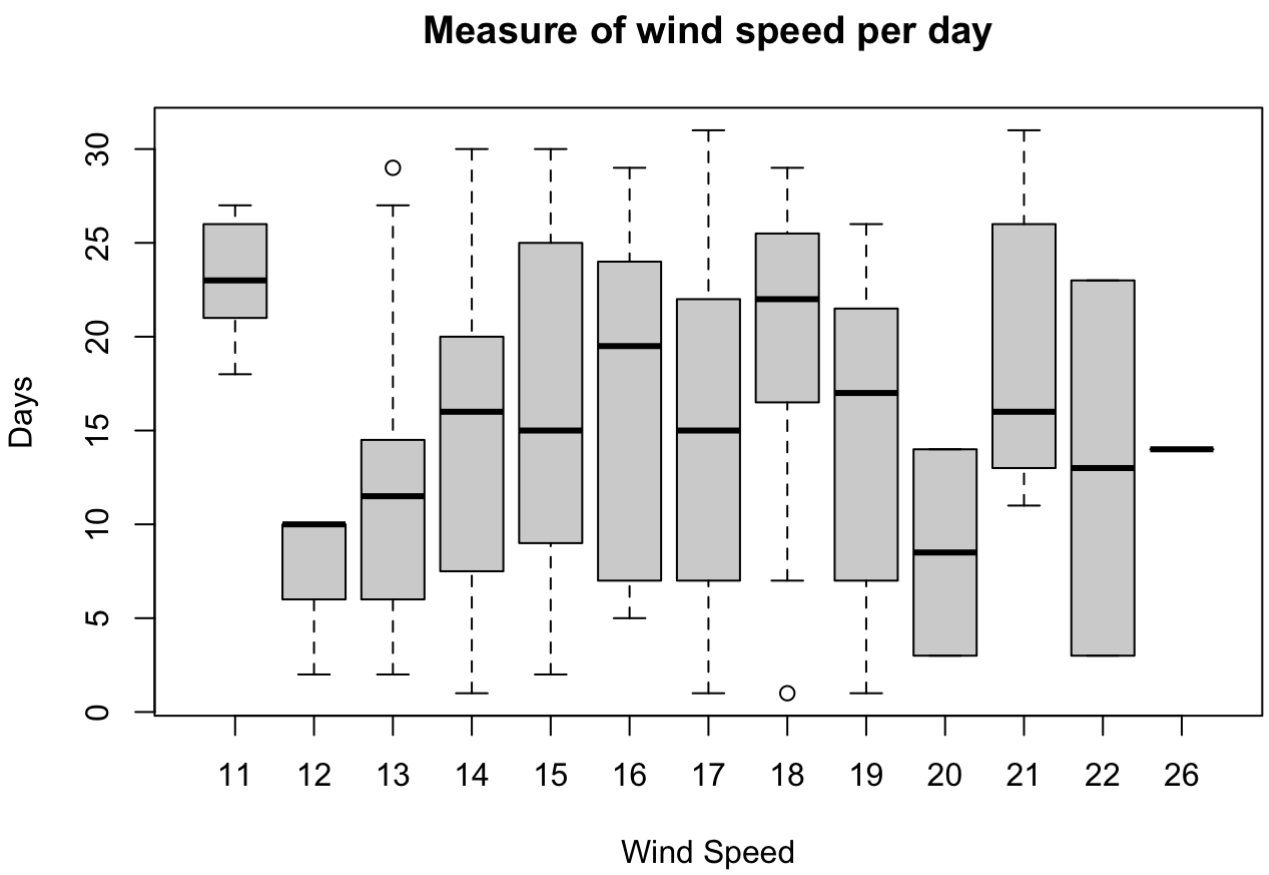
```
ws=forestfireregion1$Ws
h1=hist(ws,ylim=c(0,40))
text(h1$mids,h1$counts,labels=h1$counts, adj=c(0.5, -0.5))
```



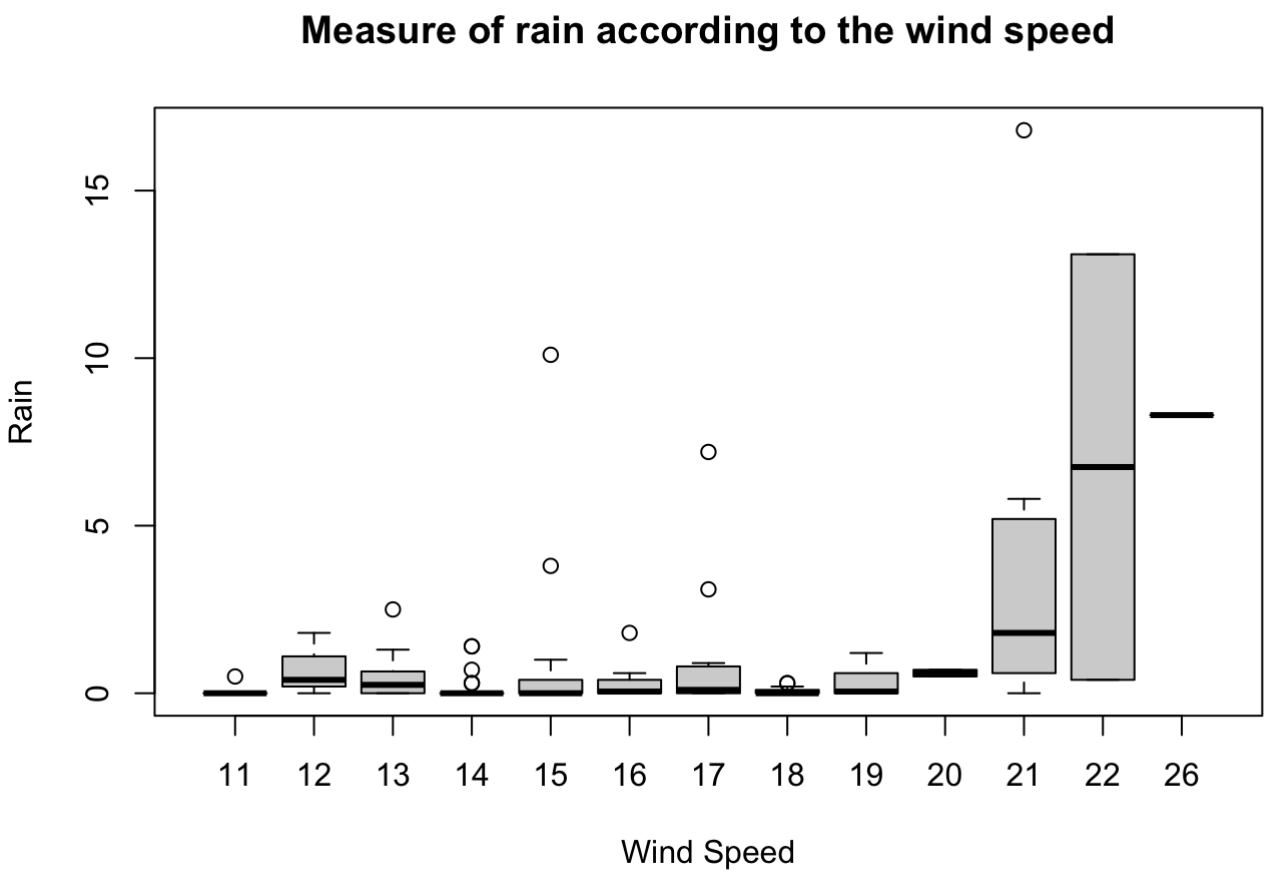

```
hist(ws,main="Maximum daily Wind Speed",xlab="Wind Speed",xlim=c(0,30),col="darkmagenta"
,freq=FALSE)
```



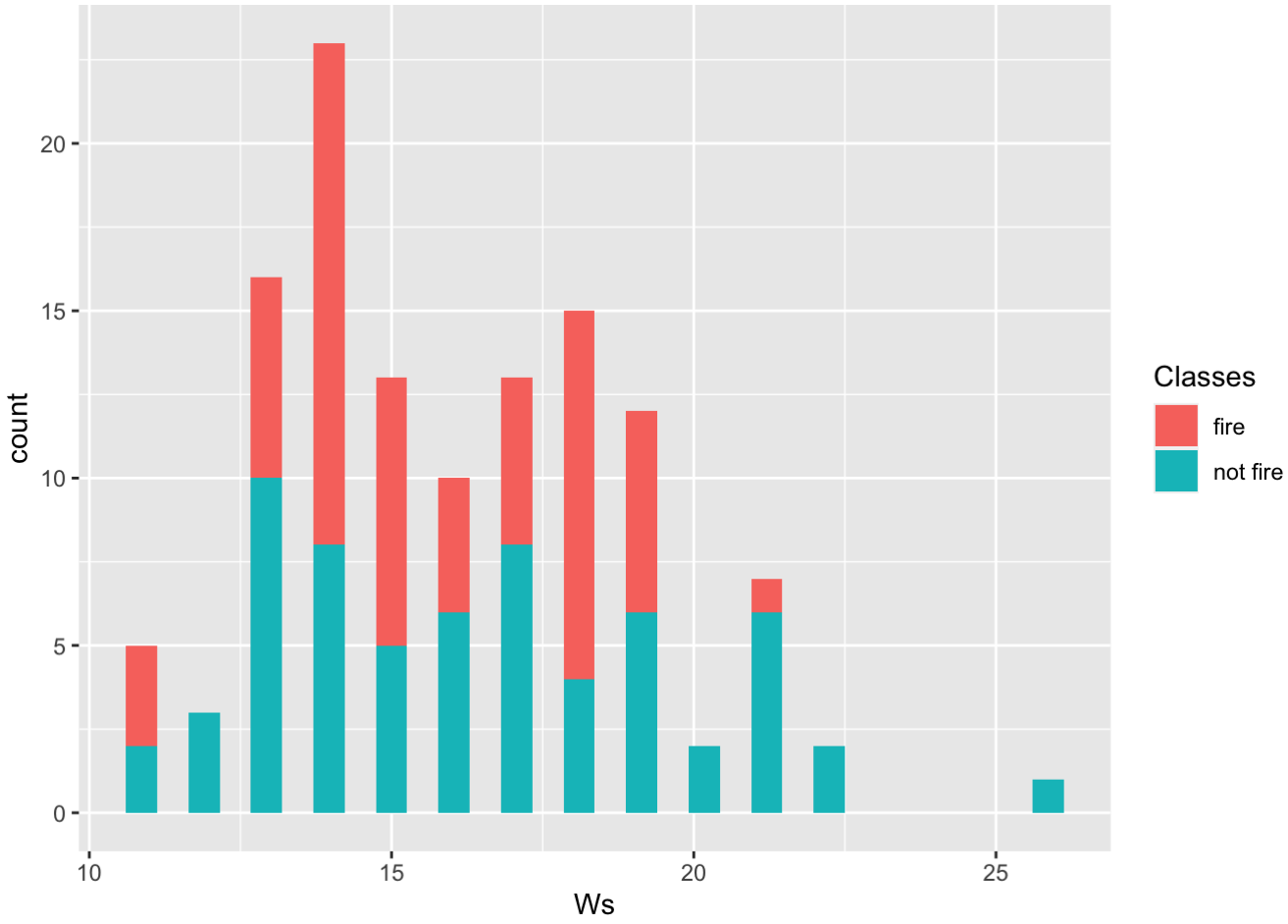

```
boxplot(day~Ws,data=forestfireregion1,main="Measure of wind speed per day",xlab="Wind Sp
eed",ylab="Days")
```




```
boxplot(Rain~Ws,data=forestfireregion1,main="Measure of rain according to the wind spee
d",xlab="Wind Speed",ylab="Rain")
```



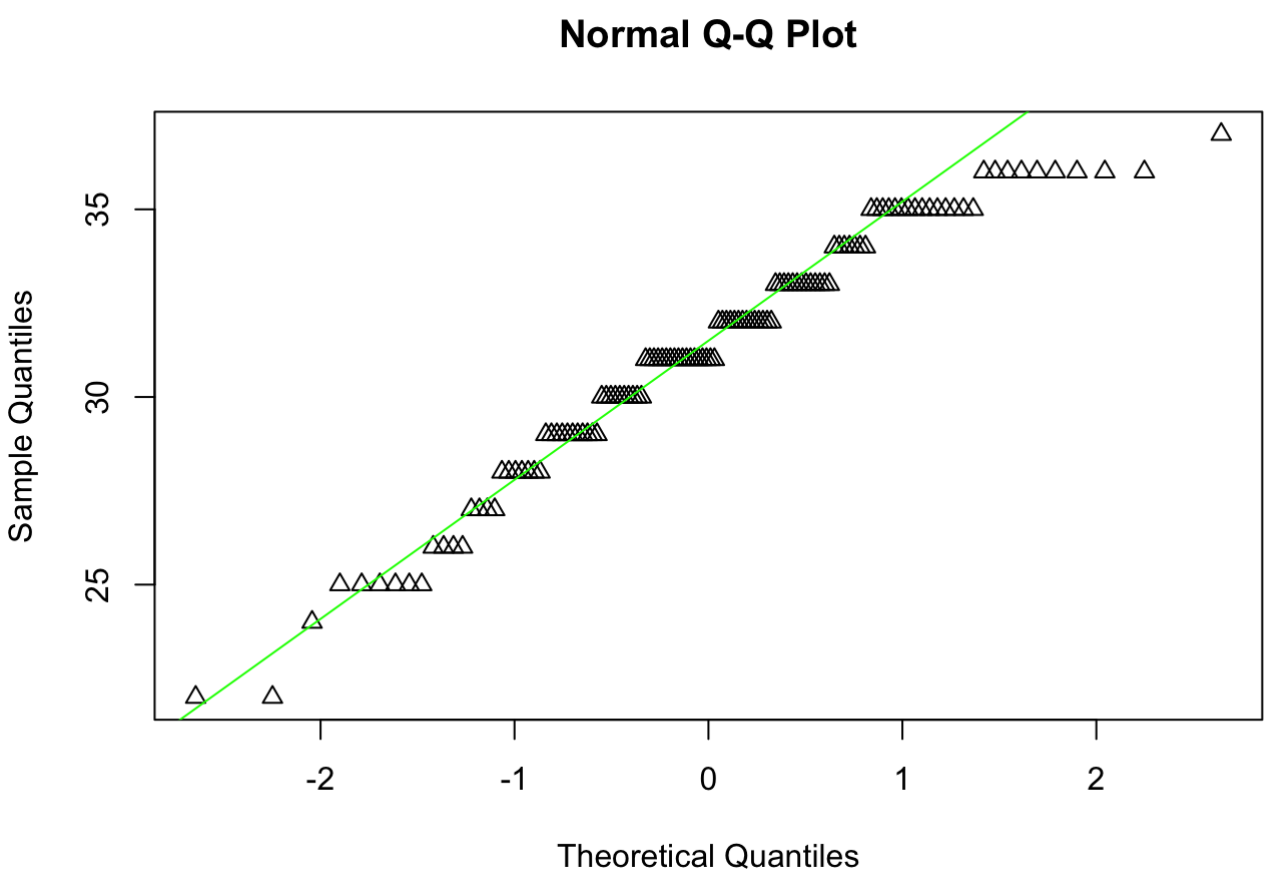

```
ggplot(data = forestfireregion1, aes(x = Ws)) +geom_histogram(aes(fill = Classes),bins =
30)
```



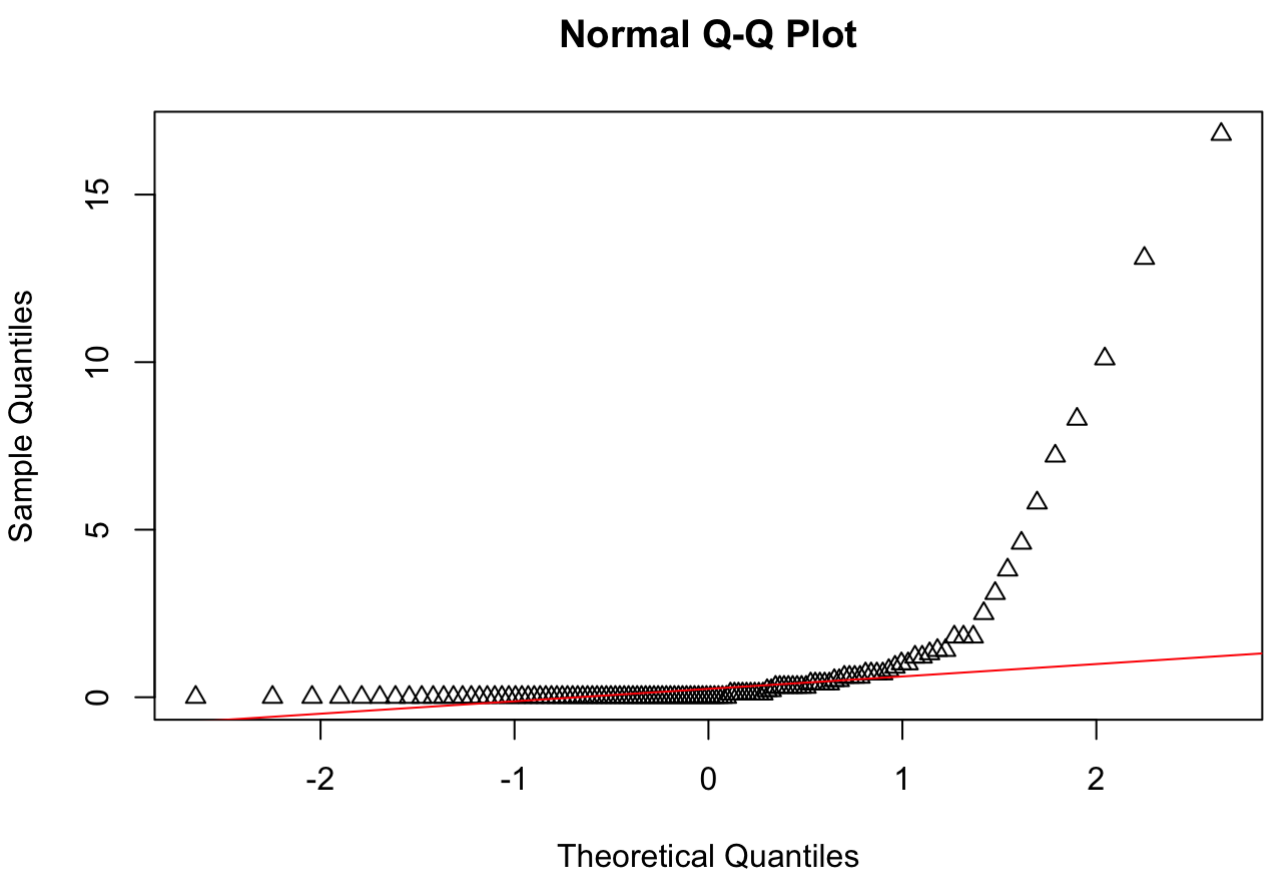
Data Analysis

QQplot

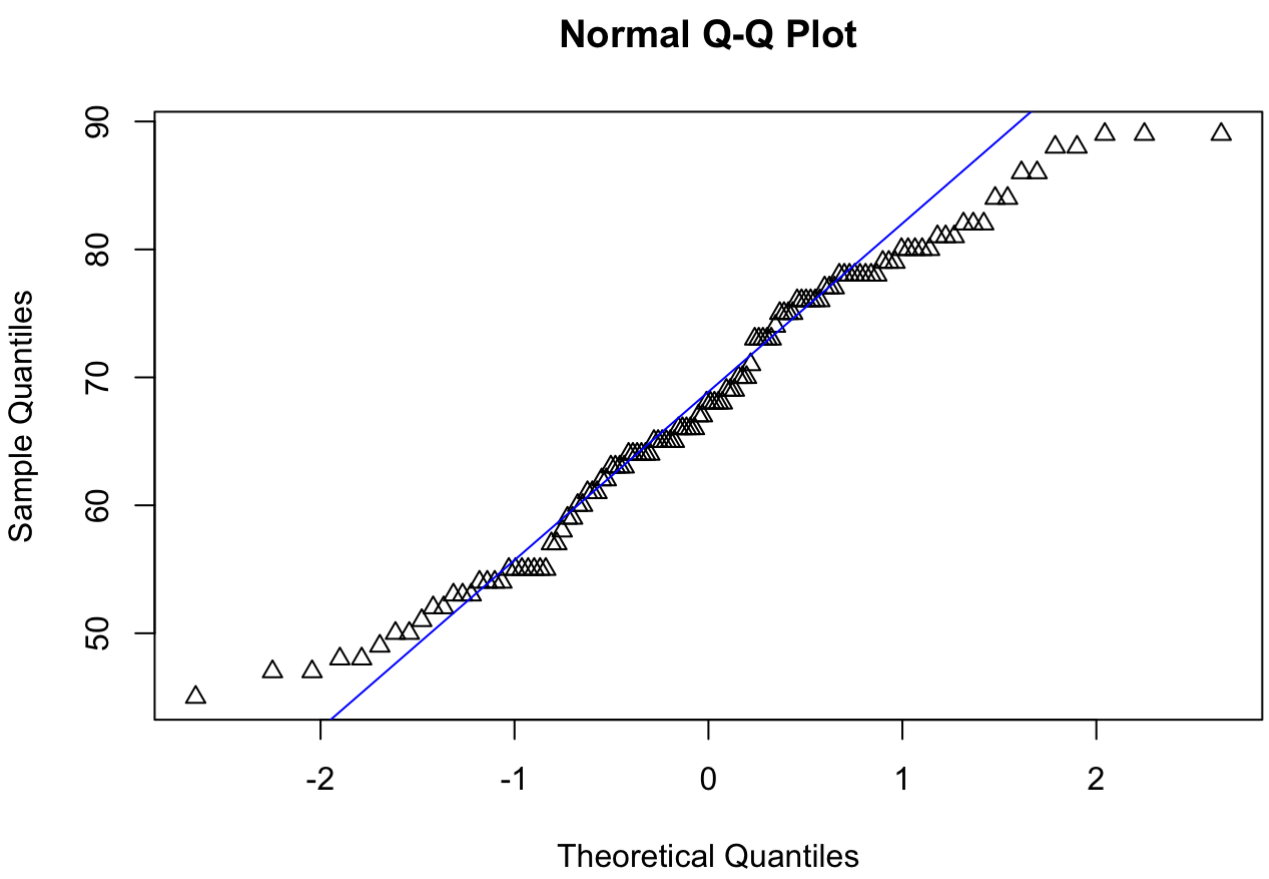
```
qqnorm(forestfireregion1$Temperature, pch=2)
qqline(forestfireregion1$Temperature, col ="green")
```



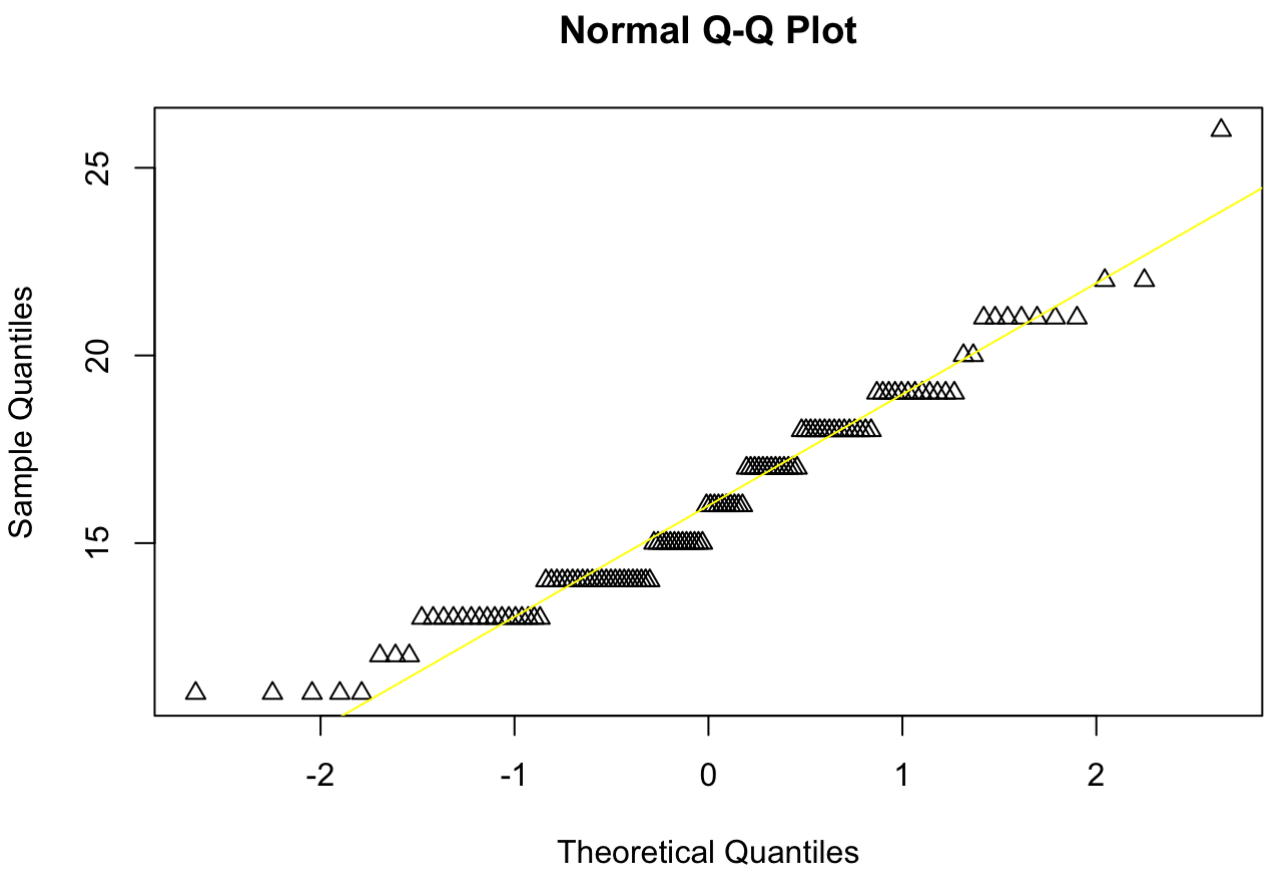

```
qqnorm(forestfireregion1$Rain, pch=2)
qqline(forestfireregion1$Rain, col ="red")
```




```
qqnorm(forestfireregion1$RH, pch=2)
qqline(forestfireregion1$RH, col ="blue")
```




```
qqnorm(forestfireregion1$Ws, pch=2)
qqline(forestfireregion1$Ws, col ="yellow")
```



qqplot gives us the idea about the distribution of data. In the above graphs it can be seen that Temperature, Rain, RH and Ws are having data, approximately normally distributed as most of the points lies on the qqline.

Hypothesis Testing:

Consider the hypothesis as given below,

H0: σ(junemonth$Rain$) $= \sigma(julymonth$Rain)

H1: σ(junemonth$Rain$) $\neq \sigma(julymonth$Rain)

```
junemonth=subset.data.frame(forestfireregion1,forestfireregion1$month==6)
summary(junemonth)
```

```
##       day            month        year       Temperature          RH
##  Min.   : 1.00   Min.   :6   Min.   :2012   Min.   :25.00   Min.   :47.00
##  1st Qu.: 8.25   1st Qu.:6   1st Qu.:2012   1st Qu.:28.25   1st Qu.:61.25
##  Median :15.50   Median :6   Median :2012   Median :30.00   Median :70.00
##  Mean   :15.50   Mean   :6   Mean   :2012   Mean   :29.80   Mean   :70.40
##  3rd Qu.:22.75   3rd Qu.:6   3rd Qu.:2012   3rd Qu.:31.00   3rd Qu.:80.00
##  Max.   :30.00   Max.   :6   Max.   :2012   Max.   :34.00   Max.   :89.00
##        Ws             Rain              FFMC            DMC
##  Min.   :12.00   Min.   : 0.0000   Min.   :28.60   Min.   : 1.10
##  1st Qu.:14.00   1st Qu.: 0.0000   1st Qu.:57.42   1st Qu.: 3.40
##  Median :15.50   Median : 0.0500   Median :79.50   Median : 6.85
##  Mean   :15.77   Mean   : 0.8167   Mean   :69.95   Mean   : 8.86
##  3rd Qu.:17.75   3rd Qu.: 0.4750   3rd Qu.:85.55   3rd Qu.:12.40
##  Max.   :22.00   Max.   :13.1000   Max.   :89.10   Max.   :25.50
##        DC             ISI              BUI              FWI
##  Min.   : 6.90   Min.   :0.00   Min.   : 1.60   Min.   : 0.000
##  1st Qu.: 7.85   1st Qu.:0.75   1st Qu.: 3.90   1st Qu.: 0.325
##  Median :31.05   Median :2.30   Median : 9.05   Median : 1.800
##  Mean   :36.34   Mean   :2.93   Mean   :10.88   Mean   : 3.847
##  3rd Qu.:55.42   3rd Qu.:5.40   3rd Qu.:15.57   3rd Qu.: 7.100
##  Max.   :92.80   Max.   :9.20   Max.   :29.70   Max.   :15.000
##    Classes
##  Length:30
##  Class :character
##  Mode  :character
##
##
##
```

```
julymonth=subset.data.frame(forestfireregion1,forestfireregion1$month==7)
summary(julymonth)
```

```
##       day           month         year       Temperature          RH
##  Min.   : 1.0   Min.   :7   Min.   :2012   Min.   :27.00   Min.   :48.00
##  1st Qu.: 8.5   1st Qu.:7   1st Qu.:2012   1st Qu.:30.50   1st Qu.:65.00
##  Median :16.0   Median :7   Median :2012   Median :32.00   Median :69.00
##  Mean   :16.0   Mean   :7   Mean   :2012   Mean   :31.84   Mean   :69.58
##  3rd Qu.:23.5   3rd Qu.:7   3rd Qu.:2012   3rd Qu.:33.00   3rd Qu.:76.00
##  Max.   :31.0   Max.   :7   Max.   :2012   Max.   :36.00   Max.   :81.00
##        Ws              Rain             FFMC            DMC
##  Min.   :13.00   Min.   :0.0000   Min.   :55.70   Min.   : 2.40
##  1st Qu.:14.00   1st Qu.:0.0000   1st Qu.:71.30   1st Qu.: 7.65
##  Median :16.00   Median :0.0000   Median :81.40   Median : 9.70
##  Mean   :16.35   Mean   :0.2258   Mean   :78.06   Mean   :12.14
##  3rd Qu.:18.50   3rd Qu.:0.3000   3rd Qu.:85.75   3rd Qu.:14.45
##  Max.   :22.00   Max.   :1.4000   Max.   :90.30   Max.   :31.90
##        DC              ISI              BUI             FWI
##  Min.   :  8.30   Min.   : 0.700   Min.   : 2.80   Min.   : 0.300
##  1st Qu.: 20.80   1st Qu.: 1.400   1st Qu.: 8.20   1st Qu.: 0.850
##  Median : 40.40   Median : 2.800   Median :12.50   Median : 3.400
##  Mean   : 54.32   Mean   : 3.623   Mean   :15.48   Mean   : 5.229
##  3rd Qu.: 79.80   3rd Qu.: 5.400   3rd Qu.:20.10   3rd Qu.: 8.400
##  Max.   :145.70   Max.   :10.000   Max.   :41.20   Max.   :15.700
##    Classes
##  Length:31
##  Class :character
##  Mode  :character
##
##
##
```

```
s1=sd(junemonth$Rain)
s2=sd(julymonth$Rain)
n1=length(junemonth$Rain)
n2=length(julymonth$Rain)
fvalue=s1^2/s2^2
fvalue
```

```
## [1] 39.04855
```

```
alpha=0.05
f1=qf(alpha/2,n1-1,n2-1)
f1
```

```
## [1] 0.4779391
```

```
f2=qf(1-alpha/2,n1-1,n2-1)
f2
```

```
## [1] 2.082661
```

```
var.test(julymonth$Rain,junemonth$Rain,alternative = "two.sided",conf.level = 0.95)
```

```
##
##  F test to compare two variances
##
## data:  julymonth$Rain and junemonth$Rain
## F = 0.025609, num df = 30, denom df = 29, p-value < 2.2e-16
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.01223961 0.05333515
## sample estimates:
## ratio of variances
##          0.02560914
```

The variances are not equal because the p value is much lesser than significance level and the fvalue doesn't lie between f1 and f2, hence we reject the hypothesis.

For unknown mean and unequal variances

> Consider the hypothesis as given below,

H0: $\mu(\text{junemonth}Rain) = \mu(julymonth\text{Rain})$

H1: $\mu(\text{junemonth}Rain) \neq \mu(julymonth\text{Rain})$

```
t.test(junemonth$Rain,julymonth$Rain,var.equal = FALSE,conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  junemonth$Rain and julymonth$Rain
## t = 1.3123, df = 30.437, p-value = 0.1992
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.3281092  1.5098296
## sample estimates:
## mean of x mean of y
## 0.8166667 0.2258065
```

We are using t.test because we have two means and the variance are unknown and are not equal. The p-value is greater than alpha i.e., 0.05. So we can accept the null hypothesis H0 and agree that the increase of total rain in the month of June is equal to the increase of total rain in the month of July. <h/5>

Hypothesis testing on one sample:

> Consider the hypothesis as given below,

H0: μ(RH) = 68

H1: μ(RH) ≠ 68

```
fr<- forestfireregion1 %>% subset(Classes == "fire", select= RH)
nofr <- forestfireregion1 %>% subset(Classes == "not fire", select= RH)
t.test(nofr, mu = 68)
```

```
##
##  One Sample t-test
##
## data:  nofr
## t = 2.7857, df = 62, p-value = 0.007077
## alternative hypothesis: true mean is not equal to 68
## 95 percent confidence interval:
##  69.13861 74.92488
## sample estimates:
## mean of x
##  72.03175
```

According to p value we should reject H0, when there is no fire average relative humidity would not be equal to 68.

Corelation

r<0.3, weak correlation

0.3<r<0.7, moderate correlation

r>0.7, high correlation

Multiple Regression Testing on Region 1

```
i <- sample(2, nrow(forestfireregion1), replace=TRUE, prob=c(0.8, 0.2))
Region1Training = forestfireregion1[i==1,]
Region1Testing = forestfireregion1[i==2,]
```

Correlation of training dataset attributes except Classes.

```
cor(subset(Region1Training, select=-Classes))
```

```
## Warning in stats::cor(x, y, ...): the standard deviation is zero
```

```
##                     day        month year Temperature          RH          Ws
## day          1.00000000  0.018338985   NA  0.19112283 -0.19201470  0.05557451
## month        0.01833898  1.000000000   NA -0.02295685 -0.04155726 -0.06902946
## year                 NA           NA    1          NA          NA          NA
## Temperature  0.19112283 -0.022956853   NA  1.00000000 -0.70040151 -0.15907256
## RH          -0.19201470 -0.041557257   NA -0.70040151  1.00000000  0.27637229
## Ws           0.05557451 -0.069029455   NA -0.15907256  0.27637229  1.00000000
## Rain        -0.07178322  0.107520484   NA -0.46297731  0.36221547  0.31424465
## FFMC         0.25829211 -0.037866065   NA  0.76164237 -0.68528277 -0.20602991
## DMC          0.52157749 -0.013382873   NA  0.57921812 -0.40713175  0.06375240
## DC           0.56376396 -0.005264077   NA  0.56928575 -0.37599922  0.13025278
## ISI          0.34202331  0.016607861   NA  0.67587681 -0.62031928 -0.05671840
## BUI          0.54218686 -0.011803296   NA  0.57784427 -0.39907817  0.08977291
## FWI          0.43599848  0.021408972   NA  0.62288779 -0.51823967  0.02761680
##                    Rain        FFMC         DMC          DC         ISI
## day         -0.07178322  0.25829211  0.52157749  0.563763962  0.34202331
## month        0.10752048 -0.03786607 -0.01338287 -0.005264077  0.01660786
## year                 NA          NA          NA          NA          NA
## Temperature -0.46297731  0.76164237  0.57921812  0.569285753  0.67587681
## RH           0.36221547 -0.68528277 -0.40713175 -0.375999217 -0.62031928
## Ws           0.31424465 -0.20602991  0.06375240  0.130252779 -0.05671840
## Rain         1.00000000 -0.57545778 -0.25492035 -0.278909546 -0.36755963
## FFMC        -0.57545778  1.00000000  0.60481345  0.566296261  0.79085694
## DMC         -0.25492035  0.60481345  1.00000000  0.956929643  0.77955184
## DC          -0.27890955  0.56629626  0.95692964  1.000000000  0.74050963
## ISI         -0.36755963  0.79085694  0.77955184  0.740509629  1.00000000
## BUI         -0.26287007  0.59412188  0.99521916  0.980149400  0.77350278
## FWI         -0.28847262  0.66526604  0.91293472  0.874733806  0.94774947
##                    BUI         FWI
## day          0.54218686  0.43599848
## month       -0.01180330  0.02140897
## year                 NA          NA
## Temperature  0.57784427  0.62288779
## RH          -0.39907817 -0.51823967
## Ws           0.08977291  0.02761680
## Rain        -0.26287007 -0.28847262
## FFMC         0.59412188  0.66526604
## DMC          0.99521916  0.91293472
## DC           0.98014940  0.87473381
## ISI          0.77350278  0.94774947
## BUI          1.00000000  0.90990129
## FWI          0.90990129  1.00000000
```

Forward Step Regression Testing ~ Temperature

```
intercept_only <- lm(Temperature ~ 1, data=Region1Training[,1:14])
all <- lm(Temperature~., data=Region1Training[,1:14])
forward <- stepAIC (intercept_only, direction='forward',scope = formula(all))
```

```
## Start:  AIC=241.51
## Temperature ~ 1
##
##           Df Sum of Sq     RSS    AIC
## + FFMC     1    645.39  467.16 157.60
## + RH       1    545.77  566.77 176.74
## + ISI      1    508.22  604.32 183.09
## + FWI      1    431.66  680.89 194.90
## + DMC      1    373.25  739.29 203.05
## + BUI      1    371.48  741.06 203.28
## + DC       1    360.56  751.98 204.73
## + Classes  1    281.10  831.45 214.68
## + Rain     1    238.47  874.07 219.63
## + day      1     40.64 1071.91 239.82
## + Ws       1     28.15 1084.39 240.97
## <none>                1112.55 241.51
## + month    1      0.59 1111.96 243.46
##
## Step:  AIC=157.6
## Temperature ~ FFMC
##
##           Df Sum of Sq    RSS    AIC
## + RH       1    66.805 400.35 144.33
## + DC       1    31.176 435.98 152.77
## + BUI      1    27.012 440.15 153.71
## + FWI      1    26.946 440.21 153.72
## + DMC      1    24.661 442.50 154.23
## + ISI      1    16.058 451.10 156.14
## + Classes  1    11.576 455.58 157.12
## <none>                467.16 157.60
## + Rain     1     1.014 466.15 159.39
## + month    1     0.039 467.12 159.59
## + day      1     0.037 467.12 159.59
## + Ws       1     0.005 467.15 159.60
##
## Step:  AIC=144.33
## Temperature ~ FFMC + RH
##
##           Df Sum of Sq    RSS    AIC
## + DC       1    33.053 367.30 137.79
## + BUI      1    28.199 372.15 139.09
## + DMC      1    25.703 374.65 139.76
## + FWI      1    18.336 382.02 141.68
## <none>                400.35 144.33
## + ISI      1     6.817 393.54 144.62
## + Classes  1     2.887 397.47 145.61
## + Ws       1     2.263 398.09 145.76
## + Rain     1     2.102 398.25 145.80
## + month    1     0.318 400.04 146.25
## + day      1     0.135 400.22 146.29
##
## Step:  AIC=137.8
```

```
## Temperature ~ FFMC + RH + DC
##
##           Df Sum of Sq    RSS    AIC
## + day      1   15.8290 351.47 135.43
## + Classes  1   10.9217 356.38 136.81
## <none>                 367.30 137.79
## + Rain     1    3.4673 363.83 138.86
## + BUI      1    1.2907 366.01 139.45
## + ISI      1    0.9839 366.32 139.53
## + DMC      1    0.7871 366.51 139.58
## + FWI      1    0.5204 366.78 139.66
## + month    1    0.4742 366.83 139.67
## + Ws       1    0.0784 367.22 139.77
##
## Step:  AIC=135.43
## Temperature ~ FFMC + RH + DC + day
##
##           Df Sum of Sq    RSS    AIC
## + Classes  1   12.4123 339.06 133.88
## <none>                 351.47 135.43
## + Rain     1    2.5252 348.95 136.72
## + ISI      1    2.1226 349.35 136.83
## + BUI      1    1.7265 349.75 136.95
## + FWI      1    1.5375 349.93 137.00
## + DMC      1    1.2171 350.25 137.09
## + month    1    0.3773 351.09 137.33
## + Ws       1    0.2292 351.24 137.37
##
## Step:  AIC=133.87
## Temperature ~ FFMC + RH + DC + day + Classes
##
##         Df Sum of Sq    RSS    AIC
## <none>               339.06 133.88
## + ISI    1   1.00795 338.05 135.58
## + Rain   1   0.94756 338.11 135.60
## + FWI    1   0.89331 338.17 135.61
## + BUI    1   0.67542 338.38 135.68
## + Ws     1   0.61038 338.45 135.70
## + month  1   0.44335 338.62 135.75
## + DMC    1   0.39381 338.67 135.76
```

```
forward$anova
```

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## Temperature ~ 1
##
## Final Model:
## Temperature ~ FFMC + RH + DC + day + Classes
##
##
##        Step Df  Deviance Resid. Df Resid. Dev      AIC
## 1                                98  1112.5455 241.5093
## 2    + FFMC  1 645.38662        97   467.1588 157.6034
## 3      + RH  1  66.80542        96   400.3534 144.3256
## 4      + DC  1  33.05257        95   367.3008 137.7951
## 5     + day  1  15.82895        94   351.4719 135.4340
## 6 + Classes  1  12.41226        93   339.0596 133.8746
```

```
ypred_forward = predict(object = forward, newdata = Region1Testing[,1:14])
MAE(y_pred = ypred_forward, y_true = Region1Testing$Temperature)
```

```
## [1] 1.803607
```

```
MSE(y_pred = ypred_forward, y_true = Region1Testing$Temperature)
```

```
## [1] 5.056672
```

Backwards Step Regression Testing ~ Temperature

```
backward <- stepAIC (all, direction='backward')
```

```
## Start:  AIC=143.89
## Temperature ~ day + month + year + RH + Ws + Rain + FFMC + DMC +
##     DC + ISI + BUI + FWI + Classes
##
##
## Step:  AIC=143.89
## Temperature ~ day + month + RH + Ws + Rain + FFMC + DMC + DC +
##     ISI + BUI + FWI + Classes
##
##           Df Sum of Sq    RSS    AIC
## - Rain     1     0.061 325.74 141.91
## - Ws       1     0.334 326.01 141.99
## - month    1     1.440 327.11 142.32
## - ISI      1     2.577 328.25 142.67
## - FWI      1     4.628 330.30 143.28
## - DMC      1     6.579 332.25 143.87
## <none>                 325.67 143.89
## - BUI      1     8.076 333.75 144.31
## - Classes  1    10.743 336.42 145.10
## - day      1    11.902 337.58 145.44
## - DC       1    12.005 337.68 145.47
## - RH       1    34.204 359.88 151.77
## - FFMC     1    35.643 361.32 152.17
##
## Step:  AIC=141.91
## Temperature ~ day + month + RH + Ws + FFMC + DMC + DC + ISI +
##     BUI + FWI + Classes
##
##           Df Sum of Sq    RSS    AIC
## - Ws       1     0.483 326.22 140.05
## - month    1     1.571 327.31 140.38
## - ISI      1     2.762 328.50 140.74
## - FWI      1     4.946 330.68 141.40
## - DMC      1     6.532 332.27 141.87
## <none>                 325.74 141.91
## - BUI      1     8.074 333.81 142.33
## - Classes  1    11.253 336.99 143.27
## - DC       1    12.124 337.86 143.52
## - day      1    12.194 337.93 143.54
## - RH       1    34.555 360.29 149.89
## - FFMC     1    45.731 371.47 152.91
##
## Step:  AIC=140.05
## Temperature ~ day + month + RH + FFMC + DMC + DC + ISI + BUI +
##     FWI + Classes
##
##           Df Sum of Sq    RSS    AIC
## - month    1     1.450 327.67 138.49
## - ISI      1     3.074 329.29 138.98
## - FWI      1     5.217 331.44 139.62
## <none>                 326.22 140.05
## - DMC      1     7.267 333.49 140.23
```

```
## - BUI      1     8.833 335.05 140.70
## - Classes  1    10.795 337.01 141.28
## - day      1    11.925 338.14 141.61
## - DC       1    12.715 338.93 141.84
## - RH       1    38.988 365.21 149.23
## - FFMC     1    47.130 373.35 151.41
##
## Step:  AIC=138.49
## Temperature ~ day + RH + FFMC + DMC + DC + ISI + BUI + FWI +
##     Classes
##
##           Df Sum of Sq    RSS    AIC
## - ISI      1     2.826 330.50 137.34
## - FWI      1     4.740 332.41 137.91
## <none>                 327.67 138.49
## - DMC      1     6.690 334.36 138.49
## - BUI      1     8.142 335.81 138.92
## - Classes  1    10.334 338.00 139.57
## - DC       1    11.913 339.58 140.03
## - day      1    12.346 340.01 140.15
## - RH       1    38.512 366.18 147.49
## - FFMC     1    46.823 374.49 149.72
##
## Step:  AIC=137.34
## Temperature ~ day + RH + FFMC + DMC + DC + BUI + FWI + Classes
##
##           Df Sum of Sq    RSS    AIC
## - FWI      1     4.088 334.58 136.56
## - DMC      1     5.223 335.72 136.89
## - BUI      1     5.984 336.48 137.12
## <none>                 330.50 137.34
## - DC       1     9.565 340.06 138.17
## - day      1    13.349 343.84 139.26
## - Classes  1    14.704 345.20 139.65
## - RH       1    36.919 367.41 145.83
## - FFMC     1    71.022 401.52 154.61
##
## Step:  AIC=136.56
## Temperature ~ day + RH + FFMC + DMC + DC + BUI + Classes
##
##           Df Sum of Sq    RSS    AIC
## - DMC      1     3.801 338.38 135.68
## - BUI      1     4.082 338.67 135.76
## <none>                 334.58 136.56
## - DC       1     7.395 341.98 136.72
## - Classes  1    10.701 345.28 137.68
## - day      1    15.741 350.32 139.11
## - RH       1    58.341 392.92 150.47
## - FFMC     1    66.939 401.52 152.61
##
## Step:  AIC=135.68
## Temperature ~ day + RH + FFMC + DC + BUI + Classes
```

```
##
##           Df Sum of Sq    RSS    AIC
## - BUI      1     0.675 339.06 133.88
## <none>                 338.38 135.68
## - DC       1     8.771 347.15 136.21
## - Classes  1    11.361 349.75 136.95
## - day      1    17.550 355.93 138.68
## - RH       1    56.520 394.90 148.97
## - FFMC     1    78.173 416.56 154.25
##
## Step:  AIC=133.87
## Temperature ~ day + RH + FFMC + DC + Classes
##
##           Df Sum of Sq    RSS    AIC
## <none>                 339.06 133.88
## - Classes  1    12.412 351.47 135.43
## - day      1    17.319 356.38 136.81
## - RH       1    56.006 395.07 147.01
## - DC       1    58.361 397.42 147.60
## - FFMC     1    77.644 416.70 152.29
```

```
backward$anova
```

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## Temperature ~ day + month + year + RH + Ws + Rain + FFMC + DMC +
##     DC + ISI + BUI + FWI + Classes
##
## Final Model:
## Temperature ~ day + RH + FFMC + DC + Classes
##
##
##      Step Df  Deviance Resid. Df Resid. Dev      AIC
## 1                              86   325.6750 143.8872
## 2  - year  0 0.0000000        86   325.6750 143.8872
## 3  - Rain  1 0.0612940        87   325.7363 141.9058
## 4    - Ws  1 0.4831406        88   326.2194 140.0526
## 5 - month  1 1.4499042        89   327.6693 138.4916
## 6   - ISI  1 2.8261313        90   330.4954 137.3418
## 7   - FWI  1 4.0881299        91   334.5836 136.5589
## 8   - DMC  1 3.8006470        92   338.3842 135.6771
## 9   - BUI  1 0.6754233        93   339.0596 133.8746
```

```
summary(backward)
```

```
##
## Call:
## lm(formula = Temperature ~ day + RH + FFMC + DC + Classes, data = Region1Training[,
##     1:14])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.7998 -1.5485 -0.0552  1.5504  5.4858
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)     28.623851   3.264020   8.770 8.19e-14 ***
## day             -0.057646   0.026448  -2.180 0.031815 *
## RH              -0.096082   0.024514  -3.919 0.000169 ***
## FFMC             0.110156   0.023870   4.615 1.26e-05 ***
## DC               0.022756   0.005688   4.001 0.000127 ***
## Classesnot fire  1.135361   0.615326   1.845 0.068200 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.909 on 93 degrees of freedom
## Multiple R-squared:  0.6952, Adjusted R-squared:  0.6789
## F-statistic: 42.43 on 5 and 93 DF,  p-value: < 2.2e-16
```

```
ypred_backward = predict(object = backward, newdata = Region1Testing[,1:14])
MAE(y_pred = ypred_backward, y_true = Region1Testing$Temperature)
```

```
## [1] 1.803607
```

```
MSE(y_pred = ypred_backward, y_true = Region1Testing$Temperature)
```

```
## [1] 5.056672
```

Response as Wind

Forward Step Regression Testing ~ Wind Speed

```
intercept_only <- lm( Ws~ 1, data=Region1Training[,1:14])
all <- lm(Ws~., data=Region1Testing[,1:14])
forward <- stepAIC (intercept_only, direction='forward',scope = formula(all))
```

```
## Start:  AIC=196.34
## Ws ~ 1
##
##               Df Sum of Sq    RSS    AIC
## + Rain         1    69.618 635.37 188.05
## + RH           1    53.848 651.14 190.48
## + FFMC         1    29.926 675.06 194.05
## + Temperature  1    17.839 687.15 195.81
## <none>                     704.99 196.34
## + DC           1    11.961 693.03 196.65
## + Classes      1    11.062 693.93 196.78
## + BUI          1     5.682 699.31 197.54
## + month        1     3.359 701.63 197.87
## + DMC          1     2.865 702.12 197.94
## + ISI          1     2.268 702.72 198.02
## + day          1     2.177 702.81 198.04
## + FWI          1     0.538 704.45 198.27
##
## Step:  AIC=188.05
## Ws ~ Rain
##
##               Df Sum of Sq    RSS    AIC
## + DC           1    36.296 599.08 184.23
## + BUI          1    22.503 612.87 186.48
## + RH           1    21.440 613.93 186.65
## + DMC          1    15.604 619.77 187.59
## <none>                     635.37 188.05
## + FWI          1    10.756 624.62 188.36
## + month        1     7.540 627.83 188.87
## + day          1     4.326 631.05 189.37
## + ISI          1     2.817 632.56 189.61
## + FFMC         1     0.669 634.70 189.95
## + Classes      1     0.286 635.09 190.00
## + Temperature  1     0.166 635.21 190.02
##
## Step:  AIC=184.23
## Ws ~ Rain + DC
##
##               Df Sum of Sq    RSS    AIC
## + RH           1    46.382 552.69 178.25
## + DMC          1    35.889 563.19 180.11
## + BUI          1    31.564 567.51 180.87
## + FFMC         1    21.079 578.00 182.68
## + Classes      1    18.493 580.58 183.12
## + Temperature  1    16.935 582.14 183.39
## + FWI          1    14.593 584.48 183.78
## + ISI          1    14.076 585.00 183.87
## <none>                     599.08 184.23
## + month        1     8.427 590.65 184.82
## + day          1     2.650 596.43 185.79
##
## Step:  AIC=178.25
```

```
## Ws ~ Rain + DC + RH
##
##               Df Sum of Sq    RSS    AIC
## + DMC          1   22.4901 530.20 176.14
## + BUI          1   19.3654 533.33 176.72
## <none>                     552.69 178.25
## + Classes      1    8.7379 543.96 178.67
## + month        1    5.5038 547.19 179.26
## + day          1    2.5330 550.16 179.79
## + FWI          1    1.2163 551.48 180.03
## + FFMC         1    0.7073 551.99 180.12
## + ISI          1    0.0840 552.61 180.23
## + Temperature  1    0.0043 552.69 180.25
##
## Step:  AIC=176.14
## Ws ~ Rain + DC + RH + DMC
##
##               Df Sum of Sq    RSS    AIC
## <none>                     530.20 176.14
## + BUI          1   10.4884 519.72 176.16
## + month        1    6.7206 523.48 176.87
## + day          1    3.9585 526.25 177.39
## + Classes      1    3.3152 526.89 177.51
## + FWI          1    2.5663 527.64 177.66
## + ISI          1    1.9726 528.23 177.77
## + FFMC         1    0.2484 529.96 178.09
## + Temperature  1    0.0780 530.13 178.12
```

```
forward$anova
```

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## Ws ~ 1
##
## Final Model:
## Ws ~ Rain + DC + RH + DMC
##
##
##     Step Df Deviance Resid. Df Resid. Dev      AIC
## 1                           98   704.9899 196.3433
## 2 + Rain  1 69.61754         97   635.3724 188.0500
## 3   + DC  1 36.29629         96   599.0761 184.2266
## 4   + RH  1 46.38161         95   552.6945 178.2489
## 5  + DMC  1 22.49010         94   530.2044 176.1361
```

```
summary(forward)
```

```
##
## Call:
## lm(formula = Ws ~ Rain + DC + RH + DMC, data = Region1Training[,
##     1:14])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.7536 -1.5909 -0.4412  1.6550  5.3184
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 10.94352    1.80996   6.046 2.98e-08 ***
## Rain         0.33592    0.10301   3.261  0.00155 **
## DC           0.04909    0.01675   2.930  0.00425 **
## RH           0.06027    0.02492   2.418  0.01753 *
## DMC         -0.15629    0.07827  -1.997  0.04874 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.375 on 94 degrees of freedom
## Multiple R-squared:  0.2479, Adjusted R-squared:  0.2159
## F-statistic: 7.747 on 4 and 94 DF,  p-value: 1.934e-05
```

```
ypred_for <- predict(object = forward, newdata = Region1Testing[,1:14])
MAE(y_pred = ypred_for, y_true = Region1Testing$Ws)
```

```
## [1] 2.723477
```

```
MSE(y_pred = ypred_for, y_true = Region1Testing$Ws)
```

```
## [1] 9.94304
```

Backwards Step Regression Testing ~ Wind Speed

```
backward <- stepAIC (all, direction='backward')
```

```
## Start:  AIC=43.04
## Ws ~ day + month + year + Temperature + RH + Rain + FFMC + DMC +
##     DC + ISI + BUI + FWI + Classes
##
##
## Step:  AIC=43.04
## Ws ~ day + month + Temperature + RH + Rain + FFMC + DMC + DC +
##     ISI + BUI + FWI + Classes
##
##               Df Sum of Sq    RSS    AIC
## - BUI          1     1.324 49.572 41.663
## <none>                     48.249 43.040
## - Temperature  1     4.569 52.818 43.121
## - DMC          1     6.403 54.652 43.906
## - month        1     9.212 57.461 45.059
## - DC           1    15.481 63.730 47.441
## - day          1    19.238 67.486 48.758
## - FFMC         1    23.706 71.955 50.232
## - Classes      1    26.101 74.349 50.985
## - RH           1    29.082 77.330 51.890
## - FWI          1    36.242 84.491 53.926
## - Rain         1    47.253 95.502 56.744
## - ISI          1    47.816 96.065 56.879
##
## Step:  AIC=41.66
## Ws ~ day + month + Temperature + RH + Rain + FFMC + DMC + DC +
##     ISI + FWI + Classes
##
##               Df Sum of Sq    RSS    AIC
## - Temperature  1     3.542 53.114 41.250
## <none>                     49.572 41.663
## - month        1     8.683 58.255 43.375
## - DMC          1    10.621 60.193 44.127
## - day          1    18.858 68.431 47.077
## - FFMC         1    22.487 72.060 48.266
## - Classes      1    26.261 75.834 49.440
## - RH           1    27.775 77.347 49.895
## - FWI          1    34.926 84.498 51.928
## - DC           1    40.013 89.585 53.273
## - Rain         1    46.098 95.670 54.785
## - ISI          1    46.712 96.284 54.932
##
## Step:  AIC=41.25
## Ws ~ day + month + RH + Rain + FFMC + DMC + DC + ISI + FWI +
##     Classes
##
##           Df Sum of Sq     RSS    AIC
## <none>                  53.114 41.250
## - DMC      1     7.182  60.296 42.167
## - month    1     9.112  62.226 42.891
## - day      1    16.538  69.651 45.484
## - FFMC     1    18.968  72.082 46.273
```

```
## - Classes  1    24.287  77.401 47.911
## - RH       1    26.648  79.762 48.602
## - FWI      1    35.250  88.364 50.957
## - DC       1    36.510  89.624 51.283
## - ISI      1    48.288 101.401 54.123
## - Rain     1    72.925 126.039 59.125
```

```
backward$anova
```

```
## Stepwise Model Path
## Analysis of Deviance Table
##
## Initial Model:
## Ws ~ day + month + year + Temperature + RH + Rain + FFMC + DMC +
##     DC + ISI + BUI + FWI + Classes
##
## Final Model:
## Ws ~ day + month + RH + Rain + FFMC + DMC + DC + ISI + FWI +
##     Classes
##
##
##            Step Df Deviance Resid. Df Resid. Dev      AIC
## 1                                  10   48.24860 43.04007
## 2        - year  0 0.000000        10   48.24860 43.04007
## 3         - BUI  1 1.323640        11   49.57224 41.66255
## 4 - Temperature  1 3.541641        12   53.11388 41.24971
```

```
summary(backward)
```

```
##
## Call:
## lm(formula = Ws ~ day + month + RH + Rain + FFMC + DMC + DC +
##     ISI + FWI + Classes, data = Region1Testing[, 1:14])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4093 -1.0167  0.1871  1.2494  2.9182
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     -3.84984   10.84114  -0.355  0.72867
## day              0.17236    0.08917   1.933  0.07718 .
## month            0.84652    0.58998   1.435  0.17688
## RH               0.14622    0.05959   2.454  0.03039 *
## Rain             1.55774    0.38377   4.059  0.00158 **
## FFMC            -0.27315    0.13195  -2.070  0.06067 .
## DMC              0.41964    0.32944   1.274  0.22685
## DC               0.11895    0.04142   2.872  0.01404 *
## ISI              7.89686    2.39084   3.303  0.00631 **
## FWI             -4.12357    1.46119  -2.822  0.01540 *
## Classesnot fire  4.85094    2.07085   2.342  0.03722 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.104 on 12 degrees of freedom
## Multiple R-squared:  0.8082, Adjusted R-squared:  0.6484
## F-statistic: 5.057 on 10 and 12 DF,  p-value: 0.005119
```

```
ypred_bac <- predict(object = backward, newdata = Region1Testing[,1:14])
MAE(y_pred = ypred_bac,y_true = Region1Testing$Ws)
```

```
## [1] 1.236526
```

```
MSE(y_pred = ypred_bac,y_true = Region1Testing$Ws)
```

```
## [1] 2.309299
```

According to MAE, MSE, R-Square and Residual Standard Error Backward stepwise model would be suitable fit over the forward stepwise model.

Confidence and Predection Intervals

```
fitLmtest1 <- lm(FFMC~ Ws, data=forestfireregion1)
new <- data.frame(Ws = 22)
predict(fitLmtest1 , new)
```

```
##        1
## 67.75645
```

```
predict(fitLmtest1, new, interval="confidence")
```

```
##        fit      lwr      upr
## 1 67.75645 61.35258 74.16033
```

```
predict(fitLmtest1, new, interval="prediction")
```

```
##        fit      lwr      upr
## 1 67.75645 36.84937 98.66353
```

Confidence and Predection Intervals

```
fitLmtest1 <- lm(FFMC~ Ws, data=forestfireregion1)
new <- data.frame(Ws = 22)
predict(fitLmtest1 , new)
```

```
##        1
## 67.75645
```

```
predict(fitLmtest1, new, interval="confidence")
```

```
##        fit      lwr      upr
## 1 67.75645 61.35258 74.16033
```

```
predict(fitLmtest1, new, interval="prediction")
```

```
##        fit      lwr      upr
## 1 67.75645 36.84937 98.66353
```

```
fitLmtest2 <- lm(Temperature~ Rain, data=forestfireregion1)
new <- data.frame(Rain = 15)
predict(fitLmtest2 , new)
```

```
##        1
## 21.67907
```

```
predict(fitLmtest2, new, interval="confidence")
```

```
##        fit      lwr      upr
## 1 21.67907 18.55489 24.80325
```

```
predict(fitLmtest2, new, interval="prediction")
```

```
##        fit      lwr      upr
## 1 21.67907 15.12112 28.23703
```

```
fitLmtest3 <- lm(RH ~ Rain, data=forestfireregion1)
new <- data.frame(Rain = 15)
predict(fitLmtest3 , new)
```

```
##        1
## 89.55115
```

```
predict(fitLmtest3, new, interval="confidence")
```

```
##        fit      lwr      upr
## 1 89.55115 78.20467 100.8976
```

```
predict(fitLmtest3, new, interval="prediction")
```

```
##        fit     lwr      upr
## 1 89.55115 65.7338 113.3685
```

Association between the Numerical Variable and Categorical Variable

```
aov1 <- aov(Rain ~Classes, data=forestfireregion1)
summary(aov1)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## Classes       1   78.5   78.53   15.11 0.000167 ***
## Residuals   120  623.8    5.20
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
aov2 <- aov(FFMC ~Classes, data=forestfireregion1)
summary(aov2)
```

```
##              Df Sum Sq Mean Sq F value Pr(>F)
## Classes       1  16198   16198   148.5 <2e-16 ***
## Residuals   120  13093     109
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
aov2 <- aov(RH ~Classes, data=forestfireregion1)
summary(aov2)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## Classes       1   2143  2143.5   19.92 1.83e-05 ***
## Residuals   120  12911   107.6
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
aov2 <- aov(Temperature ~Classes, data=forestfireregion1)
summary(aov2)
```

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## Classes       1  330.6   330.6   39.53 5.42e-09 ***
## Residuals   120 1003.4     8.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Thank you!