Kaggle-Bike Sharing Demand

Gaurav Bansal

Friday, August 15, 2014

```
#Reding the file
bike <- read.csv("train.csv")
str(bike)
```

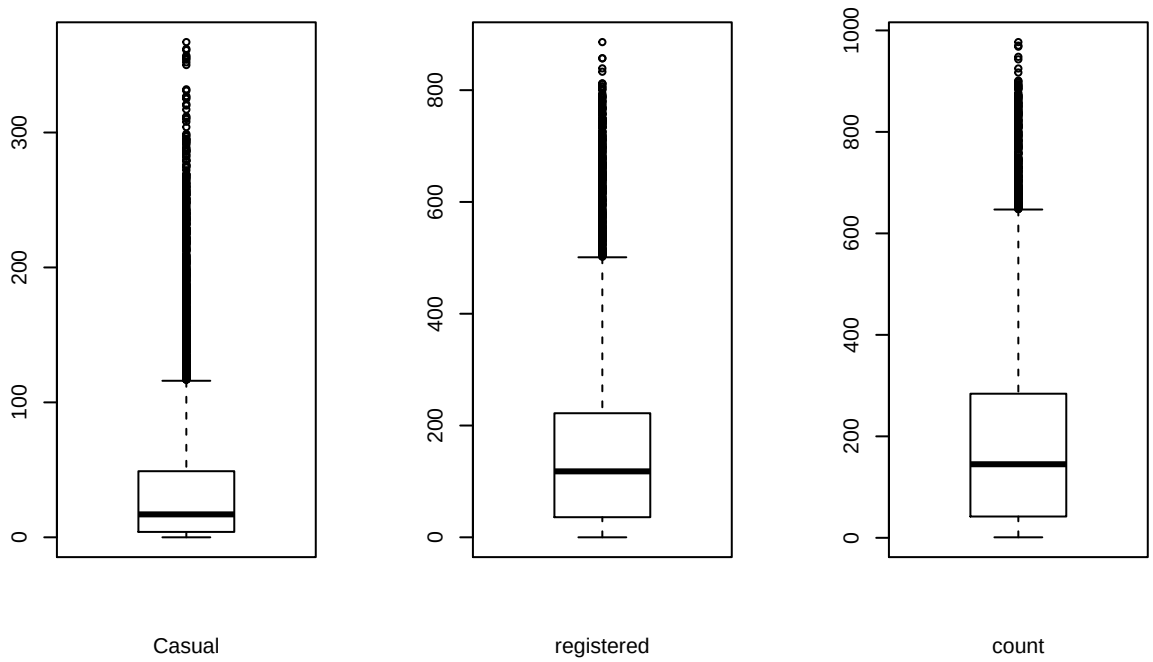
```
## 'data.frame':    10886 obs. of  12 variables:
##  $ datetime  : Factor w/ 10886 levels "2011-01-01 00:00:00",..: 1 2 3 4 5 6 7 8 9 10 ...
##  $ season    : int  1 1 1 1 1 1 1 1 1 1 ...
##  $ holiday   : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ workingday: int  0 0 0 0 0 0 0 0 0 0 ...
##  $ weather   : int  1 1 1 1 1 2 1 1 1 1 ...
##  $ temp      : num  9.84 9.02 9.02 9.84 9.84 ...
##  $ atemp     : num  14.4 13.6 13.6 14.4 14.4 ...
##  $ humidity  : int  81 80 80 75 75 75 80 86 75 76 ...
##  $ windspeed : num  0 0 0 0 0 ...
##  $ casual    : int  3 8 5 3 0 0 2 1 1 8 ...
##  $ registered: int  13 32 27 10 1 1 0 2 7 6 ...
##  $ count     : int  16 40 32 13 1 1 2 3 8 14 ...
```

```
#changing datetime format
bike$datetime <- as.POSIXlt(as.character(bike$datetime))
#basic EDA
summary(bike)
```

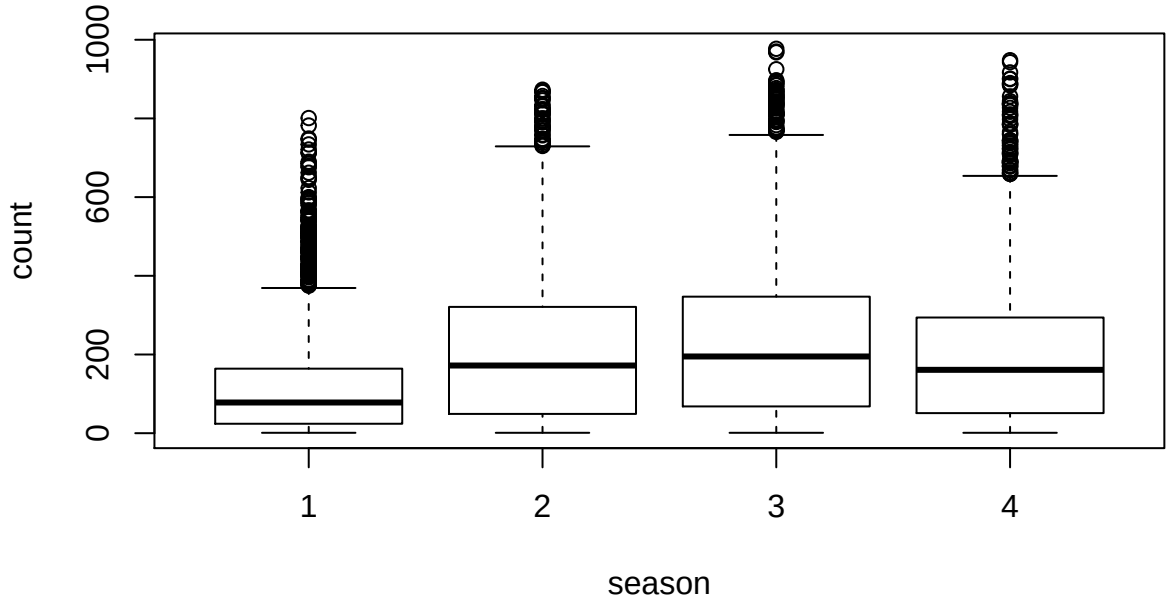
```
##     datetime                       season        holiday
##  Min.   :2011-01-01 00:00:00   Min.   :1.00   Min.   :0.0000
##  1st Qu.:2011-07-02 07:15:00   1st Qu.:2.00   1st Qu.:0.0000
##  Median :2012-01-01 20:30:00   Median :3.00   Median :0.0000
##  Mean   :2011-12-27 05:56:22   Mean   :2.51   Mean   :0.0286
##  3rd Qu.:2012-07-01 12:45:00   3rd Qu.:4.00   3rd Qu.:0.0000
##  Max.   :2012-12-19 23:00:00   Max.   :4.00   Max.   :1.0000
##    workingday       weather          temp           atemp
##  Min.   :0.000   Min.   :1.00   Min.   : 0.82   Min.   : 0.76
##  1st Qu.:0.000   1st Qu.:1.00   1st Qu.:13.94   1st Qu.:16.66
##  Median :1.000   Median :1.00   Median :20.50   Median :24.24
##  Mean   :0.681   Mean   :1.42   Mean   :20.23   Mean   :23.66
##  3rd Qu.:1.000   3rd Qu.:2.00   3rd Qu.:26.24   3rd Qu.:31.06
##  Max.   :1.000   Max.   :4.00   Max.   :41.00   Max.   :45.45
##     humidity       windspeed        casual      registered      count
##  Min.   :  0.0   Min.   : 0.0   Min.   :  0   Min.   :  0   Min.   :  1
##  1st Qu.: 47.0   1st Qu.: 7.0   1st Qu.:  4   1st Qu.: 36   1st Qu.: 42
##  Median : 62.0   Median :13.0   Median : 17   Median :118   Median :145
##  Mean   : 61.9   Mean   :12.8   Mean   : 36   Mean   :156   Mean   :192
##  3rd Qu.: 77.0   3rd Qu.:17.0   3rd Qu.: 49   3rd Qu.:222   3rd Qu.:284
##  Max.   :100.0   Max.   :57.0   Max.   :367   Max.   :886   Max.   :977
```

```
#looking for outliers
par(mfrow=c(1,3))
```

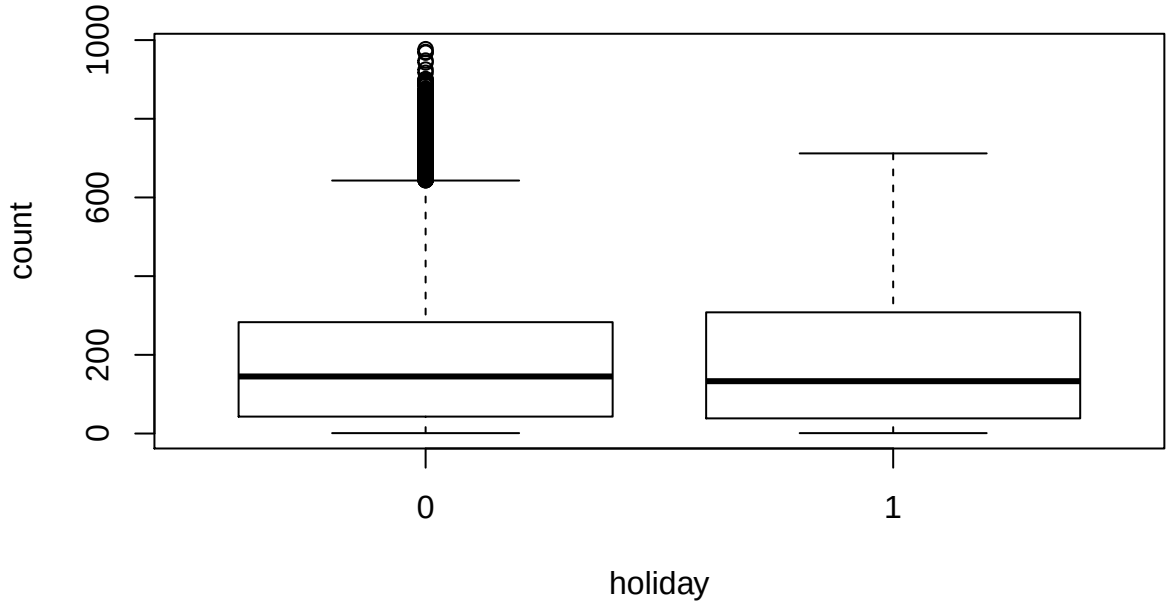
```
boxplot(bike$casual,xlab="Casual")
boxplot(bike$registered,xlab="registered")
boxplot(bike$count,xlab="count")
```




```
#looking for outliers with variables
par(mfrow=c(1,1))
boxplot(count~season,bike,xlab="season",ylab="count")
```

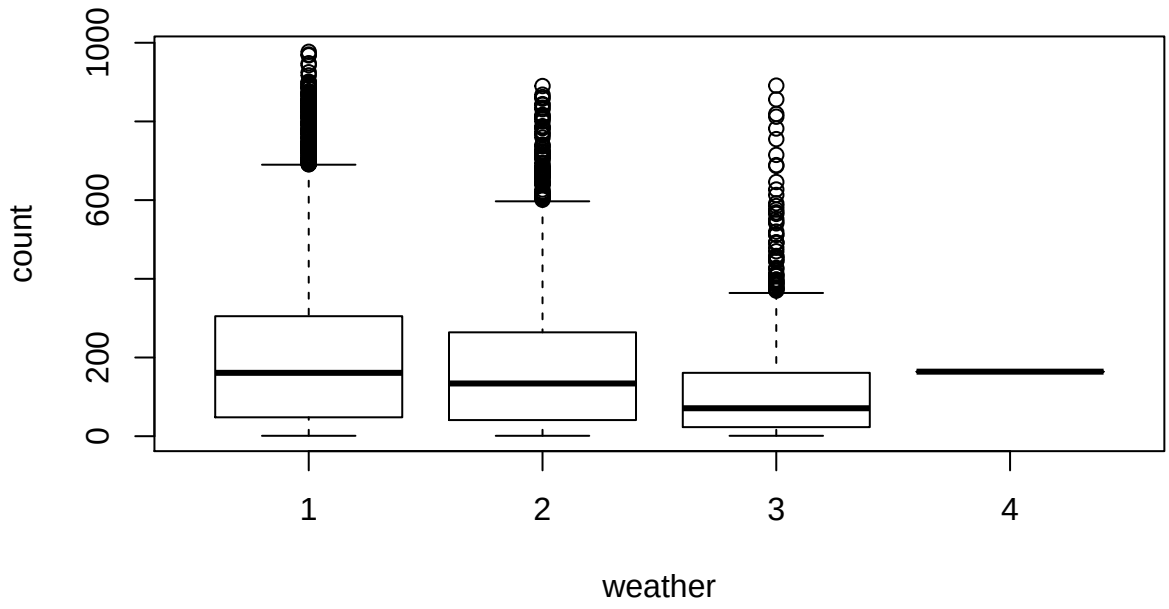



```
boxplot(count~holiday,bike,xlab="holiday",ylab="count")
```

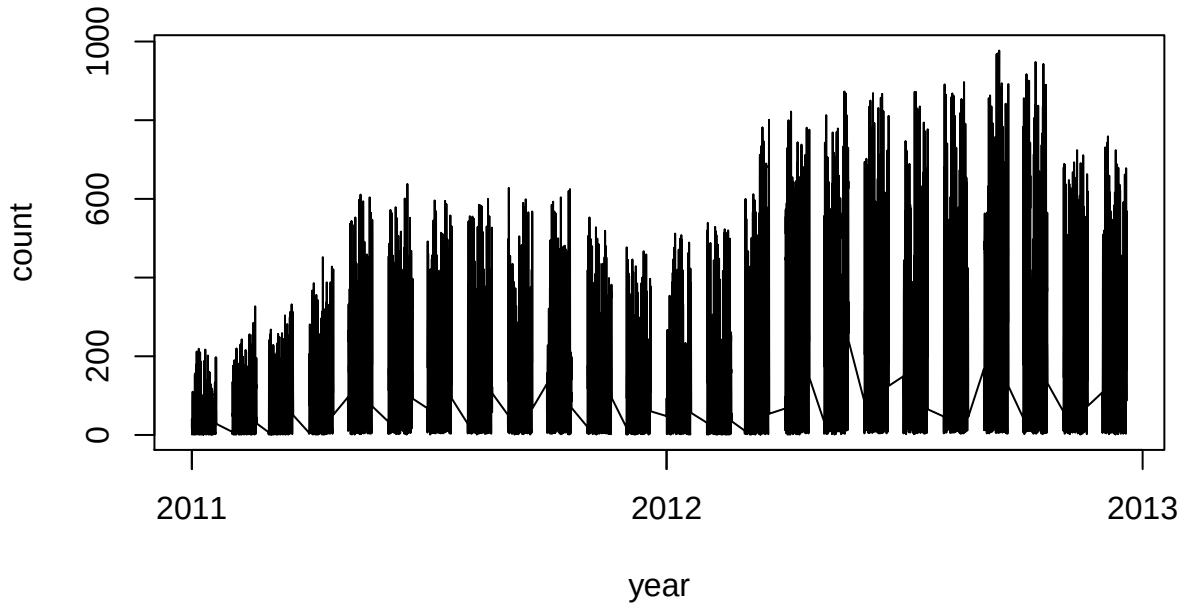


```
boxplot(count~workingday,bike,xlab="workingday",ylab="count")
```


```
boxplot(count~weather,bike,xlab="weather",ylab="count")
```



```r
plot(bike$datetime,bike$count,type="l",xlab="year",ylab="count")
```




```r
#corelation between various variables
cor(bike[,c(-1,-10,-11,-13)])
```

```
##                season    holiday workingday    weather       temp     atemp
## season      1.000000  0.0293676  -0.008126  0.008879  0.2586886  0.264744
## holiday     0.029368  1.0000000  -0.250491 -0.007074  0.0002946 -0.005215
## workingday -0.008126 -0.2504914   1.000000  0.033772  0.0299655  0.024660
## weather     0.008879 -0.0070739   0.033772  1.000000 -0.0550354 -0.055376
## temp        0.258689  0.0002946   0.029966 -0.055035  1.0000000  0.984948
## atemp       0.264744 -0.0052148   0.024660 -0.055376  0.9849481  1.000000
## humidity    0.190610  0.0019287  -0.010880  0.406244 -0.0649488 -0.043536
## windspeed  -0.147121  0.0084087   0.013373  0.007261 -0.0178520 -0.057473
## count       0.163439 -0.0053930   0.011594 -0.128655  0.3944536  0.389784
##              humidity windspeed     count
## season      0.190610 -0.147121  0.163439
## holiday     0.001929  0.008409 -0.005393
## workingday -0.010880  0.013373  0.011594
## weather     0.406244  0.007261 -0.128655
## temp       -0.064949 -0.017852  0.394454
## atemp      -0.043536 -0.057473  0.389784
## humidity    1.000000 -0.318607 -0.317371
## windspeed  -0.318607  1.000000  0.101369
## count      -0.317371  0.101369  1.000000
```