Variables present in the data-set are-

datetime - hourly date + timestamp
season - 1 = spring, 2 = summer, 3 = fall, 4 = winter
holiday - whether the day is considered a holiday
workingday - whether the day is neither a weekend nor holiday
weather - 1: Clear, Few clouds, Partly cloudy, Partly cloudy 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog
temp - temperature in Celsius
atemp - "feels like" temperature in Celsius
humidity - relative humidity
windspeed - wind speed
casual - number of non-registered user rentals initiated
registered - number of registered user rentals initiated
count - number of total rentals

Before moving ahead first of all I have factored data and introduced some new variable in the data-set.

```
> rm(list=ls(all=TRUE))  # start clean
>
> setwd("Location of your Data File")
>
> train = read.csv("train.csv", header = TRUE)
>
> # Making factors
> train$datetime <- strptime(train$datetime, format="%m/%d/%Y %H:%M")
> train$weekday <- weekdays(train$datetime)
> train$hour <- train$datetime$hour
> train_factor <- train
> train_factor$weather <- factor(train$weather)
> train_factor$holiday <- factor(train$holiday)
> train_factor$workingday <- factor(train$workingday)
> train_factor$season <- factor(train$season)
> train_factor$hour <- factor(train_factor$hour)
> train_factor$sunday[train_factor$weekday == "Sunday"] <- "1"
> train_factor$sunday[train_factor$weekday != "Sunday"] <- "0"
> train_factor$sunday <- as.factor(train_factor$sunday)
>
> train_factor$weather1[train_factor$weather == "1"] <- "1"
> train_factor$weather1[train_factor$weather != "1"] <- "0"
> train_factor$weather1 <- factor(train_factor$weather1)
>
> #convert to factor
> train_factor$sunday <- as.factor(train_factor$sunday)
> #create daypart column, default to 4 to make things easier for ourselves
> train_factor$daypart <- "4"
> train_factor$hour<- as.numeric(train_factor$hour)
> #4AM - 10AM = 1
> train_factor$daypart[(train_factor$hour < 10) & (train_factor$hour > 3)]
<- 1
> #11AM - 3PM = 2
> train_factor$daypart[(train_factor$hour < 16) & (train_factor$hour > 9)]
<- 2
> #4PM - 9PM = 3
```

```
> train_factor$daypart[(train_factor$hour < 22) & (train_factor$hour > 15)
] <- 3
> #convert daypart to factor
> train_factor$daypart <- as.factor(train_factor$daypart)
> #convert hour back to factor
> train_factor$hour <- as.factor(train_factor$hour)
> train_factor$weekday <- as.factor(train_factor$weekday)
```

The above code factors the categorical variables and introduced five new variables named weekday, hour, Sunday, wather1, day-part. These variables are extracted from the available variables.

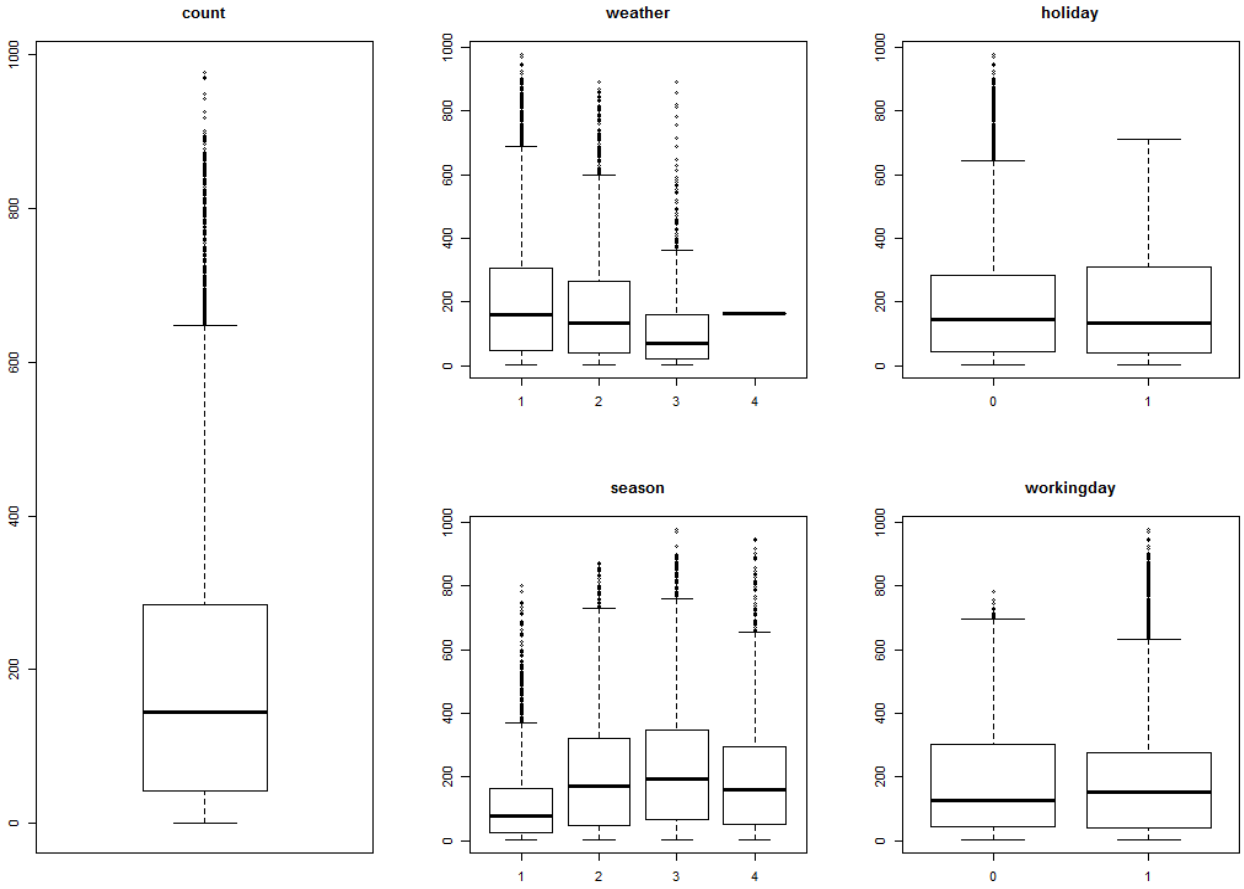
Count against different Independent variables

```
> layout(matrix(c(1,1,2,3,4,5),2,3,byrow=FALSE))

> boxplot(train$count, main="count")

> boxplot(train$count ~ train$weather, main="weather")

> boxplot(train$count ~ train$season, main="season")
> boxplot(train$count ~ train$holiday, main="holiday")
> boxplot(train$count ~ train$workingday, main="workingday")
```

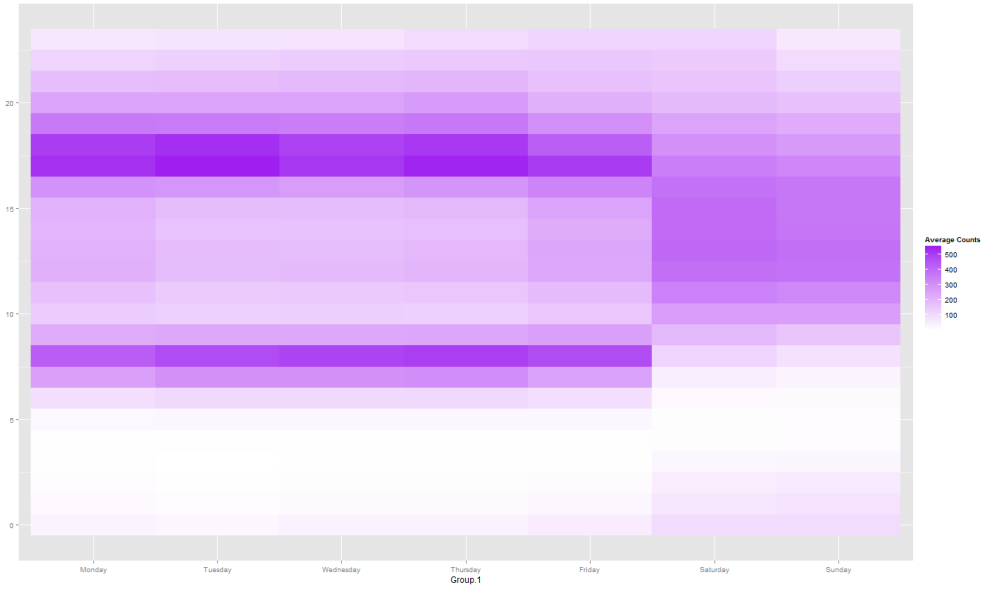


Interpretation:

Observe the five "count" boxplots above, with the larger plot being the overall "count" distribution. We see the median count hover around 150 units, and we see many outlier counts above 600. The range of counts is from 0 to under 1000 units. When stratified by "weather", besides extreme weather (==4), the median count increases, with higher usage count the better the weather. There is not much difference other than the outliers for non-"holiday" days, and also for days which are designated a "workingday". We see increases in median counts for "season" 2 and 3, summer and fall, respectively.

Impact of working day and weekend on count

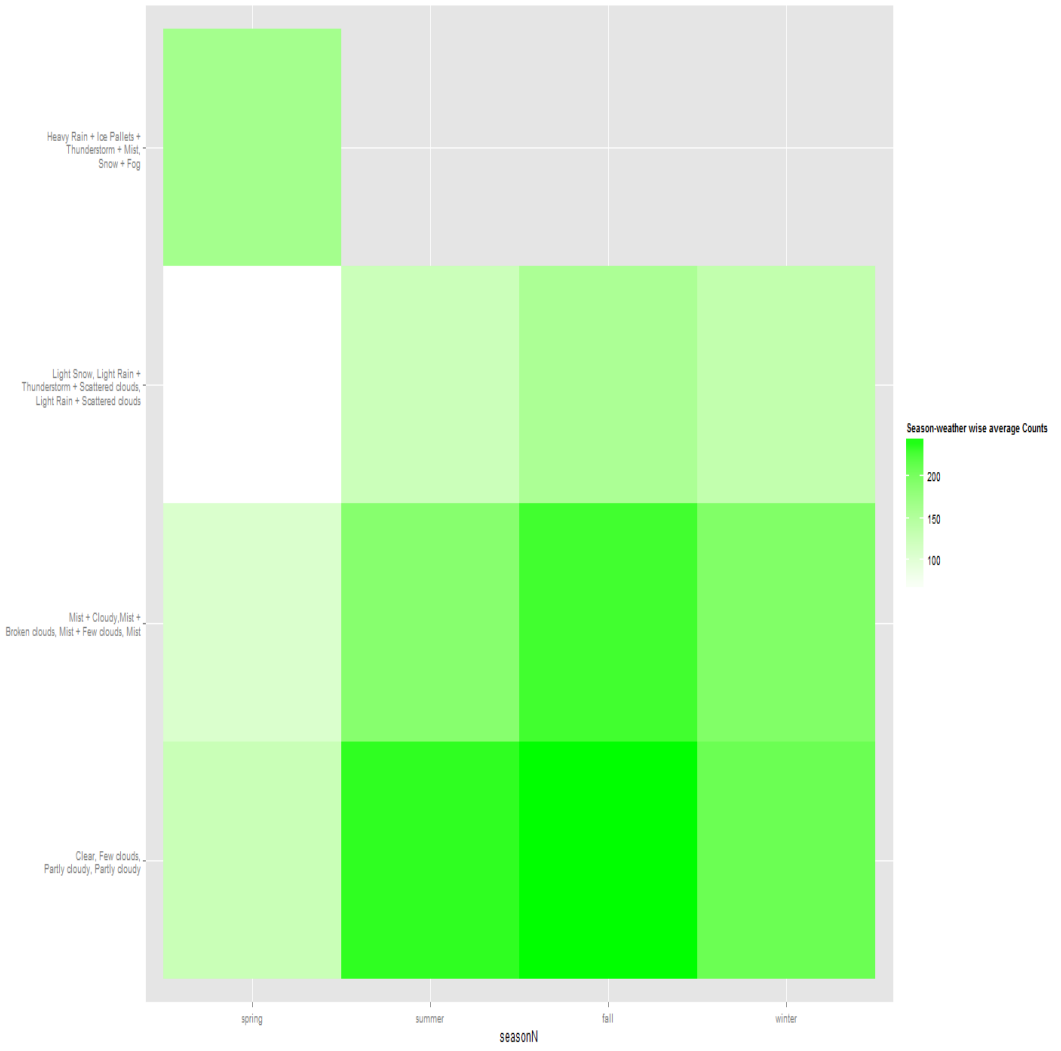
```
> day_hour_counts <- as.data.frame(aggregate(train[,"count"],
list(train$weekday, train$hour), mean))
> day_hour_counts$Group.1 <- factor(day_hour_counts$Group.1,
ordered=TRUE, levels=c("Monday", "Tuesday", "Wednesday",
"Thursday", "Friday", "Saturday", "Sunday"))
> day_hour_counts$hour <-
as.numeric(as.character(day_hour_counts$Group.2))
> # plot heat mat with ggplot
> ggplot(day_hour_counts, aes(x = Group.1, y = hour)) +
geom_tile(aes(fill = x)) + scale_fill_gradient(name="Average
Counts", low="white", high="purple") + theme(axis.title.y =
element_blank())
```



Interpretation:
By analyzing the heat map, we can see that the maximum bikes are rented during regular office hour's i.e 7-9 am when people leave for office and 5-7 pm when people leave for home from the office. So we can clearly see that the majority of people renting the bike are office going people.People are preferring to use bikes to travel to office rather than using their own car or public transport. This way they are exercising plus skipping the traffic.
In weekends the scenario is entirely opposite where bikes are rented mostly in the afternoon.

Season-wise best optimum weather condition to rent a bike

```
> season_weather_counts <-
as.data.frame(aggregate(train[,"count"], list(train$season,
train$weather), mean))
> season_weather_counts$weather <-
as.numeric(as.character(season_weather_counts$Group.2))
> season_weather_counts$season <-
as.numeric(as.character(season_weather_counts$Group.1))
> season_weather_counts$weatherN <-
factor(season_weather_counts$weather)
> season_weather_counts$weatherN <-
factor(season_weather_counts$weather,labels=c("Clear, Few
clouds,\n Partly cloudy, Partly cloudy","Mist + Cloudy,Mist
+\n Broken clouds, Mist + Few clouds, Mist","Light Snow, Light
Rain +\n Thunderstorm + Scattered clouds,\n Light Rain +
Scattered clouds","Heavy Rain + Ice Pallets +\n Thunderstorm +
Mist,\n Snow + Fog"))
> season_weather_counts$seasonN <-
factor(season_weather_counts$season)
> season_weather_counts$seasonN <-
factor(season_weather_counts$season,labels=c("spring","summer"
,"fall","winter"))
> ggplot(season_weather_counts, aes( x = seasonN,y =
weatherN)) + geom_tile(aes(fill = x)) +
scale_fill_gradient(name="Season-weather wise average Counts",
low="white", high="green") + theme(axis.title.y =
element_blank())
```


Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog
Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds
Mist + Cloudy,Mist + Broken clouds, Mist + Few clouds, Mist
Clear, Few clouds, Partly cloudy, Partly cloudy
spring
summer
fall
winter
seasonN
Season-weather wise average Counts
200
150
100

Interpretation:

We can see that the highest number of bikes are rented during summer and fall when there are few clouds and some mist in the air. And people are not preferring to rent the bike during extreme weather conditions i.e. light snow, rain and thunderstorm as there are chances of an accident to take place during such conditions. The count gets decreasing as the weather gets cooler. So people are preferring to rent a bike in a warm day rather than a cold day.