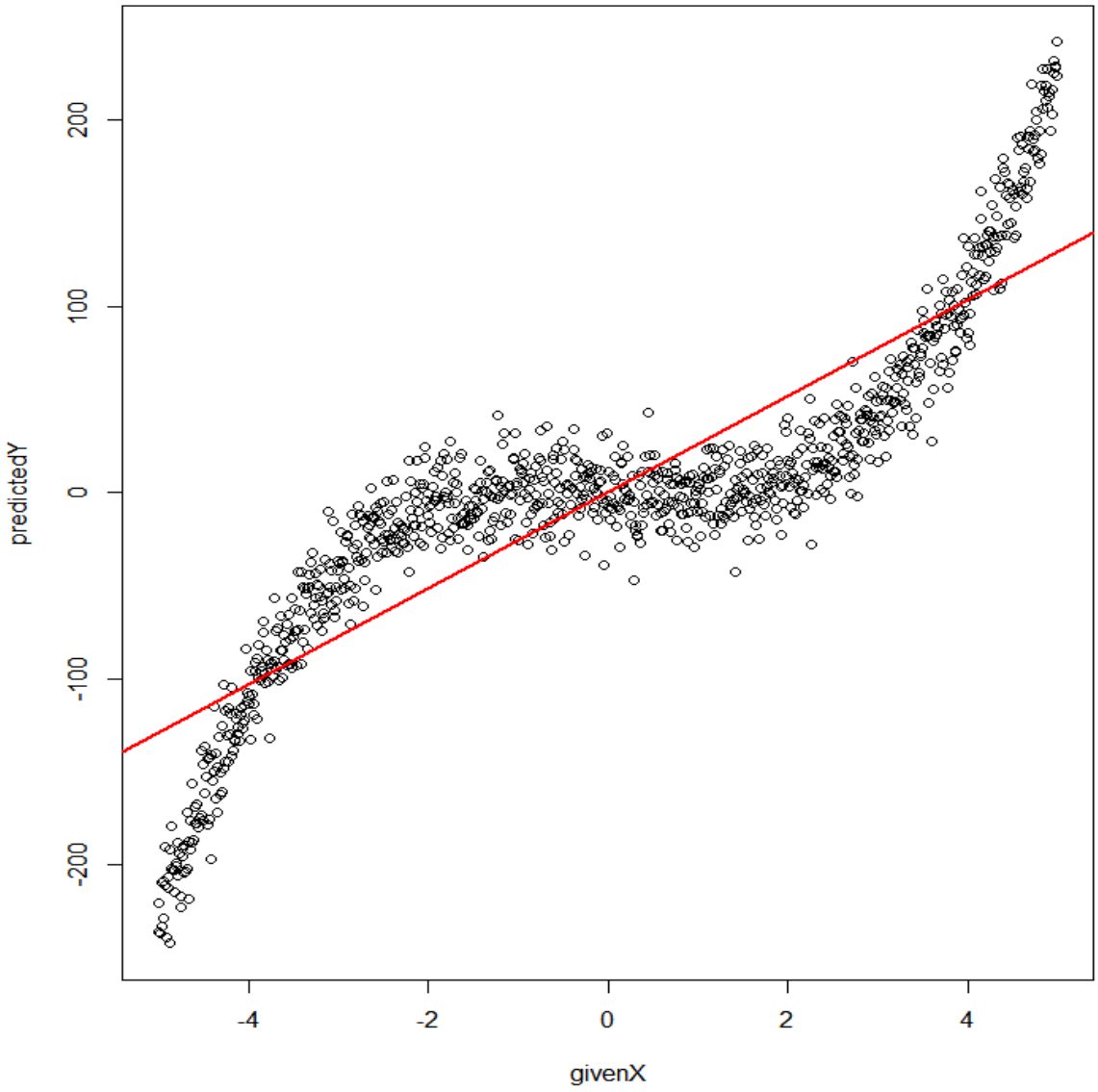
INFO304 Assignment 2 Data Modelling

Usman Shah

1006293

Linear Modelling

1. Model 1



Plot showing x , y values and the given predictions.

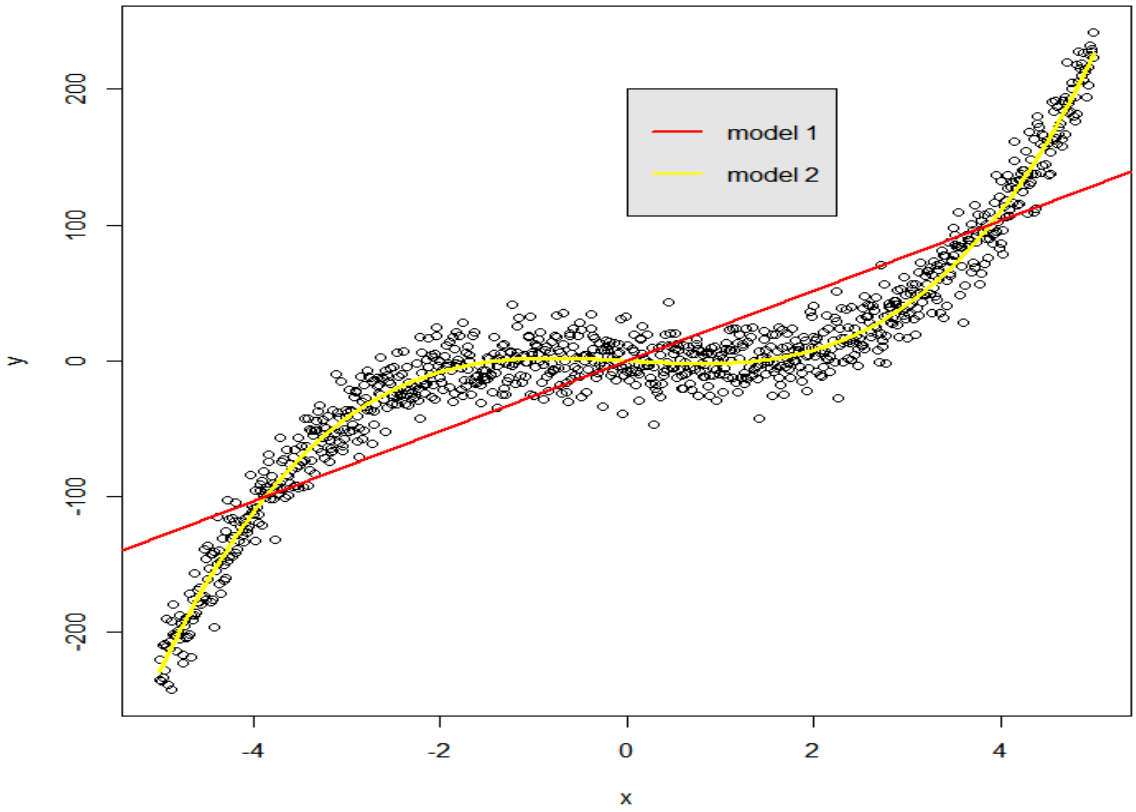
```
#attach(items)
a <- lm(y~x-1, data = items) # create a linear regression model
plot(items$x,items$y,xlab = "givenX", ylab = "predictedY", main = "model 1")
abline(a, col='red', lwd = 2) # plot the line for the best fit
summary(a)
```

Formula for Model 1

2. Model 2

```
### Linear Model 1 #####
items <- read.csv('pdata.csv',sep = ",",header=T) # load the data
names(items)
#attach(items)
a <- lm(y~x-1, data = items) # create a linear regression model
plot(items$x,items$y,xlab = "givenX", ylab = "predictedY", main = "model 1")
abline(a, col='red', lwd = 2) # plot the line for the best fit
summary(a)



### MODEL 2 ####
x <- items$x
x2 <- x^2
x3 <- x^3

b <- lm(y~x+x2+x3,data = items)
fitModel <- fitted(b) # to extract the fitted values from the object b returened by lm.
plot(x,y) # ????i was giving it data instead of specifying x and y.
lines(x,fitModel, col = "yellow",lwd=3.5)
abline(a, col='red', lwd = 2)
legend(0, 200, c("model 1", "model 2"), col = c("red","yellow"),
       text.col = "black",lty = 1, lwd = 2,
       bg = "gray90")
summary(b)
```

R Code for model 1 and model 2



Plot for both model 1 and model 2

```
Call:
lm(formula = y ~ x + x2 + x3, data = items)

Residuals:
    Min      1Q  Median      3Q     Max
-49.811  -9.852  -0.140  10.028  44.270

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.13949    0.70875  -0.197    0.844
x           -3.83238    0.40879  -9.375   <2e-16 ***
x2          -0.05195    0.06327  -0.821    0.412
x3           1.97511    0.02493  79.234   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14.95 on 997 degrees of freedom
Multiple R-squared:  0.9691,    Adjusted R-squared:  0.969
F-statistic: 1.043e+04 on 3 and 997 DF,  p-value: < 2.2e-16
```

model 2 Summary

```
Call:
lm(formula = y ~ x - 1, data = items)

Residuals:
     Min       1Q   Median       3Q      Max
-116.305  -31.866   -1.484   30.927  112.799

Coefficients:
  Estimate Std. Error t value Pr(>|t|)
x  25.8535     0.4411   58.61   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 40.33 on 1000 degrees of freedom
Multiple R-squared:  0.7745,    Adjusted R-squared:  0.7743
F-statistic:  3435 on 1 and 1000 DF,  p-value: < 2.2e-16

> summary(b)

Call:
lm(formula = y ~ x + x2 + x3, data = items)

Residuals:
    Min      1Q  Median      3Q     Max
-49.811  -9.852  -0.140  10.028  44.270

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.13949    0.70875  -0.197    0.844
x           -3.83238    0.40879  -9.375   <2e-16 ***
x2          -0.05195    0.06327  -0.821    0.412
x3           1.97511    0.02493  79.234   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 14.95 on 997 degrees of freedom
Multiple R-squared:  0.9691,    Adjusted R-squared:  0.969
F-statistic: 1.043e+04 on 3 and 997 DF,  p-value: < 2.2e-16

>
```

lm(a) vs lm(b) summary

```
> anova(a,b)
Analysis of Variance Table

Model 1: y ~ x - 1
Model 2: y ~ x + x2 + x3
  Res.Df     RSS Df Sum of Sq      F    Pr(>F)
1   1000 1626284
2    997  222806  3   1403478 2093.4 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

Compare model 1 and model 2

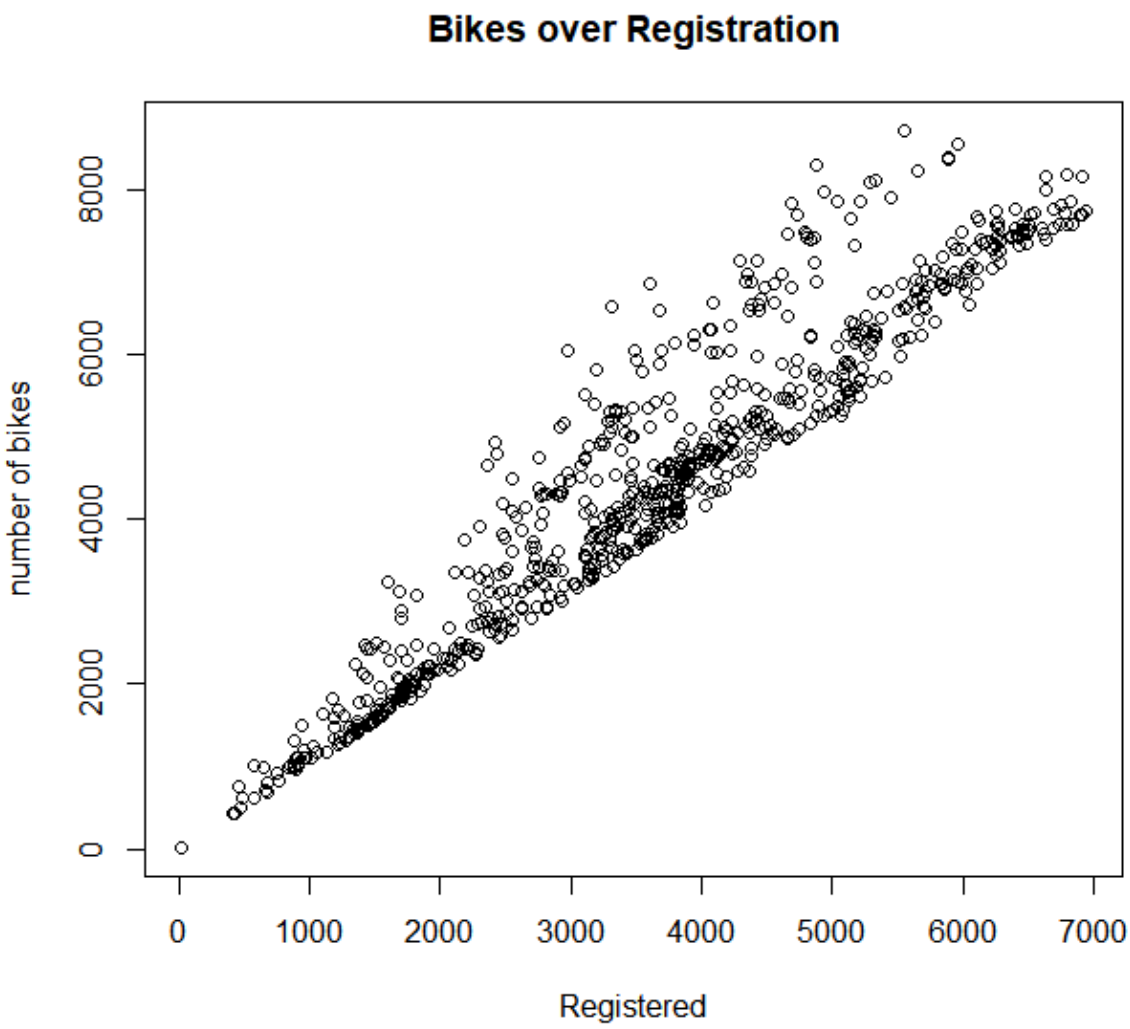
From the summary above we can examine the variables and their related p-value. The variable with largest p-value is least statistically significant, in other words the variables which has p-value very close to zero are very significant. From the summary above we can see which values are more significant and which one are less significant in model 2. Given that x and x3 seems more important in model 2 as there p-value is very close to zero on the other hand x2 is 0.412 which is not very close to zero and is not an important variable in this model. Various statistics can be used to judge the quality of the model. From the summary of linear model1 and model 2, we can see that R-squared value in model2 is very close to 1 which indicates that the model explains a very large portion of variance in the response variable. For example, model2 which uses all the three variables to predict y has an R-squared 0.9691

as compared to model1 which is only using one variable and the R-squared for that is 0.7745 indicates that model2 is a good model to represent the data.

BIKE SHARING – TIME SERIES, DECISION TREES & LINEAR MODELS (90 MARKS)

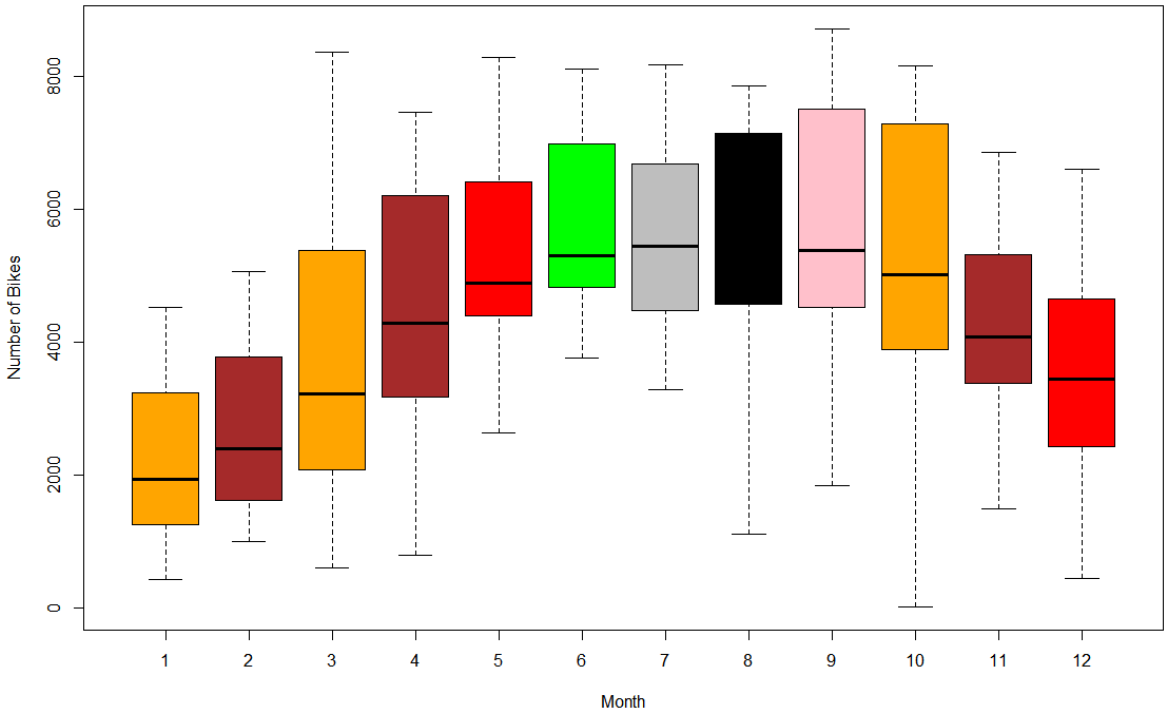
1. The data set day.csv provided to us is generated by bike sharing systems, which is an automatic system for membership, renting and returning a bike. Which makes it easier for the user to rent a bike at one spot and return it at another spot. This data set has a few characterises which are generated by these systems and are very useful for research day.csv data set has the following fields.
    - instant: record index
    - dteday : That is the date on which the bike has rented out.
    - season : four seasons 1 for winter, 2 for spring, 3 for summer and 4 for fall.
    - yr : year (0: 2011, 1:2012)
    - mnth : month January to December given numbers(1 to 12 respectively)
    - holiday : weather day is holiday or not they are extracted from the internet.
    - weekday : It just specifies day of the week(1 to 7)
    - workingday : 1 if it not a weekend nor holiday and 0 if it is a working day.
    - weathersit : 1 the weather is clear, 2 if it is a mist and cloud, 3 for light snow and 4 for heavy rain.
    - temp : Normalized temperature in Celsius.
    - windspeed: Normalized wind speed.
    - casual: count of casual users
    - registered: count of registered users
    - bikes: count of total rental bikes including both casual and registered

    In the graph bikes over registered, we can see that there is a linear relationship. The more the registered people are the more the number bikes will be and that is self explanatory more registered people means you have more people to use the system and number of bikes will increase.

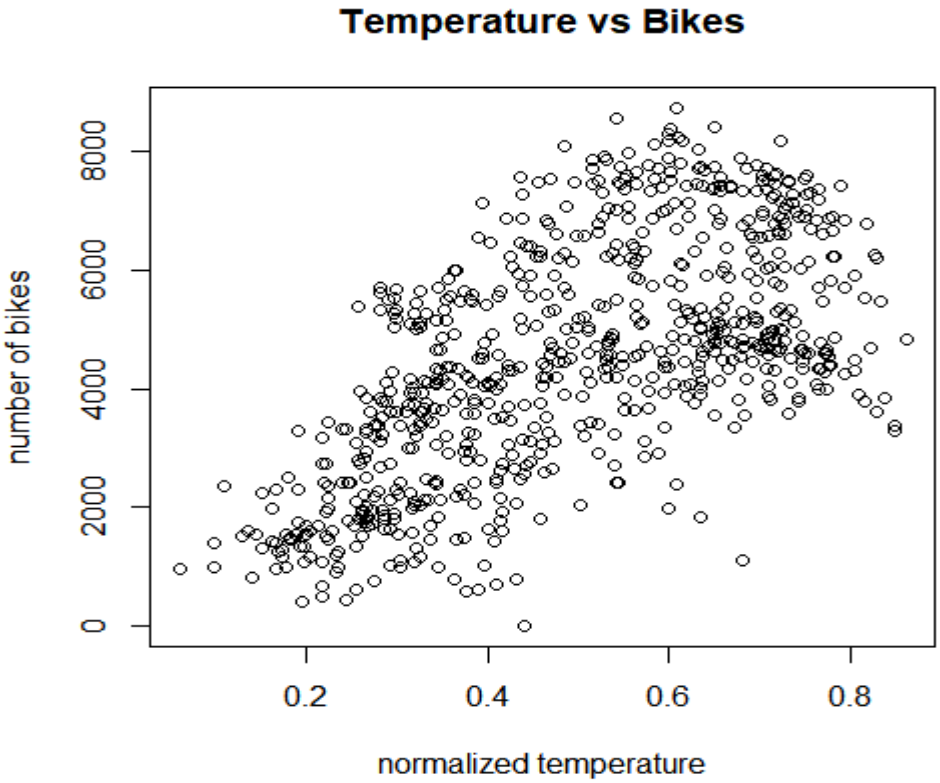


In the graph bikes vs month, we can see that during summer June and July the number of bikes rented out are quite higher as compared to winter. But during the start of fall which is September and October the bikes number is even higher than summer and they might be because in the start of fall the temperature is drops a bit and is not too hot. This is definitely

a seasonality trend in winter not a lot of people are renting bikes summer the number rises.

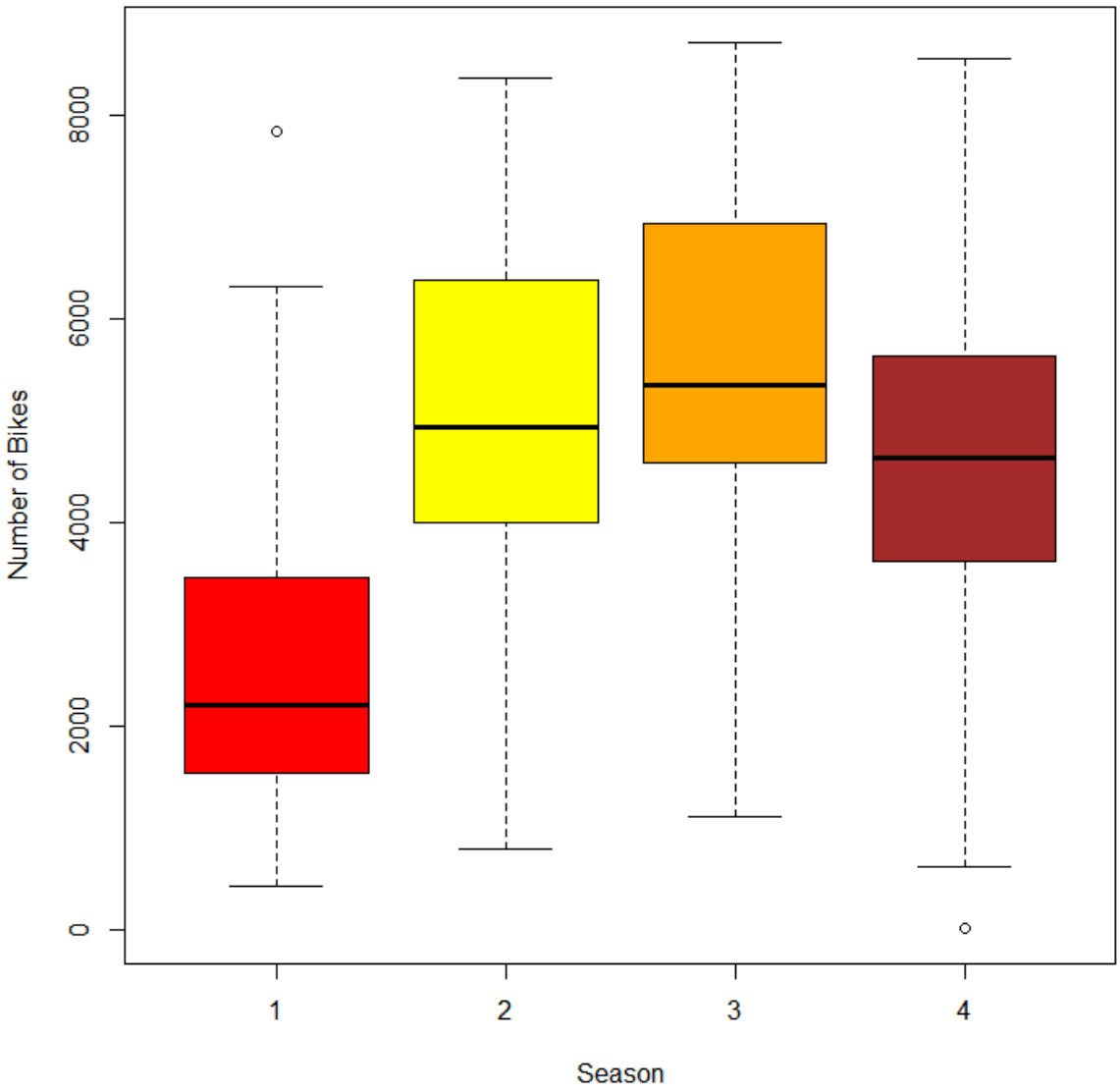


In the plot below Temperature vs Bikes we can see that there is pattern, in very high or very low temperature number of bikes are very low but when the temperature is average and is not too cold nor too hot the number of bikes increases.



plot for bikes over temperature with correlation of 0.627

Number of bikes vs season, 1 is for winter 2 is for spring 3 is for summer and 4 is for fall. We can see that during summer and spring the number of bikes is very higher as compared to winter or fall. There are two points which are out of the box 1 and 4 and they are the outliers here. Which might be the cause of an anomalous event. Like boxing day, Christmas day which is in winter or a storm which causes the number of bikes to drop.



Bikes over Season with correlation 0.4061004

Number of bikes vs year. Here we have two categories 0 which is for 2011 and 1 which is for year 2012. And the plot shows that there is an outlier in box 1 which again might be a cause of anomalous event. The plot shows that when the year increases the number of bikes increases , my assumption for this will be because every year the system is becoming more popular and that is why more people are renting the bikes.

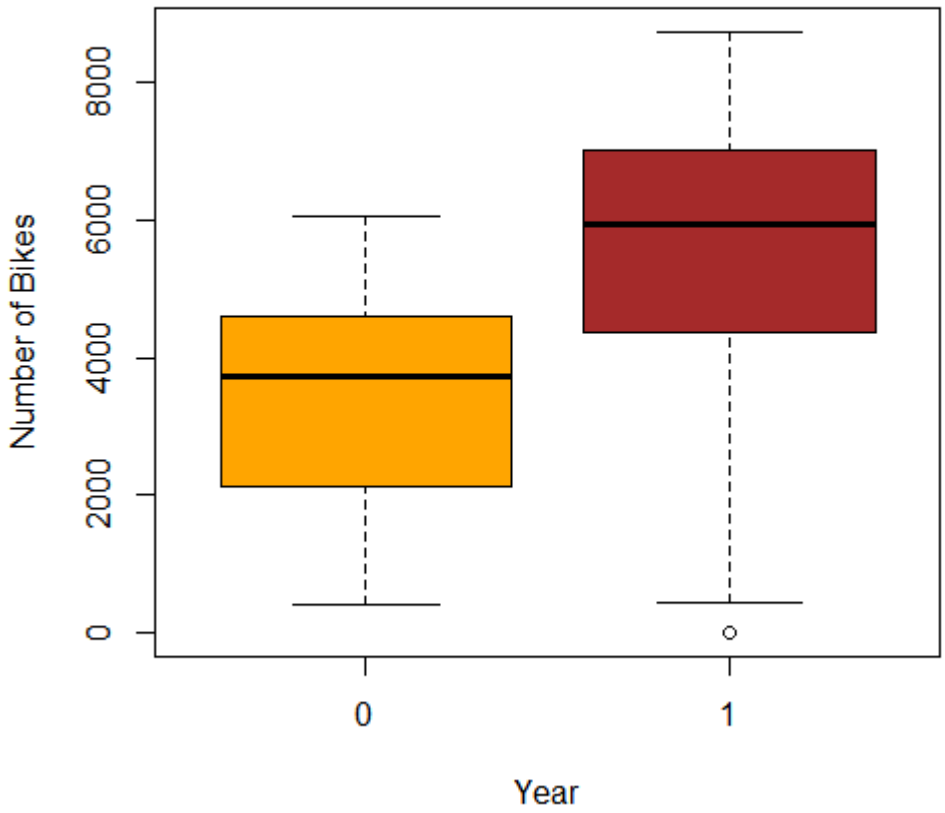


Figure 1 bikes over year with correlation value 0.5667097

Bikes over Holiday I am a bit confused and surprised that the plot has shown more bikes on when it is not a holiday that when it is a holiday. I assume that during the working day people use it as a transport for getting to work.

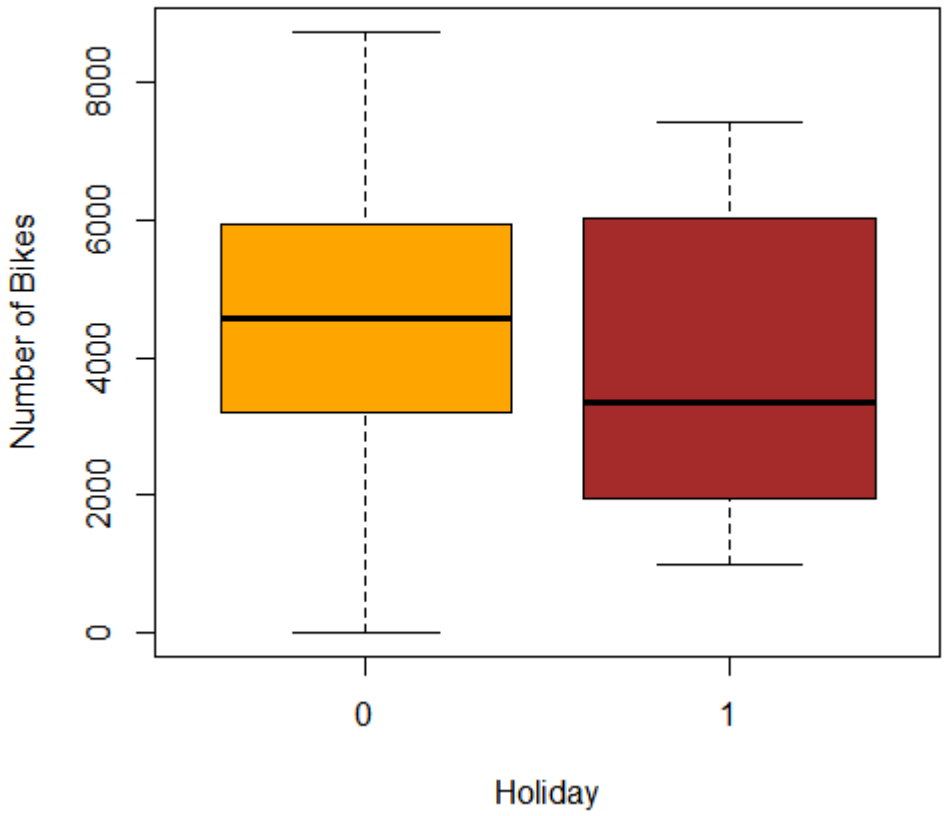
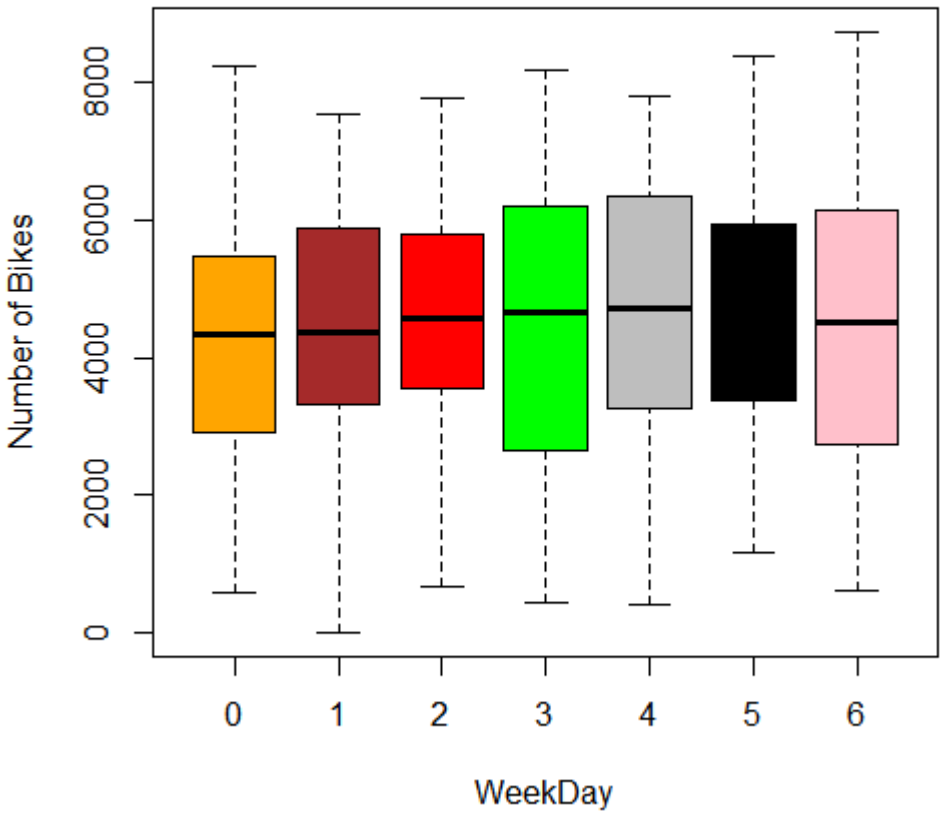


Figure 2 Bikes over Holiday with correlation value -0.06834772

Bikes over week day the over all it looks pretty similar so assume there was no big difference between the days.



Bikes Over WeekDay with correlation 0.06744341

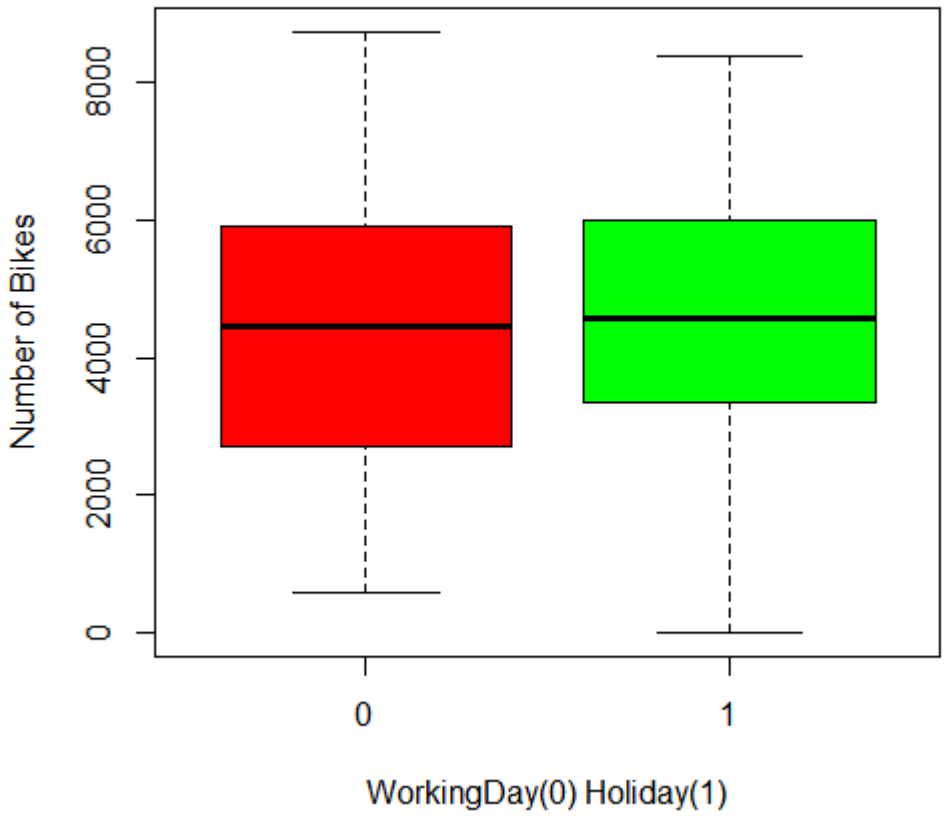


Figure 3 Bikes over Working Day and correlation is 0.06115606.

Bikes vs weather. 1 to 4 which means good to bad weather. The box plot shows that that when the weather is good the number of bikes are higher and keeps dropping down when the weather is bad. When the weather is really bad there are no bikes rented out. I mean no one want to ride a bike when it is really bad weather.

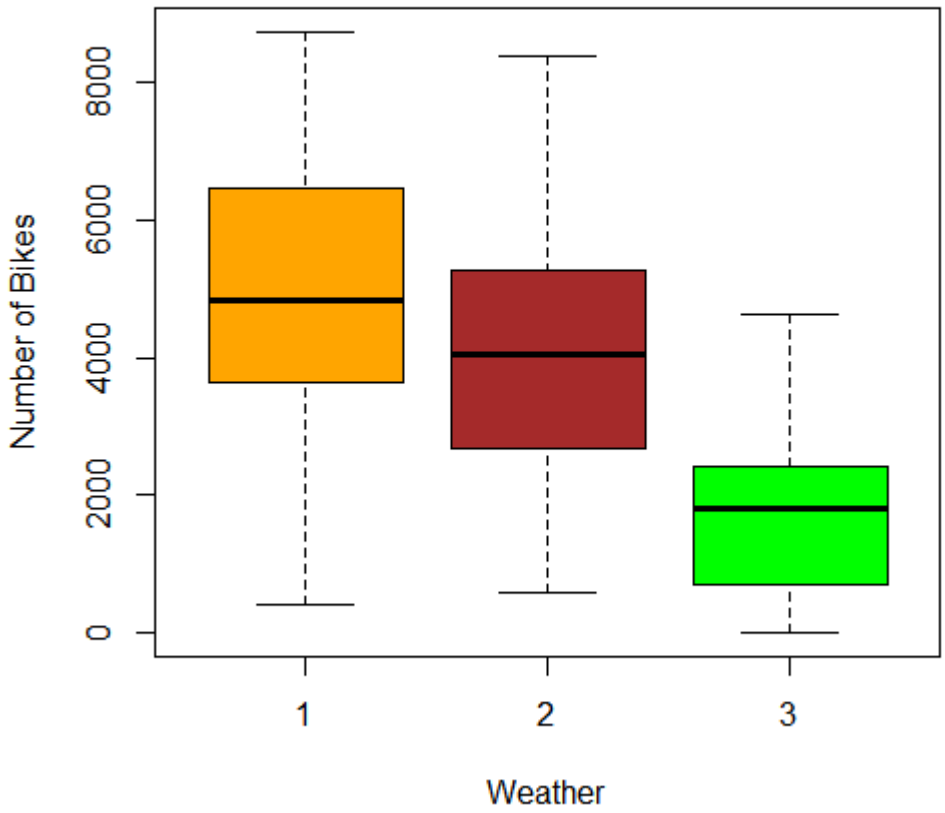


Figure 4 Bikes over Weather And correlation is -0.2973912.

Bikes over Wind. The plot shows that the lower the wind the higher the number of bikes and the higher the wind the lower the number of bikes. There is an outlier in this data which we can see from the plot.

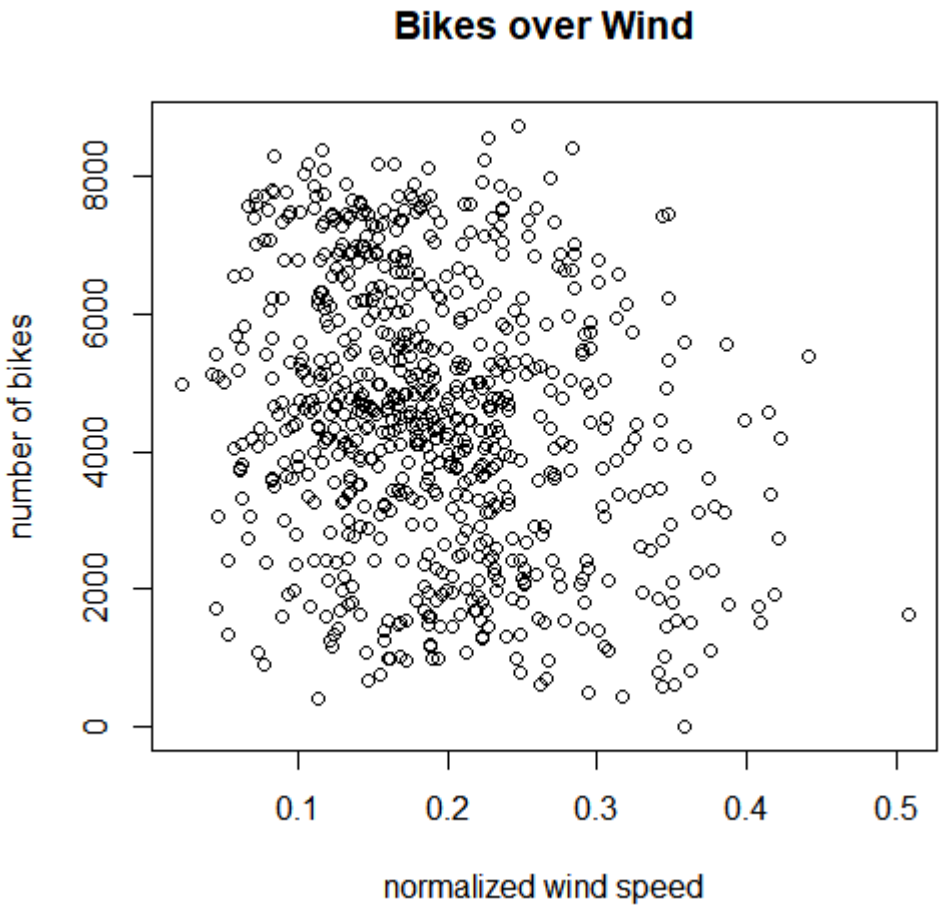


Figure 5 Bikes over Wind with correlation value -0.234545

2. From the number of bikes over days, the number of bikes rented have a very dramatic change which is it drops below 100 one day and the next day it rises over 4, 000. They are called outliers and can occurs due to measurement errors or because of momentary but drastic turbulence or may also occur due to natural fluctuations. They can adversely impact the accuracy of our results because most statistical parameters such as mean, correlation are highly sensitive to them that is why we have to remove them from our data set.

```
## Section 2: Task 2 clean the date and remove the outliers ##
count_ts = ts(data[, c('bikes')]) # to create time series object
head(count_ts)
data$clean_bikes = tsclean(count_ts) # Remove the outliers and add a new cloumn for clean_bikes
head(data)
```

R code for removing outliers

3. I have the two summaries below one with out applying the factor function and the another one after applying factor function and two figures below, one figure shows a boxplot for R^2 for 50 splits of the data with out converting any variables to factors.
In the 2nd figure we have box plot showing the result of R^2 when we factor season, weathersit and weekday. I have factored season, weathersit , month weekday because they are categorical data.

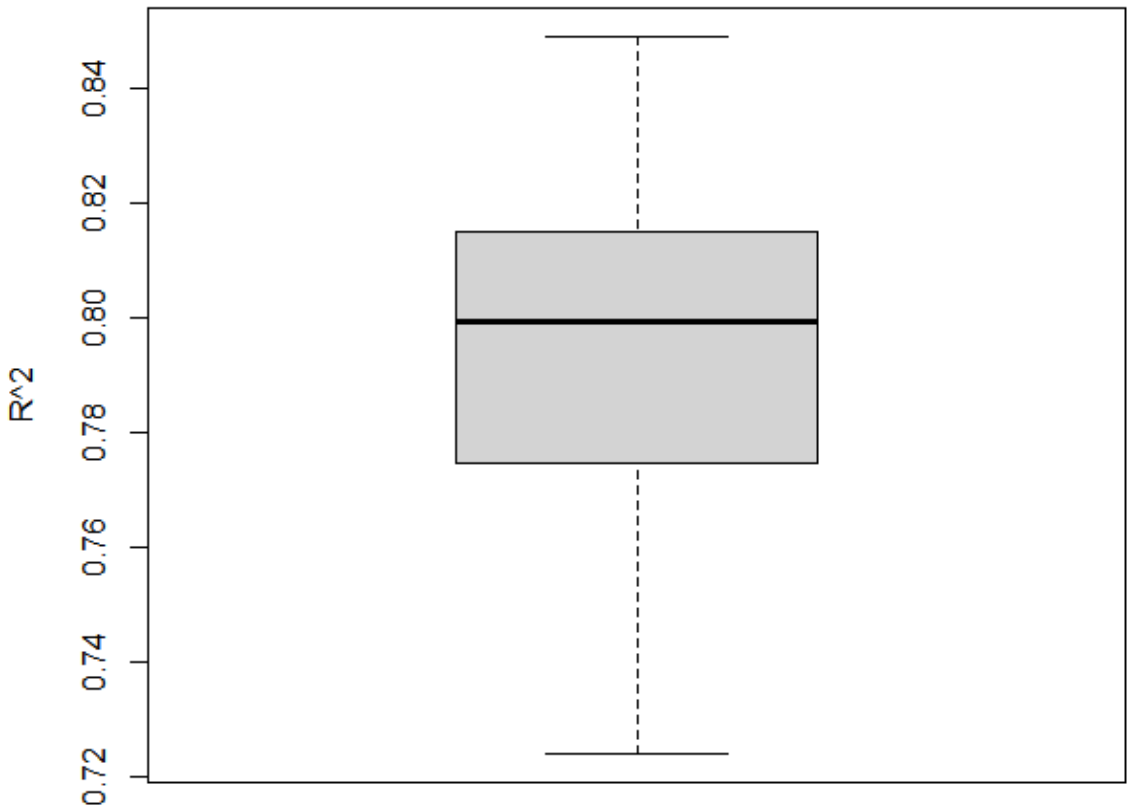
By applying the factor function to the categorical data, we can dig further into which category is more valuable in our data. For example for the season without applying the factor function we only see one variable season and their corresponding values which does not tell us anything about which season is better as it deals with all the four seasons as one variable.
And after applying factor we can see in the summary we can see every season has its only values and we can see which one season is more significant.

```
Call:
lm(formula = clean_bikes ~ ., data = trainData)

Residuals:
    Min      1Q  Median      3Q     Max
-3209.0  -458.5    37.5   516.1  2431.6

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1244.34     269.94   4.610 5.13e-06 ***
season        613.34      61.73   9.935  < 2e-16 ***
yr           2144.95      74.39  28.832  < 2e-16 ***
mnth          -64.52      18.97  -3.401 0.000725 ***
holiday      -362.59     251.05  -1.444 0.149280
weekday        57.99      18.62   3.115 0.001947 **
workingday     95.14      82.04   1.160 0.246728
weathersit   -558.38      90.63  -6.161 1.49e-09 ***
temp         2820.84    1426.89   1.977 0.048602 *
atemp        2416.91    1613.55   1.498 0.134796
hum          -891.40     352.39  -2.530 0.011728 *
windspeed   -1862.02     538.43  -3.458 0.000590 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 833.7 on 499 degrees of freedom
Multiple R-squared:  0.8132,    Adjusted R-squared:  0.8091
F-statistic: 197.5 on 11 and 499 DF,  p-value: < 2.2e-16
```

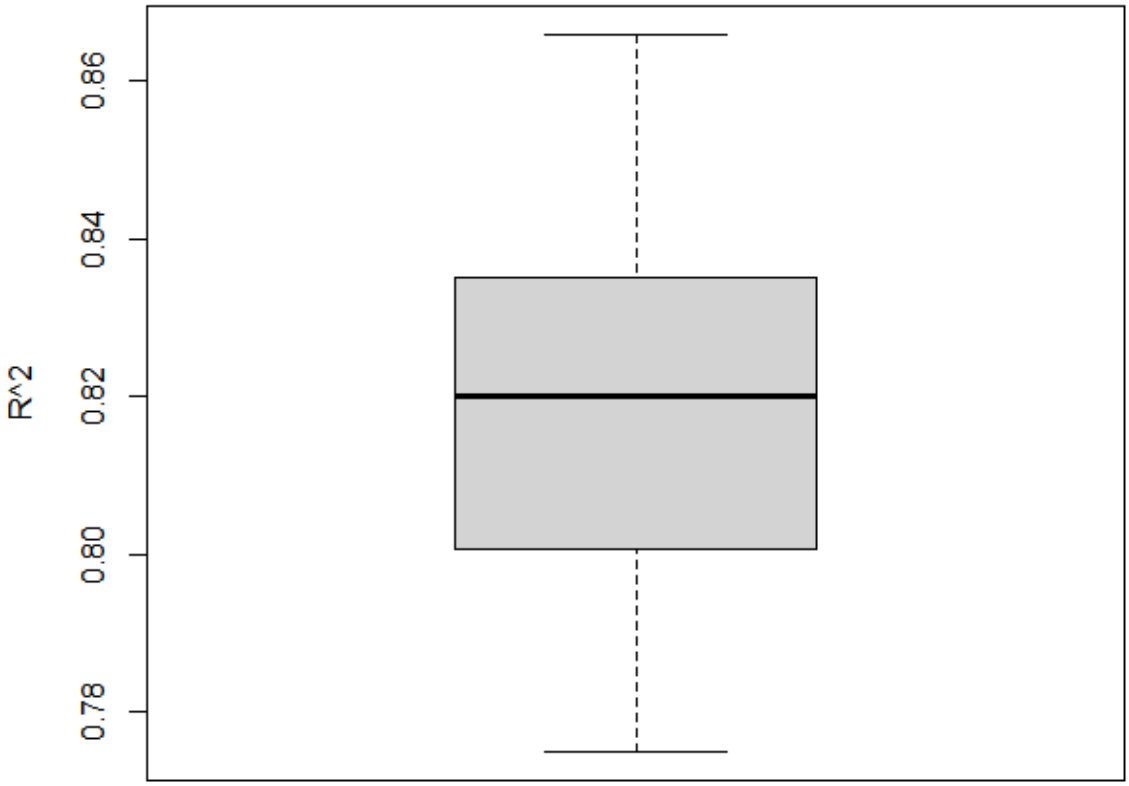
Summary of lm without factoring

```
Call:
lm(formula = clean_bikes ~ ., data = trainData)

Residuals:
    Min      1Q  Median      3Q     Max
-3186.5  -398.3    66.8   482.6  2919.1

Coefficients: (1 not defined because of singularities)
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1188.64     292.98   4.057 5.78e-05 ***
season2      1204.37     130.38   9.238  < 2e-16 ***
season3       876.69     190.27   4.608 5.20e-06 ***
season4      1752.73     174.55  10.041  < 2e-16 ***
yr           2131.32      71.97  29.615  < 2e-16 ***
mnth          -28.75      18.44  -1.560 0.119515
holiday      -500.58     207.78  -2.409 0.016355 *
weekday1      251.46     137.04   1.835 0.067123 .
weekday2      308.20     134.47   2.292 0.022333 *
weekday3      392.42     131.46   2.985 0.002976 **
weekday4      463.40     132.37   3.501 0.000506 ***
weekday5      431.43     132.07   3.267 0.001165 **
weekday6      546.39     130.77   4.178 3.48e-05 ***
workingday        NA         NA      NA       NA
weathersit2  -350.95      96.55  -3.635 0.000307 ***
weathersit3 -1457.08     241.48  -6.034 3.15e-09 ***
temp         3961.08    1429.34   2.771 0.005795 **
atemp        1404.85    1555.52   0.903 0.366894
hum         -1280.70     368.75  -3.473 0.000560 ***
windspeed   -2622.88     501.26  -5.233 2.48e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 796.3 on 492 degrees of freedom
Multiple R-squared:  0.8374,    Adjusted R-squared:  0.8314
F-statistic: 140.8 on 18 and 492 DF,  p-value: < 2.2e-16

>
```

Summary after factoring variables



with out factoring with mean value 0.7958971



After factoring variables with mean 0.8186199.

```
Call:
lm(formula = clean_bikes ~ ., data = trainData)

Residuals:
   Min     1Q Median     3Q    Max
-266.2  -59.3  -10.5   33.9 3525.4

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -7.403e+01  8.713e+01  -0.850   0.3959
season       4.173e+01  1.970e+01   2.119   0.0346 *
yr           1.500e+02  3.429e+01   4.374 1.49e-05 ***
mnth        -5.761e+00  5.733e+00  -1.005   0.3154
holiday     -9.047e+01  6.589e+01  -1.373   0.1704
weekday     -9.082e+00  5.578e+00  -1.628   0.1041
workingday  -3.593e+01  2.447e+01  -1.468   0.1426
weathersit   3.001e+01  2.713e+01   1.106   0.2692
temp        -1.051e+03  7.622e+02  -1.379   0.1684
atemp        1.498e+03  8.665e+02   1.728   0.0846 .
hum         -2.108e+01  1.055e+02  -0.200   0.8417
windspeed    1.338e+02  1.619e+02   0.826   0.4091
bikes        9.469e-01  1.281e-02  73.938  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 246.6 on 498 degrees of freedom
Multiple R-squared:  0.9838,    Adjusted R-squared:  0.9835
F-statistic:  2527 on 12 and 498 DF,  p-value: < 2.2e-16

>
```

From the summary above we can see that that the parameter estimate season = 4.173 in the multiple regression suggest that, after allowing for the effects of all other explanatory variables, a unit increase in season result an increase of 4.173 increase in the bike numbers. We can compare the multiple regression and the simple regression by using anova function to see that after allowing effect on season is there any effect of the rest of the explanatory variables on clean_bikes.

We can drop workingday as it does not have a big effect and the estimate for working day is -3.59 and mnth which has an estimate of -5.76 the lowest in all of the variables estimates so we can drop them and redo the regression. Both of these variables have a very high p-value as well which shows that they are less significant in the model.

4. From the graph the maximum R^2 value is around the depth of 8 which show that depth 8 is good option and more suitable to build the decision tree over all of the data. By comparing the R^2 values from linear model and decision tree, they summary of linear model show the value is 0.98 which is higher than the decision tree which is only 0.8 means that the linear model is the better option here.

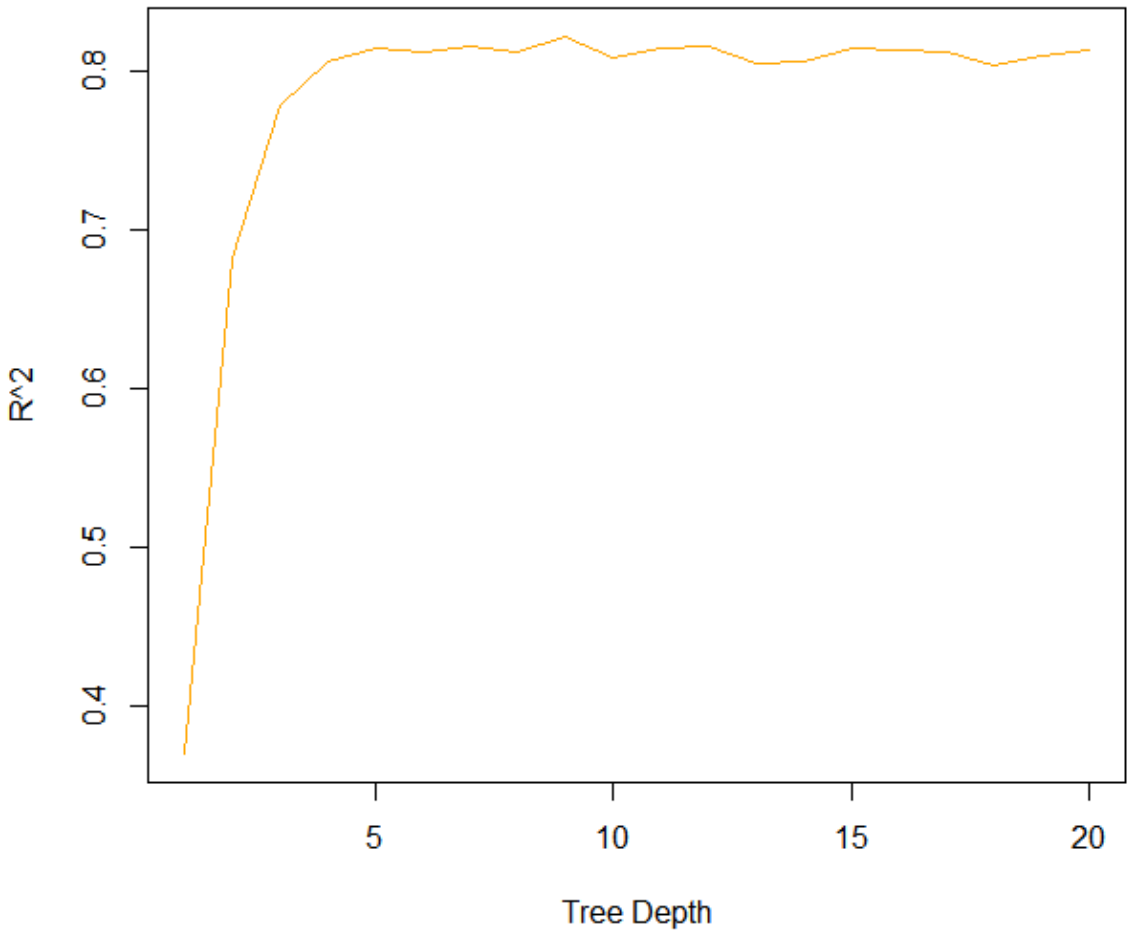


Figure showing how R^2 varies with deoth.

```r
library(rpart)
library(classifly)
library(rpart.plot)

data <- read.csv('day.csv',sep = ",",header=T) # load the data
count_ts = ts(data[, c('bikes')]) # to create time series object
head(count_ts)
data$clean_bikes = tsclean(count_ts) # Remove the outliers and add a new cloumn for clean_bikes
head(data)
dataClean <- data[,-c(1,2,14,15,16)] # remove instant, dteday, casual, registered, bikes from the data.
head(dataClean)

r2 <- matrix(ncol = 20, nrow = 50)
for (x in 1:20) {
  for (v in 1:50) {
    dataSplit <- sample.int(n=nrow(dataClean), size = floor(.70*nrow(dataClean)), replace = FALSE) # split the data
    trainData <- dataClean[dataSplit,] # get the training data
    testData <- dataClean[-dataSplit,] # get the testing data
    rp <- rpart(clean_bikes ~ ., data = trainData,control=rpart.control(maxdepth=x,minsplit=2,
                                    minbucket=2,cp=0))
    yhat <- predict(rp, newdata=testData,type="vector")
    y <- testData$clean_bikes # test data
    yhat <- predict(rp, testData) # prediction using all test data
    rsqr <- 1 - sum((y - yhat)^2) / sum((y - mean(y))^2) # R-squared measure
    r2[v,x] <- rsqr

  }

}
print(r2)
r3<- sample(0,20,replace= TRUE)
for (r in 1:20) {
 r3[r] <-  mean(r2[,r])

}

print(r3)
plot(1:20,r3, type = 'l',xlab = "Tree Depth", ylab = "R^2", col = "orange")

```

Figure 6 R code for decision tree

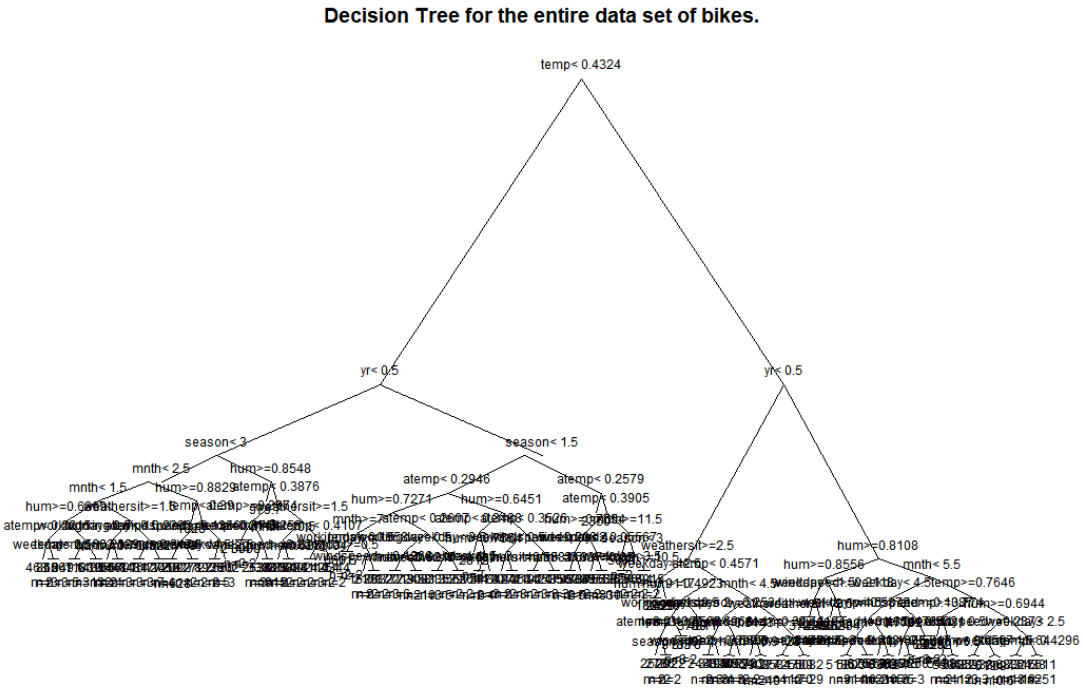
5. From the summary below of the linear models the most important and useful variables are the one with asterisk sign as they have a p-value very close to 0. For example season, yr, weather sit, weekday , windspeed are very useful variables than temp and working day.
   I could see both the model agrees mostly in terms of the relationship of the variables because we can see that it that depends on the weather, temperature and month etc have a very good p-value which means its close to zero and they are important and useful explanatory variables in this data set.
   From earlier on we saw in the first task that season is a very important variable in the data set and after looking at the summary of linear model and decision tree both of them agrees on that as well. Which is just one of example that that decision tree and linear model agree in terms of relationship with task 1.

```
(Y)
Variable importance
     atemp       temp         yr     season       mnth        hum  windspeed weathersit
        23         23         16         14         13          5          3          1
```

Summary for Decision Tree

```
Call:
lm(formula = clean_bikes ~ ., data = dataClean)

Residuals:
    Min      1Q  Median      3Q     Max
-3394.1  -461.7    59.8   524.6  2927.7

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1220.10     231.60   5.268 1.82e-07 ***
season        546.88      52.79  10.359  < 2e-16 ***
yr           2101.80      62.85  33.443  < 2e-16 ***
mnth          -44.47      16.47  -2.700 0.007087 **
holiday      -560.57     193.83  -2.892 0.003942 **
weekday        55.10      15.71   3.506 0.000483 ***
workingday    106.86      69.42   1.539 0.124185
weathersit   -537.47      75.55  -7.114 2.73e-12 ***
temp         1565.24    1353.31   1.157 0.247819
atemp        4014.35    1532.37   2.620 0.008986 **
hum          -899.13     302.73  -2.970 0.003077 **
windspeed   -2126.40     439.91  -4.834 1.64e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 841.1 on 719 degrees of freedom
Multiple R-squared:  0.8115,    Adjusted R-squared:  0.8086
F-statistic: 281.4 on 11 and 719 DF,  p-value: < 2.2e-16
```

Summary of linear model with entire data.



Decision Tree with the max depth of 8.

NOTE: It is really unclear to me what exactly you mean entire dataset, What I have assumed that use the clean data(remove the instant, dteday, casual, registered)t but do not split it into training and testing.

But to be just safe I have done the summary for linear model on entire data set and decision tree and we can see that the most important variables are registered and casual as compared to the one where we used the clean data there we had temp, year and season are the most important variables. Both these models agree with each other. Looking at the first plot in this file which is bikes over registered we can see that there is a linear relationship between bikes and registered variables. This shows that the results obtained from linear model and decision tree agree with what we had in the task 1.

```
Call:
lm(formula = bikes ~ ., data = data)

Residuals:
       Min         1Q     Median         3Q        Max
-3.383e-11 -2.450e-13  1.100e-14  2.410e-13  2.738e-11

Coefficients:
             Estimate Std. Error    t value Pr(>|t|)
(Intercept) -8.500e-11  1.237e-10 -6.870e-01 0.492176
dteday       5.488e-15  8.269e-15  6.640e-01 0.507105
season      -1.372e-12  1.317e-13 -1.041e+01  < 2e-16 ***
yr          -1.704e-12  3.052e-12 -5.580e-01 0.576767
mnth         5.244e-14  2.550e-13  2.060e-01 0.837162
holiday     -4.121e-13  4.528e-13 -9.100e-01 0.363044
weekday      1.747e-15  3.694e-14  4.700e-02 0.962286
workingday  -2.119e-12  2.739e-13 -7.737e+00 3.47e-14 ***
weathersit   9.522e-13  1.843e-13  5.168e+00 3.07e-07 ***
temp        -4.101e-12  3.152e-12 -1.301e+00 0.193617
atemp        1.234e-11  3.571e-12  3.454e+00 0.000584 ***
hum          1.409e-13  7.098e-13  1.990e-01 0.842689
windspeed    1.190e-13  1.043e-12  1.140e-01 0.909204
casual       1.000e+00  1.977e-16  5.058e+15  < 2e-16 ***
registered   1.000e+00  1.137e-16  8.799e+15  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.951e-12 on 716 degrees of freedom
Multiple R-squared:      1,     Adjusted R-squared:      1
F-statistic: 5.139e+31 on 14 and 716 DF,  p-value: < 2.2e-16

Warning message:
```

Figure 7 lm Summary for entire dataset just in case.

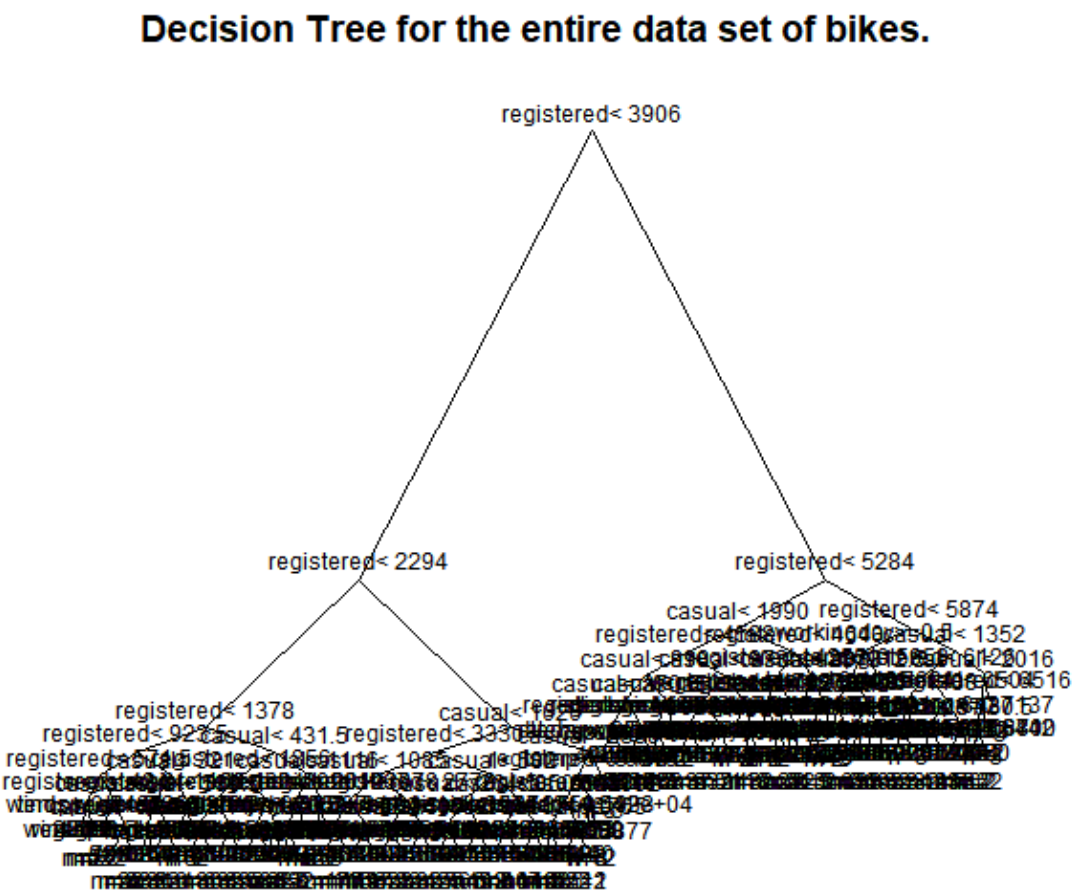


Figure 8 entire dataset with all the variables

6. Anomalous event is an event which deviates from usual data patterns with in a data set. These are the events which causes an unexpected change within data patterns or the and does not conform to the expected data pattern. I think if we look at the following data which is 19-10-212 and the event is Storm means that it happened in fall which is marked as 4 in the figure 2. Both have a relation ship because it has a very strong effect on the response in figure 1and is an outlier as shown in the following two figure 2. Which means that on that day it deviated from the usual data pattern. What I am trying to say here is that during the storm not a lot of people will be renting bikes and that is why the number of bikes in figure 2 is vey very low and the impact of this event is vey high in figure 1.
   On the other hand if we look at the Christmas day in figure 3 and the outlier in figure 2 which is in winter there is a strong deviation from usual pattern and a lot of bikes have been rented out. Which shows again the anomaly in the data. A lot of people are on holiday and out to enjoy and that is why we have a big jump in the number of bikes rented out as compared to a usual day during that season(winter).

Date	Event	Impact
29-10-2012	**Sandy**	5
30-10-2012	**Sandy**	5
19-10-2012	**Storm**	5

Figure 1

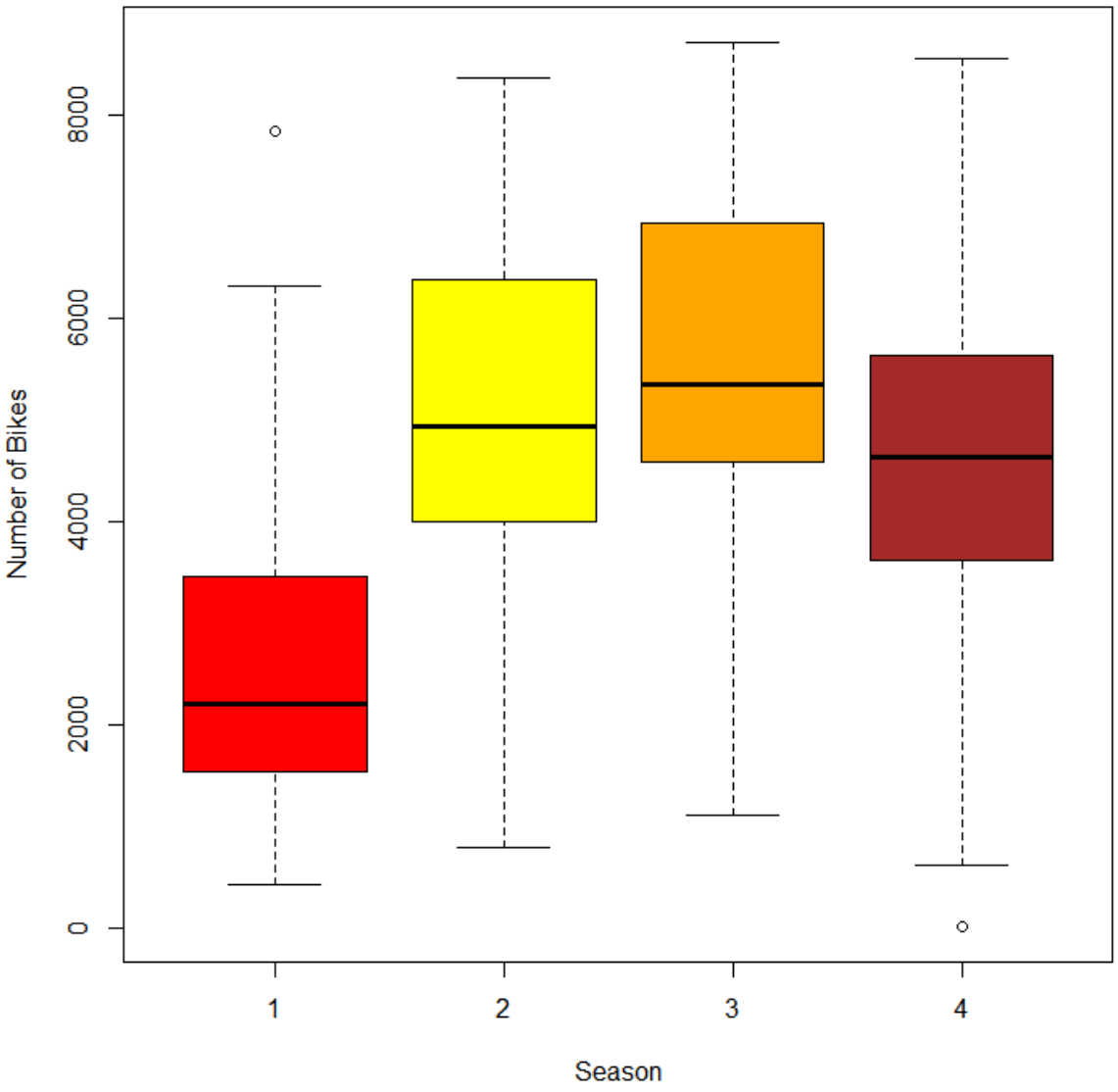


Figure 2

04-07-2012	**Washington DC fireworks**	5
23-11-2012	**Black Friday**	4
24-12-2012	**Christmas day**	4
08-10-2012	**Columbus memorial celebration**	4
27-05-2012	**Memorial day**	4

Figure 3

The impact rate which is given by the human domain specialist for Washington DC fireworks on the 04/07/2012 is 5 which is the highest means that the on this day the bikes pattern deviated from the usual pattern and the reason for that might be because a lot of people are out on that day and the number of bikes rented out on that day were very high.
Kind of the same for Black Friday which is from 23/11/2012, the impact rate given by the specialist is 4 which is still quite high and indicates strong deviation from the normal patterns, again the reason for that is clear it is a holiday, lots of people are out shopping, enjoying maybe and going places which could be the reason for more bikes rented out on that day than any other normal day.