

Seoul Bike Sharing Demand Analysis and Prediction

Ahmad Sadeed (asadeed2), Deepa Nemmili Veeravalli (deepan2), Rui Zou (ruizou4)

2022-07-22

Description of the data file

This data file contains count of public bikes rented at each hour in Seoul Bike Sharing System with the corresponding weather data and holidays information. It has 14 variables and 8760 observations. We are interested in using Rented.Bike.Count (a numeric variable) as our response variable and explore how other factors (3 categorical variables and several continuous numeric variables) affect the count of bikes rented at each hour. Among the other 13 variables which we plan to use as potential predictors, we know from intuition that some may have more importance than others, like temperature, humidity, wind speed, visibility, seasons, and holiday, etc.

Background information on the data set

The original data comes from <http://data.seoul.go.kr>. The holiday information comes from [SOUTH KOREA PUBLIC HOLIDAYS](#). A clean version can be found at [UCI Machine Learning Repository](#).

Attribute Information:

- Date : month/day/year
- Rented Bike count - Count of bikes rented at each hour
- Hour - Hour of the day
- Temperature - Temperature in Celsius
- Humidity - %
- Windspeed - m/s
- Visibility - 10m
- Dew point temperature - Celsius
- Solar radiation - MJ/m²
- Rainfall - mm
- Snowfall - cm
- Seasons - Winter, Spring, Summer, Autumn
- Holiday - Holiday, No holiday
- Functional Day - Functional or Non-functional days of rental bike system

Our Interest

This data set is interesting to us both personally and business-wise. Recently we have seen a rise in the delivery, accessibility, and usage of regular and electric rental bikes. There are clear environmental, health, and economical benefits associated with the usage of bikes as a mode of transportation. We would like to find out what factors lead to an increase in number of bikes rented and what factors have inverse effect on using

rental bikes. Learning about such factors can help a bike rental business manage its inventory and supply without any hindrance. It can also help cities plan accordingly due to an increase of bikers, e.g. opening up more bike lanes during certain days or seasons. Environmentally, we will have a better understanding of the feasibility of turning a city into a “bike city” or looking at alternative options if a city is not friendly to bikers due to harsh weather conditions.

Data in R

The data file can be successfully loaded into R. We have printed out the structure and first few rows of the data file below.

The column names in the csv file contains measurement units (like Wind speed (m/s), Solar Radiation (MJ/m²) and characters such as ° and %. We load the data using cleaned up column names.

```
columns = c("Date", "Rented.Bike.Count", "Hour", "Temperature", "Humidity",
           "Wind.Speed", "Visibility", "Dew.point.temperature",
           "Solar.Radiation", "Rainfall", "Snowfall", "Seasons", "Holiday",
           "Functioning.Day")
bike = read.csv("../data/SeoulBikeData.csv", col.names = columns)
str(bike)

## 'data.frame': 8760 obs. of 14 variables:
## $ Date : chr "01/12/2017" "01/12/2017" "01/12/2017" "01/12/2017" ...
## $ Rented.Bike.Count : int 254 204 173 107 78 100 181 460 930 490 ...
## $ Hour : int 0 1 2 3 4 5 6 7 8 9 ...
## $ Temperature : num -5.2 -5.5 -6 -6.2 -6 -6.4 -6.6 -7.4 -7.6 -6.5 ...
## $ Humidity : int 37 38 39 40 36 37 35 38 37 27 ...
## $ Wind.Speed : num 2.2 0.8 1 0.9 2.3 1.5 1.3 0.9 1.1 0.5 ...
## $ Visibility : int 2000 2000 2000 2000 2000 2000 2000 2000 2000 2000 1928 ...
## $ Dew.point.temperature: num -17.6 -17.6 -17.7 -17.6 -18.6 -18.7 -19.5 -19.3 -19.8 -22.4 ...
## $ Solar.Radiation : num 0 0 0 0 0 0 0 0.01 0.23 ...
## $ Rainfall : num 0 0 0 0 0 0 0 0 0 0 ...
## $ Snowfall : num 0 0 0 0 0 0 0 0 0 0 ...
## $ Seasons : chr "Winter" "Winter" "Winter" "Winter" ...
## $ Holiday : chr "No Holiday" "No Holiday" "No Holiday" "No Holiday" ...
## $ Functioning.Day : chr "Yes" "Yes" "Yes" "Yes" ...

head(bike)

## #> #>   Date Rented.Bike.Count Hour Temperature Humidity Wind.Speed Visibility
## #> 1 01/12/2017             254     0       -5.2      37       2.2      2000
## #> 2 01/12/2017             204     1       -5.5      38       0.8      2000
## #> 3 01/12/2017             173     2       -6.0      39       1.0      2000
## #> 4 01/12/2017             107     3       -6.2      40       0.9      2000
## #> 5 01/12/2017              78     4       -6.0      36       2.3      2000
## #> 6 01/12/2017             100     5       -6.4      37       1.5      2000
## #> #>   Dew.point.temperature Solar.Radiation Rainfall Snowfall Seasons   Holiday
## #> 1                 -17.6            0          0        0 Winter No Holiday
## #> 2                 -17.6            0          0        0 Winter No Holiday
## #> 3                 -17.7            0          0        0 Winter No Holiday
## #> 4                 -17.6            0          0        0 Winter No Holiday
## #> 5                 -18.6            0          0        0 Winter No Holiday
## #> 6                 -18.7            0          0        0 Winter No Holiday
```

```

##   Functioning.Day
## 1      Yes
## 2      Yes
## 3      Yes
## 4      Yes
## 5      Yes
## 6      Yes

library(lubridate)

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
## 
##     date, intersect, setdiff, union

bike$Date = as.Date(bike$Date, '%d/%m/%Y')

bike$year = as.numeric(format(bike$Date, '%Y'))
bike$month = as.numeric(format(bike$Date, '%m'))
bike$wday = wday(bike$Date) # Assuming Week Starts on Sunday. 1 and 7 should be weekends
bike$weekend = ifelse(bike$wday == 1 | bike$wday == 7, "Yes", "No")

table(bike$year)

##
## 2017 2018
## 744 8016

table(bike$month)

##
## 1 2 3 4 5 6 7 8 9 10 11 12
## 744 672 744 720 744 720 744 744 720 744 720 744

table(bike$wday)

##
## 1 2 3 4 5 6 7
## 1248 1248 1248 1248 1248 1272 1248

bike$Seasons = as.factor(bike$Seasons)
bike$Holiday = as.factor(bike$Holiday)
bike$Functioning.Day = as.factor(bike$Functioning.Day)
bike$year = as.factor(bike$year)
bike$month = as.factor(bike$month)
bike$wday = as.factor(bike$wday)
bike$weekend = as.factor(bike$weekend)
str(bike)

```

```

## 'data.frame': 8760 obs. of 18 variables:
## $ Date : Date, format: "2017-12-01" "2017-12-01" ...
## $ Rented.Bike.Count : int 254 204 173 107 78 100 181 460 930 490 ...
## $ Hour : int 0 1 2 3 4 5 6 7 8 9 ...
## $ Temperature : num -5.2 -5.5 -6 -6.2 -6 -6.4 -6.6 -7.4 -7.6 -6.5 ...
## $ Humidity : int 37 38 39 40 36 37 35 38 37 27 ...
## $ Wind.Speed : num 2.2 0.8 1 0.9 2.3 1.5 1.3 0.9 1.1 0.5 ...
## $ Visibility : int 2000 2000 2000 2000 2000 2000 2000 2000 2000 1928 ...
## $ Dew.point.temperature: num -17.6 -17.6 -17.7 -17.6 -18.6 -18.7 -19.5 -19.3 -19.8 -22.4 ...
## $ Solar.Radiation : num 0 0 0 0 0 0 0 0 0.01 0.23 ...
## $ Rainfall : num 0 0 0 0 0 0 0 0 0 0 ...
## $ Snowfall : num 0 0 0 0 0 0 0 0 0 0 ...
## $ Seasons : Factor w/ 4 levels "Autumn","Spring",...: 4 4 4 4 4 4 4 4 4 4 ...
## $ Holiday : Factor w/ 2 levels "Holiday","No Holiday": 2 2 2 2 2 2 2 2 2 2 ...
## $ Functioning.Day : Factor w/ 2 levels "No","Yes": 2 2 2 2 2 2 2 2 2 2 ...
## $ year : Factor w/ 2 levels "2017","2018": 1 1 1 1 1 1 1 1 1 1 ...
## $ month : Factor w/ 12 levels "1","2","3","4",...: 12 12 12 12 12 12 12 12 12 12 ...
## $ wday : Factor w/ 7 levels "1","2","3","4",...: 6 6 6 6 6 6 6 ...
## $ weekend : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...

```

```

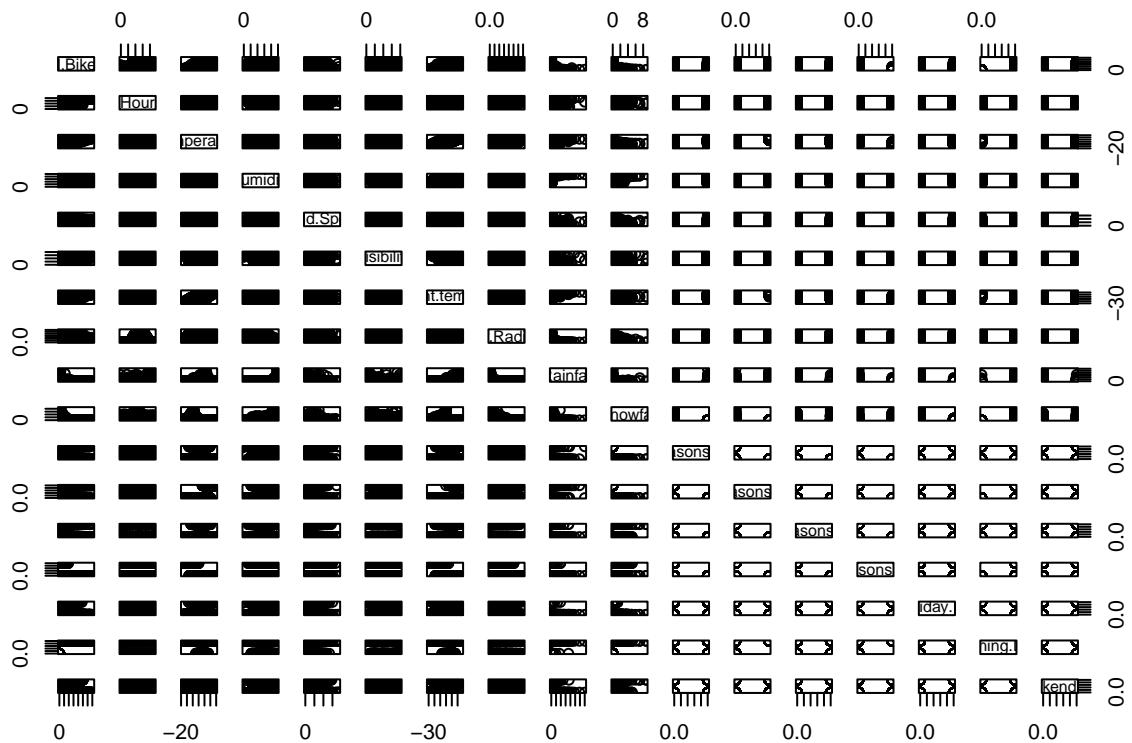
bike$Seasons.Sp = 1 * as.numeric(bike$Seasons == "Spring")
bike$Seasons.Su = 1 * as.numeric(bike$Seasons == "Summer")
bike$Seasons.Fa = 1 * as.numeric(bike$Seasons == "Autumn")
bike$Seasons.Wn = 1 * as.numeric(bike$Seasons == "Winter")

bike$Holiday.Yes = 1 * as.numeric(bike$Holiday == "Holiday")
bike$Functioning.Day.Yes = 1 * as.numeric(bike$Functioning.Day == "Yes")
bike$weekend.Yes = 1 * as.numeric(bike$weekend == "Yes")

bike_num = subset(bike, select = -c(Date, Seasons, Holiday, Functioning.Day, year, month, wday, weekend))

pairs(bike_num)

```



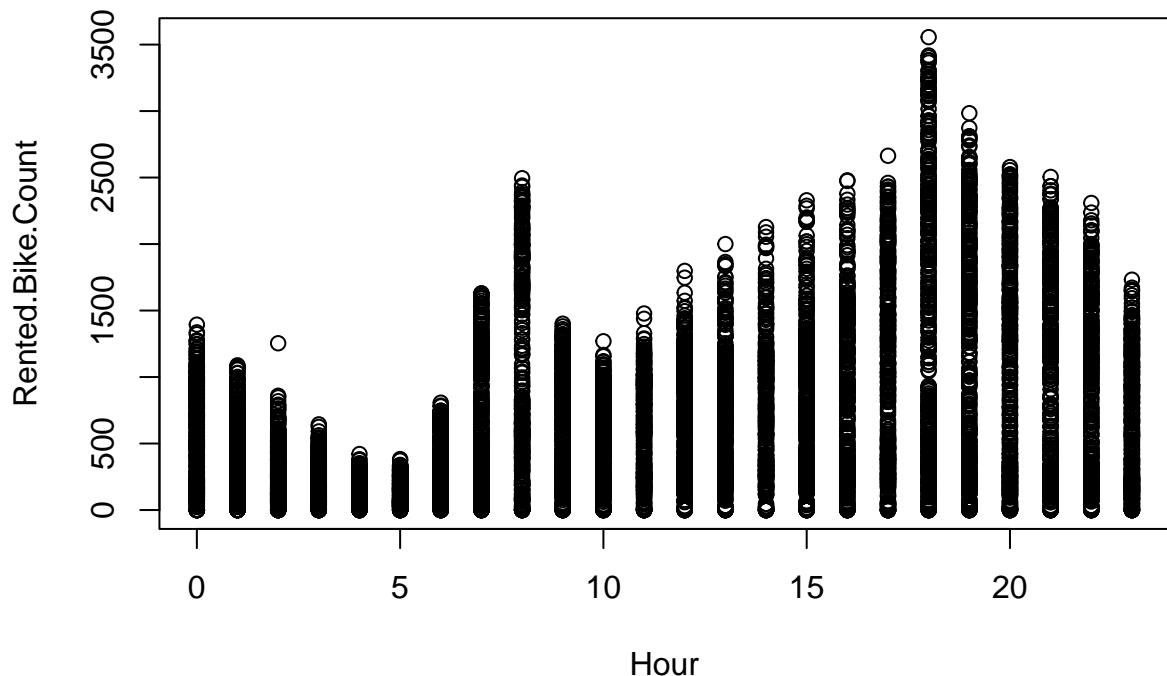
```
cor(bike_num)
```

	Rented.Bike.Count	Hour	Temperature	Humidity
## Rented.Bike.Count	1.00000	4.103e-01	0.538558	-0.19978
## Hour	0.41026	1.000e+00	0.124114	-0.24164
## Temperature	0.53856	1.241e-01	1.000000	0.15937
## Humidity	-0.19978	-2.416e-01	0.159371	1.00000
## Wind.Speed	0.12111	2.852e-01	-0.036252	-0.33668
## Visibility	0.19928	9.875e-02	0.034794	-0.54309
## Dew.point.temperature	0.37979	3.054e-03	0.912798	0.53689
## Solar.Radiation	0.26184	1.451e-01	0.353505	-0.46192
## Rainfall	-0.12307	8.715e-03	0.050282	0.23640
## Snowfall	-0.14180	-2.152e-02	-0.218405	0.10818
## Seasons.Sp	0.02289	0.000e+00	0.007960	0.01569
## Seasons.Su	0.29655	0.000e+00	0.665846	0.19259
## Seasons.Fa	0.10275	0.000e+00	0.059728	0.02837
## Seasons.Wn	-0.42493	0.000e+00	-0.738720	-0.23830
## Holiday.Yes	-0.07234	2.642e-22	-0.055931	-0.05028
## Functioning.Day.Yes	0.20394	5.439e-03	-0.050170	-0.02080
## weekend.Yes	-0.03647	0.000e+00	0.007214	-0.01695
	Wind.Speed	Visibility	Dew.point.temperature	
## Rented.Bike.Count	0.121108	0.199280	0.379788	
## Hour	0.285197	0.098753	0.003054	
## Temperature	-0.036252	0.034794	0.912798	
## Humidity	-0.336683	-0.543090	0.536894	

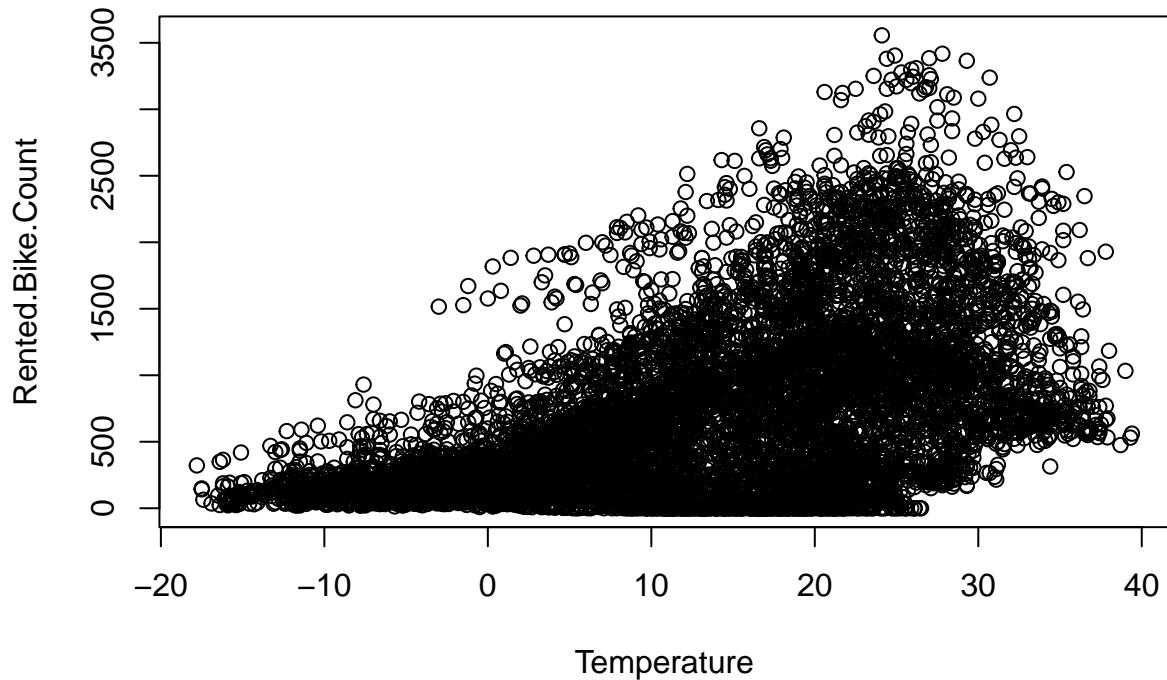
## Wind.Speed	1.000000	0.171507	-0.176486		
## Visibility	0.171507	1.000000	-0.176630		
## Dew.point.temperature	-0.176486	-0.176630	1.000000		
## Solar.Radiation	0.332274	0.149738	0.094381		
## Rainfall	-0.019674	-0.167629	0.125597		
## Snowfall	-0.003554	-0.121695	-0.150887		
## Seasons.Sp	0.083855	-0.187498	0.002056		
## Seasons.Su	-0.064698	0.061958	0.652378		
## Seasons.Fa	-0.128009	0.117413	0.062878		
## Seasons.Wn	0.109186	0.008616	-0.722366		
## Holiday.Yes	0.023017	0.031773	-0.066759		
## Functioning.Day.Yes	0.005037	-0.026000	-0.052837		
## weekend.Yes	-0.022227	-0.026762	-0.006990		
##	Solar.Radiation	Rainfall	Snowfall	Seasons.Sp	Seasons.Su
## Rented.Bike.Count	0.261837	-0.123074	-0.141804	0.022888	0.296549
## Hour	0.145131	0.008715	-0.021516	0.000000	0.000000
## Temperature	0.353505	0.050282	-0.218405	0.007960	0.665846
## Humidity	-0.461919	0.236397	0.108183	0.015694	0.192595
## Wind.Speed	0.332274	-0.019674	-0.003554	0.083855	-0.064698
## Visibility	0.149738	-0.167629	-0.121695	-0.187498	0.061958
## Dew.point.temperature	0.094381	0.125597	-0.150887	0.002056	0.652378
## Solar.Radiation	1.000000	-0.074290	-0.072301	0.079974	0.128402
## Rainfall	-0.074290	1.000000	0.008500	0.017595	0.053928
## Snowfall	-0.072301	0.008500	1.000000	-0.099785	-0.099785
## Seasons.Sp	0.079974	0.017595	-0.099785	1.000000	-0.336996
## Seasons.Su	0.128402	0.053928	-0.099785	-0.336996	1.000000
## Seasons.Fa	-0.031374	-0.013247	-0.024742	-0.334548	-0.334548
## Seasons.Wn	-0.178420	-0.058755	0.225875	-0.332099	-0.332099
## Holiday.Yes	-0.005077	-0.014269	-0.012591	-0.044791	-0.073932
## Functioning.Day.Yes	-0.007665	0.002055	0.032089	0.038413	0.108370
## weekend.Yes	0.012975	-0.014151	-0.006759	-0.002987	-0.002987
##	Seasons.Fa	Seasons.Wn	Holiday.Yes	Functioning.Day.Yes	
## Rented.Bike.Count	0.1027530	-0.424925	-7.234e-02	0.203943	
## Hour	0.0000000	0.000000	2.642e-22	0.005439	
## Temperature	0.0597283	-0.738720	-5.593e-02	-0.050170	
## Humidity	0.0283665	-0.238295	-5.028e-02	-0.020800	
## Wind.Speed	-0.1280093	0.109186	2.302e-02	0.005037	
## Visibility	0.1174133	0.008616	3.177e-02	-0.026000	
## Dew.point.temperature	0.0628783	-0.722366	-6.676e-02	-0.052837	
## Solar.Radiation	-0.0313743	-0.178420	-5.077e-03	-0.007665	
## Rainfall	-0.0132466	-0.058755	-1.427e-02	0.002055	
## Snowfall	-0.0247422	0.225875	-1.259e-02	0.032089	
## Seasons.Sp	-0.3345477	-0.332099	-4.479e-02	0.038413	
## Seasons.Su	-0.3345477	-0.332099	-7.393e-02	0.108370	
## Seasons.Fa	1.0000000	-0.329686	1.498e-02	-0.253718	
## Seasons.Wn	-0.3296859	1.000000	1.046e-01	0.106795	
## Holiday.Yes	0.0149846	0.104557	1.000e+00	-0.027624	
## Functioning.Day.Yes	-0.2537183	0.106795	-2.762e-02	1.000000	
## weekend.Yes	0.0009994	0.005016	-3.164e-02	0.040733	
##	weekend.Yes				
## Rented.Bike.Count	-0.0364674				
## Hour	0.0000000				
## Temperature	0.0072144				
## Humidity	-0.0169510				

```
## Wind.Speed           -0.0222268
## Visibility          -0.0267619
## Dew.point.temperature -0.0069896
## Solar.Radiation      0.0129755
## Rainfall             -0.0141509
## Snowfall              -0.0067586
## Seasons.Sp            -0.0029873
## Seasons.Su            -0.0029873
## Seasons.Fa             0.0009994
## Seasons.Wn             0.0050156
## Holiday.Yes           -0.0316417
## Functioning.Day.Yes    0.0407333
## weekend.Yes            1.0000000
```

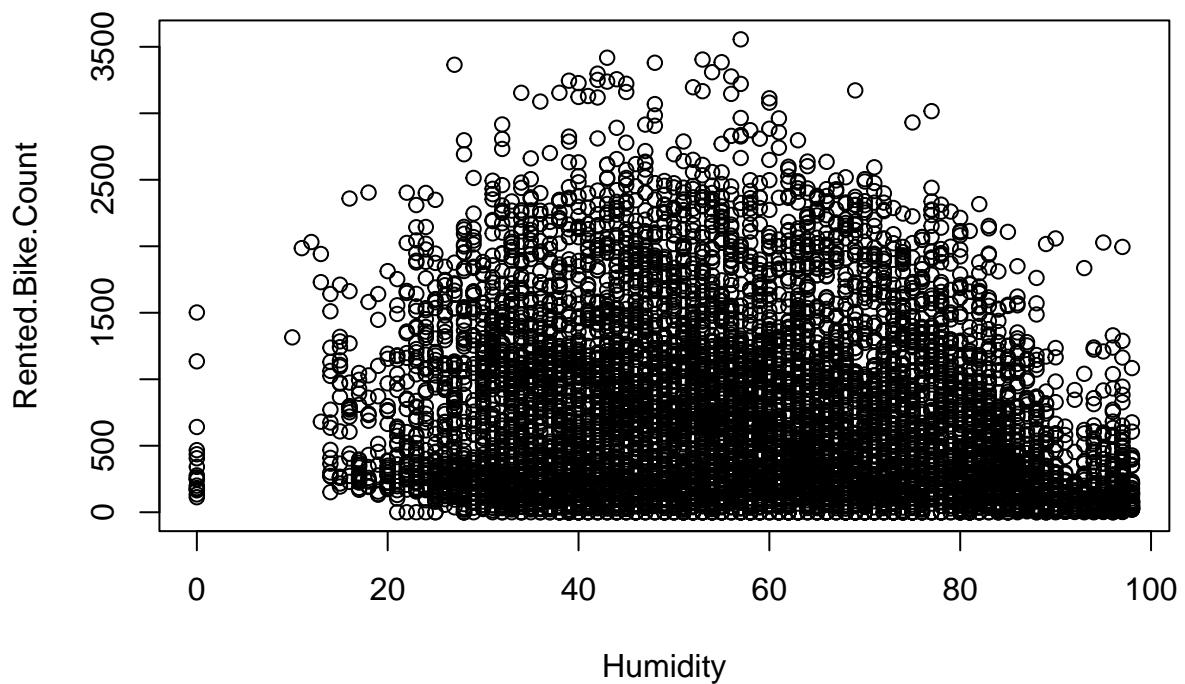
```
plot(Rented.Bike.Count ~ Hour, data = bike)
```



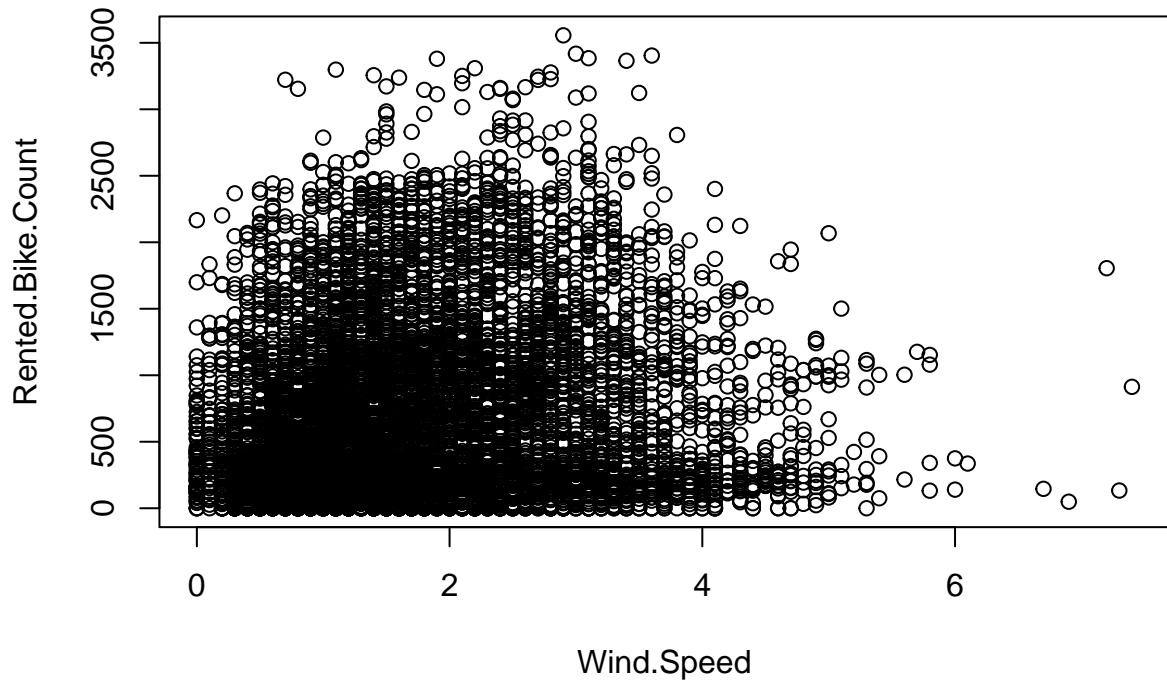
```
plot(Rented.Bike.Count ~ Temperature, data = bike)
```



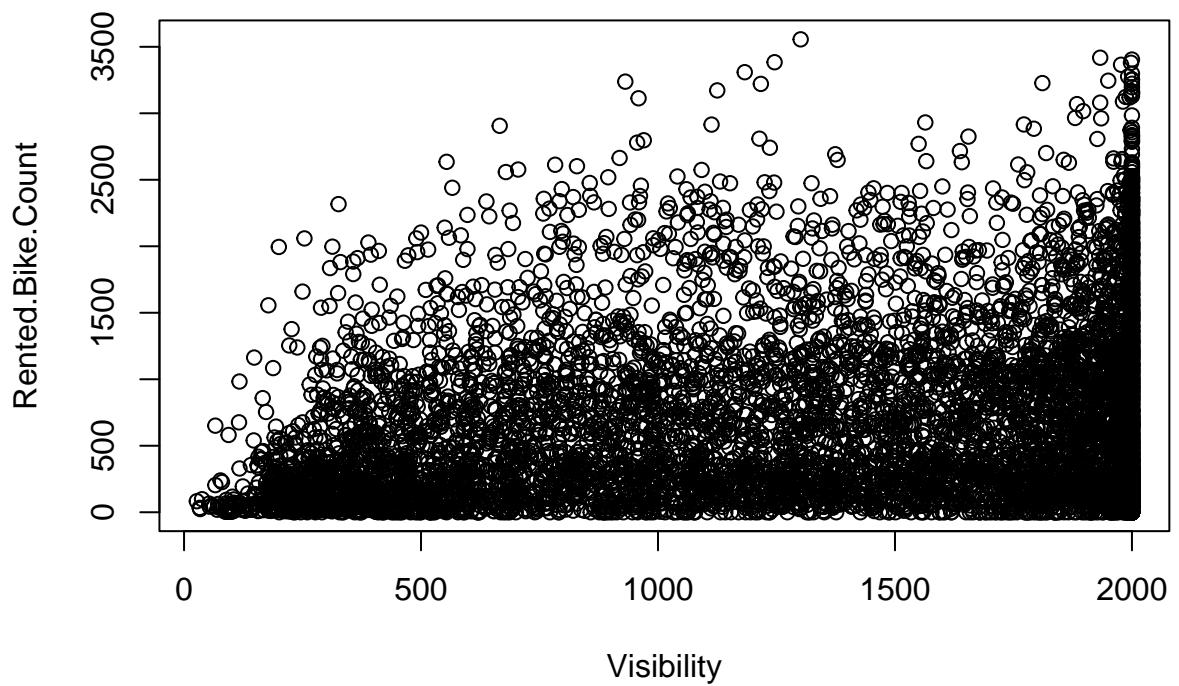
```
plot(Rented.Bike.Count ~ Humidity, data = bike)
```



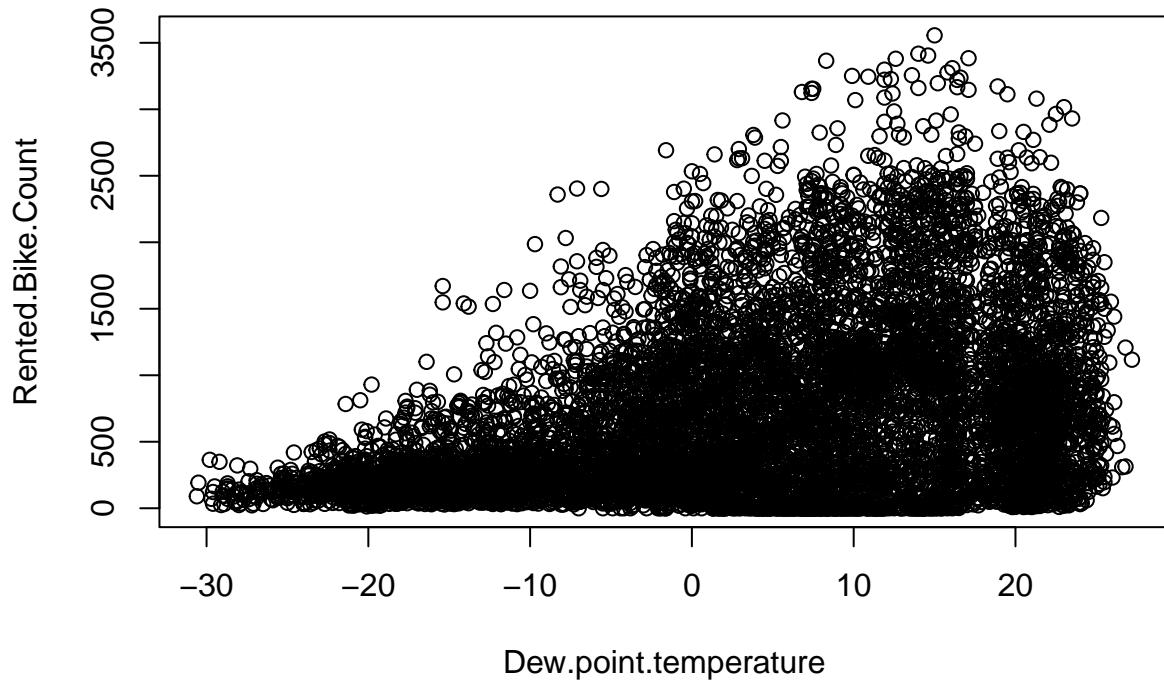
```
plot(Rented.Bike.Count ~ Wind.Speed, data = bike)
```



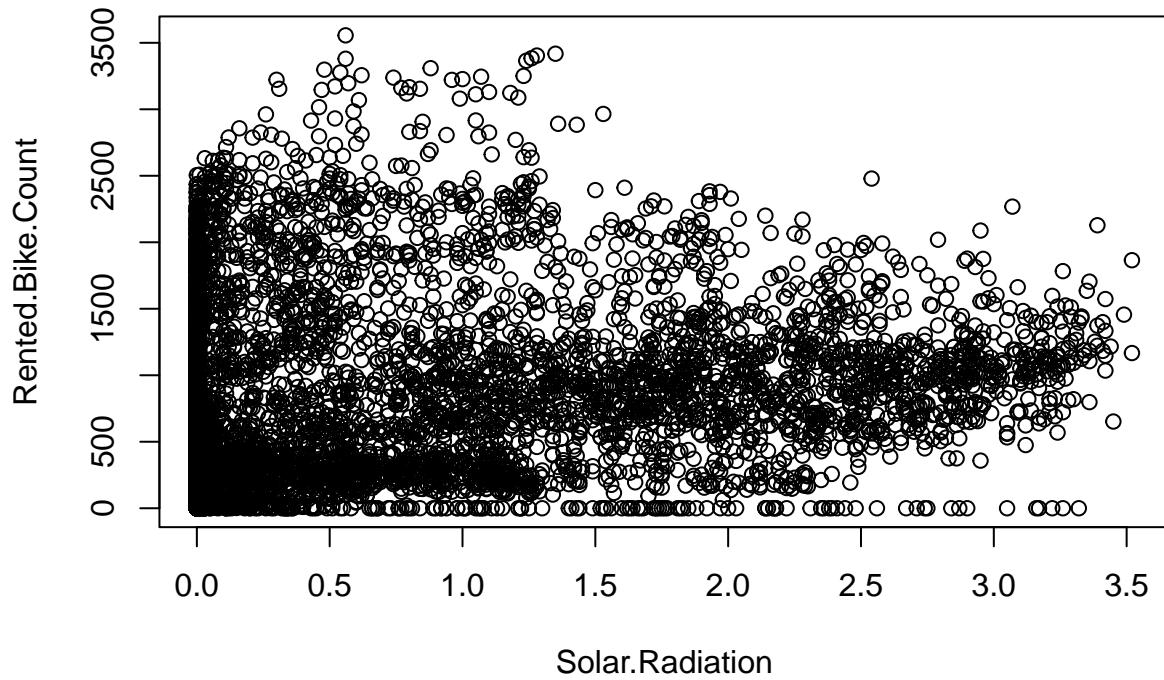
```
plot(Rented.Bike.Count ~ Visibility, data = bike)
```



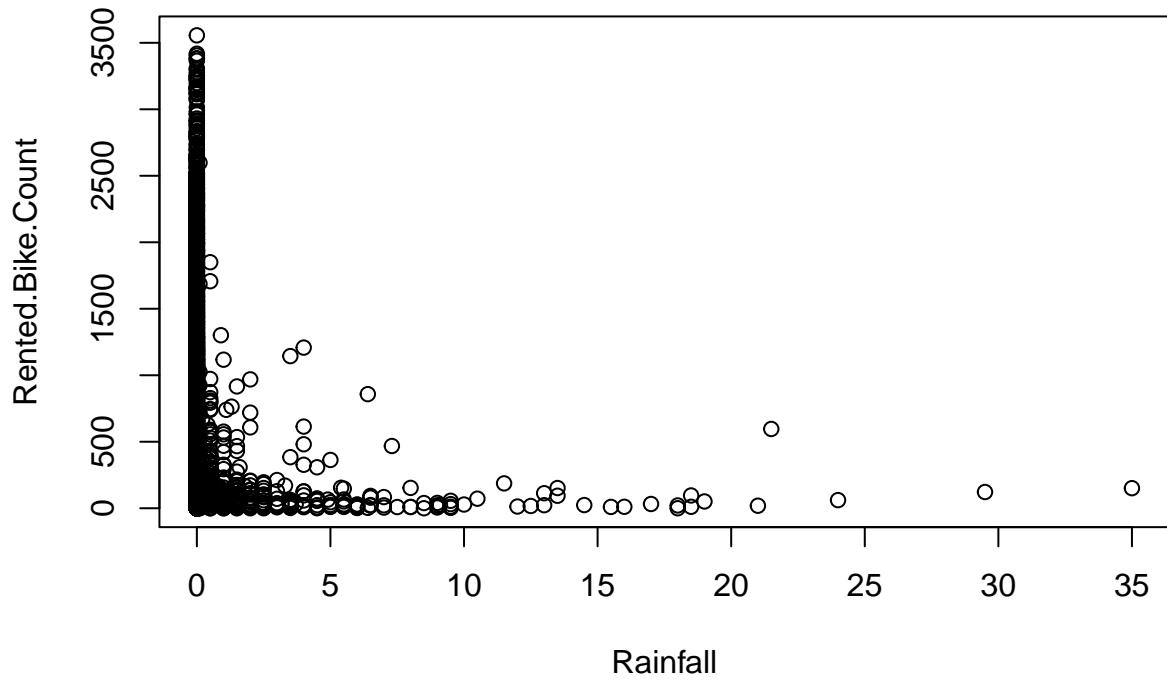
```
plot(Rented.Bike.Count ~ Dew.point.temperature, data = bike)
```



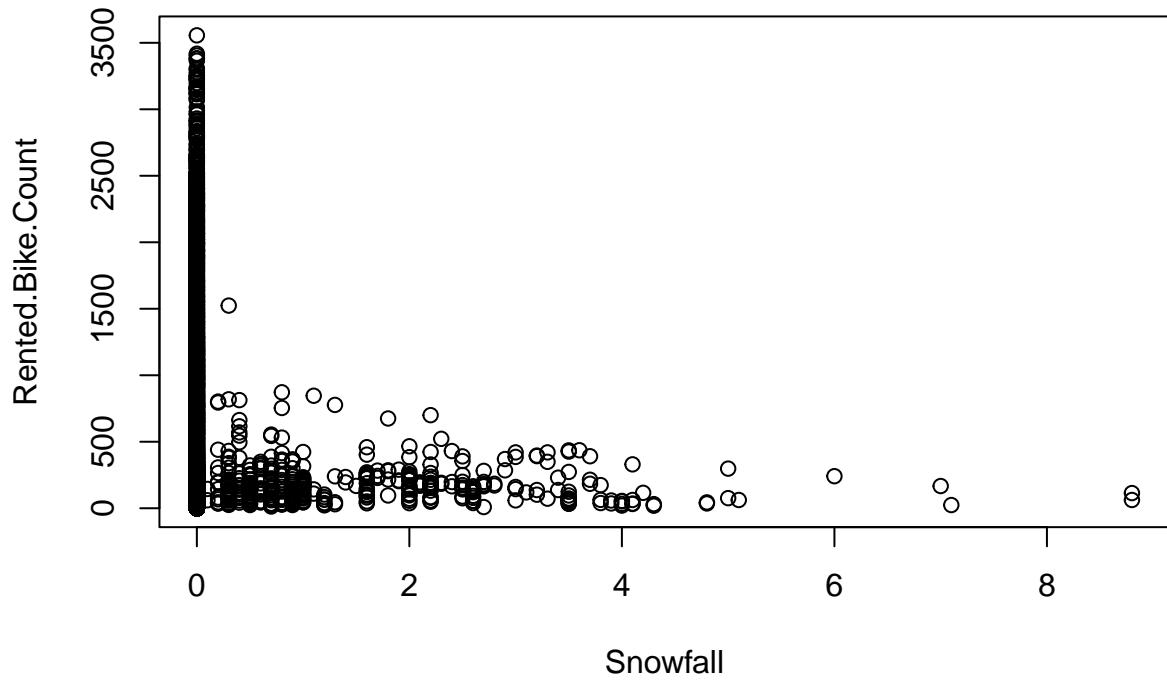
```
plot(Rented.Bike.Count ~ Solar.Radiation, data = bike)
```



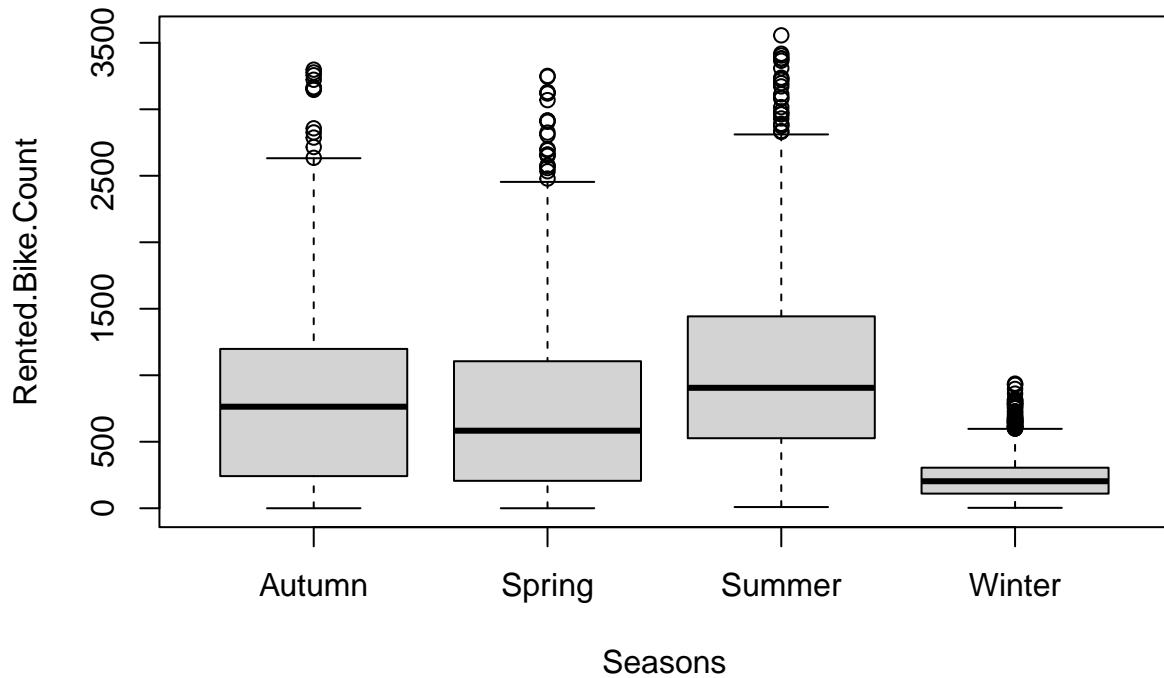
```
plot(Rented.Bike.Count ~ Rainfall, data = bike)
```



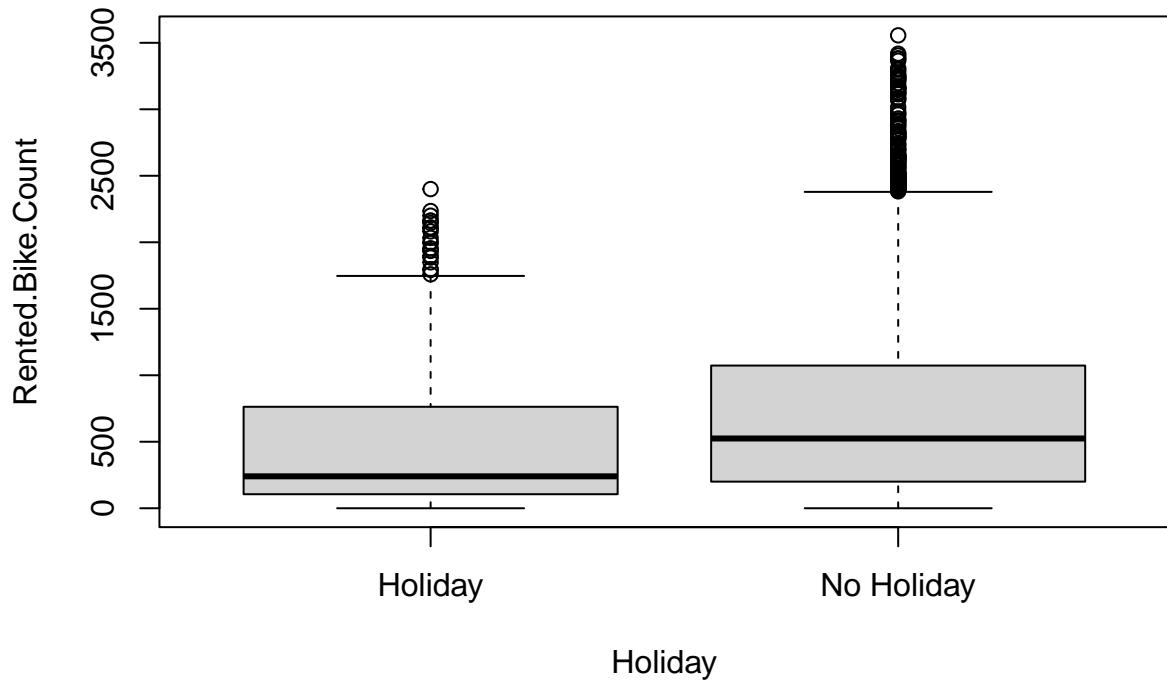
```
plot(Rented.Bike.Count ~ Snowfall, data = bike)
```



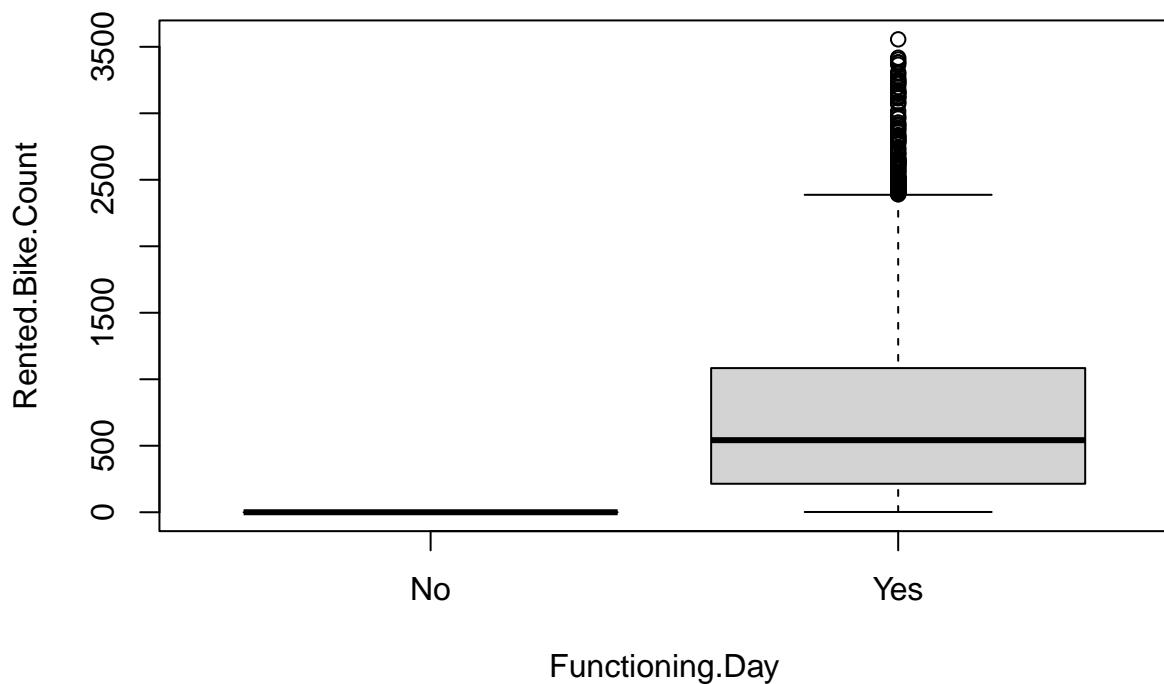
```
plot(Rented.Bike.Count ~ Seasons, data = bike)
```



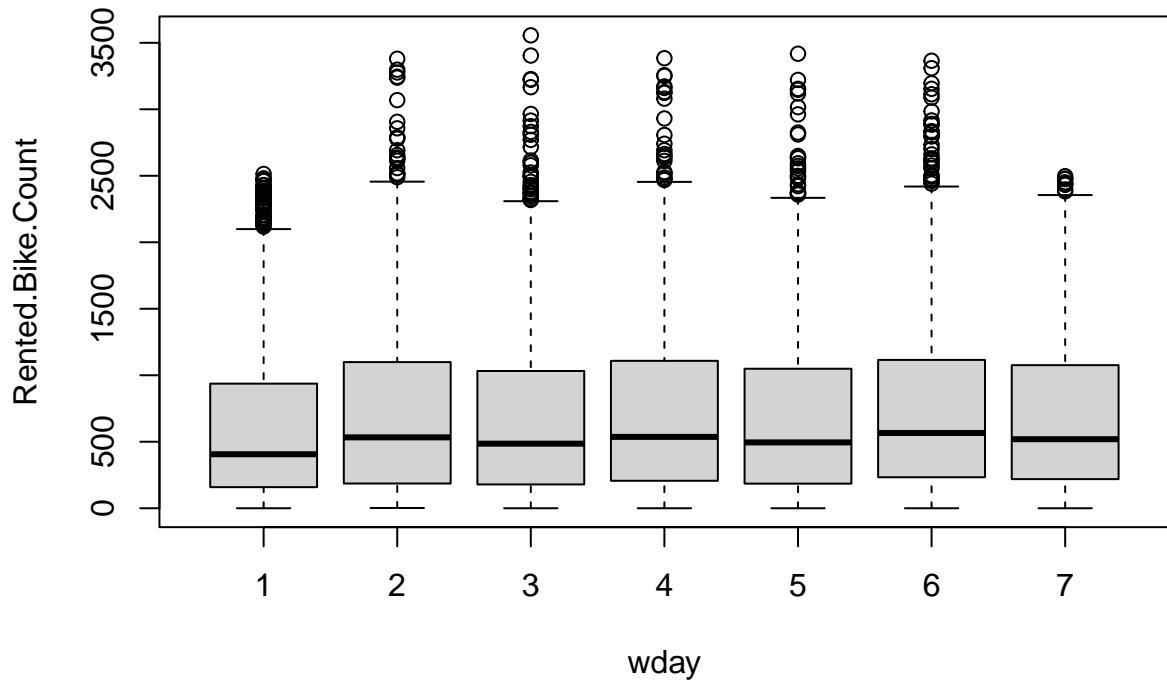
```
plot(Rented.Bike.Count ~ Holiday, data = bike)
```



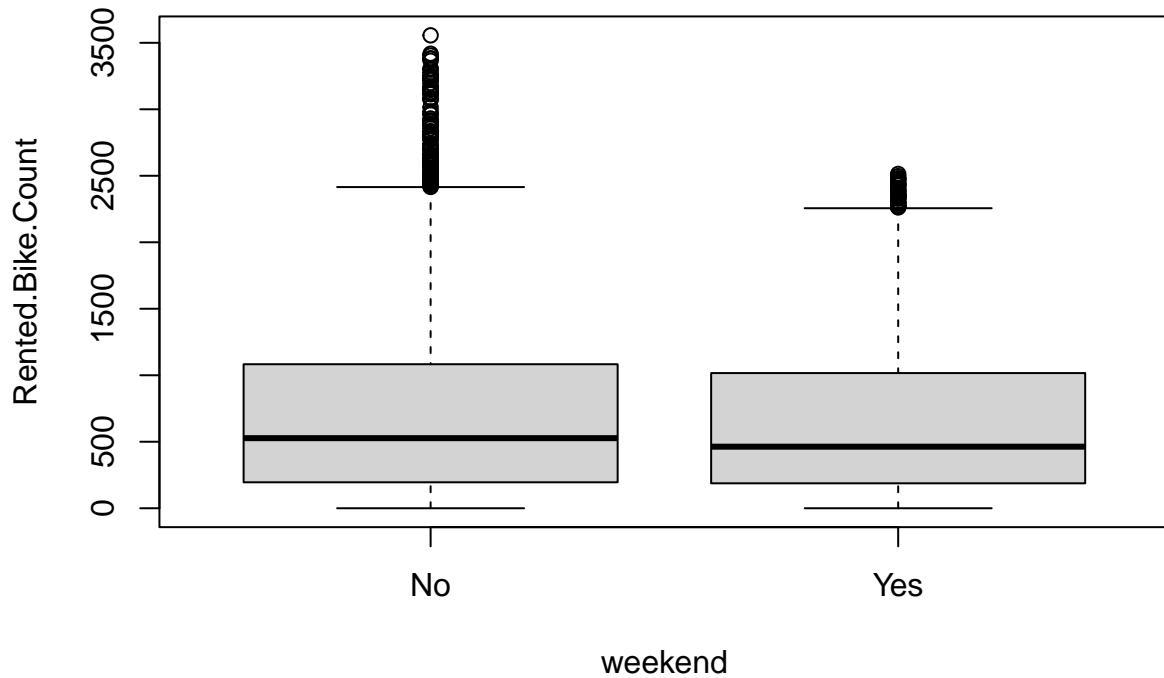
```
plot(Rented.Bike.Count ~ Functioning.Day, data = bike)
```



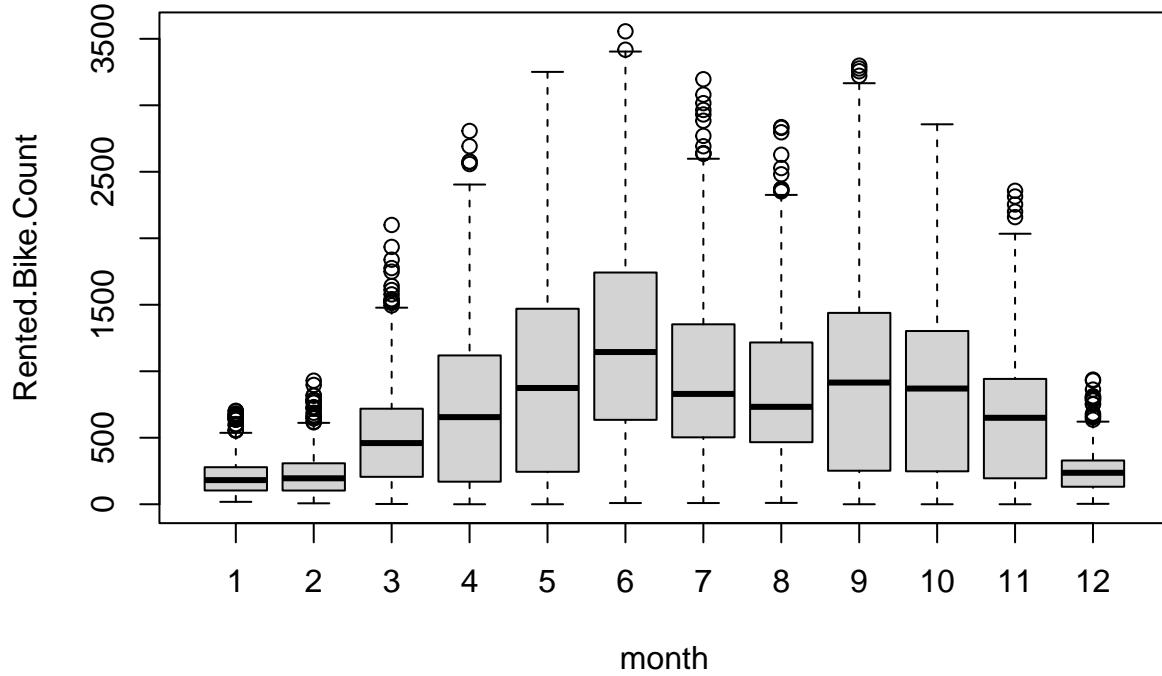
```
plot(Rented.Bike.Count ~ wday, data = bike)
```



```
plot(Rented.Bike.Count ~ weekend, data = bike)
```



```
plot(Rented.Bike.Count ~ month, data = bike)
```



```
model = lm(Rented.Bike.Count ~ . - Date, data = bike)
summary(model)
```

```
##
## Call:
## lm(formula = Rented.Bike.Count ~ . - Date, data = bike)
##
## Residuals:
##      Min       1Q   Median       3Q      Max 
## -1314.3  -264.9   -49.7  205.0  1973.0 
##
## Coefficients: (12 not defined because of singularities)
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept) -329.2616   99.4197  -3.31  0.00093 ***
## Hour         26.8242    0.7099   37.78 < 2e-16 ***
## Temperature 20.7410    3.5562    5.83  5.7e-09 ***
## Humidity     -9.9889   0.9913  -10.08 < 2e-16 ***
## Wind.Speed   21.2107   4.8782    4.35  1.4e-05 ***
## Visibility    0.0607   0.0110    5.53  3.3e-08 ***
## Dew.point.temperature 10.9079   3.7343    2.92  0.00350 ** 
## Solar.Radiation -90.2572   7.2829  -12.39 < 2e-16 ***
## Rainfall      -58.7891   4.0731  -14.43 < 2e-16 ***
## Snowfall       33.0677  10.7776    3.07  0.00216 ** 
## SeasonsSpring -0.5662   25.1718   -0.02  0.98206  
## SeasonsSummer -457.2517  35.3072  -12.95 < 2e-16 ***
```

```

## SeasonsWinter      -294.9597   25.4463  -11.59 < 2e-16 ***
## HolidayNo Holiday    138.8664   20.8072    6.67  2.6e-11 ***
## Functioning.DayYes  959.0631   25.5678   37.51 < 2e-16 ***
## year2018            -67.5897   21.7407  -3.11  0.00188 **
## month2              -35.4612   22.2289  -1.60  0.11069
## month3              -206.3619   24.7258  -8.35 < 2e-16 ***
## month4              -132.3871   22.8883  -5.78  7.5e-09 ***
## month5                  NA        NA        NA        NA
## month6              571.5151   23.8202  23.99 < 2e-16 ***
## month7              181.1382   21.5260   8.41 < 2e-16 ***
## month8                  NA        NA        NA        NA
## month9              -75.9409   29.4118  -2.58  0.00984 **
## month10             80.9523   23.6964   3.42  0.00064 ***
## month11                  NA        NA        NA        NA
## month12                  NA        NA        NA        NA
## wday2                80.3870   16.5340   4.86  1.2e-06 ***
## wday3                100.7829   16.6301   6.06  1.4e-09 ***
## wday4                125.2318   16.5884   7.55  4.8e-14 ***
## wday5                106.1122   16.5592   6.41  1.6e-10 ***
## wday6                138.7252   16.4494   8.43 < 2e-16 ***
## wday7                67.2491   16.5305   4.07  4.8e-05 ***
## weekendYes                 NA        NA        NA        NA
## Seasons.Sp                  NA        NA        NA        NA
## Seasons.Su                  NA        NA        NA        NA
## Seasons.Fa                  NA        NA        NA        NA
## Seasons.Wn                  NA        NA        NA        NA
## Holiday.Yes                 NA        NA        NA        NA
## Functioning.Day.Yes       NA        NA        NA        NA
## weekend.Yes                 NA        NA        NA        NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ',' 1
##
## Residual standard error: 412 on 8731 degrees of freedom
## Multiple R-squared:  0.594,  Adjusted R-squared:  0.593
## F-statistic:  457 on 28 and 8731 DF,  p-value: <2e-16

```