# ICS Project III: Regression Analysis

Bushra Tariq Kiyani (230204)

06-07-2023

```
library(ggplot2)
library(GGally)
```

```
## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2
```

```
library(ggpubr)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

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

```
library(reshape2)
```

```
# load the dataset
input_data <- read.csv("Bikedata.csv")
head(input_data)
```

```
##   log.Rented.Bike.Count Hour Temperature Humidity Wind.speed Visibility
## 1              6.463029    6        16.2        0        0.5       1588
## 2              6.573680   15         6.9       31        2.9       1982
## 3              4.143135    3        -1.3       80        0.7        392
## 4              7.634821   17        22.4       39        3.2       1943
## 5              6.543912   22         7.8       49        0.1        528
## 6              6.073045    2        12.8       59        0.9       1804
##   Solar.Radiation Rainfall Snowfall Seasons    Holiday
## 1            0.05        0        0  Spring No Holiday
## 2            1.21        0        0  Autumn No Holiday
## 3            0.00        0        0  Winter No Holiday
## 4            1.90        0        0  Spring No Holiday
## 5            0.00        0        0  Spring No Holiday
## 6            0.00        0        0  Autumn No Holiday
```

## Task 1. Descriptive Analysis

```
#Check  NA Values
colSums(is.na(input_data))
```

```
## log.Rented.Bike.Count                 Hour          Temperature
##                     0                    0                    0
##             Humidity           Wind.speed           Visibility
##                     0                    0                    0
##      Solar.Radiation             Rainfall             Snowfall
##                     0                    0                    0
##              Seasons              Holiday
##                     0                    0
```

```
# Convert Seasons and Holiday to factor variables
input_data$Seasons <- as.factor(input_data$Seasons)
input_data$Holiday <- as.factor(input_data$Holiday)
```

```
# Summary
summary(input_data)
```

```
##  log.Rented.Bike.Count      Hour        Temperature        Humidity
##  Min.   :0.6931        Min.   : 0.00   Min.   :-17.50   Min.   : 0.00
##  1st Qu.:5.3660        1st Qu.: 6.00   1st Qu.:  2.80   1st Qu.:42.00
##  Median :6.2971        Median :12.00   Median : 13.40   Median :57.00
##  Mean   :6.0909        Mean   :11.58   Mean   : 12.81   Mean   :57.73
##  3rd Qu.:7.0003        3rd Qu.:17.00   3rd Qu.: 22.80   3rd Qu.:74.00
##  Max.   :8.1212        Max.   :23.00   Max.   : 38.00   Max.   :98.00
##    Wind.speed      Visibility   Solar.Radiation     Rainfall
##  Min.   :0.000   Min.   :  63   Min.   :0.0000   Min.   : 0.0000
##  1st Qu.:0.900   1st Qu.: 940   1st Qu.:0.0000   1st Qu.: 0.0000
##  Median :1.500   Median :1703   Median :0.0200   Median : 0.0000
##  Mean   :1.734   Mean   :1441   Mean   :0.5753   Mean   : 0.1456
##  3rd Qu.:2.300   3rd Qu.:2000   3rd Qu.:0.9300   3rd Qu.: 0.0000
##  Max.   :7.300   Max.   :2000   Max.   :3.5200   Max.   :29.5000
##     Snowfall         Seasons          Holiday
##  Min.   :0.00000   Autumn:667   Holiday   : 148
##  1st Qu.:0.00000   Spring:721   No Holiday:2757
##  Median :0.00000   Summer:771
##  Mean   :0.08296   Winter:746
##  3rd Qu.:0.00000
##  Max.   :8.80000
```

```
#Summarise all continuous variables
standard_deviations <- rbind.data.frame(lapply(input_data  %>% dplyr::
                                                 select(-c("Seasons","Holiday")),sd))
summary_table  <- rbind.data.frame(lapply(input_data  %>% dplyr::
                                            select(-c("Seasons","Holiday")),summary))
summary_table  <-  rbind(summary_table, standard_deviations)
row.names(summary_table) <- c("Min", "Q1", "Median", "Mean", "Q3", "Max", "SD")
summary_table  <- t(round(summary_table,3))
summary_table
```

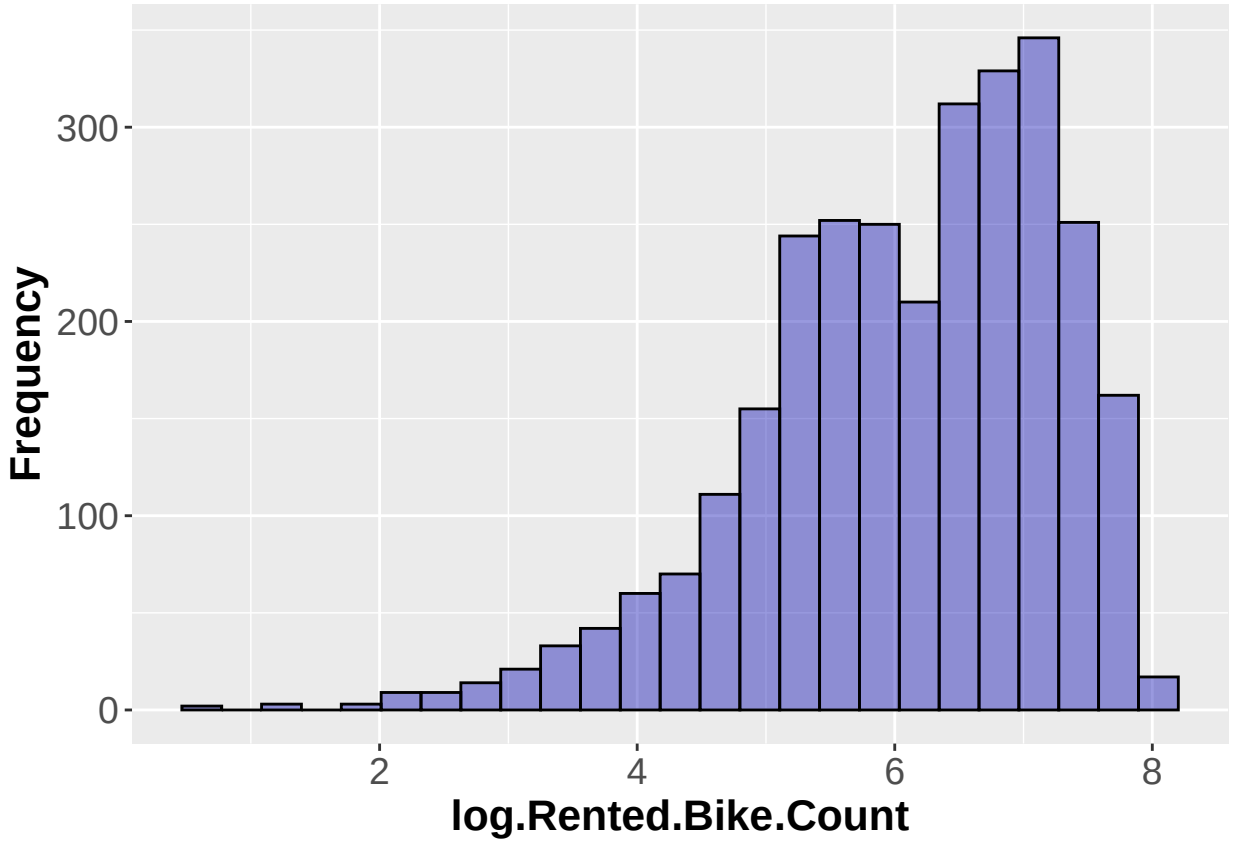
```
##                          Min      Q1   Median     Mean      Q3      Max
## log.Rented.Bike.Count   0.693   5.366    6.297    6.091    7.00    8.121
## Hour                    0.000   6.000   12.000   11.583   17.00   23.000
```

```
## Temperature            -17.500   2.800   13.400   12.807   22.80   38.000
## Humidity                 0.000  42.000   57.000   57.735   74.00   98.000
## Wind.speed               0.000   0.900    1.500    1.734    2.30    7.300
## Visibility              63.000 940.000 1703.000 1440.729 2000.00 2000.000
## Solar.Radiation          0.000   0.000    0.020    0.575    0.93    3.520
## Rainfall                 0.000   0.000    0.000    0.146    0.00   29.500
## Snowfall                 0.000   0.000    0.000    0.083    0.00    8.800
##                             SD
## log.Rented.Bike.Count    1.162
## Hour                     6.870
## Temperature             12.219
## Humidity                20.566
## Wind.speed               1.033
## Visibility             607.939
## Solar.Radiation          0.867
## Rainfall                 1.159
## Snowfall                 0.463
```

```
#Histogram to see the frequency distribution of the variable log.Rented.Bike.Count
FD = ggplot(input_data, aes(x=log.Rented.Bike.Count))+
  geom_histogram(color="black", fill="#0002AF", alpha = 0.4, bins = 25)+
  xlab("log.Rented.Bike.Count") + ylab("Frequency") +
  theme(axis.text=element_text(size=14), axis.title=element_text(size=16 , face = "bold"))
ggsave("FD.pdf", plot = FD)
```

```
## Saving 6.5 x 4.5 in image
```

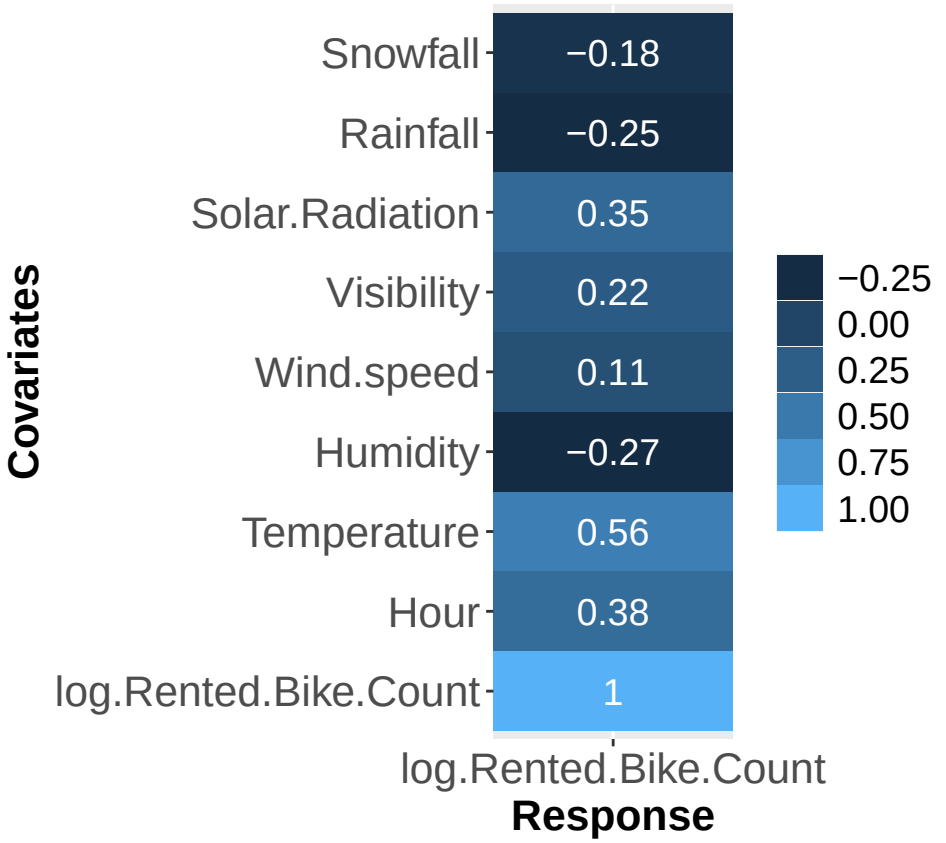
```
FD
```



```r
#Correlation Heatmap
corr_mat <- round(cor(input_data  %>% dplyr::select(-c("Seasons","Holiday"))),2)
melted_corr_mat <- melt(corr_mat) %>% dplyr::filter(Var1 == "log.Rented.Bike.Count")
out  <- ggplot(data = melted_corr_mat, aes(x=Var1, y=Var2,
                                   fill=value)) + geom_tile(aes(width = 1.5, height=1)) +
xlab("Response") + ylab("Covariates") + guides(fill=guide_legend(title="")) +
theme(legend.text = element_text(size=14),  axis.text=element_text(size=16),
      axis.title=element_text(size=16 , face = "bold"))+ coord_fixed(ratio = 2/4) +
geom_text(aes(label = value),
          color = "white", size = 5)
ggsave("out.pdf", plot = out)
```

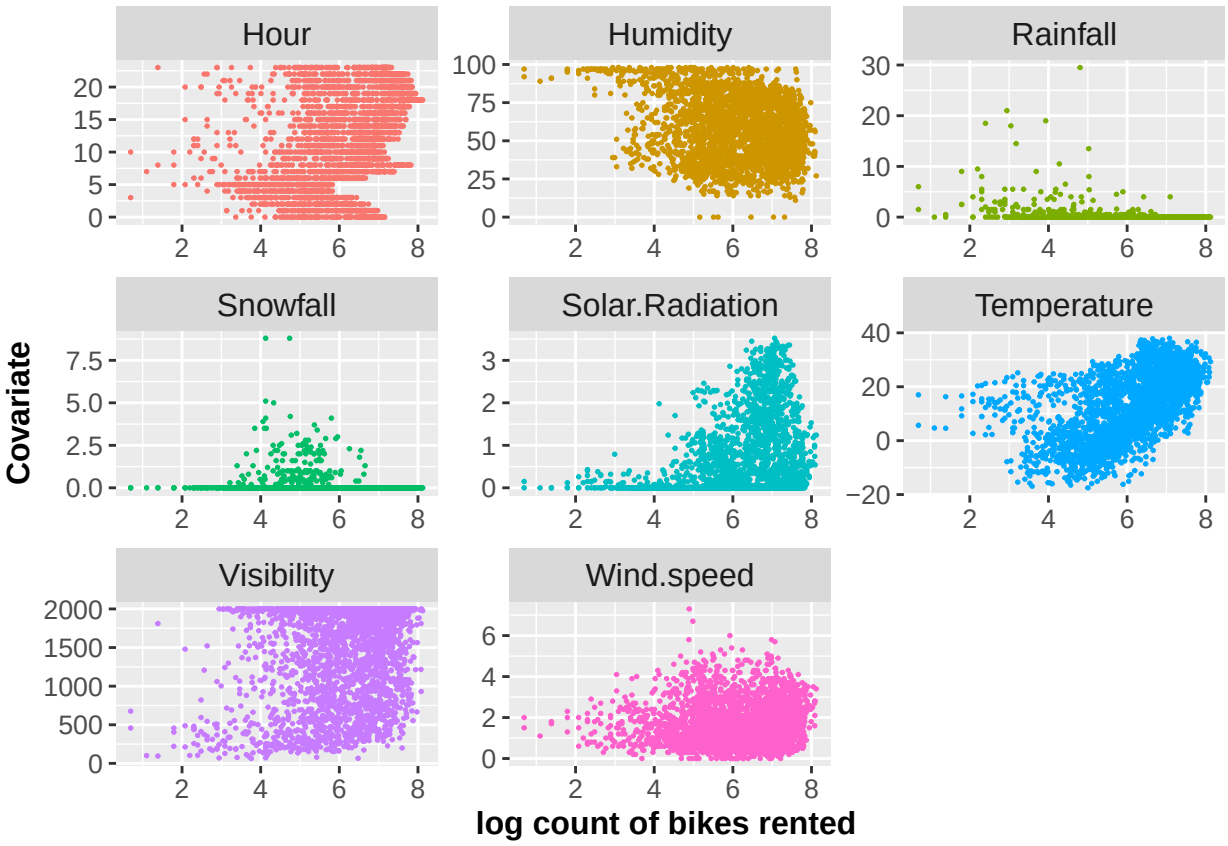
```
## Saving 6.5 x 4.5 in image
```

```r
out
```



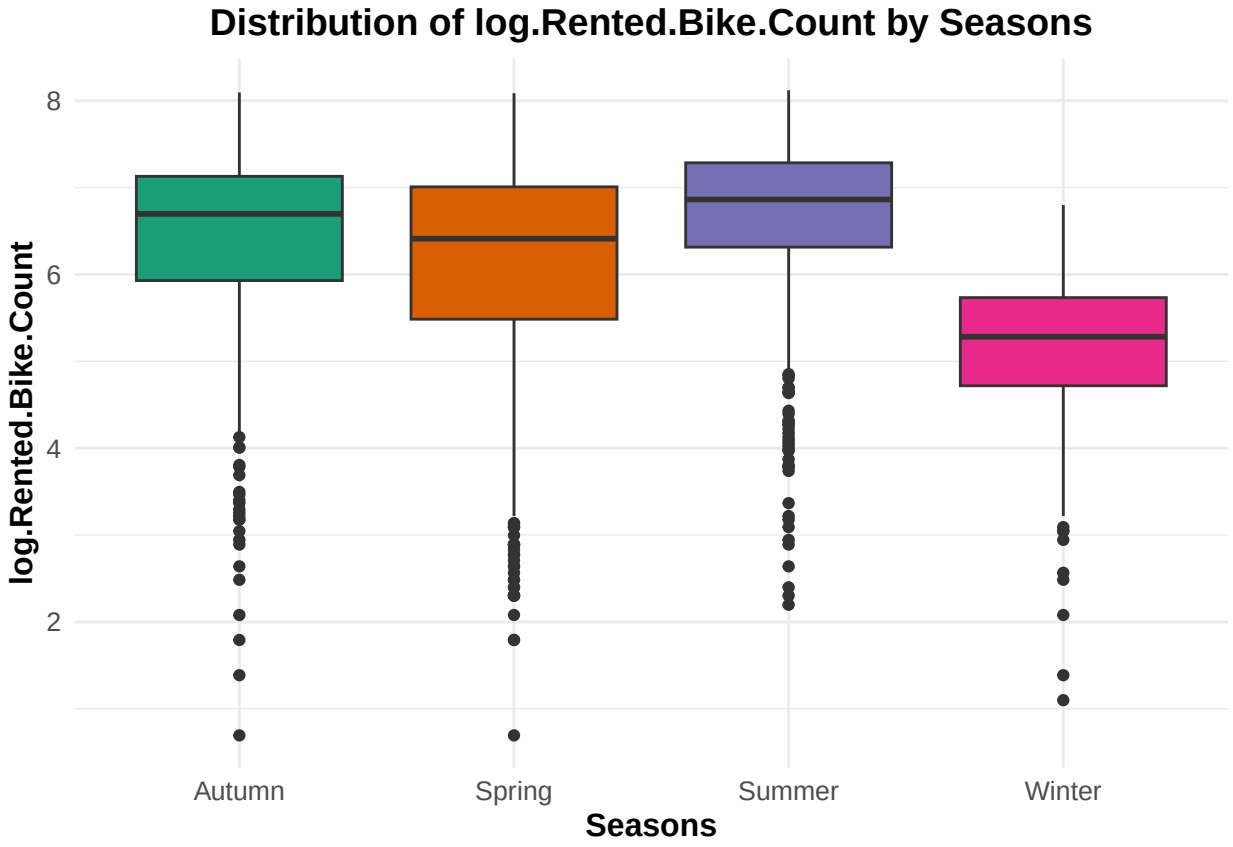

```r
#Correlation scatter plot for log.Rented.Bike.Count vs all variables
input_corr <- input_data  %>% dplyr::select(-c("Seasons","Holiday")) %>%
              tidyr::gather(key = "Covariates", value = "value",-log.Rented.Bike.Count)

corr_plot  <- ggplot(input_corr, aes(x = log.Rented.Bike.Count, y = value, col = Covariates)) +
  geom_point(size = 0.3) +
  facet_wrap(~Covariates, scales = "free") +
  xlab("log count of bikes rented ") + ylab("Covariate")  +
  theme(axis.text=element_text(size=10),
        axis.title=element_text(size=12 , face = "bold"),
        legend.title = element_text(size=12),
        legend.text = element_text(size=10),
```

```
        legend.position = "none",
        legend.direction = "horizontal",
        strip.text = element_text(size = 12)  # Add this line
      )
ggsave("Cor_plot.pdf", plot = corr_plot, width = 8.5, height = 4, units = "in")
corr_plot
```

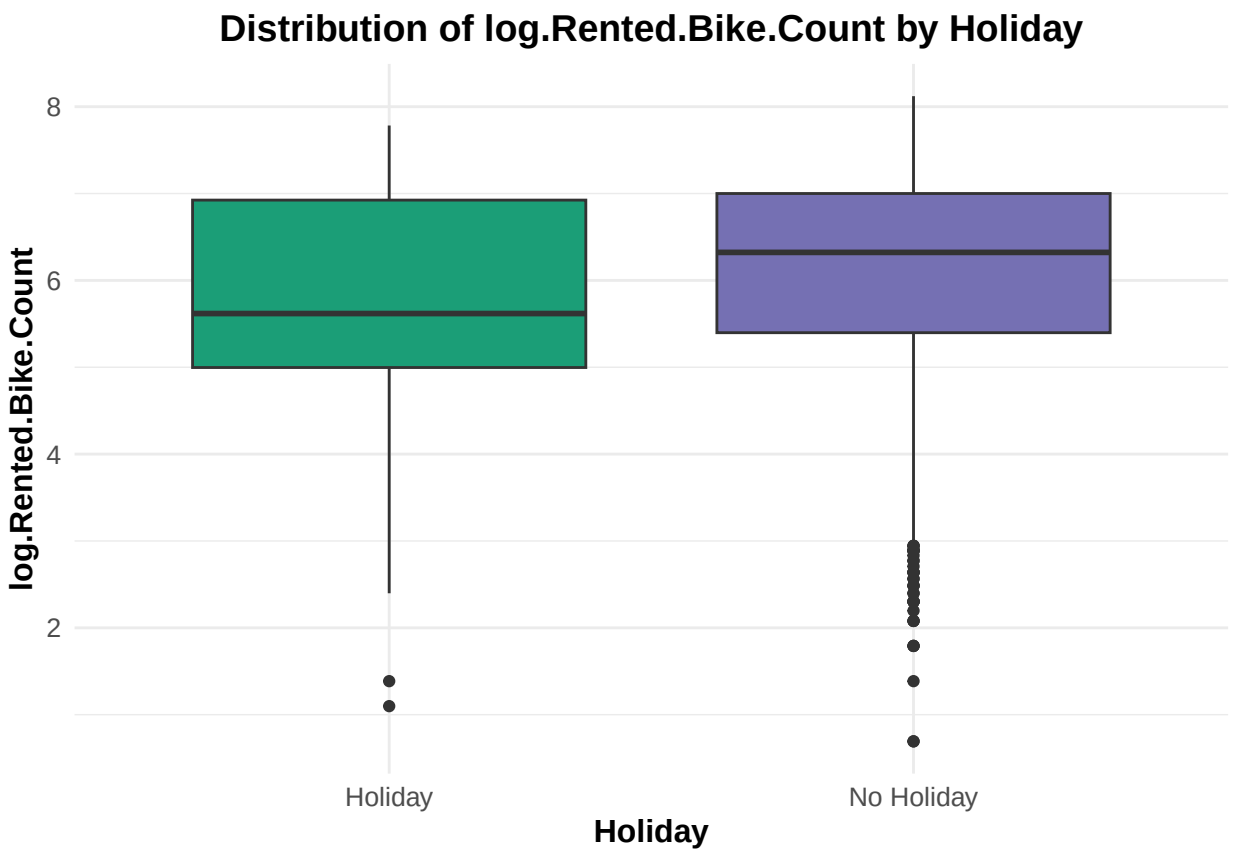



```
# Load the required packages
library(RColorBrewer)

# Create a color palette
my_palette <- brewer.pal(4, "Dark2")

# Create box plots for log.Rented.Bike.Count by Seasons
sea = ggplot(input_data, aes(x = Seasons, y = log.Rented.Bike.Count, fill = Seasons)) +
  geom_boxplot() +
  scale_fill_manual(values = my_palette) +
  xlab("Seasons") +
  ylab("log.Rented.Bike.Count") +
  ggtitle("Distribution of log.Rented.Bike.Count by Seasons") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold", size = 14),
        axis.title = element_text(face = "bold", size = 12),
        axis.text = element_text(size = 10),
        legend.position = "none")

# Create box plots for log.Rented.Bike.Count by Holiday
```

```
hol = ggplot(input_data, aes(x = Holiday, y = log.Rented.Bike.Count, fill = Holiday)) +
  geom_boxplot() +
  scale_fill_manual(values = my_palette[c(1, 3)]) +
  xlab("Holiday") +
  ylab("log.Rented.Bike.Count") +
  ggtitle("Distribution of log.Rented.Bike.Count by Holiday") +
  theme_minimal() +
  theme(plot.title = element_text(hjust = 0.5, face = "bold", size = 14),
        axis.title = element_text(face = "bold", size = 12),
        axis.text = element_text(size = 10),
        legend.position = "none")
ggsave("hol.pdf", plot = hol, width = 8.5, height = 4, units = "in")
ggsave("sea.pdf", plot = sea, width = 8.5, height = 4, units = "in")
sea
```




```
hol
```




## Task 2. Fitting a Full Linear Regression Model

```
# Selecting "Winter" as reference category for Seasons
input_data$Seasons <- relevel(factor(input_data$Seasons), ref = "Winter")
```

```
# Fit the linear regression model
model <- lm(log.Rented.Bike.Count ~ ., data = input_data)

# Display the model summary
summary(model)
```

```
##
## Call:
## lm(formula = log.Rented.Bike.Count ~ ., data = input_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.0145 -0.3517  0.0602  0.4281  5.0742
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         5.430e+00  1.156e-01  46.973  < 2e-16 ***
## Hour                4.448e-02  2.231e-03  19.940  < 2e-16 ***
## Temperature         4.094e-02  2.589e-03  15.813  < 2e-16 ***
## Humidity           -1.805e-02  1.074e-03 -16.796  < 2e-16 ***
## Wind.speed         -2.858e-02  1.534e-02  -1.864   0.0625 .
```

```
## Visibility        -1.734e-05  2.912e-05  -0.595   0.5517
## Solar.Radiation   -2.472e-02  2.200e-02  -1.124   0.2613
## Rainfall          -2.259e-01  1.227e-02 -18.407  < 2e-16 ***
## Snowfall          -6.272e-03  3.142e-02  -0.200   0.8418
## SeasonsAutumn      7.835e-01  5.808e-02  13.490  < 2e-16 ***
## SeasonsSpring      5.101e-01  5.534e-02   9.217  < 2e-16 ***
## SeasonsSummer      6.071e-01  8.335e-02   7.284 4.16e-13 ***
## HolidayNo Holiday  3.354e-01  6.356e-02   5.277 1.41e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7421 on 2892 degrees of freedom
## Multiple R-squared:  0.5939, Adjusted R-squared:  0.5922
## F-statistic: 352.5 on 12 and 2892 DF,  p-value: < 2.2e-16
```

## Task 3. Fitting Linear regression using Backward Selection Method with AIC Criteria

```
#Finding the Optimal set of Covariates using Backward Selection Method and AIC
# Load the MASS package
library(MASS)
```

```
##
## Attaching package: 'MASS'
```

```
## The following object is masked from 'package:dplyr':
##
##     select
```

```
# Apply stepAIC() with direction='backward'
AIC_model <- stepAIC(model, direction='backward')
```

```
## Start:  AIC=-1720.06
## log.Rented.Bike.Count ~ Hour + Temperature + Humidity + Wind.speed +
##     Visibility + Solar.Radiation + Rainfall + Snowfall + Seasons +
##     Holiday
##
##                   Df Sum of Sq    RSS     AIC
## - Snowfall         1     0.022 1592.6 -1722.0
## - Visibility       1     0.195 1592.8 -1721.7
## - Solar.Radiation  1     0.695 1593.3 -1720.8
## <none>                         1592.6 -1720.1
## - Wind.speed       1     1.912 1594.5 -1718.6
## - Holiday          1    15.335 1608.0 -1694.2
## - Seasons          3   121.368 1714.0 -1512.7
## - Temperature      1   137.697 1730.3 -1481.2
## - Humidity         1   155.357 1748.0 -1451.7
## - Rainfall         1   186.586 1779.2 -1400.2
## - Hour             1   218.960 1811.6 -1347.8
##
## Step:  AIC=-1722.02
## log.Rented.Bike.Count ~ Hour + Temperature + Humidity + Wind.speed +
##     Visibility + Solar.Radiation + Rainfall + Seasons + Holiday
##
```

```
##                   Df Sum of Sq    RSS     AIC
## - Visibility       1     0.189 1592.8 -1723.7
## - Solar.Radiation  1     0.729 1593.4 -1722.7
## <none>                         1592.6 -1722.0
## - Wind.speed       1     1.919 1594.6 -1720.5
## - Holiday          1    15.314 1608.0 -1696.2
## - Seasons          3   121.907 1714.5 -1513.8
## - Temperature      1   140.281 1732.9 -1478.8
## - Humidity         1   159.205 1751.8 -1447.2
## - Rainfall         1   186.788 1779.4 -1401.9
## - Hour             1   219.087 1811.7 -1349.6
##
## Step:  AIC=-1723.67
## log.Rented.Bike.Count ~ Hour + Temperature + Humidity + Wind.speed +
##     Solar.Radiation + Rainfall + Seasons + Holiday
##
##                   Df Sum of Sq    RSS     AIC
## - Solar.Radiation  1     0.600 1593.4 -1724.6
## <none>                         1592.8 -1723.7
## - Wind.speed       1     2.052 1594.9 -1721.9
## - Holiday          1    15.318 1608.2 -1697.9
## - Seasons          3   122.351 1715.2 -1514.7
## - Temperature      1   140.227 1733.1 -1480.6
## - Rainfall         1   186.604 1779.4 -1403.8
## - Hour             1   220.992 1813.8 -1348.2
## - Humidity         1   224.370 1817.2 -1342.8
##
## Step:  AIC=-1724.58
## log.Rented.Bike.Count ~ Hour + Temperature + Humidity + Wind.speed +
##     Rainfall + Seasons + Holiday
##
##               Df Sum of Sq    RSS     AIC
## <none>                     1593.4 -1724.6
## - Wind.speed   1     2.812 1596.2 -1721.5
## - Holiday      1    15.269 1608.7 -1698.9
## - Seasons      3   123.653 1717.1 -1513.5
## - Temperature  1   154.669 1748.1 -1457.5
## - Rainfall     1   187.466 1780.9 -1403.5
## - Hour         1   228.768 1822.2 -1336.9
## - Humidity     1   276.807 1870.2 -1261.3
```

```
# Print the summary of the reduced model
summary(AIC_model)
```

```
##
## Call:
## lm(formula = log.Rented.Bike.Count ~ Hour + Temperature + Humidity +
##     Wind.speed + Rainfall + Seasons + Holiday, data = input_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.0308 -0.3517  0.0619  0.4306  5.0704
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)        5.3604709  0.0858857  62.414  < 2e-16 ***
## Hour               0.0448769  0.0022012  20.387  < 2e-16 ***
## Temperature        0.0399991  0.0023861  16.763  < 2e-16 ***
## Humidity          -0.0172877  0.0007709 -22.426  < 2e-16 ***
## Wind.speed        -0.0334258  0.0147875  -2.260   0.0239 *
## Rainfall          -0.2259778  0.0122447 -18.455  < 2e-16 ***
## SeasonsAutumn      0.7842994  0.0569279  13.777  < 2e-16 ***
## SeasonsSpring      0.5144989  0.0549369   9.365  < 2e-16 ***
## SeasonsSummer      0.6109687  0.0816558   7.482 9.64e-14 ***
## HolidayNo Holiday  0.3344850  0.0635064   5.267 1.49e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7419 on 2895 degrees of freedom
## Multiple R-squared:  0.5937, Adjusted R-squared:  0.5925
## F-statistic: 470.1 on 9 and 2895 DF,  p-value: < 2.2e-16
```

## Fitting Linear Regression using Backward Selection Method with BIC Criteria

```
#Finding the Optimal set of Covariates using Backward Selection Method and BIC
BIC_model <- stepAIC(model, direction = "backward", k = log(nrow(input_data)))
```

```
## Start:  AIC=-1642.39
## log.Rented.Bike.Count ~ Hour + Temperature + Humidity + Wind.speed +
##     Visibility + Solar.Radiation + Rainfall + Snowfall + Seasons +
##     Holiday
##
##                   Df Sum of Sq    RSS     AIC
## - Snowfall         1     0.022 1592.6 -1650.3
## - Visibility       1     0.195 1592.8 -1650.0
## - Solar.Radiation  1     0.695 1593.3 -1649.1
## - Wind.speed       1     1.912 1594.5 -1646.9
## <none>                         1592.6 -1642.4
## - Holiday          1    15.335 1608.0 -1622.5
## - Seasons          3   121.368 1714.0 -1453.0
## - Temperature      1   137.697 1730.3 -1409.5
## - Humidity         1   155.357 1748.0 -1380.0
## - Rainfall         1   186.586 1779.2 -1328.5
## - Hour             1   218.960 1811.6 -1276.2
##
## Step:  AIC=-1650.33
## log.Rented.Bike.Count ~ Hour + Temperature + Humidity + Wind.speed +
##     Visibility + Solar.Radiation + Rainfall + Seasons + Holiday
##
##                   Df Sum of Sq    RSS     AIC
## - Visibility       1     0.189 1592.8 -1658.0
## - Solar.Radiation  1     0.729 1593.4 -1657.0
## - Wind.speed       1     1.919 1594.6 -1654.8
## <none>                         1592.6 -1650.3
## - Holiday          1    15.314 1608.0 -1630.5
## - Seasons          3   121.907 1714.5 -1460.0
## - Temperature      1   140.281 1732.9 -1413.1
```

```
## - Humidity         1   159.205 1751.8 -1381.5
## - Rainfall         1   186.788 1779.4 -1336.1
## - Hour             1   219.087 1811.7 -1283.9
##
## Step:  AIC=-1657.96
## log.Rented.Bike.Count ~ Hour + Temperature + Humidity + Wind.speed +
##     Solar.Radiation + Rainfall + Seasons + Holiday
##
##                   Df Sum of Sq    RSS     AIC
## - Solar.Radiation  1     0.600 1593.4 -1664.8
## - Wind.speed       1     2.052 1594.9 -1662.2
## <none>                         1592.8 -1658.0
## - Holiday          1    15.318 1608.2 -1638.1
## - Seasons          3   122.351 1715.2 -1466.9
## - Temperature      1   140.227 1733.1 -1420.8
## - Rainfall         1   186.604 1779.4 -1344.1
## - Hour             1   220.992 1813.8 -1288.5
## - Humidity         1   224.370 1817.2 -1283.1
##
## Step:  AIC=-1664.84
## log.Rented.Bike.Count ~ Hour + Temperature + Humidity + Wind.speed +
##     Rainfall + Seasons + Holiday
##
##               Df Sum of Sq    RSS     AIC
## - Wind.speed   1     2.812 1596.2 -1667.7
## <none>                     1593.4 -1664.8
## - Holiday      1    15.269 1608.7 -1645.1
## - Seasons      3   123.653 1717.1 -1471.6
## - Temperature  1   154.669 1748.1 -1403.7
## - Rainfall     1   187.466 1780.9 -1349.7
## - Hour         1   228.768 1822.2 -1283.1
## - Humidity     1   276.807 1870.2 -1207.5
##
## Step:  AIC=-1667.69
## log.Rented.Bike.Count ~ Hour + Temperature + Humidity + Rainfall +
##     Seasons + Holiday
##
##               Df Sum of Sq    RSS     AIC
## <none>                     1596.2 -1667.7
## - Holiday      1    15.179 1611.4 -1648.2
## - Seasons      3   128.205 1724.5 -1467.2
## - Temperature  1   153.583 1749.8 -1408.8
## - Rainfall     1   189.388 1785.6 -1350.0
## - Hour         1   229.276 1825.5 -1285.8
## - Humidity     1   283.272 1879.5 -1201.1
```

```
summary(BIC_model)
```

```
##
## Call:
## lm(formula = log.Rented.Bike.Count ~ Hour + Temperature + Humidity +
##     Rainfall + Seasons + Holiday, data = input_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -5.0420 -0.3540  0.0587  0.4319  5.0776
##
## Coefficients:
##                      Estimate Std. Error t value Pr(>|t|)
## (Intercept)         5.2861901  0.0794065  66.571  < 2e-16 ***
## Hour                0.0437416  0.0021447  20.395  < 2e-16 ***
## Temperature         0.0398414  0.0023868  16.692  < 2e-16 ***
## Humidity           -0.0168055  0.0007413 -22.670  < 2e-16 ***
## Rainfall           -0.2269834  0.0122453 -18.536  < 2e-16 ***
## SeasonsAutumn       0.7958984  0.0567363  14.028  < 2e-16 ***
## SeasonsSpring       0.5145910  0.0549759   9.360  < 2e-16 ***
## SeasonsSummer       0.6187082  0.0816418   7.578 4.68e-14 ***
## HolidayNo Holiday   0.3334918  0.0635499   5.248 1.65e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.7424 on 2896 degrees of freedom
## Multiple R-squared:  0.593,  Adjusted R-squared:  0.5919
## F-statistic: 527.5 on 8 and 2896 DF,  p-value: < 2.2e-16
```

### 95% confidence intervals for the coefficients of Linear Regression Model using Backward Selection Method with BIC Criteria

```
# Compute 95% confidence intervals for the coefficients
CI <- confint(BIC_model, level = 0.95)
CI
```

```
##                         2.5 %      97.5 %
## (Intercept)        5.13049109  5.44188909
## Hour               0.03953632  0.04794689
## Temperature        0.03516142  0.04452135
## Humidity          -0.01825910 -0.01535199
## Rainfall          -0.25099368 -0.20297306
## SeasonsAutumn      0.68465073  0.90714612
## SeasonsSpring      0.40679524  0.62238672
## SeasonsSummer      0.45862626  0.77879011
## HolidayNo Holiday  0.20888413  0.45809950
```

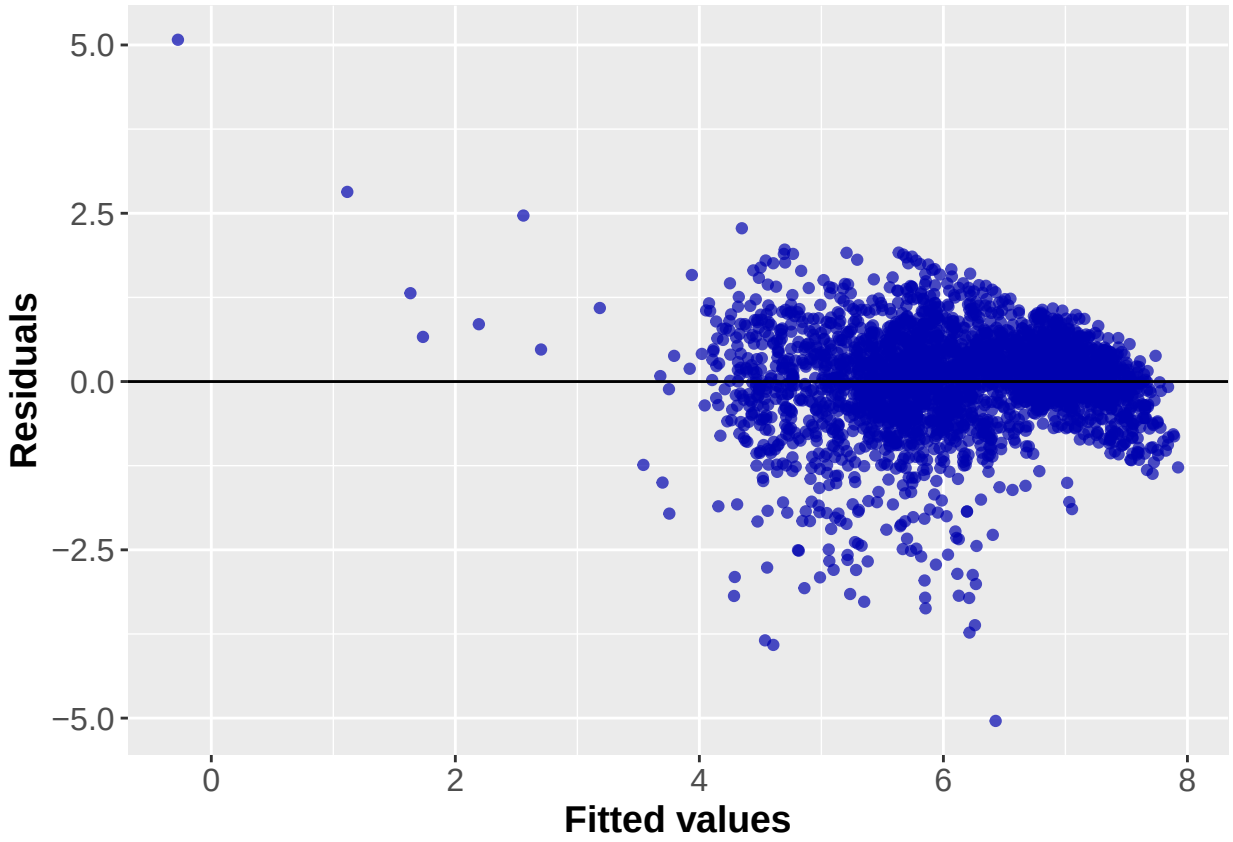
## Task 4. Verifying Model Assumptions

### Residual plot

```
#Residual plot for standardized residuals vs Fitted best subset model
res = ggplot(BIC_model, aes(x = .fitted, y = BIC_model$residuals)) +
 geom_point(size = 1.5, color = "#0002AF", alpha = 0.7) +
 geom_hline(yintercept = mean(BIC_model$residuals), color = "black", linewidth = 0.5) +
  xlab("Fitted values") + ylab("Residuals") +
  theme(axis.text=element_text(size=12), axis.title=element_text(size=14 , face = "bold"))

ggsave("res_plot.pdf", plot = res)
```

```
## Saving 6.5 x 4.5 in image
```

```
res
```

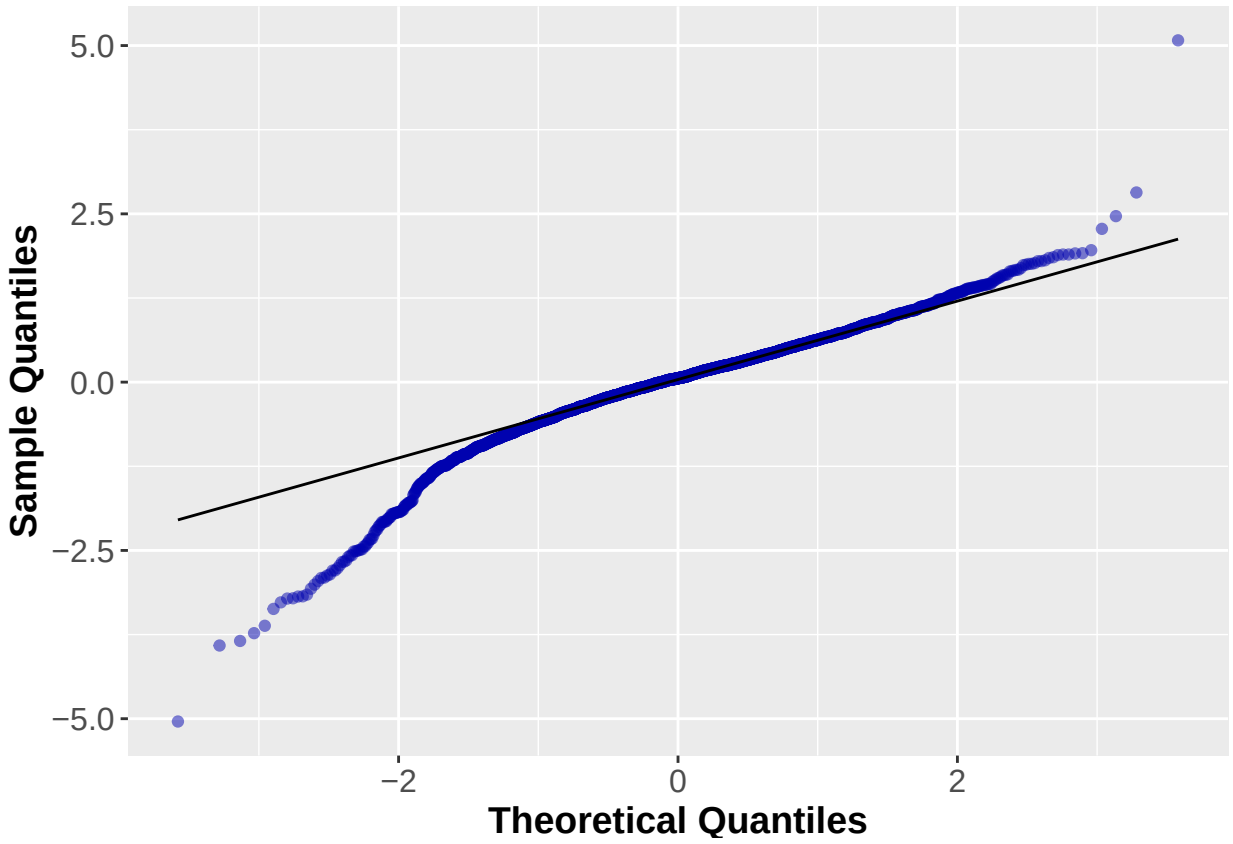



## QQ Plot

```
# QQ plot
nn = ggplot(input_data, aes(sample = BIC_model$residuals)) +
 stat_qq(distribution = stats::qnorm, color = "#0002AF", alpha = 0.5) +
 stat_qq_line() +
 xlab("Theoretical Quantiles") +
 ylab("Sample Quantiles") +
 theme(plot.title = element_text(face = "bold",hjust = 0.5, size = 12),
 axis.text=element_text(size=12),
 axis.title=element_text(size=14, face = "bold"),
 legend.text = element_text(size = 12))

ggsave("nn_plot.pdf", plot = nn)
```

```
## Saving 6.5 x 4.5 in image
```

```
nn
```




### Variance Inflation Factor (VIF)

```
library(car)
```

```
## Loading required package: carData
```

```
##
## Attaching package: 'car'
```

```
## The following object is masked from 'package:dplyr':
##
##     recode
```

```
# Calculate VIF for each predictor variable
vif_values <- vif(BIC_model)

# Print VIF values
vif_values
```

```
##                 GVIF Df GVIF^(1/(2*Df))
## Hour        1.143812  1        1.069492
## Temperature 4.480877  1        2.116808
## Humidity    1.224578  1        1.106606
## Rainfall    1.061596  1        1.030338
## Seasons     4.618603  3        1.290481
## Holiday     1.029158  1        1.014474
```

## AIC and BIC of all Models

```
# Print AIC of Models
print(paste("AIC of Full Model: ", AIC(model)))
```

```
## [1] "AIC of Full Model:  6525.9759156717"
```

```
print(paste("AIC of Backward Selection Model with AIC: ", AIC(AIC_model)))
```

```
## [1] "AIC of Backward Selection Model with AIC:  6521.45467276738"
```

```
print(paste("AIC of Backward Selection Model with BIC: ", AIC(BIC_model)))
```

```
## [1] "AIC of Backward Selection Model with BIC:  6524.57720922221"
```

```
# Print BIC of Models
print(paste("BIC of Full Model: ", BIC(model)))
```

```
## [1] "BIC of Full Model:  6609.6145570417"
```

```
print(paste("BIC of Backward Selection Model with AIC: ", BIC(AIC_model)))
```

```
## [1] "BIC of Backward Selection Model with AIC:  6587.17074812953"
```

```
print(paste("BIC of Backward Selection Model with BIC: ", BIC(BIC_model)))
```

```
## [1] "BIC of Backward Selection Model with BIC:  6584.31909591507"
```