Stats 101A final project

Roger Wilson – 805504529

03/22/23

Contents

Setup

This data set of interest is of **Capital Bikeshare** rentals in 2011 and 2012, sourced from the **UCI Machine Learning Repository**. From the site:

"Bike-sharing systems represent a new generation of traditional bike rentals, where the entire process - from membership and rental to return - has become automatic. With these systems, users can easily rent a bike from a particular location and return it to another. Currently, there are over 500 bike-sharing programs worldwide, which comprise more than 500,000 bicycles. These systems are of great interest due to their important role in addressing traffic, environmental, and health issues.

In addition to the interesting real-world applications of bike-sharing systems, the characteristics of the data generated by these systems make them attractive for research. Unlike other transport services such as buses or subways, the duration of travel, departure, and arrival positions are explicitly recorded in these systems. This feature turns bike-sharing systems into a virtual sensor network that can be used for sensing mobility in the city. Therefore, it is expected that most of the important events in the city could be detected by monitoring this data."

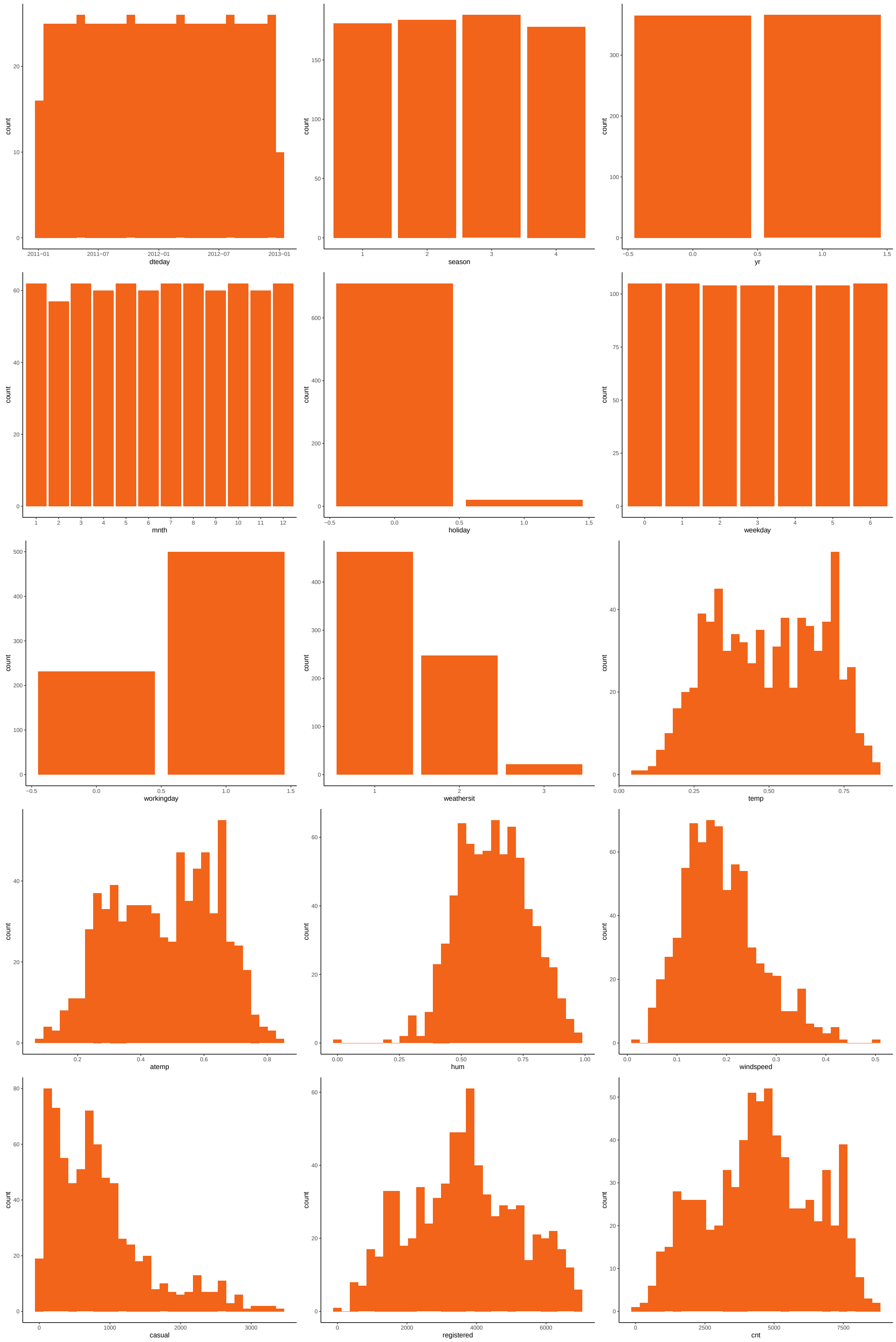
This project will analyze bike share demand in Washington D.C. from 2011-2012. The variables can be described below:

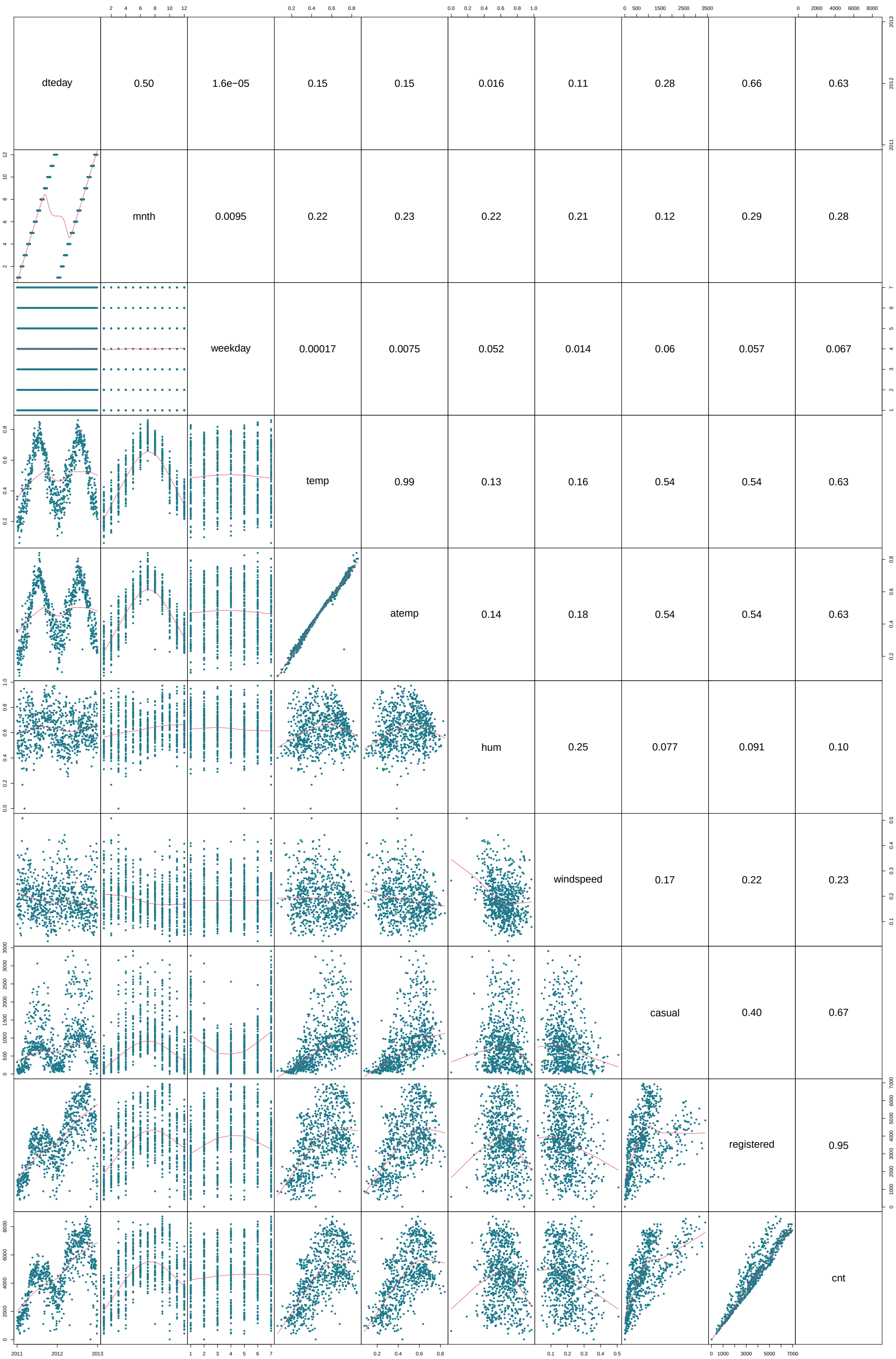
Variable	Description
instant	record index
dteday	date
season	season (1:winter, 2:spring, 3:summer, 4:fall)
yr	year (0: 2011, 1:2012)
mnth	month (1 to 12)
hr	hour (0 to 23)
holiday	weather day is holiday or not
weekday	day of the week
workingday	if day is neither weekend nor holiday is 1, otherwise is 0
weathersit	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	Normalized temperature in Celsius. The values are derived via (t-t_min)/(t_max-t_min), t_min=-8, t_max=+39
atemp	Normalized feeling temperature in Celsius. The values are derived via (t-t_min)/(t_max-t_min), t_min=-16, t_max=+50
hum	Normalized humidity. The values are divided to 100 (max)
windspeed	Normalized wind speed. The values are divided to 67 (max)
casual	count of casual users
registered	count of registered users
cnt	count of total rental bikes including both casual and registered

The index will be excluded from the model.

Exploratory Analysis

```
##      dteday               season            yr              mnth      
##  Min.   :2011-01-01   Min.   :1.000   Min.   :0.0000   Min.   : 1.00  
##  1st Qu.:2011-07-02   1st Qu.:2.000   1st Qu.:0.0000   1st Qu.: 4.00  
##  Median :2012-01-01   Median :3.000   Median :1.0000   Median : 7.00  
##  Mean   :2012-01-01   Mean   :2.497   Mean   :0.5007   Mean   : 6.52  
##  3rd Qu.:2012-07-01   3rd Qu.:3.000   3rd Qu.:1.0000   3rd Qu.:10.00  
##  Max.   :2012-12-31   Max.   :4.000   Max.   :1.0000   Max.   :12.00  
##     holiday            weekday        workingday      weathersit   
##  Min.   :0.00000   Min.   :0.000   Min.   :0.000   Min.   :1.000  
##  1st Qu.:0.00000   1st Qu.:1.000   1st Qu.:0.000   1st Qu.:1.000  
##  Median :0.00000   Median :3.000   Median :1.000   Median :1.000  
##  Mean   :0.02873   Mean   :2.997   Mean   :0.684   Mean   :1.395  
##  3rd Qu.:0.00000   3rd Qu.:5.000   3rd Qu.:1.000   3rd Qu.:2.000  
##  Max.   :1.00000   Max.   :6.000   Max.   :1.000   Max.   :3.000  
##       temp             atemp              hum           windspeed      
##  Min.   :0.05913   Min.   :0.07907   Min.   :0.0000   Min.   :0.02239  
##  1st Qu.:0.33708   1st Qu.:0.33784   1st Qu.:0.5200   1st Qu.:0.13495  
##  Median :0.49833   Median :0.48673   Median :0.6267   Median :0.18097  
##  Mean   :0.49538   Mean   :0.47435   Mean   :0.6279   Mean   :0.19049  
##  3rd Qu.:0.65542   3rd Qu.:0.60860   3rd Qu.:0.7302   3rd Qu.:0.23321  
##  Max.   :0.86167   Max.   :0.84090   Max.   :0.9725   Max.   :0.50746  
##      casual         registered        cnt      
##  Min.   :   2.0   Min.   :  20   Min.   :  22  
##  1st Qu.: 315.5   1st Qu.:2497   1st Qu.:3152  
##  Median : 713.0   Median :3662   Median :4548  
##  Mean   : 848.2   Mean   :3656   Mean   :4504  
##  3rd Qu.:1096.0   3rd Qu.:4776   3rd Qu.:5956  
##  Max.   :3410.0   Max.   :6946   Max.   :8714  
```


count
dteday
season
yr
mnth
holiday
weekday
workingday
weathersit
temp
atemp
hum
windspeed
casual
registered
cnt


dteday
mnth
weekday
temp
atemp
hum
windspeed
casual
registered
cnt
0.50
1.6e−05
0.15
0.15
0.016
0.11
0.28
0.66
0.63
0.0095
0.22
0.23
0.22
0.21
0.12
0.29
0.28
0.00017
0.0075
0.052
0.014
0.06
0.057
0.067
0.99
0.13
0.16
0.54
0.54
0.63
0.14
0.18
0.54
0.54
0.63
0.25
0.077
0.091
0.10
0.17
0.22
0.23
0.40
0.67
0.95

There is very few skewness in our variables. There are overwhelmingly more non-holidays to holidays and sunny / cloudy days to rainy days: there are also no recorded days of snow. The total count of rented bikes is equal to the sum of casual riders and registered riders, and it's interesting that casual riders is right skewed. In the correlation plots, it seems that weather-related factors such as temperature and conditions have the largest effect on the total number of rented bikes. Casual and registered, being directly related to the total, will be excluded from our models. It would be interesting to test models on these in future projects for comparison.

There does not seem to be severe collinearity between the variables, besides obviously temperature and perceived temperature.

Year should also be removed from the analysis: This model will be focusing on predicting the number rented bikes based on a single day's conditions, and there is not as much interest in differences between 2011 and 2012 numbers.

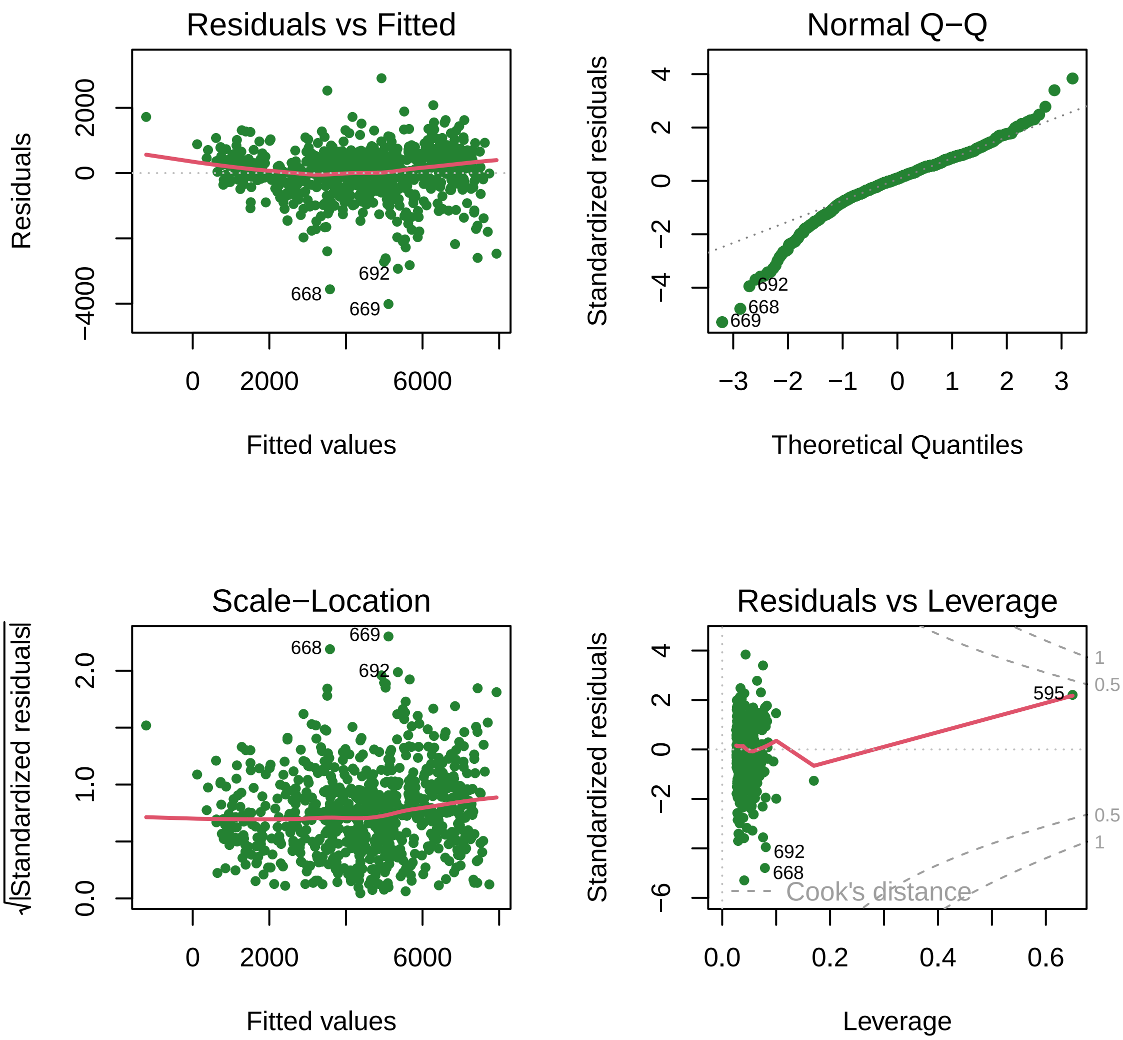
Original Model

We are just going to throw the kitchen sink at this data.

```
##
## Call:
## lm(formula = cnt ~ . - yr - casual - registered - workingday,
##     data = bikes)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -4015   -362     64    451   2907
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -80781.03    2452.81 -32.934  < 2e-16 ***
## dteday           5.49       0.16  34.324  < 2e-16 ***
## season2        878.34     180.70   4.861 1.44e-06 ***
## season3        828.14     214.56   3.860 0.000124 ***
## season4       1581.97     182.22   8.682  < 2e-16 ***
## mnth2          -32.33     144.70  -0.223 0.823261
## mnth3          219.22     166.24   1.319 0.187716
## mnth4          -35.46     248.73  -0.143 0.886669
## mnth5           85.26     268.63   0.317 0.751039
## mnth6         -304.62     282.84  -1.077 0.281858
## mnth7         -956.94     314.52  -3.042 0.002434 **
## mnth8         -711.11     304.18  -2.338 0.019677 *
## mnth9         -322.40     267.21  -1.207 0.228020
## mnth10        -985.34     245.37  -4.016 6.57e-05 ***
## mnth11       -1790.67     236.58  -7.569 1.18e-13 ***
## mnth12       -1928.86     190.73 -10.113  < 2e-16 ***
## holiday       -583.31     181.57  -3.212 0.001376 **
## weekday1       208.68     110.23   1.893 0.058752 .
## weekday2       308.52     107.85   2.861 0.004355 **
```

```
## weekday3       380.34     108.20   3.515 0.000468 ***
## weekday4       384.75     108.25   3.554 0.000404 ***
## weekday5       434.70     108.16   4.019 6.48e-05 ***
## weekday6       438.98     107.28   4.092 4.77e-05 ***
## weathersit2   -451.66      77.58  -5.822 8.86e-09 ***
## weathersit3  -1955.04     198.37  -9.855  < 2e-16 ***
## temp          2744.26    1407.80   1.949 0.051654 .
## atemp         1905.34    1472.03   1.294 0.195966
## hum          -1580.53     294.24  -5.372 1.06e-07 ***
## windspeed    -2848.83     417.30  -6.827 1.88e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 774.4 on 702 degrees of freedom
## Multiple R-squared:  0.8463, Adjusted R-squared:  0.8402
## F-statistic: 138.1 on 28 and 702 DF,  p-value: < 2.2e-16
```

This original model has an adjusted R-squared value of 0.8402. Most of the variables are statistically significant, besides a few factored months and perceived temperature.



These plots look ok. There is one point with high leverage, but doesn't have necessarily a high Cook's value. Some of the standardized residuals are left-skewed in the Q-Q plot.

```
##                     GVIF Df GVIF^(1/(2*Df))
## dteday        1.388778  1        1.178464
## season      169.912544  3        2.353446
## mnth        452.438817 11        1.320405
## holiday       1.121313  1        1.058921
## weekday       1.163200  6        1.012678
## weathersit    1.893669  2        1.173076
## temp         80.835565  1        8.990860
```

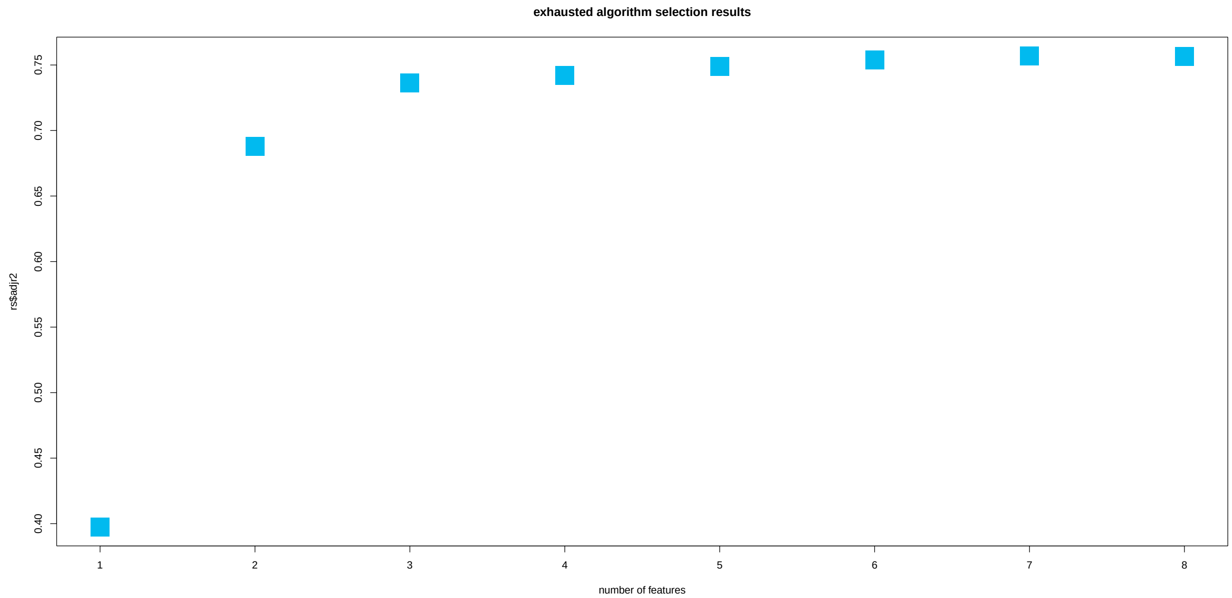
```
## atemp       70.045219  1        8.369302
## hum          2.137849  1        1.462138
## windspeed    1.273102  1        1.128318
```

However, we have VERY high variance inflation between our variables. This makes sense, particularly the relationship between `temp` and `atemp`, as well as between the weather during various seasons and months. We will remove these before performing forward feature selection.

Feature Selection

from Miles Chen 101A Nov 20 2015 youtube lecture:

```
## Subset selection object
## 8 Variables  (and intercept)
##            Forced in Forced out
## dteday         FALSE      FALSE
## season         FALSE      FALSE
## holiday        FALSE      FALSE
## weekday        FALSE      FALSE
## weathersit     FALSE      FALSE
## atemp          FALSE      FALSE
## hum            FALSE      FALSE
## windspeed      FALSE      FALSE
## 1 subsets of each size up to 8
## Selection Algorithm: exhaustive
##          dteday season holiday weekday weathersit atemp hum windspeed
## 1  ( 1 ) " "    " "    " "     " "     " "        "*"   " " " "
## 2  ( 1 ) "*"    " "    " "     " "     " "        "*"   " " " "
## 3  ( 1 ) "*"    " "    " "     " "     "*"        "*"   " " " "
## 4  ( 1 ) "*"    " "    " "     "*"     "*"        "*"   " " " "
## 5  ( 1 ) "*"    " "    " "     " "     "*"        "*"   "*" "*"
## 6  ( 1 ) "*"    " "    " "     "*"     "*"        "*"   "*" "*"
## 7  ( 1 ) "*"    " "    "*"     "*"     "*"        "*"   "*" "*"
## 8  ( 1 ) "*"    "*"    "*"     "*"     "*"        "*"   "*" "*"
```



It seems that a model with even just 3 variables would perform pretty well, and R^2 doesn't increase by much after that. It seems like date, day of the week, the weather and temperature have the greatest influence.

We'll use backward elimination to see if there's any differences.

```
## Start:  AIC=9937.07
## cnt ~ (dteday + season + yr + mnth + holiday + weekday + workingday +
##     weathersit + temp + atemp + hum + windspeed + casual + registered) -
##     yr - casual - registered - workingday - mnth - temp
##
##              Df Sum of Sq        RSS     AIC
## <none>                     559023934  9937.1
## - weekday     6  14663817  573687750  9944.0
## - holiday     1   9326879  568350812  9947.2
## - hum         1  21682806  580706740  9962.9
## - windspeed   1  23910052  582933986  9965.7
## - weathersit  2  59920188  618944121 10007.5
## - season      3  76605655  635629588 10024.9
## - atemp       1 188362382  747386316 10147.3
## - dteday      1 645459192 1204483126 10496.2
```

```
## Start:  AIC=11066.88
```

```
## cnt ~ 1
##
##              Df  Sum of Sq        RSS   AIC
## + atemp       1 1091003307 1648532085 10698
## + dteday      1 1083287662 1656247730 10701
## + season      3  950595868 1788939524 10761
## + weathersit  2  271644573 2467890819 10995
## + windspeed   1  150705556 2588829836 11028
## + hum         1   27757373 2711778019 11061
## + holiday     1   12797494 2726737898 11066
## <none>                     2739535392 11067
## + weekday     6   17659017 2721876375 11074
##
## Step:  AIC=10697.61
## cnt ~ atemp
##
##              Df Sum of Sq        RSS   AIC
## + dteday      1 795359095  853172990 10218
## + weathersit  2 163531664 1485000421 10625
## + season      3 154159967 1494372119 10632
## + hum         1  99815222 1548716864 10654
## + windspeed   1  39915579 1608616506 10682
## + holiday     1   6274901 1642257184 10697
## <none>                    1648532085 10698
## + weekday     6  15529667 1633002418 10703
##
## Step:  AIC=10218.11
## cnt ~ atemp + dteday
##
##              Df Sum of Sq       RSS   AIC
## + weathersit  2 157237928 695935062 10073
## + hum         1  96963223 756209767 10132
## + season      3  63502666 789670324 10168
```

```
## + windspeed   1  15030949 838142041 10207
## + holiday     1   9661567 843511424 10212
## + weekday     6  15373883 837799108 10217
## <none>                   853172990 10218
##
## Step:  AIC=10073.2
## cnt ~ atemp + dteday + weathersit
##
##             Df Sum of Sq       RSS   AIC
## + season     3  71420431 624514631 10000
## + hum        1  13488296 682446766 10061
## + holiday    1  13156484 682778577 10061
## + weekday    6  19701504 676233558 10064
## + windspeed  1   9883489 686051573 10065
## <none>                   695935062 10073
##
## Step:  AIC=10000.05
## cnt ~ atemp + dteday + weathersit + season
##
##             Df Sum of Sq       RSS     AIC
## + hum        1  14588837 609925794  9984.8
## + windspeed  1  13887966 610626664  9985.6
## + holiday    1  12040179 612474452  9987.8
## + weekday    6  20140870 604373761  9988.1
## <none>                   624514631 10000.1
##
## Step:  AIC=9984.77
## cnt ~ atemp + dteday + weathersit + season + hum
##
##             Df Sum of Sq       RSS    AIC
## + windspeed  1  24892109 585033685 9956.3
## + holiday    1  11688471 598237323 9972.6
## + weekday    6  17440082 592485711 9975.6
```

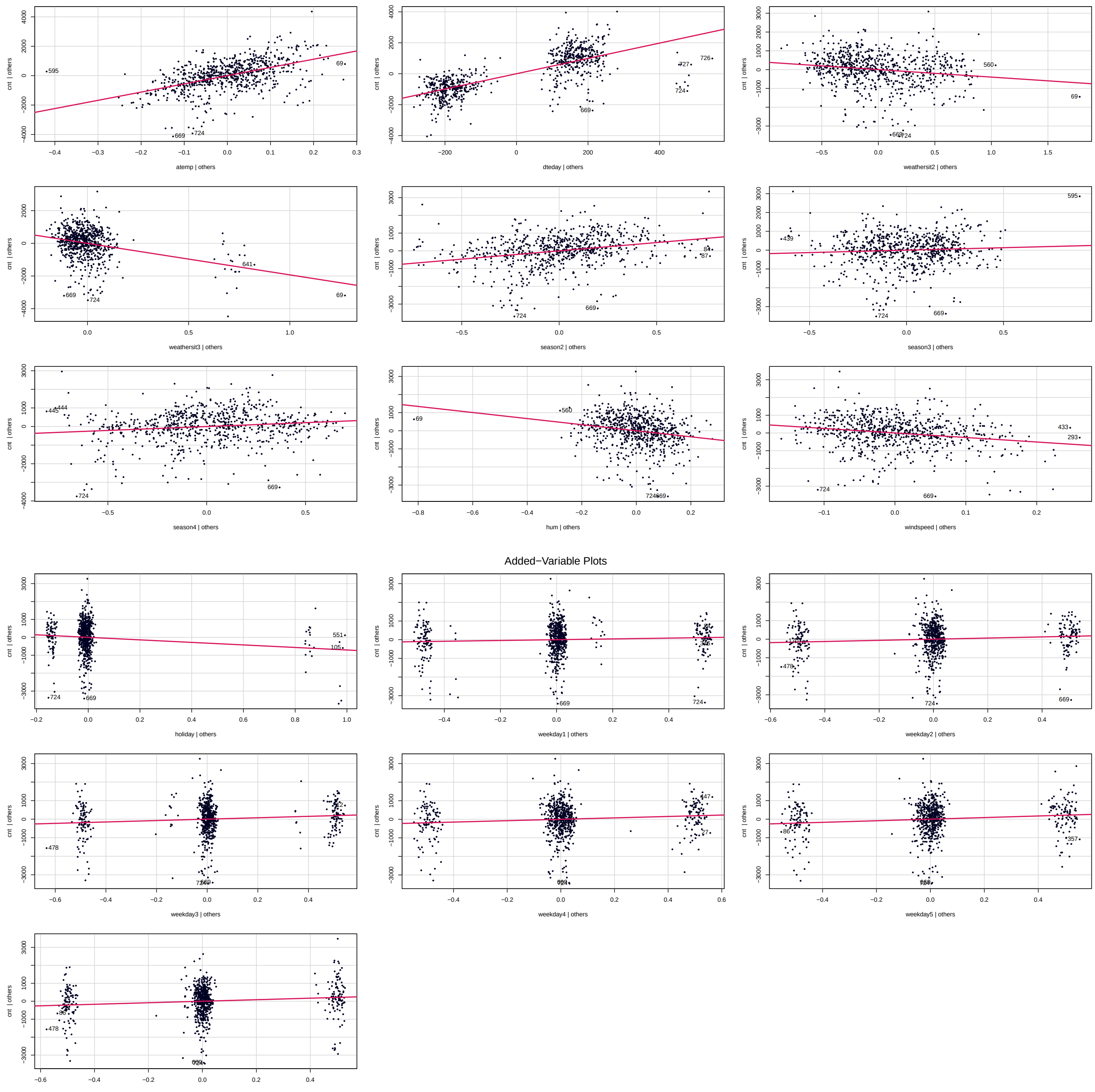
```
## <none>                    609925794 9984.8
##
## Step:  AIC=9956.31
## cnt ~ atemp + dteday + weathersit + season + hum + windspeed
##
##           Df Sum of Sq       RSS    AIC
## + holiday  1  11345935 573687750 9944.0
## + weekday  6  16682873 568350812 9947.2
## <none>                 585033685 9956.3
##
## Step:  AIC=9944
## cnt ~ atemp + dteday + weathersit + season + hum + windspeed +
##     holiday
##
##           Df Sum of Sq       RSS    AIC
## + weekday  6  14663817 559023934 9937.1
## <none>                 573687750 9944.0
##
## Step:  AIC=9937.07
## cnt ~ atemp + dteday + weathersit + season + hum + windspeed +
##     holiday + weekday
```

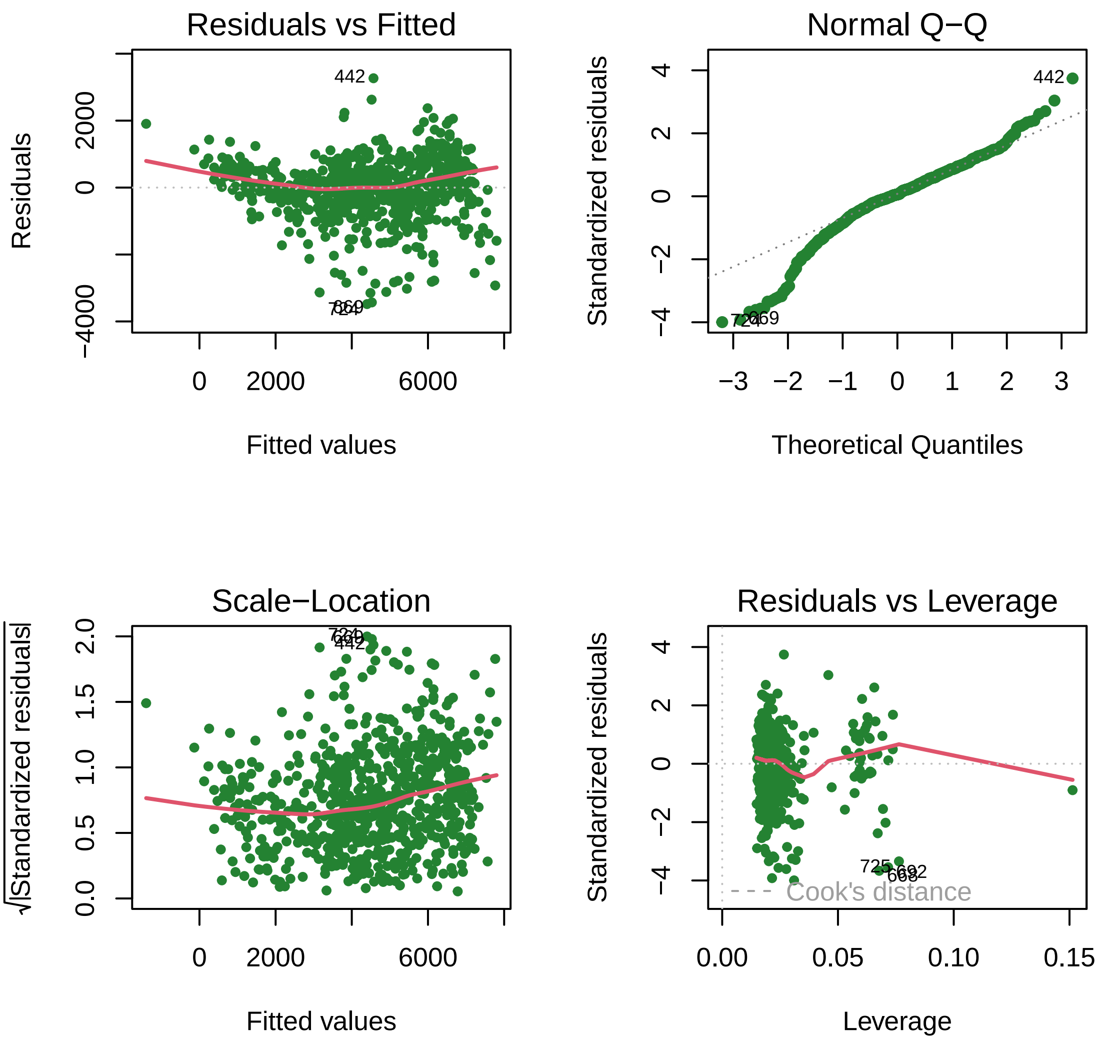
We get the same results. So, we'll keep the forward AIC model.

```
##
## Call:
## lm(formula = cnt ~ atemp + dteday + weathersit + season + hum +
##     windspeed + holiday + weekday, data = bikes)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3481.1  -382.8    53.5   525.2  3266.5
##
## Coefficients:
```

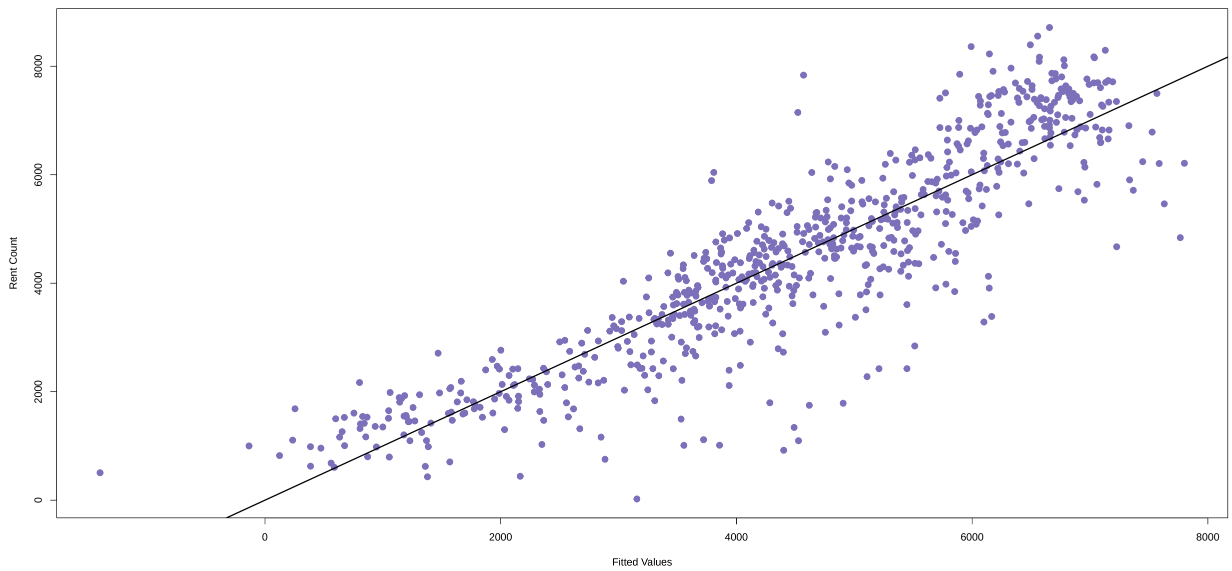
```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7.301e+04  2.647e+03 -27.585  < 2e-16 ***
## atemp        5.597e+03  3.608e+02  15.511  < 2e-16 ***
## dteday       4.948e+00  1.723e-01  28.712  < 2e-16 ***
## weathersit2 -3.976e+02  8.772e+01  -4.533 6.82e-06 ***
## weathersit3 -1.930e+03  2.246e+02  -8.591  < 2e-16 ***
## season2      9.341e+02  1.213e+02   7.703 4.45e-14 ***
## season3      2.592e+02  1.564e+02   1.657 0.098039 .
## season4      4.188e+02  1.108e+02   3.780 0.000170 ***
## hum         -1.676e+03  3.186e+02  -5.262 1.88e-07 ***
## windspeed   -2.550e+03  4.614e+02  -5.526 4.59e-08 ***
## holiday     -7.066e+02  2.047e+02  -3.451 0.000590 ***
## weekday1     2.115e+02  1.258e+02   1.681 0.093130 .
## weekday2     3.084e+02  1.229e+02   2.509 0.012334 *
## weekday3     3.781e+02  1.232e+02   3.070 0.002225 **
## weekday4     3.784e+02  1.231e+02   3.073 0.002201 **
## weekday5     4.287e+02  1.231e+02   3.482 0.000527 ***
## weekday6     4.234e+02  1.225e+02   3.457 0.000579 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 884.8 on 714 degrees of freedom
## Multiple R-squared:  0.7959, Adjusted R-squared:  0.7914
## F-statistic: 174.1 on 16 and 714 DF,  p-value: < 2.2e-16
```

Our Adjusted R-Squared is a bit lower at 0.7914, but all of our coefficients are statistically significant and there is no issues with multicolinearity.


Added−Variable Plots
cnt | others
atemp | others
dteday | others
weathersit2 | others
weathersit3 | others
season2 | others
season3 | others
season4 | others
hum | others
windspeed | others
holiday | others
weekday1 | others
weekday2 | others
weekday3 | others
weekday4 | others
weekday5 | others
weekday6 | others



No issues with our model plots.



Our fitted values are looking good too.

Reflection

```
##
## Call:
## lm(formula = cnt ~ atemp + dteday + weathersit + season + hum +
##     windspeed + holiday + weekday, data = bikes)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3481.1  -382.8    53.5   525.2  3266.5
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -7.301e+04  2.647e+03 -27.585  < 2e-16 ***
## atemp        5.597e+03  3.608e+02  15.511  < 2e-16 ***
## dteday       4.948e+00  1.723e-01  28.712  < 2e-16 ***
## weathersit2 -3.976e+02  8.772e+01  -4.533 6.82e-06 ***
## weathersit3 -1.930e+03  2.246e+02  -8.591  < 2e-16 ***
## season2      9.341e+02  1.213e+02   7.703 4.45e-14 ***
## season3      2.592e+02  1.564e+02   1.657 0.098039 .
## season4      4.188e+02  1.108e+02   3.780 0.000170 ***
## hum         -1.676e+03  3.186e+02  -5.262 1.88e-07 ***
## windspeed   -2.550e+03  4.614e+02  -5.526 4.59e-08 ***
## holiday     -7.066e+02  2.047e+02  -3.451 0.000590 ***
## weekday1     2.115e+02  1.258e+02   1.681 0.093130 .
## weekday2     3.084e+02  1.229e+02   2.509 0.012334 *
## weekday3     3.781e+02  1.232e+02   3.070 0.002225 **
## weekday4     3.784e+02  1.231e+02   3.073 0.002201 **
## weekday5     4.287e+02  1.231e+02   3.482 0.000527 ***
## weekday6     4.234e+02  1.225e+02   3.457 0.000579 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 884.8 on 714 degrees of freedom
## Multiple R-squared:  0.7959, Adjusted R-squared:  0.7914
## F-statistic: 174.1 on 16 and 714 DF,  p-value: < 2.2e-16
```

Our final model predicts the number of rented bikes based on the date, temperature feel, season, humidity, day of the week, and holiday status.

```
##   (Intercept)         atemp        dteday   weathersit2   weathersit3
## -73005.075592   5596.541431      4.947567   -397.647966  -1929.878683
##       season2       season3       season4           hum     windspeed
##    934.056020    259.153660    418.843813  -1676.437156  -2549.850027
##       holiday      weekday1      weekday2      weekday3      weekday4
##   -706.606796    211.513623    308.393894    378.087388    378.370054
##      weekday5      weekday6
##    428.734219    423.439988
```

Perceived temperature has the greatest influence on number of bikes rented: when we ran our regression subsets test, it was the very first variable chosen. A 1 degree (F) increase in the temperature correspond to 5,596 bikes rented!

Number of rented bikes also depends a fair amount on the day of the week: the farther into the week, the more predicted bikes rented.

The greatest negative influences on rental count are rain (weathersit3), high windspeed, high humidity, and holidays. These all make sense.

In all, 79% of the variation in total rented bike counts can be attributed to these variables. We can see that from our regressions subset test that a model with just the date, weather situation and temperature serves well enough, but if we have access to all of these variables, we might as well use them.

This project can prove helpful for cities that want to adopt a rental bike system and help plan accordingly for total number of stations and bikes needed. The D.C. bike company could also vary prices based on these variables to maximize profits or maximize number of bikes used for certain days or periods.

This model only predicts total number of bikes rented per day. Future analysis on hourly rental counts or location-based rental information could be interesting. It would also be interesting to create models for casual rentals and registered rentals. We assumed that there was no difference in rental preferences for the sake of this project.

Appendix: All code for this report

```r
knitr::opts_chunk$set(echo = FALSE)
bikes <- read.csv("day.csv", header = TRUE)
bikes$dteday <- as.Date(bikes$dteday)
bikes <- bikes[, -1]
summary(bikes)
bikes[, c(2, 4, 6, 8)] <- lapply(bikes[, c(2, 4, 6, 8)], factor)
library(ggplot2)
library(gridExtra)

# Create empty list to store plots
plots <- list()

# Loop through columns of bikes data frame
for (col in names(bikes)) {
  if (col %in% c("season", "yr", "mnth", "holiday", "weekday",
                 "workingday", "weathersit")) {
    p <- ggplot(bikes, aes(x = .data[[col]])) +
    geom_bar(fill = "#F26419") + theme_classic() + labs(x = col)
  }
  else  {
    p <- ggplot(bikes, aes(x = .data[[col]])) +
    geom_histogram(fill = "#F26419") + theme_classic() + labs(x = col)
  }
  plots[[col]] <- p
}

# Arrange plots in a grid
grid.arrange(grobs = plots, ncol = 3)
# Function to add correlation coefficients
panel.cor <- function(x, y, digits = 2, prefix = "", cex.cor, ...) {
    usr <- par("usr")
```

```r
    on.exit(par(usr))
    par(usr = c(0, 1, 0, 1))
    Cor <- abs(cor(x, y)) # Remove abs function if desired
    txt <- paste0(prefix, format(c(Cor, 0.123456789), digits = digits)[1])
    if(missing(cex.cor)) {
        cex.cor <- 0.4 / strwidth(txt)
    }
    text(0.5, 0.5, txt,
         cex = 2) # Resize the text by level of correlation
}

# Plotting the correlation matrix
pairs(bikes[, c(1, 4, 6, 9:15)],
      pch = 20,
      col = "#1F7A8C",
      gap = 0,
      upper.panel = panel.cor,    # Correlation panel
      lower.panel = panel.smooth) # Smoothed regression lines
summary(model <- lm(cnt ~ .  - yr - casual - registered
                    - workingday,
                    data = bikes))
par(mfrow=c(2,2))
plot(model, pch = 20, col = "#248232", lwd = 2)
library(car)
vif(model)
model <- lm(cnt ~ . -yr - casual - registered - workingday
            - mnth - temp, data = bikes)
# from Miles Chen 101A Nov 20 2015 youtube lecture
attach(bikes)
X <- cbind(dteday, season, holiday, weekday, weathersit,
           atemp, hum, windspeed)
library(leaps)
```

```r
b <- regsubsets(as.matrix(X), cnt)
rs <- summary(b)
rs
detach(bikes)
plot(1:8, rs$adjr2, xlab = "number of features",
     main = "exhausted algorithm selection results",
     pch = 15, cex = 4, col = "#01BAEF")
backwardAIC <- step(model, direction = "backward")
mint = lm(cnt ~ 1, data = bikes)
forwardAIC <- step(mint, scope = list(upper = model,
                                      lower = ~1),
                   direction = "forward")
summary(forwardAIC)
library(car)
avPlots(forwardAIC, pch = 20, col = "#000022", col.lines = "#D81159")
par(mfrow=c(2,2))
plot(forwardAIC, pch = 20, col = "#248232", lwd = 2)
plot(forwardAIC$fitted, bikes$cnt, xlab = "Fitted Values", ylab = "Rent Count",
     pch = 20, col = "#7D70BA", cex = 2)
abline(lsfit(model$fitted.values, bikes$cnt), col = "#040404", lwd = 2)
summary(model <- forwardAIC)
model$coefficients
```