Athens University of Economics and Business
Department of Management Science and Technology

Statistics for BA I
Professor: I. Ntzoufras

Bike Sharing Dataset Assignment

Marianna Konstantopoulou
A.M: P2822122
Dataset: bike_40.csv

M.Sc. Business Analytics
Part Time 2021-2023

Athens, 06/01/2022

Table of Contents

Chapter 1: Introduction - description of the problem, data, aim, background information……..……. iii

Chapter 2: Descriptive analysis and exploratory data analysis……..…… iv

Chapter 3: Pairwise comparisons……...……vi

Chapter 4: Predictive models……..……...ix

Chapter 5: Further analysis…….xiii

Chapter 6: Conclusions and Discussion…….xvii

Appendix……..….xviii

Chapter 1

Introduction
- description of the problem, data, aim, background information

The Bike sharing system is an innovative bicycle rental system that automates the entire process from membership to rental and return. With these systems, users can easily rent a bike from one location and return it to another. These types of programs have been established all around the world and consist of more than 500,000 bicycles. Due to their importance in transportation, environmental and health issues, there is currently great interest in these systems. There is a great interest coming from the research community regarding the bicycle sharing systems due to the generated data. In contrast to other transportation services such as buses and subways, these systems explicitly record travel times, departures and arrivals. This makes the bike sharing system a virtual sensor network that can be used to record urban mobility. Therefore, by monitoring this data, it is expected that most of the major events in the city can be detected. Our goal is to figure out what impacts the hourly bike rentals and additionally predict it in order to satisfy the increasing demand.

To achieve our goal we need to process data extracted by the Capital Bikeshare system, Washington D.C., USA, which are available in http://capitalbikeshare.com/system-data. The dataset includes records for years 2011 and 2012 and it contains weather conditions, seasons, day of the week, holidays etc. , factors that can affect the rental behaviors. Weather information contains hourly data and was extracted from http://www.freemeteo.com. Our results will be valuable to bike sharing companies so they can increase the bike share activity in urban centers.

Chapter 2

Descriptive analysis and exploratory data analysis

To start our analysis, we will load our data in R and check the data types. We will remove some columns that are not useful for our research: instant, X (both fields do not provide us with important information, they are indexes), dteday (which provides the same information with the separate year, month and weekday fields) and lastly casual and registered need to be removed as they are the sum of cnt (count) field. We need to change the rest of the variables' data types to the correct type. Temperature (in Celsius), feeling temperature (in Celsius), humidity, windspeed and count are converted to numeric while season, year, month, hour, holiday, weekday, working day and weather situation converted to categorical (i.e. Table 1). We need to make an extra conversion for the numeric variables, except for count. The normalized figures were extracted, so we can convert them to their usual units of measurement so it's easier for us to understand. We multiply the temperature with 41, the feeling temperature with 50, the humidity with 100 and the windspeed with 67. It is also important to perform a check for potential missing values. In our case there are no missing values in our data set.

Checking the describe output (i.e. Table 2) and histograms (i.e. Figure 1) for our numeric variables we can make some quick observations:

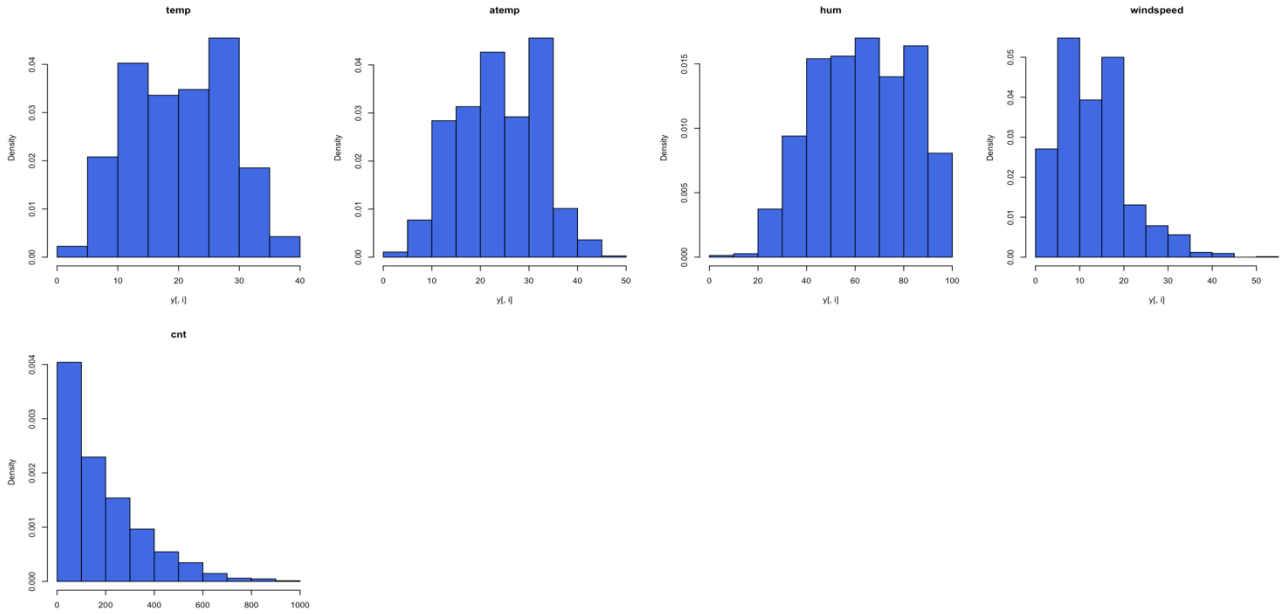


Figure 1: Histograms of numeric variables

- The average temperature is 20.3°C while the minimum and maximum temperatures are 0.8°C and 39.36°C
- The feeling temperature has an average of approximately 24°C peaking at 45.45°C and the average humidity is 63.36%
- The windspeed has a maximum of 54kph and a minimum of 0kph while the average is 12.66kph
- The average count of total rental bikes is 183.21 with a maximum 963 and a minimum of 1

Checking the bar plots (i.e. Figure 2) for our categorical variables we can make the following observations:

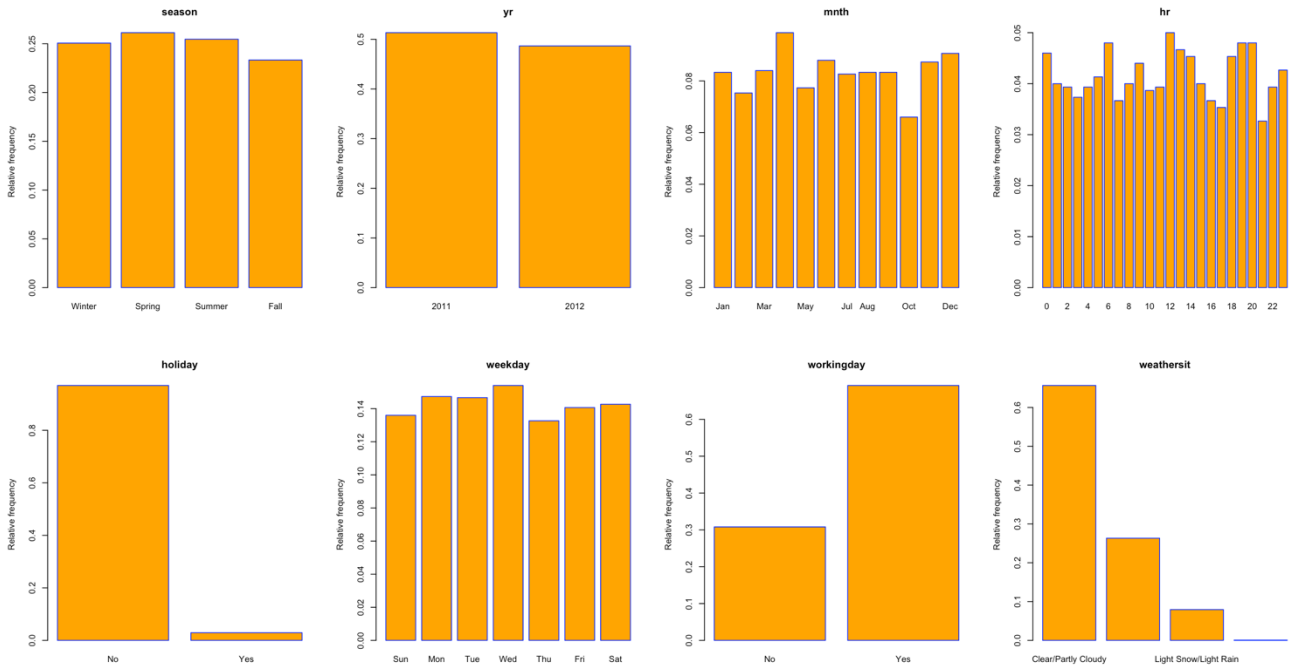


Figure 2: Bar plots for factor variables

- The most observations are in season spring and in the year 2011
- The month with the most observations is April and the hour with the most records is 12:00
- The majority of observations is during working days and not on holidays
- Wednesdays are the days with the most records and the weather with the majority of observations is Clear/Partly Cloudy

Chapter 3

Pairwise comparisons

After studying our variables separately, the next step would be to study them in pairs, so we can figure out their relationships. The relationships that are interesting to investigate are all our variables with the count variable in order to understand how these factors affect the total rental bikes as well as the association of the numeric variables with each other.

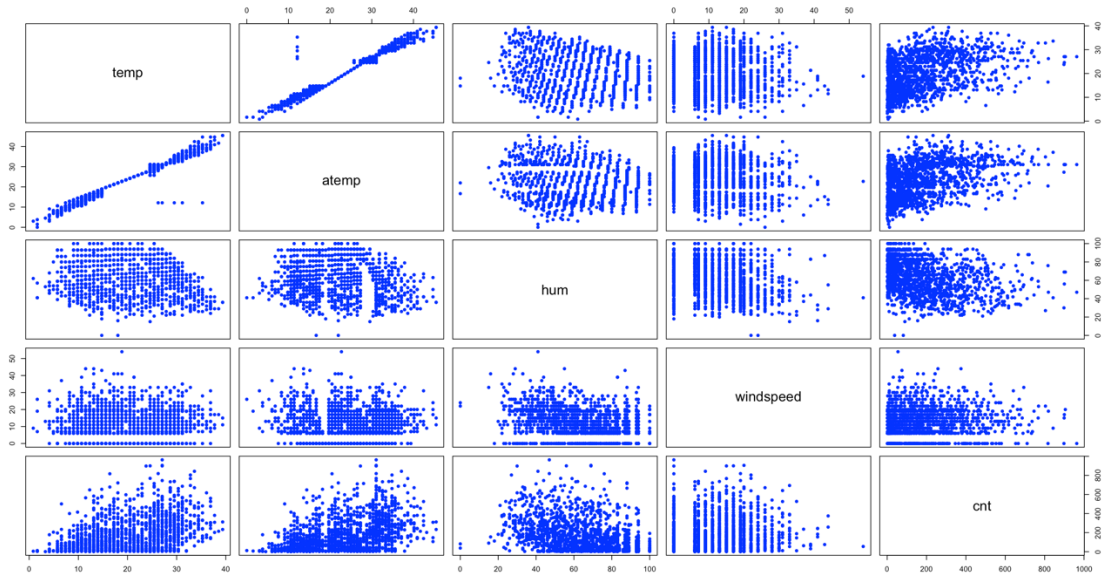


Figure 3: Scatter plots for pairwise comparisons of numeric variables

First, we can take a look at the pairwise comparisons of our numeric variables (i.e. Figure 3) to check if there are any strong relationships. We can see that temperature and feeling temperature have a strong association (linear relationship). We can't make clear assumptions for the rest of coefficients associations by looking at this plot.

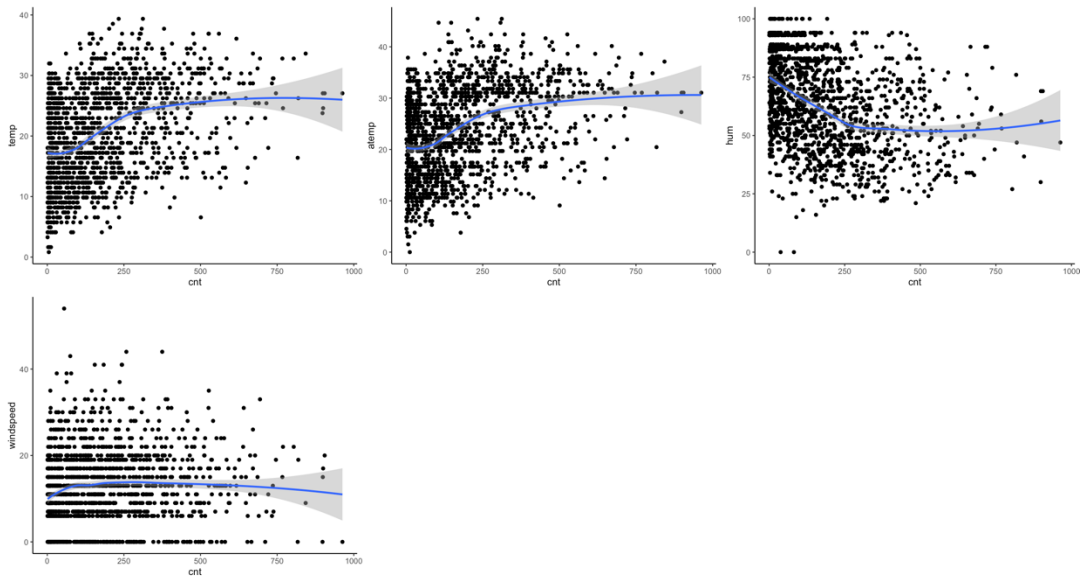


Figure 4: Scatter plots for pairwise comparisons of numeric variables with count variable

Judging from the scatterplots (i.e. Figure 4) we don't see any strong association between total rental bikes and temperature, feeling temperature, humidity and windspeed.

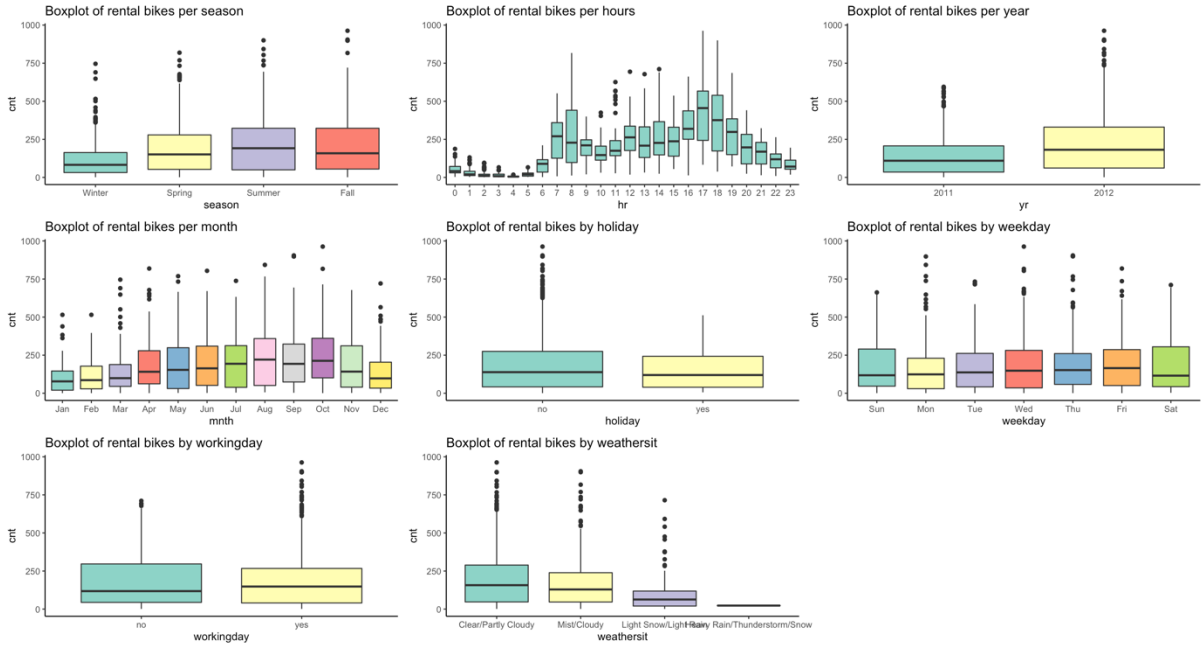


Figure 5: Box plots for pairwise comparisons of factor variables with count variable

We can notice (i.e. Figure 5) that the total bikes rentals are higher during summer and fall, the busiest hours are the commute hours like 8 in the morning and 5 in the evening, more rentals were recorded in 2012 and the most popular months, regarding the rentals, are August and October. Most rentals occur during working days and not on holidays while people tend to rent more when the weather is Clear or Partly Cloudy.

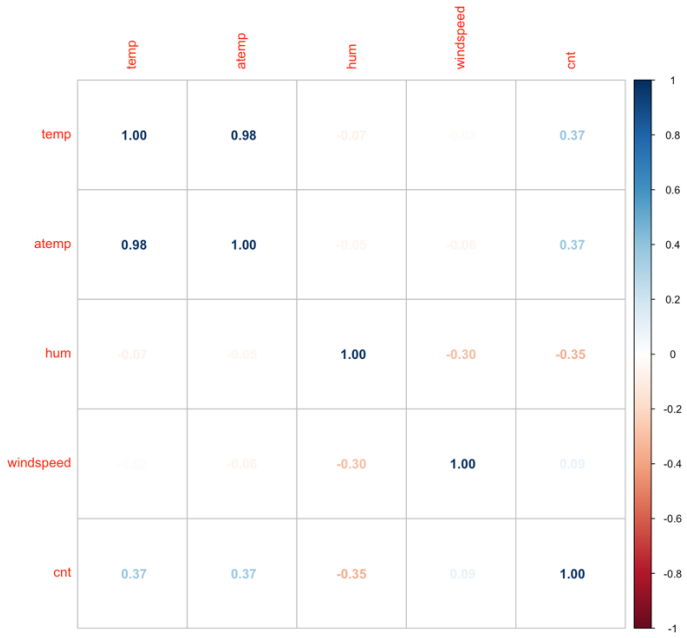


Figure 6: Correlation plot for numeric variables

Additionally, we can also use the correlation plot to check the correlation between the pairs of numeric variables. (i.e. Figure 6). Temperature and feeling temperature have a strong connection of 0.98 while the rest of numeric variables have a rather weak correlation. The strong connection between temperature and feeling temperature observed in the correlation plot complies with the linear relationship assumption we made from the scatter plot (i.e. Figure 3) earlier. The Pearson's correlation coefficient $\rho = 0.98 > 0.7$ confirms a strong positive linear relationship and indicates a multicollinearity issue between these two variables.

Chapter 4

Predictive models

Since we studied the relationships between our variables it's time to try and construct some models so that we can create a model for predicting the number of bike rentals per hour.

The full model with count as a response and all the other variables as predictors has R^2 adjusted 67%, which means that 67% of the variability is explained by the model and residual standard error 98.97 (i.e. Table 3). However, this is not a clear indication of a good fit (since sometimes a high R^2 can be misleading). The significant coefficients are season, year, hour, holiday, weekday, weather situation, humidity, windspeed and the Intercept is significant too.

We will conduct LASSO as a variable selection technique. We use cross validation and select the largest value of lambda such that error is within 1 standard error of the minimum (i.e. Figure 7). Using the lambda we selected, we receive the estimated coefficients under the lambda.1se. Our model has count as a response and as predictors all variables and intercept except for working day with R^2 adjusted and residual standard error same as the full model (i.e. Table 4).

To select our final model, we are using stepwise methods as well. We used the Stepwise procedure according to AIC (since we need that for predicting). Our final model has count as response and the predictors are season, year, month, hour, holiday, weekday, weather situation, feeling temperature, humidity and windspeed. The R^2 adjusted is still 67% and residual standard error is 99 (i.e. Table 5).

As the final model is selected, we need to check the assumptions. Particularly, we must check normality, constant variance, independence, and linearity. While conducting tests for the assumptions, they are all rejected except for independence (Shapiro-Wilk $p < .001$, KS $p < .001$, Non-constant Variance Score Test $p < .001$, Tukey's test for nonadditivity $p < .001$ and Runs Test $p = .30 > .05$) (i.e. Tables 6, 7, 8, 9 and Figures 8, 9, 10). Since the assumptions are not fulfilled, we need to do some transformations. We can start with fixing the linearity by adding log to our response and adding polynomial terms for humidity and feeling temperature until the linearity assumption is not rejected (Tukey Test $p= .80 > .05$) (i.e. Table 10). When checking homoscedasticity for this model we can see that is still rejected, so we need to make further changes. The method that worked for our model was the weighted least squares method, which can be used when the ordinary least squares assumption of constant variance in the errors is violated. This method places weights on the observations such that those with small error variance are given more weight since they contain more information compared to observations with larger error variance. We will use a weight of $1/ (\text{fitted values})^2$. When adding weight to our model the homoscedasticity is not rejected anymore (Non-constant Variance Score Test $p= .15 > .05$) (i.e. Table 11). Our transformations didn't change the independence, so it is still not rejected. (i.e. Table 12) Unfortunately normality was not fixed by any of the transformations so it might affect our predicting ability (i.e. Table 13).

This is the final model (i.e. Table 14) we selected:

$\log(cnt) = 2.083 + 0.2453 \times seasonSpring + 0.4485 \times seasonSummer + 0.6315 \times seasonFall + 0.466 \times yr2012 + 0.0659 \times mnthFeb + 0.0936 \times mnthMar + 0.1 \times mnthApr + 0.1393 \times mnthMay + 0.1034 \times mnthJun - 0.001 \times mnthJul - 0.035 \times mnthAug + 0.0045 \times mnthSep - 0.0588 \times mnthOct - 0.172 \times mnthNov - 0.139 \times mnthDec - 0.594 \times hr1 - 1.18 \times hr2 - 1.459 \times hr3 - 1.973 \times hr4 - 0.782 \times hr5 + 0.568 \times hr6 + 1.749 \times hr7 + 1.799 \times hr8 + 1.570 \times hr9 + 1.236 \times hr10 + 1.361 \times hr11 + 1.6 \times hr12 + 1.617 \times hr13 + 1.578 \times hr14 + 1.570 \times hr15 + 1.828 \times hr16 + 2.195 \times hr17 + 2.09 \times hr18 + 1.921 \times hr19 + 1.531 \times hr20 + 1.268 \times hr21 + 0.9935 \times hr22 + 0.6648 \times hr23 - 0.2012 \times holidayYes + 0.01917 \times weekdayMon + 0.0275 \times weekdayTue + 0.0624 \times weekdayWed + 0.1026 \times weekdayThu + 0.1399 \times weekdayFri + 0.1756 \times weekdaySat - 0.0611 \times weathersitMist/Cloudy - 0.0609 \times weathersitLight Snow/Light Rain + 0.7455 \times weathersitHeavy Rain/ThunderStorm/Snow + 0.0114 \times atemp + 0.02 \times hum - 0.00326 \times windspeed - 0.000294 \times hum^2 + 0.000009 \times hum^3 + 0.00224 \times atemp^2 - 0.00005 \times atemp^3 + \varepsilon$

since normality assumption is violated, we can't assume that residuals follow the normal distribution. The final R^2 adjusted of our model is 80.15%, that means that 80.15% of the variability is explained by the model and the residual standard error is 1245.

We will start to interpret our final model:

- Our intercept means that if season is Winter, year is 2011, month is January, hour is 00:00, there is no holiday, it's a Sunday, the weather is Clear/Partly Cloudy and our numeric variables atemp, humidity and windspeed are 0 then the total number of rentals are equal to $e^{2.083}$ which is almost equal to 8.
- The coefficient for windspeed -0.00326 means that if we compare two days with the same characteristics which differ only by 1 kph then the expected difference in the total rentals is equal to $(e^{-0.00326} - 1) \times 100$ which is approximately 0.32% decrease.
- The coefficient for season Spring 0.2453 means that if we compare two days with the same characteristics (for factors year 2011, month January, hour 00:00, it is not a holiday, Sunday and the weather is clear) which differ only by the season, where in this case one of them has season Winter and the other one has season Spring then the expected difference in the total rentals is equal to $(e^{0.2453} - 1) \times 100$ which is approximately 27.8% increase.
- The coefficient for season Summer 0.4485 means that if we compare two days with the same characteristics (for factors year 2011, month January, hour 00:00, it is not a holiday, Sunday and the weather is clear) which differ only by the season, where in this case one of them has season Winter and the other one has season Summer then the expected difference in the total rentals is equal to $(e^{0.4485} - 1) \times 100$ which is approximately 56.59% increase.
- The coefficient for season Fall 0.6315 means that if we compare two days with the same characteristics (for factors year 2011, month January, hour 00:00, it is not a holiday, Sunday and the weather is clear) which differ only by the season, where in this case one of them has season Winter and the other one has season Fall then the expected difference in the total rentals is equal to $(e^{0.6315} - 1) \times 100$ which is approximately 88.04% increase.
- The coefficient for year 2012 0.466 means that if we compare two days with the same characteristics (for factors season Winter, month January, hour 00:00, it is not a holiday, Sunday and the weather is clear)

which differ only by the year where in this case one of them has year 2011 and the other one has year 2012 then the expected difference in the total rentals is equal to ($e^{0.466}$– 1) x 100 which is approximately 59.36% increase.

- The coefficient for month February 0.0659 means that if we compare two days with the same characteristics (for factors year 2011, season Winter, hour 00:00, it is not a holiday, Sunday and the weather is clear) which differ only by the month where in this case one of them has month January and the other one has month February then the expected difference in the total rentals is equal to ($e^{0.0659}$– 1) x 100 which is approximately 6.81% increase.
- The coefficient for month March 0.0936 means that if we compare two days with the same characteristics (for factors year 2011, season Winter, hour 00:00, it is not a holiday, Sunday and the weather is clear) which differ only by the month where in this case one of them has month January and the other one has month March then the expected difference in the total rentals is equal to ($e^{0.0936}$– 1) x 100 which is approximately 9.81% increase. We would follow the same procedure to interpret the rest of the months.
- The coefficient for hour 01:00 -0.594 means that if we compare two days with the same characteristics (for factors year 2011, season Winter, month January, it is not a holiday, Sunday and the weather is clear) which differ only by the hour where in this case one of them has hour 00:00 and the other one has hour 01:00 then the expected difference in the total rentals is equal to ($e^{-0.594}$– 1) x 100 which is approximately 44.7% decrease. We would follow the same procedure to interpret the rest of the hours.
- The coefficient for holiday -0.2012 means that if we compare two days with the same characteristics (for factors year 2011, season Winter, month January, hour 00:00, Sunday and the weather is clear) which differ since one of them is not a holiday and the other one is, then the expected difference in the total rentals is equal to ($e^{-0.2012}$– 1) x 100 which is approximately 18.22% decrease.
- The coefficient for Monday 0.01917 means that if we compare two days with the same characteristics (for factors year 2011, season Winter, month January, hour 00:00, it is not a holiday and the weather is clear) where the one of them is a Sunday and the other one is a Monday, then the expected difference in the total rentals is equal to ($e^{0.01917}$– 1) x 100 which is approximately 1.93% increase.
- The coefficient for Tuesday 0.0275 means that if we compare two days with the same characteristics (for factors year 2011, season Winter, month January, hour 00:00, it is not a holiday and the weather is clear) where the one of them is a Sunday and the other one is a Tuesday, then the expected difference in the total rentals is equal to ($e^{0.0275}$– 1) x 100 which is approximately 2.78% increase. We would follow the same procedure to interpret the rest of the weekdays.
- The coefficient for Misty/Cloudy weather -0.0611 means that if we compare two days with the same characteristics (for factors year 2011, season Winter, month January, hour 00:00, it is not a holiday, Sunday) where the one of them has Clear/Partly Cloudy and the other one has Misty/Cloudy weather, then the expected difference in the total rentals is equal to ($e^{-0.0611}$– 1) x 100 which is approximately 5.9% decrease.
- The coefficient for Light Snow/Light Rain weather -0.0609 means that if we compare two days with the same characteristics (for factors year 2011, season Winter, month January, hour 00:00, it is not a holiday, Sunday) where the one of them has Clear/Partly Cloudy and the other one has Light Snow/Light Rain

weather, then the expected difference in the total rentals is equal to ($e^{-0.0609}$ − 1) x 100 which is approximately 5.9% decrease.

- The coefficient for Heavy Rain/ThunderStorm/Snow weather 0.7455 means that if we compare two days with the same characteristics (for factors year 2011, season Winter, month January, hour 00:00, it is not a holiday, Sunday) where the one of them has Clear/Partly Cloudy and the other one has Heavy Rain/ThunderStorm/Snow weather, then the expected difference in the total rentals is equal to ($e^{0.7455}$ − 1) x 100 which is approximately 110.7% increase (this seems like a rather abnormal observation but it could have happened in our dataset by chance).

It is not easy to interpret feeling temperature and humidity coefficients since the interpretation of polynomial factors is rather difficult. The thing we can observe is that coefficients for the polynomial terms are close to 0 so maybe for a small increase of 1% of humidity or 1 degree Celsius for feeling temperature we could ignore the polynomial terms and say the following:

- The coefficient for feeling temperature 0.0114 means that if we compare two days with the same characteristics which differ only by 1 degree Celsius then the expected difference in the total rentals is equal to ($e^{0.0114}$ − 1) x 100 which is approximately 1.14% increase.
- The coefficient for humidity 0.02 means that if we compare two days with the same characteristics which differ only by 1 percent humidity then the expected difference in the total rentals is equal to ($e^{0.02}$ − 1) x 100 which is approximately 2.02% increase.

For the error, we can say that the predicted value with these covariates is 1245. Which means that the error in the estimated log(count) will be ±2x1245=±2490 proportional change around the expected/fitted value.

Last step would be to interpret the predicting performance of our final model. We can assume we have a good fit due to R^2 adjusted which was 80.15%. In order to assess our model's predicting performance we will compare the models' predictions to the actual values of the response variable (and we will do the same for the null and full model as well) and we will use the root mean square error (RMSE) to compare our models. After calculating the RMSE for our models we can see that our final model has RMSE 85 which means that our model is off by approximately 85 total bikes rentals. RMSE for the full model is 97.2 and for the null model it is 171. Judging from the RMSE we can assume that our final model has a better predictive performance than the full or null models.

Chapter 5

Further analysis

The next step for our analysis is to use the test dataset with 500 observations to assess the out-of-sample predictive ability of the models coming from LASSO and the stepwise method as well as the full and null models. In order to be able to compare the models we will calculate the root mean square error of each model (RMSE). The results of RMSE for our models are: LASSO model has RMSE 96.2, Stepwise method model has RMSE 98.19, full model has RMSE 96.2 and null model has RMSE 182.6. We can assume that LASSO model and full model have the best out-of-sample predictive ability.

To complete our analysis, we will describe a typical day for every season in our dataset.

Winter

Figure 11: Frequency bar-plot for weather during Winter season

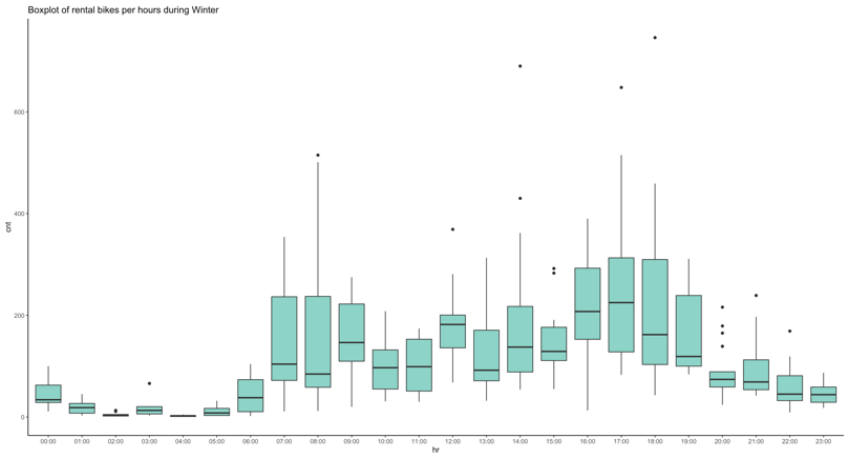


Figure 12: Boxplot of total rentals per hours during Winter season

A typical winter day (i.e. Table 15) has temperature of 12.18 degrees Celsius and the feeling temperature is 14.76 degrees. The humidity is 58.51 and the windspeed is 14.59 kph. The weather is usually clear (i.e. Figure 11) and the busiest hour of a typical day is 17:00 (i.e. Figure 12). The number of casual users is 15.36 and the registered users are 97.66 which means that the typical day of winter has 113.02 total bike rentals.

Spring

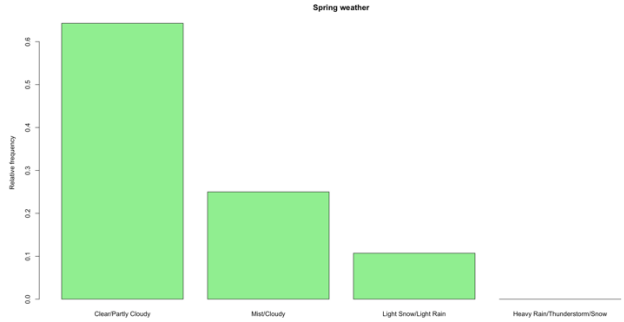


Figure 13: Frequency bar-plot for weather during Spring season

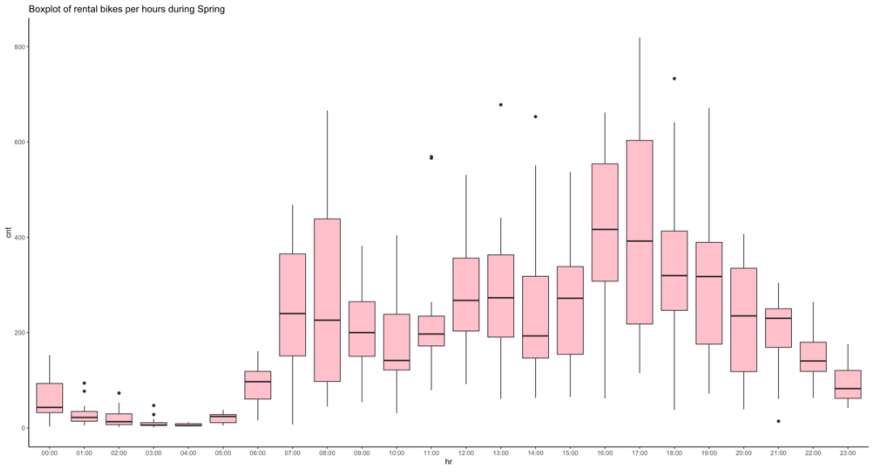


Figure 14: Boxplot of total rentals per hours during Spring season

A typical spring day (i.e. Table 16) has temperature of 22.11 degrees Celsius and the feeling temperature is 25.8 degrees. The humidity is 63.68 and the windspeed is 13.84 kph. The weather is usually clear (i.e. Figure 13) and the busiest hour of a typical day is 17:00 (i.e. Figure 14). The number of casual users is 42.67 and the registered users are 147.61 which means that the typical day of spring has 190.28 total bike rentals.

Summer

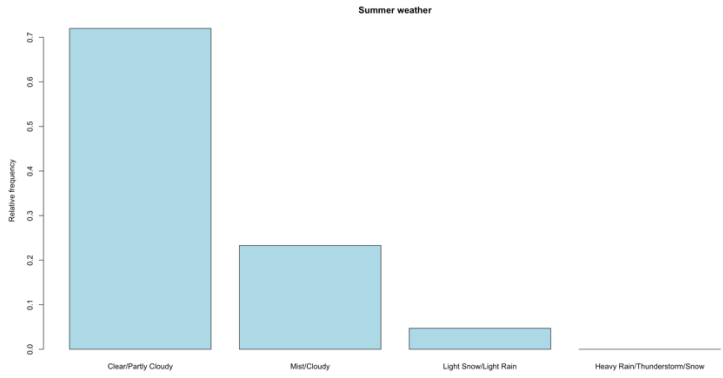


Figure 15: Frequency bar-plot for weather during Summer season

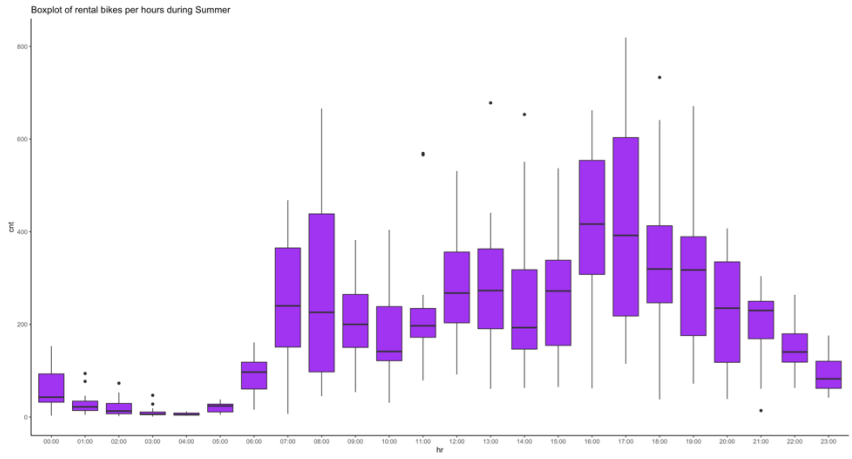


Figure 16: Boxplot of total rentals per hours during Summer season

A typical summer day (i.e. Table 17) has temperature of 28.96 degrees Celsius and the feeling temperature is 32.68 degrees. The humidity is 64.42 and the windspeed is 11.02 kph. The weather is usually clear (i.e. Figure 15) and the busiest hour of a typical day is 17:00 (i.e. Figure 16). The number of casual users is 47.31 and the registered users are 174.62 which means that the typical day of summer has 221.93 total bike rentals.

Fall

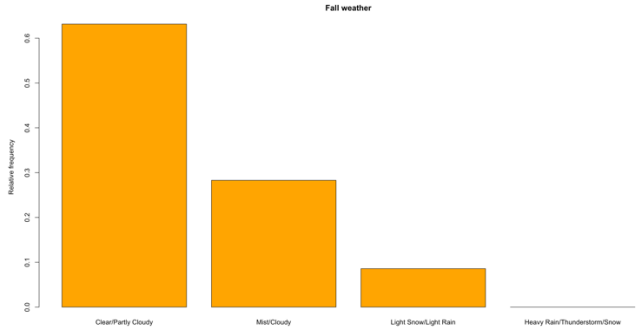


Figure 17: Frequency bar-plot for weather during Fall season

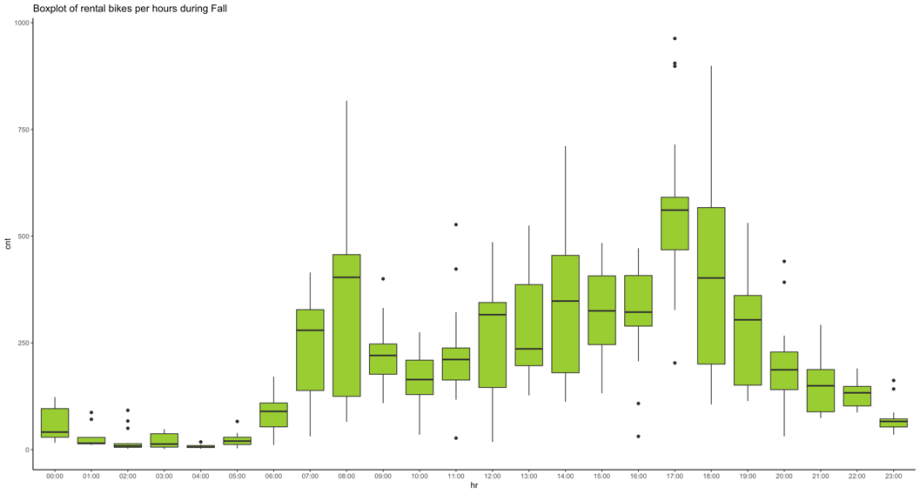


Figure 18: Boxplot of total rentals per hours during Fall season

A typical fall day (i.e. Table 18) has temperature of 17.56 degrees Celsius and the feeling temperature is 21.05 degrees. The humidity is 67.07 and the windspeed is 11.05 kph. The weather is usually clear (i.e. Figure 17) and the busiest hour of a typical day is again 17:00 (i.e. Figure 18). The number of casual users is 33.04 and the registered users are 175.40 which means that the typical day of fall has 208.44 total bike rentals.

Chapter 6

Conclusions and Discussion

The main aim of this assignment was to identify the best model for predicting the number of bike rentals per hour. The model we selected has good fit (R^2 adjusted = 80.15%) and it fulfills 3 out of 4 assumptions, specifically homoscedasticity, linearity and independence. We couldn't satisfy normality for our final model, so this might affect the future results of our analysis especially the predicting ability.

From the interpretation of our model, we can understand that during commute hours we observe a higher number of rentals and during warmer months there is a clear rise of users. Wind also clearly affects the rentals since the higher the windspeed the smaller the number of bikes is rented. In addition, people mostly use the bikeshare program during the working days and not on holidays.

Lastly, it would be very interesting to be able to analyze most recent data for the bike sharing systems since we can observe that the latest years more and more cities in the world have already deployed or are planning to implement bike sharing programs. This popularity can be mainly explained by the fact that bike sharing systems are associated with various social, environmental, and economic benefits, such as a decrease in carbon dioxide (CO^2) emissions, a reduction in various diseases (e.g., diabetes and obesity), and a decline in traffic congestion and noise pollution through the provision of alternatives to auto-commuting and an increase in public transit use (Caulfield et al., 2017; Martens, 2007; Mont, 2004).

APPENDIX

Tables

TABLE 1. Structure of our final data set

```
'data.frame':   1500 obs. of  13 variables:
 $ season    : Factor w/ 4 levels "Winter","Spring",..: 3 1 3 1 4 1 4 1 2 2 ...
 $ yr        : Factor w/ 2 levels "2011","2012": 2 1 1 2 2 2 2 2 2 2 ...
 $ mnth      : Factor w/ 12 levels "Jan","Feb","Mar",..: 8 1 7 1 11 12 10 1 5 5 ...
 $ hr        : Factor w/ 24 levels "0","1","2","3",..: 1 18 14 3 9 18 10 7 14 7 ...
 $ holiday   : Factor w/ 2 levels "No","Yes": 1 1 1 1 1 1 1 1 1 1 ...
 $ weekday   : Factor w/ 7 levels "Sun","Mon","Tue",..: 2 7 2 2 7 5 1 3 6 7 ...
 $ workingday: Factor w/ 2 levels "No","Yes": 2 1 2 2 1 2 1 2 2 1 ...
 $ weathersit: Factor w/ 4 levels "Clear/Partly Cloudy",..: 2 2 1 1 2 2 1 1 1 1 ...
 $ temp      : num  29.52 13.12 35.26 9.84 13.94 ...
 $ atemp     : num  34.8 15.2 40.9 12.1 15.2 ...
 $ hum       : num  79 36 50 56 46 52 55 59 26 88 ...
 $ windspeed : num  0 19 17 9 20 ...
 $ cnt       : num  33 83 141 5 142 257 269 85 363 31 ...
```

Final data types of our variables. There are 8 categorical variables (season, year, month, hour, holiday, weekday, working day, weather situation and 5 numeric variables temperature, feeling temperature, humidity, windspeed and count.

TABLE 2. Describe table for numeric variables

	temp	atemp	hum	windspeed	cnt
vars	1.00	2.00	3.00	4.00	5.00
n	1500.00	1500.00	1500.00	1500.00	1500.00
mean	20.30	23.68	63.36	12.66	183.21
sd	7.87	8.58	19.48	8.35	171.49
median	20.50	24.24	64.00	13.00	138.00
trimmed	20.31	23.75	63.73	12.21	158.10
mad	9.73	10.11	23.72	8.89	155.67
min	0.82	0.00	0.00	0.00	1.00
max	39.36	45.45	100.00	54.00	963.00
range	38.54	45.45	100.00	54.00	962.00
skew	-0.01	-0.08	-0.14	0.58	1.28
kurtosis	-0.94	-0.84	-0.86	0.58	1.55
se	0.20	0.22	0.50	0.22	4.43

Descriptive measures for numeric variables temperature, feeling temperature, humidity, windspeed and count.

TABLE 3. Summary of the full model

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

Residuals:
    Min      1Q  Median      3Q     Max
-291.41  -59.02   -8.22   45.33  429.57

Coefficients: (1 not defined because of singularities)
                                          Estimate Std. Error t value Pr(>|t|)
(Intercept)                               -68.8437    22.1824  -3.104 0.001949 **
seasonSpring                               22.0566    16.6467   1.325 0.185386
seasonSummer                               33.7983    18.9899   1.780 0.075318 .
seasonFall                                 95.3795    15.8401   6.021 2.19e-09 ***
yr2012                                     81.2713     5.2574  15.458  < 2e-16 ***
mnthFeb                                     4.0060    13.2073   0.303 0.761693
mnthMar                                     4.7997    14.4098   0.333 0.739115
mnthApr                                    22.2268    22.1419   1.004 0.315627
mnthMay                                    26.2128    24.0679   1.089 0.276283
mnthJun                                    10.1778    24.2705   0.419 0.675024
mnthJul                                   -18.5458    27.1890  -0.682 0.495281
mnthAug                                     6.4954    26.3517   0.246 0.805339
mnthSep                                    20.5796    23.1537   0.889 0.374244
mnthOct                                    -0.4392    21.7079  -0.020 0.983860
mnthNov                                   -29.2531    20.3627  -1.437 0.151047
mnthDec                                   -25.9345    15.8182  -1.640 0.101318
hr1                                        -3.1983    17.6704  -0.181 0.856394
hr2                                       -20.3719    17.7439  -1.148 0.251115
hr3                                       -28.2405    18.0077  -1.568 0.117041
hr4                                       -12.9943    17.7830  -0.731 0.465071
hr5                                        -7.8794    17.6562  -0.446 0.655471
hr6                                        63.0591    16.9022   3.731 0.000198 ***
hr7                                       237.8145    18.1170  13.127  < 2e-16 ***
hr8                                       270.4572    17.6621  15.313  < 2e-16 ***
hr9                                       166.1755    17.2813   9.616  < 2e-16 ***
hr10                                      113.6829    17.8986   6.351 2.85e-10 ***
hr11                                      140.6149    18.0235   7.802 1.16e-14 ***
hr12                                      190.5565    17.1205  11.130  < 2e-16 ***
hr13                                      186.8049    17.4863  10.683  < 2e-16 ***
hr14                                      182.4562    17.9754  10.150  < 2e-16 ***
hr15                                      176.1040    18.3075   9.619  < 2e-16 ***
hr16                                      239.5324    18.8357  12.717  < 2e-16 ***
hr17                                      364.7762    18.7281  19.478  < 2e-16 ***
hr18                                      322.7247    17.6032  18.333  < 2e-16 ***
hr19                                      255.0724    17.0382  14.971  < 2e-16 ***
hr20                                      165.5845    17.0519   9.711  < 2e-16 ***
hr21                                      118.7447    18.7975   6.317 3.54e-10 ***
hr22                                       88.5090    17.7713   4.980 7.11e-07 ***
hr23                                       48.4402    17.3558   2.791 0.005323 **
holidayYes                                -34.5900    16.1397  -2.143 0.032267 *
weekdayMon                                 -6.1089     9.9982  -0.611 0.541295
weekdayTue                                 -5.1799     9.8042  -0.528 0.597347
weekdayWed                                  8.1281     9.7111   0.837 0.402736
weekdayThu                                  7.0418     9.9966   0.704 0.481284
weekdayFri                                  3.5115     9.8830   0.355 0.722412
weekdaySat                                 25.5919     9.8263   2.604 0.009297 **
workingdayYes                                  NA         NA      NA       NA
weathersitMist/Cloudy                      -7.2428     6.3984  -1.132 0.257832
weathersitLight Snow/Light Rain           -63.1126    11.3221  -5.574 2.96e-08 ***
weathersitHeavy Rain/Thunderstorm/Snow     48.7557   100.5912   0.485 0.627969
temp                                        2.7097     1.9793   1.369 0.171193
atemp                                       2.3895     1.6384   1.458 0.144935
hum                                        -0.9538     0.1897  -5.027 5.60e-07 ***
windspeed                                  -0.8004     0.3507  -2.282 0.022610 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 98.97 on 1447 degrees of freedom
Multiple R-squared:  0.6785,    Adjusted R-squared:  0.6669
F-statistic: 58.72 on 52 and 1447 DF,  p-value: < 2.2e-16
```

TABLE 4. Summary of the model selected with LASSO

```
Call:
lm(formula = cnt ~ . - workingday, data = Bikes)

Residuals:
    Min      1Q  Median      3Q     Max
-291.41  -59.02   -8.22   45.33  429.57

Coefficients:
                                       Estimate Std. Error t value Pr(>|t|)
(Intercept)                            -68.8437    22.1824  -3.104 0.001949 **
seasonSpring                            22.0566    16.6467   1.325 0.185386
seasonSummer                            33.7983    18.9899   1.780 0.075318 .
seasonFall                              95.3795    15.8401   6.021 2.19e-09 ***
yr2012                                  81.2713     5.2574  15.458  < 2e-16 ***
mnthFeb                                  4.0060    13.2073   0.303 0.761693
mnthMar                                  4.7997    14.4098   0.333 0.739115
mnthApr                                 22.2268    22.1419   1.004 0.315627
mnthMay                                 26.2128    24.0679   1.089 0.276283
mnthJun                                 10.1778    24.2705   0.419 0.675024
mnthJul                                -18.5458    27.1890  -0.682 0.495281
mnthAug                                  6.4954    26.3517   0.246 0.805339
mnthSep                                 20.5796    23.1537   0.889 0.374244
mnthOct                                 -0.4392    21.7079  -0.020 0.983860
mnthNov                                -29.2531    20.3627  -1.437 0.151047
mnthDec                                -25.9345    15.8182  -1.640 0.101318
hr1                                     -3.1983    17.6704  -0.181 0.856394
hr2                                    -20.3719    17.7439  -1.148 0.251115
hr3                                    -28.2405    18.0077  -1.568 0.117041
hr4                                    -12.9943    17.7830  -0.731 0.465071
hr5                                     -7.8794    17.6562  -0.446 0.655471
hr6                                     63.0591    16.9022   3.731 0.000198 ***
hr7                                    237.8145    18.1170  13.127  < 2e-16 ***
hr8                                    270.4572    17.6621  15.313  < 2e-16 ***
hr9                                    166.1755    17.2813   9.616  < 2e-16 ***
hr10                                   113.6829    17.8986   6.351 2.85e-10 ***
hr11                                   140.6149    18.0235   7.802 1.16e-14 ***
hr12                                   190.5565    17.1205  11.130  < 2e-16 ***
hr13                                   186.8049    17.4863  10.683  < 2e-16 ***
hr14                                   182.4562    17.9754  10.150  < 2e-16 ***
hr15                                   176.1040    18.3075   9.619  < 2e-16 ***
hr16                                   239.5324    18.8357  12.717  < 2e-16 ***
hr17                                   364.7762    18.7281  19.478  < 2e-16 ***
hr18                                   322.7247    17.6032  18.333  < 2e-16 ***
hr19                                   255.0724    17.0382  14.971  < 2e-16 ***
hr20                                   165.5845    17.0519   9.711  < 2e-16 ***
hr21                                   118.7447    18.7975   6.317 3.54e-10 ***
hr22                                    88.5090    17.7713   4.980 7.11e-07 ***
hr23                                    48.4402    17.3558   2.791 0.005323 **
holidayYes                             -34.5900    16.1397  -2.143 0.032267 *
weekdayMon                              -6.1089     9.9982  -0.611 0.541295
weekdayTue                              -5.1799     9.8042  -0.528 0.597347
weekdayWed                               8.1281     9.7111   0.837 0.402736
weekdayThu                               7.0418     9.9966   0.704 0.481284
weekdayFri                               3.5115     9.8830   0.355 0.722412
weekdaySat                              25.5919     9.8263   2.604 0.009297 **
weathersitMist/Cloudy                   -7.2428     6.3984  -1.132 0.257832
weathersitLight Snow/Light Rain        -63.1126    11.3221  -5.574 2.96e-08 ***
weathersitHeavy Rain/Thunderstorm/Snow  48.7557   100.5912   0.485 0.627969
temp                                     2.7097     1.9793   1.369 0.171193
atemp                                    2.3895     1.6384   1.458 0.144935
hum                                     -0.9538     0.1897  -5.027 5.60e-07 ***
windspeed                               -0.8004     0.3507  -2.282 0.022610 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 98.97 on 1447 degrees of freedom
Multiple R-squared:  0.6785,    Adjusted R-squared:  0.6669
F-statistic: 58.72 on 52 and 1447 DF,  p-value: < 2.2e-16
```

TABLE 5. Summary of the model selected with Stepwise procedure

```
Call:
lm(formula = cnt ~ season + yr + mnth + hr + holiday + weekday +
    weathersit + atemp + hum + windspeed, data = Bikes)

Residuals:
    Min      1Q  Median      3Q     Max
-287.89  -59.29   -7.92   44.95  421.78

Coefficients:
                                      Estimate Std. Error t value Pr(>|t|)
(Intercept)                           -69.5272    22.1834  -3.134 0.001758 **
seasonSpring                           21.0484    16.6354   1.265 0.205975
seasonSummer                           34.1820    18.9936   1.800 0.072122 .
seasonFall                             95.4802    15.8447   6.026 2.13e-09 ***
yr2012                                 81.5489     5.2551  15.518  < 2e-16 ***
mnthFeb                                 4.3661    13.2086   0.331 0.741036
mnthMar                                 7.0100    14.3234   0.489 0.624626
mnthApr                                26.5484    21.9223   1.211 0.226085
mnthMay                                33.3843    23.4981   1.421 0.155612
mnthJun                                19.0559    23.3952   0.815 0.415480
mnthJul                                -8.9369    26.2754  -0.340 0.733812
mnthAug                                16.9989    25.2178   0.674 0.500366
mnthSep                                27.8038    22.5512   1.233 0.217807
mnthOct                                 3.3683    21.5355   0.156 0.875733
mnthNov                               -27.5267    20.3297  -1.354 0.175944
mnthDec                               -25.4245    15.8186  -1.607 0.108215
hr1                                    -3.0717    17.6755  -0.174 0.862061
hr2                                   -19.3387    17.7332  -1.091 0.275659
hr3                                   -28.3844    18.0128  -1.576 0.115292
hr4                                   -12.9889    17.7883  -0.730 0.465391
hr5                                    -7.2885    17.6563  -0.413 0.679815
hr6                                    63.4365    16.9050   3.753 0.000182 ***
hr7                                   237.8477    18.1225  13.124  < 2e-16 ***
hr8                                   270.7283    17.6663  15.325  < 2e-16 ***
hr9                                   166.9871    17.2763   9.666  < 2e-16 ***
hr10                                  114.3147    17.8981   6.387 2.27e-10 ***
hr11                                  142.7244    17.9629   7.946 3.86e-15 ***
hr12                                  191.9578    17.0951  11.229  < 2e-16 ***
hr13                                  188.0458    17.4681  10.765  < 2e-16 ***
hr14                                  185.4315    17.8489  10.389  < 2e-16 ***
hr15                                  178.0674    18.2567   9.754  < 2e-16 ***
hr16                                  242.2663    18.7352  12.931  < 2e-16 ***
hr17                                  366.4756    18.6925  19.605  < 2e-16 ***
hr18                                  324.4037    17.5657  18.468  < 2e-16 ***
hr19                                  256.0095    17.0296  15.033  < 2e-16 ***
hr20                                  167.2896    17.0114   9.834  < 2e-16 ***
hr21                                  119.7277    18.7894   6.372 2.50e-10 ***
hr22                                   89.7176    17.7547   5.053 4.90e-07 ***
hr23                                   49.3341    17.3487   2.844 0.004522 **
holidayYes                            -33.9582    16.1380  -2.104 0.035530 *
weekdayMon                             -5.9190    10.0003  -0.592 0.554020
weekdayTue                             -5.0262     9.8065  -0.513 0.608356
weekdayWed                              8.5883     9.7082   0.885 0.376495
weekdayThu                              7.2520     9.9985   0.725 0.468377
weekdayFri                              4.6044     9.8537   0.467 0.640373
weekdaySat                             25.7274     9.8287   2.618 0.008948 **
weathersitMist/Cloudy                  -7.2783     6.4003  -1.137 0.255647
weathersitLight Snow/Light Rain       -63.1175    11.3255  -5.573 2.98e-08 ***
weathersitHeavy Rain/Thunderstorm/Snow  49.7163   100.6191   0.494 0.621308
atemp                                   4.4510     0.6460   6.890 8.29e-12 ***
hum                                    -0.9615     0.1897  -5.069 4.52e-07 ***
windspeed                              -0.6970     0.3426  -2.035 0.042079 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 99 on 1448 degrees of freedom
Multiple R-squared:  0.678,     Adjusted R-squared:  0.6667
F-statistic:  59.8 on 51 and 1448 DF,  p-value: < 2.2e-16
```

TABLE 6. Normality tests Shapiro-Wilk and Lillie for model selected from stepwise

```
        Shapiro-Wilk normality test

data:  modelstep$residuals
W = 0.96167, p-value < 2.2e-16

        Lilliefors (Kolmogorov-Smirnov) normality test

data:  modelstep$residuals
D = 0.073446, p-value < 2.2e-16
```

Shapiro-Wilk test p-value < .001 and KS test p-value < .001 which means that normality is rejected

TABLE 7. Non-constant variance score test for model selected from stepwise

```
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 429.0659, Df = 1, p = < 2.22e-16
```

Non-constant variance test p-value < .001 which means that homoscedasticity is rejected

TABLE 8. Tukey test for model selected from stepwise

```
          Test stat Pr(>|Test stat|)
season
yr
mnth
hr
holiday
weekday
weathersit
atemp         0.7270          0.46736
hum          -2.0037          0.04529 *
windspeed    -0.9788          0.32786
Tukey test   20.5928          < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Tukey test p-value < .001 which means that linearity is rejected

TABLE 9. Runs test for model selected from stepwise

```
        Runs Test

data:  modelstep$res
statistic = 1.0331, runs = 771, n1 = 750, n2 = 750, n = 1500, p-value = 0.3015
alternative hypothesis: nonrandomness
```

Runs test p-value = .30 > .05 which means that independence is not rejected

TABLE 10. Tukey test for our final model

	Test stat	Pr(>|Test stat|)
season		
yr		
mnth		
hr		
holiday		
weekday		
weathersit		
atemp	0.7670	0.4432
hum	0.5721	0.5674
windspeed	-0.4870	0.6264
I(hum^2)	0.5247	0.5999
I(hum^3)	0.4449	0.6564
I(atemp^2)	0.7414	0.4586
I(atemp^3)	0.9246	0.3554
Tukey test	0.7684	0.4422

Tukey test p-value = .44 > .05 which means that linearity is not rejected

TABLE 11. Non-constant variance score test for our final model

```
Non-constant Variance Score Test
Variance formula: ~ fitted.values
Chisquare = 1.939088, Df = 1, p = 0.16377
```

Non-constant variance test p-value = .16 > .05 which means that homoscedasticity is not rejected

TABLE 12. Runs test for our final model

```
        Runs Test

data:  weighted_model2$res
statistic = 0, runs = 751, n1 = 750, n2 = 750, n = 1500, p-value = 1
alternative hypothesis: nonrandomness
```

Runs test p-value = 1 > .05 which means that independence is not rejected

TABLE 13. Normality tests Shapiro-Wilk and Lillie for our final model

```
        Shapiro-Wilk normality test

data:  weighted_model2$residuals
W = 0.96957, p-value < 2.2e-16

        Lilliefors (Kolmogorov-Smirnov) normality test

data:  weighted_model2$residuals
D = 0.070971, p-value < 2.2e-16
```

Shapiro-Wilk test p-value < .001 and KS test p-value < .001 which means that normality is rejected

TABLE 14. Summary of our final model

```
Call:
lm(formula = log(cnt) ~ season + yr + mnth + hr + holiday + weekday +
    weathersit + atemp + hum + windspeed + I(hum^2) + I(hum^3) +
    I(atemp^2) + I(atemp^3), data = Bikes, weights = wt)

Weighted Residuals:
    Min      1Q  Median      3Q     Max
-5.0954 -0.6639  0.0304  0.8050  3.6067

Coefficients:
                                      Estimate Std. Error t value Pr(>|t|)
(Intercept)                          2.083e+00  3.519e-01   5.920 4.01e-09 ***
seasonSpring                         2.453e-01  8.890e-02   2.760 0.005856 **
seasonSummer                         4.485e-01  9.855e-02   4.550 5.80e-06 ***
seasonFall                           6.315e-01  8.664e-02   7.289 5.13e-13 ***
yr2012                               4.660e-01  2.688e-02  17.341  < 2e-16 ***
mnthFeb                              6.585e-02  7.869e-02   0.837 0.402804
mnthMar                              9.360e-02  8.368e-02   1.119 0.263510
mnthApr                              1.001e-01  1.205e-01   0.831 0.406270
mnthMay                              1.393e-01  1.271e-01   1.096 0.273120
mnthJun                              1.034e-01  1.253e-01   0.825 0.409277
mnthJul                             -1.185e-03  1.371e-01  -0.009 0.993101
mnthAug                             -3.591e-02  1.322e-01  -0.272 0.785889
mnthSep                              4.495e-03  1.209e-01   0.037 0.970347
mnthOct                             -5.882e-02  1.168e-01  -0.503 0.614754
mnthNov                             -1.725e-01  1.121e-01  -1.540 0.123900
mnthDec                             -1.390e-01  9.188e-02  -1.513 0.130534
hr1                                 -5.940e-01  1.236e-01  -4.807 1.69e-06 ***
hr2                                 -1.180e+00  1.314e-01  -8.976  < 2e-16 ***
hr3                                 -1.459e+00  1.365e-01 -10.684  < 2e-16 ***
hr4                                 -1.973e+00  1.443e-01 -13.671  < 2e-16 ***
hr5                                 -7.825e-01  1.268e-01  -6.171 8.79e-10 ***
hr6                                  5.678e-01  1.071e-01   5.303 1.31e-07 ***
hr7                                  1.749e+00  1.006e-01  17.385  < 2e-16 ***
hr8                                  1.799e+00  9.840e-02  18.288  < 2e-16 ***
hr9                                  1.570e+00  9.789e-02  16.041  < 2e-16 ***
hr10                                 1.236e+00  1.032e-01  11.975  < 2e-16 ***
hr11                                 1.361e+00  1.011e-01  13.470  < 2e-16 ***
hr12                                 1.600e+00  9.597e-02  16.667  < 2e-16 ***
hr13                                 1.617e+00  9.858e-02  16.402  < 2e-16 ***
hr14                                 1.578e+00  9.973e-02  15.825  < 2e-16 ***
hr15                                 1.570e+00  1.008e-01  15.572  < 2e-16 ***
hr16                                 1.828e+00  1.006e-01  18.180  < 2e-16 ***
hr17                                 2.195e+00  9.746e-02  22.522  < 2e-16 ***
hr18                                 2.090e+00  9.541e-02  21.905  < 2e-16 ***
hr19                                 1.921e+00  9.413e-02  20.405  < 2e-16 ***
hr20                                 1.531e+00  9.810e-02  15.603  < 2e-16 ***
hr21                                 1.268e+00  1.073e-01  11.816  < 2e-16 ***
hr22                                 9.935e-01  1.071e-01   9.280  < 2e-16 ***
hr23                                 6.648e-01  1.072e-01   6.204 7.18e-10 ***
holidayYes                          -2.012e-01  8.214e-02  -2.449 0.014435 *
weekdayMon                           1.917e-02  5.126e-02   0.374 0.708496
weekdayTue                           2.753e-02  4.988e-02   0.552 0.581134
weekdayWed                           6.241e-02  4.867e-02   1.282 0.199941
weekdayThu                           1.026e-01  5.039e-02   2.037 0.041873 *
weekdayFri                           1.399e-01  4.975e-02   2.811 0.004999 **
weekdaySat                           1.756e-01  5.065e-02   3.467 0.000542 ***
weathersitMist/Cloudy               -6.117e-02  3.311e-02  -1.847 0.064883 .
weathersitLight Snow/Light Rain     -6.096e-01  6.445e-02  -9.458  < 2e-16 ***
weathersitHeavy Rain/Thunderstorm/Snow  7.455e-01  8.675e-01   0.859 0.390286
atemp                                1.141e-02  3.184e-02   0.358 0.720087
hum                                  2.019e-02  1.396e-02   1.446 0.148458
windspeed                           -3.268e-03  1.706e-03  -1.916 0.055588 .
I(hum^2)                            -2.946e-04  2.490e-04  -1.183 0.236911
I(hum^3)                             9.094e-07  1.403e-06   0.648 0.517134
I(atemp^2)                           2.241e-03  1.355e-03   1.654 0.098334 .
I(atemp^3)                          -5.034e-05  1.792e-05  -2.810 0.005022 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.245 on 1444 degrees of freedom
Multiple R-squared:  0.8088,    Adjusted R-squared:  0.8015
F-statistic: 111.1 on 55 and 1444 DF,  p-value: < 2.2e-16
```

TABLE 15. Describe table for Winter

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
season*	1	376	1.00	0.00	1.00	1.00	0.00	1.00	1.00	0.00	NaN	NaN	0.00
yr*	2	376	1.51	0.50	2.00	1.51	0.00	1.00	2.00	1.00	-0.03	-2.00	0.03
mnth*	3	376	3.28	3.58	2.00	2.49	1.48	1.00	12.00	11.00	1.91	1.95	0.18
hr*	4	376	13.01	6.78	13.00	13.10	8.90	1.00	24.00	23.00	-0.07	-1.14	0.35
holiday*	5	376	1.04	0.20	1.00	1.00	0.00	1.00	2.00	1.00	4.51	18.43	0.01
weekday*	6	376	3.98	2.09	4.00	3.97	2.97	1.00	7.00	6.00	0.04	-1.38	0.11
workingday*	7	376	1.64	0.48	2.00	1.68	0.00	1.00	2.00	1.00	-0.59	-1.66	0.02
weathersit*	8	376	1.45	0.65	1.00	1.34	0.00	1.00	4.00	3.00	1.18	0.46	0.03
temp	9	376	12.18	4.79	11.48	11.90	4.86	0.82	28.70	27.88	0.65	0.69	0.25
atemp	10	376	14.76	5.62	13.63	14.49	4.49	0.00	31.82	31.82	0.55	0.31	0.29
hum	11	376	58.51	19.41	56.00	58.06	17.79	0.00	100.00	100.00	0.20	-0.41	1.00
windspeed	12	376	14.59	9.26	13.00	14.15	8.90	0.00	54.00	54.00	0.66	0.79	0.48
casual	13	376	15.36	31.66	6.50	8.85	8.15	0.00	352.00	352.00	6.09	50.38	1.63
registered	14	376	97.66	95.57	72.00	83.04	75.61	1.00	648.00	647.00	1.76	4.49	4.93
cnt	15	376	113.02	113.78	82.50	94.74	87.47	1.00	746.00	745.00	1.97	5.61	5.87

TABLE 16. Describe table for Spring

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
season*	1	392	2.00	0.00	2.00	2.00	0.00	2.00	2.00	0.00	NaN	NaN	0.00
yr*	2	392	1.49	0.50	1.00	1.48	0.00	1.00	2.00	1.00	0.05	-2.00	0.03
mnth*	3	392	4.64	0.94	5.00	4.68	1.48	3.00	6.00	3.00	0.03	-0.99	0.05
hr*	4	392	12.53	7.00	13.00	12.56	8.90	1.00	24.00	23.00	-0.03	-1.24	0.35
holiday*	5	392	1.03	0.17	1.00	1.00	0.00	1.00	2.00	1.00	5.69	30.49	0.01
weekday*	6	392	4.05	1.93	4.00	4.06	2.97	1.00	7.00	6.00	0.00	-1.17	0.10
workingday*	7	392	1.72	0.45	2.00	1.78	0.00	1.00	2.00	1.00	-0.99	-1.03	0.02
weathersit*	8	392	1.46	0.68	1.00	1.33	0.00	1.00	3.00	2.00	1.15	0.00	0.03
temp	9	392	22.11	5.55	22.96	22.27	6.08	6.56	36.08	29.52	-0.26	-0.54	0.28
atemp	10	392	25.80	5.99	26.52	26.17	6.74	8.33	40.91	32.58	-0.47	-0.34	0.30
hum	11	392	63.68	21.08	65.00	64.31	26.69	20.00	100.00	80.00	-0.20	-1.04	1.06
windspeed	12	392	13.84	8.20	13.00	13.55	8.89	0.00	43.00	43.00	0.44	0.16	0.41
casual	13	392	42.67	55.38	22.00	30.56	28.17	0.00	308.00	308.00	2.22	5.19	2.80
registered	14	392	147.61	130.25	119.00	130.72	129.73	1.00	697.00	696.00	1.35	2.38	6.58
cnt	15	392	190.28	167.89	151.00	167.83	166.05	1.00	819.00	818.00	1.10	0.86	8.48

TABLE 17. Describe table for Summer

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
season*	1	382	3.00	0.00	3.00	3.00	0.00	3.00	3.00	0.00	NaN	NaN	0.00
yr*	2	382	1.51	0.50	2.00	1.51	0.00	1.00	2.00	1.00	-0.03	-2.00	0.03
mnth*	3	382	7.68	0.96	8.00	7.72	1.48	6.00	9.00	3.00	-0.11	-0.98	0.05
hr*	4	382	11.74	6.87	12.00	11.59	8.90	1.00	24.00	23.00	0.12	-1.17	0.35
holiday*	5	382	1.02	0.15	1.00	1.00	0.00	1.00	2.00	1.00	6.26	37.26	0.01
weekday*	6	382	4.05	1.94	4.00	4.07	2.97	1.00	7.00	6.00	-0.02	-1.18	0.10
workingday*	7	382	1.71	0.45	2.00	1.76	0.00	1.00	2.00	1.00	-0.93	-1.13	0.02
weathersit*	8	382	1.33	0.56	1.00	1.23	0.00	1.00	3.00	2.00	1.50	1.28	0.03
temp	9	382	28.96	3.87	28.70	28.93	3.65	18.04	39.36	21.32	0.06	0.08	0.20
atemp	10	382	32.68	4.91	33.34	32.77	4.49	12.12	45.45	33.33	-0.70	2.90	0.25
hum	11	382	64.42	17.45	67.00	65.29	17.79	24.00	100.00	76.00	-0.38	-0.81	0.89
windspeed	12	382	11.02	7.02	11.00	10.83	5.93	0.00	35.00	35.00	0.34	0.08	0.36
casual	13	382	47.31	48.20	32.50	39.43	41.51	0.00	237.00	237.00	1.49	2.29	2.47
registered	14	382	174.62	153.50	146.00	153.74	151.97	2.00	781.00	779.00	1.18	1.35	7.85
cnt	15	382	221.93	185.28	191.50	201.72	200.89	2.00	900.00	898.00	0.87	0.33	9.48

TABLE 18. Describe table for Fall

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
season*	1	350	4.00	0.00	4.00	4.00	0.00	4.00	4.00	0.00	NaN	NaN	0.00
yr*	2	350	1.44	0.50	1.00	1.43	0.00	1.00	2.00	1.00	0.24	-1.95	0.03
mnth*	3	350	10.75	0.94	11.00	10.81	1.48	9.00	12.00	3.00	-0.24	-0.85	0.05
hr*	4	350	12.86	6.87	13.00	12.94	8.90	1.00	24.00	23.00	-0.09	-1.16	0.37
holiday*	5	350	1.02	0.15	1.00	1.00	0.00	1.00	2.00	1.00	6.36	38.54	0.01
weekday*	6	350	3.89	1.99	4.00	3.86	2.97	1.00	7.00	6.00	0.10	-1.21	0.11
workingday*	7	350	1.69	0.46	2.00	1.74	0.00	1.00	2.00	1.00	-0.83	-1.32	0.02
weathersit*	8	350	1.45	0.65	1.00	1.34	0.00	1.00	3.00	2.00	1.11	0.06	0.03
temp	9	350	17.56	5.07	17.22	17.42	6.08	5.74	29.52	23.78	0.25	-0.64	0.27
atemp	10	350	21.05	5.41	21.21	20.93	5.62	8.33	33.34	25.00	0.12	-0.61	0.29
hum	11	350	67.07	18.83	68.00	67.55	23.72	28.00	100.00	72.00	-0.16	-1.11	1.01
windspeed	12	350	11.05	8.21	11.00	10.45	7.41	0.00	37.00	37.00	0.51	-0.15	0.44
casual	13	350	33.04	47.65	14.00	22.11	18.53	0.00	335.00	335.00	2.54	7.76	2.55
registered	14	350	175.40	163.89	138.00	150.94	147.52	1.00	876.00	875.00	1.51	2.73	8.76
cnt	15	350	208.44	188.95	158.50	183.24	183.84	1.00	963.00	962.00	1.21	1.41	10.10

Figures

Figure 1: Histograms of numeric variables

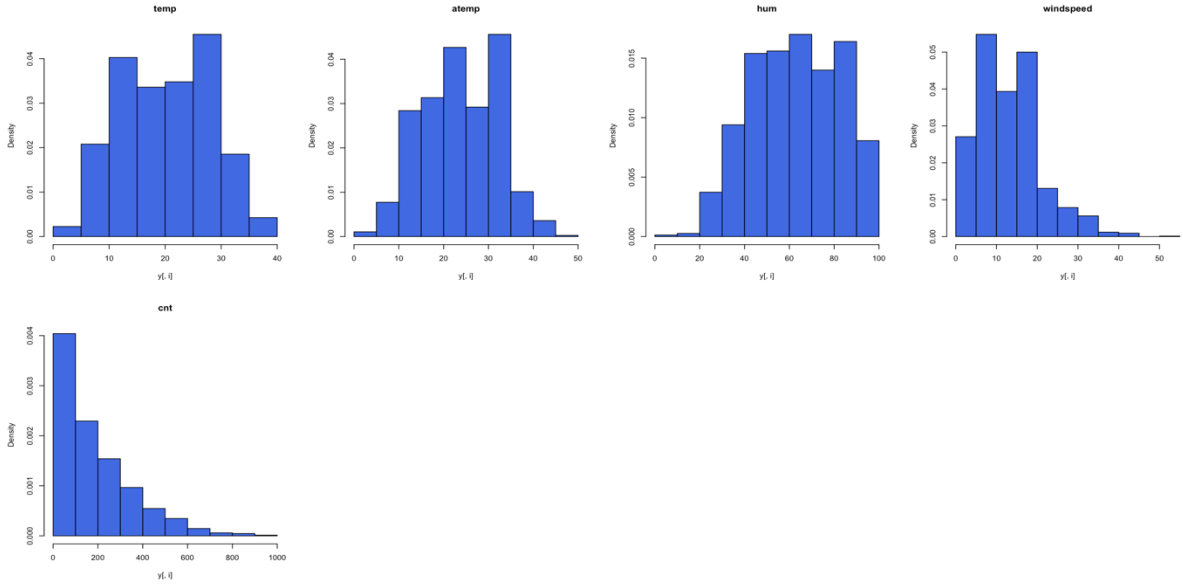


Figure 2: Bar plots for factor variables

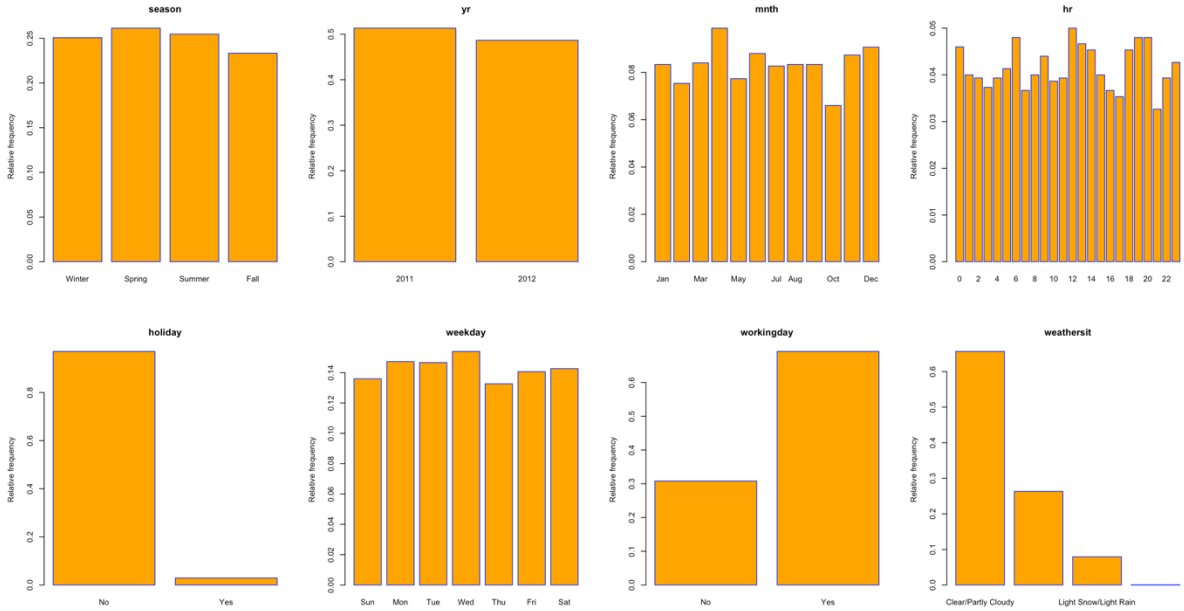


Figure 3: Scatter plots for pairwise comparisons of numeric variables

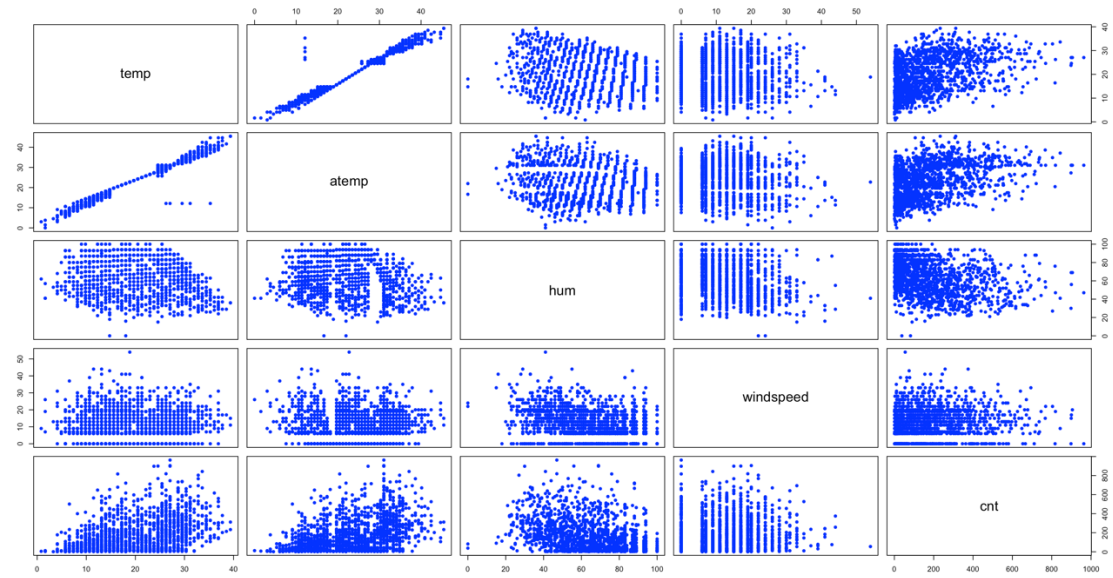


Figure 4: Scatter plots for pairwise comparisons of numeric variables with count variable

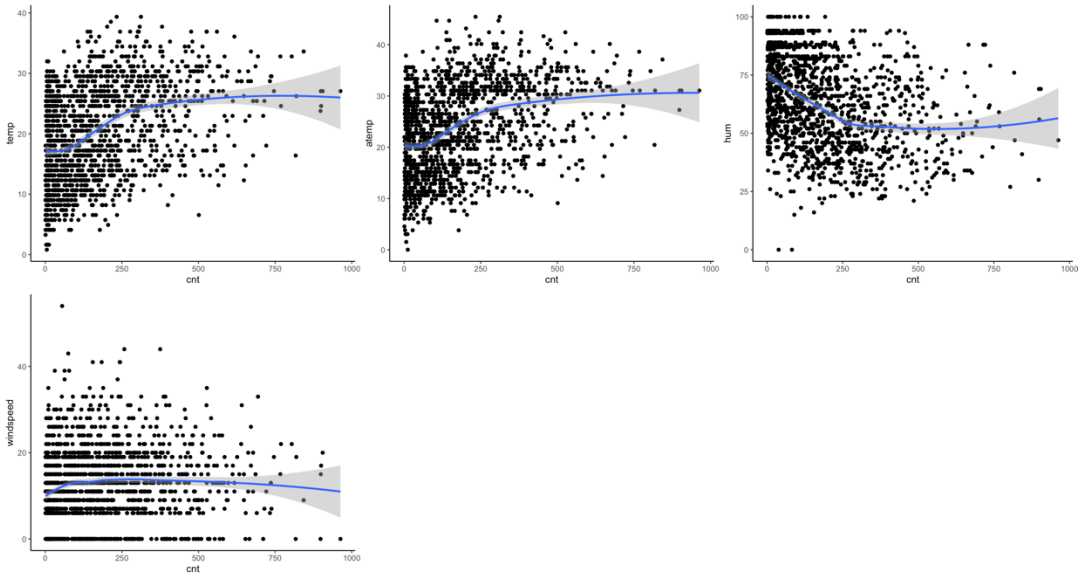


Figure 5: Box plots for pairwise comparisons of factor variables with count variable

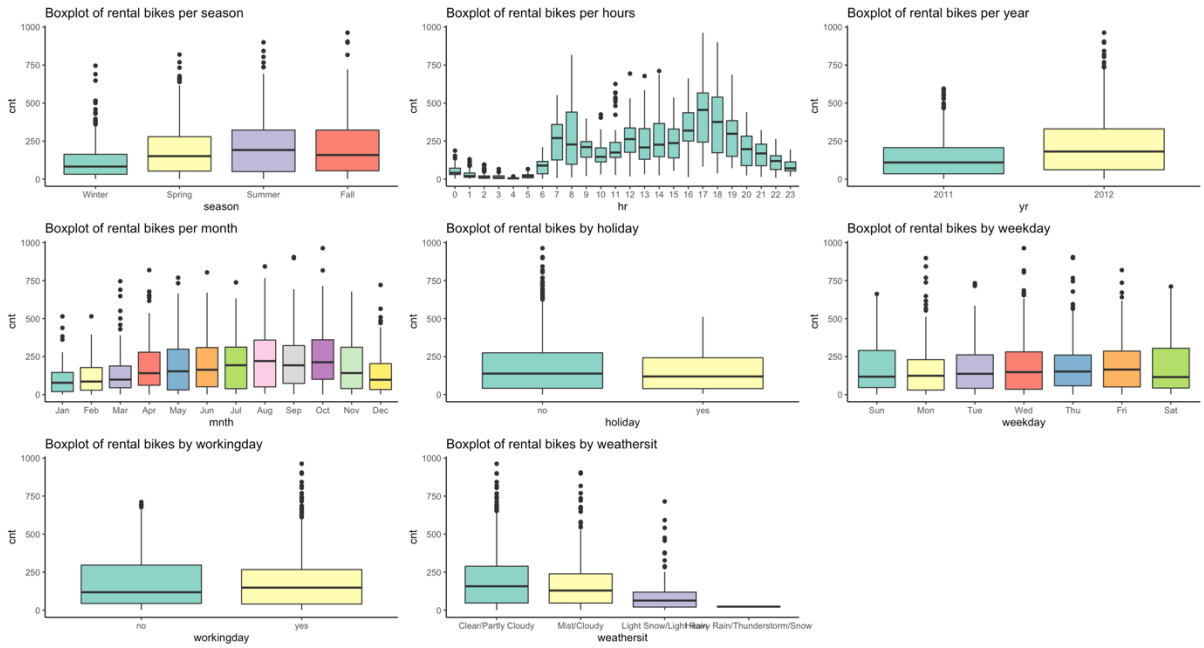


Figure 6: Correlation plot for numeric variables

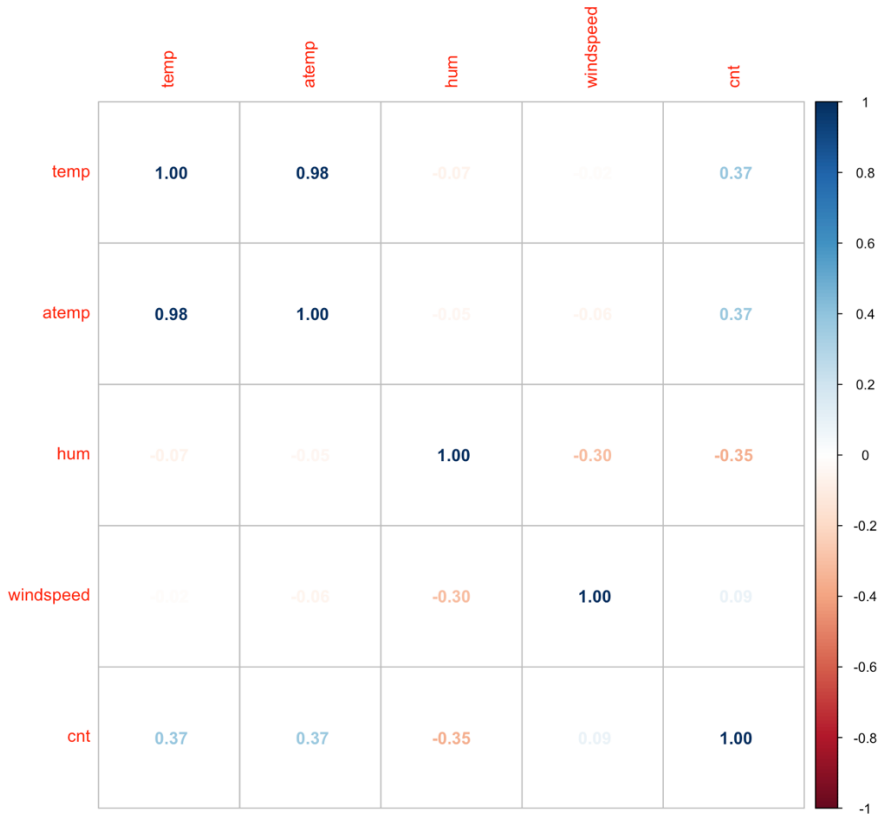
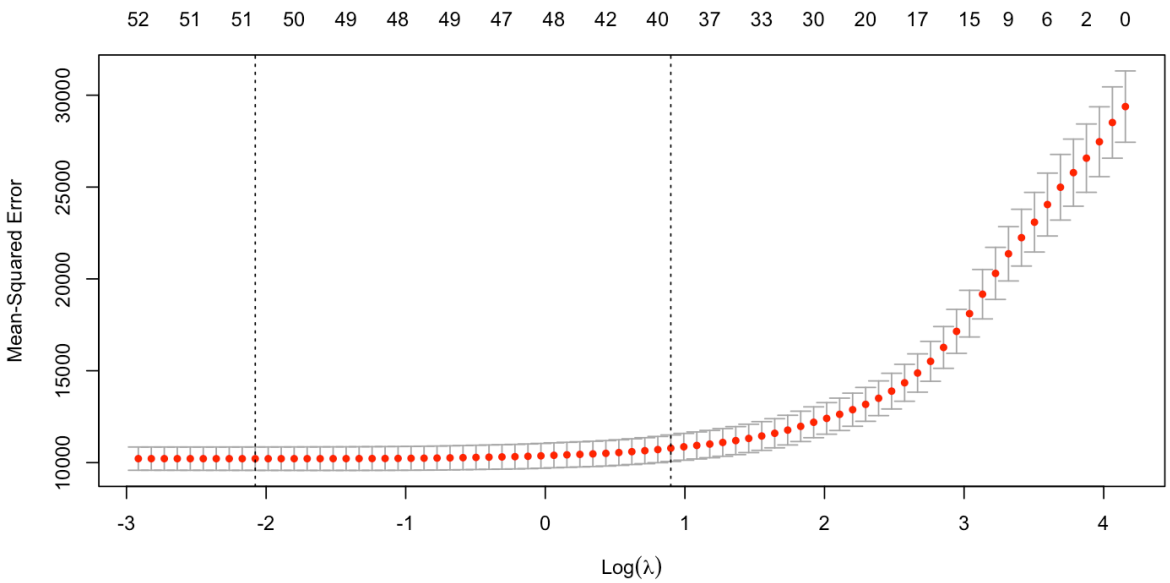


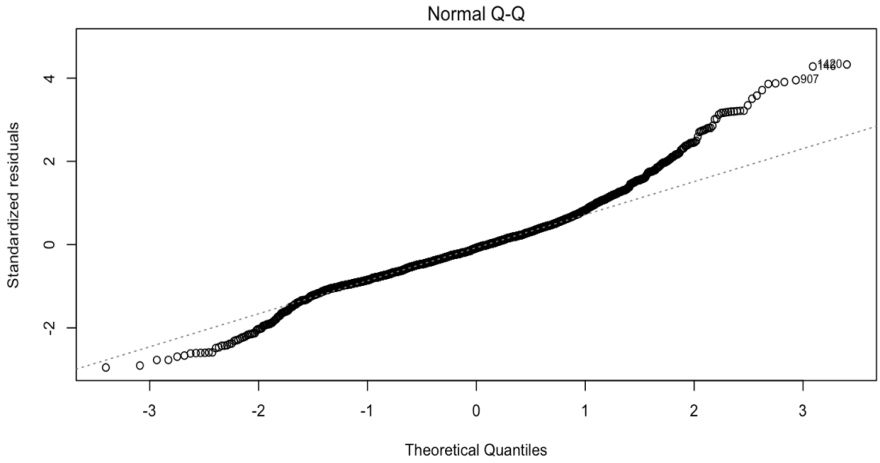
Figure 7: The cross-validation curve (red dotted line) along with upper and lower standard deviation curves along the λ sequence (error bars)



We use the cross validation to select our λ. First λ will be the lambda.min= is the one with the minimum CV-MSE and the second one is largest value of lambda such that error is within 1 standard error of the minimum. On the top part of the graph, we can see the number of coefficients dropping as log(λ) increases as well as the MSE.

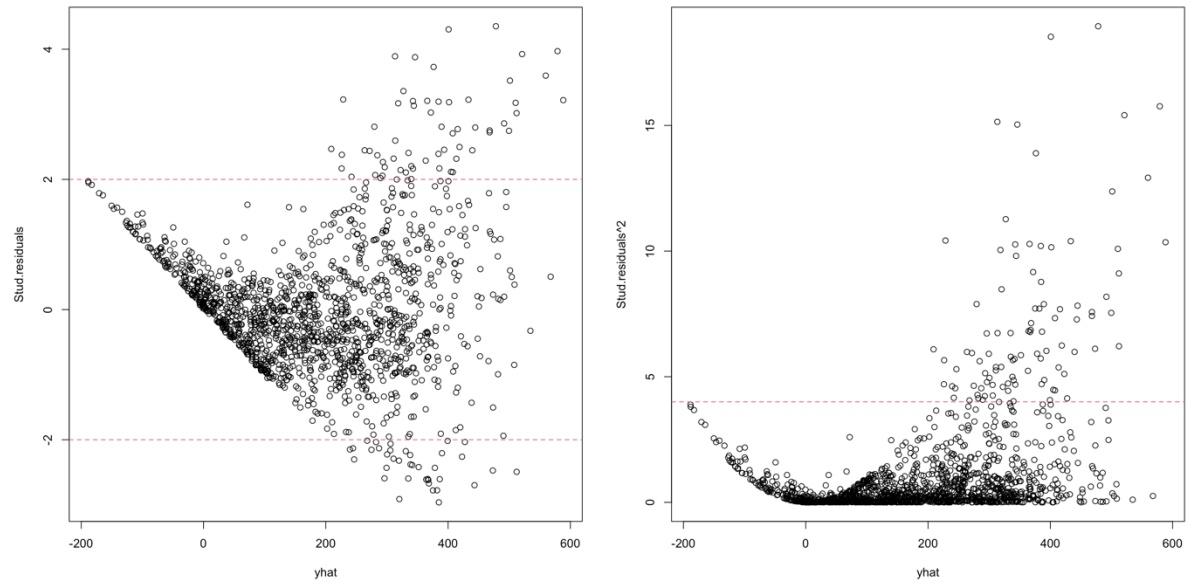
The vertical lines depict the lambdas we mentioned above, the first one is the lambda.min and the second one is the lambda.1se. It is better to choose the lambda.1se because we have less coefficients at this stage.

Figure 8: Normality QQ-plot



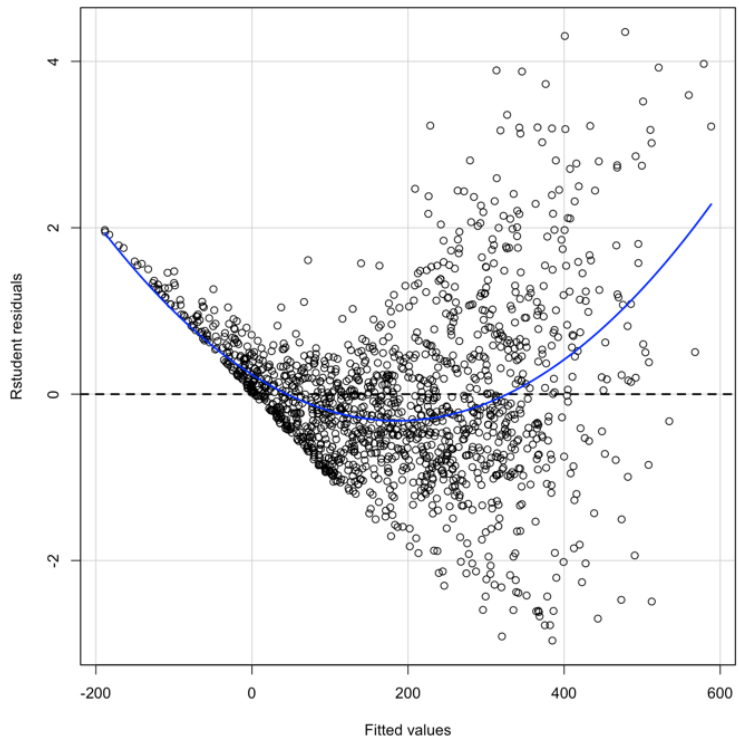
The points seem to not follow the straight diagonal line, and this aligns with the tests we conducted previously (normality is rejected)

Figure 9: Fitted values vs. standardized or studentized residuals using 95% quantiles from the correct distributions & Fitted values vs. studentized residuals squared for homoscedasticity test



We can see that the points are not evenly distributed (some of them are very close to each other and the rest are more sparse). The spread is not constant. Homoscedasticity is rejected.

Figure 10: Fitted values vs Studentized residuals for linearity check



We can assume from the plot of Residuals vs Fitted that there is no linearity since we can see a curve.

Figure 11: Frequency bar-plot for weather during Winter season

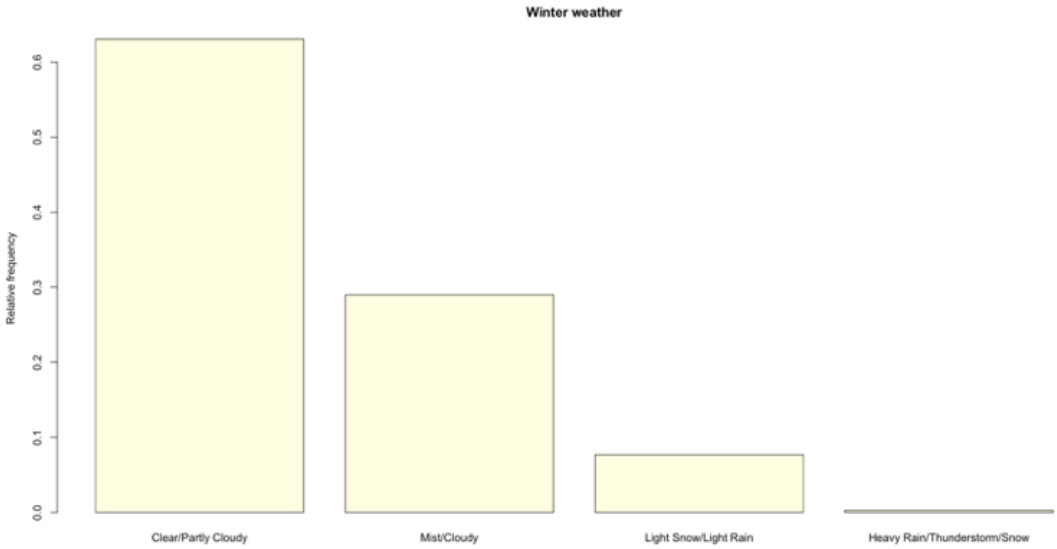


Figure 12: Boxplot for total rentals per hours during Winter season

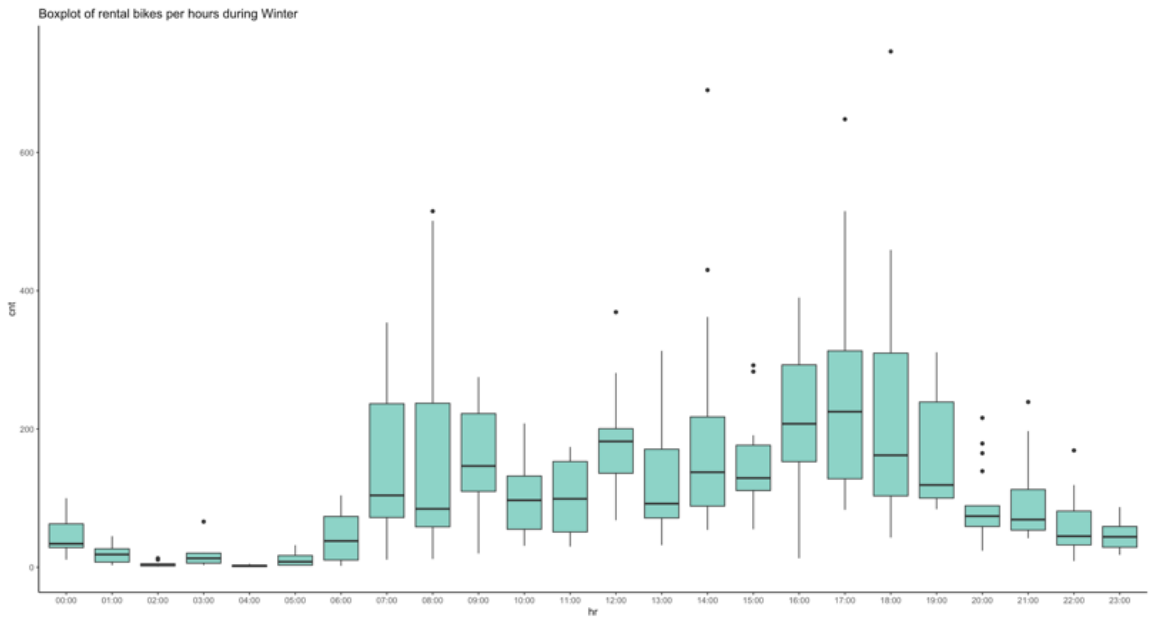


Figure 13: Frequency bar-plot for weather during Spring season

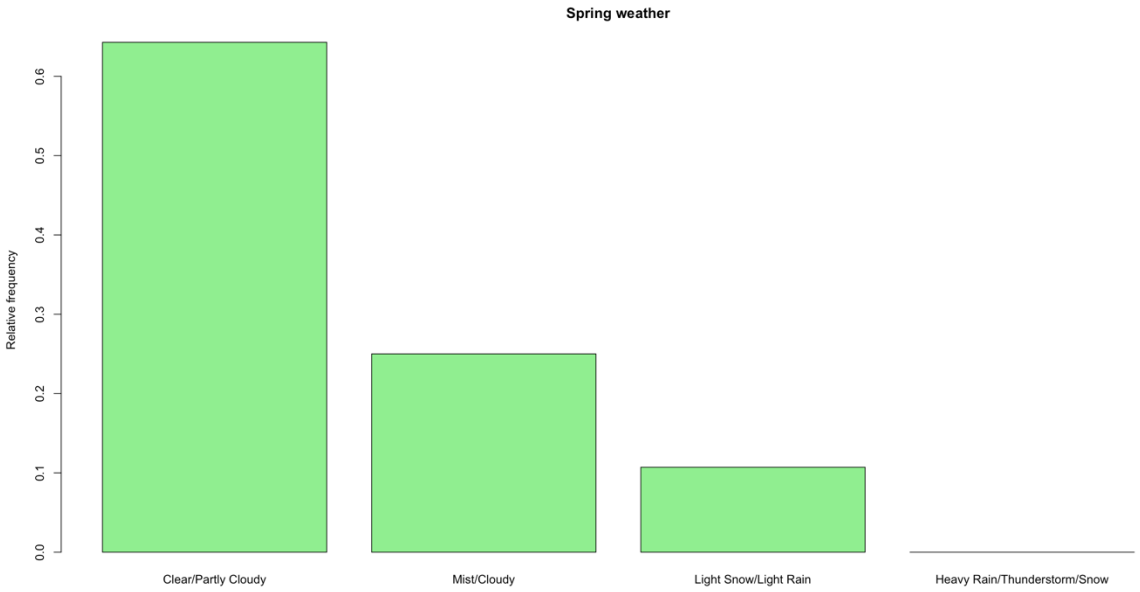


Figure 14: Boxplot for total rentals per hours during Spring season

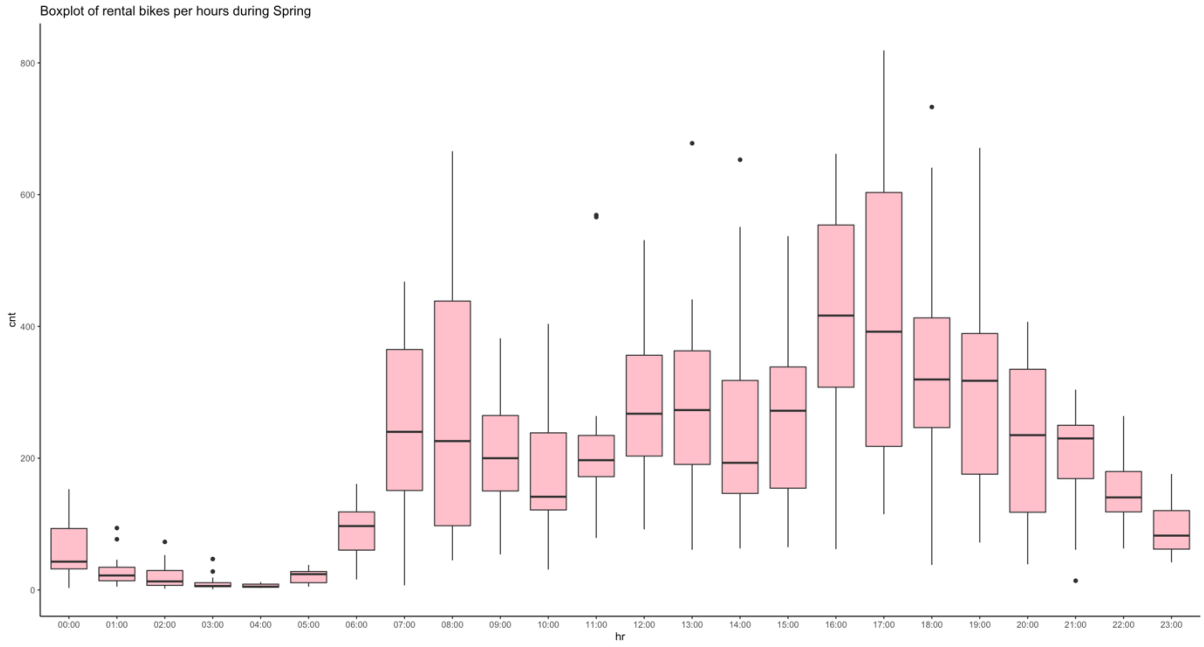


Figure 15: Frequency bar-plot for weather during Summer season

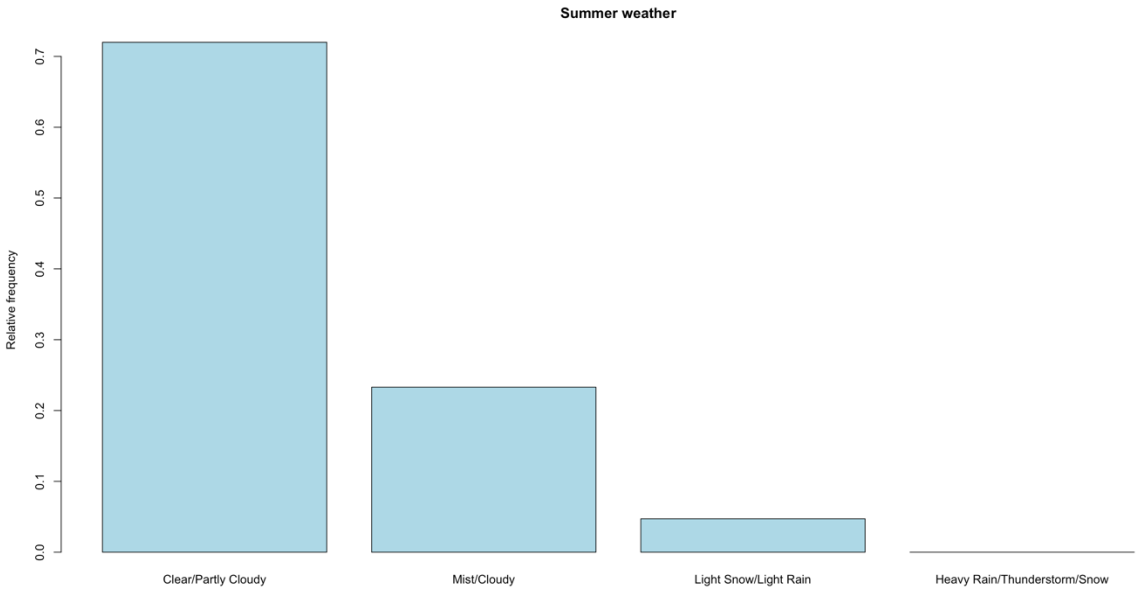


Figure 16: Boxplot for total rentals per hours during Summer season

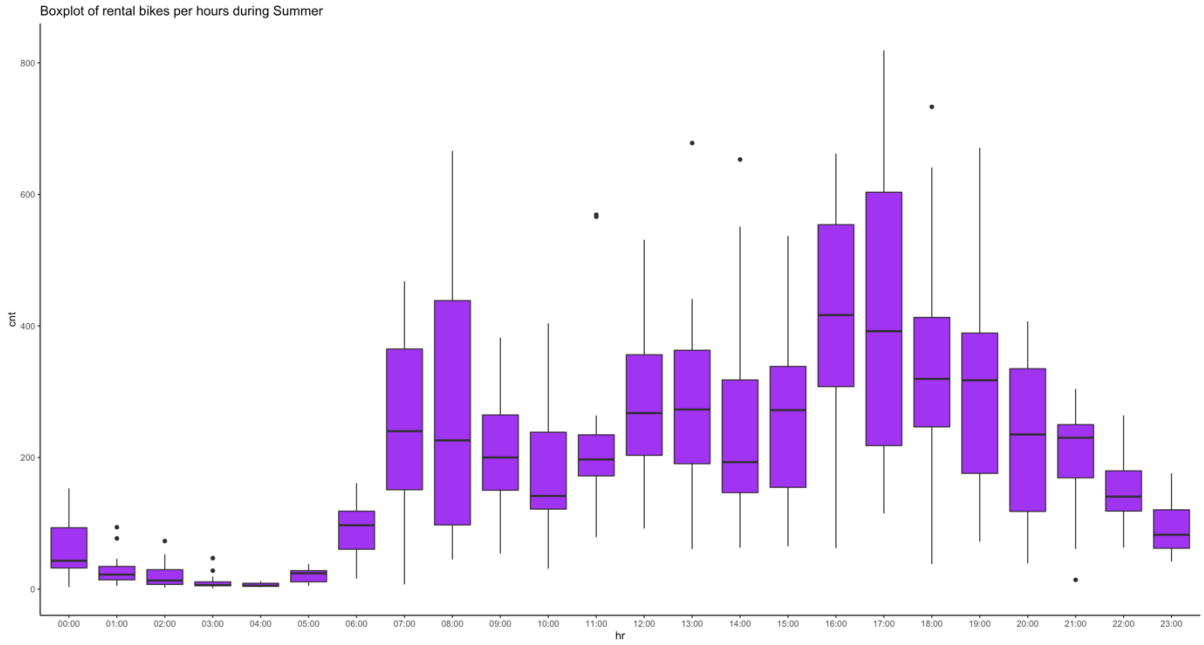


Figure 17: Frequency bar-plot for weather during Fall season

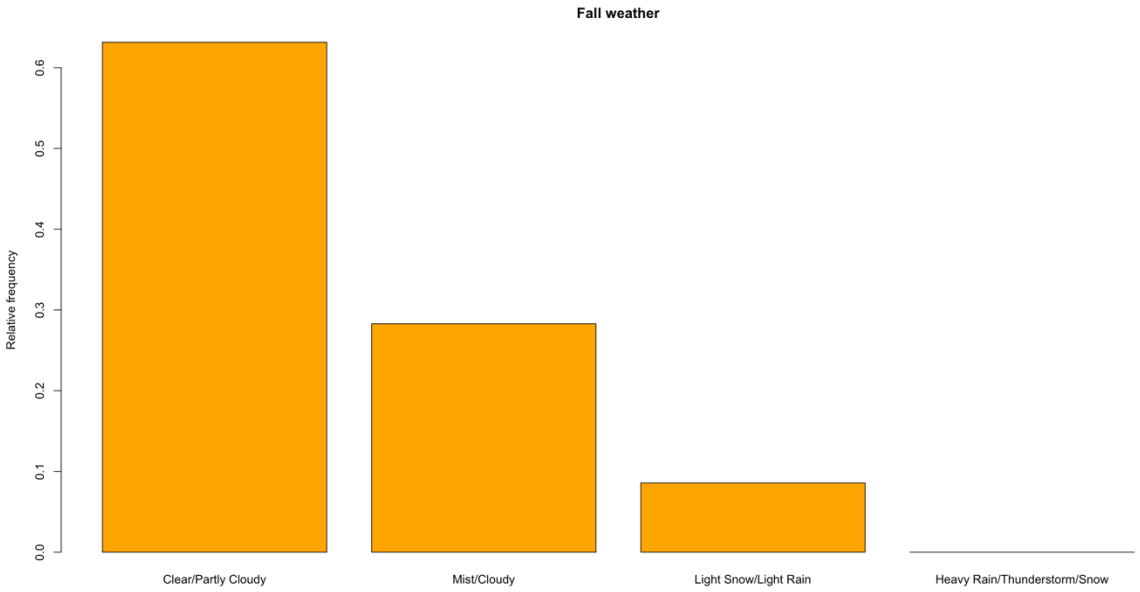


Figure 18: Boxplot for total rentals per hours during Fall season