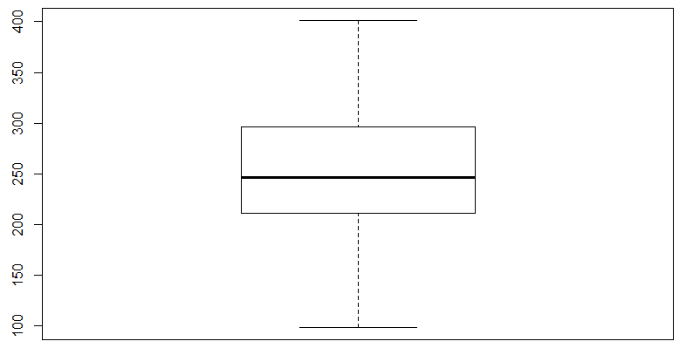
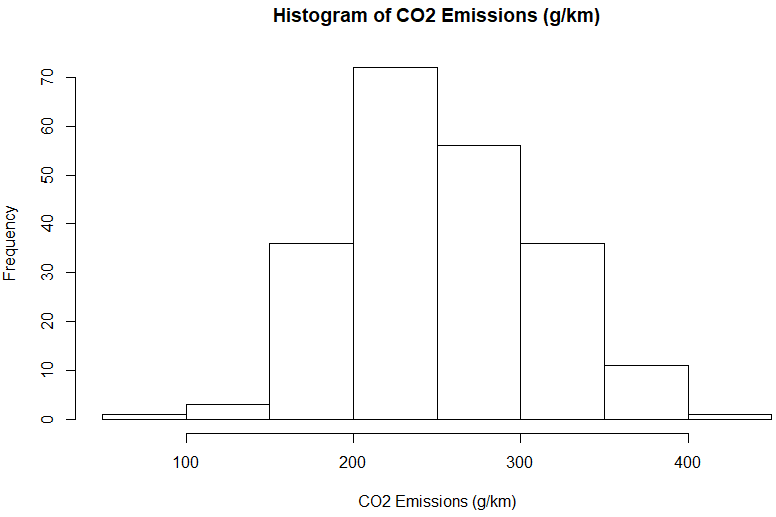
Datasets provide model-specific fuel consumption ratings and estimated carbon dioxide emissions for new light-duty vehicles for retail sale in Canada. This dataset contains 216 observations from cars manufactured in 2020 that are currently being actively used in 2020. Response variable is the “CO2 Emissions (g/kg)” and independent variables consist of three qualitative variables namely, “Make”, “Transmission”, and “Fuel Type” and three quantitative variables namely, “Cylinders”, “Engine Size (L), and “Fuel Consumption Comb (L/100 km)”.

# Response Variable

## “CO2 Emissions (g/kg)”

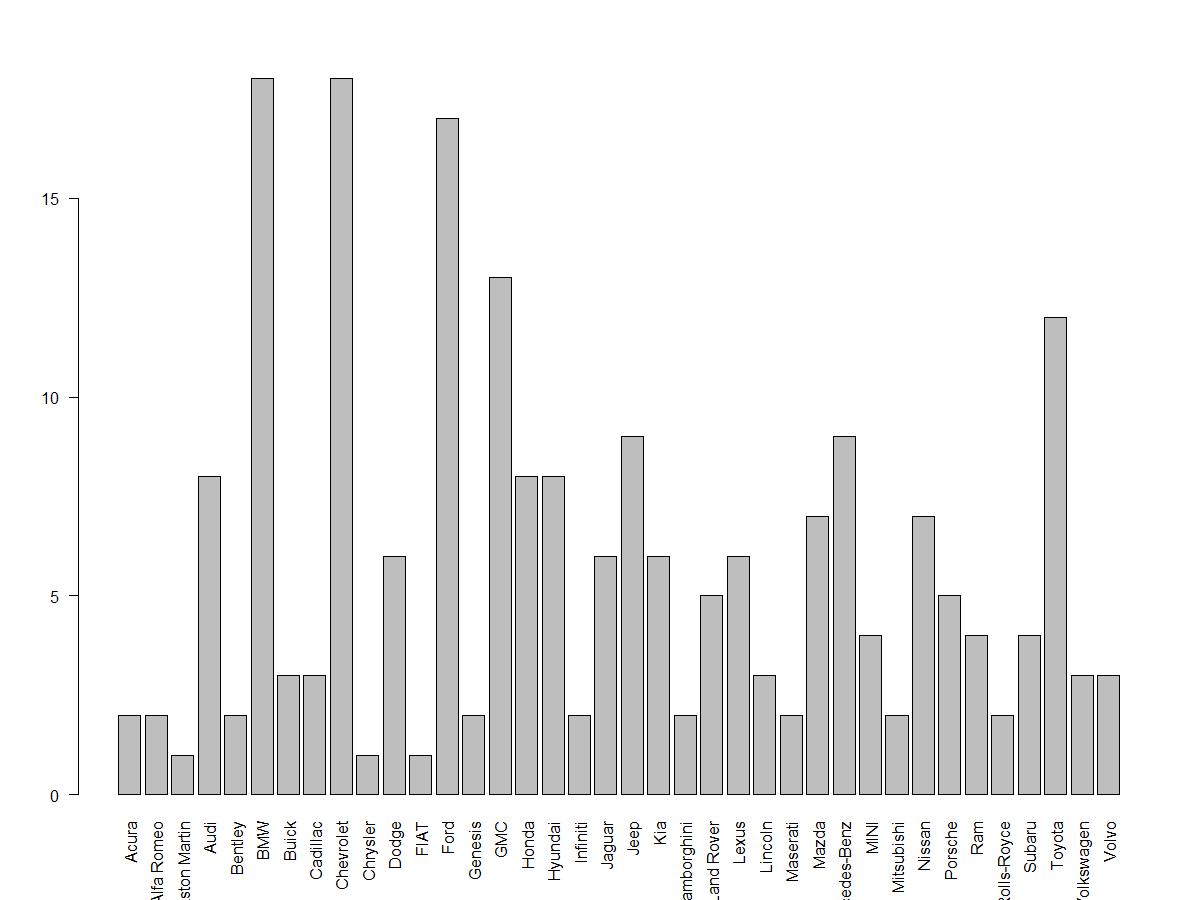


|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Mean | sd | Median | Trimmed | Mad | Min | Max | Range | Skew | Kurtosis | se |
| 252.87 | 57.94 | 246.5 | 250.76 | 58.56 | 99 | 401 | 302 | 0.32 | -0.26 | 3.94 |

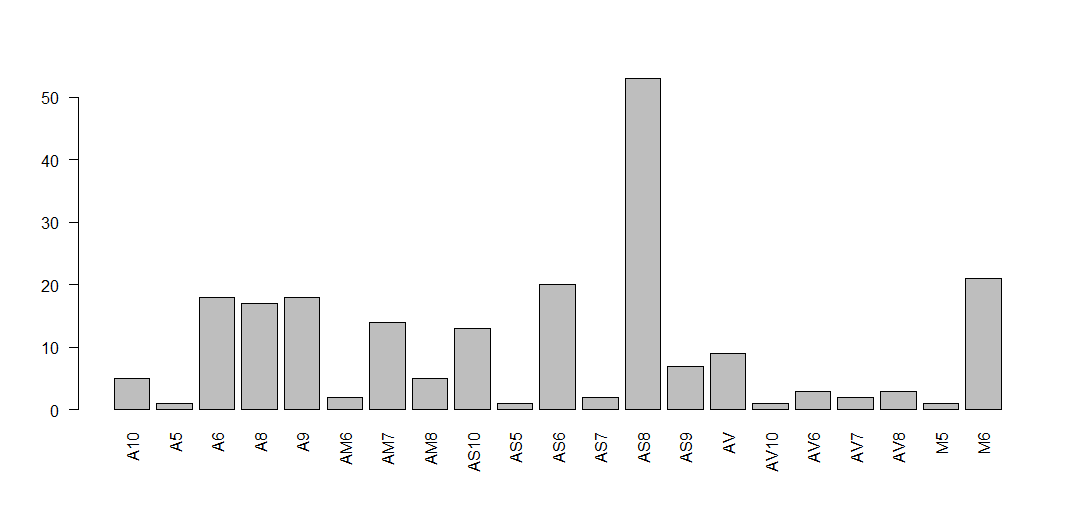
The histogram and boxplot indicate that the response variable is fairly normal and mound-shape. The mean and median value are relatively close which tells us of variable’s low skewness which could be seen from the provided skewness value. From the summary table we could also see that min and max values fall inside mean ± 3sd or (79.05,426.69) confirming absence of outliers. The negative kurtosis value indicates that the distribution has a lighter tails and flatter peak compared to standard normal distribution.

# Independent Variables: Categorical

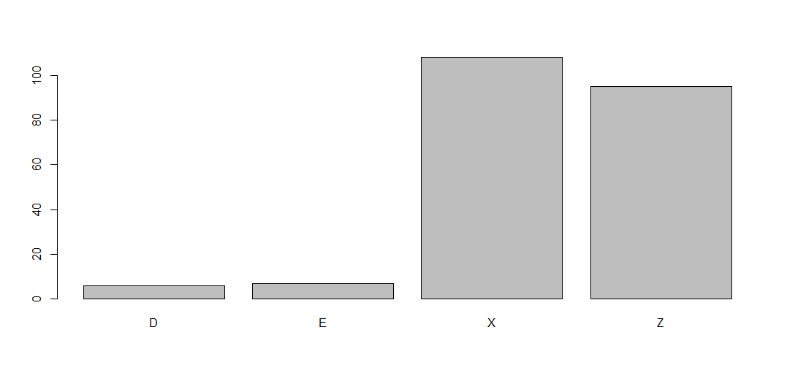
## “Make”

Make is a categorical variable denoting the brand of the manufacturer of the vehicle such as Audi or Toyota. The bar plot shows the frequency of each brand in the sample with BMW, Chevrolet, Ford, GMC, and Toyota having the most frequencies.

## “Transmission”

Transmission is a categorical variable denoting the gear type. The variable is a combination of a letter(s) plus a number which indicates the number of speeds available. The most frequent transmission type is “Automatic with Select Shift with 8 speeds.”

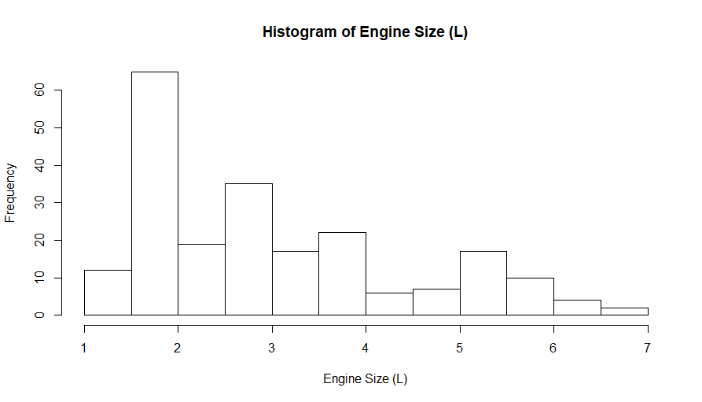
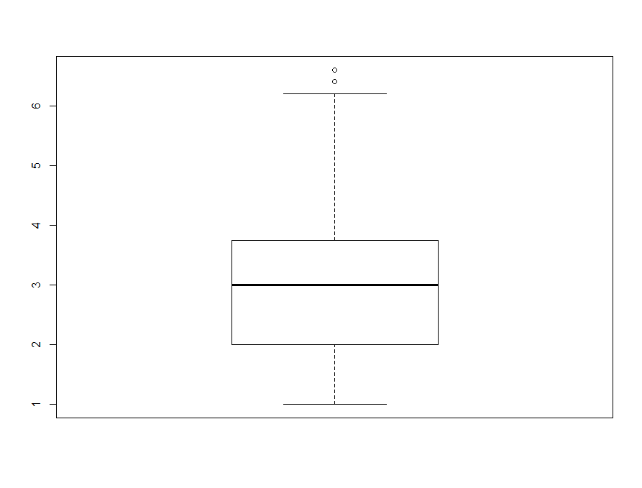
## “Fuel Type”

The most frequent fuel types are X (Regular Gasoline) and Z (Premium Gasoline).

# Dependent Variables

## “Engine Size (L)”

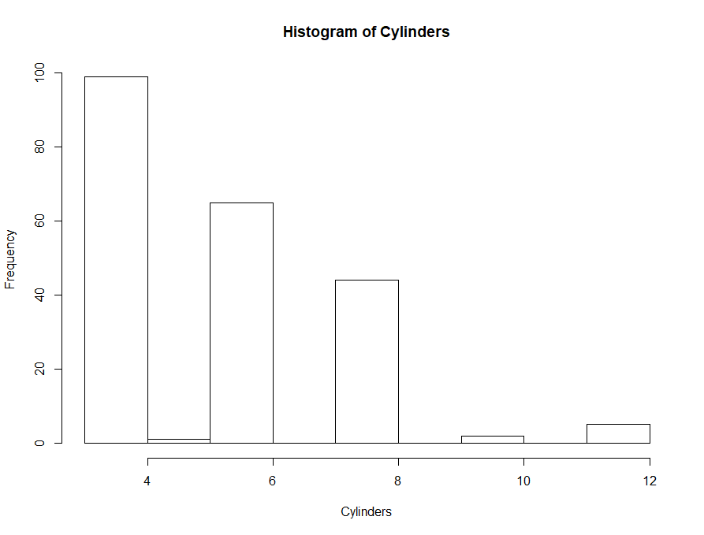
This variable denotes the size of the engine measured in liter.

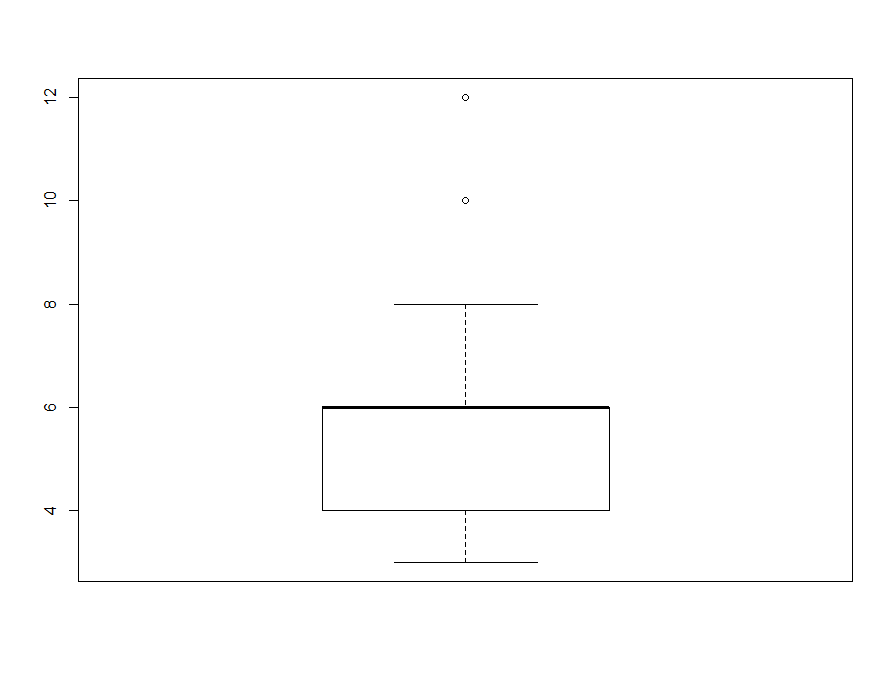


|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Mean | sd | Median | Trimmed | Mad | Min | Max | Range | Skew | Kurtosis | se |
| 3.15 | 1.35 | 3 | 3.03 | 1.48 | 1 | 6.6 | 5.6 | 0.76 | -0.46 | 0.09 |

Examining the histogram reveals that the engine size distribution in the sample is skewed to the right. The boxplot also confirms the skewness as well as existence of outliers. Looking at the statistics table, we see that mean and median are 0.15 apart which is high given the range of 5.6, further confirming the skewness, along with the actual skew statistic of 0.76. Negative kurtosis suggests flatter peak and lighter tails.

## “Cylinders”

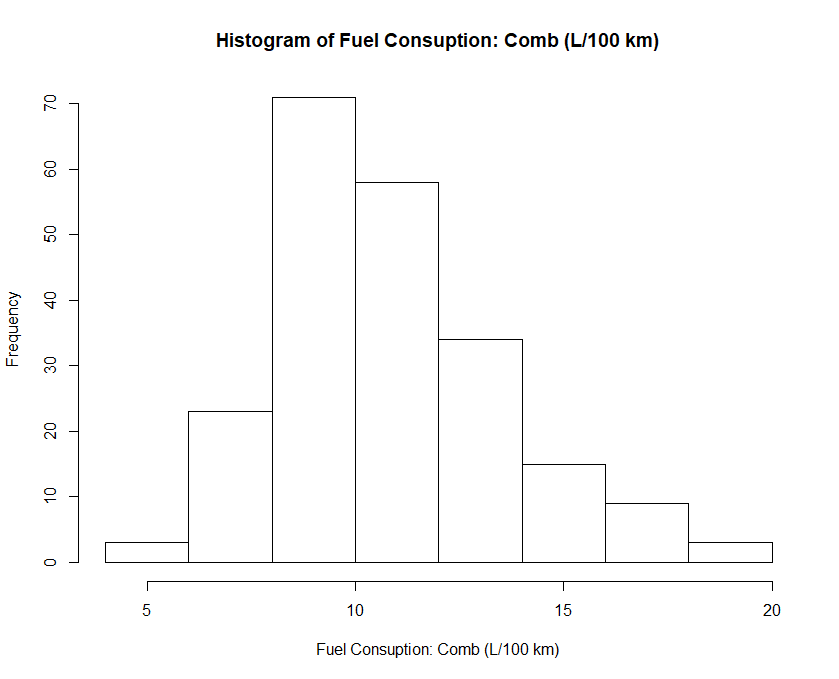
Denotes the number of cylinders in each car being sampled.



|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Mean | sd | Median | Trimmed | Mad | Min | Max | Range | Skew | Kurtosis | se |
| 5.64 | 1.9 | 6 | 5.44 | 2.97 | 3 | 12 | 9 | 1.03 | 0.94 | 0.13 |

Checking out the histogram clearly shows the distribution is completely skewed to the right. Two outliers are visible in the boxplot showing the 10 and 12 cylinders which make sense since number of cars manufactured with that many cylinders are low.

## “Fuel Consumption: Comb (L/100 km)”



|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Mean | sd | Median | Trimmed | Mad | Min | Max | Range | Skew | Kurtosis | se |
| 10.92 | 2.75 | 10.3 | 10.71 | 2.37 | 4.2 | 19.5 | 15.3 | 0.68 | 0.36 | 0.19 |