Assignment 1 - Fundamentals of Machine Learning

## For this assignment, I am using the “Cars” data from <https://corgis-edu.github.io/corgis/csv/cars/>

## Student Name: David Wilkinson

library(readr)  
cars\_sample\_data <- read\_csv("C:\\Users\\david\\OneDrive\\Documents\\Kent State University MSBA\\Fundamentals of Machine Learning June 2025\\cars\_sample\_data.csv")

## Rows: 5076 Columns: 18  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (8): Engine Information.Driveline, Engine Information.Engine Type, Engin...  
## dbl (9): Dimensions.Height, Dimensions.Length, Dimensions.Width, Engine Info...  
## lgl (1): Engine Information.Hybrid  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

View(cars\_sample\_data)

## Descriptive Statistics for Quantitative Data (City Miles Per Gallon)

library(modeest)  
#summary statistics  
summary(cars\_sample\_data$`Fuel Information.City mpg`)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 8.00 14.00 17.00 17.28 20.00 38.00

#Mode  
cars\_mode <- mlv(cars\_sample\_data$`Fuel Information.City mpg`, method='mfv')  
print(paste('Mode of city miles-per-gallon is ', cars\_mode))

## [1] "Mode of city miles-per-gallon is 17"

#Range  
range\_city\_mpg <- max(cars\_sample\_data$`Fuel Information.City mpg`)- min(cars\_sample\_data$`Fuel Information.City mpg`)  
print(paste('Range of city miles-per-gallon is ', range\_city\_mpg))

## [1] "Range of city miles-per-gallon is 30"

#Variance  
variance\_city\_mpg <- var(cars\_sample\_data$`Fuel Information.City mpg`)  
print(paste('City miles-per-gallon variance is ', variance\_city\_mpg))

## [1] "City miles-per-gallon variance is 20.0657870321847"

#Standard Deviation  
sd\_city\_mpg <- sd(cars\_sample\_data$`Fuel Information.City mpg`)  
print(paste('City miles-per-gallon standard deviation is ', sd\_city\_mpg))

## [1] "City miles-per-gallon standard deviation is 4.47948513025601"

#Interquartile Range  
iqr\_city\_mpg <- IQR(cars\_sample\_data$`Fuel Information.City mpg`)  
print(paste('City miles-per-gallon IQR is ', iqr\_city\_mpg))

## [1] "City miles-per-gallon IQR is 6"

## Descriptive Statistics for Categorical Data

This shows the frequencies of occurrence for each categorical variable, as well as their relative proportions.

driveline\_data <- table(cars\_sample\_data$`Engine Information.Driveline`)  
car\_make <- table(cars\_sample\_data$Identification.Make)  
as.data.frame(car\_make)

## Var1 Freq  
## 1 Acura 33  
## 2 AMG 10  
## 3 Aston Martin 43  
## 4 Audi 87  
## 5 Bentley 18  
## 6 BMW 133  
## 7 BMW Motorrad 58  
## 8 Buick 71  
## 9 Cadillac 200  
## 10 Chevrolet 626  
## 11 Chrysler 30  
## 12 Chrysler Group LLC 6  
## 13 Dodge 260  
## 14 Ferrari 1  
## 15 Ford 603  
## 16 GMC 368  
## 17 Grand Cherokee 48  
## 18 Honda 200  
## 19 Hyundai 188  
## 20 Infiniti 85  
## 21 Jaguar 44  
## 22 Jeep 119  
## 23 Kia 135  
## 24 Lamborghini 5  
## 25 Land Rover 35  
## 26 Lexus 52  
## 27 Lincoln 43  
## 28 Lotus 11  
## 29 Maserati 12  
## 30 Maybach 13  
## 31 Mazda 168  
## 32 Mercedes 64  
## 33 Mercedes-AMG 6  
## 34 Mercedes-Benz 7  
## 35 Mercury 28  
## 36 MINI 62  
## 37 Mitsubishi 51  
## 38 Nissan 294  
## 39 Porsche 51  
## 40 Rolls-Royce 10  
## 41 Saab 59  
## 42 Scion 26  
## 43 Subaru 105  
## 44 Suzuki 51  
## 45 Toyota 325  
## 46 Volkswagen 129  
## 47 Volvo 103

as.data.frame(driveline\_data)

## Var1 Freq  
## 1 All-wheel drive 836  
## 2 Four-wheel drive 920  
## 3 Front-wheel drive 1569  
## 4 Rear-wheel drive 1751

prop.table(car\_make)

##   
## Acura AMG Aston Martin Audi   
## 0.0065011820 0.0019700552 0.0084712372 0.0171394799   
## Bentley BMW BMW Motorrad Buick   
## 0.0035460993 0.0262017336 0.0114263199 0.0139873916   
## Cadillac Chevrolet Chrysler Chrysler Group LLC   
## 0.0394011032 0.1233254531 0.0059101655 0.0011820331   
## Dodge Ferrari Ford GMC   
## 0.0512214342 0.0001970055 0.1187943262 0.0724980299   
## Grand Cherokee Honda Hyundai Infiniti   
## 0.0094562648 0.0394011032 0.0370370370 0.0167454689   
## Jaguar Jeep Kia Lamborghini   
## 0.0086682427 0.0234436564 0.0265957447 0.0009850276   
## Land Rover Lexus Lincoln Lotus   
## 0.0068951931 0.0102442868 0.0084712372 0.0021670607   
## Maserati Maybach Mazda Mercedes   
## 0.0023640662 0.0025610717 0.0330969267 0.0126083530   
## Mercedes-AMG Mercedes-Benz Mercury MINI   
## 0.0011820331 0.0013790386 0.0055161545 0.0122143420   
## Mitsubishi Nissan Porsche Rolls-Royce   
## 0.0100472813 0.0579196217 0.0100472813 0.0019700552   
## Saab Scion Subaru Suzuki   
## 0.0116233255 0.0051221434 0.0206855792 0.0100472813   
## Toyota Volkswagen Volvo   
## 0.0640267928 0.0254137116 0.0202915682

prop.table(driveline\_data)

##   
## All-wheel drive Four-wheel drive Front-wheel drive Rear-wheel drive   
## 0.1646966 0.1812451 0.3091017 0.3449567

## Transformation of Variable

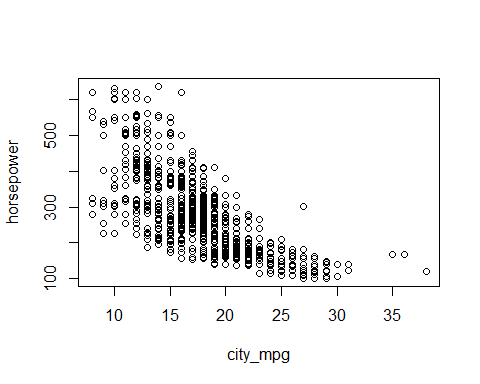
logarithm\_city\_mpg <- log(cars\_sample\_data$`Fuel Information.City mpg`)  
head(logarithm\_city\_mpg, 100)

## [1] 2.890372 3.091042 3.044522 3.044522 3.044522 2.772589 2.833213 2.564949  
## [9] 2.639057 3.091042 3.044522 2.833213 2.995732 2.890372 2.484907 2.484907  
## [17] 2.484907 2.484907 2.890372 2.890372 2.890372 2.890372 2.890372 2.890372  
## [25] 2.890372 2.995732 2.890372 2.564949 2.197225 2.564949 2.564949 2.079442  
## [33] 2.079442 2.302585 2.564949 2.708050 2.639057 2.708050 2.708050 2.708050  
## [41] 2.639057 2.890372 2.890372 2.890372 3.044522 3.044522 3.044522 3.044522  
## [49] 2.197225 2.079442 2.079442 2.079442 2.484907 2.079442 2.397895 2.302585  
## [57] 2.302585 2.302585 2.944439 2.995732 2.944439 3.044522 3.044522 3.044522  
## [65] 2.995732 2.995732 2.890372 2.890372 2.890372 2.944439 2.708050 2.772589  
## [73] 2.890372 3.044522 2.995732 2.995732 2.833213 2.890372 2.772589 2.890372  
## [81] 2.890372 2.833213 2.772589 2.772589 2.833213 2.833213 2.833213 2.833213  
## [89] 2.833213 2.772589 2.639057 2.708050 2.564949 2.890372 2.890372 2.833213  
## [97] 2.833213 2.833213 2.833213 2.833213

## ScatterPlot

This is the scatterplot illustrating the relationship between a vehicle’s horsepower and its miles per gallon in while city driving.

city\_mpg <- cars\_sample\_data$'Fuel Information.City mpg'   
horsepower <- cars\_sample\_data$`Engine Information.Engine Statistics.Horsepower`  
plot(city\_mpg, horsepower)



# Box Plot of Miles Per Gallon in City

This is a box and whisker plot of the miles per gallon (city driving) for each data point in the data set.

boxplot(city\_mpg, main= "Miles Per Gallon in the City")

