Lab5\_Kommineni\_GR022241

2025-02-27

Histogram ##

# Load the Plotly library for creating interactive visualizations.  
library(plotly)

## Loading required package: ggplot2

##   
## Attaching package: 'plotly'

## The following object is masked from 'package:ggplot2':  
##   
## last\_plot

## The following object is masked from 'package:stats':  
##   
## filter

## The following object is masked from 'package:graphics':  
##   
## layout

# Load the dplyr library for data manipulation (e.g., filtering, grouping, summarizing).  
library(dplyr)

##   
## Attaching package: 'dplyr'

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

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

#Creating a New random data

# Set a seed for reproducibility. This ensures that random data generation is consistent.  
set.seed(123)  
#Creating a data set of car sales, including mileage, price, and sales volume. The dataset will have 40 rows.  
# Create a list of real car names  
car\_names <- c("BMW M5 CS ","Toyota Camry", "Honda Accord", "Ford Mustang", "Chevrolet Silverado",   
 "Tesla Model S", "BMW X5", "Audi A4", "BMW M3 ")  
  
# Create a synthetic dataset  
car\_sales\_data <- data.frame(  
 Car\_Name = sample(car\_names, 40, replace = TRUE),   
 # This will assign car names in random   
 Year = sample(2018:2023, 40, replace = TRUE),   
 # Randomly assigning years between 2018 and 2023  
 Mileage = runif(40, min = 10, max = 50),   
 # random mileage values between 10 and 50 MPG  
 Price = runif(40, min = 20000, max = 60000),   
 # random price values between $20,000 and $60,000  
 Sales\_Volume = sample(100:500, 40, replace = TRUE)   
 # random sales volume between 100 and 500 units  
)  
# Display the first few rows of the dataset  
head(car\_sales\_data)

## Car\_Name Year Mileage Price Sales\_Volume  
## 1 Honda Accord 2018 36.91996 54545.76 448  
## 2 Honda Accord 2019 39.48311 49862.72 263  
## 3 Toyota Camry 2021 30.84543 46731.39 151  
## 4 Tesla Model S 2021 36.39354 44720.71 463  
## 5 Chevrolet Silverado 2023 42.87222 34889.52 121  
## 6 Ford Mustang 2023 41.45126 41193.43 276

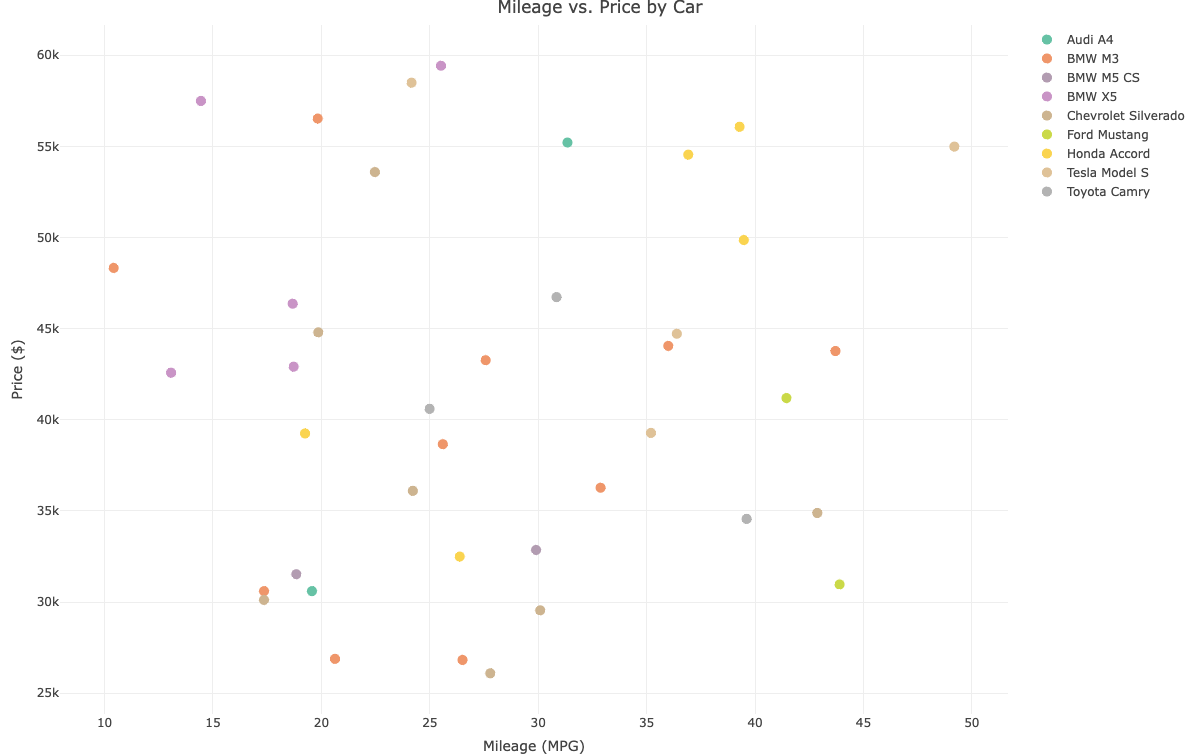
tail(car\_sales\_data)

## Car\_Name Year Mileage Price Sales\_Volume  
## 35 Toyota Camry 2020 39.61337 34563.67 490  
## 36 BMW M5 CS 2023 18.84412 31529.57 285  
## 37 BMW M3 2019 26.50984 26825.81 160  
## 38 BMW M3 2022 20.62747 26886.87 351  
## 39 Tesla Model S 2022 35.19892 39281.70 251  
## 40 Chevrolet Silverado 2020 17.35314 30118.60 418

# Create a scatter plot using Plotly.

#Scatter plot comparing mileage and price

#scatter plot comparing mileage and price  
# Creating a scatter plot using Plotly.  
scatter\_plot <- plot\_ly(  
 data = car\_sales\_data,   
 x = ~Mileage,   
 y = ~Price,   
 color = ~Car\_Name,   
 # Color points by Car\_Name to differentiate car models.  
 type = "scatter",   
 # Specify the plot type as a scatter plot.  
 mode = "markers",   
 # Use markers to represent data points.  
 text = ~paste("Car:", Car\_Name, "<br>Year:", Year),   
 # Add hover text to display car name and year.  
 marker = list(size = 10)   
 # Customize marker size for better visibility.  
) %>%  
 layout(  
 title = "Mileage vs. Price by Car",  
 # Adding title to the plot and labels.  
 xaxis = list(title = "Mileage (MPG)"),   
 yaxis = list(title = "Price ($)")   
 )  
  
# Display the scatter plot.



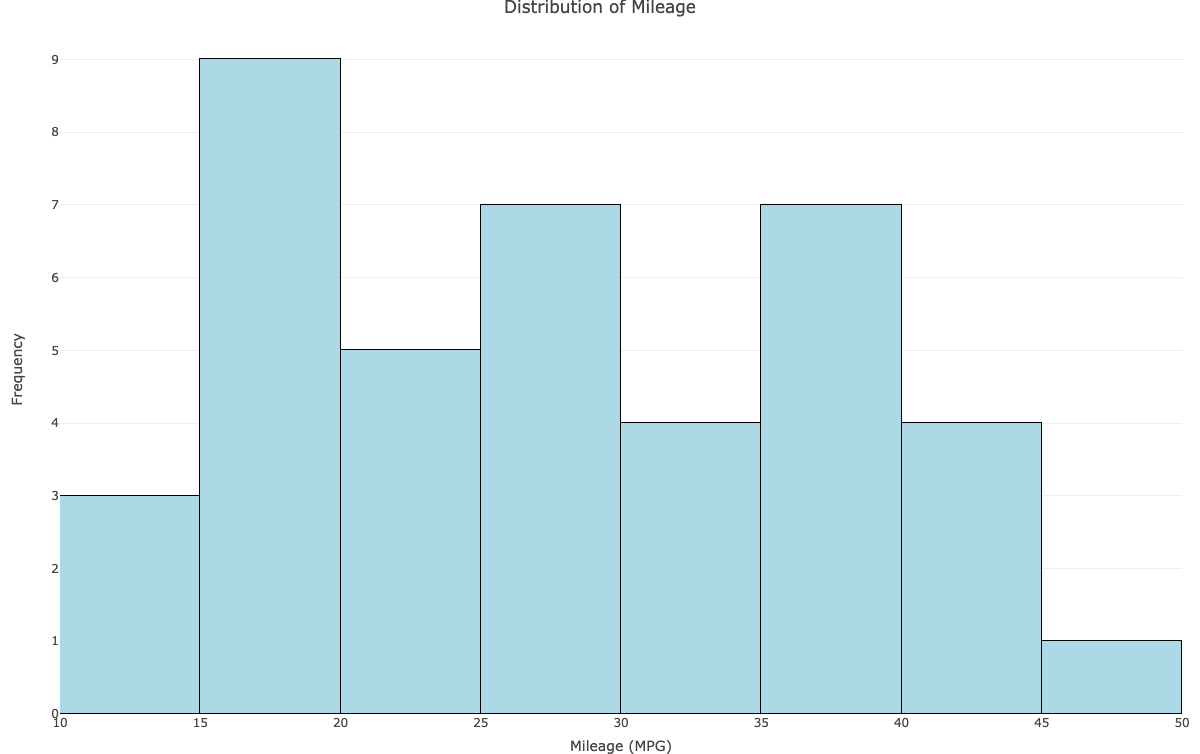
#Findings  
#Generally speaking, cars with higher mileage (better fuel efficiency) are typically less expensive. Given that fuel-efficient vehicles are frequently less expensive, this is to be expected.  
#The Tesla Model S, are electric vehicles they have high pricing and high mileage.  
#Vehicles such as the Honda Accord and Toyota Camry are grouped in the middle of the price and mileage ranges, they have good fuel economy.

#Histogram

#histogram shows of Mileage all cars.

{r}

# Create a histogram using Plotly.  
histogram <- plot\_ly(  
 data = car\_sales\_data,  
 # Using the car\_sales\_data dataset foom above .  
 x = ~Mileage, # Set the x-axis to Mileage.  
 type = "histogram", # Specify the plot type as a histogram.  
 nbinsx = 10,   
 # Set the number of bins to 10 for the histogram.  
 marker = list(color = "lightblue", line = list(color = "black", width = 1))   
 # Customize bar color and border.  
) %>%  
 layout(  
 title = "Distribution of Mileage",  
 # Add a title to the plot anf labels.  
 xaxis = list(title = "Mileage (MPG)"),   
 yaxis = list(title = "Frequency")   
 )  
# Display the histogram.  
histogram

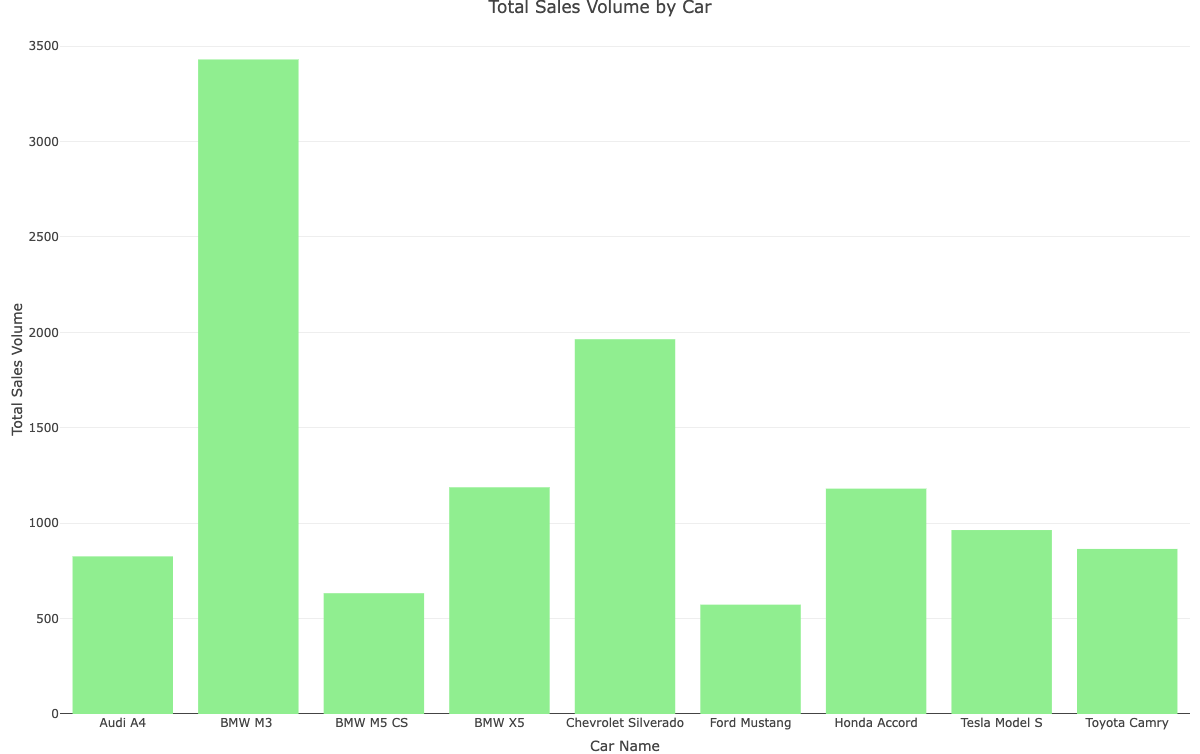


#Findings  
# Most cars have a mileage of between 25 and 40 miles per gallon, The 30-35 MPG range has the largest frequency of cars, suggesting that this is the dataset's most prevalent fuel economy level.Extreme instances are represented by a small number of cars with extremely low mileage (less than 15 MPG) or extremely high mileage (more than 45 MPG).

#Bar

Chart #Bar chart shows the total Sales of diffrent Cars

# Aggregate sales volume by car name.  
sales\_by\_car <- car\_sales\_data %>%  
 group\_by(Car\_Name) %>%   
 # Group the data by Car\_Name.  
 summarize(Total\_Sales = sum(Sales\_Volume))  
# Calculate total sales for each car.  
  
# Create a bar chart using Plotly.  
bar\_chart <- plot\_ly(  
 #total Sales\_Volume vs Car\_Name  
 data = sales\_by\_car,   
 x = ~Car\_Name,   
 y = ~Total\_Sales,   
 type = "bar",   
 # Specify the plot type as a bar chart.  
 marker = list(color = "lightgreen")   
 # bar color.  
) %>%  
 layout(  
 title = "Total Sales Volume by Car", # Adding a title to the plot and lables .  
 xaxis = list(title = "Car Name"),   
 yaxis = list(title = "Total Sales Volume")   
 )  
  
# Display the bar chart.  
bar\_chart



#Findings   
#As it is a random data the cars like BMW M3 showing huge sale despite of its less mileage