**Case Study – Exploratory Data Analysis (Jewels Wolter)**

The goal of this assignment is to analyze the Fuel Economy Dataset to understand vehicle characteristics, efficiency, and performance trends using summary statistics and visualization techniques.

Questions:

# Part 1: Summary Statistics (10pts)

1. Overall Analysis:
   1. Calculate the mean, median, and standard deviation for mpg, horsepower, and weight.

*Output:*

**Summary Statistics:**

**mpg: mean = 23.445918367346938 , median = 22.75 , std dev = 7.795045762682584**

**horsepower: mean = 104.46938775510205 , median = 93.5 ,std dev = 38.44203271442593**

**weight: mean = 2977.5841836734694 , median = 2803.5 , std dev = 848.3184465698362**

**Average mpg by origin:**

**europe 27.602941**

**japan 30.450633**

**usa 20.033469**

**Average horsepower by origin:**

**europe 80.558824**

**japan 79.835443**

**usa 119.048980**

**Average weight by origin:**

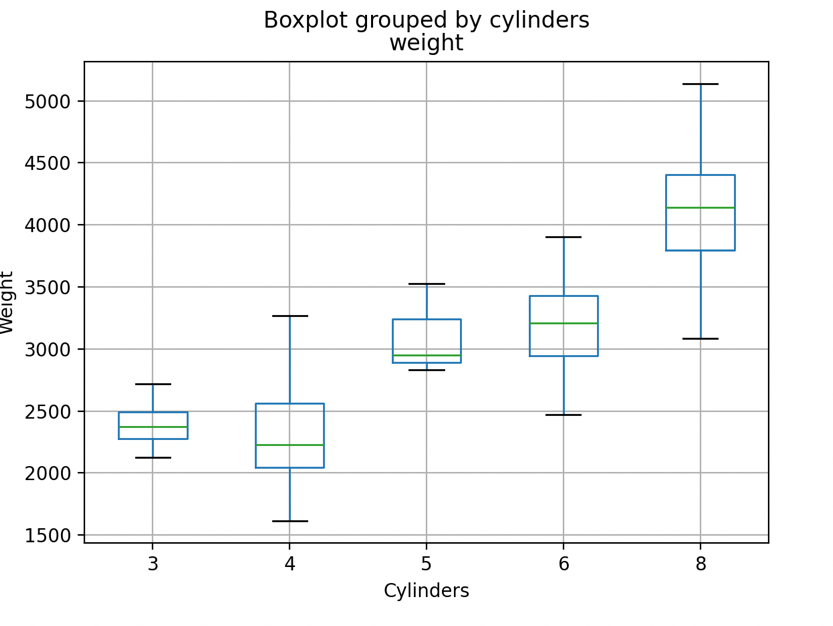
**europe 2433.470588**

**japan 2221.227848**

**usa 3372.489796**

1. Grouped Analysis:
   1. Group the data by origin and calculate the average mpg, horsepower, and weight for each region.
   2. Which region has the most fuel-efficient cars on average?
      1. **Japan, I found this by taking the average MPG for each region and finding the maximum.**

# A graph of a graph AI-generated content may be incorrect.Part 2: Data Visualization (15pts)

1. Univariate Analysis:
   1. Create a histogram of mpg. What is the most common range of fuel efficiency?
      * **16.5 to 19.1**
      * **The histogram shows a bar with the greatest height—and therefore greatest frequency—in this range.**
   2. Create a boxplot of weight grouped by cylinders. What trends do you observe in vehicle weight based on the number of cylinders?
      * **As the vehicle weight increases, the median number of cylinders also increases.**
2. A graph of pink dots

   AI-generated content may be incorrect.Bivariate Analysis:
   1. Create a scatterplot of horsepower vs mpg. Does higher horsepower generally relate to lower fuel efficiency?
      * **For the most part, the higher the horsepower, the lower the miles per gallon (and hence, fuel efficiency)**
      * **The scatter plot is right-skewed**

A diagram of weight versus mpg

AI-generated content may be incorrect.

* 1. Create a scatterplot of weight vs mpg. Discuss the relationship between vehicle weight and fuel efficiency.
     + **For the most part, the higher the weight of the vehicle the lower the miles per gallon (and hence fuel efficiency)**
     + **The scatter plot is right-skewed**

1. A graph with purple line and numbers

   Description automatically generatedTrend Analysis:
   1. According to line plot of the average mpg over model year. How has fuel efficiency changed over time?
      * **Over time, the average miles per gallon has increased almost every new model year.**

**Code Appendix:**

# Case Study - Exploratory Data Analysis

# Author - Jewels Wolter

# STAT 3010 - Statistics for Engineers and Scientists

# Import Libraries

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

# P0: Data Import

data = pd.read\_csv('mpg.csv')

[# P1: Summary Statistics](#_Part_1:_Summary)

# calculate five number summary

def fiveNumSum(x):

mean = np.mean(x)

median = np.median(x)

min = np.min(x)

max = np.max(x)

q1 = np.percentile(x, 25)

q3 = np.percentile(x, 75)

stdDev = np.std(x)

return mean, median, min, q1, q3, max, stdDev

# calculate for each column

mpg = data['mpg']

horsepower = data['horsepower']

weight = data['weight']

mpgStats = fiveNumSum(mpg)

horsepowerStats = fiveNumSum(horsepower)

weightStats = fiveNumSum(weight)

# group data by origin

origin = data.groupby('origin')

#calculate average mpg, horsepoweer, and weight for each origin

avgMpg = origin['mpg'].mean()

avgHorsepower = origin['horsepower'].mean()

avgWeight = origin['weight'].mean()

# print summary statistics

print('Summary Statistics:')

print('mpg: mean =', mpgStats[0], ', median =', mpgStats[1], ', std dev =', mpgStats[6])

print('horsepower: mean =', horsepowerStats[0], ', median =', horsepowerStats[1], ',std dev =', horsepowerStats[6])

print('weight: mean =', weightStats[0], ', median =', weightStats[1], ', std dev =', weightStats[6])

print('Average mpg by origin: \n', avgMpg)

print('Average horsepower by origin: \n', avgHorsepower)

print('Average weight by origin: \n', avgWeight)

print('Reigon with most fuel-efficient cars:', avgMpg.idxmax())

[# P2: Data Visualization](#_Part_2:_Data)

# P2.1: Univariate Analysis

# create a histogram of mpg

plt.hist(mpg, bins=15, color='mediumvioletred', edgecolor='black')

plt.title('Histogram of MPG')

plt.xticks(np.arange(10, 50, 2))

plt.xlabel('MPG')

plt.ylabel('Frequency')

plt.show()

# create a boxplot of weight grouped by cylinders

data.boxplot(column='weight', by='cylinders')

plt.xlabel('Cylinders')

plt.ylabel('Weight')

plt.show()

# # P2.2: Bivariate Analysis

# create a scatterplot of horsepower vs mpg

plt.scatter(horsepower, mpg, color='hotpink')

plt.title('Horsepower vs MPG')

plt.xlabel('Horsepower')

plt.ylabel('MPG')

plt.show()

# create a scatterplot of weight vs mpg

plt.scatter(weight, mpg, color='deeppink')

plt.title('Weight vs MPG')

plt.xlabel('Weight')

plt.ylabel('MPG')

plt.show()