Homework 1

Jayaganeshan Thanga Kumar

Introduction

This is a simple exploratory data analysis of the `mpg` data, readily available in the `ggplot2` R package. The `mpg` data contain car's fuel economy data on 38 best-selling models from 1999 through 2008. It contains a number of variables such as manufacturer, model, engine size (`displ`), year, number of cylinders (`cyl`), type of transmission (`trans`), whether front-wheel, rear-wheel, or 4-wheel (`drv`), miles per gallon in the city (`cty`), miles per gallon on the highway (`hwy`), fuel (`fl`), and car class (`class`). Primarily, I am interested in plotting **engine size versus highway fuel efficiency by car class**.

Approach

I will bring in the `mpg` data from the `ggplot2` package and use it straight away. I will examine the central variables I am interested in plotting (`displ`, `hwy`, and `class`), perform some **descriptive statistics** on those, and then draw a pretty **scatter plot** with `ggplot2`. It will enable us to visualize how highway mileage is connected with engine size and how different car classes might influence this connection. It is an easy process with the goal of providing **quick insights**.

Descriptive Statistics

Before constructing the plot, here are some descriptive statistics for the variables I will be graphing: `displ` (engine size), `hwy` (miles per gallon on the highway), and `class` (car class).

```
--- Summary for Engine Displacement (displ) ---

  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.600   2.400   3.300   3.472   4.600   7.000
```

```
--- Summary for Highway Miles Per Gallon (hwy) ---

   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  12.00   18.00   24.00   23.44   27.00   44.00


--- Frequency Table for Vehicle Class (class) ---


  2seater    compact   midsize   minivan    pickup subcompact       suv
        5         47        41        11        33         35        62


--- Head of the selected dataset structure ---
```

displ	hwy	class
1.8	29	compact
1.8	29	compact
2.0	31	compact
2.0	30	compact
2.8	26	compact
2.8	26	compact

From the statistics above, we can see the range and distribution of **engine displacement** and **highway MPG**. For example, `displ` ranges from **1.6 to 7.0 liters**, while `hwy` ranges from **12 to 44 MPG**. The **class** variable provides categorical information about the vehicle type, showing the counts for each category, with **SUV** and **compact** being the most frequent.

Plot: Engine Displacement vs. Highway MPG

The following plot is **engine displacement** (`displ`) versus **highway MPG** (`hwy`), colored by vehicle class. Engine displacement normally refers to the sum of the volume swept by all pistons of an engine. Larger engines are less efficient in burning fuel, so I would expect this to be reflected in the plot. **Highway MPG** (Miles Per Gallon) is a typical measure of a car's highway fuel economy.

Highway MPG vs. Engine Displacement by Vehicle Class
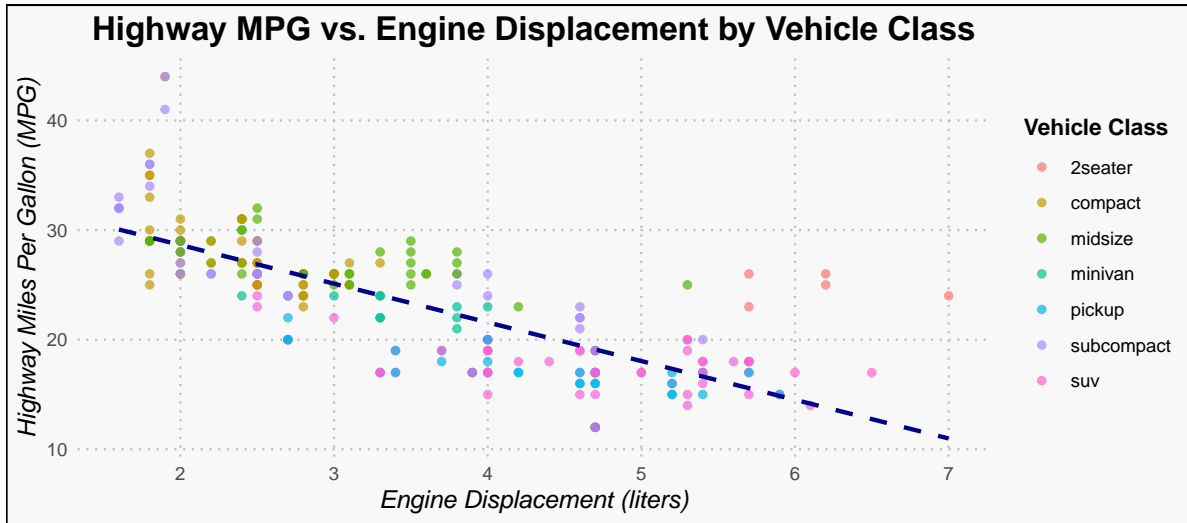
Plot Description

From the graph, we can see that **highway MPG** is negatively correlated with **engine displacement**. Vehicles with larger engine displacements (larger values of `displ`) will enjoy lower **highway MPG**. This is reasonable, of course, since larger engines will consume a larger amount of fuel to operate.

Also, **class coloring** reveals some interesting trends. **Two-seaters** and **subcompacts** achieve better highway MPG per engine size, while **pickup** and **SUV** cars achieve worse MPG, quite possibly due to their weights being somewhat larger, their frontal area being larger, and their bodies relatively less aerodynamic. A general linear relation plotting this inverse is represented in the dashed blue line.

This graph is good at showing the effect of types of cars on mileage per engine size.