MPG Data Visualization Report

Fluke Siwakorn

2026-01-01

Analyze MPG Dataset

purpose: It is primarily used to analyze the relationship between vehicle design (engine size, drive type) and environmental efficiency.

Summary of dataset

The mpg dataset, contained within the ggplot2 package, provides fuel economy data from the US Environmental Protection Agency (EPA) for popular car models between 1999 and 2008.

```
summary(mpg)
```

```
##  manufacturer        model              displ           year
##  Length:234        Length:234        Min.   :1.600   Min.   :1999
##  Class :character  Class :character  1st Qu.:2.400   1st Qu.:1999
##  Mode  :character  Mode  :character  Median :3.300   Median :2004
##                                      Mean   :3.472   Mean   :2004
##                                      3rd Qu.:4.600   3rd Qu.:2008
##                                      Max.   :7.000   Max.   :2008
##       cyl           trans               drv                cty
##  Min.   :4.000   Length:234        Length:234        Min.   : 9.00
##  1st Qu.:4.000   Class :character  Class :character  1st Qu.:14.00
##  Median :6.000   Mode  :character  Mode  :character  Median :17.00
##  Mean   :5.889                                       Mean   :16.86
##  3rd Qu.:8.000                                       3rd Qu.:19.00
##  Max.   :8.000                                       Max.   :35.00
##       hwy             fl               class
##  Min.   :12.00   Length:234        Length:234
##  1st Qu.:18.00   Class :character  Class :character
##  Median :24.00   Mode  :character  Mode  :character
##  Mean   :23.44
##  3rd Qu.:27.00
##  Max.   :44.00
```

1. Engine Size vs. Fuel Efficiency

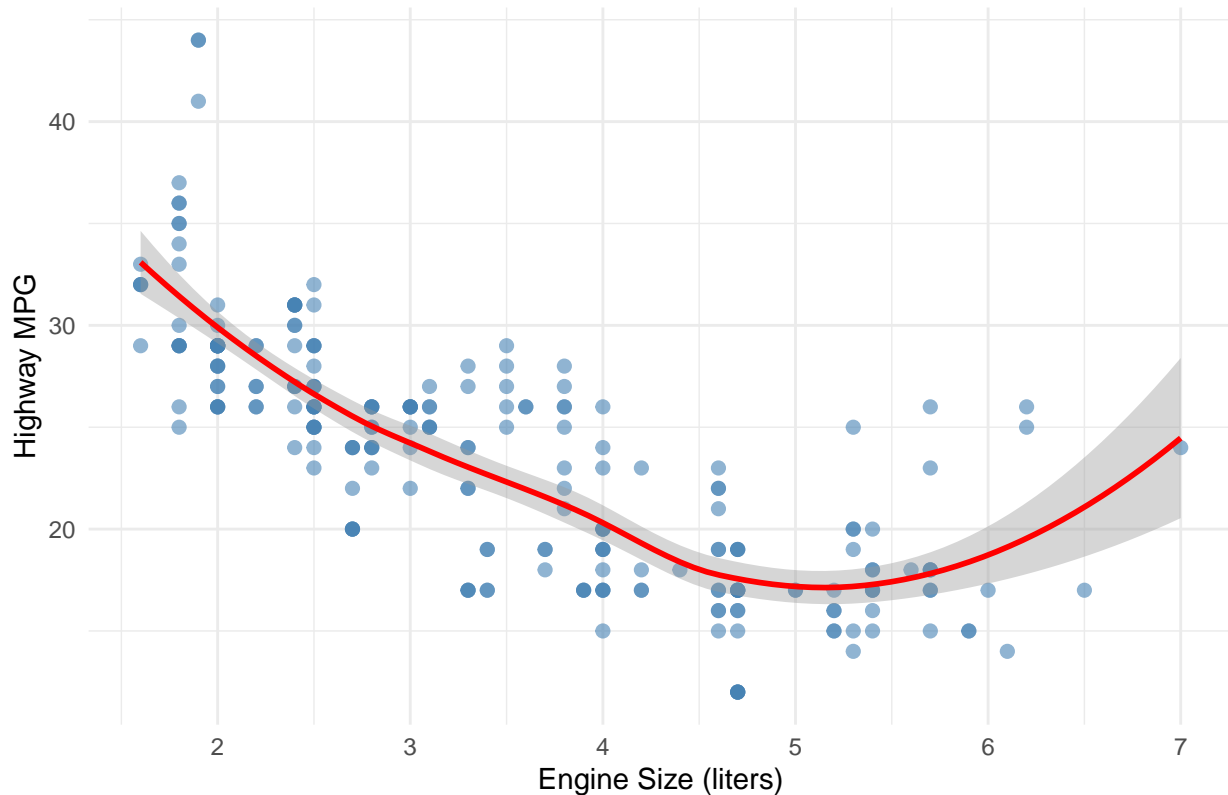**Scatter Plot**   Question: How does engine displacement relate to highway fuel economy?

- Code Strategy: Used geom_point() for relationship and geom_smooth() to visualize the trend line.

```
ggplot(data = mpg, mapping = aes(x = displ, y = hwy)) +
  geom_point(alpha = 0.6, size = 2, color = "steelblue") +
  geom_smooth(method = "loess", color = "red") +
  theme_minimal() +
```

```
    labs(title = "Engine Displacement vs. Highway MPG",
         x = "Engine Size (liters)", y = "Highway MPG")
```

`geom_smooth()` using formula = 'y ~ x'

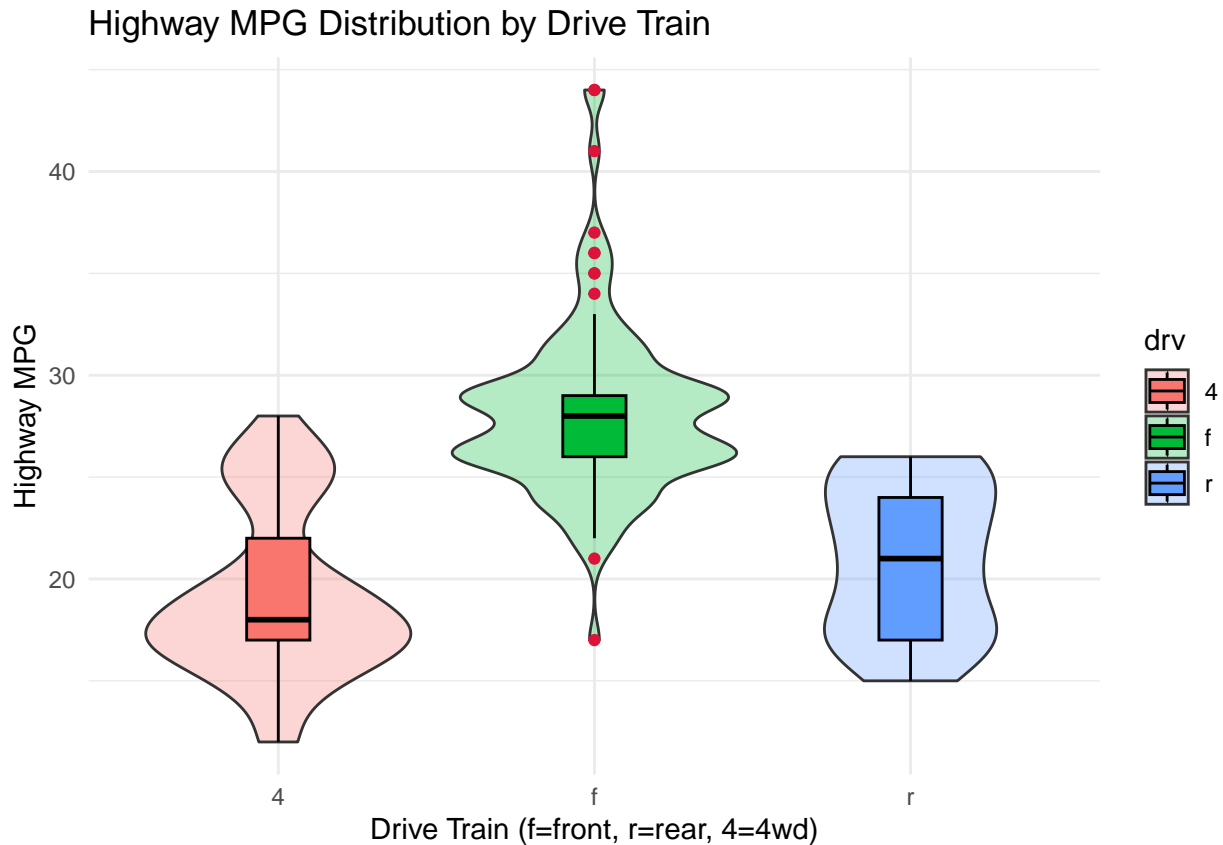### Engine Displacement vs. Highway MPG

Summary: The plot shows a clear negative correlation; as engine size increases, highway fuel efficiency tends to decrease. Smaller engines (below 2 liters) demonstrate the highest efficiency.

2. Drive Train Performance

Box & Violin Plot   Question: How does the distribution of highway MPG vary across different drive train types?

• Code Strategy: Combined geom_violin() for density shape and geom_boxplot() for statistical summaries (median/quartiles).

```
ggplot(data = mpg, mapping = aes(x = drv, y = hwy, fill = drv)) +
  geom_violin(alpha = 0.3) +
  geom_boxplot(width = 0.2, color = "black", outlier.color = "#DC143C") +
  theme_minimal() +
  labs(title = "Highway MPG Distribution by Drive Train",
       x = "Drive Train (f=front, r=rear, 4=4wd)", y = "Highway MPG")
```

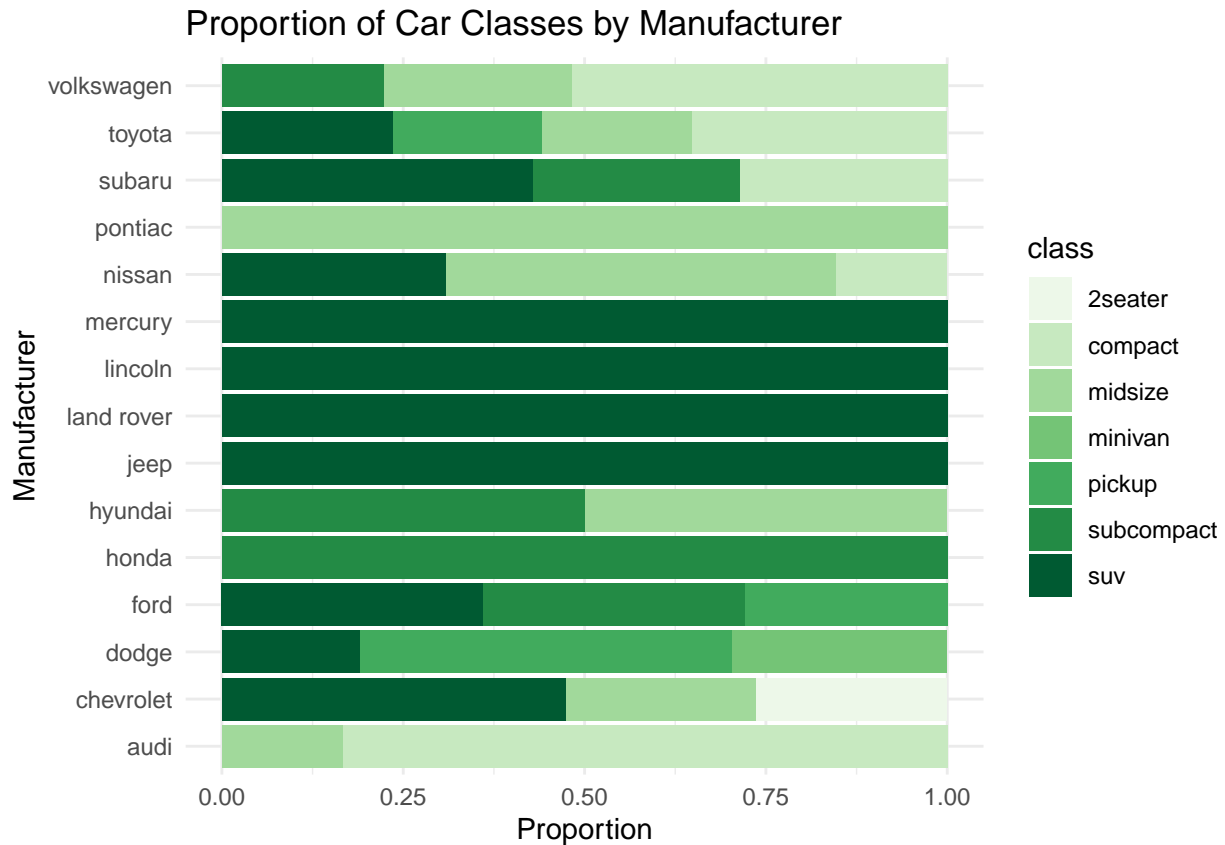## Highway MPG Distribution by Drive Train

Summary: Front-wheel drive (f) vehicles consistently outperform others in efficiency, showing both a higher median MPG and a wider range of high-performing models compared to 4-wheel (4) or rear-wheel (r) drives.

3. Brand Portfolio Composition

Stacked Bar Chart   Question: What proportion of different car classes does each manufacturer produce?

• Code Strategy: Used geom_bar(position = "fill") to compare relative proportions rather than absolute counts.

```
ggplot(data = mpg, mapping = aes(x = manufacturer, fill = class)) +
  geom_bar(position = "fill") +
  coord_flip() +
  scale_fill_brewer(palette = "Set4") +
  theme_minimal() +
  labs(title = "Proportion of Car Classes by Manufacturer",
       x = "Manufacturer", y = "Proportion")
```
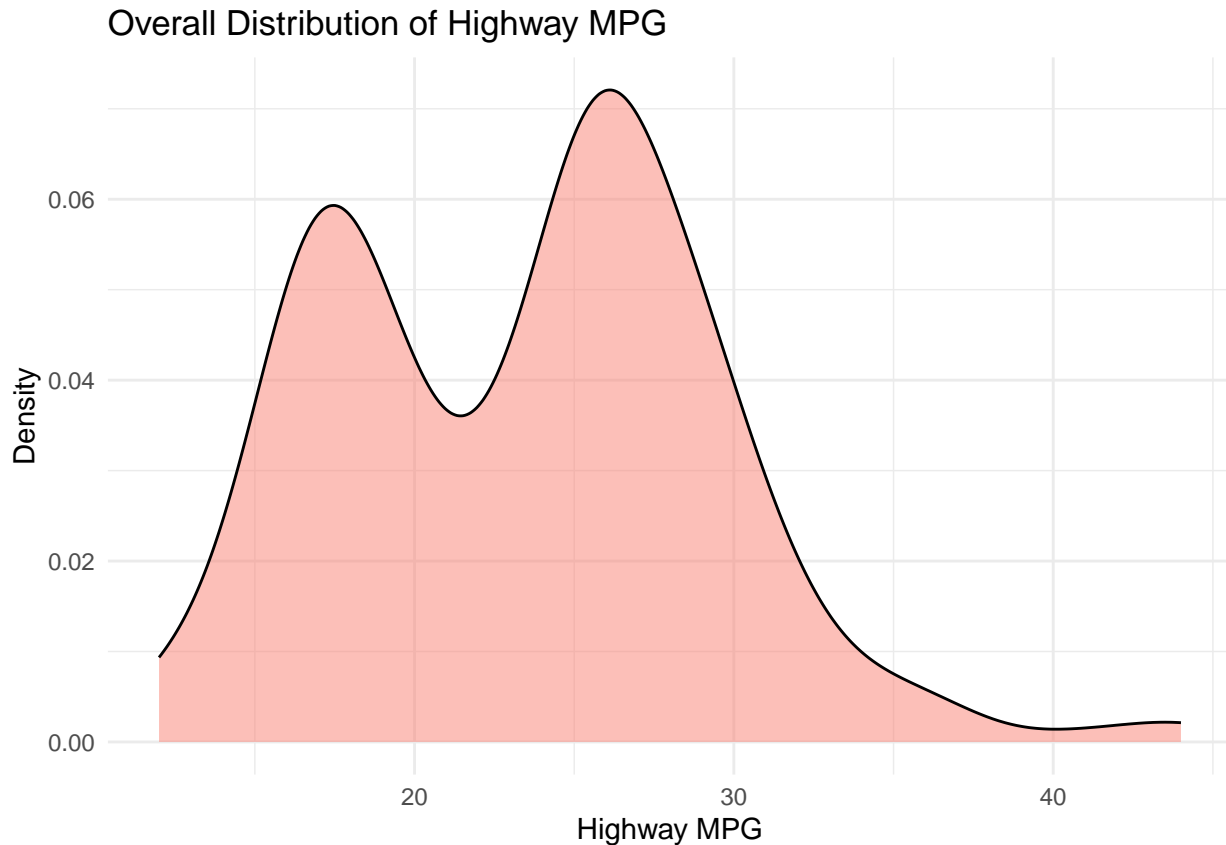
Proportion of Car Classes by Manufacturer

Summary: The chart highlights market specialization; some brands like Land Rover focus exclusively on SUVs, while others like Toyota maintain a balanced portfolio across subcompact, midsize, and SUV categories

4. Overall Efficiency Distribution

Density Plot   Question: At what MPG value does the majority of the vehicle population cluster?

- Code Strategy: Employed geom_density() to create a smooth curve of the highway MPG distribution.

```
ggplot(data = mpg, mapping = aes(x = hwy)) +
  geom_density(fill = "salmon", alpha = 0.5) +
  theme_minimal() +
  labs(title = "Overall Distribution of Highway MPG",
       x = "Highway MPG", y = "Density")
```
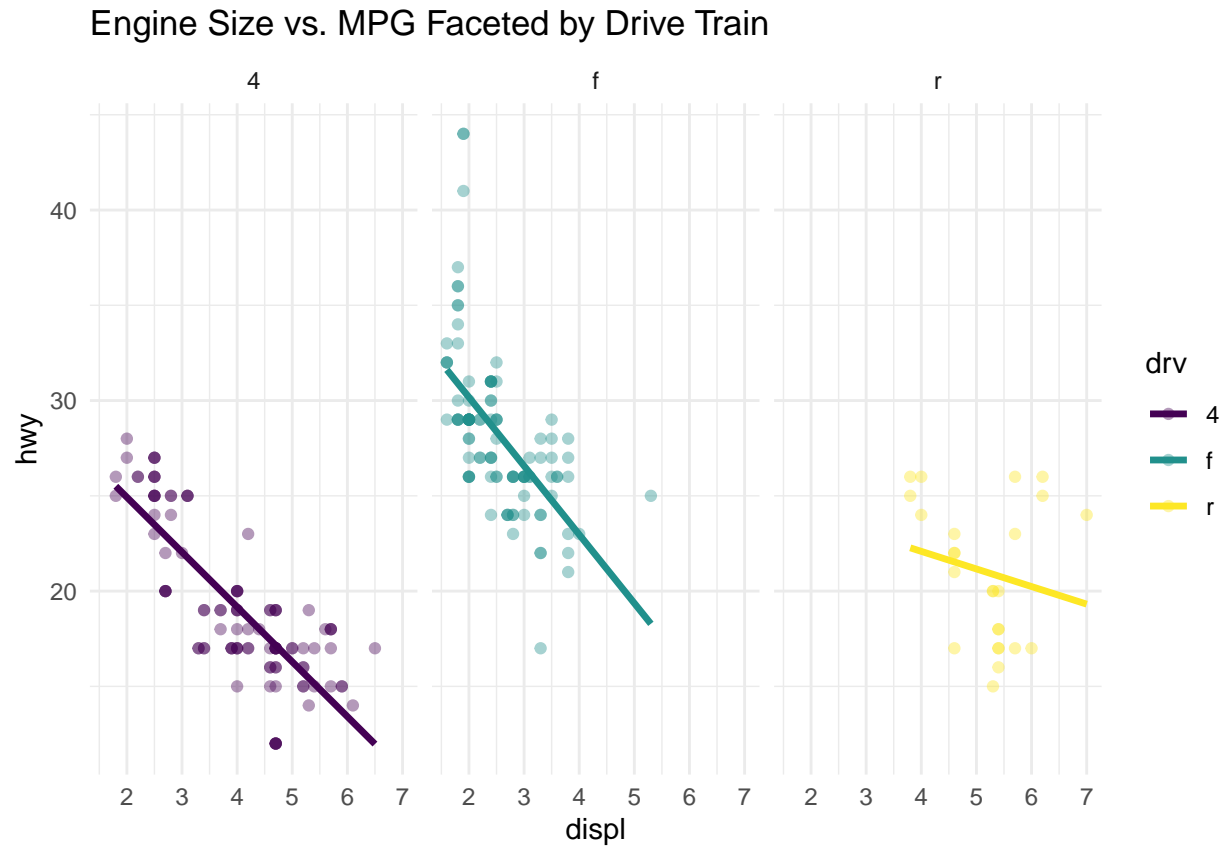
Overall Distribution of Highway MPG

Summary: The distribution is bimodal (two-peaked), suggesting the market is split into two main groups: a "heavy/powerful" group peaking around 18-20 MPG and an "economical" group peaking around 26-30 MPG.

5. Segmented Trends

Facet Wrap   Question: Does the engine-to-efficiency relationship hold true across all drive trains?

- Code Strategy: Utilized facet_wrap(~drv) to split the data into three distinct panels for direct comparison.

```
ggplot(data = mpg, mapping = aes(x = displ, y = hwy, color = drv)) +
  geom_point(alpha = 0.4) +
  facet_wrap(~drv) +
  geom_smooth(method = "lm", se = FALSE, size = 1.2) +
  scale_color_viridis_d(option = "viridis") +
  theme_minimal() +
  labs(title = "Engine Size vs. MPG Faceted by Drive Train")
```

Engine Size vs. MPG Faceted by Drive Train

Summary: Faceting shows that while the negative trend is universal, the "starting point" for efficiency is much higher for front-wheel drives. Rear-wheel drives are almost exclusively large-engine vehicles with lower efficiency.

. . .

End of Report