STAT40830 - Homework 1

Akshaykumar Musterya- 24228316

Dataset Chosen - Cars93

Overview

The *Cars93* dataset includes specifications and pricing information for **93 car models** from the year 1993. Each entry describes a unique car model with attributes such as *price*, *horsepower*, *origin* (USA or non-USA), and *type* (e.g., *Small*, *Midsize*, *Sporty*).

In this report, we explore how a car's *horsepower* relates to its *price*, and examine whether characteristics like *origin* and *vehicle type* affect pricing patterns. Our visualizations and descriptive statistics aim to uncover meaningful insights into *automotive market trends* of the early 1990s.

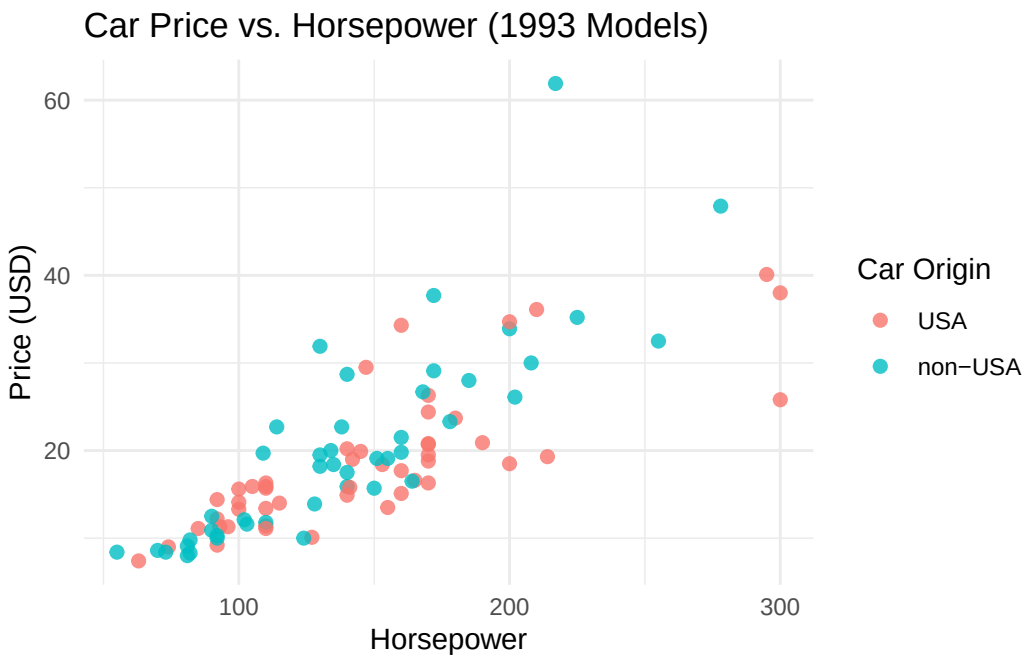
Understanding Dataset

The `Cars93` dataset (from the MASS package) contains specifications for 93 car models released in 1993. Each row represents a different car model and includes attributes such as:

- **Price** (in $1000s): Midrange market price
- **Horsepower**: Engine power output
- **Origin**: Whether the car is manufactured in the USA or not
- **Type**: Car category such as Small, Midsize, Sporty, Large, etc.

Scatterplot: Price vs. Horsepower

We visualize the relationship between **horsepower** and **price**, with points colored by **origin** (USA or non-USA). This helps us understand whether engine power and regional origin influence car prices.



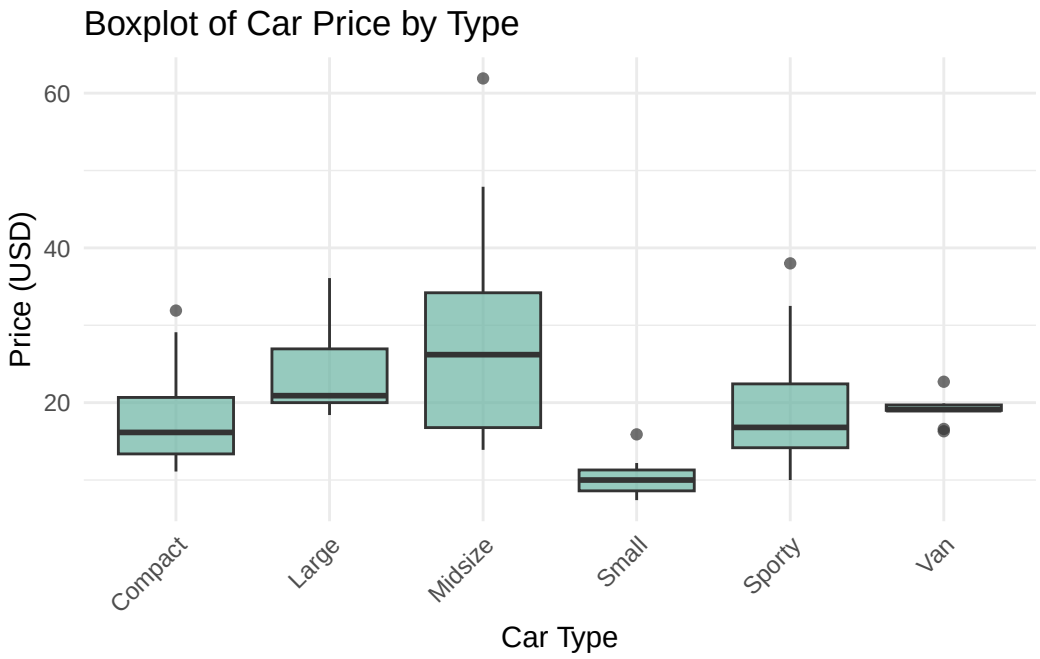
Interpretation

From the scatterplot, we observe the following patterns:

- There is a **positive relationship** between *horsepower* and *price*: cars with more powerful engines tend to cost more.
- **USA cars** show a wider spread in prices, including both affordable and premium options for similar horsepower.
- **Non-USA cars** generally cluster around the *midrange* in both price and power.
- A few **outlier models** with moderate horsepower but very high price may reflect *luxury or brand influence*.

These insights suggest that **engine power is a good predictor of car price**, but **origin** and **brand effects** also play an important role in pricing strategy.

Boxplot of Car Price by Type



Interpretation

From the boxplot, we observe the following:

- **Large** and **Sporty** cars tend to have **higher price ranges** and greater **variability**.
- **Small** and **Compact** cars are generally **more affordable** and have **tighter price distributions**.
- The boxplot also reveals **outliers**, especially in the *Large* and *Sporty* categories, suggesting the presence of **high-end luxury models**.

From the boxplot, we observe that **car type has a strong influence on price**. Larger cars and sporty models show more variability, while compact cars remain in a tighter price range. Outliers suggest luxury offerings in certain categories.