Exploratory Data Analysis in Auto-Mpg Dataset

The Outliers

Spencer Christensen, Josh Hunter, Peter Kaull, Dean Lang, Marcin Tutaj

MATH/COSC 3570 Dr. Yu 3/7/25

Description of the dataset:

This dataset contains data from cars produced between the model years of 1970 and 1982, taken from the StatLib library at Carnegie Mellon University. Attributes include, but are not limited to mpg, horsepower, weight, and cylinders.

Goal of the project:

The goal of our project is to find out what attributes correlate to longer driving performance vehicles. In this example, we have denoted “mpg” (miles per gallon) as the target variable. Through the course of this project we will try to understand what contributes most to a car’s miles per gallon.

Member’s duties:

- Each person will make one visualization or more to their liking, analyzing the correlation between their designated variable and “mpg”.
    - We will start by making a correlation plot between mpg and every variable and then each team member will take one of the top 5 correlated variables to make their visualization.
        - mpg vs. Weight, mpg vs. Cylinders, mpg vs. Acceleration, etc…
- Marcin will be in charge of coherence of the slides.
- Peter and Spencer will be in charge of analysis of the slides.
- Dean and Josh will make sure the slides look appealing.