Impact of transmission type on mileage

Executive Summary

In this report, we examine the relationship between transmission of a car and the miles per gallon. The data was extracted from the 1974 Motor Trend magazine and is available as a part of the R default datasets. After analyzing several different models, we determine that the best model for predicting miles per gallon includes transmission, weight, horsepower and an interaction between transmission and weight. It is not possible to conclude which transmission type is better in isolation. For heavier cars, an automactic transmission might be prefered. The model doesn't take into account any extrinsic features like road conditions.

Exploratory Analysis

After loading the data from the mtcars dataset, we create exploratory plots to examine the relationships between miles per gallon (mpg) and number of cylinders(cyl), transmision(am), weight(wt) and gross horsepower(hp).

Figure 1 & 2 in the appendix shows the plots. The mean mpg is also shown on the plot. The plot shows that manual transmission cars have a higher median mpg than automatic transmission cars. It is also evident that the weight, number of cylinders and gross horsepower negatively impact the miles per gallon.

Regression Models

In order to examine the relationship between mpg and transmission, a nested model approach is used. Several models are created starting with the simple model of transimission as predictor. The nested variables for the models are weight, interaction between weight and transmission and number of cylinders(or horsepower).The gross horsepower and number of cylinders have a correlation of 0.8324.Since they highly correlated, both of these variables are not included in any models.

As a result, we end up with five models. The anova function is used to analyse the models. The inclusion of weight and interaction between weight and transmision reduces the sum of squared residuals. Finally, the model with the number of cylinders and horse power are compared to determine the optimal model.

The sum of residuals squared for the model with cylinders (RSS:137.9917) is lower than the model with horsepower(RSS:146.845). However, the p-value for the model with horspower (0.0105) is smaller than the p-value for the model with the cylinder (0.0179). Therefore, we choose the model which includes the horsepower as the best fit. The Adjusted R-squared value for this model is 0.8503. The model is described as (using R style notation):

mpg ~ am + wt + am:wt + hp

Residual Analysis

The residuals are analysed by generating four different plots as shown in figure 3. We can derive the following conclusions from the plots.

1. The residuals are scattered across fitted values and there aren't any patterns in the residuals.
2. The standardized residuals follow the standard normal quantiles for smaller values. For a some of the larger values, the standardized residuals drift away from the theoretical quantiles. This indicates that there are a few outliers in the data.
3. The plot of the residuals v/s leverage confirms the presense of a few outliers that have leverage on the model.

There is a reasonable amount variation in the hatvalues. This also confirms the presence of a couple of outliers.

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0554  0.0768  0.1130  0.1560  0.2100  0.4830
```

Since the plot is part of the R standard library package, we conclude that none of the outliers are due to data entry errors and include all of them in our model.

We have also shown the plot of density of the residuals in figure 4. It closely resembles a standard normal distribution.

Results

From the model, we can conclude that transmission type is significant in determining the mileage offered by a car. Performing theoretical analysis, for a car with zero weight and zero horsepower, the expected mpg for a car with automatic transmission is 30.9473. The expected mpg for a car with manual transmission under same settings is 42.5021. This seems to indicate that manual transmissions are better than automatic transmission.

However, it also noticeable that the mpg reduces as the weight and horsepower of the car increase. The interaction between weight and transmission is also important in determining the change in mpg as other parameters change. For an automatic transmission, the mpg reduces by -2.5156 for an increase of 1000 lbs in car weight (holding horsepower constant). The decrease in mpg per 1000 lb increase in weight for a manual transmission is -6.0935. This indicates that mpg decreases more rapidly for manual transmission cars as the weight increases.

Therefore, the transmission type alone cannot determine the miles per gallon of a car. It is not possible to determine which transmission type is better since other variables play a significant role in determine the mileage.

Our model has several limitations. We highlight them below:

1. The model was built using only 32 cars in the dataset. This doesn't represent the complete set of cars available in the market.
2. The dataset was compiled together from a 1974 publication. There have been significant advances in technology which would impact miles per gallon.
3. There is no mention of how the miles per gallon was calculated in the dataset. The mileage offered by a car can vary significantly depending on road conditions, highway vs city driving & weather conditions. None of these factors were taken into account in the model.

Apendix

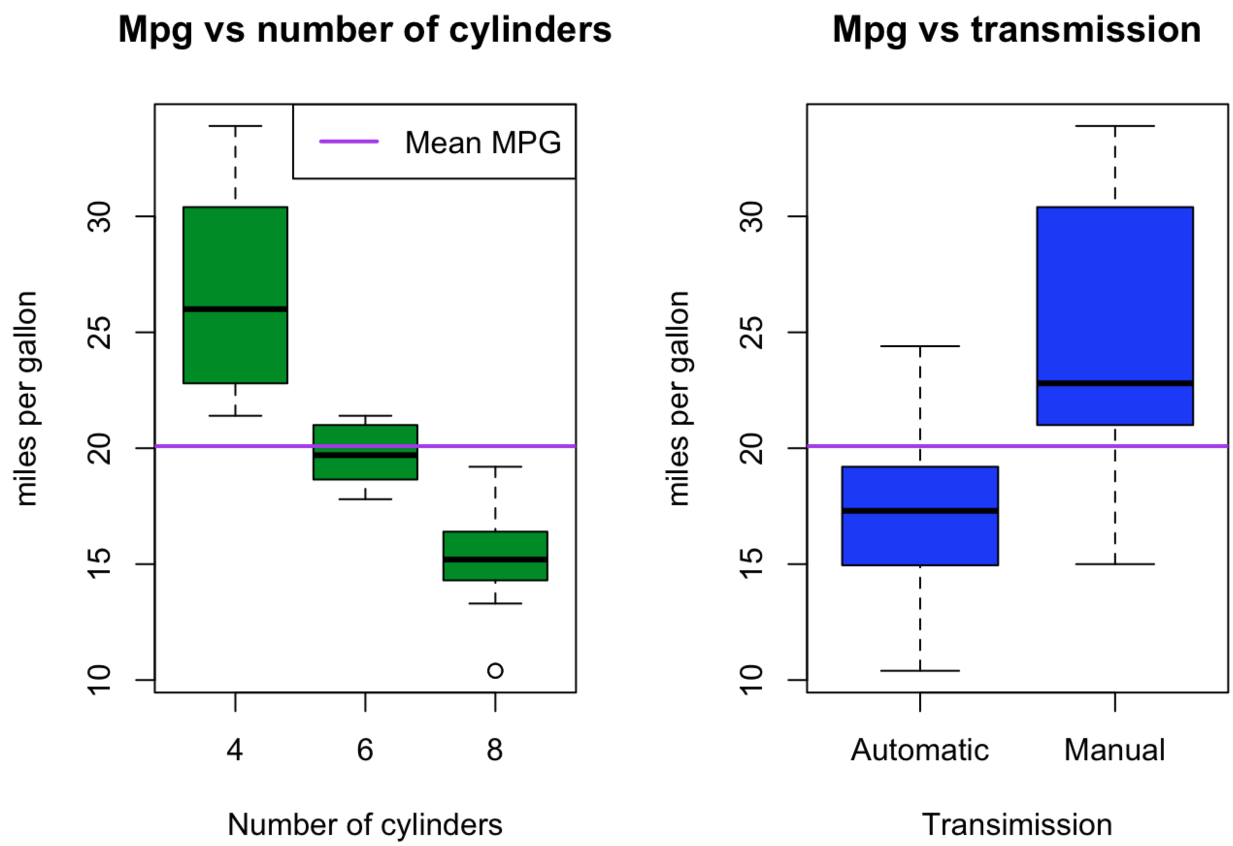

Figure 1: Exploratory analysis of different relationships

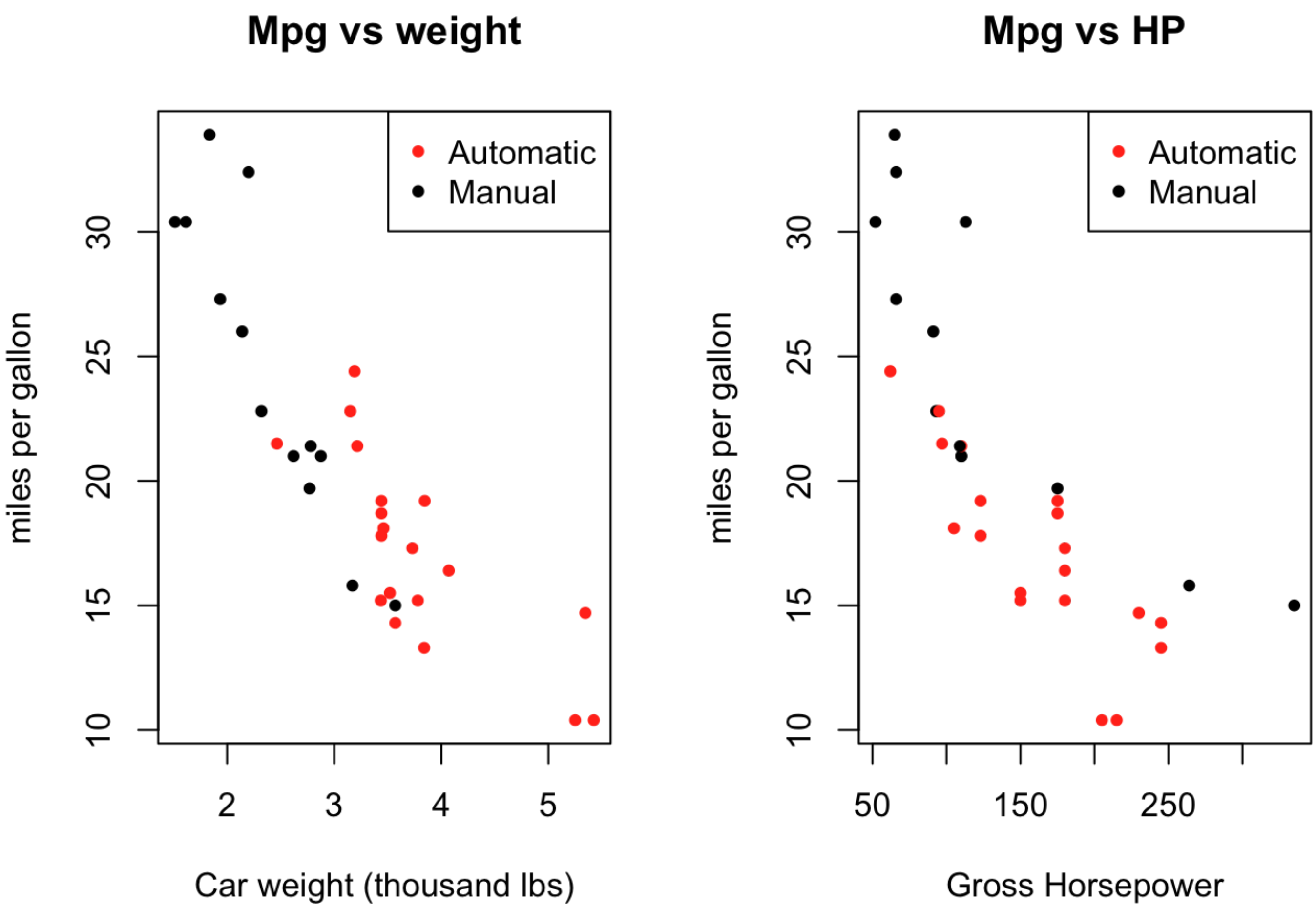

Figure 2: Exploratory analysis of different relationships

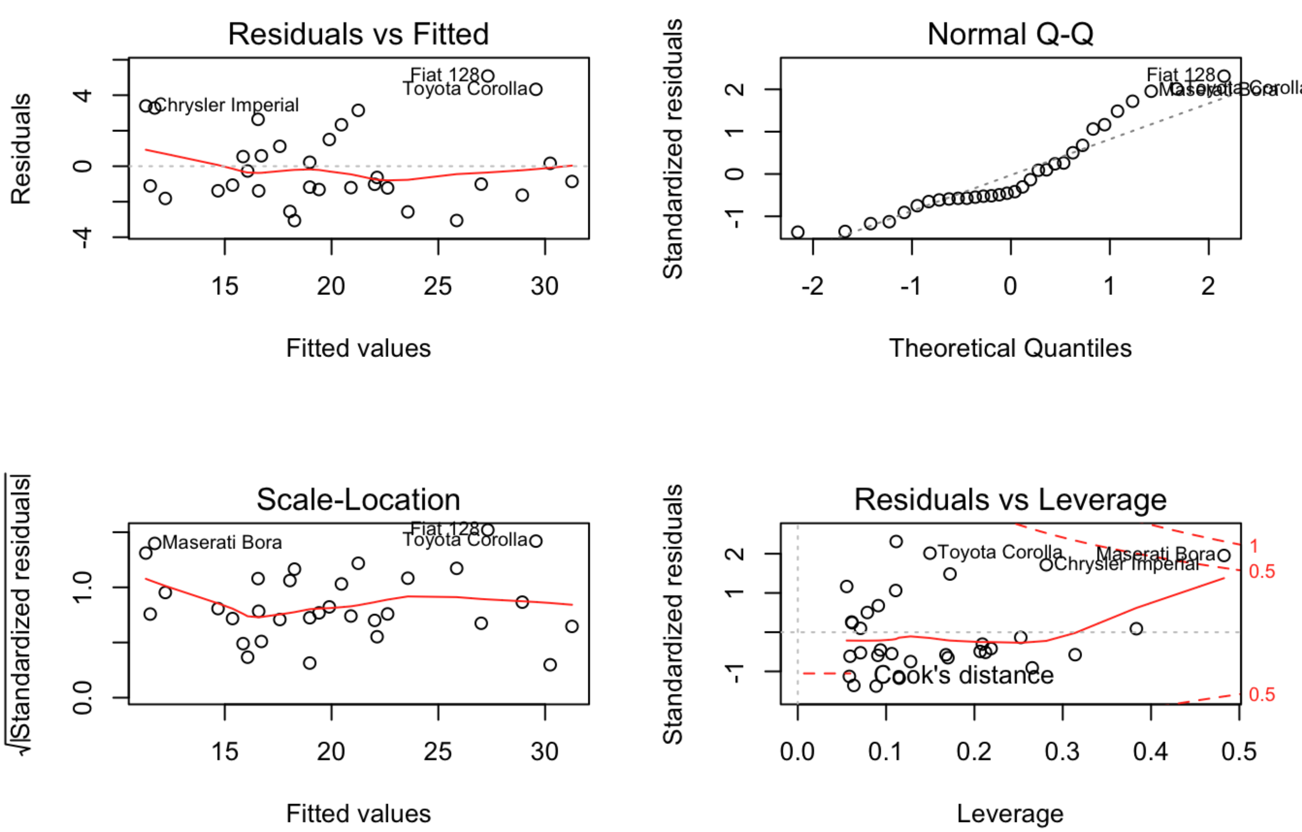

Figure 3: Residual analysis plots

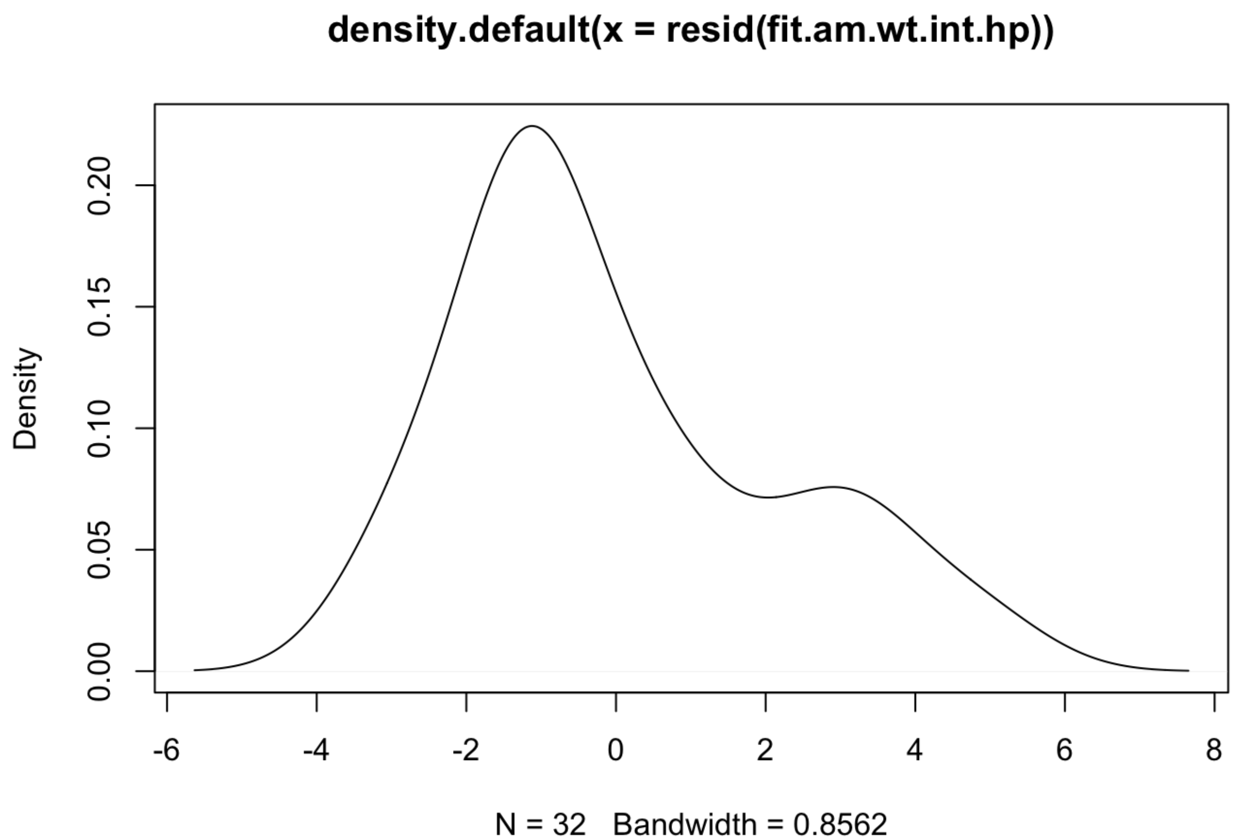

Figure 4: Denisty of residual values