# Vehicle Fuel Economy Effect of Transmission Type

## Executive Summary

1. This analysis attempts to quantify the impact of selecting a manual or an automatic transmission on vehicle fuel economy, by reference to Road and Track Magazine's list of 32 vehicles from 1974. It attempts to mathematically model the characteristics determining fuel economy of existing cars and isolate the dependence on transmission type from other factors.

2. The model predicts an improvement of 2.9 miles per gallon from a manual transmission as compared with an automatic. However, the data do not compellingly support the reliability of the model.
3. Recommendation: **Do not use the model's result**. Augment the dataset with paired manual and automatic transmission versions of the same make and model and repeat the analysis.

## Modeling the data

### Notes on Data Factorization

- The transmission type (manual or automatic) is a categorical variable.
- The number of engine cylinders is treated categorically; 4, 6, and 8 cylinder engines are the three categories. The underlying concept is that engines with differing cylinder counts may have significantly different configurations.
- Carburetors are not treated categorically as direct mathematical comparison of different carburetor counts is feasible and meaningful.

### Observations from Data Exploration (see Appendix)

- The cars with manual vs automatic transmission are not identically distributed (see also Additional MPG Plots). Note the corresponding regression lines.
- The cylinder configuration, price, interior configuration, and purpose of the vehicle (e.g. family vehicle, sports car) are not included in the dataset which could result in a biased conclusion.

### Model selection strategy

- The analysis attempts to include only highly relevant predictors for outcome MPG, minimizing the number of predictors.
- It considers two different models; but does not consider second order (interaction) terms or exclude outliers.
- Null hypothesis: None of the regressors predict vehicle fuel economy.

**Model 1: Minimize Akaike Information Criterion (AIC)** The first is the built-in R function step() which through successive iterations drops fields from the model to **minimize** the Akaike Information Criterion (AIC). Increasing the number of regressors raises the AIC; thus it attempts to balance fitting the data with minimizing regressor count, see: http://en.wikipedia.org/wiki/Akaike_information_criterion.

```
model.all <- lm(mpg ~ factor(am) + factor(cyl) + disp + hp + drat + wt + qsec + gear + carb, mtcars)
model1 <- step(model.all)
```

#### TABLE OF MPG VS MARGINAL CHANGE IN REGRESSOR VALUES WITH P-VALUES FOR MODEL 1

```
## lm(formula = mpg ~ factor(am) + wt + qsec, data = mtcars)
```

```
##                                     Estimate  Pr(>|t|)
## am: manual transmission                2.936 4.672e-02
## wt: per 1000 pounds weight            -3.917 6.953e-06
## qsec: per second of 1/4 mile time      1.226 2.162e-04
```

- All 3 predictors have a p-value below 0.05. However, the am factor (transmission type) p-value is barely below this threshold.
- The Adjusted **R squared value of 0.83** suggests the model is a good fit.
- If the AIC report is credible, we reject the null hypothesis and include 3 predictors in the model, noting the marginal statistical significance for the transmission type factor. The model then predicts manual transmission increases fuel economy by approximately 2.9 MPG, assuming we make no other changes.

**Model 2: P-Value stepwise backward selection (see Appendix)** As described in Open Stats Intro Chapter 8, the model initially includes all predictors and fits a regression line, calculating the p-value for each regressor for its relevance in the model. The method then eliminates the highest p-valued regressor from the model and refits. This iterative process continues until all predictors are below the significance threshold of 0.05.
**The stepwise backward selection resulted in the same model as AIC.**

## Inspecting the Model

### Observations from Residual Plots (see Appendix)

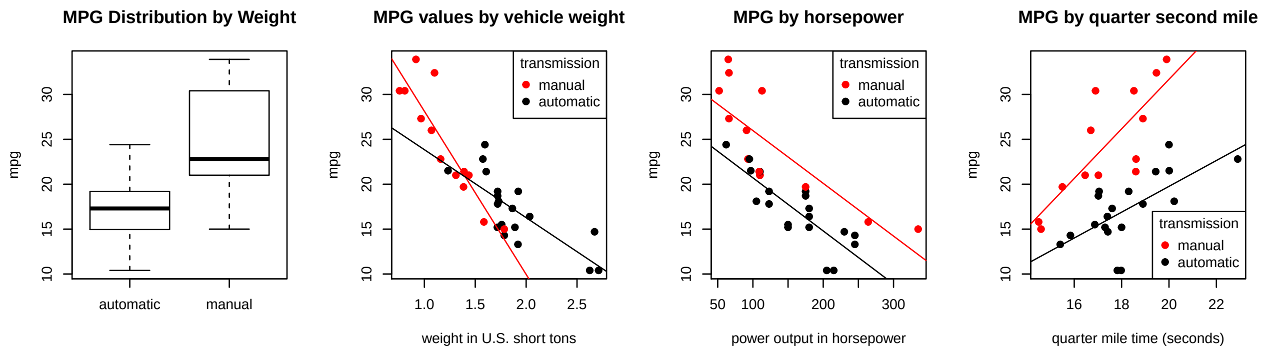
- As with the exploratory plots, the residual plots are clustered and do not appear identically distributed between automatic and manual transmission equipped cars.

# Results

- See Table of Model 1.
- The **Adjusted R squared value of 0.83** suggests the model is a good fit.
- The model predicts a **improvement to MPG of 2.94** with a manual transmission.
- The decision of whether to accept the model result is more complicated; the p-value of the transmission regressor is barely below the 0.05 significance threshold. Furthermore, plots of both absolute values and residuals show that the manual and automatic transmission values are not identically distributed.
- On balance, the author believes it is more prudent to **draw no conclusion** from the data with regard to the effect of manual or automatic transmission on vehicle fuel economy.
- Repeat the analysis when a dataset with paired manual and automatic transmission cars of the same make and model is available.

# Appendix A: Figures

## Figure A1: Data Exploration




## Figure A2: Model Residuals

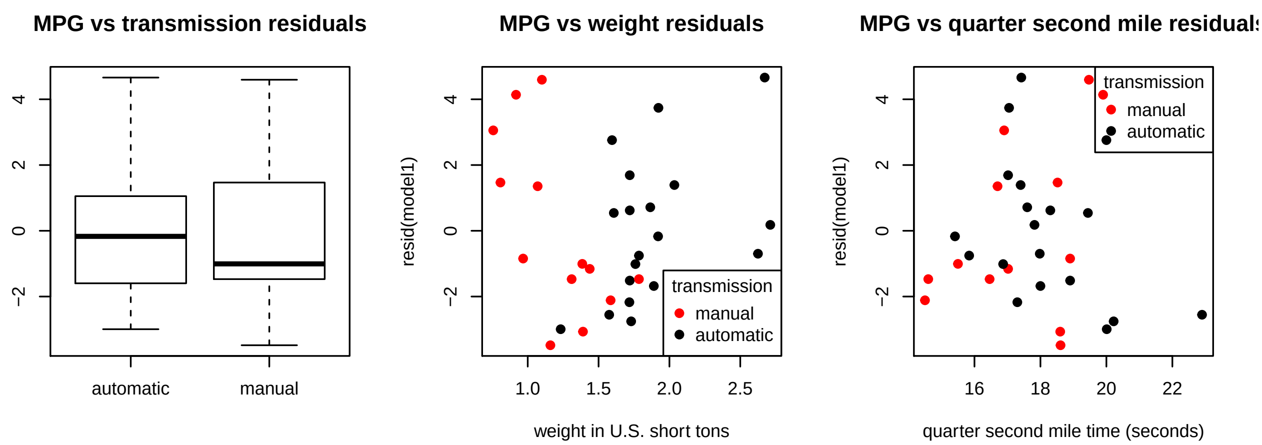


**Figure A3: Data Exploration; additional MPG plots**

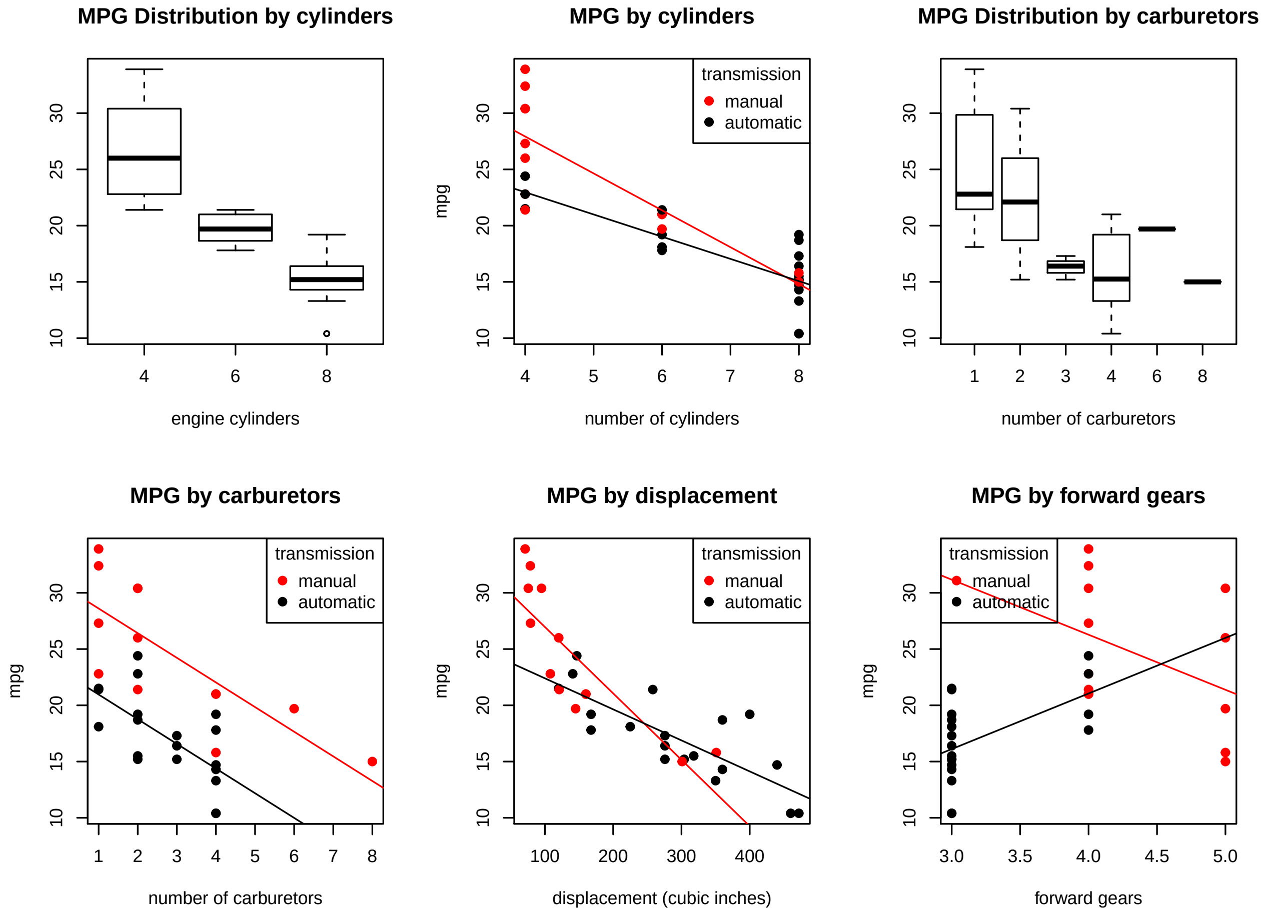



**Figure A4: Table of Backward Stepwise P-Value Iterations for Model 2**

```
model2 <- lm(mpg ~ factor(am) + factor(cyl) + disp + hp + drat + wt + qsec + gear + carb, mtcars)
model2 <- lm(mpg ~ factor(am) + disp + hp + drat + wt + qsec + gear + carb, mtcars)
model2 <- lm(mpg ~ factor(am) + disp + hp + drat + wt + qsec + gear, mtcars)
model2 <- lm(mpg ~ factor(am) + disp + hp + drat + wt + qsec, mtcars)
model2 <- lm(mpg ~ factor(am) + disp + hp + wt + qsec, mtcars)
model2 <- lm(mpg ~ factor(am) + hp + wt + qsec, mtcars)
model2 <- lm(mpg ~ factor(am) + wt + qsec, mtcars)
```

**Figure A5: Table of Influence on Model by Transmission Type**

```
##   transmission transmission dfbeta       wt     qsec
## 1    automatic              0.04002  0.06619 -0.03043
## 2       manual             -0.04722 -0.08428  0.04317
```