# Motor Trend mtcars Analysis

*James Eddy*
*August 24, 2014*

## Executive Summary

The objective of this study was to investigate the following questions:

> "Is an automatic or manual transmission better for MPG"
> "Quantify the MPG difference between automatic and manual transmissions"

After exploratory calculations and plotting to inspect apparent relationships between transmission and MPG (and between each and other variables in the data), several linear models were fit and examined. While transmission type appears to have a significant effect on MPG when viewed in isolation, further analysis revealed that this result is likely due to confounding from other variables. In particular, transmission type is negatively correlated with car weight, which has a significant negative association with MPG, thus creating the apparent (but non-significant) positive association between transmission and gas mileage.

## Exploring the Data

The data for this study can be found in the `mtcars` dataset within **R**. Exploratory analysis was performed using functions in the `dplyr` and `ggplot2` libraries. After loading the required data and libraries, the `cor()` function was used to examine the correlation between variables in the data set. This simple measure suggested that transmission type is only moderately correlated with MPG (0.60), while other values such as number of cylinders, displacement, and weight might have more of an effect on gas mileage.

Before generating some exploratory plots, the `mutate()` function was used to append an additional column `transmission` (based on the `am` column) to the `mtcars` data frame, making inspection of automatic vs. manual transmission easier. Generating boxplots of MPG values for automatic and manual transmission cars provided a simple way to compare distributions (**Figure 1**). The difference in median MPG shown in the plots suggested that manual transmission cars are better for gas mileage than automatic transmission. While there was a pronounced difference in MPG between manual and automatic transmission cars, there may be other factors correlated with transmission that are more responsible for the difference in MPG. Several variables, including cylinder, displacement, rear axle ratio, weight, and number of gears all appear to have either positive or negative associations with transmission type (**Figure 2**). These variables represent potentially "confounding" factors, which need to be controlled for when modeling the influence of transmission type on MPG.

## Modeling & Diagnostics

A simple model was first fit to directly measure the apparent influence of transmission on MPG. The first non-intercept coefficient of the model, 7.245, suggested an increase of 7.245 miles per gallon when switching from automatic to manual transmission (with p-value < 0.0001). However, the low R- squared

value (0.3598) of the model indicated that transmission alone is not a very good predictor of the variation in MPG.

```
fm1 <- formula(mpg ~ transmission)
model1 <- lm(fm1, data = mtcars)
summary(model1)
```

```
##
## Call:
## lm(formula = fm1, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.392 -3.092 -0.297  3.244  9.508
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)           17.15       1.12   15.25  1.1e-15 ***
## transmissionmanual     7.24       1.76    4.11  0.00029 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.9 on 30 degrees of freedom
## Multiple R-squared:  0.36,   Adjusted R-squared:  0.338
## F-statistic: 16.9 on 1 and 30 DF,  p-value: 0.000285
```

To examine whether additional attributes may be confounding the observed influence of transmission type on MPG, a second model was fit with three additional independent variables: rear axle ratio ( `drat` ), weight ( `wt` ), and number of gears ( `gear` ). Each of these variables was previously shown in the exploratory analysis to be highly correlated with transmission type. Notably, the resulting fit indicates that transmission is in fact not a strong predictor of MPG, with only 0.1689 (p-value = 0.94) increase in MPG for manual vs. automatic. Instead, weight seems to have a much stronger relationship to gas mileage, with a decrease of 4.89 miles-per-gallon predicted for every increase of 1000 lbs in car weight.

Even with the four indepentent variables, the second model still only led to a moderately good R-squared value of 0.7687. Furthermore, a diagnostic plot of residuals showed values non-evenly distributed around zero (**Figure 3**), indicating that the model is probably not very good. To examine whether adding additional variables could improve predictability, a third model was fit based on all variables in the data set. While R-squared was improved (0.869), the plot of residuals was still somewhat uneven (**Figure 4**). It is possible that too many variables are included in this case, which might hurt the model; alternatively, there may be interaction among some variables, which could be modeled more explicitly to improve performance. As the objective of this assignment was to explore the effect of transmission type on MPG, further optimization of a model for MPG prediction was not pursued here.

# Appendix

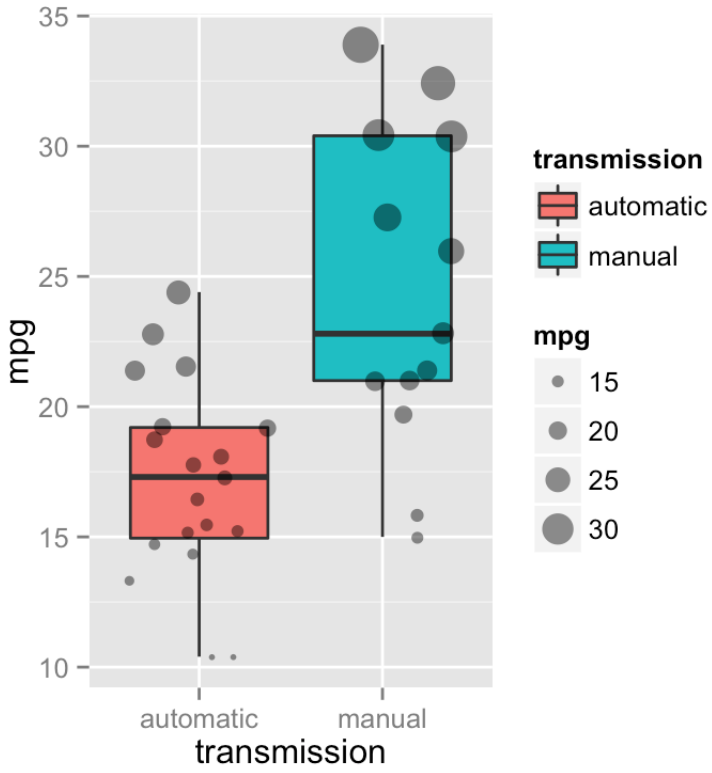


**Figure 1:** Box plot showing distribution of MPG values in manual vs. automatic transmission cars. Points are also scaled according to MPG value for emphasis.

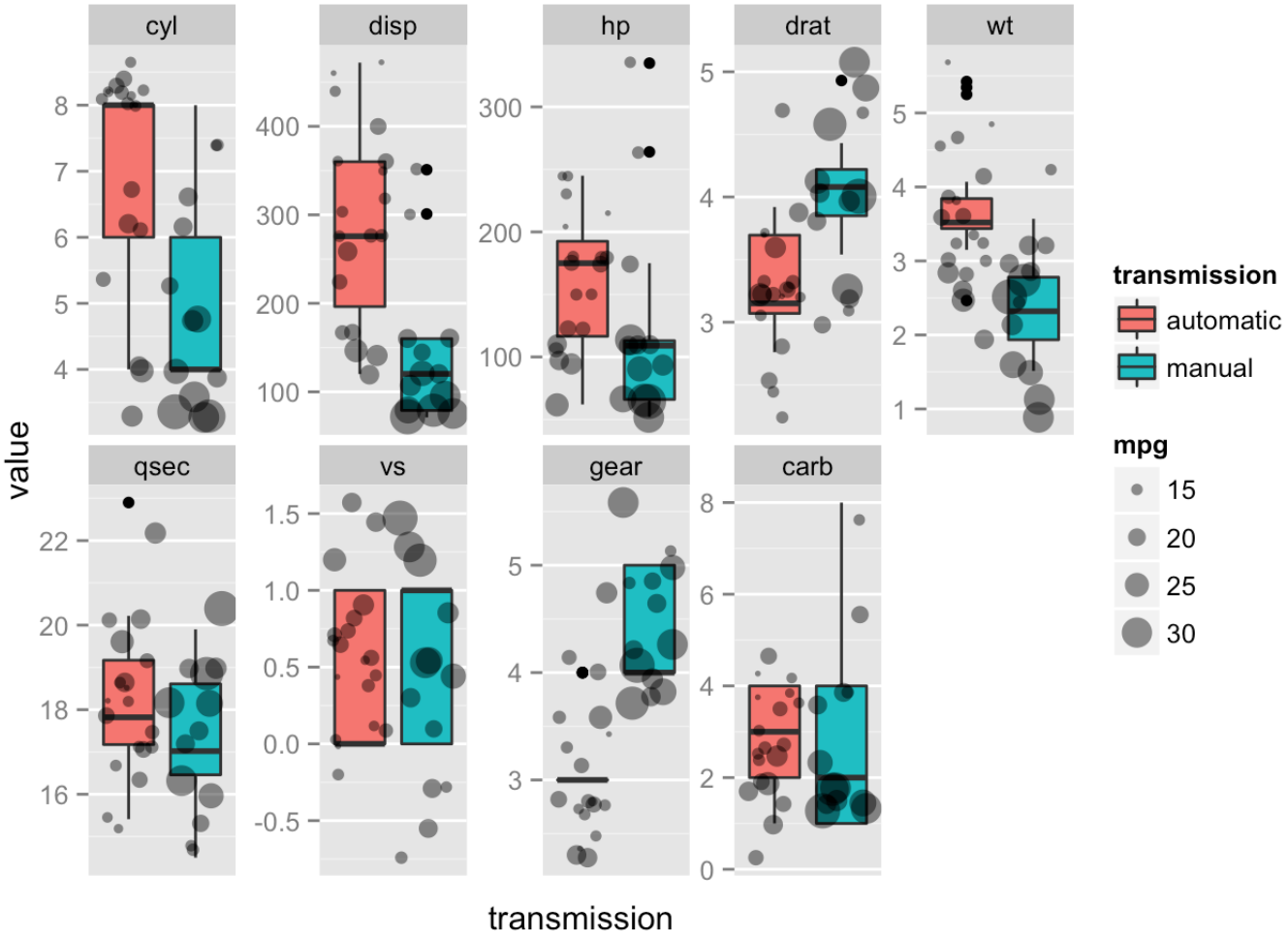


**Figure 2:** Box plots showing distributions of all attribute values in manual vs. automatic transmission cars. Points are again scaled according to MPG value to simultaneously highlight these distributions.

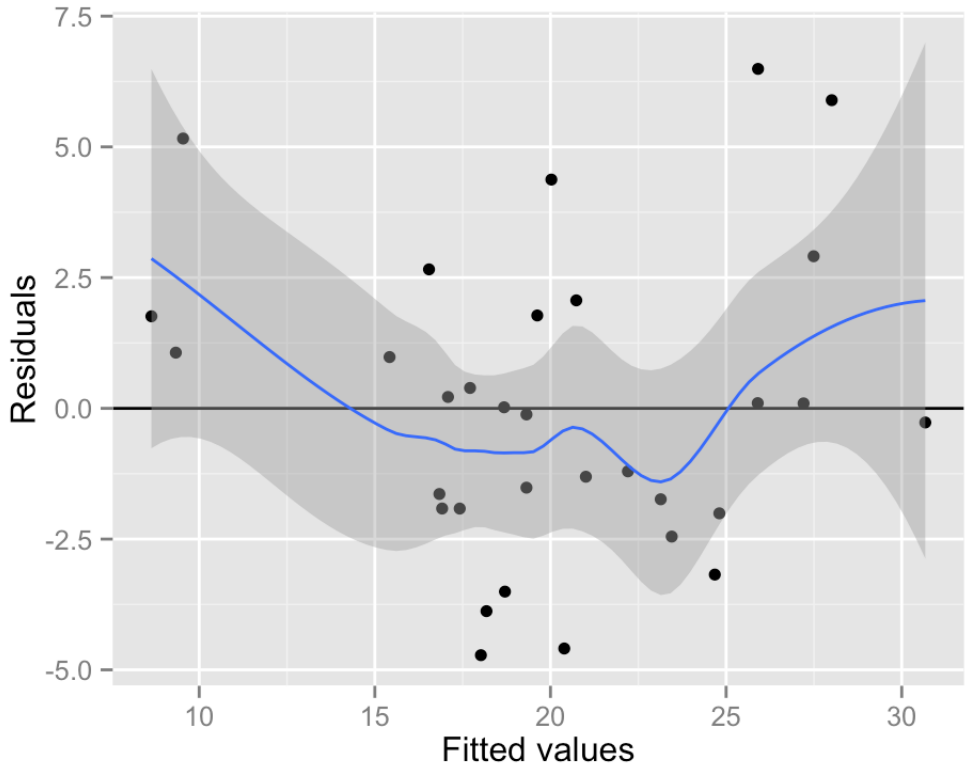

**Figure 3:** Residuals plot for model of `mpg ~ transmission + drat + wt + gear` (independent variable: MPG; dependent variables: transmission type and variables with which it is correlated).

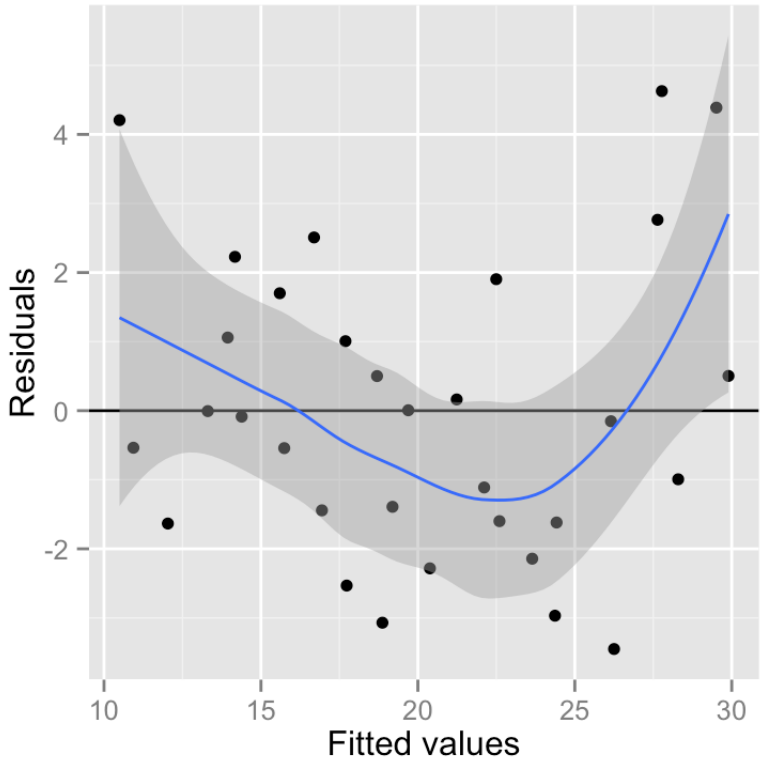


**Figure 4:** Residuals plot for model of `mpg ~ .` (independent variable: MPG; dependent variables: all other variables in the `mtcars` data set).