MPG Analysis: Regression Models Course Project

Travis Fell

August 14, 2015

Executive Summary

This analysis strives to answer the questions from data in the mtcars data set: 1. Is an automatic or manual transmission better for MPG? Manual is better for MPG, but when accounting for factors that more strongly correlate with MPG, the difference between manual and automatic transmissions is negligible. 2. What is the MPG difference between automatic and manual transmissions? When accounting for weight and displacement, the MPG range of improvement for manual transmissions is fit: 0.1777241 lwr: 0.1816018 upr:0.17384

I. Loading and Exploring the Data

First, we'll start by loading up the data and packages we need and doing some quick exploratory analysis.

```
library(ggplot2); library(GGally); library(dplyr)
data(mtcars); head(mtcars); summary(mtcars)
summarize(group_by(mtcars, am), mean(mpg), records = length(mpg))
```

Note that the am column indicates the transmission type of the care; 1 = manual, 0 = automatic. At first, the manual transmissions appear to have the edge in MPG, 24.4 to 17.1 for automatics. However, there are other variables that could also have an impact on MPG that we will examine in the following section.

II. Isolate Transmission as a Factor of MPG

As a baseline, we will create a single linear regression between transmission type and MPG. See the appendix for a graph of this regression model.

```
mpgtrans <- lm(mpg ~ am, mtcars);summary(mpgtrans)$coefficients
```

On its own, transmission type has a slope coefficient of 7.2 from automatic to manual. However, there are several other factors that could impact MPG. We'll compare transmission type to all variables to begin understanding how they impact transmission type as a predictor of MPG.

```
mpgall <- lm(mpg ~., mtcars);summary(mpgall)$coefficients
```

When creating a model with all variables, the slope coefficient for transmission decreases significantly to 2.52. But which ones have the biggest impact on MPG and how can they help us understand the true impact of transmission type on MPG?

The pairs plot in the appendix shows that transmission type has a reasonably strong positive correlation with MPG (.6). However, the variables of cylindars (-.852), displacement (-.848) and weight (-.868) have very strong negative correlations with MPG. Since displacement is a measure of total gasoline an engine would use in a single drive cycle, we'll use that variable as a measure instead of cylindars. So we will re-run our regression model with weight and displacement.

```
mpgtranswt <- lm(mpg ~ am + wt, mtcars); summary(mpgtranswt)$coefficients
mpgtransds <- lm(mpg ~ am + disp, mtcars); summary(mpgtransds)$coefficients
mpgtranswtds <- lm(mpg ~ am + wt + disp, mtcars); summary(mpgtranswtds)$coefficients
```

When using weight as a second regressor, the slope of AM changes dramatically, from 7.2 to -.02. When including displacement in the model, there is a similar reduction to .17. The fact that relationship between MPG and transmission type can be so significantly impacted by other variables should give us pause when evaluating MPG performance purely as a function of transmission type. Rather, it would probably make more sense to take weight, displacement and transmission type into account when evaluating MPG performance.

In the appendix appear regression plots of MPG by Weight and MPG by Displacement, both of which distinguish manual vs automatic transmission data points. These plots illustrate one of the conclusions of this analysis: to answer the question "Is an automatic or manual transmission better for MPG?", one should consider transmission type along with the most impactful predictors of MPG: Weight and Displacement.

III. A Quick Note on Residuals

The residuals plot in the appendix for the MPG by transmission, weight and displacement model does not seem to indicate any patterns in the data, unusual results or any other issues that would indicate some poor aspect of model fit.

IV. Predicting the MPG

To illustrate the impact of transmission type on MPG, we will predict the MPG holding weight constant at 3 tons and displacement constant at 250 while changing transmission type.

```
autompg <- predict(mpgtranswtds, data.frame(am = 0,  wt = 3, disp = 250), interval = "pre
diction")
manualmpg <- predict(mpgtranswtds, data.frame(am = 1,  wt = 3, disp = 250), interval = "p
rediction")
mpgdelta <- manualmpg - autompg
```

Using the R predict function, we find that automatic transmission MPG ranges are Fit:20.3875516 Lower: 13.9689453 Upper: 26.8061578 By comparison, manual transmission MPG ranges are Fit:20.5652757 Lower: 14.1505471 Upper: 26.9800043

While manual is slightly better, we can conclude from this data that there is not much change when weight and displacement are taken into account, the difference being Fit: 0.1777241 Lower: 0.1816018 Upper: 0.1738465

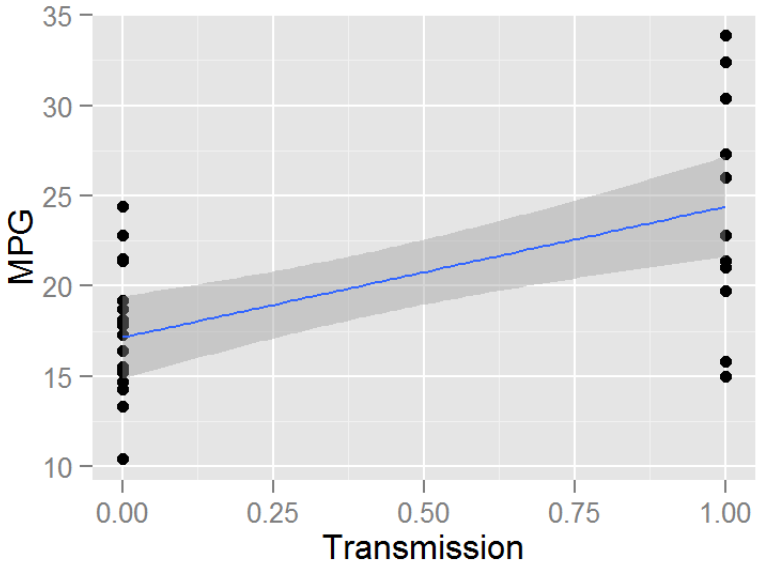
The 95% confidence interval for the variables in this model is as follows:

```
confint(mpgtranswtds)
```

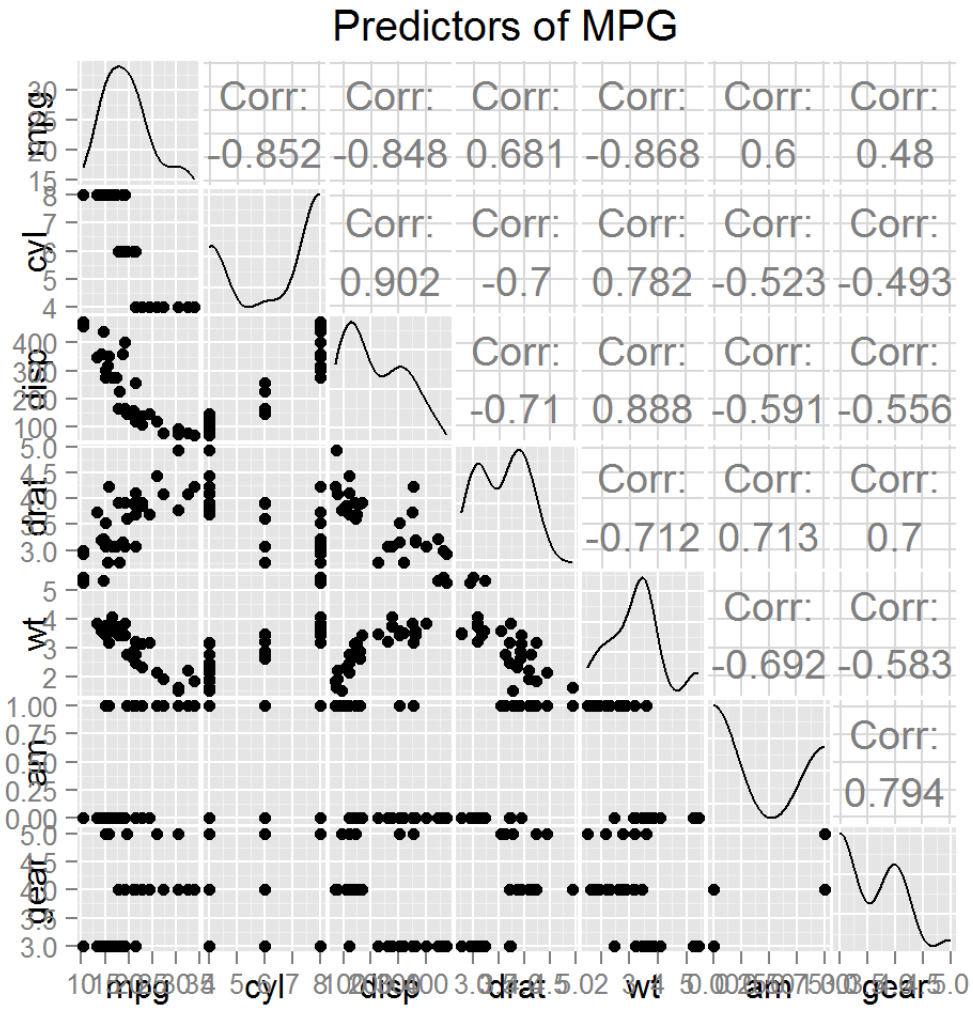
```
##                   2.5 %      97.5 %
## (Intercept) 28.03782444 41.31399731
## am          -2.86275907  3.21820735
## wt          -5.99832334 -0.55976441
## disp        -0.03700801  0.00139819
```

Appendix

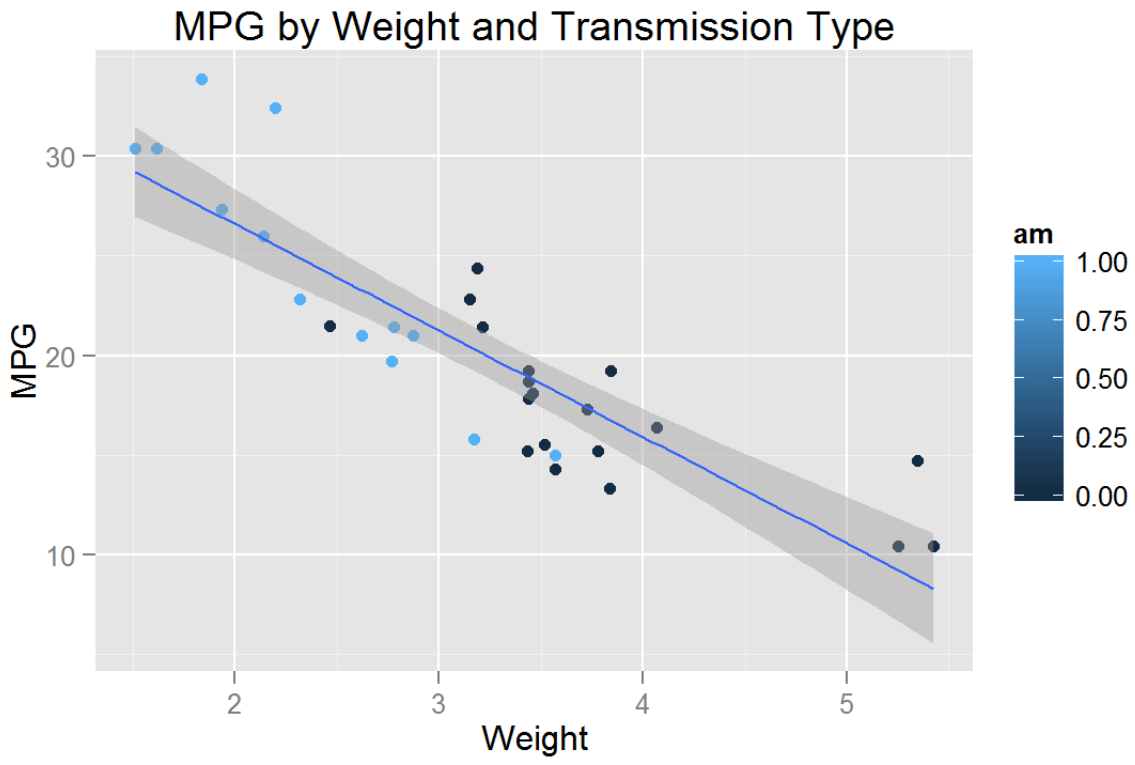
Single linear regression between MPG and transmission type.



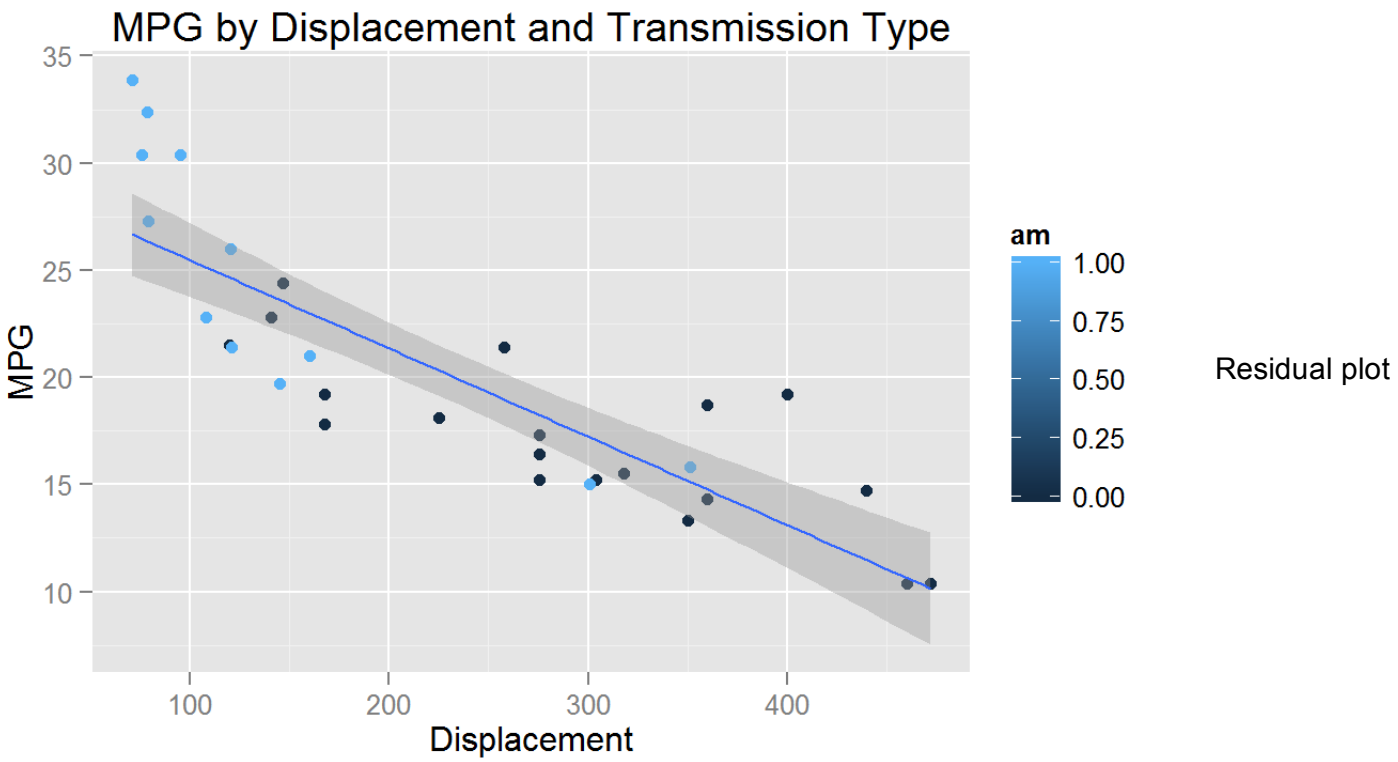
Pairs graph w/correlations between variables



Regression plot of MPG by transmission type and weight



Regression plot of MPG by transmission type and displacement



Residual plot