Motor trends: The effect of transmission type on fuel efficiency

John Lee

20 October 2015

Summary

The objective of this report is to assess the effect of automatic or manual transmission types (am) on the fuel efficiency of the motor vehicles measured in miles per gallon (mpg) in a dataset extracted from the Motor Trend US magazine (1974). While a prelimanry comparison of the two transmission types suggests that manual transmissions are substantially more efficient, this analysis is confounded by other variables in the dataset. Car weight and gross horsepower intuitively influences the fuel efficiency of a car. While some outliers exist in a model using these two variables as predictors of mpg, the model is useful in that it captures 83% of the variance of the mpg variable in the dataset. The am variable is of little use as a predictor when added to this model; however, the conclusion that tranmission type has little effect on fuel economy is rendered ambiguous by the fact that the transmission type is strongly correlated with the cars' weights.

Report

The purpose of this analysis is to assess the effect of transmission type on the MPG of various cars. Without controlling for any other variables manual transmissions are typically more fuel efficent getting more miles per gallon (mpg). Using tranmission type as a predictor for the mpg variable (see below) the mean mpg of automatic cars is 17.15 with a standard error of 1.1 and the mean manual transmission is 7.2 mpg higher with a standard error of 1.8. A hypothesis test for the difference between the two categories yields a p-value < 0.001 suggesting that the transmission type (am) strongly affects mpg; however, this description does not take into account the other confounding variables included in this dataset. Furthermore, using this model to predict mpg, only 35% of its variance is accounted for suggesting that other variables are important in understanding the fuel efficiency of any given car.

```r
summary(lm(mpg~ am,data = mtcars))[c( "coefficients","r.squared")]
```

```
## $coefficients
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## amManual     7.244939   1.764422  4.106127 2.850207e-04
##
## $r.squared
## [1] 0.3597989
```

A model using all other variables as predictors suffers from variance inflation and so a more parsimonious model was chosen. A matrix of plots pairing each of the variables demonstrates extensive correlation between many of them (see supplimentary figures 1-3). Of these the car weight (wt) shows a strong correlation with mpg. This makes sense as more fuel would be used to move heavier cars. Using wt as a regressor to predict mpg captures 75% of the variation in the dataset. An additional variable which shows a strong correlation with mpg is the gross horsepower of the car (hp). Once again this is intuitive as a more powerful car will use more fuel. A linear model including these two predictor variables, detailed below, captures almost 83% of the variance in the mpg variable.

```r
lm_fit <- lm(data = mtcars, mpg ~ wt+ hp)
summary(lm_fit)[c( "coefficients","r.squared")]
```

```
## $coefficients
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 37.22727012 1.59878754 23.284689 2.565459e-20
## wt          -3.87783074 0.63273349 -6.128695 1.119647e-06
## hp          -0.03177295 0.00902971 -3.518712 1.451229e-03
##
## $r.squared
## [1] 0.8267855
```

```r
vif(lm_fit)
```

```
##       wt       hp
## 1.766625 1.766625
```

The above model states that for every 1000lb change in the weight of a car the average decrease in the fuel efficiency of the car is 3.9 mpg (with a standard error of 0.63). Similarly for every unit increase in hp the fuel efficiency decreases by 0.03mpg (with a standard error of 0.01). The null hypotheses that wt and hp are not linearly related to mpg are rejected with $p < 0.001$ and $p < 0.01$ respectively. Power is of course correlated to a degree with the weight of the car (see supplimentary figure 2) as heavier cars will need more powerful engines but power can also independently describe cars that are not be as heavy that use their power for speed rather than carrying heavier loads (the Maserati Bora and the Ford Pantera L being good examples, see Supplimentary Figure 4). The variance inflation factors suggests that the collinearity is not high, with the standard error of the variable coefficients only 1.3 times higher than if wt and hp were uncorrelated.

While model including wt and hp captures much of the variance in mpg there are some problems that suggest it may not be useful to answer our question. A plot of the residuals (see Supplimentary Figure 3) reveals that some outliers exist in the model. Furthermore the residual plot indicate that some deviation from linearity exists in the model. This suggests that a linear model may not best describe the data or at least that the variables require scaling using an exponential function. More worringly, a plot using the variables hp,wt, am and mpg suggests that an automatic transmission was typically used in heavier cars. The fact that the two transmission types are not well distributed across all car weights makes it impossible to unambiguously compare the effect of transmission type on mpg.

With the above caveats, an attempt at calculating the effect of transmission type on mpg can be made by including it in the model. Using nested liklihood ratios for model selection suggests that the model is not improved by adding the am term ($p = 0.14$):

```r
fit_with_transmission_type <- update(lm_fit,mpg ~ wt + hp + am)
anova(lm_fit, fit_with_transmission_type)
```

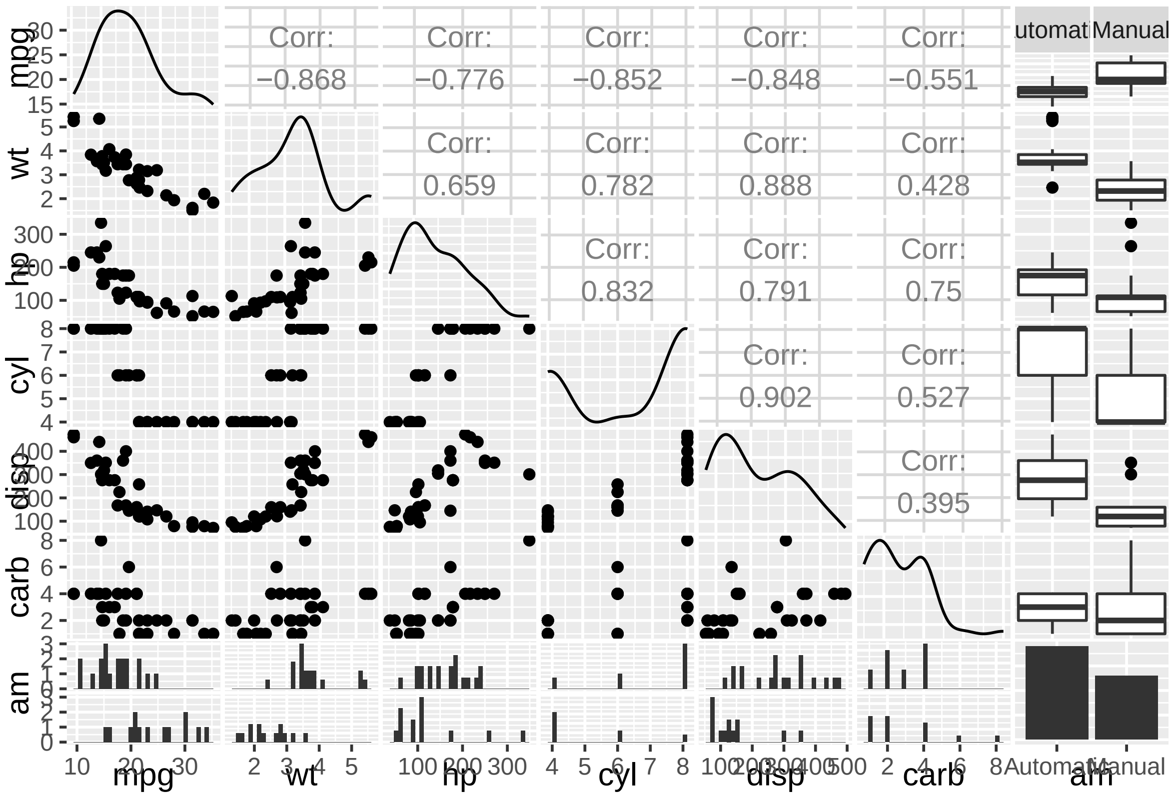
```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt + hp
## Model 2: mpg ~ wt + hp + am
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1     29 195.05
## 2     28 180.29  1    14.757 2.2918 0.1413
```

Using the model the effect of transmission type when accounting for the weight and power of the car is not significantly different from 0 and is calculated to be 2.0837101mpg higher for manual transmissions with a standard error of 1.3764202mpg. With a dataset where transmission types is more evenly distributed across the different weights and horsepowers the effect of transmission type could be more adequately addressed.

Appendix

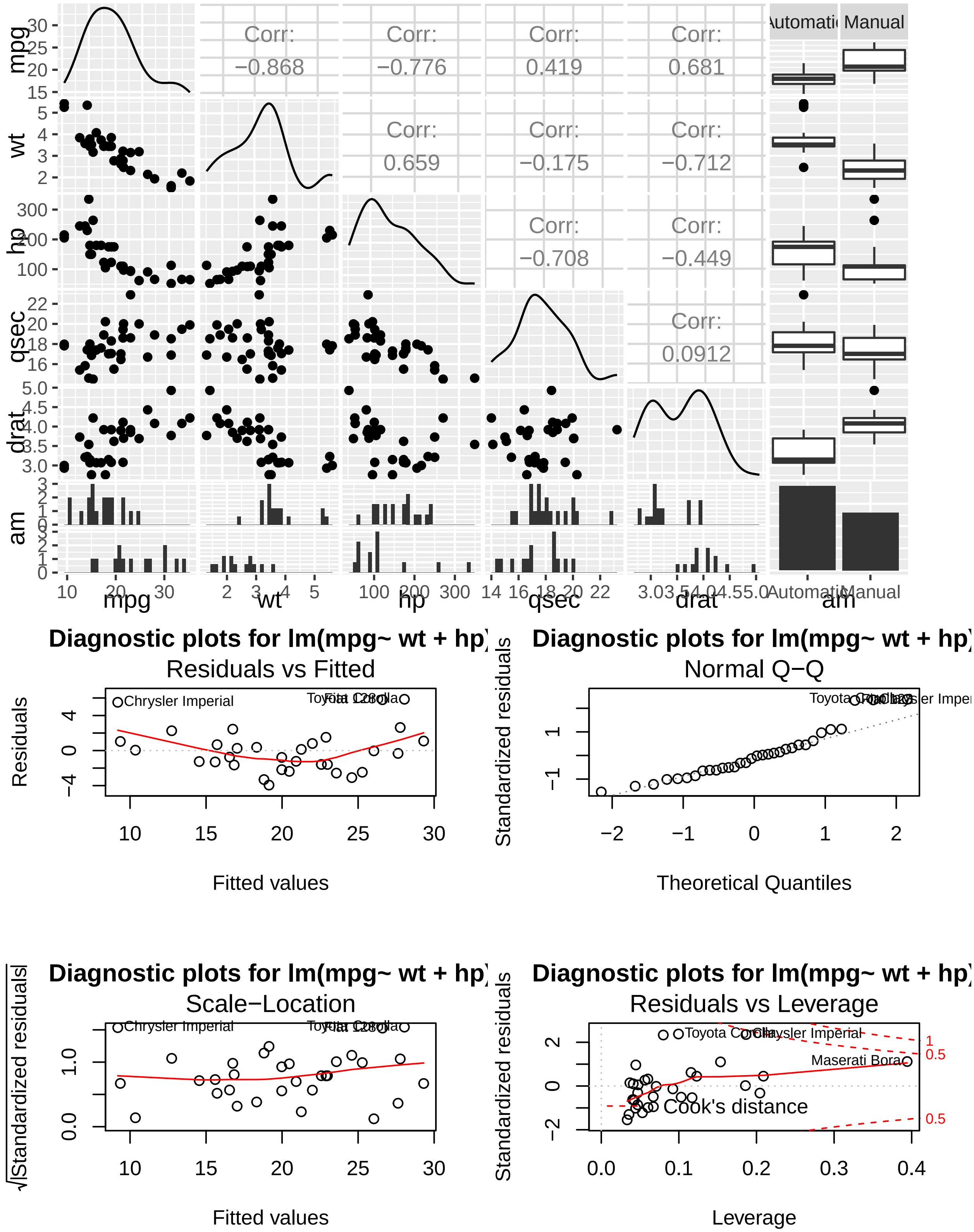
```r
par(mar= c(rep(1,4)))
ggpairs(select(mtcars,mpg,wt,hp,cyl,disp,carb,am), title = "Supp. Figure 1: Comparing variables")
```

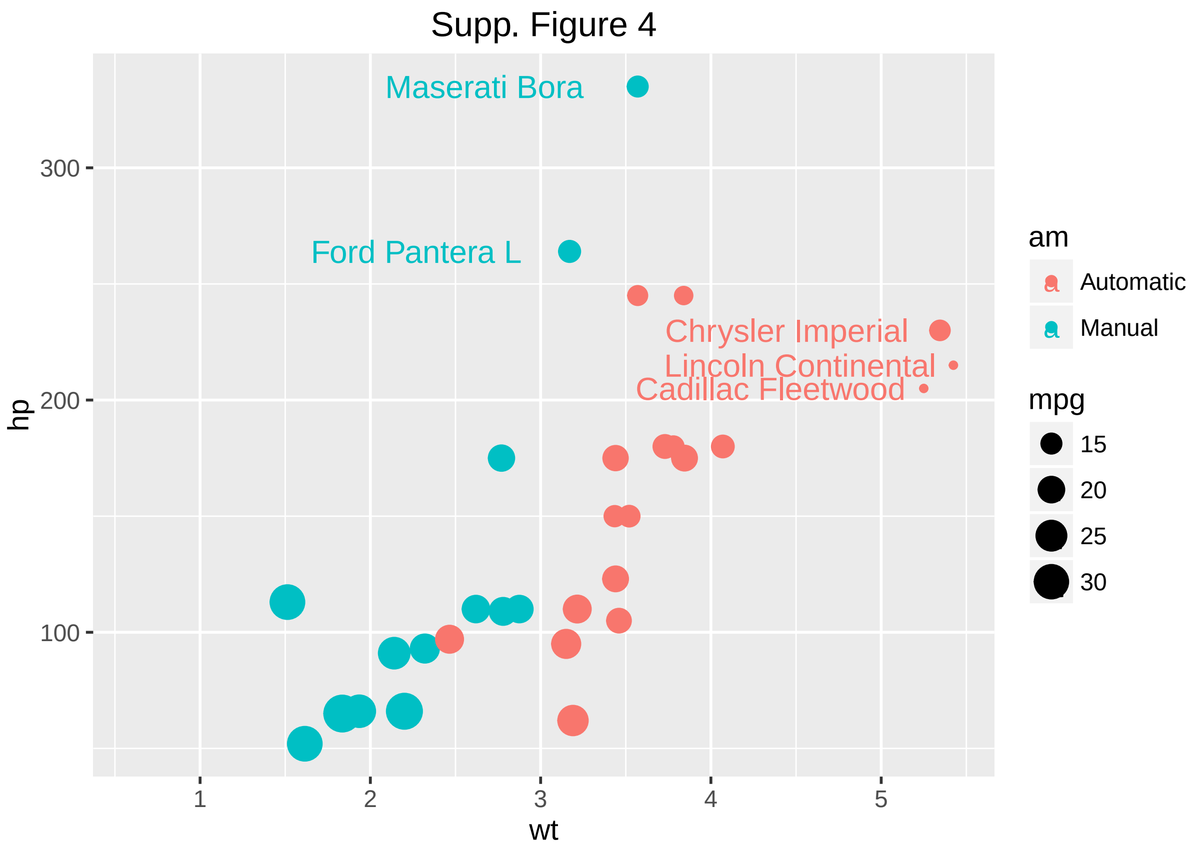
Supp. Figure 1: Comparing variables




```r
# ggpairs(select(mtcars,wt,cyl,disp,hp,carb), title = "Supp. Figure 2: Vehicle weight")
ggpairs(select(mtcars, mpg, wt, hp, qsec,drat,am), title = "Supp. Figure 2")
```

Supp. Figure 2




Supp. Figure 4
Maserati Bora
Ford Pantera L
Chrysler Imperial
Lincoln Continental
Cadillac Fleetwood
300
200
100
hp
1
2
3
4
5
wt
am
Automatic
Manual
mpg
15
20
25
30