Transmission Type and Fuel Economy in 'mtcars'

Executive Summary

A simple, two variable model of fuel economy generated from the mtcars data set produces remarkably good diagnostic results. However, the data appear to be ill-suited to determining whether transmission type has any influence on fuel economy. A more effective method for answering that question would be a pairwise comparison of the various car models, each with both transmissions.

Analysis

Weight (wt) is the variable most correlated with mpg, as shown in Table 1. So the exploratory analysis starts with a plot of these two variables.

Table 1. mtcars Correlations with mpg

```
##          cyl    disp      hp   drat      wt   qsec    vs     am   gear
## [1,] -0.8522 -0.8476 -0.7762 0.6812 -0.8677 0.4187 0.664 0.5998 0.4803
##         carb
## [1,] -0.5509
```

The curved shape of the scatter in fig. 1 suggests that the variation of mpg with respect to wt is not linear. However, a transformation of the mpg variable to gpm (gallons of fuel consumed per 100 miles) appears to be linear with respect to weight (see Fig. 2) and is more likely to satisfy error assumptions since it is in-line with the original measurement – fuel consumption over a 73-mile route (Henderson & Vellemen, 397).

```
#new gallons per mile variable
mtcars$gpm <- 100 / mtcars$mpg
```

There is very little overlap between the automatic and manual transmission cars with respect to weight, as described by Fig. 2. Nearly all the light cars have manual drivetrains while the heavy cars use automatics. This lack of heterogeneity makes me skeptical of a model derived from this particular data set that uses the transmission variable.

Seeing the differentiation between the low-gpm Civic and Fiat 128 and the higher-gpm sports cars, it does appear that a measure of how under- or over-powered a car is would add description to the model. Henderson and Velleman choose hp divided by wt. The new hp/wt (hpwt) variable is not correlated with wt, so it is acceptable to add in the first model.

```
# hp per weight measure of over/under-poweredness
mtcars$hpwt <- mtcars$hp/mtcars$wt
cor(mtcars$hpwt, mtcars$wt)
```

```
## [1] 0.05406
```

```
# Model 1
fit1 <- lm(gpm ~ hpwt + wt, mtcars)
```

Table 2. Model 1: Coefficients Summary

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept)  -0.4015   0.512044 -0.7842 4.393e-01
## hpwt          0.0240   0.007302  3.2863 2.659e-03
## wt            1.4722   0.121554 12.1113 7.237e-13
```

Both variables are staticially significant at the 0.01 level (wt is significant at the 0.001 level) gpm increases 1.47 gallons with every increase in wt of 1 (1000 lbs). The coefficient of hp/wt is 0.023997. The Adjusted R-squared value for Model 1 is 0.8379. VIFs are near 1:

```
##  hpwt    wt
## 1.003 1.003
```

Model 2 is identical to Model 1 with the addition of the transmission variable (am).

```
# include a/m transmission variable
fit2 <- lm(gpm ~ hpwt + wt + am, mtcars)
```

Table 3. Model 2: Coefficients Summary

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept) -0.50080   0.675223 -0.7417 4.645e-01
## hpwt         0.02329   0.008024  2.9030 7.130e-03
## wt           1.50236   0.179929  8.3498 4.392e-09
## ammanual     0.08369   0.362529  0.2309 8.191e-01
```

The transmission variable (am) is not statistically significant in Model 2 and does not add descriptive value to the model – the R-squared value, adjusted for degrees of freedom, is 0.8324. The addition of the a/m variable substantially raises the VIFs:

```
##  hpwt    wt    am
## 1.172 2.126 2.244
```

A review of the diagnostic plots for Model 1 (figs. 3-6) shows that the residuals are approximately normally-distributed and that three American luxury cars have a high influence on the model.

Discussion

A more effective method for measuring the possible influence of transmission type on fuel economy would be a pairwise comparison of the various car models, each with both transmissions. For example, the fuel consumption measures for a Toyota Corolla with a manual transmission could be compared with that of a Corolla with an automatic transmission.

References

Henderson, Harold V. and Paul F. Velleman. "Building Multiple Regression Models Interactively" in *Biometrics*, Vol. 37, No. 2 (Jun., 1981), pp. 391-411.

Appendix

Fig. 1. Miles per Gallon vs. Weight

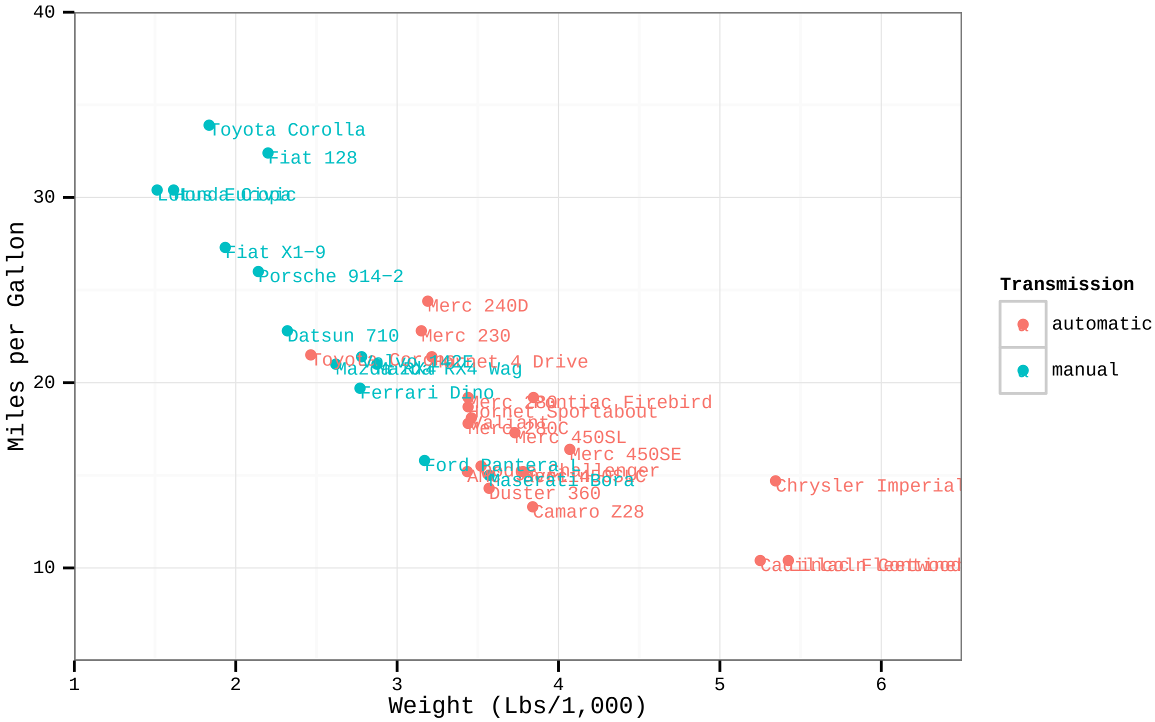

Fig. 2. Gallons per 100 Miles vs. Weight

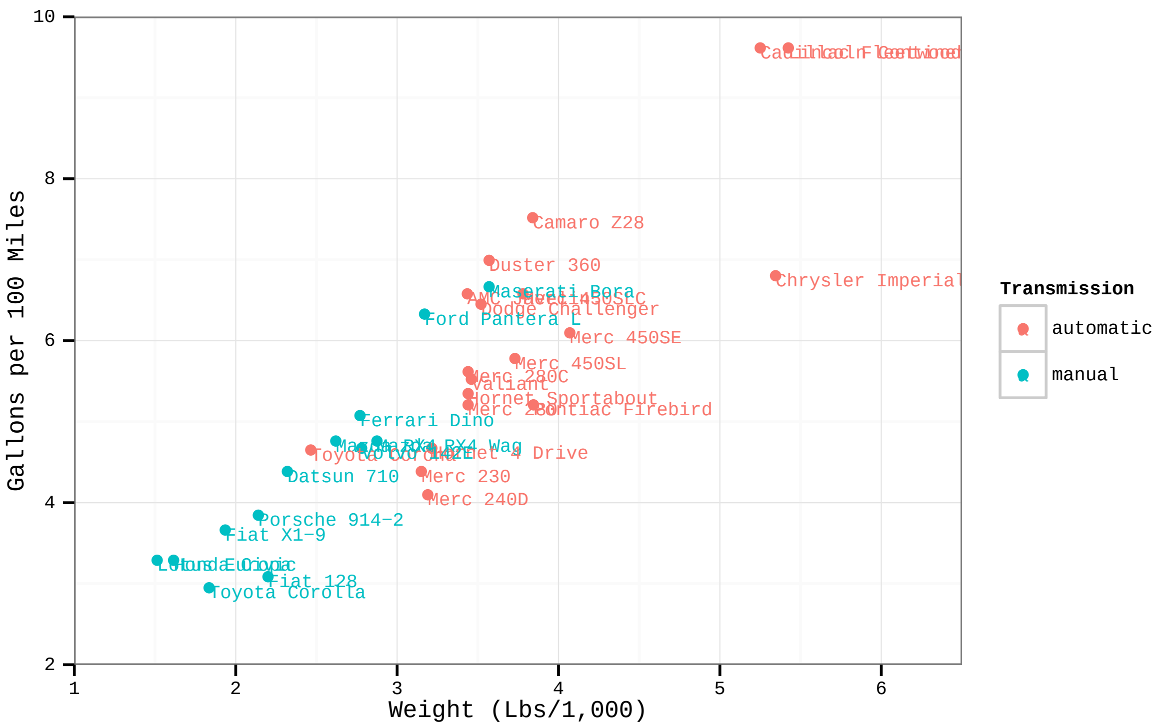

Fig. 3. Model 1: Residuals vs. Fitted Values

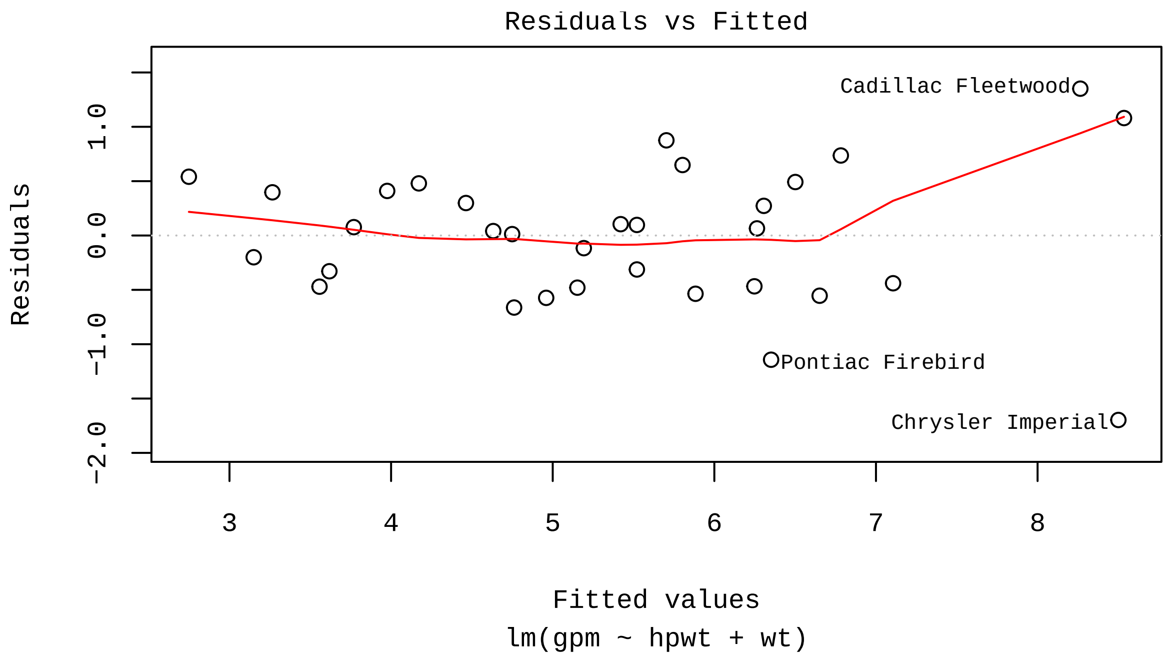

Fig. 4. Model 1: Normal Q-Q

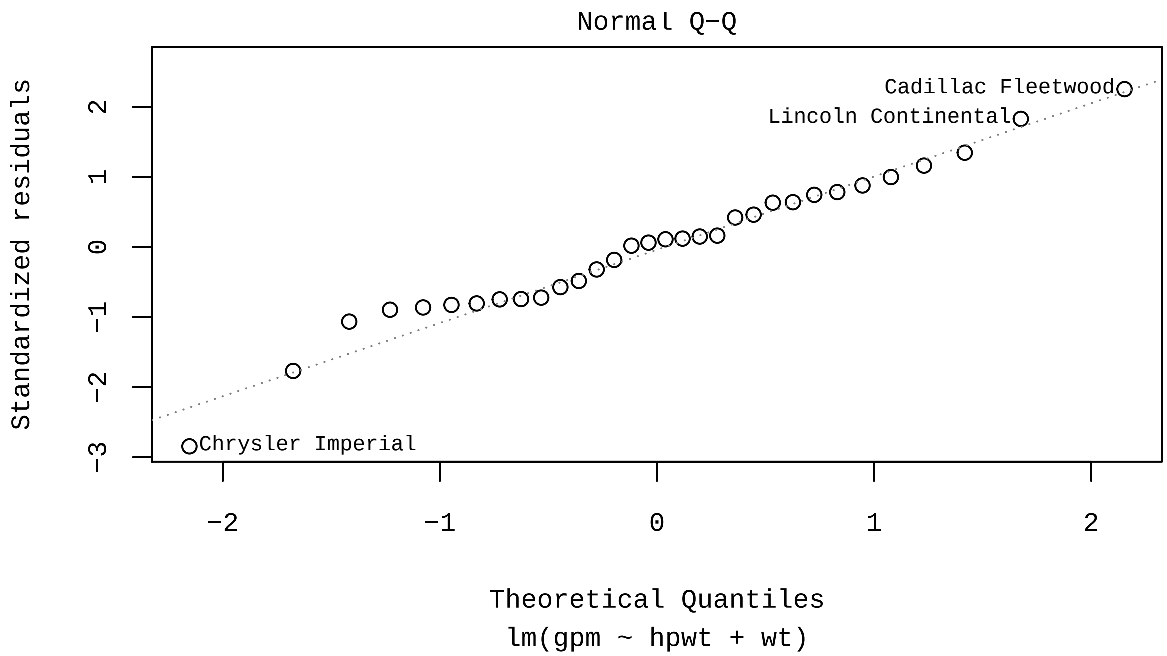

Fig. 5. Model 1: Scale-Location

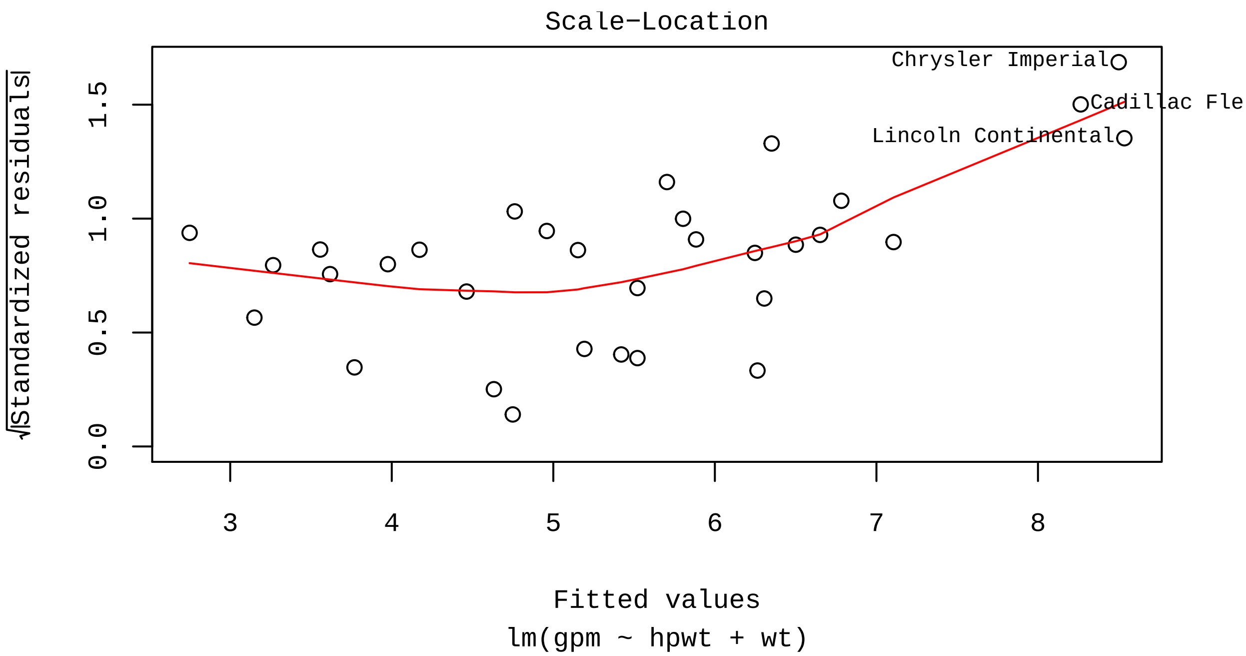

Fig. 6. Model 1: Cook's Distance