Mpg vs. Transmission Type

Executive Summary

It's a common belief that automatics get better gas mileage, but if you must use a heavy weight vehicle manuals are better. The difference when choosing automatic over manual for a given weight x in 1000s of pounds is estimated by 14.88 - 5.3x

Data Exploration

Transmission type for best gas mileage

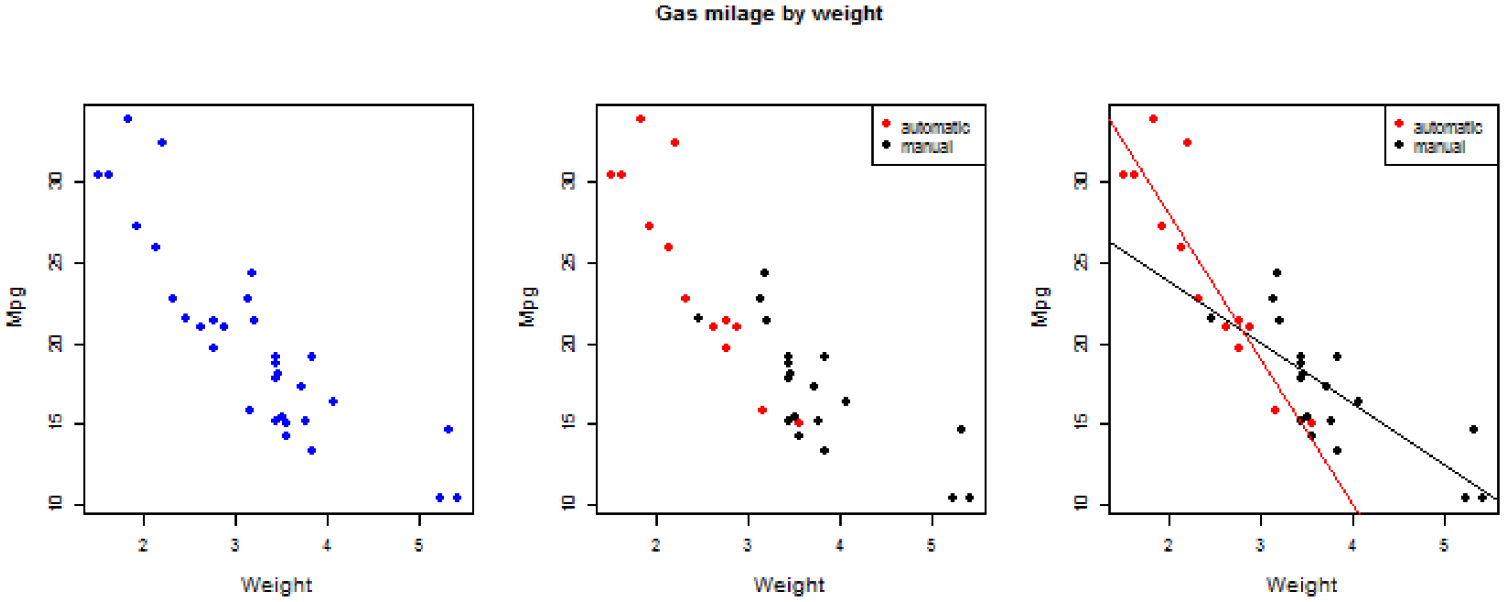
A simple regression from transmission to gas mileage says automatics are better, however, it is a misleading model. Physics suggests weight should matter most.

```
fit_all <- lm(mpg ~ ., data = mtcars)
sort(abs(summary(fit_all)$coefficients[,1]), decreasing = TRUE)[1:4]
```

```
## (Intercept)          wt          am        qsec
##      12.303       3.715       2.520       0.821
```

The coefficient of weight is -3.7153, which is the change in mpg for each increase of 1000lbs, in a model where all variables are used. It has the largest absolute value, supporting our understanding. Let's consider weight's affect on gas mileage.

```
automatics <- mtcars[mtcars$am == 1,];
manuals <- mtcars[mtcars$am == 0,]
fit_manwt <- lm(mpg ~ wt, data = manuals)
fit_autowt <- lm(mpg ~ wt, data = automatics)
```



Weight seems to explain everything, but automatics and manuals are nearly separated, which deserves exploring. The right graph shows each type is affected by weight differently, which likely explains the separation in type choice. Here we see that transmission type does not significantly add to the model:

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt
## Model 2: mpg ~ wt + am
##   Res.Df RSS Df Sum of Sq  F Pr(>F)
## 1     30 278
## 2     29 278  1   0.00224  0   0.99
```

Quantifying the difference between automatic and manual

```
intercept <- coef(fit_autowt)[1] - coef(fit_manwt)[1]
slope <- coef(fit_autowt)[2] - coef(fit_manwt)[2]
intersectionPoint <- -intercept  / slope
```

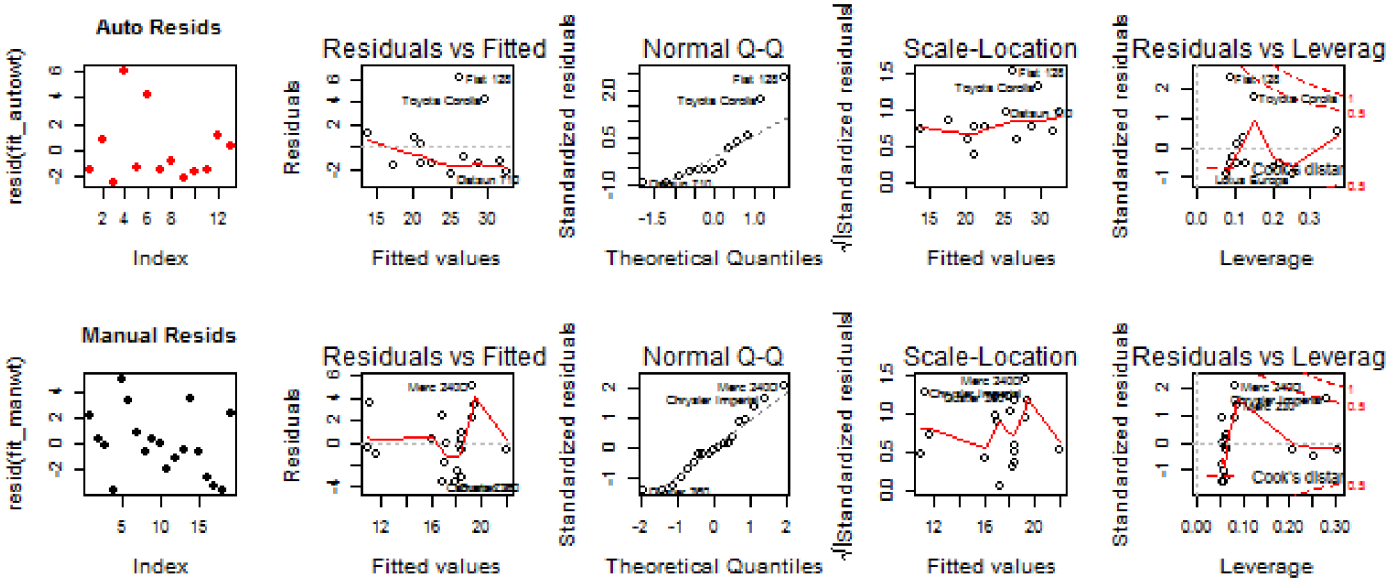
The regressions cross at 2808.12lbs. Below this, automatics are superior, but not above. If cars can use both types across all weights and the existing models would remain consistent, we can give an extrapolated difference between types by subtracing manual predictions from automatic.

```
c(intercept, slope)
```

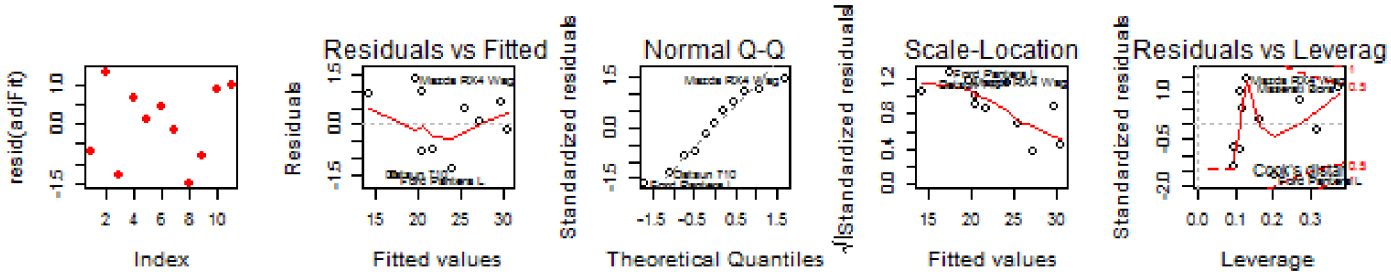
```
## (Intercept)          wt
##      14.878      -5.298
```

The equation of this difference is 14.8784 -5.2984x and we could easily expand this to incorporate ranges.

Diagnostics



Top row: automatics, bottom: manuals. The manual model is cleaner with better residual dispersion. The Toyota Corolla and Fiat 128 stand out in automatics. The Toyota Corolla has the largest dfbeta for both the intercept and the weight coefficient. The Fiat 128 has the largest dffit value. By removing them we get cleaner residuals.



We could easily change our quatified difference model to ignore these points if necessary, but it is doubtful that the difference formula will be put to much use due to the clear implication that weight is the factor to listen to when aiming for good gas mileage.

Appendix