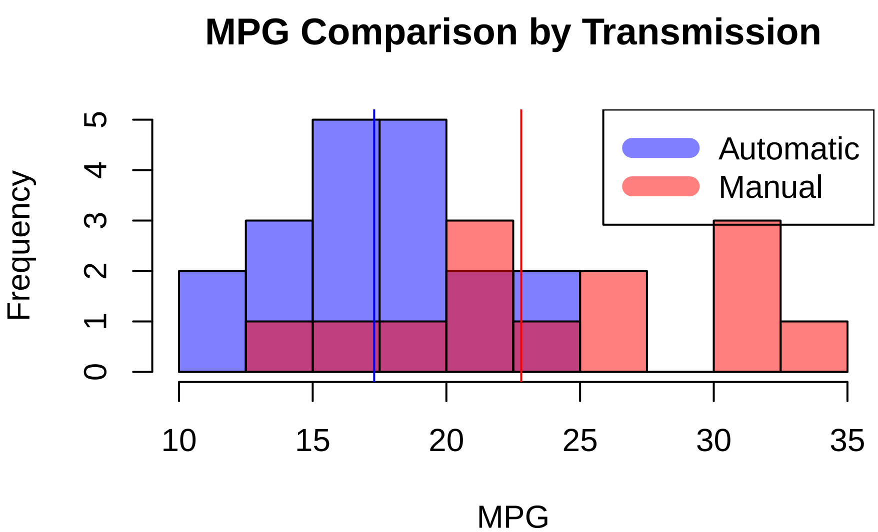
Does Transmission Type Affect MPG?

Summary

The goal of this analysis is to answer the question "Do cars with automatic and manual transmissions have differing fuel efficiencies?" And, if so, what is the difference in MPG? We use the dataset from the 1974 Motor Trend US magazine, which details 32 1973-1974 automobile models. Overall, models with automatic transmissions have a median MPG of 17.3 miles per gallon, while those with manual transmission have a median of 22.8 miles per gallon, a difference of ~25%. However, the difference can almost entirely be explained by confounding factors. Car weight alone explains most of the difference, with horsepower explaining almost all of the remaining variation. After accounting for these two variables, the effect of transmission type on fuel efficiency is not significant. The conclusion is that transmissions have at most a neglible effect on MPG after accounting for other factors.

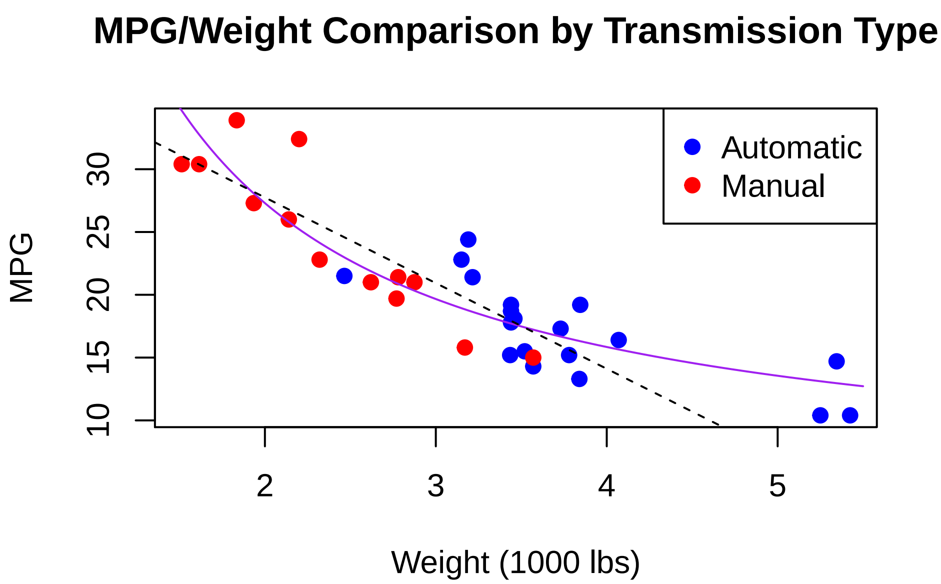
Analysis

The dataset contains 11 variables, which are Miles/gallon, number of cylinders, displacement, gross horsepower, rear axle ratio, weight, 1/4 mile time, type of engine (V vs. straight), transmission type (auto vs manual), number of forward gears, and number of carburetors. Since the two variables we are most interested in are MPG and transmission type let's start with those.



From the plot above, it's clear that fuel efficiencies differ for automatic and manual transmissions. The vertical lines show the median MPGs for the two transmission types, which differ by about 25%. However, it's possible this difference is due wholly or in part to some of the other variables.

For example, we could examine the effect on MPG from V versus straight engines (See Appendix for plot). The difference in the median MPG between engine types is even high than for transmission, at ~30%. It's possible these two effects are related, if cars with manual transmission tended to also have V-type engines, although a χ^2-test shows no significant correlation.



Possibly the strongest correlation is between weight and MPG, as shown in the plot above. The relation doesn't appear linear, which makes sense, since a linear relation would predict that MPG would become negative at some point. Instead, the relation MPG $\propto$ 1/weight, shown with a solid line, appears to explain the trend in the data without invoking transmission type. A similar case can be make for horsepower, as shown in the Appendix.

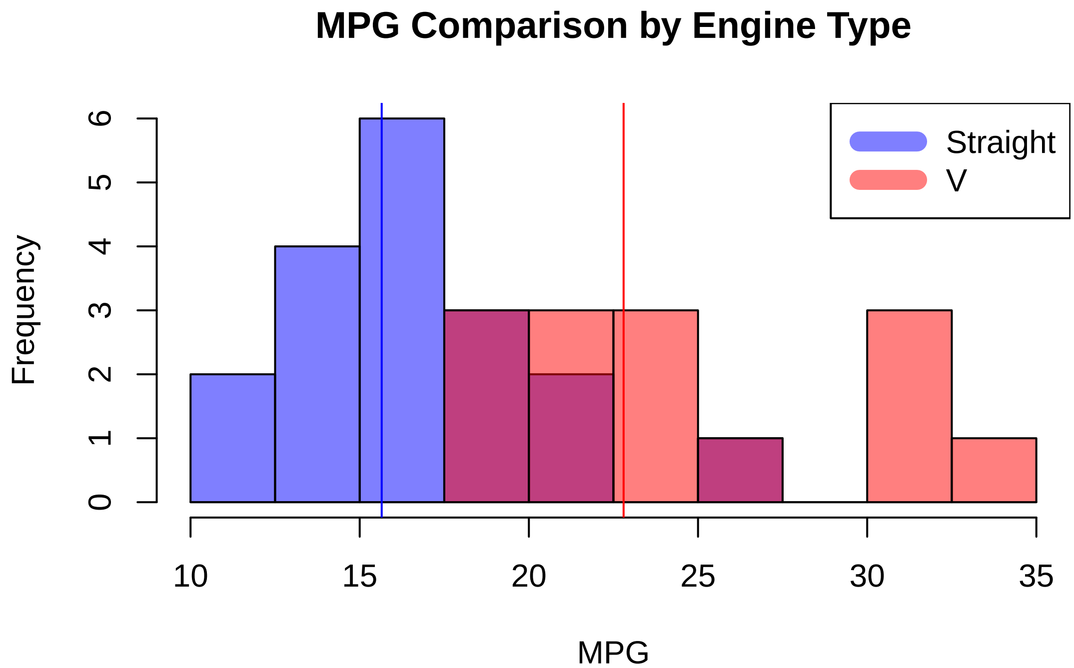
To show which variables explain trends in MPG, we use ANOVA to analyze a series of linear fits. First, we only analyze models where weight $<$ 5000 lbs, where the relation between weight and MPG is approximately linear, as evidenced by the lack of a clear pattern in the residual plot (see Appendix). This fit is shown with a dotted line above. Then, we progressively add weight, horsepower, transmission type, engine type, and everything else (excluding number of cylinders and 1/4 mile time, which are redundant with other variables, as shown in the Appendix) into the model.

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt
## Model 2: mpg ~ wt + hp
## Model 3: mpg ~ wt + hp + factor(am)
## Model 4: mpg ~ wt + hp + factor(am) + factor(vs)
## Model 5: mpg ~ wt + hp + factor(am) + factor(vs) + gear + carb + drat
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     27 201.62
## 2     26 144.22  1    57.403 9.0136 0.006787 **
## 3     25 143.99  1     0.231 0.0362 0.850912
## 4     24 138.48  1     5.507 0.8648 0.362977
## 5     21 133.74  3     4.740 0.2481 0.861741
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

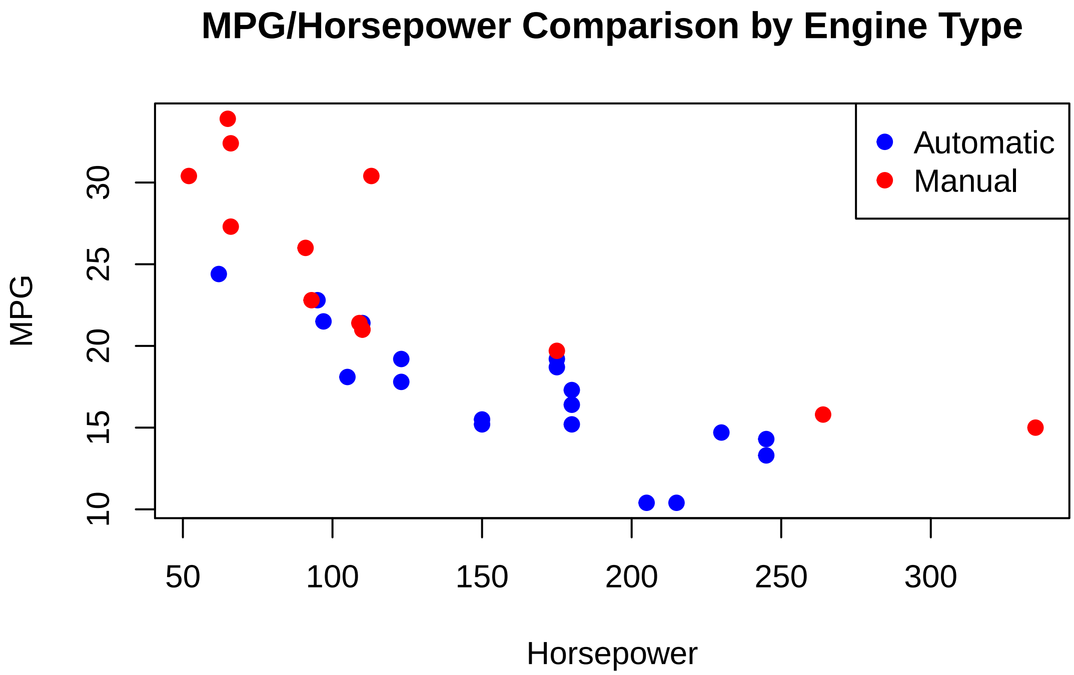
The results show that adding in horsepower has a significant effect on the fit (at the 99% level), but no further additions do. Reversing the role of weight and horsepower leads to similar results. This implies that weight and horsepower explain the majority of variation in MPG, while all other variables explain only negligible amounts. This includes transmission type, which means there is no significant difference in fuel efficieny between manual and automatic transmissions after accounting for other variables.

Appendix

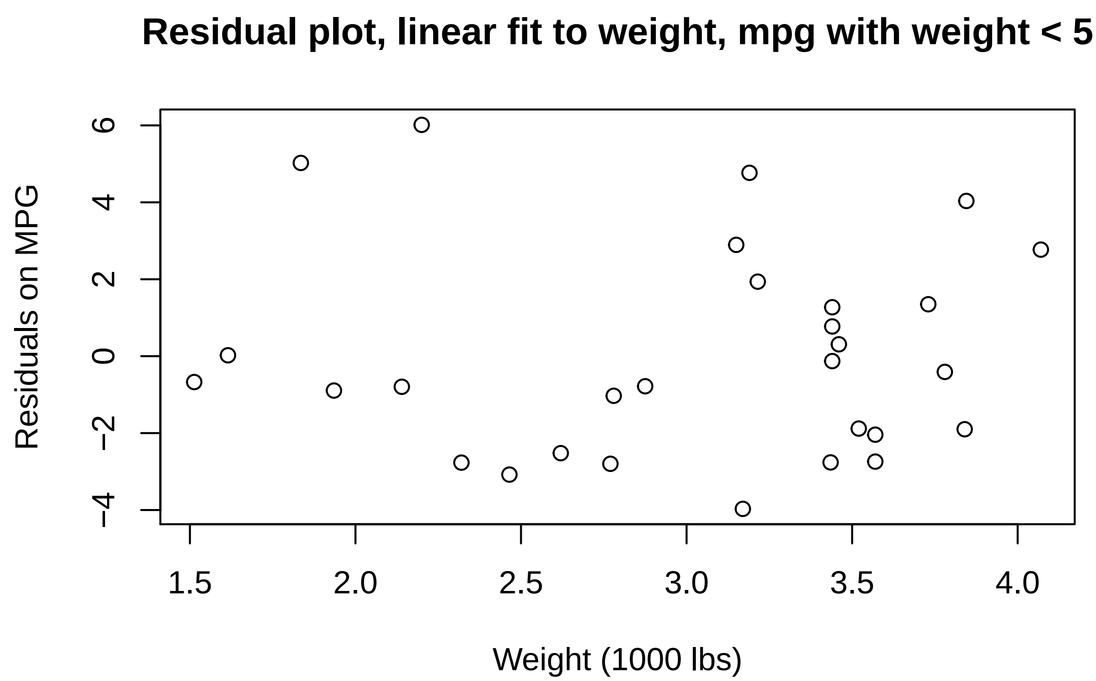
```r
hist(mtcars$mpg[mtcars$vs==0], col = rgb(0,0,1,0.5), xlim = c(10,35), breaks = seq(10,35,5/2),
     xlab = 'MPG', main = 'MPG Comparison by Engine Type')
hist(mtcars$mpg[mtcars$vs==1], col = rgb(1,0,0,0.5), xlim = c(10,35), breaks = seq(10,35,5/2),
     add = T)
abline(v = median(mtcars$mpg[mtcars$vs==0]), col='blue')
abline(v = median(mtcars$mpg[mtcars$vs==1]), col='red')
legend('topright', c('Straight','V'), col=c(rgb(0,0,1,0.5), rgb(1,0,0,0.5)), lwd=10)
```



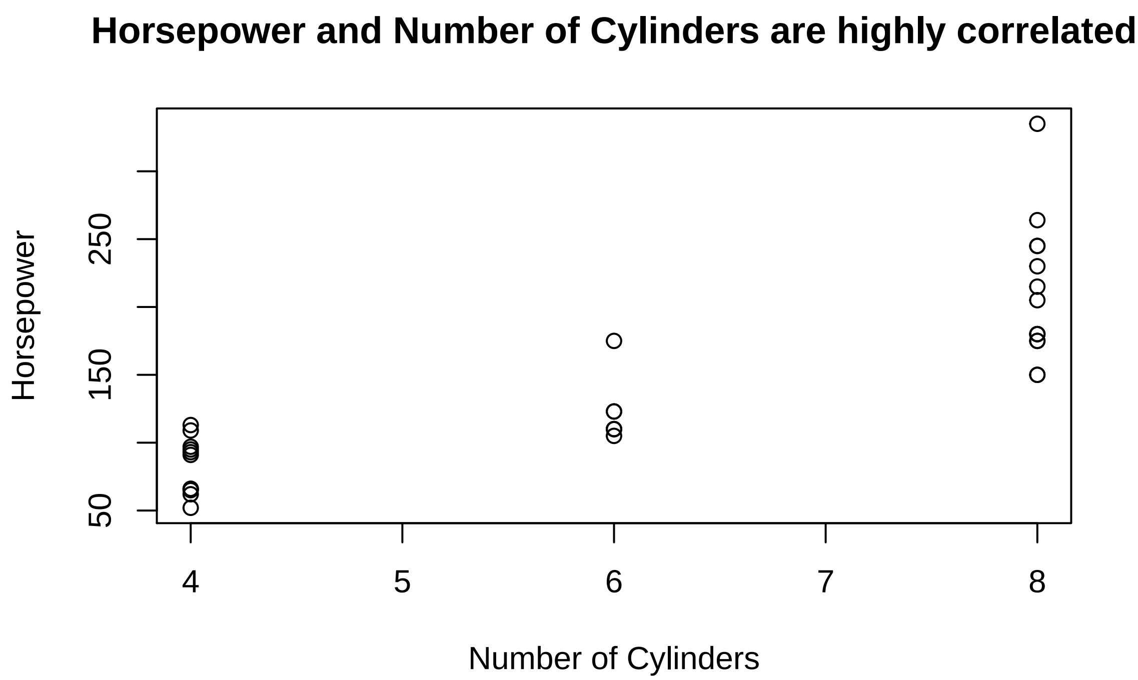

```r
plot(mpg ~ hp, data = mtcars, type = 'n', xlab = 'Horsepower', ylab = 'MPG',
     main = 'MPG/Horsepower Comparison by Engine Type')
points(mpg ~ hp, data = mtcars, subset = am == 0, pch = 19, col = 'blue')
points(mpg ~ hp, data = mtcars, subset = am == 1, pch = 19, col = 'red')
legend('topright',  c('Automatic','Manual'), col = c('blue','red'), pch = 19)
```




```r
plot(mtcut$wt, resid(fit0), xlab = 'Weight (1000 lbs)', ylab = 'Residuals on MPG',
     main = 'Residual plot, linear fit to weight, mpg with weight < 5')
```




```r
plot(hp ~ cyl, data = mtcars, xlab = 'Number of Cylinders', ylab = 'Horsepower',
     main = 'Horsepower and Number of Cylinders are highly correlated')
```




```r
plot(mtcars$hp / mtcars$wt, mtcars$qsec, xlab = 'Horsepower/Weight', ylab = '1/4 Mile Time (sec)',
     main = '1/4 mile time is redundant with Horsepower and Weight')
```