Analysis of Fuel Economy

Executive Summary

Motor Trend is interested in the effects of various car attributes on miles per gallon (mpg). Specifically, they are interested in the following two questions:

- Is an automatic or manual transmission better for MPG?
- Quantify the MPG difference between automatic and manual transmissions.

The dataset in question is stored in the `mtcars` variable, which contains 11 variables for 32 different vehicles. These variables are mpg, cyl, disp, hp, drat, wt, qsec, vs, am, gear, carb. This analysis is particularly interested in the impact of transmission type on fuel economy (mpg). The transmission type of the vehicle is stored in the 9th column, under the variable named am. This variable has two values: 0 == automatic or 1 == manual.

Figure 1 in the Appendix shows a histogram of the mile per gallon ratings for the 32 vehicles in the dataset.

A Quick Dive: Automatic or Manual

Since the impact of this analysis is to detemine which transmission gives better mpg, we will do a quick, surface level comparison.

```
average_mpg_transmission <- data.frame(c(mean(filter(mtcars, am == 0)$mpg),
  mean(filter(mtcars, am == 1)$mpg)), c("Automatic", "Manual"))
colnames(average_mpg_transmission) <- c("mpg", "Type")
average_mpg_transmission
```

```
##        mpg      Type
## 1 17.14737 Automatic
## 2 24.39231    Manual
```

Voila! Manual cars have better fuel economy than automatic vehicles, ignoring all other factors.

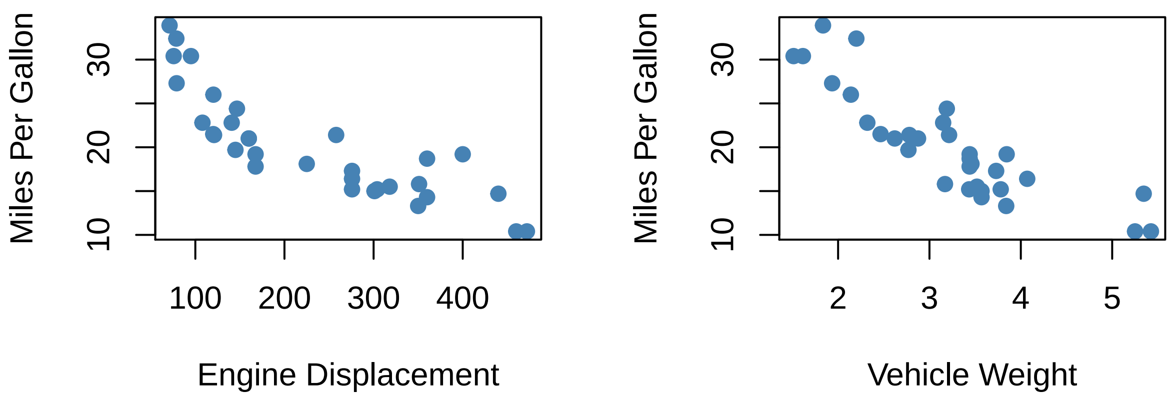
Other Factors

Obviously, there are many factors that impact the fuel economy of a vehicle. Let's looks at a few others here, with additional figures at the end of the report:

```
## Group the cars by cylinders and compute the mean mpg for each group
mtcars %>% group_by(cyl) %>% summarise("mpg" = mean(mpg))
```

```
## Source: local data frame [3 x 2]
##
##   cyl      mpg
## 1   4 26.66364
## 2   6 19.74286
## 3   8 15.10000
```

```r
## mpg vs engine size and mpg vs vehicle weight
par(mfrow = c(1, 2))
plot(mtcars$disp, mtcars$mpg, col = "steelblue", pch = 19,
     xlab = "Engine Displacement", ylab = "Miles Per Gallon")
plot(mtcars$wt, mtcars$mpg, col = "steelblue", pch = 19,
     xlab = "Vehicle Weight", ylab = "Miles Per Gallon")
```



Miles per gallon decreases as the number cyliners increases, as the size of the engine increases, and as the weight of the vehicle increases. (Admittedly, all three of these variables are likely related; more on that later.)

Linear Regressions

Basic Analysis: mpg ~ am

Moving beyond simple data analysis, we will attempt to build a linear model that fits the data. To start with, we will simply look at the linear model with `mpg` as the outcome and `am` as the predictor.

```r
mpg.am <- lm(mpg ~ am, mtcars)
summary(mpg.am)
```

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## am             7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

This model predicts automatic transmissions to get 17.1473684 mpg, and manual transmissions to get an additional 7.2449393 mpg. However, the R^2 value is fairly low, only 0.3597989. Thus, this model only accounts for a small portion of the variance. We will need to add more variables to the model to obtain a better fit.

Correlation

In order to choose what variables to add to the model, we will look at the correlation between variables, specifically for the `mpg` variable:

```
sort(cor(mtcars)[1,])
```

```
##         wt        cyl       disp         hp       carb       qsec 
## -0.8676594 -0.8521620 -0.8475514 -0.7761684 -0.5509251  0.4186840 
##       gear         am         vs       drat        mpg 
##  0.4802848  0.5998324  0.6640389  0.6811719  1.0000000
```

Here we see that weight (wt), number of cylinders (cy), and engine size (disp) all have have a similar, negative impact on mpg. Looking at these variables, we see they are all correlated with each other:

```
c(cor(mtcars$disp, mtcars$cyl), cor(mtcars$cyl, mtcars$wt), cor(mtcars$wt, mtcars$disp))
```

```
## [1] 0.9020329 0.7824958 0.8879799
```

`disp` and `cyl` are the most correlated, with `wt` and `cyl` being the least correlated. Let's add these last two to our model.

```
mpg.am.wt.cyl <- lm(mpg ~ am + wt + factor(cyl), mtcars)
summary(mpg.am.wt.cyl)
```

```
## 
## Call:
## lm(formula = mpg ~ am + wt + factor(cyl), data = mtcars)
## 
## Residuals:
##     Min      1Q  Median      3Q     Max 
## -4.4898 -1.3116 -0.5039  1.4162  5.7758 
## 
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)    
## (Intercept)   33.7536     2.8135  11.997  2.5e-12 ***
## am             0.1501     1.3002   0.115  0.90895    
## wt            -3.1496     0.9080  -3.469  0.00177 ** 
## factor(cyl)6  -4.2573     1.4112  -3.017  0.00551 ** 
## factor(cyl)8  -6.0791     1.6837  -3.611  0.00123 ** 
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 
## Residual standard error: 2.603 on 27 degrees of freedom
## Multiple R-squared:  0.8375, Adjusted R-squared:  0.8134 
## F-statistic: 34.79 on 4 and 27 DF,  p-value: 2.73e-10
```

The intercept, 33.753592, is less valuable that before, as this would be for a 4 cylinders, but weighing 0 pounds. The impact of transmission type, `am`, has also been significantly descreased, resulting in an increase of only 0.1501031 miles per gallon for a manaul transmission. As expected, heavier cars get lower mpg (-3.1495978 miles per gallon per 1,000 pounds), and the more cylinders a car has, the lower it's fuel economy.

Summary

It appears manual cars provide better gas mileage than automatic cars, showing a 0.1501031 increase in miles per gallon in our last analysis. However, the standard error on this estmiate is large (1.3002231), and the 95% confidence for transmission contains 0: [-2.5177344, 2.8179406].

Thus, it appears that the impact of transmission type on fuel economy is minimal, with other factors, such as `wt` and `cyl`, having a significantly greater influence. In fact, when these two factors are grouped by transmission type, we see that the manual cars in the data set also happen to be lighter and have few cylinders.

```r
cyl_am <- mtcars %>% group_by(am) %>% summarise("Avg. Cyl" = mean(cyl))
wt_am <- mtcars %>% group_by(am) %>% summarise("Avg. Wt" = mean(wt))

inner_join(cyl_am, wt_am, disp_am, by = "am")
```

```
## Source: local data frame [2 x 3]
##
##   am Avg. Cyl  Avg. Wt
## 1  0 6.947368 3.768895
## 2  1 5.076923 2.411000
```

Supporting Figures

```r
hist(mtcars$mpg, breaks = 18, col = "steelblue",
     main = "Figure 1: Miles Per Gallon Histogram", xlab = "Miles Per Gallon")
```

Figure 1: Miles Per Gallon Histogram

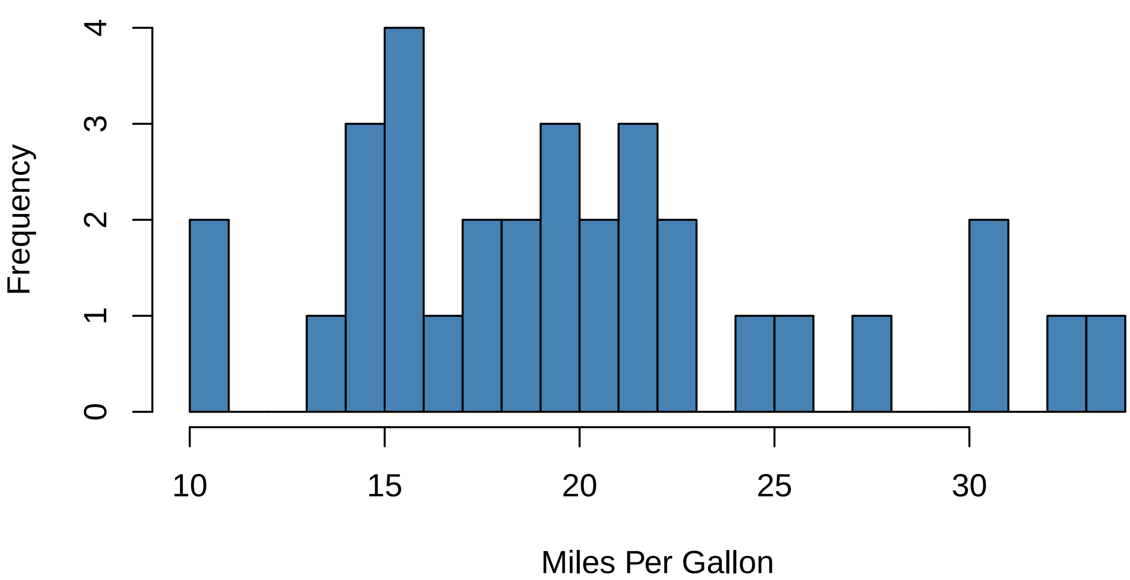


Figure 2: Summary Plots for mpg ~ am + wt + cyl

```r
par(mfrow = c(2,2))
plot(mpg.am.wt.cyl)
```