Car Analysis

Zach Christensen

May 17, 2015

You work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions:

"Is an automatic or manual transmission better for MPG"

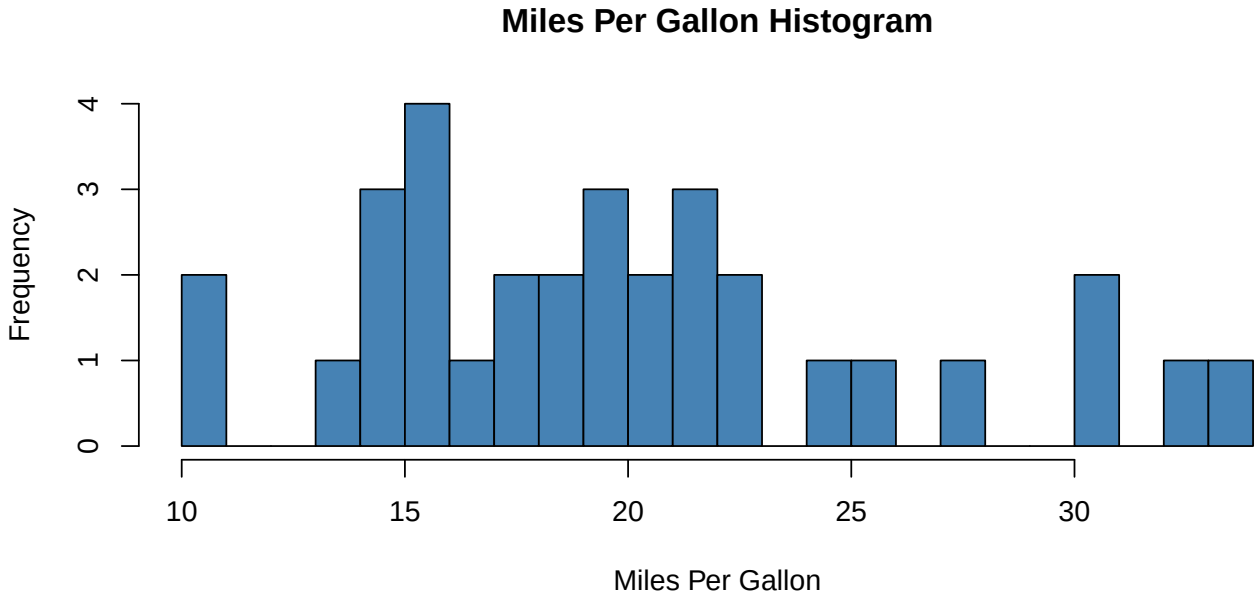
"Quantify the MPG difference between automatic and manual transmissions"

Executive Summary

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.

Below is a histogram of the mile per gallon ratings for the 32 vehicles in the dataset.

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



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. (See Figure 2 in the supporting figures section.)

```
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")

ave_mpg_plot <- ggplot(average_mpg_transmission, aes(x = Type, y = mpg, fill = Type)) +
  labs(x = "", y = "Average MPG", title = "Average MPG by Transmission Type") +
  geom_bar(stat = "identity") + theme(legend.position="none")

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:

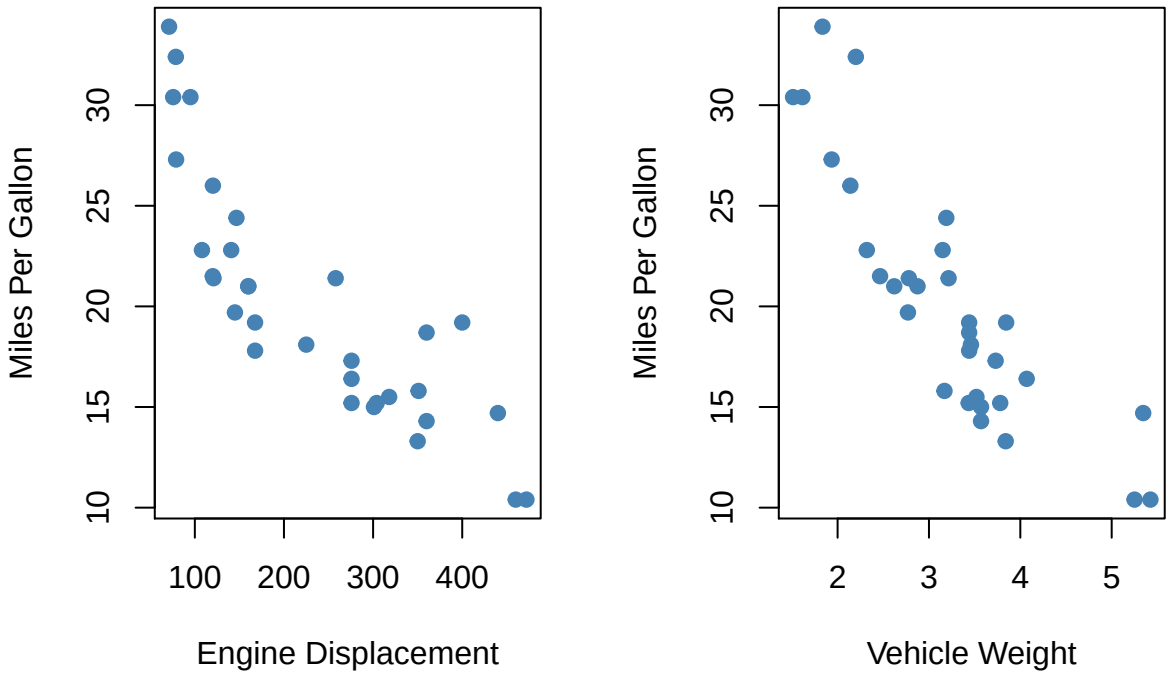
```
## Group the cars by cylinders, compute the mean mpg for each group, then plot
cyl_sum <- mtcars %>% group_by(cyl) %>% summarise("mpg" = mean(mpg))
cylPlot <- ggplot(cyl_sum, aes(x = cyl, y = mpg, fill = cyl)) + geom_bar(stat = "identity") +
  labs(x = "", y = "Average MPG", title = "Average MPG by Number of Cylinders")
cyl_sum
```

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

```
## Plot the next two beside each other
par(mfrow = c(1, 2))

## Plot the mpg against the engine size
plot(mtcars$disp, mtcars$mpg, col = "steelblue", pch = 19,
     xlab = "Engine Displacement", ylab = "Miles Per Gallon")

## Plot the mpg against vehicle weight
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.

```
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 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
```

Supporting Figures

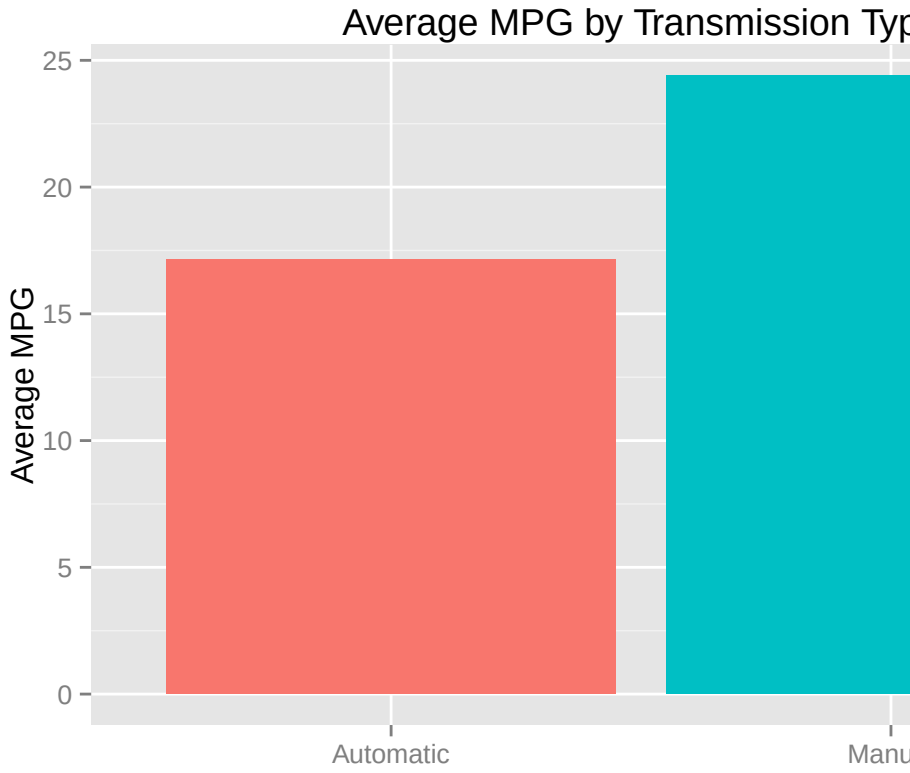


Figure 2: Average MPG By Transmission Type

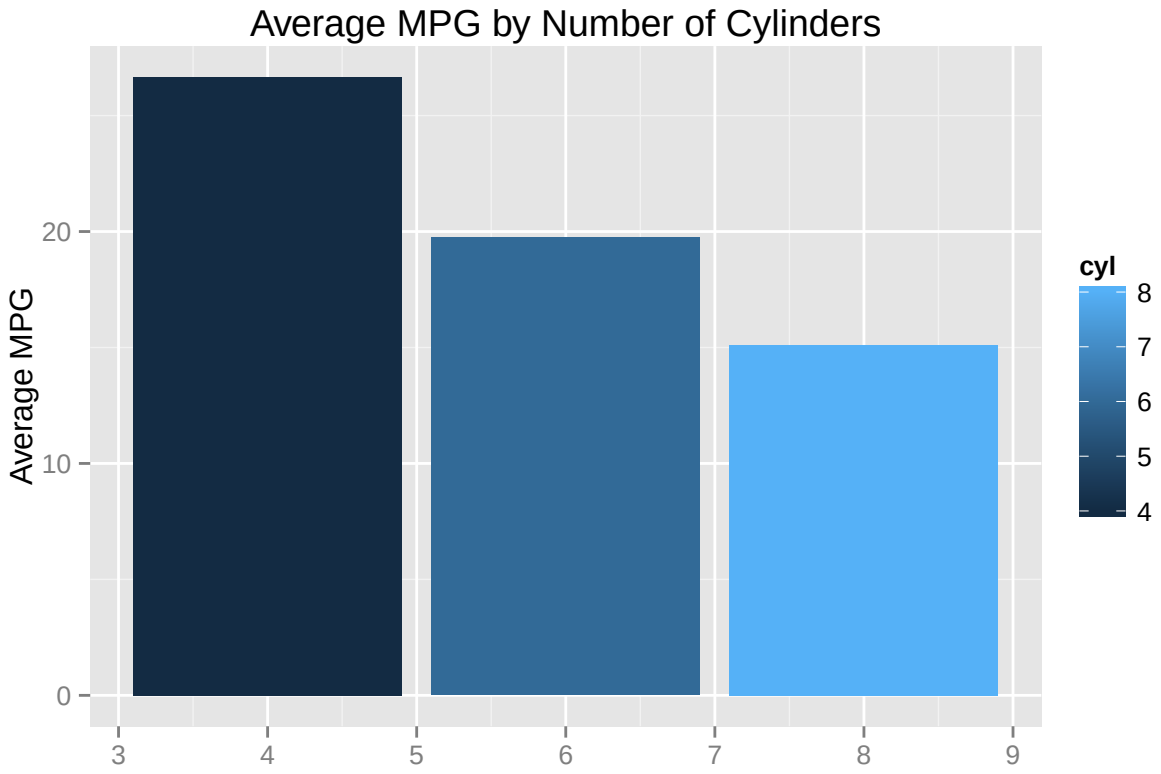

Average MPG by Number of Cylinders
Average MPG
0
10
20
3
4
5
6
7
8
9
cyl
8
7
6
5
4

Figure 3: