Factors That Affect Car Mileage

Executive Summary

With the ever increase in gas prices, it has become quite evident that understanding the real factors that impact mileage (efficiency) of a car is important for the success and growth of any car manufactures. We at Motor Trend has an obligation to enlighten these manufacturers so at to produce better cars, there by helping the consumers. For this, I utilized the 'mtcars' dataset which reflect the real world scenarios of cars with different specifications.

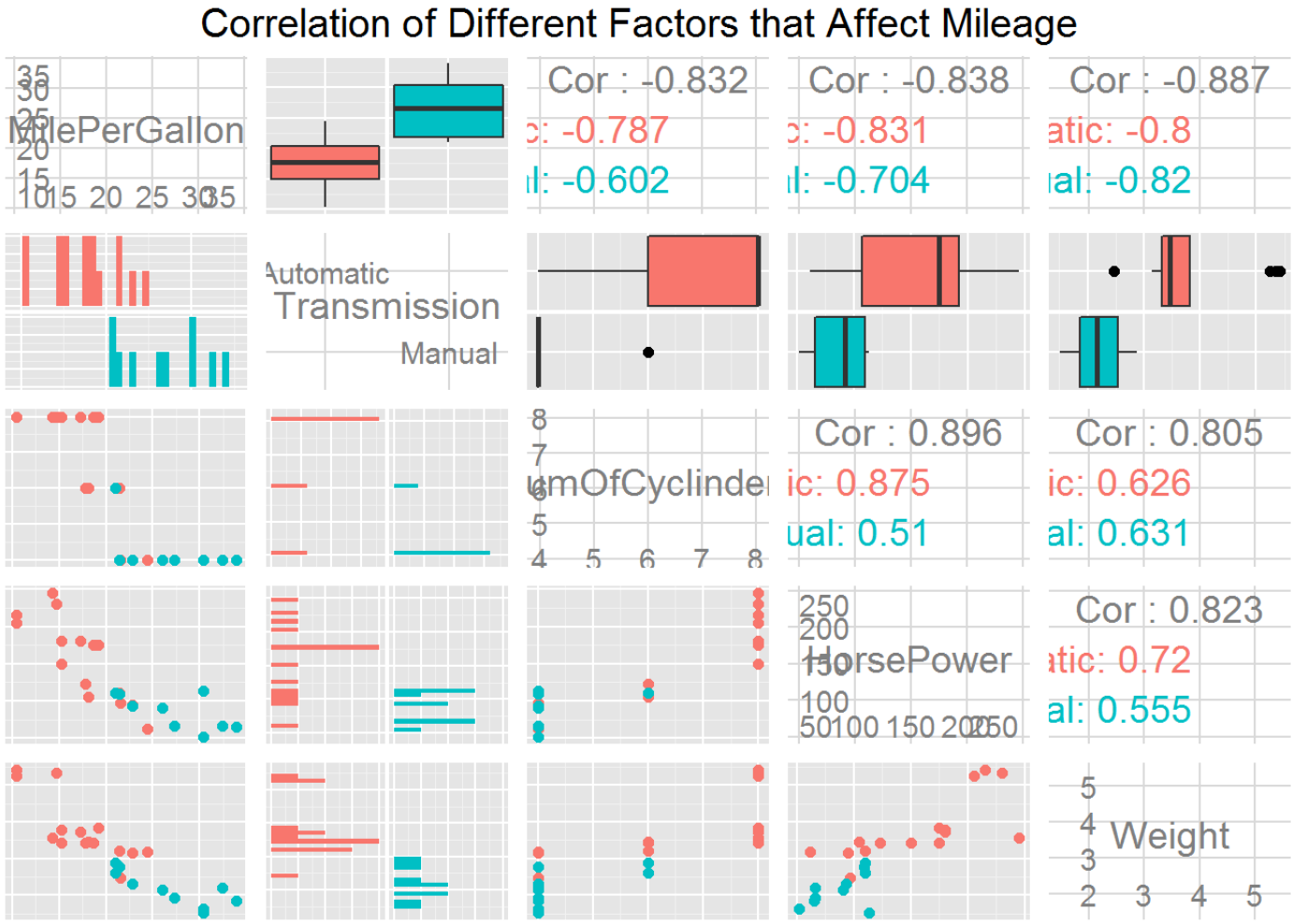
My analysis/study, which is described in detail below, has revealed that the kind of tranmission (Automatic or Manual) play an intergral part in setting the mileage of a car. However, other factors like 'NumOfCyclinders', 'Weight', 'HorsePower' etc also affect the mileage. From Figure 1 below and summary statistics learned from my model, it is quite evident that weight of car in conjunction with it's tranmission type has the greatest impact on mileage. Another interesting fact learned from the model is that, the magnitude of the negative correlation between mileage and other factors ('NumOfCyclinders', 'Weight', 'HorsePower' etc) is greater for manual tranmission. Clearly the outcomes for Model A and Model B indicates that we cannot solely dictate that tranmission is the main factor that affects mileage.

Questions Is an automatic or manual transmission better for MPG?: - Holding all other factors constant in our model, it is quite evident that manual tranmission cars yield better mileage than automatic cars.

Quantify the MPG difference between automatic and manual transmissions: - From my analysis, manaual tranmission have better mileage b/c their average weight, horsepower, num of cyclinder etc are generally smaller than automatic cars.

Modeling and Analysis

Here's a visual cue on how the different variables affect the mileage. **Figure 1**



Clearly, manual tranmission cars have more mileage than automatic tranmission. In fact, the mileage for automatic tranmission range from ~10 miles to 23 miles, while manual transmission ranges from ~ 17 miles to 35 miles.More over, other factors like number of cyclinders, horsepower, weight etc also have a negative impact on Mileages. Interestingly, this negative impact is higher for manual tranmission than automatic transmission. Figure 3 below gives you a prespective of the relationship between Mileage and other factores.

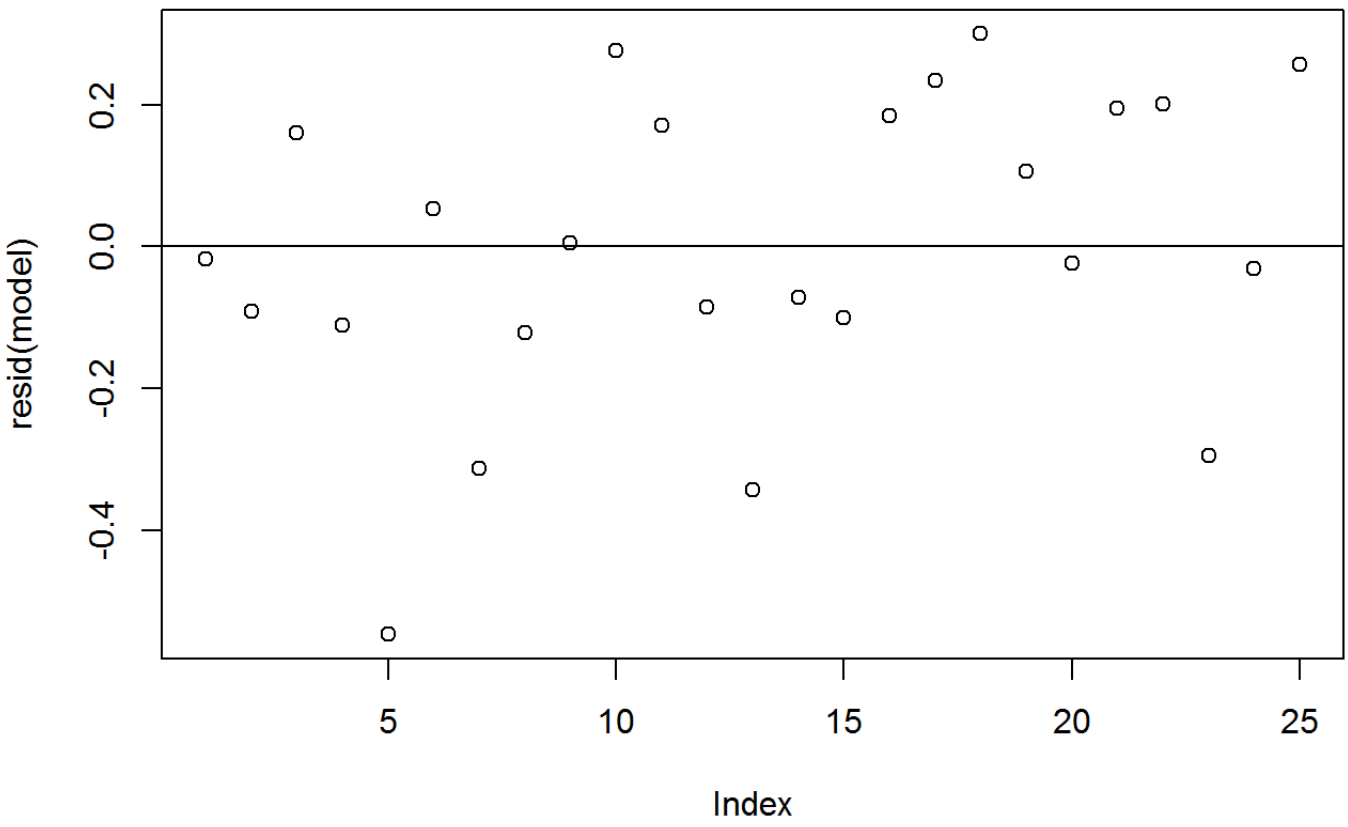
For our analysis, let's create two models:

1. Model A - consider all of these covariates: "MilePerGallon", "NumOfCyclinders", "Displacement", "HorsePower","RearAxleRatio", "Weight","qsec","V/S", "NumOfFowardGears","NumOfCarburetors"
2. Model B - consider these four covariates: "MilePerGallon", "NumOfCyclinders", "HorsePower","Weight"

Model A Results

Residual Plot and Summary

```
# Residual plot
plot(resid(model))
abline(h=0)
```




```
# Model Summary
summary(model)
```

```
##
## Call:
## lm(formula = Transmission ~ ., data = dfT)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.5473 -0.1000 -0.0169  0.1851  0.3003
##
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)       4.238991   2.651330    1.60    0.132
## MilePerGallon     0.028034   0.026063    1.08    0.300
## NumOfCyclinders  -0.100869   0.145680   -0.69    0.500
## Displacement     -0.002094   0.002341   -0.89    0.386
## HorsePower       -0.000736   0.003261   -0.23    0.825
## RearAxleRatio     0.193271   0.275088    0.70    0.494
## Weight            0.243299   0.263951    0.92    0.372
## qsec             -0.199716   0.086435   -2.31    0.037 *
## `V/S`            -0.103303   0.268882   -0.38    0.707
## NumOfFowardGears  0.018310   0.231432    0.08    0.938
## NumOfCarburetors -0.042174   0.129392   -0.33    0.749
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.284 on 14 degrees of freedom
## Multiple R-squared:  0.812,  Adjusted R-squared:  0.677
## F-statistic: 6.03 on 10 and 14 DF,  p-value: 0.00136
```

```
# The totoal residual error
deviance(model)
```

```
## [1] 1.131
```

The Residuals section provides summary statistics for the errors in our predictions. The maximum error of 1.131 suggest that the model under-predicted MilePerGallon by nearly 1 points for at least one observation.

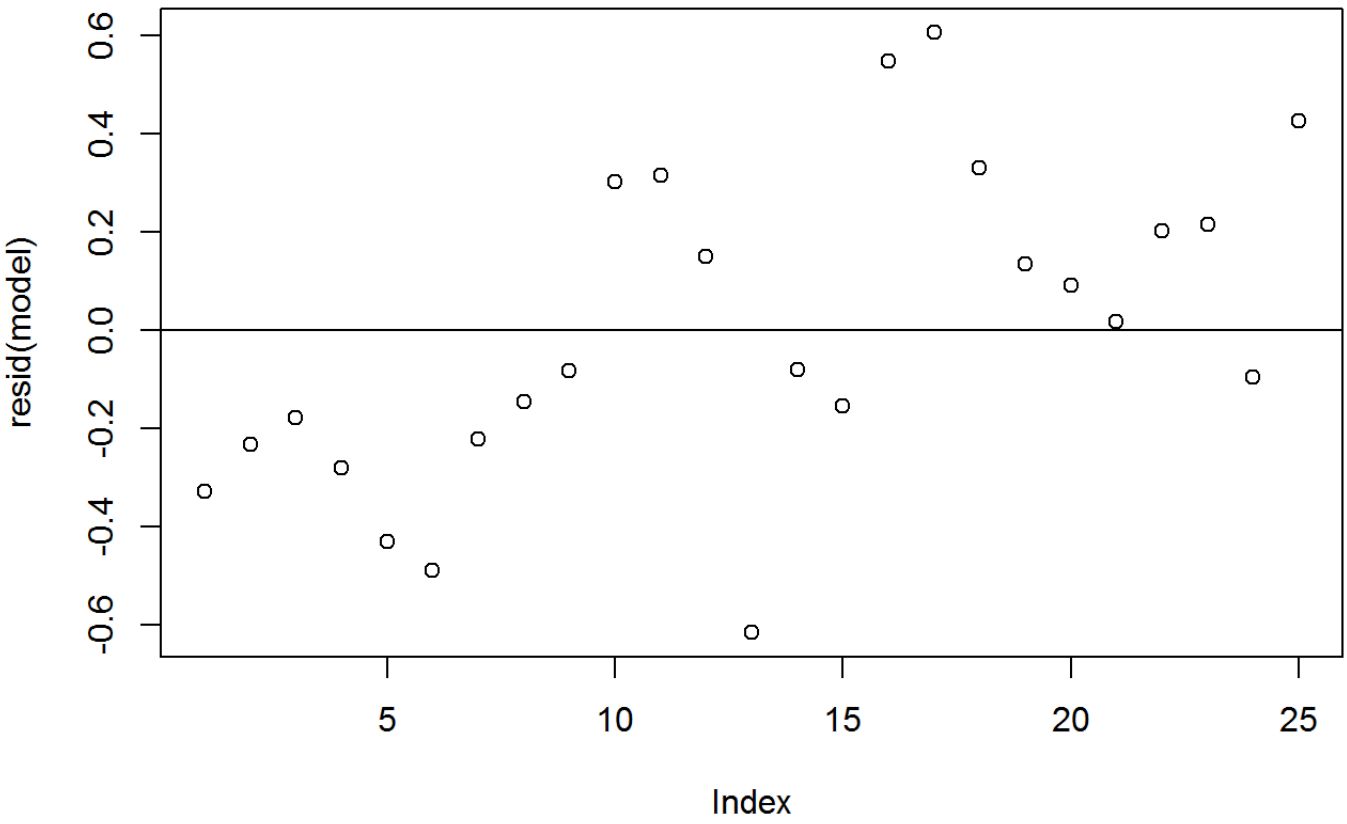
For the coefficients, the column labeled Estimate contains the estimated regression coefficients as calculated by ordinary least squares.Theoretically, if a variable's coefficient is zero then the variable is worthless; it adds nothing to the model. In our case, the estimate for 'MilePerGallon' is 0.028 and hence it does add some value to the model. Concretly speaking, we estimate the mileage to increase if the transmission is switched to manual.

R2 is a measure of the model's quality. Bigger is better. Mathematically, it is the fraction of the variance of y that is explained by the regression model. The remaining variance is not explained by the model, so it must be due to other factors. In our case, the model explains 0.812 (81.2%) of the variance of model, which is surprisingly good.

Model B Results

Residual Plot and Summary

```
# Residual plot
plot(resid(model))
abline(h=0)
```




```
# Model Summary
summary(model)
```

```
##
## Call:
## lm(formula = Transmission ~ MilePerGallon + NumOfCyclinders +
##     HorsePower + Weight, data = dfT)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.6154 -0.2225 -0.0805  0.2150  0.6049
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)      1.67558    1.09618    1.53     0.14
## MilePerGallon    0.02498    0.02747    0.91     0.37
## NumOfCyclinders -0.07502    0.09483   -0.79     0.44
## HorsePower       0.00266    0.00316    0.84     0.41
## Weight          -0.22528    0.15233   -1.48     0.15
##
## Residual standard error: 0.352 on 20 degrees of freedom
## Multiple R-squared:  0.587,  Adjusted R-squared:  0.505
## F-statistic: 7.12 on 4 and 20 DF,  p-value: 0.000982
```

```
# The totoal residual error
deviance(model) # residual sum of squares
```

```
## [1] 2.475
```

In model B, the sum of residual is pretty high 2.475. Also, R2 is 0.587, which means that this model only explains 59% of the variance. This in fact explains that all of the factors has to be consider to precisely understand why mileage varies.

Appendix

Figure 2

```
#~~~~~ frequency
ggplot(dfT, aes(x=Transmission)) + geom_histogram(aes(y=..density..),binwidth=.2, colour="black
", fill="white") +
  ggtitle("Transmission Frequency in Training Set") +
  geom_density(alpha=.2, fill="#FF6666")
```

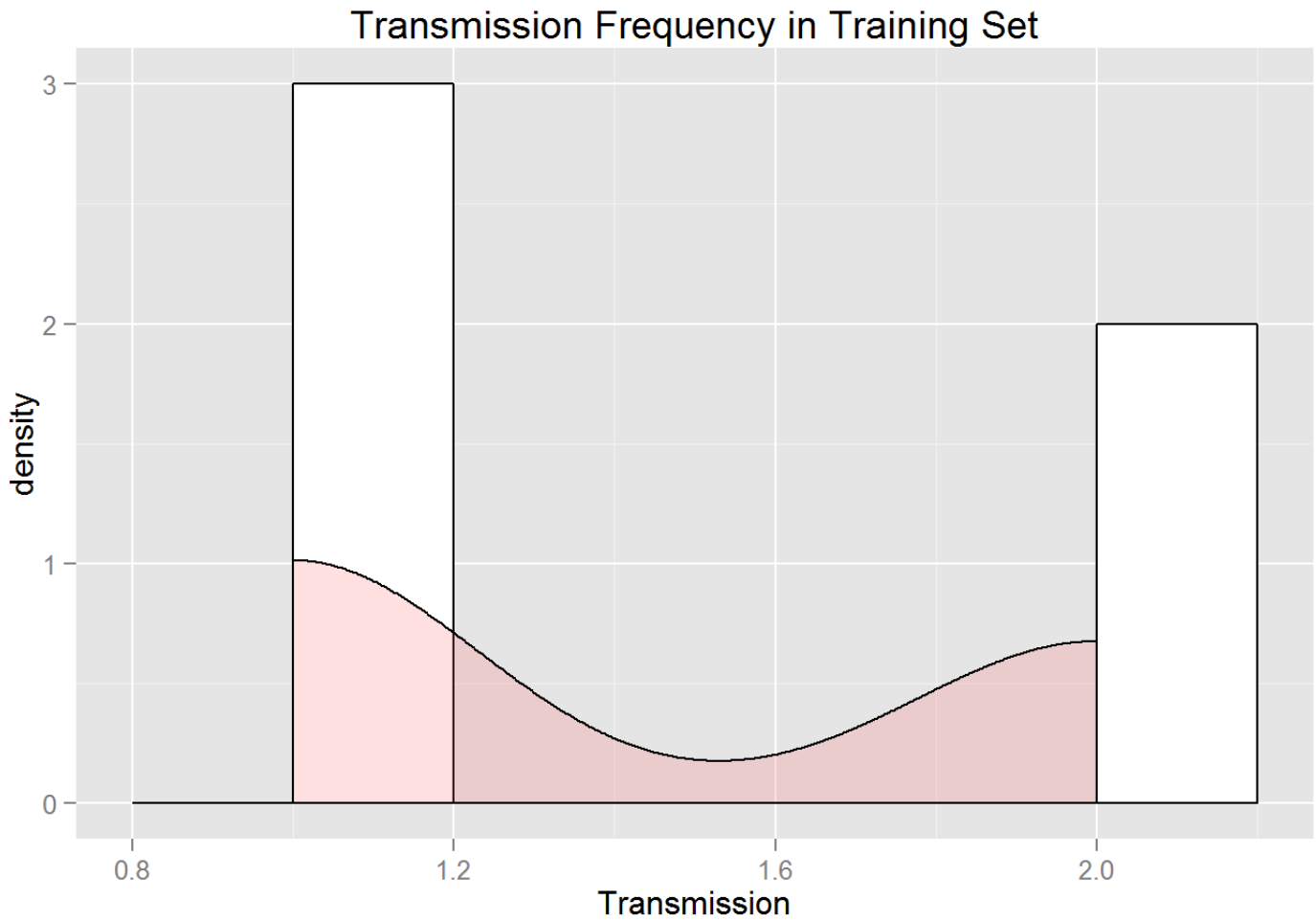


Figure 3

```
#~~~ boxplot
b1 = ggplot(dfT, aes(x=Transmission, y=MilePerGallon)) + geom_line() + geom_point()
b2 = ggplot(dfT, aes(x=HorsePower, y=MilePerGallon)) + geom_line() + geom_point()
b3 = ggplot(dfT, aes(x=NumOfCyclinders, y=MilePerGallon)) + geom_line() + geom_point()
b4 = ggplot(dfT, aes(x=Weight, y=MilePerGallon)) + geom_line() + geom_point()

grid.arrange(b1, b2, b3, b4, main = "Mileage Perspective") # notice the outliers; hoping PCA pr
ocessing should smooth that out.
```

Mileage Perspective