Automatic or Manual Transmission, Which Gets Better Mileage?

Executive Summary

Rising gas prices in the United States resulting from the formation of the OPEC cartel in the late 1970's led to heightened awareness of the fuel efficiency of automobiles. Motor Trend magazine compiled a data set of vehicles' fuel efficiency and various design aspects. In this paper, we'll look at the question of whether one of the vehicles' design aspects, the choice of a manaul versus an automatic transmission, impacts vehicles' fuel efficiency.

The paper concludes that for cars with the same engine horsepower, cars with manual transmissions are somewhere between 3.1 and 7.5 miles per gallon more fuel efficient than cars with automatic transmissions.

Please note that throughout this paper, the R-code used to produce the figures is suppressed in the interest of brevity. The R-code for each figure appears in the Appendix at the end.

mtcars Dataset

Motor Trends' `mtcars` dataset is supplied with R. The description from the mtcars help file reads:

> The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973–74 models).

The help file includes descriptions of columns in the dataset.

To make interpretaion of graph labels and analysis output easier, we'll add a few columns with label-friendly text. The augmented dataset is stored in a data.frame called `mtcars2` which is used throughout this document. Appendix A1 shows the R code for these transformations. The resulting structure, mtcars2 looks like:

```
nrow(mtcars2)
```

```
## [1] 32
```

```
names(mtcars2)
```

```
##  [1] "mpg"       "cyl"       "disp"      "hp"        "drat"
##  [6] "wt"        "qsec"      "vs"        "am"        "gear"
## [11] "carb"      "tran.type" "eng.type"  "drat.size"
```

```
head(mtcars2)
```

```
##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
```

```
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
##                   tran.type eng.type drat.size
## Mazda RX4            Manual  V-Block    Medium
## Mazda RX4 Wag        Manual  V-Block    Medium
## Datsun 710           Manual   Inline    Medium
## Hornet 4 Drive    Automatic   Inline       Low
## Hornet Sportabout Automatic  V-Block       Low
## Valiant           Automatic   Inline       Low
```

Exploratory Data Analyses

Figure 1 below shows how cars' miles per gallon (MPG) breaks out for automatic transmission cars versus cars with manual transmissions. From the figure, it appears that on average, cars with manuals get better mileage than those with automatics.

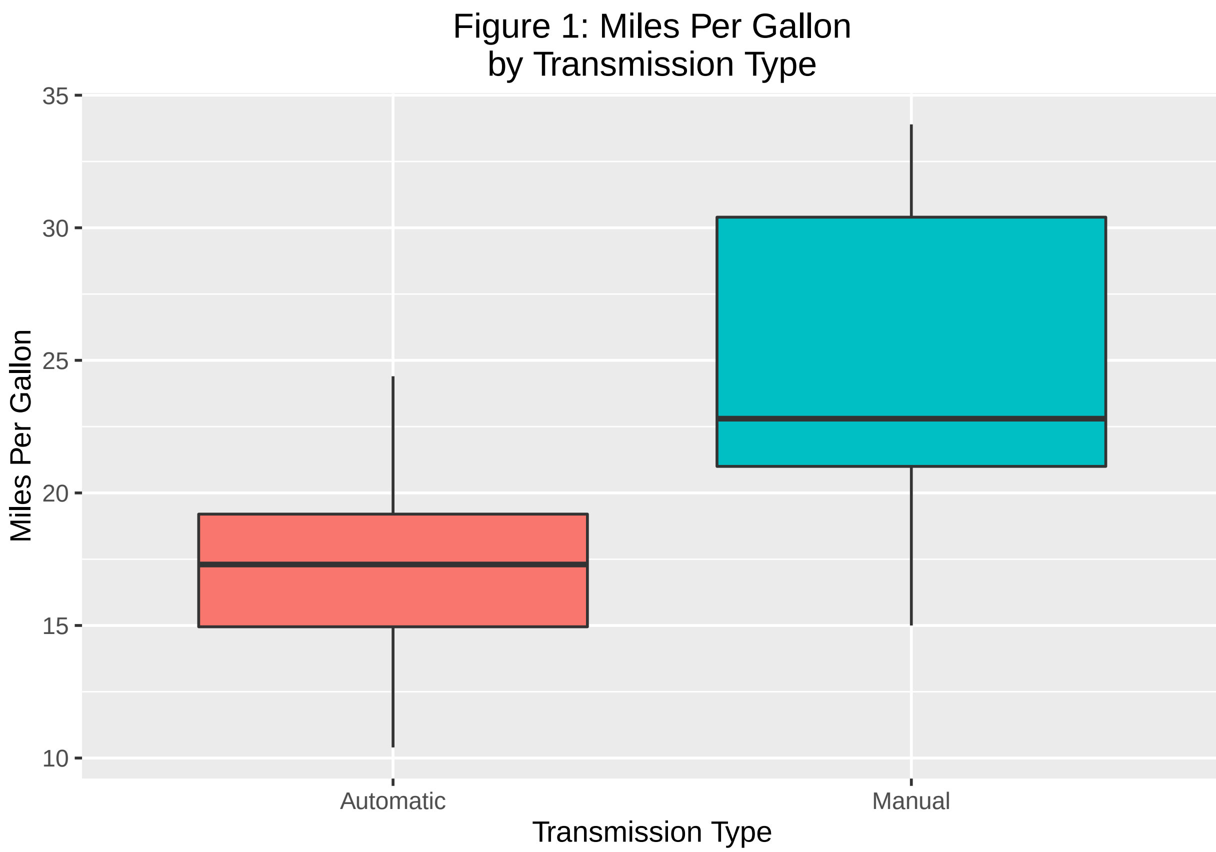
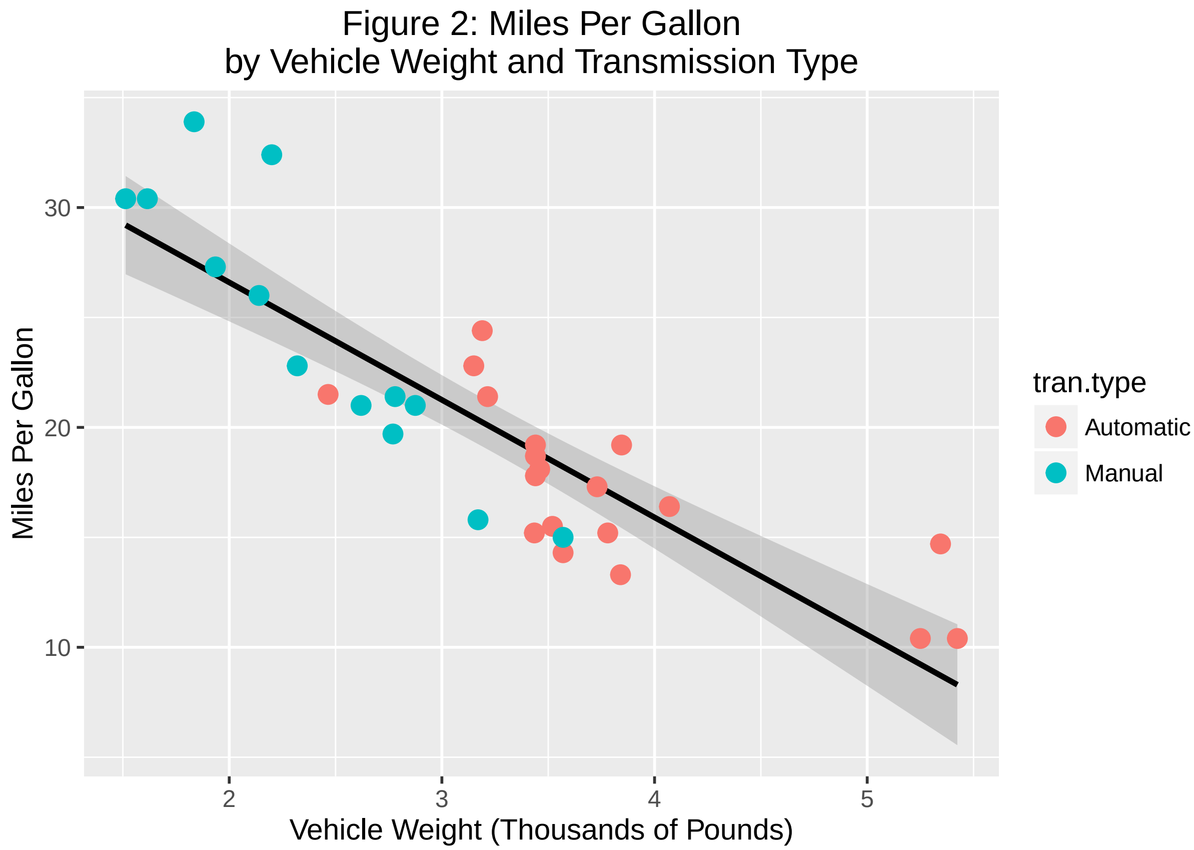


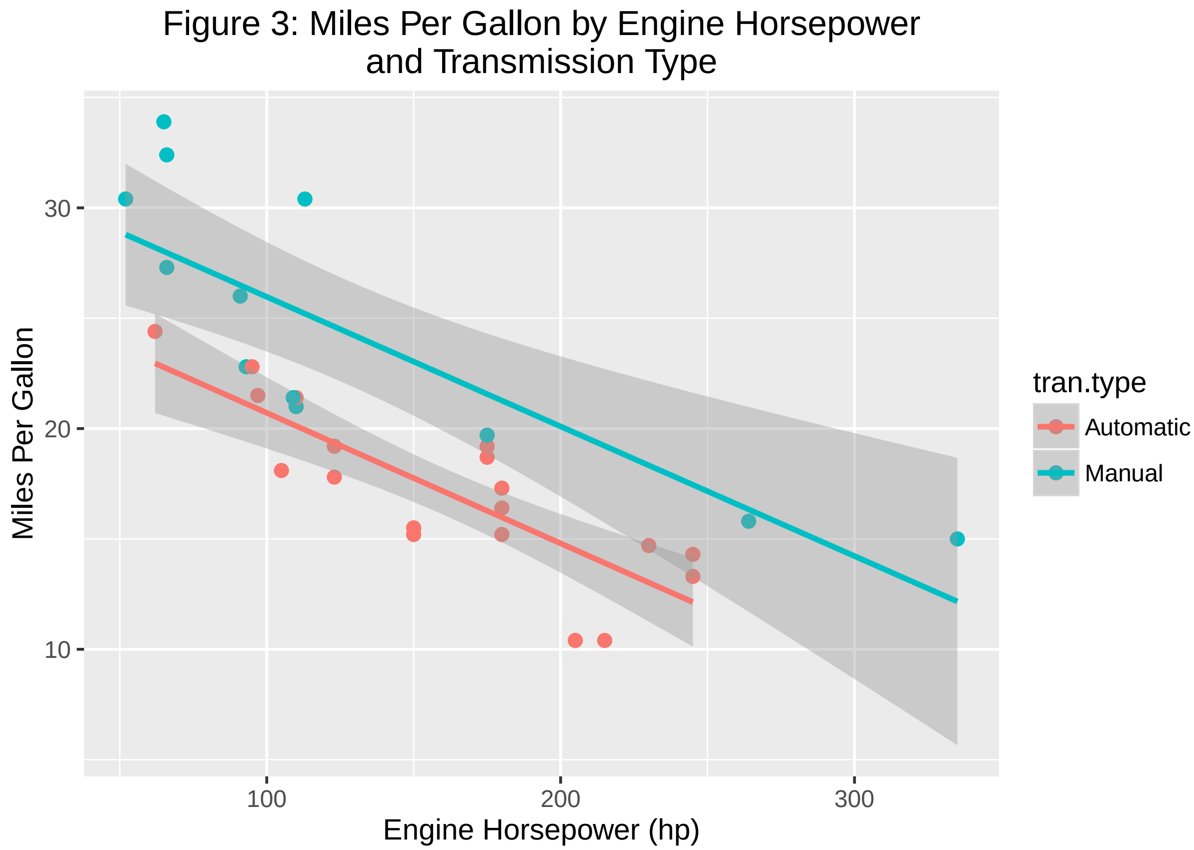
Figure 2 below shows cars' MPG as a function using vehicle weight as a possible predictor.



In a compelling fashion, Figure 2 shows a negative correlation between vehicle weight and gas mileage. The heavier the vehicle the lower the mileage.

Despite the strong negative correlation evident in Figure 2, the dispersion of automatic transmissions versus manual transmissions along the x-axis makes it difficult if not impossible to see the marginal impact of the transmission type. The difficulty is due to the fact that within the data set there are relatively few cases in which we have observations on both automatic and manual transmission cars at a given vehicle weight. Notice in Figure 2 that most of the manual transmission cars are at the lower end of the vehicle weight scale and conversely most of the automatic transmission cars are at the higher end of the weight scale. Using vehicle weight and transmission type as predictors of MPG, we won't be able to identify their isolated (marginal) impacts because they are so highly correlated.

Engine horsepower is a measure of how much power an engine is capable of producing. Since the power ultimately comes from buring gasoline, the higher power of an engine, the more gas it is likely to consume. Figure 3 shows the relationship between MPG and engine horsepower broken out (as in Figure 2) by transmission type. Here the negative correlation between MPG and horsepower is evident but perhaps not quite as strong as the relationship between MPG and vehicle weight (Figure 2).



Unlike in Figure 2, in Figure 3 the data shows a good deal of overlap between transmission type and engine horsepower. This is to say that at a given horsepower we have observations of both automatic- and manual-transmission cars' MPG. Using engine horsepower will therefore allow us to assess the marginal impact of the type of transmission.

Model Selection

Power transfers from a car's engine to its driving wheels through its transmission and rear axle gearing. To see how these factors affect gas mileage, we'll construct a three tiered nested model as follows:

1. Model 1: mpg ~ hp
2. Model 2: mpg ~ hp + tran.type
3. Model 3: mpg ~ hp + tran.type + gear + axle_ratio

and use Analysis of Variance (ANOVA) to see whether we get significantly better prediction of MPG from the more complicated models.

```r
m1=lm(mpg~hp, data=mtcars2) # Model 1
m2=update(m1, mpg~hp+tran.type) # Model 2
m3=update(m2, mpg~hp+tran.type+drat.size+factor(gear)) # Model 3
anova(m1, m2, m3)
```

```
## Analysis of Variance Table
##
```

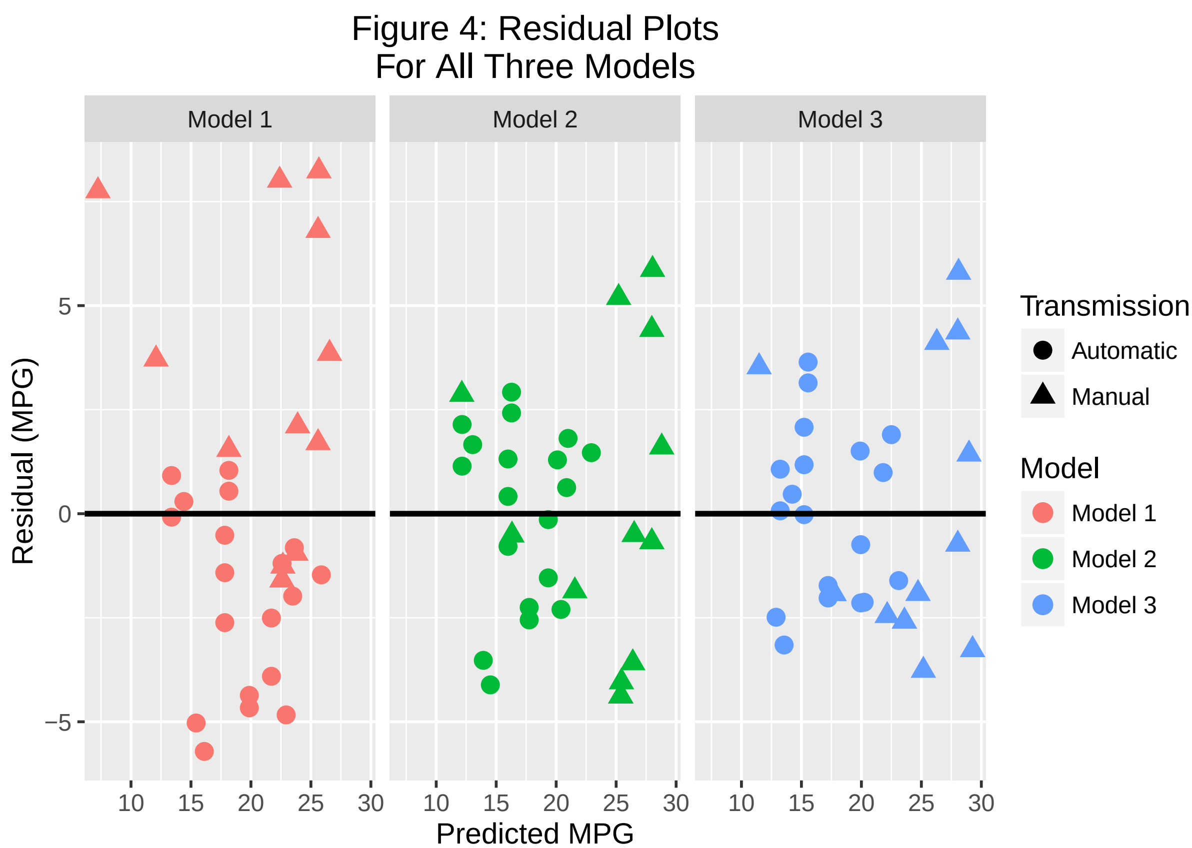
```
## Model 1: mpg ~ hp
## Model 2: mpg ~ hp + tran.type
## Model 3: mpg ~ hp + tran.type + drat.size + factor(gear)
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 447.67
## 2     29 245.44  1   202.235 24.2158 4.573e-05 ***
## 3     25 208.78  4    36.655  1.0973    0.3796
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The ANOVA output shows a big increase in explanatory power when going from Model 1 to Model 2. Since the Model 2 includes the transmission type as a predictor of MPG, the large F-statistic (and low p-value) supports the intuitive conclusion drawn from Figure 3: the transmission type does have a significant impact on mileage.

Model 3 includes the drive train gearing (transmission gears and rear axle ratio) and improves the fit of the model by statistically significant, but modest amount. Since Model 3 gives only a slight improvement in explanatory power over Model 2, we're going to exclude the transmission gears and axle ratio from the analysis. All subsequent analysis, therefore, will be based on Model 2.

Residual Analysis

Figure 4 below shows a residual plot for the three models under consideration. Notice that the residual variance decreases as we move from Model 1 to Model 2 to Model 3. Notice further that the residual variance appears to increase as the predicted MPG increases. This indicates heteroscedasticity (non-constant variance) among our residuals which violates one of the assumptions of the multivariate linear model. Heteroscedasticity is likely to be an indicator of a missing predictor or a nonlinear model, which is not surprising given the simplicity of the model.



Influence Analysis

Figure 5 shows the actual and fitted values for the two transmission types. The large shapes, the two large triangles for manual transmissions are observations that apply high leverage on the fitted lines, meaning these points moreso than the others, determine each line's slope and intercept. This is because their x-values, horsepower in this case, is far removed from the mean of the other observations' x-values.

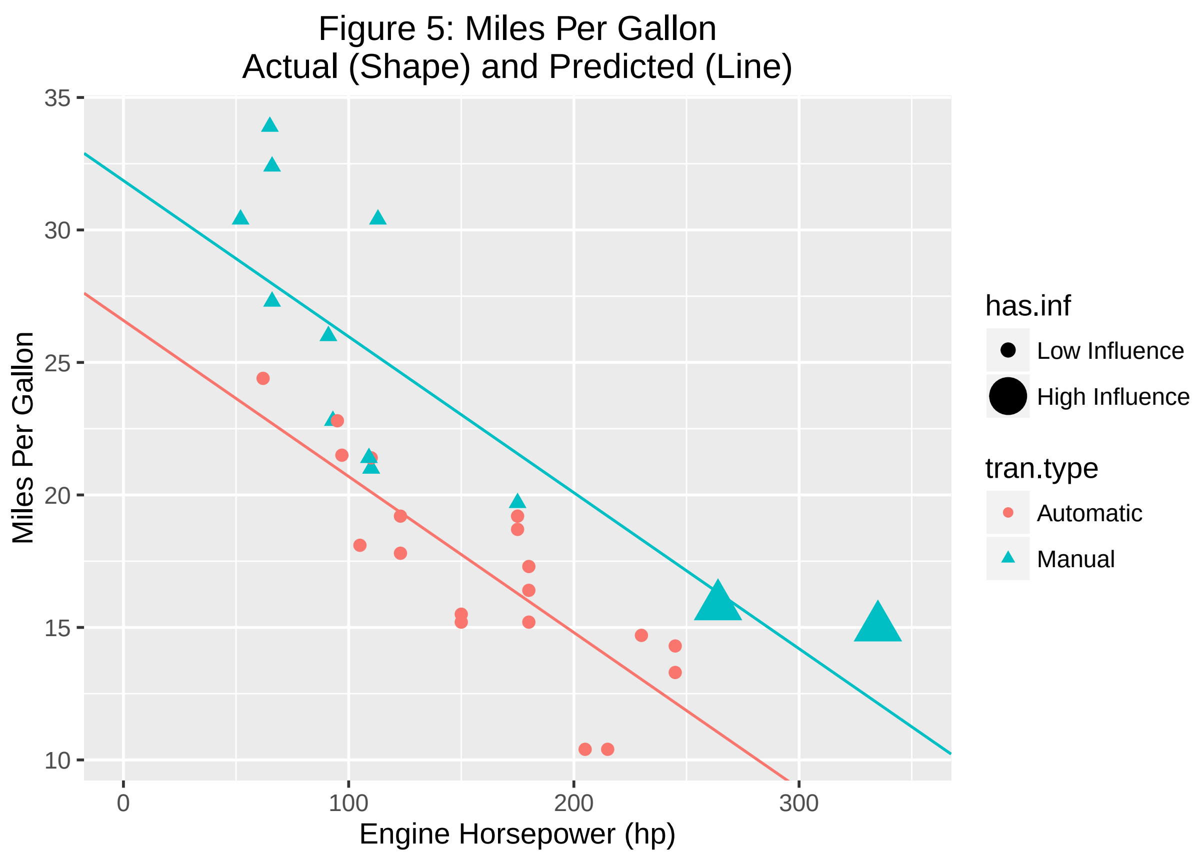
R's `influence.measures` function applies a battery of influence measure tests to a fitted linear model. If any of these tests produces a positive result, `influence.measures` considers the observation to be suspect. The following R code identifies the suspect observations that correspond to the large triangles in Figure 5.

```
im <- influence.measures(m2)
mtcars2[apply(im$is.inf,1,sum)>0,] #the rows for which any of the influence measures are true
```

```
##                  mpg cyl disp  hp drat   wt qsec vs am gear carb tran.type
## Ford Pantera L 15.8   8  351 264 4.22 3.17 14.5  0  1    5    4    Manual
## Maserati Bora  15.0   8  301 335 3.54 3.57 14.6  0  1    5    8    Manual
##                eng.type drat.size
## Ford Pantera L  V-Block      High
## Maserati Bora   V-Block    Medium
```

It is hard to imagine that a Maserati Bora producing 335 hp with 8 carburetors is going to get anywhere near 15.0 miles per gallon! The mileages of that car is probably closer to the 12.13 MPG that the model predicts.

```
## Warning: Using size for a discrete variable is not advised.
```



Conclusions

Below is the estimation summary of Model 2:

```r
summary(m2)$coef
```

```
##                    Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept)      26.5849137 1.425094292 18.654845 1.073954e-17
## hp               -0.0588878 0.007856745 -7.495191 2.920375e-08
## tran.typeManual   5.2770853 1.079540576  4.888270 3.460318e-05
```

From the large t-statistic 4.888 associated with the tran.typeManual variable and corresponding low p-value of 3.46031822026483e-05 we can **conclude that a car with a manual transmission gets different gas mileage (mpg) than a car with an automatic with a given engine horsepower.**

Interpretation of Regression Coefficients

Given the manner in which the transmission type in the `mtcars2` dataset is represented ("Automatic" versus "Manual") and the specification of the `m2` model, the "(Intercept)" row in the model summary above represents the estimation results for the fitted line for cars with automatic transmissions. The "Estimate" column is the y-intercept for the fitted line.

The "tran.typeManual" row in the summary output contains the estimation results for cars with manual transmissions. In this case however, the "Estimate" value represents the *difference* in the y-intercept of the fitted line for manual transmission cars from the y-intercept of the fitted line for automatic transmission cars. Since the "Estimate" value is positive, it represents an increase in miles per gallon. This positive difference is evident from the y-intercept values of the two fitted lines in Figure 5 above.

The model predicts an increase of 5.277 miles per gallon for a manual transmission car over an automatic transmission car with the same engine horsepower.

Uncertainty of Conclusions

The model produces a point estimate of an increase of 5.277 miles per gallon for cars with manual transmissions. The following R-code produces a 95% confidence interval for this estimated difference:

```
coefs = summary(m2)$coefficients
ci = coefs["tran.typeManual","Estimate"]+
        c(-1,1)*qt(0.975,df=m2$df)*coefs["tran.typeManual","Std. Error"]
ci
```

```
## [1] 3.069177 7.484994
```

With 95% confidence, we estimate that for two cars with the same engine horsepower, the car with a manual transmission will get between 3.069 and 7.485 more miles per gallon than the car with an automatic transmission.

Appendices

Appendix A1: Reader-friendly transformation of the mtcars data.frame

This R code adds columns that map the transmission type in the source data to the terms "Automatic" and "Manual". It also interprets the vs column to be the engine type: V-block or Inline.

```
data("mtcars")

mtcars2=mtcars

#0 means automatic, 1 means Manual
mtcars2$tran.type=factor(ifelse(mtcars$am==0,"Automatic","Manual"))

#0 means V, 1 means straight
mtcars2$eng.type = factor(ifelse(mtcars$vs == 1,"Inline","V-Block"))

#slice the rear axle ratio into 3 equal buckets
q=quantile(mtcars$drat,c(0.33,0.67))
mtcars2$drat.size=cut(mtcars2$drat,c(0,q,10),labels=c("Low","Medium","High"))
```

Appendix A2: R-code for Figure 1

```
library(ggplot2)
g=ggplot(data=mtcars2, aes(x=tran.type, y=mpg, fill=tran.type))+
        geom_boxplot() +
```

```
        labs(title="Figure 1: Miles Per Gallon\nby Transmission Type",
             y="Miles Per Gallon", x = "Transmission Type")+
        theme(legend.position = "none")
g
```

Appendix A3: R-code for Figure 2

```
library(ggplot2)

g=ggplot( data=mtcars2, aes(x=wt, y=mpg)) +
        stat_smooth(method="lm", color="Black") +
        geom_point(size=3, aes(col=tran.type)) +

        labs(title="Figure 2: Miles Per Gallon\nby Vehicle Weight and Transmission Type",
             y="Miles Per Gallon", x = "Vehicle Weight (Thousands of Pounds)") +
        scale_fill_discrete(name="Transmission Type")
g
```

Appendix A4: R-code for Figure 3

```
library(ggplot2)

g=ggplot( data=mtcars2, aes(x=hp, y=mpg, col=tran.type)) +
        geom_point(size=2) +
        stat_smooth(method="lm", size=1) +
        labs(title="Figure 3: Miles Per Gallon by Engine Horsepower\nand Transmission Type",
             y="Miles Per Gallon", x = "Engine Horsepower (hp)") +
        scale_fill_discrete(name="Transmission Type")
g
```

Appendix A5: R-code for Figure 4

```
library(ggplot2)

n=nrow(mtcars2)
df <- data.frame( predict=c(predict(m1), predict(m2), predict(m3)),
                  resid  =c(resid(m1),   resid(m2),   resid(m3)),
                  Model  =c(rep("Model 1", n), rep("Model 2",n), rep("Model 3",n)),
                  Transmission = rep(mtcars2$tran.type, 3))
g=ggplot( data=df, aes(x=predict, y=resid, col=Model)) +
        facet_wrap(~Model) +
        geom_point(size=3, aes( shape=Transmission)) +
        geom_hline(yintercept=0, size=1, col="Black")+

        labs(title="Figure 4: Residual Plots\nFor All Three Models",
             y="Residual (MPG)", x = "Predicted MPG" )
g
```

Appendix A2: R-code for Figure 5

```
library(ggplot2)
im=influence.measures(m2)
coefs=m2$coefficients
mt2=mtcars2
mt2$has.inf=factor(ifelse(apply(im$is.inf,1,sum)>0,"High Influence","Low Influence"),
                   levels=c("Low Influence", "High Influence"))

mycolors = scales::hue_pal()(4)
g=ggplot( data=mt2, aes(x=hp, y=mpg, col=tran.type)) +
        geom_point( aes(size=has.inf,shape=tran.type)) +
        geom_abline(intercept = coefs[1], slope=coefs[2], color=mycolors[1]) +
        geom_abline(intercept = coefs[1]+ coefs[3], slope=coefs[2], color=mycolors[3])+
        xlim(0,350)+
        labs(title="Figure 5: Miles Per Gallon\nActual (Shape) and Predicted (Line)",
             y="Miles Per Gallon", x = "Engine Horsepower (hp)") +
        scale_fill_discrete(name="Transmission Type")
g
```