# MPG Analysis: Regression Models Course Project

*Travis Fell*

*August 14, 2015*

# Executive Summary

This analysis strives to answer the questions from data in the mtcars data set: 1. Is an automatic or manual transmission better for MPG? Technically speaking manual is better for MPG, but when accounting for factors that more strongly correlate with MPG, the difference between manual and automatic transmissions is negligible. 2. What is the MPG difference between automatic and manual transmissions? When accounting for weight and displacement, the MPG range of improvement for manual transmissions is fit lwr upr 1 0.1777241 0.1816018 0.17384

# I. Loading and Exploring the Data

First, we'll start by loading up the data and packages we need and doing some quick exploratory analysis.

```
library(ggplot2)
library(GGally)
```

```
## Warning: package 'GGally' was built under R version 3.2.1
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:GGally':
##
##     nasa
##
## The following objects are masked from 'package:stats':
##
##     filter, lag
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
data(mtcars)
head(mtcars)
```

```
##                   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
```

```
summary(mtcars)
```

```
##       mpg             cyl             disp             hp
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0
##       drat             wt             qsec             vs
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000
##        am              gear            carb
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000
##  Median :0.0000   Median :4.000   Median :2.000
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

```
summarize(group_by(mtcars, am), mean(mpg), records = length(mpg))
```

```
## Source: local data frame [2 x 3]
##
##   am mean(mpg) records
## 1  0  17.14737      19
## 2  1  24.39231      13
```

Note that the am column indicates the transmission type of the care; 1 = manual, 0 = automatic. At first blush, the manual transmissions appear to have the edge in MPG, 24.4 to 17.1 for automatics. However, there are other variables that could also have an impact on MPG that we will examine in the following

section.

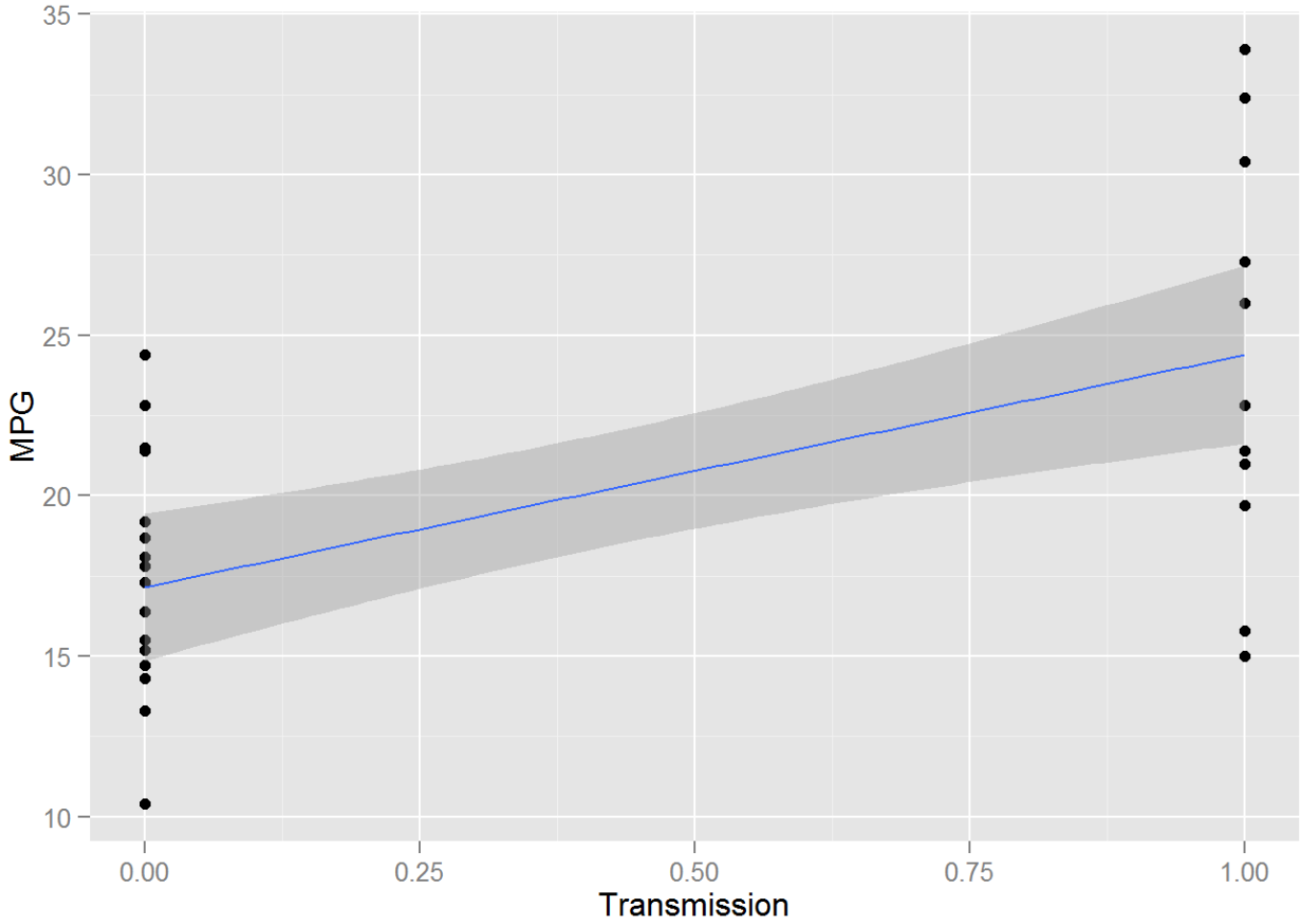
# II. Isolate Transmission as a Factor of MPG

As a baseline, we will create a single linear regression between transmission type and MPG.

```
mpgtrans <- lm(mpg ~ am, mtcars)
summary(mpgtrans)$coefficients
```

```
##             Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## am           7.244939   1.764422  4.106127 2.850207e-04
```

```
g <- ggplot(mtcars, aes(mtcars$am, mtcars$mpg))
g <- g + geom_point()
g <- g + geom_smooth(method = 'lm')
g <- g + ylab("MPG")
g <- g + xlab("Transmission")
g
```




When taken on its own transmission type has a slope coefficient of 7.2 from automatic to manual (that is, an increase in 7.2 mpg for manual transmission vehicles). However, there are several other factors that could impact MPG. We'll compare transmission type to all variables to begin understanding how they impact transmission type as a predictor of MPG.

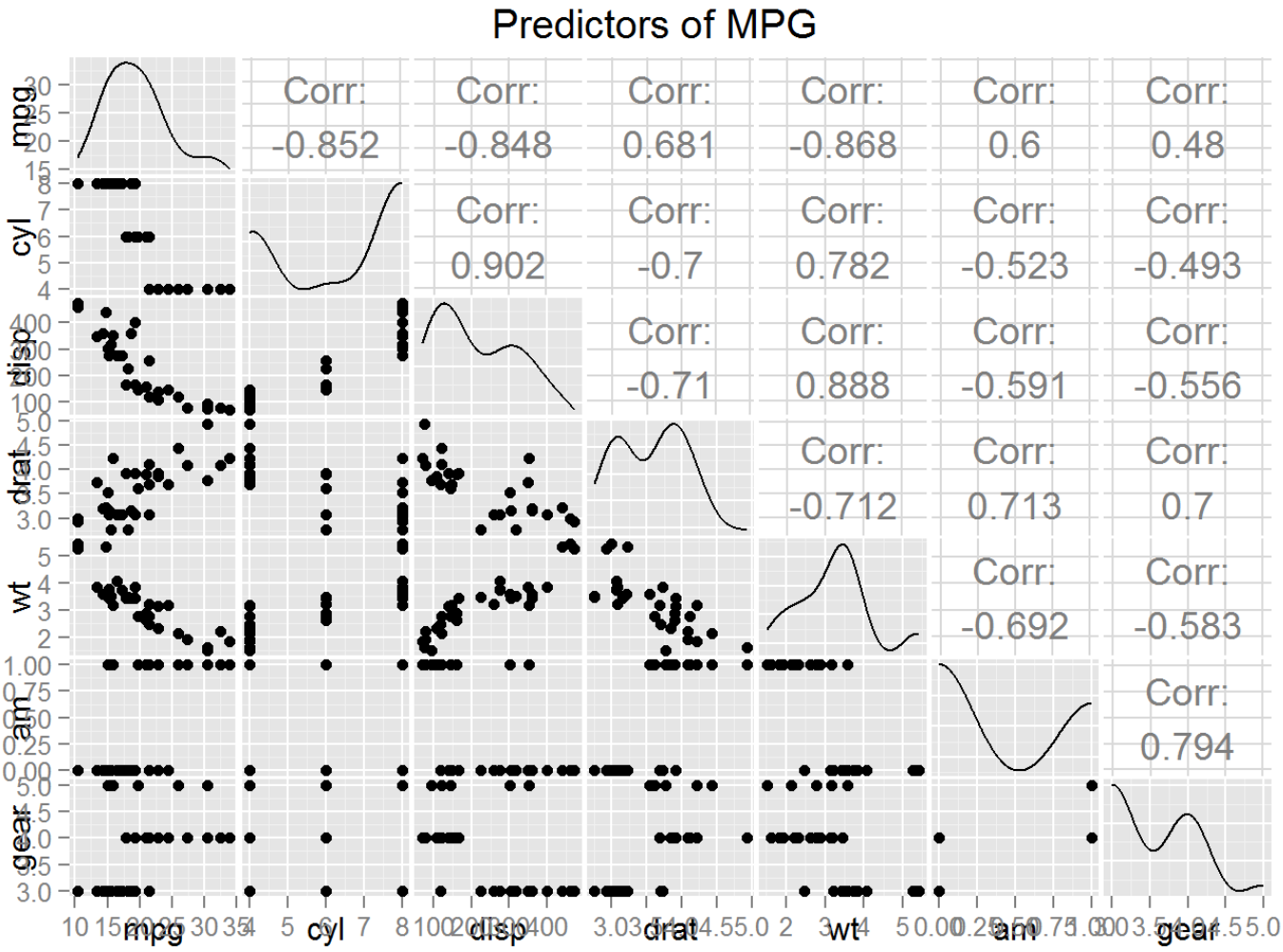
```
mpgall <- lm(mpg ~., mtcars)
summary(mpgall)$coefficients
```

```
##               Estimate  Std. Error    t value   Pr(>|t|)
## (Intercept) 12.30337416 18.71788443  0.6573058 0.51812440
## cyl         -0.11144048  1.04502336 -0.1066392 0.91608738
## disp         0.01333524  0.01785750  0.7467585 0.46348865
## hp          -0.02148212  0.02176858 -0.9868407 0.33495531
## drat         0.78711097  1.63537307  0.4813036 0.63527790
## wt          -3.71530393  1.89441430 -1.9611887 0.06325215
## qsec         0.82104075  0.73084480  1.1234133 0.27394127
## vs           0.31776281  2.10450861  0.1509915 0.88142347
## am           2.52022689  2.05665055  1.2254035 0.23398971
## gear         0.65541302  1.49325996  0.4389142 0.66520643
## carb        -0.19941925  0.82875250 -0.2406258 0.81217871
```

The first thing that sticks out when creating a model with all variables is that the slope coefficient for transmission decreases significantly to 2.52. Thus, other variables in the model have an impact on transmission type as a predictor of MPG. But which ones have the biggest impact and how can they help us understand the true impact of transmission type on MPG?

We'll examine MPG as a function of number of cylindars, displacement, real axle ratio, weight, transmission and forward gears. These factors should have a functional impact on MPG and, thus, the most relevant predictors of MPG. Once we have a feel for the factors with the strongest correlation with MPG, we'll re-run our regression with the appropriate factors.

```
p = ggpairs(mtcars, columns = c(1, 2, 3, 5, 6, 9, 10), title = "Predictors of MPG")
p
```




The pairs plot shows that transmission type has a reasonably strong positive correlation with MPG. However, the variables of cylindars, displacement and weight have very strong negative correlations with MPG. Also, since displacement is a measure of total gasoline an engine would use in a single drive cycle, we'll use that variable as a measure instead of cylindars. So will will re-run our regression model with weight and displacement.

```
mpgtranswt <- lm(mpg ~ am + wt, mtcars)
summary(mpgtranswt)$coefficients
```

```
##                Estimate Std. Error     t value     Pr(>|t|)
## (Intercept) 37.32155131  3.0546385 12.21799285 5.843477e-13
## am          -0.02361522  1.5456453 -0.01527855 9.879146e-01
## wt          -5.35281145  0.7882438 -6.79080719 1.867415e-07
```

```
mpgtransds <- lm(mpg ~ am + disp, mtcars)
summary(mpgtransds)$coefficients
```

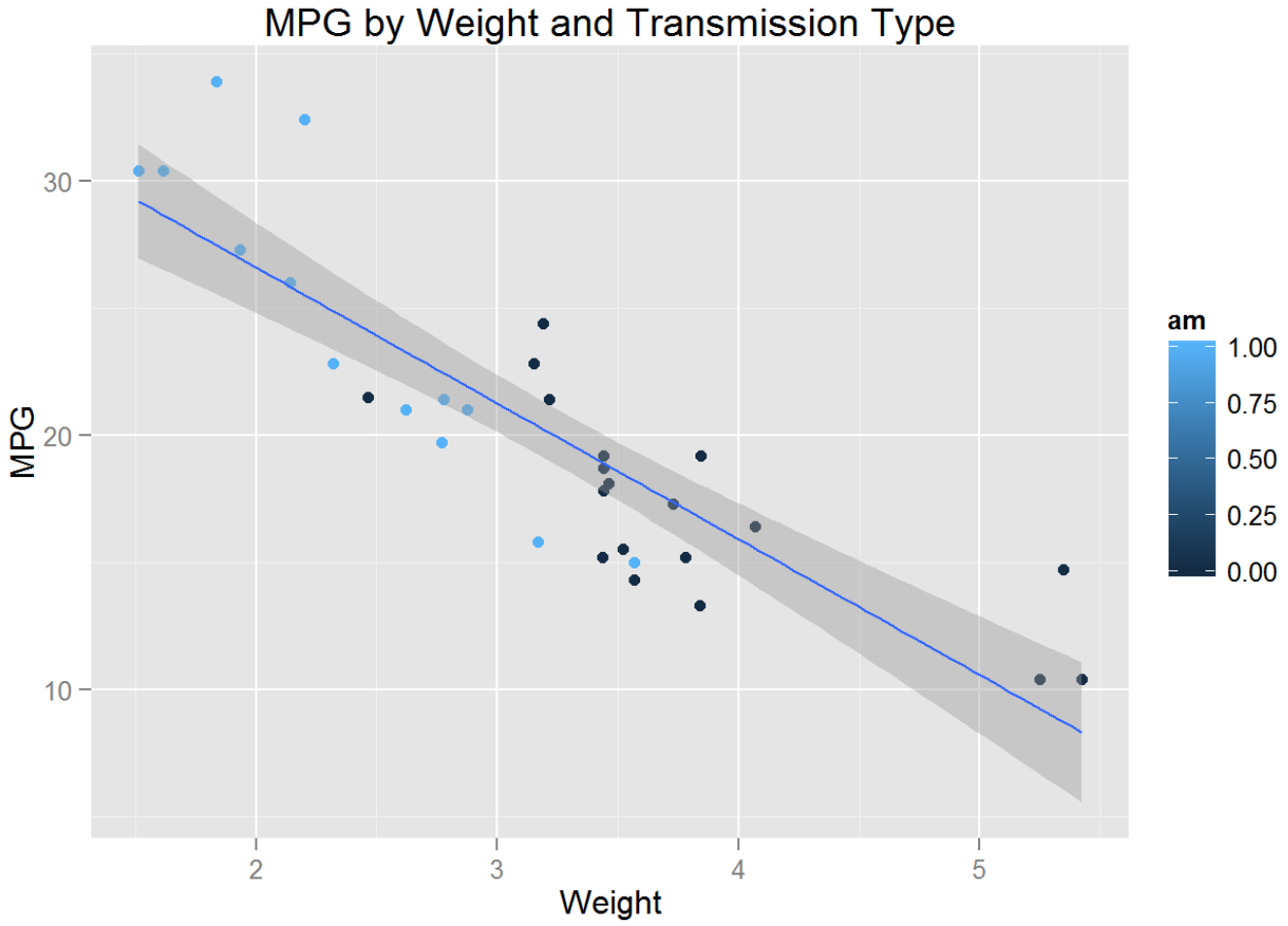
```
##                Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept) 27.84808111 1.834071377 15.183750 2.452658e-15
## am           1.83345825 1.436099585  1.276693 2.118396e-01
## disp        -0.03685086 0.005781896 -6.373490 5.747528e-07
```

```
mpgtranswtds <- lm(mpg ~ am + wt + disp, mtcars)
summary(mpgtranswtds)$coefficients
```

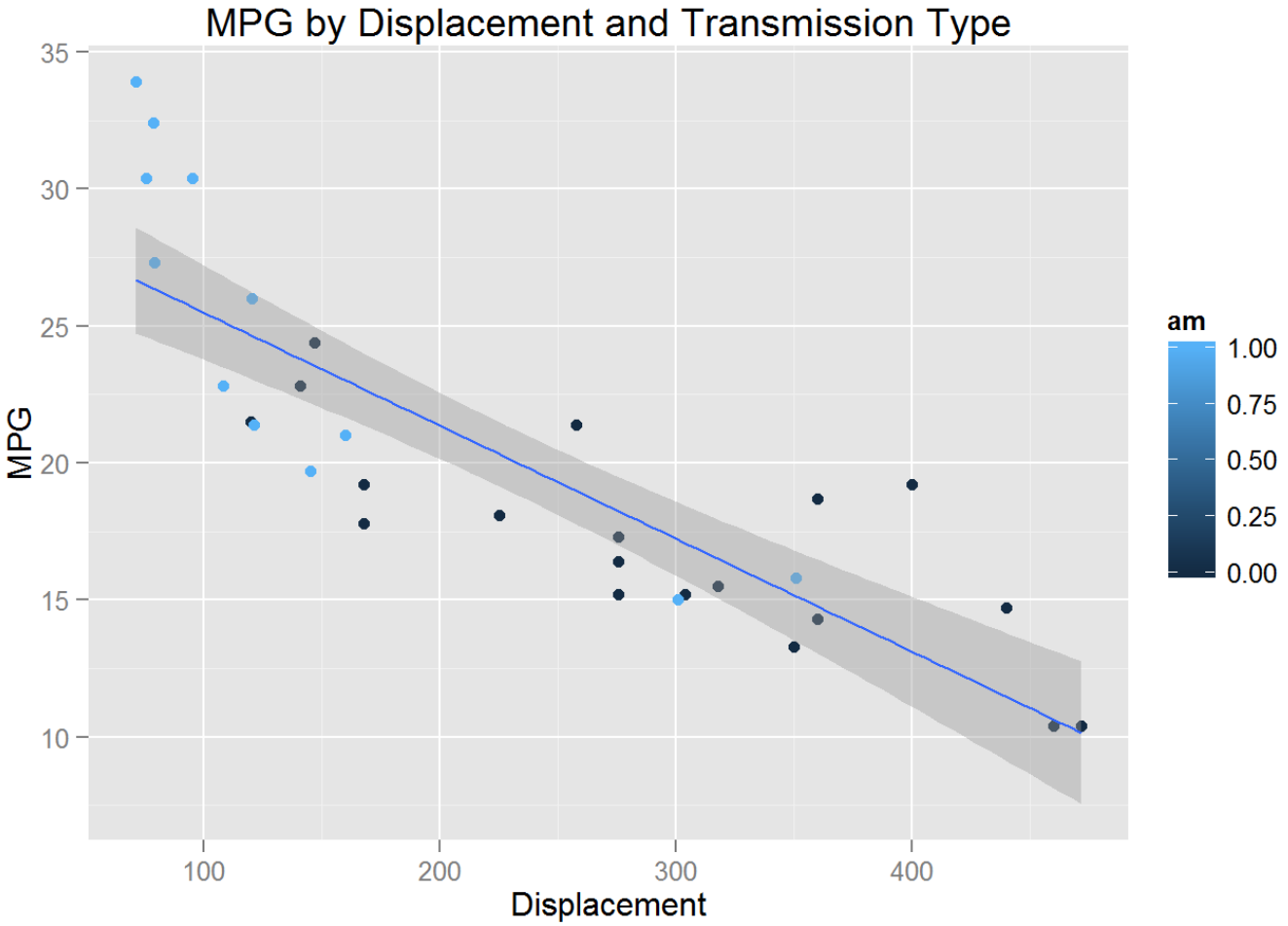
```
##                Estimate Std. Error    t value     Pr(>|t|)
## (Intercept) 34.67591088 3.24060891 10.7004306 2.115200e-11
## am           0.17772414 1.48431586  0.1197347 9.055483e-01
## wt          -3.27904388 1.32750927 -2.4700723 1.986658e-02
## disp        -0.01780491 0.00937465 -1.8992613 6.787740e-02
```

When using weight as a second regressor, the slope of AM changes dramatically, from 7.2 to -.02. When including displacement in the model, there is a similar reduction to .17. The fact that relationship between MPG and transmission type can be so significantly impacted by other variables should give us pause when evaluating MPG performance purely as a function of transmission type. Rather, it would probably make more sense to take weight, displacement and transmission type into account when evaluating MPG performance.

```
gwt <- ggplot(mtcars, aes(wt, mpg), colour = am)
gwt <- gwt + geom_point(aes(colour = am))
gwt <- gwt + geom_smooth(method = 'lm')
gwt <- gwt + ylab("MPG")
gwt <- gwt + xlab("Weight")
gwt <- gwt + ggtitle("MPG by Weight and Transmission Type")
gwt
```




```
gds <- ggplot(mtcars, aes(disp, mpg), colour = am)
gds <- gds + geom_point(aes(colour = am))
gds <- gds + geom_smooth(method = 'lm')
gds <- gds + ylab("MPG")
gds <- gds + xlab("Displacement")
gds <- gds + ggtitle("MPG by Displacement and Transmission Type")
gds
```

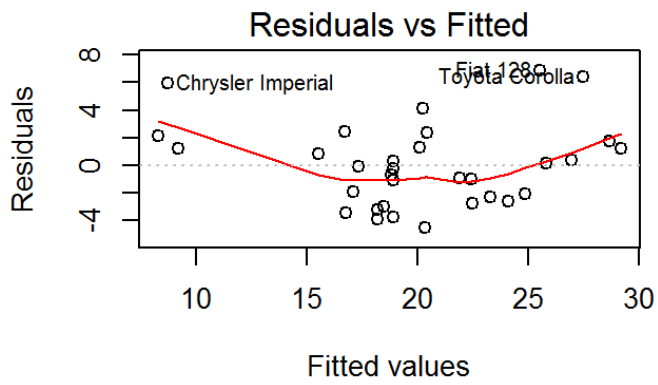


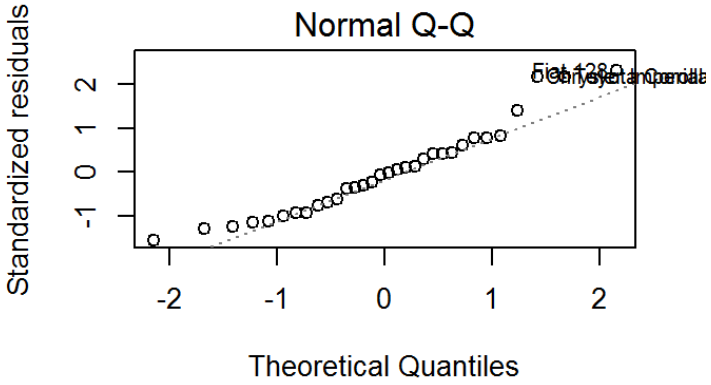

In the appendix appear regression plots of MPG by Weight and MPG by Displacement, both of which distinguish manual vs automatic transmission data points. These plots illustrate one of the conclusions of this analysis: to answer the question “Is an automatic or manual transmission better for MPG?”, one should consider transmission type along with the most impactful predictors of MPG: Weight and Displacement.

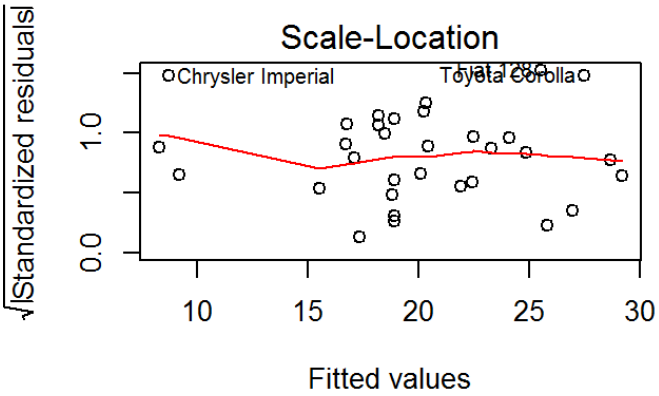
# A Quick Note on Residuals

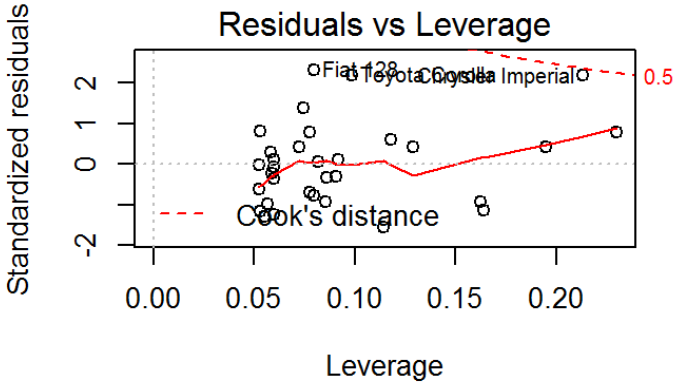
The residuals plots for the mpgtranswt model and the mpgtransds models do not seem to indicate any patterns in the data, unusual results or any other issues that would indicate some poor aspect of model fit.

```
#basic residual testing
par(mfrow = c(2,2))
plot(mpgtranswt)
```

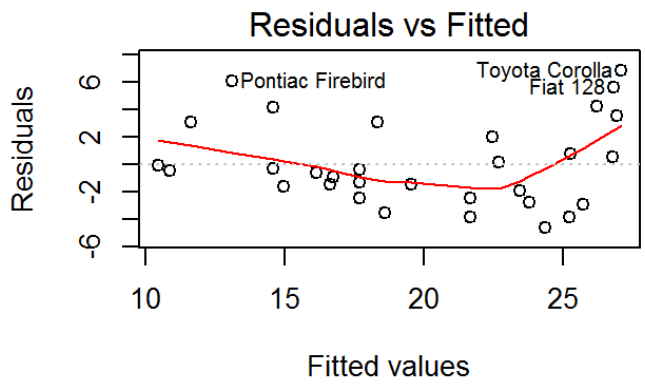


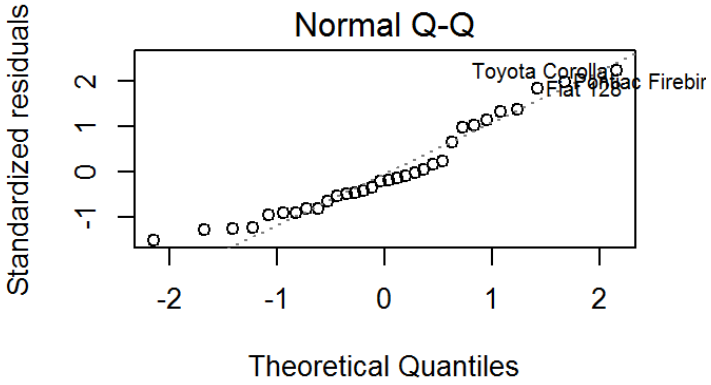



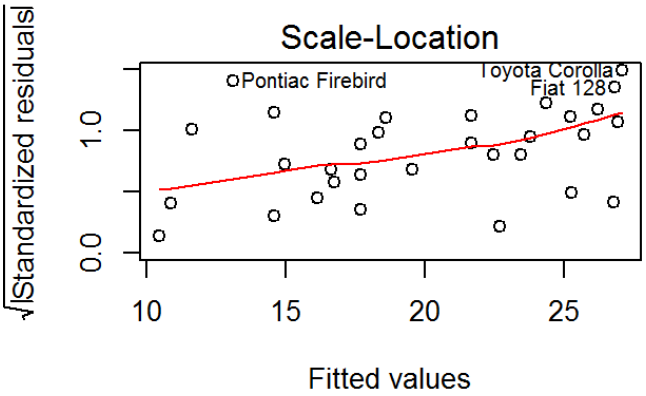



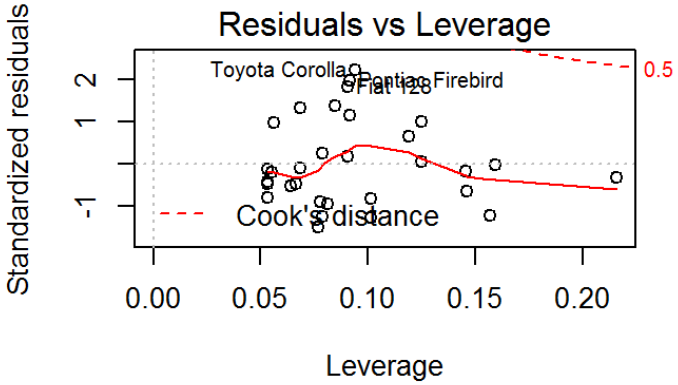




```
plot(mpgtransds)
```

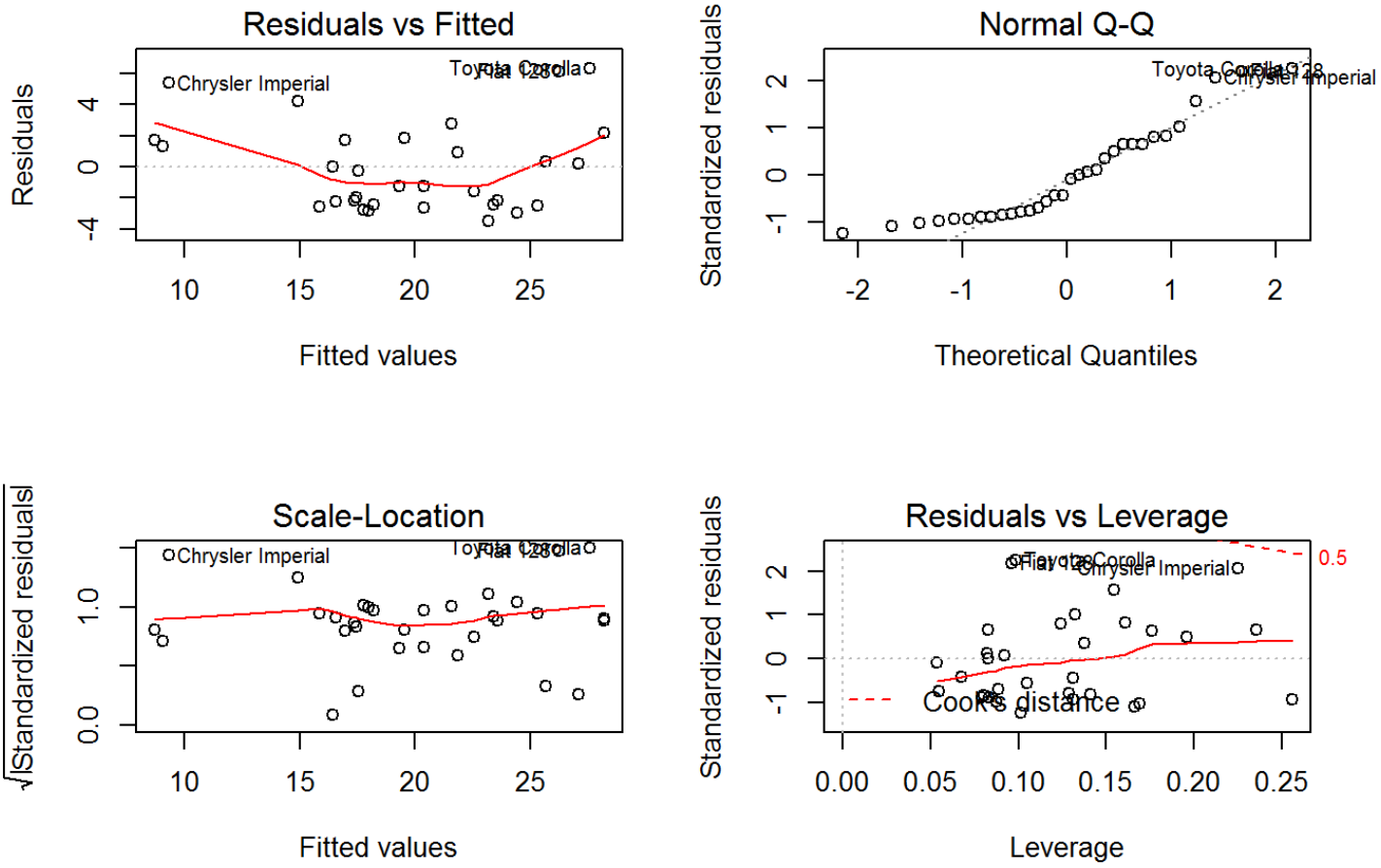


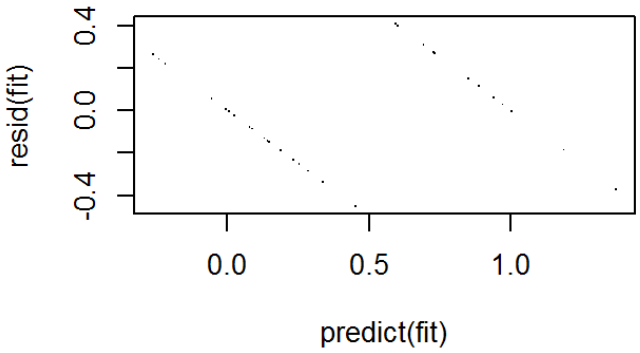




```
plot(mpgtranswtds)
```




```
# residuals vs fitted values
fit <- lm(am ~ . -1, mtcars)
plot(predict(fit), resid(fit), pch = '.')
```




#Predicting the MPG To illustrate the impact of transmission type on MPG, we will predict the MPG holding weight constant at 3 tones and displacement constant at 250 while changing transmission type.

```
autompg <- predict(mpgtranswtds, data.frame(am = 0,  wt = 3, disp = 250), interval = "pre
diction")
manualmpg <- predict(mpgtranswtds, data.frame(am = 1,  wt = 3, disp = 250), interval = "p
rediction")
mpgdelta <- manualmpg - autompg
confint(mpgtranswtds)
```

```
##                   2.5 %      97.5 %
## (Intercept) 28.03782444 41.31399731
## am          -2.86275907  3.21820735
## wt          -5.99832334 -0.55976441
## disp        -0.03700801  0.00139819
```

```
colnames(manualmpg)
```

```
## [1] "fit" "lwr" "upr"
```

```
manualmpg[1:3]
```

```
## [1] 20.56528 14.15055 26.98000
```

Using the R predict function, we find that automatic transmission MPG ranges are 20.3875516, 13.9689453, 26.8061578. By comparison, manual transmission MPG ranges are fit, lwr, upr 20.5652757, 14.1505471, 26.9800043 . While manual is slightly better, we can conclude from this data that there is not much change when weight and displacement are taken into account, the difference being 0.1777241, 0.1816018, 0.1738465.