Regression Coursera

Summary This project was done to explore the relationship between mileage in cars and transmission type (automatic vs manual)– A simple model only looking at tramsission type finds a 7.25 MPG difference between the types of transmission (p < .05)

R-squared for transission type only yields a low R-squared (.36). Further analysis shows that there are other more significant contributors to MPG such as the weight of the car. A more complete model (including weight, hp, cyl) changes the p-value of transmission type to no longer being significant (p=.3) but improves R-squared (.85) with only weight being significant at the .05 level.

```
fullmodel<-lm(mpg~ . , data=mtcars)
summary(fullmodel)
```

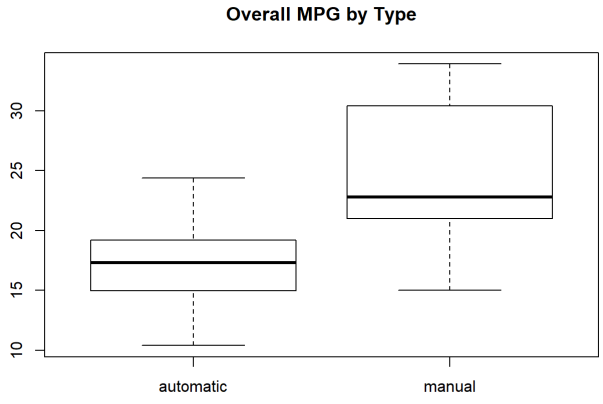
```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788   0.657   0.5181
## cyl         -0.11144    1.04502  -0.107   0.9161
## disp         0.01334    0.01786   0.747   0.4635
## hp          -0.02148    0.02177  -0.987   0.3350
## drat         0.78711    1.63537   0.481   0.6353
## wt          -3.71530    1.89441  -1.961   0.0633 .
## qsec         0.82104    0.73084   1.123   0.2739
## vs           0.31776    2.10451   0.151   0.8814
## am           2.52023    2.05665   1.225   0.2340
## gear         0.65541    1.49326   0.439   0.6652
## carb        -0.19942    0.82875  -0.241   0.8122
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869,  Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
```

```
ammod<-lm(mpg ~ am, data=mtcars)
summary(ammod)
```

```
##
## 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
```

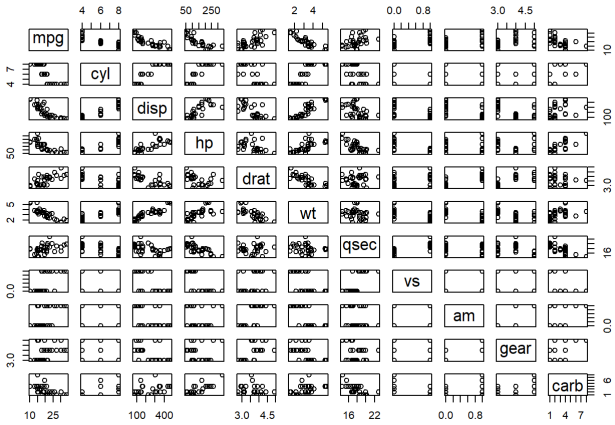
R-squared is fairly low (.36), however, transission type is significant

```
boxplot(mpg~am, data=mtcars, main= "Overall MPG by Type", names=c("automatic", "manual"))
```



at face value– manual appears to have better MPG– however, the average for automatic is contained within the error bars for automatic

```
pairs(mpg ~ ., data = mtcars)
```




```
mvr <- lm(mpg~am + wt + hp + cyl, data = mtcars)
summary(mvr)
```

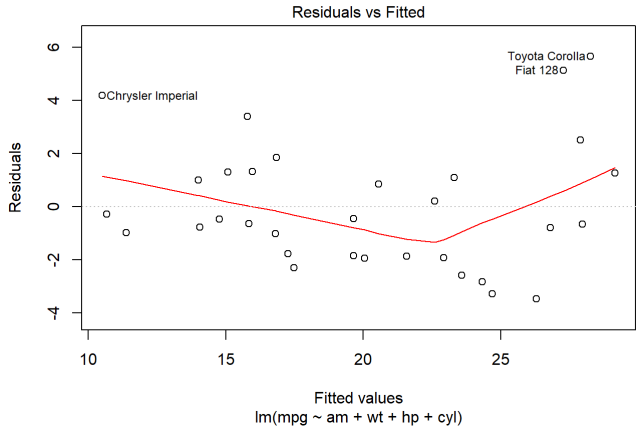
```
##
## Call:
## lm(formula = mpg ~ am + wt + hp + cyl, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4765 -1.8471 -0.5544  1.2758  5.6608
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 36.14654    3.10478  11.642 4.94e-12 ***
## am           1.47805    1.44115   1.026   0.3142
## wt          -2.60648    0.91984  -2.834   0.0086 **
## hp          -0.02495    0.01365  -1.828   0.0786 .
## cyl         -0.74516    0.58279  -1.279   0.2119
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.509 on 27 degrees of freedom
## Multiple R-squared:  0.849,  Adjusted R-squared:  0.8267
## F-statistic: 37.96 on 4 and 27 DF,  p-value: 1.025e-10
```

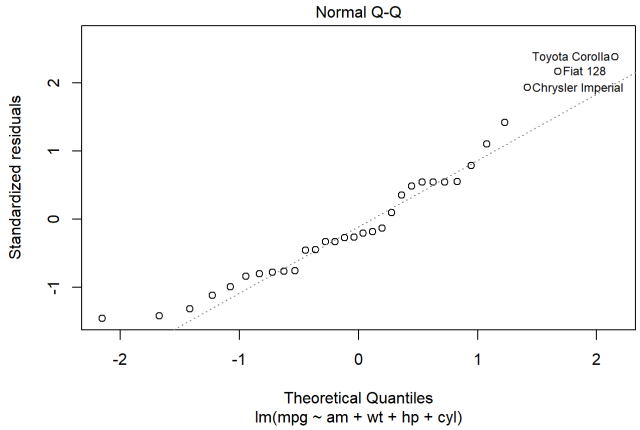
```
anova(ammod,mvr)
```

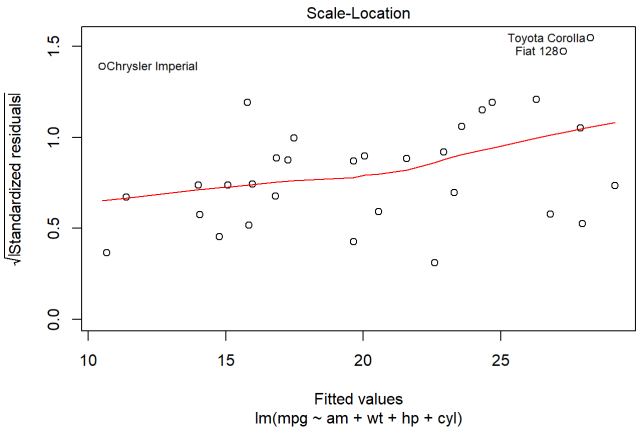
```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + wt + hp + cyl
##   Res.Df   RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.9
## 2     27 170.0  3     550.9 29.166 1.274e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

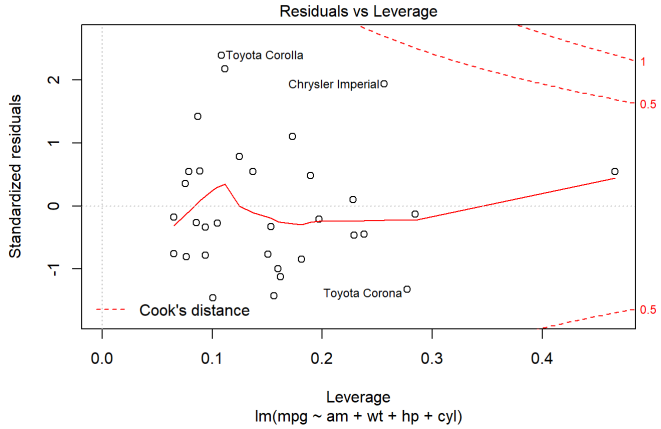
A more Inclusive Model reduces the residual sum of squares significantly

```
plot(mvr)
```

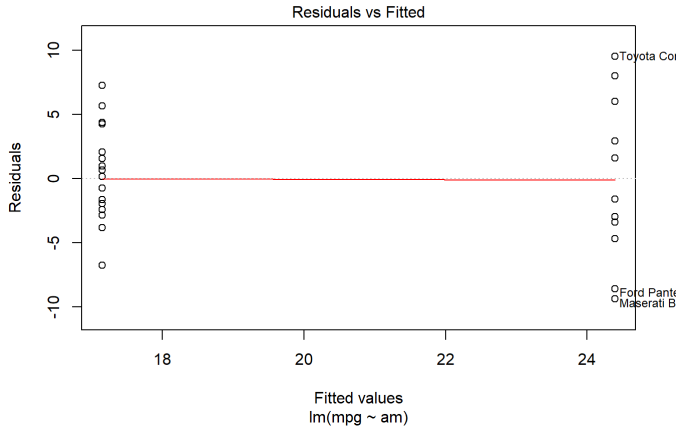

Residuals vs Fitted
Toyota Corolla
Fiat 128
Chrysler Imperial
Residuals
Fitted values
lm(mpg ~ am + wt + hp + cyl)

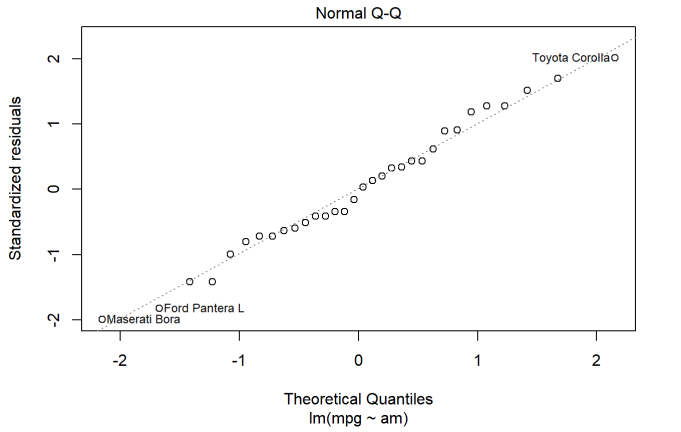

Normal Q-Q
Toyota Corolla
Fiat 128
Chrysler Imperial
Standardized residuals
Theoretical Quantiles
lm(mpg ~ am + wt + hp + cyl)

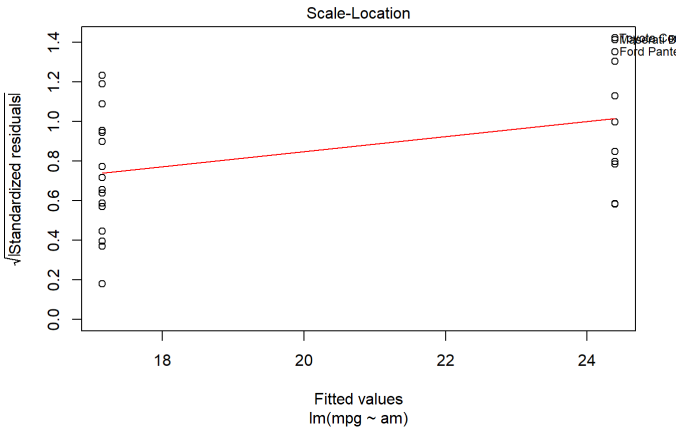

Scale-Location
Toyota Corolla
Fiat 128
Chrysler Imperial
√|Standardized residuals|
Fitted values
lm(mpg ~ am + wt + hp + cyl)

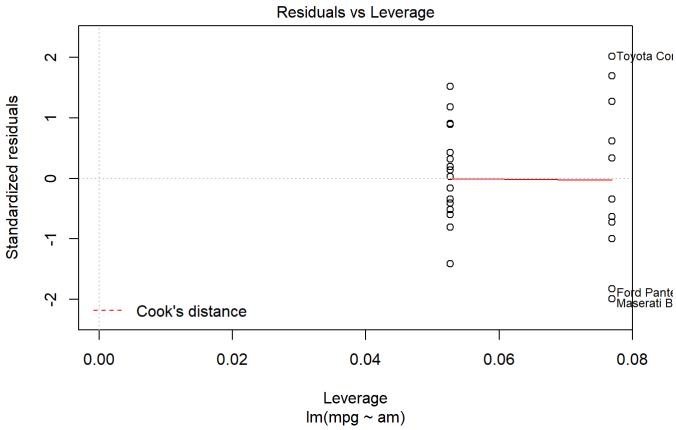

Residuals vs Leverage
Toyota Corolla
Chrysler Imperial
Toyota Corona
Cook's distance
Standardized residuals
Leverage
lm(mpg ~ am + wt + hp + cyl)

```
plot(ammod)
```


Residuals vs Fitted
Residuals
Fitted values
lm(mpg ~ am)
Toyota Co
Ford Pante
Maserati B


Normal Q-Q
Standardized residuals
Theoretical Quantiles
lm(mpg ~ am)
Toyota Corolla
Ford Pantera L
Maserati Bora


Scale-Location
√|Standardized residuals|
Fitted values
lm(mpg ~ am)


Residuals vs Leverage
Standardized residuals
2
1
0
-1
-2
Toyota Co
Ford Pante
Maserati B
Cook's distance
0.00
0.02
0.04
0.06
0.08
Leverage
lm(mpg ~ am)