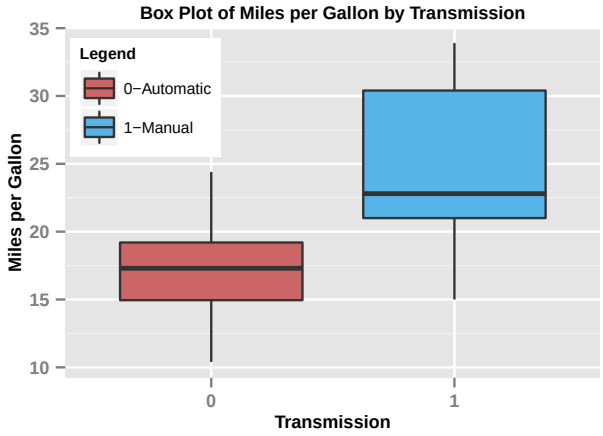
Differences in Miles per Gallon for Automatic and Manual Transmissions

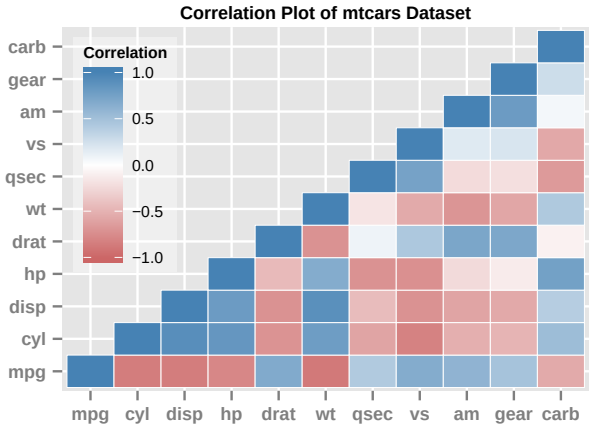
EXECUTIVE SUMMARY

This document explores the relationship between different variables and miles per gallon (mpg) with a focus on the difference on mpg for manual and automatic transmissions. Using statistical analysis and domain research into the factors contributing to fuel efficiency (mpg), it was found that the relationship between mpg and transmission is not significant. The final model constructed consists of the independent variables weight (wt) and cylinders (cyl). Residuals were plotted to check for the key assumptions of a linear model.

EXPLORATORY DATA ANALYSIS

From visual inspection of the box plot, it seems that Manual leads to a higher fuel efficiency. However, we would need to further investigate the relationship between the various variables, transmission and mpg to arrive at a conclusion. We also see that Displacement (disp), Horsepower (hp), Weight (wt) and Cylinder (cyl) are negatively correlated to mpg whereas Rear Axle Ratio (drat), Engine (vs) and Transmission (am) is somewhat positively correlated to mpg from the Correlation plot. The correlation matrix can be found in the appendix.





MODEL SELECTION

The strategy to model selection would be to consider the correlation between the different variables as well as researching into what may impact mpg and make use of the domain knowledge gained from the research.

Model 1 - Model 1 will consists of the variables listed below. At 0.05 significance level, only hp and wt are significant. Refer to the appendix (Figure 2) for the individual p values.

```
model1 <- lm(mpg ~ disp + hp + wt + cyl + drat + vs + am, data=mtcars)
```

Model 2 - By research automobiles terms (link), we note that cyl and disp are important factors contributing to mpg. In addition, wt is also related (link) to fuel econony (mpg). There was no mention about hp being directly related in both links. Hence, we will remove hp in our second model. To avoid multicollinearity (link) in the model, we will use cyl in the model. am is included to explore the relationship of am to mpg. As such,

model 2 consist of the below variables. We note that at significance level 0.05, am1 is not significant to the model. Refer to the appendix (Figure 3) for the individual p values.

```
model2 <- lm(mpg ~ wt + cyl + am, data=mtcars)
```

Model 3 - Further research indicates that am may not be directly related to fuel economy (link). Hence, we will remove am and construct model 3. Refer to the appendix (Figure 4) for the individual p values.

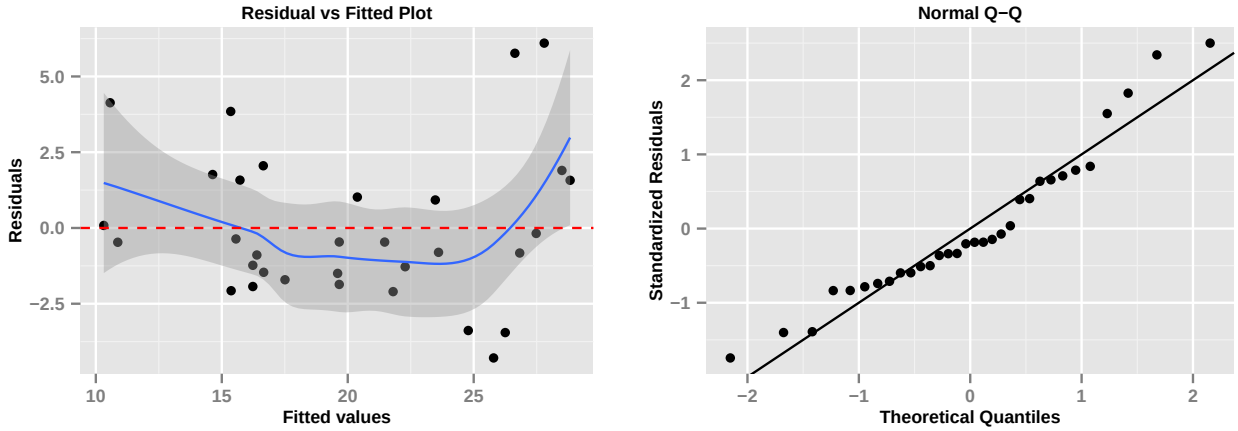
```
model3<- lm(mpg ~ wt + cyl, data=mtcars)
```

Model Comparison - The Adjusted R Squared is used as a basis for model comparison, the final model selected is the third model with the highest Adjusted R Squared (link) of 0.8185. This indicates that 81.85% of the variance in mpg is explained by the model.

```
## Model 1 Model 2 Model 3
##  0.8180  0.8122  0.8185
```

RESIDUALS AND DIAGNOSTICS

Here, we perform a residual plot and some diagnostics on the linear model (model3) to see if assumptions of a linear regression are violated. The residuals plot shows a random pattern. Hence, a linear model is appropriate for the data. The Normal Q-Q plot indicates that standardized residuals are approximately normal. The assumption of normal errors is also not violated. Refer to the appendix (Figure 5) for additional diagnostics plots.



We also did not observe any single point that has any huge effect on the estimate of a regression coefficient. Refer to the appendix (Figure 6) for the DFBETAS values.

CONCLUSION

Relationship between Transmission and MPG - From model 2, we can see that am is not significantly associated with mpg with a p-value of 0.893. Hence, we conclude that there is no significant relationship between transmission and mpg. Furthermore, from model 3, we are able to account for 81.85% of the variation (based on Adjusted R Squared) in mpg with the predictors cyl and wt.

Interpreting the coefficients - Based on model 3, holding all other variables constant, for every unit change wt, will result in -3.19 change in mpg and for every unit change in cyl, will result in -1.51 change in mpg.

Statistical Inference - At 95% confidence interval, and with F statistic of 70.91 and a p-value of less than 0.05 for model 3, we reject the null hypothesis that the slope is zero.

APPENDIX

Figure 1 - Correlation Matrix

```
##        mpg   cyl  disp    hp  drat    wt  qsec    vs    am  gear  carb
## mpg   1.00 -0.85 -0.85 -0.78  0.68 -0.87  0.42  0.66  0.60  0.48 -0.55
## cyl  -0.85  1.00  0.90  0.83 -0.70  0.78 -0.59 -0.81 -0.52 -0.49  0.53
## disp -0.85  0.90  1.00  0.79 -0.71  0.89 -0.43 -0.71 -0.59 -0.56  0.39
## hp   -0.78  0.83  0.79  1.00 -0.45  0.66 -0.71 -0.72 -0.24 -0.13  0.75
## drat  0.68 -0.70 -0.71 -0.45  1.00 -0.71  0.09  0.44  0.71  0.70 -0.09
## wt   -0.87  0.78  0.89  0.66 -0.71  1.00 -0.17 -0.55 -0.69 -0.58  0.43
## qsec  0.42 -0.59 -0.43 -0.71  0.09 -0.17  1.00  0.74 -0.23 -0.21 -0.66
## vs    0.66 -0.81 -0.71 -0.72  0.44 -0.55  0.74  1.00  0.17  0.21 -0.57
## am    0.60 -0.52 -0.59 -0.24  0.71 -0.69 -0.23  0.17  1.00  0.79  0.06
## gear  0.48 -0.49 -0.56 -0.13  0.70 -0.58 -0.21  0.21  0.79  1.00  0.27
## carb -0.55  0.53  0.39  0.75 -0.09  0.43 -0.66 -0.57  0.06  0.27  1.00
```

Figure 2 - Model 1 Coefficients

```
##
## Call:
## lm(formula = mpg ~ disp + hp + wt + cyl + drat + vs + am, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.660 -1.678 -0.417  1.371  5.312
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 32.45671    9.02383   3.597  0.00145 **
## disp         0.01348    0.01212   1.112  0.27695
## hp          -0.03032    0.01469  -2.063  0.05005 .
## wt          -3.24531    1.16754  -2.780  0.01041 *
## cyl         -0.63992    0.89674  -0.714  0.48235
## drat         0.54696    1.51009   0.362  0.72037
## vs1          1.39761    1.84843   0.756  0.45694
## am1          1.95201    1.75665   1.111  0.27749
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.571 on 24 degrees of freedom
## Multiple R-squared:  0.8591, Adjusted R-squared:  0.818
## F-statistic: 20.91 on 7 and 24 DF,  p-value: 9.089e-09
```

Figure 3 - Model 2 Coefficients

```
##
## Call:
## lm(formula = mpg ~ wt + cyl + am, data = mtcars)
##
```

```
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.1735 -1.5340 -0.5386  1.5864  6.0812
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  39.4179     2.6415  14.923 7.42e-15 ***
## wt           -3.1251     0.9109  -3.431  0.00189 **
## cyl          -1.5102     0.4223  -3.576  0.00129 **
## am1           0.1765     1.3045   0.135  0.89334
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.612 on 28 degrees of freedom
## Multiple R-squared:  0.8303, Adjusted R-squared:  0.8122
## F-statistic: 45.68 on 3 and 28 DF,  p-value: 6.51e-11
```

Figure 4 - Model 3 Coefficients

```
##
## Call:
## lm(formula = mpg ~ wt + cyl, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.2893 -1.5512 -0.4684  1.5743  6.1004
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  39.6863     1.7150  23.141  < 2e-16 ***
## wt           -3.1910     0.7569  -4.216 0.000222 ***
## cyl          -1.5078     0.4147  -3.636 0.001064 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.568 on 29 degrees of freedom
## Multiple R-squared:  0.8302, Adjusted R-squared:  0.8185
## F-statistic: 70.91 on 2 and 29 DF,  p-value: 6.809e-12
```

Figure 5 - Additional Diagnostics Plot

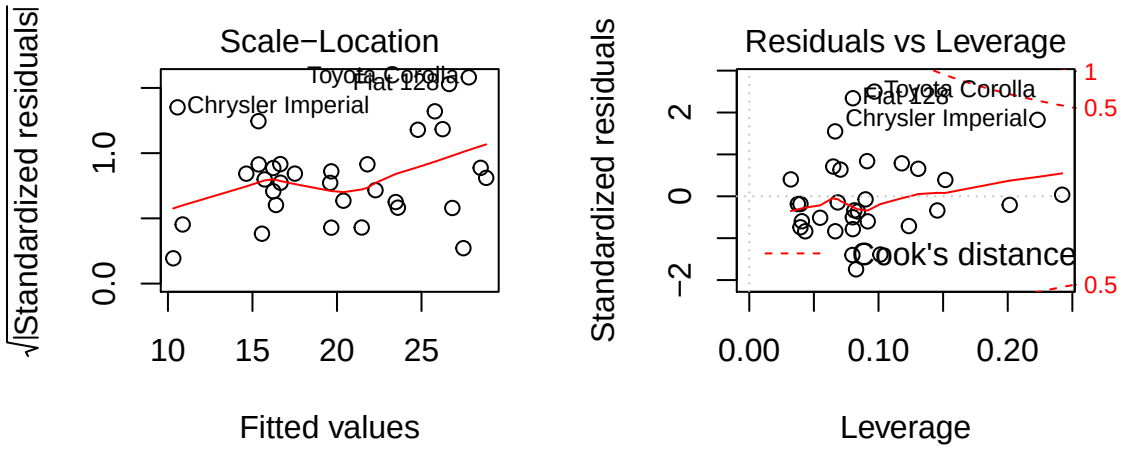


Figure 6 - DFBETAS

```
##                     (Intercept)          wt         cyl
## Mazda RX4          -0.053237261  0.079278772 -0.055926520
## Mazda RX4 Wag      -0.015479896  0.014296567 -0.009012663
## Datsun 710         -0.369864784 -0.017774216  0.217583190
## Hornet 4 Drive      0.023693902  0.009311472 -0.012032565
## Hornet Sportabout  -0.072034982 -0.143014091  0.211242595
## Valiant            -0.024432351 -0.057307617  0.051906281
## Duster 360          0.074633635  0.101628775 -0.171741855
## Merc 240D           0.077263805  0.112208295 -0.145061758
## Merc 230           -0.067709483 -0.092541841  0.121806994
## Merc 280           -0.007760996 -0.016480956  0.015061621
## Merc 280C          -0.031532428 -0.066961068  0.061194401
## Merc 450SE         -0.093624241  0.016252299  0.069919958
## Merc 450SL         -0.067815115 -0.050994986  0.114190165
## Merc 450SLC         0.015962344  0.009399100 -0.024241733
## Cadillac Fleetwood  0.048663076 -0.083681991  0.039824506
## Lincoln Continental -0.009665873  0.017864085 -0.009168135
## Chrysler Imperial  -0.480801336  0.861419380 -0.428527327
## Fiat 128            0.686041200 -0.062387471 -0.316036811
## Honda Civic         0.250248213 -0.163103469  0.013570292
## Toyota Corolla      0.829946808 -0.382392720 -0.100548368
## Toyota Corona      -0.450332511 -0.103459346  0.340037555
## Dodge Challenger    0.032927567  0.052209217 -0.083233710
## AMC Javelin         0.050817406  0.102256391 -0.150405432
## Camaro Z28          0.095974853  0.039083287 -0.128109809
## Pontiac Firebird   -0.184447112 -0.072438903  0.243499406
## Fiat X1-9          -0.021036429  0.007710838  0.004367838
## Porsche 914-2      -0.090364890  0.014180745  0.036153448
## Lotus Europa        0.214901357 -0.157636752  0.027785578
## Ford Pantera L      0.047014762  0.182881129 -0.228521082
## Ferrari Dino       -0.077660140  0.091993150 -0.061997518
## Maserati Bora       0.047290986  0.064396233 -0.108822806
## Volvo 142E         -0.320695123 -0.220151415  0.376628991
```

The rmd file is available here: link