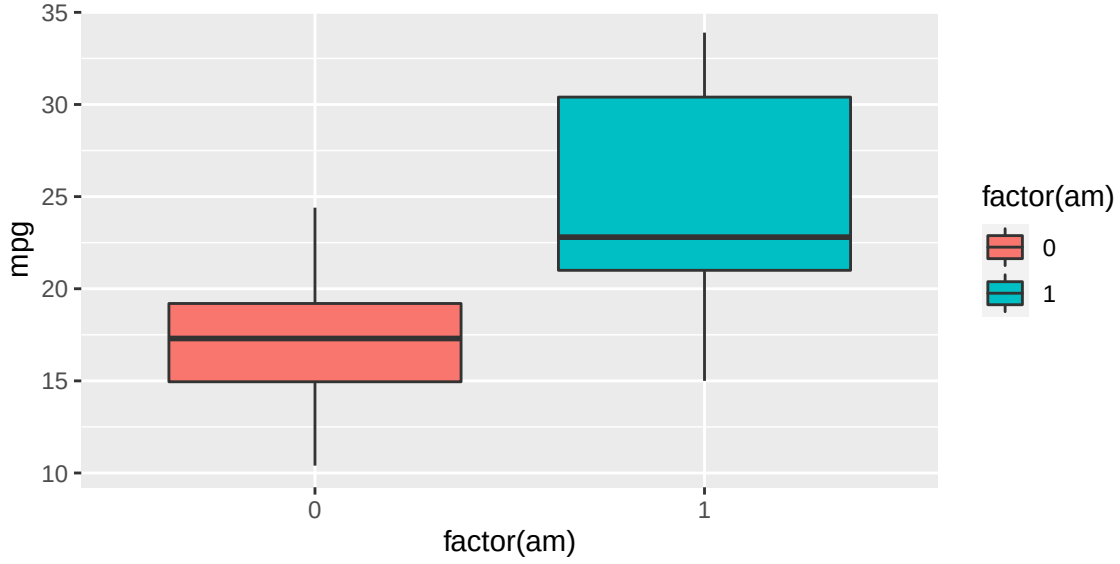
RM Course Project: Manual or Automatic Transmission for MPG?

Executive Summary

This document will be an exploration of whether manual (am = 1) or automatic (am = 0) transmissions are better for gas mileage, according to the mtcars dataset. Exploratory data analysis and regression models will be used to determine whether the gas mileage, measured in miles per gallon, is higher for cars with an automatic transmission, or cars with a manual transmission.

Exploratory Data Analysis

A quick examination of the data in mtcars suggests that the mean mpg for cars depends significantly on the type of transmission. It appears that the difference in mean mpg is 7.24, with manual transmissions having the higher mpg. This on its own suggests that manual transmissions improve gas mileage, but more careful examination needs to be done.




```
## [1] "The mean of mpg for automatic transmission is 17.15"
```

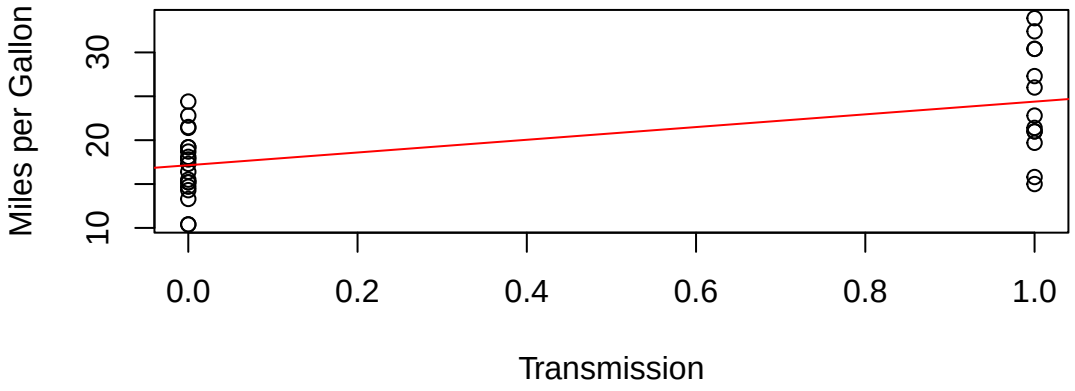
```
## [1] "The mean of mpg for manual transmission is 24.39"
```

Initial Regression Models

Next we will try to make some linear regression models to see whether am is a good way of predicting the value of mpg. The initial simple model is one looking at the effect of am alone on mpg.

```
## [1] "mpg~am"
```

```
##              Estimate     Pr(>|t|)
## (Intercept) 17.147368 1.133983e-15
## am           7.244939 2.850207e-04
```



In the simplest fit looking only at am, the coefficient for the intercept is 17.15, suggesting that when am is 0 (an automatic transmission), the estimate for the mpg is 17.15 miles per gallon. The coefficient for the am variable is 7.24, suggesting that a switch from automatic to manual transmission improves gas mileage my 7.24 miles per gallon up to 24.39 miles per gallon. This is consistent with the difference in mean calculated above, suggesting that indeed, a manual transmission is better for gas mileage.

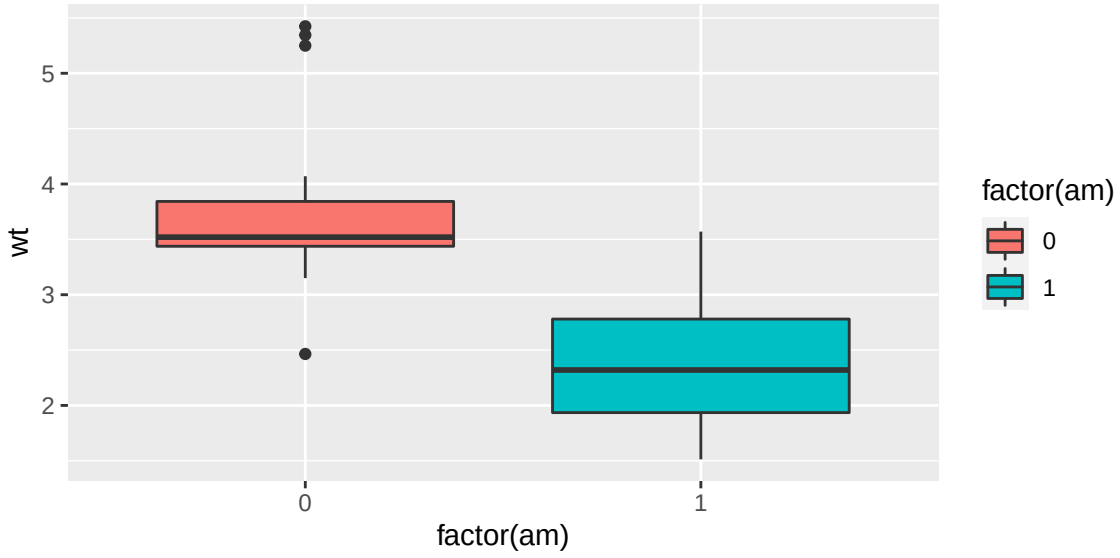
Next, we will look at a model of mpg predicted by all of the variables in the data set to see the effect of am when the other variables have been controlled for, as well as to see what variables are the ones that have the most influence over am.

```
## [1] "mpg~."
```

```
##                Estimate   Pr(>|t|)
## (Intercept) 12.30337416 0.51812440
## cyl         -0.11144048 0.91608738
## disp         0.01333524 0.46348865
## hp          -0.02148212 0.33495531
## drat         0.78711097 0.63527790
## wt          -3.71530393 0.06325215
## qsec         0.82104075 0.27394127
## vs           0.31776281 0.88142347
## am           2.52022689 0.23398971
## gear         0.65541302 0.66520643
## carb        -0.19941925 0.81217871
```

We can see that when all of the variables are taken into consideration in this second model, the coefficient for am changes significantly to 2.52, and the significance is low. It may be that the effect observed in the simple model is actually being caused in large part by other variables. The variable that has the most effect on mpg is wt, it has the largest coefficient and the smallest p value, so this variable should be examined. It is also the only variable in this model that has a significant effect with an alpha = 0.10.

Considering Weight




```
## [1] "The mean of weight for automatic transmission is 3.77"
```

```
## [1] "The mean of weight for manual transmission is 2.41"
```

The first thing to check is whether there is a relationship between am and wt. It can be seen above that there is a significant difference in wt between automatic and manual cars, with manual cars being 1.35 (1000 lbs) heavier on average. Given this and the fact that wt had the most significant and largest slope estimate from the total model, it makes sense then for the next model to be one that examines the effect of am and wt, as well as another model that also includes their interaction term. The data from the two models are shown below.

```
## [1] "mpg~wt+am"
```

```
##                Estimate     Pr(>|t|)
## (Intercept) 37.32155131 5.843477e-13
## wt          -5.35281145 1.867415e-07
## factor(am)1 -0.02361522 9.879146e-01
```

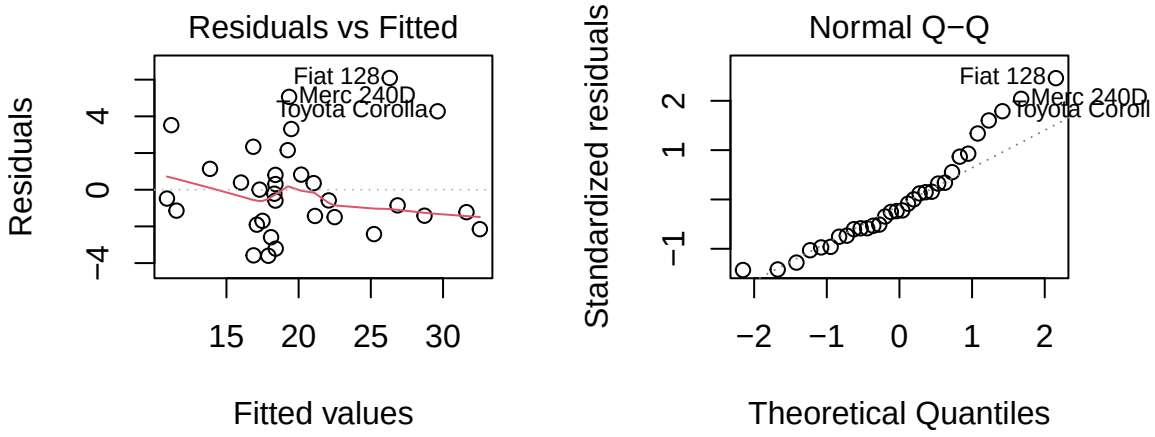
```
## [1] "mpg~wt*am"
```

```
##                 Estimate     Pr(>|t|)
## (Intercept)    31.416055 4.001043e-11
## wt             -3.785908 4.551182e-05
## factor(am)1    14.878423 1.621034e-03
## wt:factor(am)1 -5.298360 1.017148e-03
```

In the first model the estimate of the slope for am is very small, and the p is large. Initially it appears that after correcting for wt, am no longer has a significant effect on mpg, so this model does not work well as a way to predict mpg based on am. However, looking at the model that includes the interaction term, we can see that the estimates are all statistically significant, suggesting this the is best model seen so far. This is not surprising given the significant relationship between am and mpg shown above.

Evaluating the Model

The final model should be checked for any significant issues. A plot of the residuals vs the fitted values and a normal QQ plot show that the model does not have any enormous issues, however it is slightly heteroscedastic, since the residuals are spread out less in the larger fitted values.



Analysis of Final Model

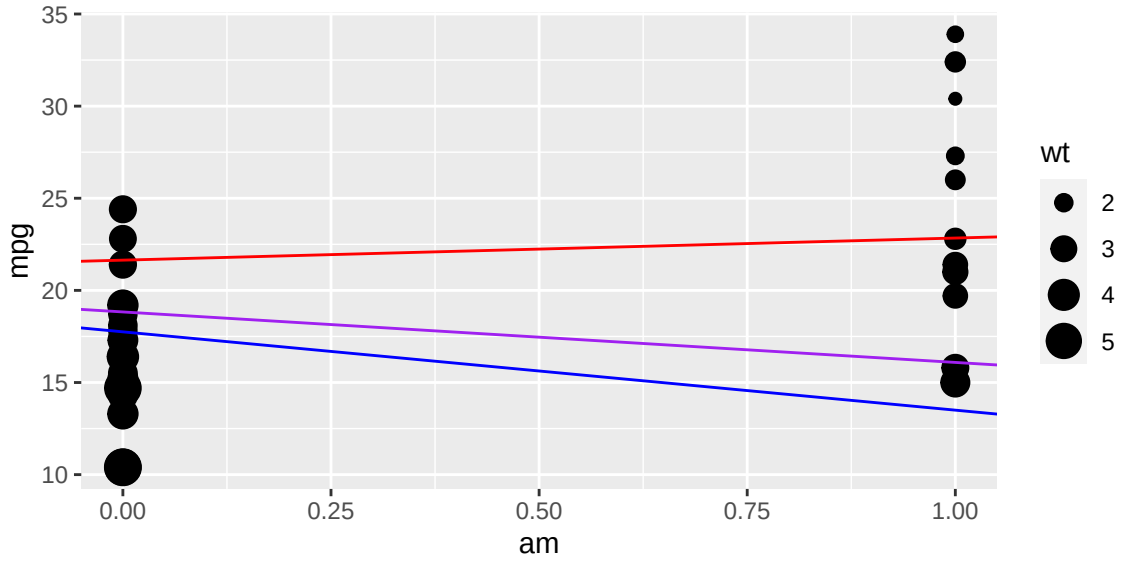
The interpretation of the coefficient estimates from the final model does not lead to a straightforward answer to whether a manual or automatic transmission is better for gas mileage. The model, with coefficients inserted, is:

```
## [1] "mpg = 31.42 - 3.79*wt + 14.88*am - 5.30*wt*am"
```

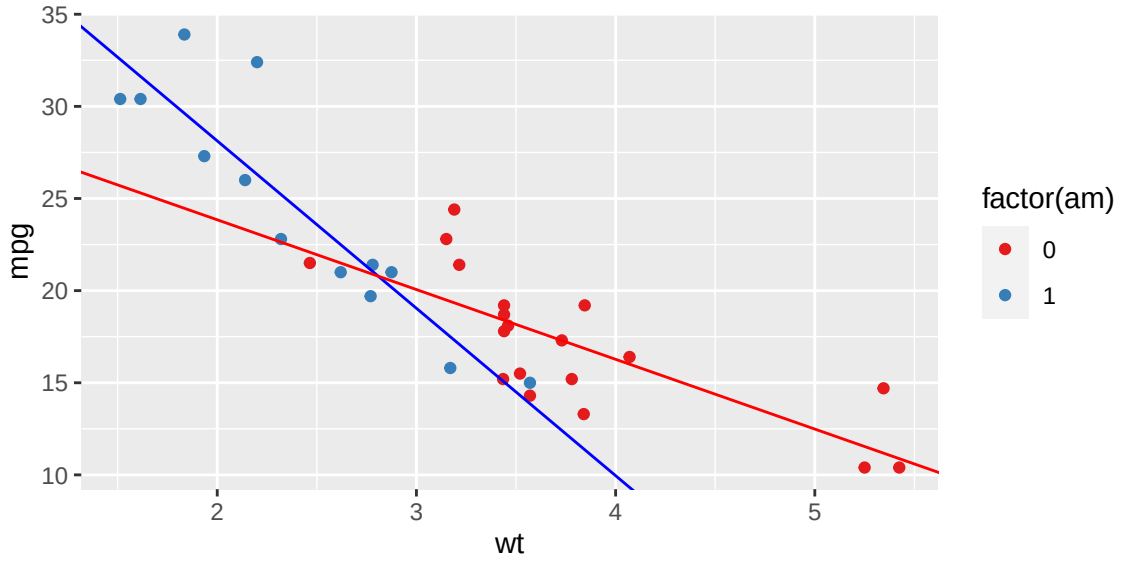
An examination of the regression equation shows that different values of wt leads to different conclusions for the effect of am on mpg: for low weights, manual transmissions increase mpg, but for high weights, manual transmissions decrease mpg. With wt fixed at its first, second and third quartiles, we can see the intercepts and slopes for the mpg~am graph in the table below.

Weight	Intercept	Slope
2.58(Q1)	21.64	1.20
3.33(Q2)	18.83	-2.74
3.61(Q3)	17.75	-4.25

These slopes show us that the effect of am on mpg depends strongly on the weight, and in fact, not only is the effect weaker than the one we saw in the simple mpg~am model - which was 7.245 - the effect is negative for most values of weight, suggesting that for average or heavier cars, a manual transmission reduces gas mileage, the opposite of the conclusion we obtained from the simpler model. Below we see the graph of mpg and am, with the regressions plotted with wt held fixed at its three quartiles.



The plot below shows us how the final model predicts mpg based on wt with the am held fixed. The mpg of a car decreases as weight increases, but the effect of weight on mpg is larger for manual transmissions with a slope of -9.0843 miles per gallon per 1000lbs, compared to the smaller effect on automatic transmissions of -3.7859 miles per gallon per 1000 lbs.



Final Conclusion

Despite the initial appearance that a manual transmission is better for gas mileage, the final model suggests that in fact for most cars, automatic transmissions are better. However it is strongly effected by weight, and for light cars the manual transmission is better.