Investigating the effect of transmission on mpg by Malna Polya

Executive summary

In this study, we are examining the effect of manual versus automatic transmission on the mpg of cars in the mtcars dataset. The result of our examination shows that cars with manual transmission have a higher mpg by 7.24 on average, the manual transmission is better. However, this result might be distorted by other variables, in case of a multivariable regression, the am variable tends to be insignificant, therefore more examination with a broader dataset is recommended.

Initial exploratory analysis of the dataset

We can produce a boxplot which compares the mpg of the cars with automatic and manual transmission. The resulting plot is shown in the Appendix. It is clear from this that manual transmission cars have a better mpg than automatic transmission cars do.
However, if we explore the relationship including the weight as an additional independent variable, we can see that this relationship between the transmission and the mpg might be existing because of the relationship between transmission and mpg (see Appendix). Automatic transmission cars tend to have a higher weight and manual transmission a lower weight so this could result in the increased performance of the manual cars.

Conducting single and multivariable regressions

Single variable regression

To recognize the relationship between the transmission and the mpg, we conduct a linear regression between the two variables. The coefficient of the dummy variable “am” will show the difference between the average performance of a manual and an automatic car. Note that am is 0 if the transmission is automatic and 1 if the transmission is manual.

```
data(mtcars)
fit<-lm(mpg~am,data=mtcars)
round(summary(fit)$coef,2)
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    17.15       1.12   15.25        0
## am              7.24       1.76    4.11        0
```

We can verify that the manual car performs better in terms of mpg because am has a positive coefficient of 7.24. This means that manual cars have on average an mpg that is 7.24 higher than that of automatic cars. We can verify these results by subtracting the average automatic car mpg from the average manual car mpg. Also, from the previous table we can note that these results are significant.

```
round(mean(mtcars$mpg[mtcars$am==1])-mean(mtcars$mpg[mtcars$am==0]),2)
```

```
## [1] 7.24
```

This results in the same value, 7.24.

Empirical multivariable regression

However, if we conduct a multivariable regression, including weight as well, we get a different result. The criteria for choosing weight comes from practical experience: that heavier cars tend to have lower mpg's.

```
fit<-lm(mpg~am+wt,data=mtcars)
round(summary(fit)$coef,2)
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    37.32       3.05   12.22     0.00
## am             -0.02       1.55   -0.02     0.99
## wt             -5.35       0.79   -6.79     0.00
```

The sign of the coefficient changed along with the significance of the transmission. According to this result, the transmission is not at all significant but the weight is.

Multivariable regression with a strategic approach

We can also fit a model with all the variables and then strategically get rid of the most insignificant variables (to save space several at a time without showing the results).

```
fit<-lm(mpg~.,data=mtcars)
round(summary(fit)$coef,2)
```

We leave the 6, then the 3 most significant variables in the model.

```
fit<-lm(mpg~disp+hp+wt+qsec+am+gear,data=mtcars)
```

```
fit<-lm(mpg~wt+qsec+am,data=mtcars)
round(summary(fit)$coef,2)
```

```
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)     9.62       6.96    1.38     0.18
## wt             -3.92       0.71   -5.51     0.00
## qsec            1.23       0.29    4.25     0.00
## am              2.94       1.41    2.08     0.05
```

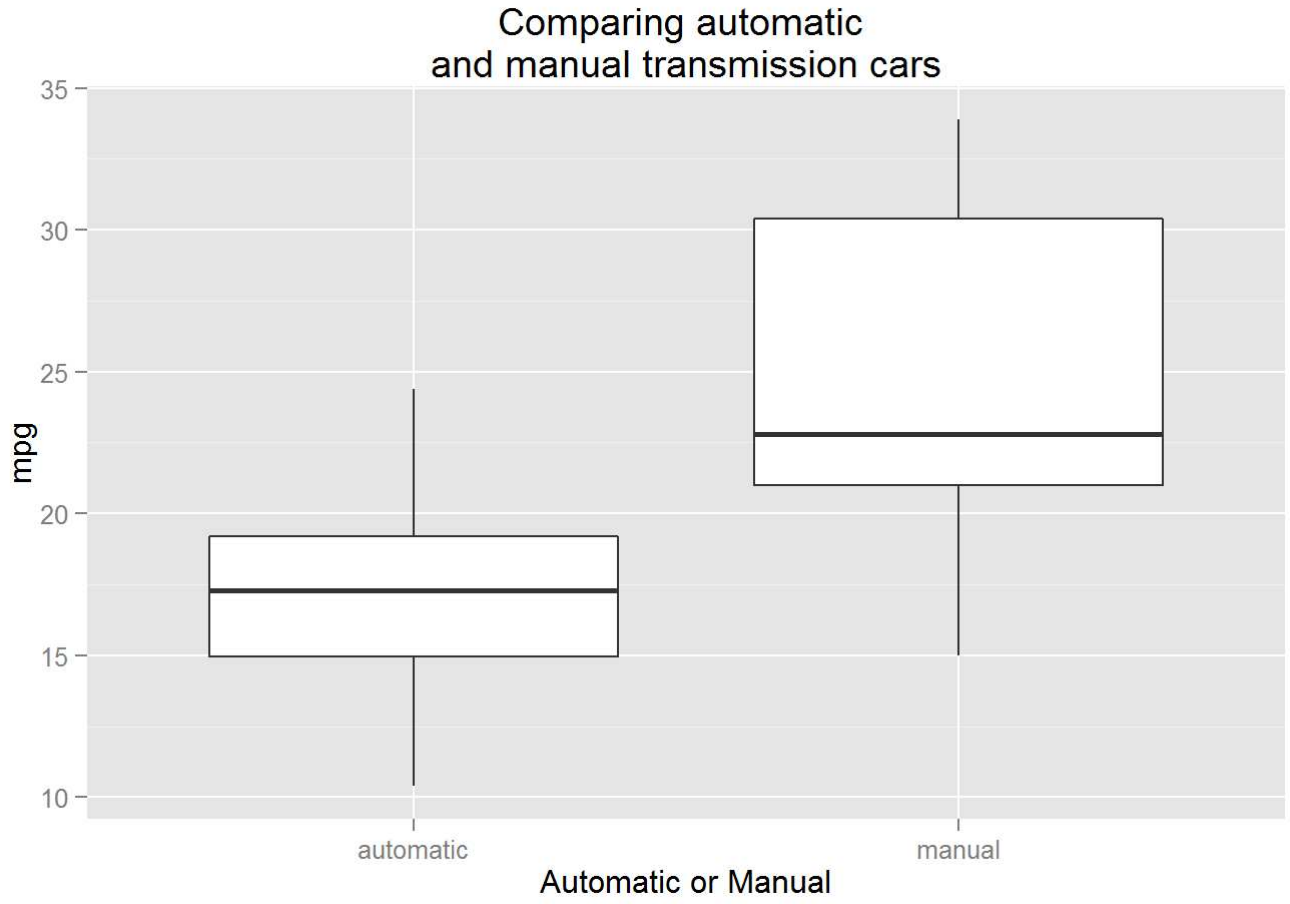
In this case, the am variable seems significant with a coefficient 2.94. This means that only by changing the transmission from automatic to manual, the mpg increases by 2.94. The sign of the relationship is the same as in the initial single variable model. We can see the residuals plot in the appendix where we can see that they are randomly distributed.

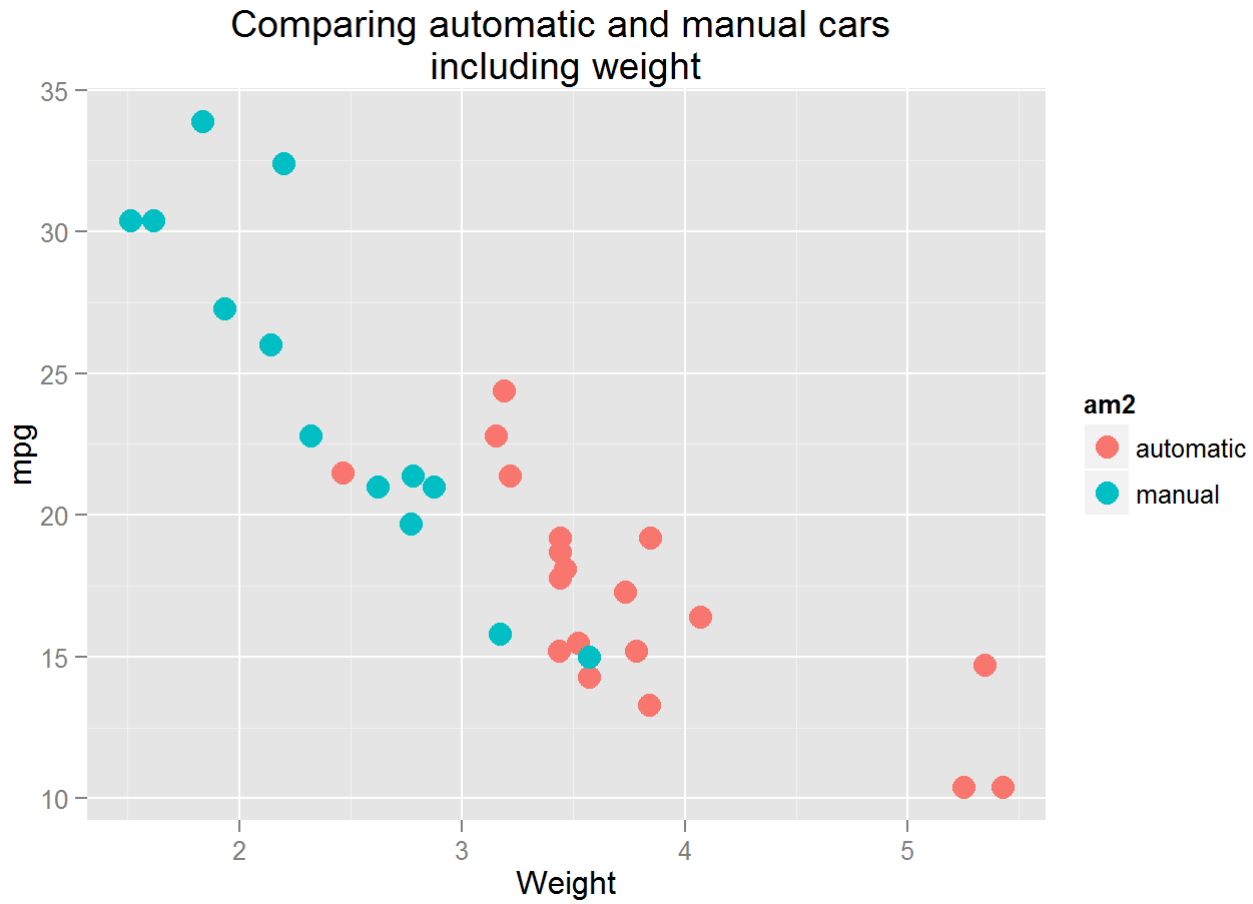
Results

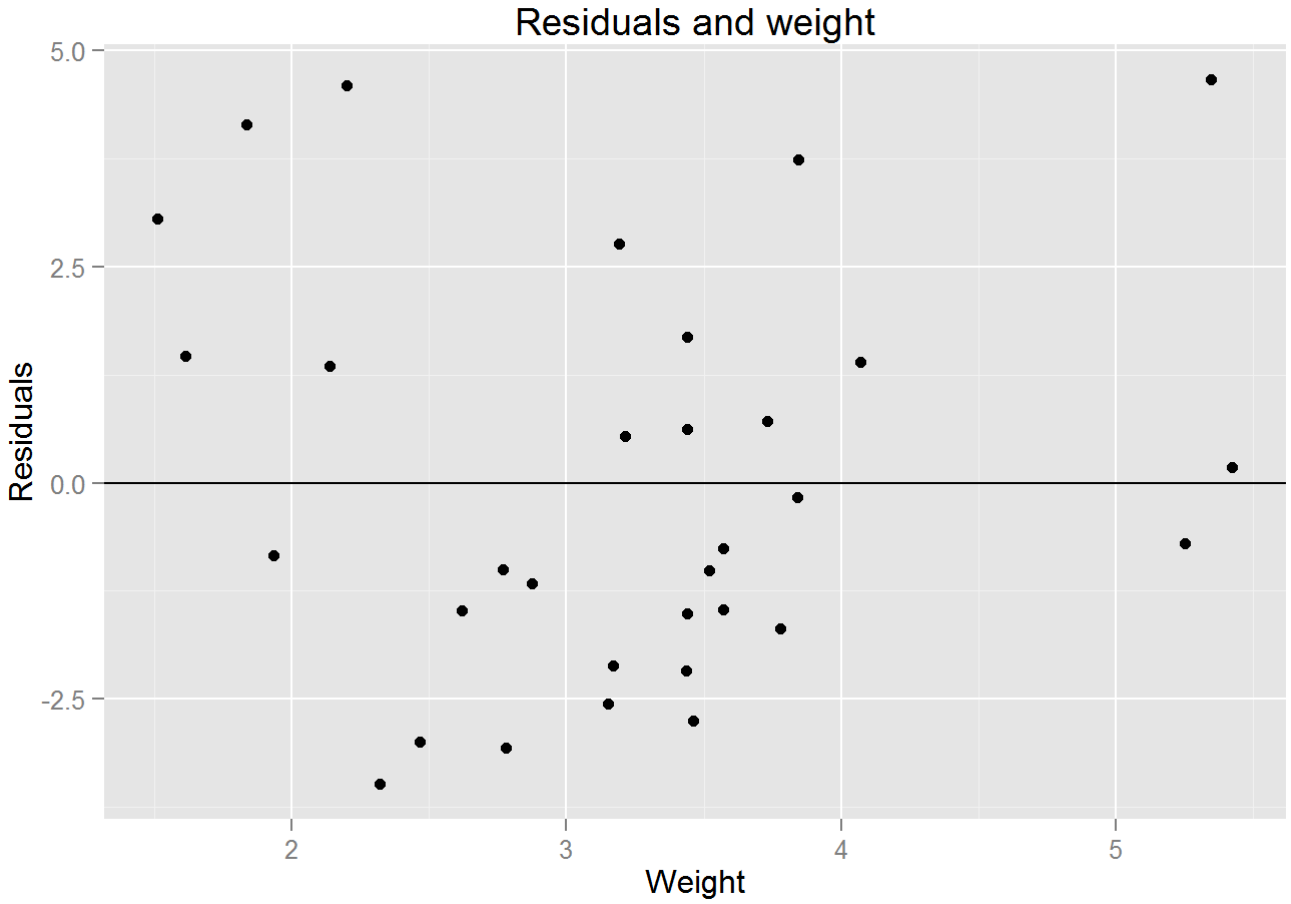
The result of our examination is that based on the single variable relationship between the transmission and the mpg, a manual transmission car has on average 7.24 better mpg than an automatic one. However, examining the multivariable relationship between mpg as dependent and transmission and weight as independent variables, transmission becomes insignificant and there is no real difference between a car with automatic and manual transmission per se. If we examine a model that includes mpg as dependent and am, qsec and wt as independent variables, the coefficient of am is positive and a bit lower than in the single variable model: 2.94.

Appendix

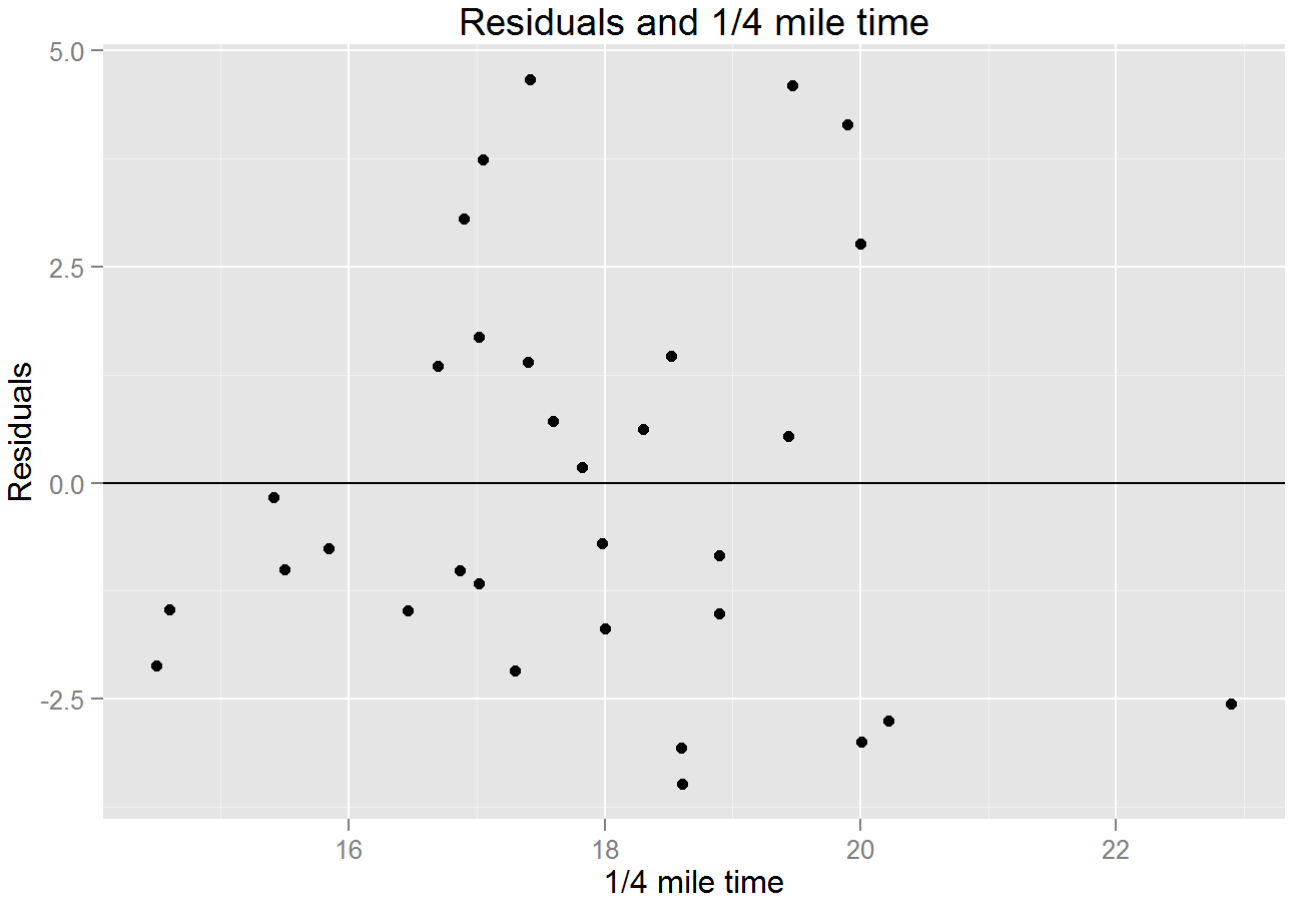
```
## Warning: package 'ggplot2' was built under R version 3.1.3
```






Residuals and weight
Residuals
5.0
2.5
0.0
-2.5
2
3
4
5
Weight


Residuals and Transmission
Residuals
5.0
2.5
0.0
-2.5
automatic
manual
Transmission


Residuals and 1/4 mile time
5.0
2.5
0.0
-2.5
Residuals
16
18
20
22
1/4 mile time