

Analysis on Transmission Type

Justin Pizzino

8/15/2019

1. Executive Summary

Purpose of this research paper is to determine if an automatic or manual transmissions is better for MPG, and quantify that difference. To do this we will use the Motor Trend data.

3. Load the data

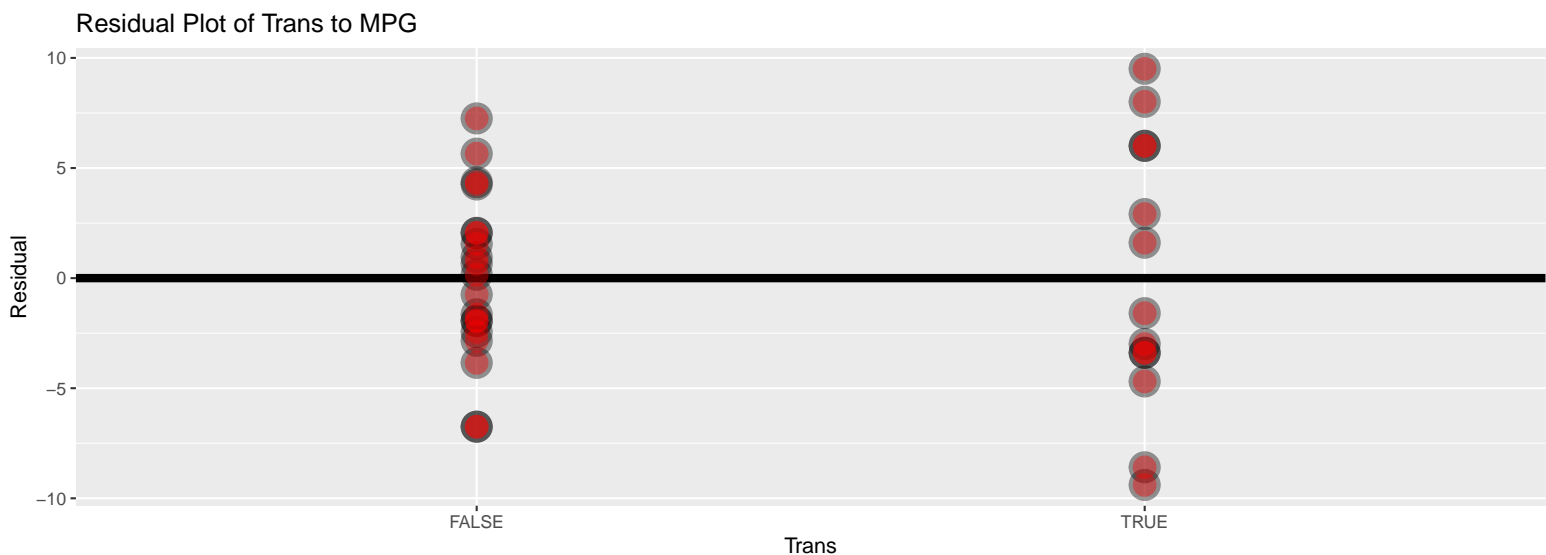
First load the data, then convert the Transmission type (am) to Boolean, and split out cylinder into dummy variables. Details are in the appendix.

4. Exploratory Data Analysis

Easy way to start is to do a direct compare of mpg to the transmission type. From here it would appear that **a manual transmission would add about 7.2 MPG**. We can see below that transmission is **statistically significant with low p value of 0.000285**, but it's **adjusted r squared value is low, at only 0.3384589** meaning only **33.8% of the data** is explained by this model. Further more if we plot the residuals we can see there is something else going on with with the manual transmission. Something else is impacting the results. Coefficients below:

```
summary(lm(mpg~am,mtcars))$coef
```

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

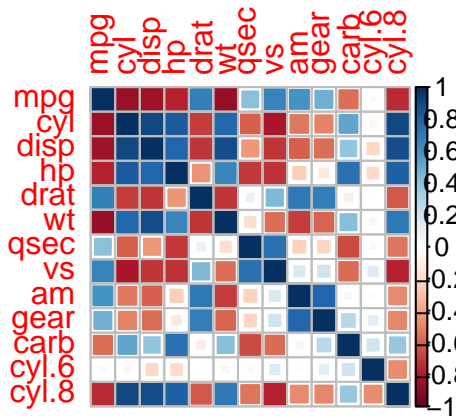


5. Additional exploratory analysis using fit models & define strategy

We know we should include other data, but need to determine which data points. Let's do some exploratory analysis. I compared the correlation of each data point to it self.

Here we can see that many other values that are correlated to MPG. We need to factors those in to determine transmissions impact. Weight, displacement, number of cyl, and engine shape (vs) all are very correlated to MPG, but also to each other.

```
corrplot(cor(mtcars),method="square")
```



As weight has the highest correlation to MPG I choose that value to add, and created a new model.

```
summary(lm(mpg~am+wt,mtcars))$coef
```

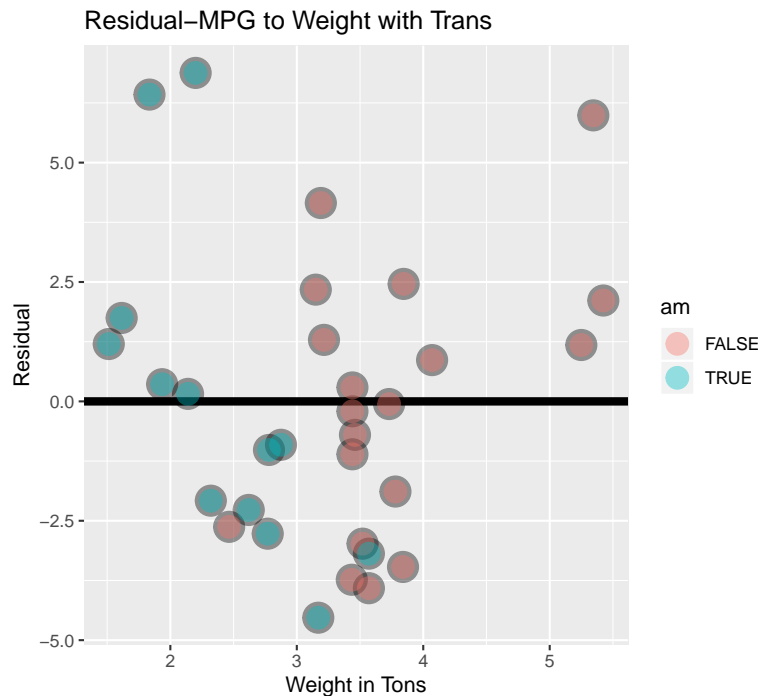
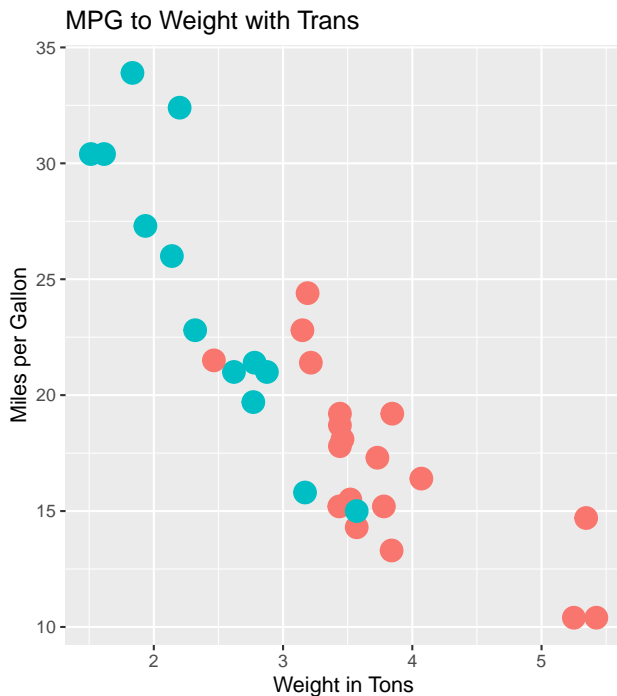
```
##              Estimate Std. Error    t value    Pr(>|t|)
## (Intercept) 37.32155131  3.0546385 12.21799285 5.843477e-13
## amTRUE      -0.02361522  1.5456453 -0.01527855 9.879146e-01
## wt          -5.35281145  0.7882438 -6.79080719 1.867415e-07
```

```
summary(lm(mpg~am+wt,mtcars))$adj.r.squared
```

```
## [1] 0.7357889
```

Based on this model it would appear a manual transmission would reduce MPG by 0.02. Also again, **looking at the coefficient, each ton of weight decreases MPG by about 5.35.**

Now that seems strange, our rss value went up, which is great, showing greater correlation between our model and observed data. **But the statistical significant of transmission type went down.** To understand why, I created a plot of weight to mpg, with color coding for the two trans types. On the left we can see that Auto trans tend to be in lighter cars, and thus it appears the impact is larger with the weight then the trans type. On the right we can see a small pattern of cars weighing 3-4 tons having MPG below our fitted line. There could be other factors here, but our focus is on Transmission type, and not weight.



With this plot we can now see that while MPG does go down with manual transmission, we also see that weight goes up, and appears to be the larger factor in impacting MPG.

6. Answer to the two questions

Based on the results, I would say, there is **not enough data to determine the impact** of transmission on MPG. The **p value of 0.98 too low making it not significant**. Plus the majority of the change can be explained away by weight of the vehicle. The coefficient shows it's possible an automatic transmission is better for MPG, by 0.02 MPG, but again there is not enough data to support that.

7. Appendix

Libraries used

```
library(stats)
library(ggplot2)
library(car)
library(corrplot)
```

Code for the plot

```
# Residual Trans to MPG
ggplot(data.frame(x=as.factor(mtcars$am), y=resid(lm(mpg~as.factor(am),mtcars))),
  aes(x=x,y=y))+
  geom_hline(yintercept = 0,size=2)+
  geom_point(size=7, colour="black", alpha=0.4)+
  geom_point(size=5,colour="red",alpha=0.4)+
  xlab("Trans")+ylab("Residual")+labs(title="Residual Plot of Trans to MPG")
```

```
# MPG to Weight with Trans
ggplot(mtcars, aes(wt, mpg))+
  geom_point(aes(col=as.factor(am)),size=5)+
  labs(title = "MPG to Weight with Trans",
    x="Weight in Tons",
    y="Miles per Gallon")+
  scale_colour_discrete(name="Trans",
    labels=c("Manual","Auto"))
```

```
data(mtcars)
mtcars<-mtcars
mtcars$am<-as.logical(mtcars$am)
mtcars$cyl.6<-as.logical(FALSE)
mtcars$cyl.6[ which(mtcars$cyl==6)]<-as.logical(TRUE)
mtcars$cyl.8<-as.logical(FALSE)
mtcars$cyl.8[ which(mtcars$cyl==8)]<-as.logical(TRUE)
```