

Analysis on Transmission Type

Justin Pizzino

8/15/2019

R Markdown

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

```
data(mtcars)
#which items should be factors?
mtcars<-mtcars
mtcars$am<-as.logical(mtcars$am)
#V2mtcars$cyl<-as.factor(V2mtcars$cyl)
#mtcars$cyl.4<-as.logical(FALSE) #There is a better way to do this
#mtcars$cyl.4[cyl=4]<-as.logical(TRUE)
mtcars$cyl.6<-as.logical(FALSE) #There is a better way to do this
mtcars$cyl.6[ which(mtcars$cyl==6)]<-as.logical(TRUE)
mtcars$cyl.8<-as.logical(FALSE) #There is a better way to do this
mtcars$cyl.8[ which(mtcars$cyl==8)]<-as.logical(TRUE)
```

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, but it's adjusted r squared value is low, at only .338 meaning only 33.8% of the data is explained by this model. So let's add more data points in to create a stronger model.

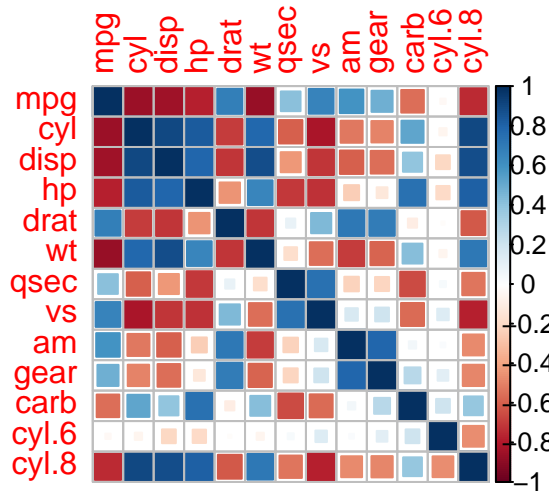
```
summary(lm(mpg~am,mtcars))

##
## 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 ***
## amTRUE         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
```

We know we should include other data, but need to determine which data points. So 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 to the model.

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

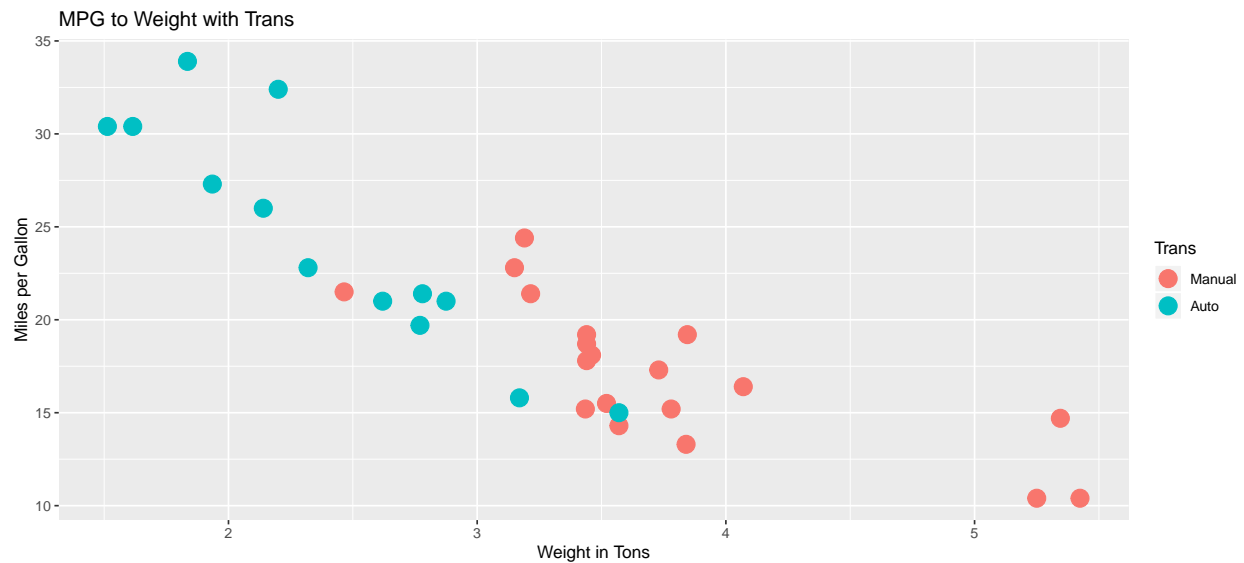
```
##
## Call:
## lm(formula = mpg ~ am + wt, data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.5295 -2.3619 -0.1317  1.4025  6.8782
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  37.32155    3.05464  12.218 5.84e-13 ***
## amTRUE       -0.02362    1.54565  -0.015  0.988
## wt          -5.35281    0.78824  -6.791 1.87e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.098 on 29 degrees of freedom
## Multiple R-squared:  0.7528, Adjusted R-squared:  0.7358
## F-statistic: 44.17 on 2 and 29 DF,  p-value: 1.579e-09
```

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, created a plot of weight to mpg, with color coding for the two trans types.

```
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"))
```



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.

Result

Based on the analysis done I would say it's possible an automatic transmission is better for MPG, by 0.02 MPG. But based on the results, I would say, there is not enough data to determine the impact of transmission on MPG. The p value is too low, and a majority of the change can be explained away by weight of the vehicle.