mtcars

Umercia

March 25, 2017

Executive summary

Having a look at the mtcars data (car database), we are interested to understand what drive the fuel consumption of those cars. We are particulary interested if the transmission (automatic or manual) has an important role. A first raw comparison on this parameter seems to show that automatic cars have in general higher fuel consumption. But after a deeper investigation, including other parameters, it is shown that this is more due to the fact that in general, automatic cars are heavier and have more cylinder than manual transmission. At the end, the effect of the transmission "only" is not proven.

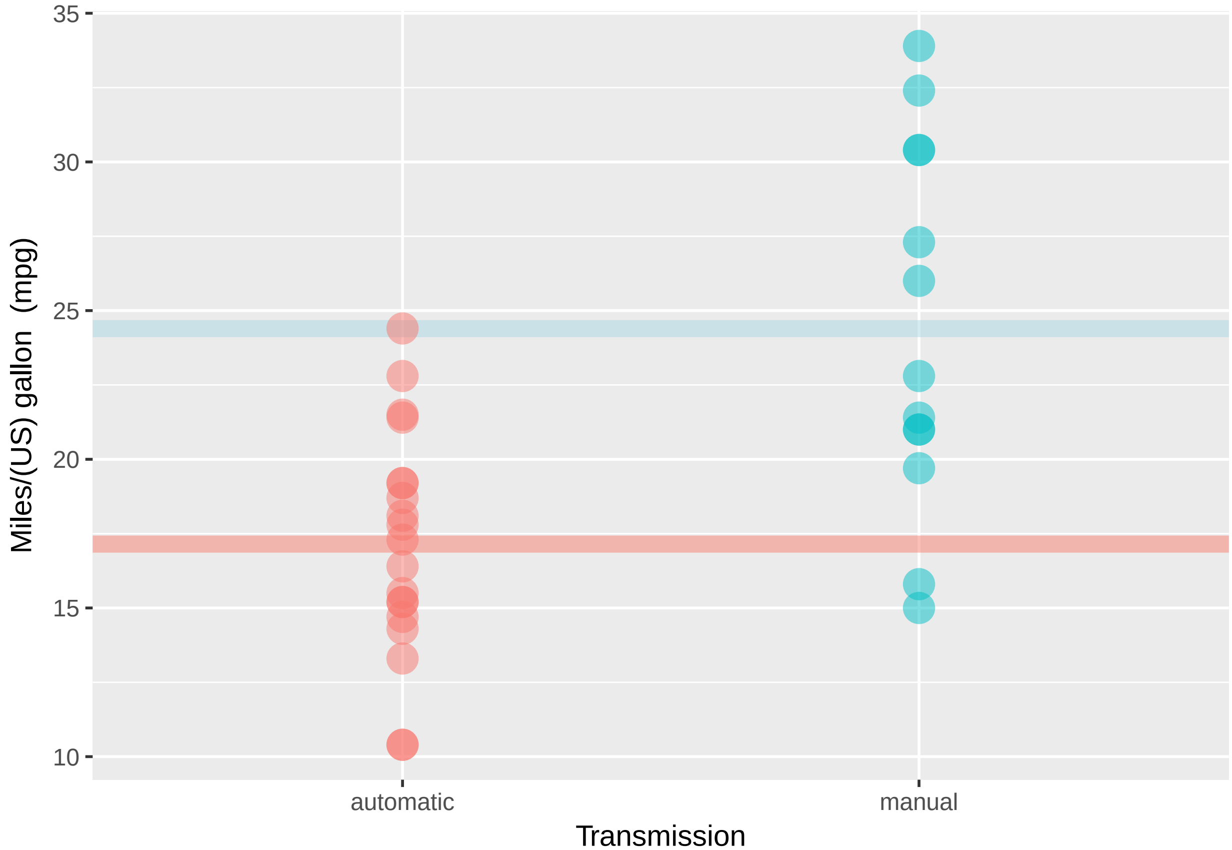
Exploratory data analysis

The data from "mtcars"" data package was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 models). A data frame with 32 observations on 11 variables:
- **mpg: Miles/(US) gallon**
- cyl: Number of cylinders
- disp: Displacement (cu.in.)
- hp: Gross horsepower
- drat: Rear axle ratio
- wt: Weight (1000 lbs)
- qsec: 1/4 mile time
- vs: V/S
- **am: Transmission (0 = automatic, 1 = manual)**
- gear: Number of forward gears
- carb: Number of carburetors

```r
data(mtcars)
head(mtcars)
```

```
##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

Below a plot of the mpg versus the transmission

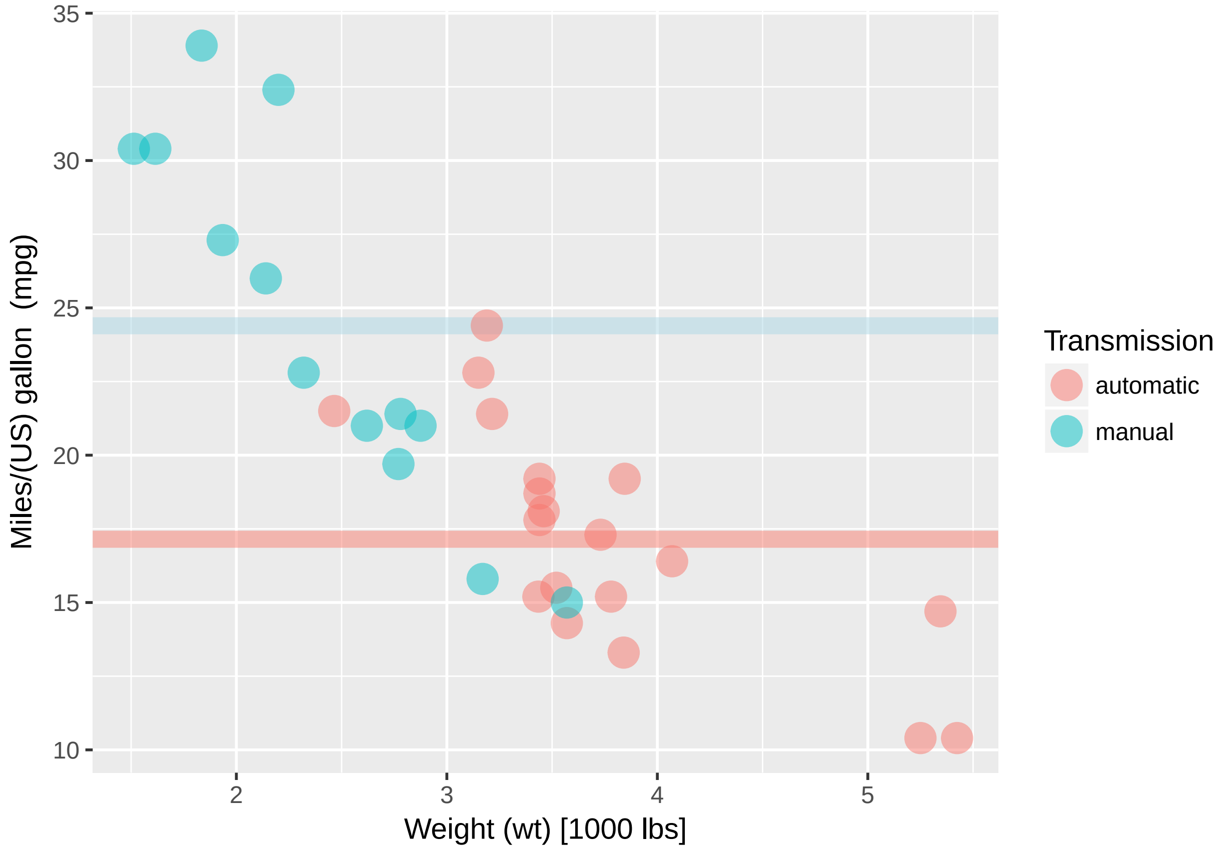


In the above graph, the elements are as follows:

- salmon line = mean of mpg for the group "automatic"
- light blue line = mean of mpg for the group "manual"

At first view, we can see that the group of automatic transmission has lower mpg (higher fuel consumption) than the manual. The question is now: Does the transmission is the reason of higher consumption or is there other(s) reason(s) behind this?

If we add the weight variable to the plot:



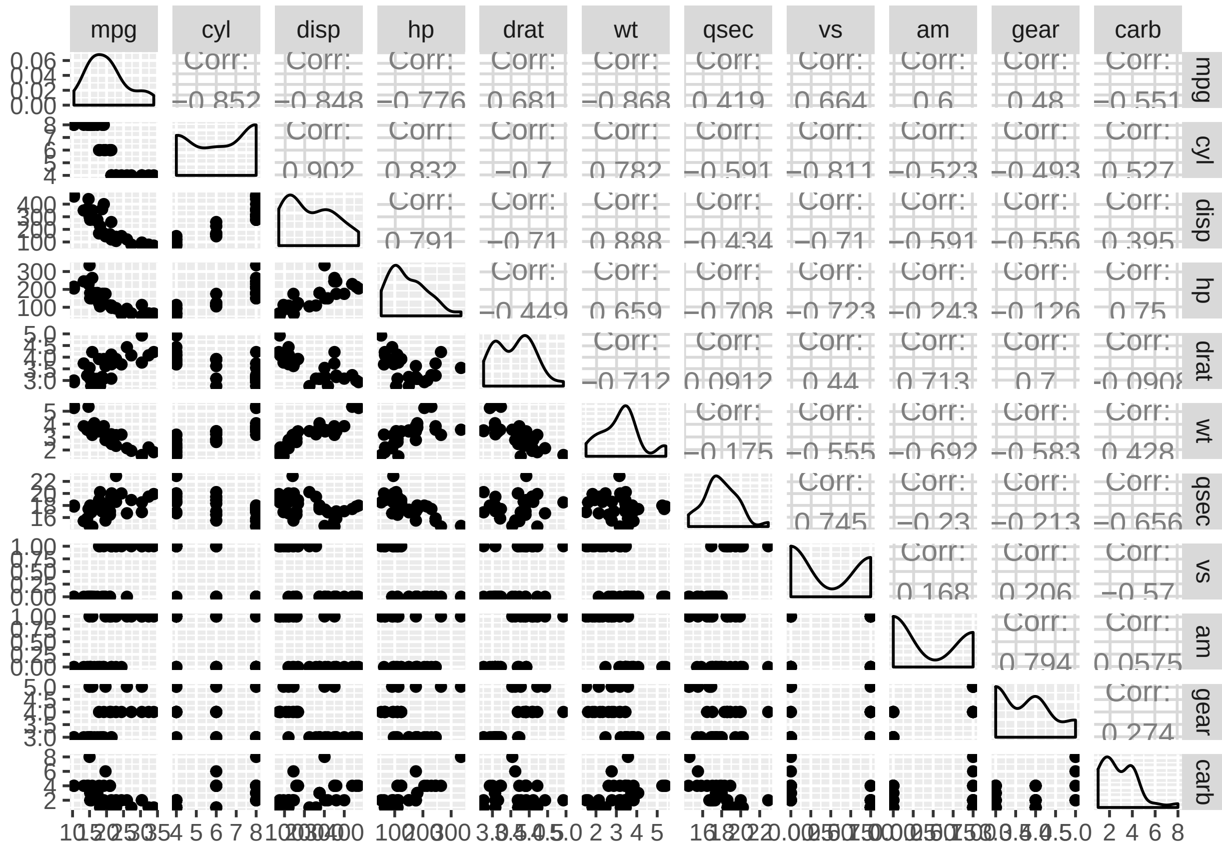
This last plot offers a new angle of vision:

- clear relationship in between weight and mpg.
- the transmission does not seems to change the relationship weigth / mpg
- In general, most of the cars with a weight above 3 (1000 lbs) have automatic transmission, whereas the ones below have a manual transmission.
- The transmission groups (salmon vs light blue) have very few overlapping in weight. At this point, it is then difficult to conclude if automatic or manual transmission better for MPG.

Model

Let's try to see wich variables have the most impact on mpg.

```
library(GGally)
ggpairs(data = mtcars)
```



From the graph above:

- There is a good correlation of mpg with the following variables: wt, cyl, dips, hp and am.
- Some of those variables seems also to be correlated in beetween them.

To answer the question of what variable's selection make sense, we will build several models by adding variables one by one. Then we will analyse the variance (anova) to select the best model.

```
fit1 <- lm(mpg ~ wt, data = mtcars)
fit2 <- lm(mpg ~ wt+cyl, data = mtcars)
fit3 <- lm(mpg ~ wt+cyl+am, data = mtcars)
fit4 <- lm(mpg ~ wt+cyl+am+disp, data = mtcars)
fit5 <- lm(mpg ~ wt+cyl+am+disp+hp, data = mtcars)

anova(fit1,fit2,fit3,fit4,fit5)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ wt
## Model 2: mpg ~ wt + cyl
## Model 3: mpg ~ wt + cyl + am
## Model 4: mpg ~ wt + cyl + am + disp
## Model 5: mpg ~ wt + cyl + am + disp + hp
##   Res.Df    RSS Df Sum of Sq       F    Pr(>F)
## 1     30 278.32
## 2     29 191.17  1    87.150 13.8910 0.0009487 ***
## 3     28 191.05  1     0.125  0.0199 0.8888792
```

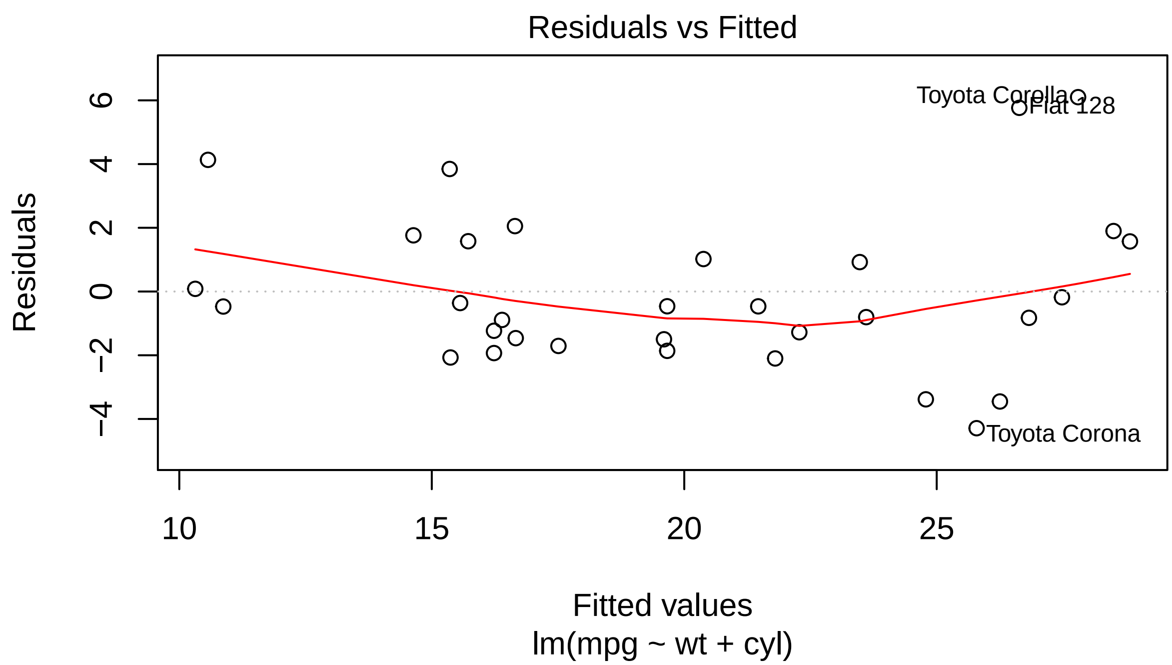
```
## 4     27 188.43  1     2.621  0.4178 0.5236992
## 5     26 163.12  1    25.306  4.0336 0.0550966 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From the above results:

- fit1 to fit2 results in significant reduction in Residual Sum of Square (RSS) and better model fits.
- Adding the transmission (am) in fit3 does not really reduce the RSS. Therefore, this variable might not be relevant here.
- The main parameters seems to be the weight (wt) and the number of cylinder (cyl).

Let's have a look on the residuals:

```
plot(fit2, which=1)
```



No special pattern present. We can also have a deeper look on the model 3 where we had the transmission variable (am).

```
summary(fit3)$coef
```

```
##                Estimate Std. Error    t value     Pr(>|t|)
## (Intercept) 39.4179334  2.6414573 14.9227979 7.424998e-15
## wt          -3.1251422  0.9108827 -3.4308942 1.885894e-03
## cyl         -1.5102457  0.4222792 -3.5764148 1.291605e-03
## am           0.1764932  1.3044515  0.1353007 8.933421e-01
```

From those coefficients:

- We expect 0.176 increase of average mpg from manual transmission to automatic.
- But since the p-value is 0.89 the t-test for $H0 : \beta_{am} = 0$ (transmission has no impact on mpg) versus $H_a : \beta_{am} \neq 0$ (transmission has an impact on mpg) is not significant.