The Influence of Transmission Type on MPG

Andrey Bondarenko

23 May 2015

Executive Summary

Motor Trend magazine is interested in exploring the relationship between a set of variables (characteristics of cars) and miles per gallon (MPG). This report answers three questions:

1. Is an automatic or manual transmission better for MPG?
2. What‘s the quantified difference between automatic and manual transmissions?
3. What‘s the relationship between a set of variables and MPG?

This research uses a `mtcars` dataset.

Data Processing

We need some variables as factors for the analysis:

```
mtcars$cyl  <- factor(mtcars$cyl)
mtcars$vs   <- factor(mtcars$vs)
mtcars$am   <- factor(mtcars$am, labels = c("Automatic", "Manual"))
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
```

Exploratory Data Analysis

Than we could explore some relations in the dataset. Figure 1 in Appendix shows, that manual transmission cars have higher MPG than automatic. It’s evident, that other variables (regressors) also have an influence on MPG. You could see this on Figure 2: for example, the higher displacement a car has, the less miles you can drive. There are also some correlation coefficients on this figure and you could discover, that a weight of a car has more influence on MPG, than a displacement. In fact, a weight is the highest correlated predictor for MPG.

Regression analysis

One of the goals is to investigate the relation between two variables: `mpg` and `am`. This problem could be solved by ***simple linear regression***:

```
fit0 <- lm(mpg ~ am, data = mtcars)
summary(fit0)$coef
```

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

An intercept here is the mean MPG for automatic transmissions (you could see it on Figure 1), and the slope is the mean increase in MPG for manual transmissions. To calculate the 95% confidence interval for the mean MPG difference we use t-statistic:

```
n <- length(mtcars$mpg)
alpha <- 0.05
tstat <- qt(1 - alpha/2, n - 2)
B0 <-   summary(fit0)$coef[2,1]
seB0 <- summary(fit0)$coef[2,2]
B0 + c(-1, 1) * (seB0 * tstat)
```

```
## [1]  3.64151 10.84837
```

Let's turn to ***multivarible regression*** to explore an influence of other predictors. By chosing the model via step-wise model selection (AIC algorithm) we select significant predictors for the analysis:

```
library(MASS)
fit1 <- lm(mpg~., data = mtcars)
fit2 <- stepAIC(fit1, direction = "both")
```

The best model `fit2` that includes `cyl`, `wt` and `hp` as confounders and `am` as the independent variable is shown below:

```
summary(fit2)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## hp          -0.03211    0.01369  -2.345  0.02693 *
## wt          -2.49683    0.88559  -2.819  0.00908 **
## amManual     1.80921    1.39630   1.296  0.20646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

Conclusions

1. A manual transmission is better for MPG.
2. The mean increase in mpg for manual transmission is 7.245 gallons.
3. Number of cyllinders, weigth and horsepower have the biggest influence on MPG: the smaller they are, the greater is MPG. This result is shown on Figure 4.

Appendix

Figure 1. Bar plot.

```
library(ggplot2)
g <- ggplot(mtcars, aes(x = am, y = mpg, color = am)) +
    geom_boxplot(outlier.colour = "green") +
    labs(title = "Miles per Gallon by Transmission Type\n") +
    theme(axis.title = element_blank(), legend.position = 'none',
          title = element_text(size=10))
print(g)
```

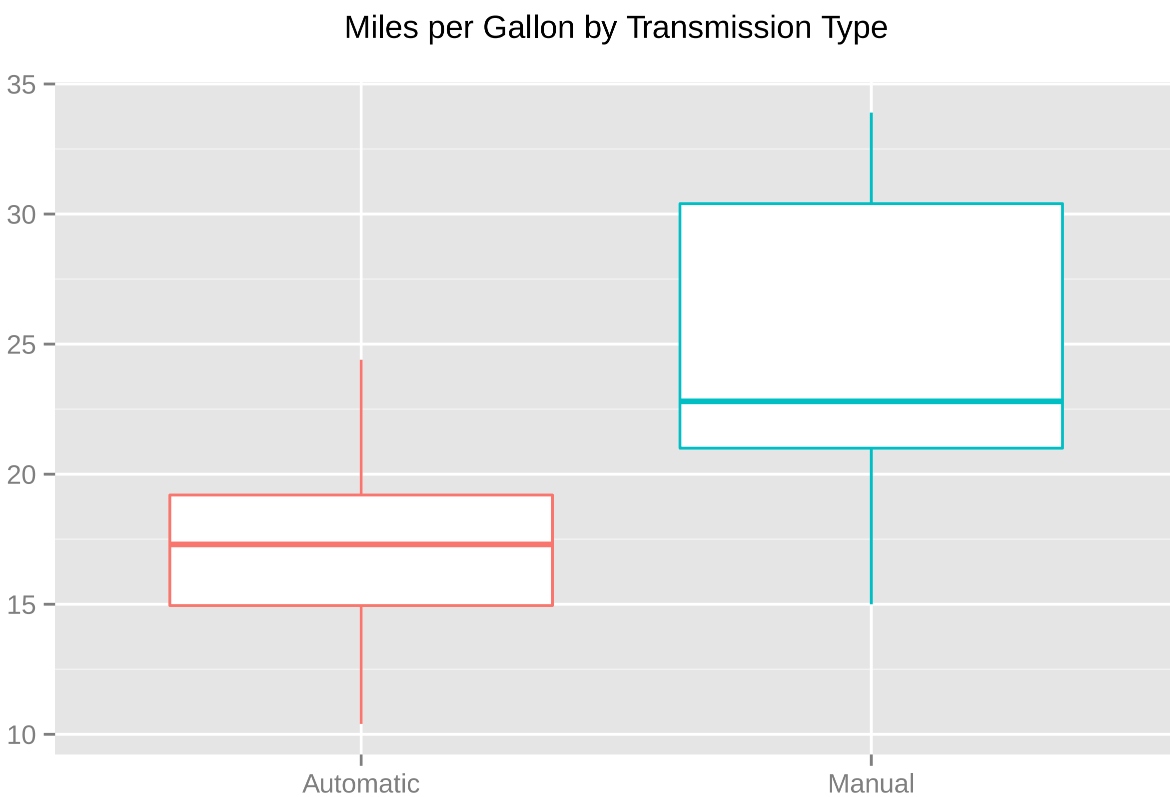

Figure 2. Pairs plot.

```
library(GGally)
g <- ggpairs(mtcars[,c("mpg","cyl","disp","wt")],
             diag = list(continuous = "blank"), title = "Pairs plot")
print(g)
```

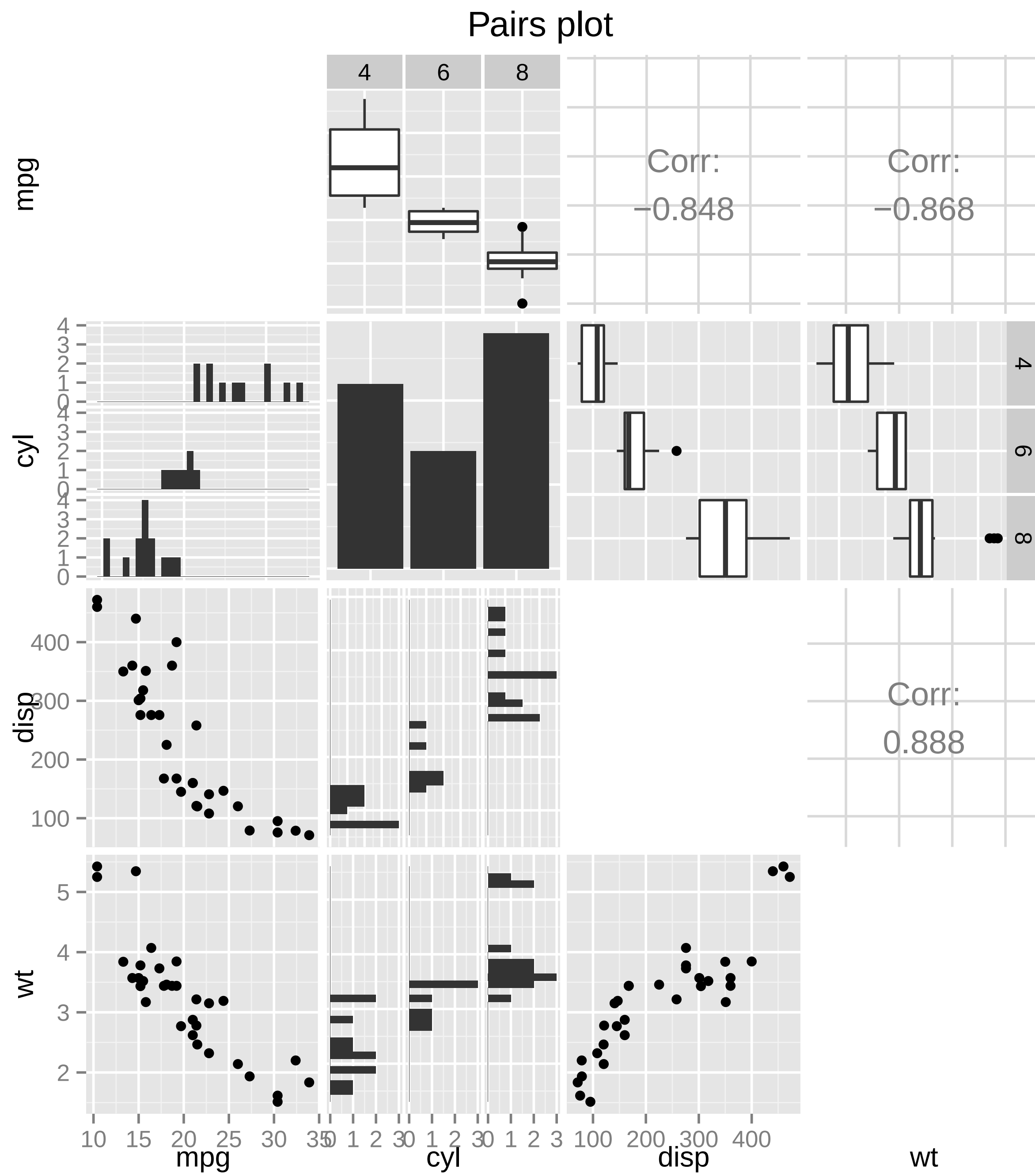

Figure 3. Risidual plots.

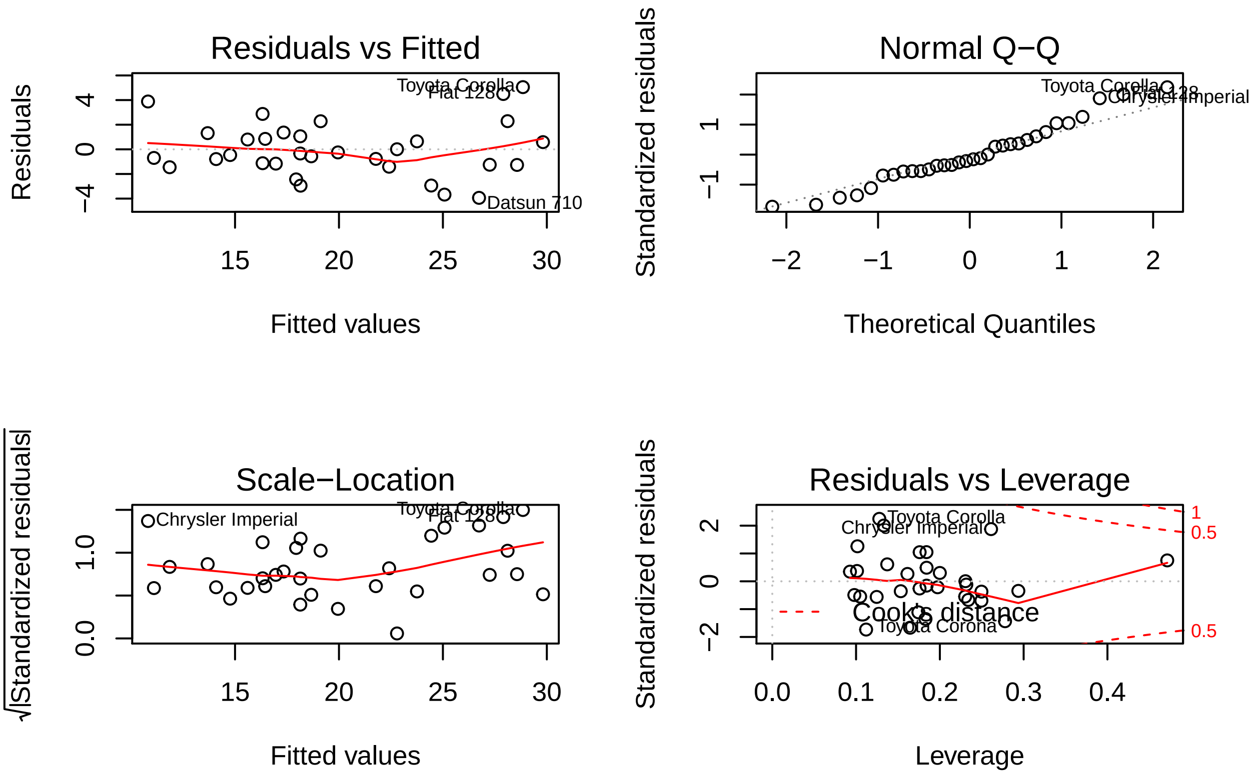

Figure 4. MPG with the most correlated predictors.