Efficiency comparisons between car transmission types

Kevin Tham

May 12, 2018

Overview

In this report we will explore the relationship between a set of variables and miles per gallon (MPG) from a data set of a collection of cars. Specifically, we would like to answer the following two questions:

1. How different is the MPG between automatic and manual transmissions?
2. Is an automatic of manual transmission better for MPG?

Using the dataset `mtcars` we shall embark on a statistical study to address the above two questions.

Exploratary Data Analysis

We begin the study by conducting some exploratory data analysis. First we load in required libraries:

```
if (!require("pacman"))
  install.packages("pacman", repos = "http://cran.us.r-project.org")
pacman::p_load(knitr, dplyr, ggplot2, GGally, tidyr, grid, gridExtra, car)
```

Next we import and examine the dataset:

```
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
```

```
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

Some of the variables are in the wrong data type and require coercion to the correct data type:

```
mtcars$am <- factor(mtcars$am, labels = c('automatic',' manual'))
mtcars$vs <- factor(mtcars$vs, labels = c('V-shaped', 'straight'))
```

```r
mtcars$cyl <- ordered(mtcars$cyl)
mtcars$gear <- ordered(mtcars$gear)
```

We can make a direct comparison between the transmission type and MPG with a boxplot:

From the boxplot in Figure 1 we can conclude that from the dataset, cars with a manual transmission have a larger median MPG than cars with an automatic transmission. The MPG for cars with a manual transmission also appear to have a larger spread between the first and third quartiles.

In order to visualise the relationship of MPG and transmission type with the other variables we can utilise a pairplot, shown in Figure 2. From the pairplot we can observe that many of the variables are fairly correlated with each other. This suggests that it will be difficult to intepret linear regression results to answer question 2 due to confounding variables.

Statistical Analysis

Quantifying the MPG difference between transmission types

For this section we will examine a linear model of MPG `mpg` regressed on transmission type `am` in order to answer question 1. Since we are only concerned with the bulk difference between automatic and manual transmissions, we will not consider other variables here.

```r
fit1 <- lm(mpg ~ am, data=mtcars)
summary(fit1)
```

```
##
## 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 ***
## am manual      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
```

```r
fit2 <- lm(mpg ~ ., data=mtcars[, sample(1:11)])
summary(fit2)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars[, sample(1:11)])
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.2015 -1.2319  0.1033  1.1953  4.3085
##
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 15.73290   16.55442   0.950   0.3539
## qsec         0.76801    0.75222   1.021   0.3201
## am manual    3.34736    2.28948   1.462   0.1601
## hp          -0.05712    0.03175  -1.799   0.0879 .
## carb         0.78703    1.03599   0.760   0.4568
## drat         0.73577    1.98461   0.371   0.7149
## cyl.L        2.16015    3.41523   0.633   0.5346
## cyl.Q        2.22647    1.43687   1.550   0.1378
## gear.L       0.75275    2.14062   0.352   0.7290
## gear.Q       1.25046    1.80855   0.691   0.4977
## vsstraight   2.48849    2.54015   0.980   0.3396
## wt          -3.54512    1.90895  -1.857   0.0789 .
## disp         0.01257    0.01774   0.708   0.4873
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.616 on 19 degrees of freedom
## Multiple R-squared:  0.8845, Adjusted R-squared:  0.8116
## F-statistic: 12.13 on 12 and 19 DF,  p-value: 1.764e-06
```

```
anova(fit2)
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq F value    Pr(>F)
## qsec       1 197.39  197.39 28.8382 3.502e-05 ***
## am         1 576.02  576.02 84.1547 2.073e-08 ***
## hp         1 114.07  114.07 16.6656 0.0006348 ***
## carb       1  14.99   14.99  2.1896 0.1553362
## drat       1  19.95   19.95  2.9145 0.1040761
## cyl        2  13.96    6.98  1.0197 0.3796221
## gear       2  10.34    5.17  0.7556 0.4833506
## vs         1  16.80   16.80  2.4545 0.1336915
## wt         1  29.03   29.03  4.2418 0.0534107 .
## disp       1   3.44    3.44  0.5019 0.4872665
## Residuals 19 130.05    6.84
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
vif(fit2)
```

```
##           GVIF Df GVIF^(1/(2*Df))
## qsec  8.182966  1        2.860588
## am    5.910988  1        2.431252
## hp   21.456428  1        4.632108
## carb 12.681439  1        3.561101
## drat  5.099622  1        2.258234
## cyl  44.446614  2        2.582020
## gear 25.668180  2        2.250861
## vs    7.423472  1        2.724605
## wt   15.800677  1        3.975007
## disp 21.894422  1        4.679148
```

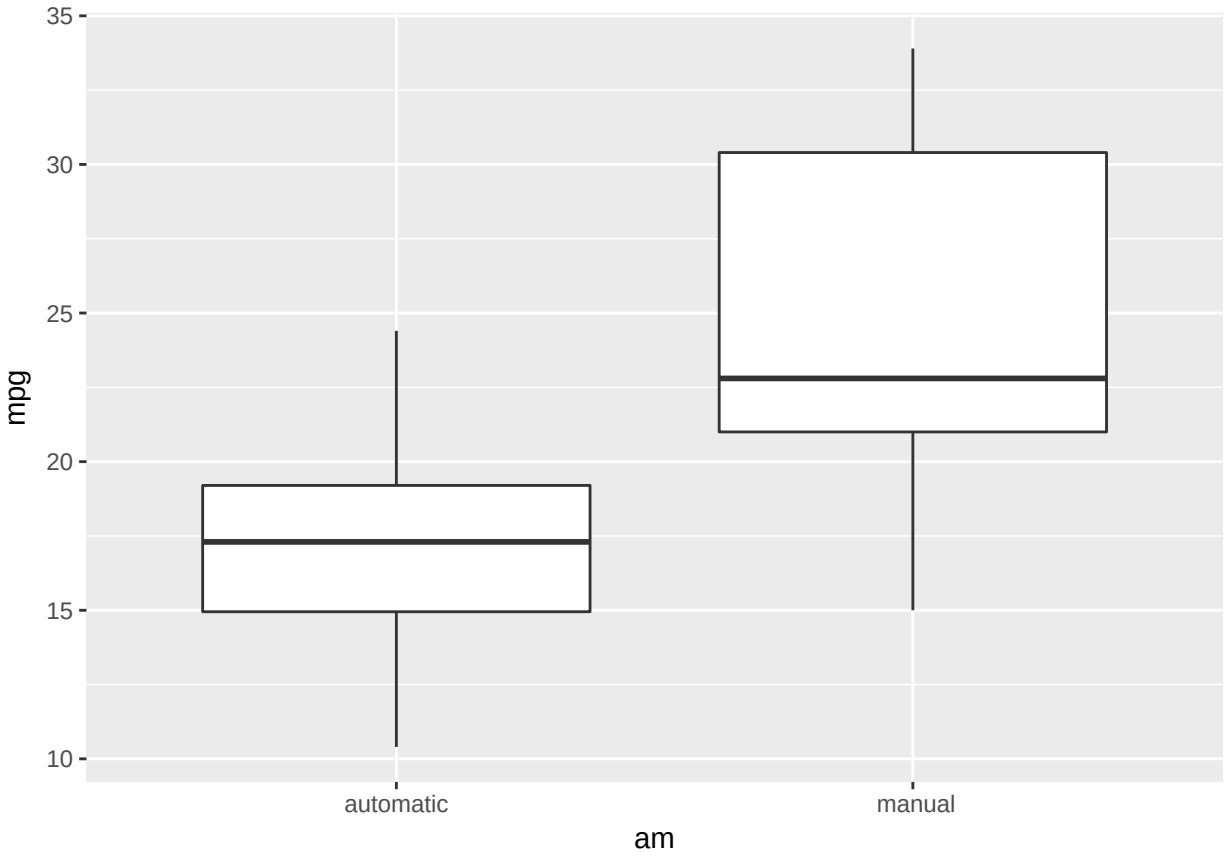

Figure 1: Box plot of MPG against transmission type.

Appendix

```
ggplot(mtcars, aes(x=am,y=mpg)) +
  geom_boxplot()
```

```
ggpairs(mtcars, lower=list(combo=wrap('facethist',binwidth=0.8)))
```

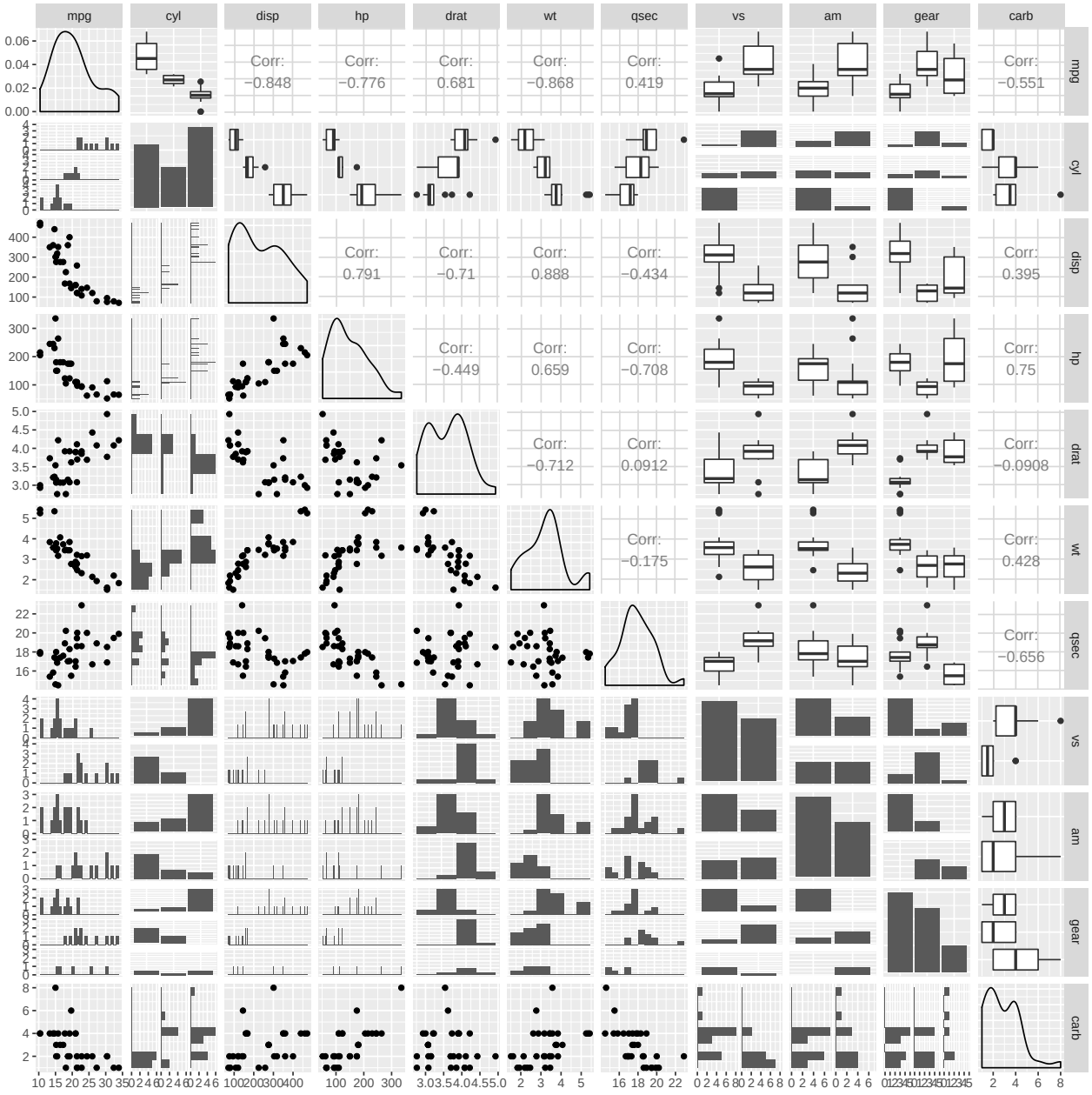

Figure 2: Pair plot of variables from mtcars dataset.