MPG comparisons between car transmission types

Kevin Tham

May 12, 2018

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

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 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 ...
```

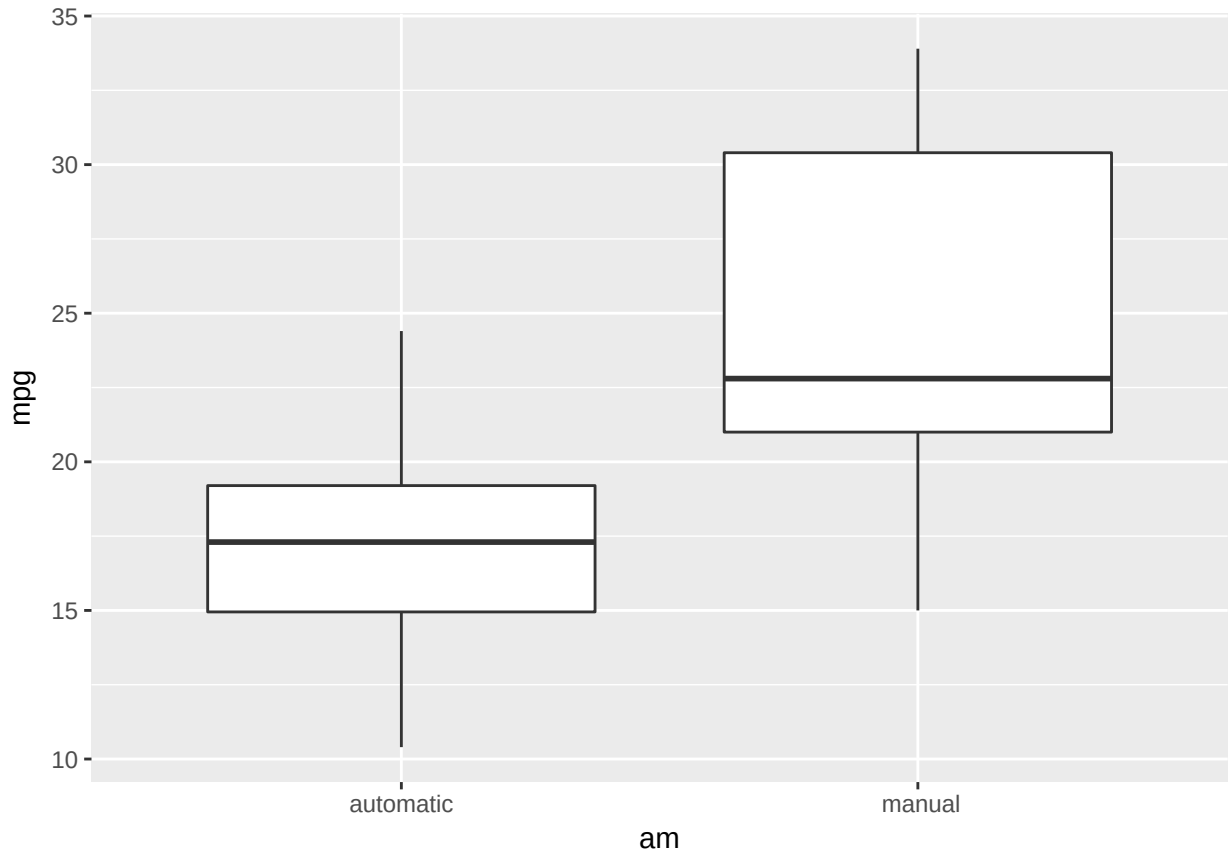
Some of the variables are in the wrong data type and require coercion to the correct data type:

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

```
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:

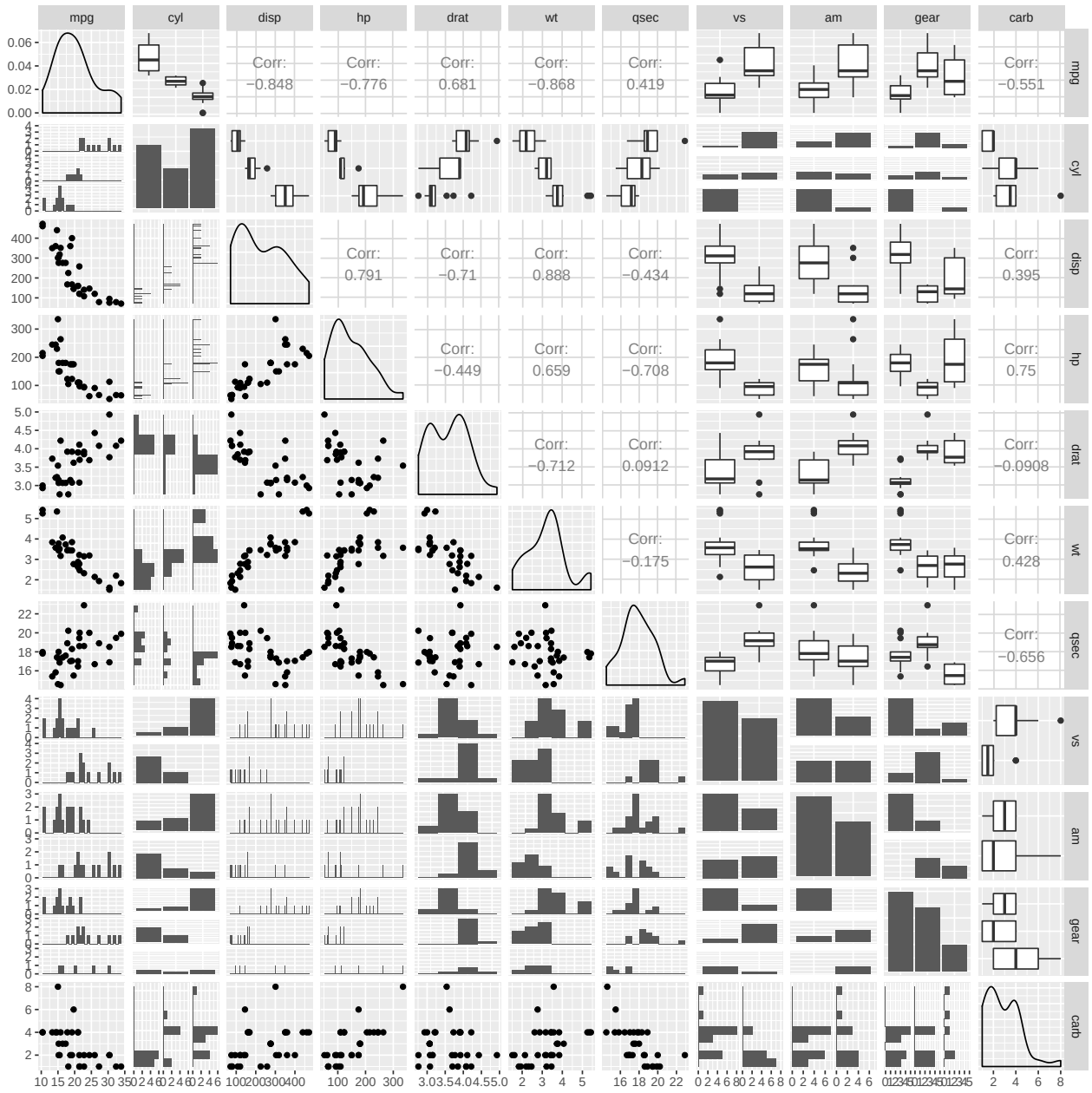
```
ggplot(mtcars, aes(x=am,y=mpg)) +
  geom_boxplot()
```



From the boxplot 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:

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




```
#ggpairs(select(mtcars, mpg, cyl, disp, hp, drat, wt, qsec, am),
#        lower=list(combo=wrap('facethist',binwidth=0.8)))
```

```
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
```

```
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
## drat         0.73577    1.98461   0.371   0.7149
## gear.L       0.75275    2.14062   0.352   0.7290
## gear.Q       1.25046    1.80855   0.691   0.4977
## disp         0.01257    0.01774   0.708   0.4873
## qsec         0.76801    0.75222   1.021   0.3201
## carb         0.78703    1.03599   0.760   0.4568
## hp          -0.05712    0.03175  -1.799   0.0879 .
## vsstraight   2.48849    2.54015   0.980   0.3396
## am manual    3.34736    2.28948   1.462   0.1601
## wt          -3.54512    1.90895  -1.857   0.0789 .
## cyl.L        2.16015    3.41523   0.633   0.5346
## cyl.Q        2.22647    1.43687   1.550   0.1378
## ---
## 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)
## drat       1 522.48  522.48 76.3324 4.421e-08 ***
## gear       2  37.25   18.62  2.7208  0.091396 .
## disp       1 270.88  270.88 39.5752 4.869e-06 ***
## qsec       1  14.89   14.89  2.1759  0.156569
## carb       1  89.84   89.84 13.1248  0.001812 **
## hp         1   1.68    1.68  0.2448  0.626411
```

```
## vs         1   0.12    0.12  0.0173  0.896662
## am         1  17.47   17.47  2.5530  0.126584
## wt         1  24.94   24.94  3.6429  0.071532 .
## cyl        2  16.45    8.22  1.2016  0.322549
## 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))
## drat  5.099622  1        2.258234
## gear 25.668180  2        2.250861
## disp 21.894422  1        4.679148
## qsec  8.182966  1        2.860588
## carb 12.681439  1        3.561101
## hp   21.456428  1        4.632108
## vs    7.423472  1        2.724605
## am    5.910988  1        2.431252
## wt   15.800677  1        3.975007
## cyl  44.446614  2        2.582020
```