

# Investigation of MTCARS Data

Author: Pieter Overdevest

## Executive Summary

Motor Trend, a magazine about the automobile industry is interested in exploring the relationship between car variables and fuel consumption (miles per gallon). The study below suggests that 4 and 6 cylinder manual transmission engines get more miles out of a gallon than then same cylinder automatic transmission engines. Though, for 8 cylinder engines there seems no difference between the two engine transmission types. In fact, 4 cylinder manual transmission engines will take you 5 miles further per gallon than automatic transmission engines with the same number of cylinders.

## Exploratory Data Analyses

First, we open a few libraries and the mtcars dataset. Than we show a few summary statistics,

```
library(ggplot2)
library(plyr)
library(gridExtra)
```

```
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 ...
```

This shows all attributes are of class “numerical”. For further assessment, we convert the levels of the transmission data to “automatic” and “manual”, as this is more informative than 0 and 1 respectively. In addition, the class of the column is converted to factor,

```
mtcars$am[mtcars$am == 0] <- "automatic"
mtcars$am[mtcars$am == 1] <- "manual"

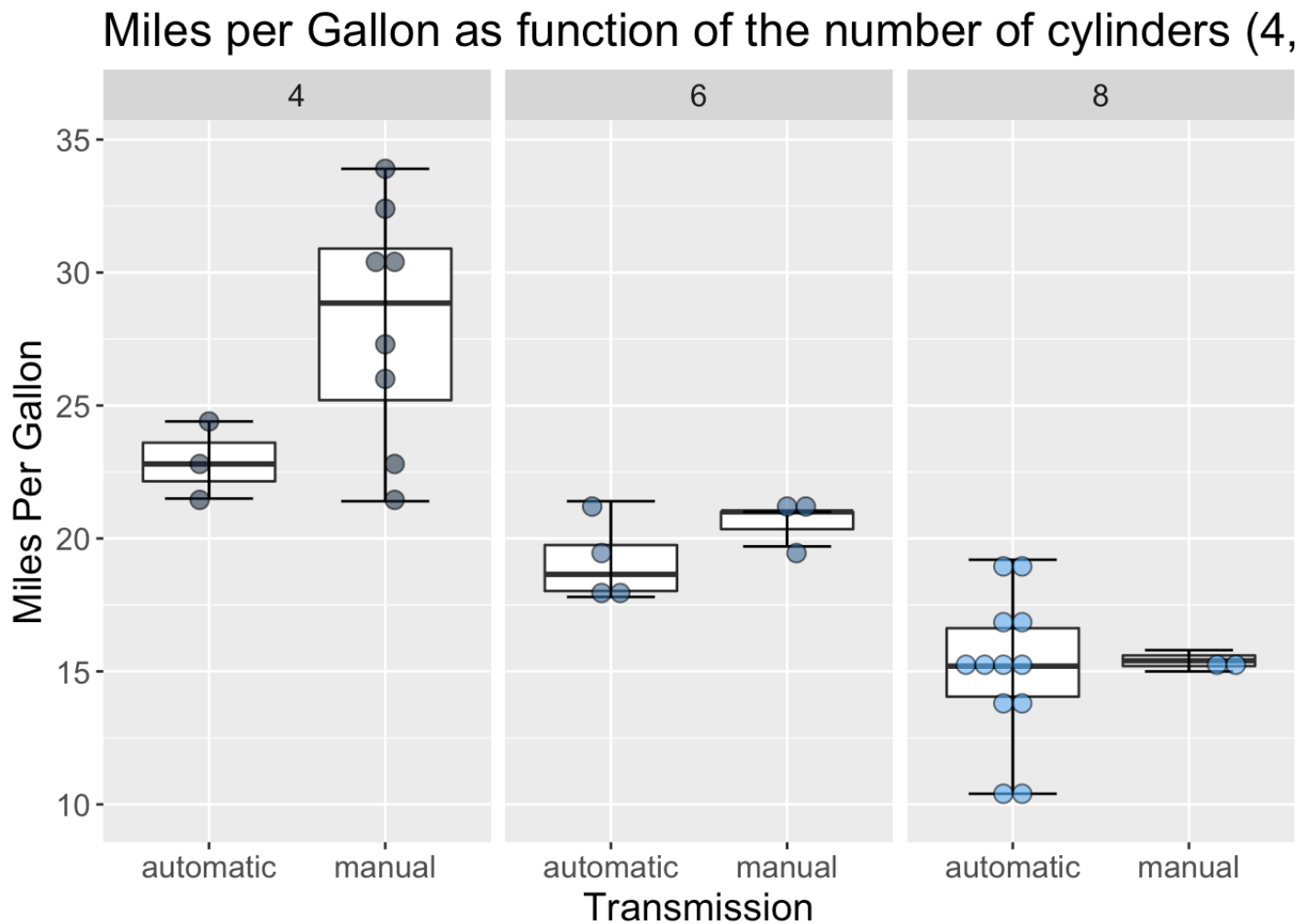
mtcars$am <- as.factor(mtcars$am)
```

Next, we do some further exploratory data analyses on the MPG data vs transmission and the number of cylinders.

```
binwidth <- (max(mtcars$mpg) - min(mtcars$mpg))/20

g <- ggplot(data = mtcars, aes_string(x = "am", y = "mpg")) +
  geom_boxplot() +
  stat_boxplot(geom = "errorbar", width = 0.5) +
  facet_wrap(~ factor(cyl)) +
  labs(x = "Transmission", y = "Miles Per Gallon", title =
    "Miles per Gallon as function of the number of cylinders (4,
6 or 8) and transmission.") +
  theme(text = element_text(size = 15),
    legend.position="none") +
  geom_dotplot(binaxis = "y", stackdir = "center", stackgroups = TRUE,
    binpositions="all", binwidth = binwidth,
    dotsize = 0.6, alpha = 0.6, aes(fill = cyl))

g
```



The boxplots show that manuals transmissions with 4 cylinders get more miles out of a gallon than automatic transmissions with 4 cylinders. The difference between the two transmission types diminishes towards 8 cylinder engines.

## Is an automatic or manual transmission better for MPG?

This will be investigated through linear modeling of the MPG data for engines with automatic vs manual transmissions. In addition, the effect of the number of cylinders will be investigated. It is anticipated that engines with a higher number cylinders will consume more gas per mile than engines with a fewer number of cylinders.

```
fit.automatic <- lm(mpg ~ cyl, data = mtcars[mtcars$am == "automatic",])
fit.manual <- lm(mpg ~ cyl, data = mtcars[mtcars$am == "manual",])

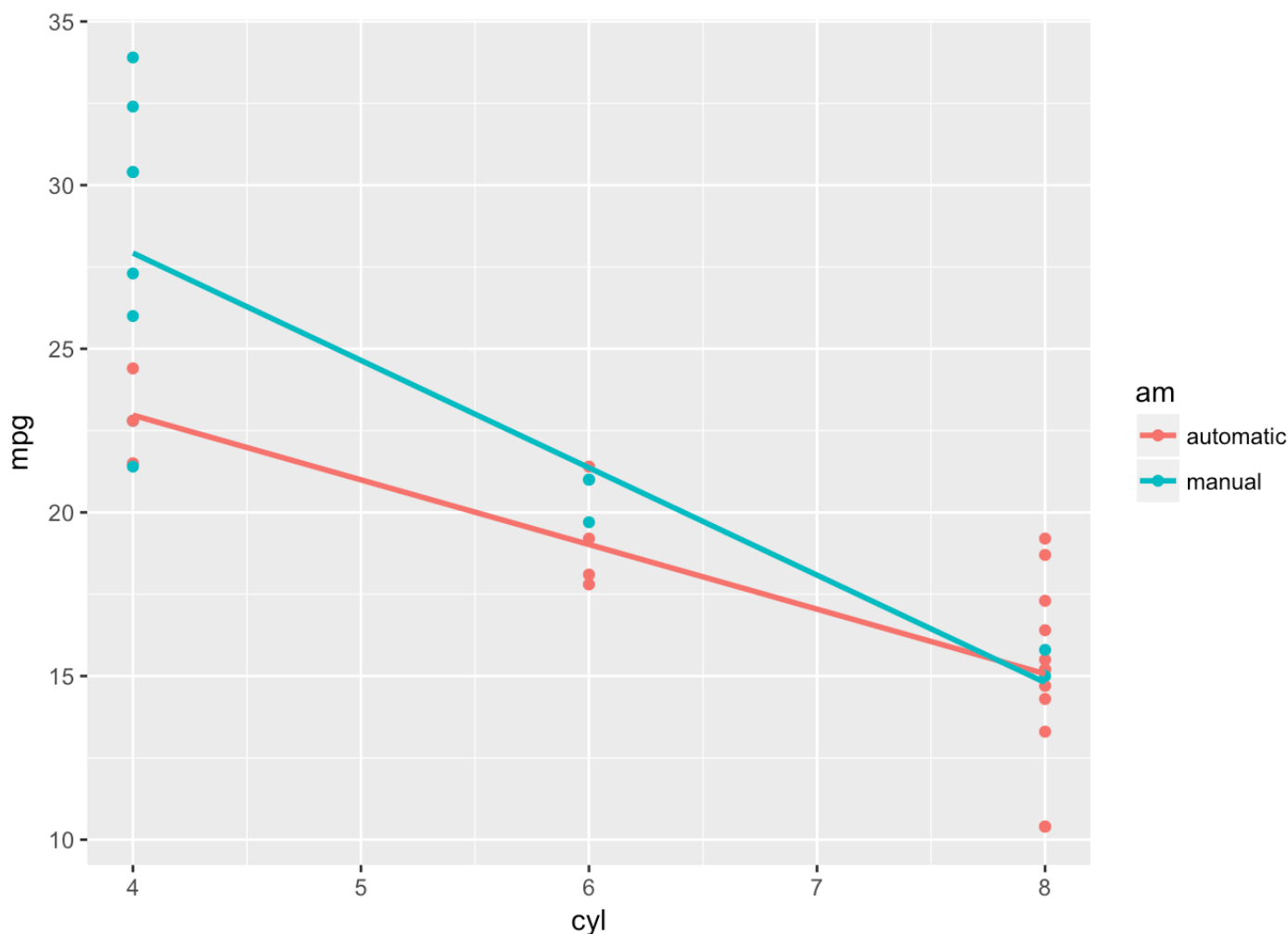
summary(fit.automatic)
```

```
##
## Call:
## lm(formula = mpg ~ cyl, data = mtcars[mtcars$am == "automatic",
##      ])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.6676 -1.0691  0.1324  1.3809  4.1324
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  30.8735      2.5901   11.920 1.11e-09 ***
## cyl         -1.9757      0.3644   -5.422 4.58e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.388 on 17 degrees of freedom
## Multiple R-squared:  0.6336, Adjusted R-squared:  0.6121
## F-statistic: 29.4 on 1 and 17 DF, p-value: 4.576e-05
```

```
summary(fit.manual)
```

```
##
## Call:
## lm(formula = mpg ~ cyl, data = mtcars[mtcars$am == "manual",
##      ])
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.5255 -1.6638 -0.3638  2.4745  5.9745
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  41.0489      3.5720   11.49 1.81e-07 ***
## cyl         -3.2809      0.6751   -4.86 0.000503 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.63 on 11 degrees of freedom
## Multiple R-squared:  0.6823, Adjusted R-squared:  0.6534
## F-statistic: 23.62 on 1 and 11 DF, p-value: 0.0005026
```

```
h <- ggplot(data = mtcars, aes(x = cyl, y = mpg, col = am),
            cex=4, alpha = 0.5, width = 0.05) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE)
h
```



The summary tables and the figure above suggest that at a low number of cylinders in the engine the miles per gallon is higher for manual transmission engines than for automatic transmission engines. Notice, the intercepts are 41 MPG and 31 MPG resp. for a theoretical engine with 0 cylinders. However, as the number of miles per gallon decreases with the number of cylinders, the decrease is steeper for manual transmissions than for automatic transmissions (3.2 MPG per cylinder vs 2.0 MPG per cylinder).

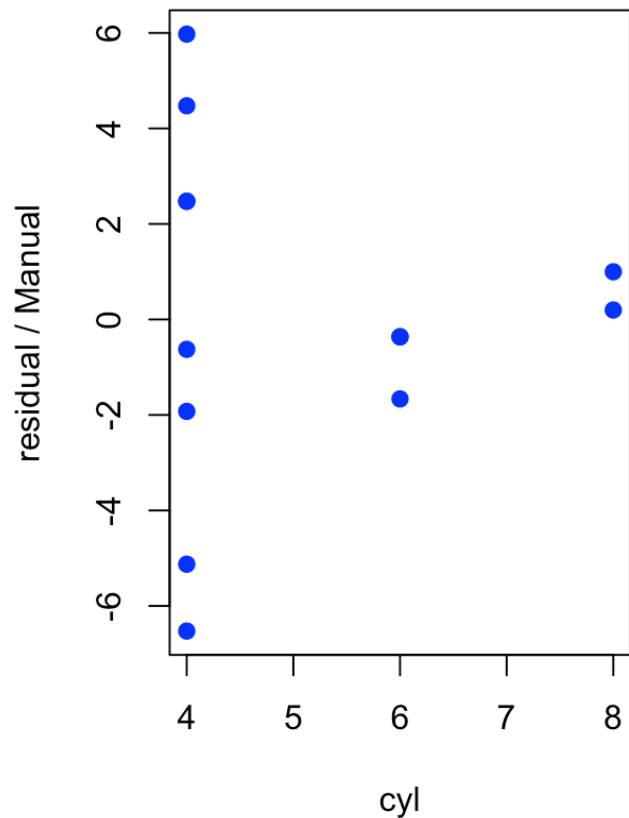
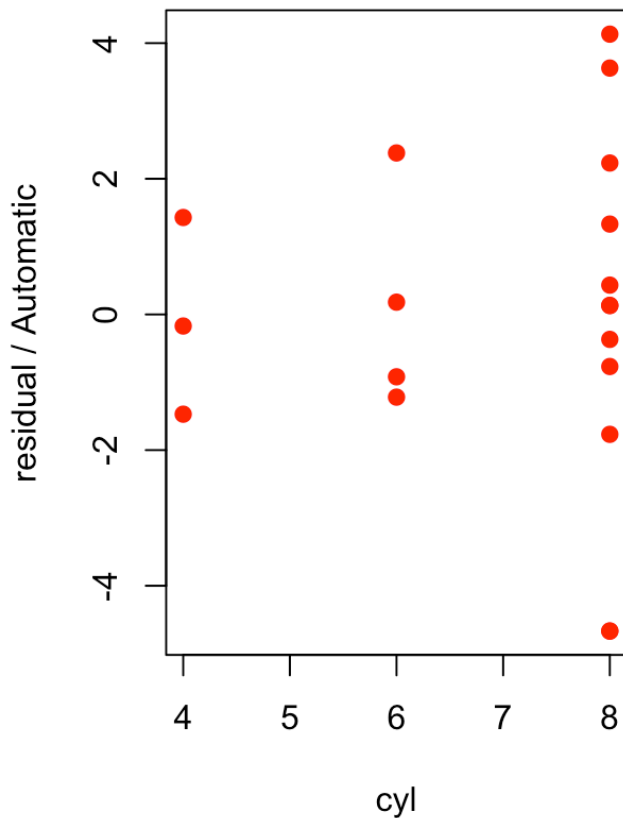
## Quantify the uncertainty

The slopes are significant in both models. The p statistic is smaller than 0.001 in both cases.

## Plot the residuals

Plotting the residuals shows they are reasonably around zero, see figure below,

```
par(mfrow=c(1,2))
plot(mtcars$cyl[mtcars$am=="automatic"], fit.automatic$residuals, pch=19, col="red",
     xlab="cyl", ylab="residual / Automatic")
plot(mtcars$cyl[mtcars$am=="manual"], fit.manual$residuals, pch=19, col="blue", xlab="cyl",
     ylab="residual / Manual")
```



## Quantify the MPG difference between automatic and manual transmissions

Listing the Miles Per Gallon,

```
cbind(mtcars$cyl[mtcars$am=="automatic"], predict.lm(fit.automatic))
```

```
##           [,1]      [,2]
## Hornet 4 Drive      6 19.01912
## Hornet Sportabout   8 15.06765
## Valiant              6 19.01912
## Duster 360          8 15.06765
## Merc 240D           4 22.97059
## Merc 230            4 22.97059
## Merc 280            6 19.01912
## Merc 280C           6 19.01912
## Merc 450SE          8 15.06765
## Merc 450SL          8 15.06765
## Merc 450SLC         8 15.06765
## Cadillac Fleetwood  8 15.06765
## Lincoln Continental  8 15.06765
## Chrysler Imperial   8 15.06765
## Toyota Corona       4 22.97059
## Dodge Challenger    8 15.06765
## AMC Javelin         8 15.06765
## Camaro Z28          8 15.06765
## Pontiac Firebird    8 15.06765
```

```
cbind(mtcars$cyl[mtcars$am=="manual"], predict.lm(fit.manual))
```

```
##           [,1]      [,2]
## Mazda RX4      6 21.36383
## Mazda RX4 Wag  6 21.36383
## Datsun 710      4 27.92553
## Fiat 128        4 27.92553
## Honda Civic     4 27.92553
## Toyota Corolla  4 27.92553
## Fiat X1-9       4 27.92553
## Porsche 914-2   4 27.92553
## Lotus Europa    4 27.92553
## Ford Pantera L  8 14.80213
## Ferrari Dino    6 21.36383
## Maserati Bora   8 14.80213
## Volvo 142E      4 27.92553
```

The numbers in the tables above indicates that on average the MPG difference between manual and automatic transmission engines with 4, 6 and 8 cylinders is resp.,

```
27.92553 - 22.97059
```

```
## [1] 4.95494
```

```
21.36383 - 19.01912
```

```
## [1] 2.34471
```

```
14.80213 - 15.06765
```

```
## [1] -0.26552
```