# Gear Shift Type correlation to Fuel Efficiency from mtcars dataset

*Odin Matanguihan*

*May 10, 2017*

## Objective

Instructions

Determine whether automatic or manual transmission better for MPG.

Quantify the MPG difference between automatic and manual transmissions.

```
#load the data
library(car)
library(dplyr)
library(ggplot2)
library(gridExtra)
data(mtcars)
```

The main item of interest in this data set is the effect between transmission type and fuel efficiency.

```
names(mtcars)
```

```
##  [1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"   "qsec" "vs"   "am"   "gear"
## [11] "carb"
```

Info regarding this data set can be found in this link.

A simple boxplot can show that there is a noticeable difference in fuel efficiency between auto and manual transmission types. The boxplot can be found in plot1 at the bottom.

It would also be helpful to take a look at the pairwise plots of all the variables. It is noteworthy that many of the variables have a clear correlation, they may not be independent variables. We can try to use all the variables and check the variance inflation.

```
vif(lm(mpg~., mtcars))
```

```
##       cyl      disp        hp      drat        wt      qsec        vs
## 15.373833 21.620241  9.832037  3.374620 15.164887  7.527958  4.965873
##        am      gear      carb
##  4.648487  5.357452  7.908747
```

## Selecting best fit model

Due to the high number of correlated variables, and due to the high variance inflation, we need a better model. A step function would try several models by trying different combinations of variables determine a best fit.

```
best <- step(lm(mpg ~ ., mtcars),trace=0)
best$coefficients
```

```
## (Intercept)          wt        qsec          am
##    9.617781   -3.916504    1.225886    2.935837
```

The best model that came from the function shows that the best models for predicting efficiency(mpg) includes just wt, qsec, and am as the predictors. The variance inflation factor is also much lower than before.

```r
vif(best)
```

```
##       wt     qsec       am
## 2.482952 1.364339 2.541437
```

## Quantifiying the difference

We can check the difference in means as well as whether there is a true difference in means.

```r
a <- t.test(mtcars$mpg~mtcars$am,conf.level=0.95)
diff <- a$estimate[2] - a$estimate[1]
cat("Automatic transmission has better mileage with a mean diff of",
    + diff, "miles per gallon.")
```

```
## Automatic transmission has better mileage with a mean diff of 7.244939 miles per gallon.
```

```r
cat("The difference has a p-value of", a$p.value, ".")
```

```
## The difference has a p-value of 0.001373638 .
```

With such a low p-value, we reject the null hypothesis that there is no true difference in means.

## Checking for outliers

Next we can do a residual analysis to see if there is any high leverage or high influence data that could skew the model one way or another.

```r
mtcars.resid <- resid(step(lm(mpg ~ ., mtcars),trace=0))
```

The graph of the residuals can be found be found here. There is no obviously high influence or high leverage value. The scale on the graph does show that manual and automatic transmission vehicles on this data set are on different weight ranges.

## Summary

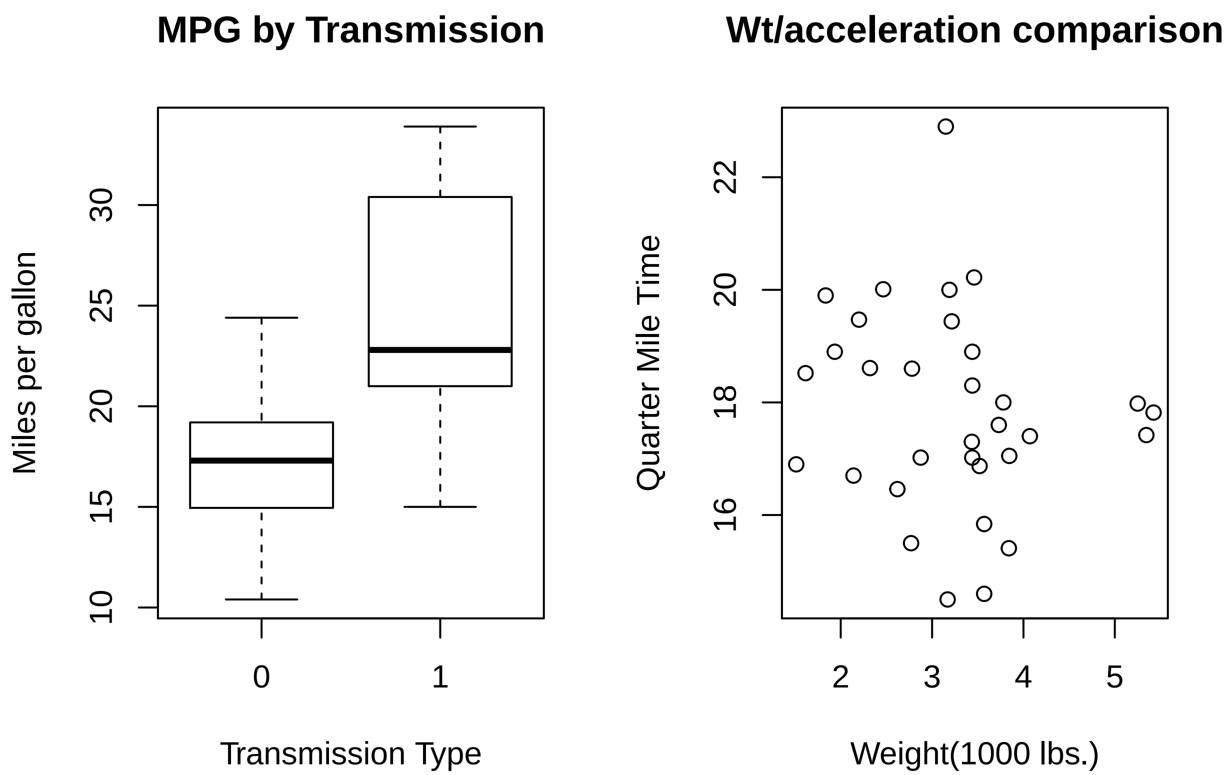
This analysis of the dataset shows that automatic transmission does have better mileage than manual transmission. Considering that weight is the biggest predictor, and that automatic transmission is on a different weight range from manual transmission, it cannot be used as a strong predictor and may just be an incidental factor.

## Plots

### Boxplot MPG by Transmission Type

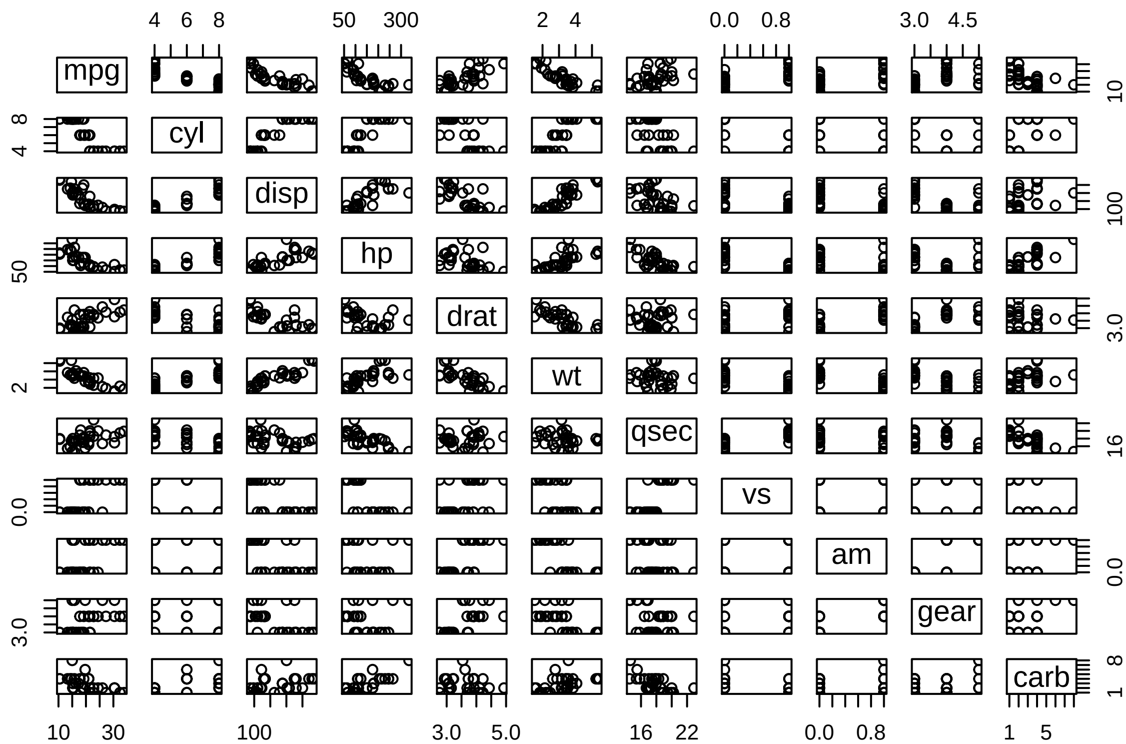
```
par(mfcol=c(1,2))

boxplot(mpg ~ am, data = mtcars, xlab = "Transmission Type", ylab = "Miles per gallon",
        main="MPG by Transmission")
plot(mtcars$wt, mtcars$qsec, main = "Wt/acceleration comparison", xlab = "Weight(1000 lbs.)",
      ylab = "Quarter Mile Time")
abline(mtcars$wt, mtcars$qsec)
```



## Pairwise plots of variables

```r
pairs(mpg ~ ., data = mtcars)
```



## Residual plots

```
x <- mtcars$am == 1
p1 <- ggplot(mtcars[x,], aes(x=wt, y=qsec)) + geom_point(shape=1) + geom_smooth()
p1 <- p1 + xlab("wt(in tons) - auto transmission")
p2 <- ggplot(mtcars[!x,], aes(x=wt, y=qsec)) + geom_point(shape=1) + geom_smooth()
p2 <- p2 + xlab("wt(in tons) - manual transmission")
p3 <- ggplot(mtcars[x,], aes(x=wt, y=mtcars.resid[x])) + geom_point(shape=1)
p3 <- p3 + geom_smooth()  + ylab("residuals") + xlab("wt(in tons) - auto transmission")
p4 <- ggplot(mtcars[!x,], aes(x=wt, y=mtcars.resid[!x])) + geom_point(shape=1)
p4 <- p4 + geom_smooth() + ylab("residuals")+ xlab("wt(in tons) - manual transmission")
p5 <- ggplot(mtcars[x,], aes(x=qsec, y=mtcars.resid[x])) + geom_point(shape=1)
p5 <- p5 + geom_smooth()   + ylab("residuals") + xlab("qsec(auto transmission)")
p6 <- ggplot(mtcars[!x,], aes(x=qsec, y=mtcars.resid[!x])) + geom_point(shape=1)
p6 <- p6 + geom_smooth() + ylab("residuals") + xlab("qsec(manual transmission)")
grid.arrange(p1, p2, p3, p4, p5, p6, ncol =2, nrow=3)
```