Effict of Transmission Type on MPG

Introduction

In this report aim to explore the relationship between set of variables and miles per gallon (MPG) (outcome).

Data Processing

Some data has to be prepared

```
library(ggplot2)
data(mtcars)

mcars <- mtcars
mcars[mcars$am == 0, c("tm")] <- gsub("0", "automatic", mcars[mcars$am == 0,
    c("am")])
mcars[mcars$am == 1, c("tm")] <- gsub("1", "manual", mcars[mcars$am == 1, c("am")])
mcars$tm <- as.factor(mcars$tm)

auto <- subset(mcars, am == 0)
manual <- subset(mcars, am == 1)
```

variables correlation

First we take a look at variables correlation with MPG column

```
sort(abs(cor(mtcars)[, c("mpg")]), decreasing = T)
```

```
##   mpg    wt   cyl  disp    hp  drat    vs    am  carb  gear
## 1.0000 0.8677 0.8522 0.8476 0.7762 0.6812 0.6640 0.5998 0.5509 0.4803
##   qsec
## 0.4187
```

Exploratory Analysis

Exploratory Analysis had been done and showed relations between variables Check the appendix

Model selection

Attempt to fit a model with the most correlated variables

```
carsfit <- lm(mpg ~ tm + as.factor(cyl) + wt + disp + hp - 1, data = mcars)
```

Now We Exclude Number of Cylenders and dispalcement variables because they are not statistically segnificant.

and fit a model with significant values.

```
carsfit <- lm(mpg ~ tm + wt + hp - 1 + tm * wt * hp, data = mcars)
```

But that didn't worked, it has many many insignificant values.

So, let's try it with less details.

```
carsfit <- lm(mpg ~ tm + wt + hp - 1, data = mcars)
summary(carsfit)
```

```
##
```

```
## Call:
## lm(formula = mpg ~ tm + wt + hp - 1, data = mcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.422 -1.792 -0.379  1.225  5.532
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## tmautomatic 34.00288    2.64266   12.87  2.8e-13 ***
## tmmanual    36.08659    1.73634   20.78  < 2e-16 ***
## wt          -2.87858    0.90497   -3.18  0.00357 **
## hp          -0.03748    0.00961   -3.90  0.00055 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.54 on 28 degrees of freedom
## Multiple R-squared: 0.987,   Adjusted R-squared: 0.985
## F-statistic:  538 on 4 and 28 DF,  p-value: <2e-16
```

Now we got a sutable model.

Conclusion

The Answer

From the model we can say that manual transmission cars are little bit better for MPG by an average of **2.0837** miled per US galon.

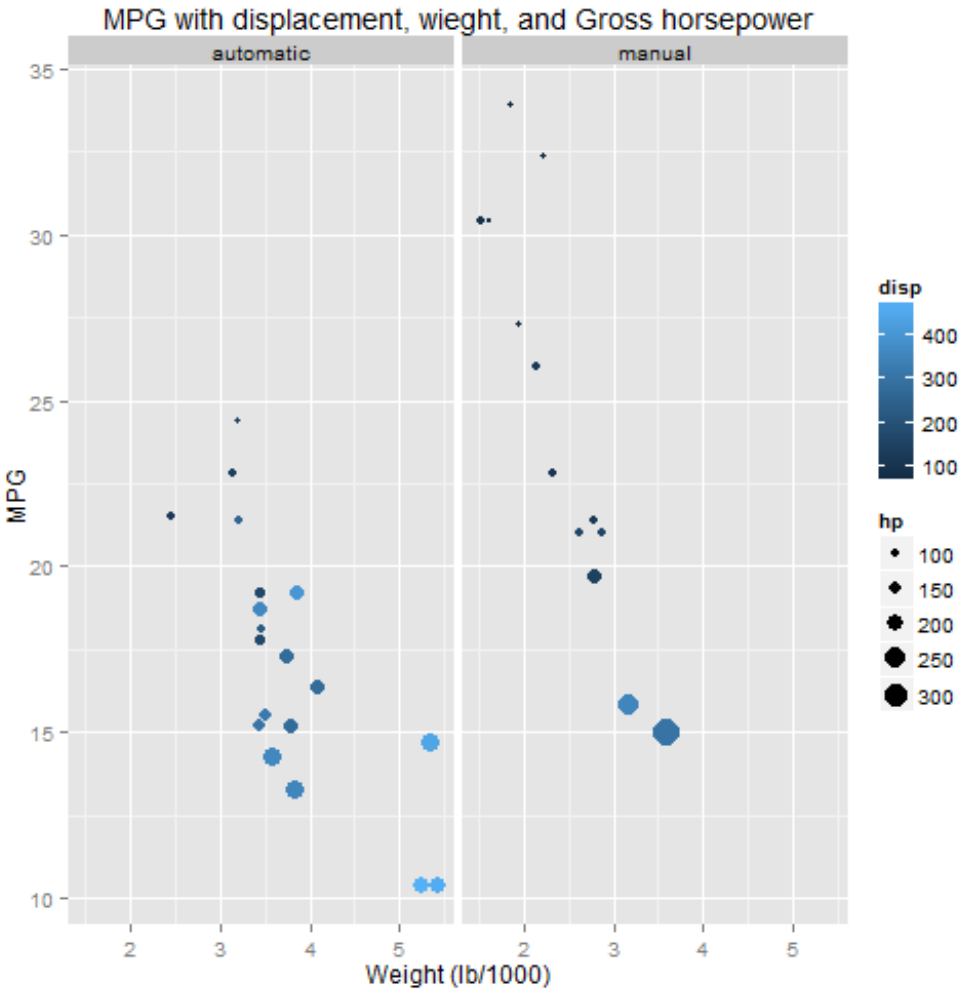
Model Uncertainty

The model used is very statistically significant one, with 0.9853 Adjusted R-squared. the data given is too small data and that is why we couldn't build a model with more details and variables.

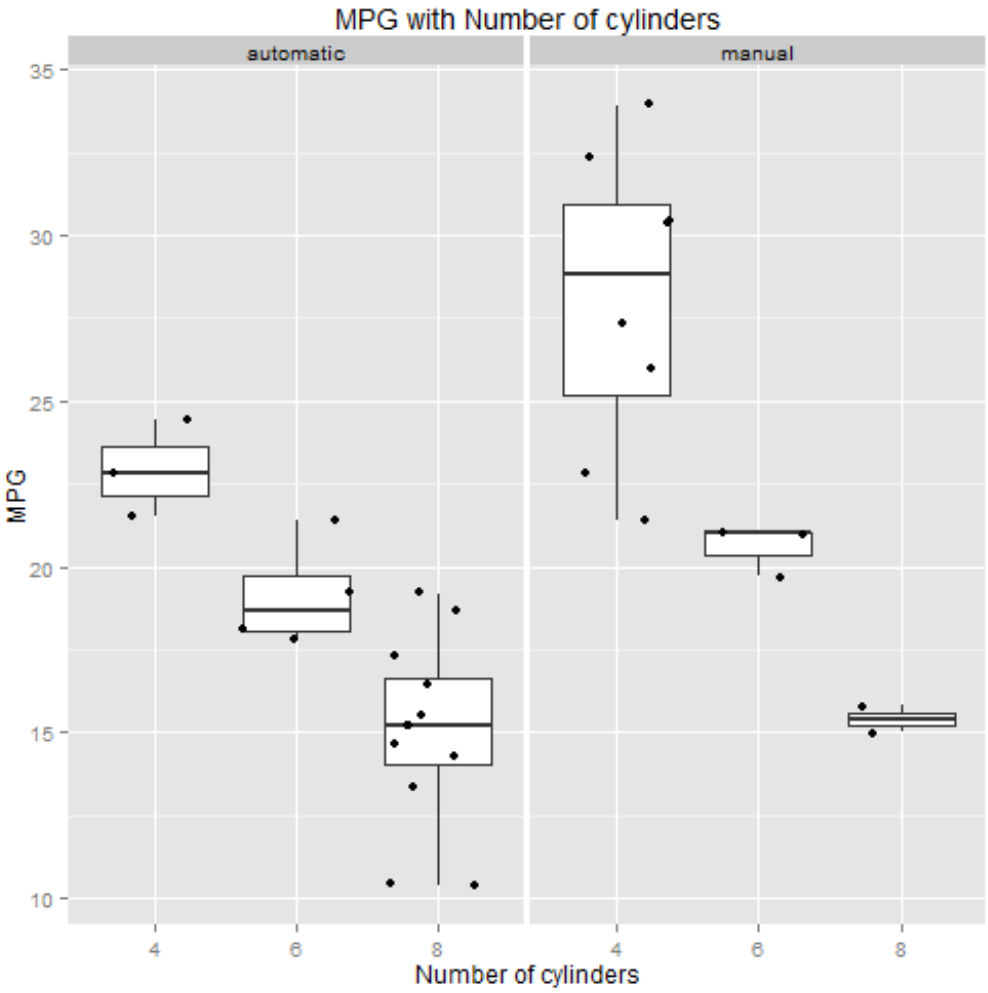
Appendix

Exploratory analysis Plots

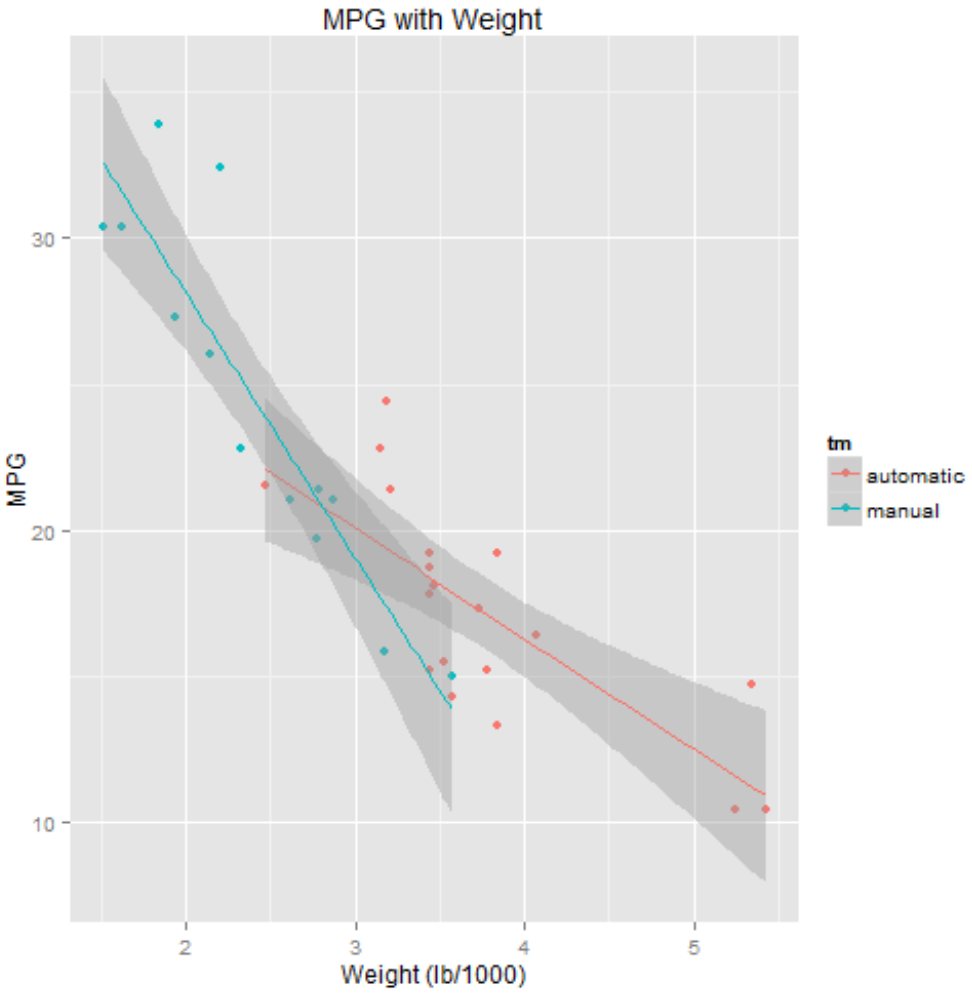
```
qplot(y = mpg, x = wt, data = mcars, colour = disp, size = hp, facets = . ~
    tm, geom = c("point"), ylab = "MPG", xlab = "Weight (lb/1000)", main = "MPG with displacement, wieght, and Gross
horsepower")
```



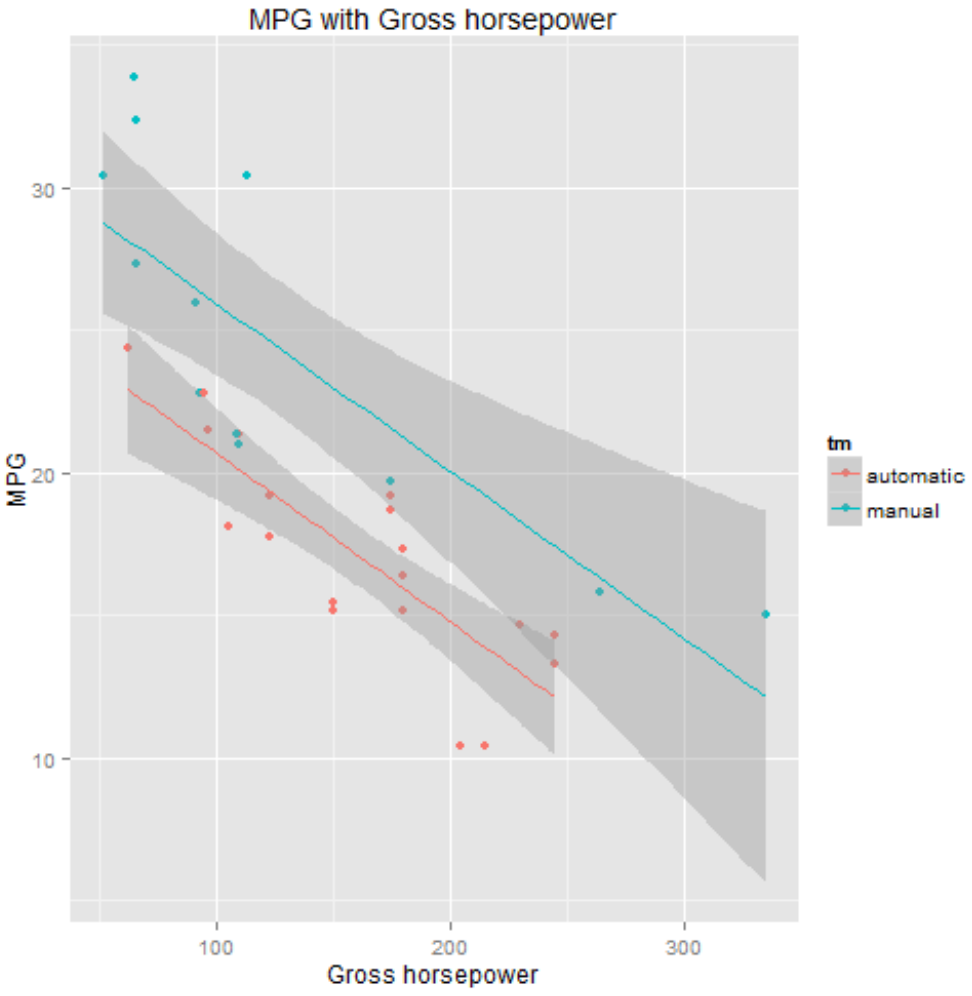

```
qplot(y = mpg, x = as.factor(cyl), data = mcars, geom = c("boxplot", "jitter"),
   facets = . ~ tm, main = "MPG with Number of cylinders", ylab = "MPG", xlab = "Number of cylinders")
```




```
qplot(y = mpg, x = wt, data = mcars, color = tm, geom = c("point", "smooth"),
   method = "lm", main = "MPG with Weight", ylab = "MPG", xlab = "Weight (lb/1000)")
```



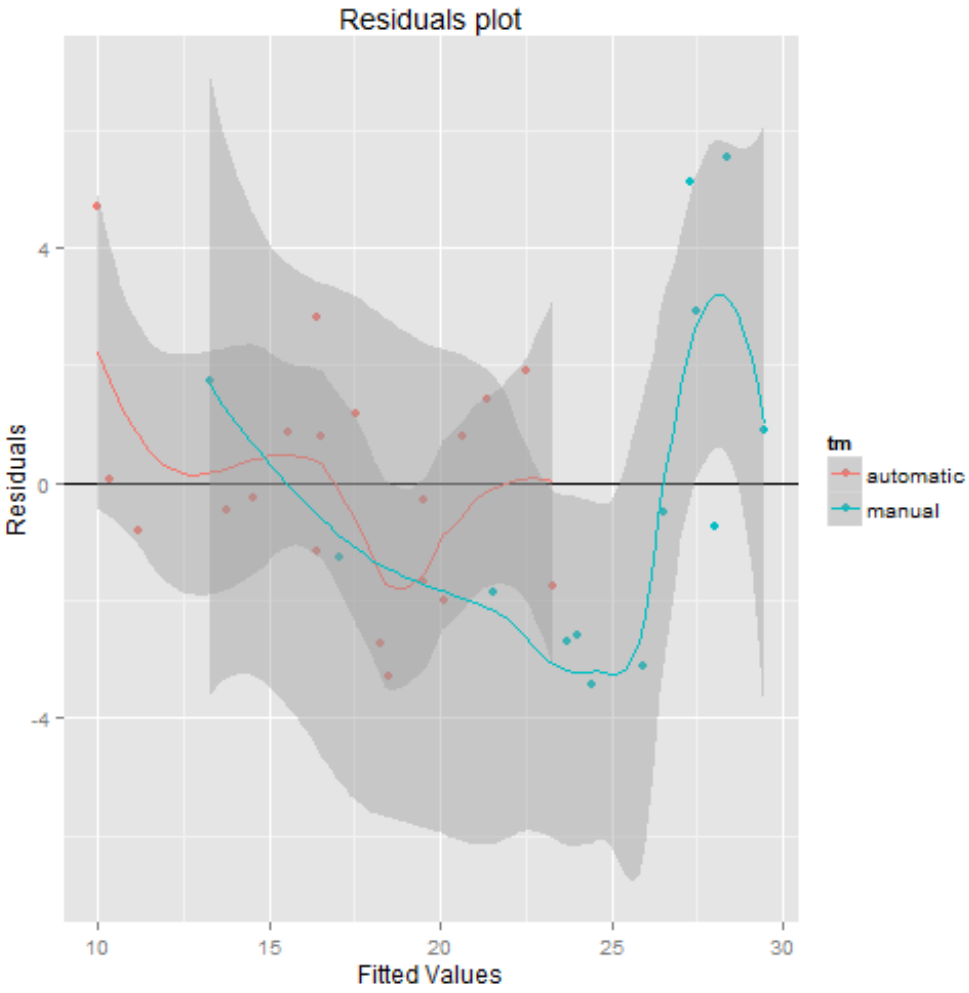

```
qplot(y = mpg, x = hp, data = mcars, color = tm, geom = c("point", "smooth"),
   method = "lm", main = "MPG with Gross horsepower", ylab = "MPG", xlab = "Gross horsepower")
```



Rsiduals

```
qplot(y = resid(carsfit), x = carsfit$fitted.values, data = mcars, color = tm,
   xlab = "Fitted Values", ylab = "Residuals", main = "Residuals plot") + geom_abline(slope = 0,
   intercept = 0) + geom_smooth()
```

```
## geom_smooth: method="auto" and size of largest group is <1000, so using loess. Use 'method = x' to change the smoothing method.
```


Residuals plot
Residuals
4
0
-4
10
15
20
25
30
Fitted Values
tm
automatic
manual