What gives better mpg, Automatic or Manual transmission?

Nitin Aggarwal

September 27, 2015

Executive Summary

To answer the dependence of mpg on Automatic vs Manual transmission, `mtcars` data is studied. After exploration a few models are derived and a best model is chosen from which it was found that the dependence of `mpg`, miles/gallon for different cars is confounded on `hp`, `wt`, and `cyl`, i.e. the horsepower, weight, and number of cylinders. The final linear models shows that the Manual cars have **1.809MPGS** more than that of Automatic cars.

Introduction

We are working with the `mtcars` data which can be loaded through `datasets` as follows:

```
data("mtcars")
```

Let us also convert the qualitative features as factor variables.

A call to the R help shows the following description to the data:

The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 models).

Goal As the title suggests the main goal of this analysis is to study the effect of type of transmission, automatic vs manual, on mpg.

Exploration

In **Appendix 1** we have a sequence of plots ploting `mpg` against other features with respect to transmission type.

Let us also get the correlations of different features present in `mtcars` as follows:

```
ind<-sapply(1:11,function(i)!is.factor(mtcars[1:nrow(mtcars),i]))
mtcars.corr<-cor(mtcars[,ind])
mtcars.corr
```

```
##         mpg   disp     hp    drat     wt    qsec    carb
## mpg   1.000 -0.848 -0.776  0.6812 -0.868  0.4187 -0.5509
## disp -0.848  1.000  0.791 -0.7102  0.888 -0.4337  0.3950
## hp   -0.776  0.791  1.000 -0.4488  0.659 -0.7082  0.7498
## drat  0.681 -0.710 -0.449  1.0000 -0.712  0.0912 -0.0908
## wt   -0.868  0.888  0.659 -0.7124  1.000 -0.1747  0.4276
## qsec  0.419 -0.434 -0.708  0.0912 -0.175  1.0000 -0.6562
## carb -0.551  0.395  0.750 -0.0908  0.428 -0.6562  1.0000
```

From these one can infer that `mpg` has high level of positive or negative correlation with features in columns `disp`, `hp`, `wt`, `dart`, `carb`. Of these `disp`, `quesc`, `carb` shows high correlation with `hp`, giving a possibility of them being redundant. Also, from the graphs in **Appendix 1** we notice that among the factor columns, `cyl`, `vs`, and `gear`, all of them are shown to affect the `mpg`, but the columns `cyl` and `gear` are also showing interaction with `am` column.

Linear Regression

We will fit the following models we found from our analysis, and choose the best of them.

1. `mpg` against just `am`
2. `mpg` against `am`, `cyl`
3. `mpg` against `am`, `cyl`, `gear`
4. `mpg` against interactions between `am` and `cyl`
5. `mpg` against interactions between `am` and `gear`
6. `mpg` against interactions between `am` and `cyl`, and between `am` and `gear`
7. `mpg` against `hp` with best of above models.
8. above model with `wt` included

Let us create the first six models first:

Let us choose the best of top three first.

```
anova(fit[[1]], fit[[2]], fit[[3]])
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + cyl
## Model 3: mpg ~ am + cyl + gear
##   Res.Df RSS Df Sum of Sq     F  Pr(>F)
## 1     30 721
## 2     28 264  2       456 23.02 1.8e-06 ***
## 3     26 258  2         7  0.34    0.72
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Choosing the second model, let us compare it with the fourth and sixth model, ignoring fifth because of absense of `cyl` in it.

```
anova(fit[[2]], fit[[4]], fit[[6]])
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am + cyl
## Model 2: mpg ~ am * cyl
## Model 3: mpg ~ am * cyl + am * gear
```

```
##   Res.Df RSS Df Sum of Sq    F Pr(>F)
## 1     28 264
## 2     26 239  2     25.44 1.28   0.30
## 3     24 239  2      0.19 0.01   0.99
```

From this result because of high p values we keep the second of above six model and compare it with the updated by `hp` and then that with `wt`. These models are tested against the second model above in one line as follows:

```r
anova(fit[[2]],
      update(fit[[2]], .~. + hp),
      update(fit[[2]], .~. + hp + wt))
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am + cyl
## Model 2: mpg ~ am + cyl + hp
## Model 3: mpg ~ am + cyl + hp + wt
##   Res.Df RSS Df Sum of Sq     F Pr(>F)
## 1     28 264
## 2     27 197  1      67.3 11.59 0.0022 **
## 3     26 151  1      46.2  7.95 0.0091 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

From this we choose the third of these models. Following presents the details about this model:

```r
fitf<-update(fit[[2]], .~. + hp + wt)
fitf
```

```
##
## Call:
## lm(formula = mpg ~ am + cyl + hp + wt, data = mtcars)
##
## Coefficients:
## (Intercept)          am1         cyl6         cyl8           hp
##     33.7083       1.8092      -3.0313      -2.1637      -0.0321
##          wt
##     -2.4968
```

Let us now quantify this plot using a diagnosis using residuals:

Residuals and diagnositcs

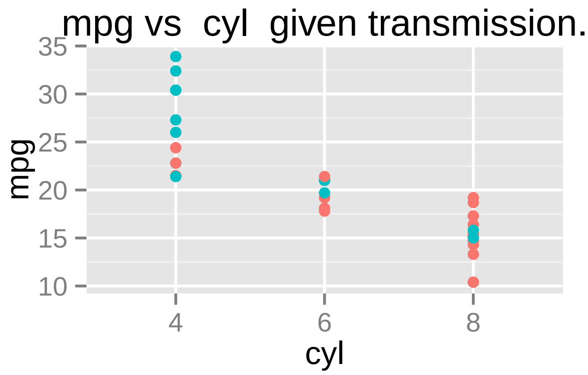
The residuals and the Normal Q-Q plot in `Appendix 2` shows that the data is free of heteroscedasticity and is normal.

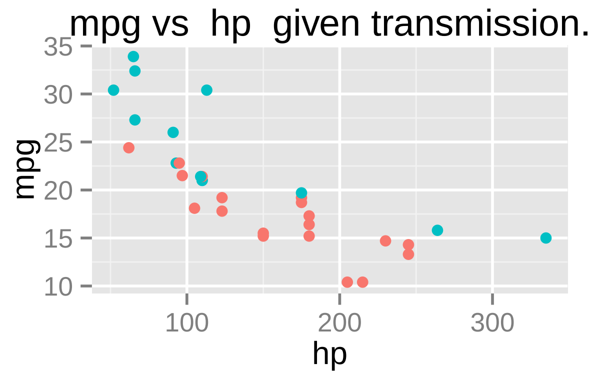
Conclusion

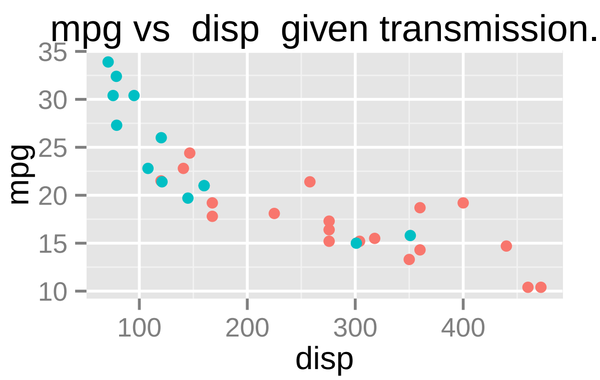
We get following two results from our analysis:

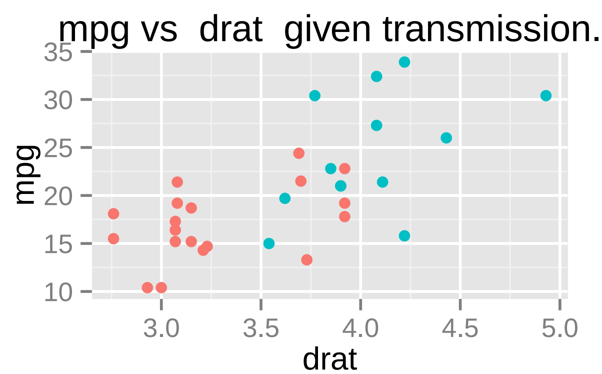
1. Manual transmission gives more mileage per gallon given the fixed values of cyl, hp, and wt.
2. Manual transmission increases the mileage by 1.809 miles per gallon.

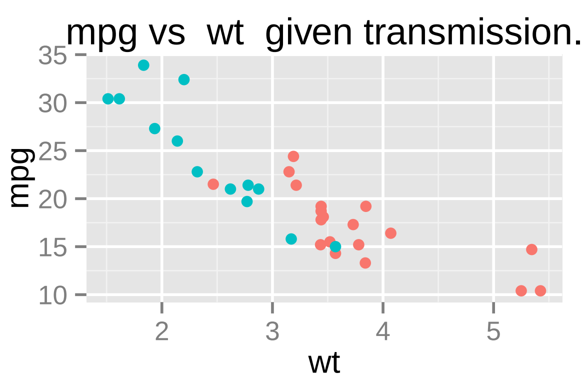
Appendix 1

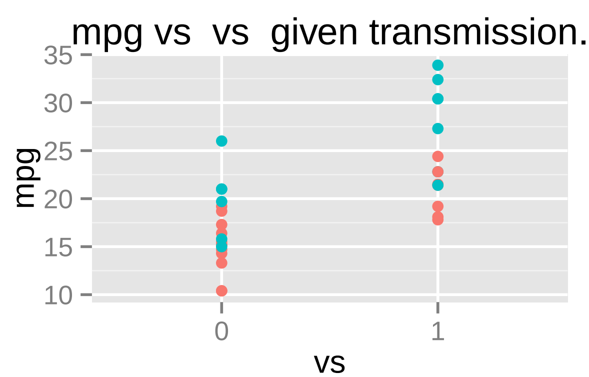

mpg vs cyl given transmission.
mpg
35
30
25
20
15
10
4
6
8
cyl

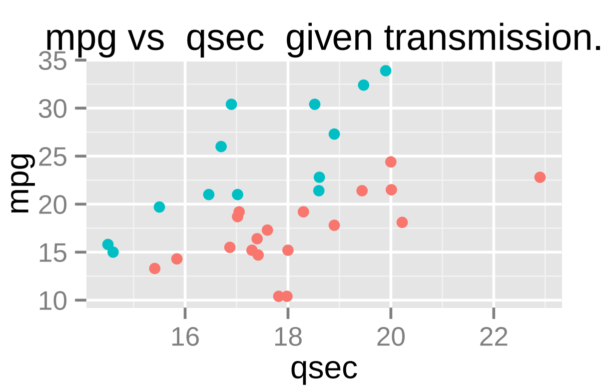

mpg vs hp given transmission.
mpg
35
30
25
20
15
10
100
200
300
hp

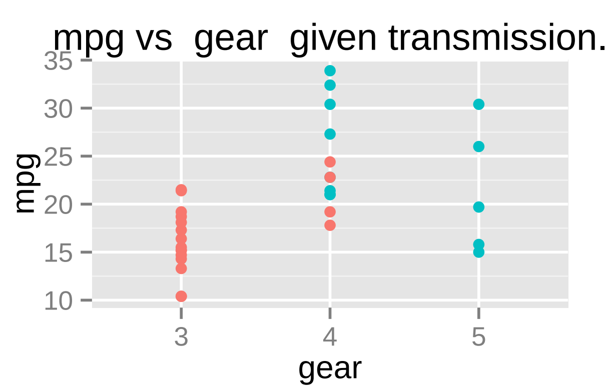

mpg vs disp given transmission.
mpg
35
30
25
20
15
10
100
200
300
400
disp

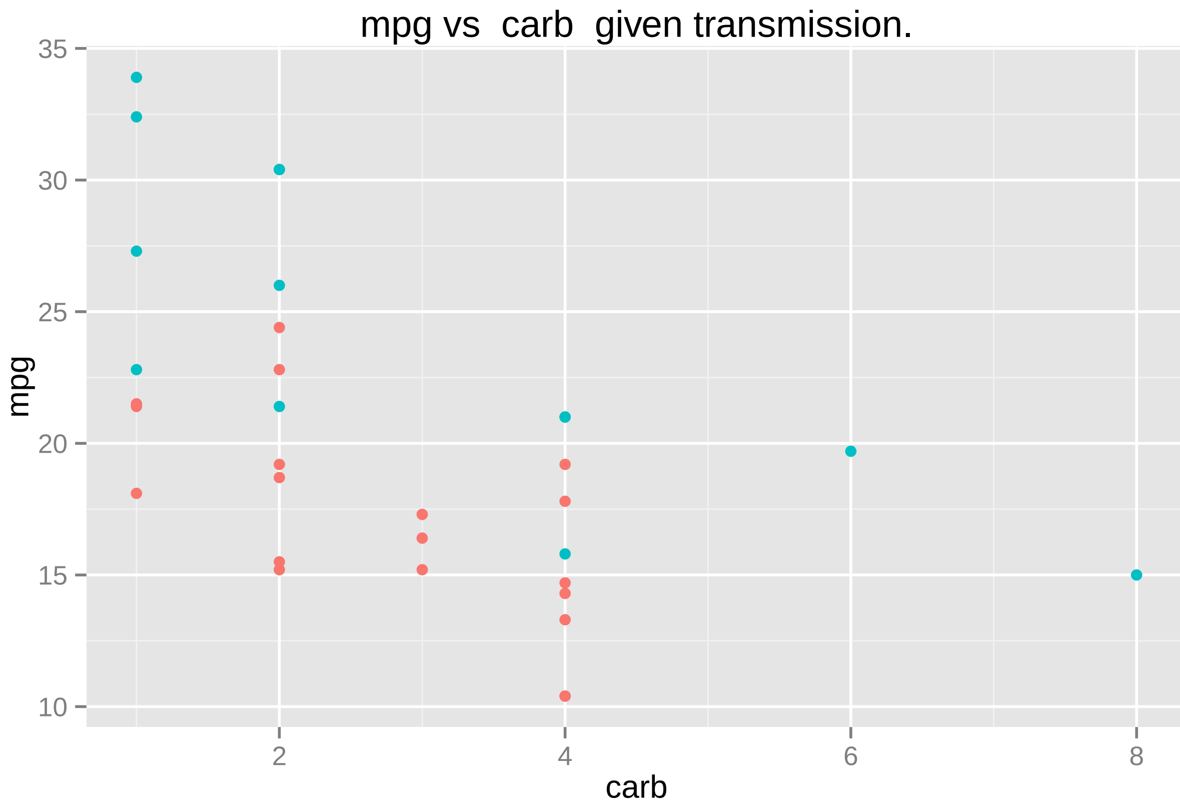

mpg vs drat given transmission.
mpg
35
30
25
20
15
10
3.0
3.5
4.0
4.5
5.0
drat


mpg vs wt given transmission.
mpg
35
30
25
20
15
10
2
3
4
5
wt

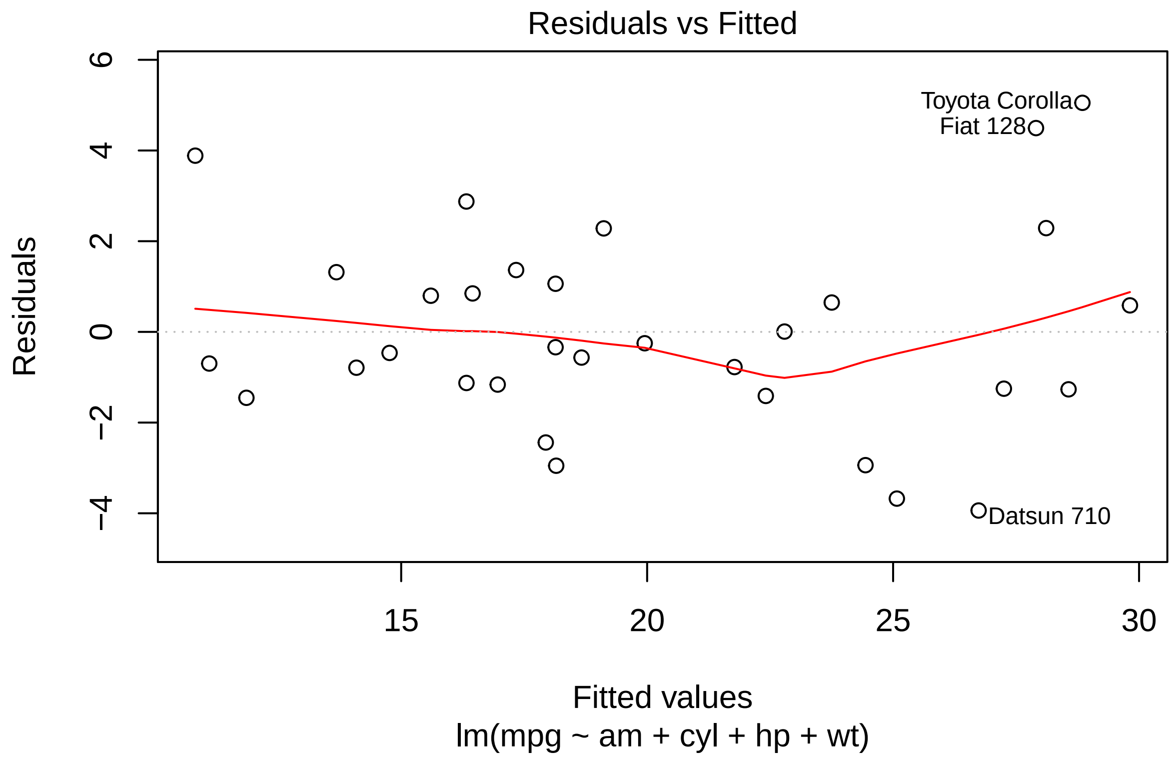

mpg vs vs given transmission.
mpg
35
30
25
20
15
10
0
1
vs


mpg vs qsec given transmission.
mpg
35
30
25
20
15
10
16
18
20
22
qsec

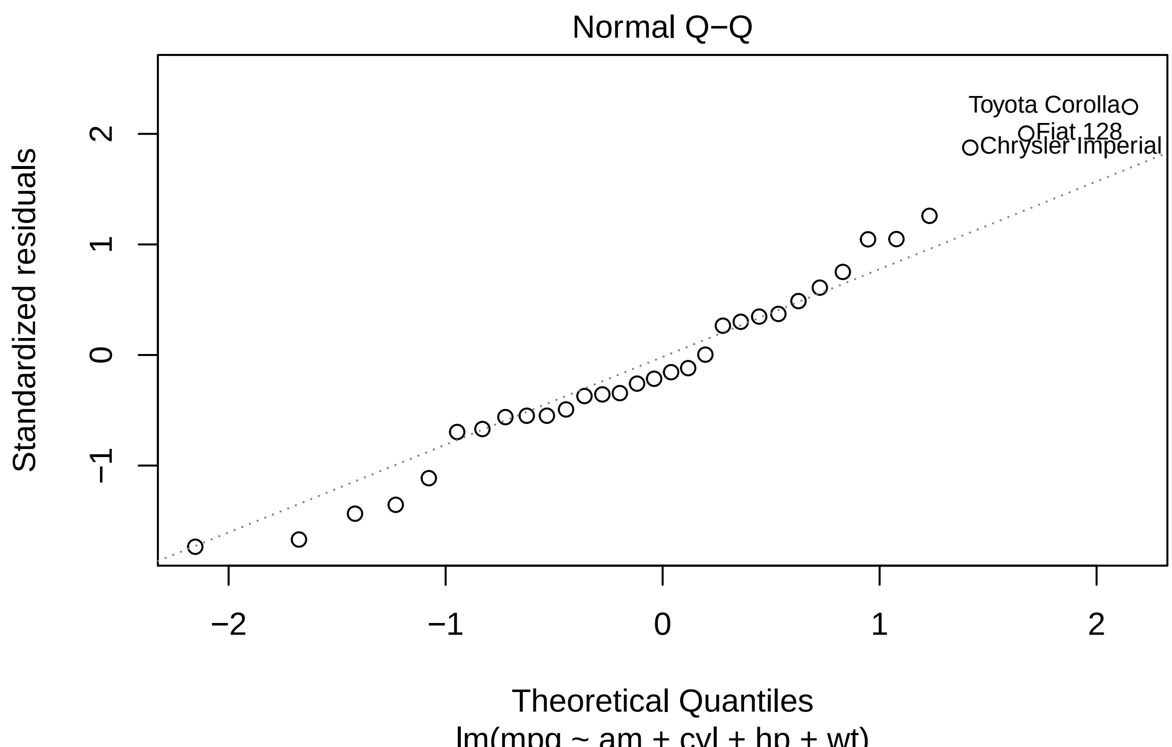

mpg vs gear given transmission.
mpg
35
30
25
20
15
10
3
4
5
gear


mpg vs carb given transmission.
35
30
25
20
15
10
mpg
2
4
6
8
carb

Appendix 2



Residuals vs Fitted
Toyota Corolla
Fiat 128
Datsun 710
Residuals
6
4
2
0
−2
−4
15
20
25
30
Fitted values
lm(mpg ~ am + cyl + hp + wt)



Normal Q−Q
Toyota Corolla
Fiat 128
Chrysler Imperial
Standardized residuals
2
1
0
−1
−2
−1
0
1
2
Theoretical Quantiles
lm(mpg ~ am + cyl + hp + wt)