Is an automatic or manual transmission better for MPG?

Lukas Nick

17 Auguscht 2014

Is an automatice or manual transmission better for MPG?

Executive summary

Question

1. Is an automatic transmission better for MPG?
2. How big is the difference in MPG between automatic and manual transmission?
3. Does the answers to 1) and 2) depend on other variables?

- number of cylinders
- weight
- displacementnumber of forward gears
- . . . ?

Data

I try to answer these questions based on the dataset mtcars. It includes 32 cars, each with

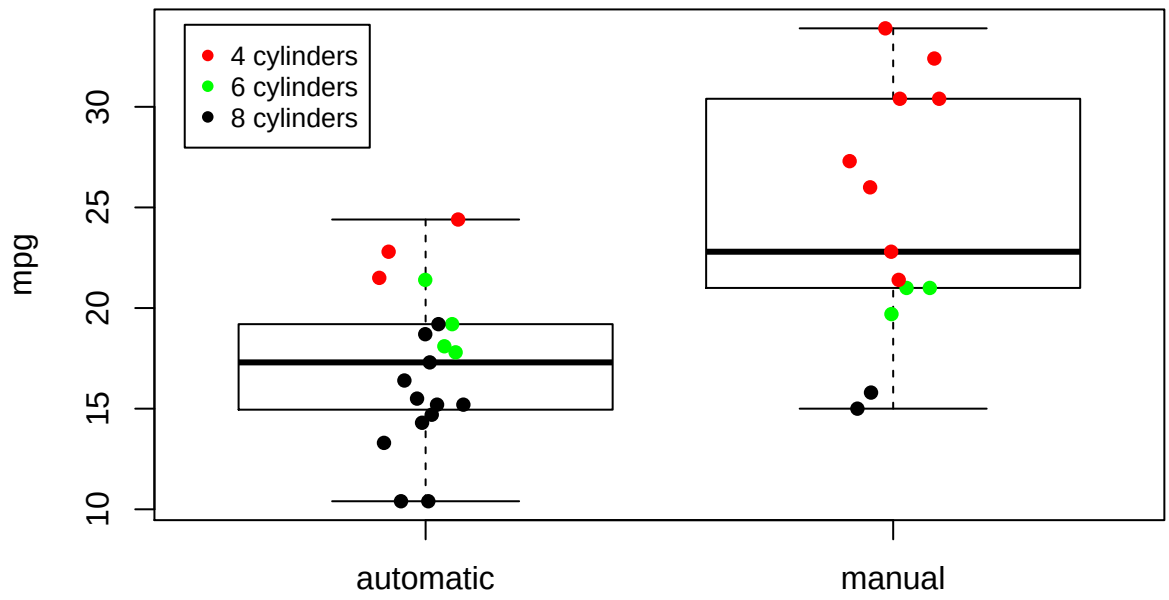
- miles per hours,
- number of cylinders
- diplacement
- gross horse power
- rear axis ratio
- weight
- 1/4 mile time
- straight or v-engine
- transmission (automatic vs manual)
- number of forward gears
- number of carburetors

Results

Exploratory results

Display the mpg for both automatic and manual transmission cars. Since - in theory - the number of cylinder might have an influence on the relationship between transmission and mpg, color the data points by the number of cylinders:

```
plot(cars$am, cars$mpg, ylab="mpg")
points(jitter(as.numeric(cars$am), factor=0.5), cars$mpg, col=sapply(cars$cyl, switch, 'red', 'green',
legend(x="topleft", legend=c('4 cylinders', '6 cylinders', '8 cylinders'), pch=16, col=c('red', 'green'
```



It looks like manual transmission cars are doing better in terms of mpg.

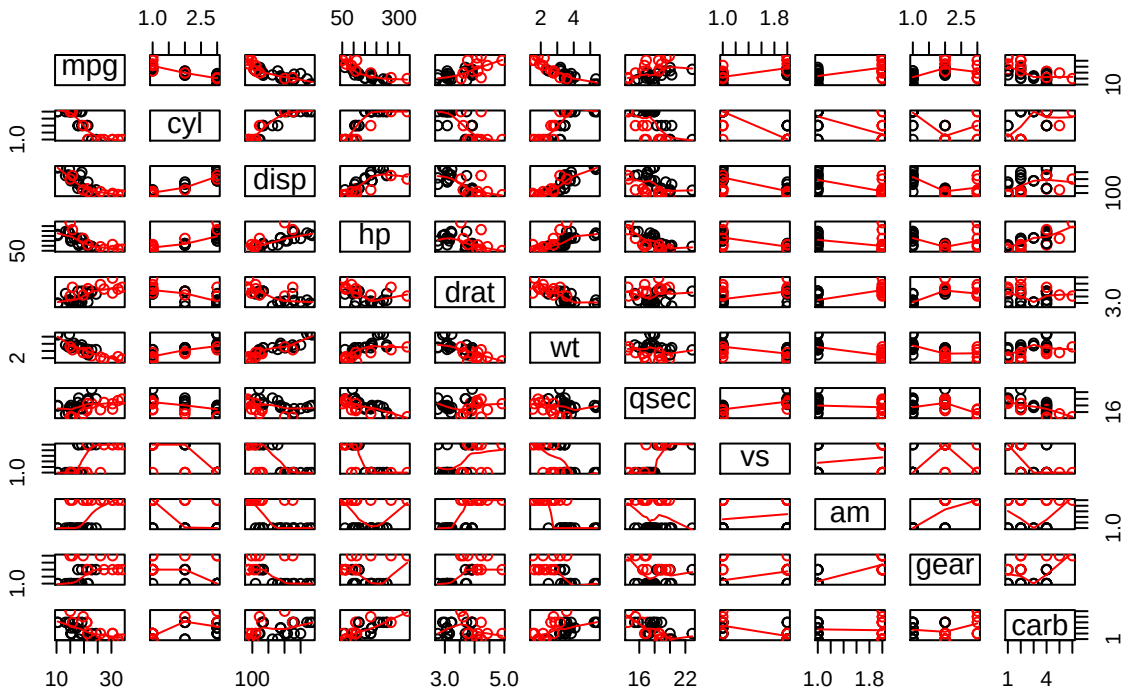
A simple linear regression of *mpg* on *am* confirms this:

```
fit.simple <- lm( mpg ~ am, data=cars)
summary(fit.simple)
```

```
##
## Call:
## lm(formula = mpg ~ am, data = cars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.392 -3.092 -0.297  3.244  9.508
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    17.15       1.12   15.25  1.1e-15 ***
## ammanual        7.24       1.76    4.11  0.00029 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.9 on 30 degrees of freedom
## Multiple R-squared:  0.36,   Adjusted R-squared:  0.338
## F-statistic: 16.9 on 1 and 30 DF,  p-value: 0.000285
```

Get a sense of how variables correlate with each other:

```
pairs(cars,panel=panel.smooth,main="mtcars", col=ifelse(cars$am=='manual', 'red', 'black'))
```



In order to find variables that we should adjust for, find out which variables *am* depends on.

First the categorial variables:

```
sapply( c("cyl", "vs", "gear", "carb"), function(cat){ summary( table(cars[,cat], cars$am) )} )
```

```
##           cyl      vs      gear       carb
## n.vars    2        2       2          2
## n.cases   32       32      32         32
## statistic 8.741    0.9069  20.94      6.237
## parameter 2        1       2          5
## approx.ok FALSE    TRUE    FALSE      FALSE
## p.value   0.01265  0.3409  2.831e-05  0.2838
## call      NULL     NULL    NULL       NULL
```

So the variables *cyl* and *gear* differ, depending on *am*.

Influence of the numerical variables:

```
sapply( c("disp", "hp", "drat", "wt", "qsec"), function(var){ t.test( cars[,var] ~ cars$am)[3] } )
```

```
## $disp.p.value
## [1] 0.00023
```

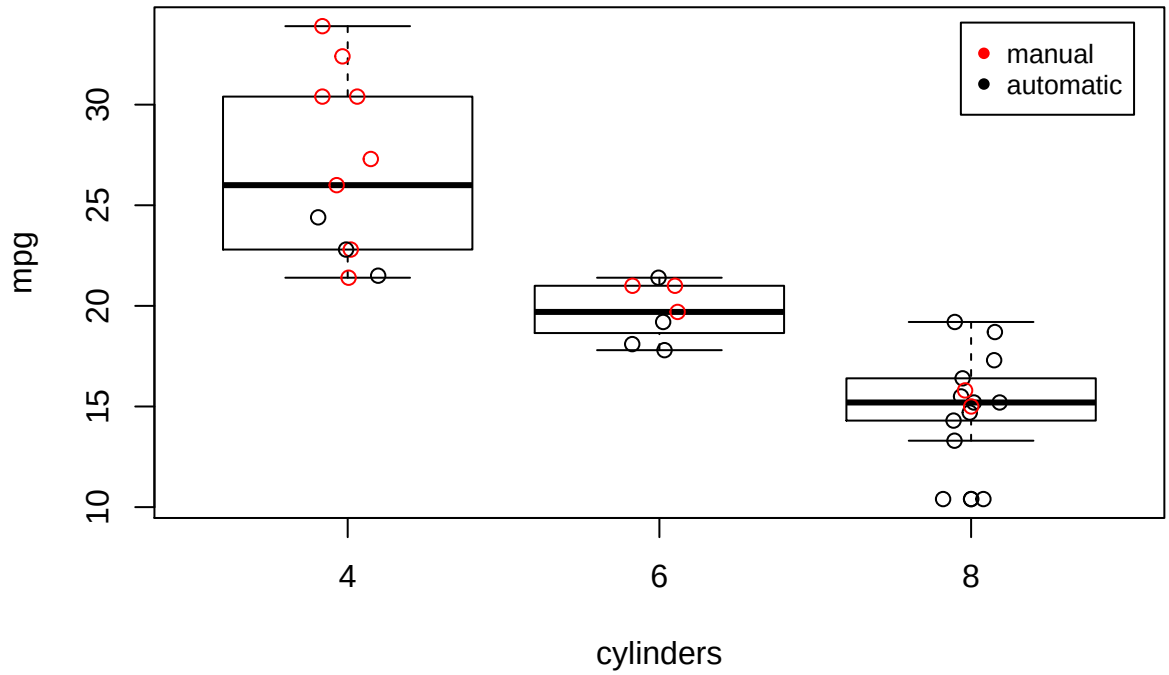
```
##
## $hp.p.value
## [1] 0.221
##
## $drat.p.value
## [1] 5.267e-06
##
## $wt.p.value
## [1] 6.272e-06
##
## $qsec.p.value
## [1] 0.2093
```

So the variables *disp*, *drat*, and *wt* differ depending on *am*.

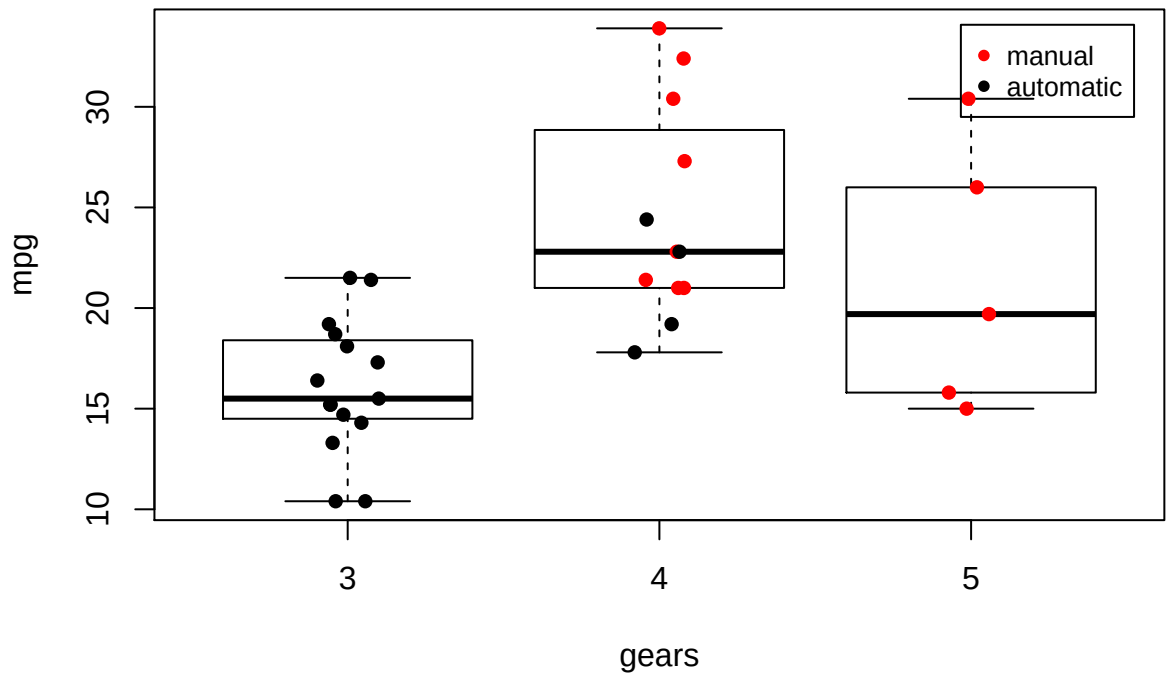
Let's look at some plots to get a feeling of the data distributions:

Categorical variables *cyl* and *gear*:

```
plot(cars$cyl, cars$mpg, xlab="cylinders", ylab="mpg")
points(jitter(as.numeric(cars$cyl), factor=0.5), cars$mpg, col=ifelse(cars$am=='manual', 'red', 'black'
legend(x="topright", legend=c('manual', 'automatic'), pch=16, col=c('red', 'black'), cex=0.8, inset=0.0
```

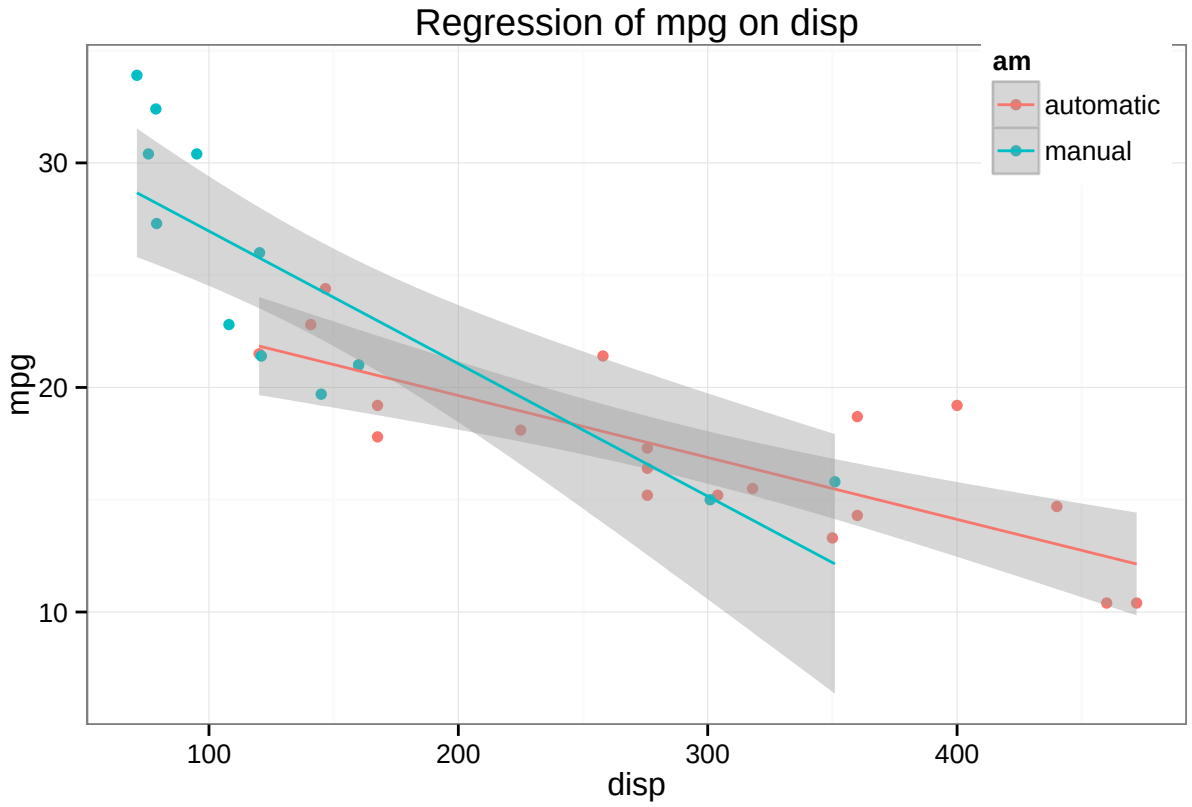



```
plot(cars$gear, cars$mpg, xlab="gears", ylab="mpg")
points(jitter(as.numeric(cars$gear), factor=0.5), cars$mpg, col=ifelse(cars$am=='manual', 'red', 'black
legend(x="topright", legend=c('manual', 'automatic'), pch=16, col=c('red', 'black'), cex=0.8, inset=0.0
```

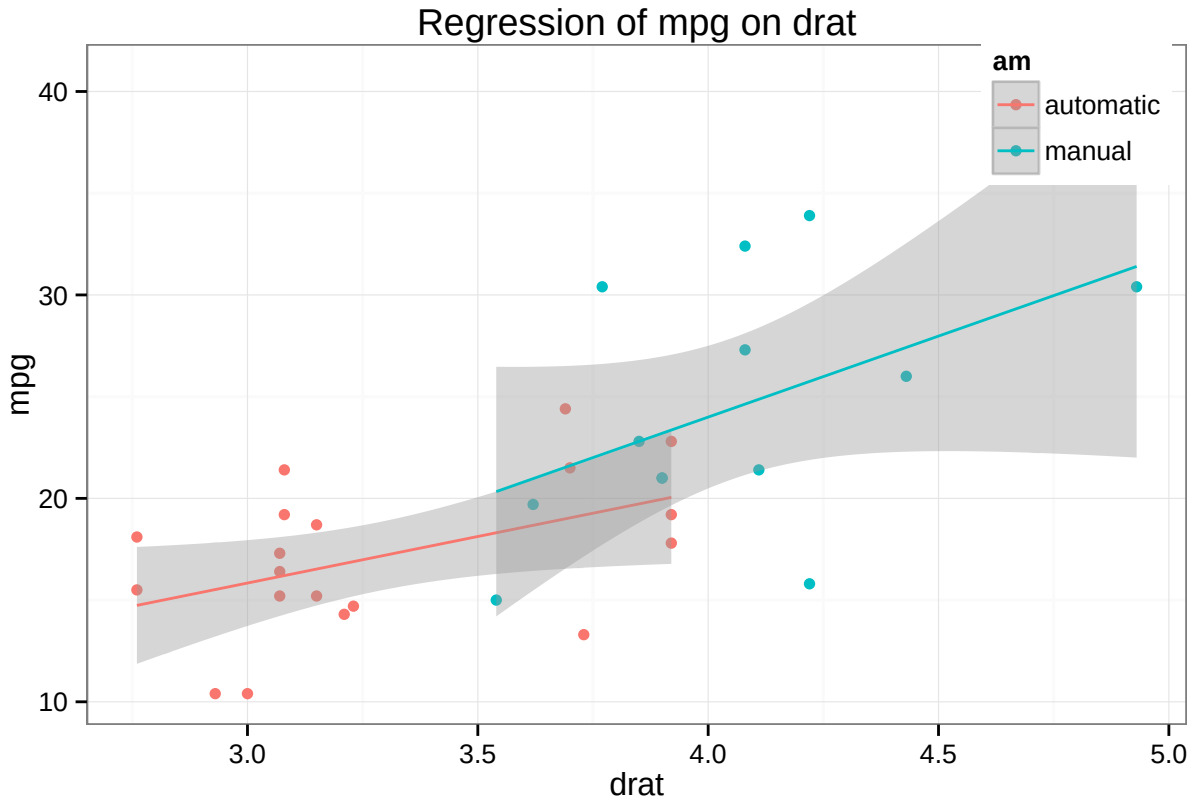


Numerical variables *disp*, *drat*, and *wt*:

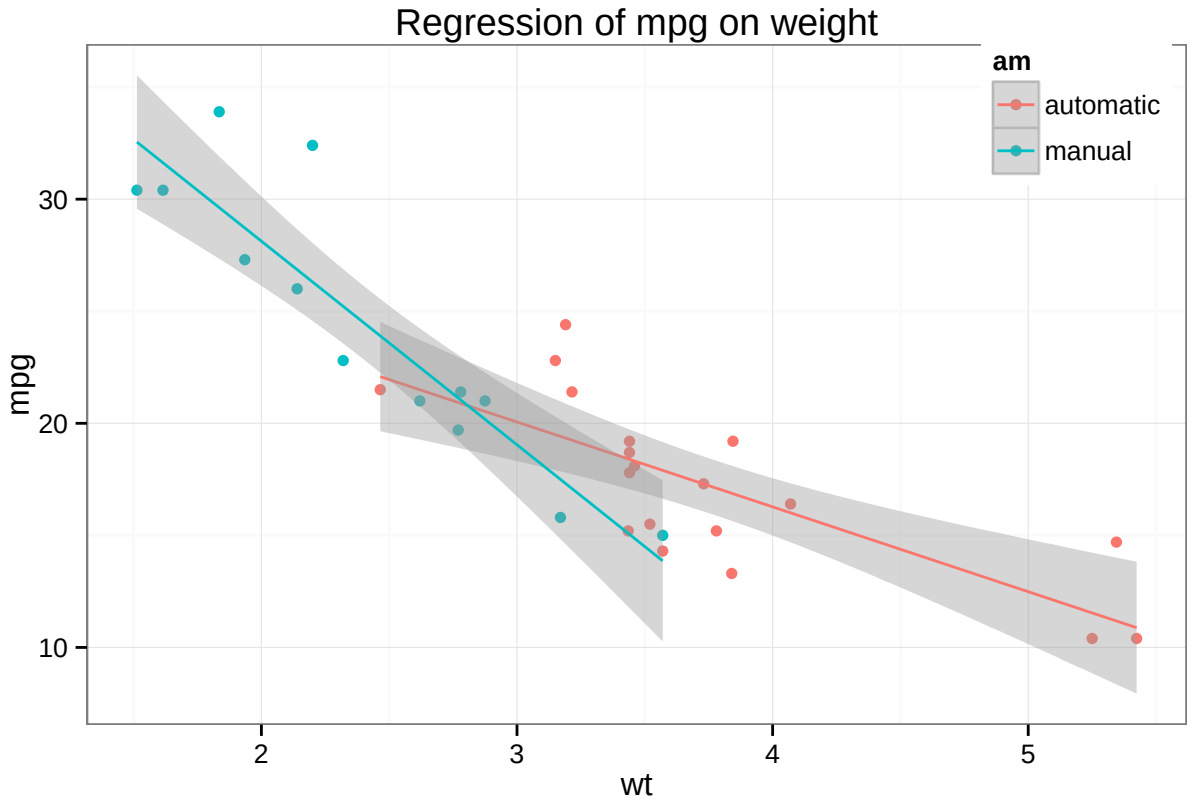
```
par(mfrow=c(2,2))
library(ggplot2)
qplot(disp, mpg, data=cars, geom=c("point", "smooth"),
      method="lm", formula=y~x, color=am,
      main="Regression of mpg on disp", xlab="disp", ylab="mpg") + theme_bw() + theme(legend.position =
```




```
qplot(drat, mpg, data=cars, geom=c("point", "smooth"),
      method="lm", formula=y~x, color=am,
      main="Regression of mpg on drat", xlab="drat", ylab="mpg") + theme_bw() + theme(legend.position =
```




```
qplot(wt, mpg, data=cars, geom=c("point", "smooth"),
      method="lm", formula=y~x, color=am,
      main="Regression of mpg on weight", xlab="wt", ylab="mpg") + theme_bw() + theme(legend.position =
```



multiple regression: summary(lm(Fertility ~ . , data = swiss))

- Make a boxplot for expl.Anal
- also just a scatter plot, with differently colored subgroups (red=manual, black=automatic, e.g.) possible to print both lm-lines in same plot (after fitting two different models, one for manual, one for automatic), see 02_02_c, p.28 also see p.29 for 2 lines in same model

Parameter interpretation: 02_02, p.5

Achtung: auch die 'unadjusted' Parameter anschauen, dh nur mpg ~ var1, ohne korrigienden Einfluesse der anderen Variablen. Die Zusammenhaenge koennen sich drehen (vgl Agriculture on Fertility in Swiss)!

Modell-Wahl:

- 02_02, p8
- 02_05, p14: nested testing of a model with anova & update

Need to use Dummy Variable!

Diagnostics:

residualplot: plot(fit, which=1)

Is an automatic transmission better for MPG?

How big is the difference in MPG between automatic and manual transmission?

Conclusions