JHU Regression Models Project

Ahmed Assal

Executive Summary

This study shows that the fuel consumption for manual transmission are better than automatic transmission by a factor of 1.8. Cars of manual transmission can gain up to 7.24 more miles per galon than automatic transmission cars. This conclusion has been reached using linear regression analysis.

Study

Context

Motor Trend, a magazine about the automobile industry Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome).

The Questions

They are particularly interested in the following two questions:
"Is an automatic or manual transmission better for MPG"
"Quantify the MPG difference between automatic and manual transmissions"

Take the mtcars data set and write up an analysis to answer their question using regression models and exploratory data analyses.

Data

Description

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).

Format

A data frame with 32 observations on 11 variables.

[, 1] mpg Miles/(US) gallon
[, 2] cyl Number of cylinders
[, 3] disp Displacement (cu.in.)
[, 4] hp Gross horsepower
[, 5] drat Rear axle ratio
[, 6] wt Weight (lb/1000)
[, 7] qsec 1/4 mile time
[, 8] vs V/S
[, 9] am Transmission (0 = automatic, 1 = manual)
[,10] gear Number of forward gears
[,11] carb Number of carburetors

Data Processing

Packages Loading

```
library(ggplot2)
library(caret)
```

```
## Loading required package: lattice
```

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(bitops)
library(RCurl)
```

Data Loading and Transofrmation

We load the data and convert am, gear, cyl, carb and vs variables to factor variables.

```
data(mtcars)
mtcarsOld = mtcars
mtcars$am   = factor(mtcars$am,labels=c("automatic","manual"))
mtcars$gear = factor(mtcars$gear)
mtcars$cyl  = factor(mtcars$cyl)
mtcars$carb = factor(mtcars$carb)
mtcars$vs   = factor(mtcars$vs)
```

Exploratory Analysis

Are the vairables correlated? To explore the relationship of the variable under study, mpg, with the other variables in the dataset, we plot mpg against each of the other variables(see appendix - fig.1). We find out that most of these variableas are highly correlated with mpg as pairs, i.e. mpg vs. a preictor variable.

Finding highly correlated variables We try find the highly correlated variables which should not be included in the model collectively as predictors.

```
c=cor(mtcarsOld)
corIndices = findCorrelation(c, cutoff = 0.75) ; print(corIndices)
```

```
## [1] 2 3 1 9
```

We find that variables of indices 2, 3, 1, 9 which are cyl, disp, mpg, and am are highly correlated. mpg is the output variable. am is the variable under study so we will include both of mpg and am while cyl, and disp should not be both included in the model. We will leave the step function to fine tune the selection process.

The effect of transmission type on the fuel consumption Also, to compare the effect of the type of transmission of the fuel consumption, we plot the different distributions of mpg vs transmission type(see appendix - fig.2). The figure shows that manual transmission cars tend to consume less fuel.

Regression Analysis

First, we build a base regression model that include all variablea as predictors for the MPG then we use the step function to build further linear models to select the most approprate variables as predictors while eliminating those that do not add much to the model.

```
baseModel = lm(mpg ~., data= mtcars)
bestModel = step(baseModel, direction = "both")
```

As we can see, the algorithm has selected cyl, hp, wt and am as predictors.

```
summary(bestModel)
```

```
##
## Call:
## lm(formula = mpg ~ cyl + hp + wt + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## hp          -0.03211    0.01369  -2.345  0.02693 *
## wt          -2.49683    0.88559  -2.819  0.00908 **
## ammanual     1.80921    1.39630   1.296  0.20646
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

Residuals and diagonistics

To check residuals for normaility and homoskedasticity, we plot the residuals(see appendix - fig.3). We find that the residuals are normally distributed and homoskedastic.

Results

Explanation of the variability of the data

This model, bestModel, has successflly explained more than 83% of the variablilty of the data.

Interpretation of the best model coefficients

mpg vs transmission control

Miles per gallon increases by a factor of 1.8 (1.8) with manual transmission.

mpg vs hp

Miles per gallon decreases by a factor of 0.03 (-0.03) as horspower increases.

mpg vs hp

Miles per gallon decreases by a factor of 2.5 (-2.5) for every increase of 1000lb in the weight.

mpg vs cyl6/cyl8

Miles per gallon decreases by a factor of 3.03 (-3.03) for 6 cylinders and by a factor of 2.16 (-2.16) for 8 cylinders.

Intercept

The intercept is at 33.7 mpg.

P-value

The overall p-value is very small (1.506 x 10 ^ -10) which means that there are a very small uncertainty with the sign of the coefficients.

Answers to the study questions

"Is an automatic or manual transmission better for MPG"

As shown in this study, the manual transmission is better for fuel consumption, miles per gallon. This is demonestrated by the best fit model which estimates the decrease of fuel consumption by a factor of 1.8 for manual transmission.

"Quantify the MPG difference between automatic and manual transmissions"

Fuel consumption, mpg, for manual transmission increases by a factor of 1.8 over automatic transmission.

```r
manualData = mtcarsOld[ mtcarsOld$am == 0, 1]
automaticData = mtcarsOld[mtcarsOld$am == 1, 1]
manualCI = t.test(manualData, mu = 0)$conf.int
automaticCI = t.test(automaticData, mu = 0)$conf.int
print(manualCI);print(automaticCI)
```

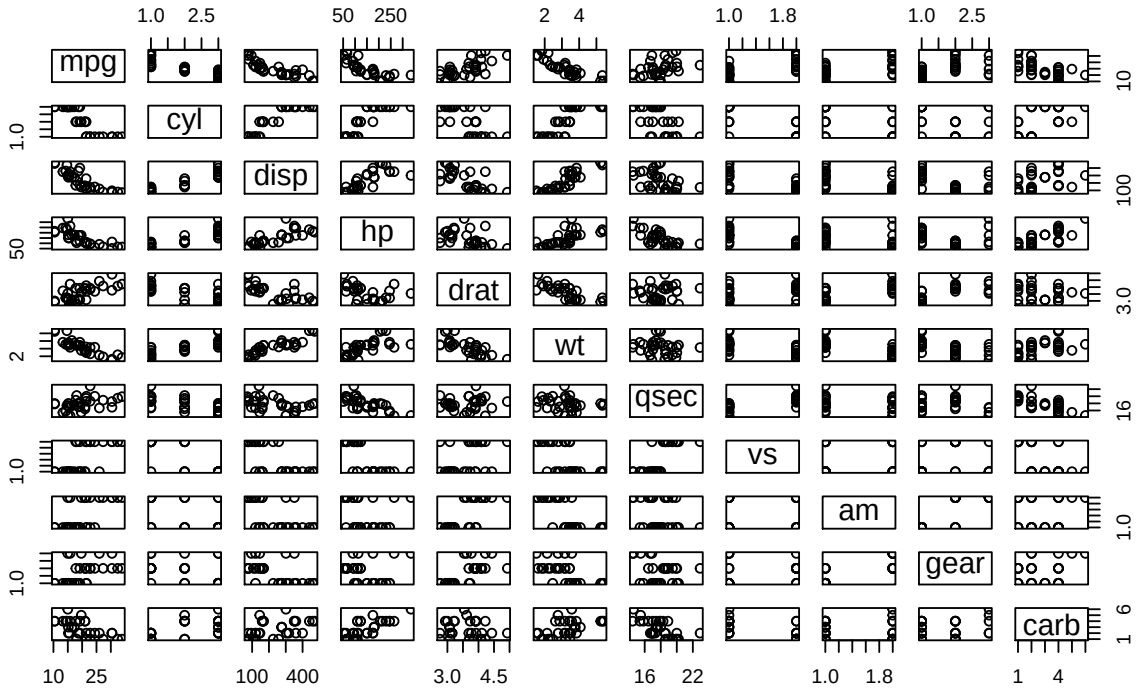
```
## [1] 15.29946 18.99528
## attr(,"conf.level")
## [1] 0.95
```

```
## [1] 20.66593 28.11869
## attr(,"conf.level")
## [1] 0.95
```

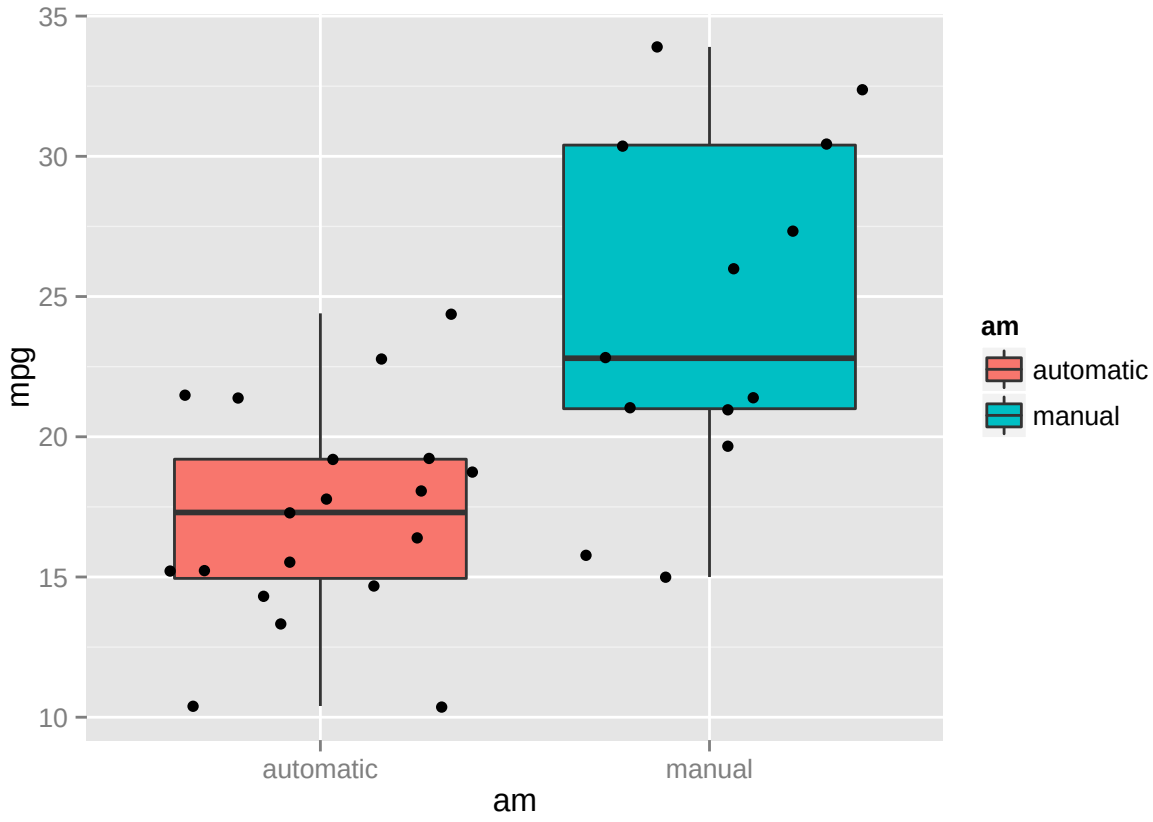
The 95% confidence interval for mpg for automatic transmission (15.3, 19) while it is (20.67, 28.12) for the manual transmission. This is an evidence that manual transmission consumes less fuel. Manual transmission travels around 7.24 more miles per gallon, (24.39mpg), than automatic transmission (17.15 mpg).

Appendix

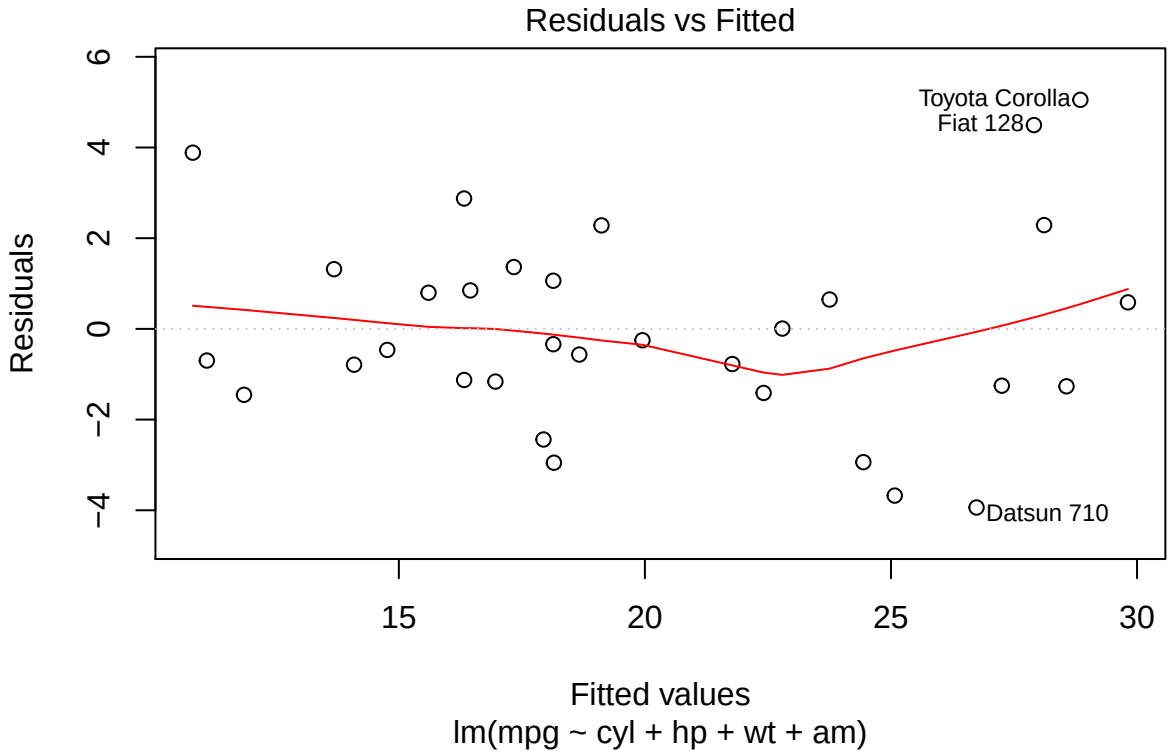
Exploratory features plot

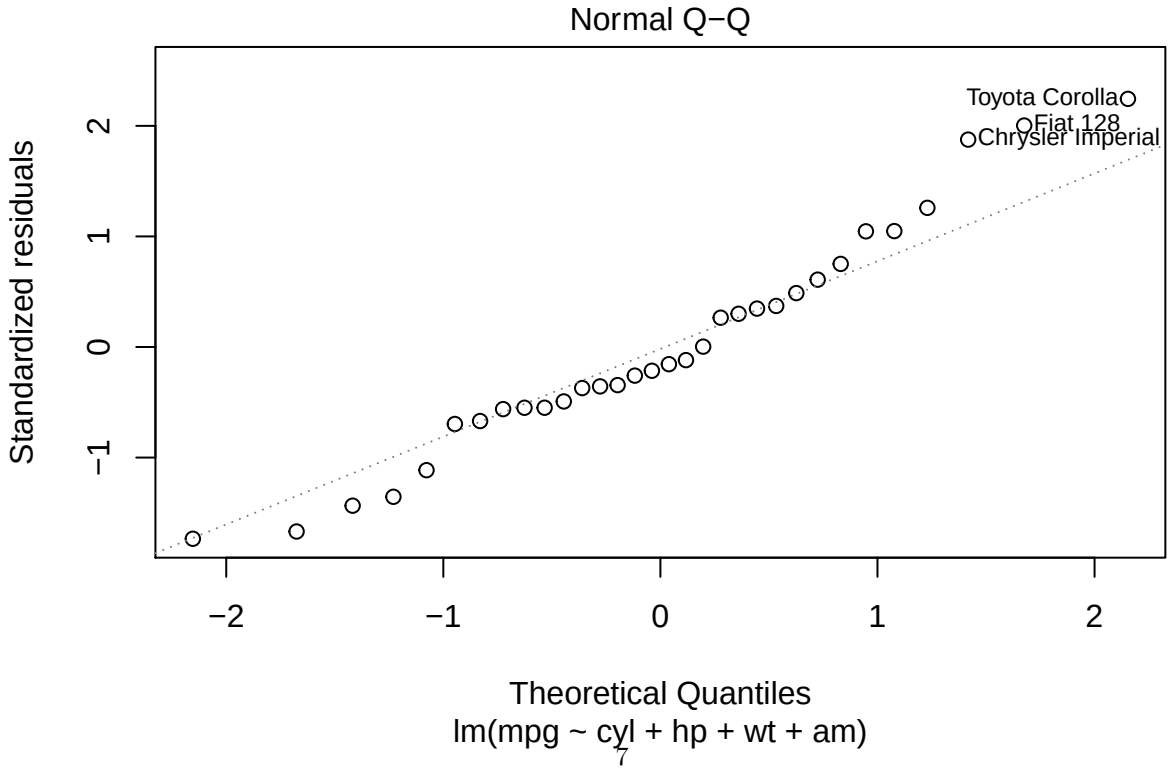


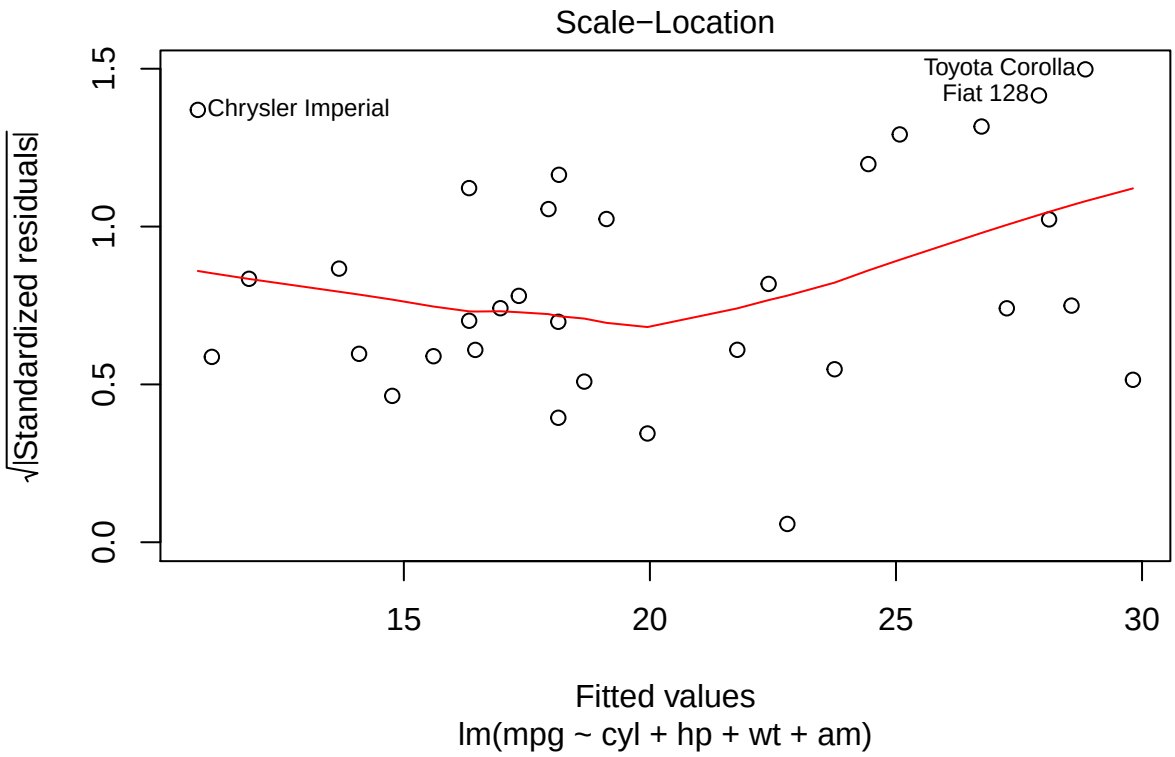
Miles per gallon vs. transmission type

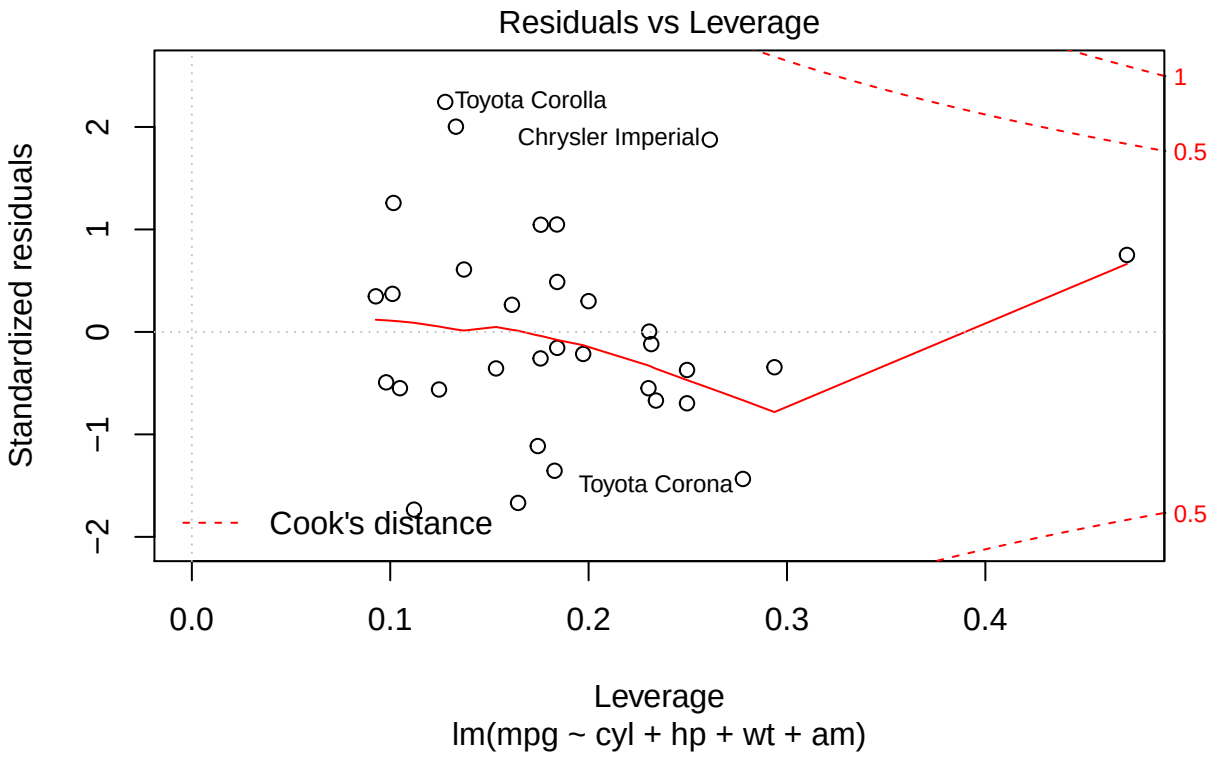


Residuals plot

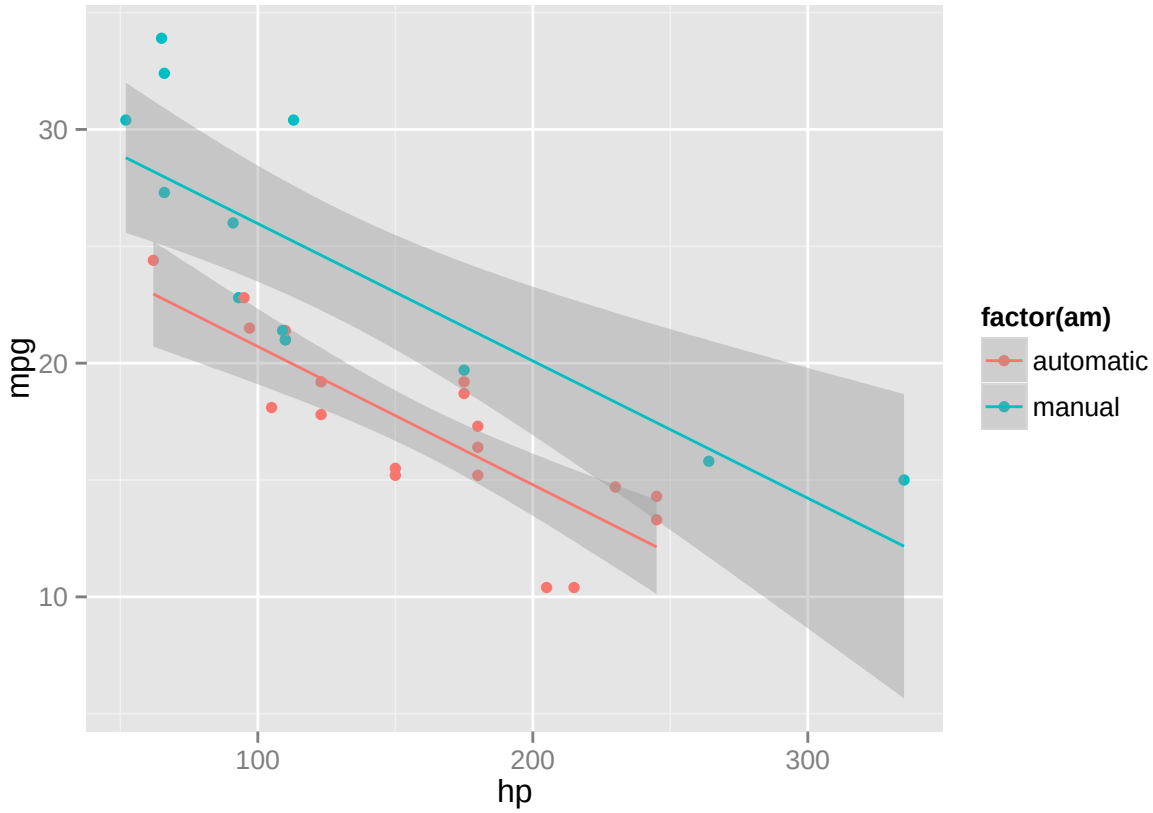


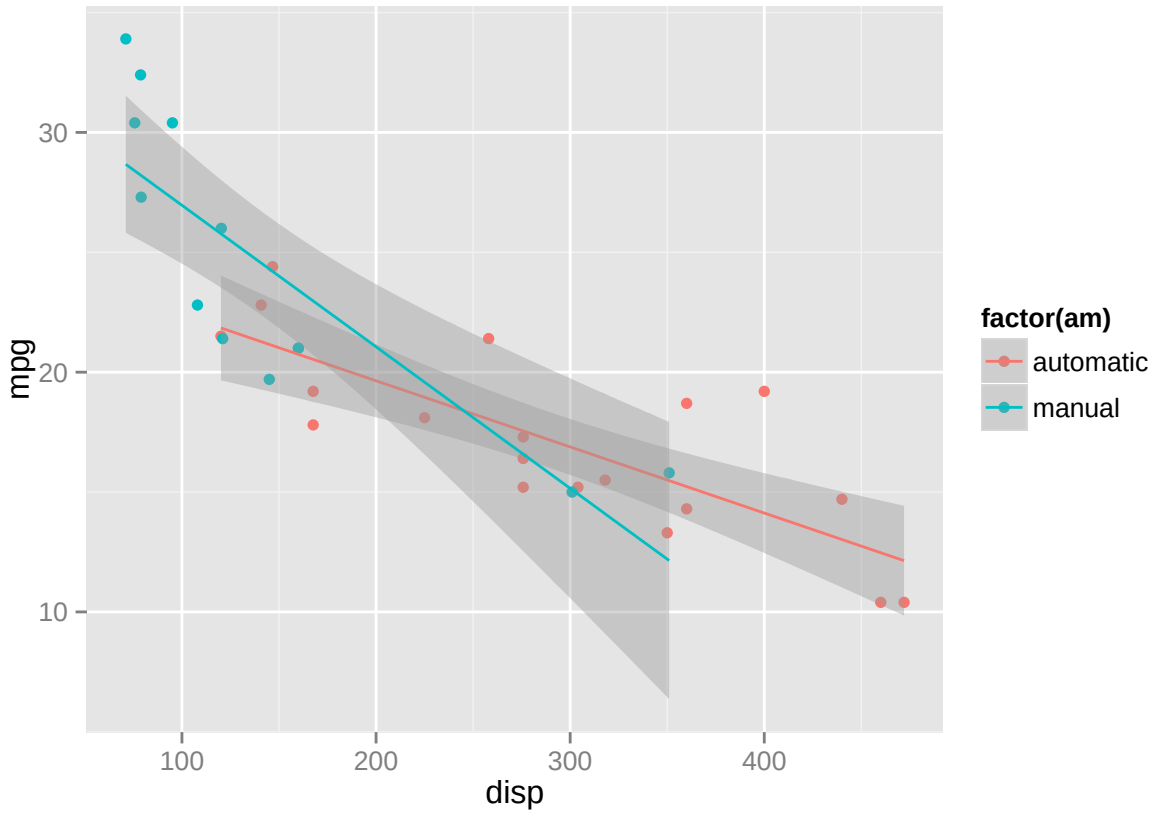


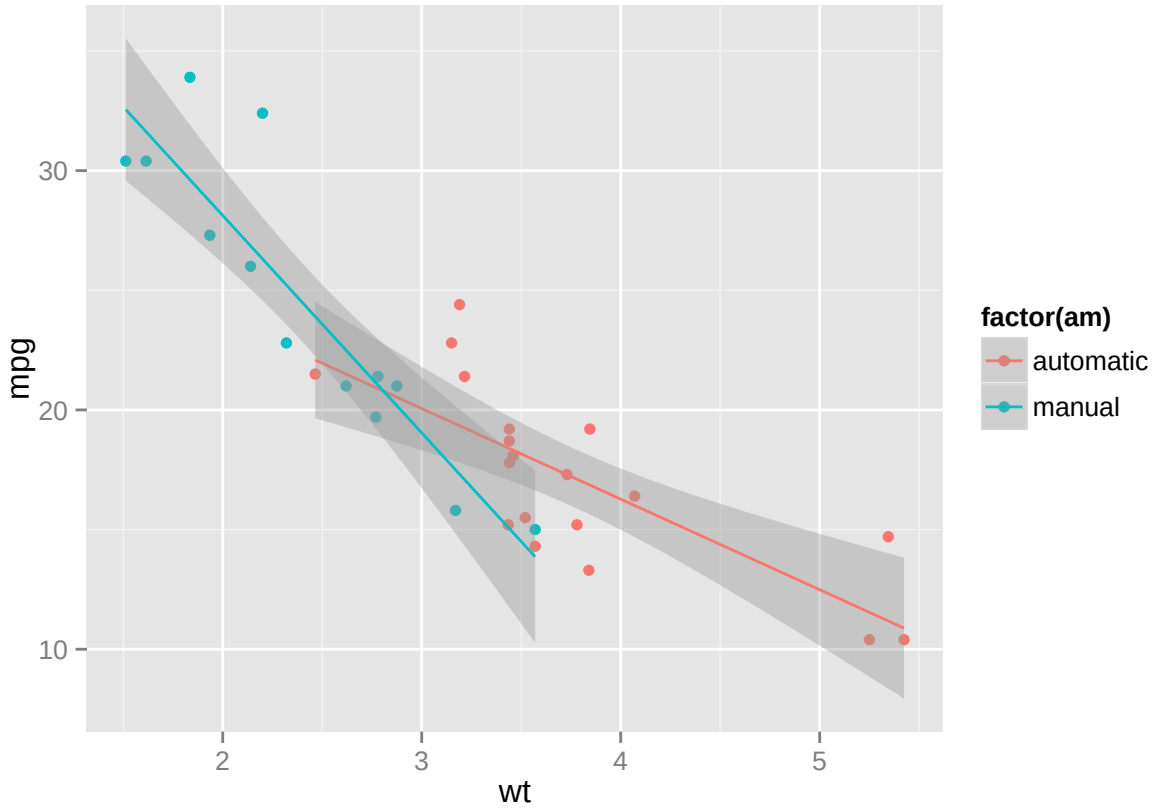

Scale−Location
Chrysler Imperial
Toyota Corolla
Fiat 128
√|Standardized residuals|
0.0
0.5
1.0
1.5
15
20
25
30
Fitted values
lm(mpg ~ cyl + hp + wt + am)


Residuals vs Leverage
Standardized residuals
Leverage
lm(mpg ~ cyl + hp + wt + am)
Toyota Corolla
Chrysler Imperial
Toyota Corona
Cook's distance
1
0.5
0.5
2
1
0
−1
−2
0.0
0.1
0.2
0.3
0.4

Exploration of other linear models






mpg
30
20
10
2
3
4
5
wt
factor(am)
automatic
manual