Regression Models Project - Motor Trend Data 'mtcars' Miles Per Gallon Analysis

james c walmsley

12/1/2016

Executive Summary:

```
The mean "MPG of automtic transmisions is 17.15 mpg, and manual
transmisions 24.38 mpg for a difference of 7.24 mpg in favor
of the manual transmission.

Exploring the other relationships we found that the vehicle weight
is signficantly correlated (-86%) with mpg ratings. Addtionally we
also discovered that transmission type is correlated with vehicle wt
at (69%). In other words lighter vehicles tend to have manual
transmissions where as heavier vehicles tend to have automatic
trasnmissions.

To answer the question of wheter automatic or manual transmissions are
better for gas mileage we must vehicles roughly the same in as many
aspects (regressors) as possible using this given data.
than transmission.




### Add after completing analysis
NOTE: include some info on cor, confint, ChisSq?, VIF
```

Problem Statement & Questions to Answer:

```
## Q1 "Is an automatic or manual transmission better for 'mpg'"
## Q2 "Quantify the MPG difference between automatic and manual transmissions"
```

Grading - Criteria (remove on completion)!!!

```
####???? Did the report include an executive summary?
####???? Did the student answer the questions of interest or detail why the question(s) is (are) no
####???? Did the student quantify the uncertainty in their conclusions and/or perform an inference
####???? Was the report brief (about 2 pages long) for the main body of the report and no longer th


####YES!!!! Did the student interpret the coefficients correctly?
####YES!!!! Did the student do some exploratory data analyses?
```

```
####YES!!!! Did the student fit multiple models and detail their strategy for model selection?
####Needs some more work


####YES!!!! Did the student do a residual plot and some diagnostics?
####YES!!!! Was the report done in Rmd (knitr) with pdf output?
```

Analysis Considerations:

```
Descriptive - (dim, mean, sd, sigma^2, str & summary) statistics

Exploratory - pairs, histograms, QQ, fitted, residualplots, boxplots
& (multiple plots); T-Test

Analysis -  OLS, simple linear regression, statistical linear regression,
multivariate regression & model selection, logistic regression, pValues,
adjustments, residuals, residual fit, predict fit, hatvalues, variance, & dfbetas, R^2,
diagnostics; ANOVA, coeficients, confint, correlation, covariance, variance inflation
```

Software Environment: & System - session Info:

```
Set the Working Directory then get System & Session Info
```

Accessing Data & Raw Data Overview: Motor Trend 'mtcars' data set:

```
Clean up the work space & get the data:
```

A data frame with 32 observations on 11 variables.

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

Process Data: factor columns 2 & 8:11 (cyl,vs,am,gear,carb)

Descriptive Statistics

Exploratory Analysis:

```
See figures 1:4 in Appendix A
Add narrative here!!
```

Statistical Modeling, Regression & Model Fit:

```
Assumptions:
        A A correlation to mpg ratings may exist among multiple variables
        B
```

Simple Linear Regression Plot

Bivariate Linear Model: # Note: manual trans & eng shape / vs are significant

Multivariate Linear Model (all vars) Fitted Values, Quantiles & Leverage Plots

Note; vars with lowest vif = drat, vs & am

Multivariate Linear Model Adjusted

(Note: the variable qsec appears to be significant at this point)

Multivariate LM Nested & ANOVA table

With nested modeling method, models 3 & 6 each add a significate variable to the fitted model and in this case model 3 is disp & model 6 is wt. Now use the vif (variable inflation) test on model six to check for any variance inflation among the varialbes of this model. The results show that "hp", "drat", and "wt" all have square rooted infaltion values less than **3.0** so these can be accepted into the model of best fit (fbf1) along with the main variable of the study "am"

Best Fit Modeling

Based on the nested modeling process followed by the anova table check then followed by the vif and sqrt(vif) test we decide to go with the following model labeled (fbf1) and notice that all of the sqrt(vif) values are less than **2.0** indicating a good model fit.

DIAGNOSTICS A

DIAGNOSTICS B

It turns out that the confint for "am" 17.33614 - 42.71801 is rather wide and does not include zero meaning it may not be the best overall predictor of "mpg" rating values. On the other hand the much narrower confint of "drat", "hp", & "wt" indicate that they are likely to be good predictors of "mpg" # however the confint for drat, hp and wt each include zero

NEXT PROCESS: Prediction !!!!

Preliminary Findings:

```
Questions of Interest:
        A What other regressors if any correlated with mpg rating and transmission type?
        B
Interpretation of Results:
        A Using ANOVA table with Nested Multivariate Regression fit it is clear that the variable w
        B Based on the
        C
```

Inference:

```
Hypothesis':
        A HO = The difference between Automatic and Manual transmission MPG = 0
        B Ha = The difference between Automatic and Manual transmission MPG != 0
        C Desired confidence interval = .95 (one sided) ??
```

Conclusions / Recommendations:

```
    A
    B
```

Are there other alternative analyses?

```
    A VIF
    B Challenge the results ?
    C Measures of uncertainty 'e'
```


# Appendix A: "Exploratory Graphical Analysis"

Pairs Plot:

Interpretation: MPG decreases as the; # of cylinders increases, the engine displacement increases, the horsepower increases, the weight increases, the rear axle ratio decreases, qsec

time decreases, the engine is a V-block, # of gears decreases, and the # of carburators increases ~ conversely the MPG increases as the: rear axle ratio increases, qsec time increases, the engine is a Straight-block, transmission is manual, # of gears increases and the # of carburators decreases
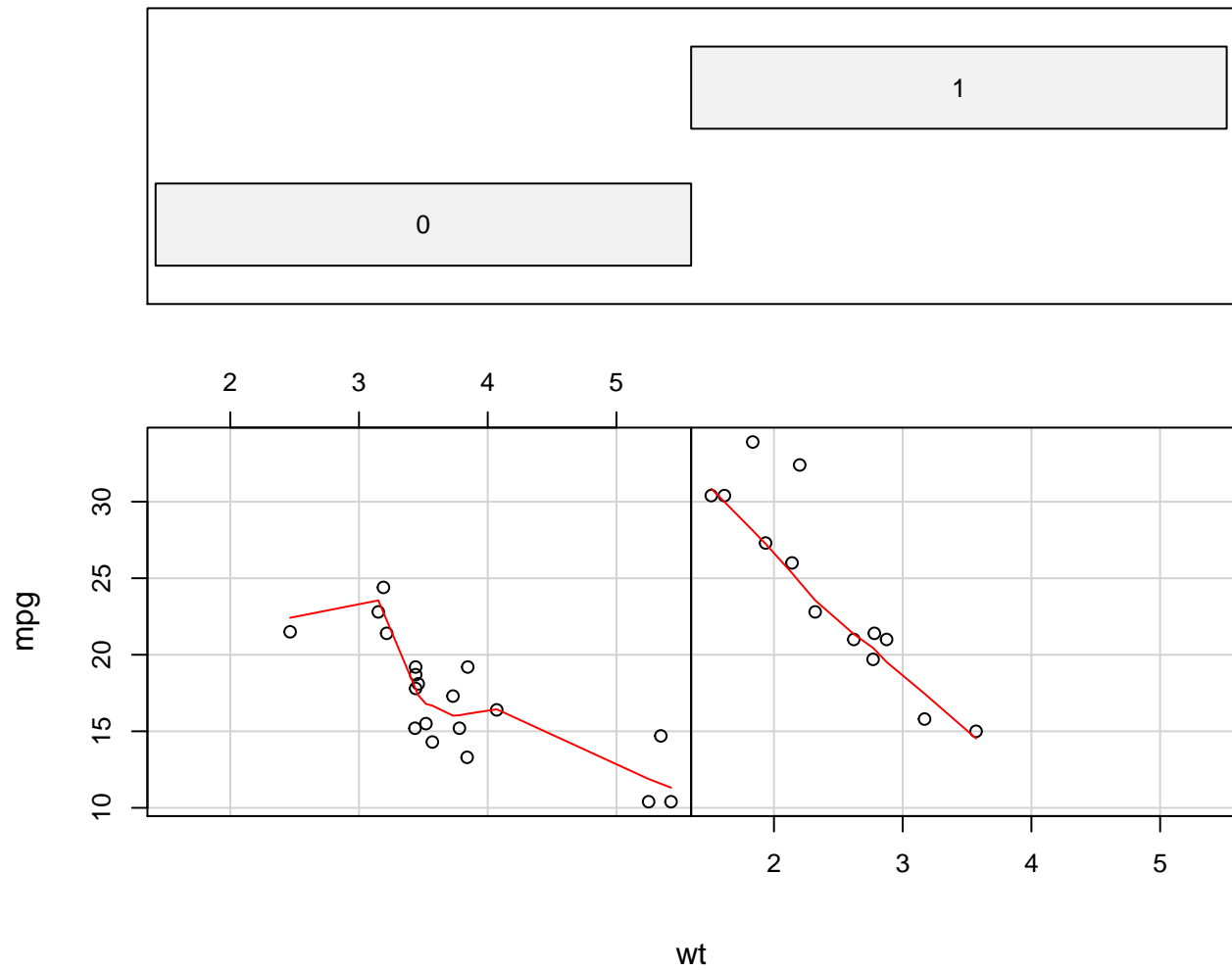
mtcars data
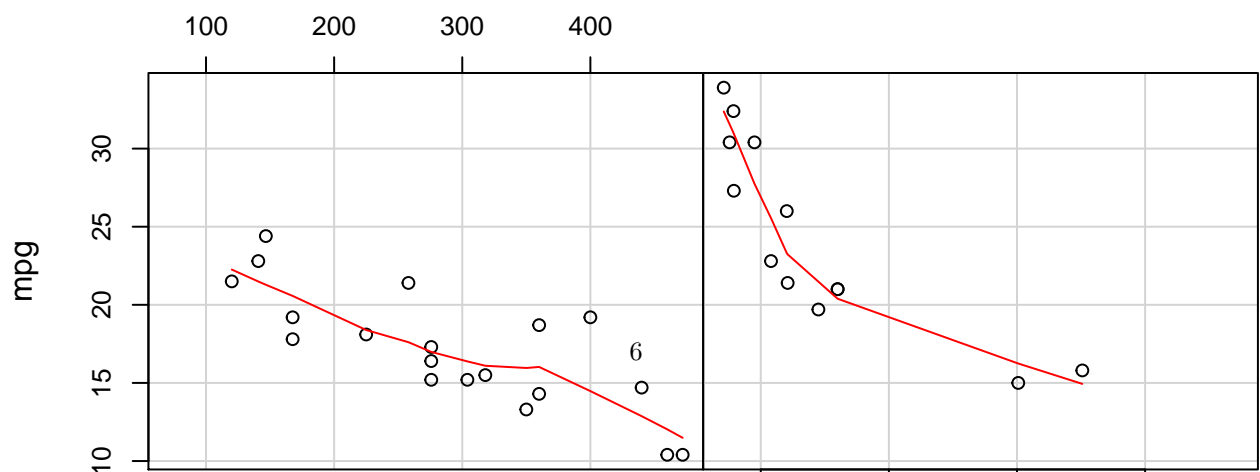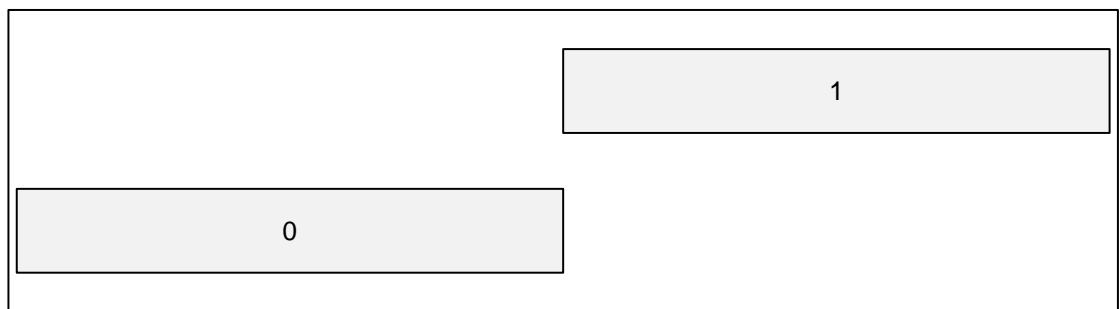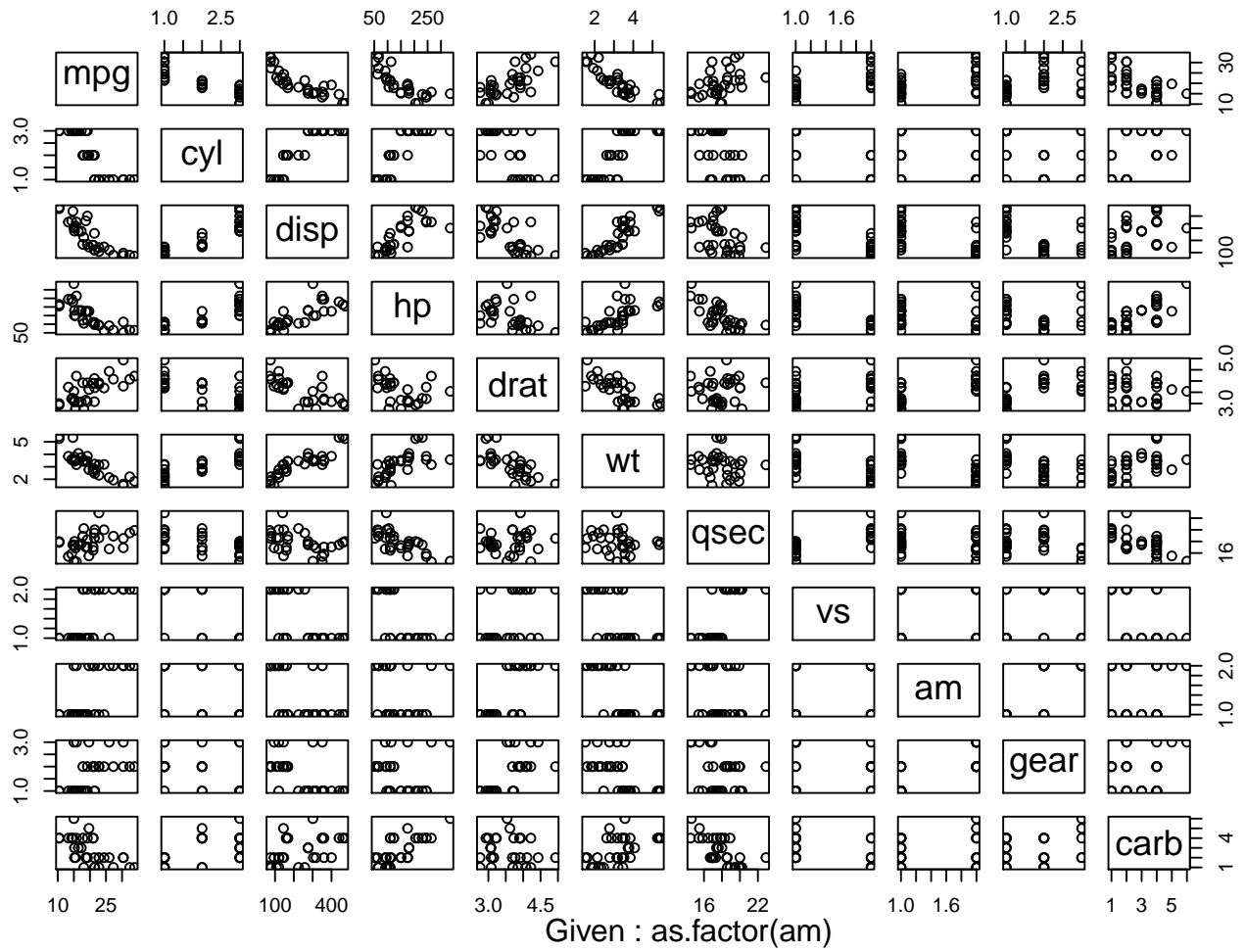
Given : as.factor(am)

mtcars data
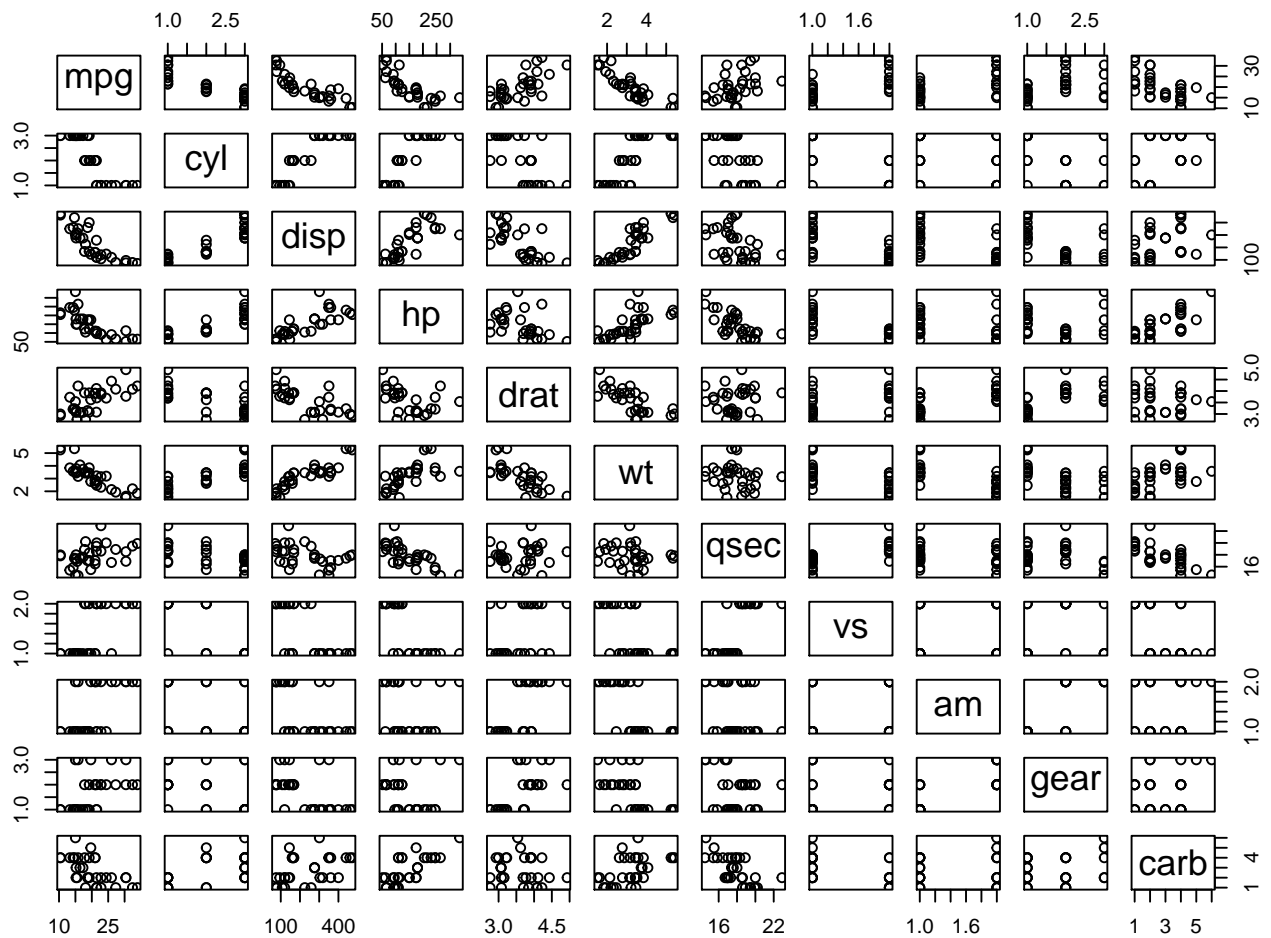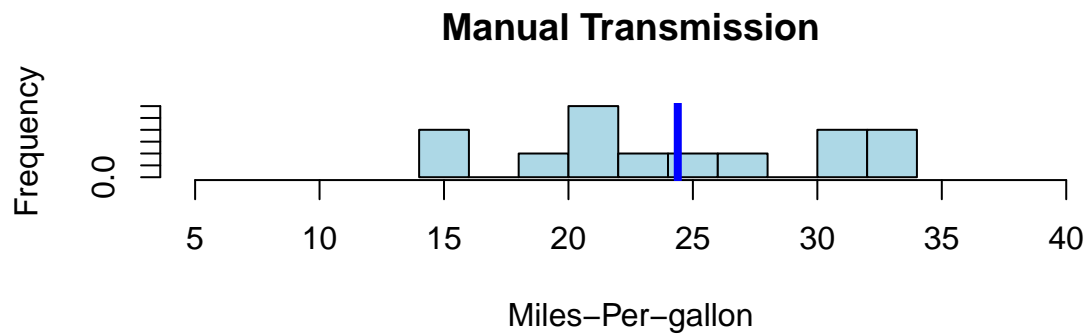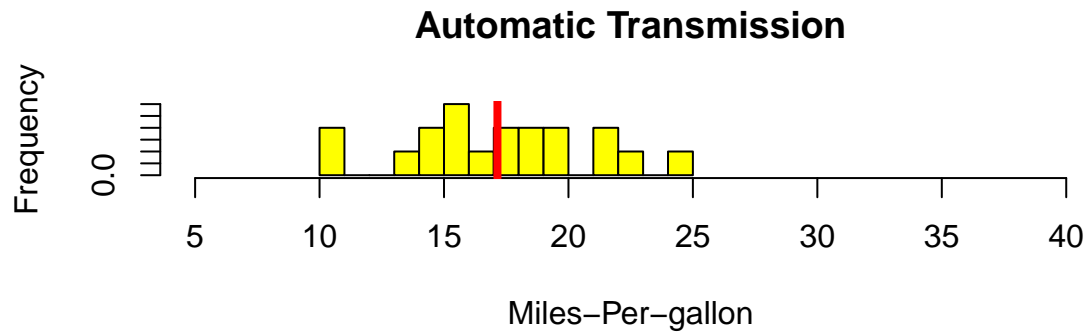
Given : as.factor(am)

## Histograms Plot

```
##               mpg cyl disp  hp drat    wt  qsec      vs     am gear carb
## Mazda RX4     21.0   6  160 110 3.90 2.620 16.46 V-block Manual    4    4
## Mazda RX4 Wag 21.0   6  160 110 3.90 2.875 17.02 V-block Manual    4    4
## Datsun 710    22.8   4  108  93 3.85 2.320 18.61 S-block Manual    4    1
```

Automatic Transmission

Manual Transmission

Box Plot

```
##                     mpg cyl disp  hp drat    wt  qsec      vs     am gear carb
## Mazda RX4          21.0   6  160 110 3.90 2.620 16.46 V-block Manual    4    4
## Mazda RX4 Wag      21.0   6  160 110 3.90 2.875 17.02 V-block Manual    4    4
## Datsun 710         22.8   4  108  93 3.85 2.320 18.61 S-block Manual    4    1
```

Automatic vs Manual Transmission, Miles Per Gallon Ratings
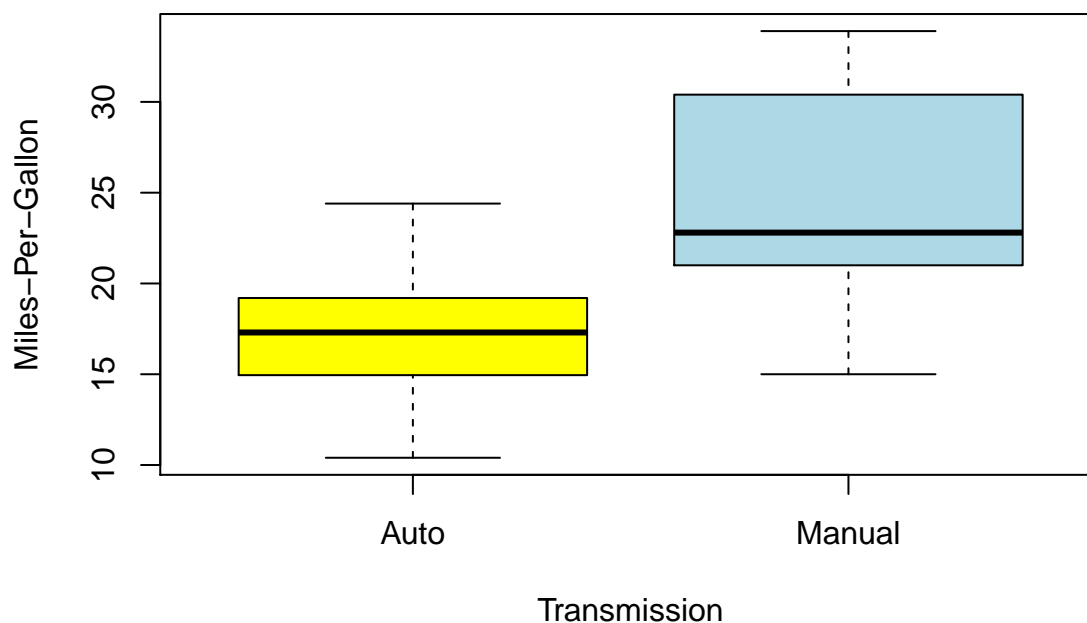
NOTE: use the cut function by 3 on MPG

=== END ===