

Regression Models Project - Motor Trend Data 'mtcars'

Miles Per Gallon Analysis

james c walmsley

12/1/2016

I. Executive Summary:

Add after completing analysis

NOTE: include some info on cor, confint, ChisSq?, VIF

II. 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) not answerable?

???? Did the student quantify the uncertainty in their conclusions and/or perform an inference correctly?

???? Was the report brief (about 2 pages long) for the main body of the report and no longer than 5 with supporting appendix of figures?

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?

YES!!!! Did the student do a residual plot and some diagnostics?

YES!!!! Was the report done in Rmd (knitr) with pdf output?

III. Analysis Considerations:

Descriptive - (dim, mean, sd, sigma², 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, R²,
diagnostics; ANOVA, coefficients, confint, correlation, covariance, variance inflation

IV. Software Environment: & System - session Info:

Set the Working Directory then get System & Session Info

```
setwd("~/Desktop/Coursera_R/7_Regression Models/RM_proj_MPG_MotorTrendData")
sessionInfo()
```

V. Accessing Data & Raw Data Overview: Motor Trend ‘mtcars’ data set:

Clean up the work space & get the data:

```
rm(list=ls());library(car);library(dplyr); data("mtcars");any(is.na(mtcars));data(mtcars)
```

```
##
## Attaching package: 'dplyr'

## The following object is masked from 'package:car':
##
##      recode

## The following objects are masked from 'package:stats':
##
##      filter, lag

## The following objects are masked from 'package:base':
##
##      intersect, setdiff, setequal, union
```

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)
```

VI. Process Data: Cconvert columns 2 & 8:11 (cyl=(4,6,8), vs = engine block shape (0=V, 1=S), am into transmission type (0=automatic & 1=manual), gear=(3:5) & carb=(1,2,3,4,6,8) factor levels

VII. Descriptive Statistics

VIII. Exploratory Analysis:

Add narrative here!!

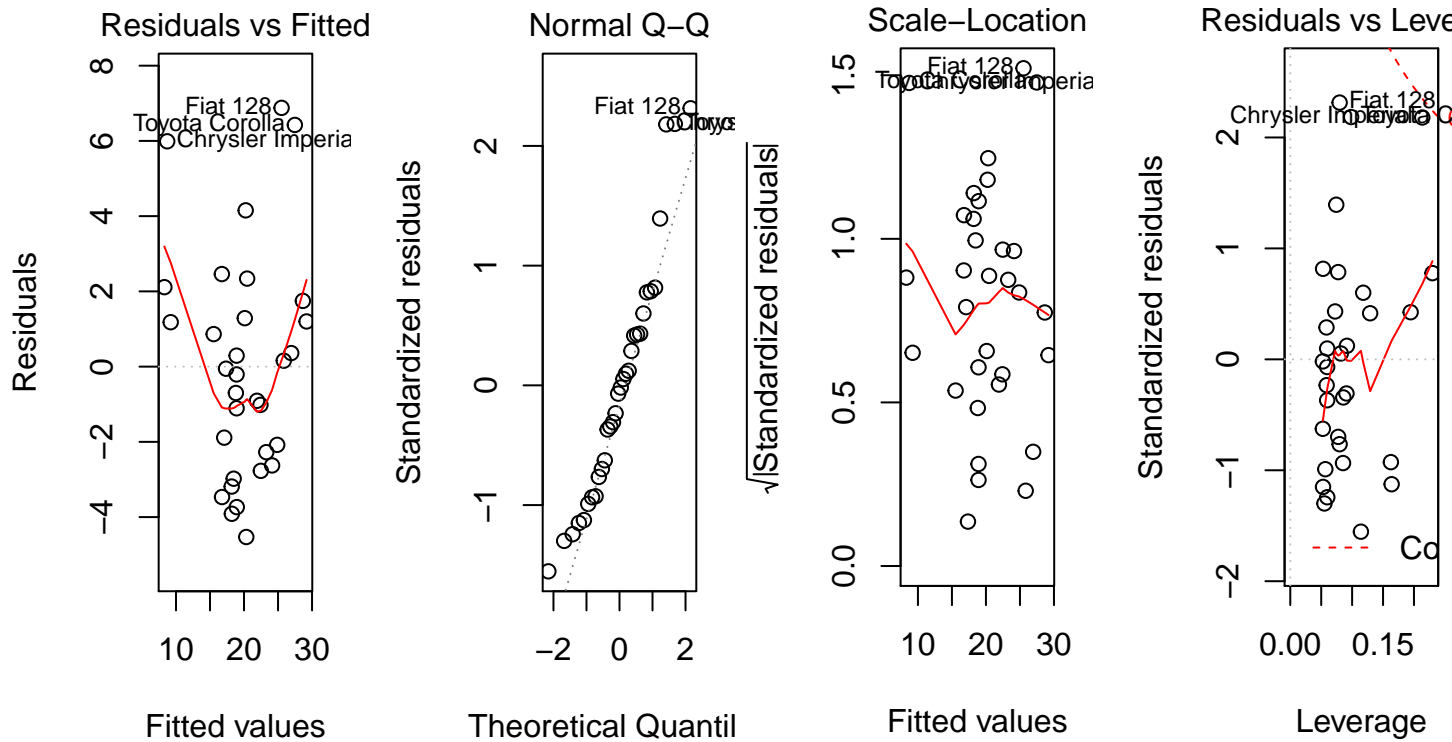
For additional exploratory graphical analysis see Appendix A, Figures 1:4

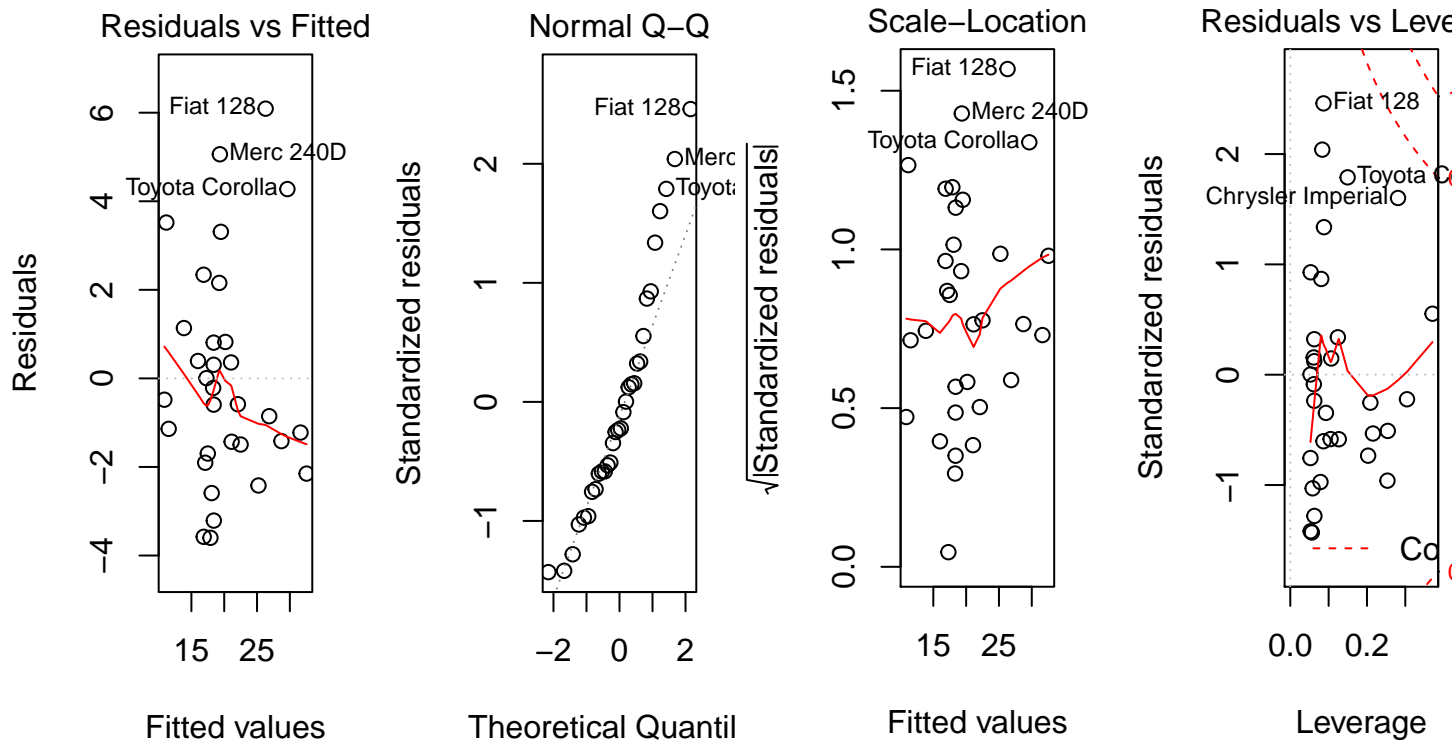
IX. Statistical Modeling, Regression & Model Fit:

Assumptions:

- A A correlation to mpg ratings may exist among multiple variables
- B

Bivariate Linear Model





Multivariate LM (all vars) Fitted Plot

Multivariate LM (all vars) Fitted & Adjusted

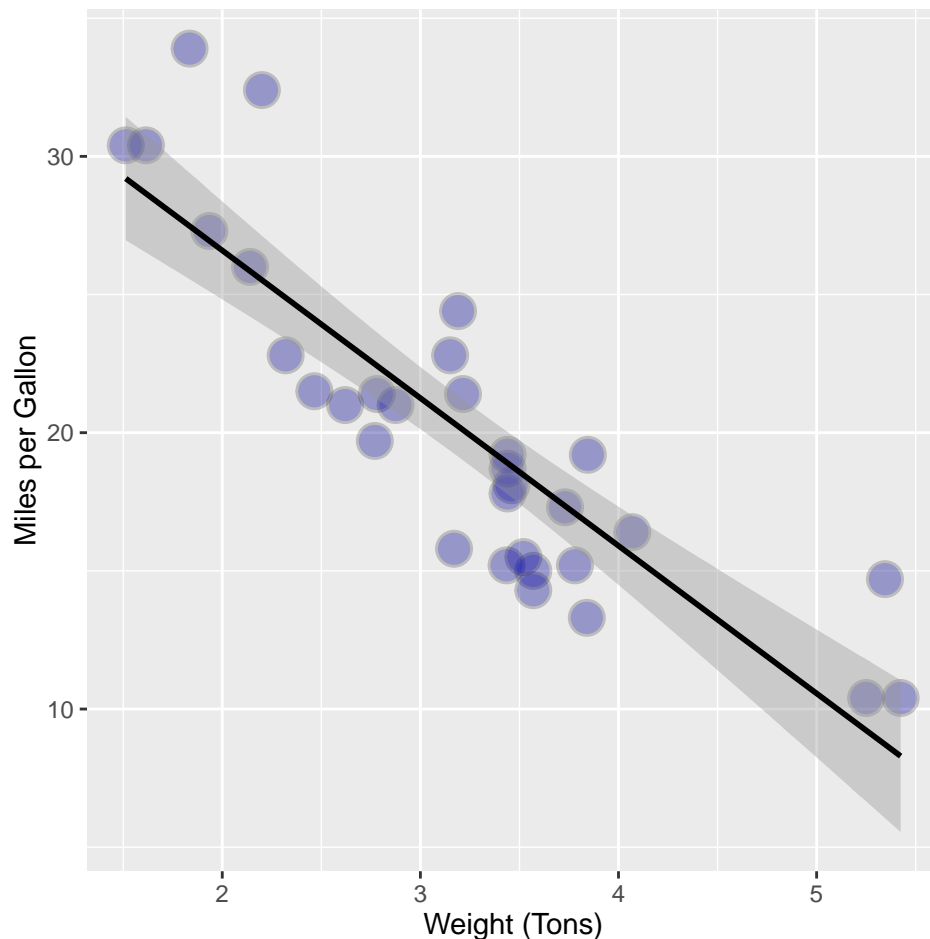
(Note: the variable qsec appears to be significant at this point)

MultivariateLM(allvars)VIF(varianceInflationFactor)

(Note: at this point qsec no longer appears significant; but drat now appears to be instead, and notice the variance inflation figures of the other variables now seem to warrant removing them from the model)

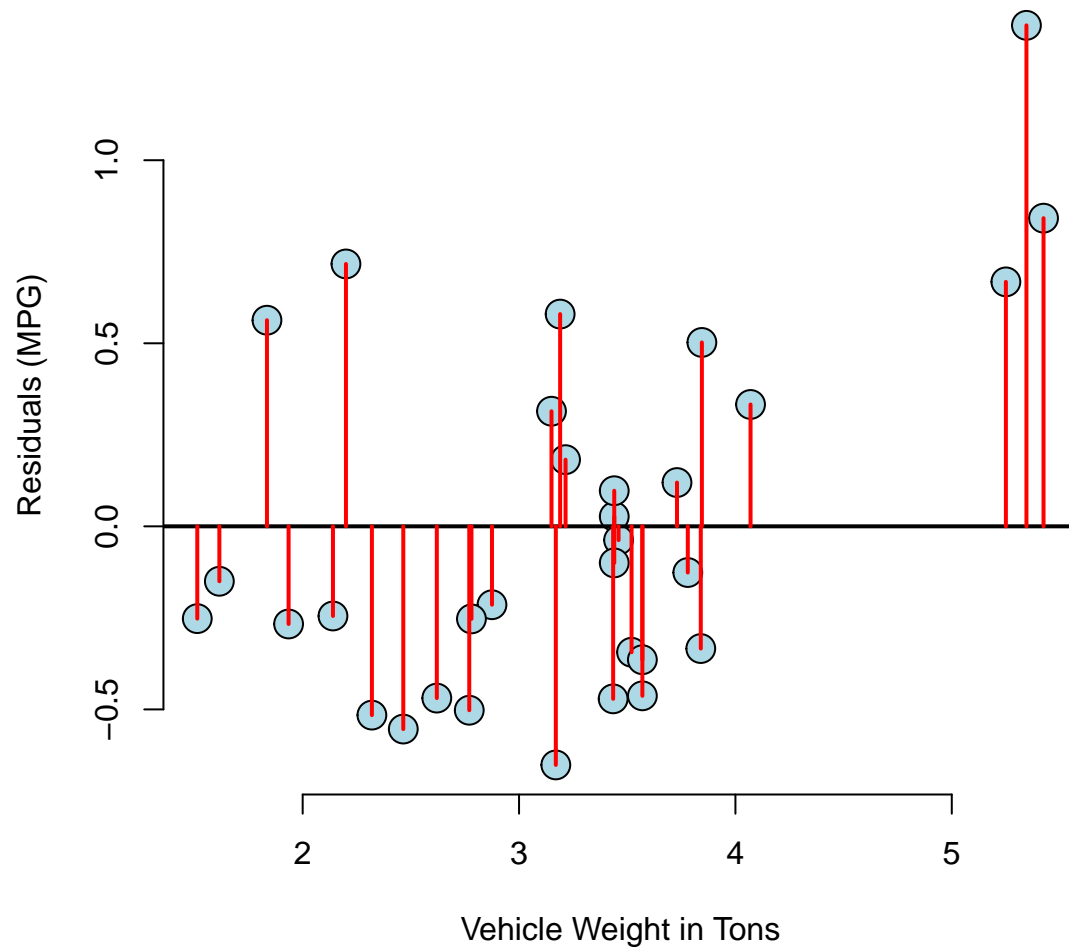
Multivariate LM Nested & ANOVA table (Note: here that with the nested model approach that models 2 & 4 are indicated to be significant)

Best Fit Modeling (Note: comparing the following three models and using anova table test we find that model fbf3 is highly significant and therefore discard the other two models)



NEXT PROCESS TO CHECK!!!! IS RESIDUALS TO IDENTIFY POOR MODEL FIT??

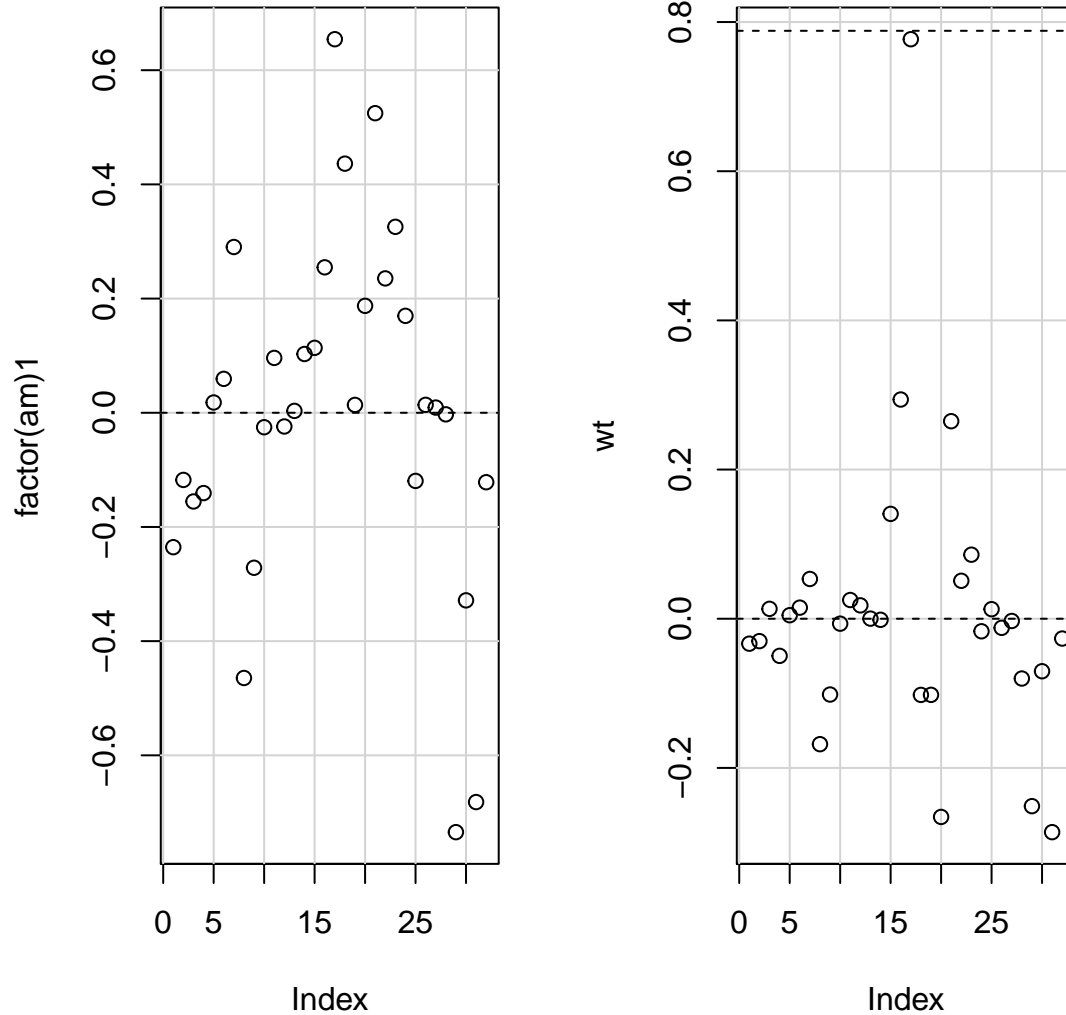
```
y <- mtcars$mpg
x <- mtcars$wt
n <- length(y)
fw <- lm(x ~ y, data = mtcars)
e <- resid(fw) # ;e;plot(e);sum(e)
plot(x,e,
      xlab = "Vehicle Weight in Tons",
      ylab = "Residuals (MPG)",
      bg = "lightblue",
      col = "black", cex = 2, pch = 21, frame = FALSE)
abline(h = 0, lwd = 2)
for(i in 1:n)
  lines(c(x[i], x[i]), c(e[i], 0), col = "red", lwd = 2)
```



NEXT PROCESS: R^2 the percentage of total variability explained by the linear relationship with the predictor !!!!

```
## Warning: In lm.fit(x, y, offset = offset, singular.ok = singular.ok, ...) :
## extra argument 'family' will be disregarded
```

dfbeta Plots



ISSUE NEEDS RESOLUTION

	Estimate	Std. Error	t value	Pr(> t)
## (Intercept)	37.32155131	3.0546385	12.21799285	5.843477e-13
## am	-0.02361522	1.5456453	-0.01527855	9.879146e-01
## wt	-5.35281145	0.7882438	-6.79080719	1.867415e-07

[1] 31.07411 43.56899

[1] -3.184815 3.137584

Find something with levels

X. 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

XI. Inference:

Hypothesis':
A H_0 = The difference between Automatic and Manual transmission MPG = 0
B H_a = The difference between Automatic and Manual transmission MPG \neq 0
C Desired confidence interval = .95 (one sided) ??

XII. Conclusions / Recommendations:

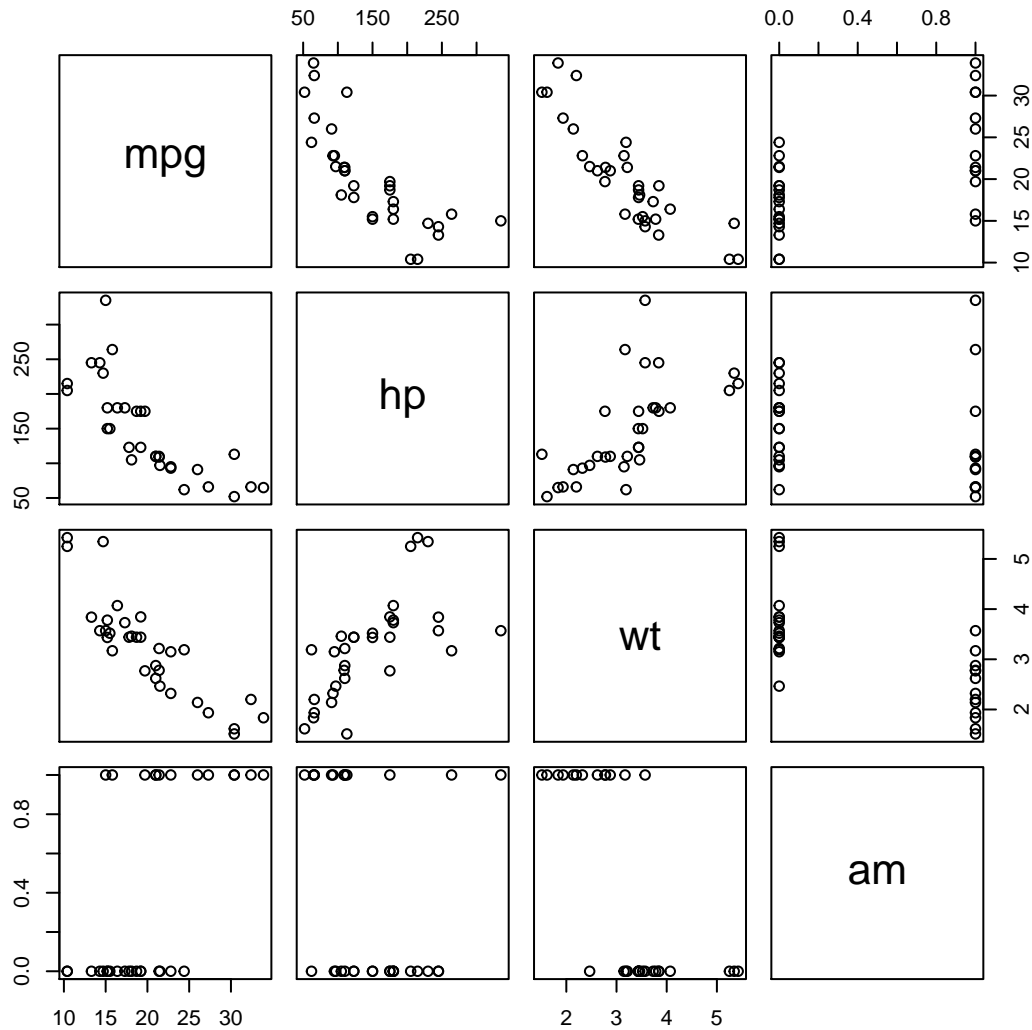
A
B

XIII. Are there other alternative analyses?

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

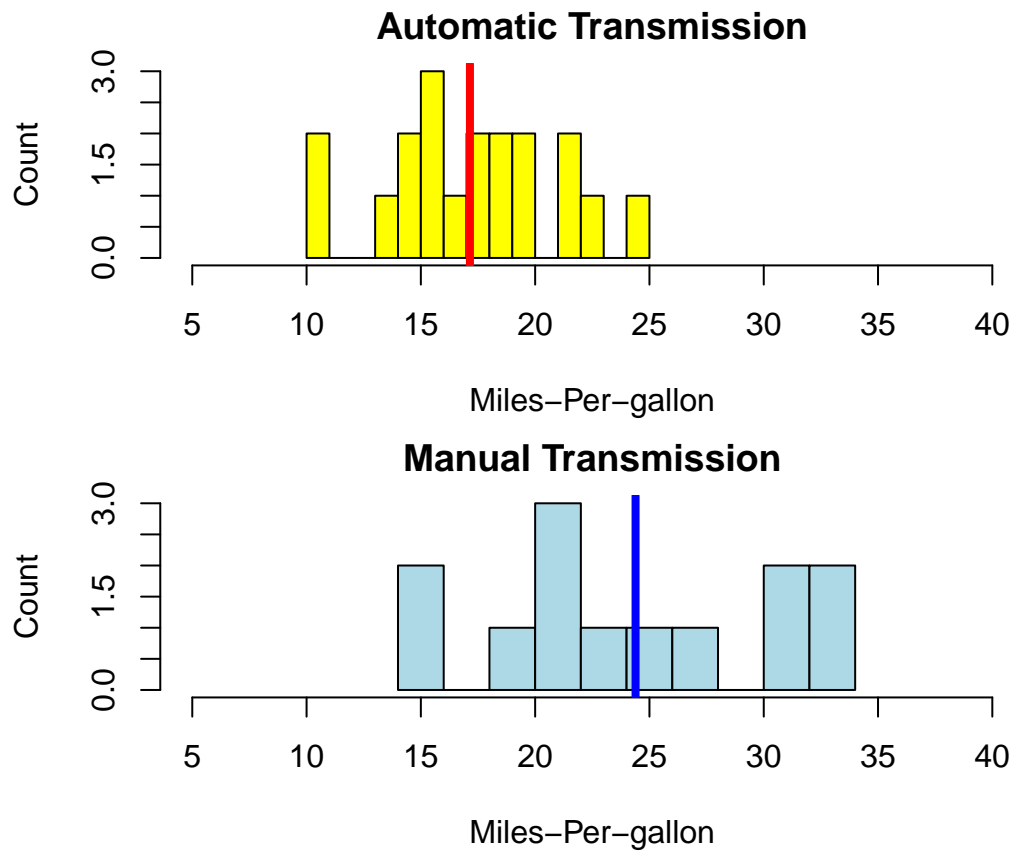
XIV. Appendix A: “Exploratory Graphical Analysis”

Pairs Plot



Histograms Plot

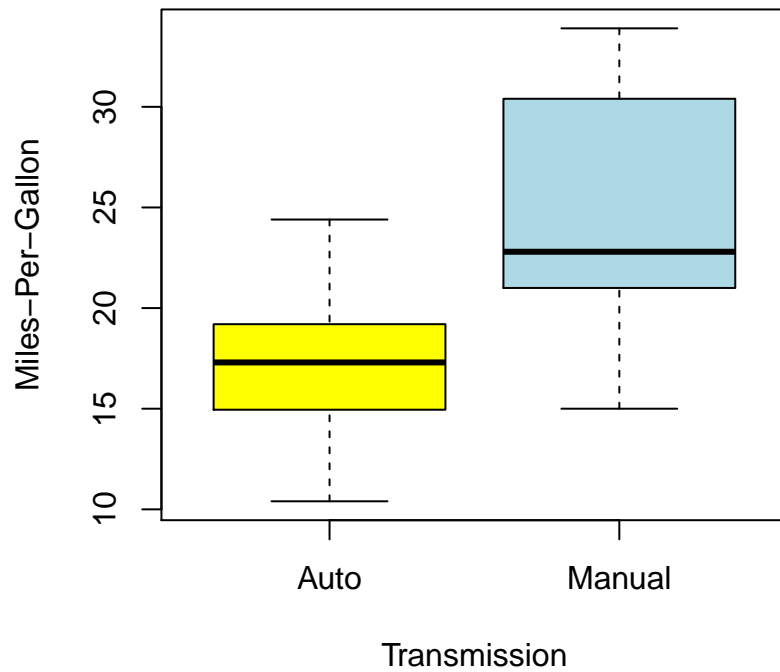
```
##          mpg cyl disp  hp drat   wt  qsec    vs  am gear carb
## Mazda RX4      21.0   6  160  110 3.90 2.620 16.46 V-block Manual    4    4
## Mazda RX4 Wag  21.0   6  160  110 3.90 2.875 17.02 V-block Manual    4    4
## Datsun 710     22.8   4  108   93 3.85 2.320 18.61 S-block  Manual    4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



Box Plot

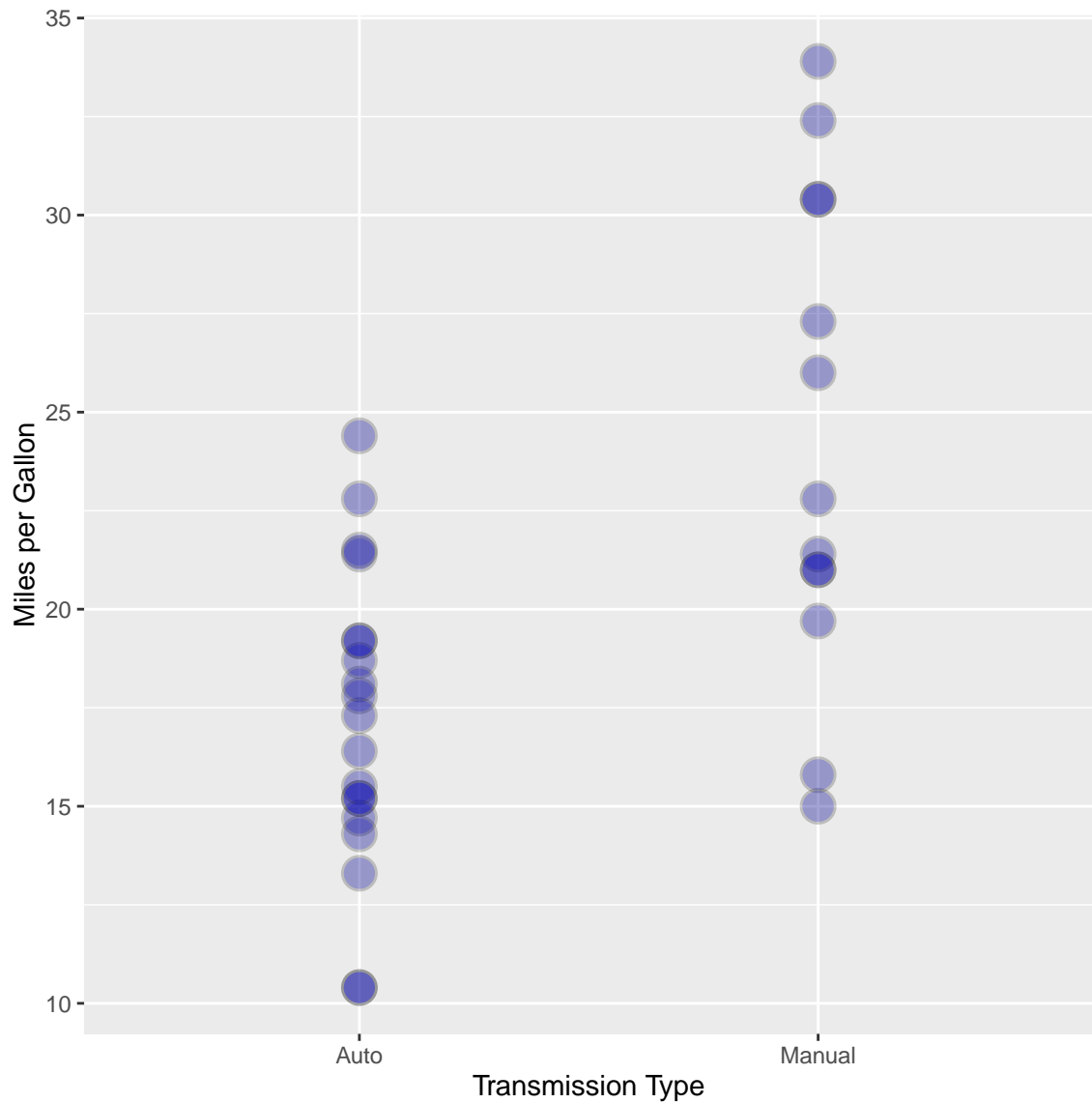
##	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
## Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	V-block	Manual	4	4
## Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	V-block	Manual	4	4
## Datsun 710	22.8	4	108	93	3.85	2.320	18.61	S-block	Manual	4	1

Automatic vs Manual Transmission, Miles Per Gallon



Simple Linear Regression Plot

```
##           Estimate Std. Error  t value    Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## amManual     7.244939   1.764422  4.106127 2.850207e-04
```



NOTE: use the cut function by 3 on MPG

=== END ===

##