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?
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?

III. 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
```

IV. Software Environment: & System - session Info:

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

V. 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)

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 (view first & last three rows & factor variables 2 & 8:11)

```
ordmtcars <<- mtcars[order(mtcars$mpg,decreasing = FALSE),]
rbind(head(ordmtcars,3),tail(ordmtcars,3))
ordmtcars
```

VIII. Exploratory Analysis:

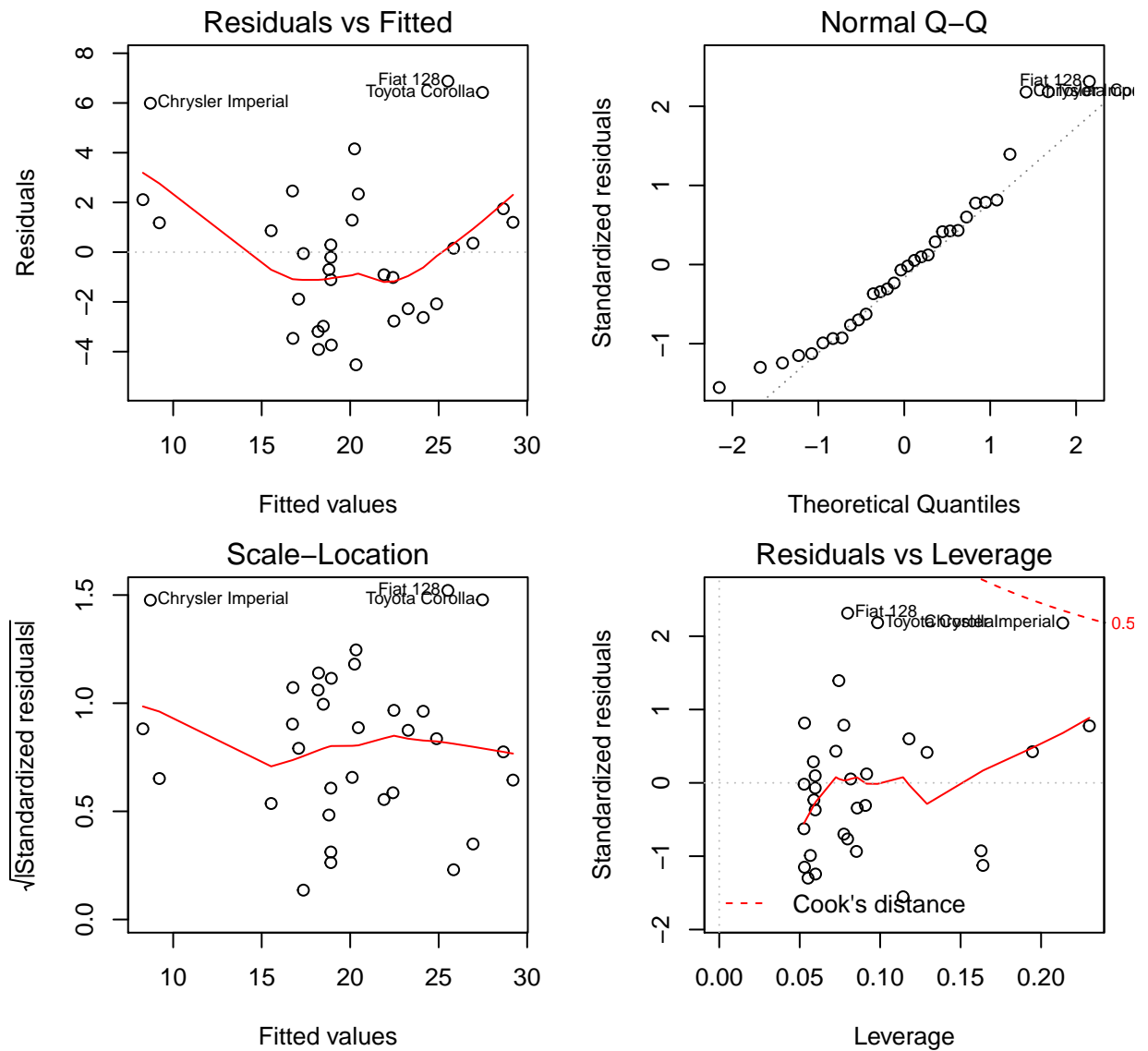```
See Appendix A, Figures 1:4
Add narrative here!!
```

---

## IX. Statistical Modeling, Regression & Model Fit:
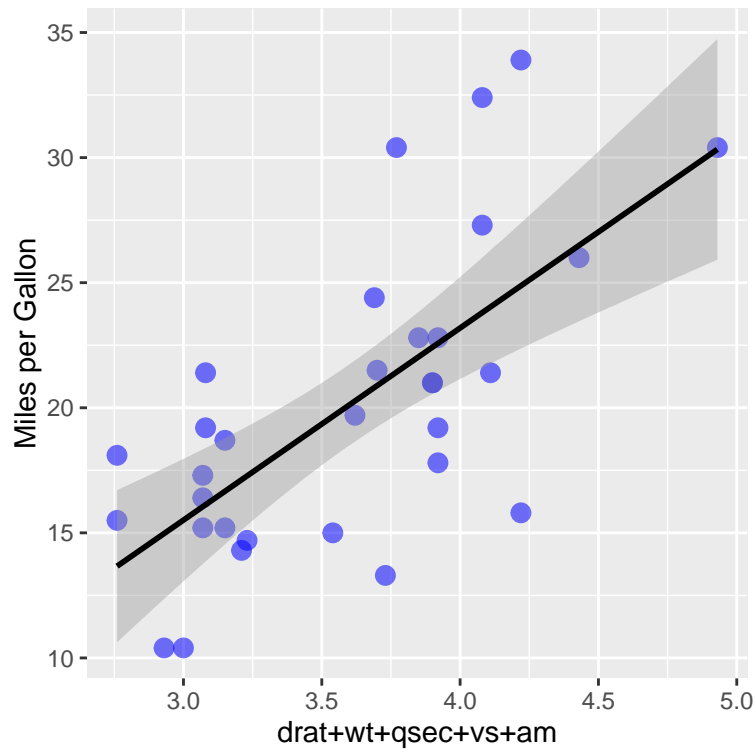
```
Assumptions:
        A A correlation to mpg ratings may exist among multiple variables
        B
```

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

## Residuals vs Fitted

Chrysler Imperial
Fiat 128
Toyota Corolla

Residuals

Fitted values

## Normal Q–Q

Fiat 128
Chrysler Imperial

Standardized residuals

Theoretical Quantiles

## Scale–Location

Chrysler Imperial
Fiat 128
Toyota Corolla

$\sqrt{|\text{Standardized residuals}|}$

Fitted values

## Residuals vs Leverage

Fiat 128
Toyota Corolla Chrysler Imperial
0.5

Standardized residuals

- - - Cook's distance

Leverage

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



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

# Multivariate Linear Model Adjusted
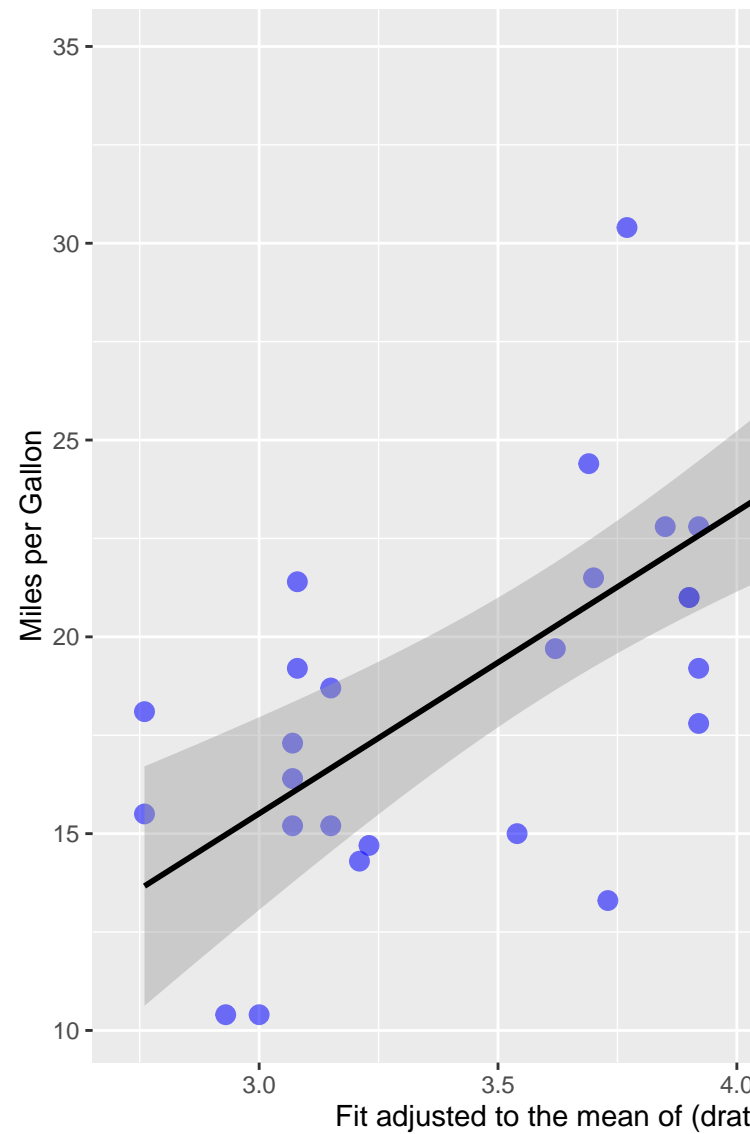


(Note: the variable qsec appears to be significant at this point)

Multivariate LM Nested & ANOVA table (Note: 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 (Note: 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.

```r
library(car)
data("mtcars")
fbf1 <- lm(mpg ~ hp + drat + am, data = mtcars)
vif(fbf1)
sqrt(vif(fbf1))
library(ggplot2)
par(mfrow = c(1,1), mar = c(4,4,2,2)) # set margin
g <- ggplot(mtcars, aes(x = hp + drat +  wt + am, y = mpg),)
g <- g + xlab(">hp + drat + am<")
g <- g + ylab("Miles per Gallon")
g <- g + geom_point(size = 3.0, col = "blue", alpha = 0.55)
g <- g + geom_smooth(method = "lm", col = "black")
g
```
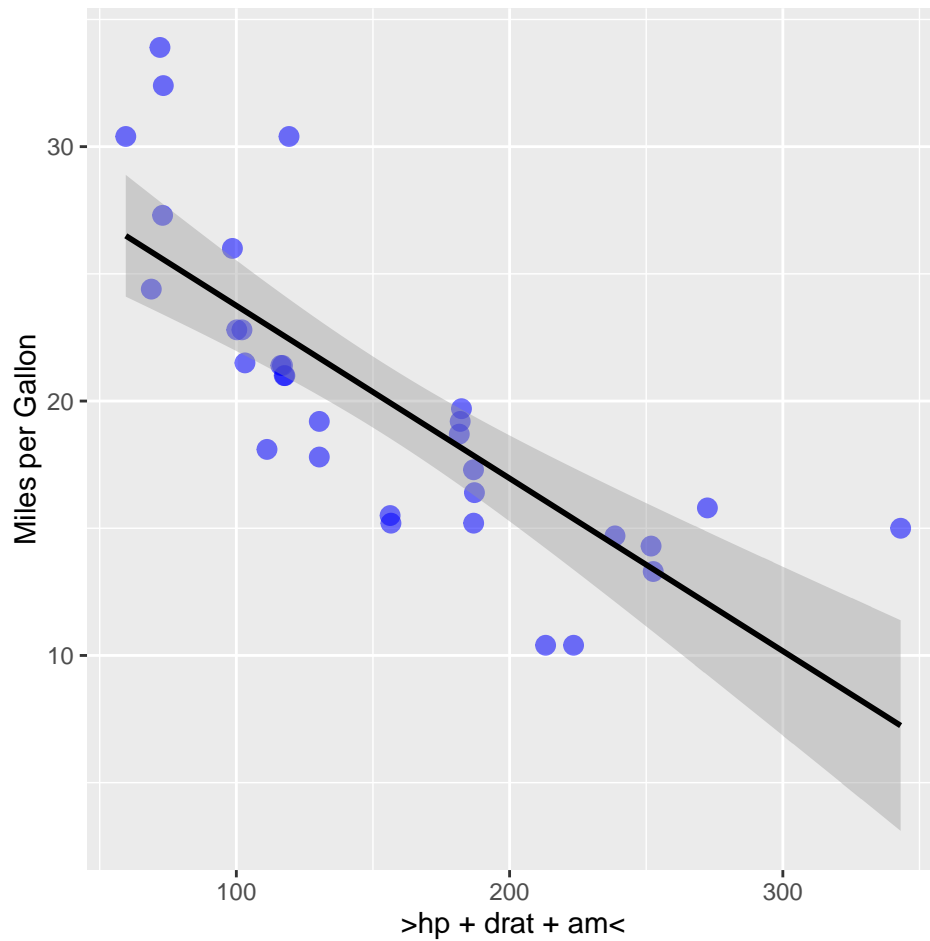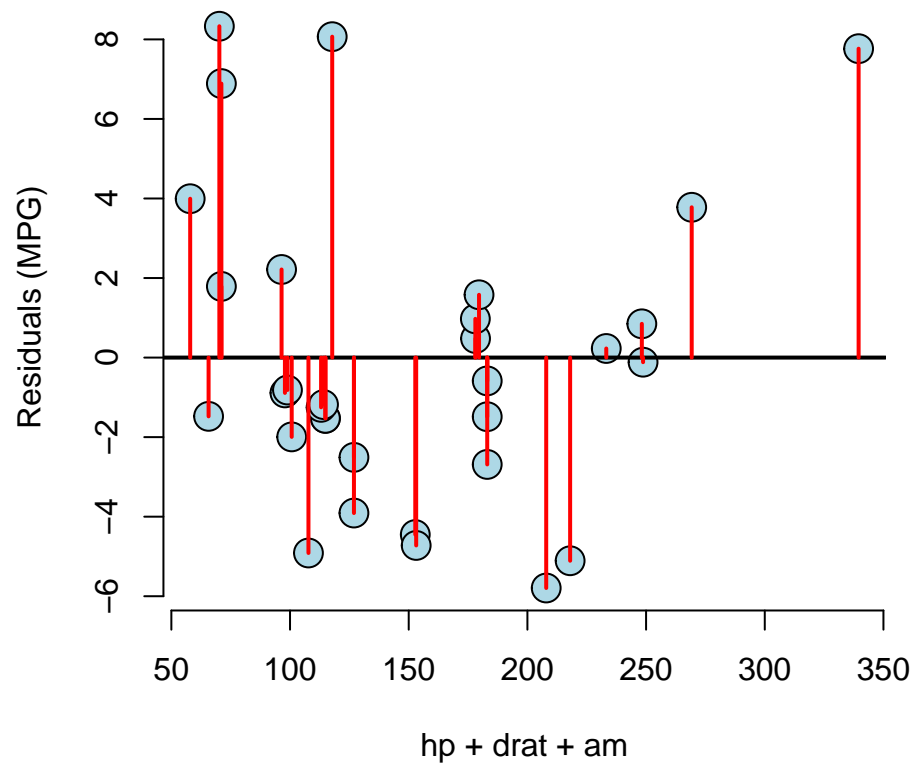
DIAGNOSTICS A

```r
y <- mtcars$mpg
x <- (mtcars$hp + mtcars$drat + mtcars$am)
n <- length(y)
fw <- lm(y ~ x, data = mtcars)
e <- resid(fw)# ;e;plot(e);sum(e)
plot(x,e,
     xlab = "hp + drat + am",
     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)
```
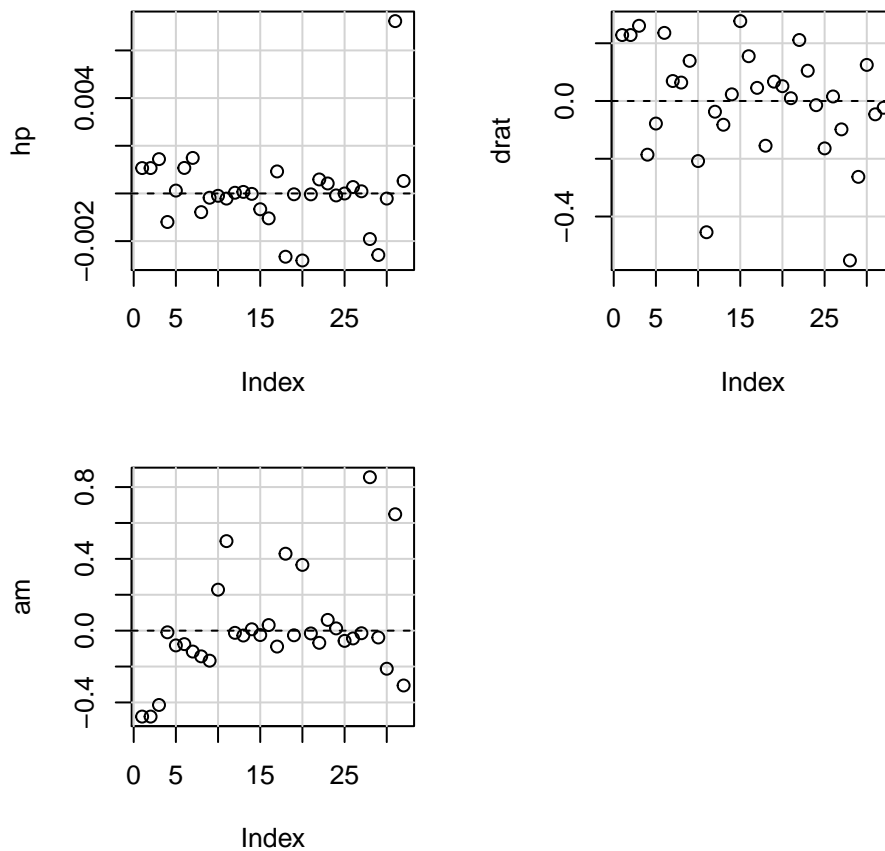
DIAGNOSTICS B

```r
dfbetaPlots(lm(mpg ~ hp + drat + am, data = mtcars, family = binomial))
```

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

dfbeta Plots


```
par(mfrow = c(4, 4), mar = c(4.0, 3.0, 3.0, 3.0))
```

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

```
fbf3 <- lm(mpg ~ hp + drat + factor(am) , data = mtcars)
coef(summary(fbf3))
```

```
##                  Estimate  Std. Error   t value     Pr(>|t|)
## (Intercept) 19.66504055 5.687565112   3.457550 1.760144e-03
## hp          -0.05456235 0.008507916  -6.413128 6.065931e-07
## drat         1.89472426 1.508614795   1.255936 2.195182e-01
## factor(am)1  3.97465803 1.489315986   2.668781 1.252064e-02
```

```
confint(fbf3)
```

```
##                     2.5 %      97.5 %
## (Intercept)   8.01459156 31.31548955
## hp           -0.07199002 -0.03713467
## drat         -1.19553306  4.98498158
## factor(am)1   0.92393253  7.02538353
```

NEXT PROCESS: Prediction !!!!

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 we
            B Based on the
            C
```

XI. Inference:

```
    Hypothesis':
            A H0 = 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) ??
```
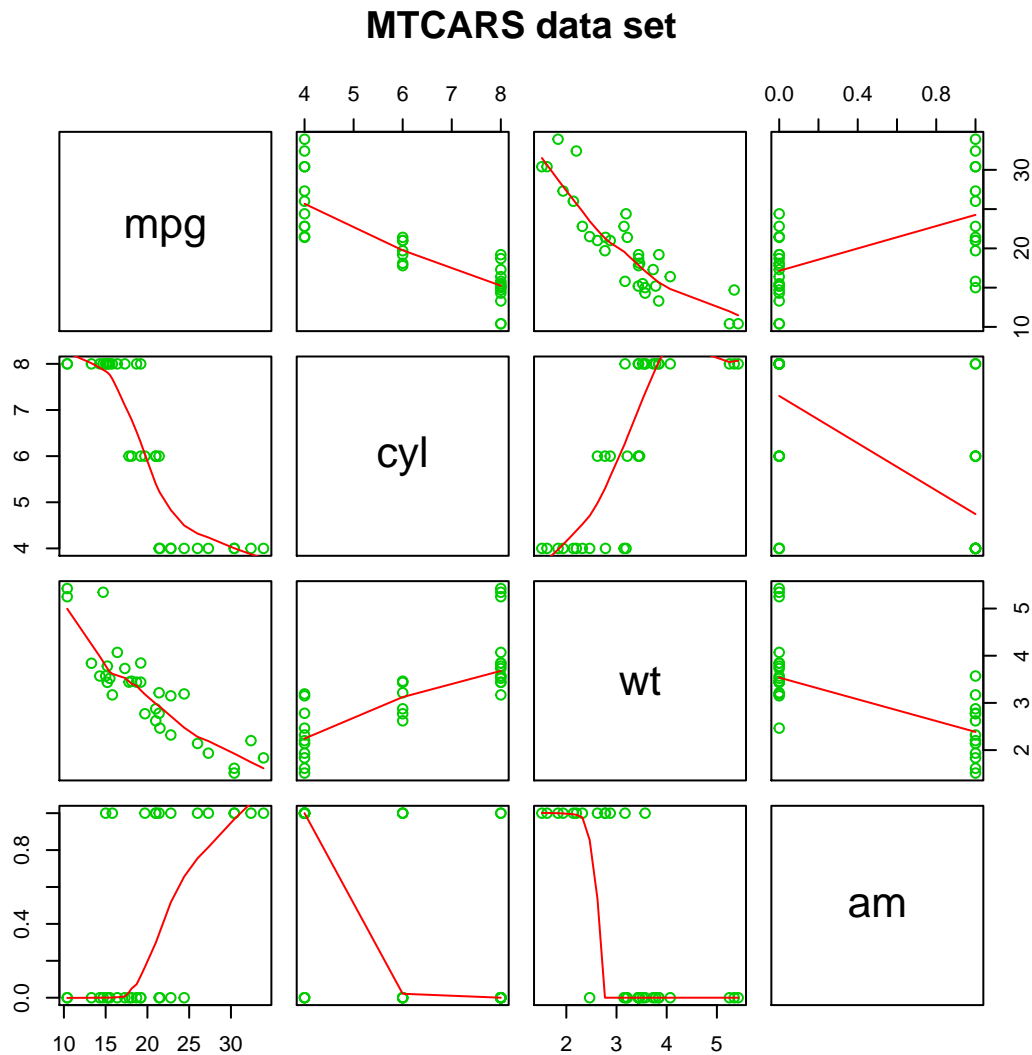
XII. Conclusions / Recommendations:

```
    A
    B
```

XIII. Are there other alternative analyses?

```
    A VIF
    B Challenge the results ?
    C Measures of uncertainty 'e'
```
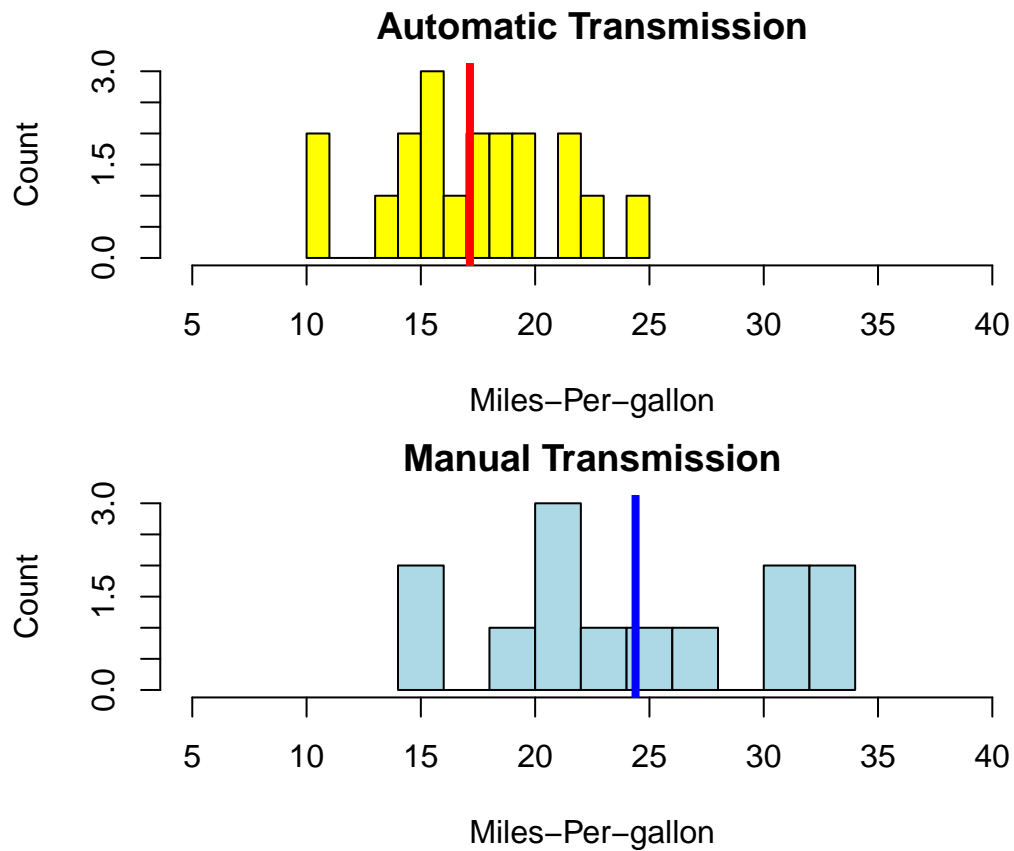
XIV. Appendix A: "Exploratory Graphical Analysis""

Pairs Plot

MTCARS data set

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
```

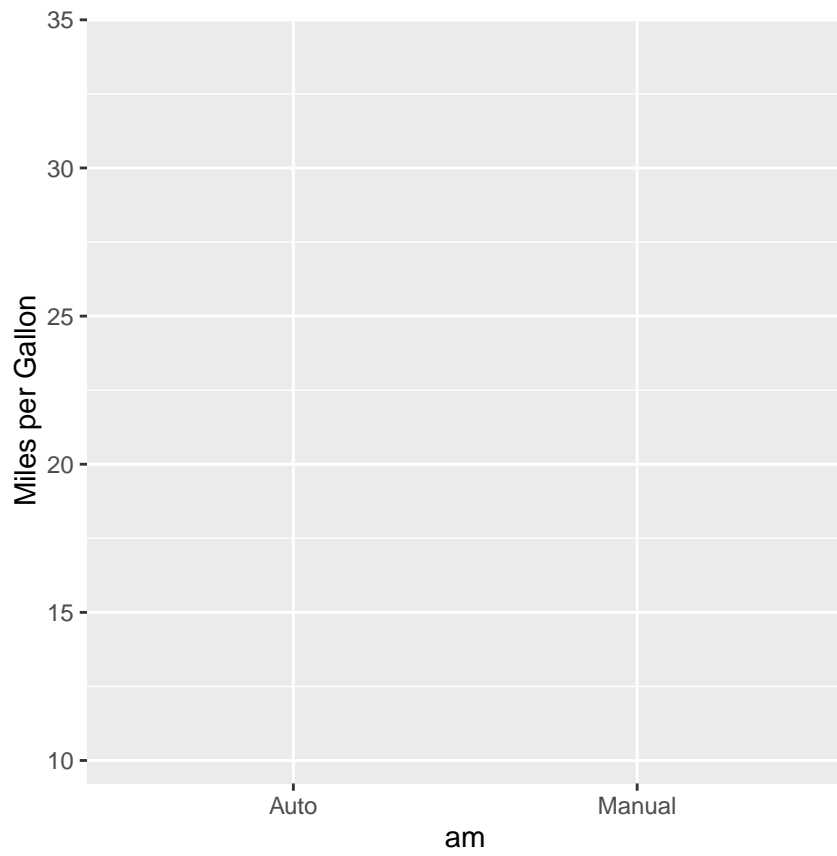# ɔmatic vs Manual Transmission, Miles Per Gallon
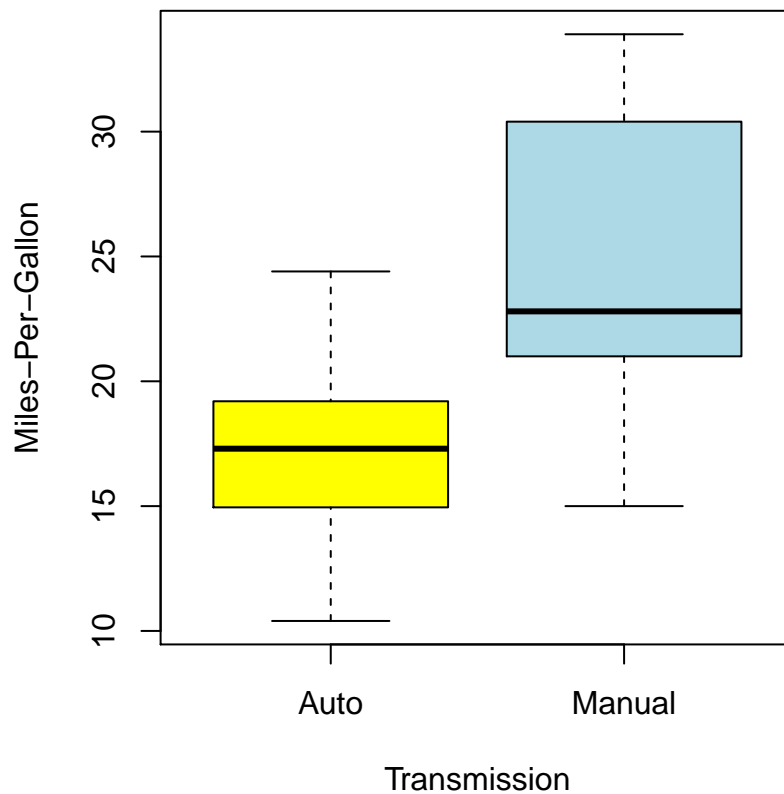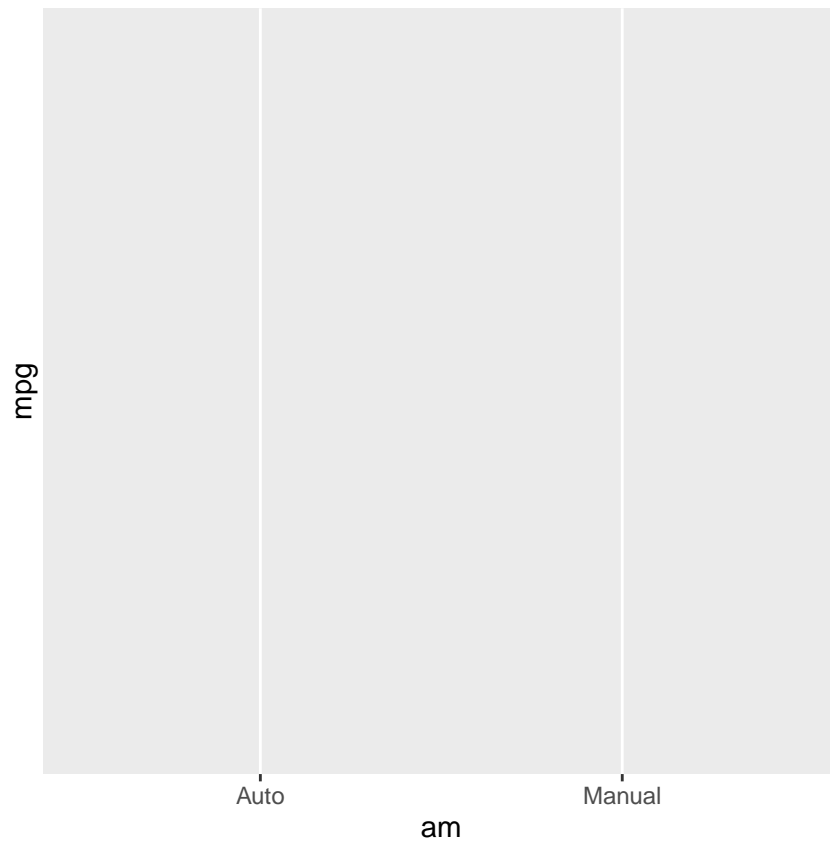

Simple Linear Regression Plot

```
library(ggplot2)
fit <- lm(mpg ~ factor(am), data = mtcars)
par(mfrow = c(1,1), mar = c(4,4,2,2)) # set margin
g <- ggplot(mtcars, aes(x = am, y = mpg),)
g + xlab("Transmission Type")
```
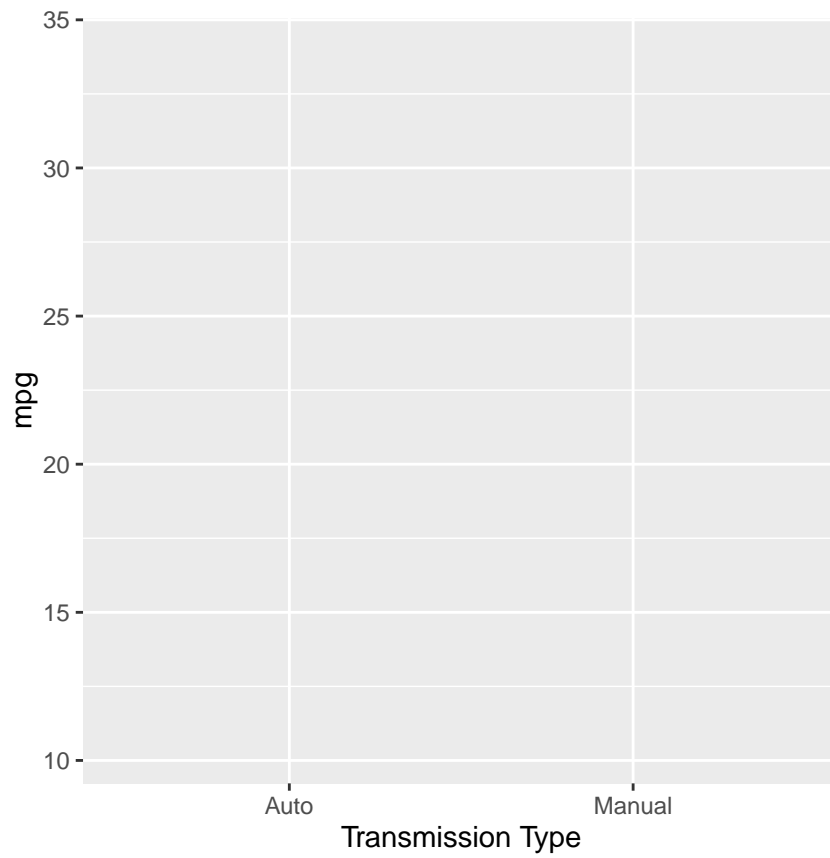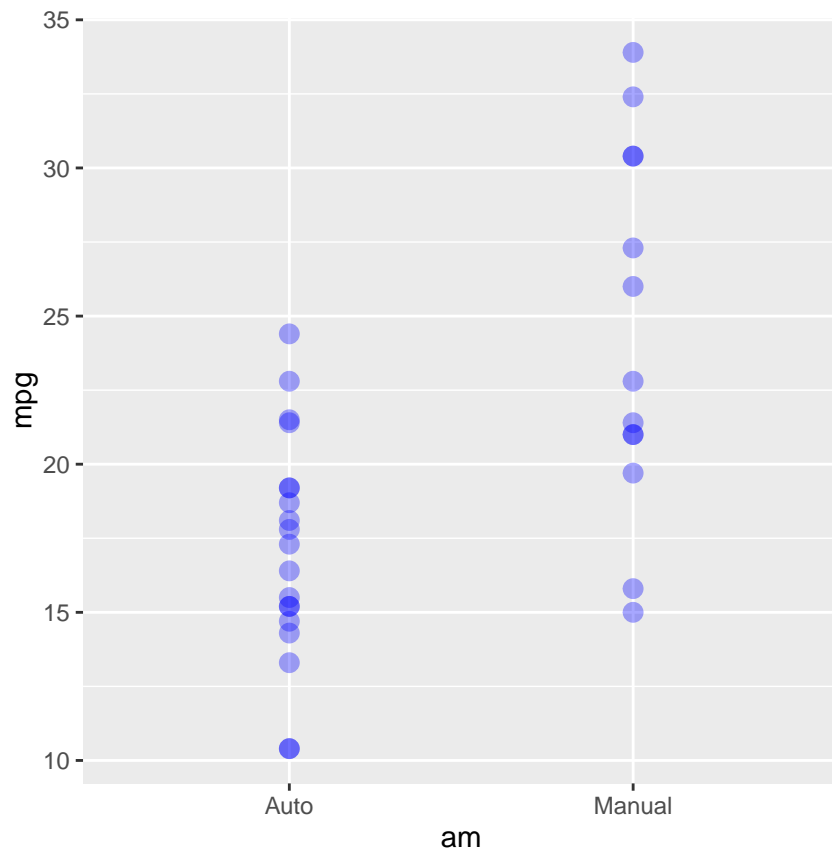
```
g + ylab("Miles per Gallon")
```

```
g + geom_smooth(method = "lm", col = "black")
```
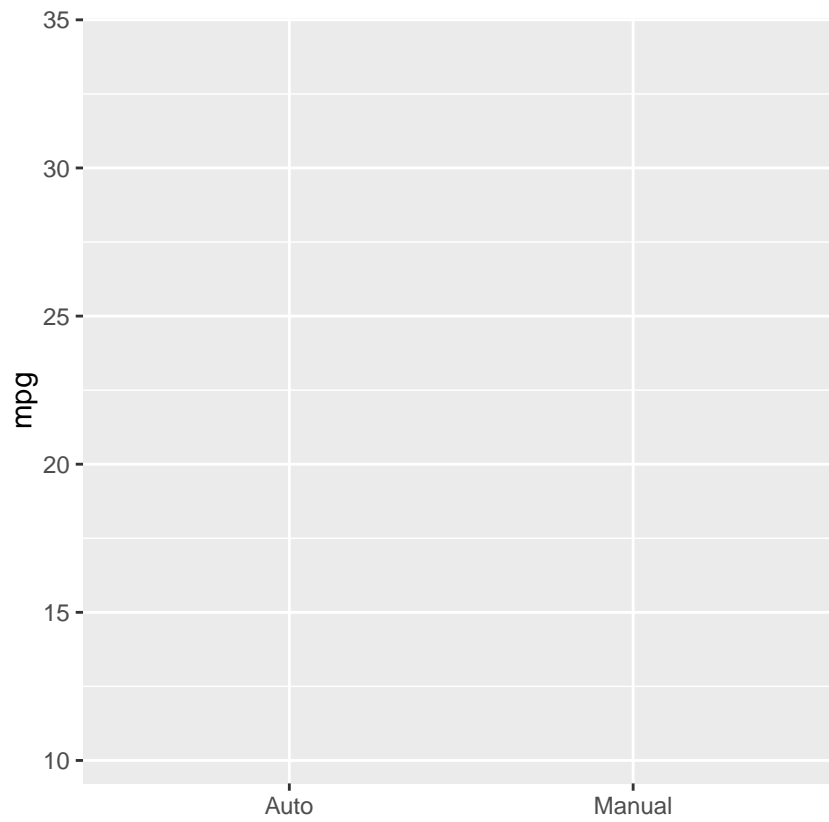
mpg

Auto   Manual

am

```r
g + geom_point(size = 3, col = "blue", alpha = 0.35)
```

g

NOTE: use the cut function by 3 on MPG

=== END ===