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)!!!

YES!!!! Did the student interpret the coefficients correctly?

YES!!!! Did the student do some exploratory data analyses?

Did the student fit multiple models and detail their strategy for model selection?

Did the student answer the questions of interest or detail why the question(s) is (are) not answerable?

Did the student do a residual plot and some diagnostics?

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?

Did the report include an executive summary?

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,        residu
diagnostics; ANOVA, coeficients, confint, correlation, covariance, variance inflation
```

IV. Software Environment:

Set working directory:

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

System - session Info:

```
sessionInfo()
```

V. Accessing Data:

Clean up the work space & get the data:

```
rm(list=ls()); data("mtcars");mtcars
```

VI. Raw Data Overview: Motor Trend 'mtcars' data set: Descriptive Statistics

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 (1000 lbs)
[, 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

VII. Process Data:

Transformationns = Factor variables 8:11; modify variable names for col 8 & 9

```
data("mtcars");for(i in c(2,8:11))mtcars[,i] <- as.factor(mtcars[,i]);mtcars[,]
```

VIII. Exploratory Analysis:

Add narrative here!!

```
data(mtcars);library(car);library(dplyr)
h1 <- sapply(mtcars, range, simplify = TRUE);h1;mu1 <- sapply(mtcars, mean, simplify = TRUE);mu1
```

```
##       mpg cyl  disp  hp drat    wt qsec vs am gear carb
## [1,] 10.4   4  71.1  52 2.76 1.513 14.5  0  0    3    1
## [2,] 33.9   8 472.0 335 4.93 5.424 22.9  1  1    5    8
```

```
##        mpg        cyl       disp         hp       drat         wt
##  20.090625   6.187500 230.721875 146.687500   3.596563   3.217250
##       qsec         vs         am       gear       carb
##  17.848750   0.437500   0.406250   3.687500   2.812500
```

```
c1 <- sapply(mtcars, sd, simplify = TRUE);c1;c2 <- sapply(mtcars, var, simplify = TRUE);c2
```

```
##         mpg         cyl        disp          hp        drat          wt
##   6.0269481   1.7859216 123.9386938  68.5628685   0.5346787   0.9784574
##        qsec          vs          am        gear        carb
##   1.7869432   0.5040161   0.4989909   0.7378041   1.6152000
```

```
##          mpg          cyl         disp           hp         drat
## 3.632410e+01 3.189516e+00 1.536080e+04 4.700867e+03 2.858814e-01
##           wt         qsec           vs           am         gear
## 9.573790e-01 3.193166e+00 2.540323e-01 2.489919e-01 5.443548e-01
##         carb
## 2.608871e+00
```

```
ordmtcars <- mtcars[order(mtcars$mpg,decreasing = TRUE),]
ordmtcars2 <- mtcars[order(mtcars$mpg,decreasing = FALSE),]
rbind(head(ordmtcars2), head(ordmtcars))
```

```
##                     mpg cyl  disp  hp drat    wt  qsec vs am gear carb
## Cadillac Fleetwood  10.4   8 472.0 205 2.93 5.250 17.98  0  0    3    4
## Lincoln Continental 10.4   8 460.0 215 3.00 5.424 17.82  0  0    3    4
## Camaro Z28          13.3   8 350.0 245 3.73 3.840 15.41  0  0    3    4
## Duster 360          14.3   8 360.0 245 3.21 3.570 15.84  0  0    3    4
## Chrysler Imperial   14.7   8 440.0 230 3.23 5.345 17.42  0  0    3    4
## Maserati Bora       15.0   8 301.0 335 3.54 3.570 14.60  0  1    5    8
## Toyota Corolla      33.9   4  71.1  65 4.22 1.835 19.90  1  1    4    1
## Fiat 128            32.4   4  78.7  66 4.08 2.200 19.47  1  1    4    1
## Honda Civic         30.4   4  75.7  52 4.93 1.615 18.52  1  1    4    2
## Lotus Europa        30.4   4  95.1 113 3.77 1.513 16.90  1  1    5    2
## Fiat X1-9           27.3   4  79.0  66 4.08 1.935 18.90  1  1    4    1
## Porsche 914-2       26.0   4 120.3  91 4.43 2.140 16.70  0  1    5    2
```

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

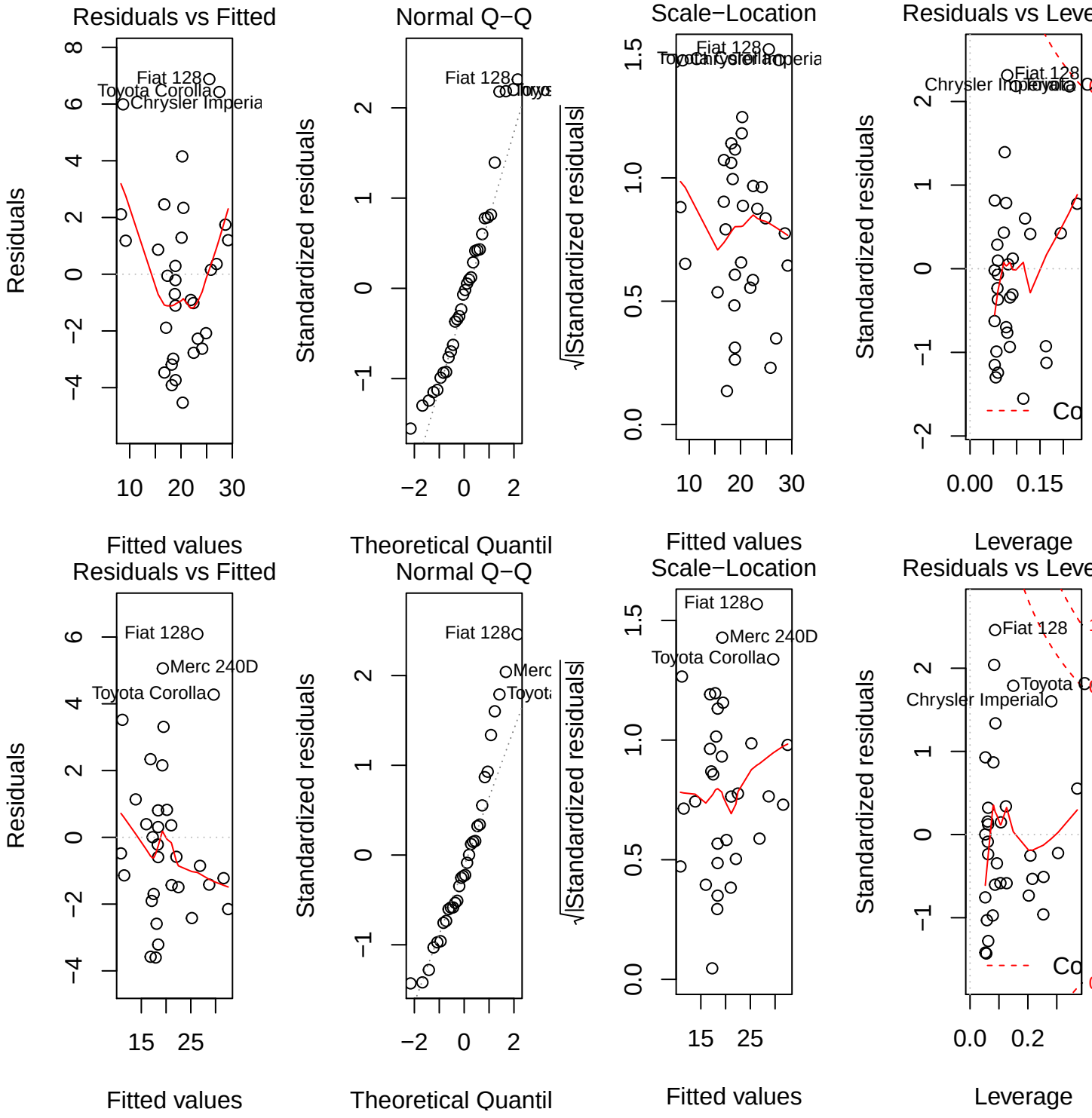
IX. Statistical Modeling, Regression & Model Fit:

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

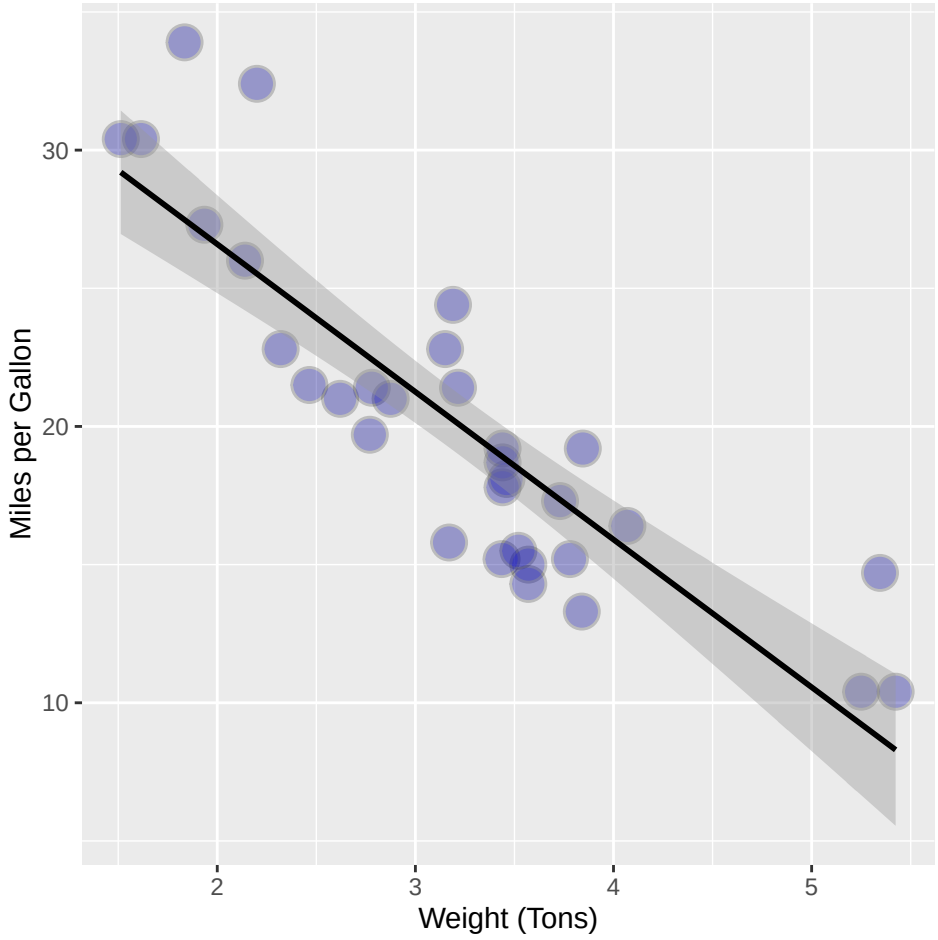
(Note: the variable qsec appears to be significant at this point)

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 varables now seem to warrant removing them from the model)

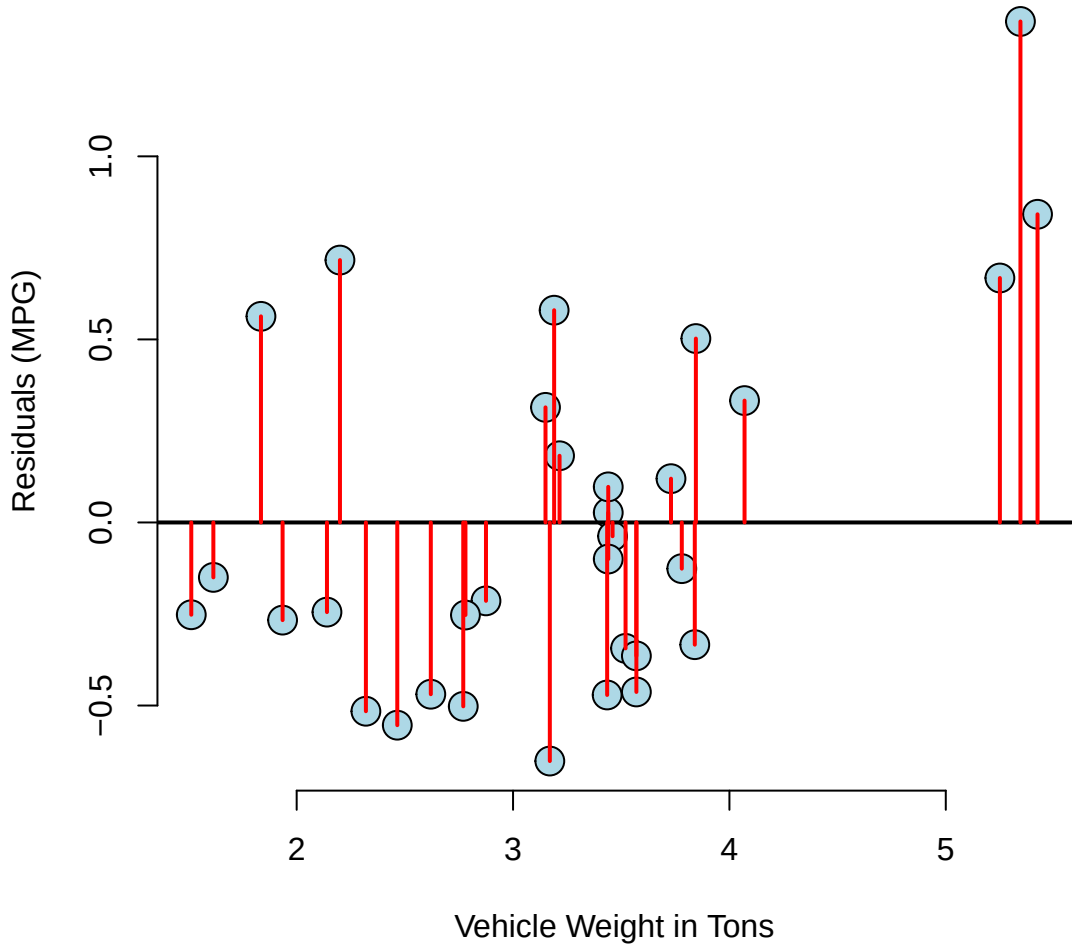
Multivariate LM Nested & ANOVA table (Note: here that with the nested model approach that models 2 & 4 are indicated to be significant)

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)



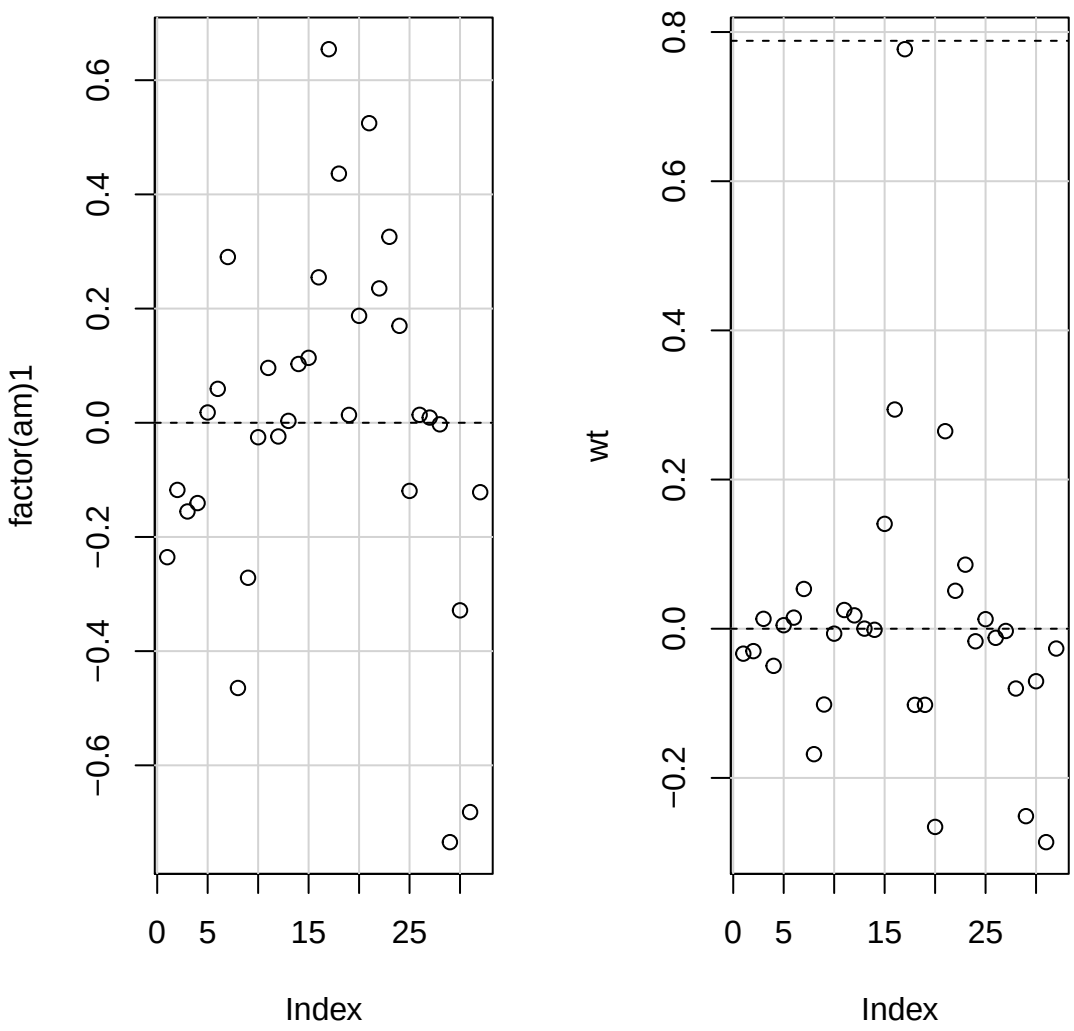
NEXT PROCESS TO CHECK!!!!! IS RESIDUALS TO IDENTIFY POOR MODEL FIT??

```r
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
```



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 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:

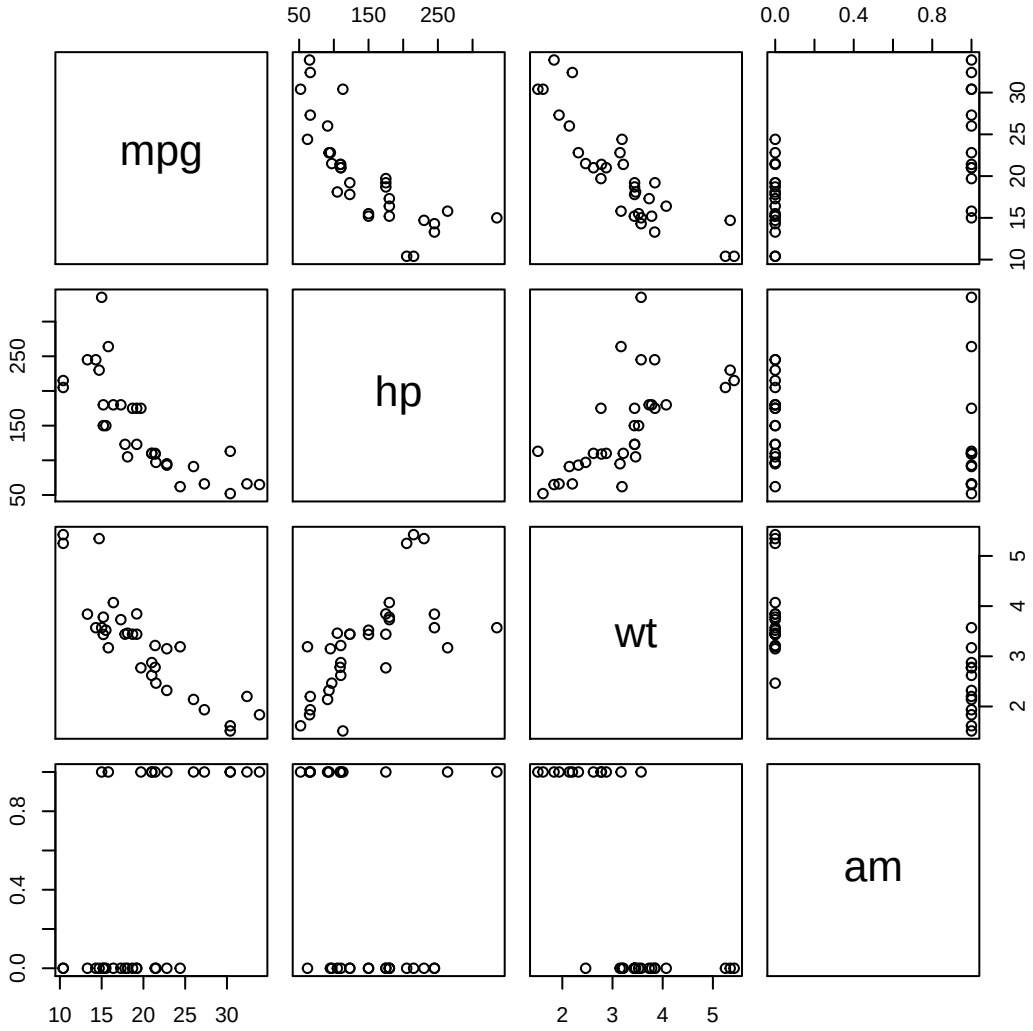
```
A
B
```

XIII. Are there other alternative analyses?

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

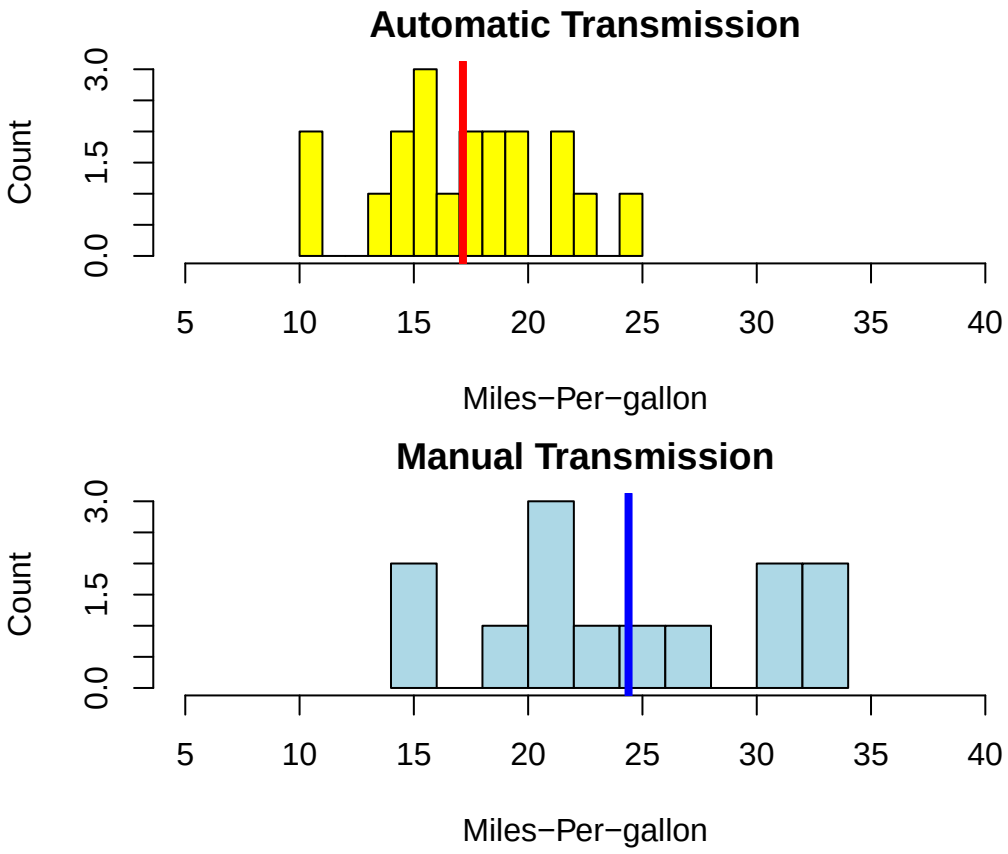
XIV. Appendix A: “Exploratory Graphical Analysis”"

Pairs Plot



Histograms Plot

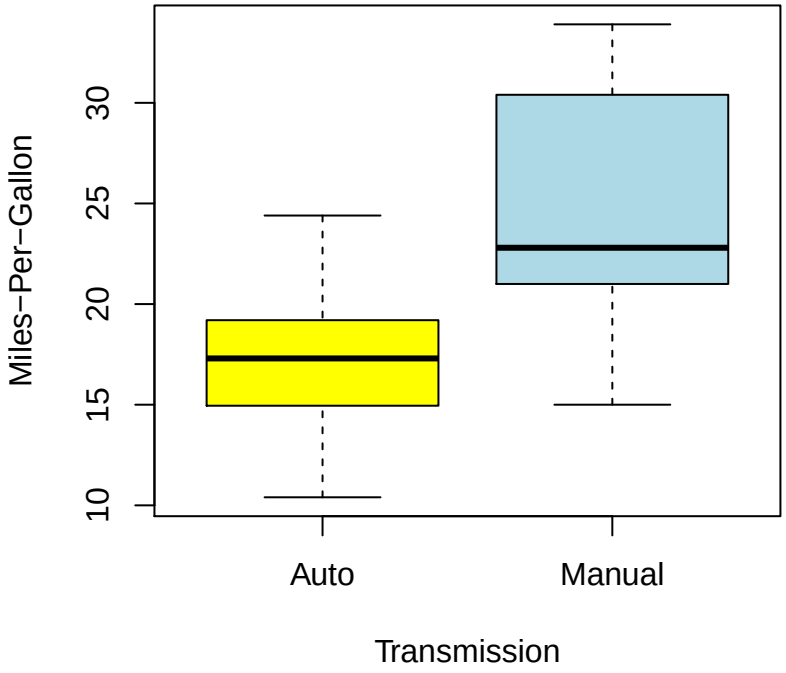
```
##                mpg cyl disp  hp drat    wt  qsec      vs     am gear carb
## Mazda RX4     21.0   6  160 110 3.90 2.620 16.46 V-block Manual    4    4
## Mazda RX4 Wag 21.0   6  160 110 3.90 2.875 17.02 V-block Manual    4    4
## Datsun 710    22.8   4  108  93 3.85 2.320 18.61 S-block Manual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



Box Plot

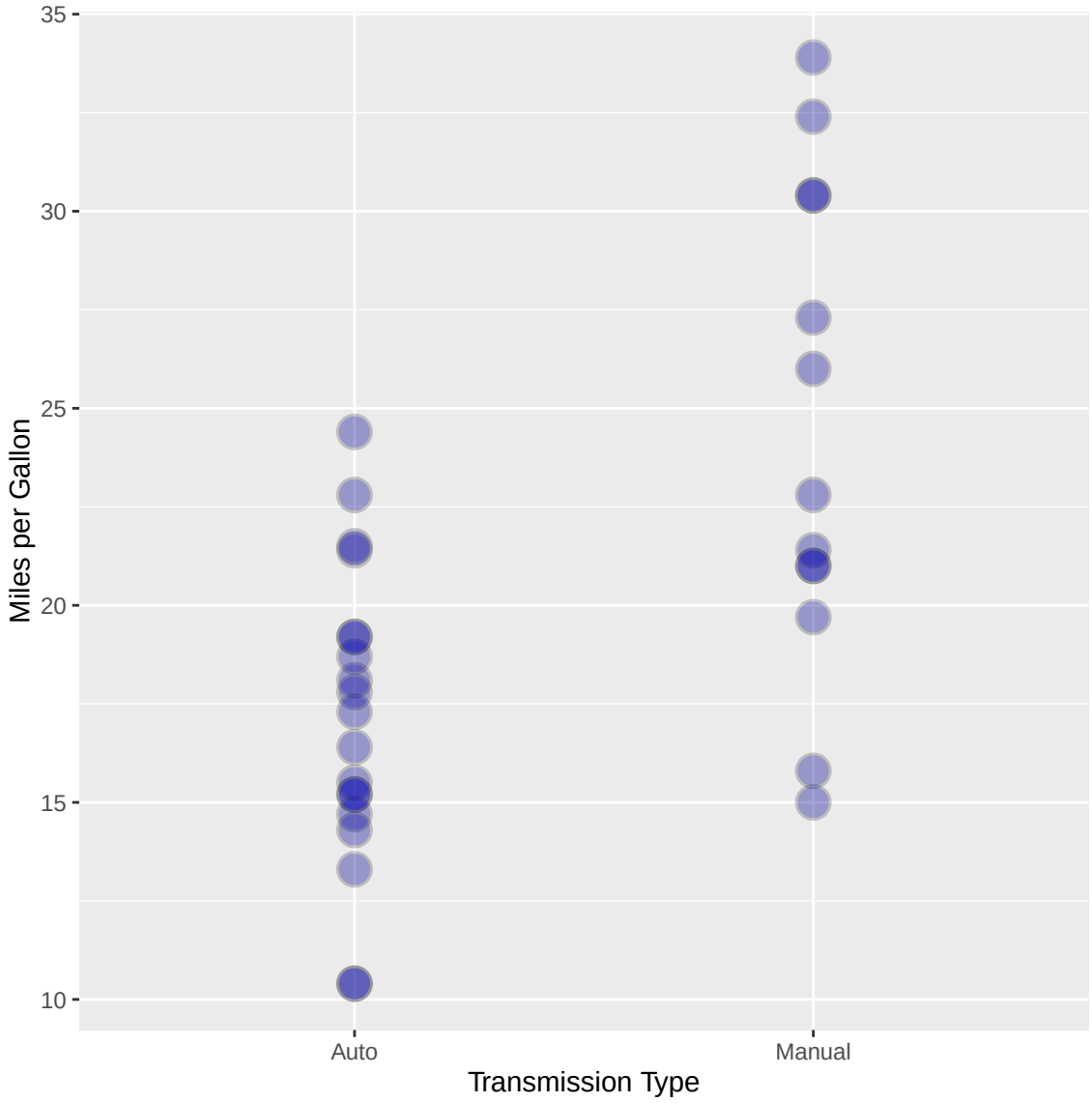
```
##                mpg cyl disp  hp drat    wt  qsec      vs     am gear carb
## Mazda RX4     21.0   6  160 110 3.90 2.620 16.46 V-block Manual    4    4
## Mazda RX4 Wag 21.0   6  160 110 3.90 2.875 17.02 V-block Manual    4    4
## Datsun 710    22.8   4  108  93 3.85 2.320 18.61 S-block Manual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

ɔmatic 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
```


35
30
25
20
15
10
Miles per Gallon
Auto
Manual
Transmission Type

=== END ===