Regression Models Class Project (JHU via Coursera) - July 2014

Leif Ulstrup

July 22, 2014

Executive Summary

The graphical analysis of the relationship of mpg to automobile transmission type using the 1974 Motor Trend data on 32 cars gives some hints that manual transmissions get better mileage for cars of equivalent horsepower; however the linear regression analysis shows that the combination of weight + horsepower + horsepower*wt explains 89% of the predicted mpg and the introduction of transmission type adds no significant additional factor to the prediction of mpg.

The only way to absolutely confirm whether automatic or manual transmissions are more fuel efficient is to compare the same vehicles outfitted with each option for a side by side comparison. Most auto makers provide this comparison information on the web for their cars.

Introduction

Our project assignment for this class (Regression Models at Johns Hopkins U via Coursera - July 2014) is to assume we work for Motor Trend Magazine and are asked to use data from a 1974 Motor Trend magazine survey of car performance for 32 automobiles to answer two questions. The information about this data source "mtcars" can be found here: http://stat.ethz.ch/R-manual/R-devel/library/datasets/html/mtcars.html (http://stat.ethz.ch/R-manual/R-devel/library/datasets/html/mtcars.html)

The questions are:

1. Is an automatic or manual transmission better for MPG (miles per gallon)?
2. Quantify the MPG difference between automatic and manual transmissions?

The challenge in answering these questions is that the data do NOT include automatic AND manual transmission versions of the same cars for comparison. Solely using the data in the mtcars data set to answer this question by trying to normalize the physical characteristics such as weight, horsepower, acceleration, cylinders, etc and then testing for differences in MPG for automatic vs. manual transmissions will likely have a low confidence factor.

The performance factors in the "mtcars" data set are ultimately governed by physics (Newton's Laws), car designer engineering trade-offs, and how the car was driven in testing. Given that, one can refer to the physics that drive the major factors in MPG performance to think through ways to normalize the physical characteristics in answering the question. Per this reference: http://www.driverside.com/auto-library/top_10_factors_contributing_to_fuel_economy-317 (http://www.driverside.com/auto-library/top_10_factors_contributing_to_fuel_economy-317)

Four of these major physical factors are:

1. engine efficiency in converting the energy in a gallon of gasoline into power,
2. mechanical heat loss due to inefficiencies in the power transmission to the tires,
3. rolling friction losses (these are function of the weight of the car and coefficient of rolling friction of the tires), and, most importantly at high speeds (>50mph),
4. drag on the car where the power (energy/unit time) required to overcome

drag = (1/2)*(fluid density of air)*(front surface "drag area" of the car)*(drag coefficient of the car body)* velocity^3

If one assumes that the test driving for the mtcars data approximated the same velocities, then the major difference in energy consumption is based on the front surface of the car * the car's drag coefficient. The drag area and drag coefficients can vary by multiples and create major differences in fuel economy per this reference and these tables: http://en.wikipedia.org/wiki/Automobile_drag_coefficient (http://en.wikipedia.org/wiki/Automobile_drag_coefficient). We do NOT have this critical drag related data in the mtcars data set.

The exploratory data analysis below explores the relationship between MPG and transmission type by first exploring the relationships between MPG and Weight, MPG and Horsepower (HP), Quarter Mile time vs HP, and Quarter Mile time vs Power Ratio (Weight/HP). In each case the data are plotted noting whether the car has an automatic (am=0) or manual transmission (am=1). The analysis shows a number of differences between the automatic and manual car performance but also a number of overlaps in physical parameters where the data show little variation between the two types of transmissions.

The most striking diagram "MPG vs HP" seems to show that manual transmission cars are on a trend line of ~4 MPG superior performance when horsepower is the assumed factor affecting MPG; however, the data is limited and high horsepower manual transmission cars are relatively low weight cars (sports cars) vs. the automatic transmission set.

The 13 of 32 manual transmission cars average 24.39 mpg (with a standard deviation of 6.17) vs the automatics which are 19 of 32 cars and average 17.15 mpg (with a standard deviation of 3.834); however, it is also easy to see on the MPG vs Weight chart that the majority of the manual transmission cars are the much lighter cars. 11 of the 13 manual transmission cars are below 3000 lbs while only 1 of the 19 automatic cars is below 3000 lbs. The MPG vs Weight trend lines would lead one to the conclusion that manual transmissions have much worse mpg than automatic for cars above 3000 lbs but again there are only 2 of 13 data points in that region. The 95% (t-test) intervals for automatic is 8.7 mpg to 25.59 mpg and manual is 10.82 mpg to 38.0 mpg. The mean weight of the manual transmission cars is 64% of the automatic cars.

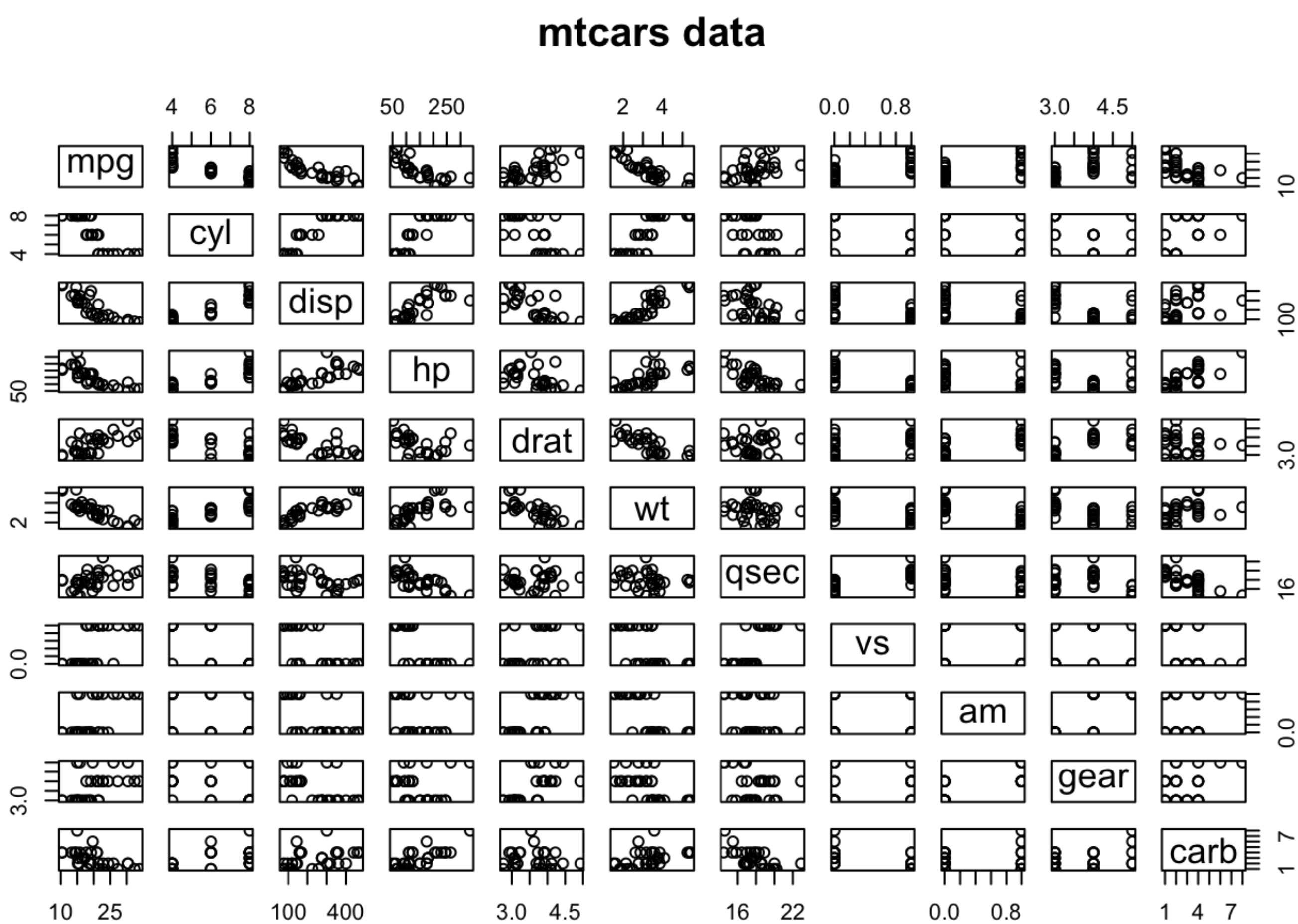
The exporatory data analysis below explains the process used to evaluate mtcars.

Exploratory Data Analysis (Including Appendix of Figures)

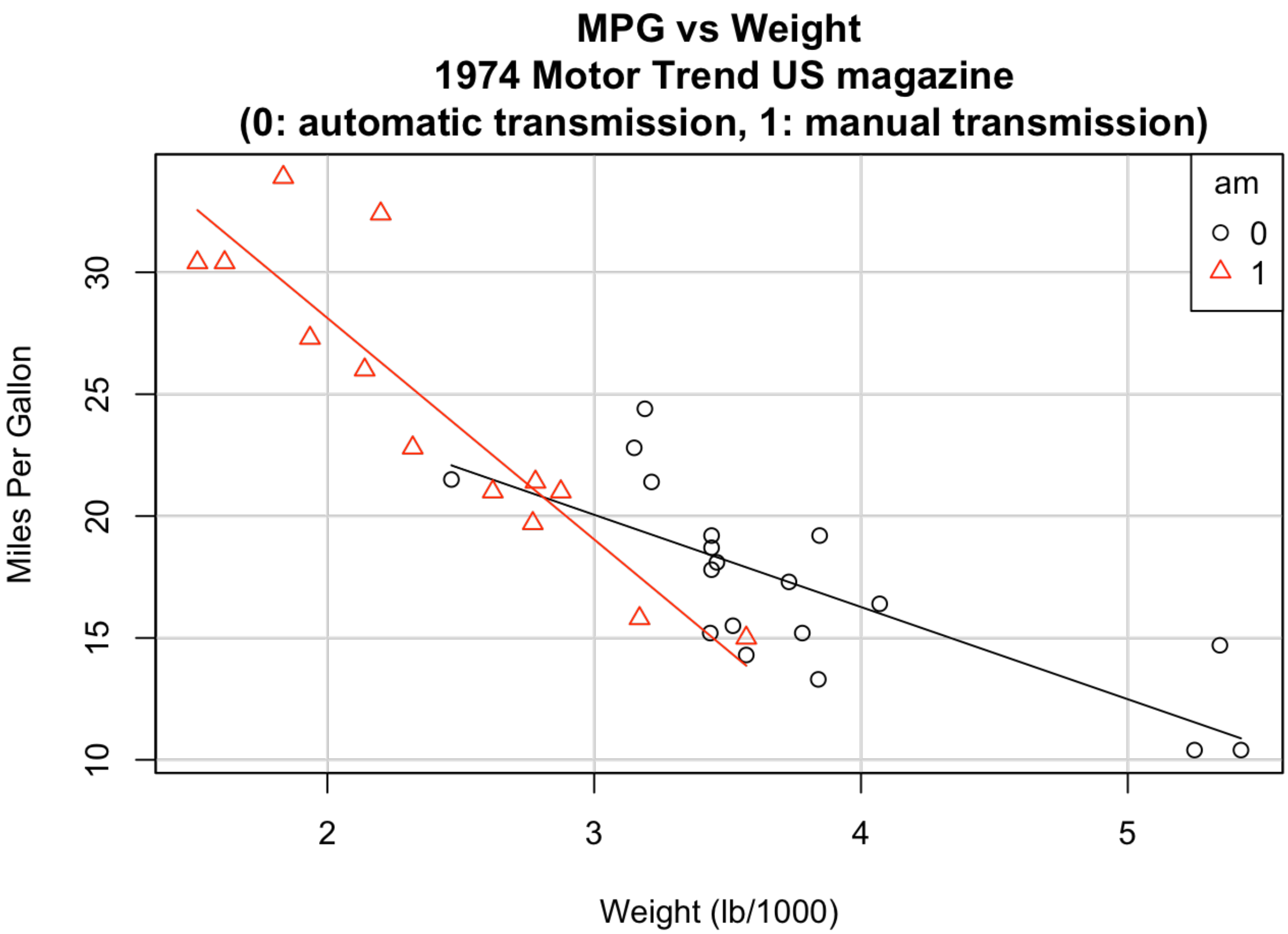
```
require(car) # this package includes the function scatterplot() which is useful for the vis
uals below
```

```
## Loading required package: car
```

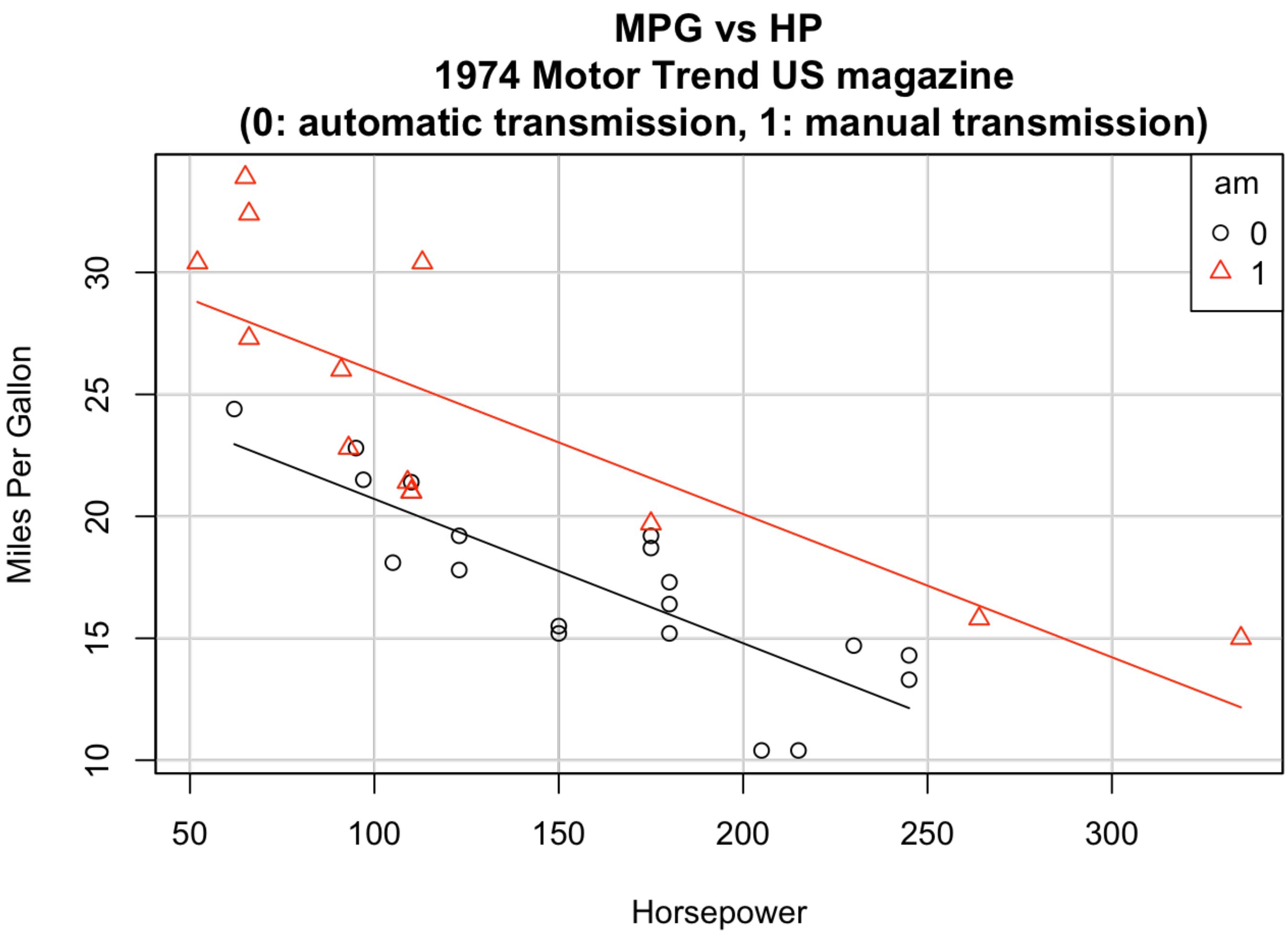
```
require(graphics)

data(mtcars) # confirm this is refreshed

# do a quick check of the relationships using the pairs() function to display
pairs(mtcars, main = "mtcars data")
```



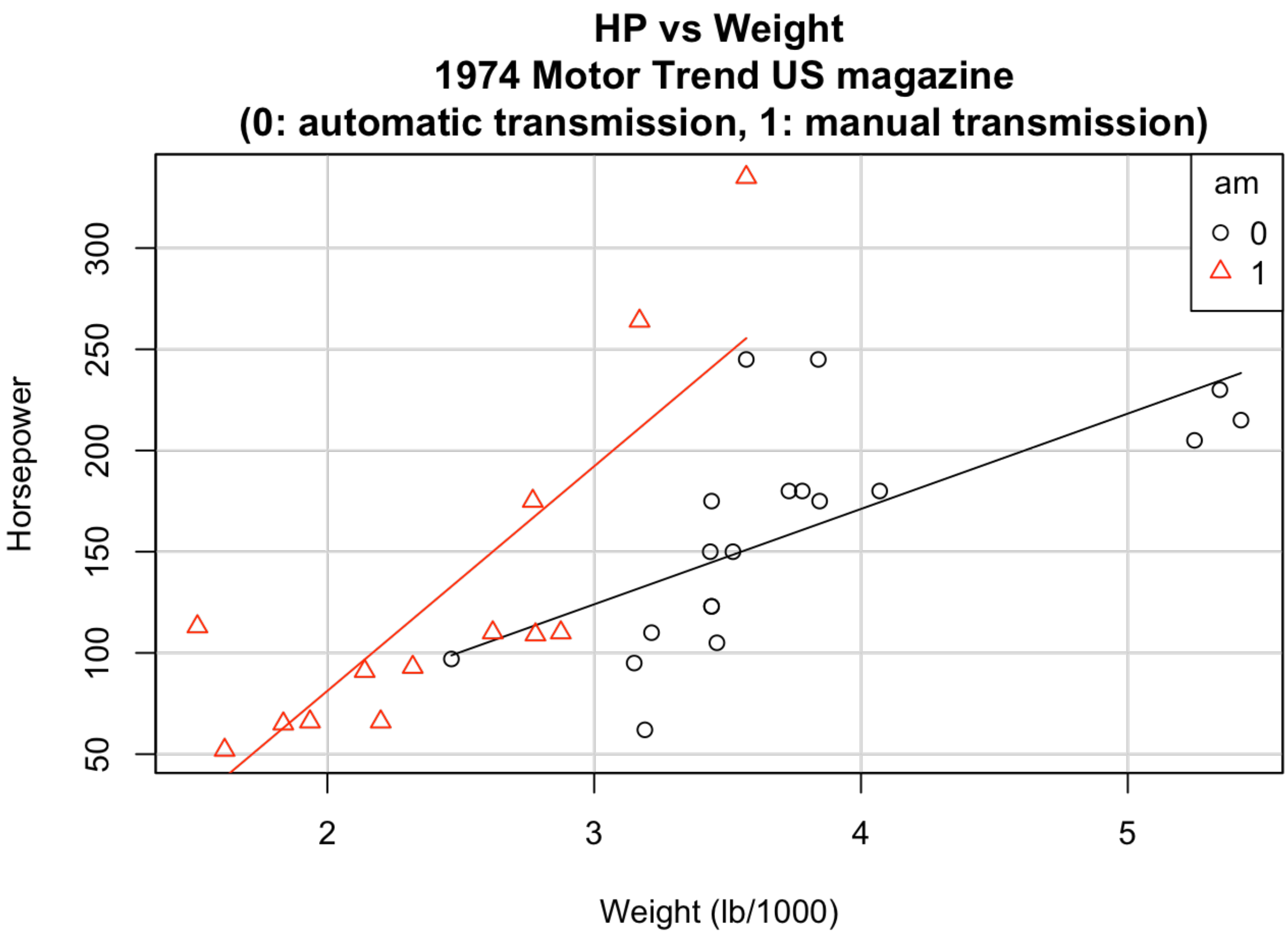
```
# do basic analysis on the data set
autoMeanMPG <- mean(mtcars[mtcars$am == 0,]$mpg)  # am=0 automatic transmission
autosdMPG <- sd(mtcars[mtcars$am == 0,]$mpg)  # am=0 automatic transmission
nauto <- nrow(mtcars[mtcars$am == 0,])
automaticRange <- autoMeanMPG+c(-1,1)*qt(.975, df=nauto-2)*autosdMPG # 95% confidence range
wtMeanauto <- mean(mtcars[mtcars$am == 0,]$wt)

manualMeanMPG <- mean(mtcars[mtcars$am == 1,]$mpg)  # am=1 automatic transmission
manualsdMPG <- sd(mtcars[mtcars$am == 1,]$mpg)   # am=1 automatic transmission
nmanual <- nrow(mtcars[mtcars$am == 1,])
manualRange <- manualMeanMPG+c(-1,1)*qt(.975, df=nmanual-2)*manualsdMPG # 95% confidence ra
nge
wtMeanmanual <- mean(mtcars[mtcars$am == 1,]$wt)

manualWTtoautoWTratio <- wtMeanmanual/wtMeanauto

# dive in and start looking at relationships visually
scatterplot(mpg ~ wt | am, data=mtcars, xlab="Weight (lb/1000)", ylab="Miles Per Gallon", m
ain="MPG vs Weight\n 1974 Motor Trend US magazine\n (0: automatic transmission, 1: manual t
ransmission)", labels=row.names(mtcars), legend.coords="topright", smoother=FALSE)
```



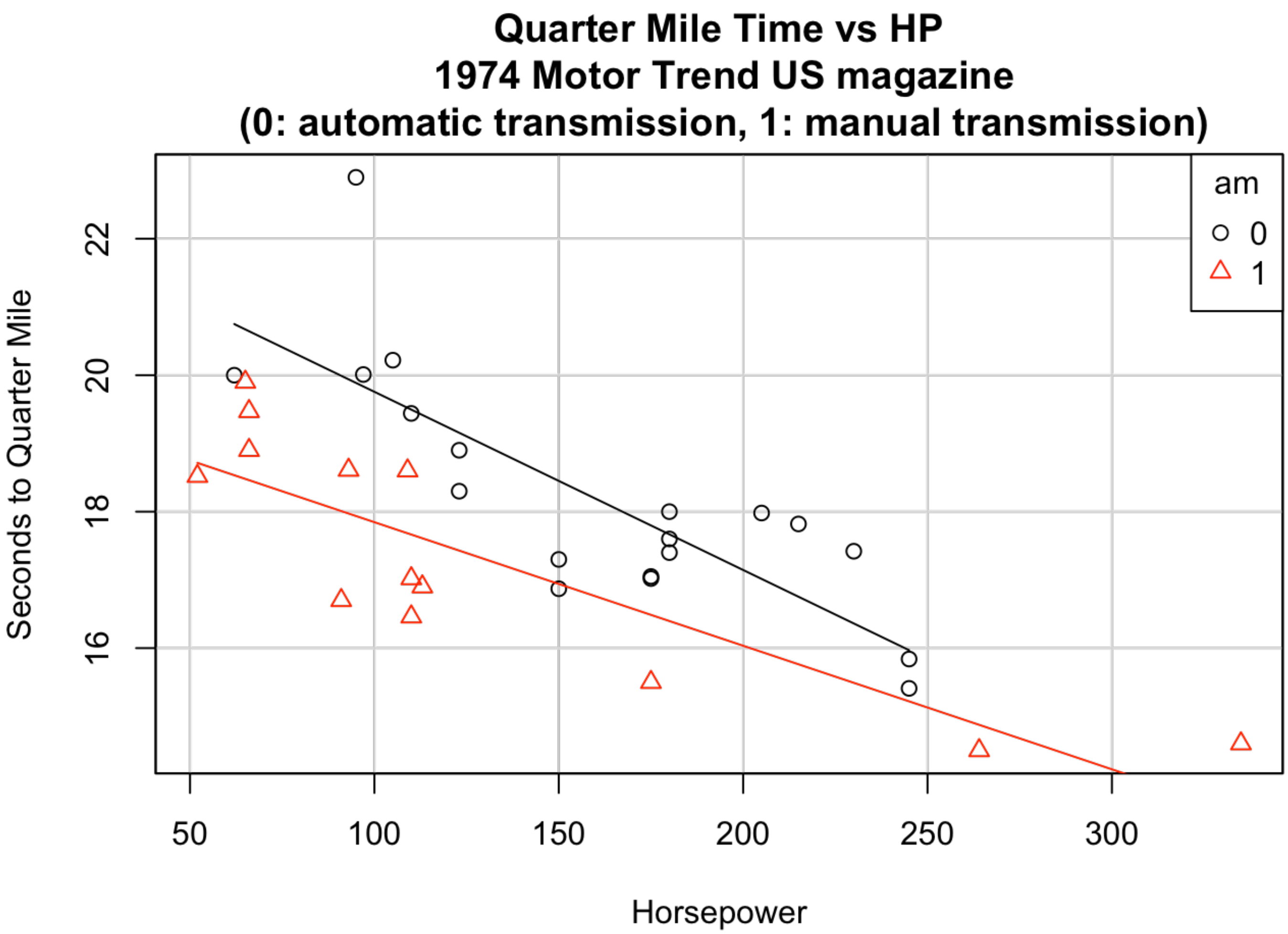

```
scatterplot(mpg ~ hp | am, data=mtcars, xlab="Horsepower", ylab="Miles Per Gallon", main="M
PG vs HP\n 1974 Motor Trend US magazine\n (0: automatic transmission, 1: manual transmissio
n)", labels=row.names(mtcars), legend.coords="topright", smoother=FALSE)
```



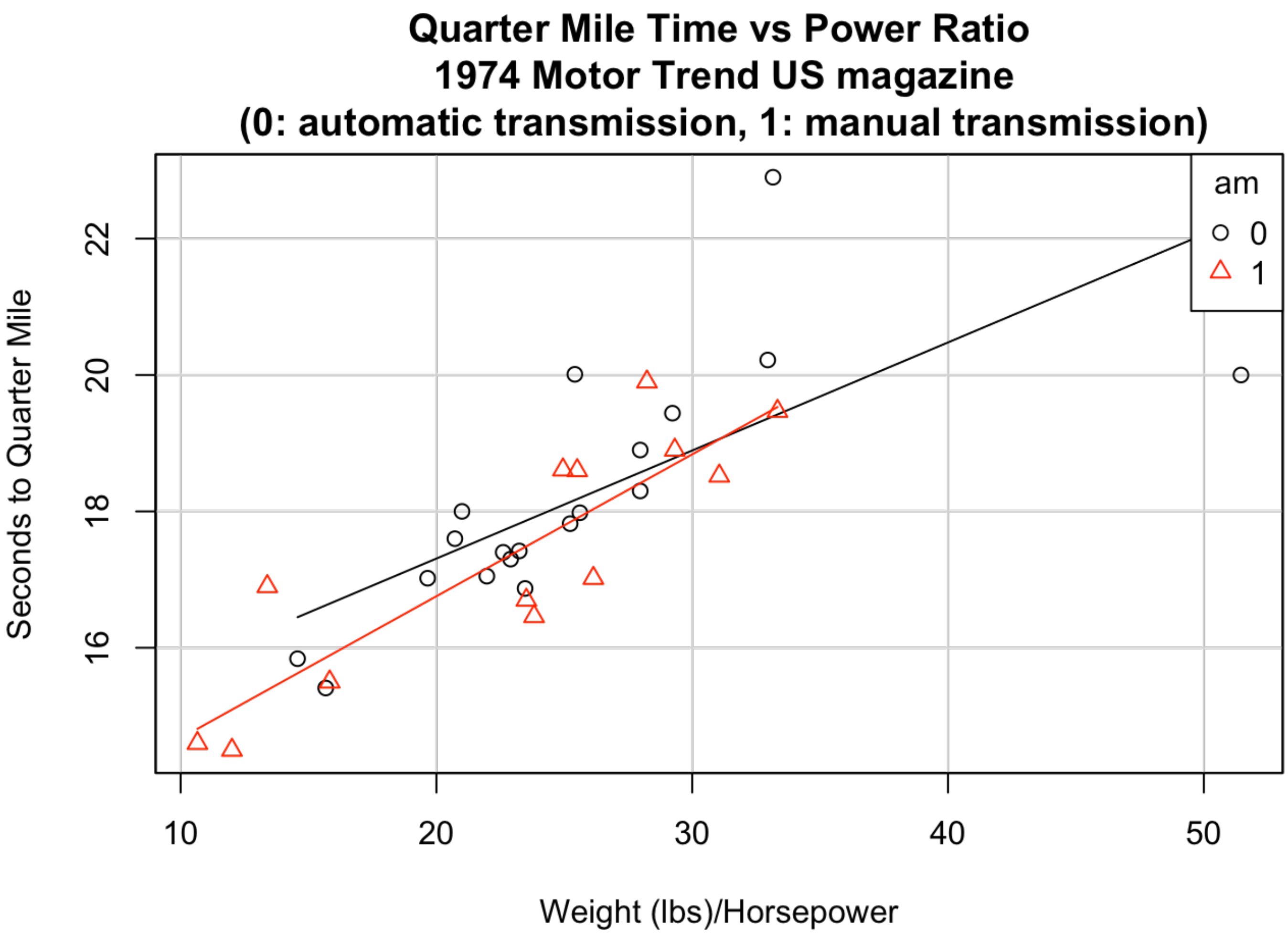

```
scatterplot(hp ~ wt | am, data=mtcars, xlab="Weight (lb/1000)", ylab="Horsepower", main="HP
 vs Weight\n 1974 Motor Trend US magazine\n (0: automatic transmission, 1: manual transmiss
ion)", labels=row.names(mtcars), legend.coords="topright", smoother=FALSE)
```




```
scatterplot(qsec ~ hp | am, data=mtcars, xlab="Horsepower", ylab="Seconds to Quarter Mile",
 main="Quarter Mile Time vs HP\n 1974 Motor Trend US magazine\n (0: automatic transmission,
 1: manual transmission)", labels=row.names(mtcars), legend.coords="topright", smoother=FAL
SE)
```




```
mtcars$powerRatio <- mtcars$wt*1000/mtcars$hp

scatterplot(qsec ~ powerRatio | am, data=mtcars, xlab="Weight (lbs)/Horsepower", ylab="Seconds to Quarter Mile", main="Quarter Mile Time vs Power Ratio\n 1974 Motor Trend US magazine\n (0: automatic transmission, 1: manual transmission)", labels=row.names(mtcars), legend.coords="topright", smoother=TRUE)
```




```
#
# Using external sources for reference to compare to a larger body of car performance data
# using data from "Calculators for 1/4-Mile ET & MPH vs. HP and Weight" http://www.stealth3
16.com/2-calc-hp-et-mph.htm

mtcars$MPHestimate <- 215.39*(mtcars$hp/(mtcars$wt*1000))^(0.3018)
mtcars$QuarterMileTimeEstimate <- 7.3571*((mtcars$wt*1000)/mtcars$hp)^(0.2574)
mtcars$HPestimate <- (mtcars$wt*1000)/(mtcars$qsec/7.3571)^3.885
mtcars$HPestimateToActualHPRatio <- mtcars$HPestimate/mtcars$hp

#test transmission type (am) as a predictor
summary(fitam <- lm(mpg ~ am, data=mtcars))
```

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.392 -3.092 -0.297  3.244  9.508
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    17.15       1.12   15.25  1.1e-15 ***
## am              7.24       1.76    4.11  0.00029 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.9 on 30 degrees of freedom
## Multiple R-squared:  0.36,   Adjusted R-squared:  0.338
## F-statistic: 16.9 on 1 and 30 DF,  p-value: 0.000285
```

```
confint(fitam)
```

```
##              2.5 % 97.5 %
## (Intercept) 14.851  19.44
## am           3.642  10.85
```

```
# test weight (wt) as the best predictor of MPG
summary(model1auto <- lm(mpg ~ wt, data=mtcars[mtcars$am==0,])) # test manual relationship
to weight
```

```
##
## Call:
## lm(formula = mpg ~ wt, data = mtcars[mtcars$am == 0, ])
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.600 -1.523 -0.217  1.482  5.061
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   31.416      2.947   10.66    6e-09 ***
## wt            -3.786      0.767   -4.94  0.00012 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.53 on 17 degrees of freedom
## Multiple R-squared:  0.589,  Adjusted R-squared:  0.565
## F-statistic: 24.4 on 1 and 17 DF,  p-value: 0.000125
```

```
confint(model1auto)
```

```
##              2.5 % 97.5 %
## (Intercept) 25.199 37.633
## wt          -5.403 -2.169
```

```
summary(model1manual <- lm(mpg ~ wt, data=mtcars[mtcars$am==1,])) # test automatic relation
ship to weight
```

```
##
## Call:
## lm(formula = mpg ~ wt, data = mtcars[mtcars$am == 1, ])
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -2.419 -1.494 -1.223  0.823  6.091
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    46.29       3.12   14.84  1.3e-08 ***
## wt             -9.08       1.26   -7.23  1.7e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.69 on 11 degrees of freedom
## Multiple R-squared:  0.826,  Adjusted R-squared:  0.81
## F-statistic: 52.3 on 1 and 11 DF,  p-value: 1.69e-05
```

```
confint(model1manual)
```

```
##              2.5 % 97.5 %
## (Intercept)  39.43 53.161
## wt          -11.85 -6.319
```

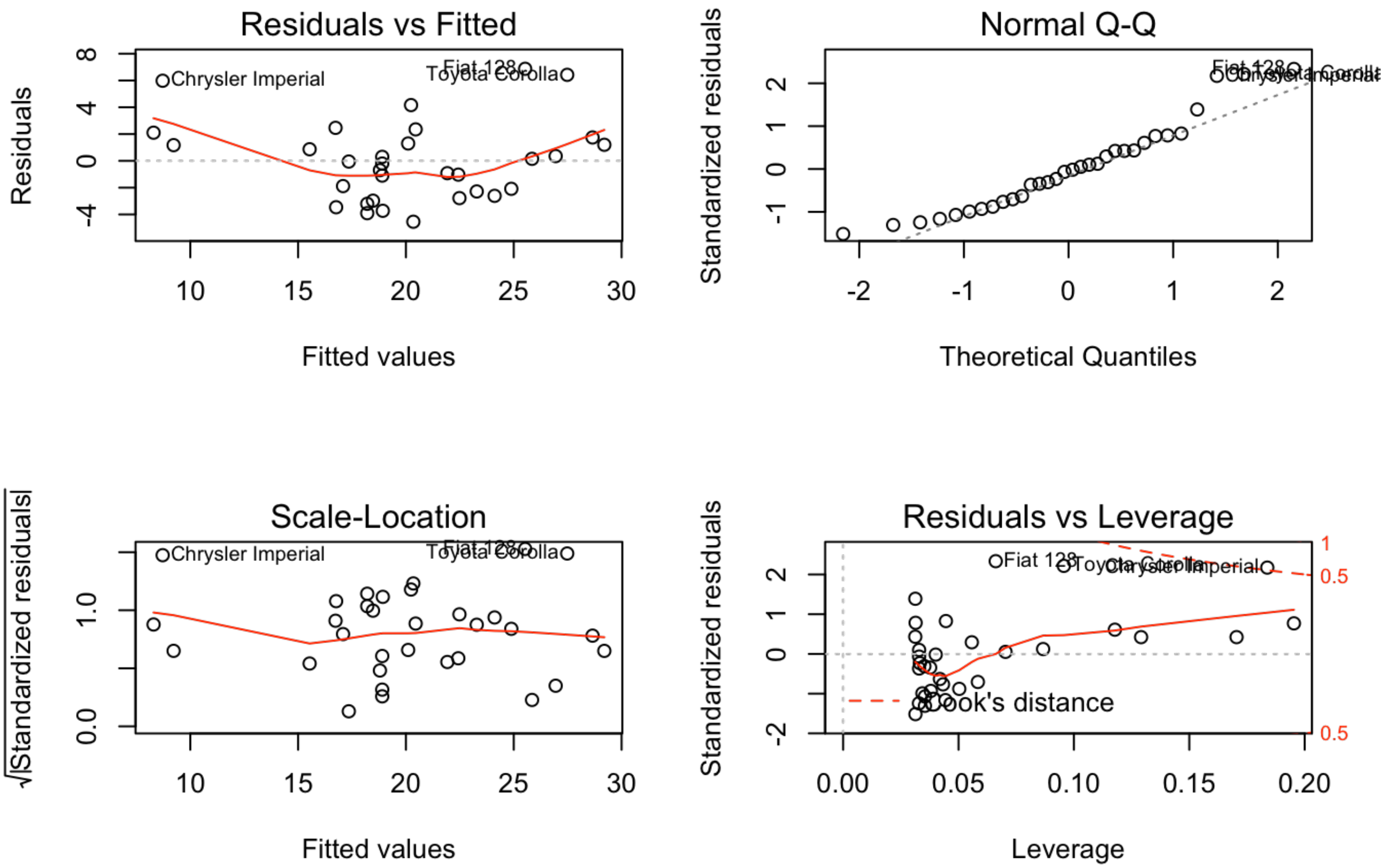
```
summary(model1 <- lm(mpg ~ wt, data=mtcars))
```

```
##
## Call:
## lm(formula = mpg ~ wt, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.543 -2.365 -0.125  1.410  6.873
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   37.285      1.878   19.86  < 2e-16 ***
## wt            -5.344      0.559   -9.56  1.3e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.05 on 30 degrees of freedom
## Multiple R-squared:  0.753,  Adjusted R-squared:  0.745
## F-statistic: 91.4 on 1 and 30 DF,  p-value: 1.29e-10
```

```
confint(model1)
```

```
##              2.5 % 97.5 %
## (Intercept) 33.450 41.120
## wt          -6.486 -4.203
```

```
#plot(mtcars$wt, model1$residuals)
#abline(h=0)
par(mfrow=c(2,2))
plot(model1)
```




```
#test horsepower(hp) as the best predictor of MPG
summary(model2auto <- lm(mpg ~ hp, data=mtcars[mtcars$am==0,])) # test manual relationship
to hp
```

```
##
## Call:
## lm(formula = mpg ~ hp, data = mtcars[mtcars$am == 0, ])
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.102 -1.903  0.611  1.559  2.924
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 26.62485    1.61588   16.48  6.9e-12 ***
## hp          -0.05914    0.00958   -6.17  1.0e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.19 on 17 degrees of freedom
## Multiple R-squared:  0.691,  Adjusted R-squared:  0.673
## F-statistic: 38.1 on 1 and 17 DF,  p-value: 1.02e-05
```

```
confint(model2auto)
```

```
##                2.5 %   97.5 %
## (Intercept) 23.21563 30.03406
## hp          -0.07935 -0.03892
```

```
summary(model2manual <- lm(mpg ~ hp, data=mtcars[mtcars$am==1,])) # test automatic relation
ship to hp
```

```
##
## Call:
## lm(formula = mpg ~ hp, data = mtcars[mtcars$am == 1, ])
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -4.382 -3.580 -0.537  2.833  5.875
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  31.8425     1.9926   15.98  5.8e-09 ***
## hp           -0.0587     0.0133   -4.43    0.001 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.86 on 11 degrees of freedom
## Multiple R-squared:  0.641,  Adjusted R-squared:  0.608
## F-statistic: 19.6 on 1 and 11 DF,  p-value: 0.00101
```

```
confint(model2manual)
```

```
##               2.5 %   97.5 %
## (Intercept) 27.4568 36.22822
## hp          -0.0879 -0.02957
```

```
summary(model2 <- lm(mpg ~ hp, data=mtcars))
```

```
##
## Call:
## lm(formula = mpg ~ hp, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.712 -2.112 -0.885  1.582  8.236
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  30.0989     1.6339   18.42  < 2e-16 ***
## hp           -0.0682     0.0101   -6.74  1.8e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.86 on 30 degrees of freedom
## Multiple R-squared:  0.602,  Adjusted R-squared:  0.589
## F-statistic: 45.5 on 1 and 30 DF,  p-value: 1.79e-07
```

```
confint(model2)
```

```
##                2.5 %   97.5 %
## (Intercept) 26.76195 33.43577
## hp          -0.08889 -0.04756
```

```
#plot(mtcars$hp, model2$residuals)
#abline(h=0)
par(mfrow=c(2,2))
plot(model2)

# source "R in Action" by Robert Kabacoff, page 186-188
# recommends that one explore the interrelationship between wt and hp together as a factor
# so...
fithpANDwt <- lm(mpg ~ hp + wt + hp:wt + am, data = mtcars)
summary(fithpANDwt)
```

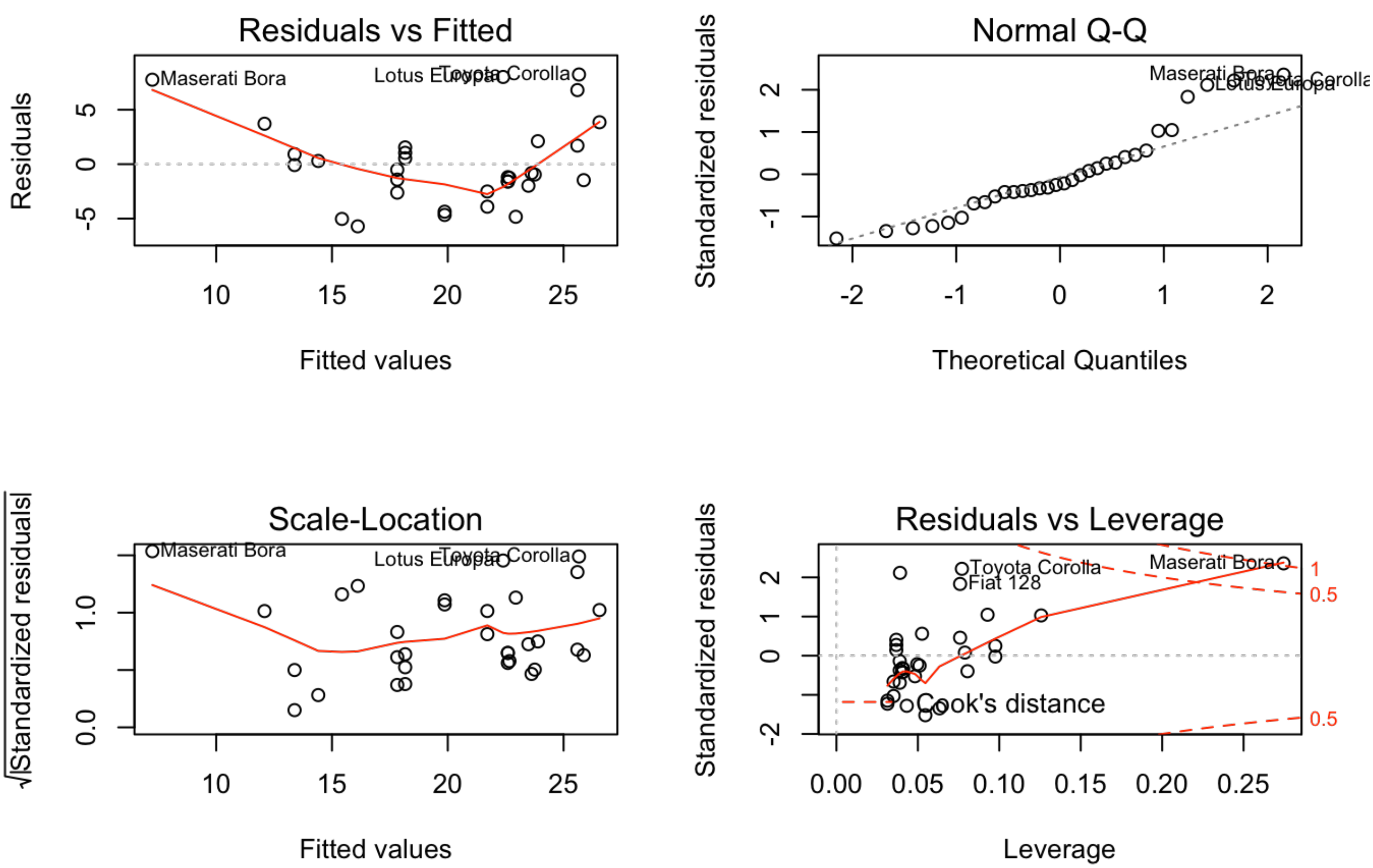
```
##
## Call:
## lm(formula = mpg ~ hp + wt + hp:wt + am, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -2.985 -1.658 -0.741  1.436  4.527
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 49.45224    5.28073    9.36  5.7e-10 ***
## hp          -0.11930    0.02655   -4.49  0.00012 ***
## wt          -8.10056    1.78933   -4.53  0.00011 ***
## am           0.12511    1.33343    0.09  0.92594
## hp:wt        0.02749    0.00847    3.24  0.00313 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.19 on 27 degrees of freedom
## Multiple R-squared:  0.885,  Adjusted R-squared:  0.868
## F-statistic: 51.8 on 4 and 27 DF,  p-value: 2.76e-12
```

```
confint(fithpANDwt)
```

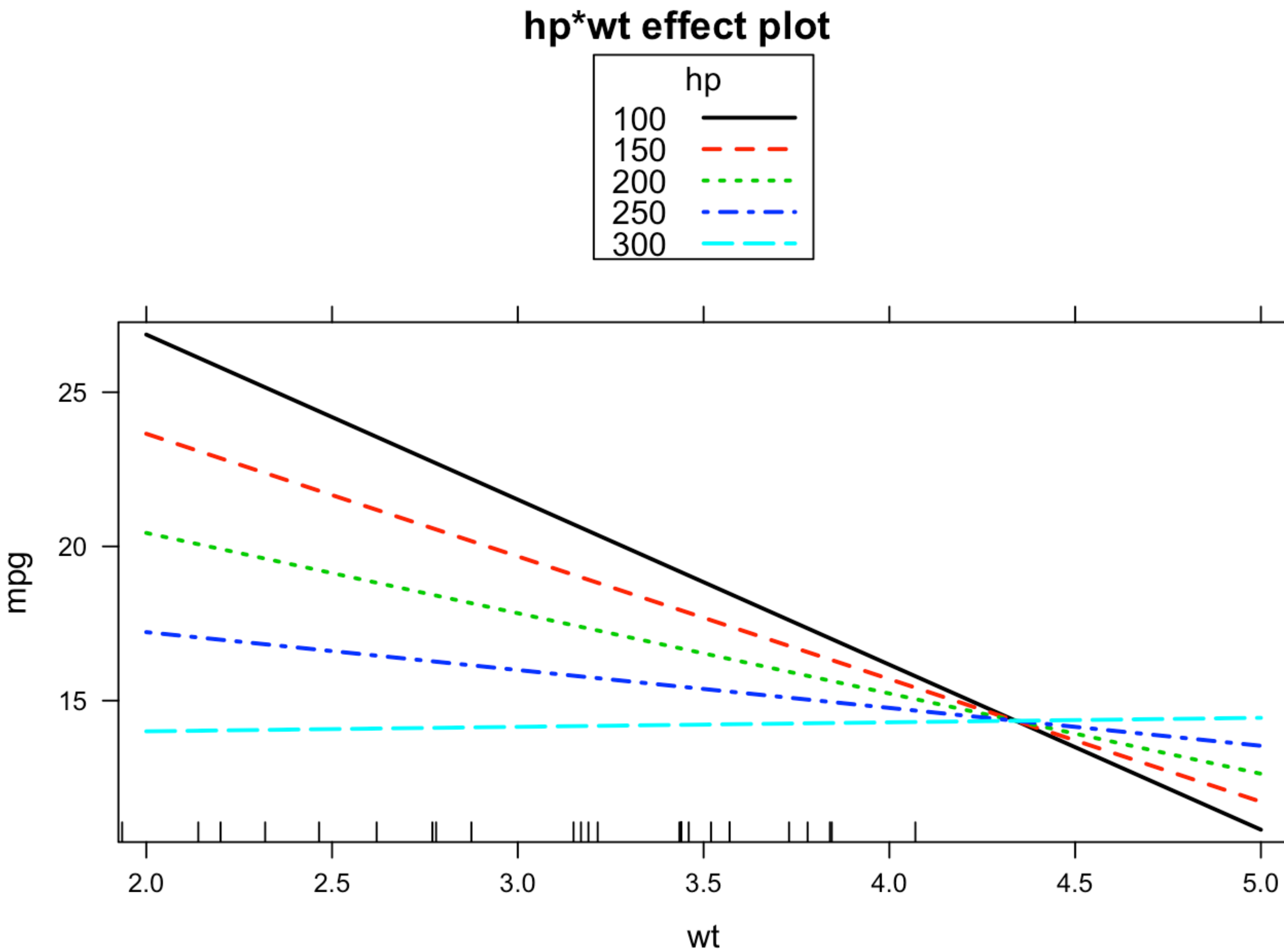
```
##                2.5 %   97.5 %
## (Intercept)  38.6171 60.28741
## hp           -0.1738 -0.06483
## wt          -11.7719 -4.42917
## am           -2.6109  2.86108
## hp:wt         0.0101  0.04487
```

```
require(effects)
```

```
## Loading required package: effects
## Loading required package: lattice
## Loading required package: grid
## Loading required package: colorspace
##
## Attaching package: 'effects'
##
## The following object is masked from 'package:car':
##
##     Prestige
```




```
plot(effect("hp:wt", fithpANDwt,list(wt=c(2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0))), multiline=T
RUE)
```



Conclusions

There are clearly different MPG performance characteristics between the cars with manual vs automatic transmission as evidenced by the graphs and trend lines; however, there are several issues with the data (e.g., manual transmissions are mostly on the lightest cars with presumably less rolling friction and smaller drag area cross sections). The MPG vs HP graph seems to show that the manual transmissions have as good or better mpg than their equivalent HP automatic comparisons, but the confidence level is very low once you add the weight and weight * hp factors.

A simple linear regression on transmission type (am) as a predictor only explains 36% of the MPG difference (R^2) and has a slope confidence interval that is very broad (from 3.6 to 10.9) while weight (wt) explains 75%, hp explains 60%, and weight + hp + wt * hp explains 89% of the MPG prediction. Adding the transmission type has no effect and has a p-value of .93.

For Fun - Analysis of This Question Using Modern Car Capabilities

This question about whether manual or automatic transmissions provide superior MPG and other advantages continues to this day. Most car analysts suggest that very sophisticated modern day automatic transmissions (especially “continuous variable”) are as good are even better than manual transmissions for fuel economy. There is a funny version of this discussion with the famous “Frick and Frack Brothers Car Talk Series on NPR” dated 8/21/12 and located here: http://www.cartalk.com/content/today-manual-transmission-myths-debunked (http://www.cartalk.com/content/today-manual-transmission-myths-debunked) . Here is another source that shows that it is a toss up also though the title indicates otherwise: http://www.consumerreports.org/cro/2012/01/save-gas-and-money-with-a-manual-transmission/index.htm (http://www.consumerreports.org/cro/2012/01/save-gas-and-money-with-a-manual-transmission/index.htm)

Here is another reference exploring statistical analysis of a broad range of automobile performance characteristics starting with regression analysis from the 1950’s to today. The analysis does not explore the automatic vs. manual discussion but does look at the ratios between quarter mile time, quarter mile MPH, weight, horsepower, Power Ratio, etc. It can be found here: “Calculators for 1/4-Mile ET & MPH vs. HP and Weight” http://www.stealth316.com/2-calc-hp-et-mph.htm (http://www.stealth316.com/2-calc-hp-et-mph.htm)