Regression Models Project

Prateek Sarangi

Executive Summary

When you just consider automatic vs. manual transmission, the manual transmission is more fuel efficient. When you adjust for horsepower and weight, the difference between the two types of transmission is greatly reduced and is no longer statistically significant. I analyzed this by plotting the residuals against mpg as I added terms to the model. The last stage was to use the anova function to compare the model. The final conclusion is that the transmission type doesn't affect fuel efficiency.

##Project description.

Using the data in the mtcars dataset, determine whether manual cars have lower mpg than automatic. Derive an estimate of the difference and interpret the difference.

##Considerations

Manual and automatic transmission cars may differ for reasons other than the transmission type. Exploratory data analysis will allow us to find these differences and interpret them. All numeric variables will be plotted as two boxplots comparing manual transmission in one box against automatic transmission in the other box. Plot 1 in Appendix.

The mpg of AM=1 (manual transmission) is clearly higher. At a first glance, this is our answer. Manual transmissions have higher mpg (more gas efficiency). But look at weight in Plot 2 in the Appendix.

The weight of AM=1 (manual) is clearly lower, so the weight varies with transmission type. We might expect weight to be related to mpg (see Plot 3).

And, yes, heavier cars have lower mpg.

Is Horsepower related to mpg? (See Plot 4)

Yes

Now we do a simple (or marginal) fit of mpg to transmission type.

The fit gives us a beta1 of 7.2449393; this the change in mpg when switching from automatic to manual transmission. This model explains 36 % of the variance.

The residuals are not random with respect to mpg, horsepower or weight. The residuals basically make a pair of straight lines with mpg, so something significant is missing. The residuals are lower (negative) at higher weight and higher horsepower. Now, we can color in the residuals vs mpg plot with transmission type, weight, displacement, horsepower, gear, or cylinders (see Plots 5-10)

All of these variables show different distribution between automatic and manual transmission. So we can add them one by one in a nested set of fits. First, we will try with adding horsepower (See plots 11-12). ###Fit1

The beta1 coefficient against transmission type is 5.2770853. This is different than the beta1 coefficient we got before. This is the change in mpg going from an automatic to a manual transmission, holding horsepower constant.

This model explains 78 % of the variance.

We can use the diagnostics to look for outliers. Maserati Bora is the most extreme. The fitted mpg (when fitted without this car) is greater than the actual by the most. Dodge Challenger is extreme. The fitted mpg (when fitted without this car) is less than the actual value by the most.

###Fit 2 Let's see what the interaction between transmission and horsepower reveals. See plots 13-15.

Now the interpretation of the beta1 is a bit more complex, the beta1 of 5.2176534 is the difference between automatic and manual transmissions for a car with 0 hp. We would rather look at the mean of reasonable car horsepower, 147 hp. At that horsepower, the change in mpg from automatic to manual is 5.2767524.

This model explains 78 % of the variance. This is not a big improvement, so interaction is not the best explanation. This fit will not be used for any further analysis.

Maserati Bora is the most extreme value.

###Fit 3

This model has 3 terms: transmission, horsepower and weight. This model explains 84 % of the variance. See plots 16 and 17.

The transmission type does not significantly affect the mpg, after adjustment for horsepower and weight. Beta1 is 2.0837101.

Maserati Bora is the most extreme value.

You can explain even more of the variance using an interaction term with horsepower and weight.

###Fit 4

This model has four terms: transmission, horsepower, weight and the interaction between horsepower and weight. This model explains 88 % of the variance. To compare the fits (which are nested, meaning each one contains the terms of the previous), we use nested model testing in R.

Beta1 is 0.1251069.

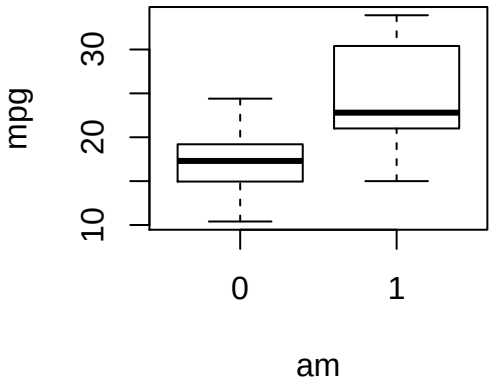
```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + hp
## Model 3: mpg ~ am + hp + wt
## Model 4: mpg ~ am + hp * wt
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     29 245.44  1    475.46 98.963 1.588e-10 ***
## 3     28 180.29  1     65.15 13.560  0.001019 **
## 4     27 129.72  1     50.57 10.526  0.003130 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

In this comparison of fits, we look at the decrease in RSS to see how much the additional terms improve the fit, versus just add complexity to the model. One rule of thumb is that the RSS should approximately halve with each term. These four fits have that relationship, so the fourth one is the best.

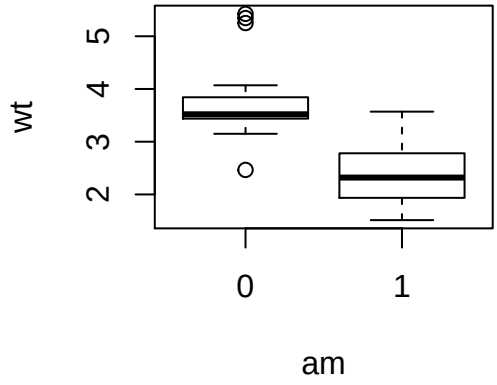
The transmission type does not significantly affect the mpg, after adjustment for horsepower and weight.

Appendix Plots are found here.

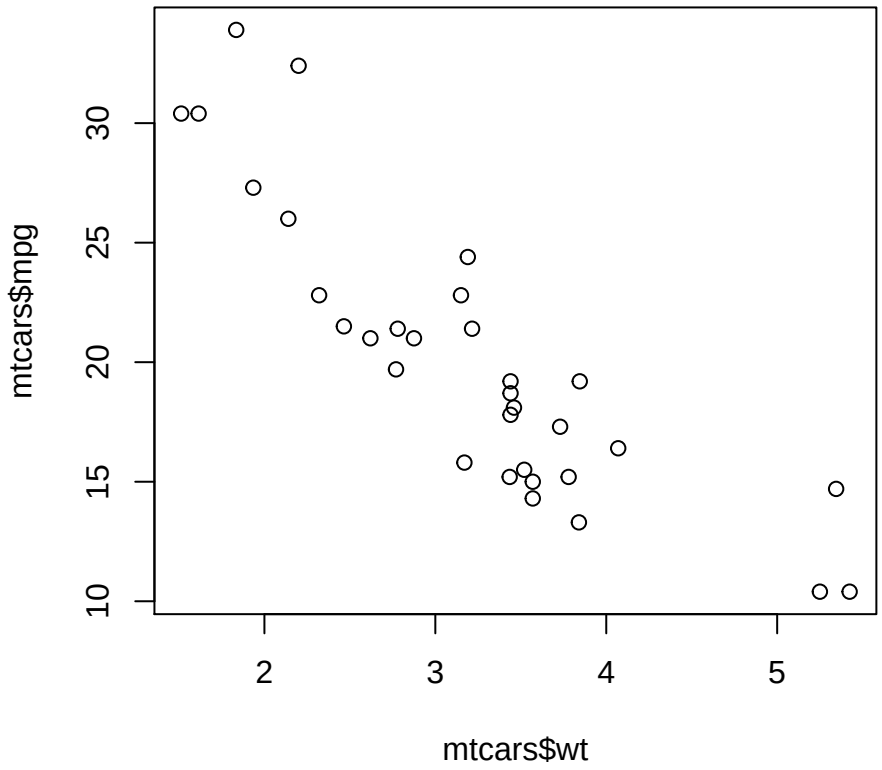
Plot 1



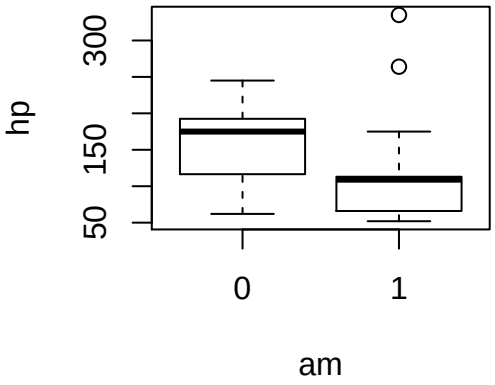
Plot 2

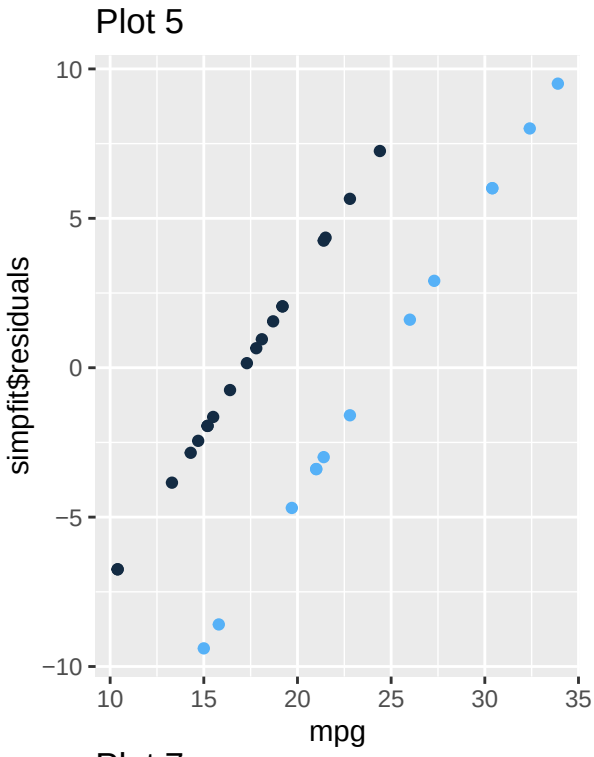


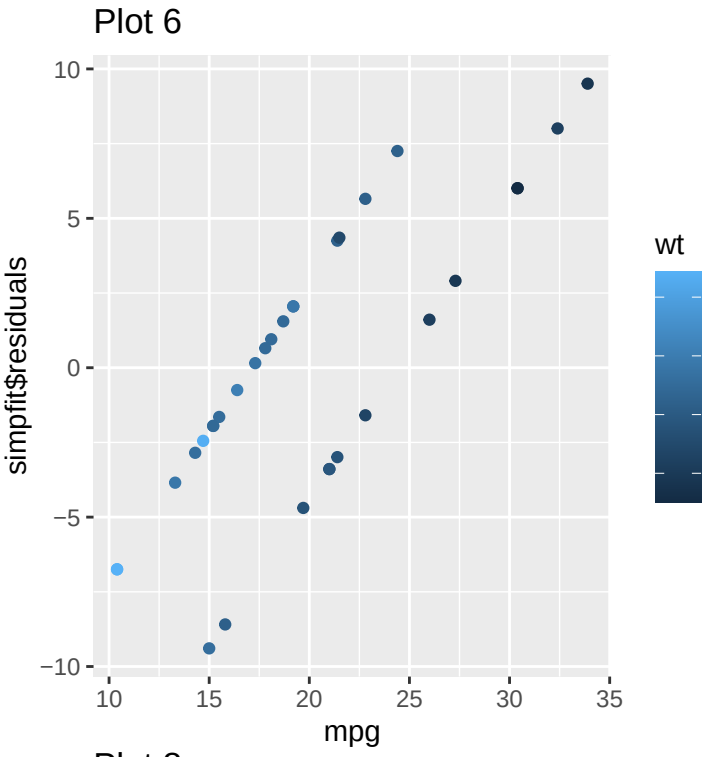
Plot 3

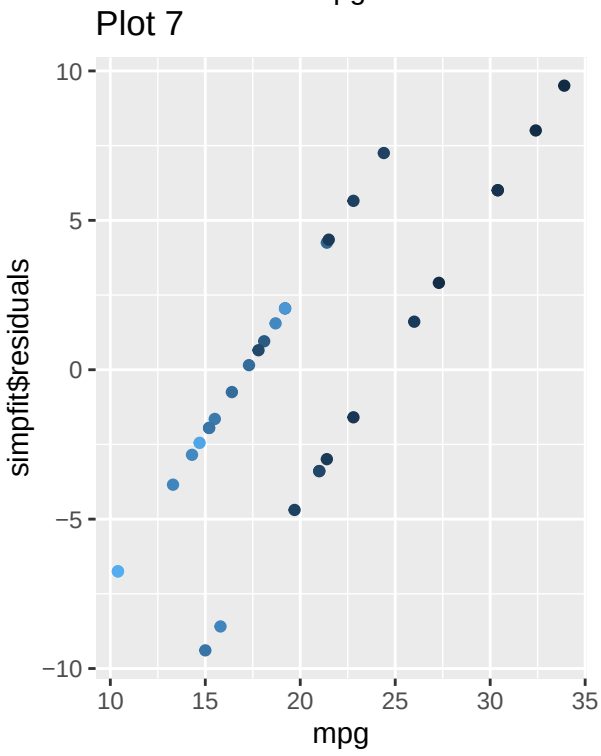


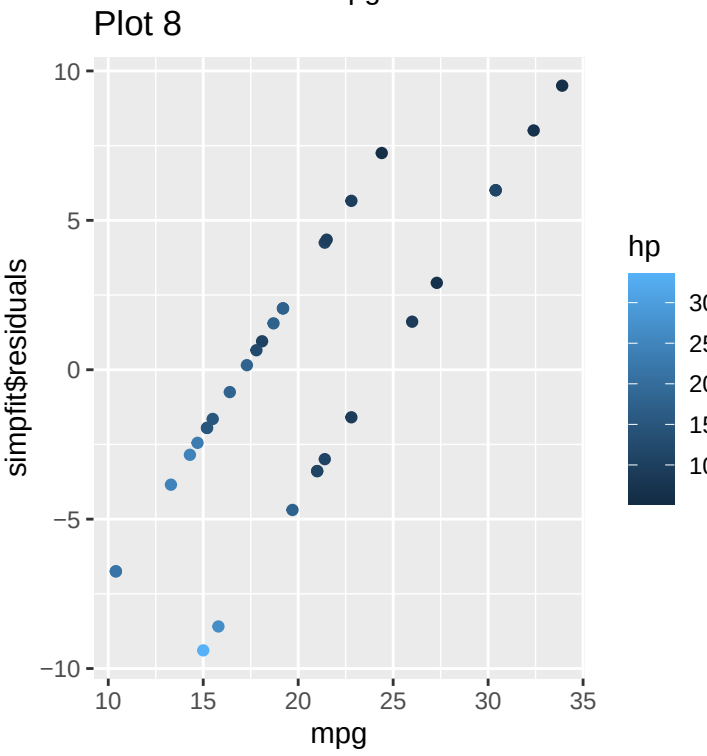
Plot 4

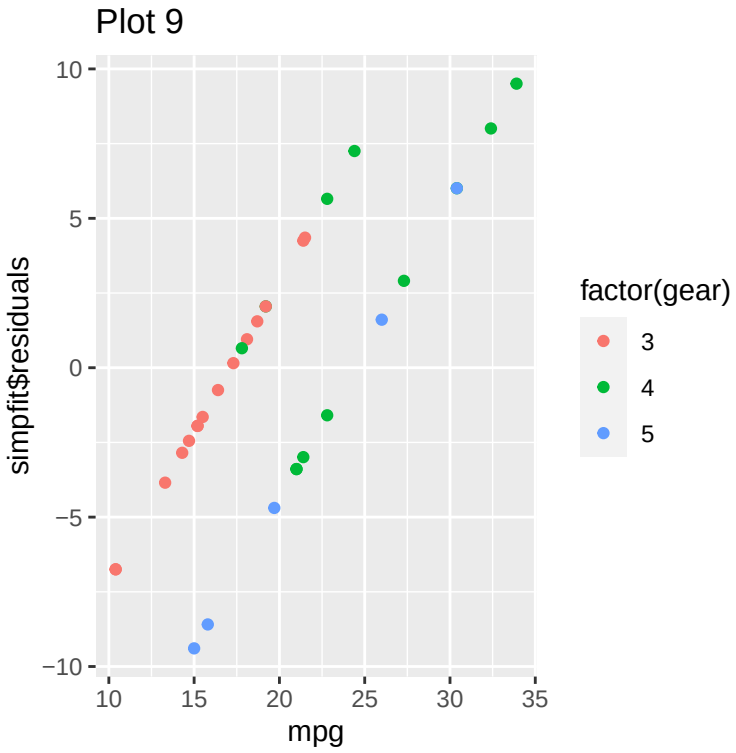


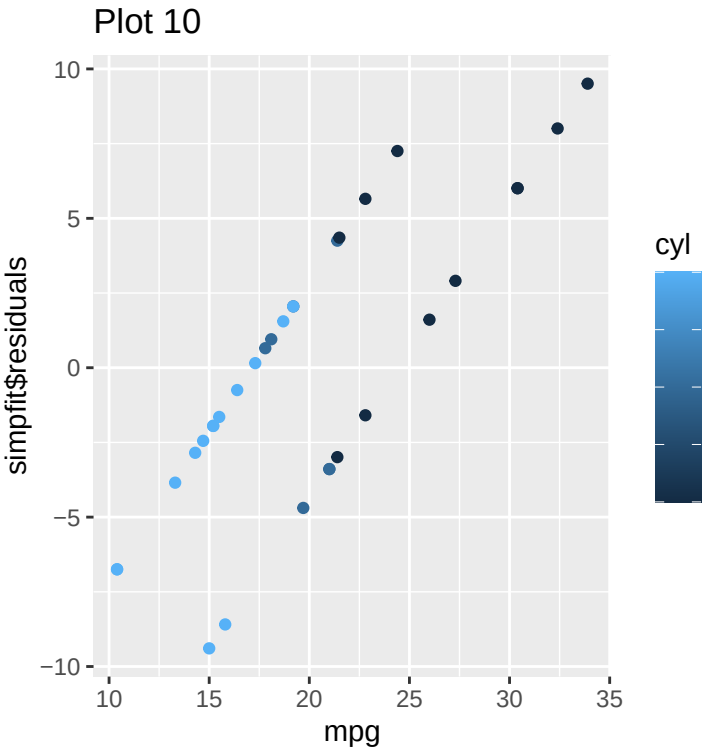

Plot 5
simpfit$residuals
mpg
am
1.00
0.75
0.50
0.25
0.00

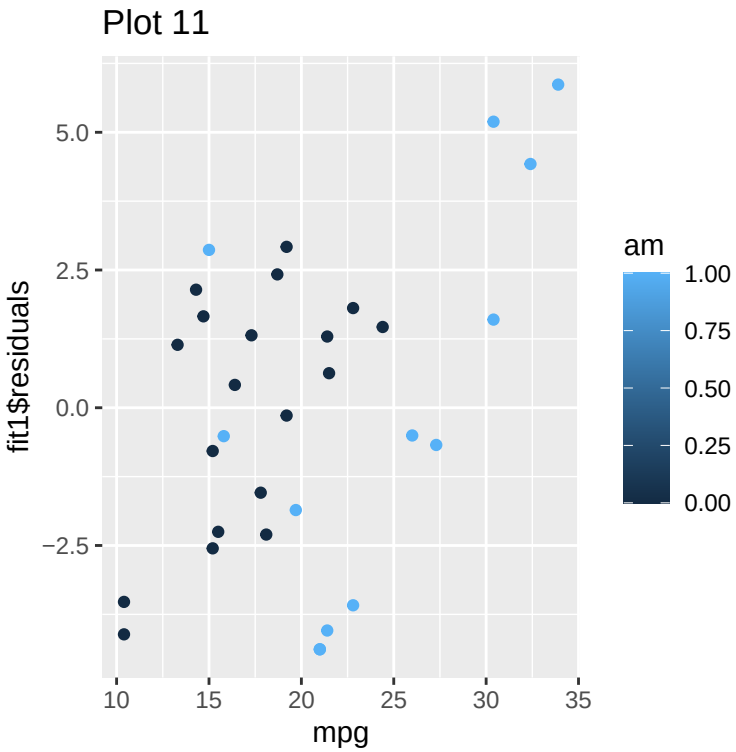

Plot 6
simpfit$residuals
mpg
wt

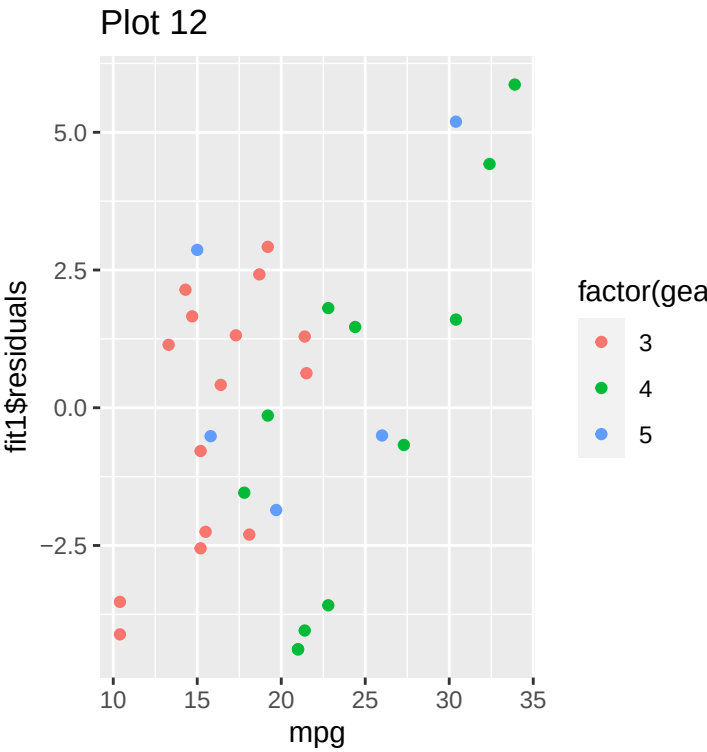

Plot 7
simpfit$residuals
mpg
disp
400
300
200
100


Plot 8
simpfit$residuals
mpg
hp

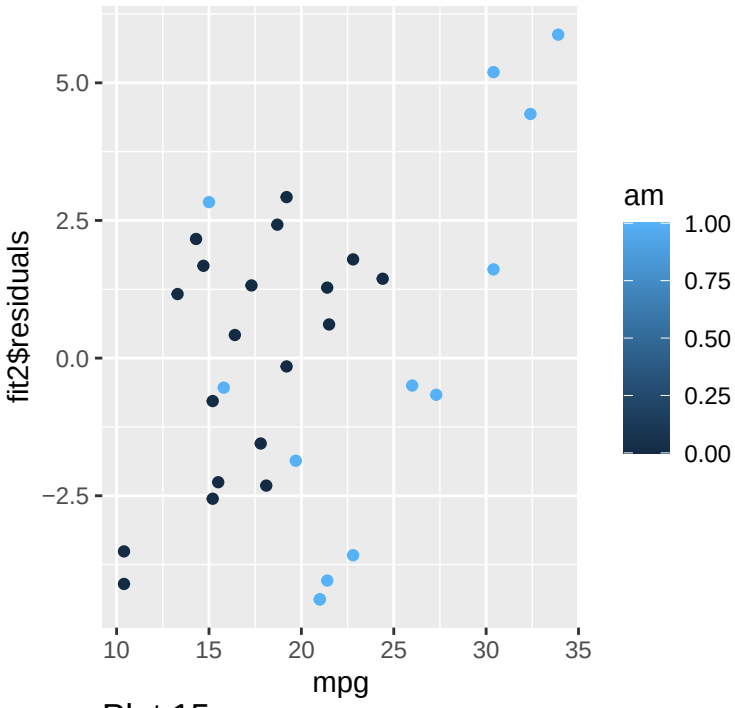

Plot 9
simpfit$residuals
mpg
factor(gear)
3
4
5


Plot 10
simpfit$residuals
mpg
cyl

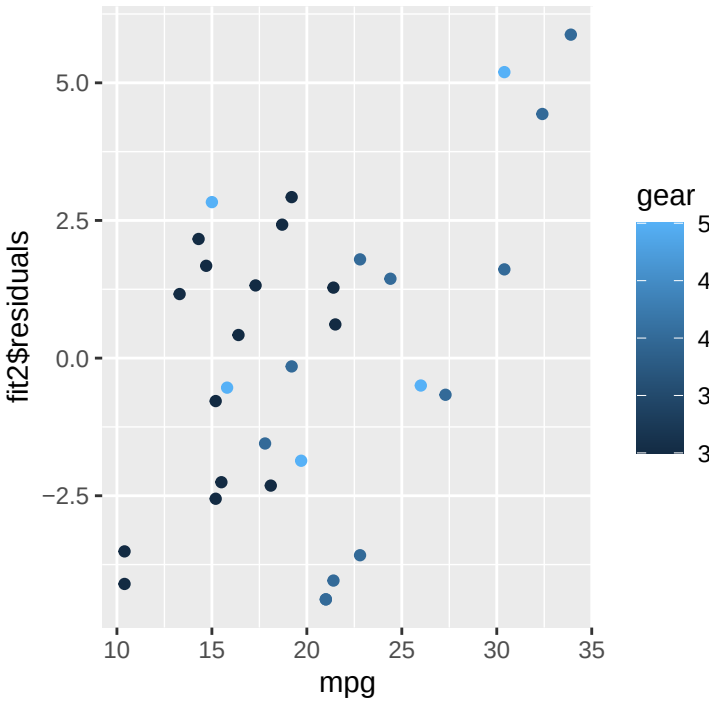

Plot 11
fit1$residuals
mpg
am
1.00
0.75
0.50
0.25
0.00


Plot 12
fit1$residuals
mpg
factor(gea
3
4
5

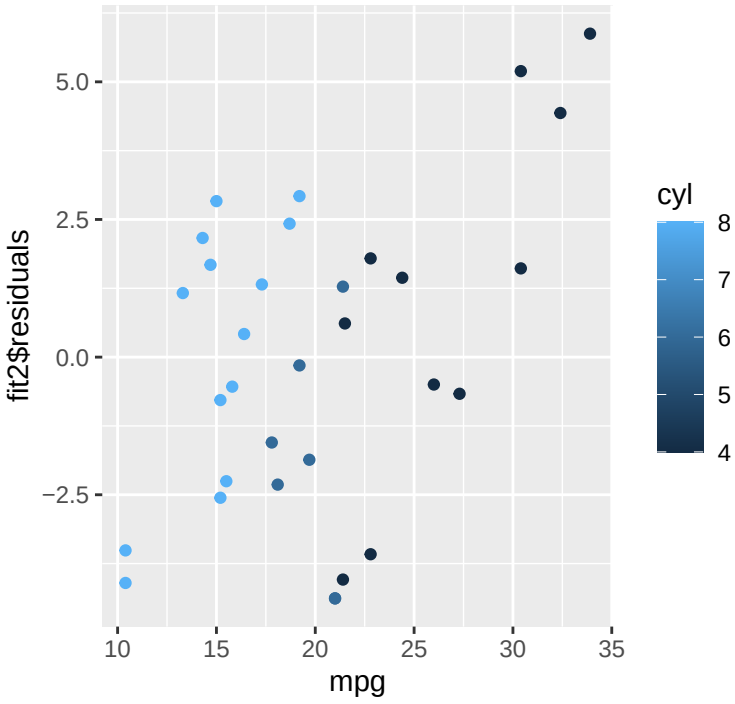
Plot 13



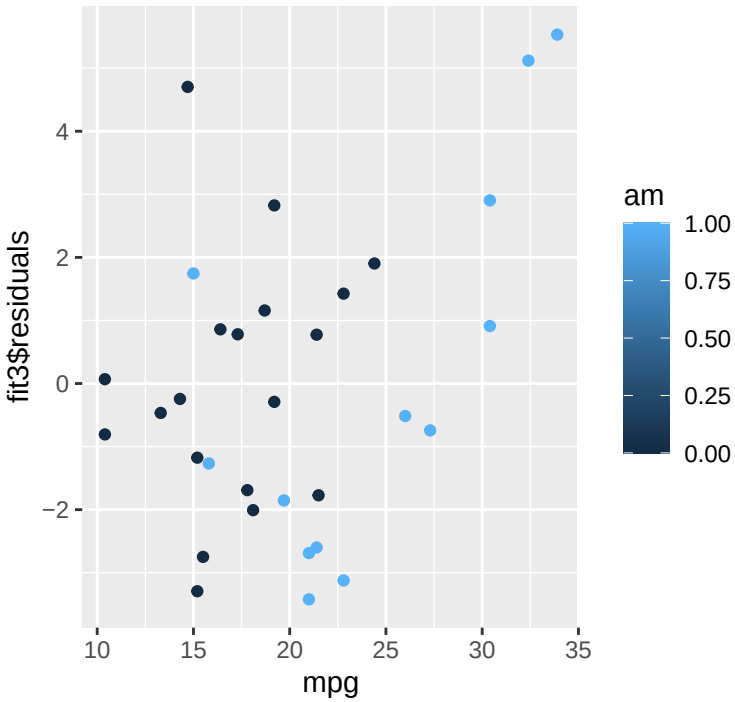
Plot 14



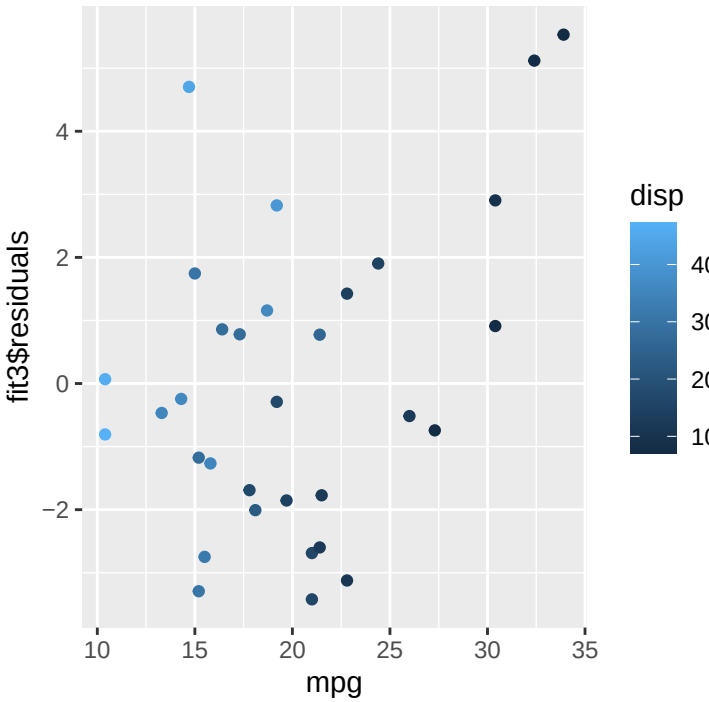
Plot 15



Plot 16



Plot 17



##Version Information

```
## [1] "x86_64-apple-darwin15.6.0"
```

```
## [1] "R version 3.6.3 (2020-02-29)"
```

Data is originally from the 1974 Motor Trend US magazine.