Project

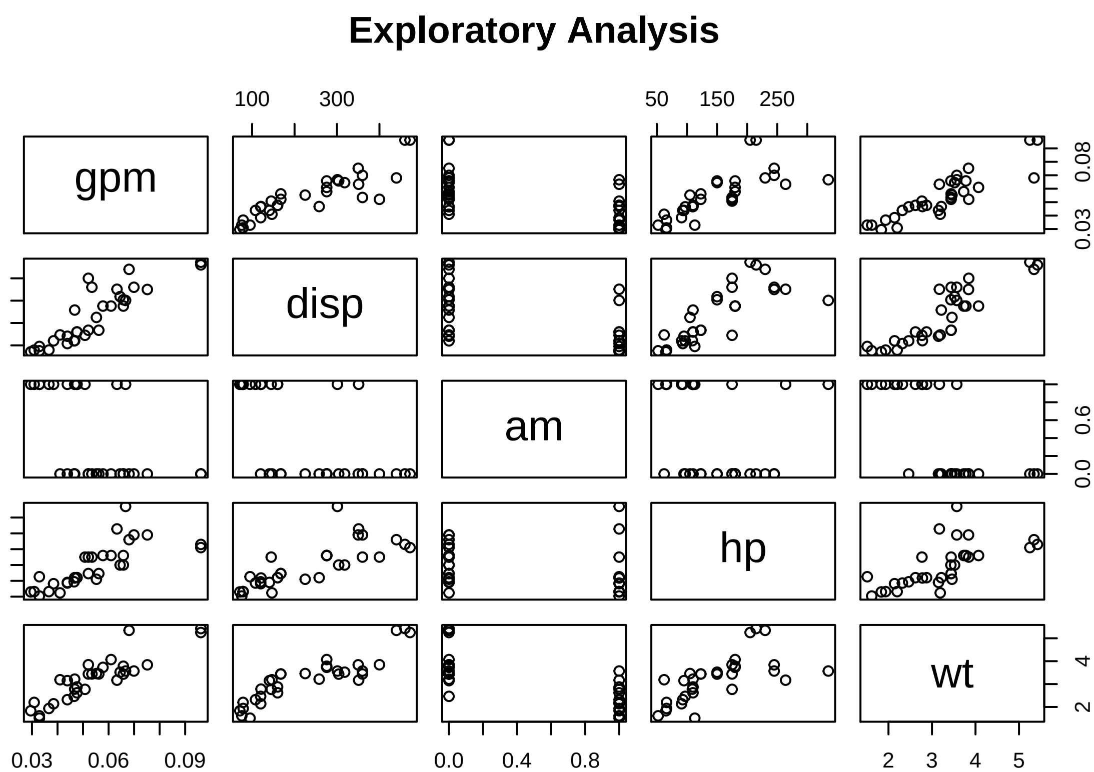
Winston

November 9, 2014

Executive Summary

exploratory data analysis

```r
data(mtcars)
library(plyr)

mtcars2=mutate(mtcars, gpm=1/mpg)

pairs(~gpm+disp+am+hp+wt,data=mtcars2,
   main="Exploratory Analysis")
```



There are data for 32 different cars in the data set.
The data present are *mpg, cyl, disp, hp, drat, wt, qsec, vs, am, gear, carb.*

mpg Miles/(US) gallon cyl Number of cylinders disp Displacement (cu.in.) hp Gross horsepower drat Rear axle ratio wt Weight (lb/1000) qsec 1/4 mile time vs V/S am Transmission (0 = automatic, 1 = manual) gear Number of forward gears carb Number of carburetors

An exploratory analysis of the dependcies of mpg on several of the factors in the mtcars dataset show there are multiple dependencies. To understand how transmission type affects mpg we need account for these other factors.

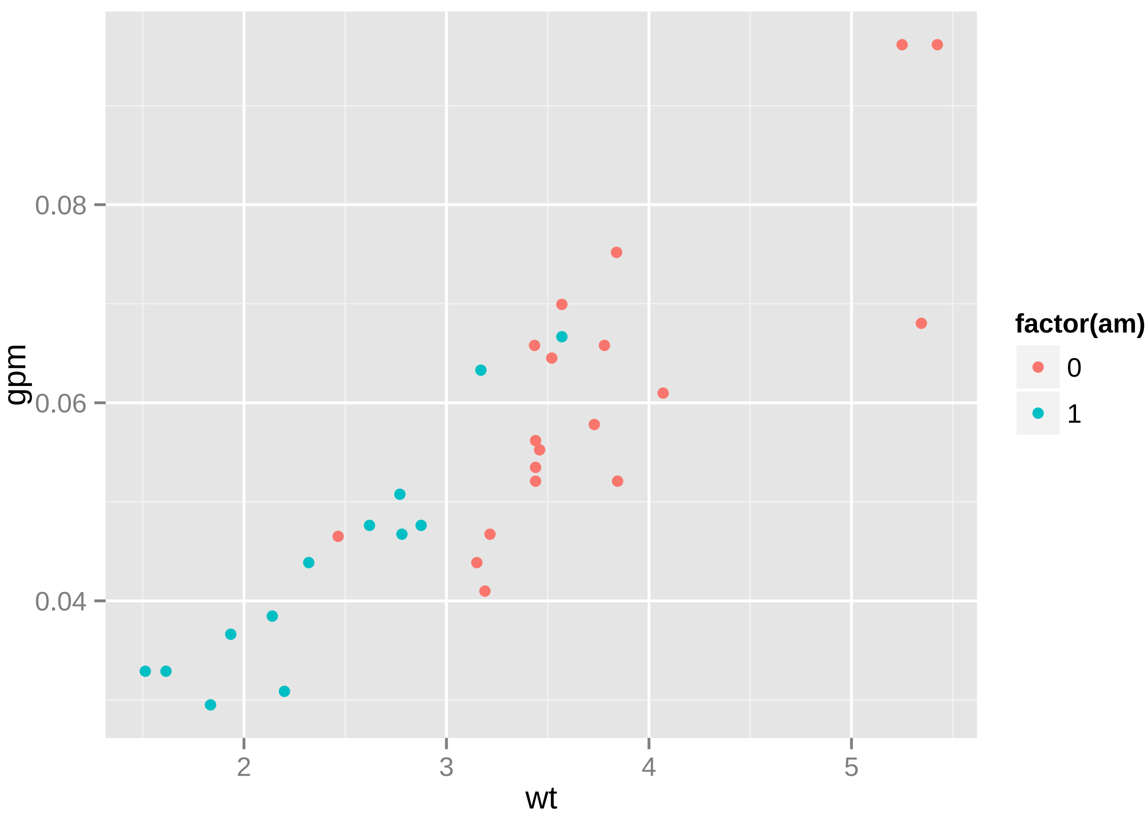
As you can see from the graph, displacement and horsepower tend to follow weigth. So we'll focus on correcting for the weight factor in mpg and look at residual.

linear model

The linear model shown in the graph above clearly does not fit the observed behavior.
The reason the linear fit doesn't work well is because mpg is a kind of *reciprocal* energy (gallons of gasoline are equivalent to a measure of energy). The energy to accelerate a mass m to speed v is proportional to m so it makes sense to look at a reciprocal.

This suggests an inverse relationship between *mpg* and *weight* should be explored.

```
library(plyr)
library(ggplot2)

p<-ggplot(mtcars2, aes(wt, gpm, color=factor(am))) + geom_point()

p
```



nonlinear model of mpg versus wt

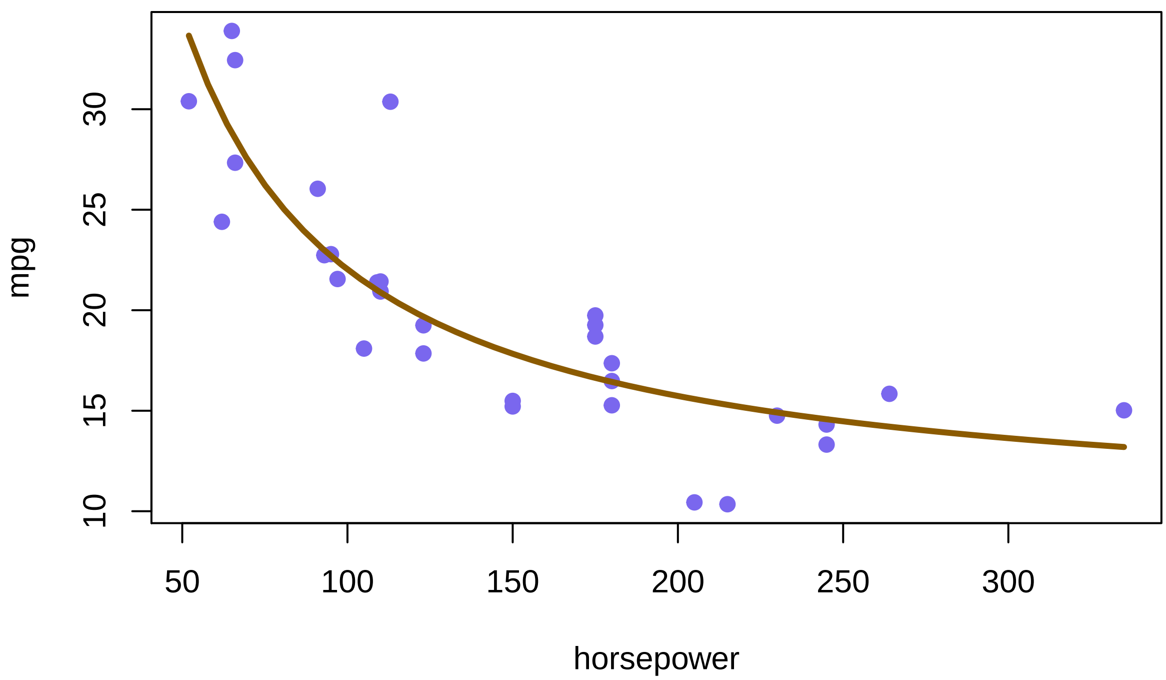
To fit the data try a functional relationship $mpg\sim a + b/wt$ with a and b constants to be determined. The graph below shows the fit.

```
mtcarsm <- mutate
```

```r
carnls <- nls(mpg~a1 + b1/hp, mtcars, start = list(a1=1, b1=50))


carnls
```

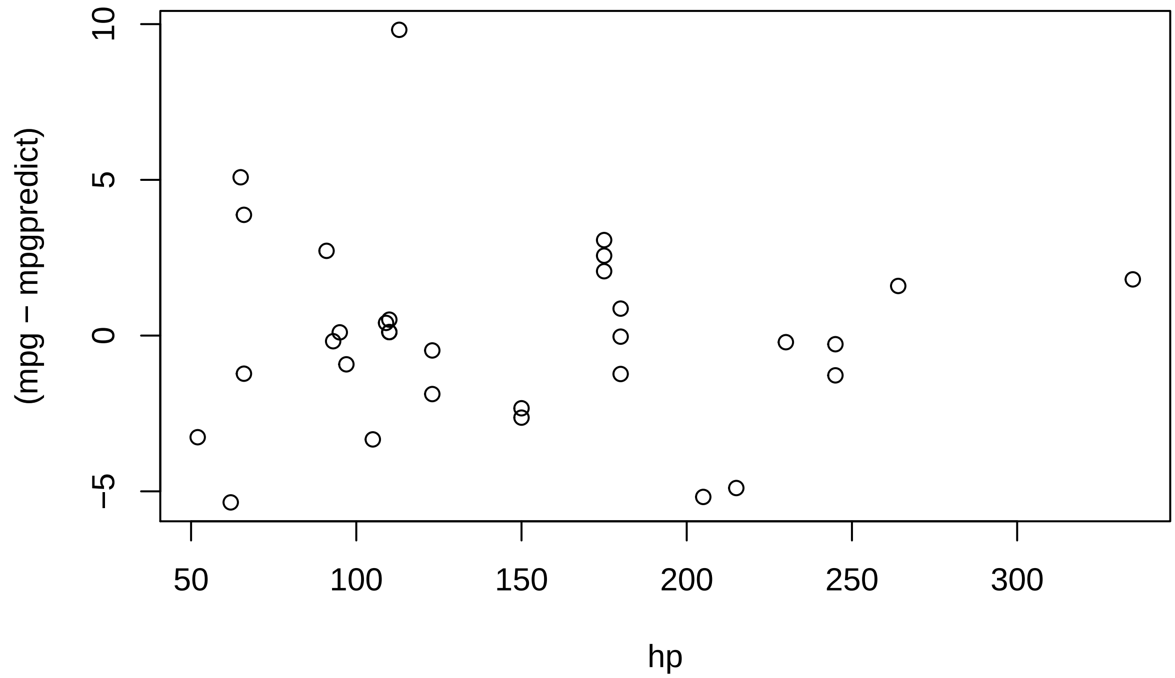
```
## Nonlinear regression model
##   model: mpg ~ a1 + b1/hp
##    data: mtcars
##       a1       b1
##    9.434 1259.881
##  residual sum-of-squares: 294.9
##
## Number of iterations to convergence: 1
## Achieved convergence tolerance: 3.841e-08
```

```r
ss<-seq(from = min(mtcars$hp), to = max(mtcars$hp), length = 50)


plot(jitter(mpg,4)~hp, data=mtcars, pch=19, col="slateblue2", xlab="horsepower", ylab="mpg")


yy <- predict(carnls,list(hp = ss))

lines(ss, yy, col = "orange4", lwd=3)
```

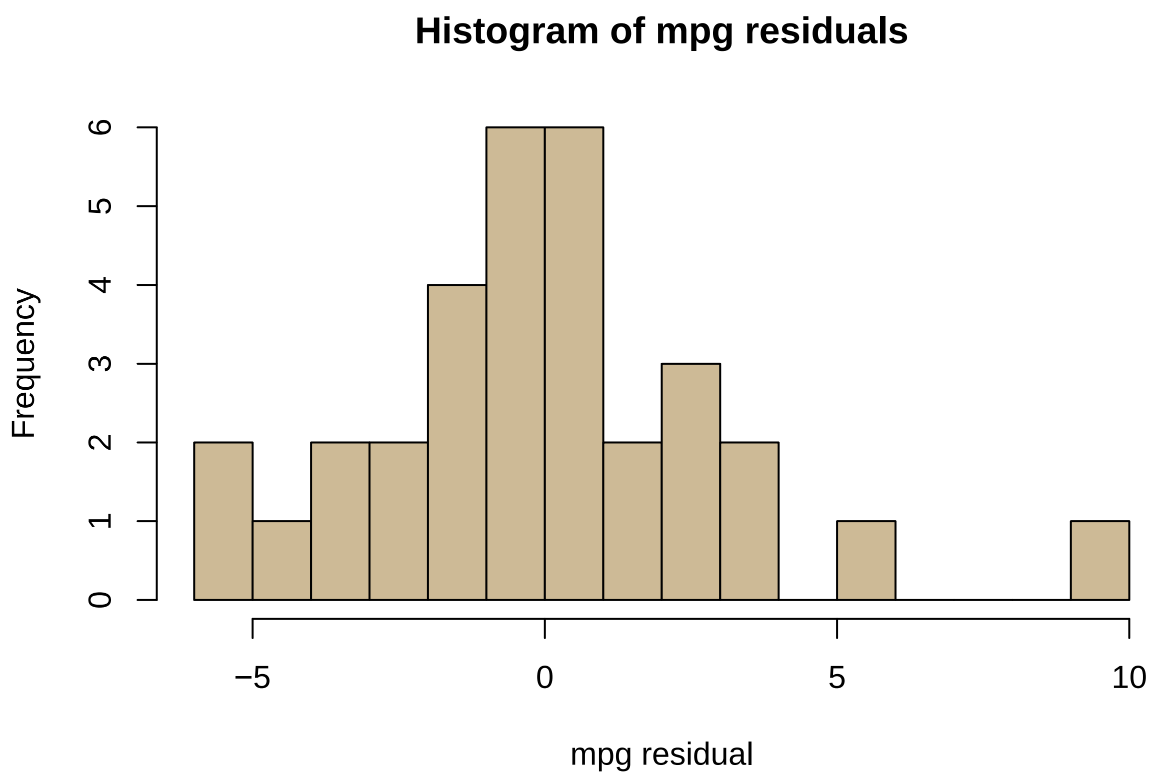


Model Residuals

```r
mpgpredict<-predict(carnls)

plot((mpg-mpgpredict)~hp, data=mtcars)
```

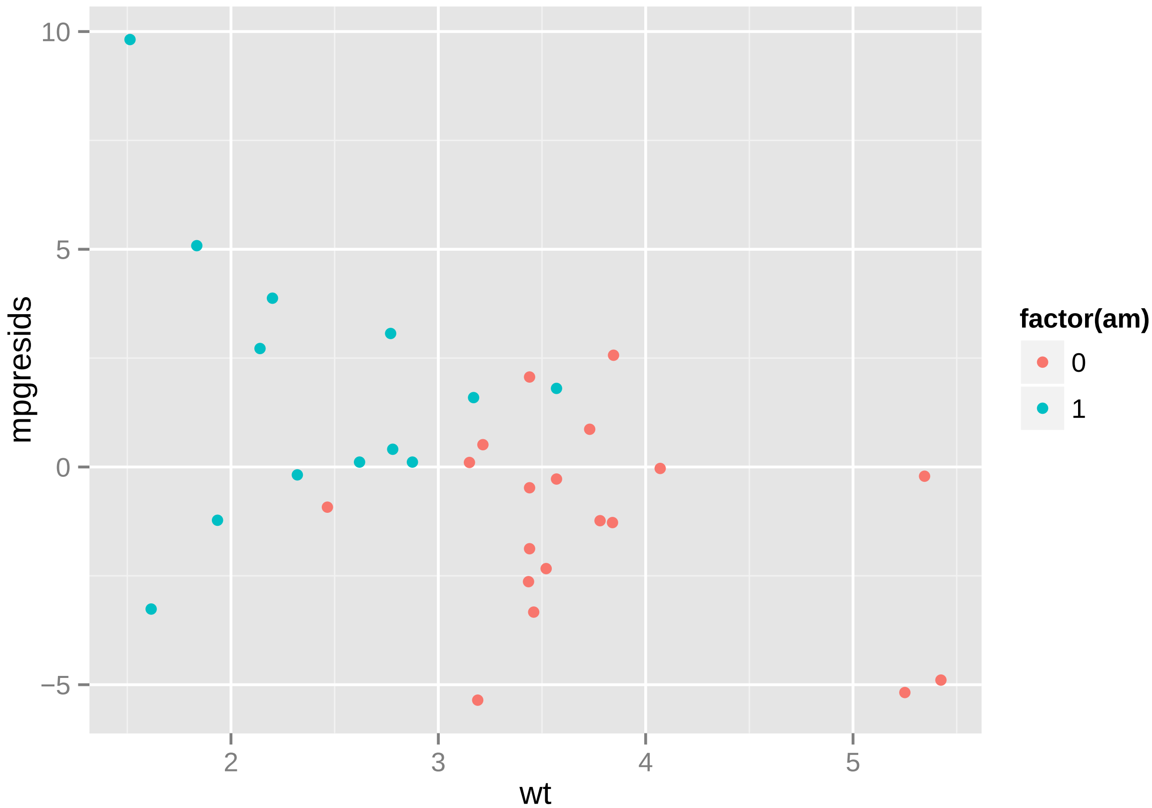



```
hist(mtcars$mpg-mpgpredict, breaks=18, col="wheat3", xlab="mpg residual", main="Histogram of mpg residu
```

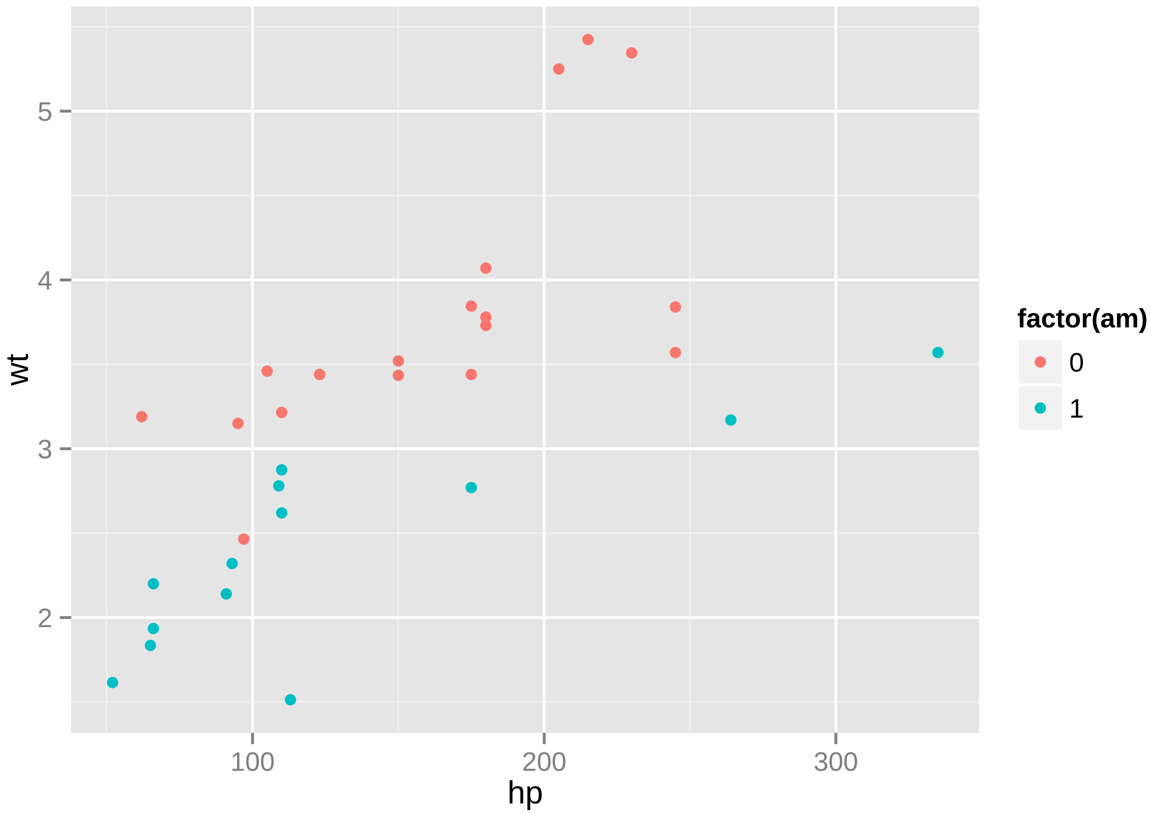



```
library(plyr)

mpgresids <- mtcars$mpg-mpgpredict

mtcars2<-cbind(mtcars, mpgresids)
```

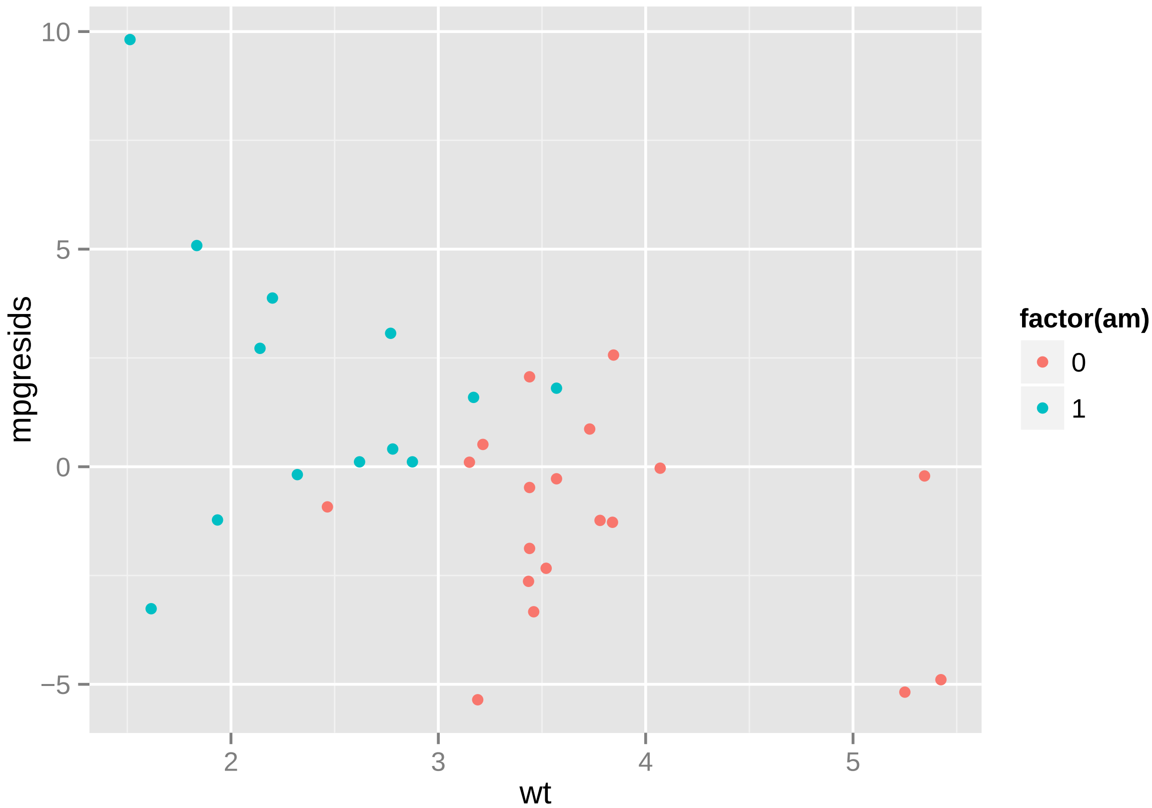
```
p<-ggplot(mtcars2, aes(wt, mpgresids, color=factor(am))) + geom_point()

p
```



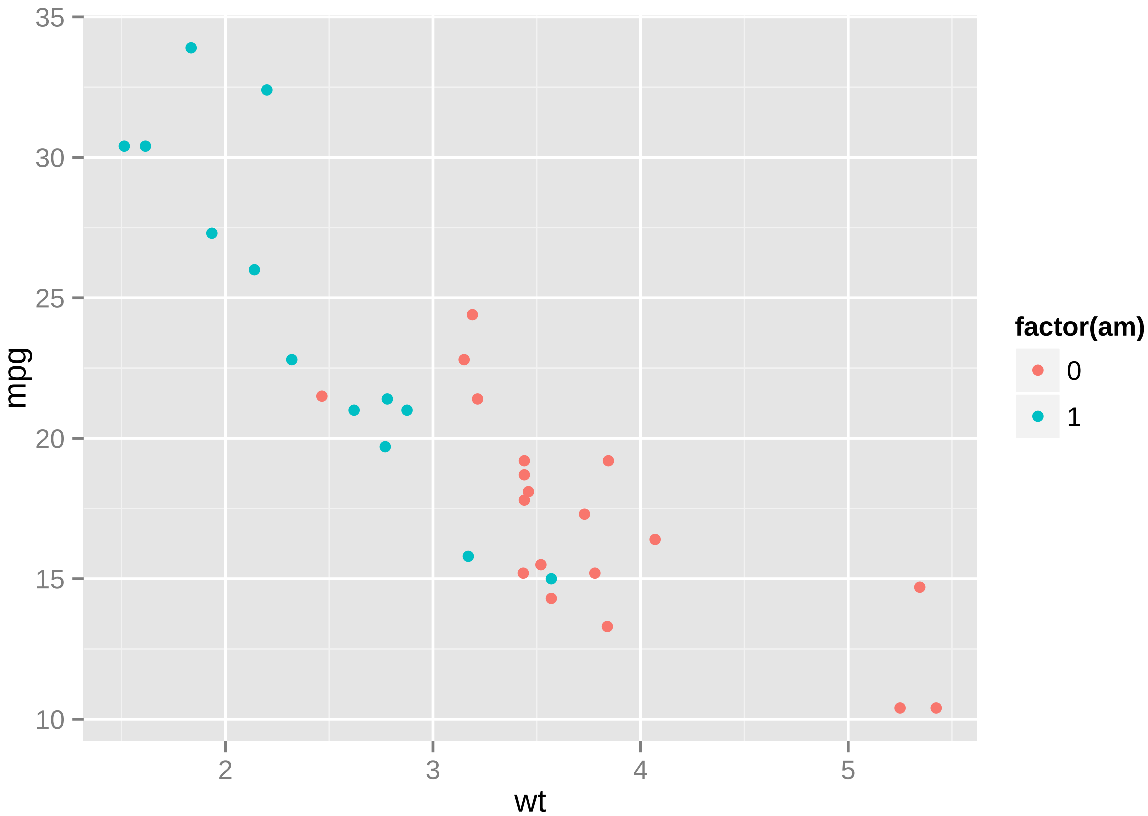

```
p<-ggplot(mtcars, aes(hp, wt, color=factor(am))) + geom_point()

p
```




```
library(plyr)

## calculate residuals ad bind to mtcars to do some plotting

mpgresids <- mtcars$mpg-mpgpredict

mtcars2<-cbind(mtcars, mpgresids)

p<-ggplot(mtcars2, aes(wt, mpgresids, color=factor(am))) + geom_point()

p
```




```
p<-ggplot(mtcars, aes(wt, mpg, color=factor(am))) + geom_point()

p
```




```
##here we do statistics on the residuals


autoresids <- mtcars2[mtcars$am==0,]
manresids <- mtcars2[mtcars$am==1,]

meanautoresids<-mean(autoresids$mpgresids)
sdautoresids<-sd(autoresids$mpgresids)
nautoresids <- nrow(autoresids)

meanmanresids<-mean(manresids$mpgresids)
sdmanresids<-sd(manresids$mpgresids)
nmanresids <- nrow(manresids)


aaa<-t.test(autoresids$mpgresids, manresids$mpgresids, alternative="less")
pv<-aaa$p.value
lc<-aaa$conf.int[1]
uc<-aaa$conf.int[2]
```