Virtual Lab - Diagnostics

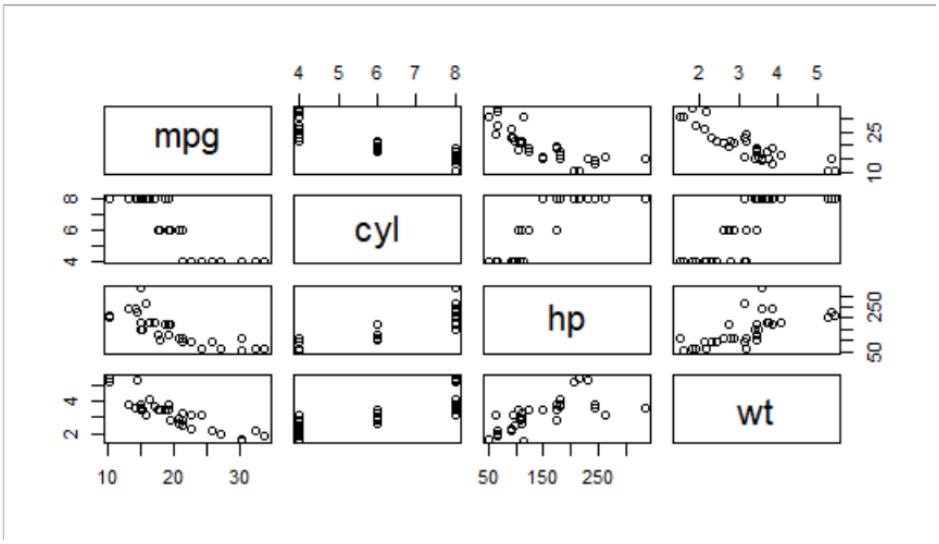
We will use of the mtcars data set in R (in the car package). For this problem set, use mpg as your response and take a look at the model with cyl, wt and hp. Look at residuals to see if you have normality, misspecified model, homoscedastic variance, etc. Try to find the most ***appropriate*** model using ONLY these three variables.

$$\hat{Y}_i = 55.82 - 2x_{cyl} - 0.14x_{hp} - 8.1x_{wt} + 0.66x_{wt}^2 + 0.02x_{cyl}x_{hp}$$

AIC = 146.365

lab.dat=mtcars[,c(1,2,4,6)]

pairs(lab.dat)



> empty.model=lm(mpg~1,lab.dat)

> full.model=lm(mpg~.^2 + I(hp^2)+I(wt^2),data=lab.dat)

>for.model=step(empty.model,scope=list(lower=empty.model,upper=full.model),direction='forward',data=lab.dat)

mpg ~ wt + cyl + hp + wt:cyl + cyl:hp

>back.model=step(full.model,scope=list(lower=empty.model,upper=full.model),direction="backward",data=lab.dat)

mpg ~ cyl + hp + wt + I(wt^2) + cyl:hp

> mod.1=lm(mpg~.+I(wt^2)+cyl:hp,data=lab.dat)

> summary(mod.1)

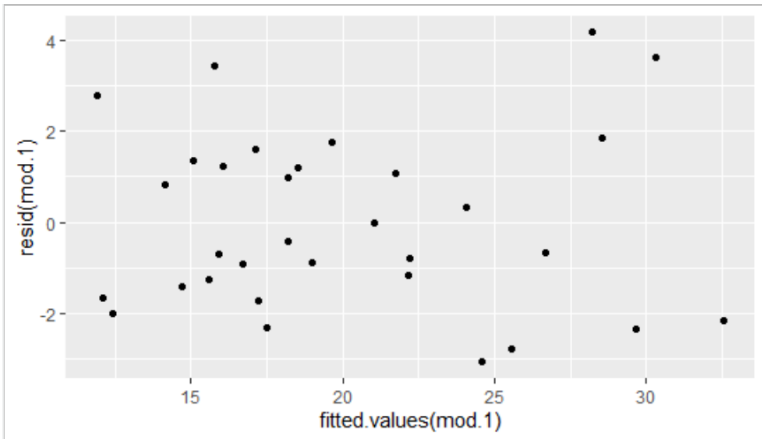
```
Estimate Std. Error t value Pr(>|t|)
(Intercept) 55.816981   5.020960  11.117 2.24e-11 ***
cyl         -1.996473   0.842682  -2.369  0.02554 *
hp          -0.135699   0.051254  -2.648  0.01359 *
wt          -8.095960   2.536614  -3.192  0.00368 **
I(wt^2)      0.662544   0.327277   2.024  0.05331 .
cyl:hp       0.015011   0.006588   2.278  0.03116 *
```

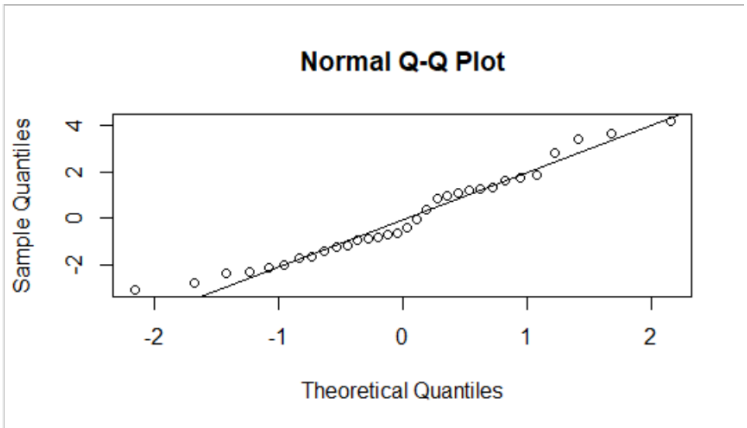
> ggplot(mod.1,aes(x=fitted.values(mod.1),y=resid(mod.1)))+geom_point()

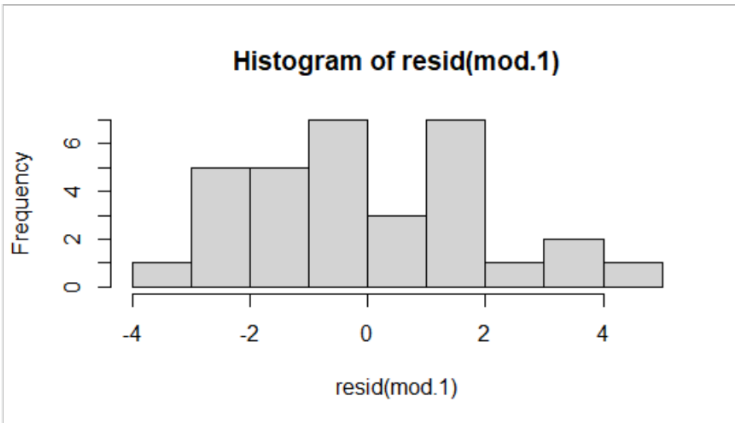
> qqnorm(resid(mod.1))

> qqline(resid(mod.1))

> hist(resid(mod.1))








```
> shapiro.test(resid(mod.1))

        Shapiro-Wilk normality test

data:  resid(mod.1)
W = 0.95822, p-value = 0.2454


> AIC(mod.1)
[1] 146.365
> mod.2=lm(mpg~.+I(wt^2)+cyl:hp+wt:cyl ,data=lab.dat)
> AIC(mod.2)
[1] 148.3523
```