Plot Comparison

JC

9/9/2017

Note

Plot Comparison based on plan.

Data & Libs

```r
library(tidyverse)
library(broom)
library(gridExtra)

# Traditional Model
lm.1 <- lm(mpg ~ wt,data=mtcars)

# Generic name
My.Mod <- lm.1
# Tidy output
Tidy.Mod <- augment(My.Mod)

# Attributes
class(My.Mod)
```

```
## [1] "lm"
```

```r
attributes(My.Mod)
```

```
## $names
##  [1] "coefficients"  "residuals"     "effects"       "rank"
##  [5] "fitted.values" "assign"        "qr"            "df.residual"
##  [9] "xlevels"       "call"          "terms"         "model"
##
## $class
## [1] "lm"
```

```r
head(Tidy.Mod,4)
```

```
##        .rownames  mpg    wt  .fitted   .se.fit     .resid       .hat
## 1      Mazda RX4 21.0 2.620 23.28261 0.6335798 -2.2826106 0.04326896
## 2  Mazda RX4 Wag 21.0 2.875 21.91977 0.5714319 -0.9197704 0.03519677
## 3     Datsun 710 22.8 2.320 24.88595 0.7359177 -2.0859521 0.05837573
## 4 Hornet 4 Drive 21.4 3.215 20.10265 0.5384424  1.2973499 0.03125017
##     .sigma     .cooksd .std.resid
## 1 3.067494 0.013274072 -0.7661677
## 2 3.093068 0.001723963 -0.3074305
## 3 3.072127 0.015439367 -0.7057525
## 4 3.088268 0.003020558  0.4327511
```

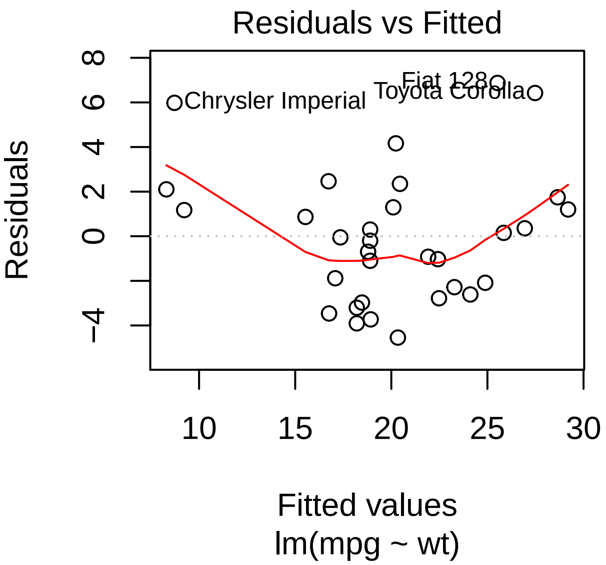
Notes axis scale/ aspect slightly different Text annotations logic ~ highlighting "Wigglyness" of line different ~ slightly diff parameter?

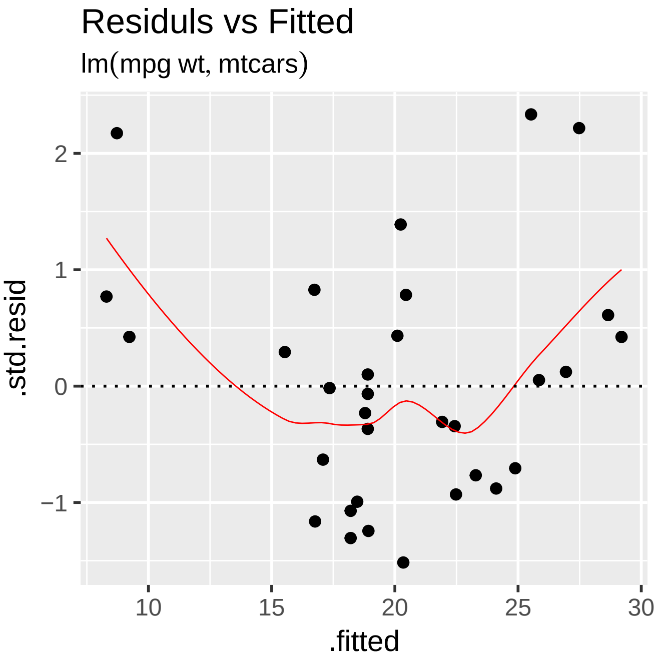
Plot1: Residules v Fitted

```
plot(My.Mod,which=1)

D1  <- augment(My.Mod) %>%
  ggplot(aes(x=.fitted,y=.std.resid)) +
  geom_point() +
  geom_smooth(se=FALSE,colour="red",size=.25) +
  geom_hline(yintercept=0,linetype=3) +
  labs(title="Residuls vs Fitted",subtitle=My.Mod$call)

D1
```



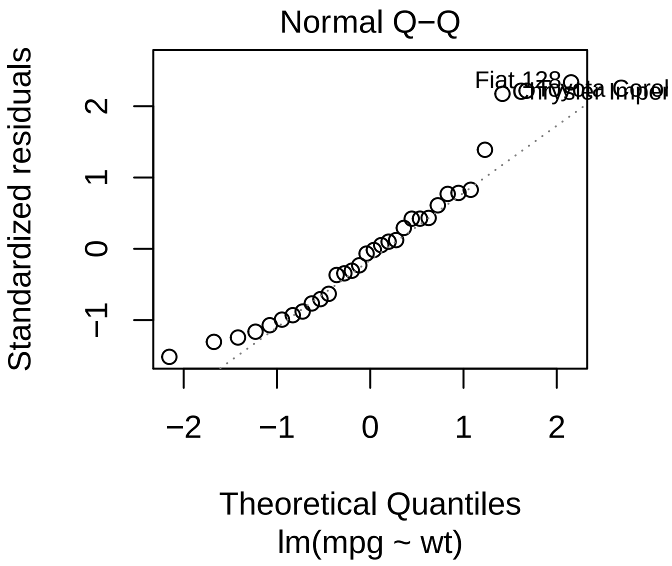


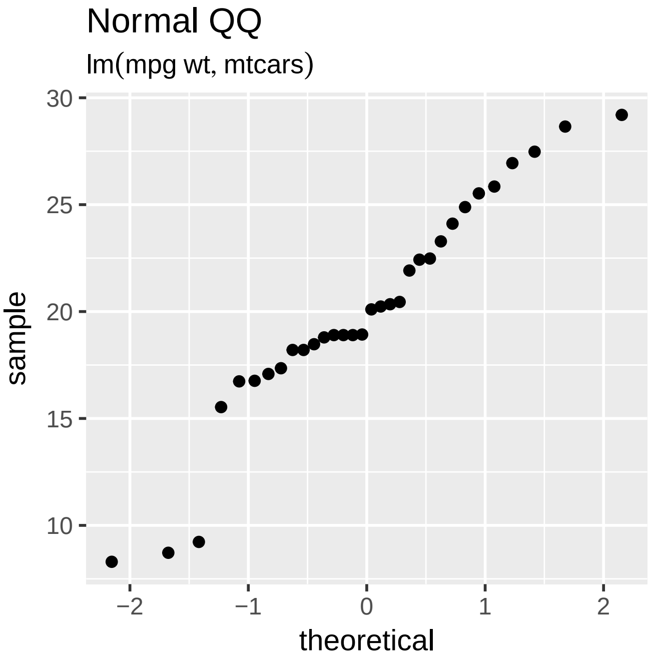
Plot2: Normal QQ

```
plot(My.Mod,which=2)

D2  <- augment(My.Mod) %>%
  ggplot(aes(sample=.fitted)) +
  stat_qq() +
  labs(title="Normal QQ",subtitle=My.Mod$call)

D2
```



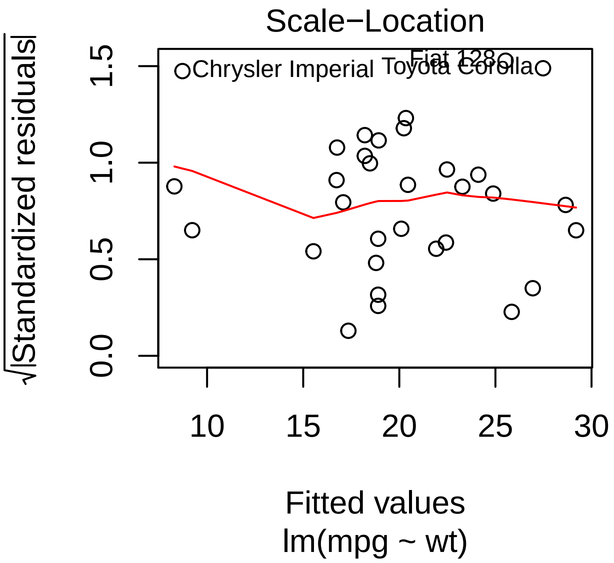


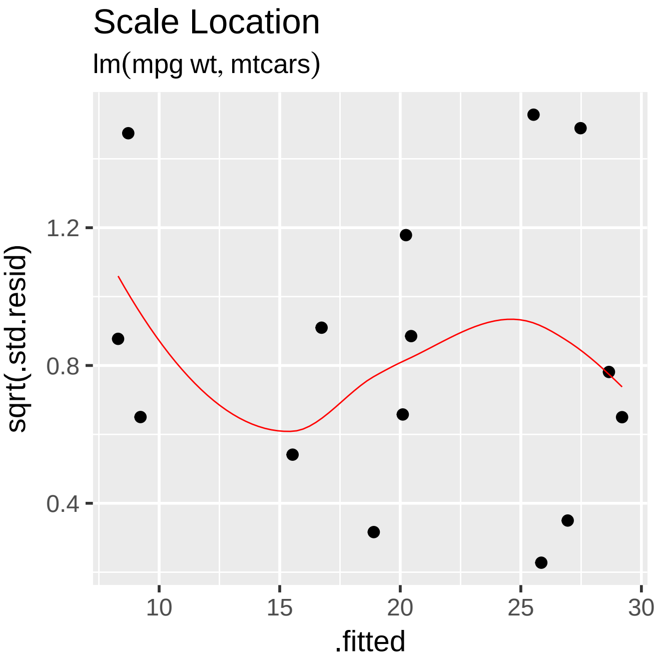
Plot3: Scale Location

```
plot(My.Mod,which=3)

D3  <- augment(My.Mod) %>%
  ggplot(aes(x=.fitted, y=sqrt(.std.resid))) +
  geom_point()+
  labs(title="Scale Location",subtitle=My.Mod$call) +
  geom_smooth(se=FALSE,colour="red",size=.25)

D3
```





Questions…

Missing Values? Why?
Line wigliness.What parameters?

Missing Values

```
# Raw
Tidy.Mod$.std.resid
```

```
##  [1] -0.76616765 -0.30743051 -0.70575249  0.43275114 -0.06681879
##  [6] -0.23148309 -1.30552216  1.38889709  0.78392687  0.10010803
## [11] -0.36728706  0.29288651 -0.01683789 -0.63159969  0.42296071
## [16]  0.76979873  2.17353314  2.33490215  0.61035691  2.21708271
## [21] -0.87964013 -0.99313634 -1.24418015 -1.16279098  0.82771968
## [26]  0.12244407  0.05177187  0.42254270 -1.51549710 -0.93086929
## [31] -1.07151943 -0.34388215
```

```
# Why so many NAN's?
sqrt(Tidy.Mod$.std.resid)
```

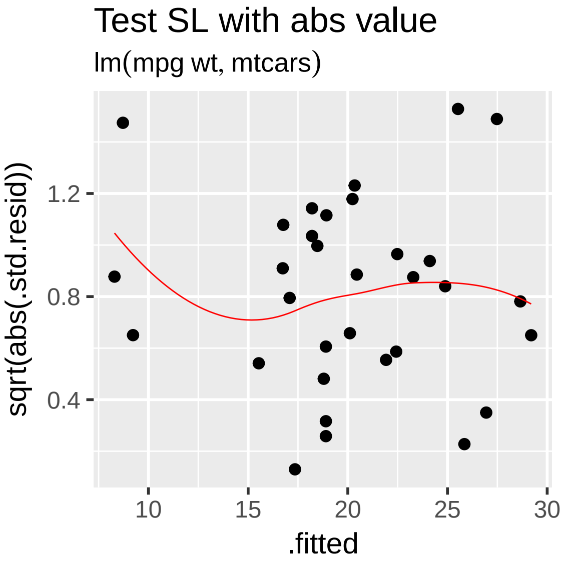
```
##  [1]       NaN       NaN       NaN 0.6578382       NaN       NaN       NaN
##  [8] 1.1785148 0.8853964 0.3163985       NaN 0.5411899       NaN       NaN
## [15] 0.6503543 0.8773817 1.4742907 1.5280387 0.7812534 1.4889871       NaN
## [22]       NaN       NaN       NaN 0.9097910 0.3499201 0.2275343 0.6500328
## [29]       NaN       NaN       NaN       NaN
```

```
#Try Abs value
sqrt(abs(Tidy.Mod$.std.resid))
```

```
##  [1] 0.8753100 0.5544642 0.8400908 0.6578382 0.2584933 0.4811269 1.1425945
##  [8] 1.1785148 0.8853964 0.3163985 0.6060421 0.5411899 0.1297609 0.7947325
## [15] 0.6503543 0.8773817 1.4742907 1.5280387 0.7812534 1.4889871 0.9378913
## [22] 0.9965623 1.1154282 1.0783279 0.9097910 0.3499201 0.2275343 0.6500328
## [29] 1.2310553 0.9648157 1.0351422 0.5864147
```

```
D3abs <- augment(My.Mod) %>%
  ggplot(aes(x=.fitted, y=sqrt(abs(.std.resid)))) +
  geom_point()+
  labs(title="Test SL with abs value",subtitle=My.Mod$call) +
  geom_smooth(se=FALSE,colour="red",size=.25)

D3abs
```



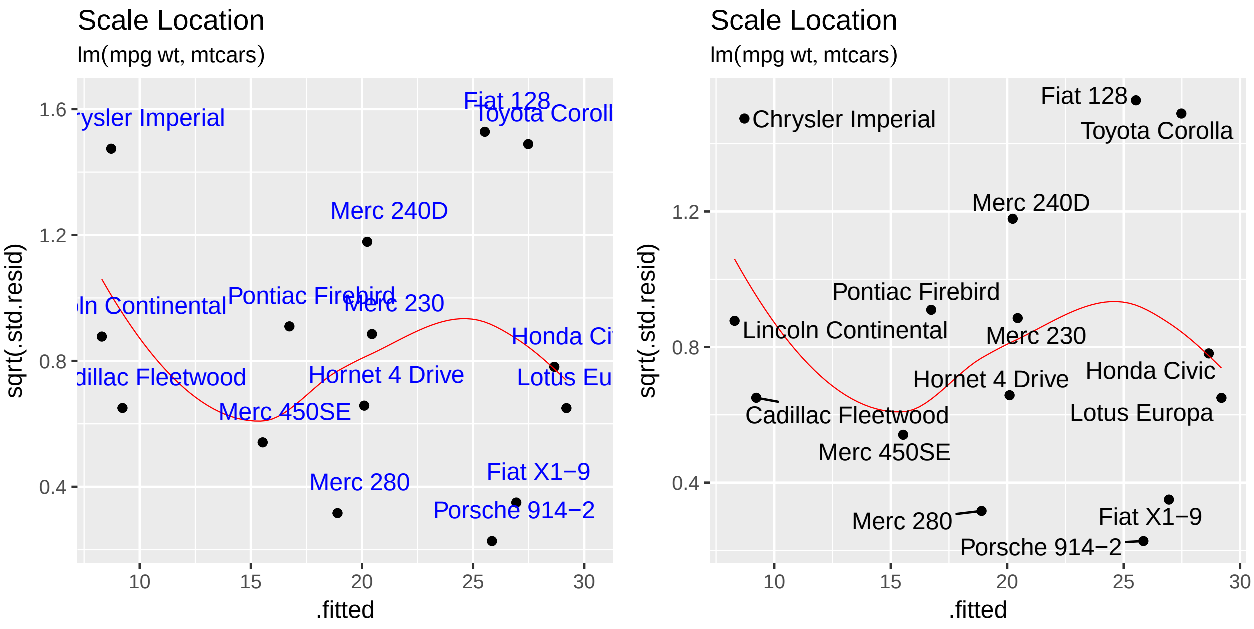
with text labels added

```
## text?
D3c <- D3 + geom_text(aes(label=.rownames),
               nudge_x = 1, nudge_y = .1,
               colour="blue")

# extra txt library
library(ggrepel)

D3d <- D3 + geom_text_repel(aes(label = .rownames))

#grid extra
grid.arrange(D3c, D3d, ncol=2)
```

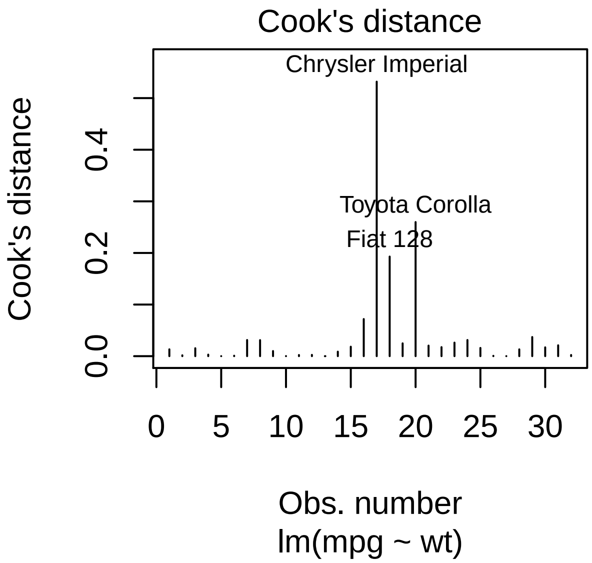


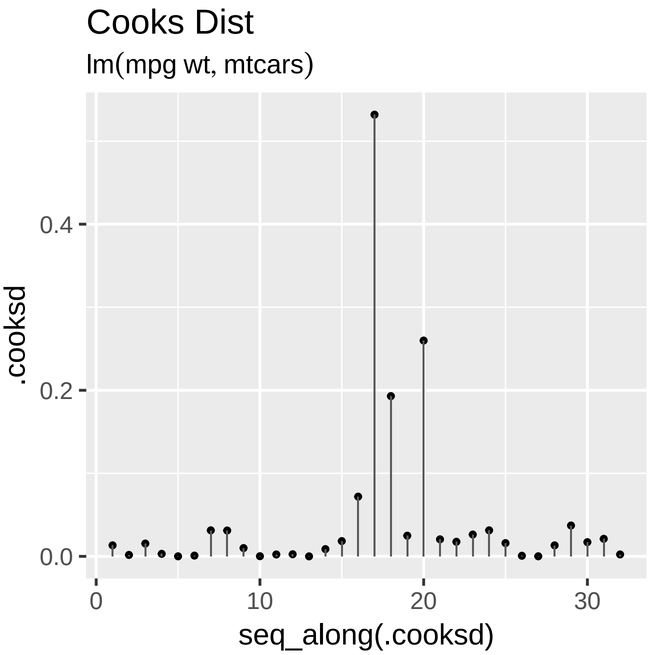
Plot4: Cooks Distance

```
#Cooks Distane
plot(My.Mod,which=4)

D4 <- augment(My.Mod) %>%
  ggplot(aes(x=seq_along(.cooksd), y=.cooksd)) +
  geom_point(size = .75) +
  geom_col(width = .1) +
  labs(title="Cooks Dist",subtitle=My.Mod$call)

D4
```



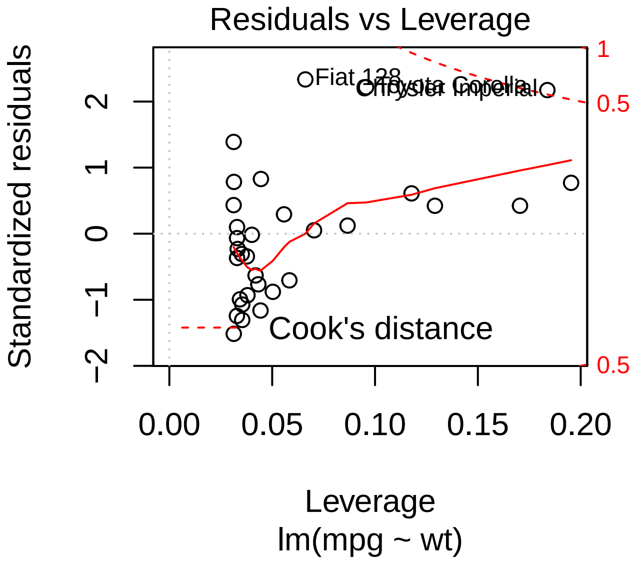


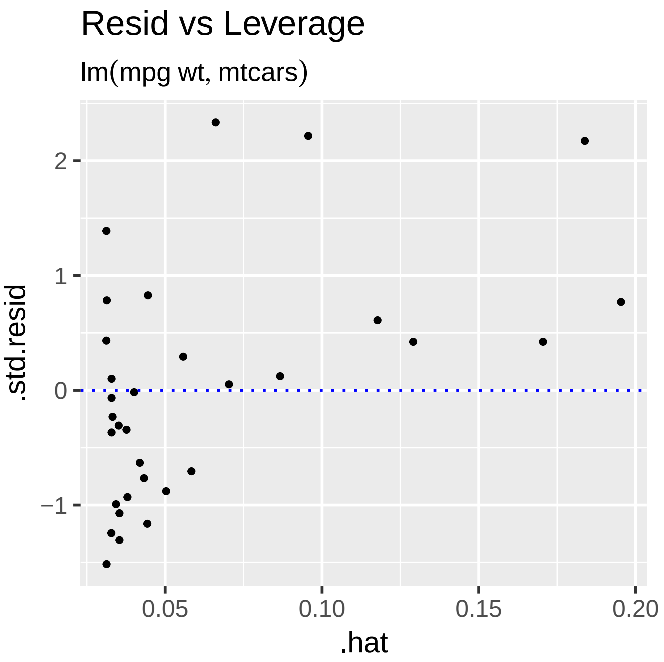
Plot5: Resid vs Leverage

```
# Resid vs Leverage
plot(My.Mod,which=5)

D5 <- augment(My.Mod) %>%
  ggplot(aes(x=.hat, y=.std.resid)) +
  geom_point(size=.75) +
  geom_hline(yintercept=0,linetype=3,colour="blue") +
  labs(title="Resid vs Leverage",subtitle=My.Mod$call)

D5
```


Residuals vs Leverage
Standardized residuals
−2
−1
0
1
2
0.00
0.05
0.10
0.15
0.20
Leverage
lm(mpg ~ wt)
Fiat 128
Toyota Corolla
Chrysler Imperial
Cook's distance
1
0.5
0.5

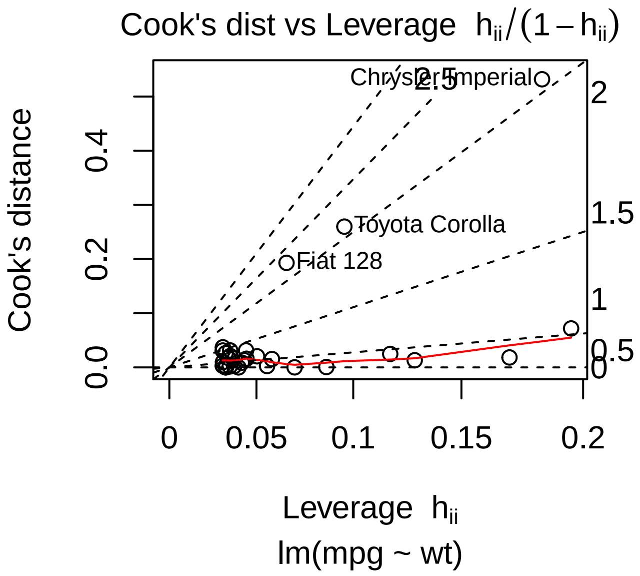

Resid vs Leverage
lm(mpg wt, mtcars)
.std.resid
−1
0
1
2
0.05
0.10
0.15
0.20
.hat

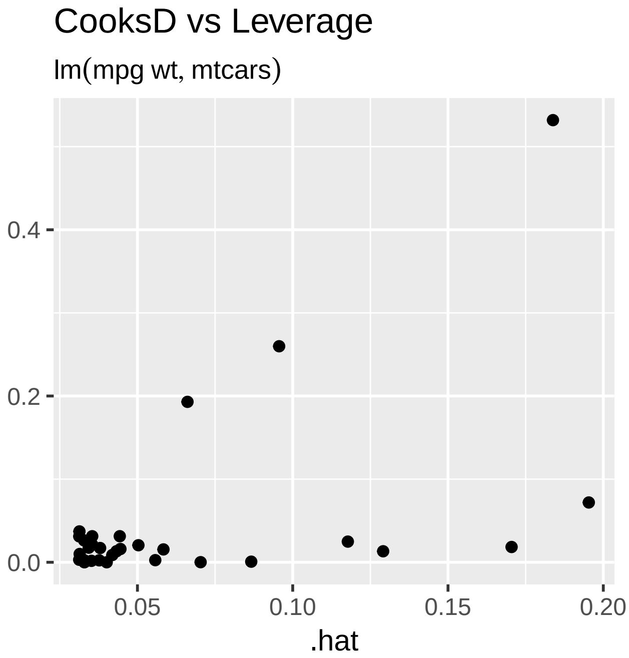
Plot6: CooksSD vs Leverage

```
plot(My.Mod,which=6)

D6 <- augment(My.Mod) %>%
  ggplot(aes(x=.hat, y=.cooksd)) +
  geom_point() +
  labs(title="CooksD vs Leverage",subtitle=My.Mod$call)

D6
```

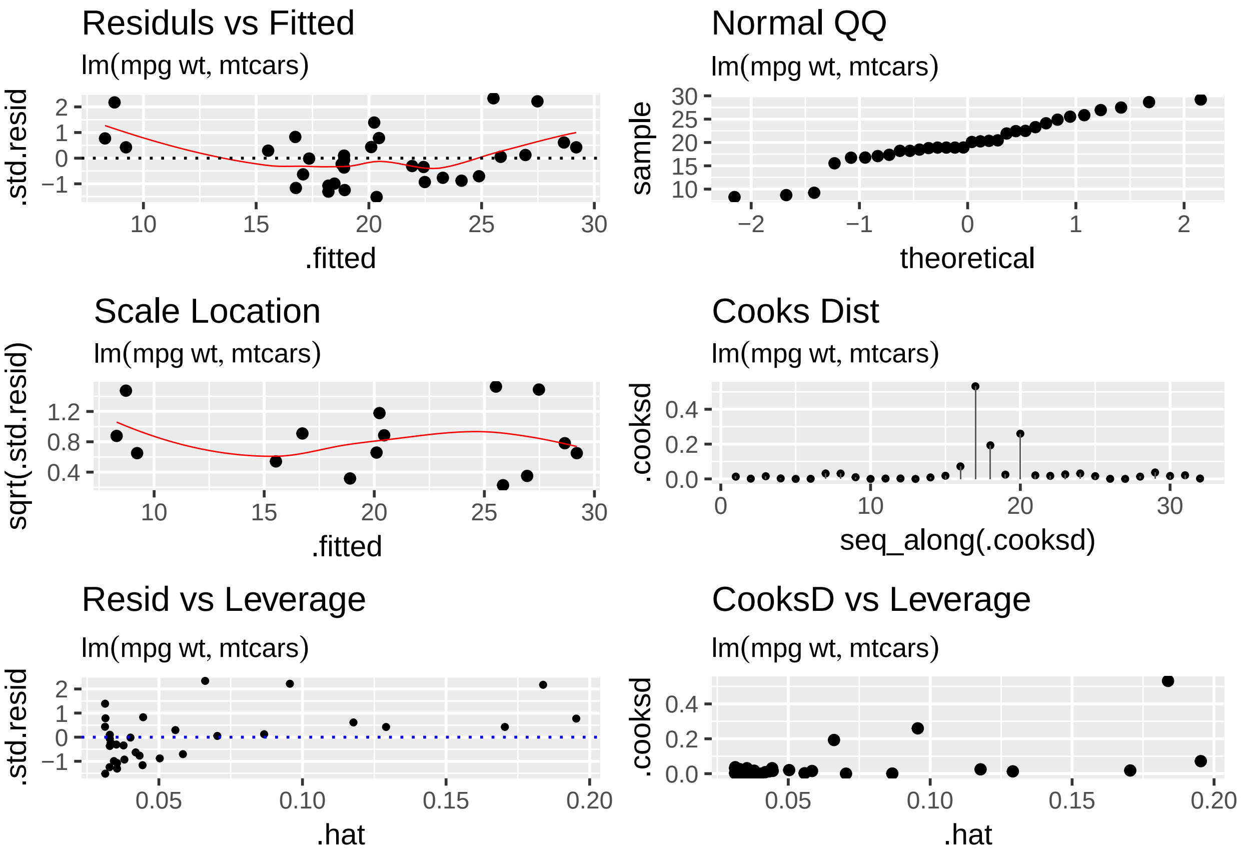




Panel

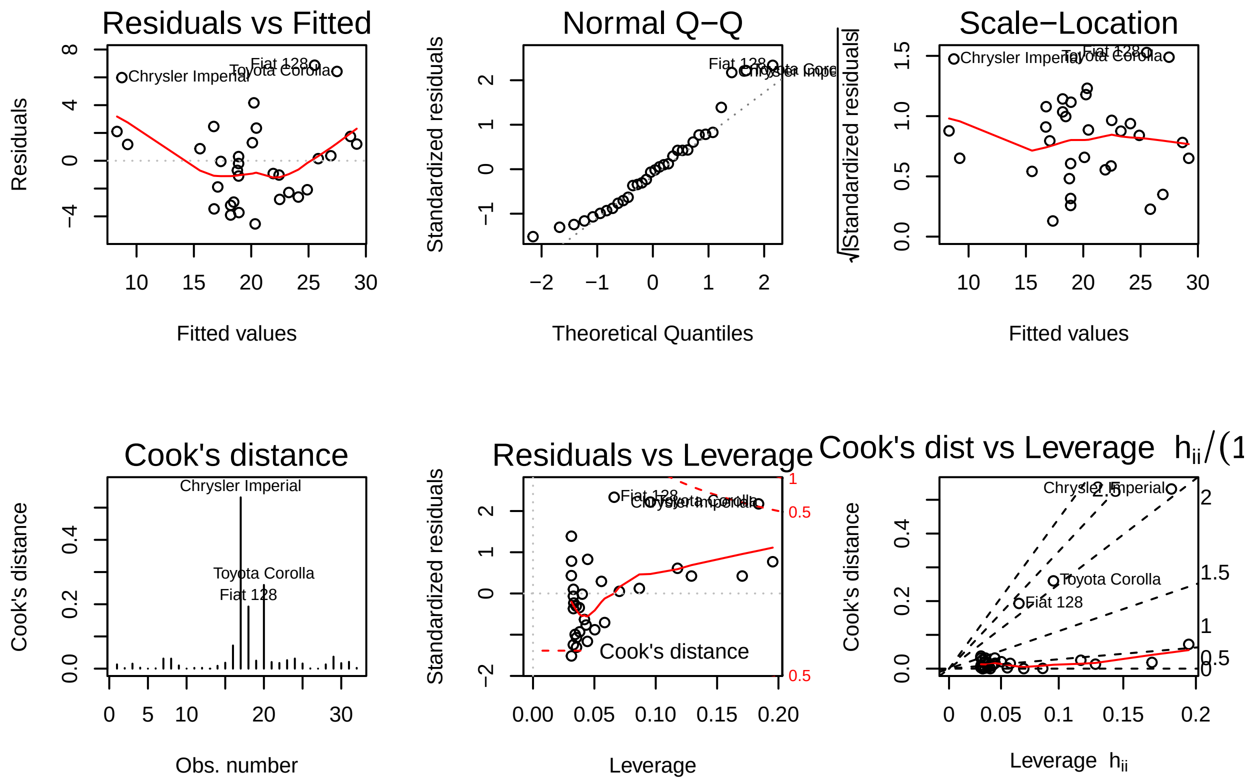
```
#library(gridExtra)

D.Panel <- grid.arrange(D1, D2, D3, D4, D5, D6, ncol=2)
D.Panel
```



```
## TableGrob (3 x 2) "arrange": 6 grobs
##   z     cells    name           grob
## 1 1 (1-1,1-1) arrange gtable[layout]
## 2 2 (1-1,2-2) arrange gtable[layout]
## 3 3 (2-2,1-1) arrange gtable[layout]
## 4 4 (2-2,2-2) arrange gtable[layout]
## 5 5 (3-3,1-1) arrange gtable[layout]
## 6 6 (3-3,2-2) arrange gtable[layout]
```

```r
par(mfrow=c(2,3))
plot(My.Mod,which=1:6)
```


Residuals vs Fitted
Residuals
Fitted values
Chrysler Imperial
Toyota Corolla
Fiat 128
−4
0
4
8
10
15
20
25
30
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Fiat 128
Toyota Coro
Chrysler Impe
−1
0
1
2
−2
−1
0
1
2
Scale−Location
√|Standardized residuals|
Fitted values
Chrysler Imperial
Toyota Corolla
Fiat 128
0.0
0.5
1.0
1.5
10
15
20
25
30
Cook's distance
Cook's distance
Obs. number
Chrysler Imperial
Toyota Corolla
Fiat 128
0.0
0.2
0.4
0
5
10
20
30
Residuals vs Leverage
Standardized residuals
Leverage
Fiat 128
Toyota Corolla
Chrysler Imperial
Cook's distance
1
0.5
0.5
−2
0
1
2
0.00
0.05
0.10
0.15
0.20
Cook's dist vs Leverage $h_{ii}/(1$
Cook's distance
Leverage h_{ii}
Chrysler Imperial
Toyota Corolla
Fiat 128
2.5
2
1.5
1
0.5
0
0.0
0.2
0.4
0
0.05
0.1
0.15
0.2