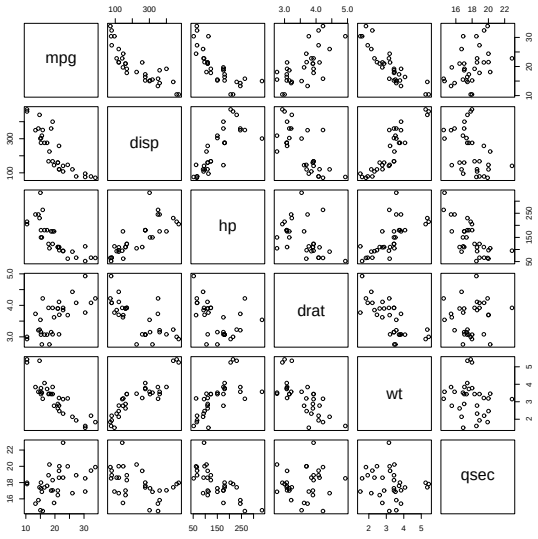
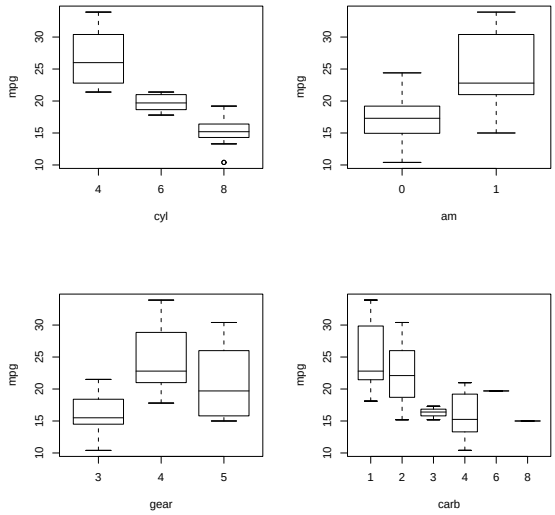
Motor Trend Car Tests

```
> names(mtcars)
 [1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"   "qsec" "vs"
 [9] "am"   "gear" "carb"
> plot(mtcars[,c("mpg","disp","hp","drat","wt","qsec")])
```



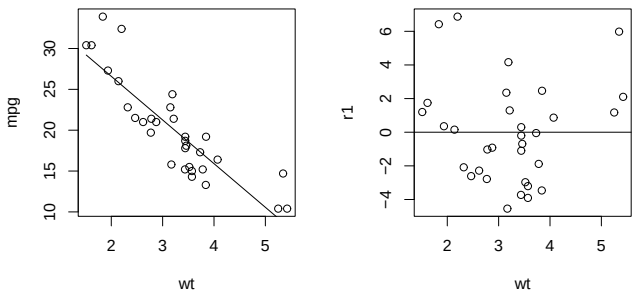
Motor Trend Car Tests

```
> for (i in c("cyl","am","gear","carb")) {
+   mtcars[,i] <- factor(mtcars[,i])
+ }
> attach(mtcars)
> plot(mpg ~ cyl)
> plot(mpg ~ am)
> plot(mpg ~ gear)
> plot(mpg ~ carb)
```



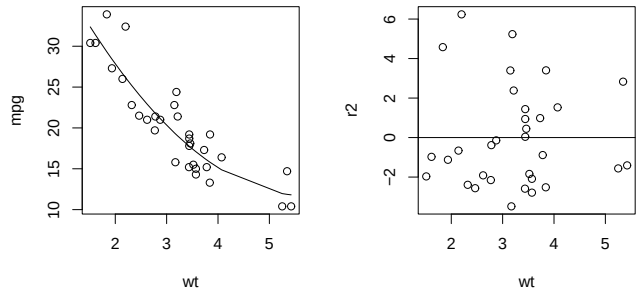
Regression on Weight

```
> cor(mtcars$mpg, mtcars[,c("disp","hp","drat","wt","qsec")])
          disp         hp      drat         wt     qsec
[1,] -0.8475514 -0.7761684 0.6811719 -0.8676594 0.418684
> l1 <- lm(mpg ~ wt, data=mtcars)
> summary(l1)
[ . . . ]
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  37.2851     1.8776  19.858  < 2e-16 ***
wt           -5.3445     0.5591  -9.559 1.29e-10 ***
[ . . . ]
Residual standard error: 3.046 on 30 degrees of freedom
Multiple R-Squared: 0.7528,Adjusted R-squared: 0.7446
F-statistic: 91.38 on 1 and 30 DF,  p-value: 1.294e-10
> plot(mpg ~ wt)
> o <- order(wt); lines(wt[o],fitted(l1)[o])
> plot(resid(l1) ~ wt); abline(h=0)
```



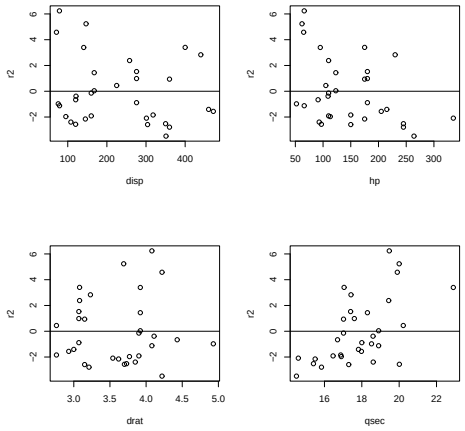
Quadratic Fit

```
> l2 <- update(l1, .~.+I(wt^2))
> anova(l1,l2)
[ . . . ]
Model 1: mpg ~ wt
Model 2: mpg ~ wt + I(wt^2)
  Res.Df     RSS Df Sum of Sq      F  Pr(>F)
1     30 278.322
2     29 203.745  1    74.576 10.615 0.00286 **
> plot(mpg ~ wt)
> o <- order(wt); lines(wt[o],fitted(l2)[o])
> plot(resid(l2) ~ wt); abline(h=0)
```



Residuals vs. Other Vars

```
> r2 <- resid(l2)
> plot(r2 ~ disp); abline(h=0)
> plot(r2 ~ hp); abline(h=0)
> plot(r2 ~ drat); abline(h=0)
> plot(r2 ~ qsec); abline(h=0)
```



One-Term Additions with `add1`

The `add1` function can be used to "grow" a model by one term within a given scope:

```
> add1(l2, ~.+disp+hp+drat+qsec, test="F")
Single term additions

Model:
mpg ~ wt + I(wt^2)
       Df Sum of Sq     RSS     AIC F value     Pr(F)
<none>              203.745  65.236
disp    1    30.705 173.040  62.009  4.9685 0.0340156 *
hp      1    59.452 144.293  56.196 11.5367 0.0020606 **
drat    1     0.276 203.470  67.193  0.0379 0.8470048
qsec    1    70.431 133.315  53.663 14.7925 0.0006339 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> 32*log(203.745/32)+2*3
[1] 65.23627
>
```

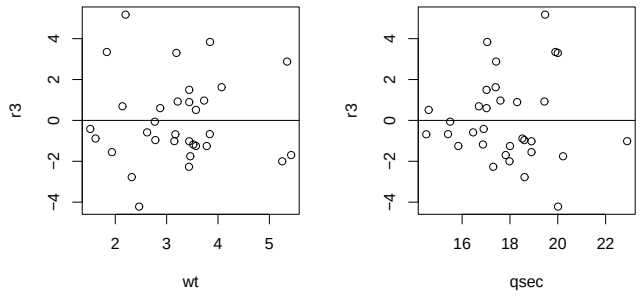
Besides comparing RSS values, you can look at the

$$\text{AIC} = n \log\left(\frac{\text{RSS}}{n}\right) + 2k$$

where n is number of obs and k is number of terms (including intercept) in the model. A lower AIC means a "better" model.

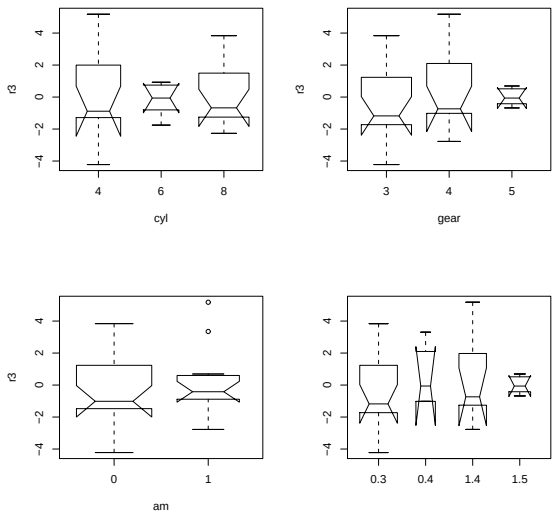
Weight and Quarter-Sec

```
> l3 <- update(l2,~.+qsec)
> summary(l3)
[ . . . ]
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  32.6418     5.6768   5.750 3.59e-06 ***
wt          -12.4331     2.0842  -5.965 2.01e-06 ***
I(wt^2)       1.0730     0.2970   3.613 0.001174 **
qsec          0.8599     0.2236   3.846 0.000634 ***
[ . . . ]
> anova(l3, update(l3, .~.+wt:qsec))
Model 1: mpg ~ wt + I(wt^2) + qsec
Model 2: mpg ~ wt + I(wt^2) + qsec + wt:qsec
  Res.Df     RSS Df Sum of Sq      F Pr(>F)
1     28 133.315
2     27 128.033  1     5.282 1.1138 0.3006
>
```



Useful Factors?

```
> r3 <- resid(l3)
> plot(r3 ~ cyl, notch=T, varwidth=T)
> plot(r3 ~ gear, notch=T, varwidth=T)
> plot(r3 ~ am, notch=T, varwidth=T)
> boxplot(r3 ~ am*gear, notch=T, varwidth=T)
```



Useful Factors?

```
> anova(l3, update(l3, .~cyl/.))
Analysis of Variance Table

Model 1: mpg ~ wt + I(wt^2) + qsec
Model 2: mpg ~ cyl + cyl:wt + cyl:I(wt^2) + cyl:qsec
  Res.Df     RSS Df Sum of Sq      F Pr(>F)
1     28 133.315
2     20 122.355  8    10.959 0.2239 0.9822
> anova(l3, update(l3, .~(am*gear)/.))
Analysis of Variance Table

Model 1: mpg ~ wt + I(wt^2) + qsec
Model 2: mpg ~ am + gear + am:gear + am:gear:wt
         + am:gear:I(wt^2) + am:gear:qsec
  Res.Df     RSS Df Sum of Sq     F Pr(>F)
1     28 133.315
2     16  86.693 12    46.622 0.717 0.7167
>
```

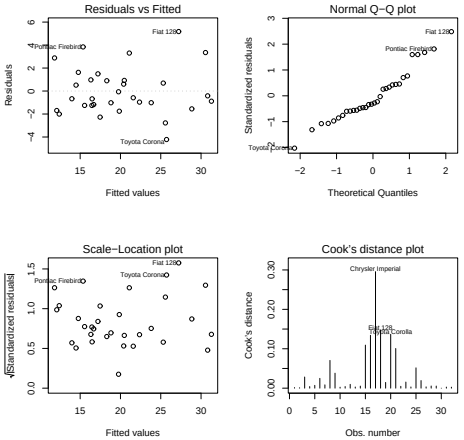

Quick Up/Down Search

Double-check that there are no "extra" terms that help:

```
> add1(l3, ~ .+disp+hp+drat+I(wt^3)+I(qsec^2), test="F")
[ . . . ]
mpg ~ wt + I(wt^2) + qsec
          Df Sum of Sq     RSS     AIC F value  Pr(F)
<none>                  133.315  53.663
disp       1     0.216 133.098  55.611  0.0439 0.8356
hp         1     2.722 130.593  55.003  0.5628 0.4596
drat       1     1.374 131.941  55.332  0.2811 0.6003
I(wt^3)    1     1.805 131.510  55.227  0.3706 0.5477
I(qsec^2)  1     0.913 132.402  55.443  0.1862 0.6695
```

Double-check that there are no "useless" terms still in the model:

```
> drop1(l3, test="F")
[ . . . ]
mpg ~ wt + I(wt^2) + qsec
        Df Sum of Sq     RSS     AIC F value     Pr(F)
<none>              133.315  53.663
wt       1   169.437 302.752  77.910  35.587 2.008e-06 ***
I(wt^2)  1    62.149 195.464  63.908  13.053 0.0011739 **
qsec     1    70.431 203.745  65.236  14.793 0.0006339 ***
```


"Final" Model

```
> summary(l3)
lm(formula = mpg ~ wt + I(wt^2) + qsec, data = mtcars)
[ . . . ]
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  32.6418     5.6768   5.750 3.59e-06 ***
wt          -12.4331     2.0842  -5.965 2.01e-06 ***
I(wt^2)       1.0730     0.2970   3.613 0.001174 **
qsec          0.8599     0.2236   3.846 0.000634 ***
[ . . . ]
Residual standard error: 2.182 on 28 degrees of freedom
Multiple R-Squared: 0.8816,Adjusted R-squared: 0.8689
F-statistic:  69.5 on 3 and 28 DF,  p-value: 4.345e-13
> plot(l3)
```