

R version 3.0.1 (2013-05-16) -- "Good Sport"

```
> ed = read.table(file="http://www-stat.stanford.edu/~rag/stat209/coleman.dat", header = T)
> # read in the Coleman data from the posted data file
```

```
> cor(ed)
      ssal      whcol      ses      tverb      momed      vach
ssal  1.0000000  0.18113980  0.2296278  0.50266385  0.1967731  0.1922916
whcol  0.1811398  1.00000000  0.8271829  0.05105812  0.9271008  0.7534008
ses    0.2296278  0.82718291  1.0000000  0.18332924  0.8190633  0.9271611
tverb  0.5026638  0.05105812  0.1833292  1.00000000  0.1238087  0.3336495
momed  0.1967731  0.92710081  0.8190633  0.12380866  1.0000000  0.7329859
vach   0.1922916  0.75340081  0.9271611  0.33364951  0.7329859  1.0000000
> # in reg ex we made .733 big and negative
```

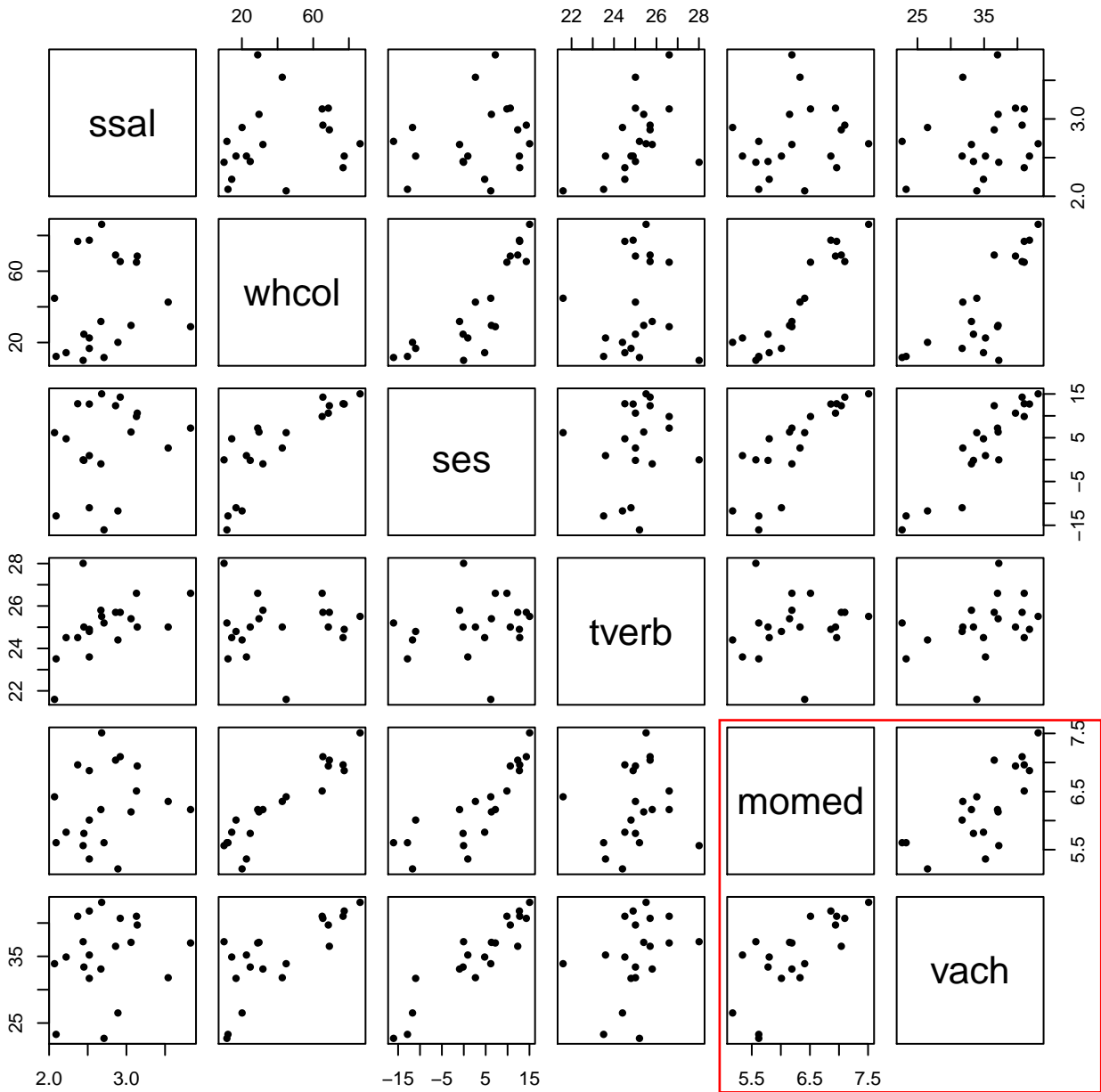
```
> pairs(ed, pch = 20)
> pairs(ed, panel = panel.smooth, pch = 20)
> #overlay a smoother on each pairwise scatterplot
```

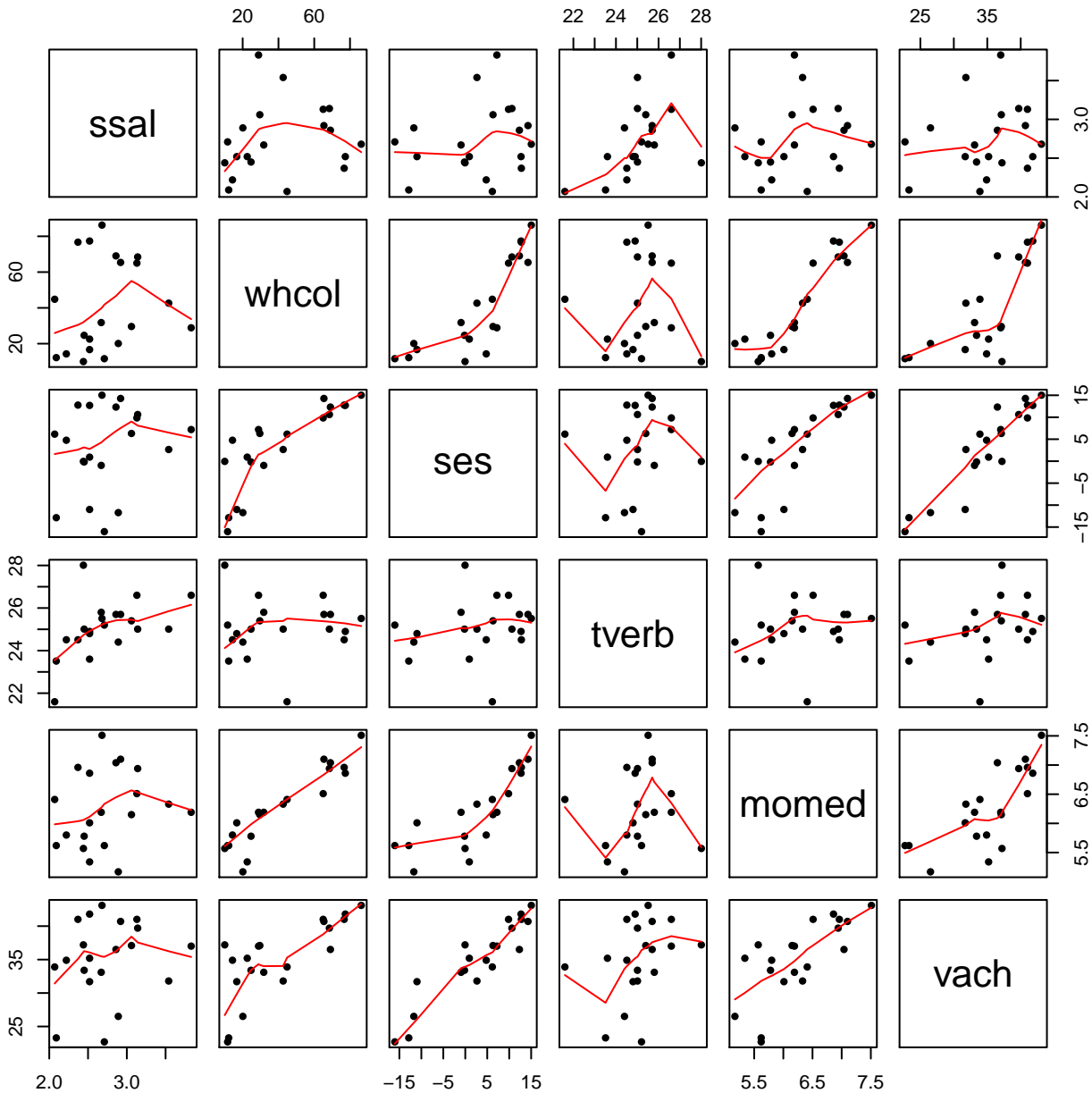
```
> ed
  ssal whcol      ses tverb momed  vach
1  3.83 28.87   7.20 26.60  6.19 37.01
2  2.89 20.10 -11.71 24.40  5.17 26.51
3  2.86 69.05  12.32 25.70  7.04 36.51
4  2.92 65.40  14.28 25.70  7.10 40.70
5  3.06 29.59   6.31 25.40  6.15 37.10
6  2.07 44.82   6.16 21.60  6.41 33.90
7  2.52 77.37  12.70 24.90  6.86 41.80
8  2.45 24.67  -0.17 25.01  5.78 33.40
9  3.13 65.01   9.85 26.60  6.51 41.01
10 2.44  9.99  -0.05 28.01  5.57 37.20
11 2.09 12.20 -12.86 23.51  5.62 23.30
12 2.52 22.55   0.92 23.60  5.34 35.20
13 2.22 14.30   4.77 24.51  5.80 34.90
14 2.67 31.79  -0.96 25.80  6.19 33.10
15 2.71 11.60 -16.04 25.20  5.62 22.70
16 3.14 68.47  10.62 25.01  6.94 39.70
17 3.54 42.64   2.66 25.01  6.33 31.80
18 2.52 16.70 -10.99 24.80  6.01 31.70
19 2.68 86.27  15.03 25.51  7.51 43.10
20 2.37 76.73  12.77 24.51  6.96 41.01
```

```
> colMeans(ed)
  ssal  whcol      ses  tverb  momed  vach
2.7315 40.9060  3.1405 25.0690  6.2550 35.0825
```

```
> ?cov
```

```
> var(ed)
      ssal      whcol      ses      tverb      momed      vach
ssal  0.20620289  2.130280  1.003668  0.2998858  0.05846579  0.5079382
whcol  2.13028000 670.734846 206.202997 1.7372853 15.71053684 113.5023000
ses    1.00366763 206.202997  92.647984 2.3183584  5.15850789  51.9130250
tverb  0.29988579  1.737285  2.318358  1.7260832  0.10643158  2.5499079
momed  0.05846579 15.710537  5.158508 0.1064316  0.42813158  2.7898921
vach   0.50793816 113.502300 51.913025 2.5499079  2.78989211 33.8381250
```





```
# As the Coleman data is only 20 schools (with 5 predictors),
# for expository purposes I created a larger artificial data set,
# with 320 rows (16x) , sample having the same means and
# covariances as the n=20 sample.
```

```
> library(MASS)
> ?mvrnorm

> eddat320 = mvrnorm(n = 320, colMeans(ed), var(ed), empirical = TRUE)
> dim(eddat320)
[1] 320 6

> cor(eddat320)
      ssal      whcol      ses      tverb      momed      vach
ssal  1.0000000  0.1811398  0.2296278  0.50266385  0.1967731  0.1922916
whcol  0.1811398  1.0000000  0.8271829  0.05105812  0.9271008  0.7534008
ses    0.2296278  0.8271829  1.0000000  0.18332924  0.8190633  0.9271611
tverb  0.5026638  0.05105812  0.1833292  1.00000000  0.1238087  0.3336495
momed  0.1967731  0.92710081  0.8190633  0.12380866  1.0000000  0.7329859
vach   0.1922916  0.75340081  0.9271611  0.33364951  0.7329859  1.0000000

> colMeans(ed)
      ssal      whcol      ses      tverb      momed      vach
 2.7315 40.9060  3.1405 25.0690  6.2550 35.0825
> colMeans(eddat320)
      ssal      whcol      ses      tverb      momed      vach
 2.7315 40.9060  3.1405 25.0690  6.2550 35.0825

> head(eddat320)
      ssal      whcol      ses      tverb      momed      vach
[1,] 2.427460 43.949136 -1.708884 23.23557 6.470510 31.20838
[2,] 1.830226 -2.438887 -5.066753 23.22068 5.281980 30.21031
[3,] 2.198009 50.203696 16.143299 26.99223 6.444350 42.57380
[4,] 3.775413 58.727910  5.364175 26.05464 6.809575 35.33795
[5,] 2.967065  8.568443 -14.931575 25.38875 5.671992 21.00682
[6,] 3.086095 80.980756 10.666565 26.78154 6.723911 40.75565

> eddat320F = as.data.frame(eddat320) # make a data frame to run regression
> edreg320 = lm(vach ~ ssal+ whcol+ ses+ tverb+ momed, data = eddat320F)

> summary(edreg320)
Call:
lm(formula = vach ~ ssal + whcol + ses + tverb + momed, data = eddat320F)

Residuals:
    Min       1Q   Median       3Q      Max
-4.8445 -1.2318 -0.1038  1.2294  6.3920

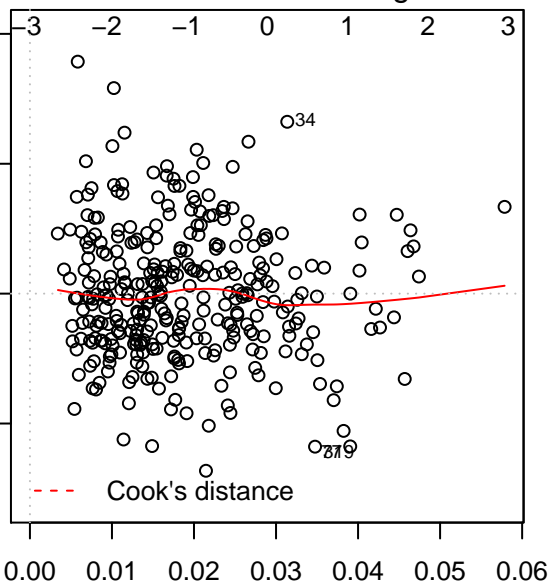
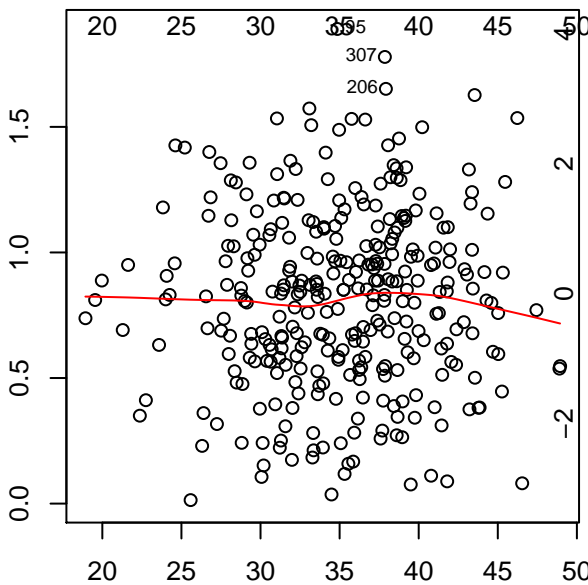
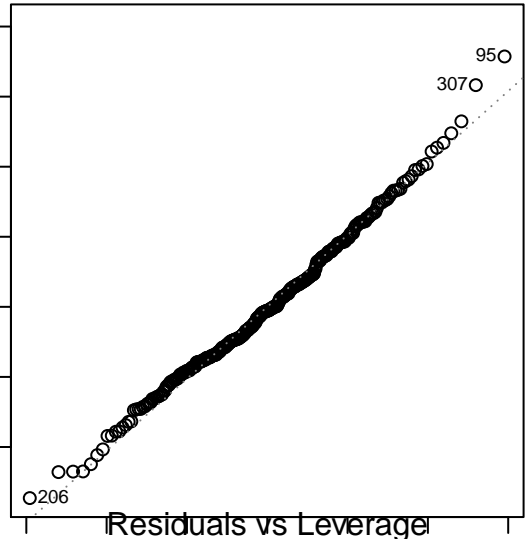
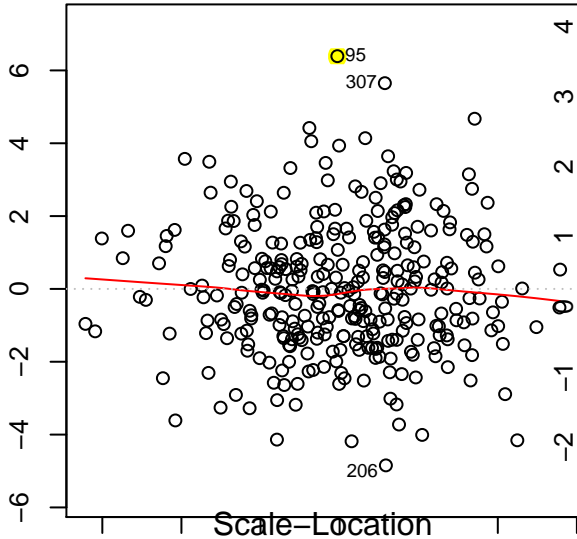
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 19.94857    2.87759   6.932 2.36e-11 ***
ssal        -1.79333    0.26044  -6.886 3.14e-11 ***
whcol         0.04360    0.01125   3.877 0.000129 ***
ses           0.55576    0.01963  28.315 < 2e-16 ***
tverb         1.11017    0.09159  12.121 < 2e-16 ***
momed        -1.81092    0.42809  -4.230 3.07e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.795 on 314 degrees of freedom
Multiple R-squared:  0.9063,    Adjusted R-squared:  0.9048
F-statistic: 607.5 on 5 and 314 DF,  p-value: < 2.2e-16

> ?plot.lm
> par(mfrow = c(2, 2)) # make a 2x2 array, see plot
> plot(edreg320)
```

lm(vach ~ ssa1 + whcol + ses + tverb + momed)

Residuals vs Fitted



?plot.lm