

Statistics 5041

4. Graphical Methods

Gary W. Oehlert
School of Statistics
313B Ford Hall
612-625-1557
gary@stat.umn.edu

Data set of 50 colleges and universities. Variables are average SAT, percent accepted, dollars spent per student, percent of students from top 10 percent of high school class, percent of faculty with Ph.D. degrees, and percent of students who graduate.

```
Cmd> readdata("schools",sat,accpt,cost,\  
top10,phd,grad)
```

```
Cmd> schools <- hconcat(sat,accpt,cost,\  
top10,phd,grad)
```

Read in by columns and then paste together.

```
Cmd> schools <- matrix(vcread("schools"),6)'
```

```
Cmd> schools <- matrix(schools,50,\  
labels:structure("(" ,vector("sat","accpt",\  
"$ /stud","top10","phd","grad")))
```

Or read into a matrix at once. Here I've also added some nice labels

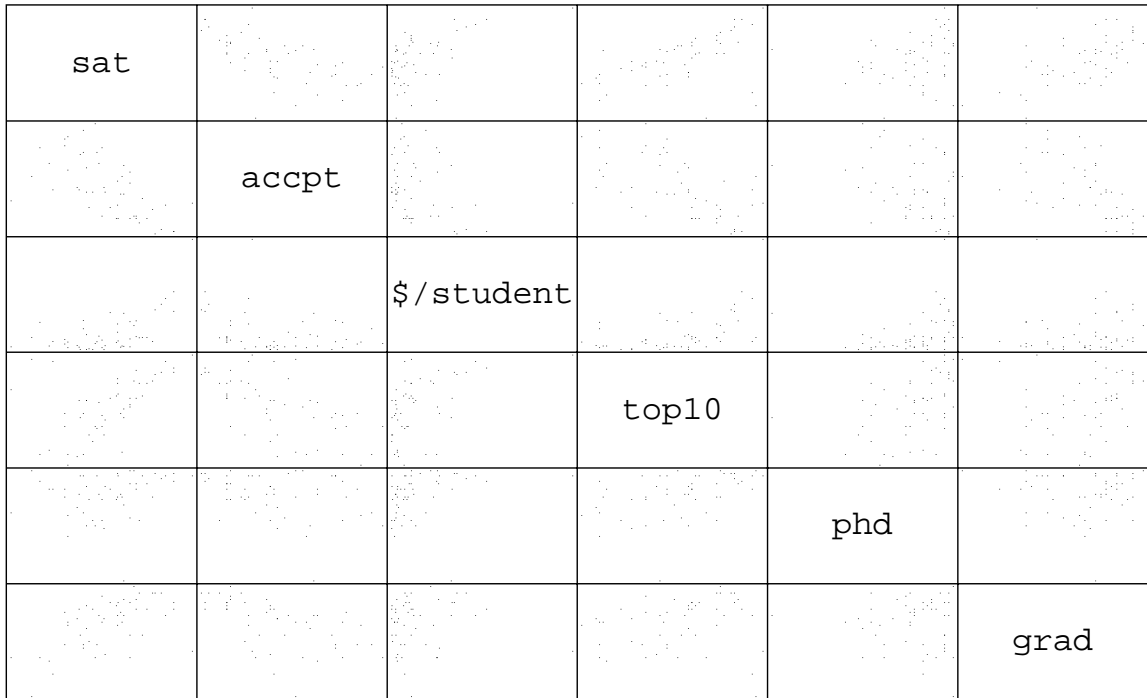
	sat	accpt	\$/stud	top10	phd	grad
(1)	1315	22	26636	85	81	93
(2)	1310	24	27487	78	93	88
(3)	1336	28	23772	86	90	93
(4)	1300	24	25703	78	95	90
(5)	1250	49	27879	76	91	86
(6)	1320	33	26668	79	98	80
(7)	1290	35	19948	73	87	91
(8)	1255	25	24718	65	89	92
(9)	1195	57	25271	65	90	87
(10)	1230	36	17721	77	94	89
(11)	1287	43	20179	53	90	84
(12)	1300	40	19504	75	82	80
(13)	1260	36	20377	68	94	74
(14)	1247	54	23591	64	98	77
(15)	1234	29	17998	61	89	78
(16)	1244	67	22301	65	79	73
(17)	1200	61	23358	47	83	83
(18)	1200	46	18872	52	75	84
(19)	1215	38	20722	51	86	85
(20)	1240	36	17554	58	81	88
(21)	1285	35	19418	71	91	87

(22)	1258	38	17520	61	78	85
(23)	1255	56	18847	70	81	84
(24)	1170	49	20192	54	93	72
(25)	1220	53	17653	69	98	80
(26)	1370	18	46918	90	99	90
(27)	1370	18	61921	92	96	88
(28)	1350	19	52468	90	97	93
(29)	1340	17	48123	89	99	93
(30)	1400	31	102262	98	98	75
(31)	1357	30	56766	95	98	86
(32)	1310	25	39504	91	95	91
(33)	1306	25	35804	86	100	95
(34)	1280	30	37137	85	90	83
(35)	1268	29	45879	78	93	90
(36)	1300	45	38937	74	100	73
(37)	1281	24	24201	80	98	90
(38)	1280	41	30882	87	99	86
(39)	1176	37	23665	95	93	68
(40)	1290	48	45460	69	58	86
(41)	1327	24	26730	85	95	88
(42)	1142	43	26859	96	100	61
(43)	1218	37	19365	77	91	88
(44)	1278	24	23115	79	89	89
(45)	1109	32	19684	82	84	73
(46)	1195	60	21853	71	93	77
(47)	1225	64	33607	52	84	77
(48)	1230	47	28851	77	79	82
(49)	1225	54	39883	71	98	76
(50)	1155	56	38597	52	96	73

A *scatterplot matrix* is an array of bivariate scatter plots, with one plot for each (possibly ordered) pair of variables in the data set.

```
Cmd> plotmatrix(schools)
```

schools



schools



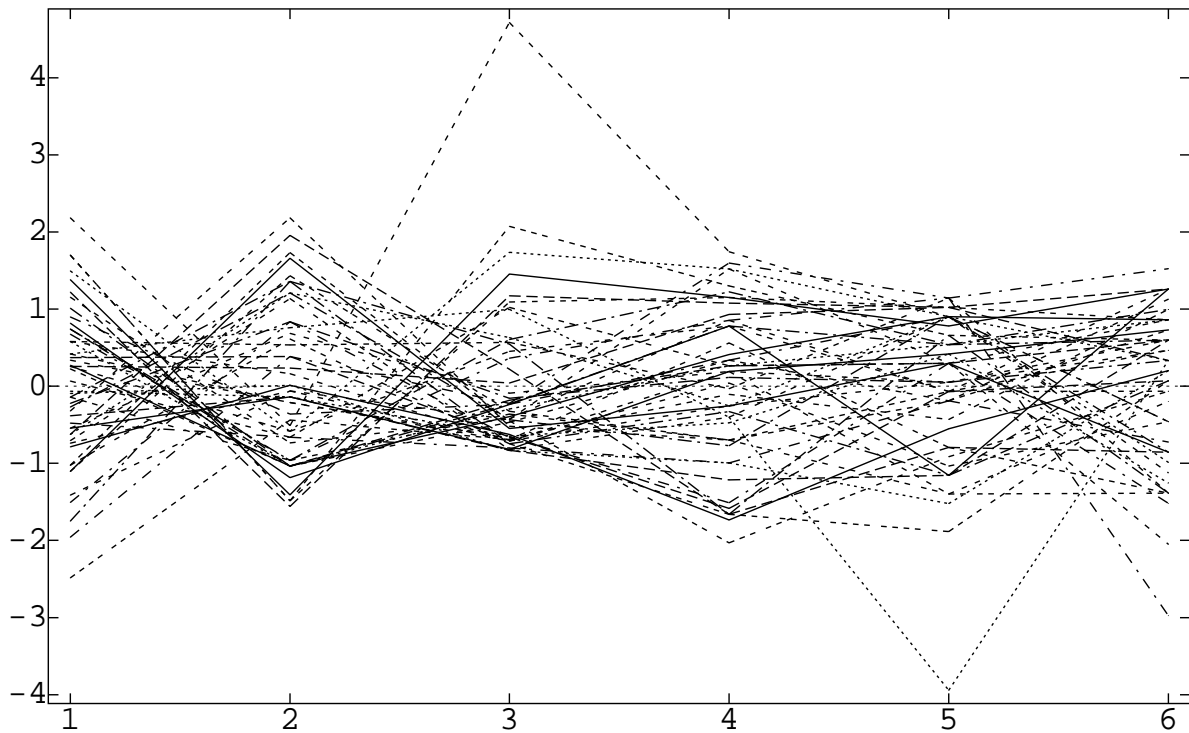
Profile plots are simply “connect the dots” plots for each unit. Look for patterns and relationships.

Often helps to put all the data on the same scale, perhaps by standardization or transformation.

```
Cmd> ys <- schools - describe(schools, mean:T)'
```

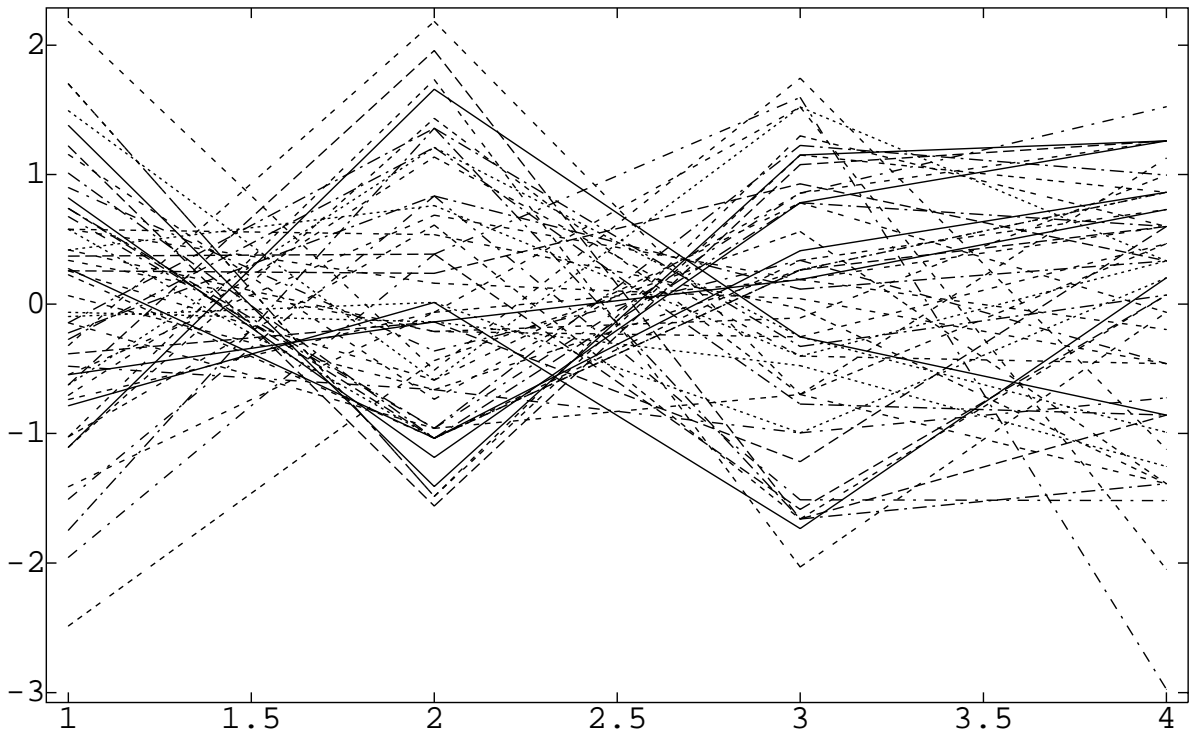
```
Cmd> ys <- schools/sqrt(describe(schools, var:T))'
```

```
Cmd> plot(run(6), ys', lines:T, symbols:"\7")
```



Variables 3 and 5 seem a bit funny, so let's try again without them.

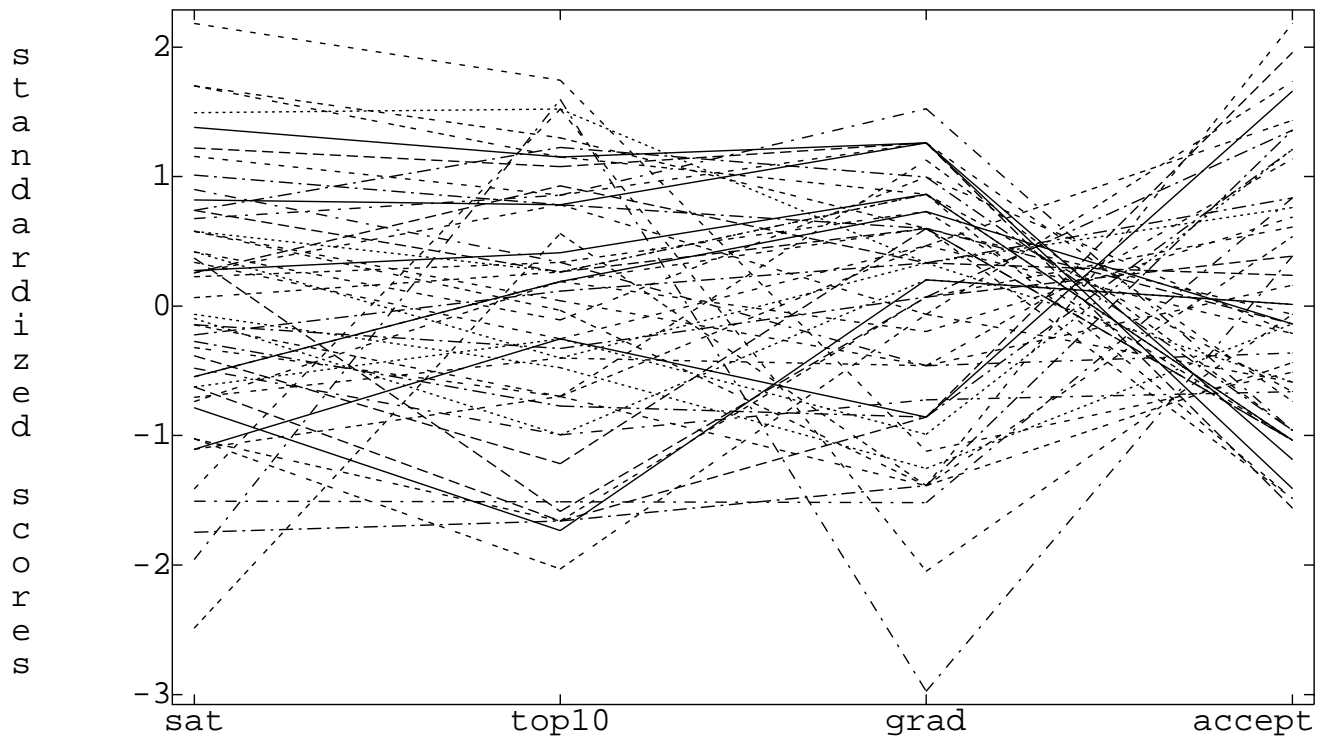
```
Cmd> plot(run(4), ys[, -vector(3,5)]', \  
lines:T, symbols:"\7", xaxis:F)
```



Variable 2 seems different from the others, so let's put it last and add some nice labeling.

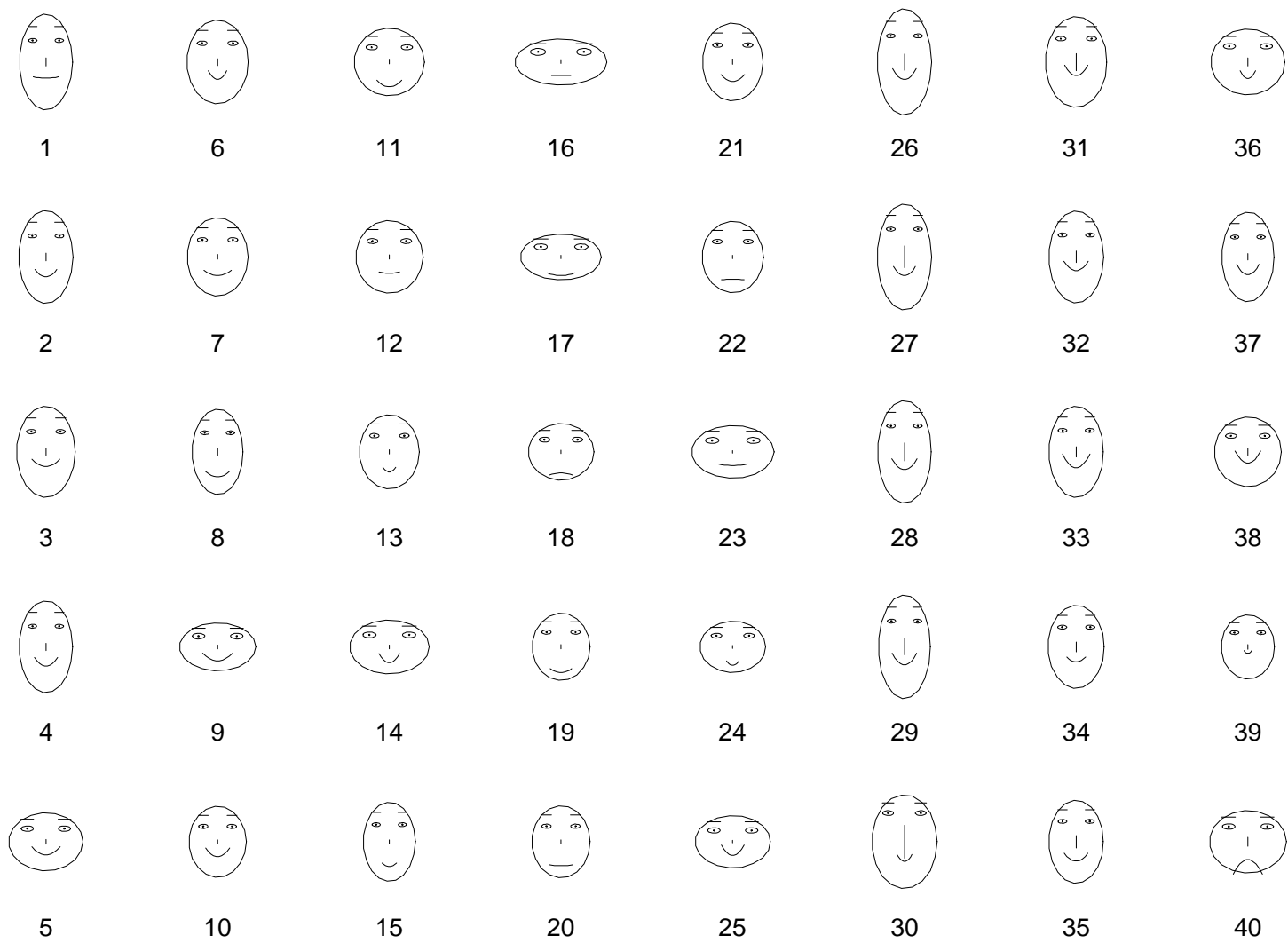
```
Cmd> plot(run(4),ys[,vector(1,4,6,2)]',\  
lines:T,symbols:"\7",xaxis:F,\  
xticks:vector(1,2,3,4),\  
xticklabs:vector("sat","top10","grad","accept"),\  
title:"School profiles",ylab:"standardized scores")
```

School profiles



Chernoff faces let us look at multivariate data by associating data levels with aspects of faces. Variables determine 1-area of face; 2-shape of face; 3-length of nose; 4-location of mouth; 5-curve of smile; 6-width of mouth; 7, 8, 9, 10, 11-location, separation, angle, shape and width of eyes; 12-location of pupil; 13, 14, 15-location, angle and width of eyebrow.

Not in MacAnova.



Order of variables affects the way faces look. Here we reverse the order of the six variables.



1



7



13



19



25



31



2



8



14



20



26



32



3



9



15



21



27



33



4



10



16



22



28



34



5



11



17



23



29



35



6



12



18



24



30



36