STAT 400 Data Analysis Oct 3, 2005

Note:
For November – you will be required to give a 7 minute presentation of a topic of your choice. Define the item – say why it is useful – give an example. Any topic from this course is possible but please check with me first – I want each student to do a different topic. You can use data from the course or your own data (in the latter case, important that you describe the source of data, and that you do not use an example from another textbook or from another course).
There will be a few (3-4) more assignments like the ones you have done already)
There will be a final exam.

Back to Cleveland: Ch 4 - Trivariate Data Here is a TOC of the chapter …

Important Chapter – start reading through – questions in the margin …

Recall Slicing: way of looking at dependence of one Y on an X.

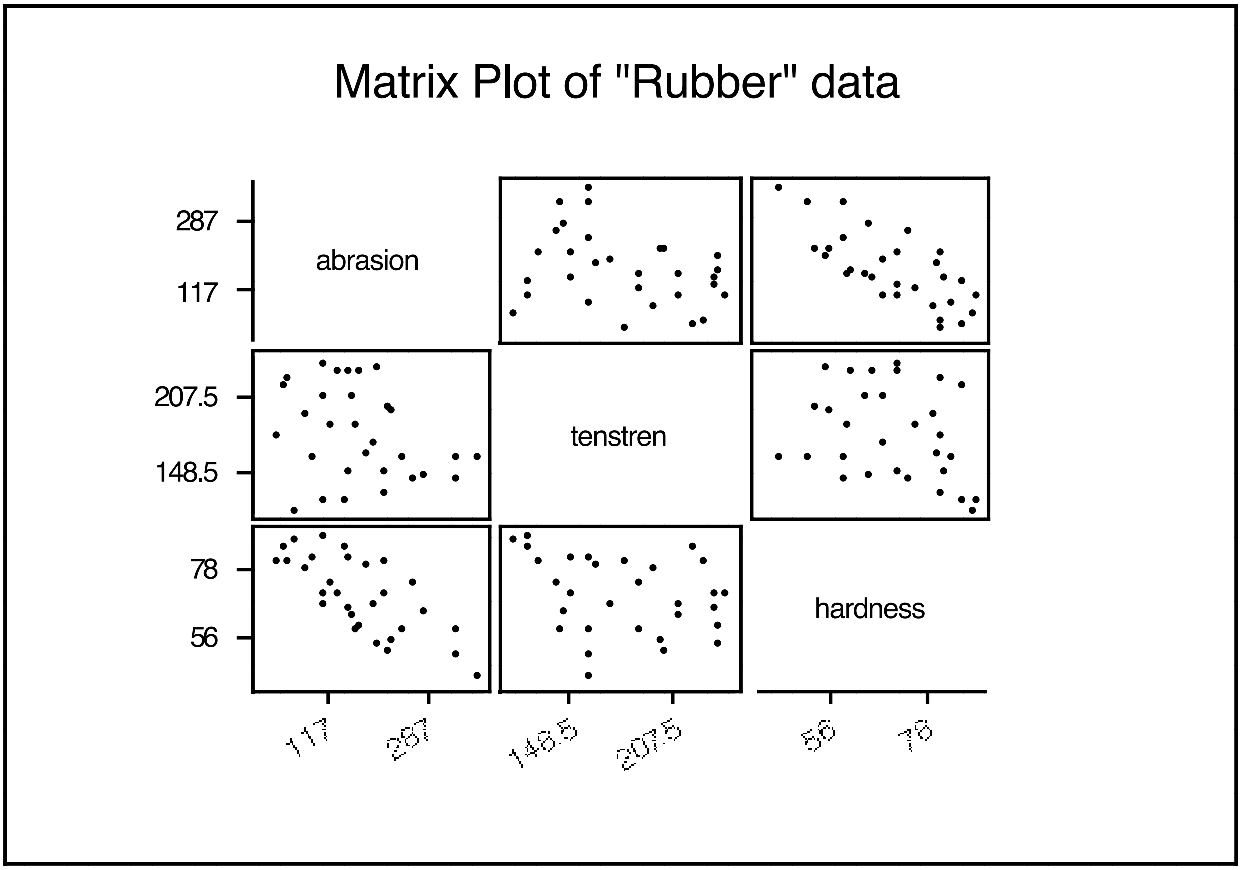
Suppose now another variable Z
Would like to look at the Y-X relationship for various values of Z. Example: See "rubber" data on p 180.
Matrix plot is useful for some purposes but interaction is hard to see.

What is interaction? **Interaction** is a quality of predictor variables in the context of prediction of a response variable. The predictor variables are said to **interact** in their effect on the response if the change in the response due to a change in one predictor depends on the value of the other predictor. In this situation we cannot talk of the effect of changes in factor A on response R without specifying the level of Factor B, because the change in R due to a change in A is different for different levels of B.

Conditioning Plots – Coplots – Allow us to visualize interactions in a very flexible way.

An example for the rubber data is shown on p 185.
Ordering of panels
Use of Loess
Slicing of Z (Hardness in Fig 4.3)
Visualization of Interaction
Verbalization of Interaction
Rubber Data: (see p 180 for matrix plot of this data): 30 rubber specimens
Random sample? Yes & No.

Row	Hardness	Tenstrth	Abrasion
1	45	162	372
2	51	161	341
3	53	203	221
4	55	233	206
5	56	200	228
6	59	161	249
7	59	146	340
8	60	189	166
9	61	232	175
10	64	210	164
11	65	148	283
12	66	231	154
13	68	210	113
14	68	173	196
15	71	231	136
16	71	237	112
17	71	151	219
18	74	144	267
19	75	188	128
20	79	196	82
21	80	165	186
22	81	224	55
23	81	180	32
24	81	134	215
25	82	151	155
26	83	161	97
27	86	219	45
28	86	127	148
29	88	119	64
30	89	128	114



p 180 Note the pairs of graphs
Can you see any dependence of Abrasion Loss on Tensile Strength and Hardness?

P 181 Note role of variables – 1 dependent, 2 independent
Coordinate system for panels (i,j) i is number of panel left to right
j is number of panel bottom to top
(just as in usual coordinate system)

p 182 concept of slice of one variable (hardness) to examine relationship between two other variables (abrasion loss and tensile strength). Conditioning plot = "coplot"

p 183 Does the relationship between abrasion loss and tensile strength depend on hardness? i.e. Do Tensile Strength and Hardness interact in determining Abrasion Loss?

pp 184-5 Putting several slices together. Panels show response to increasing hardness, from lower left to upper right. (1,1), (2,1), (3,1), (1,2), (2,2), (3,2) is the order of the plots corresponding to the increasing slices.

Note robust fitting of loess curves to each slice. Reveals possible interaction. What if highest hardness range were missing?

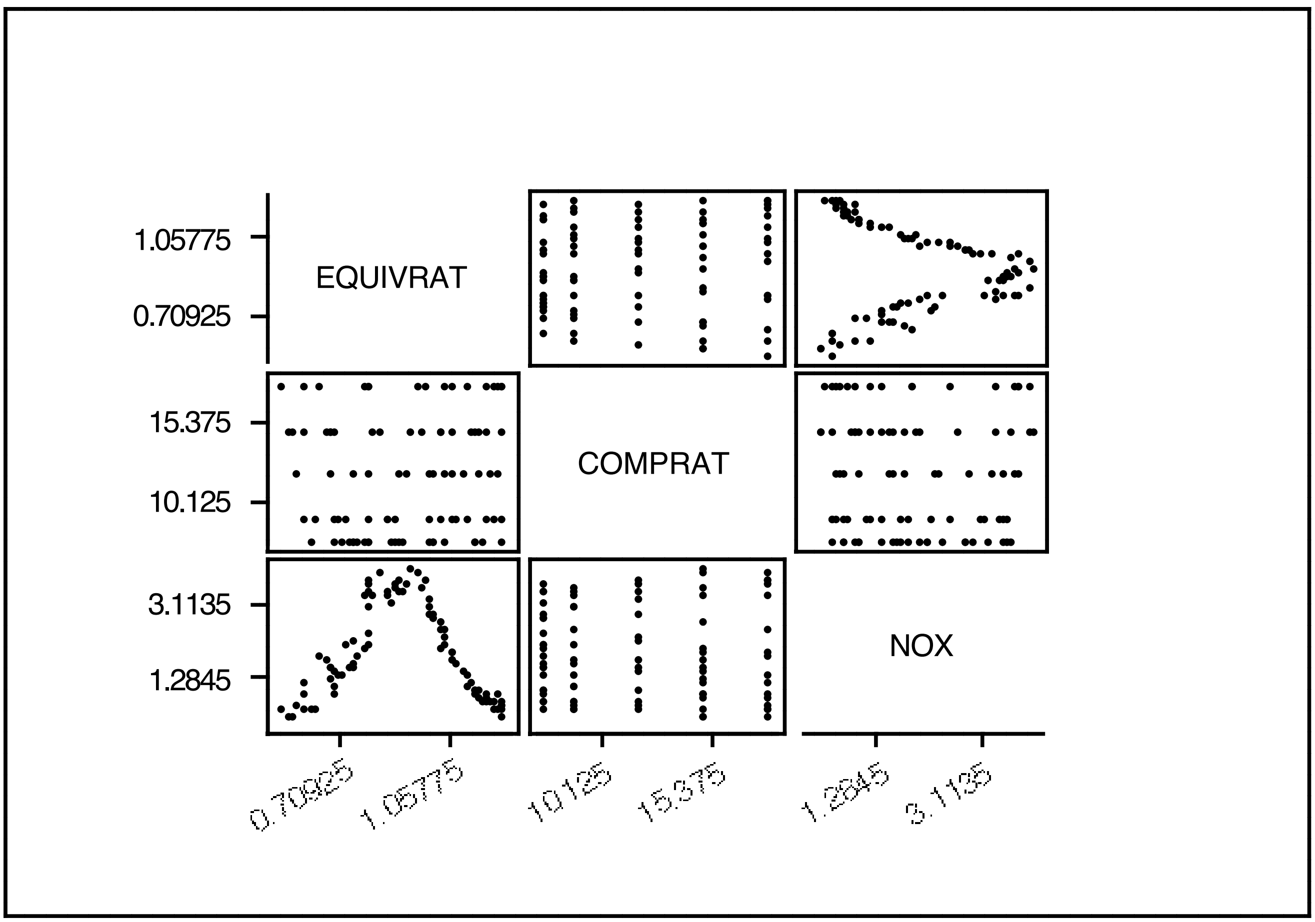
pp 186-7 This time condition on Tensile Strength. Slight interaction less obvious.

p188 Ethanol Data:

Row	NOX	CR	ER
1	3.741	12.0	0.907
2	2.295	12.0	0.761
3	1.498	12.0	1.108
4	2.881	12.0	1.016
5	0.760	12.0	1.189
6	3.120	9.0	1.001
7	0.638	9.0	1.231
8	1.170	9.0	1.123
9	2.358	12.0	1.042
10	0.606	12.0	1.215
11	3.669	12.0	0.930
12	1.000	12.0	1.152
13	0.981	15.0	1.138
14	1.192	18.0	0.601
15	0.926	7.5	0.696
16	1.590	12.0	0.686
17	1.806	12.0	1.072
18	1.962	15.0	1.074
19	4.028	15.0	0.934
20	3.148	9.0	0.808
21	1.836	9.0	1.071
22	2.845	7.5	1.009
23	1.013	7.5	1.142
24	0.414	18.0	1.229
25	0.812	18.0	1.175
26	0.374	15.0	0.568
27	3.623	15.0	0.977
28	1.869	7.5	0.767
29	2.836	7.5	1.006
30	3.567	9.0	0.893
31	0.866	15.0	1.152
32	1.369	15.0	0.693
33	0.542	15.0	1.232

34	2.739	15.0	1.036
35	1.200	15.0	1.125
36	1.719	9.0	1.081
37	3.423	9.0	0.868
38	1.634	7.5	0.762
39	1.021	7.5	1.144
40	2.157	7.5	1.045
41	3.361	18.0	0.797
42	1.390	18.0	1.115
43	1.947	18.0	1.070
44	0.962	18.0	1.219
45	0.571	9.0	0.637
46	2.219	9.0	0.733
47	1.419	9.0	0.715
48	3.519	9.0	0.872
49	1.732	7.5	0.765
50	3.206	7.5	0.878
51	2.471	7.5	0.811
52	1.777	15.0	0.676
53	2.571	18.0	1.045
54	3.952	18.0	0.968
55	3.931	15.0	0.846
56	1.587	15.0	0.684
57	1.397	7.5	0.729
58	3.536	7.5	0.911
59	2.202	7.5	0.808
60	0.756	7.5	1.168
61	1.620	7.5	0.749
62	3.656	7.5	0.892
63	2.964	7.5	1.002
64	3.760	18.0	0.812
65	0.672	18.0	1.230
66	3.677	18.0	0.804
67	3.517	12.0	0.813
68	3.290	12.0	1.002
69	1.139	9.0	0.696
70	0.727	9.0	1.199
71	2.581	9.0	1.030
72	0.923	15.0	0.602
73	1.527	15.0	0.694

74	3.388	15.0	0.816
75	2.085	15.0	1.037
76	0.966	15.0	1.181
77	3.488	7.5	0.899
78	0.754	7.5	1.227
79	0.797	9.0	1.180
80	2.064	7.5	0.795
81	3.732	18.0	0.990
82	0.586	18.0	1.201
83	0.561	7.5	0.629
84	0.563	9.0	0.608
85	0.678	12.0	0.584
86	0.370	15.0	0.562
87	0.530	18.0	0.535
88	1.900	18.0	0.655



Understand the graph ….

Note that labeling could be improved.

NOX is dependent, ER and CR are independent.

EQ – richness of mixture
CR – max volume of cylinder/min volume of cylinder

Fig 4.6 : 9 slices of ER, f=.25 (overlap), NOX vs CR examined. Big interaction here.
Fig 4.7: 5 slices (not really slices in usual sense) NOX vs ER examined. Max increases.

Continuation of Ch 4 next time ….