Johns Hopkins Engineering

625.464 Computational Statistics

Viewing Multivariate Data

Module 13 Lecture 13D

JOHNS HOPKINS
WHITING SCHOOL
of ENGINEERING

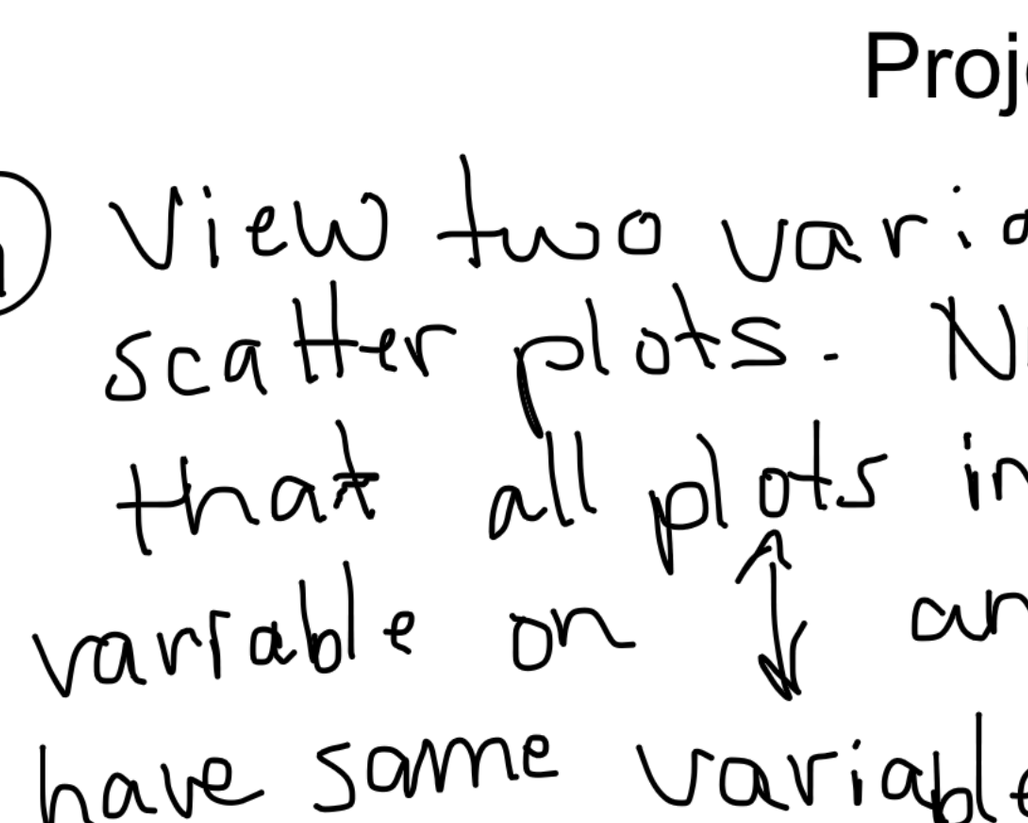
Viewing Multivariate Data

(1) use multiple 2-d views / proj into cartesian plane $O(d^2)$

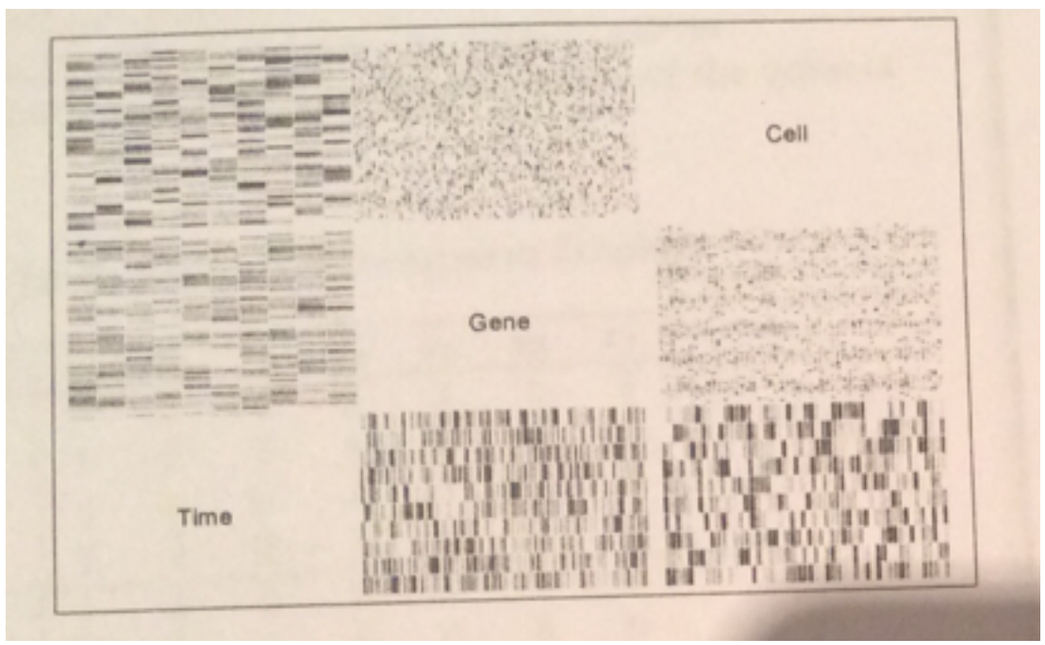
(2) use other types of graphical obj w/ char assoc. w/ each of the variables small dataset

Projections

(i) View two variables at a time using scatter plots. Nice to arrange plots so that all plots in same row have same variable on ↕ and all plots in same col have same variable on ↔ SPLOM



(ii) IMPLOM



Non Cartesian Displays

- Each obs is rep as a more complex obj. than just a point w/ the values of the ind variables x_i (in $X = (x_1, \ldots, x_n)$) rep. by some aspect of the object.
- only useful on small datasets
- sometimes hard to see relationships at first

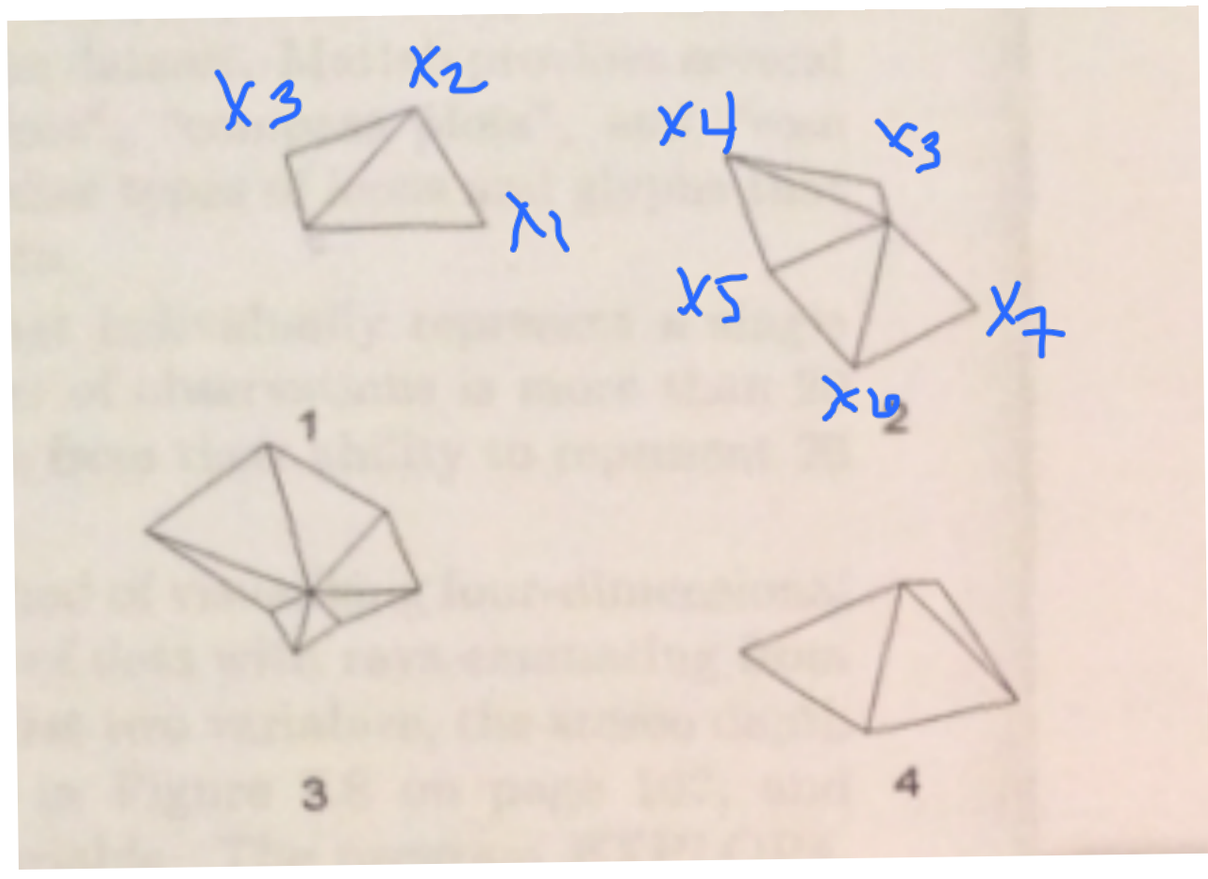
Observation	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	6	5	4	3	2	1	2
2	1	2	3	4	5	6	5
3	4	4	6	4	3	3	3
4	2	2	2	3	6	6	6
Minimum	1	2	2	3	2	1	2
Maximum	6	5	6	4	6	6	6

Glyphs and Icons

① Star Diagrams

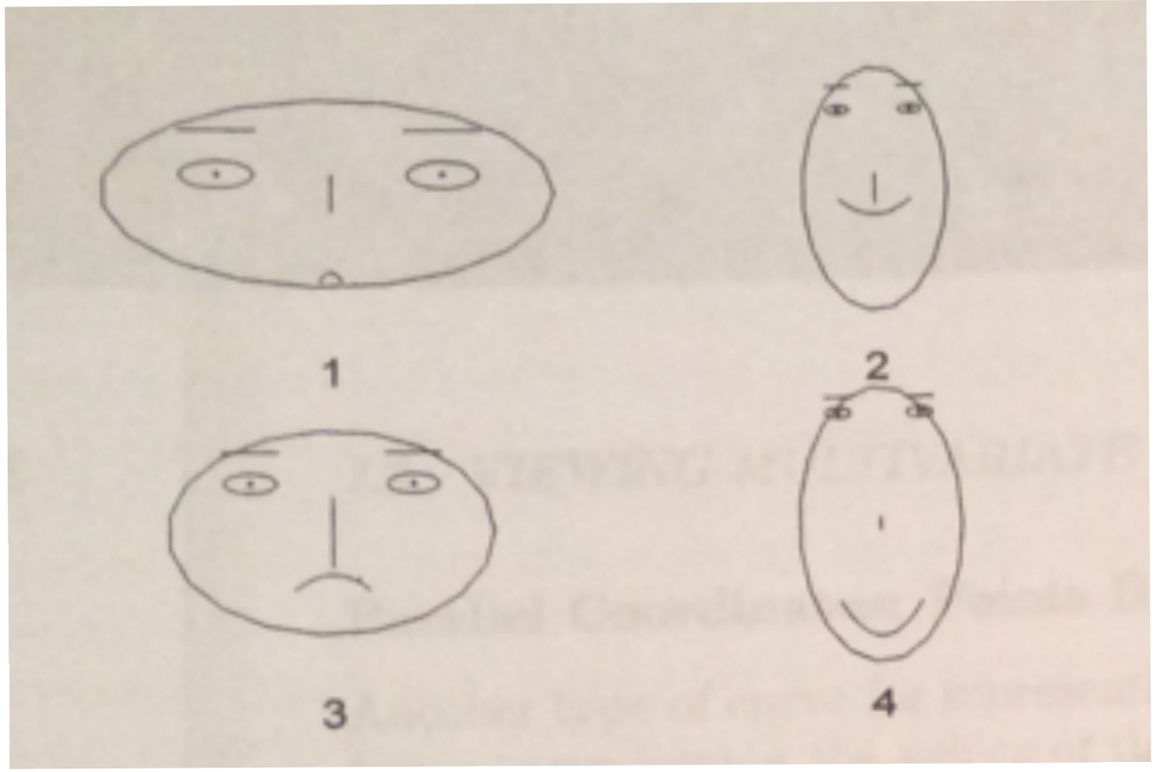
- To represent an obs of d-dim data you use rays pointing out from a central point in d equally spaced directions

Observation	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	6	5	4	3	2	1	2
2	1	2	3	4	5	6	5
3	4	4	6	4	3	3	3
4	2	2	2	3	6	6	6
Minimum	1	2	2	3	2	1	2
Maximum	6	5	6	4	6	6	6



Chernoff Faces

Observation	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	6	5	4	3	2	1	2
2	1	2	3	4	5	6	5
3	4	4	6	4	3	3	3
4	2	2	2	3	6	6	6
Minimum	1	2	2	3	2	1	2
Maximum	6	5	6	4	6	6	6



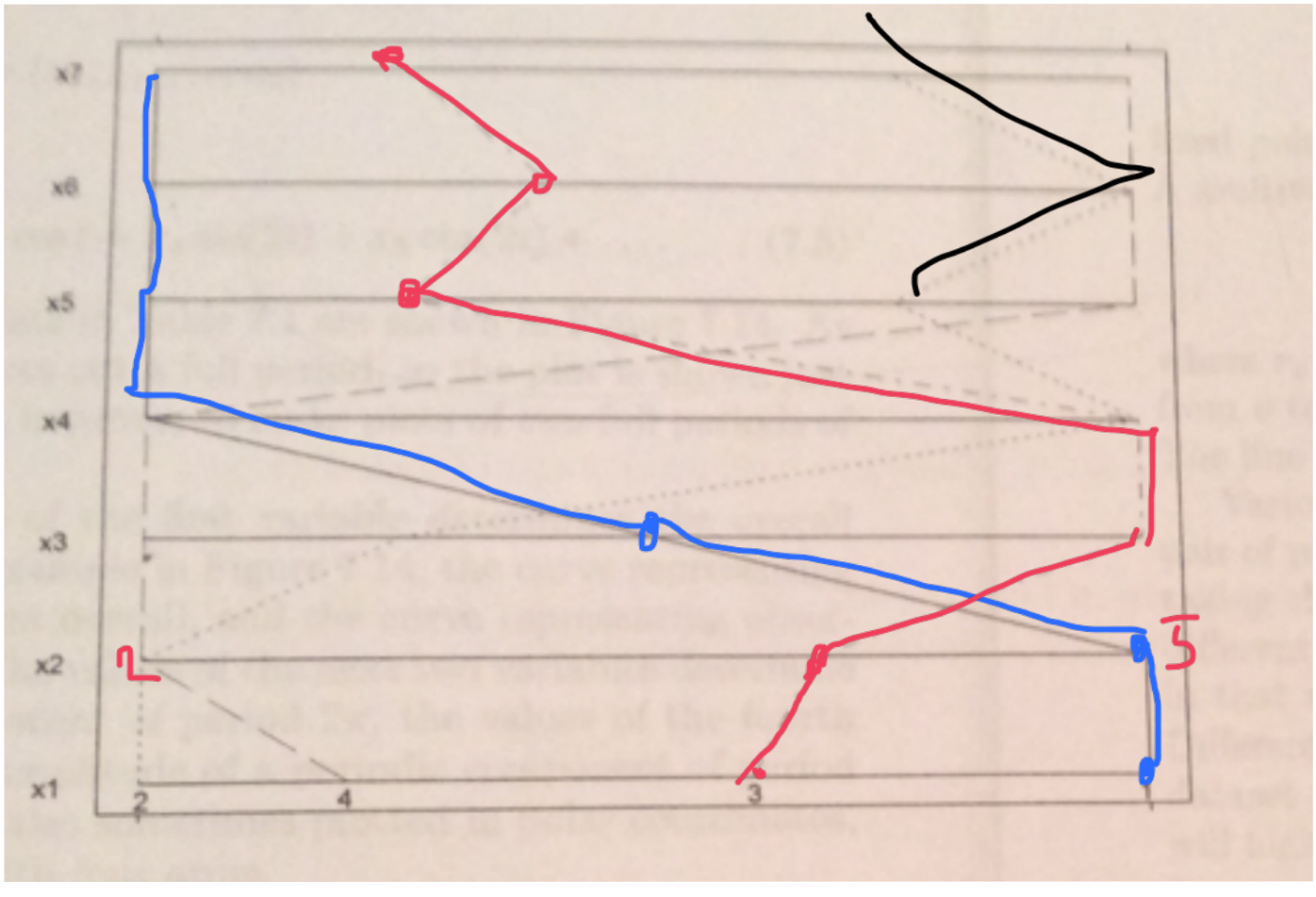
x_1 - area of face
x_2 - shape of face
x_3 - length of nose
x_4 - location of mouth
x_5 - curve of smile
x_6 - width of mouth
x_7 - location of eyes

# of obs ≤ 20 or 30

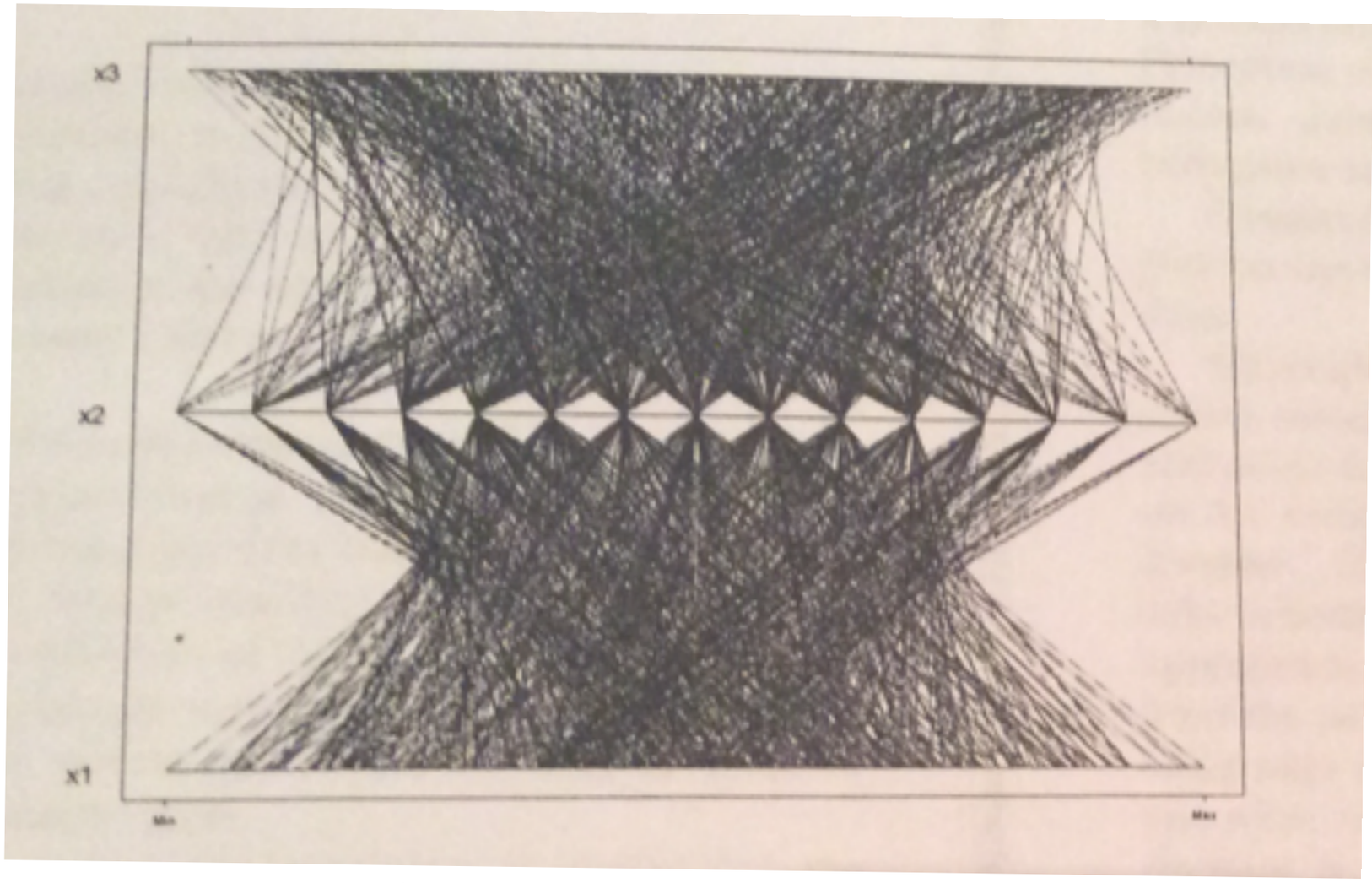
Parallel Coordinates

- a piecewise linear curve joining the values of the variables on a set of parallel axes, where each axis rep the values of a given variable.

Observation	x_1	x_2	x_3	x_4	x_5	x_6	x_7
1	6	5	4	3	2	1	2
2	1	2	3	4	5	6	5
3	4	4	6	4	3	3	3
4	2	2	2	3	6	6	6
Minimum	1	2	2	3	2	1	2
Maximum	6	5	6	4	6	6	6



Displaying Large Data Sets



- look at subsets
- jittering
- grayscale
- color