
Information Visualization
CMSC 838B – Spring 2003

Multidimensional Data Visualization

Benjamin B. Bederson
University of Maryland
www.cs.umd.edu/~bederson

This presentation adapted from [John Stasko](#)

Data Sets

- Data comes in many different forms
 - Typically, not in the way you want it

 - How is stored?
-

Example

- Cars
 - make
 - model
 - year
 - miles per gallon
 - cost
 - number of cylinders
 - weights
 - ...
-

Data Tables

- Often, we take raw data and transform it into a form that is more workable
- Main idea:
 - Individual items are called *cases*
 - Cases have *variables* (attributes)

Example

Baseball statistics

Player	At Bats	Hits	Home Runs	Runs	Stolen Bases	Walks	Years in Major League	At Bats	Hits	Home Runs	Runs	Stolen Bases	Walks	Years in Major League
1. Babe Ruth	297	66	1	30	29	14	1	293	66	1	30	29	14	1
2. Andy Anderson	293	66	1	30	29	14	1	293	66	1	30	29	14	1
3. Alan Ashby	315	81	7	24	38	39	14	3449	836	7	24	38	39	14
4. Alan Davis	479	130	18	68	72	76	3	1624	467	18	68	72	76	3
5. Andre Dawson	486	141	20	65	79	37	11	5638	1676	20	65	79	37	11
6. Andres Galarraga	321	87	10	39	42	30	2	396	101	10	39	42	30	2
7. Ardeus Griffin	584	169	4	74	51	36	11	4489	1133	4	74	51	36	11
8. Al Newman	185	37	1	23	8	21	2	214	42	1	23	8	21	2
9. August Galaz	298	73	0	24	24	7	2	169	108	0	24	24	7	2
10. Andrew Thomas	323	81	6	26	32	8	2	341	86	6	26	32	8	2
11. Andre Thornton	401	82	17	49	66	65	13	5206	1332	17	49	66	65	13
12. Alan Trammell	524	169	21	107	75	69	10	4011	1300	21	107	75	69	10
13. Alex Trevino	352	53	4	31	26	27	9	1876	467	4	31	26	27	9
14. Andy Van Slyke	418	113	13	48	61	47	4	1912	362	13	48	61	47	4
15. Alan Wiggins	229	60	0	30	11	22	4	1941	510	0	30	11	22	4
16. Bill Almon	196	43	7	29	27	30	13	3231	826	7	29	27	30	13
17. Billy Brann	183	39	3	20	15	11	3	201	42	3	20	15	11	3
18. Buddy Bell	568	158	20	89	75	73	16	6068	2273	20	89	75	73	16
19. Buddy Bianchi	190	46	2	24	8	15	5	479	102	2	24	8	15	5
20. Brock Boesche	407	104	6	57	43	66	12	5233	1476	6	57	43	66	12

Metadata

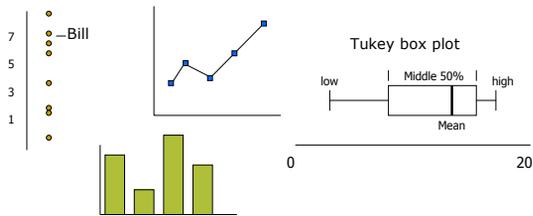
- Descriptive information about the data
 - Might be something as simple as the type of a variable, or could be more complex
 - For times when the table itself just isn't enough
 - Example: if variable1 is "I", then variable3 can only be 3, 7 or 16

How Many Variables?

- Data sets of dimensions 1,2,3 are common
- Number of variables per class
 - 1 - Univariate data
 - 2 - Bivariate data
 - 3 - Trivariate data
 - >3 - Hypervariate/Multivariate data

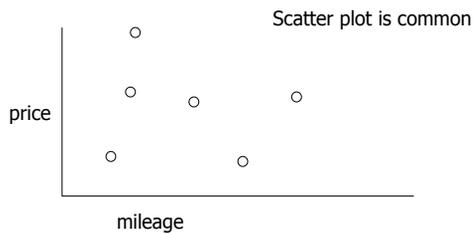
Univariate Data

■ Representations



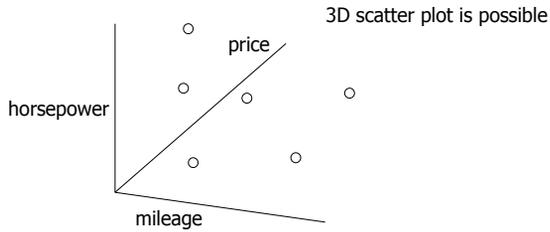
Bivariate Data

■ Representations



Trivariate Data

■ Representations



Multivariate Data

■ Number of well-known visualization techniques exist for data sets of 1-3 dimensions

- line graphs, bar graphs, scatter plots OK
- We see a 3-D world (4-D with time)

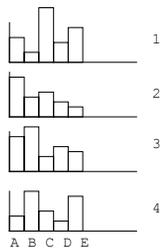
■ What about data sets with more than 3 variables?

- Often the interesting ones

Multiple Views

Give each variable its own display

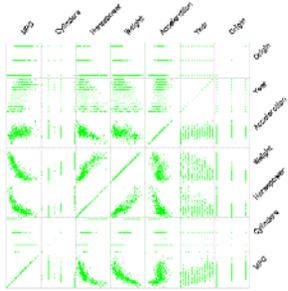
	A	B	C	D	E
1	4	1	8	3	5
2	6	3	4	2	1
3	5	7	2	4	3
4	2	6	3	1	5



Scatterplot Matrix

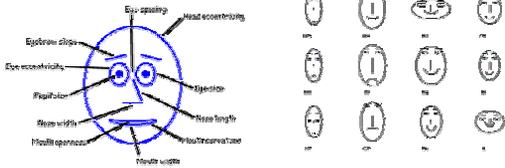
Represent each possible pair of variables in their own 2-D scatterplot

Useful for what?
Misses what?



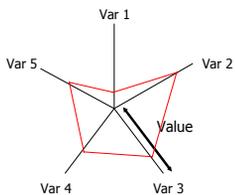
Chernoff Faces

Encode different variables' values in characteristics of human face



Cute applets: <http://www.cs.uchicago.edu/~wiseman/chernoff/>
<http://hesketh.com/schampeo/projects/Faces/chernoff.html>

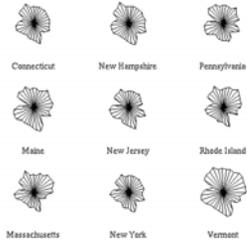
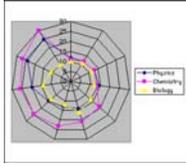
Star Plots



Space out the n variables at equal angles around a circle

Each "spoke" encodes a variable's value

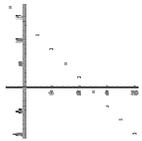
Star Plot examples



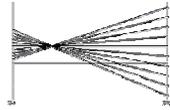
<http://seamonkey.ed.asu.edu/~behrens/asu/reports/compre/compl.html>

Parallel Coordinates

- Encode variables along a horizontal row
- Vertical line specifies values

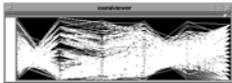


Dataset in a Cartesian graph

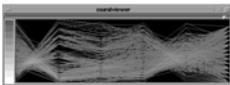


Same dataset in parallel coordinates

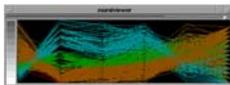
Parallel Coords Example



Basic



Grayscale



Color

Limitations and Issues

- **Complexity**
 - Many of these systems seem only appropriate for expert use
- **User testing**
 - Minimal evidence of user testing in most cases
