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

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

Tukey box plot

low   Middle 50%   high

Mean

0   20

Bivariate Data

- Representations

Scatter plot is common

price

mileage

Trivariate Data

- Representations

3D scatter plot is possible

price

horsepower

mileage
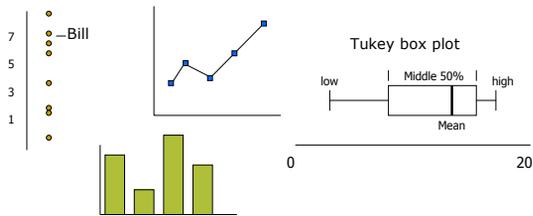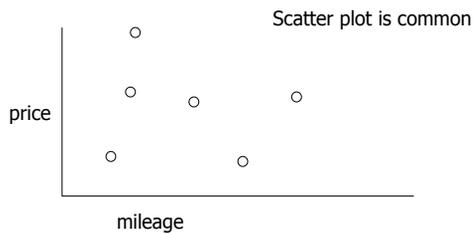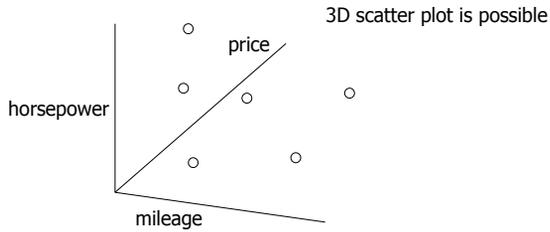
Multivariate Data

- 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

1

2

3

4

A B C D E
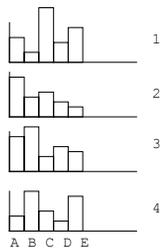
Scatterplot Matrix

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

Useful for what?
Misses what?
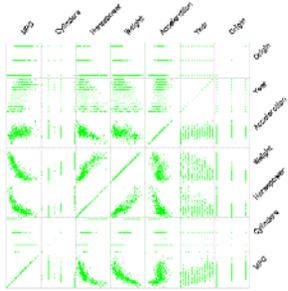
Chernoff Faces

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
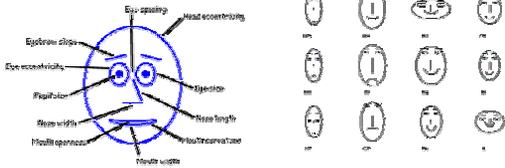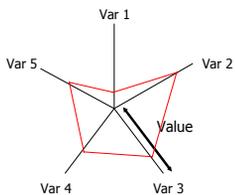
Star Plots

Space out the n
variables at equal
angles around a
circle

Each "spoke" encodes
a variable's value

Star Plot examples
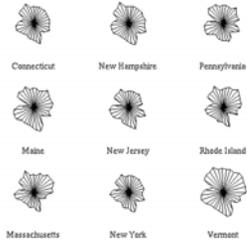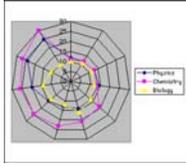
Parallel Coordinates

- Encode variables along a horizontal row
- Vertical line specifies values
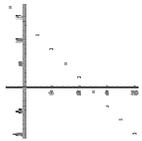
Dataset in a Cartesian graph
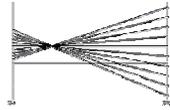
Same dataset in parallel coordinates

Parallel Coords Example
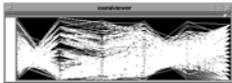
Basic
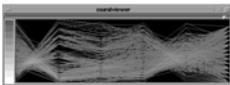
Grayscale
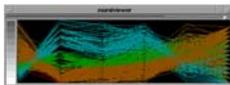
Color

Limitations and Issues

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