Multivariate Visual Representations

CS 7450 - Information Visualization
Jan. 27, 2011
John Stasko

Agenda

- General representation techniques for multivariate (>3) variables per data case
  - But not lots of variables yet…

1

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
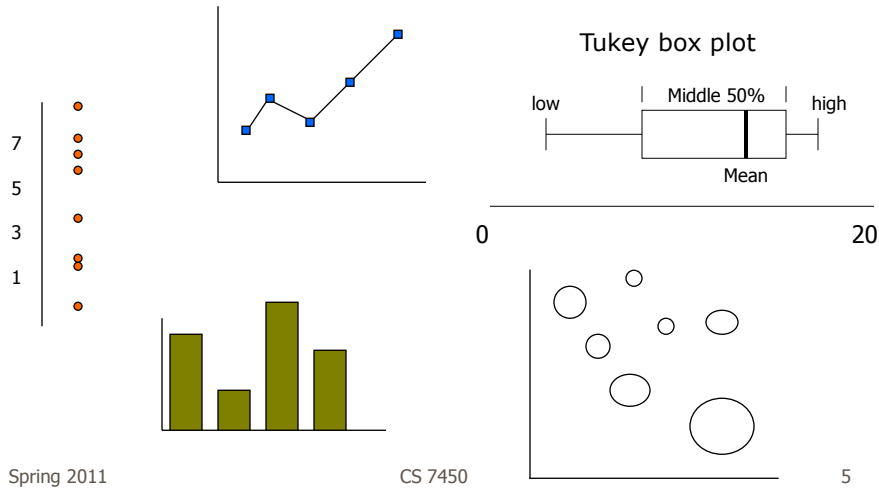 - >3 - Hypervariate data Focus Today

Earlier

- We examined a number of tried-and-true techniques/visualizations for presenting multivariate (typically <=3) data sets
  - Hinted at how to go above 3 dimensions

Representations

Some standard ways for low-d data

Hypervariate Data

- How about 4 to 20 or so variables (for instance)?
  - Lower-dimensional hypervariate data
  - (Much higher dimensions next week)
  - Many data sets fall into this category

More Dimensions

- Fundamentally, we have 2 geometric (position) display dimensions
- For data sets with >2 variables, we must project data down to 2D
- Come up with visual mapping that locates each dimension into 2D plane

- Computer graphics: 3D->2D projections

Wait a Second

- A spreadsheet already does that
  - Each variable is positioned into a column
  - Data cases in rows
  - This is a projection (mapping)

- What about some other techniques?
  - Already seen a couple

Multiple Views
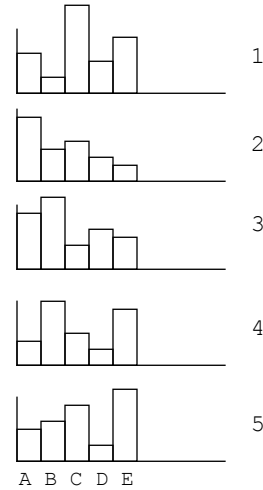
Give each variable its own display

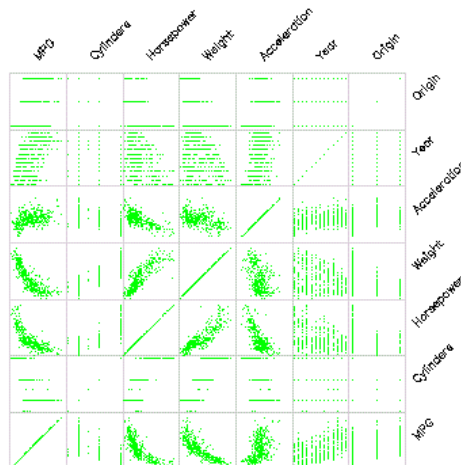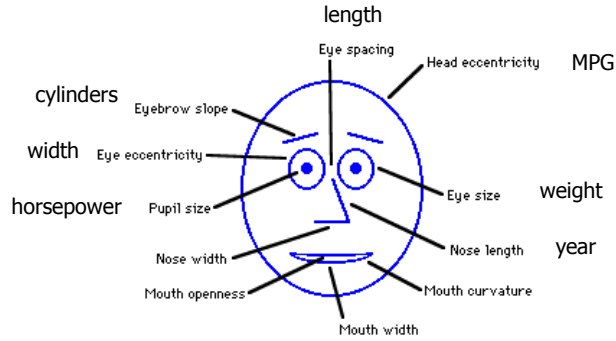|	A	B	C	D	E
1	4	1	8	3	5
2	6	3	4	2	1
3	5	7	2	4	3
4	2	6	3	1	5
5	3	4	5	1	7

Scatterplot Matrix

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

5

Chernoff Faces

Encode different variables' values in characteristics
of human face

length

Eye spacing

Head eccentricity

MPG

cylinders
Eyebrow slope

width   Eye eccentricity

horsepower   Pupil size

Eye size

weight

Nose length

year

Nose width

Mouth openness

Mouth curvature

Mouth width

Examples

ME   MA   NY   OH

NH   RI   NJ   IN

VT   CT   PA   IL

Cute applets: `http://www.cs.uchicago.edu/~wiseman/chernoff/`
`http://hesketh.com/schampeo/projects/Faces/chernoff.html`
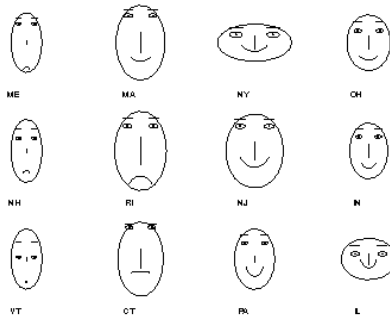
Table Lens

- Spreadsheet is certainly one hypervariate data presentation
- Idea: Make the text more visual and symbolic
- Just leverage basic bar chart idea

Rao & Card
CHI '94

Visual Mapping

	A	B	C	D	E	F
1	Sales rep	Quota	Variance to quota	% of quota	Forecast	Actual bookings
2	Albright, Gary	200,000	-16,062	92	205,000	183,938
3	Brown, Sheryll	150,000	84,983	157	260,000	234,983
4	Cartwright, Bonnie	100,000	-56,125	44	50,000	43,875
5	Caruthers, Michael	300,000	-25,125	92	324,000	274,875
6	Garibaldi, John	250,000	143,774	158	410,000	393,774
7	Girard, Jean	75,000	-48,117	36	50,000	26,883
8	Jones, Suzanne	140,000	-5,204	96	149,000	134,796
9	Larson, Terri	350,000	238,388	168	600,000	588,388
10	LeShan, George	200,000	-75,126	62	132,000	124,874
11	Levenson, Bernard	175,000	-9,267	95	193,000	165,733
12	Mulligan, Robert	225,000	34,383	115	275,000	259,383
13	Tetracelli, Sheila	50,000	-1,263	97	50,000	48,737
14	Wotisek, Gillian	190,000	-3,648	98	210,000	186,352
15						

Change quantitative values to bars

7

Tricky Part

What do you do for
nominal data?

Instantiation
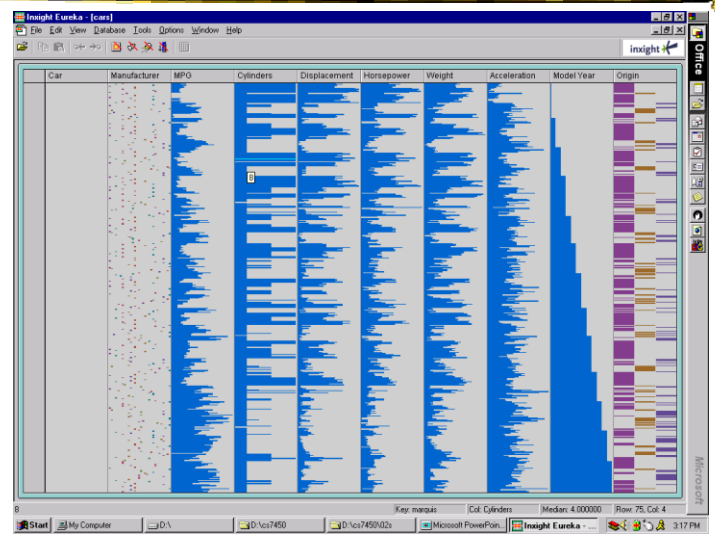
Details

Focus on item(s) while showing the context

See It

http://www.open-video.org/details.php?videoid=8304

Video

FOCUS

- Feature-Oriented Catalog User Interface
- Leverages spreadsheet metaphor again
- Items in columns, attributes in rows
- Uses bars and other representations for attribute values

Spenke, Beilken, & Berlage
UIST '96

Figure 1: An overview of the printer table.

10

Characteristics

- Can sort on any attribute (row)
- Focus on an attribute value (show only cases having that value) by double-clicking on it
- Can type in queries on different attributes to limit what is presented too

Limit by Query

Figure 4: A disjunction.

Manifestation

InfoZoom

Commercial product to be demo'ed coming up

Categorical data?

- How about multivariate categorical data?

- Students
  - Gender: Female, male
  - Eye color: Brown, blue, green, hazel
  - Hair color: Black, red, brown, blonde, gray
  - Home country: USA, China, Italy, India, …

Mosaic Plot

Mosaic Plot

Women   Men

13

Mosaic Plot

Brown

Hazel

Green

Blue

Women   Men

Mosaic Plot

Black Red Brown Blond

Brown

Hazel

Green

Blue

Women   Men

14

Attr Explorer

- General hypervariate data representation combined with flexible interaction

Spence & Tweedie
Inter w Computers '98

Characteristics

- Multiple histogram views, one per attribute (like trellis)
- Each data case represented by a square
- Square is positioned relative to that case's value on that attribute
- Selecting case in one view lights it up in others
- Query sliders for narrowing
- Use shading to indicate level of query match (darkest for full match)

Features

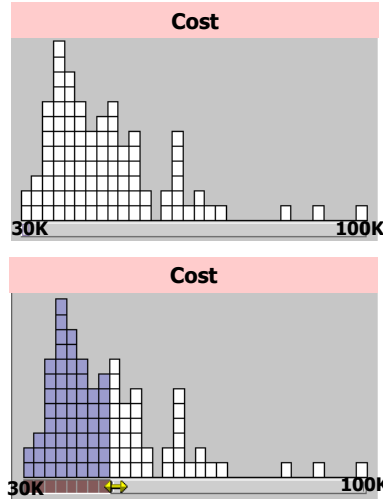- Attribute histogram
- All objects on all attribute scales

Cost

- Interaction with attributes limits

Cost

Features

- Inter-relations between attributes – brushing

Cost

Bedroom

Journey time

16
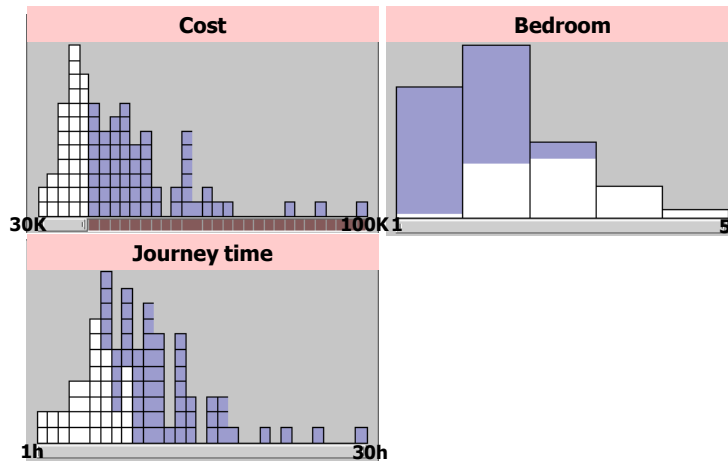
Features

- Color-encoded sensitivity
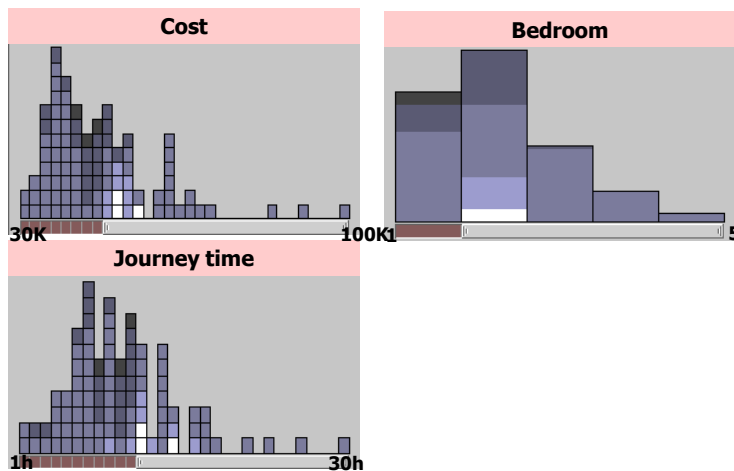
Attribute Explorer

Video

http://www.open-video.org/details.php?videoid=8162

Summary

- Summary
  - Attribute histogram
  - Attribute relationship
  - Sensitivity information
  - Especially useful in "zero-hits" situations or when you are not familiar with the data at all

- Limitations
  - Limits on the number of attributes

MultiNav

- Each different attribute is placed in a different row
- Sort the values of each row
  - Thus, a particular item is not just in one column
- Want to support browsing
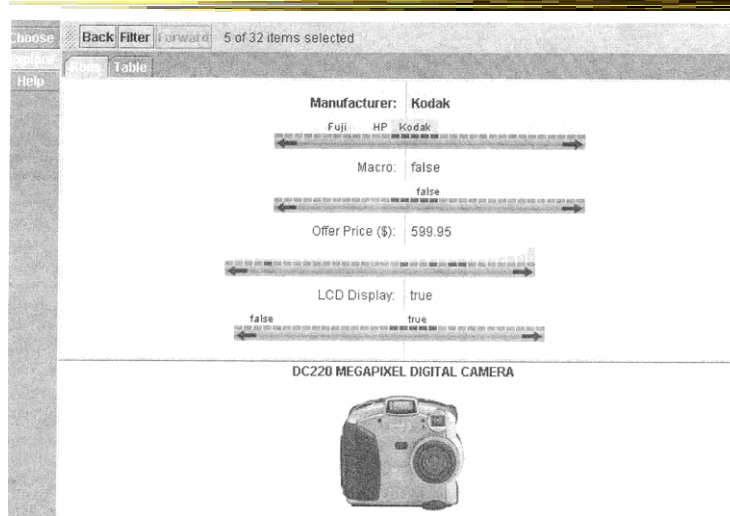
Lanning et al
AVI '00

Interface
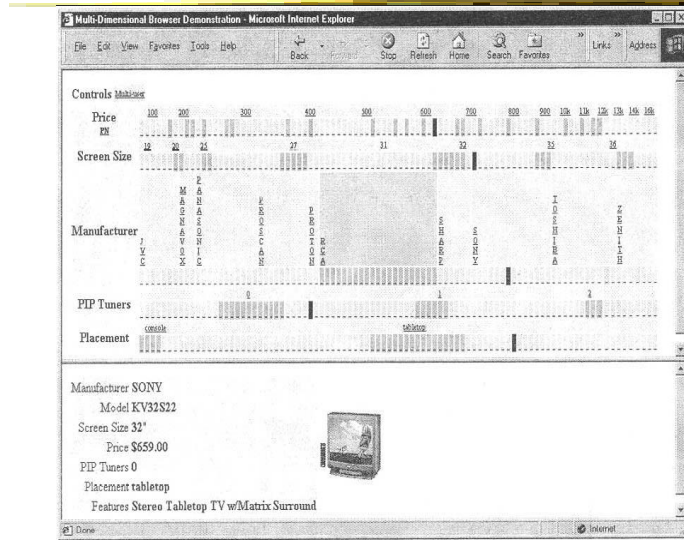
Alternate UI

- Can slide the values in a row horizontally
- A particular data case then can be lined up in one column, but the rows are pushed unequally left and right

19

Attributes as Sliding Rods

Information-Seeking Dialog
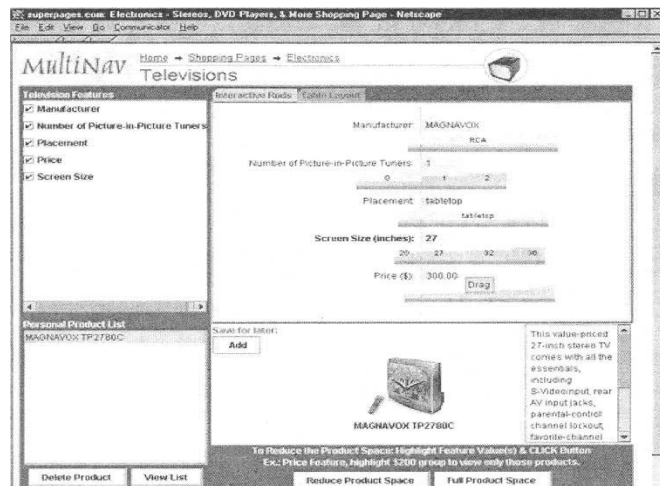
20

Instantiation
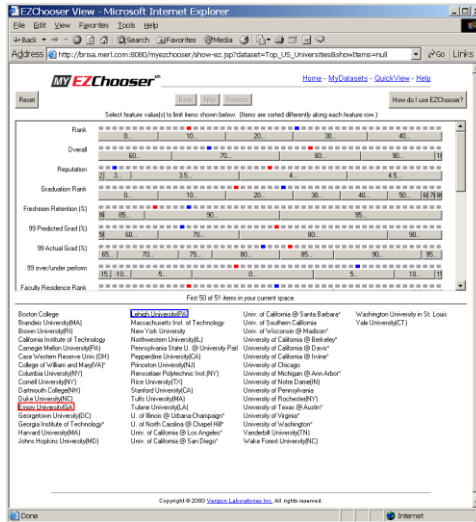
Demo

Limitations

- Number of cases (horizontal space)
- Nominal & textual attributes don't work quite as well

Dust & Magnet

- Altogether different metaphor
- Data cases represented as small bits of iron dust
- Different attributes given physical manifestation as magnets
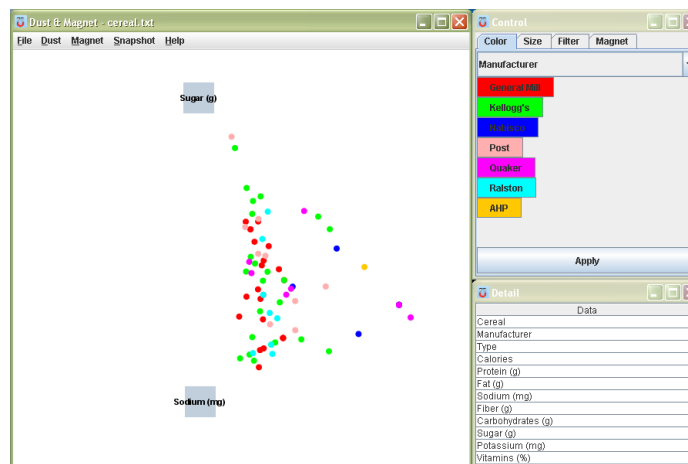- Interact with objects to explore data

Yi, Melton, Stasko & Jacko
Info Vis '05

Interface

Interaction

- Iron bits (data) are drawn toward magents (attributes) proportional to that data element's value in that attribute
  - Higher values attracted more strongly
- All magnets present on display affect position of all dust
- Individual power of magnets can be changed
- Dust's color and size can connected to attributes as well
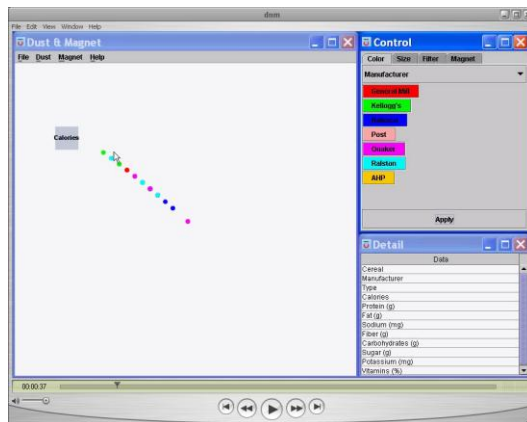
Interaction

- Moving a magnet makes all the dust move
  - Also command for shaking dust
- Different strategies for how to position magnets in order to explore the data

See It Live

Video & Demo

Parallel Coordinates

- What are they?
  - Explain...

Parallel Coordinates

	V1	V2	V3	V4	V5
D1	7	3	4	8	1
D2	2	7	6	3	4
D3	9	8	1	4	2
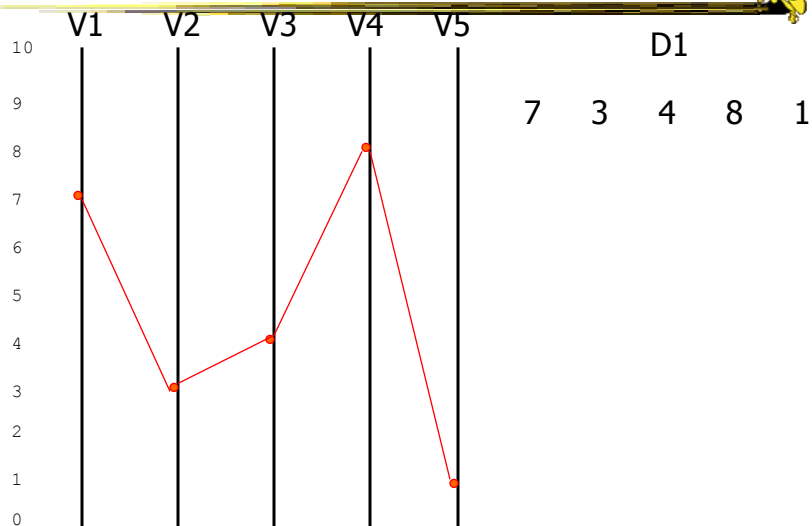
Parallel Coordinates

D1

7   3   4   8   1

25

Parallel Coordinates

V1	V2	V3	V4	V5	D2
2	7	6	3	4	

Parallel Coordinates

V1	V2	V3	V4	V5	D3
9	8	1	4	2	
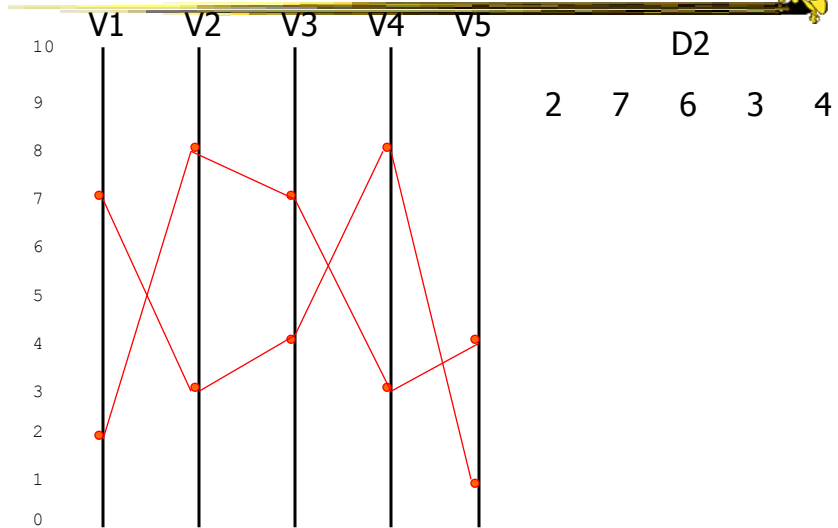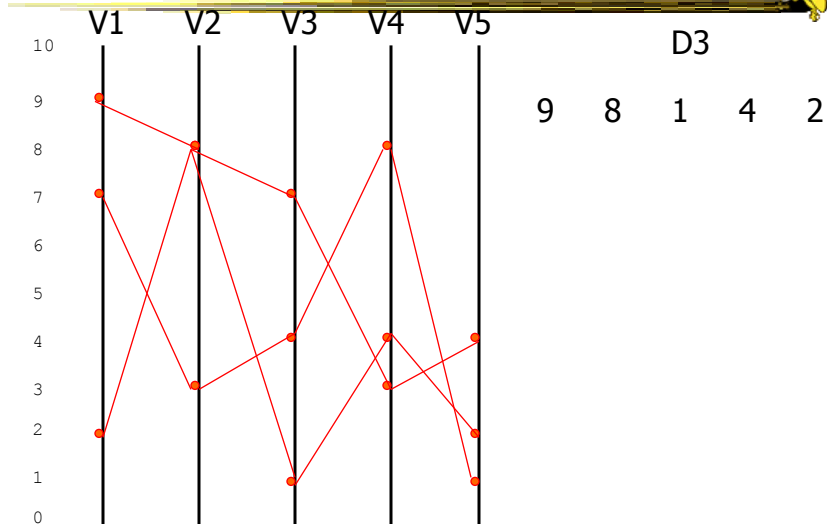
Parallel Coordinates
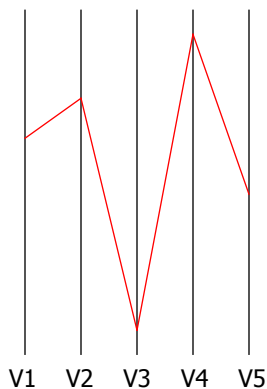

Encode variables along
a horizontal row

Vertical line specifies different
values that variable can take

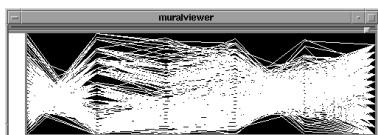Data point represented as a
polyline

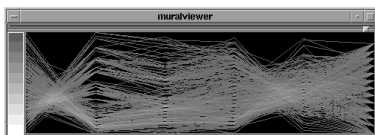V1   V2   V3   V4   V5

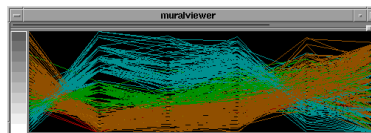# Parallel Coords Example

Basic

Grayscale

Color

Issue

- Different variables can have values taking on quite different ranges
- Must normalize all down (e.g., 0->1)

Application

- System that uses parallel coordinates for information analysis and discovery
- Interactive tool
  - Can focus on certain data items
  - Color

Taken from:

A. Inselberg, "Multidimensional Detective" InfoVis '97, 1997.

Discuss

- What was their domain?
- What was their problem?
- What were their data sets?

The Problem

- VLSI chip manufacture
- Want high quality chips (high speed) and a high yield batch (% of useful chips)
- Able to track defects
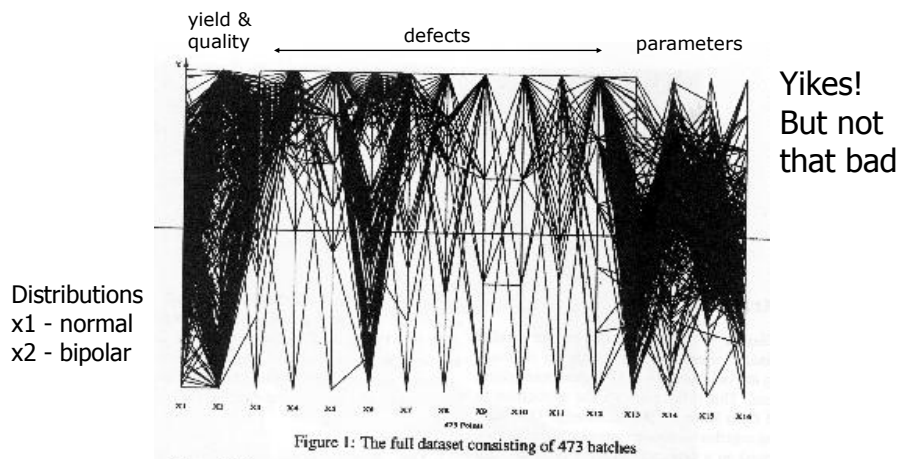- Hypothesis: No defects gives desired chip types
- 473 batches of data

The Data

- 16 variables
  - X1 - yield
  - X2 - quality
  - X3-X12 - # defects (inverted)
  - X13-X16 - physical parameters

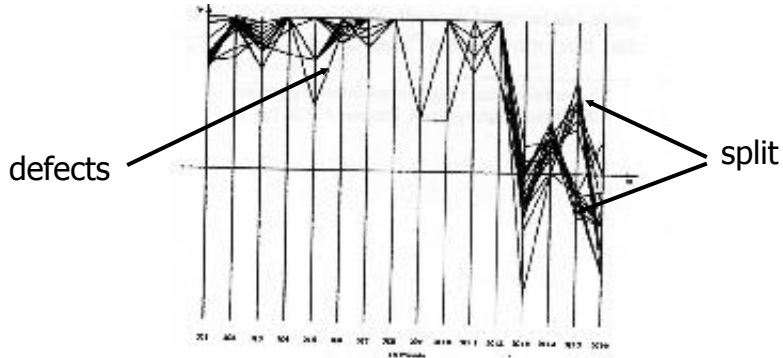# Parallel Coordinate Display

yield &
quality   defects   parameters

Yikes!
But not
that bad

Distributions
x1 - normal
x2 - bipolar

Figure 1: The full dataset consisting of 473 batches

Top Yield & Quality

defects

split

Figure 2: The batches high in Yield, $X1$, and Quality, $X2$,

Have some defects

Minimal Defects

Not the highest yields and quality
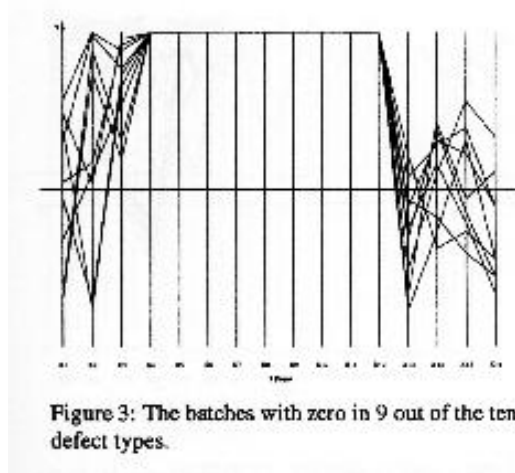
Figure 3: The batches with zero in 9 out of the ten defect types.

Best Yields

Appears that
some defects
are necessary
to produce
the best chips

Non-intuitive!

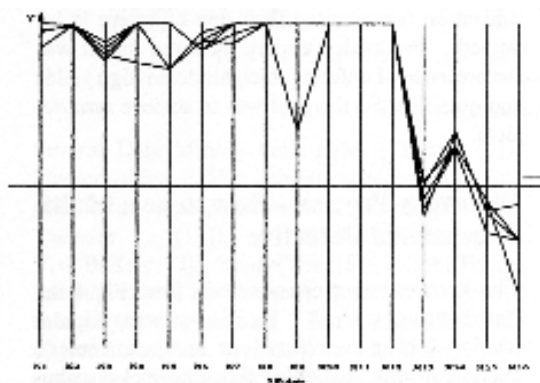Figure 6: Batches with the highest Yields do not
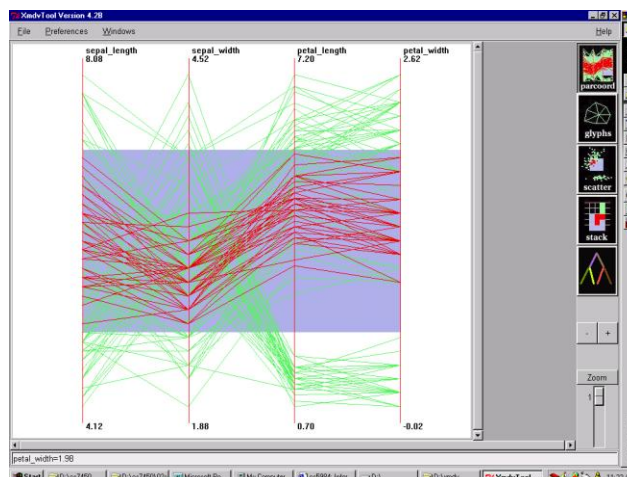have the lowest defects in $X3$ and $X6$.

XmdvTool

Toolsuite created
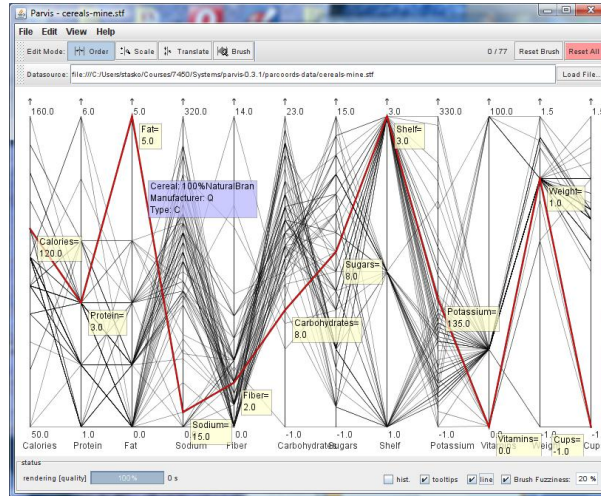by Matthew Ward
of WPI

Includes parallel
coordinate views

ParVis System

Demo

Challenges

Too much data
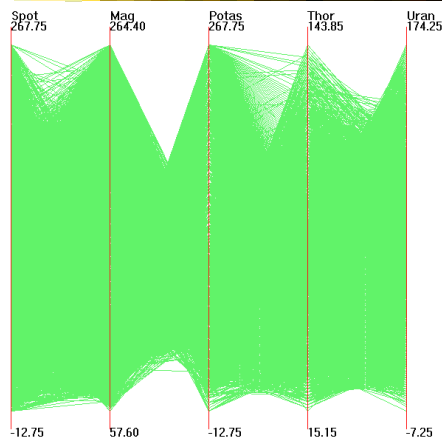
Out5d dataset (5 dimensions, 16384 data items)

(courtesy of J. Yang)

Dimensional Reordering

Which dimensions are most like each other?
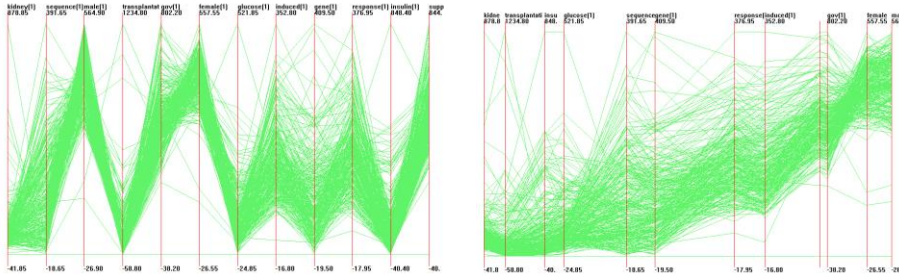
Same dimensions ordered according to similarity

Yang et al
InfoVis '03

Dimensional Reordering

Can you reduce
clutter and highlight
other interesting
features in data by
changing order of
dimensions?

Figure 1: Parallel coordinates visualization of Cars dataset. Outliers are highlighted with red in (b).

Figure 2: Parallel coordinates visualization of Cars dataset after clutter-based dimension reordering. Outliers are highlighted with red in (b).

Peng et al
InfoVis '04

Reducing Density
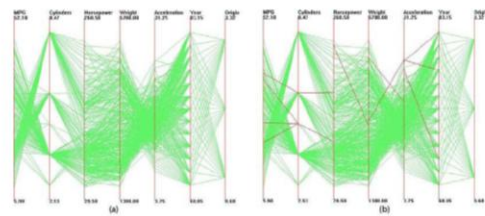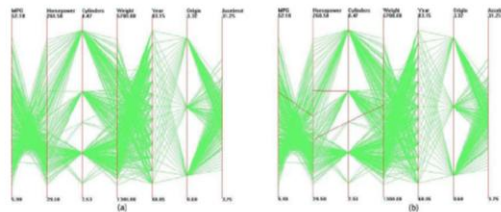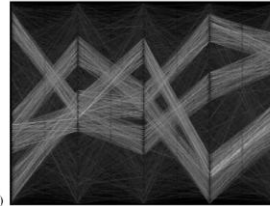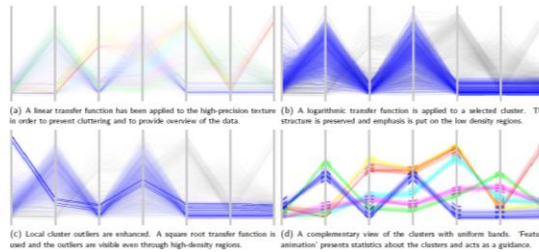
Figure 1 – Parallel Coordinates visualization of the *Sint1* data set
(7,500 five-attribute records).

Artero et al, 04

(a) A linear transfer function has been applied to the high-precision texture in order to prevent cluttering and to provide overview of the data.

(b) A logarithmic transfer function is applied to a selected cluster. The structure is preserved and emphasis is put on the low density regions.

(c) Local cluster outliers are enhanced. A square root transfer function is used and the outliers are visible even through high-density regions.

(d) A complementary view of the clusters with uniform bands. 'Feature animation' presents statistics about the clusters and acts as a guidance.

Johansson et al, '05

Improved Interaction

- How do we let the user select items of interest?
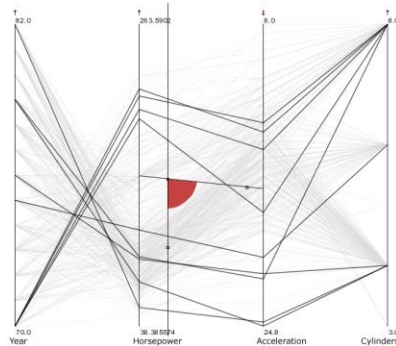- Obvious notion of clicking on one of the polylines, but how about something more than that

Attribute Ratios

- Angular Brushing
    - Select subsets which exhibit a correlation along 2 axes by specifying angle of interest

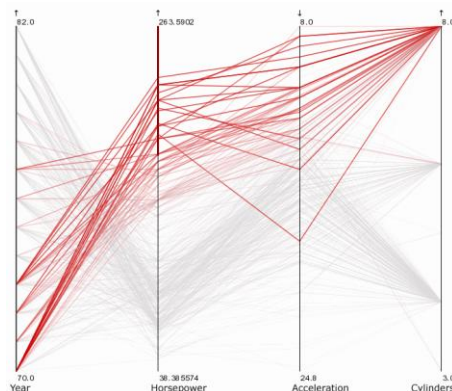Hauser, Ledermann, & Doleisch
InfoVis '02

(earlier demo)

Range Focus

- Smooth Brushing
    - Specify a region of interest along one axis

Combining

- Composite Brushing
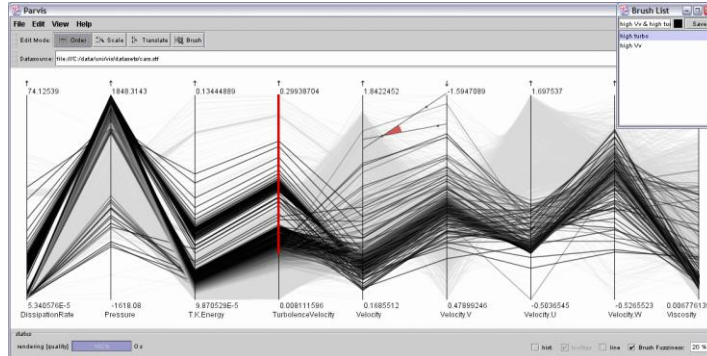 - Combine brushes and DOI functions using logical operators
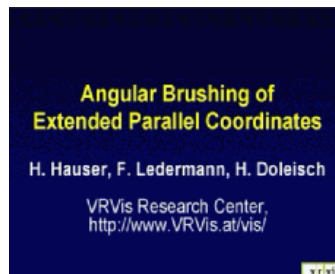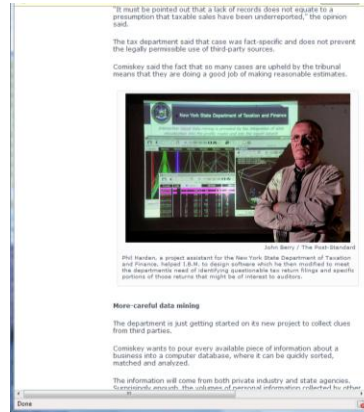
Video

Angular Brushing of Extended Parallel Coordinates

H. Hauser, F. Ledermann, H. Doleisch

VRVis Research Center,
http://www.VRVis.at/vis/

http://www.vrvis.at/via/research/ang-brush/parvis4.mov

Application

Different Kinds of Data

- How about categorical data?
  - Can parallel coordinates handle that well?

Parallel Sets

- Visualization method adopting parallel coordinates layout but uses frequency-based representation
- Visual metaphor
  - Layout similar to parallel coordinates
  - Continuous axes replaced with boxes
- Interaction
  - User-driven: User can create new classifications
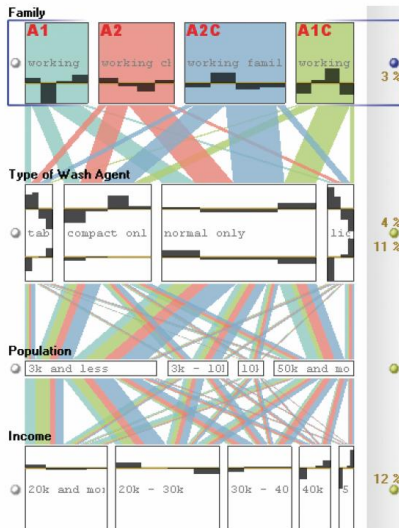
Kosara, Bendix, & Hauser
TVCG '05

Spring 2011   CS 7450   77

Representation

Color used for different categories

Those values flow into the other variables

Spring 2011   CS 7450   78
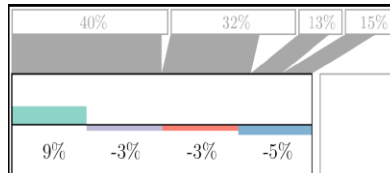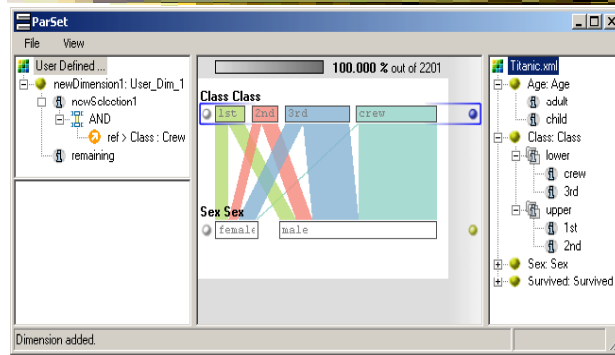
Example

Titanic passengers data set

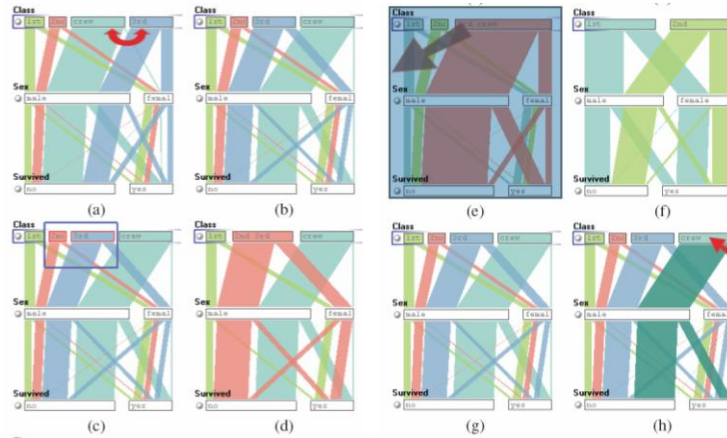| Class	Sex		
	female	male	
first	145 44.6% 30.8% 6.6%	180 55.4% 10.4% 8.2%	325 14.8%
second	106 37.2% 22.6% 4.8%	179 62.8% 10.4% 8.1%	285 12.9%
third	196 27.8% 41.7% 8.9%	510 72.2% 29.5% 23.2%	706 32.1%
crew	23 2.6% 4.9% 1.1%	862 97.4% 49.8% 39.1%	885 40.2%
	470 21.4%	1731 78.6%	2201 100%

Titanic Data Set

Interactions

Fig. 7. Basic interaction elements in Parallel Sets: reordering categories (a, b) helps to generate a more meaningful layout; grouping categories (c, d) enables a hierarchical analysis/exploration; excluding categories from the visualization (e, f) allows for interactive filtering; and category highlighting (g, h) enables the selective investigation of high-dimensional relations.

Video

InfoVis '05
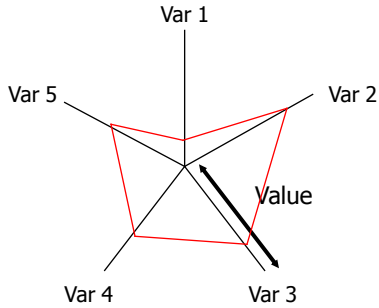
41

Star Plots

Var 1

Var 5

Var 2

Value

Var 4

Var 3
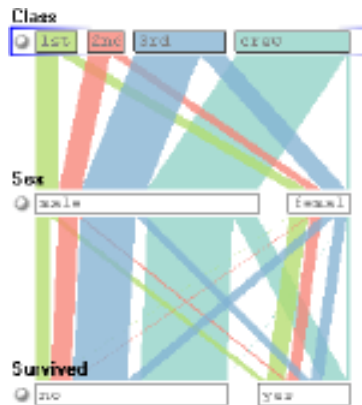
Space out the n
variables at equal
angles around a
circle

Each "spoke" encodes
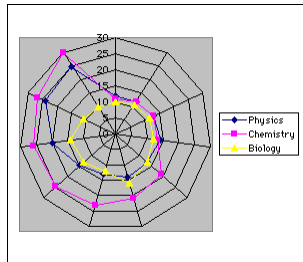a variable's value

Alternative Rep.

Data point is now a "shape"

Star Plot examples

Connecticut   New Hampshire   Pennsylvania

Maine   New Jersey   Rhode Island

Massachusetts   New York   Vermont

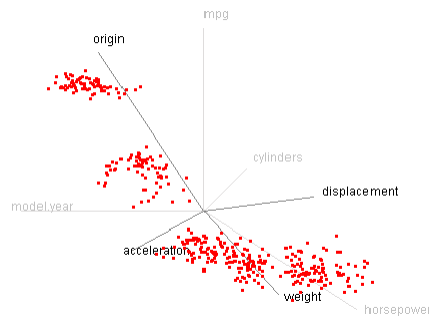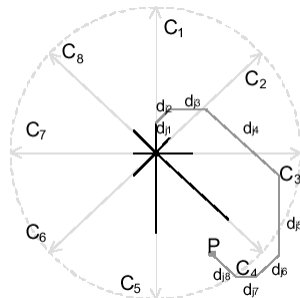http://seamonkey.ed.asu.edu/~behrens/asu/reports/compre/comp1.html

Star Coordinates

- Same ideas as star plot
- Rather than represent point as polyline, just accumulate values along a vector parallel to particular axis
- Data case then becomes a point

Star Coordinates

E. Kandogan, "Star Coordinates: A Multi-dimensional Visualization Technique with Uniform Treatment of Dimensions", InfoVis 2000 Late-Breaking Hot Topics, Oct. 2000

Demo

Star Coordinates

- Data cases with similar values will lead to clusters of points
- (What's the problem though?)
- Multi-dimensional scaling or projection down to 2D

Parallel Coordinates

- Technique
  - Strengths?
  - Weaknesses?

HW 3

- Analysis and Visual Design
  - Three datasets, choose one
  - Construct three analytic queries
  - Design visualization (sketch)
- Due Tues Feb 8

Upcoming

- Overview and Detail (Focus + Context)
  - Reading:
    Bederson et al, '04

- Interaction (2 days)
  - Reading
    Ward chapters 10, 11
    Few chapter 4
    Yi et al, '07