Topic Notes

Multivariate Visual Representations 1

CS 7450 - Information Visualization
Sep. 11, 2013
John Stasko

Agenda

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

Quick Quiz

- What type of dataset has three variables per case?
- What is a scatterplot matrix?

Revisit

How Many Variables?

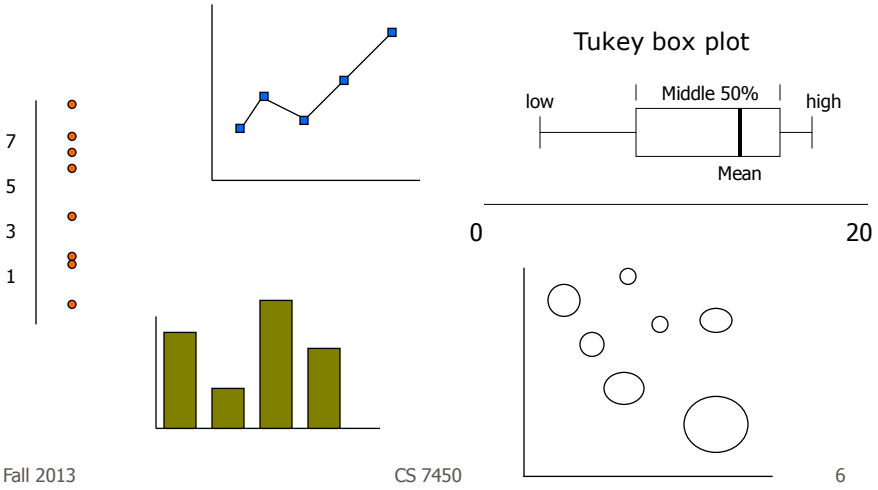
- Data sets of dimensions 1, 2, 3 are common
- Number of variables per class
  - 1 - Univariate data
  - 2 - Bivariate data
  - 3 - Trivariate data
  - >3 - Hypervariate data Focus Today

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
  - 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

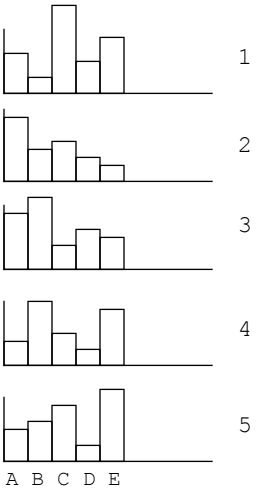
- A spreadsheet already does that
  - Each variable is positioned into a column
  - Data cases in rows
  - This is a projection (mapping)
- What about some other techniques?
  - Already seen a couple

Revisit

Multiple Views

Give each variable its own display

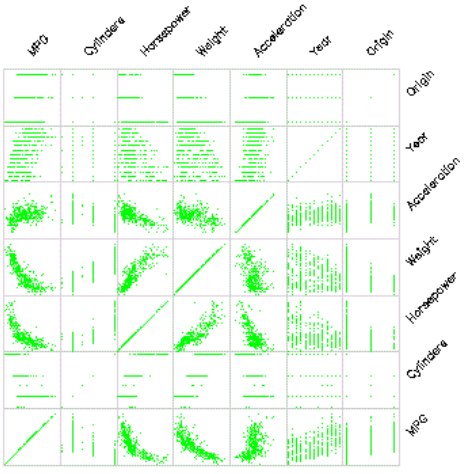
	A	B	C	D	E
1	4	1	8	3	5
2	6	3	4	2	1
3	5	7	2	4	3
4	2	6	3	1	5
5	3	4	5	1	7



Revisit

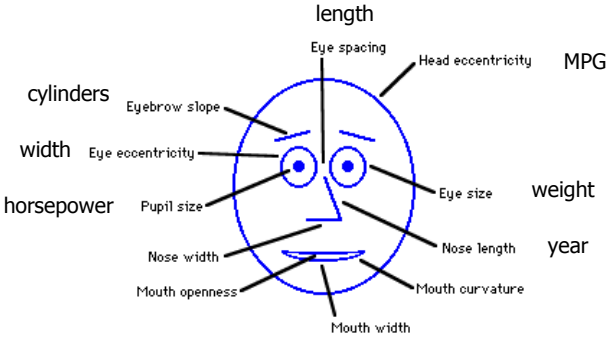
Scatterplot Matrix

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

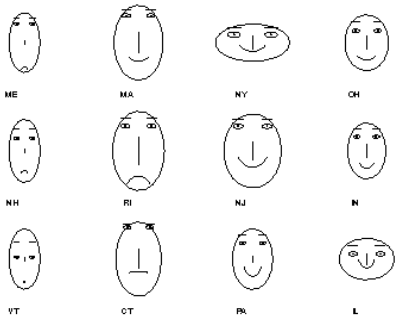


Chernoff Faces

Encode different variables' values in characteristics of human face



Examples



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

Tricky Part

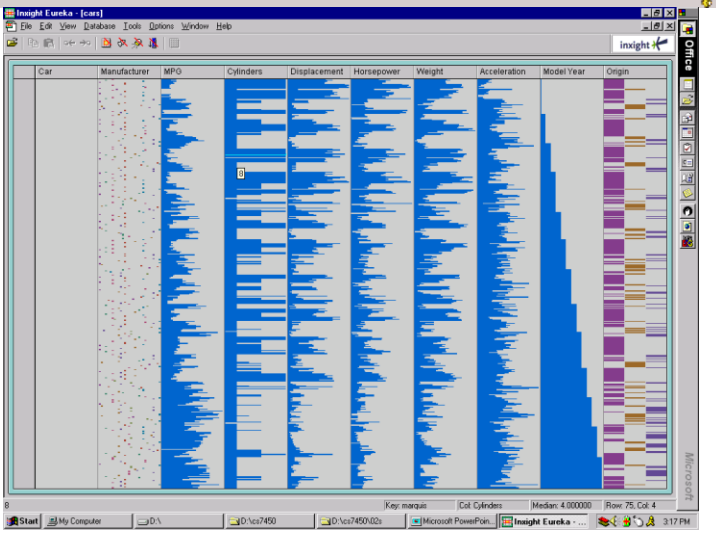
a1-cereals [Read-Only] [Compatibility Mode]

	A	B	C	D	E	F	G	H	I
1	Cereal	Manufactu	Type	Calories	Protein	Fat	Sodium	Fiber	Carbol
2	Frosted Mini-Wheats	K	C	100	3	0	0	3	
3	Raisin Squares	K	C	90	2	0	0	2	
4	Shredded Wheat	N	C	80	2	0	0	3	
5	Shredded Wheat 'n'Bran	N	C	90	3	0	0	4	
6	Shredded Wheat spoon s	N	C	90	3	0	0	3	
7	Puffed Rice	Q	C	50	1	0	0	0	
8	Puffed Wheat	Q	C	50	2	0	0	1	
9	Maypo	A	H	100	4	1	0	0	
10	Quaker Oatmeal	Q	H	100	5	2	0	2.7	
11	Strawberry Fruit Wheats	N	C	90	2	0	15	3	
12	100% Natural Bran	Q	C	120	3	5	15	2	
13	Golden Crisp	P	C	100	2	0	45	0	
14	Smacks	K	C	110	2	1	70	1	
15	Great Grains Pecan	P	C	120	3	3	75	3	
16	Cream of Wheat (Quick)	N	H	100	3	0	80	1	
17	Corn Pops	K	C	110	1	0	90	1	
18	Muesli Raisins, Dates, &	R	C	150	4	3	95	3	
19	Apple Jacks	K	C	110	2	0	125	1	

cereals

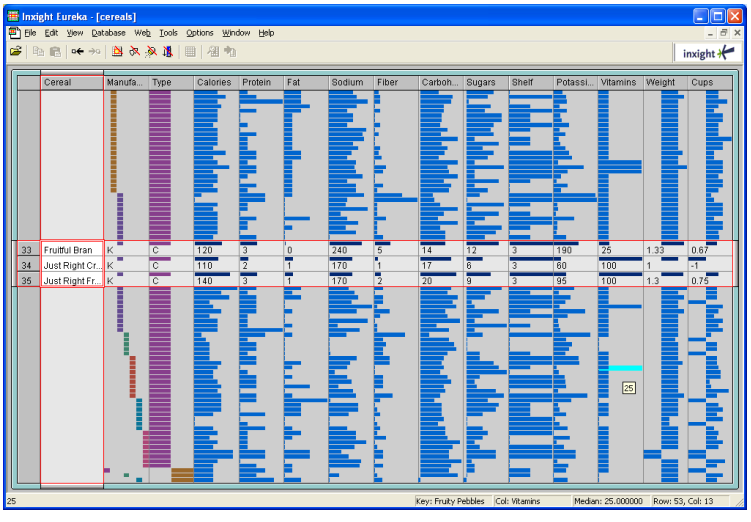
What do you do for nominal data?

Instantiation

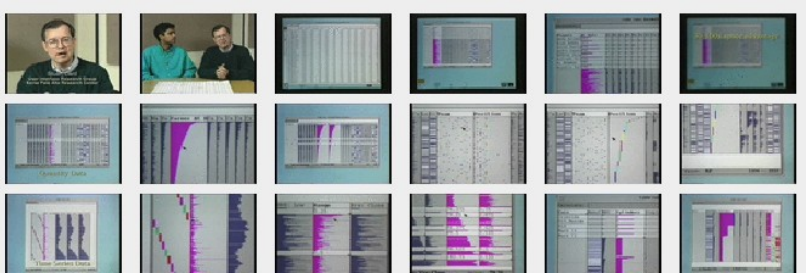


Details

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

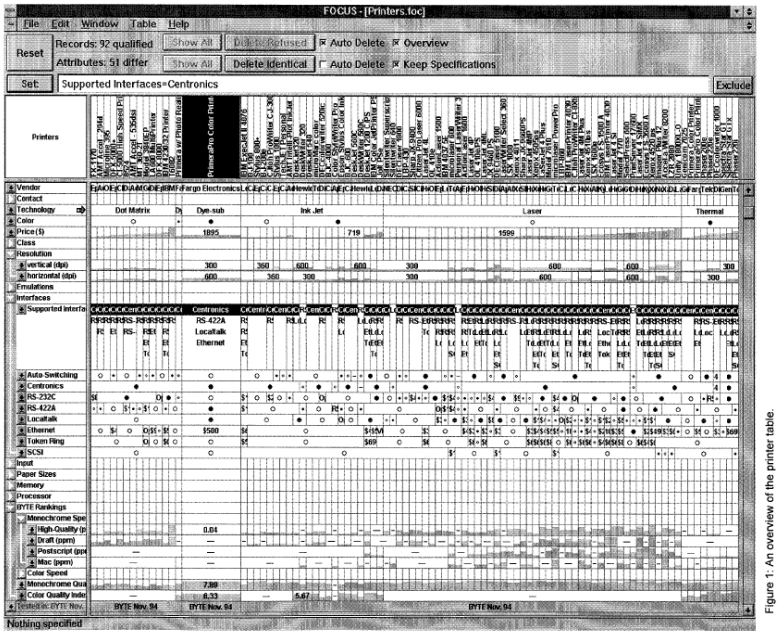


Figure 1: An overview of the printer table.

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

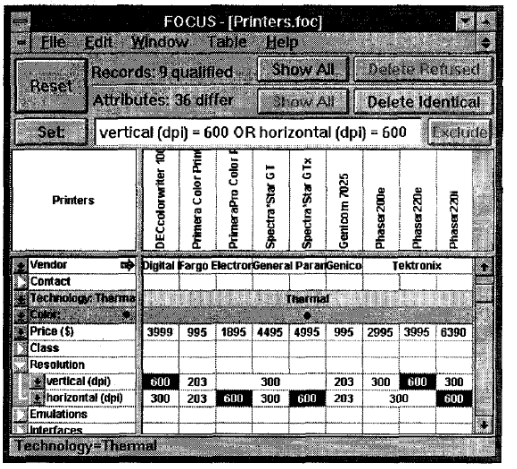
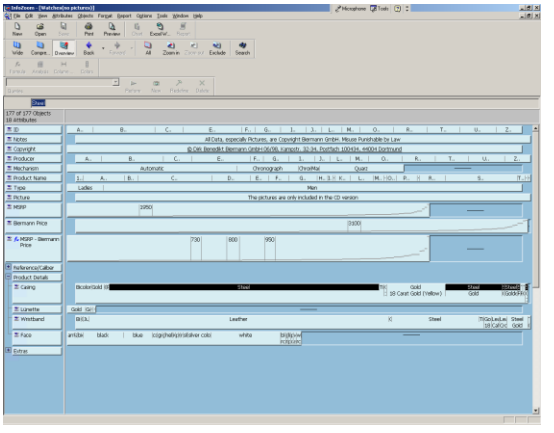


Figure 4: A disjunction.

Manifestation

InfoZoom



Commercial product to be demo'ed coming up

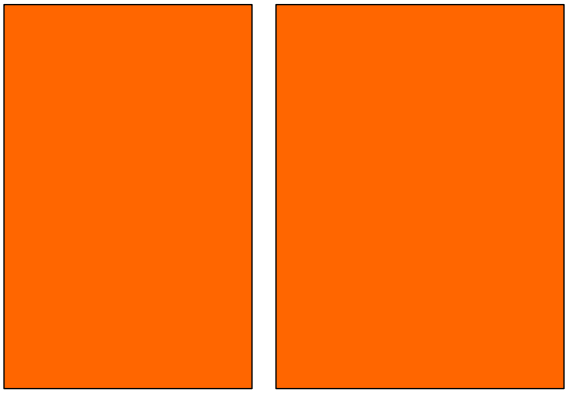
Categorical data?

- How about multivariate categorical data?

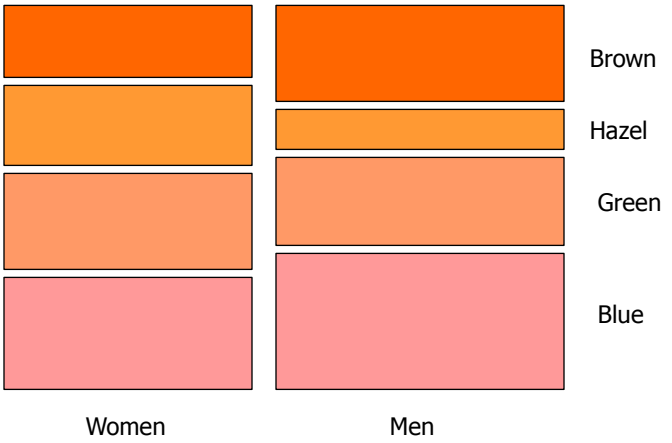
- Students
  - Gender: Female, male
  - Eye color: Brown, blue, green, hazel
  - Hair color: Black, red, brown, blonde, gray
  - Home country: USA, China, Italy, India, …

Mosaic Plot

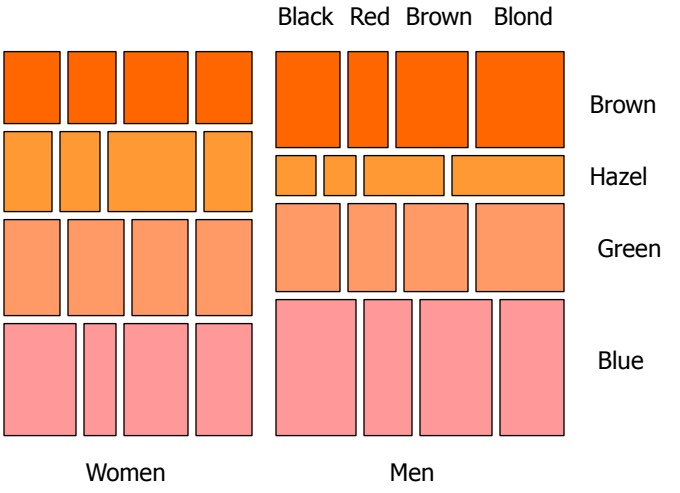
Mosaic Plot



Mosaic Plot



Mosaic Plot



Attribute Explorer

- General hypervariate data representation combined with flexible interaction

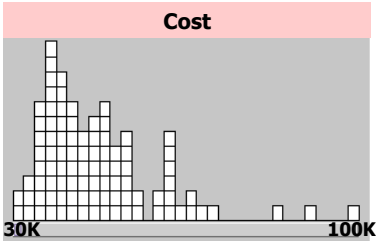
Spence & Tweedie
Inter w Computers '98

Characteristics

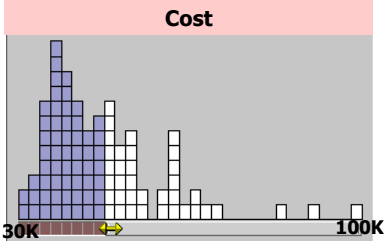
- Multiple histogram views, one per attribute (like trellis)
- Each data case represented by a square
- Square is positioned relative to that case's value on that attribute
- Selecting case in one view lights it up in others
- Query sliders for narrowing
- Use shading to indicate level of query match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

Features

- Attribute histogram
- All objects on all attribute scales

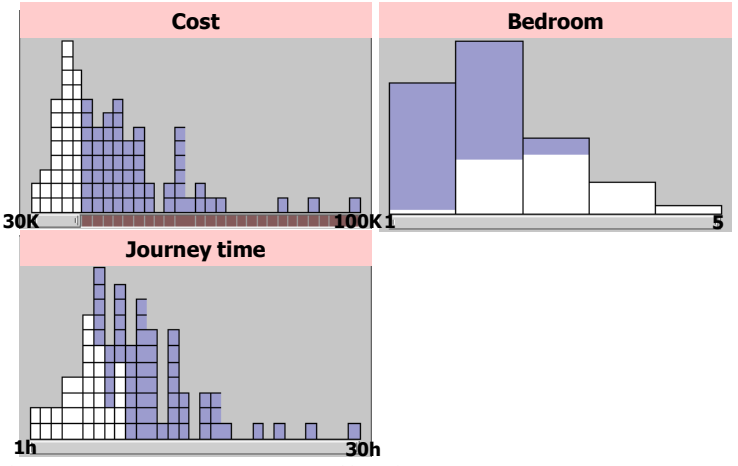


- Interaction with attributes limits



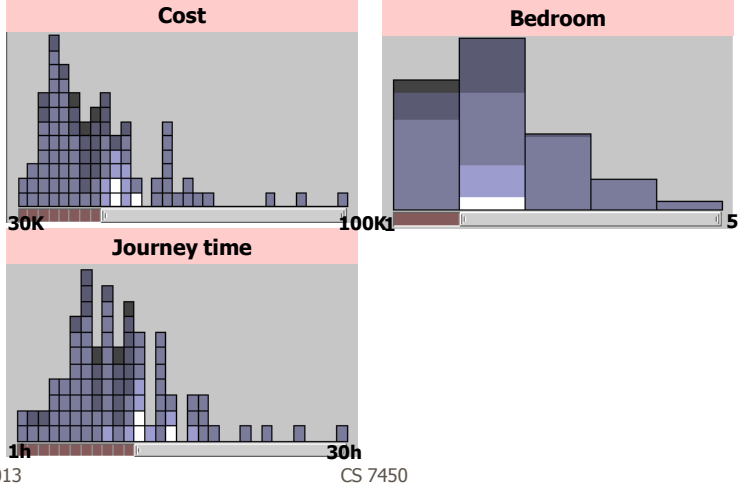
Features

- Inter-relations between attributes – brushing



Features

- Color-encoded sensitivity



Attribute Explorer

Video

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

Summary

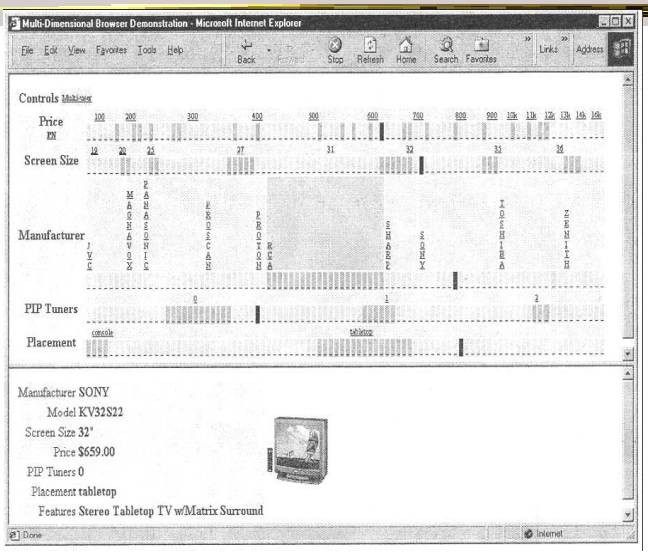
- Summary
  - Attribute histogram
  - Attribute relationship
  - Sensitivity information
  - Especially useful in "zero-hits" situations or when you are not familiar with the data at all
- Limitations
  - Limits on the number of attributes

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

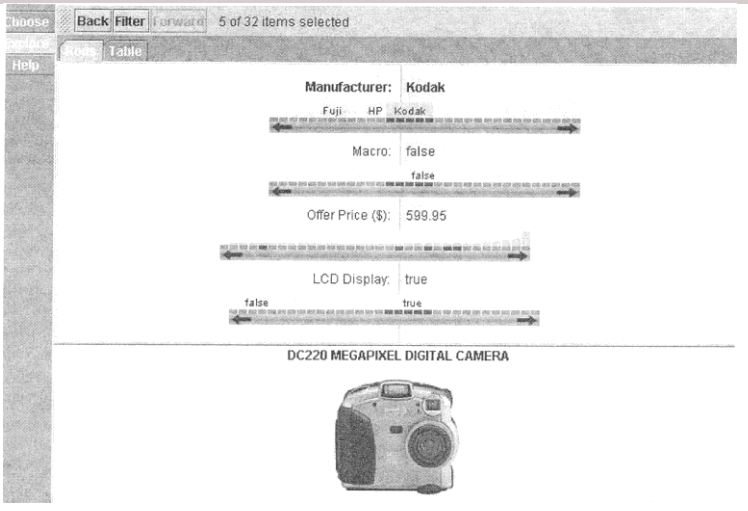
Interface



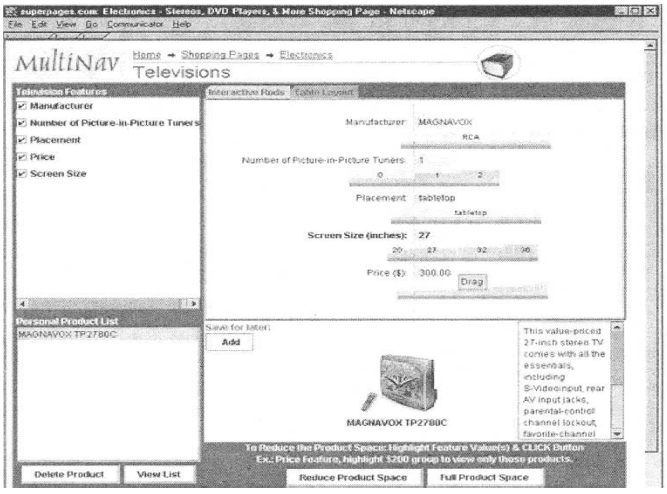
Alternate UI

- 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

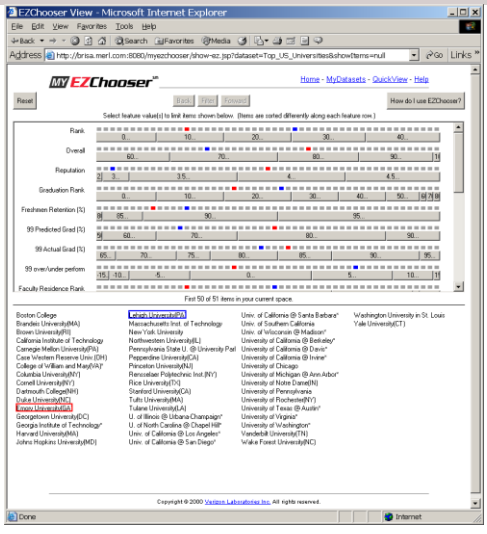
Attributes as Sliding Rods



Information-Seeking Dialog



Instantiation



Demo

Limitations

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

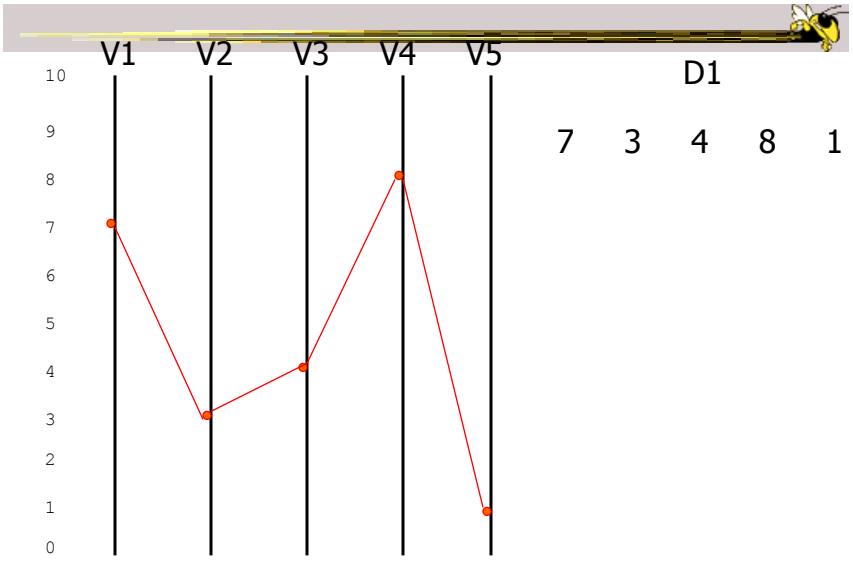
Parallel Coordinates

- What are they?
  - Explain…

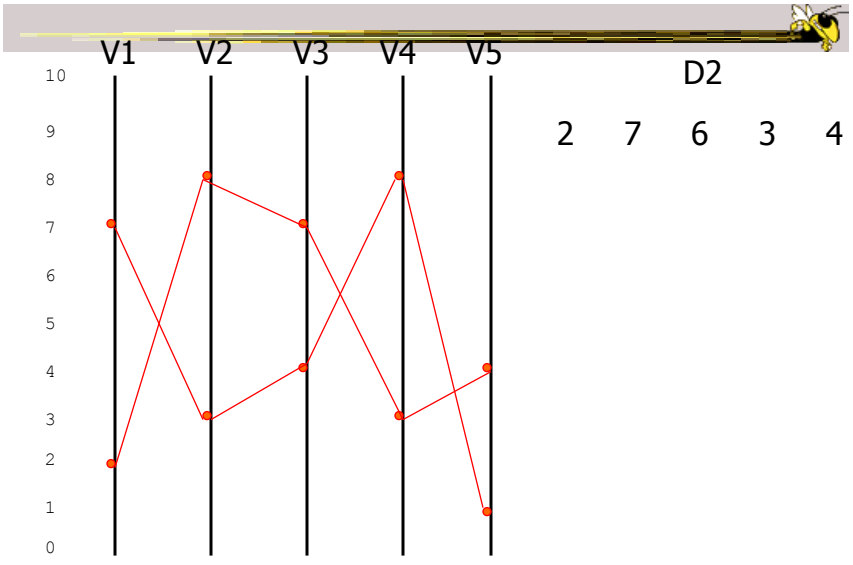
Parallel Coordinates

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

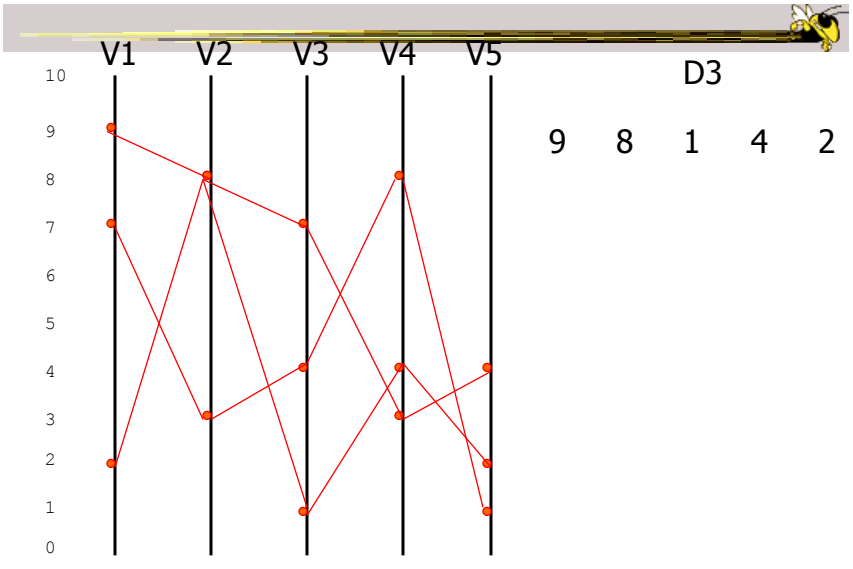
Parallel Coordinates



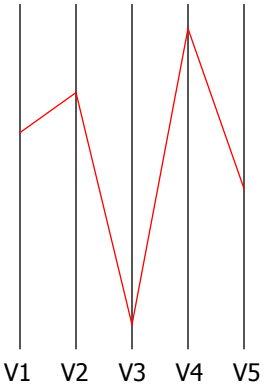
Parallel Coordinates



Parallel Coordinates



Parallel Coordinates

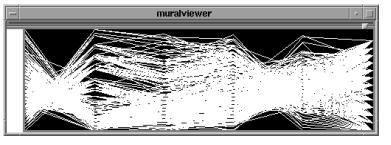


Encode variables along a horizontal row

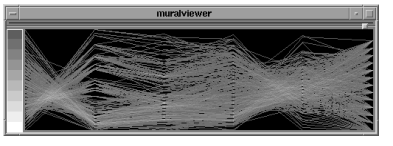
Vertical line specifies different values that variable can take

Data point represented as a polyline

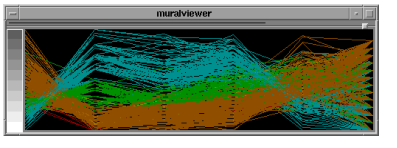
Parallel Coords Example



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?

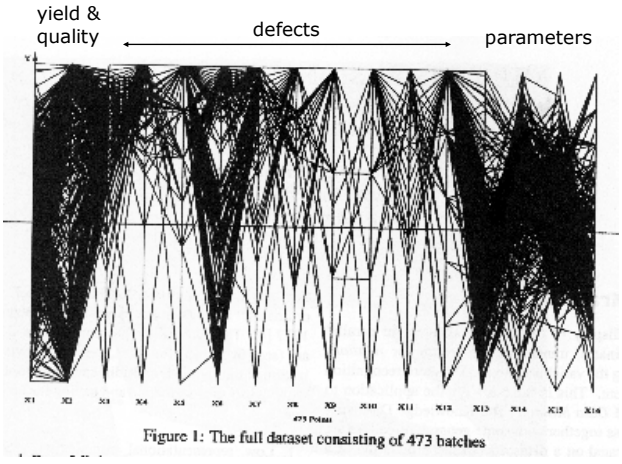
The Problem

- 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



Yikes!
But not
that bad

Distributions
x1 - normal
x2 - bipolar

Top Yield & Quality

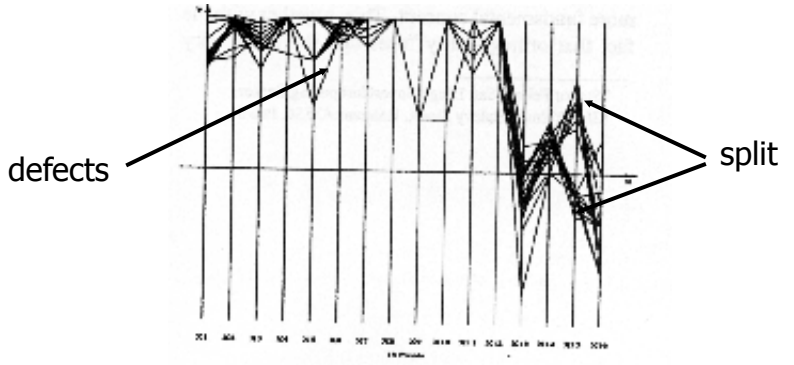


Figure 2: The batches high in Yield, $X1$, and Quality, $X2$.

Have some defects

Minimal Defects

Not the highest yields and quality

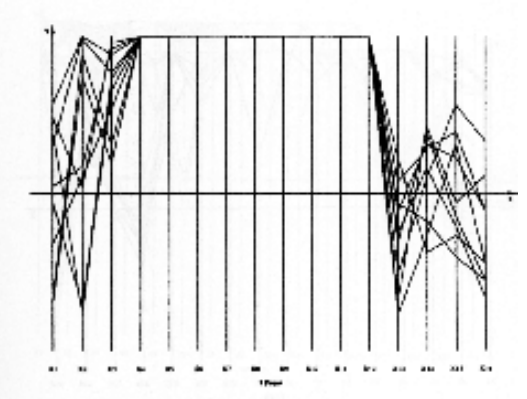


Figure 3: The batches with zero in 9 out of the ten defect types.

Best Yields

Appears that some defects are necessary to produce the best chips

Non-intuitive!

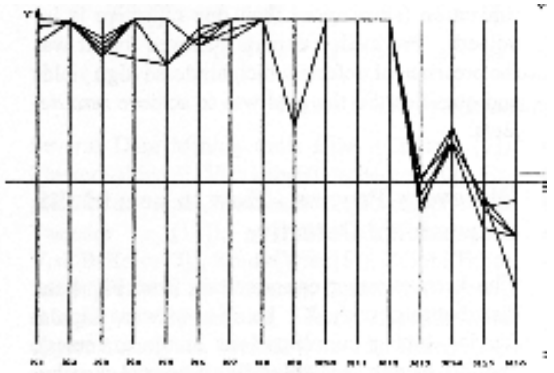
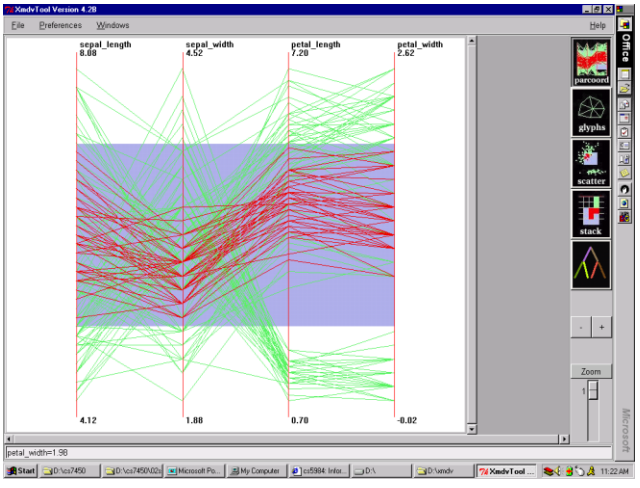


Figure 6: Batches with the highest Yields do not have the lowest defects in $X3$ and $X6$.

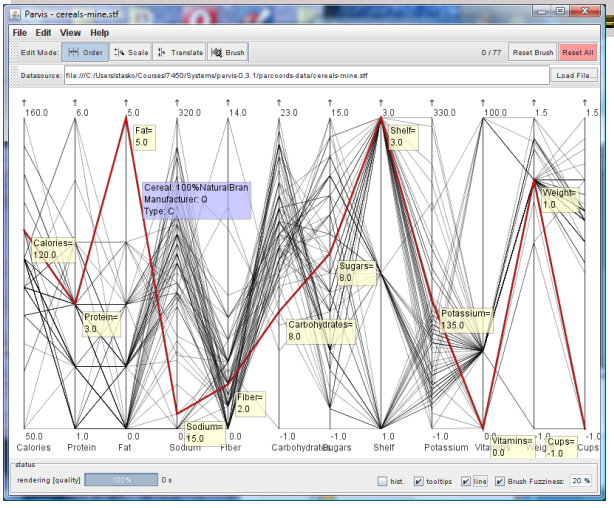
XmdvTool

Toolsuite created by Matthew Ward of WPI

Includes parallel coordinate views



ParVis System

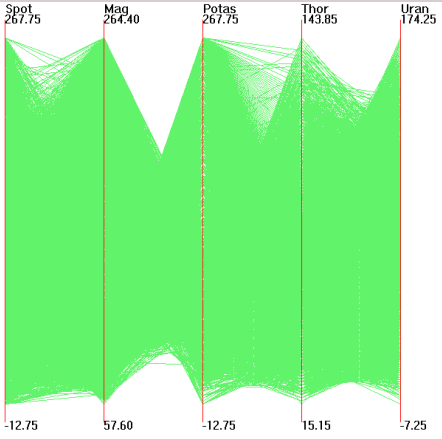


Demo

http://www.mediavirus.org/parvis/

Challenges

Too much data

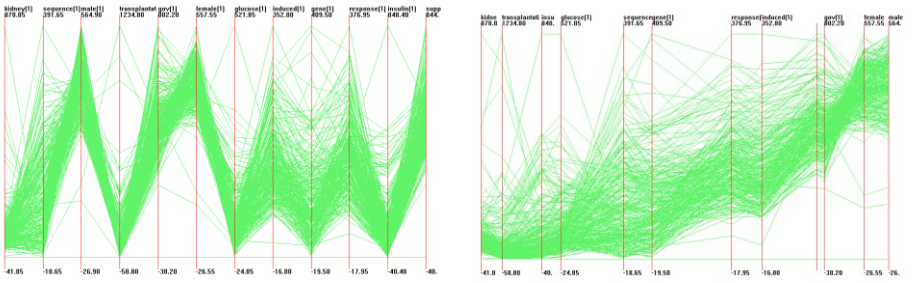


Out5d dataset (5 dimensions, 16384 data items)

(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?

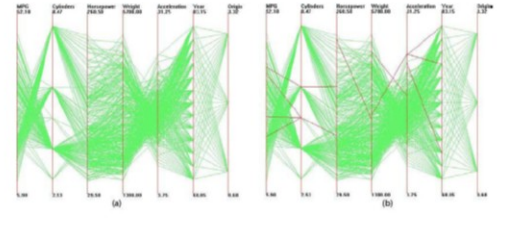


Figure 1: Parallel coordinates visualization of Cars dataset. Outliers are highlighted with red in (b).

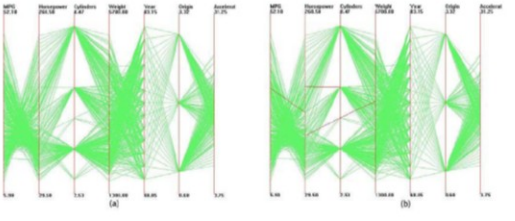


Figure 2: Parallel coordinates visualization of Cars dataset after clutter-based dimension reordering. Outliers are highlighted with red in (b).

Peng et al
InfoVis '04

Reducing Density

Jerding and Stasko, '95, '98
Wegman & Luo, '96

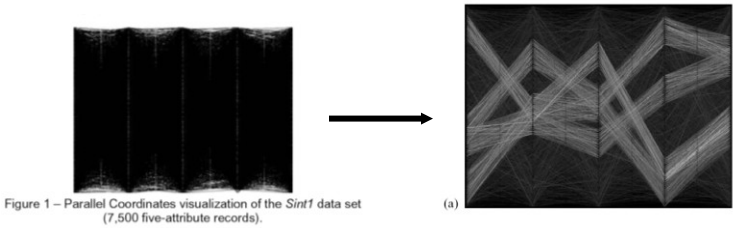
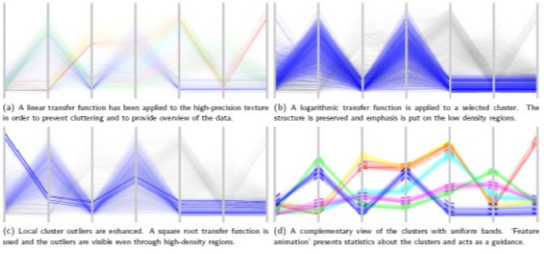


Figure 1 – Parallel Coordinates visualization of the *Sint1* data set (7,500 five-attribute records).

(a)

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.

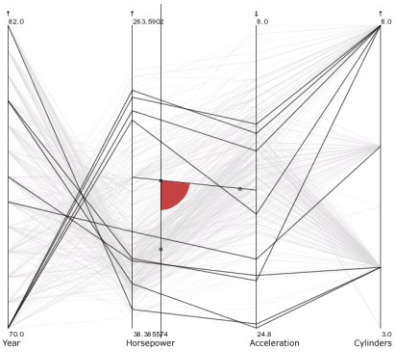
Johansson et al, '05

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

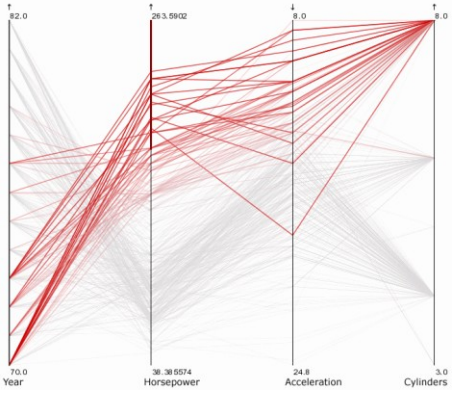


Hauser, Ledermann, & Doleisch
InfoVis '02

(earlier demo)

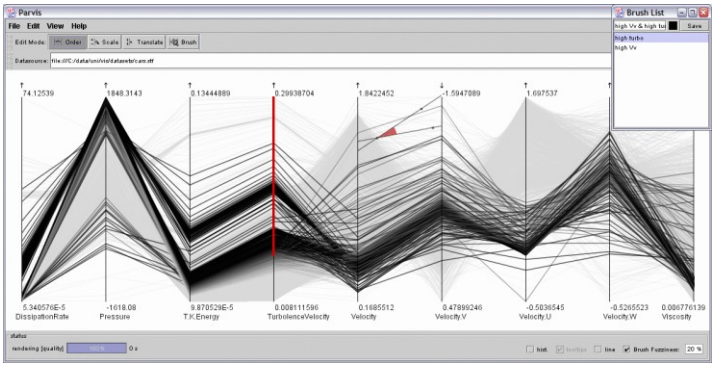
Range Focus

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

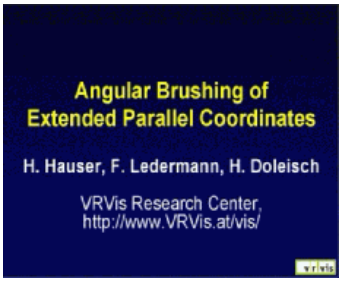


Combining

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

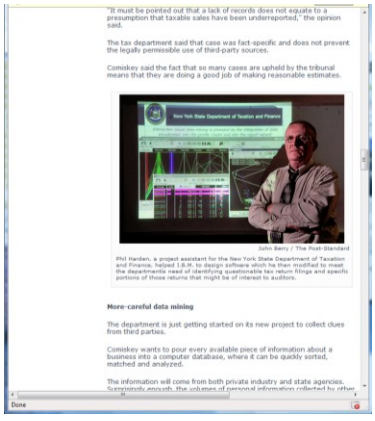


Video



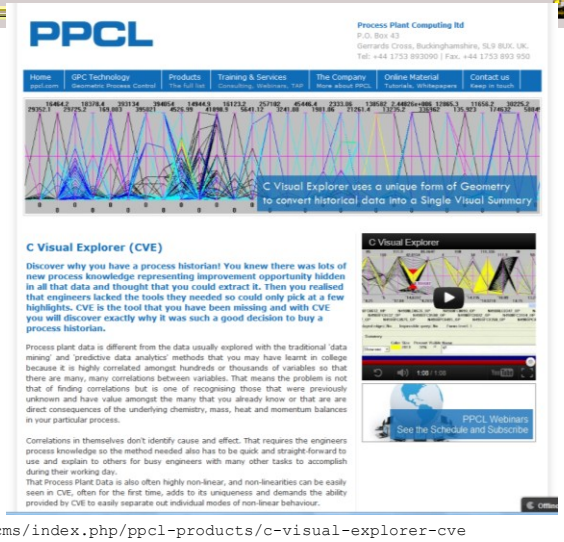
http://www.vrvis.at/via/research/ang-brush/parvis4.mov

Application



http://www.syracuse.com/news/index.ssf/2010/01/data_mining_helps_new_york_cat.html

Application



http://www.ppcl.com/cms/index.php/ppcl-products/c-visual-explorer-cve

Different Kinds of Data

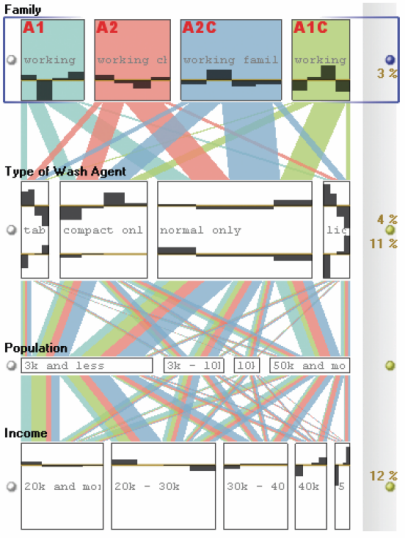
- How about categorical data?
  - Can parallel coordinates handle that well?

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

Representation



Color used for different categories

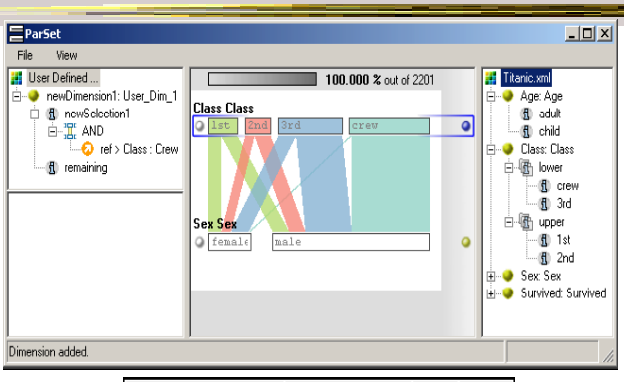
Those values flow into the other variables

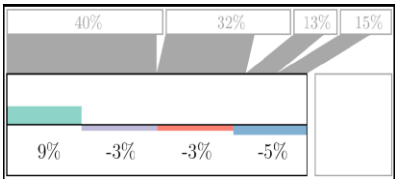
Example

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

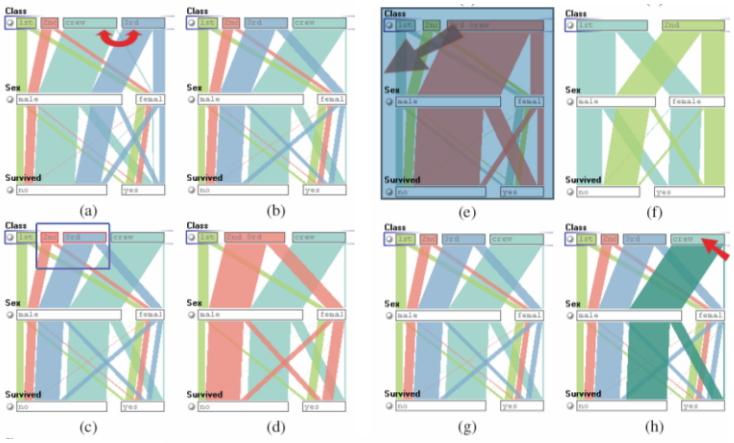
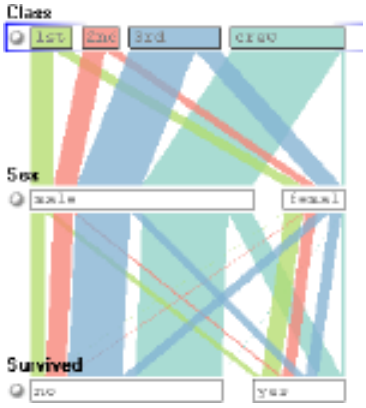


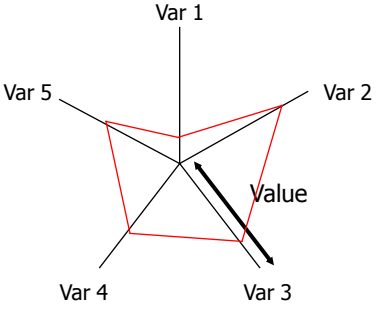
Fig. 7. Basic interaction elements in Parallel Sets: reordering categories (a, b) helps to generate a more meaningful layout; grouping categories (c, d) enables a hierarchical analysis/exploration; excluding categories from the visualization (e, f) allows for interactive filtering; and category highlighting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

Video



InfoVis '05

Star Plots



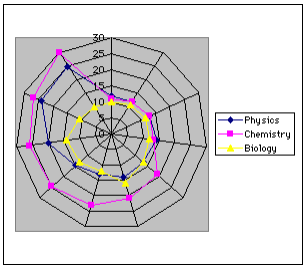
Space out the n variables at equal angles around a circle

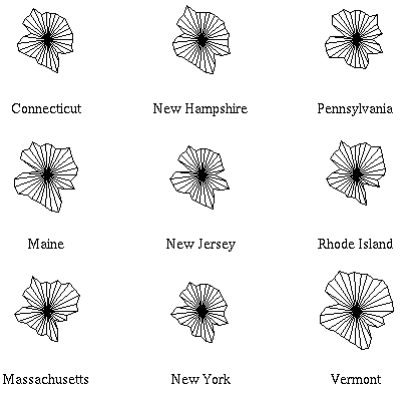
Each "spoke" encodes a variable's value

Alternative Rep.

Data point is now a "shape"

Star Plot examples



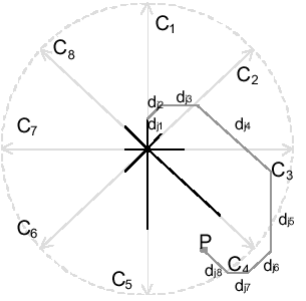


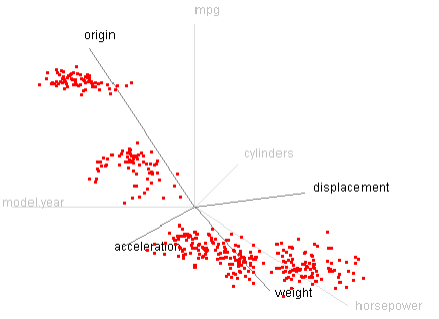
http://seamonkey.ed.asu.edu/~behrens/asu/reports/compre/comp1.html

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

Generalizing the Principles

- General & flexible framework for axis-based visualizations
  - Scatterplots, par coords, etc.
- User can position, orient, and stretch axes
- Axes can be linked

Claessen & van Wijk
TVCG (InfoVis) '11

FLINA View

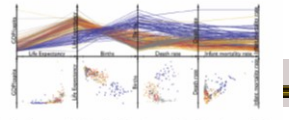


Fig. 6. Demographic data for different countries. Asia: brown; Africa: blue; North America: red; South America: green; Oceania: orange; Europe: gray.

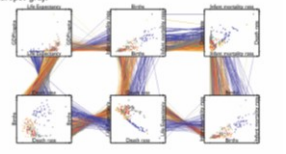
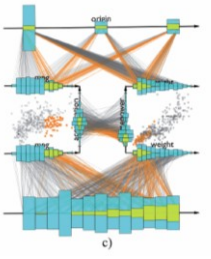
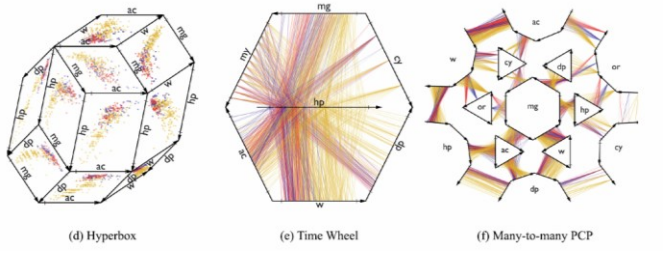


Fig. 7. Alternative lay-out for demographic data



c)



(d) Hyperbox (e) Time Wheel (f) Many-to-many PCP

Video

Parallel Coordinates

- Technique
  - Strengths?
  - Weaknesses?

Design Challenge

year	os	units
2007	Symbian	77.7
2007	RIM	11.8
2007	iPhone	3.3
2007	Windows	14.7
2007	Android	0
2007	Other	14.9
2008	Symbian	72.9
2008	RIM	23.1
2008	iPhone	11.4
2008	Windows	16.5
2008	Android	0.6
2008	Other	15.3
2009	Symbian	80.9
2009	RIM	34.3
2009	iPhone	24.9
2009	Windows	15
2009	Android	6.8
2009	Other	10.4
2010	Symbian	107.7
2010	RIM	46.9
2010	iPhone	41.5
2010	Windows	12.7
2010	Android	47.5
2010	Other	12.6
2011	Symbian	141.3
2011	RIM	62.2
2011	iPhone	70.7
2011	Windows	21.3
2011	Android	91.9
2011	Other	26

Smart Phones sold by OS

Challenge: Help someone understand the competitive landscape in this area

Projections

Source: Gartner

Project

- Teams & Topics due Monday
  - Bring 2 copies
- More topic ideas

Upcoming

- Multivariate Visual Representations 2
  - Reading:
    Keim et al, '02
- User Tasks & Analysis
  - Reading
    Amar & Stasko, '05