Multivariate Visual Representations 1

CS 7450 - Information Visualization
Sep. 14, 2016
John Stasko

Learning Objectives

- For the following visualization techniques/systems, be able to describe each and its visual encoding, know what type of data it's best for, know its strengths and limitations, and understand how to apply it
  - Iconic representatons (Chernoff faces), Table Lens, InfoZoom, Mosaic plot, Attribute Explorer, Parallel Sets, Star plots, Star coordinates
- Explain the visual encoding and design issues of Parallel Coordinates, as well as their utility and limitations
- Understand how the different types of variables in a multivariate data set influence the visualization technique that should be chosen to represent the data
- Be able to apply any of these techniques to a data set that is an appropriate match for them

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 – Hyper/Multivariate data Focus Today

Earlier

- We examined a number of tried-and-true techniques/visualizations for presenting multivariate (typically <=3) data sets
  - Bar graph, line graph, pie chart, scatterplot, box plot, trellis display, crosstab, radar graph, heatmap
- Hinted at how to go above 3 dimensions

Hypervariate Data

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

Design Challenge

- Data set of 500 cases
- Attributes
  - 5 quantitative
  - 4 nominal
  - 2 ordinal
- Design a visualization

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)

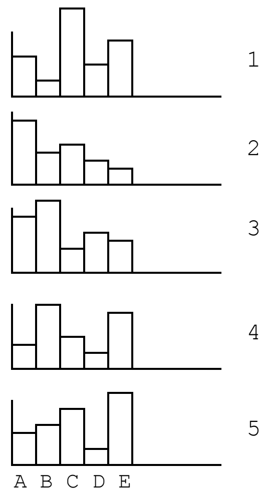
- What about some other techniques?
  - Already seen a couple

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

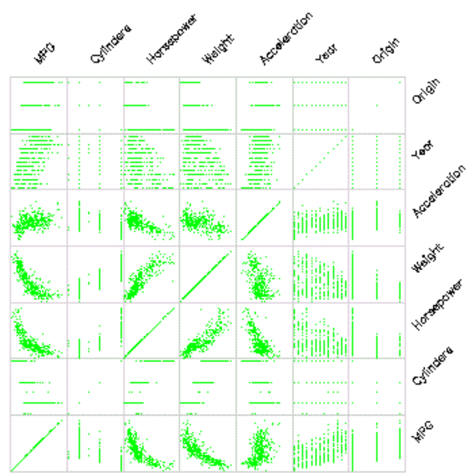


Revisit

Scatterplot Matrix

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

If pairwise correlation is key



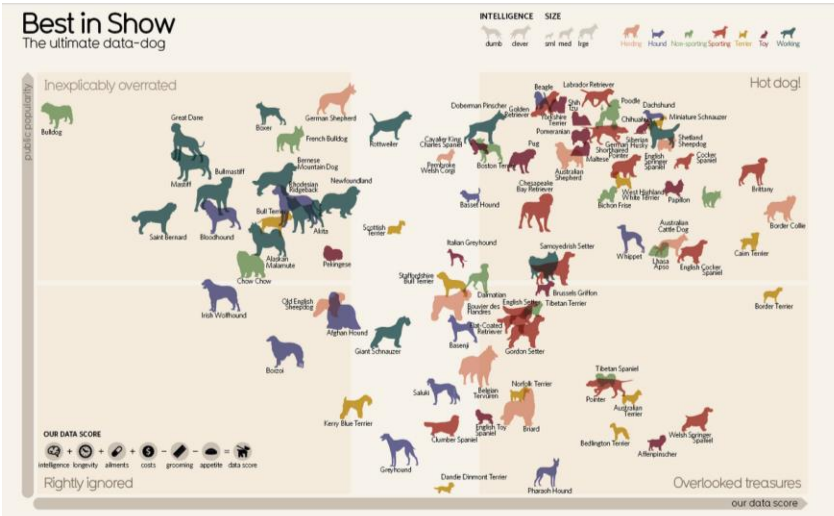
Key Principle (today)

- Handle all data sets generically
  - Examine techniques not specific to some data or domain
  - Technique can generally handle all data sets

Iconic Representations

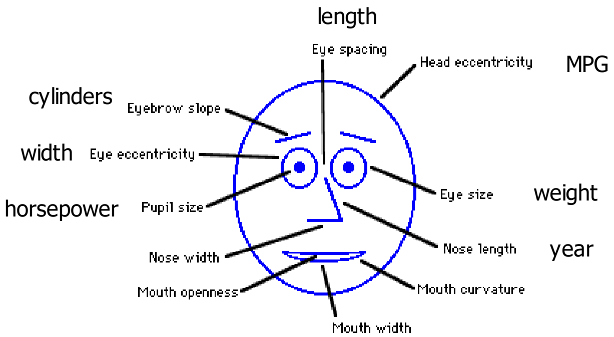
- Glyph (graphical object) represents a data case
- Visual properties of glyph represent different variables

Remember?

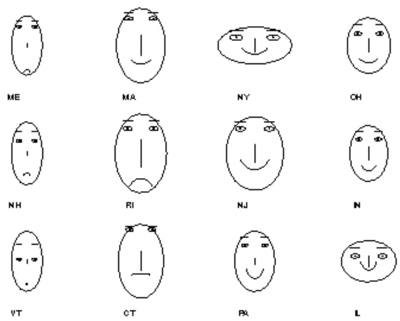


Chernoff Faces

Encode different variables' values in characteristics of human face



Examples



Cute applet: `http://www.cs.uchicago.edu/~wiseman/chernoff/`

Table Lens

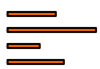
- Spreadsheet is certainly one hypervariate data presentation
- Idea: Make the text more visual and symbolic
- Just leverage basic bar chart idea

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

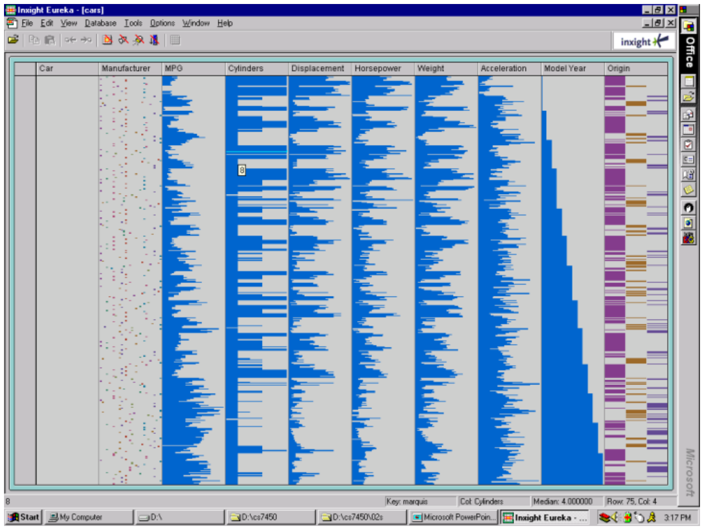
a1-cereals [Read-Only] [Compatibility Mode]

	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

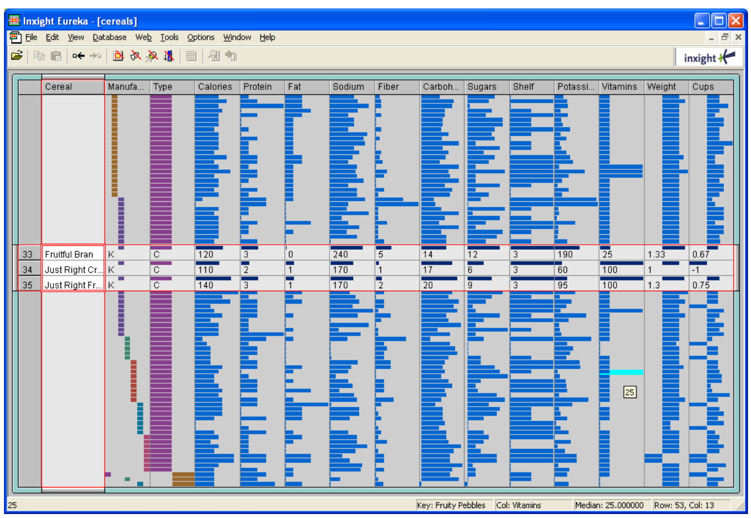
What do you do for nominal data?

Instantiation

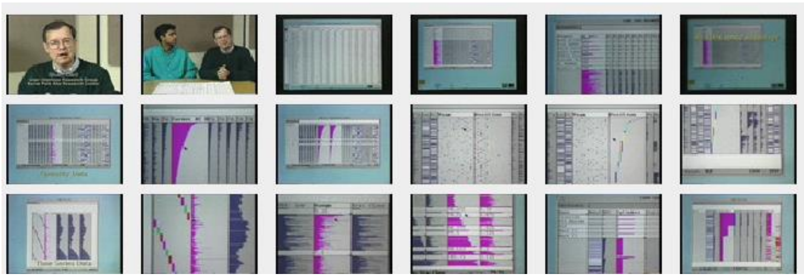


Details

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

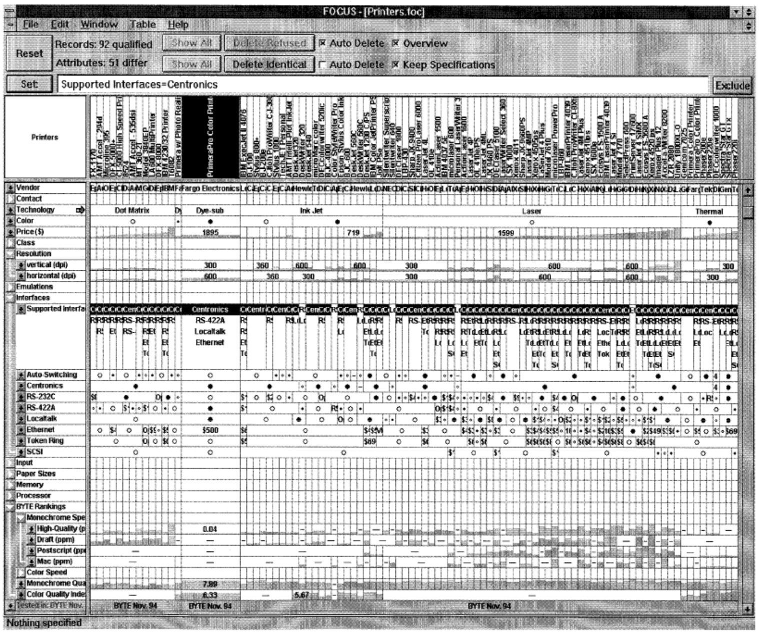


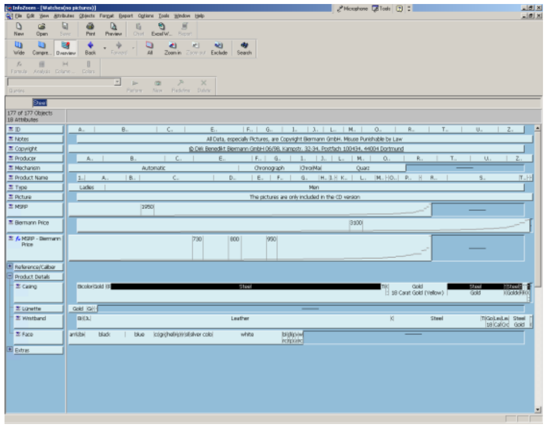
Figure 1: An overview of the printer table.

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

Manifestation

InfoZoom



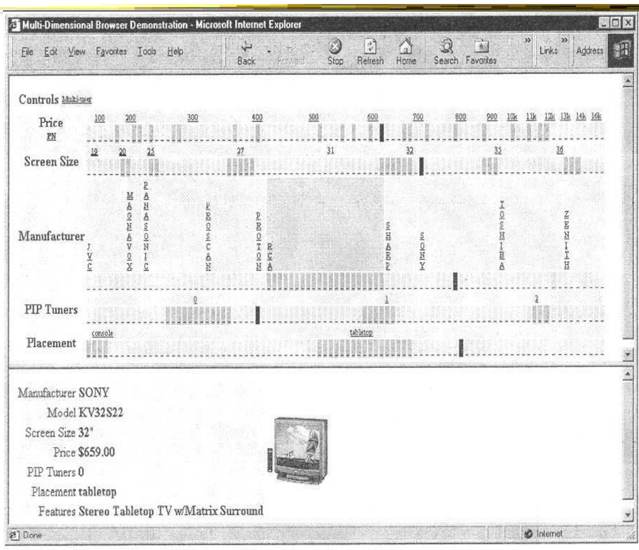
Commercial product
Demo/video

MultiNav

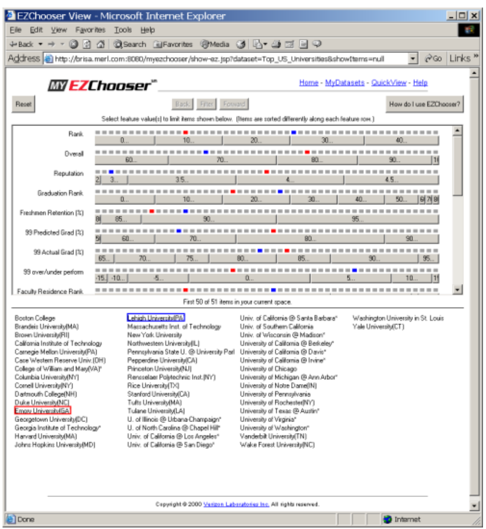
- Each different attribute is placed in a different row
- Sort the values of each row
  - Thus, a particular item is not just in one column
- Want to support browsing

Lanning et al
AVI '00

Interface



Instantiation

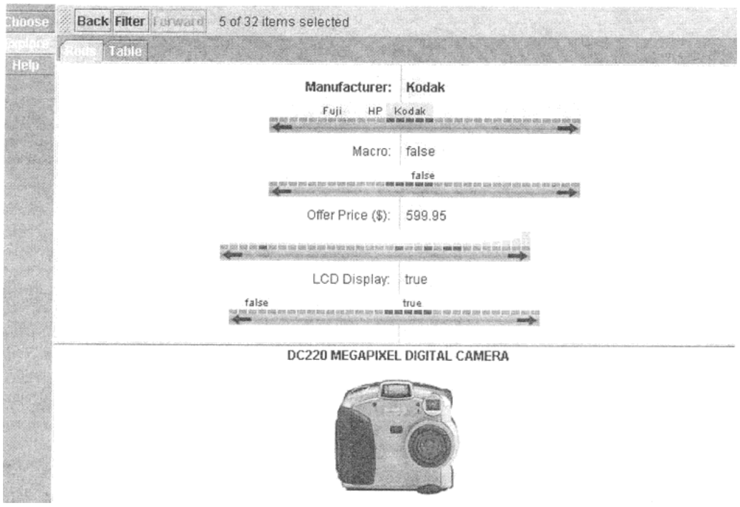


Video

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



Limitations

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

An Application

- What if you cared about ranking items?
  - Think of the attributes per item as contributing to some score or value for it
- Apply the representations we've seen earlier

LineUp

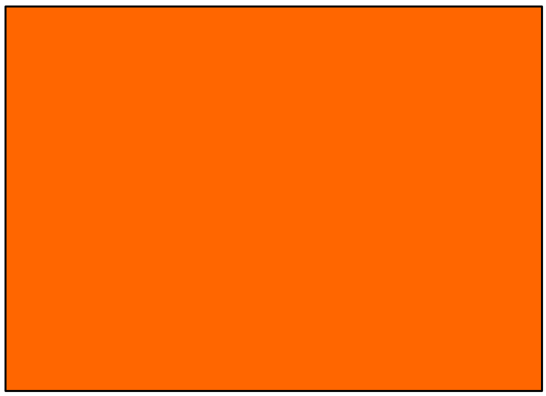
Video

Gratzl et al
TVCG (InfoVis) '13

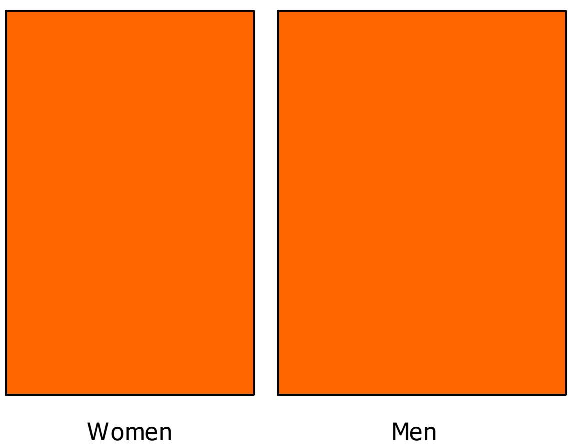
Categorical data?

- 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, …

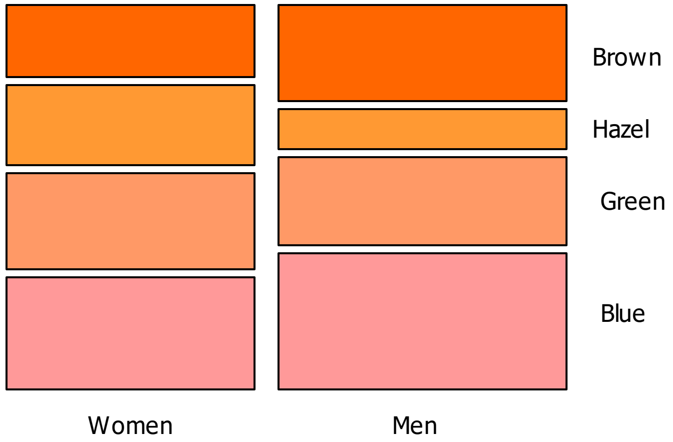
Mosaic Plot



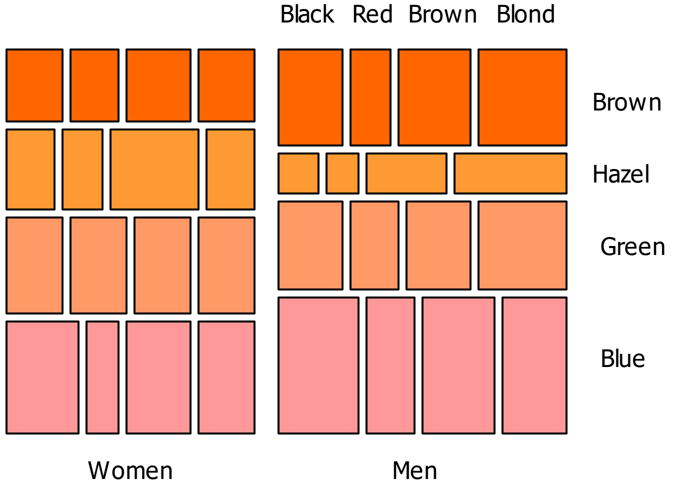
Mosaic Plot



Mosaic Plot

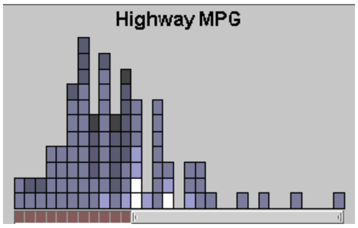


Mosaic Plot



Attribute Explorer

- General hypervariate data representation combined with flexible interaction



Spence & Tweedie
Inter w Computers '98

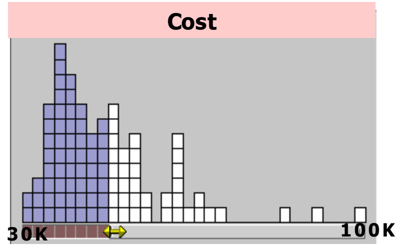
Characteristics

- 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

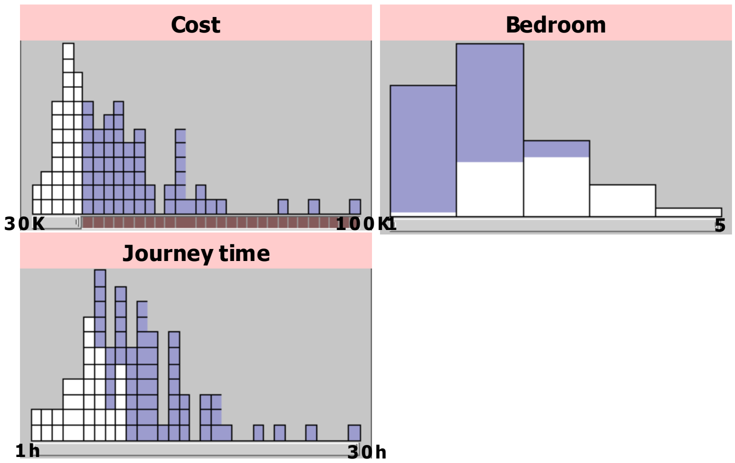
- Attribute histogram
- All objects on all attribute scales

- Interaction with attributes limits



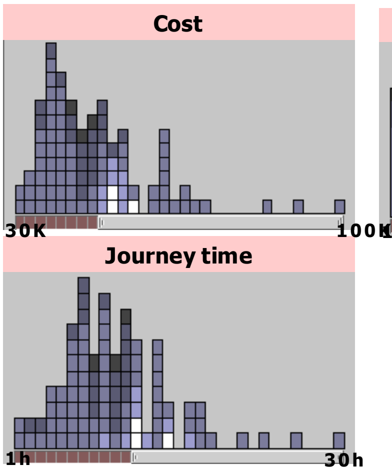
Features

- Inter-relations between attributes – brushing



Features

- Color-encoded sensitivity



Attribute Explorer

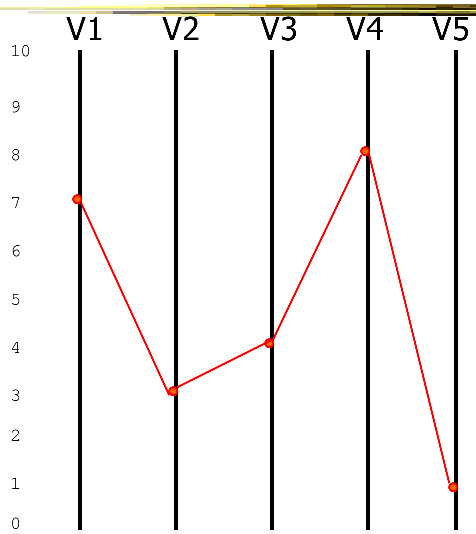
Video

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

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

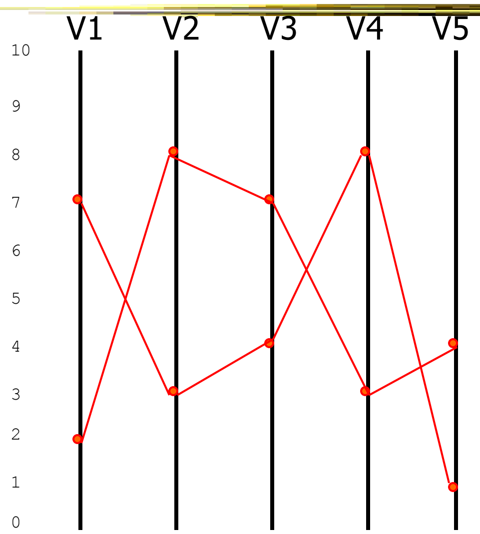
Parallel Coordinates



D1

7 3 4 8 1

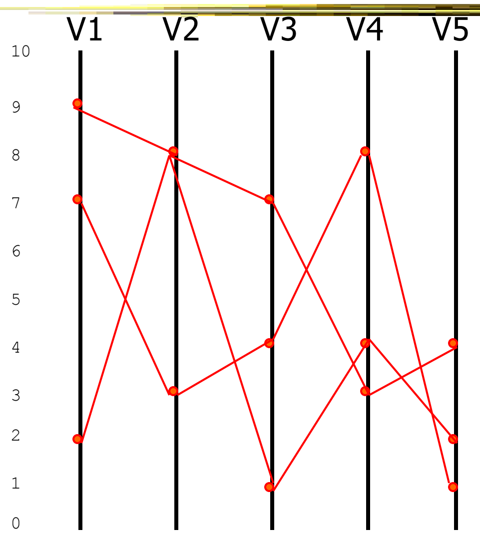
Parallel Coordinates



D2

2 7 6 3 4

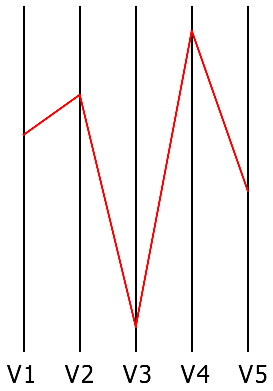
Parallel Coordinates



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

Issue

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

"Multidimensional Detective"
A. Inselberg, InfoVis '97

Application

- 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

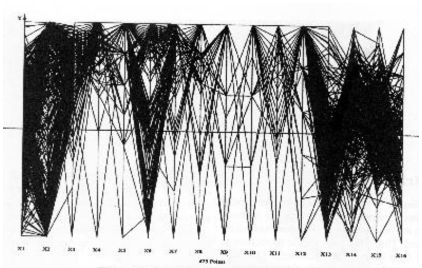
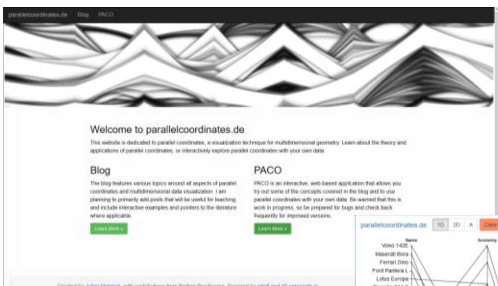
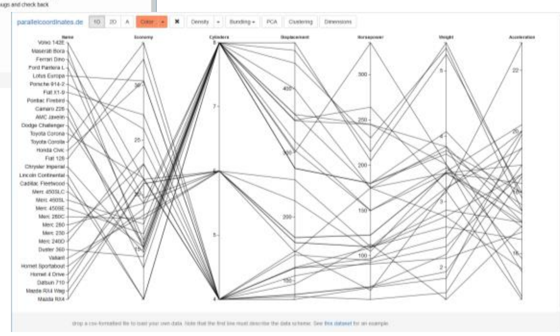


Figure 1: The full dataset consisting of 473 batches

To Learn More

Great site that's all about parallel coords

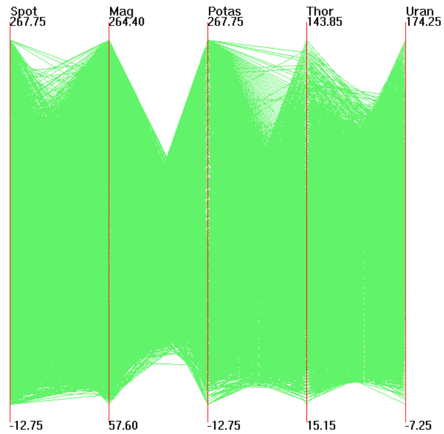




http://www.parallelcoordinates.de

Challenges

Too much data



Out5d dataset (5 dimensions, 16384 data items)

(courtesy of J. Yang)

Reducing Density

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

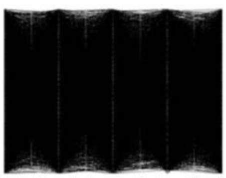
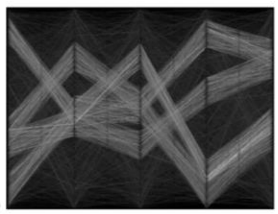
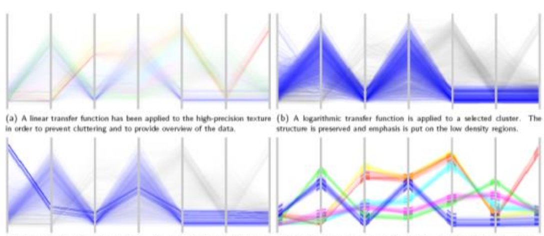


Figure 1 – Parallel Coordinates visualization of the *Sint1* data set (7,500 five-attribute records).



(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

Dimensional Reordering

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

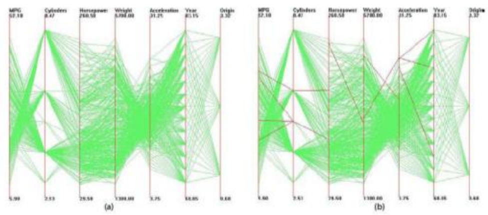


Figure 1: Parallel coordinates visualization of Cars dataset. Outliers are highlighted with red in (b).

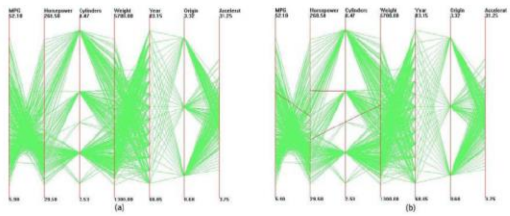
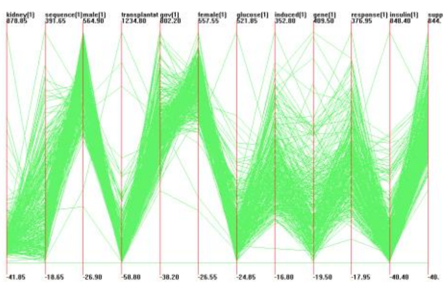


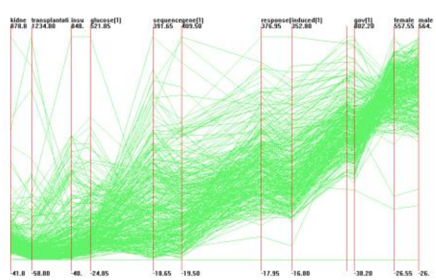
Figure 2: Parallel coordinates visualization of Cars dataset after clutter-based dimension reordering. Outliers are highlighted with red in (b).

Peng et al
InfoVis '04

Dimensional Reordering

Which dimensions are most like each other?





Same dimensions ordered according to similarity

Yang et al
InfoVis '03

Different Kinds of Data

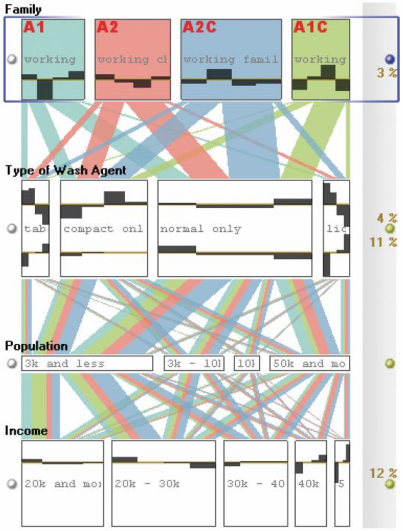
- How about categorical data?
  - Can parallel coordinates handle that well?

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

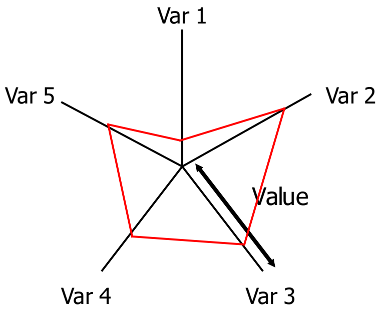
Representation



Color used for different categories

Those values flow into the other variables

Star Plots (Radar Chart)



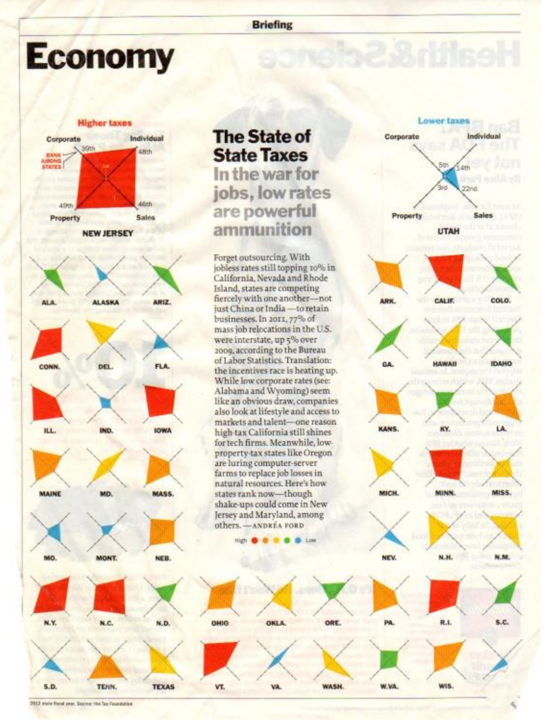
Space out the n variables at equal angles around a circle

Each "spoke" encodes a variable's value

Alternative Rep.

Data point is now a "shape"

Example


Briefing

Economy

The State of State Taxes

In the war for jobs, low rates are powerful ammunition

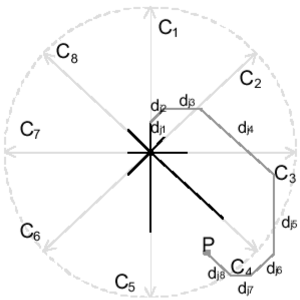
Forget outsourcing. With jobless rates still topping 10% in California, Nevada and Rhode Island, states are competing fiercely with one another—not just China or India—to retain businesses. In 2011, 77% of mass job relocations in the U.S. were interstate, up 5% over 2009, according to the Bureau of Labor Statistics. Translation: the incentives race is heating up. While low corporate rates (see: Alabama and Wyoming) seem like an obvious draw, companies also look at lifestyle and access to markets and talent—one reason high-tax California still shines for tech firms. Meanwhile, low-property-tax states like Oregon are luring computer-server farms to replace job losses in natural resources. Here's how states rank now—though shake-ups could come in New Jersey and Maryland, among others. —ANDRÉA FORD

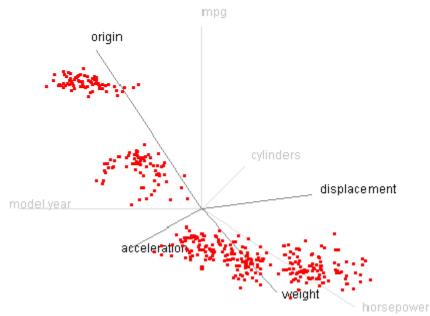
Time
April 16, 2012

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
Late-Breaking Hot Topics, InfoVis '00

Demo

Star Coordinates

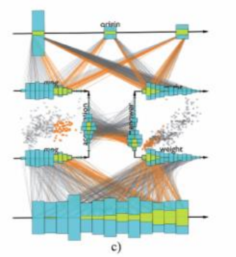
- Data cases with similar values will lead to clusters of points
- (What's the problem though?)
- Multi-dimensional scaling or projection down to 2D

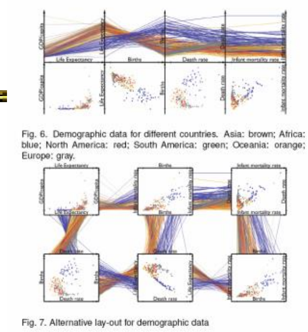
Generalizing the Principles

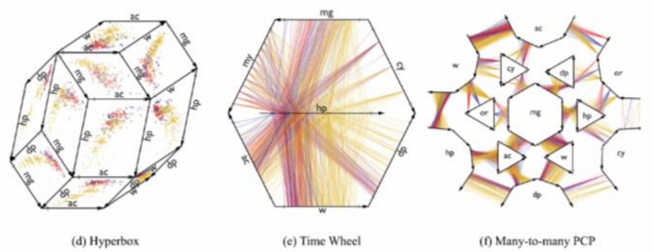
- General & flexible framework for axis-based visualizations
  - Scatterplots, par coords, etc.
- User can position, orient, and stretch axes
- Axes can be linked

Claessen & van Wijk
TVCG (InfoVis) '11

FLINA View







Video

Learning Objectives

- For the following visualization techniques/systems, be able to describe each and its visual encoding, know what type of data it's best for, know its strengths and limitations, and understand how to apply it
  - Iconic representatons (Chernoff faces), Table Lens, InfoZoom, Mosaic plot, Attribute Explorer, Parallel Sets, Star plots, Star coordinates
- Explain the visual encoding and design issues of Parallel Coordinates, as well as their utility and limitations
- Understand how the different types of variables in a multivariate data set influence the visualization technique that should be chosen to represent the data
- Be able to apply any of these techniques to a data set that is an appropriate match for them

Readings

- Inselberg, InfoVis '97 paper
- Browse Heinrich http://www.parallelcoordinates.de website (try out demo)

Reminder

- Processing tutorial session
  - Thursday 11-12 in GVU Café
- HW 3 due a week from today

Upcoming

- Multivariate Visual Representations 2
- InfoVis Systems & Toolkits