

# Multivariate Visual Representations 1



CS 7450 - Information Visualization  
Aug. 31, 2015  
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?

Fall 2015

CS 7450

3

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**

Fall 2015

CS 7450

4

## Earlier



- We examined a number of tried-and-true techniques/visualizations for presenting multivariate (typically  $\leq 3$ ) 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

## 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- $\rightarrow$ 2D projections

Fall 2015

CS 7450

7

## 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

Fall 2015

CS 7450

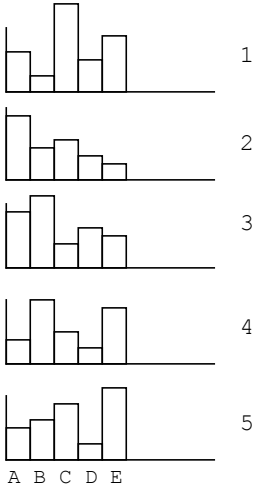
8

# Multiple Views



Give each variable its own display

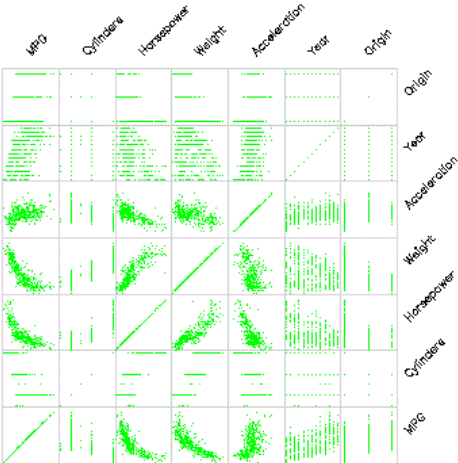
	A	B	C	D	E
1	4	1	8	3	5
2	6	3	4	2	1
3	5	7	2	4	3
4	2	6	3	1	5
5	3	4	5	1	7



# Scatterplot Matrix



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



# Key Principle (today)

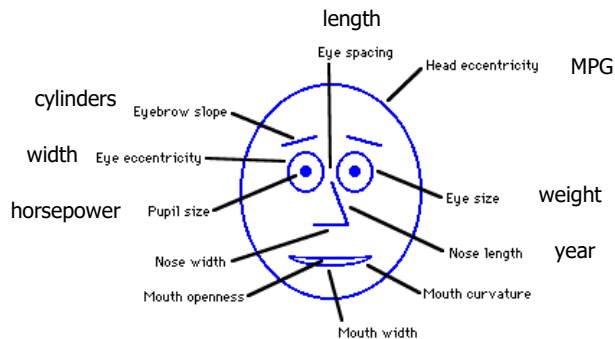


- Handle all data sets generically
  - Examine techniques not specific to some data or domain
  - Technique can generally handle all data sets

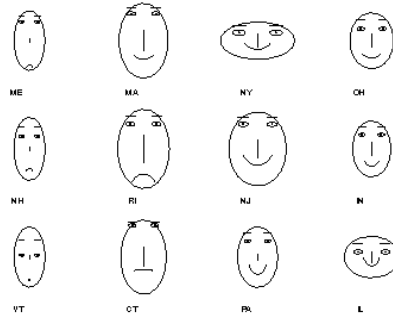
# Chernoff Faces



Encode different variables' values in characteristics of human face



# Examples



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

Fall 2015

CS 7450

13

# 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

Fall 2015

CS 7450

14

# 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

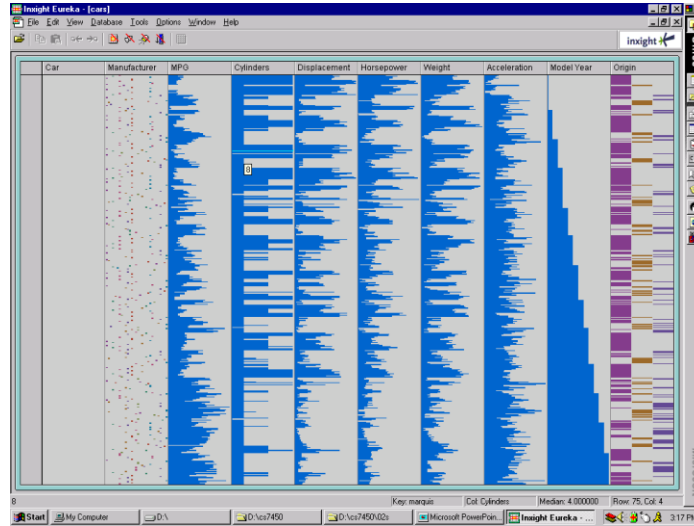


	A	B	C	D	E	F	G	H	I
1	Cereal	Manufacture	Type	Calories	Protein	Fat	Sodium	Fiber	Carbol
2	Frosted Mini-Wheats	K	C	100	3	0	0	3	
3	Raisin Squares	K	C	90	2	0	0	2	
4	Shredded Wheat	N	C	80	2	0	0	3	
5	Shredded Wheat 'n Bran	N	C	90	3	0	0	4	
6	Shredded Wheat spoon s	N	C	90	3	0	0	3	
7	Puffed Rice	Q	C	50	1	0	0	0	
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	C	150	4	3	95	3	
19	Anna Marie	K	C	110	2	0	125	1	

What do you do for nominal data?



# Instantiation



Fall 2015

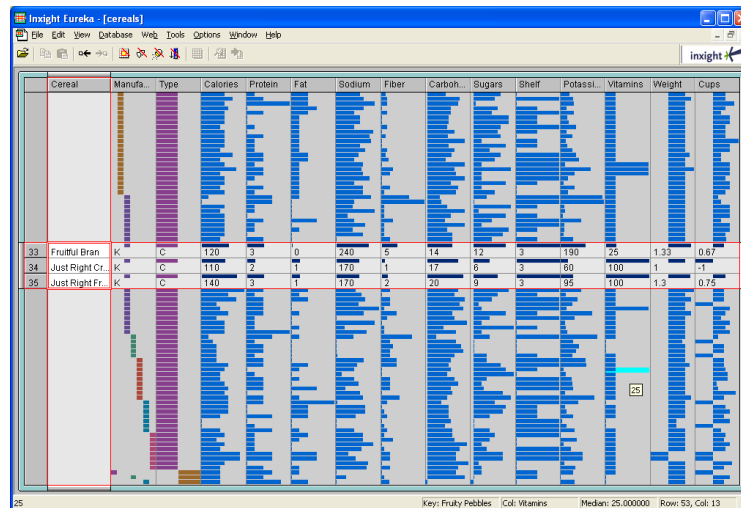
CS 7450

17

# Details



Focus on item(s) while showing the context

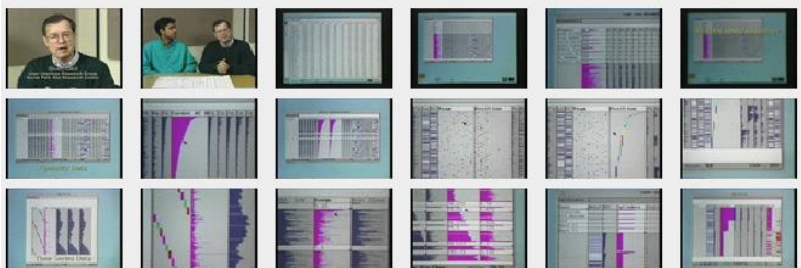


Fall 2015

CS 7450

18

# See It



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

Video

Fall 2015

CS 7450

19

# 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

Fall 2015

CS 7450

20

FOCUS - (Printers.tac)

Records: 92 qualified  
Attributes: 51 differ

Supported Interfaces=Centronics

Printer	Vendor	Technology	Color	Price (\$)	Class	Resolution	Vertical (dpi)	Horizontal (dpi)	Emulations	Interfaces	Auto Switching	Centronics	LocalTalk	Ethernet	Other
HP LaserJet 4000	HP	Laser	•	1599	•	600	600	600	•	•	•	•	•	•	•
HP LaserJet 4050	HP	Laser	•	1599	•	600	600	600	•	•	•	•	•	•	•
HP LaserJet 4100	HP	Laser	•	1599	•	600	600	600	•	•	•	•	•	•	•

Nothing specified



Figure 1: An overview of the printer table.

## Characteristics

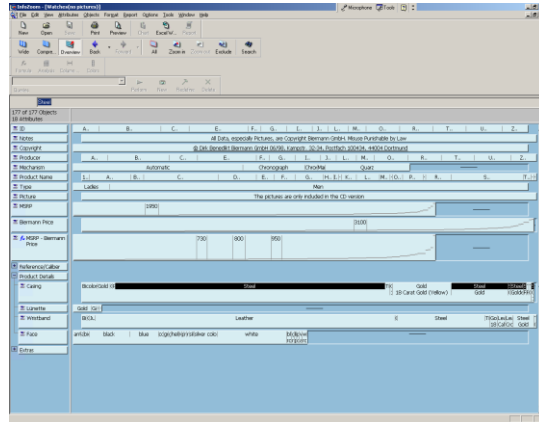
- Can sort on any attribute (row)
- Focus on an attribute value (show only cases having that value) by double-clicking on it
- Can type in queries on different attributes to limit what is presented too



# Manifestation



InfoZoom



Commercial product to be demo'ed coming up

Fall 2015

CS 7450

23

# 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

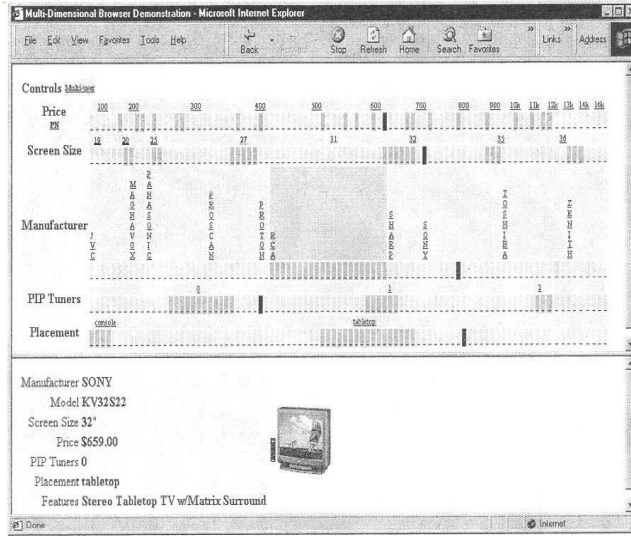
Lanning et al  
AVI '00

Fall 2015

CS 7450

24

# Interface

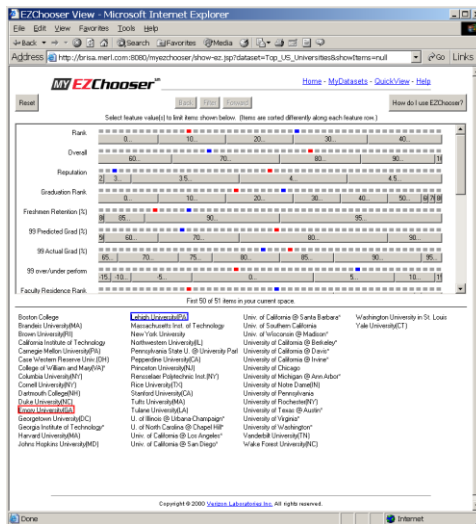


Fall 2015

CS 7450

25

# Instantiation



Fall 2015

CS 7450

Demo

26

# Alternate UI



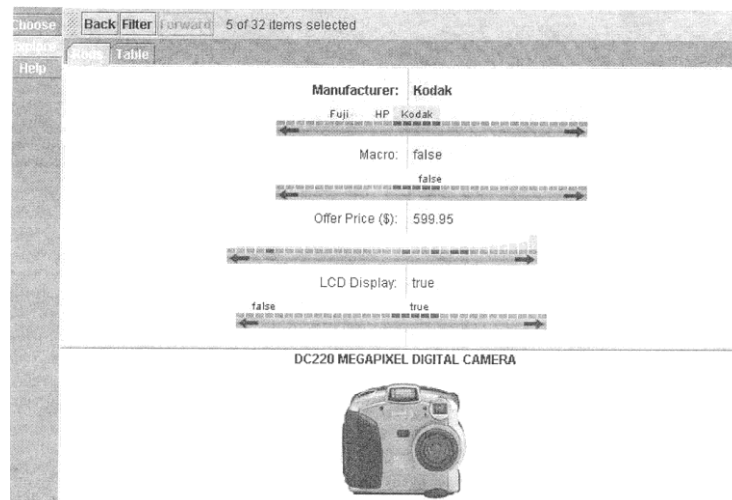
- Can slide the values in a row horizontally
- A particular data case then can be lined up in one column, but the rows are pushed unequally left and right

Fall 2015

CS 7450

27

# Attributes as Sliding Rods



Fall 2015

CS 7450

28

# 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



Fall 2015 CS 7450 Gratzl et al TVCG (InfoVis) '13 31

## 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, ...

Fall 2015 CS 7450 32



# Mosaic Plot

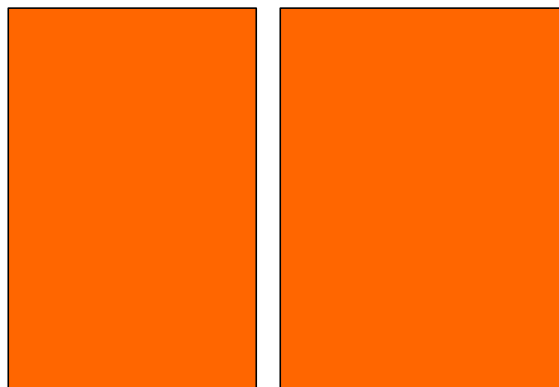


Fall 2015

CS 7450

33

# Mosaic Plot



Women

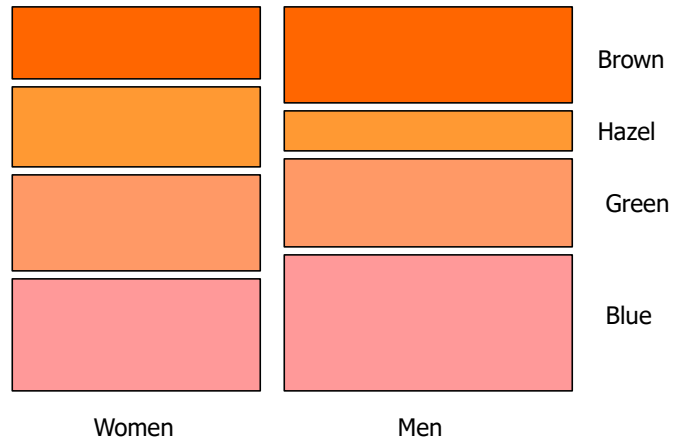
Men

Fall 2015

CS 7450

34

# Mosaic Plot

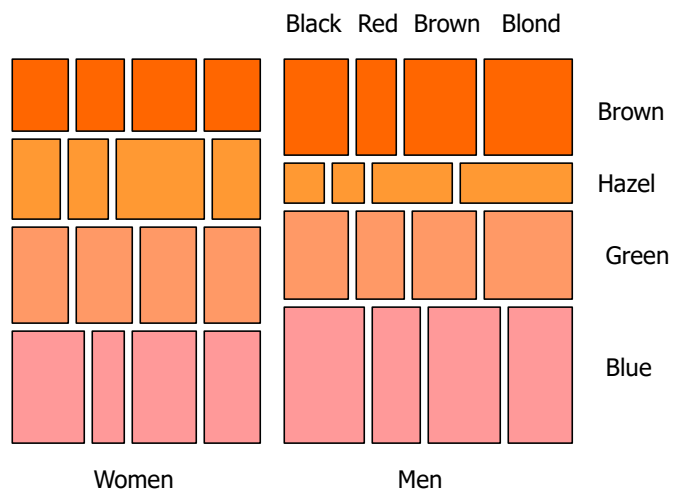


Fall 2015

CS 7450

35

# Mosaic Plot



Fall 2015

CS 7450

36

# Attribute Explorer



- General hypervariate data representation combined with flexible interaction

Spence & Tweedie  
*Inter w Computers '98*

Fall 2015

CS 7450

37

# 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)

Fall 2015

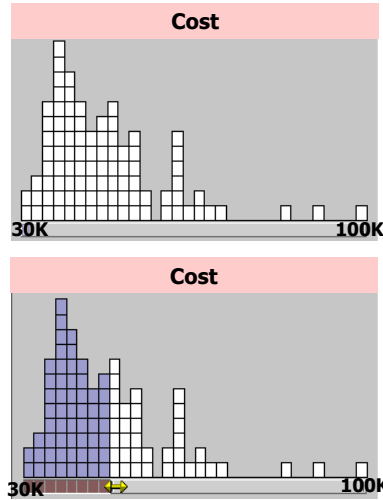
CS 7450

38

# Features



- Attribute histogram
- All objects on all attribute scales
  
- Interaction with attributes limits



Fall 2015

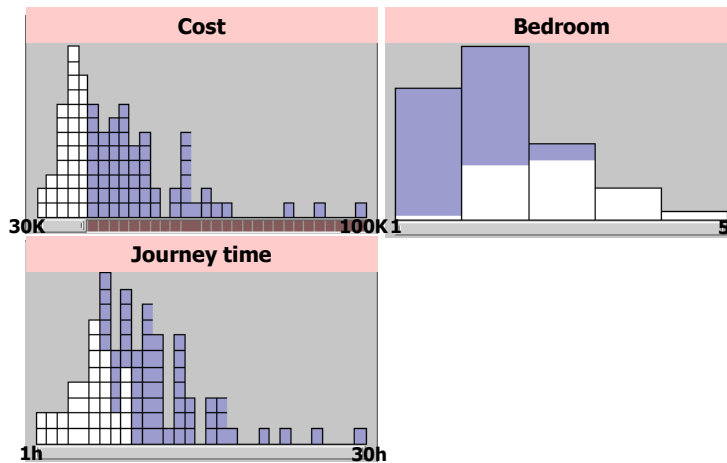
CS 7450

39

# Features



- Inter-relations between attributes – brushing



Fall 2015

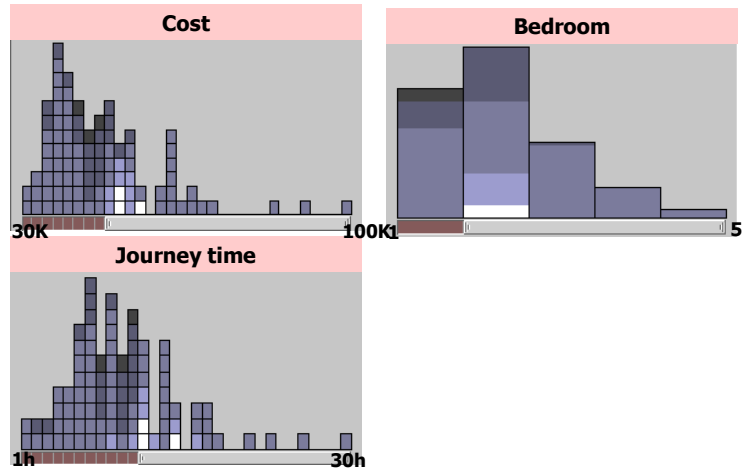
CS 7450

40

# Features



- Color-encoded sensitivity



Fall 2015

CS 7450

41

# Attribute Explorer



Video

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

Fall 2015

CS 7450

42

# 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

# Parallel Coordinates



- What are they?
  - Explain...

# Parallel Coordinates



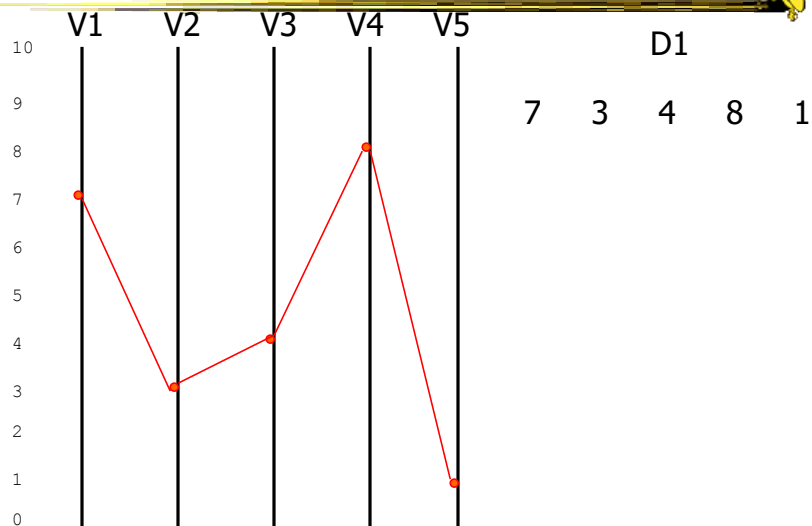
	V1	V2	V3	V4	V5
D1	7	3	4	8	1
D2	2	7	6	3	4
D3	9	8	1	4	2

Fall 2015

CS 7450

45

# Parallel Coordinates

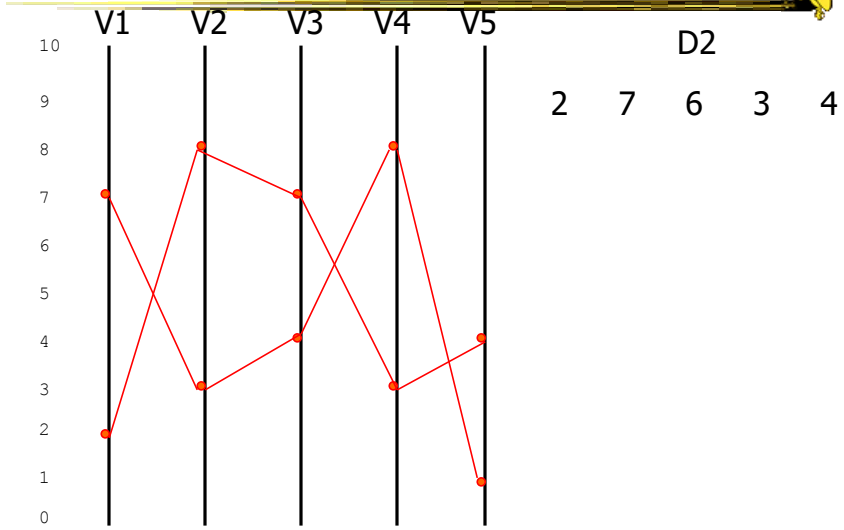


Fall 2015

CS 7450

46

# Parallel Coordinates



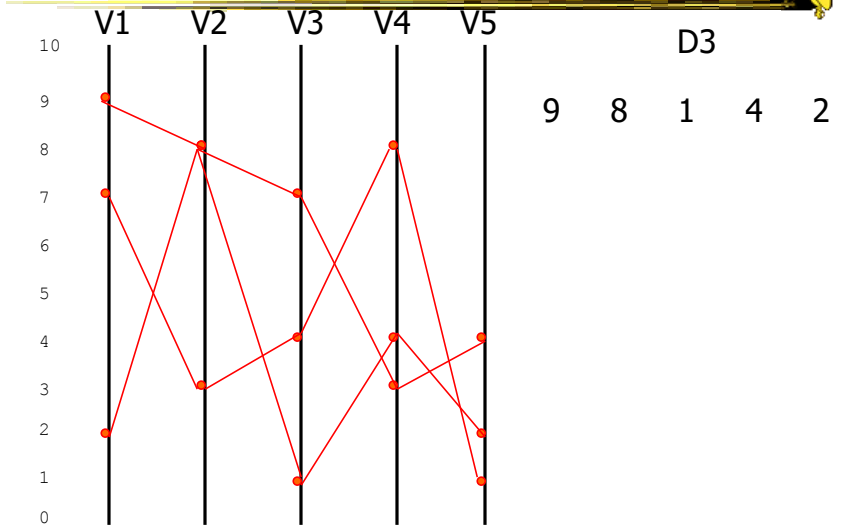
D2  
2 7 6 3 4

Fall 2015

CS 7450

47

# Parallel Coordinates



D3  
9 8 1 4 2

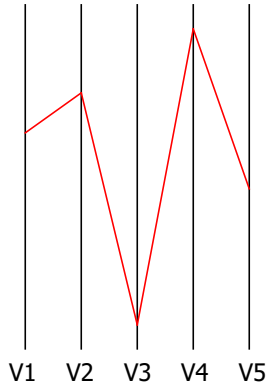
Fall 2015

CS 7450

48



# Parallel Coordinates



Encode variables along a horizontal row

Vertical line specifies different values that variable can take

Data point represented as a polyline

Fall 2015

CS 7450

49

## Issue



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

Fall 2015

CS 7450

50

# Application



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

Taken from:  
"Multidimensional Detective"  
A. Inselberg, InfoVis '97

Fall 2015

CS 7450

51

# The Problem



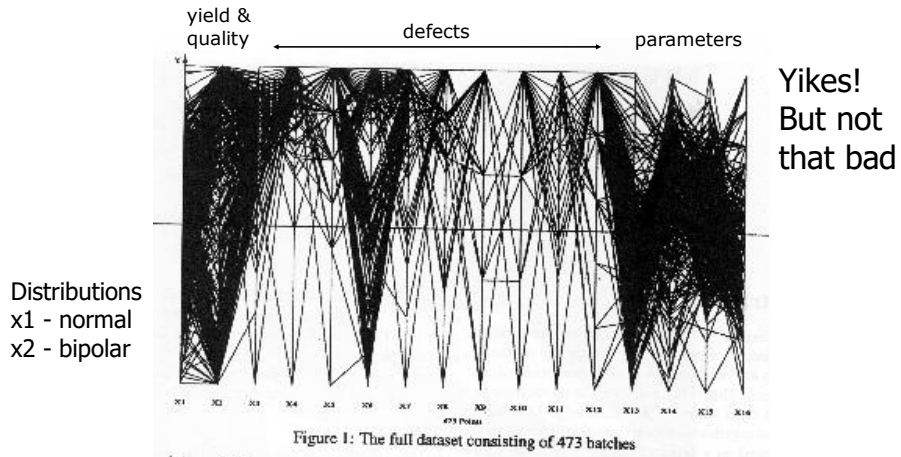
- VLSI chip manufacture
- Want high quality chips (high speed) and a high yield batch (% of useful chips)
- Able to track defects
- Hypothesis: No defects gives desired chip types
- 473 batches of data

Fall 2015

CS 7450

52

# Parallel Coordinate Display

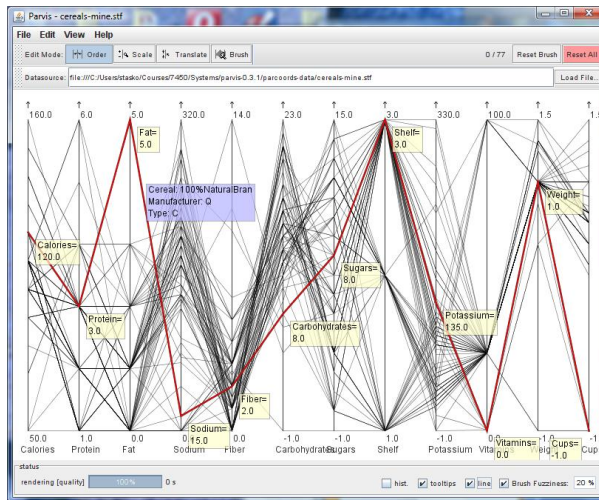


Fall 2015

CS 7450

53

# ParVis System



Demo

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

Fall 2015

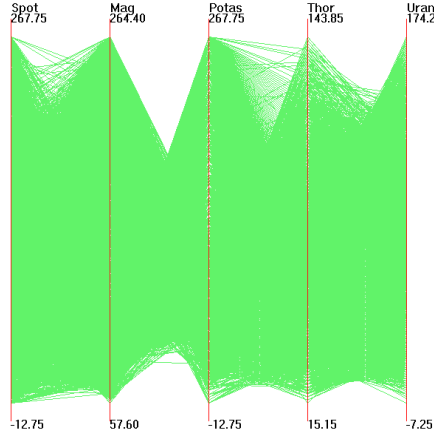
CS 7450

54

# Challenges



Too much data



Out5d dataset (5 dimensions, 16384 data items)

Fall 2015

CS 7450

(courtesy of J. Yang)

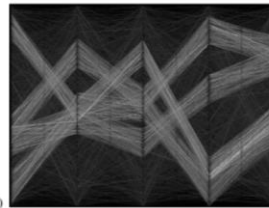
55

# Reducing Density

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

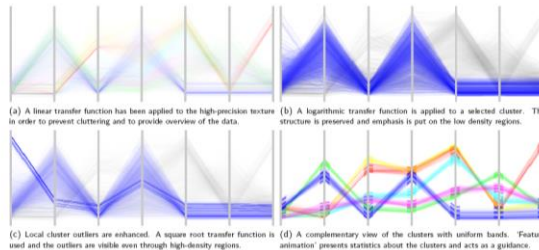


Figure 1 – Parallel Coordinates visualization of the *Sirt1* data set (7,500 five-attribute records).



(a)

Artero et al, '04



(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

Fall 2015

CS 7450

56

# Dimensional Reordering



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

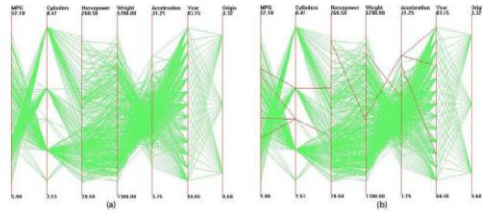


Figure 1: Parallel coordinates visualization of Cars dataset. Outliers are highlighted with red in (b).

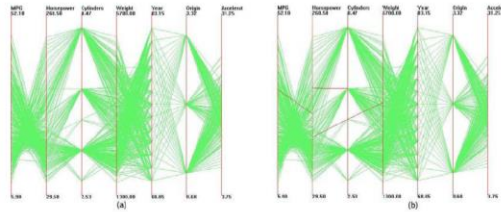


Figure 2: Parallel coordinates visualization of Cars dataset after clutter-based dimension reordering. Outliers are highlighted with red in (b).

Peng et al  
InfoVis '04

Fall 2015

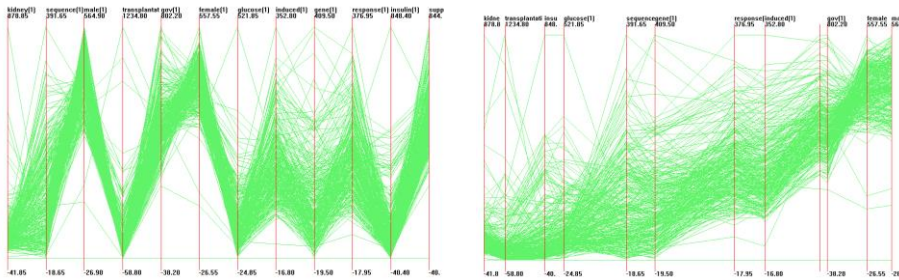
CS 7450

57

# Dimensional Reordering



Which dimensions are most like each other?



Same dimensions ordered according to similarity

Yang et al  
InfoVis '03

Fall 2015

CS 7450

58

# Different Kinds of Data



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

Fall 2015

CS 7450

59

# 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

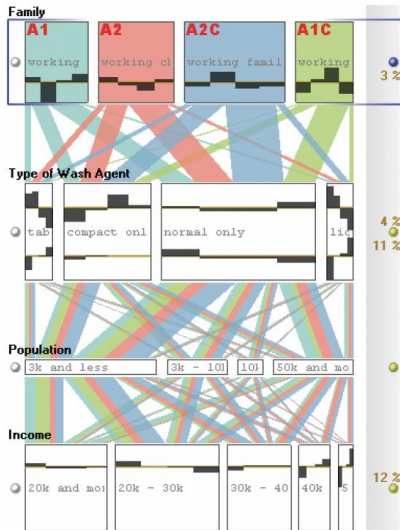
Kosara, Bendix, & Hauser  
*TVCG*'05

Fall 2015

CS 7450

60

# Representation



Color used for different categories

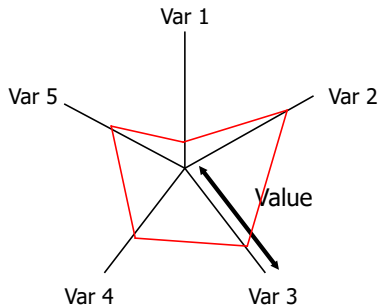
Those values flow into the other variables

Fall 2015

CS 7450

61

# 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"

Fall 2015

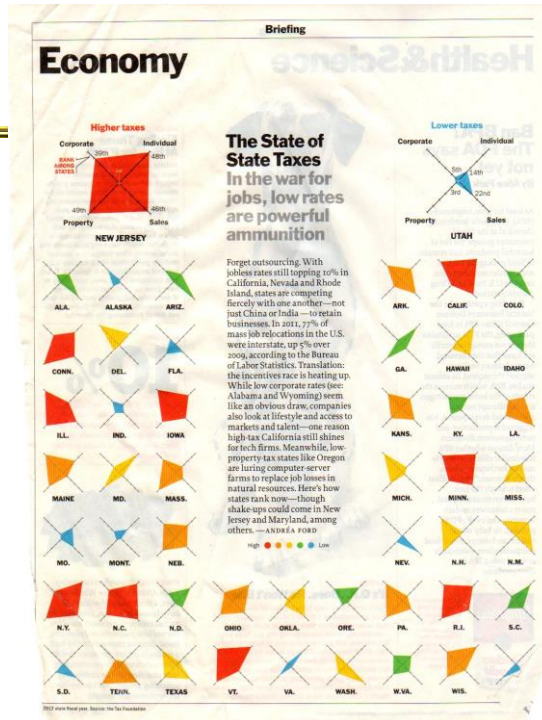
CS 7450

62

# Example

Time  
April 16, 2012

Fall 2015



## 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

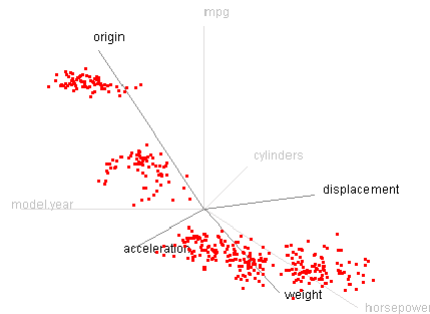
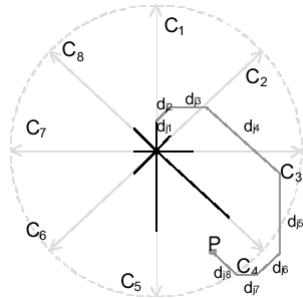
Fall 2015

CS 7450

64



# Star Coordinates



E. Kandogan  
Late-Breaking Hot Topics, InfoVis '00

Demo

Fall 2015

CS 7450

65

# 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

Fall 2015

CS 7450

66

# 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

Fall 2015

CS 7450

67

## FLINA View

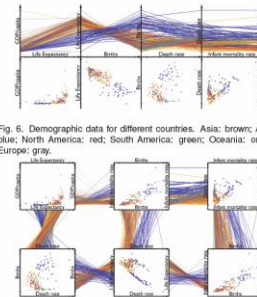
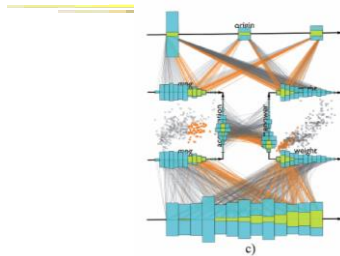
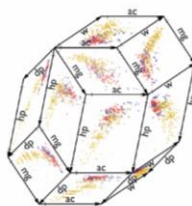
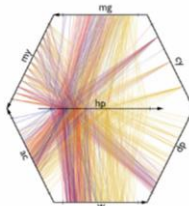


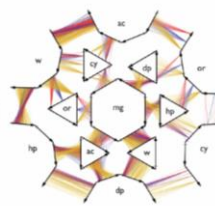
Fig. 7. Alternative lay-out for demographic data



(d) Hyperbox



(e) Time Wheel



(f) Many-to-many PCP

Video

Fall 2015

CS 7450

68

# Parallel Coordinates



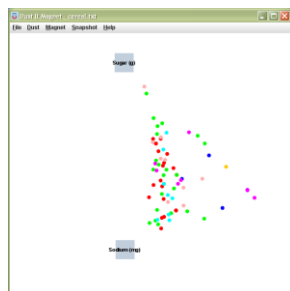
- Technique
  - Strengths?
  - Weaknesses?

Fall 2015

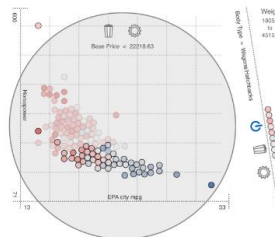
CS 7450

69

# More to Come



Dust and Magnet



Kinetica

and others...

On Interaction day

Fall 2015

CS 7450

70

# 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

Fall 2015

CS 7450

71

# Project



- Teams & Topics due 14th
  - Bring 3 copies
- Next time
  - Topic ideas
  - Help with team formation
  - On t-square wiki

Fall 2015

CS 7450

72

# Upcoming



- Multivariate Visual Representations 2
  - Reading:  
Keim et al, '02
- Labor Day holiday
  
- Visualization Programming Tutorial
  - Reading  
Murray online book