

# Multivariate Visual Representations 1



CS 4460 – Intro. to Information Visualization  
Sep. 11, 2014  
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 2014

CS 4460

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 2014

CS 4460

4

# Earlier



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

Fall 2014

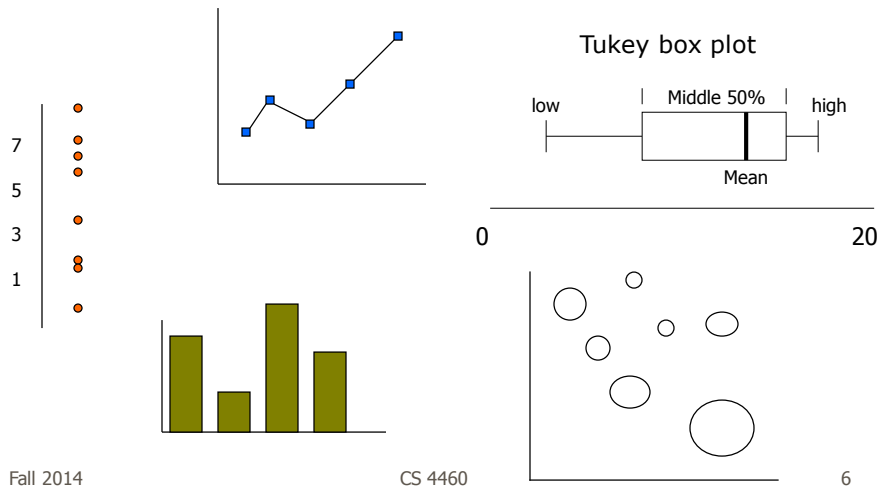
CS 4460

5

# Representations



Some standard ways for low-d data



Fall 2014

CS 4460

6

# Hypervariate Data



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

# 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

# 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

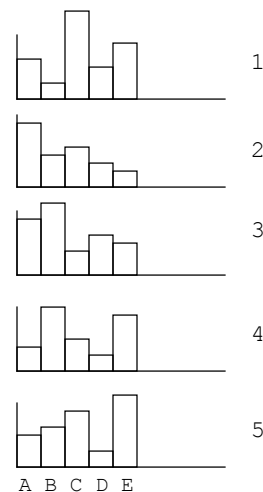
Revisit

# Multiple Views



Give each variable its own display

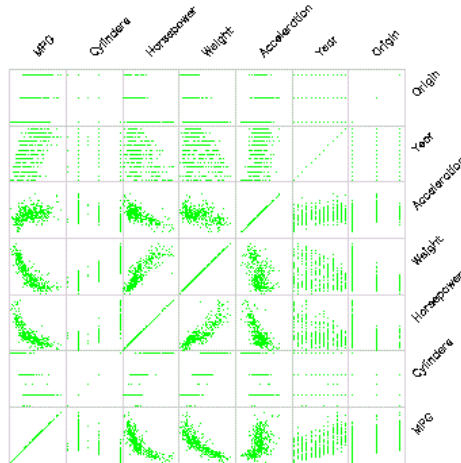
	A	B	C	D	E
1	4	1	8	3	5
2	6	3	4	2	1
3	5	7	2	4	3
4	2	6	3	1	5
5	3	4	5	1	7



# Scatterplot Matrix



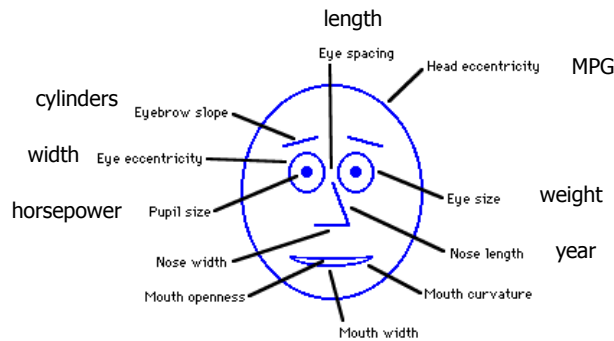
Represent each possible pair of variables in their own 2-D scatterplot



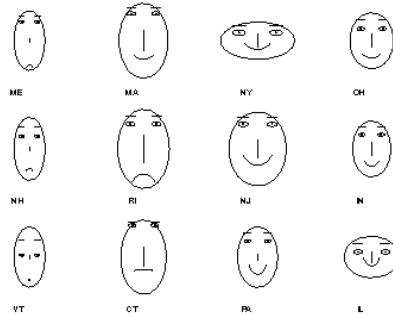
# Chernoff Faces



Encode different variables' values in characteristics of human face



# Examples



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

Fall 2014

CS 4460

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 2014

CS 4460

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

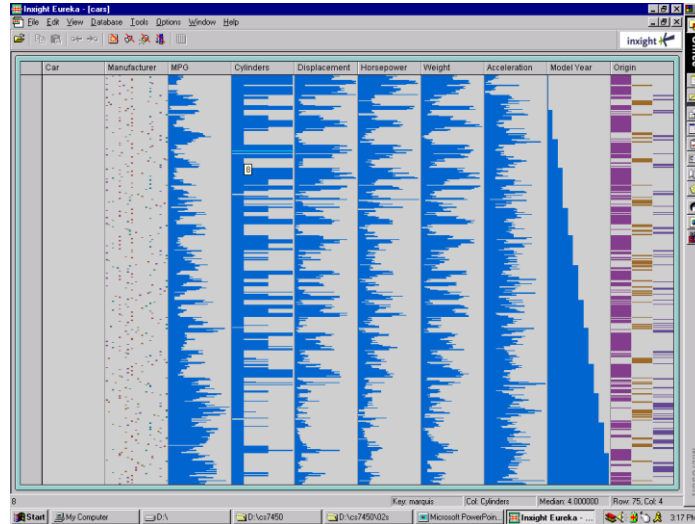


	A	B	C	D	E	F	G	H	I
1	Cereal	Manufacture	Type	Calories	Protein	Fat	Sodium	Fiber	Carbol
2	Frosted Mini-Wheats	K	C	100	3	0	0	3	
3	Raisin Squares	K	C	90	2	0	0	2	
4	Shredded Wheat	N	C	80	2	0	0	3	
5	Shredded Wheat 'n Bran	N	C	90	3	0	0	4	
6	Shredded Wheat spoon s	N	C	90	3	0	0	3	
7	Puffed Rice	Q	C	50	1	0	0	0	
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 2014

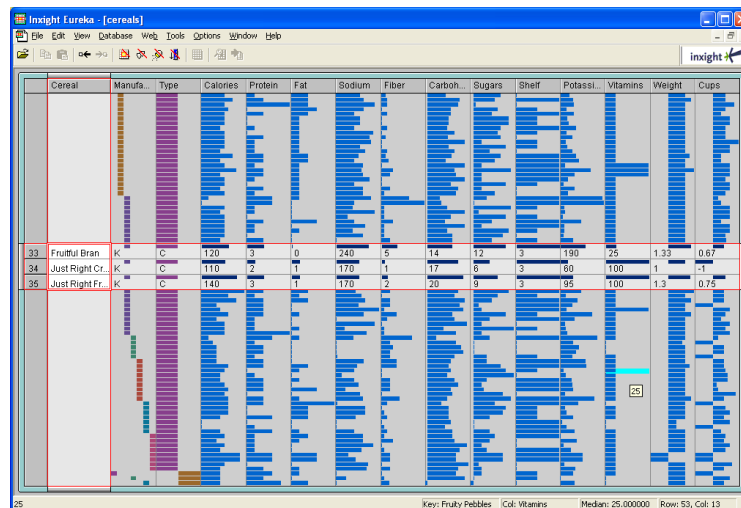
CS 4460

17

# Details



Focus on item(s) while showing the context

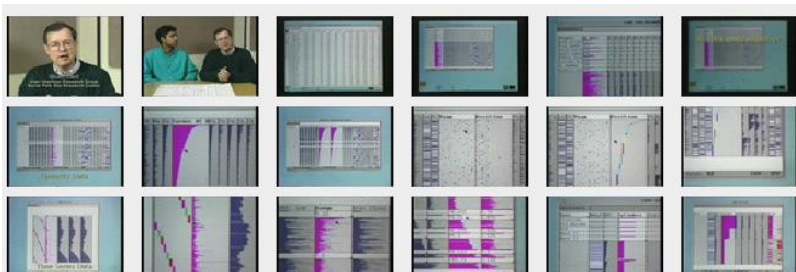


Fall 2014

CS 4460

18

# See It



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

Video

Fall 2014

CS 4460

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 2014

CS 4460

20

FOCUS - (Printers.tac)

Records: 92 qualified  
Attributes: 51 differ

Supported Interfaces=Centronics

Printer	Vendor	Technology	Color	Price (\$)	Resolution	Vertical (dpi)	Horizontal (dpi)	Emulations	Interfaces	Auto Switching	Centronics	RS-422A	LocalTalk	Ethernet	Other
HP LaserJet 4000	HP	Laser	•	1599	600	600	600	HP LaserJet 4000	Centronics	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
HP LaserJet 4050	HP	Laser	•	1599	600	600	600	HP LaserJet 4050	Centronics	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
HP LaserJet 4100	HP	Laser	•	1599	600	600	600	HP LaserJet 4100	Centronics	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>



Figure 1: An overview of the printer table.

## Characteristics

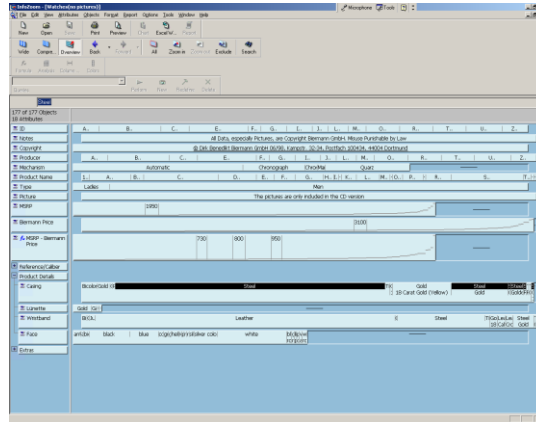
- Can sort on any attribute (row)
- Focus on an attribute value (show only cases having that value) by double-clicking on it
- Can type in queries on different attributes to limit what is presented too



# Manifestation



InfoZoom



Commercial product to be demo'ed coming up

Fall 2014

CS 4460

23

# 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 2014

CS 4460

24

# Mosaic Plot

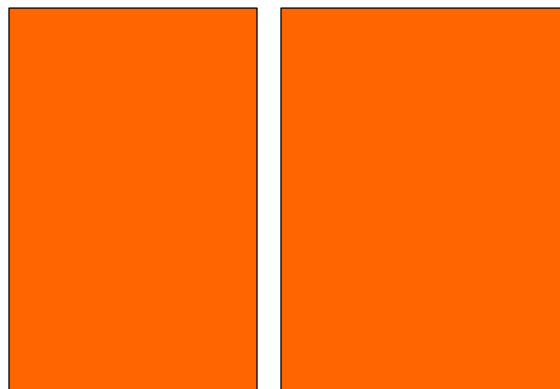


Fall 2014

CS 4460

25

# Mosaic Plot



Women

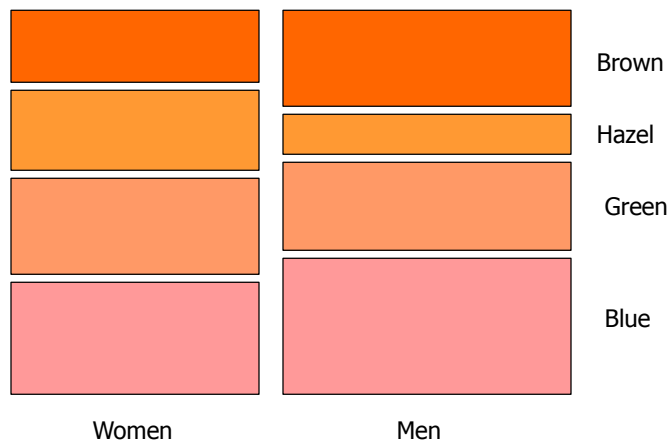
Men

Fall 2014

CS 4460

26

# Mosaic Plot

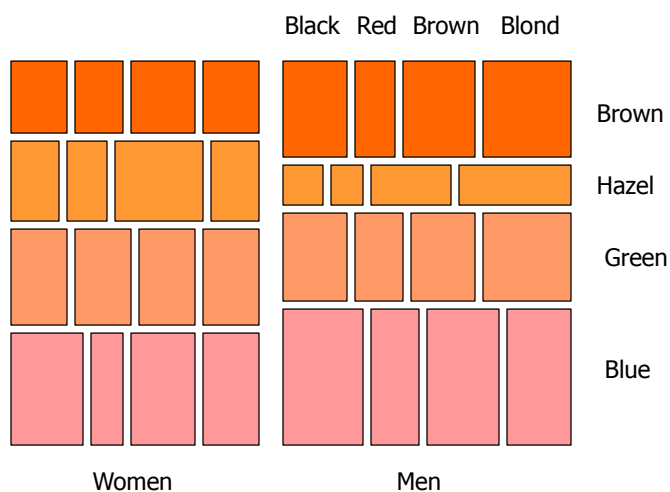


Fall 2014

CS 4460

27

# Mosaic Plot



Fall 2014

CS 4460

28

# Attribute Explorer



- General hypervariate data representation combined with flexible interaction

Spence & Tweedie  
Inter w Computers '98

Fall 2014

CS 4460

29

# 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 2014

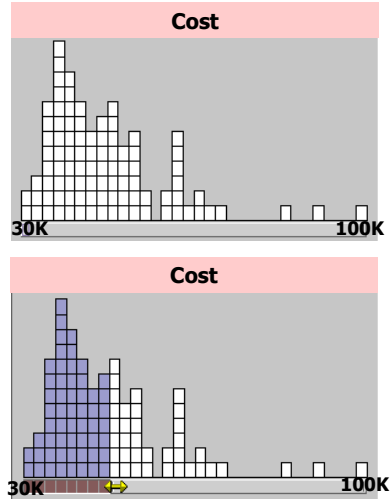
CS 4460

30

# Features



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



Fall 2014

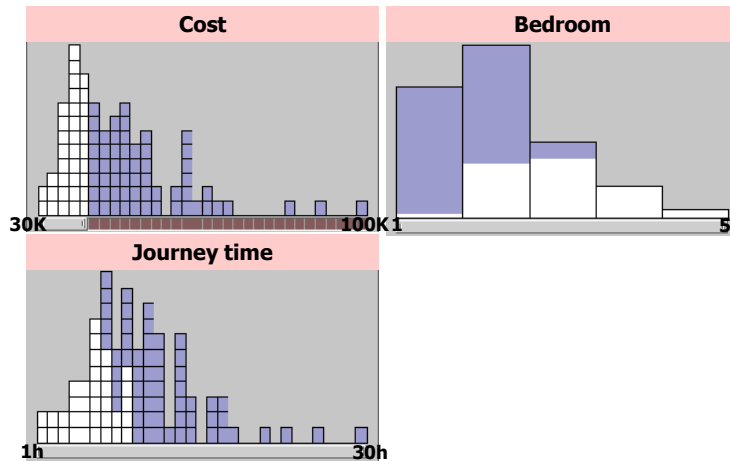
CS 4460

31

# Features



- Inter-relations between attributes – brushing



Fall 2014

CS 4460

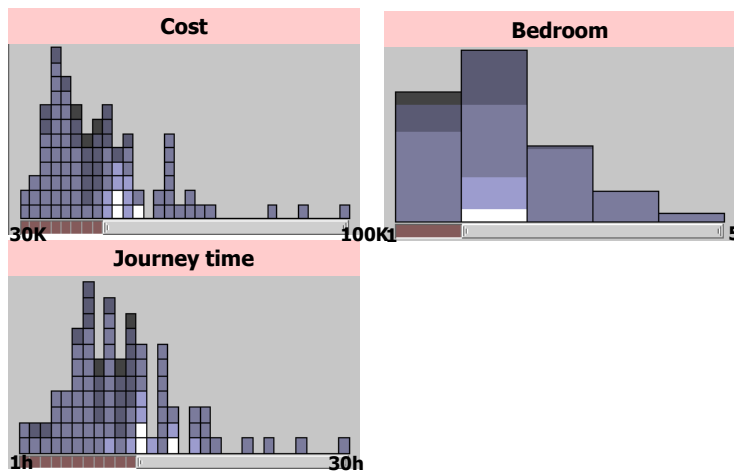
32



# Features



- Color-encoded sensitivity



Fall 2014

CS 4460

33

# Attribute Explorer



Video

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

Fall 2014

CS 4460

34

# Summary



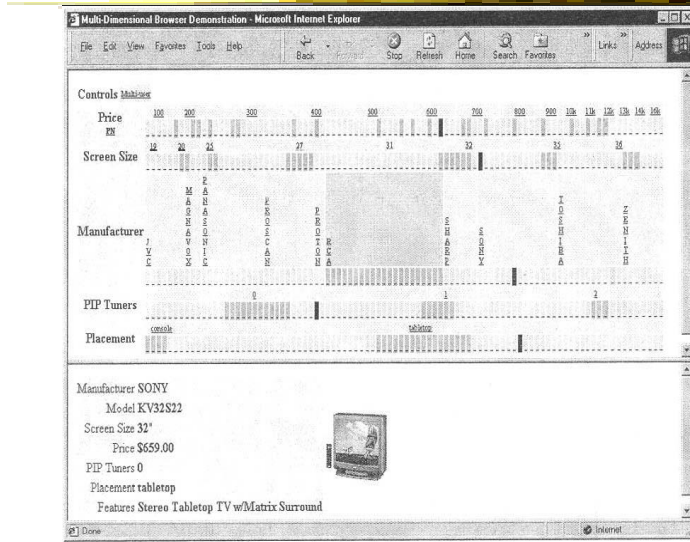
- Summary
  - Attribute histogram
  - Attribute relationship
  - Sensitivity information
  - Especially useful in “zero-hits” situations or when you are not familiar with the data at all
- Limitations
  - Limits on the number of attributes

# 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

# Interface



Fall 2014

CS 4460

37

# 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 2014

CS 4460

38

# Attributes as Sliding Rods



Fall 2014

CS 4460

39

# Instantiation



Video

<https://www.youtube.com/watch?v=GEBx-XTTrGps>

Fall 2014

CS 4460

40

# Limitations



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

# Parallel Coordinates



- What are they?
  - Explain...

# Parallel Coordinates



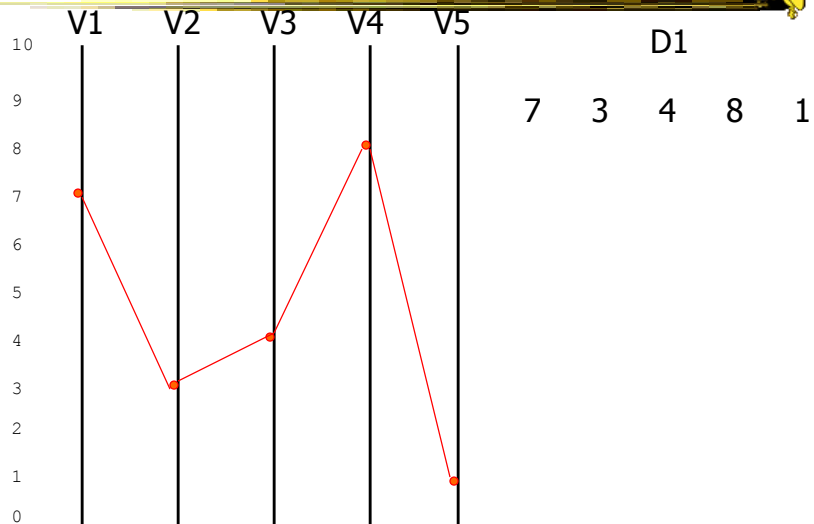
	V1	V2	V3	V4	V5
D1	7	3	4	8	1
D2	2	7	6	3	4
D3	9	8	1	4	2

Fall 2014

CS 4460

43

# Parallel Coordinates

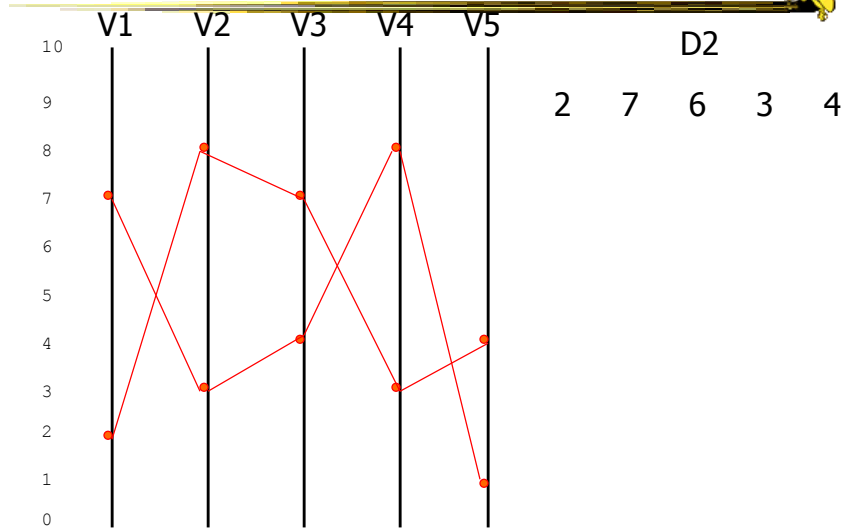


Fall 2014

CS 4460

44

# Parallel Coordinates

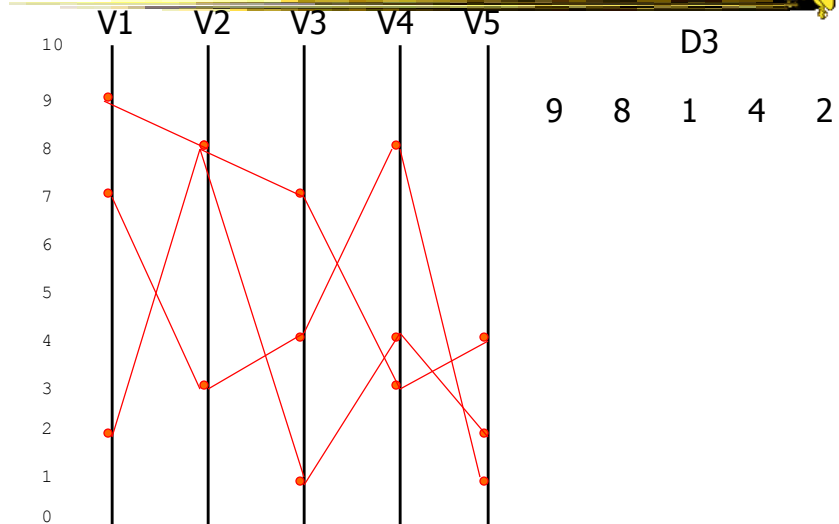


Fall 2014

CS 4460

45

# Parallel Coordinates

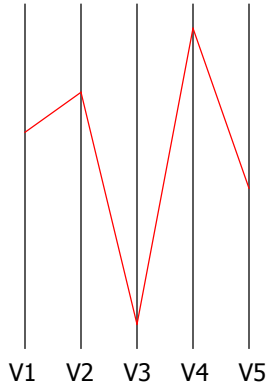


Fall 2014

CS 4460

46

# Parallel Coordinates



Encode variables along a horizontal row

Vertical line specifies different values that variable can take

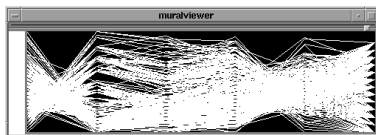
Data point represented as a polyline

Fall 2014

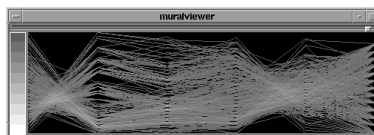
CS 4460

47

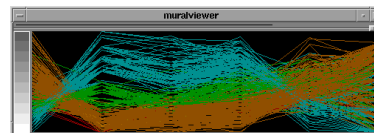
# Parallel Coords Example



Basic



Grayscale



Color

Fall 2014

CS 4460

48



# Issue



- Different variables can have values taking on quite different ranges
- Must normalize all down (e.g., 0- $\rightarrow$ 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



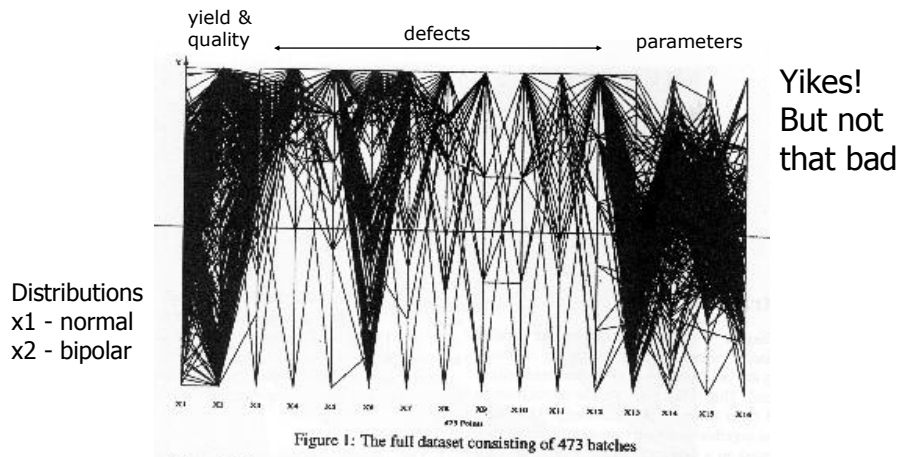
- VLSI chip manufacture
- Want high quality chips (high speed) and a high yield batch (% of useful chips)
- Able to track defects
- Hypothesis: No defects gives desired chip types
- 473 batches of data

# The Data



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

# Parallel Coordinate Display



# Top Yield & Quality

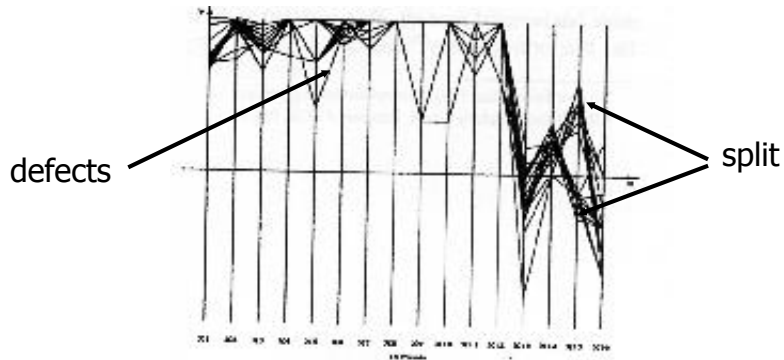


Figure 2: The batches high in Yield,  $X_1$ , and Quality,  $X_2$ .

Have some defects

Fall 2014

CS 4460

55

# Minimal Defects



Not the highest yields and quality

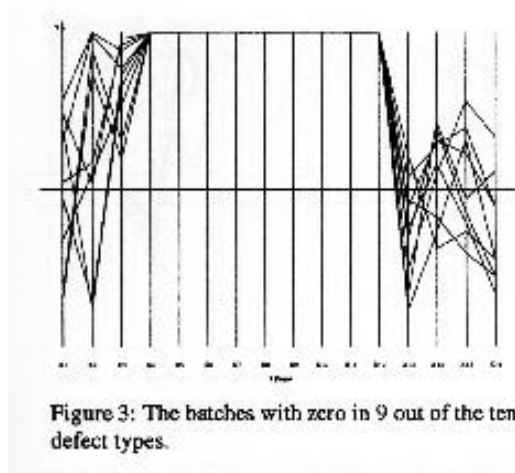


Figure 3: The batches with zero in 9 out of the ten defect types.

Fall 2014

CS 4460

56

# Best Yields



Appears that some defects are necessary to produce the best chips

Non-intuitive!

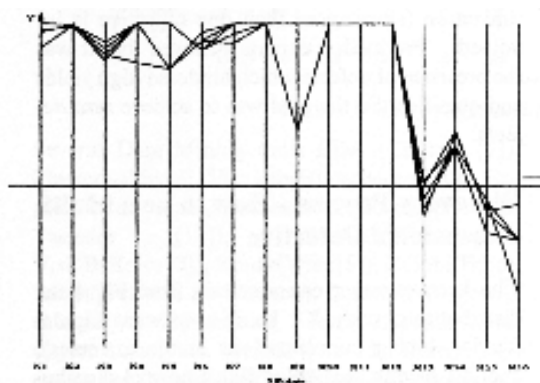


Figure 6: Batches with the highest Yields do not have the lowest defects in X3 and X6.

Fall 2014

CS 4460

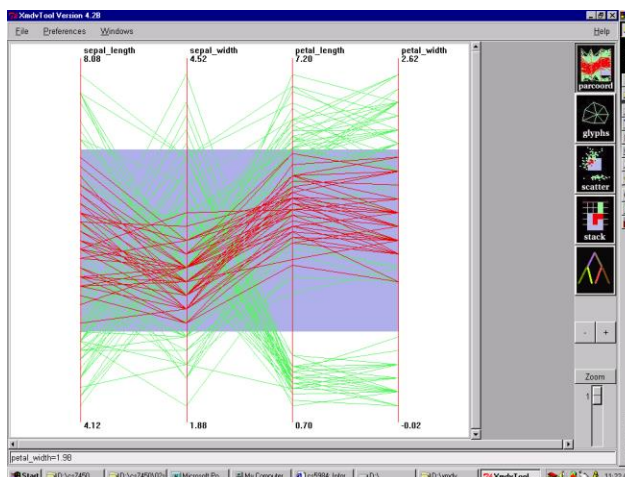
57

# XmdvTool



Toolsuite created by Matthew Ward of WPI

Includes parallel coordinate views

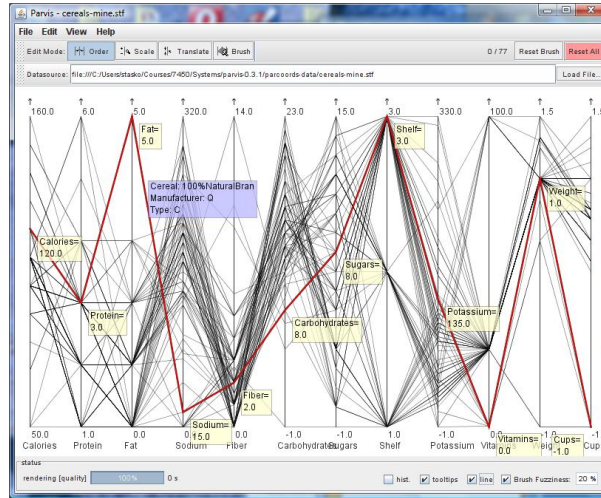


Fall 2014

CS 4460

58

# ParVis System



Demo

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

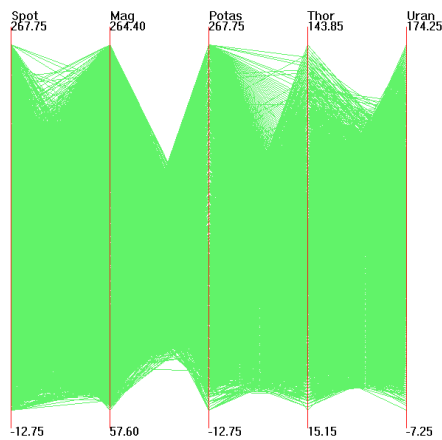
Fall 2014

CS 4460

59

# Challenges

Too much data



Out5d dataset (5 dimensions, 16384 data items)

Fall 2014

CS 4460

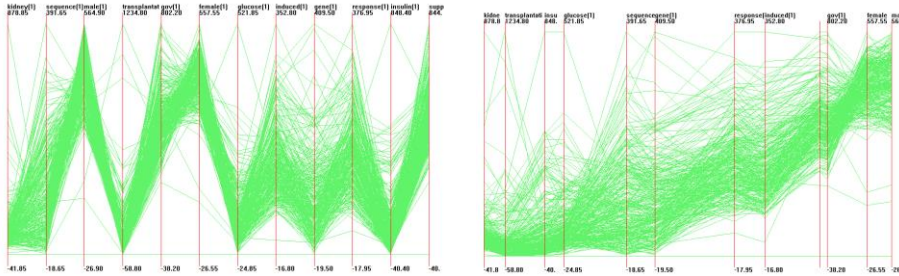
(courtesy of J. Yang)

60

# Dimensional Reordering



Which dimensions are most like each other?



Same dimensions ordered according to similarity

Yang et al  
InfoVis '03

Fall 2014

CS 4460

61

# Dimensional Reordering



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

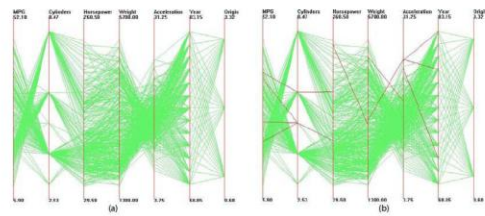


Figure 1: Parallel coordinates visualization of Cars dataset. Outliers are highlighted with red in (b).

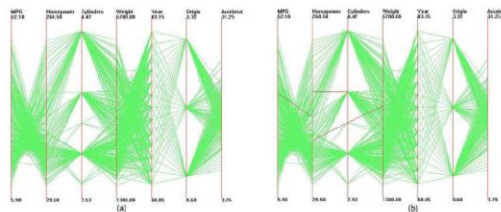


Figure 2: Parallel coordinates visualization of Cars dataset after clutter-based dimension reordering. Outliers are highlighted with red in (b).

Peng et al  
InfoVis '04

Fall 2014

CS 4460

62

# Reducing Density

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

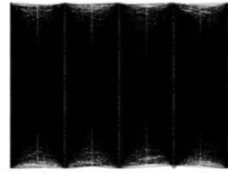
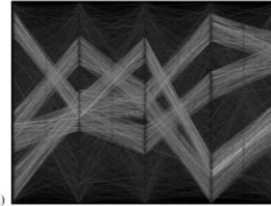
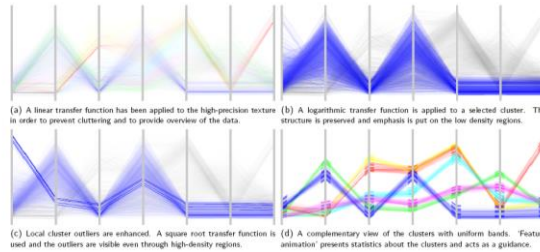


Figure 1 – Parallel Coordinates visualization of the *Sirt1* data set (7,500 five-attribute records).



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 bases. 'Feature animation' presents statistics about the clusters and acts as a guidance.

Johansson et al, '05

Fall 2014

CS 4460

63

# 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

Fall 2014

CS 4460

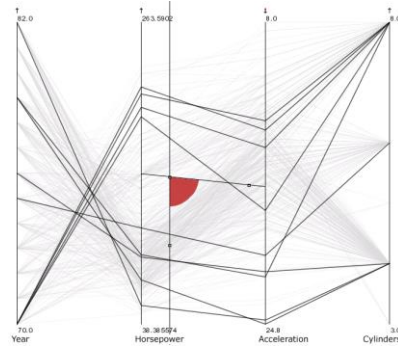
64



# 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)

Fall 2014

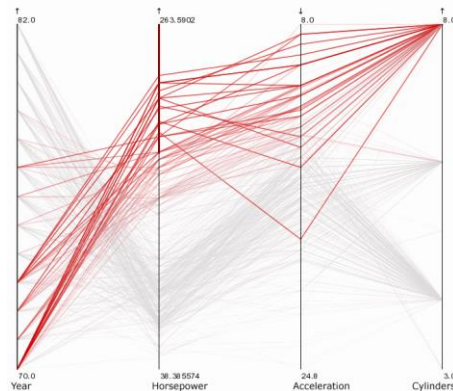
CS 4460

65

# Range Focus



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

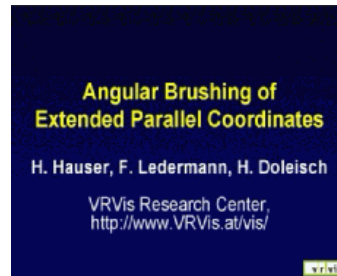


Fall 2014

CS 4460

66

# Video



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

Fall 2014

CS 4460

67

# Different Kinds of Data



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

Fall 2014

CS 4460

68

# 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

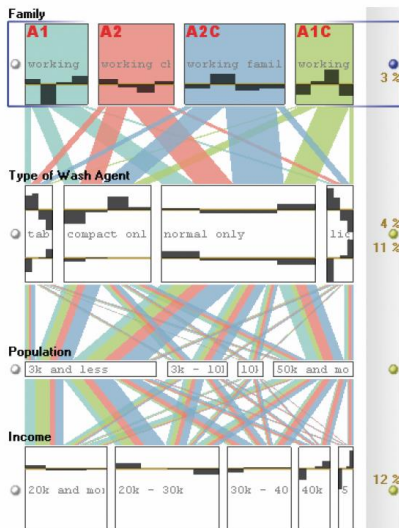
Kosara, Bendix, & Hauser  
TVCG '05

Fall 2014

CS 4460

69

# Representation



Color used for different categories

Those values flow into the other variables

Fall 2014

CS 4460

70

# Example



Titanic passengers data set

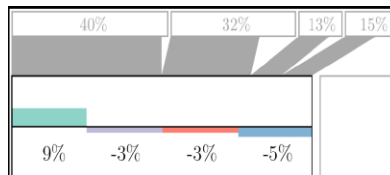
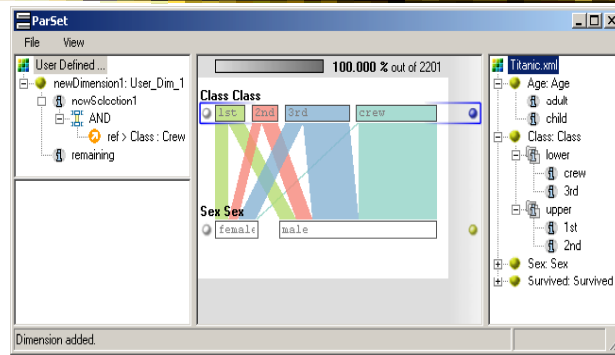
Class	Sex				
	female		male		
first	145	44.6%	180	55.4%	325
	30.8%	6.6%	10.4%	8.2%	14.8%
second	106	37.2%	179	62.8%	285
	22.6%	4.8%	10.4%	8.1%	12.9%
third	196	27.8%	510	72.2%	706
	41.7%	8.9%	29.5%	23.2%	32.1%
crew	23	2.6%	862	97.4%	885
	4.9%	1.1%	49.8%	39.1%	40.2%
	470		1731		2201
		21.4%		78.6%	100%

Fall 2014

CS 4460

71

# Titanic Data Set



Fall 2014

CS 4460

72

# Interactions

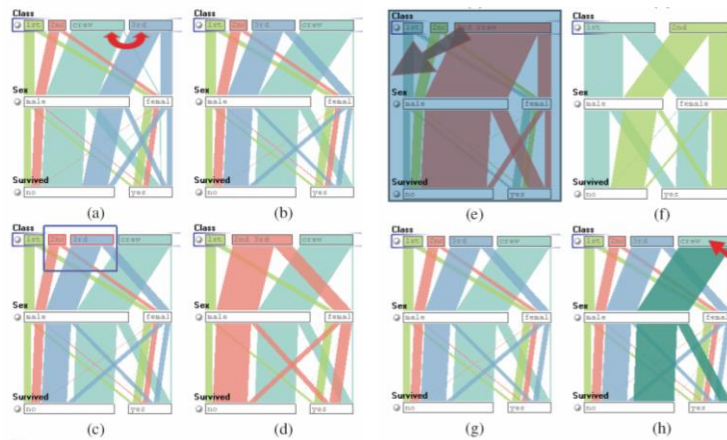
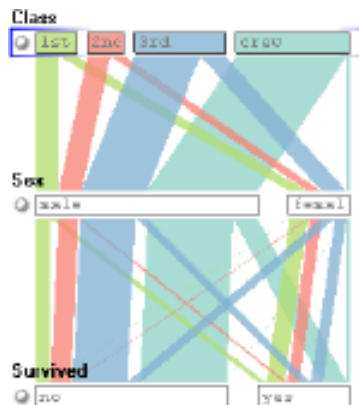


Fig. 7. Basic interaction elements in Parallel Sets: reordering categories (a, b) helps to generate a more meaningful layout; grouping categories (c, d) enables a hierarchical analysis/exploration; excluding categories from the visualization (e, f) allows for interactive filtering; and category highlighting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

Fall 2014

73

# Video



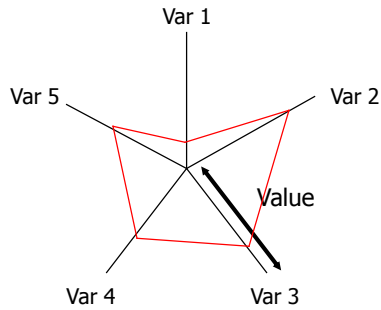
Fall 2014

CS 4460

InfoVis '05

74

# Star Plots



Space out the  $n$  variables at equal angles around a circle

Each "spoke" encodes a variable's value

Alternative Rep.

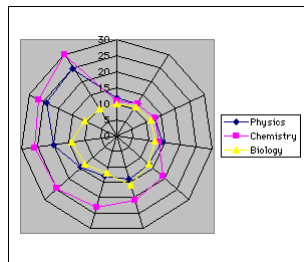
Data point is now a "shape"

Fall 2014

CS 4460

75

# 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>

Fall 2014

CS 4460

76

# Star Coordinates



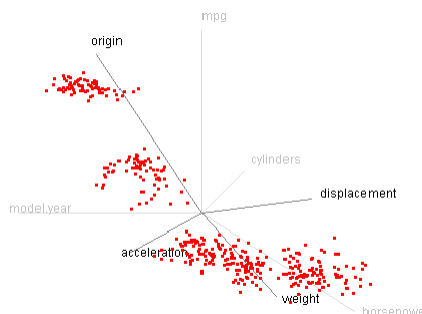
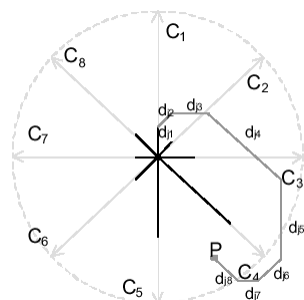
- Same ideas as star plot
- Rather than represent point as polyline, just accumulate values along a vector parallel to particular axis
- Data case then becomes a point

Fall 2014

CS 4460

77

# 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](#)

Fall 2014

CS 4460

78

# 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

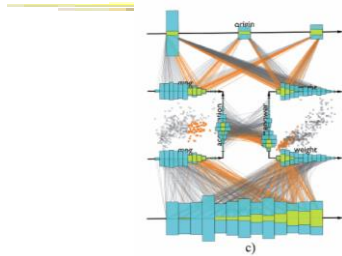
# Generalizing the Principles



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



# FLINA View



c)

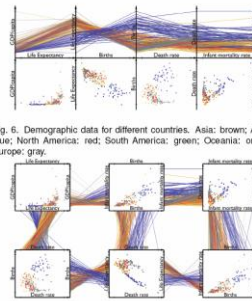
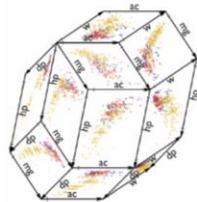
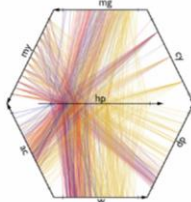


Fig. 6. Demographic data for different countries. Asia: brown; Africa: blue; North America: red; South America: green; Oceania: orange; Europe: gray.

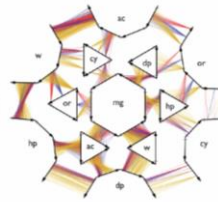
Fig. 7. Alternative lay-out for demographic data



(d) Hyperbox



(e) Time Wheel



(f) Many-to-many PCP

[Video](#)

Fall 2014

CS 4460

81

# Parallel Coordinates

- Technique
  - Strengths?
  - Weaknesses?

Fall 2014

CS 4460

82

# Project



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

Fall 2014

CS 4460

83

# HW 2



- Design table & graph
- Due Tuesday
  - Bring 2 hardcopies

Fall 2014

CS 4460

84

# Upcoming



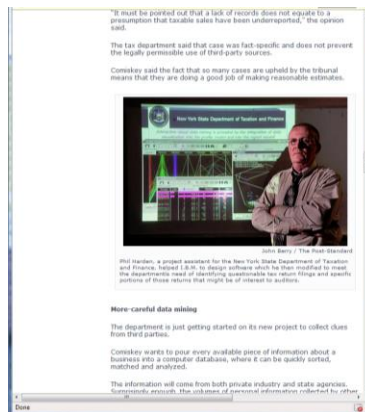
- Multivariate Visual Representations 2
  - Reading:  
Munzner chapter 12
- D3 tutorial
  - Reading  
*Interactive Data Visualizations for the Web*,  
chapters 3 and 5

Fall 2014

CS 4460

85

# Application



[http://www.syracuse.com/news/index.ssf/2010/01/data\\_mining\\_helps\\_new\\_york\\_cat.html](http://www.syracuse.com/news/index.ssf/2010/01/data_mining_helps_new_york_cat.html)

Fall 2014

CS 4460

86

# Application



**PPCL** Process Plant Computing Ltd  
P.O. Box 43  
Gerrards Cross, Buckinghamshire, SL9 8JX, UK  
Tel: +44 (1753) 852000 | Fax: +44 (1753) 892959

Home | **IPC Technology** | Products | Training & Services | The Company | Online Material | Contact us

C Visual Explorer uses a unique form of Geometry to convert historical data into a Single Visual Summary

### C Visual Explorer (CVE)

Discover why you have a process historian! You knew there was lots of new process knowledge representing improvement opportunity hidden in all that data and thought that you could extract it. Then you realised that engineers lacked the tools they needed so could only pick at a few highlights. CVE is the tool that you have been missing and with CVE you will discover exactly why it was such a good decision to buy a process historian.

Process plant data is different from the data usually explored with the traditional 'data mining' and 'predictive data analytics' methods that you may have learnt in college because it is highly correlated amongst hundreds or thousands of variables so that there are many, many correlations between variables. That means the problem is not that of finding correlations but is one of recognising those that were previously unknown and have value amongst the many that you already know or that are direct consequences of the underlying chemistry, mass, heat and momentum balances in your particular process.

Correlations in themselves don't identify cause and effect. That requires the engineers process knowledge so the method needed also has to be quick and straight-forward to use and explain to others for busy engineers with many other tasks to accomplish during their working day.

That Process Plant Data is also often highly non-linear, and non-linearities can be easily seen in CVE, often for the first time, adds to its uniqueness and demands the ability provided by CVE to easily separate out individual modes of non-linear behaviour.

PPCL Webinars  
Set the Schedule and Book Here

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

Fall 2014

CS 4460

87