Multivariate Visual Representations, Parts 1&2

CS 4460 - Introduction to Information Visualization

Fall, 2018

Alex Endert

Today

- General representation techniques for multivariate (>3) variables per data case
  - high-dimensional data

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

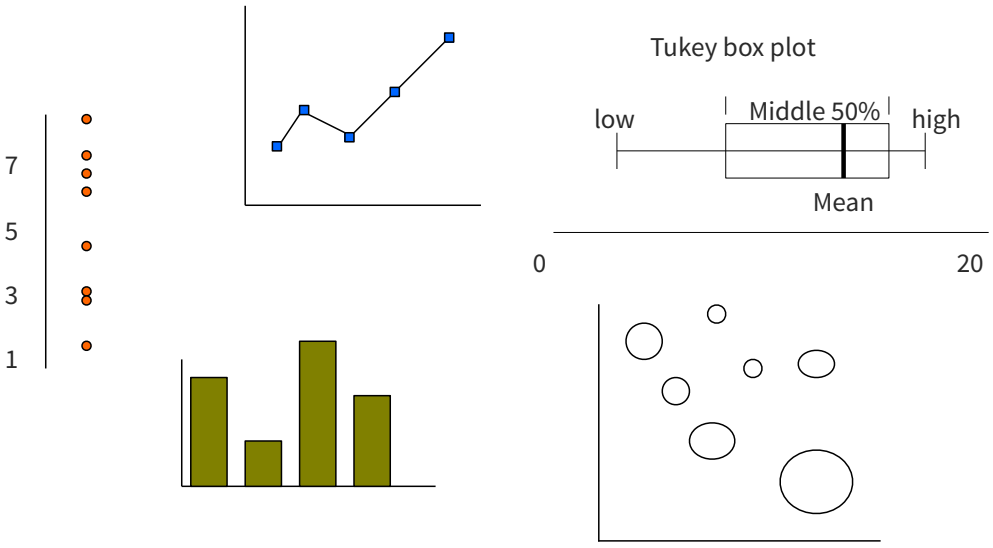
Earlier

- 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

Challenge of more dimensions: how to project to 2D

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

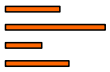
Spreadsheets?

- 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
  - You also watched the TableLens Video again.
  - **Anyone want to share how TableLens works?**

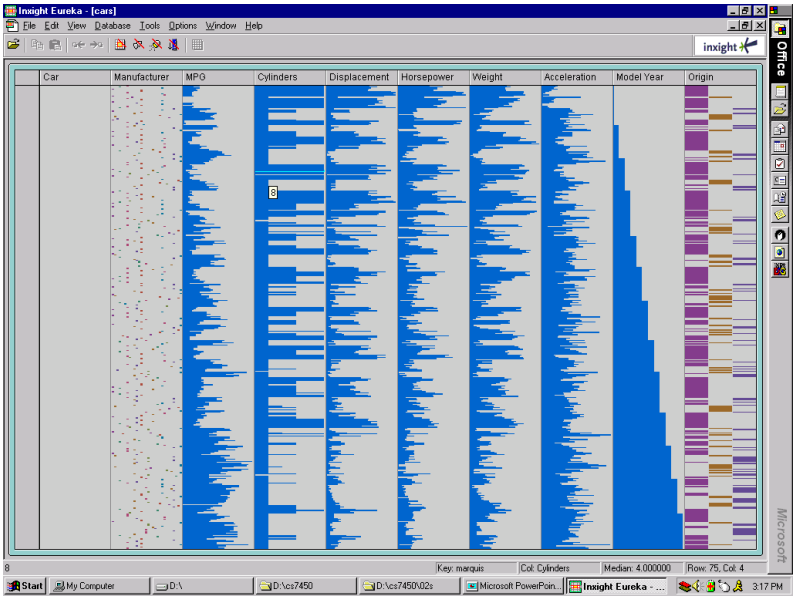
Visual Mapping

	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



Instantiation

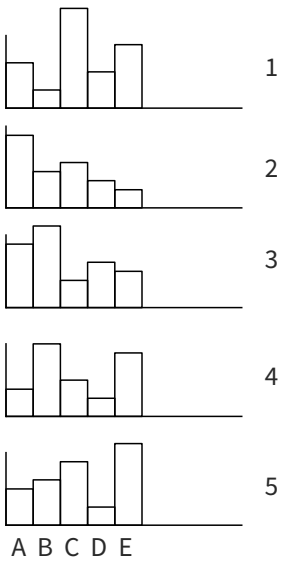


Revisit

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

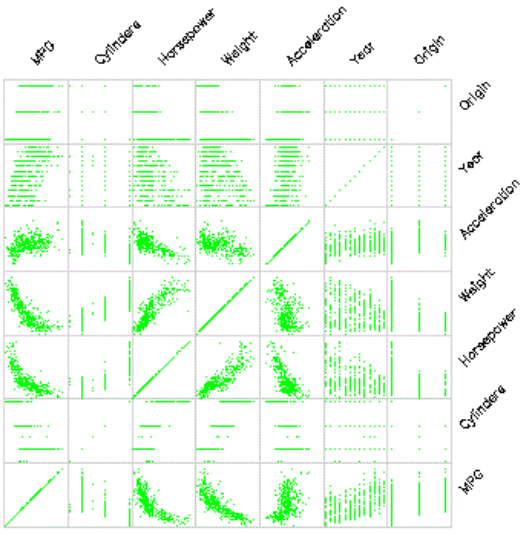


Revisit

Scatterplot Matrix

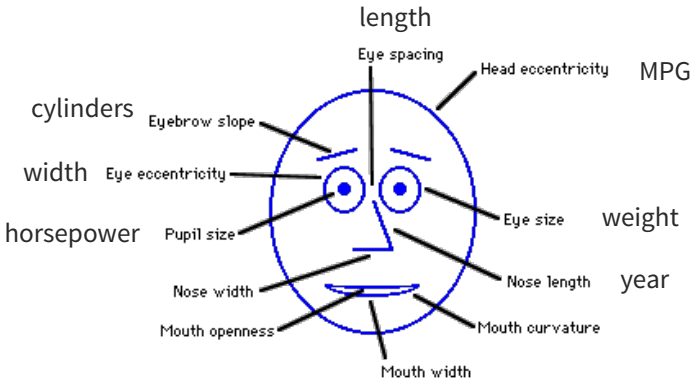
Pairwise scatter plot combination

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

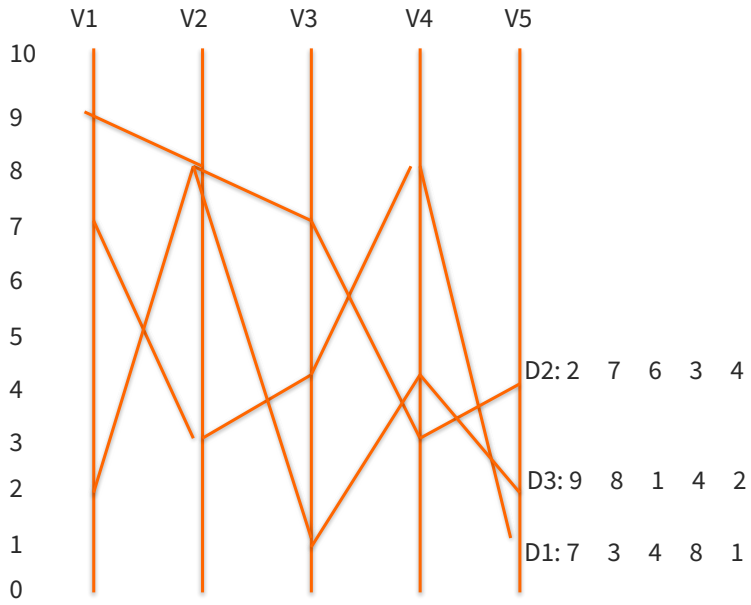


Chernoff Faces

Encode different variables' values in characteristics of human face

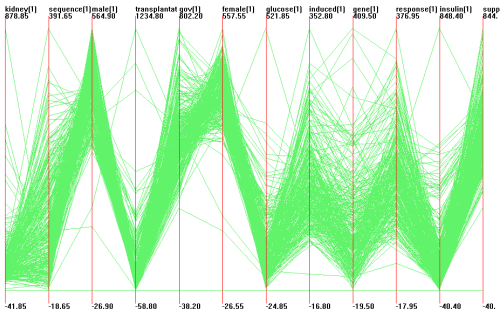


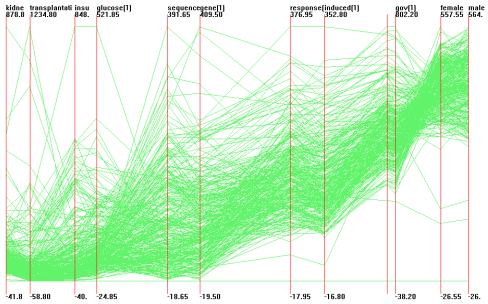
Parallel Coordinates



Remember, Dimensional Reordering

Which dimensions are most like each other?

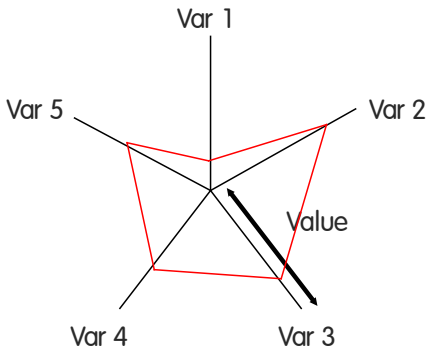




Same dimensions ordered according to similarity

Yang et al
InfoVis '03

Star Plots



Alternative Rep.

Space out the n variables at equal angles around a circle

Each "spoke" encodes a variable's value

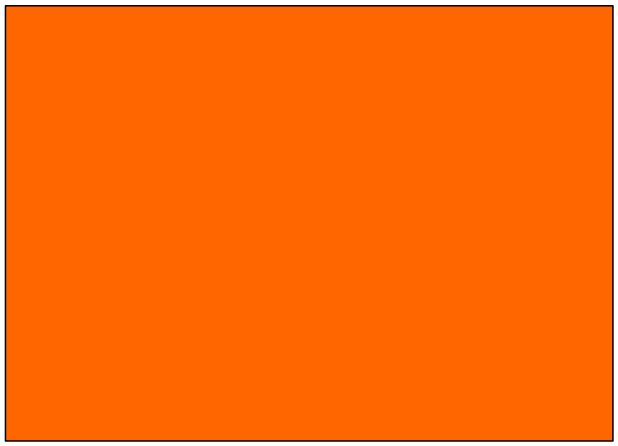
Data point is now a "shape"

Let's try one, warmup

- Create either a **star plot** or **parallel coordinates plot** from the data table
- Get in small groups
- Draw either plot
- I'll call on one of each plot to share

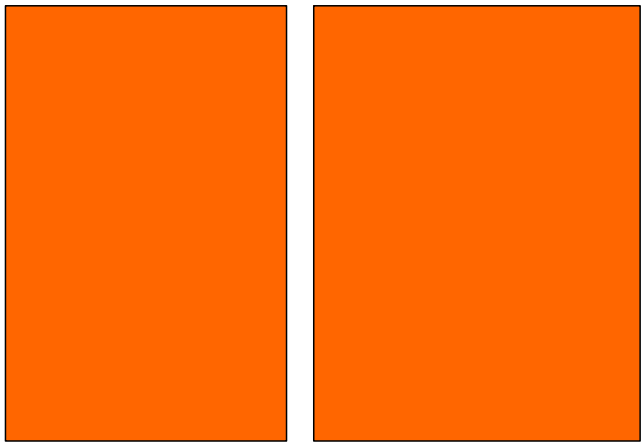
ID	Price	Sq. Foot	# Beds
House A	$200,000	2,250	3
House B	$300,000	3,000	4
House C	$175,000	1,500	1
House D	$500,000	5,500	5

Mosaic Plot



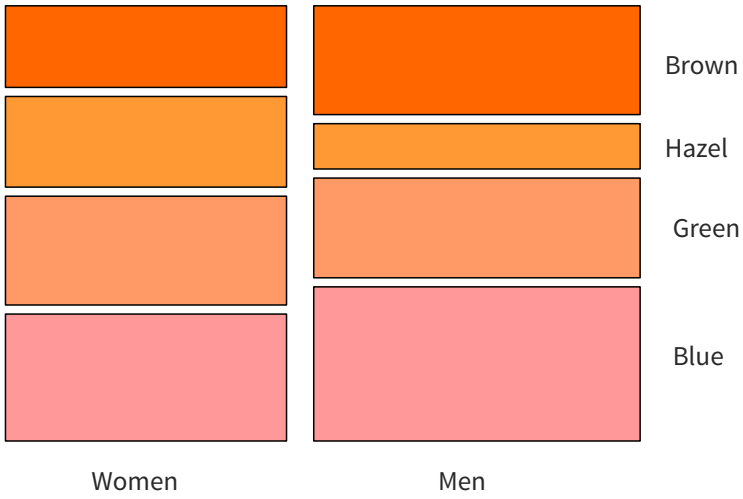
Mosaic Plot

encode **gender**



Mosaic Plot

encode **gender, eye-color**

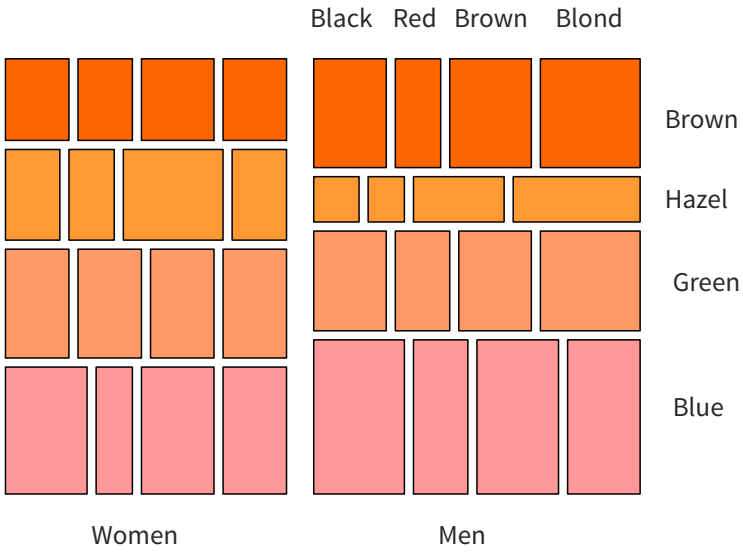


Mosaic Plot

encode **gender, eye-color, hair color**

Get more variables and sub-divide the whole area further

Very good for "proportion" tasks, but gives no insight on individual data.



Attribute Explorer

- General hypervariate data representation combined with flexible interaction

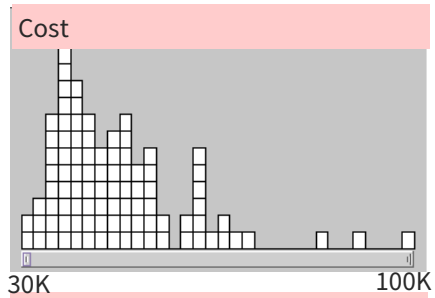
Spence & Tweedie
Inter w Computers '98

Characteristics

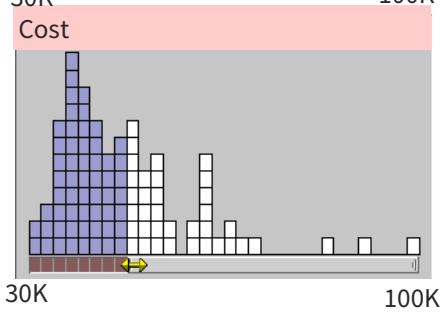
- Multiple histogram views, one per attribute (like trellis)
- Each data **item** represented by a **square**
- Square is positioned relative to that case's value on that attribute
- Selecting **case in one view lights it up in others**
- Query sliders for narrowing
- Use shading to indicate level of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

Features

- Attribute histogram
- All objects on all attribute scales



- Interaction with attributes limits

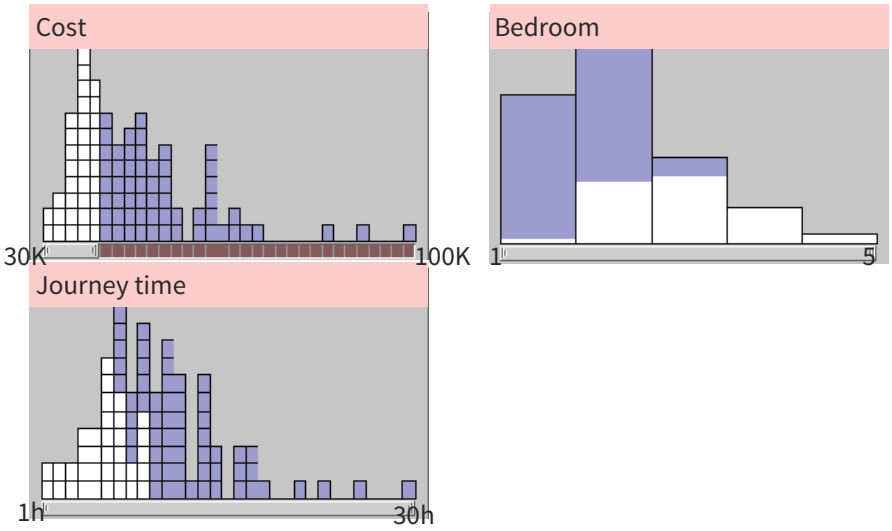


Features

Shows distribution of data that falls into a particular categorization of an attribute.

- Inter-relations between attributes – brushing

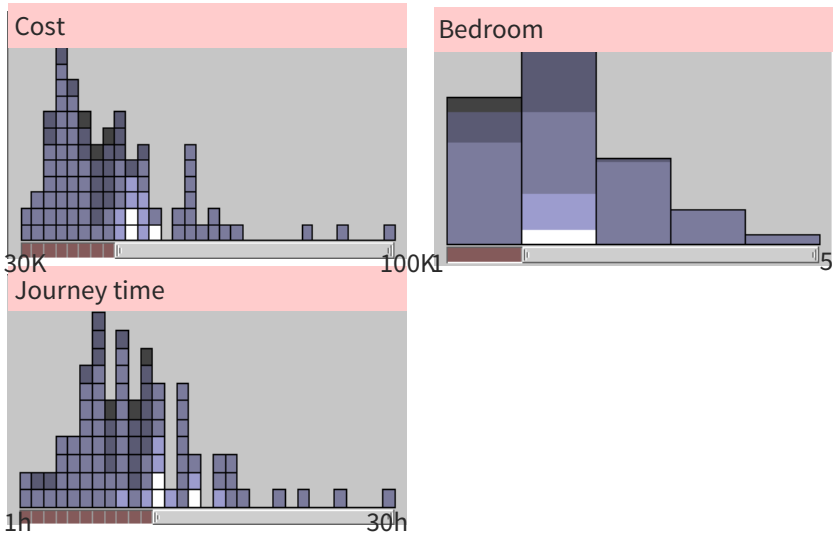
>_<: you have to create one of these for each one of the attributes.



Features

^_^: One distribution histogram using squares with different colors that layer up.

- Color-encoded sensitivity



Different Kinds of Data

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

Parallel Sets

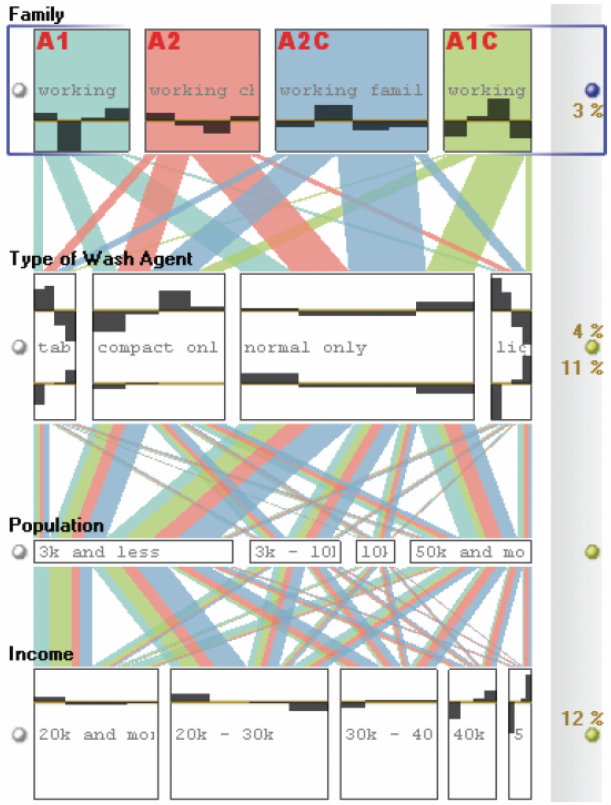
- Visualization method **adopting parallel coordinates layout but uses frequency-based representation**
- Visual metaphor
  - Layout similar to parallel coordinates
  - Continuous axes replaced with boxes
- Interaction
  - User-driven: User can create new classifications

Kosara, Bendix, & Hauser
TVCG '05

Representation

Color used for different **categories**

Those values flow into the other variables



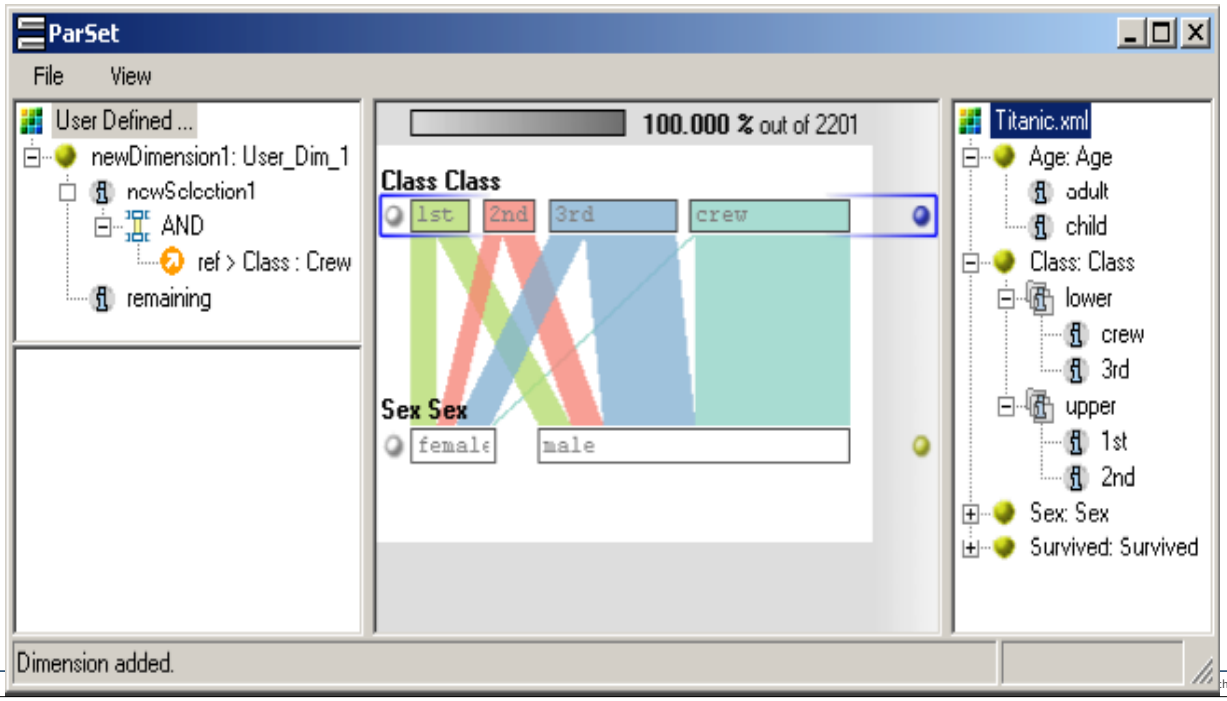
distribution of values encoded as bars within each color category

Example

Titanic passengers data set

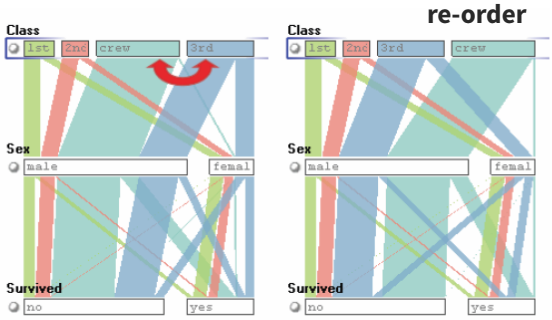
Class	Sex		
	female	male	
first	145 44.6% 30.8% 6.6%	180 55.4% 10.4% 8.2%	325 14.8%
second	106 37.2% 22.6% 4.8%	179 62.8% 10.4% 8.1%	285 12.9%
third	196 27.8% 41.7% 8.9%	510 72.2% 29.5% 23.2%	706 32.1%
crew	23 2.6% 4.9% 1.1%	862 97.4% 49.8% 39.1%	885 40.2%
	470 21.4%	1731 78.6%	2201 100%

Titanic Data Set

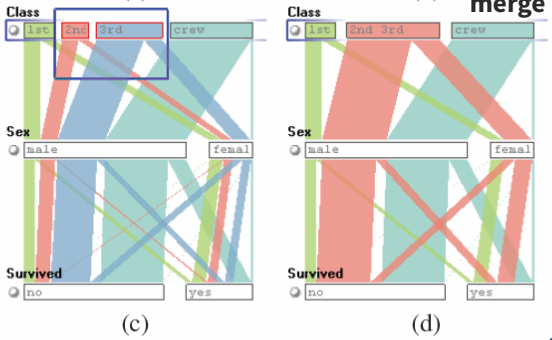


Interactions

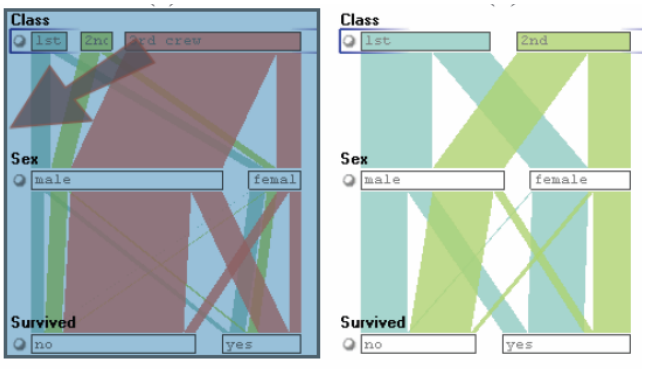
re-order



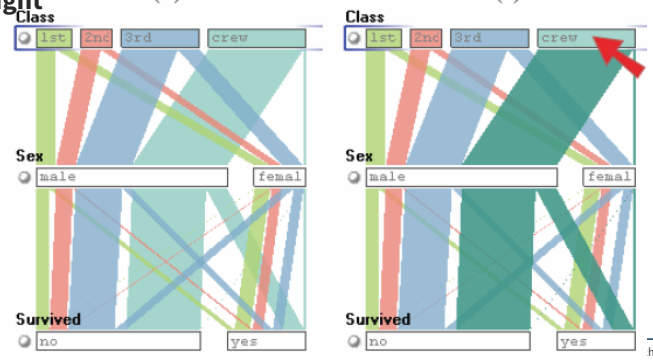
merge



delete/filter

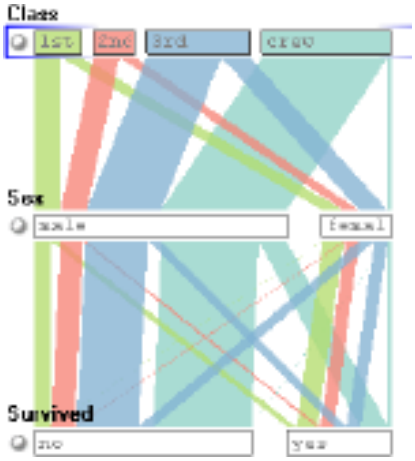


highlight



Video

https://www.youtube.com/watch?v=SphrIOU76o0

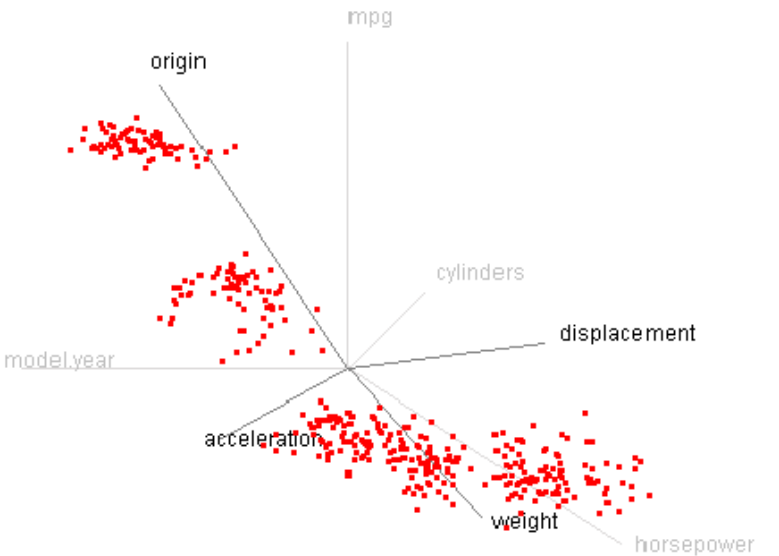


InfoVis '05

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
- Vector from center for each data dimension can be customized

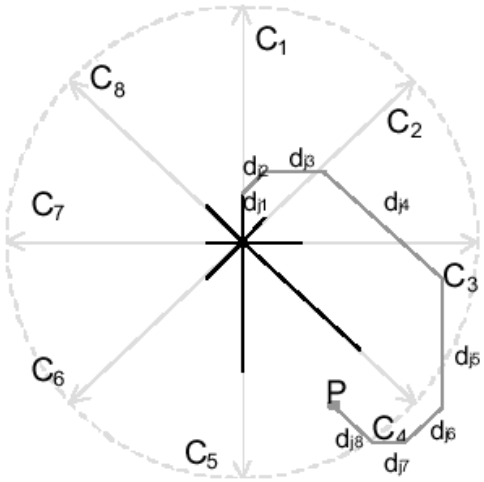
Star Coordinates



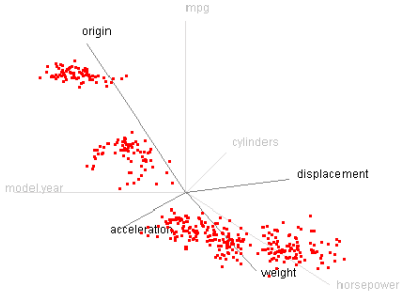
E. Kandogan, "Star Coordinates: A Multi-dimensional Visualization Technique with Uniform Treatment of Dimensions", InfoVis 2000 Late-Breaking Hot Topics, Oct. 2000

https://youtu.be/9U487xShhUs?t=54s

Star Coordinates



process



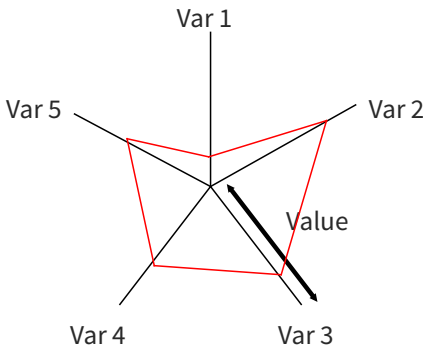
result

align each dimension as a vector
move your data point along the vector (distance = value)
sum up all movements, and that's where you plot your point
do this for all points

Star Coordinates

- Data cases with **similar values will lead to clusters of points**
- Naive approach to projection or multidimensional scaling
  - Multi-dimensional scaling or projection down to 2D

Star Plots; recall



similar to **star coordinates** plot

Let's try it again

- Create a **star coordinate plot**
- Get in small groups
- Think about **design challenges**
  - What worked?
  - What did you have to decide on?
  - What decisions might change the effectiveness.
  - How would this technique scale (both **items** and **attributes**)?

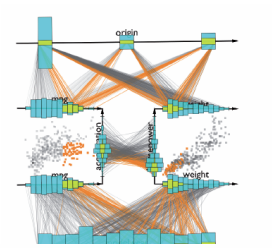
ID	Price	Sq. Foot	# Beds	Distance to work
House A	$200,000	2,250	3	4 miles
House B	$300,000	3,000	4	23 miles
House C	$175,000	1,500	1	5 miles
House D	$500,000	5,500	5	35 miles

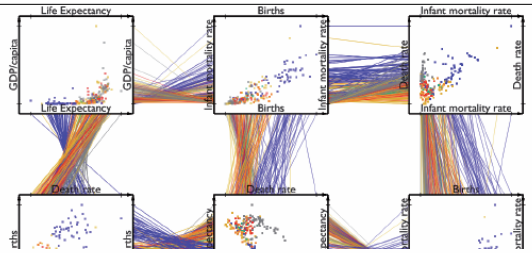
Generalizing the Principles

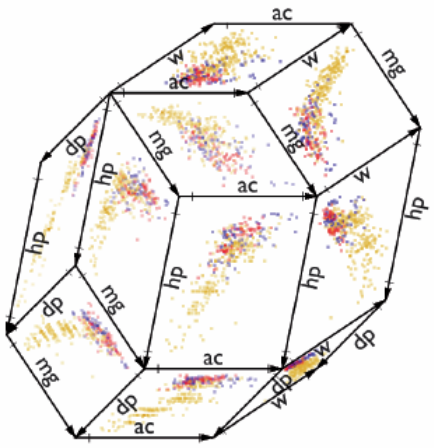
- General & flexible framework for axis-based visualizations
  - Scatterplots, par coords, etc.
- User can position, orient, and stretch axes
- Axes can be linked

Claessen & van Wijk
TVCG (InfoVis) '11

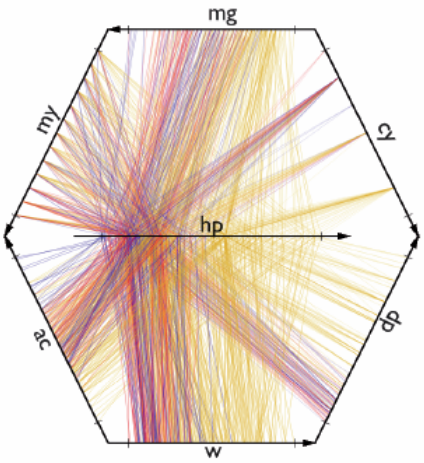
FLINA View



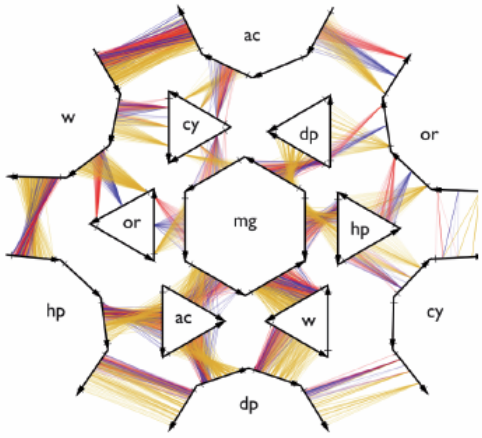




(d) Hyperbox



(e) Time Wheel



(f) Many-to-many PCP

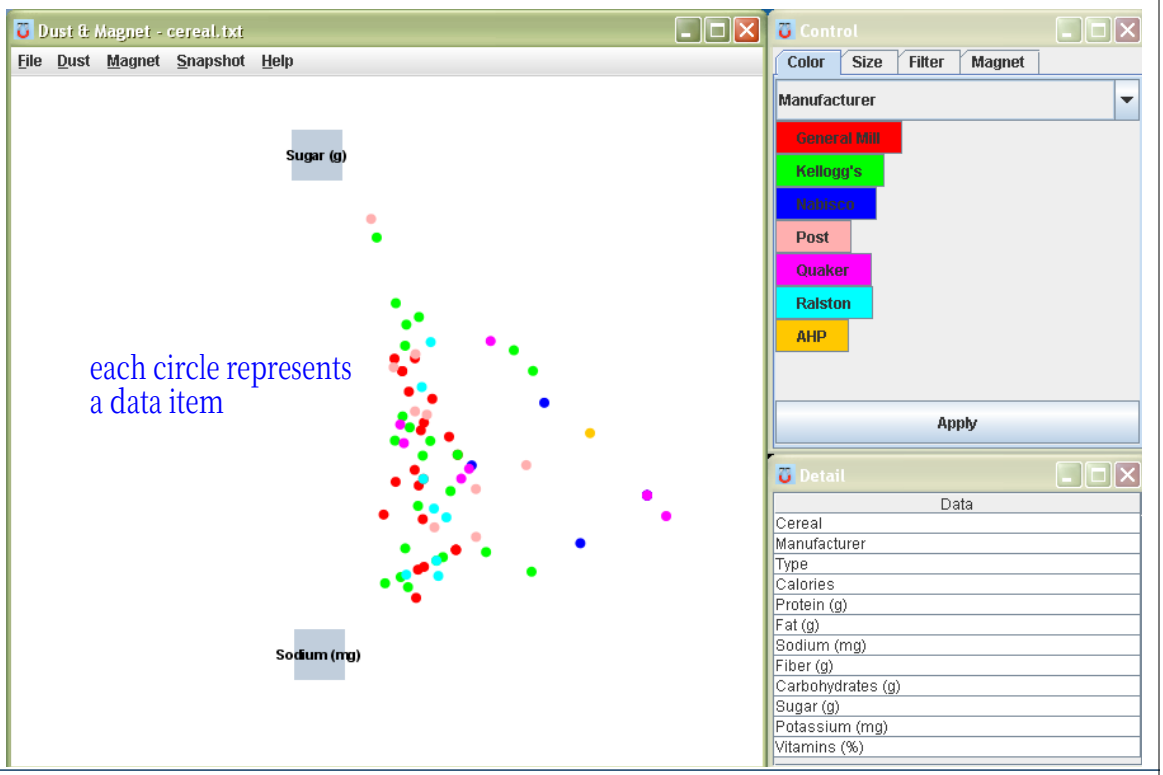
Dust & Magnet

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

Place attributes in the space, data points that are highly representative of certain attribute to be attracted to the corresponding magnet.

Yi, Melton, Stasko & Jacko
Information Visualization '05

Interface



Interaction

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

Interaction

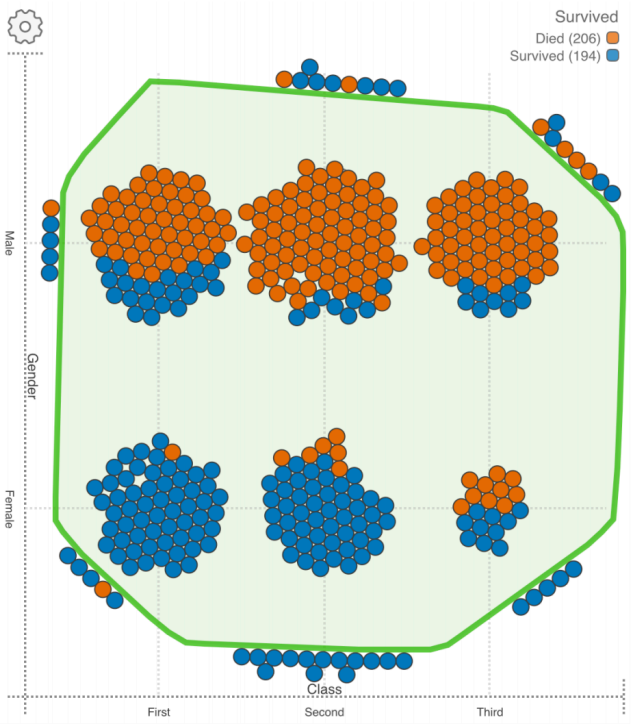
- Moving a magnet makes all the dust move
  - Also command for shaking dust
- Different strategies for how to position magnets in order to explore the data

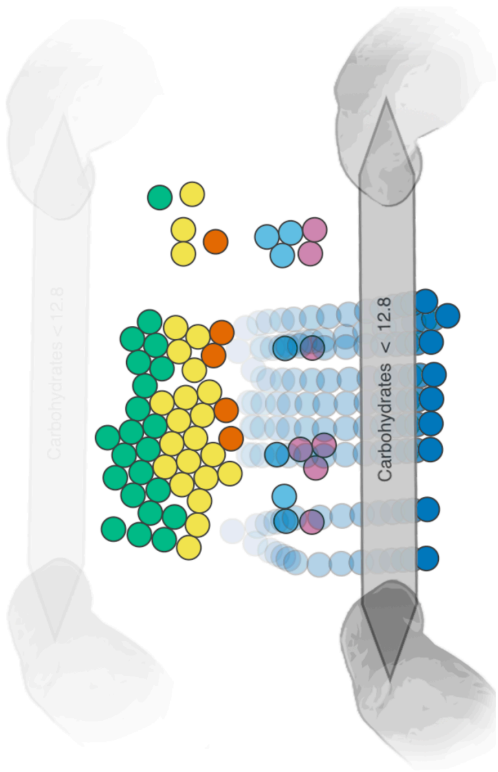
video

another one

- Kinetica
  - tablet visualization with physics-based metaphor for interactions
  - https://www.youtube.com/watch?v=7OYcGiKrmEg
  -

physical-based metaphor for interactions



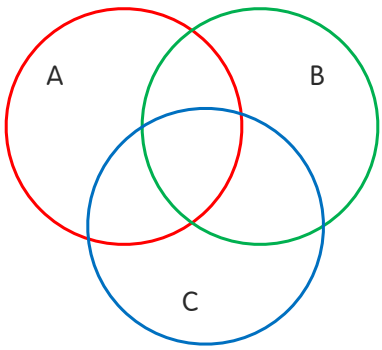


Set Operations

- Different type of problem
  - Large set of items, each item can be in one or more sets
  - How do we visually represent the set membership?

Standard Technique

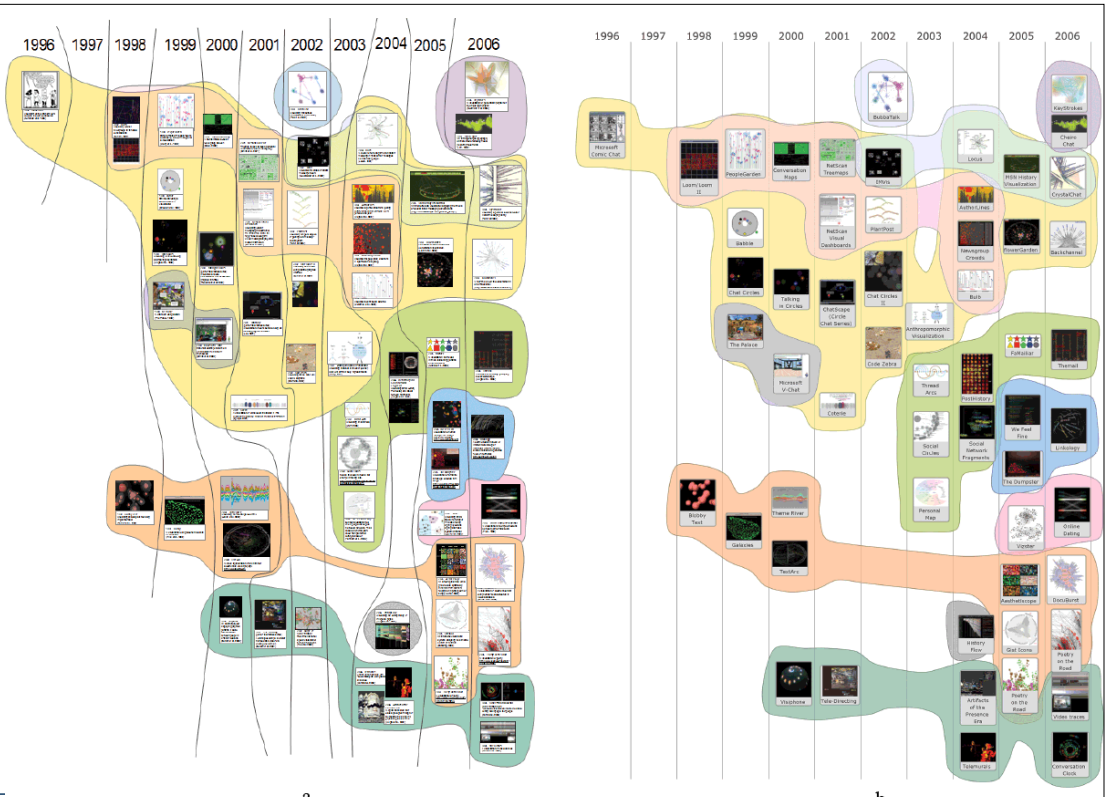
Venn Diagram



Contains all possible zones of overlap

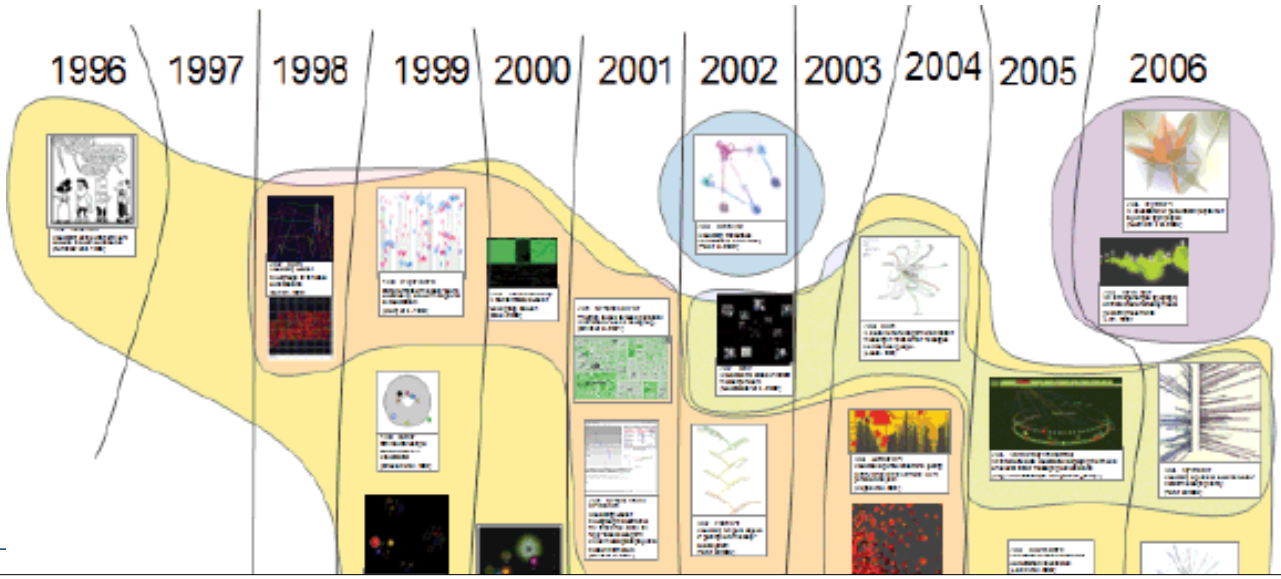
Bubble Sets

Collins et al
TVCG (InfoVis) '09



Bubble Sets

Collins et al
TVCG (InfoVis) '09



Georgia Tech

Minimum Possible?

- We have data cases with variables
- What's the smallest visual representation we can use?

… a pixel! Dense Pixel Display

- Represent **data case** or a variable as a **pixel**
- Million or more per display
- Seems to rely on use of color
- Can pack lots in

- Challenge: What's the layout?

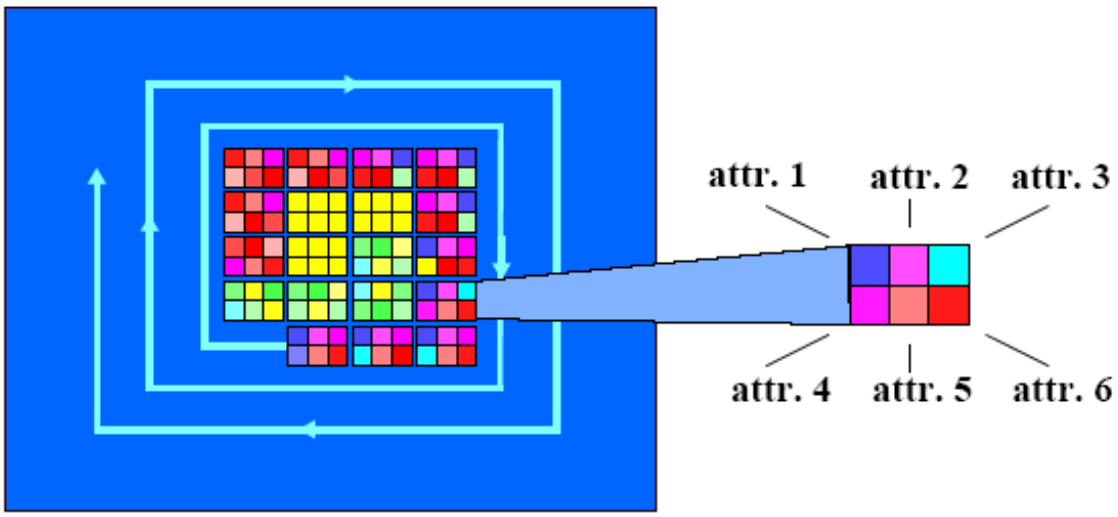
One Representation

- Grouping arrangement
- One pixel per variable
- Each data case has its own small rectangular icon
- Plot out variables for data point in that icon using a grid or spiral layout

Uses color scale

Illustration

schematic representation of 6-dim. data



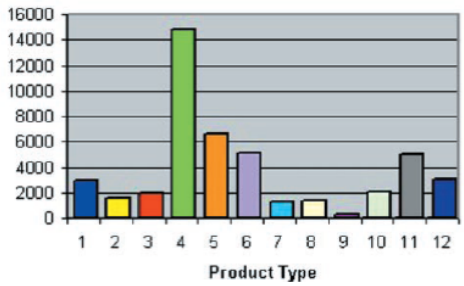
Levkowitz
Vis '91

Related Idea

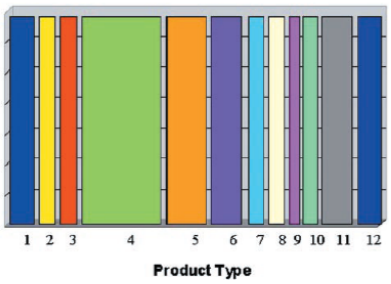
- Pixel Bar Chart
- Overload typical bar chart with more information about individual elements

Keim et al
Information Visualization '02

Idea 1



Height encodes quantity



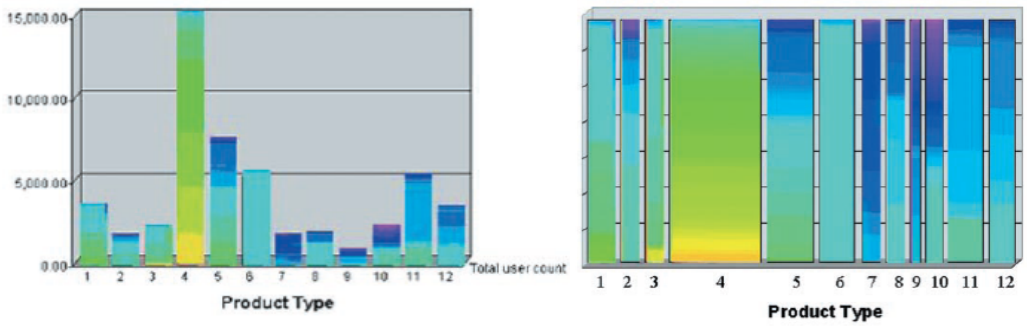
Width encodes quantity

space splitting/filling, technique, similar to mosaic.

Idea 2

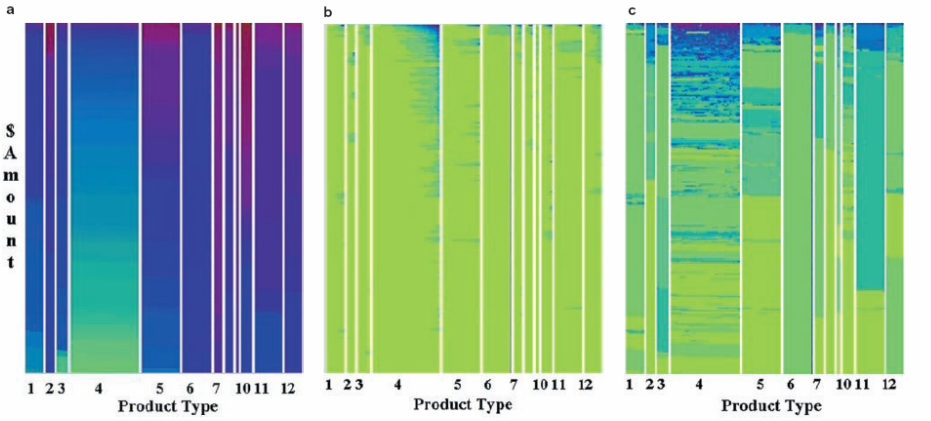
- Make each **pixel** within a bar correspond to **a data point** in that group represented by the bar
  - Can do millions that way
- Color the pixel to represent the value of one of the data point's variables

Idea 3



Each pixel is a customer
Color encodes amount spent by that person
High-bright, Low-dark
Ordered by that color attribute too
Right one shows more customers

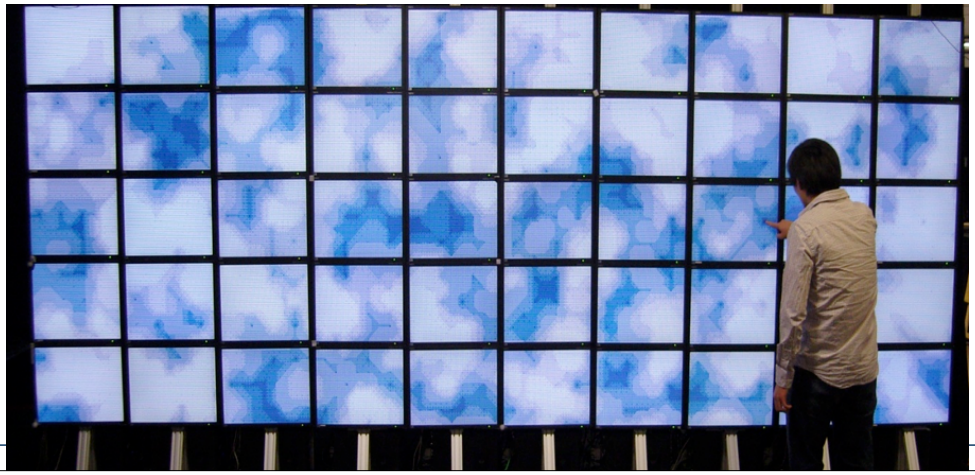
Idea 4



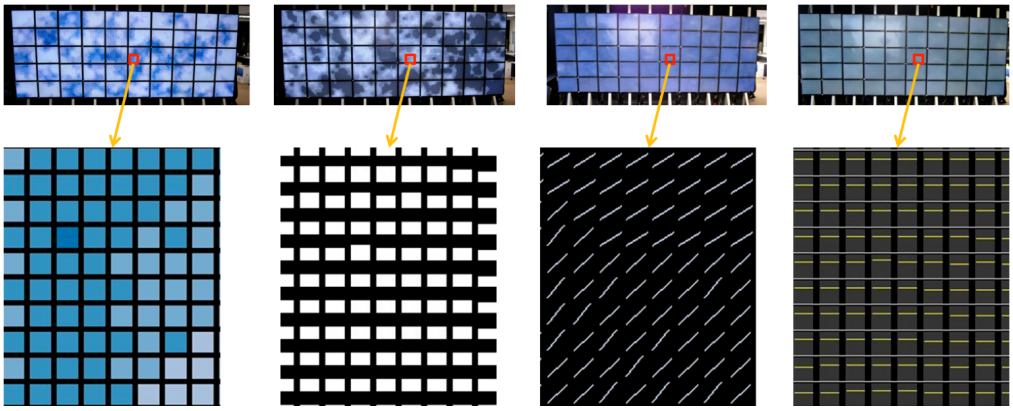
Product type is x-axis divider
Customers ordered by
y-axis: dollar amount
x-axis: number of visits
Color is (a) dollar amount spent, (b) number of visits, (c) sales quantity

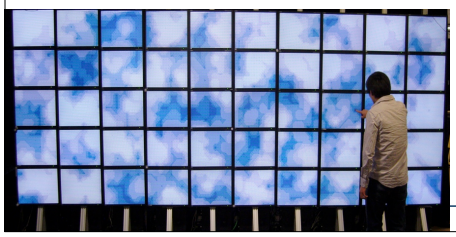
What about scaling up?

- study the effectiveness of specific visual encodings for large displays
  - Endert et al., 2011



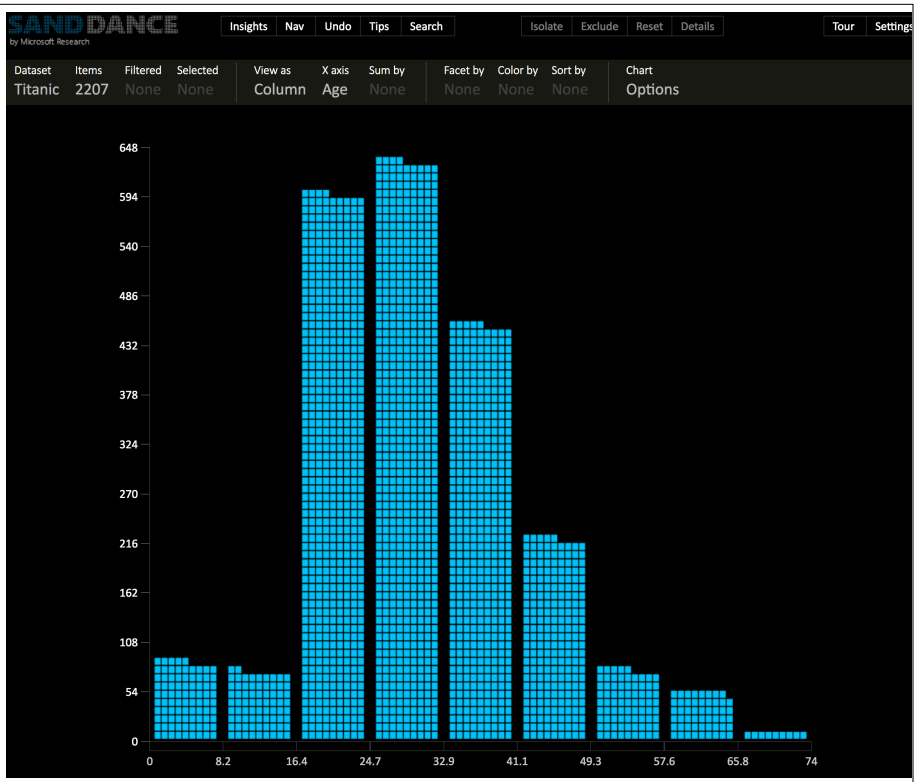
What does this do for visual encodings?





Sand Dance

https://sanddance.azurewebsites.net/BeachPartyApp/BeachPartyApp.html



Same data, different vis

- P5 examples
- Read over the description if you haven't
- Here are a few examples from previous years
  - http://vis.gatech.edu/demo/value/vis133/
  - http://vis.gatech.edu/demo/value/vis460/
  - http://vis.gatech.edu/demo/value/vis745/
  -

Upcoming

- Design Principles