Multivariate Visual Representations 2

CS 4460 – Intro. to Information Visualization
Sep. 20, 2017
John Stasko

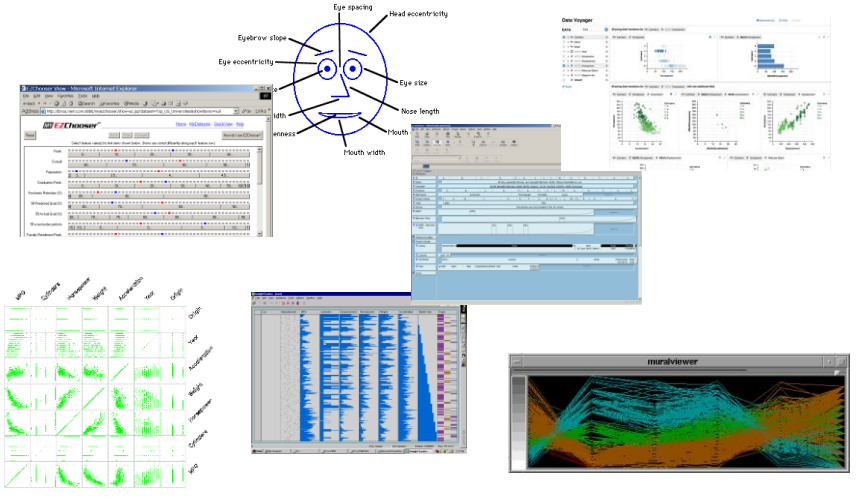
Learning Objectives

- Explain the concept of dense pixel/small glyph visualization techniques
- Describe each of the following examples of that technique and list their unique properties
  - Pixel bar chart, Dust 'n Magnet, Kinetica, SandDance
- Explain the potential benefits and drawbacks of these approaches
- Describe the "set visualization" problem and explain what a Venn Diagram and an Euler Diagram are
- Describe different approaches for set visualization when the number of sets and elements grow larger
- Understand where to turn for assistance with visualizing "Big data"

Recap

- We examined a number of techniques for projecting >2 variables (modest number of dimensions) down onto the 2D plane
  - Iconic displays
  - Table lens
  - Parallel coordinates
  - etc.

Variety of Techniques



Can We Make a Taxonomy?

- D. Keim proposes a taxonomy of techniques
  - Standard 2D/3D display
    - Bar charts, scatterplots
  - Geometrically transformed display
    - Parallel coordinates
  - Iconic display
    - Needle icons, Chernoff faces
  - Dense pixel display
    - What we're about to see…
  - Stacked display
    - Treemaps, dimensional stacking (coming later…)

TVCG '02

Minimum Possible?

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

Dense Pixel Display

- Represent data case or a variable as a pixel (or as a small glyph such as a circle)
- Million or more per display
- Seems to rely on use of color
- Can pack lots in

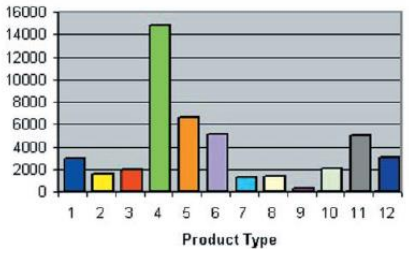
- Challenge: What's the layout? What does position mean?

Related Idea

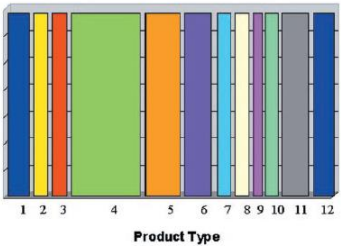
- Pixel Bar Chart
- Overload typical bar chart with more information about individual elements

Keim et al
Information Visualization `02

Idea 1



Height encodes quantity



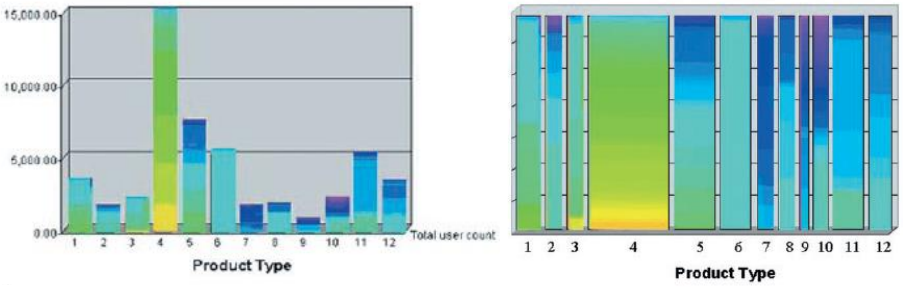
Width encodes quantity

What is the negative?

Idea 2

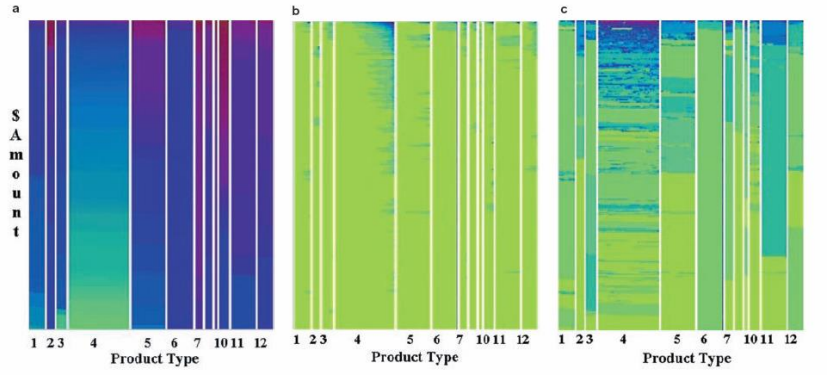
- Make each pixel within a bar correspond to a data point in that group represented by the bar
  - Can do millions that way
- Color the pixel to represent the value of one of the data point's variables

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

Next Step

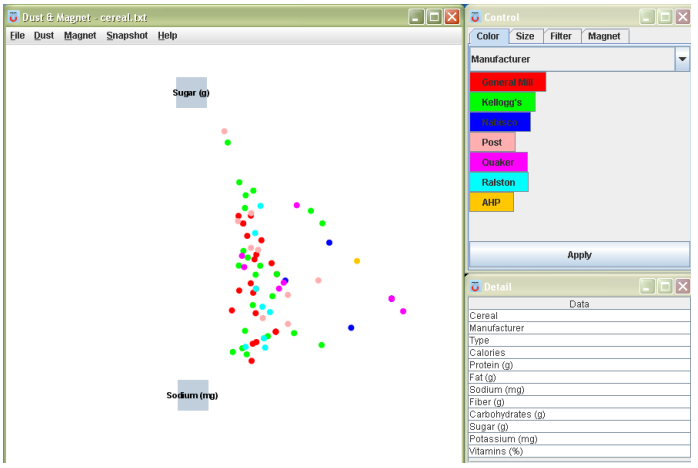
- Use a little more room to represent each data case
  - Make each a small glyph such as a circle
- Position of each still important
- Interaction likely becomes a crucial part of the visualization

Dust & Magnet

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

Yi, Melton, Stasko & Jacko
Information Visualization '05

Interface



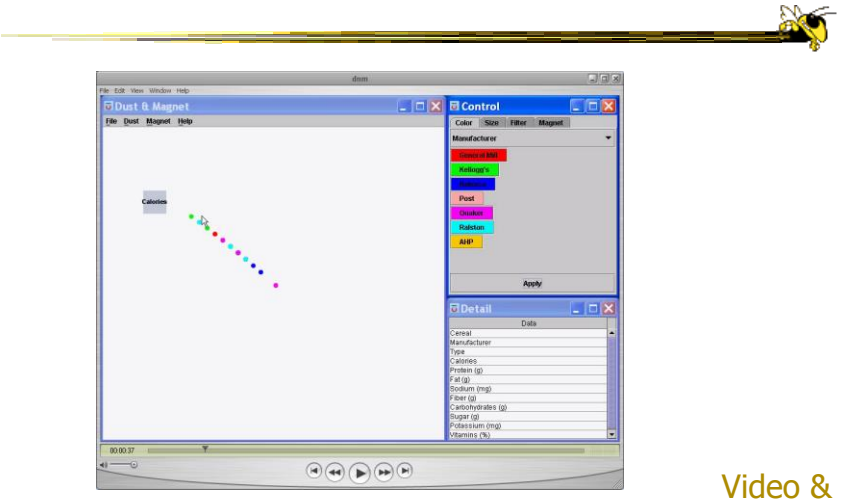
Interaction

- 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

See It Live

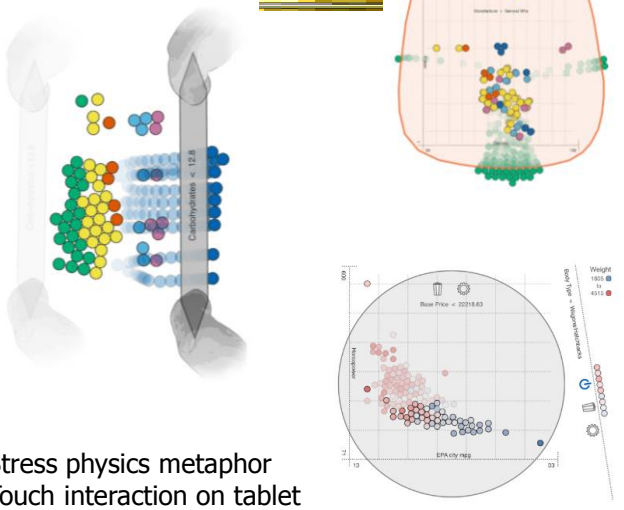


ftp://ftp.cc.gatech.edu/pub/people/stasko/movies/dnm.mov

Video & Demo

Kinetica

Video



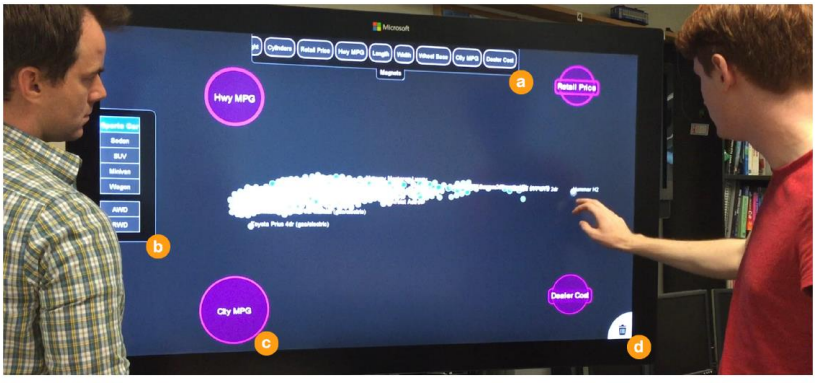
Stress physics metaphor
Touch interaction on tablet

Rzeszotarski & Kittur
CHI '14

Go Big

Video

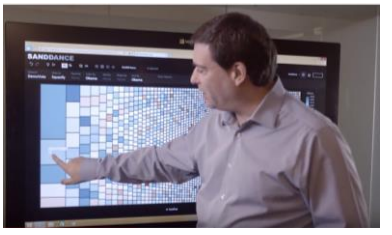
Dust & Magnet on a large multitouch display



Dai, Sadana, Stolper & Stasko
InfoVis '15 Poster

Sand Dance

- Data items as small squares
- Can position and color based on different attributes
- Multiple layouts provided
- Slick animated transitions

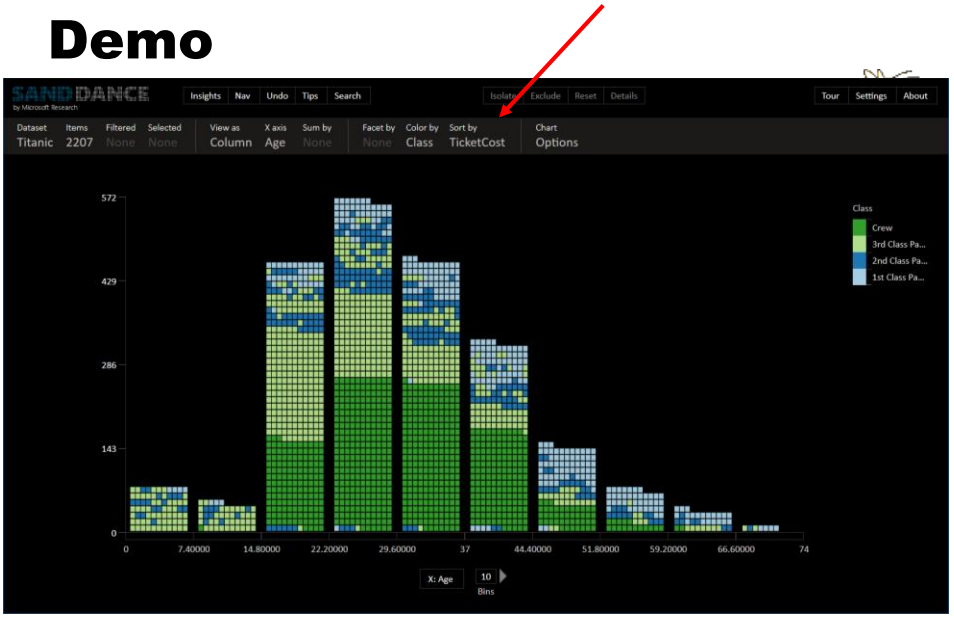


https://www.microsoft.com/en-us/research/project/sanddance/
https://www.youtube.com/watch?v=15Hns2igiag

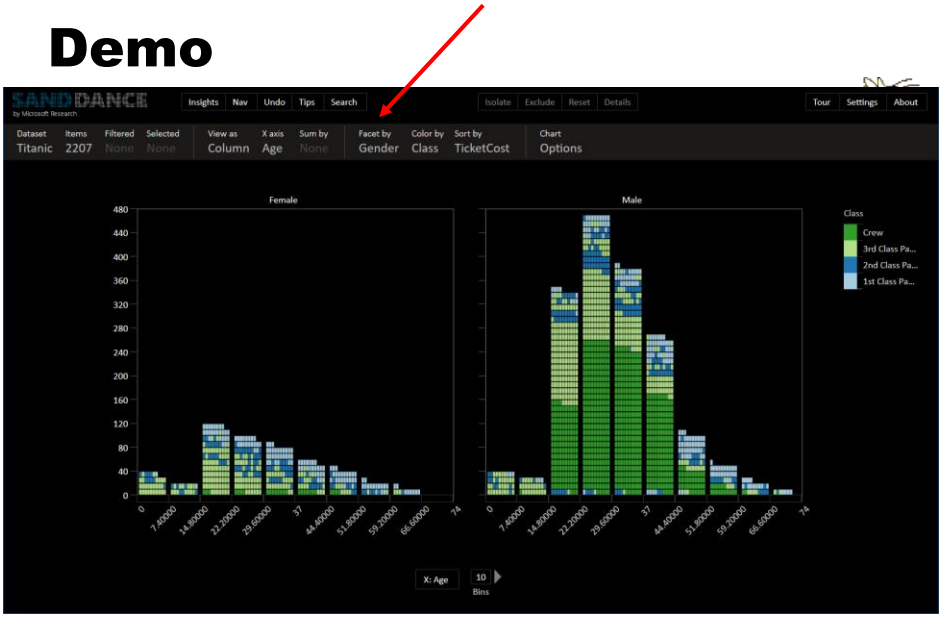
Demo

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

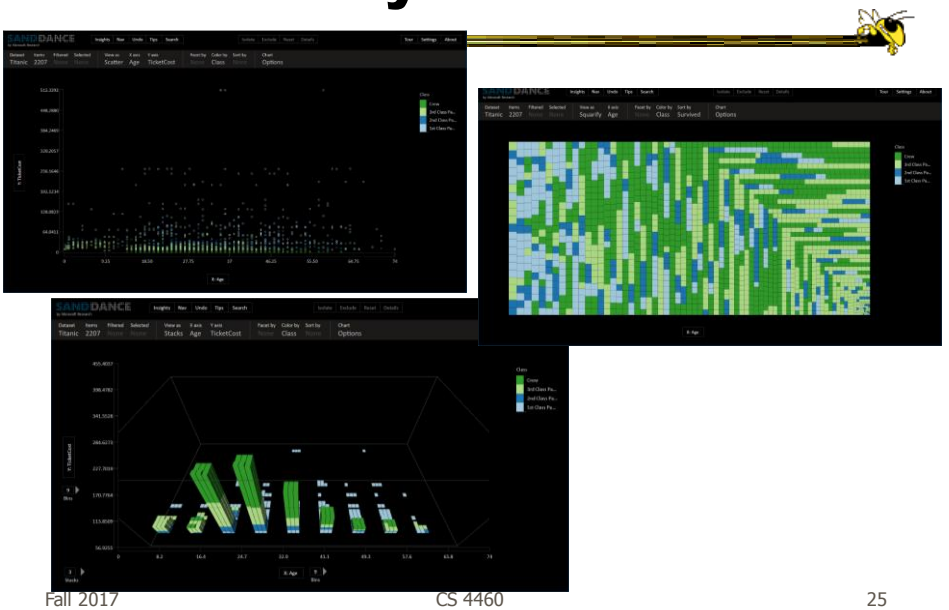
Demo


SANDDANCE
by Microsoft Research
Insights
Nav
Undo
Tips
Search
Isolate
Exclude
Reset
Details
Tour
Settings
About
Dataset
Items
Filtered
Selected
View as
X axis
Sum by
Facet by
Color by
Sort by
Chart
Titanic
2207
None
None
Column
Age
None
None
Class
TicketCost
Options
572
429
286
143
0
0
7.40000
14.80000
22.20000
29.60000
37
44.40000
51.80000
59.20000
66.60000
74
Class
Crew
3rd Class Pa...
2nd Class Pa...
1st Class Pa...
X: Age
10
Bins

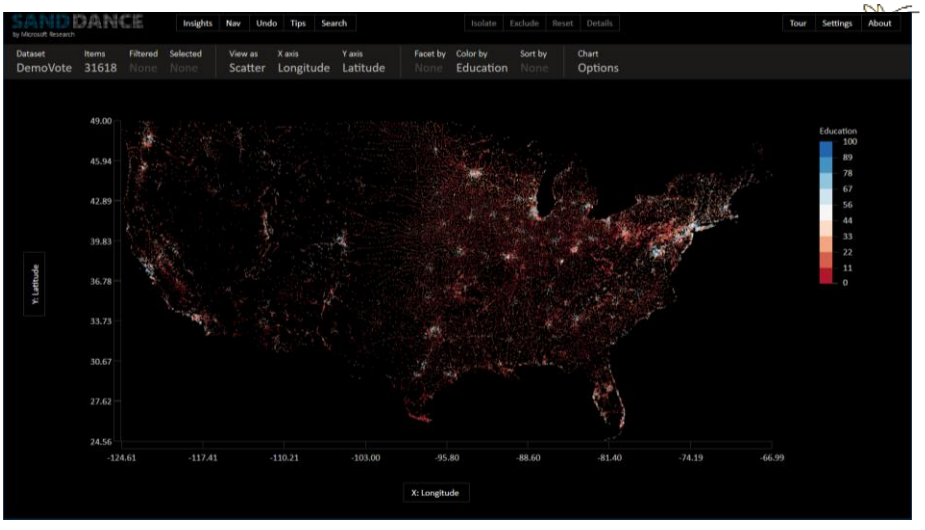
Demo


SANDDANCE
by Microsoft Research
Insights
Nav
Undo
Tips
Search
Isolate
Exclude
Reset
Details
Tour
Settings
About
Dataset
Items
Filtered
Selected
View as
X axis
Sum by
Facet by
Color by
Sort by
Chart
Titanic
2207
None
None
Column
Age
None
Gender
Class
TicketCost
Options
Female
Male
480
440
400
360
320
280
240
200
160
120
80
40
0
0
7.40000
14.80000
22.20000
29.60000
37
44.40000
51.80000
59.20000
66.60000
74
Class
Crew
3rd Class Pa...
2nd Class Pa...
1st Class Pa...
X: Age
10
Bins

Different Layouts



Geo for "Free"



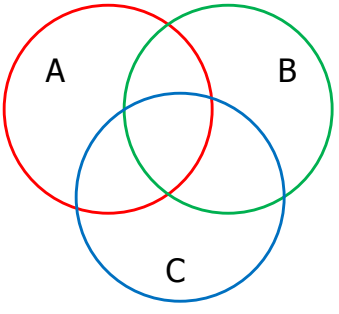
Scatterplot with x-longitude, y-latitude

Set Data & Operations

- Different type of problem
  - Large set of items, each can be in one or more sets
  - How do we visually represent the set membership?
  - What's the well-known technique?

Standard Technique

Venn Diagram

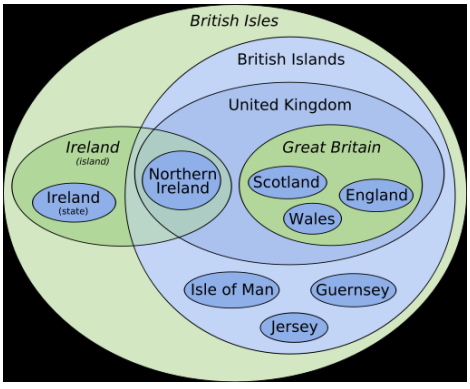


Contains all possible zones of overlap

Alternately

Euler Diagram

Does not necessarily show all possible overlap zones



http://en.wikipedia.org/wiki/File:British_Isles_Euler_diagram_15.svg

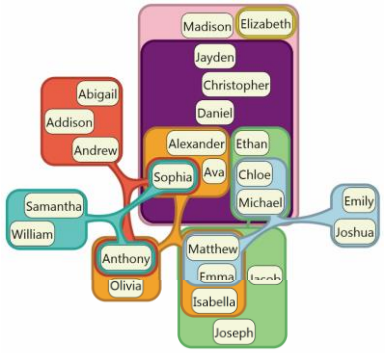
But what's the problem?

Bubble Sets

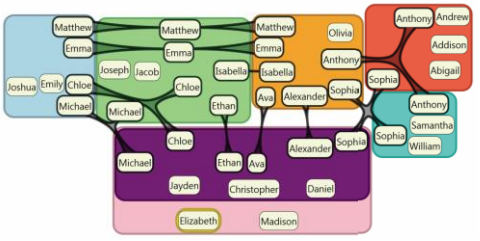
Video

Collins et al
TVCG (InfoVis) '09

ComED & DupED



Item appears once



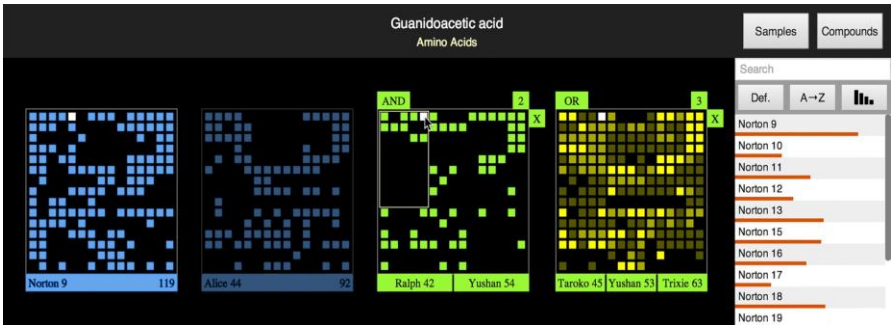
Item can appear more than once

Video

Riche & Dwyer
TVCG (InfoVis) '10

Another Set Problem

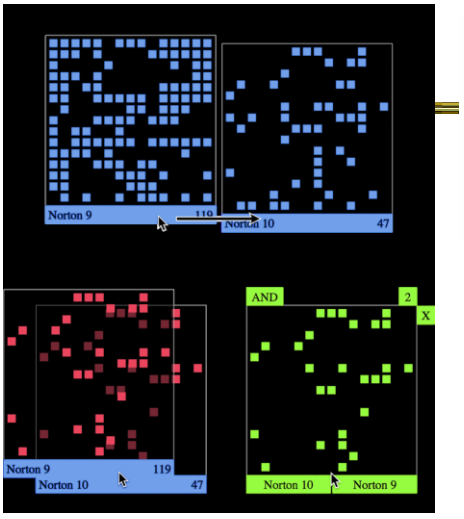
OnSet



Represent set as a box, elements are spots in that box
Use interaction to do set union, intersection

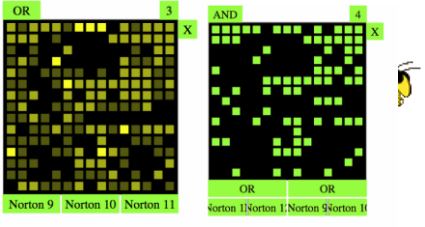
Sadana, Major, Dove & Stasko
TVCG (InfoVis) '14

Demo/video

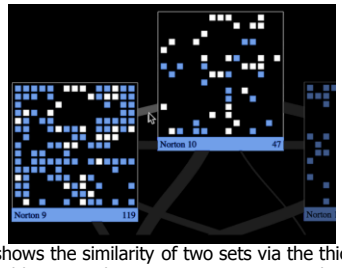


Dragging and dropping a PixelLayer to create a new AND MultiLayer.

`http://www.cc.gatech.edu/gvu/ii/setvis`



A MultiLayer OR with three sets.
A MultiLayer AND of nested OR layers.



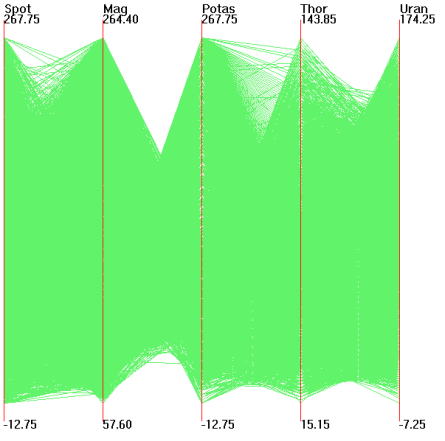
OnSet shows the similarity of two sets via the thickness of a band between them. Hovering over a similarity band highlights the common elements between two sets.

Step Back

- Most of the techniques we've examined work for a modest number of data cases or variables
  - What happens when you have lots and lots of data cases and/or variables?

Many Cases

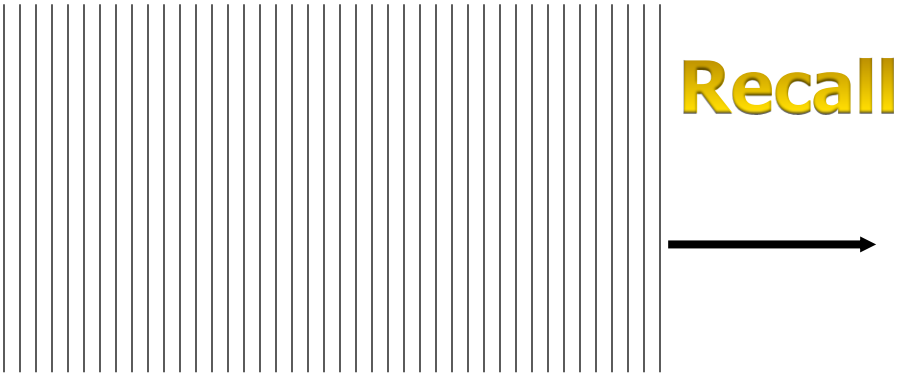
Recall



Out5d dataset (5 dimensions, 16384 data items)

(courtesy of J. Yang)

Many Variables



Strategies

- How are we going to deal with such big datasets with so many variables per case?
- Ideas?

General Notion

- Data that is similar in most dimensions ought to be drawn together
  - Cluster at high dimensions
- Need to project the data down into the plane and give it some ultra-simplified representation
- Or perhaps only look at certain aspects of the data at any one time

Mathematical Assistance 1

- There exist many techniques for clustering high-dimensional data with respect to all those dimensions
  - Affinity propagation
  - k-means
  - Expectation maximization
  - Hierarchical clustering

Mathematical Assistance 2

- There exist many techniques for projecting n-dimensions down to 2-D (dimensionality reduction)
  - Multi-dimensional scaling (MDS)
  - Principal component analysis
  - Linear discriminant analysis
  - Factor analysis

Comput Sci & Eng courses
Data & Visual Analytics, Prof. Chau

Data mining
Knowledge discovery

Other Techniques

- Other techniques exist to manage scale
  - Sampling – We only include every so many data cases or variables
  - Aggregation – We combine many data cases or variables
  - Interaction (later)
    - Employ user interaction rather than special renderings to help manage scale

Use?

- What kinds of questions/tasks would you want such techniques to address?
  - Clusters of similar data cases
  - Useless dimensions
  - Dimensions similar to each other
  - Outlier data cases
  - …
- Think about the "cognitive tasks" we want to accomplish

Recap

- We've seen many general techniques for multivariate data these past two days
  - Know strengths and limitations of each
  - Know which ones are good for which circumstances
  - We still haven't explored interaction much

HW 3

- Due Friday
- Submit one copy
- Questions?

Quick Tip

- If you send me an email, include "4460" in the Subject line somewhere
  - Just easier for me to find these then

Learning Objectives

- Explain the concept of dense pixel/small glyph visualization techniques
- Describe each of the following examples of that technique and list their unique properties
  - Pixel bar chart, Dust 'n Magnet, Kinetica, SandDance
- Explain the potential benefits and drawbacks of these approaches
- Describe the "set visualization" problem and explain what a Venn Diagram and an Euler Diagram are
- Describe different approaches for set visualization when the number of sets and elements grow larger
- Understand where to turn for assistance with visualizing "Big data"

Upcoming

- Lab 2: SVG
  - Prep: Murray section on SVG, Soueidan article
- Systems & Toolkits