Information Visualization

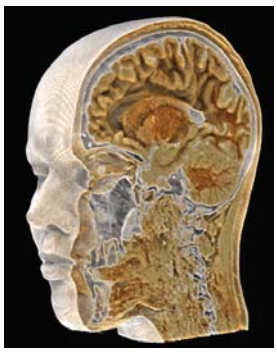
Matt Cooper

Visualization?

- **New Oxford Dictionary of English, 1999**
- **visualize** - ***verb*** **[with *obj.*]**
  1. **form a mental image of; imagine: *it is not easy to visualize the future.***
  2. **make (something) visible to the eye: *the DNA as visualized by staining with ethidium bromide.***

Ceci n'est pas une visualization!

- This is not a visualization!
- This is a picture!
  - A very useful picture!
- The visualization is all in your mind.

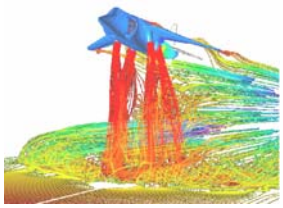


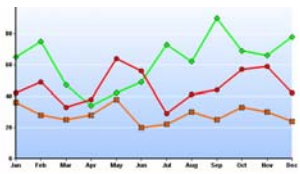
Information Visualization?

- Techniques to enable the visualization of Information
- So what is Scientific Visualization?
  - Scientific data doesn't contain information?!?!?!
- What's the difference?
  - The application areas are different?
  - The data is different?
  - The needs of the users are different?

So why two disciplines?

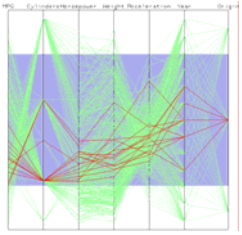
- Information Visualization exists as a discipline due to a superiority complex.
  - "My pictures are nicer than your pictures"





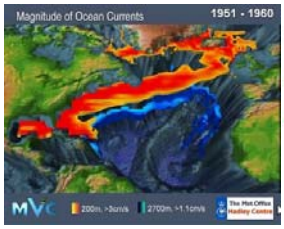
So why two disciplines?

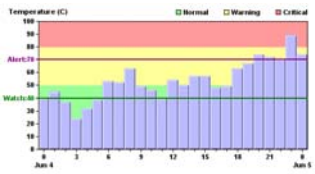
- Information Visualization exists as a discipline due to a superiority complex.
  - "My pictures are nicer than your pictures"



So why two disciplines?

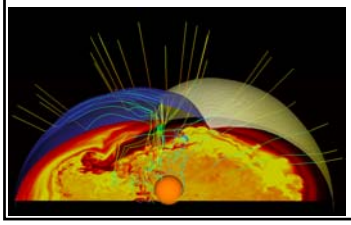
- Information Visualization exists as a discipline due to a superiority complex.
  - "My pictures are nicer than your pictures"





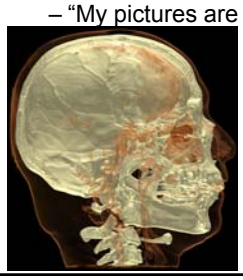
So why two disciplines?

- Information Visualization exists as a discipline due to a superiority complex.
  - "My pictures are nicer than your pictures"



So why two disciplines?

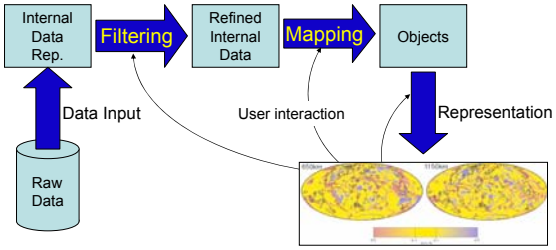
- Information Visualization exists as a discipline due to a superiority complex.
  - "My pictures are nicer than your pictures"



Two disciplines

- So, we end up with two "distinct" disciplines in the same area.
  - Doing the same thing
  - Focussing in different ways
- How are they really separated?
  - The application areas are usually different
  - The needs of the users are usually different
  - The data is different

The Visualization Pipeline



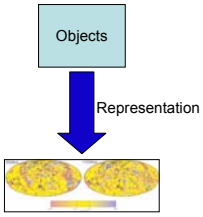
The Viz pipeline: Filtering

- Selecting data to seek specific relations
- Reducing data to manageable sizes
- Modifying the internal structure

The Viz pipeline: Mapping

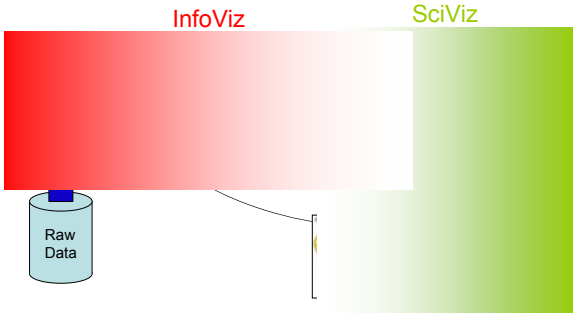
- This is the most complex phase
- Involves the definition of relationship representations
- Driven by models of human perception…

The Viz pipeline: Display



- Rendering of (simple) shapes.
- Simple for ease of interpretation.
- And ease of interaction
  - Important for exploring

So who gets clever where?



The (re-)merging of the disciplines

- SciViz is rapidly heading into InfoViz territory:
  - Sophisticated data analysis and pattern identification
  - Complex data representations
  - Sophisticated interaction
- Infoviz taking up methods from SciViz:
  - Volumetric rendering (translucent blocks of data)
  - Surfacing of data structures (clusters)
  - Tensor representations (high-dimensional data)

Why the different approaches?

- SciViz data is:
  - Small
  - Well understood
    - The data values are low dimensional
    - The geometry is often regular and always known
  - Clean (or at least predictably dirty)
  - Filled with well understood relationships
    - Simulated or measured data
    - Created or gathered for a purpose

Why the different approaches?

- InfoViz data is:
  - Big (hundreds of gigabytes is common)
  - May not be well understood
    - Data items have tens or hundreds of dimensions
    - The ‘geometry’ is usually next to non-existant.
  - Often very dirty (and unpredictably so)
  - Filled with (apparent) relationships
    - Correlations and anticorrelations
    - Patterns which exist only in certain regions in the data

A Comparative example: Heart disease

- SciViz problem: Is this patient sick?
  - Examine (scan) hearts of sick and well patients
  - Examine their hearts by visualizing features of the data:
    - muscle, fluid, blood flow, fatty tissue…
  - Determine what features correlate with heart disease
- Test new patient for those features

A Comparative example: Heart disease

- InfoViz problem: Who's going to get sick?
  - Collect detailed multi-generation histories of similar-seeming patients who are/aren't sick
    - Think of the endless possible measurements!
  - Explore that data looking for patterns of factors which correlate with sickness and not wellness
  - Create a model of behaviour(s) which produce a relatively high risk
- Tell media – who then terrify the public

So The InfoVis problem is…

- Finding (very) complex relationships…
- Which may not be present everywhere
- Within very large…
- High-dimensional…
- Data sets…
- That you may not have known you were looking for in the first place

A 'haystack' problem:

"See that haystack? Go and see if there's anything interesting about it."

InfoVis example application

- Dr. John Snow's Cholera diagram
- Exploratory Data Analysis
  - Also Visual Data Mining
- Done without knowing what caused Cholera!

Graphical Displays

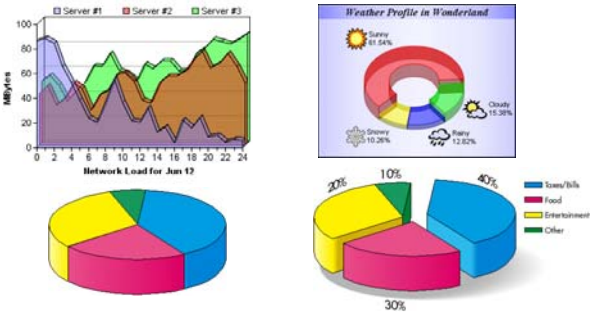
- Line plots
- Scatter plots
- Bar charts
- Pie charts
- Stacked charts
- Star plots
- Donut charts
- Various glyphs
- Maps – very trendy!
- Parallel coordinates

- Almost always 2D
  - 3D is trouble!
- Combined in diverse ways

Displays

- Simplistic – as simple a picture as possible
- Minimalist – as little UI as possible
- Highly interactive
  - Complex selection mechanisms
  - Filtering
- Highly coordinated (multiple views)

3D bad!...

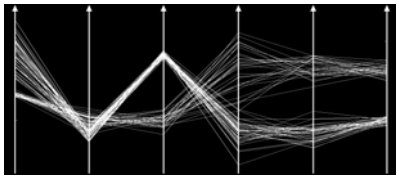


…when there is no logical mapping from the data to a 3D space

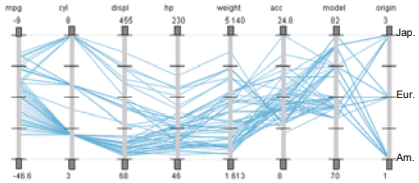
Interaction

- Project carried out by Daniel Ericsson
  - Exjobb project based on a PhD student's work
- Coordinated and multiple views
- Using parallel coordinates and animation to convey changes when selecting and filtering

Parallel Coordinates — Clusters (6D example)

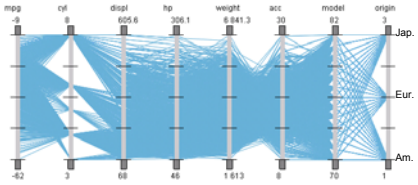


Parallel Coordinates



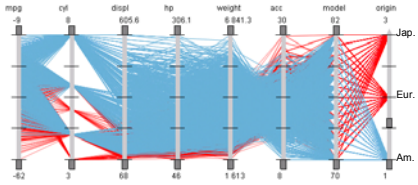
- 50 objects from the data set of cars
- PC shows structure of the data

Parallel Coordinates



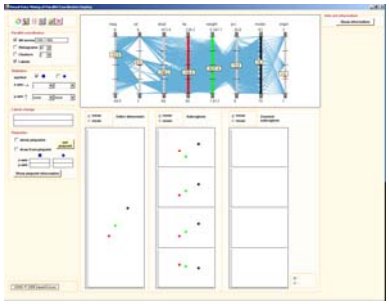
- 3000 objects
- Overlapping lines
- Cluttered PC display

Parallel Coordinates



- Good for detecting correlations, but
- Hard to see overall distribution of values
- Does not always scale well with large data sets.
- Changes are hard to track

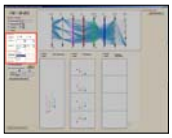
Visual Data Mining Display (VDMD)



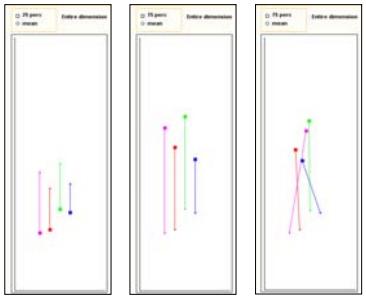
Application and GUI for examining and exploring data displayed with parallel coordinates.

Extract and display statistics from selected dimensions in a separate view.

Change of variables



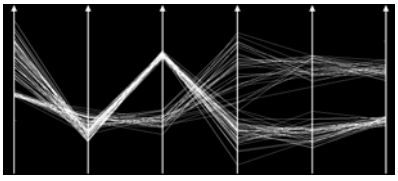
x-axis: median



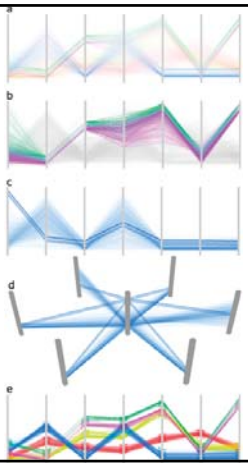
Volume rendering in IV

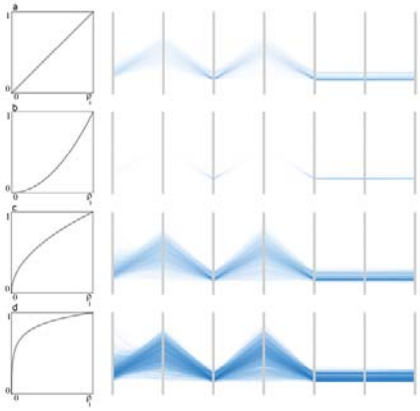
- Jimmy Johansson – PhD student
- Using techniques of volume rendering to examine large, time-varying data sets.
- Used fragment shaders and graphics hardware to render data sets and then manipulate them in real time
- Very large data sets in real time.

Parallel Coordinates — Clusters (6D example)



Jimmy's PC clusters…

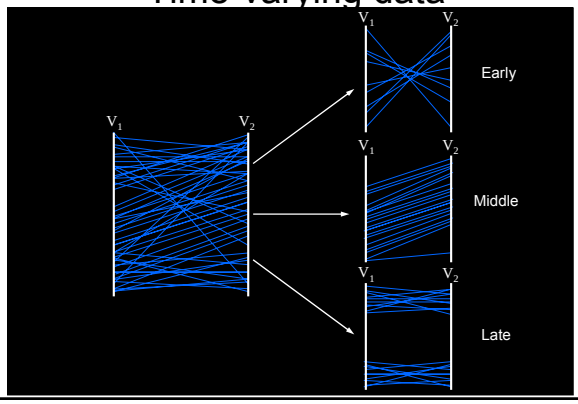




Time-varying data

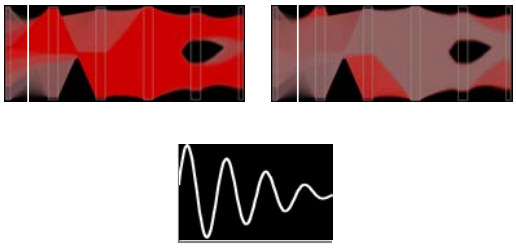
- Time-varying data is a big challenge in visualization
  - Expands data sizes
  - Often hard to represent
  - People are bad at remembering features
- Volumetric representation allows multiple time steps to simultaneously displayed

Time-varying data



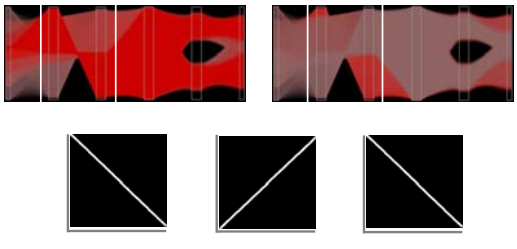
Results (100 items, 500 time steps)

- Depth cue parallel coordinates



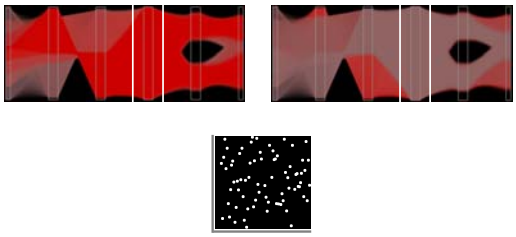
Results (100 items, 500 time steps)

- Depth cue parallel coordinates



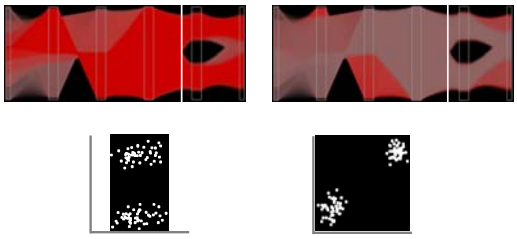
Results (100 items, 500 time steps)

- Depth cue parallel coordinates

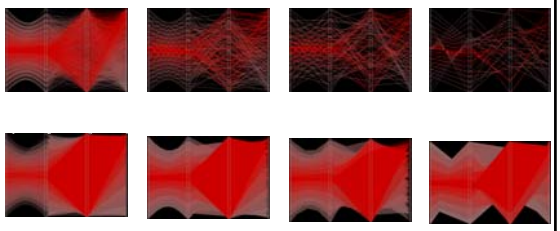


Results (100 items, 500 time steps)

- Depth cue parallel coordinates



Also allows data reduction



PC through volume graphics

- Lets us:
  - Visualize actual data values at each time step
  - Visualize changes between time steps
- Allows Interactive analysis of large multivariate data sets over thousands of time steps
- Permits us to reduce the data with less loss of information

Data Mining

- Approach to find relationships in data
  - That you don't know are there
  - That the data was not collected for
- 'Post-Analysis'
- Very interesting from the point of view of many sciences, businesses and other fields.

Visual Data Mining

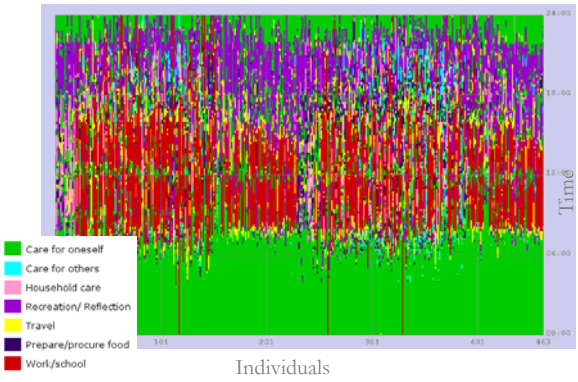
- The process of data mining can be
  - Done by InfoViz – Visual Data Exploration
    - Not really practical though – small data only
  - Vastly aided by visualization
    - Patterns extracted are the representations to show
    - InfoViz helps you interpret and analyse the patterns
      - Valid/Invalid
      - Interesting/Uninteresting

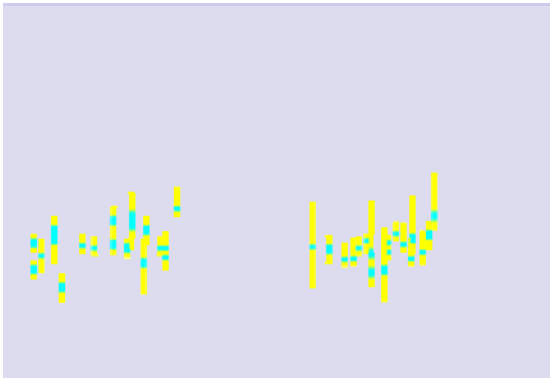
Everyday Life Discoveries:
Mining and Visualizing Activity Patterns in Social Science Diary Data

Katerina Vrotsou, NVIS, Linköping University, Sweden
Kajsa Ellegård, Tema-T, Linköping University, Sweden
Matt Cooper, NVIS, Linköping University, Sweden

Patterns in Social Science

- Everyday life exploration
  - Individuals, groups, populations
- Time-use data, activities diaries
  - Series of activities
  - Large databases





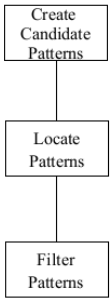
Patterns in Social Science

- Activity patterns
  - Activity sequences that aim to achieve larger projects
  - "drive car" →"drop off child" → "drive car"
    ⇨ *get child to school*
  - Work time and division of labour

Our Objectives

- To combine:
  - Sequential data mining
  - Visualization and
  - Interaction
- to enable the user to study the everyday lives of populations and analyse their activity patterns.

Algorithm – adapted AprioriAll

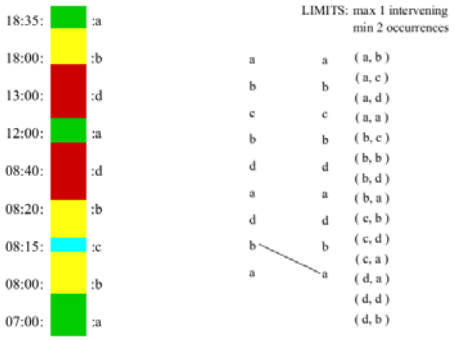


- **Join** existing patterns to create **new valid** candidate patterns
- Locate candidates in the activity data – **match**
- Filter matched patterns according to **user restrictions**

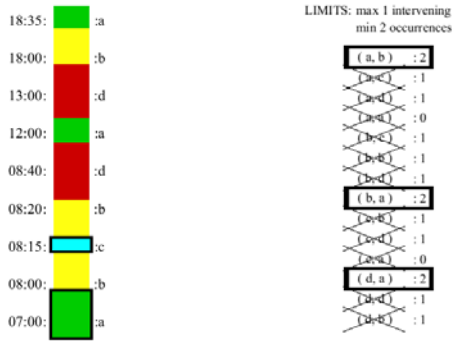
Filtering

- Time of Day Limit
- Pattern Duration Limit
- Minimum – maximum intervening activities
- Occurrences Limit
- Active People Limit

Algorithm: singles → doubles



Algorithm: singles → doubles

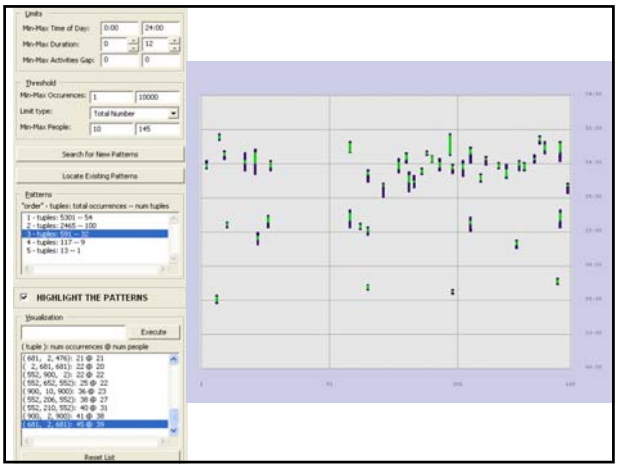


(a, a, b) (a, a, c) (a, a, d) (a, b, a) (a, b, b) (a, b, c) (a, b, d) (a, c, a) (a, c, b) (a, c, d) (a, d, a) (a, d, b) (a, d, c) (a, d, d)
(b, a, a) (b, a, b) (b, a, c) (b, a, d) (b, b, a) (b, [illegible]) (b, [illegible]) (b, c, a) (b, [illegible]) (b, c, d) (b, d, a) (b, d, b) (b, d, c) (b, d, d)
(c, a, a) (c, a, b) (c, a, c) (c, a, d) (c, b, a) (c, [illegible]) (c, b, c) (c, [illegible]) (c, [illegible]) (c, d, b) (c, d, d)
(d, a, a) (d, a, b) (d, a, c) (d, a, d) (d, b, a) (d, b, b) (d, b, c) (d, b, d) (d, c, b) (d, c, d) (d, d, a) (d, d, b)

51

Choice of Algorithm

- Sequential data
  - Patterns = activity sequences
- Finds **all** patterns
  - Discover the unexpected
- Computation time
  - Candidate creation
  - Constraints



Conclusions

- Large number of candidate patterns
  - Large datasets → time consuming
- Filtering
  - Trivial patterns
  - *"sleep" – "wake up"*
- Alternative search algorithms
- Alternative representations

Conclusions to this talk

- InfoViz is far less focussed on computer graphics than SciViz
- More focussed on:
  - (combinations of) simple representations
  - Sophisticated representations
  - Sophisticated interaction
- It's often more mathematical and deals with even larger problems than SciViz
- The two are merging once again, however.