Multivariate Data Visualization

VOTech/Universities of Leeds & Edinburgh

Richard Holbrey

Outline

- Focus on data exploration

- Joined VOTech in June, so ...
- ... informal introduction
- Examine some definitions and assumptions
- Describe visual & data mining approaches in general

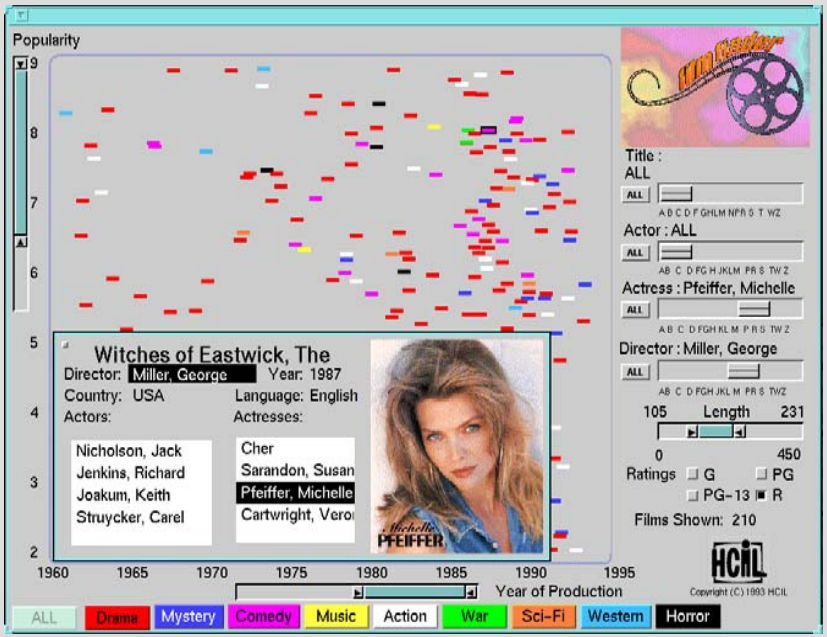
- Open discussion
- Some demos of simple vis tools
- Feedback

Definition: Visualization

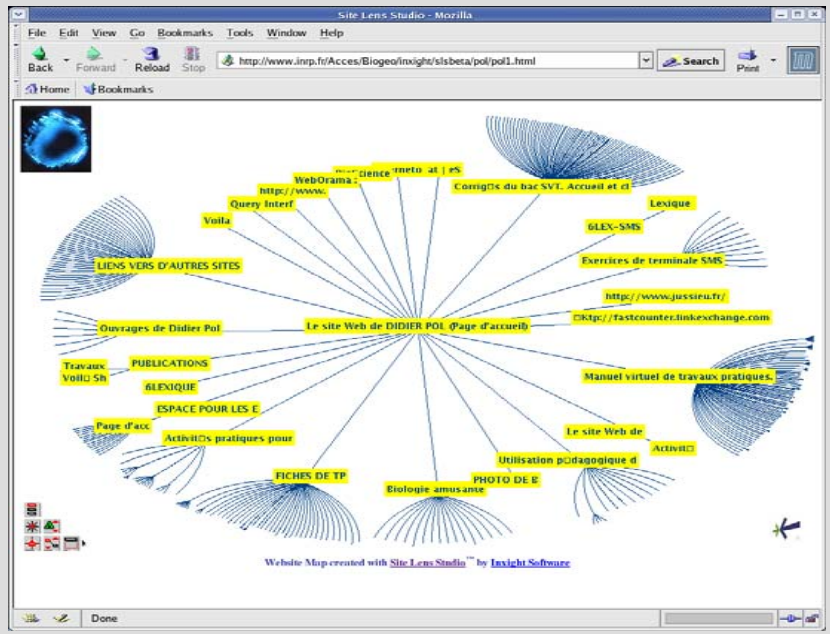
- Born as a computing discipline in 1987 with publication of NSF Report
- Gurus tell us:
  - “use of computer-supported, interactive, visual representations of data to amplify cognition”
    - (Card, McKinlay, Shneiderman)
- Emphasises human brain as pattern recognition tool
- ...but can only accept so much input
  - eg. active region of the eye
  - update rate (c. 25Hz eye, 1000Hz touch)

Building mental maps

- HCI
- FilmFinder tool
  - (Shneiderman et al)



- Web maps
  - www.inxight.com



Data mining

- Cynical view
  - Statistics with better marketing
- More mainstream
  - Practical way of dealing with large data sets
  - employing computing power to handle complexity (n^?)

In practice

- Classification rules/decision tree
  - eg. if *a* and *b* then *x*
- Association/Dissociation rules
  - eg. ‘loyalty’ cards
  - multiple, categorical data items
- Numeric data
  - clustering
  - regression
  - neural nets (‘black box’)
- Statistical modelling needed where uncertainty occurs

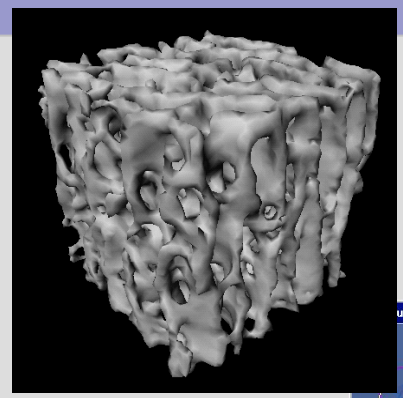
Straw poll

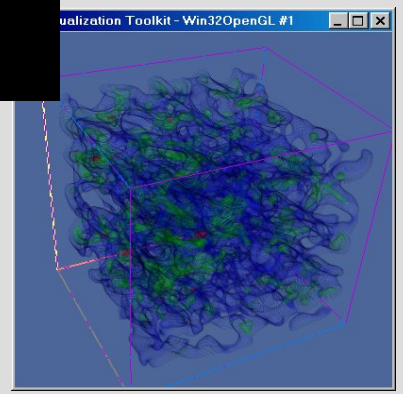
- I do pattern matching ‘by eye’
  - likely many spurious correlations otherwise
- Some tools popular with astronomers
  - IDL in the UK
- Mirage cf Weka cf Topcat (RM & MA)
  - need to
    - handle various formats
    - manipulate/add columns to table data
    - overlay sources and models graphically
- SRM doc
  - need for
    - models, function setting
    - visualization as an endpoint

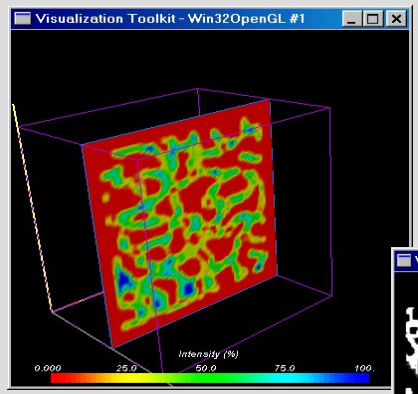
My assumptions

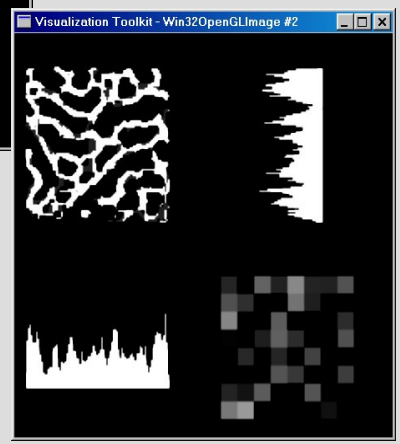
- Background
  - bone density study
  - and real-time VR surgical simulation
- Working in 3D or higher difficult
  - 6dof, focus, context
- 2D much more obvious...
- ..but can't lose 3D

- Sound? Haptics?







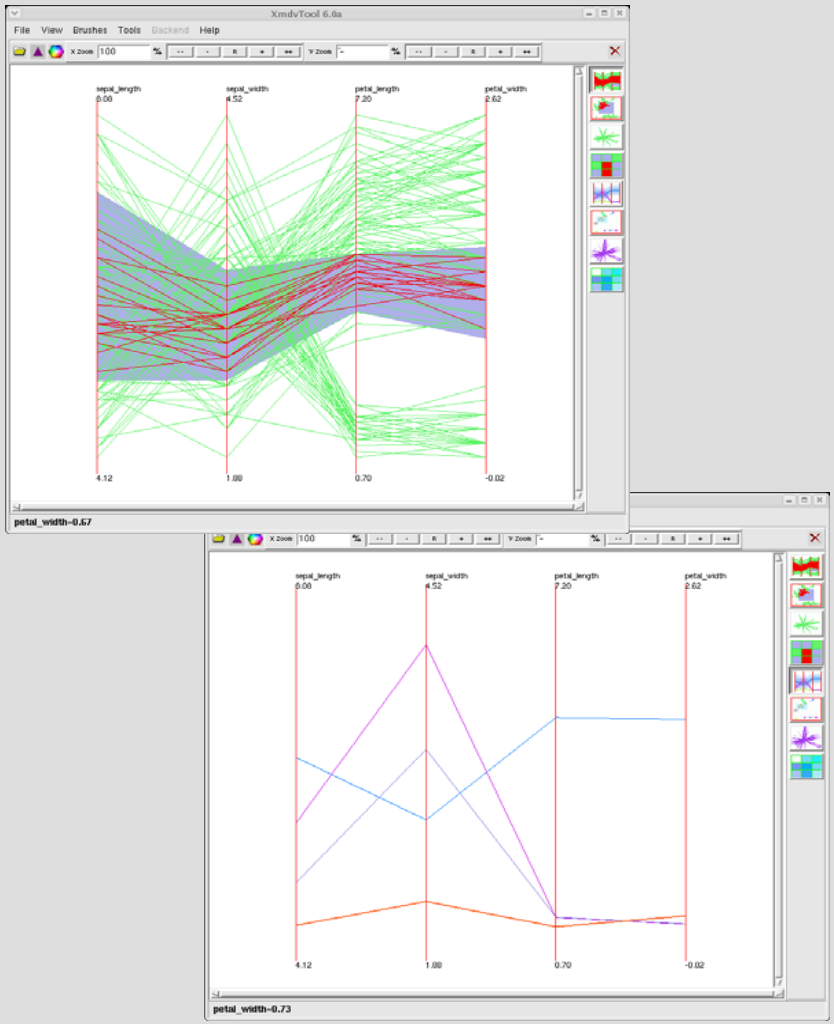


Some example techniques

Projections from nD ➐ 2D

- One of the oldest vis tricks
- Visual techniques
  - Parallel coordinates, dendrograms, MDS, stacking
- Projections
  - PCA, Kohonen maps, Sammon maps
  - Problem: tend to lose relationship with original variates
- Class-preserving
  - CViz approach

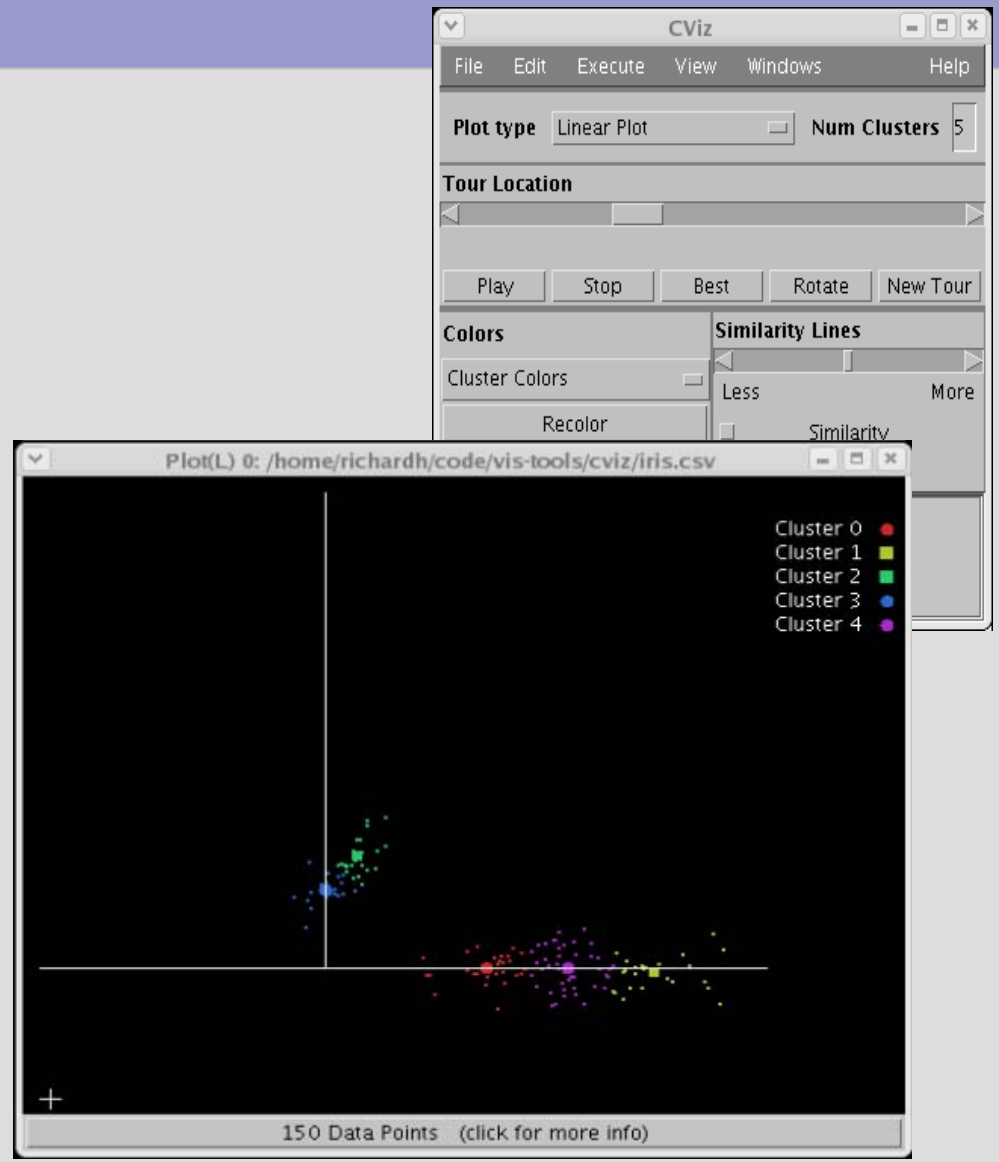
Parallel coordinates



- Xmdvtool
  - originally limited to 20 columns of data
  - practical limit
- Brushing and clustering
  - to compensate for screen real-estate

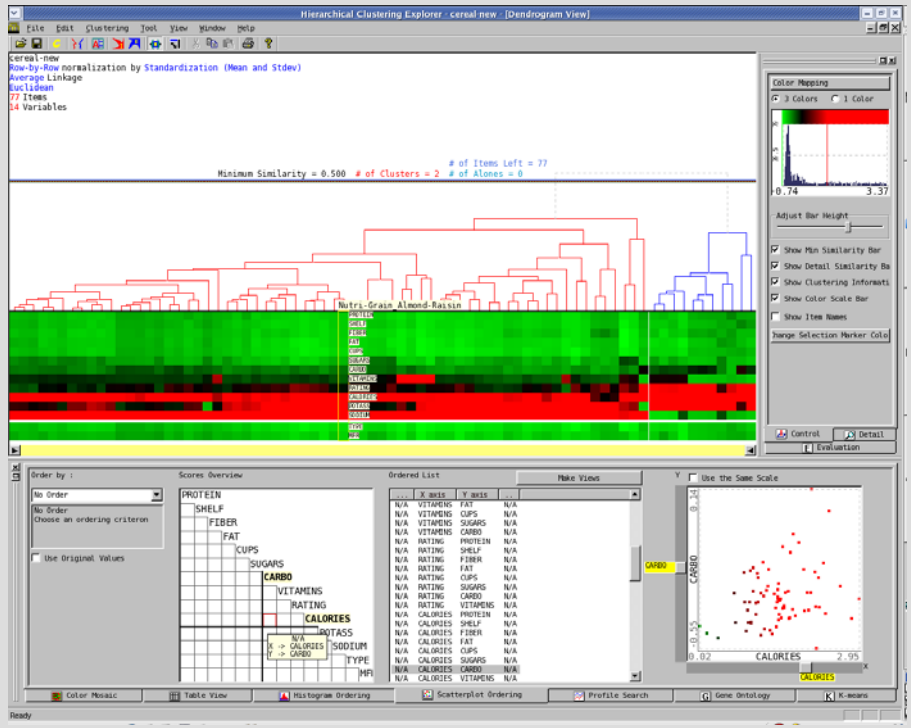
Class-projection

- CViz
  - aim is to tour through the data, through a series of 2D projections
  - only a sample of the data is shown
  - Authors tested 26 variables with some difficulty



Clustering-centred

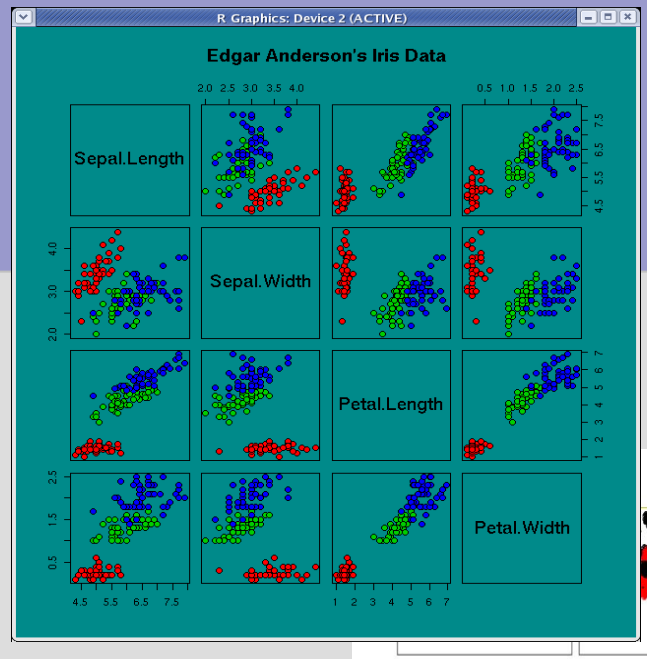
- HCE: Hierarchical Clustering Explorer
  - linked trees and display tools
- Copes well with ~10 variates

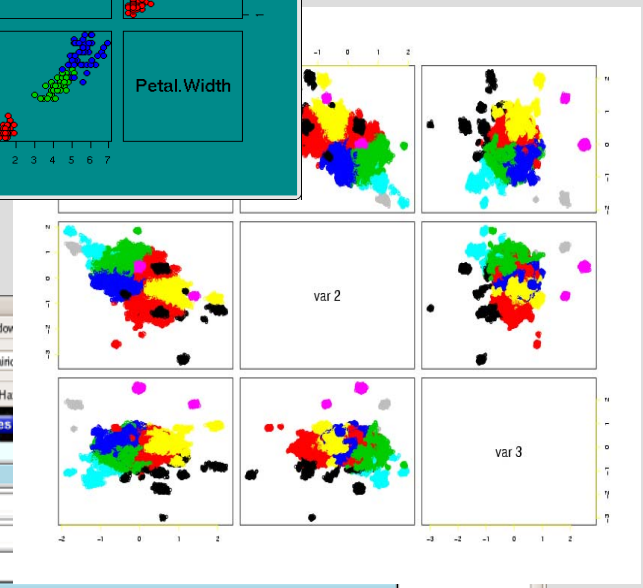


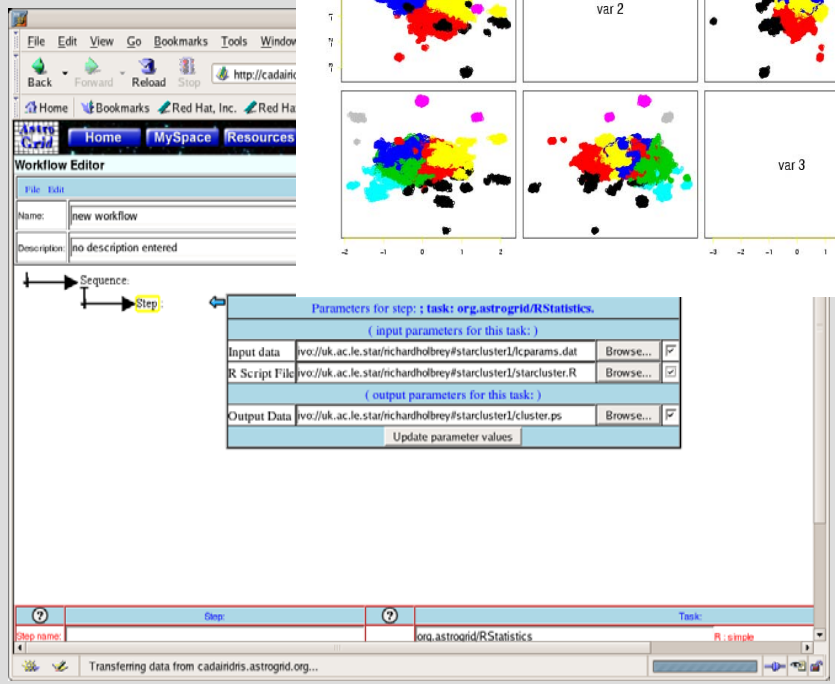
R

(http://cran.r-project.org)

- Large statistically-oriented library with many contributed packages
- Powerful vectorized scripting – can do almost anything
- Scales to large number of variates
- Command-line oriented: graphics, but non-interactive in the usual sense
- Can run under CEA!

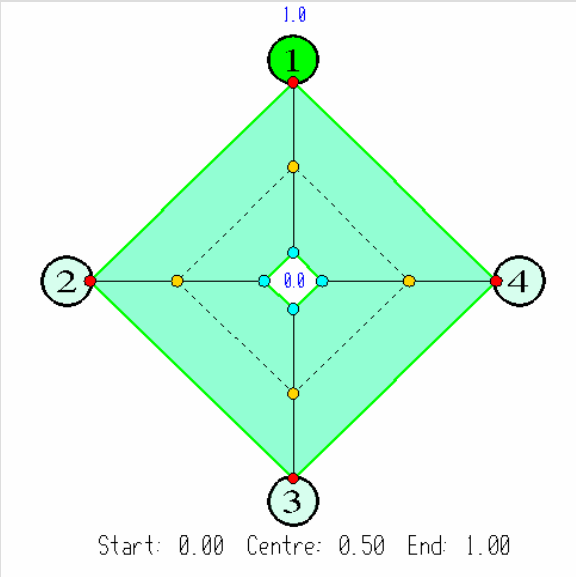




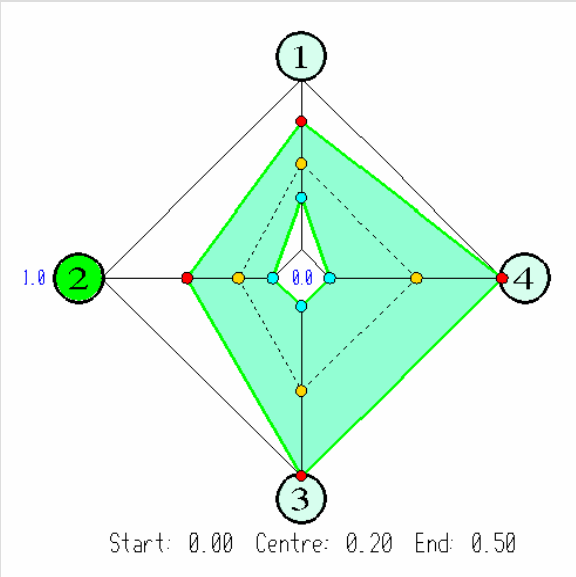


Work at Leeds - Hypercell

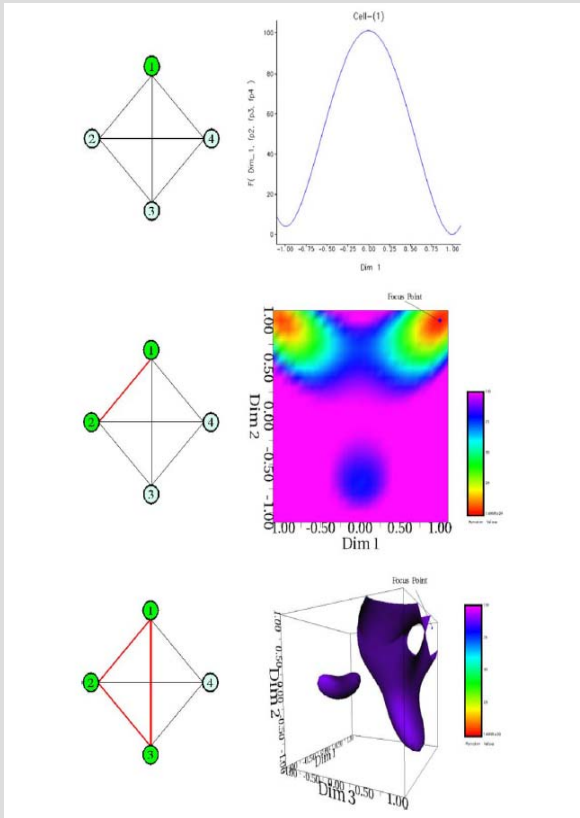
- Extends hyperslice method (2D plots of all variates)



Each attribute has a range of interest and a focus value

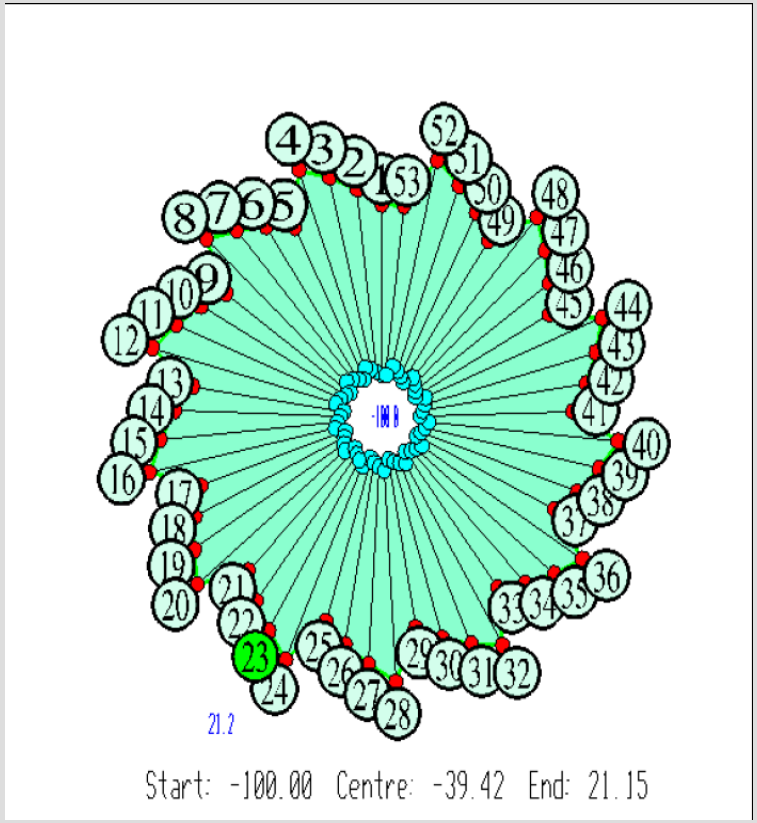


These values can be dynamically changed



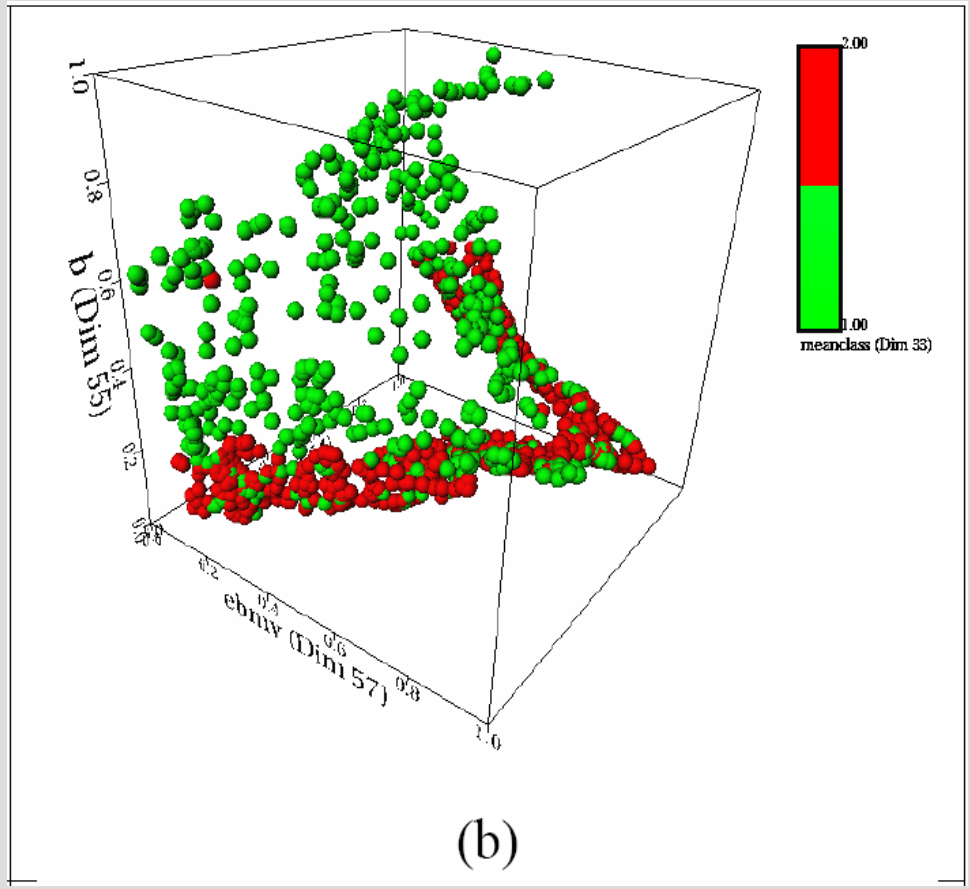
Re-thinking

- Joint study Leeds/Bob Mann
- SuperCOSMOS archive
- ..made us rethink some ideas!
- Only looked at subset of 57 attributes and 1000 observations
- Analytical task:
  - Calibration of SSA data
  - Look for expected and unexpected correlations



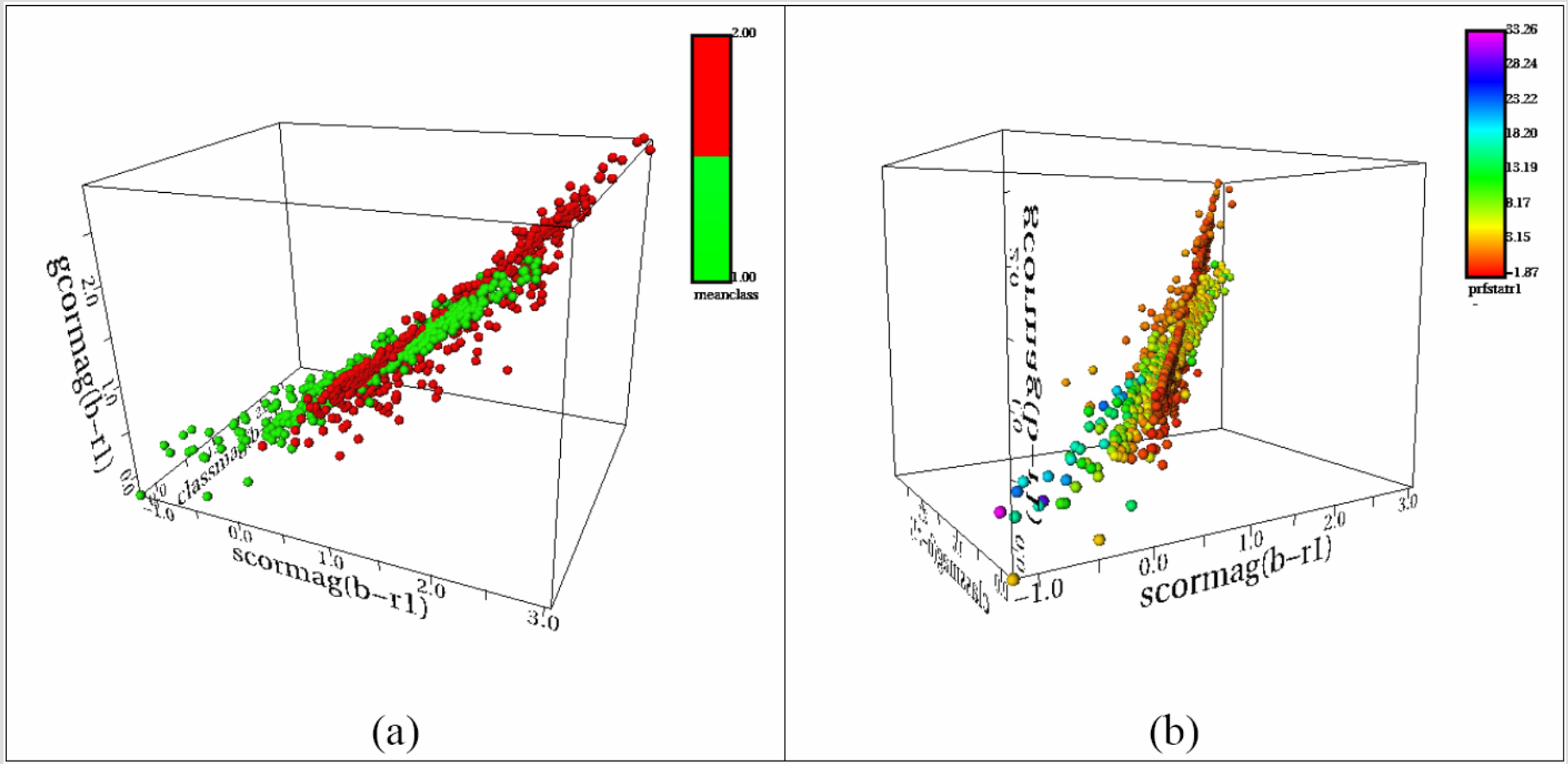
A result..

- Subspace (l, b, ebmv)
- with colouring by
- meanclass attribute
- An outlier is evident



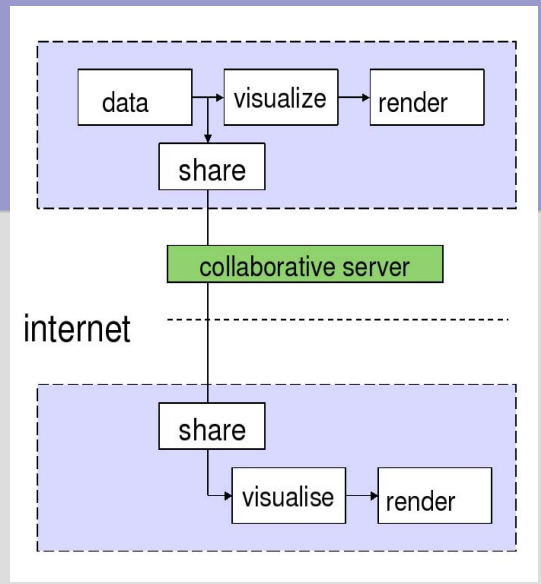
..and a plane in 3D

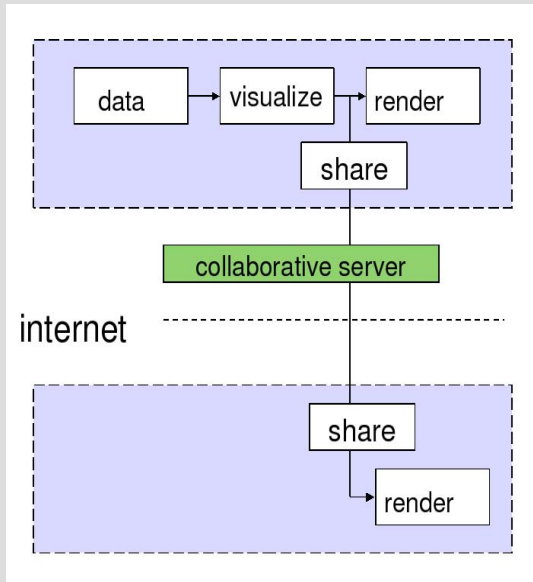
Subspace defined by (*classmag(b-r1), gcormag(b-r1), scormag(b-r1)*)



Also, gViz

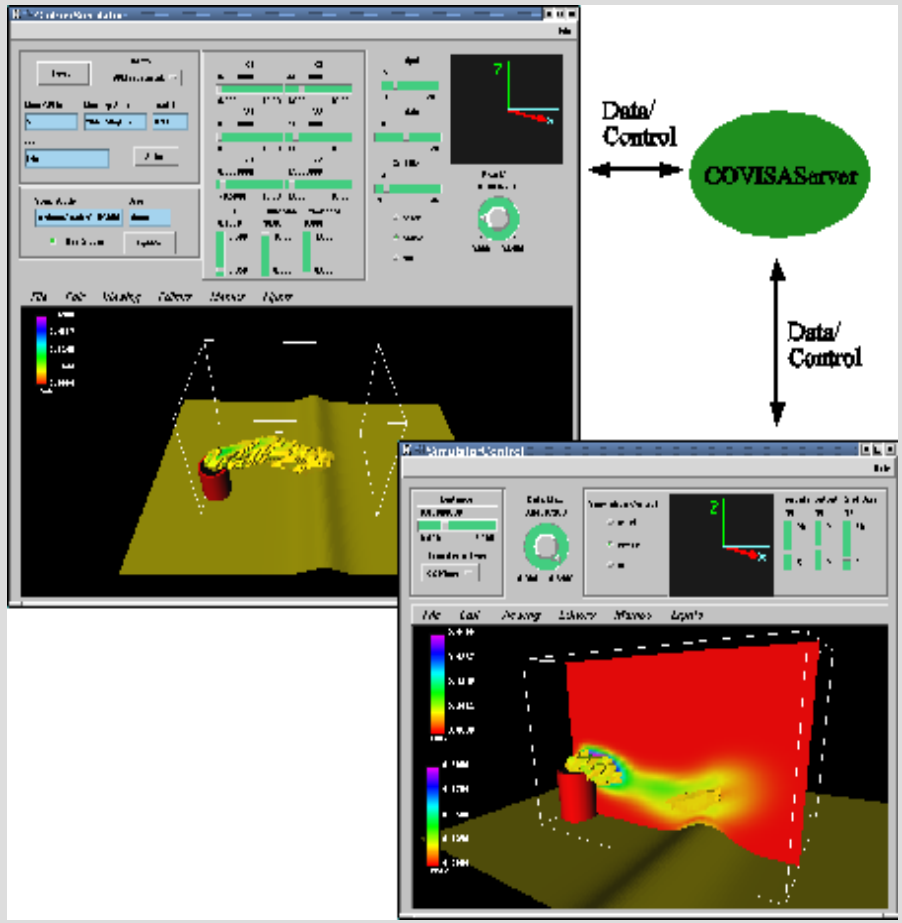
- Developed at Leeds
- Grid-based collaborative visualization
- ..but not tied to grid
- Can push geometry or data
- In principle, client can be any capable renderer (eg. Iris explorer, Matlab, VTK, ...)





gViz

- Allowing steering
- or collaboration



Discuss....

- These and other tools?
- 3D, 4D, hyperslice?
- Black-box techniques?
- Intuitive nature of interface?

Some demos

Conclusions?

- Visualization demands high level of interaction (and good HCI)
  - Interactivity on AG does tied to specific applications
  - could we make use of pipe/SVG model?
- 3D can be challenging, but can't afford to lose
  - focus, context, manipulation
- Using R within AG demonstrated several problems
  - passing files, security, interactivity
- Gap between moving from 100s of variates to 10s
  - can we prioritize data?
- Preliminary survey of data mining tools:
- http://wiki.eurovotech.org/bin/view/VOTech/DS6Preliminary