

Multidimensional Interactive Visualization

Cecilia R. Aragon

1247

UC Berkeley

Spring 2010

Acknowledgments

- Thanks to Marti Hearst, Tamara Munzner for the slides

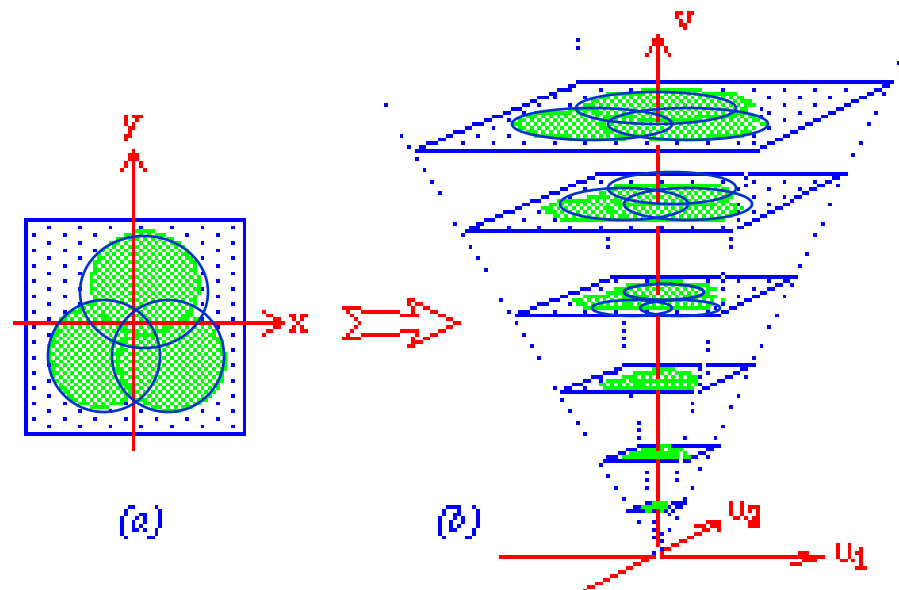
Today

- Finish panning and zooming from last time: space-scale diagrams (Furnas & Bederson, 95) and smooth and efficient panning and zooming (van Wijk et al, 03)
- Then move on to multidimensional interactive visualization

Space-Scale Diagrams

(Furnas & Bederson 95)

- Original figure, shown at various scales
- Horizontal axis is standard, vertical is scale

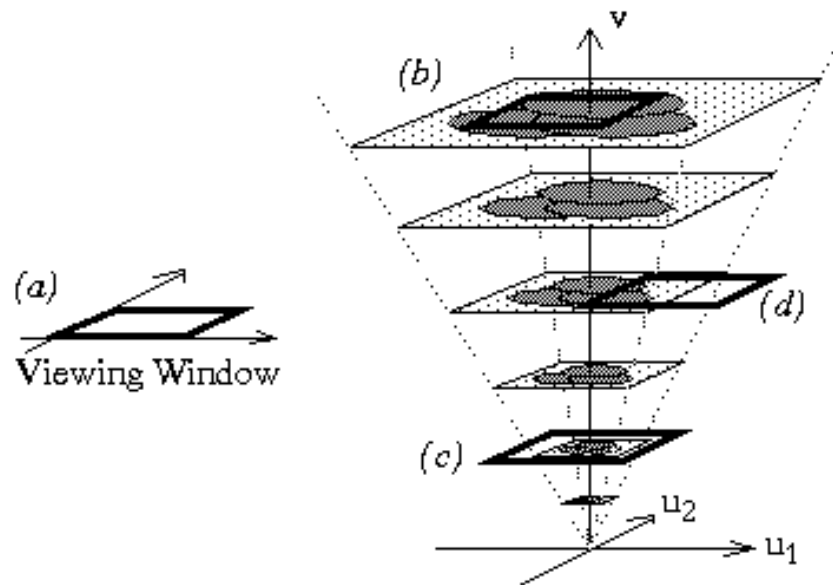


Space-Scale diagram

Space-Scale Diagrams

(Furnas & Bederson 95)

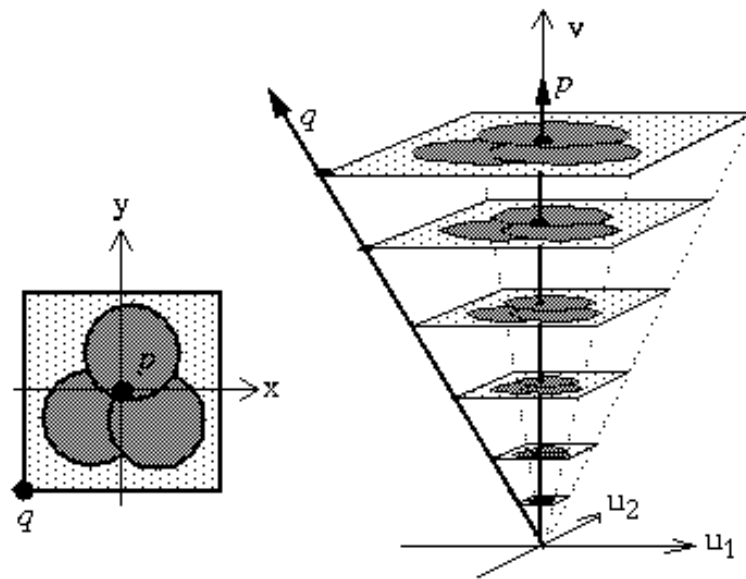
- User has a fixed-sized viewing window
- Moving it through 3D space yields all possible sequences of pan & zoom



Space-Scale Diagrams

(Furnas & Bederson 95)

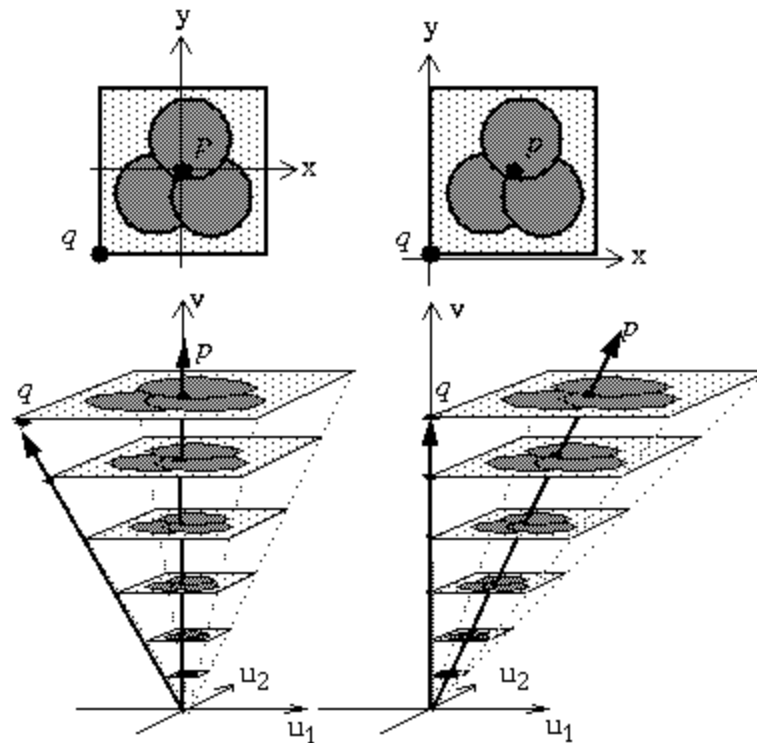
- A point is transformed to a ray
- Circular regions become cones



Space-Scale Diagrams

(Furnas & Bederson 95)

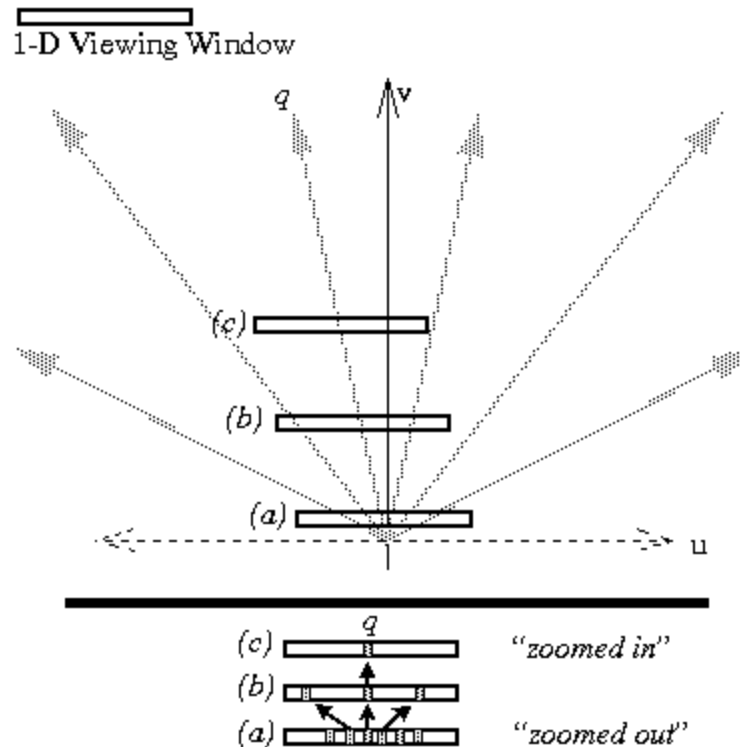
- If you move the origin of the 2D plane, the properties of the original 2D picture do not change
- Therefore, the absolute angles between the rays should not be assigned any meaning



Space-Scale Diagrams

(Furnas & Bederson 95)

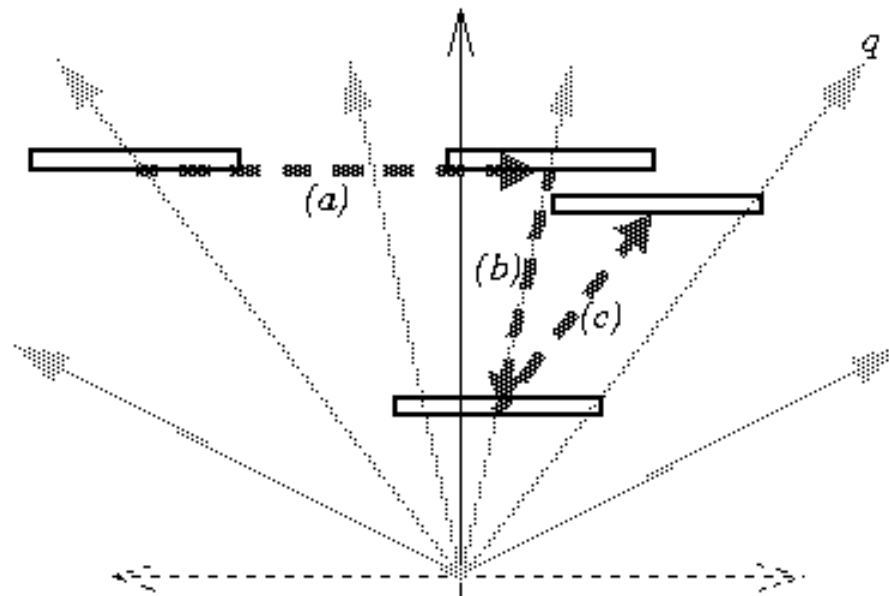
- We can think of this in terms of 1D too
- When zoomed out, you can see wider set of points



Space-Scale Diagrams

(Furnas & Bederson 95)

- Pure pan (a)
- Pure zoom (b)
- Pan and zoom keeping q in same position in the viewing window (c)



Space-Scale Diagrams

(Furnas & Bederson 95)

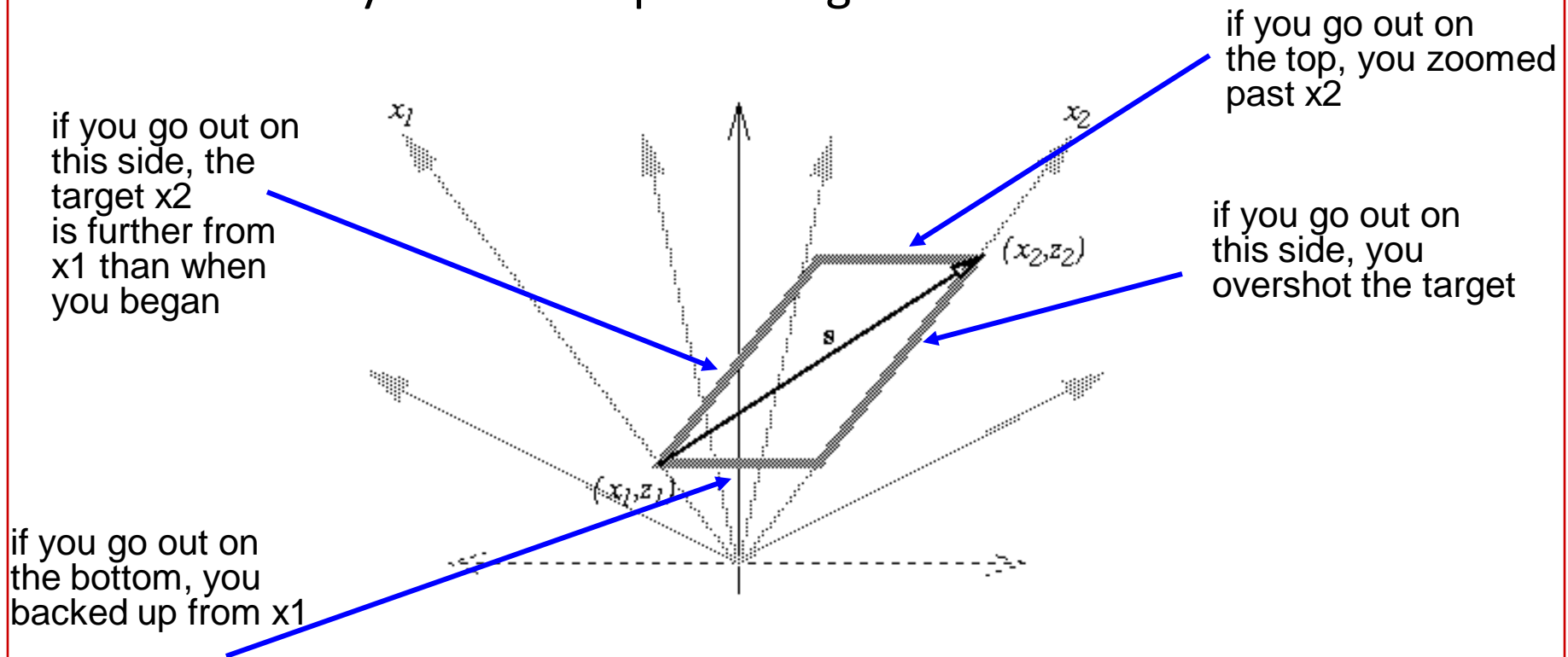
What about panning and zooming at the same time?

- Panning is linear
- Zooming is logarithmic
- The two effects interact
 - If you compute the two separately and run them in parallel you get problems
 - When zooming in, things go exponentially fast
 - Panning can't keep up
 - The target “runs away” out of view

Space-Scale Diagrams

(Furnas & Bederson 95)

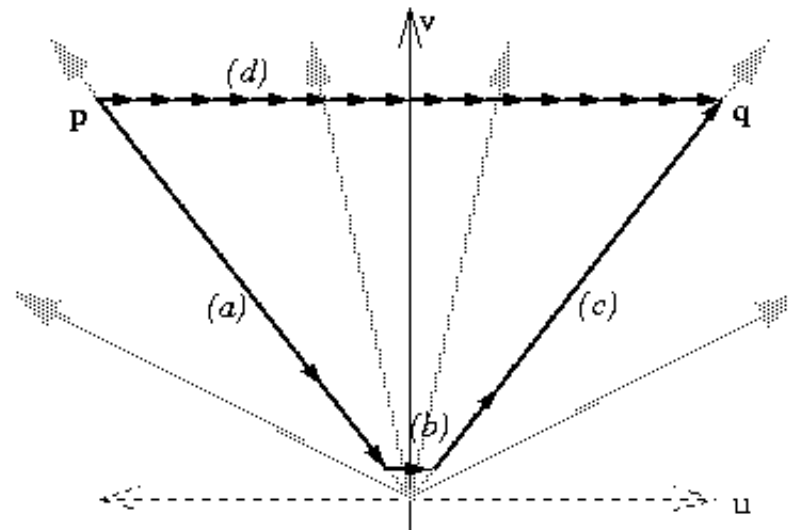
- Solution: space-scale diagram
- Goal: go from (x_1, z_1) to (x_2, z_2)
- Must stay within the parallelogram



Space-Scale Diagrams

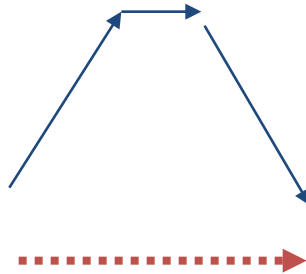
(Furnas & Bederson 95)

- Zooming covers more ground faster than panning
 - zooming is logarithmic, panning is linear
- Alternative way to navigate:
 - Instead of a long pan
 - Do a big zoom, a short pan, a big zoom
 - (count the number of arrows each way)



Navigation in Pad++

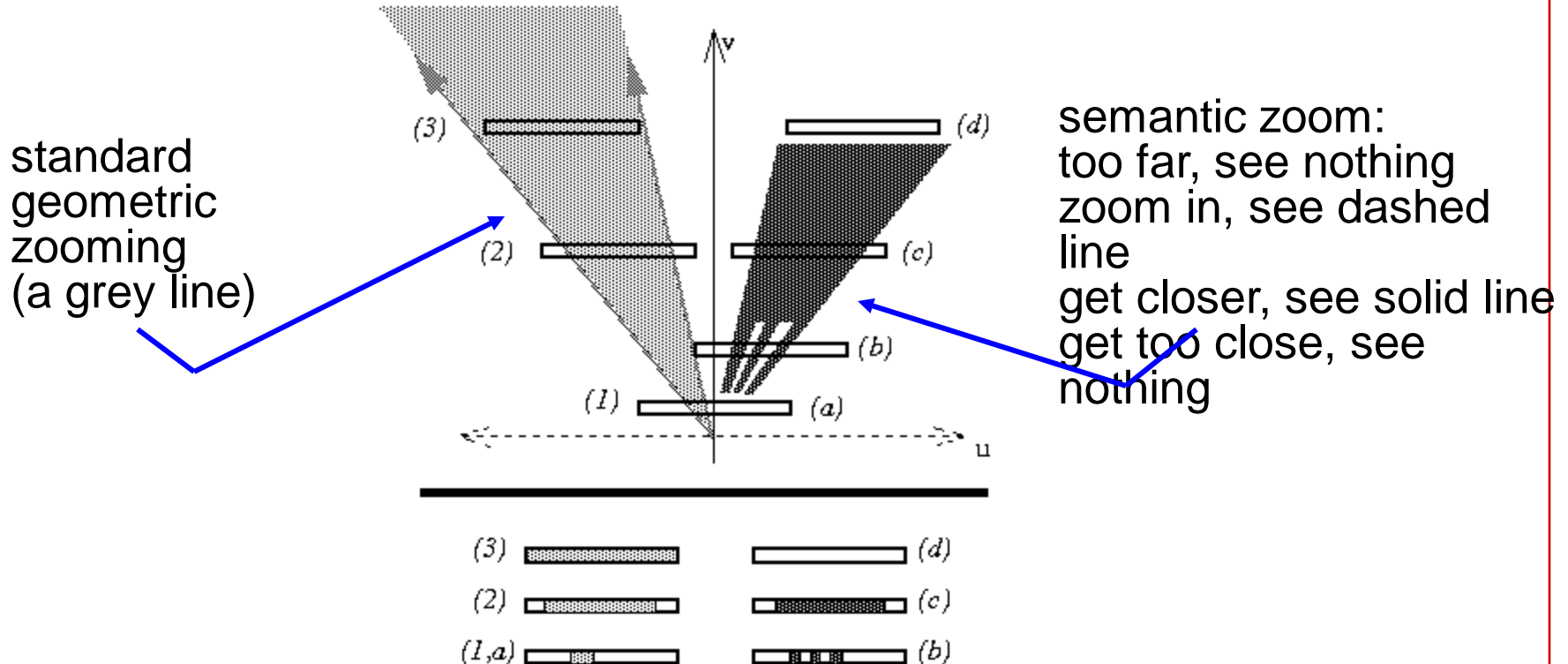
- How to keep from getting lost?
 - Animate the traversal from one object to another using “hyperlinks”
 - If the target is more than one screen away, zoom out, pan over, and zoom back in
 - Goal: help viewer maintain context



Space-Scale Diagrams

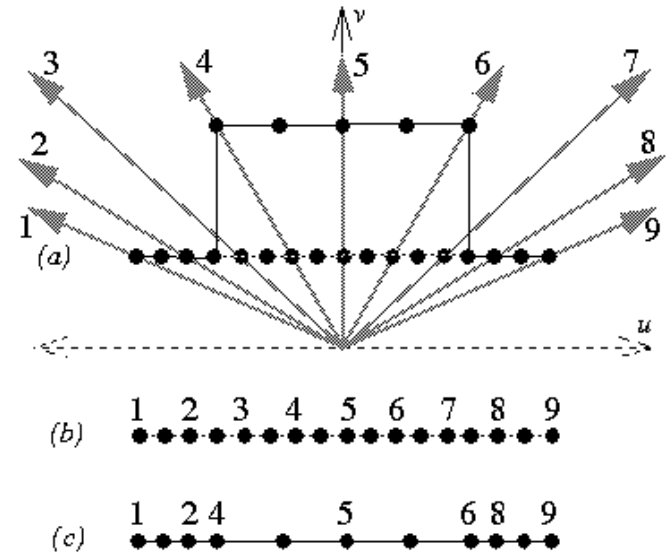
(Furnas & Bederson 95)

- Implementing semantic zooming



Space-Scale Diagrams (Furnas & Bederson 95)

- How space-scale diagrams represent overview + details views (2 levels of magnification with a sharp transition between)
 - Instead of a horizontal slice through scale-space, take a step up and a step down
 - The points in the middle have more room; those on the periphery are squished together



Smooth and Efficient Zooming and Panning (vanWijk, Nuij 2003)

Smooth and Efficient Zooming and Panning. Jack J. van Wijk and Wim A.A. Nuij, Proc. InfoVis 2003, p. 15-22.

- Multi-scale viewing
- What is the optimal path for combining zooming and panning?
- Define it as “smooth and efficient”
- Define a metric to measure the combination of zooming and panning based on differential geometry

Smooth and Efficient Zooming and Panning (vanWijk, Nuij 2003)

Smooth and Efficient Zooming and Panning. Jack J. van Wijk and Wim A.A. Nuij, Proc. InfoVis 2003, p. 15-22.

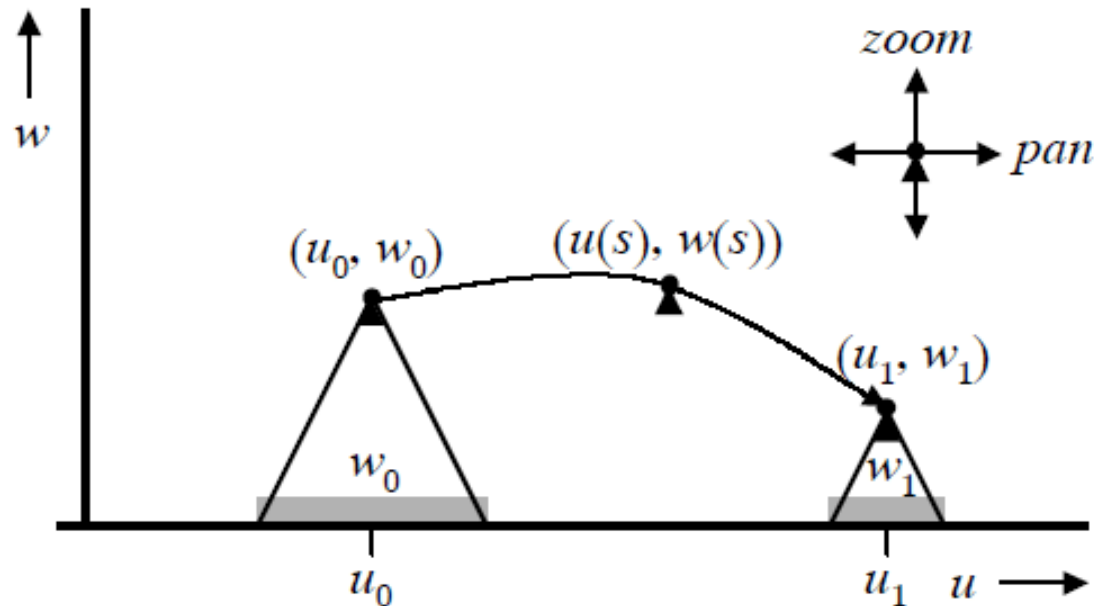


Figure 2: u, w space diagram

Smooth and Efficient Zooming and Panning (vanWijk, Nuij 2003)

Metric: perceived average optic flow in image window
window

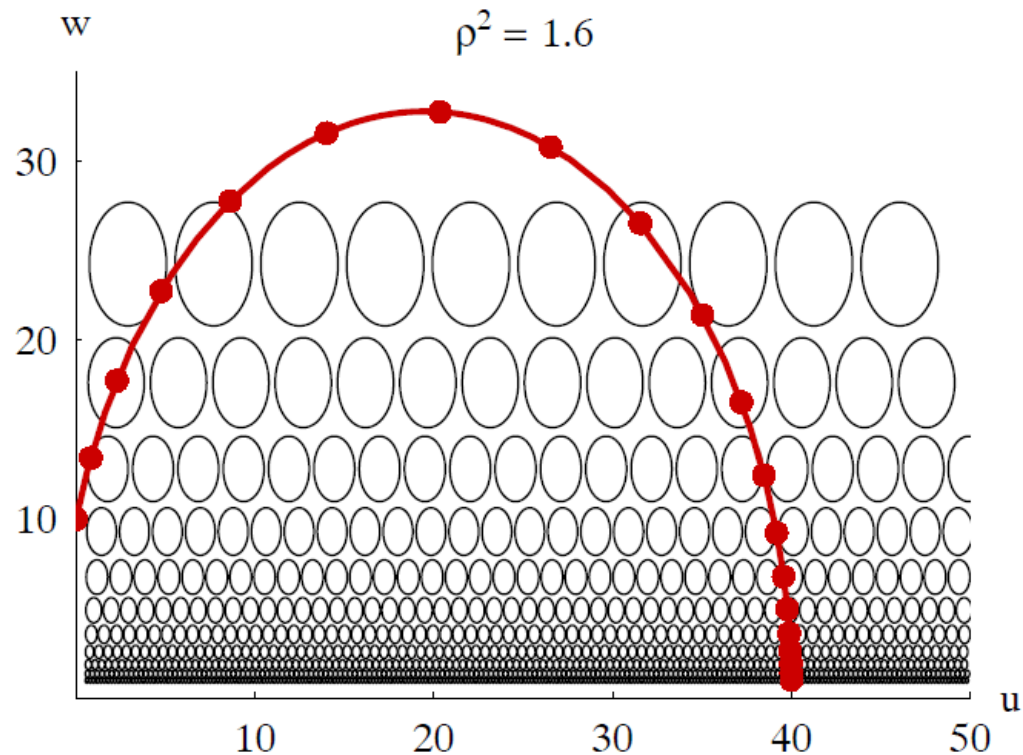


Figure 3: Metric in (u, w) space

Panning and Zooming

- Is it actually useful?
 - Is it better to show multiple simultaneous views?
 - Is it better to use distortion techniques?
- Would keeping a separate global overview help with navigation?
 - The research literature suggests yes, that overview+detail is usually better than pan & zoom.

Multidimensional Interactive Visualization

- Most data sets have multiple dimensions; how do we compare them?
- Should we automate dimensionality analysis/reduction?
- Ask what-if questions spontaneously while working through a problem
- Control the exploration of subsets of data from different viewpoints

Interactive Advantages

- *“Automatic methods do not leverage human expertise and can only find solutions that are optimal with regard to an invariably oversimplified problem description. Furthermore, users must understand the generated solutions in order to implement, justify or modify them.”*

-- Scott et al (2002)

SCOTT, S.D., LESH, N. and KLAU, G.W. (2002) Investigating human-computer optimization, ACM, Proc. CHI'02, pp.155-162.

Problem Statement

- How to effectively present more than 3 dimensions of information in a visual display with 2 (to 3) dimensions?
- How to effectively visualize “inherently abstract” data?
- How to effectively visualize very large, often complex data sets?
- How to effectively display results – when you don’t know what those results will be?

Another Statement of Goals

Visualization of multidimensional data

- Without loss of information
- With:
 - Minimal complexity
 - Any number of dimensions
 - Variables treated uniformly
 - Objects remain recognizable across transformations
 - Easy / intuitive conveyance of information
 - Mathematically / algorithmically rigorous

(Adapted from Inselberg)

Characteristics

- “Data-dense displays” (large number of dimensions and/or values)
 - Often combine color with position / proximity representing relevance “distance”
 - Often provide multiple views
- Build on concepts from previous weeks:
 - Retinal properties of marks
 - Gestalt concepts, e.g., grouping
 - Direct manipulation / interactive queries
 - Incremental construction of queries
 - Dynamic feedback
- Some require specialized input devices or unique gesture vocabulary

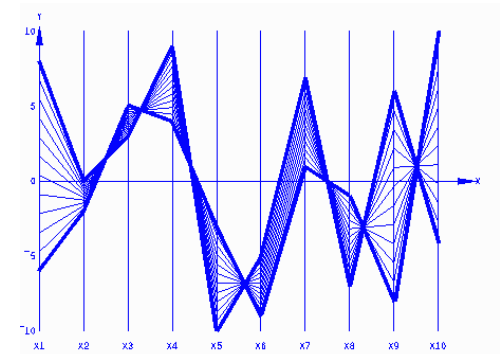
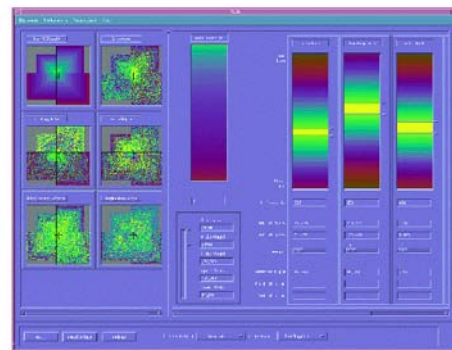
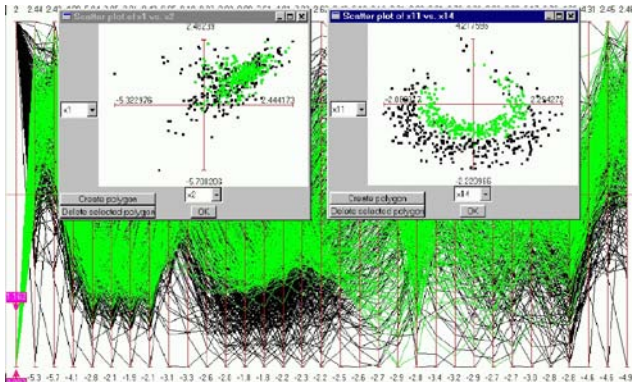
Approaches to Multidimensional Visualization

- Parallel coordinates (Inselberg 97)
- Dynamic queries, brushing and linking (Influence Explorer/Prosection Matrix, Tweedie 95)
- Dimensionality reduction (Dimension ordering, Peng 03)
- Other approaches (Semantic substrates, PaperLens, VisDB)

Examples

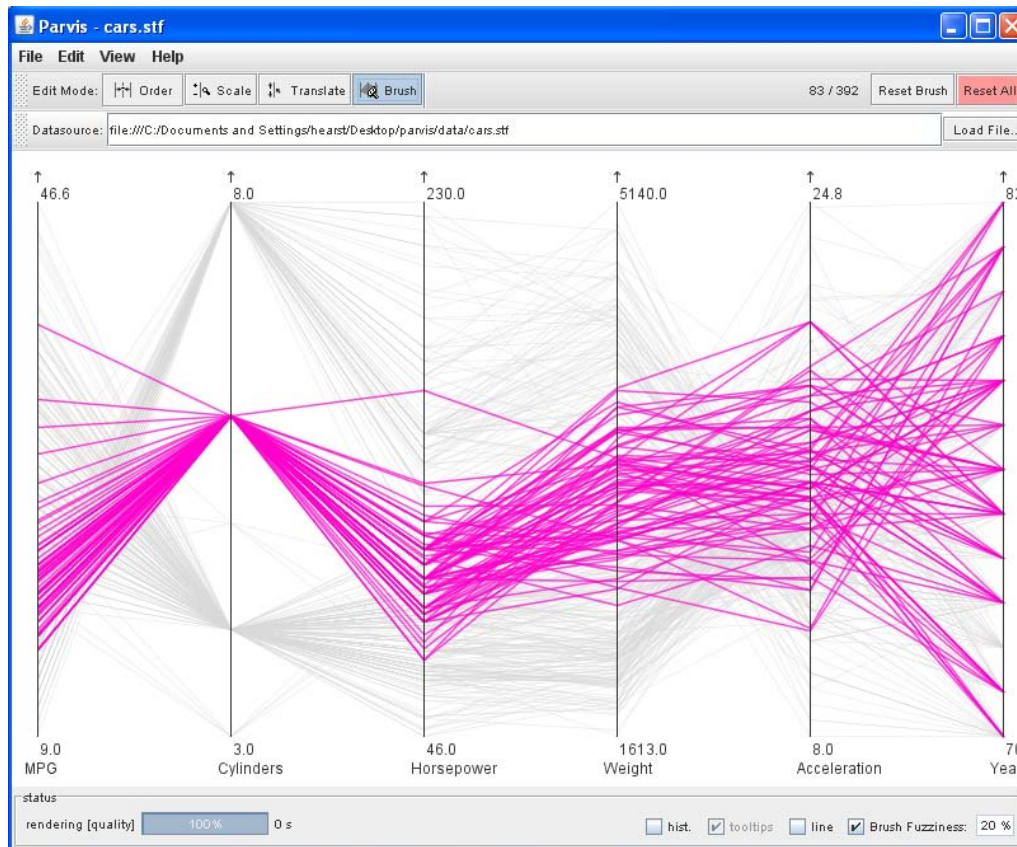
Warning: These visualizations are not easy to grasp at “first glance”!

DON'T PANIC



Highlighting and Brushing: Parallel Coordinates by Inselberg

- Visual Data Detective



Multidimensional Detective

A. Inselberg, Multidimensional Detective,
Proceedings of IEEE Symposium on Information
Visualization (InfoVis '97), 1997.

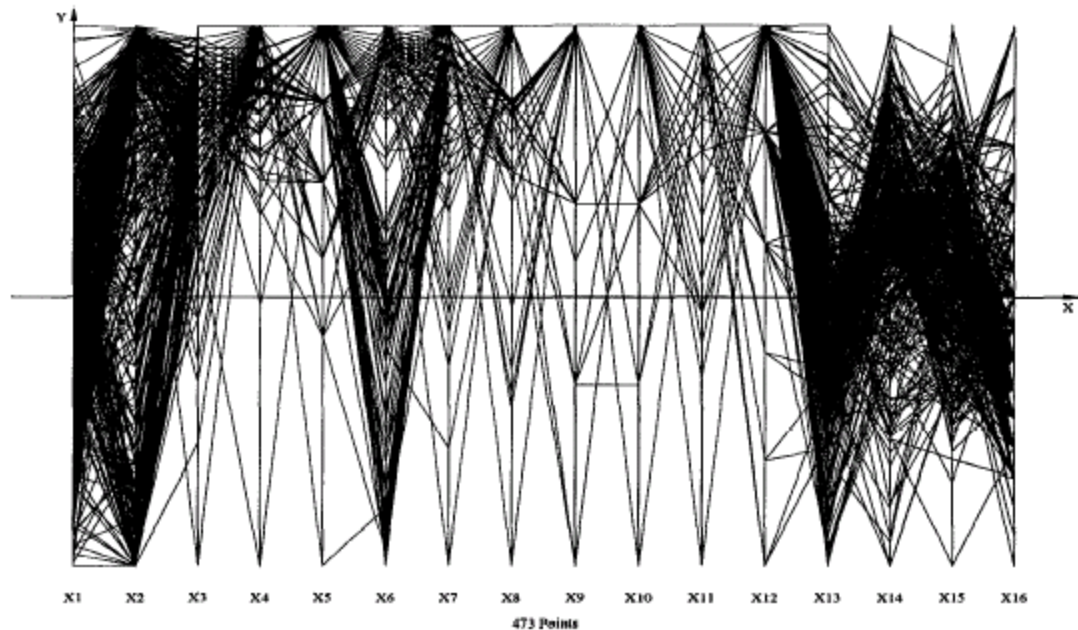


Figure 1: The full dataset consisting of 473 batches

Inselberg's Principles

A. Inselberg, Multidimensional Detective, Proceedings of IEEE Symposium on Information Visualization (InfoVis '97), 1997

1. Do not let the picture scare you
2. Understand your objectives
 - Use them to obtain visual cues
3. Carefully scrutinize the picture
4. Test your assumptions, especially the “I am really sure of’s”
5. You can't be unlucky all the time!

A Detective Story

A. Inselberg, Multidimensional Detective, Proceedings of IEEE Symposium on Information Visualization (InfoVis '97), 1997

- The Dataset:
 - Production data for 473 batches of a VLSI chip
 - 16 process parameters:
 - **X1:** The yield: % of produced chips that are useful
 - **X2:** The quality of the produced chips (speed)
 - **X3 ... X12:** 10 types of defects (zero defects shown at top)
 - **X13 ... X16:** 4 physical parameters
- The Objective:
 - Raise the yield (X1) and maintain high quality (X2)

Multidimensional Detective

- Each line represents the values for one batch of chips
- This figure shows what happens when only those batches with both high X_1 and high X_2 are chosen
- Notice the separation in values at X_{15}
- Also, some batches with few X_3 defects are not in this high-yield/high-quality group.

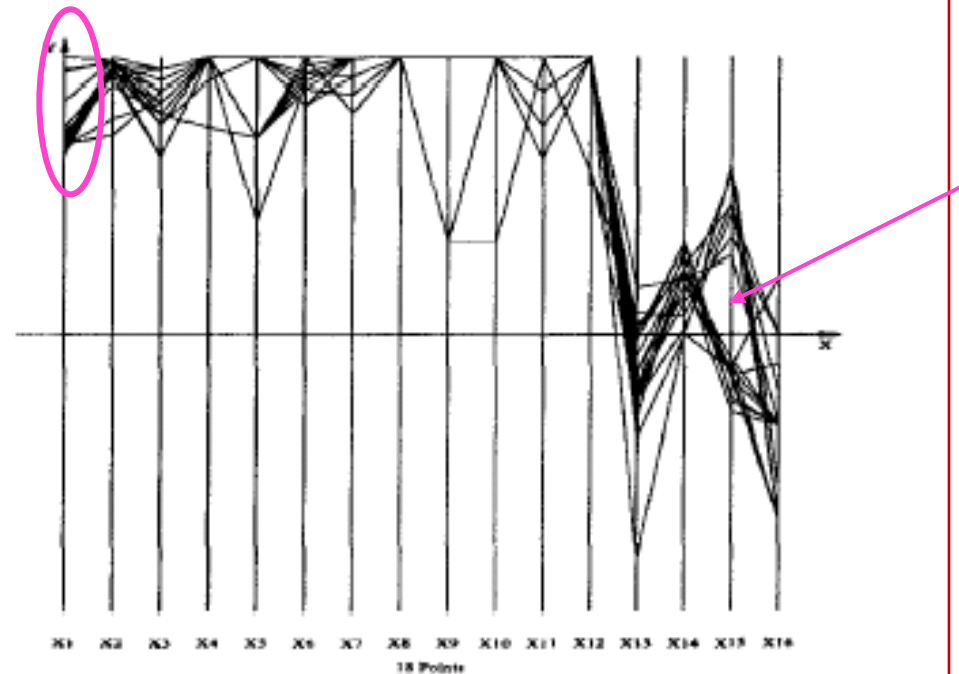


Figure 2: The batches high in Yield, X_1 , and Quality, X_2 .

Multidimensional Detective

- Now look for batches which have *nearly* zero defects.
 - For 9 out of 10 defect categories
- Most of these have low yields
- This is surprising because we know from the first diagram that some defects are ok.

Go back to first diagram, looking at defect categories.

Notice that X6 behaves differently than the rest.

Allow two defects, where one defect is X6.

This results in the very best batch appearing.

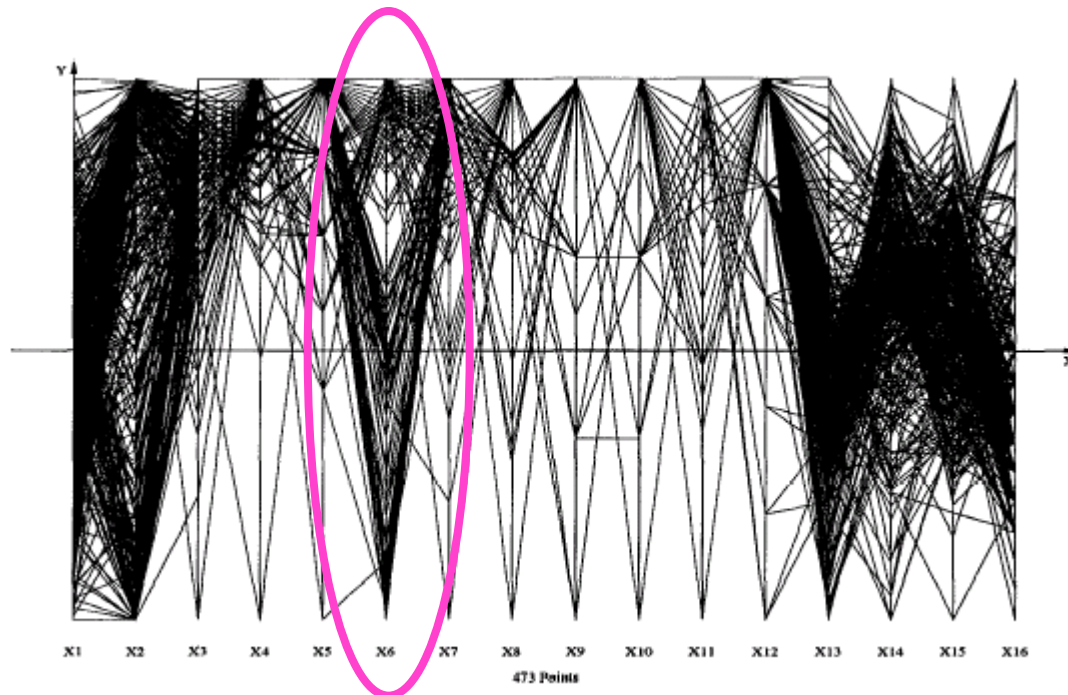


Figure 1: The full dataset consisting of 473 batches

Multidimensional Detective

- Fig 5 and 6 show that high yield batches don't have non-zero values for defects of type X3 and X6 (Don't believe your assumptions ...)
- Looking now at X15 we see the separation is important
 - Lower values of this property end up in the better yield batches

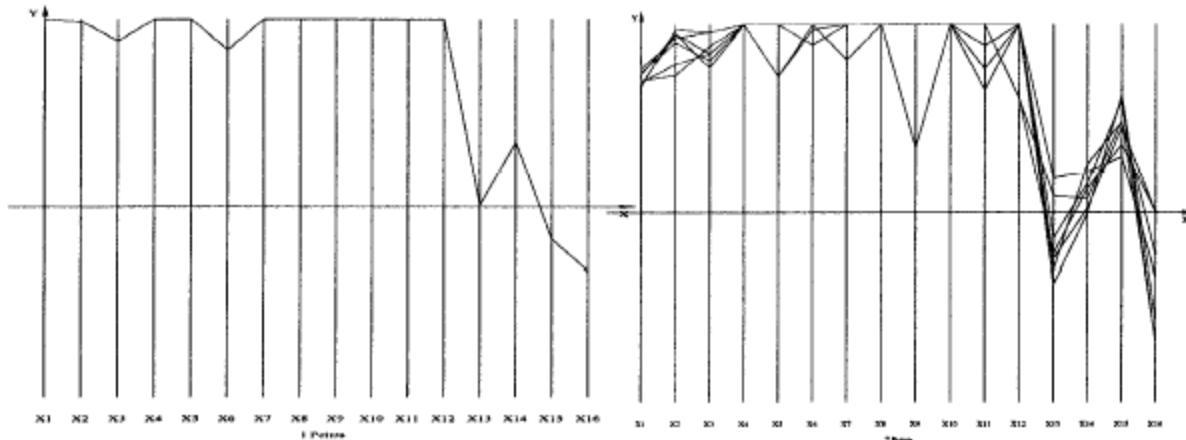
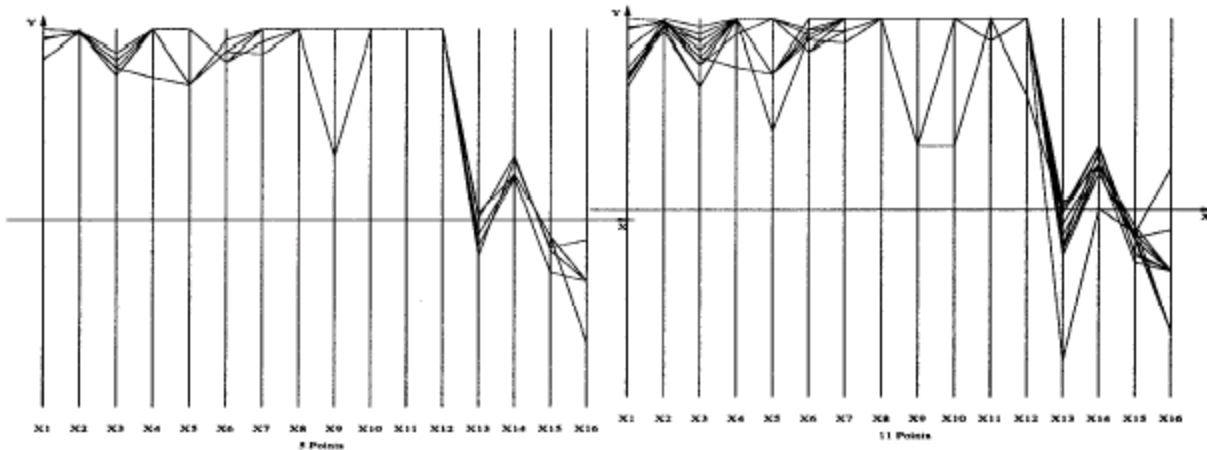


Figure 5: The best batch. Highest in Yield, X1, and very high in Quality, X2.

Figure 7: Upper range of split in X15



Influence Explorer / Prosection Matrix

(Tweedie et. al. 95)

- <http://www.open-video.org/details.php?videoid=5015>
- Abstract one-way mathematical models: multiple parameters, multiple variables.
- Data for visualization comes from sampling
- Visualization of non-obvious underlying structures in models
- Color coding, attention to near misses

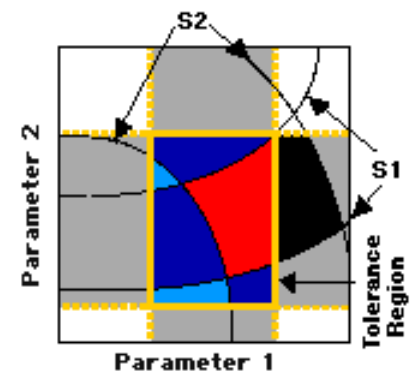
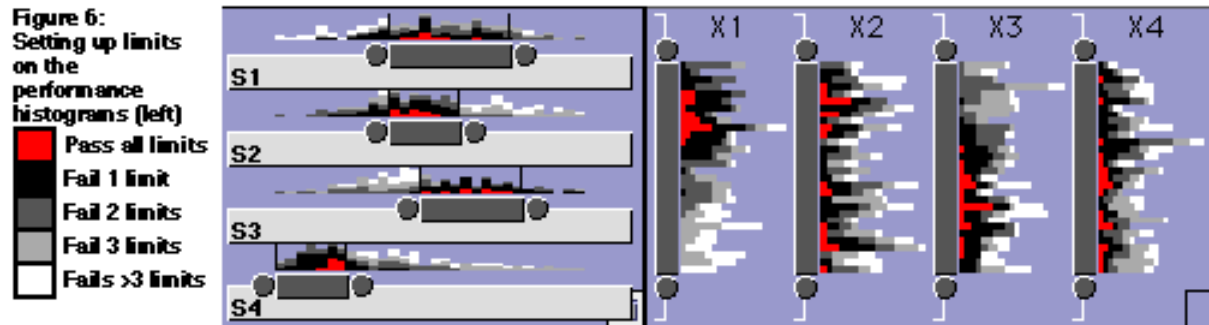


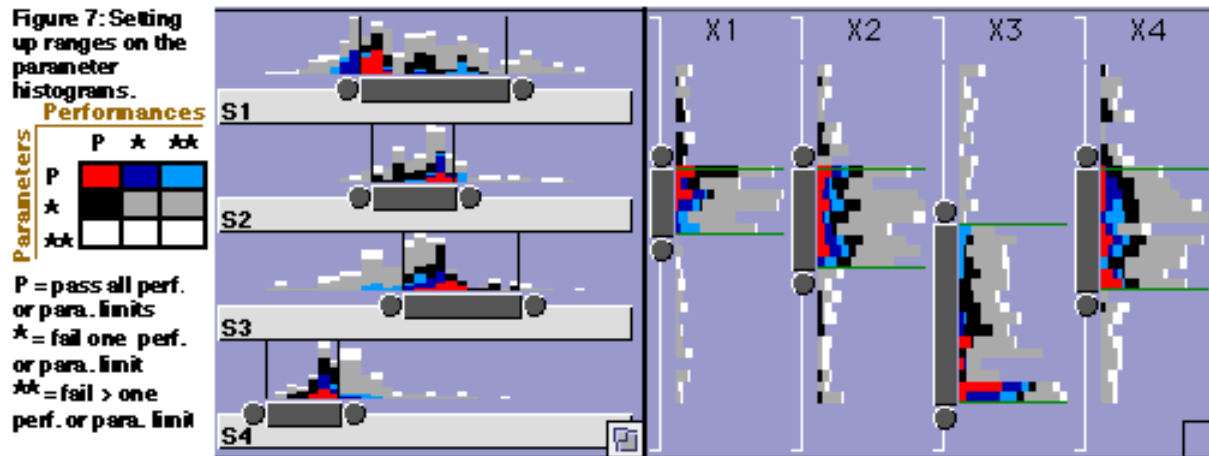
Figure 8: The colour coding for the tolerance setting (Figure 7) shown in 2 parameter space.

Influence Explorer / Prosection Matrix (Tweedie et. al.)



- Use the sliders to set performance limits.
- Color coding gives immediate feedback as to effects of changes—both for ‘perfect’ scores and for near-misses.
- Can also highlight individual values across histograms, show parallel coordinates.
- Interactive querying!

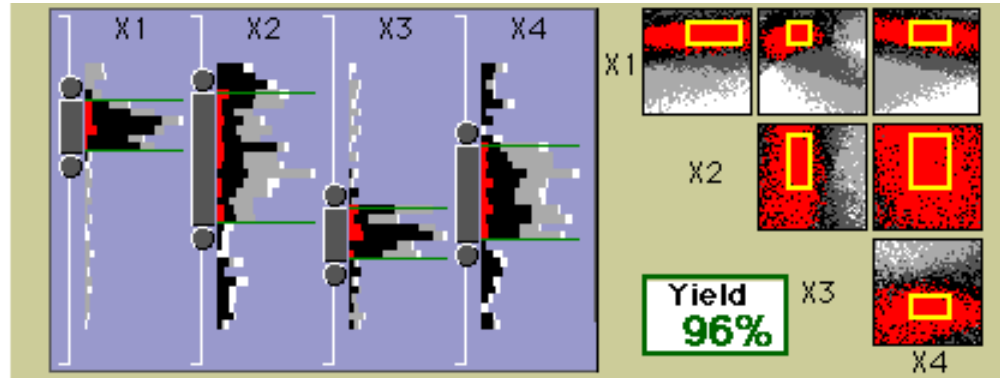
Influence Explorer / Prosection Matrix (Tweedie et. al.)



- In this view we can shift parameter ranges in addition to performance limits.
- Red is still a perfect score—blacks miss one parameter limit, blues one or two performance limits.
- Does this color scheme make sense? Would another work better?

Influence Explorer / Prosection Matrix (Tweedie et. al.)

Figure 15:
Maximising the widths of the parameter ranges to get high yield and minimum cost. (the same specification is used as in Figure 6). Note how the rectangular tolerance region just fits inside the red region of acceptability.



- Prosection matrix (on right) = scatter plots for pairs of parameters.
- Color coding matches histograms.
- Fitting tolerance region (yellow box) to acceptability (red region) gives high yield for minimum cost
- Or: Make the red bit as big as possible!
- This aspect closely tuned to task at hand: manufacturing and similar.

Dimensionality Reduction

- Mapping multidimensional space into space of fewer dimensions
 - Filter subset of original dimensions
 - Generate new synthetic dimensions
- Why is lower-dimensional approximation useful?
 - Assume intrinsic dimensionality of dataset is much lower than measured dimensionality
- Why?
 - Sparse data in verbose space
 - Ex.: documents: word occurrence vectors. 10K+ dimensions, want dozens of topic clusters

Adapted from Munzner

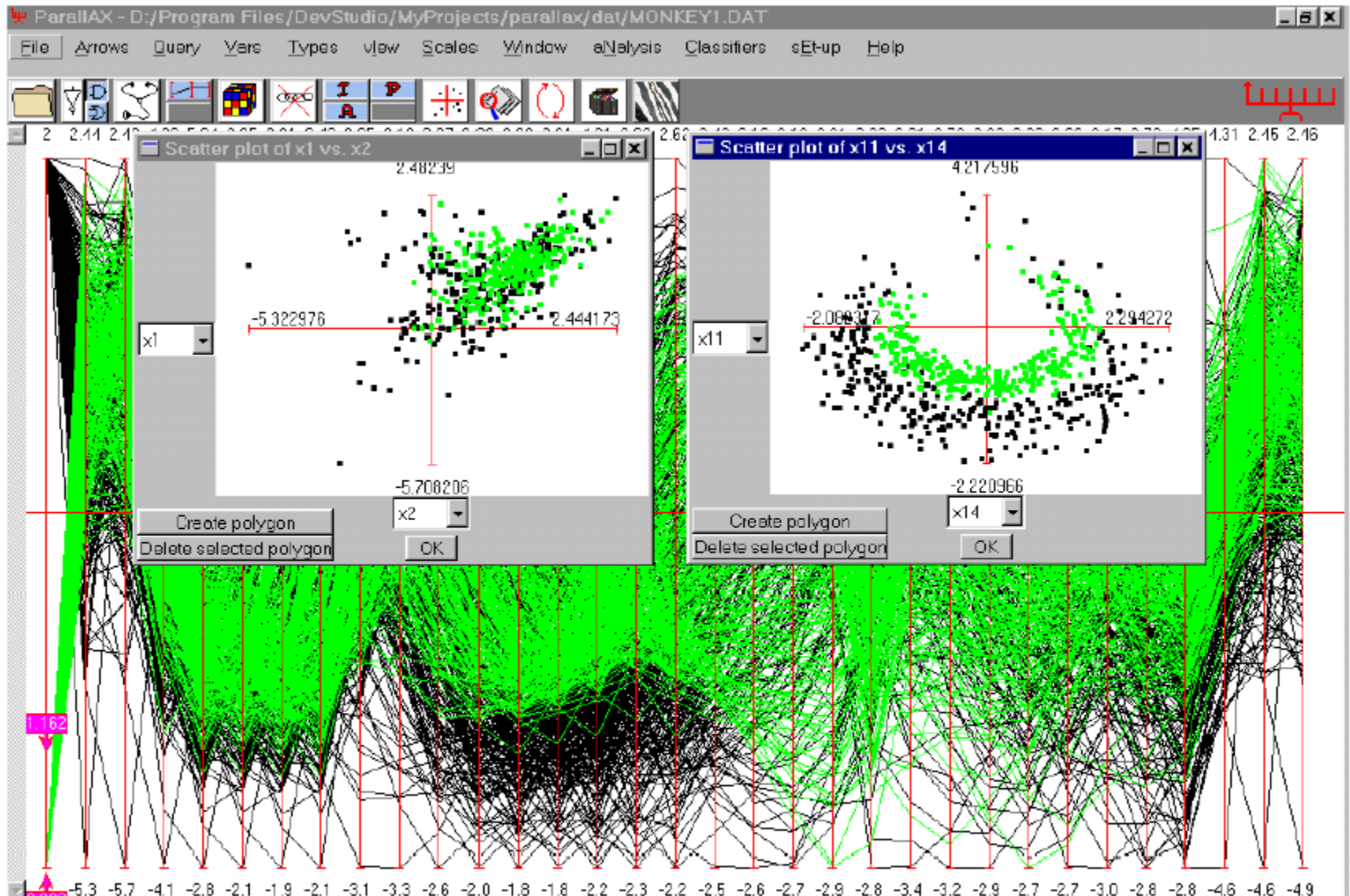
Dimensionality Reduction

- Large number of methods from statistical learning theory:
 - Principal component analysis
 - Multi-dimensional scaling
 - Minimize differences between interpoint distances in high and low dimensions
 - Distance scaling: minimize objective function

Adapted from Munzner

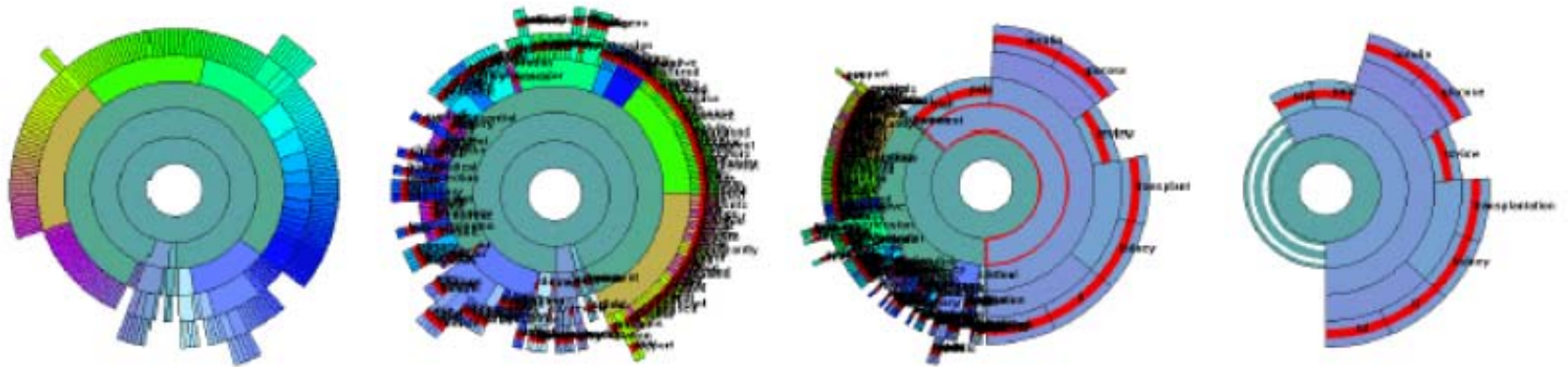
Automated Analysis

A. Inselberg, Automated Knowledge Discovery using Parallel Coordinates, INFOVIS '99



Dimension Ordering

- Raw, ordered, distorted, rollup (filter)

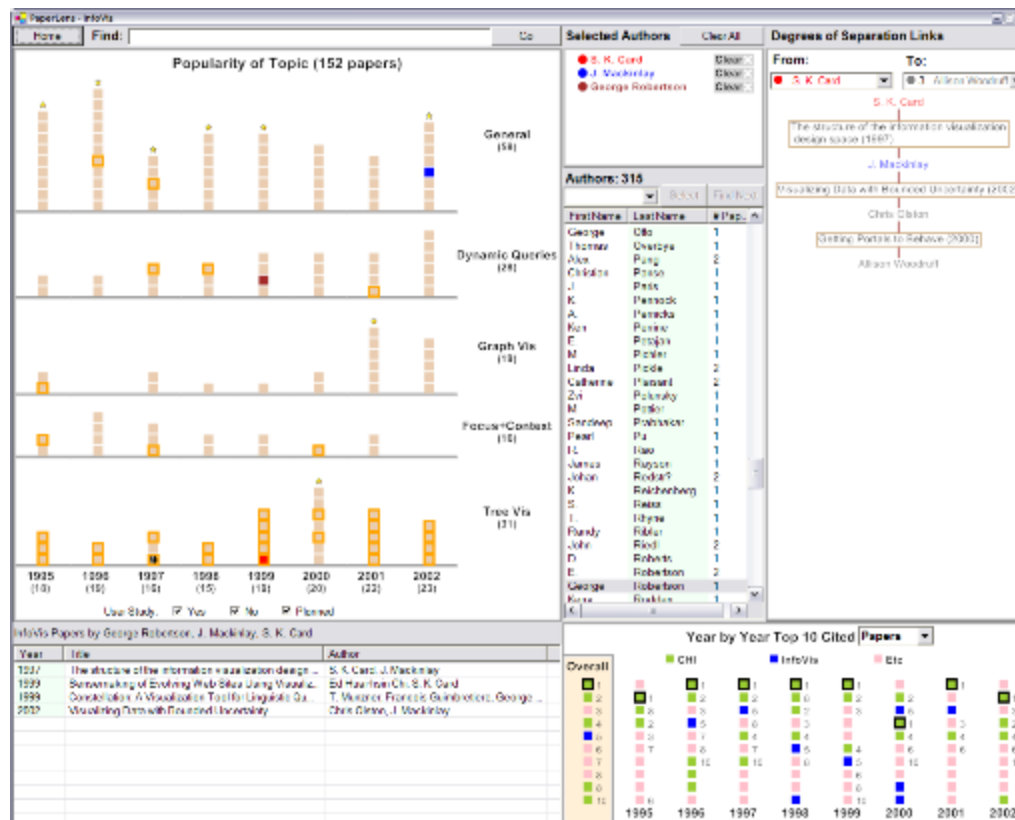


[Interactive Hierarchical Dimension Ordering, Spacing and Filtering for Exploration Of High Dimensional Datasets. Yang Peng, Ward, and Rundensteiner. Proc. InfoVis 2003]

Adapted from Munzner

PaperLens

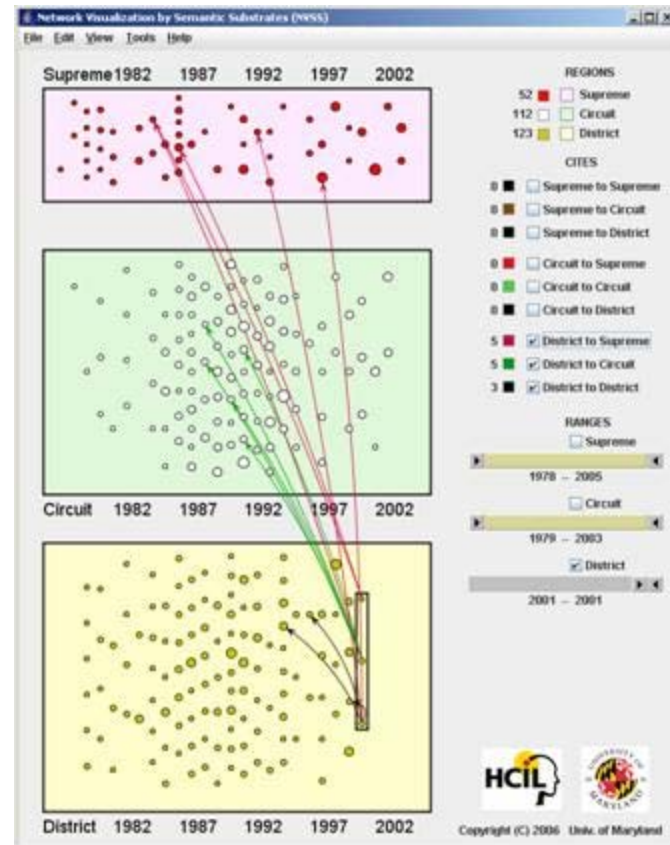
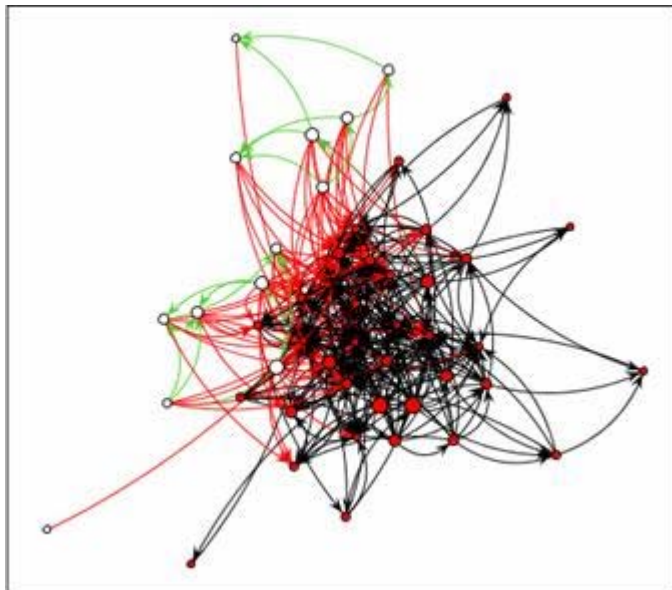
- Understanding research trends in conferences using PaperLens Lee et al., CHI'05 extended abstracts
- <http://www.cs.umd.edu/hcil/paperlens/PaperLens-Video.mov>



Alternative Network Viz

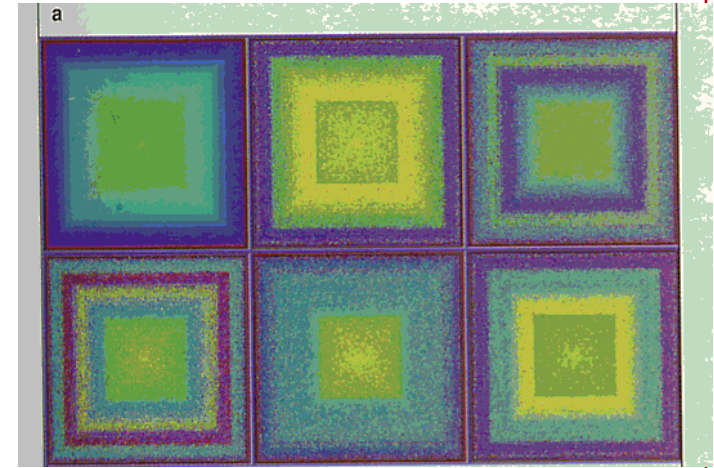
(Legal cases)

- Network Visualization by Semantic Substrates, Shneiderman & Aris, IEEE TVCG 2006.
- <http://hcil.cs.umd.edu/video/2006/substrates.mpg>

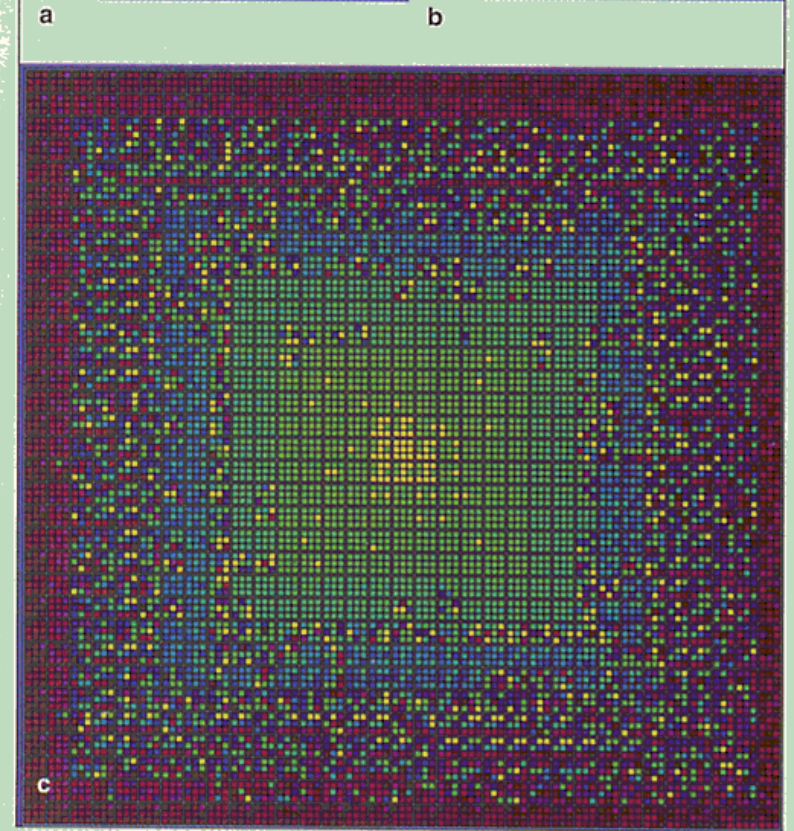
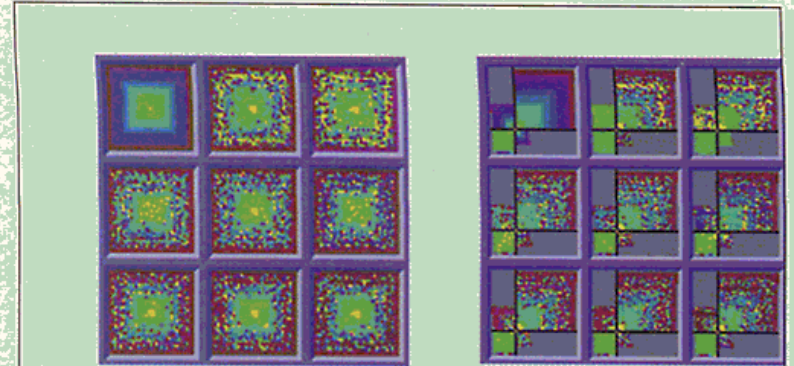
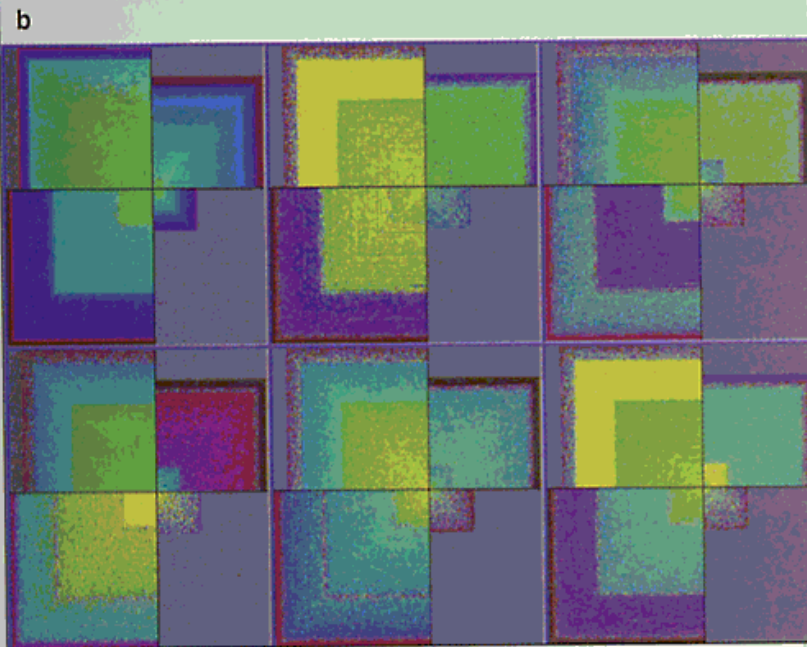
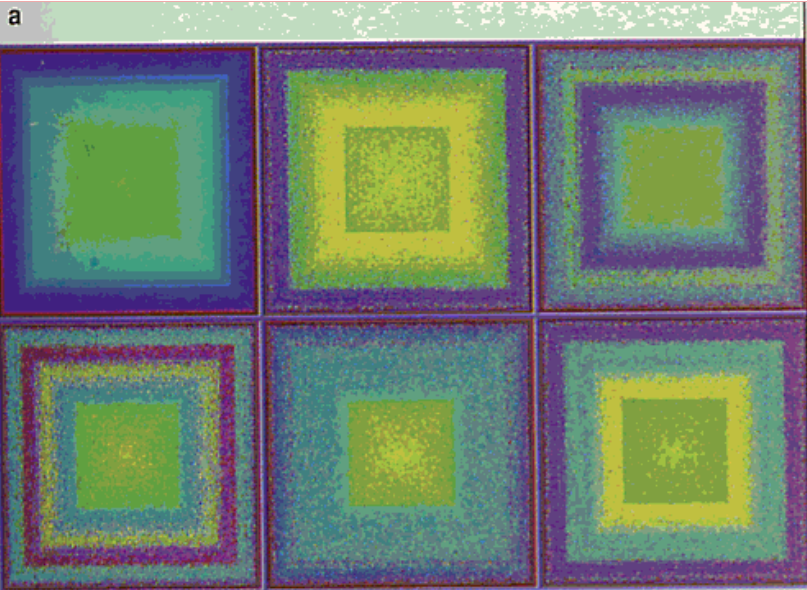


VisDB

(Keim & Kriegel 95)

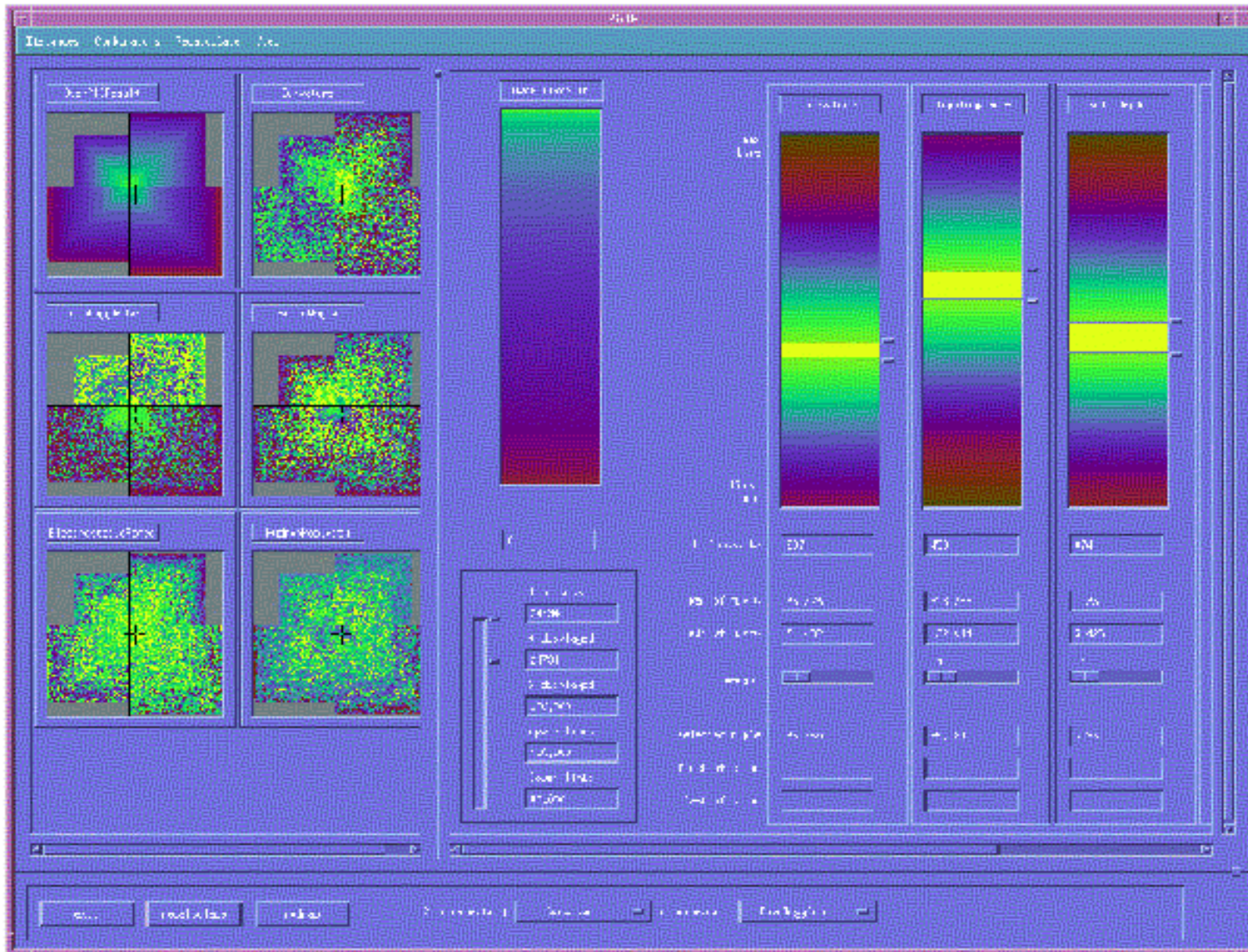


- Mapping entries from relational database to pixels on the screen
- Include “approximate” answers, with placement and color-coding based on relevance
- Data points laid out in:
 - Rectangular spiral
 - Or, with axes representing positive/negative values for two selected dimensions
 - Or, group dimensions together (easier to interpret than very large number of dimensions)



VisDB

- Relevance calculation based on “distance” of each variable from query specification
- Distance calculation depends on data type
 - Numeric: mathematical
 - String: character/substring matching, lexical, phonetic?, syntactic?
 - Nominal: predefined distance matrix
 - Possibly other “domain-specific” distance metrics



Limitations and Issues

- Complexity
- Abstract data
 - These visualizations are oriented toward abstract data
 - For “naturally” two or three-dimensional data (things that vary over time or space, e.g., geographic data) visualizations which exploit those properties may exist and be more effective

Generalized Selection via Interactive Query Relaxation (Heer, Agarwala, Willett 2008)

- Couple declarative selection with *query relaxation*
 - Enable users to interactively generalize their selections
 - http://vis.berkeley.edu/papers/generalized_selection/

