Automated Methods
in Information Visualization

Helwig Hauser (Univ. of Bergen)

Visualization? Information Visualization?

- Vis … Visualization
- InfoVis (or IV) … Information Visualization
  (SoftVis … Software Visualization, …)
- SciVis … Scientific Visualization
  (VolVis … Volume Visualization, FlowVis … Flow Visualization, …)
- VA (or VAST [or IVA]) … Visual Analytics
- GeoVis … Geographic Visualization
- Statistical Graphics, Statistical Visualization
- HCI … Human-Computer Interaction
- Graph Drawing
- Information Design, Information Graphics
- … (Architectual Visualization, …)

Integrated Methods

- Clustering
  - k-means
  - hierarchical clustering methods
  - etc.
- Projections (embeddings), e.g., for dimension reduction
  - PCA
  - MDS
  - etc.
- Classification, regression
  - decision trees
  - SVM
  - etc.
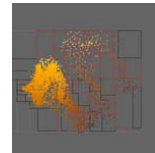- Etc.

Levels of Integration

- In most of Information Visualization:
    - no integration

- Then you find:
    - the visualization of ML results

- You also find:
    - approaches to make ML interactive

- Very rarly you find:
    - tight integration of ML and InfoVis to solve problems
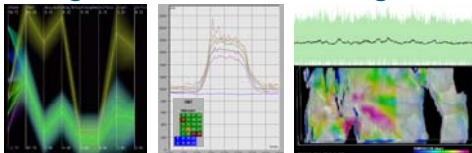
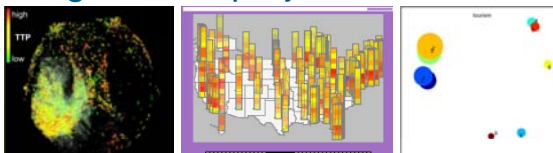[Maniyar & Nabney; MDM 2006]
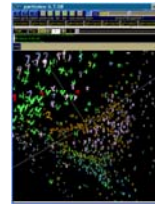
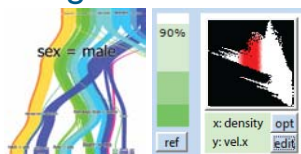[Williams & Munzner; InfoVis 2004]

Some Examples

- Integration of clustering
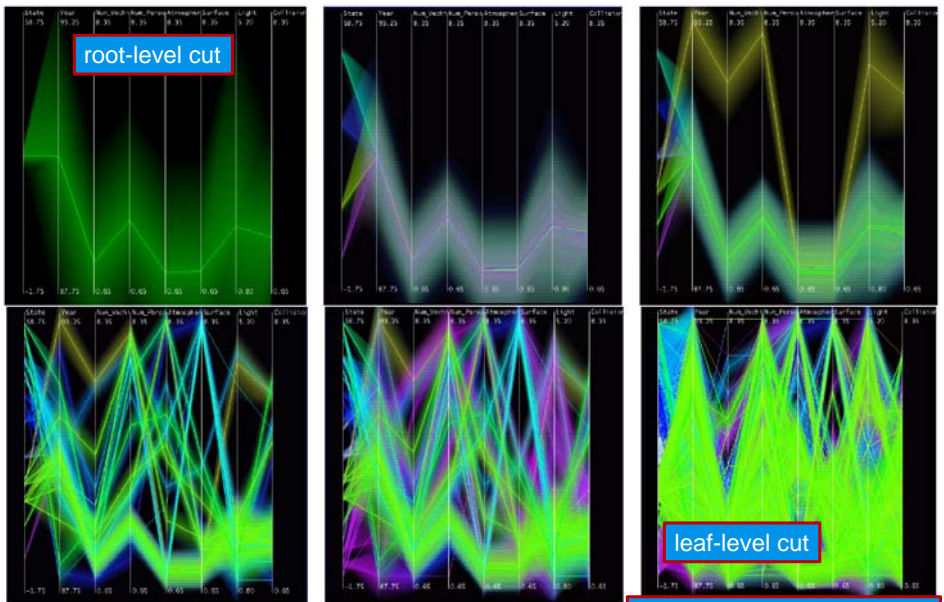

- Integration of projection/embedding

- Integration of classifcation/learning

Hierarchical Parallel Coordinates

root-level cut

leaf-level cut

[Fua, Ward, Rundensteiner; Vis 1999]

hierarchical clustering used

Calendar view example

power demand by ECN

outlier

special days

winter

summer

weekend
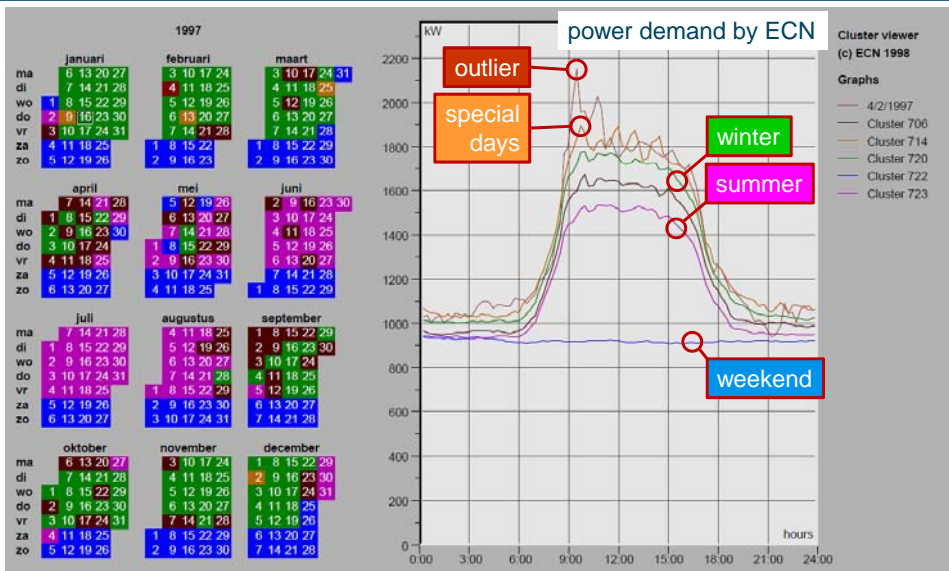
[van Wijk & van Selow; InfoVis 1999]

hierarchical clustering used

Time-varying Multivariate Climate Data

low-activity   high-activity

time-activity curves

importance curves

(a) low-activity temperature TACs   (b) medium-activity temperature TACs   (c) high-activity temperature TACs

(d) low-activity temperature ICs   (e) medium-activity temperature ICs   (f) high-activity temperature ICs
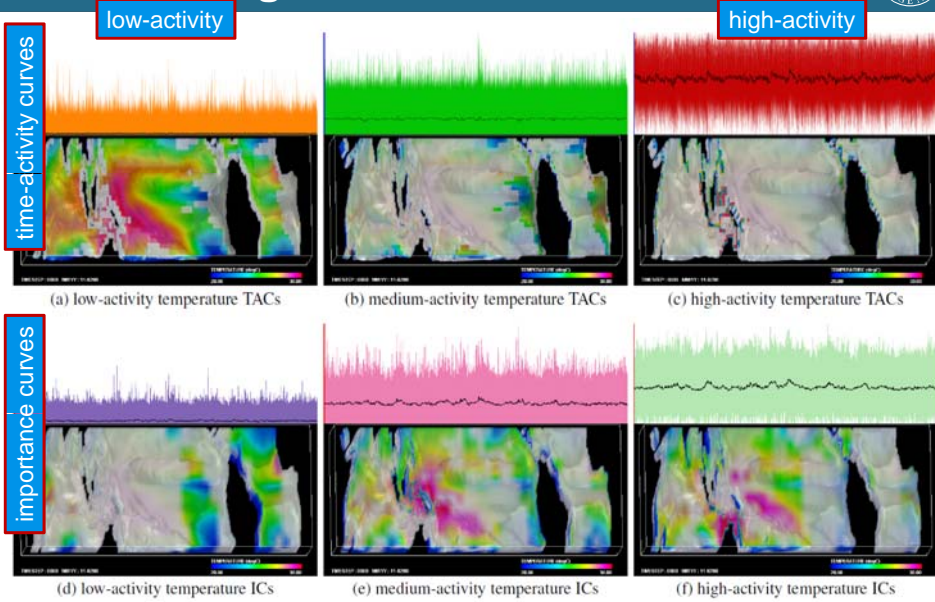
[Sukharev, Wang, Ma, Wittenberg; PacificVis 2009]   k-means clustering used

Perfusion data IVA

visualizing "tissue at risk" (c & e)

(a)   (c)   (e)

ischemic stroke

(b)   (d)   PC1, PC2 (f)

[Oeltze, Doleisch, Hauser, Muigg, Preim; TVCG 2007]

Using SOMs

low crime rates

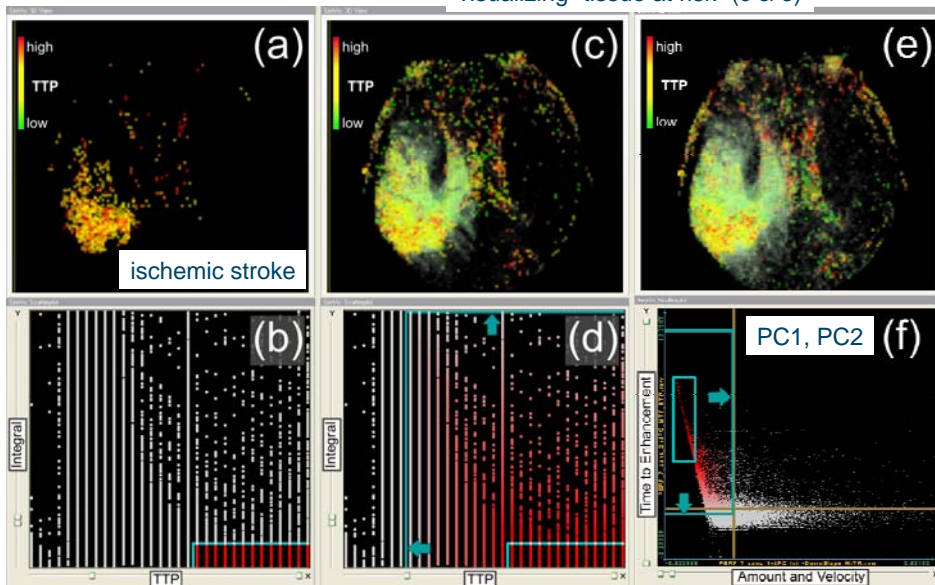space-in-time SOM

drop in 2000

high crime rates

1960 ▮▮▮▮▮▮▮▮▮▮ 2000

[Andrienko, Andrienko, Bremm, Schreck,
von Landesberger, Bak, Keim; CGF 2009]

− mean +
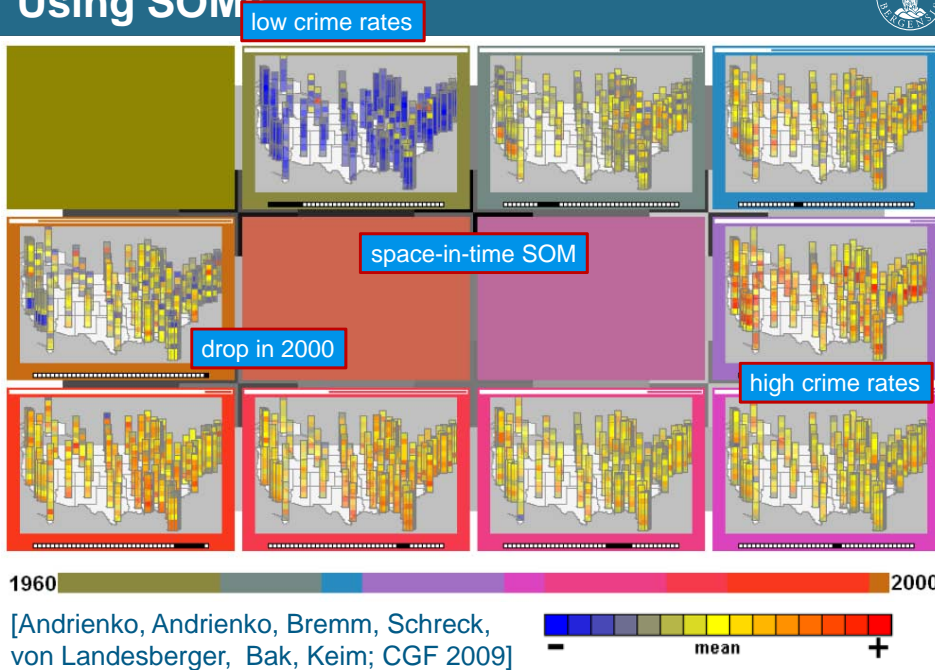
Web search visualization

nine search engines, MDS embedding

company
1. altavista
2. alltheweb
3. ask
4. google
5. lycos
6. live
7. yahoo
8. aol
9. dogpile

university

celebrity

google,
AOL

clusters

questions
1. altavista
2. alltheweb
3. ask
4. google
5. lycos
6. live
7. yahoo
8. aol
9. dogpile

sports

tourism

yahoo

[Sun, Lebanon, Collins-Thompson; WWW 2010]
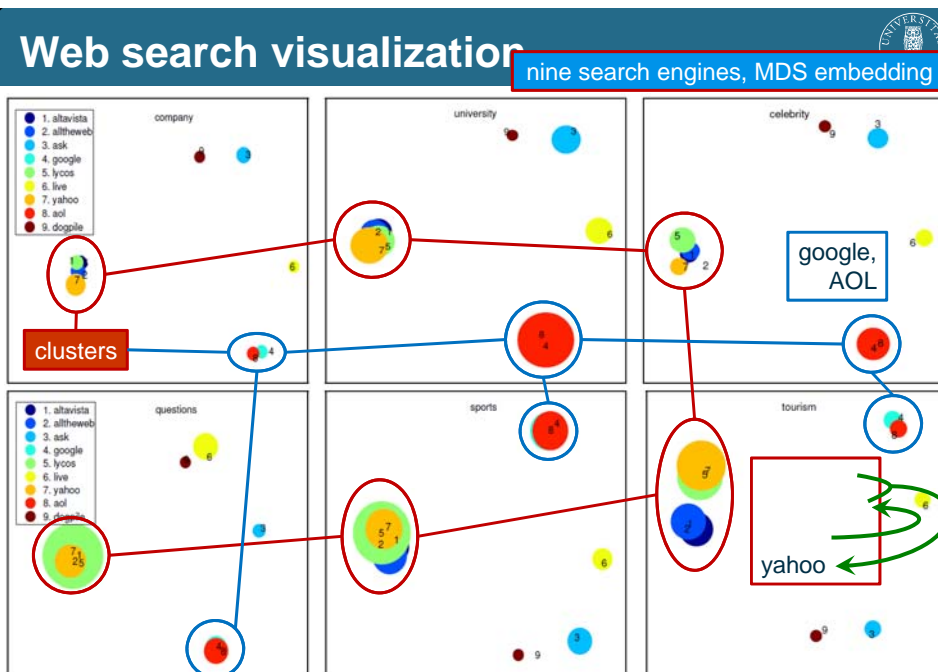
BaobabViews (decision trees)

tumor location classification

same classification, manual sex-split

easy to separate

neck = no — lung

axillar = no
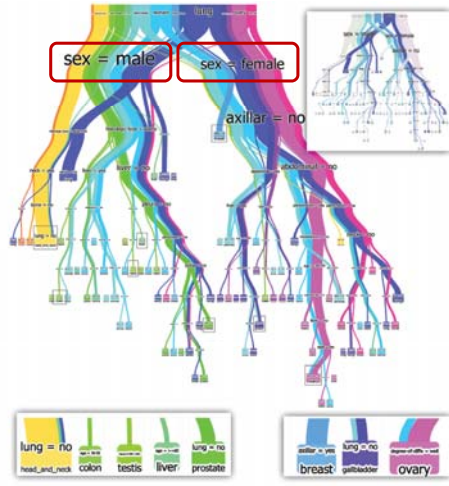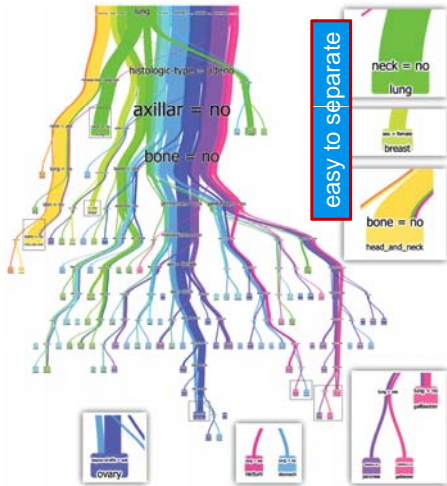
breast

bone = no

bone = no — head_and_neck

sex = male   sex = female

axillar = no

misclassification

hard to separate

typical male

typical female

lung = no — head_and_neck   colon   testis   liver   prostate — lung = no
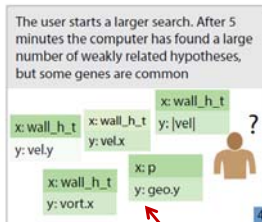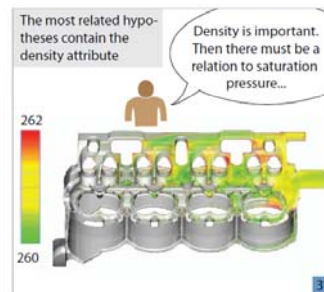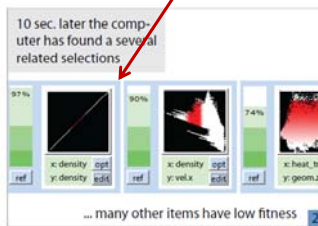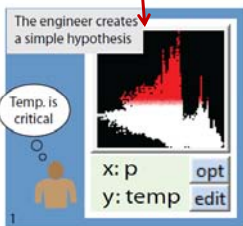
axillar = yes — breast   lung = no — gallbladder   ovary

[van den Elzen, van Wijk; VAST 2011]

Visual Human+Machine Learning

start focus: high temp.

ML tell: density is related!

The engineer creates a simple hypothesis

Temp. is critical

x: p   opt
y: temp   edit

10 sec. later the computer has found a several related selections

97 %   90 %   74 %

x: density   opt   x: density   opt   x: heat_tr
y: density   edit   y: vel.x   edit   y: geom.z

ref   ref   ref

... many other items have low fitness

The most related hypotheses contain the density attribute

Density is important. Then there must be a relation to saturation pressure...

262

260

The user starts a larger search. After 5 minutes the computer has found a large number of weakly related hypotheses, but some genes are common

x: wall_h_t
y: vel.y   x: wall_h_t   x: wall_h_t
y: vel.x   y: |vel| ?

x: wall_h_t   x: p
y: vort.x   y: geo.y

Visual analysis and optimization lead to three hypotheses

High temperatures occur when ...

A: density is medium and pressure is high

B: heat convection from the wall and recirculation

C: low pressure and lambda$_2$ negative

larger search: more results…

three interpretations (A, B, C)

[Fuchs, Waser, Gröller; TVCG 2009]

Problem solving with ML and InfoVis?

- Problem solving with ML and InfoVis
  - it's not about ML, it's not about InfoVis
  - it's about solving problems
    (with ML & InfoVis)

- Inner vs. outer integration

- Concept of a toolbox of enabling technologies
  - integrate visualization
    - show to the user
    - let the user interact
  - integrate ML
    - give feedback quickly
    - be explained

- A form of visual analytics (or visual computing)

Acknowledgements

- **You** – thank you for your attention!
  Question?

- Çağatay Turkay

- Many!