IEEE Visualization 2004 Conference

Information Visualization and Visual Discovery

Georges Grinstein
UMass Lowell
Daniel Keim
University of Konstanz
Matthew Ward
Worcester Polytechnic Institute

Emails and URLs

- Daniel A. Keim
  - keim@informatik.uni-konstanz.de
  - http://www.fmi.uni-konstanz.de/~keim
- George Grinstein
  - grinstein@cs.uml.edu
- Matthew Ward
  - ward@cs.wpi.edu

Goals of Tutorial

- Look at history of Visualization and Visual Data Mining (ScDV, mDV, EDA)
- Understand the issues in interactive data visualization
- Examine numerous visualization techniques and systems
- Look at a specific application example area (bioinformatics, …)
- Explore the future of visualization

outline

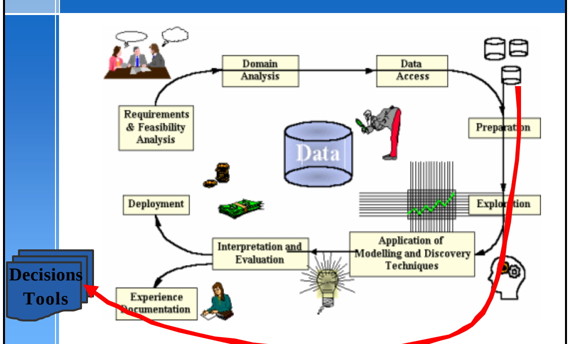
Part I: Introduction
1. *Goals of Information Visualization*
2. Definition of Information Visualization
3. History of Information Visualization

Part II: Foundations
1. Visualization Pipeline
2. Data Foundations
3. Perceptual Foundations
4. Visualization Foundations / Theory

Part III: Visualization Techniques
1. Classification
2. Visual Data Exploration Techniques
3. Distortion and Interaction Techniques
4. Visual Data Mining Systems

Part VI: Specific Visual Data Mining Techniques
1. Association Rules
2. Classification
3. Clustering
4. Text Mining
5. Tightly Integrated Visualization

The Knowledge Discovery Process

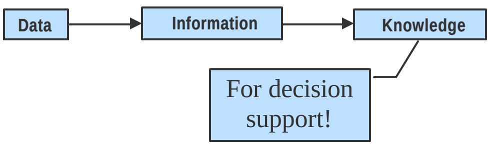


Data Exploration

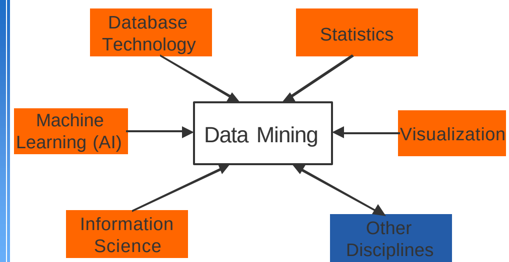
- Data Exploration is the process of searching and analyzing databases (or data sets) to discover implicit but potentially useful information
- Note the use of the terms
  - Implicit
  - Useful

Goals of Data Exploration

- Convey information
- Discover new knowledge
- Identify structure, patterns, anomalies, trends, relationships



A Confluence of Multiple Disciplines



Data Mining Tasks & Techniques

Major Data Mining Tasks
- Summarization
- Association
- Classification
- Prediction
- Clustering

using →

based on ←

Statistical Tools
- Missing Value Imputation
- Normalizations
- Error & Variational Analysis
- Confidence Estimates

Major Techniques
- Linear Regression Trees
- Non-Linear Regression
- MARS
- Naïve Bayes
- K-Means and K-Median
- Neural Networks
- Association Rules
- Decision Trees
- Principal Curve Analysis
- Support Vector Machines
- Genetic Algorithms

Why so many?

- Almost all tasks are NP-hard!
- KDD2001 CUP
  - Thrombosis data set
  - Over 200 submissions
  - Over 100 different techniques
  - Many combined techniques
- KDD2002 CUP
  - Creativity

Introduction to Visualization

- Many visualization techniques
- Various interaction techniques
- Lots of applications
- A number of visualization platforms
- A number of data mining platforms
- Few integrated platforms
- One can consider these visual data mining systems - if they are done the right way (☺)

Visualization Techniques

Pure
- 2D and 3D Scatterplots
- Matrix of Scatterplots
- Statistical Charts
- Line and Multi-line Graphs
- Parallel Coordinates
- Circle Segment
- Polar Charts
- Survey Plots
- Heatmaps
- Height Maps
- Iconographic Displays
- RadViz and PolyViz

Integrated with Analysis
- Projection Pursuit
- Dimensional Stacking
- Sammon Plots
- Multi-Dimensional Scaling
- PCA and Principal Curves
- Self Organizing Maps

Interactions
- Selection
- Probing, Querying
- Grand Tours
- Non-linear Zooms

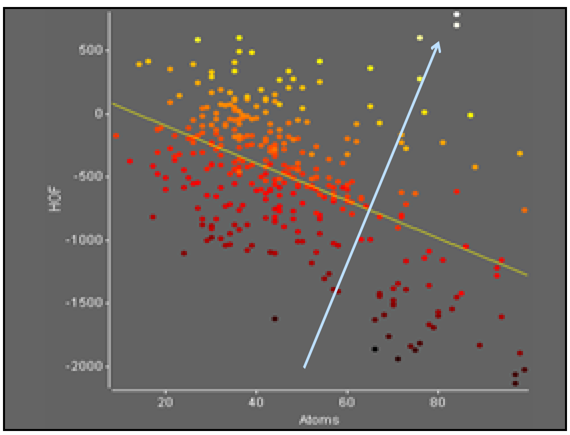
What do you do with Visualization?

- Explore data
- Explore algorithms
- Develop presentations
- Develop new visualizations
  - for statistics, missing values, noise, …
- Build theories

- In all cases remember that visualization is an interface technology to harness your human visual and intuitive capabilities
- In any analysis it supports the analysis steps – it is not a new discovery tool on its own

Goals of Visualization Techniques

- Presentation
  - Facts to be presented are fixed a priori
  - Need to select appropriate presentation techniques and parameters
  - Result is a high-quality visualization to present the known facts
- Confirmatory Analysis
  - Have some close to specific hypotheses about the data
  - Goal-oriented examination of these hypotheses
  - Result is a visualization of data that confirms, rejects or provides more information on the hypotheses
- Exploratory Analysis
  - Have no hypotheses or very broad ones about the data
  - Need to explore interactively, usually as undirected searches for structures, trends, ...
  - Result is a visualization of data that hopefully leads to hypotheses about the data



outline

What is Visualization?

- "Visualization is a method of computing. It transforms the symbolic into the geometric, enabling researchers to observe their simulations and computations. Visualization offers a method for seeing the unseen." (from McCormick87)

- Visualization is an interface technology to data, algorithms, and any human task

  (Grinstein)

- Visualization now includes other data representations
  - Auditory, haptic and tactile

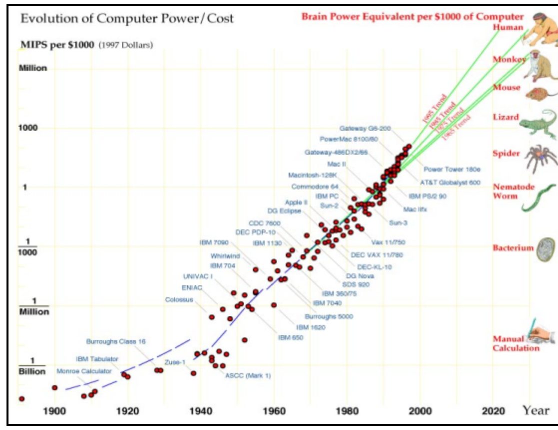
The Great Demand for Visualization

- Fueled by technological advancements
  - Displays
  - High performance computers
  - Large storage systems
  - Personal computers
  - Sensor technology
  - Communication
- Fueled by a growing user awareness and experience
  - Interfaces
  - Programming tools
  - The Web

Global Computing Applications

- 48-hour Weather Forecast
- 2D Airfoil
- Oil Reservoir Model
- Climate Monitoring
- Vehicle Signature

- Plasma Modeling
- Chemical Dynamics
- Stock Market Prediction
- Drug Discovery



A Definition of Visualization

- It is the Visual Interface to the data and mining tools
- It is a method of interacting with data and algorithms
  - supports the user through all the knowledge discovery steps
  - uses selections, queries, probes, and view transformations
- It is completely separable from the analysis methods
  - Data can be analyzed using many different algorithms
  - Each result can be viewed in a different visualization
  - Each visualization thus provides a different view of the results

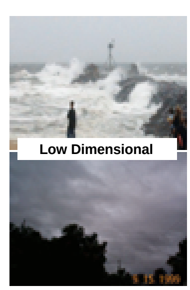
What are the Key Data Factors?

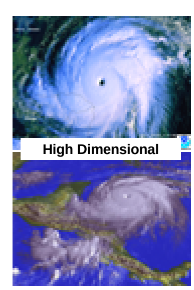
- Very large number of parameters
  - more than 100
- Very large data sets
  - more than 10^7
- Multiple data types
  - discrete and continuous
- Noisy data
  - often not uniform
- Missing values
  - could be important
- Lots of different tasks

What is High Dimensional?

# of Variables	Dimensionality
˜ 10	Low
˜ 100	Medium
˜ 1000	High
> 1000	Very High

A Complete Data View

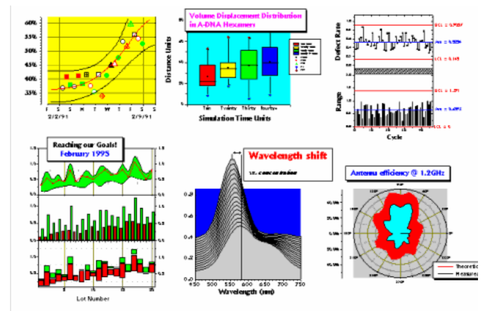




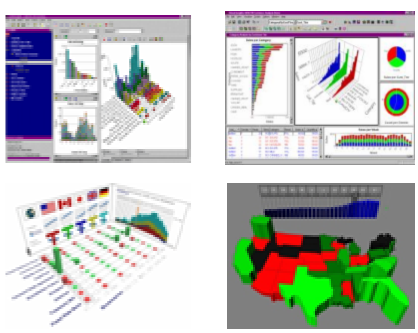
Visual Data Mining

- Definition
  Data Mining is the process of searching and analyzing databases to find implicit but potentially useful information
- More formally
  Data Mining is the process of finding a subset D' of the database D and hypotheses H_U(D',C) that a user *U* considers useful in an application context C

Classic Visualization Techniques

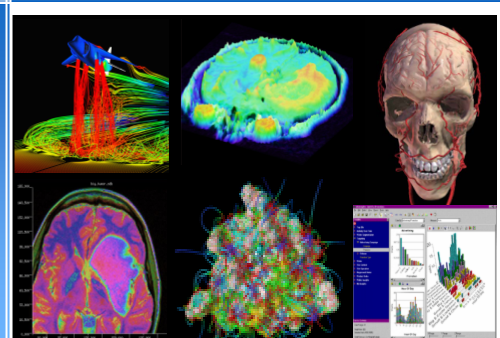


Standard 2D/3D Displays

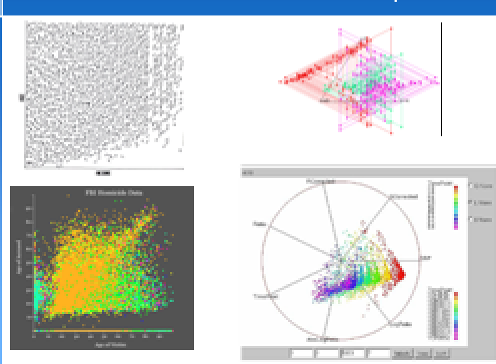


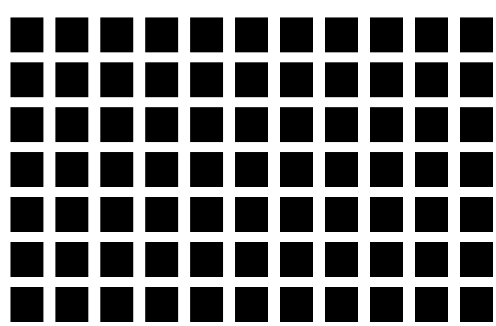
Examples from theVisualInsights WebPage

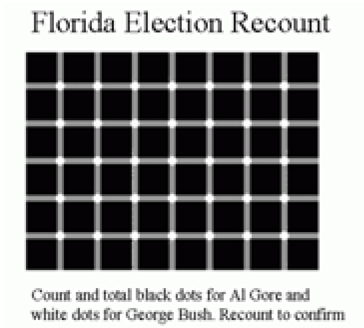
Modern Visualization Techniques



Ultra Modern Visualization Techniques







outline

A History of Visualization

- Pictures
  - From hieroglyphics to spreadsheets
  - From lines to surface and volumes
  - From scatterplots to HDVs
  - From static to dynamic images
  - From simple to complex integrated analysis
- Slides

5000 BC

2000 AD

Brief Historical Overview of Exploratory Data Visualization Techniques

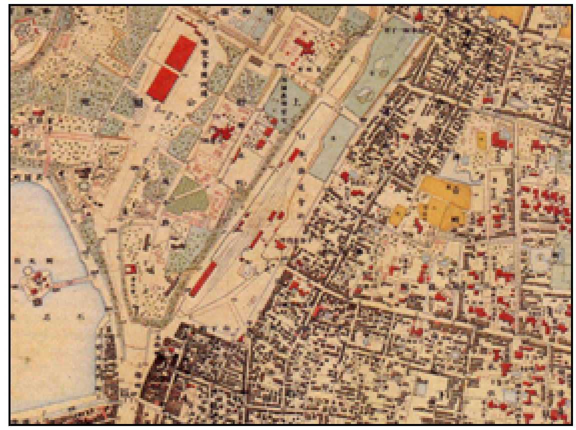
- Pioneering work of Tufte [Tuf 83, Tuf 90] and Bertin [Ber 81] focused on presentation
  - visualization of data with inherent 2D and 3D semantics
  - general rules for layout, color composition, attribute mapping, etc.
- Development of visualization techniques for different types of data with an underlying physical model for confirmatory visualizations
  - geographic data, CAD data, flow data, image data, voxel data, etc.
- Development of visualization techniques for arbitrary multidimensional data (without an underlying physical model) for exploratory visualizations
  - applicable to databases and other information resources

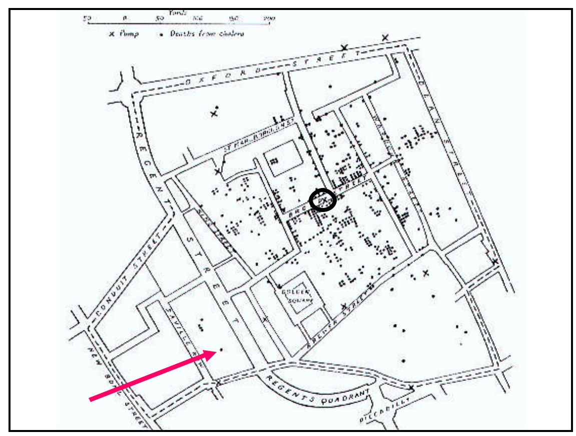
1-10 Variables

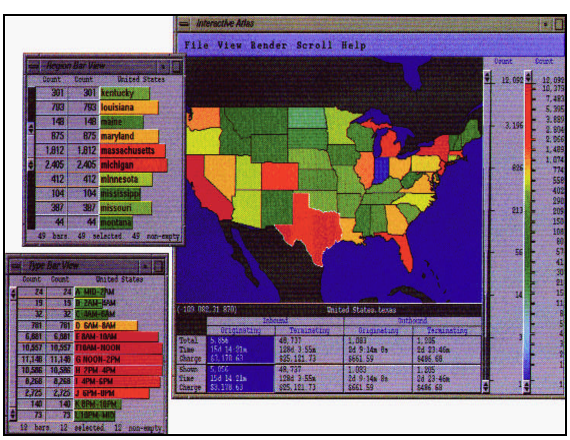
Maps and Earthy Colors (1800s)

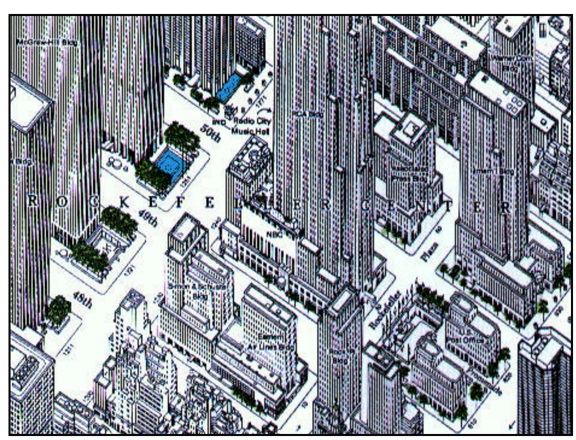
Maps

- Valuable
  - Save time, money, lives
- Provide an anchoring image
  - Experience base
  - Reasoning base
- Understandable









Early Visualization Fuels

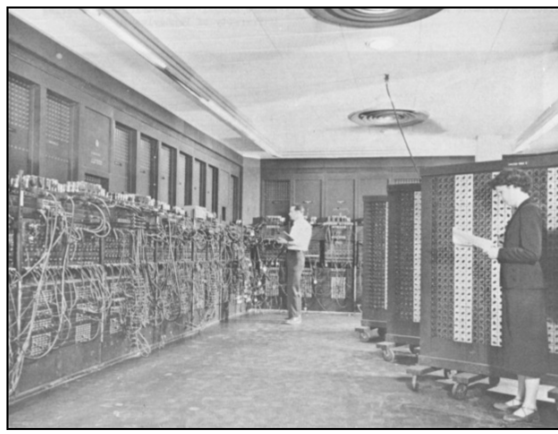
- Military
- Aerospace and Automotive
- Entertainment

- Scientific Data Visualization
- GIS

NASA Movie

Classic Science

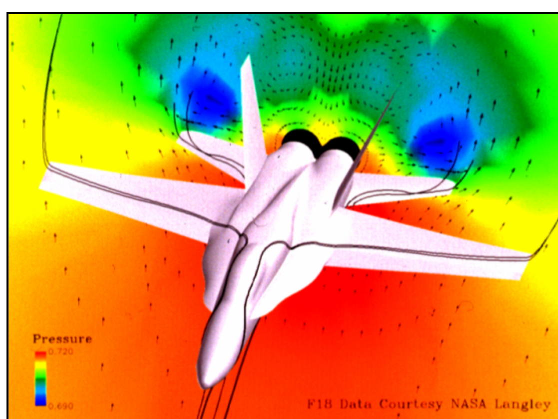
- Build Model
- Validate Model using Real Data
- Repeat

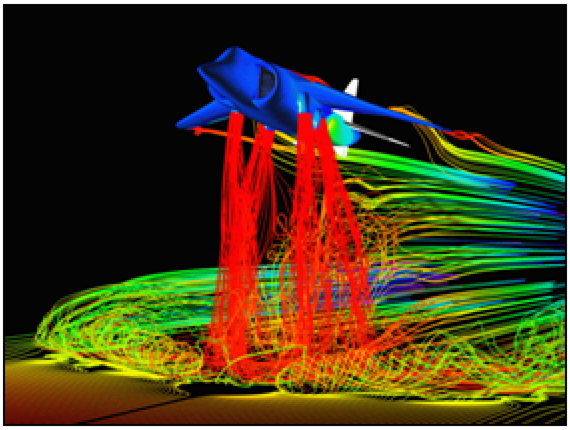


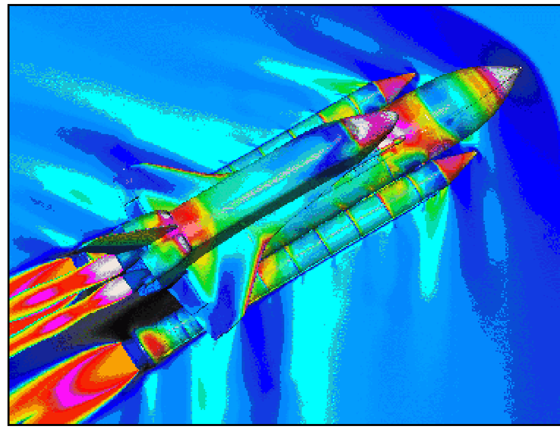
Aerospace

Aircraft Data

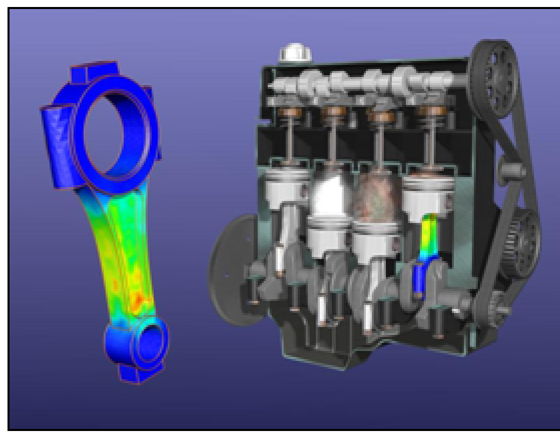
- Velocity = 165 knots
- Wing Area = 29 m²
- Wing Span = 16 m
- Mean Aerodynamic Chord = 2 m
- Weight = 8000 kg
- Chord Reynolds Number = 1.18×10^7







Computer-Aided Design



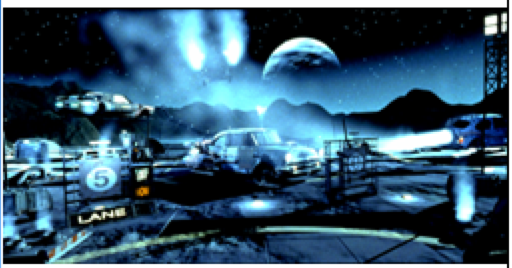
Modern Visualization Fuels

- Entertainment
- Medicine
- Architecture
- Art

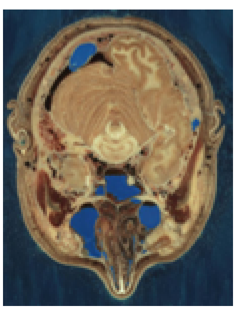
- The Web and public demand

- New approach to discovery
  - Collect lots of data
  - Look for patterns in that data
  - See if that pattern is meaningful (and new)

Film and Entertainment

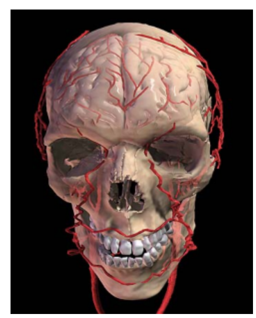


Visible Digital Human (2D Scans)



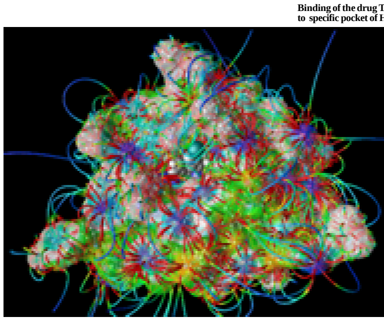
- Head, including cerebellum
- Cerebral cortex, brainstem
- Nasal passages from Head subset

Medical Imagery (3D Scans)



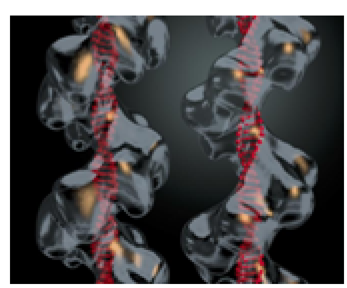
HIV-1 Target

Binding of the drug TIBO R86183 to specific pocket of HIV -I enzyme

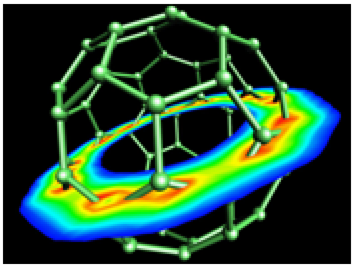


DNA (Electron Microscopy)

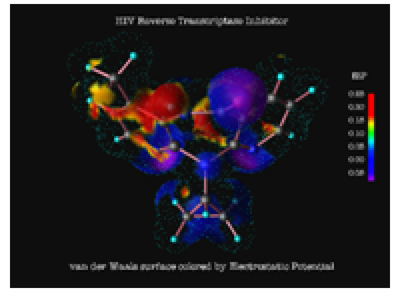
Bacterial RecA and eukaryotic Rad51 Proteins form similar filaments on DNA

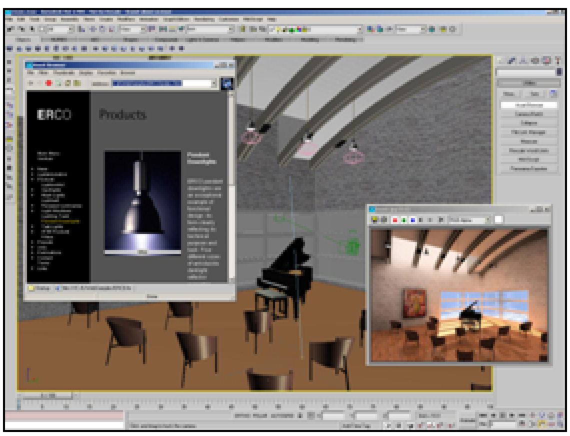


Bucky's Ball (C60)



Electrostatic Potential of the HIV Reverse Transcriptase Inhibitor





½ Gram of Fat

outline

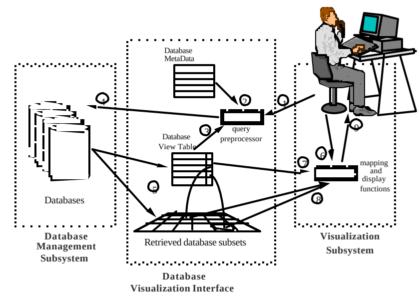
Part I: Introduction
1. Goals of Information Visualization
2. Definition of Information Visualization
3. History of Information Visualization

Part II: Foundations
1. *Visualization Pipeline*
2. Data Foundations
3. Perceptual Foundations
4. Visualization Foundations / Theory

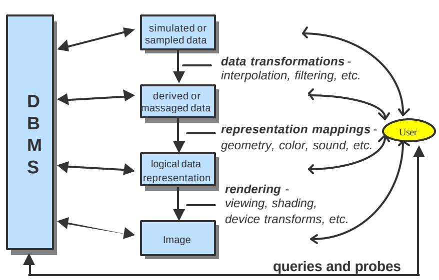
Part III: Visualization Techniques
1. Classification
2. Visual Data Exploration Techniques
3. Distortion and Interaction Techniques
4. Visual Data Mining Systems

Part VI: Specific Visual Data Mining Techniques
1. Association Rules
2. Classification
3. Clustering
4. Text Mining
5. Tightly Integrated Visualization

Visualization Architecture



The Visualization Pipeline



Output Devices

- CRTs and LCDs
- Projection Systems
- Stereoscopic Displays
- Head Mounted Displays
- 3-D devices
- Aural and Haptic Displays

Input Devices

- Keyboards
- Pointers
  - light pen (1960-1970)
  - mouse (1960-1990)
  - wand (1990)
- Touchscreens
- 3D and 6D
  - Spaceball, Flying mouse, …
- Gloves
- Haptic and Tactile devices
- Digitizers/scanners
- Cameras

Distributed Systems Architectures

- Client - Server Architecture
- Homogeneous Systems
  - Small scale parallelism
  - Massively Parallel (Processors and Workstations)
- Heterogeneous Systems
  - Transparent parallelism

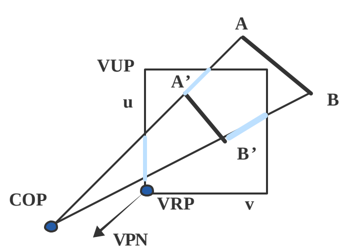
Rendering Scenes

- Define scene
  - objects, attributes, light sources and types
- Specify viewing parameters
- Render using
  - The camera model
  - Light models
  - Rendering models

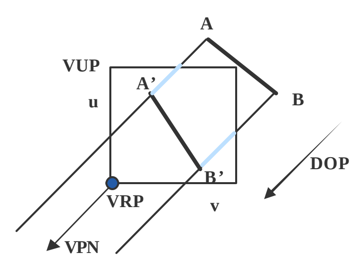
The Camera Model

- Projection
  - parallel
  - perspective
  - fisheye
- Mathematics
  - center of projection
  - display (screen) surface
    - u,v coordinate system
    - clipping region
  - view up vector
  - front and back clipping planes

Perspective Projection and Parameters



Parallel Projection and Parameters



Shading Fundamentals

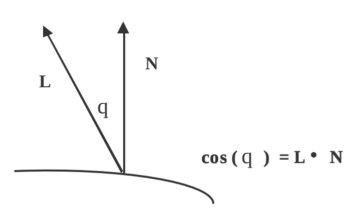
- $I = k_a I_a + k_d I_d + k_s I_s$ where
  I = the intensity at a pixel on the screen and

  I_a = the contribution from ambient light sources
  I_d = the contribution from diffuse light sources
  I_s = the contribution from specular light sources
  and the k_i represent the object's shading coefficients

- I is computed for each color component R, G, B

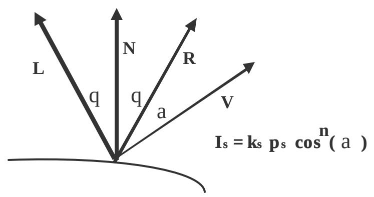
Simple Shading Models: Diffuse Reflection

- Lambert's law: $I_d = k_d I_p \cos(\theta)$ where θ is the angle between the normal vector and the light source



Simple Shading Models: Specular Reflection

- $I_s = I_p \, W(\theta) \cos^n(\alpha)$ where α is the angle between the reflected vector and viewer ray



Polygon Rendering

- Constant shading
  - each triangle is rendered with the object's constant color
- Flat shading
  - each triangle is rendered with a constant color determined by the normal vector to the triangle
- Gouraud shading
  - each pixel in each triangle is rendered with its color determined by interpolating the vertices' colors (trilinearly)
- Phong shading
  - each pixel in each triangle is rendered with its color determined by interpolating the vertices' normal vectors, then determining the color from the interpolated vectors

Scene Rendering Parameters

- Geometric primitives
  - points, lines, surfaces, volumes
  - curves, meshes
  - polyhedra, splines, NURBS
- Object representation
  - Boundary representation
  - Constructive solid geometry
- Displays
  - projections
  - multiple views
  - stereoscopic
  - clipping
  - dynamics
- Anti-aliasing and filters
- Hidden object removal
  - points, lines, surfaces
- Surface properties
  - bumps, scars, burns
  - color
  - texture
  - transparency
  - reflection, refraction
- Shadows
- Motion blur, depth cueing, ...
- Special effects
  - clouds
  - Fire
  - Water
  - trees, ...

Rendering Approaches

- Painter's algorithm
- Z-buffer algorithm
- Hidden object removal
- Spatial subdivision techniques
- Ray tracing
- Radiosity
- Hybrid techniques
  - radiosity + specular + shadows

Graphics API

- Application programmer's Interfaces
- Control screen surface drawing areas
- Provide for points, lines, curves, surfaces
- Provide attributes and grouping controls
- Support lighting, shading, cameras...

Goals of Visualization Techniques

- Presentation
  - Facts to be presented are fixed a priori
  - Select appropriate presentation techniques
  - Result is a high-quality visualization of the data to present known facts
- Confirmatory Analysis
  - Have specific hypotheses about the data
  - Goal-oriented examination of the hypotheses
  - Result is a visualization of data that confirms, rejects or provides more information on the hypotheses
- Exploratory Analysis
  - Have no hypotheses about the data
  - Explore interactively, usually as undirected searches for structures, trends, ...
  - Result is visualization of data that hopefully leads to hypotheses about the data

outline

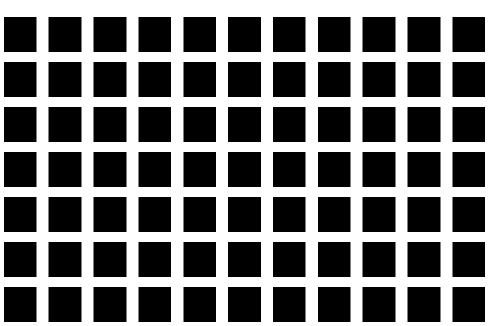
Data Preprocessing Techniques Techniques for Dimensional Reduction

- d-dim data → k-dim data (k<<d)
- Principal Component Analysis [DE 82]
  - Determines a minimal set of principal components (linear combinations of the original dimensions) which explain the main variations in the data
- Factor Analysis [Har 67]
  - Determines a set of unobservable common factors which explain the main variations in the data. The original dimensions are linear combinations of the common factors.
- Multidimensional Scaling [SRN 72]
  - Use the similarity (or dissimilarity) matrix of the data to define coordinate axes in multidimensional space. Similarity can be determined by the Euclidean distance in that space.
- Fastmap [FL 95]
  - Using a similarity matrix iteratively reduce the number of dimensions while preserving the distances as much as possible

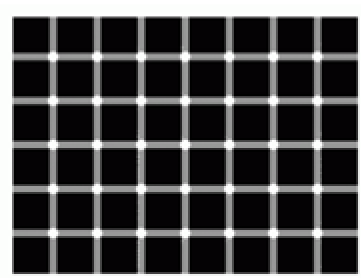
Data Preprocessing Techniques

- Subsetting Techniques
  - Set of Data Items → Subset of Data Items
  - Sampling determines a representative subset of a database
  - Querying determines a certain, usually a-priori fixed subset of the database
- Segmentation Techniques
  - Set of Data-Items -> Set of (Set of Data Items))
  - Segmentation based upon attribute values or attribute ranges
- Aggregation Techniques
  - Set of Data-Items -> Set of Aggregate Values)
  - Aggregation (sum, count, min, max,...) based upon
    - attribute values
    - topological properties, etc.
  - Visualization of Aggregations
    - Histograms
    - Pie Charts, Bar Charts, Line Graphs, etc.

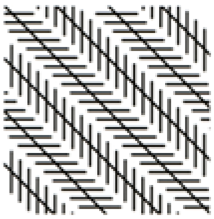
outline

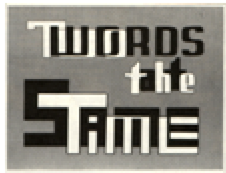


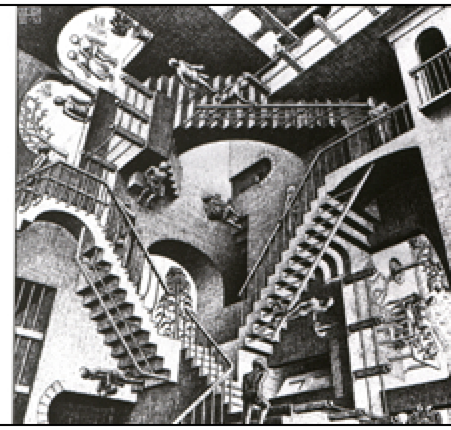
Florida Election Recount

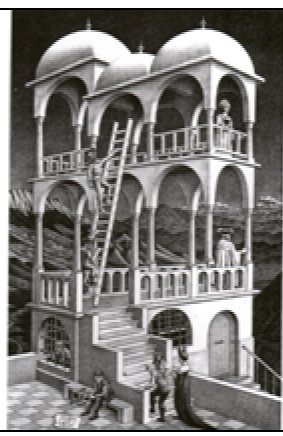


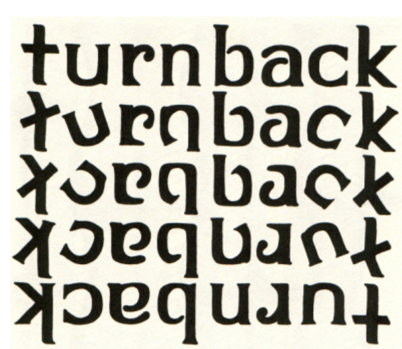
Count and total the black dots for Al Gore and the white dots for George Bush. Recount to confirm.

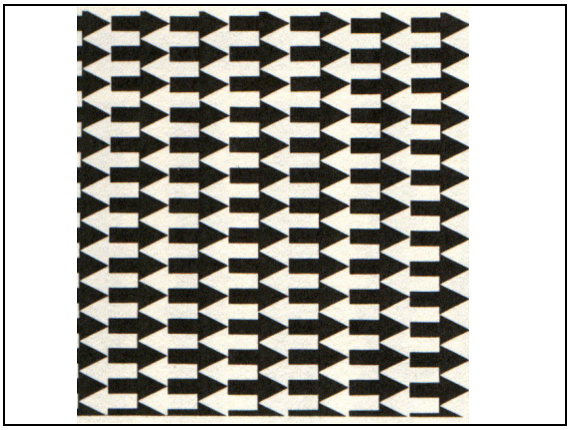


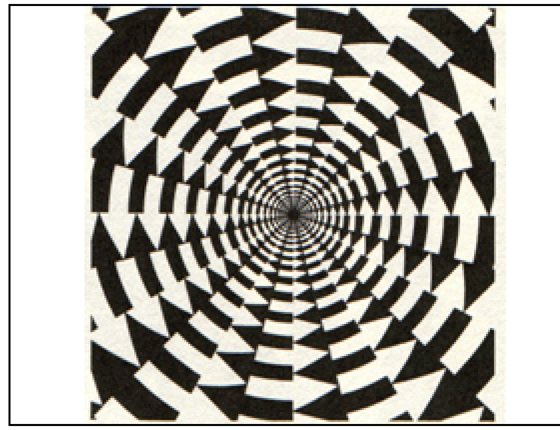

WORDS
the
SAME

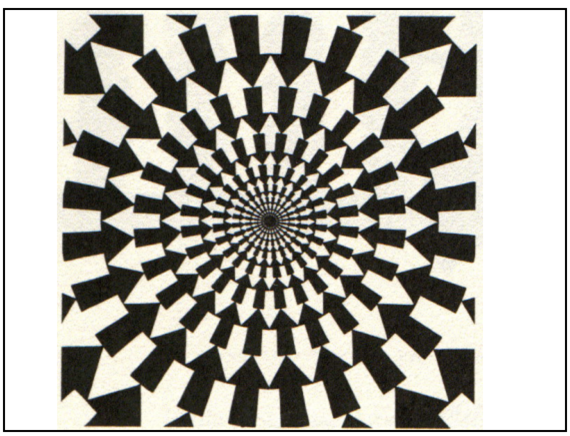





turnback
turnback








MIRROR IMAGES
The paintings shown here are confusing swirls of lines and colors. What do they mean? In fact, they have been painted like this on purpose.

How well do you read?
How well do you see colors?

Which is faster?

Perception Tests

Perception Examples

Perception Tests

Reading 1

Aoccdrnig to a rscheearch at an Elingsh uinervtisy, it deosn't mttaer in waht oredr the ltteers in a wrod are, the olny iprmoetnt tihng is taht frist and lsat ltteer is at the rghit pclae. The rset can be a toatl mses and you can sitll raed it wouthit porbelm. Tihs is bcuseae we do not raed ervey lteter by it slef but the wrod as a wlohe

Wow!

An article a few months ago was circulating around reporting on the claim that

scrambled words are legible

as long as first and last letters are in place

Reading 2

Anidroccg to crad cniyrrag lcitsiugnis planoissefors at an uemannd, utisreviny in Bsitirh Cibmuloa, and crartnoy to the duoibus cmials of the ueticnd rcraeseh, a slpmie, macinahcel ioisrevnn of ianretnl cretcarahs araepps sneiciffut to csufnoe the eadyrevy oekoolnr

Reading 2

According to card carrying linguistics planoissefors at an uemannd, utisreviny in Bsitirh Cibmuloa, and crartnoy to the duoibus cmials of the ueticnd rcraeseh, a slpmie, macinahcel ioisrevnn of ianretnl cretcarahs araepps sneiciffut to csufnoe the eadyrevy oekoolnr

Reading 2

According to card carrying linguistics professionals at an unnamed, university in Bsitirh Cibmuloa, and crartnoy to the duoibus cmials of the ueticnd rcraeseh, a slpmie, macinahcel ioisrevnn of ianretnl cretcarahs araepps sneiciffut to csufnoe the eadyrevy oekoolnr

Reading 2

According to card carrying linguistics professionals at an unnamed, university in British Columbia, and crartnoy to the duoibus cmials of the ueticnd rcraeseh, a slpmie, macinahcel ioisrevnn of ianretnl cretcarahs araepps sneiciffut to csufnoe the eadyrevy oekoolnr

Reading 2

According to card carrying linguistics professionals at an unnamed, university in British Columbia, and crartnoy to the duoibus cmials of the ueticnd rcraeseh, a slpmie, macinahcel ioisrevnn of ianretnl cretcarahs araepps sneiciffut to csufnoe the eadyrevy oekoolnr

Reading 2

According to card carrying linguistics professionals at an unnamed, university in British Columbia, and contrary to the duoibus cmials of the ueticnd rcraeseh, a slpmie, macinahcel ioisrevnn of ianretnl cretcarahs araepps sneiciffut to csufnoe the eadyrevy oekoolnr

Reading 2

According to card carrying linguistics professionals at an unnamed, university in British Columbia, and contrary to the dubious claims of the ueticnd research, a slpmie, macinahcel ioisrevnn of ianretnl cretcarahs araepps sneiciffut to csufnoe the eadyrevy oekoolnr

Reading 2

According to card carrying linguistics professionals at an unnamed, university in British Columbia, and contrary to the dubious claims of the ueticnd research, a slpmie, macinahcel ioisrevnn of ianretnl cretcarahs araepps sneiciffut to csufnoe the eadyrevy oekoolnr

Reading 2

According to card carrying linguistics professionals at an unnamed, university in British Columbia, and contrary to the dubious claims of the uncited research, a slpmie, macinahcel ioisrevnn of ianretnl cretcarahs araepps sneiciffut to csufnoe the eadyrevy oekoolnr

Reading 2

According to card carrying linguistics professionals at an unnamed, university in British Columbia, and contrary to the dubious claims of the uncited research, a simple, mechanical inversion of ianretnl cretcarahs araepps sneiciffut to csufnoe the eadyrevy oekoolnr

Reading 2

According to card carrying linguistics professionals at an unnamed, university in British Columbia, and contrary to the dubious claims of the uncited research, a simple, mechanical inversion of internal characters appears sufficient to csufnoe the eadyrevy oekoolnr

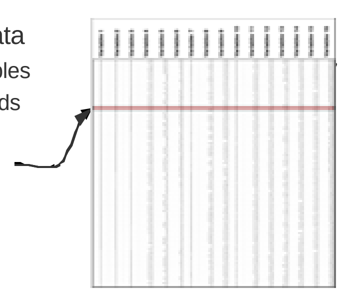
Reading 2

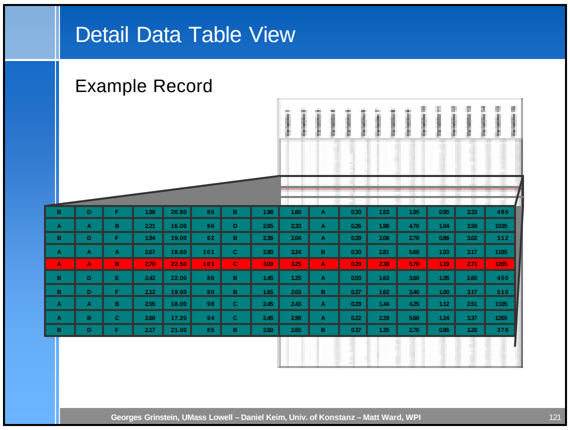
According to card carrying linguistics professionals at an unnamed, university in British Columbia, and contrary to the dubious claims of the uncited research, a simple, mechanical inversion of internal characters appears sufficient to confuse the everyday onlooker

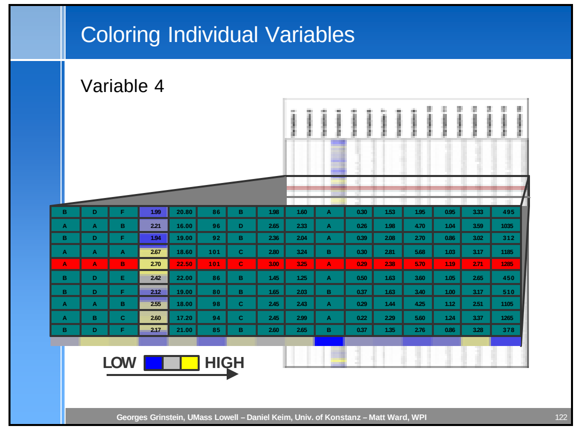
outline

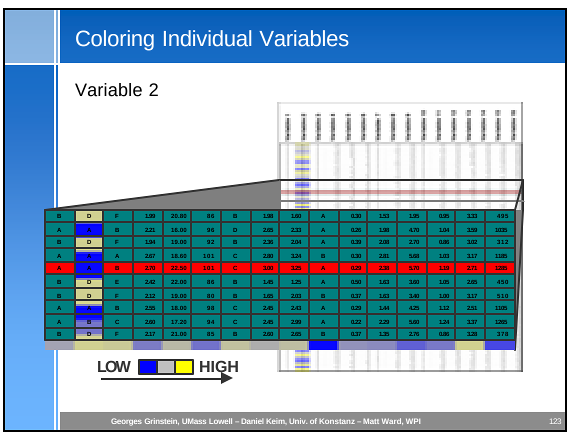
Input Data Table

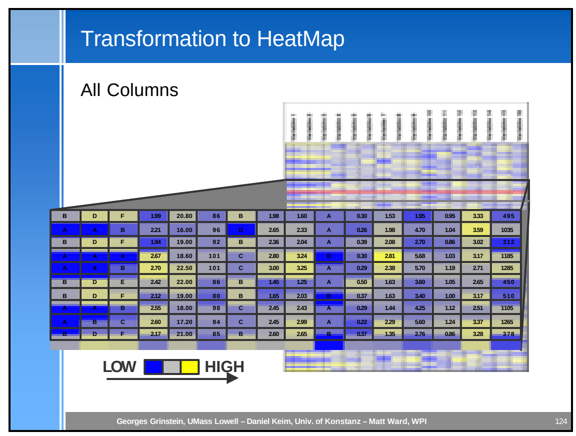
Example Data
- 16 Variables
- 98 Records

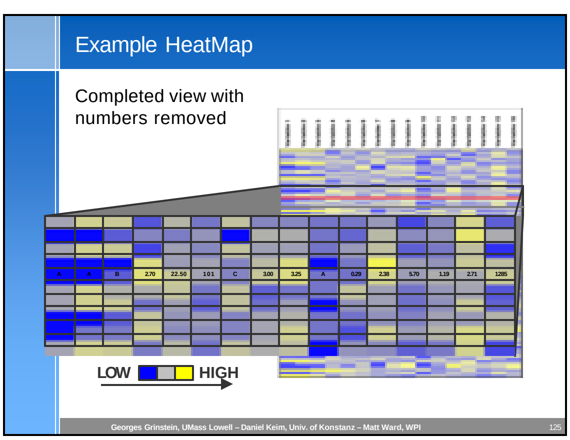


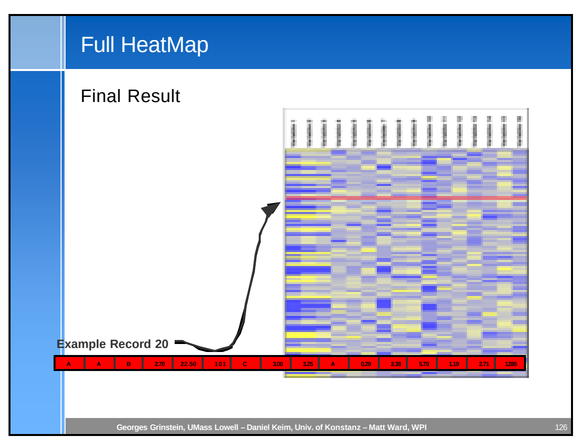

Detail Data Table View
Example Record
Georges Grinstein, UMass Lowell – Daniel Keim, Univ. of Konstanz – Matt Ward, WPI
121


Coloring Individual Variables
Variable 4
LOW
HIGH
Georges Grinstein, UMass Lowell – Daniel Keim, Univ. of Konstanz – Matt Ward, WPI
122


Coloring Individual Variables
Variable 2
LOW
HIGH
Georges Grinstein, UMass Lowell – Daniel Keim, Univ. of Konstanz – Matt Ward, WPI
123

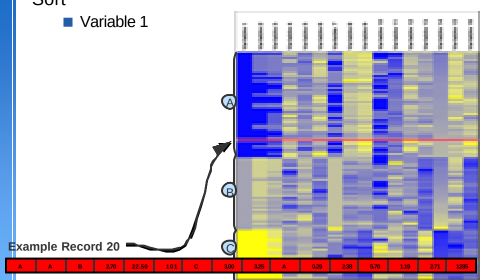

Transformation to HeatMap
All Columns
LOW
HIGH
Georges Grinstein, UMass Lowell – Daniel Keim, Univ. of Konstanz – Matt Ward, WPI
124


Example HeatMap
Completed view with
numbers removed
LOW
HIGH
Georges Grinstein, UMass Lowell – Daniel Keim, Univ. of Konstanz – Matt Ward, WPI
125


Full HeatMap
Final Result
Example Record 20
Georges Grinstein, UMass Lowell – Daniel Keim, Univ. of Konstanz – Matt Ward, WPI
126

HeatMap to Parallel Coordinates

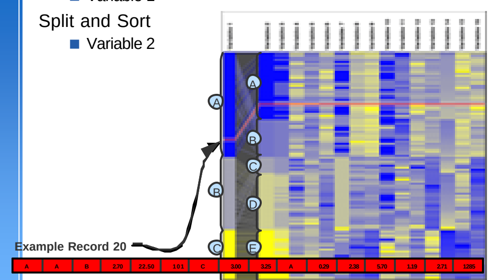
Sort
- Variable 1



First Table Split

- Variable 1

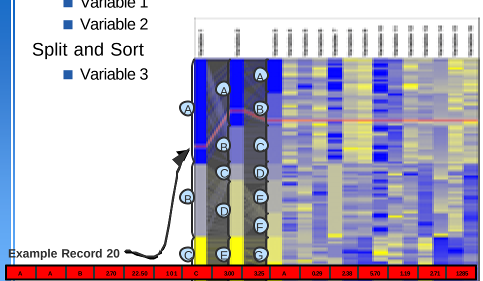
Split and Sort
- Variable 2



Second Table Split

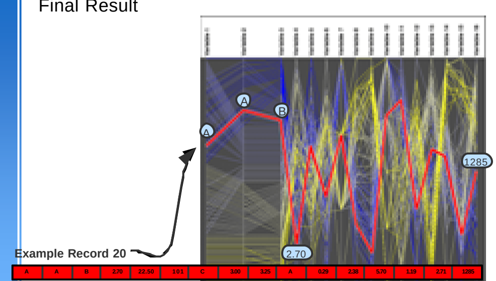
- Variable 1
- Variable 2

Split and Sort
- Variable 3



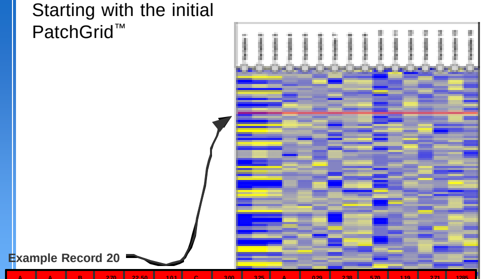
Parallel Coordinates

Final Result



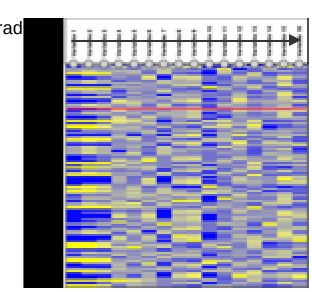
HeatMap to RadViz™

Starting with the initial PatchGrid™



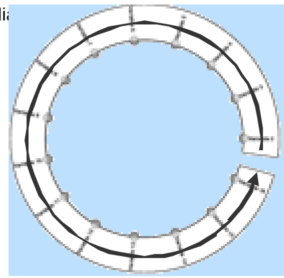
Radial Layout

Transforming a rectangular to a rad layout



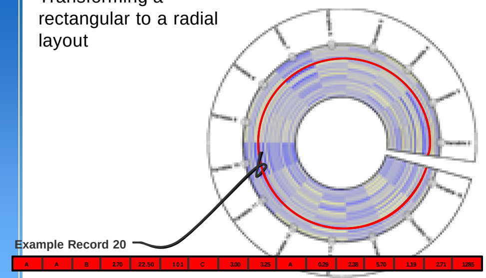
Radial Layout

Transforming a rectangular to a radial layout



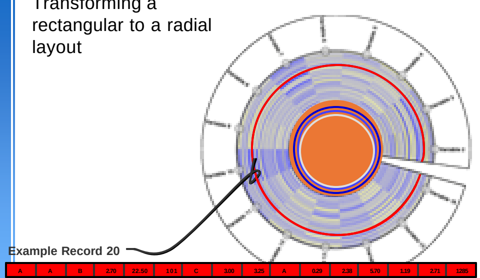
Radial Layout

Transforming a rectangular to a radial layout



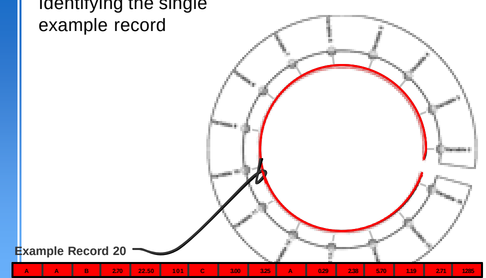
Radial Layout

Transforming a rectangular to a radial layout



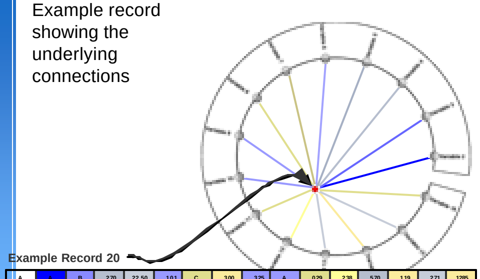
Radial Layout

Identifying the single example record

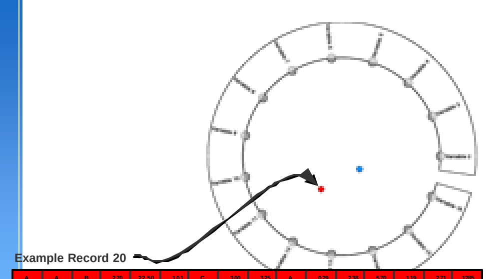


RadViz™ Fundamentals

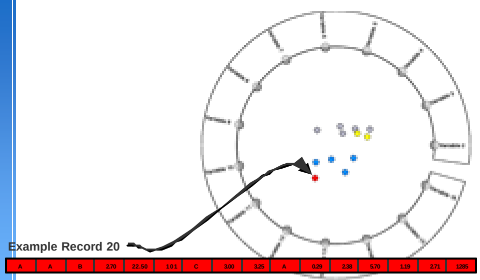
Example record showing the underlying connections



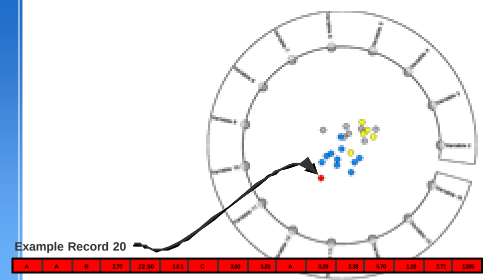
RadViz™



RadViz™

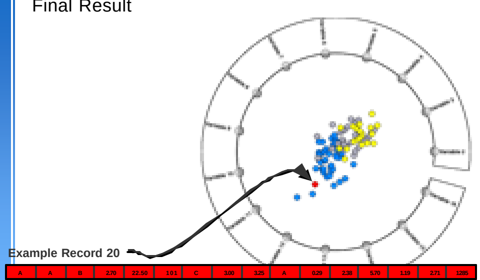


RadViz™



RadViz™

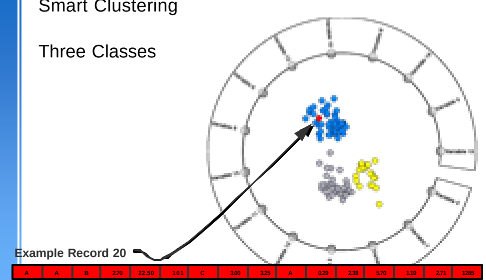
Final Result



RadViz™ Visual Classifier

Smart Clustering

Three Classes



Outline

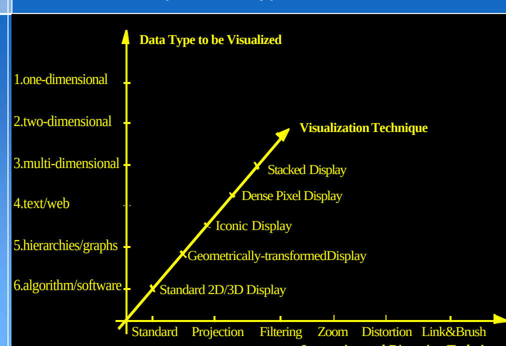
Part I: Introduction
1. Goals of Information Visualization
2. Definition of Information Visualization
3. History of Information Visualization

Part II: Foundations
1. Visualization Pipeline
2. Data Foundations
3. Perceptual Foundations
4. Visualization Foundations / Theory

Part III: Visualization Techniques
1. *Classification*
2. Visual Data Exploration Techniques
3. Distortion and Interaction Techniques
4. Visual Data Mining Systems

Part VI: Specific Visual Data Mining Techniques
1. Association Rules
2. Classification
3. Clustering
4. Text Mining
5. Tightly Integrated Visualization

Classification (Taxonomy)



Outline

Part I: Introduction
1. Goals of Information Visualization
2. Definition of Information Visualization
3. History of Information Visualization

Part II: Foundations
1. Visualization Pipeline
2. Data Foundations
3. Perceptual Foundations
4. Visualization Foundations / Theory

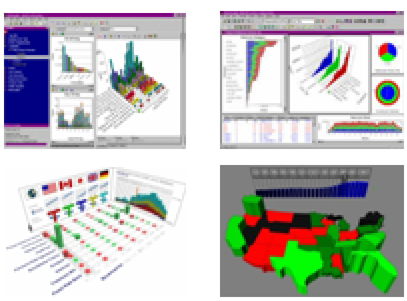
Part III: Visualization Techniques
1. Classification
2. *Visual Data Exploration Techniques*
3. Distortion and Interaction Techniques
4. Visual Data Mining Systems

Part VI: Specific Visual Data Mining Techniques
1. Association Rules
2. Classification
3. Clustering
4. Text Mining
5. Tightly Integrated Visualization

Visual Data Exploration Techniques

- Standard 2D/3D Displays
- Geometric Transformations
- Iconic Displays
- Dense Pixel Displays
- Stacked Displays

Standard 2D/3D Displays



Examples from theVisualInsights WebPage

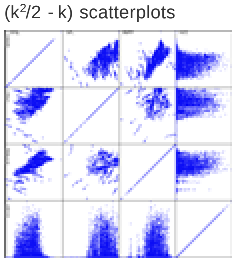
Geometric Transformations

- Basic Idea
  Visualization of geometric transformations and projections of the data
- Scatterplot-Matrices [And 72, Cle 93]
- Landscapes [Wis 95]
- Projection Pursuit Techniques [Hub 85] (a techniques for finding meaningful projections of multidimensional data)
- Prosection Views [FB 94, STDS 95
- Hyperslice [WL 93]
- Parallel Coordinates [Ins 85, ID 90]

Geometric Transformations Scatterplot Matrices [Cle 93]

- matrix of scatterplots
- (x,y) diagrams of the k-dim. data
- A total of ($k^2/2$ - k) scatterplots

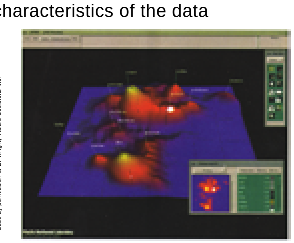
Used by permission of M. Ward, Worcester Polytechnic Institute



Geometric Transformations Lanscapes [Wis 95]

- Visualization of the data as a perspective landscape
- Goal is to transform the data into a (possibly artificial) 2D spatial representation which preserves characteristics of the data

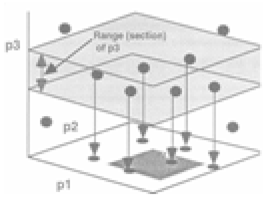
Used by permission of B. Wright, Visible Decisions Inc.

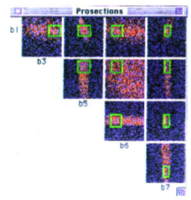


news articles visualized as a landscape

Geometric Transformations
Prosection Views [FB 94, STDS 95]

- schematic representation example
- matrix of all orthogonal projections where the result of
- the selected multidimensional range is colored differently (combination of selections and projections)



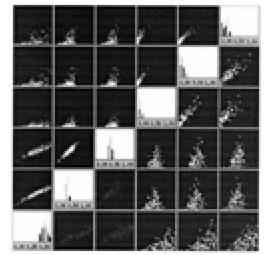


used by permission of R. Spence, Imperial College London

Geometric Transformations
Hyperslice [93]

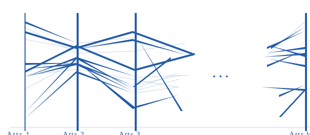
- Matrix of k² slices through the k-dim. data (the slices are determined interactively)

used by permission of J. J. van Wijk

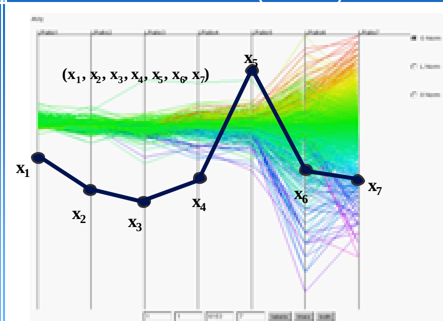


Geometric Transformations
Parallel Coordinates [Ins 85, ID 90]

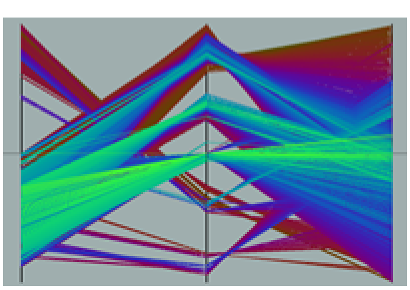
- n equidistant axes which are parallel to one of the screen axes and correspond to the attributest
- the axes are scaled to the [minimum, maximum] - range of the corresponding attribute
- every data item corresponds to a polygonal line which intersects each of the axes at the point which corresponds to the value for the attribute



Geometric Transformations
Parallel Coordinates (details)

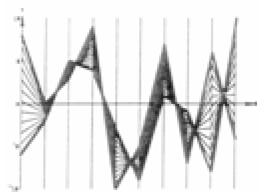


Geometric Transformations
Parallel Coordinates (cont'd)

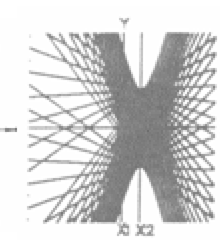


Geometric Transformations
Parallel Coordinates Patterns

used by permission of A. Inselberg, Tel Aviv University, Israel

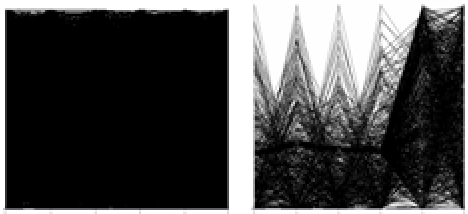


points on a line in 10D



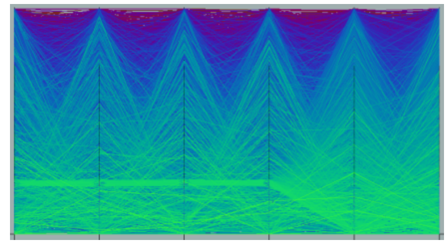
points on a circle in 2D space

Geometric Transformations
Parallel Coordinates Issues



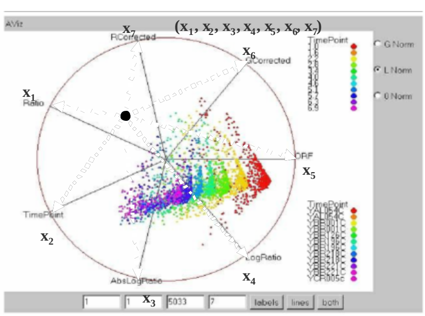
15,000 data items — 5 % of the data

Geometric Transformations
Parallel Coordinates Issues (cont'd)



15.000 data items with a query-dependent coloring

Geometric Transformations
RadViz

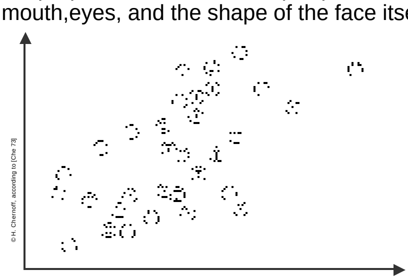


Iconic Displays

- Basic Idea
  Visualization of the data values as features of graphical icons (often called glyphs)
- Examples
  - Chernoff-Faces [Che 73, Tuf 83]
  - Stick Figures [Pic 70, PG 88]
  - Shape Coding [Bed 90]
  - Color Icons [Lev91, KK94]
  - TileBars [Hea 95] small icons representing the relevance feature vectors in document retrieval

Iconic Displays
Chernoff Faces [Che 73, Tuf 83]

- visualization of multidimensional data using the properties of a face icon (shape of nose, mouth,eyes, and the shape of the face itself)

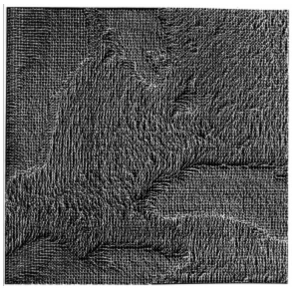


Iconic Displays
Stick Figures [Pic 70, PG 88]

- visualization of the data using the properties of a line icon
  - two attributes of the data are mapped to the display axes and the remaining attributes are mapped to the angle and/or length of the limbs
  - texture patterns in the visualization show certain data characteristics

Stick Figure Icon — A Family of Stick Figures

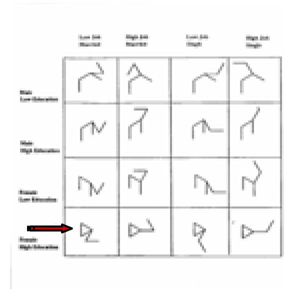
Iconic Displays
Stick Figures (cont'd)



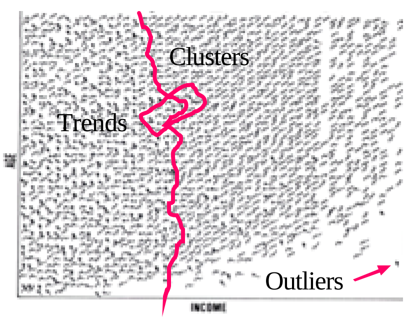
used by permission of G. Grinstein, University of Massachusetts at Lowell

5D image data from the great lakes region

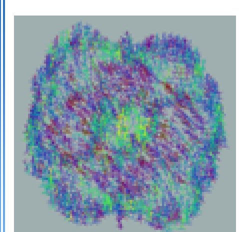
Iconic Displays

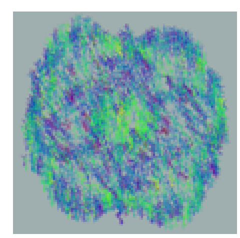


Iconic Displays
Census Data New England Engineers 1988



Iconic Displays
Stick Figures (cont'd)

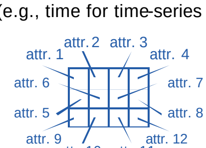




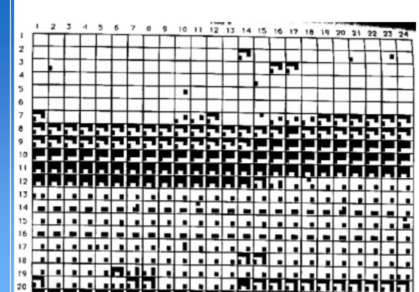
properties of the triangulation of molecule data

Iconic Displays
Shape Coding [Bed 90]

- the data are visualized using small arrays of fields
  - Each field represents one attribute value
  - Fields are arranged in some order based on attributes
  - Arrays are arranged line by line based on a given sorting (e.g., time for time-series data)



Iconic Displays
Shape Coding (cont'd)



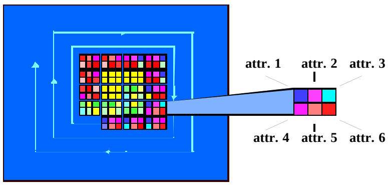
time series of NASA earth observation data

used by permission of J. Beddow

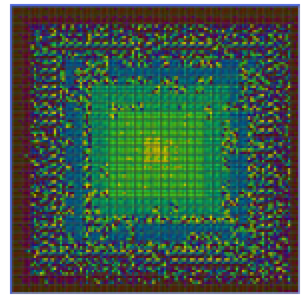
Iconic Displays
Color Icons [Lev 91, KK 94]

- visualization of the data using color icons
- color icons are arrays of color fields representing the attribute values
- arrangement is query-dependent (e.g., spiral)

schematic representation of 6D data

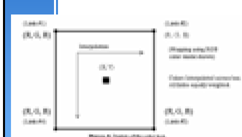


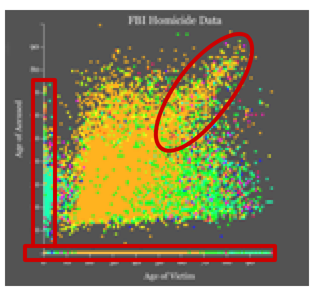
Iconic Displays
Colro Icons (cont'd)



random data containing several clusters

Iconic Displays
Color Icons

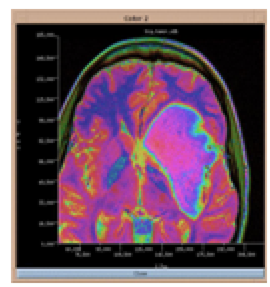




14 Dimensions

Iconic Displays
Colro Icons

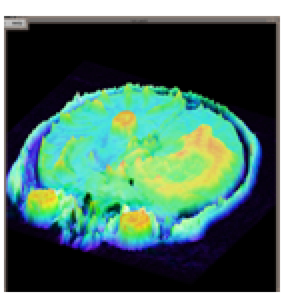
used by permission of G. Grinstein, University of Massachusetts at Lowell



Fused MRI Image data

Iconic Displays
3D Color Icon

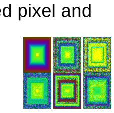
used by permission of G. Grinstein, University of Massachusetts at Lowell



Fused MRI Image data

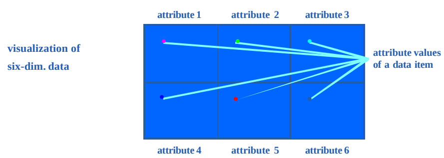
Dense Pixel Displays

- Task: data exploration and analysis
  - look at very large amounts of multidimensional data
- Principle: harness perceptual human capabilities
  - Produce adequate presentation of as much information as possible
- Goal: use each pixel of the display to visualize one data value
  - about 1.3 million data values may be displayed at one time
- Idea: map each data value to a colored pixel and arrange them adequately
  - pixel-oriented visualization techniques



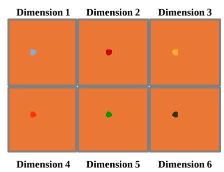
Dense Pixel Displays

- Basic Idea
  - each attribute value is represented by one colored pixel (the value ranges of the attributes are mapped to a fixed colormap)
  - the attribute values for each attribute are presented in separate subwindows



Four Questions

- How should the *pixels be arranged* within the subwindows?
- Are alternative *shapes* of the subwindows possible?
- How can an appropriate ordering of the dimensions be achieved?
- What can be done with *geometry-related data*?



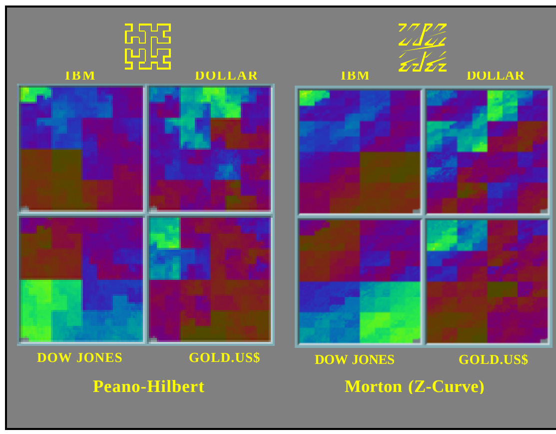
Arrangement of Pixels

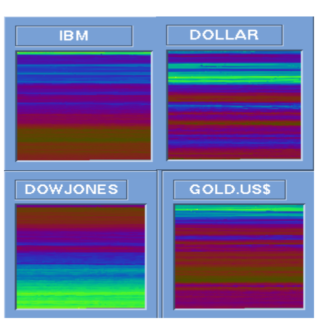
Given: An ordered Set of n data items {a1, ..., an} consisting of k data values each $\left(a_i^1, \cdots, a_i^k\right)$

Goal: Two-dim. arrangement of the data values, i.e. bijective mapping $f: \{1 \ldots n\} \rightarrow \{1 \ldots b\} \times \{1 \ldots h\}$ $(n \leq b * h)$, such that the function

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \left| d\left(f(i), f(j)\right) - d\left((0,0), \left(b \cdot \sqrt{\frac{|i-j|}{n}}, h \cdot \sqrt{\frac{|i-j|}{n}}\right)\right) \right|$$

is minimal, where $d(f(i), f(j))$ is the L^p-distance ($p=1,2$) of the pixels belonging to a_i and a_j

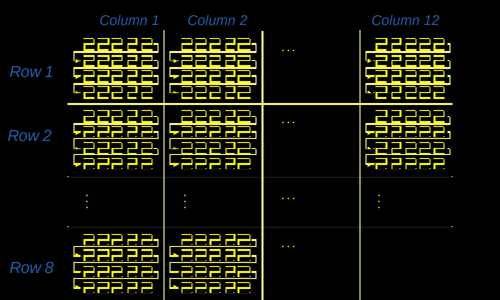




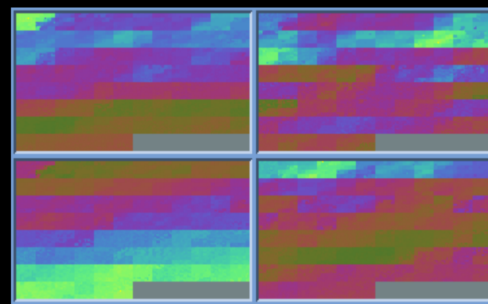
Recursive Pattern Technique

- Idea: r recursive generalization of line- and column-oriented arrangements
  - semantic arrangement by allowing user interaction to determine the height h_i and width w_i for each recusion level
- Algorithm for recursion level *i*
  Draw w_i pattern of recursion level *(i-1)* in *left-right* direction and repeat this h_i times in *top-down* direction

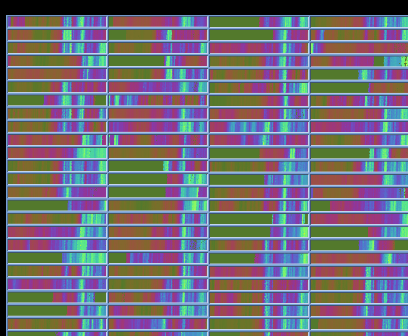
Example of a Structured Arrangement



$(w_1, h_1)=(3,3), (w_2, h_2)=(5,1), (w_3, h_3)=(1,4), (w_4, h_4)= (12, 1)$ and $(w_5, h_5)=(1,8)$



FAZ Stock Index (Jan. '74 - Apr. '95)



FAZ Aktien Index (Jan. '74 - Apr. '95)

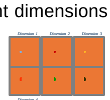
Shape of Subwindows

- Idea: Pixel, which belong to the k data values of the same data item, should have a small distance.

Goal: Shape of the Subwindows, such that

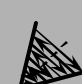
$$\frac{1}{n}\sum_{k=1}^{n}\left(\frac{1}{k}\sum_{i=1}^{k}\sum_{j=1}^{k} d\left(f(a_k^i), f(a_k^j)\right)\right) \text{ is minimal}$$

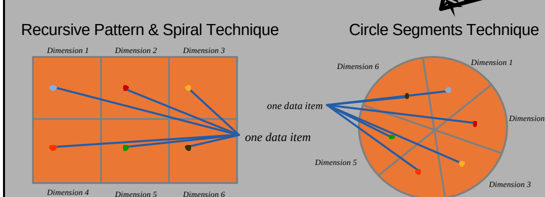
where $d\left(f(a_k^i), f(a_k^j)\right)$ is the L^p-distance of two pixels a_k^i and a_k^j belonging to two different dimensions



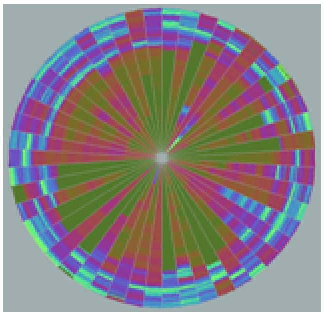
Shape of Subwindows

- Two Possibilities for the Shape of the Subwindows:
  - two-dim. array of rectangles
  - segmented circle



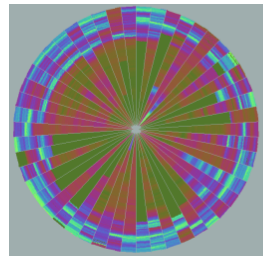


50 Stocks of the FAZ Stock Index

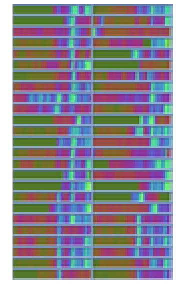


Georges Grinstein, UMass Lowell – Daniel Keim, Univ. of Konstanz – Matt Ward, WPI 187

Circle Segment and Recursive Pattern Technique Comparison



Circle Segments



Recursive Pattern

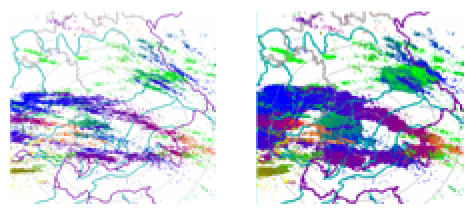
Georges Grinstein, UMass Lowell – Daniel Keim, Univ. of Konstanz – Matt Ward, WPI 188

Ordering the Dimensions

Geometry-related Visualizations

- Task: Visualizing Spatial Data
  - Problem: Data Overlap

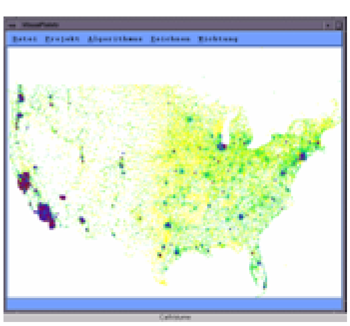
Example: Lightning Strikes in Southern Germany



Georges Grinstein, UMass Lowell – Daniel Keim, Univ. of Konstanz – Matt Ward, WPI 190

The **VisualPoints** System

Application of **Gridfit** Alg. to Telcom Data



Georges Grinstein, UMass Lowell – Daniel Keim, Univ. of Konstanz – Matt Ward, WPI 191

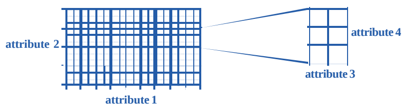
Stacked Displays

- **Basic Idea**: Visualization of the data using a hierarchical partitioning into subspaces.
- **Overview**
  - Dimensional Stacking [LWW 90]
  - Worlds-within-Worlds [FB 90a/b]
  - Treemap [Shn 92, Joh 93]
  - Cone Trees [RMC 91]
  - InfoCube [RG 93]

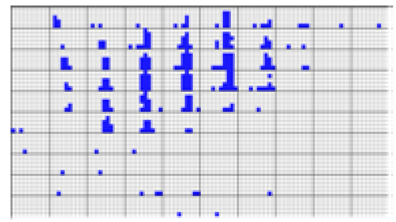
Georges Grinstein, UMass Lowell – Daniel Keim, Univ. of Konstanz – Matt Ward, WPI 192

Stacked Displays
Dimensional Stacking [LWW 90]

- Partitioning of the n-dimensional attribute space in 2-dimensional subspaces which are ‘stacked’ into each other
- Partitioning of the attribute value ranges into classes
- The important attributes should be used on the outer levels
- Adequate especially for data with ordinal attributes of low cardinality



Stacked Displays
Dimensional Stacking (cont‘d)



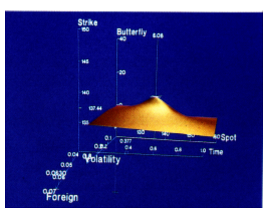
Visualization of oil mining data with longitude and latitude mapped to the outer x-,y-axes and ore grade and depth mapped to the inner x-,y-axes

Used by permission of M. Ward, Worcester Polytechnic Institute

Stacked Displays
Worlds within Worlds [FB 90a/b]

- Partitioning of the n-dim space into 3-dim subspaces
  - e.g. a six-dim object is displayed by having a new coordinate system for the last three dimensions sit inside the coordinate system for the first three

used by permission of C. Beshers, S. Feiner, Columbia University

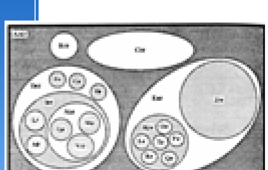


visualization of a six-dim. function

Stacked Displays
Treemaps [JS 91, Shn 92, Joh 93]

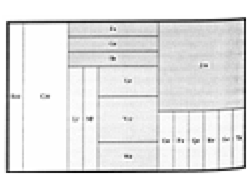
- Screen-filling method which uses a hierarchical partitioning of the screen into regions depending on the attribute values
- The x- and y-dimension of the screen are partitioned alternately according to the attribute values (the attribute value ranges have to be partitioned into classes)
- The attributes used for the partitioning and their ordering are user-defined (the most important attributes should be used first)
- The color of the regions may correspond to an additional attribute
- Suitable to get an overview over large amounts of hierarchical data (e.g., file system) and for data with multiple ordinal attributes (e.g., census data)

Stacked Displays
Treemaps



Used by permission of B. Shneiderman, University of Maryland

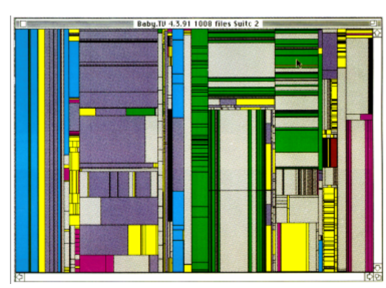
Venn Diagram



Used by permission of B. Shneiderman, University of Maryland

Tree-Map

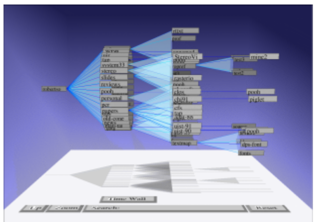
Stacked Displays
Treemaps (cont‘d)



Treemap of a file system containing about 1000 files

Used by permission of B. Shneiderman, University of Maryland

Stacked Displays
Cone Trees [RMS 91, CK 95]

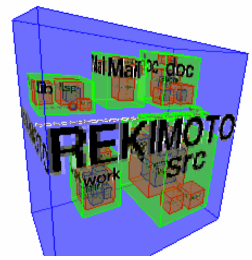


File system visualized as a cone tree

Used by permission of S. Card, Xerox Parc

Stacked Displays
InfoCube [RG 93]

Used by permission of J. Rekimoto, Sony CS Lab Inc.



3D visualization of hierarchical file system data using transport boxes

Hybrid Techniques

- **Basic Idea**: Integrated use of multiple techniques in one or multiple windows to enhance the expressiveness of the visualizations.
  - linking diverse visualization techniques may provide additional information
  - virtually all visualization techniques combine dynamics & interactivity
- **Examples:**
  - IVEE [AW 95a/b] uses *Starfield Displays* [AS 94] which are scatterplots of icons with dynamic zooming and mapping (combination of geometric, icon-based, and dynamic techniques)
  - XmDv [War 94] allows to dynamically link and brush scatterplot matrices, star icons, parallel coordinates, and dimensional stacking combination of geometric, icon-based, hierarchical and dynamic techniques)
  - Many others

Comparison of the Technique

- **Criteria for Comparison [KK 96]**
- **comparison of the described information visualization techniques based on their suitability for certain**
  - **data characteristics**
    (e.g., no. of variates, no. of data items, categorical data, ...)
  - **task characteristics**
    (e.g., clustering, multi variate hot spots, ...)
- **visualization characteristics**
  (e.g., visual overlap, learning curve, ...)

- **Disclaimer:** The following comparison table expresses my personal opinion obtained from reading the literature and experimenting with several of the described techniques. Many of the ratings are arguable and largely depend on the considered data, the exploration task, experience of the user, etc. In addition, implementations of the techniques in real systems usually avoid the drawbacks of a single technique by combining it with other techniques, which is also not reflected in the ratings.

Comparison of the Techniques

		cluster-ing	multi-variate hot spot	no. of variates	no. of data items	cate-gorical data	visual overlap	learning curve
Geometric Transformations	Scatterplot Matrices	++	++	+	+	-	o	++
	Landscapes	+	+	-	o	o	+	+
	Prosection Views	++	++	+	+	-	o	+
	Hyperslices	+	+	+	+	-	o	o
	Parallel Coordinates	o	++	++	-	o	–	o
Iconic Displays	Stick Figures	o	o	+	-	-	-	o
	Shape Coding	o	-	++	+	-	+	-
	Color Icon	o	-	++	+	-	+	-
Pixel Displays	Query-Independent	+	+	++	++	-	++	+
	Query-Dependent	+	+	++	++	-	++	-
Stacked Displays	Dimensional Stacking	+	+	o	o	++	o	o
	Worlds-within-Worlds	o	o	o	+	o	o	o
	Treemaps	+	o	+	o	++	+	o
	Cone Trees	+	+	o	+	o	+	+
	InfoCube	o	o	-	-	o	o	+

Outline

Distortion and Interaction Techniques

- Projection
- Filtering
- Zooming
- Linking and Brushing
- Distortion

Interactive Projections

- Dynamic or interactive variation of the projections
- Visualization of the remaining parameters in 2D or 3D
- Automatic variation of the data results in an animation
- Examples
  - GrandTour [Asi 85]
  - S Plus [BCW88]
  - XGobi [SCB92, BCS 96]
  - Influence & Attribute Explorer [STDS95, SDTS 95]
  - ...

Interactive Filtering

- Dynamic or interactive determination of database subsets
- Distinction between
  - *Selection*: direct selection of the desired subset
  - *Querying*: specification of the desired subset
- Specific problem
  - specification of complex Boolean conditions
- Examples
  - Magic Lenses [BIE 93]
  - Moveable Filter [FS 95]
  - Filter-Flow Model [YS 93]
  - InfoCrystal [Spo 93]
  - DeVise [Lic 97]
  - Dynamic Queries [As 94, Eic 94, GR 94]

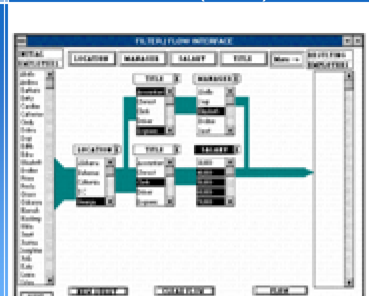
Interactive Filtering
Magic Lens/Moveable Filter [Bie 93, SFB 94, FS 95]

- interactive selection using lens-like tools which selectively filter the data in the considered areas
- multiple lenses / moveable filters can be used for a multi-level filtering (allowing complex conditions)

Interactive Filtering
Filter-Flow Model [YS 93]

- Selection based on a dataflow-oriented model: the data flows through filter-units which reduce the flow
- Especially useful for an intuitive specification of complex boolean queries:
  - AND-connected query portions may be specified using multiple filter units in a pipeline fashion
  - OR-connected query portions may be specified using multiple independent flows which reunite into a single bigger flow

Interactive Filtering
Filter-Flow Model (cont'd)

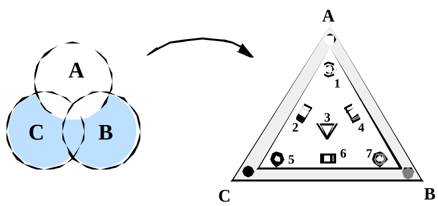


Complex boolean query:

Find accountants or engineers from Georgia managed by Elizabeth or clerks from Georgia who make more than $30,000!

Interactive Filtering
InfoCrystal [Spo 93]

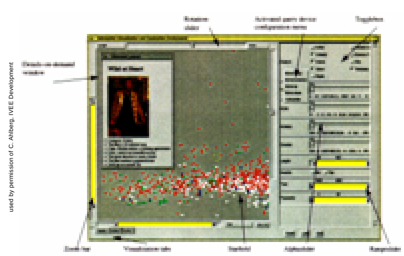
- Specification of complex boolean queries using an intuitive model for specifying complex subsets
  - basic idea



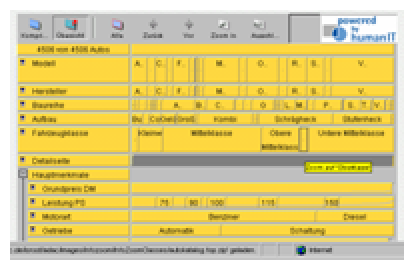
Interactive Zooming

- visualization of large amounts of data in reduced form to provide an overview of the data
- variable zooming of the data with automatic changes of the visualization modes to present more details
- examples:
  - PAD++ [PF 93, Bed 94, BH 94]
  - IVEE [AW 95a/b]
  - DataSpace [ADLP 95]
  - ...
- a comparison of fisheye and zooming techniques can be found in [Sch 93]

Interactive Zooming
IVEE/Spotfire [AW 95a/b]



Interactive Zooming
InfoZoom [Hum 01]

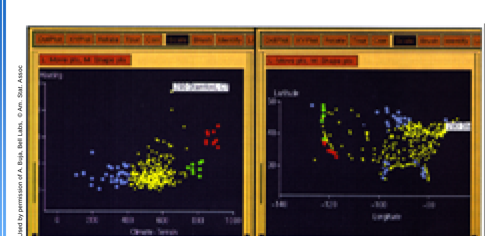


web-pages of the German Automobile Association (ADAC)

Interactive Linking and Brushing

- prerequisite: multiple visualizations of the same data (e.g., visualizations of different projections)
- interactive changes made in one visualization are automatically reflected in the other visualizations
- examples:
  - Xmdv-Tool [War 94]
  - S Plus [BCW 88]
  - XGobi [SCB 92, BCS 96]
  - DataDesk [Vel 92, WUT 95]
  - UVP
  - ...

Interactive Linking and Brushing
XGobi [SCB 92, BSC 96]



climate and housing data of the US

Distortion Techniques

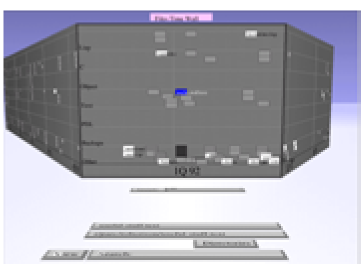
- **Basic Idea**: Distortion of the image to allow a visualization of larger amounts of data
- **Overview [LA 94]**
  - Perspective Wall MRC 91]
  - Bifocal Displays [SA 82]
  - TableLens [RC 94]
  - Graphical Fisheye Views [Fur 86, SB 94]
  - Hyperbolic Representations [LR 94, LRP 95]
  - 3D-Hyperbolic Representation [MB 95]
  - Hyperbox [AC 91]

Distortion Techniques
Perspective Wall [MRC 91]

- presentation of the data on a perspective wall
  - the data outside the focal area are perspectively reduced in size
  - the perspective wall is a variant of the bifocal lens display [SA 82] which horizontally compresses the sides of the workspace by direct scaling

Distortion Techniques
Perspective Wall (cont'd)

used by permission of S. Card, Xerox Parc

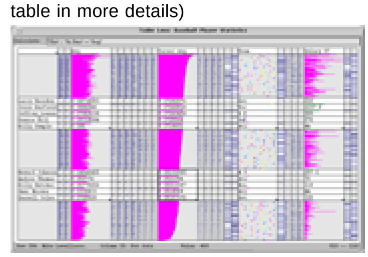


documents arranged on a perspective wall

Distortion Techniques
Table Kens [RC 94]

- Compact visualization of a table (spreadsheet/ database) with local zooming (viewing portions of the table in more details)

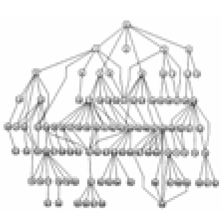
used by permission of R. Rao, Xerox PARC



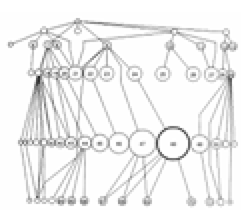
Visualization of the baseball database with a few rows selected for full detail

Distortion Techniques
Fisheye View [Fur 86, SB 94]

- Shows an area of interest quite large and with detail and other areas successively smaller with less detail
- graph visualization using a fisheye perspective



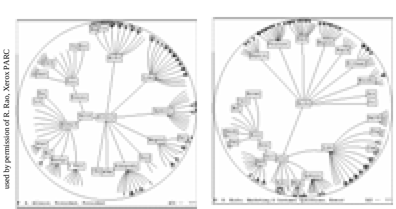
original graph



fisheye view of the graph

Distortion Techniques
Hyperbolic Trees [LR 94, LRP 95]

used by permission of R. Rao, Xerox PARC

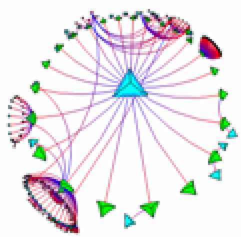


visualization of a large organizational hierarchy

visualization of tree structure in hyperbolic space with different foci

Distortion Techniques
Hyperbolic Trees [LR 94, LRP 95]

used by permission of T. Munzner, Stanford University

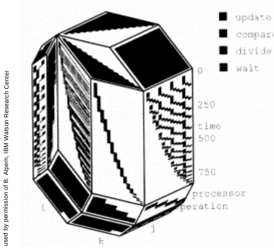


visualization of a large number of connected web-pages

visualization of a graph in 3D hyperbolic conetree-like representation

Distortion Techniques
Hyperbox [AC 91]

used by permission of B. Alpern, IBM Watson Research Center



Parallel processing performance data visualized as a hyperbox

mapping of scatterplots onto a hyperbox

Outline

Visual Data Mining Systems

- Overview
  - Statistics-oriented Systems
  - Visualization-oriented Systems
  - Database-oriented Systems
  - Special Purpose Systems

Visual Data Mining Systems

- **Statistics-oriented Systems**
  - S Plus [BCW 88] / Trellis [BCS 96]
    - generic system for statistical analysis and visualization
  - XGobi [XGobi, SCB 92, BSC 96]
    - extensible lisp-based system for statistical analysis and visualization
  - Data Desk [Vel 92, WUT 95]
    - commercial system for statistical analysis and visualization
    - features dynamic linking & brushing of scatterplots and histograms
  - Diamond (SPSS)
    - commercial system for statistical analysis and visualization
    - features dynamic linking & brushing of scatterplots, parallel coordinates, etc.)
  - DataSpace [ADLP 95]
    - 3D-arrangement of a large number of arbitrary visualizations

Visual Data Mining Systems
Visualization Oriented Systems

- ExVis [GPW 89]
  - stick figure and other icon-based techniques)
- Parallel Visual Explorer (IBM)
  - parallel coordinates with query-based coloring, and more
- XmDv [War 94, MW 95]
  - scatterplot matrices, star icons, parallel coordinates, dimensional stacking, dynamic linking and brushing)
- Influence & Attribute Explorer [STDS 95, SDTS 95]
  - scatterplot and prosection matrices, histograms, dynamic linking and brushing
- Information Visualizer (Xerox) [HC 86, CRY 96]
  - diverse information visualization techniques including perspective wall, table lens, cone trees)

Visual Data Mining Systems
Database Oriented Systems

- Hy+ [CM 93]
  - query and visualizations of hygraphs)
- TreeViz [Joh 93]
  - treemap technique)
- VisDB [KK 94, KK 95]
  - system for interactive slider-based exploration of very large databases
  - stick figure, parallel coordinate, and pixel-oriented techniques)
- IVEE [AW 95a/b] / Spotfire
  - commercial system for database exploration
  - generic interactive slider-based visualization environment
- DEVise [Liv 97]
  - system for the generation of interactive special purpose database visualizations

Visual Data Mining Systems
Special Purpose Visualization Systems

- **Software & Algorithm Visualization**
  - SeeSoft [ESS 92, BE 96] - a listing of Software & Algorithm Interfaces can be found under http://wwwbroy.informatik.tu-muenchen.de/~trilk/sv.html
  - for an overview paper see [SP 92])
- **Web Visualization**
  - Narcissus [HDWB 95], WebBook and WebForager [CRY 96] - a listing of Web Visualization Interfaces can be found under "http://www.geog.ucl.ac.uk/casa/martin/geography_of_cyberspace.html"
- **Visual Information Retrieval**
  - Vibe [Ols 93] - a bibliography of Information Retrieval Interfaces can be found under "http://www.pitt.edu/~korfhage/viri_bib.htm";
  - for an overview paper see [Car 96])

Visual Data Mining Systems

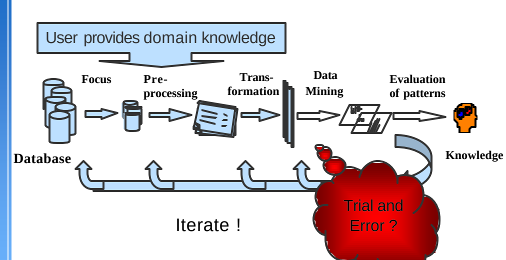
Examples/Demos

Outline

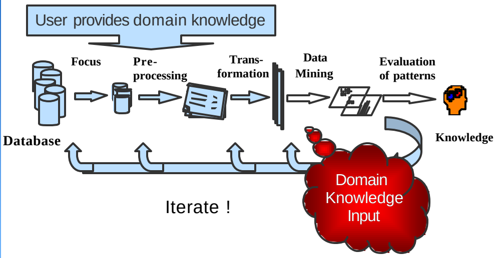
Specific Visual Data Mining Techniques

- Association Rules
- Classification
- Clustering
- Text Mining
- Tightly Integrated Visualization

The Human Role in The KDD Process



The Human's Role in the KDD Process
Human-Centered



Data Mining ? Visualization
Visual Data Mining

	Data Mining Algorithms	Visualization
Actionable	+	-
Evaluation	+	-
Flexibility	-	+
User Interaction	-	+

Human Involvement

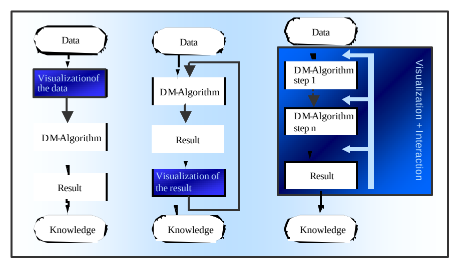
- When ?

- Right before the data mining step
  - Display initial data
  - Focus on/ narrow relevant search space

- During the data mining step
  - Display intermediate results
  - Direct the search

- After the data mining step
  - Display the result

Overview



Preceding Visualization (PV) **Subsequent Visualization (SV)** **Tightly integrated Visualization (TIV)**

outline

Association Rules

Definitions

$I = \{i_1,\ldots,i_m\}$ — I *Items* ,
$t \subseteq I$ — t *Transactions,*
$D = \{t_1,\ldots,t_N\}$, $t_i \subseteq I$ — D Database,
$X, Y \subset I$

Support of X, s(X): $\frac{|\{t \in D : X \subseteq t\}|}{|D|}$

Confidence of X and Y, c(X,Y): $\frac{s(X \cup Y)}{s(X)}$

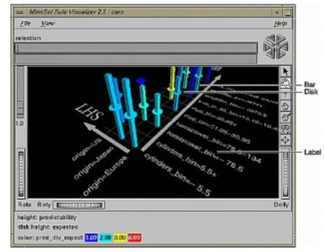
Association Rules Problem Description

Find all association rules X -> Y with

$s(X \cup Y) \geq smin$

and $c(X, Y) \geq cmin$

Subsequent Visualization: Association Rules
Rule Visualizer (MineSet) [Min 01]

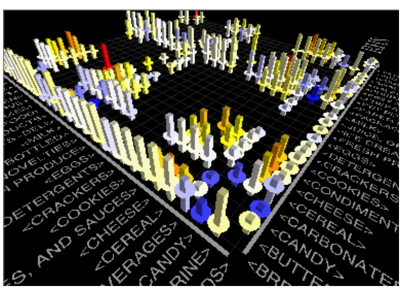


LHS and RHS items are mapped to x-, y-axis

Confidence, support correspond to height of the bar or disc, respectively

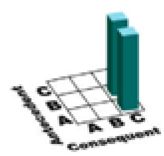
Interestingness is mapped to Color

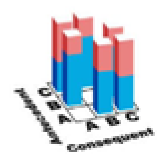
Subsequent Visualization: Association Rules
Rule Visualizer (MineSet) [Min 01]



Subsequent Visualization: Association Rules
Rule-to-item Visualization [WWT 99]

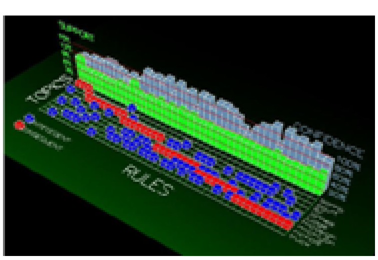
Limitations of item-to-item visualizations:





A->C, B->C or A+B -> C ? **Object occlusion ? metadata**

Subsequent Visualization: Association Rules
Rule-to-item Visualization [WWT 99]

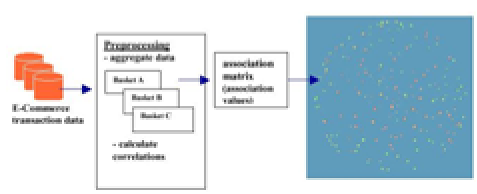


Items are mapped to one axis

Rules can be sorted according to a criterion

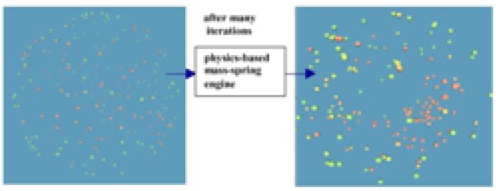
Subsequent Visualization: Association Rules
Market Basket Analysis Visualizer [HHD+ 01]

Initialization



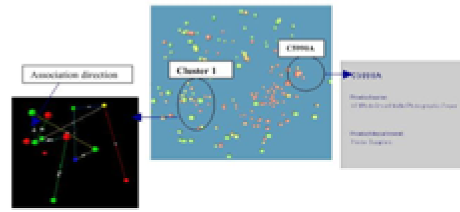
Subsequent Visualization: Association Rules
Market Basket Analysis Visualizer [HHD+ 01]

Relaxation

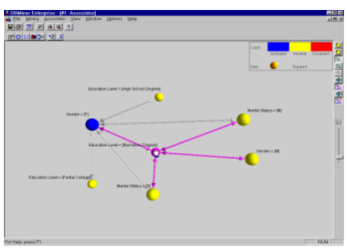


Subsequent Visualization: Association Rules
Market Basket Analysis Visualizer [HHD+ 01]

Navigation

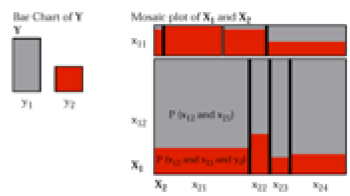


Subsequent Visualization: Association Rules
Association Ball Graph (DBMiner) [DBM 01]



Items are visualized as balls
Arrows indicate rule implication
Size represents support

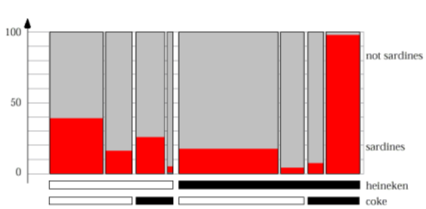
Subsequent Visualization: Association Rules
Interactive Mosaic Plots [HSW 00]



Visualization of contingency table of attributes within a rule

Recursive height/width splitting

Subsequent Visualization: Association Rules
Double Decker Plots [HSW 00]



Recursive width splitting

outline

Part I: Introduction
1. Goals of Information Visualization
2. Definition of Information Visualization
3. History of Information Visualization

Part II: Foundations
1. Visualization Pipeline
2. Data Foundations
3. Perceptual Foundations
4. Visualization Foundations / Theory

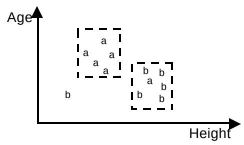
Part III: Visualization Techniques
1. Classification
2. Visual Data Exploration Techniques
3. Distortion and Interaction Techniques
4. Visual Data Mining Systems

Part VI: Specific Visual Data Mining Techniques
1. Association Rules
2. Classification
3. Clustering
4. Text Mining
5. Tightly Integrated Visualization

Classification Problem Description

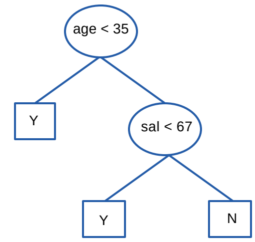
Given a set of objects with known class labels.

- Description
  - Build model describing the data with respect to the class
- Prediction
  - Use model to predict the class label of objects

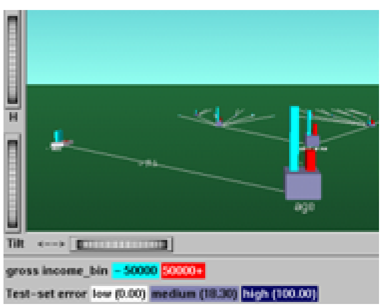


Classification – Decision Trees

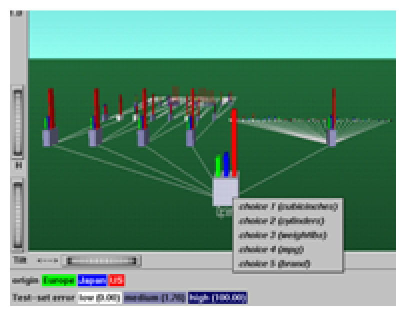
Age	Salary	Sex	**Class**
25	15	M	**Y**
42	40	M	**N**
29	63	F	**Y**
81	45	F	**N**
57	89	M	**Y**



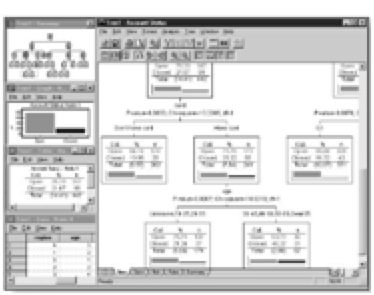
Subsequent Visualization: Classification Decision Tree Visualizer (MineSet) [MIN 01]



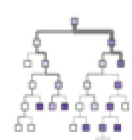
Subsequent Visualization: Classification Option Tree Visualizer (MineSet) [MIN 01]



Subsequent Visualization: Classification SPSS AnswerTree [SPS+ 01]



Subsequent Visualization: Classification SAS EM [SAS 01] Tree Viewer

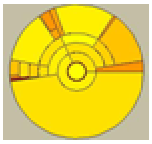


Color corresponds to relative frequency of a class in a node

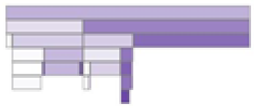
Branch line thickness is proportional to the square root of the objects

Subsequent Visualization: Classification
SAS EM [SAS 01] Tree Ring and Map

Tree Ring



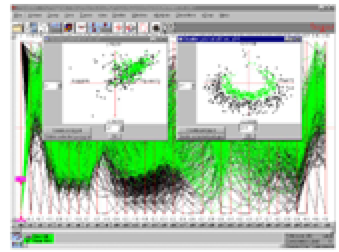
Tree Map



Color corresponds to relative frequency of a class in a node

Number of objects in a node are reflected proportionally

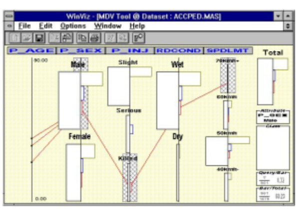
Subsequent Visualization: Classification
ParallAX [IA 00]



Select and order subset of predicting attributes

Visualize the result based on the parallel coordinates technique

Subsequent Visualization: Classification
WinViz [LO 96]



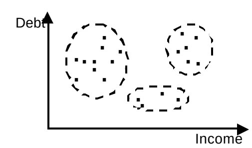
Left of attribute value: width of box indicates number of objects

Right of attribute value: class histograms

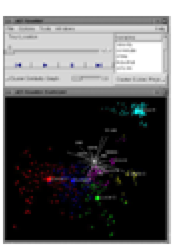
Cluster Analysis Problem Description

Given a set of objects.

- Group data into *clusters* so that objects within a cluster are very similar
- and objects not in the same cluster are dissimilar

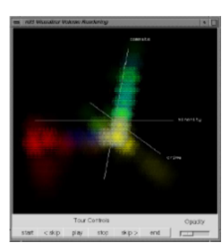


Subsequent Visualization: Cluster Analysis
3D Dynamic Projections [Yan 00]



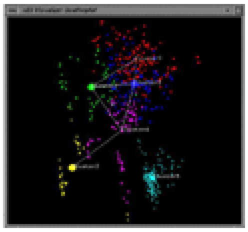
- 3D subspace is determined by centroids of 4 clusters 0, 1, 3, 5
- Projection preserves inter-cluster distances
- Projection-determining cluster centroids are visualized as big spheres
- Other cluster centroids are represented as small cubes

Subsequent Visualization: Cluster Analysis
3D Dynamic Projections [Yan 00]



- Volume rendering (by splatting) of multi-dimensional volume data to overcome clutter

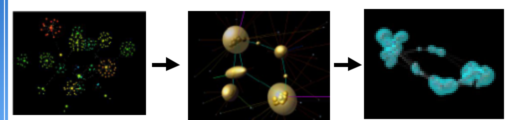
Subsequent Visualization: Cluster Analysis
3D Dynamic Projections [Yan 00]



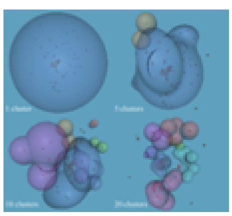
- Cluster similarity graph can be overlaid on to data projections
- User-defined threshold for distance between two cluster centroids

Subsequent Visualization: Cluster Analysis
H-BLOB (Hierarchical BLOB) [SBG 00]

Motivation

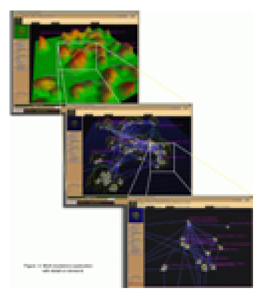


Subsequent Visualization: Cluster Analysis
H-BLOB (Hierarchical BLOB) [SBG 00]



- Cluster hierarchies are shown for 1, 5, 10 and 20 clusters

Subsequent Visualization: Cluster Analysis
VxInsight [VCI 02]

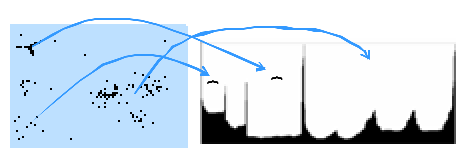


- Clusters are visualized as hills
- SQL query to database
- Multi-resolution exploration

Subsequent Visualization: Cluster Analysis
Optics [ABKS 99]

- OPTICS
  Ordering Points To Identify the Clustering Structure
- Insensitive to Parameters
- Augmented Cluster Ordering
- Reachability-distance: Basis for Interactive Cluster Analysis

Subsequent Visualization: Cluster Analysis
Optics [ABKS 99] - Reachability



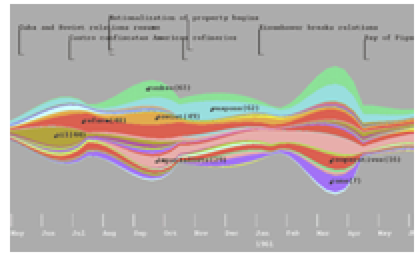
Represents the density-based clustering structure
Easy to analyze
Independent of the dimension of the data

Text Mining Problem Description

Given unstructured documents

- Cluster documents
- Summarize documents
- Classify documents

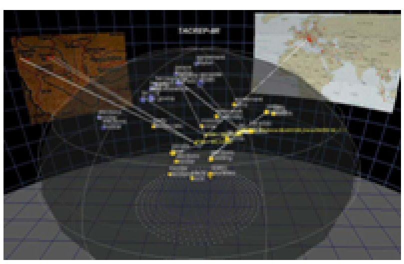
Subsequent Visualization: Text Mining Theme River [WCF+ 00]



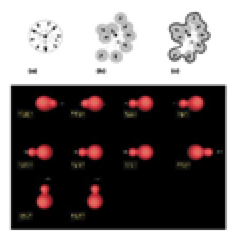
Visualization of thematic changes in documents

Vertical distance indicates collective strength of the themes

Subsequent Visualization: Text Mining Starlight [STA 01]

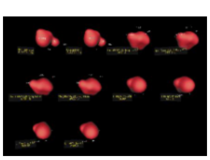


Subsequent Visualization: Text Retrieval Shape-base Visual Interface [RSE 99]



- Document terms are features
- Term vector proportional to term weight
- Term vectors are spread evenly about the sphere (for all documents)
- Spherical density source field is used to form a surface for each document

Subsequent Visualization: Text Retrieval Shape-based Visual Interface



Shape-based visualization of query result (*Macbeth* query by example)

Close-up of King Lear

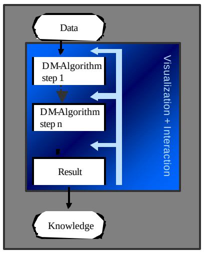
Subsequent Visualization

... and a lot more ... !

e.g. Visualization group at PNNL
http://www.pnl.gov/InfoViz/technologies.html

or SOM-based Visualization:
http://www.cis.hut.fi/~juuso/

Tightly Integrated Visualization (TIV)



- Visualization of algorithmic decisions
  - Data and patterns are better understood
  - User can make decisions based on perception
  - User can make decisions based on domain knowledge
  - Visualization of result enables user specified feedback for next algorithmic run or iteration

TIV: Visual Classification [AEK 00]

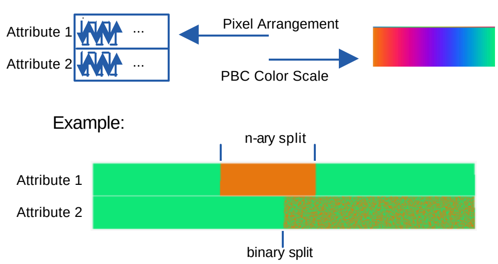
- Each attribute is sorted and visualized separately
- Each attribute value is mapped onto a unique pixel
- The color of a pixel is determined by the class label of the object
- The order is reflected by the arrangement of the pixels

attr.1	attr.2	...	class
0.3	23.3	...	Y
2.4	2.0	...	N
⋮	⋮	⋮	⋮

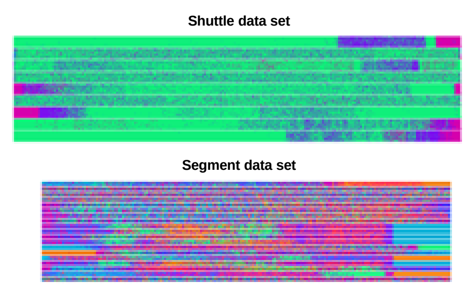
attr. 1	class
0.2	Y
0.3	Y
0.3	Y
0.5	N
1.1	Y
⋮	⋮

attr. 2	class
0.5	N
1.3	Y
2.0	N
2.5	Y
5.1	N
⋮	⋮

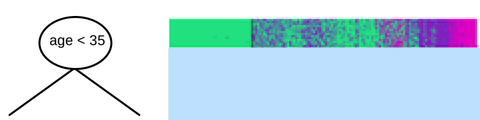
TIV: Visual Classification [AEK 00]



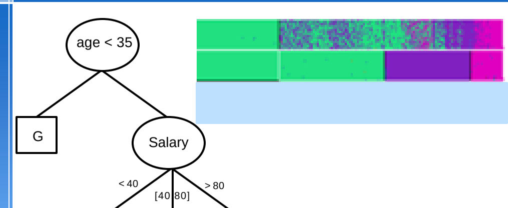
TIV: Visual Classification [AEK 00]



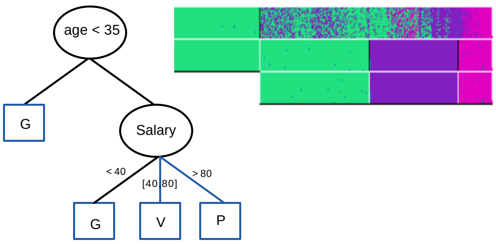
Visual Classification
Alternative Visualization of Decision Trees



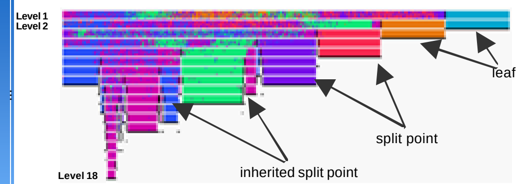
Visual Classification
Alternative Visualization of Decision Trees 2



Visual Classification
Alternative Visualization of Decision Trees 3



Visual Classification
Decision Tree for Segment Data



TIV: HD-Eye [HKW 99]
OptiGrid Algorithm

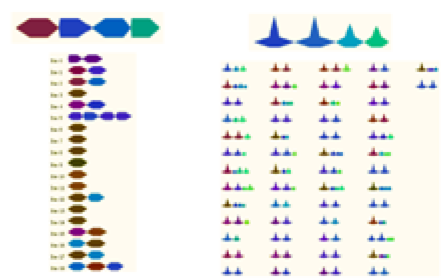
The OptiGrid Algorithm maps problem space of finding projections and separators onto visualization space

- Determine a set of contracting projections $\{P_0,\ldots,P_k\}$
- Determine the best q separators $\{H_0,\ldots H_q\}$ in the projections
- If there are no good separators exit;

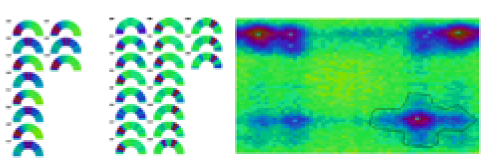
Otherwise

- Determine a multi-dimensional grid based on $\{H_0,\ldots, H_k\}$
- Find Clusters C_i in the grid by determining highly-populated grid cells
- For each C_i: OptiGrid(C_i)

TIV: HD-Eye [HKW 99]
Finding Contracting Projections

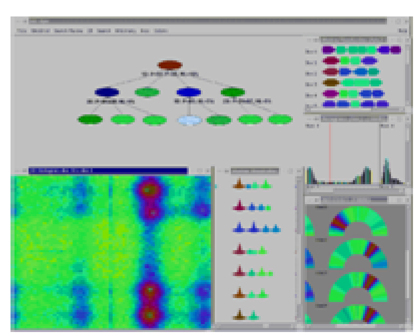


TIV: HD-Eye [HKW 99]
Finding Separators



in 1D projections **in 2D projections**

TIV: HD-Eye [HKW 99]



Summary
Visual Data Mining Architectures

Benefits of Visualization/ Interaction

	Preceding Visualization	Subsequent Visualization	Tightly int. Visualization
Present/ display patterns		●	●
Search problem space with perception	●		●
Incorporate domain knowledge	●		●
Provide trust and understandability of patterns		●	●

Summary and Conclusions

Integrated Analysis and Visualization Systems

- KXEN
- NeuroGenetic
- Optimizer
- AnswerTree
- Clementine
- Alice
- CART, MARS
- Cubist
- See5 (C5.0)
- R, S-Plus
- SAS
- Matlab
- CrossGraphs
- Intelligent Miner
- DecisionSite
- Partek
- SOMine
- OmniVis

Bio/ Cheminformatics Tools

- Examples
  - Sequence matching
  - Gene expression analysis
  - QSAR and 3D-QSAR modeling
  - ADME and toxicology prediction
  - Pattern recognition
  - Molecular similarity analysis
  - Diversity analysis
    - Population
    - Structure
  - …

Bioinformatics Case Study

Gene Expression Array Analysis

Array Instrumentation

- Array chips
  - Silicon-based, Oligos, 10,000+ genes
  - Custom or potted cDNA arrays, glass, 1,000-5,000+ genes
  - PE glass beads
  - Fiber optics
- Lots of other arrays

Simple Array Data

- Each assay provides expression of genes
- Each measures different conditions
- The array is a matrix of images which are converted to a matrix of numbers
- Informatics activities
  - Manipulate, convert, store, validate, compute, identify, search, mine, present, …

Computations

- Statistics that are meaningful are hard to get – need replicates
- Goal is to come up with hypotheses
  - Classes, clusters, relationships, outliers
- In many cases only a small number of genes change expression levels significantly
- Lots of noise
- Some missing values

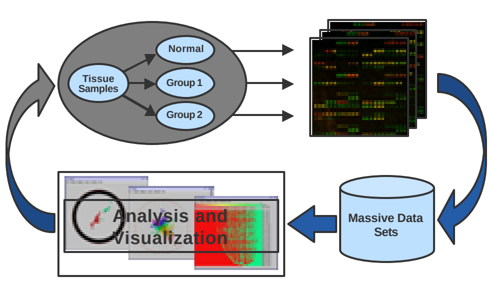
Define Expression Levels

- Threshold differences
- N-fold ratio change
- Statistics (f, T, non-parametric)
  - T assumes normal distributions
  - Large data set requires Bonferroni correction
  - Too small (p = .05, 10,000 genes, yields p/tests= .05/10,000 = .000005)
- Use novel techniques
- Use known-data

Compute

- Analyze and mine
  - Define similarity and other metrics
  - Use Euclidean or other cost functions
  - Reduce dimensionality (sigh)
- Use external information (databases, structures, domain experts)
- Predict or classify or …
  - Cluster
  - Identify associations
  - Delineate outliers

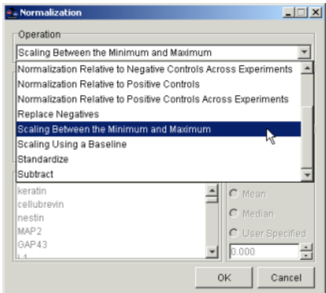
In Summary



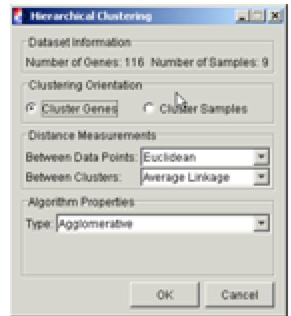
Molecular Mining

http://www.molecularmining.com/GeneLinker/Gold/

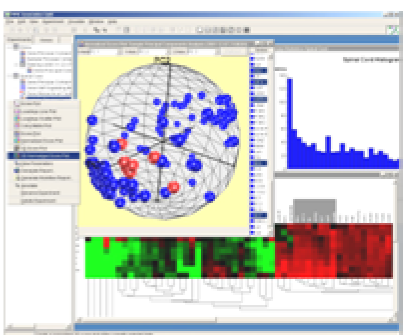
Normalization



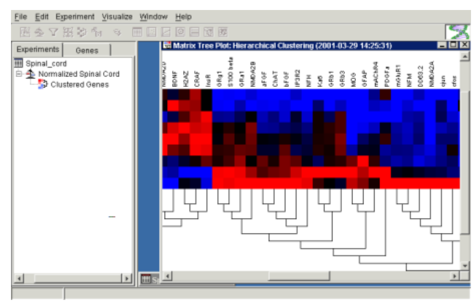
Clustering



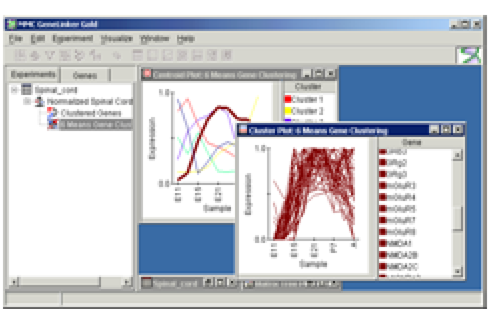
Gene-Linker



Matrix Tree Plot



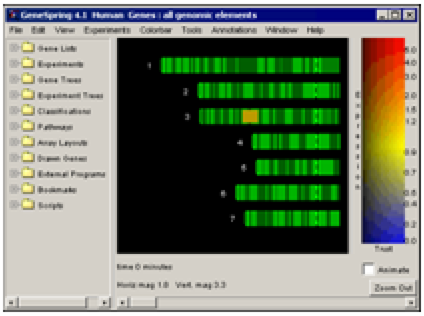
Centroid and Clustering Plots



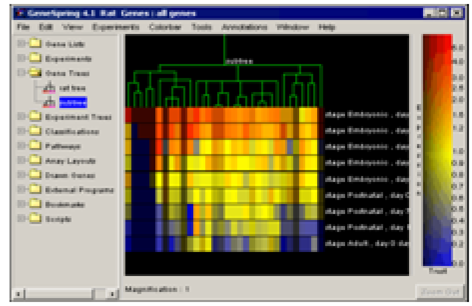
GeneSpring

http://www.sigenetics.com/

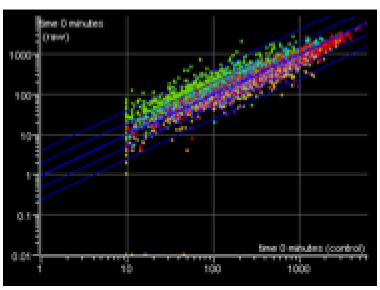
Physical Position



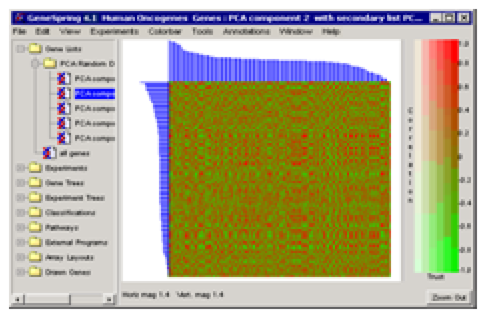
Hierarchical Clustering



Scatterplot



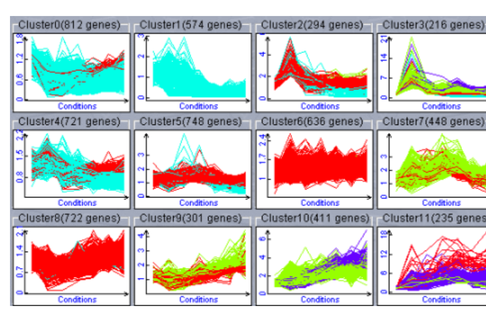
Gene Comparison View



GeneSight

http://www.biodiscovery.com/

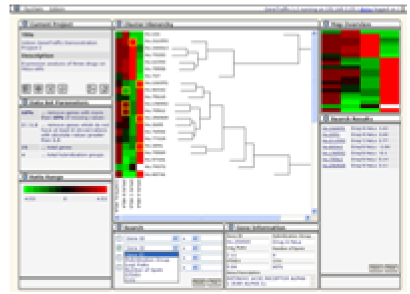
2D Self Organizing Maps



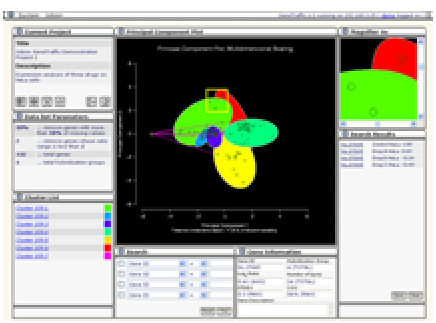
GeneTraffic

http://www.iobion.com/

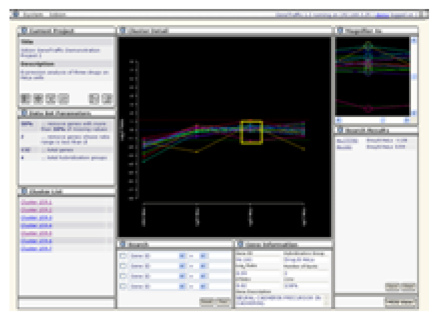
Hierarchical Clustering



Multidimensional Scaling View



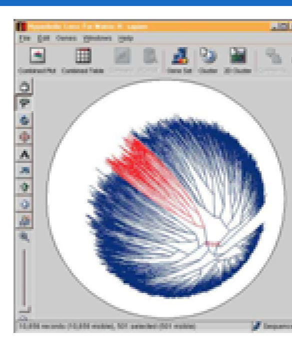
K-means Clustering View



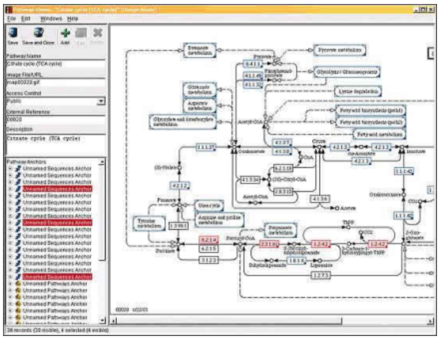
Rosetta BioSciences

http://www.rosettabio.com/products/resolver/

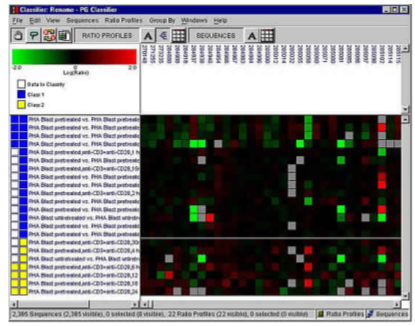
Hyperbolic Lens Viewer



Pathway Viewer



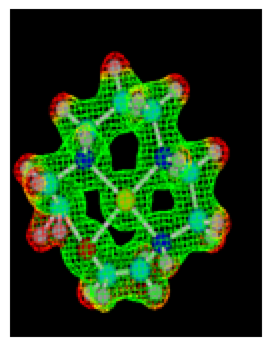
Class Prediction Algorithms Bayesian Classifiers



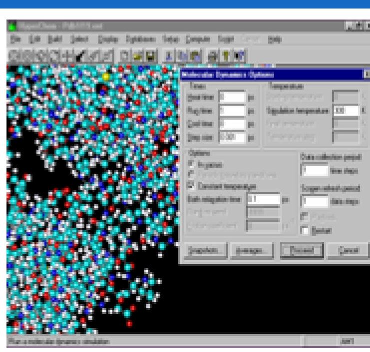
HyperChem

http://www.hplc1.com/hyperchem/hc5_features.html

Examples: Electrostatic Potential



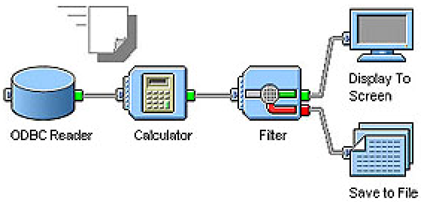
Molecular Dynamics



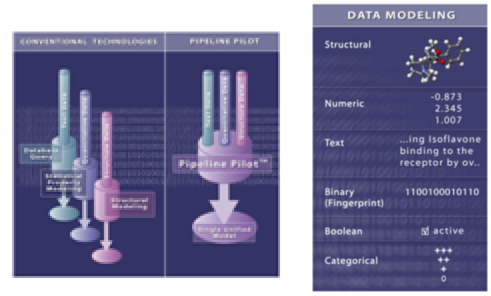
Scitegic

http://www.scitegic.com/

Data Pipelining



Data Modeling



generate a model of biological activity using the hits and failures from a high- throughput screening (HTS) run

Summary and Conclusions

Summary (1)

- Lots of commercial and academic products
- Key differentiating factors include
  - Usability
  - Number of tools
  - Flexibility
  - Robustness
  - Sensitivity
  - Interoperability

Summary (2)

- Some have newer and more novel visualization techniques which are applicable to database exploration
- Different techniques apply to different types of data (relational tables, hierarchies, graphs, etc.)
  - no guarantee of success
  - many of the techniques are applicable to traditional relational information sources
- Customization of tools is still necessary

Research Issues

- Develop integrated information visualization and exploration systems
  - with techniques from statistics, machine learning, databases, ...
- Perform in-depth evaluations and comparisons of visualization techniques for database exploration
  - there are possibilities for improvement
- Use more dynamics & interaction to steer the mining process
- Perform more case studies in a variety of application areas

Bibliography

[ABKS 99] Ankerst M., Breunig M., Kriegel H.P., Sander J.:"*OPTICS: Ordering Points To Identify the Clustering Structure*", Proc. ACM SIGMOD '99, Int. Conf on Management of Data, Philadelphia, PA, 1999, pp.49-60.

[AC 91] Alpern B., Carter L.: *'Hyperbox'*, Proc. Visualization '91, San Diego, CA, 1991, pp.133-139.

[ADLP 95] Anupam V., Dar S., Leibfried T., Petajan E.: *'DataSpace: 3-D Visualization of Large Databases'*, Proc. Int. Symp. on Information Visualization, Atlanta, GA, 1995, pp.82-88.

[AEK 00] Ankerst M., Ester M., Kriegel H.P.:"Towards an Effective Cooperation of the Computer and the User for Classification", Proc. ACM SIGKDD Int. Conf. On Knowledge Discovery & Data Mining (KDD 2000), Boston, MA, 2000, pp. 179-188.

[AKK 96] Ankerst M., Keim D.A., Kriegel H.P.: *'Circle Segments: A Technique for Visually Exploring Large Multidimensional Data Sets'*, Visualization '96, Hot Topic Session, San Francisco, CA, 1996.

[And 72] Andrews D. F.: *'Plots of High-Dimensional Data'*, in: Biometrics, Vol. 29, 1972, pp.125-136.

[And 95] Andrews K.: *'Visualizing Cyberspace: Information Visualization in the Harmony Internet Browser'*, Proc. Int. Symp. on Information Visualization, Atlanta, GA, 1995, pp.97-104.

[AS 94] Ahlberg C., Shneiderman B.: *'Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays'*, Proc. Human Factors in Computing Systems CHI '94 Conf., Boston, MA, 1994, pp. 313-317.

[AS 95] Ayers E. Z., Stasko J. T.: *'Using Graphic History in Browsing the World Wide Web'*, Proc. 4th Int. World Wide Web Conf., Boston, MA, 1995.

[Asi85] Asimov D.: *'The Grand Tour: A Tool For Viewing Multidimensional Data'*, SIAM Journal of Science & Stat. Comp., Vol. 6, 1985, pp.128-143.

[AW 95a] Ahlberg C., Wistrand E.: *'IVEE: An Environment for Automatic Creation of Dynamic Queries Applications'*, Proc. Human Factors in Computing Systems CHI '95 Conf., Demo Program, Denver, CO, 1995.

[AW 95b] Ahlberg C., Wistrand E.: *'IVEE: An Information Visualization and Exploration Environment*, Proc. Int. Symp. on Information Visualization, Atlanta, GA, 1995, pp.66-73.

Bibliography

[BCS 96] Becker R. A., Cleveland W. S., Shyu M.-J.: 'The Visual Design and Control of Trellis Display', Journal of Computational and Graphical Statistics, Vol. 5, No. 2, 1996, pp.123-155.

[BSC 96] Buja A., Swayne D. F., Cook D.: 'Interactive High-Dimensional Data Visualization', Journal of Computational and Graphical Statistics, Vol. 5, No.1, 1996, pp.78-99.

[BCW 88] Becker R., Chambers J. M., Wilks A. R.: 'The New S Language', Wadsworth & Brooks/Cole Advanced Books and Software, Pacific Grove, CA, 1988.

[BE 96] Ball T., Eick S.: 'Software Visualization in the Large', Computer, Vol. 29, No. 4, 1996, pp.33-43.

[Bed 90] Beddow J.: 'Shape Coding of Multidimensional Data on a Mircocomputer Display', Visualization '90, San Francisco, CA, 1990, pp.238-246.

[Bed 94] Bederson B.: 'Pad++: Advances in Multiscale Interfaces', Proc. Human Factors in Computing Systems CHI '94 Conf., Boston, MA, 1994, p. 315.

[BEMW 90] Becker R. A., Eick S. G., Miller E. O., Wilks A. R.: 'Dynamic Graphics for Network Visualization', Proc. Visualization '90, San Francisco, CA, 1990, pp.93-95.

[Ben 95] Benford S., Snowdon D., Greenhaigh C., Ingram R., Knox I., Brown C.: 'Vr-vibe: A Virtual Environment for Co-operative Information Retrieval', Proc. Eurographics'95, Maastricht, The Netherlands, 1995, pp. 349-360.

[Ber 81] Bertin J.: 'Graphics and Graphic Information Processing', Berlin, 1981.

[BETT 94] Battista G. D., Eades P., Tamassia R., Tollis I.: 'Annotated Bibliography on Graph Drawing Algorithms', Computational Geometry: Theory and Applications, Vol. 4, 1994, pp. 235-282.

[BEW 95] Becker R. A., Eick S. G., Wilks A. R.: 'Visualizing Network Data', Transactions on Visualization and Computer Graphics, Vol. 1, No. 1, 1995, pp.16-28.

[BF 93] Beshers C., Feiner S.: 'AutoVisual: Rule-Based Design of Interactive Multivariate Visualizations', IEEE Computer Graphics and Applications, Vol. 13, No. 4, 1993, pp.41-49.

Bibliography

[BFN 86] Boecker H.-D., Fischer G., Nieper H.: 'The Enhancement of Unterstanding through Visual Representations', Proc. Human Factors in Computing Systems CHI '86 Conf., Boston, MA, 1986, pp.44-58.

[BH 86] Bryce D., Hull R.: 'SNAP: A Graphics Based Schema Manager', Proc. 2nd Int. Conf. on Data Engineering, Los Angeles, CA, 1986, pp.151-164.

[BH 91] Brown M. H., Hershberger J.: 'Color and Sound in Algorithm Animation', Proc. IEEE Symp. on Visual Languages, 1991, pp.10-17.

[BH 94] Bederson B. B., Hollan J. D.: 'Pad++: A Zooming Graphical Interface for Exploring Alternate Interface Physics', Proc. UIST, 1994, pp.17-26.

[Bie 93] Bier E. A., Stone M. C., Pier K., Buxton W., DeRose T.: 'Toolglass and Magic Lenses: The see-through Interface', Proc. SIGGRAPH'93, Anaheim, CA, 1993, pp.73-80.

[BMMS 91] Buja A., McDonald J.A., Michalak J., Stuetzle W.: 'Interactive Data Visualization Using Focusing and Linking', Visualization '91, San Diego, CA, 1991, pp.156-163.

[BN 93] Brown M. H., Najork M. A.: 'Algorithm Animation Using 3D Interactive Graphics', Proc. UIST, 1993, pp.93-100.

[Bro 91] Brown M. H.: 'Zeus: A System for Algorithm Animation and Multi-View Editing', Proc. IEEE Symp. On Visual Languages, 1991, pp.4-9.

[Car 96] Card S. K.: 'Visualizing Retrieved Information: A Survey', IEEE CG&A, 1996, pp. 63-67.

[CE95] Cox K. C., Eick S. G.: '3D Displays of Internet Traffic', Proc. Symp. on Information Visualization, Atlanta, GA, 1995, pp.129-131.

[CELR 94] Cohen R. F., Eades P., Lin T., Ruskey F.: 'Three-dimensional Graph Drawing', Proc. Graph Drawing, 1994, in: Lecture Notes in Computer Science, Vol. 894, Springer, 1994, pp.1-11.

[Che 73] Chernoff H.: 'The Use of Faces to Represent Points in k-Dimensional Space Graphically', Journal Amer. Statistical Association, Vol. 68, pp.361-368.

Bibliography

[CK 95] Carrière J., Kazman R.: 'Interacting with Huge Hierarchies: Beyond Cone Trees', Proc. Symp. On Information Visualization, Atlanta, GA, 1995, pp.90-96.

[Cle 93] Cleveland W. S.: 'Visualizing Data', AT&T Bell Laboratories, Murray Hill, NJ, Hobart Press, Summit NJ, 1993.

[Cle 94] Cleveland W. S.: 'The Elements of Graphing Data', Hobart Press, Summit, NJ, revised edition.

[CM93] Consens M. P., Mendelzon A. O.: 'Hy+: A Hygraph-based Query and Visualization System', Proc. ACM SIGMOD Int. Conf. on Management of Data, Washington, DC, 1993, pp.511-516.

[Con 94] Consens M. P., Eigler F. Ch., Hasan M. Z., Mendelzon A. O., Noik E. G., Ryman A. G., Vista D.: 'Architecture and Applications of the Hy+ Visualization System', IBM Systems Journal, Vol. 33, No.3, 1994, pp.458-476.

[CPS 95] Chan M. C., Pacifici G., Stadler R.: 'Managing Real-Time Services in Multimedia Networks Using Dynamic Visualization and high-level Controls', Proc. Multimedia'95, San Francisco, CA, 1995.

[CR 92] Cox K. C., Roman G.-C.: 'Abstraction in Algorithm Animation', Proc. Symp. on Visual Languages, 1992, pp.18-23.

[CRMK 95] Chuah M. C., Roth S. F., Mattis J., Kolojejchick J.: 'SDM: Malleable Information Graphics', Proc. Symp. on Information Visualization'95, Atlanta, GA, 1995, pp.36-42.

[CRY 96] Card S. K., Robertson G. G., York W.: 'The WebBook and the WebForager: An Information Workspace for the World Wide Web', Proc. Human Factors in Computing Systems CHI '96 Conf., 1996.

[CT 94] Cruz I. F., Tamassia R.: 'How to Visualize a Graph: Specification and Algorithms', 1994.

[DBM 01] http://www.dbminer.com

[DE82] Dunn G., Everitt B.: 'An Introduction to Mathematical Taxonomy', Cambridge University Press, Cambridge, MA, 1982.

[DH 95] Drew N. S., Hendley R. J.: 'Visualizing Complex Interacting Systems', Proc. Human Factors in Computing Systems CHI'95 Conf., Denver, CO, 1995, pp.204-205.

Bibliography

[Doe 94] Doemel P.: 'Webmap - A Graphical Hypertext Navigation Tool', Proc. 2nd Int. World Wide Web Conf., Chicago, IL, 1994.

[Eic 94] Eick S. G.: 'Data Visualization Sliders', Proc. ACM UIST, 1994, pp.119-120.

[Eic 96] Eick S. G.: 'Aspects of Network Visualization', Proc. IEEE CG&A, 1996, pp.69-72.

[ES 92] Eick S. G., Steffen J. L.: 'Visualizing Code Profiling Line Oriented Statistics', Proc. Visualization'92, Boston, MA, 1992, pp.210-217.

[ESS 92] Eick S. G., Steffen J.L., Sumner E. E.: 'SeeSoft - A Tool for Visualizing Software', IEEE Transactions on Software Engineering, Vol. 18, No.11, 1992, pp. 957-968.

[EW 93] Eick S. G., Wills G. J.: 'Navigating Large Networks with Hierarchies', Visualization '93, San Jose, CA, 1993, pp.204-210.

[FB 90a] Feiner S., Beshers C.: 'Visualizing n-Dimensional Virtual Worlds with n-Vision', Computer Graphics, Vol. 24, No. 2, 1990, pp.37-38.

[FB 90b] Feiner S., Beshers C.: 'World within World: Metaphors for Exploring n-dimensional Virtual Worlds', Proc. UIST, 1990, pp.76-83.

[FB 94] Furnas G. W., Buja A.: 'Prosections Views: Dimensional Inference through Sections and Projections', Journal of Computational and Graphical Statistics, Vol. 3, No.4, 1994, pp.323-353.

[FB 95] Furnas G. W., Bederson B. B.: 'Space-Scale Diagrams: Unterstanding Multiscale Interfaces', Proc. Human Factors in Computing Systems CHI '95 Conf., Denver, CO, 1995.

[FL 95] Faloutsos C., Lin K.: 'Fastmap: A fast Algorithm for Indexing, Data-Mining and Visualization of Traditional and Multimedia Datasets', Proc. ACM SIGMOD Int. Conf. on Management of Data, San Jose, CA, 1995, pp.163-174.

[FPF 88] Fairchild K., Poltrok S., Furnas G.: 'SemNet: Three Dimensional Graphic Representations of Large Knowledge Bases', Lawrence Erlbaum (ed.), 1988, pp.201-233.

Bibliography

[FS 95] Fishkin K., Stone M. C.: 'Enhanced Dynamic Queries via Movable Filters', Proc. Human Factors in Computing Systems CHI '95 Conf., Denver, CO, 1995, pp.415-420.

[Fur 86] Furnas G.: 'Generalized Fisheye Views', Proc. Human Factors in Computing Systems CHI'86 Conf., Boston, MA, 1986, pp.18-23.

[FW 94] Frolich M., Werner M.: 'Demonstration of the Interactive Graph Visualization System da vinci', Proc. Graph Drawing, 1994, in: Lecture Notes in Computer Science, Vol. 894, Springer, 1994, pp.266-269.

[FZ 94] Furnas G. W., Zacks J.: 'Multitrees: Enriching and Reusing Hierarchical Structures', Proc. Human Factors in Computing Systems CHI '94 Conf., Boston, MA, 1994, pp.330-336.

[Ger 95] Gershon N., LeVasseur J., Winstead J., Croall J., Pernick A., Ruh W.: 'Visualizing Internet Resources', Proc. Symp on Information Visualization'95, Atlanta, GA, 1995, pp.122-128.

[Ger 96] Gershon N.: 'Moving Happily Through the World Wide Web', Proc. IEEE CG&A, 1996, pp.72-74.

[GKNV93] Gansner E. R., Koutsofois E., North S. C., Vo K.-P.: 'A Technique for Drawing Directed Graphs', IEEE Transactions on Software Engineering, Vol. 19, No. 3, 1993, pp.214-229.

[GPW 89] Grinstein G, Pickett R., Williams M. G.: 'EXVIS: An Exploratory Visualization Environment, Proc. Graphics Interface '89, London, Ontario, Canada, 1989.

[GR 94] Goldstein J., Roth S. F.: 'Using Aggregation and Dynamic Queries for Exploring Large Data Sets', Proc. Human Factors in Computing Systems CHI '94 Conf., Boston, MA, 1994, pp.23-29.

[GT 96] Garg A., Tamassia R.: 'Giotto3d: A System for Visualizing Hierarchical Structures in 3d', Proc. Graph Drawing, '96, in: Lecture Notes in Computer Science, Springer, 1996.

[Har 67] Harman H. H.: 'Modern Factor Analysis', University of Chicago Press, 1967.

Bibliography

[HDWB95] Hendley R. J., Drew N. S., Wood A. M., Beale R.: 'Narcissus: Visualizing Information', Proc. Int. Symp. on Information Visualization, Atlanta, GA, 1995, pp.90-94.

[HDD+ 01] Hao M., Hsu M., Dayal U., Wei S.F., Sprenger T., Holenstein T.:"Market Basket Analysis Visualization on a Spherical Surface", http://www.hpl.hp.com/techreports/2001/HPL-2001-3.pdf

[Hea 95] Hearst M.: 'TileBars: Visualization of Term Distribution Information in Full Text Information Access', Proc. of ACM Human Factors in Computing Systems Conf. (CHI'95), 1995, pp. 59-66.

[Him 94] Himsolt M.: 'Graphed: A Graphical Platform for the Implementation of Graph Algorithms', Proc. Graph Drawing '94, in: Lecture Notes in Computer Science, Vol. 894, Springer, pp.182-193.

[HKW 99] Hinneburg A., Keim D.A., Wawryniuk M.: „HD-Eye: Visual Mining of High-Dimensional Data", IEEE Computer Graphics and Applications, Vol. 19, No. 5, 1999.

[HSW 00] Hofmann H., Siebes A., Wilhelm A.:"Visualizing Association Rules with Interactive Mosaic Plots", Proc. ACM SIGKDD Int. Conf. On Knowledge Discovery & Data Mining (KDD 2000), Boston, MA, 2000.

[Hub 85] Huber P. J.: 'Projection Pursuit', The Annals of Statistics, Vol. 13, No.2, 1985, pp.435-474.

[Hum 01] HumanIT Web Page: „http://www.humanit.de/infozoom", Feb. 2001.

[IA 00] Inselberg A., Avidan T.:"Classification and Visualization for High-Dimensional Data", Proc. ACM SIGKDD Int. Conf. On Knowledge Discovery & Data Mining (KDD 2000), Boston, MA, 2000.

[Ins 85] Inselberg A.: 'The Plane with Parallel Coordinates, Special Issue on Computational Geometry', The Visual Computer, Vol. 1, 1985, pp.69-97.

[ID 90] Inselberg A., Dimsdale B.: 'Parallel Coordinates: A Tool for Visualizing Multi-Dimensional

Bibliography

[JS91] Johnson B., Shneiderman B.: 'Treemaps: A Space-filling Approach to the Visualization of Hierarchical Information', Proc. Visualization'91 Conf., 1991, pp.284-291.

[JS95] Jerding D. F., Stasko J. T.: 'The Information Mural: A Technique for Displaying and Navigating Large Information Spaces', Proc. Symp. on Information Visualization'95, Atlanta, GA, 1995, pp.43-50.

[Kei 94] D. A. Keim: 'Visual Support for Query Specification and Data Mining', Shaker-Verlag, 1994, ISBN 3-8265-0594-8.

[Kei 95] D. A. Keim: 'Enhancing the Visual Clustering of Query-dependent Database Visualization Techniques using Screen-FillingCurves', Proc. Workshop on Database Issues for Data Visualization, Atlanta, GA, 1995, Springer, in: Lecture Notes in Computer Science, No. 1183, Springer, pp. 101-110.

[Kei 95] D. A. Keim: 'Enhancing the Visual Clustering of Query-dependent Database Visualization Techniques using Screen-FillingCurves', Proc. Workshop on Database Issues for Data Visualization, Atlanta, GA, 1995, Springer, in: Lecture Notes in Computer Science, No. 1183, Springer, pp. 101-110.

[Kei 96] Keim D. A.: 'Pixel-oriented Visualization Techniques for Exploring Very Large Databases', Journal of Computational and Graphical Statistics, Vol. 5, No. 1, 1996, pp.58-77.

[KK 94] Keim D. A., Kriegel H.-P.: 'VisDB: Database Exploration using Multidimensional Visualization', Computer Graphics & Applications, Sept. 1994, pp.40-49.

[KK 95] Keim D. A., Kriegel H.-P.: 'VisDB: A System for Visualizing Large Databases', System Demonstration, Proc. ACM SIGMOD Int. Conf. on Management of Data, San Jose, CA, 1995, p.482.

[KKA 95] Keim D. A., Kriegel H.-P., Ankerst M.: 'Recursive Pattern: A Technique for Visualizing Very Large Amounts of Data', Proc. Visualization '95, Atlanta, GA, 1995, pp.279-286.

[KKS 93] D. A. Keim, H.-P. Kriegel, T. Seidl: 'Visual Feedback in Querying Large Databases', Proc. Visualization '93, San Jose, CA, 1993, pp.158-165.

Bibliography

[KK 96] Keim D. A., Kriegel H.-P.: 'Visualization Techniques for Mining Large Databases: A Comparison', Transactions on Knowledge and Data Engineering, Vol. 8, No. 6, Dec. 1996, pp. 923-938.

[KL 93] Keim D. A., Lum V.: 'GRADI: A Graphical Database Interface for a Multimedia DBMS', Proc. Int. Workshop on Interfaces to Databases, Glasgow, England, 1992, in: Workshops in Computing, Springer, 1993, pp.95-112.

[KRB94] Kaugars K., Reinfelds J., Brazma A.: 'A Simple Algorithm for Drawing Large Graphs on Small Screens', Proc. Graph Drawing '94, in: Lecture Notes in Computer Science, Vol. 894, Springer, 1994, pp.278-281.

[LA94] Leung Y., Apperley M.: 'A Review and Taxonomy of Distortion-oriented Presentation Techniques', Proc. Human Factors in Computing Systems CHI '94 Conf., Boston, MA, 1994, pp. 126-160.

[Lev 91] Levkowitz H.: 'Color icons: Merging color and texture perception for integrated visualization of multiple parameters', In Visualization '91, San Diego, CA, October 22-25 1991.

[Liv 97] M. Livny, R. Ramakrishnan, K. Beyer, G. Chen, D. Donjerkovic, S. Lawande, J. Myllymaki, and K. Wenger: 'DEVise: Integrated Querying and Visual Exploration of Large Data sets', Proc. of ACM SIGMOD, May, 1997.

[LO 96] Lee H.-Y., Ong K.-L.:"Visualization Support for Data Mining", IEEE Expert, Vol. 11, I. 5, pp.69-75, 1996.

[LR 94] Lamping J., Rao R.: 'Laying out and Visualizing Large Trees Using a Hyperbolic Space', Proc. UIST, 1994, pp.13-14.

[LRP95] Lamping J., Rao R., Pirolli P.: 'A Focus + Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies', Proc. Human Factors in Computing Systems CHI '95 Conf., Denver, CO, 1995, pp. 401-408.

[LSR 96] Lamm S. E., Scullin W. H., Reed D. A.: 'Real-time Geographic Visualization of World Wide Web Traffic', Proc. WWW, 1996

Bibliography

[MB 95] Munzner T., Burchard P.: 'Visualizing the Structure of the World Wide Web in 3D Hyperbolic Space', Proc. VRML '95 Symp, San Diego, CA, 1995, pp.33-38.

[MF 95] Mukherjea S., Foley J. D.: 'Showing the Context of Nodes in the World-Wide Web', Proc. ACM SIGCHI, Denver, CO, 1995.

[MF 95] Mukherjea S., Foley J. D.: 'Visualizing Complex Hypermedia Networks through Multiple Hierarchical Views', Proc. ACM SIGCHI, Denver, CO, 1995.

[MF 95] Mukherjea S., Foley J. D.: 'Visualizing the World-Wide Web with the Navigational View Builder', Proc. 3rd Int. World Wide Web Conf., Darmstadt, Germany, 1995.

[MGTS 90] Mihalisin T., Gawlinski E., Timlin J., Schwendler J.: 'Visualizing A Scalar Field on an N-dimensional Lattice', Visualization '90, San Francisco, CA, 1990, pp.255-262.

[MHCF 96] Munzner T., Hoffman E., Claffy K., Fenner B.: 'Visualizing the Global Topology of the mbone', Proc. Symp. on Information Visualization, 1996.

[Min 01] http://www.sgi.com/software/mineset/

[MMIK 95] Masui T., Minakuchi M., IV G. R., Kashiwagi K.: 'Multiple-View Approach for Smooth Information Retrieval', Proc. UIST, 1995, pp.199-206.

[MRC 91] Mackinlay J. D., Robertson G. G., Card S. K.: 'The Perspective Wall: Detail and Context Smoothly Integrated', Proc. Human Factors in Computing Systems CHI '91 Conf., New Orleans, LA, 1991, pp.173-179.

[MSG 94] McCreary C., Shieh F.-S., Gill H.: 'Cg: A Graph Drawing System Using Graph-Grammar Parsing', Proc. Graph Drawing '94, in: Lecture Notes in Computer Science, Vol. 894, Springer, 1994, pp.270-273.

[MW 95] Martin A. R., Ward M. O.: 'High Dimensional Brushing for Interactive Exploration of MultivariateData', Visualization '95, Atlanta, GA, 1995, pp.271-278.

Bibliography

[Mye 94] Myers B. A.: 'Visual Programming, Programming by Example and Program Visualization: A Taxonomy', Proc. Human Factors in Computing Systems CHI '94 Conf., Boston, MA, 1994, pp.59-66.

[NK94] North S. C., Koutsofois E.: 'Applications of Graph Visualization', Proc. Graphics Interface, 1994, pp. 235-245.

[Noi 93] Noik E. G.: 'Layout-independent Fisheye Views of Nested Graphs', Proc. Symp. on Visual Languages, 1993, pp.336-341.

[Noi 94] Noik E. G.: 'Encoding Presentation Emphasis Algorithms for Graphs', Proc. Graph Drawing '94, in: Lecture Notes in Computer Science, Vol. 894, Springer, 1994, pp.428-235.

[Noi 94] Noik E. G.: 'A Space of Presentation Emphasis Techniques for Visualizing Graphs', Proc. Graphics Interface, 1994, pp.225-233.

[Ols 93] Olsen K.A. et al.: 'Visualization of a Document Collection: The VIBE System', Information Processing and Management, Vol. 29, No.1, 1993, pp.69-81.

[OW 93] Ozoyoglu G., Wang H.: 'Example-Based Graphical Database Query Languages', Computer, Vol. 26, No.5, 1993, pp.25-38.

[PB 94] Pitkow J., Bharat K.: 'Webviz: A Tool for World-Wide Web Access Log Visualization', Proc. 1st Int. World Wide Web Conf., Geneva, Switzerland, 1994.

[PF 93] Perlin K., Fox D.: 'Pad: An Alternative Approach to the Computer Interface', Proc. SIGGRAPH, Anaheim, CA, 1993, pp.57-64.

[PG 88] Pickett R. M., Grinstein G. G.: 'Iconographic Displays for Visualizing Multidimensional Data', Proc. IEEE Conf. on Systems, Man and Cybernetics, IEEE Press, Piscataway, NJ, 1988, pp.514-519.

[Pic 70] Pickett R. M.: 'Visual Analyses of Texture in the Detection and Recognition of Objects', in: Picture Processing and Psycho-Pictorics, Lipkin B. S., Rosenfeld A. (eds.), Academic Press, New York, 1970.

Bibliography

[Rao 92] Rao R., Card S. K., Jellinek H. D., Mackinlay J. D., Robertson G. G.: 'The Information Grid: A Framework for Information Retrieval and Retrieval-centered Applications', Proc. UIST, 1992, pp.23-32.

[RC 94] Rao R., Card S. K.: 'The Table Lens: Merging Graphical and Symbolic Representation in an Interactive Focus+Context Visualization for Tabular Information', Proc. Human Factors in Computing Systems CHI '94 Conf., Boston, MA, 1994, pp. 318-322.

[RCM93] Robertson G. G., Card S. K., Mackinlay J. D.: 'Information Visualization Using 3D Interactive Animation', Communications of the ACM, Vol. 36, No. 4, 1993, pp.57-71.

[Reg 93] Regan B.: 'Two Algorithms for Drawing Trees in Three Dimensions', Proc. Graph Drawing, 1993.

[Rei 93] Reiss S. P.: 'A Framework for Abstract 3D Visualization', Proc. Symp. on Visual Languages, 1993, pp.108-115.

[Rei 94] Reiss S. P.: '3-D Visualization of Program Information', Proc. Graph Drawing, 1994, in: Lecture Notes in Computer Science, Vol. 894, Springer, 1994, pp.12-24.

[Ren 94] Rennison E.: 'Galaxy of News: An Approach to Visualizing and Understanding Expansive News Landscapes', Proc. UIST, 1994, pp.3-12.

[RG 93] Rekimoto J., Green M.: 'The Information Cube: Using Transparency in 3d Information Visualization', Proc. 3rd Annual Workshop on Information Technologies & Systems (WITS '93), 1993, pp.125-132.

[RKJ+ 99] Ribarsky W., Katz J., Jiang F., Holland A.: „Discovery Visualization using Fast Clustering", IEEE Computer Graphics and Applications, Vol. 19, No. 5, 1999.

[RM93] Robertson G. G., Mackinlay J. D.: 'The Document Lens', Proc. UIST, 1993, pp.101-108.

[RMC 91] Robertson G. G., Mackinlay J. D., Card S. K.: 'Cone Trees: Animated 3D Visualizations of Hierarchical Information', Proc. Human Factors in Computing Systems CHI '91 Conf., New Orleans, LA, 1991, pp. 189-194.

Bibliography

[SA 82] Spence R., Apperley M.: 'Data Base Navigation: An Office Environment for the Professional', Behaviour and Information Technology, Vol. 1, No.1, pp. 43-54.

[SAS 01] http://www.sas.com

[San 94] Sander G.: 'Graph Layout through the VCG Tool', Proc. Graph Drawing, 1994, in: Lecture Notes in Computer Science, Springer, Vol. 894, Springer, 1994, pp. 194-205.

[SB 94] Sarkar M., Brown M.: 'Graphical Fisheye Views', Communications of the ACM, Vol. 37, No. 12, 1994, pp.73-84.

[SCB92] Swayne D. F., Cook D., Buja A.: 'User's Manual for XGobi: A Dynamic Graphics Program for Data Analysis', Bellcore Technical Memorandum, 1992.

[Sch 93] Schaffer, Doug, Zuo, Zhengping, Bartram, Lyn, Dill, John, Dubs, Shelli, Greenberg, Saul, Roseman, Mark: Comparing Fisheye and Full-Zoom Techniques for Navigation of Hierarchically Clustered Networks', Proc. Graphics Interface (GI '93), Toronto, Ontario, 1993, in: Canadian Information Processing Soc., Toronto, Ontario, Graphics Press, Cheshire, CT, 1993, pp. 87-96.

[SDTS 95] Su H., Dawkes H., Tweedie L., Spence R.: 'An Interactive Visualization Tool for Tolerance Design', Technical Report, Imperial College, London, 1995.

[SFB 94] Stone M. C., Fishkin K., Bier E. A.: 'The Movable Filter as a User Interface Tool', Proc. Human Factors in Computing Systems CHI '94 Conf., Boston, MA, 1994, pp. 306-312.

[SGB 00] Sprenger T.C., Brunella R., Gross M.H.: "Hierarchical Visual Clustering Method using Implicit Surfaces", ETH Zurich, Tech report, http://graphics.ethz.ch, 2000

[Shn 92] Shneiderman B.: 'Tree Visualization with Treemaps: A 2D Space-Filling Approach', ACM Transactions on Graphics, Vol. 11, No.1, pp.92-99, 1992.

[SM 94] Sugiyama K., Misue K.: 'A Simple and Unified Method for Drawing Graphs: Magnetic-spring Algorithm', Proc. Graph Drawing, 1994, in: Lecture Notes in Computer Science, Vol. 894, Springer, 1994, pp.364-375.

Bibliography

[SP 92] Stasko, John T. and Patterson, Charles, "Understanding and Characterizing Software Visualization Systems", Proc. of the 1992 IEEE Workshop on Visual Languages, Seattle, WA, September 1992, pp. 3-10.

[Spo 93] Spoerri A.: 'InfoCrystal: A Visual Tool for Information Retrieval', Visualization '93, San Jose, CA, 1993, pp.150-157.

[SPS+ 01] http://www.spsscom

[STA 01] http://www.pnl.gov/nsd/commercial/starlight, 2001

[STDS95] Spence R., Tweedie L., Dawkes H., Su H.: 'Visualization for Functional Design', Proc. Int. Symp. on Information Visualization (InfoVis '95), Atlanta, GA, 1995, pp.4-10.

[SRN 72] Shepard R. N., Romney A. K., Nerlove S. B.: 'Multidimensional Scaling', Seminar Press, New York, 1972.

[SW 93] Stasko J. T., Wehrli J. F.: 'Three-dimensional Computation Visualization', Proc. Symp. on Visual Languages, Bergen, Norway, 1993, pp.100-107.

[TD 95] Tal A., Dobkin D.: 'Visualization of Geometric Algorithms', Transactions on Visualization and Computer Graphics, Vol. 1, No. 2, 1995, pp.194-204.

[Tuf83] Tufte E. R.: 'The Visual Display of Quantitative Information', Graphics Press, Cheshire, CT, 1983.

[Tuf90] Tufte E. R.: 'Envisioning Information', Graphics Press, Cheshire, CT, 1990.

[Vas 94] Vasudevan V.: 'Supporting High Bandwidth Navigation in Object-Bases', Proc. 10th Int. Conf. on Data Engineering, Houston, TX, 1994, pp.294-301.

[Vel 92] Velleman P. F: 'Data Desk 4.2: Data Description', Ithaca, NY, 1992.

[VXI 02] http://www.cs.sandia.gov/projects/VxInsight.html, 2001

[War 94] Ward M. O.: 'XmdvTool: Integrating Multiple Methods for Visualizing Multivariate Data', Visualization'94

Bibliography

[WCF+ 00] Wong P., Cowley W., Foote H., Jurrus E., Thomas J.: "Visualizing Sequential Patterns for Text Mining", InfoVis '00.

[Wis 95] Wise J. A., Thomas J. J., Pennock K., Lantrip D., Pottier M., Schur A., Crow V.: *'Visualizing the Non-Visual: Spatial Analysis and Interaction with Information from Text Documents',* Proc. Symp. On Information Visualization, Atlanta, GA, 1995, pp.51-58.

[WB95] Wong P.C., Bergeron R. D.: *'30 Years of Multidimensional Multivariate Visualization',* Proc. Workshop on Scientific Visualization, IEEE Computer Society Press, 1995.

[WDBH 95] Wood A., Drew N., Beale R., Hendley B.: *'Hyperspace: Web Browsing with Visualisation',* Proc. 3rd Int. World-Wide Web Conf., Poster, 1995, Darmstadt, Germany.

[WL 93] van Wijk J. J., van Liere R.. D.: *'Hyperslice',* Proc. Visualization '93, San Jose, CA, 1993, pp.119-125.

[Wri 95] Wright W.: *'Information Animation Applications in the Capital Markets'*, Proc. Int. Symp. on Information Visualization, Atlanta, GA, 1995, pp.19-25.

[WUT 95] Wilhelm A., Unwin A.R., Theus M.: *'Software for Interactive Statistical Graphics - A Review'*, Proc. Int. Softstat '95 Conf., Heidelberg, Germany, 1995.

[WWT 99] Wong P., Whitney P., Thomas J.: "*Visualizing Association Rules for Text Mining*", InfoVis '99.

[XGobi] XGobi Web-Page including pointers to publications and the most recent release of the XGobi software: "http://www.research.att.com/~andreas/xgobi/".

[Yan 00] Yang L.: "*Interactive Exploration of Very Large Relational Data Sets Through 3D Dynamic Projections*", Proc. ACM SIGKDD Int. Conf. On Knowledge Discovery & Data Mining (KDD 2000), Boston, MA, 2000.

[YS 93] Young D. and Shneiderman B.: 'A graphical filter/flow model for boolean queries: An implementation and experiment', Journal of the American Society for Information Science Vol. 44, No. 6, 1993, pp. 327-339.

[Zlo 77] Zloof M. M.: *'Query-By-Example: A Data Base Language',* IBM Systems Journal, Vol. 4, 1977, pp.324-343.