Tutorial Notes

An Introduction to
Information Visualization
Techniques
for Exploring Large Databases

Daniel A. Keim

Institute for Computer Science
University of Halle

A pdf version of the tutorial notes is available under
“http://www.informatik.uni-halle.de/~keim”
Overview

1. Introduction ..........................................................2
2. Data Preprocessing Techniques.................................9
3. Visual Data Exploration Techniques .....................11
   • Geometric Techniques ........................................11
   • Icon-based Techniques .....................................21
   • Pixel-Oriented Techniques ................................31
   • Hierarchical Techniques ....................................51
   • Graph-based Techniques ..................................60
   • Hybrid Techniques ..........................................72
4. Distortion Techniques ............................................73
5. Dynamic / Interaction Techniques ......................81
6. Comparison of the Techniques ..........................96
7. Database Exploration and Visualization Systems ...............98
8. Summary and Conclusion .............................103
Bibliography..................................................105

The screens dumps of systems other than the VisDB System are copyright of the respective authors and are taken from the cited publications.
Goals of Visualization Techniques

☐ Explorative Analysis
  • starting point: data without hypotheses about the data
  • process: interactive, usually undirected search for structures, trends, etc.
  • result: visualization of the data, which provides hypotheses about the data

☐ Confirmative Analysis
  • starting point: hypotheses about the data
  • process: goal-oriented examination of the hypotheses
  • result: visualization of the data, which allows the confirmation or rejection of the hypotheses

☐ Presentation
  • starting point: facts to be presented are fixed a priori
  • process: choice of an appropriate presentation technique
  • result: high-quality visualization of the data presenting the facts

Data Exploration

Definition

Data Exploration is the process of searching and analyzing databases to find implicit but potentially useful information.

more formally:

Data Exploration 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 $C$. 
Introduction

Comparison of the Abilities of Humans and Computers

- Data Storage
- Numerical Computation
- Searching
- Planning
- Diagnosis
- Prediction
- Perception
- Creativity
- General Knowledge

Introduction

Historical Overview of Exploratory Data Visualization Techniques (cf. [WB 95])

- pioneering work of Tufte [Tuf 83, Tuf 90] and Bertin [Ber 81] focuses on visualization of data with inherent 2D-/3D-semantics and general rules for layout, color composition, attribute mapping, etc.

- development of visualization techniques for different types of data with an underlying physical model, such as geographic data, CAD data, flow data, image data, voxel data, etc.

- development of visualization techniques for arbitrary multidimensional data (without an underlying physical model) and applicable to databases and other information resources.
**Classification of Data Visualization Techniques**

- **Geometric Techniques:** Scatterplots, Landscapes, Projection Pursuit, Projection Views, Hyperslice, *Parallel Coordinates*, ...
- **Icon-based Techniques:** Chernoff Faces, *Stick Figures*, Shape-Coding, Color Icons, TileBars, ...
- **Pixel-oriented Techniques:** *Recursive Pattern Technique, Circle Segments Technique, Spiral- & Axes-Techniques*, ...
- **Hierarchical Techniques:** Dimensional Stacking, Worlds-within-Worlds, *Treemap*, Cone Trees, InfoCube, ...
- **Graph-Based Techniques:** Basic Graphs (Straight-Line, Polyline, Curved-Line, ..) Specific Graphs (e.g., DAG, Symmetric, Cluster, ..) Systems (e.g., Tom Sawyer, Hy⁺, *SeeNet, Narcissus*, ..)
- **Hybrid Techniques:** arbitrary combinations from above
Introduction

Distortion and Dynamic / Interaction Techniques

- **Distortion Techniques**
  - Simple Distortion (e.g., *Perspective Wall*, Bifocal Lenses, *TableLens*, *Graphical Fisheye Views*, ...)
  - Complex Distortion (e.g., Hyperbolic Repr., *Hyperbox*, ...)

- **Dynamic / Interaction Techniques**
  - Data-to-Visualization Mapping (e.g., AutoVisual, S Plus, *XGobi*, *IVEE*, ...)
  - Projections (e.g., GrandTour, S Plus, *XGobi*, ...)
  - Filtering (Selection, Querying) (e.g., *MagicLens*, *Filter/Flow Queries*, *InfoCrystal*, ...)
  - Linking & Brushing (e.g., *Xmdv-Tool*, *XGobi*, DataDesk, ...)
  - Zooming (e.g., PAD++, *IVEE*, DataSpace, ...)
  - Detail on Demand (e.g., *IVEE*, TableLens, *MagicLens*, *VisDB*, ...)

Data Preprocessing Techniques

- **Techniques for Dimension Reduction**
  (Set of d-dim Data Items -> Set of k-dim. Data Items; 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 of the data.
  - Factor Analysis [Har 67]
    Determines a set of unobservable common factors which explain the main variations of the data. The original dimensions are linear combinations of the common factors.
  - Multidimensional Scaling [SRN 72]
    Uses the similarity (or dissimilarity) matrix of the data as defining coordinate axes in multidimensional space. The Euclidean distance in that space is a measure of the similarity of the data items.
  - Fastmap [FL 95]
    Fastmap also operates on a given similarity matrix and iteratively reduces 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 the 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.
  - Visualizations of Aggregations:
    - Histograms
    - Pie Charts, Bar Charts, Line Graphs, etc.

Geometric Techniques

**Basic Idea:** Visualization of geometric transformations and projections of the data.

**Overview**

- Scatterplot-Matrices [And72, Cle93]
- Landscapes [Wis95]
- Projection Pursuit Techniques [Hub85]
  (techniques for finding meaningful projections of multidimensional data)
- Prosection Views [FB94, STDS95]
- Hyperslice [WL93]
- Parallel Coordinates [Ins85, ID90]
Geometric Techniques

Scatterplot-Matrices [Cle 93]

Matrix of scatterplots (x-y-diagrams) of the k-dim. data [total of \((k^2/2 - k)\) scatterplots]

Geometric Techniques

Landscapes [Wis 95]

Visualization of the data as perspective landscape

The data needs to be transformed into a (possibly artificial) 2D spatial representation which preserves the characteristics of the data
Geometric Techniques

Prosection Views [FB 94, STDS 95]

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

Hyperslice [WL 93]

matrix of $k^2$ slices through the k-dim. data (the slices are determined interactively)
Parallel Coordinates [Ins 85, ID 90]

- $n$ equidistant axes which are parallel to one of the screen axes and correspond to the attributes
- the axes are scaled to the $[\text{minimum}, \text{maximum}]$ 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

![Diagram of Parallel Coordinates]

Parallel Coordinates (cont’d)
Parallel Coordinates (cont’d)

- Points on a line in 10-dimensional space
- Points on a circle in 2-dimensional space

15,000 data items with noise
5% of the data (750 data items)

Daniel A. Keim
Page 18
Information Visualization Techniques
Geometric Techniques

Parallel Coordinates (cont’d)

15,000 data items with a query-dependent coloring

Icon-based Techniques

Basic Idea: Visualization of the data values as features of icons.

Overview

- Chernoff-Faces [Che 73, Tuf 83]
- Stick Figures [Pic 70, PG 88]
- Shape Coding [Bed 90]
- Color Icons [Lev 91, KK 94]
- TileBars [Hea 95]
  (use of small icons representing the relevance feature vectors in document retrieval)
Icon-based Techniques

Chernoff-Faces [Che 73, Tuf 83]

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

Icon-based Techniques

Stick Figures [Pic 70, PG 88]

- visualization of the multidimensional data using stick figure icons
- 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
Icon-based Techniques

Stick Figures (cont’d)

5-dim. image data from the great lake region

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

Icon-based Techniques

Stick Figures (cont’d)

census data showing age, income, sex, education, etc.

used by permission of G. Grimm, University of Massachusetts at Lowell
Icon-based Techniques

Stick Figures (cont’d)

properties of the triangulation of molecule data

Icon-based Techniques

Shape Coding [Bed 90]

- ↳ the data are visualized using small arrays of fields
- ↳ each field represents one attribute value
- ↳ arrangement of attribute fields (e.g., 12-dimensional data):

  ![Array Diagram]

- ↳ arrays are arranged line-by-line according to a given sorting (e.g., the time attribute for time-series data)
Icon-based Techniques

Shape Coding (cont’d)

time series of
NASA earth
observation data

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)
Icon-based Techniques

Color Icons (cont’d)

random data containing several clusters

Pixel-oriented Techniques

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
- example:
Pixel-oriented Techniques

Overview [Kei 96]

Simple Techniques

Query-Independent Techniques

Space-Filling Curves

Recursive Pattern Technique

Query-Dependent Techniques

Spiral Technique

Axes Technique

Circle Segments

Pixel-oriented Techniques

Query-Independent Techniques:
Space-Filling Curve Arrangements

Peano-Hilbert

Morton (Z-Curve)
Pixel-oriented Techniques

Space-Filling Curve Arrangements

IBM    DOLLAR
   DOW JONES     GOLD.US$
Peano-Hilbert

IBM    DOLLAR
   DOW JONES     GOLD.US$
Morton (Z-Curve)

Pixel-oriented Techniques

Query-Independent Techniques:
Recursive Pattern Technique [KKA 95]

• recursive generalization of iterated line- and column-based arrangements

• the user may specify the height $h_i$ and width $w_i$ for each recursion level

• on recursion level $i$, $w_i$ patterns of recursion level $(i-1)$ are drawn in left-right direction and this is repeated $h_i$ times in top-down direction

⇒ the pattern on recursion level $i$ consists of $w_i \times h_i$ patterns of recursion level $(i-1)$
Pixel-oriented Techniques

Recursive Pattern: Possible Arrangements

- line-by-line loop
- back-and-forth loop

Recursive Pattern: Example of a Structured Arrangement

(row 1) (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)
Pixel-oriented Techniques

Recursive Pattern: Example of Financial Data

IBM

DOLLAR

time series of financial data

DOW JONES

GOLD (US$)

Recursive Pattern: FAZ-Index (Jan. ‘74 - Apr. ‘95)

time series of the 100 stocks in the Frankfurt Stock Index
Pixel-oriented Techniques

Query-Dependent Techniques: Basic Idea

• data items \((a_1, a_2, \ldots, a_m)\) & query \((q_1, q_2, \ldots, q_m)\)

  \(\Rightarrow\) distances \((d_1, d_2, \ldots d_m)\)

• extend distances by overall distance \((d_{m+1})\)

• determine data items with lowest overall distances

• map distances to color (for each attribute)

• visualize each distance value \(d_i\) by one colored pixel

Pixel-oriented Techniques

Calculating the Visualizations

1. Read Data into Main Memory
2. Calculate Distances: Datatype \(\rightarrow\) Real
3. Normalize Distances: \([d_{min}, d_{max}] \rightarrow [0, 255]\)
4. Determine the Desired Percentage of Data Items with Lowest Combined Distances: \(\rightarrow \alpha\)-quantile
5. Normalize Value Ranges: \([d_{min}, d_{max}] \rightarrow [0, 255]\)
6. Sort the Data Items: \(\rightarrow\) bucket sort
7. Visualize Relevance Factors and Distances
Pixel-oriented Techniques

Query-Dependent Techniques:
Spiral Technique [KKS 93, KK 94]

arrangement in spiral form according to the overall distance

example of the overall distance

Spiral Technique (cont’d)

• the values for each of the attributes are presented in a separate subwindows
• the arrangement inside the subwindows is according to the overall distance
• example:

visualization of six-dim. data

attribute values of a data item

attribute 1 attribute 2 attribute 3
attribute 4 attribute 5 attribute 6
**Pixel-oriented Techniques**

**Spiral Technique (cont’d)**

Result of a complex query

**Pixel-oriented Techniques**

**Axes Technique [KK 94]**

Attribute j

- pos
- neg

Attribute i

- neg
- pos

Arrangement in partial spirals in each quadrant

Example of the overall distance
Pixel-oriented Techniques

Spiral and Axes Techniques [KK 94]

random data containing several clusters

Pixel-oriented Techniques

Spiral, Axes, and Color Icon Techniques [KK 94]

Spiral Technique

Axes Technique

Color Icon Technique
Pixel-oriented Techniques

Generalized Spiral Technique [Kei 95]

- Combination of Spiral Technique and Space-Filling Curves

Spiral Technique

Snake-Spiral Technique

Circle Segments Technique [AKK 96]

- Arrangement of Attributes on the Segments of a Circle

Arrangement of 8-dim. Data

Arrangement of 15-dim. Data
Pixel-oriented Techniques

Circle Segments Technique (cont’d)

![Circle Segments Technique](image)

time series of 50 stocks of the Frankfurt Stock Index

Hierarchical Techniques

**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]
Hierarchical Techniques

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

Hierarchical Techniques

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
Hierarchical Techniques

**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)

**Treemap [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)
Hierarchical Techniques

Treemap (cont’d)

Venn Diagram

Tree-Map

Treemap (cont’d)

A treemap of a file system containing about 1000 files.

used by permission of B. Shneiderman, University of Maryland
Hierarchical Techniques

Cone Trees [RMC 91, CK 95]

- animated 3D visualizations of hierarchical data

file system structure visualized as a cone tree

Hierarchical Techniques

InfoCube [RG 93]

- 3D visualization of hierarchical data using transparent boxes

visualization of a file system structure
Graph-based Techniques

**Basic Idea:** Visualization of large graphs using techniques to convey the meaning of the graph clearly and quickly.

**Overview [CT 94, BETT 94]**

- **Basic Graphs** (e.g., Straight-Line, Polylines, Curved-Line, Orthogonal Graphs, ...)
- **Specific Graphs** (e.g., Directed Acyclic, Cluster-Optimized, Symmetry-Optimized Graphs, Hygraphs, ...)
- **Systems** (e.g., Tom Sawyer, Hy⁺ [CM 93, Con 94], SeeNet [EW 93, BEW 95], Narcissus [HDWB 95], ...)

Graph-based Techniques

**2D-Graph Drawings**

- Polyline Drawing
- Straight-Line Drawing
- Curved-Line Drawing
Graph-based Techniques

Properties of 2D-Graph Drawings

- planarity (no line crossings)
- orthogonality (only orthogonal lines)
- grid property (coordinates of vertices are integers)

Aesthetics Properties (Optimization Goals)

- minimal number of line crossings
- optimal display of symmetries
- optimal display of clusters
- minimal number of bends in polyline graphs
- uniform distribution of vertices
- uniform edge lengths

Graph-based Techniques

2D-Graph Drawings (Examples)

Orthogonal Graph

Symmetry-Optimized Graph
Graph-based Techniques

2D-Graph Drawings (Examples)

Cluster-Optimized Graph

Directed Acyclic Graph

Graph-based Techniques

2D-Graph Drawings: Open Problems [BETT 94]

- Performance Bounds (e.g. for planarization, ...)
- Dynamic Algorithms
- Parallel Algorithms
- Complexity of Bend Minimization
- Angular Resolution Constraints
- Three-dimensional Graph Drawings
Graph-based Techniques

3D-Graph Drawings (cf. [CELR 94])

Ball-like Graph

Torrus-like Graph

Graph-based Techniques

3D-Graph Drawings (cont’d)

Cluster-Optimized 3D-Graph
Graph-based Techniques

Hygraphs (cont’d)

- multi-resolution visualization of hygraphs allowing
  an interactive manipulation using Graphlog

SeeNet [EW 93, BEW 95]

- visualization of hierarchical networks with weighted links

- special features of SeeNet:
  - semantic node placement
    (minimizing the distance of nodes with high-weighted links)
  - attributes are mapped to size and color of nodes and links
  - interactivity for:
    - changing the mappings
    - expanding or collapsing nodes within the hierarchy
    - getting additional information, etc.

- mappings in the example:
  - size of nodes: number of e-mail messages of a person
  - color of nodes: function of staff members
  - size of links: number of e-mail messages of the link
  - color of links: blue for few through green and yellow to red for many messages
Graph-based Techniques

SeeNet (cont’d)

visualization of all e-mail connections in a department over a period of time

Narcissus [HDWB 95]

visualization of complex highly interconnected data (e.g., graphs such as the web)
### 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 are combined with dynamics & interactivity

**Examples:** IVEE [AW 95a/b] uses *Starfield Displays* [AS 94] which are scatter-plots 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)

### Distortion Techniques

**Basic Idea:** Distortion of the image to allow a visualization of larger amounts of data

**Overview [LA 94]**

- **Simple:**
  - Perspective Wall [MRC 91]
  - Bifocal Displays [SA 82]
  - TableLens [RC 94]
  - Graph. Fisheye Views [Fur 86, SB 94]
  - Hyperbolic Repr. [LR 94, LRP 95]

- **Complex:**
  - Hyperbolic Repr. [LR 94, LRP 95]
  - 3D-Hyperbolic Repr. [MB 95]
  - Hyperbox [AC 91]
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

Perspective Wall (cont’d)

documents arranged on a perspective wall
Distortion Techniques

### Table Lens [RC 94]

- **Visualization**: Compact visualization of a table (spreadsheet / database) with the possibility of viewing portions of the table in more detail.
- **Example**: Visualization of a baseball database with a few rows being selected in full detail.

### Fisheye View [Fur 86, SB 94]

- **Visualization**: Graph visualization using a fisheye perspective.
- **Example**: Shows an area of interest quite large and with detail and the other areas successively smaller and in less detail.
Distortion Techniques

Hyperbolic Trees [LR 94, LRP 95]

- Visualization of a tree structure in hyperbolic space with different foci
- Visualization of a large organizational hierarchy

Distortion Techniques

3D-Hyperbolic Representation [MB 95, MHCF 96]

- Visualization of a graph in 3D hyperbolic conetree-like representation
- Visualization of a large number of connected web-pages
Distortion Techniques

Hyperbox [AC 91]

parallel processing
performance data visualized as a hyperbox

(mapping of scatterplots onto a hyperbox)

Dynamic / Interaction Techniques

Basic Idea: Dynamic generation of the visualizations or interaction with the visualization for a more effective exploration of the data.

Overview

- Data-to-Visualization Mapping
- Projections
- Filtering (Selection, Querying)
- Linking & Brushing
- Zooming
- Detail on Demand
Dynamic / Interaction Techniques

- **Dynamic / Interactive Data-to-Visualization Mapping**
  - Dynamic or interactive mapping of the data attributes to the parameters of the visualization
  - Parameters of the visualization are
    - x-, y-, and z-axes
    - Color and size of icons, links, etc.
  - Examples:
    - AutoVisual [BF 93]
    - S Plus [BCW 88]
    - XGobi [SCB 92, BCS 96]
    - IVEE / Spotfire [AW 95a/b]
    - SDM [CRMK 95], ...

- **Dynamic / Interactive Projections**
  - Dynamic or interactive variation of the projections
  - Visualization of the remaining parameters in 2D or 3D
  - Automatic variation results in an animation of the data
  - Examples:
    - GrandTour [Asi 85]
    - S Plus [BCW 88]
    - XGobi [SCB 92, BCS 96]
    - Influence & Attribute Explorer [STDS 95, SDTS 95]
    - ...
Dynamic / Interaction Techniques

Dynamic / Interactive Filtering

- dynamic or interactive determination of subsets of the database
- distinction between
  - selection: direct selection of the desired subset
  - querying: specification of properties 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 [Liv 97]
  - Dynamic Queries [AS 94, Eic 94, GR 94]
  - ...

Magic Lenses / 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)
**Dynamic / Interaction Techniques**

**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

**Filter-Flow Model (cont’d)**

complex boolean query:

Find the accountants or engineers from Georgia who are managed by Elizabeth, or the clerks from Georgia who make more than 30,000!
Dynamic / Interaction Techniques

InfoCrystal [Spo 93]

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

![Diagram of Venn diagram with sets A, B, and C]

Dynamic / Interaction Techniques

DEVise [Liv 97]

- tool set for the construction of interactive visualizations
- interactive selection of data items in the upper two subwindows

![Diagram of interactive visualization interface]
Dynamic / Interaction Techniques

- **Dynamic / Interactive Linking & 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]
    - ...

---

Dynamic / Interaction Techniques

**XGobi [SCB 92, BSC 96]**

![climatemap.png](attachment:climatemap.png)

XGobi is used by permission of B. Bell Labs. © Am. Stat. Assoc.

climate and housing data of the US
taken from [BSC 96]
Dynamic / Interaction Techniques

- Dynamic / 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 Details on Demand

- the possibility to interactively obtain more details of the visualized data
- details are, for example, the attribute values corresponding to an icon or additional attribute values of a data item
- examples:
  - IVEE / Spotfire [AW 95a/b]
  - Table Lens [RC 94]
  - Magic Lens [Bie 93]
  - VisDB [KK 94, KK 95]
  - ...

VisDB [Kei 94]
### 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, multivariate 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: An Attempt

<table>
<thead>
<tr>
<th>Comparison</th>
<th>clustering</th>
<th>multivariate hot spot</th>
<th>no. of variates</th>
<th>no. of data items</th>
<th>categorical data</th>
<th>visual overlap</th>
<th>learning curve</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Geometric Techniques</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scatterplot Matrices</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>o</td>
<td>++</td>
</tr>
<tr>
<td>Landscapes</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>o</td>
<td>o</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Prosection Views</td>
<td>++</td>
<td>++</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>o</td>
<td>+</td>
</tr>
<tr>
<td>Hyperslice</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Parallel Coordinates</td>
<td>0</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>o</td>
<td>--</td>
<td>o</td>
</tr>
<tr>
<td><strong>Icon-based Techniques</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stick Figure</td>
<td>o</td>
<td>o</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>o</td>
</tr>
<tr>
<td>Shape Coding</td>
<td>o</td>
<td>-</td>
<td>++</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Color Icon</td>
<td>o</td>
<td>-</td>
<td>++</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td><strong>Pixel-oriented Techniques</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Query-Independent</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>Query-Dependent</td>
<td>+</td>
<td>+</td>
<td>++</td>
<td>++</td>
<td>-</td>
<td>++</td>
<td>-</td>
</tr>
<tr>
<td><strong>Hierarchical Techniques</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dimensional Stacking</td>
<td>+</td>
<td>+</td>
<td>o</td>
<td>o</td>
<td>++</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Worlds-within-Worlds</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>+</td>
<td>o</td>
<td>o</td>
<td>o</td>
</tr>
<tr>
<td>Treemap</td>
<td>+</td>
<td>o</td>
<td>+</td>
<td>o</td>
<td>++</td>
<td>+</td>
<td>o</td>
</tr>
<tr>
<td>Cone Trees</td>
<td>+</td>
<td>+</td>
<td>o</td>
<td>+</td>
<td>o</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>InfoCube</td>
<td>o</td>
<td>o</td>
<td>-</td>
<td>-</td>
<td>o</td>
<td>o</td>
<td>+</td>
</tr>
</tbody>
</table>
Database Visualization Systems

Overview

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

Database Visualization 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: dyn. linking & brushing of scatterplots and histograms)
  - Diamond (SPSS)
    - (-> commercial system for statistical analysis and visualization; features: dyn. linking & brushing of scatterplots, parallel coordinates, etc.)
  - DataSpace [ADLP 95]
    - (-> 3D-arrangement of a large number of arbitrary visualizations)
**Database Visualization Systems**

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

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

- **Special Purpose Visualization Systems**
  - Software & Algorithm Visualization
    (e.g., 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
    (e.g., 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”)
  - Visualization in Information Retrieval
    (e.g., Vibe [Ols 93] - a listing of Information Retrieval Interfaces can be found under “http://www-cui.darmstadt.gmd.de/visit/Activities/Viri/visual.html”;
    for an overview paper see [Car 96])

**Summary and Conclusions**

**Summary**

- there are a number of recently developed visualization techniques which are applicable to database exploration
- there are different techniques for different types of data (relational tables, hierarchies, graphs, etc.)
- many of the techniques are applicable to traditional relational information sources
- there are a number of research prototypes and commercial systems available
Summary and Conclusions

Research Issues

- development of integrated information visualization and exploration systems (integration with techniques from statistics, machine learning, databases, ...)

- in-depth evaluation and comparison of visualization techniques for database exploration (-> possibilities for improvement)

- using more dynamics & interaction to steer the mining process

- case studies in a variety of application areas

Bibliography


Bibliography


Bibliography


### Bibliography

<table>
<thead>
<tr>
<th>Reference</th>
<th>Description</th>
</tr>
</thead>
</table>
Bibliography


### Bibliography

<table>
<thead>
<tr>
<th>Reference</th>
<th>Title and Authors</th>
</tr>
</thead>
</table>

---

<table>
<thead>
<tr>
<th>Reference</th>
<th>Title and Authors</th>
</tr>
</thead>
</table>
Bibliography


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


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