A Survey on Multivariate Data Visualization

Winnie Chan
August 2, 2006
Outline

- Introduction
- Concepts and Terminology
- Classification of Techniques
  - Geometric Projection
  - Pixel-Oriented Techniques
  - Hierarchical Display
  - Iconography
- Discussion and Conclusion
Multivariate data visualization is a specific type of information visualization that deals with multivariate data. The data to be visualized are of high dimensionality in which the correlations between these many attributes are of interest.
Motivations

• Multivariate data are encountered in all aspects by researchers, scientists, engineers, manufactures, financial managers and analysts

• Visualization is motivated by the many situation when they try to obtain an integrated understanding of the data
Challenges

• **Mapping**
  - Bad mapping of data attributes to graphical features may overwhelm observer’s ability
  - Conjunction of several elements in the representations may induce cognition overload to the users
  - A simple example:
Challenges (2)

- **Dimensionality**
  - Resulting display is dense, making it hard for the users to
    - Explore the data space intuitively
    - Discriminate individual dimensions
  - Different ordering of dimensions $\rightarrow$ different conclusions to be drawn
Challenges (3)

- **Design Tradeoffs**
  - Details of each attributes are hardly shown due to the high dimensionality of the data
  - There is a tradeoff between amount of information, simplicity and accuracy

- **Assessment of Effectiveness**
  - We cannot assess the effectiveness of a particular visualization technique
Outline

• Introduction

• Concepts and Terminology

• Classification of Techniques
  - Geometric Projection
  - Pixel-Oriented Techniques
  - Hierarchical Display
  - Iconography

• Discussion and Conclusion
Dimensionality

• Refers to the **number of attributes** that presents in the data
  - 1: one-dimensional 1D / univariate
  - 2: two-dimensional 2D / bivariate
  - 3: three-dimensional 3D / trivariate
  - ≥3: multidimensional / hypervariate / multivariate

• Boundary between high and low dimensionality not clear, generally high dimensionality has >4 variables
## Terminology

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimensions</td>
<td>attributes that are <strong>independent</strong> of each other</td>
</tr>
<tr>
<td>Variables</td>
<td>attributes that are <strong>dependent</strong> of each other</td>
</tr>
<tr>
<td>Multidimensional</td>
<td>dimensionality of the independent dimensions</td>
</tr>
<tr>
<td>Multivariate</td>
<td>dimensionality of the dependent variables</td>
</tr>
</tbody>
</table>

- A more appropriate term: **Multidimensional multivariate data visualization**
Outline

• Introduction
• Concepts and Terminology
  • Classification of Techniques
    - Geometric Projection
    - Pixel-Oriented Techniques
    - Hierarchical Display
    - Iconography
• Discussion and Conclusion
Classifications

- Based on the overall approaches taken to generate the resulting visualizations
- Taxonomy
  - Geometric Projection
  - Pixel-Oriented
  - Hierarchical Display
  - Iconography
Outline

• Introduction
• Concepts and Terminology
• Classification of Techniques
  - Geometric Projection
  - Pixel-Oriented Techniques
  - Hierarchical Display
  - Iconography
• Discussion and Conclusion
Geometric Projection

- Informative projections and transformations of multidimensional datasets
- Maps attributes to 1-3D or arbitrary space
  - Effective in detecting outliers and correlation amongst different dimensions
  - Can handle huge datasets when appropriate interaction techniques are introduced
- Data attributes are treated equally, but may not be perceived equally; rearrangement is important if the display should not be biased
- Potential visual cluttering and record overlapping that overwhelm the user’s perception capabilities
Scatterplot Matrix

- Scatterplot: 2 attributes projected along the x- and y-axis
- Collection of scatterplots is organized in a matrix
  - Straightforward
  - Important patterns in higher dimensions barely recognized
  - Chaotic when number of data items too large
Prosection Matrix

- Prosection: Orthogonal projections of 2D data
- Data items lie in the selected multi-dimensional range are colored differently
- Can indicate tolerances on parameter values (yellow rectangle)
- Less information about correlations between >2 attributes
HyberSlice

- Matrix graphics representing a scalar function of the variables
- Allows data navigation around a user defined focal point
- Targets at continuous scalar functions rather than discrete data
Hyberbox

- Plots constructed as $n$-dimensional box instead of a matrix
  
  ➕ Can map variables to both size and shape of each face
  
  ➕ Can emphasize or de-emphasize some variables
  
  ➖ $n$-Dimensional box modeled in 2D → arbitrary length and orientation which may convey wrong information
Parallel Coordinates

- Attributes represented by parallel vertical axes scaled within the data range
- Each data item represented by a polygonal line that intersects each axis at the attribute data value
  - Correlations among attributes studied by spotting the locations of the intersection points
  - Effective for revealing data distributions and functional dependencies
    - Visual clutter due to limited space available for each parallel axis
    - Axes packed very closely when dimensionality is high
Varied Parallel Coordinates

- Circular Parallel Coordinates
  - Adopts a radial arrangement of axes

- Hierarchical Parallel Coordinates
  - Displays aggregation information derived from a hierarchical clustering of the data, at different levels of abstraction
Andrews Curve

- Similar to Parallel Coordinates with each data item plotted as a curved line, like a Fourier transform of data point
  - Close points, similar curves; distant points, distinct curves → useful for detecting clusters and outliers
  - Computationally expensive for large datasets
Radical Coordinates

- Lines associated with attributes emanate radically from the center of the circle
- Spring constants attached to attribute values define positions of data points along the lines
- Points with approximately equal or similar dimensional values lie close to the center
Star Coordinates

- Scatterplots for higher dimensions: attribute as axis on a circle, data item as point
- Change the length of axis → alters contribution of attribute
- Change the direction of axis → angles not equal, adjusts correlations between attributes

👍 Useful for gaining insight into hierarchically clustered datasets and for multi-factor analysis for decision-making
Table Lens

- Represents rows as data items and columns as attributes
- Each column viewed as histogram or plot
- Information along rows or columns interrelated

👍 Uses the familiar concept “table”
Outline

• Introduction
• Concepts and Terminology
• Classification of Techniques
  - Geometric Projection
  - Pixel-Oriented Techniques
  - Hierarchical Display
  - Iconography
• Discussion and Conclusion
Pixel-Based Techniques

- Each attribute value of a data item represented by one pixel, based on some color scale
- \( n \) colored pixels needed to represent one data item for \( n \)-dimensional data, with each attribute values being placed in separate sub-windows
  - Relationships between attributes detected by relating corresponding regions in the multiple windows
  - Record overlap and visual cluttering not likely because each data item is uniquely mapped to a pixel
  - Not straightforward
Two subgroups:

- **Query-independent**
  - Favored by data with natural ordering according to one attribute
  - Absolute values are mapped to colors

- **Query-dependent**
  - Appropriate if the feedback to query is of interest
  - Distances of attribute values to the query are mapped to colors
Space Filling Curves

- Query-independent
- Pixels representing a data attribute arranged on curves in their sub-windows

 harms
- Provides a better clustering of closely related data items

- Peano and Hibert curves
- Morton or Z-Curve
Recursive Pattern

- Query-independent
- Based on generic recursive scheme performed iteratively

👍 Allows users to influence the arrangement of data items on-the-fly
Spiral Technique

• Query-dependent
• Arranges pixels in spiral form according to the overall distance from the query

- Yellow center represents the data items satisfying the user specified query
- Additional window (top left one) showing overall distance, i.e. the color scheme encoding distance from query results
Axes Technique

- Query-dependent
- Arranges pixels in partial spirals in each quadrant, i.e. two attributes are assigned to the axes and data items are arranged according to the displacement from the query.

- Additional window (top left one) showing overall displacement, i.e. the color scheme encoding displacement from query results.

- Yellow center represents the data items satisfying the user specified query.
Circle Segment

- Query-dependent
- Assigns attributes on the segments of a circle
- Single data item appears in the same position at different segments
- Ordering and colors of the pixels similarly determined by overall distance to the query
Pixel Bar Chart

- Derived from regular bar chart
- Bars can be
  - Histogram plotting one attribute against its values
  - x-y diagram plotting one attribute against another
- Pixel color used to encode the values of another attribute
- Multi-pixel bar charts used for higher-dimensional data

- Equal-width pixel bar chart
- Equal-height pixel bar chart
• 3 pixel bar charts all plotting product type (1-12) against amount of money
• Color in each pixel bar chart encodes different attributes $i$, $j$ and $k$ respectively
Outline

• Introduction
• Concepts and Terminology
• Classification of Techniques
  - Geometric Projection
  - Pixel-Oriented Techniques
    - Hierarchical Display
    - Iconography
• Discussion and Conclusion
Hierarchical Display

- Subdivides the data space and presents subspaces in a hierarchical fashion

👍 Effective for visualizing hierarchical data, or data in which several attributes are more important or of more interest

👎 Attributes are treated differently → different mappings produce different views of the underlying data

👎 Interpretation of results requires training
Hierarchical Axis

- Axes laid out horizontally in a hierarchical fashion
- Can plot many attributes in one screen
- Example: Histograms within histograms
Dimensional Stacking

- Partitions the data space into 2D subspaces which are stacked into each other
- Important attributes chosen for outer levels
  - Adequate for discrete categorical or binned ordinal values
Worlds Within Worlds

- Also known as \( n \)-Vision
- Subdivides data space into 3D subspaces
  - Generates interactive hierarchy display instead of static objects
Treemap

- Uses a hierarchical partitioning of the screen into regions, depending on the attribute values
- Sizes of the nested rectangles represent the attribute values
- Suitable to obtain an overview on large datasets with multiple ordinal attributes
- Fully utilizes the available display space
Outline

• Introduction
• Concepts and Terminology
• Classification of Techniques
  - Geometric Projection
  - Pixel-Oriented Techniques
  - Hierarchical Display
  - Iconography
• Discussion and Conclusion
Iconography

• Maps each multidimensional data item to an icon/glyph that contains several graphical parameters

👍 Observations of graphical features are pre-attentive

👎 Biases in interpreting the result
  - Some features are more salient than others
  - Adjacent elements are easier to be related
  - Accuracy of perceiving different graphical attributes varies between humans
Chernoff Faces

- Two attributes mapped to the 2D position of a face
- Remaining attributes mapped to properties of the face, e.g. shape of nose, mouth, eyes and face
  - Different visual features are not quite comparable to each other
  - Can only visualize a limited amount of data items
  - Semantic relation to the task has significant impact on the perceptive effectiveness
Star Glyph

- Dimensions represented as equal angular axes radiating from the center of a circle
- An outer line connects the data value points on each axis
- Each data item presented by one star glyph

Display becomes overwhelming when the number of data items increases
Maps two attributes to the display axes and the remaining to the rotation angle, length, thickness or color of the limbs

Packed icons exhibit some texture patterns showing some data features

Visual discernment of an important pattern is highly dependent upon the selection of an appropriate graphical attribute
Shape Coding

- Each data item represented by one array of pixels
- Pixels are mapped to a color scale according to the attribute values

👍 Arrays can contain an arbitrary number of pixels
Color Icon

- Combines pixel-based spiral axes and icon-based shape coding
- Color, shape, size, orientation, boundaries and area sub-dividers can be used to map the data
- Merges color, shape and texture perception for iconographic integration
Texture

- 3 dominant visual dimensions: orientation, size and contrast

👍 Contains various visual dimensions that human can distinguish effectively and pre-attentively

👍 Outcomes are more engaging and aesthetic that are more attractive and favorable
Problems remain in finding a suitable mapping from data attributes to texture features.

Contrast illusions are induced when we are comparing the scale and orientation of textures.

Would you believe that the two central circles are identical? Yes, you are fooled by your eyes (again).
Natural Texture

- Data attributes mapped to texture features including
  - Hue
  - Luminance
  - Scale
  - Regularity
  - Periodicity
  - Directionality
  - Homogeneity
  - Transparency
  - Fuzziness
  - Level of abstraction
Nonphotorealistic Texture

- Uses perceptually-based brush strokes
- Data attribute mapped to a nonphotorealistic property such as color, orientation, coverage, size, coarseness and weight
- Attributes values identified from the different visual appearances of the brush strokes
Outline

• Introduction
• Concepts and Terminology
• Classification of Techniques
  - Geometric Projection
  - Pixel-Oriented Techniques
  - Hierarchical Display
  - Iconography

• Discussion and Conclusion
Discussion and Conclusion

• Motivations and challenges of visualizing high-dimensional multivariate data
• Terminology of multidimensional multivariate data visualization
• Taxonomy that categorized multivariate data visualization techniques into four broad classes
  - Geometric projection
  - Pixel-oriented techniques
  - Hierarchical display
  - Iconography
Discussion and Conclusion (2)

• **Geometric projection**
  - ✅ Does not require us much effort to understand the representations
  - ❌ Becomes problematic when the dimensionality of the data increases, only 3 dimensions are mapped to a 3D space

• **Pixel-oriented techniques**
  - ✅ Corresponding pixels appear at the same position in each respective window
  - ❌ Outcomes not straightforward; requires training

• **Hierarchical displays**
  - ✅ Effective in visualizing hierarchical data
  - ❌ Outcomes not straightforward; requires training

• **Iconography**
  - ✅ Has numerous graphical properties that data attributes can map to
  - ✅ Shows overall features and relationships in the data when packed and exhibit some texture patterns
  - ❌ Has difficulties in finding a good mapping from data attributes to graphical features
A picture is worth a thousand words.

-- The End --
Thank you!