CIS 4930/6930-902
Scientific Visualization

Tabular Data
Paul Rosen
Assistant Professor
University of South Florida

slides credits Miriah Meyer (U of Utah)
DATASET TYPES

- **Tables**
  - Items
  - Attributes

- **Networks & Trees**
  - Items (nodes)
  - Links
  - Attributes

- **Fields**
  - Grids
  - Positions
  - Attributes

- **Geometry**
  - Items
  - Positions

- **Clusters, Sets, Lists**
  - Items
Arrange Tables

1. Express Values

2. Separate, Order, Align Regions
   - Separate
   - Order
   - Align
     - 1 Key List
     - 2 Keys Matrix
     - 3 Keys Volume
     - Many Keys Recursive Subdivision

3. Axis Orientation
   - Rectilinear
   - Parallel
   - Radial

4. Layout Density
   - Dense
   - Space-Filling
ARRANGE is the focus of all four design choices for tabular data.
Spatial channels are the most effective for all attribute types.
SINGLE KEY, SINGLE VALUE
ENCODE ONE KEY ATTRIBUTE

BAR, DOT, & LINE CHARTS
DONT’T USE LINE CHARTS FOR CATEGORICAL ATTRIBUTES!

ok: “Men are taller than women (on average)”

bad: “The more male a person is, the taller he/she is”

ok: “Twelve year olds are taller than ten year olds”

ok: “Height increases with age”
BANKING TO 45°

The aspect ratio of a graph is an important factor for judging rate of change.

**perceptual principle**: most accurate angle judgment is at 45°
RESULTS

people use two different strategies to estimate slope—angle and height. Slope angle accuracy NOT minimized at 45°.
**Tick Placement**

Ticks help in user interpretation of data, but too much may hinder
AUTOMATIC TICKS

optimization of label formatting, font size, and orientation

placement based on simplicity, coverage, granularity, and legibility
PIE CHARTS: TAKE CARE WITH ACCURACY
2 KEYS, 1 VALUE
ENCODE USING TWO KEYS: HEATMAP

- uses heatmap representation
- matrix layout using keys
- encode values with color
- often augmented with clustering
ENCODE USING TWO KEYS:

HEATMAP

uses heatmap representation
matrix layout using keys
encode values with color
often augmented with clustering

<p>| | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0.7</td>
<td>0.8</td>
<td>1</td>
<td>1</td>
<td>0.8</td>
<td>0.6</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.2</td>
<td>0.5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td>0.8</td>
<td>0.5</td>
<td>0.3</td>
<td>0.5</td>
<td>0.8</td>
</tr>
<tr>
<td>0.7</td>
<td>0.5</td>
<td>0.8</td>
<td>0.7</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0.3</td>
<td>0.4</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>0.7</td>
<td>0.5</td>
<td>0.3</td>
</tr>
</tbody>
</table>
ENCODE USING TWO KEYS:

HEATMAP

uses heatmap representation
matrix layout using keys
encode values with color
often augmented with clustering
ENCODE USING TWO KEYS:

HEATMAP

uses heatmap representation
matrix layout using keys
encode values with color
often augmented with clustering
here, used on genomic data

EISEN 1998
Interactively Exploring Hierarchical Clustering Results

The Hierarchical Clustering Explorer provides a dendrogram and color mosaic linked to two-dimensional scattergrams, a variety of visualization options, and dynamic query controls for use in genomic microarray data analysis.

Molecular biologists and geneticists seek to understand the function of genes, including the more than 6,000 genes in the yeast genome and the estimated 40,000 genes in the human genome. Recently developed for genome analysis, DNA microarrays—also known as gene arrays or gene chips—usually consist of glass or nylon substrates that measure 1 x 3 inches or smaller. These chips contain specific DNA gene samples spotted in an array by a robotic printing device. Researchers spread fluorescently labeled messenger RNA (mRNA) from an experimental condition onto the DNA gene samples in the array. This mRNA binds (hybridizes) strongly with some DNA makes it impossible to display a large microarray experiment—on one screen.

Researchers also struggle to understand the implications of a specific clustering result. Because the clusters occupy a high-dimensional space and involve so many experimental conditions, researchers find it difficult to view patterns on a 2D or even a 3D display. Further, data can contain hundreds of variously sized clusters, which makes spotting the meaningful clusters a challenge, especially when using a static display. Users need an efficient interactive visualization tool to facilitate pattern extraction from microarray data sets.

Hierarchical clustering has been shown to be effective in microarray data analysis for identifying
SCALABILITY THROUGH INTERACTION

interactively controlled reduction
aggregation, filtering, and navigation

view coordination
overview+detail, small multiples, side-by-side multiform views
with linked highlighting
CRITIQUE: WHAT DO YOU THINK?
MULTIPLE ATTRIBUTES
ENCODE USING SCATTERPLOTS
ENCODE USING STACKED BAR CHART
ALIGN USING MULTIPLE KEYS

LineUp: Visual Analysis of Multi-Attribute Rankings

Samuel Gratl, Alexander Lex, Nils Gehlenborg, Hanspeter Pfister and Marc Streit

Fig. 1. LineUp showing a ranking of the top Universities according to the QS World University Ranking 2012 dataset with custom attributes and weights, compared to the official ranking.

Abstract—Rankings are a popular and universal approach to structuring otherwise unorganized collections of items by computing a rank for each item based on the value of one or more of its attributes. This allows us, for example, to prioritize tasks or to evaluate the performance of products relative to each other. While the visualization of a ranking itself is straightforward, its interpretation is not.
**Challenge**

Rankings based on single attribute are trivial to display when based on multiple attributes: not clear how attributes contribute to ranking, not clear how changes to multiple attributes will affect ranking, different contexts/people/situations will rank on multiple attributes differently.
LineUp
Visual Analysis of Multi-Attribute Rankings

Samuel Gratzl, Alexander Lex, Nils Gehlenborg, Hanspeter Pfister and Marc Streit
CRITIQUE: WHAT DO YOU THINK?
SPLOMs: SCATTERPLOT MATRICES

Nine characteristics of Abalone (sea snails)
**PARALLEL COORDINATES**

**scatterplot limitation:** visual representation with orthogonal axes can show only two attributes with spatial position channel

**alternative:** line up axes in parallel to show many attributes with position channel

item encoded with a line with $n$ segments

$n$ is the number of attributes shown
PARALLEL COORDINATES

cylinders: 9

displacement: 455 sq in

weight: 5141 lbs

horsepower: 231 hp

acceleration (0-60 mph): 7.5 sec

mileage: 7.5 mpg

year: 83

Protovis
**EXAMPLE**

<table>
<thead>
<tr>
<th></th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
<th>V4</th>
<th>V5</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>D2</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>D3</td>
<td>9</td>
<td>8</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>
EXAMPLE

V1  V2  V3  V4  V5

D1
7  3  4  8  1
EXAMPLE
EXAMPLE
PARALLEL COORDINATES TASK

show correlation
positive correlation
straight lines
negative correlation
lines cross at a single point

Figure 3. Parallel Coordinate Plot of Six-Dimensional Data Illustrating Correlations of $p = 1, .8, .2, 0, -.2, -.8, \text{ and } -1$. WEGMAN 1990
PARALLEL COORDINATES TASK

do you see any correlations?
PARALLEL COORDINATES TASK

visible patterns only between neighboring axis pairs

how to pick axis order?

usual solution: reorderable axes, interactive exploration

same weakness as many other techniques
downside: human-powered search

not directly addressed in HPC paper
Hierarchical Parallel Coordinates for Exploration of Large Datasets

Ying-Huey Fua, Matthew O. Ward and Elke A. Rundensteiner
Computer Science Department
Worcester Polytechnic Institute
Worcester, MA 01609
{yingfua,matt,runden}@cs.wpi.edu

Abstract

Our ability to accumulate large, complex (multivariate) data sets has far exceeded our ability to effectively process them in search of patterns, anomalies, and other interesting features. Conventional multivariate visualization techniques generally do not scale well with respect to the size of the data set. The focus of this paper is on the interactive visualization of large multivariate data sets based on a number of novel extensions to the parallel coordinates display technique. We develop a multiresolutional view of the data via hierarchical clustering, and use a variation on parallel coordinates to convey aggregation information for the resulting clusters. Users can then navigate the resulting structure until the desired focus region and level of detail is reached, using our suite of navigational and filtering tools. We describe the design and implementation of our hierarchical parallel coordinates system which is based on extending the XmdVTool system. Lastly, we show examples of the tools and techniques applied to large (hundreds of thousands of records) multivariate data sets.

Keywords: Large-scale multivariate data visualization, hierarchical data exploration, parallel coordinates.

1 Introduction

- Dimensional embedding techniques, such as dimensional stacking [16] and worlds within worlds [6].
- Dimensional subsetting, such as scatterplots [5].
- Dimensional reduction techniques, such as multidimensional scaling [20, 15, 29], principal component analysis [12] and self-organizing maps [14].

Most of these techniques do not scale well with respect to the size of the data set. As a generalization, we postulate that any method that displays a single entity per data point invariably results in overlapped elements and a convoluted display that is not suited for the visualization of large data sets. The quantification of the term "large" varies and is subject to revision in sync with the state of computing power. For our present application, we define a large data set to contain $10^6$ to $10^8$ data elements or more.

Our research focus extends beyond just data display, incorporating the process of data exploration, with the goal of interactively uncovering patterns or anomalies not immediately obvious or comprehensible. Our goal is thus to support an active process of discovery as opposed to passive display. We believe that it is only through data exploration that meaningful ideas, relations, and subsequent inferences may be extracted from the data. The major hurdles we need to overcome are the problems of display density/clutter (too many overlapping elements), and the lack of efficient means to "zoom" into an interesting area.
HIERARCHICAL PARALLEL COORDINATES

goal: scale up parallel coordinates to large datasets
challenge: overplotting/occlusion
HPC: ENCODING DERIVED DATA

visual representation: variable-width opacity bands
show whole cluster, not just single item
min / max: spatial position
cluster density: transparency
mean: opaque
HPC: INTERACTING WITH DERIVED DATA

interactively change level of detail to navigate cluster hierarchy
RADIAL LAYOUTS USE POLAR COORDINATES
RADAR PLOT & STAR GRAPH

“parallel” dimensions in polar coordinate space
best if same units apply to each axis
**CRITIQUE: WHAT DO YOU THINK?**

[Image: A radar chart showing the percentage of owners using specific services on smartphones and tablets. Services include Email, Social Networks, Internet Banking, News, Sports, Search Engine, View Shopping Sites, Paying Online, Buy Online, Stream Music, Online Gaming, Navigation, Offline Gaming, Photo Video, Reading, TV/Movies Streaming, Listen Music, Listen Radio, Sending Money, Watch TV, Other.]

DENSE PIXEL DISPLAY: \textit{VisDB}

represent each data item, or each attribute in an item as a single pixel
can fit as many items on the screen as there are pixels, on the order of millions
relies heavily on color coding
challenge: what’s the layout?
CRITIQUE: WHAT DO YOU THINK?