Part III

Geometrically-transformed displays
Outline

7 Introduction

8 Scatter plot matrix
   - Interaction

9 Parallel Coordinates
   - Overlapping
   - Variable order
Geometrically-transformed display

A class of Visual mapping transformation:

- for table data
- mostly data independent
- maps one object to a set of points (and maybe lines) in 2D or in 3D
- examples:
  - scatter plot matrix
  - Andrews’ curves
  - parallel coordinates
Outline

7 Introduction

8 Scatter plot matrix
   - Interaction

9 Parallel Coordinates
   - Overlapping
   - Variable order
Scatter plot matrix

- A matrix of scatter plots

<table>
<thead>
<tr>
<th></th>
<th>$X_1$ alone</th>
<th>$(X_2, X_1)$</th>
<th>$(X_3, X_1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(X_1, X_2)$</td>
<td></td>
<td>$X_2$ alone</td>
<td>$(X_3, X_2)$</td>
</tr>
<tr>
<td>$(X_1, X_3)$</td>
<td>$(X_2, X_3)$</td>
<td></td>
<td>$X_3$ alone</td>
</tr>
</tbody>
</table>

- Pros:
  - almost no learning curve
  - inherits scatter plot extensions

- Cons:
  - worsen scatter plot overlapping problem: $p$ variables $\Rightarrow p^2$ scatter plots $\Rightarrow p^2 n$ values ($n$ objects)
  - links between plots are difficult to understand
Number of pixels per object and per attribute on a 2 MP display

Number of attributes

Number of objects

1
5
10
20
50

F. Rossi (INRIA)
Scalability

Anderson’s/Fisher’s Iris

4+1 variables, 150 objects
Scalability

8+1 variables, 4177 objects
Scalability

13+1 variables, 178 objects
Scalability via interaction

- **User choice:**
  - variable selection
  - variable ordering

- **Instant zoom (popup)**

- **Linking and brushing:**
  - **brushing:**
    - selection of a subset of the objects
    - display of the selection (hue based)
  - **linking:**
    - brush in one view
    - display the results in all views
    - views don’t have to be scatter plots

- In previous slides: class information!
### Italian Wines

<table>
<thead>
<tr>
<th>Alcohol</th>
<th>Malic acid</th>
<th>Ash</th>
<th>Alcalinity</th>
<th>Magnesium</th>
<th>Total phenols</th>
<th>Flavanoids</th>
<th>Nonflavanoid phenols</th>
<th>Proanthocyanins</th>
<th>Color intensity</th>
<th>Hue</th>
<th>OD280/OD315 ratio</th>
<th>Proline</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 2 3 4 5 6</td>
<td>10 20 30</td>
<td>1.0</td>
<td>2.0</td>
<td>3.0</td>
<td>0.2</td>
<td>0.4</td>
<td>0.6</td>
<td>2</td>
<td>6</td>
<td>10</td>
<td>1.5</td>
<td>2.5</td>
</tr>
<tr>
<td>11 13 15</td>
<td>1.1 1.2 1.3</td>
<td>1.5</td>
<td>1.6</td>
<td>1.7</td>
<td>0.6</td>
<td>0.8</td>
<td>1.0</td>
<td>1.1</td>
<td>1.2</td>
<td>1.3</td>
<td>1.4</td>
<td>1.5</td>
</tr>
<tr>
<td>2 8</td>
<td>0.2 0.6</td>
<td>1.1</td>
<td>1.2</td>
<td>1.3</td>
<td>0.6</td>
<td>0.8</td>
<td>1.0</td>
<td>1.1</td>
<td>1.2</td>
<td>1.3</td>
<td>1.4</td>
<td>1.5</td>
</tr>
<tr>
<td>11 13 15</td>
<td>1.5 2.5</td>
<td>80</td>
<td>120</td>
<td>160</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>0.5</td>
<td>1.5</td>
<td>2.5</td>
</tr>
<tr>
<td>11 13 15</td>
<td>1.1 1.2 1.3</td>
<td>400</td>
<td>1000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Linking and brushing (Ggobi)
Linking and brushing (Ggobi)
Linking and brushing (Ggobi)
Linking and brushing (Ggobi)
Linking and brushing (Ggobi)
Linking and brushing (Ggobi)
Outline

7 Introduction

8 Scatter plot matrix
   - Interaction

9 Parallel Coordinates
   - Overlapping
   - Variable order
Parallel Coordinates

Proposed in 1985 by A. Inselberg

- A variable = a vertical axis
- An object = a piecewise linear curve
- \((x_1, \ldots, x_p)\) is mapped to the polyline that joins \((1, x_1), (2, x_2), \ldots, (p, x_p)\)
Iris dataset

Anderson's/Fisher's Iris

Sepal.Length Sepal.Width Petal.Length Petal.Width Species

F. Rossi (INRIA)

Infovis & ML

27-28/02/2007 95 / 172
Parallel Coordinates

Pros:
- reasonable variable scalability
- powerful after proper training

Cons:
- learning curve
- major overlapping problem
- variable order

Reducing the problems:
- rendering (transparency)
- interactivity (brushing)
- simplification (data clustering)
- variable ordering (user guided and/or automated)
Overlapping Spheres
Overlapping Spheres

Spheres

V1 V2 V3
Transparency

Spheres

V1 V2 V3
Transparency

Spheres

V1 V2 V3
Brushing
Brushing
Brushing
Brushing and linking

[Image of a parallel coordinates plot with brushing and linking features]
Brushing and linking

F. Rossi (INRIA)
Infovis & ML
27-28/02/2007
100 / 172
Brushing and linking
Brushing and linking
Brushing and linking
Clustering

Abalone (UCI)

Length Diameter Height Whole.weight Shucked.weight Shell.weight Rings

F. Rossi (INRIA)
Infovis & ML
27-28/02/2007
Clustering

Abalone (UCI)

Length Diameter Height Whole.weight Shucked.weight Shell.weight Rings

F. Rossi (INRIA)
Clustering

Abalone (UCI)

Length Diameter Height Whole.weight Shucked.weight Shell.weight Rings

F. Rossi (INRIA)  Infovis & ML  27-28/02/2007
Clustering

Abalone (UCI)

Length  Diameter  Height  Whole.weight  Shucked.weight  Shell.weight  Rings
Clustering

Abalone (UCI)

Length  Diameter  Height  Whole.weight  Shucked.weight  Shell.weight  Rings

F. Rossi (INRIA)

Infovis & ML

27-28/02/2007  101 / 172
Are petal variables correlated?
Variable order

Anderson's/Fisher's Iris

Are petal variables correlated?
Automated ordering

- resemblance measure between variables (e.g., correlation): \( s(i, j) \)
- variable permutation \( \sigma : \{1, \ldots, p\} \rightarrow \{1, \ldots, p\} \)
- optimal \( \sigma \) with respect to:

\[
Q(\sigma) = \sum_{k=1}^{p-1} s(\sigma(k), \sigma(k + 1))
\]

- NP-complete (Ankerst, Berchtold & Keim, 1998) \( \Rightarrow \) heuristics:
  - traveling salesman analogy: optimization methods (e.g., Ant colony)
  - variable clustering
  - etc.

- an instance of the seriation problem: finding an optimal order among objects
Projection based approach

Projection to 1D:

- $d(i,j)$ dissimilarity between variables
- map a variable $i$ to a real number $y_i$
- choose $(y_i)_{1 \leq i \leq p}$ such that $(y_i - y_j)^2 \approx d(i,j)^2$:
  - many possibilities
  - see Prof. Lee’s lecture
- order variables according to $(y_i)_{1 \leq i \leq p}$
- this is exactly a projection problem
Ordering by Multi Dimensional Scaling

Anderson's/Fisher's Iris

Sepal.Width  Petal.Length  Petal.Width  Sepal.Length

F. Rossi (INRIA)

Infovis & ML  27-28/02/2007  105 / 172
Ordering by Multi Dimensional Scaling
Ordering by Multi Dimensional Scaling
Clustering based approach

- Cluster variables according to resemblance measure
- Use clustering results to produce an order:
  - One dimensional Self Organizing Map
  - Hierarchical clustering:
    - with some heuristics
    - with optimal leaf ordering ($O(p^4)$, Bar-Joseph, Gifford and Jaakkola, 2001)
- Can be used to reduce the number of variables
Hierarchical clustering with heuristics

Wine variables

- Ash
- Alkalinity of ash
- Malic acid
- Hue
- Nonflavanoid phenols
- Proanthocyanins
- OD280/OD315 of diluted wines
- Total phenols
- Flavonoids
- Magnesium
- Color intensity
- Alcohol
- Proline
Hierarchical clustering with heuristics
Hierarchical clustering with heuristics

Proanthocyanins  Total.phenols  Hue  Malic.acid  Alcalinity.of.ash  Alcohol  Magnesium  Ash