Multivariate Visual Representations 1

CS 4460 – Intro. to Information Visualization
Sep. 18, 2017
John Stasko

Learning Objectives

- For the following visualization techniques/systems, be able to describe each and its visual encoding, know what type of data it's best for, know its strengths and limitations, and understand how to apply it
  - Iconic representations (Chernoff faces), Table Lens, InfoZoom, EZChooser, Mosaic plot, Star plots
- Explain the visual encoding and design issues of Parallel Coordinates, as well as their utility and limitations
- Understand how the different types of variables in a multivariate data set influence the visualization technique that should be chosen to represent the data
- Be able to apply any of these techniques to a data set that is an appropriate match for them
How Many Variables?

- Data sets of dimensions 1, 2, 3 are common
- Number of variables per class
  - 1 - Univariate data
  - 2 - Bivariate data
  - 3 - Trivariate data
  - >3 – Hyper/Multivariate data

Focus This Week

Earlier

- We examined a number of tried-and-true techniques/visualizations for presenting multivariate (typically <=3) data sets
  - Bar graph, line graph, pie chart, scatterplot, box plot, trellis display, crosstab, radar graph, heatmap

- Hinted at how to go above 3 dimensions
Hypervariate Data

- How about 4 to 20 or so variables (for instance)?
  - Lower-dimensional hypervariate data
  - Many data sets fall into this category

How would you handle that?

One Approach

- Stay with standard views, but use lots and lots of them
More Dimensions

- Fundamentally, we have 2 geometric (position) display dimensions
- For data sets with >2 variables, we must project data down to 2D
- Come up with visual mapping that locates each dimension into 2D plane
- Computer graphics: 3D->2D projections

Wait a Second

- A spreadsheet already does that
  - Each variable is positioned into a column
  - Data cases in rows
  - This is a projection (mapping)
- What about some other techniques?
  - Already seen a couple
Scatterplot Matrix

Represent each possible pair of variables in their own 2-D scatterplot

If pairwise correlation is key

Key Principle (today)

- Handle all data sets generically
  - Examine techniques not specific to some data or domain
  - Technique can generally handle all data sets
Iconic Representations

- Glyph (graphical object) represents a data case
- Visual properties of glyph represent different variables

Remember?
Chernoff Faces

Encode different variables’ values in characteristics of human face

Examples

Cute applet:  http://www.cs.uchicago.edu/~wiseman/chernoff/
Table Lens

- Spreadsheet is certainly one hypervariate data presentation
- Idea: Make the text more visual and symbolic
- Just leverage basic bar chart idea

Visual Mapping

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Name</td>
<td>Quota</td>
<td>Variance to quota</td>
<td>% of quota</td>
<td>Forecast</td>
</tr>
<tr>
<td>1</td>
<td>Sales rep</td>
<td>200,000</td>
<td>100%</td>
<td>0%</td>
<td>300,000</td>
</tr>
<tr>
<td>2</td>
<td>Allbright, Guy</td>
<td>200,000</td>
<td>100%</td>
<td>-10,000</td>
<td>200,000</td>
</tr>
<tr>
<td>3</td>
<td>Deen, Cheryl</td>
<td>200,000</td>
<td>100%</td>
<td>10,000</td>
<td>200,000</td>
</tr>
<tr>
<td>4</td>
<td>Cartwright, Bonnie</td>
<td>100,000</td>
<td>50%</td>
<td>0%</td>
<td>50,000</td>
</tr>
<tr>
<td>5</td>
<td>Colburn, Michael</td>
<td>100,000</td>
<td>50%</td>
<td>0%</td>
<td>50,000</td>
</tr>
<tr>
<td>6</td>
<td>Spalding, John</td>
<td>250,000</td>
<td>125%</td>
<td>50%</td>
<td>425,000</td>
</tr>
<tr>
<td>7</td>
<td>Dew, Jean</td>
<td>75,000</td>
<td>37.5%</td>
<td>-12,500</td>
<td>62,500</td>
</tr>
<tr>
<td>8</td>
<td>Jones, Suzanne</td>
<td>100,000</td>
<td>50%</td>
<td>0%</td>
<td>50,000</td>
</tr>
<tr>
<td>9</td>
<td>Lennon, Terri</td>
<td>100,000</td>
<td>50%</td>
<td>0%</td>
<td>50,000</td>
</tr>
<tr>
<td>10</td>
<td>Stenberg, George</td>
<td>200,000</td>
<td>100%</td>
<td>0%</td>
<td>200,000</td>
</tr>
<tr>
<td>11</td>
<td>Leutzon, Ingrid</td>
<td>175,000</td>
<td>87.5%</td>
<td>-12,500</td>
<td>192,500</td>
</tr>
<tr>
<td>12</td>
<td>Hulpman, Paul</td>
<td>225,000</td>
<td>112.5%</td>
<td>-25,000</td>
<td>250,000</td>
</tr>
<tr>
<td>13</td>
<td>Tresch, Sheila</td>
<td>50,000</td>
<td>25%</td>
<td>-25,000</td>
<td>25,000</td>
</tr>
<tr>
<td>14</td>
<td>Wettrock, Glenn</td>
<td>150,000</td>
<td>75%</td>
<td>-50,000</td>
<td>100,000</td>
</tr>
</tbody>
</table>

Change quantitative values to bars

How did they handle nominal data?
Instantiation

Details

Focus on item(s) while showing the context
Alternative

What if you
Flipped rows and columns
Sorted each row

Result

InfoZoom

Commercial product
Demo/video
Similar Idea

EZChooser

Alternate UI

- Can slide the values in a row horizontally
- A particular data case then can be lined up in one column, but the rows are pushed unequally left and right
Attributes as Sliding Rods

Limitations?

- Number of cases (horizontal space)
- Nominal & textual attributes don’t work quite as well
An Application

• What if you cared about ranking items?
  – Think of the attributes per item as contributing to some score or value for it
• Apply the Table Lens representation we saw earlier

LineUp

Gratzl et al
TVCG (InfoVis) ’13
Categorical data?

• How about multivariate categorical data?

• Students
  – Gender: Female, male
  – Eye color: Brown, blue, green, hazel
  – Hair color: Black, red, brown, blonde, gray
  – Home country: USA, China, Italy, India, ...
Mosaic Plot

Fundamental Limitation

- Run out of rows/columns for lots of data cases
- How about an alternative generic representation?
### Parallel Coordinates

<table>
<thead>
<tr>
<th>variables</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>data cases</td>
<td>D2</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>D3</td>
<td>9</td>
<td>8</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>
Parallel Coordinates

Encode variables along a horizontal row

Vertical line specifies different values that variable can take

Data point represented as a polyline

Questions

What do two correlated variables look like?

What do two inversely correlated variables look like?
Issue

- Different variables can have values taking on quite different ranges
- Must normalize all down (e.g., 0->1)

To Learn More

Great site that's all about parallel coords

http://www.parallelcoordinates.de
Challenges

- Too much data
- Order of dimensions really matters

Star Plots (Radar Chart)

* Space out the $n$ variables at equal angles around a circle
* Each “spoke” encodes a variable’s value

Alternative Rep.

Data point is now a “shape”
Generalizing the Principles

- General & flexible framework for axis-based visualizations
  - Scatterplots, par coords, etc.
- User can position, orient, and stretch axes
- Axes can be linked
Learning Objectives

- For the following visualization techniques/systems, be able to describe each and its visual encoding, know what type of data it's best for, know its strengths and limitations, and understand how to apply it
  - Iconic representations (Chernoff faces), Table Lens, InfoZoom, EZChooser, Mosaic plot, Star plots
- Explain the visual encoding and design issues of Parallel Coordinates, as well as their utility and limitations
- Understand how the different types of variables in a multivariate data set influence the visualization technique that should be chosen to represent the data
- Be able to apply any of these techniques to a data set that is an appropriate match for them
HW 3

• Due Friday

• Questions?

Upcoming

• Multivariate Visual Representations 2
  – Prep: Dust & Magnet video

• Lab 2: SVG