Housekeeping

- Second assignment due today
- Important Dates on Webpage
Agenda

• Data forms and representations
• Basic representation techniques
• Multivariate (>3) techniques

Data Sets

• Data comes in many different forms
• Typically, not in the way you want it

• How is stored (in the raw)?
Example

- Cars
  - make
  - model
  - year
  - miles per gallon
  - cost
  - number of cylinders
  - weights
  - ...

Data Tables

- Often, we take raw data and transform it into a form that is more workable
- Main idea:
  - Individual items are called cases
  - Cases have variables (attributes)
### Data Table Format

<table>
<thead>
<tr>
<th>Dimensions</th>
<th>Case\textsubscript{1}</th>
<th>Case\textsubscript{2}</th>
<th>Case\textsubscript{3}</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable\textsubscript{1}</td>
<td>Value\textsubscript{11}</td>
<td>Value\textsubscript{21}</td>
<td>Value\textsubscript{31}</td>
<td></td>
</tr>
<tr>
<td>Variable\textsubscript{2}</td>
<td>Value\textsubscript{12}</td>
<td>Value\textsubscript{22}</td>
<td>Value\textsubscript{32}</td>
<td></td>
</tr>
<tr>
<td>Variable\textsubscript{3}</td>
<td>Value\textsubscript{13}</td>
<td>Value\textsubscript{23}</td>
<td>Value\textsubscript{33}</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Think of as a function 
\( f(\text{case}_1) = \langle \text{Val}_{11}, \text{Val}_{12}, \ldots \rangle \)

### Example

<table>
<thead>
<tr>
<th>People in class</th>
<th>Mary</th>
<th>Jim</th>
<th>Sally</th>
<th>Mitch</th>
<th>...</th>
</tr>
</thead>
<tbody>
<tr>
<td>M-Nr.</td>
<td>145</td>
<td>294</td>
<td>563</td>
<td>823</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>23</td>
<td>17</td>
<td>47</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>Hair</td>
<td>brown</td>
<td>black</td>
<td>blonde</td>
<td>red</td>
<td></td>
</tr>
<tr>
<td>Grade</td>
<td>2.9</td>
<td>3.7</td>
<td>3.4</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Variable Types

- Three main types of variables
  - N-Nominal (equal or not equal to other values)
    Example: gender
  - O-Ordinal (obeys < relation, ordered set)
    Example: degrees - Bachelor, Master, PhD
  - Q-Quantitative (can do math on them)
    Example: age
**Metadata**

- Descriptive information about the data
  - Might be something as simple as the type of a variable, or could be more complex
  - For times when the table itself just isn’t enough
  - Example: if variable1 is “l”, then variable3 can only be 3, 7 or 16

**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 - Hypervariate data
Representation

• What’s a common way of visually representing multivariate data sets?
• Graphs! (not the vertex-edge ones)
Basic Symbolic Displays

- Graphs
- Charts
- Maps
- Diagrams

From:

1. Graph

Showing the relationships between variables’ values in a data table
Properties

- Graph
  - Visual display that illustrates one or more relationships among entities
  - Shorthand way to present information
  - Allows a trend, pattern or comparison to be easily comprehended

Issues

- Critical to remain task-centric
  - Why do you need a graph?
  - What questions are being answered?
  - What data is needed to answer those questions?
  - Who is the audience?
Graph Components

• Framework
  – Measurement types, scale
• Content
  – Marks, lines, points
• Labels
  – Title, axes, ticks

Other Symbolic Displays

• Chart
• Map
• Diagram
2. Chart

- Structure is important, relates entities to each other
- Primarily uses lines, enclosure, position to link entities

Examples: flowchart, family tree, org chart, ...

3. Map

- Representation of spatial relations
- Locations identified by labels
**Choropleth Map**

Areas are filled and colored differently to indicate some attribute of that region.

**Cartography**

- Cartographers and map-makers have a wealth of knowledge about the design and creation of visual information artifacts
  - Labeling, color, layout, ...
- Information visualization researchers should learn from this older, existing area.
4. Diagram

- Schematic picture of object or entity
- Parts are symbolic

Details

- What are the constituent pieces of these four symbolic displays?
- What are the building blocks?
Visual Structures

- Composed of
  - Spatial substrate
  - Marks
  - Graphical properties of marks

Space

- Visually dominant
- Often put axes on space to assist
- Use techniques of composition, alignment, folding, recursion, overloading to
  1) increase use of space
  2) do data encodings
**Marks**

- Things that occur in space
  - Points
  - Lines
  - Areas
  - Volumes

**Graphical Properties**

- Size, shape, color, orientation...

<table>
<thead>
<tr>
<th>Spatial properties</th>
<th>Object properties</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td>Grayscale</td>
</tr>
<tr>
<td>Size</td>
<td></td>
</tr>
<tr>
<td>Orientation</td>
<td>Color</td>
</tr>
<tr>
<td></td>
<td>Shape</td>
</tr>
<tr>
<td></td>
<td>Texture</td>
</tr>
</tbody>
</table>
**Back to Data**

- What were the different types of data sets?
- Number of variables per class
  - 1 - Univariate data
  - 2 - Bivariate data
  - 3 - Trivariate data
  - >3 - Hypervariate data

---

**Univariate Data**

- Representations

![Tukey box plot and bar chart](image)

- Bill
- Tukey box plot
- Mean
- Low
- Middle 50%
- High

---
What goes where

- In univariate representations, we often think of the data case as being shown along one dimension, and the value in another

<table>
<thead>
<tr>
<th>Line graph</th>
<th>Bar graph</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y-axis is quantitative variable</td>
<td></td>
</tr>
<tr>
<td>See changes over consecutive values</td>
<td></td>
</tr>
<tr>
<td>Y-axis is quantitative variable</td>
<td></td>
</tr>
<tr>
<td>Compare relative point values</td>
<td></td>
</tr>
</tbody>
</table>

Alternative View

- We may think of graph as representing independent (data case) and dependent (value) variables
- Guideline:
  - Independent vs. dependent variables
    - Put independent on x-axis
    - See resultant dependent variables along y-axis
**Bivariate Data**

- **Representations**

  Scatter plot is common

  ![Graph](image)

  Each mark is now a data case

  Is there a linear, curved or random pattern?

---

**Trivariate Data**

- **Representations**

  3D scatter plot is possible

  ![Graph](image)

  Summer term 2008
Still use 2D but have mark property represent third variable

Represent each variable in its own explicit way
**Hypervariate Data**

- Ahhh, the tough one
- Number of well-known visualization techniques exist for data sets of 1-3 dimensions
  - line graphs, bar graphs, scatter plots OK
  - We see a 3-D world (4-D with time)
- What about data sets with more than 3 variables?
  - Often the interesting, challenging ones

---

**Multiple Views**

Give each variable its own display

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

---
**Scatterplot Matrix**

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

Useful for what?
Misses what?

---

**Chernoff Faces**

Encode different variables’ values in characteristics of human face

Paper Recap

“Multidimensional Information Visualization Through Sliding Rods”
Tom Lanning, Kent Wittenburg, Micheal Heinrichs, Christina Fyock, Glenn Li
AVI 2000

Introduction

• Two types of interaction paradigms for Web Information Finding
  – Browsing
  – Query/Response
• Motivation for MultiNav
  – Easy to use techniques for multidimensional visualization
  – Integrate attribute info. with individual item browsing
MultiNav

Attributes as sliding rods
Sources Used

CMS book
Referenced articles
Marti Hearst SIMS 247 lectures
Kosslyn ‘89 article
A. Marcus, Graphic Design for Electronic Documents and User Interfaces
M. Monmonier, How to Lie with Maps
W. Cleveland, The Elements of Graphing Data
C. H. Yu, Visualization Techniques of Different Dimensions
http://seamonkey.ed.asu.edu/~behrens/asu/reports/compre/compl.html
http://www.csc.ncsu.edu/faculty/healey/PP/PP.html