IAT 355
Visual Analytics

Visual Encoding and Image Models

Lyn Bartram
Image models

• Representation
  • The visual, aural or haptic (i.e. sensory) encoding of the data
  • This is often termed *mapping*

• Presentation
  • Selection, layout and organisation of encoded data
  • May involve *multiple representations*

• Interaction
  • Manipulation to acquire different views of the data
Representation

- What’s a common way of visually representing multivariate data sets?
- Graphs! (not the vertex-edge ones)
- More accurately, symbolic display
Types of Symbolic Displays (Kosslyn 89)

- Graphs
- Charts
- Maps
- Diagrams
Types of Symbolic Displays

- **Graphs**
  - at least two scales required
  - values associated by a symmetric “paired with” relation
    - Examples: scatter-plot, bar-chart, layer-graph
Types of Symbolic Displays

• **Charts**
  • discrete relations among discrete entities
  • structure relates entities to one another
  • lines and relative position serve as links

• **Examples:**
  • Family tree
  • Flow chart
  • Network diagram

*Slide adapted from Marti Hearst*
Types of Symbolic Displays

- **Maps**
  - Internal relations determined (in part) by the spatial relations of what is pictured
  - Labels paired with locations

- **Examples:**
  - Map of census data
  - Topographic maps
Map

- Internal relations determined (in part) by the spatial relations of what is pictured
  - Grid: *geometric* metadata
- Locations identified by labels
- *Nominal* metadata

Examples:
- Map of census data
- Topographic maps
Types of Symbolic Displays

- **Diagrams**
  - Schematic pictures of objects or entities
  - Parts are symbolic (unlike photographs)
    - how-to illustrations
    - figures in a manual

What is the “real” taxonomy for visual representations?

- An empirical investigation by Lohse et al.’ 94
  - (Only used static, 2D graphics)
  - 16 participants
    - Half had a graphic design background
  - First, looked at 60 images and scored them along 10 scales.
    - These were used to compute statistical similarity

- Then, organized the 60 images into categories according to similarity.
  - Were asked to name the groups
  - Then they grouped these into higher-level groups, repeatedly, until they were in one large group.

Lohse, G L; Biolsi, K; Walker, N and H H Rueter, A Classification of Visual Representations, CACM, Vol. 37, No. 12, pp 36-49, 1994
## Scales that Participants Used
(and percentage of variance explained)

<table>
<thead>
<tr>
<th>Scale Description</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>emphasizes whole – parts</td>
<td>16.0</td>
</tr>
<tr>
<td>spatial – nonspatial</td>
<td>11.3</td>
</tr>
<tr>
<td>static structure – dynamic structure</td>
<td>10.6</td>
</tr>
<tr>
<td>continuous – discrete</td>
<td>10.5</td>
</tr>
<tr>
<td>attractive – unattractive</td>
<td>10.3</td>
</tr>
<tr>
<td>nontemporal – temporal</td>
<td>10.1</td>
</tr>
<tr>
<td>concrete – abstract</td>
<td>9.9</td>
</tr>
<tr>
<td>hard to understand – easy</td>
<td>9.6</td>
</tr>
<tr>
<td>nonnumeric – numeric</td>
<td>9.5</td>
</tr>
<tr>
<td>conveys a lot of info – conveys little</td>
<td>2.2</td>
</tr>
<tr>
<td>nontemporal – temporal</td>
<td></td>
</tr>
<tr>
<td>concrete – abstract</td>
<td></td>
</tr>
<tr>
<td>hard to understand – easy</td>
<td></td>
</tr>
<tr>
<td>nonnumeric – numeric</td>
<td></td>
</tr>
<tr>
<td>conveys a lot of info – conveys little</td>
<td></td>
</tr>
</tbody>
</table>
Resulting Categories (Lohse et al. 94)

- Graphs
- Tables (numerical)
- Tables (graphical)
- Charts (time)
- Charts (network)
- Diagrams (structure)
- Diagrams (network)
- Maps
- Cartograms
- Icons
- Photo-realistic images
Graphs

- Encode quantitative information using position and magnitude of geometric objects.
- Examples: scatter plots, bar charts.
Tables

- An arrangement of words, numbers, signs, or combinations of them to exhibit a set of facts or relationships in a compact fashion.
- Less abstract symbolic notation than graphs.
  - Graphical tables and numerical tables
Time Charts

- Display temporal data.
  - Gantt chart, time schedule.
Network Charts

- Show the relationships among components
- Symbols indicate the presence or absence of components.
- Correspondences are shown by lines, arrows, proximity, similarity, or containment.
  - Flow charts, org charts, pert charts, decision trees.

18. data model

36. organizational chart
Structure Diagrams

- A static description of a physical object.
- Spatial layout expresses true coordinate dimensions of the object.
  - Cross-sections
Process Diagrams

• Describe interrelationships and processes associated with physical objects.
• Spatial layout expresses dynamic, continuous, or temporal relationships among the objects.

41. nervous system
27. nitrogen cycle
54. equinox cycle
Maps

- Symbolic representations of physical geography.
  - Marine charts, topo maps, projections of world maps.
- Differ from cartograms in that cartograms super-impose quantitative data over a base map.
Cartograms

- Spatial maps that show quantitative data.
- Show more quantitative information than structure diagrams.
  - Chloropleths, dot maps, flow maps.
Icons

• Impart a single interpretation or meaning for a picture; a unique label for a visual representation.
Put into Multiple Categories

- No real agreement on these.
Where should these go?

- 56. star diagrams
- 60. diskette cross section
- 39. music
- 51. chess board
- 35. periodic table
- 29. bus routes
• How do you decide what kind of chart is best for what kind of data?

• Image models and visual language
• Semiology [Bertin] : an image is perceived as a set of signs and “retinal variables”
Visual encoding variables

- Position (x 2)
- Size
- Value
- Texture
- Color
- Orientation
- Shape
Visual encoding variables

- Position
- Length
- Area
- Volume
- Value
- Texture
- Color
- Orientation
- Shape
- Transparency
- Blur / Focus ……
Bertin’s Graphical Vocabulary

- **Position**
  - Points
  - Lines
  - Areas

- **Marks**
  - Points
  - Lines
  - Areas

- **Retinal variables**
  - Color
  - Size
  - Shape

- **Grayscale**
- **Orientation**
- **Texture**
Key visual encoding tasks/characteristics

1. Selection/Discrimination:
   • Is A different from B?

2. Association:
   • Are A and B similar (related in some way)?

3. Order
   • Is A > B?

4. Quantification: a number can be deduced from differences
   • How much bigger is A than B?
Visual characteristics

5. Capacity (length) [Carpendale]
   • The number of distinctions possible using the variable
   • How many different things can we represent with this variable?
Figure 3.41
Interpretation of Bertin’s guidance regarding the suitability of various encoding methods to support common tasks

- **Association**: The marks can be perceived as SIMILAR
- **Selection**: The marks are perceived as DIFFERENT, forming families
- **Order**: The marks are perceived as ORDERED
- **Quantity**: The marks are perceived as PROPORTIONAL to each other

**Visual Encoding**

- **Size**
- **Value**
- **Texture**
- **Colour**
- **Orientation**
- **Shape**
Interpretations of Graphical Vocabulary

Some properties can be discriminated more accurately but don’t have intrinsic meaning

(Senay & Ingatious 97, Kosslyn, others)

- **Density (Greyscale)**
  Darker -> More

- **Size / Length / Area**
  Larger -> More

- **Position**
  Leftmost -> first,
  Topmost -> first

- **Hue**
  no intrinsic meaning;
  good for highlighting

- **Slope / Shape**
  no intrinsic meaning;
  good for contrast

Visual Encoding | IAT 355
Accuracy Ranking of Quantitative Perceptual Tasks

Estimated; only pairwise comparisons have been validated.

(Mackinlay 88 from Cleveland & McGill)
Visual variables: selectivity

**Selectivity**: Different values are easily seen as different

“Is A different from B?”

Worst case: visual properties of all objects need to be looked at one by one
Visual variables: Associativity

- **Associativity**: Similar values can easily be grouped together
  
  “Is A similar to B?”

Positioning > {size, brightness} > {color, orientation (for points)} > texture > shape
Visual variables: Order

**Order:** Different values are perceived as ordered

"Is A more/greater/bigger than B?"

- Size and brightness are ordered
- Orientation, shape, texture are not ordered
- Hue is "not really" ordered
  - Some visual culture of progression
Visual variables: quantity

**Quantity:** A number can be deduced from differences

- “How much is the difference between A and B?”
  - Position is quantitative, size is somewhat quantitative
  - The other variables are not quantitative
Visual variables: capacity

**Length:** The number of distinctions possible using the variable

- “How many different things can we represent with this variable?”

- Shape, Texture: infinite, but …
- Brightness, hue: 7 (Association) – 10 (Distinction)
- Size: 5 (Association) -20 (Distinction)
- Orientation: 4
Figure 3.45 Mackinlay’s guidance for the encoding of quantitative, ordinal and categorical data
### Few’s Table:

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Quantitative</th>
<th>Qualitative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Line length</td>
<td>✜</td>
<td></td>
</tr>
<tr>
<td>2-D position</td>
<td>✜</td>
<td></td>
</tr>
<tr>
<td>Orientation</td>
<td>✜</td>
<td></td>
</tr>
<tr>
<td>Line width</td>
<td>✜</td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>✜</td>
<td></td>
</tr>
<tr>
<td>Shape</td>
<td>✜</td>
<td></td>
</tr>
<tr>
<td>Curvature</td>
<td>✜</td>
<td></td>
</tr>
<tr>
<td>Added marks</td>
<td>✜</td>
<td></td>
</tr>
<tr>
<td>Enclosure</td>
<td>✜</td>
<td></td>
</tr>
<tr>
<td>Hue</td>
<td>✜</td>
<td></td>
</tr>
<tr>
<td>Intensity</td>
<td>✜</td>
<td></td>
</tr>
</tbody>
</table>
Design Space of Visual Encodings
Choosing Visual Encodings

Principle of Consistency
• The properties of the image should match the properties of the data

Principle of Importance Ordering
• Encode the most important information in the most important way
Univariate Data

Tukey box plot

Jan 21, 2011

IAT 355
What goes where

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

Graph shows change in Y over continuous range X

Y Axis is quantitative

Graph shows value of Y for 4 cases

Y Axis is quantitative
Bivariate Data

• Representations
  • Scatter plot
  • Each mark is a data case
  • Want to see relationship between two variables
  • What is the pattern?
  • Note both variables are *continuous* data

![Scatter plot example](example.png)
Trivariate Data

- 3D scatter plot may work
  - Must have 3D cues
    - 3D blobs
    - motion parallax
    - Stereoscopy
  - But there are other coding dimensions
Three variables

- Use blob attribute for another variable

![Graph showing two variables: Price and Mileage.](image)

![Graph showing another variable: size, colour, shape, etc.](image)
Hypervariate Data

• 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

Each variable on its own line

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>6</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>9</td>
<td>12</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>8</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>8</td>
</tr>
</tbody>
</table>
Scatterplot Matrix

- Represent each possible pair of variables in their own 2-D scatterplot
- Useful for what?
- Misses what?
Using visual variables

• “Sameness of a visual element implies sameness of what the visual element represents” (Tufte, 2006)

• Characteristics of visual variables determine their use
  • e.g. Ordered values have to be represented by ordered visual variables

• Be consistent concerning relations of similarity, proportion and configuration

• Adhere to conventional uses of visual variables
  e.g. in cartography use blue color for water

• Scales should be made up of visually equidistant values of a variable
Using visual variables (2)

- The full range of a visual variable should be used
  - e.g. when using shades of gray, use from white to black

- The number of visual variables of a visualization should correspond to the dimensionality of the represented information
  - But sometimes dual encoding can be useful
Using visual variables

When combining two visual variables, if people should be able to analyze the two attributes independently, then separable variables should be used.

Perception!!!! Next week.