CIS 4930/6930-902
Scientific Visualization

Data Abstraction

Paul Rosen
Assistant Professor
University of South Florida

slides credits Miriah Meyer (U of Utah)
ADMINISTRATIVE

first assignment out today
check the project presentation schedule
LAST TIME . . .
Tufte’s Integrity Principles

Clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity. The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented. Show data variation, not design variation.
Tufte’s Design Principles

- maximize the data-ink ratio
- avoid chart junk (sometimes)
- use multifunctioning elements
- layer information
- maximize the data density
- shrink the graphics
- maximize the amount of data shown (sometimes)
TODAY...
DATA ABSTRACTION

the what part of an analysis that pertains to the data
translation of domain-specific terms into words that are as generic as possible
DATA TYPES

- Items
- Attributes
- Links
- Positions
- Grids
## DATASET TYPES

<table>
<thead>
<tr>
<th>Tables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items</td>
</tr>
<tr>
<td>Attributes</td>
</tr>
<tr>
<td>Order ID</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>32</td>
</tr>
<tr>
<td>32</td>
</tr>
<tr>
<td>32</td>
</tr>
<tr>
<td>35</td>
</tr>
<tr>
<td>35</td>
</tr>
<tr>
<td>36</td>
</tr>
<tr>
<td>65</td>
</tr>
<tr>
<td>66</td>
</tr>
<tr>
<td>69</td>
</tr>
<tr>
<td>70</td>
</tr>
<tr>
<td>70</td>
</tr>
<tr>
<td>96</td>
</tr>
<tr>
<td>97</td>
</tr>
<tr>
<td>129</td>
</tr>
<tr>
<td>130</td>
</tr>
<tr>
<td>130</td>
</tr>
<tr>
<td>132</td>
</tr>
<tr>
<td>132</td>
</tr>
<tr>
<td>134</td>
</tr>
<tr>
<td>135</td>
</tr>
<tr>
<td>166</td>
</tr>
<tr>
<td>193</td>
</tr>
<tr>
<td>194</td>
</tr>
</tbody>
</table>

**Field: attribute, item, cell**
DATASET TYPES

Tables

- Items
- Attributes

*Multidimensional Table*
## Dataset Types

<table>
<thead>
<tr>
<th>Tables</th>
<th>Networks &amp; Trees</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items</td>
<td>Items (nodes)</td>
</tr>
<tr>
<td>Attributes</td>
<td>Links</td>
</tr>
<tr>
<td></td>
<td>Attributes</td>
</tr>
</tbody>
</table>

- **Multidimensional Table**
  - Items (rows)
  - Attributes (columns)
  - Cell containing value

- **Trees**
  - Node (item)
  - Link

![Diagram of a multidimensional table and a tree structure](image)
DATASET TYPES

Tables
- Items
- Attributes

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

Fields
- Grids
- Positions
- Attributes

→ Multidimensional Table
→ Trees
→ Grid of positions
GRID TYPES

uniform
rectilinear
structured unstructured
GRID CHOICES IMPACT HOW CONTINUOUS DATA IS INTERPRETED

two key considerations:

sampling, or the choice of where attributes are measured

interpolation, or how to model the attributes in the rest of space
DATASET TYPES

- **Tables**
  - Items
  - Attributes

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

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

- **Multidimensional Table**
- **Trees**
- **Grid of positions**
## Dataset Types

<table>
<thead>
<tr>
<th>Tables</th>
<th>Networks &amp; Trees</th>
<th>Fields</th>
</tr>
</thead>
<tbody>
<tr>
<td>Items</td>
<td>Items (nodes)</td>
<td>Grids</td>
</tr>
<tr>
<td>Attributes</td>
<td>Links</td>
<td>Positions</td>
</tr>
</tbody>
</table>

### Examples

- **Multidimensional Table**
  - Items (rows)
  - Attributes (columns)
  - Cell containing value

- **Trees**
  - Nodes (items)
  - Links

- **Grid of positions**
  - Attributes (columns)
  - Value in cell
DATASET TYPES

- **Tables**
  - Items
  - Attributes

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

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

**Vector**
- Multidimensional Table
- Trees
  - Link
  - Node (item)
  - Grid of positions

**Scalar**
- Attributes (columns)
- Cell containing value
- Items (rows)
  - Key
  - Value in cell
  - Value in cell
DATASET TYPES

- **Tables**
  - Items
  - Attributes

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

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

**Multidimensional Table**

**Trees**

**Grid of positions**

**Scalar**

**Vector**

**Tensor**
### DATASET TYPES

**Tables**
- Items
- Attributes

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

**Fields**
- Grids
- Positions
- Attributes

**Geometry**
- Items
- Positions

- **Multidimensional Table**
- **Trees**
- **Grid of positions**
- **Position**

**Key**
- Key 1
- Value in cell
- Attributes

**Cell**
- Node (item)
- Link
DATASET TYPES

- **Tables**
  - Items
  - Attributes

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

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

- **Geometry**
  - Items
  - Positions

- **Clusters, Sets, Lists**
  - Items

> Multidimensional Table

> Trees
ATTRIBUTE TYPES

Categorical
no implicit ordering
ATTRIBUTE TYPES

Categorical
no implicit ordering
ATTRIBUTE TYPES

Categorical
no implicit ordering

Ordered

Ordinal

Quantitative
ATTRIBUTE TYPES

Categorical
no implicit ordering

Ordered

Ordinal
meaningful magnitude
(can do arithmetic)

Quantitative
ATTRIBUTE TYPES

Categorical
no implicit ordering

+  ⬤  □  △

Ordered

Ordinal

Quantitative
meaningful magnitude (can do arithmetic)

★ ★ ★ ★ ★
★ ★ ★ ★ ★
★ ★ ★ ★ ★
★ ★ ★ ★ ★
★ ★ ★ ★ ★
ATTRIBUTE TYPES

Categorical
no implicit ordering

Ordered

Quantitative
meaningful magnitude
(can do arithmetic)
<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>S</th>
<th>T</th>
<th>U</th>
</tr>
</thead>
<tbody>
<tr>
<td>Order ID</td>
<td>Order Date</td>
<td>Order Priority</td>
<td>Product Container</td>
<td>Product Base Margin</td>
<td>Ship Date</td>
</tr>
<tr>
<td>3</td>
<td>10/14/06</td>
<td>5-Low</td>
<td>Large Box</td>
<td>0.8</td>
<td>10/21/06</td>
</tr>
<tr>
<td>6</td>
<td>2/21/08</td>
<td>4-Not Specified</td>
<td>Small Pack</td>
<td>0.55</td>
<td>2/22/08</td>
</tr>
<tr>
<td>32</td>
<td>7/16/07</td>
<td>2-High</td>
<td>Small Pack</td>
<td>0.79</td>
<td>7/17/07</td>
</tr>
<tr>
<td>32</td>
<td>7/16/07</td>
<td>2-High</td>
<td>Jumbo Box</td>
<td>0.72</td>
<td>7/17/07</td>
</tr>
<tr>
<td>32</td>
<td>7/16/07</td>
<td>2-High</td>
<td>Medium Box</td>
<td>0.6</td>
<td>7/18/07</td>
</tr>
<tr>
<td>32</td>
<td>7/16/07</td>
<td>2-High</td>
<td>Medium Box</td>
<td>0.65</td>
<td>7/18/07</td>
</tr>
<tr>
<td>35</td>
<td>10/23/07</td>
<td>4-Not Specified</td>
<td>Wrap Bag</td>
<td>0.52</td>
<td>10/24/07</td>
</tr>
<tr>
<td>35</td>
<td>10/23/07</td>
<td>4-Not Specified</td>
<td>Small Box</td>
<td>0.58</td>
<td>10/25/07</td>
</tr>
<tr>
<td>36</td>
<td>11/3/07</td>
<td>1-Urgent</td>
<td>Small Box</td>
<td>0.55</td>
<td>11/3/07</td>
</tr>
<tr>
<td>65</td>
<td>3/18/07</td>
<td>1-Urgent</td>
<td>Small Pack</td>
<td>0.49</td>
<td>3/19/07</td>
</tr>
<tr>
<td>66</td>
<td>1/20/05</td>
<td>5-Low</td>
<td>Wrap Bag</td>
<td>0.56</td>
<td>1/20/05</td>
</tr>
<tr>
<td>69</td>
<td>6/4/05</td>
<td>4-Not Specified</td>
<td>Small Pack</td>
<td>0.44</td>
<td>6/6/05</td>
</tr>
<tr>
<td>69</td>
<td>6/4/05</td>
<td>4-Not Specified</td>
<td>Small Pack</td>
<td>0.44</td>
<td>6/6/05</td>
</tr>
<tr>
<td>70</td>
<td>12/18/06</td>
<td>5-Low</td>
<td>Small Box</td>
<td>0.59</td>
<td>12/23/06</td>
</tr>
<tr>
<td>70</td>
<td>12/18/06</td>
<td>5-Low</td>
<td>Small Box</td>
<td>0.82</td>
<td>12/23/06</td>
</tr>
<tr>
<td>96</td>
<td>4/17/05</td>
<td>2-High</td>
<td>Small Box</td>
<td>0.55</td>
<td>4/19/05</td>
</tr>
<tr>
<td>97</td>
<td>1/29/06</td>
<td>3-Medium</td>
<td>Small Box</td>
<td>0.38</td>
<td>1/30/06</td>
</tr>
<tr>
<td>129</td>
<td>11/19/08</td>
<td>5-Medium</td>
<td>Small Box</td>
<td>0.37</td>
<td>11/28/08</td>
</tr>
<tr>
<td>130</td>
<td>5/8/08</td>
<td>2-High</td>
<td>Medium Box</td>
<td>0.37</td>
<td>5/9/08</td>
</tr>
<tr>
<td>130</td>
<td>5/8/08</td>
<td>2-High</td>
<td>Small Box</td>
<td>0.38</td>
<td>5/10/08</td>
</tr>
<tr>
<td>130</td>
<td>5/8/08</td>
<td>2-High</td>
<td>Small Box</td>
<td>0.6</td>
<td>5/10/08</td>
</tr>
<tr>
<td>132</td>
<td>6/11/06</td>
<td>3-Medium</td>
<td>Medium Box</td>
<td>0.6</td>
<td>6/12/06</td>
</tr>
<tr>
<td>132</td>
<td>6/11/06</td>
<td>3-Medium</td>
<td>Medium Box</td>
<td>0.69</td>
<td>6/14/06</td>
</tr>
<tr>
<td>134</td>
<td>5/1/08</td>
<td>4-Not Specified</td>
<td>Large Box</td>
<td>0.82</td>
<td>5/3/08</td>
</tr>
<tr>
<td>134</td>
<td>5/1/08</td>
<td>4-Not Specified</td>
<td>Large Box</td>
<td>0.82</td>
<td>5/3/08</td>
</tr>
<tr>
<td>135</td>
<td>10/21/07</td>
<td>4-Not Specified</td>
<td>Small Pack</td>
<td>0.64</td>
<td>10/23/07</td>
</tr>
<tr>
<td>166</td>
<td>9/12/07</td>
<td>2-High</td>
<td>Small Box</td>
<td>0.55</td>
<td>9/14/07</td>
</tr>
<tr>
<td>193</td>
<td>8/8/06</td>
<td>1-Urgent</td>
<td>Medium Box</td>
<td>0.57</td>
<td>8/10/06</td>
</tr>
<tr>
<td>194</td>
<td>4/5/08</td>
<td>3-Medium</td>
<td>Wrap Bag</td>
<td>0.42</td>
<td>4/7/08</td>
</tr>
<tr>
<td>Order ID</td>
<td>Order Date</td>
<td>Order Priority</td>
<td>Product Container</td>
<td>Product Base Margin</td>
<td>Ship Date</td>
</tr>
<tr>
<td>----------</td>
<td>------------</td>
<td>----------------</td>
<td>-------------------</td>
<td>---------------------</td>
<td>-----------</td>
</tr>
<tr>
<td>3</td>
<td>10/14/06</td>
<td>5-Low</td>
<td>Large Box</td>
<td>0.8</td>
<td>10/21/06</td>
</tr>
<tr>
<td>6</td>
<td>2/21/08</td>
<td>4-Not Specified</td>
<td>Small Pack</td>
<td>0.55</td>
<td>2/22/08</td>
</tr>
<tr>
<td>32</td>
<td>7/16/07</td>
<td>2-High</td>
<td>Small Pack</td>
<td>0.79</td>
<td>7/17/07</td>
</tr>
<tr>
<td>32</td>
<td>7/16/07</td>
<td>2-High</td>
<td>Jumbo Box</td>
<td>0.72</td>
<td>7/17/07</td>
</tr>
<tr>
<td>32</td>
<td>7/16/07</td>
<td>2-High</td>
<td>Medium Box</td>
<td>0.6</td>
<td>7/18/07</td>
</tr>
<tr>
<td>32</td>
<td>7/16/07</td>
<td>2-High</td>
<td>Medium Box</td>
<td>0.65</td>
<td>7/18/07</td>
</tr>
<tr>
<td>35</td>
<td>10/23/07</td>
<td>4-Not Specified</td>
<td>Wrap Bag</td>
<td>0.52</td>
<td>10/24/07</td>
</tr>
<tr>
<td>35</td>
<td>10/23/07</td>
<td>4-Not Specified</td>
<td>Small Box</td>
<td>0.58</td>
<td>10/25/07</td>
</tr>
<tr>
<td>36</td>
<td>11/3/07</td>
<td>1-Urgent</td>
<td>Small Box</td>
<td>0.55</td>
<td>11/3/07</td>
</tr>
<tr>
<td>65</td>
<td>3/18/07</td>
<td>1-Urgent</td>
<td>Small Pack</td>
<td>0.49</td>
<td>3/19/07</td>
</tr>
<tr>
<td>66</td>
<td>1/20/05</td>
<td>5-Low</td>
<td>Wrap Bag</td>
<td>0.56</td>
<td>1/20/05</td>
</tr>
<tr>
<td>69</td>
<td>6/4/05</td>
<td>4-Not Specified</td>
<td>Small Pack</td>
<td>0.44</td>
<td>6/6/05</td>
</tr>
<tr>
<td>69</td>
<td>6/4/05</td>
<td>4-Not Specified</td>
<td>Small Box</td>
<td>0.6</td>
<td>6/6/05</td>
</tr>
<tr>
<td>70</td>
<td>12/18/06</td>
<td>5-Low</td>
<td></td>
<td>0.59</td>
<td>12/23/06</td>
</tr>
<tr>
<td>70</td>
<td>12/18/06</td>
<td>5-Low</td>
<td></td>
<td>0.82</td>
<td>12/23/06</td>
</tr>
<tr>
<td>96</td>
<td>4/17/05</td>
<td>2-High</td>
<td></td>
<td>0.55</td>
<td>4/19/05</td>
</tr>
<tr>
<td>97</td>
<td>1/29/06</td>
<td>3-Medium</td>
<td></td>
<td>0.38</td>
<td>1/30/06</td>
</tr>
<tr>
<td>129</td>
<td>11/19/08</td>
<td>5-Low</td>
<td></td>
<td>0.37</td>
<td>11/28/08</td>
</tr>
<tr>
<td>130</td>
<td>5/8/08</td>
<td>2-High</td>
<td>Small Box</td>
<td>0.37</td>
<td>5/9/08</td>
</tr>
<tr>
<td>130</td>
<td>5/8/08</td>
<td>2-High</td>
<td>Medium Box</td>
<td>0.38</td>
<td>5/10/08</td>
</tr>
<tr>
<td>130</td>
<td>5/8/08</td>
<td>2-High</td>
<td>Small Box</td>
<td>0.6</td>
<td>5/11/08</td>
</tr>
<tr>
<td>132</td>
<td>6/11/06</td>
<td>3-Medium</td>
<td>Medium Box</td>
<td>0.6</td>
<td>6/12/06</td>
</tr>
<tr>
<td>132</td>
<td>6/11/06</td>
<td>3-Medium</td>
<td>Jumbo Box</td>
<td>0.69</td>
<td>6/14/06</td>
</tr>
<tr>
<td>134</td>
<td>5/11/08</td>
<td>4-Not Specified</td>
<td>Large Box</td>
<td>0.82</td>
<td>5/3/08</td>
</tr>
<tr>
<td>135</td>
<td>10/21/07</td>
<td>4-Not Specified</td>
<td>Small Pack</td>
<td>0.64</td>
<td>10/23/07</td>
</tr>
<tr>
<td>166</td>
<td>9/12/07</td>
<td>2-High</td>
<td>Small Box</td>
<td>0.55</td>
<td>9/14/07</td>
</tr>
<tr>
<td>193</td>
<td>8/8/06</td>
<td>1-Urgent</td>
<td>Medium Box</td>
<td>0.57</td>
<td>8/10/06</td>
</tr>
<tr>
<td>194</td>
<td>4/5/08</td>
<td>3-Medium</td>
<td>Wrap Bag</td>
<td>0.42</td>
<td>4/7/08</td>
</tr>
<tr>
<td>219</td>
<td>1/15/08</td>
<td>1-Urgent</td>
<td>Small Pack</td>
<td>0.6</td>
<td>1/16/08</td>
</tr>
<tr>
<td>222</td>
<td>11/18/07</td>
<td>3-Medium</td>
<td>Medium Box</td>
<td>0.6</td>
<td>11/19/07</td>
</tr>
</tbody>
</table>
ATTRIBUTE TYPES

Categorical
no implicit ordering

+  ●  □  △

Hierarchical

Ordered

Ordinal

Quantitative
meaningful magnitude (can do arithmetic)

Sequential

Divergent

Cyclic
DERIVED ATTRIBUTES

derived attribute: computed from originals

- simple change of type
- acquire additional data
- complex transformation

transformation is abstraction choice

original data

\[ f(x), g(x) \]

derived data

\[ h(x) = f(x) - g(x) \]
DATA MODEL vs CONCEPTUAL MODEL

data model: mathematical abstraction (data abstraction)
    set with operations, e.g. floats with * / - +

conceptual model: mental construction (semantics)
    includes semantics, supports reasoning

conceptual model motivates data abstraction choices
EXAMPLE

from data model . . .
- 32.52, 54.06, -17.35, . . . (floats)

using conceptual model . . .
temperature

to new data abstraction...
continuous to 2 significant figures (quantitative)
hot, warm, cold (ordinal)
above freezing, below freezing (categorical)
ABSTRACTION EXERCISE . . .
PROCESS AND PITFALLS IN WRITING INFORMATION VISUALIZATION RESEARCH PAPERS

Discuss 1-2 examples of what each paper type from Section 2 might “look like”.

Discuss the pitfalls from Section 3. Which might you have been likely to commit before this course?