From Data To Image

Pat Hanrahan

Topics

The properties of the data or information
The properties of the image
The rules mapping data to images

Bertin 101
The Data

Taxonomy by Data Type

- 1D (sets and sequences)
- Temporal
- 2D (maps)
- 3D (shapes)
- nD (relational)
- Trees (hierarchies)
- Networks (graphs)
- Text and documents [mine]

B. Schneiderman, The eyes have it: A task by data type taxonomy for information visualization, 1996
Data Models vs. Conceptual Models

Data models are mathematical abstractions
- Sets with operations on them
  For example, integers with + and × operators

Conceptual models are mental constructions
- Include semantics and support reasoning
  For example, navigating through a city using landmarks

Examples (data vs. conceptual):
- 1D vs. Time
- nD vs. Space

Types of Data Models

Discrete
- Relations
- Topology

Continuous
- Fields*
- Manifolds

* Treinish, A function-based datamodel for visualization
Relational Data Model

Records are fixed-length tuples
Each column of a tuple has a domain (type)
Relation is a schema plus a table of tuples
Database is a collection of relations

Example: Digital cameras

Relational Algebra [Codd]

Data transformations (SQL)
- Selection (SELECT)
- Projection (WHERE)
- Sorting (ORDER BY)
- Aggregation (GROUP BY, SUM, MIN, ...)
- Set operations (UNION, ...)
- Join (INNER JOIN)
Statistical Data Model

Variables or Measurements
Categories or Factors
Observations or Cases

<table>
<thead>
<tr>
<th>Month</th>
<th>Control</th>
<th>Placebo</th>
<th>300 mg</th>
<th>450 mg</th>
</tr>
</thead>
<tbody>
<tr>
<td>March</td>
<td>165</td>
<td>163</td>
<td>166</td>
<td>168</td>
</tr>
<tr>
<td>April</td>
<td>162</td>
<td>159</td>
<td>161</td>
<td>163</td>
</tr>
<tr>
<td>May</td>
<td>164</td>
<td>158</td>
<td>161</td>
<td>153</td>
</tr>
<tr>
<td>June</td>
<td>162</td>
<td>161</td>
<td>158</td>
<td>160</td>
</tr>
<tr>
<td>July</td>
<td>166</td>
<td>158</td>
<td>160</td>
<td>148</td>
</tr>
<tr>
<td>August</td>
<td>163</td>
<td>158</td>
<td>157</td>
<td>150</td>
</tr>
</tbody>
</table>

Blood Pressure Study (4 treatments, 6 months)

Sepal and petal lengths and widths for three species of Iris (Fisher).
Sepal and petal lengths and widths for three species of iris (Fisher).

<table>
<thead>
<tr>
<th>ID</th>
<th>Case</th>
<th>Species</th>
<th>Organ</th>
<th>Width</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>1</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>11</td>
<td>10</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>12</td>
<td>11</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>13</td>
<td>12</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>14</td>
<td>13</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>15</td>
<td>14</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>16</td>
<td>15</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>17</td>
<td>16</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>18</td>
<td>17</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>19</td>
<td>18</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>20</td>
<td>19</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>21</td>
<td>20</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>22</td>
<td>21</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>23</td>
<td>22</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>24</td>
<td>23</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>25</td>
<td>24</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>26</td>
<td>25</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>27</td>
<td>26</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>28</td>
<td>27</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>29</td>
<td>28</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>30</td>
<td>29</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>31</td>
<td>30</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>32</td>
<td>31</td>
<td>Setosa</td>
<td>Petal</td>
<td>2</td>
<td>14</td>
</tr>
</tbody>
</table>

Sepal and petal lengths and widths for three species of iris (Fisher).
### Types

#### Physical types
- Characterized by storage
- Characterized by machine operations

**Example:**
- `bool`, `short`, `int32`, `float`, `double`, `string`, ...

#### Abstract types
- Characterized by methods/attributes
- Organized into a class hierarchy

**Example:**
- `nominal`, `ordinal`, `cardinal`, ..., `plants`, `animals`, `metazoans`, ...

---

**Format of the data in Appendix 14, pp. 365-366**

Chambers, Cleveland, Kleiner, Tukey, *Graphical Methods for Data Analysis*
Measurements

N - Nominal (labels or types)
- Fruits: Apples, oranges, ...

O - Ordered
- Days: Mon, Tue, Wed, Thu, Fri, Sat, Sun
- Quality of meat: Grade A, AA, AAA

Q - Interval (Location of 0 arbitrary)
- Periods of time: second, minute, ...

Q - Ratio (0 fixed)
- Counts
- Physical measurement: Kelvin, L, M, R, ...

S. S. Stevens, On the theory of scales of measurements, 1946
Dimensions and Measures

Independent vs. dependent variables
Example: $y = f(x, a)$
Infer causality
- Response $\sim$ factors
- Functional dependency in databases [Ullman]

Extrinsic vs. intrinsic variables
Example: mass vs. density (mass/vol)
Summarize
- Groupby dimensions and aggregate measures

Data Cube

Measure
Width
Length
Species
I. setosa
I. versicolor
I. virginica
Petal
Sepal
Organ
Summary of Basic Properties

- Multidimensional
  - Number of columns
- Type
  - Type of column (N, O, Q)
- Cardinality (levels)
  - Number of different column values

The Image
Image Information

Graphical primitives and attributes (Marks)
Attributes are parameters that control the appearance of geometric primitives

Visual channels
Separable channels of information flowing from the retina to the brain

Visual Language is a Sign System

Image is perceived as a set of signs

Sender encodes information in these signs

Receiver decodes information from these signs
8 Visual Variables

J. Bertin, Semiology of Graphics, 1967

- Position
- Size
- Value
- Color
- Texture
- Orientation
- Shape

Note: Bertin does not consider 3D or time
Note: Card and Mackinlay extend the number of vars.
Information in Position

1. A, B, C are distinguishable
2. B is between A and C.
3. BC is twice as long as AB.

"Resemblance, order and proportional are the three signfields in graphics." - Bertin

Information in Color and Value

Value is perceived as ordered

∴ Encode ordinal variables (O)

∴ Encode continuous variables (Q) [not as well]

Hue is normally perceived as unordered

∴ Encode nominal variables (N) using color
Bertins’ “Levels of Organization”

<table>
<thead>
<tr>
<th>Attribute</th>
<th>N</th>
<th>O</th>
<th>Q</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td>N</td>
<td>O</td>
<td>Q</td>
</tr>
<tr>
<td>Size</td>
<td>N</td>
<td>O</td>
<td>Q</td>
</tr>
<tr>
<td>Value</td>
<td>N</td>
<td>O</td>
<td>Q</td>
</tr>
<tr>
<td>Texture</td>
<td>N</td>
<td>O</td>
<td></td>
</tr>
<tr>
<td>Color</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Orientation</td>
<td>N</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shape</td>
<td>N</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

N Nominal  
O Ordered  
Q Quantitative

Note: Bertin actually breaks visual variables down into differentiating (≠) and associating (≡)

The Rules
Design Space = Visual Metaphors

Bertin’s Specification

[ Bertin, Semiology, 1967 ]
Fields Create Tables and Graphs

Ordinal fields: interpret field as a sequence that partitions table into rows and columns:
Quarter = \{(Qtr1),(Qtr2),(Qtr3),(Qtr4)\} →

<table>
<thead>
<tr>
<th>Qtr1</th>
<th>Qtr2</th>
<th>Qtr3</th>
<th>Qtr4</th>
</tr>
</thead>
<tbody>
<tr>
<td>95092</td>
<td>501760</td>
<td>105282</td>
<td>98225</td>
</tr>
</tbody>
</table>

Quantitative fields: treat field as single element sequence and encode as an axis:
Profit = \{\{Profit\}\} →

<p>| | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>⋯</td>
<td>⋯</td>
<td>⋯</td>
<td>⋯</td>
<td>⋯</td>
<td>⋯</td>
<td>⋯</td>
<td>⋯</td>
</tr>
<tr>
<td>⋯</td>
<td>⋯</td>
<td>300</td>
<td>400</td>
<td>500</td>
<td>600</td>
<td>⋯</td>
<td>⋯</td>
</tr>
<tr>
<td>0</td>
<td>-100</td>
<td>-200</td>
<td>-300</td>
<td>⋯</td>
<td>⋯</td>
<td>⋯</td>
<td>⋯</td>
</tr>
</tbody>
</table>

Profit
**Union (+) Operator**

**Quarter + ProductType**

\[ \{(Qtr1),(Qtr2),(Qtr3),(Qtr4)\} + \{(Coffee),(Espresso)\} = \{(Qtr1),(Qtr2),(Qtr3),(Qtr4),(Coffee),(Espresso)\} \]

<table>
<thead>
<tr>
<th>Qtr1</th>
<th>Qtr2</th>
<th>Qtr3</th>
<th>Qtr4</th>
<th>Coffee</th>
<th>Espresso</th>
</tr>
</thead>
<tbody>
<tr>
<td>48</td>
<td>59</td>
<td>57</td>
<td>55</td>
<td>151</td>
<td>21</td>
</tr>
</tbody>
</table>

**Profit + Sales**

\[ \{(Profit),(Sales)\} \]

**Cross (×) Operator**

**Quarter × ProductType**

\[ \{(Qtr1, Coffee), (Qtr1, Tea), (Qtr2, Coffee), (Qtr2, Tea), (Qtr3, Coffee), (Qtr3, Tea), (Qtr4, Coffee), (Qtr4, Tea)\} \]

<table>
<thead>
<tr>
<th>Qtr1</th>
<th>Qtr2</th>
<th>Qtr3</th>
<th>Qtr4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coffee</td>
<td>Espresso</td>
<td>Coffee</td>
<td>Espresso</td>
</tr>
<tr>
<td>131</td>
<td>19</td>
<td>160</td>
<td>20</td>
</tr>
</tbody>
</table>

**ProductType × Profit = \{(Coffee, Profit), (Tea, Profit)\}**
Combinatorics of Encodings

Challenge:
Pick the best encoding from the exponential number of possibilities $(n+1)^8$

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 effective way.

Mackinlay’s Expressiveness Criteria

Expressiveness
A set of facts is expressible in a visual language if the sentences (i.e. the visualizations) in the language express all the facts in the set of data, and only the facts in the data.
Cannot Express the Facts

A 1 → N relation cannot be expressed in a single horizontal dot plot because multiple tuples are mapped to the same position.

Expresses Facts Not in the Data

A length is interpreted as a quantitative value; ∴ Length says something untrue about N data.

---

Fig 11. Incorrect use of a bar chart for the Nation relation. The lengths of the bars suggest an ordering on the vertical axis, so if the USA cars were longer or better than the other cars, which is not true for the Nation relation.

[Mackinlay, APT, 1986]
Mackinlay’s Criteria 2

Effectiveness

A visualization is more effective than another visualization if the information conveyed by one visualization is more readily perceived than the information in the other visualization.

The subject of the next lecture.

Summary

Formal approach to picture specification
- Declare the picture you want to see
- Compile query, analysis, and rendering commands needed to make the picture
- Automatically generate presentations by searching over the space of designs

Bertin’s vision still not complete
- Formalize data model
- Formalize the specifications
- Experimentally test perceptual assumptions

Much more research to be done in this area ...