Data and Image Models

Maneesh Agrawala

CS 294-10: Visualization
Spring 2010

Last Time: The Purpose of Visualization

Three functions of visualizations

Record information
- Photographs, blueprints, …

Support reasoning about information (analyze)
- Process and calculate
- Reason about data
- Feedback and interaction

Convey information to others (present)
- Share and persuade
- Collaborate and revise
- Emphasize important aspects of data

Record information

Gallop, Bay Horse “Daisy” [Muybridge 1884-86]
Analysis: Challenger

Visualizations drawn by Tufte show how low temperatures damage O-rings [Tufte 97]

Communicate: Exports and Imports

[Playfair 1786]

Announcements

Auditors, please enroll in the class (1 unit, P/NP)
- Requirements: Come to class and participate (online as well)
- Requirements: Assignment 1 and 2

Class participation requirements
- Complete readings before class
- In-class discussion
- Post at least 1 discussion substantive comment/question on wiki within a week of each lecture

All, add yourself to participants page on the wiki

Class wiki
http://vis.berkeley.edu/courses/cs294-10-sp10Wiki/

Assignment 1: Good and Bad Vis.

Find two visualizations one good and one bad

Use original sources
- Journals
- Science magazines
- Newspapers
- Textbooks

Make wiki page
- Clearly mark as good or bad
- Provide short explanation
- Be prepared to succinctly describe in class on Mon Sep. 8

Due before class today
Assignment 2: Visualization Design
Due before class on Feb 8, 2010

Data and Image Models

The big picture

- task
- data
  - physical type
    - int, float, etc.
    - abstract type
      - nominal, ordinal, etc.
- domain
  - metadata
  - semantics
  - conceptual model

Processing

- algorithms

Image

- visual channel
- retinal variables

Mapping

- visual encoding
- visual metaphor

Topics

- Properties of data or information
- Properties of the image
- Mapping data to images

[based on slide from Munzner]
Data models vs. Conceptual models

Data models are low level descriptions of the data
- Math: Sets with operations on them
  - Example: integers with + and \times operators

Conceptual models are mental constructions
- Include semantics and support reasoning

Examples (data vs. conceptual)
- (1D floats) vs. Temperature
- (3D vector of floats) vs. Space

Taxonomy
- 1D (sets and sequences)
- Temporal
- 2D (maps)
- 3D (shapes)
- nD (relational)
- Trees (hierarchies)
- Networks (graphs)

Are there others?  

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

Example:  
- bool, short, int32, float, double, string, …

Abstract types
- Provide descriptions of the data
- May be characterized by methods/attributes
- May be organized into a hierarchy

Example:  
- plants, animals, metazoans, …
Nominal, ordinal and quantitative

N - Nominal (labels)
- Fruits: Apples, oranges, …

O - Ordered
- Quality of meat: Grade A, AA, AAA

Q - Interval (Location of zero arbitrary)
- Dates: Jan, 19, 2006; Location: (LAT 33.98, LONG -118.45)
- Like a geometric point. Cannot compare directly
- Only differences (i.e. intervals) may be compared

Q - Ratio (zero fixed)
- Physical measurement: Length, Mass, Temp, …
- Counts and amounts
- Like a geometric vector, origin is meaningful

S. S. Stevens, On the theory of scales of measurements, 1946

From data model to N,O,Q data type

Data model
- 32.5, 54.0, -17.3, …
- floats

Conceptual model
- Temperature

Data type
- Burned vs. Not burned (N)
- Hot, warm, cold (O)
- Continuous range of values (Q)

[based on slide from Munzner]
### Relational data model

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

### 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 or dimensions
- Observations or cases
Statistical data model

Variables or measurements
Categories or factors or dimensions
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)

Dimensions and measures

Independent vs. dependent variables
- Example: \( y = f(x,a) \)
- Dimensions: Domain(x) x Domain(a)
- Measures: Range(y)

Dimensions:
- Discrete variables describing data
dates, categories of values (independent vars.)

Measures:
- Data values that can be aggregated
numbers to be analyzed (dependent vars)

Aggregations:
- sum, count, average, std. dev.

Data cube

Measure
Width
Length
Species

I. setosa
I. versicolor
I. virginica

Petal
Sepal
Organ
Projections summarize data

Multiscale visualization using data cubes [Stolte et al. 02]

Visual language is a sign system

Images perceived as a set of signs
Sender encodes information in signs
Receiver decodes information from signs

Semiology of Graphics, 1983

Information in position

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

:. Encode quantitative variables

"Resemblance, order and proportional are the three signfields in graphics." - Bertin
Note: Bertin does not consider 3D or time
Note: Card and Mackinlay extend the number of vars.

Information in color and value

Value is perceived as ordered
- Encode ordinal variables (O)

Hue is normally perceived as unordered
- Encode nominal variables (N) using color

Bertins’ “Levels of Organization”

<table>
<thead>
<tr>
<th>Variable</th>
<th>N Nominal</th>
<th>O Ordered</th>
<th>Q Quantitative</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>Q</td>
</tr>
<tr>
<td>Color</td>
<td>N</td>
<td>Q</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>

Note: Bertin actually breaks visual variables down into differentiating (↓) and associating (→)
Encoding rules

Univariate data

A B C

1

Variable

Bivariate data

A B C

1 2

Scatter plot is common
Trivariate data

A B C

1 2 3

3D scatter plot is possible

Three variables

Two variables [x,y] can map to points
- Scatterplots, maps, …

Third variable [z] must use …
- Color, size, shape, …

Large design space (visual metaphors)

[ Bertin, Graphics and Graphic Information Processing, 1981 ]

Multidimensional data

How many variables can be depicted in an image?

[ A B C ]
Multidimensional data

How many variables can be depicted in an image?

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

"With up to three rows, a data table can be constructed directly as a single image.... However, an image has only three dimensions. And this barrier is impassible."

Bertin

Deconstructions

Stock chart from the late 90s

- x-axis: time (Q)
- y-axis: price (Q)
Playfair 1786

- x-axis: year (Q)
- y-axis: currency (Q)
- color: imports/exports (N, O)

Wattenberg 1998

- rectangle size: market cap (Q)
- rectangle position: market sector (N), market cap (Q)
- color hue: loss vs. gain (N, O)
- color value: magnitude of loss or gain (Q)

http://www.smartmoney.com/marketmap/
Minard 1869: Napoleon’s march

Single axis composition

Mark composition

y-axis: temperature (Q)

x-axis: time (Q)

= temp over time (Q x Q)

Mark composition

y-axis: longitude (Q)

x-axis: latitude (Q)

width: army size (Q)

= army position (Q x Q) and army size (Q)
Minard 1869: Napoleon’s march

Depicts at least 5 quantitative variables
Any others?

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 (visual variables) 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 one-to-many (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; therefore, the length of a bar says something untrue about N data.

---

**Mackinlay’s effectiveness criteria**

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.

---

**Subject of perception lecture**

[Mackinlay, APT, 1986]
Mackinlay’s ranking

<table>
<thead>
<tr>
<th>Quantitative</th>
<th>Ordinal</th>
<th>Nominal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Position</td>
<td>Position</td>
<td>Position</td>
</tr>
<tr>
<td>Length</td>
<td>Density</td>
<td>Hue</td>
</tr>
<tr>
<td>Angle</td>
<td>Saturation</td>
<td>Texture</td>
</tr>
<tr>
<td>Slope</td>
<td>Connection</td>
<td>Containment</td>
</tr>
<tr>
<td>Area</td>
<td>Length</td>
<td>Shape</td>
</tr>
<tr>
<td>Volume</td>
<td>Angle</td>
<td>Shape</td>
</tr>
<tr>
<td>Density</td>
<td>Slope</td>
<td>Shape</td>
</tr>
<tr>
<td>Saturation</td>
<td>Area</td>
<td>Shape</td>
</tr>
<tr>
<td>Hue</td>
<td>Volume</td>
<td>Shape</td>
</tr>
<tr>
<td>Texture</td>
<td>Volume</td>
<td>Shape</td>
</tr>
<tr>
<td>Containment</td>
<td>Shape</td>
<td>Shape</td>
</tr>
<tr>
<td>Shape</td>
<td>Shape</td>
<td>Shape</td>
</tr>
</tbody>
</table>

Conjectured effectiveness of the encoding

Mackinlay’s design algorithm

- User formally specifies data model and type
- APT searches over design space
  - Tests expressiveness of each visual encoding
  - Generates image for encodings that pass test
  - Tests perceptual effectiveness of resulting image
- Outputs most effective visualization

Limitations

- Does not cover many visualization techniques
  - Bertin and others discuss networks, maps, diagrams
  - They do not consider 3D, animation, illustration, photography, …
- Does not model interaction

Summary

- Formal specification
  - Data model
  - Image model
  - Encodings mapping data to image
- Choose expressive and effective encodings
  - Formal test of expressiveness
  - Experimental tests of perceptual effectiveness