Chapter 2: Data Principles

[graphics.stanford.edu/papers/polaris]
Further Readings


The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. Ben Shneiderman, Proc. 1996 IEEE Visual Languages, also Maryland HCIL TR 96-13. [citeseer.ist.psu.edu/shneiderman96eyes.html]


Dataset Types

DATASET TYPES

- tables
  - spatial fields
  - abstract tables
- networks
  - trees
  - node-link graphs
- text/logs
- file/stream
Tables

b) TABLES

columns = attributes

rows = items
Networks

- **NODES**
  - node attributes
  - nodes
    - node table
  - links
    - link attributes
    - link table
Attribute Types

- Categorical
- Ordered
  - Ordinal
  - Quantitative
- Sequential
- Diverging

amendment: sequential/diverging crosscuts ordered
Attribute Types

- continuous (quantitative)
  - 10 inches, 17 inches, 23 inches
Attribute Types

- **continuous (quantitative)**
  - 10 inches, 17 inches, 23 inches

- **ordered (ordinal)**
  - small, medium, large
  - days: Sun, Mon, Tue, ...

---

[graphics.stanford.edu/papers/polaris]
Attribute Types

- **continuous (quantitative)**
  - 10 inches, 17 inches, 23 inches

- **ordered (ordinal)**
  - small, medium, large
  - days: Sun, Mon, Tue, ...

- **categorical (nominal)**
  - apples, oranges, bananas

[graphics.stanford.edu/papers/polaris]
More Attribute Types: Stevens

- further subdivision of quantitative
  - interval: 0 location arbitrary
    - time: seconds, minutes
  - ratio: 0 fixed
    - physical measurements: Kelvin temp

Attribute Semantics

- spatial/nonspatial
- temporal/nontemporal
- independent/dependent
- continuous/discrete

- dimensions/measures
  - dimensions: categorical
  - measures: quantitative
  - databases vocab, used in Polaris
Dataset Semantics

- spatial/abstract
  - spatial fields if independent spatial attrs
  - abstract otherwise: must choose spatial layout
Attributes: Multidimensional Tables

MULTIDIMENSIONAL TABLES

ind1 \rightarrow \text{item} (\text{dep1}, \text{dep2})

ind: index/independent attrib

dep: dependent/value attrib
Attributes: Spatial Fields

Spatial Fields

dependent attrnb count

1-dp attr
2-dp attrnb
3-dp attrnb

scalar
vector
tensor
Dataset Semantics

- spatial/abstract
  - spatial fields if independent spatial attrs
  - abstract otherwise: must choose spatial layout
- static/timevarying
  - timevarying if independent temporal dimension
  - tv different than static/dynamic dataset types
  - time series data: simple special case
Other Data Taxonomies

- Shneiderman’s data+task taxonomy: data
  - 1D, 2D, 3D, temporal, nD, trees, networks
- Hanrahan’s addition:
  - text and documents

[Shneiderman, The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. Proc. 1996 IEEE Visual Languages]
Derived Attributes and Spaces

- derived attribute: compute from originals
  - simple change of type
  - complex transformation using global information

- derived spaces
  - dataset with derived attributes
  - may be only derived attrs, or derived+original attrs
  - dataset transformation as abstraction choice
Data Models vs. Conceptual Models

- data model: mathematical abstraction
  - set with operations
  - e.g. integers or floats with *, +

- conceptual model: mental construction
  - includes semantics, support data
  - e.g. navigating through city using landmarks

- conceptual model motivates derived data

[Hanrahan, graphics.stanford.edu/courses/cs448b-04-winter/lectures/encoding/walk005.html]

Derived Attributes Example

- data model
  - 17, 25, -4, 28.6
  - (floats)
Derived Attributes Example

- data model
  - 17, 25, -4, 28.6
  - (floats)
- conceptual model
  - temperature

depending on task, transform to data type
- making toast
  - burned vs. not burned (N)
- classifying showers
  - hot, warm, cold (O)
- finding anomalies in local weather patterns
  - continuous to 4 sig figures (Q)
Derived Attributes Example

- data model
  - 17, 25, -4, 28.6
  - (floats)

- conceptual model
  - temperature

- depending on task, transform to data type
  - making toast
    - burned vs. not burned (N)
  - classifying showers
    - hot, warm, cold (O)
  - finding anomalies in local weather patterns
    - continuous to 4 sig figures (Q)
Derived Space: Strahler Numbers for Trees

Derived: Feature Detection in Fluids

Derived: Graph-Theoretic Scagnostics

- SPLOM: scatterplot matrix

Scagnostic Measures

- scatterplot measures: monotonic, stringy, skinny, convex, striated, sparse, clumpy, skewed, outlying

Scagnostics Measures

Time

- **2D+T vs. 3D**
  - same or different? depends on POV
    - input side vs. output side
  - same
    - input: time as just one kind of abstract input dimension
  - different
    - input: semantics (time steps of dynamically changing data)
    - output: visual encoding channel of temporal change very different than spatial position change
- processing might be different
  - e.g. interpolate differently across timesteps than across spatial position
Polaris

- infovis spreadsheet
- table cell
  - not just numbers: graphical elements
  - wide range of retinal variables and marks
- table algebra $\leftrightarrow$ interactive interface
  - formal language
- influenced by Wilkinson’s Grammar of Graphics
- commercialized as Tableau Software
  - good sandbox for projects!

Polaris: Gantt Bar, Country/Time

Polaris: Circles, Lat/Long

Polaris: Circles, Profit/State:Months

Fields Create Tables and Graphs

- Ordinal fields: interpret field as 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>95892</td>
<td>101760</td>
<td>105282</td>
<td>98225</td>
</tr>
</tbody>
</table>

- Quantitative fields: treat field as single element sequence and encode as axes:
  - Profit = (Profit) ⇔

[Hanrahan, graphics.stanford.edu/courses/cs448b-04-winter/lectures/encoding]