Visualization Design

Maneesh Agrawala

CS 448B: Visualization
Spring 2016

Last Time: 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

[based on slide from Munzner]

Nominal, ordinal and quantitative

N - Nominal (labels)
Fruits: Apples, oranges, ...
Operations: =, ≠

O - Ordered
Quality of meat: Grade A, AA, AAA
Operations: =, ≠, <, ≤, ≥

Q - Interval (location of zero arbitrary)
Dates: Jan, 19, 2006; Loc.: (LAT 33.98, LON -118.45)
Like a geometric point. Cannot compare directly
Only differences (i.e. intervals) may be compared
Operations: =, ≠, <, >, ≤, ≥,

Q - Ratio (location of zero fixed)
Physical measurement: Length, Mass, Temp, ...
Counts and amounts
Like a geometric vector, origin is meaningful
Operations: =, ≠, <, >, ≤, ≥, ÷
Marks and Visual Variables

Marks: geometric primitives
- points
- lines
- areas

Visual Variables: control mark appearance
- Position (2x)
- Size
- Value
- Texture
- Color
- Orientation
- Shape

Semiology of Graphics
J. Bertin, 1967

Playfair 1786/1801

- Time → x-position (Q, linear)
- Exports/Imports Values → y-position (Q, linear)
- Exports/Imports → color (N, O, nominal)
- Balance for/against → area (maybe length??) (Q, linear)
- Balance for/against → color (N, O, nominal)
### Bertins’ “Levels of Organization”

<table>
<thead>
<tr>
<th>Position</th>
<th>N</th>
<th>O</th>
<th>Q</th>
</tr>
</thead>
<tbody>
<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 (≡)

Note: Q < O < N

### 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; .:. Length of bar 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, as 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 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’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>Hue</td>
<td>Connection</td>
</tr>
<tr>
<td>Area</td>
<td>Texture</td>
<td>Containment</td>
</tr>
<tr>
<td>Volume</td>
<td>Connection</td>
<td>Length</td>
</tr>
<tr>
<td>Density</td>
<td>Saturation</td>
<td>Angle</td>
</tr>
<tr>
<td>Saturation</td>
<td>Hue</td>
<td>Slope</td>
</tr>
<tr>
<td>Hue</td>
<td>Texture</td>
<td>Area</td>
</tr>
<tr>
<td>Texture</td>
<td>Connection</td>
<td>Slope</td>
</tr>
<tr>
<td>Connection</td>
<td>Containment</td>
<td>Area</td>
</tr>
<tr>
<td>Containment</td>
<td>Length</td>
<td>Shape</td>
</tr>
<tr>
<td>Shape</td>
<td></td>
<td>Volume</td>
</tr>
</tbody>
</table>

Conjectured effectiveness of the encoding
Graphical Perception

Most accurate

- Position (common) scale
- Position (non-aligned) scale
- Length
- Slope
- Angle
- Area
- Volume

Least accurate

- Color hue-saturation-density

APT: Automatic Chart Construction

Encode most important data using highest ranking visual variable for the data type

<table>
<thead>
<tr>
<th>Year</th>
<th>Exports</th>
<th>Imports</th>
</tr>
</thead>
<tbody>
<tr>
<td>1700</td>
<td>100,000</td>
<td>200,000</td>
</tr>
<tr>
<td>1701</td>
<td>110,000</td>
<td>210,000</td>
</tr>
<tr>
<td>1702</td>
<td>120,000</td>
<td>220,000</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

mark: lines

Year → x-pos (Q)
Exports → y-pos (Q)
Imports → y-pos (Q)

Automating the design of graphical presentation of relational information
J. Mackinlay, 1986
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
Announcements

Class participation requirements

- Complete readings before class
- In-class discussion
- Post at least 1 lecture/reading response by noon day after lecture

Class wiki
http://web.stanford.edu/class/cs448b/
Assignment 2: Exploratory Data Analysis

Use **Tableau** to formulate & answer questions

**First steps**
- Step 1: Pick a domain
- Step 2: Pose questions
- Step 3: Find data
- Iterate

**Create visualizations**
- Interact with data
- Question will evolve
- Tableau

**Make wiki notebook**
- Keep record of all steps you took to answer the questions

Due before class on Oct 17, 2016

---

Assignment 1: Visualization Design

**Breakfast Cereals**

*Number of cases: 77*

**Variable Names:**
- name: Name of cereal
- manufacturer: Manufacturer of cereal
- shelf: Display shelf of cereal (1, 2, or 3, counting from the floor upwards)
- weight: Weight of one serving in ounces
- cups: Number of cups in one serving
- type: Cold or hot cereal
- calories: Calories per serving
- protein: Grams of protein per serving
- fat: Grams of fat per serving
- sodium: Milligrams of sodium per serving
- fiber: Grams of dietary fiber per serving
- carbohydrates: Grams of complex carbohydrates per serving
- sugar: Grams of sugar per serving
- potassium: Milligrams of potassium per serving
- vitamins: Vitamins and minerals: 0, 20, or 100, indicating the typical percentage of FDA recommended per serving
- rating: A healthiness rating of the cereal, derived from their nutrients

A value of "-" indicates a missing observation.

We've cleaned up the cereal dataset and formatted it as a csv file for this assignment.
### Design Considerations

**Expressiveness**
- Do the mappings show the facts and only the facts?
- Are visual mappings consistent? (e.g., respect color mappings)

**Effectiveness**
- Are perceptually effective encodings used?
- Are the most important data mapped to the most effective visual variables?

**Cognitive Load (Efficiency)**
- Are there extraneous visual elements?

**Data Transformation**
- Are transformations (filter, sort, derive, aggregate) appropriate?

**Guides (Non-Data Elements)**
- Descriptive, consistent: Title, Label, Caption, Source, Annotations
- Meaningful references: Gridlines, Legend

### Design Space of A1 Submissions

<table>
<thead>
<tr>
<th>Spatial Encoding</th>
<th>Bar charts, Scatterplots, Pie charts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color Encoding</td>
<td>Mostly nominal (manufacturers), Quantitative (dual encoding)</td>
</tr>
<tr>
<td>Data Transformation</td>
<td>Often normalized data in some way</td>
</tr>
<tr>
<td>Labeling</td>
<td>Title, Caption, Axis labels, Legends Not many annotations</td>
</tr>
</tbody>
</table>
Average Nutrient Content of a Cup of Cereal
Broken Down by Manufacturer

Shelf 2, the “hot zone”, is notoriously unhealthy Options in this shelf are predominantly unhealthy High average sugar levels...

... and high calories, make this an unhealthy spot
How Healthy is Your Cereal?

By Do-Hyoung Park
CS 4486 • Fall 2016

Nutrition

Manufacturers

Shredded Wheat (Nabisco)
100% Bran (Nabisco)
Shredded Wheat w/ Bran (Nabisco)
Shredded Wheat, Spice-Size (Nabisco)
All-Bran with Extra Fiber (Kellogg’s)

Healthiest
Unhealthiest

By Manufacturer

Nabisco
American Home Food Products
Quaker Cuts
Post
Kellogg’s
Farrnor Purina
General Mills

Healthy
Very healthy
Devin healthy

16-15 15-14 14-13 13-12 12-11 11-10 10-9 9-8 8-7 7-6 6-5 5-4 4-3 3-2 2-1 1-0
In-Class Review

Procedure

Break into groups of 2
Present your visualization – in order by last name – 5 min each to describe what your visualization shows, and design choices
We will keep time and tell you to switch

Critique in order by last name – rubric on next slide (~10 min each)
- Write down feedback and score on rubric (be specific). We will collect critiques.
- Author takes notes (post critique notes/feedback to wiki after class) and keeps them

Turn in critiques you wrote for others
Authors add feedback notes to wiki page for A1

In-Class Review

Expressiveness
- Do the mappings show the facts and only the facts?
- Are visual mappings consistent? (e.g., respect color mappings)

Effectiveness
- Are perceptually effective encodings used?
- Are the most important data mapped to the most effective visual variables?

Cognitive Load (Efficiency)
- Are there extraneous visual elements?

Data Transformation
- Are transformations (filter, sort, derive, aggregate) appropriate?

Guides (Non-Data Elements)
- Descriptive, consistent: Title, Label, Caption, Source, Annotations
- Meaningful references: Gridlines, Legend