Visualization Design

Maneesh Agrawala

CS 448B: Visualization
Fall 2017

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: ≠, ≠, <, >, ≤, ≥, ±

---

On the theory of scales of measurements

S. S. Stevens, 1946
 Marks and Visual Variables

**Marks:** geometric primitives
- points
- lines
- areas

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

---

**Playfair 1786/1801**

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

<table>
<thead>
<tr>
<th></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: Q < O < N

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

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.
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>Density</td>
<td>Position</td>
</tr>
<tr>
<td>Length</td>
<td>Saturation</td>
<td>Hue</td>
</tr>
<tr>
<td>Angle</td>
<td>Texture</td>
<td>Texture</td>
</tr>
<tr>
<td>Slope</td>
<td>Connection</td>
<td>Containment</td>
</tr>
<tr>
<td>Area</td>
<td>Length</td>
<td>Density</td>
</tr>
<tr>
<td>Volume</td>
<td>Angle</td>
<td>Saturation</td>
</tr>
<tr>
<td>Density</td>
<td>Slope</td>
<td>Shape</td>
</tr>
<tr>
<td>Saturation</td>
<td>Area</td>
<td>Length</td>
</tr>
<tr>
<td>Hue</td>
<td>Volume</td>
<td>Angle</td>
</tr>
<tr>
<td>Texture</td>
<td>Shape</td>
<td>Slope</td>
</tr>
<tr>
<td>Connection</td>
<td>Length</td>
<td>Area</td>
</tr>
<tr>
<td>Containment</td>
<td>Angle</td>
<td>Volume</td>
</tr>
<tr>
<td>Shape</td>
<td></td>
<td></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

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>170,000</td>
<td>300,000</td>
</tr>
<tr>
<td>1701</td>
<td>171,000</td>
<td>302,000</td>
</tr>
<tr>
<td>1702</td>
<td>176,000</td>
<td>303,000</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

1. Year (Q)  2. Exports (Q)  3. Imports (Q)

mark: lines
Year \(\rightarrow\) x-pos (Q)
Exports \(\rightarrow\) y-pos (Q)
Imports \(\rightarrow\) 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 discussion substantive comment/question by noon the day after lecture (short paragraph)

Office hours on website

Class wiki
https://magrawala.github.io/cs448b-fa17
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 16, 2017

Assignment 1: Visualization Design

**Barley Yield Data**

In 1931 and 1932 Minnesota collected data on the yield in bushels per acre of 10 varieties of barley grown in 1/40 acre plots at University Farm, St. Paul, and at the five branch experiment stations located at Waseca, Morris, Crookston, Grand Rapids, and Duluth (all in Minnesota). The varieties were grown in three randomized blocks at each of the six stations during 1931 and 1932, different land being used each year of the test.

Number of records: 120
Variable Names:
- Site: Crookston, Duluth, Grand Rapids, Morris, University Farm, Waseca
- Variety: Gladiator, Mandura, No 457, No 462, No 478, Peatland, Swansota, Trebl, Velvet, Wisc. No 38
- Yield: bushels/acre
- Year: 1931, 1932

We’ve cleaned up this dataset and posted in csv format: barley2.csv

**Barley Yields**

Due by noon on Mon Oct 2
Submissions of PDF via Canvas, **bring printout to class**
### 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, Maps, Scatterplot, Pie</th>
</tr>
</thead>
<tbody>
<tr>
<td>Color Encoding</td>
<td>Mostly ordered or nominal (year, loc.), Quantitative (dual encoding)</td>
</tr>
<tr>
<td>Data Transformation</td>
<td>Aggregation (avg. yield across variety)</td>
</tr>
<tr>
<td>Labeling</td>
<td>Title, Caption, Axis labels, Legends Not many annotations</td>
</tr>
</tbody>
</table>
Barley Yield in Minnesota

Minneapolis Test: Average Yields 1931 and 1932

Across all locations and barley varieties, the average yield was 37.1 bushels/acre in 1931 and 31.8 bushels/acre in 1932.

*Morris was the only location where the total average was higher in 1932 than in 1931.

Different shades on Minnesota map represent different climates. Source: https://www.climategen.org/blog/minnesotac--changing-climate-updates-may-2021/
In-Class Review

Procedure

Break into groups of 3
Present your visualization – in order by last name – 3 min each to describe what your visualization shows, and design choices.
Others should write down critique on sheet
We will keep time and tell you to switch

Critique in order by last name – rubric on next slide (~5 min each)
- Tell author your critique.
- Give critiques to author

Author take photos of critiques and add to A1 along with a short response (1 paragraph) to the feedback.

---

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