Information Visualization

SWE 432, Fall 2016
Design and Implementation of Software for the Web
Today

• What types of information visualization are there?
  • Which one should you choose?
• What does usability mean for information visualizations?
• Cool reference from NYT on election visualization:
  • http://www.nytimes.com/interactive/2016/11/01/upshot/many-ways-to-map-election-results.html?_r=0
Recall: Cholera Epidemic in London, 1854

- >500 fatal attacks of cholera in 10 days
  - Concentrated in Broad Street area of London
  - Many died in a few hours
- Dominant theory of disease: caused by noxious odors
- Afflicted streets deserted by >75% inhabitants
Investigation and aftermath

• Based on visualization, did case by case investigation
• Found that **61 / 83** positive identified as using well water from Broad Street pump
• Board ordered pump-handle to be removed from well
• Epidemic soon ended
• Solved centuries old question of how cholera spread
Methods used by Snow

• Placed data in appropriate **context** for assessing cause & effect
  • Plotted on map, included well location
  • Reveals proximity as cause
• Made quantitative **comparisons**
  • Fewer deaths closer to brewery, could investigate cause
• Considered **alternative** explanations & contrary cases
  • Investigated cases not close to pump, often found connection to pump
• Assessment of possible **errors** in numbers
How information visualization amplifies cognition.

### Increased Resources
- **High-bandwidth hierarchical interaction**
  - The human moving gaze system partitions limited channel capacity so that it combines high spatial resolution and wide aperture in sensing visual environments (Resnikoff, 1987).
- **Parallel perceptual processing**
  - Some attributes of visualizations can be processed in parallel compared to text, which is aerial.
- **Offload work from cognitive to perceptual system**
  - Some cognitive inferences done symbolically can be recoded into inferences done with simple perceptual operations (Larkin and Simon, 1987).
- **Expanded working memory**
  - Visualizations can expand the working memory available for solving a problem (Norman, 1993).
- **Expanded storage of information**
  - Visualizations can be used to store massive amounts of information in a quickly accessible form (e.g., maps).

### Reduced Search
- **Locality of processing**
  - Visualizations group information used together, reducing search (Larkin and Simon, 1987).
- **High data density**
  - Visualizations can often represent a large amount of data in a small space (Tufte, 1983).
- **Spatially indexed addressing**
  - By grouping data about an object, visualizations can avoid symbolic labels (Larkin and Simon, 1987).

### Enhanced Recognition of Patterns
- **Recognition instead of recall**
  - Recognizing information generated by a visualization is easier than recalling that information by the user.
- **Abstraction and aggregation**
  - Visualizations simplify and organize information, supplying higher centers with aggregated forms of information through abstraction and selective omission (Card, Robertson, and Mackinlay, 1991; Resnikoff, 1987).
- **Visual schemata for organization**
  - Visually organizing data by structural relationships (e.g., by time) enhances patterns.
- **Value, relationship, trend**
  - Visualizations can be constructed to enhance patterns at all three levels (Bertin, 1977/1981).

### Perceptual Inference
- **Visual representations make some problems obvious**
  - Visualizations can support a large number of perceptual inferences that are extremely easy for humans (Larkin and Simon, 1987).
- **Graphical computations**
  - Visualizations can enable complex specialized graphical computations (Hutchins, 1996).

### Perceptual Monitoring
- **Visualizations can allow for the monitoring of a large number of potential events if the display is organized so that these stand out by appearance or motion.**

### Manipulable Medium
- **Unlike static diagrams, visualizations can allow exploration of a space of parameter values and can amplify user operations.**
Mapping data to visual form
Designing an information visualization

Raw Data: idiosyncratic formats
Data Tables: relations (cases by variables) + metadata
Visual Structures: spatial substrates + marks + graphical properties
Views: graphical parameters (position, scaling, clipping, ...)

Data

Data Transformations

Data Tables

Visual Mappings

Visual Structures

View Transformations

Views

Human Interaction

task
Types of raw data

- Nominal - unordered set
- Ordinal - ordered set
- Quantitative - numeric range
Data transformations

• Classing / binning: Quantitative —> ordinal
  • Maps ranges onto **classes** of variables
  • Can also count # of items in each class w/ histogram

• Sorting: Nominal —> ordinal
  • Add order between items in sets

• Descriptive statistics: mean, average, median, max, min, …
Visual structures

• 3 components
  • spatial substrate
  • marks
  • marks’ graphical properties
Spatial substrate

- Axes that divide space
- Types of axes - unstructured, nominal, ordinal, quantitative
- Composition - use of multiple orthogonal axes (e.g., 2D scatterplot, 3D)
Folding

• continuing an axis by continuing in different space

---

LaToza/Bell
GMU SWE 432 Fall 2016
Marks

• Points (0D)
• Lines (1D)
• Areas (2D)
• Volumes (3D)
Marks’ graphical properties

- Quantitative (Q), Ordinal (O), Nominal (N)
- Filled circle - good; open circle - bad
Effectiveness of graphical properties

- Quantitative (Q), Ordinal (O), Nominal (N)
- Filled circle - good; open circle - bad

<table>
<thead>
<tr>
<th></th>
<th>Spatial</th>
<th>Q</th>
<th>O</th>
<th>N</th>
<th>Object</th>
<th>Q</th>
<th>O</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extent</td>
<td>(Position)</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>Grayscale</td>
<td>○</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td></td>
<td>Size</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Differential</td>
<td>Orientation</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>Color</td>
<td>○</td>
<td>○</td>
<td>●</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Texture</td>
<td>○</td>
<td>○</td>
<td>●</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Shape</td>
<td>○</td>
<td>○</td>
<td>●</td>
</tr>
</tbody>
</table>
Animation

- Visualization can change over time
- Could be used to encode data as a function of time
  - But often not effective as makes direct comparisons hard
- Can be more effective to animate transition from before to after as user configures visualization
Examples of visualizations
Time-series data
Stacked graph

- Supports visual summation of multiple components
Small multiples

- supports separate comparison of data series
- may have better legibility than placing all in single plot
Maps
Choropleth map

• Groups data by area, maps to color
Cartograms

- Encodes two variables with size and color
Hierarchies
Node link diagram
Dendrogram

- leaf nodes of hierarchy on edges of circle
Treemaps
Networks
Force-directed layout

- edges function as springs, find least energy configuration
Arc diagram

- can support identifying cliques & bridges w/ right order
Adjacency matrix
Design considerations
Tufte’s principles of graphical excellence

• show the **data**
• induce the viewer to think about the substance rather than the methodology
• avoid distorting what the data have to say
• present **many** numbers in a small space
• make large data sets **coherent**
• encourage the eye to **compare** different pieces of data
• reveal data at several levels of detail, from overview to fine structure
• serve reasonable clear **purpose**: description, exploration, tabulation, decoration
Distortions in visualizations

• Visualizations may distort the underlying data, making it harder for reader to understand truth
• Use of **design** variation to try to falsely communicate **data** variation
Example

Operating Revenues

- 1970: $3,549,385
- 1971: $4,520,962
- 1972: $4,916,444
- 1973: $6,814,503
- 1974: $8,114,014

Net Income (Loss)

- 1970: $(11,014)
- 1971: $371,747
- 1972: $521,943
- 1973: $1,435,102
- 1974: $1,647,001

Exploration & Development Expenditures

- 1970: $351,341
- 1971: $85,149
- 1972: $75,243
- 1973: $329,421
- 1974: $1,226,007
Example

Nobel Prizes Awarded in Science, for Selected Countries, 1901-1974

(Number of Prizes)

United States

United Kingdom

Germany

U.S.S.R.

France

Example (corrected)
Example
Traditional Electoral Map
Weighted Electoral Map
Data-ink

- Data-ink - non-redundant ink encoding data information

\[
\text{Data-ink ratio} = \frac{\text{Data-ink}}{\text{Total ink used to print the graphic}}
\]

- proportion of a graphic’s ink devoted to the non-redundant display of data-information

- \( 1.0 - \text{proportion of a graphic that can be erased} \)
Examples of data-ink ratio
Design principles for data-ink

• (a.k.a. aesthetics & minimalism / elegance & simplicity)
• **Above all else show the data**
  • Erase non-data-ink, within reason
    • Often not valuable and distracting
    • Redundancy not usually useful
Example
Example (revised)
Interacting with visualizations
Interactive visualizations

• Users often use iterative process of making **sense** of the data
• Answers lead to new questions
• Interactivity helps user constantly change display of information to answer new questions
• Should offer visualization that offers best view of data moment to **moment** as desired view **changes**
Shneiderman’s visualization tasks

- Overview: gain an overview of entire collection
- Zoom: zoom in on items of interest
- Filter: filter out uninteresting items
- Details on demand: select an item or group and get details
- Relate: view relationships between items
- History: support undo, replay, progressive refinement
- Extract: allow extraction of sub-collections through queries
In Class Activity
Design an information visualization

• In groups of 2 or 3
  • Select a set of data to visualize and two or more representative questions to answer using this data
  • Design an **interactive** information visualization
    • Create sketches showing the design of the information visualization
    • Should have multiple views of data, interactions to configure and move between views