Information Visualization

SWE 632, Spring 2018
Today

• What types of information visualization are there?
  • Which one should you choose?
• What principles and guidelines inform the design of information visualizations?
• How can interactivity be used to design better information visualizations?
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
Charles Minard’s Map of Napoleon’s Russian Campaign of 1812

Drawn in Paris, November 20, 1869.

**Carte Figurative des pertes successives en hommes de l’Armée Française dans la campagne de Russie 1812-1813.**


The numbers of men present are represented by the length of the lines colored to show the number of soldiers. The colored lines represent the number of men present in Russia, the blue lines representing the men who remained in Russia until the end of the campaign.

**TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au-dessous de zéro.**

Les pyramides sont au volant le 12-8, au-dessous.

- 26° le 7 XI.
- 28° le 8 XI.
- 30° le 6 XI.

- 15° le 14 XI.
- 22° le 14 XII.
- 28° le 28 XII.
- 30° le 27 XII.

- 30° le 30 XI.
- 30° le 15 XII.
- 30° le 15 XIII.

The map shows the progression of Napoleon's army through Russia, with the number of soldiers remaining in each city or region represented by the length of the line.
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, ...)
Types of raw data

- Nominal - unordered set **without** a quantitative value
  - Gender: male, female
  - Hair color: brown, black, blonde, gray, orange, ...
- Ordinal - **ordered** set, with no meaning assigned to differences
  - How do you feel today: very unhappy, unhappy, ok, happy, very happy
  - Undefined how much better happy is than ok
- Quantitative - **numeric** value
  - Height, weight, distance, ...
Data transformations

- Classing / binning: Quantitative $\rightarrow$ ordinal
  - Maps ranges onto **classes** of variables
  - Can also count # of items in each class w/ histogram
- Sorting: Nominal $\rightarrow$ ordinal
  - Add order between items in sets
- Descriptive statistics: mean, average, median, max, min, …
Example uses of a data transformation
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)
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></th>
<th></th>
<th>Object</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Q</td>
<td>O</td>
<td>N</td>
<td>Grayscale</td>
<td>Q</td>
<td>O</td>
</tr>
<tr>
<td>Extent (Position)</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>●</td>
</tr>
<tr>
<td>Size</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>●</td>
</tr>
<tr>
<td>Differential Orientation</td>
<td>○</td>
<td>○</td>
<td>●</td>
<td>●</td>
<td>○</td>
<td>●</td>
</tr>
<tr>
<td></td>
<td>Color</td>
<td>Texture</td>
<td>Shape</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

18
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 & 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

Net Income (Loss)

Exploration & Development Expenditures


$(11,014)$
Example
Example (corrected)
Example

Purchasing Power of the Diminishing Dollar
Source: Labor Department
Weighted Electoral Map
Data-ink

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

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 – 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
Information 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
Global tweet map

https://www.mapd.com/demos/tweetmap/
Is It Better to Rent or Buy?

By MIKE BOSTOCK, SHAN CARTER and ARCHIE TSE

The choice between buying a home and renting one is among the biggest financial decisions that many adults make. But the costs of buying are more varied and complicated than for renting, making it hard to tell which is a better deal. To help you answer this question, our calculator takes the most important costs associated with buying a house and computes the equivalent monthly rent.

---

**Home Price**
A very important factor, but not the only one. Our estimate will improve as you enter more details below.

$208,000

---

**How Long Do You Plan to Stay?**
Buying tends to be better the longer you stay because the upfront fees are spread out over many years.

9 years

---

**What Are Your Mortgage Details?**
In addition to the interest rate and down payment, the calculator takes into account the mortgage-interest tax deduction.

3.67% Mortgage rate

---

If you can rent a similar home for less than ...

$751 PER MONTH

... then renting is better.

<table>
<thead>
<tr>
<th>Costs after 9 years</th>
<th>Rent</th>
<th>Buy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial costs</td>
<td>$751</td>
<td>$49,920</td>
</tr>
<tr>
<td>Recurring costs</td>
<td>$90,937</td>
<td>$137,912</td>
</tr>
<tr>
<td>Opportunity costs</td>
<td>$13,092</td>
<td>$37,376</td>
</tr>
<tr>
<td>Net proceeds</td>
<td>-$751</td>
<td>-$121,180</td>
</tr>
<tr>
<td>Total</td>
<td>$104,029</td>
<td>$104,029</td>
</tr>
</tbody>
</table>

*How to Read the Charts* Charts that are relatively flat indicate factors that are not particularly important to the outcome. Conversely, the factors that have steep slopes have a large impact.

---

https://www.nytimes.com/interactive/2014/upshot/buy-rent-calculator.html?_r=0
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