Administrivia

• HW6 due today

• HW7 due next week

• Project presentations in class in 2 weeks
Overview of Information Visualization
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
Amplifying Cognition

- Information Visualization can amplify cognition by:
  
  1. *Increasing* the *memory* and *processing* resources available to users
  2. *Reducing* the *search* for information
  3. *Using* visual representations to *enhance* the *detection of patterns*
  4. *Enabling* perceptual inference
  5. *Using* perceptual *attention mechanisms* for *monitoring*
  6. *Encoding* Information in a *manipulable* medium
Charles Minard’s Map of Napoleon’s Russian Campaign of 1812

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

Drawn by M. Minard, Statuary General of Lens to Cluny, and Retired

Paris, the 20th of November 1869.

In the number of men present on the map, the increase in the number of men in Russia, the total, and the per cent of each division of death in the various phases of the campaign, including the march of the army to Moscow and the retreat. The map shows the temperature at different points on the route. The scale is given at the bottom.

Drawn in Paris, November 20, 1869.
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, ...)

Diagram shows the process from raw data to visual form with human interaction.
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, …
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

<table>
<thead>
<tr>
<th></th>
<th>Spatial</th>
<th>Object</th>
<th>Differential</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extent</td>
<td>Position</td>
<td>Grayscale</td>
<td>Orientation</td>
</tr>
<tr>
<td></td>
<td>Q</td>
<td>O</td>
<td>Q</td>
</tr>
<tr>
<td>Size</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td></td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td>Orientation</td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
<tr>
<td></td>
<td>●</td>
<td>●</td>
<td>●</td>
</tr>
</tbody>
</table>

- Quantitative (Q), Ordinal (O), Nominal (N)
- Filled circle - good; open circle - bad
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 w/ size & color
Cartograms

- Encodes two variables w/ size & color
Hierarchies
Node Link Diagram
Dendrogram

- Leaf nodes of hierarchy on edges of circle
Treemaps
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
Example
Example (corrected)
Example
Weighted Electoral Map

306 BIDEN

270 to win

232 TRUMP
Data-ink

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

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

\[
= \frac{\text{proportion of a graphic’s ink devoted to the non-redundant display of data-information}}{\text{1.0 – proportion of a graphic that can be erased}}
\]
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
US Electricity Sources

https://www.carbonbrief.org/mapped-how-the-us-generates-electricity/
Renting vs. Buying Utility

Is It Better to Rent or Buy?

BY WYNNE NOTICIO, SHAY CARTER AND MIRONE TIE

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 payments, the calculator takes into account the mortgage-interest tax deduction.

3.67% Mortgage rate @ $100 per month

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

$751

... then renting is better.

<table>
<thead>
<tr>
<th>Costs after 5 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>Renting costs</td>
<td>$80,097</td>
<td>$17,912</td>
</tr>
<tr>
<td>Opportunity costs</td>
<td>$33,092</td>
<td>$17,376</td>
</tr>
<tr>
<td>Net present value</td>
<td>$751</td>
<td>$121,189</td>
</tr>
<tr>
<td>Total</td>
<td>$104,129</td>
<td>$104,129</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 boxers that have steep slopes have a larger impact.

https://www.nytimes.com/interactive/2014/upshot/buy-rent-calculator.html?_r=0
10 Minute Break
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
• Deliverables: 2+ questions you support, sketches with annotations explaining how users would use visualization to answer questions
• Due by 6:25pm today