Today

- **Before class**
  - Reading: Ch 1 - What's Vis, and Why Do It?

- **During class**
  - Highlight and discuss Ch 1
  - *We will not cover everything that you are responsible for during class time.*
Topic Objectives

- Define visualization.
- Explain the importance of humans in the visualization process.
- Explain why human vision is particularly well-suited for information transfer.
- Give an example of a visualization idiom.
- Explain why it is best to consider multiple alternatives for vis before selecting a solution.
- Explain at a high-level the "what-why-how" framework for analyzing visualization use.
- Differentiate between R, D3, and Tableau and describe the type of tasks for which each tool might be most appropriate.

What is visualization?

- "The communication of information using graphical representations"
  - Ward, Grinstein, Keim

- "The use of computer-supported interactive visual representations of data to amplify cognition"
  - Card, Mackinlay, Shneiderman, Readings in Information Visualization: Using Vision to Think

- "The purpose of visualization is insight, not pictures."
  - Ben Shneiderman
Where have you seen a visualization today?

What's vis?

- Visualization is suitable when there is a need to augment human capabilities rather than replace people with computational decision-making methods.

- The design space of possible vis idioms is huge, and includes the considerations of both how to create and how to interact with visual representations.

- Vis design is full of tradeoffs, and most possibilities in the design space are ineffective for a particular task, so validating the effectiveness of a design is both necessary and difficult.

- Vis designers must take into account three very different kinds of resource limitations: those of computers, of humans, and of displays.

- Vis usage can be analyzed in terms of why the user needs it, what data is shown, and how the idiom is designed.
Visualization is suitable when there is a **need to augment human capabilities** rather than replace people with computational decision-making methods.

**Why have a human in the loop?**

- Vis allows people to analyze data when they don't know exactly what questions to ask in advance.

- **Best path - put a human in the loop**
  - exploit the pattern detection properties of human vision
Humans are great at pattern recognition

Create visualizations that lets computers do what computers do well and lets humans do what humans do well.

Uses of vis tools

- Transitional
  - vis works itself out of a job
- Long-term
  - exploratory analysis
- Presentation
  - visual explanations

https://www.geovista.psu.edu/research/healthvisualization/

Why have a computer in the loop?

- augment human capabilities

Why use an external representation?

- Vis allows people to offload cognition and memory usage to make space for other operations.

- Diagrams as external representations
  - information can be organized by spatial location
    - search - grouping items needed for problem-solving in one location
    - recognition - grouping relevant info for one item in the same location
Visualization can extend your memory

What is 57 x 48?

<table>
<thead>
<tr>
<th>paper</th>
<th>mental buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>57 x 48</td>
<td>[8 * 7 = 56]</td>
</tr>
<tr>
<td>1 456</td>
<td>[8 * 5 = 40 + 5 = 45]</td>
</tr>
<tr>
<td>228</td>
<td>[4 * 7 = 28]</td>
</tr>
<tr>
<td><strong>2736</strong></td>
<td>[4 * 5 = 20 + 2 = 22]</td>
</tr>
<tr>
<td></td>
<td>[8 + 5 = 13]</td>
</tr>
<tr>
<td></td>
<td>[4 + 2 + 1 = 7]</td>
</tr>
</tbody>
</table>

Example courtesy Tamara Munzner, Univ. of British Columbia

augment human capabilities

Why depend on vision?

- Visual system provides a high-bandwidth channel to our brains.
- Significant amount of visual information processing occurs in parallel at the pre-conscious level.
Can you find the red dot?

preattentive processing

Which state had the highest marriage rate?

- Florida - 7.5
- Connecticut - 6.9
- Colorado - 6.3
- Delaware - 5.4
- District of Columbia - 4.7

http://www.csc.ncsu.edu/faculty/healey/PP/index.html
Which state had the highest marriage rate?

[Bar chart showing marriage rates by state]

Why show the data in detail?

- Vis tools can allow people to explore data to find patterns or to determine if a statistical model actually fits the data.

- Look out for questionable data.
  - "just because it's numbers doesn't mean it's true"
  - is it a typo or something interesting?
    - "make sure you know which one it is"
Anscombe's Quartet

Anscombe’s Quartet: Raw Data

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>y</td>
<td>x</td>
<td>y</td>
<td>x</td>
</tr>
<tr>
<td>10.0</td>
<td>8.04</td>
<td>10.0</td>
<td>9.14</td>
<td>10.0</td>
</tr>
<tr>
<td>8.0</td>
<td>6.95</td>
<td>8.0</td>
<td>8.14</td>
<td>8.0</td>
</tr>
<tr>
<td>13.0</td>
<td>7.58</td>
<td>13.0</td>
<td>8.74</td>
<td>13.0</td>
</tr>
<tr>
<td>9.0</td>
<td>8.81</td>
<td>9.0</td>
<td>8.77</td>
<td>9.0</td>
</tr>
<tr>
<td>11.0</td>
<td>8.33</td>
<td>11.0</td>
<td>9.26</td>
<td>11.0</td>
</tr>
<tr>
<td>14.0</td>
<td>9.96</td>
<td>14.0</td>
<td>8.10</td>
<td>14.0</td>
</tr>
<tr>
<td>6.0</td>
<td>7.24</td>
<td>6.0</td>
<td>6.13</td>
<td>6.0</td>
</tr>
<tr>
<td>4.0</td>
<td>4.26</td>
<td>4.0</td>
<td>3.10</td>
<td>4.0</td>
</tr>
<tr>
<td>12.0</td>
<td>10.84</td>
<td>12.0</td>
<td>9.13</td>
<td>12.0</td>
</tr>
<tr>
<td>7.0</td>
<td>4.82</td>
<td>7.0</td>
<td>7.26</td>
<td>7.0</td>
</tr>
<tr>
<td>5.0</td>
<td>5.68</td>
<td>5.0</td>
<td>4.74</td>
<td>5.0</td>
</tr>
<tr>
<td>mean</td>
<td>9.0</td>
<td>7.5</td>
<td>9.0</td>
<td>7.5</td>
</tr>
<tr>
<td>var.</td>
<td>10.0</td>
<td>3.75</td>
<td>10.0</td>
<td>3.75</td>
</tr>
<tr>
<td>corr.</td>
<td>0.816</td>
<td>0.816</td>
<td>0.816</td>
<td>0.816</td>
</tr>
</tbody>
</table>

Munzner, Figure 1.3

The four data sets are not the same

"Graphics reveal data"
- Edward Tufte, *The Visual Display of Quantitative Information*
The design space of possible vis idioms is huge, and includes the considerations of both how to create and how to interact with visual representations.
Why is the idiom design space huge?

- Vis idioms - approaches to creating and manipulating visual representations

- Simple examples: scatterplots, bar charts, line charts

45 Ways to Communicate Two Quantities

[Diagram: Multiple ways to communicate two quantities]

https://visual.ly/blog/45-ways-to-communicate-two-quantities/
The design space of possible visualization idioms is huge.

Why use interactivity?

- Interaction allows for
  - handling complexity
  - displaying multiple aspects of a dataset

Vis design is full of tradeoffs, and most possibilities in the design space are ineffective for a particular task, so validating the effectiveness of a design is both necessary and difficult.

Why focus on tasks?

- The intended task is just as important as the data to be visualized.

- Four categories of tasks
  - presentation
  - discovery
  - enjoyment of information
  - producing more information for later use
Why focus on effectiveness?

- Effectiveness is an important measure for understanding if the user task was supported.
  - "The purpose of visualization is insight, not pictures." - Ben Shneiderman

- But, no picture can tell the truth, the whole truth, and nothing but the truth.

Why are most designs ineffective?

- Design may not match with human perception

- Design may not match with intended task

![Pie chart](image)

- Which color comprises the greatest portion?
- What is the percentage of the green region?
1. What's Vis, and Why Do It?

Consideration space
Proposal space

<table>
<thead>
<tr>
<th>Good!</th>
<th>Bad!</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space of possible solutions</td>
<td>Space of possible solutions</td>
</tr>
<tr>
<td>Known space</td>
<td>Proposal space</td>
</tr>
<tr>
<td>Selected solution</td>
<td>Consideration space</td>
</tr>
<tr>
<td>x Good solution</td>
<td>x Good solution</td>
</tr>
<tr>
<td>o OK solution</td>
<td>o OK solution</td>
</tr>
<tr>
<td>□ Poor Solution</td>
<td>□ Poor Solution</td>
</tr>
</tbody>
</table>

The diagram shows five spaces, each of which is progressively smaller than the previous. First, there is the space of all possible solutions, including potential solutions that nobody has ever thought of before. Next, there is the set of possibilities that are known to you, the vis designer. Of course, this set might be small if you are a novice designer who is not aware of the full array of methods that have been proposed in the past. If you're in that situation, one of the goals of this book is to enlarge the set of methods that you know about. The next set is the consideration space of the solutions that you actively consider. This set is necessarily smaller than the known space, because you can't consider what you don't know. An even smaller set is the proposal space of possibilities that you investigate in detail. Finally, one of these becomes the selected solution.

The diagram in Figure 1.5 contrasts a good strategy on the left, where the known and consideration spaces are large, with a bad strategy on the right, where these spaces are small. The problem of a small consideration space is the higher probability of only considering ok or poor solutions and missing a good one. A fundamental principle of design is to consider multiple alternatives and then choose the best, rather than to immediately fixate on one solution without considering any alternatives. One way to ensure that more than one possibility is considered is to explicitly generate multiple ideas in parallel. This book is intended to help you as a designer entertain a broad consideration space by systematically considering many alternatives, and also to help you rule out some parts of the space by noting when there are mismatches of possibilities with human capabilities or the intended task.

As with all design problems, vis design cannot be easily handled as a simple process of optimization because trade-offs abound. A design that does well by one measure will rate poorly on another. The characterization of trade-offs in the vis design space is a very open problem at the frontier of research.

Why is validation difficult?

- How do you know if your visualization "works"?
  - How do you measure insight?

- How do you argue that one design is better than another?
  - What does "better" mean? faster? more fun? more effective?
  - What does "effectively" mean?
Vis designers must take into account three very different kinds of resource limitations: those of computers, of humans, and of displays.

Why are there resource limitations?

- computational capacity
- human perceptual and cognitive capacity
- display capacity

Vis usage can be **analyzed in terms of why** the user needs it, **what** data is shown, **and how** the idiom is designed.

**Why analyze vis?**

- Analyzing existing systems is a good stepping stone to designing new ones.
- High-level framework for analyzing vis use
  - **what** data the user sees
  - **why** the user intends to use a vis tool
  - **how** the visual encoding and interaction idioms are constructed in terms of design choices
Tools

Workflow

- **What**
  - data gathering
  - data wrangling

- **Why**
  - developing questions
  - initial analysis

- **How**
  - charts for analysis
  - charts for presentation
What: Data Gathering


- Tabula – extract tables from PDFs
- Beautiful Soup – extract data from webpages

What: Data Wrangling

- Data is often messy

- Tools
  - Excel
  - OpenRefine – filter and clean data files

*Much more on this next week*
Why

- Developing questions
  - which states have the highest marriage rates?
  - which states have the highest divorce rates? is that correlated to marriage rate?
  - which states have the highest birth rates? is that correlated to marriage rate?

- Initial analysis
  - Excel
  - Google Sheets, Google Charts
  - Tableau
  - R

---

How

- Charts for analysis

- Charts for presentation

[Image of maps and charts]

https://www.geovista.psu.edu/research/healthvisualization/

Excel

http://chandoo.org/wp/2008/09/03/6-charts-to-never-use/

http://www.juiceanalytics.com/writing/recreating-ny-times-cancer-graph/

---

R

http://www.r-project.org
Tableau

https://www.tableau.com/academic/students

Seattle Real Estate: Overview

Number of Home Sales

Months of Supply

Change in Median Price

D3

http://d3js.org
Vega and Vega-Lite

- Vega is a visualization grammar, built on top of D3
  https://vega.github.io/vega/examples/bar-chart/

- Vega-Lite is a higher-level language built on top of Vega
  https://vega.github.io/vega-lite/examples/bar.html

Observable

https://observablehq.com
My Observable Notebooks

Choosing Tools – datavisualization.ch
Choosing Tools – chartmaker directory

<table>
<thead>
<tr>
<th></th>
<th>Amazon QuickSight</th>
<th>AntGIS</th>
<th>ChartJS</th>
<th>D3.js</th>
<th>Datawrapper</th>
<th>FusionCharts</th>
<th>Gephi</th>
<th>Google Charts</th>
<th>Google Data Studio</th>
<th>Highcharts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bar chart</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>Clustered bar chart</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>Bullet chart</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>Dot plot</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>Connected dot plot</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>Pictogram</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
<td>●</td>
</tr>
</tbody>
</table>

http://chartmaker.visualisingdata.com

What I Learned Recreating One Chart Using 24 Tools

Tools for Analysis vs. Presentation


Flexibility of Tools

Apps vs. Libraries and Static vs. Interactive

<table>
<thead>
<tr>
<th>Static Apps</th>
<th>Static Charting Libraries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Illustrator, NodeBox, Excel, Polestar, Raw</td>
<td></td>
</tr>
<tr>
<td>Ggplot2, Matplotlib, R, Seaborn, Bokeh, Processing</td>
<td></td>
</tr>
<tr>
<td>Web-Interactive Apps</td>
<td>Web-Interactive Charting Libraries</td>
</tr>
<tr>
<td>Highcharts Cloud, Quadrigram, Easychrt, Datawrapper, Tableau, Plotly, Google Sheets</td>
<td></td>
</tr>
<tr>
<td>D3, D4, C3, NVD3, Ggviz, Highcharts, Shiny, Vega, Vega-Lite</td>
<td></td>
</tr>
</tbody>
</table>


One Chart, Nine Tools – Revisited

https://lisacharlotterost.github.io/datavistools-revisited
Define visualization.

Explain the importance of humans in the visualization process.

Explain why human vision is particularly well-suited for information transfer.

Give an example of a visualization idiom.

Explain why it is best to consider multiple alternatives for vis before selecting a solution.

Explain at a high-level the "what-why-how" framework for analyzing visualization use.

Differentiate between R, D3, and Tableau and describe the type of tasks for which each tool might be most appropriate.