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 visualization 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?

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 $57 \times 48$?

<table>
<thead>
<tr>
<th></th>
<th>paper</th>
<th>mental buffer</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{3}{2}$</td>
<td>57</td>
<td>$[8 \times 7 = 56]$</td>
</tr>
<tr>
<td>$\times 48$</td>
<td>1</td>
<td>$[8 \times 5 = 40 + 5 = 45]$</td>
</tr>
<tr>
<td>1</td>
<td>456</td>
<td>$[4 \times 7 = 28]$</td>
</tr>
<tr>
<td></td>
<td>228</td>
<td>$[4 \times 5 = 20 + 2 = 22]$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$[8 + 5 = 13]$</td>
</tr>
<tr>
<td>$\frac{2}{7}$</td>
<td>36</td>
<td>$[4 + 2 + 1 = 7]$</td>
</tr>
</tbody>
</table>

Example courtesy Tamara Munzner, Univ. of British Columbia

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 - 6.8
- 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?

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>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>
</tr>
<tr>
<td>10.0</td>
<td>8.04</td>
<td>10.0</td>
<td>9.14</td>
</tr>
<tr>
<td>8.0</td>
<td>6.95</td>
<td>8.0</td>
<td>8.14</td>
</tr>
<tr>
<td>13.0</td>
<td>7.58</td>
<td>13.0</td>
<td>8.74</td>
</tr>
<tr>
<td>9.0</td>
<td>8.81</td>
<td>9.0</td>
<td>8.77</td>
</tr>
<tr>
<td>11.0</td>
<td>8.33</td>
<td>11.0</td>
<td>9.26</td>
</tr>
<tr>
<td>14.0</td>
<td>9.96</td>
<td>14.0</td>
<td>8.10</td>
</tr>
<tr>
<td>6.0</td>
<td>7.24</td>
<td>6.0</td>
<td>6.13</td>
</tr>
<tr>
<td>4.0</td>
<td>4.26</td>
<td>4.0</td>
<td>3.10</td>
</tr>
<tr>
<td>12.0</td>
<td>10.84</td>
<td>12.0</td>
<td>9.13</td>
</tr>
<tr>
<td>7.0</td>
<td>4.82</td>
<td>7.0</td>
<td>7.26</td>
</tr>
<tr>
<td>5.0</td>
<td>5.68</td>
<td>5.0</td>
<td>4.74</td>
</tr>
</tbody>
</table>

mean 9.0  7.5  9.0  7.5  9.0  7.5  9.0  7.5
var. 10.0 3.75 10.0 3.75 10.0 3.75 10.0 3.75
corr. 0.816 0.816 0.816 0.816

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

https://visual.ly/blog/45-ways-to-communicate-two-quantities/
**Why use interactivity?**

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

[source](http://adilmoujahid.com/posts/2016/08/interactive-data-visualization-geospatial-d3-dc-leaflet-python/)
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

Which color comprises the greatest portion?

What is the percentage of the green region?
Search space metaphor for vis design

![Diagram showing search space metaphor for vis design]

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
  - Open Refine – 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

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

Excel

![Excel chart example]

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

---

![R plot example]

43 44
Tableau

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

Seattle Real Estate: Overview

Number of Home Sales

Table

<table>
<thead>
<tr>
<th>County</th>
<th>Number of Home Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>King</td>
<td></td>
</tr>
<tr>
<td>Pierce</td>
<td></td>
</tr>
<tr>
<td>Snohom</td>
<td></td>
</tr>
<tr>
<td>Skagit</td>
<td></td>
</tr>
<tr>
<td>Island</td>
<td></td>
</tr>
<tr>
<td>Whatcom</td>
<td></td>
</tr>
</tbody>
</table>

Springs

Many examples use D3v3. Current version is D3v5. In many instances, these are not compatible.

D3

http://d3js.org
Choosing Tools – datavisualization.ch

http://selection.datavisualization.ch

Choosing Tools – chartmaker directory

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

One Chart, Nine Tools – Revisited

Class Tools

- **ODU-CS Gitlab Community**
  - [https://git-community.cs.odu.edu/](https://git-community.cs.odu.edu/)

- **Bl.ocks.org**
  - [https://bl.ocks.org](https://bl.ocks.org)
  - My blocks at bl.ocks.org: [https://bl.ocks.org/weiglemc](https://bl.ocks.org/weiglemc)

- **BlockBuilder**
  - [https://blockbuilder.org](https://blockbuilder.org)
  - My blocks: [https://blockbuilder.org/search?user=weiglemc](https://blockbuilder.org/search?user=weiglemc)

- **Observable**
  - [https://beta.observablehq.com](https://beta.observablehq.com)
  - My notebooks: [https://beta.observablehq.com/@weiglemc](https://beta.observablehq.com/@weiglemc)
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.