Data Visualization (DSC 530/CIS 602-01)

Definition & Web Programming

Dr. David Koop
Definition of Visualization

“Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.”

— T. Munzner
"Computer-based visualization systems provide visual representations of **datasets** designed to help people carry out **tasks** more effectively."
Definition

“Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.”
“Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.”
Definition

“Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively.”
Why People?

• Certain tasks can be totally *automated*
  - Statistical computations
  - Machine learning algorithms
  - We don’t need visualization for these tasks (although perhaps for debugging them…)

• Analysis problems are often *ill-specified*
  - What is the correct question?
  - Exploit human visual system, pattern detection capabilities
  - Goal may be an automated solution or a visual analysis system

• Presentation
  - It is often easier to show someone something than to tell them a bunch of facts about the data (and let them explore it)
Why Computers?

[Cerebral, Barsky et al., 2007]
Why Computers?

[Cerebral, Barsky et al., 2007]
Resource Limitations

- Memory and space constraints
- How many pixels do I have?
- Information Density
Administrivia

- Course Web Site
- Syllabus
  - Plagiarism
  - Accommodations
- Books:
  - Munzner (VAD)
  - Murray (IDV)
- Assignments
- Exams
- Project: Create an interactive vis
- Registration:
  - Add/Drop is Friday
Do not cheat!
Do not plagiarize

• It is **cheating**. It violates the Academic Honesty Policy at UMassD.
• Do your own work
• Do not copy anyone else's work, text, sentences, …
  - Anyone = another student, an internet source, book, blog, …
• Never quote text unless there is a specific need.
  - Usually, only famous quotes or very specific definitions
  - "I think there is a world market for maybe five computers."
    —Thomas Watson (1874-1956), Chairman of IBM, 1943
• **Cite** sources that back up your claims or reflect the origin of an idea
  - Vertex cover is an NP-Complete problem [1].

  …

Do not cheat

• Cheating on assignments, projects, and exams is not allowed
• You will receive a zero on the assignment/project/exam
• It will be reported to the department and university
• If it repeats, you will fail the course
• You can be kicked out of the university
Do ask questions!
Do ask questions

• If you are stuck on a specific issue with an assignment:
  - Do email me with **specific** questions
  - Do consult books, online documentation, tutorials
  - Do discuss that specific issue with a classmate

• If you are asked about a question:
  - Do not share your code
  - If the questioner is trying to cheat, walk away
  - If you see an obvious mistake, kindly point it out
  - Suggest a specific function or library that may be useful
“Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively”
### Why Visual?

<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>
</tbody>
</table>

[F. J. Anscombe]
Why Visual?

<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>10.0</td>
<td>8.04</td>
<td>10.0</td>
<td>7.46</td>
</tr>
<tr>
<td></td>
<td>8.0</td>
<td>6.95</td>
<td>8.0</td>
<td>6.77</td>
</tr>
<tr>
<td></td>
<td>13.0</td>
<td>7.58</td>
<td>13.0</td>
<td>12.74</td>
</tr>
<tr>
<td></td>
<td>9.0</td>
<td>8.81</td>
<td>9.0</td>
<td>7.11</td>
</tr>
<tr>
<td></td>
<td>11.0</td>
<td>8.33</td>
<td>11.0</td>
<td>7.81</td>
</tr>
<tr>
<td></td>
<td>14.0</td>
<td>9.96</td>
<td>14.0</td>
<td>8.84</td>
</tr>
<tr>
<td></td>
<td>6.0</td>
<td>7.24</td>
<td>6.0</td>
<td>6.08</td>
</tr>
<tr>
<td></td>
<td>4.0</td>
<td>4.26</td>
<td>4.0</td>
<td>5.39</td>
</tr>
<tr>
<td></td>
<td>12.0</td>
<td>10.84</td>
<td>9.13</td>
<td>8.15</td>
</tr>
<tr>
<td></td>
<td>7.0</td>
<td>4.82</td>
<td>7.0</td>
<td>6.42</td>
</tr>
<tr>
<td></td>
<td>5.0</td>
<td>5.68</td>
<td>5.0</td>
<td>5.73</td>
</tr>
</tbody>
</table>

Mean of x: 9
Variance of x: 11
Mean of y: 7.50
Variance of y: 4.122
Correlation: 0.816

[F. J. Anscombe]
Why Visual?

[F. J. Anscombe]
Visual Pop-out
Visual Pop-out
Visual Pop-out
Visual Perception Limitations
Visual Perception Limitations

[C. G. Healey]
Another Test

- https://www.youtube.com/watch?v=0grANlx7y2E
Other Human Limitations

• Visual working memory is **small**

• **Change blindness:** Large changes go unnoticed when we are working on something else in our view
“Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively”
Design Iteration

[k. Quealy, 2013]
Design Iteration

<table>
<thead>
<tr>
<th>Team</th>
<th>Quarterbacks</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York Giants</td>
<td>Eli Manning</td>
</tr>
<tr>
<td>Indianapolis Colts</td>
<td>Peyton Manning, Andrew Luck</td>
</tr>
<tr>
<td>San Diego Chargers</td>
<td>Drew Brees, Philip Rivers, Joe Flacco</td>
</tr>
<tr>
<td>Baltimore Ravens</td>
<td>Kyle Boller, Steve McNair, Joe Flacco</td>
</tr>
<tr>
<td>New England Patriots</td>
<td>Tom Brady, Matt Cassel, Tom Brady</td>
</tr>
<tr>
<td>Green Bay Packers</td>
<td>Brett Favre, Aaron Rodgers, Aaron Rodgers</td>
</tr>
<tr>
<td>New Orleans Saints</td>
<td>Aaron Brooks, Drew Brees, Drew Brees</td>
</tr>
<tr>
<td>Atlanta Falcons</td>
<td>Michael Vick, Matt Ryan, Mark Sanchez</td>
</tr>
<tr>
<td>New York Jets</td>
<td>Chad Pennington, Brett Favre, Mark Sanchez</td>
</tr>
<tr>
<td>Cincinnati Bengals</td>
<td>Carson Palmer, Ryan Fitzp, Carson Palmer, Andy Dalton</td>
</tr>
<tr>
<td>Houston Texans</td>
<td>David Carr, Matt Schaub, Matt Schaub</td>
</tr>
<tr>
<td>Carolina Panthers</td>
<td>Jake Delhomme, Cam Newton, Tim Tebow, Peyton Manning</td>
</tr>
<tr>
<td>Denver Broncos</td>
<td>Jake Plummer, Jay Cutler, Tyler Orton, Matt Schaub</td>
</tr>
<tr>
<td>Arizona Cardinals</td>
<td>Matt Leinart, Kurt Warner, Kyle Boller, Jon Kitna</td>
</tr>
<tr>
<td>Jacksonville Jaguars</td>
<td>Byron Leftwich, David Garrard, Blaine Gabbert</td>
</tr>
<tr>
<td>Detroit Lions</td>
<td>Joey Harrington, Jon Kitna, Matthew Stafford</td>
</tr>
<tr>
<td>Tampa Bay Buccaneers</td>
<td>Chris Simms, Bruce Gradkowski, Jeff Garcia, Josh Freeman</td>
</tr>
<tr>
<td>Dallas Cowboys</td>
<td>Drew Bledsoe, Tony Romo, Tony Romo</td>
</tr>
</tbody>
</table>

[K. Quealy, 2013]
Design Iteration

Each streak shows consecutive starts by a quarterback for a single team. Streaks include playoffs.

Only two players have longer streaks: Brett Favre (275) and Eli’s brother, Peyton (227).

Among active players, Philip Rivers (122) and Joe Flacco (96) are closest behind Eli.

Find a quarterback

Eli Manning (149)

[K. Quealy, 2013]
Another Design Example

Each solid circle represents a bee species active in Carlinville, Ill., in both the late 1800s and 2010.

Hatching represents a bee species active in the 1800s but now locally extinct.

The spot where each block rests on the circle indicates one of 26 plant species frequently by these bees.

In the 1880s scientists observed the following about the bee-plant encounters:

- Present
- Frequent
- Abundant

Studies in 2009 and 2010 showed many bee-plant interactions had changed:

- Lost
- Persisted
- New

[M. Stefaner, 2013]
Design Studies

- Design Study Methodology, Sedlmair et al., 2012

Design study: "[A] way to explore the choices made when applying visualization techniques to a particular application area"

Definition:
- a project, not a paper
- a specific real-world problem, with real users and real data
- design a visualization system: gather requirements, examine multiple ideas
- validate the design: problem characterization through final tool
- reflect about lessons learned: improve design guidelines
Design Study

• Steps:
  - Analyzing the problem
  - Abstracting data and tasks
  - Designing and implementing a visualization solution
  - Evaluating the solution with real users
  - Writing up the findings
A design study is a project in which visualization researchers analyze a specific real-world problem faced by domain experts, design a visualization system that supports solving this problem, validate the design, and reflect about lessons learned in order to refine visualization design guidelines.

[Design Study Methodology, Sedlmair et al., 2012]
When are Design Studies Appropriate?

We introduce two axes, task clarity and information location, as shown in Figure 1. The two axes can be used as a way to think and reason about potential contributions generated by design studies. Note that movement along one axis often causes similar movement along the other: increased task clarity can facilitate a better understanding of information location, and vice versa. The axes can also associate visualization with, and differentiate it from, more algorithmic approaches. The blue triangular area on the top right is also dangerous territory: automatic algorithmic solutions such as machine learning are not likely to be possible; we provide practical advice based on our own experience, and our own study of the field. The general layout of the framework is linear to suggest that one stage follows another. Certain actions rely on artifacts from earlier stages organized into three categories: a core phase depicting the analytical reasoning at the end. For each stage, we provide practical advice based on our own experience, and our own study of the field. The axes can also associate visualization with, and differentiate it from, more algorithmic approaches. The blue triangular area on the top right is also dangerous territory: automatic algorithmic solutions such as machine learning are not likely to be possible; we provide practical advice based on our own experience, and our own study of the field. The general layout of the framework is linear to suggest that one stage follows another. Certain actions rely on artifacts from earlier stages organized into three categories: a core phase depicting the analytical reasoning at the end. For each stage, we provide practical advice based on our own experience, and our own study of the field.
A Nine-Stage Framework

[Design Study Methodology, Sedlmair et al., 2012]
Collaborator Winnowing

- Initial conversation
- Further meetings
- Prototyping
- Full collaboration

Talk with many, stay with few!

[Design Study Methodology, Sedlmair et al., 2012]
Design Space: Think Broad

[Design Study Methodology, Sedlmair et al., 2012]
Analysis

• Want to contribute understanding of the way visualization can be used to address specific problem
• How does this relate to other situations?
• Does this align with existing guidelines? Does it challenge them?
• Is there anything that can still be improved?
Definition

“Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively”
Why Effectiveness?

• “It’s not just about pretty pictures”

• Any depiction of data requires the designer to make choices about how that data is visually represented
  - Analogy to photography
  - Lots of possibilities (see quarterback study)

• Effectiveness measures how well the visualization helps a person with their tasks
  - How? insight, engagement, efficiency?
  - Benchmarks and user studies
Design: Focus on only the y-axis

Average Annual Global Temperature in Fahrenheit
1880-2015

[S. Hayward, 2015]
Design: Year on the y-axis
Design: Different y-axis

Average Annual Global Temperature in Fahrenheit
1880-2015

[S. Hayward, 2015]
Analyzing Visualization

[Why?]

[How?]

[What?]

[Munzner (ill. Maguire), 2014]
What languages do we use on the Web?
Languages of the Web

- HTML
- CSS
- SVG
- JavaScript
  - Versions of Javascript: ES6, ES2015, ES2017…
  - Specific frameworks: react, jQuery, bootstrap, D3
Web Programming Tools

• Basic: Text editor and Modern Browser
• Developer Tools: Built in to browsers (e.g. Chrome Developer Tools)
• Web Environments: JSFiddle, Liveweave, CodePen, etc.
• IDEs: WebStorm, etc.
Hyper Text Markup Language (HTML)

• Markup languages allow users to encode the **semantics** of text
• Tags define the boundaries of the structures of the content
  - Tags are enclosed in angle brackets (e.g. `<html>`)
  - Most of the time, you have a start and end tag
  - End tags are just like start tags except that they have forward slash after the open bracket (e.g. `</html>`)
  - Tags may be nested but not mismatched
    • `<p>A `<strong>`<em>very</em>`</strong> cool example</p>`
    • `<p>A `<strong>`very `<em>cool</em>`</strong>` example</p>`
  - What about `<img src="mypicture.png" alt="My Image">`?
HTML Elements and Attributes

• Tags denote **elements** of the content (e.g. sections, paragraphs, images)

• Each element may have **attributes** which define other information about the element
  
  - An attribute has a **key** and **value** (*key=*“*value*”)
  
  - e.g. `<img src="mypicture.png" alt="My Image">`

• Many different elements available
  
  - Common: headers (h1, ..., h6), paragraph (p), lists (ul, ol, li), emphasis (em, strong), link (a), spans & divisions (span, div)
  
  - Lots more (e.g. abbr): see [MDN Documentation](https://developer.mozilla.org)

• Many different attributes available
  
  - See [MDN Documentation](https://developer.mozilla.org): note that some are legacy due to CSS
HTML Element Structure & Naming

• Elements structure a document
  - Document Object Model (DOM)
  - We can visualize this information
  - More importantly, we can navigate this tree

• Identifying and Classifying elements: id and class attributes
  - id identifies a **single** element—use for a unique case
  - class may identify **multiple** elements—use for common cases
  - Each element may have multiple classes, separate by spaces
  - Use normal identifiers: don’t start the name with a number
Other HTML Trivia

• `<`, `>`, `&`, and " are special characters, escape with `&lt;`, `&gt;`, `&amp;`, and `&quot;` (note the semi-colon)

• Comments are enclosed by `<!--` and `-->`
  - `<!-- This is a comment -->`

• HTML Documents begin with a DOCTYPE declaration
  - For HTML5, this is easier `<!DOCTYPE html>`

• meta tag: `<meta charset="UTF-8"/>`

• HTML has audio and video tags, math equation support, and more