Data Visualization (CIS 490/680)

Definition & Web Programming

Dr. David Koop
Exploration <-> Communication Spectrum

Consecutive Starts by a Quarterback for a Single Team

- Exploration
- Questions
- Communication
- Answers/Persuasion

[K. Quealy, 2013]
The Power of Interactive Visualization

[Music Timeline, Google Research]
Definition of Visualization

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

— T. Munzner
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.”
“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
Why Computers?

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

[Cerebral, Barsky et al., 2007]
Administrivia

- Course Web Site
- Syllabus
  - Plagiarism
  - Accommodations
- Textbook:
  - Required: Munzner (VAD)
  - Recommend: Murray, 2nd ed. (IDV)
- Assignments
- Exams: Midterm (Oct. 17) and Final (Dec. 11)
- Registration
Assignment 1
Do not cheat!
Do not plagiarize

• It is Academic Misconduct
• 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
Do not cheat!
Questions?
“Computer-based visualization systems provide visual representations of datasets designed to help people carry out tasks more effectively”
Why do we visualize data?

Figures are richer; provide more information with less clutter and in less space. Figures provide the gestalt effect: they give an overview; make structure more visible.

Figures are more accessible, easier to understand, faster to grasp, more comprehensible, more memorable, more fun, and less formal.

List adapted from: [Stasko et al. 1998]
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>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 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?

- $y_1$ vs. $x_1$
- $y_2$ vs. $x_2$
- $y_3$ vs. $x_3$
- $y_4$ vs. $x_4$

[F. J. Anscombe]
Why Visual?

![Graphs of four datasets with linear relationships and statistical properties:](image)

<table>
<thead>
<tr>
<th></th>
<th>x</th>
<th>y</th>
<th>Mean of x</th>
<th>Variance of x</th>
<th>Mean of y</th>
<th>Variance of y</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dataset 1</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>11</td>
<td>7.50</td>
<td>4.122</td>
<td>0.816</td>
</tr>
<tr>
<td>Dataset 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dataset 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dataset 4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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

[C. G. Healey]
Visual Perception Limitations
Visual Perception Limitations

[C. G. Healey]
Human Perception

[Inside NOVA: Change Blindness]
Human Perception

[Inside NOVA: Change Blindness]
Not Uncommon
Not Uncommon
Other Human Limitations

• Visual working memory is **small**

• **Change blindness:** A failure to notice a change in our view

• **Inattentional blindness:** A failure to notice something else going on in our view while focusing on a particular task

• "The goal of vision is not to build a complete photograph or model of the world in your mind. The goal of vision is to make sense of the meaning of the world around you." - D. Simons
Definition

“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</td>
</tr>
<tr>
<td>San Diego Chargers</td>
<td>Drew Brees, Drew Brees, Philip Rivers</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, 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, Michael Vick, Matt Ryan, Matt Ryan</td>
</tr>
<tr>
<td>New York Jets</td>
<td>Chad Pennington, Brett Favre, Mark Sanchez, Mark Sanchez</td>
</tr>
<tr>
<td>Cincinnati Bengals</td>
<td>Carson Palmer, Carson Palmer</td>
</tr>
<tr>
<td>Houston Texans</td>
<td>David Carr, Matt Schaub</td>
</tr>
<tr>
<td>Carolina Panthers</td>
<td>Jake Delhomme, Jake Delhomme</td>
</tr>
<tr>
<td>Denver Broncos</td>
<td>Jake Plummer, Jay Cutler, Kyle Orton, Tim Tebow, Peyton Manning</td>
</tr>
<tr>
<td>Arizona Cardinals</td>
<td>Matt Leininger, Kurt Warner</td>
</tr>
<tr>
<td>Jacksonville Jaguars</td>
<td>Byron Leftwich, David Garrard, David Garrard, Blaine Gabbert</td>
</tr>
<tr>
<td>Detroit Lions</td>
<td>Joey Harrington, Jon Kitna</td>
</tr>
<tr>
<td>Tampa Bay Buccaneers</td>
<td>Chris Simms, Bruce Grad, 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 (140)

[K. Quealy, 2013]
Another Design Example

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

[Another Design Example](https://example.com)  

[M. Stefaner, 2013](https://example.com)
Design Space: Think Broad

[Design Study Methodology, Sedlmair et al., 2012]
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
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: CodePen, JSFiddle, Liveweave, CodeSandbox, 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</strong></em> 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. \( <\text{img src="mypicture.png" alt="My Image">} \)

• 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

• Many different attributes available
  - See MDN Documentation: 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
Basic HTML File

<!DOCTYPE html>
<html>
  <head>
    <title>A Basic Web Page</title>
  </head>
  <body>
    <h1>My Wicked Awesome Web Page</h1>
    <p><em>This is <strong>cool</strong>. What about <u><strong>this?</strong></u></em></p>
  </body>
</html>

- https://codepen.io/dakoop/pen/PdRKEtL
What is CSS?
Cascading Style Sheets (CSS)

• Separate from content, just specifies how to style the content
• Style information can appear in three places:
  - External file
  - In a style element at the beginning of the HTML file
  - In a specific element in the body of a document (least preferable)
• Why Cascading?
  - Don’t want to have to specify everything over and over
  - Often want to use the same characteristics in a region of the DOM
  - Use inheritance: properties that apply to children cascade down
CSS Selectors

• How do we specify what part(s) of the page we want to style?
• The **element types** themselves (the HTML tag)
  - `strong { color: red; }`
• **Classes** of elements (ties to HTML `class` attribute)
  - `.cool { color: blue; }`
• A **specific** element (ties to HTML `id` attribute)
  - `#main-section { color: green; }`
• Relationships
  - Descendant: `p em { color: yellow; }`
  - Child: `p > em { color: orange; }`
• **Pseudo-classes**: `a:hover { color: purple; }`
Other CSS Bits

• Comments: /* This is a comment in CSS */

• Grouping Selectors: p, li { font-size: 12pt; }

• Multiple Classes: .cool.temp { color: blue; }

• Colors:
  - Names (Level 1, 2, & 3): red, orange, antiquewhite
  - Dash notation (3- & 6-character): #fff, #00ff00
  - Integer or % RGB and HSL Functions: rgb(255, 0, 0), rgb(50%, 50%, 0%), hsl(120, 100%, 50%)
  - Also background-color

• Watch out for multiple rules (look at how a web browser parses)

• Again, much more documentation at MDN
Example CSS

body {
    font-face: sans-serif;
    font-size: 12pt;
}
em { color: green; }
em u { color: red; }
em > strong { color: blue; }
img { border: 4px solid red; }

• What colors are displayed for this HTML (with the above stylesheet)?
  - <em>This is <strong>cool</strong>. What about <u><strong>this</strong></u>?</em>
• https://codepen.io/dakoop/pen/ErNJvJ
CSS Specificity

• Example:
  - CSS:
    
    ```
    p.exciting { color: red; }
    p { color: blue; }
    ```
  - What is the color of the paragraph
    ```
    <p class="exciting">Cool</p>
    ```

• Generally, last rule listed overrides previous rules
• …but anytime a selector is more specific, it has precedence
• p.exciting is a more specific selector
• When in doubt, test it in a browser
• https://codepen.io/dakoop/pen/MLbRQz
How to add CSS to HTML

• External: a separate file via a link element (in the <head> section):
  - `<link rel="stylesheet" href="styles.css">`

• Embedded: in the header:
  - `<style type="text/css"> … </style>`

• Inline: for a specific element: (Discouraged!)
  - `<p style="font-weight: bold;">Some text</p>`