Data Visualization (CIS 468)

Web Programming

Dr. David Koop
What is Data Visualization?
Exploration <-> Communication Spectrum

Consecutive Starts by a Quarterback for a Single Team

[K. Quealy, 2013]
Exploration <-> Communication Spectrum

Consecutive Starts by a Quarterback for a Single Team

Questions

Answers/Persuasion

[K. Quealy, 2013]
Administrivia

- Course Web Site
- Syllabus
  - Plagiarism
  - Accommodations
- Textbook:
  - Required: Munzner (VAD)
  - Rec'd: Murray, 2nd ed. (IDV)
- Assignments
- Registration:
  - Add/Drop is **Wednesday**
  - Make sure you are registered
Important Dates

• **Check these now!**

• Quizzes:
  - October 2 (in class)
  - November 20 (in class)

• Midterm: October 23 (in class)

• Final: December 13 (8-11am)
Accommodation Policy

• Please contact me at the **beginning** of the semester and provide the appropriate paperwork from the Center for Access and Success.

• Please update me if anything changes during the semester.

• Center for Access and Success: Pine Dale Hall Room 7136, x8711, access_success@umassd.edu
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.”
“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.”

NYC Subway Fare Data

Find Interesting NYC Subway Ridership Patterns
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

![Diagram of tree representations A, B, C]

- [McGuffin & Robert, 2010]
Definition

“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></th>
<th>II</th>
<th></th>
<th>III</th>
<th></th>
<th>IV</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>x</td>
<td>y</td>
<td>x</td>
<td>y</td>
<td>x</td>
<td>y</td>
<td>x</td>
<td>y</td>
<td></td>
</tr>
<tr>
<td>10.0</td>
<td>8.04</td>
<td>10.0</td>
<td>9.14</td>
<td>10.0</td>
<td>7.46</td>
<td>8.0</td>
<td>6.58</td>
<td></td>
</tr>
<tr>
<td>8.0</td>
<td>6.95</td>
<td>8.0</td>
<td>8.14</td>
<td>8.0</td>
<td>6.77</td>
<td>8.0</td>
<td>5.76</td>
<td></td>
</tr>
<tr>
<td>13.0</td>
<td>7.58</td>
<td>13.0</td>
<td>8.74</td>
<td>13.0</td>
<td>12.74</td>
<td>8.0</td>
<td>7.71</td>
<td></td>
</tr>
<tr>
<td>9.0</td>
<td>8.81</td>
<td>9.0</td>
<td>8.77</td>
<td>9.0</td>
<td>7.11</td>
<td>8.0</td>
<td>8.84</td>
<td></td>
</tr>
<tr>
<td>11.0</td>
<td>8.33</td>
<td>11.0</td>
<td>9.26</td>
<td>11.0</td>
<td>7.81</td>
<td>8.0</td>
<td>8.47</td>
<td></td>
</tr>
<tr>
<td>14.0</td>
<td>9.96</td>
<td>14.0</td>
<td>8.10</td>
<td>14.0</td>
<td>8.84</td>
<td>8.0</td>
<td>7.04</td>
<td></td>
</tr>
<tr>
<td>6.0</td>
<td>7.24</td>
<td>6.0</td>
<td>6.13</td>
<td>6.0</td>
<td>6.08</td>
<td>8.0</td>
<td>5.25</td>
<td></td>
</tr>
<tr>
<td>4.0</td>
<td>4.26</td>
<td>4.0</td>
<td>3.10</td>
<td>4.0</td>
<td>5.39</td>
<td>19.0</td>
<td>12.50</td>
<td></td>
</tr>
<tr>
<td>12.0</td>
<td>10.84</td>
<td>12.0</td>
<td>9.13</td>
<td>12.0</td>
<td>8.15</td>
<td>8.0</td>
<td>5.56</td>
<td></td>
</tr>
<tr>
<td>7.0</td>
<td>4.82</td>
<td>7.0</td>
<td>7.26</td>
<td>7.0</td>
<td>6.42</td>
<td>8.0</td>
<td>7.91</td>
<td></td>
</tr>
<tr>
<td>5.0</td>
<td>5.68</td>
<td>5.0</td>
<td>4.74</td>
<td>5.0</td>
<td>5.73</td>
<td>8.0</td>
<td>6.89</td>
<td></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?

\[ F. J. \text{ Anscombe} \]
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
- **Inattentitional 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
“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>Players</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, Philip Rivers</td>
</tr>
<tr>
<td>Baltimore Ravens</td>
<td>Kyle Boller, 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</td>
</tr>
<tr>
<td>Atlanta Falcons</td>
<td>Michael Vick, Matt Ryan</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 Fitz, Carson Palmer, Andy Dalton</td>
</tr>
<tr>
<td>Houston Texans</td>
<td>David Carr, Matt Schaub</td>
</tr>
<tr>
<td>Carolina Panthers</td>
<td>Jake Delhomme, Cam Newton</td>
</tr>
<tr>
<td>Denver Broncos</td>
<td>Jake Plummer, Kyle Orton, Tim Tebow, Peyton Manning</td>
</tr>
<tr>
<td>Arizona Cardinals</td>
<td>Matt Leinart, Kurt Warner</td>
</tr>
<tr>
<td>Jacksonville Jaguars</td>
<td>Byron Leftwich, David Garrard, Blaine Gabbert</td>
</tr>
<tr>
<td>Detroit Lions</td>
<td>Joey Harrington, Matthew Stafford</td>
</tr>
<tr>
<td>Tampa Bay Buccaneers</td>
<td>Chris Simms, Bruce Gradkowski, Jeff Garcia</td>
</tr>
<tr>
<td>Dallas Cowboys</td>
<td>Drew Bledsoe, Tony Romo, 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 frequented 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]
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
  - Insight, engagement, efficiency
  - Human visual processing is not foolproof…
  - Use benchmarks and user studies
Simple World Assumption

[D. Szafir, 2017]
Problems with Simple World Assumption

Visualizations violate three CIELAB assumptions:

- Isolation Assumption
- Geometric Assumption


[D. Szafir, 2017]
Isolation Assumption

[D. Szafir, 2017]
What colors are in this graphic?
What colors are in this graphic?

Red, yellow, blue

Purple, orange do not exist!

[A. Kitaoka]
How do we create modern visualizations?
Tools

• Desktop Applications:
  - Excel (see excelcharts.com)
  - Tableau
  - …

• Programming Frameworks/Languages
  - Processing
  - d3.js
  - deck.gl, MapboxGL
  - vega-lite, …

• Advantages and disadvantages
  - Speed, customization, understanding, dissemination
D3.js is a JavaScript library for manipulating documents based on data. D3 helps you bring data to life using HTML, SVG, and CSS. D3’s emphasis on web standards gives you the full capabilities of modern browsers without tying yourself to a proprietary framework, combining powerful visualization components and a data-driven approach to DOM manipulation.
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, Observable, etc.
• IDEs: WebStorm, etc.
What is HTML?
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>` 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/en-US/docs/Web/HTML/Element)
- Many different attributes available
  - See [MDN Documentation](https://developer.mozilla.org/en-US/docs/Web/HTML/Attribute): 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
<!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>
    <img src="patriots.jpg"/>
  </body>
</html>
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