Data Visualization (DSC 530/CIS 568)

Definition & Web Programming

Dr. David Koop
Exploration <-> Communication Spectrum

Consecutive Starts by a Quarterback for a Single Team

[Image of a graph showing the number of games started over time for different teams, with labels for various quarterbacks and their start streaks.]

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

Consecutive Starts by a Quarterback for a Single Team

Questions

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.”
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

[McGuffin & Robert, 2010]
Administrivia

- Course Web Site
- Syllabus
  - Plagiarism
  - Accommodations
- Textbook:
  - Required: Munzner (VAD)
  - Rec'd: Murray, 2nd ed. (IDV)
- Assignments
- Tests: March 6 and April 10
- Registration:
  - Add/Drop is Wednesday
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
Definition

“Computer-based visualization systems provide **visual** representations of datasets designed to help people carry out tasks more effectively”
Why do we visualize data?

**Total Bandwidth**
(millions of bits per second)

0 1 2 3 4 5 6 7 8 9 10

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]

[viA. Lex]
### 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?

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>II</td>
<td>III</td>
<td>IV</td>
<td></td>
</tr>
<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>

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]
Human Perception

[Inside NOVA: Change Blindness]
Human Perception
Not Uncommon

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

- Visual working memory is small
- Change blindness: A failure to notice a change in our view
- Inattentinal 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, Peyton 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, Andrew Luck</td>
</tr>
<tr>
<td>Baltimore Ravens</td>
<td>Kyle Bolin, 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, Aaron Rodgers</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 Fitzp, 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 Leina, 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, Jon Kitna, Matthew Stafford</td>
</tr>
<tr>
<td>Tampa Bay Buccaneers</td>
<td>Chris Simms, Bruce Gra, Josh Freeman, Josh Freeman</td>
</tr>
<tr>
<td>Dallas Cowboys</td>
<td>Drew Bledsoe, Tony Romo, Tony Romo</td>
</tr>
</tbody>
</table>
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

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

We introduce two axes, task clarity and information location, as shown in Figure 1. When we move from left to right, we move from low-level tasks to high-level tasks. When we move vertically, we move from abstract tasks to concrete tasks.

3.2 Task Clarity and Information Location Axes

The red and the blue areas at the periphery represent situations for which design study methodology is a suitable choice. This rough characterization is not applicable in all cases, and it is not always possible to determine whether design study methodology is suitable just by looking at a single task. However, we can use this characterization to guide our decisions about whether to use design study methodology in a particular situation.

3.3 Design Study Methodology Suitability

In Section 4.1.2, we describe how to use the axes to determine whether design study methodology is suitable. The axes can also associate visualization with, and differentiate it from, other methods. For example, we can use the axes to determine whether a design study is appropriate when a researcher is trying to develop a new visualization tool. The axes can also help us determine whether a design study is appropriate when a researcher is trying to improve an existing visualization tool.

4.1.2 Design Study Methodology Suitability

The axes can also be used to determine whether design study methodology is suitable when a researcher is trying to develop a new visualization tool. For example, we can use the axes to determine whether a design study is appropriate when a researcher is trying to develop a new visualization tool for a particular domain. The axes can also be used to determine whether a design study is suitable when a researcher is trying to improve an existing visualization tool. For example, we can use the axes to determine whether a design study is appropriate when a researcher is trying to improve an existing visualization tool for a particular domain.

Figure 1: Design Study Methodology Suitability

[Design Study Methodology, Sedlmair et al., 2012]
A Nine-Stage Framework

PRECONDITION  
personal validation

CORE  
inward-facing validation

ANALYSIS  
outward-facing validation

[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
Simple World Assumption

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

- Viewing Distance
- Environmental Surround
- Ambient Illumination
- Direct Illumination
- Viewing Population
- Gamma, Whitepoint, Resolution, Peak Color Outputs

Visualizations violate three CIELAB assumptions:

- Isolation Assumption
- Geometric Assumption

Crowdsourced Sampling

- Szafir, Stone, & Gleicher, 2014
- Reinecke, Flatla, & Brooks, 2016

[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]
Analyzing Visualization

Why?
How?
What?

Why?
How?
What?

Why?
How?
What?

Dependency

[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