The Value of Visualization

Jeffrey Heer  University of Washington
How much data (bytes) did we produce in 2010?
2010: 1,200 exabytes
10x increase over 5 years

Gantz et al, 2008, 2010
Physical Sensors

Image courtesy cabspotting.org
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart Rate (HR)</td>
<td>97 bpm</td>
</tr>
<tr>
<td>Systolic Blood</td>
<td>82 mmHg</td>
</tr>
<tr>
<td>Diastolic Blood</td>
<td>60 mmHg</td>
</tr>
<tr>
<td>SpO2</td>
<td>99%</td>
</tr>
<tr>
<td>RR/C02</td>
<td>1/1 min</td>
</tr>
<tr>
<td>Heart Rate (HR)</td>
<td>79 bpm</td>
</tr>
<tr>
<td>Systolic Blood</td>
<td>152 mmHg</td>
</tr>
<tr>
<td>Diastolic Blood</td>
<td>79 mmHg</td>
</tr>
<tr>
<td>SpO2</td>
<td>95%</td>
</tr>
<tr>
<td>RR/C02</td>
<td>1/1 min</td>
</tr>
<tr>
<td>Heart Rate (HR)</td>
<td>64 bpm</td>
</tr>
<tr>
<td>Systolic Blood</td>
<td>93 mmHg</td>
</tr>
<tr>
<td>Diastolic Blood</td>
<td>55 mmHg</td>
</tr>
<tr>
<td>SpO2</td>
<td>99%</td>
</tr>
<tr>
<td>RR/Imp</td>
<td>1/1 min</td>
</tr>
</tbody>
</table>
Records of Human Activity
Wikipedia History Flow (IBM)
The ability to take data—to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it—that’s going to be a hugely important skill in the next decades, … because now we really do have essentially free and ubiquitous data. So the complimentary scarce factor is the ability to understand that data and extract value from it.

Hal Varian, Google’s Chief Economist

The McKinsey Quarterly, Jan 2009
A Poverty of Attention

“What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the overabundance of information sources that might consume it.”

Herb Simon
as quoted by Hal Varian
Scientific American
September 1995
What is Visualization?

“Transformation of the symbolic into the geometric”
[McCormick et al. 1987]

“... finding the artificial memory that best supports our natural means of perception.” [Bertin 1967]

“The use of computer-generated, interactive, visual representations of data to amplify cognition.”
[Card, Mackinlay, & Shneiderman 1999]
<table>
<thead>
<tr>
<th>Set A</th>
<th>Set B</th>
<th>Set C</th>
<th>Set D</th>
</tr>
</thead>
<tbody>
<tr>
<td>X Y</td>
<td>X Y</td>
<td>X Y</td>
<td>X Y</td>
</tr>
<tr>
<td>10 8.04</td>
<td>10 9.14</td>
<td>10 7.46</td>
<td>8 6.58</td>
</tr>
<tr>
<td>8 6.95</td>
<td>8 8.14</td>
<td>8 6.77</td>
<td>8 5.76</td>
</tr>
<tr>
<td>13 7.58</td>
<td>13 8.74</td>
<td>13 12.74</td>
<td>8 7.71</td>
</tr>
<tr>
<td>9 8.81</td>
<td>9 8.77</td>
<td>9 7.11</td>
<td>8 8.84</td>
</tr>
<tr>
<td>11 8.33</td>
<td>11 9.26</td>
<td>11 7.81</td>
<td>8 8.47</td>
</tr>
<tr>
<td>14 9.96</td>
<td>14 8.1</td>
<td>14 8.84</td>
<td>8 7.04</td>
</tr>
<tr>
<td>6 7.24</td>
<td>6 6.13</td>
<td>6 6.08</td>
<td>8 5.25</td>
</tr>
<tr>
<td>4 4.26</td>
<td>4 3.1</td>
<td>4 5.39</td>
<td>19 12.5</td>
</tr>
<tr>
<td>12 10.84</td>
<td>12 9.11</td>
<td>12 8.15</td>
<td>8 5.56</td>
</tr>
<tr>
<td>7 4.82</td>
<td>7 7.26</td>
<td>7 6.42</td>
<td>8 7.91</td>
</tr>
<tr>
<td>5 5.68</td>
<td>5 4.74</td>
<td>5 5.73</td>
<td>8 6.89</td>
</tr>
</tbody>
</table>

**Summary Statistics**

\[
\begin{align*}
u_X &= 9.0 & \sigma_X &= 3.317 \\
u_Y &= 7.5 & \sigma_Y &= 2.03
\end{align*}
\]

**Linear Regression**

\[
Y = 3 + 0.5 \, X
\]

\[
R^2 = 0.67
\]

[Anscombe 1973]
Why Create Visualizations?
Why Create Visualizations?

Answer questions (or discover them)
Make decisions
See data in context
Expand memory
Support graphical calculation
Find patterns
Present argument or tell a story
Inspire
Record Information
Gallop, Bay Horse "Daisy" [Muybridge 1884-86]
1. Marey’s sphygmograph in use, 1860. La méthode graphique dans les sciences expérimentales et principalement en physiologie et en médecine.
Support Reasoning
### HISTORY OF O-RING DAMAGE ON SRM FIELD JOINTS

<table>
<thead>
<tr>
<th>SRM No.</th>
<th>Erosion Depth (in.)</th>
<th>Perimeter Affected (deg)</th>
<th>Nominal Dia. (in.)</th>
<th>Length Of Max Erosion (in.)</th>
<th>Total Heat Affected Length (in.)</th>
<th>Clamping Location (deg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>22A</td>
<td>None</td>
<td>None</td>
<td>0.280</td>
<td>None</td>
<td>None</td>
<td>36° – 66°</td>
</tr>
<tr>
<td>22A</td>
<td>None</td>
<td>NONE</td>
<td>0.280</td>
<td>NONE</td>
<td>NONE</td>
<td>336° – 18°</td>
</tr>
<tr>
<td>15A</td>
<td>0.010</td>
<td>154.0</td>
<td>0.280</td>
<td>4.25</td>
<td>5.25</td>
<td>163</td>
</tr>
<tr>
<td>15B</td>
<td>0.038</td>
<td>130.0</td>
<td>0.280</td>
<td>12.50</td>
<td>58.75</td>
<td>354</td>
</tr>
<tr>
<td>15B</td>
<td>None</td>
<td>45.0</td>
<td>0.280</td>
<td>None</td>
<td>29.50</td>
<td>354</td>
</tr>
<tr>
<td>15B</td>
<td>0.028</td>
<td>110.0</td>
<td>0.280</td>
<td>3.00</td>
<td>None</td>
<td>275</td>
</tr>
<tr>
<td>11A</td>
<td>None</td>
<td>None</td>
<td>0.280</td>
<td>None</td>
<td>None</td>
<td>--</td>
</tr>
<tr>
<td>10A</td>
<td>0.040</td>
<td>217.0</td>
<td>0.280</td>
<td>3.00</td>
<td>14.50</td>
<td>351</td>
</tr>
<tr>
<td>2B</td>
<td>0.053</td>
<td>116.0</td>
<td>0.280</td>
<td>--</td>
<td>--</td>
<td>90</td>
</tr>
</tbody>
</table>

*Hot gas path detected in putty. Indication of heat on O-ring, but no damage.
**Soot behind primary O-ring.
***Soot behind primary O-ring, heat affected secondary O-ring.

Clocking location of leak check port – 0 deg.

Other SRM-15 field joints had no blowholes in putty and no soot near or beyond the primary O-ring.

SRM-22 forward field joint had putty path to primary O-ring, but no O-ring erosion and no soot blowby. Other SRM-22 field joints had no blowholes in putty.

#### Blow By History

**SRM-15 Worst Blow By**
- 2 case joints (50°, 110°) arc
- Much worse visually than SRM-22

**SRM-22 Blow By**
- 2 case joints (30°-40°)

**SRM-13A, 15, 16A, 18, 23A, 24A**
- Nozzle Blow By

#### HISTORY OF O-RING TEMPERATURES (DEGREES F)

<table>
<thead>
<tr>
<th>MOTOR</th>
<th>MMT</th>
<th>AMB</th>
<th>O-RING</th>
<th>WIND</th>
</tr>
</thead>
<tbody>
<tr>
<td>DM-1</td>
<td>68</td>
<td>36</td>
<td>47</td>
<td>10 mph</td>
</tr>
<tr>
<td>DM-2</td>
<td>76</td>
<td>45</td>
<td>52</td>
<td>10 mph</td>
</tr>
<tr>
<td>GM-3</td>
<td>72.5</td>
<td>40</td>
<td>48</td>
<td>10 mph</td>
</tr>
<tr>
<td>GM-4</td>
<td>76</td>
<td>48</td>
<td>51</td>
<td>10 mph</td>
</tr>
<tr>
<td>SRM-15</td>
<td>52</td>
<td>64</td>
<td>53</td>
<td>10 mph</td>
</tr>
<tr>
<td>SRM-22</td>
<td>77</td>
<td>78</td>
<td>75</td>
<td>10 mph</td>
</tr>
<tr>
<td>SRM-25</td>
<td>55</td>
<td>26</td>
<td>29</td>
<td>25 mph</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Make a Decision: Challenger
Make a Decision: Challenger

Visualizations drawn by Tufte show how low temperatures damage O-rings [Tufte 97]
In 1854 John Snow plotted the position of each cholera case on a map. [from Tufte 83]
Data in Context: Cholera Outbreak

Used map to hypothesize that pump on Broad St. was the cause. [from Tufte 83]
Expand Memory: Multiplication

Class Exercise!
Expand Memory: Multiplication

34
x 72
Expand Memory: Multiplication

34
x 72
---
68
2380
2448

Time (Sec.)
0 28 55 83 110

- Mental
- Paper & Pencil
Find Patterns: NYC Weather

The Most Powerful Brain?

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
<th>Body Weight</th>
<th>Brain Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Lesser Short-tailed Shrew</td>
<td>5</td>
<td>0.14</td>
</tr>
<tr>
<td>2</td>
<td>Little Brown Bat</td>
<td>10</td>
<td>0.25</td>
</tr>
<tr>
<td>3</td>
<td>Mouse</td>
<td>23</td>
<td>0.3</td>
</tr>
<tr>
<td>4</td>
<td>Big Brown Bat</td>
<td>23</td>
<td>0.4</td>
</tr>
<tr>
<td>5</td>
<td>Musk Shrew</td>
<td>48</td>
<td>0.33</td>
</tr>
<tr>
<td>6</td>
<td>Star Nosed Mole</td>
<td>60</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>Eastern American Mole</td>
<td>75</td>
<td>1.2</td>
</tr>
<tr>
<td>8</td>
<td>Ground Squirrel</td>
<td>101</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>Tree Shrew</td>
<td>104</td>
<td>2.5</td>
</tr>
<tr>
<td>10</td>
<td>Golden Hamster</td>
<td>120</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>Mole Rate</td>
<td>122</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>Galago</td>
<td>200</td>
<td>5</td>
</tr>
<tr>
<td>13</td>
<td>Rat</td>
<td>280</td>
<td>1.9</td>
</tr>
<tr>
<td>14</td>
<td>Chinchilla</td>
<td>425</td>
<td>6.4</td>
</tr>
<tr>
<td>15</td>
<td>Desert Hedgehog</td>
<td>550</td>
<td>2.4</td>
</tr>
<tr>
<td>16</td>
<td>Rock Hyrax (a)</td>
<td>750</td>
<td>12.3</td>
</tr>
<tr>
<td>17</td>
<td>European Hedgehog</td>
<td>785</td>
<td>3.5</td>
</tr>
<tr>
<td>18</td>
<td>Tenrec</td>
<td>900</td>
<td>2.6</td>
</tr>
<tr>
<td>19</td>
<td>Arctic Ground Squirrel</td>
<td>920</td>
<td>5.7</td>
</tr>
<tr>
<td>20</td>
<td>African Giant Pouched Rat</td>
<td>1000</td>
<td>6.6</td>
</tr>
<tr>
<td>21</td>
<td>Guinea Pig</td>
<td>1040</td>
<td>5.5</td>
</tr>
<tr>
<td>22</td>
<td>Mountain Beaver</td>
<td>1350</td>
<td>8.1</td>
</tr>
<tr>
<td>23</td>
<td>Slow Loris</td>
<td>1400</td>
<td>12.5</td>
</tr>
<tr>
<td>24</td>
<td>Genet</td>
<td>1410</td>
<td>17.5</td>
</tr>
<tr>
<td>25</td>
<td>Phalanger</td>
<td>1620</td>
<td>11.4</td>
</tr>
</tbody>
</table>
The Elements of Graphing Data
[Cleveland]
Convey Information to Others
Inspire

Bones in hand [from 1918 edition]

Double helix model [Watson and Crick 53]
“to affect thro’ the Eyes what we fail to convey to the public through their word-proof ears”

1856 “Coxcomb” of Crimean War Deaths, Florence Nightingale
The Value of Visualization

**Record** information
- Blueprints, photographs, seismographs, ...

**Analyze** data to support reasoning
- Develop and assess hypotheses
- Find patterns / Discover errors in data
- Expand memory

**Communicate** information to others
- Share and persuade
- Collaborate and revise
Goals of Visualization Research

1. **Understand** how visualizations convey information
   - What do people perceive/comprehend?
   - How do visualizations inform mental models?

2. **Develop principles and techniques** for creating effective visualizations and supporting analysis
   - Leverage perception & augment cognition
   - Improve ties between visualization & mental model
Course Topics
Data and Image Models

Sémiologie Graphique [Bertin 67]
Visualization (Re-)Design

Problematic design

Redesign
Exploratory Data Analysis
Visualization Software

D3: Data-Driven Documents
Interaction

Friday, December 12, 2008
154 reports

Crimespotting.org
Graphical Perception

The psychophysics of sensory function [Stevens 61]
CIE 1931 xy chromaticity diagram showing the gamut of the sRGB and Adobe RGB color spaces including the Planckian locus, with temperatures indicated. Wavelengths of monochromatic light are shown in blue.
Animation

Animated transitions in statistical data graphics [Heer & Robertson 07]
Networks

Zephoria

User ID: 21721
Friends: 286
Age: N/A
Gender: Female
Status: Single
Location: San Francisco, CA
Hometown: Lancaster, PA
Occupation: researcher, social networks, identity, context
Interests: apophenia, observing people, culture, questioning power, reading, Buddhism, Moby, computer-mediated communication, social networks, technology, anthropology, standing
TV Shows: ??
Member Since: ??
Last Login: 2003-12-21
Last Updated: 2003-12-21
About: Some know me as danah... I'm a geek, an activist and an academic, fascinated by people and society. I see life as a very large playground and enjoy exploring its intricacies. I revel in the chaos, while simultaneously providing my own insane element.

My musings:
http://www.zephoria.org/thoughts/

Want to Meet: Someone who makes life's complex less so or simply elegant.
A partner in understanding...
Visualizations : Word tree / Alberto Gonzales

Creator: Martin Wattenberg
Tags:

Search: i don't

Data source: CQ Transcript Wire via the Washington Post

Comments (4)
Uncertainty
Recent elections have placed a heavy emphasis on “swing states” — Ohio, Florida and the other competitive states. You can see how much these states have shifted between the Democratic and Republican parties. A look at how the states have shifted over past elections.

Obama Re-elected
The country voted about 5 percentage points more Republican in 2012 than in 2008. Obama lost North Carolina and Indiana, but won every tossup except Florida, which remains too close to call.

As Goes Ohio
Ohio, which has voted for the winner in every election since 1964, provided the decisive electoral votes in 2004, and it is the state likeliest to play that role again this year, according to the FiveThirtyEight model.
Course Mechanics
You should expect to:

1. Evaluate and critique visualization designs
2. Learn visualization techniques & theory
3. Implement interactive data visualizations
4. Develop a substantial visualization project
Instructors

Jeffrey Heer  
Assoc Prof, CSE  
OH: Tu 10:15-11:15a, 642 CSE  
http://jheer.org

Assistants

Leilani Battle  
OH: Fri 3:30-4:30p, CSE 2nd Floor

Shobhit Hathi  
OH: Mon 1-2p, CSE 5th Floor

Halden Lin  
OH: Mon 1-2p, CSE 5th Floor
Leilani Battle
Office Hours 3:30-4:30pm Fridays
CSE 2nd Floor Breakout
(Out of town 4/20)

I’m a postdoc working on making big data analysis systems fast and easy to use. I often build user behavior models for visualization prediction and recommendation.
Hi! I’m a combined BS/MS student in my first quarter as a graduate student. I took data visualization last Spring, and I’m really excited to be a part of 512 this quarter!

My office hours will be 1-2 Mondays in the 5th floor breakout.
Hello!

I'm a combined B.S./M.S. Computer Science student with research interests in automated visualization design, currently working with Kanit 'Ham' Wongsuphasawat and Dominik Moritz in Jeff's lab (IDL).

I'm also a huge fan of visual and user experience design.

Always happy to chat.
Textbooks

See also: www.edwardtufte.com
Optional Textbook

Interactive Data Visualization for the Web, 2nd Edition

For learning D3!
Book available online.
Code / examples on GitHub.

We will be using D3 v5.
https://d3js.org
Readings

Some from textbooks, also many articles. Material in class will loosely follow readings. Readings should be read by start of class. Post discussion comments on class Canvas forum.

One comment per week (up through week 8). Comments must be posted by Friday 11:59pm. You have 1 “pass” for the quarter.
Assignments

Class Participation (10%)

A1: Visualization Design (10%) - Due 4/2

A2: Exploratory Data Analysis (15%) - Due 4/13

A3: Interactive Prototype (25%) - Due 4/30

Peer Evaluation - Due 5/7

FP: Final Project (40%)

Proposal - Due 5/10

Initial Prototype - Due 5/21

Project Deliverables - Due 5/30 (tentative)
Final Project

Visualization research project on topic of choice

Initial prototype and peer evaluation

Design reviews and final presentation

Submit and publish on GitHub

Projects from previous classes have been:

• Published as research papers
• Featured in the New York Times
• Released as successful open source projects
Visualizing Galaxy Merger Trees

S. Loebman, J. Ortiz, L. Orr, M. Balazinska, T. Quinn et al. [SIGMOD '14]
Perfopticon Distributed Query Performance

Dominik Moritz et al. [EuroVis ’15]
Protovis: A Graphical Toolkit for Visualization

Mike Bostock
var army = pd.nest(napoleon.army, "dir", "group");
var vis = new pv.Panel();

var lines = vis.add(pv.Panel).data(army);
lines.add(pv.Line)
  .data(function() army[this.idx])
  .left(lon).top(lat).size(function(d) d.size/8000)
  .strokeStyle(function() color[army[paneIndex][0].dir]);

vis.add(pv.Label).data(napoleon.cities)
  .left(lon).top(lat)
  .text(function(d) d.city).font("italic 10px Georgia")
  .textAlign("center").textBaseline("middle");

vis.add(pv.Label).data(napoleon.temp)
  .left(lon).top(tmp).strokeStyle("#0")
  .add(pv.Label)
  .top(function(d) 5 + tmp(d))
  .text(function(d) d.temp+"° " + d.date.substr(0,6))
Visualizing the Republic of Letters

Daniel Chang, Yuankai Ge, Shiwei Song

Republic of Letters

1700

Filter by Author

- Damien Desormes
- Daniel Cornabs
- Daniel de Pury
- Daniel Defoe
- Daniel Mathus
- Daniel Marc Antoine Chardon
- Daniel Muller

Top Cities and Authors

- London: 346 letters received, 357 letters sent
- Oates: 304 letters received, 250 letters sent
- Dublin: 208 letters received, 154 letters sent
- Paris: 238 letters received, 112 letters sent
- Twickenham: 18 letters received, 101 letters sent
- John Locke: 350 letters received, 253 letters sent
- Joseph Addison: 30 letters received, 244 letters sent
- Voltaire: 26 letters received, 231 letters sent
- Jonathan Swift: 28 letters received, 159 letters sent
- Alexander Pope: 28 letters received, 150 letters sent
Questions?
A1: Visualization Design

Design a static visualization for a data set.

Every 10 years, the census documents the demographic make-up of the U.S., influencing congressional districting and social services. This dataset contains a summary of census data for two years a century apart: 1900 and 2000.

You must choose the message you want to convey. What question(s) do you want to answer? What insight do you want to communicate?
A1: Visualization Design

Pick a **guiding question**, use it to title your vis. Design a **static visualization** for that question. You are free to **use any tools** (inc. pen & paper).

**Deliverables** (upload via Canvas; see A1 page)
- Image of your visualization (PNG or JPG format)
- Short description + design rationale (≤ 4 paragraphs)

Due by **8:00 pm, Monday April 2**.