CSE 512 - Data Visualization

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
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*
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]
## Summary Statistics

<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</td>
<td>Y</td>
<td>X</td>
<td>Y</td>
</tr>
<tr>
<td>10</td>
<td>8.04</td>
<td>10</td>
<td>9.14</td>
</tr>
<tr>
<td>8</td>
<td>6.95</td>
<td>8</td>
<td>8.14</td>
</tr>
<tr>
<td>13</td>
<td>7.58</td>
<td>13</td>
<td>8.74</td>
</tr>
<tr>
<td>9</td>
<td>8.81</td>
<td>9</td>
<td>8.77</td>
</tr>
<tr>
<td>11</td>
<td>8.33</td>
<td>11</td>
<td>9.26</td>
</tr>
<tr>
<td>14</td>
<td>9.96</td>
<td>14</td>
<td>8.1</td>
</tr>
<tr>
<td>6</td>
<td>7.24</td>
<td>6</td>
<td>6.13</td>
</tr>
<tr>
<td>4</td>
<td>4.26</td>
<td>4</td>
<td>3.1</td>
</tr>
<tr>
<td>12</td>
<td>10.84</td>
<td>12</td>
<td>9.11</td>
</tr>
<tr>
<td>7</td>
<td>4.82</td>
<td>7</td>
<td>7.26</td>
</tr>
<tr>
<td>5</td>
<td>5.68</td>
<td>5</td>
<td>4.74</td>
</tr>
</tbody>
</table>

\[ \bar{X} = 9.0 \quad \sigma_X = 3.317 \]

\[ \bar{Y} = 7.5 \quad \sigma_Y = 2.03 \]

## Linear Regression

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

\[ R^2 = 0.67 \]

[Anscombe 1973]
Set A

Set B

Set C

Set D
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]
E.J. Marey’s sphygmograph [from Braun 83]
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>Clocking Location (deg.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>61A LH Center Field**</td>
<td>22A</td>
<td>None</td>
<td>None</td>
<td>0.220</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>61A LH CENTER FIELD**</td>
<td>22A</td>
<td>None</td>
<td>None</td>
<td>0.220</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>51C LH Forward Field**</td>
<td>15A</td>
<td>0.010</td>
<td>104.0</td>
<td>0.280</td>
<td>12.50</td>
<td>50.75</td>
</tr>
<tr>
<td>51C LH Forward Field**</td>
<td>15B</td>
<td>0.038</td>
<td>130.0</td>
<td>0.280</td>
<td>12.50</td>
<td>50.75</td>
</tr>
<tr>
<td>51C LH Forward Field**</td>
<td>15B</td>
<td>0.040</td>
<td>110.0</td>
<td>0.280</td>
<td>3.00</td>
<td>None</td>
</tr>
<tr>
<td>41D RH Forward Field</td>
<td>18</td>
<td>None</td>
<td>None</td>
<td>0.280</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>41D RH Long Field*</td>
<td>11A</td>
<td>None</td>
<td>None</td>
<td>0.280</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>41B RH Forward Field</td>
<td>10A</td>
<td>0.053</td>
<td>116.0</td>
<td>0.280</td>
<td>3.00</td>
<td>14.50</td>
</tr>
<tr>
<td>STS-2 RH Aft Field</td>
<td>2B</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</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 seal 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
- 10 Case Joints (50°, 110°) Arc
- Much worse visually than SRM-22

SRM-22 Blow By
- 10 Case Joints (30°-40°)
- Nozzle Blow By

---

**History of O-Ring Temperatures (Degrees - F)**

<table>
<thead>
<tr>
<th>SRM</th>
<th>Motor</th>
<th>MHT</th>
<th>Amb</th>
<th>O-Ring</th>
<th>Wind</th>
</tr>
</thead>
<tbody>
<tr>
<td>DM-1</td>
<td>67</td>
<td>36</td>
<td>47</td>
<td>10 mph</td>
<td></td>
</tr>
<tr>
<td>DM-2</td>
<td>76</td>
<td>45</td>
<td>52</td>
<td>10 mph</td>
<td></td>
</tr>
<tr>
<td>DM-3</td>
<td>72.5</td>
<td>40</td>
<td>48</td>
<td>10 mph</td>
<td></td>
</tr>
<tr>
<td>DM-4</td>
<td>76</td>
<td>45</td>
<td>52</td>
<td>10 mph</td>
<td></td>
</tr>
<tr>
<td>SRM-15</td>
<td>52</td>
<td>64</td>
<td>53</td>
<td>10 mph</td>
<td></td>
</tr>
<tr>
<td>SRM-22</td>
<td>77</td>
<td>78</td>
<td>75</td>
<td>10 mph</td>
<td></td>
</tr>
<tr>
<td>SRM-25</td>
<td>55</td>
<td>26</td>
<td>29</td>
<td>10 mph</td>
<td></td>
</tr>
<tr>
<td>SRM-26</td>
<td>27</td>
<td>10 MPH</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

2 of 13 pages of material faxed to NASA by Morton Thiokol [from Tufte 1997]
Make a Decision: Challenger
Visualizations drawn by Tufte show how low temperatures damage O-rings [Tufte 97]
Data in Context: Cholera Outbreak

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 \times 72

\[ \begin{align*}
68 \\
2380 \\
2448
\end{align*} \]

<table>
<thead>
<tr>
<th>Time (Sec.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>110</td>
</tr>
<tr>
<td>83</td>
</tr>
<tr>
<td>55</td>
</tr>
<tr>
<td>28</td>
</tr>
<tr>
<td>0</td>
</tr>
</tbody>
</table>

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”
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
Visualization Research
Challenge

More and more unseen data
Faster creation and collection
Challenge

More and more unseen data
Faster creation and collection

Urban development planning
www.urbansim.org

Fluid flow
ctr.stanford.edu
Challenge

More and more unseen data
Faster creation and collection

Sloan digital sky survey  
www.sdss.org

Sensor networks  [Hill 02]  
www.xbow.com

Digital photography

Sensing
Challenge

More and more unseen data
Faster creation and collection; Faster dissemination

Photo sharing/annotation
flickr.com

Group Authoring
wikipedia.org

Map of the Internet [Cheswick 99]
research.lumeta.com

Internet
Challenge

More and more unseen data
Faster creation and collection; Faster dissemination

5 exabytes of new information in 2002 [Lyman 03]
161 exabytes in 2006 [Gantz 07]
1,200 exabytes in 2010 [Gantz 10]

Necessitates **better tools and algorithms** for visually conveying information
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
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
   Amplify perception and cognition
   Improve ties between visualization & mental model
Course Topics
### Data and Image Models

**Sémiologie Graphique [Bertin 67]**

#### Les Variables de l'Image

- **XY 2 Dimensions du Plan**
  - Points
  - Lignes
  - Zones

- **Z Taille**
  - Valeur

- **Les Variables de Séparation des Images**
  - Grain
  - Couleur
  - Orientation
Visualization (Re-)Design

SlicerDicers' Sales Compared to Other Products

Problematic design

Sales of SlicerDicers Compared to Other Products
July - December, 2003
(SlicerDicers’ sales are displayed as black reference lines of 100%, the red bars represent the average monthly sales percentages for July through December.)

Redesign
Visualization Software

D3: Data-Driven Documents
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]
Mapping & Cartography

Dymaxion Maps [Fuller 46]
Graphs and Trees
I don't recall want to know believe think have
Where have all the dentists gone?
Course Mechanics
You should expect to:

1. Evaluate and critique visualization designs
2. Implement interactive data visualizations
3. Gain an overview of research & techniques
4. Develop a substantial visualization project
Instructors

_instructor_

Jeffrey Heer  Assoc Prof, CSE  OH: Tue 2-3pm, 642 CSE  http://jheer.org

Assistants

_Michael Correll_  OH: Mon 2-3pm, 278 CSE
_Jane Hoffswell_  OH: Fri 2-3pm, 218 CSE
Michael Correll

Office: CSE 278
Office Hours: Monday 2-3
Jane Hoffswell

OH: Fri 2-3pm CSE218

Graduated from Harvey Mudd College 2014
2nd year PhD student working with Jeff Heer

Research interests: visualizing program behavior
Textbooks

See also: www.edwardtufte.com
Readings

Some from textbooks, also many papers
Material in class will loosely follow readings
Readings should be read by start of class
Post discussion comments on class Canvas forum
  Comments must be posted within 1 day of lecture
You have 2 “passes” for the quarter
Assignments

Class Participation (10%)
A1: Visualization Design (10%)
A2: Exploratory Data Analysis (15%)
A3: Interactive Visualization Software (25%)
FP: Final Project (40%)
Final Project

Visualization research project on topic of choice

Project write-up in form of a short research paper

Project check-ins and final poster/demo show

Projects from previous classes have been:

- Published (e.g., at the IEEE InfoVis conference)
- Featured in the New York Times
- Released as successful open source projects
Divided Edge Bundling – David Selassie
Visualizing Galaxy Merger Trees

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

Physical Query Plan:

Overview / Operators inside fragment 1

Query time contribution collapse/expand

Detailed execution

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");
Questions?
Assignment 1: Visualization Design

Design a static visualization for a data set.

College admissions can play a profound role in determining one’s future life and career. We’ve collected admissions data (grouped by gender) for selected departments at a major university.

You must choose the message you want to convey. What question(s) do you want to answer? What insight do you want to communicate?
Assignment 1: 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 **5:00 pm, Monday April 4**.