CSE 442 - 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
Abortion

In its most commonly used sense, abortion refers to the deliberate early termination of a pregnancy, resulting in the death of the fetus. In medical terms, the term also refers to early termination of a pregnancy by nature (spontaneous abortion) or medically induced abortion. In the United States, approximately 1 in 4 pregnancies end in abortion. Abortion is a legal right in many countries, and the legality and availability of abortion vary widely around the world.

Methods

Depending on the stage of pregnancy, abortion can be performed by a number of different methods. The earliest terminations (before nine weeks) are often done with a chemical abortion, which involves using medications to induce labor. After the first trimester, a surgical abortion is usually performed. This can be done using a vacuum aspiration procedure, which involves using a suction device to remove the contents of the uterus. Alternatively, a dilatation and evacuation (D&E) procedure may be used.

The controversy

The morality and legality of abortion is a complex issue, discussed by ethicists, religious, and political leaders. The ethics of abortion are also debated by sociologists and historians.

Abortion has been common in most societies throughout history. In the United States and Europe, abortion became a common and acceptable procedure during the second half of the 20th century. Additionally, abortion is accepted in China, India, and other countries. The Church of England officially recognizes abortion as a legitimate and necessary medical procedure, although it remains opposed to the procedure, however, in other countries, notably the United States and the (predominantly Catholic) Republic of Ireland, the controversy remains extremely active, to the extent that even the respective positions are subject to political debate. While those on both sides of the issue are generally peaceful, if heated, in their defence of their positions, the debate is sometimes characterized by violence.

The central question

The central question in the abortion debate is the clash of presumed or perceived rights. On one hand, a fetus (sometimes called the prolife/pro-life/anti-abortion advocates) a human right to life, and if so, at what point in the pregnancy (abortion) becomes human? On the other hand, from the moment of conception (abortion), when the cells of the fetus start to divide, or the moment of birth (abortion), when the fetus is born?
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]
## Summary Statistics and Linear Regression

<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>

### Linear Regression

- **Summary Statistics**
  - $u_X = 9.0$, $\sigma_X = 3.317$
  - $u_Y = 7.5$, $\sigma_Y = 2.03$

- **Linear Regression**
  - \( Y = 3 + 0.5 \times X \)
  - $R^2 = 0.67$

[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]
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>22A</td>
<td>0.000</td>
<td>None</td>
<td>0.280</td>
<td>None</td>
<td>None</td>
<td>36°-66°</td>
</tr>
<tr>
<td>22A</td>
<td>0.000</td>
<td>None</td>
<td>0.280</td>
<td>None</td>
<td>None</td>
<td>160°</td>
</tr>
<tr>
<td>15A</td>
<td>0.010</td>
<td>154.0</td>
<td>0.280</td>
<td>4.25</td>
<td>50.75</td>
<td>335°</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>0.000</td>
<td>45.0</td>
<td>0.280</td>
<td>None</td>
<td>None</td>
<td>29.50</td>
</tr>
<tr>
<td>15B</td>
<td>0.000</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>0.000</td>
<td>None</td>
<td>0.280</td>
<td>None</td>
<td>None</td>
<td>--</td>
</tr>
<tr>
<td>11A</td>
<td>0.000</td>
<td>None</td>
<td>0.280</td>
<td>3.00</td>
<td>14.50</td>
<td>351°</td>
</tr>
<tr>
<td>10A</td>
<td>0.040</td>
<td>217.0</td>
<td>0.280</td>
<td>3.00</td>
<td>None</td>
<td>275°</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°)
  - Much worse visually than SRM-22

- **SRM-22 BLOW-BY**
  - 2 case joints (30-40°)

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

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

<table>
<thead>
<tr>
<th>MOTOR</th>
<th>MGT</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>10 MPH</td>
</tr>
<tr>
<td></td>
<td>27</td>
<td>25</td>
<td></td>
<td>10 MPH</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]
Data in Context: Cholera Outbreak

In 1854 John Snow plotted the position of each cholera case on a map. [from Tufte 83]
Used map to hypothesize that pump on Broad St. was the cause. [from Tufte 83]
Expand Memory: Multiplication

Class Exercise!
34
x 72

Expand Memory: Multiplication
Expand Memory: Multiplication

\[
\begin{array}{c}
34 \\
\times 72 \\
\hline
68 \\
2380 \\
\hline
2448
\end{array}
\]

Time (Sec.)

\[
\begin{array}{c|c|c}
& Mental & Paper & Pencil \\
\hline
0 & 110 & 0 \\
28 & & \\
55 & & \\
83 & & \\
110 & & \\
\end{array}
\]
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>13.9</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]

<table>
<thead>
<tr>
<th>Variables of the Image</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>XY</td>
<td></td>
</tr>
<tr>
<td>2 Dimensions</td>
<td></td>
</tr>
<tr>
<td>Du Plan</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variables of Separation of Images</th>
<th>Grain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taille</td>
<td></td>
</tr>
<tr>
<td>Valeur</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Les Variables de l'Image</th>
<th>Lignes</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Les Variables de Séparation des Images</th>
<th>Couleur</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Orientation</th>
<th></th>
</tr>
</thead>
</table>

Sémiologie Graphique [Bertin 67]
Visualization Design

SlicerDicers' Sales Compared to Other Products

- Ahluts
- NervousNellies
- RingaDingies
- RoundTuits
- SlicerDicers
- SweetNuthins
- ThingamaGigs
- Whatchamacallits
- WileyWidgets

Problematic design

Redesign

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 percentage for July through December.)
Exploratory Data Analysis
Visualization Software

D3: Data-Driven Documents
Interaction

Crimespotting.org

Friday, December 12, 2008
154 reports
The psychophysics of sensory function [Stevens 61]
Color Brewer

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.
Uncertainty
Recent elections have placed a heavy emphasis on “swing states” — Ohio, Florida and the other competitive states. Yet in the past many more states shifted between the Democratic and Republican parties. A look at how the states shifted left or right 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.
Dymaxion Maps [Fuller 46]
Animation

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

Zephoria

User ID: 21721
Friends: 256
Age: 44
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, insomniac, computer-mediated communication, social networks, technology, anthropology, stamping
Music: Portamento, Ozone, Infected Mushroom, Sun Kil Moon, Digital Structures, Ani Difranco, downtempo, Thievery Corporation, Beth Orton, Manciells, Vase, White Pines
Books: Kathryn Otto, glitter, Stanley Milgram, Jeanette Winter, Eric Schlosser, Leslie Feinberg, Dorothy Allison, Itoz Calvino, Hermann Hesse
TV Shows: ??
Movies: Amélie, Walking Dead, Tank Girl, The Matrix, Clockwork Orange, American Beauty, Fight Club, Boys Don't Cry
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 life's chaos, while simultaneously providing my own insane element.

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

Want to Meet: Someone who makes life's complexity seem simply elegant. A partner in analyzing life's complexities.
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
<table>
<thead>
<tr>
<th>Instructor</th>
<th>OH:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jeffrey Heer</td>
<td>Tue 9:00-10:00am, 642 CSE</td>
</tr>
<tr>
<td>Prof, CSE</td>
<td><a href="http://jheer.org">http://jheer.org</a></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Assistants</th>
<th>OH:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Halden Lin</td>
<td>Mon 1:30-2:30pm, 4th Floor</td>
</tr>
<tr>
<td>Younghoon Kim</td>
<td>Tue 3:00-4:00pm, 4th Floor</td>
</tr>
<tr>
<td>Zening Qu</td>
<td>Thu 3:30-4:30pm, 4th Floor</td>
</tr>
<tr>
<td>Sherry Wu</td>
<td>Fri 1:30-2:30pm, 5th Floor</td>
</tr>
</tbody>
</table>
Hi!

I’m a master’s student working on visualization recommendation systems and visualization for NLP.
Younghoon Kim
yhkim01@cs.washington.edu

Office Hour
Tuesday 3:00 - 4:00 p.m.
4th Floor Breakout

Hi!
I’m a 4th year Ph.D. student and interested in algorithms for visualization recommendation and data-driven storytelling!
Tongshuang Wu (Sherry)

OH: Fri 1:30-2:30pm
CSE 5th Floor Breakout

I’m a third year Ph.D. student in IDL working on interactive machine learning. Most recently, I’m building visual tools for error analysis of machine learning models.

Always happy to chat!
Hi, I am Zening Qu

I like to **build tools** to assist people in **data exploration** and **presentation**.

Libra: when you create dashboards, it checks guidelines & suggests revisions.

(vis recommender)
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 D3 book, others from papers & web. 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 (ending week 8). Comments posted by Monday 11:59pm. You have 1 “pass” for the quarter.
Assignments

Class Participation (10%)

A1: Visualization Design (10%) - Due 10/1

A2: Exploratory Data Analysis (15%) - Due 10/16

A3: Interactive Prototype (25%) - Due 10/30

Peer Evaluation - Due 11/6

FP: Final Project (40%)

Initial Prototypes - Due 11/27

Project Deliverables - Due 12/6
Final Project

Produce **interactive web-based visualizations**

Initial **prototype** and **design review**

**Final deliverables** and **video 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
Final Project Theme

Data Visualization for Social Good

Goal: find data of social or scientific import, design visualizations to explore or communicate it effectively.

The specific data domain is open-ended. Possibilities include transportation, housing, public health, education, climate, campaign finance, scientific research, and so on...

You must identify a target audience. May be general (citizens, voters) or specialized (scientists, policy makers).

Use Assignment 2 to explore a data set of interest prior to committing to final project teams and topic!
Inspiration...
Change In Times (CSE 442, Spring 2017)
Gunnar Olson, Halden Lin, Lilian Liang, and Shobhbit Hathi
Locations of each train on the red, blue, and orange lines at 5:46 am. Hover over the diagram to the right to display trains at a different time.

Trains are on the right side of the track relative to the direction they are moving.

See the morning rush-hour, midday lull, afternoon rush-hour, and the evening lull.

Service starts at 5AM on Monday morning. Each line represents the path of one train. Time continues downward, so steeper lines indicate slower trains.

Since the red line splits, we show the Ashmont branch first then the Braintree branch. Train on the Braintree branch "jump over" the Ashmont branch.

Train frequency increases around 6:30AM as morning rush hour begins.
Visualizing Galaxy Merger Trees

S. Loebman, J. Ortiz, L. Orr, M. Balazinska, T. Quinn et al. [SIGMOD ’14]
Visualizing the Republic of Letters

Daniel Chang, Yuankai Ge, Shiwei Song
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 **11:59 pm, Monday October 1**.