CSE 412 - Intro to Data Visualization

The Value of Visualization

Jane Hoffswell  University of Washington
How much data (bytes) did we produce in 2010?
2010: 1,200 EB (1.2 ZB)
Exponential growth...

Gantz et al., 2008, 2010
Data Created & Consumed
Source: IDC Digital Universe

~2x every 2 years

Data (Zettabytes):
- 1.2 ZB (2010)
- 4.4 ZB (2013)
- 44.0 ZB (2020)
But what is all this data?

Data Created & Consumed
Source: IDC Digital Universe

~2x every 2 years

1.2 ZB
4.4 ZB
44.0 ZB
Physical Sensors

Image courtesy cabspotting.org
Records of Human Activity
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*
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. …to whose benefit?

Hal Varian, Google’s Chief Economist
The McKinsey Quarterly, Jan 2009
High potential for data abuse...
Inequality

Rise of the racist robots - how AI is learning all our worst impulses

There is a saying in computer science: garbage in, garbage out. When we feed machines data that reflects our prejudices, they mimic them - from antisemitic chatbots to racially biased software. Does a horrifying future await people forced to live at the mercy of algorithms?

...amplified by “big data” and ML systems.
How might we use **visualization** to **empower understanding** of data and analysis processes?
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 )</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>

**Summary Statistics**

- \( u_X = 9.0 \) \( \sigma_X = 3.317 \)
- \( u_Y = 7.5 \) \( \sigma_Y = 2.03 \)

**Linear Regression**

- \( Y = 3 + 0.5 \times X \)
- \( R^2 = 0.67 \)
Abortion

"Abortion," in its most commonly used sense, refers to the deliberate early termination of a pregnancy, resulting in the death of the fetus. (Informally, the term also refers to the early termination of a pregnancy by natural means such as spontaneous abortion or miscarriage, which occurs in about 15% of all pregnancies, usually within the first 12 weeks of pregnancy, or to the cessation of normal growth of the fetus in the body and organ. What follows is a discussion of the issues related to deliberate or induced abortion.

Methods

Depending on the stage of pregnancy and the method used, abortion is performed by a number of different methods. The earliest terminations (before nine weeks of pregnancy) are typically handled with a medical abortion (the usual method, the combination of two medications). In many countries, this is the only legal method for terminating a pregnancy. However, research has uncovered similar outcomes using a combination of methotrexate and misoprostol. Concerns with chemical abortion have prompted a call for vacuum abortion to be used instead. Vacuum aspiration around the fifteen week (suction aspirated) is usually the most common procedure. From the fifteenth week up until around the eighteenth week, a surgical dilation and evacuation (D & E) is used.

As the fetus size increases other techniques are used to secure abortion in the third trimester. A common technique is the dilation and evacuation of the uterus (D & E) or a hysterotomy abortion, similar to a cesarean section.

The controversy

The morality and legality of abortion is a highly charged and important topic in applied ethics and is also discussed by legal scholars and religious and political leaders. The important facts about abortion are also relevant to discussions by socialists and historians.

Abortion has been common in most societies throughout history, although it has often been opposed by some religious groups. In the United States, the availability of abortion became less common after the Roe v. Wade decision in 1973. Additionally, abortion is legal in the United States and the United Kingdom, although it is more common in China, India, and other countries. Abortion remains a controversial issue, with passionate advocates and opponents on both sides of the issue. The debate is sometimes characterized by violence, though true of sides, is more likely to be expressed in the form of opposition to abortion, because of the gravity and urgency of their views.

The central question

The central question in the abortion debate is whether pregnancy does the fetus become human at some point in time, if so, at what point in pregnancy does the fetus become human? This is a highly controversial issue, with passionate advocates and opponents on both sides of the issue. The debate is sometimes characterized by violence, though true of sides, is more likely to be expressed in the form of opposition to abortion, because of the gravity and urgency of their views.
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]
You Draw It: How Family Income Predicts Children’s College Chances

You Draw It: How Family Income Predicts Children’s College Chances

Support Reasoning
HISTORY OF O-RING DAMAGE ON SRM FIELD JOINTS

<table>
<thead>
<tr>
<th>Cross Sectional View</th>
<th>Top View</th>
<th>Cladding Location (deg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRM No.</td>
<td>Erosion Depth (in.)</td>
<td>Perimeter Affected (deg)</td>
</tr>
<tr>
<td>22A</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>22A</td>
<td>NONE</td>
<td>NONE</td>
</tr>
<tr>
<td>15A</td>
<td>0.010</td>
<td>154.0</td>
</tr>
<tr>
<td>15A</td>
<td>0.038</td>
<td>130.0</td>
</tr>
<tr>
<td>15B</td>
<td>None</td>
<td>45.0</td>
</tr>
<tr>
<td>15B</td>
<td>0.028</td>
<td>110.0</td>
</tr>
<tr>
<td>11A</td>
<td>None</td>
<td>None</td>
</tr>
<tr>
<td>10A</td>
<td>0.040</td>
<td>217.0</td>
</tr>
<tr>
<td>28A</td>
<td>0.053</td>
<td>116.0</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-12 Blow-By
- 2 Case Joints (30-40°)

SRM-13 A, 15, 16A, 18, 23A, 24A
- Nozzle Blow-By
Make Decisions: Challenger
Make Decisions: Challenger

But wait! What is an appropriate “damage index”? Which temperatures, O-ring or outside air?

Chart of temperatures vs. O-ring damage [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]
Find Patterns: NYC Weather

### Answer Questions: Brain Power?

<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 Dragons of Eden [Carl Sagan]
The Elements of Graphing Data
[Cleveland]
Convey Information
“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
Communicate, Inform, Inspire

Visualizing Black America, Du Bois et al. 1900

Bones in hand, Gray’s Anatomy 1918 ed.
New deaths attributed to Covid-19 in European Union, United States, Brazil and United Kingdom

Seven-day rolling average of new deaths, by number of days since 3 average daily deaths first recorded

Source: Financial Times analysis of data from the European Centre for Disease Prevention and Control, the Covid Tracking Project, the UK Dept of Health & Social Care and the Spanish Ministry of Health.
Data updated September 25 2020 12.46pm BST. Interactive version: ft.com/covid19

Coronavirus Tracked, John Burn-Murdoch & Financial Times
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

**Convey** information
  Communicate, inform, inspire
  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 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><strong>Instructors</strong></th>
<th><strong>cse412@cs</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Jane Hoffswell</td>
<td>OH: Wed After Lecture</td>
</tr>
<tr>
<td>Kevin Chang</td>
<td>OH: Mon 2-3pm</td>
</tr>
<tr>
<td>Aayush Chhabra</td>
<td>OH: By Appointment</td>
</tr>
<tr>
<td>Naveena Karusala</td>
<td>OH: By Appointment</td>
</tr>
<tr>
<td>Sonya Lao</td>
<td>OH: Fri 2-3pm</td>
</tr>
<tr>
<td>Kalyani Marathe</td>
<td>OH: Thur 5-6pm</td>
</tr>
<tr>
<td>Yueqian Zhang</td>
<td>OH: Tue 5-6pm</td>
</tr>
</tbody>
</table>
Kevin Chang

OH: Monday 2–3 pm

I’m a 5th year master’s student interested in ML and data visualization.

My 442 final project:
Aayush Chhabra

OH: By Appointment

Hi. I’m a senior graduating in June 2021.

**Major:** CS
**Minor:** Applied Math

My interests include:
- Data Viz
- Artificial Intelligence
- Machine Learning
- Robotics
- and all kinds of cool stuff with data.

Most recently, I interned as a Quantitative Researcher at a hedge fund (Citadel).

Looking forward to learning and working with all of you :)
Naveena Karusala
OH: By Appointment

- 4th year PhD student in the ICTD Lab
- Research in HCI and global development, specifically on health messaging for underserved populations
- Excited to talk about: data ethics, using data as a tool for empowerment and advocacy
Hi, I’m Sonya!

I’m a senior majoring in computer science, and this will be my last quarter. This is my first time TAing for a CSE course, but I used to teach Excel and R workshops for UW Learning Technologies. I’m really excited for the quarter & am looking forward to seeing what you create!

Outside of data visualization, I’m also interested in accessible technology and computer security.

Feel free to talk to me about anything or my latest crochet and baking projects 😊
Kalyani Marathe

OH : Thursday 5:00 pm - 6:00 pm

- **Background**: Master’s student at GRAIL Lab, interested in Computer Vision, Machine Learning, Data Science, and Data Visualization
- **Work Experience**: 2 years as a Software Engineer at IDeaS, A SAS Company
- **Technical Experience**: Python, Java, Groovy, ScikitLearn, NumPy, SciPy, ScikitImage, OpenCV
- **Fun**: Quilling art and playing music
Yueqian Zhang

OH: Tuesday 5 pm - 6 pm

I am a junior majoring in CS and minoring in Education.

I am interested in HCI and Computing Education.

Outside of CS I enjoy anime and drawing.

Come talk to me! I’d love to meet everyone :) 

Some of my artworks:
Important Links

Course Website:
https://courses.cs.washington.edu/courses/cse412/21wi/

Canvas:
https://canvas.uw.edu/courses/1431892

Ed Discussion:
https://edstem.org/us/courses/3116/discussion/
Interactive Data Visualization for the Web, 2nd Edition

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

We will be using D3 v6.
https://d3js.org
Interactive Vega-Lite Notebooks

Hands-on engagement with course concepts and tools using Observable (JavaScript) notebooks.
Readings

From books, notebooks, and linked articles. Material in class will loosely follow readings. Readings should be read by start of class. Post quizzes / comments on class forum. One comment per week (up through week 8). Post comments by Monday 11:59pm.

You have 1 “pass” for the quarter.
Assignments

**CP**  Class Participation (10%)

**A1**  Visualization Design (10%) - Due 1/11 (next Monday)

**A2**  Exploratory Data Analysis (15%) - Due 1/25

**A3**  Ethical & Deceptive Visualization (20%) - Due 2/8

Peer Evaluation (5%) - Due 2/15

**FP**  Final Project (40%)

Proposal - Due 2/12

Milestone Prototype - Due 2/26

Demonstration Video - Due 3/10

Final Prototype - Due 3/15
Weekly Discussion Post and Quiz

Share substantive discussion post on Ed about course readings or lecture material.

Short quizzes reinforce course concepts. Quizzes are graded on participation not score.

First discussion thread and quiz will be posted on Ed later this afternoon, due Mon 1/11, 11:59pm.
Final Project

Produce **interactive web-based visualizations**

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

**Final deliverables** and **video presentation**

Submit and **publish online** (GitHub)

Projects from **previous classes** have been:

- Published as research papers
- Shared widely (some 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 (residents, voters) or specialized (scientists, policy makers).

Use Assignment 2 and 3 to explore a data set of interest prior to committing to final project teams and topic!
Inspiration...
Change In Times (CSE 442, Spring ’17)
Gunnar Olson, Halden Lin, Lilian Liang, and Shobhit 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. Trains on the Braintree branch "jump over" the Ashmont branch.

Train frequency increases around 6:30AM as morning rush hour begins.
Questions?
A1: Visualization Design

Design a static visualization for a data set.

The climate of a place can have a tremendous impact on people's lived experience. You will examine average monthly climate measurements for six major U.S. cities, roughly covering the edges of the continental United States.

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 ($\leq 4$ paragraphs)

Due by **11:59 pm, Monday January 11**.