The Value of Visualization

Leilani Battle  University of Washington
Testing poll everywhere
setup

Respond here: pollev.com/leibatt
How much data (bytes) did we produce in 2010?

Respond here: pollev.com/leibatt
2010: 1,200 exabytes and exponential growth...

Gantz et al., 2008, 2010
1.2 ZB

4.4 ZB

44.0 ZB

Data Created & Consumed

Source: IDC Digital Universe

~2x every 2 years
But *what* is in all this data?

Data Created & Consumed

Source: IDC Digital Universe

~2x every 2 years
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. "ubiquitous" about whom? "free" to whom? ...to whose benefit?

Hal Varian, Google’s Chief Economist
The McKinsey Quarterly, Jan 2009
High potential for data abuse…
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.
We move from data to information to knowledge to wisdom, and separating one from the other, being able to distinguish among and between them that is, knowing the limitations and the danger of exercising one without the others while respecting each category of intelligence, is generally what serious education is about.

Toni Morrison, American Novelist

*The Source of Self Regard*
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**  
$u_X = 9.0$  
$\sigma_X = 3.32$  
$u_Y = 7.5$  
$\sigma_Y = 2.03$  

**Linear Regression**  
$Y = 3 + 0.5 \, X$  
$R^2 = 0.67$  

[Anscombe 1973]
Set A

Set B

Set C

Set D

[Anscombe 1973]
Abortion

(Revision as of 12:56, 4 Jun 2003)

"Abortion," in its most commonly used sense, refers to the deliberate early termination of pregnancy, resulting in the death of the fetus. Medically, the term also refers to early termination of a pregnancy by natural (spontaneous abortion) or medicated (induced abortion) means. 1 in 5 of all pregnancies, usually within the first trimester, result in the spontaneous loss of the developing fetus.

Methods

Depending on the stage of pregnancy and the choice of method, abortion may be performed by a number of different procedures. In the first trimester, before the fetus is viable, a chemical abortion is the usual method, involving the use of medications such as mifepristone and misoprostol, usually taken orally. Around the fifteenth week, suction aspiration and vacuum abortion is the most common approach, replacing the more risky dilation and curettage (D&C). From the fourteenth week up to about the eighteenth week, a surgical dilation and curettage (D & C) is used.

As the fetus size increases other techniques used to end the pregnancy include abortion by forceps and abortion using suction aspirators. In more advanced pregnancies, the fetus can be expelled with considerable difficulty by injecting the amniotic fluid with saline or a solution. Very late abortion can be brought about by the controversial ultrasound-guided destructive and aspiration (D & X) or a hormonal abortion, similar to the early abortion described above.

The controversy

The morality and legality of abortion is an important topic in applied ethics and is often discussed by local, national, and international political leaders. Important facts about abortion are also researched by the database of laws.

Abortion has been common in most societies, although it has often been opposed by some institutionalized religions and governments. Abortion was legal in the United States until 1973 when abortion became commonly accepted by the 21st century. Additionally, abortion is accepted in China, India, and other populous countries. The Catholic Church remains against the procedure, however, and in other countries notably the United States and the (predominantly Catholic) Republic of Ireland, the controversy is extremely active, to the extent that events of the previous day are subject to debate. While those on both sides of the issue are generally peaceful, it is thought that the debate is sometimes characterized by violence. Though true on both sides, this is more marked on the side of those opposed to abortion, because of what they view as the gravity and urgency of their views.

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 "unborn child") is frequently described as human with a right to life. On the other hand, if at what point a pregnancy does the fetus become human, and whether or not it is a right of a woman to have an abortion, is not clear.
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
E.J. Marey’s sphygmograph [from Braun 83]
Gallop, Bay Horse “Daisy”  [Muybridge]
You Draw It: How Family Income Predicts Children’s College Chances

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

Support Reasoning
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]
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 Rat</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 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
The coronavirus crisis is different

Job growth (or loss) since each recession began, based on weekly earnings

1990 recession

2001 recession

2008 recession

Coronavirus crisis

Notes: Based on a three-month average to show the trend in volatile data. Source: Labor Department via IPUMS, with methodology assistance from Ernie Tedeschi of Evercore ISI.
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 Topics
Data and Image Models

<table>
<thead>
<tr>
<th>LES VARIABLES DE L'IMAGE</th>
<th>POINTS</th>
<th>LIGNES</th>
<th>ZONES</th>
</tr>
</thead>
<tbody>
<tr>
<td>XY 2 DIMENSIONS DU PLAN</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Z</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TAILLE</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VALEUR</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LES VARIABLES DE SÉPARATION DES IMAGES</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>GRAIN</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COULEUR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ORIENTATION</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sémiologie Graphique [Bertin 67]
Visualization Design

Problematic design

Redesign

SlicerDicers' Sales Compared to Other Products

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

Sales of SlicerDicers Compared to Sales of Other Products
July - December, 2011

Respond here: pollev.com/leibatt
Exploratory Data Analysis
Maps

Dymaxion Maps [Fuller 46]
Animated transitions in statistical data graphics [Heer & Robertson 07]
Color Brewer
Graphical Perception

The psychophysics of sensory function [Stevens 61]
Hierarchies

Degree-Of-Interest Trees [Heer & Card 04]
Scalability

Interactive querying of 1.7B stars (1.2TB) in Falcon [Moritz et al. 2019]
Recent elections have placed a heavy emphasis on “swing states” — Ohio, Florida and the other competitive states. Yet, there is no clear trend of which party is leading 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
Lectures & Office Hours

Watch the pre-recorded video before class on Thursdays!

Tues = Lectures. Thurs = in-class activities.

All Tues lectures will be in-person + recorded. We will use PollEV to reinforce important concepts for the homework assignments.

Please attend in person but **NOT** if you feel ill.

Office hours will be held in person or on Zoom.

Links are on Canvas for virtual office hours.

We strongly encourage using Ed to post questions and seek help!
There is no one universal textbook on visualization!

So we will draw on books, notebooks, and linked articles.

Material in class will loosely follow readings. Readings should be read by start of class.
Interactive Data Visualization for the Web, 2nd Edition

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

We will be using D3 v7.
https://d3js.org
Interactive Vega-Lite Notebooks

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

**CP** Class Participation (10%)

**A1** Expository Visualization (10%) - *Due 10/6*

**A2** Deceptive Visualization (15%) - *Due 10/18*
  Peer Review (5%) - *Due 10/24*

**A3** Interactive Prototype (20%) - *Due 11/7*
  Peer Review (5%) - *Due 11/14*

**FP** Final Project (35%)
  Proposal - *Due 11/15*
  Prototype - *Due 11/28*
  Demonstration Video - *Due 12/5*
  Final Prototype - *Due 12/11*
Grading Philosophy

A great submission gets a great grade (A- to A, 3.6 - 3.8), but an exceptional grade (A+, 3.9 - 4.0) requires exceptional effort.

Example: Typical A1 grades (out of 10 points).
Everyone starts with a high score (9/10). Then, we deduct points for errors. We also add points for going above and beyond the assignment requirements.

The median score for A1 is typically 8.5 out of 10 (considered an A-).
Final Project

Produce an **explorable visual explanation**
Initial **prototype** and **design review**
**Final deliverables** and **video presentation**
Submit and **publish online** (GitLab)

Projects from **previous classes** have been:
• Published as research papers
• Shared widely (some in the New York Times!)
• Released as successful open source projects
Why outbreaks like coronavirus spread exponentially, and how to “flatten the curve”

Harry Stevens, Washington Post 2020

<table>
<thead>
<tr>
<th>Count</th>
<th>Change over time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recovered</td>
<td>73</td>
</tr>
<tr>
<td>Healthy</td>
<td>0</td>
</tr>
<tr>
<td>Sick</td>
<td>127</td>
</tr>
</tbody>
</table>
Locations of each train on the red, blue, and orange lines at 5:13 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.
KEYBOARD WALKING

Passwords with a "keyboard walking" pattern start at an arbitrary key, then move in a direction (usually right or down) while continuing to hit keys. Sometimes this is combined with holding down the $\text{SHIFT}$ key, so that some characters are uppercase or symbols to improve complexity.

While the generated password may seem to be random and unhackable, password crackers check for these keyboard patterns and guess them early on.

Many passwords in the leaked passwords dataset have a spatial pattern. Other than the numeric passwords like 123456, common keyboard walking offenders include qwerty and 1qaz@wsx.

Semantic Passwords
Vishal Devireddy  (CSE 512, Spring ’21)
Course Participation

Tue Lectures - PolIEV questions
Thur Excercises - Team submissions
Online quizzes - Submitted on Canvas
Online Practice Quizzes

We assign practice quizzes in weeks 2 - 8 to emphasize important concepts.

Quizzes are due each Friday by 11:59pm, starting next week. They can be retaken to get full points.

Quizzes only count towards course participation.
Coming Up Soon!
Thur Oct 7: In-Class Activity

We will try designing our own data visualizations in class!

You need to watch the pre-recorded lecture video beforehand. (We will post them soon!)

We will have a quick poll on PollEV before diving into the activity.
Observable + Data Tutorial

This Friday Sept. 29, 3:30-5pm. Virtual.

Introduction to Observable notebooks, JavaScript basics, and data management and transformation, led by Katherine and Amanda.

Zoom link is available on Canvas. The tutorial will be recorded.
A1: Expository Visualization

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: Expository Visualization

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 on Gradescope; see A1 page)
- Image of your visualization (PNG or JPG format)
- Short description + design rationale (≤ 4 paragraphs)

Due by **11:59 pm, Wed Oct 6**.
Seeking Help From Course Staff

The fastest way to reach us is through the Ed Discussion Board

Email us ASAP if you need access to edstem.org!

We also hold virtual and in-person office hours each week (schedule on the next slide).

We can also be reached over email at cse442@cs.washington.edu
Instructors

Leilani Battle
Assistant Professor, CSE

OH: Wed 2-3pm (virtual)

Teaching Assistants

Katherine Juarez
OH: Online / Ed

Catalina Martinez
OH: Fri 2:30pm-3:30pm (in person)

Kai Nylund
OH: Mon 9am-10am (virtual)

Ron Pechuk
OH: Mon 1pm-2pm (virtual)

Krithika Satish
OH: Online / Ed

Hamsa Shankar
OH: Fri 10:30am-11:30am (virtual)

Wei Jun Tan
OH: Online / Ed

Yuanjie ‘Tukey’ Tu
OH: Thu 9am-10am (virtual)

Amanda Worthy
OH: Tue 4pm-5pm (in person)

Yifan Zhang
OH: Online / Ed
Leilani Battle (she/her)
Assistant Professor, UW CSE
Co-Director, CSE Interactive Data Lab
https://homes.cs.washington.edu/~leibatt/

Visualization / HCI / Data management / Data Science

I model how people interact with data analysis systems.

I use these models to build **behavior-driven** optimizations, UI features, and performance benchmarks for interactive data analysis

Hobbies: disc golf, reading, cooking, travel, board games, etc.
Katherine Juarez

kajuarez@cs.washington.edu

- Third Year PhD Student
- Research Interests: Human-Computer Interaction, CS Education
- Hobbies:
  - CrossFit
  - Kayaking
  - Hiking
  - Foodie
Catalina Martinez
catamtz3@cs.washington.edu

Year: 4th Year Senior
From: Bridgeport, WA
Hobbies: Gym, Basketball, Drawing
Interests: Data Science, Astronautics
Kai Nylund (he/him)
knylund@cs.washington.edu

Year: 6th (BS/MS)

From: Fall City, WA

Hobbies: drawing, climbing

Interests: NLP, visualization
Ron Pechuk
ft. his cats

- Contact: rpechuk@cs.washington.edu
- Senior studying Computer Science
  - w/ Minor in Data Science
- Hobbies
  - Basketball
  - Hiking
  - Board Games
  - Bowling
- Interests
  - Front-end Development
  - Data Science
Krithika Satish  (she/her)
ksatish@cs.washington.edu

- Senior in Computer Science
- From Fremont, CA
- Hobbies
  - Hiking, cooking, traveling
- Interests
  - Data Science, NLP
Hamsa Shankar
hamsas@cs.washington.edu

Year: BS/MS
From: Redmond, WA
Prior TA experience: 10-time TA (332, 351, 455)
Hobbies: Musicals, Art
Interests: Animation, AR/VR, Graphics
WeiJun Tan
wj428@cs.washington.edu

- From Selangor, Malaysia
- 4th Year BS/MS in CS + BS in Stat
- 4th time TA Data Visualization
- Interests: System / ML / Data Science
- Hobbies: Traveling / Chess / Table Tennis
Yuanjie (Tukey) Tu
yuanjt2@uw.edu

- From Jiangxi, China
- PhD candidate
- Research: Self-driving vehicles
- Hobbies: Hiking, archery, Go
Amanda Worthy
aworthy@cs.washington.edu (She/ Her)

Civil Engineering Ph.D. Student, Data Science Option

TA experience: CSE 412 (Spring 2023)

From: Boulder, CO (sko buffs)

Interests: Data Science, Urban Building Energy Systems, Air Quality

Hobbies: Nordic Skiing and Swimming
Yifan Zhang  (She/Her)
yifanz47@cs.washington.edu

- Senior in Computer Science
- Interests:
  - Software Development
  - NLP
- Hobbies:
  - Hiking
  - Bouldering
  - Working Out
  - Exploring Food
Questions?