The Purpose of Visualization

Maneesh Agrawala

CS 448B: Visualization
Fall 2021

How much data (bytes) did we produce in 2020?
2020: 64.2 zetabytes

[IDC 2021]

10x increase over 5 years

[IDC 2021]
Physical Sensors

Image courtesy cabspotting.org

Health & Medicine
Abortion

Abortion is the ending of a pregnancy by removal or expulsion of an embryo or fetus before it can survive outside the uterus. An abortion that occurs without intervention is known as a miscarriage or spontaneous abortion. When deliberate steps are taken to end a pregnancy, it is called an induced abortion, or less frequently "induced miscarriage." The unmodified term abortion generally refers to an induced abortion. A similar procedure after the fetus has potential to survive outside the womb is known as a "late termination of pregnancy" or less accurately as a "late term abortion." When properly done, abortion is one of the safest procedures in medicine, but unsafe abortion is a major cause of maternal death, especially in the developing world. Making safe abortion legal and accessible reduces maternal deaths. It is safer than childbirth, which has a 14 times higher risk of death in the United States. Modern methods use medication or surgery for abortions. The drug mifepristone in combination with prostaglandins appears to be as safe and effective as surgery during the first and second trimester of pregnancy. The most common surgical technique involves dilating the cervix and using a suction device. Birth control, such as the pill or intrauterine devices, can be used immediately following abortion. When performed legally and safely on a woman who desires it, induced abortions do not increase the risk of long-term mental or physical problems. In contrast, unsafe abortions (those performed by unskilled individuals, with hazardous equipment, or in unsanitary facilities) cause 47,000 deaths and 5 million hospital admissions each year. The World Health Organization recommends safe and legal abortion be available to all women worldwide. Around 56 million abortions are performed each year in the world, with about 45% done legally, 23% abortion rates have changed little between 2000 and 2008, before which they decreased for at least two decades as access to family planning and birth control increased. As of 2008, 40% of the world's women had access to legal abortions without limits as they reason. Countries that permit abortions have different limits on how late in pregnancy abortion is allowed.

Historically, abortions have been attempted using herbal medicines, sharp tools, formalin, massage, or through other traditional methods. Abortion laws and cultural or religious views of abortions are different around the world. In some areas abortion is legal only in specific cases such as rape, problems with the fetus, poverty, risk to a woman's health, or incest. There are debates over the moral, ethical, and legal issues of abortion. Those who oppose abortion often argue that an embryo or fetus is a human with a right to life, and they may compare abortion to murder. Those who support the legality of abortion often hold that it is part of a woman's right to make decisions about her own body. Others favor legal and accessible abortion as a public health strategy.
“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
“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*  
January 2009

What is visualization?
Examples

Ages of first-time mothers in 1990

Examples

[Image of an anatomical diagram of the ear]
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>

Linear Regression

\[ Y = 3 + 0.5X \]

\[ R^2 = 0.67 \]

[Anscombe 73]
Why do we create visualizations?

- Answer questions (or discover them)
- Make decisions
- See data in context
- Expand memory
- Support graphical calculation
- Find patterns
- Present argument
- Tell a story
- Inspire
The Purpose of Visualization

Record information
- Photographs, blueprints, ...

Support reasoning about information (analyze)
- Process and calculate
- Reason about data
- Expand memory

Convey information to others (present)
- Share and persuade
- Emphasize important aspects of data

Record Information
Answer question

Gallop, Bay Horse “Daisy” [Muybridge 1884-86]

30

Answer question

Gallop, Bay Horse “Daisy” [Muybridge 1884-86]

31
Photographs: Phases of the moon

Drawing: Phases of the moon

Galileo's drawings of the phases of the moon from 1616
http://galileo.rice.edu/sci/observations/moon.html
Support Reasoning

Make a decision: Challenger

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]
Make a decision: Challenger

Visualizations drawn by Tufte show how low temperatures damage O-rings [Tufte 97]

See data in context: Cholera outbreak

In 1854 John Snow plotted the position of each cholera case on a map. [from Tufte 83]
See data in context: Cholera outbreak

Used map to support hypothesis Broad St. pump was the cause. [from Tufte 83]

Expand memory: Multiplication

Class Exercise
Expand memory: Multiplication

34
x 87

238
2720
2958
Expand memory: Multiplication

74
\times 48

Time (Sec.)

<table>
<thead>
<tr>
<th>Mental</th>
<th>Paper &amp; Pencil</th>
</tr>
</thead>
<tbody>
<tr>
<td>592</td>
<td>0</td>
</tr>
<tr>
<td>2960</td>
<td>120</td>
</tr>
<tr>
<td>3552</td>
<td>120</td>
</tr>
</tbody>
</table>
Most powerful brain?

![Graph showing brain weights compared to body weights across various species.]

The Dragons of Eden [Carl Sagan]
Tell a story: Most powerful brain?

Convey Information to Others
Most powerful brain?

![Brain Mass vs. Body Mass Diagram]

Beautiful Evidence [Tufte]

Present argument

![Crimean War Deaths Diagram]

“to affect thro’ the eyes what we fail to convey to the public through their word-proof ears”

Crimean War Deaths [Nightingale 1858]
Inspire

X-ray crystallography of DNA [Franklin 52]

Bones in hand [from 1918 edition]

58

Inspire

Double helix model [Watson and Crick 53]

Bones in hand [from 1918 edition]

59
The Purpose of Visualization

Record information
- Photographs, blueprints, ...

Support reasoning about information (analyze)
- Process and calculate
- Reason about data
- Expand memory

Convey information to others (present)
- Share and persuade
- Emphasize important aspects of data

Goals of visualization research

1. Understand how visualizations convey information
   - What do people perceive/comprehend?
   - How do visualizations correspond with mental models of data?

2. Develop principles and techniques for creating effective visualizations and supporting analysis
   - Leverage perception and cognition
   - Strengthen connection between visualization and mental models
Course Topics

Data and Image Models

[Bertin, Graphics and Graphic Information Processing 1981]
Visualization Design & ReDesign

SlicerDicers’ Sales Compared to Other Products

Problemsitic design

Redesign

Exploratory Data Analysis

Tableau – based on Polaris [Stolle, Tang, Hanrahan]
Using Space Effectively

Dorling and Dorling-like Cartograms

Graduated Symbol Map
Demers Cartogram
Dorling Cartogram

http://www.ncgia.ucsb.edu/projects/Cartogram_Central/types.html

Interaction

Oakland Crimespotting (crimespotting.org) [Stamen]
Introduction to D3

D3: Data Driven Documents [Bostock 2011]

Perception

The psychophysics of sensory function [Stevens 61]
**Visual Explainers**

Gapminder [Rosling]

---

**Color**

[from Cynthia Brewer http://www.personal.psu.edu/faculty/c/a/cab38/ ]
Animation

Animated Transitions [Heer 07]

Networks

Degree-of-Interest Trees [Heer 2004]
You should expect to

1. Design, evaluate and critique visualizations
2. Explore data using existing visualization tools
3. Implement interactive data visualizations
4. Develop a substantial visualization project

Course Mechanics
Course Assistant: Shana Hadi

Shana Hadi

Projects:

- Live Digital Handwriting at the C.S. Sanyal Home
- ML: Currently learning to model SQL via Cloud Nine Text's Software Engineering training at the C.S. Sanyal Home, among other things, including handwriting implementation for the creation of ML models to aid in handwriting recognition

Outside of OH use Slack to connect with us

https://canvas.stanford.edu/courses/144332/external_tools/11232

Office Hours

Maneesh: 2-3pm Wed, Coupa Café Y2E2
Dae Hyun: 10-11:00am Thu, outside CEMEX Aud
Shana Hadi: 7-8:00pm Sun, via Canvas/Zoom

Happy to schedule other OH by appointment

Outside of OH use Slack to connect with us

https://canvas.stanford.edu/courses/144332/external_tools/11232
Textbooks

See also: www.edwardtufte.com

Interactive Notebooks

Hands-on engagement with course concepts and modern visualization tools (Vega-Lite / D3), in JavaScript (Observable)
Optional Textbook

For additional theory and depth

Optional Book

For learning D3!

- Book available online
- Code/examples on GitHub

We will be using D3 v7

https://d3js.org
Readings

- From books, notebooks and linked articles
  Many open to public, some may require SUNetID/Password
- Material in class will be loosely based on readings
- Readings should be read by start of class

- Post comments (about reading, notebooks or lecture) using link on class webpage
  One comment per week through week 9
  Must post by end of the week
  You have 1 pass for the quarter

Class home page
https://magrawala.github.io/cs448b-fa21/

Reading/Notebook/Lecture Responses

Good responses typically exhibit one or more
- Critiques of arguments made in the papers/lectures
- Analysis of implications or future directions for ideas in readings/lectures
- Insightful questions about the readings/lectures

Responses should not be summaries
Discussion

Discussion is essential for effective design, evaluation and critique of visualizations

Attendance is very highly recommended

Assignments

Class participation (10%)

Assignment 1: Visualization Design (10%) due 9/27

Assignment 2: Exploratory Data Analysis (15%) due 10/11
   Learn to use Tableau

Assignment 3: Interactive Prototype (25%) due 10/25
   Should be familiar with Javascript (start now if you are not)
   Will cover basics of Vega-Lite and D3 in class

Final Project (40%) proposal due 11/3, design review 11/29, 12/1,
   final submission 12/10
Final project

Choice of project type

- Create an extended visual explainer
- Small visualization research project

Projects from previous classes have been:

- Gone viral on blogs
- Published as research papers
- Released as open source projects

Structure of Musicals

Lyrical themes in Hamilton [Townley-Smith, Sterman, Cook 2016]
Visualization of Narrative Structure

Character interactions and sentiment in The Hobbit [Bilenko, Miyakawa 2013]

deepviz: Visualizing Convolutional NNs

1) Filter details 2) Image selector 3) Network overview 4) Filter visualization 5) Visualization selector 6) Selection helper 7) Animation slider [Bruckner, Rosen, Sparks 2013]
Assignment 1: Visualization Design

Design a static visualization for a data set.

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

Data: Stanford Undergraduate Majors

The Stanford Daily publishes a variety of datasets through the Stanford Open Data Portal. They have published a data table containing information about the number of Stanford students majoring in 70 different subject areas from 2011-2019. We have filtered and wrangled this data to the top 10 majors over the time period to produce a dataset with the following variables:

Number of records:
Variable Names:
Subject: Subject areas in which students majored.
Number of Students: Number of students majoring in the area.

The extracted dataset is available in csv format: StanfordTopTenMajors2010s.csv

Due by 7am on Mon Sep 27