The Purpose of Visualization

Maneesh Agrawala

CS 448B: Visualization
Fall 2017

How much data (bytes) did we produce in 2013?
2013: 4400 exabytes
10x increase over 5 years

[Image courtesy cabspotting.org]
Health & Medicine

Records of Human Activity
“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?
Examples

![Chart showing crime distribution]

Observation: A large number of auto thefts occur in the University district, even though the area ranks relatively low in total crime.

Examples

![Diagram of human ear]
Examples

What is visualization?

Definition [www.oed.com]

1. The action or fact of visualizing; the power or process of forming a mental picture or vision of something not actually present to the sight; a picture thus formed.

2. The action or process of rendering visible.
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 Linear Regression

\[ u_x = 9.0 \quad \sigma_x = 3.317 \quad Y = 3 + 0.5 \, X \]
\[ u_y = 7.5 \quad \sigma_y = 2.03 \quad R^2 = 0.67 \]

[Anscombe 73]
Why do we create visualizations?
Why do we create visualizations?

- To answer a question about the data
- To provide a frame for a narrative
- To be able to communicate (to show information)
- To engage a viewer (interactivity)
- To bring out structure in the data
- Take the work out of getting the point
- To create a flashy thing

Three functions of visualizations

**Record information**
- Photographs, blueprints, …

**Support reasoning about information (analyze)**
- Process and calculate
- Reason about data
- Feedback and interaction

**Convey information to others (present)**
- Share and persuade
- Collaborate and revise
- Emphasize important aspects of data
Record Information

Answer question

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

Gallop, Bay Horse "Daisy" [Muybridge 1884-86]

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

Other recording instruments

Marey’s sphygmograph [from Braun 83]
Support Reasoning

Find patterns: New York weather

From the New York Times 1981
Make a decision: Challenger
Visualizations drawn by Tufte show how low temperatures damage O-rings [Tufte 97]
In 1854 John Snow plotted the position of each cholera case on a map. [from Tufte 83]

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

34
x 87

34
x 87
**Expand memory: Multiplication**

\[
\begin{array}{c}
34 \\
\times 87 \\
2958
\end{array}
\]

<table>
<thead>
<tr>
<th>Time (Sec.)</th>
<th>Mental</th>
<th>Paper &amp; Pencil</th>
</tr>
</thead>
<tbody>
<tr>
<td>120</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Graphical calculation: Evaporation**

Johannes Lambert used graphs to study the rate of water evaporation as function of temperature [from Tufte 83]
Graphical calculation: Evaporation

Johannes Lambert used graphs to study the rate of water evaporation as function of temperature [from Tufte 83]

Convey Information to Others
Present argument

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

Crimean War Deaths [Nightingale 1858]

Inspire

Bones in hand [from 1918 edition]  Double helix model [Watson and Crick 53]
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

Simulation

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

Internet

Photo sharing/annotation
flickr.com

Group Authored Encyclopedia
wikipedia.org

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

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]
1800 exabytes in 2011 [Gantz 11]
4400 exabytes in 2013 [Gantz 14]

Need better tools and algorithms for visually conveying information
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

Goals of visualization research

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

2. Develop principles and techniques for creating effective visualizations
   - Amplify perception and cognition
   - Strengthen connection between visualization and mental models of data
1. Data and image models

2. Visualization Design

The psychophysics of sensory function [Stevens 61]
4. Interaction

Oakland Crimespotting (crimespotting.org) [Stamen]

5. Building interactive visualizations with D3

D3: Data Driven Documents [Bostock 2011]
6. Color

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

7. Spatial Layout

London underground [Beck 33]
8. Animation

Animated Transitions [Heer 07]

9. Trees and graphs

Degree-of-Interest Trees [Heer 2004]
10. Text visualization

Course Goals

1. Design, evaluate and critique visualizations
2. Explore data using existing visualization tools
3. Implement interactive data visualizations
4. Gain an overview of research and techniques
5. Develop a substantial visualization project
Course Mechanics

Instructor: Maneesh Agrawala
Course Assistants
Alec Glassford
Zach Maurer

Piazza is the best way to interact with us
http://piazza.com/stanford/fall2017/cs448b

Laptops
Textbooks

See also: www.edwardtufte.com

Readings

- Some from textbooks, also many papers
  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 discussion comment (about reading or lecture) using link on class webpage
  Must post by noon the day after the lecture
  You have 2 passes for the quarter

Class home page
https://magrawala.github.io/cs448b-fa17
Lecture/Reading 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

Requirements

Class participation (10%)

Assignment 1: Visualization Design (10%)

Assignment 2: Exploratory Data Analysis (15%)
  - Learn to use Tableau will show you a bit in class, but expect to pick it up on your own

Assignment 3: Creating Interactive Visualization Software (25%)
  - Should be familiar with Javascript (start now if you are not)
  - Will cover basics of D3 in class, but expect you will also pick it up on your own

Final Project (40%)
Assignment 1: Visualization Design

Barley Yield Data

In 1931 and 1932 Minnesota collected data on the yield in bushels per acre of 10 varieties of barley grown in 1/40 acre plots at University Farm, St. Paul, and at the five branch experiment stations located at Waseca, Morris, Crookston, Grand Rapids, and Duluth (all in Minnesota). The varieties were grown in three randomized blocks at each of the six stations during 1931 and 1932, different land being used each year of the test.

Number of records: 120
Variable Names:
- Site: Crockston, Duluth, Grand Rapids, Morris, University Farm, Waseca
- Variety: Glabron, Manchuria, No 457, No 462, No 475, Peatland, Sivasota, Trebi, Velvet, Wasc. No 38
- Yield: bushels/acre
- Year: 1931, 1932

We’ve cleaned up this dataset and posted in csv format: barley2.csv

Barley Yields
Due by noon on Mon Oct 2

Final project

- Visualization research project on topic of your choice
- Last 4 weeks of course
- Project write-up in form of a research paper (6-8 pages)
- Two in-class project presentations
  1. Initial in-class status report (dates TBD – likely week before Thanksgiving)
  2. Final poster presentation (dates TBD)

Projects from previous classes have been published
- IEEE Visualization
- IEEE Information Visualization
- SIGGRAPH
Structure of Musicals

**Lyrical Themes in Hamilton**

- Schuyler Sisters
- Alexander Hamilton
- The Revolution
- King George
- My Shot
- I Know
- Farewell
- Alien
- New York
- One Last Time
- The Room

**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]