The Value of Visualization

Jeffrey Heer  University of Washington
How much data (bytes) did we produce in 2010?
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?
Physical Sensors

Image courtesy cabspotting.org
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR/ECG</td>
<td>97</td>
</tr>
<tr>
<td>Art</td>
<td>82/60</td>
</tr>
<tr>
<td>SpO2</td>
<td>99</td>
</tr>
<tr>
<td>RR/CO2</td>
<td>------</td>
</tr>
<tr>
<td>HR/ECG</td>
<td>79</td>
</tr>
<tr>
<td>Art</td>
<td>152/79</td>
</tr>
<tr>
<td>SpO2</td>
<td>95</td>
</tr>
<tr>
<td>RR/CO2</td>
<td>------</td>
</tr>
<tr>
<td>HR/ECG</td>
<td>64</td>
</tr>
<tr>
<td>Art</td>
<td>93/55</td>
</tr>
<tr>
<td>SpO2</td>
<td>99</td>
</tr>
<tr>
<td>RR/Imp</td>
<td>------</td>
</tr>
</tbody>
</table>
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...
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>X</th>
<th>Y</th>
<th>Set B</th>
<th>X</th>
<th>Y</th>
<th>Set C</th>
<th>X</th>
<th>Y</th>
<th>Set D</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>8.04</td>
<td></td>
<td>10</td>
<td>9.14</td>
<td></td>
<td>10</td>
<td>7.46</td>
<td></td>
<td>8</td>
<td>6.58</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>6.95</td>
<td></td>
<td>8</td>
<td>8.14</td>
<td></td>
<td>8</td>
<td>6.77</td>
<td></td>
<td>8</td>
<td>5.76</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>7.58</td>
<td></td>
<td>13</td>
<td>8.74</td>
<td></td>
<td>13</td>
<td>12.74</td>
<td></td>
<td>8</td>
<td>7.71</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>8.81</td>
<td></td>
<td>9</td>
<td>8.77</td>
<td></td>
<td>9</td>
<td>7.11</td>
<td></td>
<td>8</td>
<td>8.84</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>8.33</td>
<td></td>
<td>11</td>
<td>9.26</td>
<td></td>
<td>11</td>
<td>7.81</td>
<td></td>
<td>8</td>
<td>8.47</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>9.96</td>
<td></td>
<td>14</td>
<td>8.1</td>
<td></td>
<td>14</td>
<td>8.84</td>
<td></td>
<td>8</td>
<td>7.04</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7.24</td>
<td></td>
<td>6</td>
<td>6.13</td>
<td></td>
<td>6</td>
<td>6.08</td>
<td></td>
<td>8</td>
<td>5.25</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4.26</td>
<td></td>
<td>4</td>
<td>3.1</td>
<td></td>
<td>4</td>
<td>5.39</td>
<td></td>
<td>8</td>
<td>5.56</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>10.84</td>
<td></td>
<td>12</td>
<td>9.11</td>
<td></td>
<td>12</td>
<td>8.15</td>
<td></td>
<td>8</td>
<td>5.56</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>4.82</td>
<td></td>
<td>7</td>
<td>7.26</td>
<td></td>
<td>7</td>
<td>6.42</td>
<td></td>
<td>8</td>
<td>7.91</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5.68</td>
<td></td>
<td>5</td>
<td>4.74</td>
<td></td>
<td>5</td>
<td>5.73</td>
<td></td>
<td>8</td>
<td>6.89</td>
<td></td>
</tr>
</tbody>
</table>

**Summary Statistics**

- $u_X = 9.0$  $\sigma_X = 3.317$
- $u_Y = 7.5$  $\sigma_Y = 2.03$

**Linear Regression**

- $Y = 3 + 0.5X$
- $R^2 = 0.67$

[Anscombe 1973]
Set A

Set B

Set C

Set D

[Anscombe 1973]
Wikipedia History Flow [Viegas & Wattenberg]
Edit War...

Wikipedia History Flow [Viegas & Wattenberg]
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]
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>Clocking Location (deg)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SRM No.</strong></td>
<td><strong>Erosion Depth (in.)</strong></td>
<td><strong>Perimeter Affected (deg.)</strong></td>
</tr>
<tr>
<td>61A LH Center Field**</td>
<td>22A</td>
<td>None</td>
</tr>
<tr>
<td>61A LH CENTER FIELD**</td>
<td>22A</td>
<td>NONE</td>
</tr>
<tr>
<td>51C LH Forward Field**</td>
<td>15A</td>
<td>0.010</td>
</tr>
<tr>
<td>51C LH Center Field (prim)**</td>
<td>15B</td>
<td>0.038</td>
</tr>
<tr>
<td>51C LH Center Field (sec)**</td>
<td>15B</td>
<td>None</td>
</tr>
<tr>
<td>41D RH Forward Field</td>
<td>13B</td>
<td>0.028</td>
</tr>
<tr>
<td>41C LH Aft Field**</td>
<td>11A</td>
<td>None</td>
</tr>
<tr>
<td>41B LH Forward Field</td>
<td>10A</td>
<td>0.040</td>
</tr>
<tr>
<td>STS-2 RH Aft Field</td>
<td>2B</td>
<td>0.053</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 32 blow-by
  - 2 case joints (30-40°)

- SRM-13 A, 15, 16A, 18, 23 A, 24 A
  - Nozzle blow-by

---

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

<table>
<thead>
<tr>
<th>MOTOR</th>
<th>MT</th>
<th>AMB</th>
<th>O-RING</th>
</tr>
</thead>
<tbody>
<tr>
<td>DM-1</td>
<td>68</td>
<td>36</td>
<td>47</td>
</tr>
<tr>
<td>DM-2</td>
<td>76</td>
<td>45</td>
<td>52</td>
</tr>
<tr>
<td>QM-3</td>
<td>72.5</td>
<td>40</td>
<td>48</td>
</tr>
<tr>
<td>QM-4</td>
<td>76</td>
<td>48</td>
<td>51</td>
</tr>
<tr>
<td>SRM-15</td>
<td>52</td>
<td>54</td>
<td>53</td>
</tr>
<tr>
<td>SRM-22</td>
<td>77</td>
<td>78</td>
<td>75</td>
</tr>
<tr>
<td>SRM-25</td>
<td>55</td>
<td>26</td>
<td>29</td>
</tr>
</tbody>
</table>

Wind:
- 10 mph
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]
Expand Memory: Multiplication
Expand Memory: Multiplication

34
x 72
Expand Memory: Multiplication

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

Time (Sec.)

<table>
<thead>
<tr>
<th></th>
<th>Mental</th>
<th>Paper &amp; Pencil</th>
</tr>
</thead>
<tbody>
<tr>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>83</td>
<td></td>
<td></td>
</tr>
<tr>
<td>110</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

2448
### 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>Artic 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 WASHINGTON POST
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
LES VARIABLES DE L'IMAGE

XY
2 DIMENSIONS
DU PLAN

POINTS

Z

TAILLE

VALEUR

LIGNES

ZONES

LES VARIABLES DE SÉPARATION DES IMAGES

GRAIN

COULEUR

ORIENTATION

Sémiologie Graphique [Bertin 67]
Visualization Design

Problematic design

Redesign
Exploratory Data Analysis
Maps

Dymaxion Maps [Fuller 46]
Recent elections have placed a heavy emphasis on “swing states” — Ohio, Florida and the other competitive states. You can see how many swing states shifted between the Democratic and Republican parties. A look at how the states shifted left or right between 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.
Visualization Software

D3: Data-Driven Documents

Vega-Lite / Altair
Animated transitions in statistical data graphics [Heer & Robertson 07]
Color Brewer
The psychophysics of sensory function [Stevens 61]
Hierarchies

Degree-Of-Interest Trees [Heer & Card 04]
Networks
Uncertainty
Interactive querying of 1.7B stars (1.2TB) in Falcon [Moritz et al. 2019]
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
Instructors

**Instructor**

**Jeffrey Heer**
OH: Tue 10-11am

**Teaching Assistants**

**Abhishek Babu**
OH: Mon 1:30-2:30pm

**Vishal Devireddy**
OH: Online / Ed

**Naveena Karusala**
OH: By appointment

**Xia Su**
OH: Online / Ed

**Firn Teanklin**
OH: Wed 1:30-2:30pm

**Yueqian Zhang**
OH: Fri 2:30-3:30pm
Abhishek Babu
babua @ cs

Office hours: Mondays 1:30 - 2:30 PM

I’m an MS student in the BS/MS program and I’m interested in machine learning and data engineering.

Relevant technical skills: JavaScript, TypeScript, HTML/CSS, D3.js, general web dev (and other data wrangling libraries in Python)

I enjoy Spikeball as well! (check out Husky Roundnet Club)
Vishal Devireddy

Email  vishald@cs
Office hours  by appointment / Ed

I’m a BS/MS student interested in interactive web development, HCI, and cool graphs.

Ask me about
- JavaScript + CSS
- Web design
- Web accessibility
- Observable
- D3
- Idyll
Nussara (Firn) Tieanklin
nussara@cs.washington.edu
Office hour: Wed 1:30 - 2:30 PM

Research
• Seattle Community Networks
  providing internet access to resource-constrained community

  • Evaluate the impact of Air Pollution on Grab motorcycle drivers/riders

Technical Experience
  User research, Design process, Data Management, Web-programming

(Covid) Hobbies
• Gardening
• Badminton
• Playing computer games
• Travelling
Naveena

- 5th year PhD student in the ICTD Lab

- Research in HCI and global development, specifically on health messaging to underserved populations

- Excited to talk about: design of visualizations
Hi! I’m a first year PhD student researching into HCI with special interest in design and creativity.

Have background in design. Feel free to discuss your designs with me!

No office hour, but I manage the online discussions with Vishal. So chat with me online!

Will lead a Tableau Tutorial with Naveena on Oct. 8th.

xiasu@cs.washington.edu
Yueqian Zhang
OH: Friday 2:30 pm - 3:30 pm

I am a senior majoring in CS and minoring in Education.
I am interested in Accessibility and Graphics.
Outside of CS I enjoy design and drawing.
Come talk to me! I’d love to meet everyone :)

Some of my artworks:
Readings

From books, notebooks, and linked articles. Material in class will loosely follow readings. Readings should be read by start of class. Complete non-scored quizzes on Canvas. Post comments on class forum.

One comment or peer reviews each week. Post comments by Friday 11:59pm.

You have 1 discussion “pass” for the quarter.
Textbook

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 10/11

**A2** Deceptive Visualization (15%) - Due 10/20
Peer Review (5%) - Due 10/25

**A3** Interactive Prototype (20%) - Due 11/8
Peer Review (5%) - Due 11/15

**FP** Final Project (35%)
Proposal - Due 11/19
Demonstration Video - Due 12/8
Final Prototype - Due 12/14
Final Project

Produce an **explorable visual explanation**

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
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 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)
Questions?
Observable + Data Tutorial

Friday Oct. 1, 4:30-6pm on Zoom

Introduction to Observable notebooks, JavaScript basics, and data management and transformation, led by Firn & Yueqian.

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 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 11.