How much data (bytes) did we produce in 2010?
2010: 1,200 exabytes
10x increase over 5 years

Gantz et al, 2008, 2010
Records of Human Activity
Wikipedia History Flow (IBM)
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*
A Poverty of Attention

“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?

“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></th>
<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>
<td>X</td>
</tr>
<tr>
<td>10</td>
<td>8.04</td>
<td>10</td>
<td>9.14</td>
<td>10</td>
</tr>
<tr>
<td>8</td>
<td>6.95</td>
<td>8</td>
<td>8.14</td>
<td>8</td>
</tr>
<tr>
<td>13</td>
<td>7.58</td>
<td>13</td>
<td>8.74</td>
<td>13</td>
</tr>
<tr>
<td>9</td>
<td>8.81</td>
<td>9</td>
<td>8.77</td>
<td>9</td>
</tr>
<tr>
<td>11</td>
<td>8.33</td>
<td>11</td>
<td>9.26</td>
<td>11</td>
</tr>
<tr>
<td>14</td>
<td>9.96</td>
<td>14</td>
<td>8.1</td>
<td>14</td>
</tr>
<tr>
<td>6</td>
<td>7.24</td>
<td>6</td>
<td>6.13</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>4.26</td>
<td>4</td>
<td>3.1</td>
<td>4</td>
</tr>
<tr>
<td>12</td>
<td>10.84</td>
<td>12</td>
<td>9.11</td>
<td>12</td>
</tr>
<tr>
<td>7</td>
<td>4.82</td>
<td>7</td>
<td>7.26</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>5.68</td>
<td>5</td>
<td>4.74</td>
<td>5</td>
</tr>
</tbody>
</table>

**Summary Statistics**

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

**Linear Regression**

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

[Anscombe 1973]
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]
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 RH Center Field (prim)**</td>
<td>15B</td>
<td>0.038</td>
</tr>
<tr>
<td>51C RH 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 22 Blow-By**
- 2 case joints (30-40°)

**SRM-13A, 15, 16A, 18, 23A, 24A**
- Nozzle Blow-By

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

<table>
<thead>
<tr>
<th>MOTOR</th>
<th>MGT</th>
<th>AMB</th>
<th>O-RING</th>
<th>WIND</th>
</tr>
</thead>
<tbody>
<tr>
<td>DM-1</td>
<td>68</td>
<td>36</td>
<td>47</td>
<td>10 mph</td>
</tr>
<tr>
<td>DM-2</td>
<td>76</td>
<td>45</td>
<td>52</td>
<td>10 mph</td>
</tr>
<tr>
<td>QM-3</td>
<td>72.5</td>
<td>40</td>
<td>48</td>
<td>10 mph</td>
</tr>
<tr>
<td>QM-4</td>
<td>76</td>
<td>48</td>
<td>51</td>
<td>10 mph</td>
</tr>
<tr>
<td>SRM-15</td>
<td>52</td>
<td>64</td>
<td>53</td>
<td>10 mph</td>
</tr>
<tr>
<td>SRM-22</td>
<td>77</td>
<td>78</td>
<td>75</td>
<td>10 mph</td>
</tr>
<tr>
<td>SRM-25</td>
<td>55</td>
<td>26</td>
<td>29</td>
<td>25 mph</td>
</tr>
</tbody>
</table>
Make a Decision: Challenger
Visualizations drawn by Tufte show how low temperatures damage O-rings [Tufte 97]
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]
Expand Memory: Multiplication

Class Exercise!
Expand Memory: Multiplication

34
x 72
Expand Memory: Multiplication

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

Time (Sec.)

- Mental: 110
- Paper & Pencil: 0
Find Patterns: NYC Weather

The Most Powerful Brain?

<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 Elements of Graphing Data
[Cleveland]
Convey Information to Others
Inspire

Bones in hand [from 1918 edition]

Double helix model [Watson, Crick, Franklin]
“to affect thro’ the Eyes what we fail to convey to the public through their word-proof ears”
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

**Communicate** information to others
   Share and persuade
   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

Sémiologie Graphique [Bertin 67]
Visualization (Re-)Design

Problematic design

Redesign
Exploratory Data Analysis
Visualization Software

**D3**: Data-Driven Documents

Vega-Lite / Altair
Recent elections have placed a heavy emphasis on “swing states” — Ohio, Florida and the other competitive states. You can see how some states shifted 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.
Maps

Dymaxion Maps [Fuller 46]
CIE 1931 xy chromaticity diagram showing the gamut of the sRGB and Adobe RGB color spaces including the Planckian locus, with temperatures indicated. Wavelengths of monochromatic light are shown in blue.
Graphical Perception

The psychophysics of sensory function [Stevens 61]
Animation

Animated transitions in statistical data graphics [Heer & Robertson 07]
Hierarchies

Degree-Of-Interest Trees [Heer & Card 04]
Uncertainty
Visualizations: Word tree / Alberto Gonzales

Creator: Martin Wattenberg
Tags:

Search: i don't

118 hits

i don't

recall

want to

know

believe

think

have


Data source: CQ Transcript Wire via the Washington Post

Comments (4)
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
Professor, CSE

OH: Tue 10-11:15am, CSE2 302
http://jheer.org

Assistants

Matt Conlen
OH: Mon 11am-12p, CSE2 152

Sherry Wu
OH: Thu 2:30-3:30p, CSE2 152

Yang Liu
OH: Thu 2:30-3:30p, CSE2 152

Halden Lin
OH: Online / Canvas

cse512@cs
Matthew Conlen
mconlen@cs.washington.edu
OH: Monday 11-12pm GATES 152 or by appointment

Research on interactive documents and data-driven storytelling

Experience as data journalist
FiveThirtyEight, CNN, The New Yorker

Much experience with JavaScript, D3, and general web programming
Tongshuang Wu (Sherry)
wts Huang@cs.washington.edu
3rd year Ph.D. student working on HCI for machine learning model training & evaluation.
Always happy to chat!
Yang Liu

Office Hours:
Thur 2:30 - 3:30 pm (Out of town 6/6)
Gates Center 152

I’m a PhD student working on visualization and HCl. I took 512 last Spring, and I am looking forward to being part of 512 this year!
Hello! I’m a CSE masters student with interests in **Automating Visualization Design** and, more broadly, in **Visualization for Data Science**.
“Textbooks”

See also: www.edwardtufte.com
Interactive Notebooks

Hands-on engagement with course concepts and visualization tools (Vega-Lite / Altair), in both JavaScript (Observable) and Python (Jupyter)!
Optional Book

Interactive Data Visualization for the Web, 2nd Edition

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

We will be using D3 v5.
https://d3js.org
Readings

From books, notebooks, and linked articles. Material in class will loosely follow readings. Readings should be read by start of class. Post discussion comments on class Canvas forum. One comment per week (up through week 8). Comments must be posted by Friday 11:59pm. You have 1 “pass” for the quarter.
Assignments

Class Participation (10%)

A1: Visualization Design (10%) - Due 4/8

A2: Exploratory Data Analysis (15%) - Due 4/22

A3: Interactive Prototype (25%) - Due 5/6

Peer Evaluation - Due 5/13

FP: Final Project (40%)

Proposal - Due 5/16

Milestone Prototype - Due 6/3

Project Deliverables - Due 6/11
Final Project

Visualization research project on topic of choice

Initial prototype and peer evaluation

Design reviews and final presentation

Submit and publish online (if feasible)

Projects from previous classes have been:

- Published as research papers
- Featured in the New York Times
- Released as successful open source projects
Visualizing Galaxy Merger Trees

S. Loebman, J. Ortiz, L. Orr, M. Balazinska, T. Quinn et al. [SIGMOD '14]
Perfopticon: Distributed Query Performance

Dominik Moritz et al. [EuroVis '15]
A browser-based tool for visualization and analysis of diffusion MRI data

Jason D. Yeatman, Adam Richie-Halford, Josh K. Smith, Anisha Keshavan & Ariel Rokem
Protovis: A Graphical Toolkit for Visualization

Mike Bostock
```javascript
var army = pd.nest(napoleon.army, "dir", "group");
var vis = new pv.Panel();

var lines = vis.add(pv.Panel).data(army);
lines.add(pv.Line)
  .data(function() army[this.idx])
  .left(lon).top(lat).size(function(d) d.size/8000)
  .strokeStyle(function() color[army[paneIndex][0].dir]);

vis.add(pv.Rule).data([0,-10,-20,-30])
  .top(function(d) 300 - 2*d - 0.5).left(200).right(150)
  .lineWidth(1).strokeStyle("#ccc")
  .anchor("right").add(pv.Label)
  .text(function(d) d.temp+"°").textBaseline("center");

vis.add(pv.Label).data(napoleon.cities)
  .left(lon).top(lat)
  .text(function(d) d.city).font("italic 10px Georgia")
  .textAlign("center").textBaseline("middle");

vis.add(pv.Line).data(napoleon.temp)
  .left(lon).top(tmp).strokeStyle("#0")
  .add(pv.Label)
  .top(function(d) 5 + tmp(d))
  .text(function(d) d.temp+"° +=d.date.substr(0,6))
```
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 (≤ 4 paragraphs)

Due by **8:00 pm, Monday April 8**.