

The Purpose of Visualization

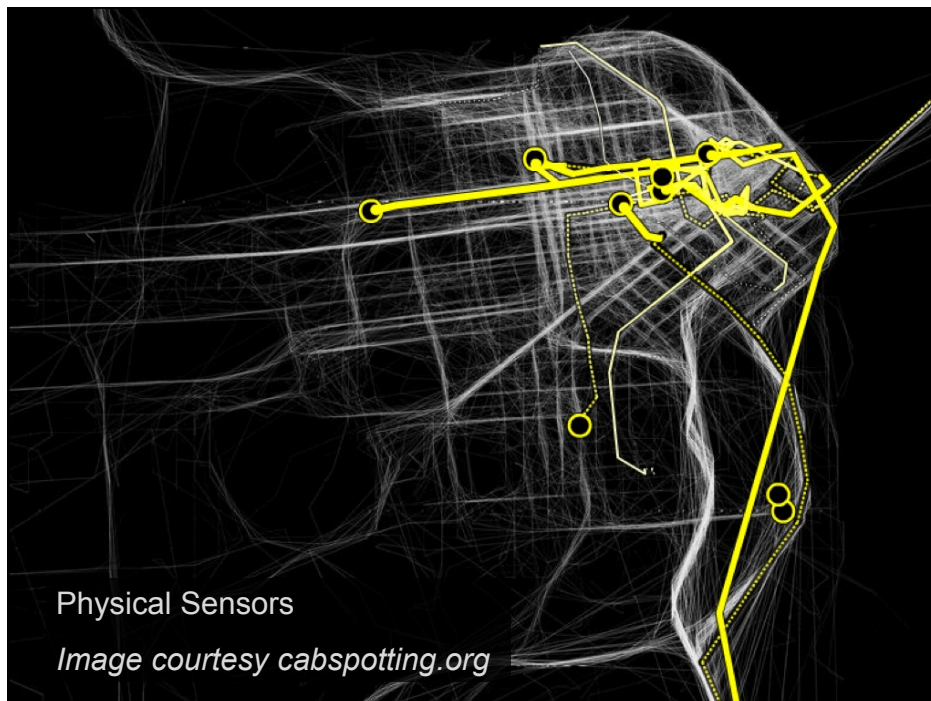
Maneesh Agrawala

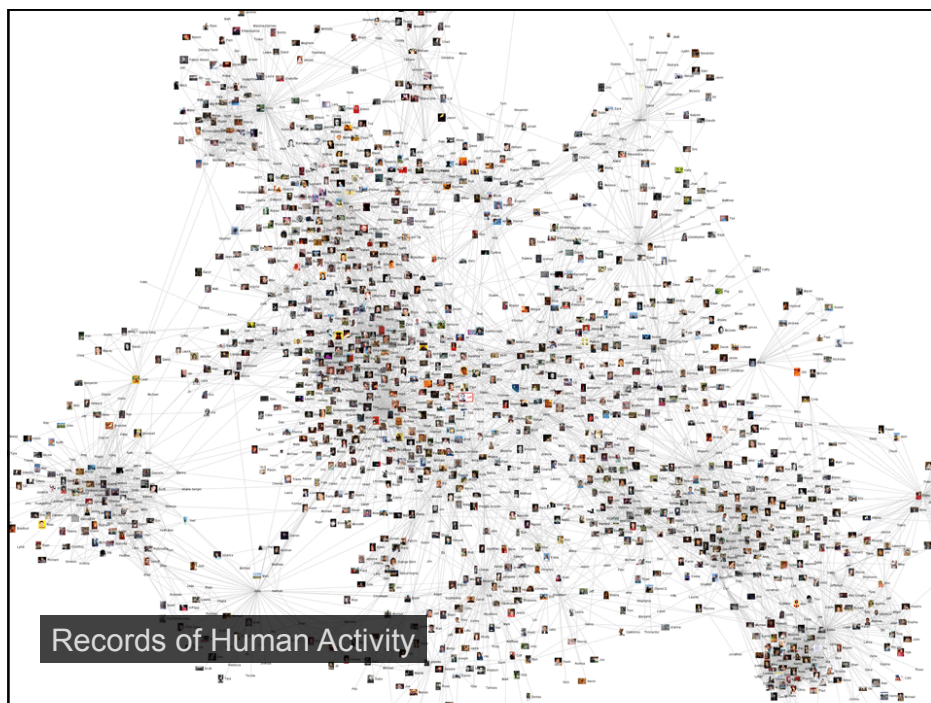
CS 448B: Visualization
Fall 2017

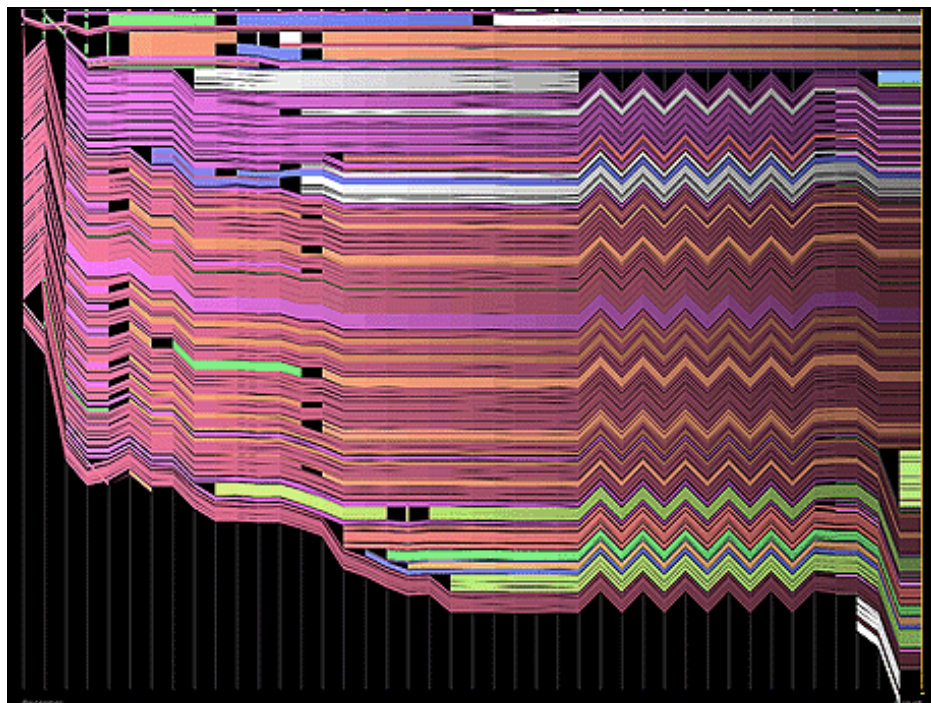
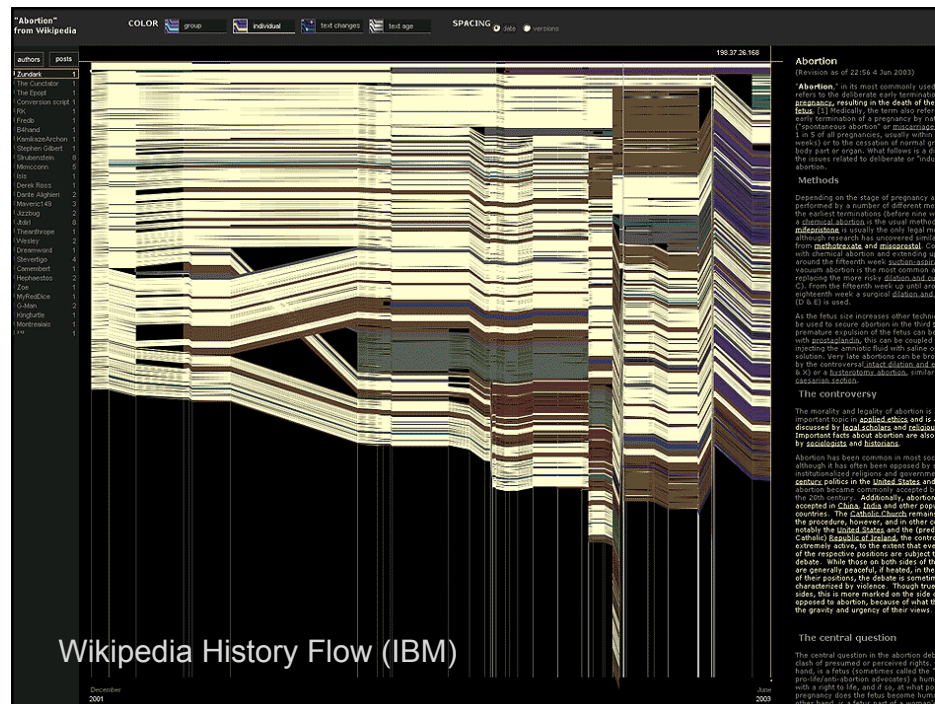
How much data (bytes)
did we produce in 2013?

2013: 4400 exabytes
10x increase over 5 years

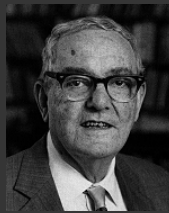
[Gantz 2014]







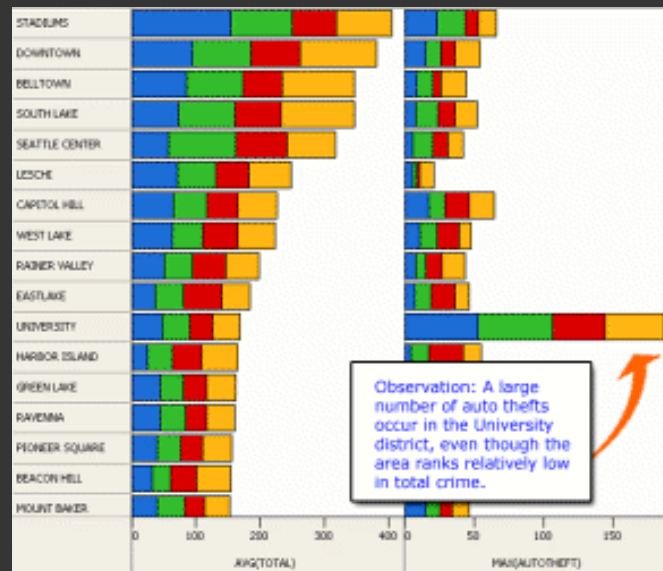
“What information consumes is rather obvious:
it consumes the attention of its recipients.
Hence a wealth of information creates a poverty
of attention, and a need to allocate that
attention efficiently among the overabundance of
information sources that might consume it.”



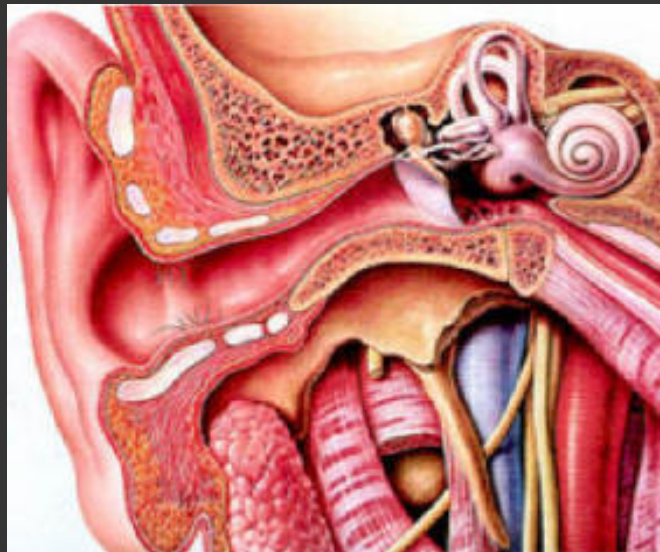
~Herb Simon
as quoted by Hal Varian
Scientific American
September 1995

What is visualization?

Examples



Examples



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.”

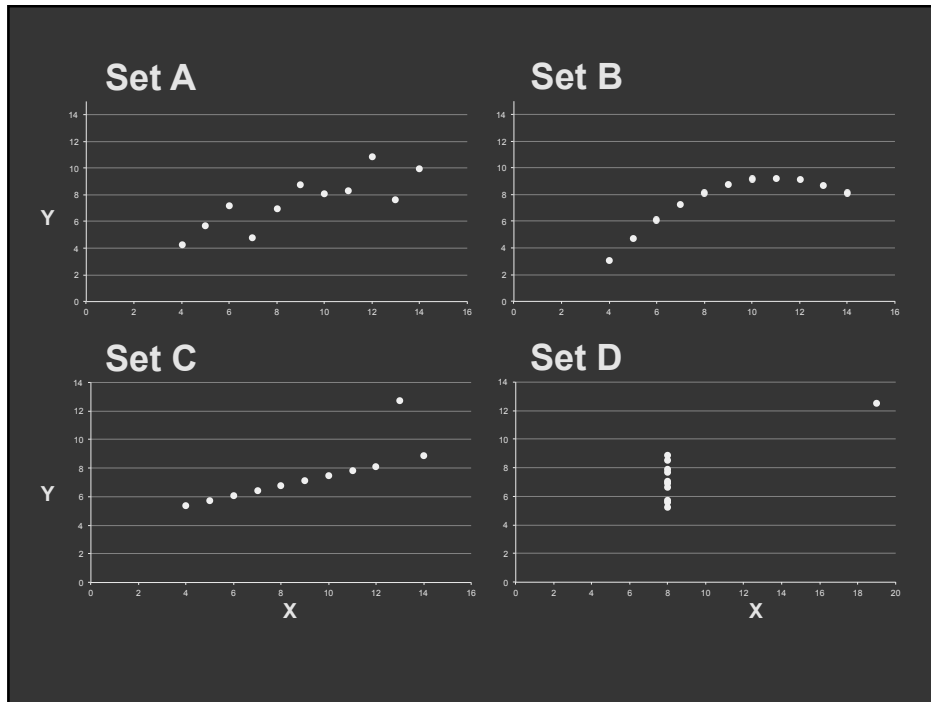
[Card, Mackinlay, & Shneiderman 1999]

Set A		Set B		Set C		Set D	
X	Y	X	Y	X	Y	X	Y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.11	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

Summary Statistics Linear Regression

$u_X = 9.0$ $\sigma_X = 3.317$ $Y = 3 + 0.5 X$
 $u_Y = 7.5$ $\sigma_Y = 2.03$ $R^2 = 0.67$

[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

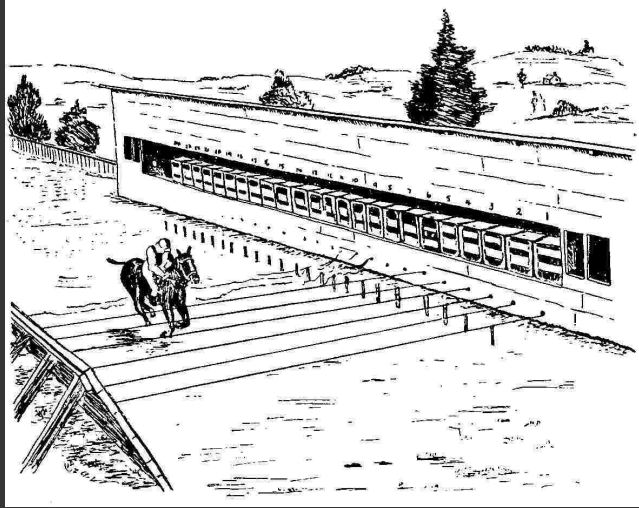
Record Information

Answer question



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

Answer question

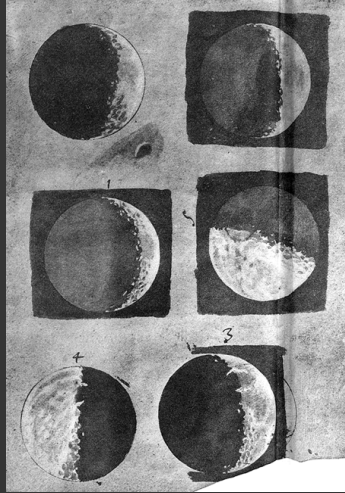


Gallop, Bay Horse "Daisy" [Muybridge 1884-86]

Photographs: Phases of the moon

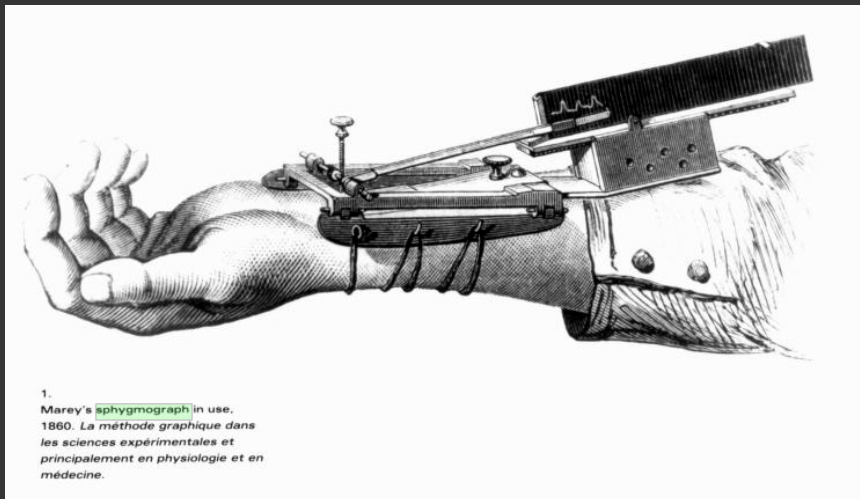


Drawing: Phases of the moon



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

Other recording instruments

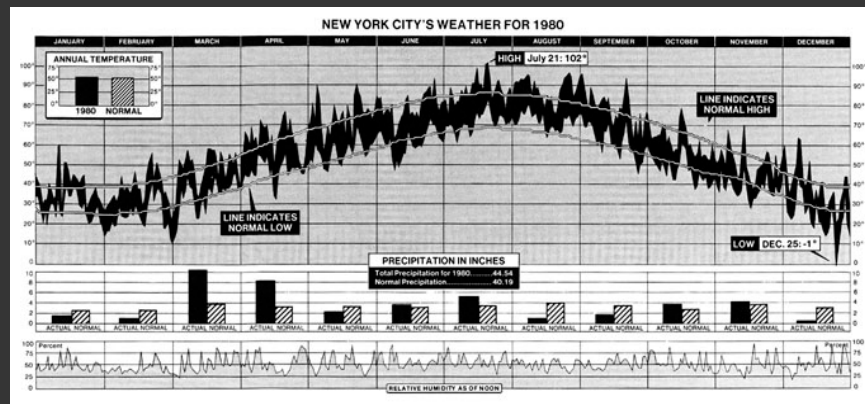


1.
Marey's sphygmograph in use.
1860. *La méthode graphique dans
les sciences expérimentales et
principalement en physiologie et en
médecine.*

Marey's sphygmograph [from Braun 83]

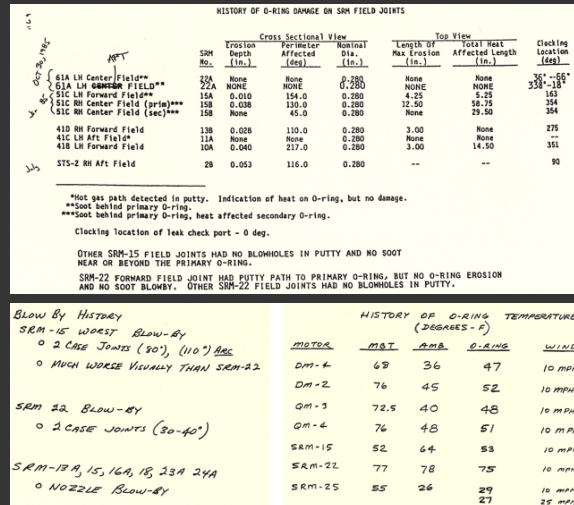
Support Reasoning

Find patterns: New York weather



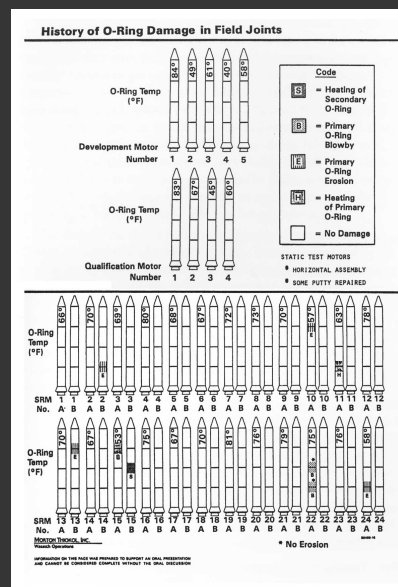
From the New York Times 1981

Make a decision: Challenger

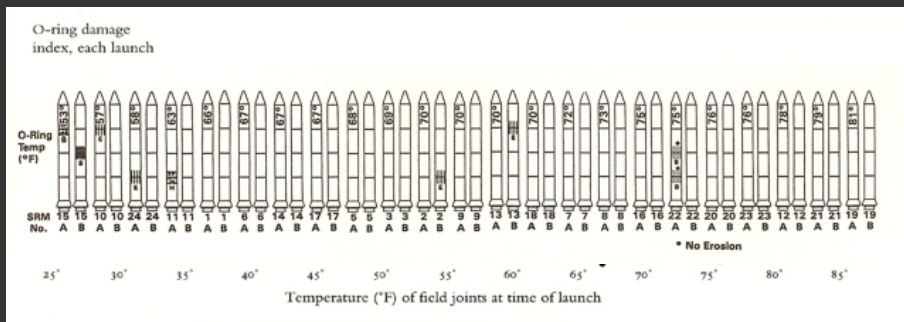


2 of 13 pages of material faxed to NASA by Morton Thiokol [from Tufte 1997]

Make a decision: Challenger

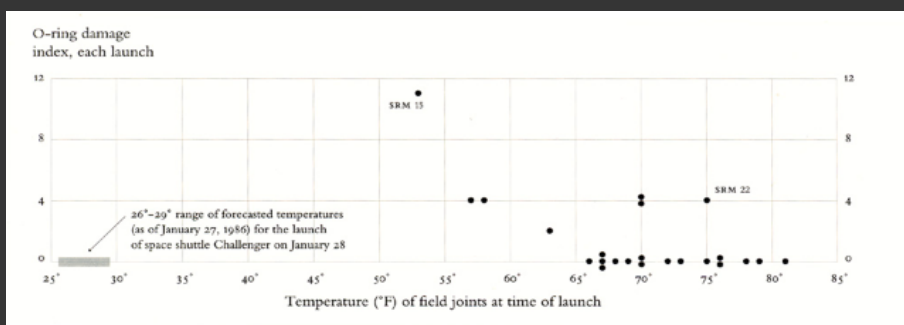


Make a decision: Challenger



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]

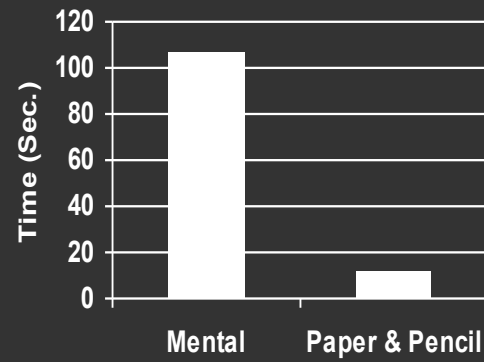
Expand memory: Multiplication

Class Exercise

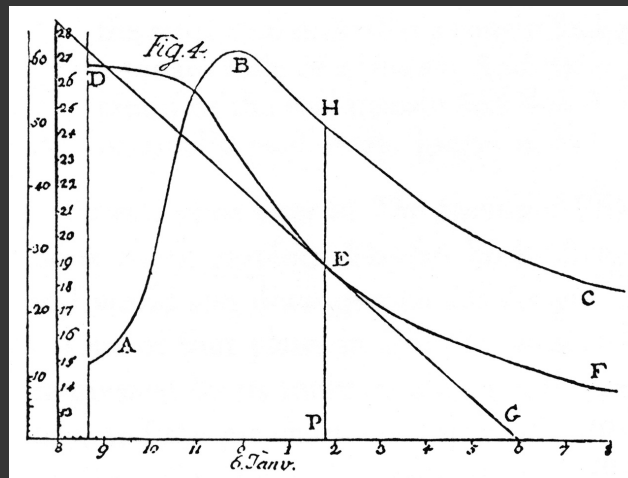
Expand memory: Multiplication

$$\begin{array}{r} 34 \\ \times 87 \\ \hline \end{array}$$

Expand memory: Multiplication

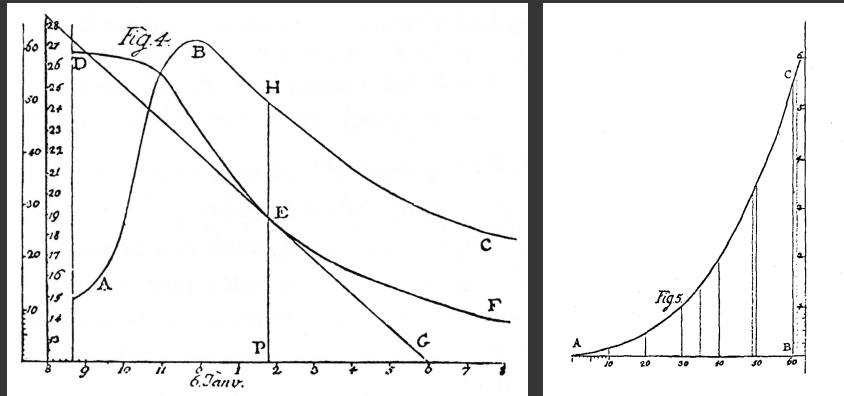
$$\begin{array}{r} 34 \\ \times 87 \\ \hline 238 \\ 2720 \\ \hline 2958 \end{array}$$


Graphical calculation: Evaporation



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

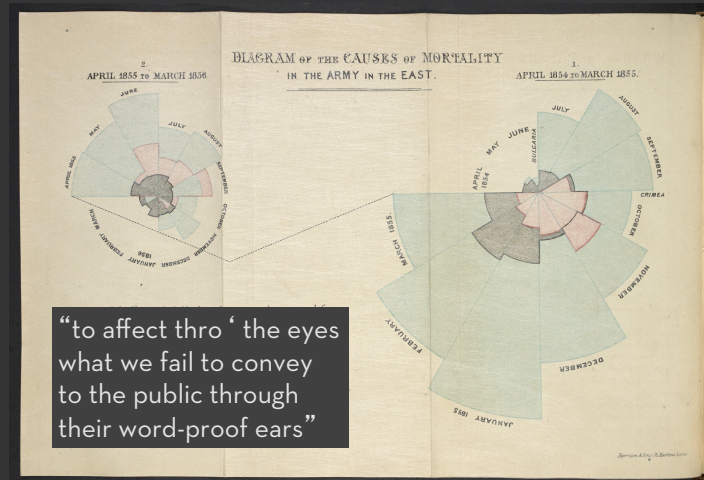
Graphical calculation: Evaporation



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

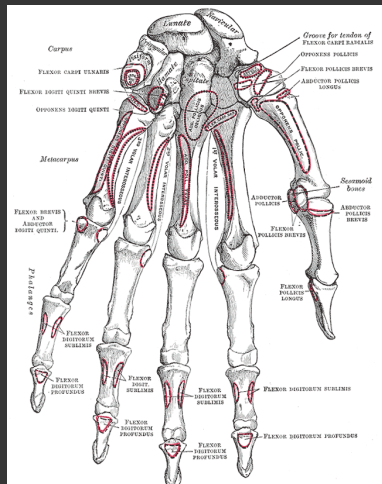
Convey Information to Others

Present argument



Crimean War Deaths [Nightingale 1858]

Inspire



Bones in hand [from 1918 edition]



Double helix model [Watson and Crick 53]

Visualization Research

Challenge

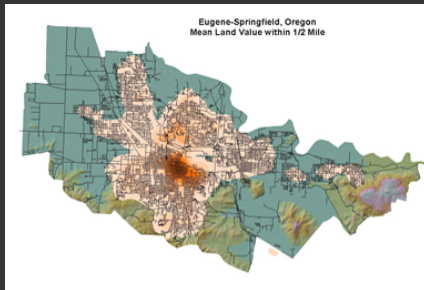
More and more unseen data

- Faster creation and collection

Challenge

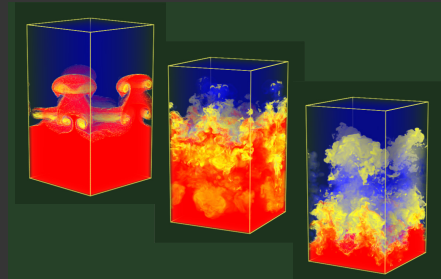
More and more unseen data

- Faster creation and collection



Urban development planning

www.urbansim.org



Fluid flow

ctr.stanford.edu

Simulation

Challenge

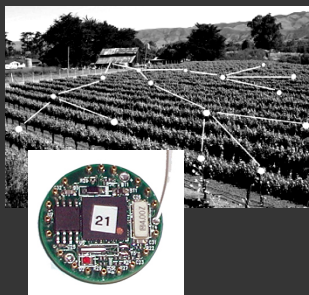
More and more unseen data

- 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

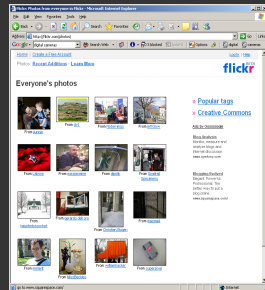
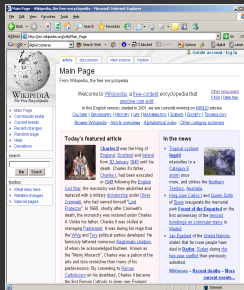
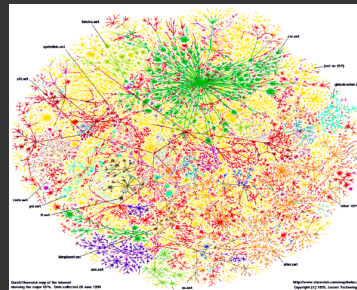


Photo sharing/annotation
flickr.com



Group Authored
Encyclopedia
wikipedia.org



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

Internet

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

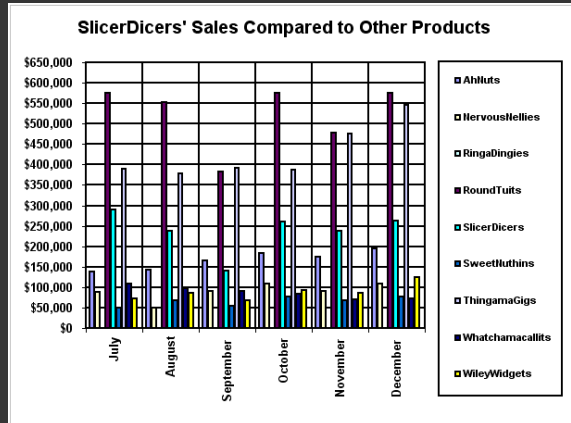
Topics

1. Data and image models

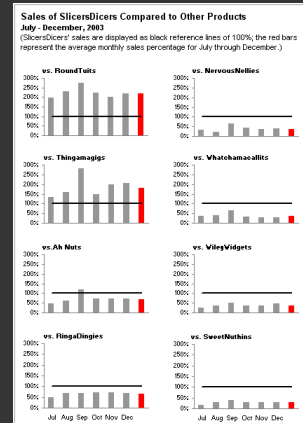
LES VARIABLES DE L'IMAGE											
POINTS			LIGNES			ZONES					
XY 2 DIMENSIONS DU PLAN	x	x	x	/	2	/	14 15 9 14 15 9	2 18 2 1 21 15 14 15 9	OQ	≠	
Z TAILLE	■	■	■	/	2	/	■	■	OQ	≠	
VALEUR	■	■	■	/	2	/	■	■	O	≠	
LES VARIABLES DE SÉPARATION DES IMAGES											
GRAIN	■	■	■	/	2	/	■	■	O	≠	
COULEUR	■	■	■	/	2	/	■	■	≡	≠	
ORIENTATION	■	■	■	/	2	/	■	■	≡	≠	
FORME	■	■	■	/	2	/	■	■	≡	≠	

[Bertin, Graphics and Graphic Information Processing 1981]

2. Visualization Design

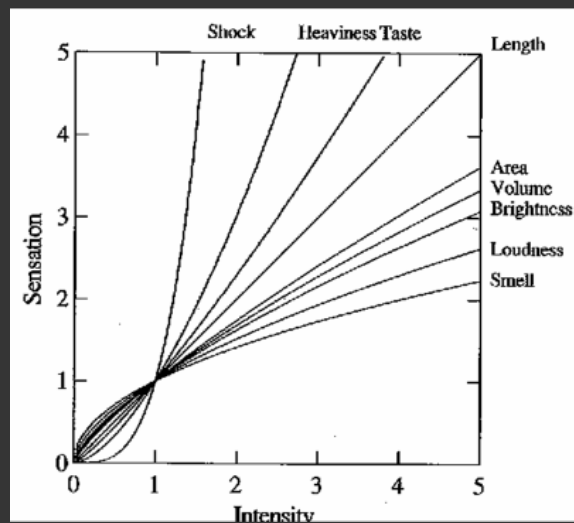


Problematic design



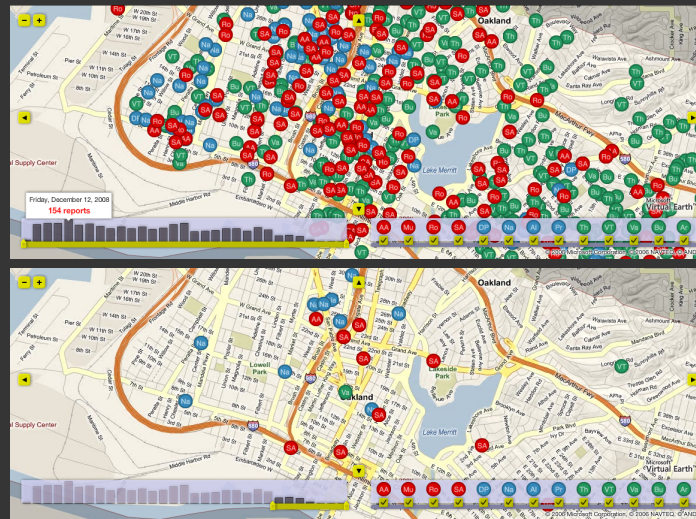
Redesign

3. Perception



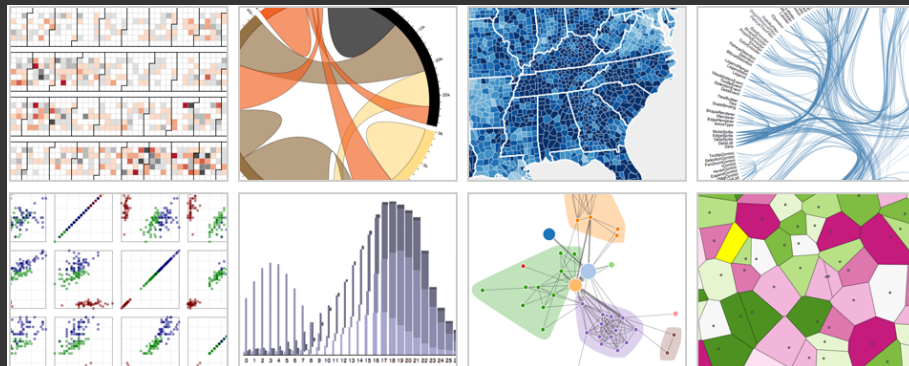
The psychophysics of sensory function [Stevens 61]

4. Interaction



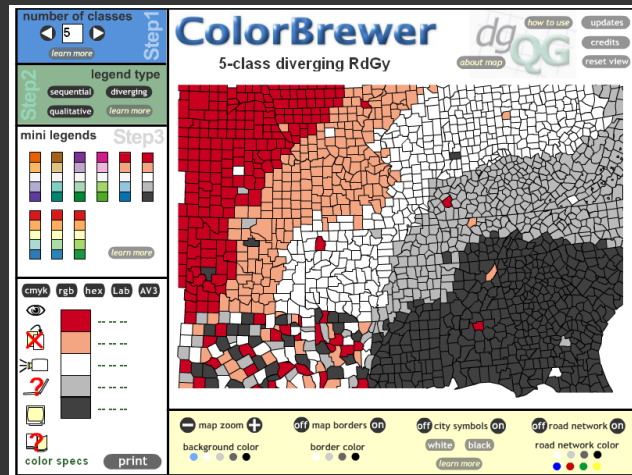
Oakland Crimespotting (crimespotting.org) [Stamen]

5. Building interactive visualizations with D3



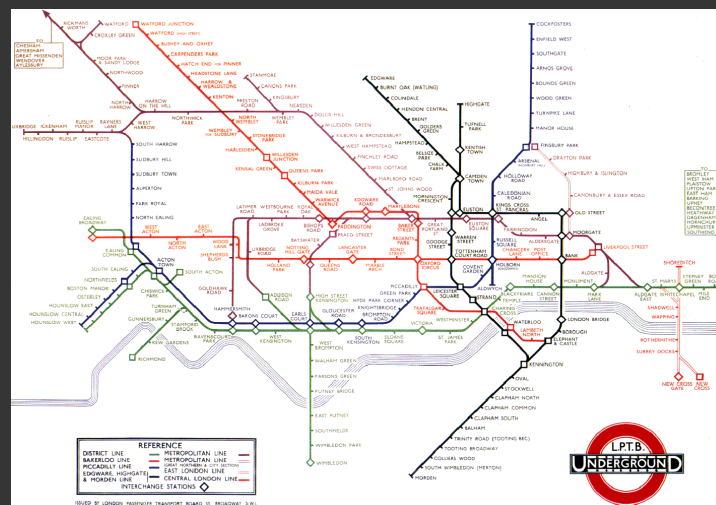
D3: Data Driven Documents [Bostock 2011]

6. Color



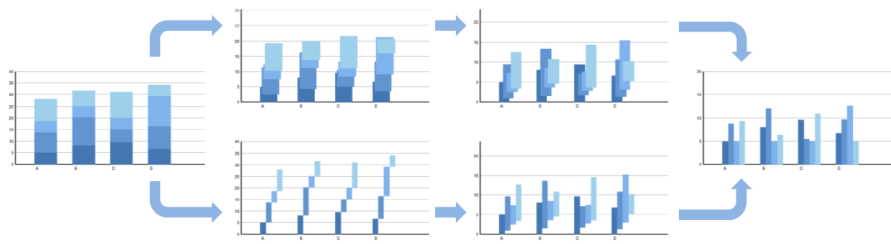
[from Cynthia Brewer <http://www.personal.psu.edu/faculty/c/a/cab38/>]

7. Spatial Layout



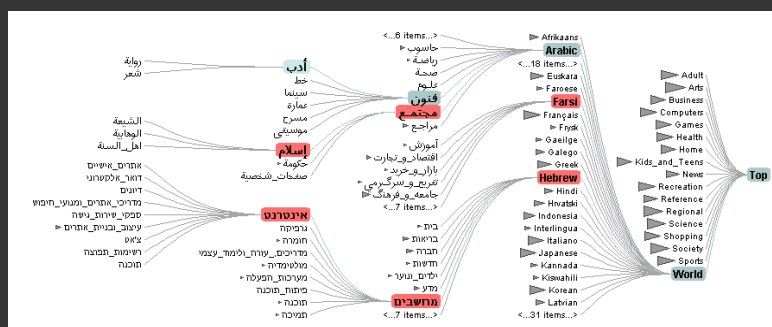
London underground [Beck 33]

8. Animation



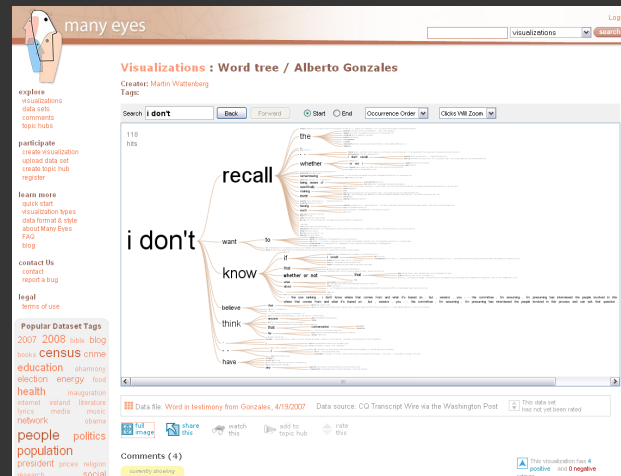
Animated Transitions [Heer 07]

9. Trees and graphs



Degree-of-Interest Trees [Heer 2004]

10. Text visualization



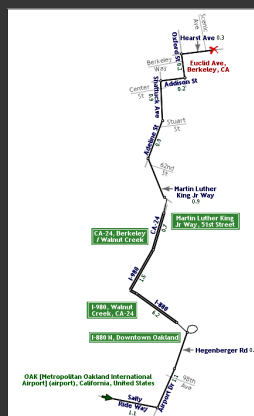
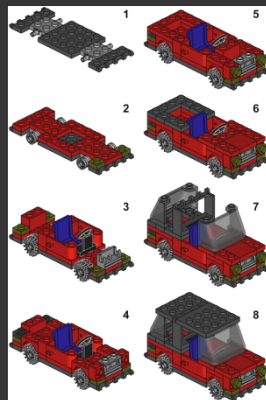
Word Trees [Wattenberg 2008]

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

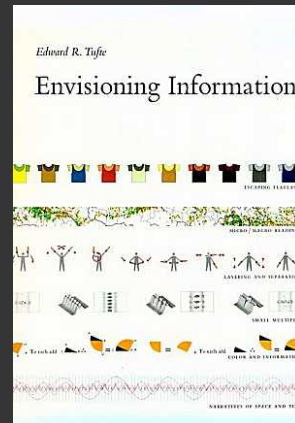
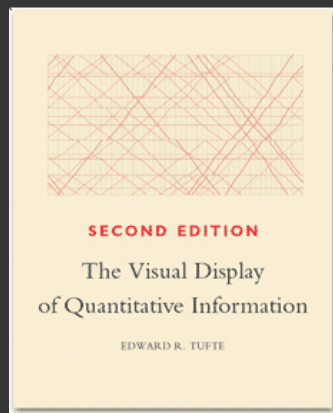
Zach Maurer

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: Crookston, Duluth, Grand Rapids, Morris, University Farm, Waseca

Variety: Glabron, Manchuria, No 457, No 462, No 475, Peatland, Svansota, Trebi, Velvet, Wisc. 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

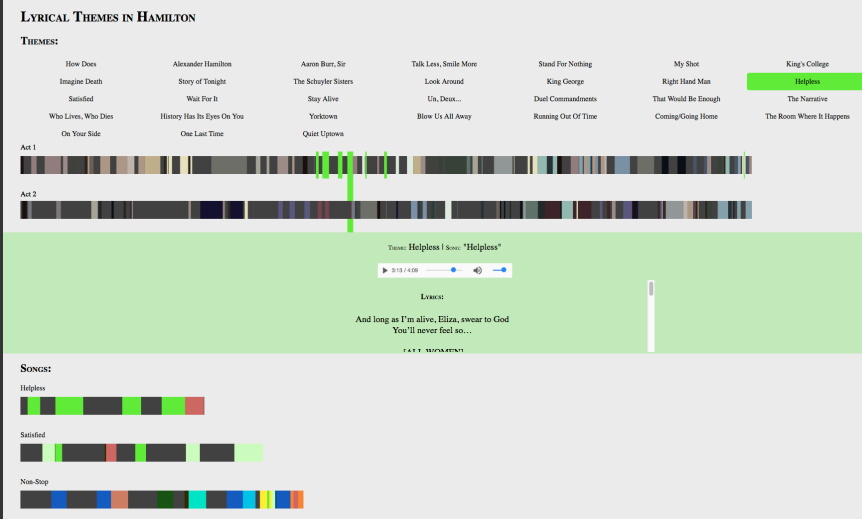
Final project

- 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

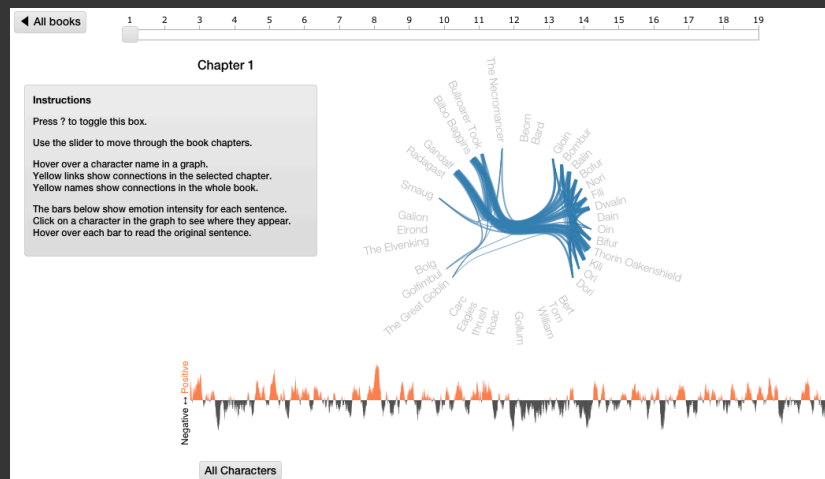
- IEEE Visualization
- IEEE Information Visualization
- SIGGRAPH

Structure of Musicals



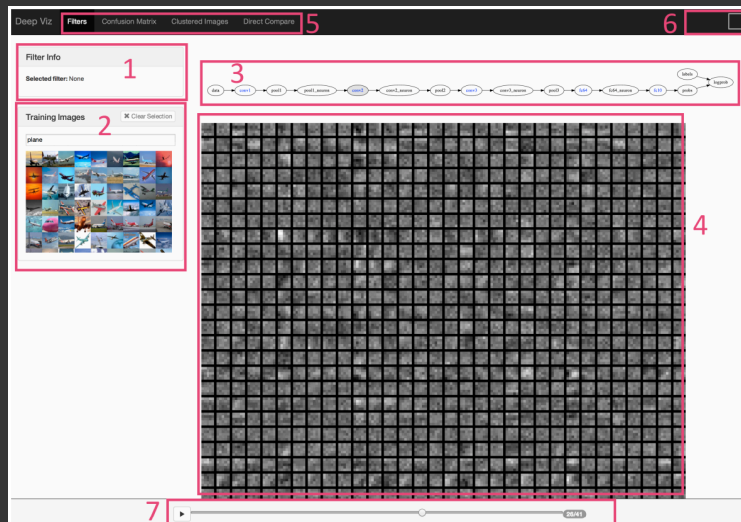
Lyrical themes in Hamilton [Townley-Smith, Sterman, Cook 2016]

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]