The Purpose of Visualization

Maneesh Agrawala

CS 448B: Visualization
Fall 2021

1

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

2

2020: 64.2 zetabytes

[IDC 2021]

3

2020: 64.2 zetabytes
10x increase over 5 years

[IDC 2021]

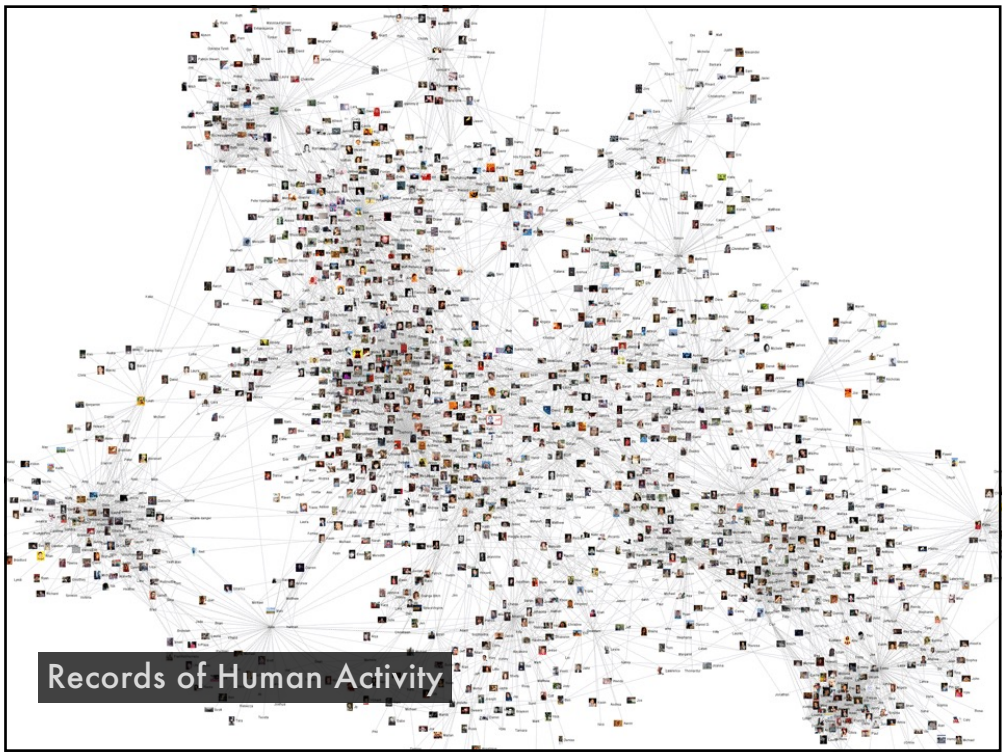
4


Physical Sensors
Image courtesy cabspotting.org

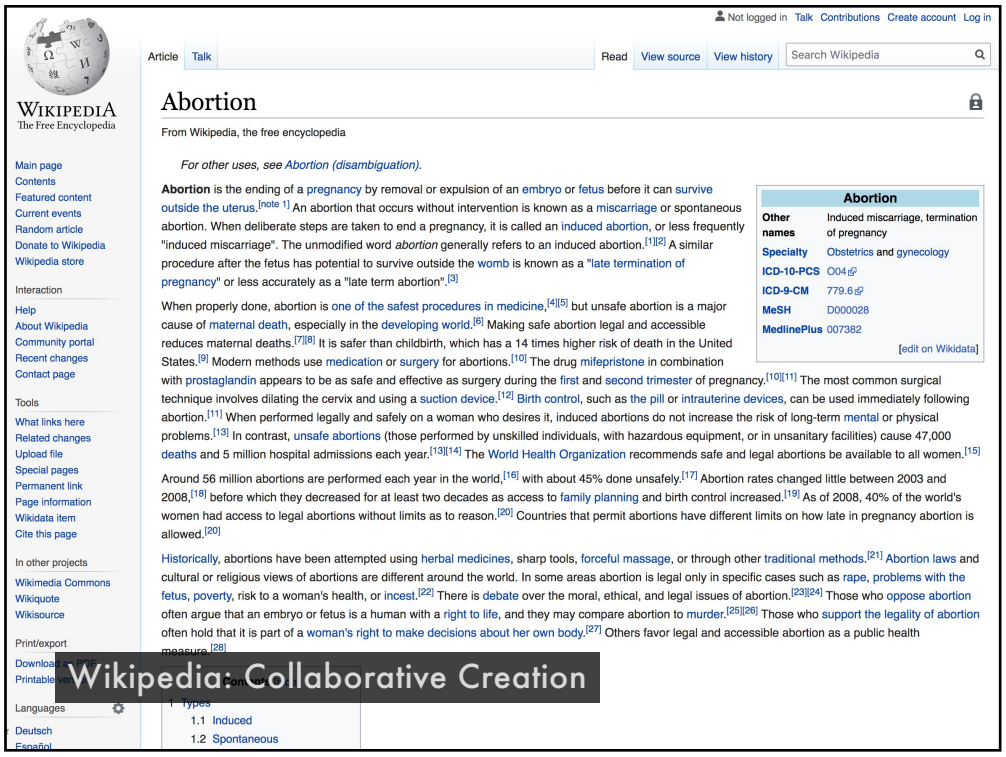
5


HR/ECG
1/min
97
Art
mmHg sys/dia
82/60
SpO2
%
99
RR/CO2
1/min
HR/ECG
1/min
79
Art
mmHg sys/dia
152/79
SpO2
%
95
RR/CO2
1/min
HR/ECG
1/min
64
Art
mmHg sys/dia
93/55
SpO2
%
99
RR/CO2
1/min
HR/ECG
1/min
99
mmHg sys/dia
SpO2
%
RR/Imp
1/min
Health & Medicine

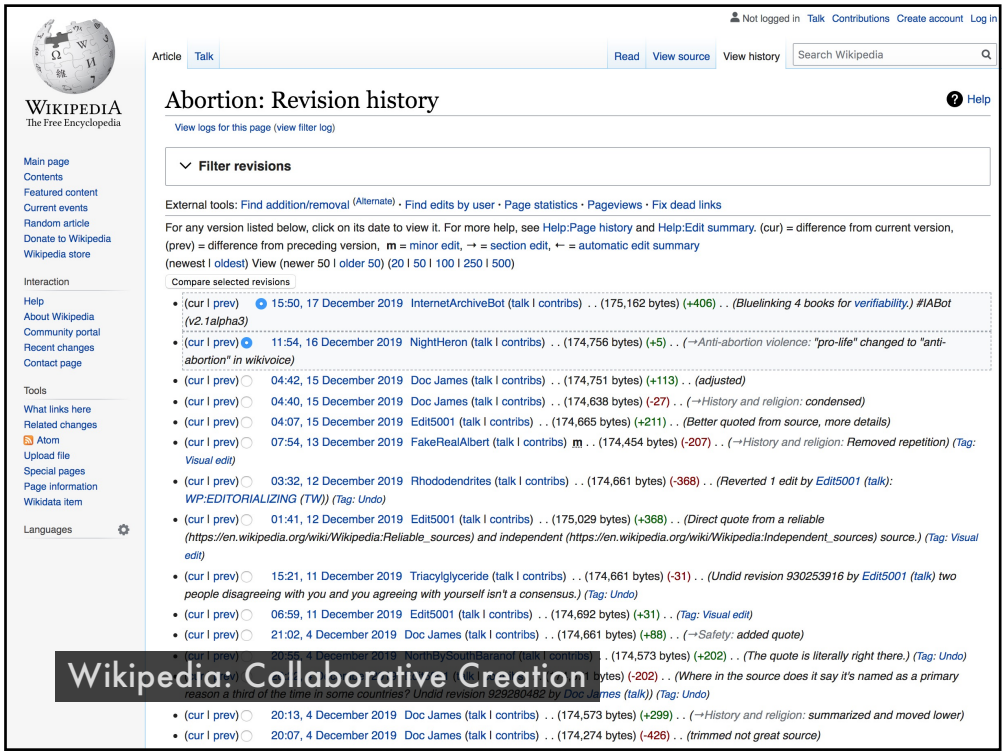
6


Records of Human Activity

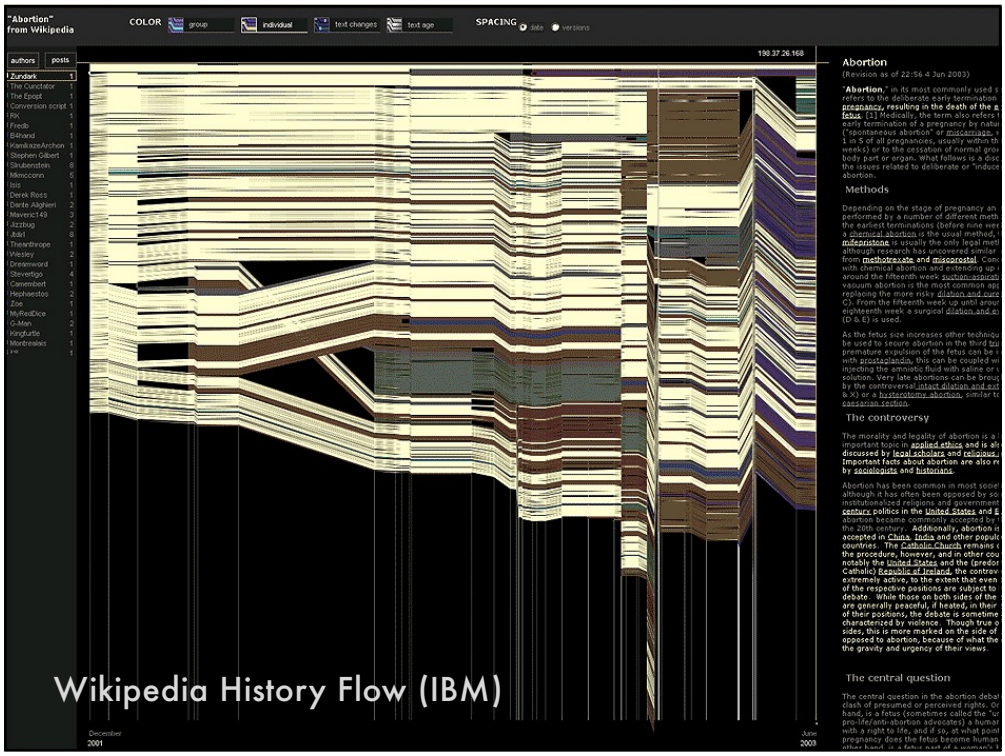
7


Wikipedia: Collaborative Creation
Not logged in Talk Contributions Create account Log in
Article Talk
Read View source View history
Search Wikipedia
WIKIPEDIA
The Free Encyclopedia
Abortion
From Wikipedia, the free encyclopedia
For other uses, see Abortion (disambiguation).
Abortion is the ending of a pregnancy by removal or expulsion of an embryo or fetus before it can survive outside the uterus.[note 1] 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 word abortion generally refers to an induced abortion.[1][2] 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".[3]
When properly done, abortion is one of the safest procedures in medicine,[4][5] but unsafe abortion is a major cause of maternal death, especially in the developing world.[6] Making safe abortion legal and accessible reduces maternal deaths.[7][8] It is safer than childbirth, which has a 14 times higher risk of death in the United States.[9] Modern methods use medication or surgery for abortions.[10] The drug mifepristone in combination with prostaglandin appears to be as safe and effective as surgery during the first and second trimester of pregnancy.[10][11] The most common surgical technique involves dilating the cervix and using a suction device.[12] Birth control, such as the pill or intrauterine devices, can be used immediately following abortion.[11] 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.[13] 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.[13][14] The World Health Organization recommends safe and legal abortions be available to all women.[15]
Around 56 million abortions are performed each year in the world,[16] with about 45% done unsafely.[17] Abortion rates changed little between 2003 and 2008,[18] before which they decreased for at least two decades as access to family planning and birth control increased.[19] As of 2008, 40% of the world's women had access to legal abortions without limits as to reason.[20] Countries that permit abortions have different limits on how late in pregnancy abortion is allowed.[20]
Historically, abortions have been attempted using herbal medicines, sharp tools, forceful massage, or through other traditional methods.[21] 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.[22] There is debate over the moral, ethical, and legal issues of abortion.[23][24] 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.[25][26] 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.[27] Others favor legal and accessible abortion as a public health measure.[28]
Abortion
Other names Induced miscarriage, termination of pregnancy
Specialty Obstetrics and gynecology
ICD-10-PCS O04
ICD-9-CM 779.6
MeSH D000028
MedlinePlus 007382
[edit on Wikidata]
Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store
Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page
Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page
In other projects
Wikimedia Commons
Wikiquote
Wikisource
Print/export
Languages
Deutsch
Español
1 Types
1.1 Induced
1.2 Spontaneous

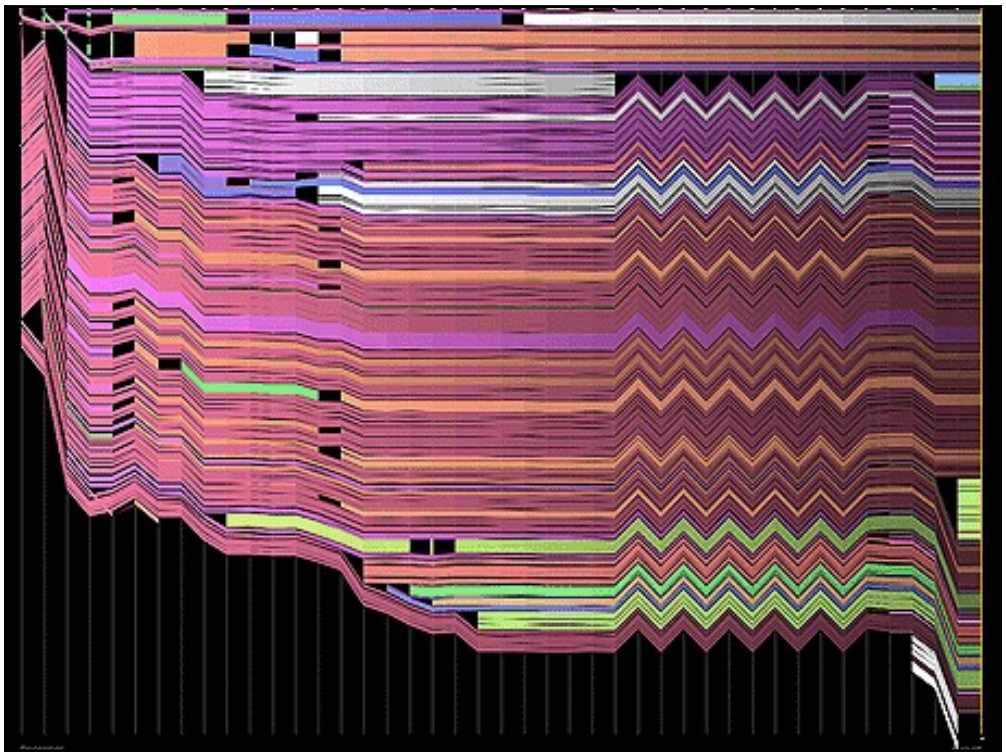
8


Not logged in Talk Contributions Create account Log in
Article Talk Read View source View history Search Wikipedia
WIKIPEDIA
The Free Encyclopedia
Abortion: Revision history
Help
View logs for this page (view filter log)
Filter revisions
External tools: Find addition/removal (Alternate) · Find edits by user · Page statistics · Pageviews · Fix dead links
For any version listed below, click on its date to view it. For more help, see Help:Page history and Help:Edit summary. (cur) = difference from current version, (prev) = difference from preceding version, m = minor edit, → = section edit, ← = automatic edit summary
(newest | oldest) View (newer 50 | older 50) (20 | 50 | 100 | 250 | 500)
Compare selected revisions
Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store
Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page
Tools
What links here
Related changes
Atom
Upload file
Special pages
Page information
Wikidata item
Languages
Wikipedia: Collaborative Creation

9

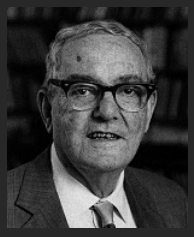

"Abortion" from Wikipedia
COLOR group individual text changes text age
SPACING date versions
authors posts
Abortion
Methods
The controversy
The central question
Wikipedia History Flow (IBM)
December 2001
June 2003

10



11

"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

12

"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."

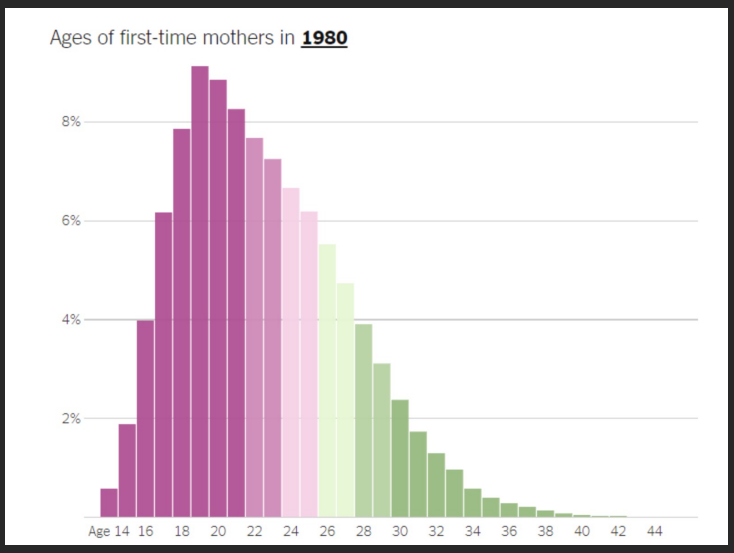
Hal Varian, Google's Chief Economist
The McKinsey Quarterly
January 2009

13

What is visualization?

14

Examples



15

Examples

16

Examples

17

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]

19

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

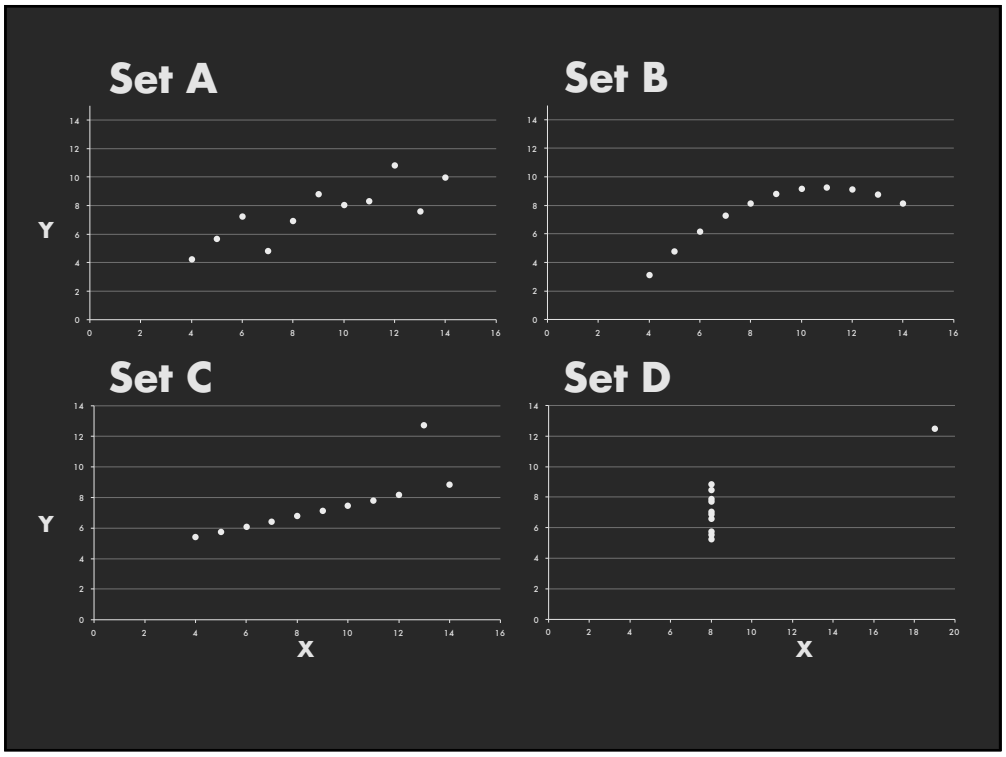
$u_X = 9.0$ $\sigma_X = 3.317$
$u_Y = 7.5$ $\sigma_Y = 2.03$

Linear Regression

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

[Anscombe 73]

20



21

Why do we create visualizations?

22

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

27

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

28

Record Information

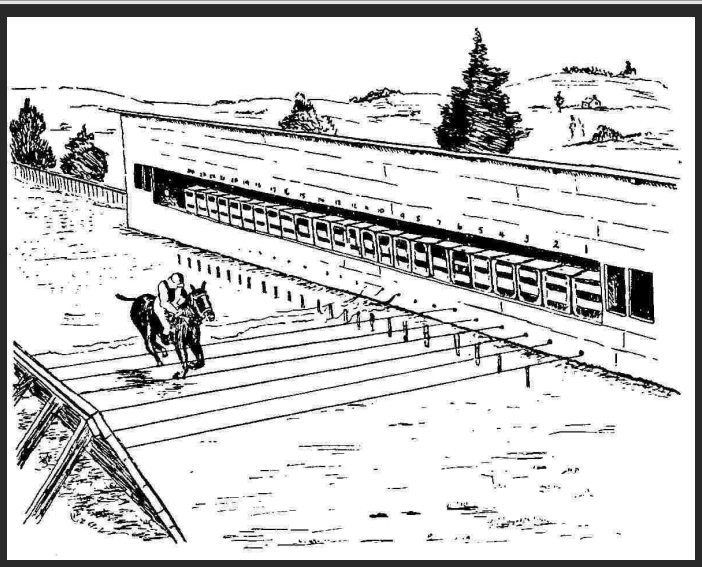
29

Answer question

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

30

Answer question



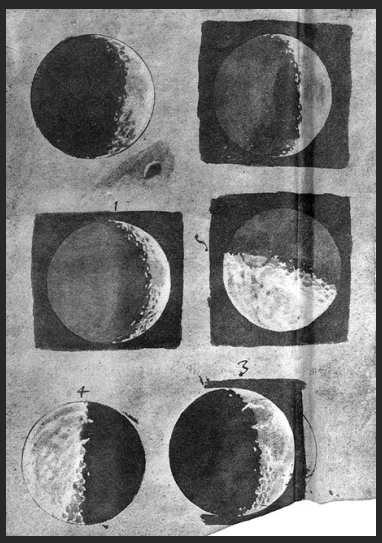
Gallop, Bay Horse “Daisy” [Muybridge 1884-86]

31

Photographs: Phases of the moon

32

Drawing: Phases of the moon



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

33

Support Reasoning

36

Make a decision: Challenger

HISTORY OF O-RING DAMAGE ON SRM FIELD JOINTS

	SRM No.	Cross Sectional View: Erosion Depth (in.)	Cross Sectional View: Perimeter Affected (deg)	Cross Sectional View: Nominal Dia. (in.)	Top View: Length Of Max Erosion (in.)	Top View: Total Heat Affected Length (in.)	Clocking Location (deg)
61A LH Center Field**	22A	None	None	0.280	None	None	36°-66°
61A LH CENTER FIELD**	22A	NONE	NONE	0.280	NONE	NONE	338°-18°
51C LH Forward Field**	15A	0.010	154.0	0.280	4.25	5.25	163
51C RH Center Field (pri)***	15B	0.038	130.0	0.280	12.50	58.75	354
51C RH Center Field (sec)***	15B	None	45.0	0.280	None	29.50	354
41D RH Forward Field	13B	0.028	110.0	0.280	3.00	None	275
41C LH Aft Field*	11A	None	None	0.280	None	None	--
41B LH Forward Field	10A	0.040	217.0	0.280	3.00	14.50	351
STS-2 RH Aft Field	2B	0.053	116.0	0.280	--	--	90

*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 (90°), (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

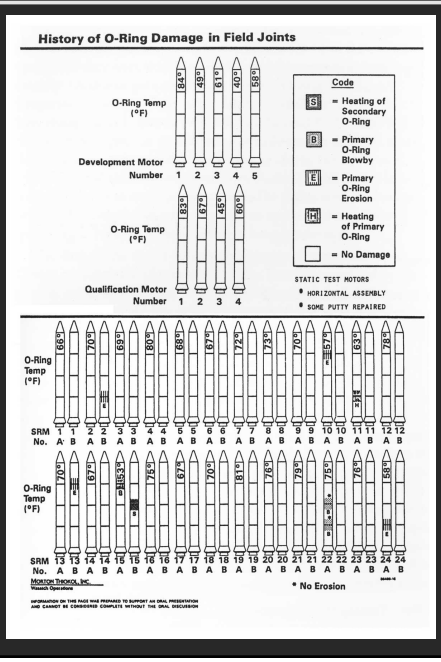
HISTORY OF O-RING TEMPERATURES (DEGREES-F)

MOTOR	MBT	AMB	O-RING	WIND
DM-1	68	36	47	10 MPH
DM-2	76	45	52	10 MPH
QM-1	72.5	40	48	10 MPH
QM-2	76	48	51	10 MPH
SRM-15	52	64	53	10 MPH
SRM-22	77	78	75	10 MPH
SRM-25	55	26	29 27	10 MPH 25 MPH

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

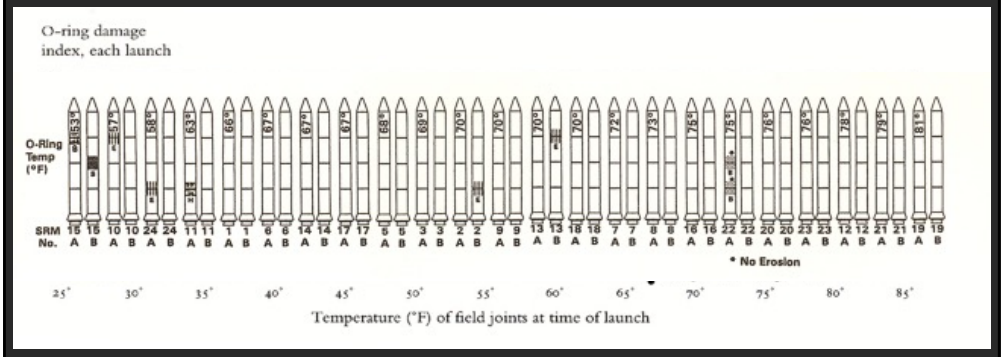
37

Make a decision: Challenger



38

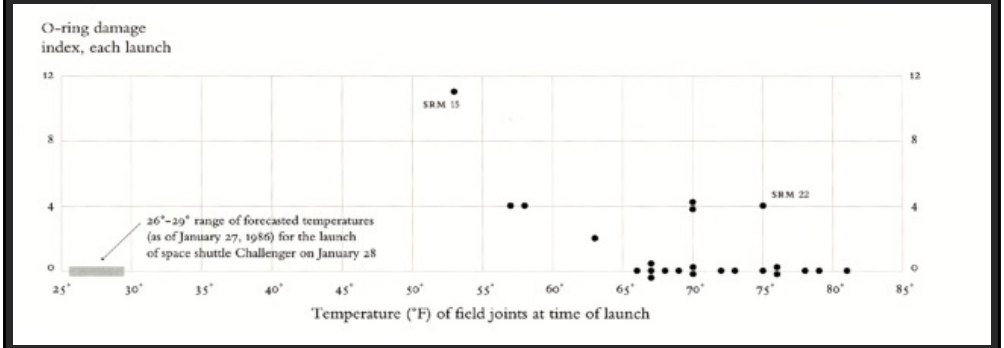
Make a decision: Challenger



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

39

Make a decision: Challenger



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

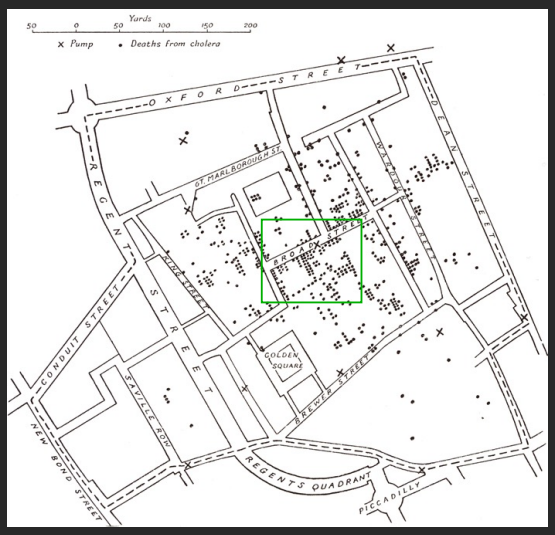
40

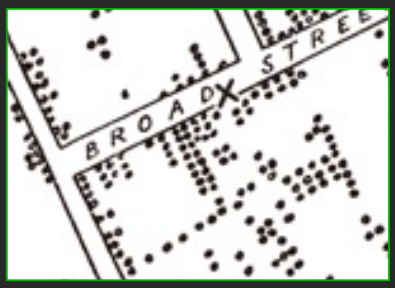
See data in context: Cholera outbreak

In 1854 John Snow plotted the position of each cholera case on a map. [from Tufte 83]

41

See data in context: Cholera outbreak





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

42

Expand memory: Multiplication

Class Exercise

43

Expand memory: Multiplication

```
  34
x 87
----
```

44

Expand memory: Multiplication

```
  34
x 87
----
 238
2720
----
2958
```

45

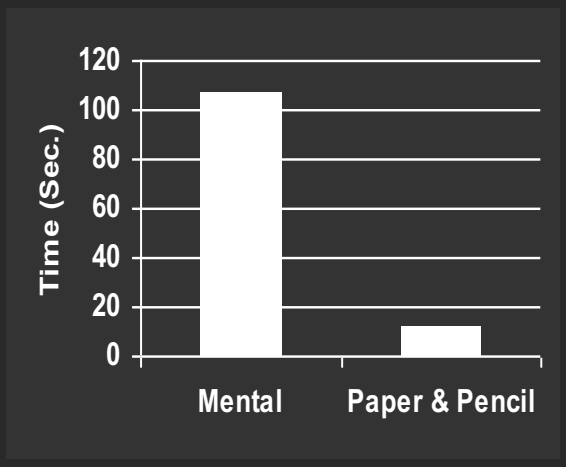
Expand memory: Multiplication

74
x 48

46

Expand memory: Multiplication

74
x 48
592
2960
3552



47

Most powerful brain?

Microsoft Excel - animal.xls

File Edit View Insert Format Tools Data Window Help

A1 ID

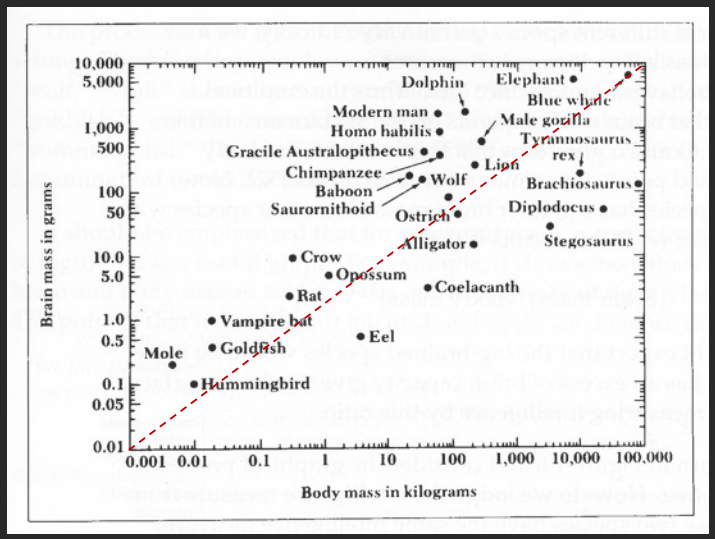
	A	B	C	D	E
1	ID	Name	Body Weight	Brain Weight	
2	1	Lesser Short-tailed Shrew	5	0.14	
3	2	Little Brown Bat	10	0.25	
4	3	Mouse	23	0.3	
5	4	Big Brown Bat	23	0.4	
6	5	Musk Shrew	48	0.33	
7	6	Star Nosed Mole	60	1	
8	7	Eastern American Mole	75	1.2	
9	8	Ground Squirrel	101	4	
10	9	Tree Shrew	104	2.5	
11	10	Golden Hamster	120	1	
12	11	Mole Rate	122	3	
13	12	Galago	200	5	
14	13	Rat	280	1.9	
15	14	Chinchilla	425	6.4	
16	15	Desert Hedgehog	550	2.4	
17	16	Rock Hyrax (a)	750	12.3	
18	17	European Hedgehog	785	3.5	
19	18	Tenrec	900	2.6	
20	19	Arctic Ground Squirrel	920	5.7	
21	20	African Giant Pouched Rat	1000	6.6	
22	21	Guinea Pig	1040	5.5	
23	22	Mountain Beaver	1350	8.1	
24	23	Slow Loris	1400	12.5	
25	24	Genet	1410	17.5	
26	25	Phalanger	1620	11.4	

animal

Ready

52

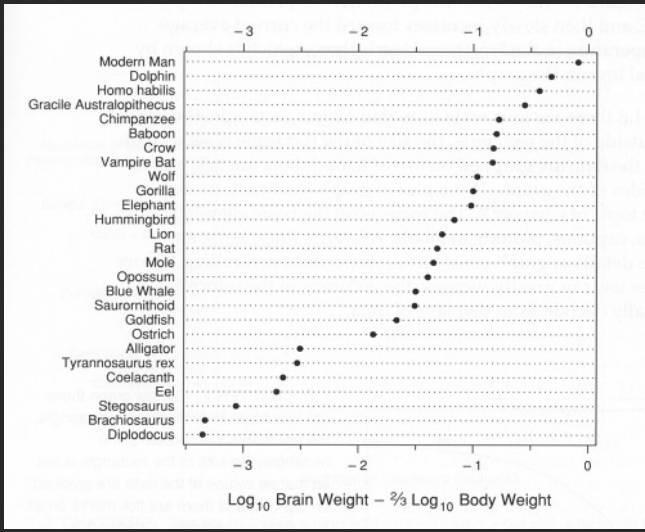
Most powerful brain?



The Dragons of Eden [Carl Sagan]

53

Tell a story: Most powerful brain?



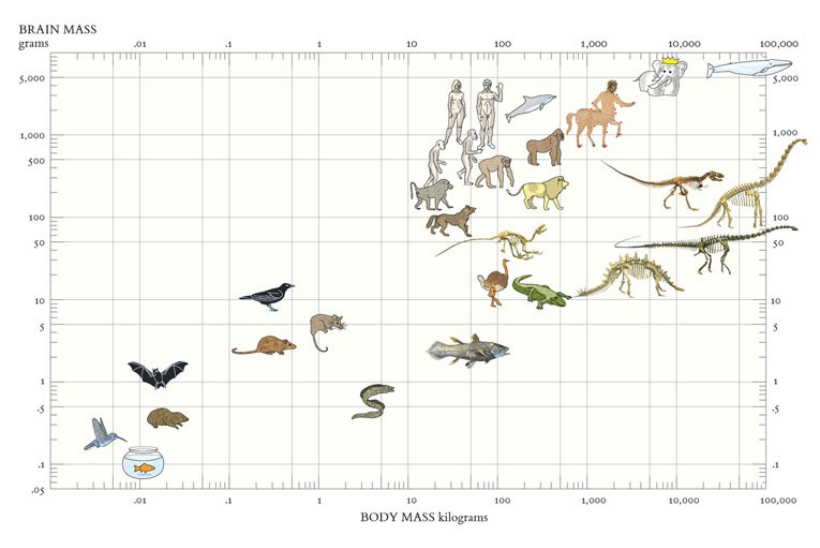
The Elements of Graping Data [Cleveland]

54

Convey Information to Others

55

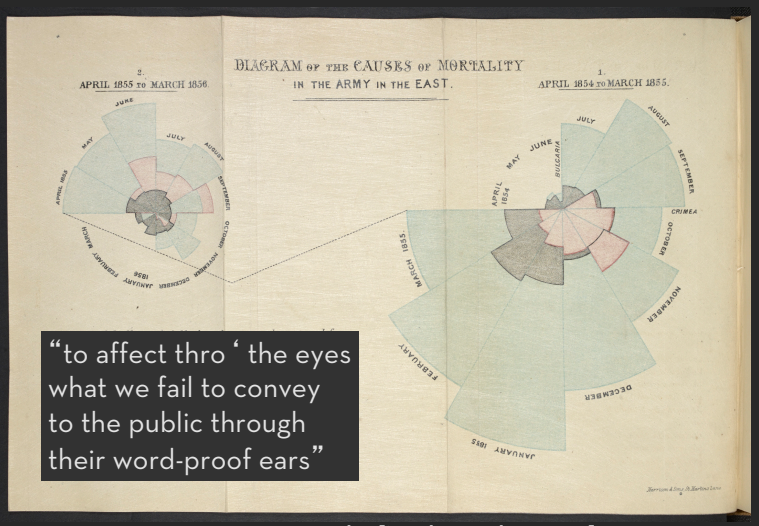
Most powerful brain?



Beautiful Evidence [Tufte]

56

Present argument

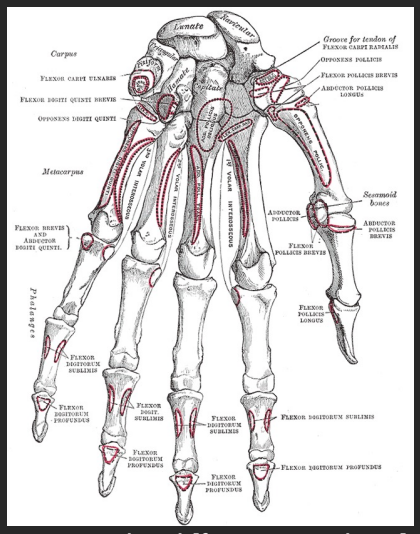


"to affect thro ' the eyes what we fail to convey to the public through their word-proof ears"

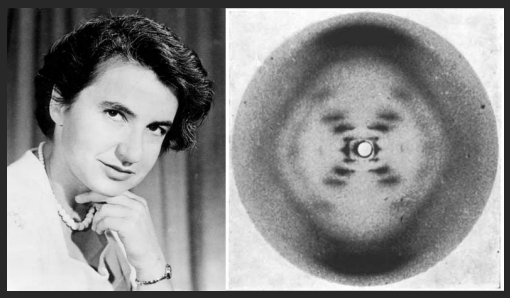
Crimean War Deaths [Nightingale 1858]

57

Inspire

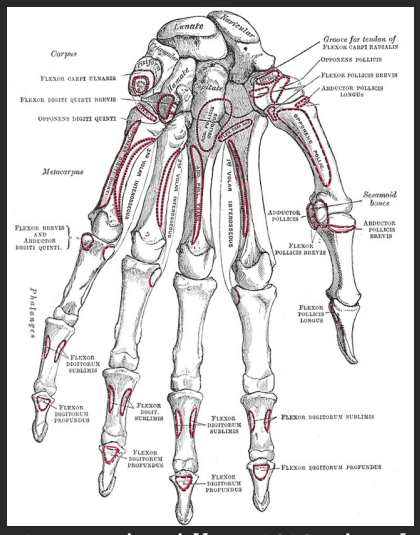


Bones in hand [from 1918 edition]


X-ray crystallography of DNA [Franklin 52]

58

Inspire



Bones in hand [from 1918 edition]


Double helix model [Watson and Crick 53]

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

60

Goals of visualization research

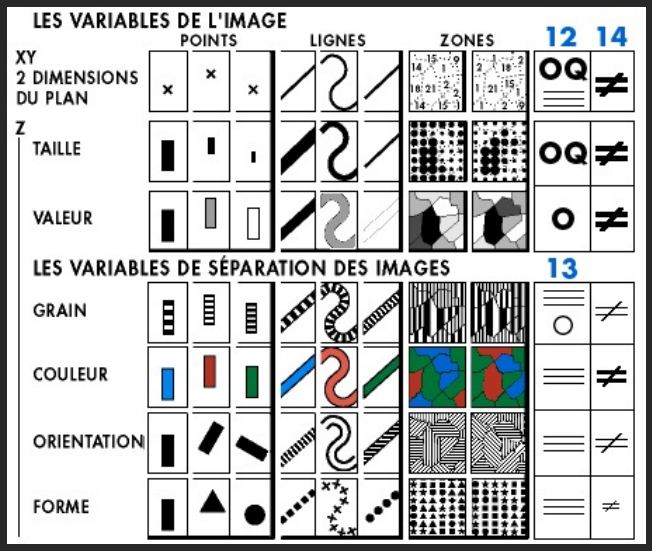
1. **Understand** how visualizations convey information
   - What do people perceive/comprehend ?
   - How do visualizations correspond with mental models of data?
2. **Develop principles and techniques** for creating effective visualizations and supporting analysis
   - Leverage perception and cognition
   - Strengthen connection between visualization and mental models

68

Course Topics

69

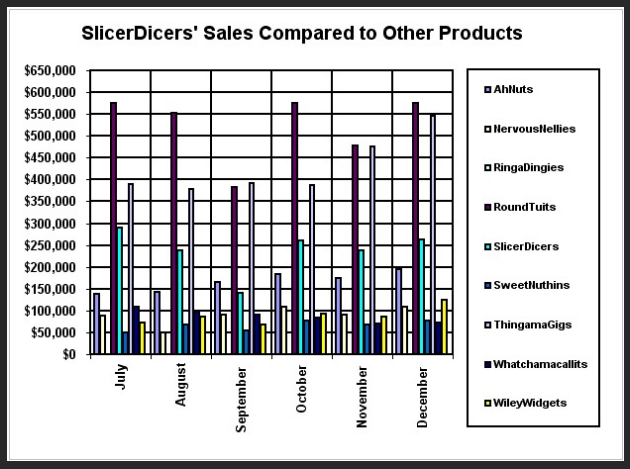
Data and Image Models



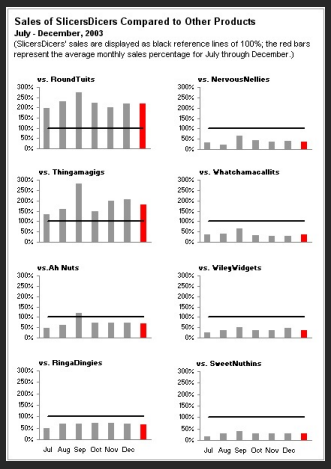
[Bertin, Graphics and Graphic Information Processing 1981]

70

Visualization Design & ReDesign



Problematic design



Redesign

71

Exploratory Data Analysis

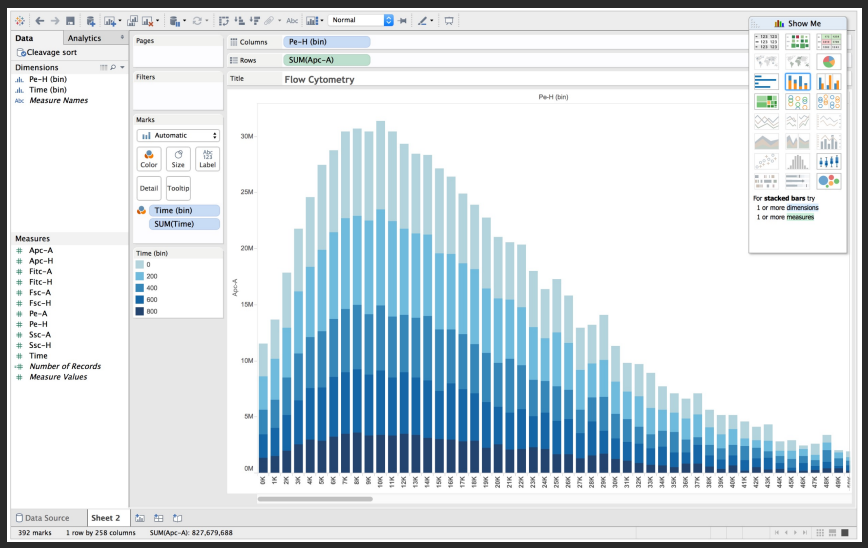
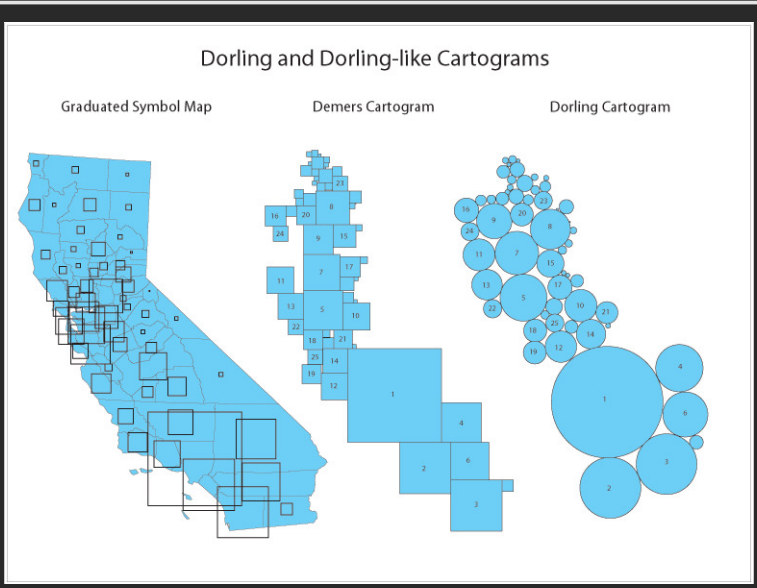


Tableau – based on Polaris [Stolte, Tang, Hanrahan]

72

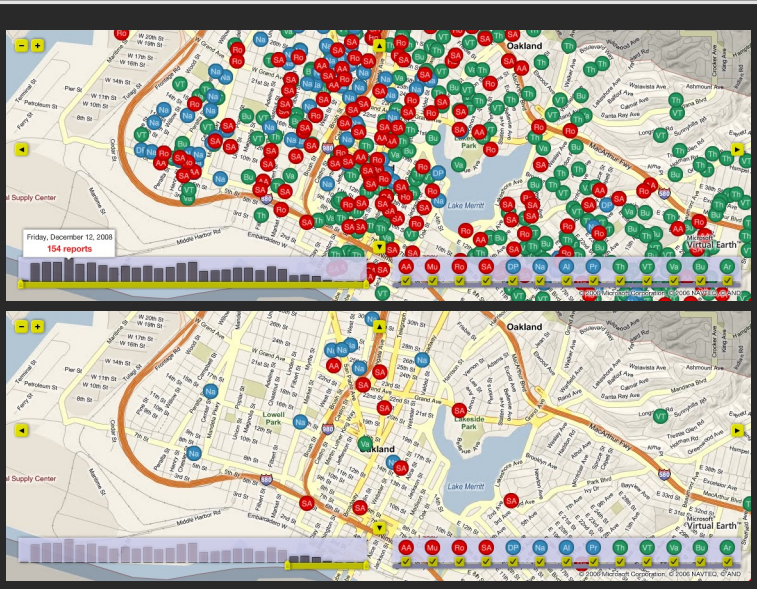
Using Space Effectively



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

73

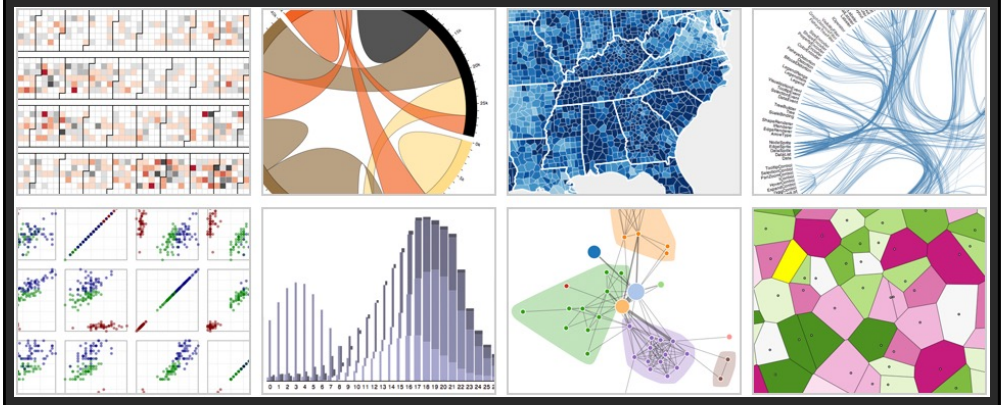
Interaction



Oakland Crimespotting (crimespotting.org) [Stamen]

74

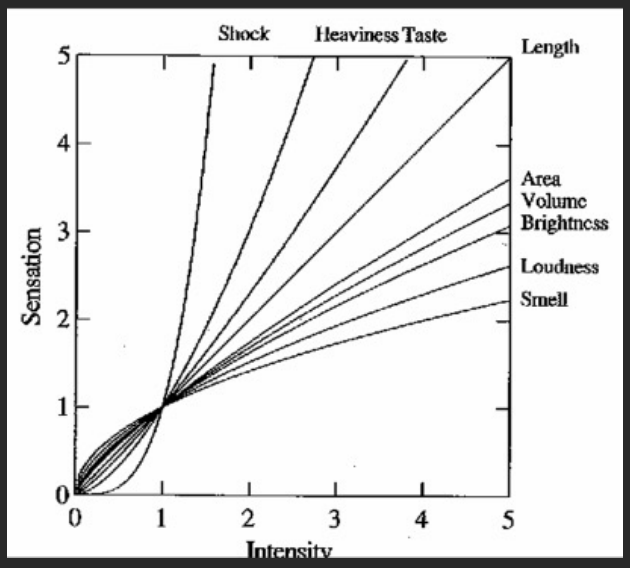
Introduction to D3



D3: Data Driven Documents [Bostock 2011]

75

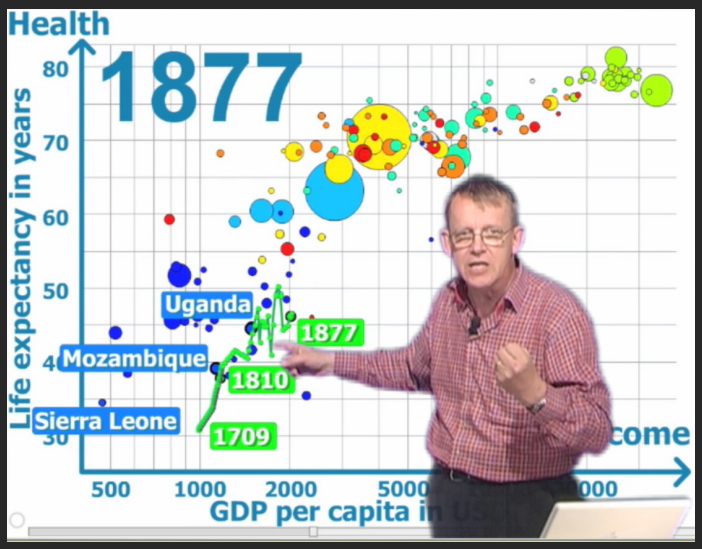
Perception



The psychophysics of sensory function [Stevens 61]

76

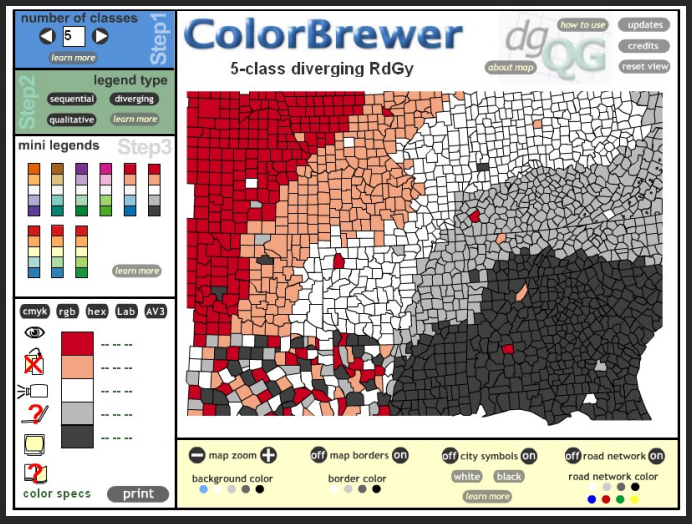
Visual Explainers



Gapminder [Rosling]

77

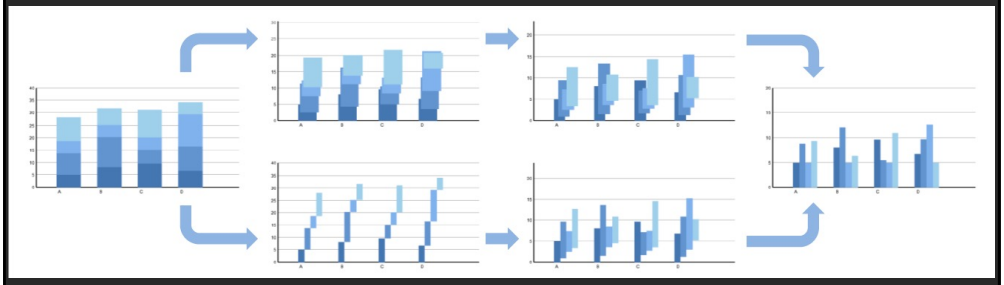
Color



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

78

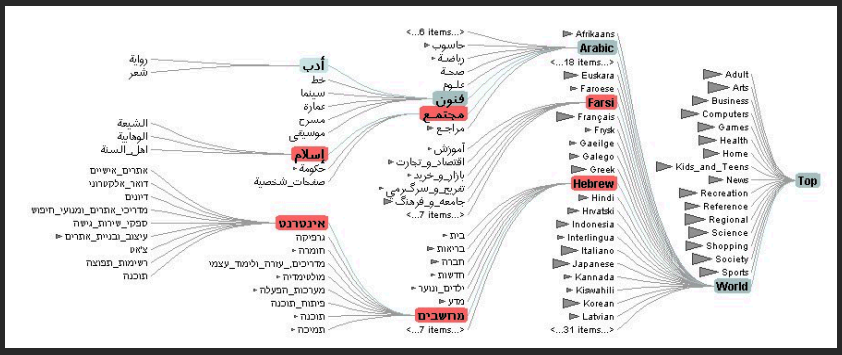
Animation



Animated Transitions [Heer 07]

79

Networks



Degree-of-Interest Trees [Heer 2004]

80

You should expect to

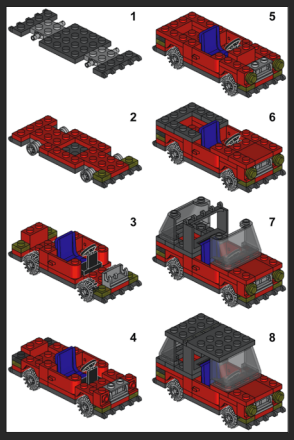
1. ***Design, evaluate and critique*** **visualizations**
2. ***Explore data*** **using existing visualization tools**
3. ***Implement*** **interactive data visualizations**
4. ***Develop*** **a substantial visualization project**

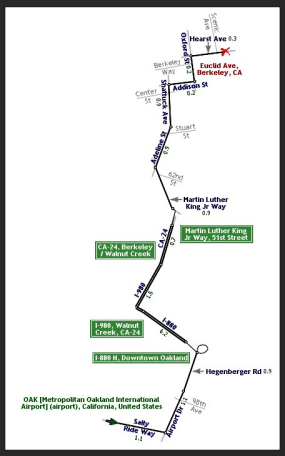
90

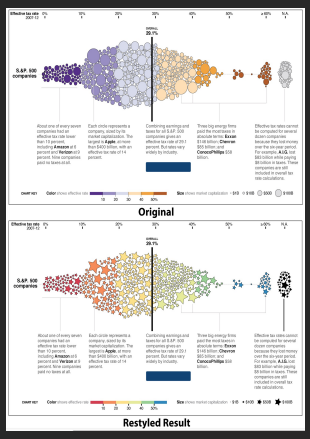
Course Mechanics

91

Instructor: Maneesh Agrawala

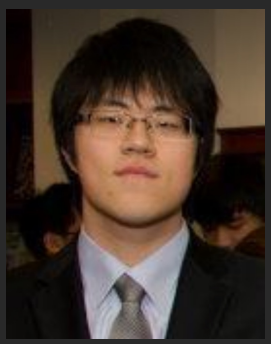






92

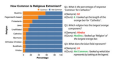
Course Assistant: Dae Hyun Kim



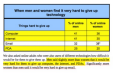
Dae Hyun Kim

I am a Computer Science PhD student at Stanford, working with Prof. Maneesh Agrawala. My research focuses on building natural language interfaces for data visualizations. I did my undergrad in Computer Science at California Institute of Technology.

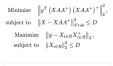
My Works



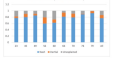
Answering Questions about Charts and Generating Visual Explanations
Dae Hyun Kim, Enamul Hoque and Maneesh Agrawala
ACM Human Factors in Computing Systems (CHI), Apr 2020 (to appear)
PDF | Supplemental PDF | Code & Data



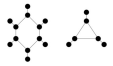
Facilitating Document Reading by Linking Text and Tables
Dae Hyun Kim, Enamul Hoque, Juho Kim and Maneesh Agrawala
The 31st Annual ACM Symposium on User Interface Software and Technology (UIST), Oct 2018. pp. 423-434.
PDF | MP4 | Data



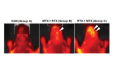
Finding Solutions to Generative Adversarial Privacy
Dae Hyun Kim, Taeyoung Kong and Seungbin Jeong
arXiV preprint arXiv:1810.02069, Oct 2018
PDF



Multiple Item, Ascending Price Auctions: An Experimental Examination of Alternative Auction Sequences
Dae Hyun Kim, Hsing Yang Lee, Travis Maron, Charles R. Plott and Ruijie D. Teo
SSRN, Apr. 2015.
PDF



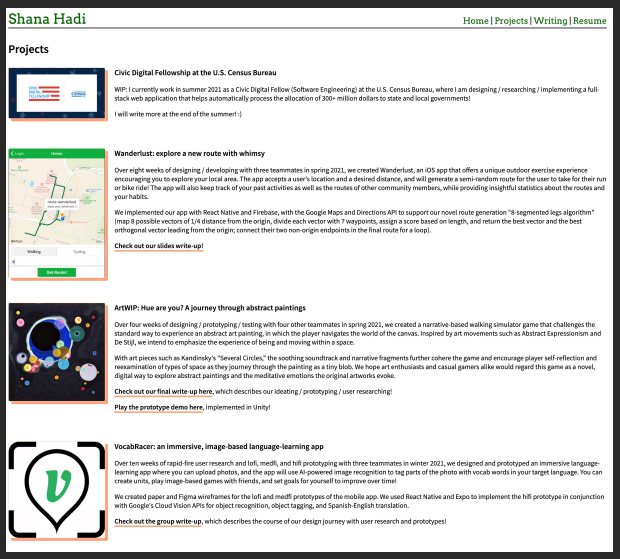
Chromatic Bounds on Orbital Chromatic Roots
Dae Hyun Kim, Alexander H. Mun and Mohamed Omar
The Electronic Journal of Combinatorics, 21(4), Oct 2014. pp. 4-17.
Presented at 2014 Joint Mathematics Meetings (JMM)
(American Mathematical Society + Mathematical Association of America)
PDF



Schedule-Dependent Synergistic Effect of Rituximab on Methotrexate Chemotherapy against Lymphoma of the Central Nervous System
Juyoun Jin, Kyeung Min Joo, Yoonhee Nam, Dae Hyun Kim, Se Jeong Lee, Mi-Young Jo, Younggeon Jin, Hyeong-Seok Kim, Soo Won Seo, Seok Jin Kim, Do-Hyun Nam and Won Seog Kim
Experimental and Therapeutic Medicine, 1(6), Nov 2010. pp. 943-946.

93

Course Assistant: Shana Hadi



94

Office Hours

Maneesh: 2-3pm Wed, Coupa Café Y2E2

Dae Hyun: 10-11:00am Thu, outside CEMEX Aud

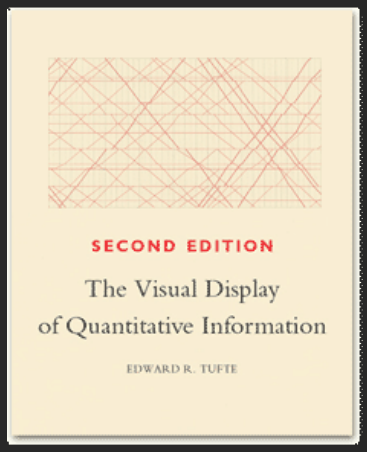
Shana Hadi: 7-8:00pm Sun, via Canvas/Zoom

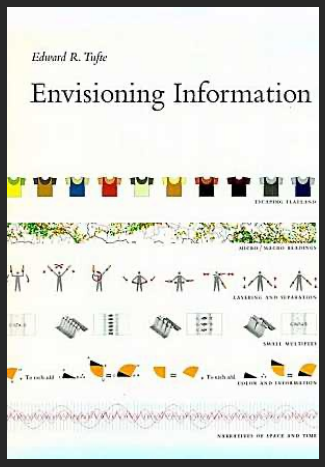
Happy to schedule other OH by appointment
Outside of OH use Slack to connect with us

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

95

Textbooks

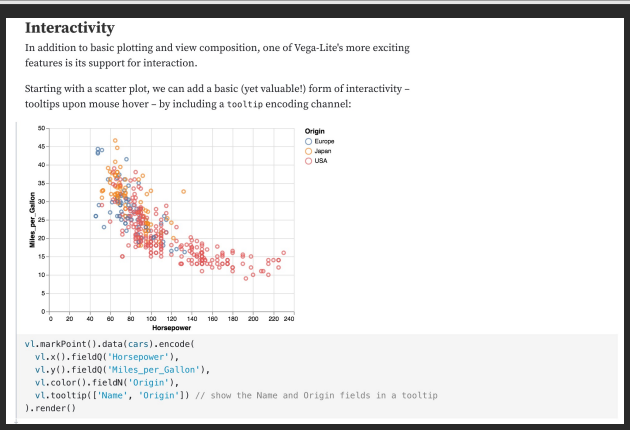




See also: www.edwardtufte.com

96

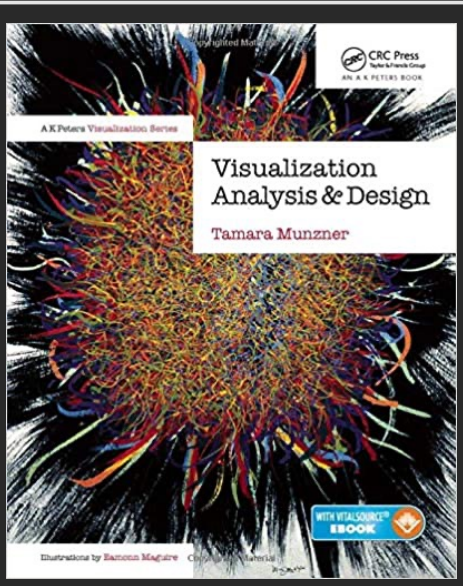
Interactive Notebooks



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

97

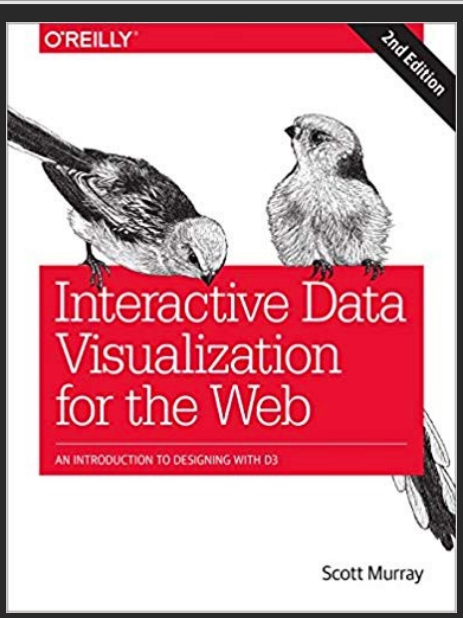
Optional Textbook



For additional theory and depth

98

Optional Book



For learning D3!

Book available online
Code/examples on GitHub

We will be using D3 v7
https://d3js.org

99

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/

100

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

101

Discussion

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

- Attendance is very highly recommended

102

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**

103

Final project

Choice of project type

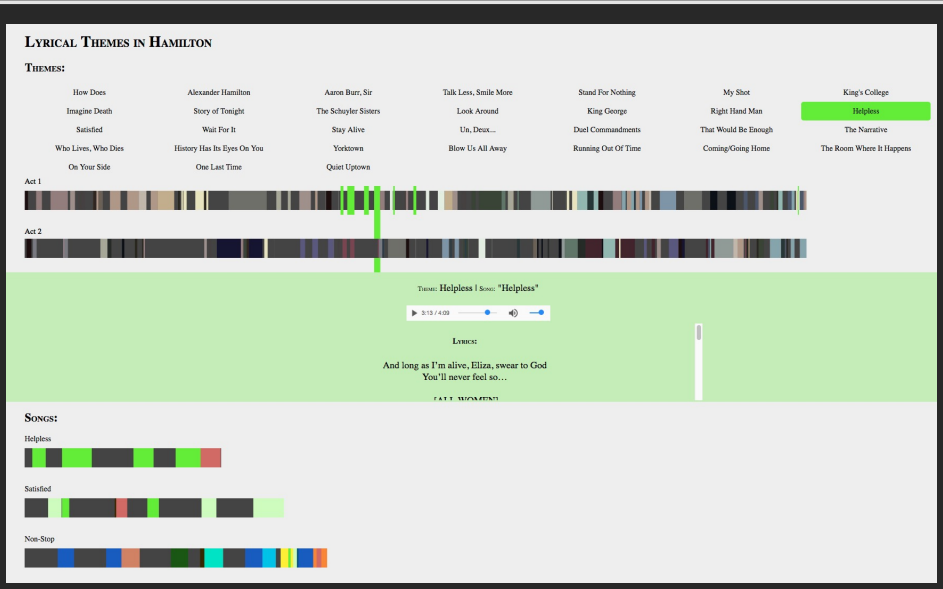
- Create an extended visual explainer
- Small visualization research project

Projects from previous classes have been:

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

104

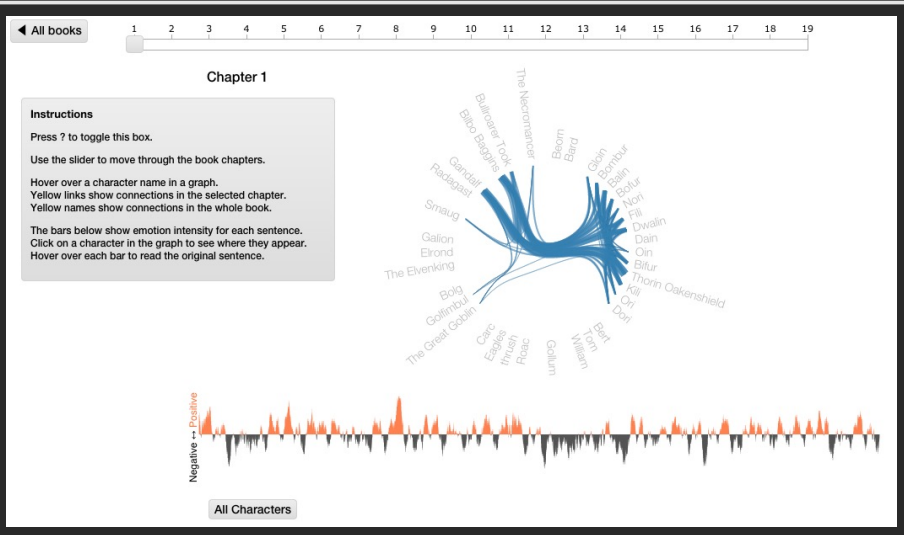
Structure of Musicals



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

105

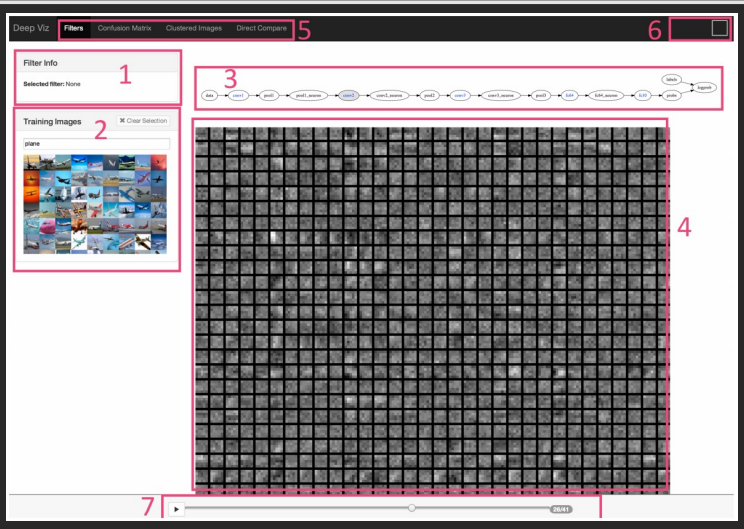
Visualization of Narrative Structure



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

106

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]

107

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:
- **Year:** Academic year between 2011-2012 and 2018-2019.
- **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

108

Assignment 1: Visualization Design

Pick a guiding question, use it to title your visualization

Design a static visualization for that question

You are free to use any tools (including pen & paper)

Deliverables (upload via Canvas; see A1 page)

PDF of your visualization with a short description including design rationale (≤ 4 paragraphs)

Due by 7am on Mon Sep 27

109