The Purpose of Visualization

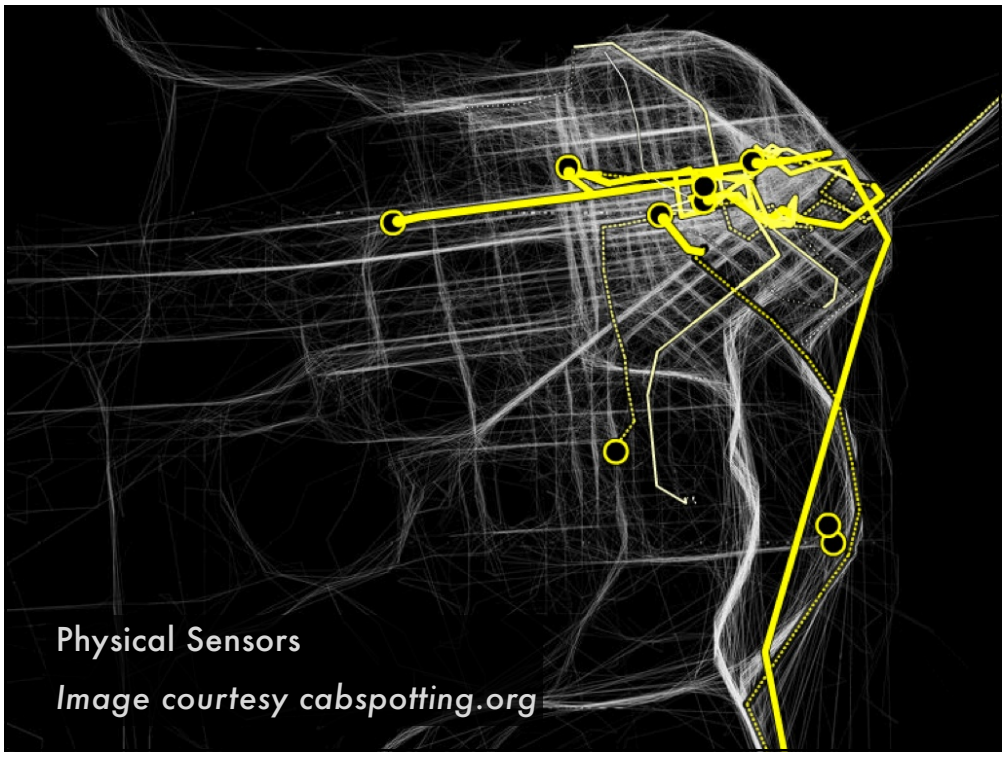
Maneesh Agrawala

CS 448B: Visualization
Fall 2018

How much data (bytes) did we produce in 2016?

2016: 16.1 zetabytes
10x increase over 5 years

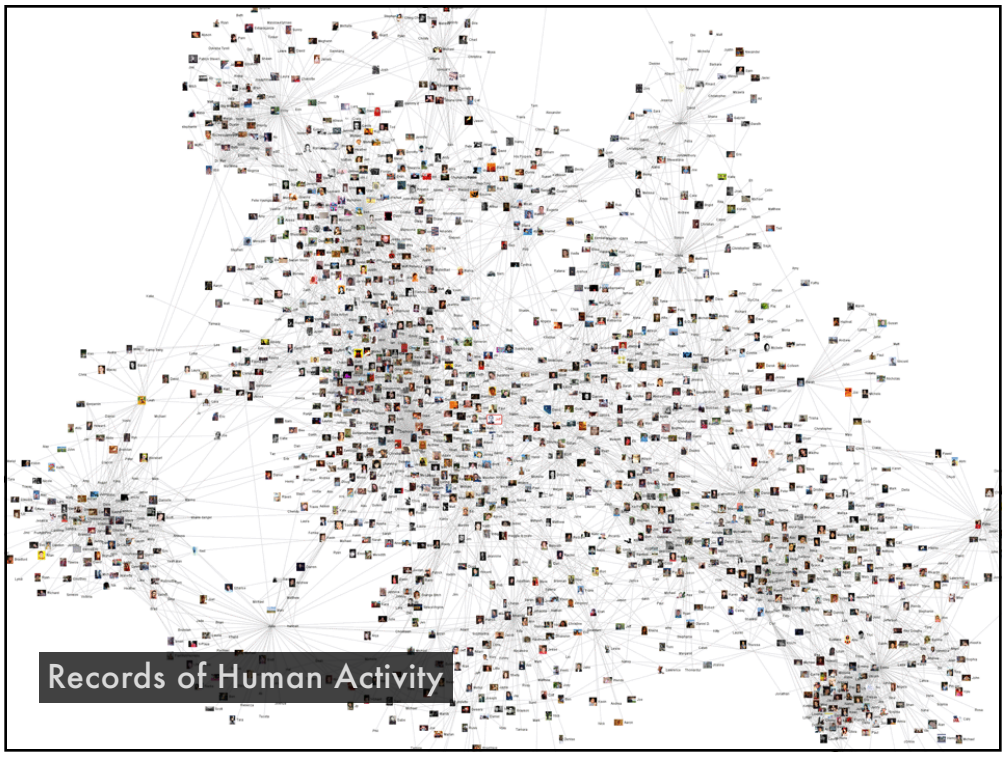
[Gantz 2017]

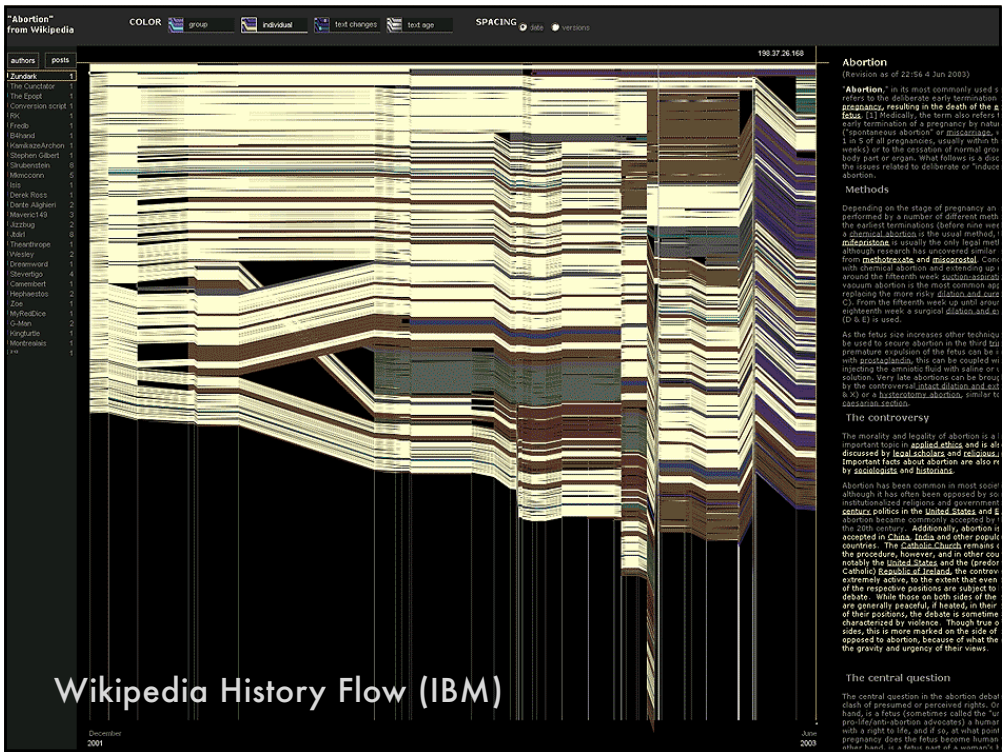


Physical Sensors

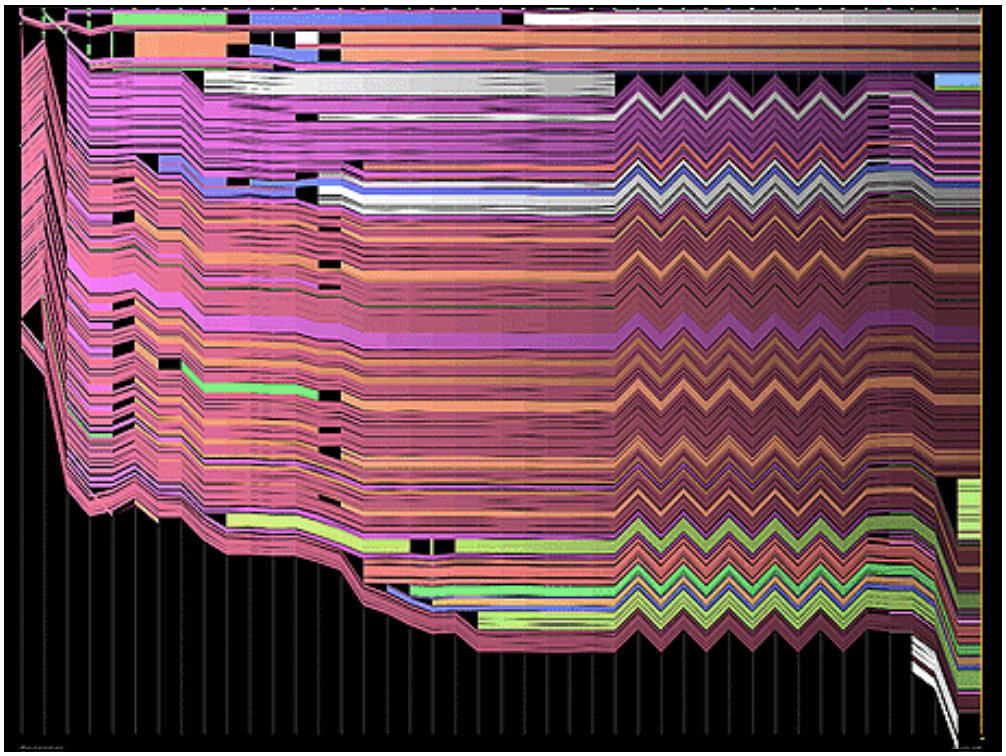
Image courtesy cabspotting.org


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
SpO2
%
RR/Imp
1/min
Health & Medicine

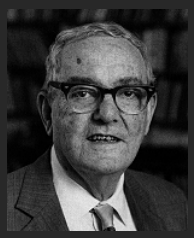

Records of Human Activity



Wikipedia History Flow (IBM)



"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."

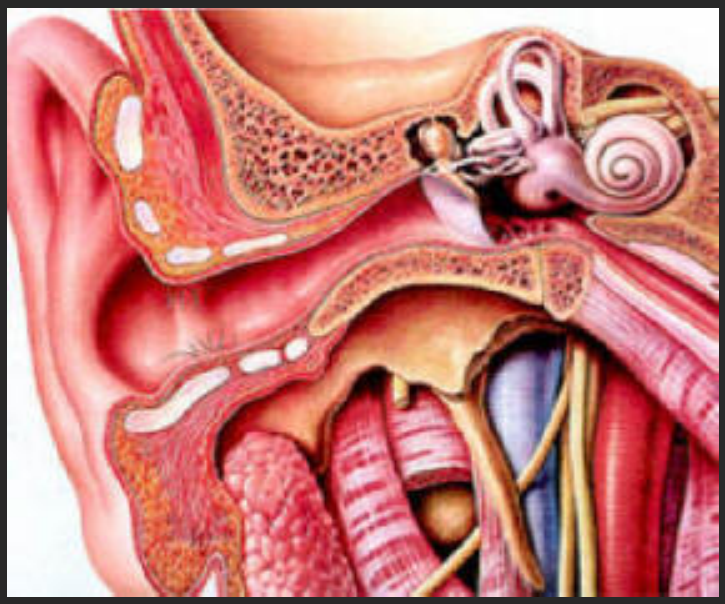


~Herb Simon
as quoted by Hal Varian
Scientific American
September 1995

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

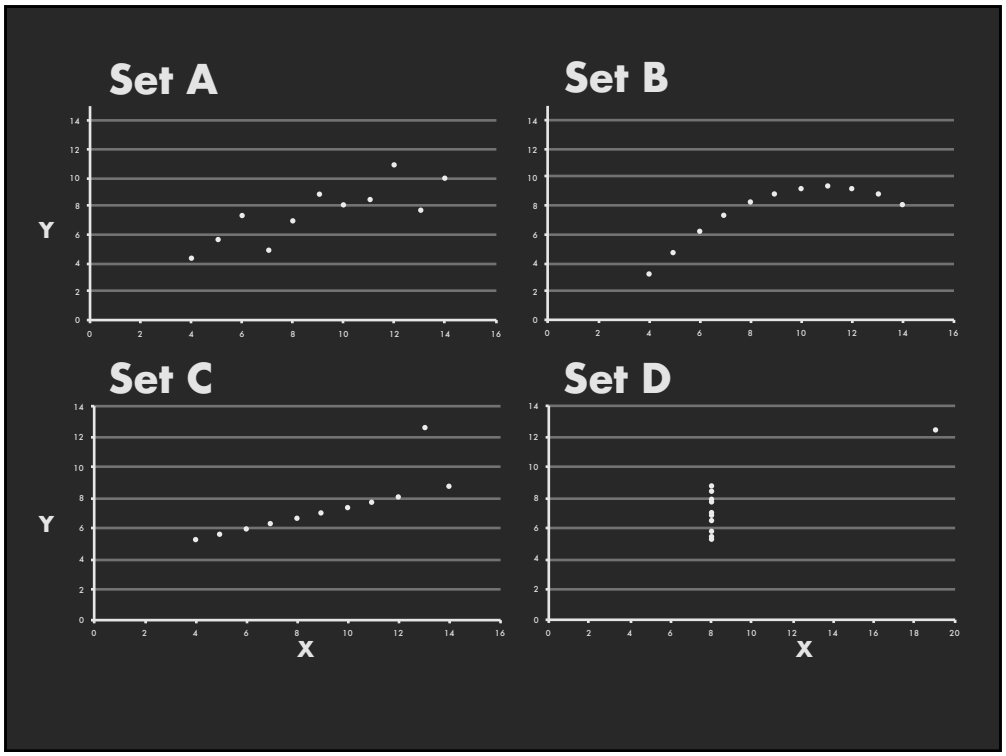
Summary Statistics

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

Linear Regression

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

[Anscombe 73]



Why do we create visualizations?

Why do we create visualizations?

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)

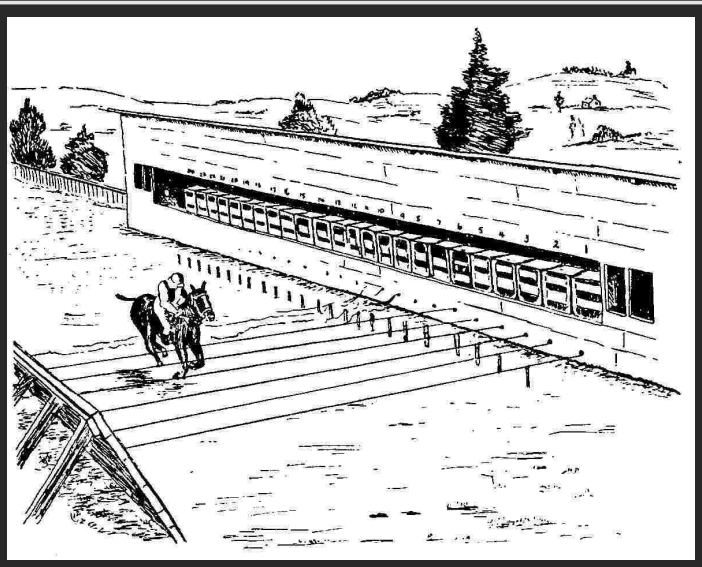
- Share and persuade
- Collaborate and revise
- Emphasize important aspects of data

Record Information

Answer question

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

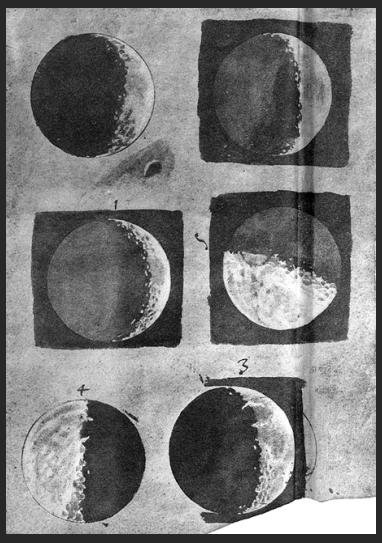
Answer question



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

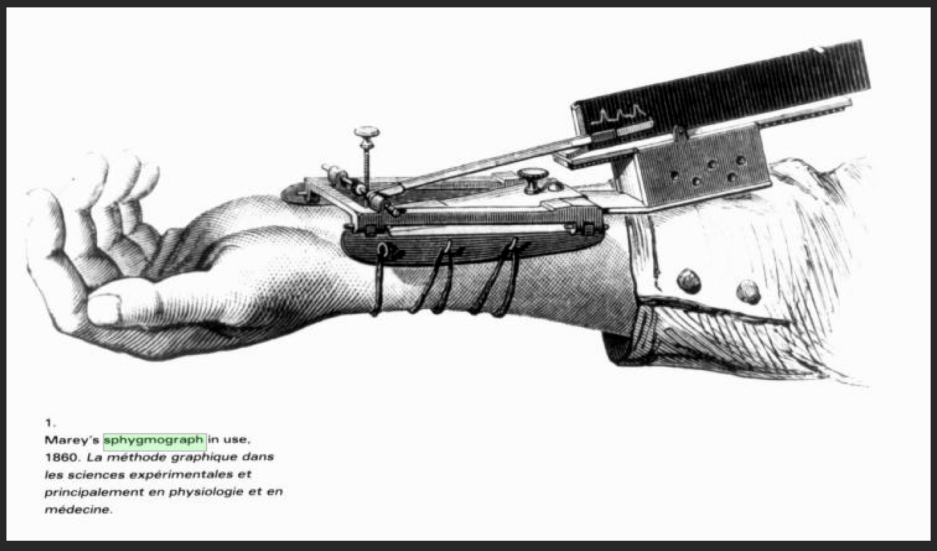
Photographs: Phases of the moon

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

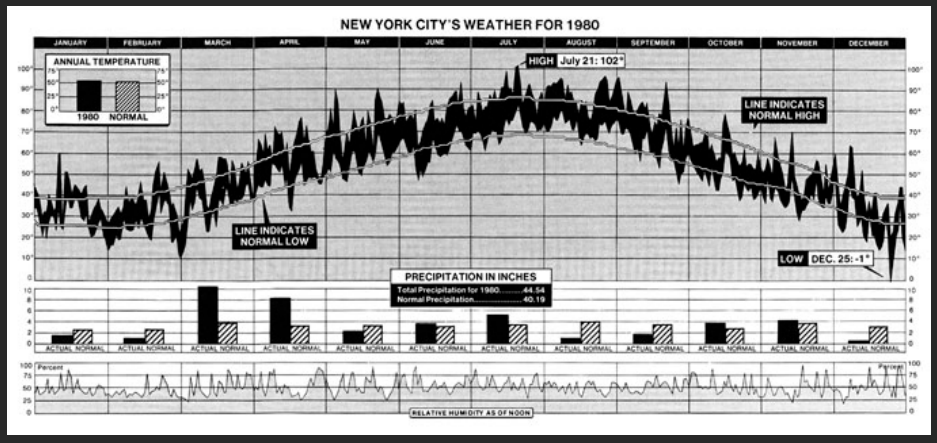


1.
Marey's sphygmograph in use.
1860. *La méthode graphique dans les sciences expérimentales et principalement en physiologie et en médecine.*

Marey's sphygmograph [from Braun 83]

Support Reasoning

Find patterns: New York weather



From the New York Times 1981

Make a decision: Challenger

HISTORY OF O-RING DAMAGE ON SRM FIELD JOINTS

SRM No.	Erosion Depth (in.)	Perimeter Affected (deg)	Nominal Dia. (in.)	Length Of Max Erosion (in.)	Total Heat Affected Length (in.)	Clocking Location (deg)
22A	None	None	0.280	None	None	36°-66°
22A	NONE	NONE	0.280	NONE	NONE	338°-18°
15A	0.010	154.0	0.280	4.25	5.25	163
15B	0.038	130.0	0.280	12.50	58.75	354
15B	None	45.0	0.280	None	29.50	354
13B	0.028	110.0	0.280	3.00	None	275
11A	None	None	0.280	None	None	--
10A	0.040	217.0	0.280	3.00	14.50	351
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°)

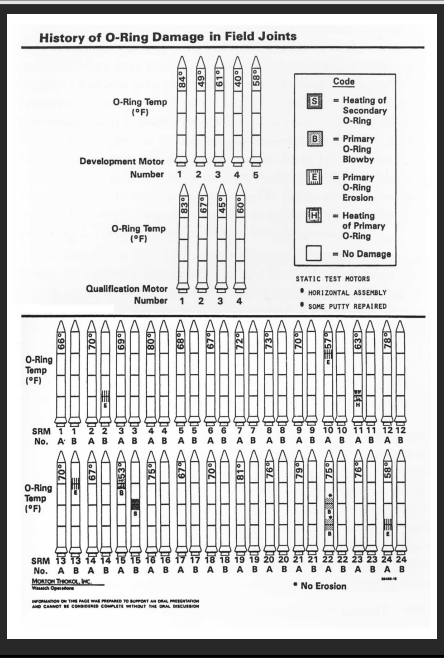
SRM-13A, 15, 16A, 18, 23A 24A
- NOZZLE BLOW-BY

HISTORY OF O-RING TEMPERATURES (DEGREES-F)

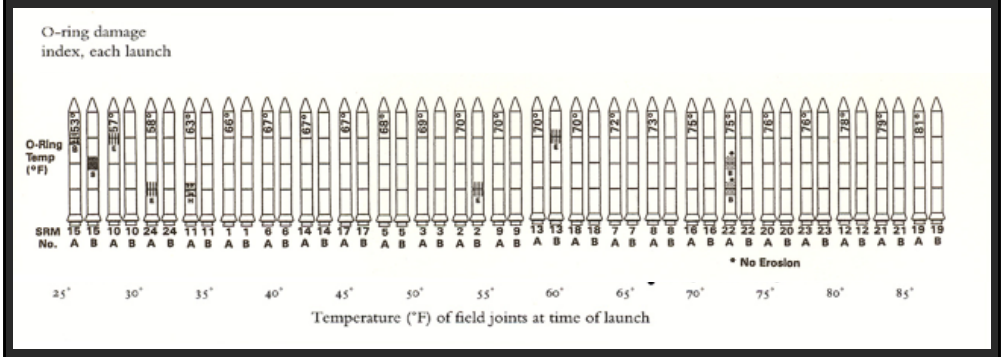
MOTOR	MBT	AMB	O-RING	WIND
DM-1	68	36	47	10 MPH
DM-2	76	45	52	10 MPH
QM-3	72.5	40	48	10 MPH
QM-4	76	48	51	10 MPH
SRM-15	52	64	53	10 MPH
SRM-22	77	78	75	10 MPH
SRM-25	55	26	29 27	10 MPH 25 MPH

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

Make a decision: Challenger

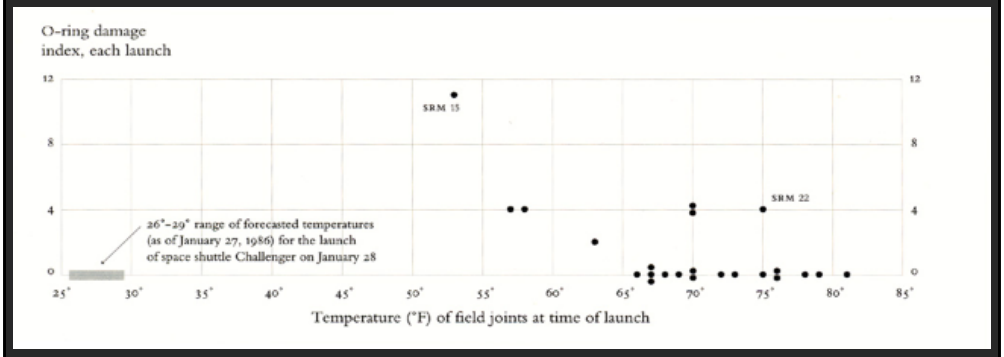


Make a decision: Challenger



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

Make a decision: Challenger

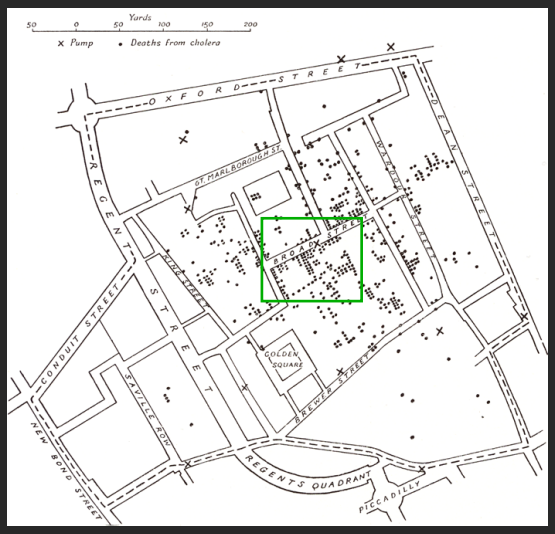


Visualizations drawn by Tufte show how low temperatures damage O-rings [Tufte 97]

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

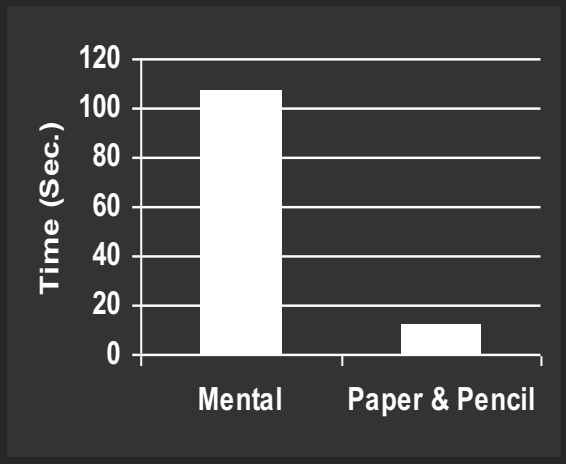
Class Exercise

Expand memory: Multiplication

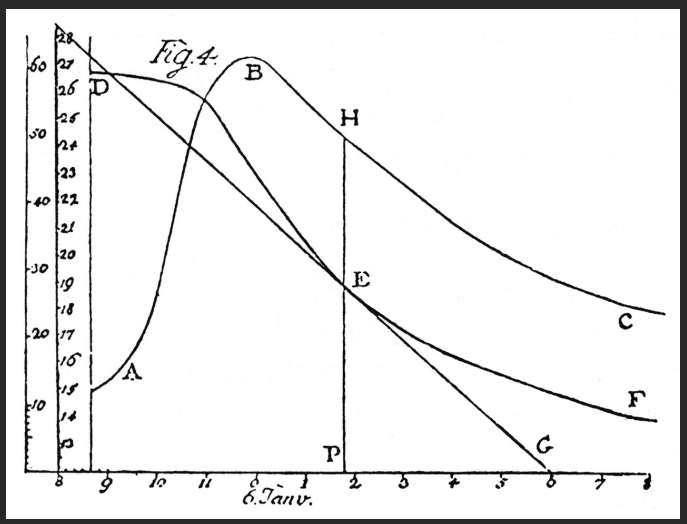
34
x 87

Expand memory: Multiplication

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

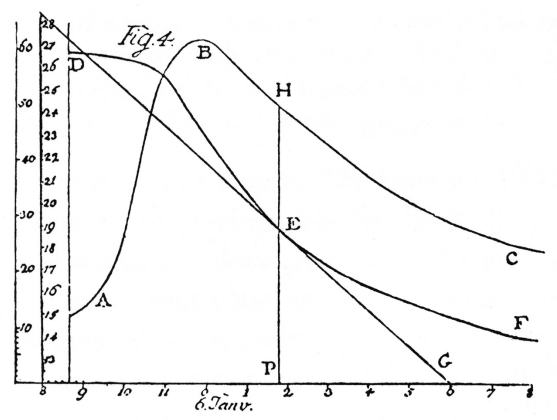


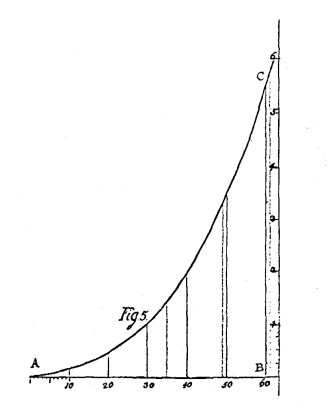
Graphical calculation: Evaporation



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]

Most powerful brain?

Microsoft Excel - animal.xls

File Edit View Insert Format Tools Data Window Help

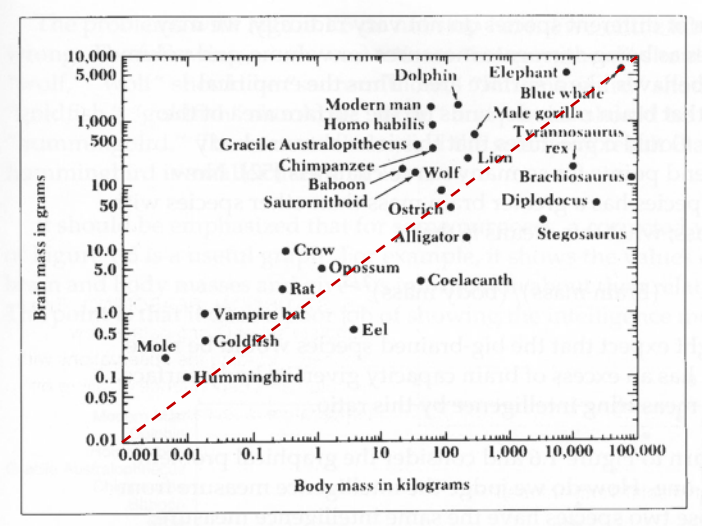
A1 ID

	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

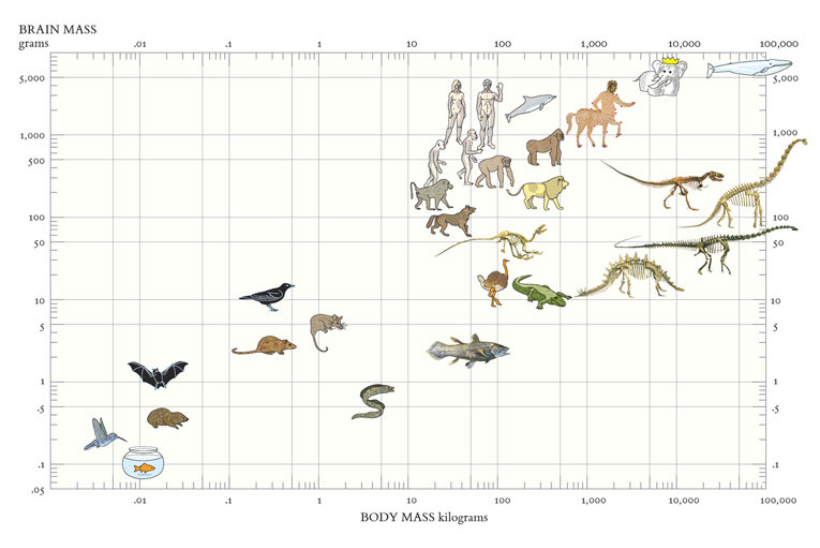
Most powerful brain?



The Dragons of Eden [Carl Sagan]

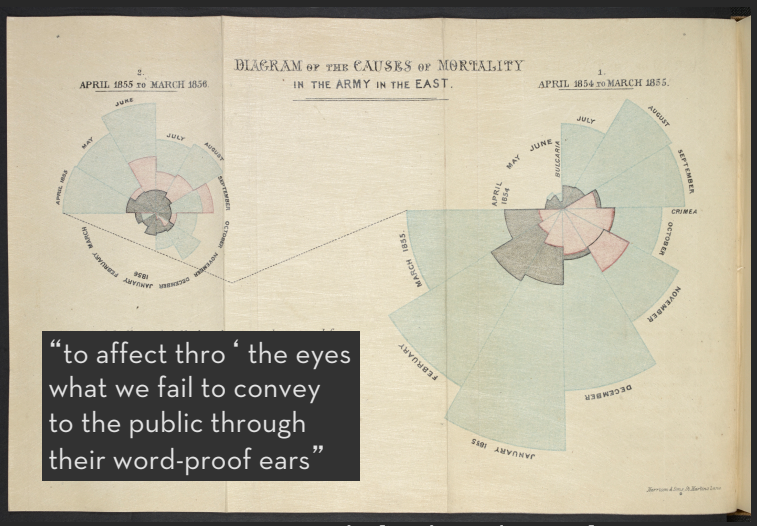
Convey Information to Others

Most powerful brain?



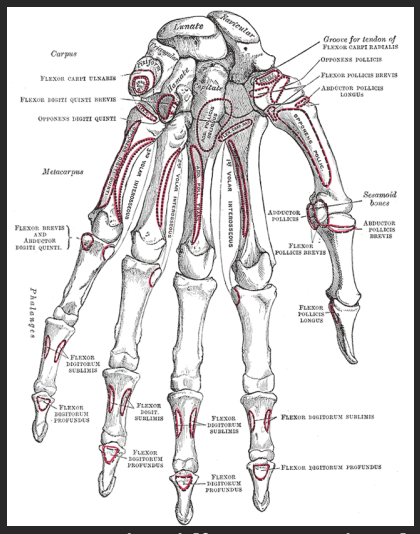
Beautiful Evidence [Tufte]

Present argument

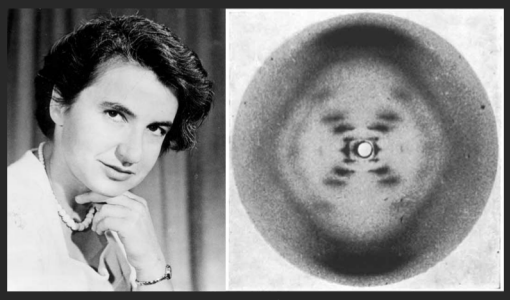


Crimean War Deaths [Nightingale 1858]

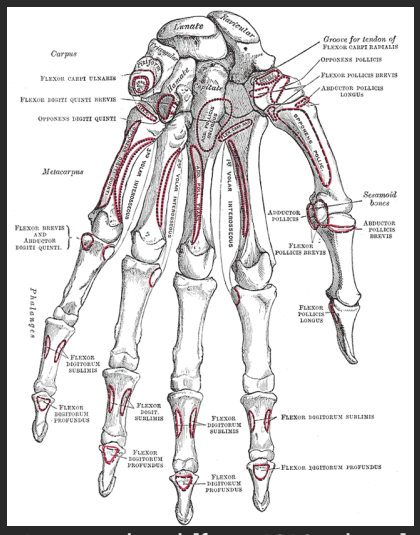
Inspire



Bones in hand [from 1918 edition]


X-ray crystallography of DNA [Franklin 52]

Inspire



Bones in hand [from 1918 edition]


Double helix model [Watson and Crick 53]

Visualization Research

Challenge

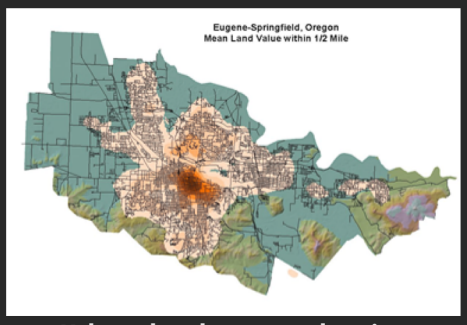
More and more unseen data

- Faster creation and collection

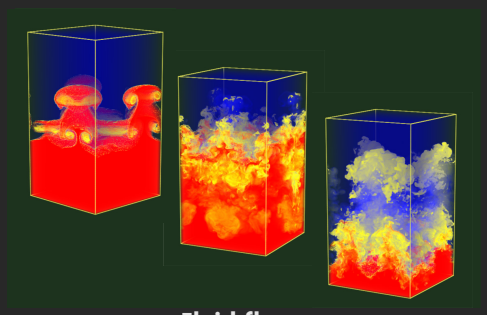
Challenge

More and more unseen data

- Faster creation and collection



Urban development planning
www.urbansim.org



Fluid flow
ctr.stanford.edu

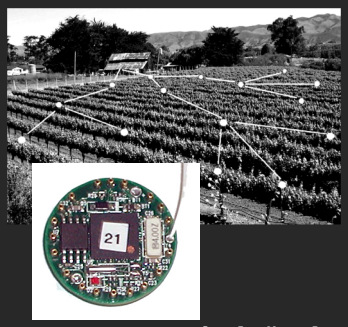
Simulation

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

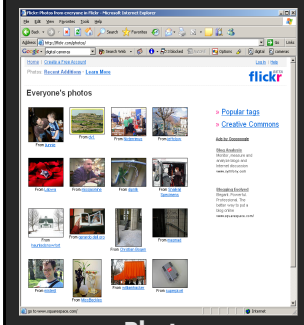
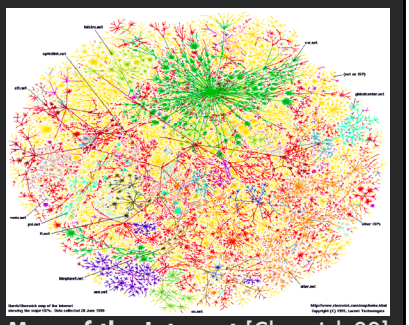


Photo sharing/annotation
flickr.com

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]
16100 exabytes in 2016 [IDC 17]

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.

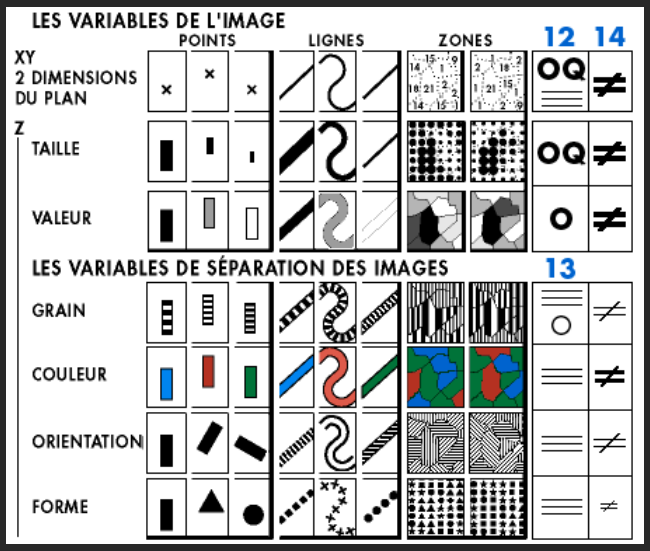
Hal Varian, Google's Chief Economist
The McKinsey Quarterly, Jan 2009

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

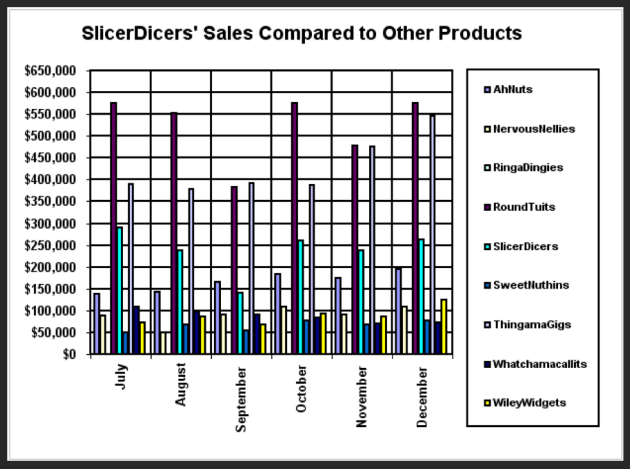
Topics

1. Data and image models

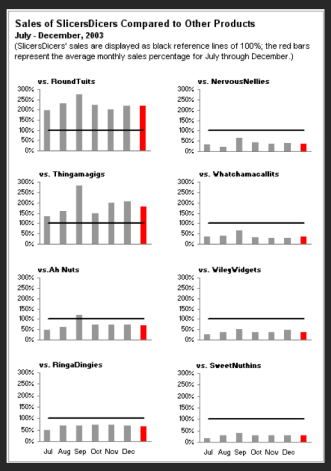


[Bertin, Graphics and Graphic Information Processing 1981]

2. Visualization Design

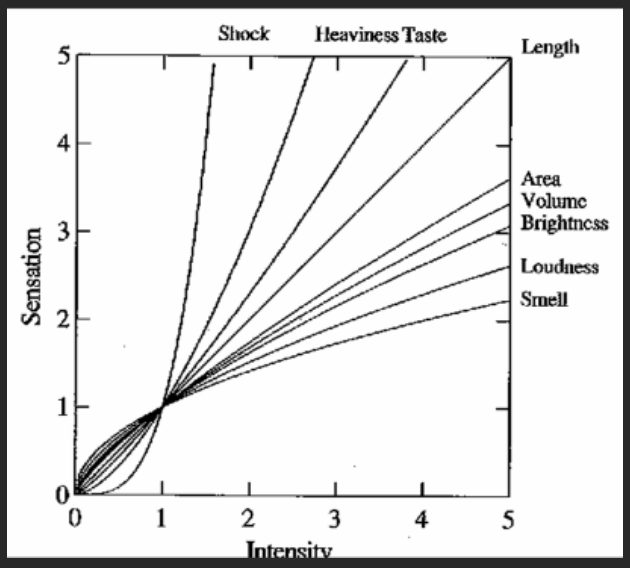


Problematic design



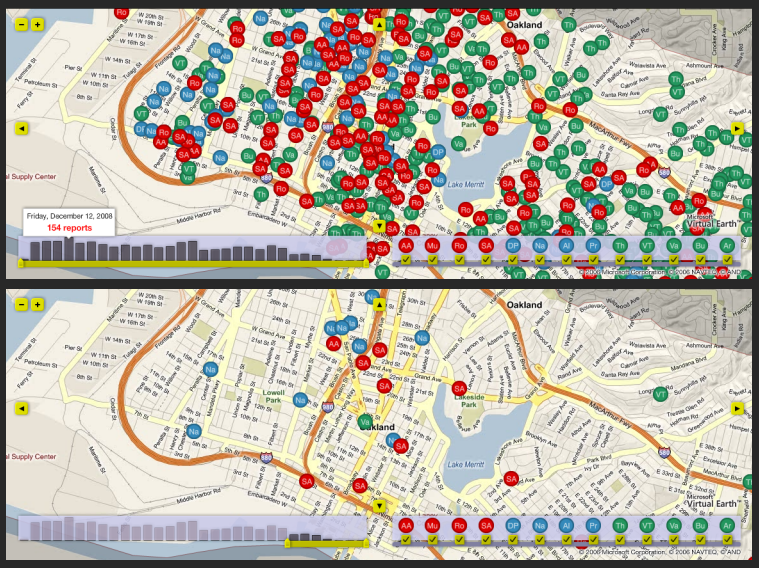
Redesign

3. Perception



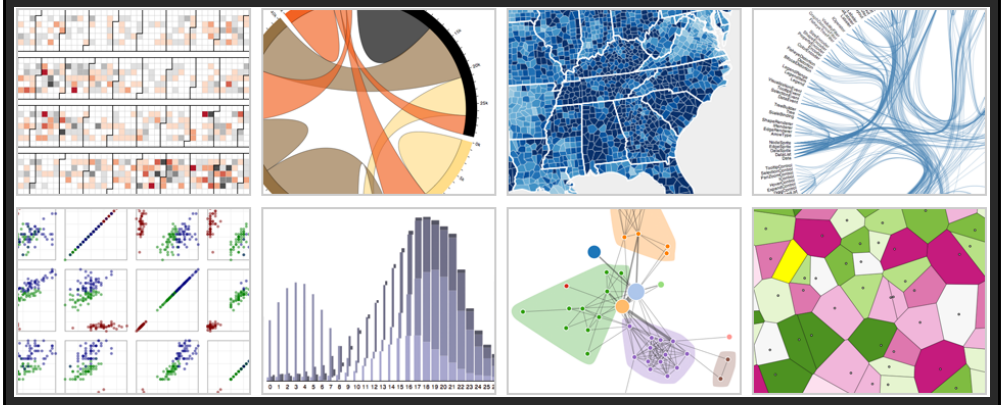
The psychophysics of sensory function [Stevens 61]

4. Interaction



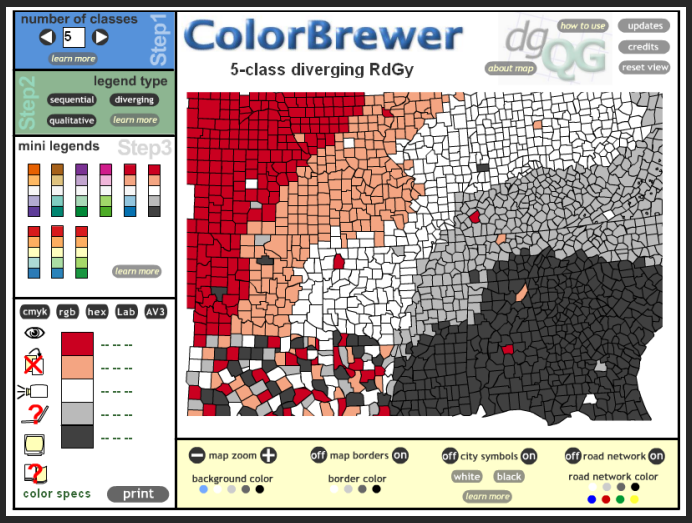
Oakland Crimespotting (crimespotting.org) [Stamen]

5. Interactive visualizations with D3



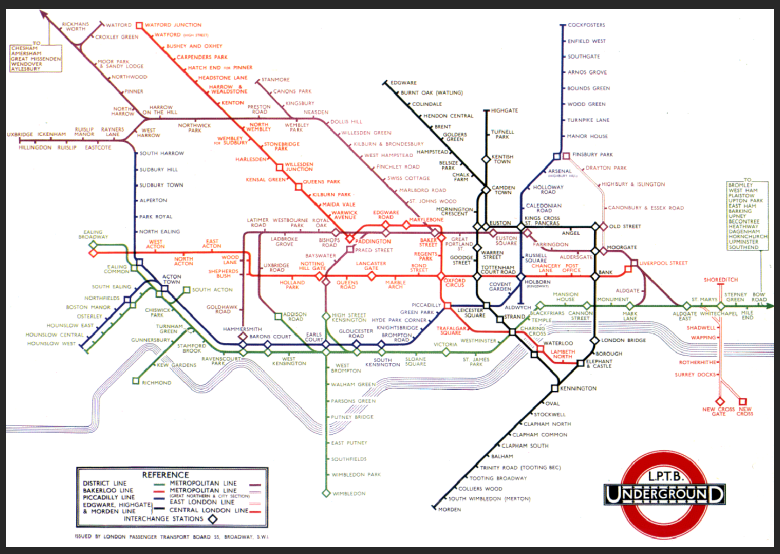
D3: Data Driven Documents [Bostock 2011]

6. Color



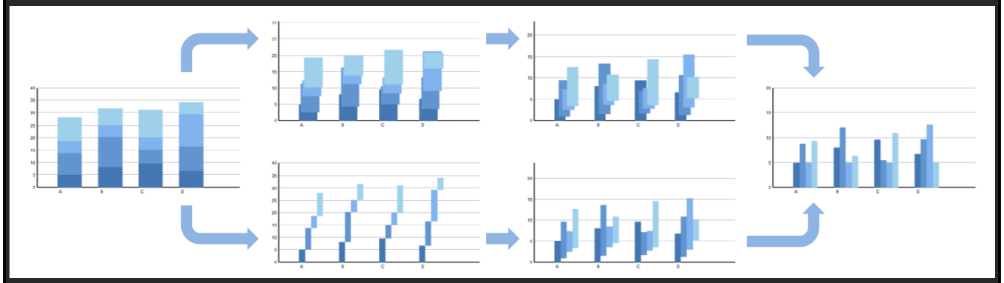
[from Cynthia Brewer http://www.personal.psu.edu/faculty/c/a/cab38/]

7. Spatial Layout



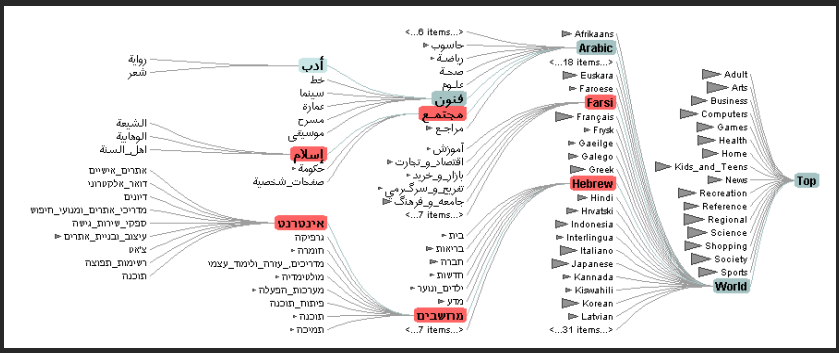
London underground [Beck 33]

8. Animation



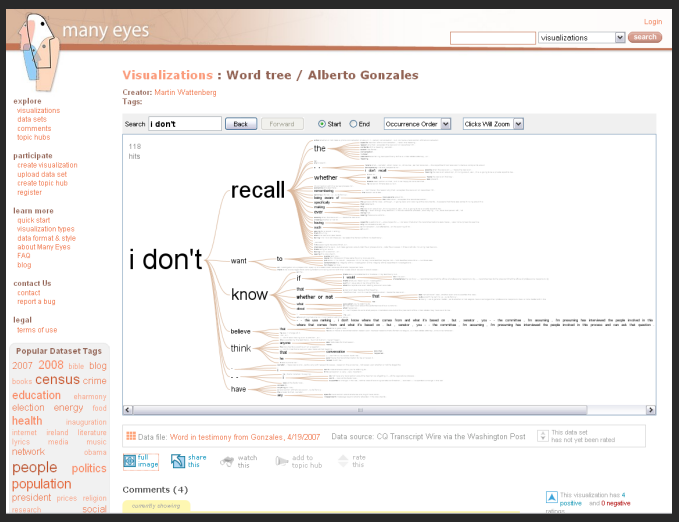
Animated Transitions [Heer 07]

9. Trees and graphs



Degree-of-Interest Trees [Heer 2004]

10. Text visualization



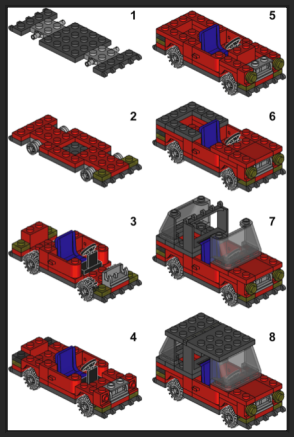
Word Trees [Wattenberg 2008]

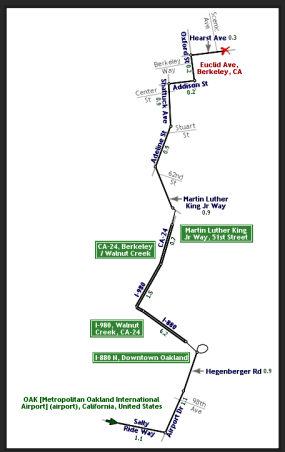
Course Goals

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

Vera Lin
Gracie Young

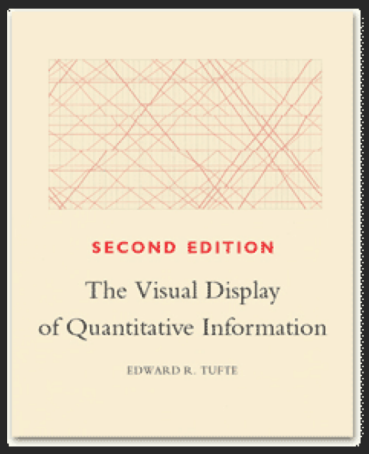
Piazza is the best way to interact with us
http://piazza.com/stanford/fall2018/cs448b

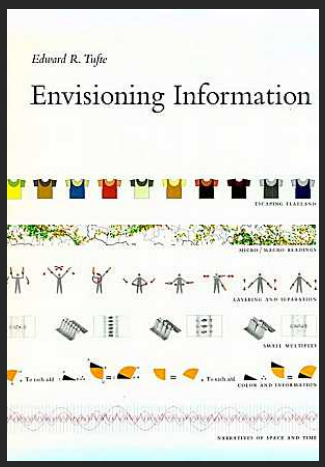
Office Hours

Maneesh: 10:00-11a Mon, Gates 364
Gracie: 9:30-10:30a Tue, Lathrop Tech Lounge
Vera: 4:30-5:30p Thu, Huang basement

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-fa18

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

Simpsons Episodes Data

The site data.world has collected a data set describing the first 600 episodes of the Simpsons. For each episode the data set contains the following information.

Number of records: 600
Variable Names:

- **id:** Episode number
- **image_url:** Link to image for the episode
- **imdb_rating:** Rating from IMDB
- **imdb_votes:** Votes from IMDB
- **number_in_season:** Number of episodes in season
- **number_in_series:** Episode number
- **original_air_date:** Date of first airing
- **original_air_year:** Year of first airing
- **original_air_year:** Year of first airing
- **production_code:**
- **season:** Season number
- **title:** Episode title
- **us_viewers_in_millions:** Number of viewers
- **video_url:** Link to episode online
- **views:** Number of views for online episode

We've cleaned up this dataset and posted in csv format: simpsons_episodes.csv

Simpsons Episodes

Due by noon on Mon Oct 1

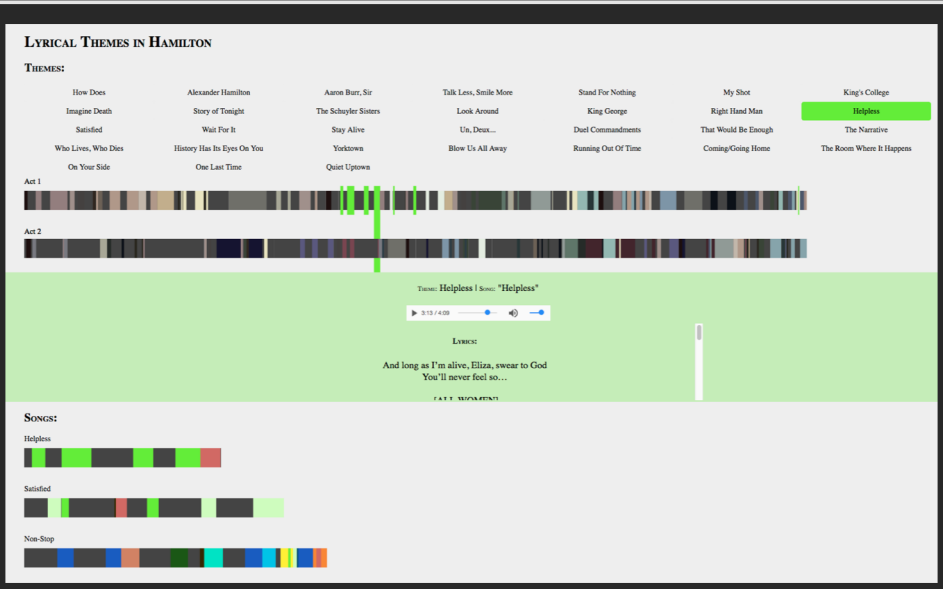
Final project

- **Visualization project on topic of your choice**
- **Last 4 weeks of course**
- **Project write-up (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

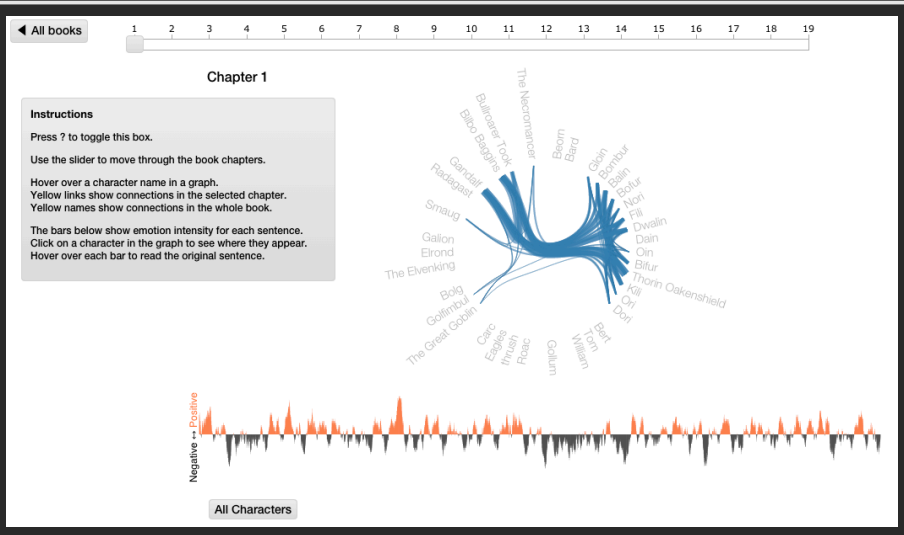
- IEEE Visualization
- IEEE Information Visualization
- SIGGRAPH

Structure of Musicals



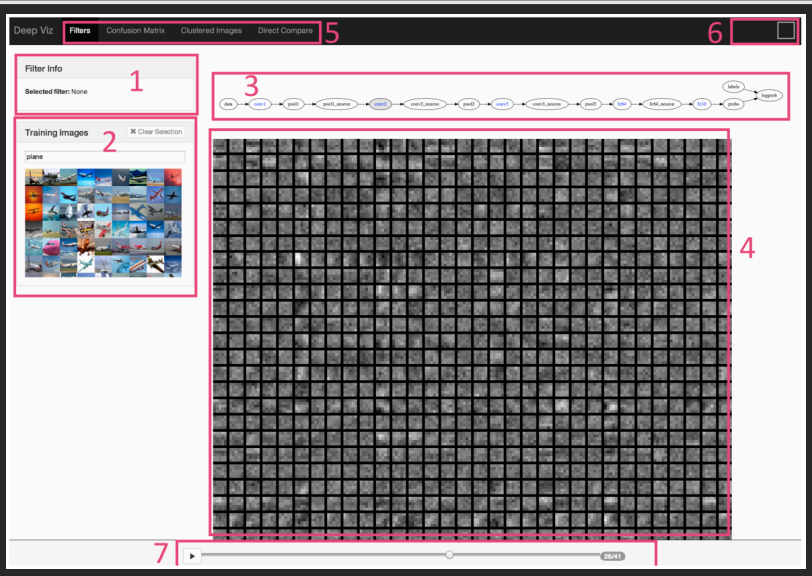
Lyrical themes in Hamilton [Townley-Smith, Sterman, Cook 2016]

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]