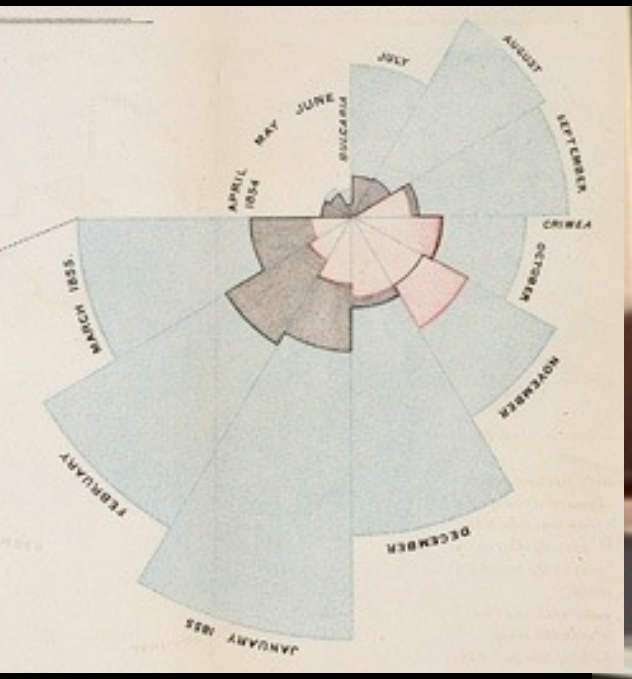
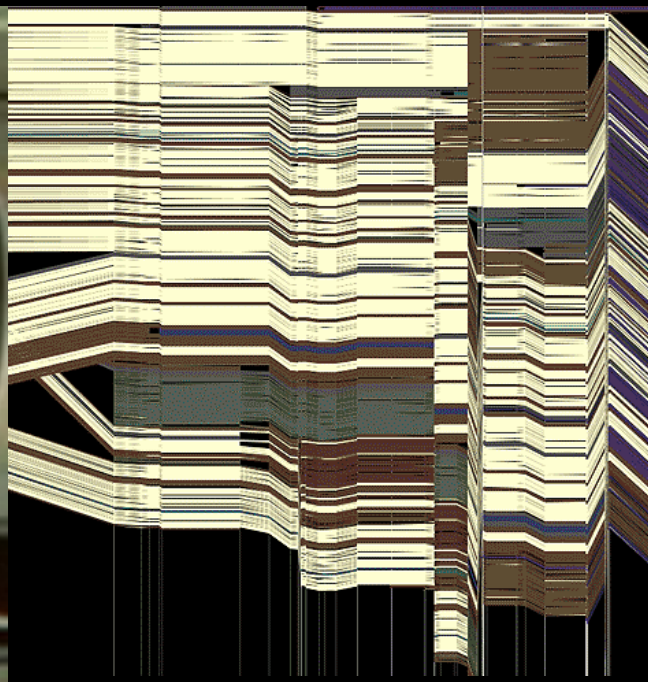
CSE 512 - Data Visualization

The Value of Visualization





Jeffrey Heer University of Washington

How much data (bytes) did we produce in 2010?

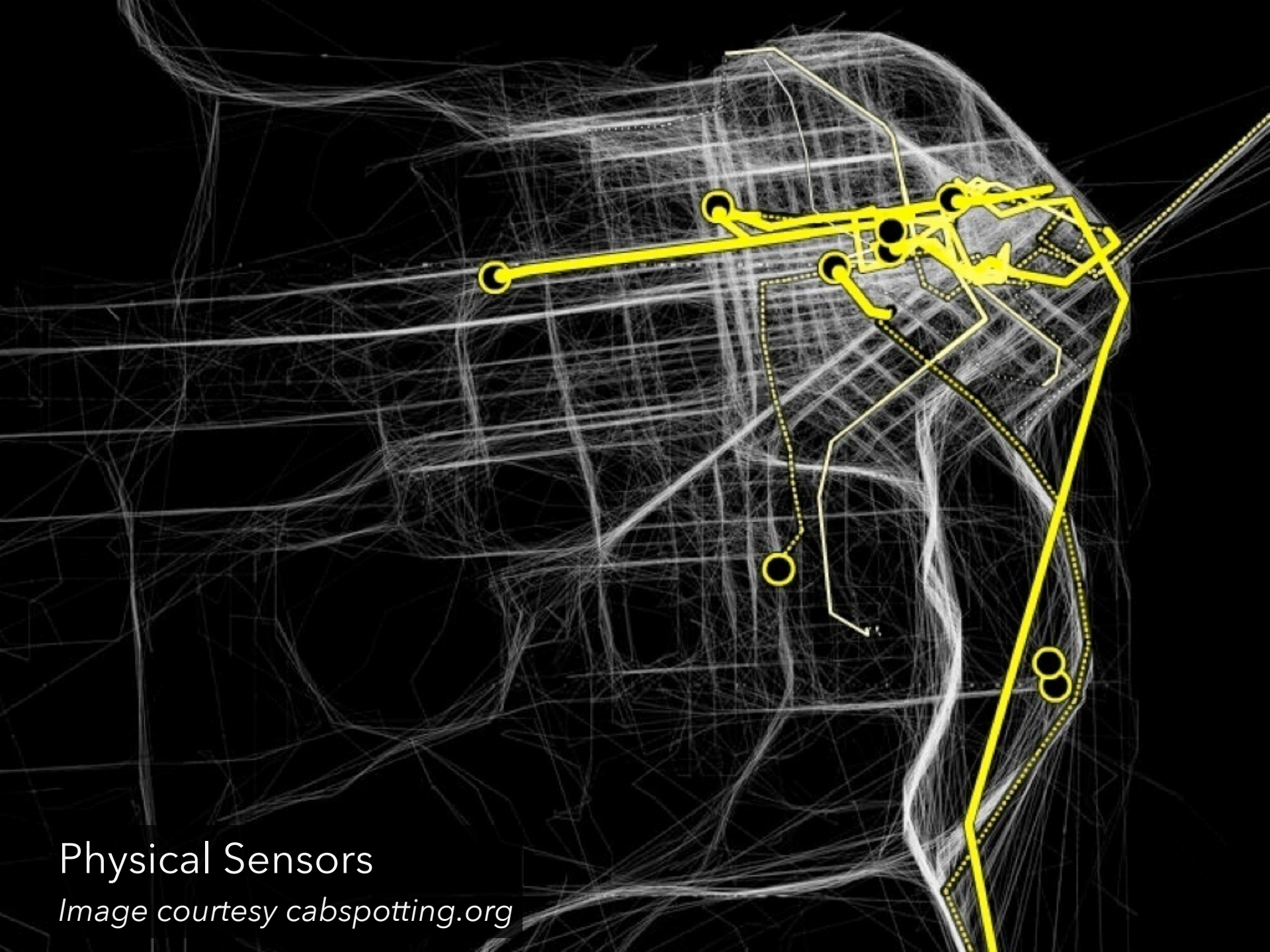
2010: 1,200 exabytes

Gantz et al, 2008, 2010

2010: 1,200 exabytes

10x increase over 5 years

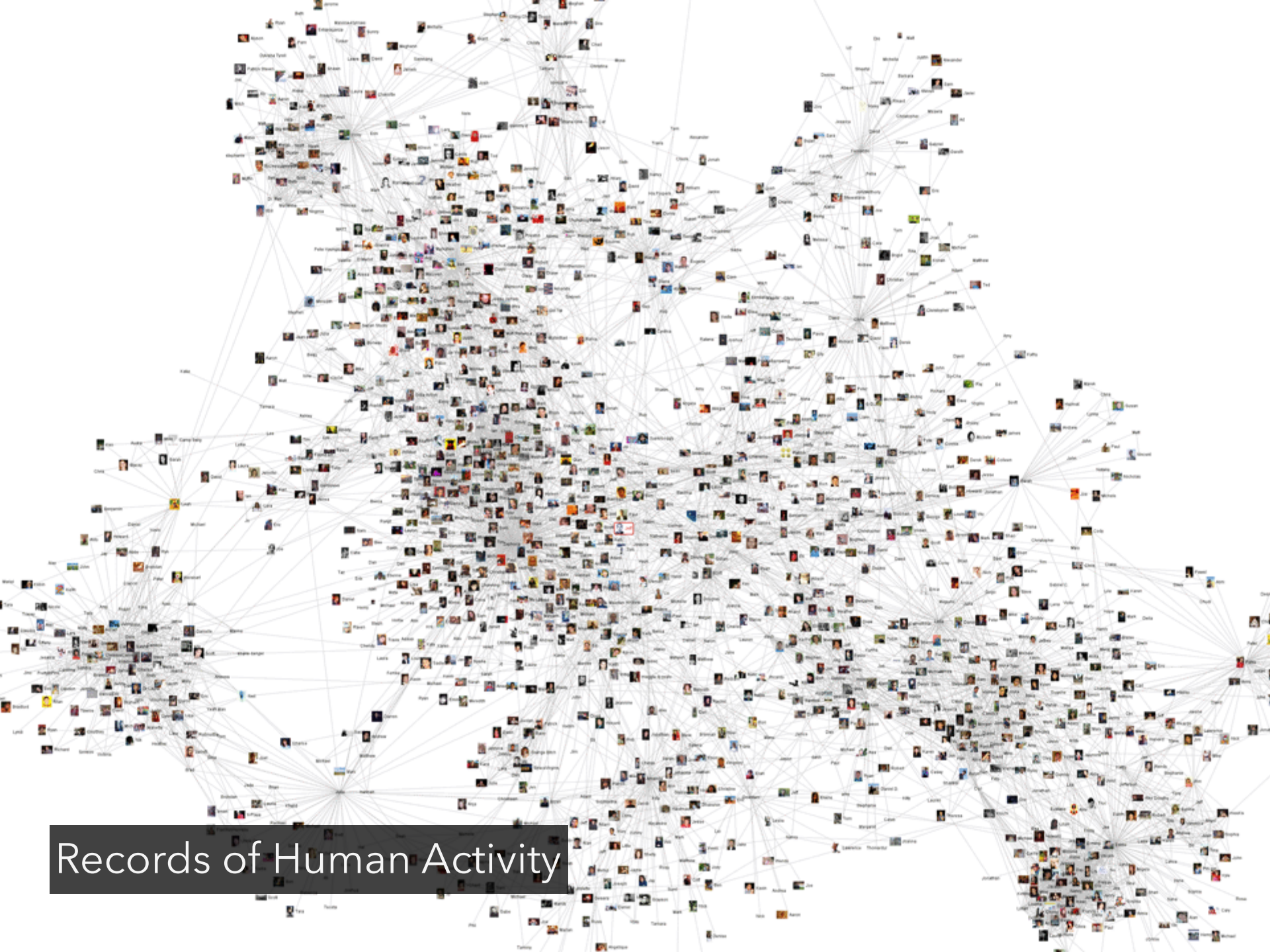
Gantz et al, 2008, 2010

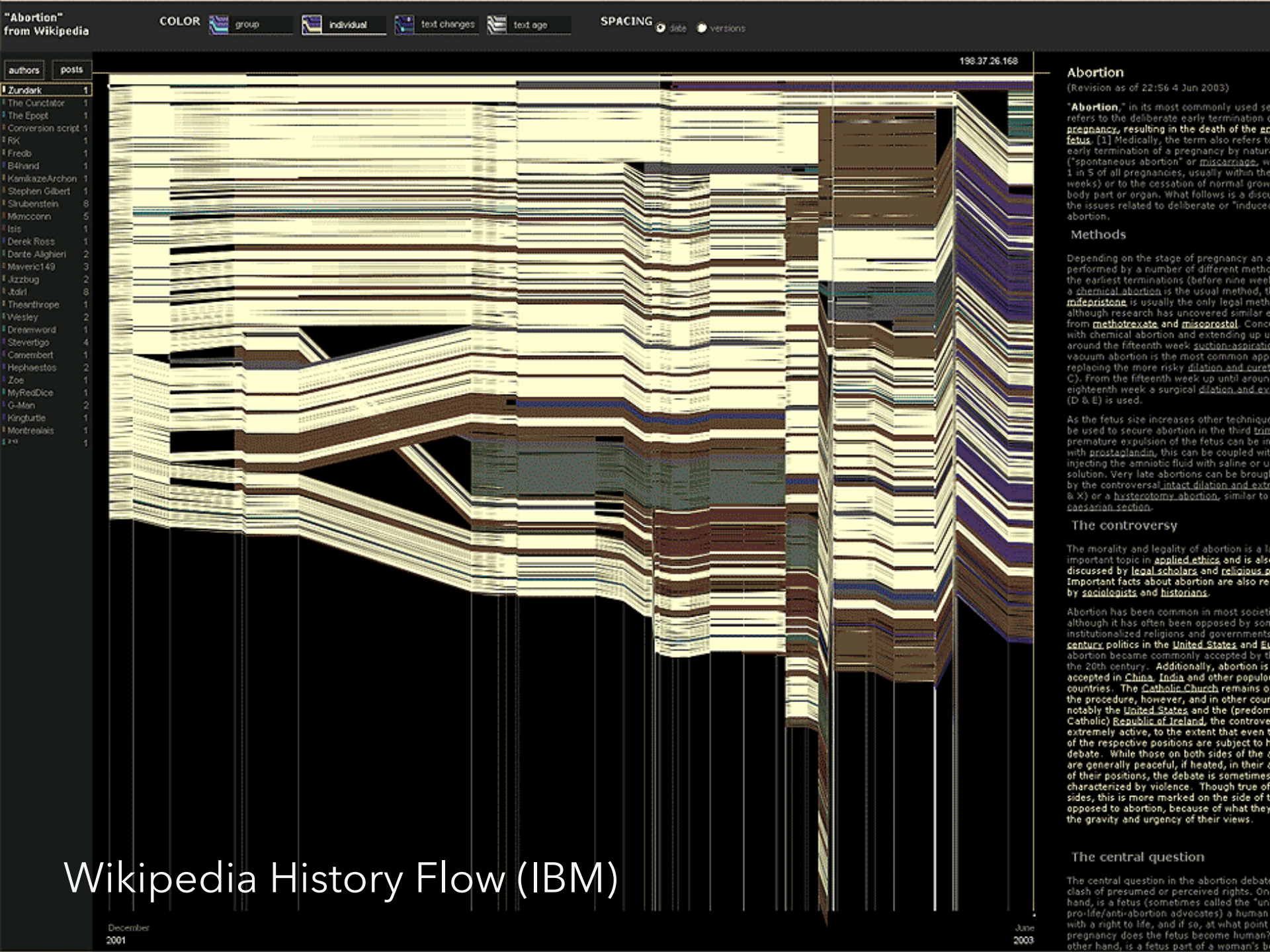


Physical Sensors

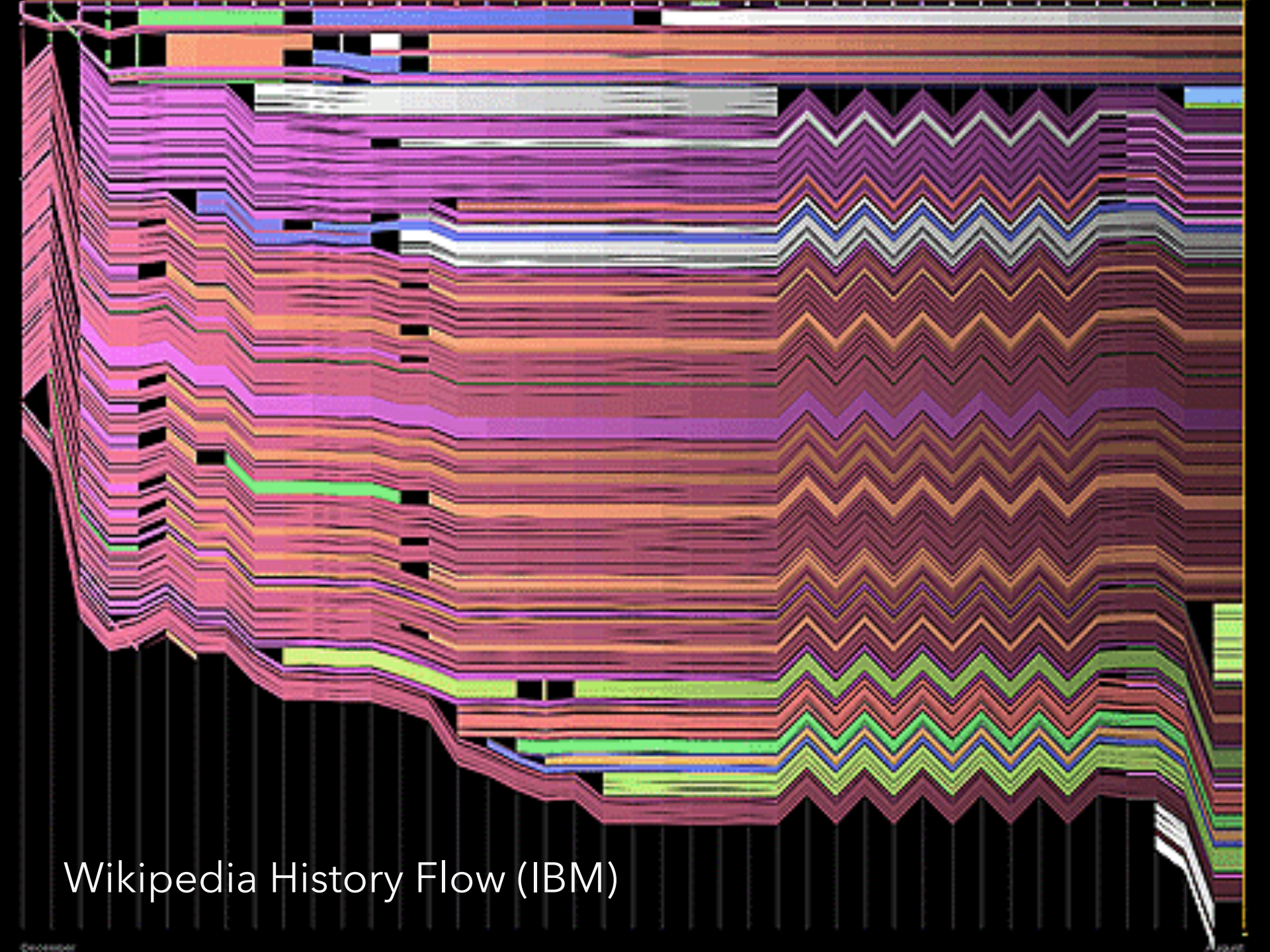
Image courtesy cabspotting.org

Health & Medicine


Records of Human Activity



Wikipedia History Flow (IBM)


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

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

X	Y
10	8.04
8	6.95
13	7.58
9	8.81
11	8.33
14	9.96
6	7.24
4	4.26
12	10.84
7	4.82
5	5.68

Set B

X	Y
10	9.14
8	8.14
13	8.74
9	8.77
11	9.26
14	8.1
6	6.13
4	3.1
12	9.11
7	7.26
5	4.74

Set C

X	Y
10	7.46
8	6.77
13	12.74
9	7.11
11	7.81
14	8.84
6	6.08
4	5.39
12	8.15
7	6.42
5	5.73

Set D

X	Y
8	6.58
8	5.76
8	7.71
8	8.84
8	8.47
8	7.04
8	5.25
19	12.5
8	5.56
8	7.91
8	6.89

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

$u_X = 9.0$ $\sigma_X = 3.317$

$u_Y = 7.5$ $\sigma_Y = 2.03$

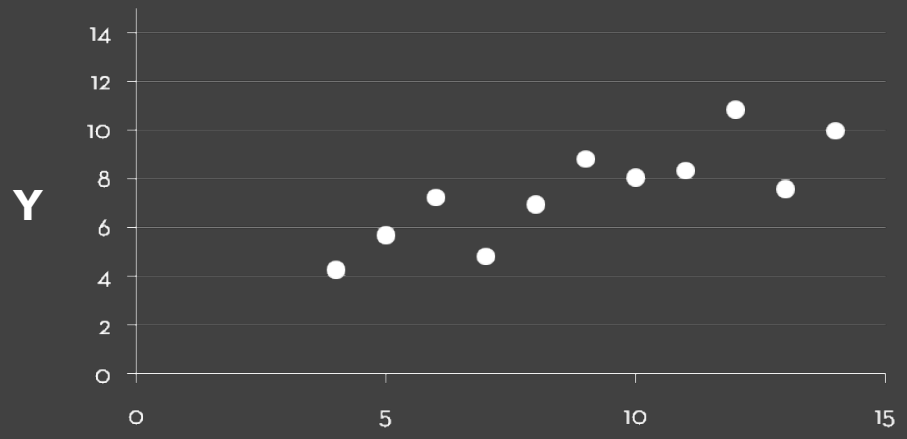
Linear Regression

$Y = 3 + 0.5\ X$

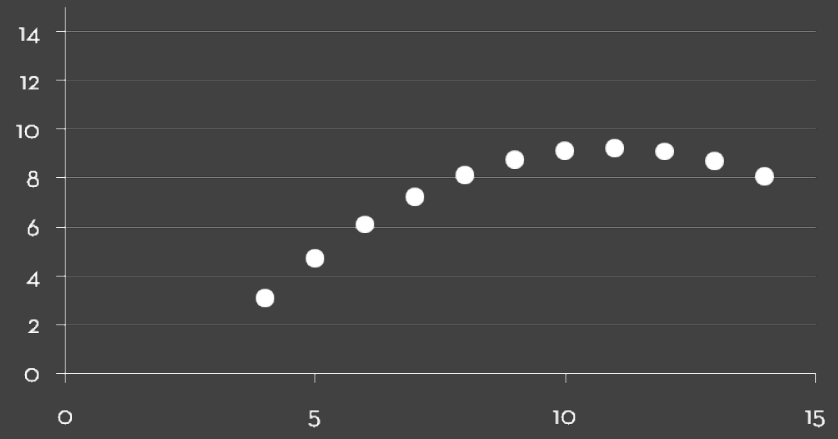
$R^2 = 0.67$

[Anscombe 1973]

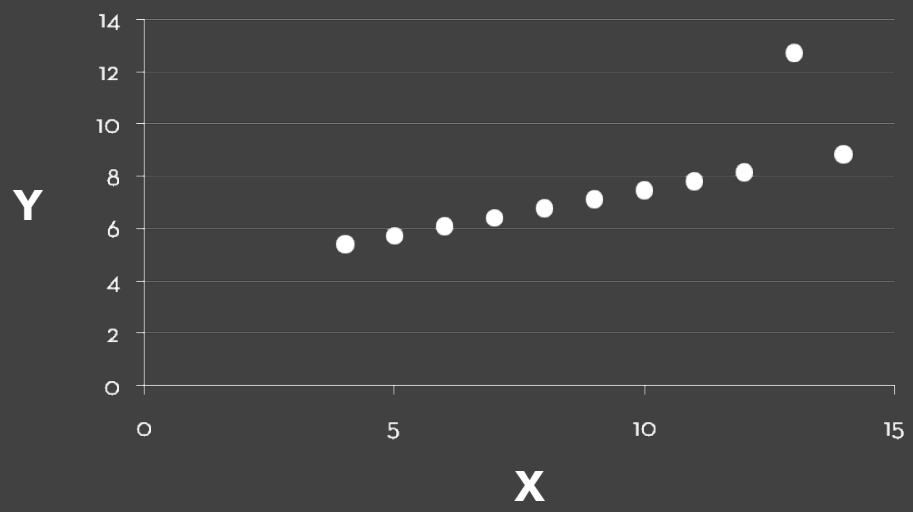
Set A



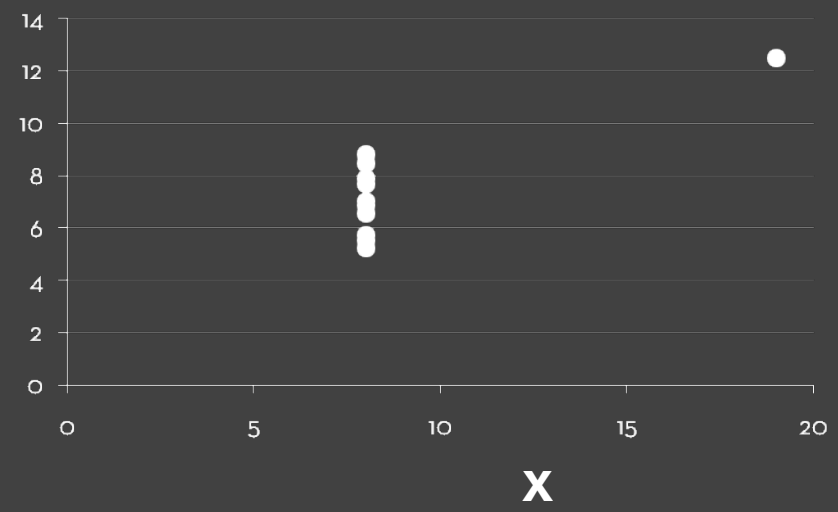
Set B



Set C



Set D

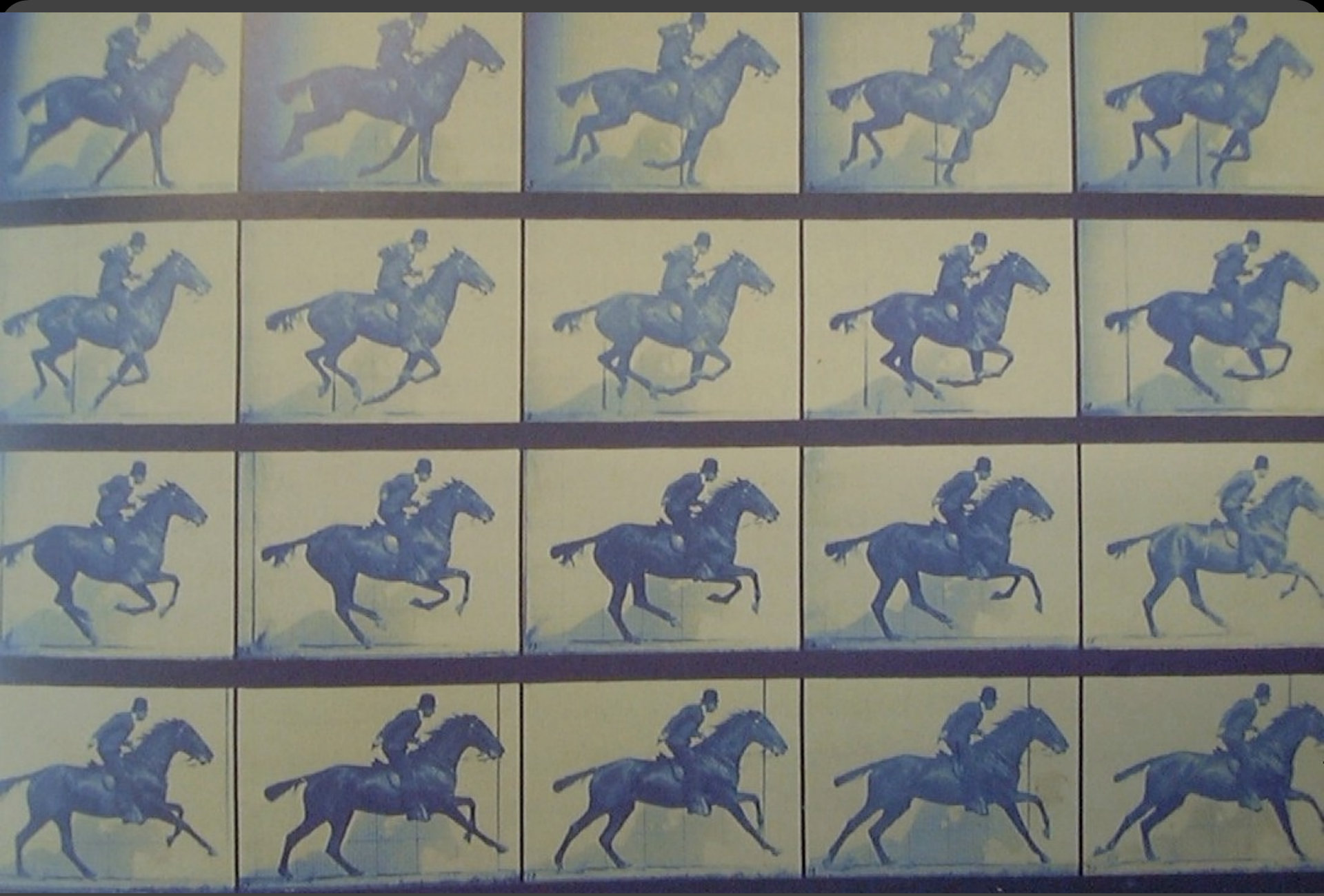


Why Create Visualizations?

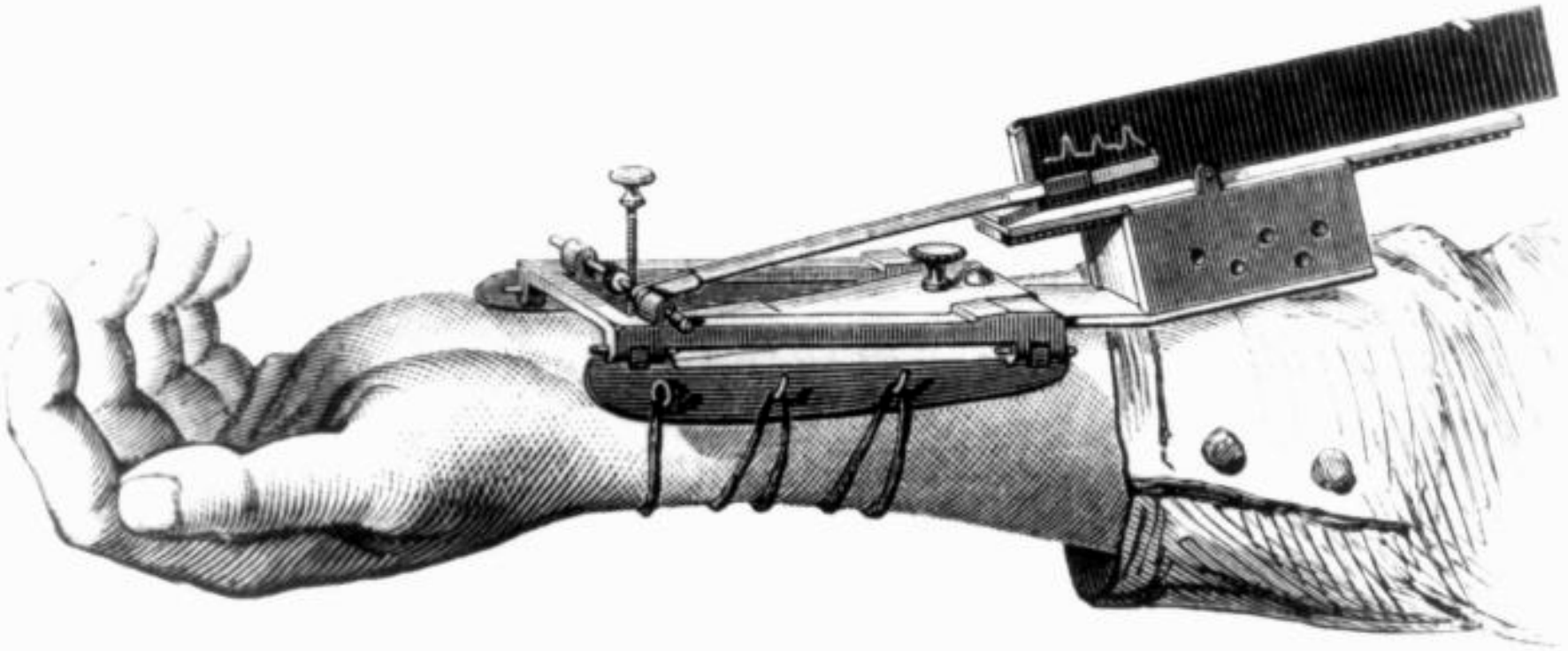
Why Create Visualizations?

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

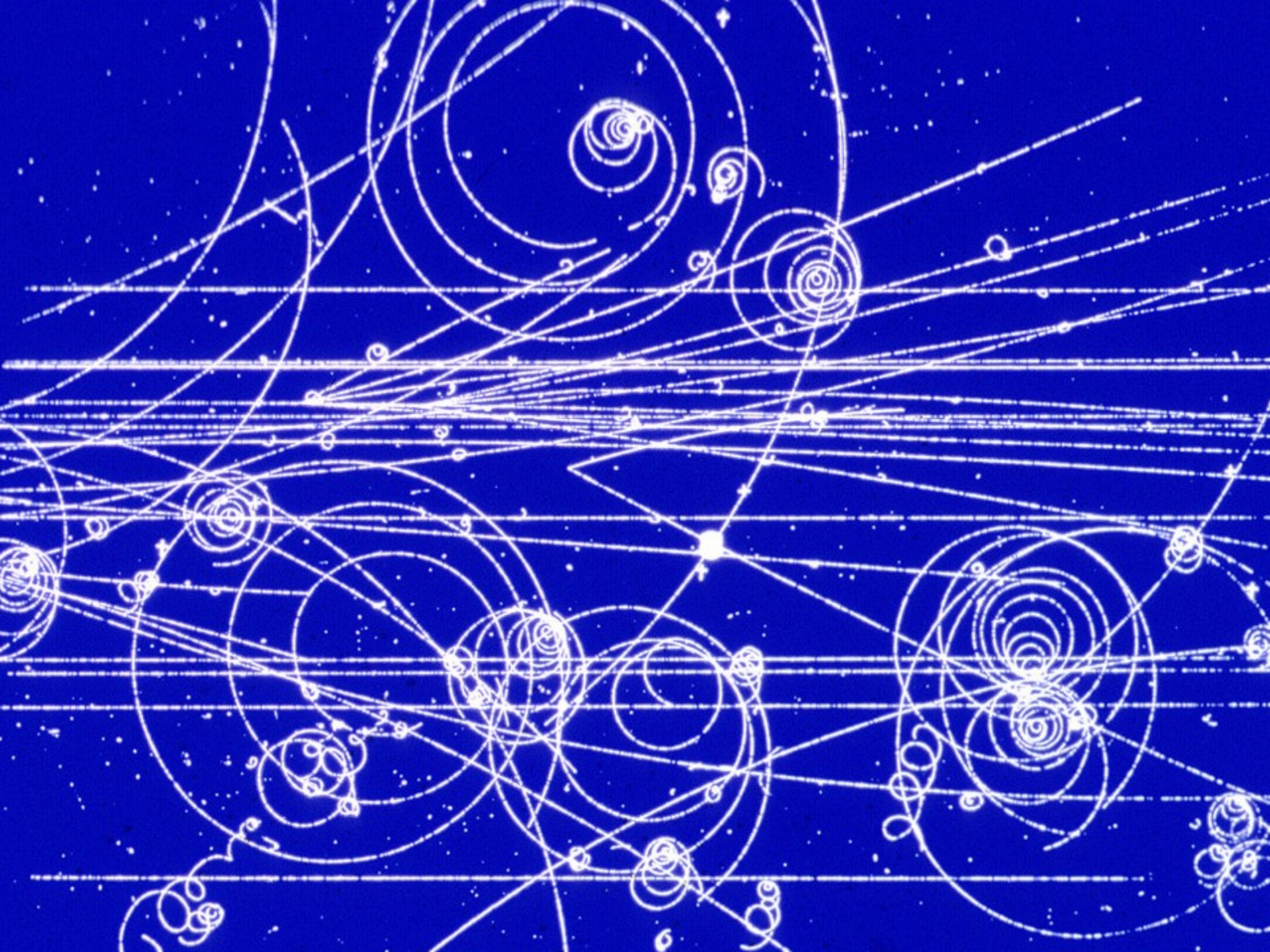


Gallop, Bay Horse “Daisy” [Muybridge 1884-86]

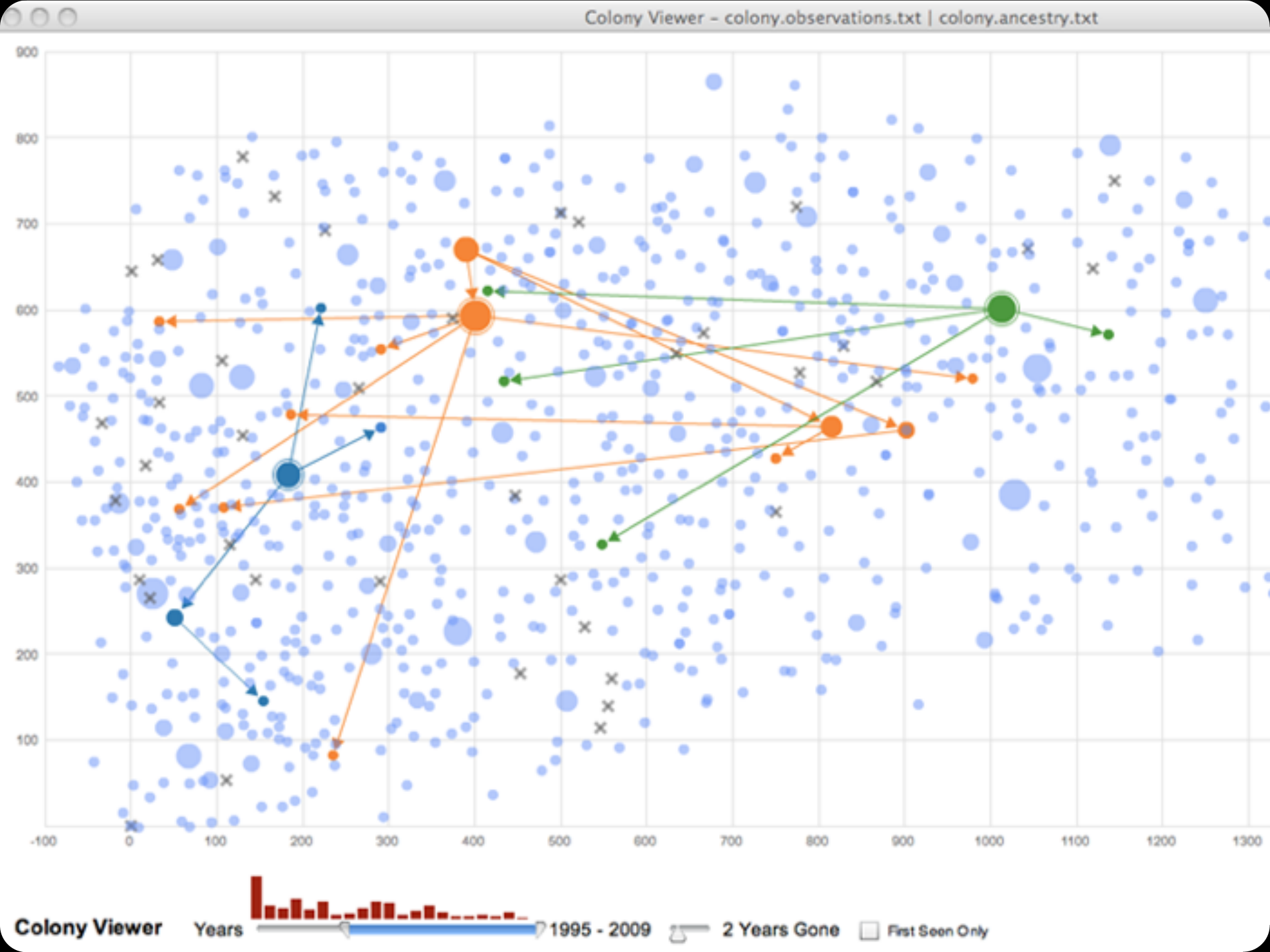


1.
Marey's sphygmograph in use, 1860. *La méthode graphique dans les sciences expérimentales et principalement en physiologie et en médecine.*

E.J. Marey's sphygmograph [from Braun 83]




Expected live in 06


Colony Viewer – colony.observations.txt | colony.ancestry.txt
900
800
700
600
500
400
300
200
100
-100
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
Colony Viewer
Years
1995 - 2009
2 Years Gone
First Seen Only

Support Reasoning

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 (prim)***	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 (80°), (110°) arc
- much worse visually than SRM-22

SRM 22 Blow-By
- 2 case joints (30-40°)

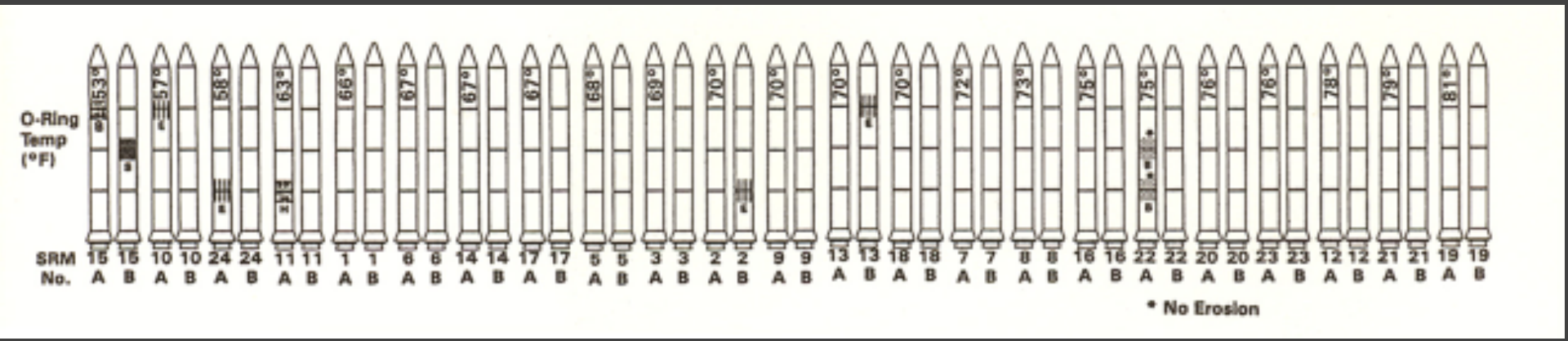
SRM-13A, 15, 16A, 18, 23A 24A
- nozzle Blow-By

History of O-ring Temperatures (degrees - F)

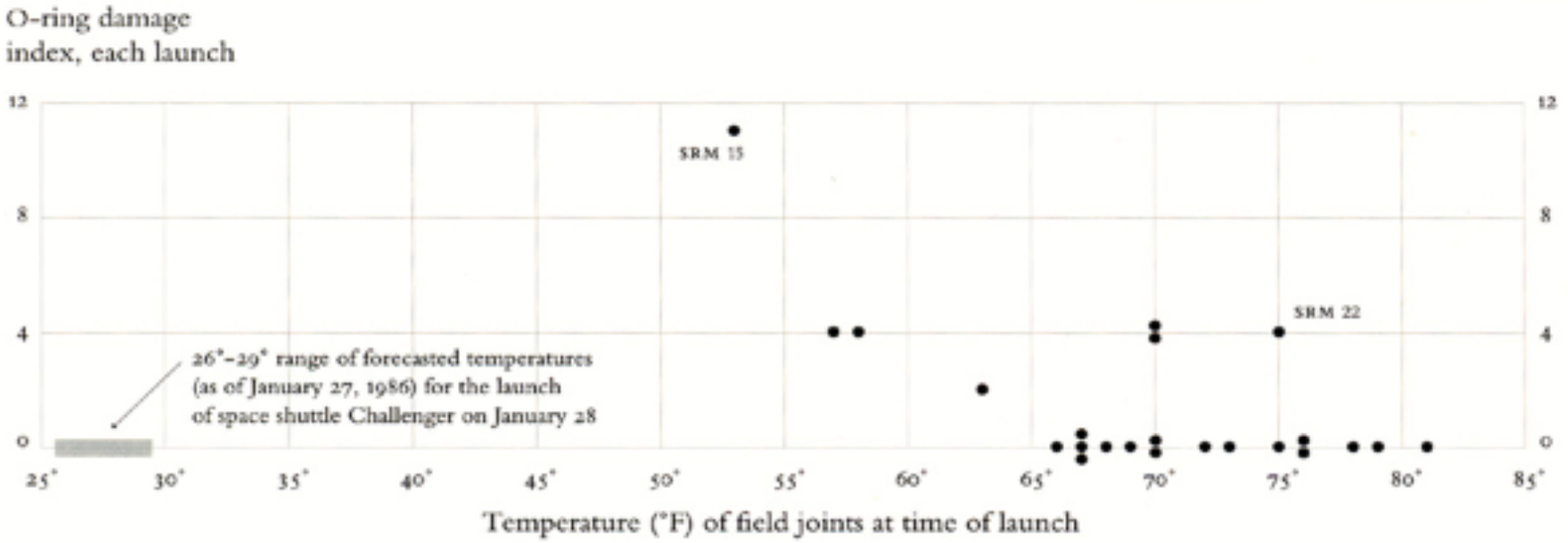
Motor	MBT	Amb	O-ring	Wind
DM-1	68	36	47	10 mph
DM-2	76	45	52	10 mph
QM-3	72.5	40	48	10 mph
QM-4	76	48	51	10 mph
SRM-15	52	64	53	10 mph
SRM-22	77	78	75	10 mph
SRM-25	55	26	29 27	10 mph 25 mph

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]

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

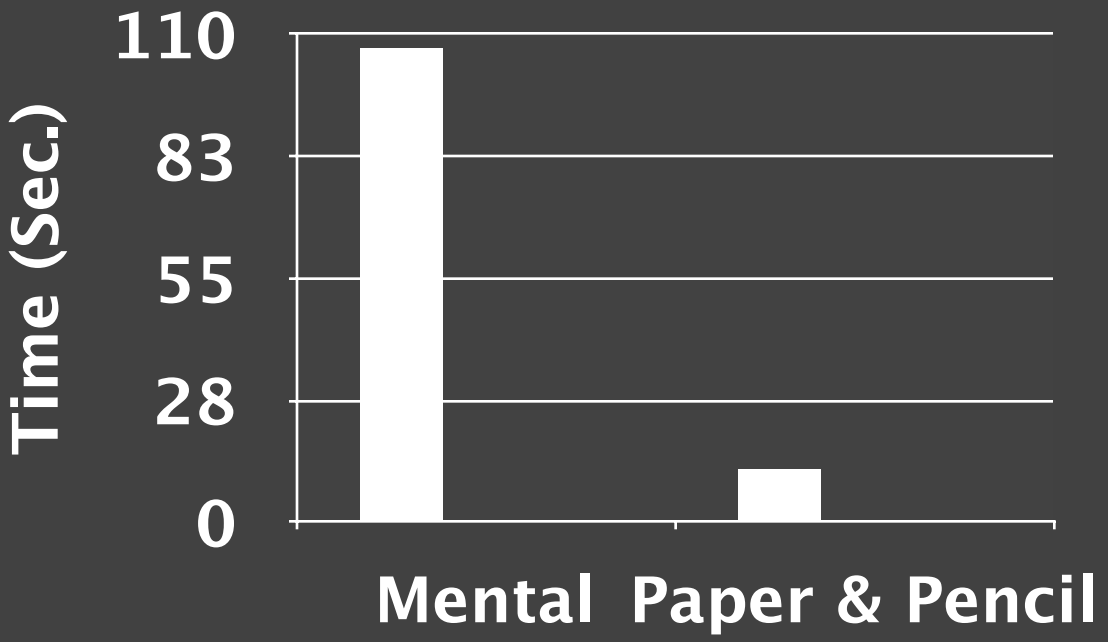
Class Exercise

Expand Memory: Multiplication

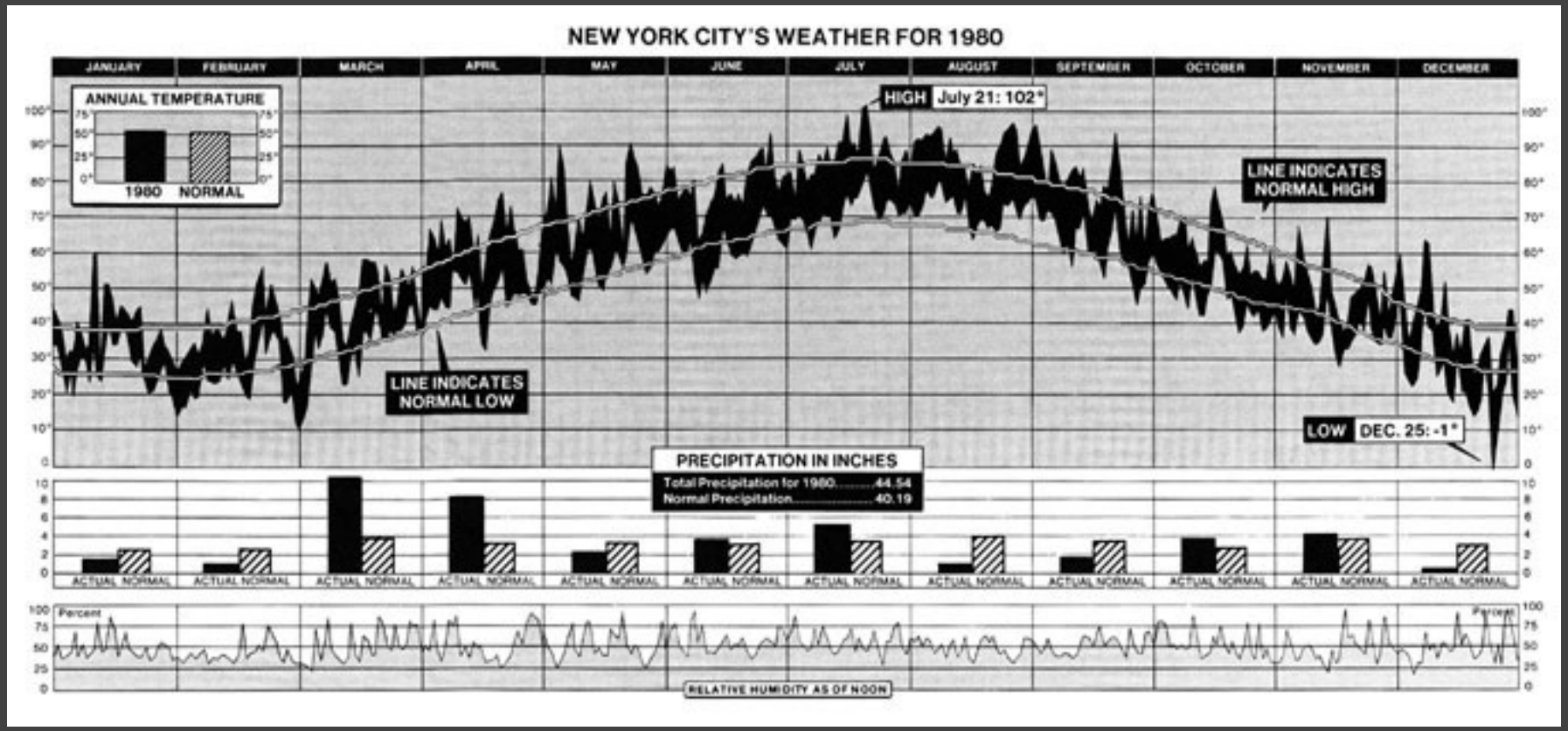
```
   34
x  72
```

Expand Memory: Multiplication

```
   34
x  72
------
   68
 2380
------
 2448
```

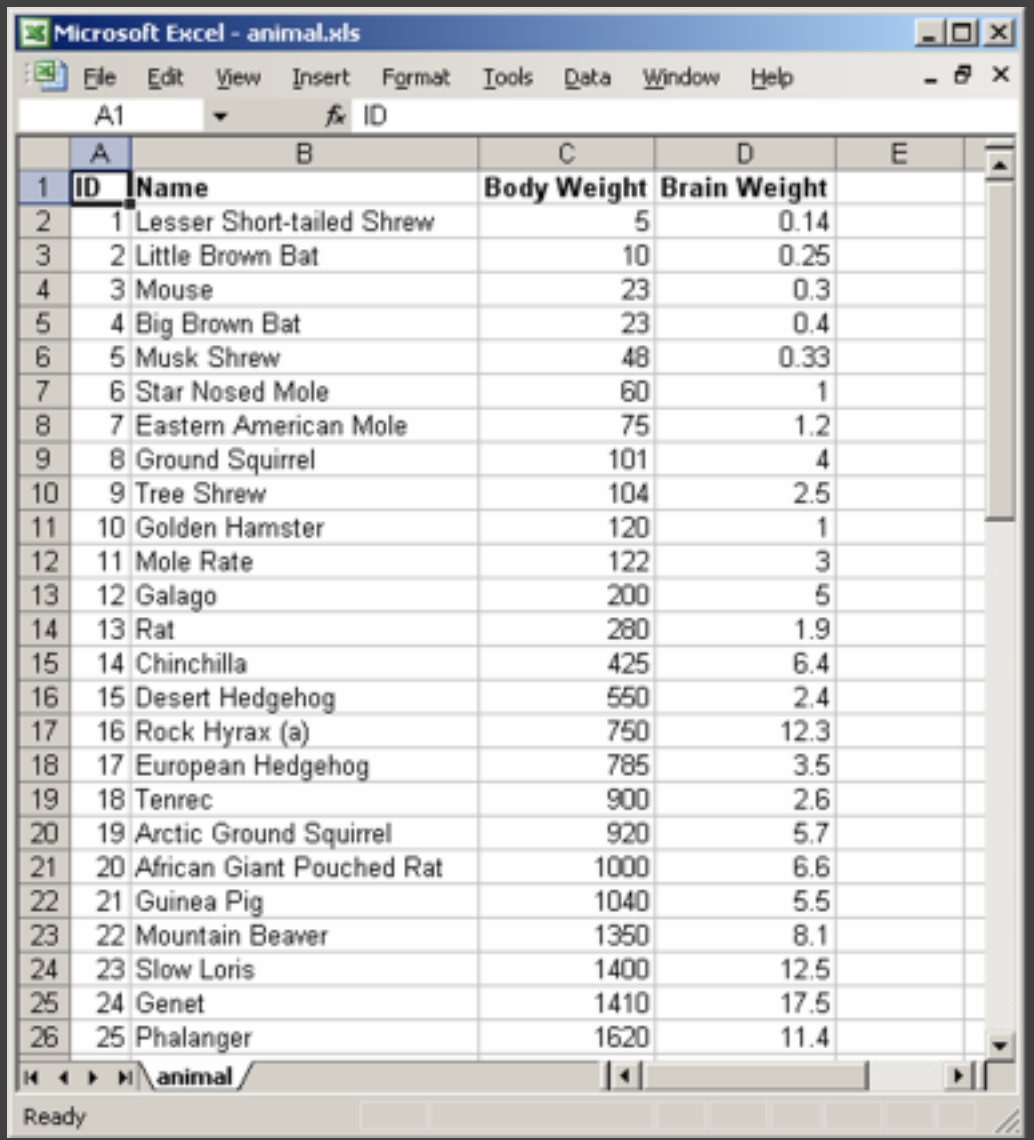


Find Patterns: NYC Weather



[New York Times 1981]

The Most Powerful Brain?


Microsoft Excel - animal.xls

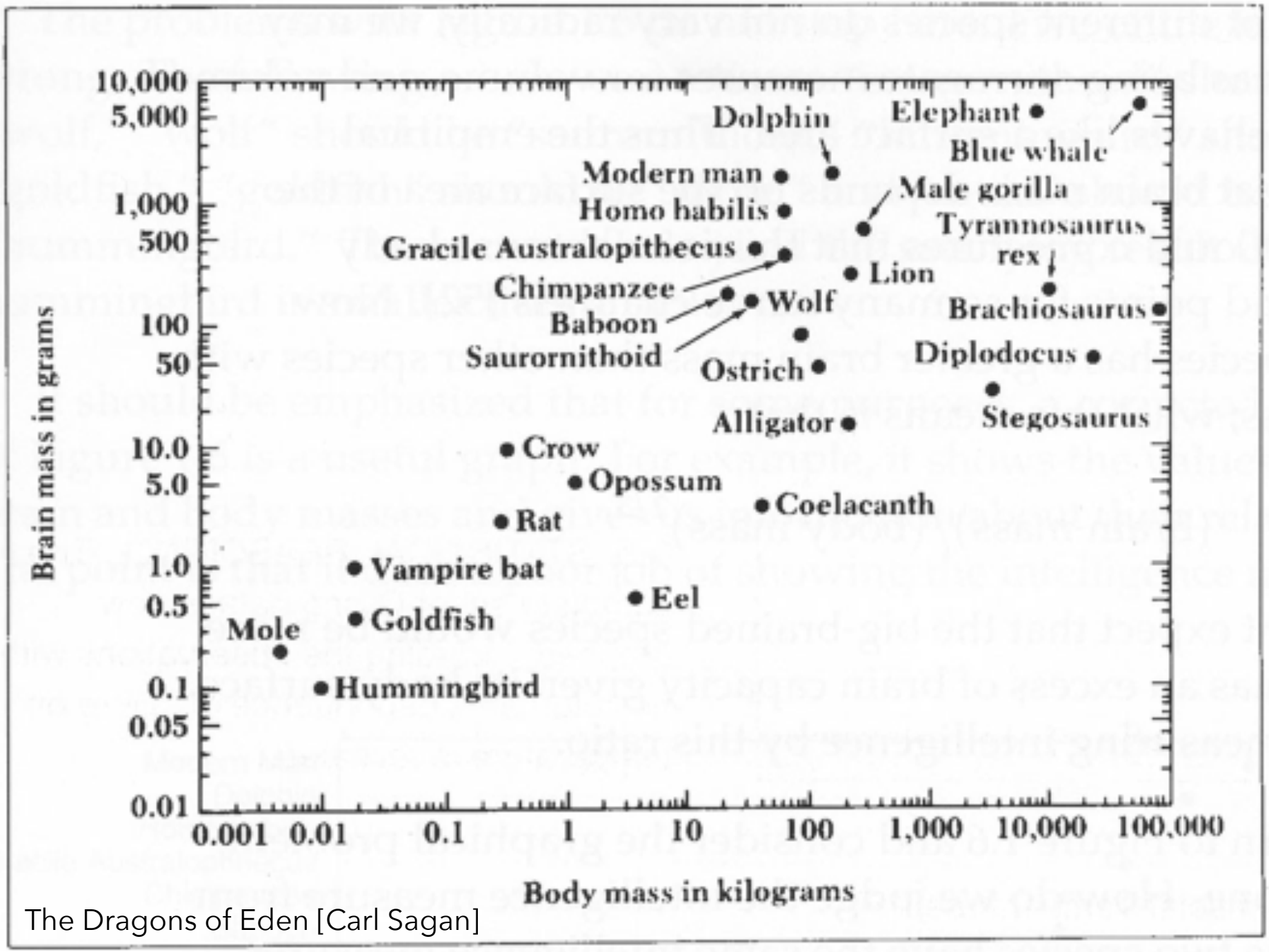
File Edit View Insert Format Tools Data Window Help

A1 ID

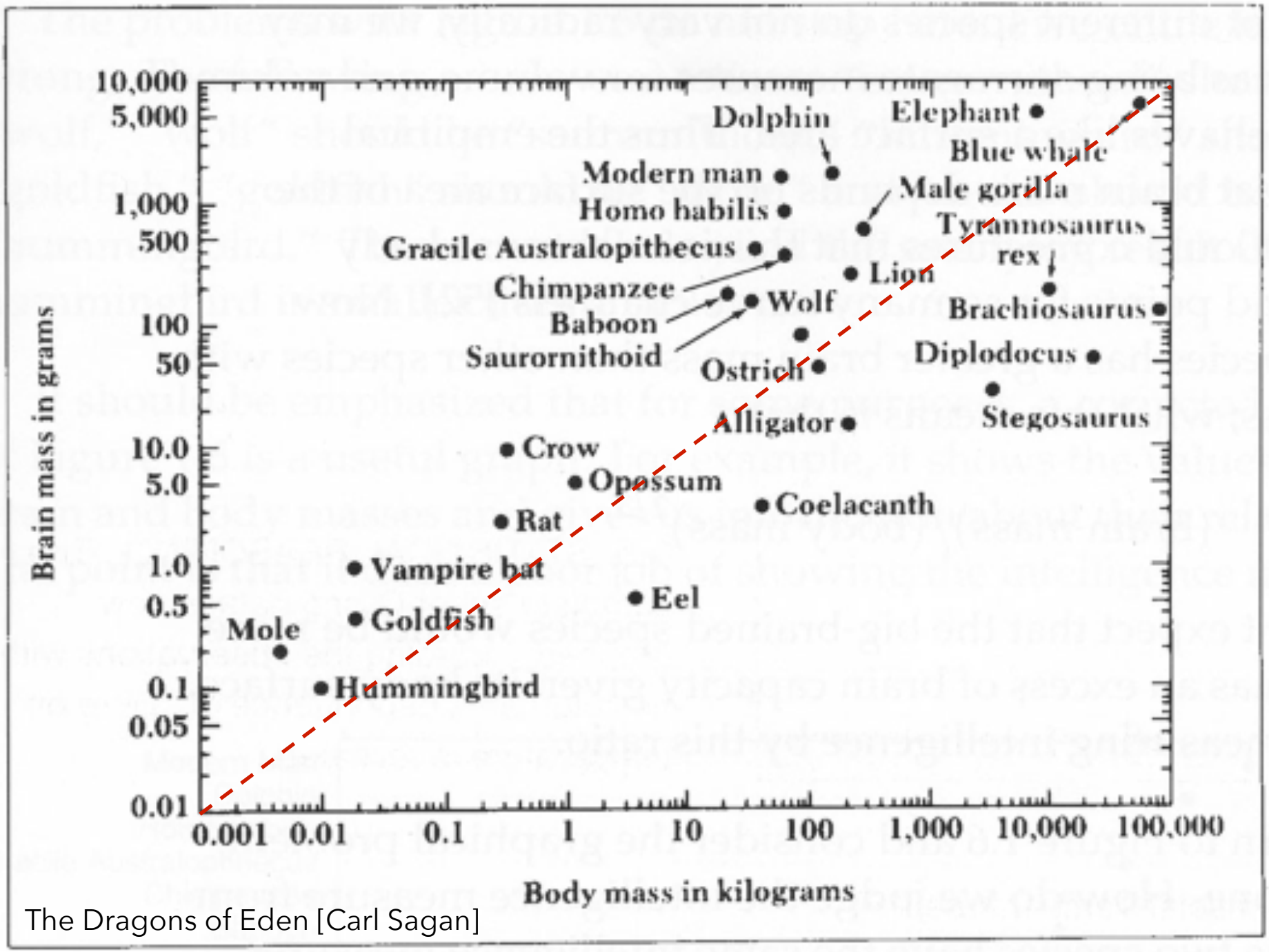
	A	B	C	D	E
1	ID	Name	Body Weight	Brain Weight	
2	1	Lesser Short-tailed Shrew	5	0.14	
3	2	Little Brown Bat	10	0.25	
4	3	Mouse	23	0.3	
5	4	Big Brown Bat	23	0.4	
6	5	Musk Shrew	48	0.33	
7	6	Star Nosed Mole	60	1	
8	7	Eastern American Mole	75	1.2	
9	8	Ground Squirrel	101	4	
10	9	Tree Shrew	104	2.5	
11	10	Golden 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

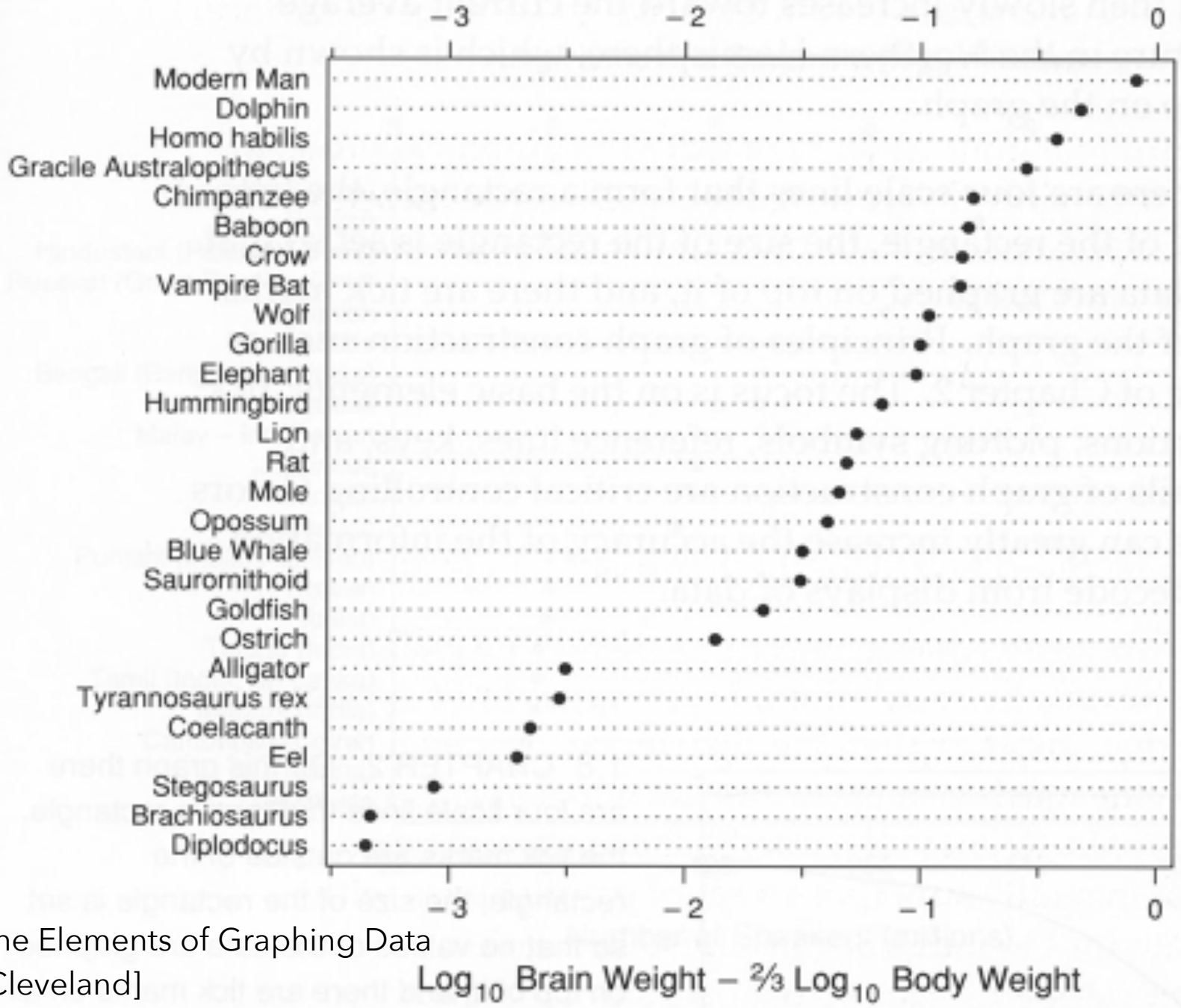
Ready



The Dragons of Eden [Carl Sagan]



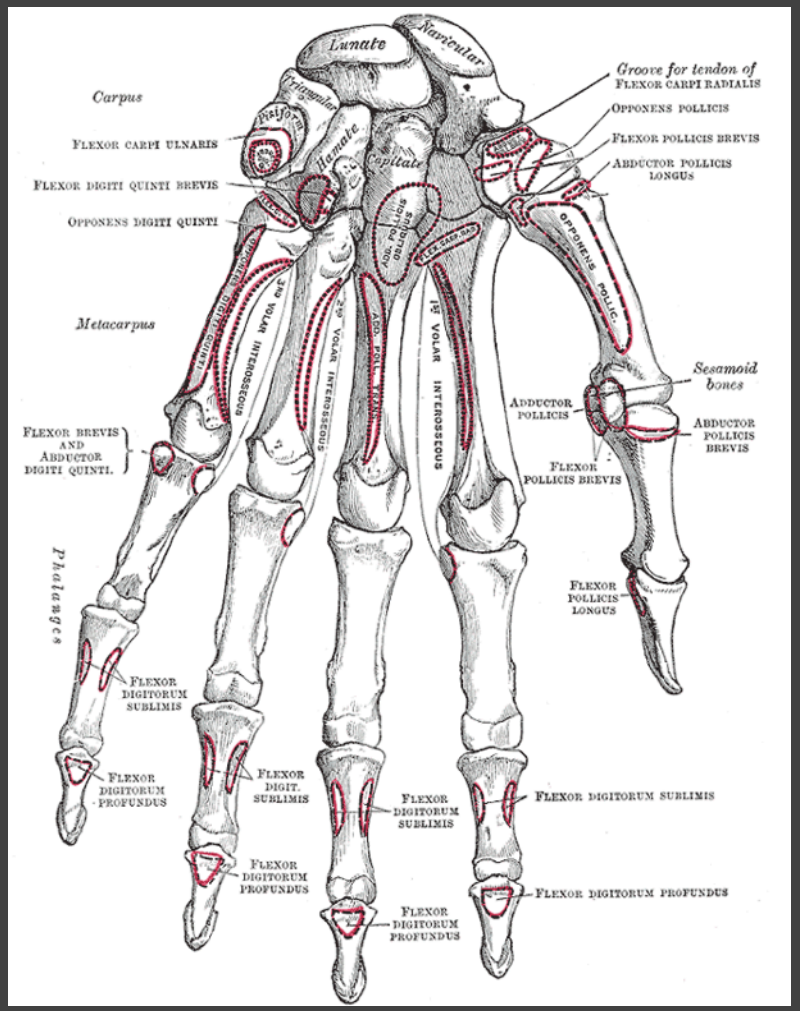
The Dragons of Eden [Carl Sagan]



The Elements of Graphing Data
[Cleveland]

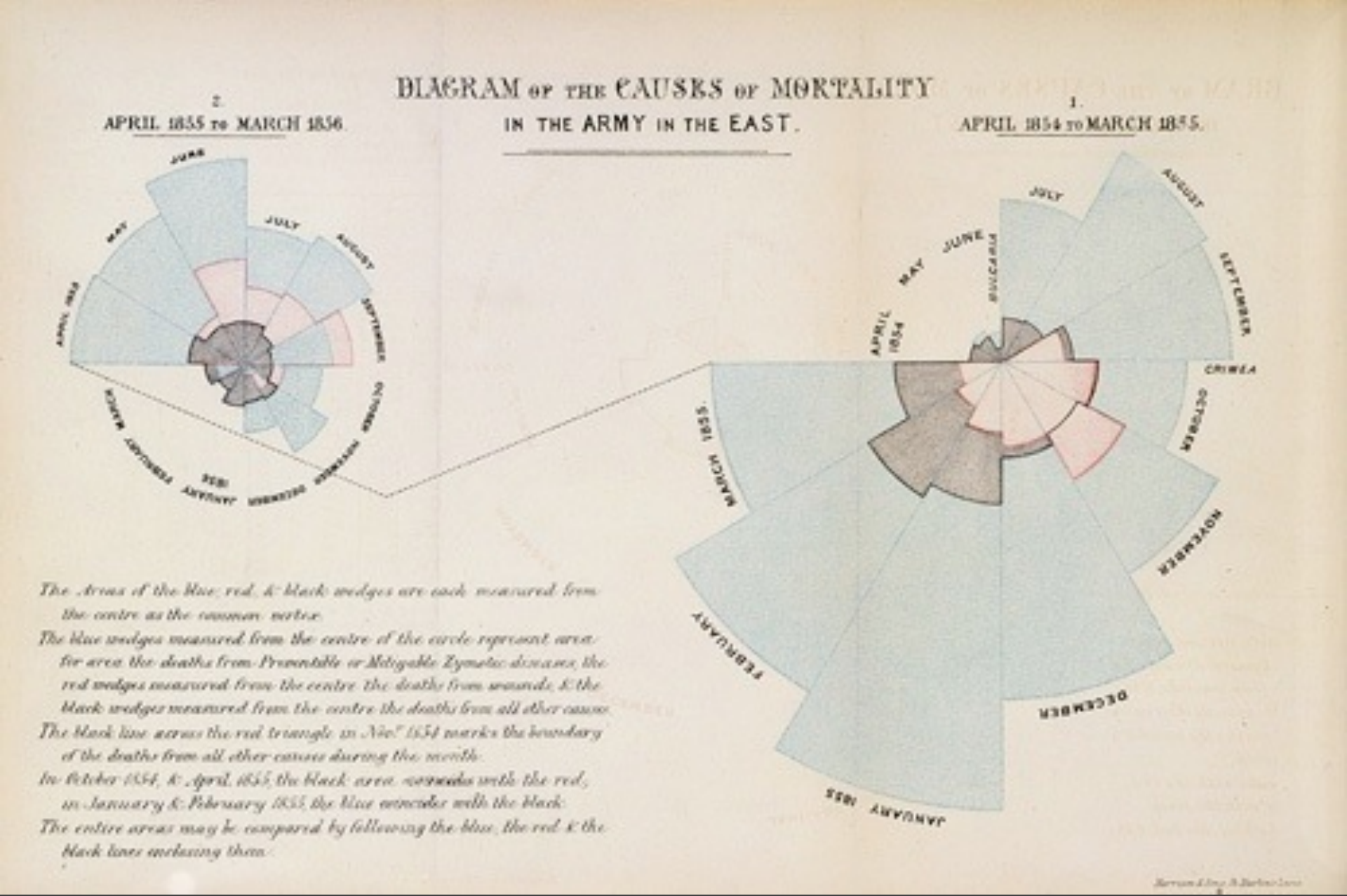
Convey Information to Others

Inspire

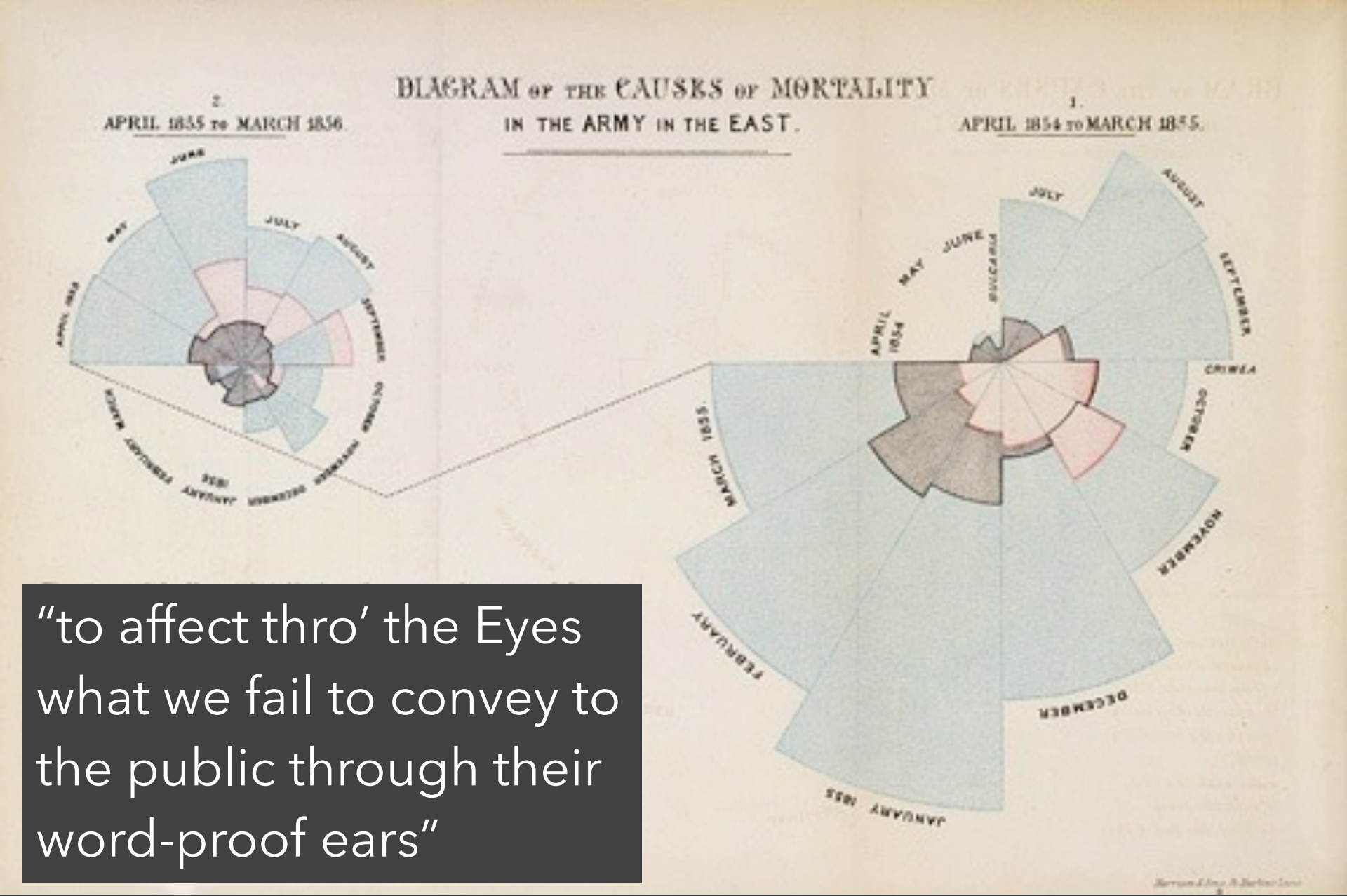


Bones in hand [from 1918 edition]

Double helix model [Watson and Crick 53]



1856 “Coxcomb” of Crimean War Deaths, Florence Nightingale



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

1856 "Coxcomb" of Crimean War Deaths, Florence Nightingale

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

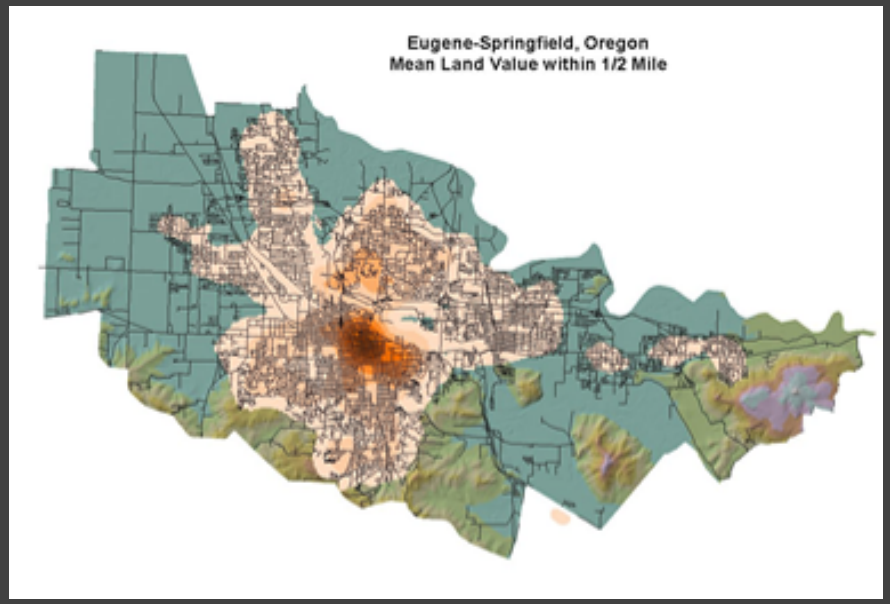
Visualization Research

Challenge

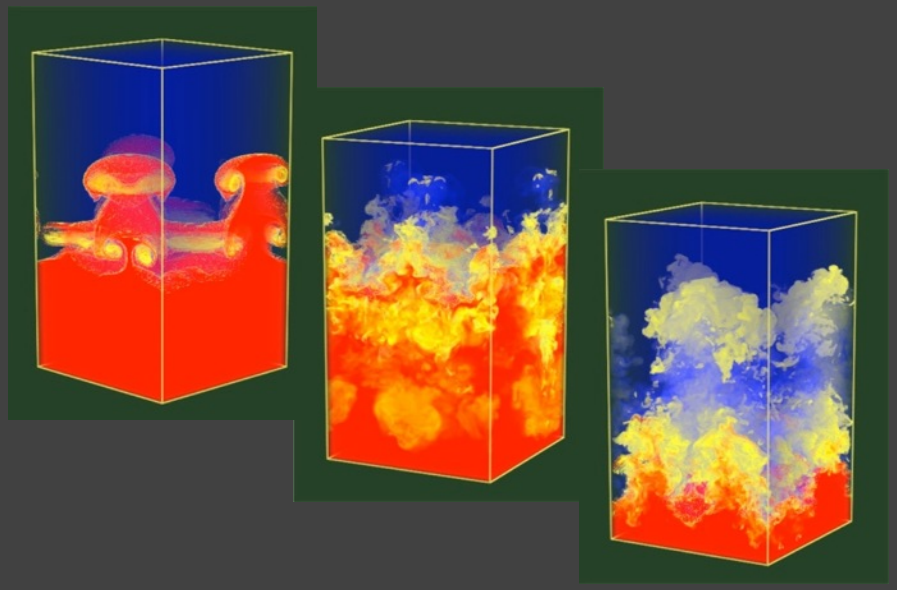
More and more unseen data
Faster creation and collection

Challenge

More and more unseen data
Faster creation and collection



Urban development planning
www.urbansim.org



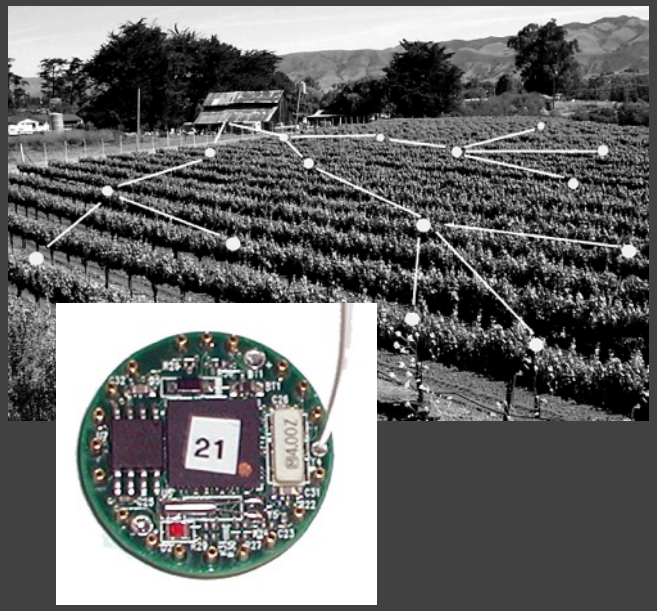
Fluid flow
ctr.stanford.edu

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

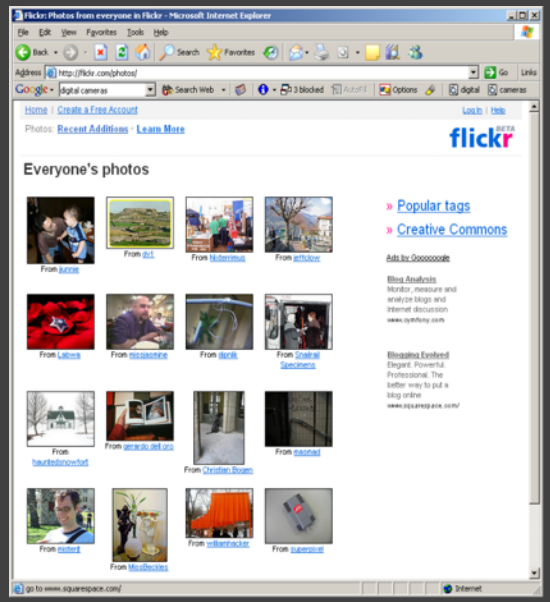
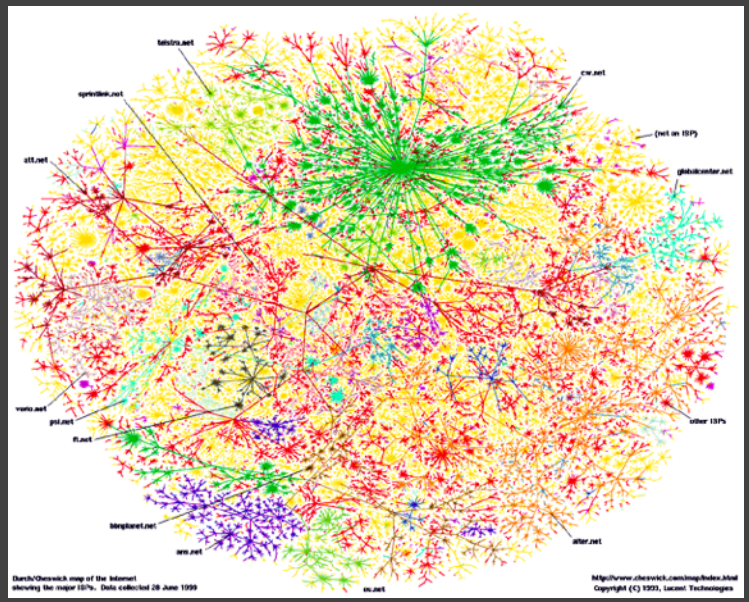


Photo sharing/annotation
flickr.com

Group Authoring
wikipedia.org



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

Internet

Challenge

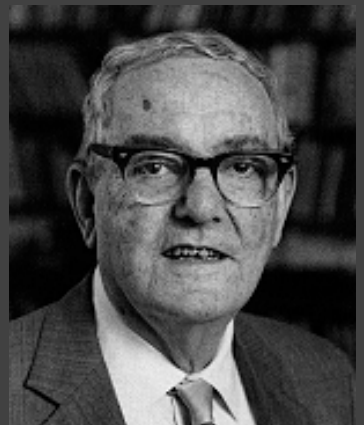
More and more unseen data
Faster creation and collection; Faster dissemination

5 exabytes of new information in 2002 [Lyman 03]
161 exabytes in 2006 [Gantz 07]
1,200 exabytes in 2010 [Gantz 10]

Necessitates **better tools and algorithms** for **visually conveying information**

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

Goals of Visualization Research

1 **Understand** how visualizations convey information

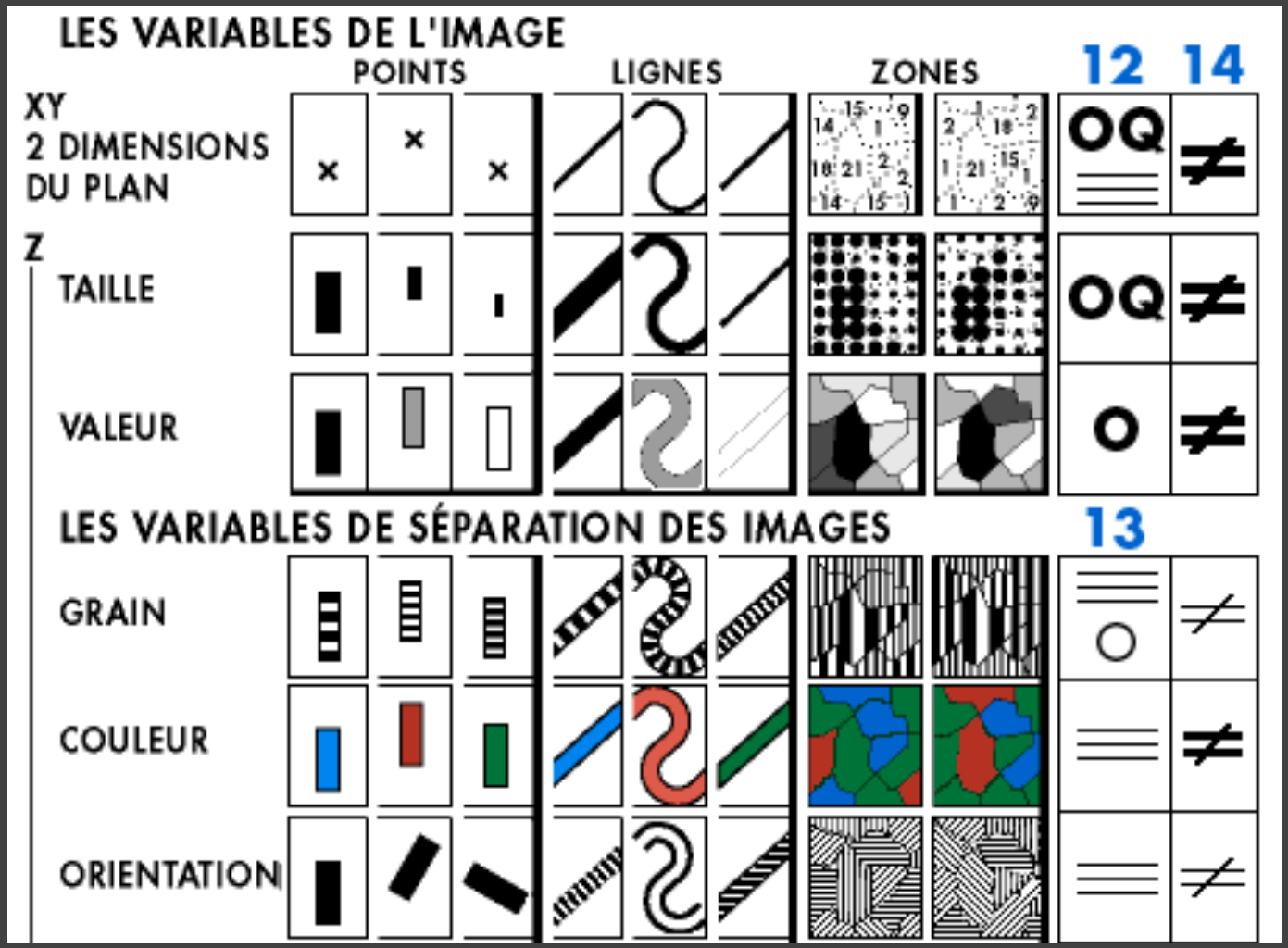
What do people perceive/comprehend?
How do visualizations inform mental models?

2 **Develop principles and techniques** for creating effective visualizations and supporting analysis

Amplify perception and cognition
Improve ties between visualization & mental model

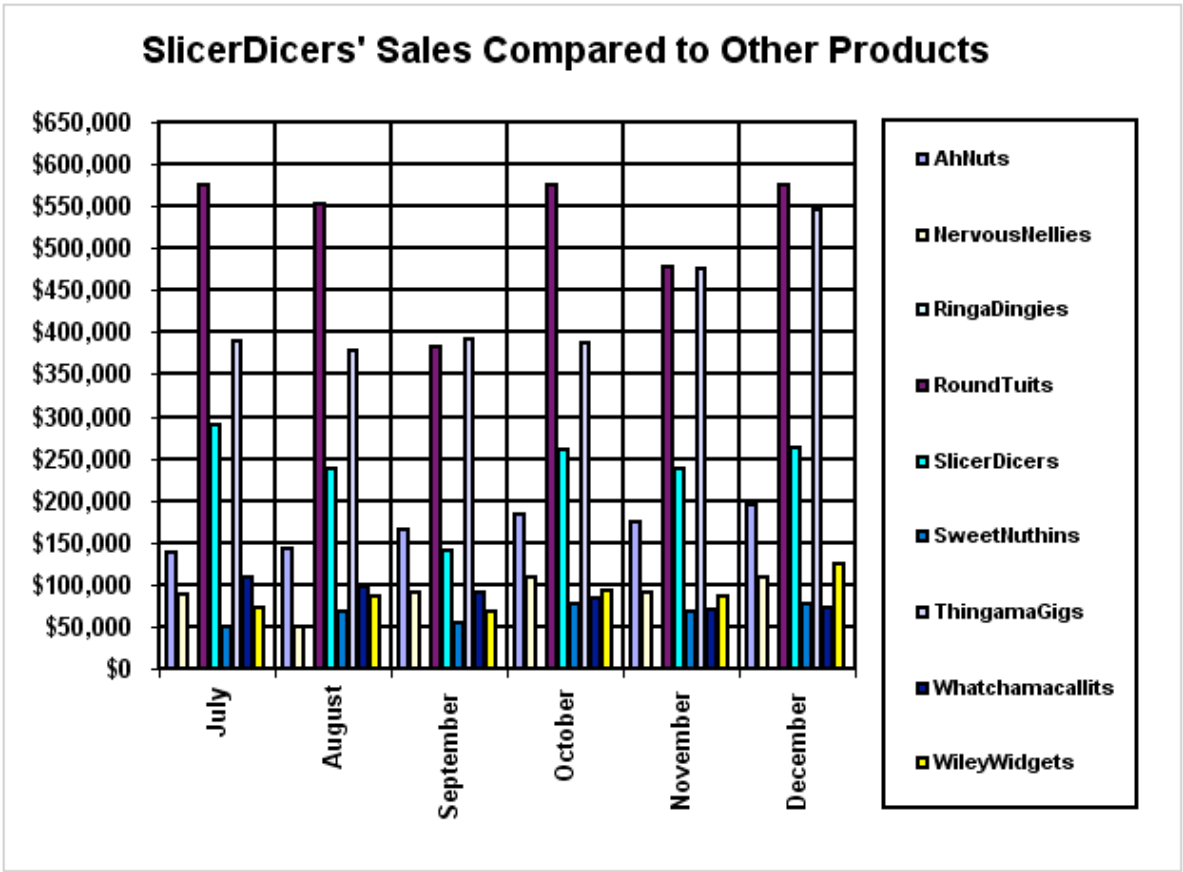
Course Topics

Data and Image Models

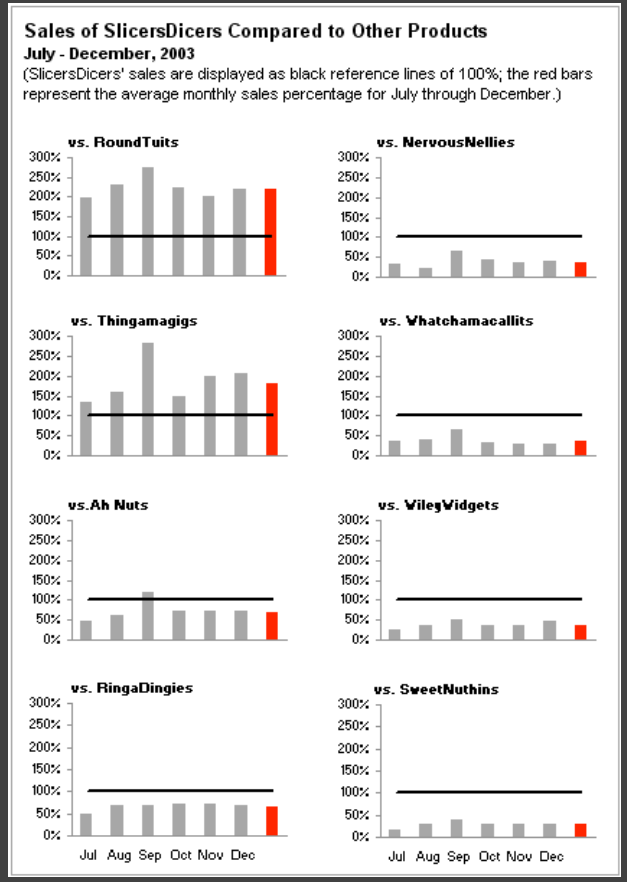


Sémiologie Graphique [Bertin 67]

Visualization (Re-)Design

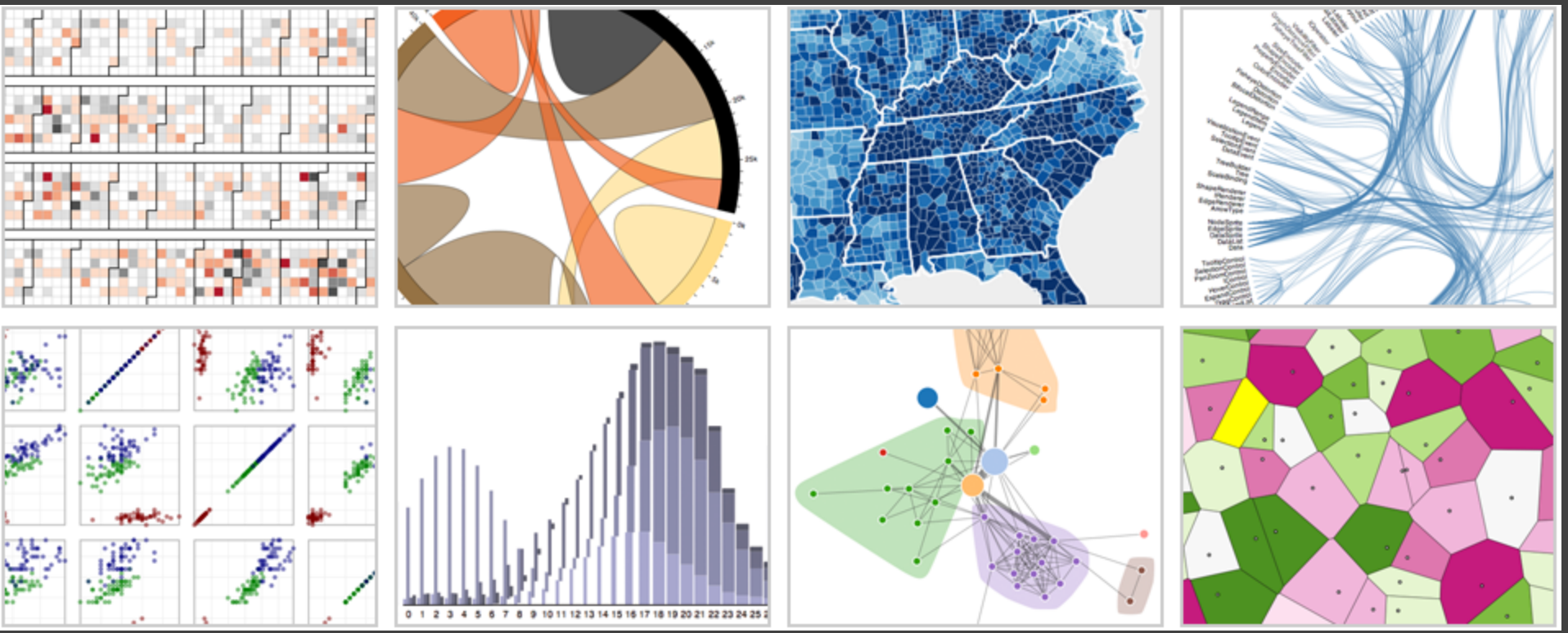


Problematic design



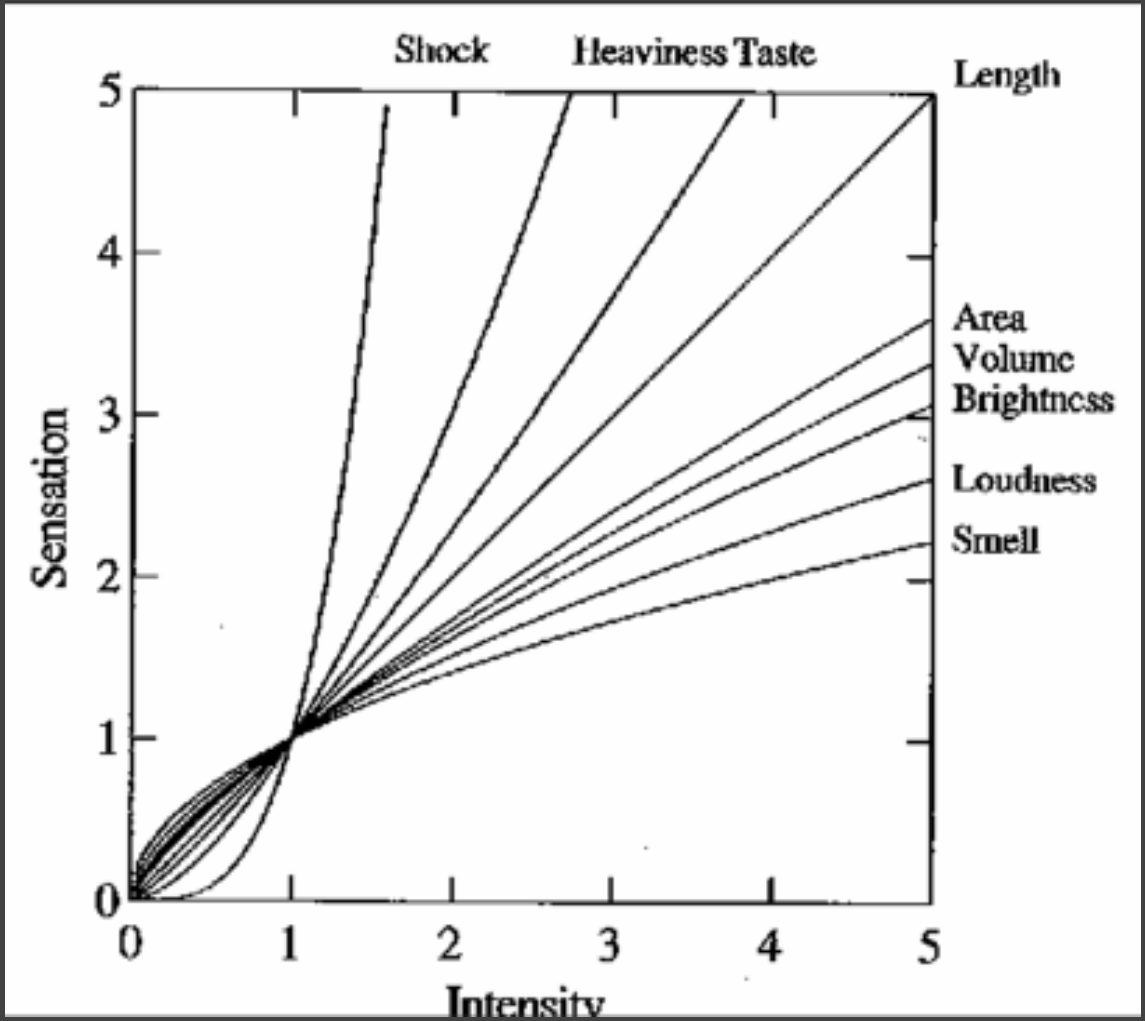
Redesign

Visualization Software



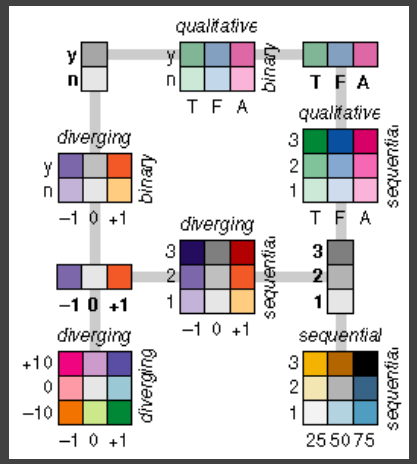
D3: Data-Driven Documents

Graphical Perception

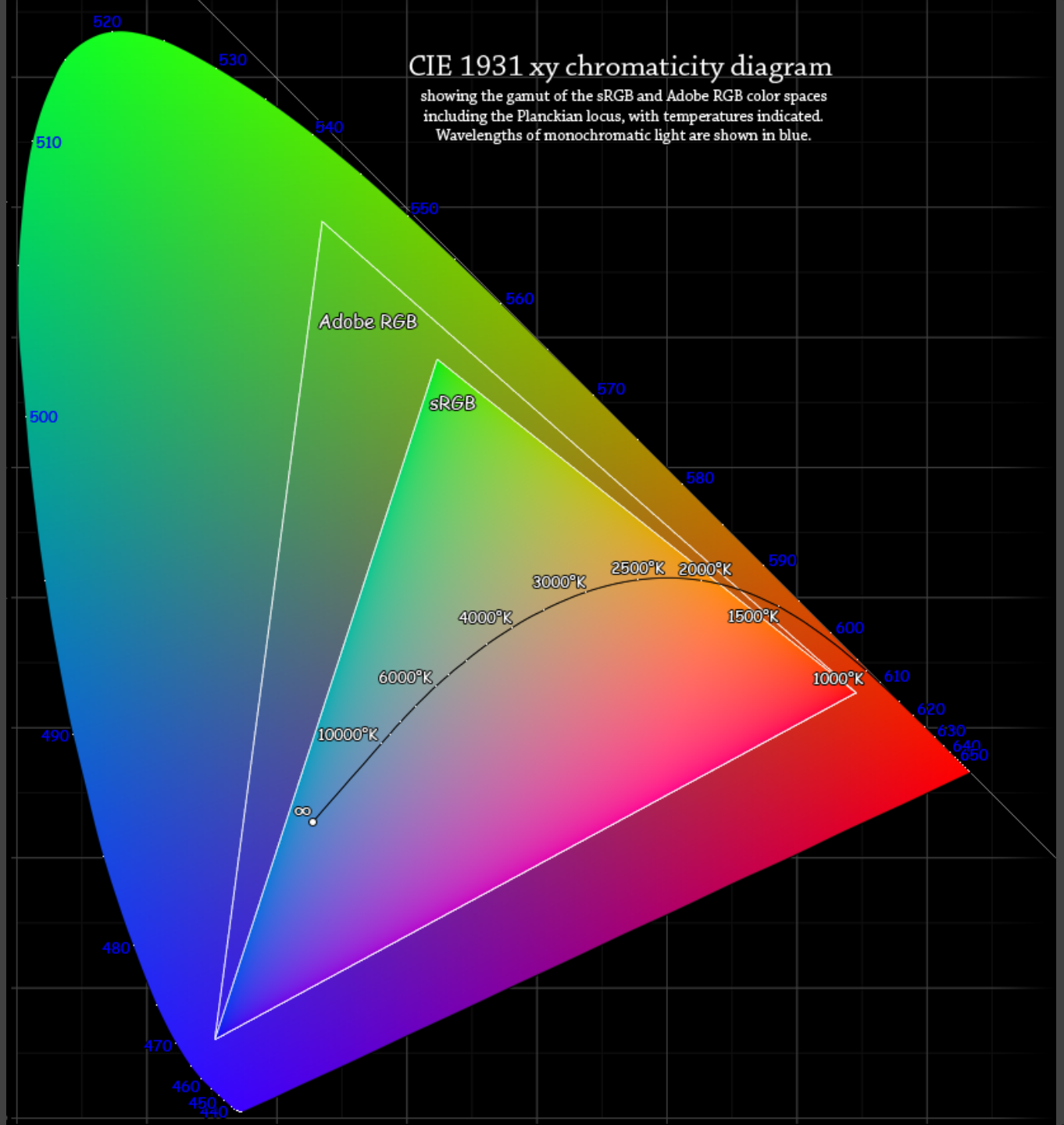


The psychophysics of sensory function [Stevens 61]

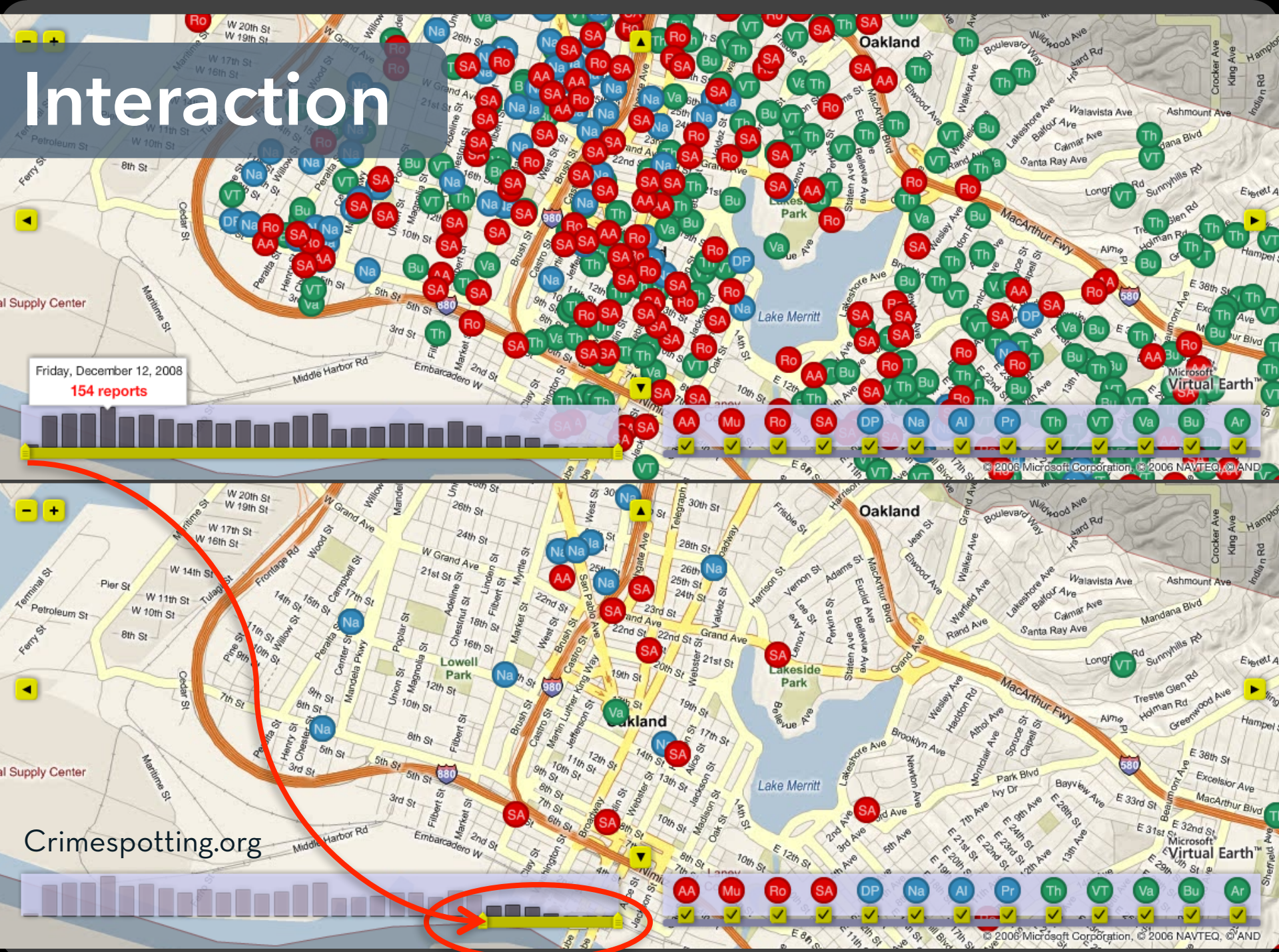
Color



Color Brewer

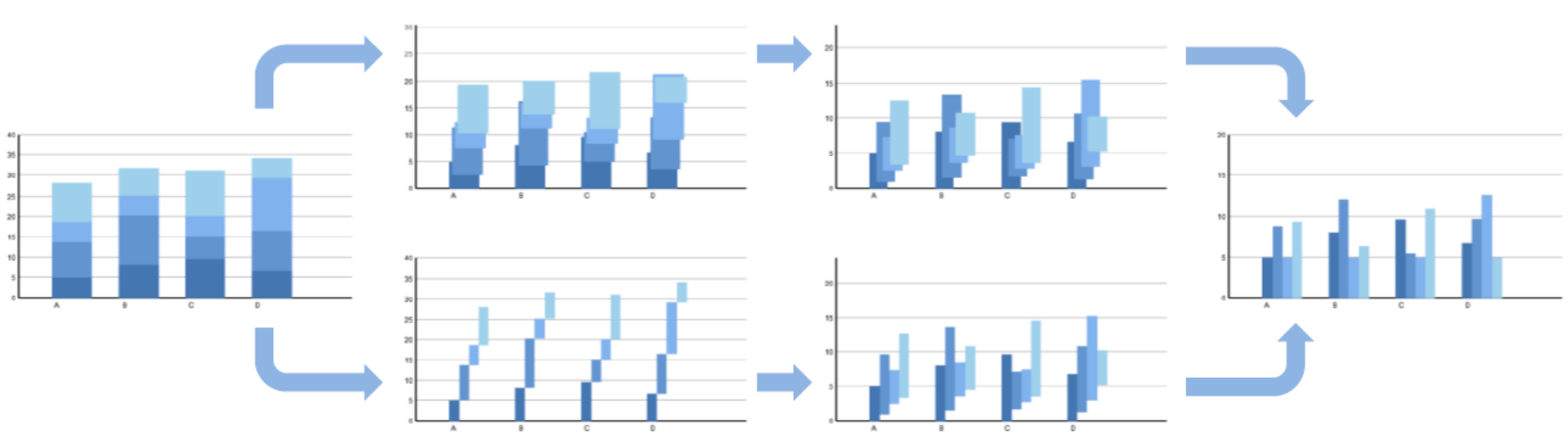


Interaction



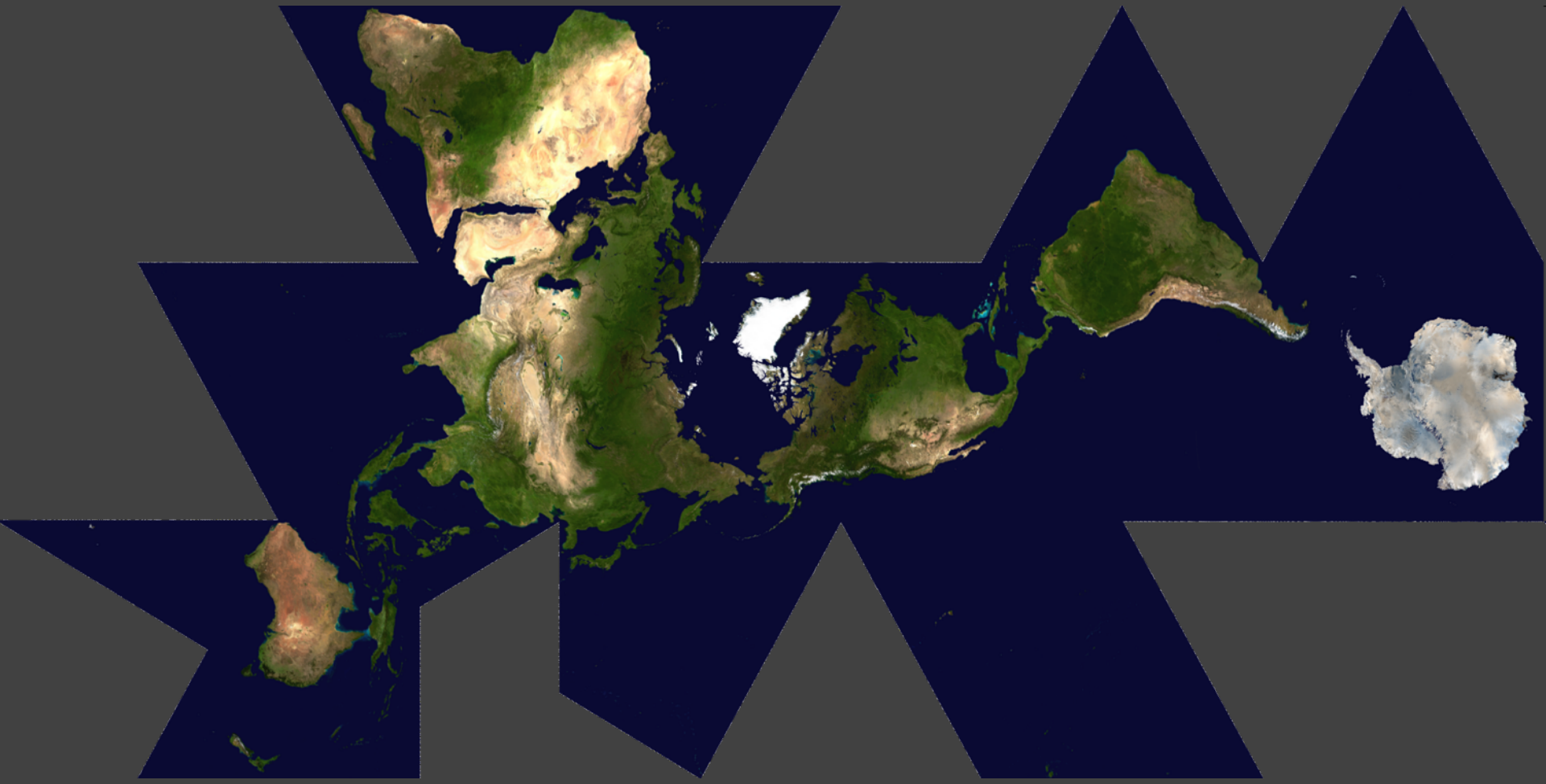
Crimespotting.org

Animation



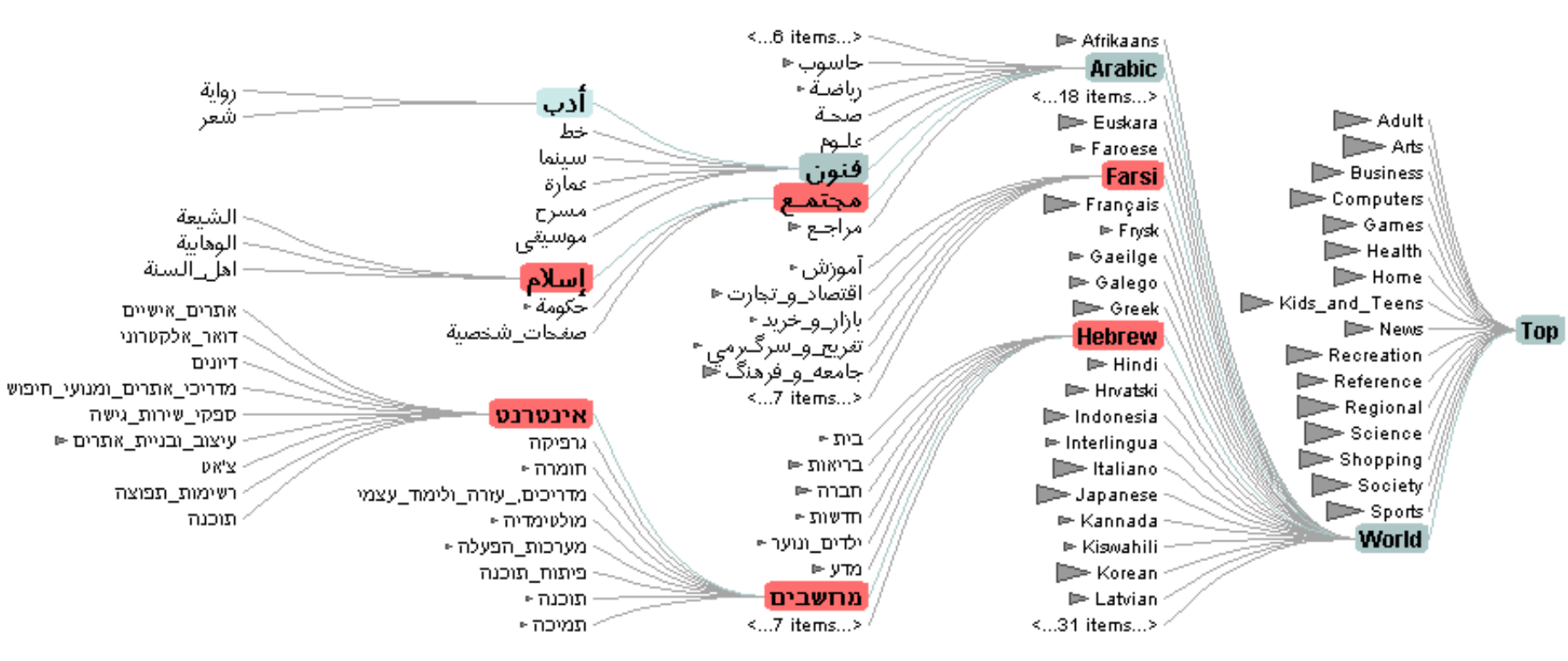
Animated transitions in statistical data graphics [Heer & Robertson 07]

Mapping & Cartography



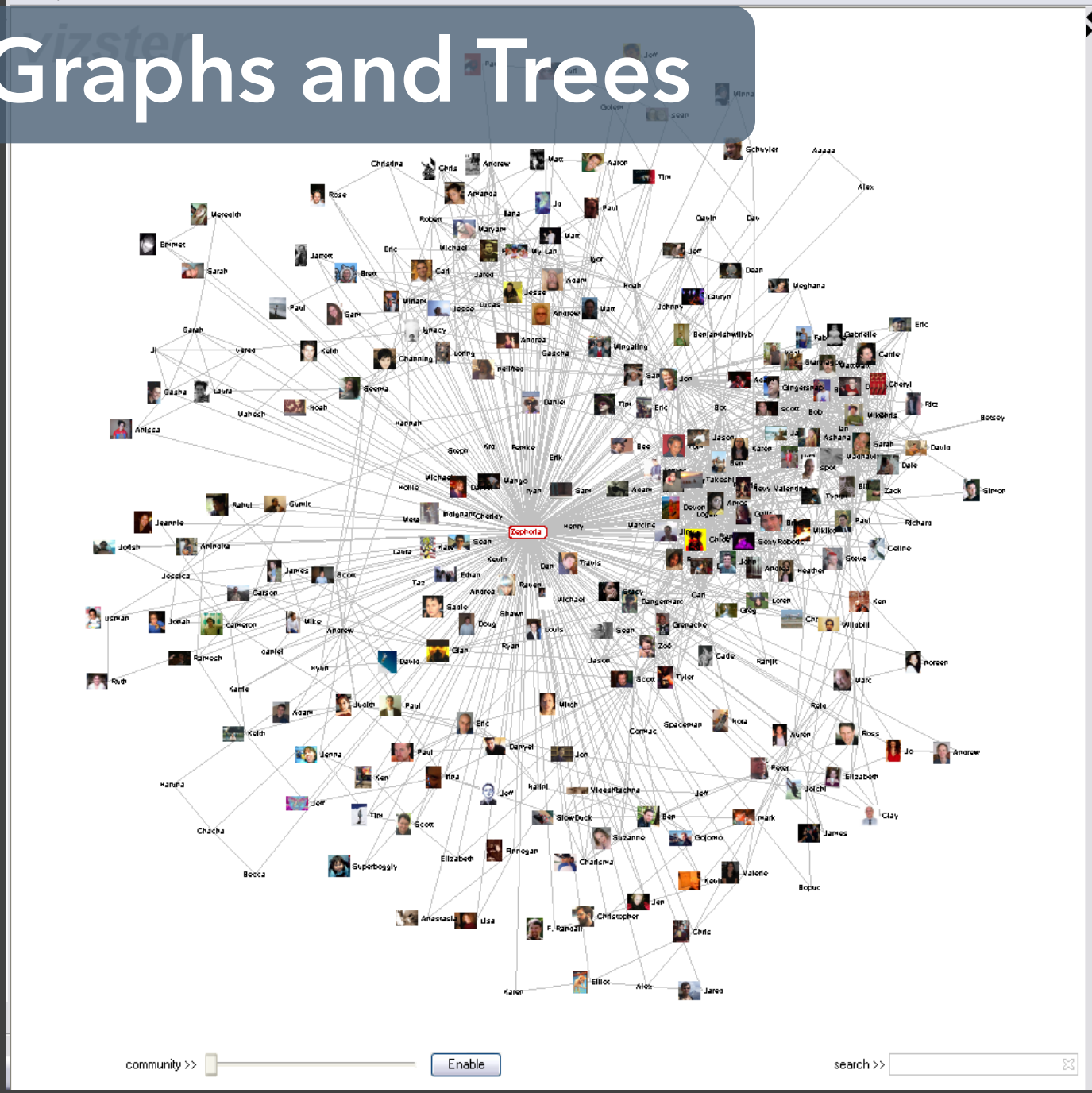
Dymaxion Maps [Fuller 46]

Graphs and Trees



Degree-Of-Interest Trees [Heer & Card 04]

Graphs and Trees



Zephoria

User ID	21721
Friends	266
Age	??
Gender	Female
Status	Single
Location	San Francisco, CA
Hometown	Lancaster, PA
Occupation	researcher: social networks, identity, context
Interests	apophenia, observing people, culture, questioning power, reading, buddhism, ipseity, computer-mediated communication, social networks, technology, anthropology, stomping
Music	psytrance/goa/trance [Infected Mushroom, Son Kite... Iboga/Digital Structures], Ani Difranco, downtempo, Thievery Corporation, Beth Orton, Morcheeba, Ween, White Stripes
Books	Authors: Erving Goffman, Stanley Milgram, Jeanette Winterson, Eric Schlosser, Leslie Feinberg, Dorothy Allison, Italo Calvino, Hermann Hesse
TV Shows	??
Movies	Koyaanisqatsi, Amelie, Waking Life, Tank Girl, The Matrix, Clockwork Orange, American Beauty, Fight Club, Boys Don't Cry
Member Since	??
Last Login	2003-10-21
Last Updated	2003-10-21
About	[Some know me as danah...] I'm a geek, an activist and an academic, fascinated by people and society. I see life as a very large playground and enjoy exploring its intricacies. I revel in life's chaos, while simultaneously providing my own insane element. My musings: http://www.zephoria.org/thoughts/
Want to Meet	Someone who makes life's complexities seem simply elegant.

Text Visualization

Visualizations / Word tree / Alberto Gonzales

Creator: Martin Wattenberg
Tags:

explore
visualizations
data sets
comments
topic hubs

participate
create visualization
upload data set
create topic hub
register

learn more
quick start
visualization types
data format & style
about Many Eyes
FAQ
blog

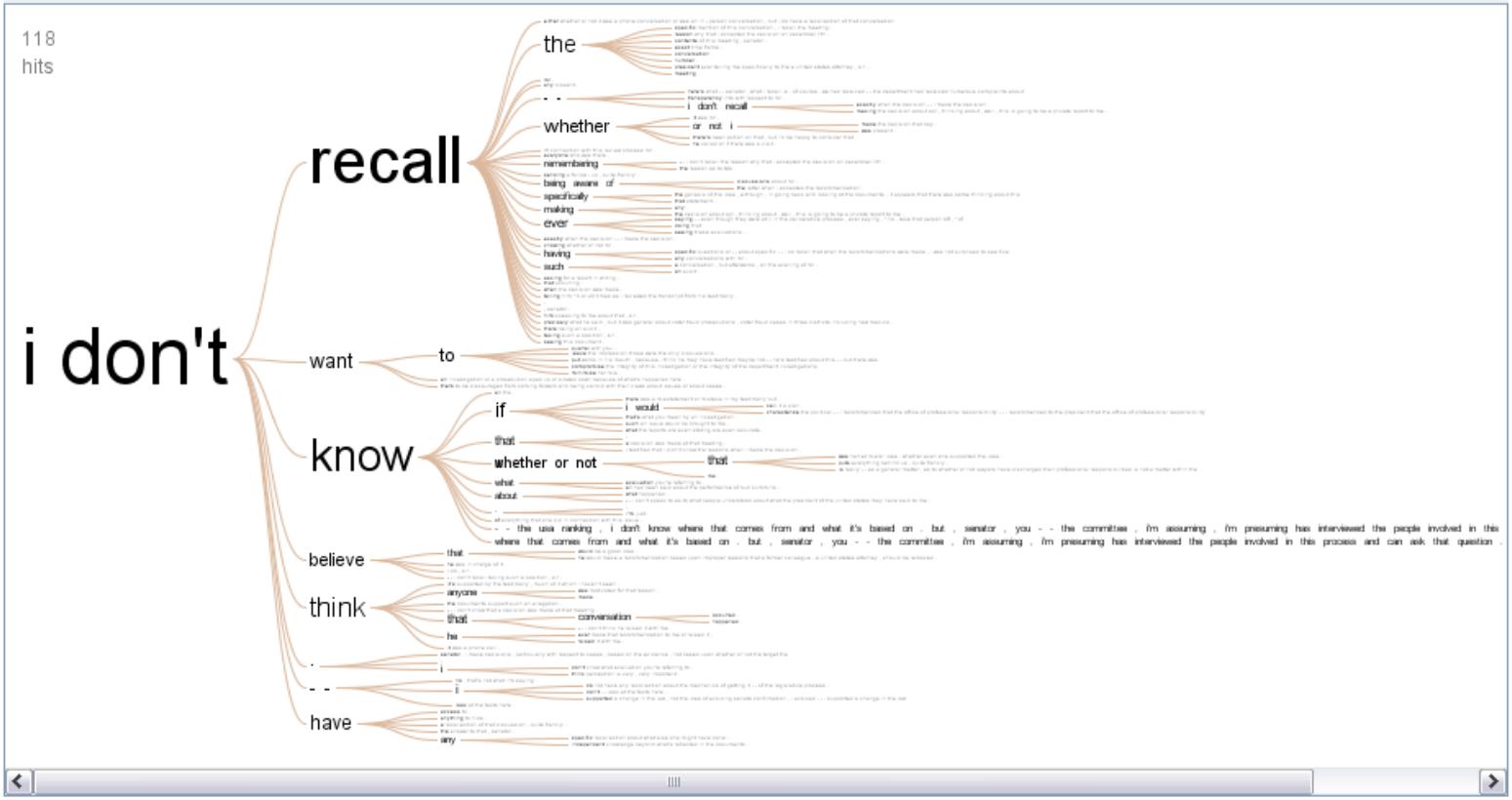
contact Us
contact
report a bug

legal
terms of use

Popular Dataset Tags

2007 2008 bible blog books census crime education eharmony election energy food health inauguration internet ireland literature lyrics media music network obama people politics population president prices religion

Search i don't | Back | Forward | Start | End | Occurrence Order | Clicks Will Zoom



Data file: Word in testimony from Gonzales, 4/19/2007 Data source: CQ Transcript Wire via the Washington Post This data set has not yet been rated

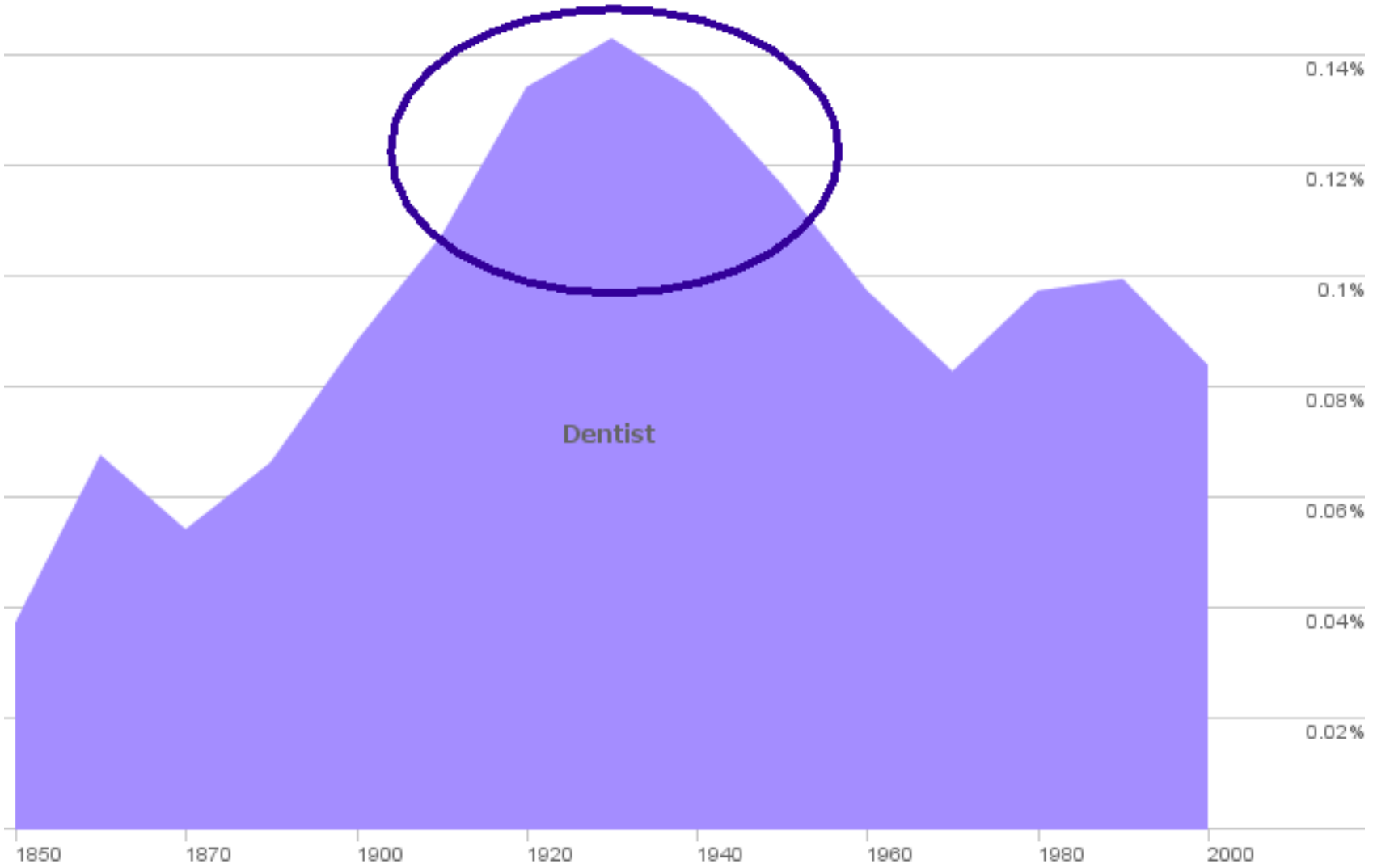
full image | share this | watch this | add to topic hub | rate this

Comments (4)

currently showing

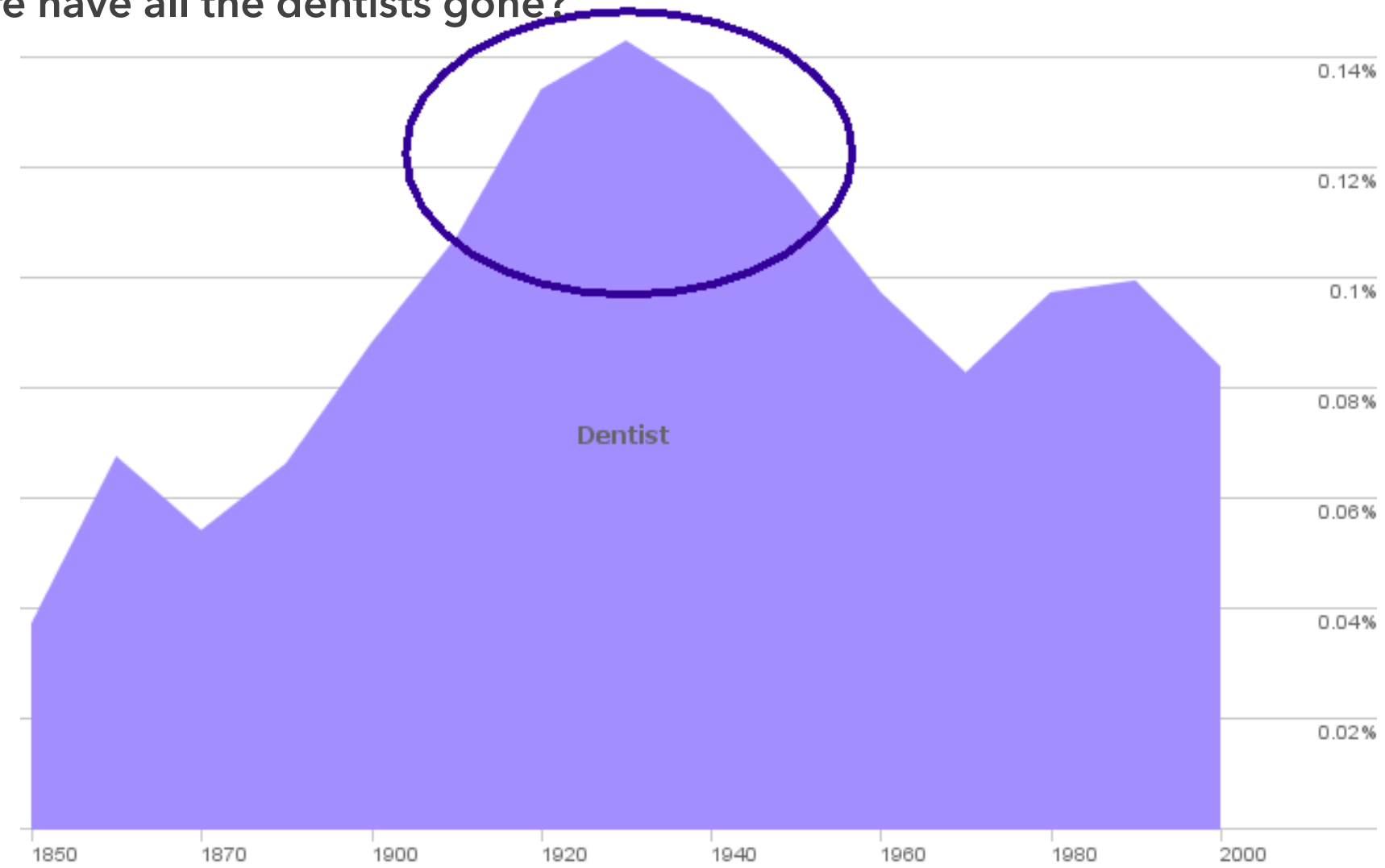
This visualization has 4 positive and 0 negative

Collaboration and History

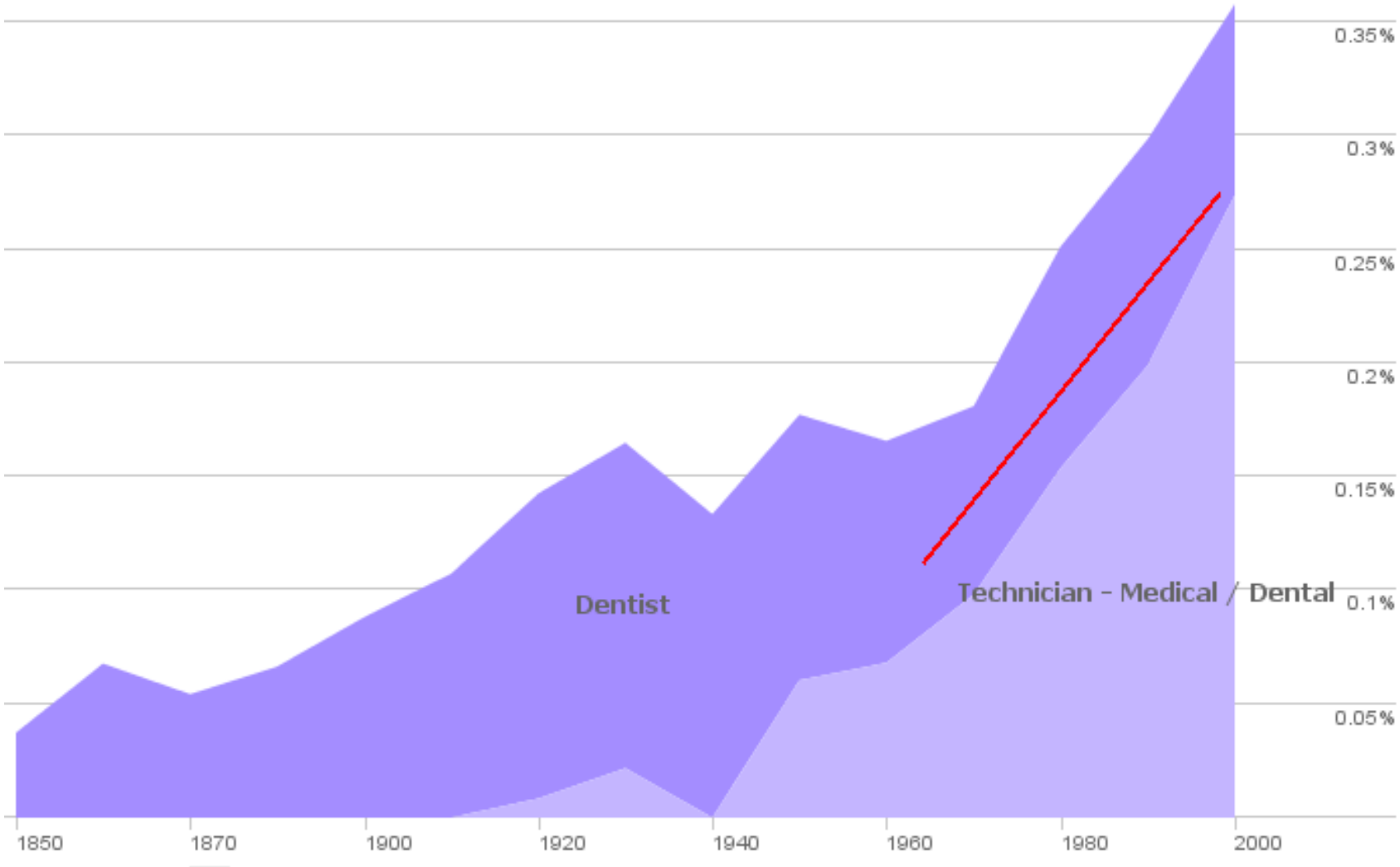


Collaboration and History

Where have all the dentists gone?



Collaboration and History



Course Mechanics

You should expect to:

1. *Evaluate and critique* visualization designs
2. *Implement* interactive data visualizations
3. *Gain* an overview of research & techniques
4. *Develop* a substantial visualization project

Instructors

cse512@cs

Instructor

Jeffrey Heer — OH: *Th 10-11am, 642 CSE*
Assoc Prof, CSE — http://jheer.org

Assistants

Dominik Moritz — OH: *Fri 1:30-2:30, 218 CSE*
Jeff Snyder — OH: *Fri 1:30-2:30, 218 CSE*

Dominik Moritz

@domoritz

- Graduated from University of Potsdam
- 1½ year PhD student with Bill Howe & Jeff Heer
- Databases + Visualization
- homes.cs.washington.edu/~domoritz

Google

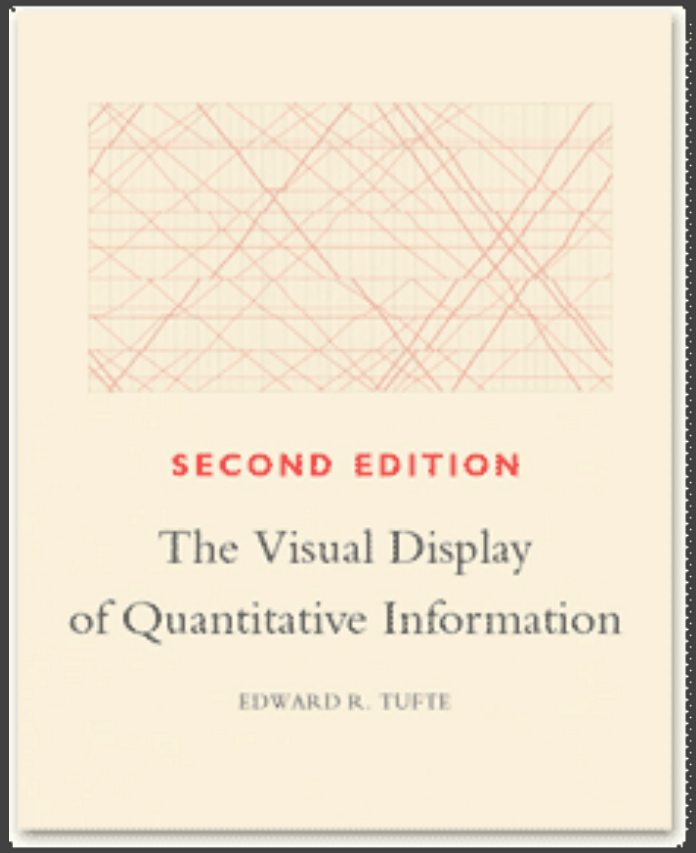
Hello
my name is...

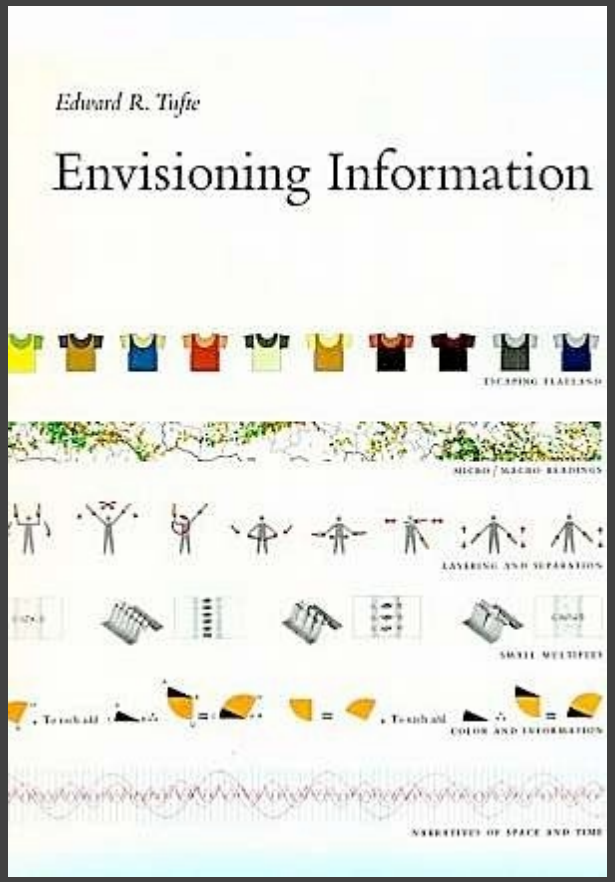
Jeff Snyder

Office Hours: F 1:30-2:30 CSE 218

- graduated Princeton in 2013
- 2nd year PhD student with Jeff Heer
- part time at Google Seattle, working on coLaboratory
- loves javascript, warts and all
- more at http://jsnyd.es

Textbooks





See also: www.edwardtufte.com

Readings

Some from textbooks, also many papers

Material in class will loosely follow readings

Readings should be read by start of class

Post discussion comments on class Canvas forum

- Comments must be posted within 1 day of lecture
- You have 2 “passes” for the quarter

Assignments

Class Participation (10%)

A1: Visualization Design (10%)

A2: Exploratory Data Analysis (15%)

A3: Interactive Visualization Software (25%)

FP: Final Project (40%)

Final Project

Visualization research project on topic of choice

Project write-up in form of a **short research paper**

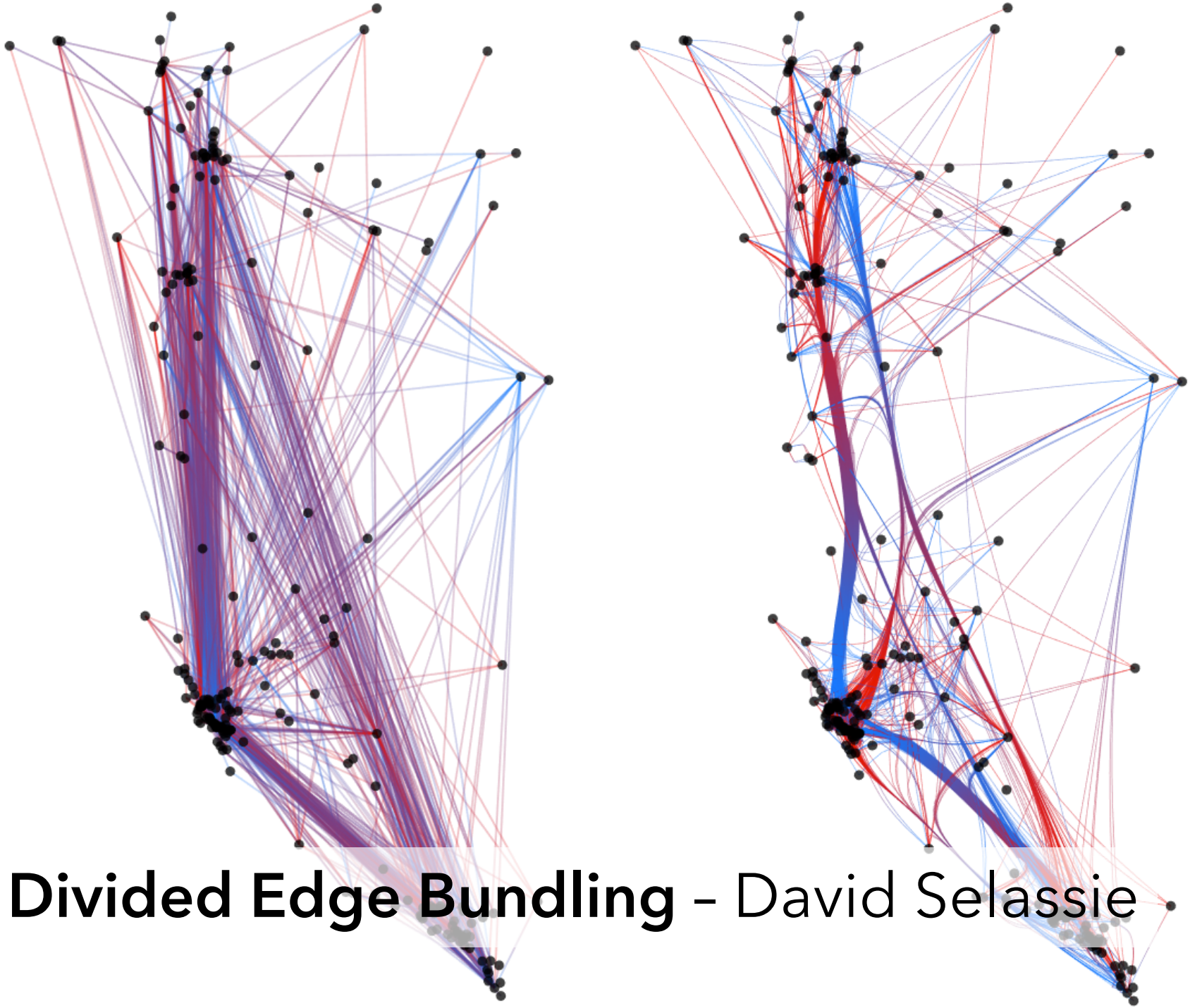
Project **check-ins** and **final poster/demo show**

Projects from **previous classes** have been:

- Published (e.g., at the IEEE InfoVis conference)
- Featured in the New York Times
- Released as successful open source projects

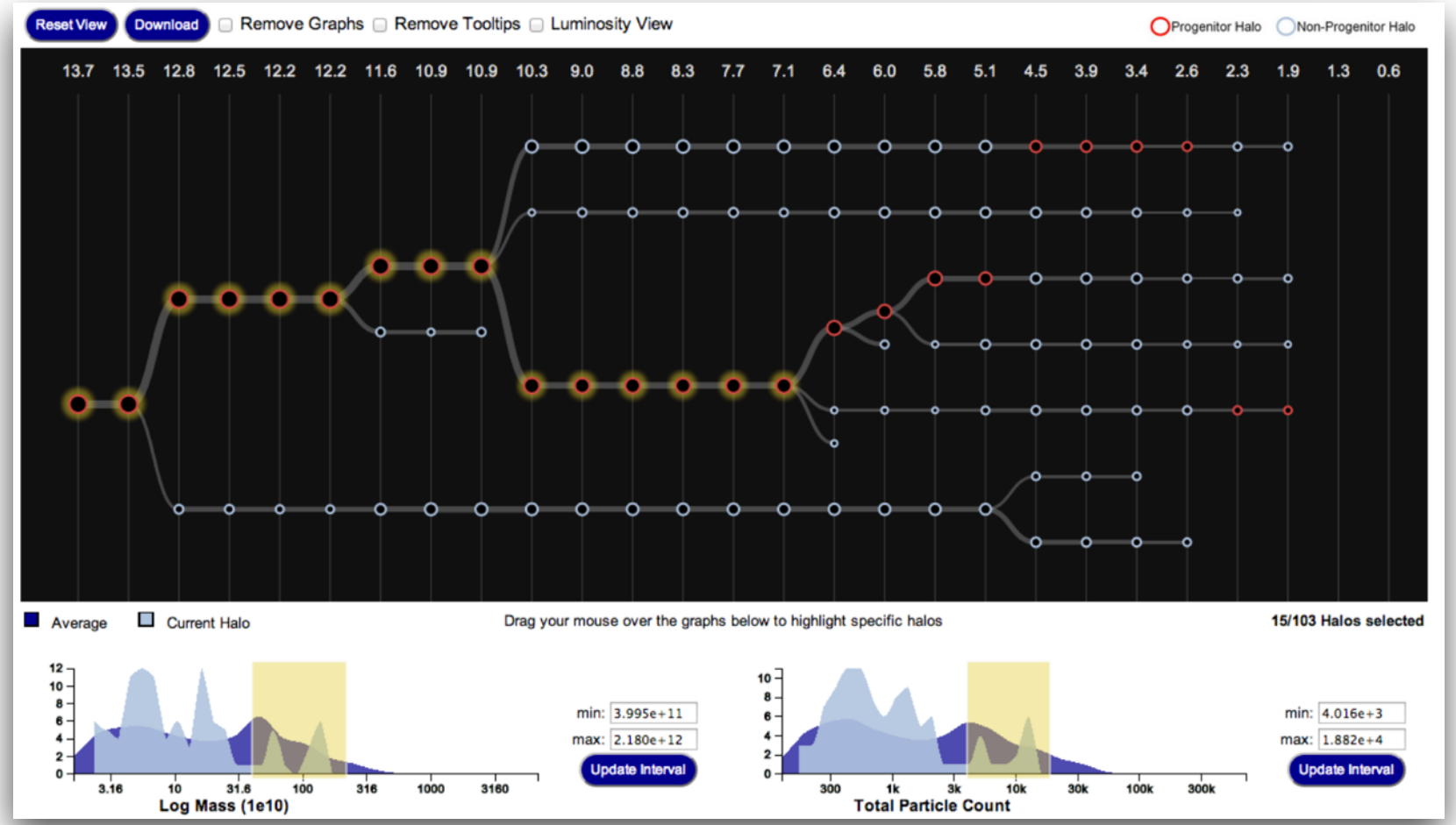
RunMonster

Troy Brant & Steve Marmon



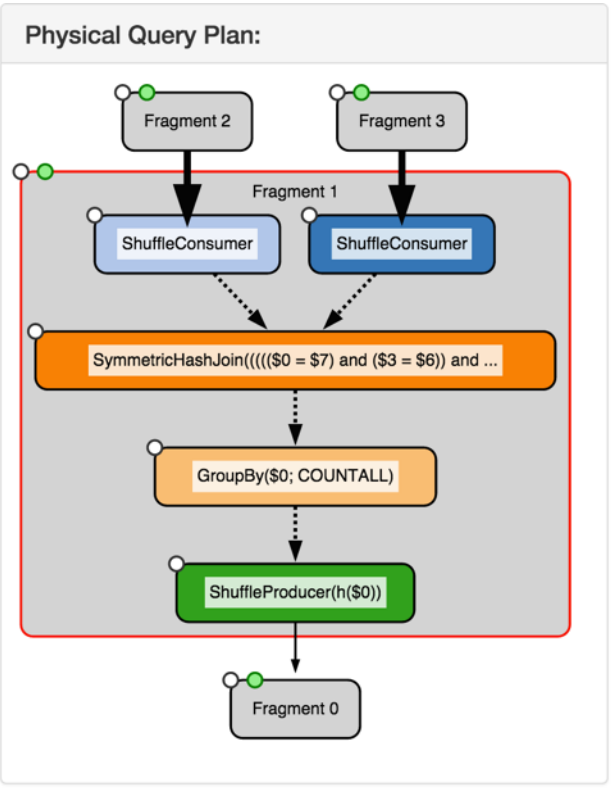
Divided Edge Bundling – David Selassie

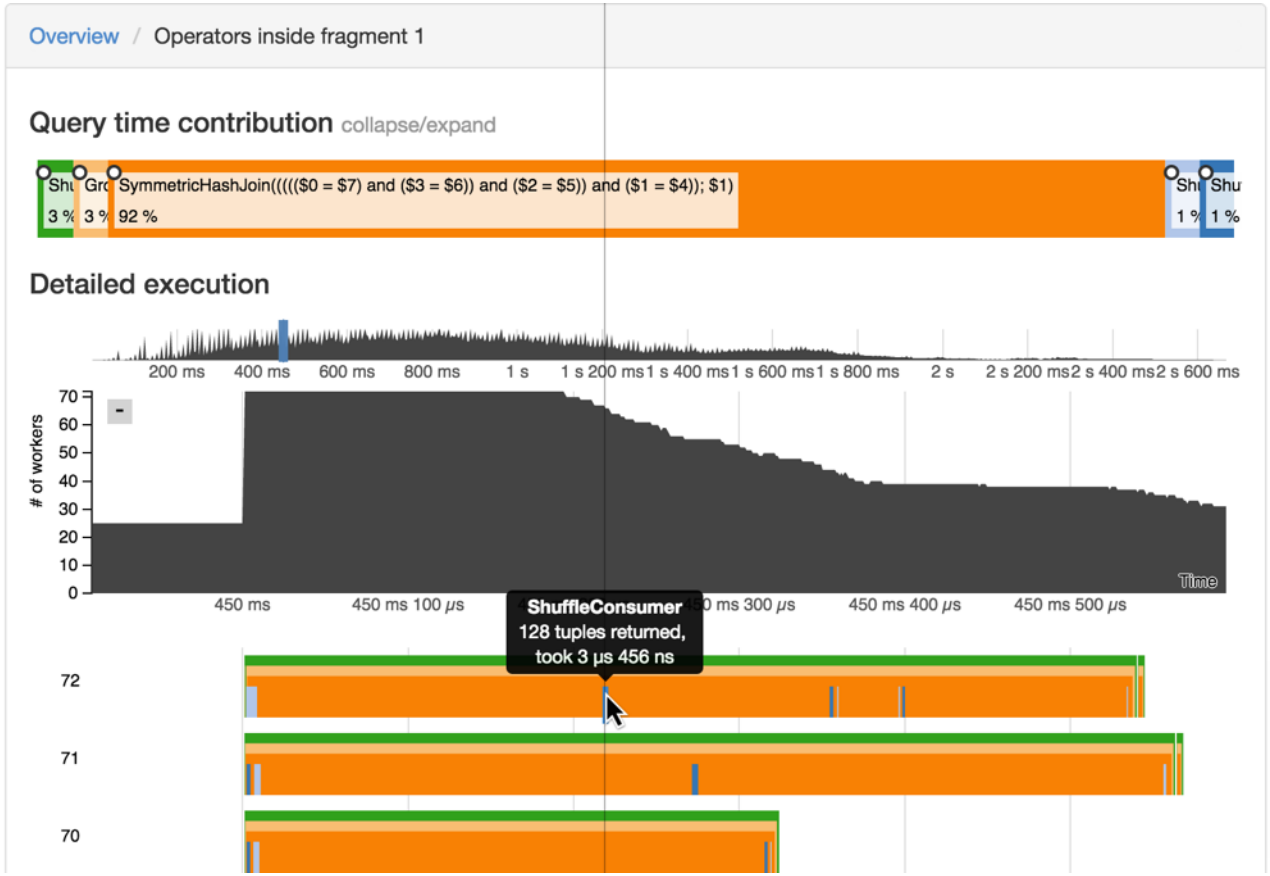
Visualizing Galaxy Merger Trees



S. Loebman, **J. Ortiz**, **L. Orr**, M. Balazinska, T. Quinn et al. [SIGMOD '14]

Perfopticon Distributed Query Performance

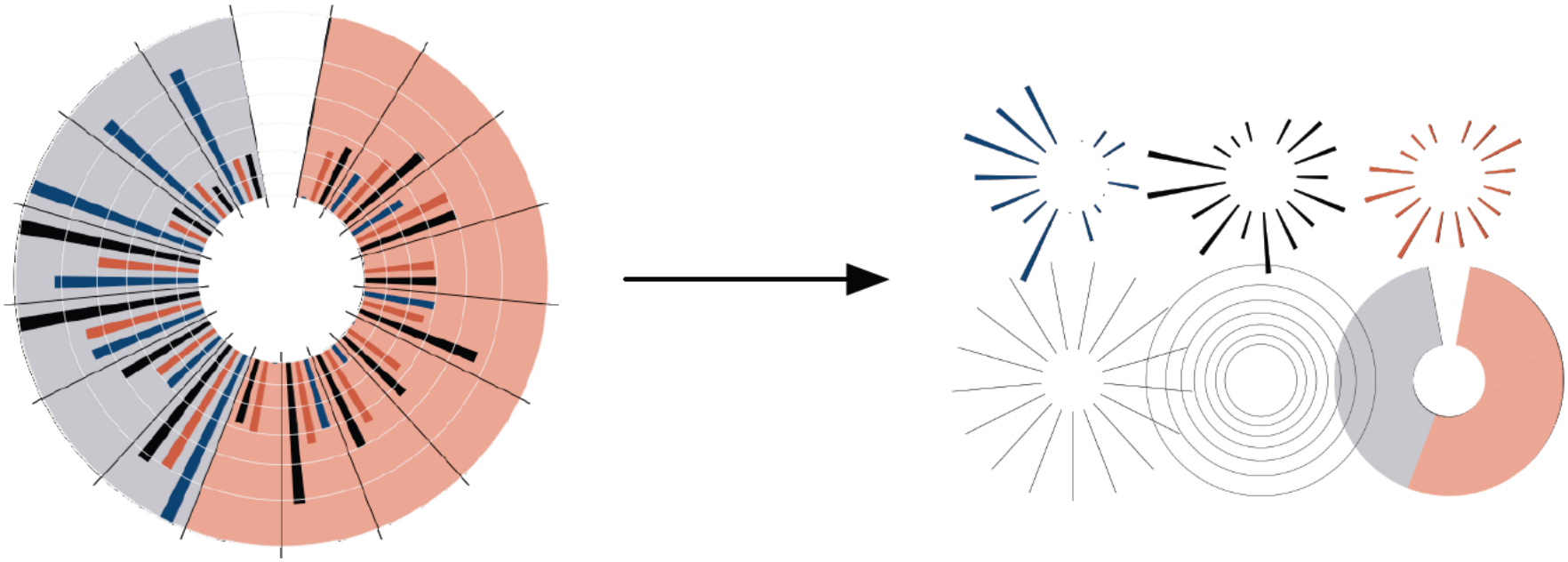


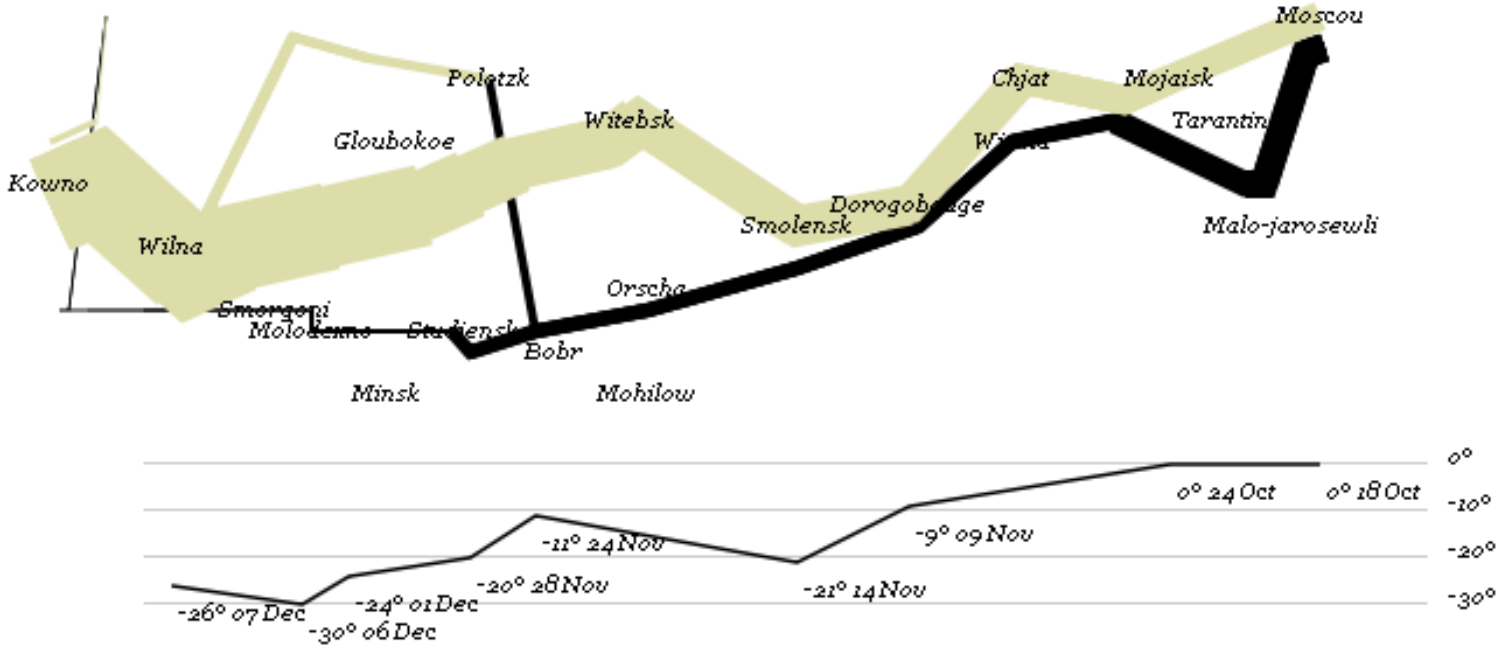


Dominik Moritz et al. [EuroVis '15]

Protovis: A Graphical Toolkit for Visualization

Mike Bostock



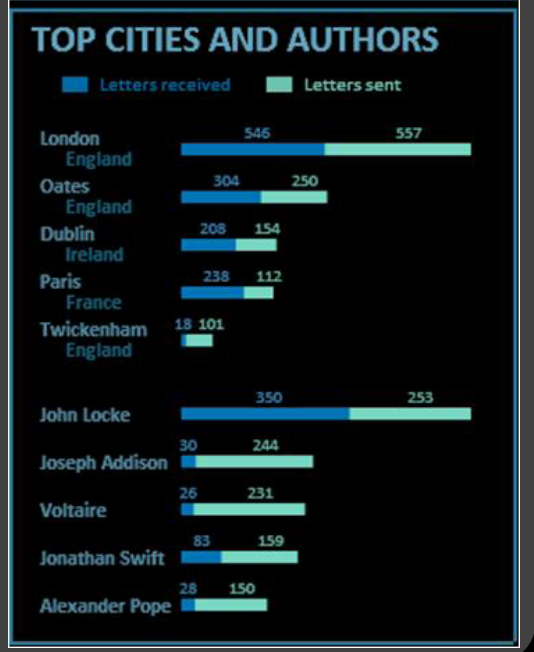



```
var army = pd.nest(napoleon.army, "dir", "group“);
var vis = new pv.Panel();

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

```
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)
     .font("italic 10px Georgia")
     .text(function(d) d+"°").textBaseline("center");

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))
    .textBaseline("top").font("italic 10px Georgia");
```

Visualizing the Republic of Letters

Daniel Chang, Yuankai Ge, Shiwei Song



Questions?

Assignment 1: Visualization Design

Design a static visualization for a data set.

After the World War II, antibiotics were considered “wonder drugs.” To learn which drug is most effective for which bacterial infection, performance of the three most popular antibiotics were gathered.

You must choose the message you want to convey. What task do you want to support? What insight do you want to communicate?

Assignment 1: Visualization Design

Design a **static visualization** for the data set.
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 **5:00 pm, Monday April 6**.