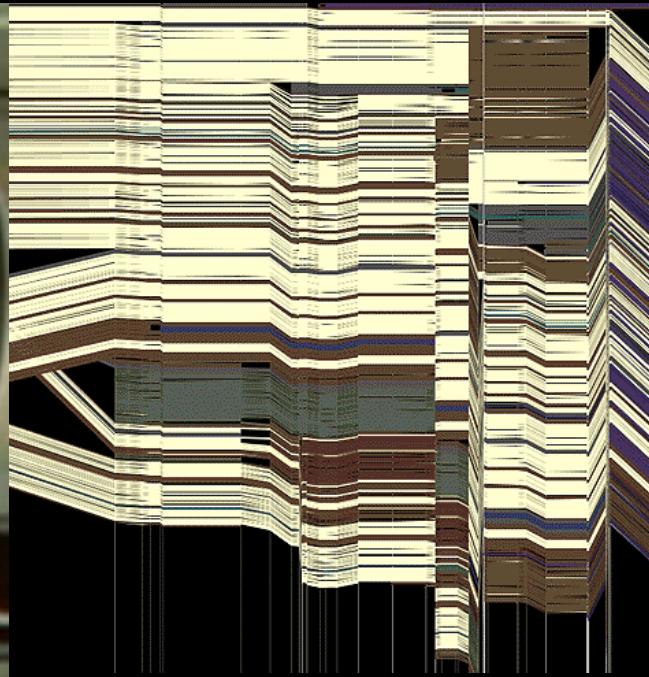
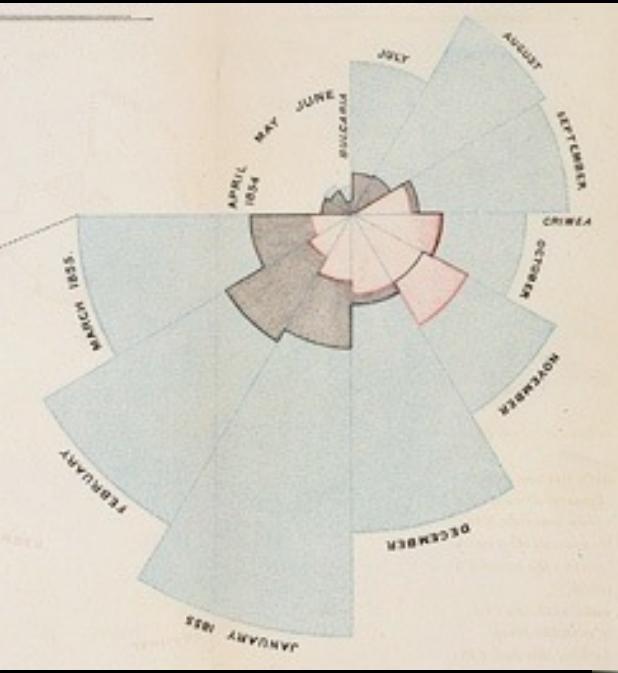


CSE 512 - Data Visualization

The Value of Visualization



Jeffrey Heer University of Washington

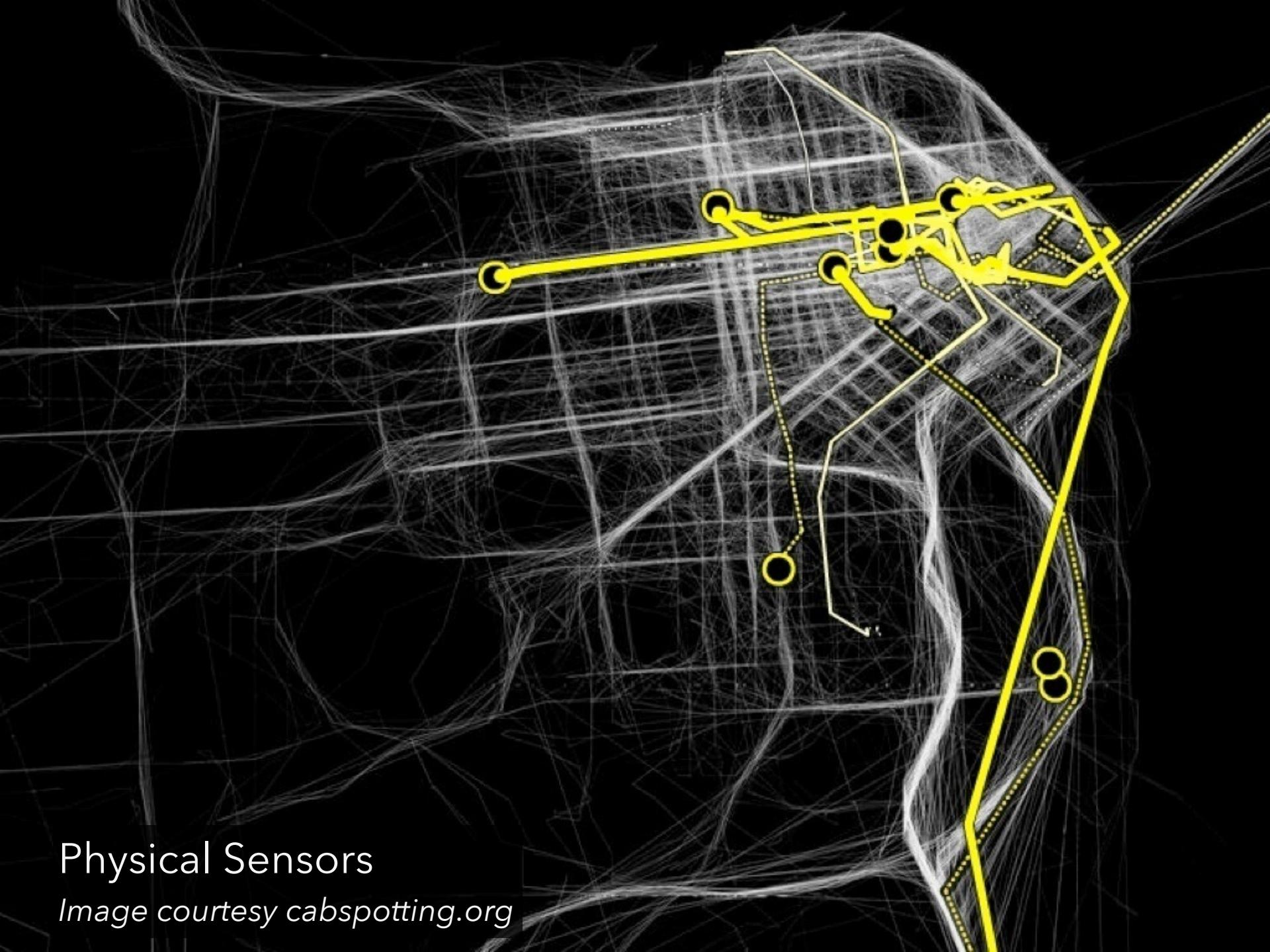
How much data (bytes)
did we produce in 2010?

2010: 1,200 exabytes

Gantz et al, 2008, 2010

2010: 1,200 exabytes
10x increase over 5 years

Gantz et al, 2008, 2010

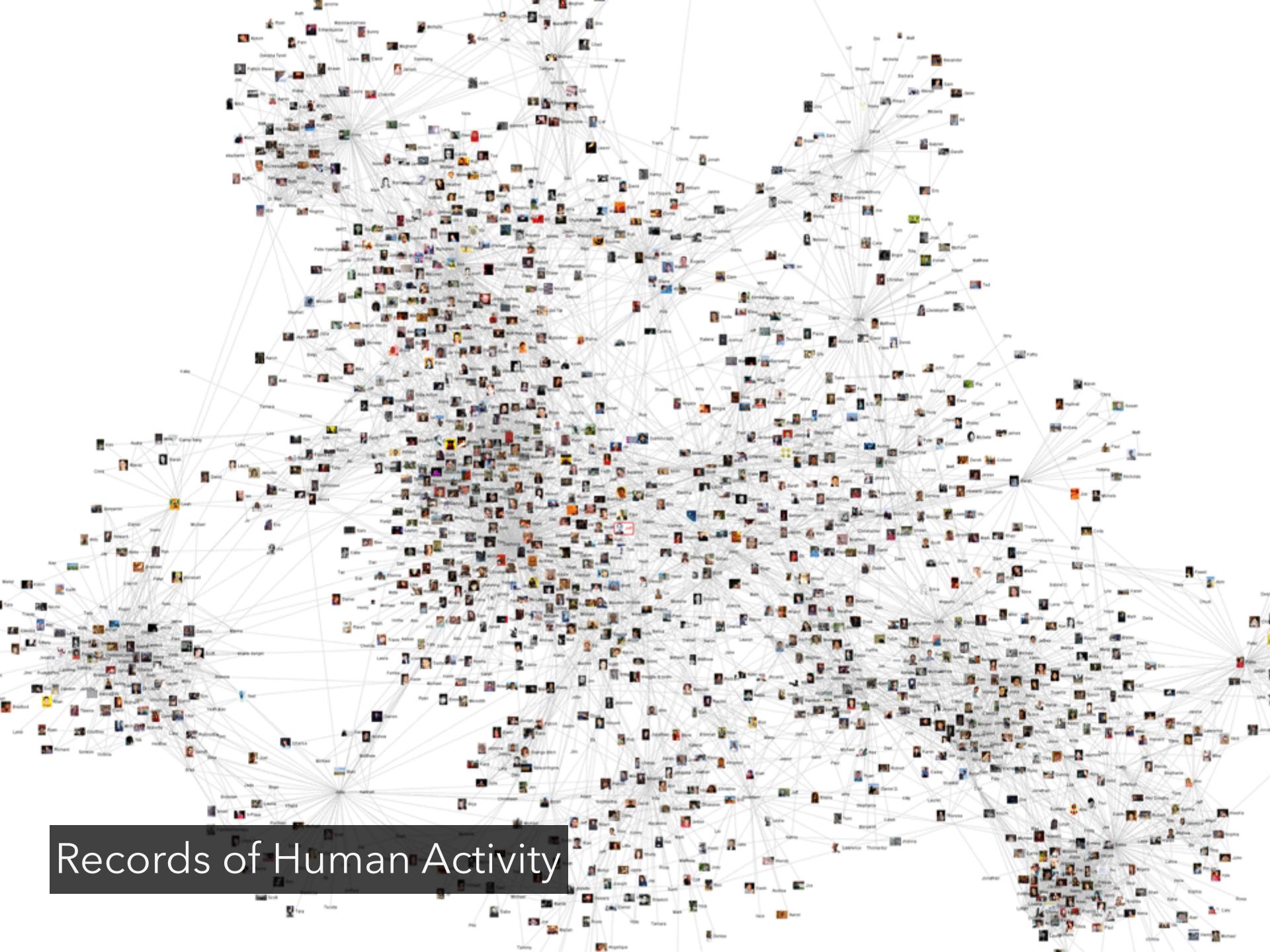


Physical Sensors

Image courtesy cabspotting.org



Records of Human Activity



Abortion

(Revision as of 22:56 4 Jun 2003)

"**Abortion**," in its most commonly used sense, refers to the deliberate early termination of a pregnancy, resulting in the death of the embryo or fetus. [1] Medically, the term also refers to the early termination of a pregnancy by natural ("spontaneous abortion" or "miscarriage," which occurs in 1 in 5 of all pregnancies, usually within the first 12 weeks) or to the cessation of normal growth of a body part or organ. What follows is a discussion of the issues related to deliberate or "induced" abortion.

Methods

Depending on the stage of pregnancy an abortion can be performed by a number of different methods. The earliest terminations (before nine weeks gestation) a chemical abortion is the usual method, though **methotrexate** is usually the only legal method, although research has uncovered similar effects from **methotrexate** and **mifepristone**. Combined with chemical abortion and extending up until around the fifteenth week **suction-aspiration** and **vacuum abortion** is the most common approach, replacing the more risky **dilation and curettage** (D & C). From the fifteenth week up until around the eighteenth week a surgical **dilation and extraction** (D & X) or a **hysterotomy abortion**, similar to a **caesarian section**.

As the fetus size increases other techniques can be used to secure abortion in the third trimester. Premature expulsion of the fetus can be induced with **prostaglandin**, this can be coupled with injecting the amniotic fluid with saline or urea solution. Very late abortions can be brought about by the controversial **intact dilation and extraction** (I & X) or a **hysterotomy abortion**, similar to a **caesarian section**.

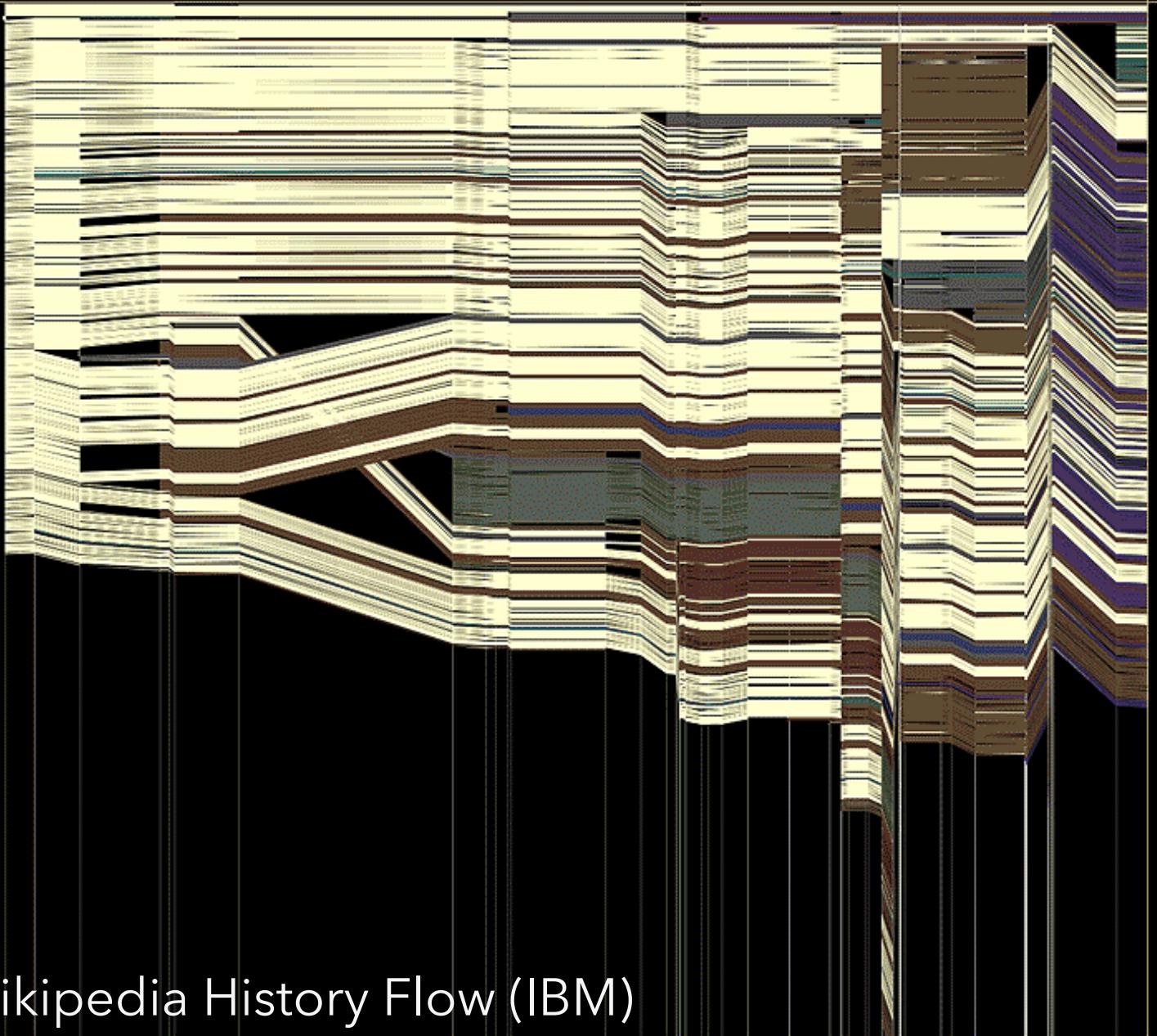
The controversy

The morality and legality of abortion is a highly important topic in **applied ethics**, and is also discussed by **legal scholars** and **religious groups**. Important facts about abortion are also reported by **sociologists** and **historians**.

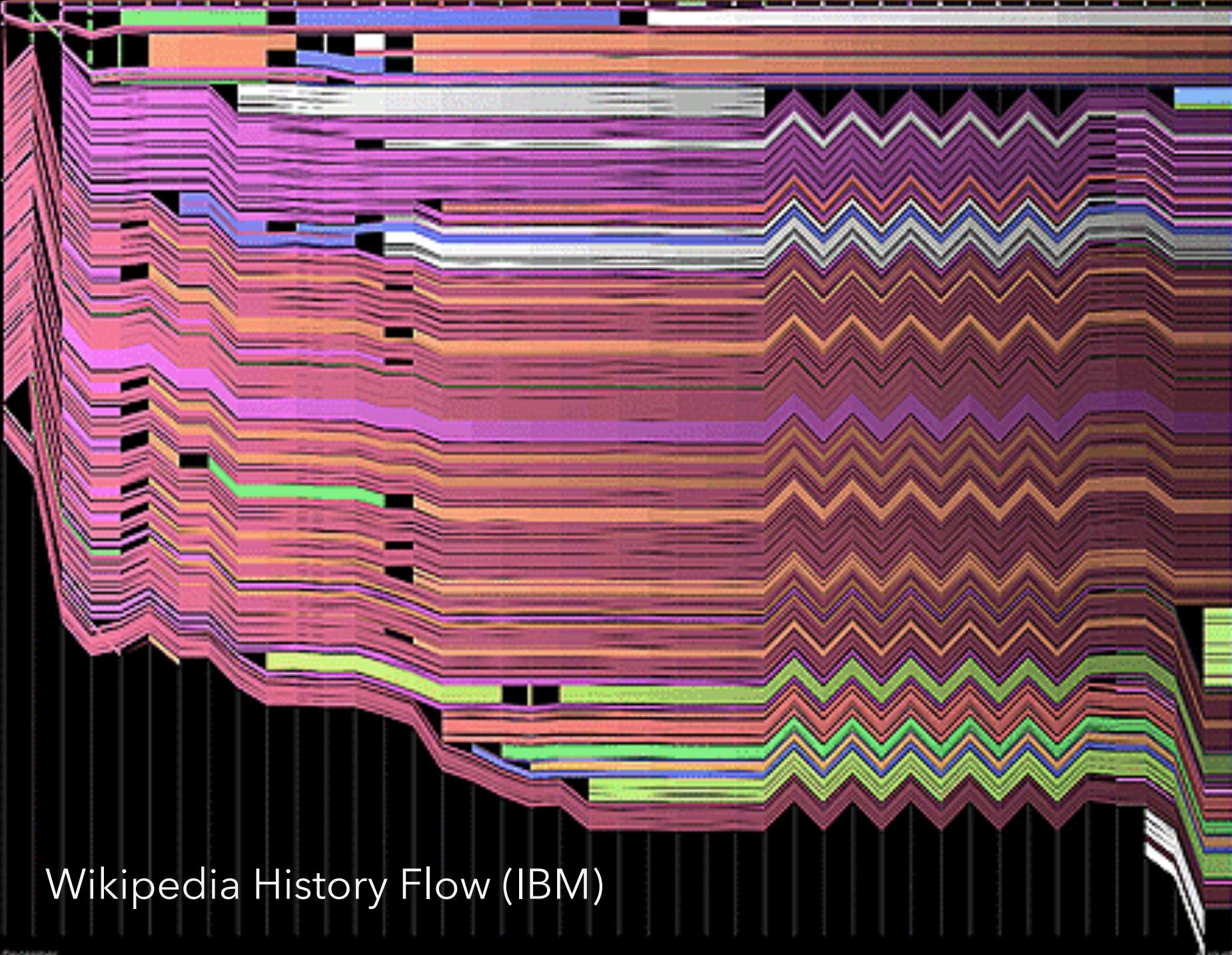
Abortion has been common in most societies, although it has often been opposed by some institutionalized religions and governments. In the **United States** and **Europe**, abortion became commonly accepted by the end of the 20th century. Additionally, abortion is accepted in **China**, **India** and other populous countries. The **Catholic Church** remains opposed to the procedure, however, and in other countries, notably the **United States** and the (predominantly Catholic) **Republic of Ireland**, the controversy is extremely active, to the extent that even the respective positions are subject to heated debate. While those on both sides of the issue are generally peaceful, if heated, in their expression of their positions, the debate is sometimes characterized by violence. Though true of both sides, this is more marked on the side of those who are opposed to abortion, because of what they see as the gravity and urgency of their views.

The central question

The central question in the abortion debate is the clash of presumed or perceived rights. On one hand, is a fetus (sometimes called the "unborn") a person, or is it a "pro-life/anti-abortion advocate's" a human being with a right to life, and if so, at what point in pregnancy does the fetus become human? On the other hand, is a fetus part of a woman's body?



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

$$\mu_X = 9.0 \quad \sigma_X = 3.317$$

$$\mu_Y = 7.5 \quad \sigma_Y = 2.03$$

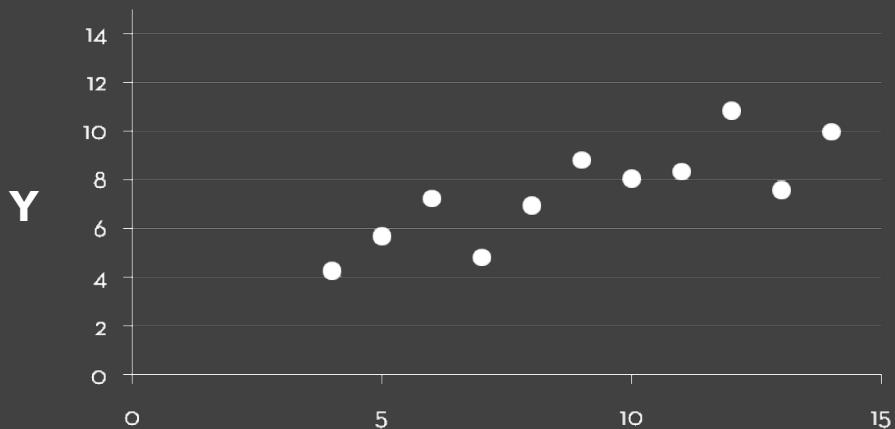
Linear Regression

$$Y = 3 + 0.5 X$$

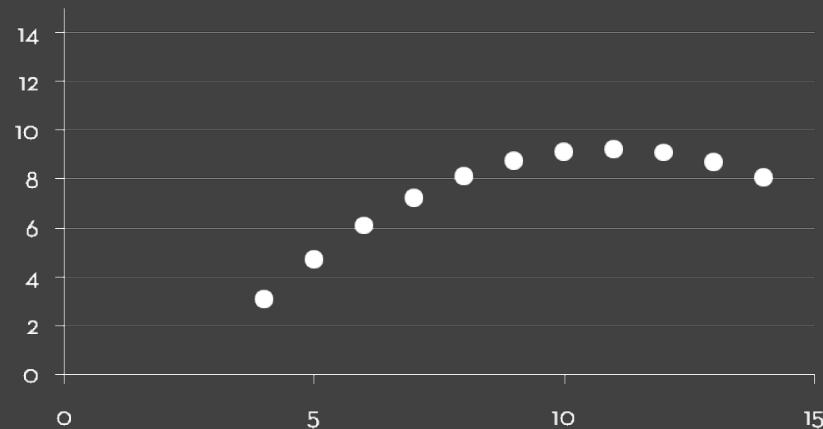
$$R^2 = 0.67$$

[Anscombe 1973]

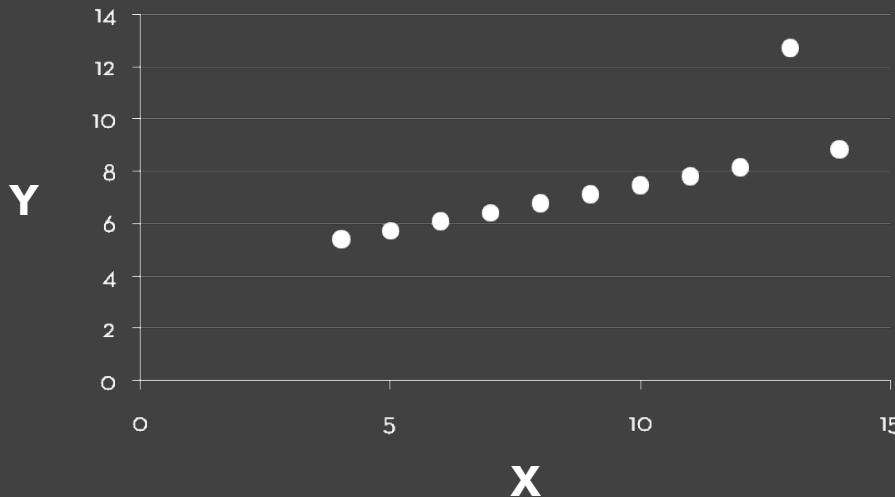
Set A



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

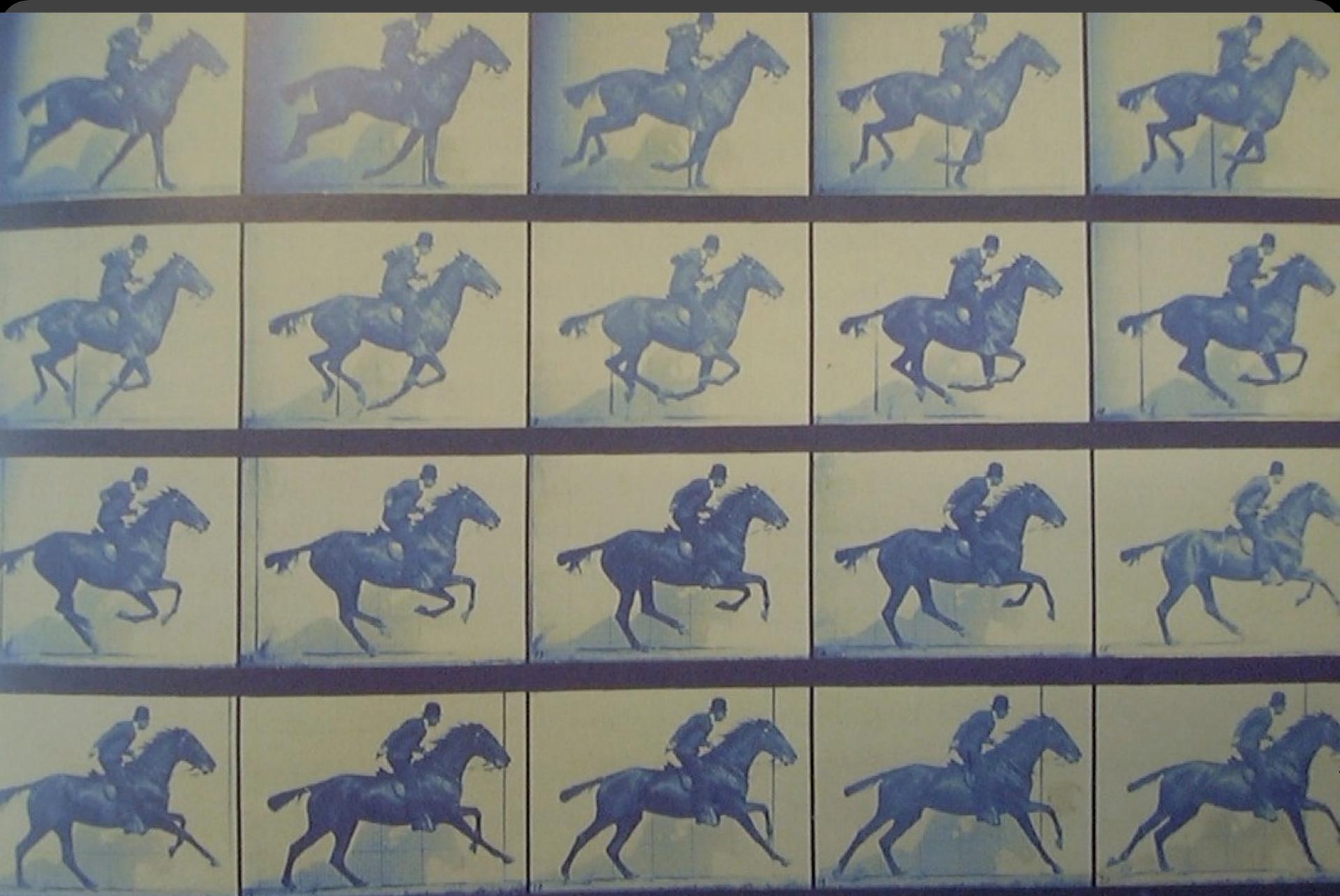
Support graphical calculation

Find patterns

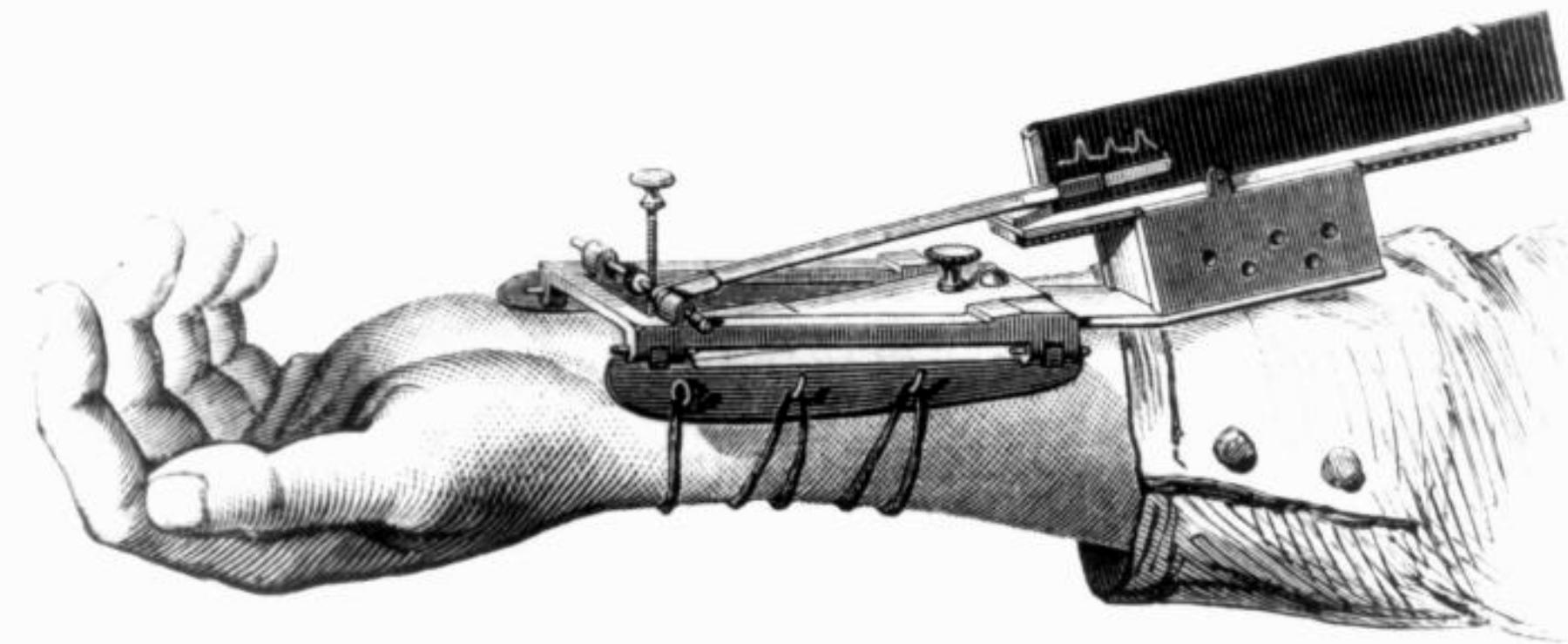
Present argument or tell a story

Inspire

Record Information



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

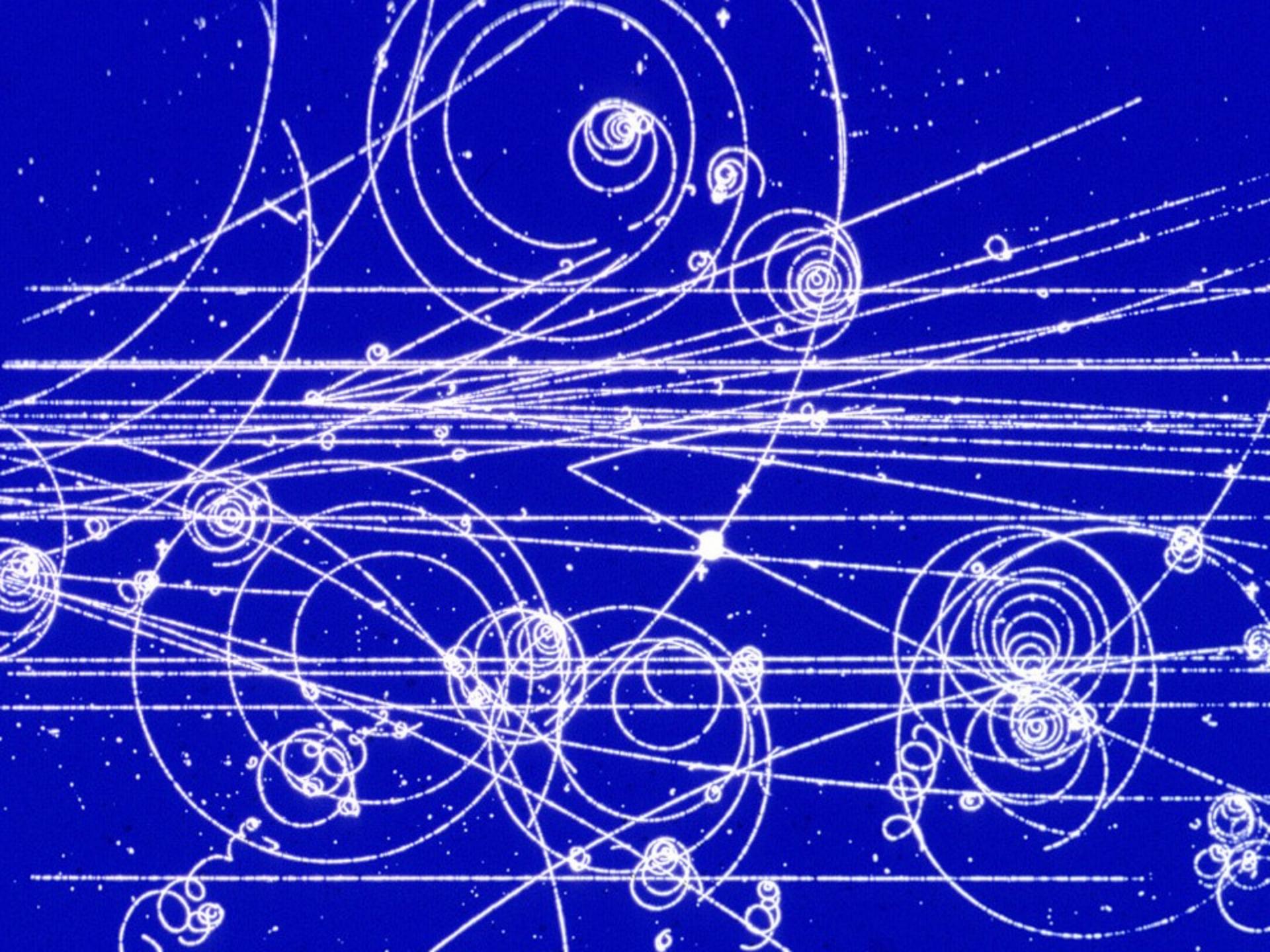


1.

Marey's **sphygmograph** in use.

1860. *La méthode graphique dans les sciences expérimentales et principalement en physiologie et en médecine.*

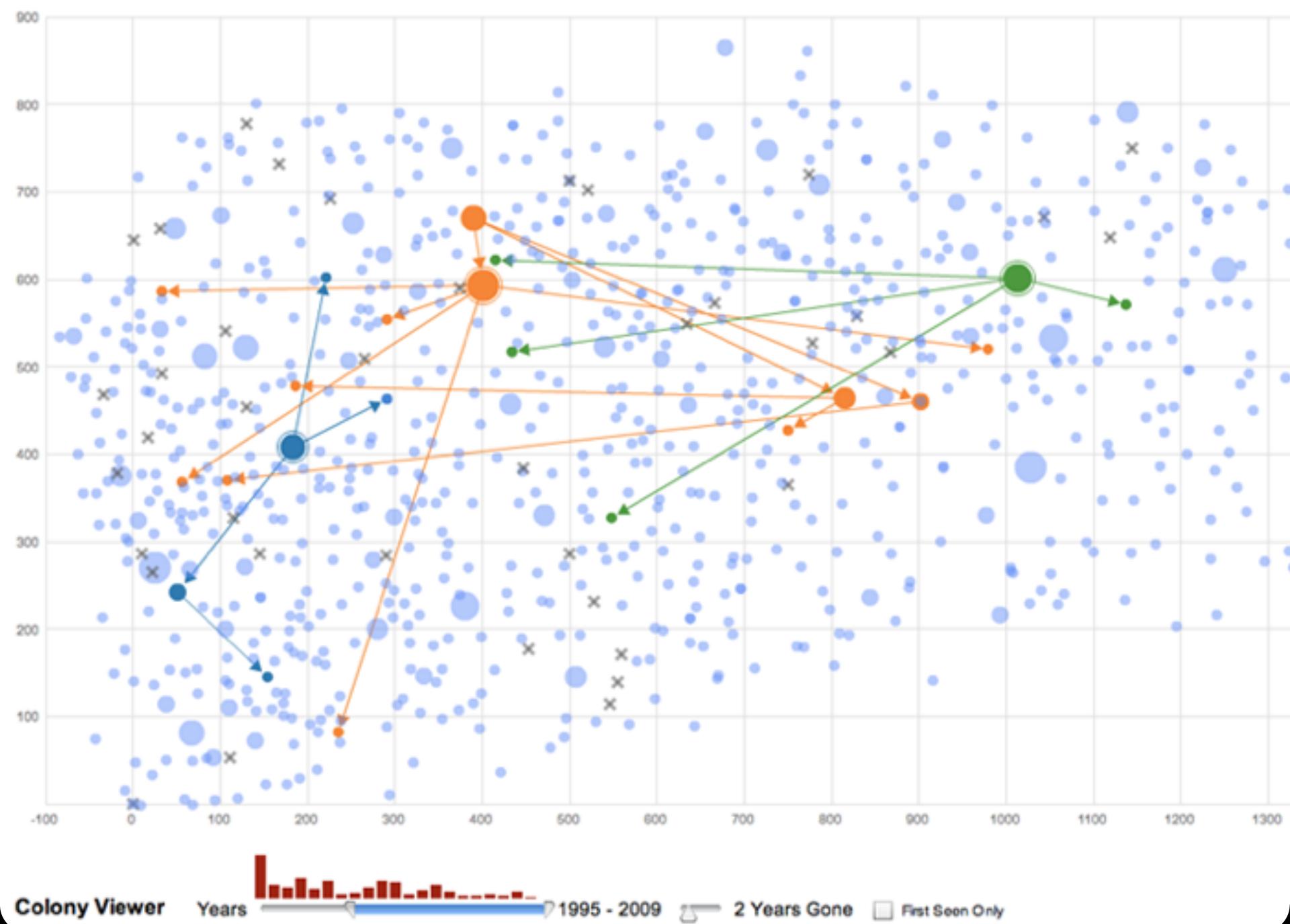
E.J. Marey's sphygmograph [from Braun 83]





Expected live in 06			
841	942	462	20
858	656	580	787
962	903	253	806
867	337	223	740
330	666	545	519
635	750	252	55289
12	46	497	948
47	532	341	303
4849	634	3341	181
6	785	577	786
300	469	728	722
280	732	786	7949
35-69	74	786	35-69
695	871	736	734
275125	940	735	577
37	729	919	924
168	347	436	957
739770	769	912	993
549	890	911	924
848	941	839	997
587	32	895	43594
43	273	804	792
945	940	810	313
621	729	934	995
683	171	934	53
493	98	63	758
870	99	699	310
495	18	64	432
553	628834	24	965
562	590	221	920
963	887	686	631
482	651	686	536
520	939	783	6641
801	590	742	537
908	600	737	68
36	600	600	999
30	737	538	996
576	62	61	645
567	3	61	763
594	61	61	482
415	61	61	520
901	61	61	801

Colony Viewer - colony.observations.txt | colony.ancestry.txt



Support Reasoning

HISTORY OF O-RING DAMAGE ON SRM FIELD JOINTS

	SRM No.	Cross Sectional View			Top View		Clocking Location (deg)
		Erosion Depth (in.)	Perimeter Affected (deg)	Nominal Dia. (in.)	Length Of Max Erosion (in.)	Total Heat Affected Length (in.)	
0	61A LH Center Field**	22A	None	0.280	None	None	36°--66°
	61A LH CENTER FIELD**	22A	NONE	0.280	NONE	NONE	338°-18°
0	51C LH Forward Field**	15A	0.010	154.0	0.280	4.25	163
	51C RH Center Field (prim)***	15B	0.038	130.0	0.280	12.50	354
y	51C RH Center Field (sec)***	15B	None	45.0	0.280	None	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°)

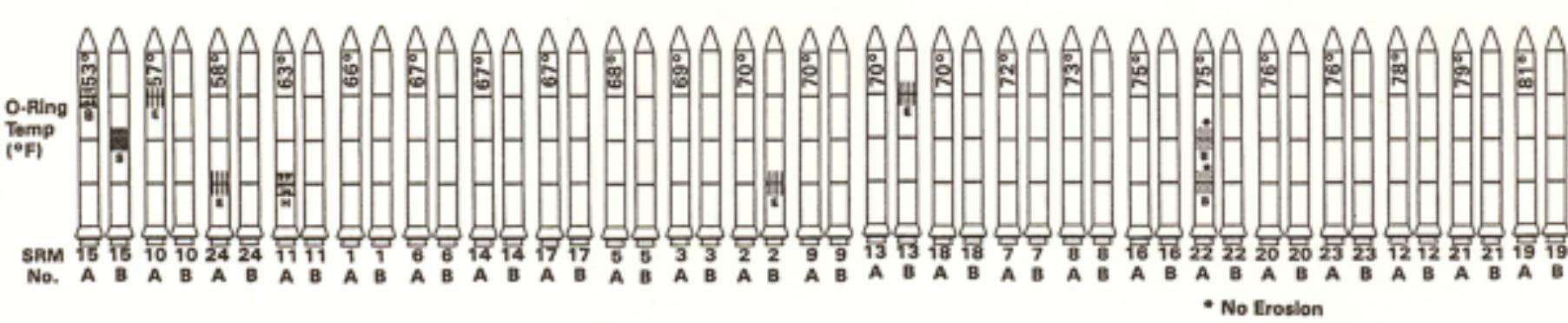
SRM-13A, 15, 16A, 18, 23A 24A

- NOZZLE Blow-by

HISTORY OF O-RING TEMPERATURES (DEGREES - F)

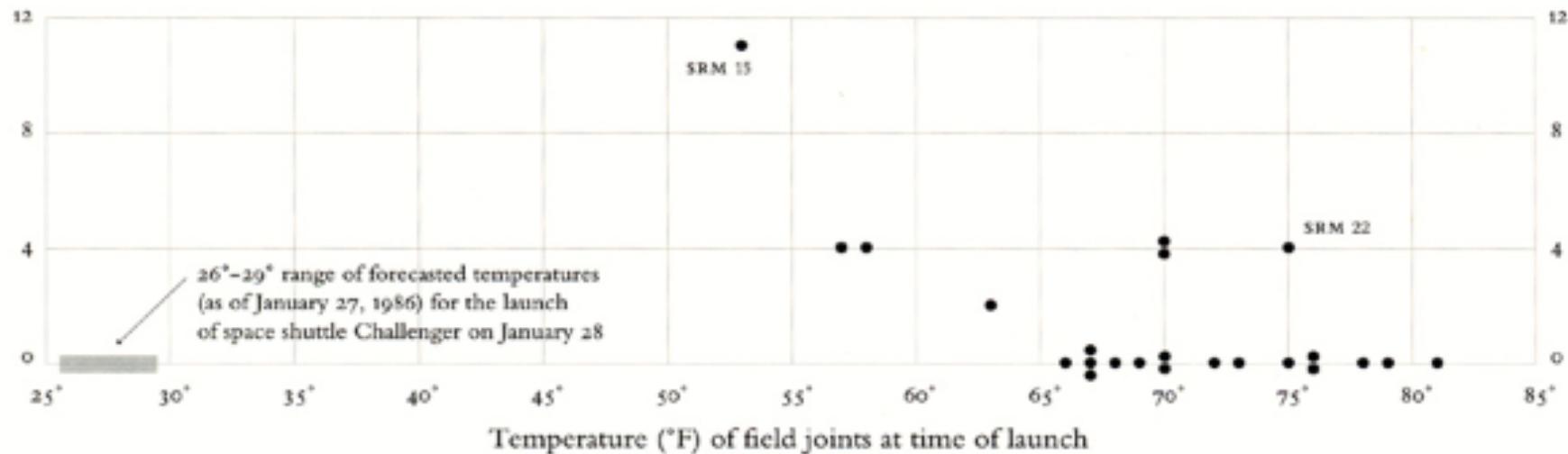
MOTOR	MBT	AMB	O-RING	WIND
DM-4	68	36	47	10 MPH
DM-2	76	45	52	10 MPH
QM-3	72.5	40	48	10 MPH
QM-4	76	48	51	10 MPH
SRM-15	52	64	53	10 MPH
SRM-22	77	78	75	10 MPH
SRM-25	55	26	29	10 MPH
			27	25 MPH

Make a Decision: Challenger



Make a Decision: Challenger

O-ring damage
index, each launch



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

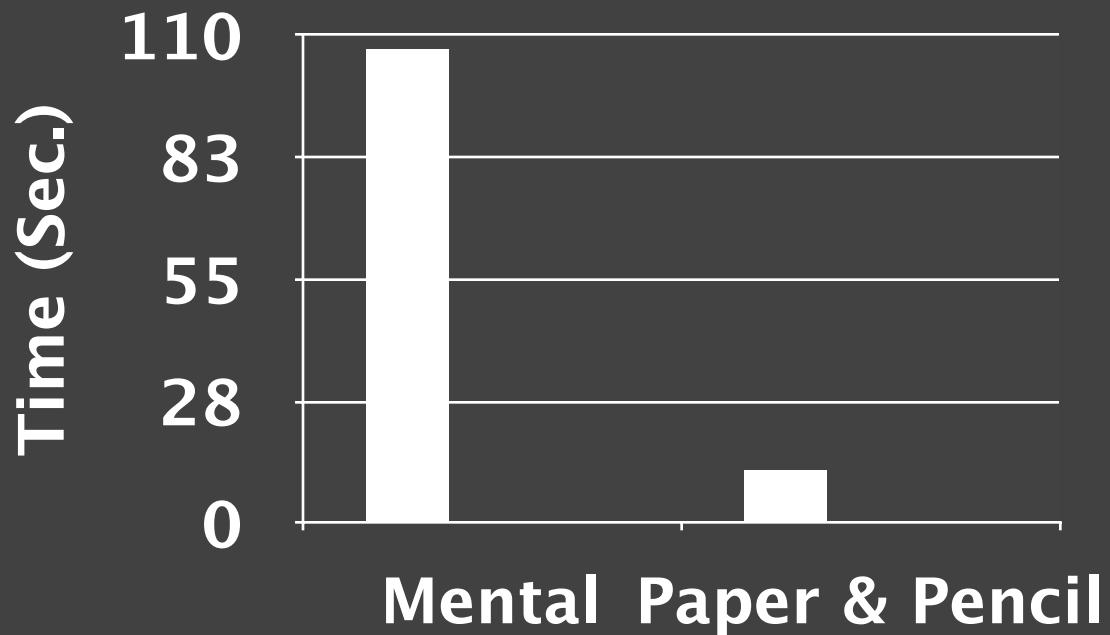
Class Exercise

Expand Memory: Multiplication

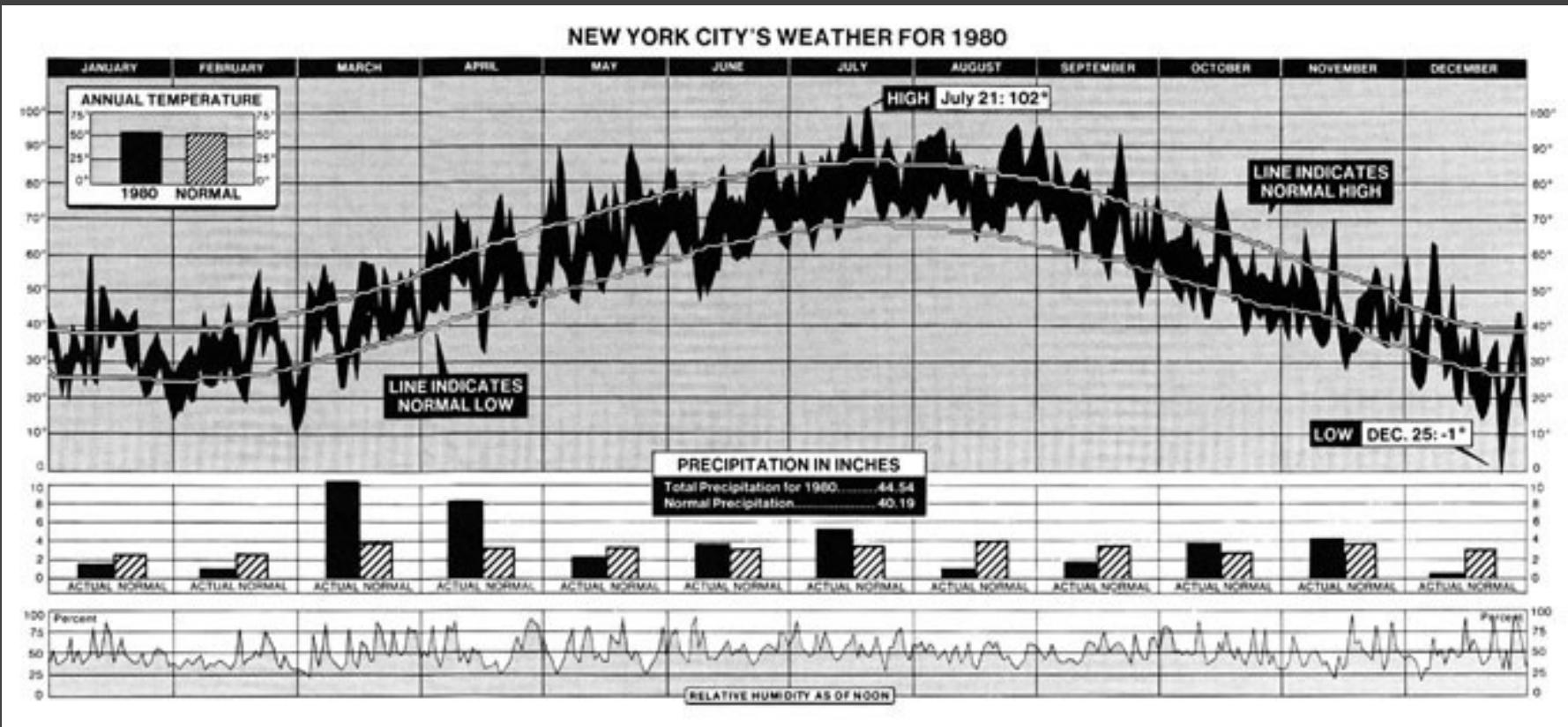
$$\begin{array}{r} 34 \\ \times 72 \\ \hline \end{array}$$

Expand Memory: Multiplication

$$\begin{array}{r} 34 \\ \times 72 \\ \hline 68 \\ 2380 \\ \hline 2448 \end{array}$$

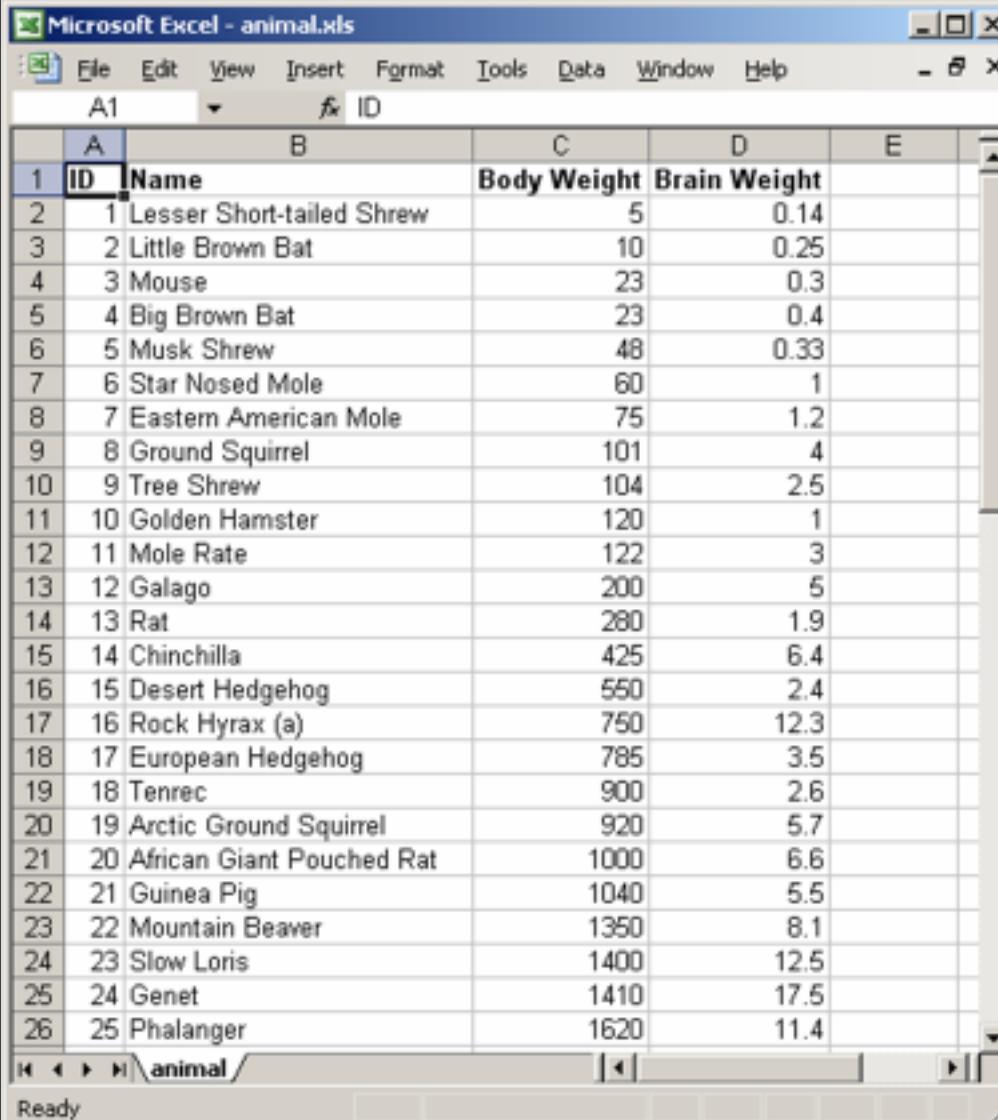


Find Patterns: NYC Weather



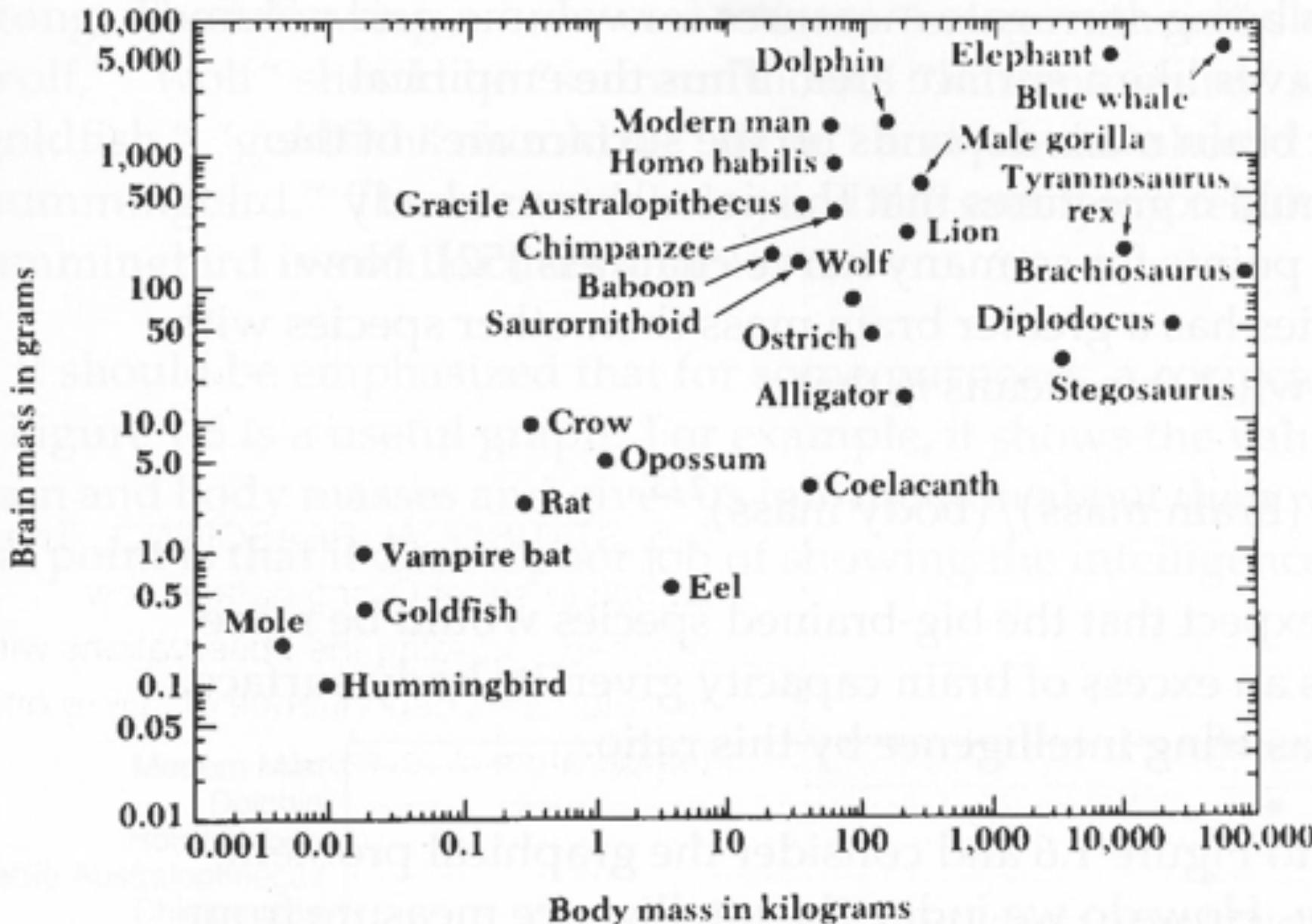
[New York Times 1981]

The Most Powerful Brain?

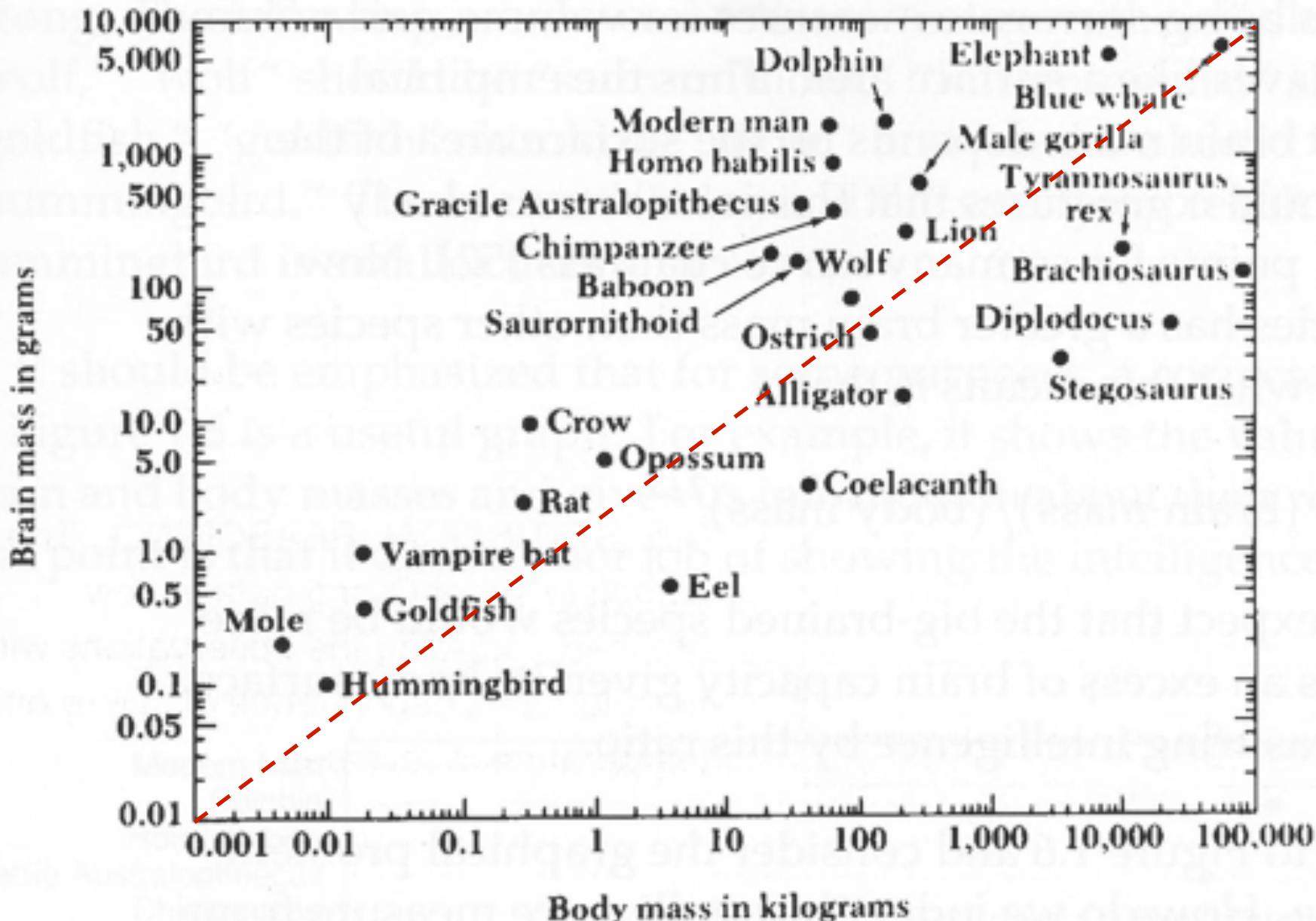


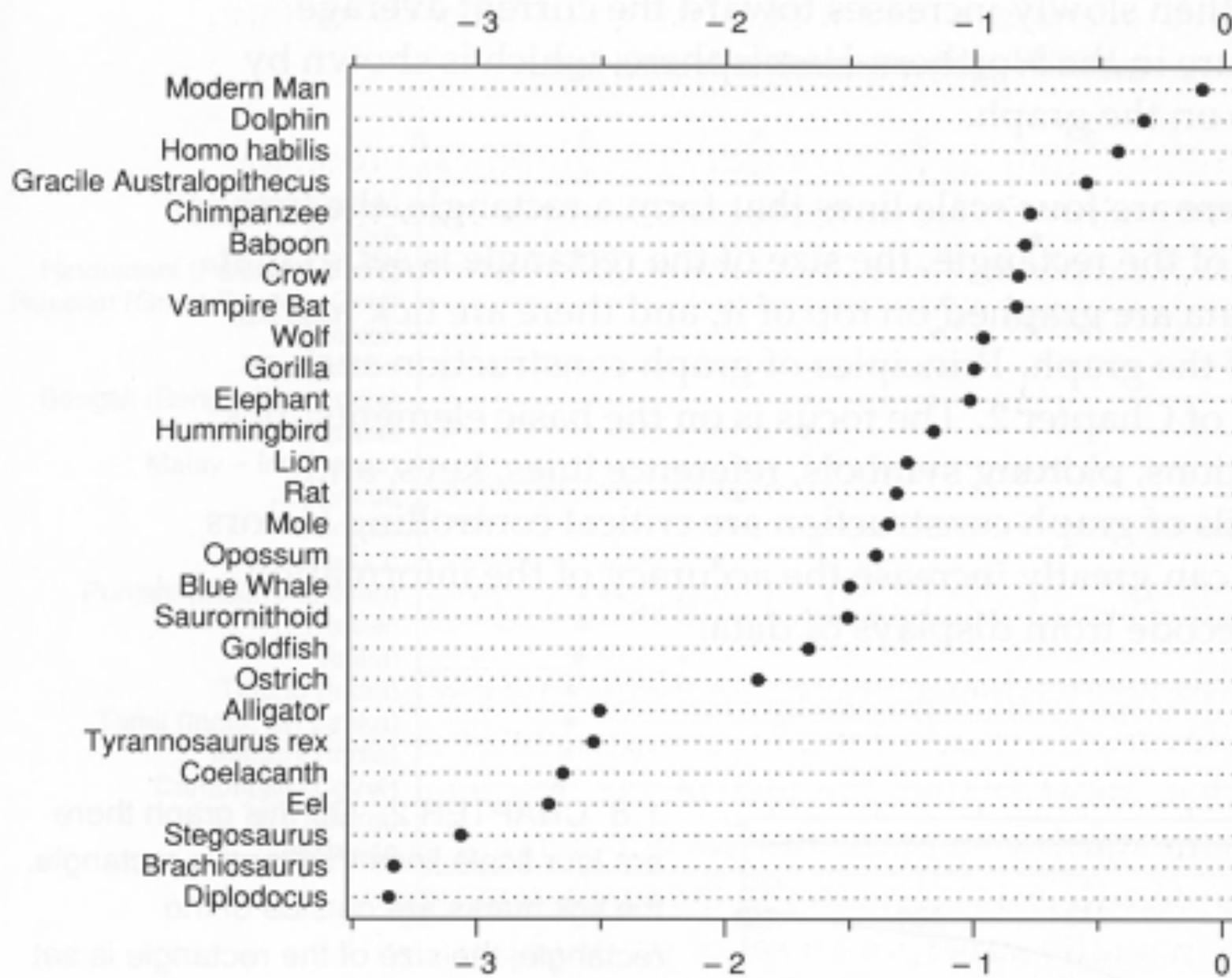
A screenshot of Microsoft Excel version 2003 displaying a spreadsheet titled "animal.xls". The window title bar reads "Microsoft Excel - animal.xls". The menu bar includes File, Edit, View, Insert, Format, Tools, Data, Window, and Help. The ribbon tabs are visible at the top right. The active cell is A1, and the formula bar shows the formula for cell ID. The spreadsheet contains data for 26 animals, organized into columns A, B, C, D, and E. Column A lists IDs from 1 to 26. Column B lists the animal names. Column C lists body weight in grams. Column D lists brain weight in grams. The data shows a clear positive correlation between body size and brain size.

	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 Rat	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	



The Dragons of Eden [Carl Sagan]





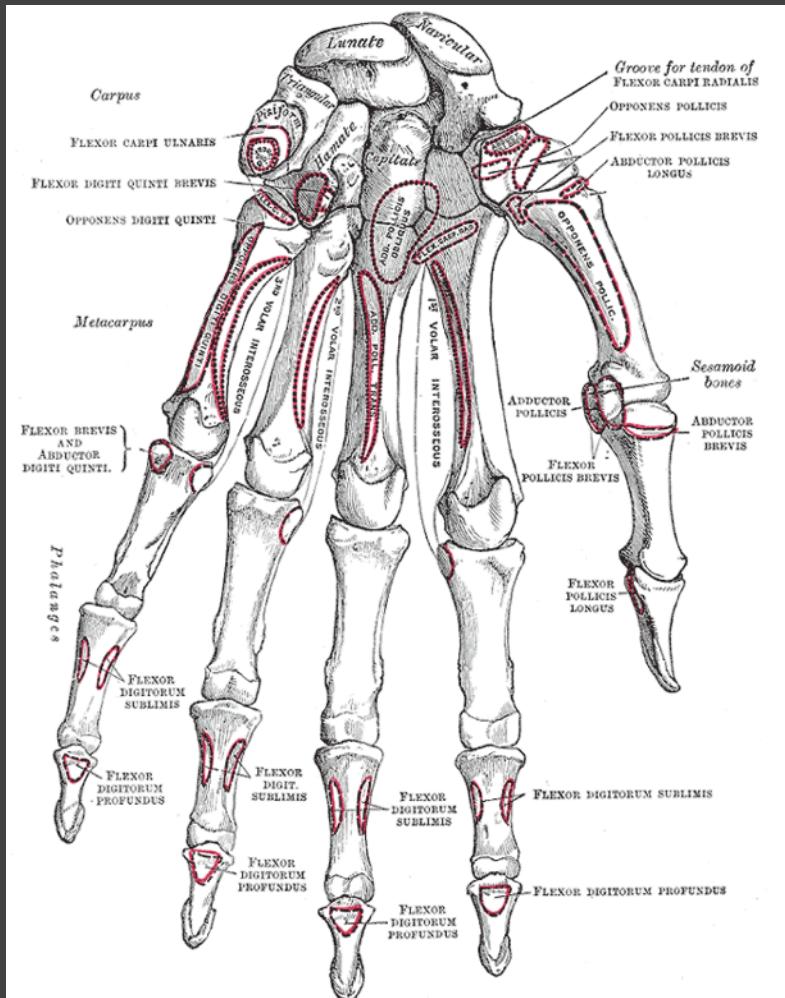
The Elements of Graphing Data

[Cleveland]

Log₁₀ Brain Weight – $\frac{2}{3}$ Log₁₀ Body Weight

**Convey Information
to Others**

Inspire



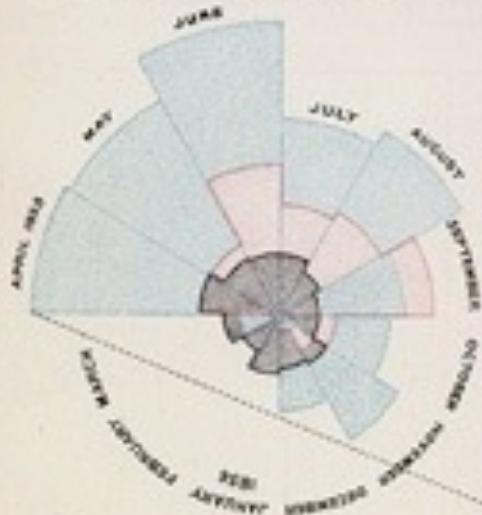
Bones in hand [from 1918 edition]



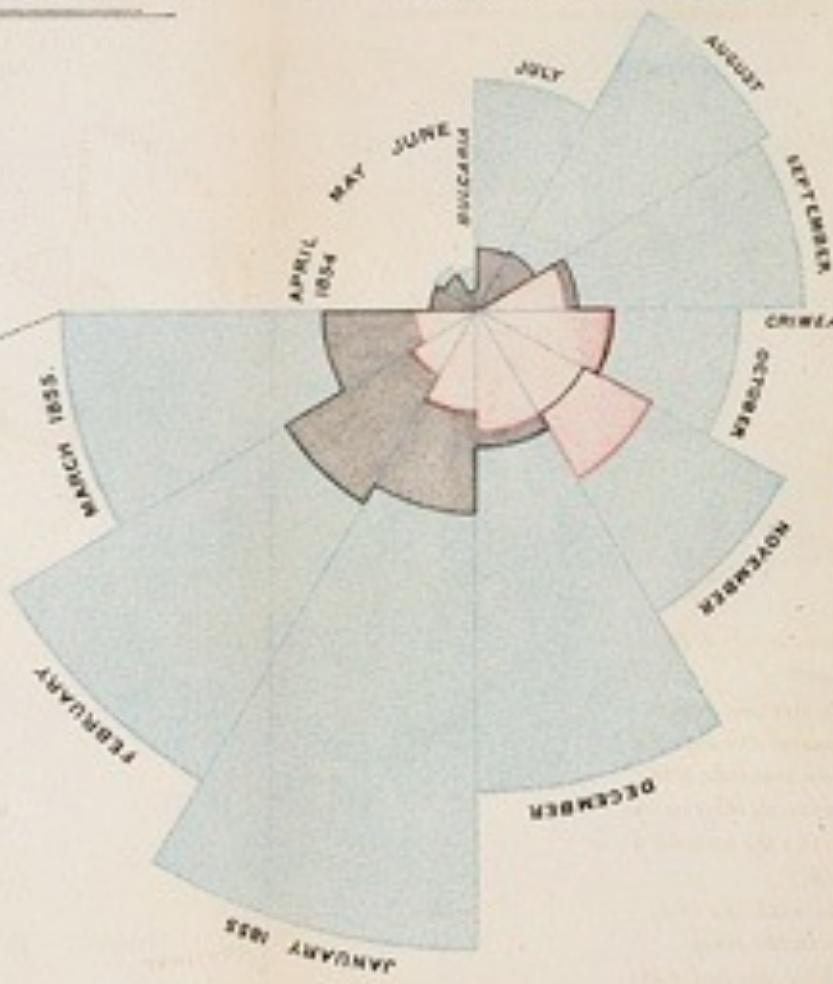
Double helix model [Watson and Crick 53]

DIAGRAM OF THE CAUSES OF MORTALITY
IN THE ARMY IN THE EAST.

APRIL 1855 TO MARCH 1856.



APRIL 1854 TO MARCH 1855.



The areas of the blue, red, & black wedges are each measured from the centre as the common vertex.

The blue wedges measured from the centre of the circle represent area for area the deaths from Preventable or Meligable Zymotic diseases, the red wedges measured from the centre the deaths from wounds, & the black wedges measured from the centre the deaths from all other causes. The black line across the red triangle in No. 1634 marks the boundary of the deaths from all other causes during the month.

In October 1854, & April 1855, the black area coincides with the red;

in January & February 1855, the blue coincides with the black.

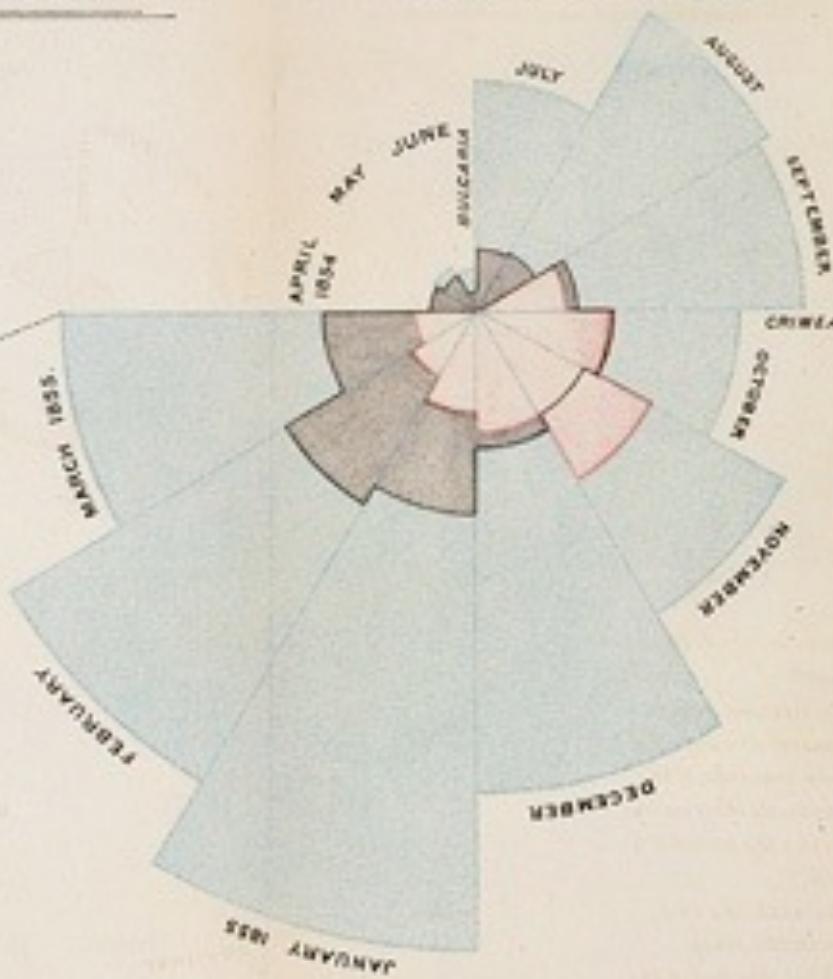
The entire areas may be compared by following the blue, the red & the black lines enclosing them.

2.
APRIL 1855 TO MARCH 1856.

DIAGRAM OF THE CAUSES OF MORTALITY
IN THE ARMY IN THE EAST.



1.
APRIL 1854 TO MARCH 1855.



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

© National Gallery of Victoria

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

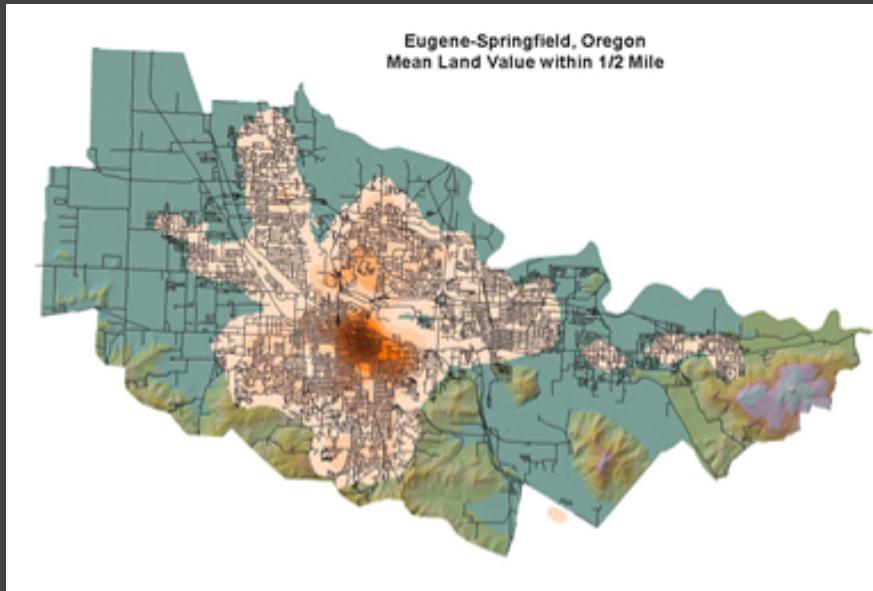
Challenge

More and more unseen data

Faster creation and collection

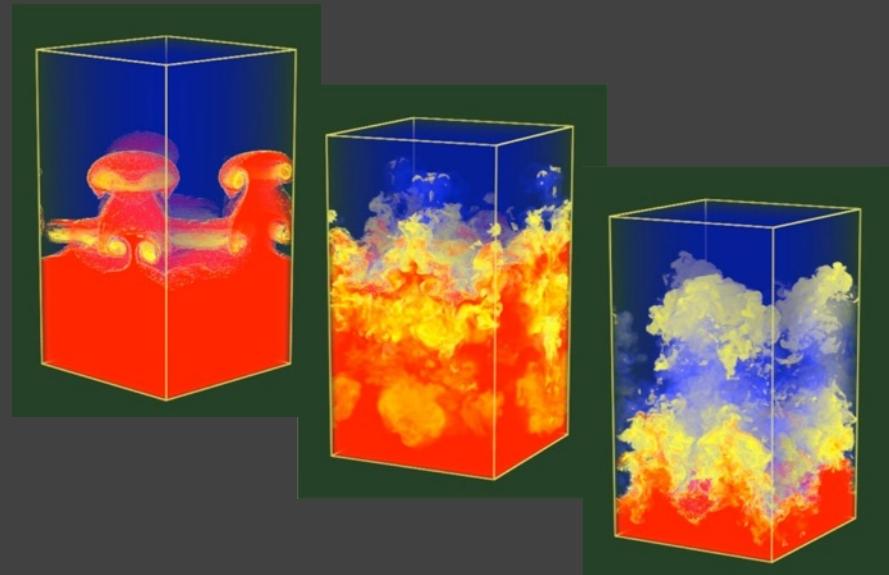
Challenge

More and more unseen data
Faster creation and collection



Urban development planning
www.urbansim.org

Simulation



ctr.stanford.edu

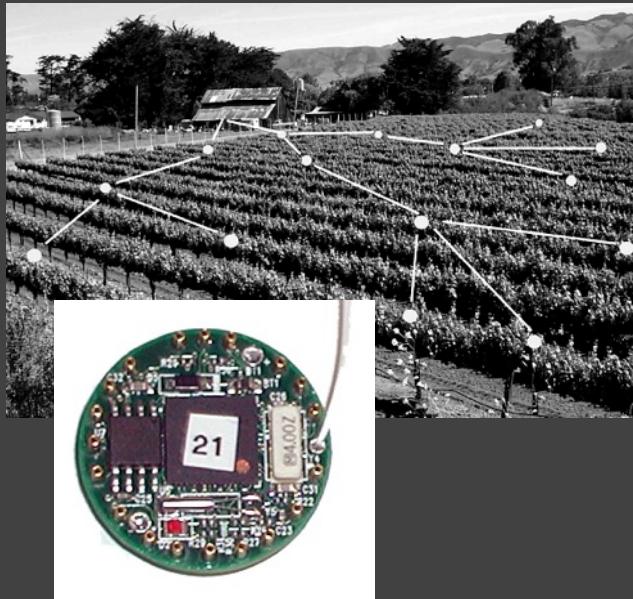
Challenge

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

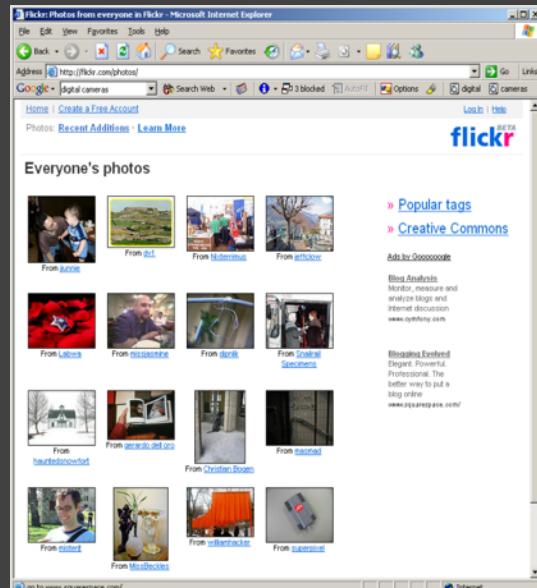
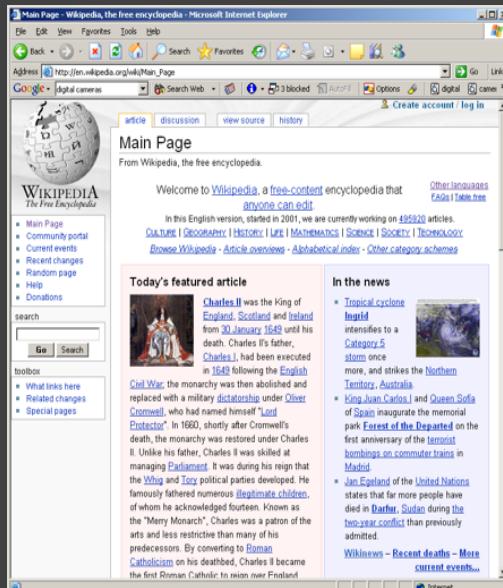
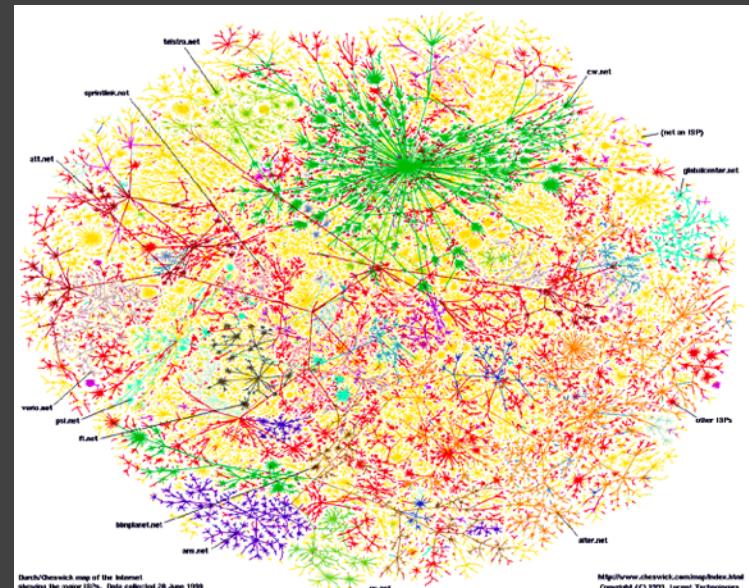


Photo sharing/annotation
flickr.com



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

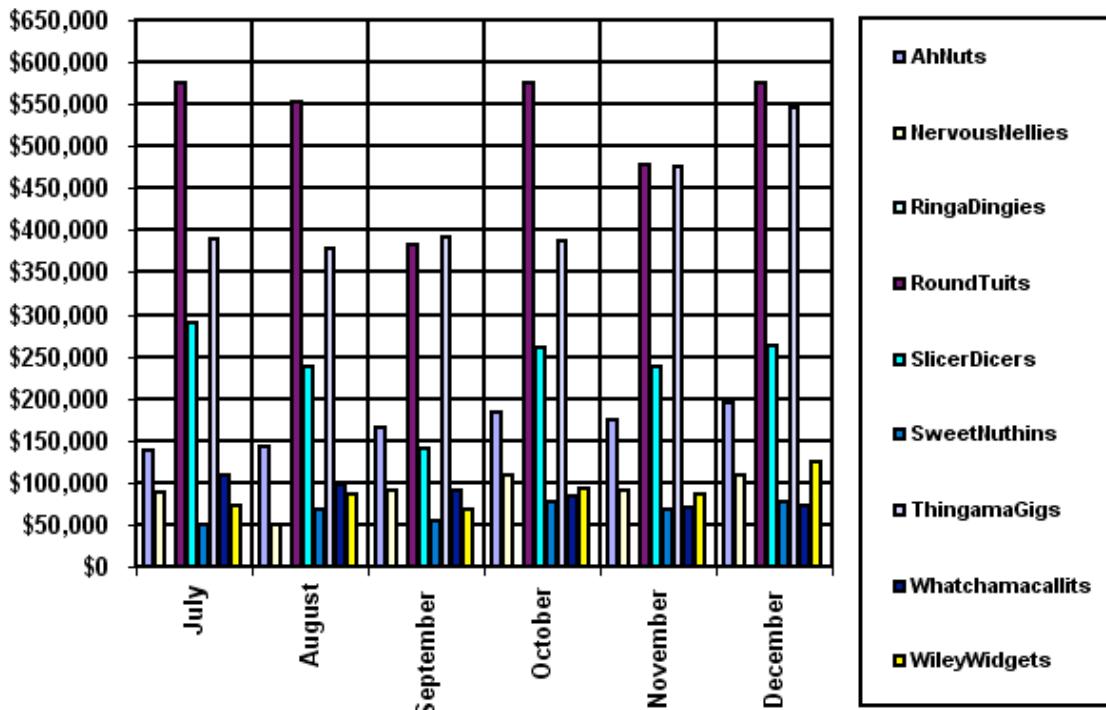
Course Topics

Data and Image Models

LES VARIABLES DE L'IMAGE								12	14
XY 2 DIMENSIONS DU PLAN	POINTS								
Z TAILLE									
VALEUR									
LES VARIABLES DE SÉPARATION DES IMAGES								13	
GRAIN									
COULEUR									
ORIENTATION									

Visualization (Re-)Design

SlicerDicers' Sales Compared to Other Products

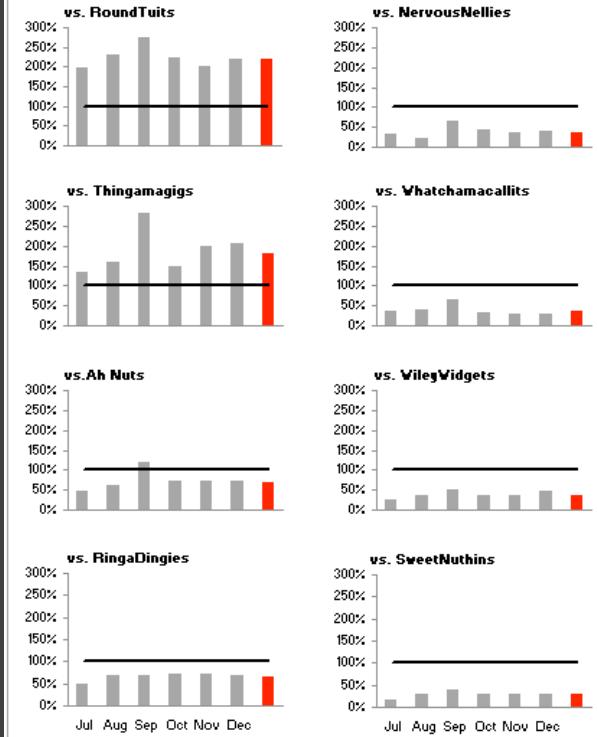


Problematic design

Sales of SlicerDicers Compared to Other Products

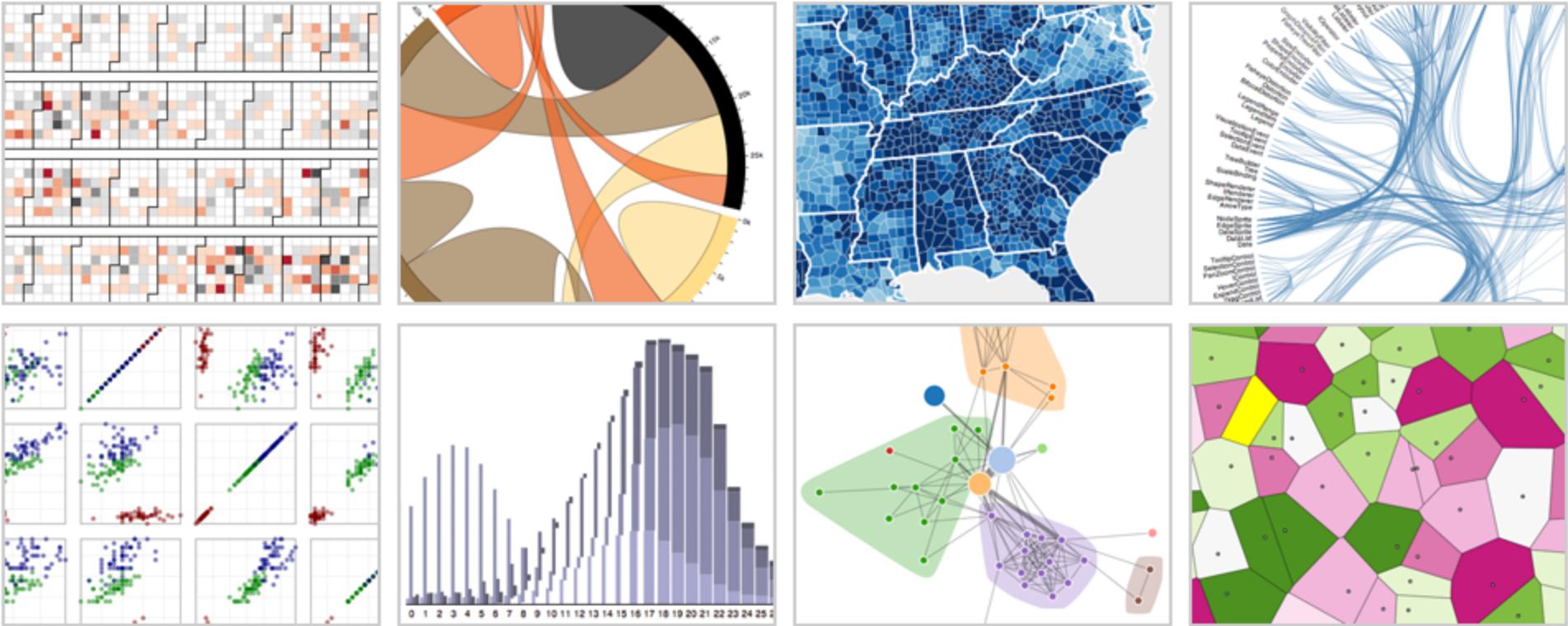
July - December, 2003

(SlicerDicers' sales are displayed as black reference lines of 100%; the red bars represent the average monthly sales percentage for July through December.)



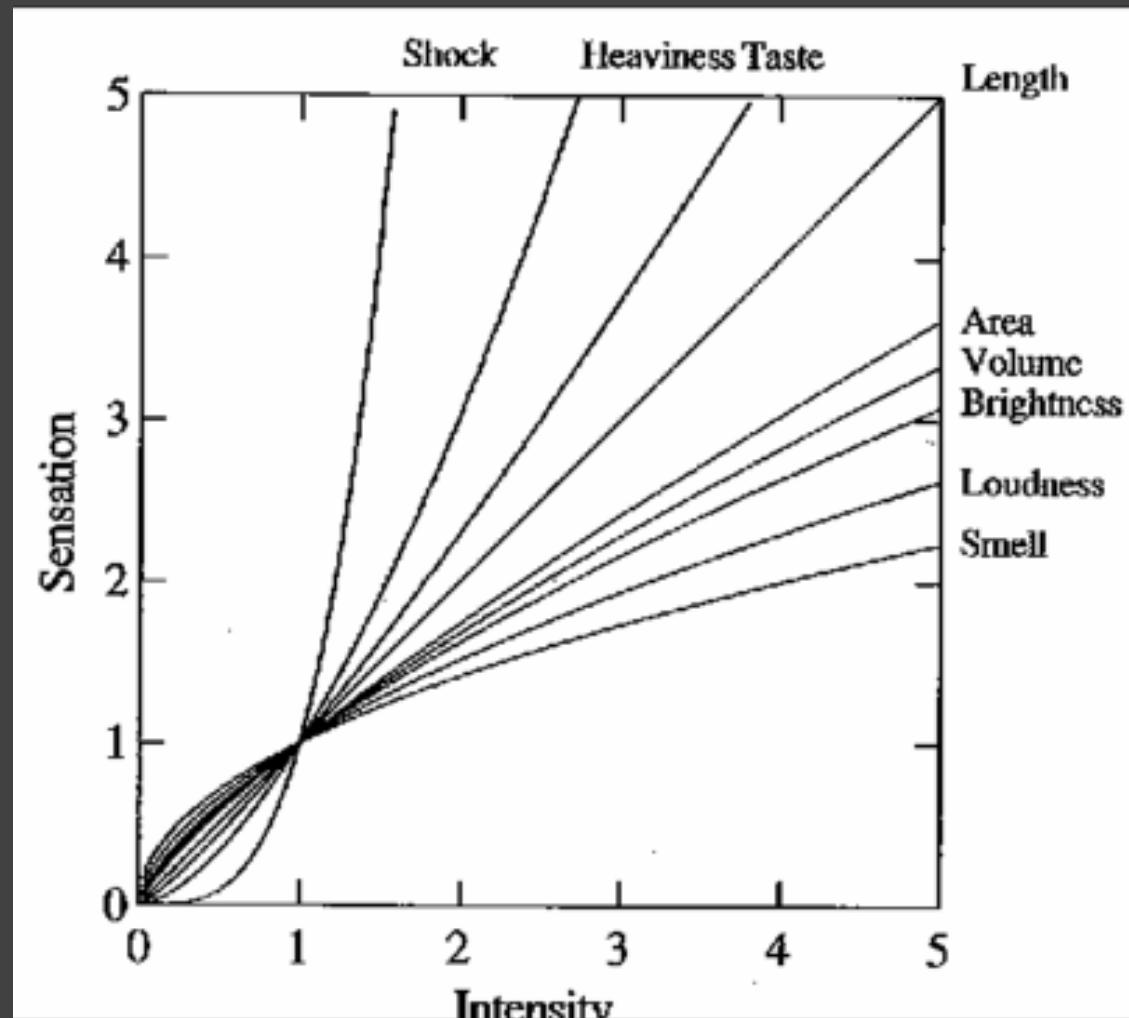
Redesign

Visualization Software



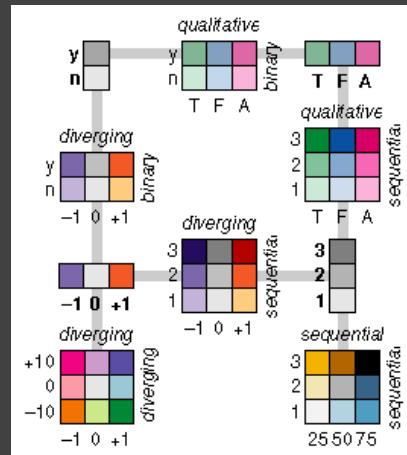
D3: Data-Driven Documents

Graphical Perception

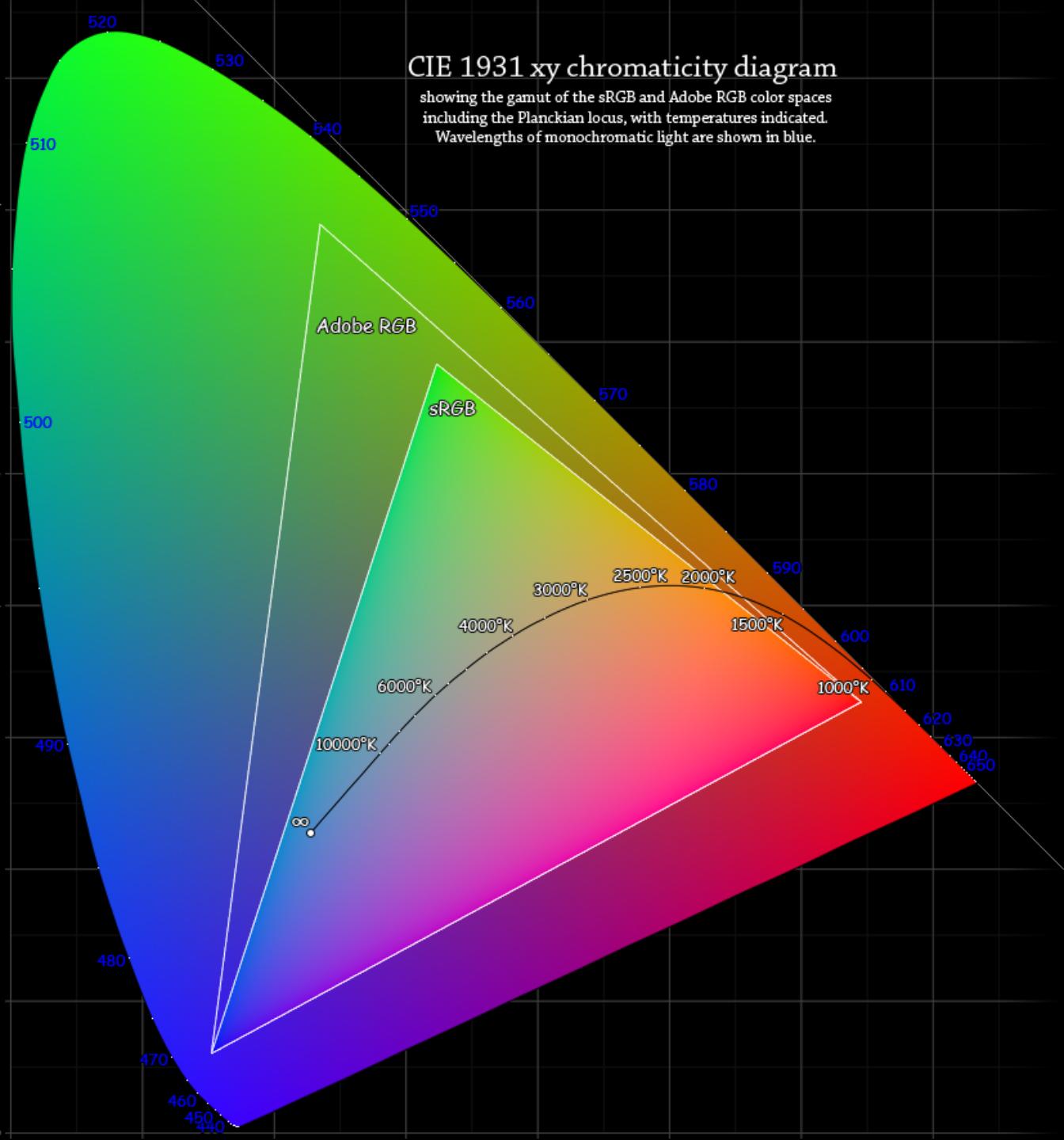


The psychophysics of sensory function [Stevens 61]

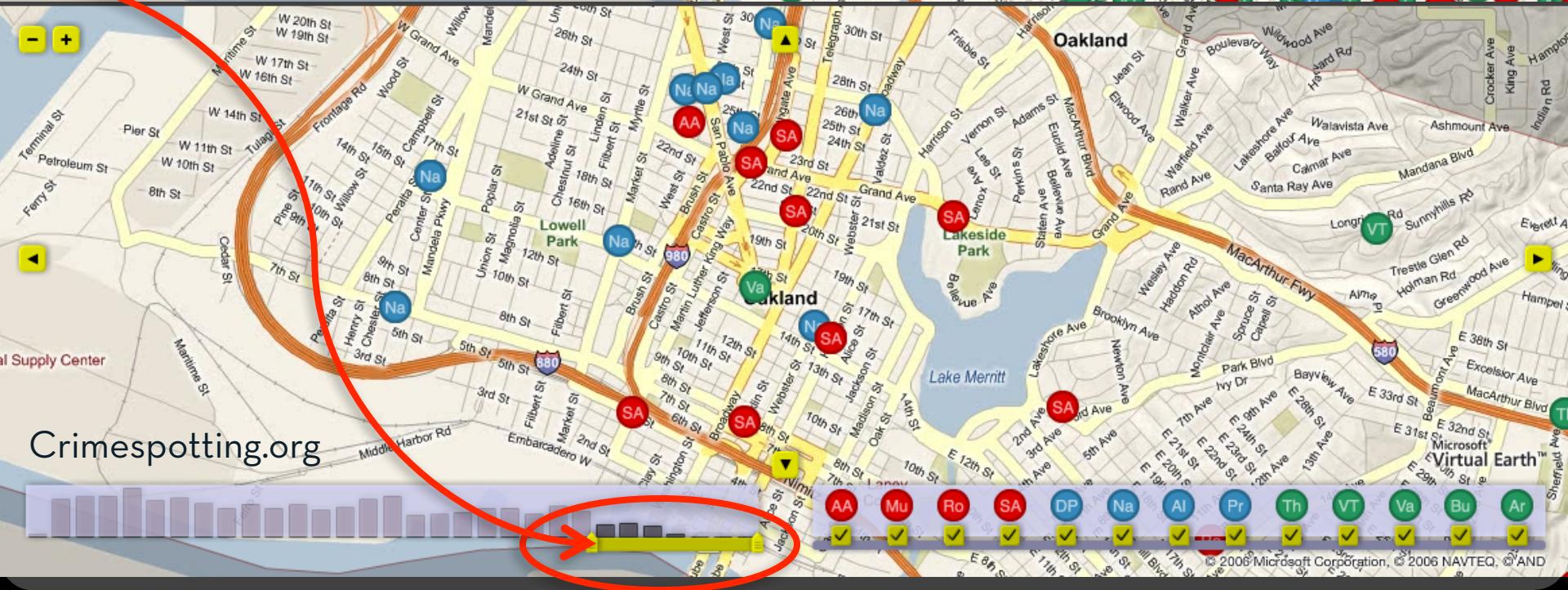
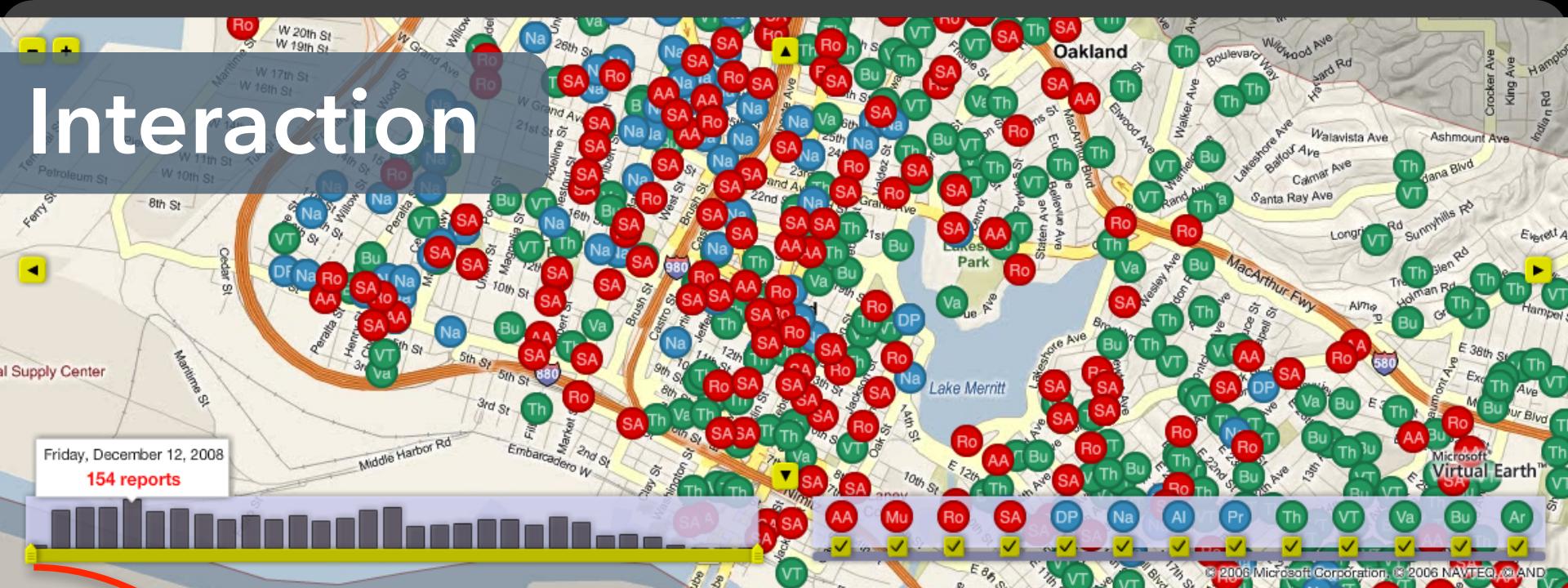
Color



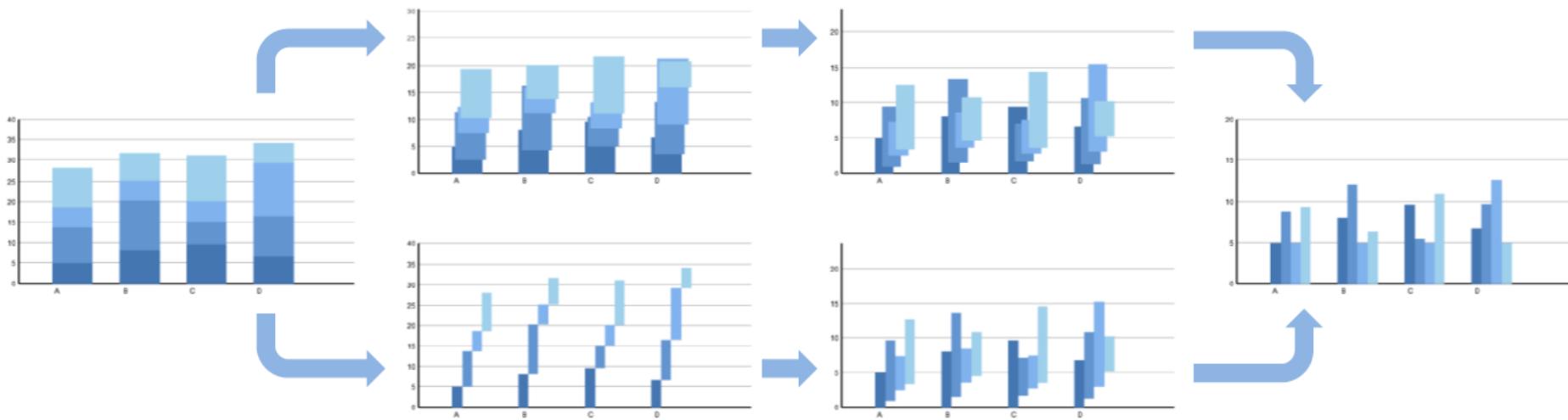
Color Brewer



Interaction

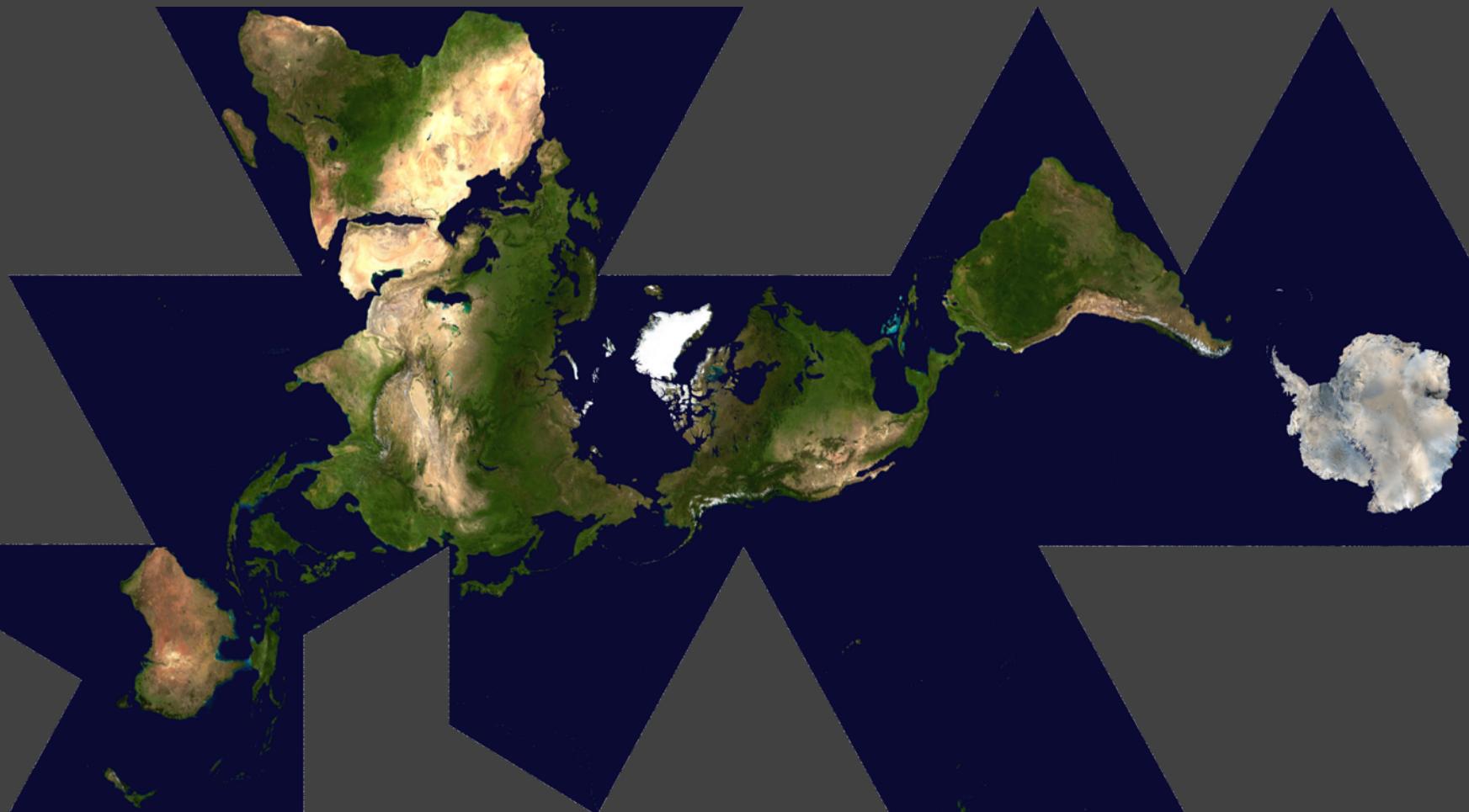


Animation



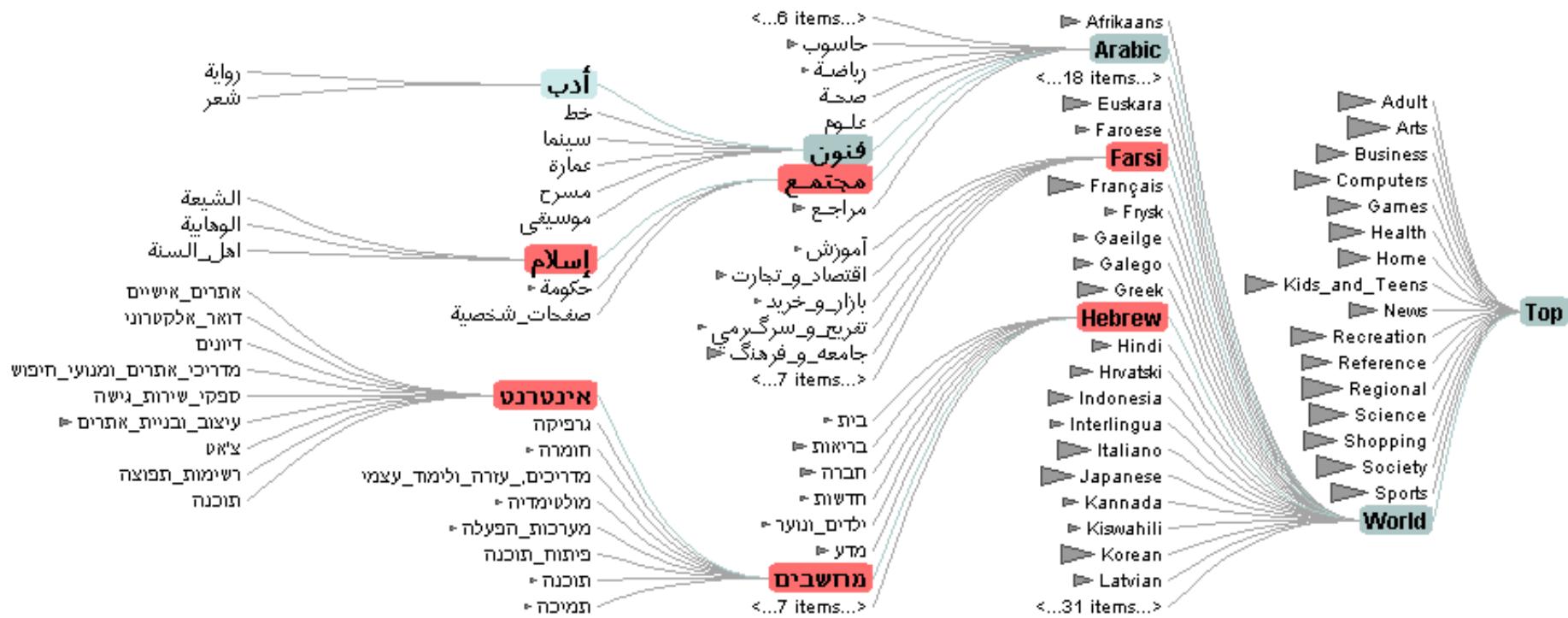
Animated transitions in statistical data graphics [Heer & Robertson 07]

Mapping & Cartography



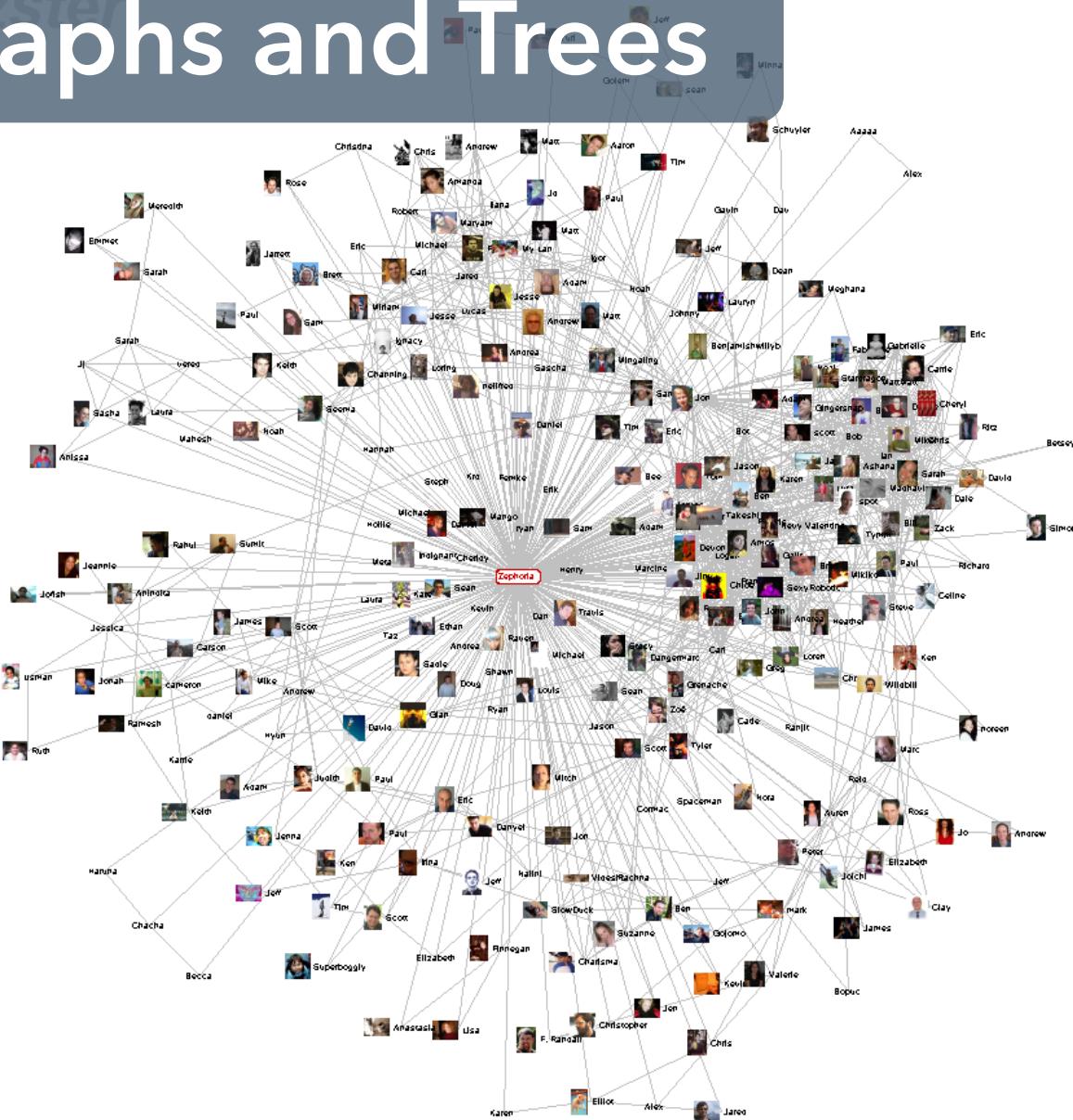
Dymaxion Maps [Fuller 46]

Graphs and Trees



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

Gizster Graphs and Trees



community >>



Enable

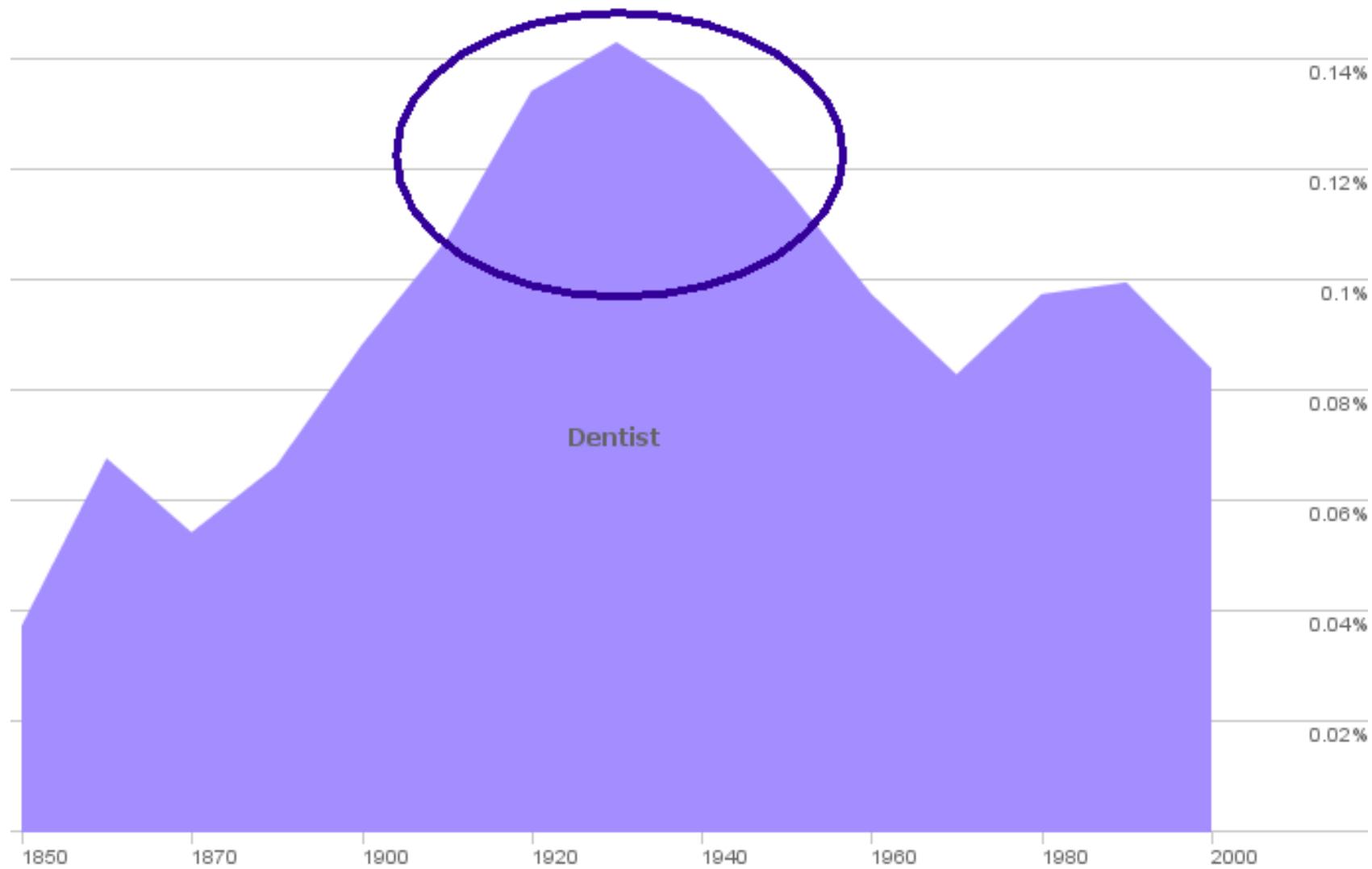
search >>

Want to Meet

Zephoria

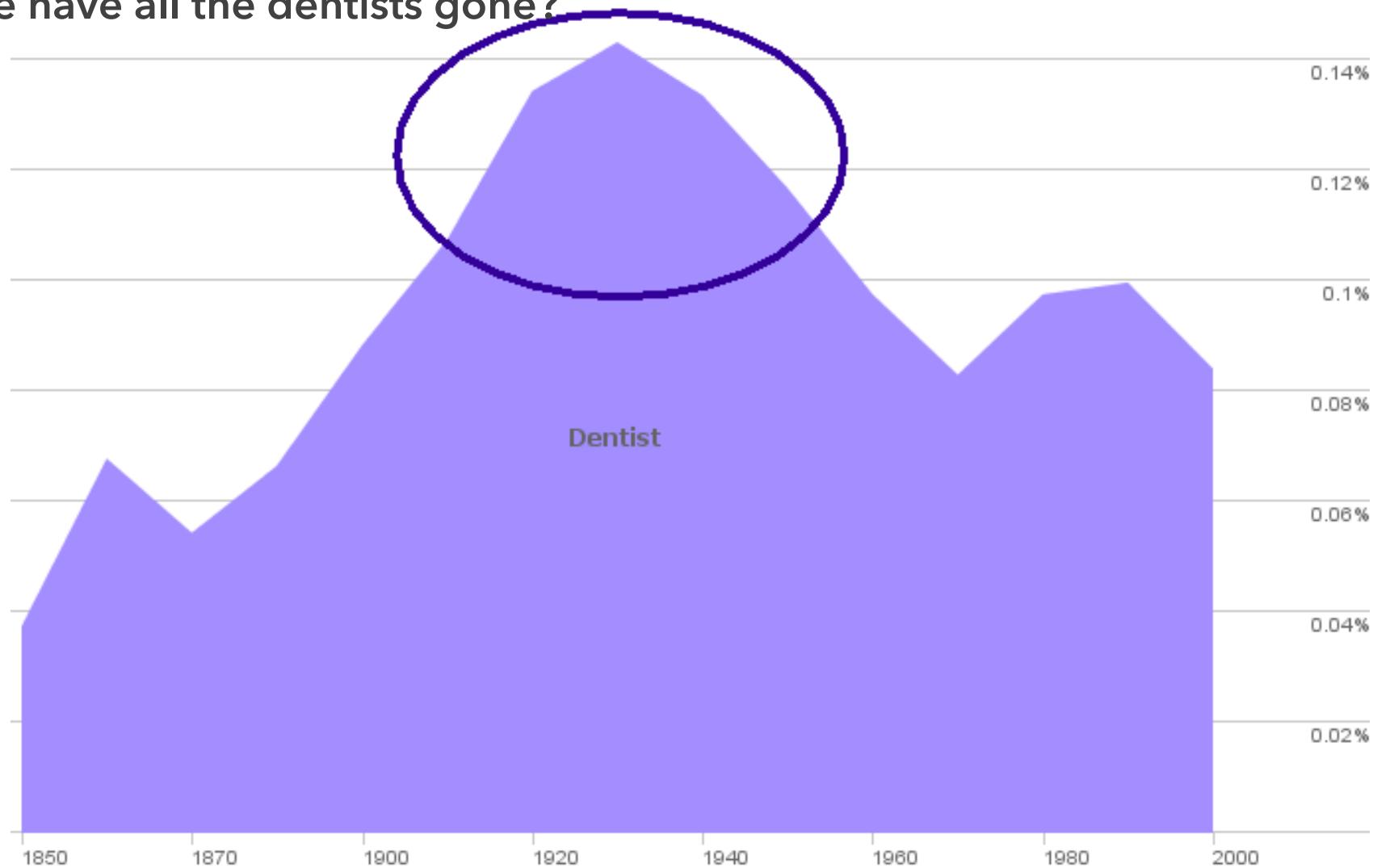
User ID: 21721
 Friends: 266
 Age: ??
 Gender: Female
 Status: Single
 Location: San Francisco, CA
 Hometown: Lancaster, PA
 Occupation: researcher: social networks, identity, context
 Interests: apophenia, observing people, culture, questioning power, reading, buddhism, ipseity, computer-mediated communication, social networks, technology, anthropology, stomping, psytrance/goa/trance [Infected Mushroom, Son Kite... Iboga/Digital Structures], Ani Difranco, downtempo, Thievery Corporation, Beth Orton, Morcheeba, Ween, White Stripes
 Music: Authors: Erving Goffman, Stanley Milgram, Jeanette Winterson, Eric Schlosser, Leslie Feinberg, Dorothy Allison, Italo Calvino, Hermann Hesse
 Books: ??
 TV Shows: 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/>
 Someone who makes life's complexities seem simply elegant.
 A partner in crime with an

Collaboration and History

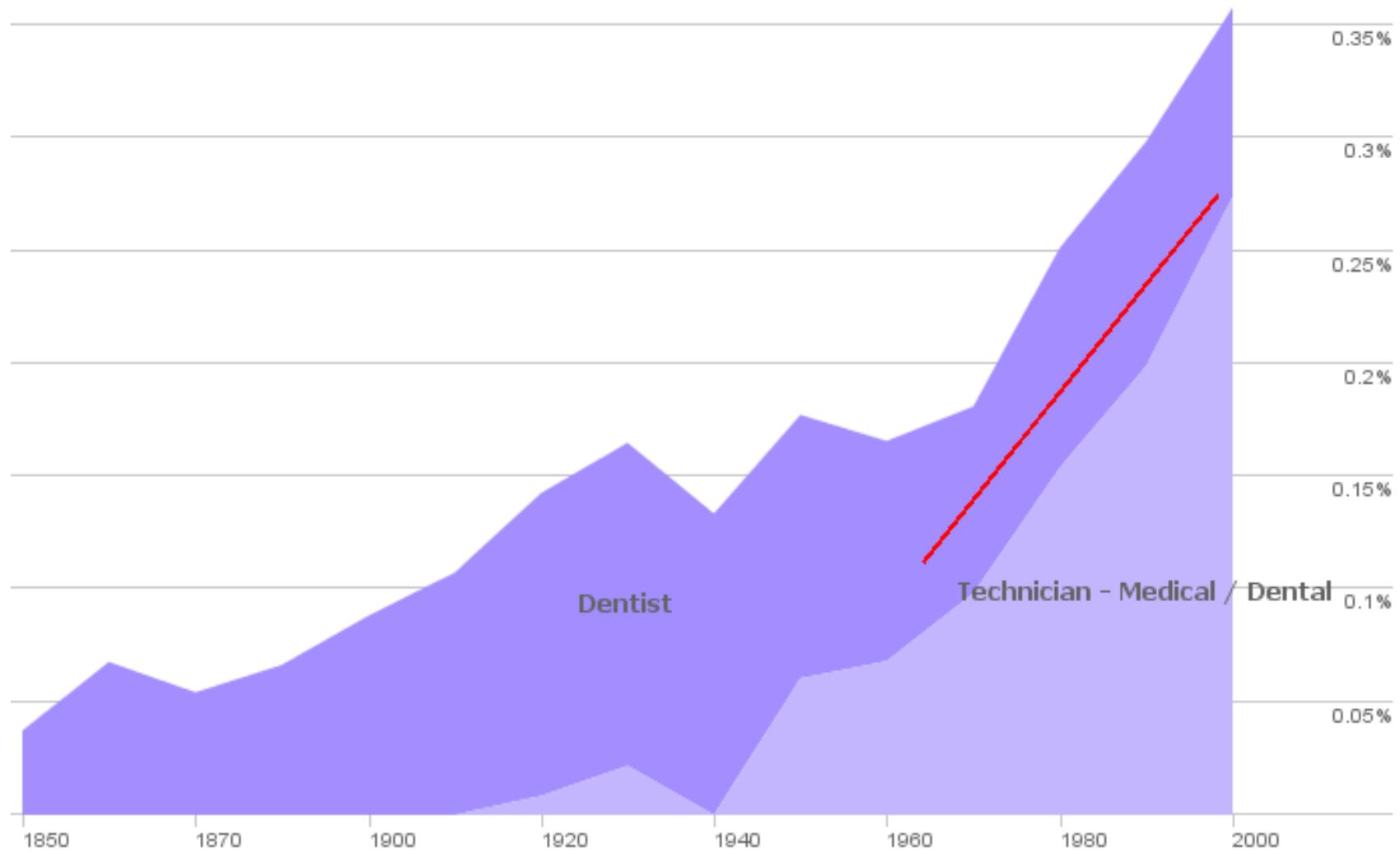


Collaboration and History

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

Assoc Prof, CSE

OH: *Th 10-11am, 642 CSE*

<http://jheer.org>

Assistants

Dominik Moritz

Jeff Snyder

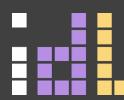
OH: *Fri 1:30-2:30, 218 CSE*

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





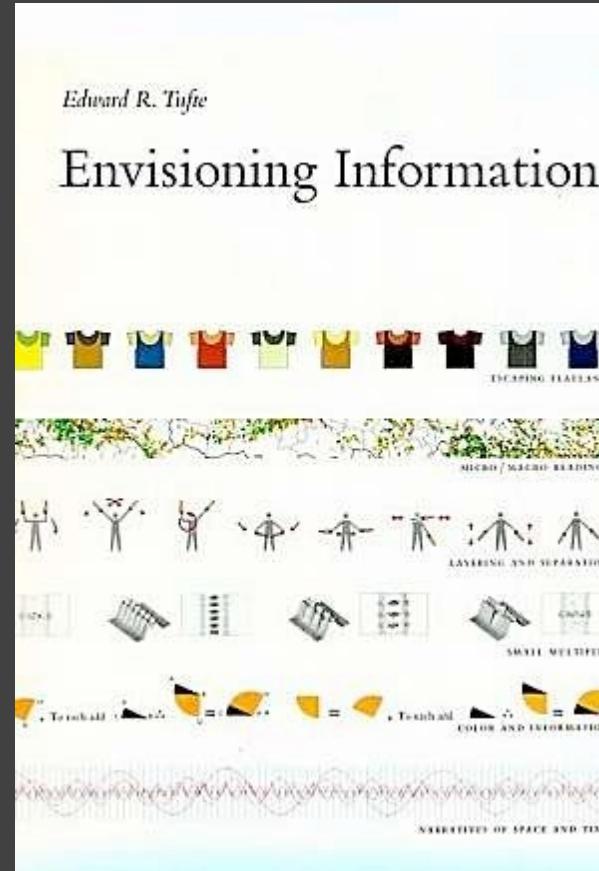
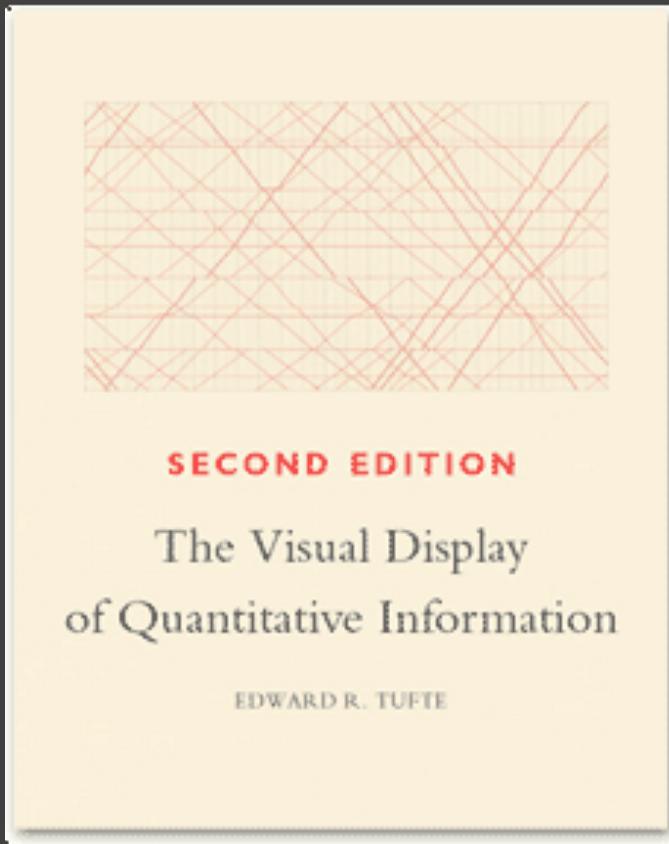
Google



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:

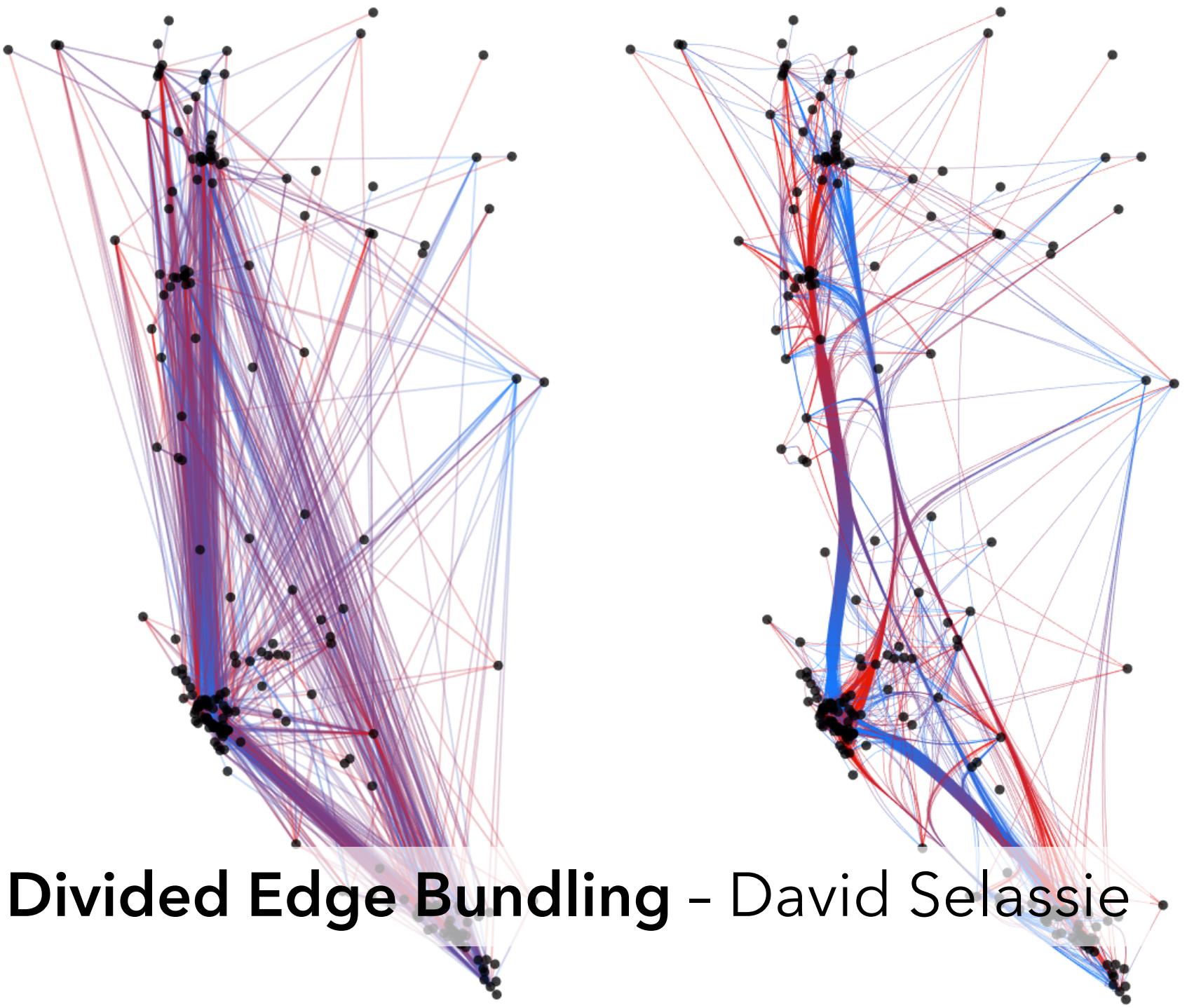
- Published (e.g., at the IEEE InfoVis conference)
- Featured in the New York Times
- Released as successful open source projects



RunMonster

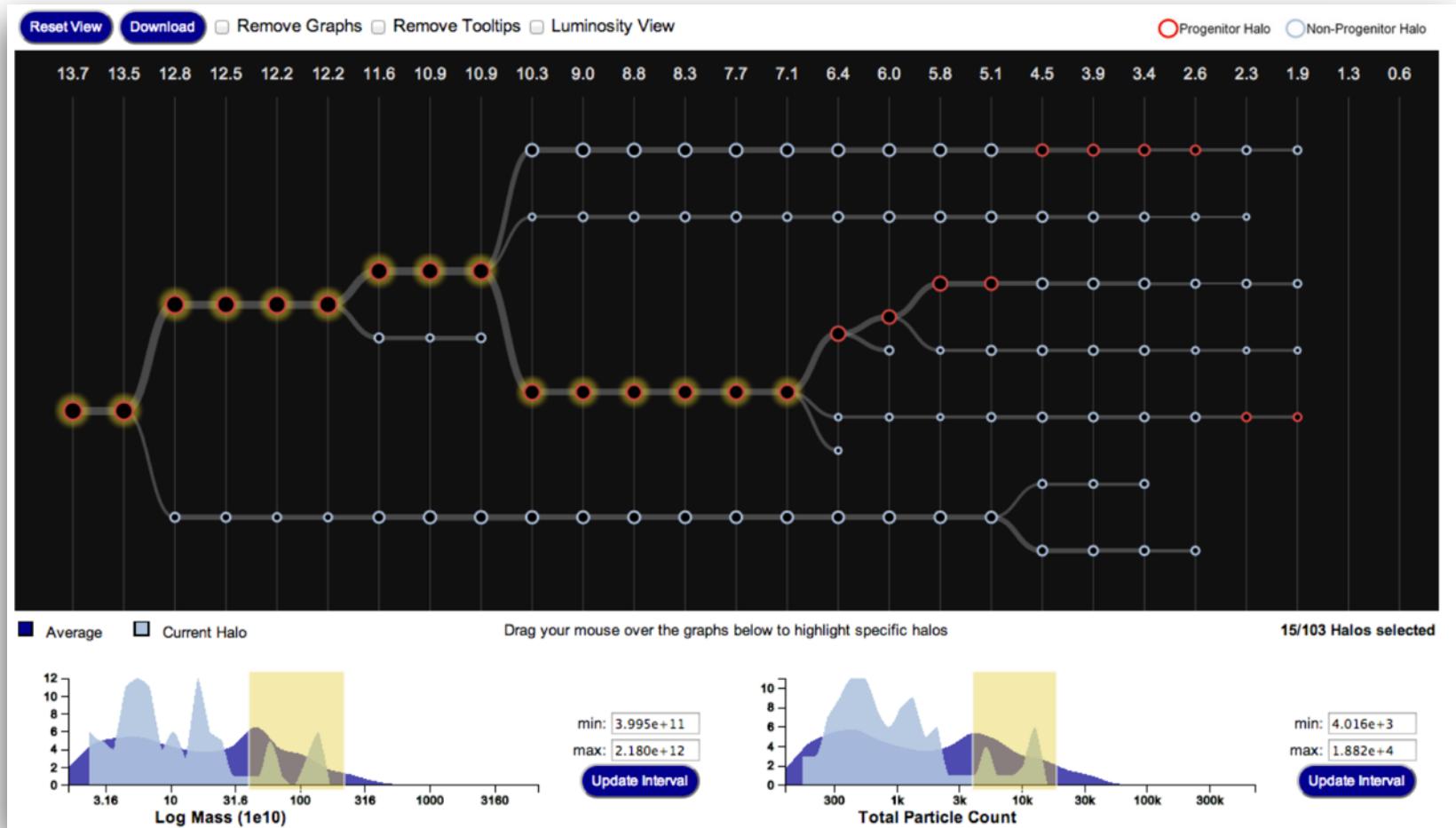
Troy Brant & Steve Marmon





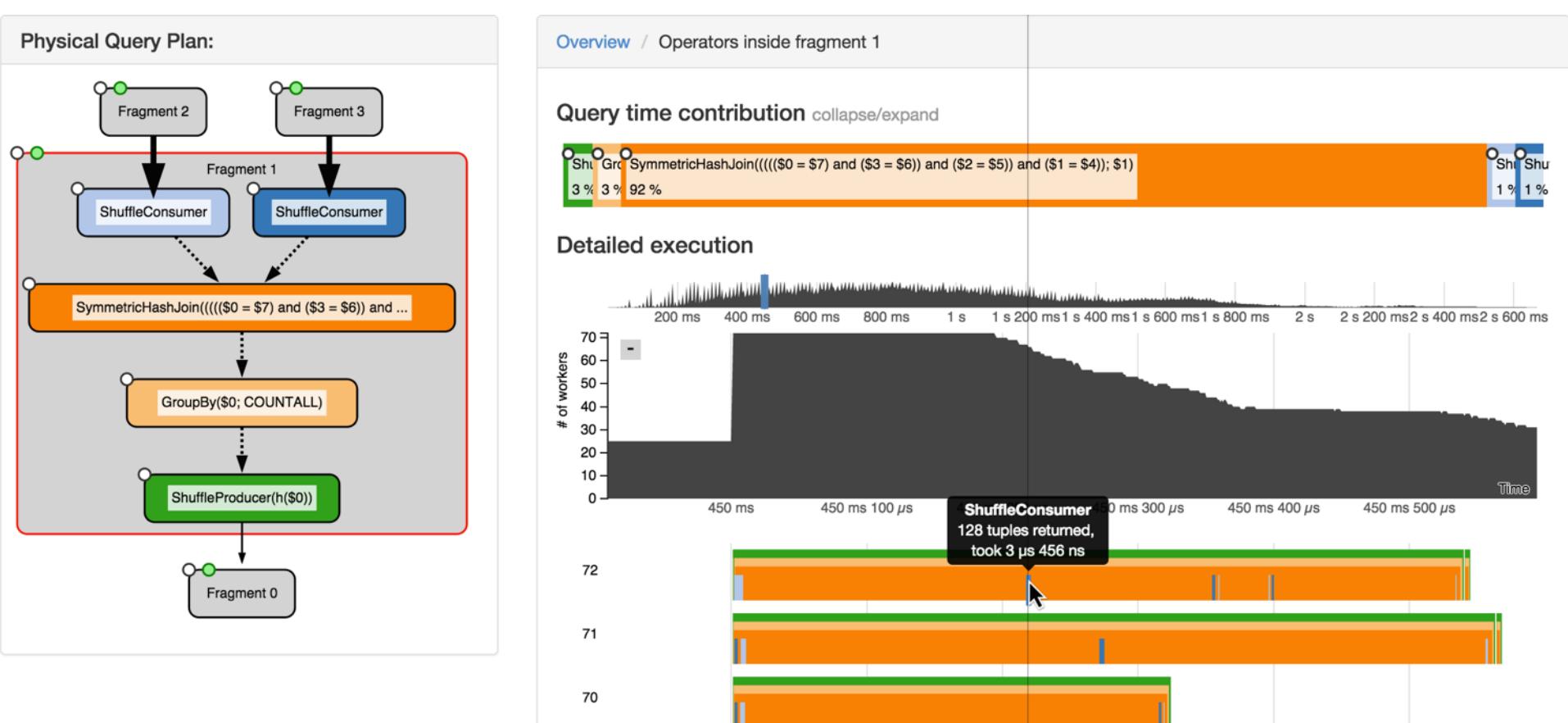
Divided Edge Bundling - David Selassie

Visualizing Galaxy Merger Trees



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

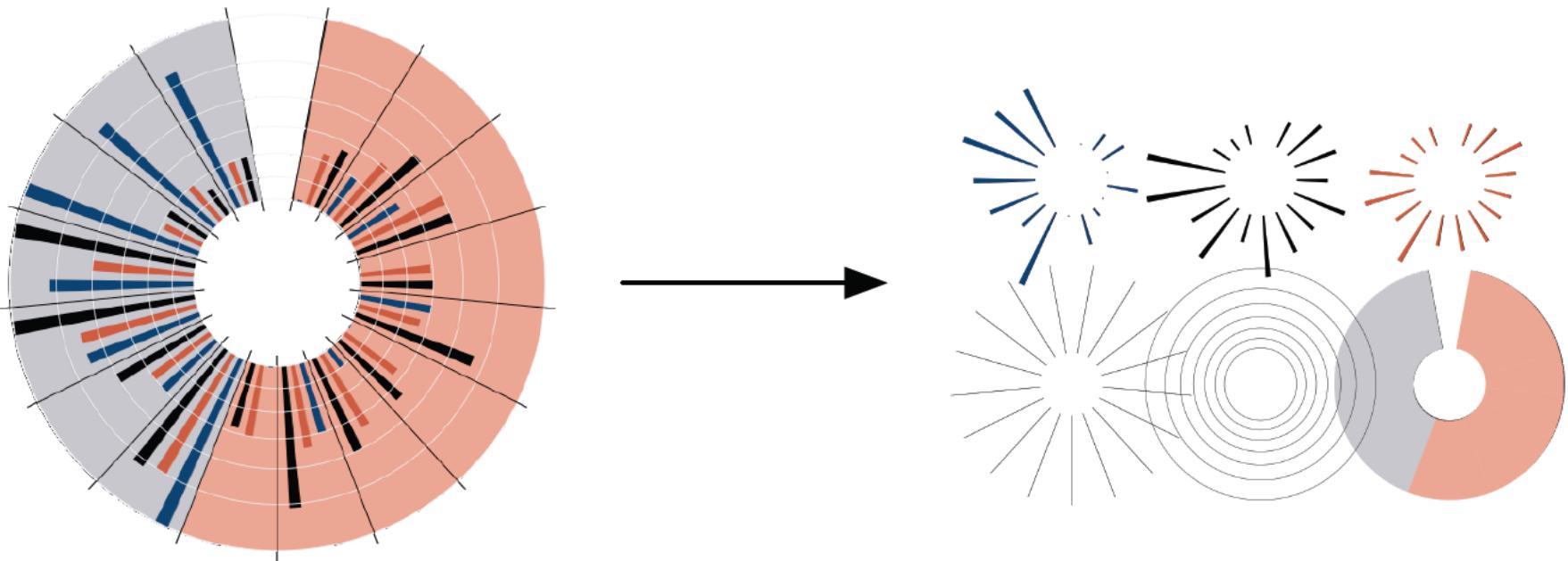
Perfopticon Distributed Query Performance

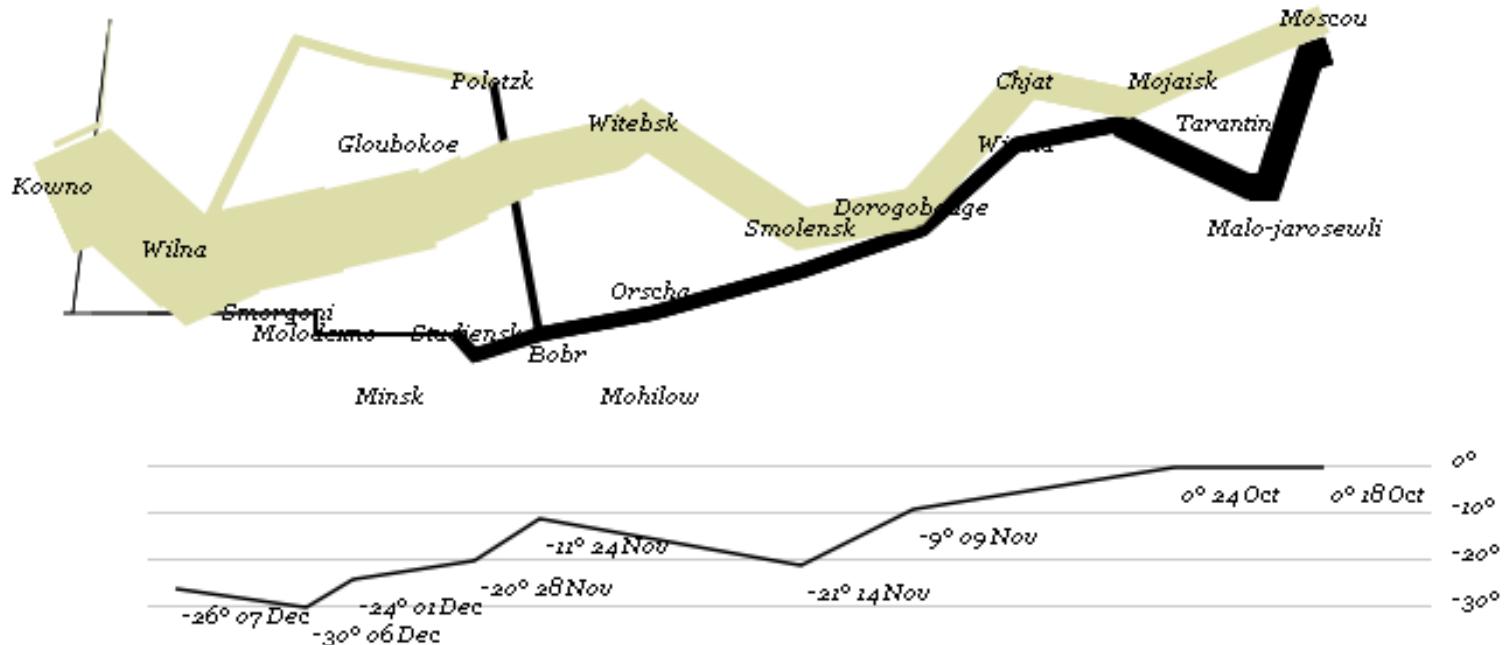


Dominik Moritz et al. [EuroVis '15]

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[panelIndex][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

Republic of Letters

1700

United Kingdom

Denmark

Ireland

France

Spain

Germany

Poland

Italy

Letters: (97 max)

68

44

11

5

3

1

0



FILTER BY AUTHOR

[Clear](#) [All](#)

Damien Desormes

Daniel Cornabs

Daniel de Pury

Daniel Defoe

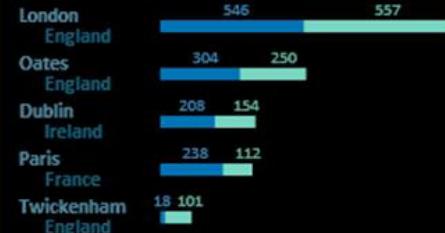
Daniel Malthus

Daniel Marc Antoine Chardon

Daniel Muller

TOP CITIES AND AUTHORS

█ Letters received █ Letters sent



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 (\leq 4 paragraphs)

Due by **5:00 pm, Monday April 6.**