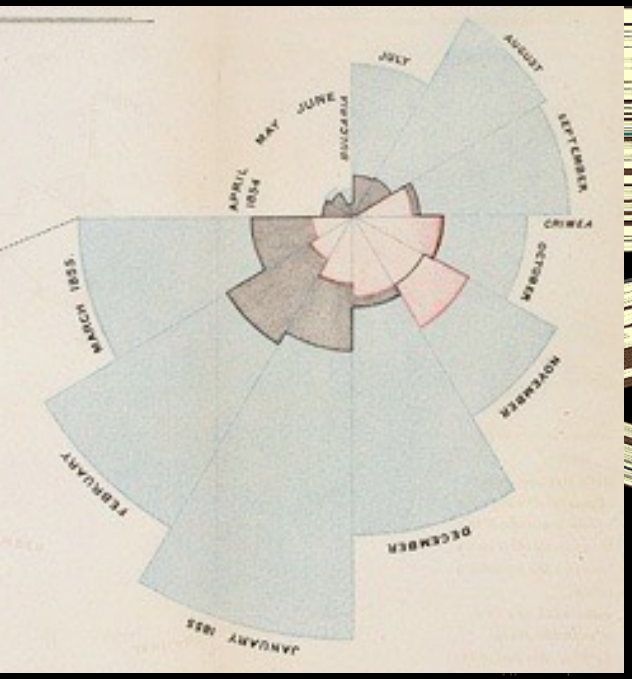
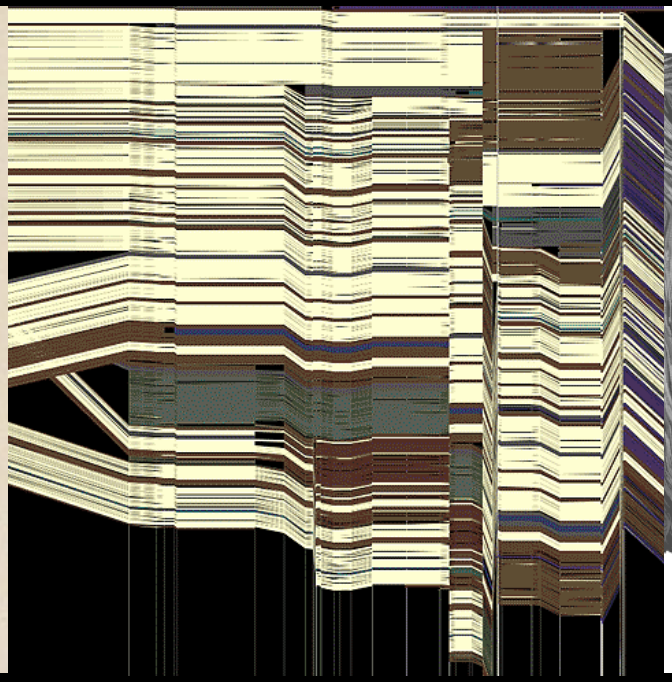
CSE 442 - Data Visualization

The Value of Visualization





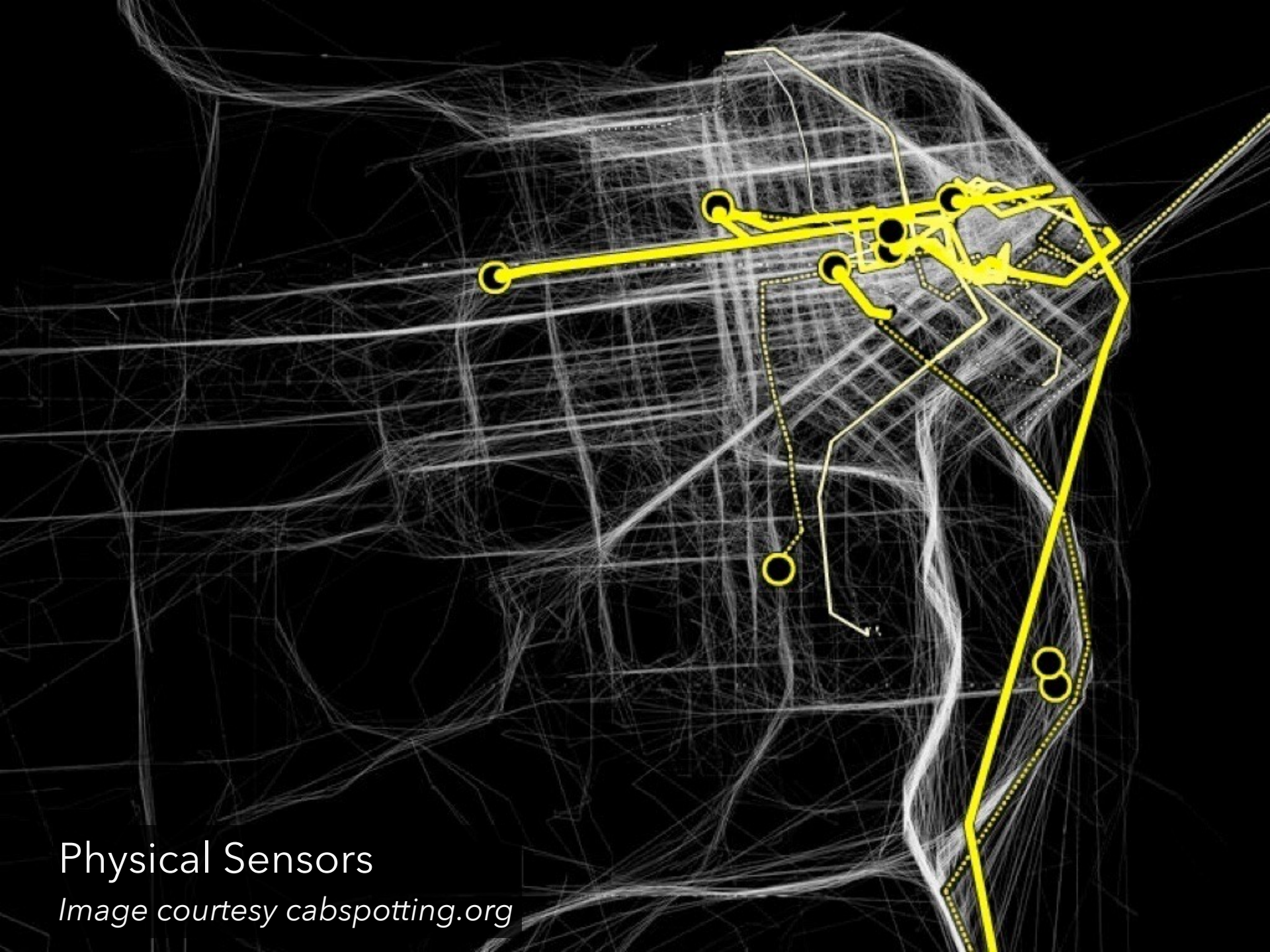
Jeffrey Heer University of Washington

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

2010: 1,200 exabytes

10x increase over 5 years

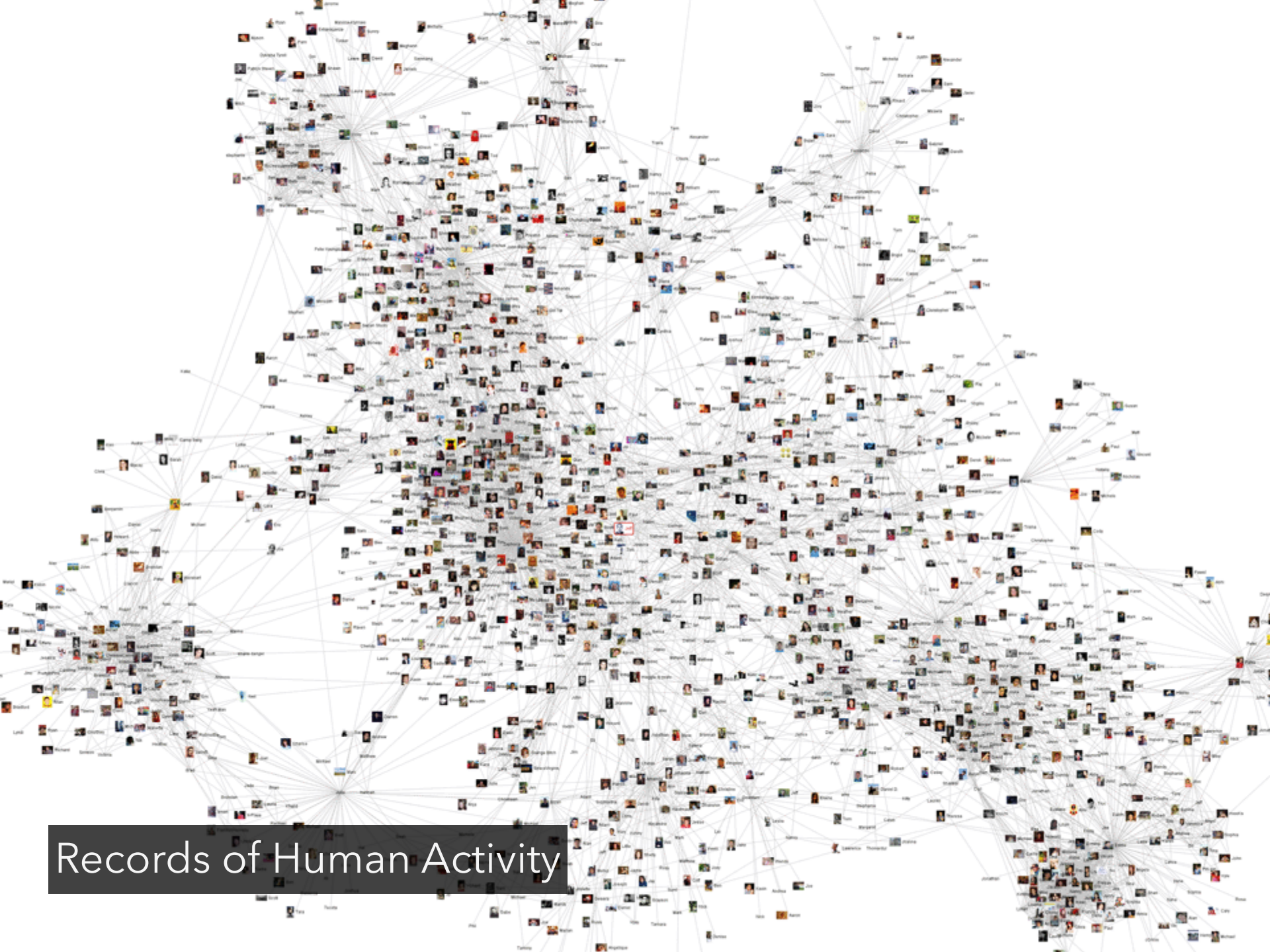
Gantz et al, 2008, 2010

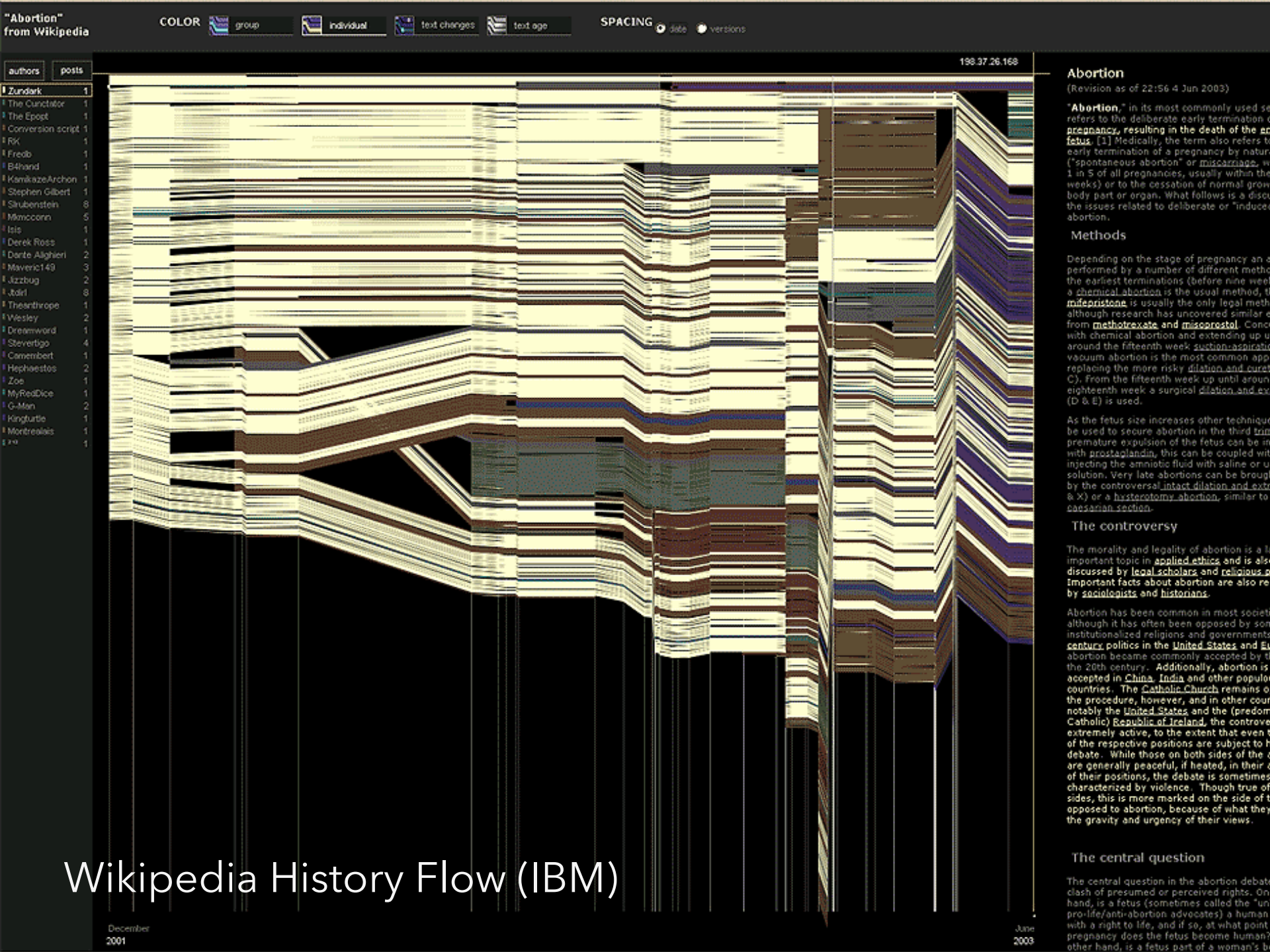


Physical Sensors

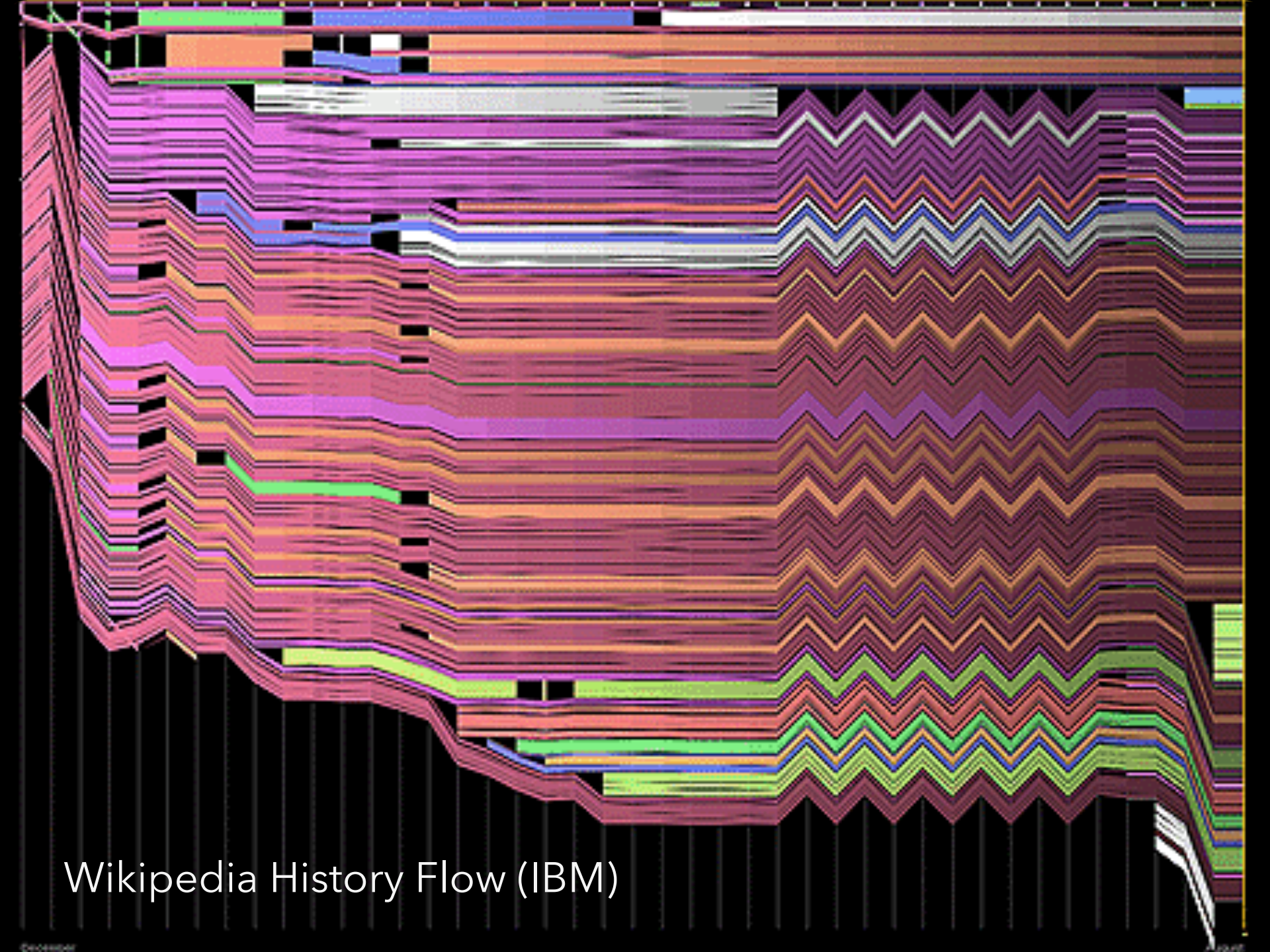
Image courtesy cabspotting.org

Health & Medicine


Records of Human Activity



Wikipedia History Flow (IBM)

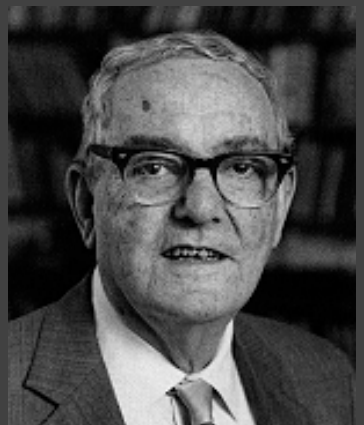

Wikipedia History Flow (IBM)

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

A Poverty of 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

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

$u_X = 9.0$ $\sigma_X = 3.317$

$u_Y = 7.5$ $\sigma_Y = 2.03$

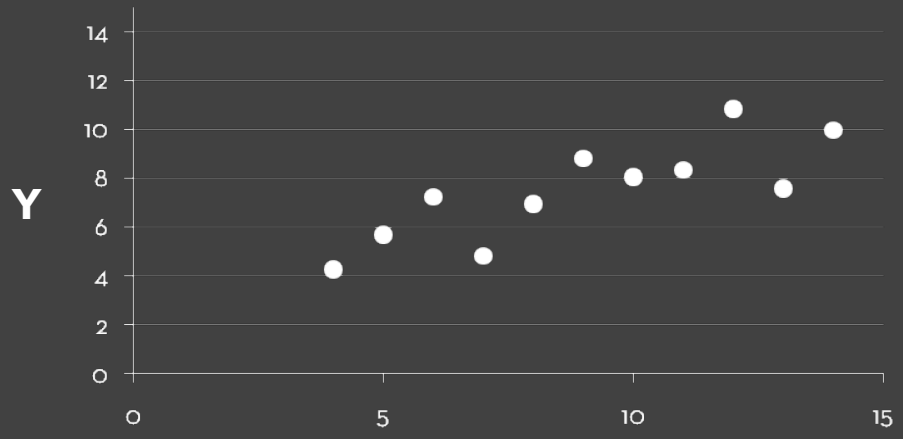
Linear Regression

$Y = 3 + 0.5\ X$

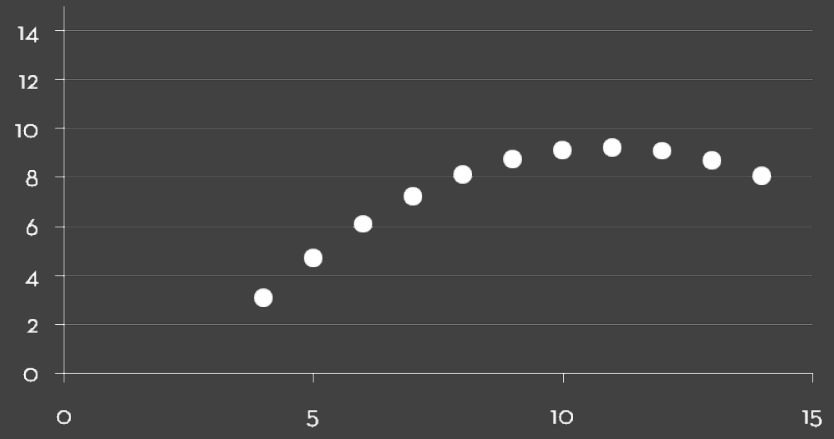
$R^2 = 0.67$

[Anscombe 1973]

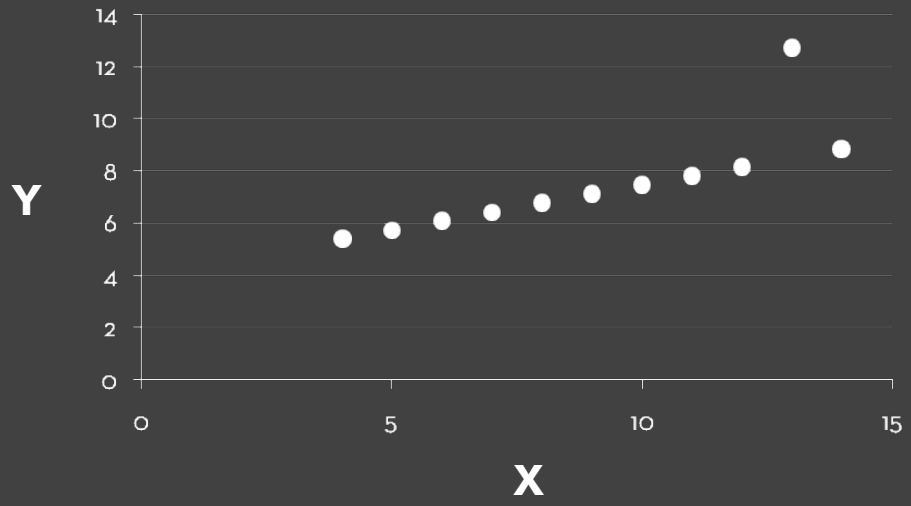
Set A



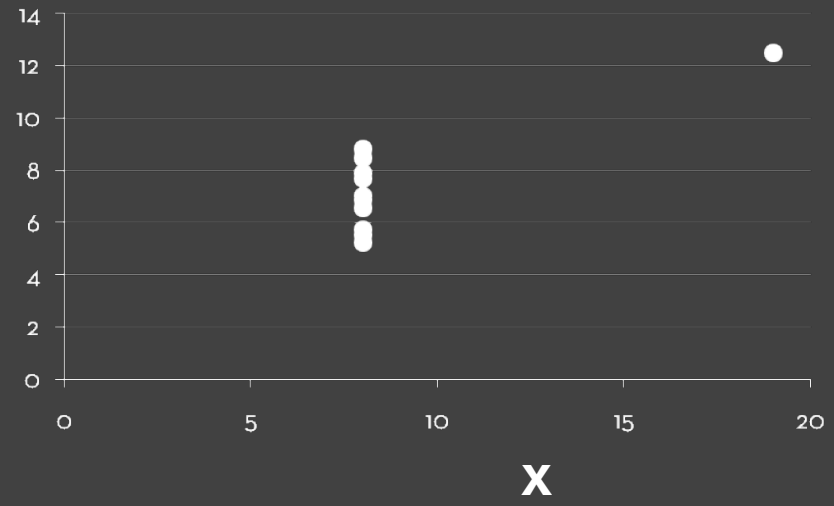
Set B



Set C



Set D

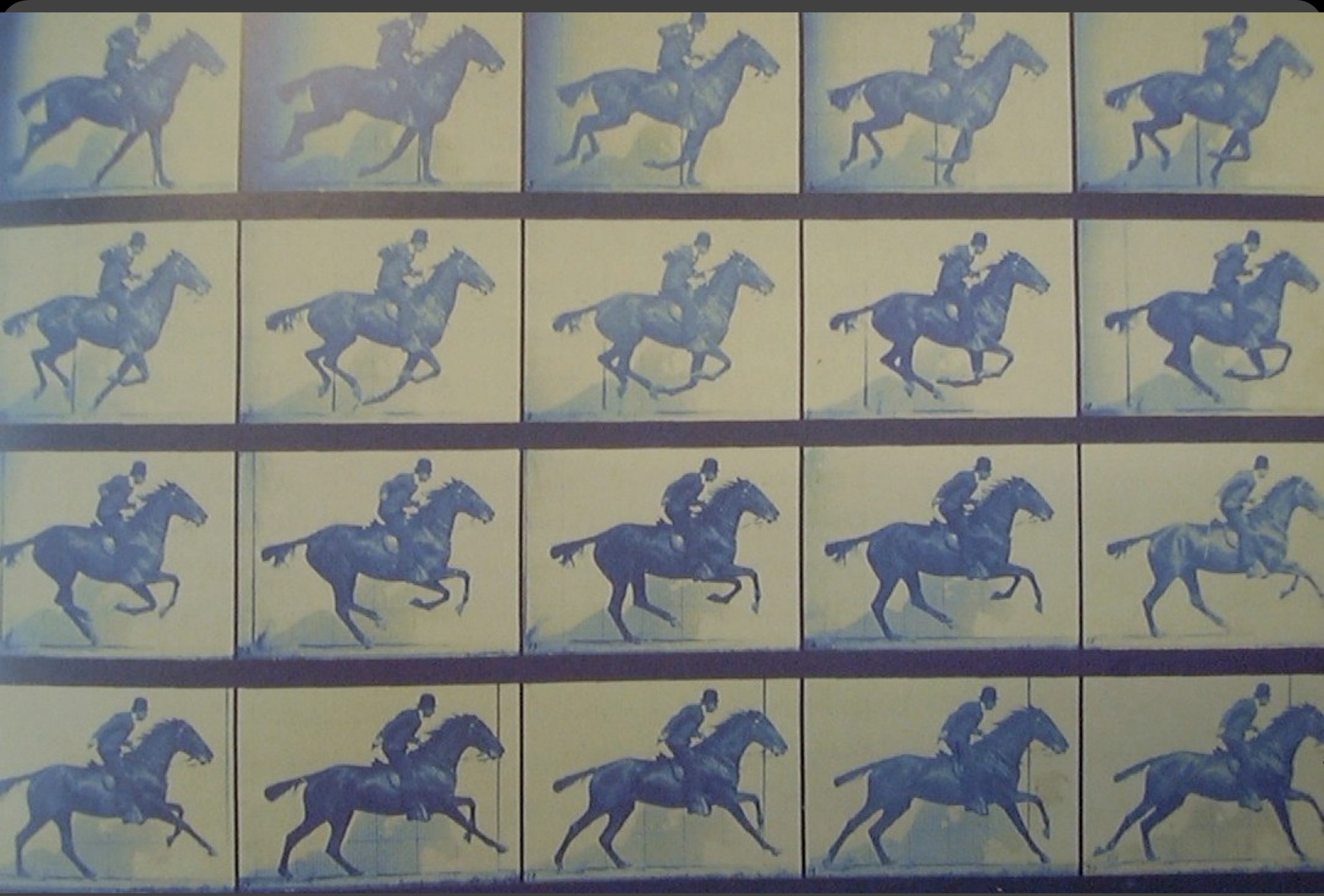


Why Create Visualizations?

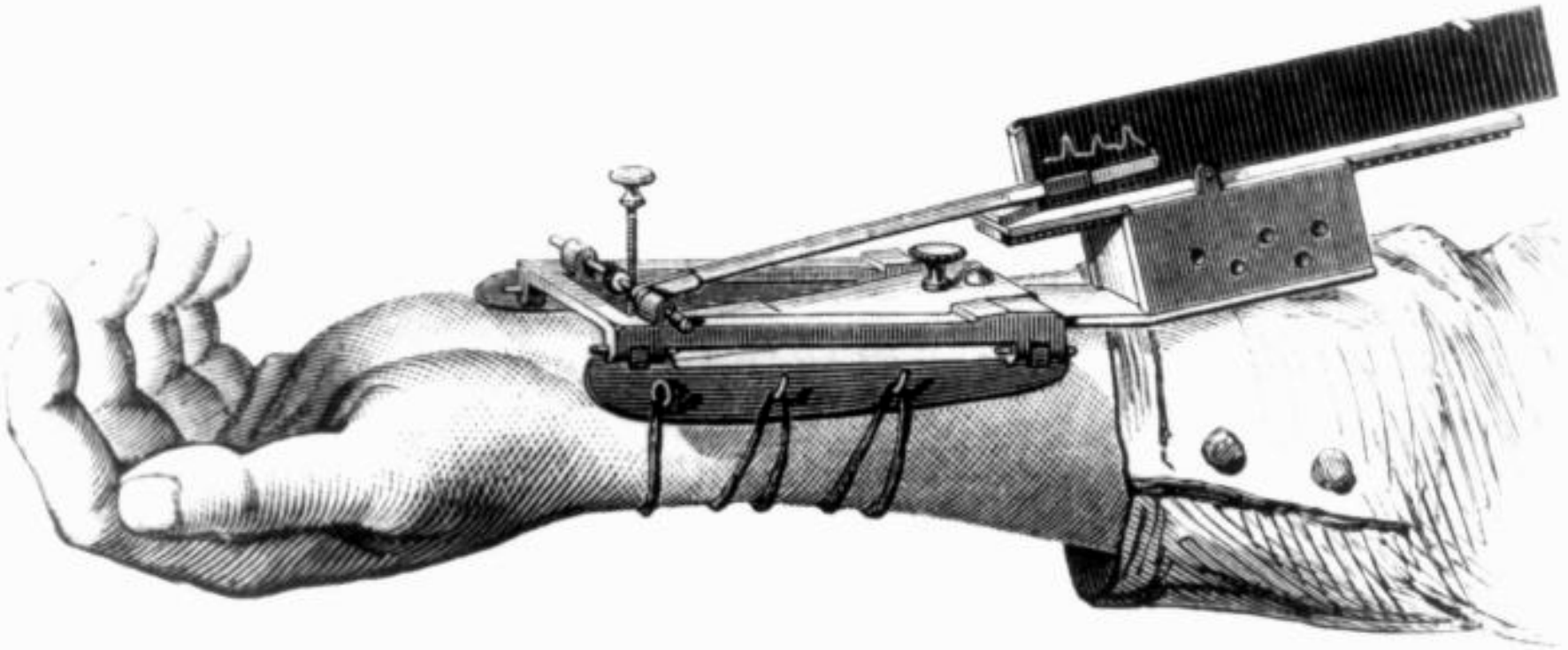
Why Create Visualizations?

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

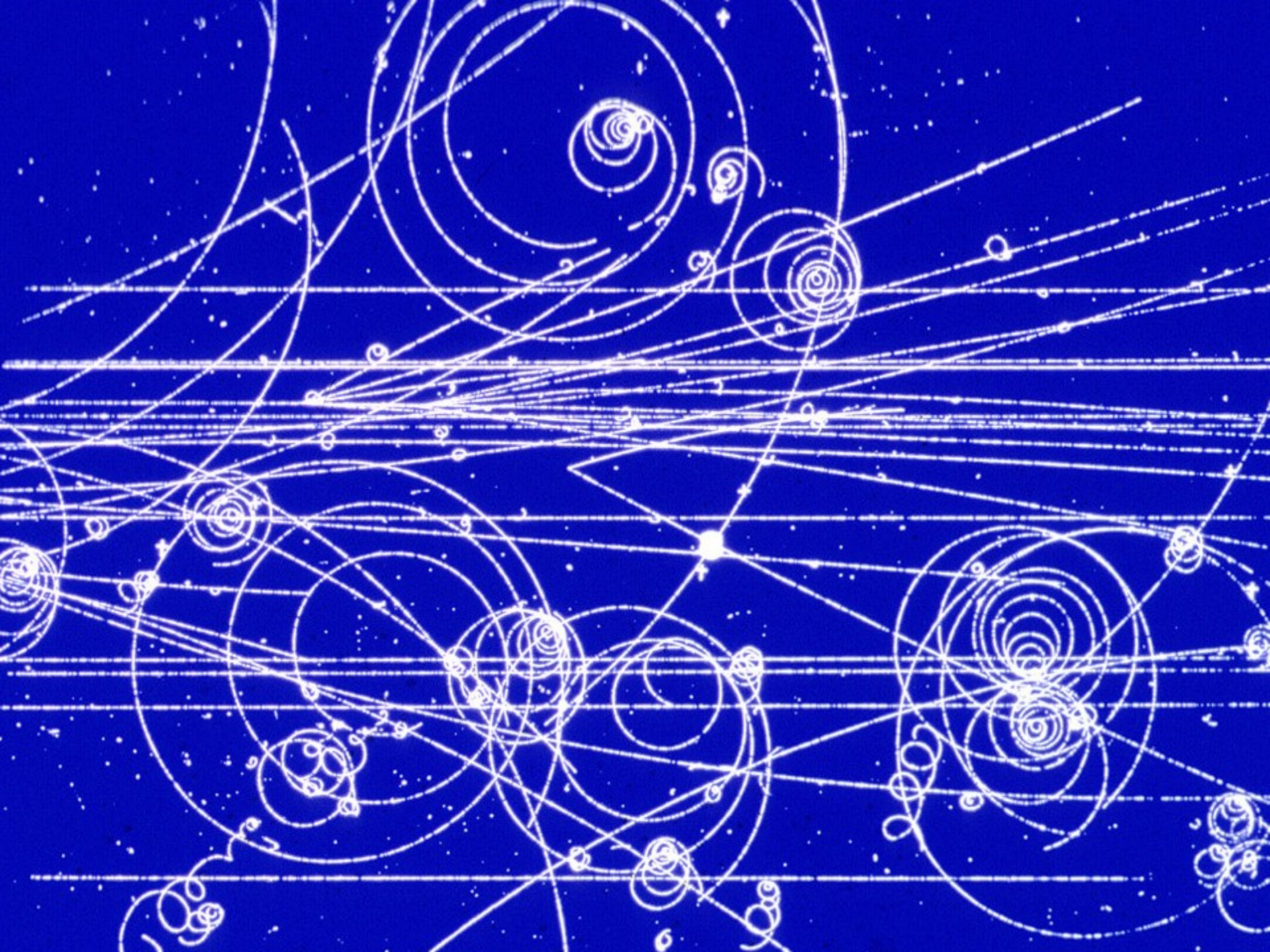


Gallop, Bay Horse “Daisy” [Muybridge 1884-86]

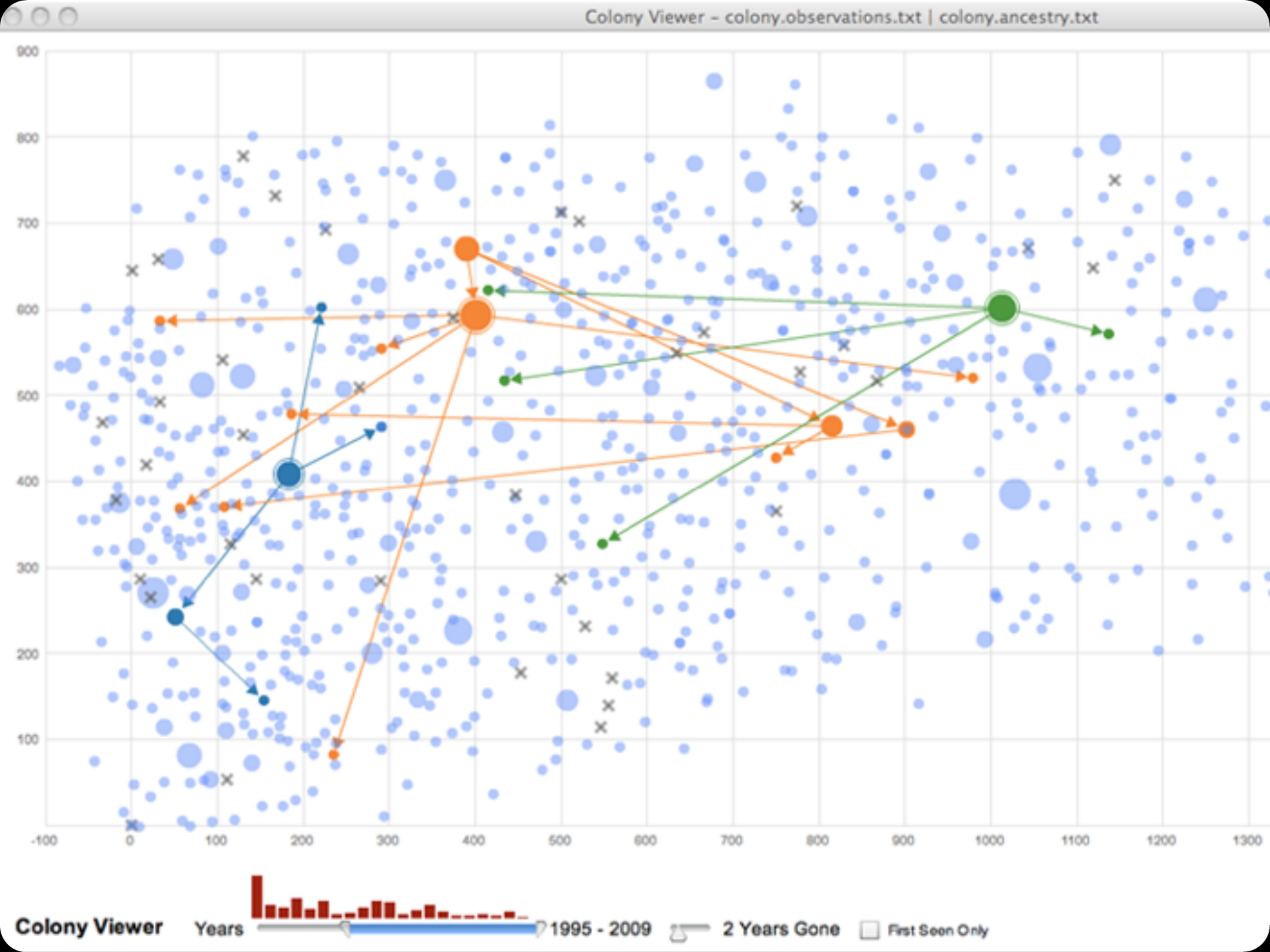


1.
Marey's sphygmograph in use, 1860. *La méthode graphique dans les sciences expérimentales et principalement en physiologie et en médecine.*

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°)

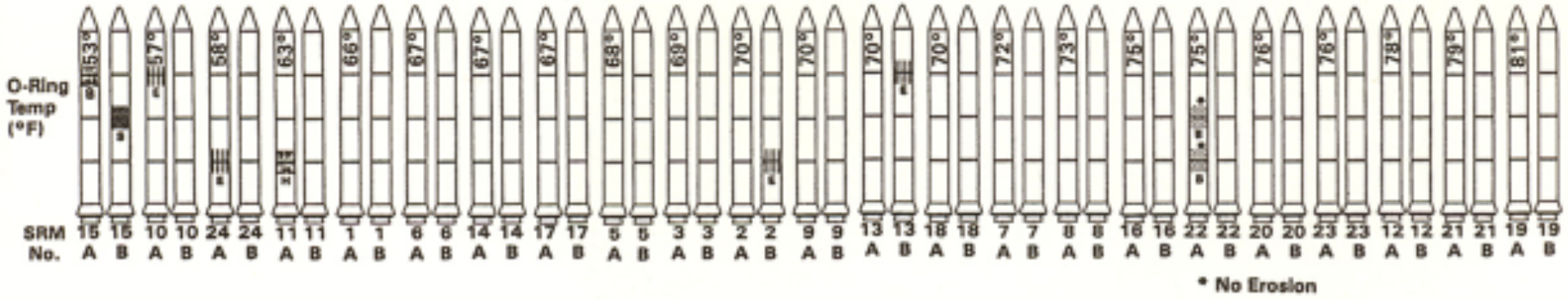
SRM-13A, 15, 16A, 18, 23A 24A
- nozzle Blow-By

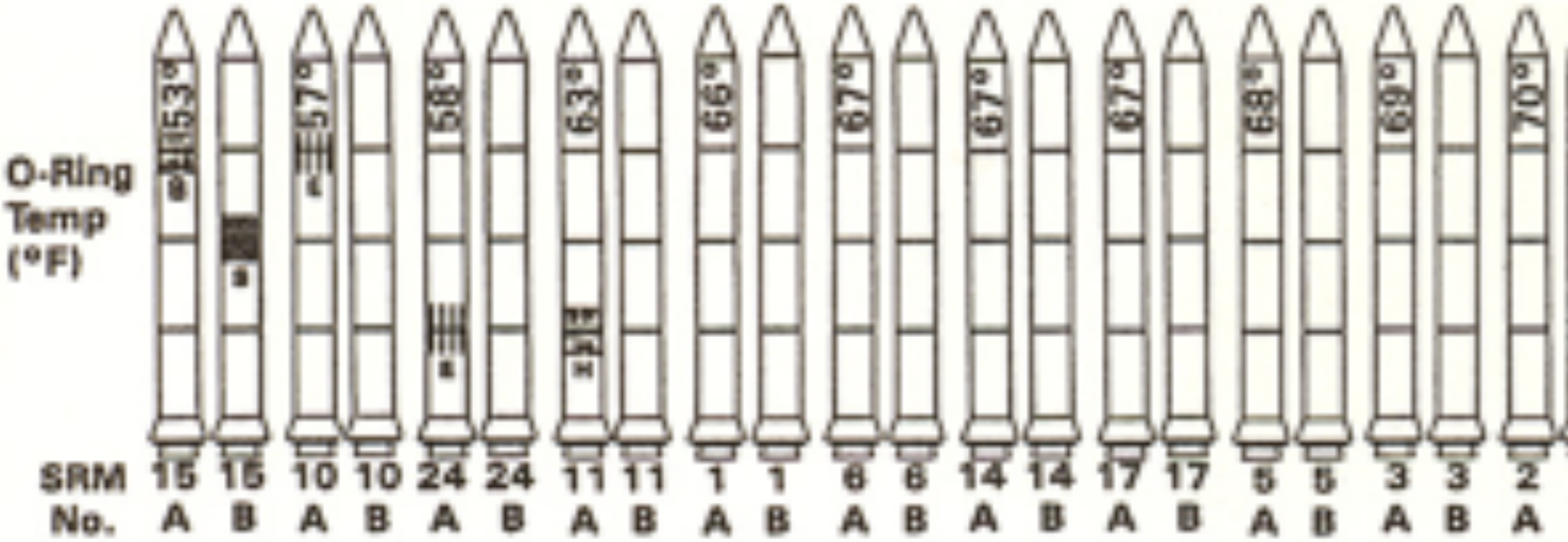
History of O-ring Temperatures (Degrees - F)

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

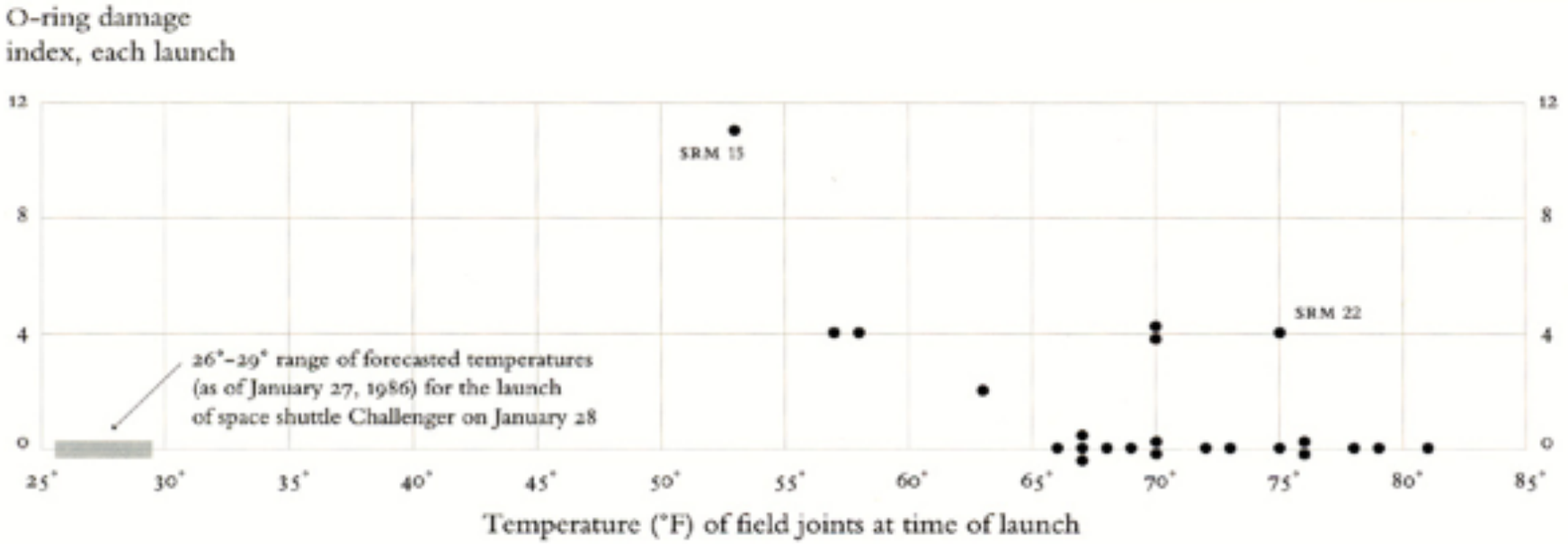
2 of 13 pages of material faxed to NASA by Morton Thiokol [from Tufte 1997]

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]

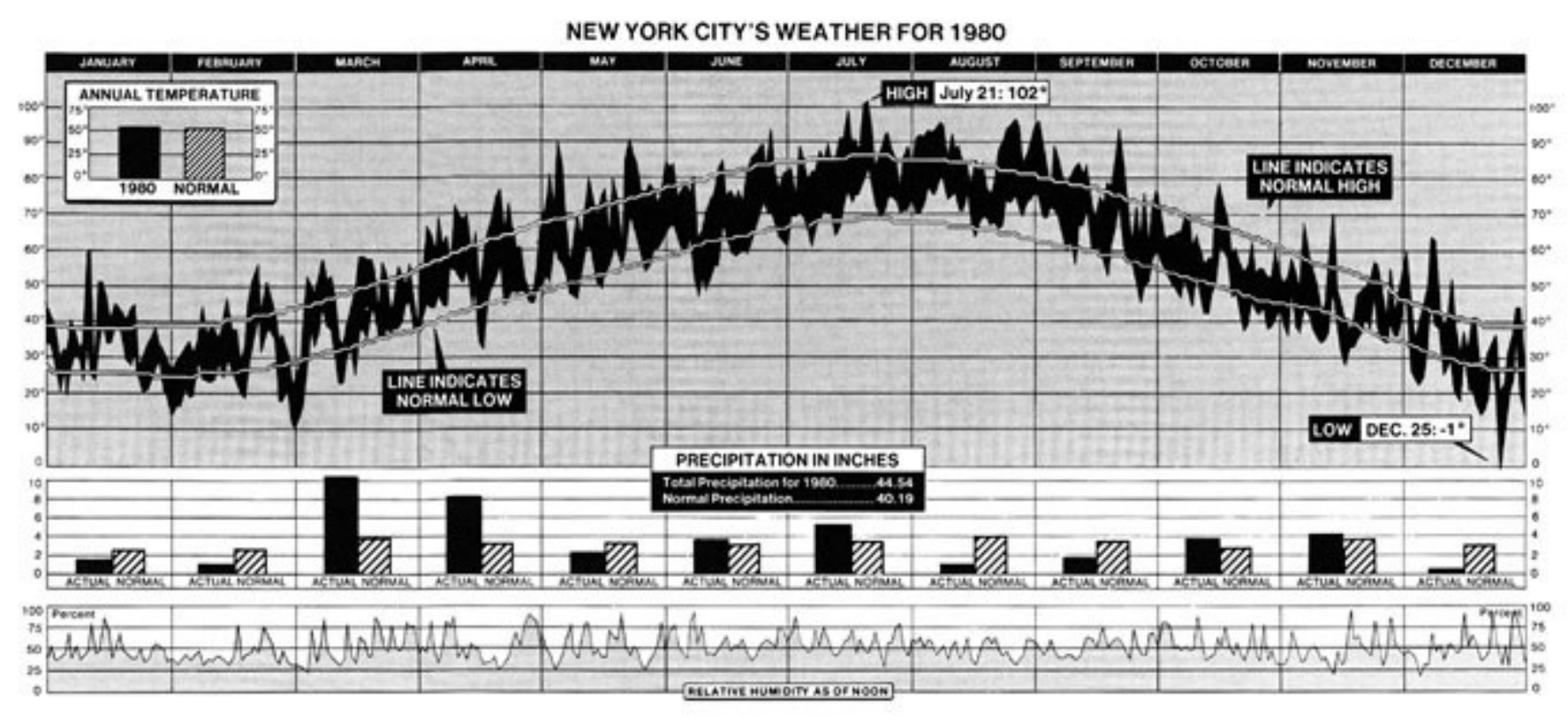
Data in Context: Cholera Outbreak

Used map to hypothesize that pump on Broad St. was the cause. [from Tufte 83]

Expand Memory: Multiplication

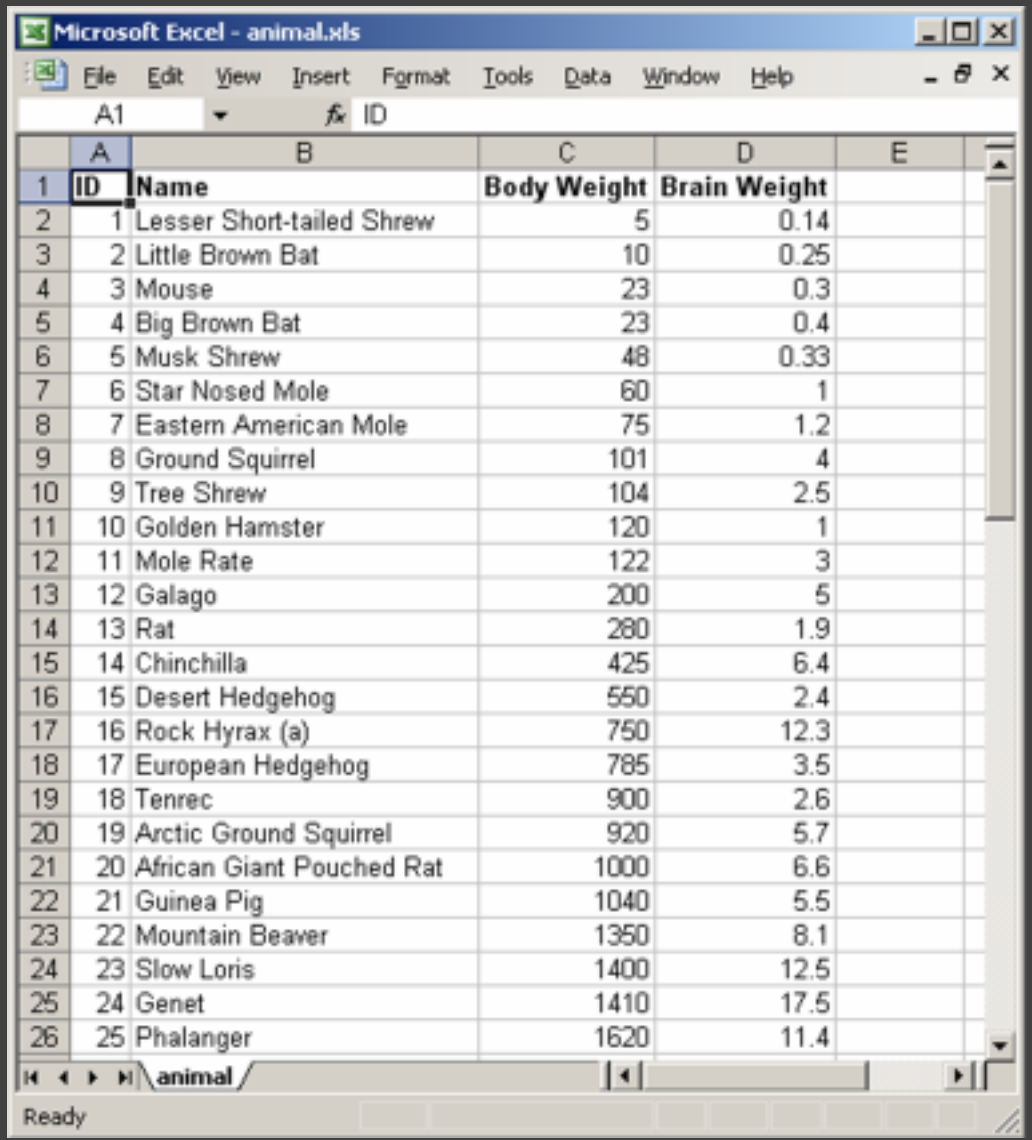
Class Exercise!

Find Patterns: NYC Weather

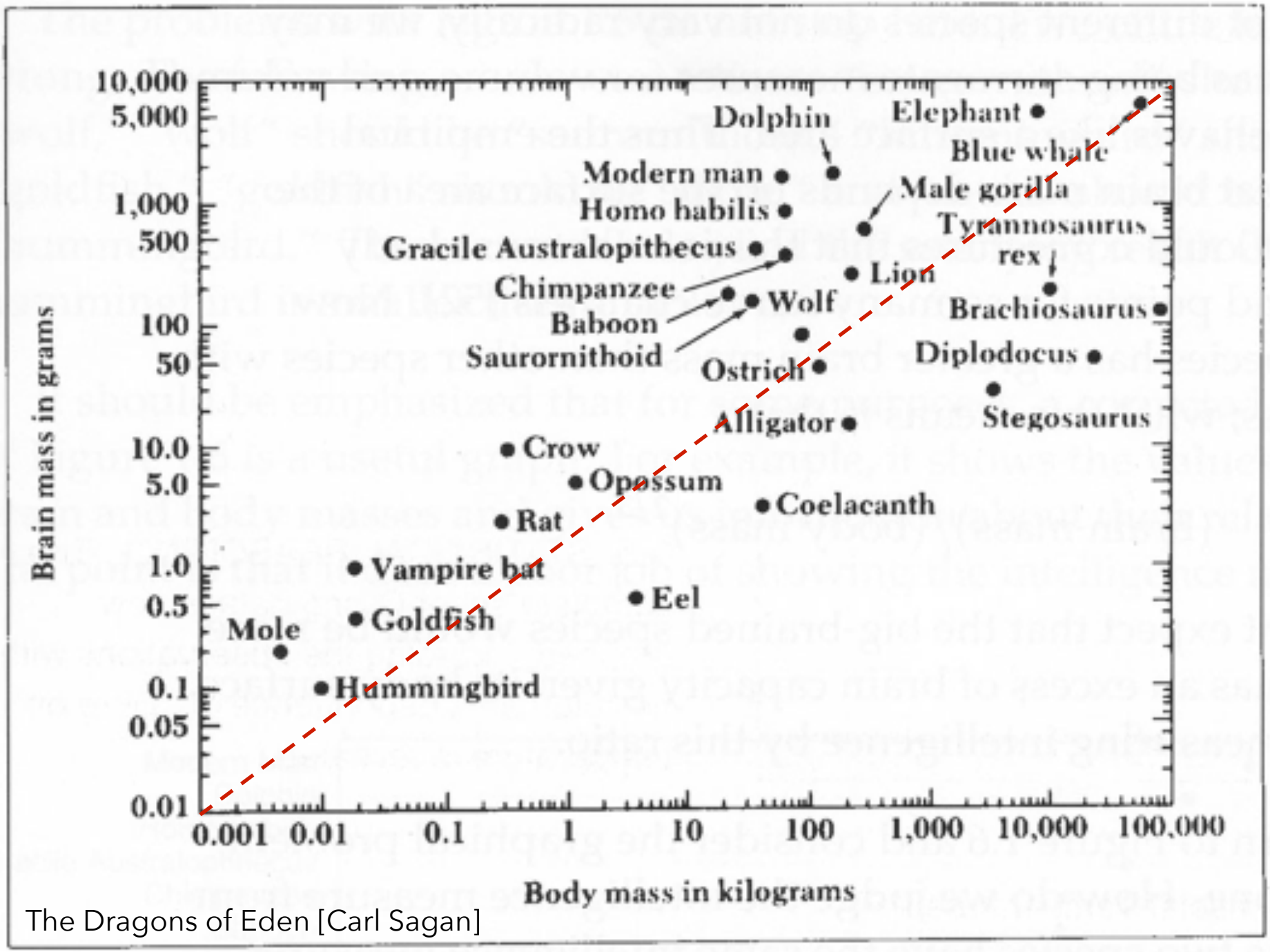


[New York Times 1981]

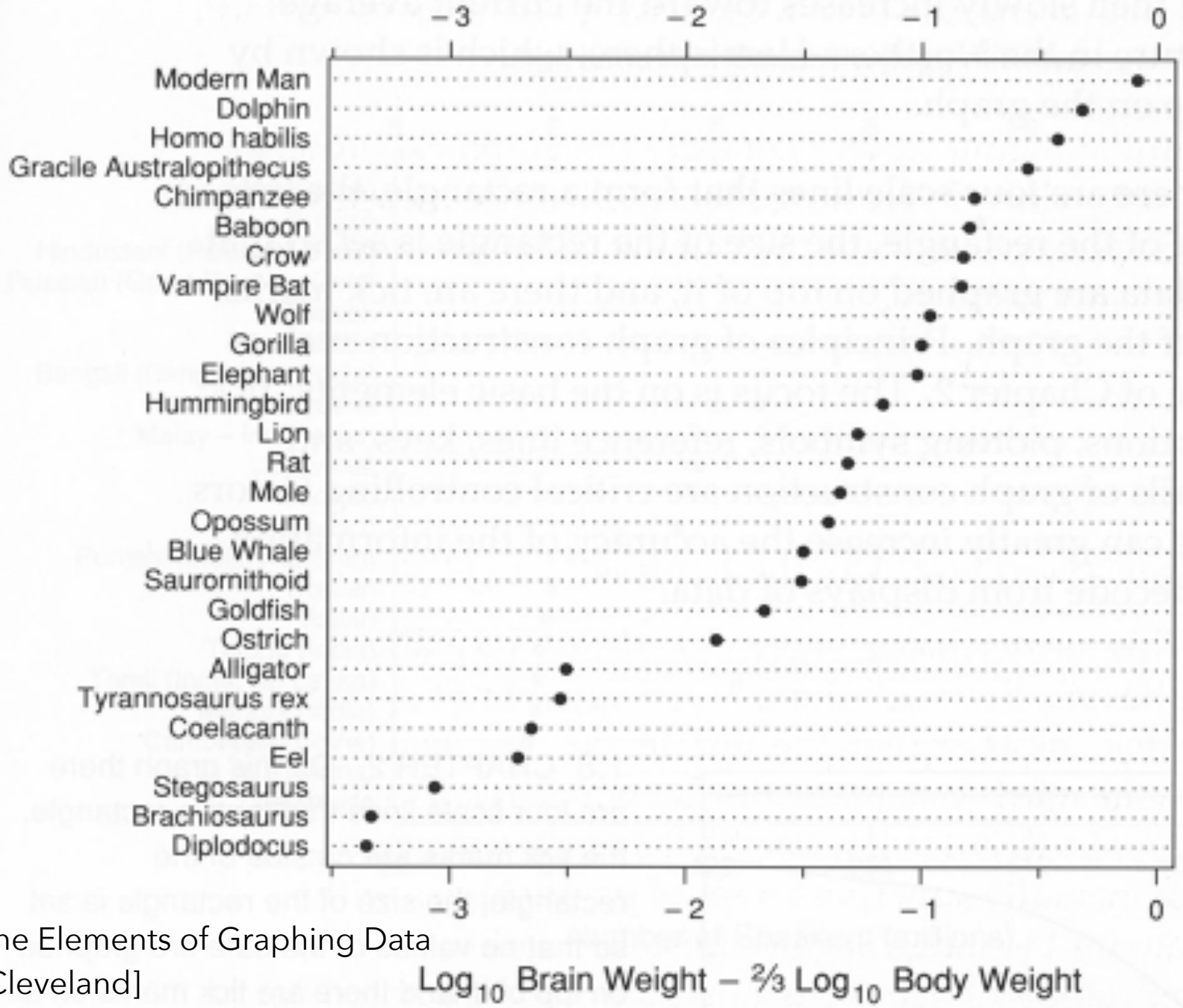
The Most Powerful Brain?



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



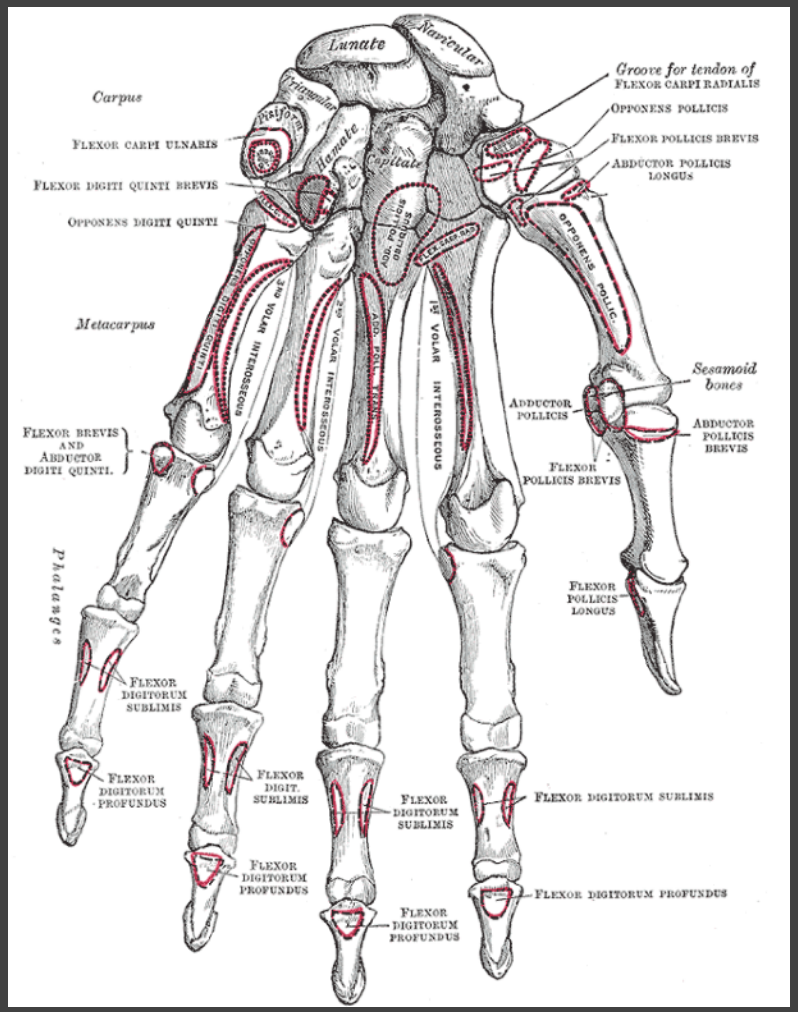
The Dragons of Eden [Carl Sagan]



The Elements of Graphing Data
[Cleveland]

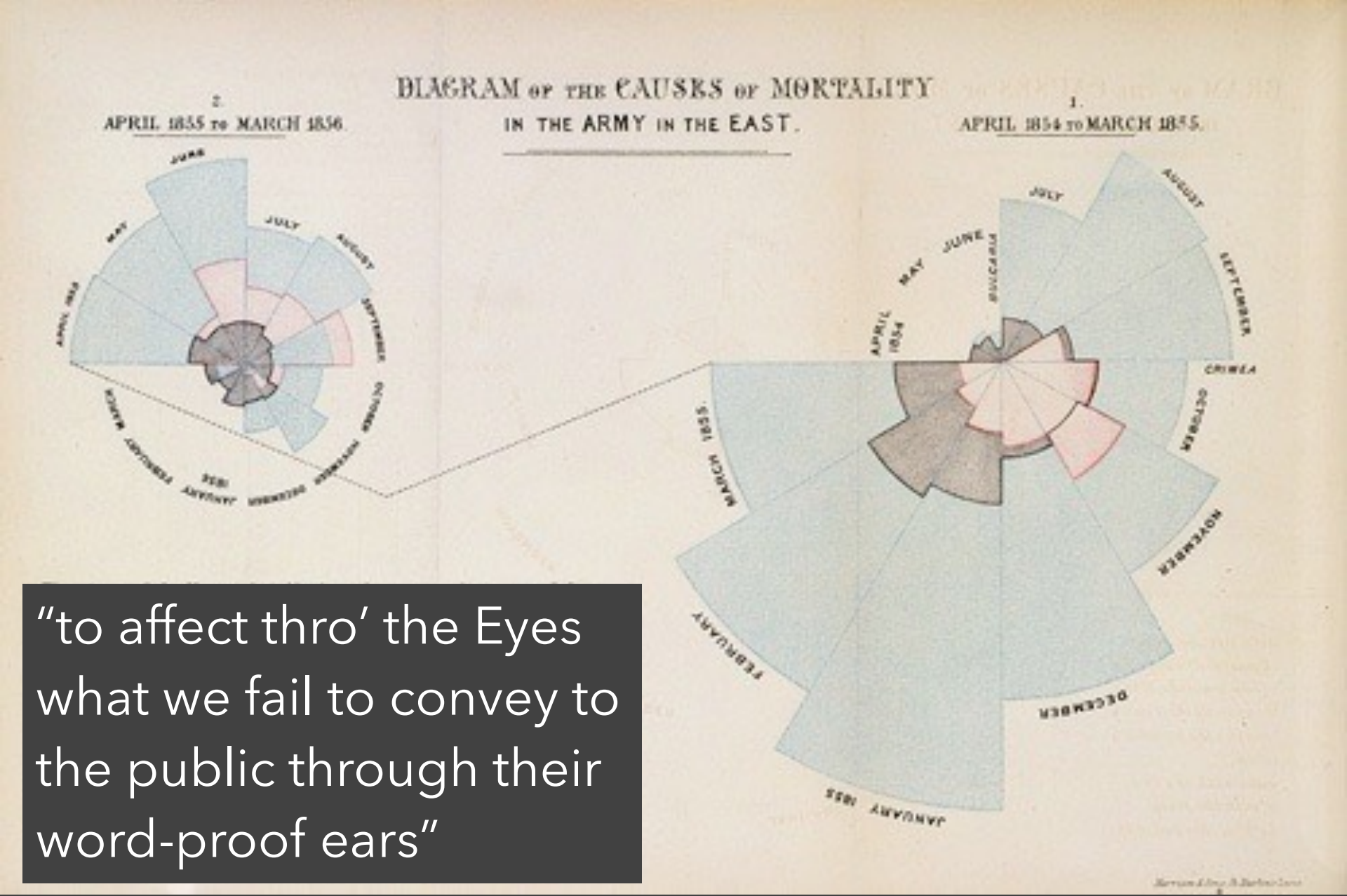
Convey Information to Others

Inspire



Bones in hand [from 1918 edition]

Double helix model [Watson and Crick 53]



"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

Goals of Visualization Research

1. **Understand** how visualizations convey information

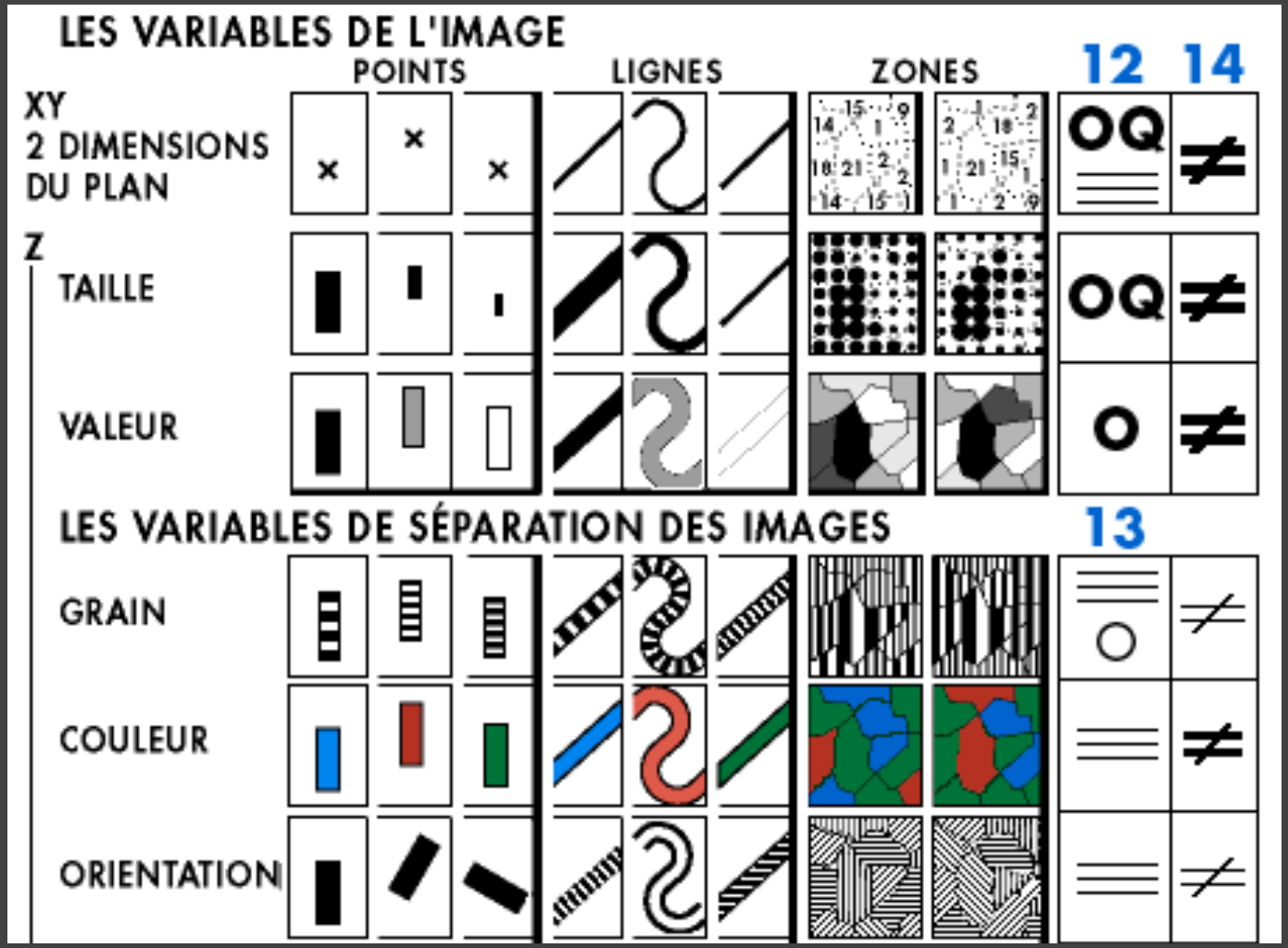
   What do people perceive/comprehend?
   How do visualizations inform mental models?

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

   Leverage perception & augment cognition
   Improve ties between visualization & mental model

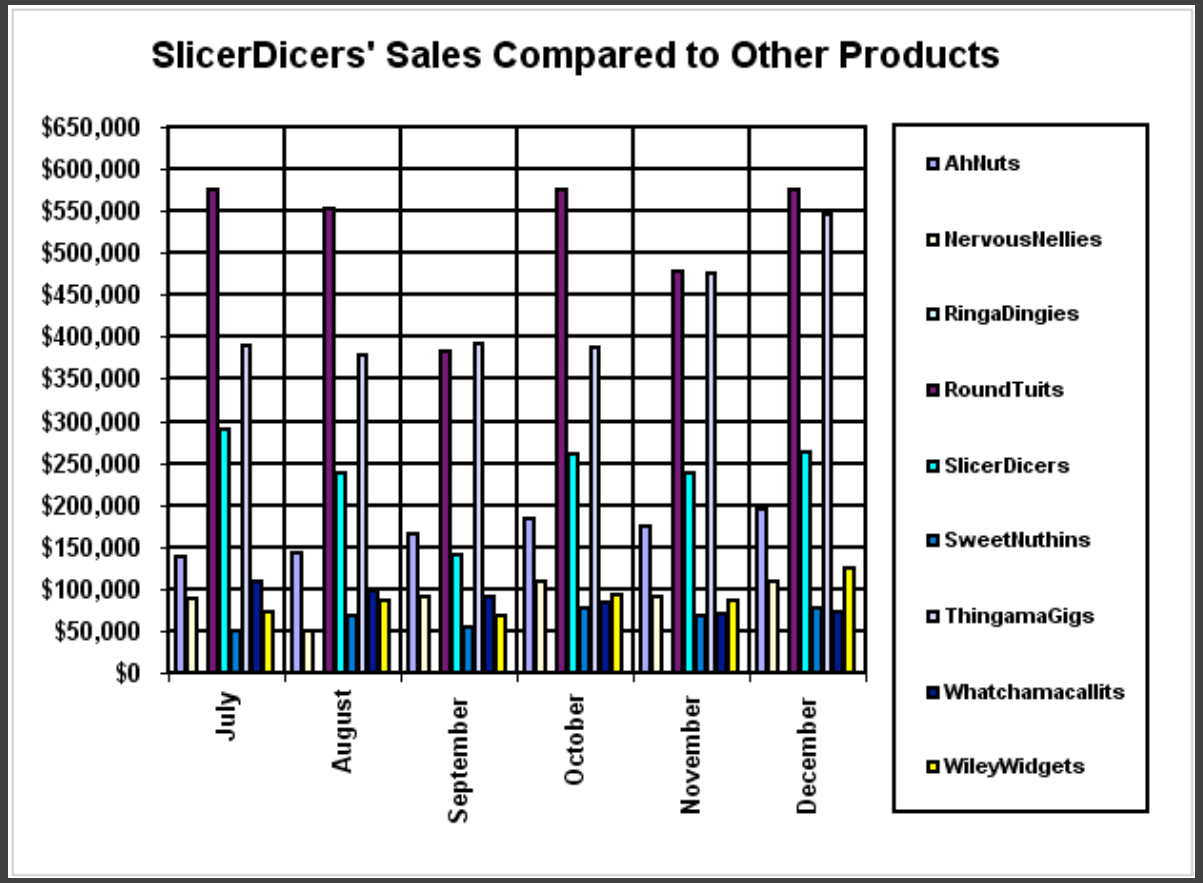
Course Topics

Data and Image Models

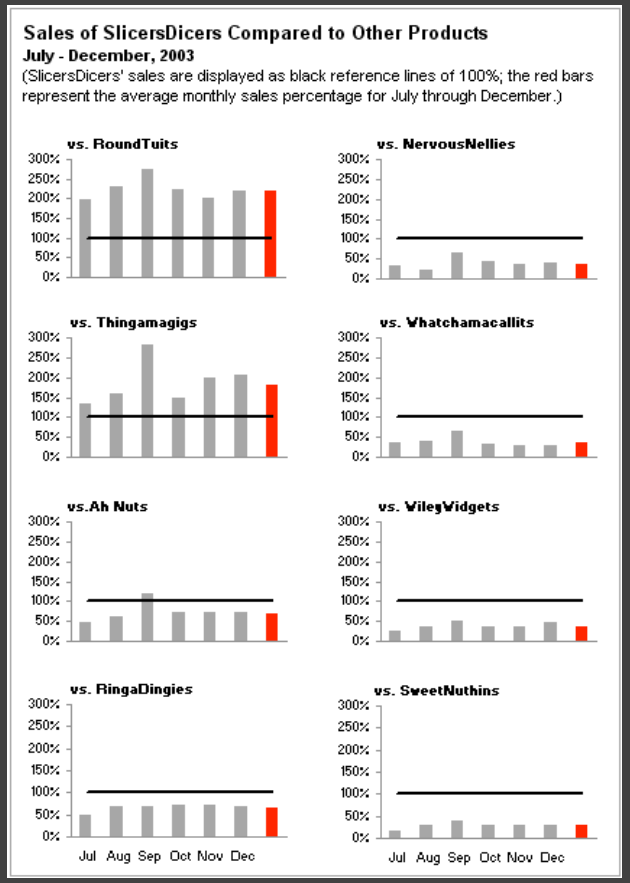


Sémiologie Graphique [Bertin 67]

Visualization Design

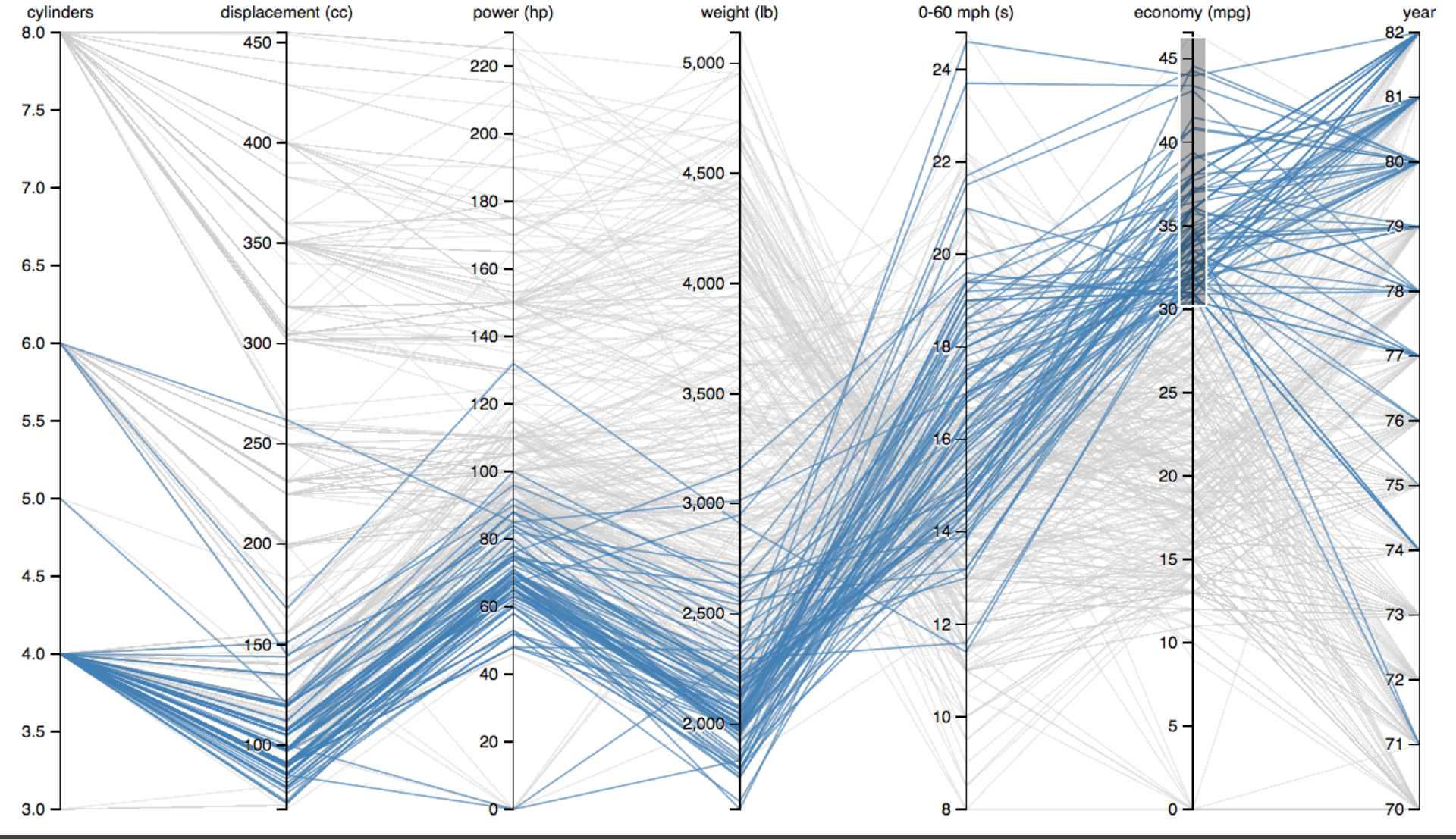


Problematic design

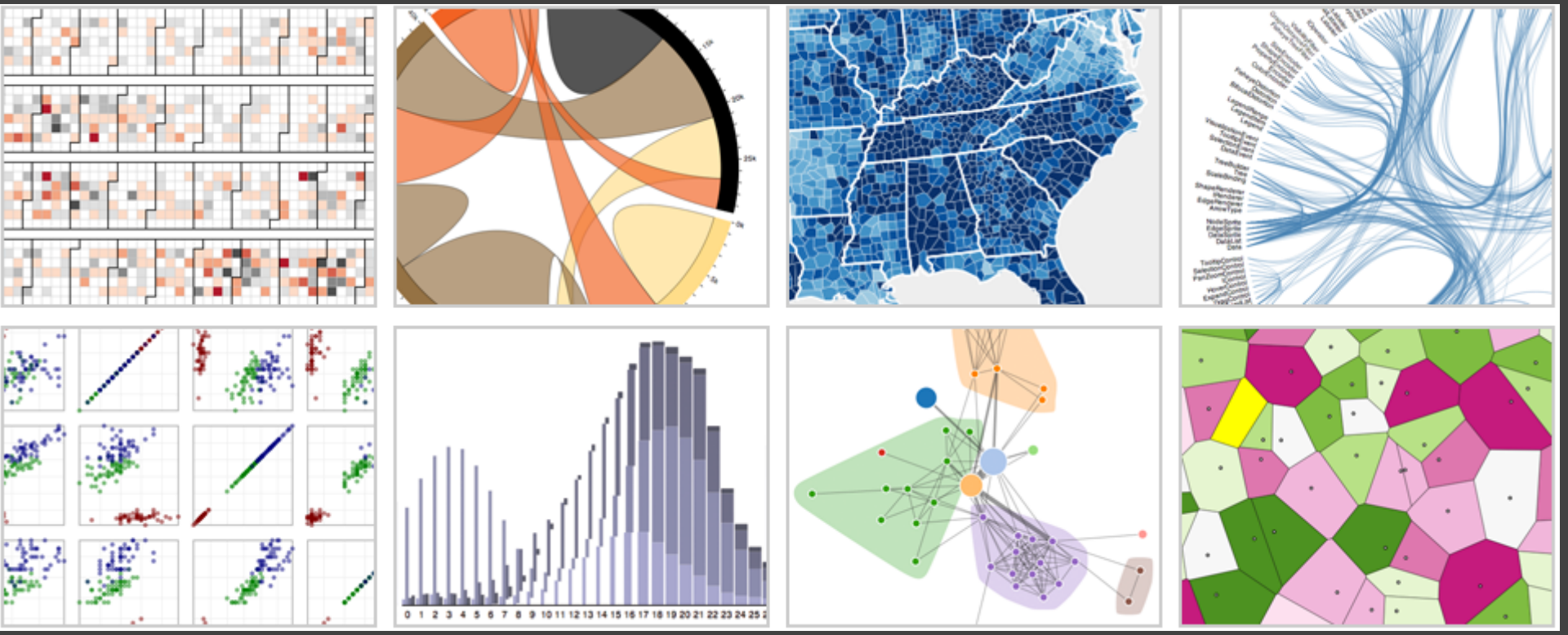


Redesign

Exploratory Data Analysis

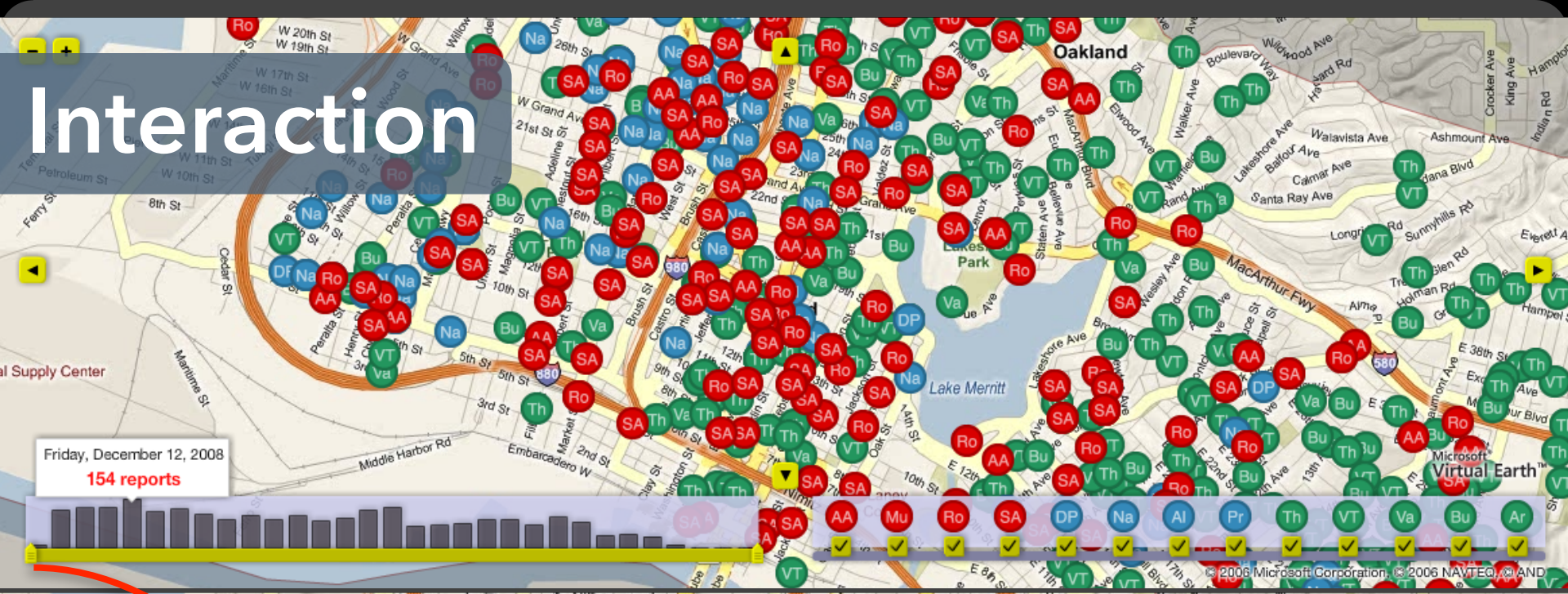


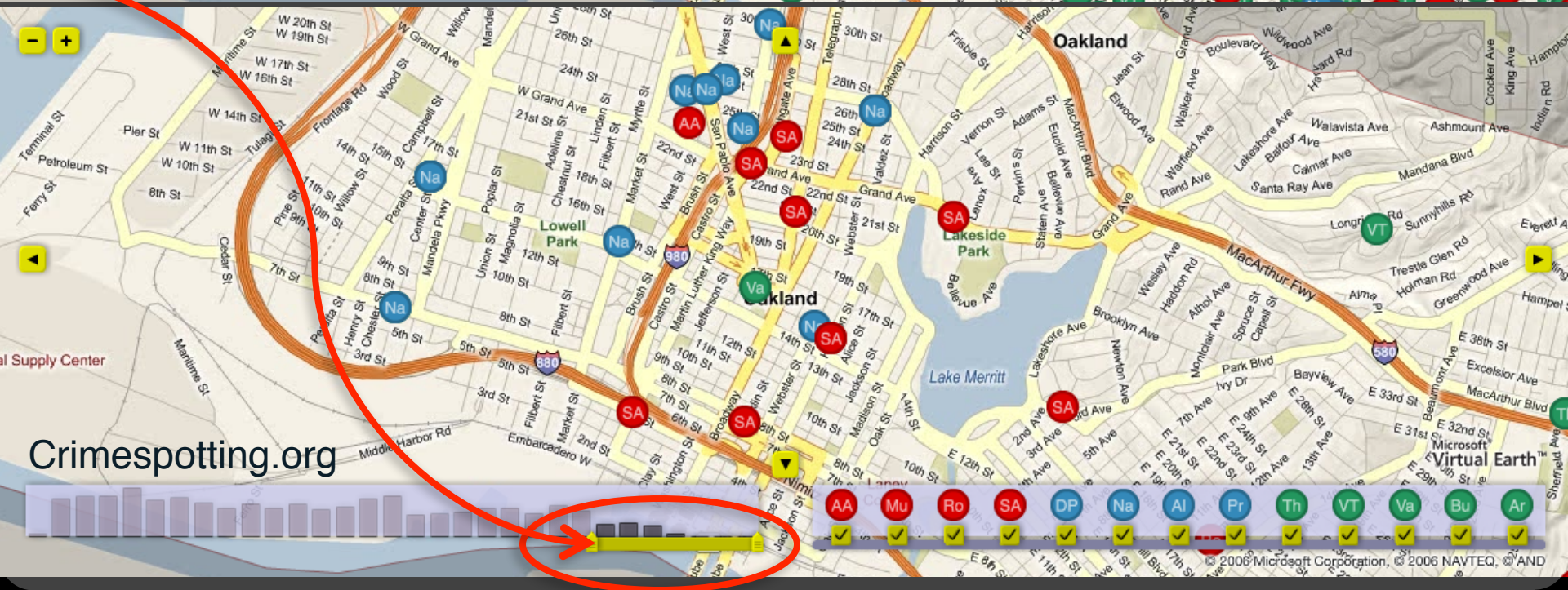
Visualization Software



D3: Data-Driven Documents

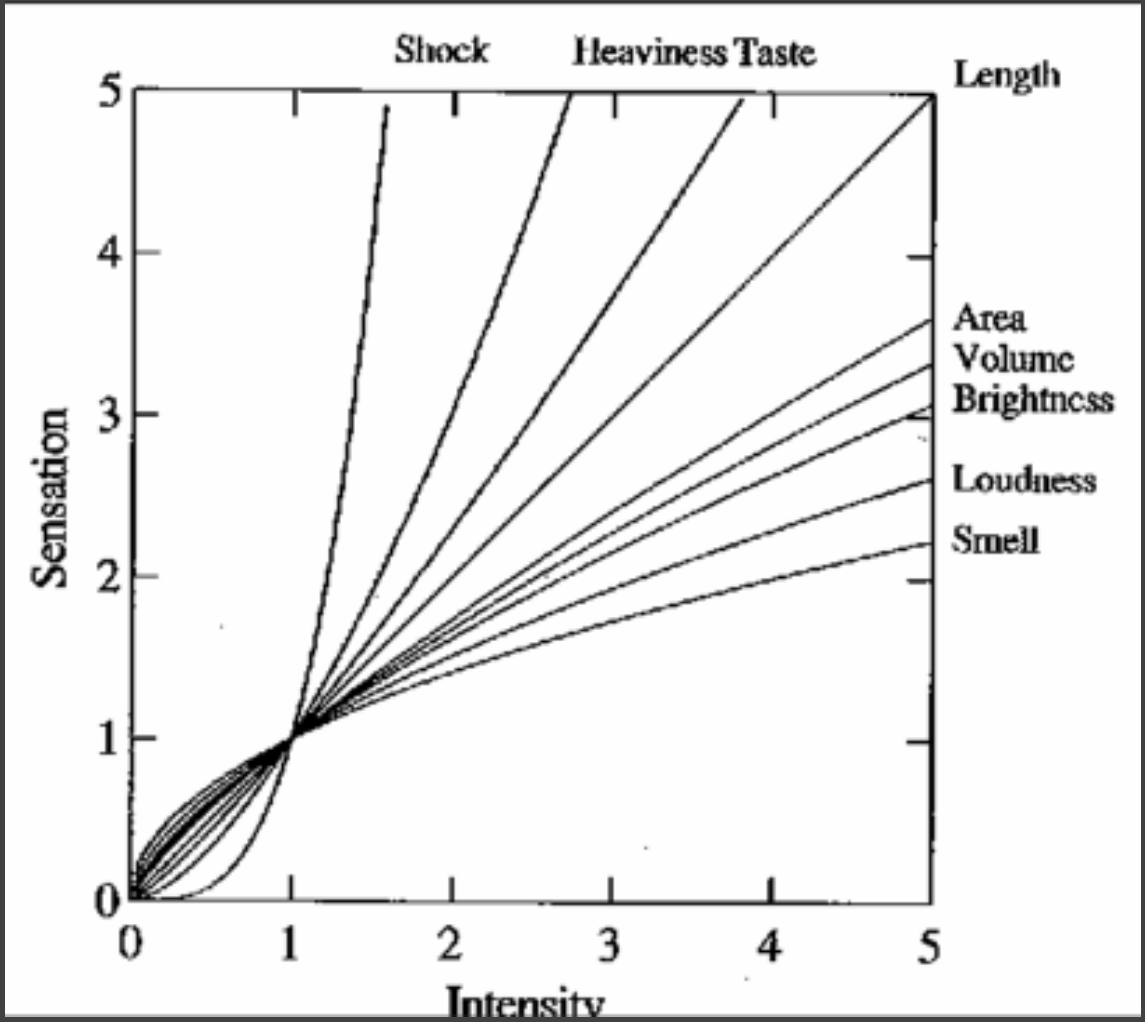
Interaction





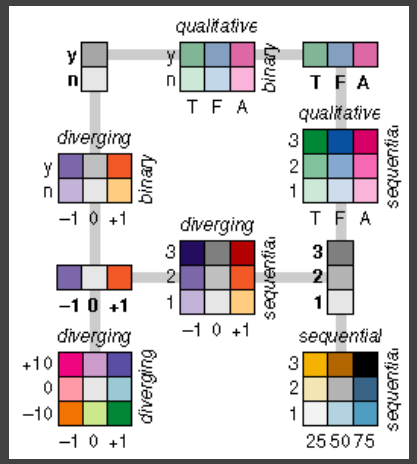
Crimespotting.org

Graphical Perception

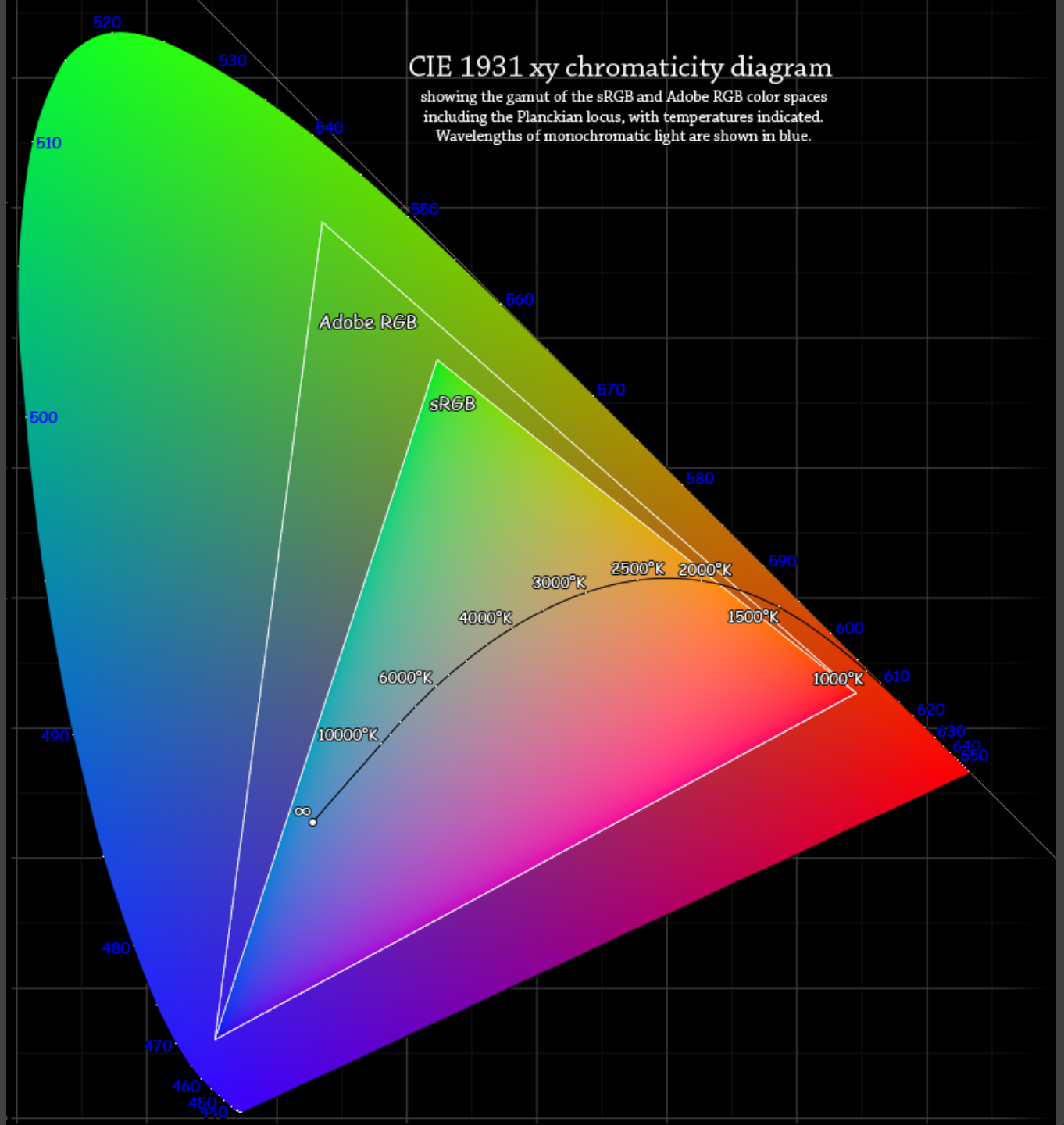


The psychophysics of sensory function [Stevens 61]

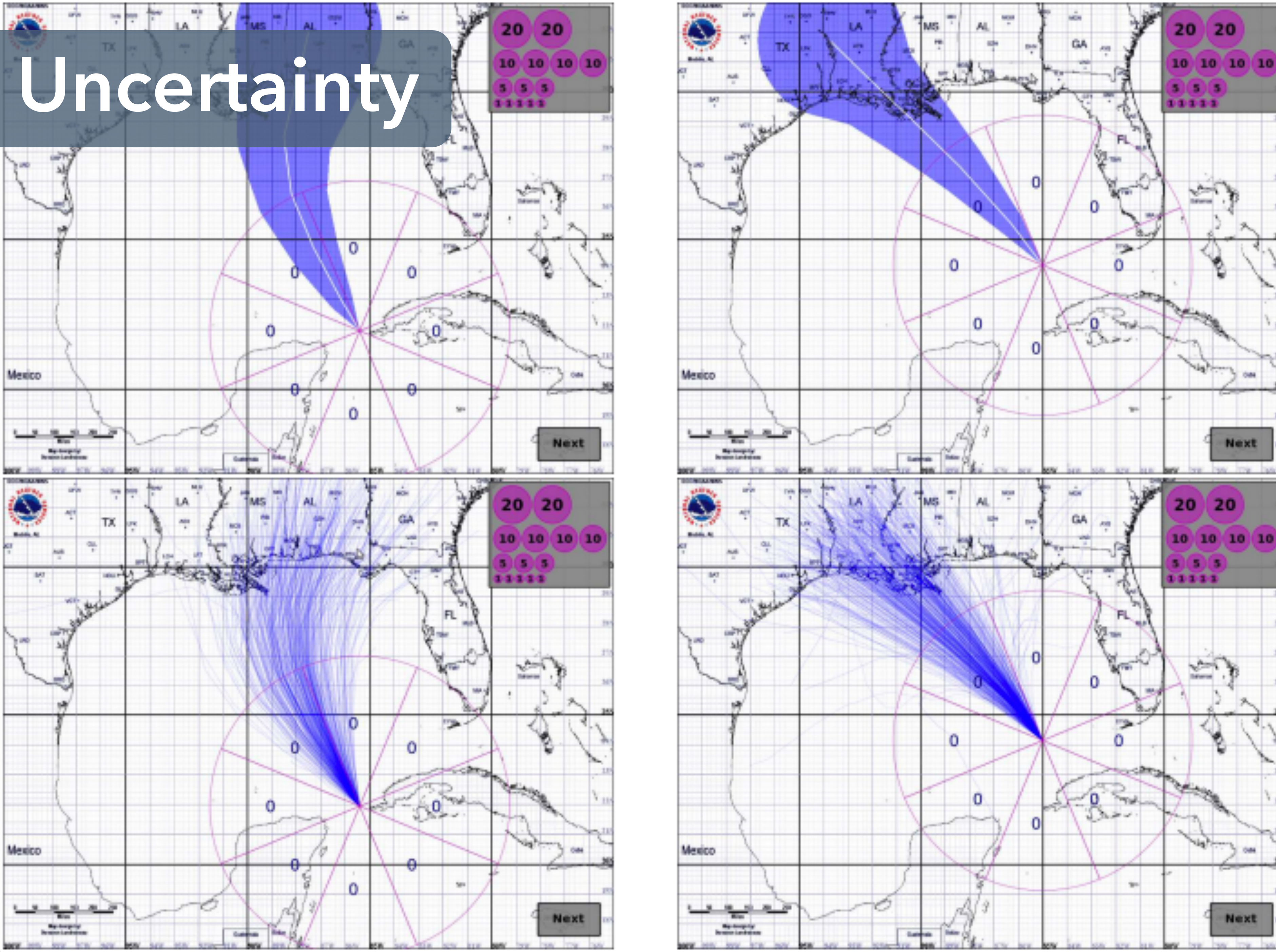
Color



Color Brewer



Uncertainty



Narrative

Recent elections have placed a heavy emphasis on "swing states" — Ohio, Florida and the other competitive states. Yet in the past, many more states shifted between the Democratic and Republican parties. A look at how the states stuck with their political traditions and how they have shifted over past elections.

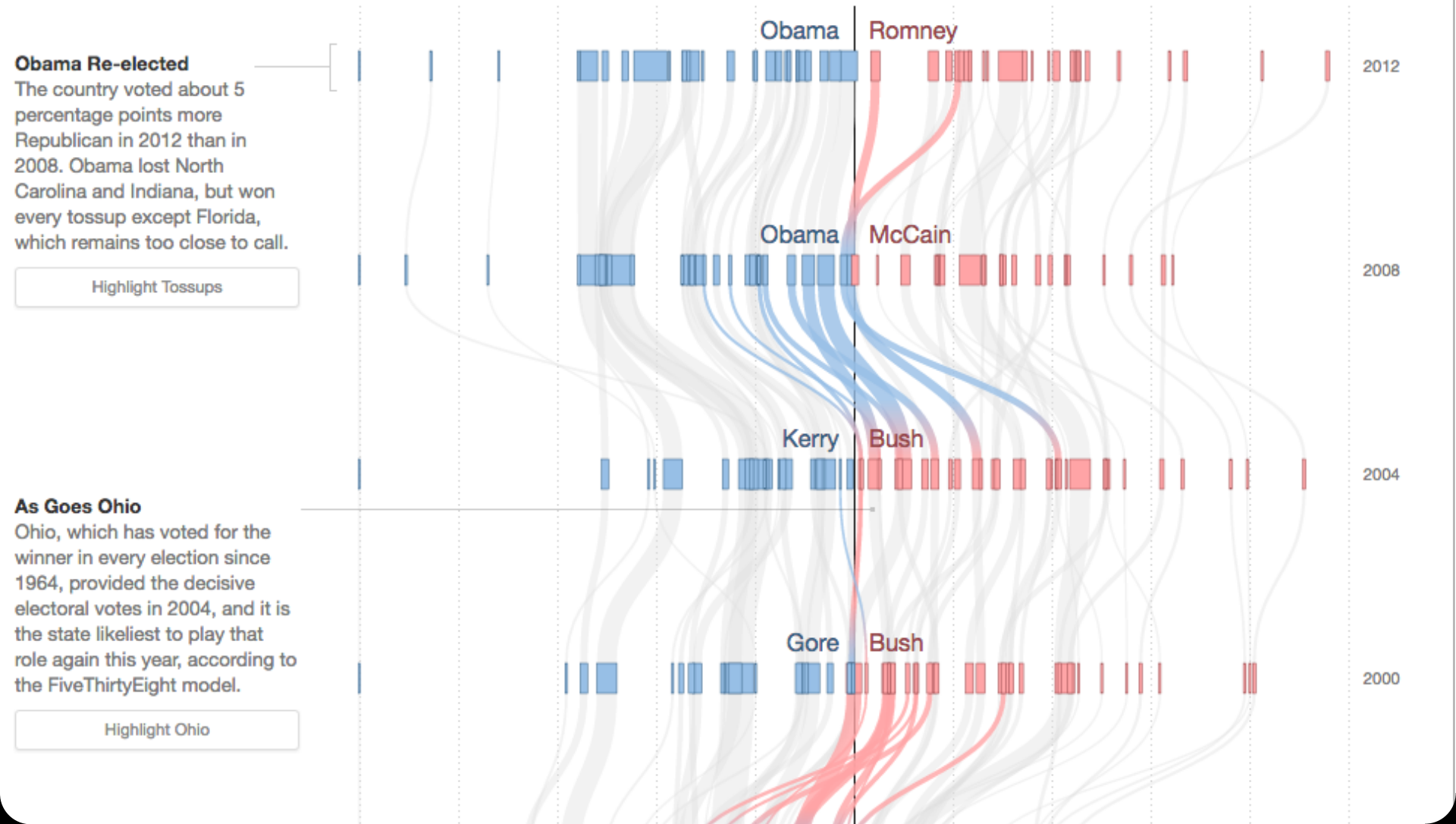
Each box represents a state sized by number of electoral votes.

Each curve shows how much it shifted left or right between elections

Chart **Size of Lead** | Chart **Electoral Votes**

← MORE DEMOCRATIC | MORE REPUBLICAN →

≥50% +40% +30% +20% +10% | +10% +20% +30% +40% ≥50%



Obama Re-elected

The country voted about 5 percentage points more Republican in 2012 than in 2008. Obama lost North Carolina and Indiana, but won every tossup except Florida, which remains too close to call.

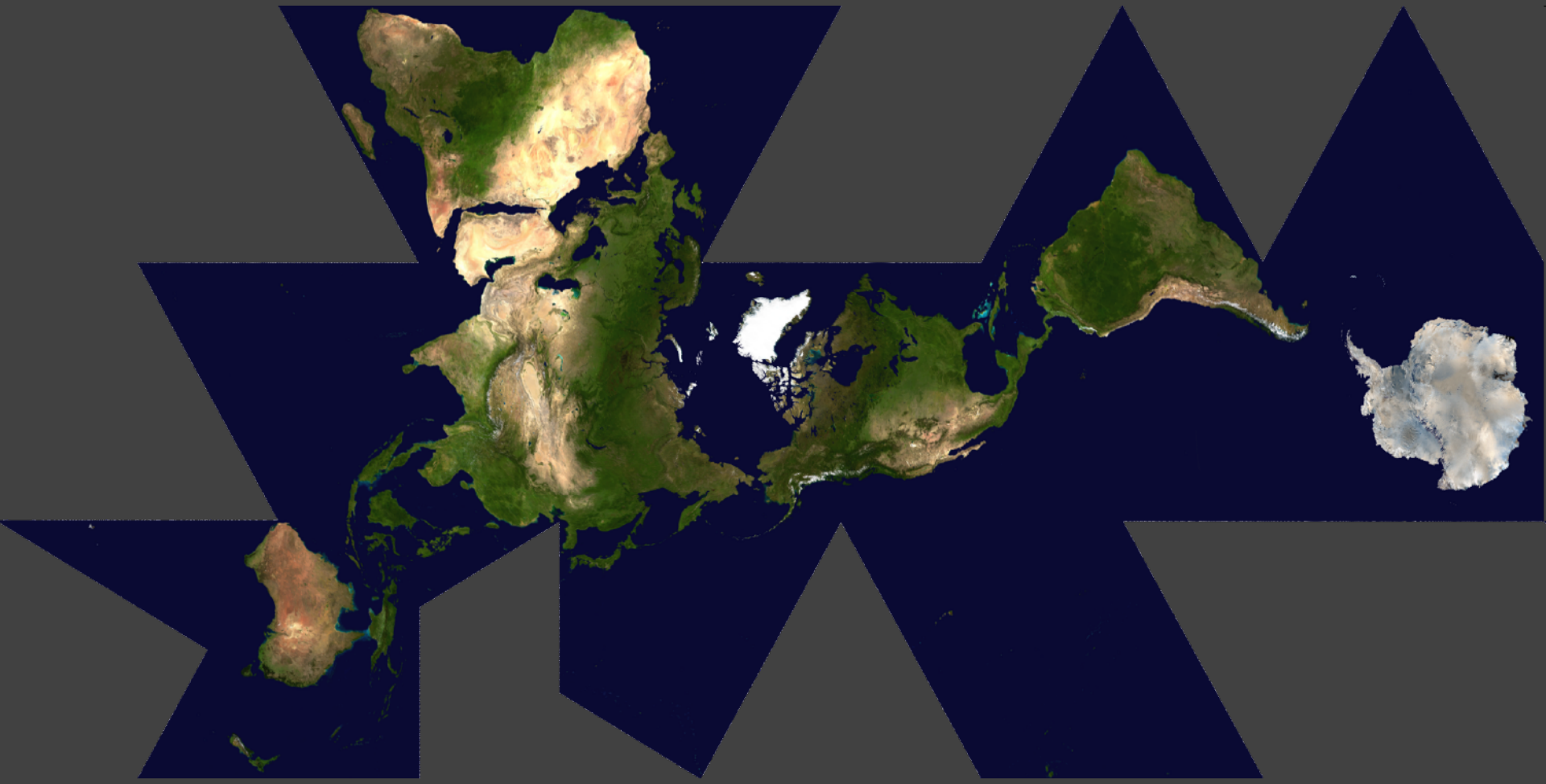
Highlight Tossups

As Goes Ohio

Ohio, which has voted for the winner in every election since 1964, provided the decisive electoral votes in 2004, and it is the state likeliest to play that role again this year, according to the FiveThirtyEight model.

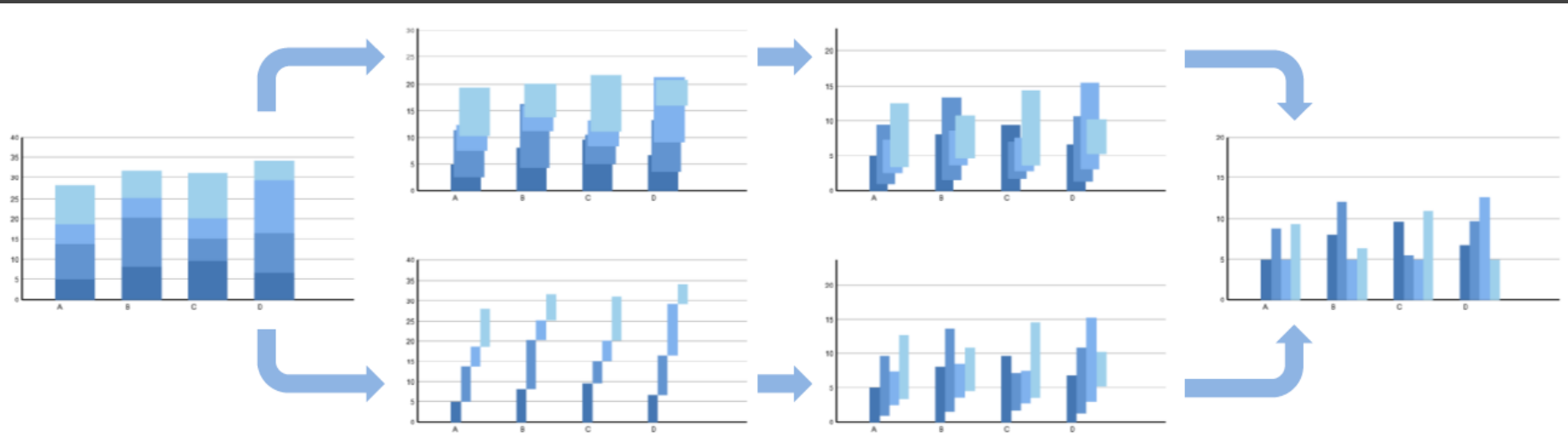
Highlight Ohio

Maps



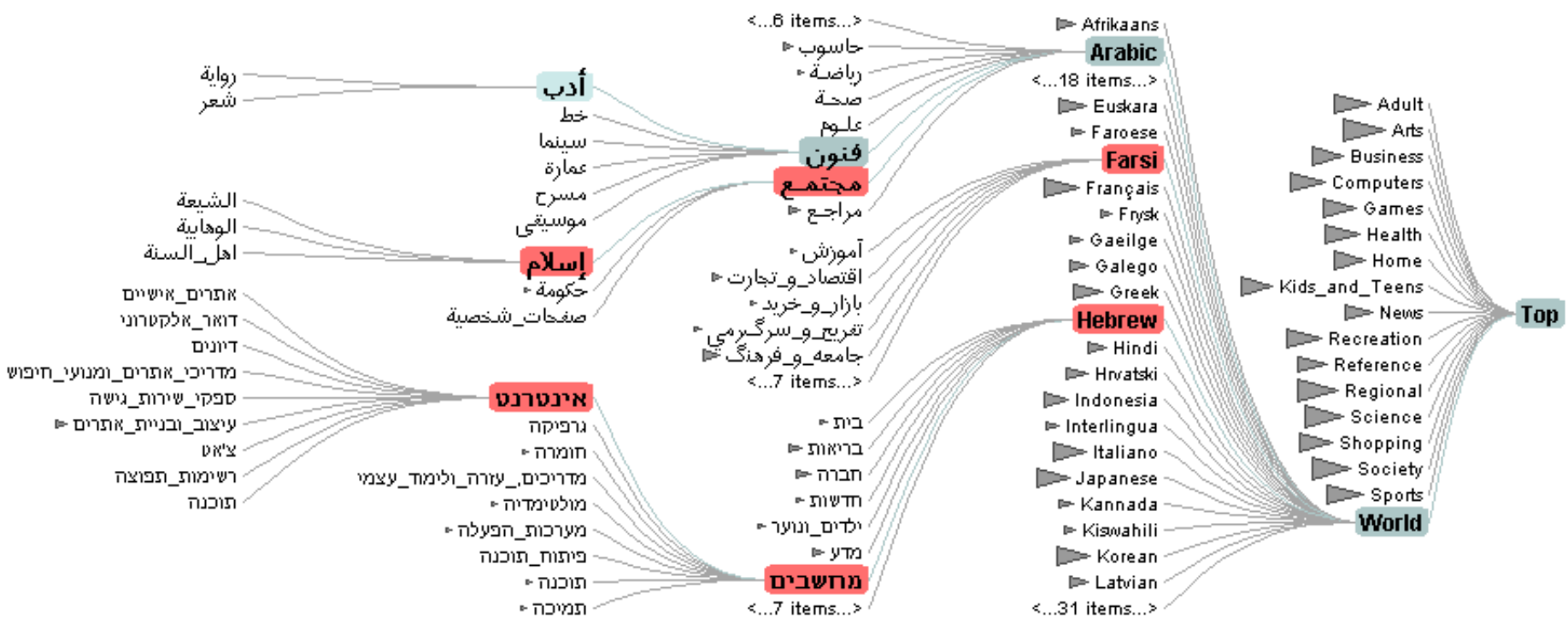
Dymaxion Maps [Fuller 46]

Animation



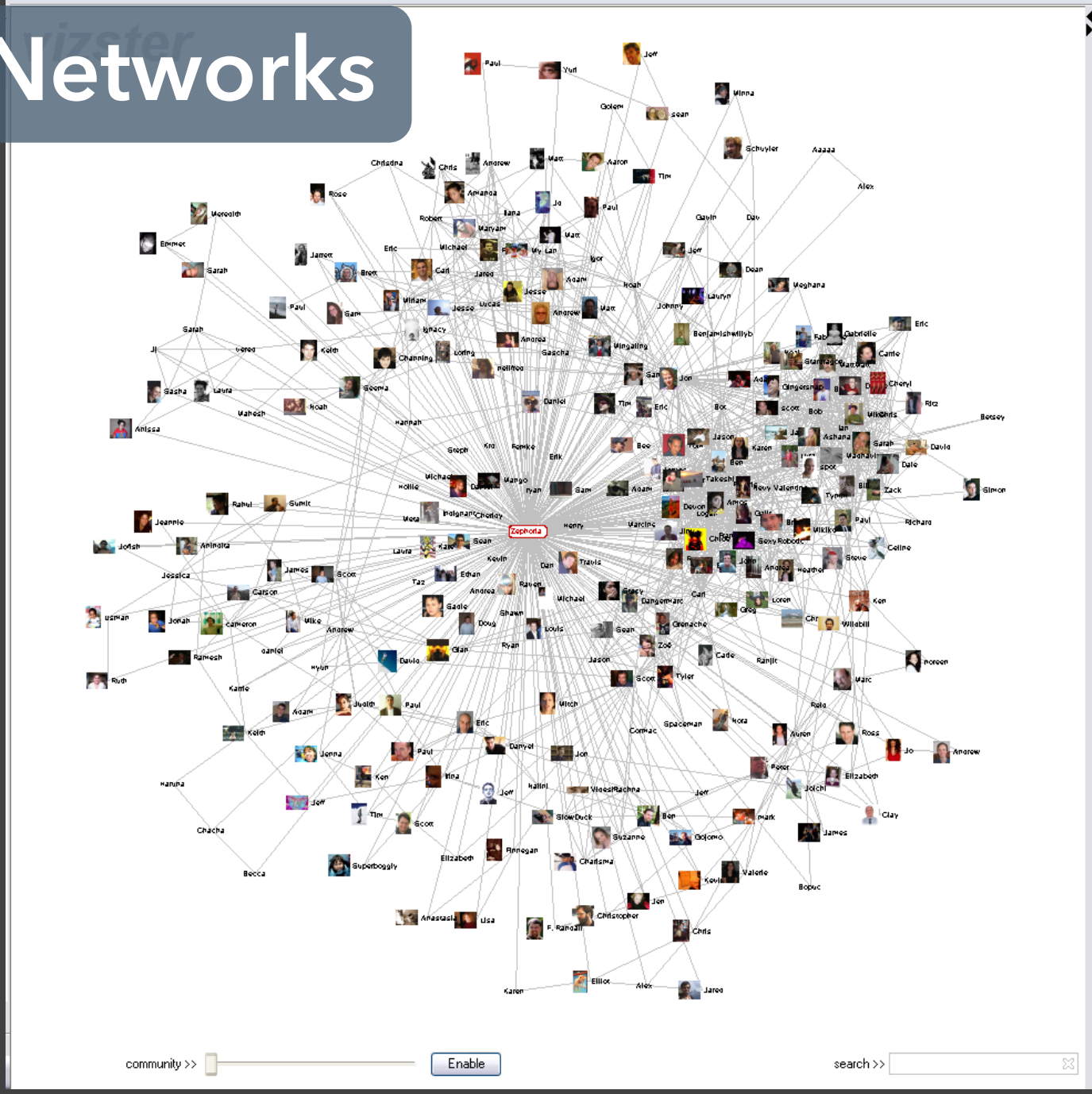
Animated transitions in statistical data graphics [Heer & Robertson 07]

Hierarchies



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

Networks



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

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

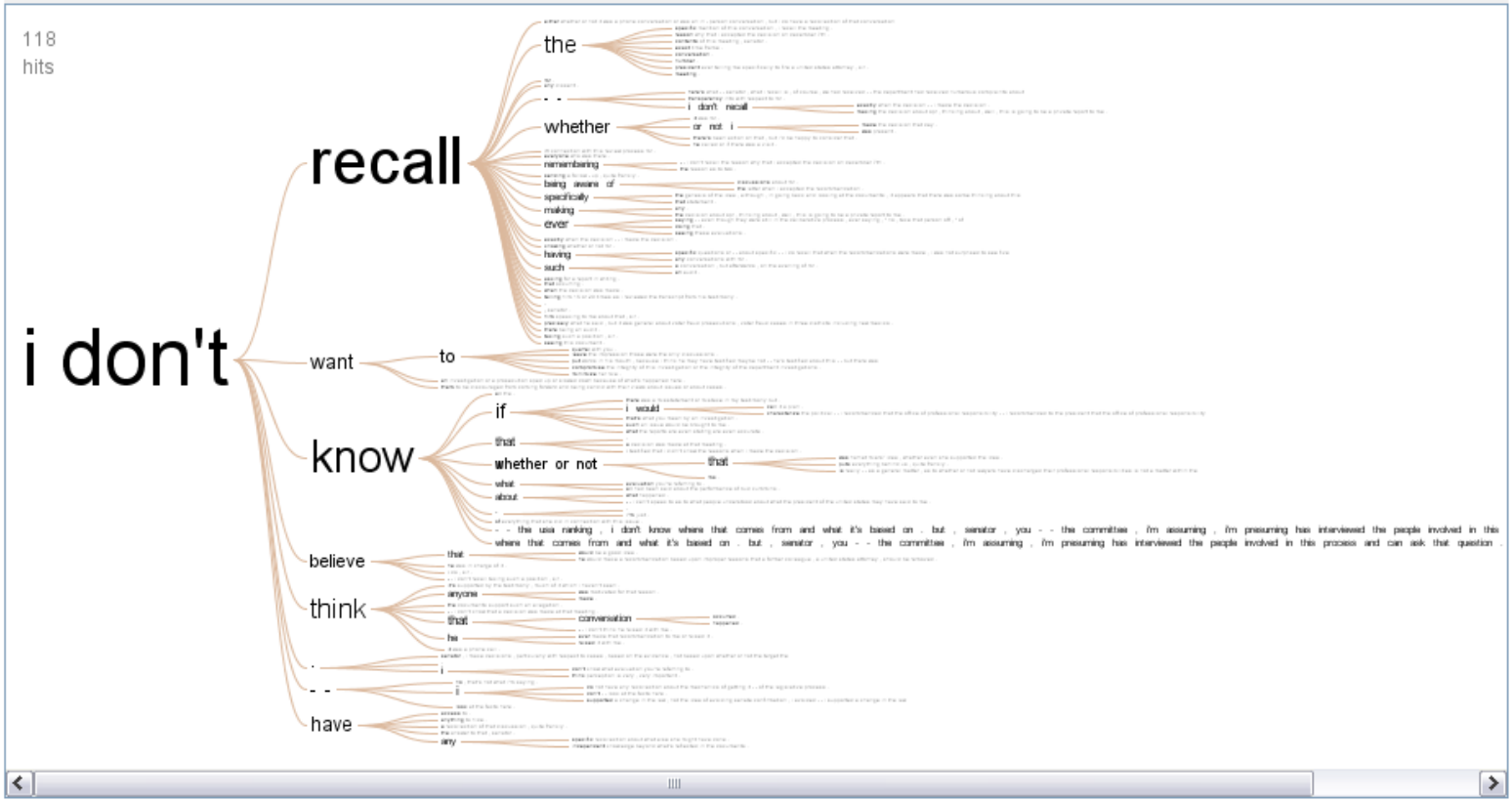
legal
terms of use

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

Visualizations : Word tree / Alberto Gonzales

Creator: Martin Wattenberg
Tags:

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

Course Mechanics

You should expect to:

1. *Evaluate and critique* visualization designs
2. *Learn* visualization techniques & theory
3. *Implement* interactive data visualizations
4. *Develop* a substantial visualization project

Instructors

cse442@cs

Instructor

Jeffrey Heer — OH: *Th 9:15-10:15a, 642 CSE*
Assoc Prof, CSE — http://jheer.org

Assistants

Matt Conlen — OH: *Mon 1-2pm, 4th Floor CSE*

Younghoon Kim — OH: *Fri 5:30-6:30pm, 3rd Floor*

Kaitlyn Zhou — OH: *By appointment*

Matthew Conlen

mconlen@cs.washington.edu // @mathisonian
OH: Monday 1-2p CSE 4th Floor Breakout

Experience as a computational journalist @ FiveThirtyEight

Research on interactive documents: https://idyll-lang.github.io/

Lots of experience with JavaScript, D3, and general web programming

Younghoon Kim

yhkim01@cs.washington.edu

Office Hours

Friday 5:30 - 6:30 p.m.
3rd Floor Breakout
(Out of Town 10/6)

Hi! I'm a 3rd year Ph.D. student interested in algorithms for visualization recommendation and data storytelling!

Kaitlyn Zhou

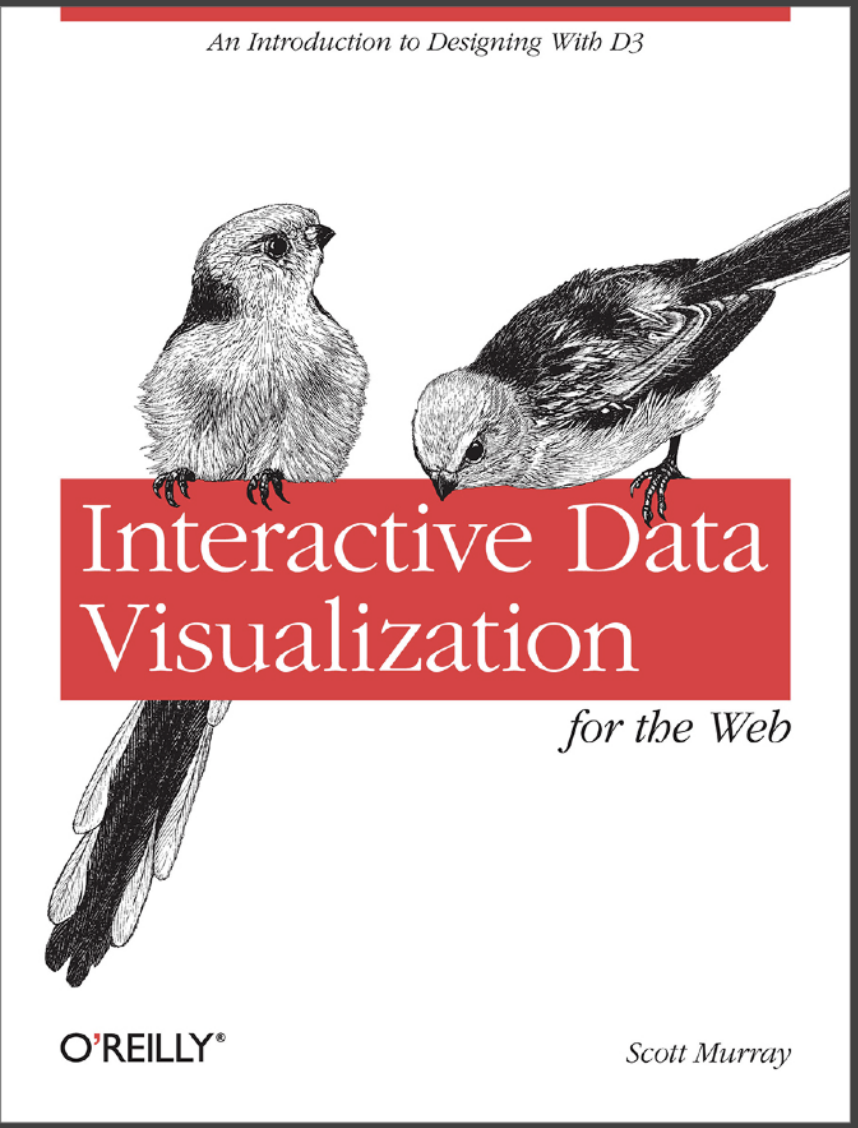
Senior studying CS & HCDE.

My 442 final project investigated inequalities in Seattle public schools.

I am helping to start a Computer Science Student Advisory Council - would love to see you all involved!

Fun fact: I went on a road trip from Seattle to Atlanta this past summer.

Textbook



Interactive Data Visualization for the Web, 2nd Edition

For learning D3!
Book available online.
Code / examples on GitHub.

We will be using **D3 v4**.
https://d3js.org

Readings

Some from D3 book, others from papers & web.

Material in class will loosely follow readings.

Readings should be read by start of class.

Post discussion comments on class Canvas forum.

- One comment per week (ending week 8).
- Comments must be posted by Friday 11:59pm.
- You have 1 “pass” for the quarter.

Assignments

Class Participation (10%)

A1: Visualization Design (10%) - *Due 10/2*

A2: Exploratory Data Analysis (15%) - *Due 10/16*

A3: Interactive Prototype (25%) - *Due 10/30*

- Peer Evaluation - *Due 11/6*

FP: Final Project (40%)

- Initial Prototypes - *Due 11/20*
- Peer Evaluation - *Due 11/27*
- Project Deliverables - *Due 12/6*

Final Project

Produce **interactive web-based visualizations**

Initial **prototype** and **peer evaluation**

Design reviews and **final presentation**

Submit and publish on GitHub

Projects from **previous classes** have been:

- Published as research papers
- Featured in the New York Times
- Released as successful open source projects

Final Project Theme

The UW Atlas of Illustrated Algorithms

Goal: use data visualization to produce an interactive narrative that explains the workings of a chosen algorithm. This can include both behavior and runtime analysis.

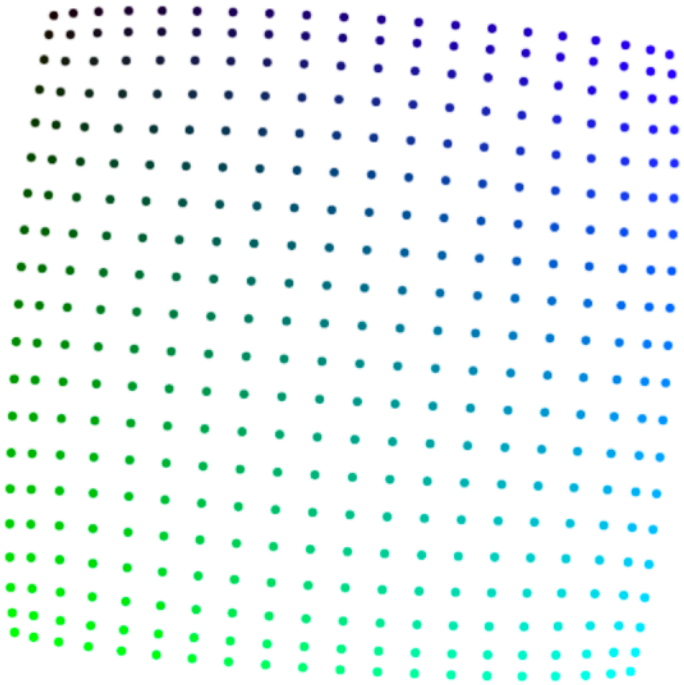
Potential examples include: network protocols, scheduling protocols, randomized algorithms, the Fast Fourier Transform, Voronoi tesselation, backpropagation, …

We will collectively produce an online book as a resource for people to better understand algorithms.

Inspiration…

How to Use t-SNE Effectively

Although extremely useful for visualizing high-dimensional data, t-SNE plots can sometimes be mysterious or misleading. By exploring how it behaves in simple cases, we can learn to use it more effectively.



Step 1,910

A square grid with equal spacing between points. Try convergence at different sizes.

Points Per Side 20

Perplexity 10

Epsilon 5

MARTIN WATTENBERG Google Brain
FERNANDA VIÉGAS Google Brain
IAN JOHNSON Google Cloud

Oct. 13 2016

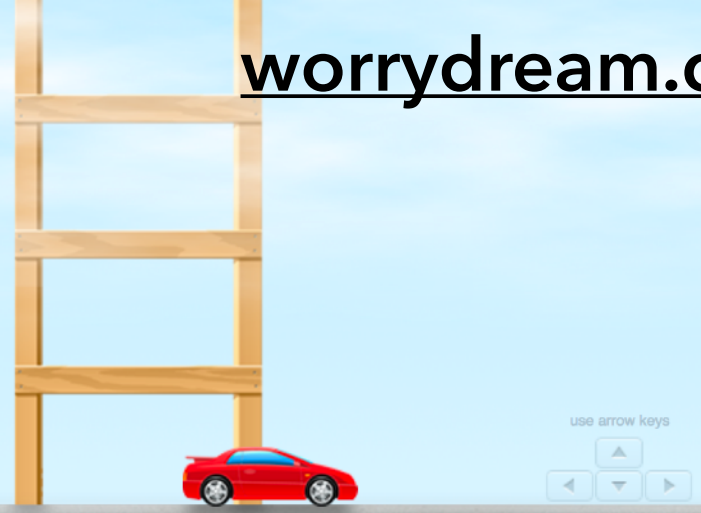
Citation: Wattenberg, et al., 2016

distill.pub

worrydream.com

Up and Down the Ladder of Abstraction

A Systematic Approach to Interactive Visualization



Bret Victor / October, 2011

"In science, if you know what you are doing, you should not be doing it. In engineering, if you do not know what you are doing, you should not be doing it. Of course, you seldom, if ever, see either pure state."

—Richard Hamming, *The Art of Doing Science and Engineering*

How can we design systems when we don't know what we're doing?

The most exciting engineering challenges lie on the **boundary of theory and the unknown**. Not so unknown that they're hopeless, but not enough theory to predict the results of our decisions. Systems at this boundary often rely on *emergent behavior* — high-level effects that arise indirectly from low-level interactions.

When designing at this boundary, the challenge lies not in constructing the system, but in understanding it. In the absence of theory, we must develop an *intuition* to guide our decisions. The design process is thus one of exploration and discovery.

How do we explore? If you move to a new city, you might learn the territory by walking around. Or you might peruse a map. But far more effective than either is *both together* — a street-level experience with higher-level guidance.

Likewise, the most powerful way to gain insight into a system is by *moving between levels of abstraction*. Many designers do this instinctively. But it's easy to get stuck on the ground, experiencing concrete systems with no higher-level view. It's also easy to get stuck in the clouds, working entirely with abstract equations or aggregate statistics.

This interactive essay presents the **ladder of abstraction**, a technique for thinking *explicitly* about these levels, so a designer can move among them consciously and confidently.

I believe that an essential skill of the modern system designer will be using the *interactive medium* to move fluidly around the ladder of abstraction.

Standing on Concrete

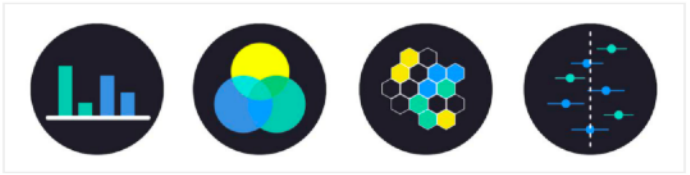
We'll start with an in-depth example — designing the control system for a **simple car simulation**. Our goal is to write a set of

Explorable Explanations

★ **Back to the Future of Handwriting Recognition**

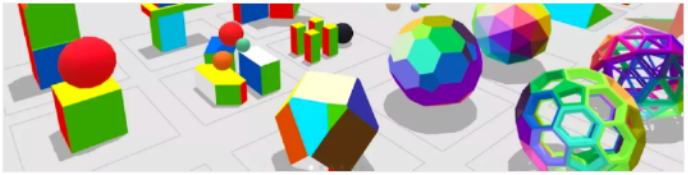
programming

an interactive history of handwriting recognition software



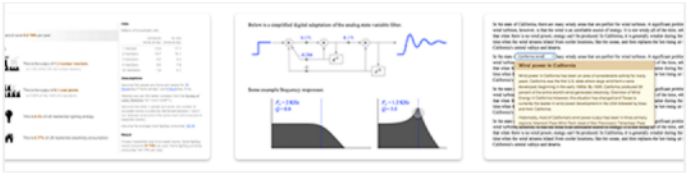
★ **Seeing Theory** math

a visual introduction to probability and stats



★ **4D Toys** math

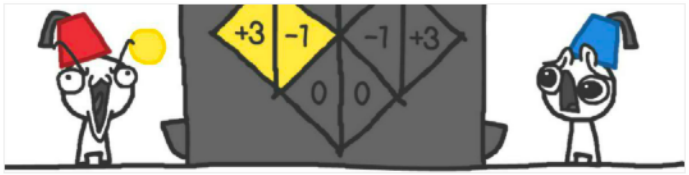
an interactive toy for 4D children



★ **Explorable Explanations**

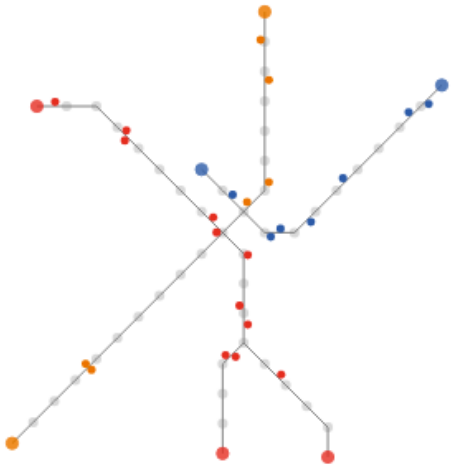
further reading

the 2011 essay that coined the phrase 'Explorable Explanation'



★ **The Evolution of Trust** social science

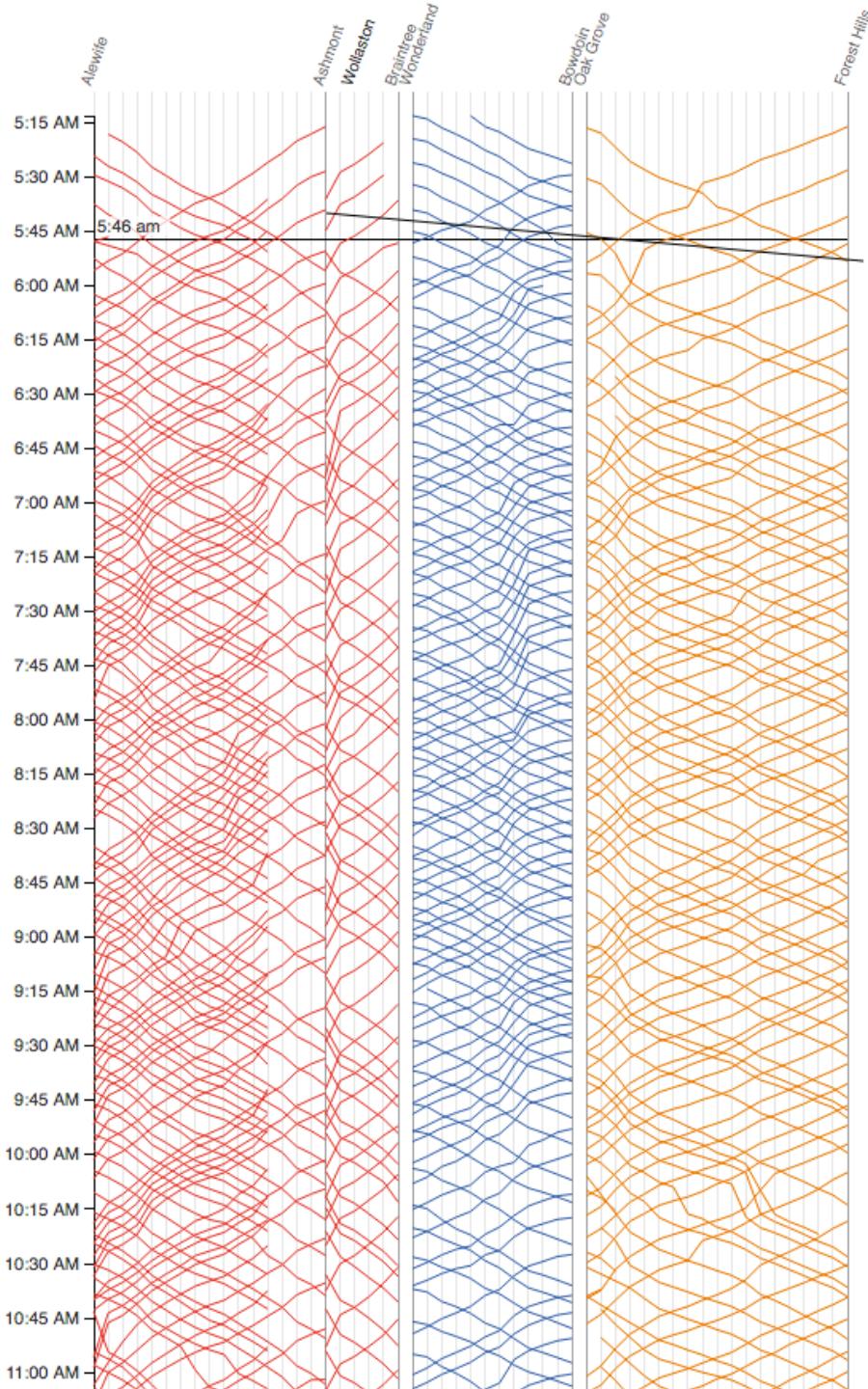
★ **Fireflies** biology



Locations of each train on the red, blue, and orange lines at 5:46 am. Hover over the diagram to the right to display trains at a different time.

Trains are on the right side of the track relative to the direction they are moving.

See the morning rush-hour, midday lull, afternoon rush-hour, and the evening lull.



Service starts at 5AM on Monday morning. Each line represents the path of one train. Time continues downward, so steeper lines indicate slower trains.

Since the red line splits, we show the Ashmont branch first then the Braintree branch. Trains on the Braintree branch "jump over" the Ashmont branch.

Train frequency increases around 6:30AM as morning rush hour begins.

MBTA Viz

Barry & Card

Questions?

A1: Visualization Design

Design a static visualization for a data set.

Every 10 years, the census documents the demographic make-up of the U.S., influencing congressional districting and social services. This dataset contains a summary of census data for two years a century apart: 1900 and 2000.

You must choose the message you want to convey. What question(s) do you want to answer? What insight do you want to communicate?

A1: Visualization Design

Pick a **guiding question**, use it to title your vis.
Design a **static visualization** for that question.
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 **8:00 pm, Monday October 2**.

Course Waitlist

Hoping to enroll in the course?

CSE Undergrad Advising will release a link for an overload list on **Friday 9/28**.

If you are enrolled late, we can grant extensions for discussion forum posts and A1. However, we encourage you to complete A1 on time – it's a useful and short exercise regardless!