

CSE 442 - Data Visualization

# The Value of Visualization

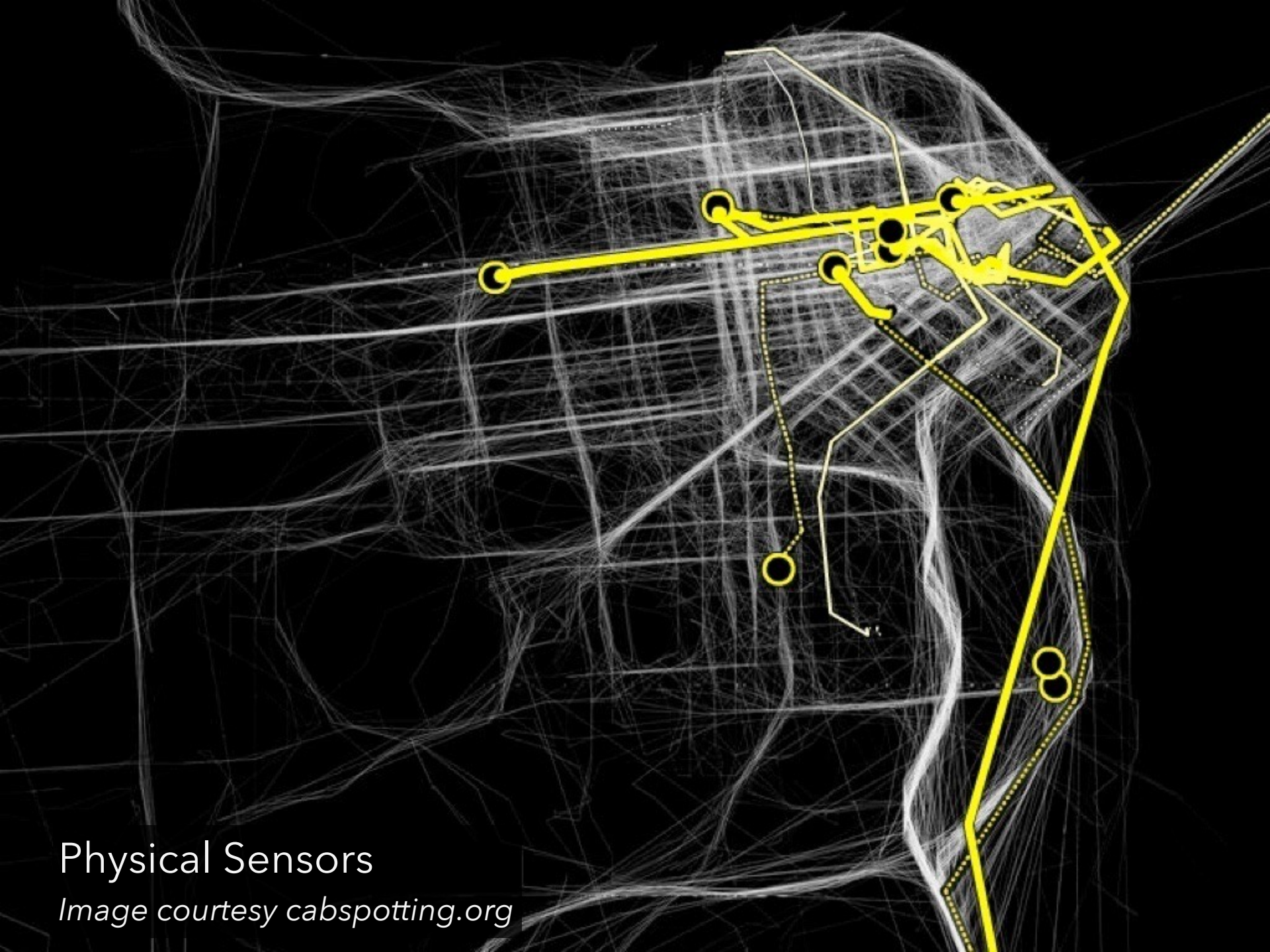


Jeffrey Heer University of Washington

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

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

Gantz et al, 2008, 2010

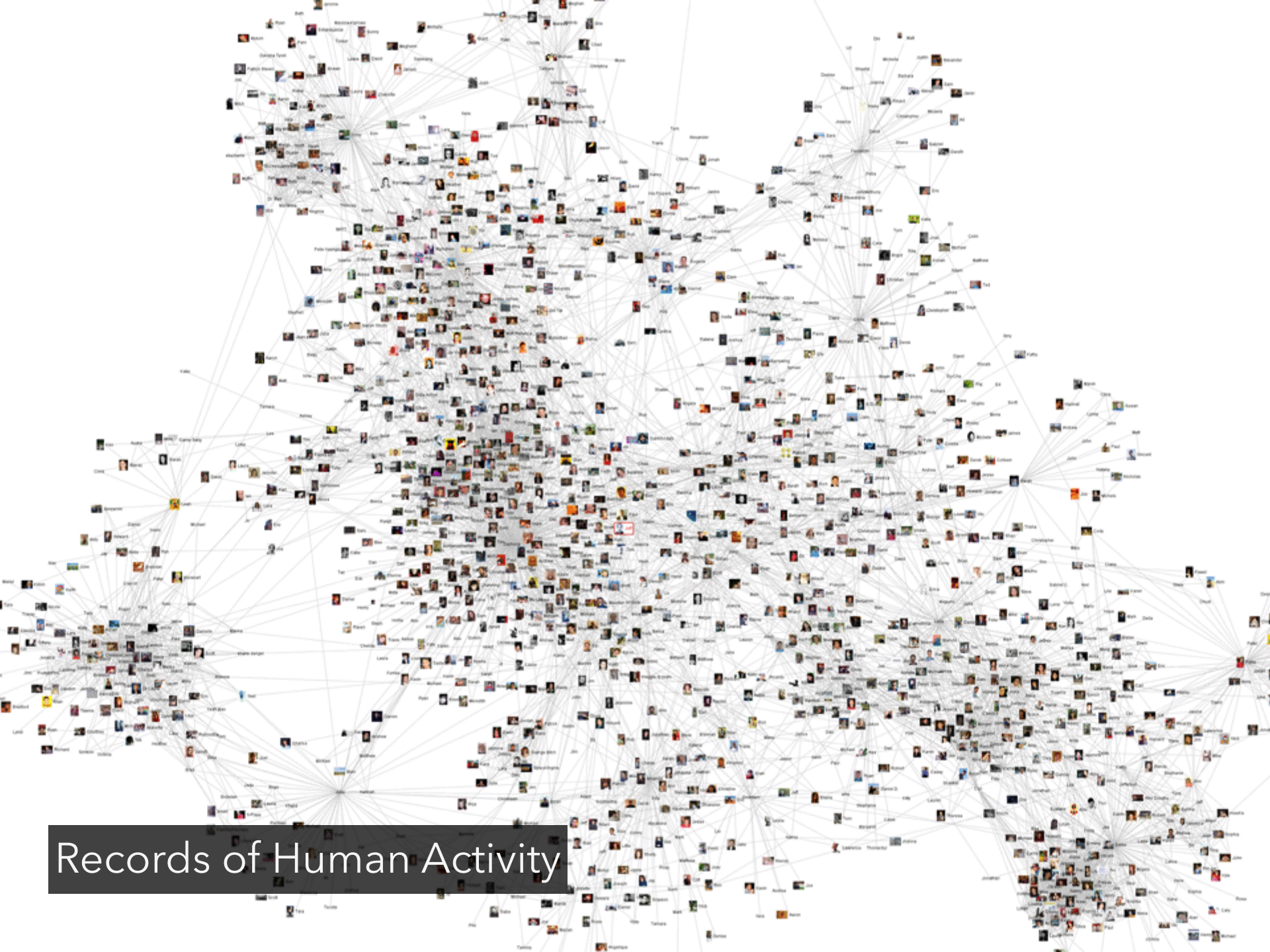


# Physical Sensors

*Image courtesy cabspotting.org*



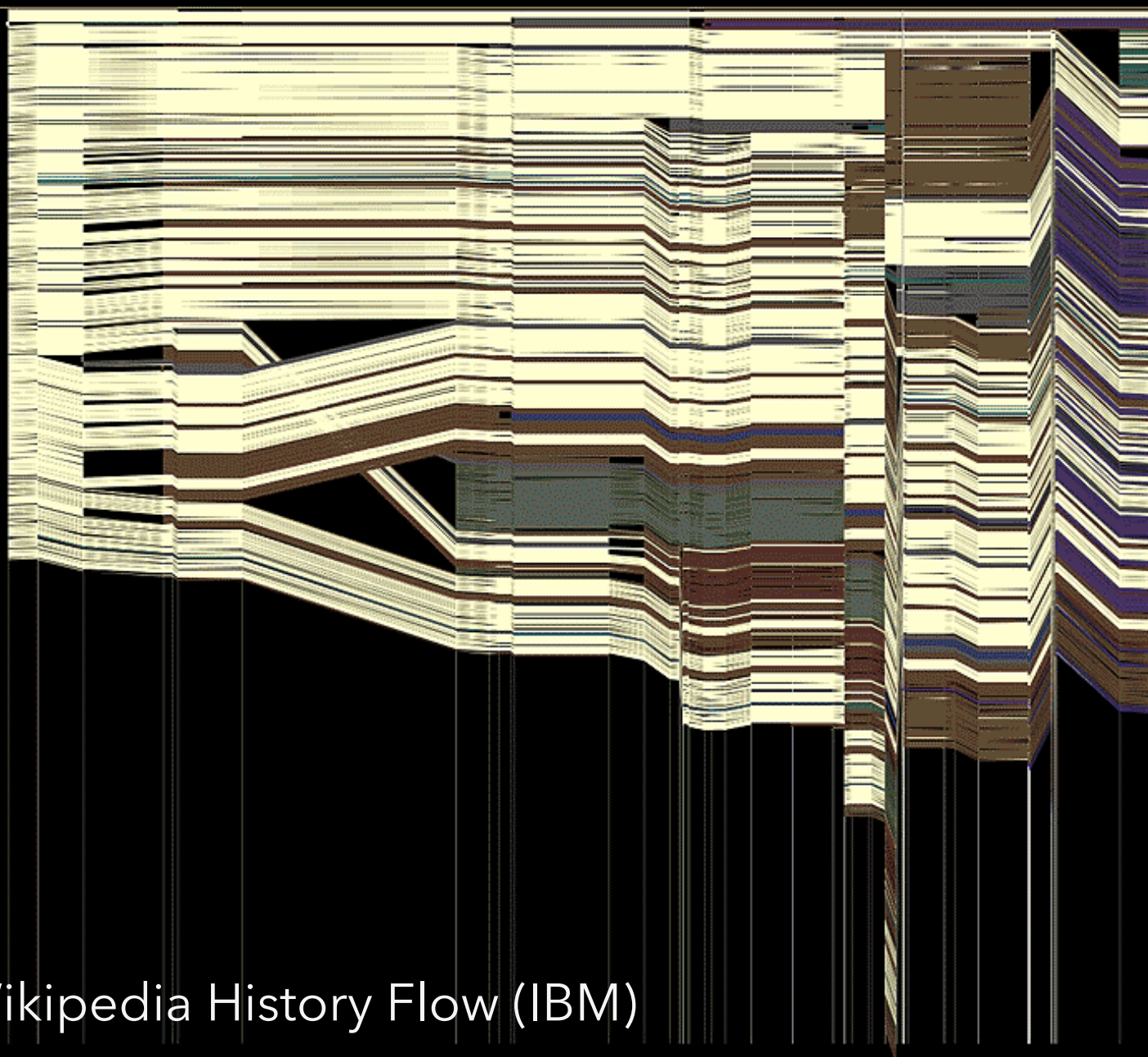
Health & Medicine



Records of Human Activity

authors posts

Zundark	1
The Cunctator	1
The Epost	1
Conversion script	1
RK	1
Freob	1
B4hand	1
KamikazeArchon	1
Stephen Gilbert	1
Slubenstein	8
Mimccorn	5
Iels	1
Derek Ross	1
Dante Alighieri	2
Maverick149	3
Jazzbug	2
Jzdrl	8
Theanthrope	1
Wesley	2
Dreamword	1
Stevrigo	4
Canembert	1
Hephaestus	2
Zoe	1
MyRedDice	1
G-Man	2
Kingturtle	1
Montrealais	1
...	1



### 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**, **fetus**, **[1]** Medically, the term also refers to early termination of a pregnancy by nature ("spontaneous abortion" or **miscarriage**), 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 is performed by a number of different methods. The earliest terminations (before nine weeks) are **chemical abortions** is the usual method, though **mifepristone** is usually the only legal method, although research has uncovered similar effects from **methotrexate** and **misoprostol**. Concern with chemical abortion and extending up until around the fifteenth week **suction-aspiration** 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 & E)** is used.

As the fetus size increases other techniques may 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 (D & X)** or a **hysterotomy abortion**, similar to **Caesarian section**.

#### The controversy

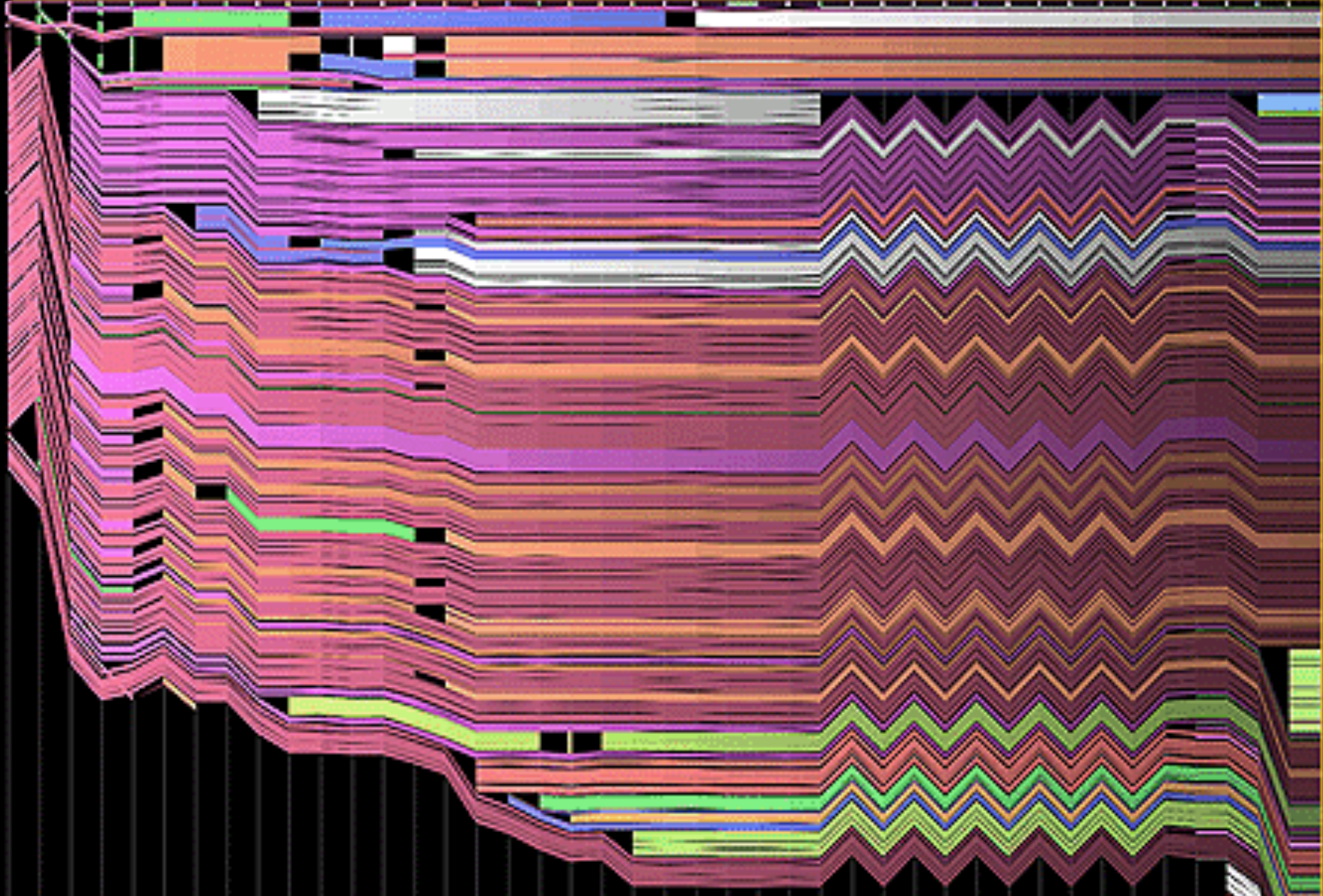
The morality and legality of abortion is a long important topic in **applied ethics**, and is also discussed by **legal scholars** and **religious philosophers**. Important facts about abortion are also recorded by **sociologists** and **historians**.

Abortion has been common in most societies, although it has often been opposed by some institutionalized religions and governments. **century politics in the United States and Europe** abortion became commonly accepted by 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 debate are generally peaceful, if heated, in their defense of their positions, the debate is sometimes characterized by violence. Though true of both sides, this is more marked on the side of those 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 a clash of presumed or perceived rights. On the one hand, is a fetus (sometimes called the "unborn" by **pro-life/anti-abortion** advocates) a human 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)



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

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

### Summary Statistics

$$u_X = 9.0 \quad \sigma_X = 3.317$$

$$u_Y = 7.5 \quad \sigma_Y = 2.03$$

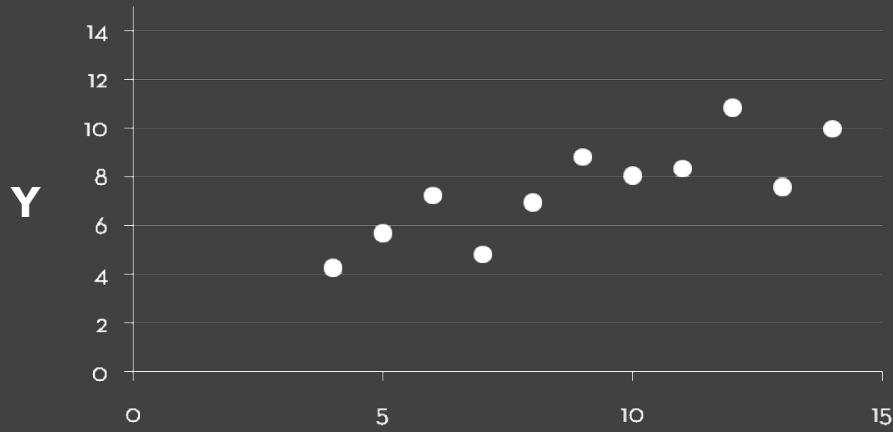
### Linear Regression

$$Y = 3 + 0.5 X$$

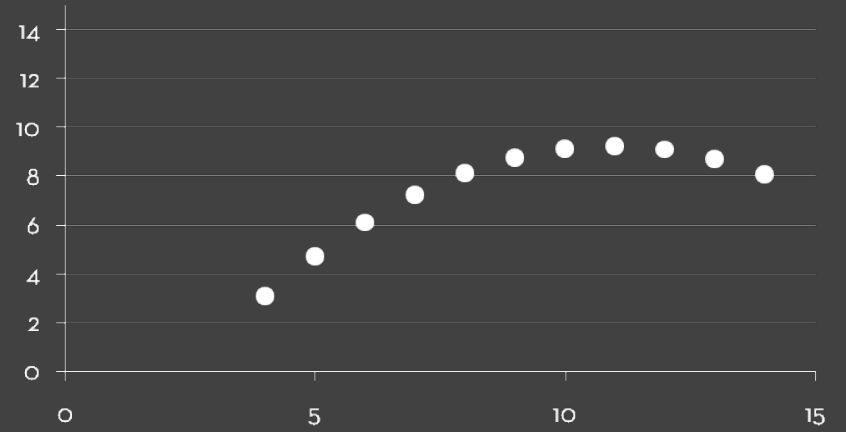
$$R^2 = 0.67$$

[Anscombe 1973]

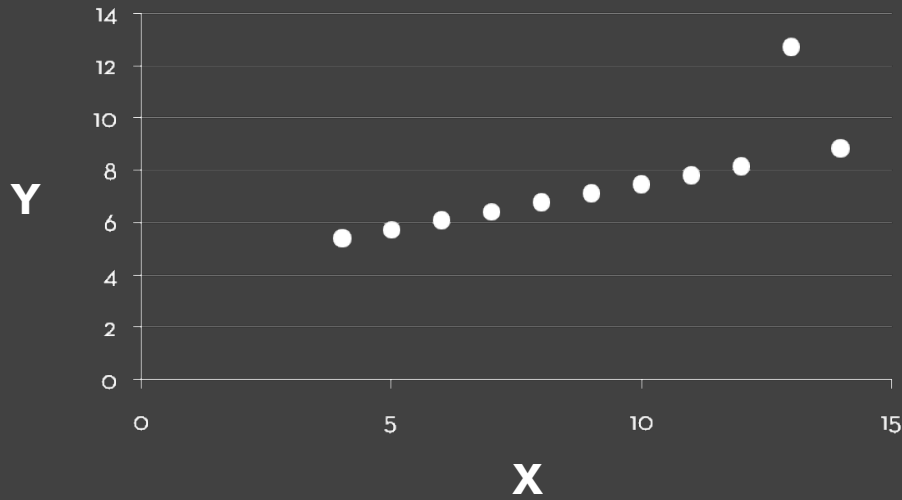
# Set A



# 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

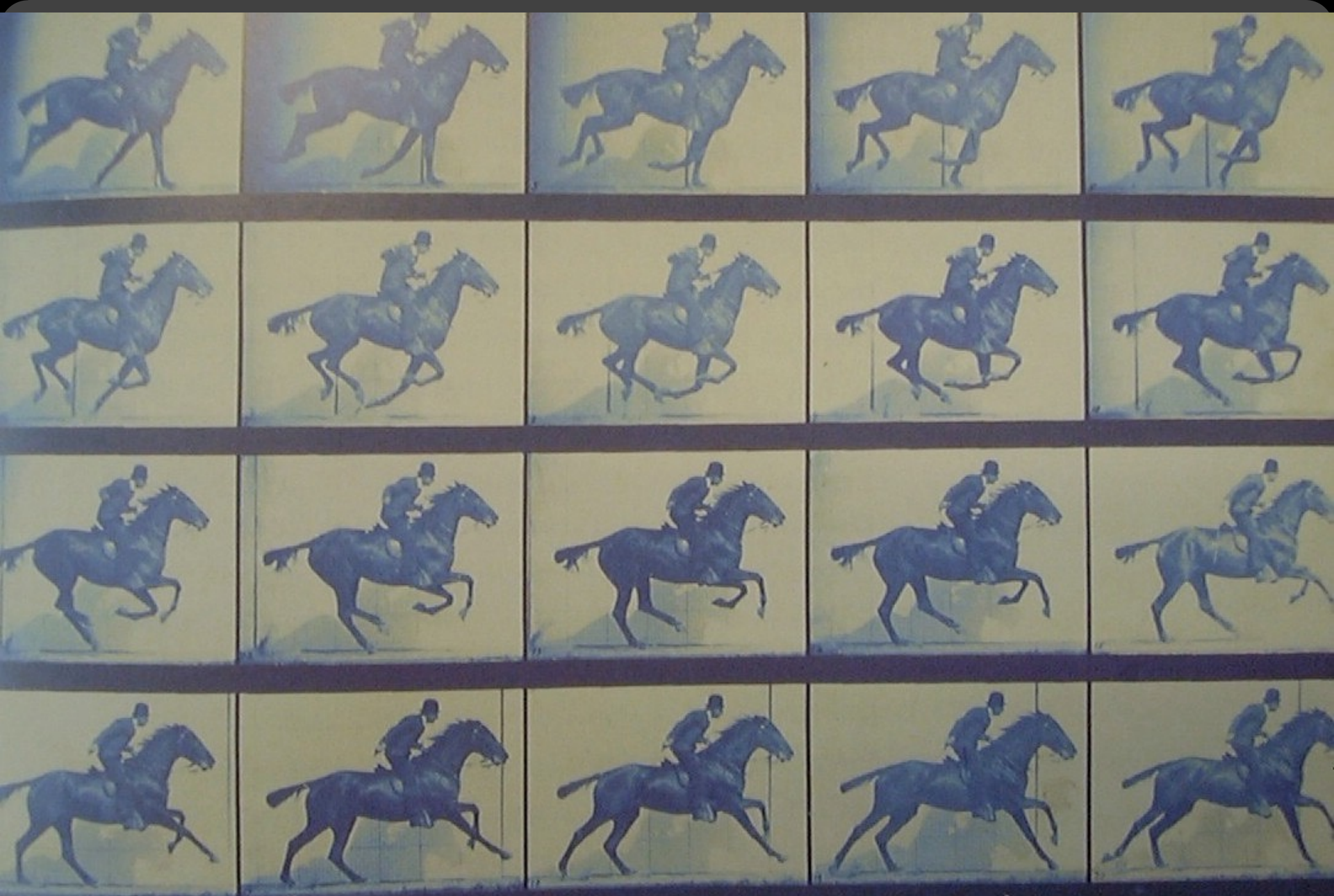
Find patterns

Present argument or tell a story

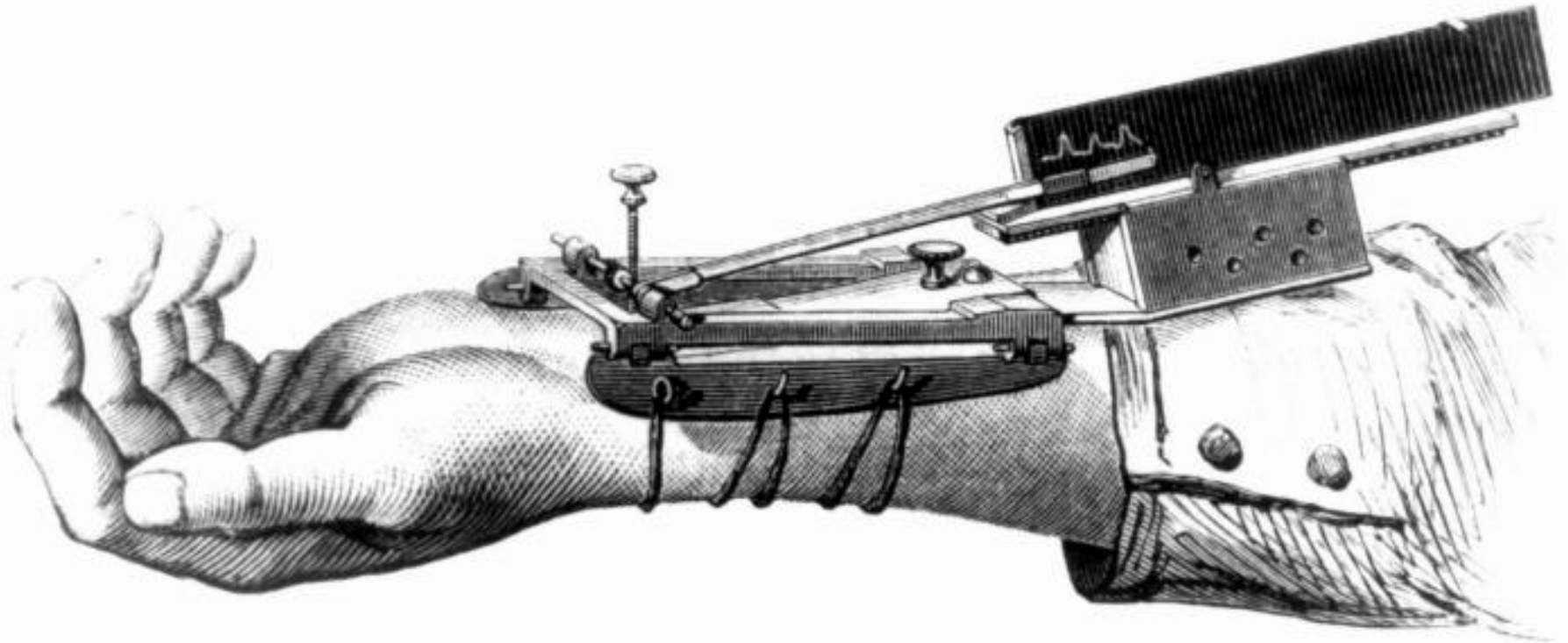
Inspire

# Record Information





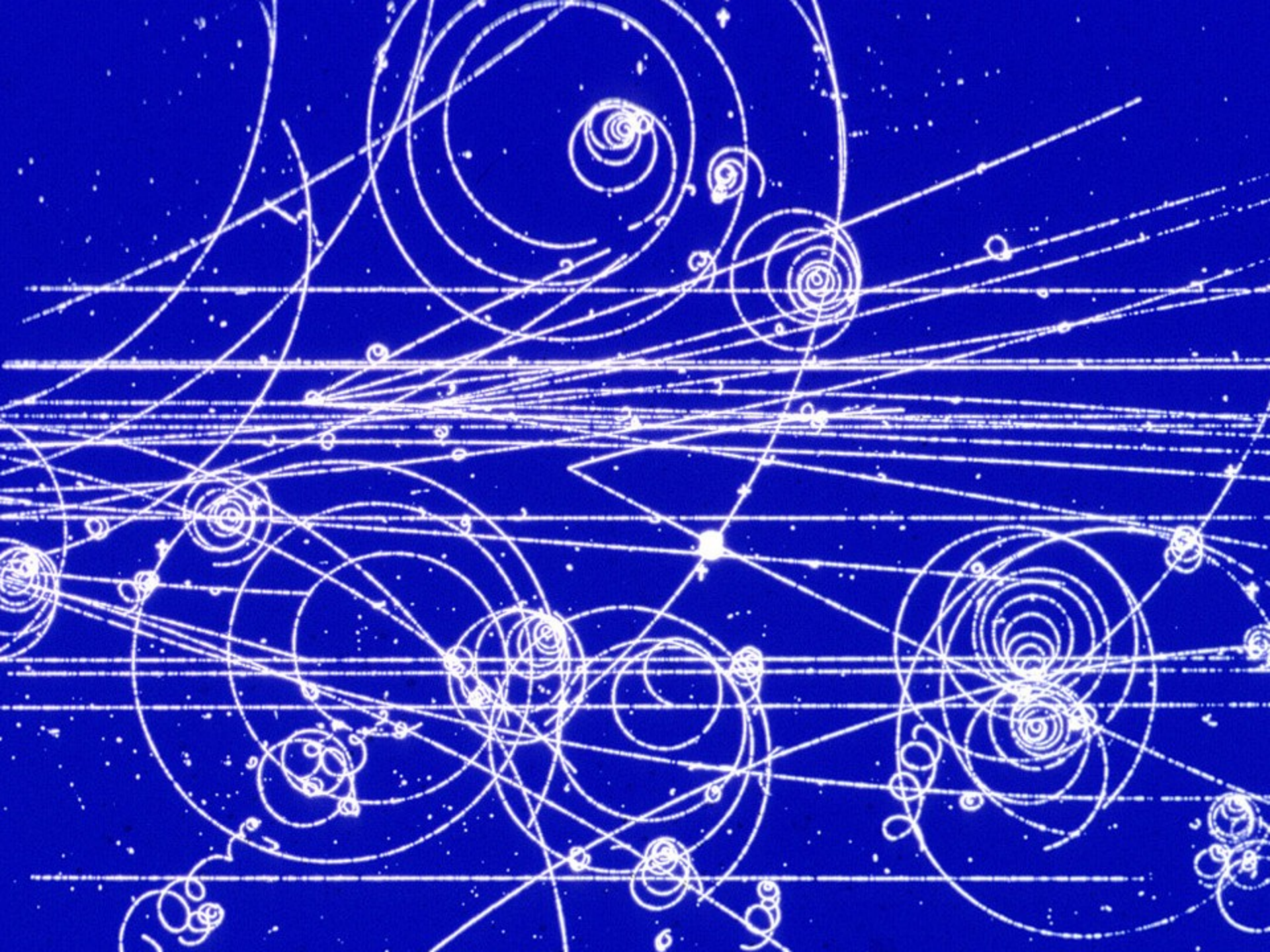
Gallop, Bay Horse "Daisy" [Muybridge 1884-86]



1.

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

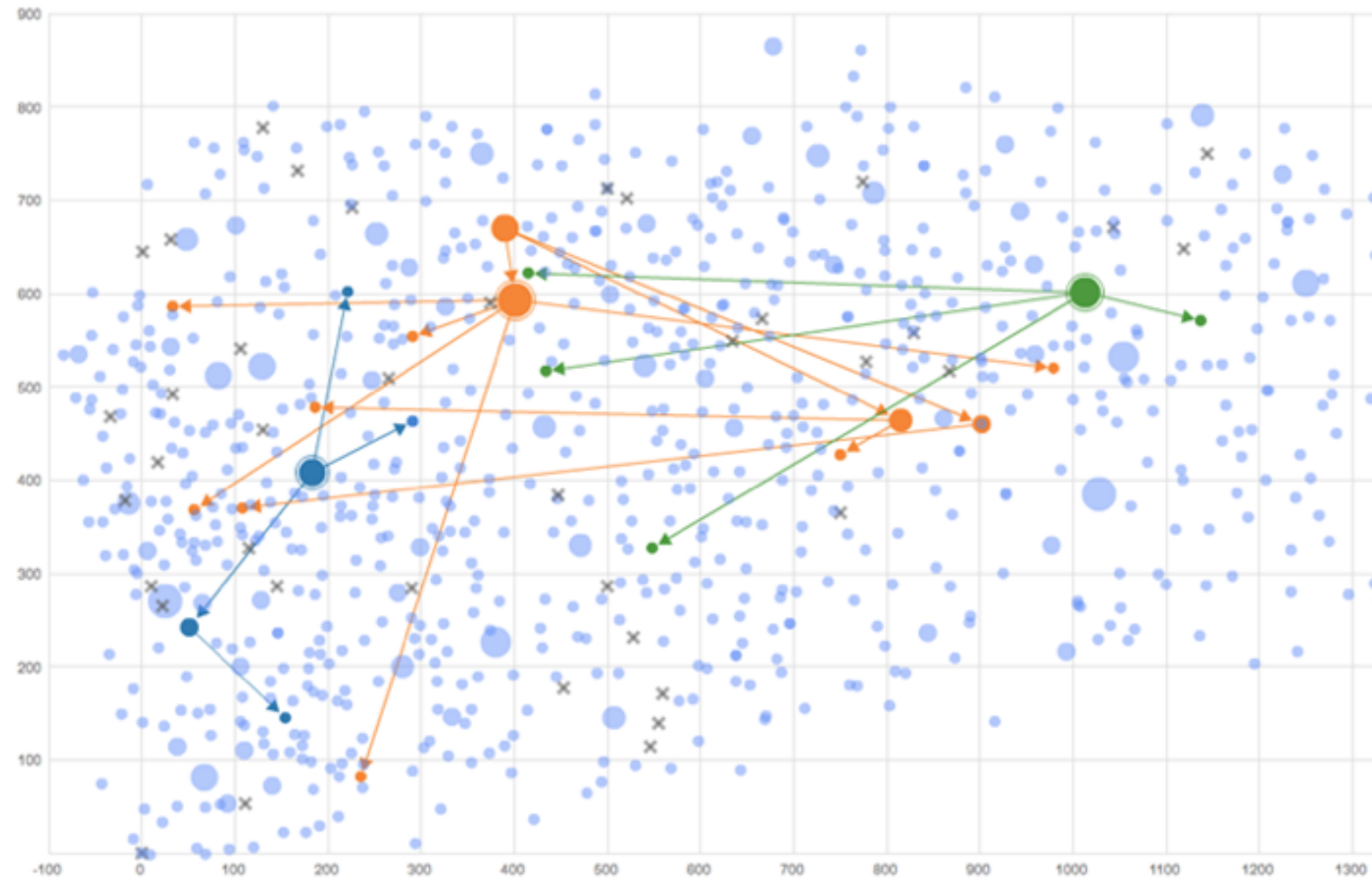
E.J. Marey's sphygmograph [from Braun 83]





Expected live in 06





# Support Reasoning

HISTORY OF O-RING DAMAGE ON SRM FIELD JOINTS

1161  
Oct 30, 1985  
y

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.)		
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  
 o 2 CASE JOINTS (80°), (110°) ARC  
 o MUCH WORSE VISUALLY THAN SRM-22

SRM 22 BLOW-BY  
 o 2 CASE JOINTS (30-40°)

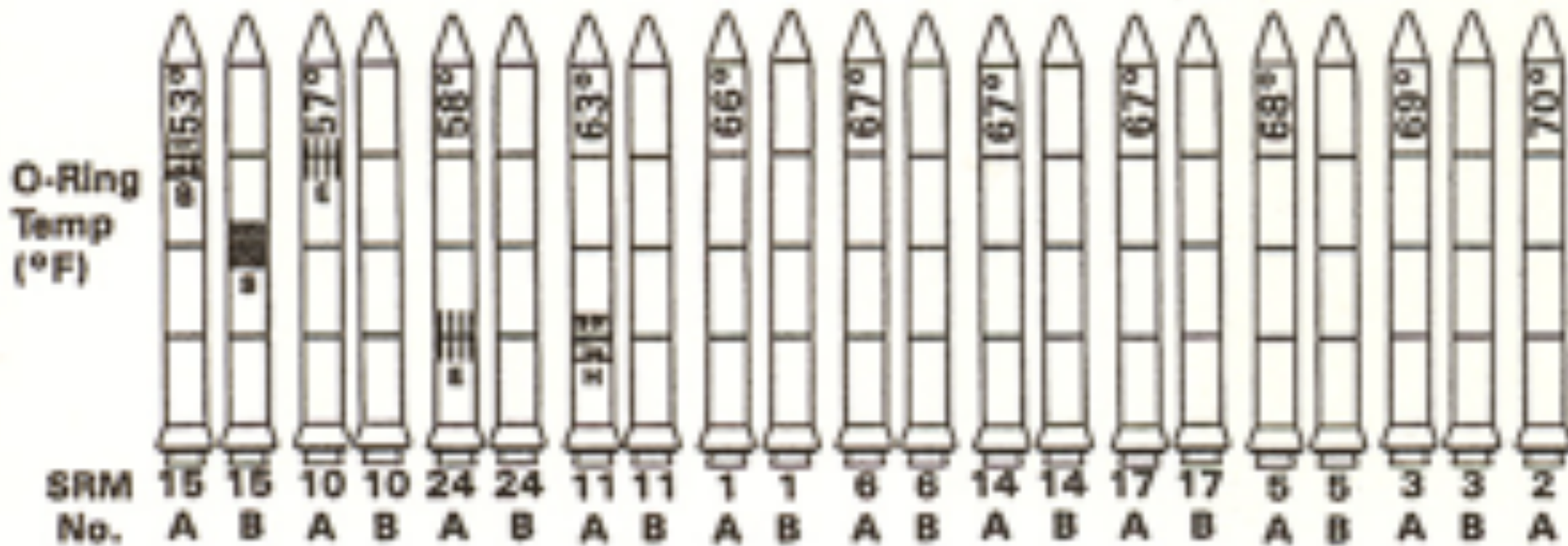
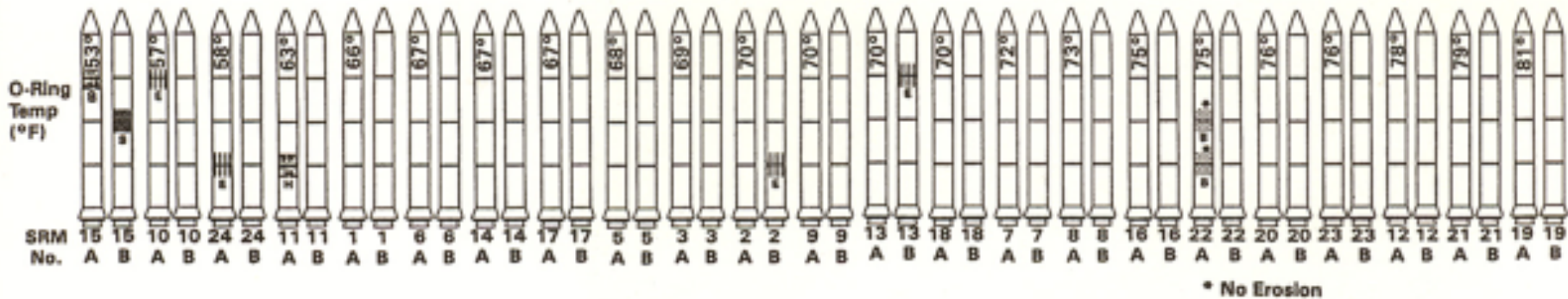
SRM-13A, 15, 16A, 18, 23A 24A  
 o NOZZLE BLOW-BY

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

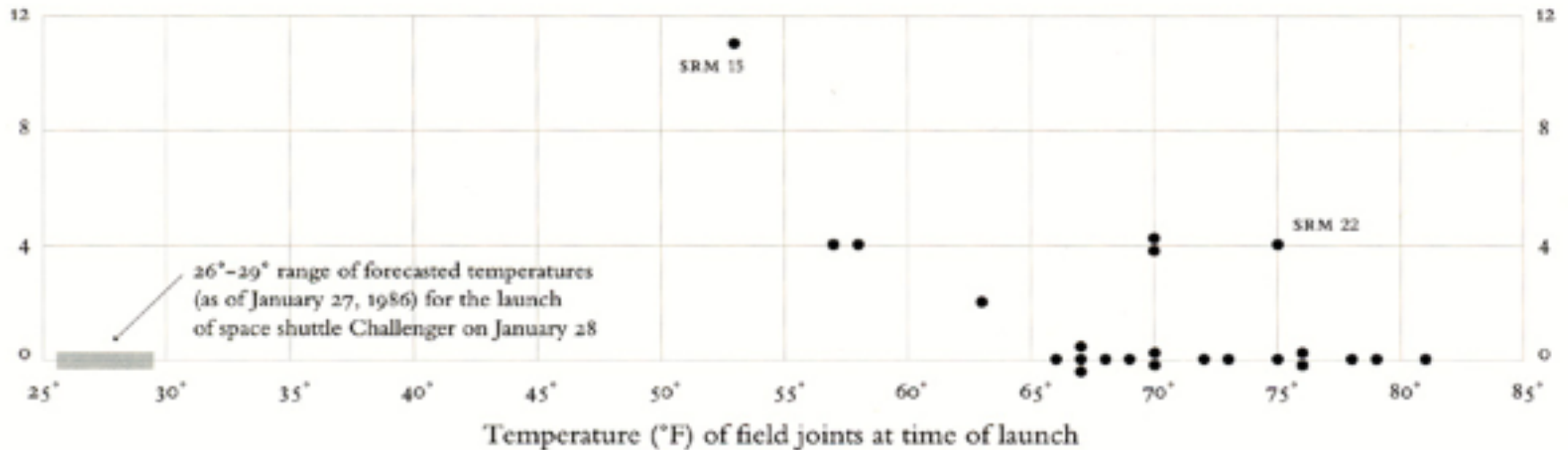


# Make a Decision: Challenger



# 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

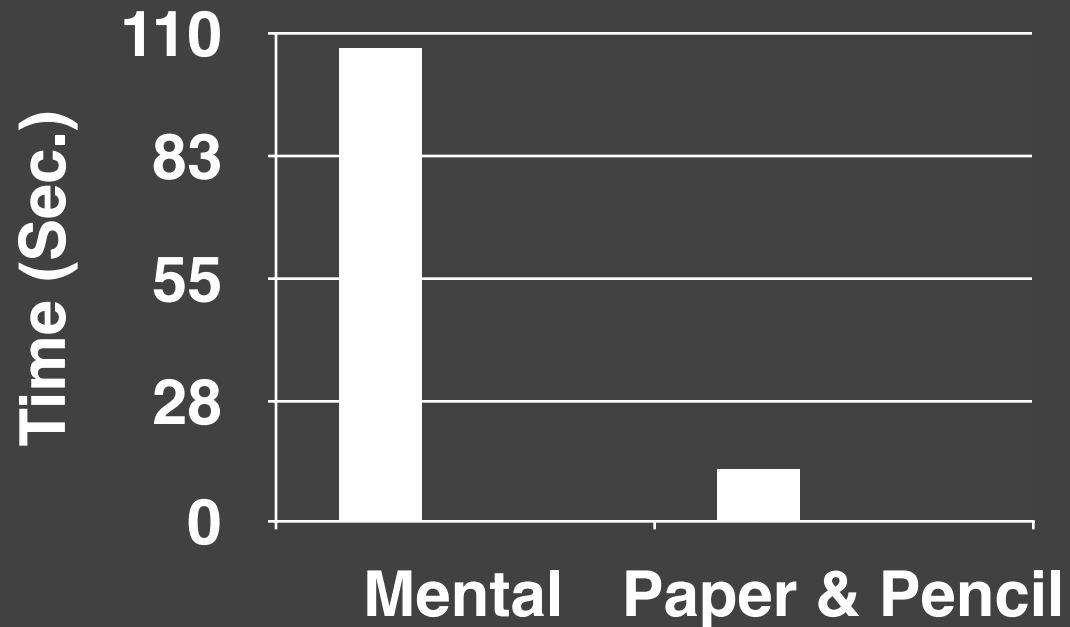
Class Exercise!

# Expand Memory: Multiplication

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

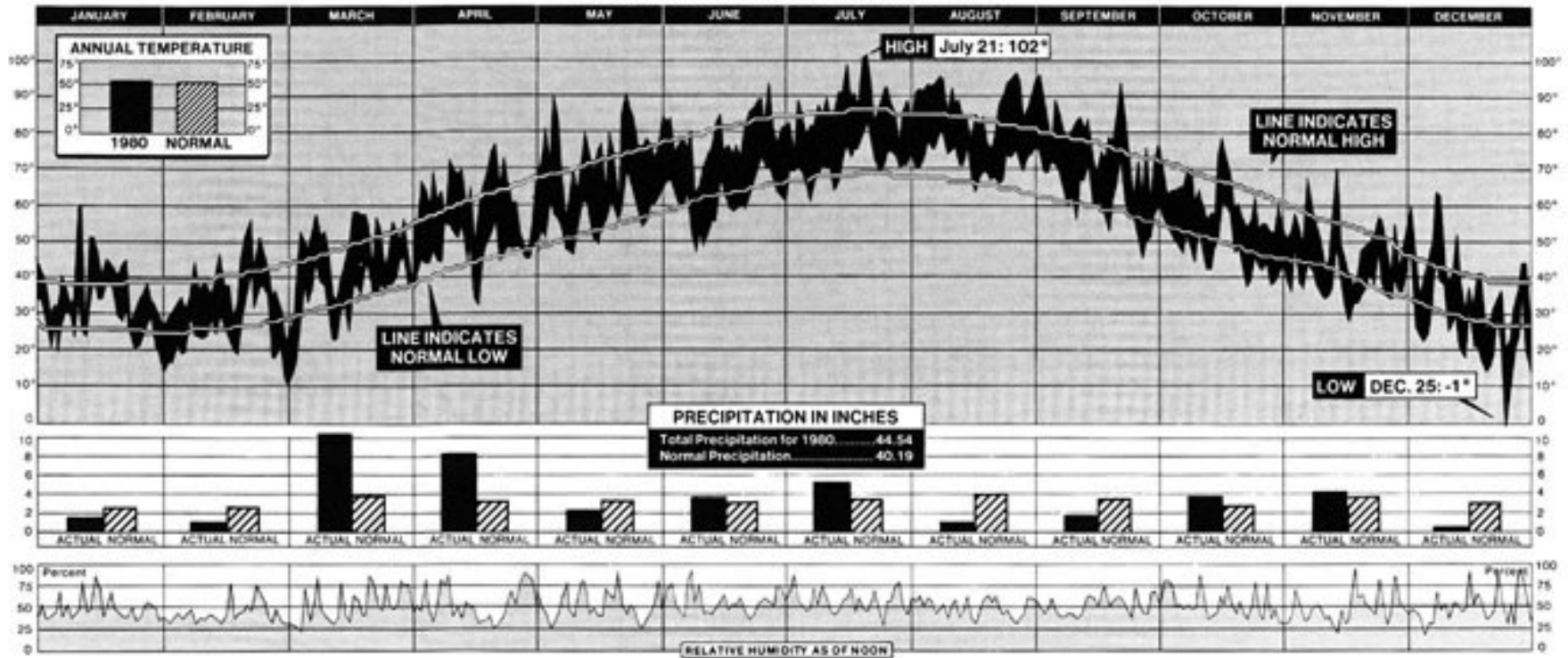
# Expand Memory: Multiplication

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



# Find Patterns: NYC Weather

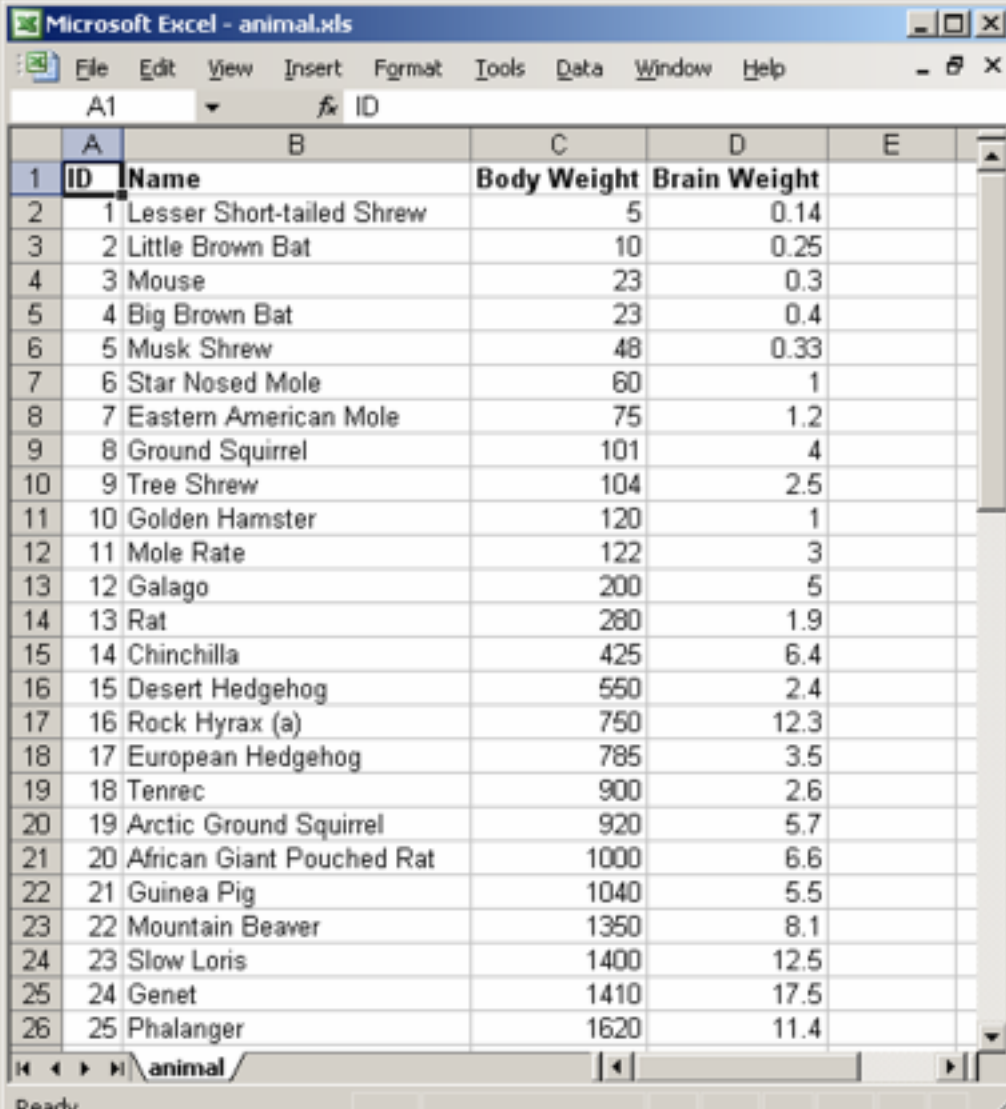
NEW YORK CITY'S WEATHER FOR 1980



[New York Times 1981]

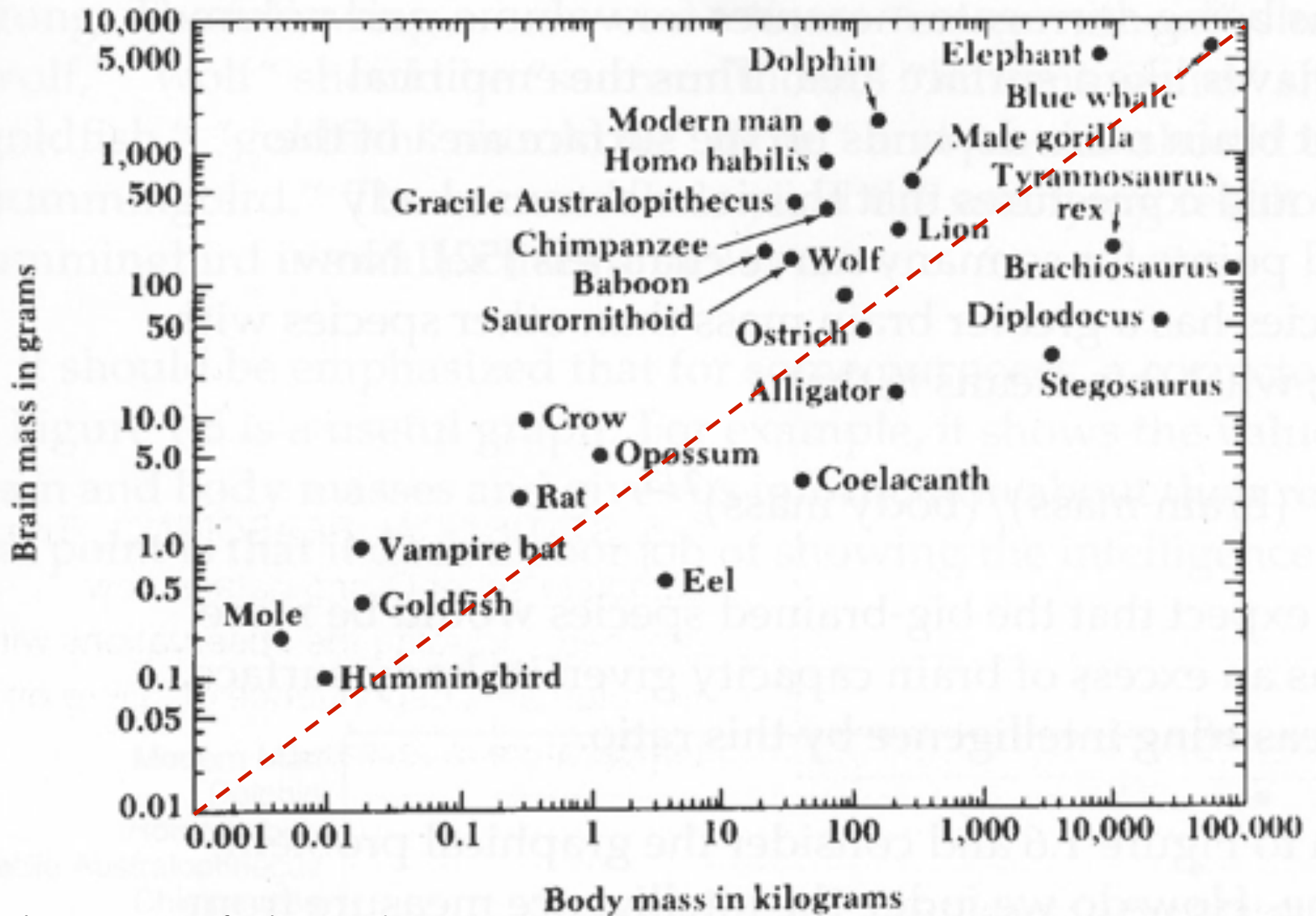


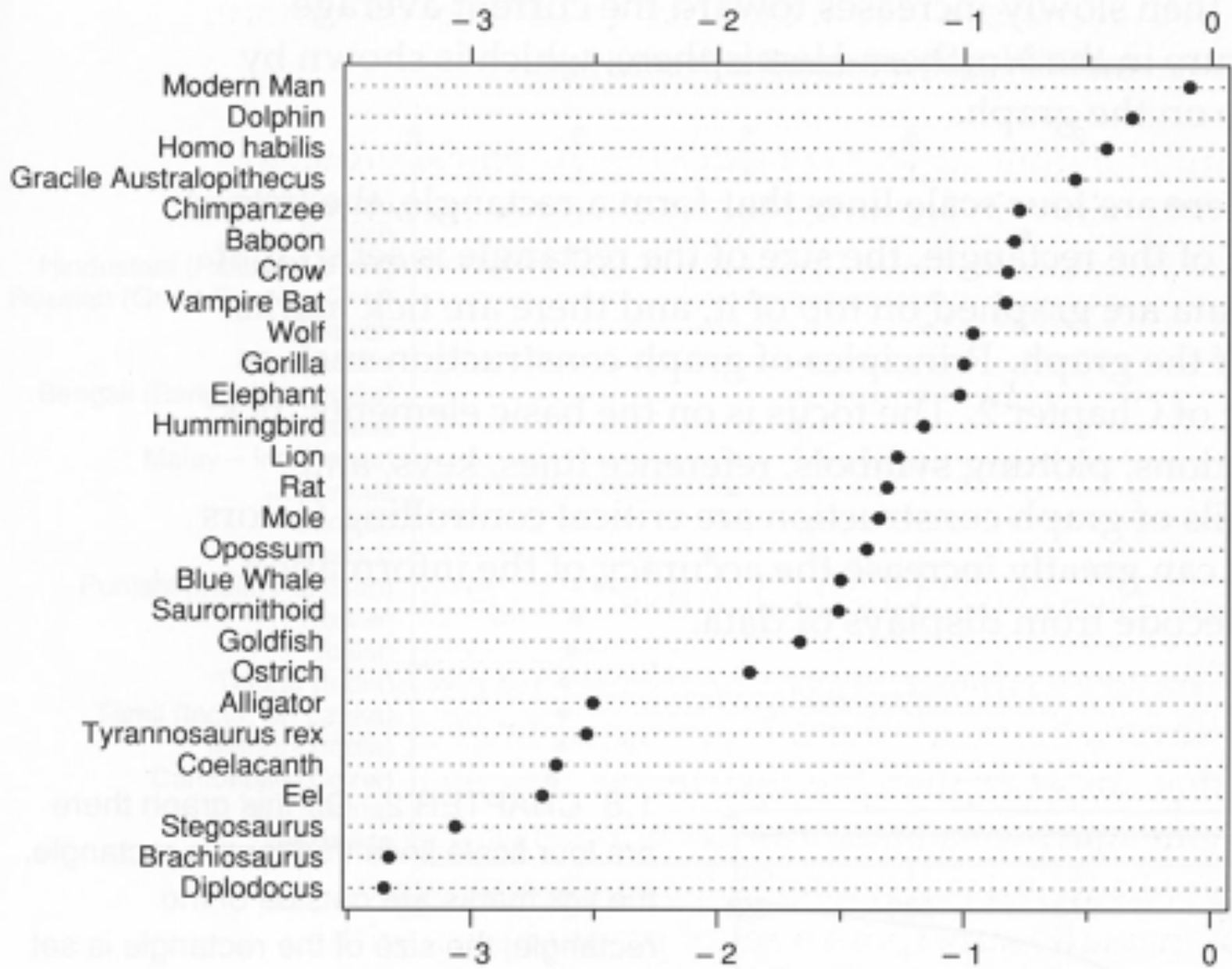
# The Most Powerful Brain?



The image shows a screenshot of a Microsoft Excel spreadsheet titled "animal.xls". The spreadsheet contains a table with four columns: ID, Name, Body Weight, and Brain Weight. The data is sorted by Body Weight in ascending order. The status bar at the bottom indicates the file is "Ready".

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





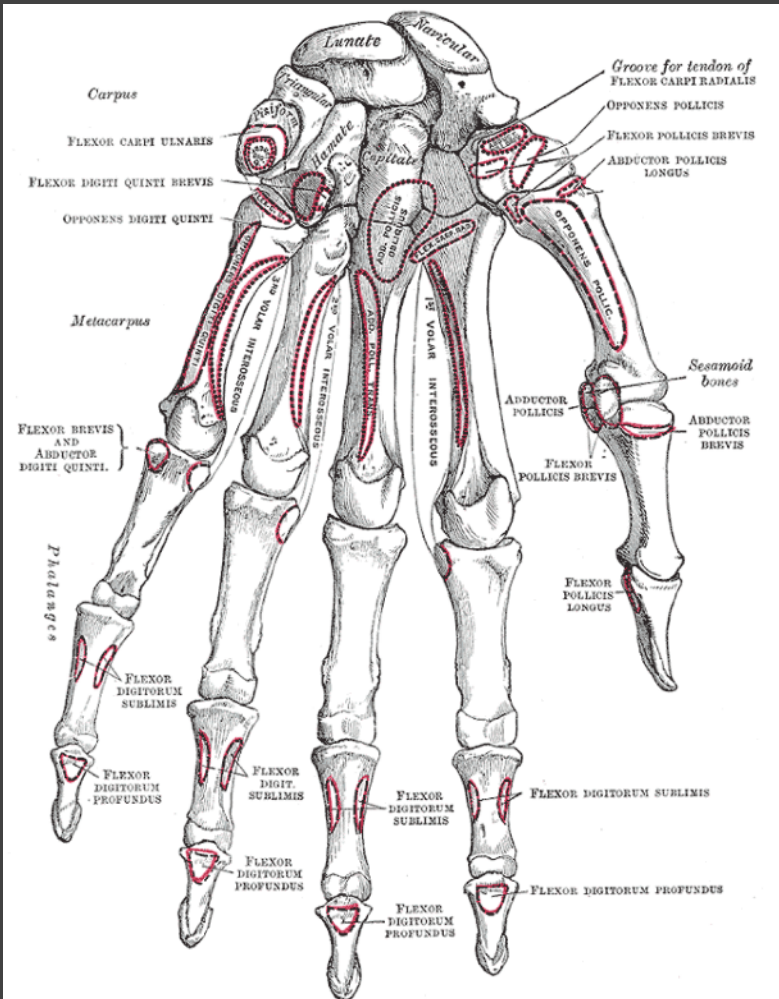
The Elements of Graphing Data

[Cleveland]

$\log_{10} \text{Brain Weight} - \frac{2}{3} \log_{10} \text{Body Weight}$

**Convey Information  
to Others**

# Inspire



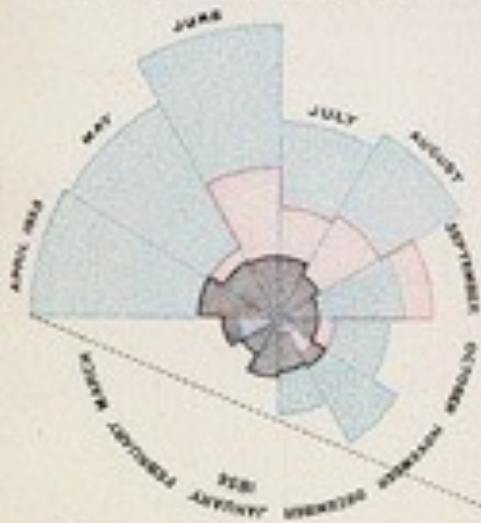
Bones in hand [from 1918 edition]



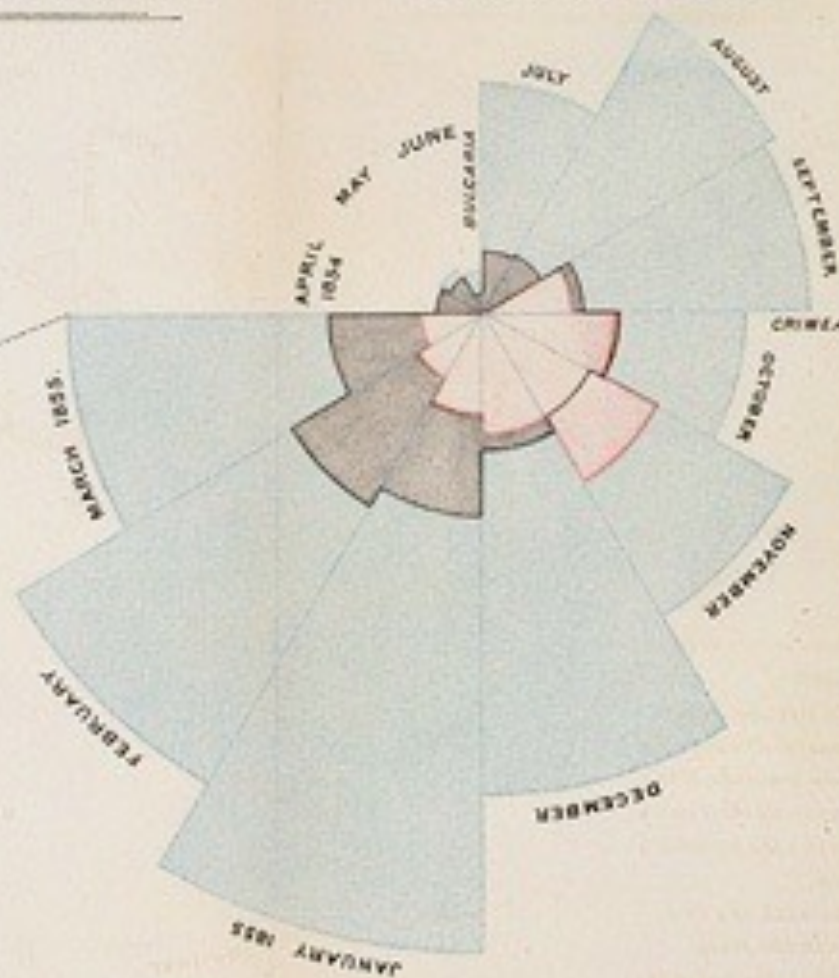
Double helix model [Watson and Crick 53]

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

2.  
APRIL 1855 TO MARCH 1856.



1.  
APRIL 1854 TO MARCH 1855.



“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



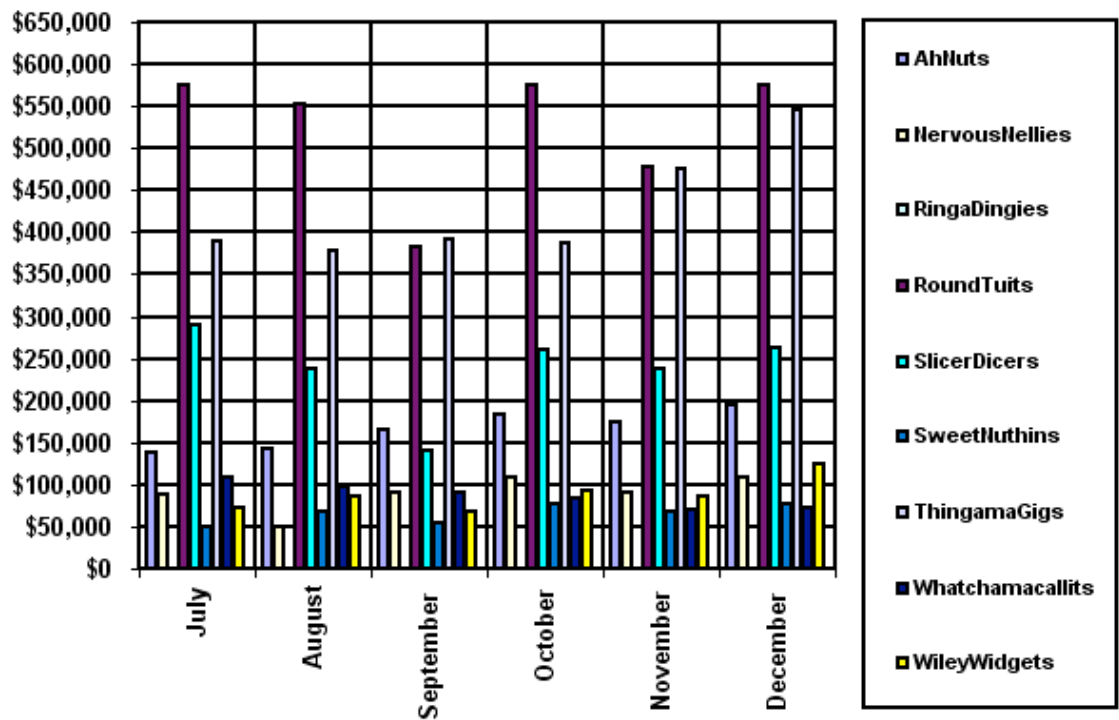
# Course Topics

# Data and Image Models

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

# Visualization Design

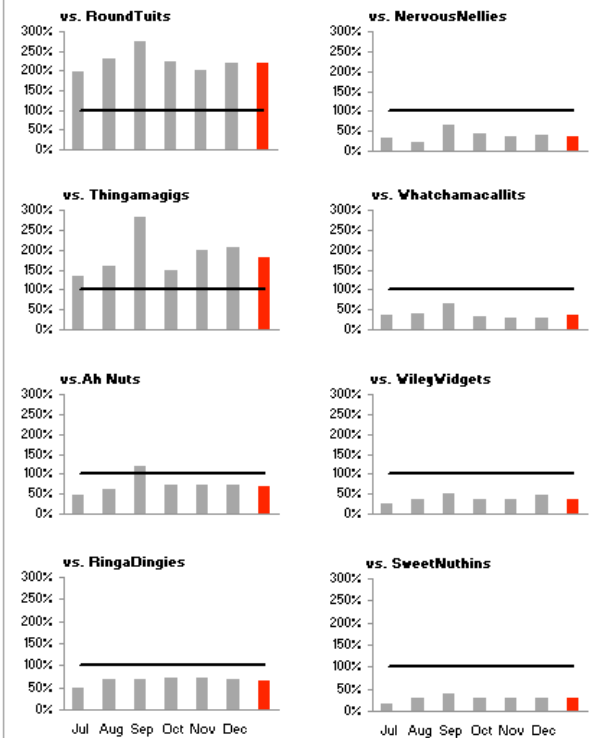
## SlicerDicers' Sales Compared to Other Products



Problematic design

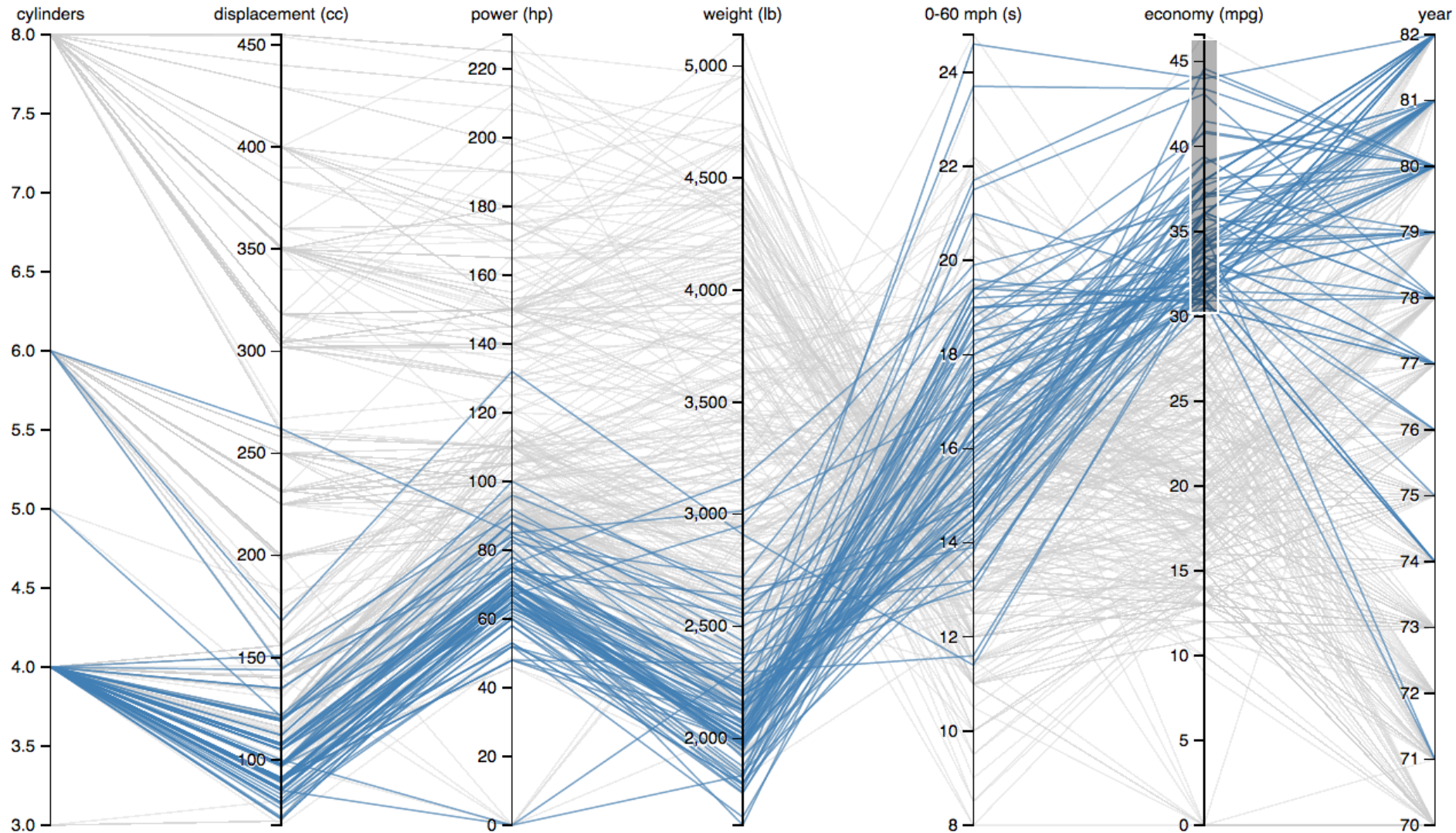
## Sales of SlicersDicers Compared to Other Products

July - December, 2003  
 (SlicersDicers' sales are displayed as black reference lines of 100%; the red bars represent the average monthly sales percentage for July through December.)

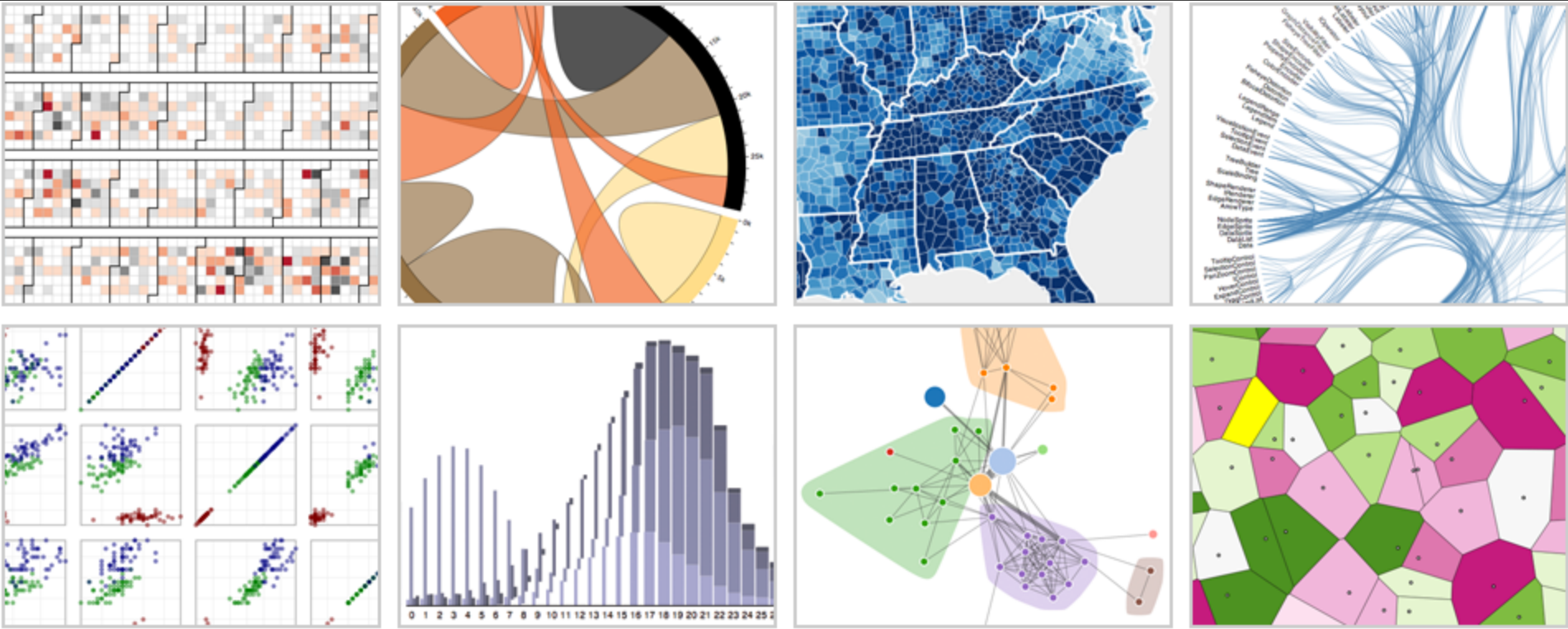


Redesign

# Exploratory Data Analysis

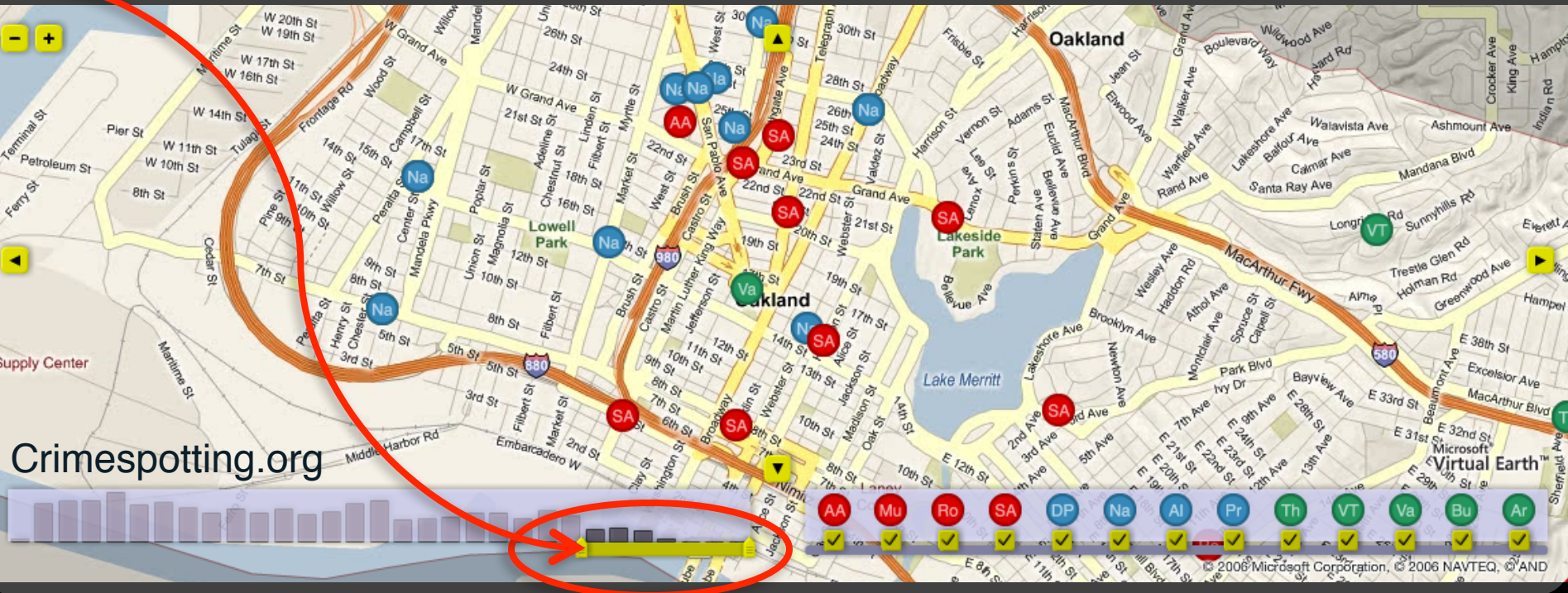
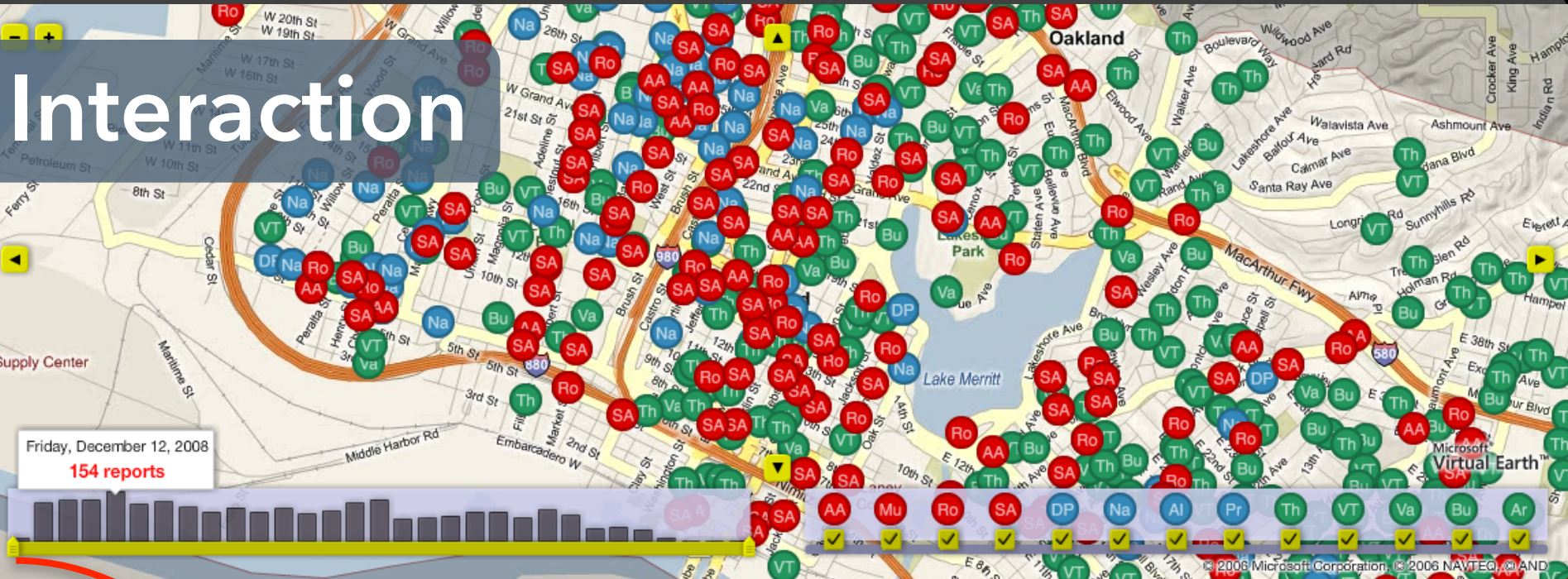


# Visualization Software

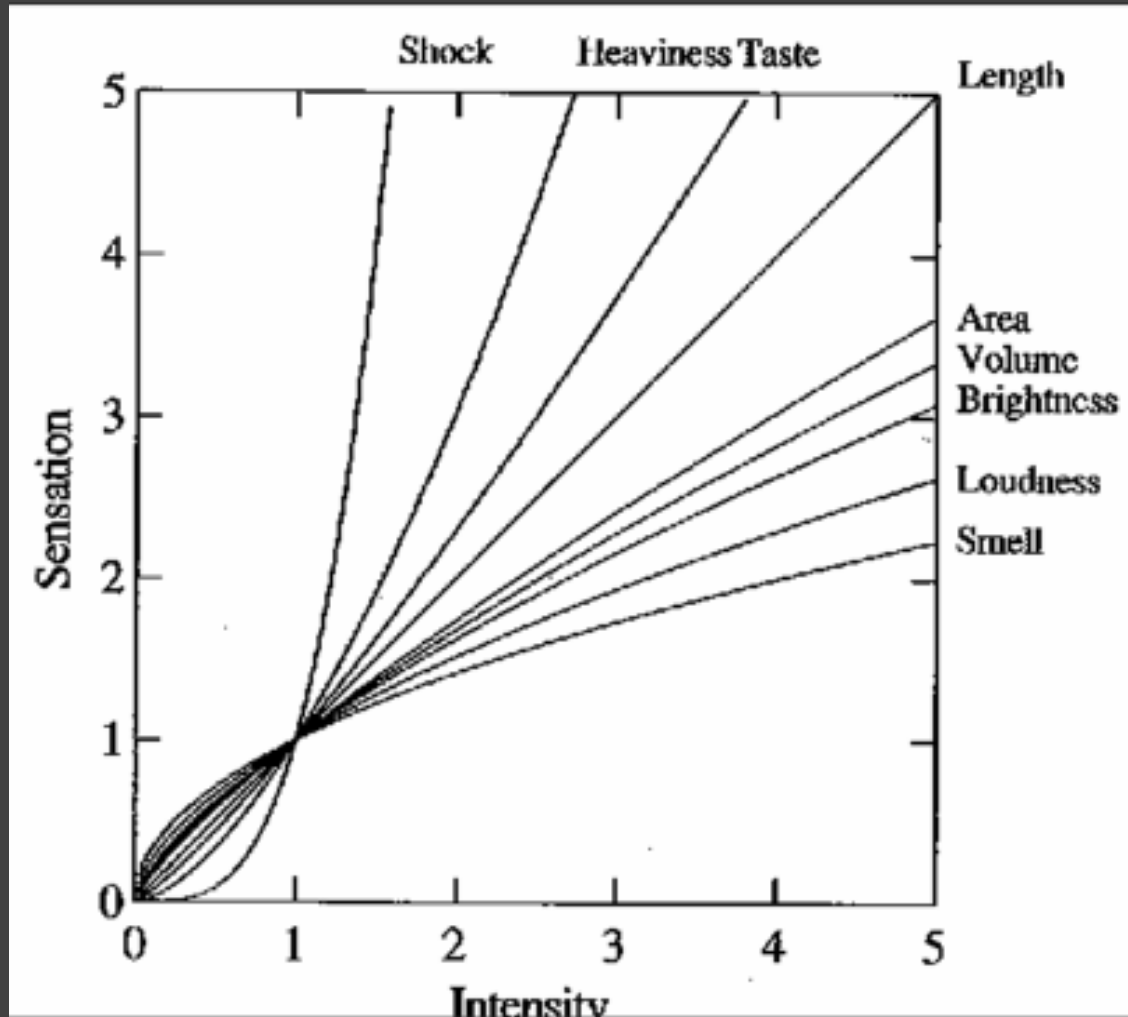


D3: Data-Driven Documents

# Interaction



# Graphical Perception

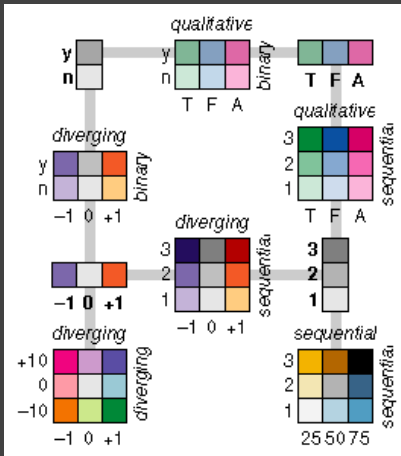
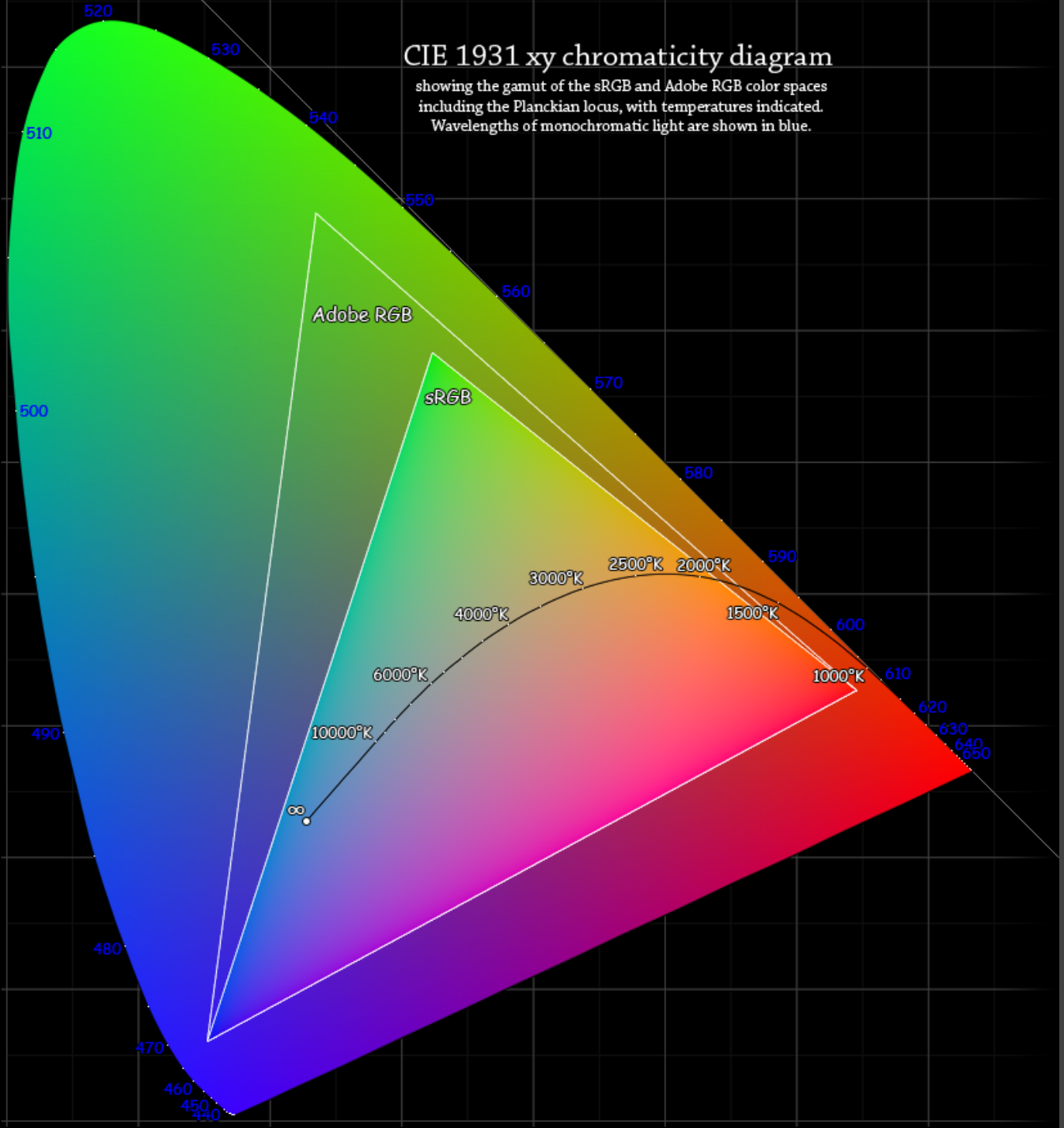


The psychophysics of sensory function [Stevens 61]

# Color

## CIE 1931 xy chromaticity diagram

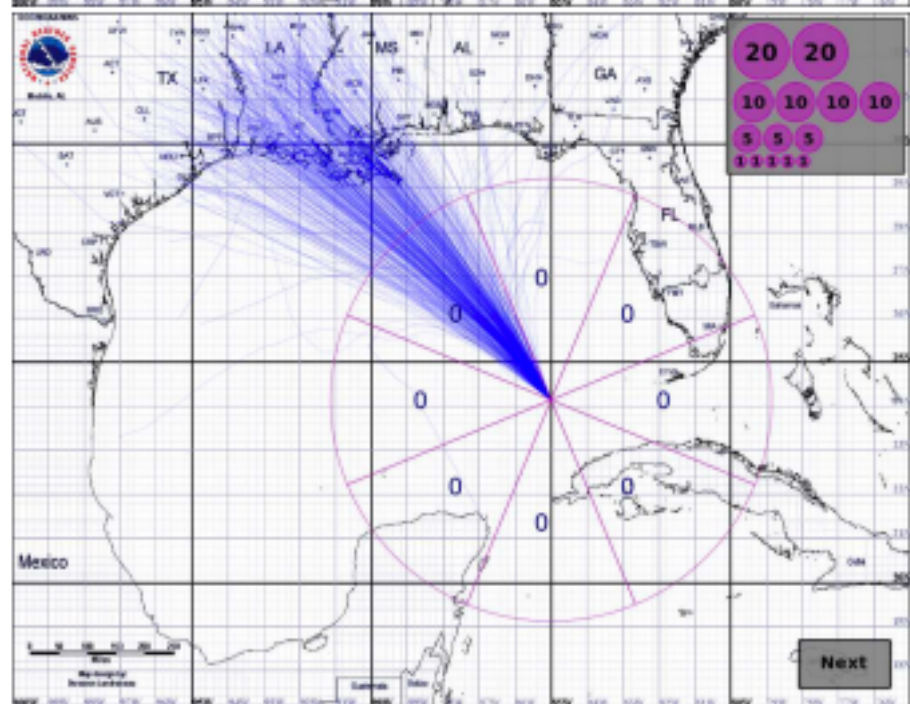
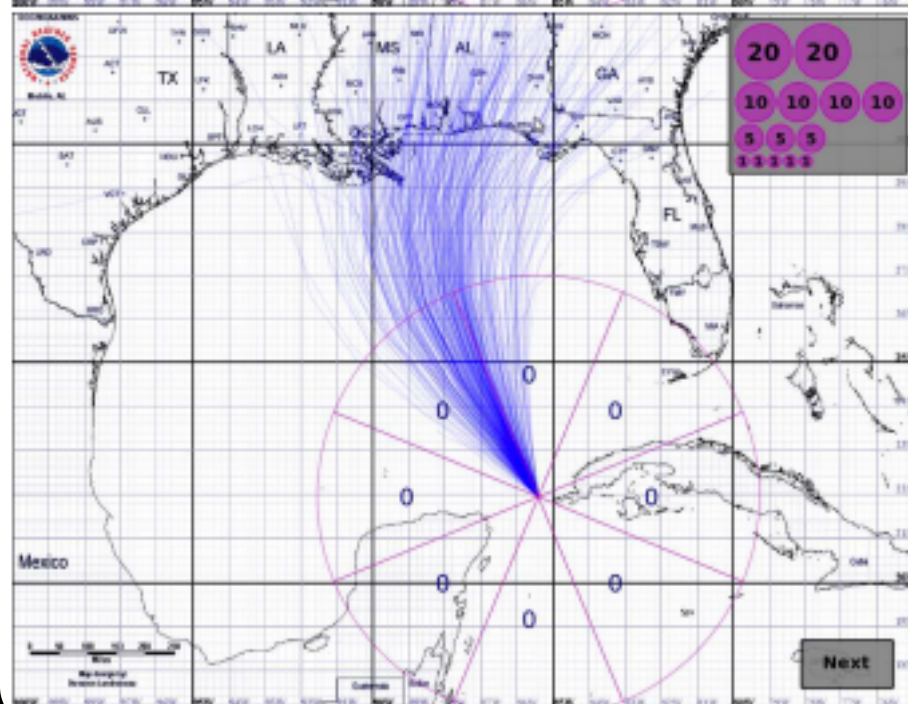
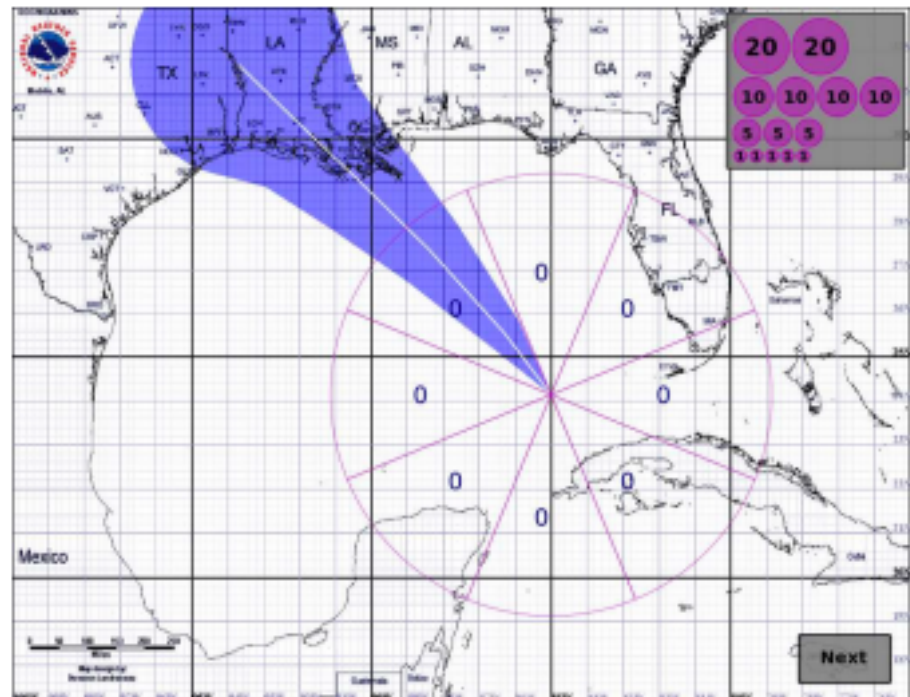
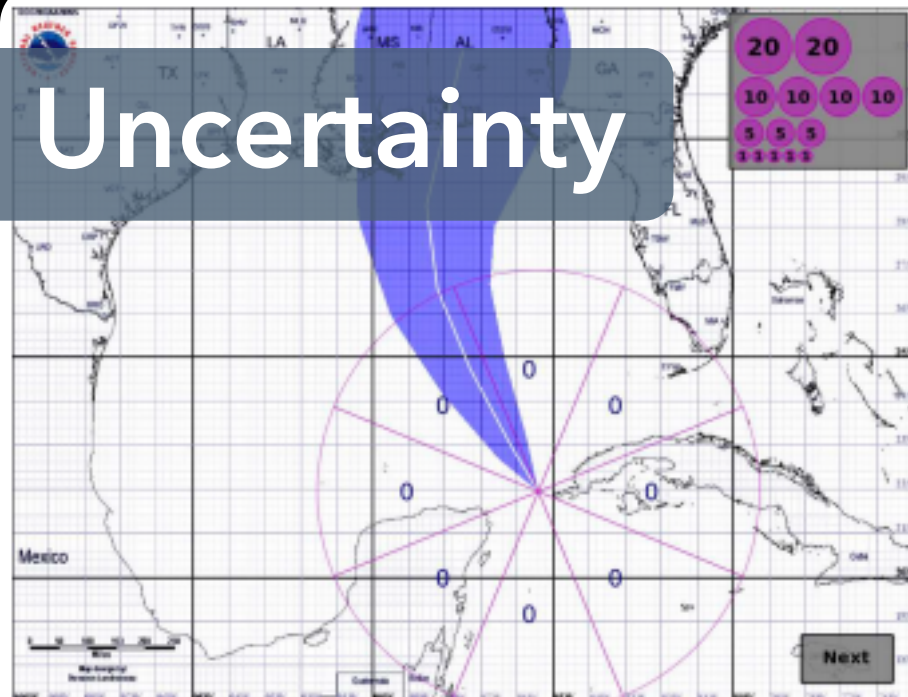
showing the gamut of the sRGB and Adobe RGB color spaces including the Planckian locus, with temperatures indicated. Wavelengths of monochromatic light are shown in blue.



## Color Brewer



# Uncertainty



Recent elections have placed a heavy emphasis on “swing states” — Ohio, Florida and the other competitive states. In the past, many more states shifted between the Democratic and Republican parties. A look at how the states stacked up in the election and how they have shifted over past elections.

# Narrative

- 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

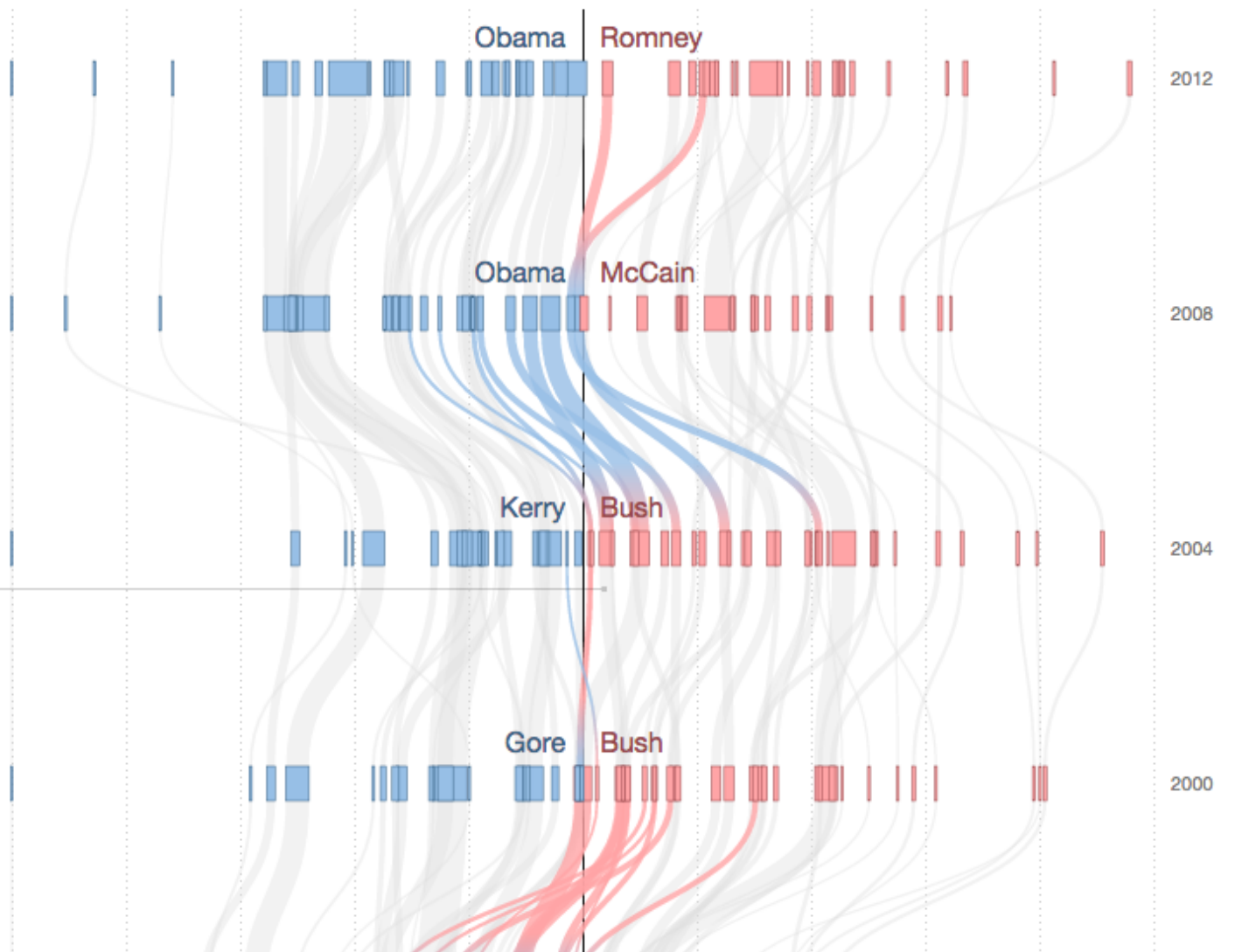
← MORE DEMOCRATIC | MORE REPUBLICAN →  
 ≥50% +40% +30% +20% +10% +10% +20% +30% +40% ≥50%

**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.

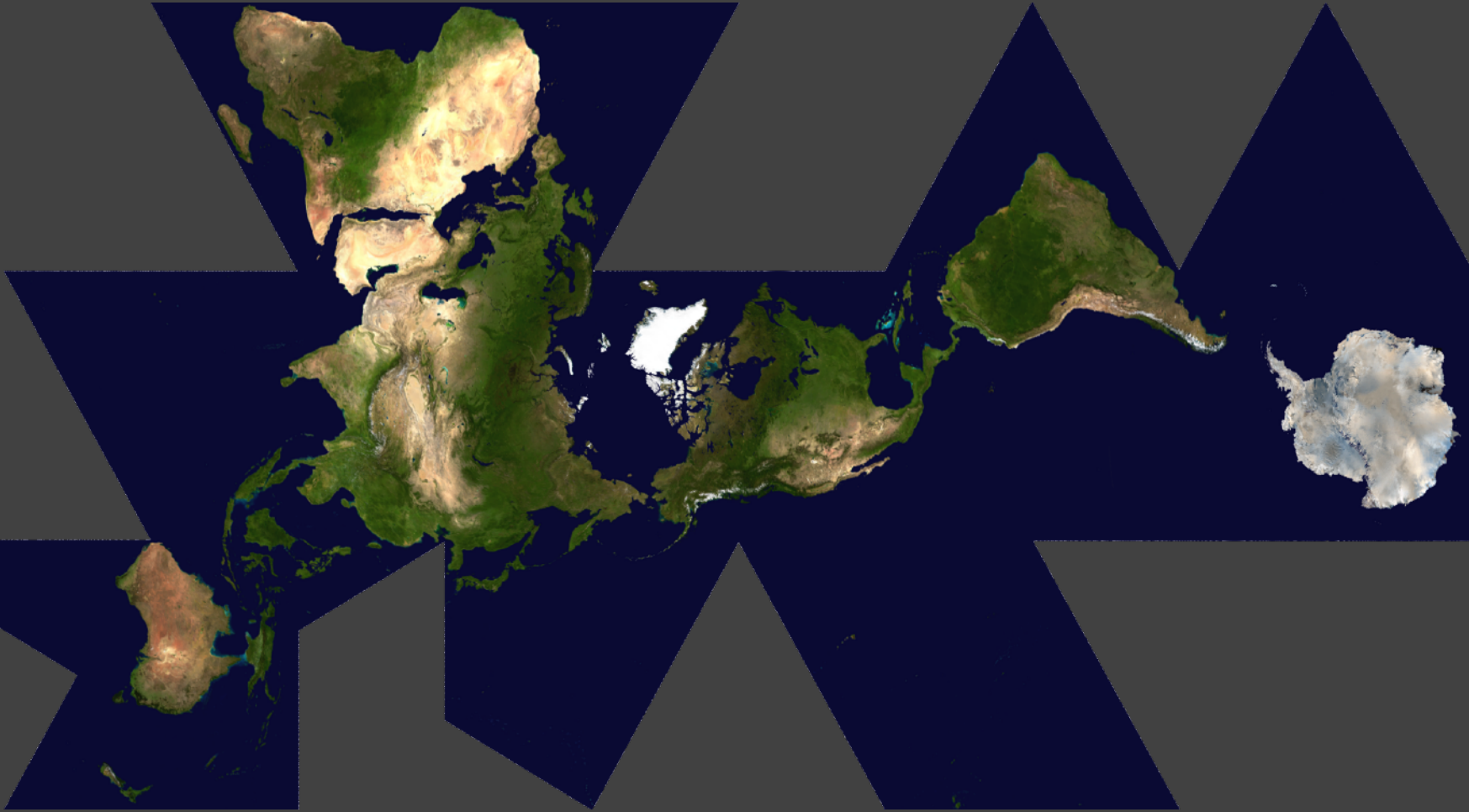
Highlight Tossups

**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

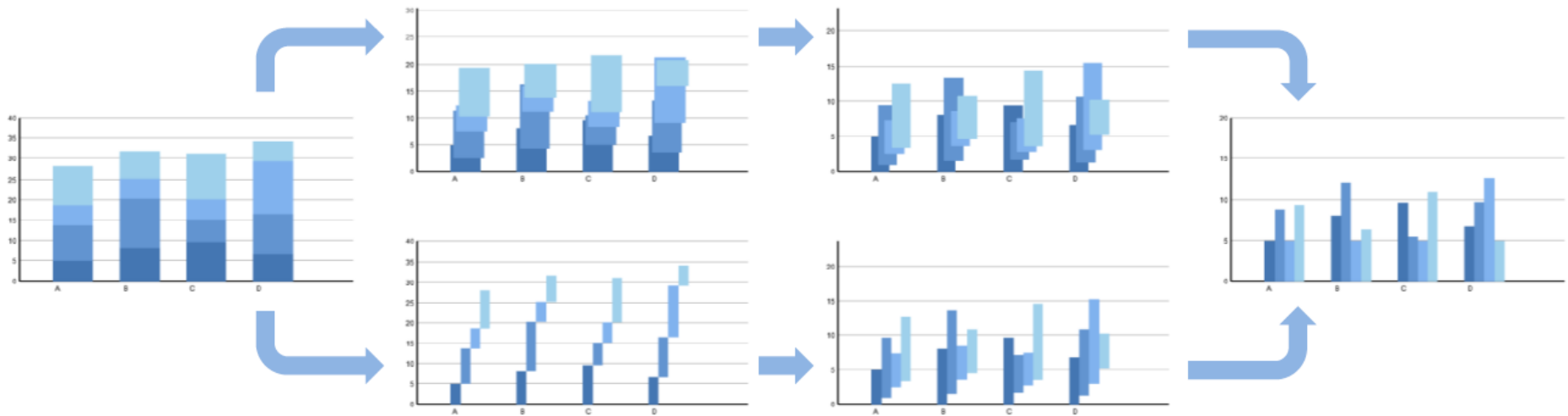


# Maps



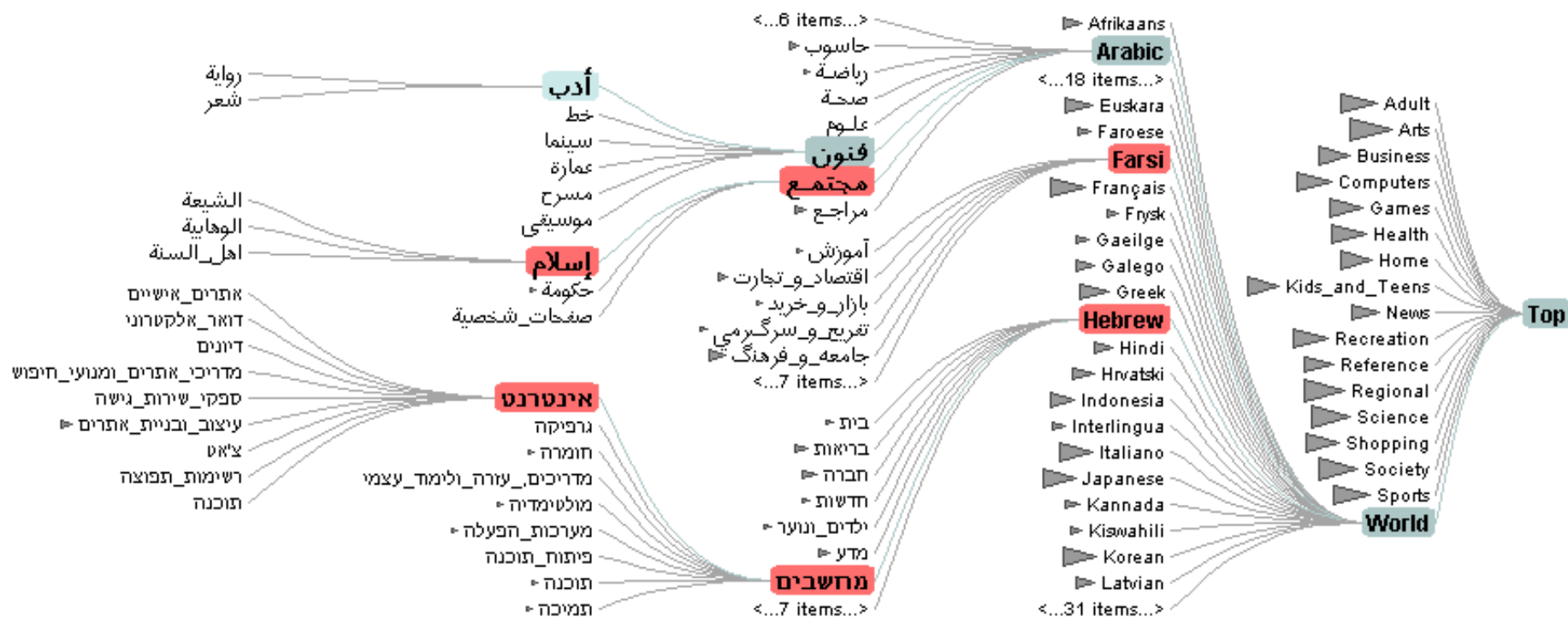
Dymaxion Maps [Fuller 46]

# Animation



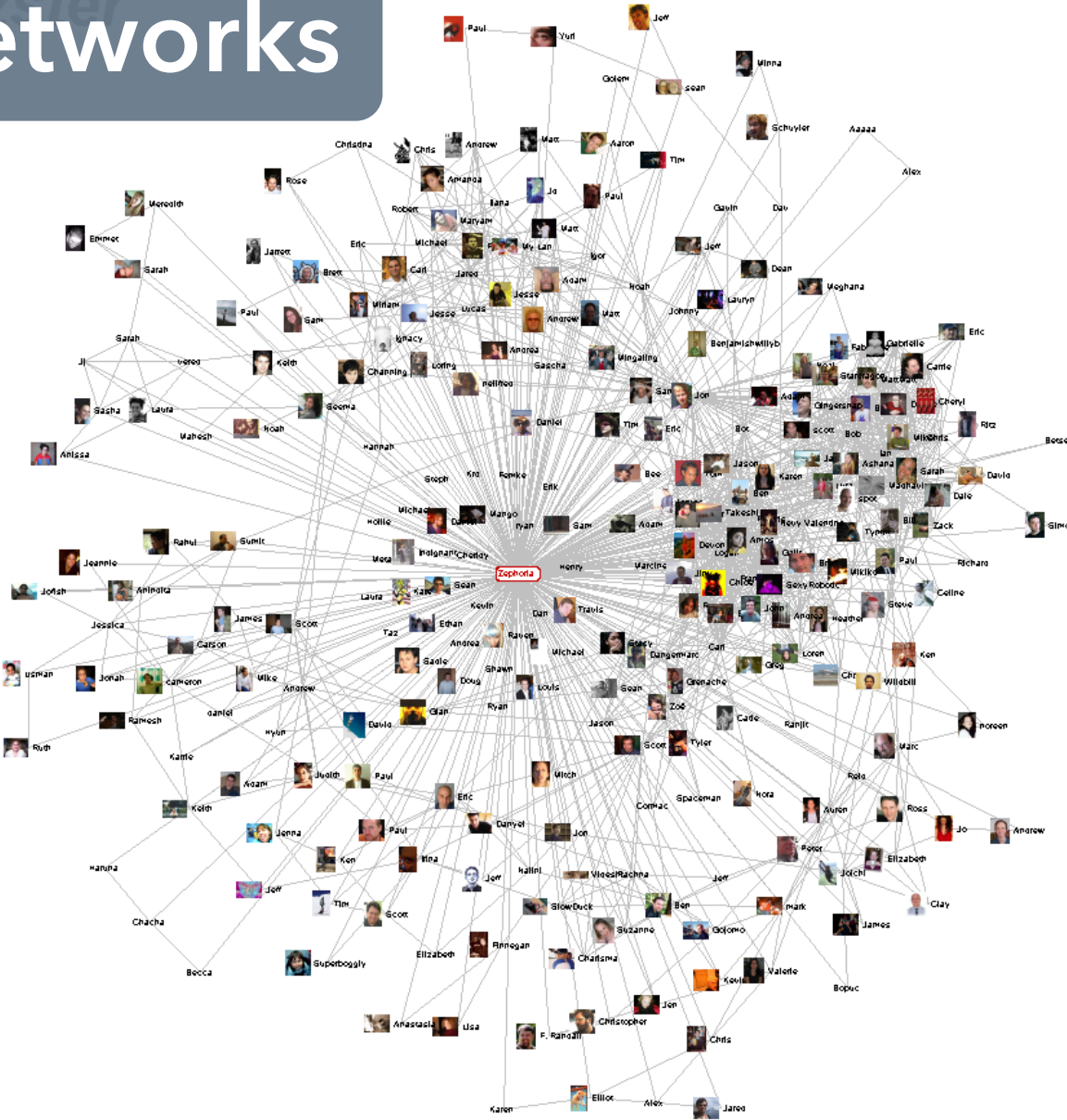
Animated transitions in statistical data graphics [Heer & Robertson 07]

# Hierarchies



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

# Networks



community >>

Enable

search >>

## Zephoria

<b>User ID</b>	21721
<b>Friends</b>	<input type="checkbox"/> 266
<b>Age</b>	??
<b>Gender</b>	<input type="checkbox"/> Female
<b>Status</b>	<input type="checkbox"/> Single
<b>Location</b>	San Francisco, CA
<b>Hometown</b>	Lancaster, PA
<b>Occupation</b>	researcher: social networks, identity, context
<b>Interests</b>	apophenia, observing people, culture, questioning power, reading, buddhism, ipseity, computer-mediated communication, social networks, technology, anthropology, stumping
<b>Music</b>	psytrance/goa/trance [Infected Mushroom, Son Kite... Iboga/Digital Structures], Ani Difranco, downtempo, Thievery Corporation, Beth Orton, Morcheeba, Ween, White Stripes
<b>Books</b>	Authors: Erving Goffman, Stanley Milgram, Jeanette Winterson, Eric Schlosser, Leslie Feinberg, Dorothy Allison, Italo Calvino, Hermann Hesse
<b>TV Shows</b>	??
<b>Movies</b>	Koyaanisqatsi, Amelie, Waking Life, Tank Girl, The Matrix, Clockwork Orange, American Beauty, Fight Club, Boys Don't Cry
<b>Member Since</b>	??
<b>Last Login</b>	2003-10-21
<b>Last Updated</b>	2003-10-21
<b>About</b>	[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.

# 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**

Prof, CSE

OH: *Tue 9:00-10:00am, 642 CSE*

<http://jheer.org>

## *Assistants*

**Halden Lin**

OH: *Mon 1:30-2:30pm, 4th Floor*

**Younghoon Kim**

OH: *Tue 3:00-4:00pm, 4th Floor*

**Zening Qu**

OH: *Thu 3:30-4:30pm, 4th Floor*

**Sherry Wu**

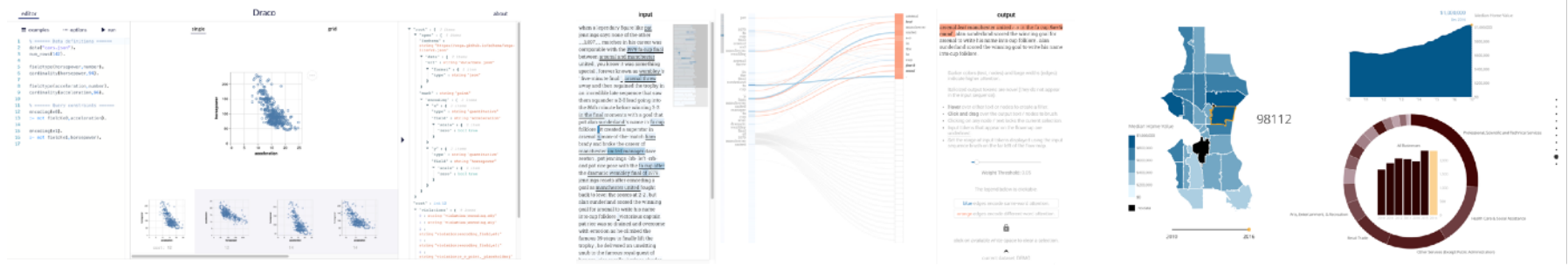
OH: *Fri 1:30-2:30pm, 5th Floor*

# Halden Lin

[haldenl@cs.washington.edu](mailto:haldenl@cs.washington.edu)

Hi!

I'm a master's student working on **visualization recommendation systems** and **visualization for NLP**.



# Younghoon Kim

[yhkim01@cs.washington.edu](mailto:yhkim01@cs.washington.edu)

## **Office Hour**

Tuesday 3:00 - 4:00 p.m.  
4th Floor Breakout

Hi!

I'm a 4th year Ph.D. student and interested in algorithms for visualization recommendation and data-driven storytelling!



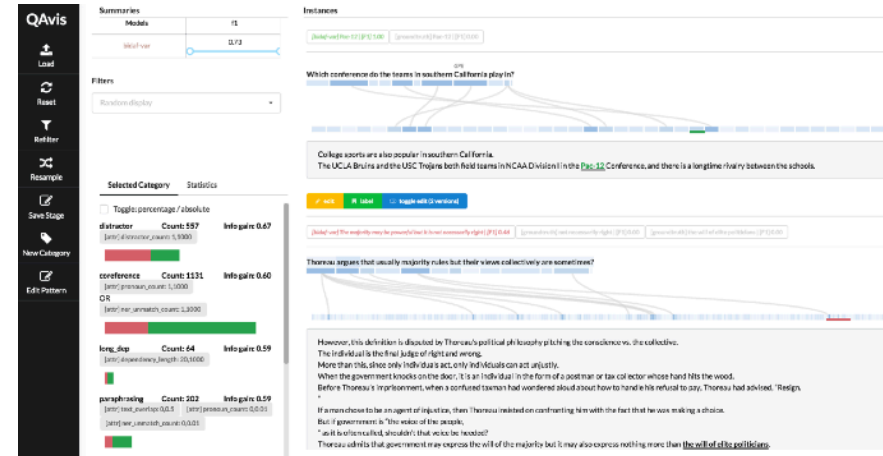
# Tongshuang Wu (Sherry)

OH: Fri 1:30-2:30pm

CSE 5th Floor Breakout

I'm a third year Ph.D. student in IDL working on interactive machine learning. Most recently, I'm building visual tools for error analysis of machine learning models.

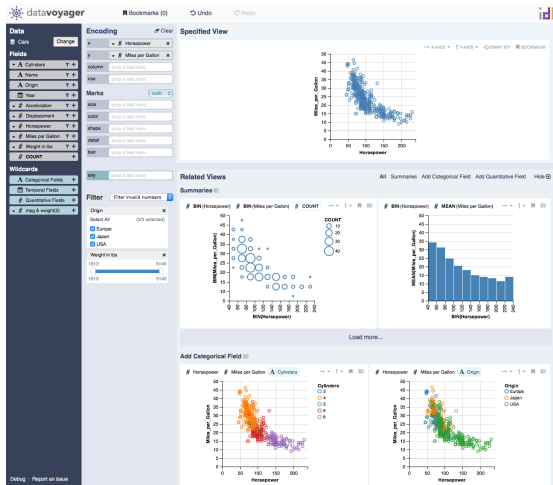
Always happy to chat!



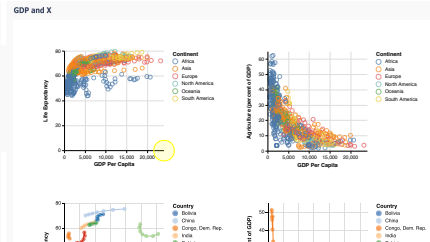


# Hi, I am Zening Qu

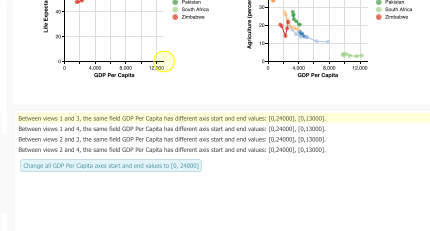
I like to **build tools** to assist people in **data exploration** and **presentation**.



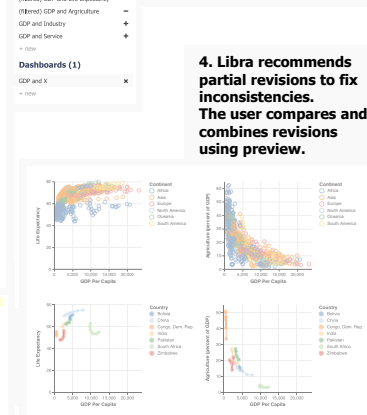
1. A user creates a chart by dragging data to encoding.



3. Libra detects inconsistencies and shows grouped & ranked text warnings. The user hovers over and sees visual highlights.




2. The user creates a multi-view dashboard using four charts.



4. Libra recommends partial revisions to fix inconsistencies. The user compares and combines revisions using preview.

 **datavoyager**  
(vis recommender)

 **Libra:** when you create dashboards, it checks guidelines & suggests revisions.

# Textbook

*An Introduction to Designing With D3*



O'REILLY®

*Scott Murray*

## Interactive Data Visualization for the Web, 2nd Edition

*For learning D3!*

Book available online.

Code / examples on GitHub.

We will be using **D3 v5**.

<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 posted by **Monday 11:59pm**.

You have 1 "pass" for the quarter.

# Assignments

Class Participation (10%)

A1: Visualization Design (10%) - *Due 10/1*

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 11/27*

Project Deliverables - *Due 12/6*



# Final Project

Produce **interactive web-based visualizations**

Initial **prototype** and **design review**

**Final deliverables** and **video 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

## **Data Visualization for Social Good**

*Goal: find data of social or scientific import, design visualizations to explore or communicate it effectively.*

The specific data domain is open-ended. Possibilities include transportation, housing, public health, education, climate, campaign finance, scientific research, and so on...

You must identify a target audience. May be general (citizens, voters) or specialized (scientists, policy makers).

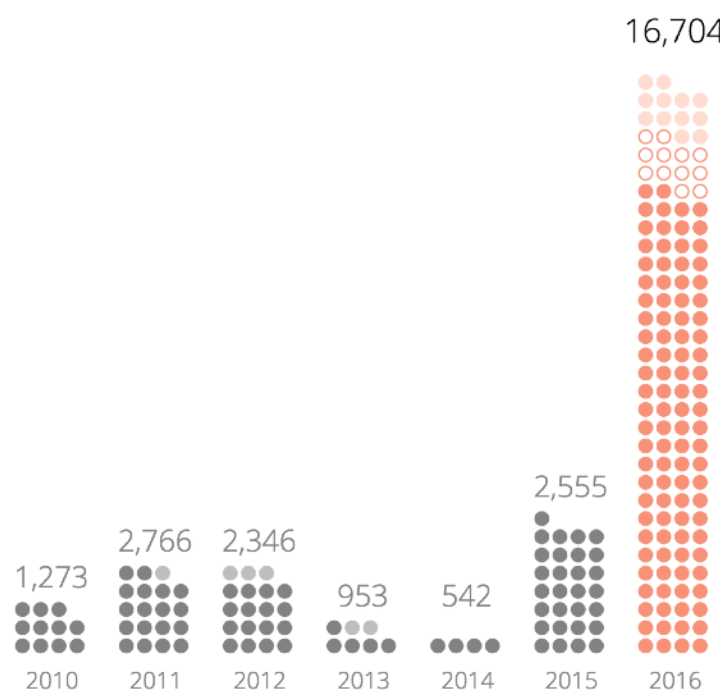
Use Assignment 2 to explore a data set of interest prior to committing to final project teams and topic!

**Inspiration...**

**Professional, Scientific and Technical Services** ●

● approx. 131 businesses

- Transportation and Warehousing
- Other Services (Except Public Administration)
- Retail Trade
- Construction
- Health Care & Social Assistance
- Arts, Entertainment, & Recreation
- Accommodation & Food Services
- Administrative & Support & Waste
- Wholesale Trade
- Manufacturing
- Real Estate, Rental & Leasing
- Information
- Educational Services
- Finance and Insurance
- Public Administration
- Management of Companies and Enterprises
- Agriculture, Forestry, Fishing and Hunting
- Utilities
- Mining
- Unclassified



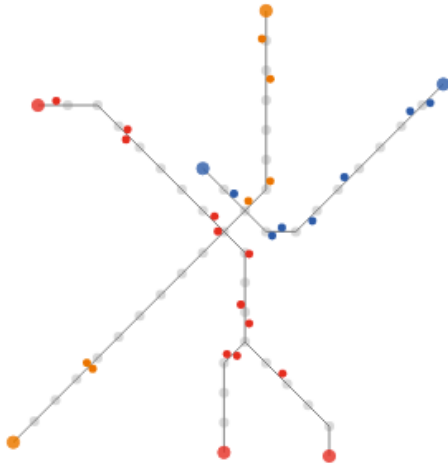
● new businesses ○ old businesses (records appearing in that year) ● old businesses  
 ● new businesses that got left behind ● old businesses that got left behind

Business Count



# Change In Times (CSE 442, Spring 2017)

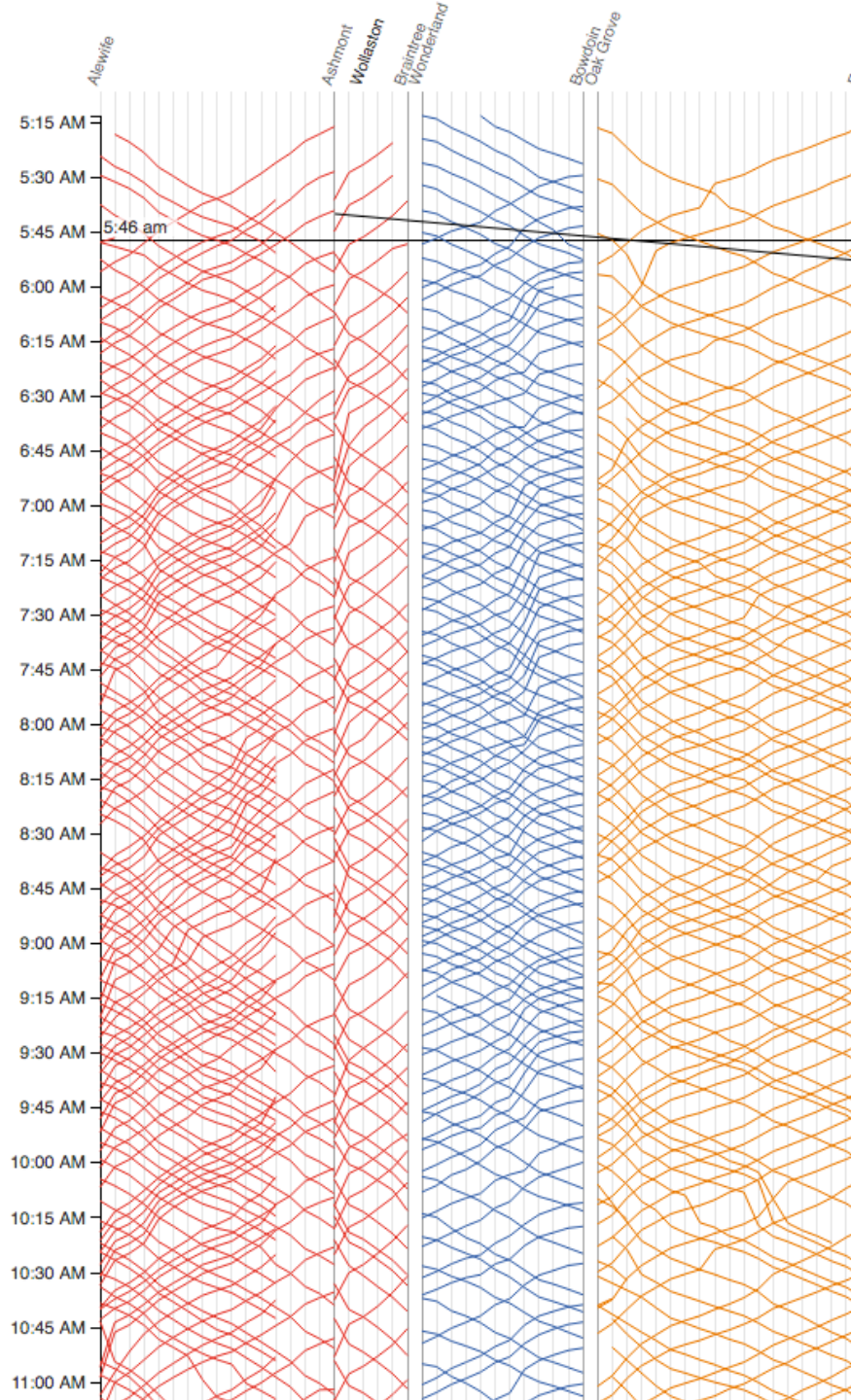
Gunnar Olson, Halden Lin, Lilian Liang, and Shobhit Hathi



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.

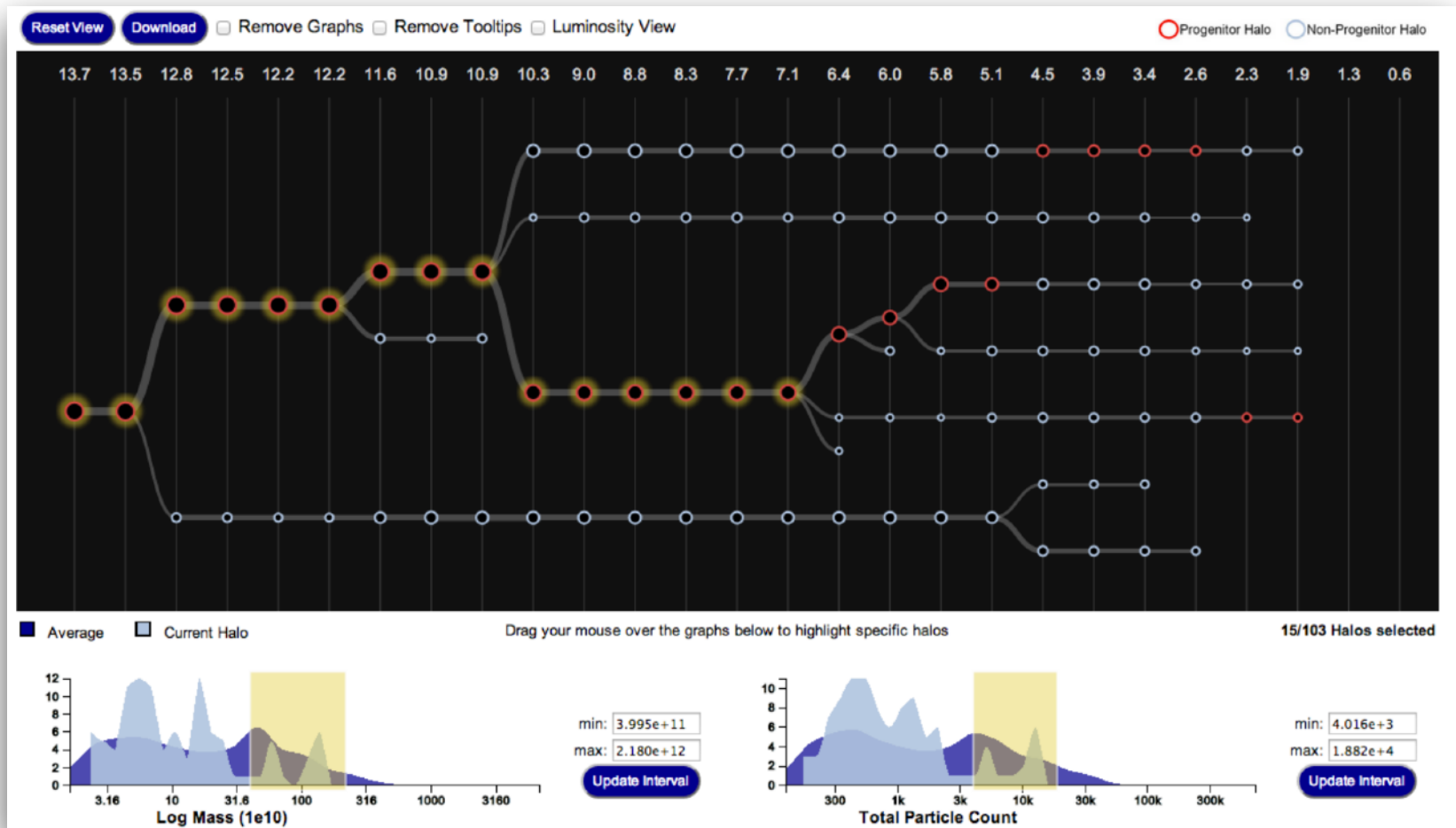
Since the red line splits, we show the Ashmont branch first then the Braintree branch. Trains on the Braintree branch "jump over" the Ashmont branch.

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

# MBTA Viz

Barry & Card

# Visualizing Galaxy Merger Trees



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

# Visualizing the Republic of Letters

Daniel Chang, Yuankai Ge, Shiwei Song

## Republic of Letters

1700



## 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



**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 ( $\leq 4$  paragraphs)

Due by **11:59 pm, Monday October 1.**