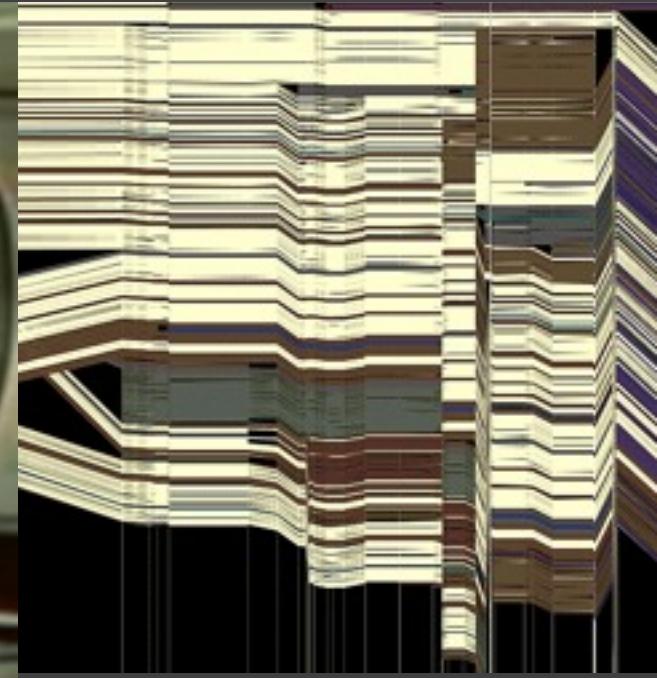
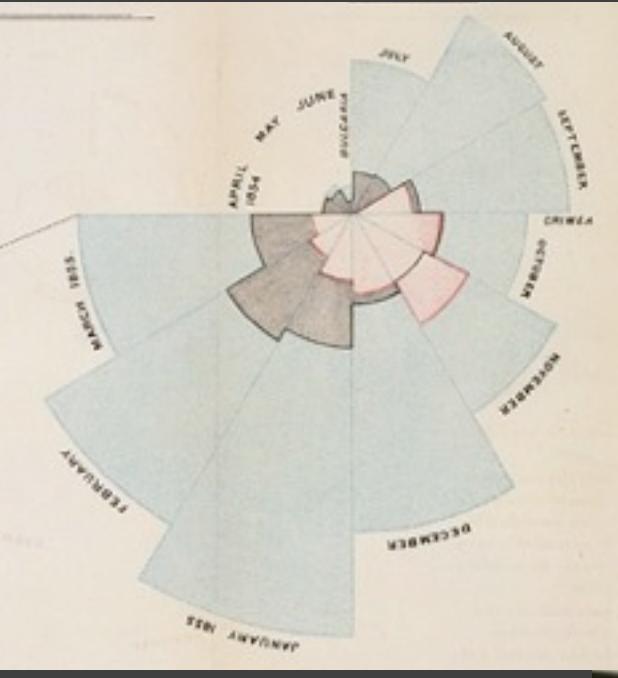


CSE512 :: 7 Jan 2014

# 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

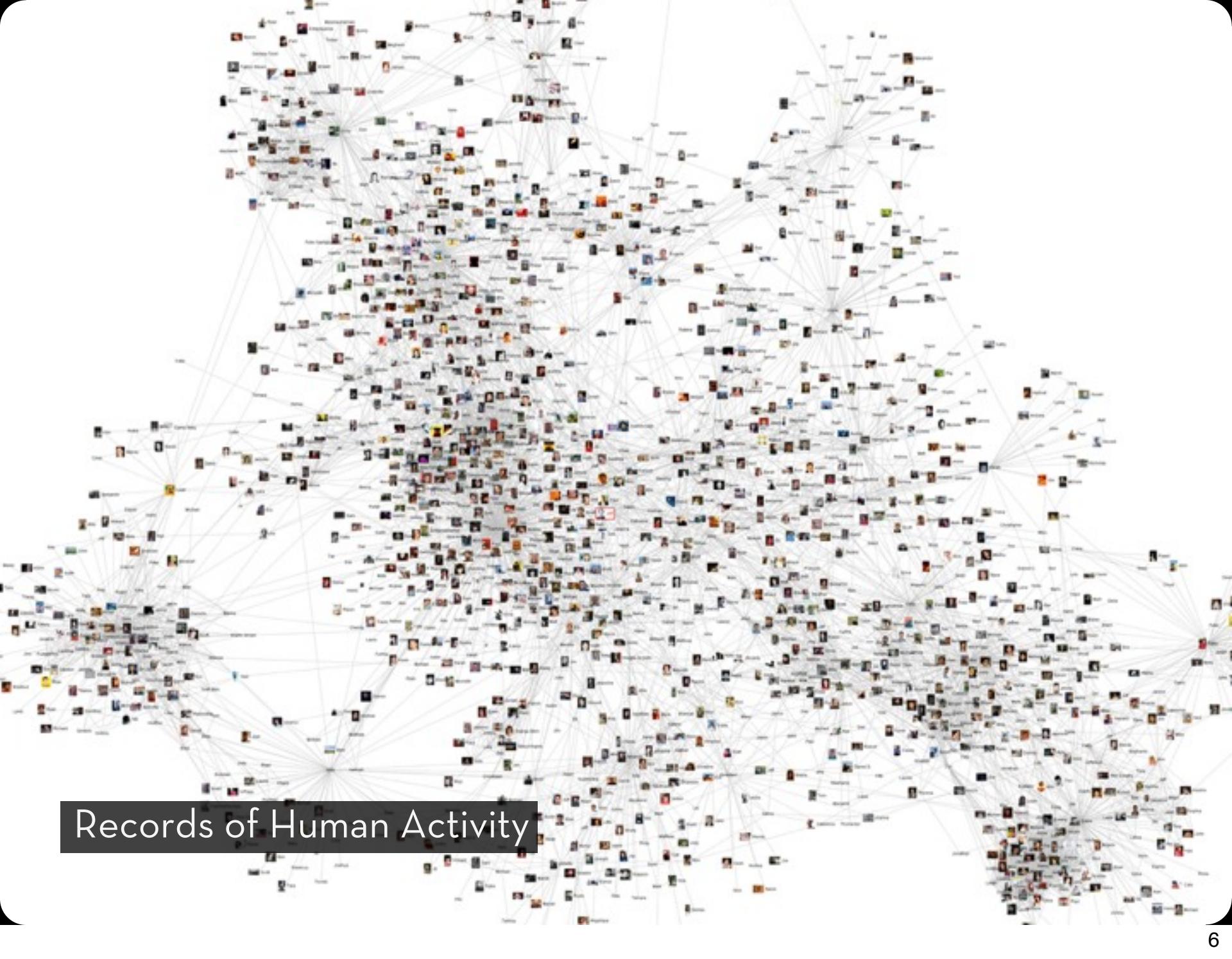


## Physical Sensors

*Image courtesy cabspotting.org*



Health & Medicine



Records of Human Activity

## Abortion

(Revision as of 22:56 4 Jan 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 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) are usually a **chemical abortion**, the usual method, though **incomplete abortion** is usually the only legal method, although research has uncovered similar rates of success from **methotrexate** and **misoprostol**. Concomitant with chemical abortion and exceeding up to about the thirteenth week is **suction-abortion**. After around the thirteenth week **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 **c-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 **postabortal dilation and extraction** (P&X) or a **hysterotomy abortion**, similar to a **c-section**.

### The controversy

The morality and legality of abortion is a highly controversial topic in **applied ethics** and is also discussed by **legal scholars** and **religious scholars**. 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. **Controversy politics** in the **United States** and elsewhere has led to abortion becoming commonly accepted by the general public. 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 positions of the respective positions are subject to hot debate. While those on both sides of the issue are generally peaceful, if heated, in their defense of their position, 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.

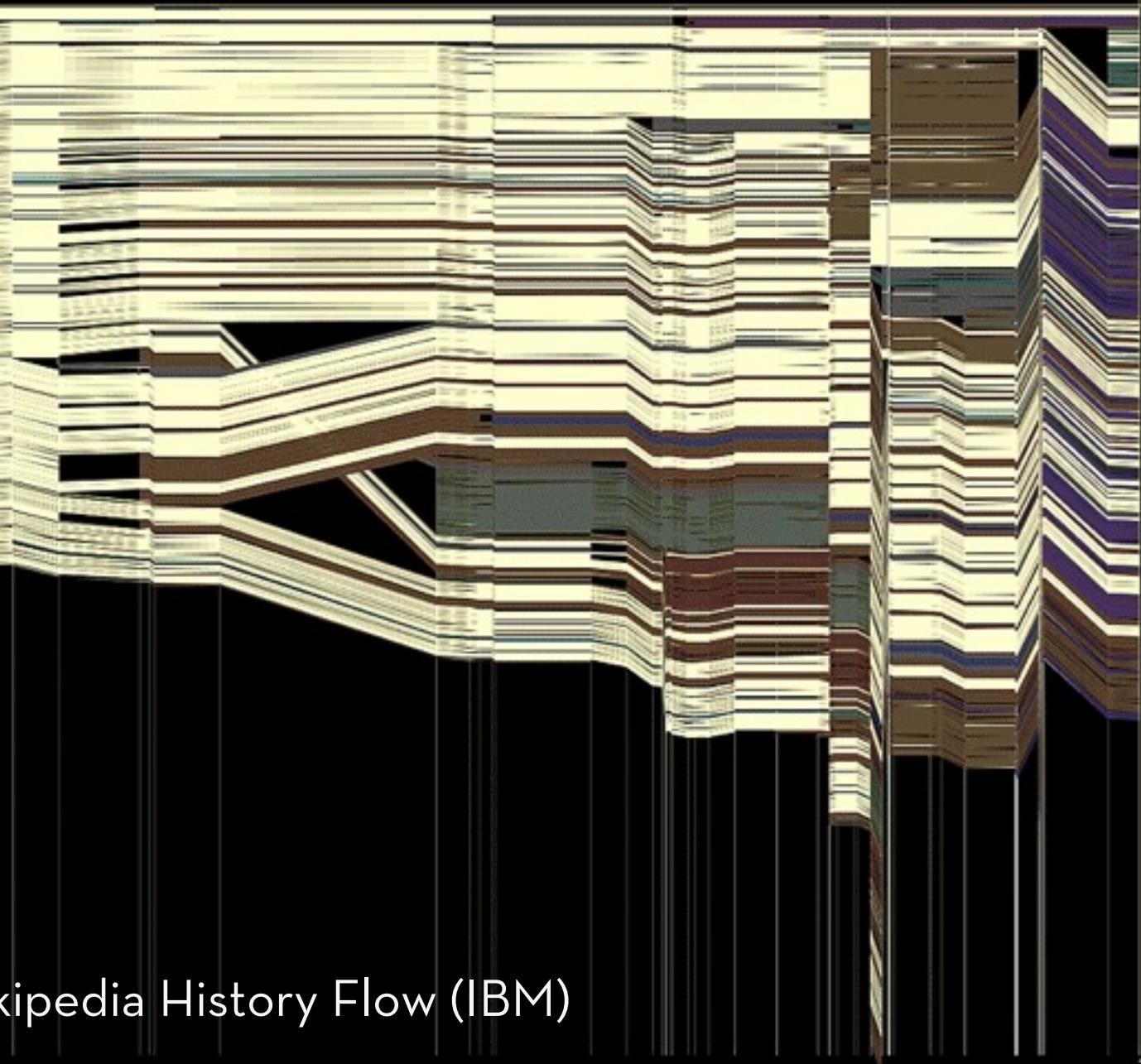
### The central question

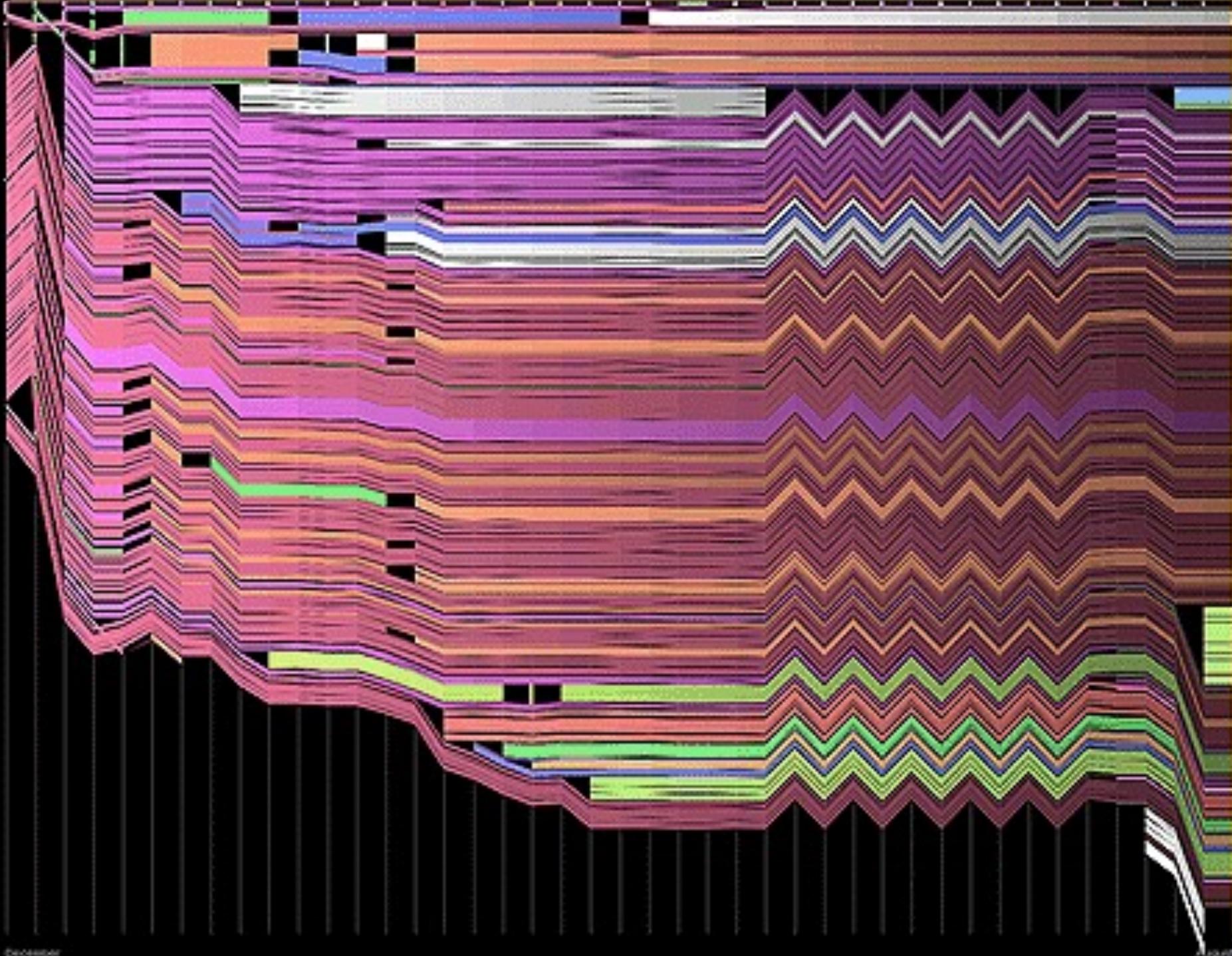
The central question in the abortion debate is the clash of presumed or perceived rights. On one hand, is a fetus (sometimes called the "unborn") a person? pro-life/anti-abortion advocates) a human being with a right to life, and if so, at what point does pregnancy does the fetus become human? On the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

# Wikipedia History Flow (IBM)

December  
2001

June  
2009





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

### Summary Statistics

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

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

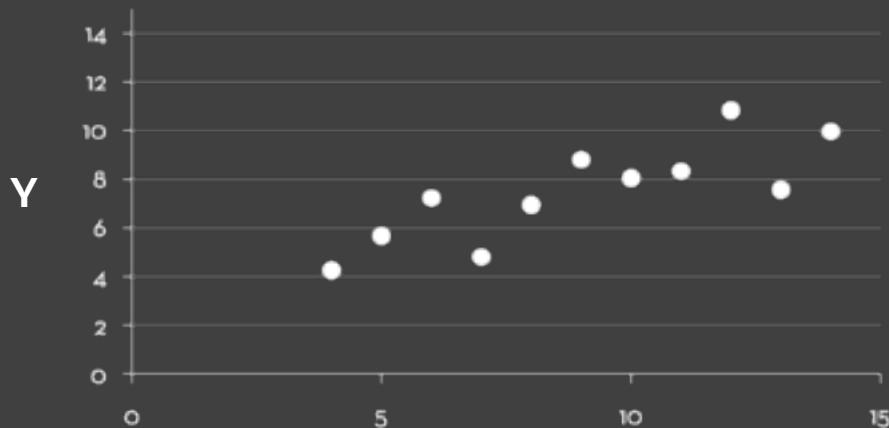
### Linear Regression

$$Y = 3 + 0.5 X$$

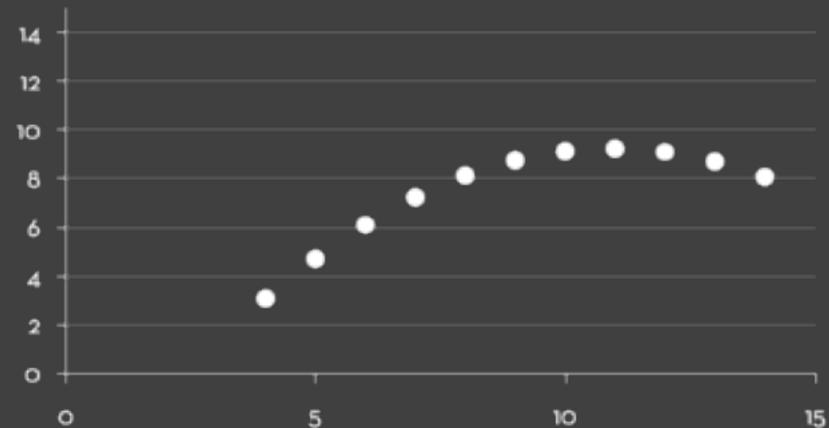
$$R^2 = 0.67$$

[Anscombe 73]

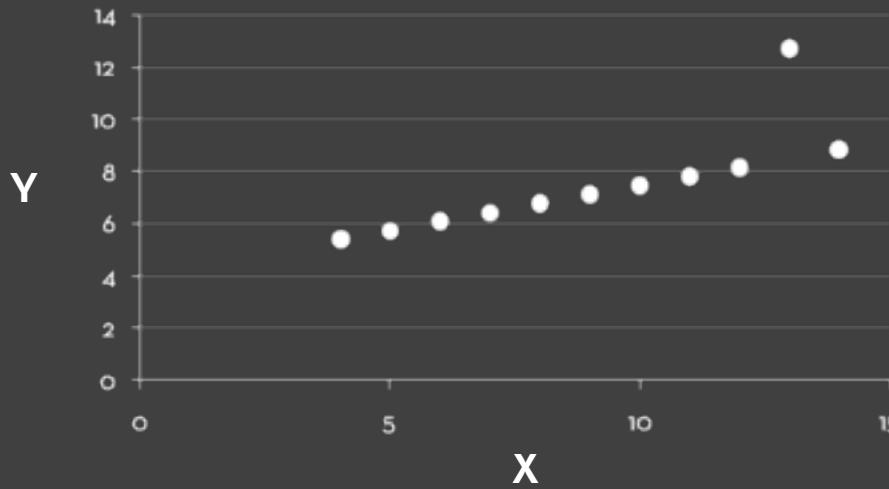
# Set A



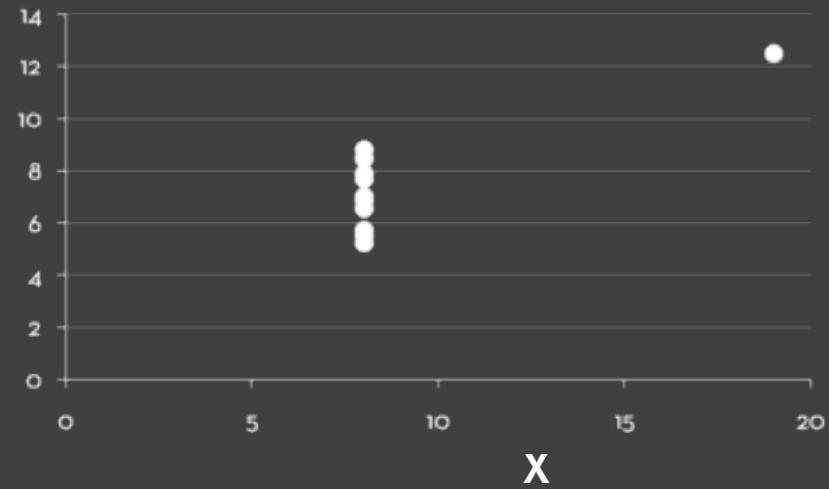
# Set B



# Set C



# Set D



# Why create visualizations?

# 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

# The Value of Visualization

## **Record** information

Blueprints, photographs, seismographs, ...

## **Analyze** data to support reasoning

Develop and assess hypotheses

Discover errors in data

Expand memory

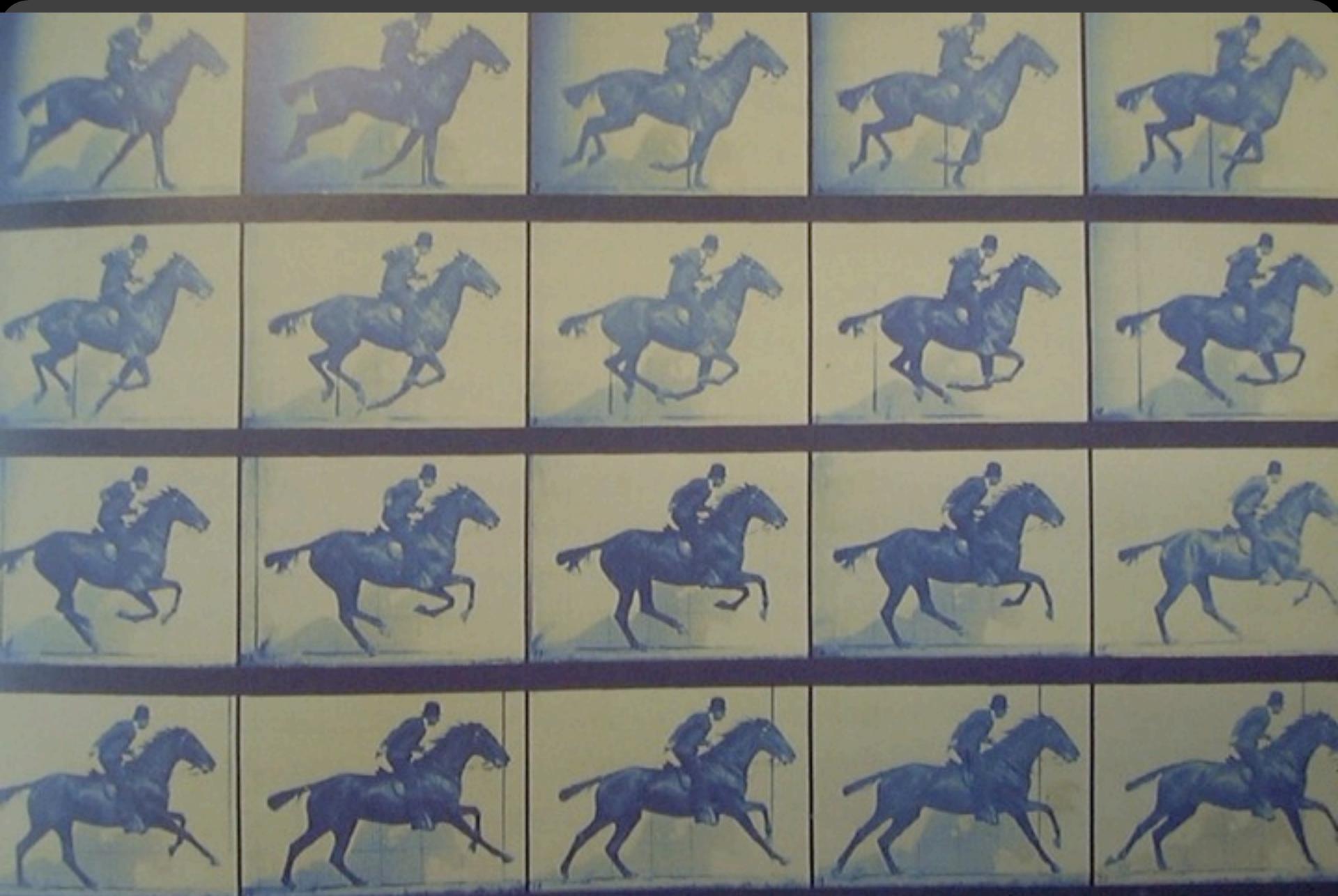
Find patterns

## **Communicate** information to others

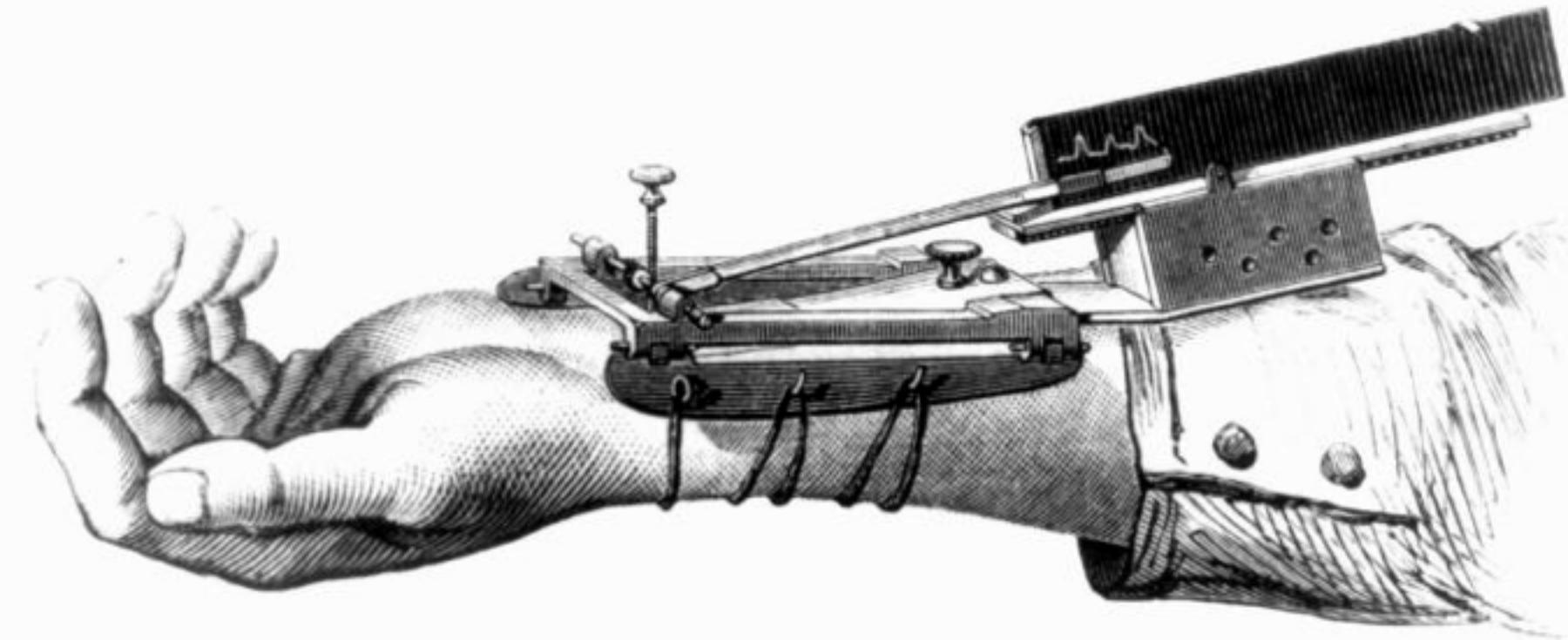
Share and persuade

Collaborate and revise

# Record Information



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

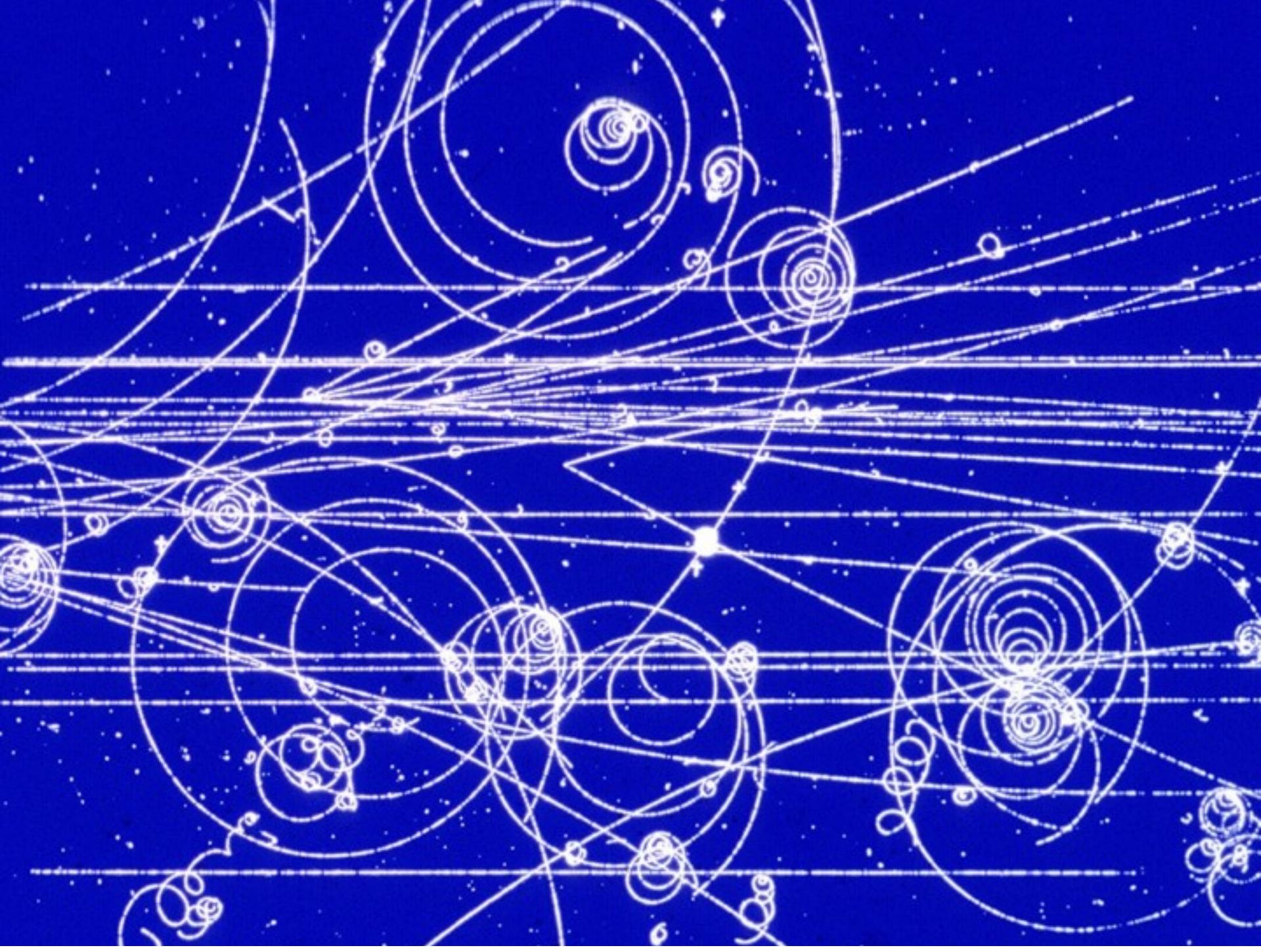


1.

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		642	462	20	11	503	45	540	429	429	479	215	67	799
	858	656	580	806	740	519	181	302	320	17	425	429	479	205	579
	962	903	253	223	802	948	303	28	691	503	230	608	671	547	597
	867	337	545	7949	341	35-6974	786	14	685	886	643	935	671	547	582
330				252	497	3341	577	728	856	872	951	885	285	10	306
635	750	5339	342	785	592	941	32	273	890	736	924	993	957	833	283
12	46	332	47	4849	634	6	469	871	895	940	729	919	436	912	271
	642			300	280	275	25	37	168	825	720	838	52	839	57
51	726	50	552	859	391	308	848	795	559	639	549	895	15	997	43594
	936						945	587	43	18	99	274	610	698	792313
482	599	259	646	869	944	493	870	621	683	495	887	651	590	549	965
	963	755	539	562	553						628	834	939	63	432
	520										742	783	686	24	84
	901										600	737	686	221	920
	415														

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.)	
22A	None	None	0.280	None	None	36°--66°
22A	NONE	NONE	0.280	NONE	NONE	330°-18°
15A	0.010	154.0	0.280	4.25	5.25	163
15B	0.038	130.0	0.280	12.50	58.75	354
15B	None	45.0	0.280	None	29.50	354
41D RH Forward Field	0.028	110.0	0.280	3.00	None	275
41C LH Aft Field*	None	None	0.280	None	None	--
41B LH Forward Field	0.040	217.0	0.280	3.00	14.50	351
STS-2 RH Aft Field	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 (30°), (110°) ARC
- MUCH WORSE VISUALLY THAN SRM-22

HISTORY OF O-RING TEMPERATURES (DEGREES - F)

MOTOR	MGT	AMB	O-RING	WIND
DM-4	68	36	47	10 MPH
DM-2	76	45	52	10 MPH

SRM-22 BLOW-BY

- 2 CASE JOINTS (30-40°)

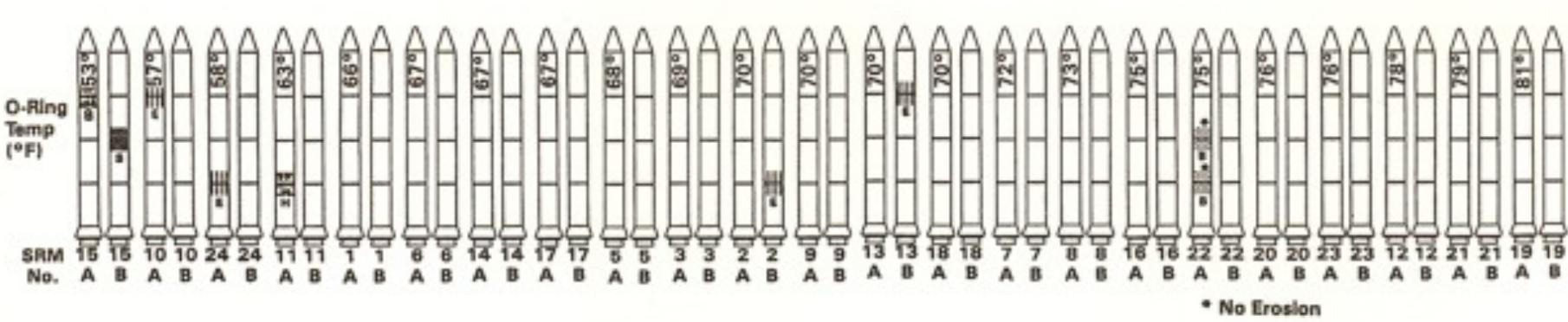
QM-3	72.5	40	48	10 MPH
QM-4	76	48	51	10 MPH

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

- NOZZLE Blow-by

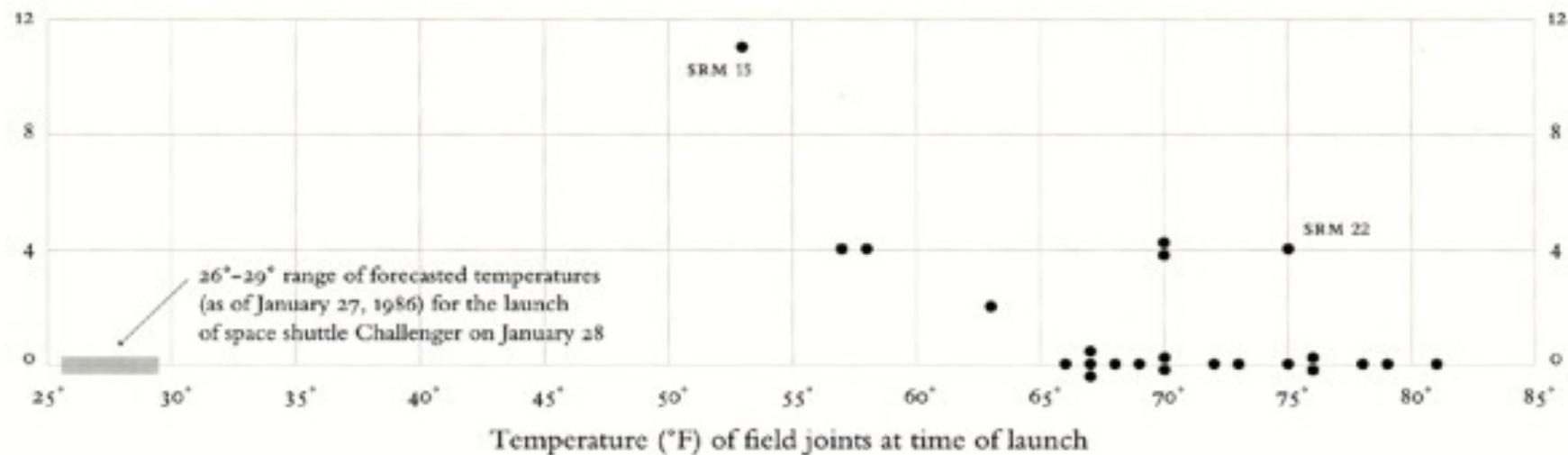
SRM-15	52	64	53	10 MPH
SRM-22	77	78	75	10 MPH
SRM-25	55	26	29 27	10 MPH 25 MPH

# Make a decision: Challenger



# Make a decision: Challenger

O-ring damage  
index, each launch



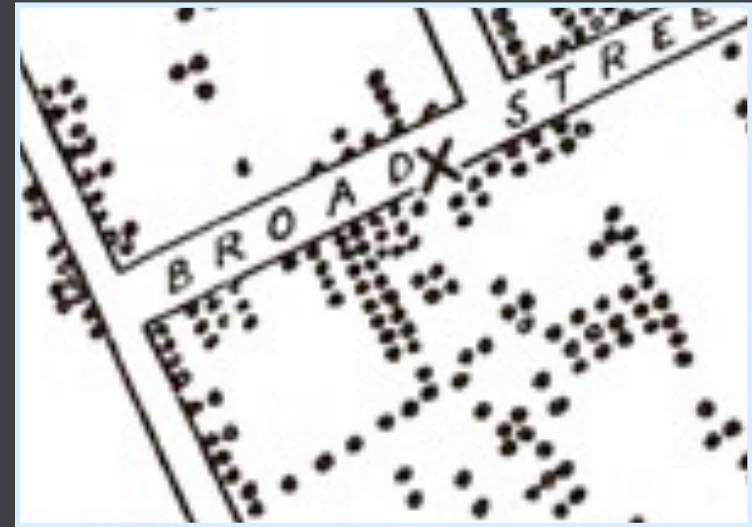
Visualizations drawn by Tufte show how low temperatures damage O-rings [Tufte 97]

# 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

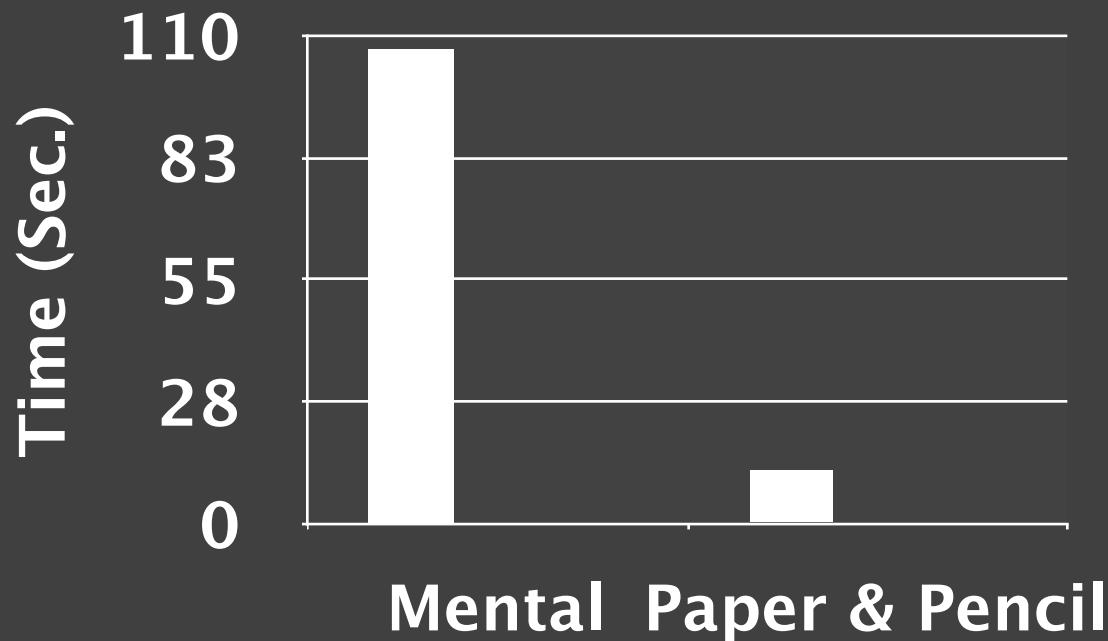
Class Exercise

# Expand memory: Multiplication

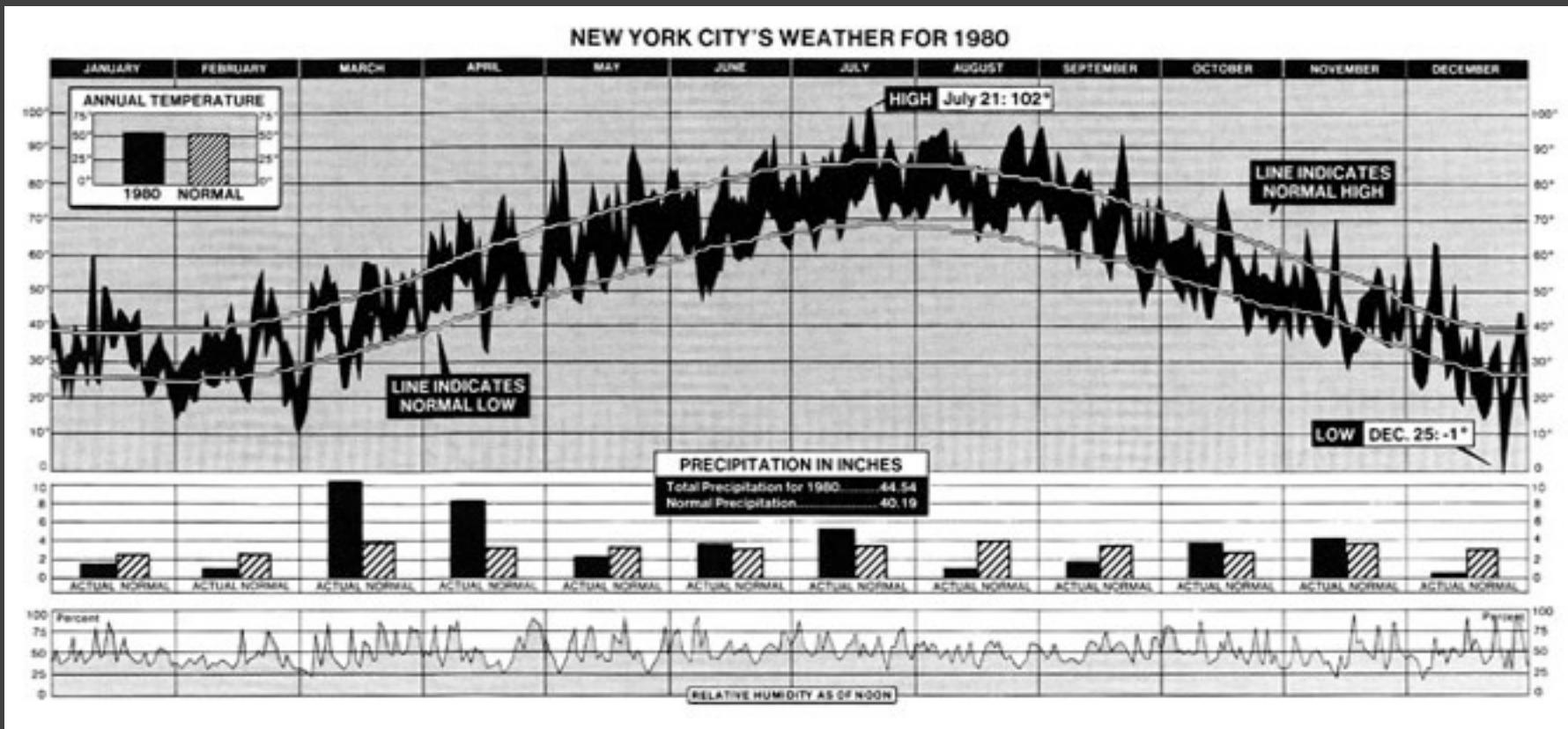
$$\begin{array}{r} 34 \\ \times 72 \\ \hline \end{array}$$

# Expand memory: Multiplication

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

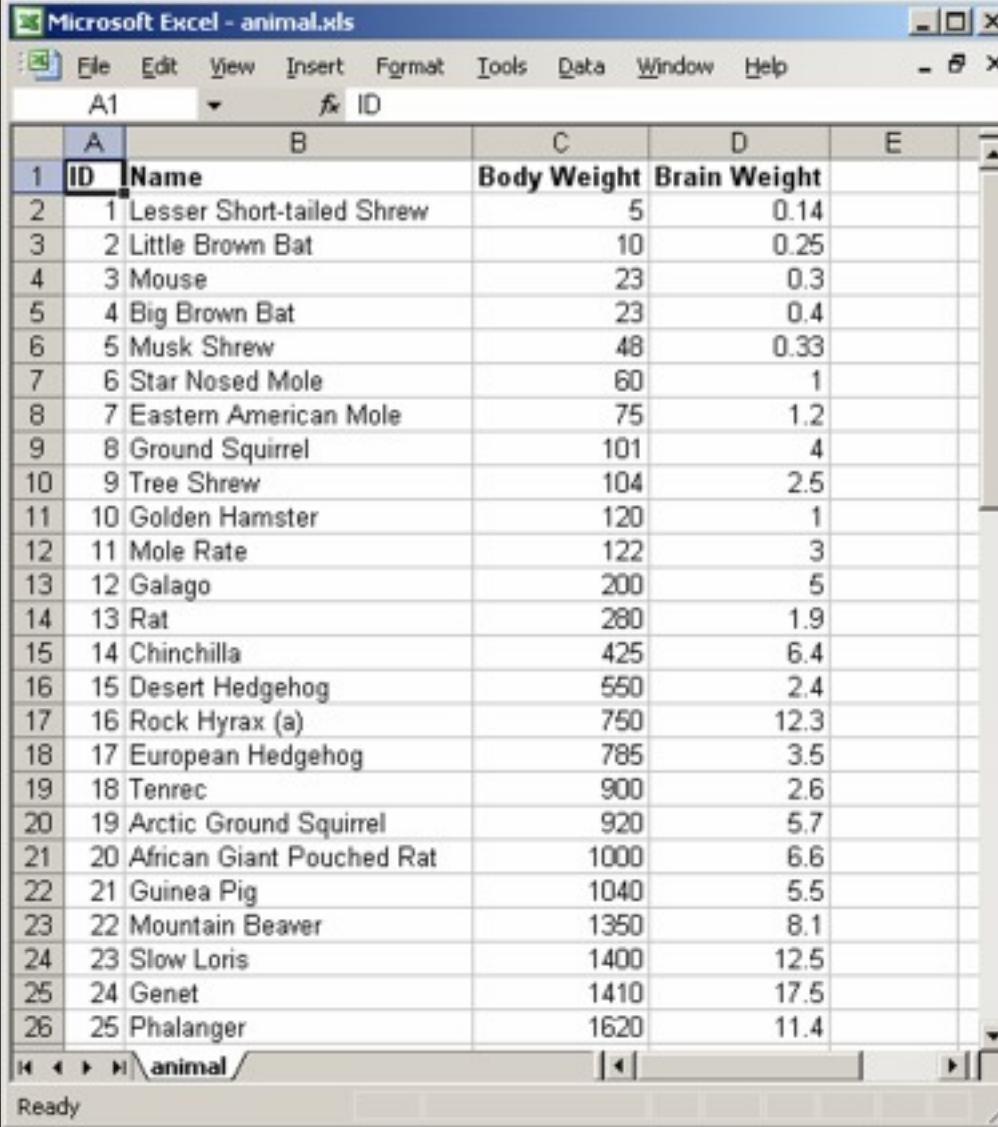


# Find patterns: NYC weather



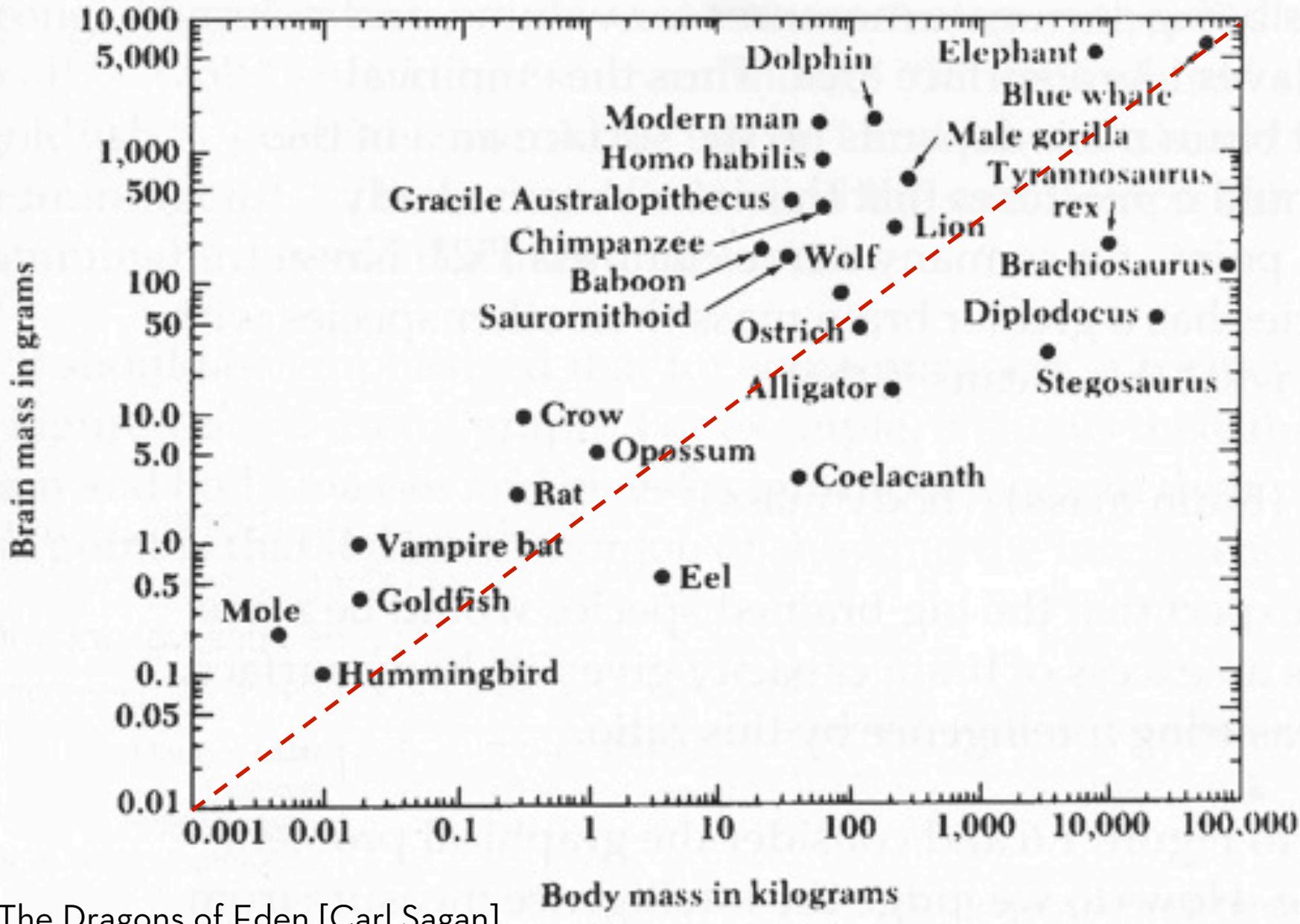
From the New York Times 1981

# The most powerful brain?

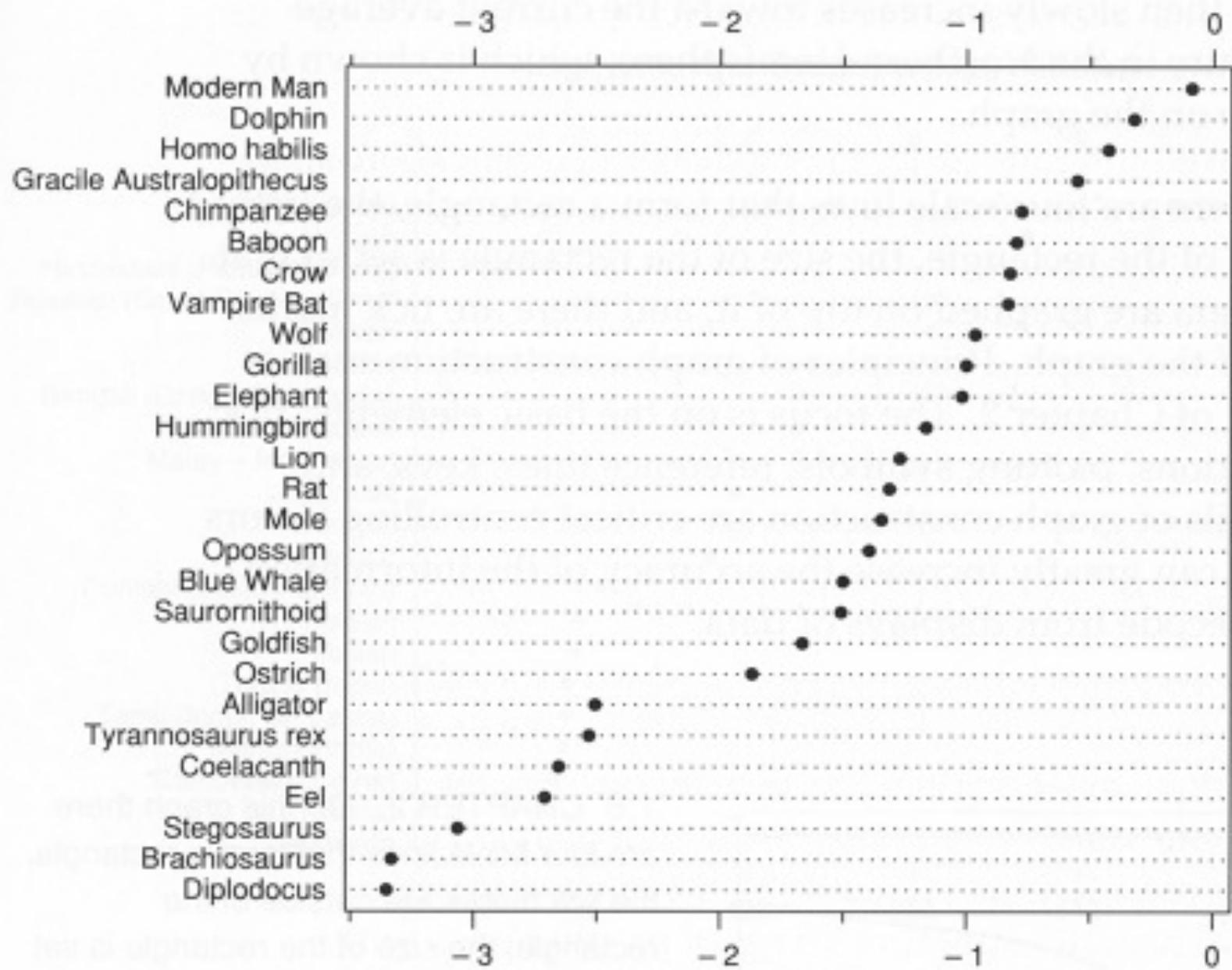


A screenshot of a Microsoft Excel spreadsheet titled "animal.xls". The spreadsheet contains data for various animals, organized into columns A through E. Column A is labeled "ID", column B is labeled "Name", column C is labeled "Body Weight", and column D is labeled "Brain Weight". The data includes entries for a Lesser Short-tailed Shrew (ID 1), a Little Brown Bat (ID 2), a Mouse (ID 3), a Big Brown Bat (ID 4), a Musk Shrew (ID 5), a Star Nosed Mole (ID 6), an Eastern American Mole (ID 7), a Ground Squirrel (ID 8), a Tree Shrew (ID 9), a Golden Hamster (ID 10), a Mole Rat (ID 11), a Galago (ID 12), a Rat (ID 13), a Chinchilla (ID 14), a Desert Hedgehog (ID 15), a Rock Hyrax (ID 16), a European Hedgehog (ID 17), a Tenrec (ID 18), an Arctic Ground Squirrel (ID 19), an African Giant Pouched Rat (ID 20), a Guinea Pig (ID 21), a Mountain Beaver (ID 22), a Slow Loris (ID 23), a Genet (ID 24), and a Phalanger (ID 25). The "Body Weight" column shows values ranging from 5 to 1620, while the "Brain Weight" column shows values ranging from 0.14 to 17.5.

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 Rat	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]

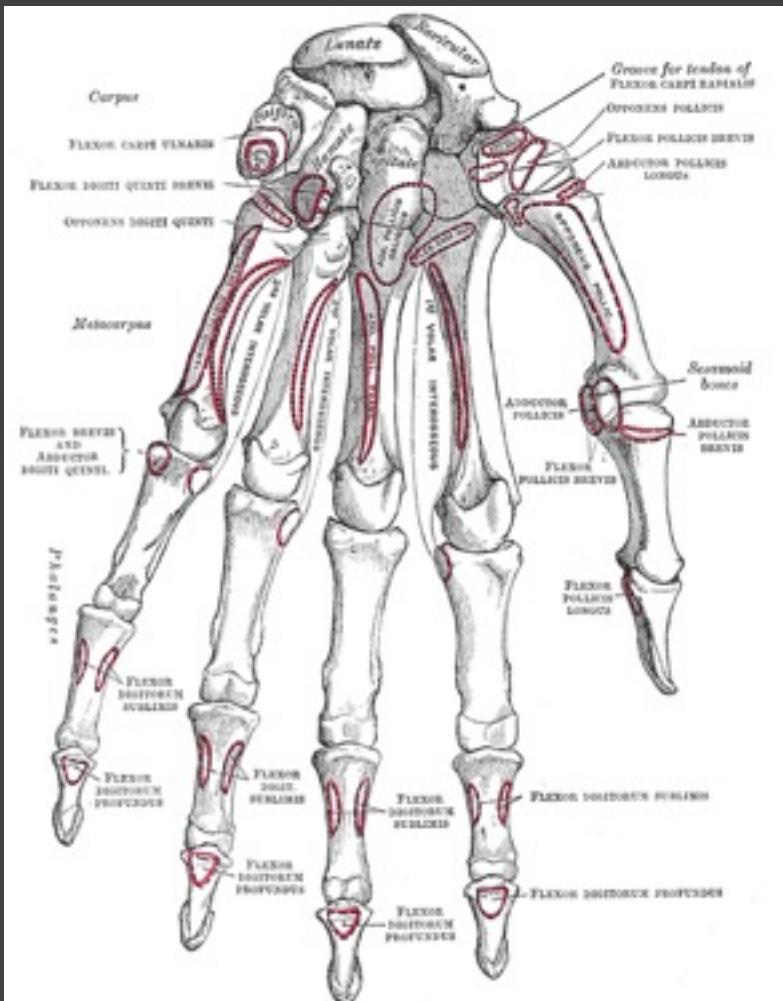


The Elements of Graphing Data  
[Cleveland]

Log<sub>10</sub> Brain Weight –  $\frac{2}{3}$  Log<sub>10</sub> Body Weight

# Convey Information to Others

# Inspire



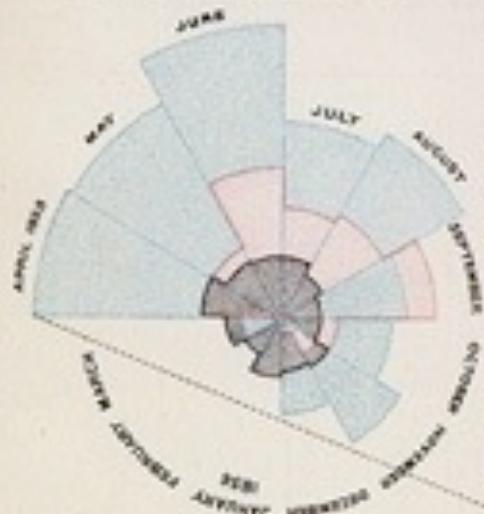
## Bones in hand [from 1918 edition]



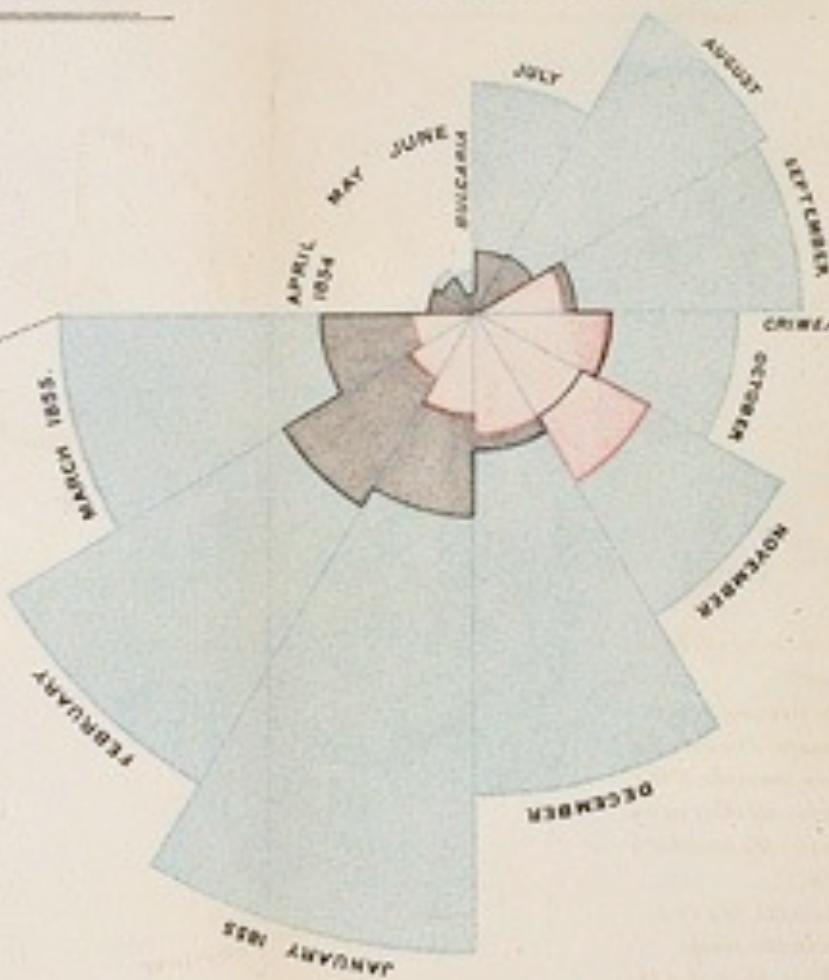
## Double helix model [Watson and Crick 53]

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”

Harrison & Sons, 32, Queen Street, London.

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

Discover errors in data

Expand memory

Find patterns

**Communicate** information to others

Share and persuade

Collaborate and revise

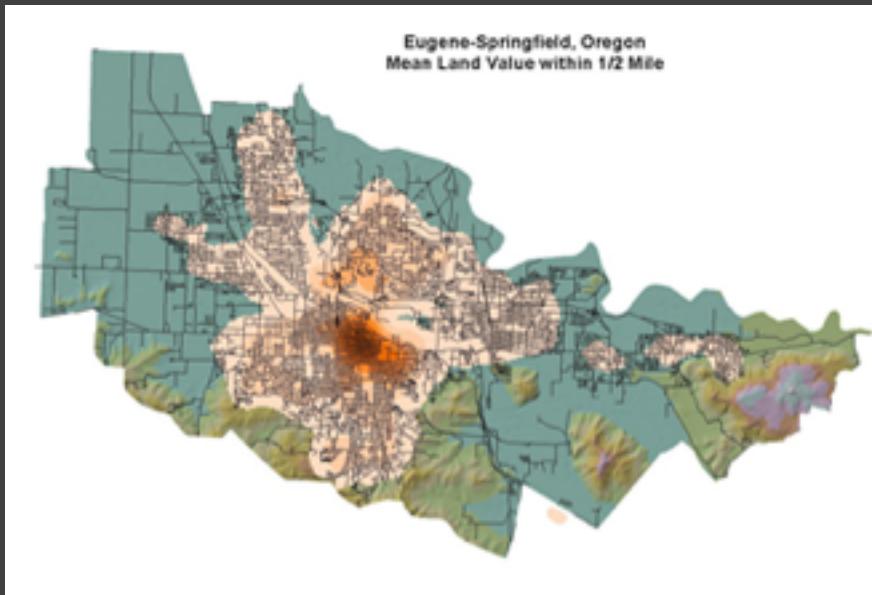
# Visualization Research

# Challenge

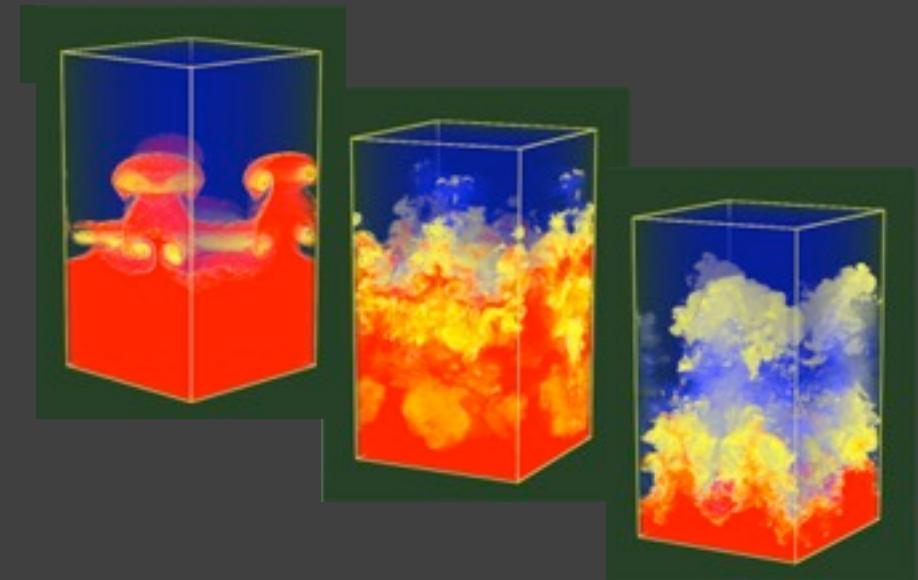
- More and more unseen data
  - Faster creation and collection

# Challenge

- More and more unseen data
  - Faster creation and collection



**Urban development planning**  
[www.urbansim.org](http://www.urbansim.org)



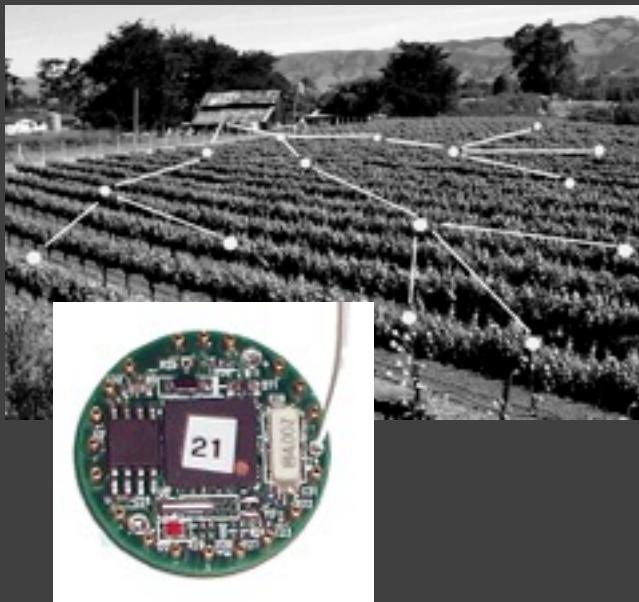
**Simulation**

# Challenge

- More and more unseen data
  - Faster creation and collection



**Sloan digital sky survey**  
[www.sdss.org](http://www.sdss.org)



**Sensor networks [Hill 02]**  
[www.xbow.com](http://www.xbow.com)

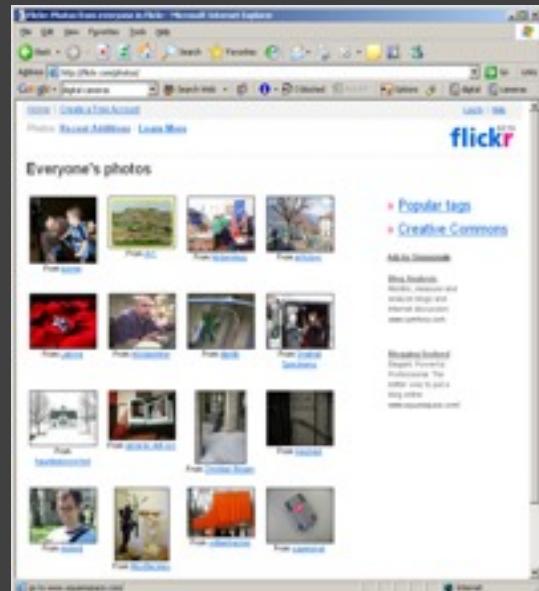


**Digital photography**

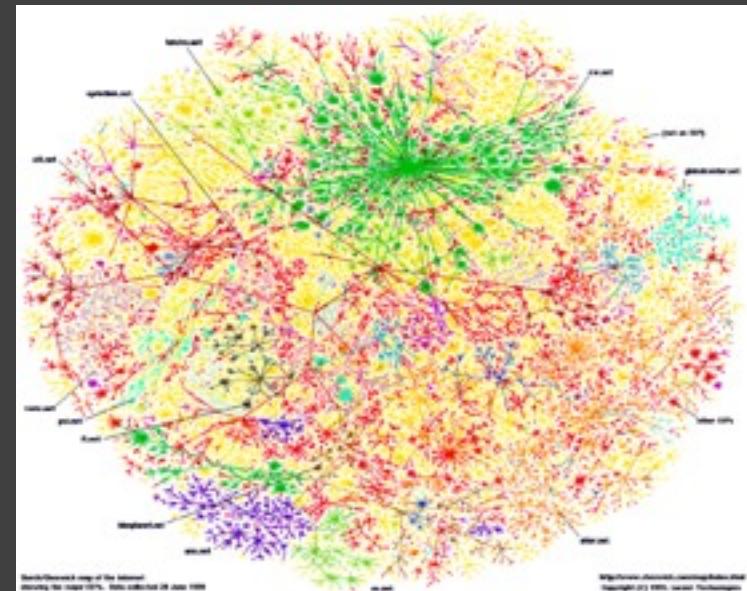
**Sensing**

# Challenge

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



**Photo sharing/annotation**  
[flickr.com](http://flickr.com)



**Group Authoring**  
[wikipedia.org](http://wikipedia.org)

Internet

**Map of the Internet [Cheswick 99]**  
[research.lumeta.com](http://research.lumeta.com)

# 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 correspond with mental models?
- 2 **Develop principles and techniques** for creating effective visualizations and supporting analysis
  - Amplify perception and cognition
  - Strengthen tie between visualization and mental models

# Course Topics

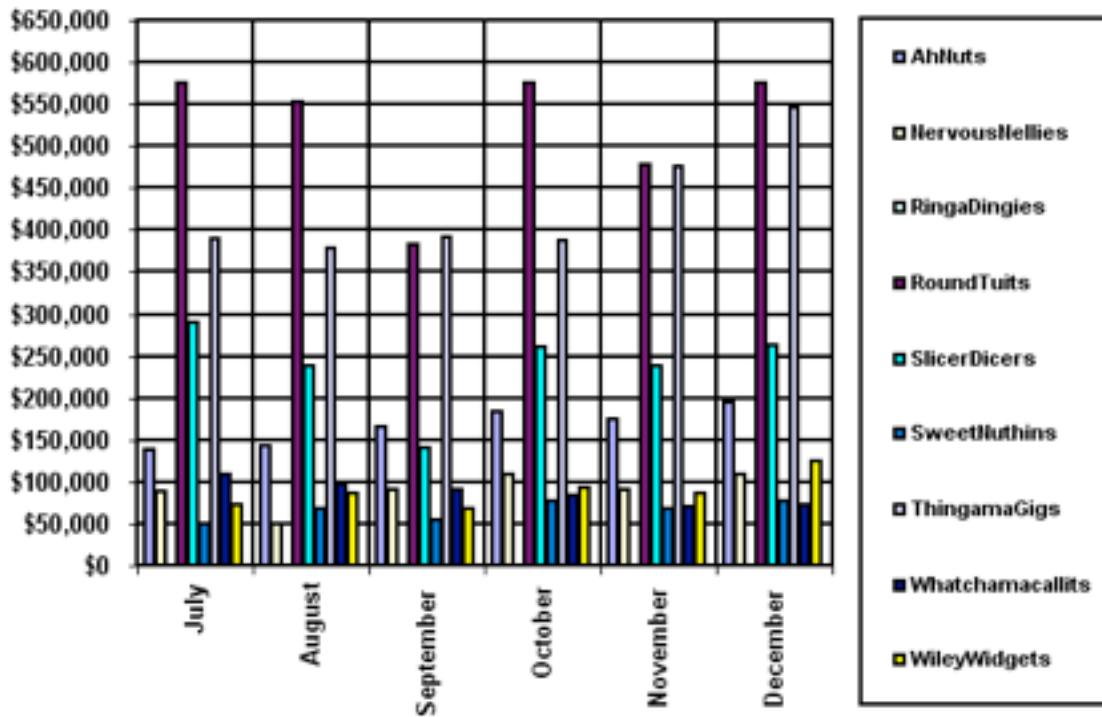
# Data and Image Models

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

Sémiologie Graphique [Bertin 67]

# Visualization (Re-)Design

SlicerDicers' Sales Compared to Other Products

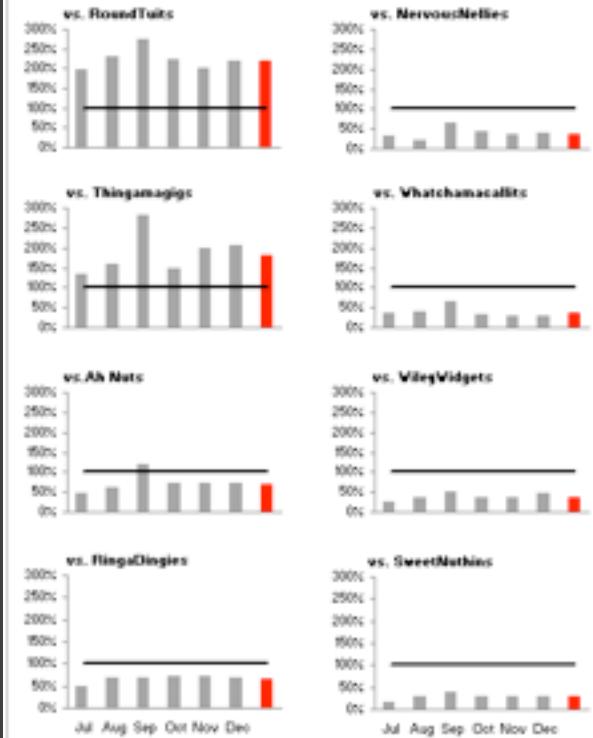


Problematic design

Sales of SlicerDicers Compared to Other Products

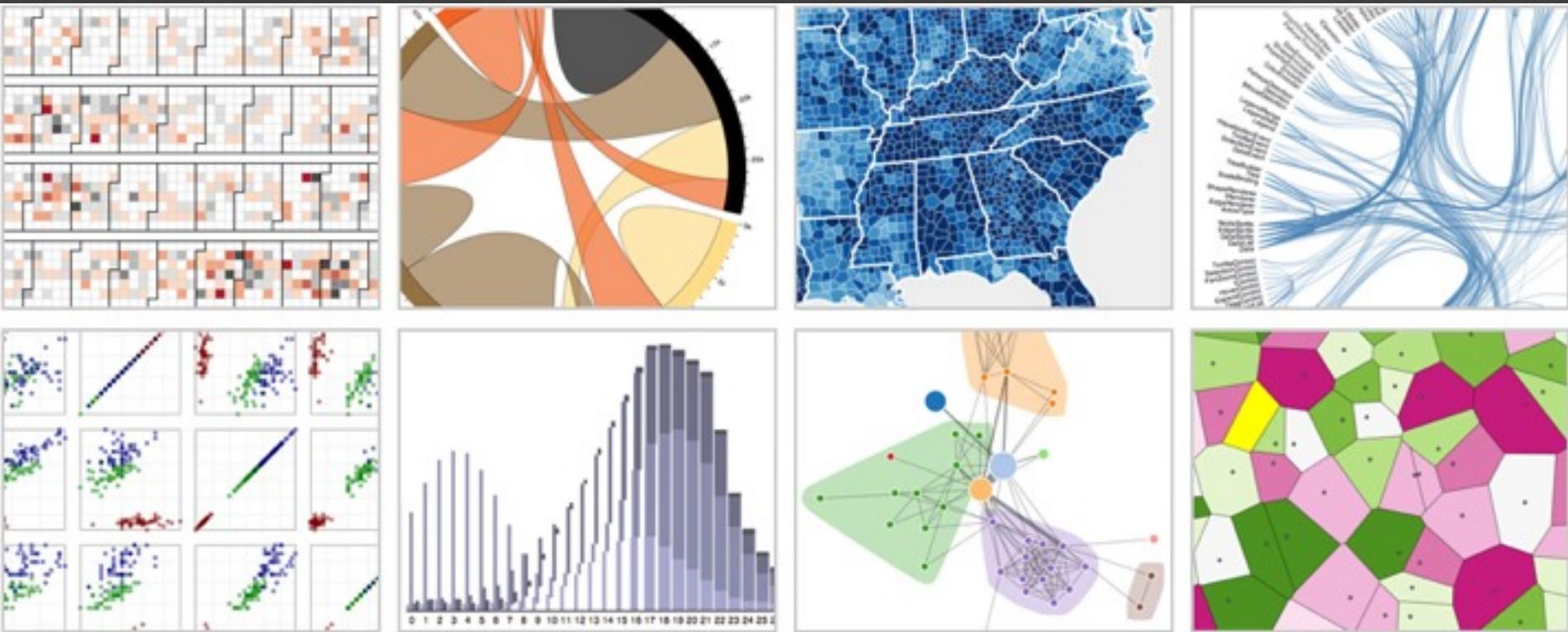
July - December, 2003

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



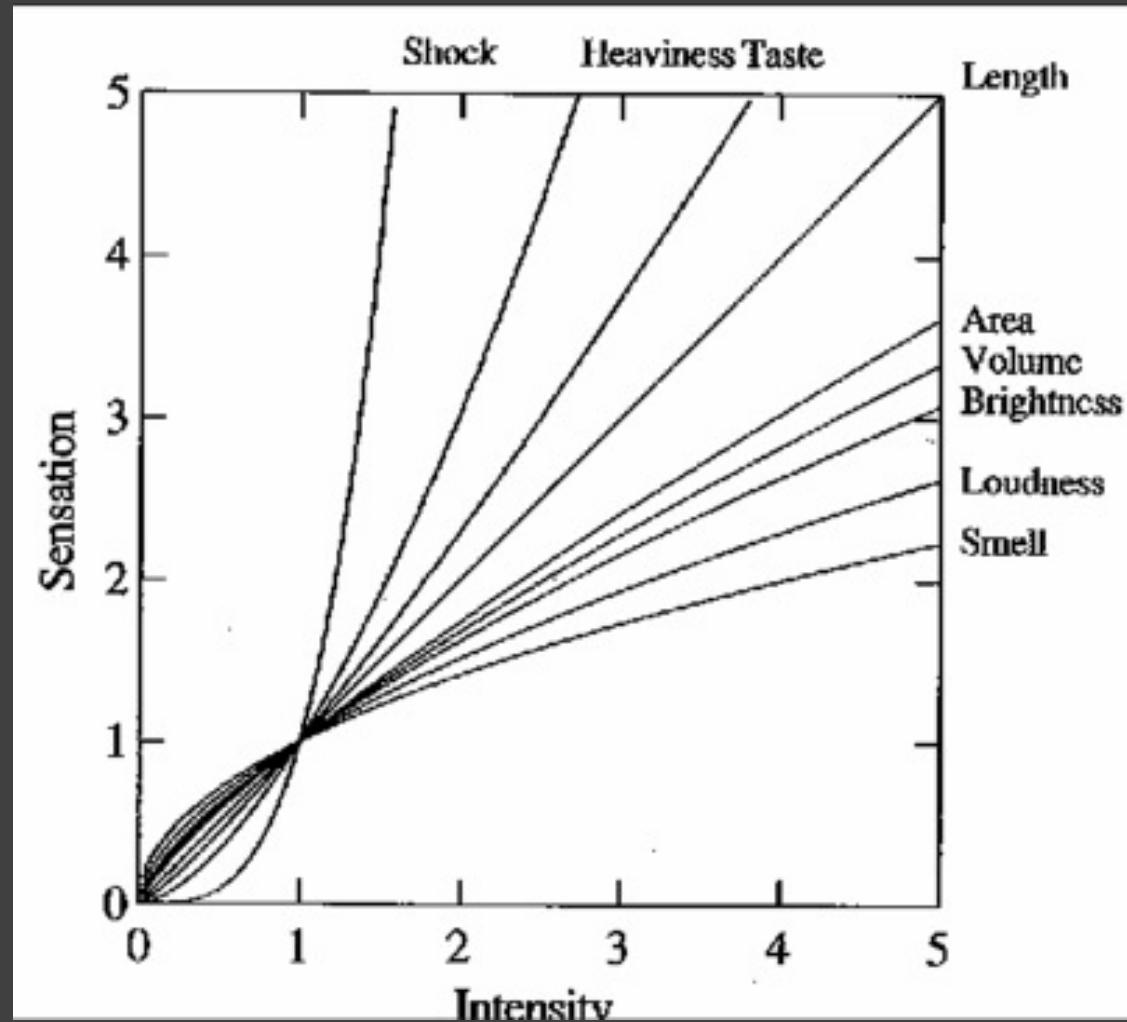
Redesign

# Visualization Software



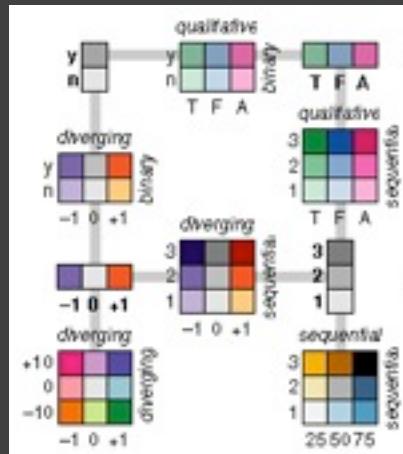
D3: Data-Driven Documents

# Graphical Perception

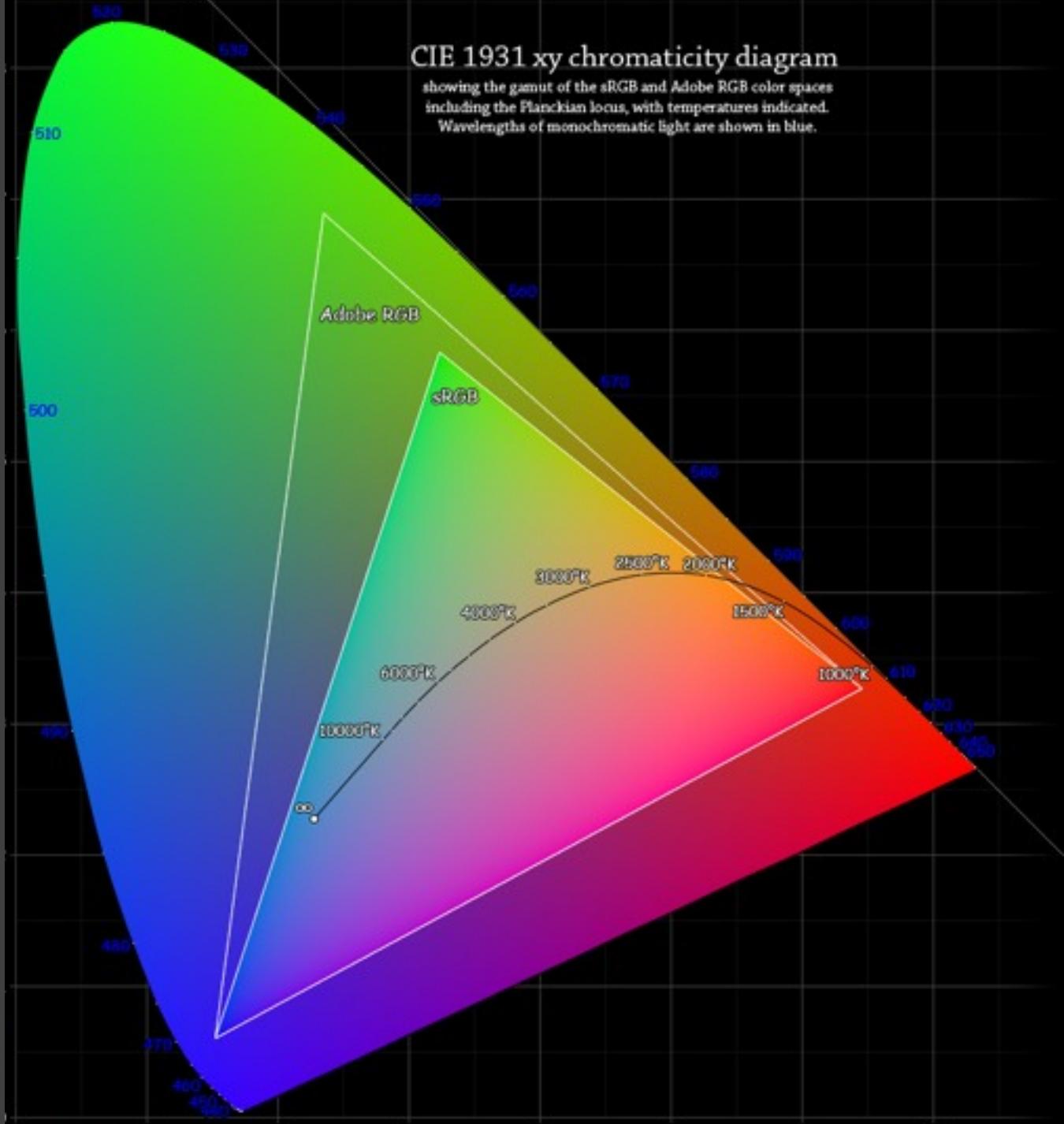


The psychophysics of sensory function [Stevens 61]

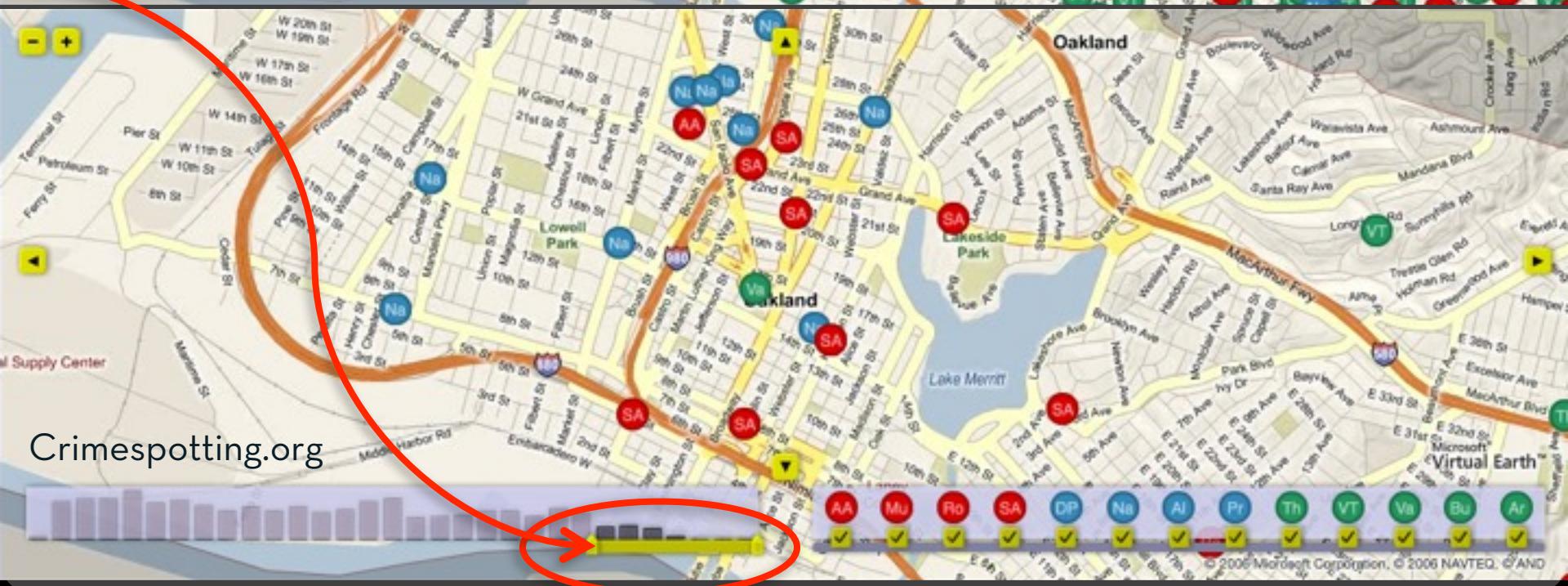
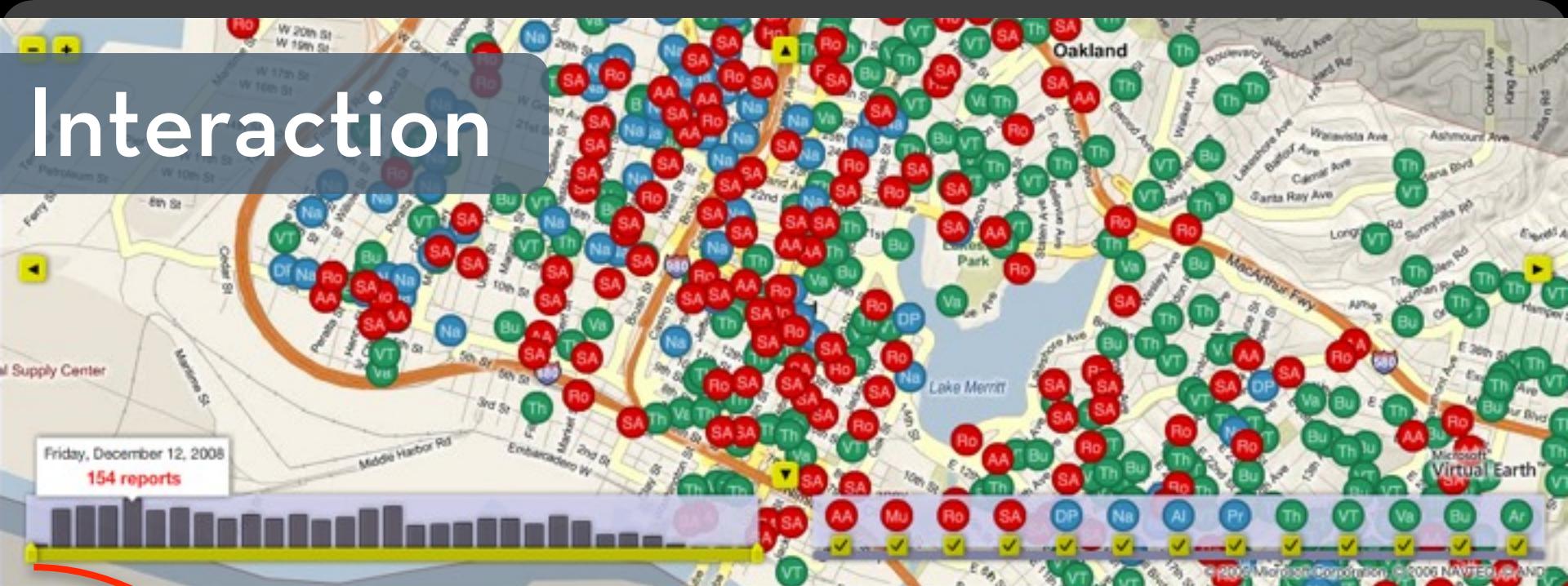
# Color



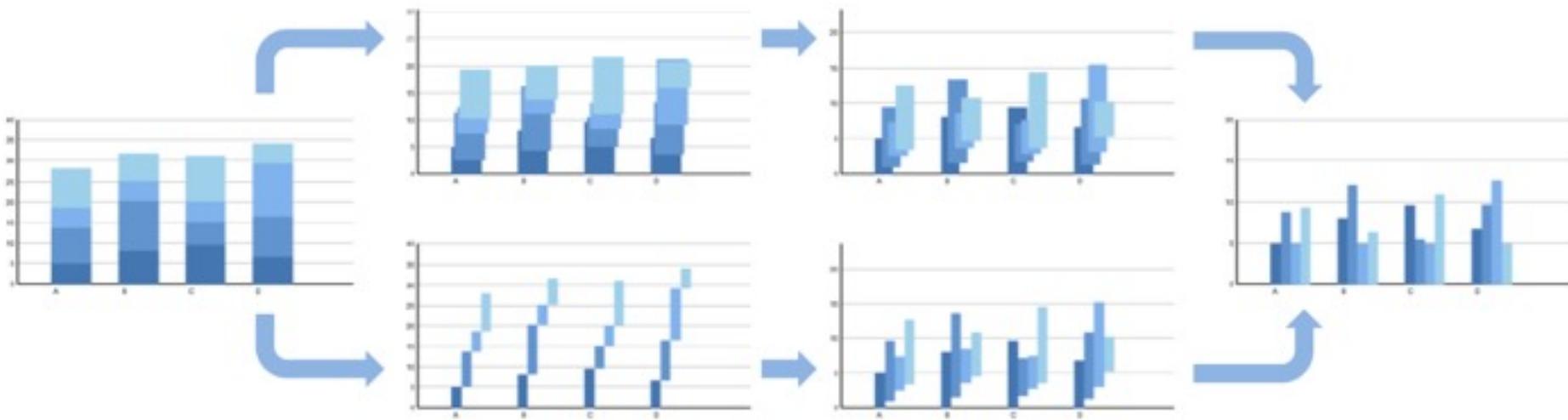
Color Brewer



# Interaction

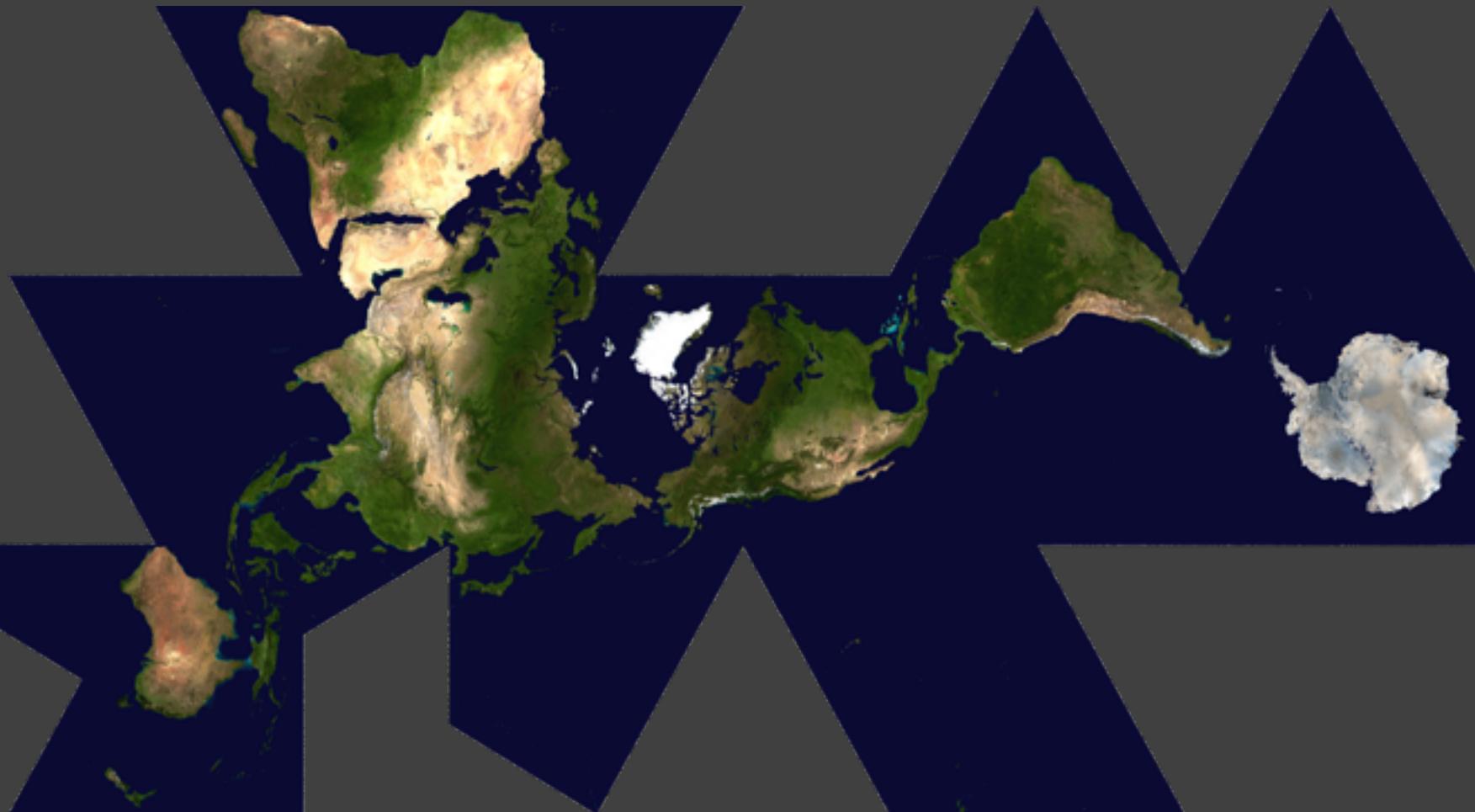


# Animation



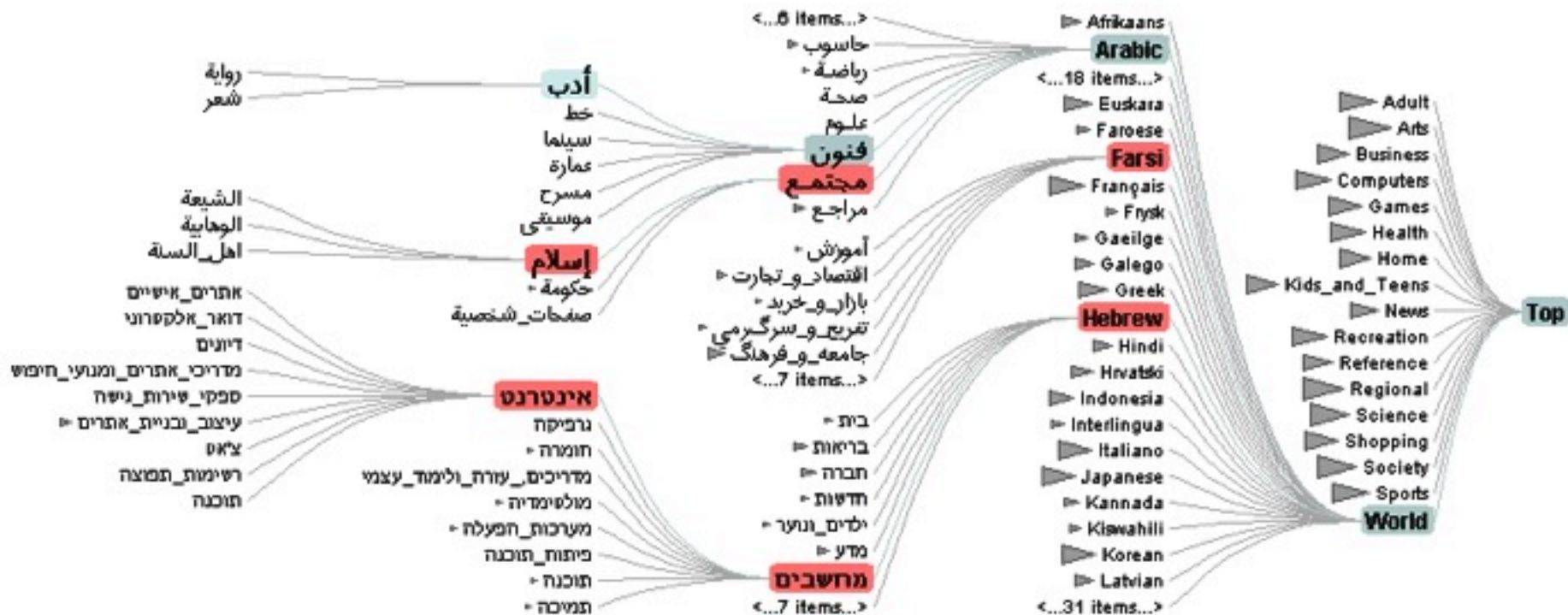
Animated transitions in statistical data graphics [Heer & Robertson 07]

# Mapping / Cartography



Dymaxion Maps [Fuller 46]

# Graphs and Trees



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

# Gizster Graphs and Trees



community &gt;&gt;



Enable

search &gt;&gt;

Want to Meet

## Zephoria

User ID: 21721  
 Friends: 286  
 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: Music  
 Books: Books  
 Authors: Erving Goffman, Stanley Milgram, Jeanette Winterson, Eric Schlosser, Leslie Feinberg, Dorothy Allison, Italo Calvino, Hermann Hesse  
 TV Shows: TV Shows  
 Movies: Movies  
 Member Since: ??  
 Last Login: 2003-10-21  
 Last Updated: 2003-10-21  
 About: 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.



# Text Visualization

Visualization / Word tree / Alberto Gonzales

Creator: Martin Wattenberg

Tags:

Search **i don't**

Back

Forward

Start

End

Occurrence Order

Clicks Will Zoom

118  
hits

**i don't**

want  
know  
believe  
think  
have



&lt;

33

&gt;

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



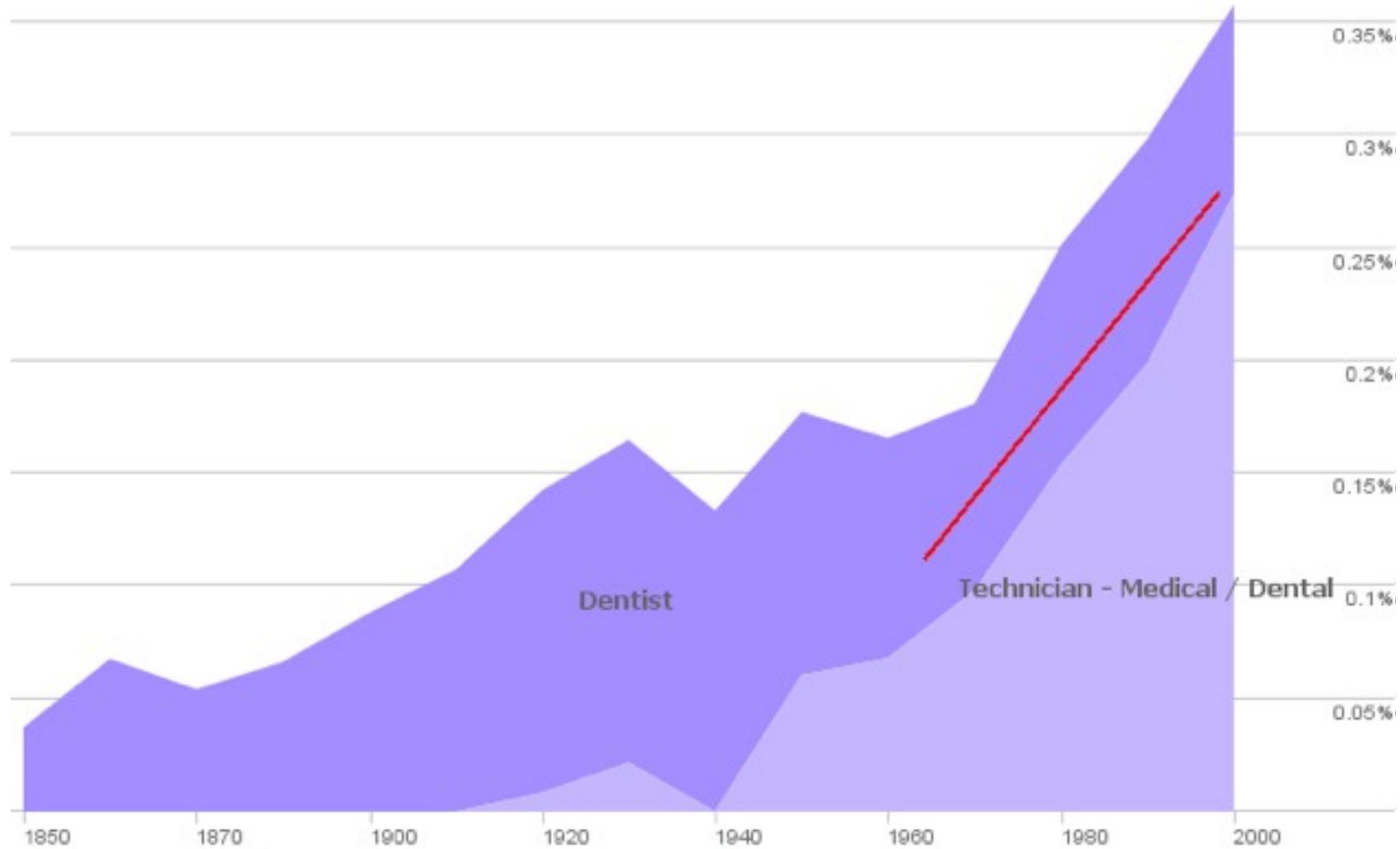
Comments (4)

currently showing

This visualization has 4 positive and 0 negative

# Collaboration and History

Where have all the dentists gone?



# Course Mechanics

# You should expect to:

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

# Instructors

cse512@cs

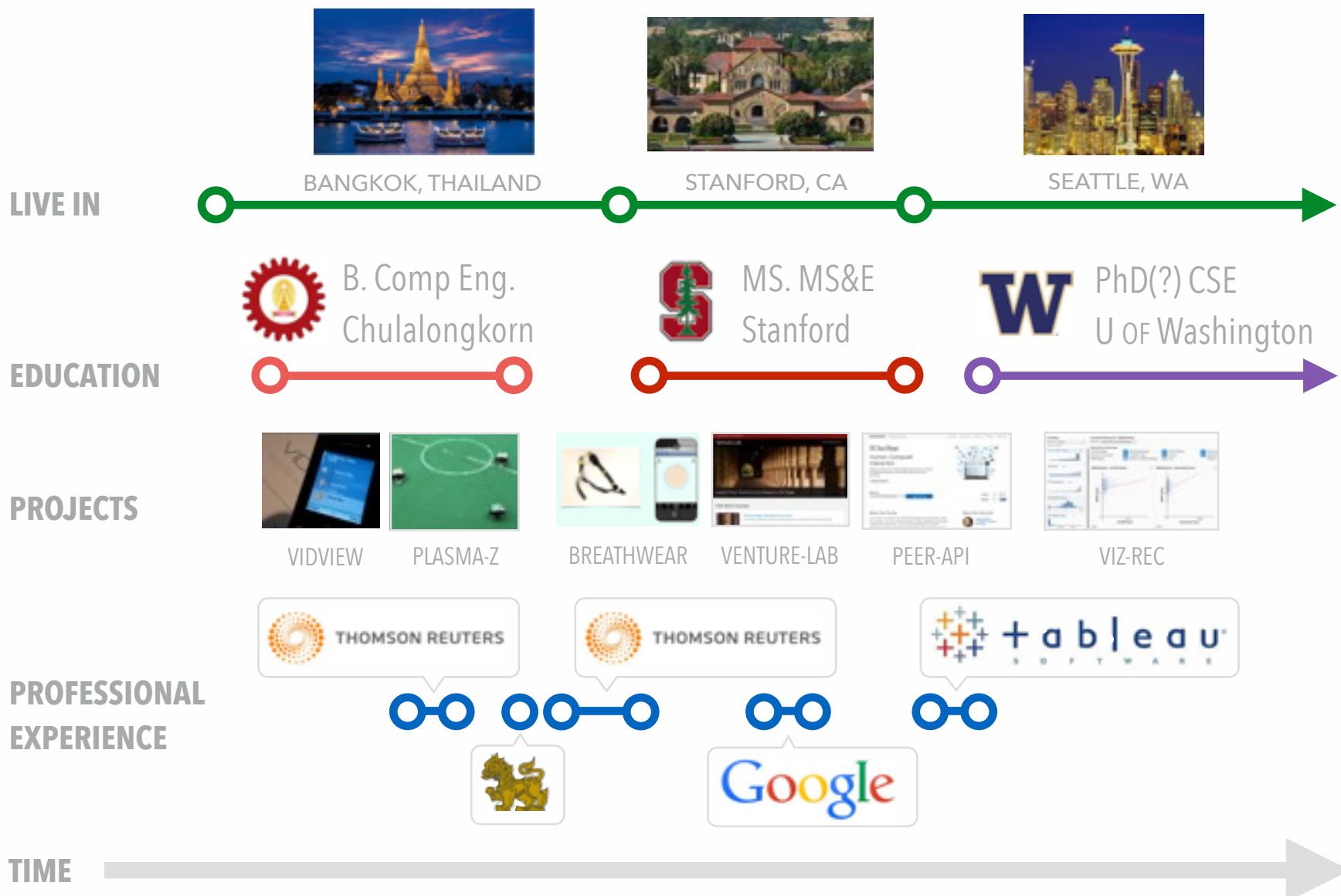
## *Instructor*

**Jeffrey Heer** OH: *Tue 9:30-10:30, 642 CSE*  
Assoc Prof, CSE <http://jheer.org>

## *Teaching Assistant*

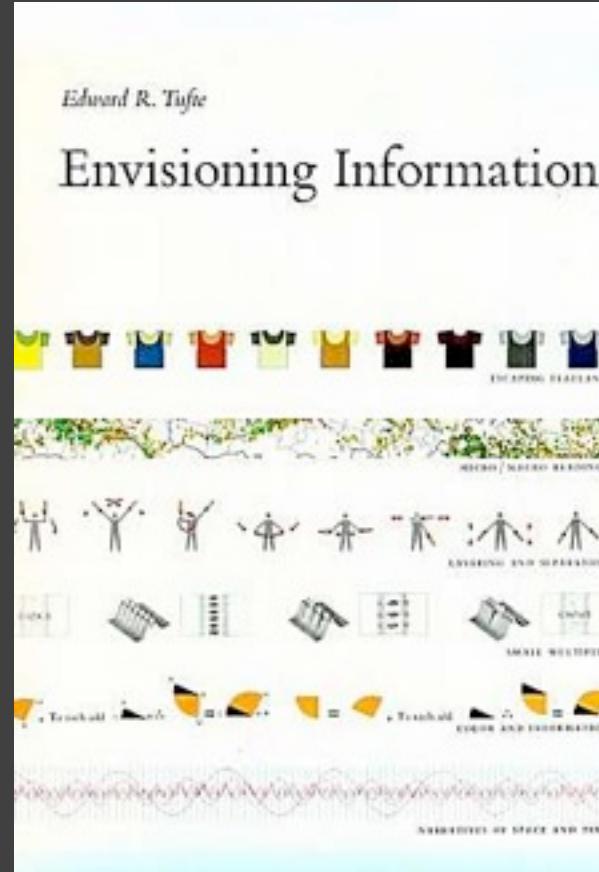
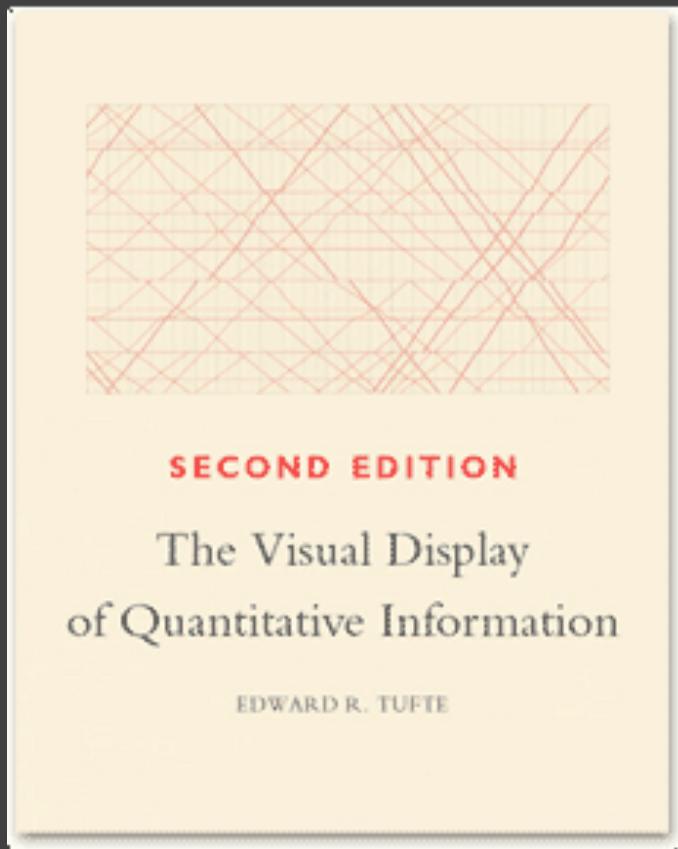
**Kanit “Ham” Wongsuphasawat** OH: *Thu 9:45-10:45, 218 CSE*  
<http://kanitw.yellowpigz.com>

# Kanit "Ham" Wongsuphasawat @kanitw



See my portfolio at [kanitw.yellowpigz.com](http://kanitw.yellowpigz.com) or [bit.ly/knowham](http://bit.ly/knowham)

# Textbooks



See also: [www.edwardtufte.com](http://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 Piazza forum

Comments must be posted within **1 day** of lecture

You have 2 “passes” for the quarter

**Complete enrollment form on the website!**

We will then invite you to the Piazza forum.

# Requirements

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 your choice

Project write-up in form of a short research paper

Two project presentations

1. Initial in-class status report (2/27)
2. Final poster presentation (tentatively 3/13)

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



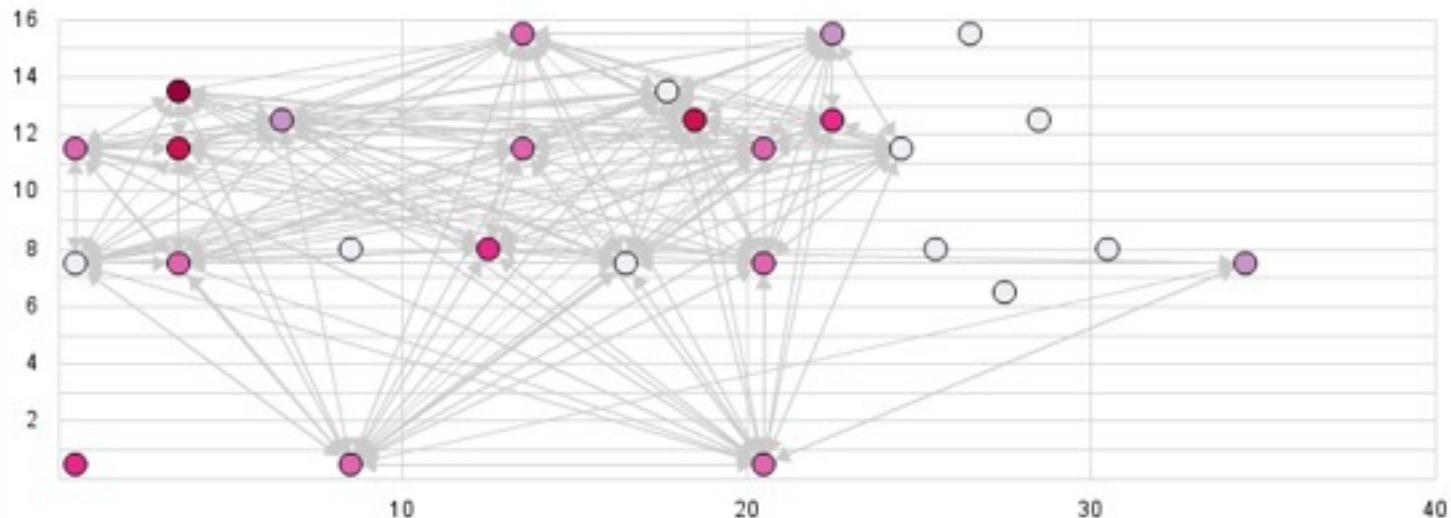
## Node Noisefloor

- low
- high

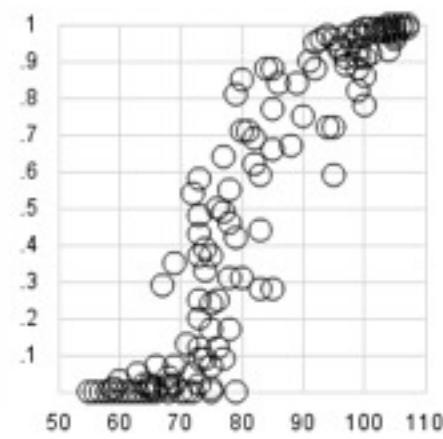
Click on a node to select its edges.

Filter edges by:

- incoming
- outgoing
- both

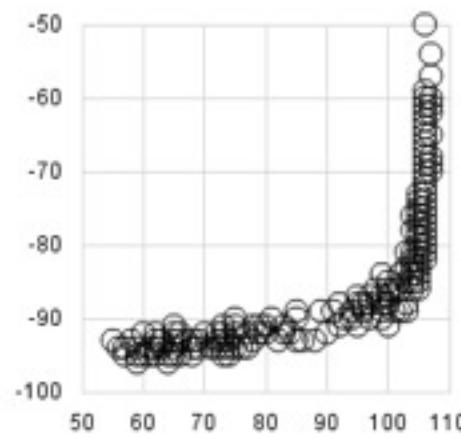


PRR vs CCI



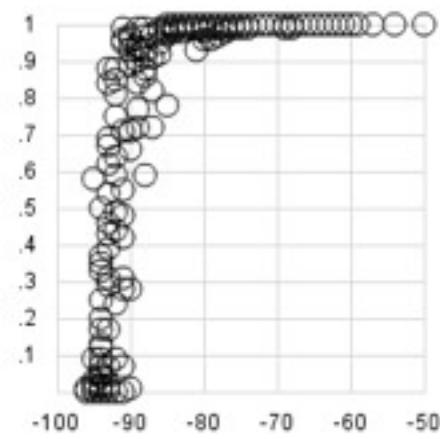
zoom on/off

RSSI vs CCI



zoom on/off

PRR vs RSSI

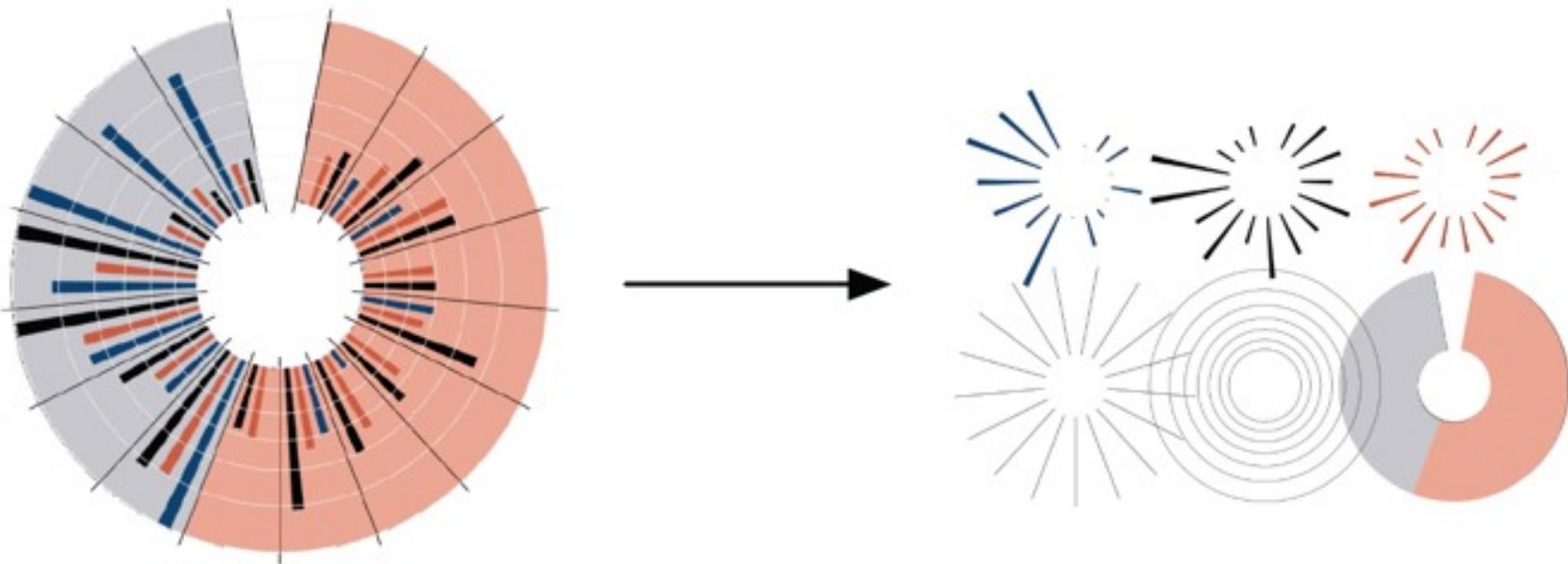


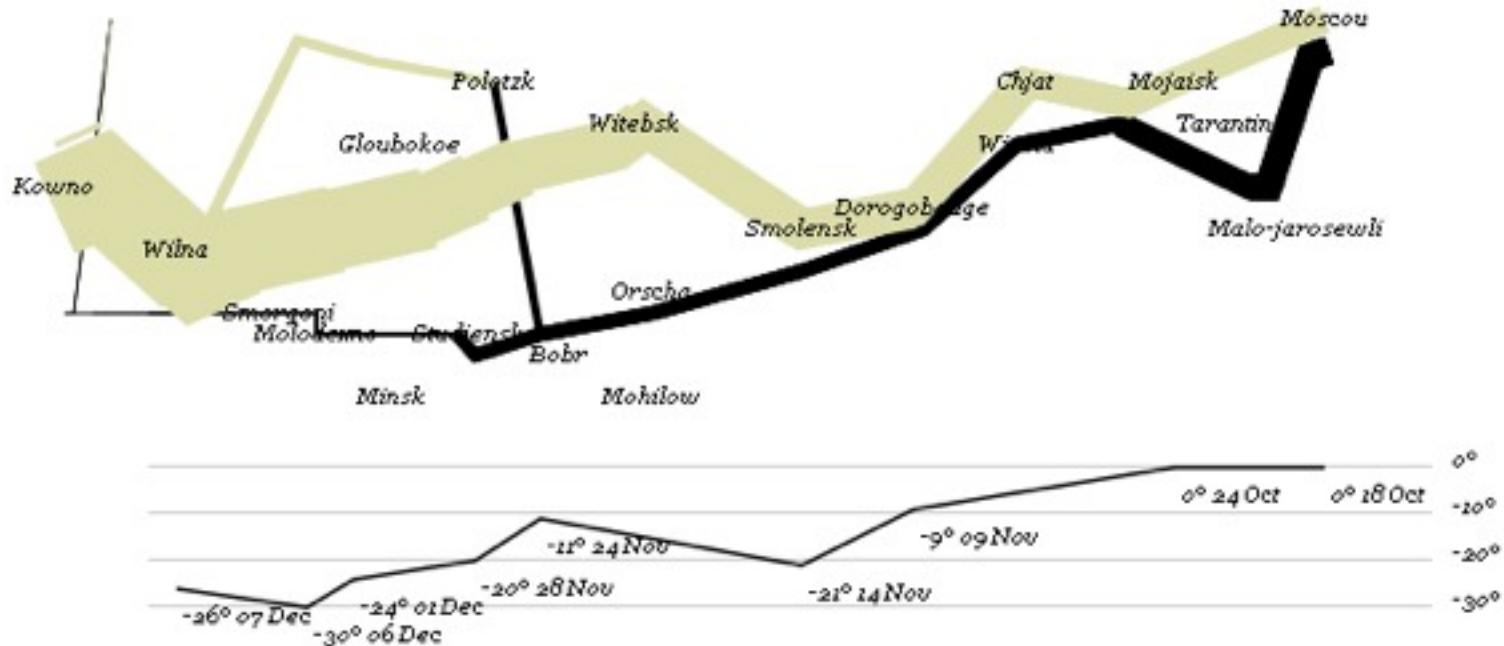
zoom on/off

Stanford Network Analysis Tool - Nick Briggs & Maria Kazandjieva

# 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");

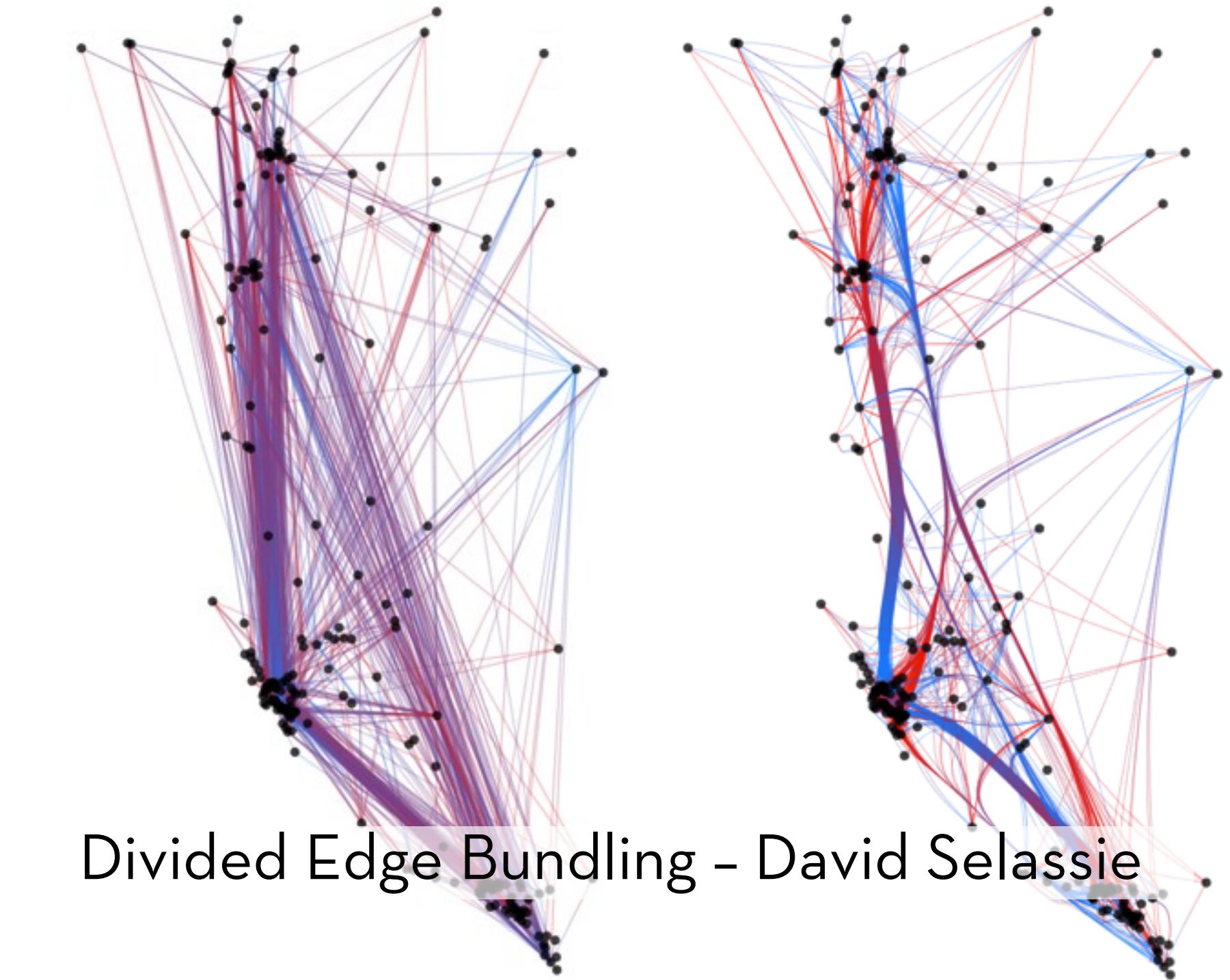
```

```

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");

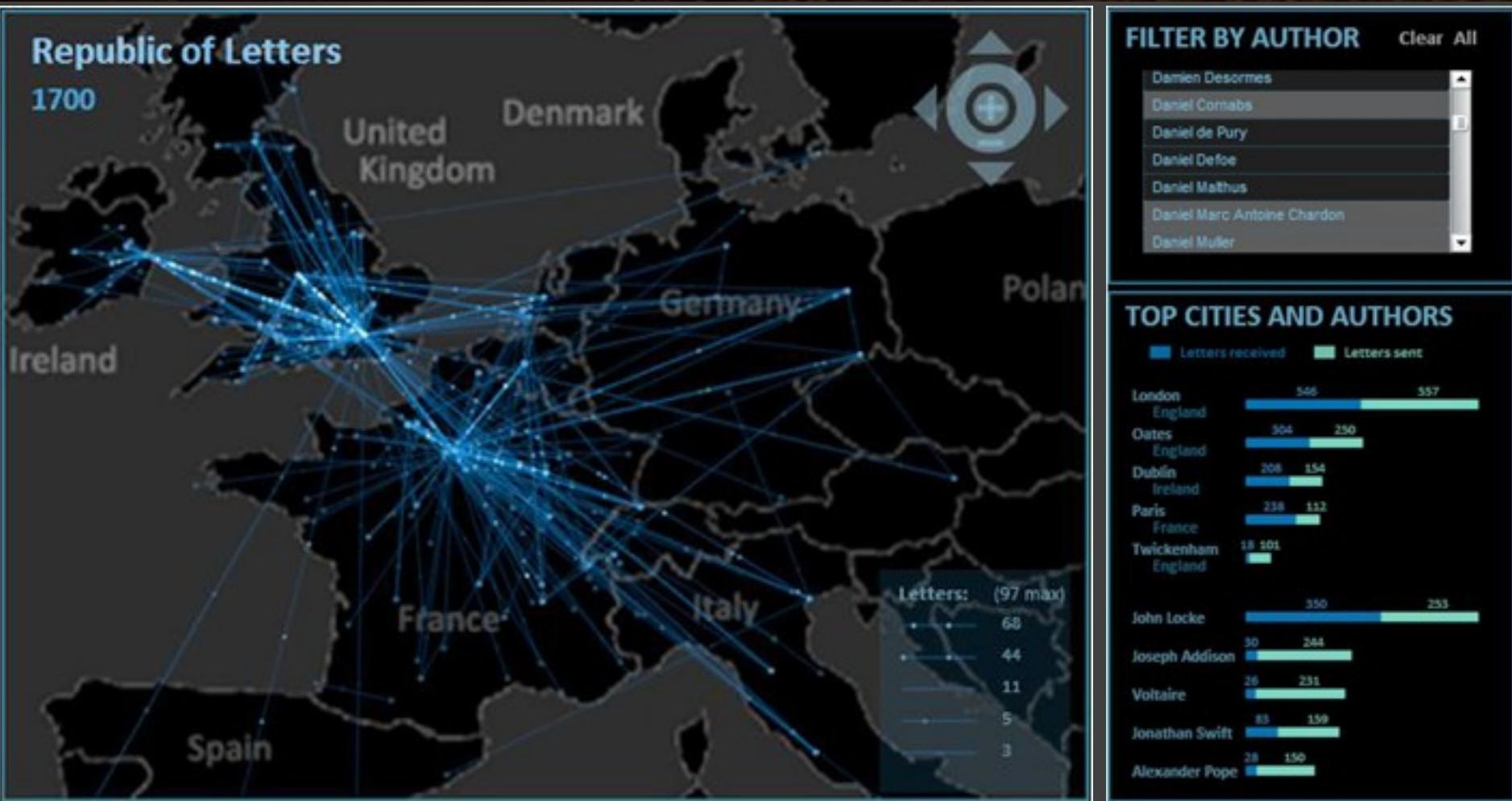
vis.add(pv.Line).data(napoleon.temp)
  .left(lon).top(tmp).strokeStyle("#0")
  .add(pv.Label)
  .top(function(d) 5 + tmp(d))
  .text(function(d) d.temp+"° "+d.date.substr(0,6))
  .textBaseline("top").font("italic 10px Georgia");

```



# Visualizing the Republic of Letters

Daniel Chang, Yuankai Ge, Shiwei Song



# 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 Catalyst; see A1 webpage)

- Image of your visualization (e.g., PNG, GIF, JPG)
- Short description and design rationale ( $\leq$  4 paragraphs)

Due by 5:00 pm, **Monday Jan 13.**