

---

# DATA VISUALIZATION

Lecturer: Yang Yue, Min Lu  
2019 Autumn



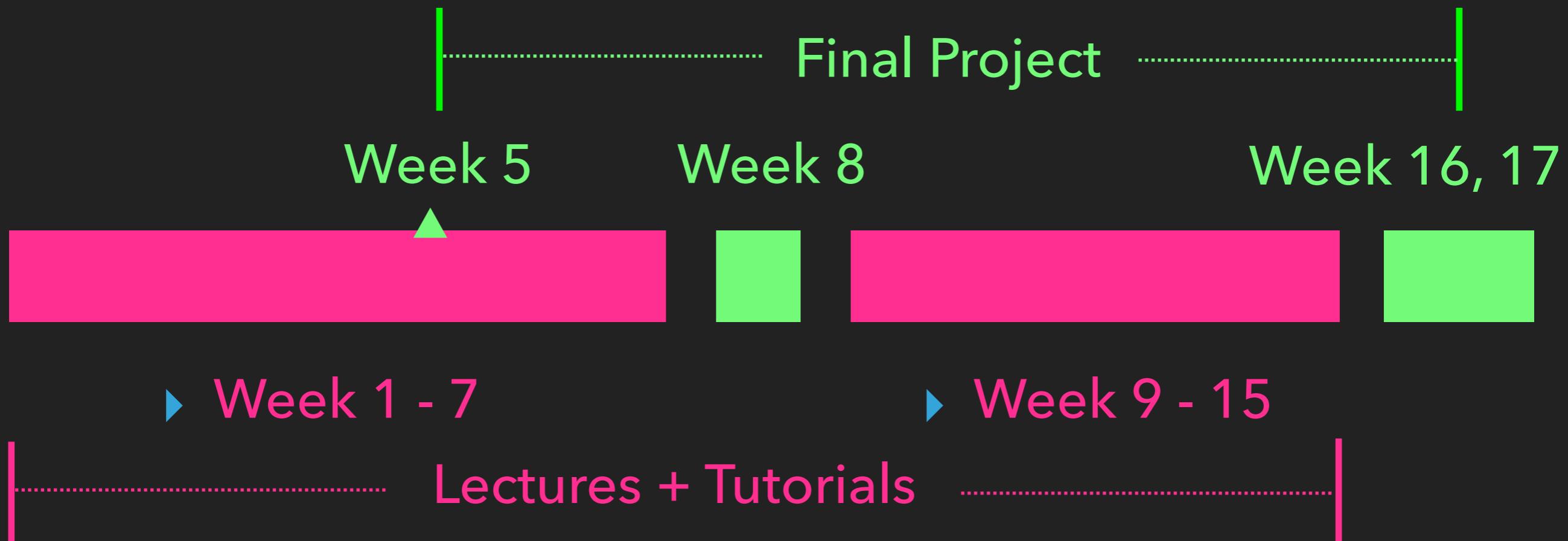
# Course Overview

# Keywords

Fun!

Practice : )

# Agenda



# Assignments

40% + 60%

# Assignments (cont.)

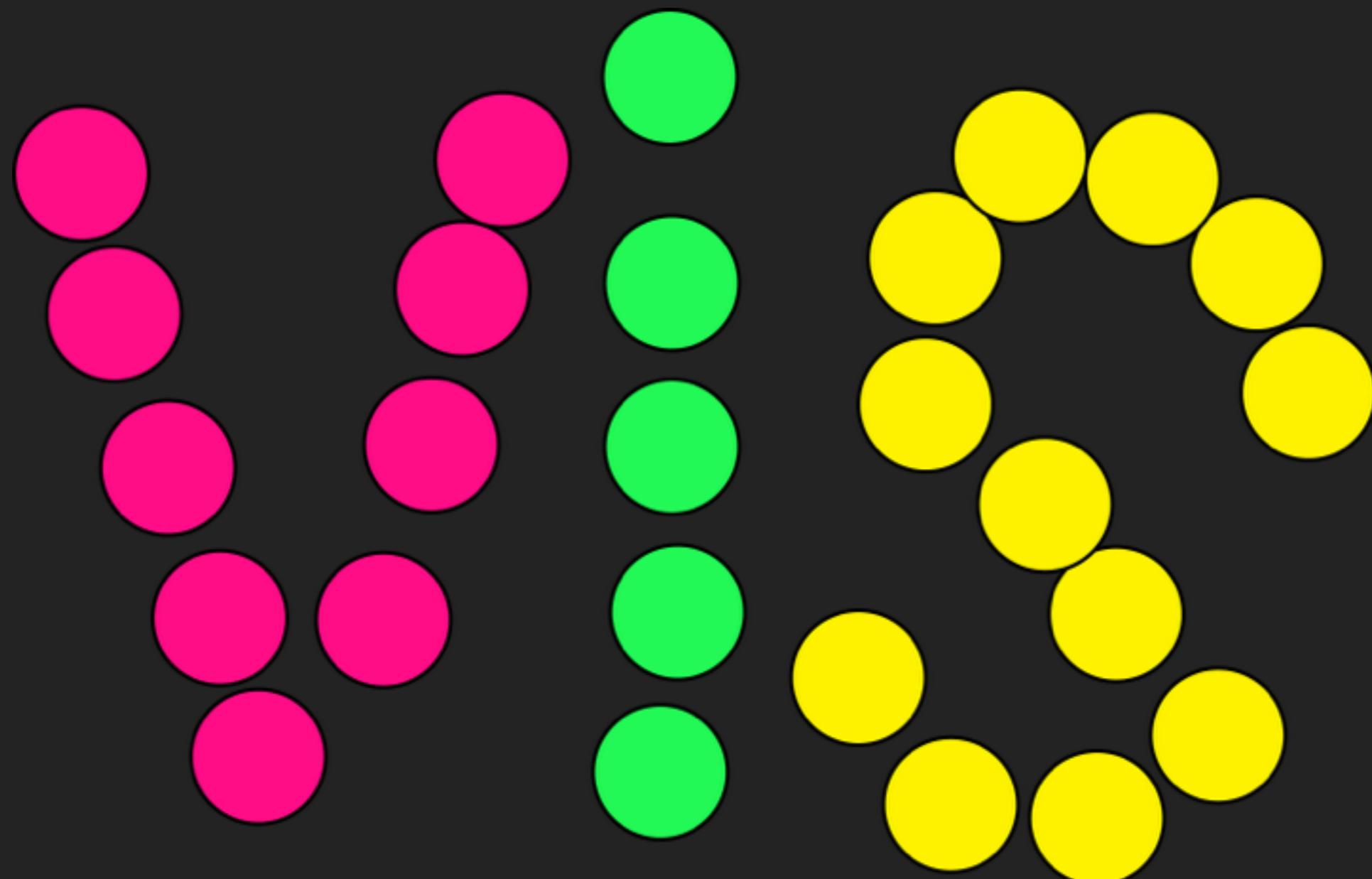
- ▶ Class Participation (10%)
- ▶ A1: Good/Bad Visualization Evaluation (10%) - Due Sept. 27
- ▶ A2: Interactive Prototype (20%) - Due Oct. 11
- ▶ FP: Final Project (60%) - Due Dec. 12
  - ▶ Design - Due Oct. 28
  - ▶ Implementation - Due Dec. 12

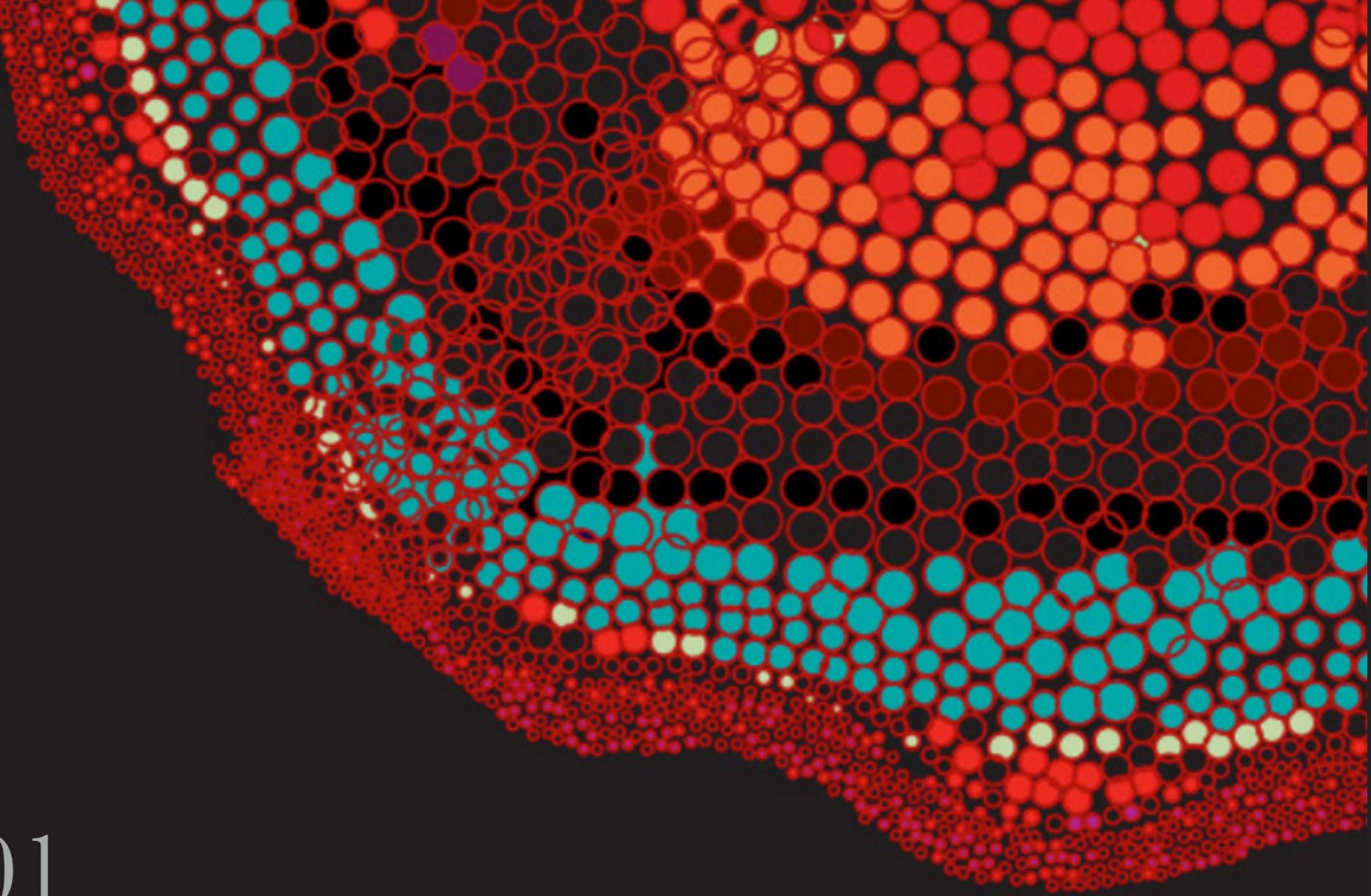
# Who Am I



- ▶ 陆旻 Min Lu, Ph.D, visualization, visual analytics
- ▶ Homepage: <https://deardeer.github.io/>
- ▶ Email: [lumin.vis@gmail.com](mailto:lumin.vis@gmail.com)
- ▶ Office: ScienceTec-building 1402

# Who Are You





---

LECTURE 01

---

# The Value Of Visualization

How Much Data (Bytes)  
Do We Create Everyday?

$10^{18}$

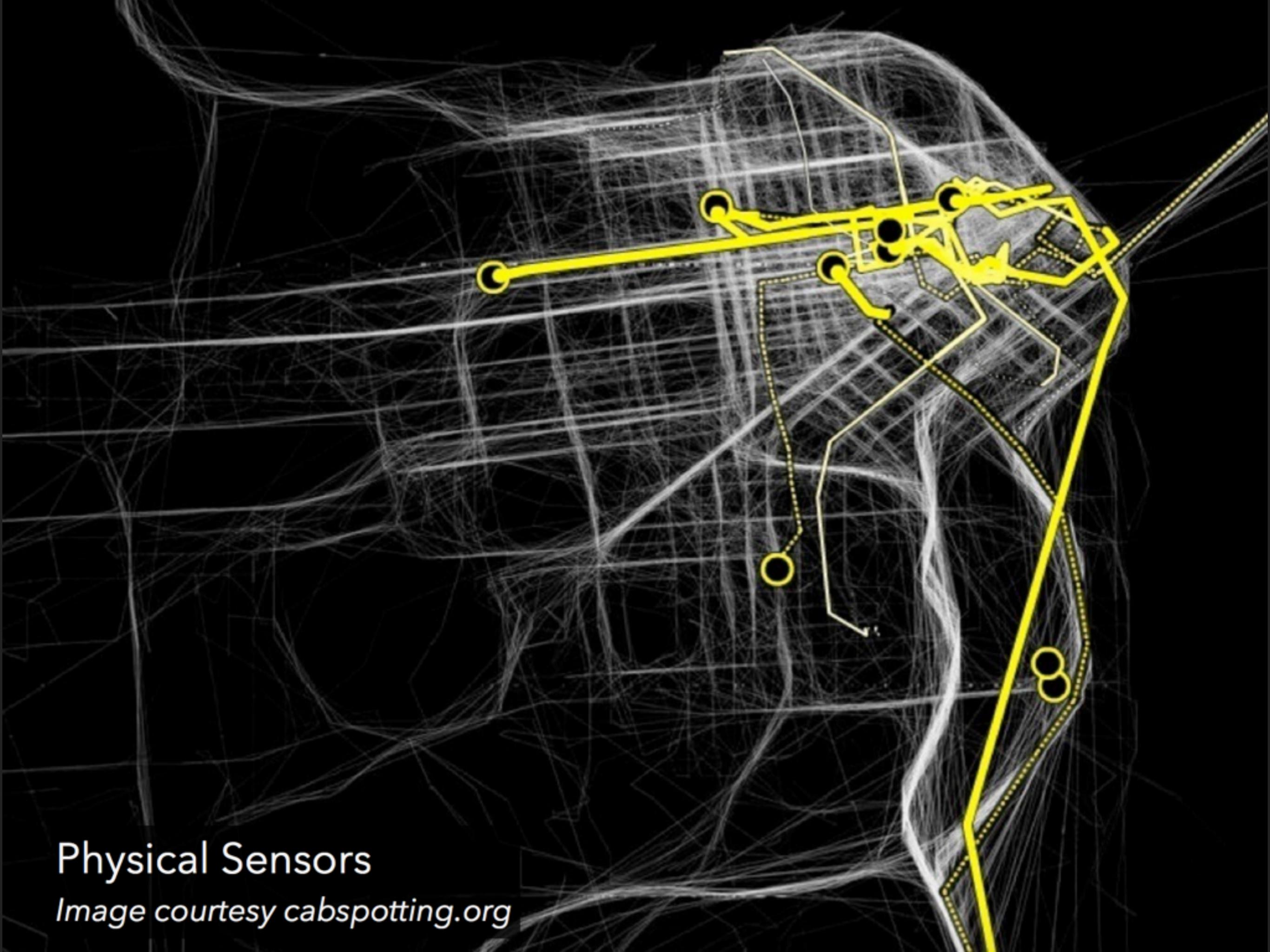
REFERENCE: <https://www.forbes.com/sites/bernardmarr/2018/05/21/how-much-data-do-we-create-every-day-the-mind-blowing-stats-everyone-should-read/#1836349e60ba>

$$1\text{G} = 10^9$$

$$10^{18} = 10^9 * 1\text{G}$$

INFORMATION TECHNOLOGY  
ADVERTISING  
SOCIAL MEDIA  
SERVICES  
WORK  
INTERNET  
BRANDS  
SOLUTIONS  
SOLOS  
B2B TARGET  
PLANNING  
MEDIA  
BEHAVIOUR  
BUZZ  
PROJECTS  
CONTENT  
TARGET  
CONSUMER  
ORGANIZATION  
EVENTS  
PROGRAMMING  
BIG DATA  
WEB SERVICES  
MOBILE  
INFORMATION  
DIGITAL  
PROMOTION  
PLAN  
COMMUNICATION  
BRANDING  
CONSUMER DEMAND  
WEB MARKETING  
SOCIAL STATISTICS  
APPSS  
VISION  
ENGINEERING  
WEB DEV  
STRATEGY  
WORLDWIDE  
ORGANIZATION  
PRICING  
SEGMENTATION  
SOCIAL NETWORKS  
INFORMATION  
DIGITAL  
E-MARKETING  
COMPUTER  
DEMAND  
MARKETING  
INTERNET  
DATA  
SERVICES  
PROJECTS  
INTERNET  
SOCIAL  
STATISTICS  
APPSS  
VISION  
ENGINEERING  
WEB DEV  
STRATEGY  
WORLDWIDE  
ORGANIZATION  
PRICING  
SEGMENTATION  
SOCIAL NETWORKS  
INFORMATION  
DIGITAL  
E-MARKETING  
COMPUTER  
DEMAND  
MARKETING

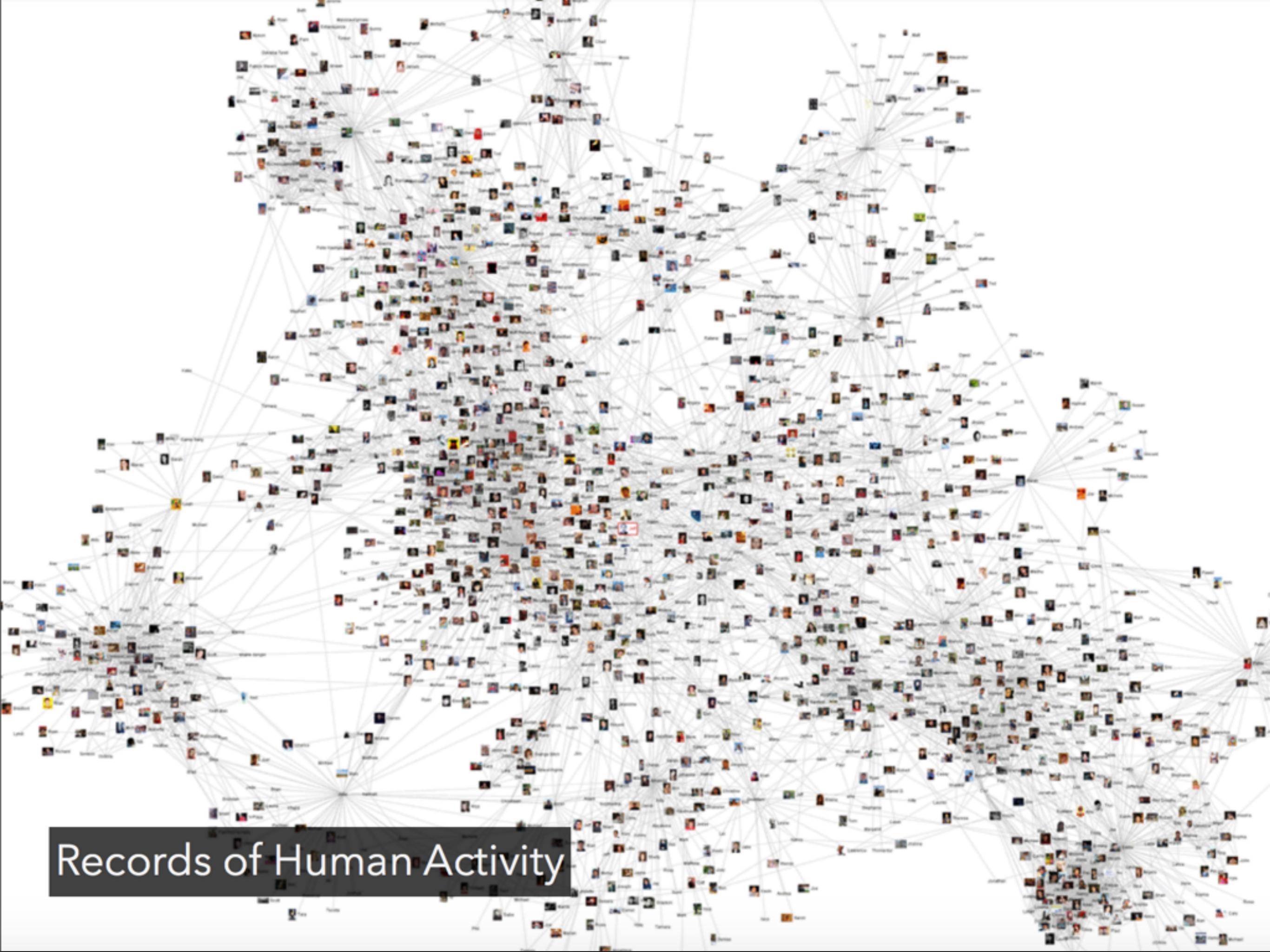
# BIG DATA



## Physical Sensors

*Image courtesy cabspotting.org*





Records of Human Activity

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 complementary 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**

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

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

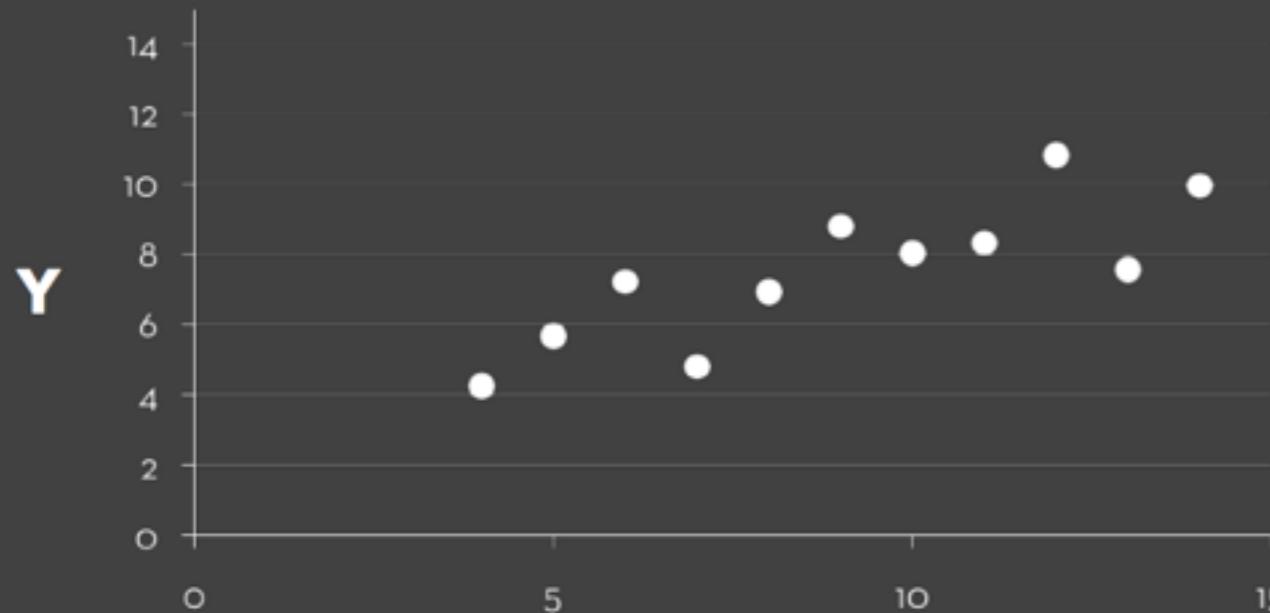
**Linear Regression**

$$Y = 3 + 0.5X$$

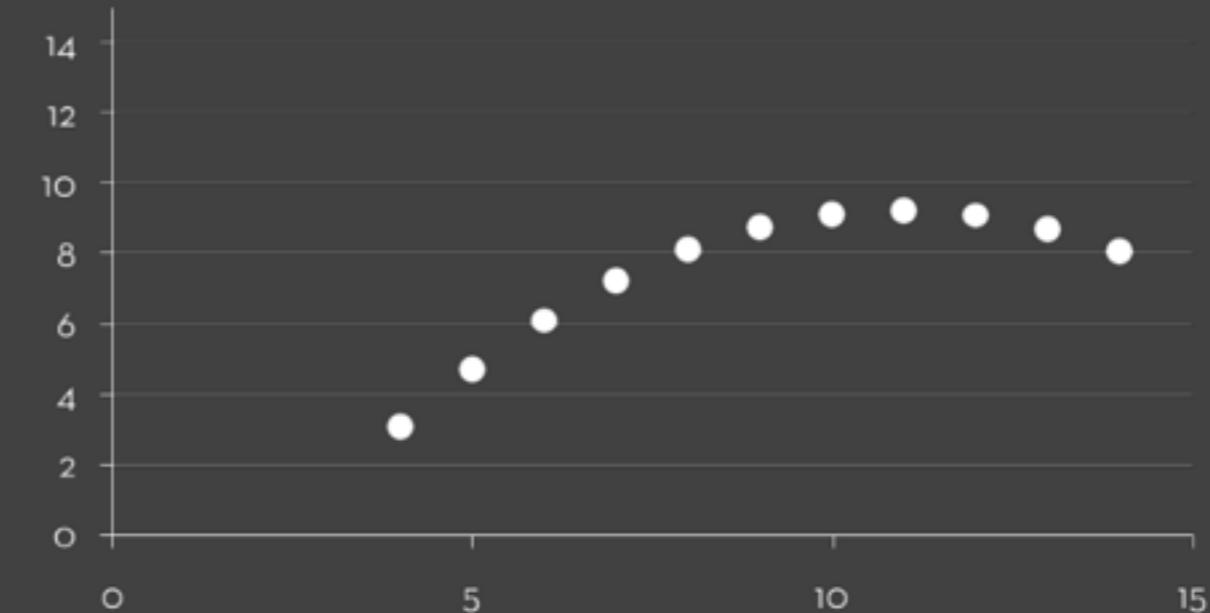
$$R^2 = 0.67$$

[Anscombe 1973]

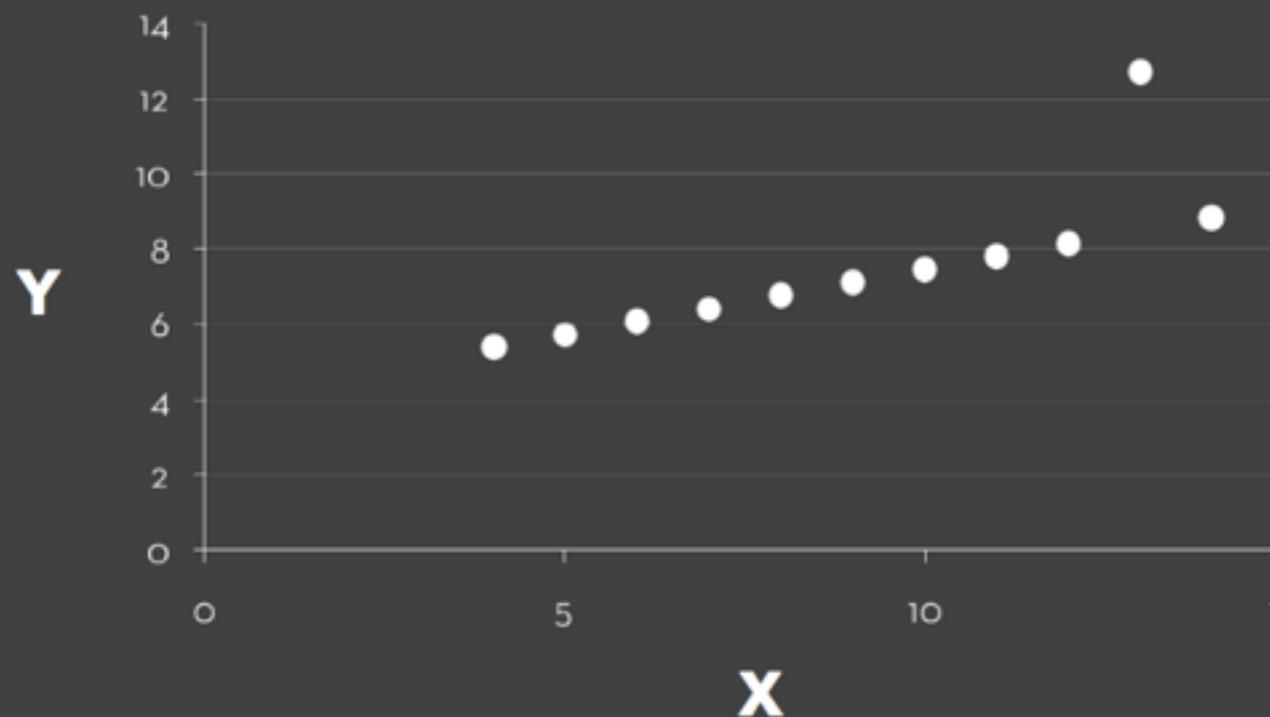
Set A



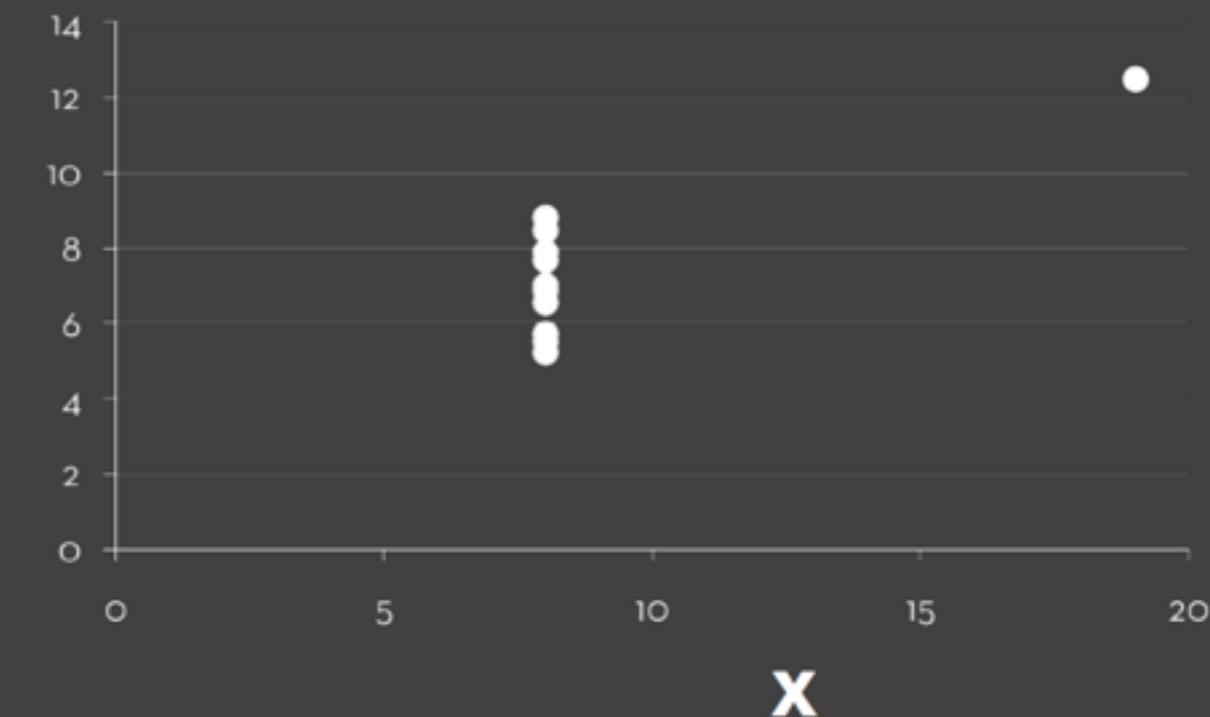
Set B



Set C



Set D



# Why Create Visualizations?

Answer questions (or discover them)

Make decisions

See data in context

Expand memory

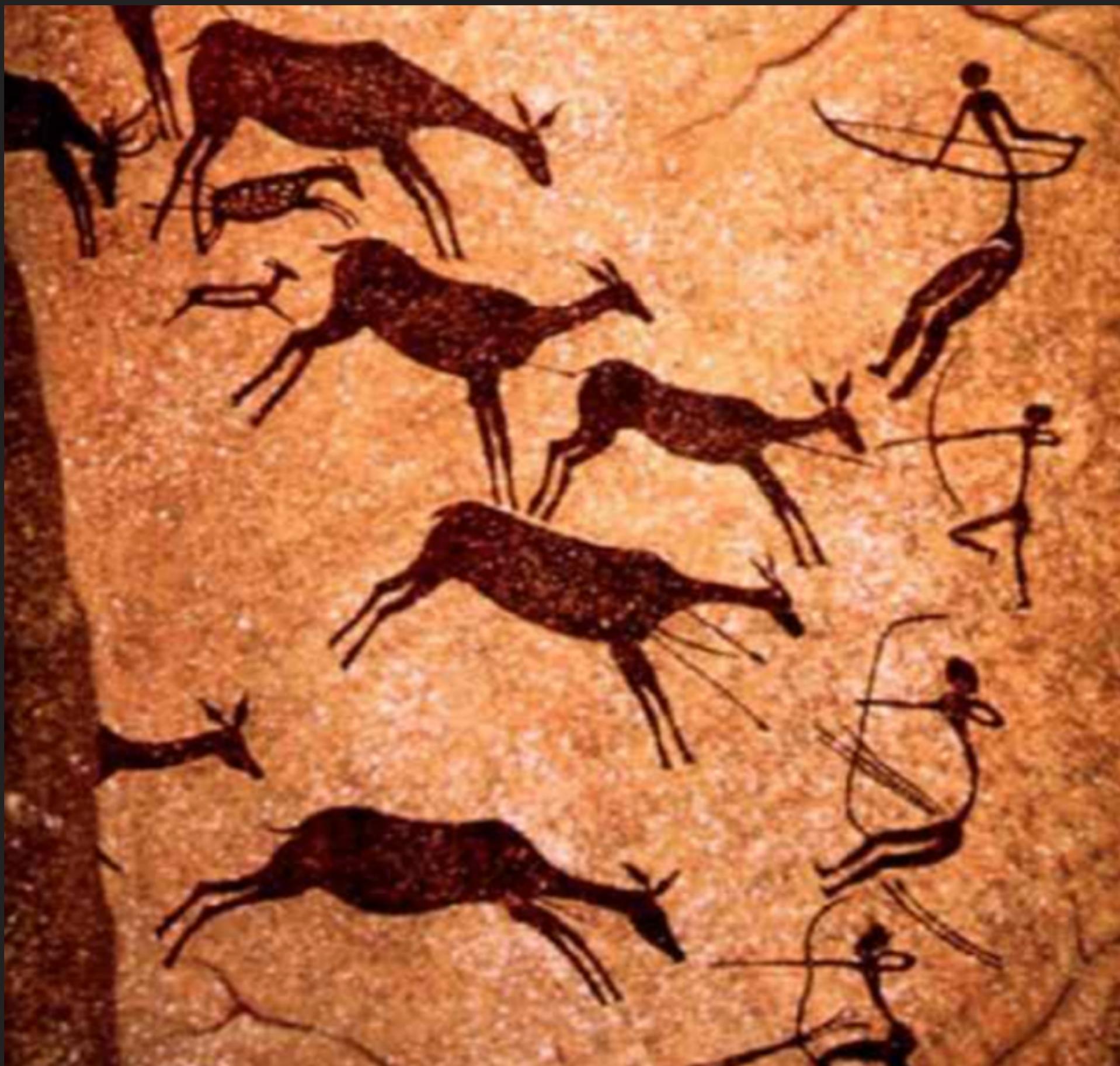
Support graphical calculation

Find patterns

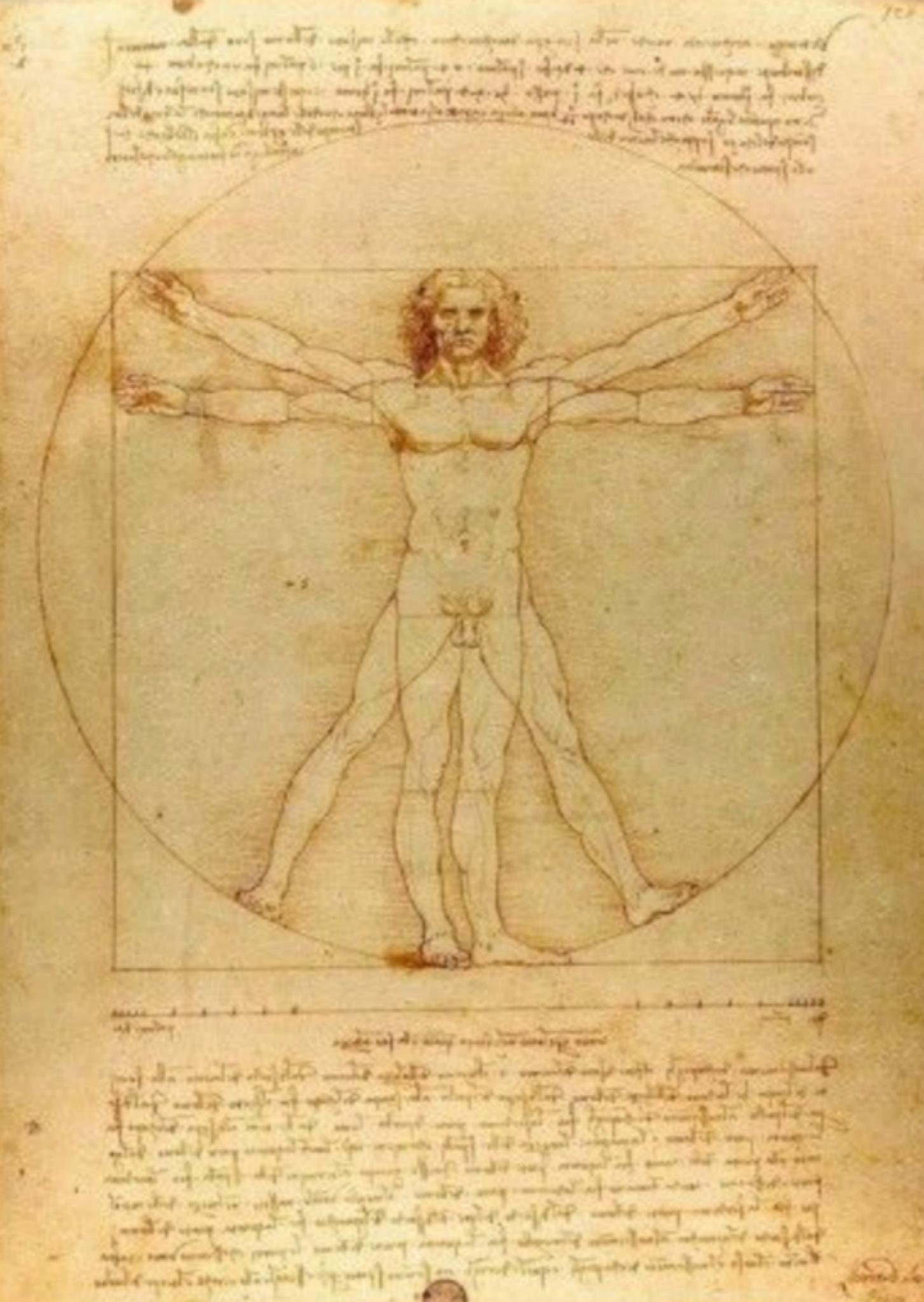
Present argument or tell a story

Inspire

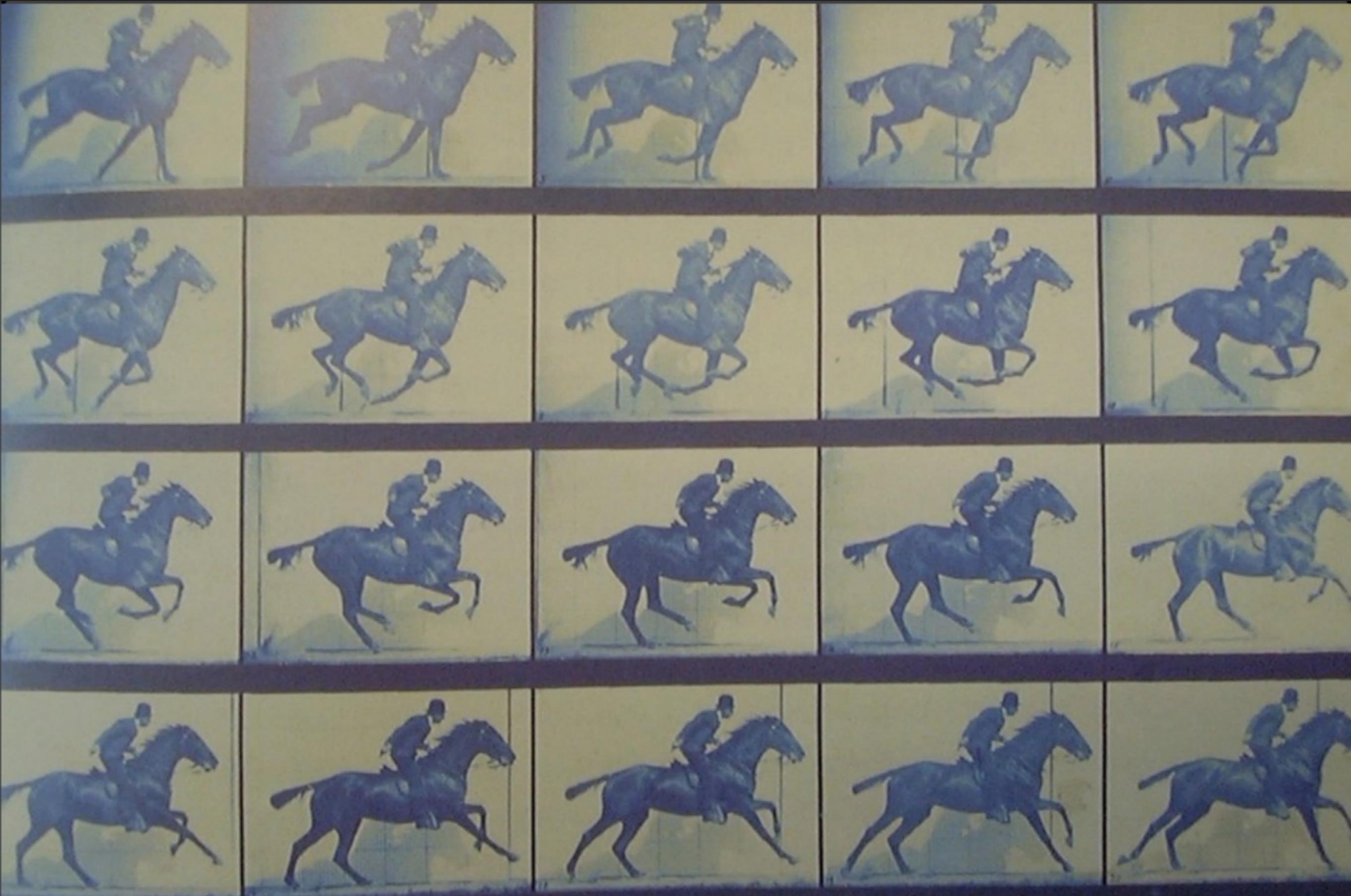
# Record Information



Spanish Los Caballos

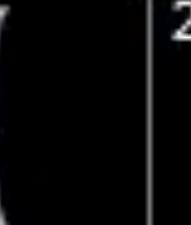
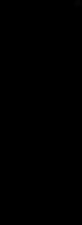


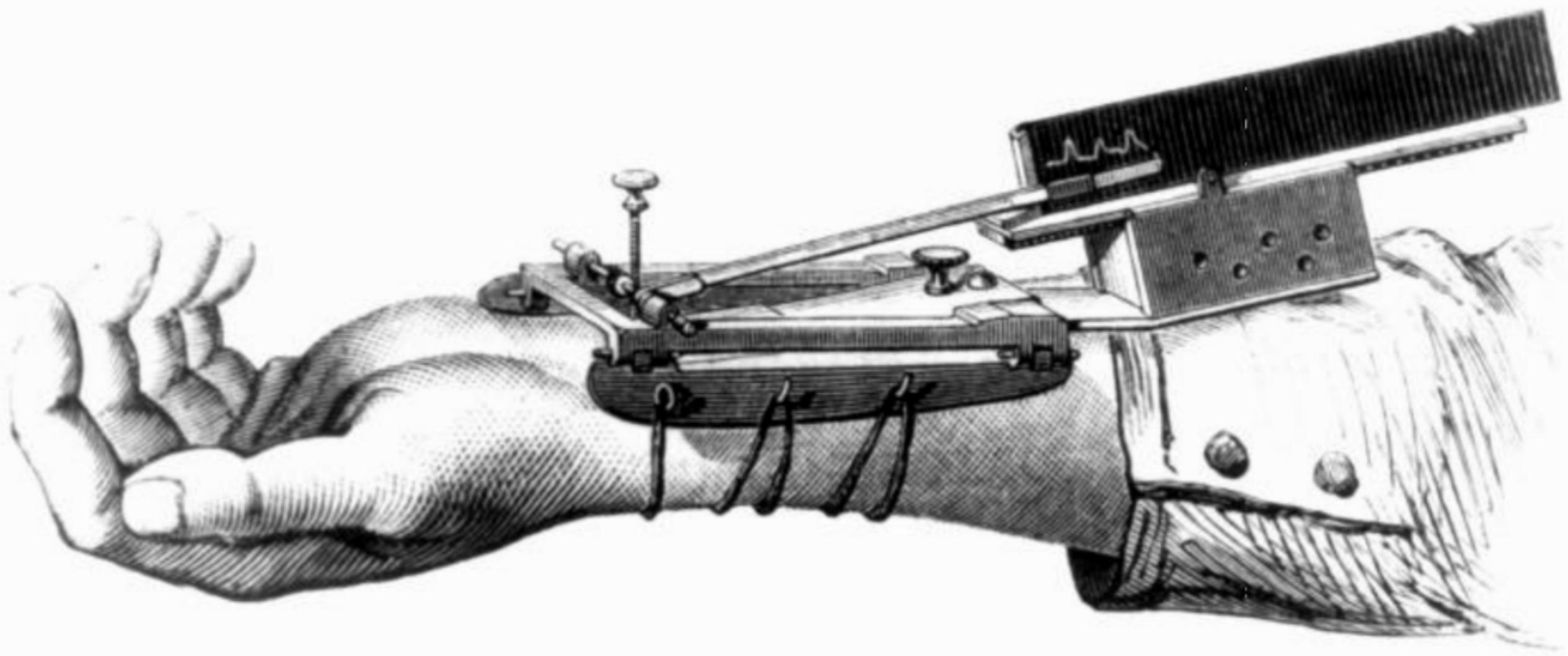
The Vitruvian Man, by  
Leonardo da Vinci, 1487



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

**May 2011**

Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
1 	2 	3 NM 	4 	5 	6 	7 
8 	9 	10 FQ 	11 	12 	13 	14 
15 	16 	17 FM 	18 	19 	20 	21 
22 	23 	24 LQ 	25 	26 	27 	28 
29 	30 	31 				



1.

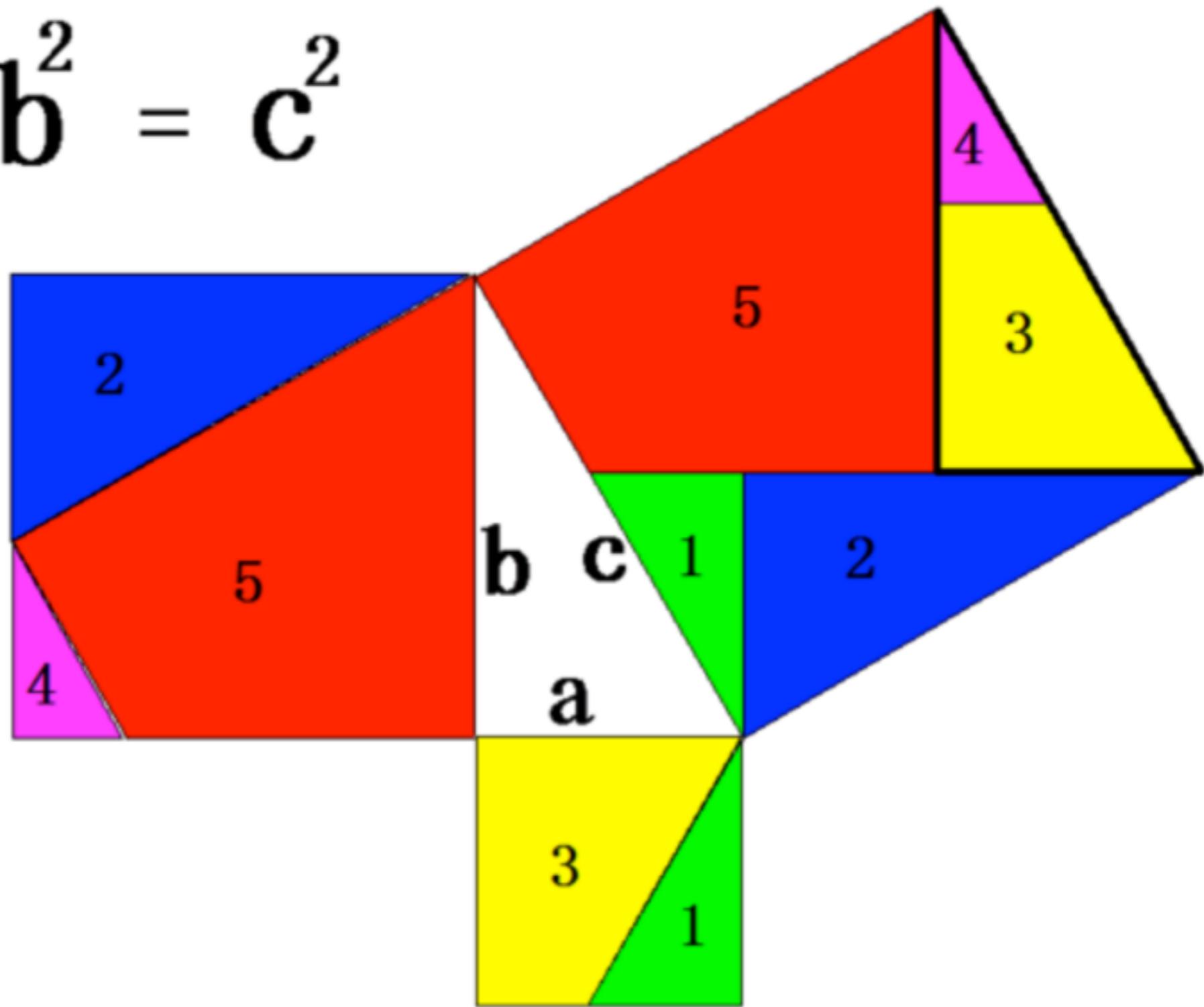
Marey's sphygmograph in use.

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

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

# Support Reasoning

$$a^2 + b^2 = c^2$$





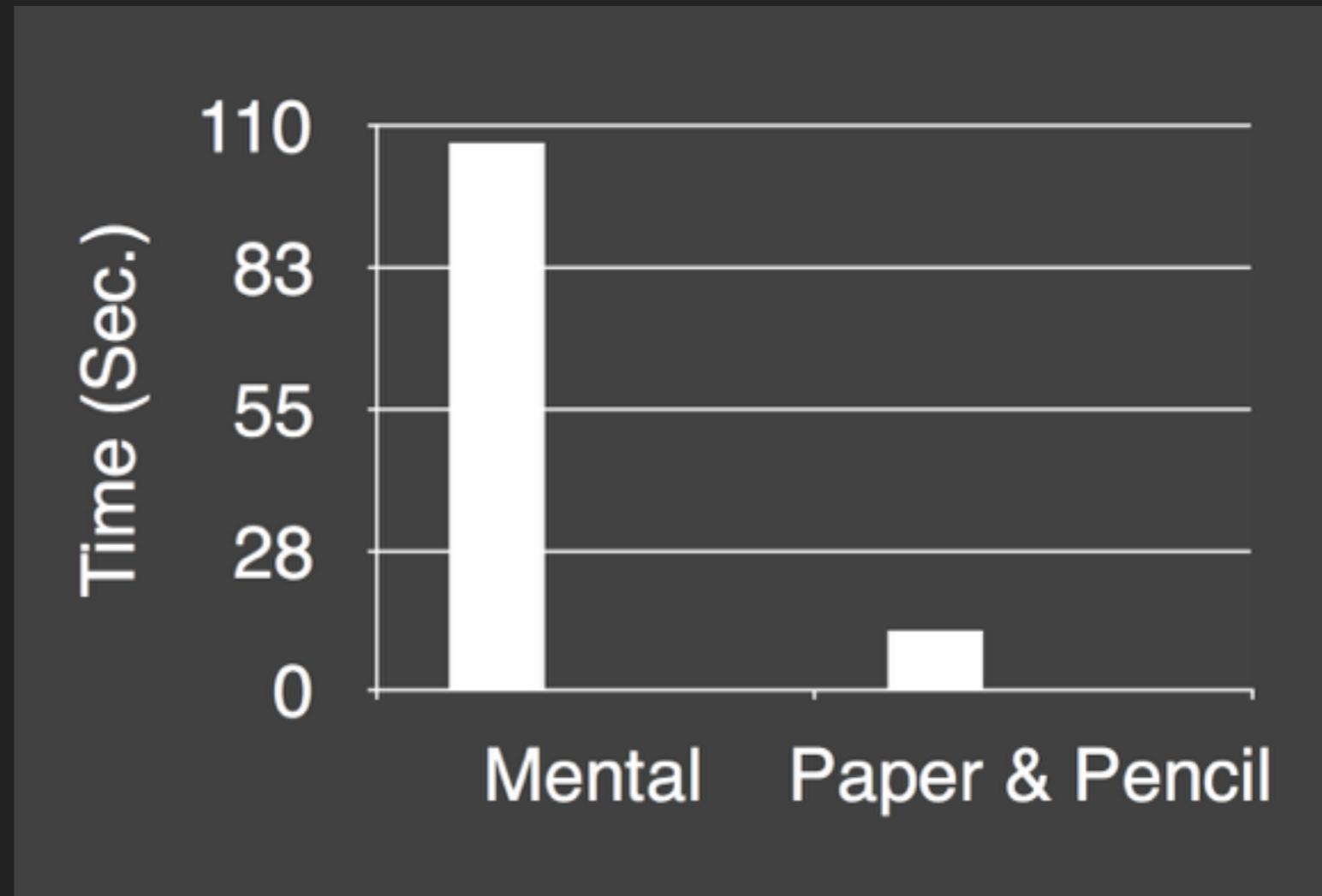
In 1854 John Snow plotted the position of each cholera case on a map. [from Tufte 83]

# Expand Memory

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

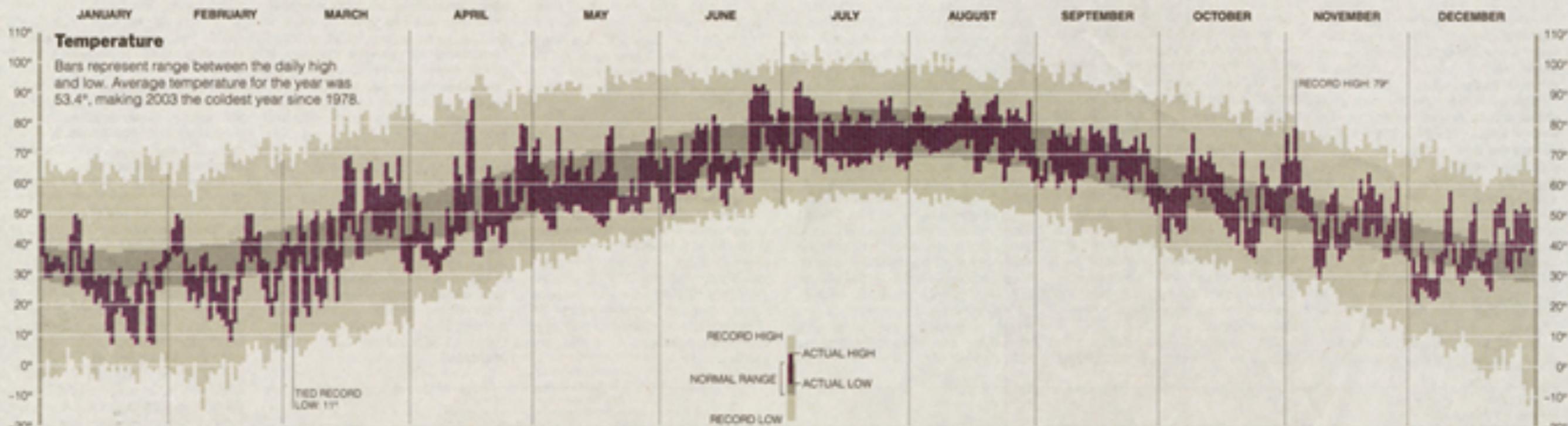
?

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

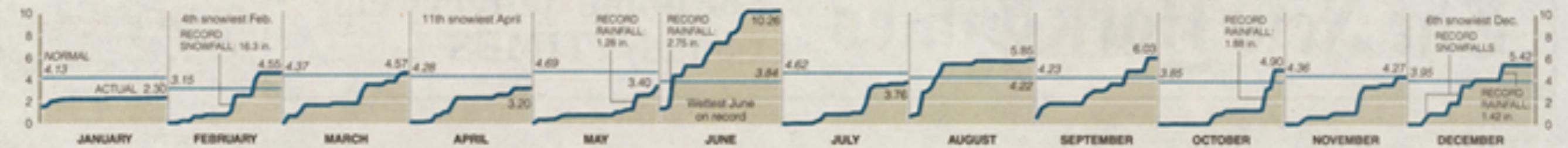


# Find Patterns

## New York City's Weather in 2003



**Precipitation** Cumulative monthly precipitation in inches compared with normal monthly precipitation. Total precipitation in 2003 was 58.51 inches, 8.82 inches more than normal, which makes the year the sixth wettest on record.



Microsoft Excel - animal.xls

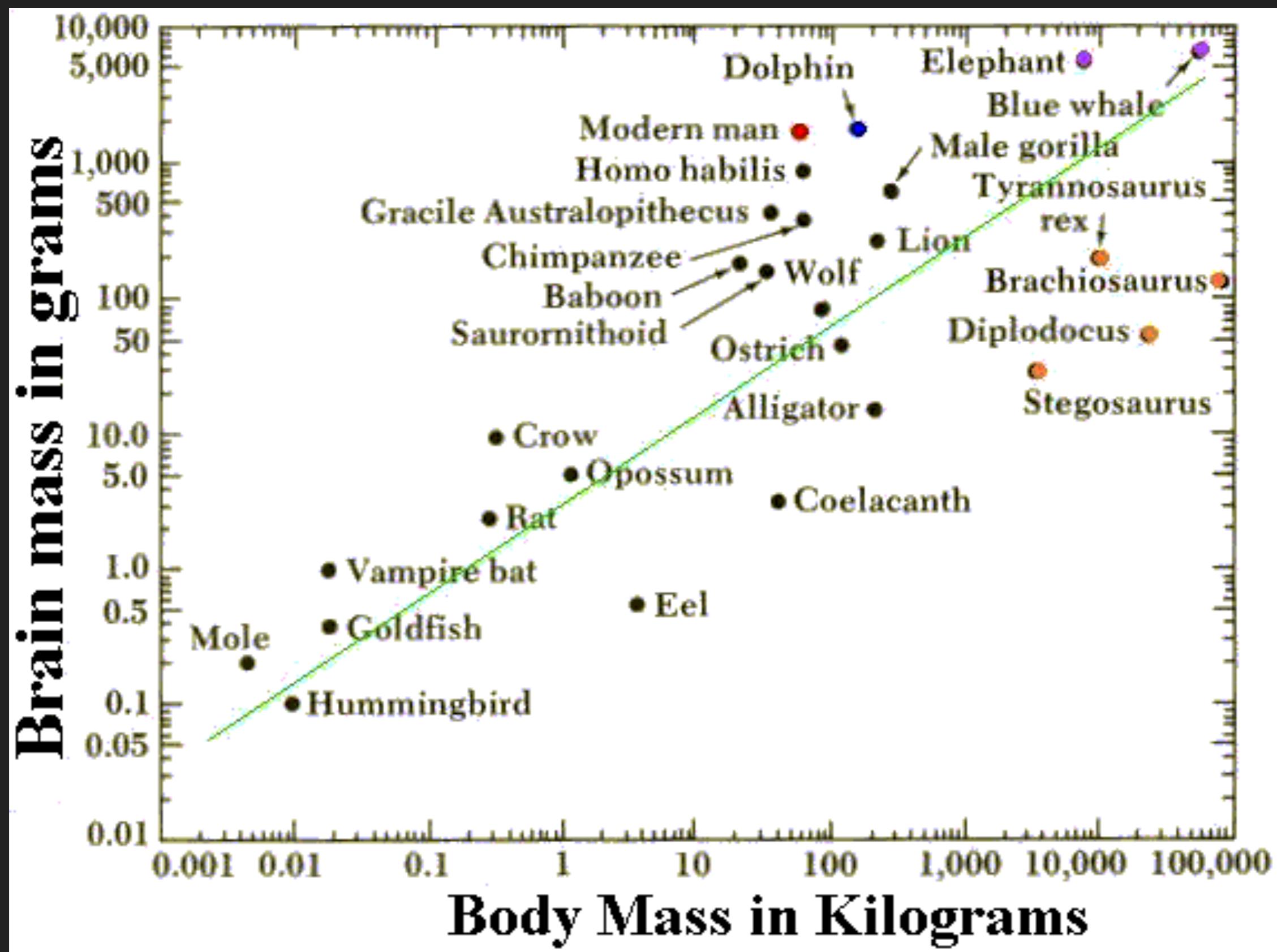
A1 ID

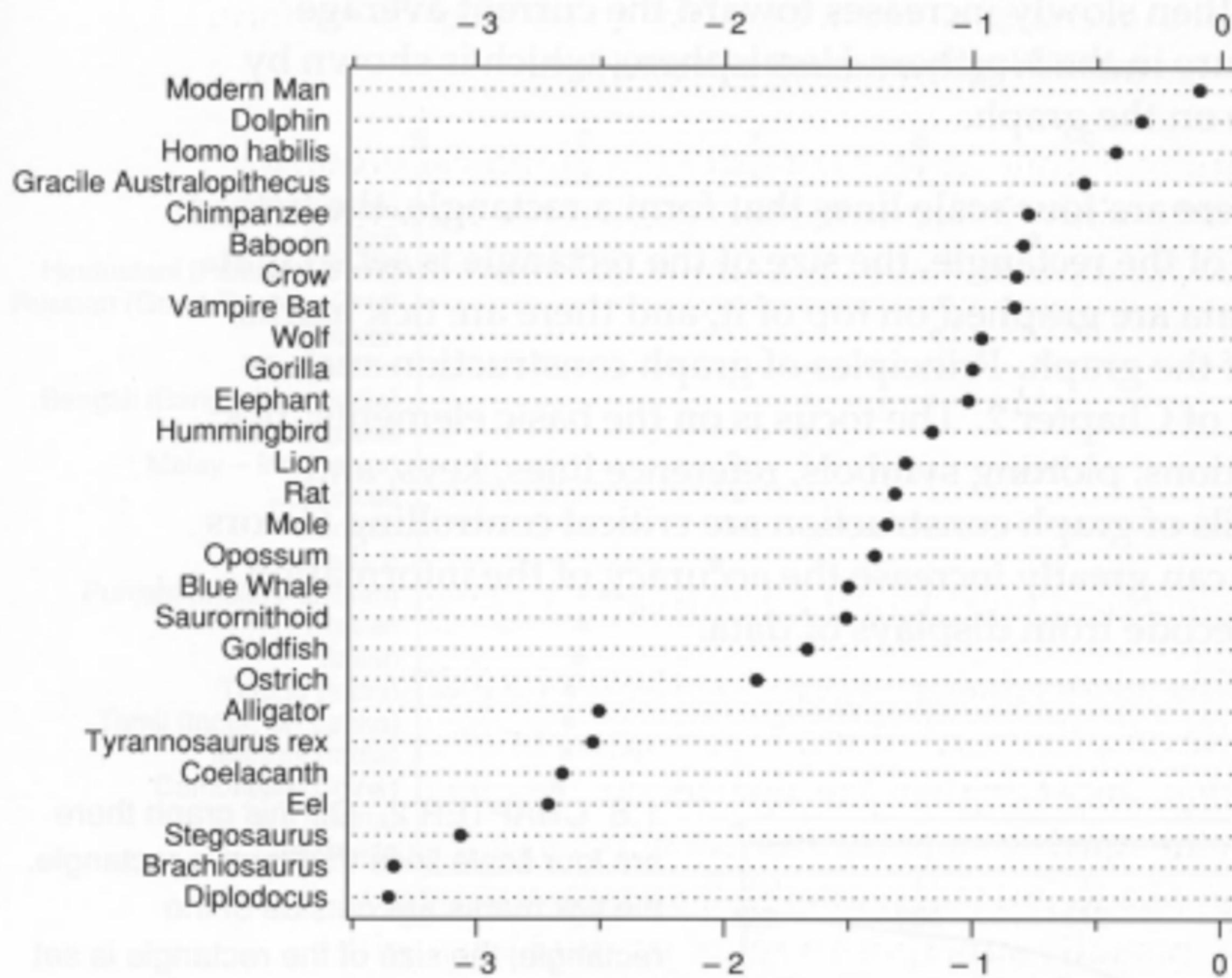
	A	B	C	D	E
1	ID	Name	Body Weight	Brain Weight	
2	1	Lesser Short-tailed Shrew	5	0.14	
3	2	Little Brown Bat	10	0.25	
4	3	Mouse	23	0.3	
5	4	Big Brown Bat	23	0.4	
6	5	Musk Shrew	48	0.33	
7	6	Star Nosed Mole	60	1	
8	7	Eastern American Mole	75	1.2	
9	8	Ground Squirrel	101	4	
10	9	Tree Shrew	104	2.5	
11	10	Golden Hamster	120	1	
12	11	Mole Rat	122	3	
13	12	Galago	200	5	
14	13	Rat	280	1.9	
15	14	Chinchilla	425	6.4	
16	15	Desert Hedgehog	550	2.4	
17	16	Rock Hyrax (a)	750	12.3	
18	17	European Hedgehog	785	3.5	
19	18	Tenrec	900	2.6	
20	19	Arctic Ground Squirrel	920	5.7	
21	20	African Giant Pouched Rat	1000	6.6	
22	21	Guinea Pig	1040	5.5	
23	22	Mountain Beaver	1350	8.1	
24	23	Slow Loris	1400	12.5	
25	24	Genet	1410	17.5	
26	25	Phalanger	1620	11.4	

Ready

Which animal is smartest?

- Carl Segan





The Elements of Graphing Data  
[Cleveland]

Log<sub>10</sub> Brain Weight - 2/3 Log<sub>10</sub> Body Weight

Convey Information To  
Others

# Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite à Paris, le 20 Novembre 1869.

The number of men is represented by the width of the coloured bands, where 1 mm equals 6000 men. Furthermore, they're written across the bands. The red numbers represent the men entering Russia while the black numbers represent the men coming out. The information used to make this map were taken from the works of M.M. Chiers, deSécur, deFezensac, deChambray and deJacob's unpublished journal, army pharmacist since the 28th of October. To better gauge the decrease of the army, I supposed that the bodies of Prince Jérôme du Maréchal Davousi that were released on Minsk and Mobilow and/or rejoined around Orscha and Witebsk had always walked with the army.

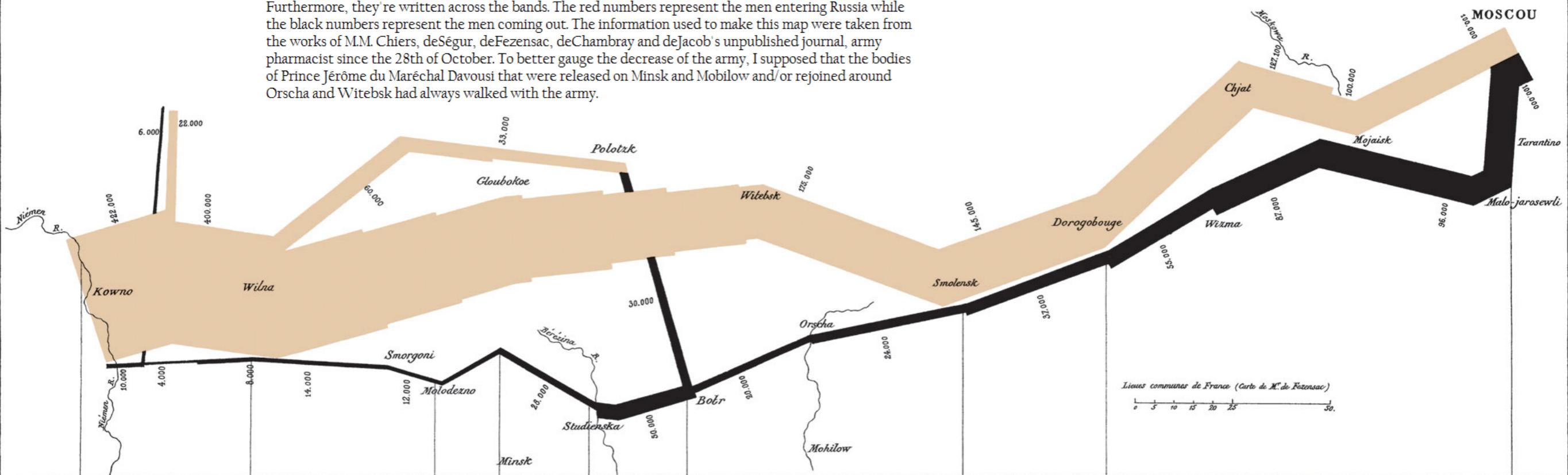
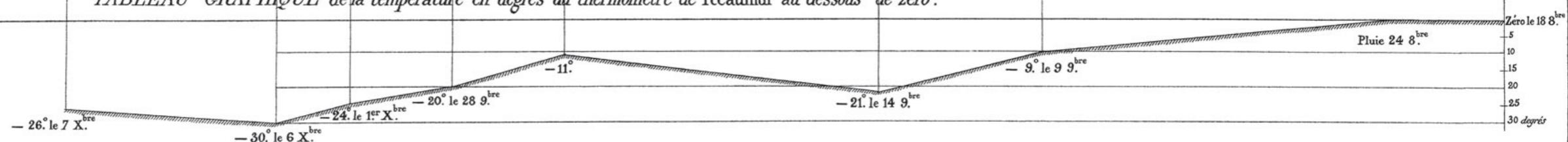


TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

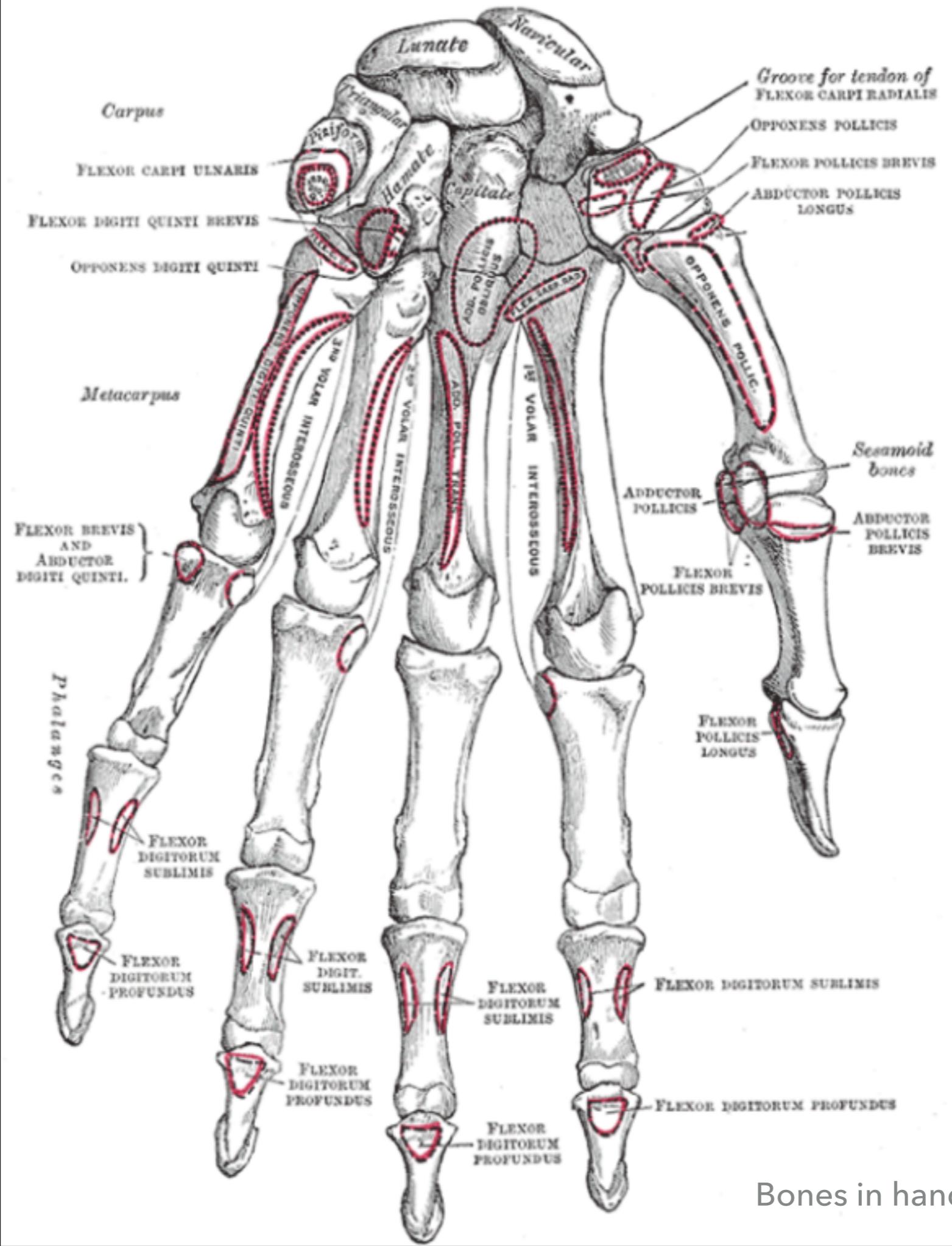
Les Cosaques passent au galop  
le Niemen gelé.



Autog. par Regnier, 8. Pas. S<sup>e</sup> Marie S<sup>t</sup> Gain à Paris.

Imp. Lith. Regnier et Dourdet.

Napoleon's army in the Russian campaign of 1812 by Charles Joseph Minard



Bones in hand [from 1918 edition]



Double helix model [Watson, Crick, Franklin]

# 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

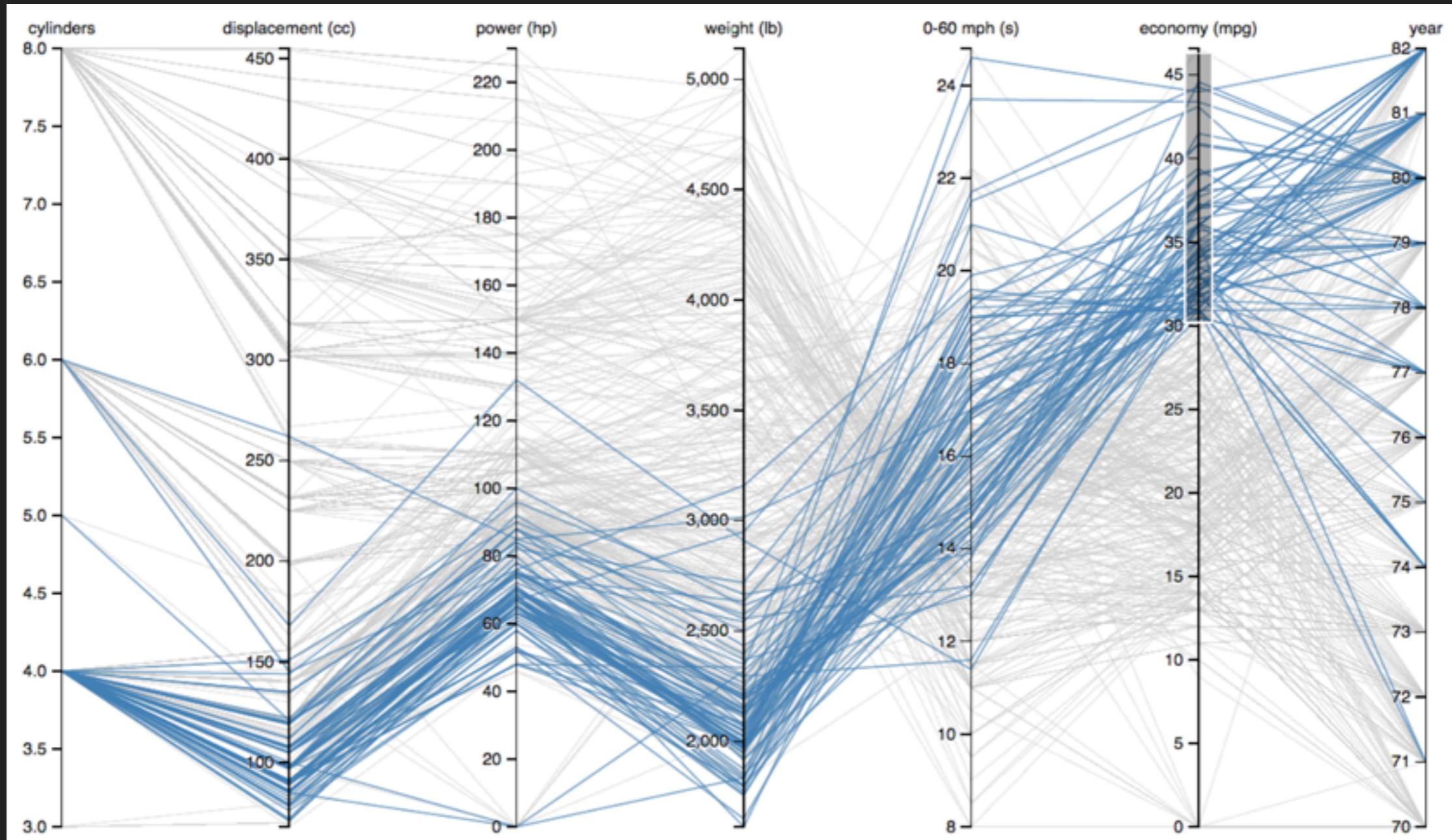
Improve ties between visualization & mental model

# Topics In This Course

# Visual Encoding: Marks & Channels

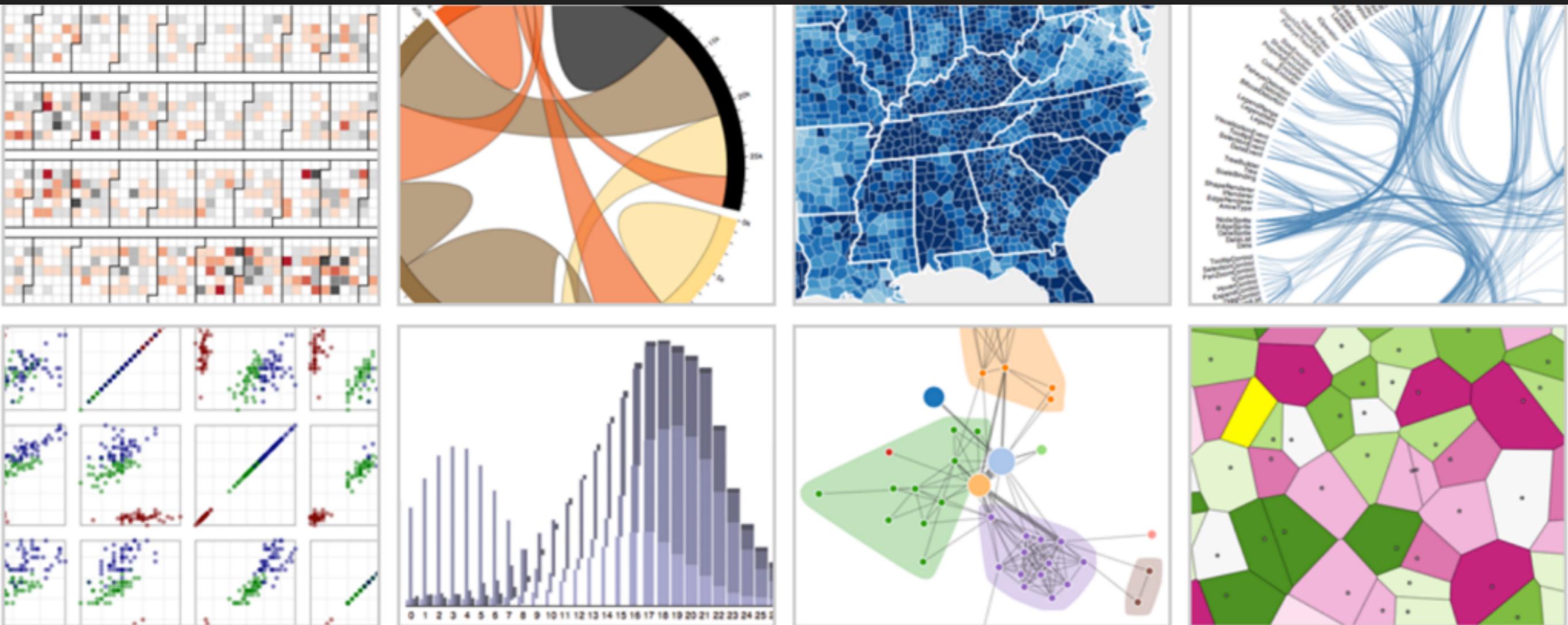
LES VARIABLES DE L'IMAGE							
	POINTS		LIGNES		ZONES		12 14
XY 2 DIMENSIONS DU PLAN	x	x	x	2	2	2	OQ
Z TAILLE	■	■	.	2	2	2	OQ
VALEUR	■	■	■	2	2	2	O
LES VARIABLES DE SÉPARATION DES IMAGES							
GRAIN	■■■	■■■	■■■	2	2	2	13
COULEUR	■	■	■	2	2	2	■
ORIENTATION	■	■	■	2	2	2	■

# Exploratory Data Analysis: Data & Task Abstractions



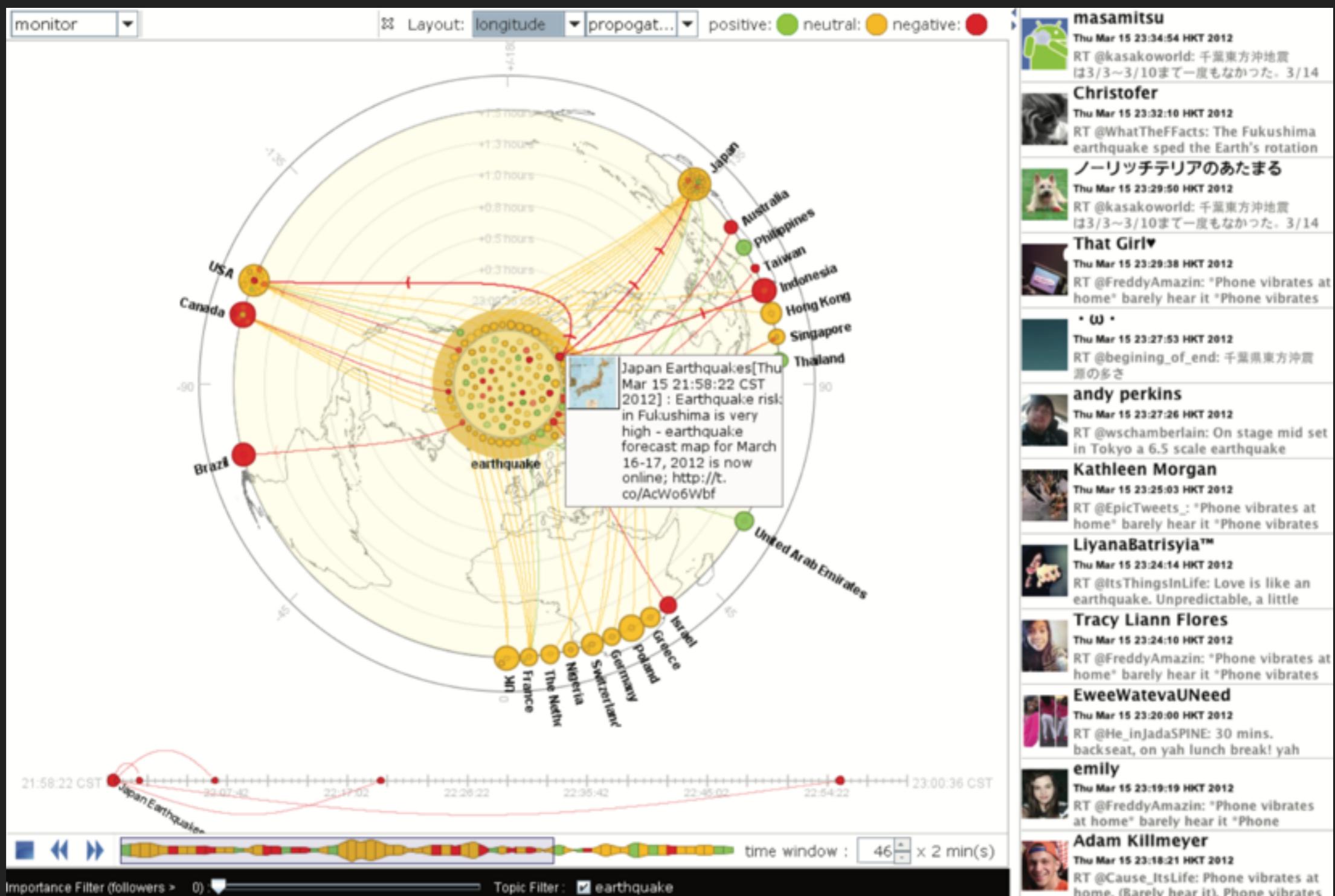
Sémiologie Graphique [Bertin 67]

# Data Representation: Classic Visualizations

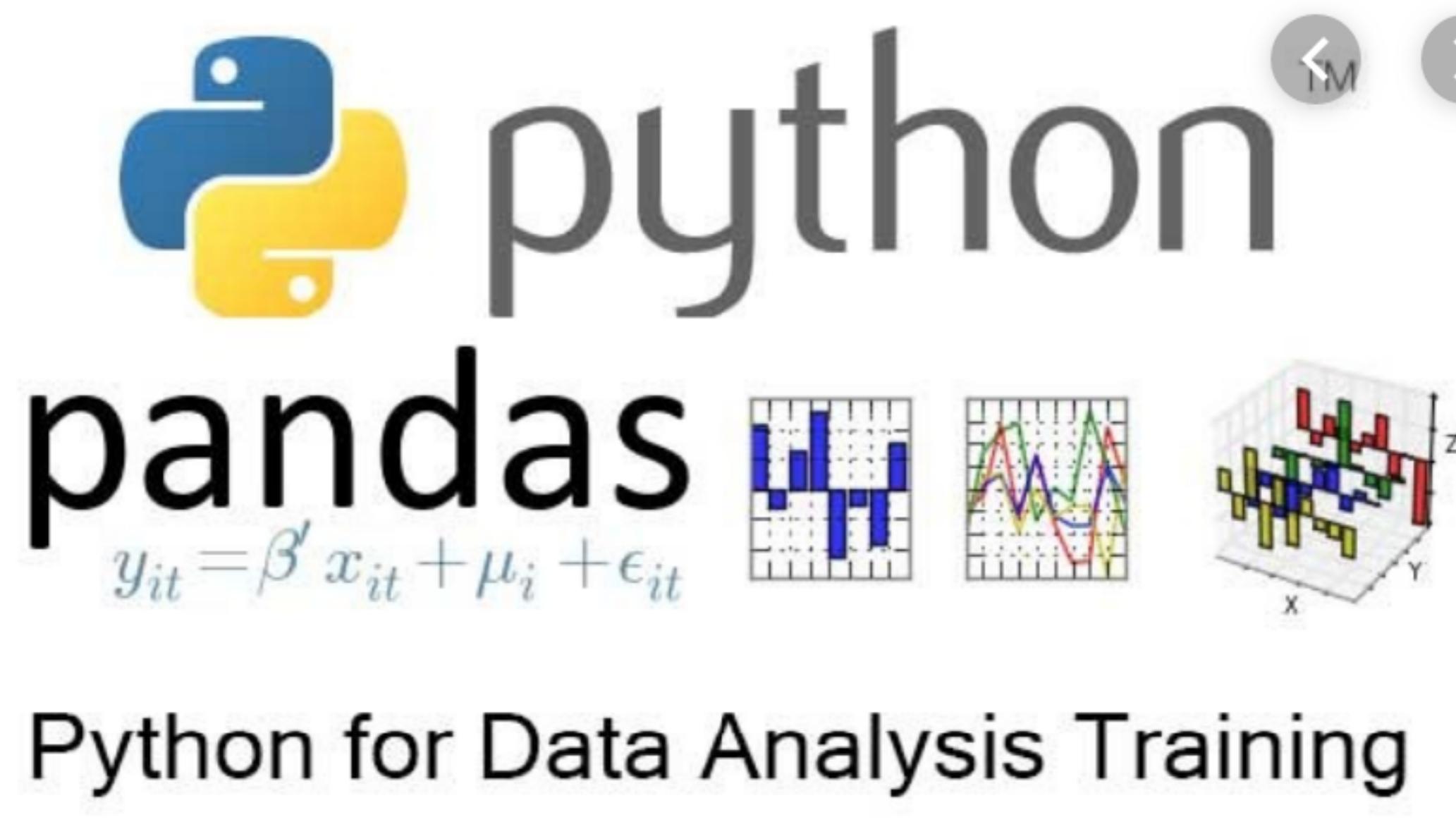


D3: Data-driven Visualization

# Facet Into Multiple Views



# Data Analysis



The image features the Python logo (a blue and yellow snake icon) and the word "python" in a grey sans-serif font. A small circular "TM" symbol is located above the letter "o". Below this, the word "pandas" is written in a large, bold, black sans-serif font. To the right of the text are three small square icons: a heatmap with blue bars, a line graph with multiple colored lines, and a 3D scatter plot with points in red, green, and blue across three axes labeled X, Y, and Z.

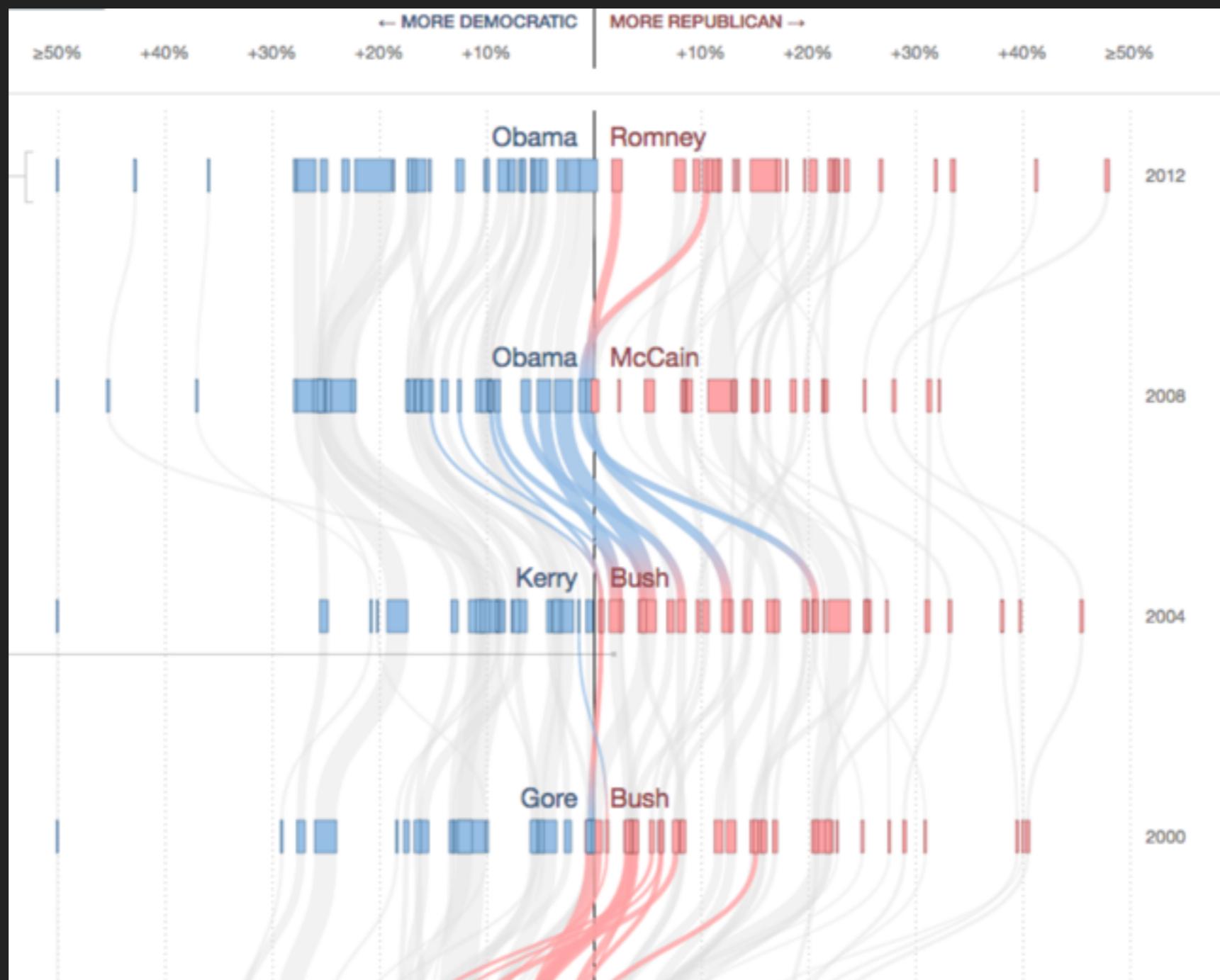
$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$

Python for Data Analysis Training

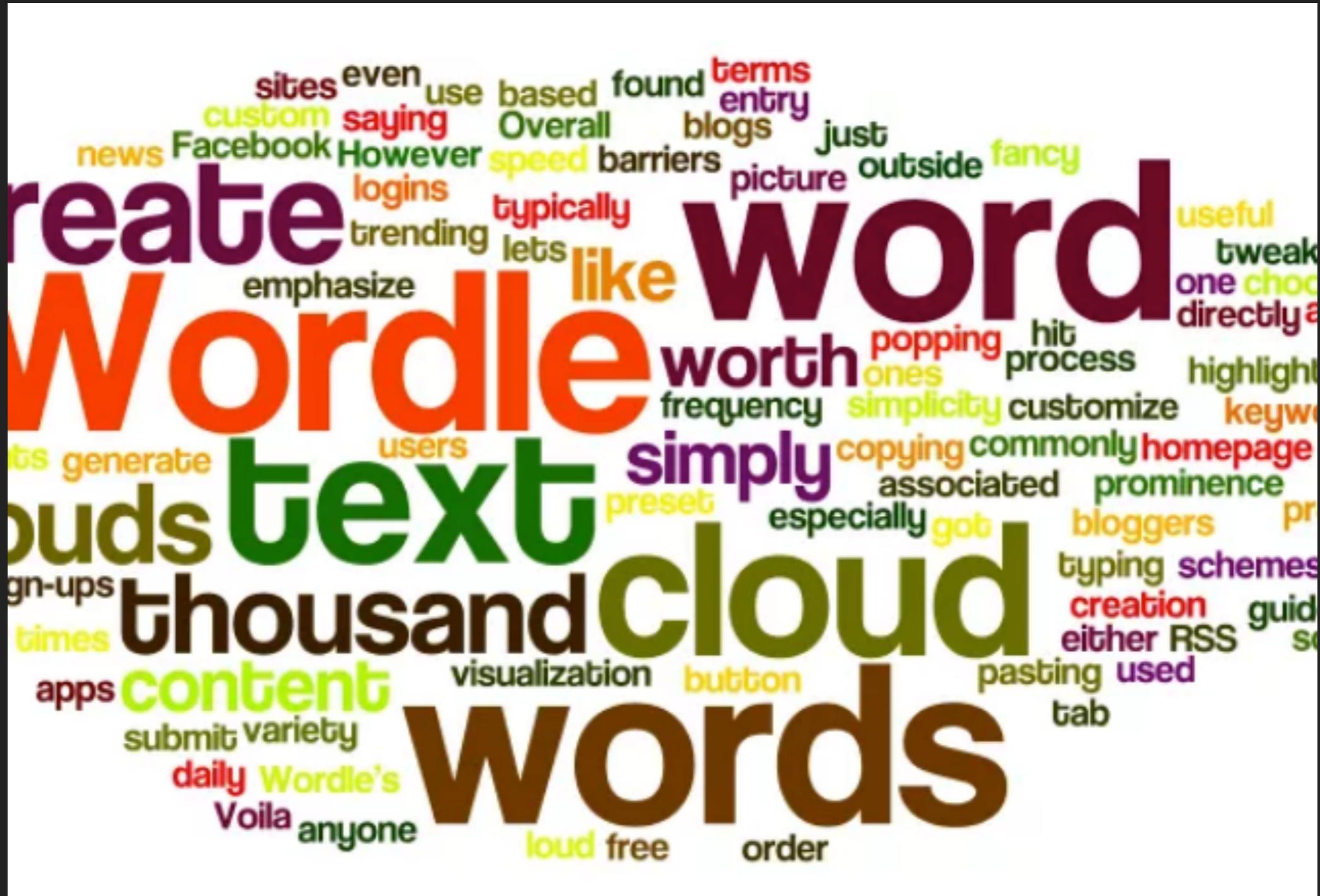
# Cartographic Visualization



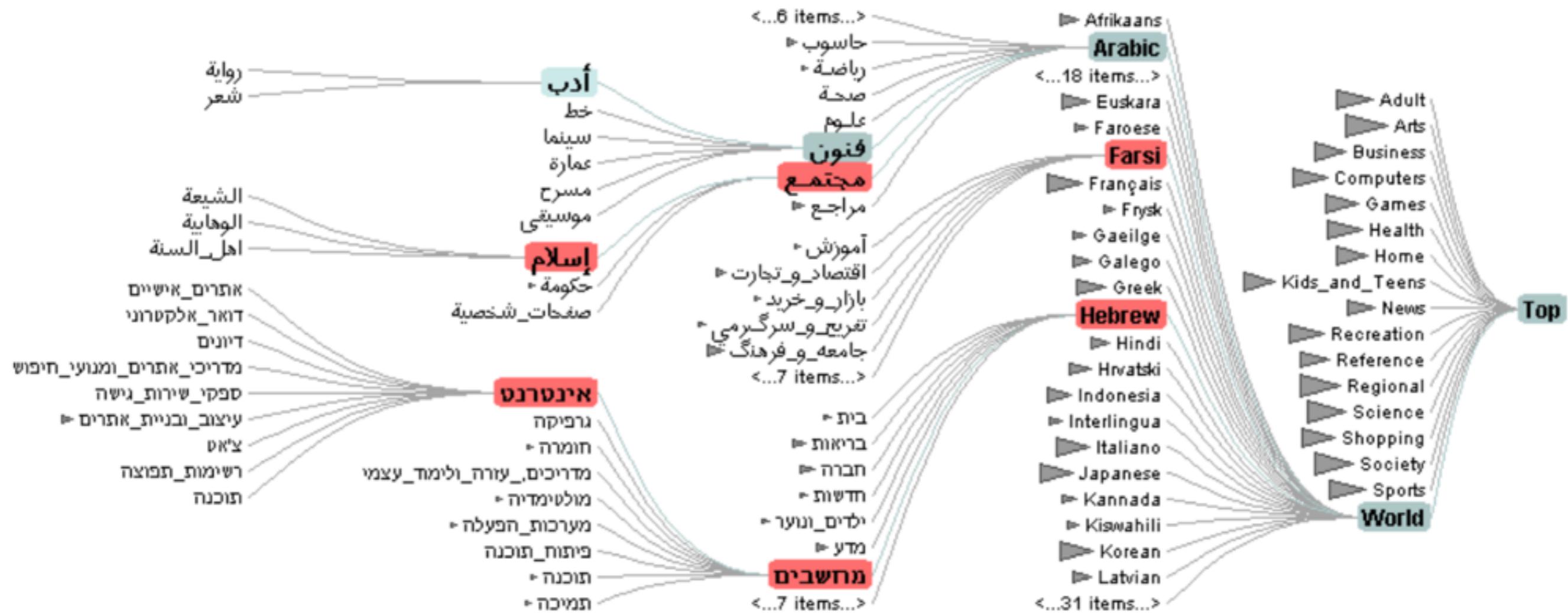
# Temporal Visualization



# Text Visualization

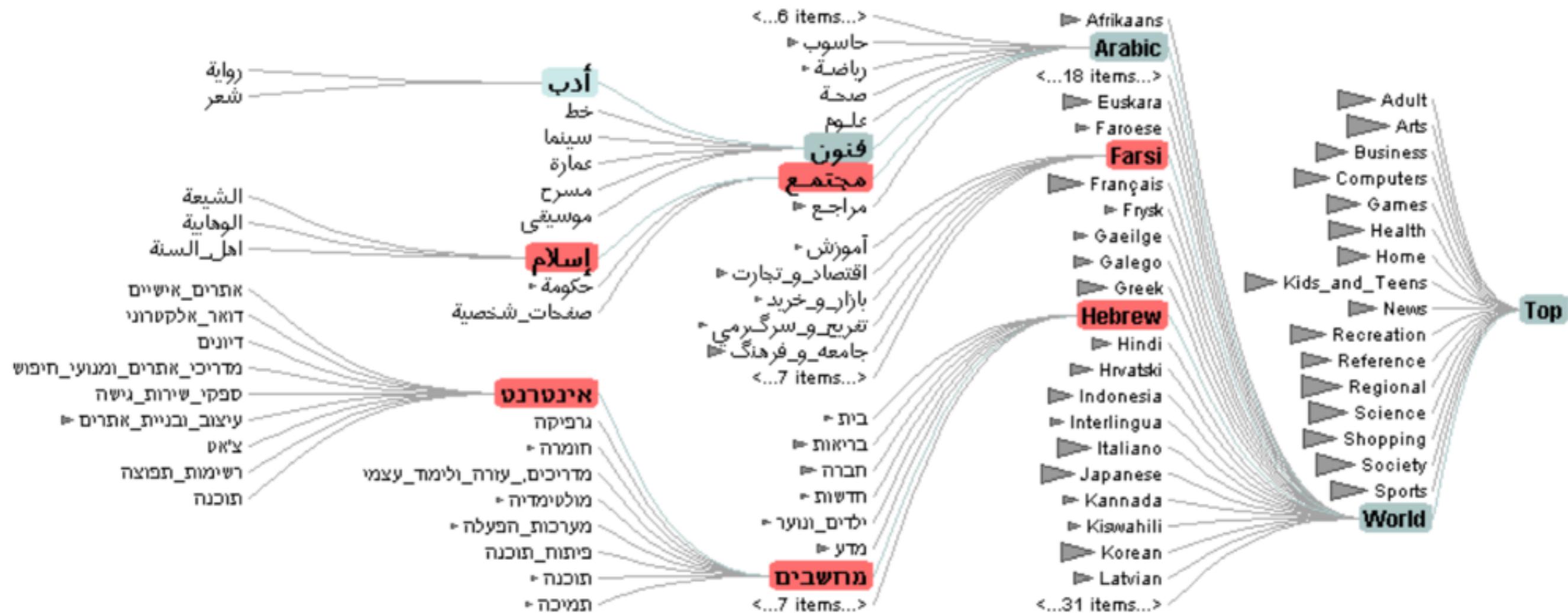


# Network & Tree Visualization



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

# High Dimensional Data Visualization



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

D3.js