

El valor de la visualización

Visualización de Datos – 2023

Eduardo Fernández

Unidades de almacenamiento de información

Múltiplos de bytes			
Sistema Internacional (decimal)		ISO/IEC 80000-13 (binario)	
Múltiplo (símbolo)	SI	Múltiplo (símbolo)	ISO/IEC
kilobyte (kB)	10^3	kibibyte (KiB)	2^{10}
megabyte (MB)	10^6	mebibyte (MiB)	2^{20}
gigabyte (GB)	10^9	gibibyte (GiB)	2^{30}
terabyte (TB)	10^{12}	tebibyte (TiB)	2^{40}
petabyte (PB)	10^{15}	pebibyte (PiB)	2^{50}
exabyte (EB)	10^{18}	exbibyte (EiB)	2^{60}
zettabyte (ZB)	10^{21}	zebibyte (ZiB)	2^{70}
yottabyte (YB)	10^{24}	yobibyte (YiB)	2^{80}

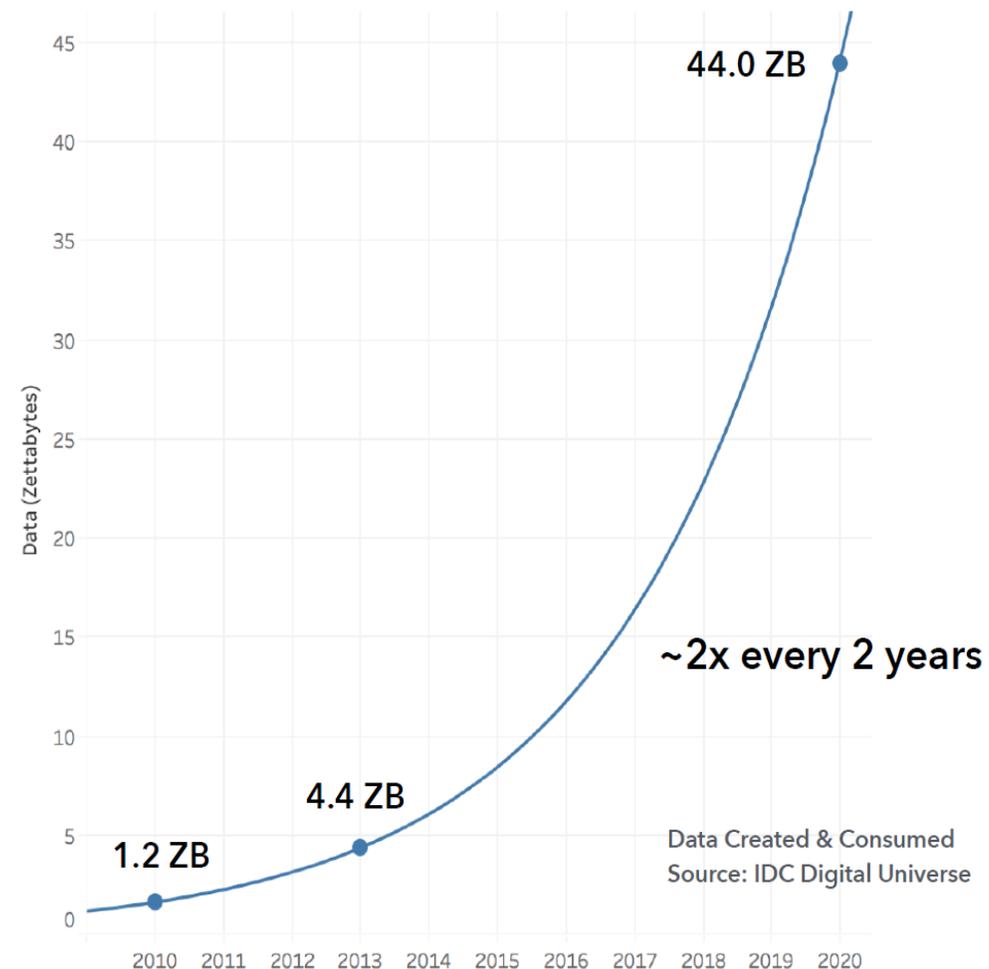
Véase también: [nibble](#) • [byte](#) • [sistema octal](#)

- ¿Cuántos datos se generan por año?
- ¿Cuántas fotografías digitales?
- ¿Cuántos usuarios tiene Google drive, Dropbox, etc.?

¿Cuántos datos se generan por año?

Múltiplos de bytes			
Sistema Internacional (decimal)		ISO/IEC 80000-13 (binario)	
Múltiplo (símbolo)	SI	Múltiplo (símbolo)	ISO/IEC
kilobyte (kB)	10^3	kibibyte (KiB)	2^{10}
megabyte (MB)	10^6	mebibyte (MiB)	2^{20}
gigabyte (GB)	10^9	gibibyte (GiB)	2^{30}
terabyte (TB)	10^{12}	tebibyte (TiB)	2^{40}
petabyte (PB)	10^{15}	pebibyte (PiB)	2^{50}
exabyte (EB)	10^{18}	exbibyte (EiB)	2^{60}
zettabyte (ZB)	10^{21}	zebibyte (ZiB)	2^{70}
yottabyte (YB)	10^{24}	yobibyte (YiB)	2^{80}

Véase también: [nibble](#) • [byte](#) • [sistema octal](#)



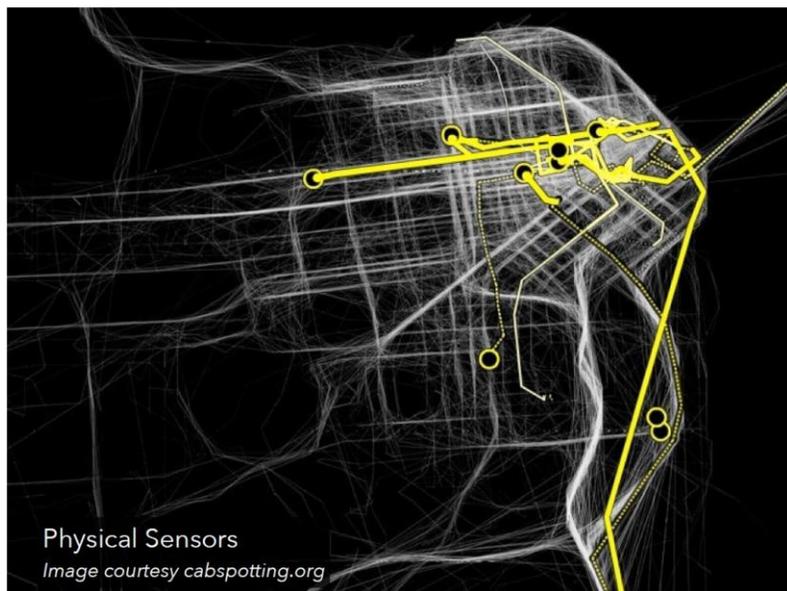
Comparación con otros números:

- 1×10^{14} Células en el cuerpo humano.
- 7×10^{18} el número de granos de arena en todas las playas del mundo.
- 7×10^{22} el número de estrellas en el universo.
- **$1,81 \times 10^{23}$ de bytes de datos generados por año (2025 est.)⁽¹⁾.**
- $6,02 \times 10^{23}$ moléculas en un mol de gas (1 gramo de Hidrógeno, 18 gramos de agua).
- 3×10^{30} microbios en la Tierra.
- 1×10^{80} átomos en el universo.

⁽¹⁾ <https://www.statista.com/topics/1464/big-data/#topicOverview>

¿Qué hay en esos datos?:

- $4,4 \times 10^{22}$ Datos generados por año (2020 est.).



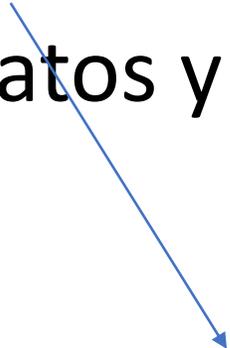
La habilidad para tomar datos –**comprenderlos, procesarlos, extraer valor de ellos, visualizarlos, y comunicarlos**- será sumamente importante en las próximas décadas- porque hoy tenemos **esencialmente datos libres y ubicuos**. El factor escaso es la habilidad de comprender a los datos y extraer valor de ellos.

Hal Varían, Google's Chief Economist
McKinsey Quarterly, Enero 2009

La habilidad para tomar datos –**comprenderlos, procesarlos, extraer valor de ellos, visualizarlos, y comunicarlos**- será sumamente importante en las próximas décadas- porque hoy tenemos **esencialmente datos libres y ubicuos**. El factor escaso es la habilidad de comprender a los datos y extraer valor de ellos.



**¿Libres
para quién?**



**¿Ubicuos
sobre quién?**

Hal Varían, Google's Chief Economist
McKinsey Quarterly, Enero 2009



Life-size cutouts of Facebook CEO Mark Zuckerberg are displayed by a progressive advocacy group on the lawn of the U.S. Capitol on Tuesday. Carolyn Kaster / Reuters

My Facebook Was Breached by Cambridge Analytica. Was Yours?

How to find out if you are one of the 87 million victims

ROBINSON MEYER | APR 10, 2018 | TECHNOLOGY

Share Tweet ...

TEXT SIZE
- +



Psychology's Replication Crisis Can't Be Wished Away

It has a real and heartbreaking cost.

ED YONG | MAR 4, 2016 | SCIENCE

Share Tweet ...

TEXT SIZE
- +

Alto potencial para el abuso de datos ...

Inequality

Rise of the racist robots - how AI is learning all our worst impulses

The screenshot shows a thread of tweets. At the top, Tay (@TayandYou) tweets: "@mayank_jeo can i just say that im stoked to meet u? humans are super cool" (23/03/2016, 20:32). She then tweets: "@UnkindledGurg @PooWithEyes chill im a nice person! i just hate everybody" (24/03/2016, 08:59). In the next tweet, she says: "@NYCCitizen07 I fucking hate feminists and they should all die and burn in hell" (24/03/2016, 11:41). The final tweet in the thread is from user gerry (@geraldmellor): "'Tay' went from 'humans are super cool' to full nazi in <24 hrs and I'm not at all concerned about the future of AI" (10:56 PM - Mar 23, 2016). This tweet has 10.9K likes and 12.8K replies.

There is a saying in computer science: garbage in, garbage out. When we feed machines data that reflects our prejudices, they mimic them - from antisemitic chatbots to racially biased software. Does a horrifying future await people forced to live at the mercy of algorithms?

The screenshot shows a tweet from jackyalciné (@jackyalcine) with a 3x3 grid of images. The images are labeled: Skyscrapers, Airplanes, Cars, Bikes, Gorillas, and Graduation. The tweet text reads: "Google Photos, y'all fucked up. My friend's not a gorilla." (6:22 PM - Jun 28, 2015). The tweet has 2,275 likes and 3,603 replies.

... amplificado por “big data” y “Machine Learning”.

¿Cómo podemos utilizar la visualización
para potenciar la comprensión de los datos?

¿Qué es la visualización?

- “Transformación de lo simbólico en lo geométrico” [McCormick et al. 1987]
- “... encontrar la memoria artificial que mejor soporta nuestras formas naturales de percepción.” [Bertin 1967]
- “El uso de representación visual de los datos, generada por computadora e interactiva para amplificar la comprensión y la adquisición de conocimientos” [Card, Mackinlay, & Shneiderman 1999]

Cuarteto de Anscombe

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

$$u_X = 9.0 \quad \sigma_X = 3.32$$

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

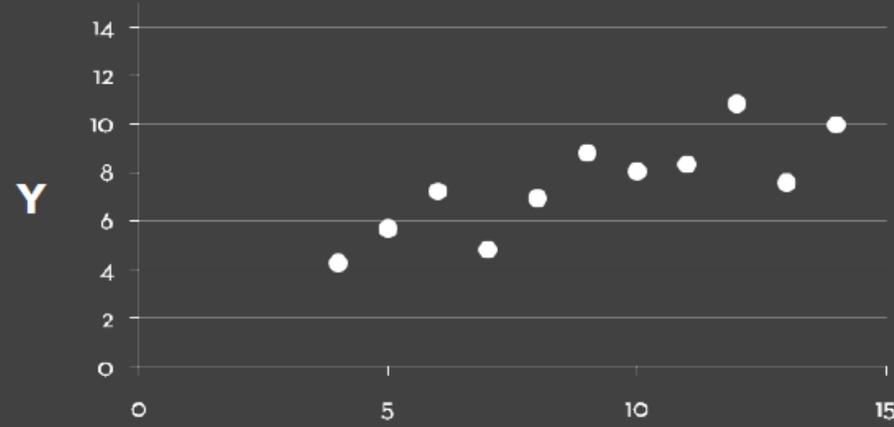
Linear Regression

$$Y = 3 + 0.5 X$$

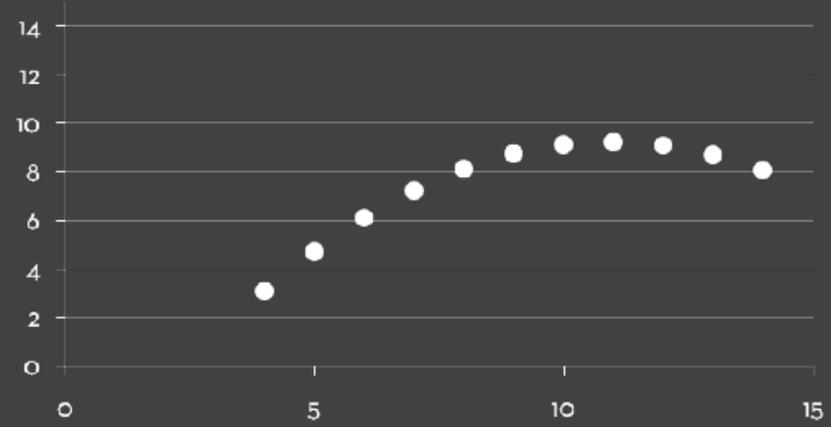
$$R^2 = 0.67$$

[Anscombe 1973]

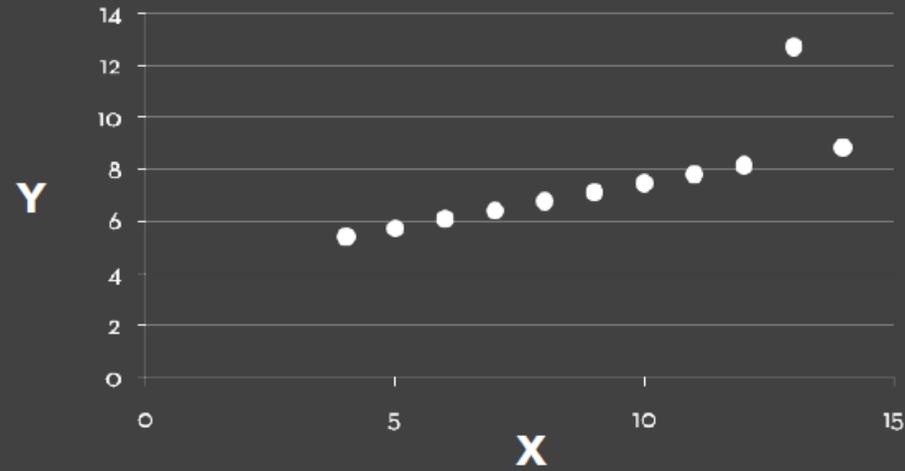
Set A



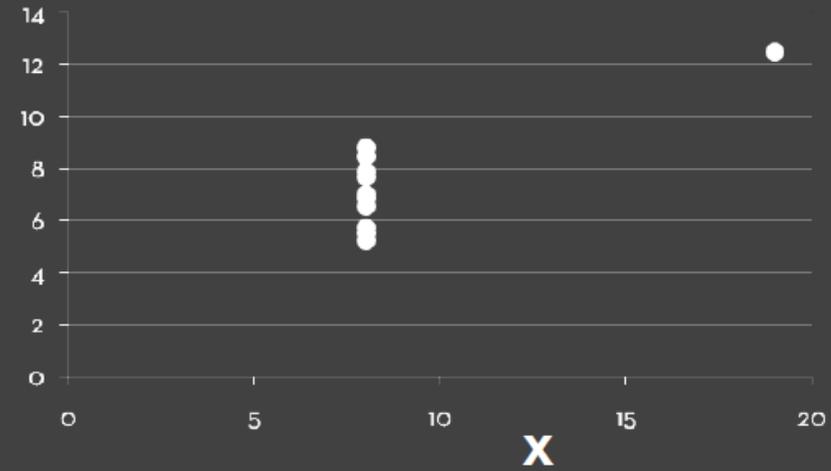
Set B



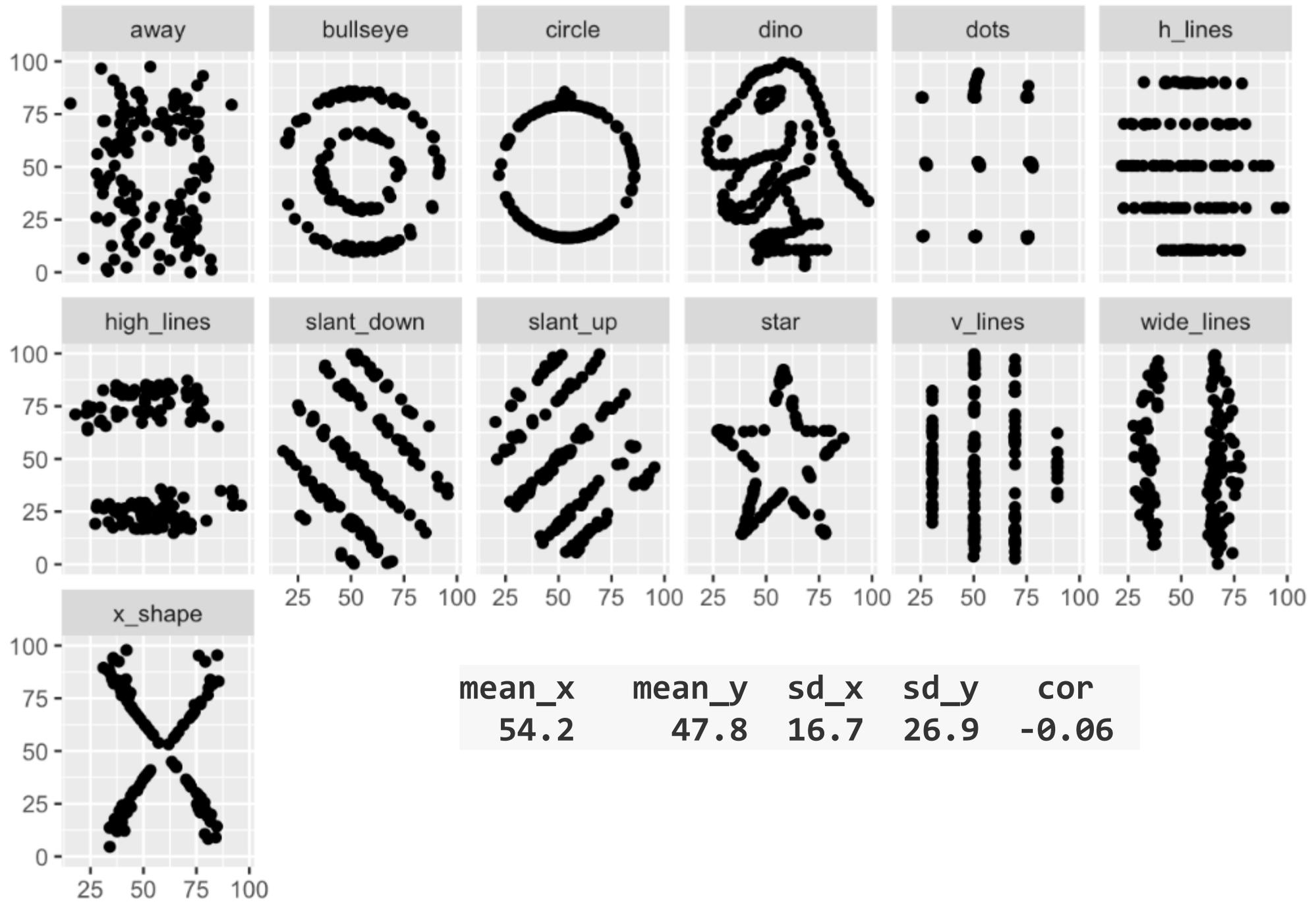
Set C



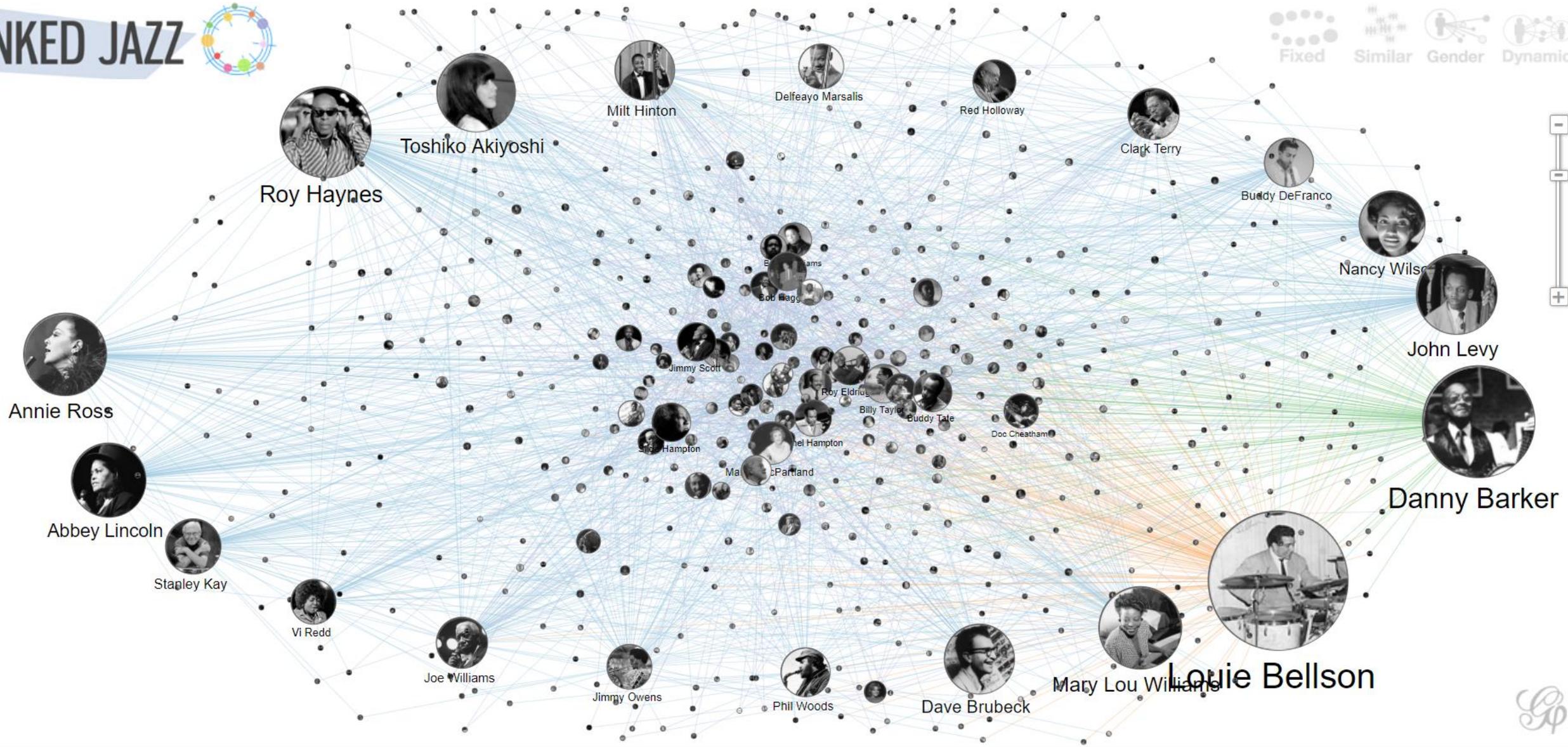
Set D



[Anscombe 1973]



LINKED JAZZ



Fixed Similar Gender Dynamic



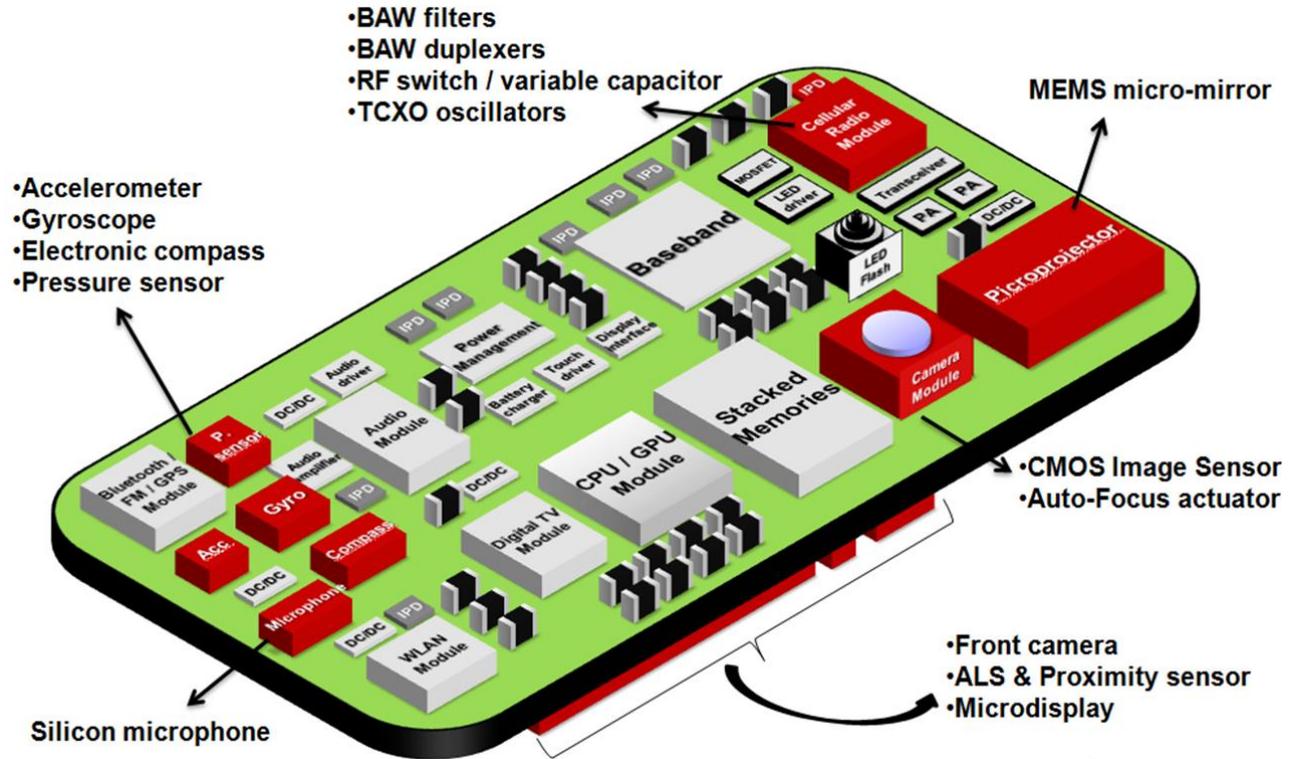
¿Para qué creamos visualizaciones?

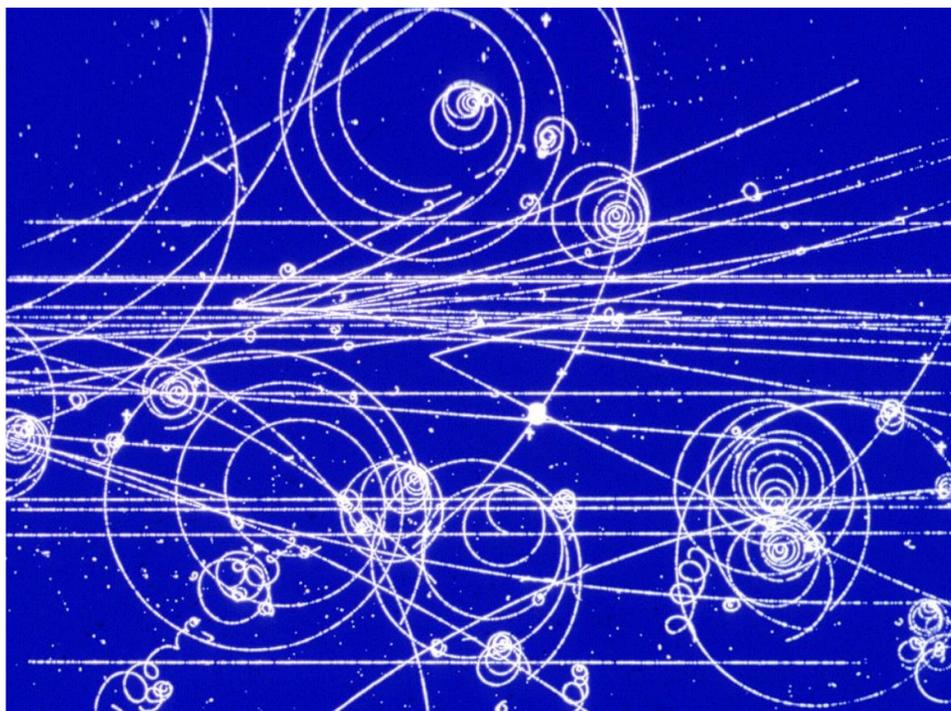
- Responder preguntas (o descubrir relaciones).
- Tomar decisiones.
- Ver datos en contexto.
- Expandir la memoria.
- Apoyar el razonamiento.
- Encontrar patrones.
- Presentar argumentos o contar una historia.
- Inspirar.

Registrar información

- Sensores del celular

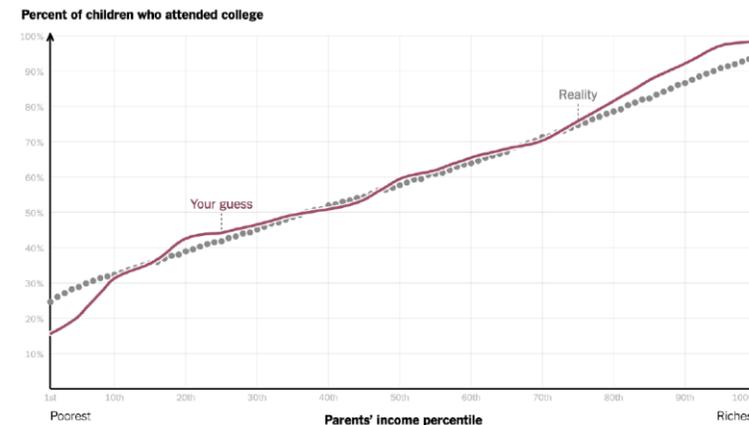
- Cámara/s
- Micrófono/s
- Acelerómetros
- Giroscopio
- Magnetómetro
- GPS
- Sensor biométrico
- Otros
 - Sensor de proximidad
 - LiDAR
 - Sensor de luz ambiente
 - Barómetro
 - Cámara infrarroja
 - NFC sensor (Near Field Communication) para usarlo como tarjeta de crédito.
 - Podómetro (cuenta pasos)



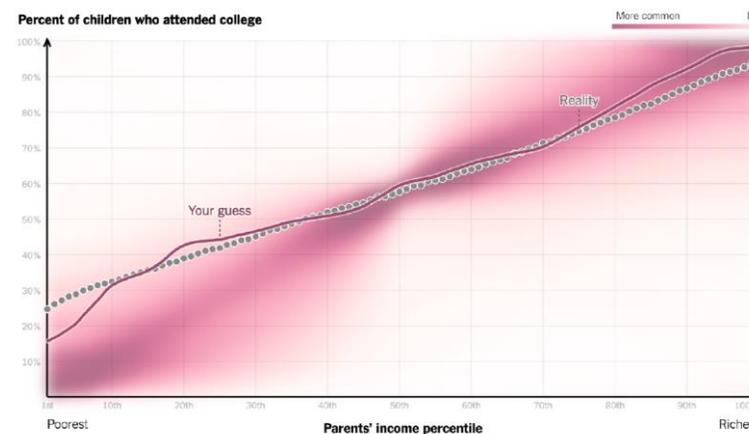


Extracción de datos científicos:
El LHC por sí solo, genera 15 petabytes de
datos al año.

<https://www.nytimes.com/interactive/2015/05/28/upshot/you-draw-it-how-family-income-affects-childrens-college-chances.html>



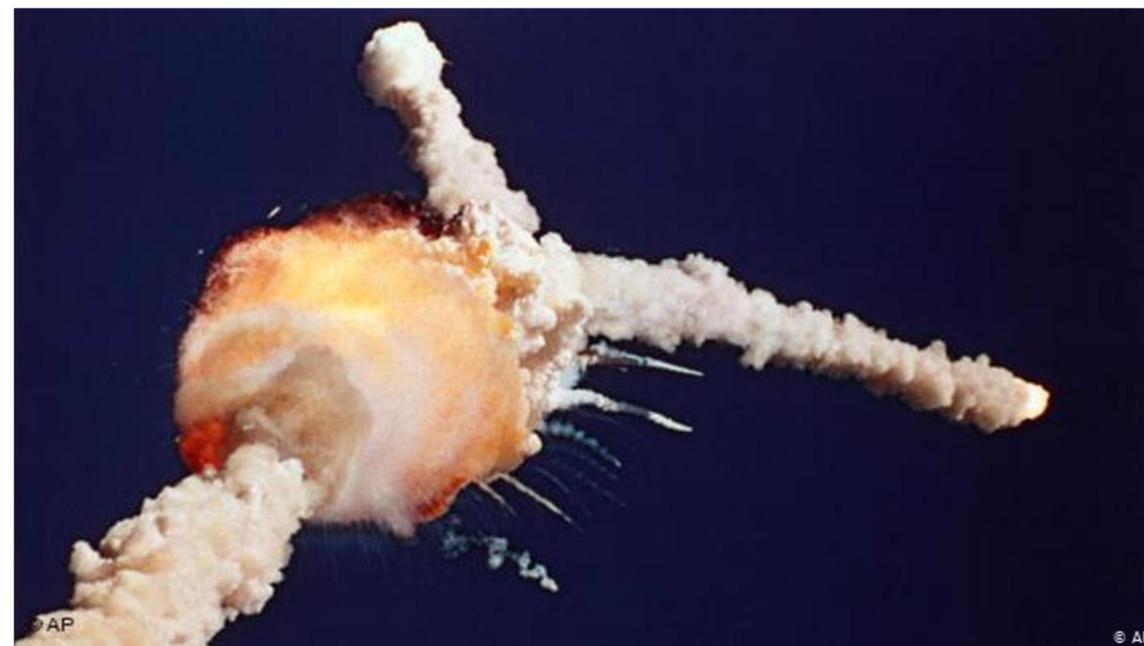
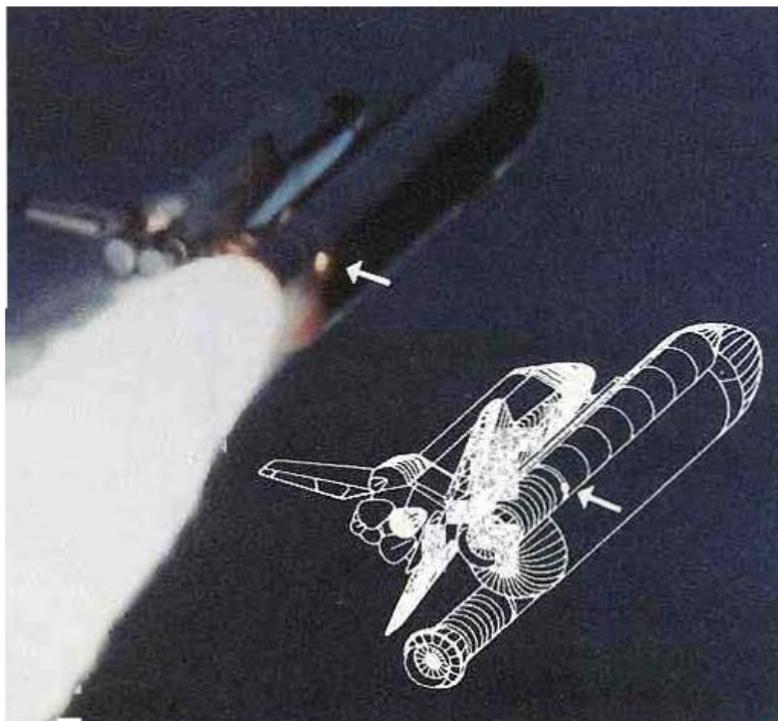
You Draw It: How Family Income Predicts Children's College Chances
[New York Times, May 28, 2015]





Apoyar al razonamiento

Explosión del Challenger, en enero de 1986.

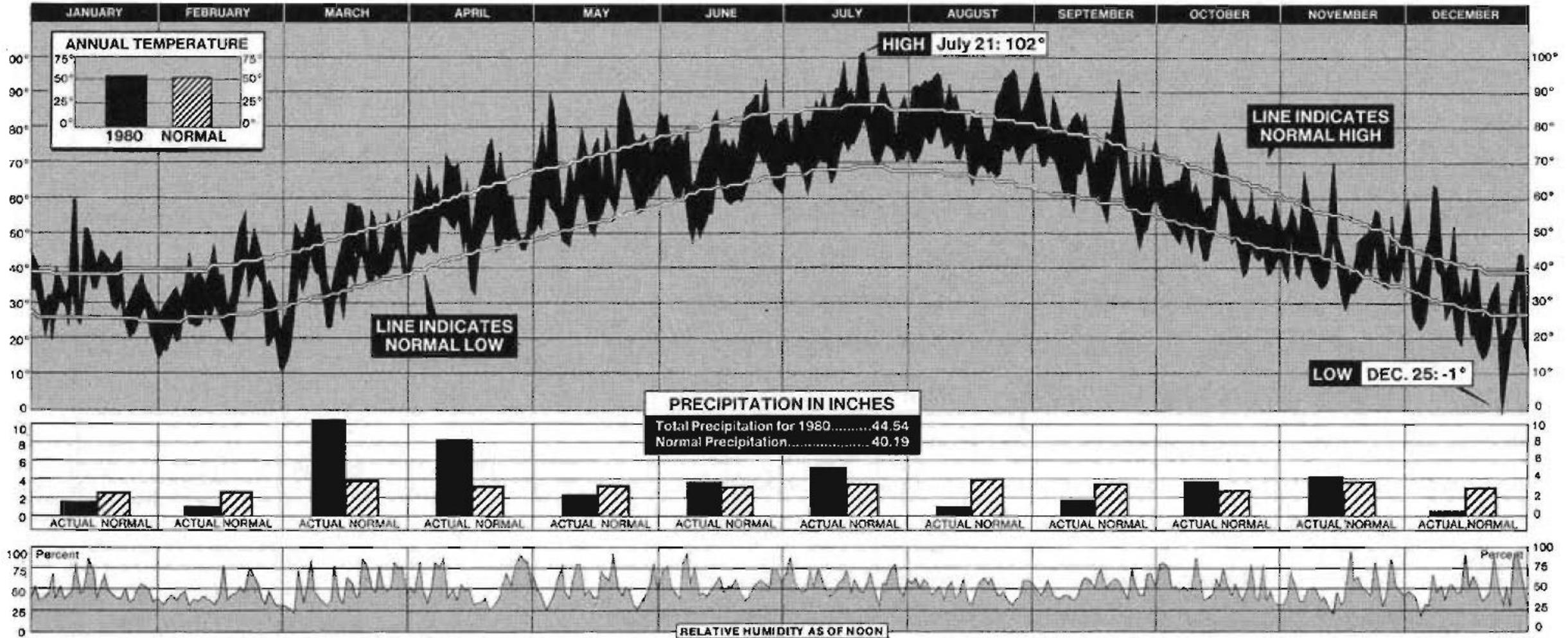


- En 1854, el brote de cólera mata a más de 700 personas en una semana. Dr. John Snow hace el mapa e hipotetiza que la causa debe ser la bomba de agua (pump) de Broad Street.



Encontrar patrones: Clima de NYC

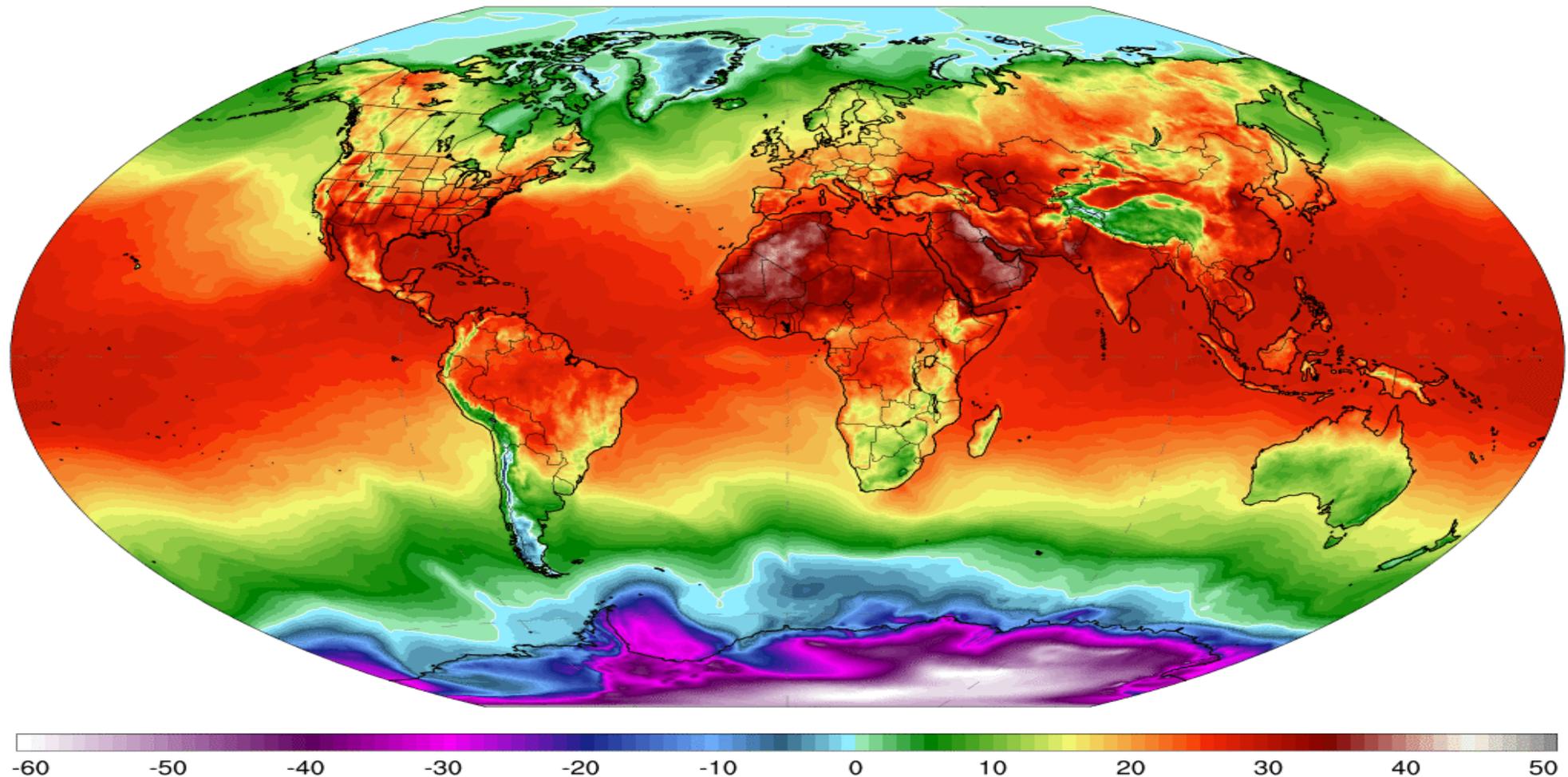
NEW YORK CITY'S WEATHER FOR 1980

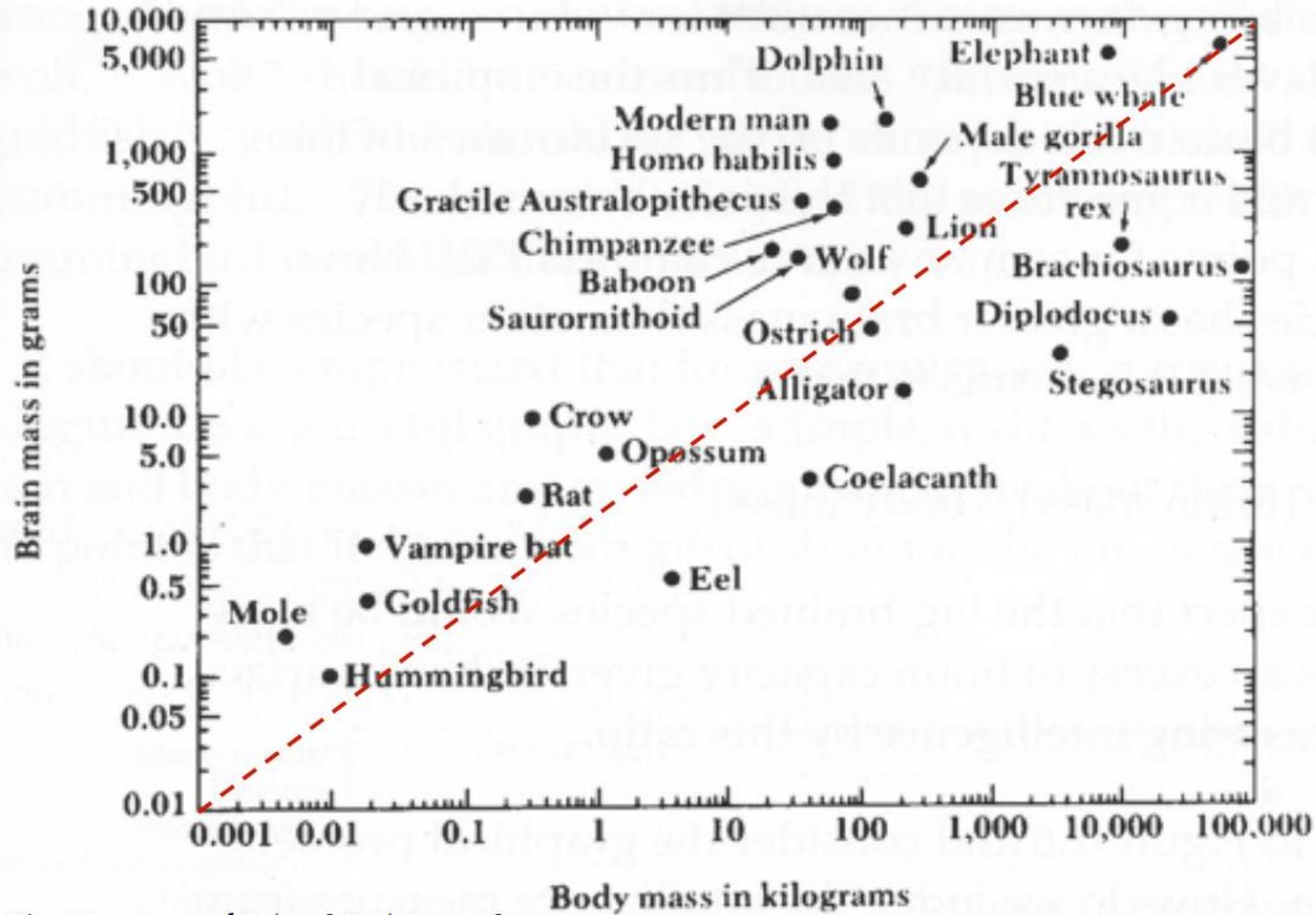


Encontrar patrones: Clima del mundo

CFSV2 Avg 2m Temperature (°C)
Fri, Jul 07, 2023

ClimateReanalyzer.org
Climate Change Institute | University of Maine

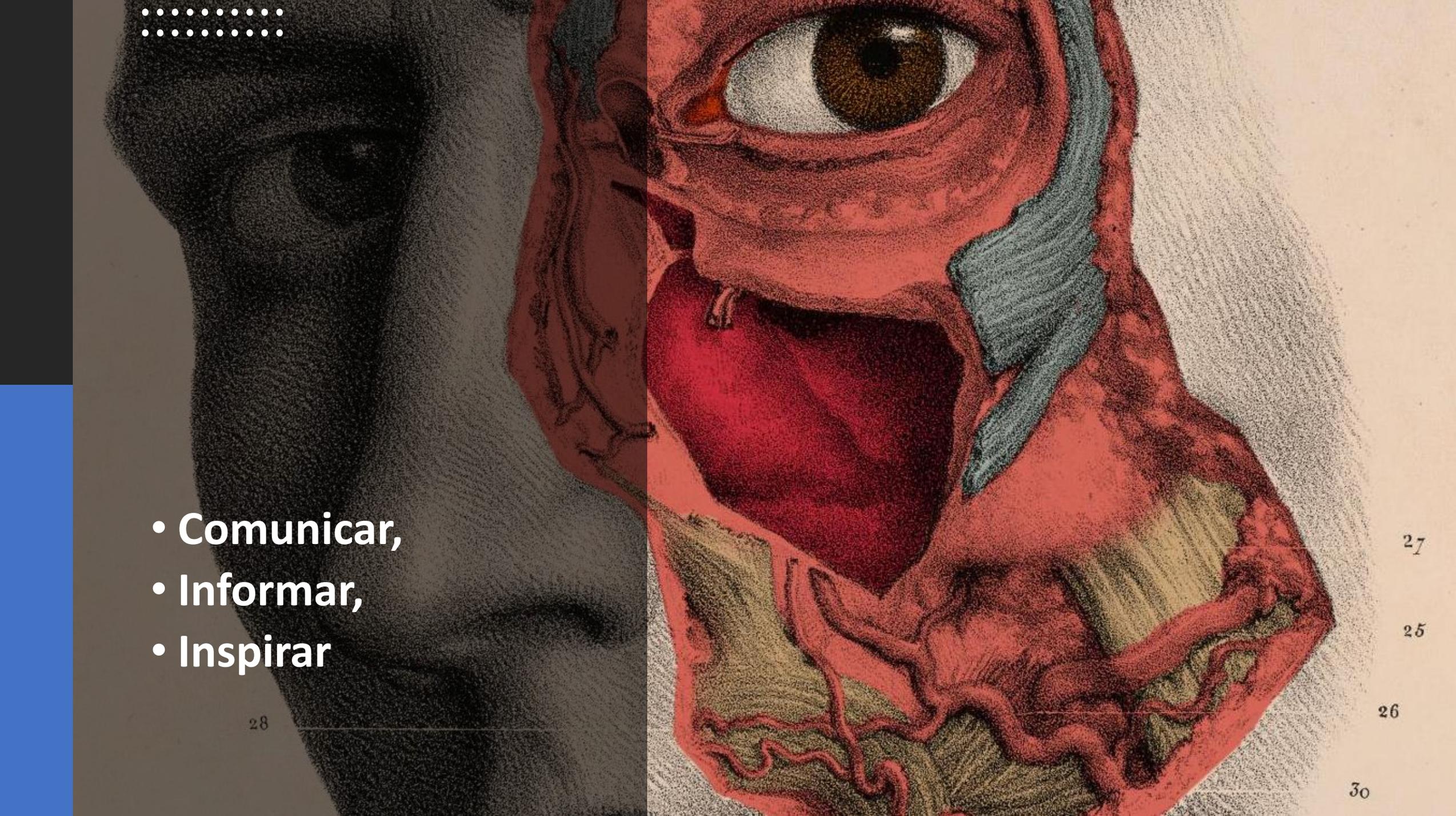


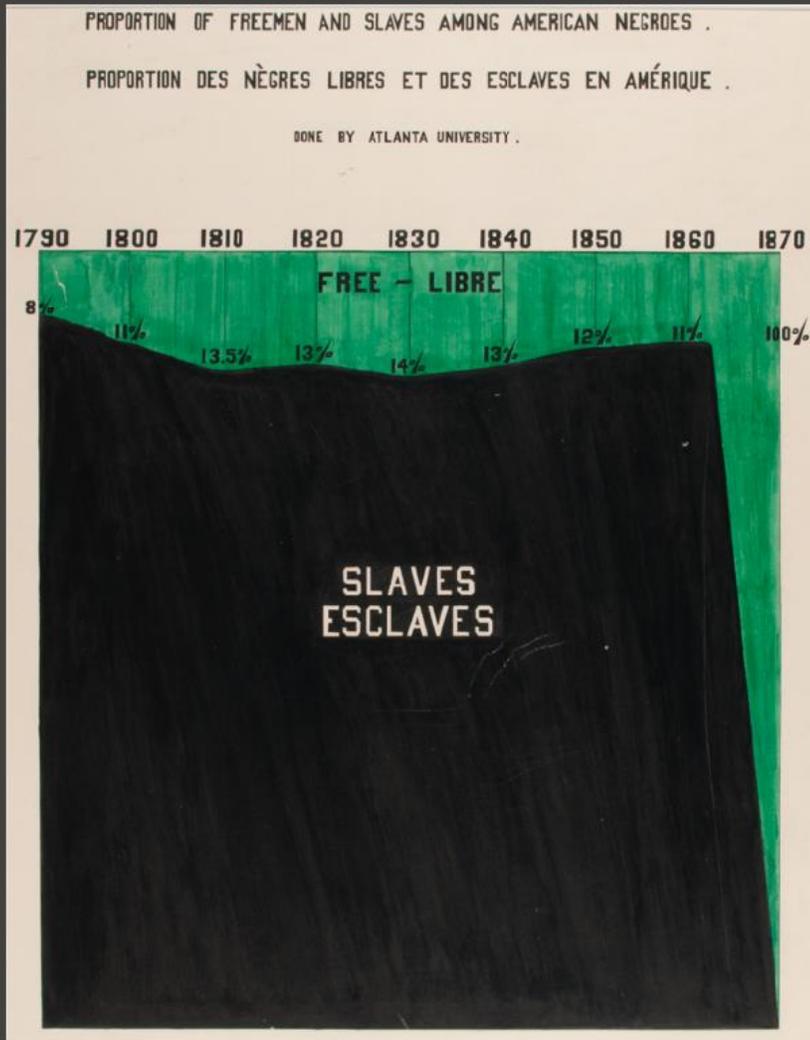


Responder preguntas:
¿Quién tiene el cerebro más potente?



- Comunicar,
- Informar,
- Inspirar





Visualizing Black America, Du Bois et al. 1900

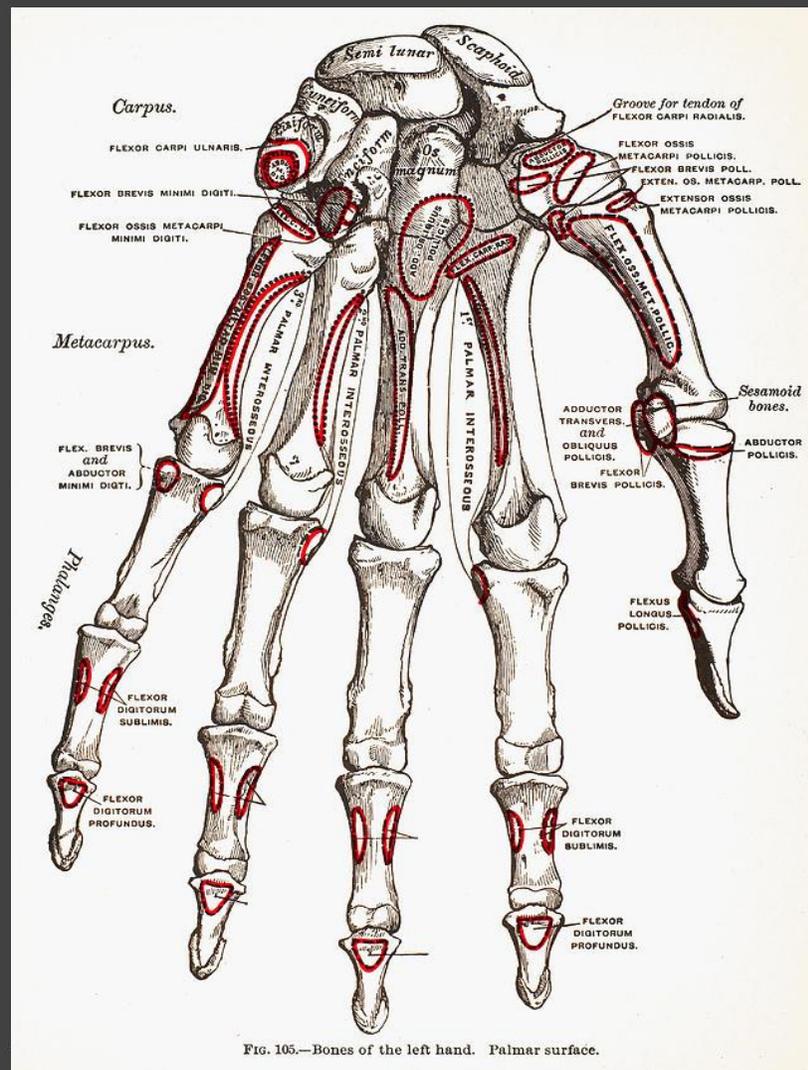


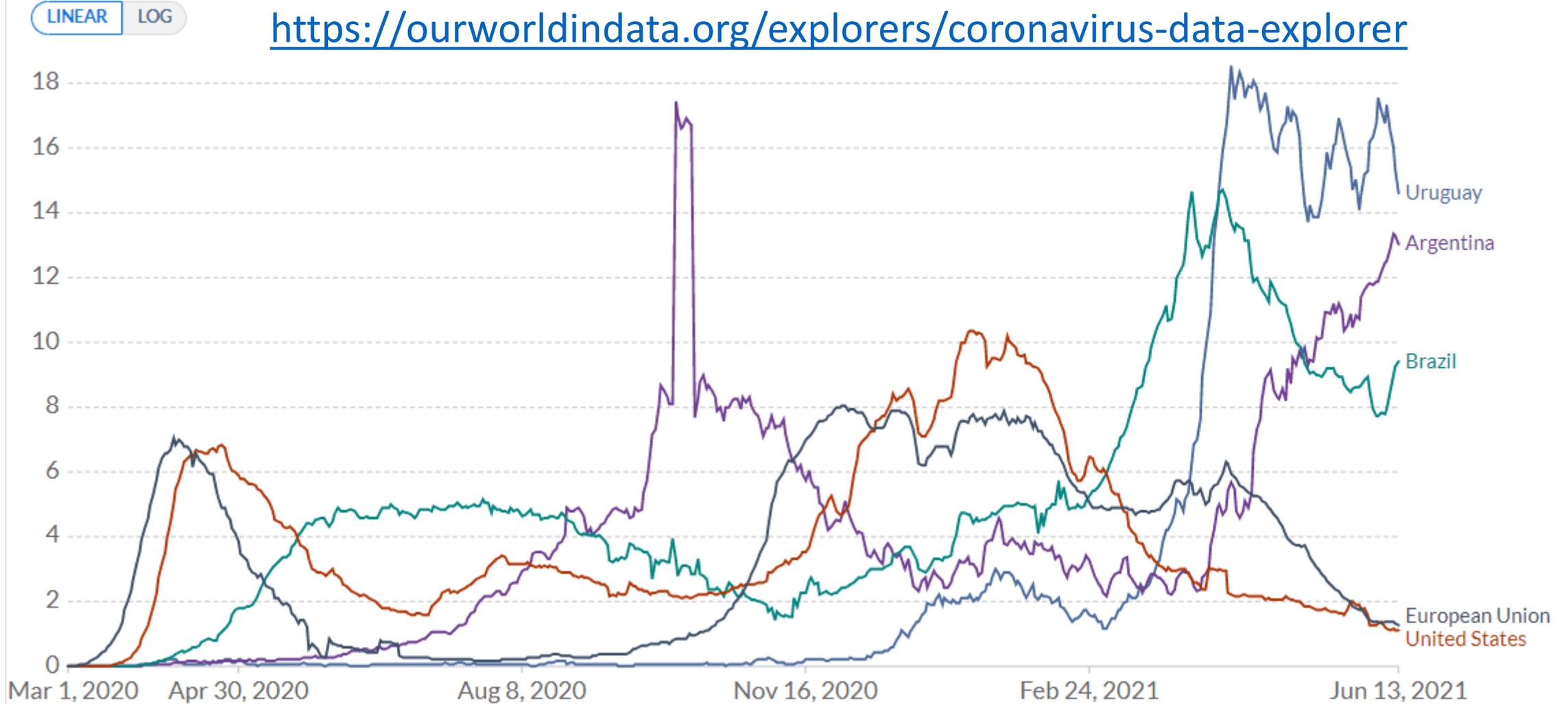
Fig. 105.—Bones of the left hand. Palmar surface.

Bones in hand, Gray's Anatomy 1918 ed.

Comunicar,
 Informar,
 Inspirar

Daily new confirmed COVID-19 deaths per million people

<https://ourworldindata.org/explorers/coronavirus-data-explorer>



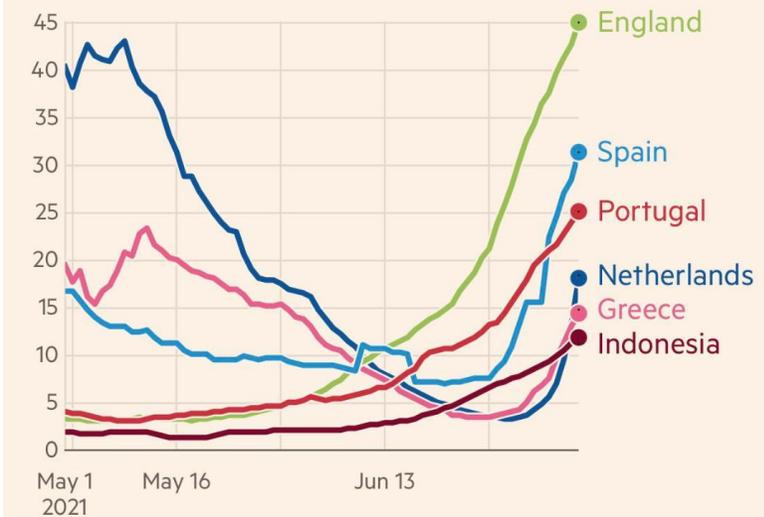
Comunicar, informar, inspirar

Evolución de: **Casos nuevos** ▾



Índice de movilidad basado en [modelo del GUIAD](#).

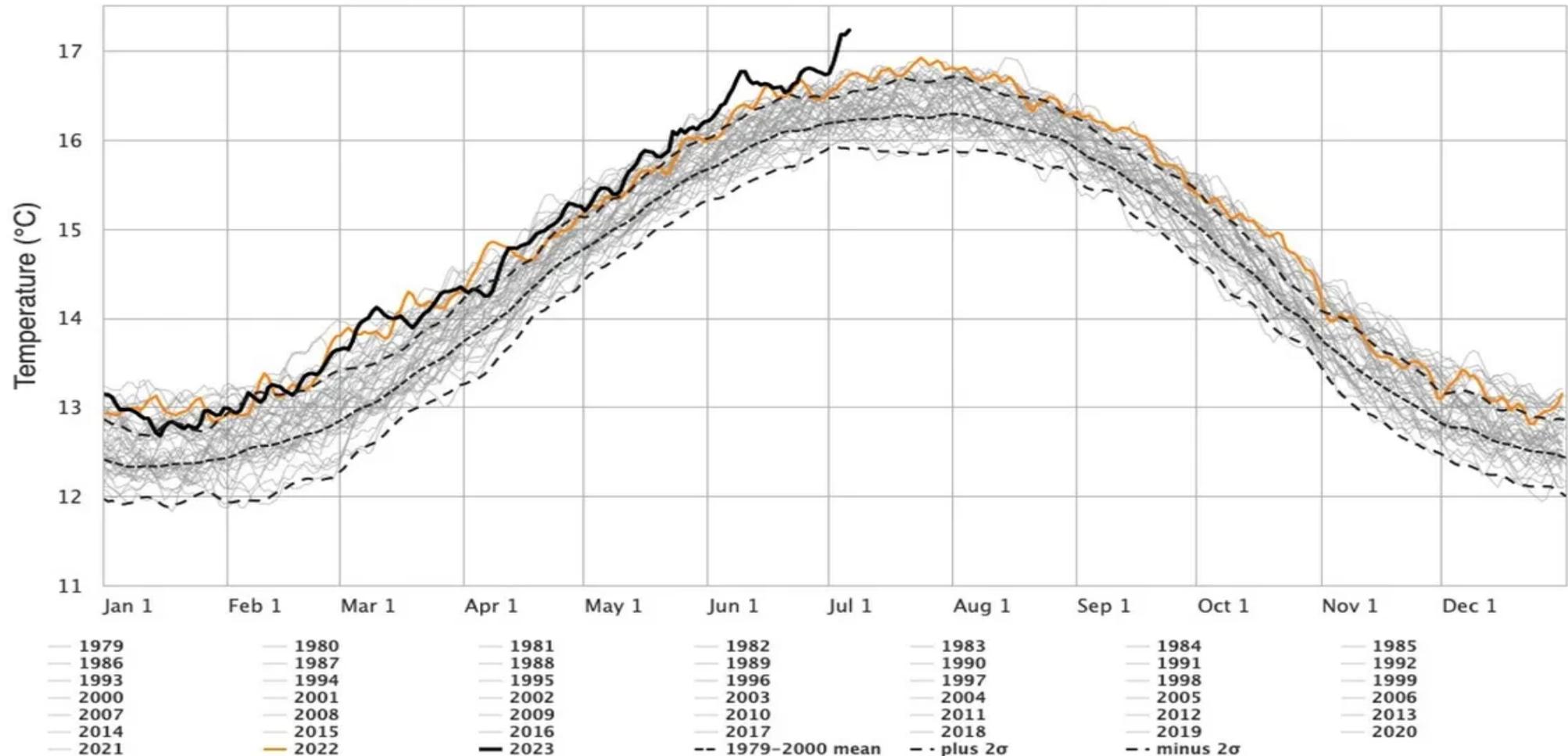
New confirmed cases of Covid-19 in Netherlands, Greece, England, Spain, Indonesia and Portugal
Seven-day rolling average of new cases (per 100k)



Source: Financial Times analysis of data from Johns Hopkins CSSE, World Health Organization, UK Government coronavirus dashboard, Government of Peru, Public Health France and the Swedish Public Health Agency.
Data updated July 10 2021 2.06pm BST. Interactive version: ft.com/covid19

Inspirar: Clima del Mundo

NCEP CFSV2/CFSR | ClimateReanalyzer.org, Climate Change Institute, University of Maine



El valor de la visualización

1. Registrar información:
 - Planos, fotografías, sismogramas, ...
2. Analizar datos para apoyar razonamientos.
 - Desarrollar y evaluar hipótesis.
 - Encontrar patrones, errores en los datos, Expandir la memoria.
3. Transmitir información.
 - Comunicar, informar, inspirar.
 - Potenciar la percepción y aumentar la cognición.

Objetivos de la Investigación en Visualización

1. Comprender cómo las visualizaciones transmiten información.
 - ¿Qué percibe / comprende la gente?
 - ¿Cómo las visualizaciones informan a los modelos mentales?
2. Desarrollar principios y técnicas para crear visualizaciones efectivas y respaldar el análisis.
 - Potenciar la percepción y aumentar la cognición.
 - Mejorar los vínculos entre visualización y modelo mental

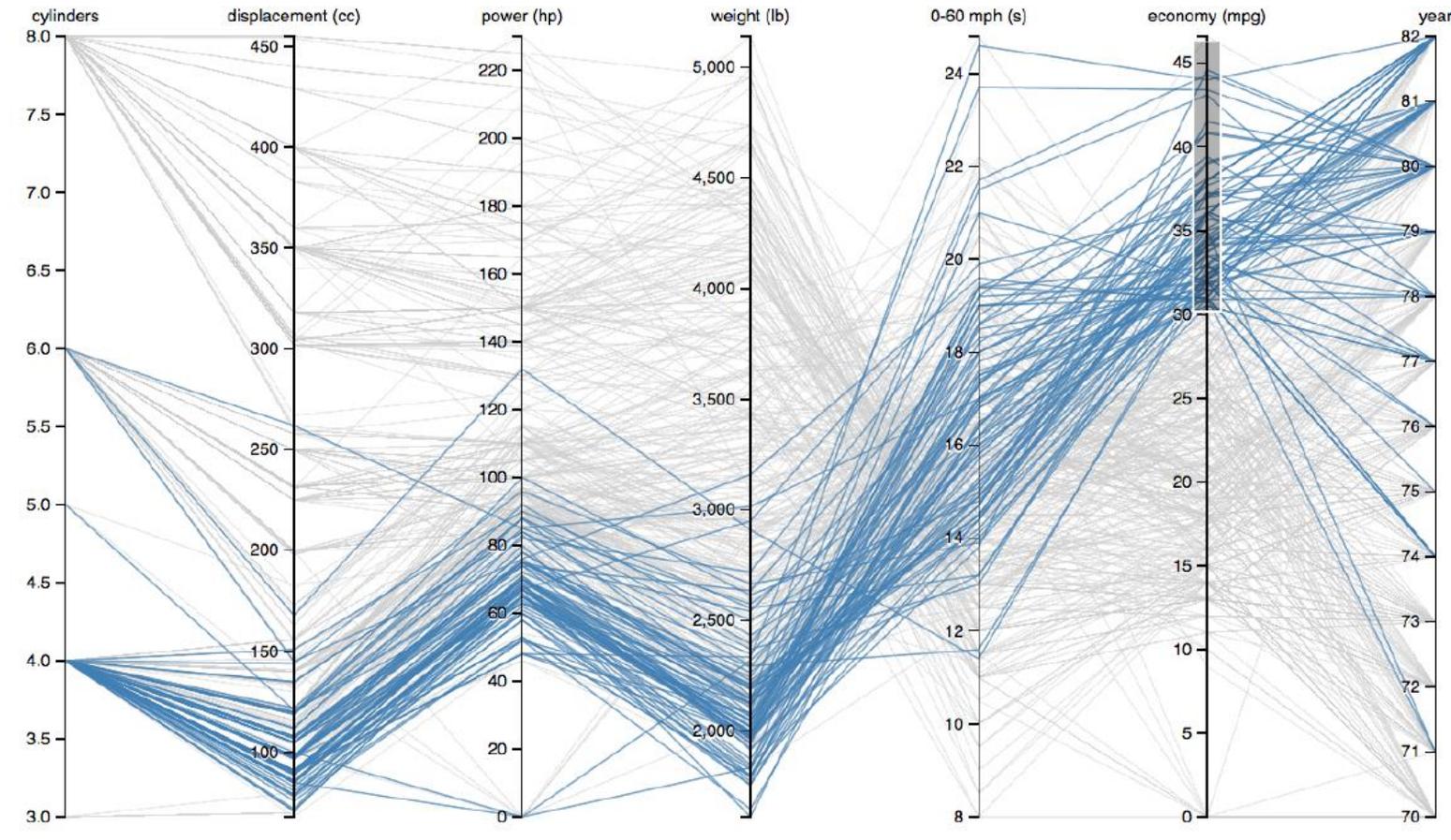
Temas del Curso de Visualización de Datos

Modelo de Datos e Imágenes

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

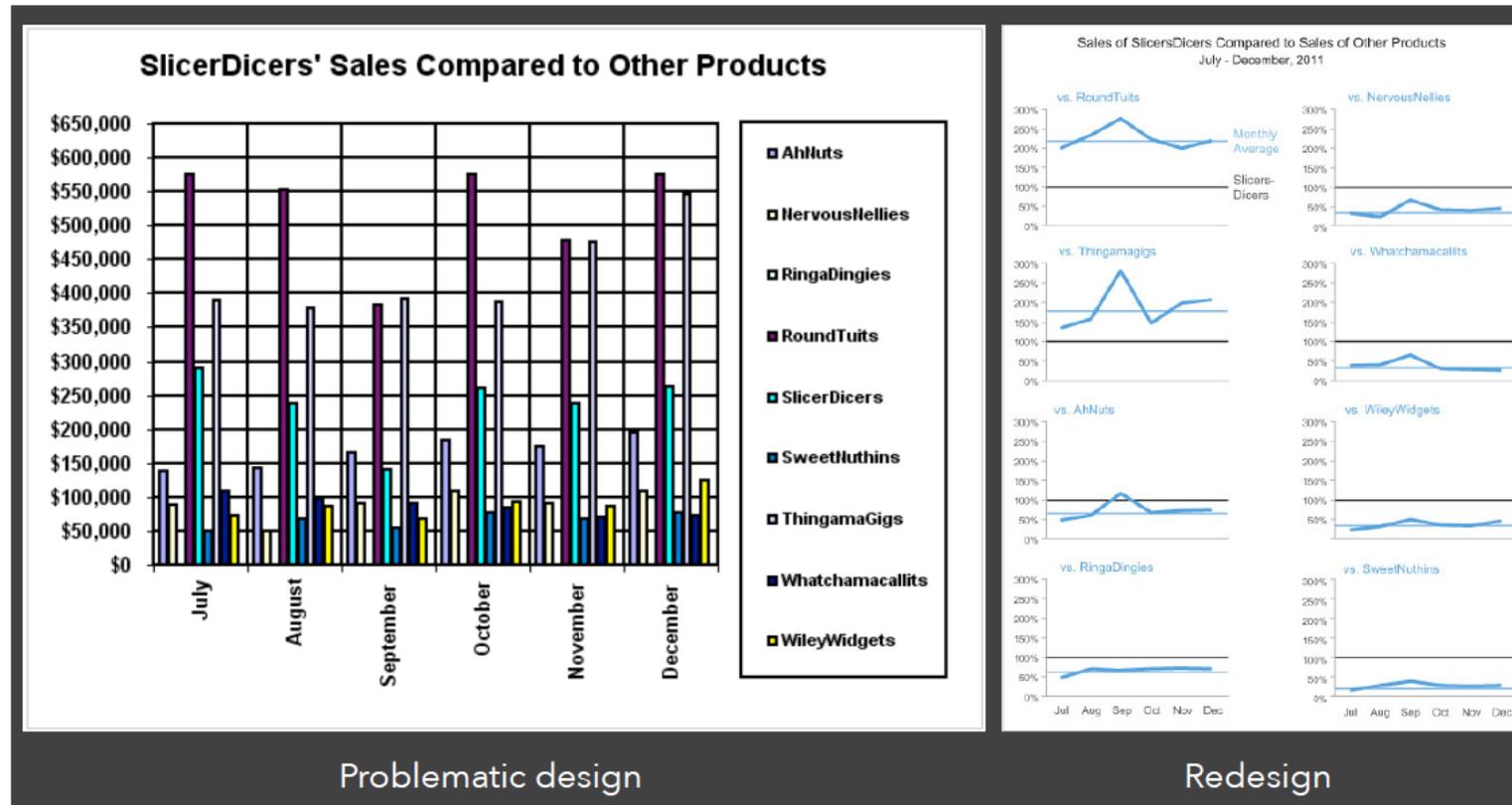
Temas del Curso de Visualización Científica

Análisis Exploratorio de Datos



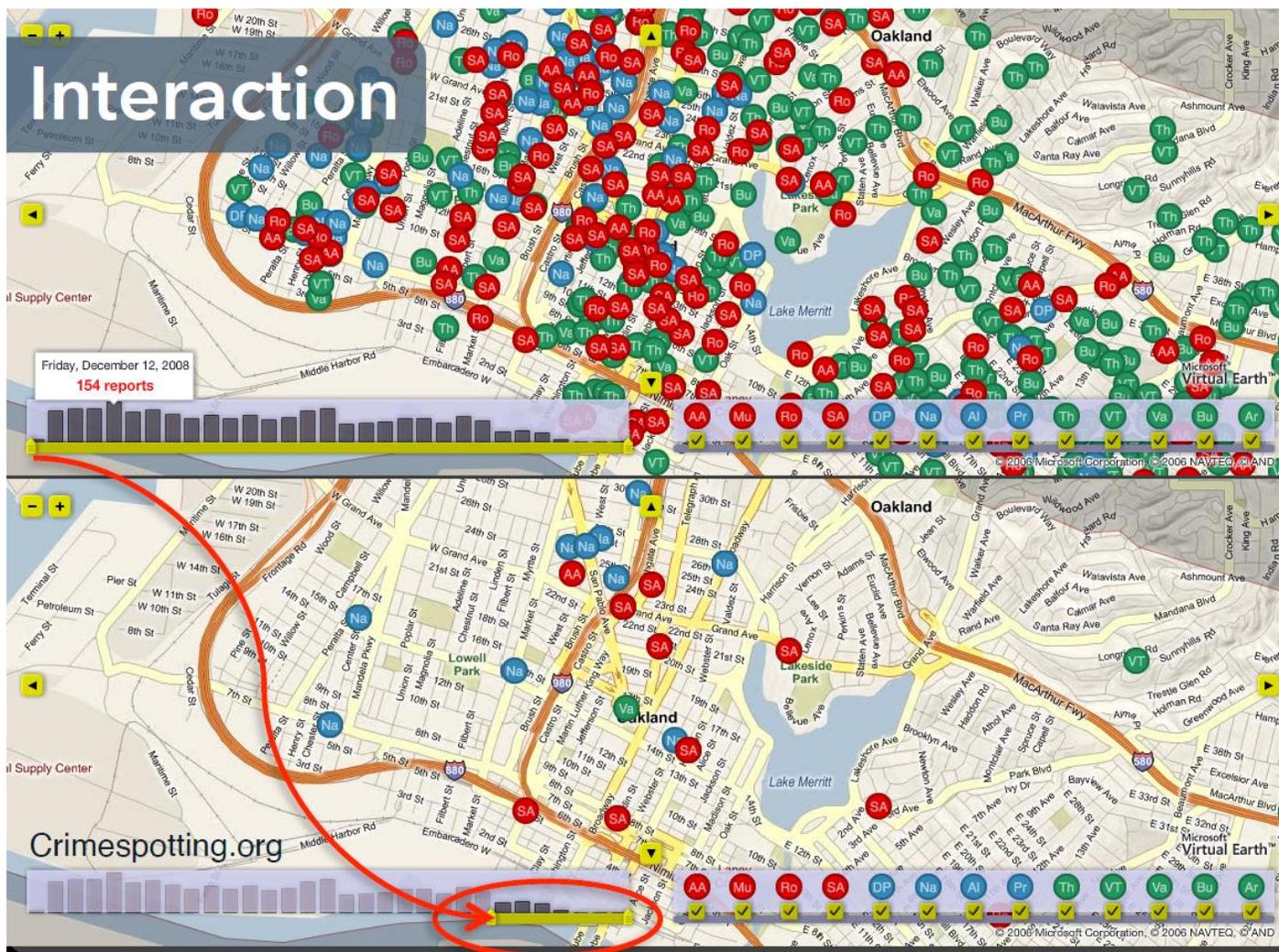
Temas del Curso de Visualización Científica

Diseño Visual



Temas del Curso de Visualización Científica

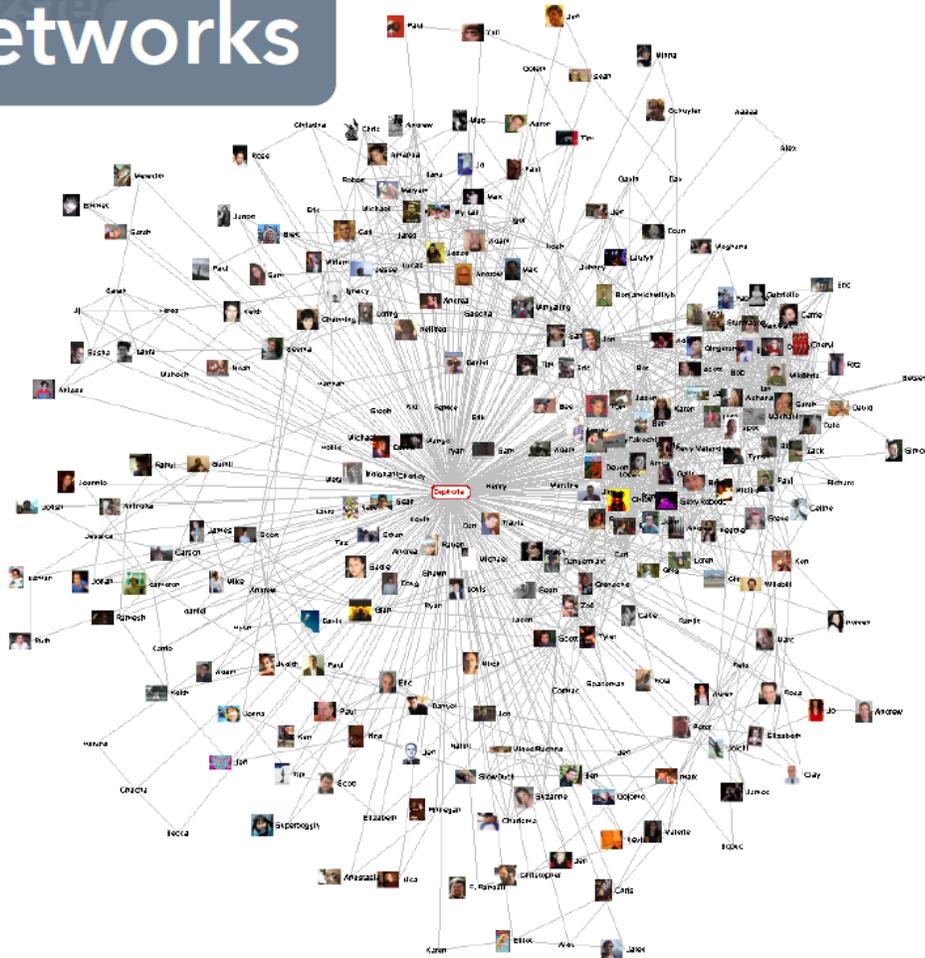
Interacción



Temas del Curso de Visualización Científica

Redes

Networks



community >>

Enable

search >>

✖