

# Visualización de Datos y Storytelling

Magíster en Data Science - Facultad de Ingeniería UDD



Carlos Elías Pérez Pizarro

Agradecimientos: Eduardo Graells-Garrido

# Método Científico

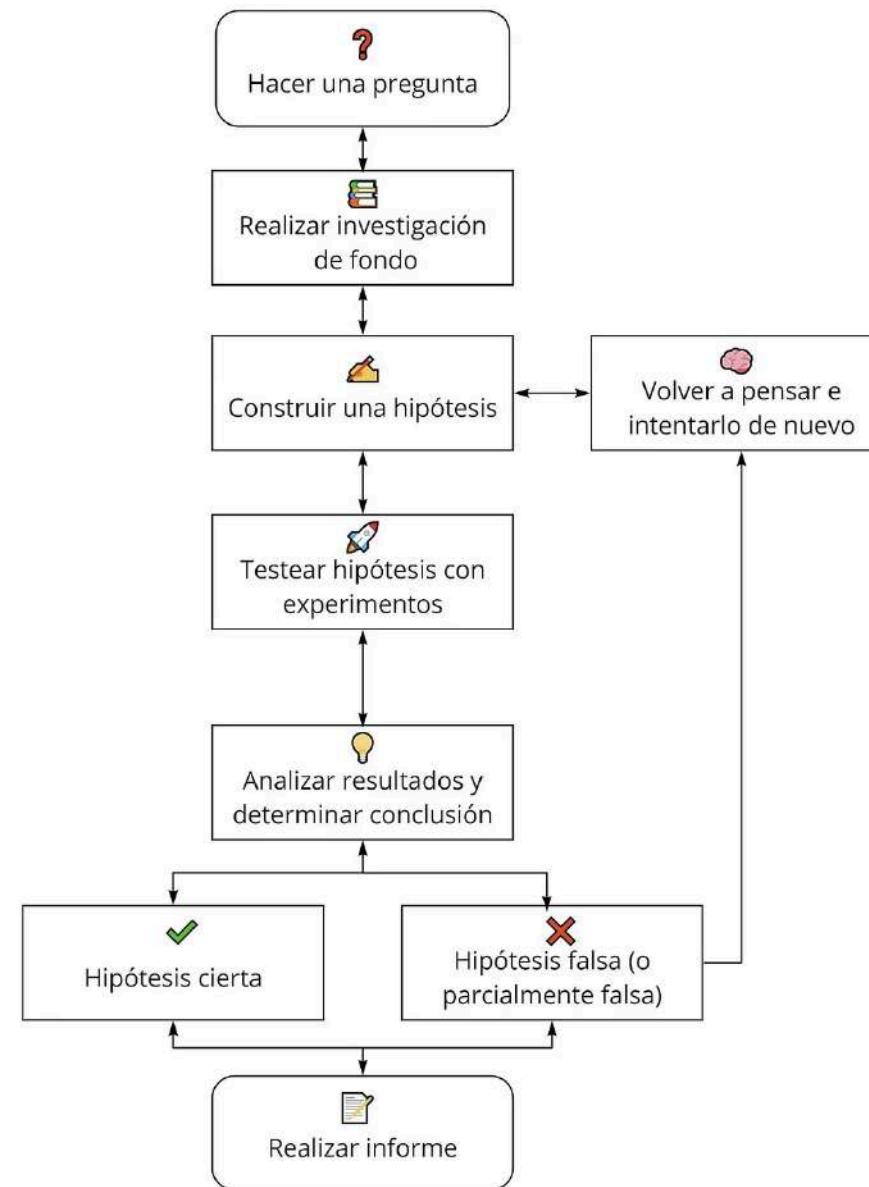
Más que el conocimiento en sí mismo,

Asociamos a la ciencia con un método

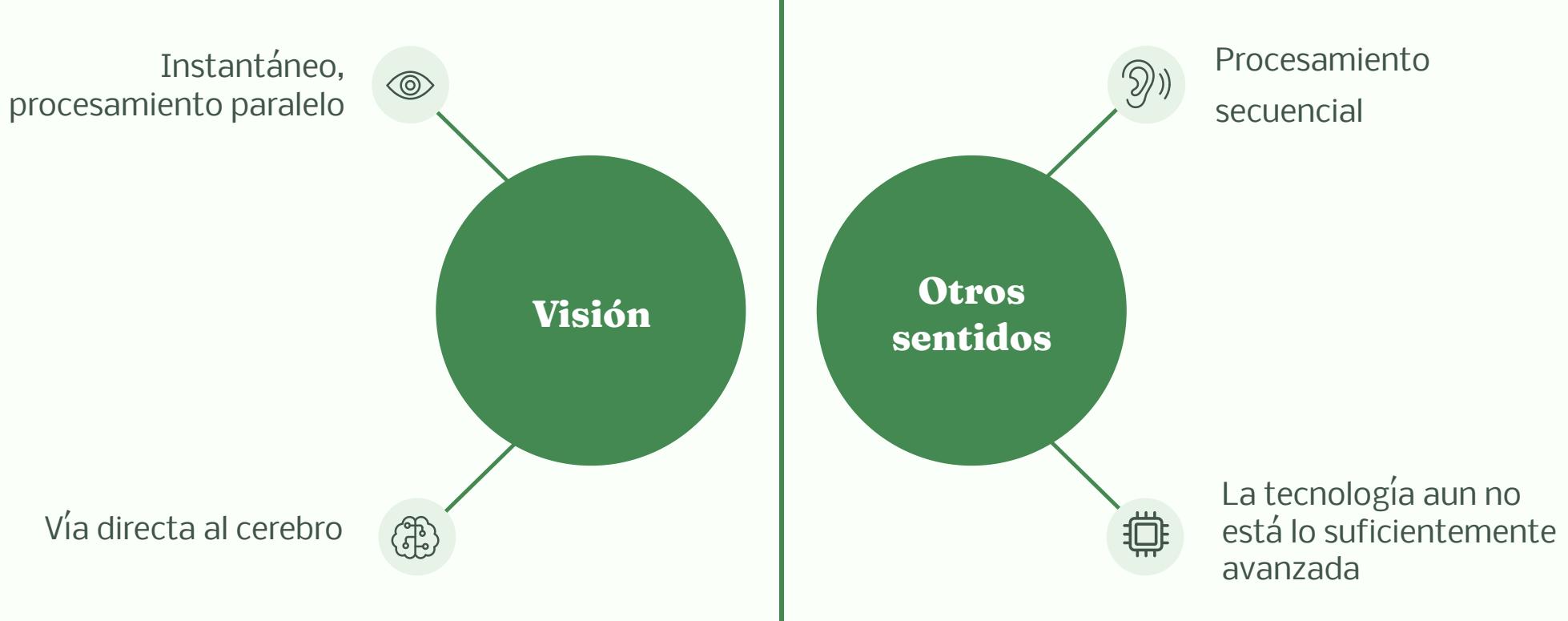
Para **conseguir dicho conocimiento**

Noten que **el método científico ya incluye datos**: son parte de los experimentos y del análisis de resultados

Modelo simplificado de las etapas del método científico

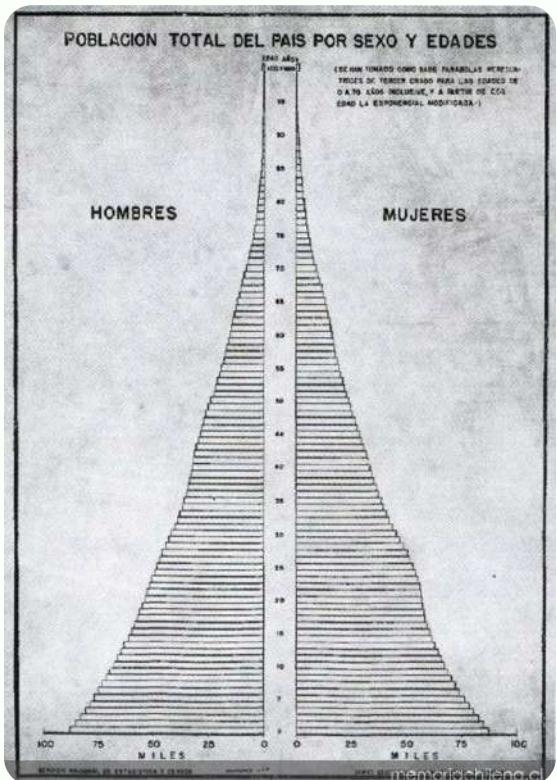


# ¿Por qué depender de la visión?



# Datos no son algo "nuevo"

## **Censos de 1865 y 1940**



REPÚBLICA DE CHILE.

REPÚBLICA DE CHILE.

**Resumen de las profesiones en toda la República, con especificación del sexo.**

P R O F E S I O N E S .	H O M B R E S .	M U J E R E S .
e pianos.	2	
e juegos pirotecnicos.	75	
e fideos.	97	
e pólvora.	516	66
e tejidos.	4	
o jarcia.	106	10
	25	
	206	
	18	
	5	
	16	
	4	
	2	
	2	
	958	
	55	
	4	21
	9	
	63	
	97	33
	375	
	178766	189
	11	
	32	
	3561	
	150	
	115	37400
	418	
	148	15
	39	1
	32	
	84	9
	101	
	159	3
	199	28
	3533	
	82	
	1	28904
	360	6
	2	
	659	13
	9	
	230	2726
	3812	
	453	
		468,
	8	
	81	
	239	
	23625	118
	7081	
culto disidente.	1	467
	866	
		107
	520	310
	3	

# Siempre visualiza tus datos

*“Sin datos, eres simplemente otra persona opinando”*

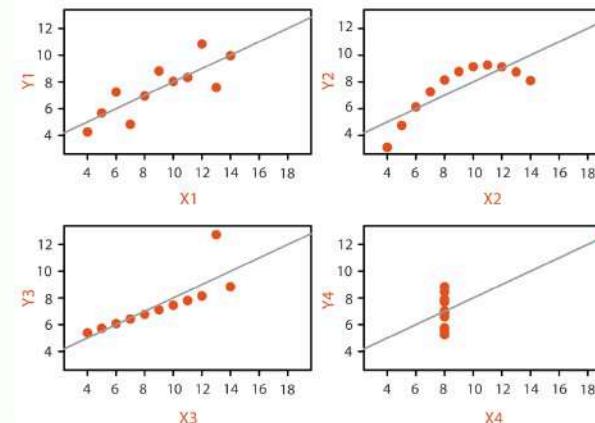
- W. Edwards Deming

## Dataset

Un dataset es un conjunto de observaciones o registros que contiene variables y atributos, usualmente numéricos.

Derecha: **El Cuarteto de Anscombe**.  
Presenta 4 subdatasets que tienen las mismas propiedades estadísticas básicas.

	1		2		3		4	
	X	Y	X	Y	X	Y	X	Y
	10.0	8.04	10.0	9.14	10.0	7.46	8.0	6.58
	8.0	6.95	8.0	8.14	8.0	6.77	8.0	5.76
	13.0	7.58	13.0	8.74	13.0	12.74	8.0	7.71
	9.0	8.81	9.0	8.77	9.0	7.11	8.0	8.84
	11.0	8.33	11.0	9.26	11.0	7.81	8.0	8.47
	14.0	9.96	14.0	8.10	14.0	8.84	8.0	7.04
	6.0	7.24	6.0	6.13	6.0	6.08	8.0	5.25
	4.0	4.26	4.0	3.10	4.0	5.39	19.0	12.50
	12.0	10.84	12.0	9.13	12.0	8.15	8.0	5.56
	7.0	4.82	7.0	7.26	7.0	6.42	8.0	7.91
	5.0	5.68	5.0	4.74	5.0	5.73	8.0	6.89
Mean	9.0	7.5	9.0	7.5	9.0	7.5	9.0	7.5
Variance	10.0	3.75	10.0	3.75	10.0	3.75	10.0	3.75
Correlation	0.816		0.816		0.816		0.816	



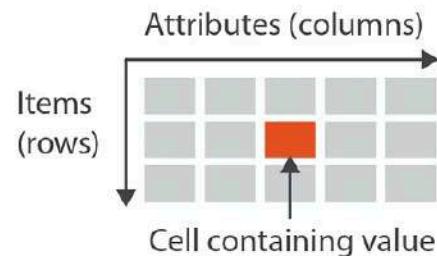
# Datasaurus dozen



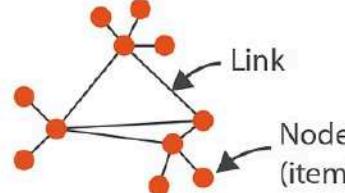
Dataset	Summary statistics					Regression results	
	Mean x	Mean y	Std Dev x	Std Dev y	Corr x,y	Intercept	Coefficients
Away	54.27	47.83	16.77	26.94	-0.06	53.43	-0.10
Bullseye	54.27	47.83	16.77	26.94	-0.07	53.81	-0.11
Circle	54.27	47.84	16.76	26.93	-0.07	53.80	-0.11
Dots	54.26	47.83	16.77	26.93	-0.06	53.45	-0.10
Dots	54.26	47.84	16.77	26.93	-0.06	53.10	-0.10
H_lines	54.26	47.83	16.77	26.94	-0.06	53.21	-0.10
High_lines	54.27	47.84	16.77	26.94	-0.07	53.81	-0.11
Slant_down	54.27	47.84	16.77	26.94	-0.07	53.85	-0.11
Slant_up	54.27	47.83	16.77	26.94	-0.07	53.81	-0.11
Star	54.27	47.84	16.77	26.93	-0.06	53.33	-0.10
V_lines	54.27	47.84	16.77	26.94	-0.07	53.89	-0.11
Wide_lines	54.27	47.83	16.77	26.94	-0.07	53.63	-0.11
X_shape	54.26	47.84	16.77	26.93	-0.07	53.55	-0.11

# Tipos de Datos

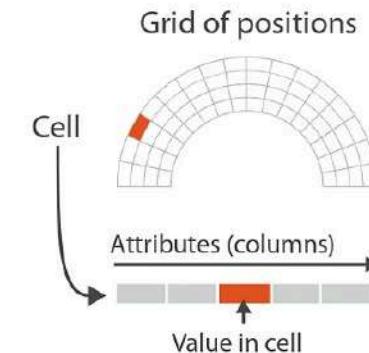
→ Tables



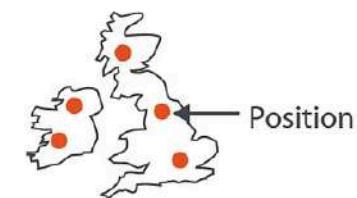
→ Networks



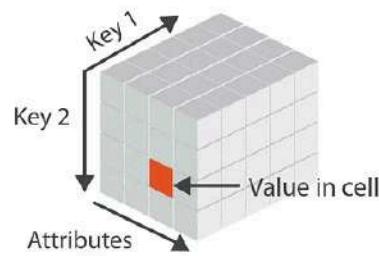
→ Fields (Continuous)



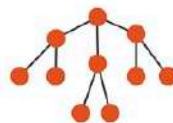
→ Geometry (Spatial)



→ Multidimensional Table

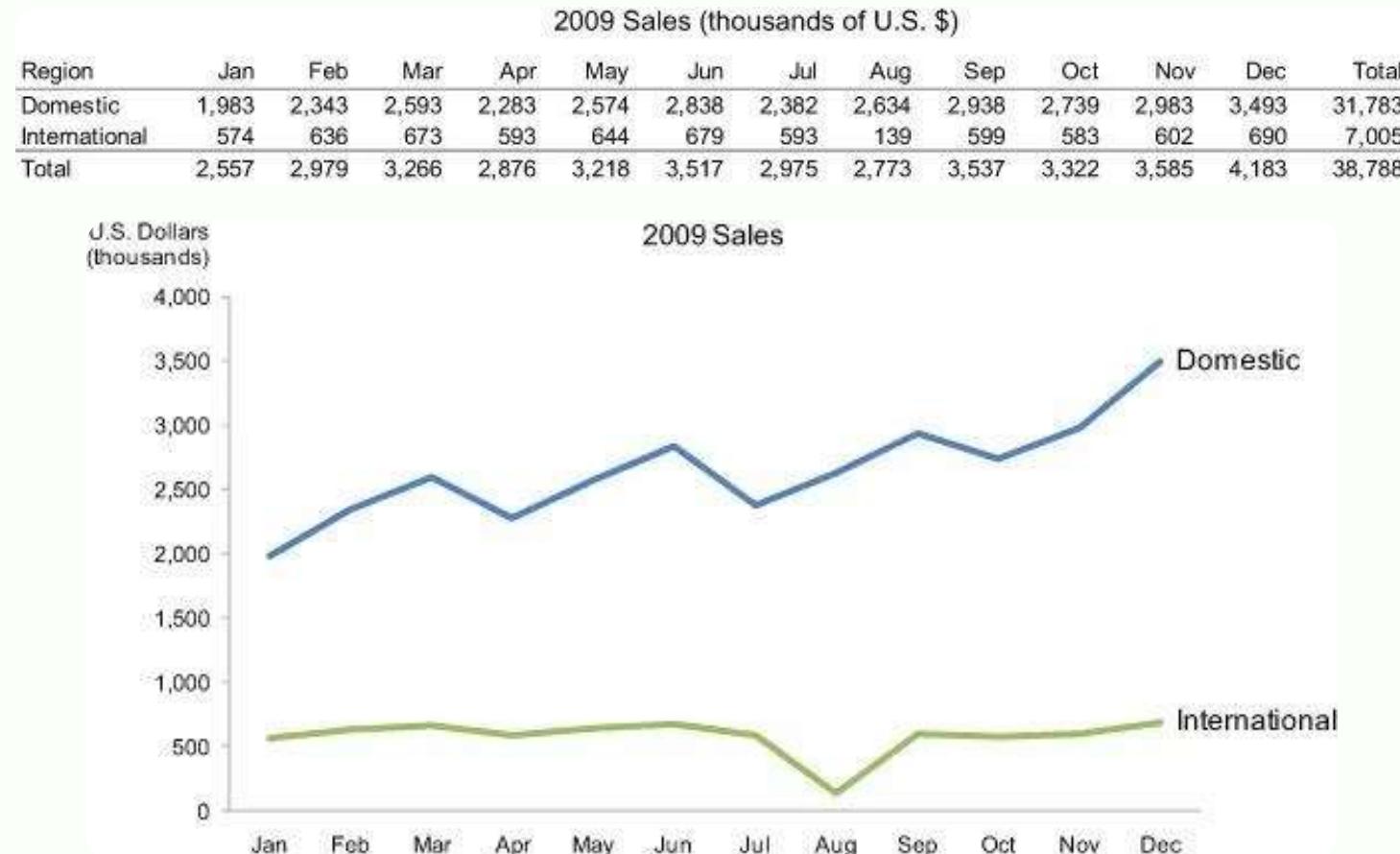


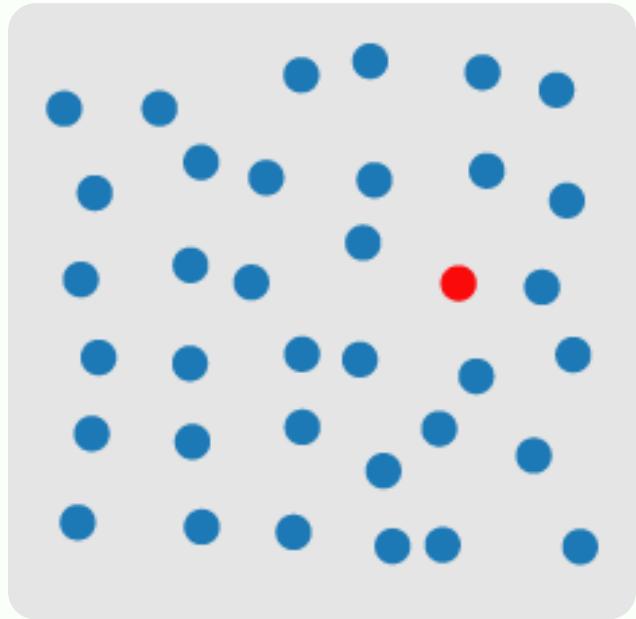
→ Trees



# Nos da superpoderes

Una visualización nos puede organizar la información de manera que podamos realizar consultas de manera eficiente.





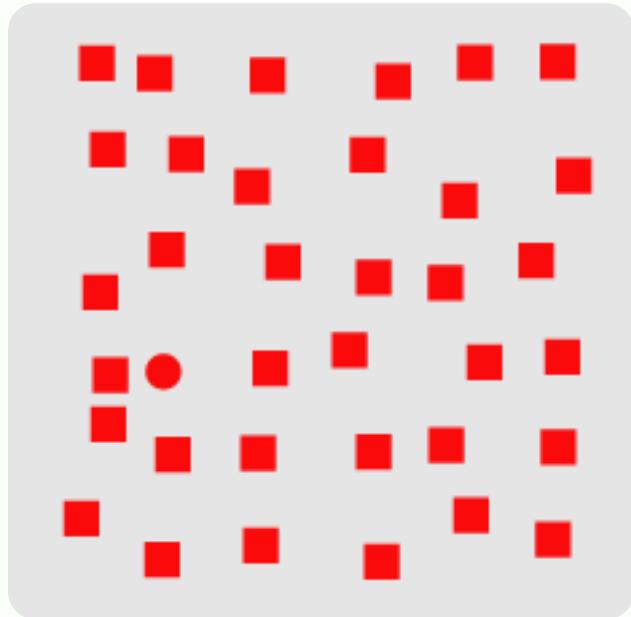
# ¿Dónde está el punto rojo?

Fuente:

Perception in Visualization

Christopher G. Healy

<https://www.csc2.ncsu.edu/faculty/healey/PP/>



# ¿Dónde está el punto rojo?

Fuente:

Perception in Visualization

Christopher G. Healy

<https://www.csc2.ncsu.edu/faculty/healey/PP/>

# El mundo es una traza digital

## Trazas Digitales

transbank.  
APOYANDO NEGOCIOS

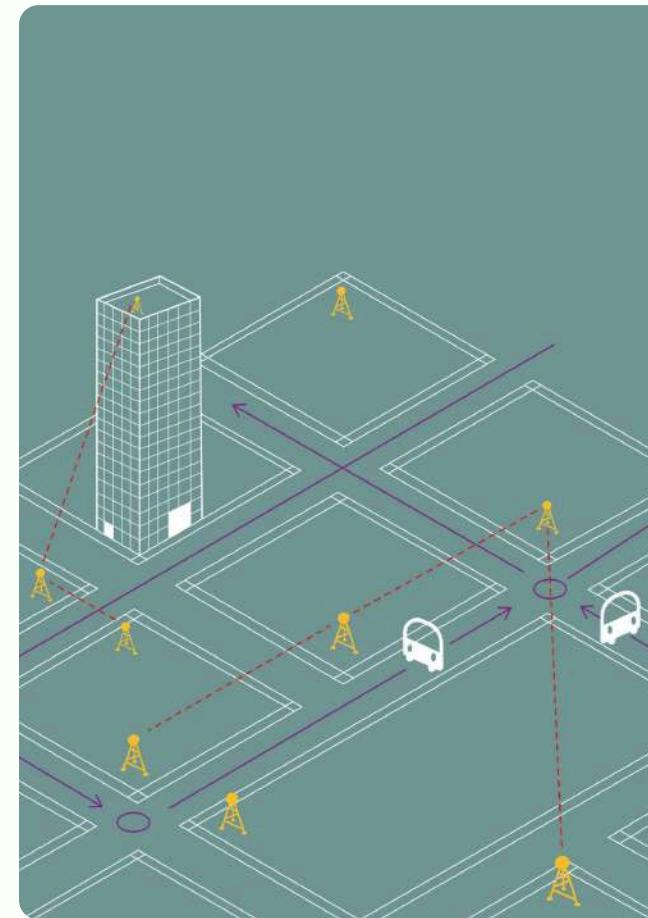


Cornershop



Google

# Mejora nuestro entendimiento del mundo



# Muchos datos?

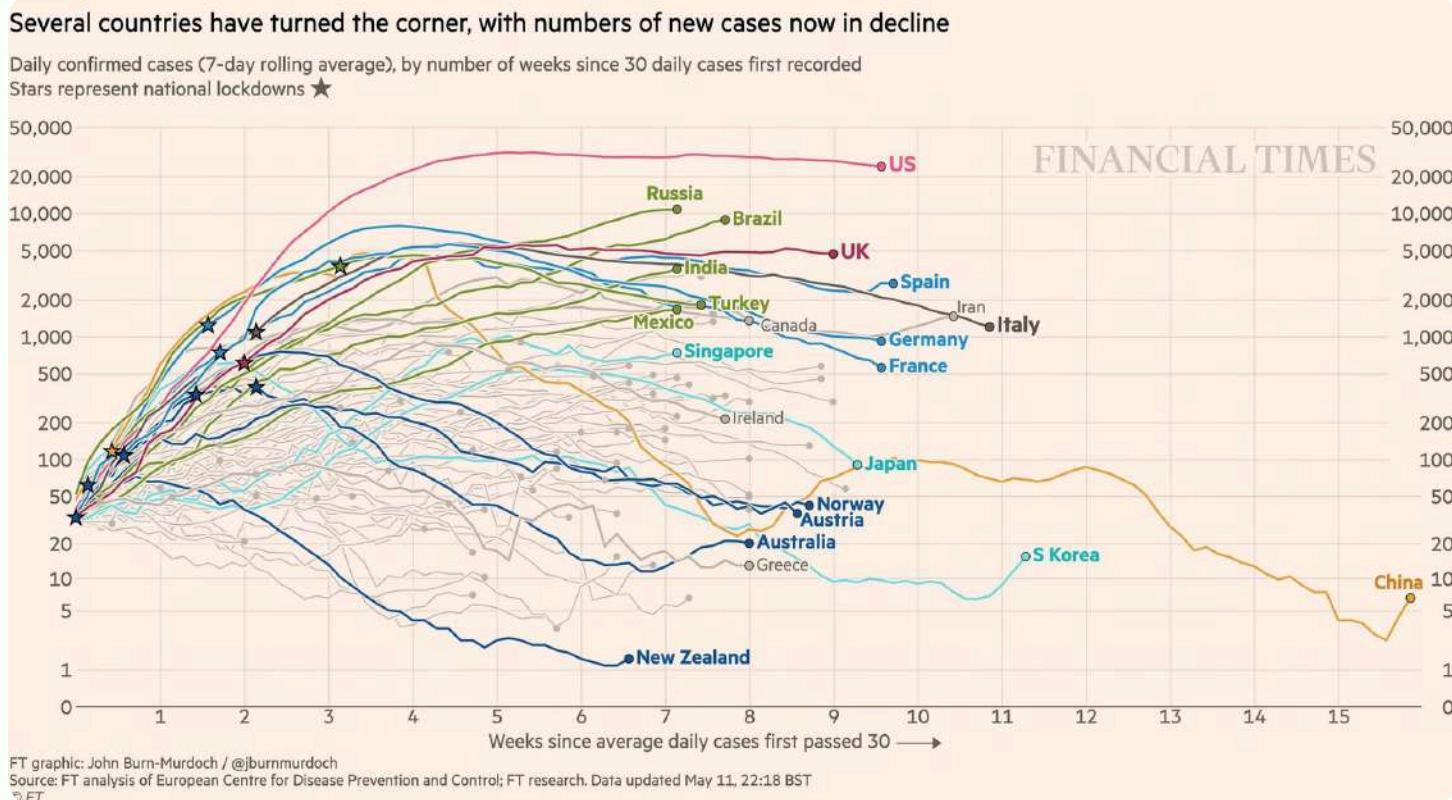


Gráfico por John Burn-Murdoch sobre la evolución del COVID-19.