

# Visualización de Datos

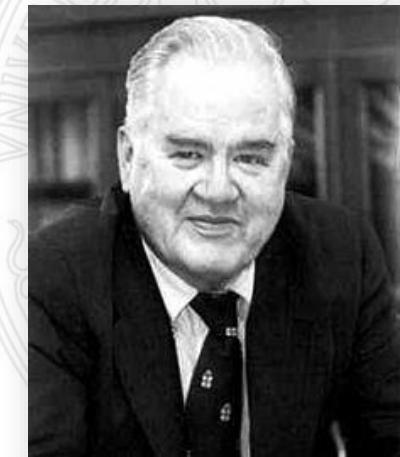
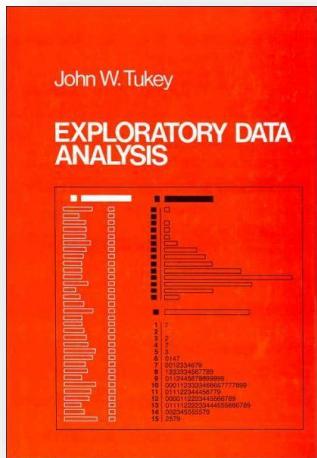


UNIVERSIDAD DE  
**MURCIA**

# Tema 1. Introducción a la visualización de datos

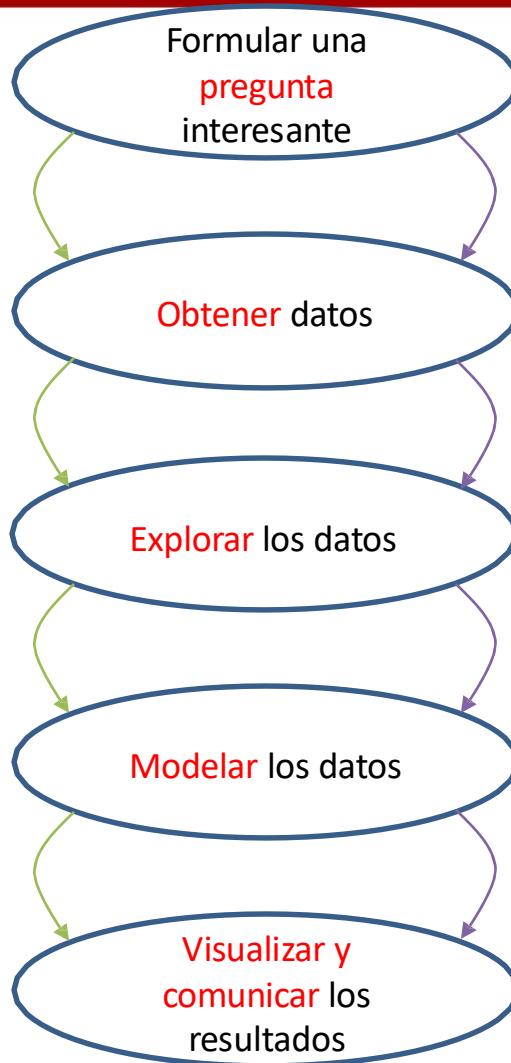
“The greatest value of a picture is when it forces us to notice what we never expected to see”

— John W. Tukey, *Exploratory Data Analysis*, 1977



- Exploración
  - **Problema:** Disponemos de datos, pero un escaso conocimiento sobre su comportamiento
  - **Tarea:** Generar hipótesis a partir de su visualización
- Comunicación
  - **Problema:** Disponemos de una serie de conocimientos sobre los datos
  - **Tarea:** Presentarlos visualmente de una manera comprensiva e intuitiva
- Confirmación
  - **Problema:** Se formula alguna hipótesis sobre el comportamiento que subyace a los datos
  - **Tarea:** Confirmar o rechazar hipótesis

# El método científico



¿Cuál es nuestra hipótesis?  
¿Cómo explicamos lo observado?  
¿Qué se pretende predecir o estimar?

¿Cómo obtenemos los datos? ¿Qué datos obtenemos?  
¿Qué datos son relevantes? ¿Qué hacemos con los que faltan o son defectuosos?

¿Cómo dibujamos los datos?  
¿Se observan patrones?  
¿Se observan anomalías?

Proponemos y construimos un modelo  
Ajustamos el modelo a los datos  
Validamos el modelo

¿Qué comportamientos subyacen en los datos?  
¿Qué relato construimos a partir de los datos?

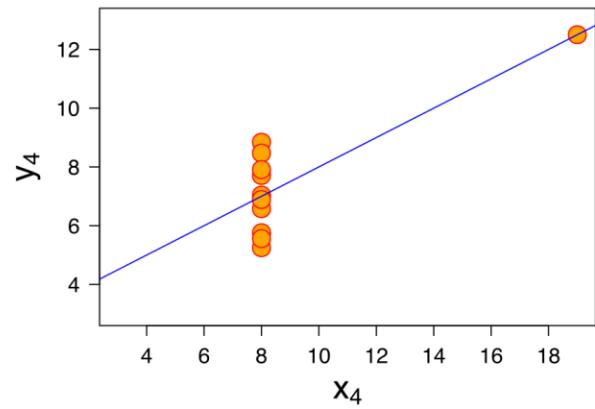
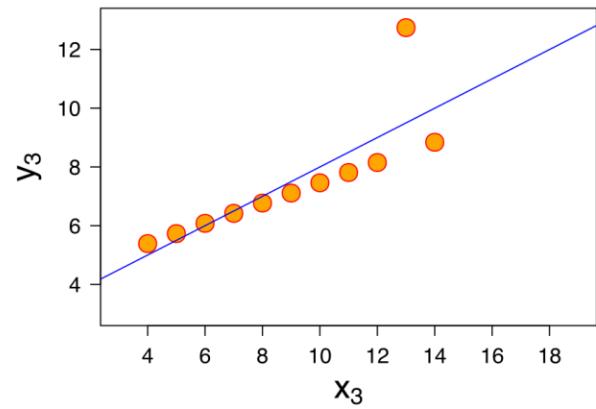
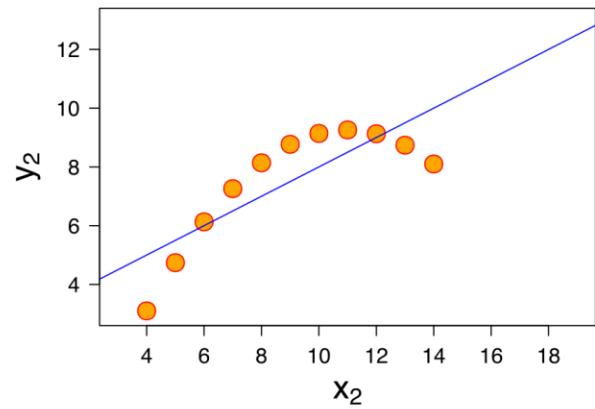
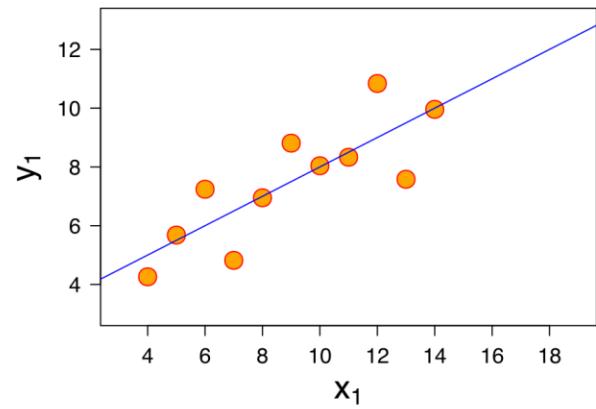
# El cuarteto de Anscombe (1/2)

- Cuatro conjuntos de datos con las **mismas propiedades estadísticas**

I		II		III		IV	
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

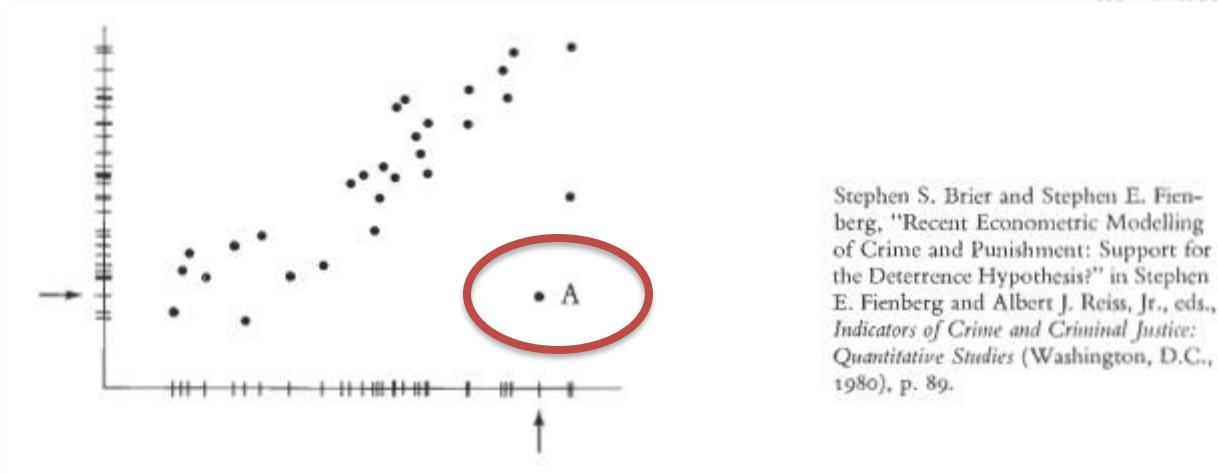
$N = 11$	$y = 3.00 + 0.500x$
$\bar{X} = 9.0$	$r = 0.816$
$\bar{Y} = 7.5$	$r^2 = 0.67$

## El cuarteto de Anscombe (2/2)



# La visualización revela información

- El punto A constituye claramente un punto fuera de rango, y sin embargo puede pasar inadvertido en las correspondientes distribuciones marginales

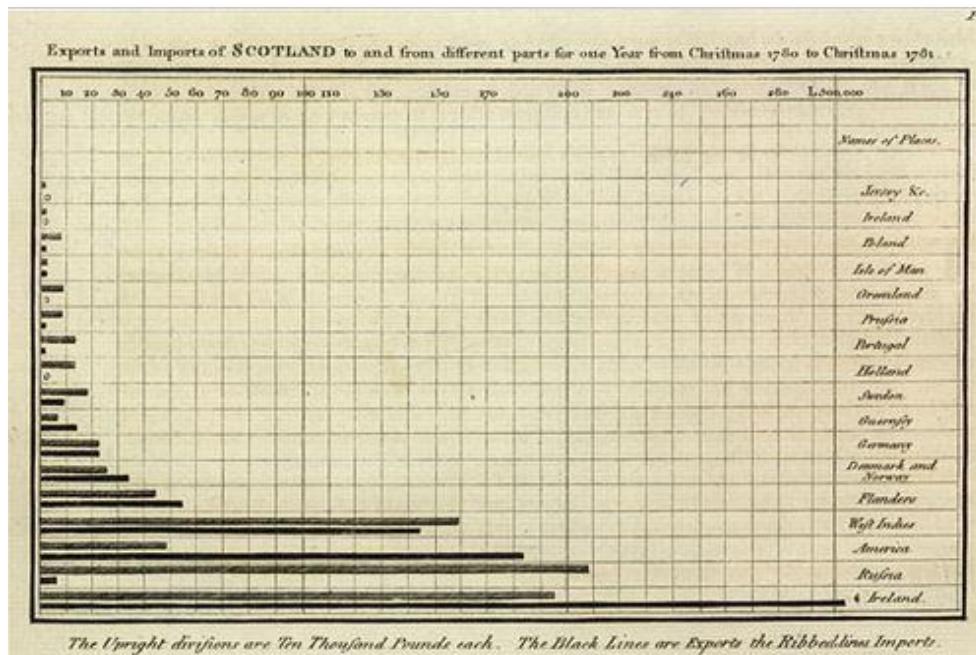


Stephen S. Brier and Stephen E. Fienberg, "Recent Econometric Modelling of Crime and Punishment: Support for the Deterrence Hypothesis?" in Stephen E. Fienberg and Albert J. Reiss, Jr., eds., *Indicators of Crime and Criminal Justice: Quantitative Studies* (Washington, D.C., 1980), p. 89.

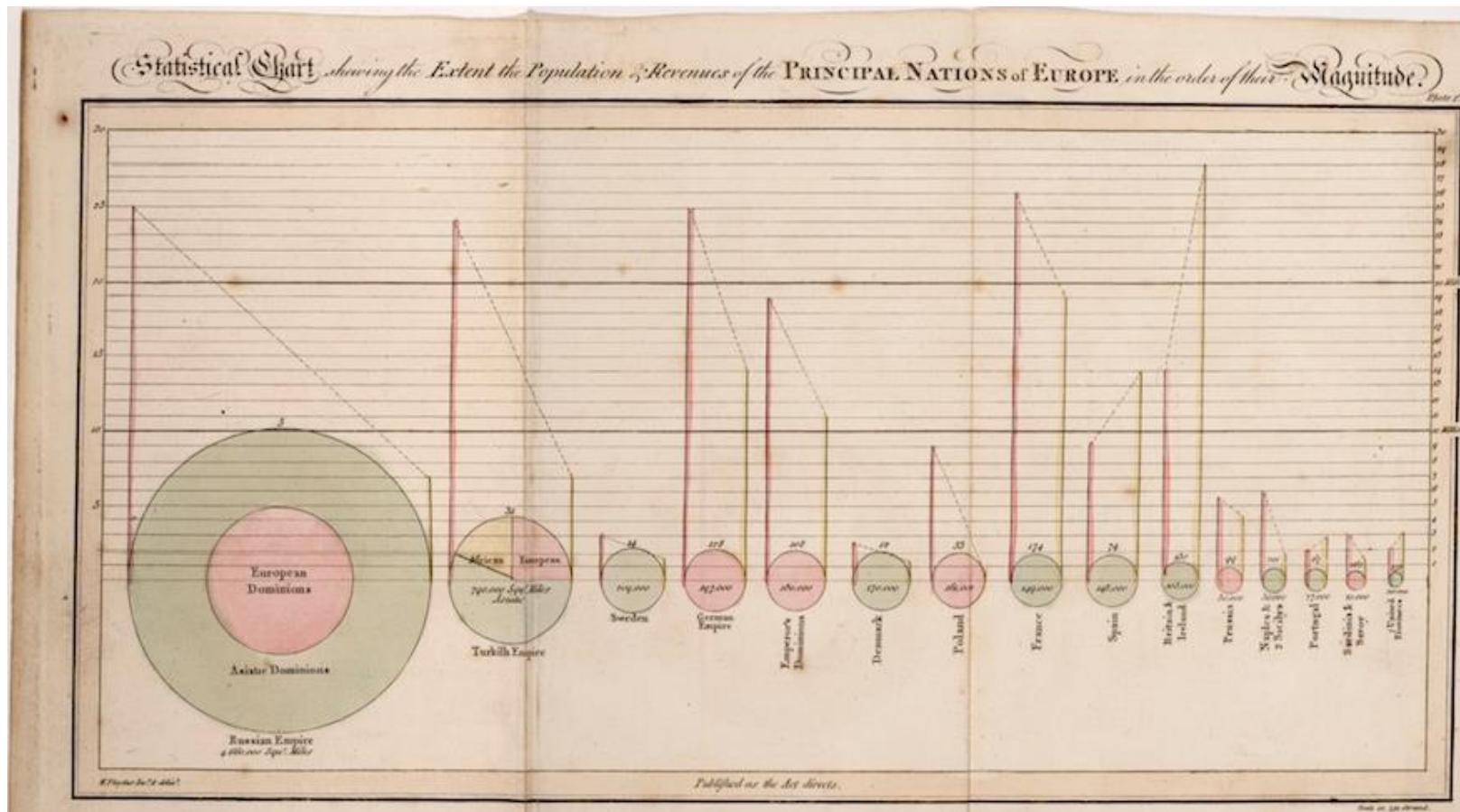
# Características de una representación gráfica

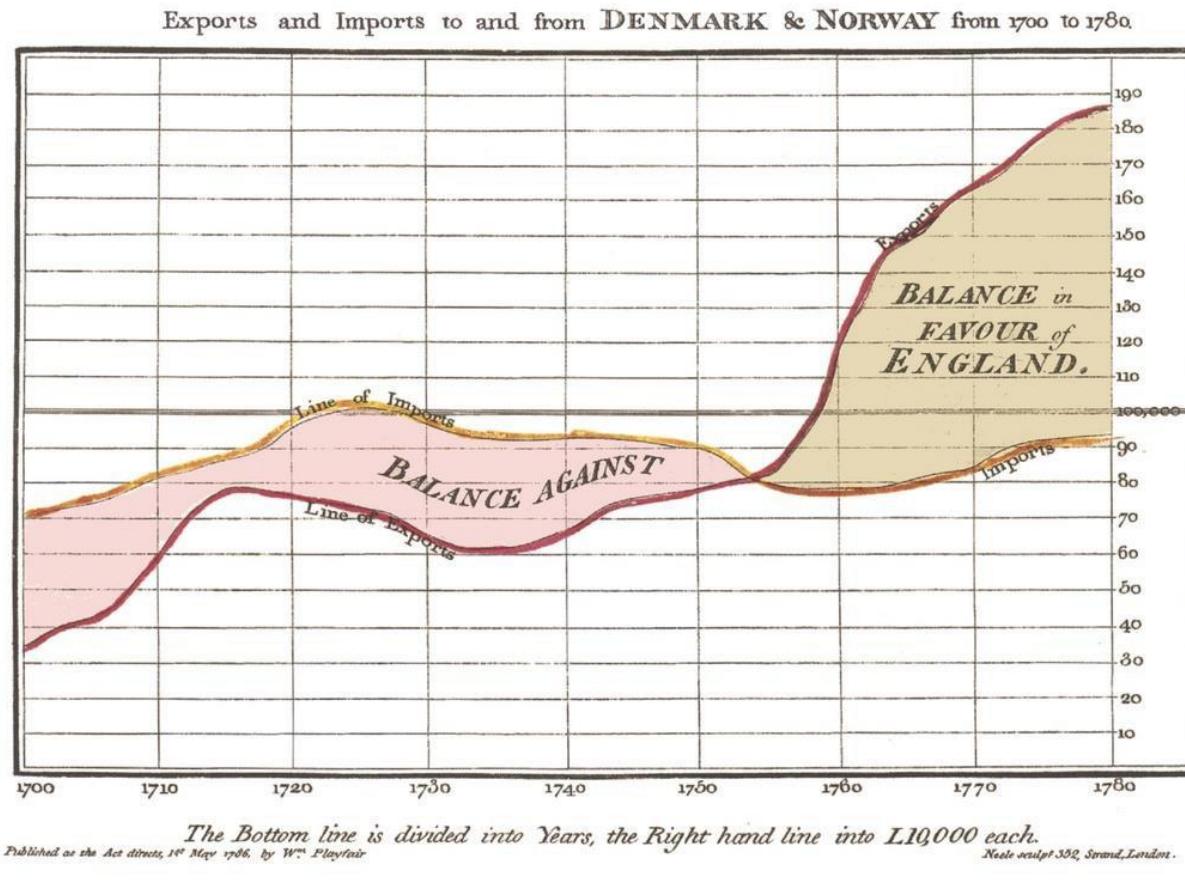
1. Mostrar los datos y ser **fiel** a los mismos
2. Tener un **objetivo claro**: describir, explorar, comparar, descubrir, etc.
3. Llevar al observador a **razonar sobre lo substancial**, no sobre la metodología de dibujo, o sobre tecnología informática, o cualquier otro asunto
4. Ser capaz de **sintetizar** un gran volumen de datos
5. **Enfatizar** aquello que se quiere expresar en los datos
6. **Revelar** los múltiples niveles de detalle de los datos
7. Estar íntimamente ligada a **otras descripciones** de los datos: estadísticas, verbales, etc.

- A William Playfair se atribuyen los primeros gráficos de carácter estadístico: gráficos de barras, gráficos de sectores y series temporales

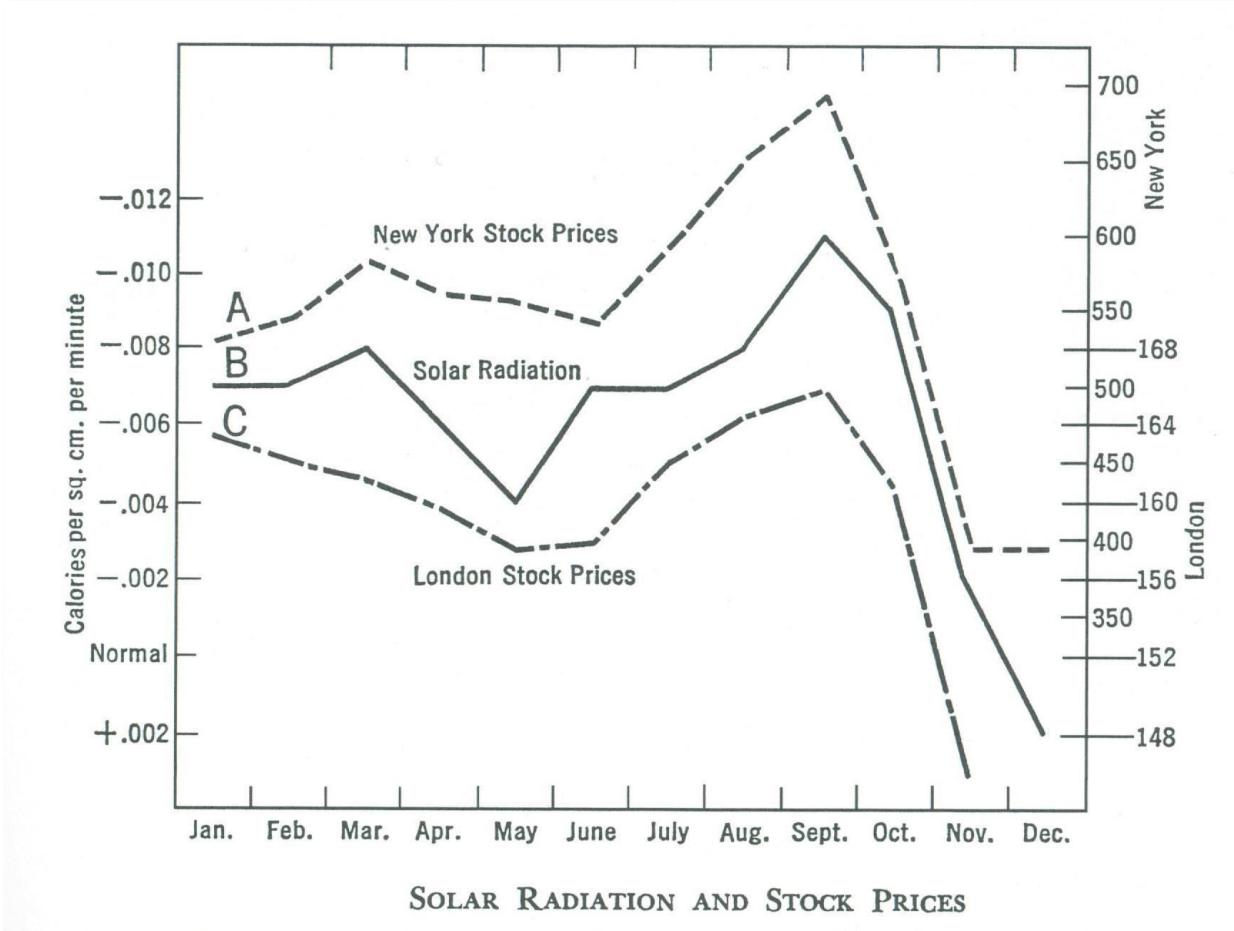


## Siglo XVIII (2/3)





# Si la hipótesis es absurda, su representación gráfica también lo es



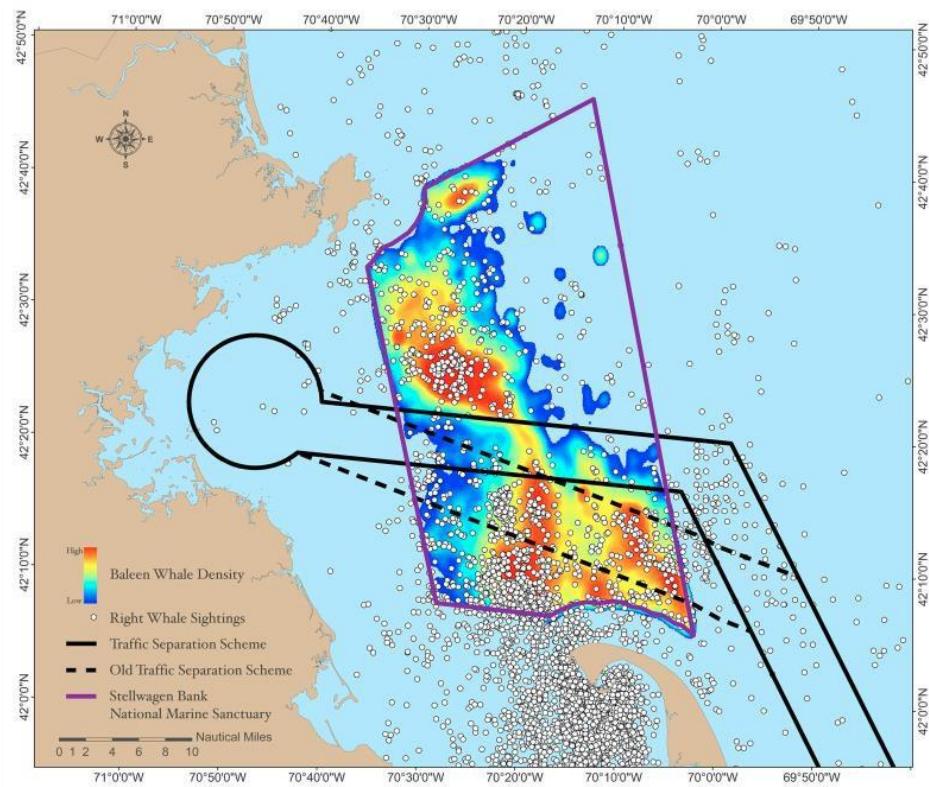
# Una nueva forma de razonar (1/4)

- Mapa realizado por el doctor John Snow durante la epidemia de cólera que sufrió Londres en 1854



## Una nueva forma de razonar (2/4)

- Desplazamiento del tráfico marino de Boston para evitar la colisión con ballenas barbadas



# Una nueva forma de razonar (3/4)

- En 1951, el famoso diseñador gráfico Will Burtin asumió el reto de representar gráficamente la eficacia de 3 antibióticos sobre 16 tipos de bacterias

Bacteria	Penicillin	Antibiotic Streptomycin	Neomycin	Gram stain
<i>Aerobacter aerogenes</i>	870	1	1.6	-
<i>Brucella abortus</i>	1	2	0.02	-
<i>Bacillus anthracis</i>	0.001	0.01	0.007	+
<i>Diplococcus pneumoniae</i>	0.005	11	10	+
<i>Escherichia coli</i>	100	0.4	0.1	-
<i>Klebsiella pneumoniae</i>	850	1.2	1	-
<i>Mycobacterium tuberculosis</i>	800	5	2	-
<i>Proteus vulgaris</i>	3	0.1	0.1	-
<i>Pseudomonas aeruginosa</i>	850	2	0.4	-
<i>Salmonella (Eberthella) typhosa</i>	1	0.4	0.008	-
<i>Salmonella schottmuelleri</i>	10	0.8	0.09	-
<i>Staphylococcus albus</i>	0.007	0.1	0.001	+
<i>Staphylococcus aureus</i>	0.03	0.03	0.001	+
<i>Streptococcus fecalis</i>	1	1	0.1	+
<i>Streptococcus hemolyticus</i>	0.001	14	10	+
<i>Streptococcus viridans</i>	0.005	10	40	+

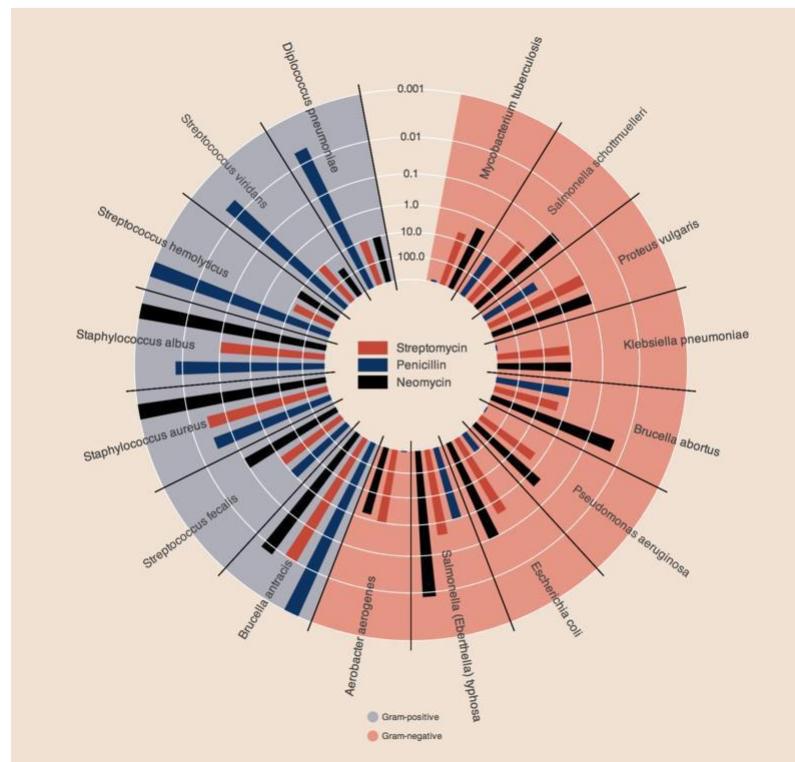
Concentración mínima inhibitoria

Grupo

# Una nueva forma de razonar (4/4)

- Este gráfico responde a una pregunta:  
**¿Cómo comparar la efectividad de los antibióticos?**

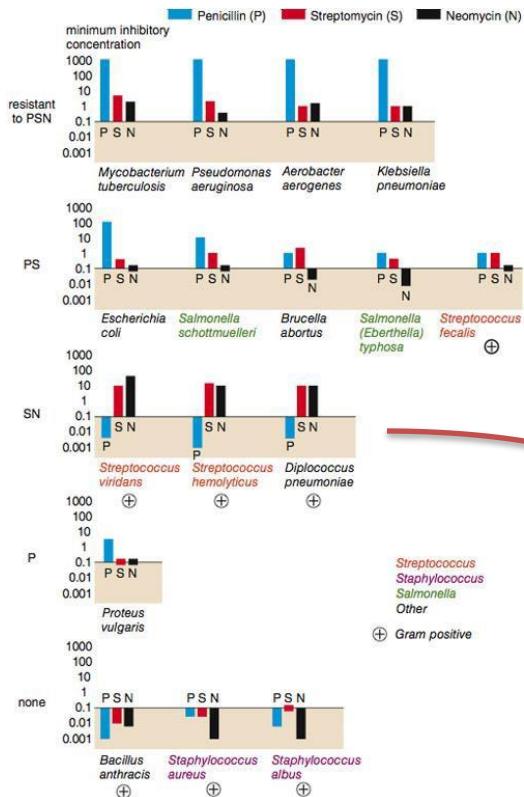
Si la bacteria es **Gram positiva**, entonces es más eficaz una combinación de *Penicilina* y *Estreptomicina*



Si la bacteria es **Gram negativa**, entonces la *Neomicina* es más eficaz

# Hagamos otras preguntas

- Este gráfico responde a otra pregunta:  
**¿Cómo agrupar las bacterias?**



Se observó que el *Streptococcus fecalis* no obedecía a algunas características de los estreptococos, y acabó cambiando su nombre en 1984 por *Enterococcus faecalis*

Se observó que el *Diplococcus pneumoniae* se comportaba como un estreptococo, y en 1974 se cambió su nombre por *Streptococcus pneumoniae*

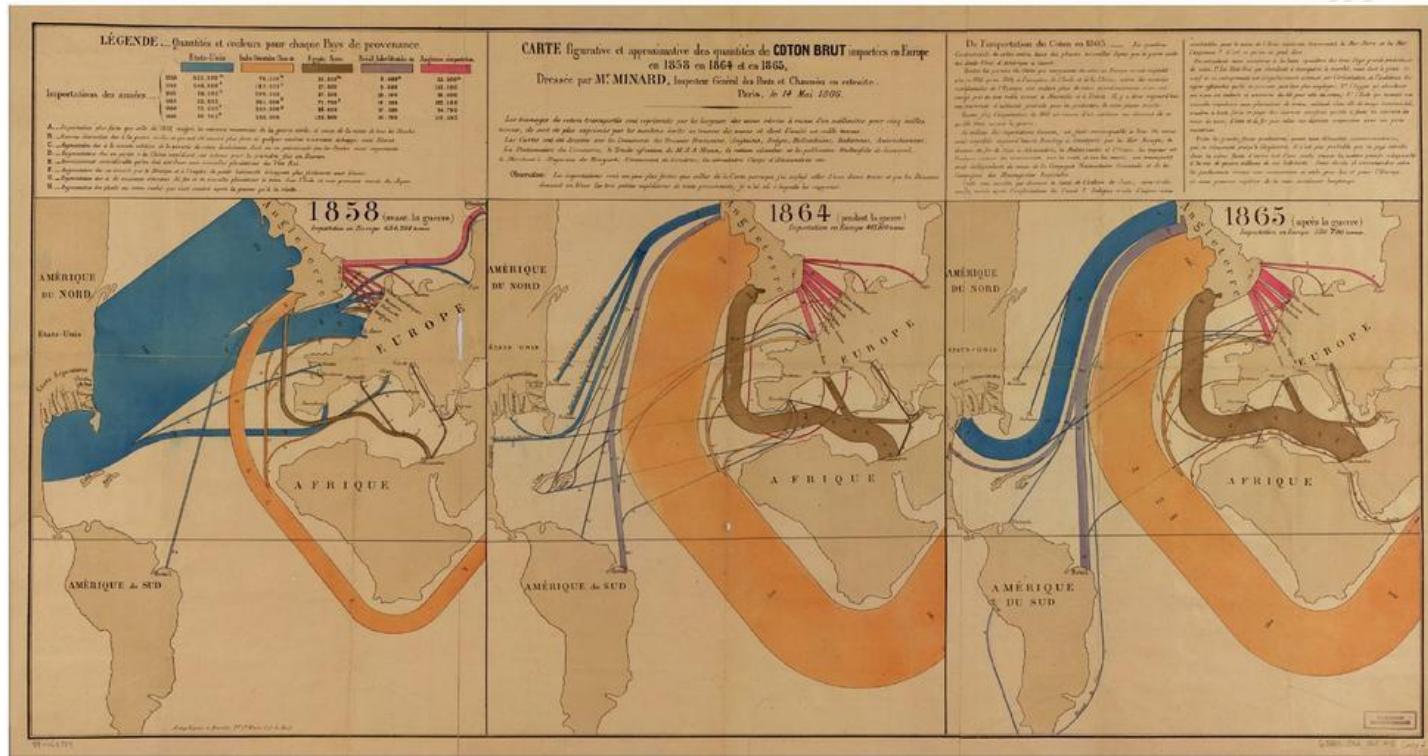
# Un gráfico, una pregunta



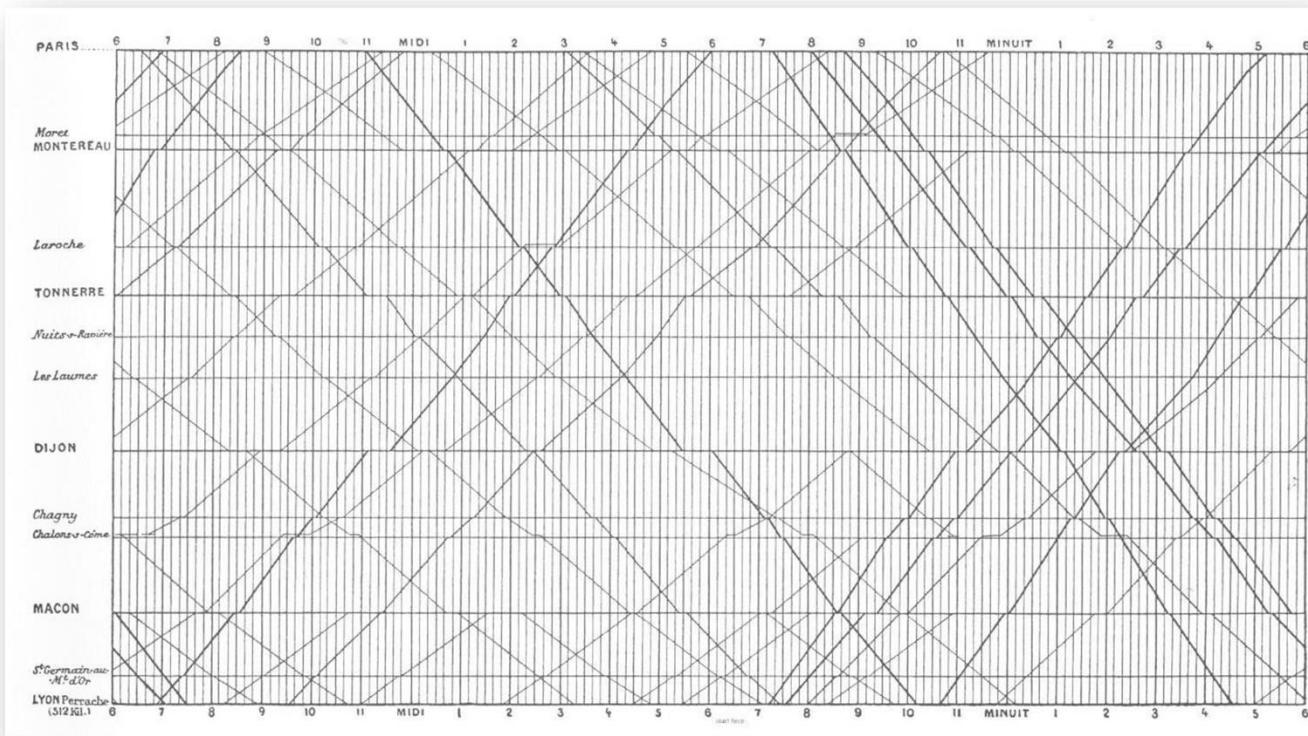
Gráficos de Jacques Bertin para exemplificar la presentación de información gráfica de diferentes maneras para diferentes propósitos

El mapa de la parte superior muestra los productos básicos producidos en cada estado. La matriz de abajo muestra los estados que producen cada mercancía

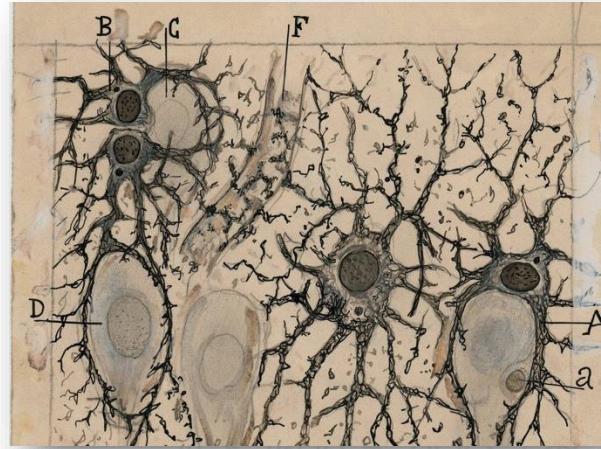
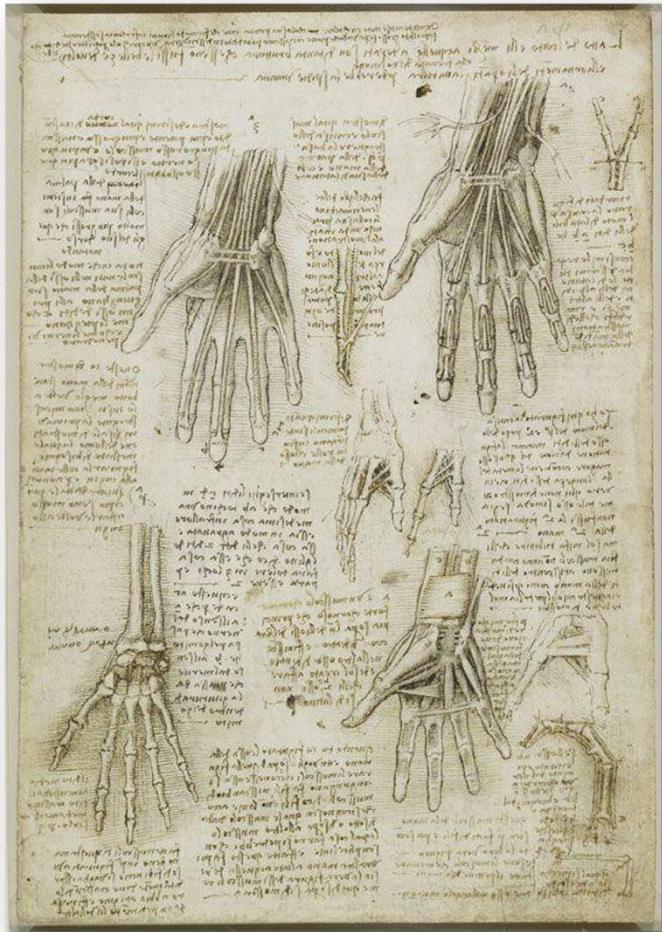
- Mapas del francés C.J. Minard para representar las importaciones de algodón en Europa en distintas fechas



- Horario de trenes de París a Lyon de 1880, publicado por E.J. Marey



# Visualizar para comprender

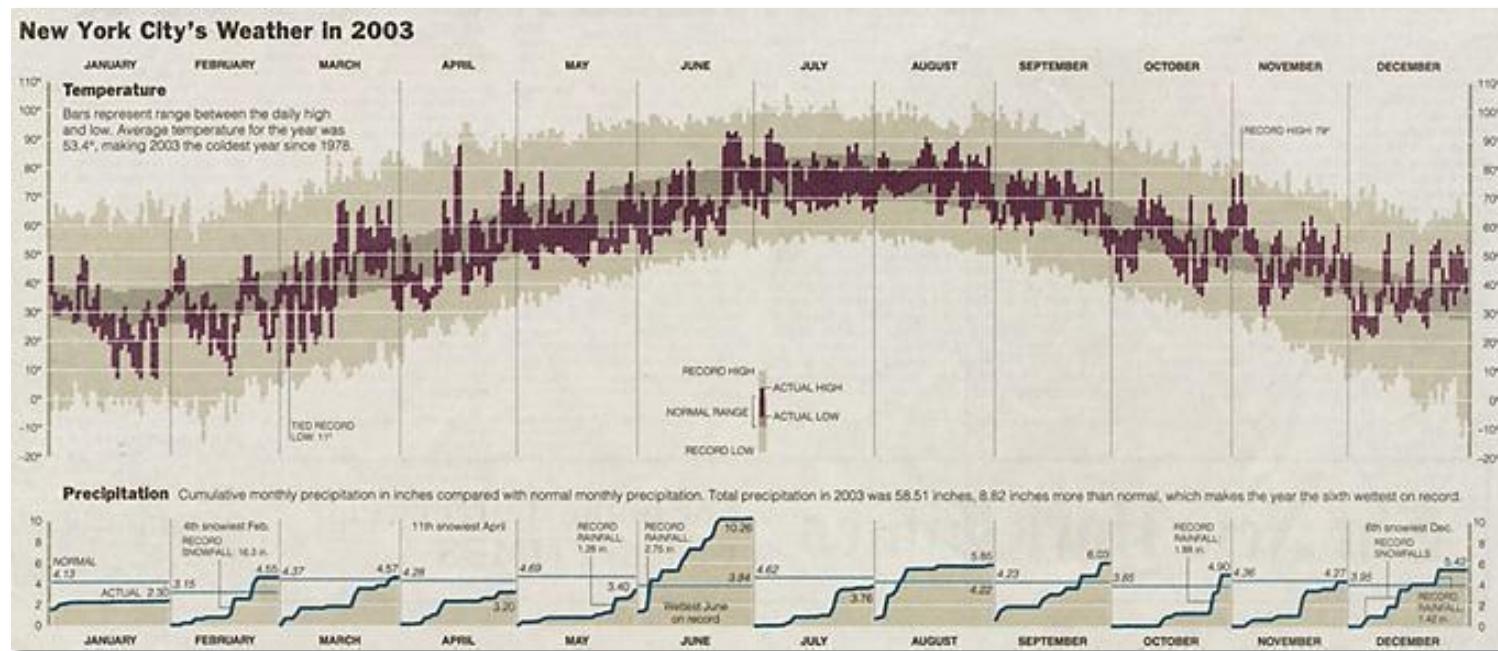


Cajal resume muchas de las propiedades de los astrocitos en este dibujo del hipocampo de una persona

Leonardo da Vinci realizó en el siglo XV numerosos estudios de anatomía que incluían ilustraciones a las que Leonardo llamó “demostraciones”

# Representando miles de datos (1/3)

- New York Times, 4 de enero de 2004



## Representando miles de datos (2/3)

UNIVERSIDAD DE  
**MURCIA**

- Proyecto del *conectoma* humano
    - Busca realizar el mapa de conexiones de los sistemas neuronales del cerebro a partir del flujo de moléculas de agua en imagen de resonancia magnética

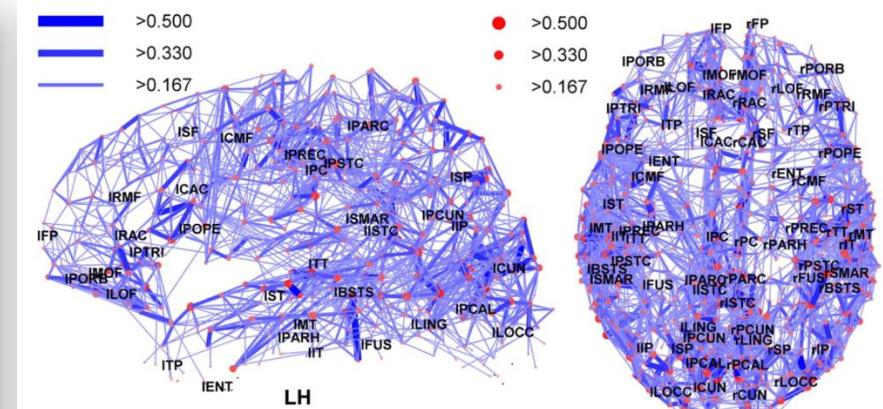
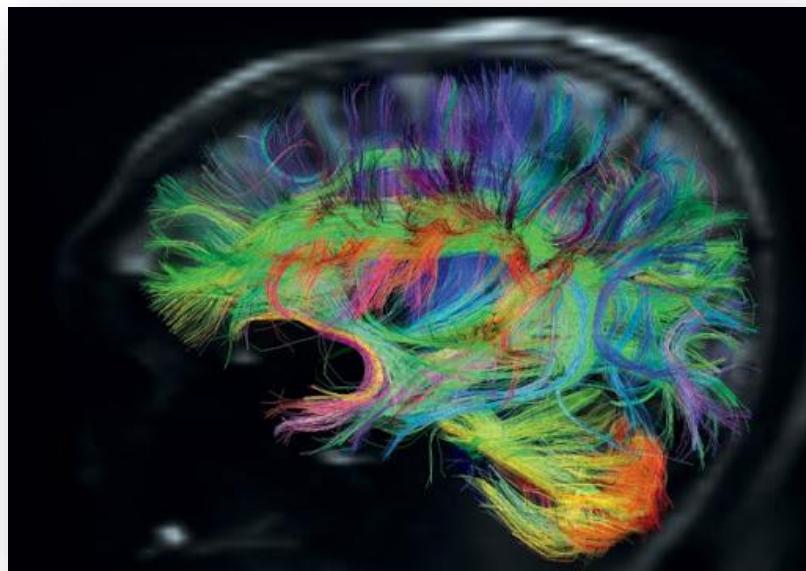
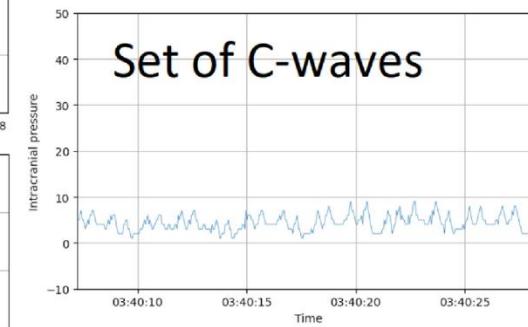
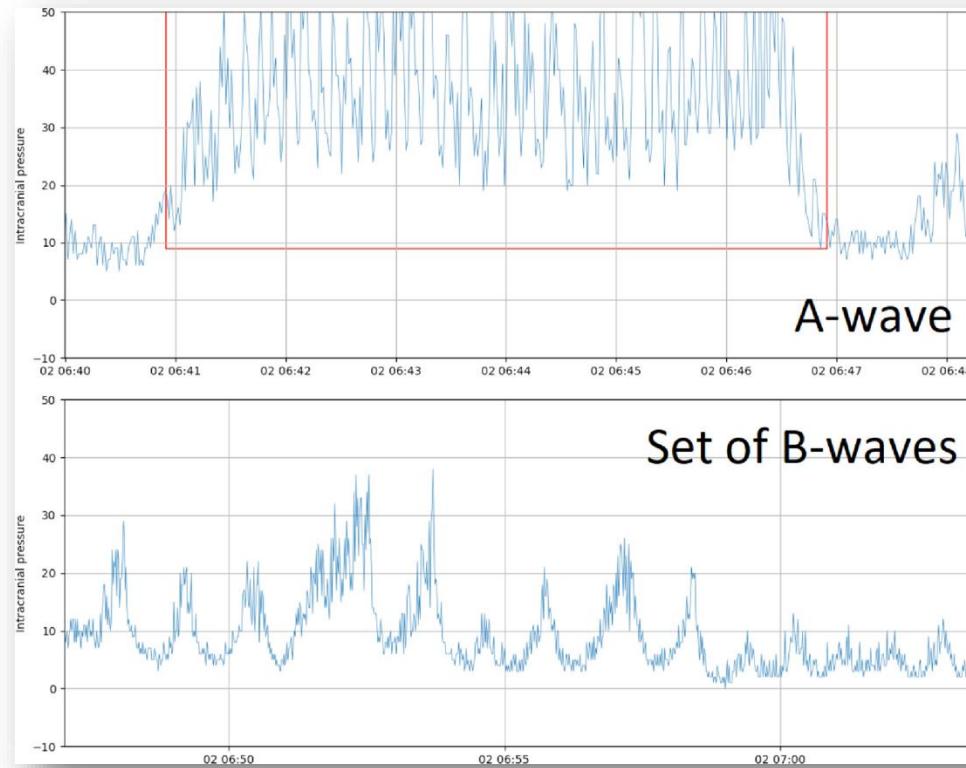


Figura 1.- Vistas dorsal y lateral de la conectividad del cerebro humano. Las etiquetas indican sub-regiones anatómicas ubicadas en sus respectivos centros de masa. Los nodos se codifican según la fuerza, y las aristas están codificadas de acuerdo al peso de conexión. (cc) Hagmann y col. (2008) Mapping the structural core of human cerebral cortex. PLoS Biology, 6, 7, e159.

## Representando miles de datos (3/3)

- Monitorización de la presión intracranal
  - Ejemplo de ondas de presión intracranal. Eje horizontal: tiempo. Eje vertical: PIC medida en mmHg



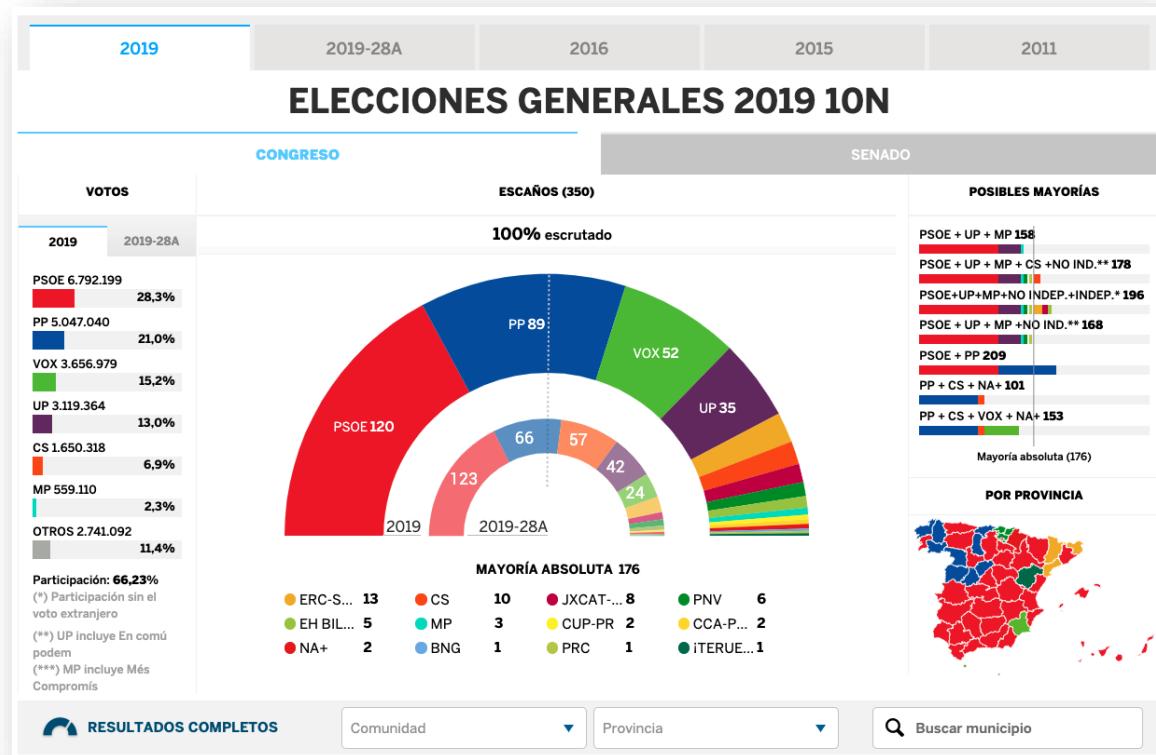
# Herramienta de comunicación



# Comunicación de resultados (1/3)

## • Elecciones

- <https://resultados.elpais.com/elecciones/generales.html>



# Comunicación de resultados (2/3)

## Economía

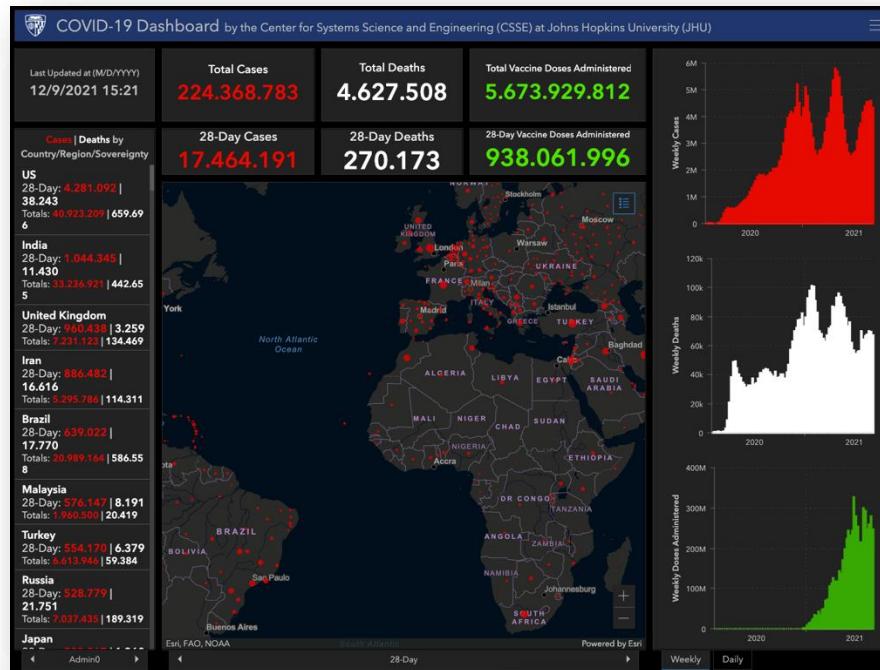
- <http://www.nytimes.com/interactive/2013/05/25/sunday-review/corporate-taxes.html>



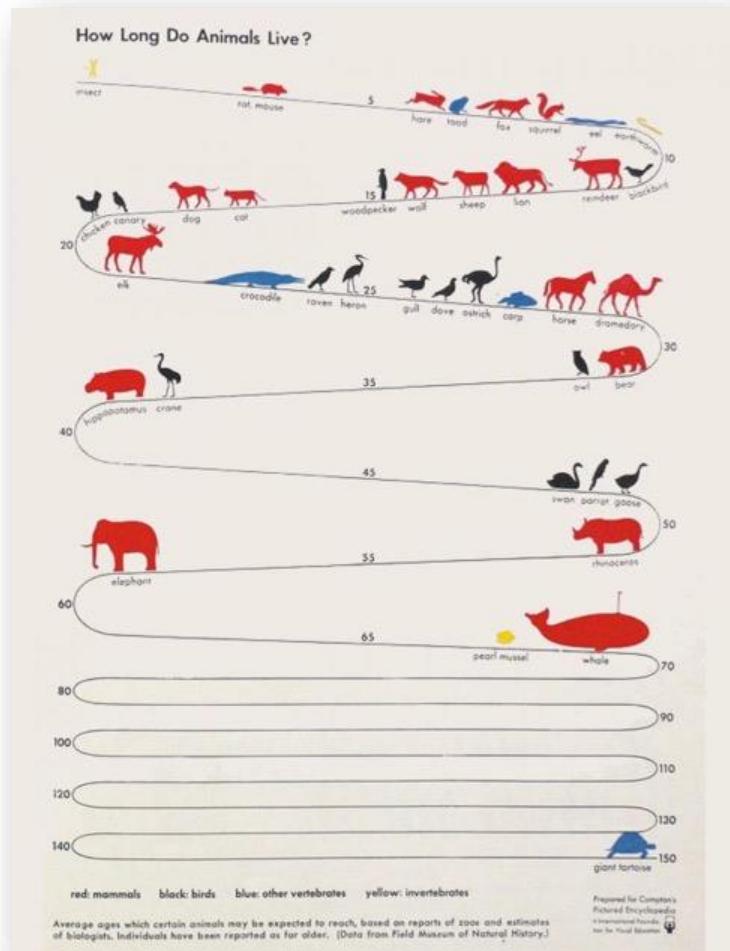
# Comunicación de resultados (3/3)

## • Salud

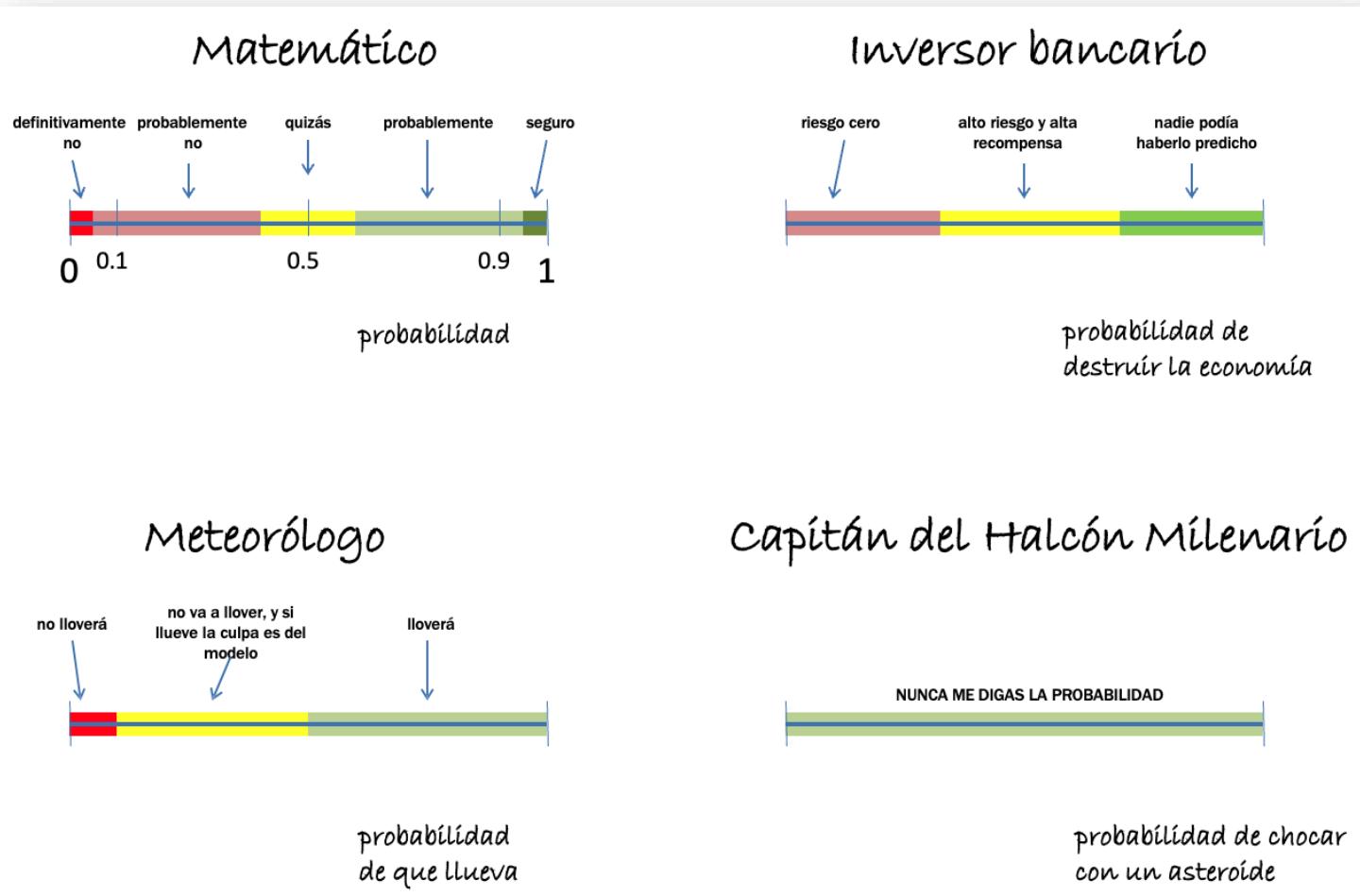
- <https://www.arcgis.com/apps/dashboards/bda7594740fd40299423467b48e9e cf6>



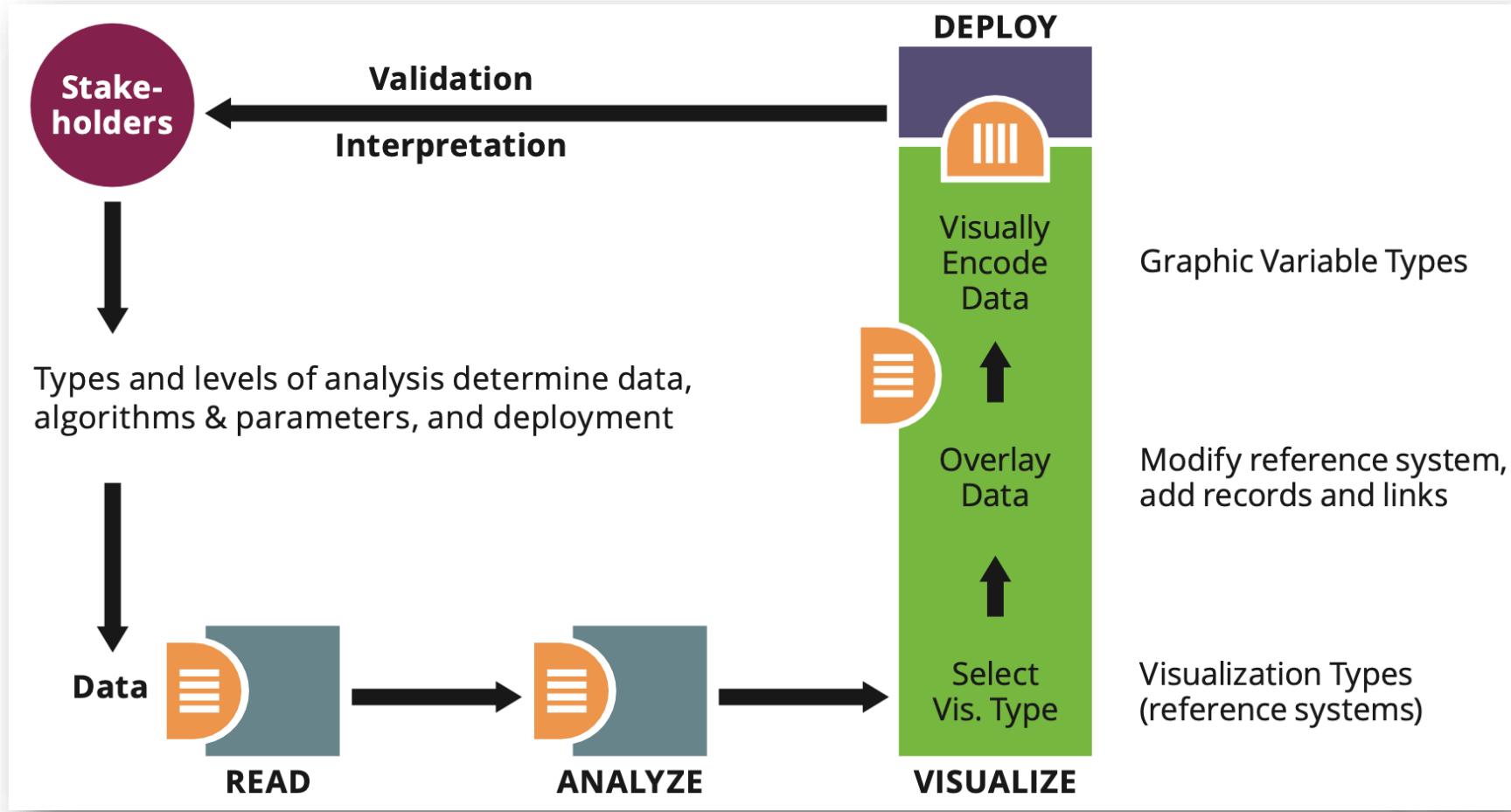
# La representación exige imaginación



# La representación de la subjetividad



# Diseño orientado a preguntas (1/2)



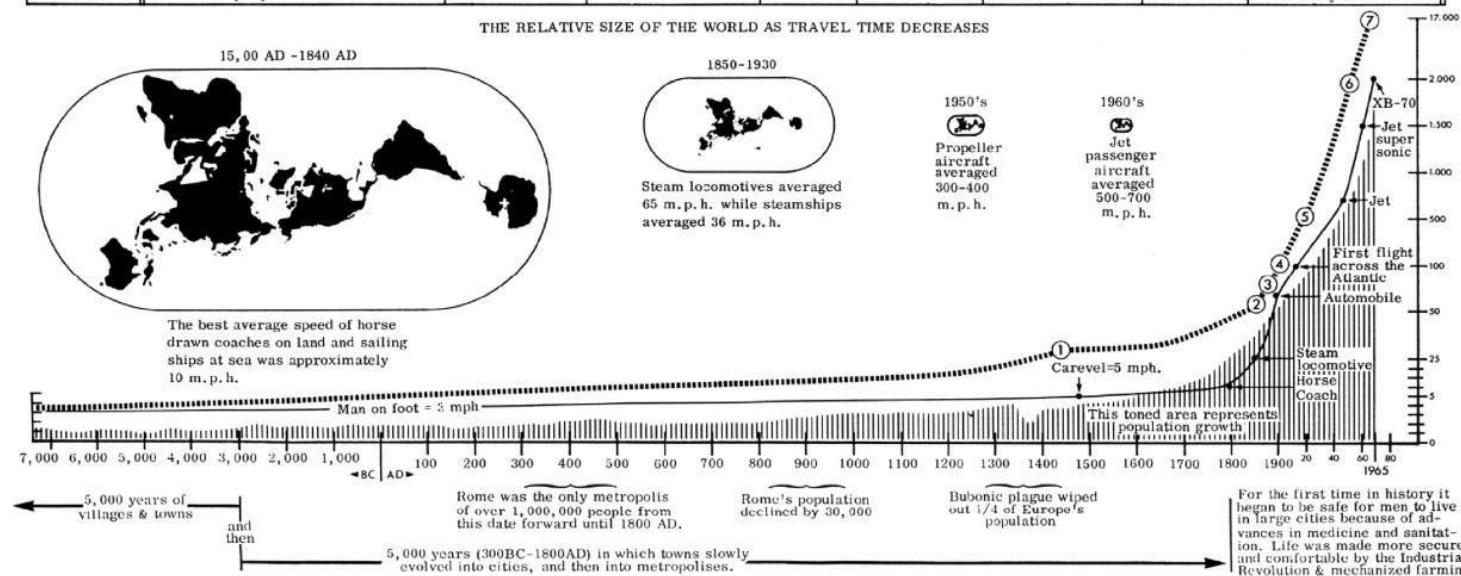
- Hay cinco tipos de análisis posibles: *análisis estadístico/profiling*, análisis *temporal*, análisis *geoespacial*, análisis *temático* y análisis de *redes*
- Cada uno de estos tipos de análisis trata de responder a un tipo de pregunta específico

Tipo de análisis	Tipo de pregunta
Temporal	<i>Cuándo</i>
Geoespacial	<i>Dónde</i>
Temático/textual	<i>Qué</i>
Red/árbol	<i>Con quién</i>

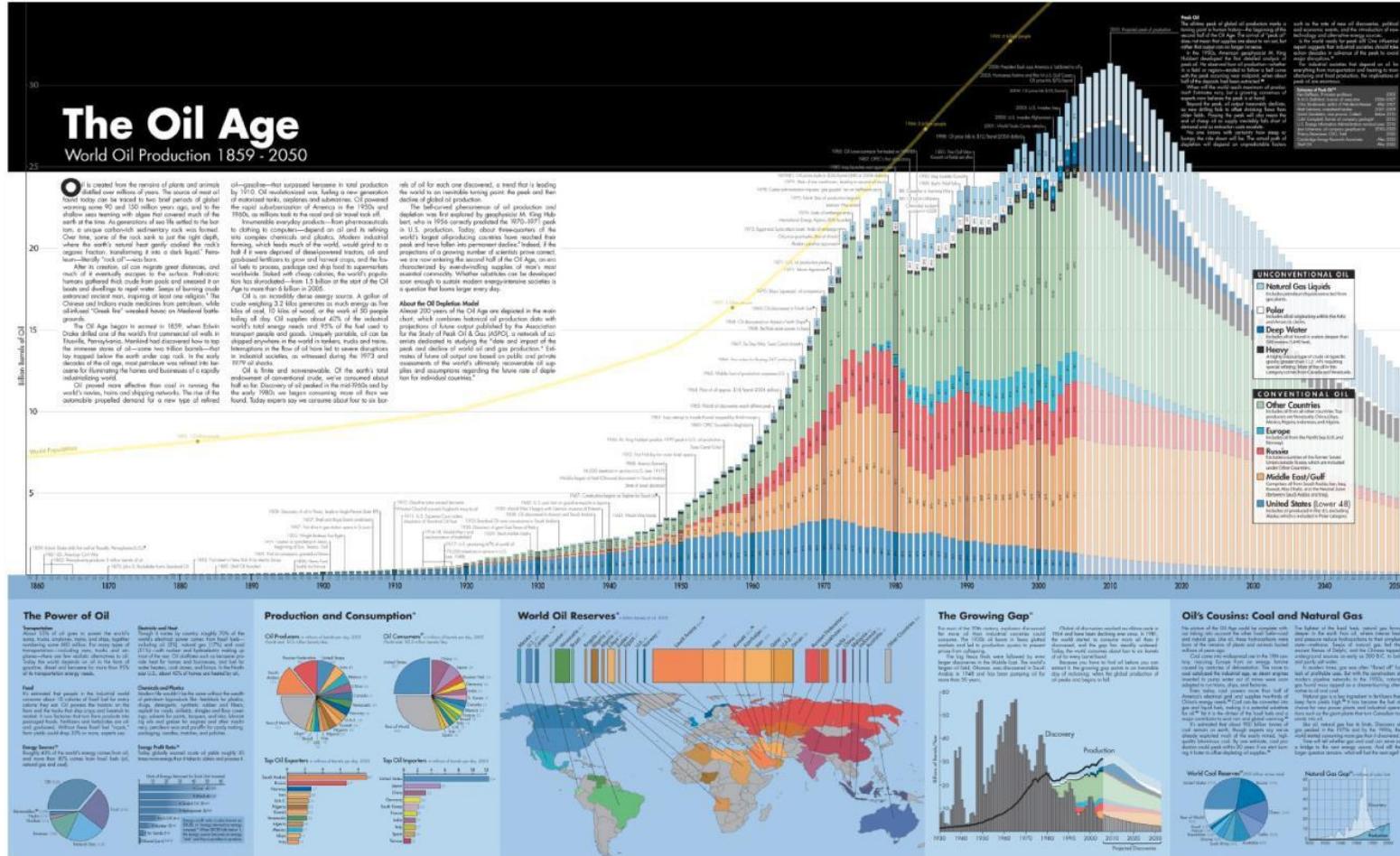
# Diseño orientado a preguntas

## ¿Cuándo? (1/2)

Shrinking of Our Planet by Man's Increased Travel and Communication Speeds Around the Globe									
Year	500,000 BC	20,000 BC	300 BC	500 BC	1,500 AD	1900 AD	1925	1950	1965
Required time to travel around the globe	A few hundred thousand years	A few thousand years	A few hundred years	A few tens of years	A few years	A few months	A few weeks	A few days	A few hours
Means of transportation	Human on foot (over, ice bridges)	On foot and by canoe	Canoe with small sail or paddles or relays of runners	Large sail boats with oars, pack animals, and horse chariots	Big sailing ships (with compass), horse teams, and coaches	Steam boats and railroads (Suez and Panama Canals)	Steamships, transcontinental railways, autos, and airplanes	Steamships, railroads, auto jet and rocket aircraft	Atomic steamship, high speed railway, auto, and rocket-jet aircraft
Distance per day (land)	15 miles	15-20 miles	20 miles	15-25 miles	20-25 miles	Rail 300-900 miles	400-900 miles	Rail 500-1,500	Rail 1000-2000
Distance per day (sea or air)		20 by sea	40 miles by sea	135 miles by sea	175 miles by sea	250 miles by sea	3,000-6000 air	6000-9500 air	408,000 air
Potential state size	None	A small valley in the vicinity of a small lake	Small part of a continent	Large area of a continent with coastal colonies	Great parts of a continent with trans-oceanic colonies	Large parts of a continent with transoceanic colonies	Full continents & Transoceanic Commonwealths	The Globe	The globe and more
Communications	Word of mouth, drums, smoke, relay runners, and hand printed manuscripts prior to 1441 A.D.	① The Gutenberg 1441 printing press ② The rapid print Web 1863 newspaper press	③ The Bell 1865 telephone	④ The Marconi 1895 telegraph	⑤ First commercial 1929 radio broadcast ⑥ National 1950 television	⑦ Transcontinental T.V. 1965	with the introduction of Early Bird satellite		



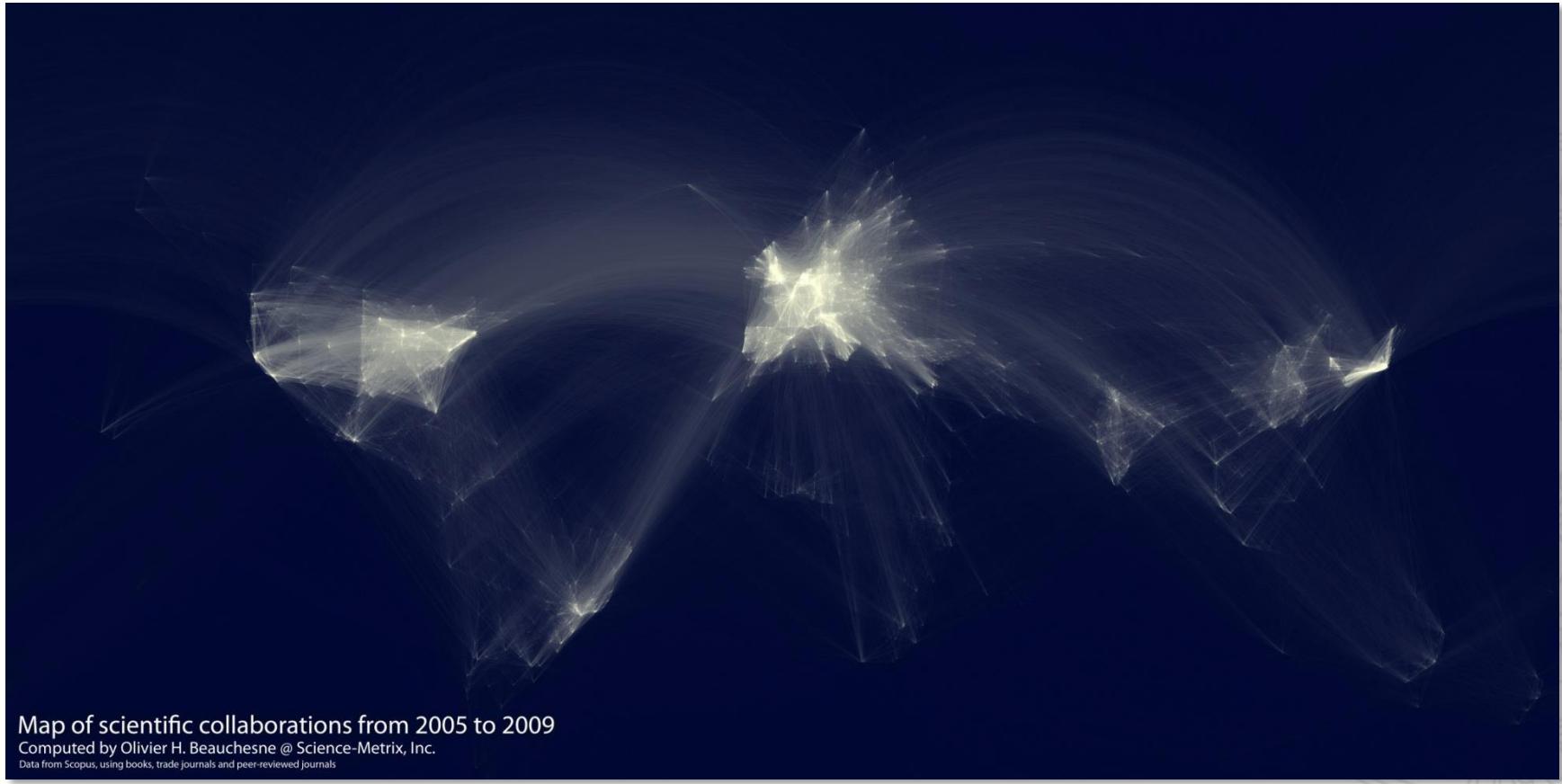
# Diseño orientado a preguntas ¿Cuándo? (2/2)



# Diseño orientado a preguntas

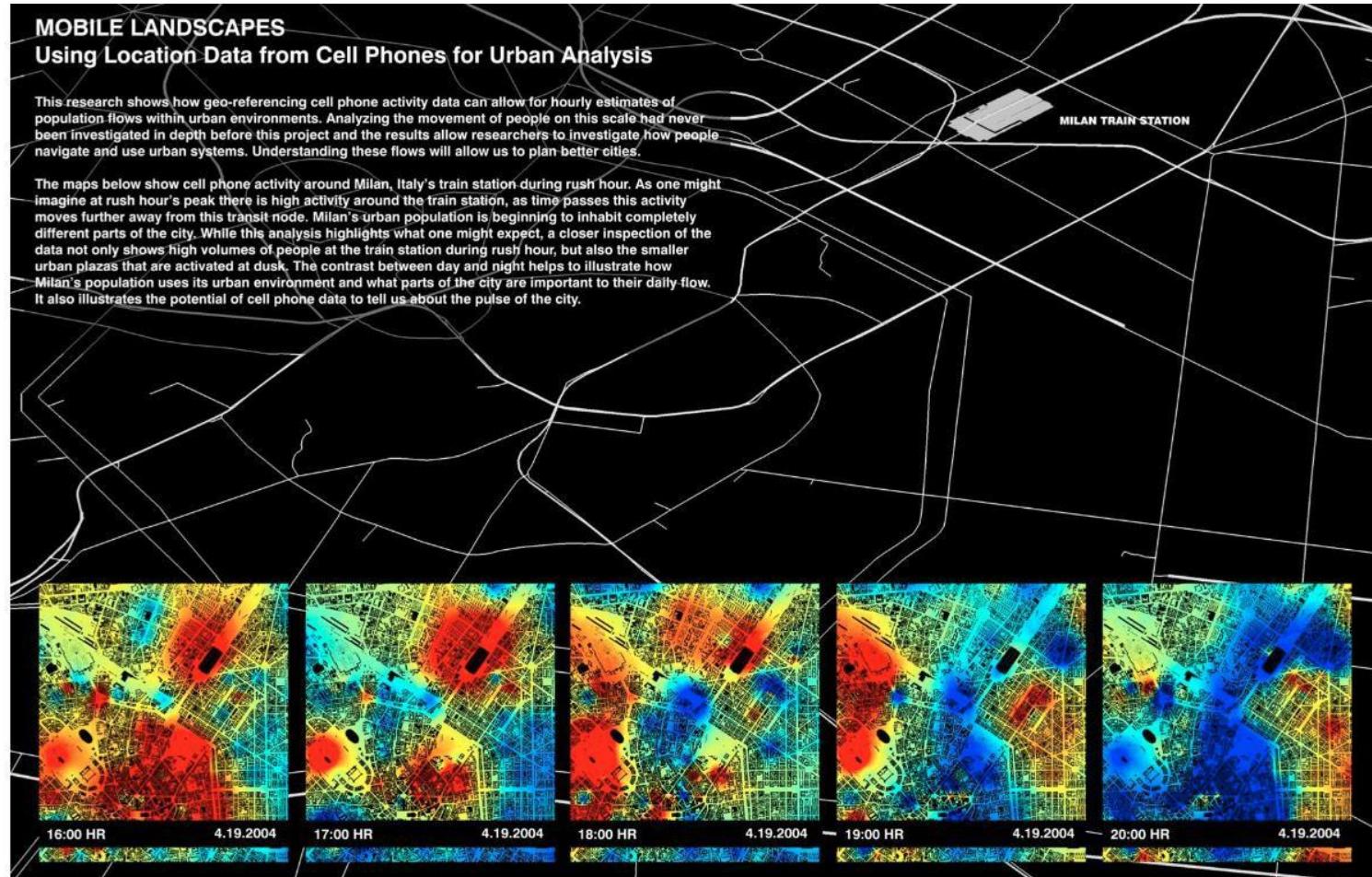
## ¿Dónde? (1/2)

UNIVERSIDAD DE  
MURCIA



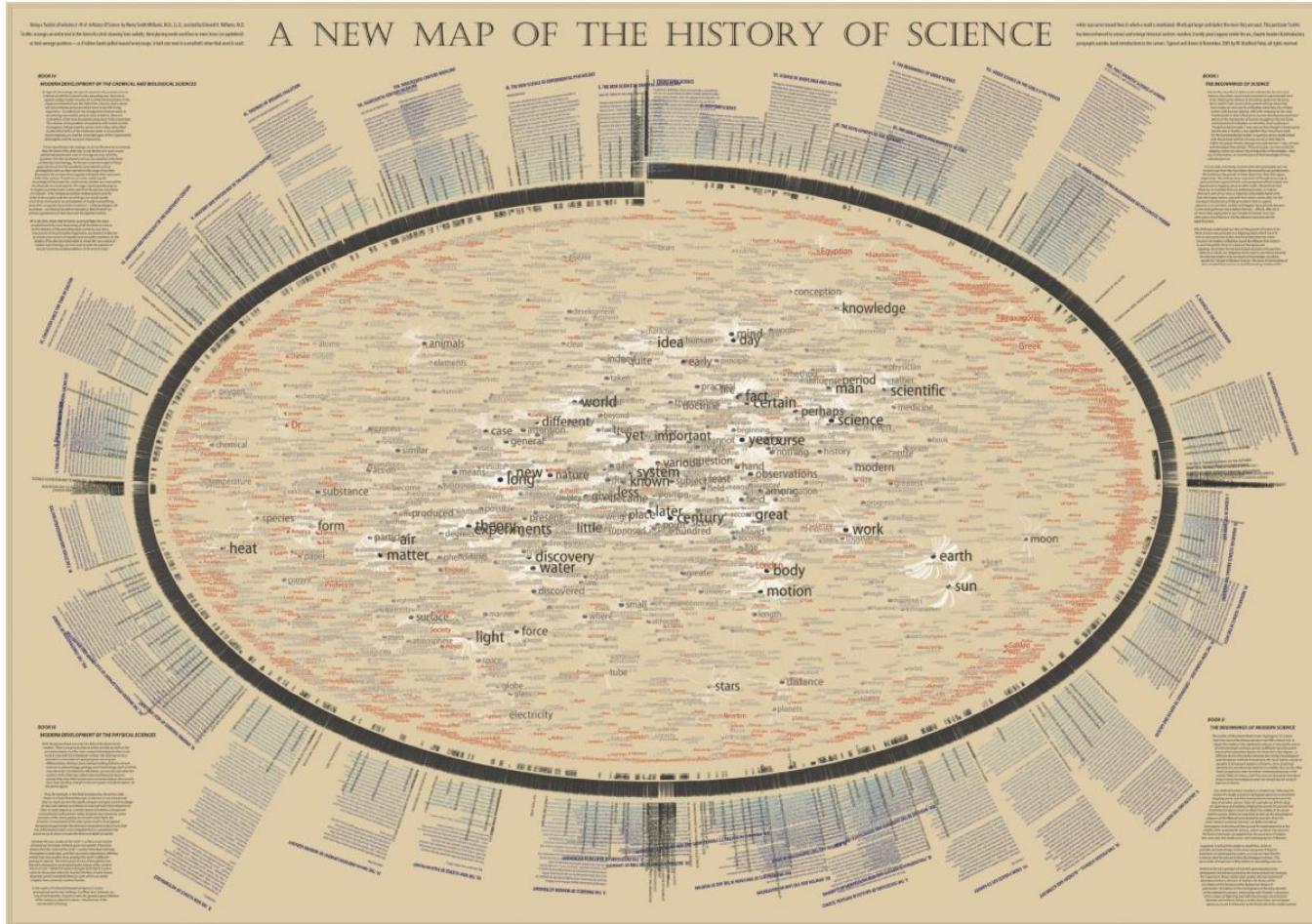
# Diseño orientado a preguntas

## ¿Dónde? (2/2)



# Diseño orientado a preguntas

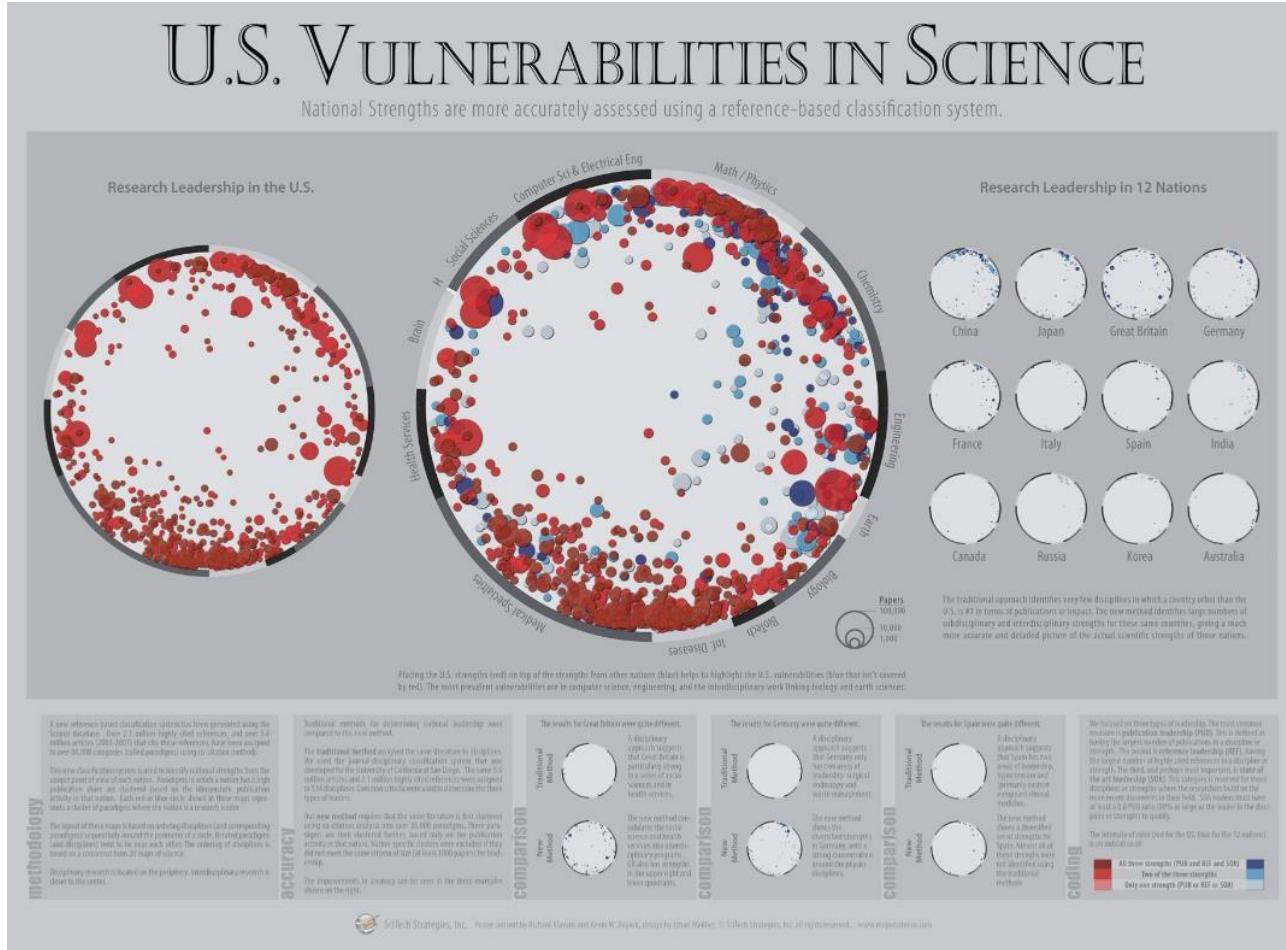
## ¿Qué? (1/2)



# Diseño orientado a preguntas

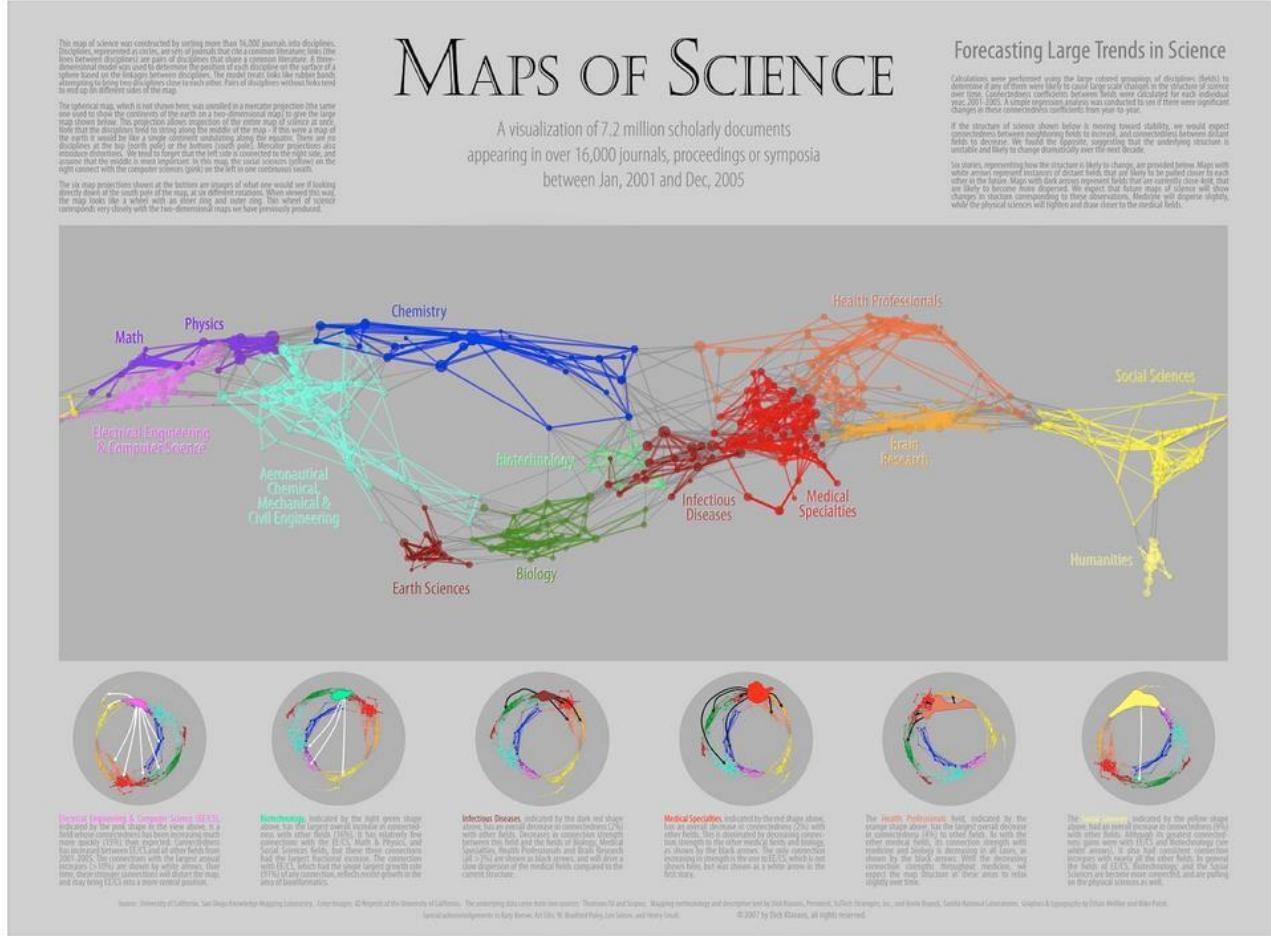
## ¿Qué? (2/2)

UNIVERSIDAD DE  
MURCIA



# Diseño orientado a preguntas

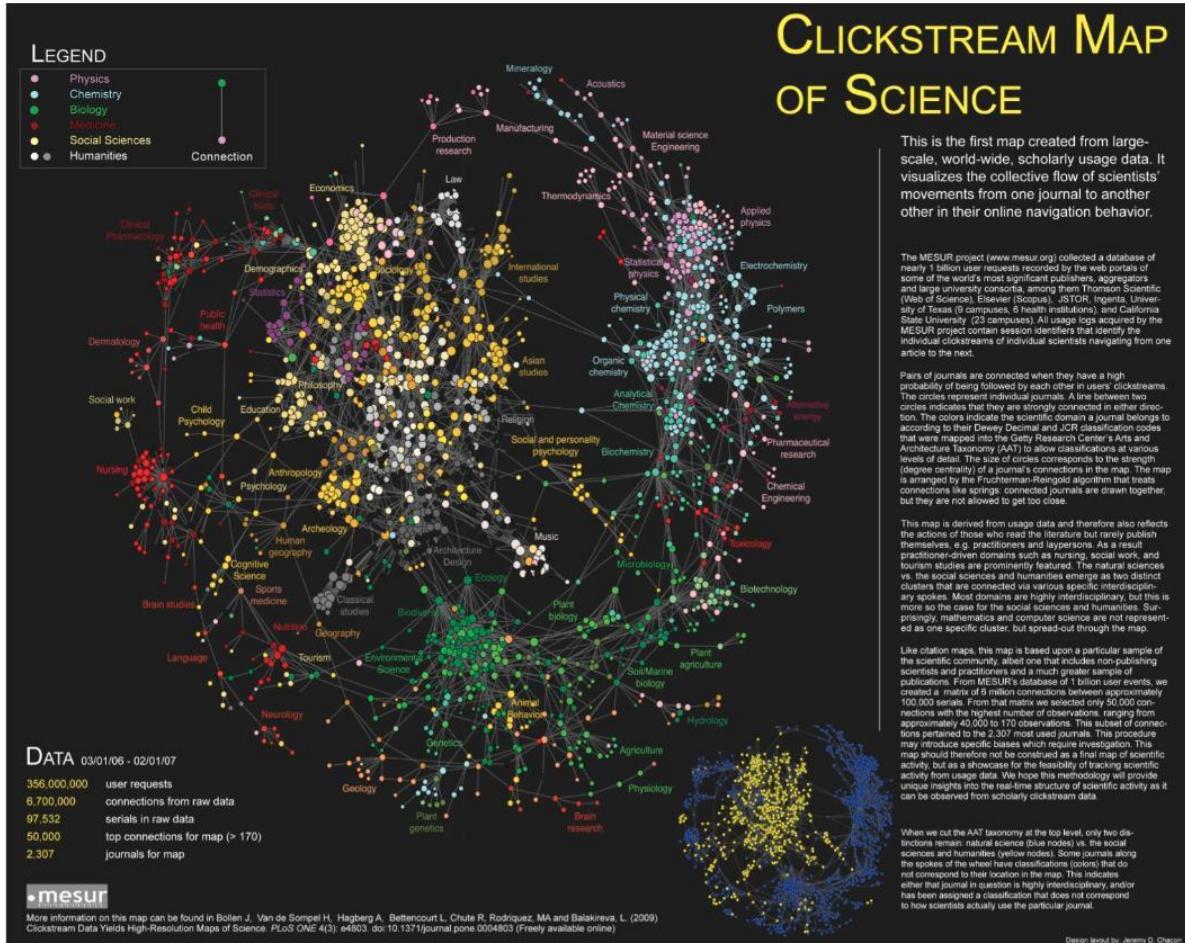
## ¿Con quién? (1/2)



# Diseño orientado a preguntas

## ¿Con quién? (2/2)

UNIVERSIDAD DE  
MURCIA



- Sobre malos ejemplos
  - WTF Visualizations (<https://viz.wtf/>)
- Formación y consultoría
  - PolicyViz (<https://policyviz.com/>)



- K. Börner and D.E. Polley. *Visual insights*. The MIT Press, 2014
- A. Cairo. *El arte funcional*. Alamut, 2011
- E.R. Tufte. *The visual display of quantitative information*. Second Edition. Graphic Press, 2001
- H. Wainer and S. Lysen. That's funny... *American Scientist*, 97(4):272, 2009