

# Visualización de datos

## Tema 1: Introducción



**Paulo Félix Lamas**  
Área de Ciencias da Computación e Intelixencia Artificial

**Departamento de Electrónica e Computación**

# Una frase

---

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

John Tukey

# Objetivos de la visualización

---

## ▷ 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

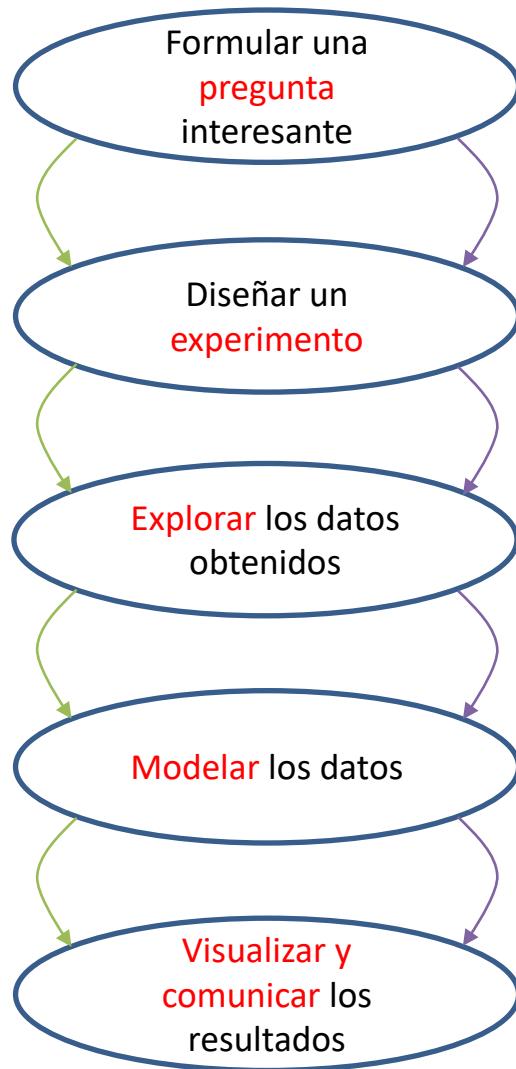
**TAREA:** Presentarlos visualmente de una manera comprensiva e intuitiva

## ▷ CONFIRMACIÓN

**PROBLEMA:** Se formula alguna hipótesis sobre el comportamiento que subyace a los datos

**TAREA:** Confirmar o falsificar hipótesis

# El método científico



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

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

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

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

$$N = 11$$

Ecuación de la recta de regresión:  $Y = 3.00 + 0.500X$

$$\bar{X} = 9.0$$

Coeficiente de correlación = 0.816

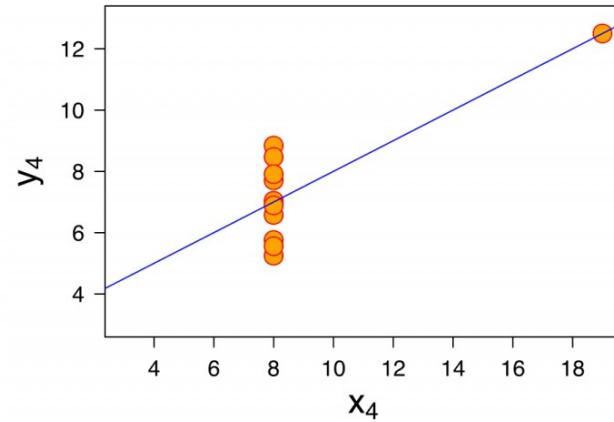
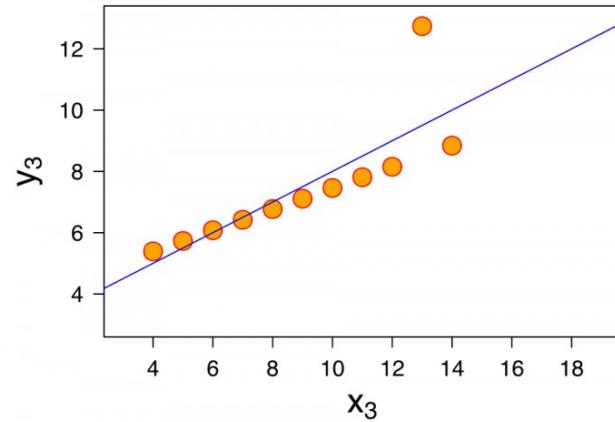
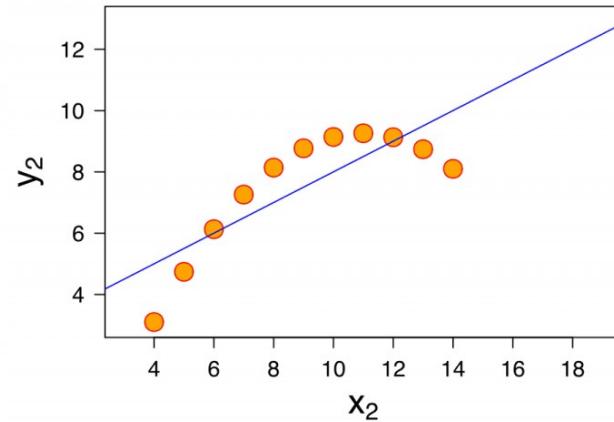
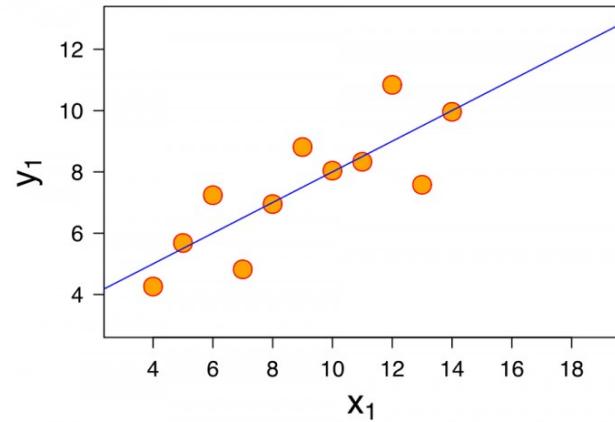
$$\bar{Y} = 7.5$$

$$r^2 = 0.67$$

Los cuatro conjuntos de datos se describen de manera exacta mediante el mismo modelo lineal!

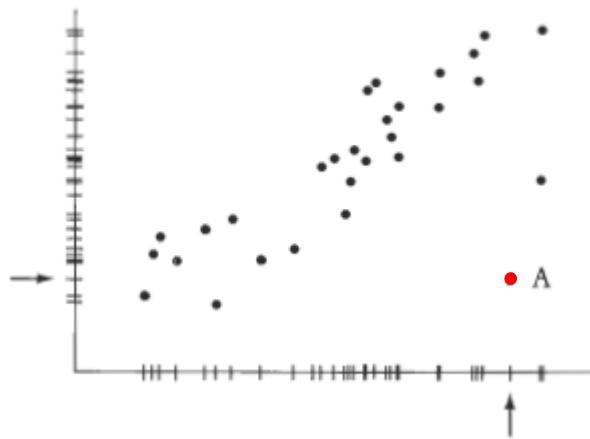
# El cuarteto de Anscombe

---



# La visualización revela información

---



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.

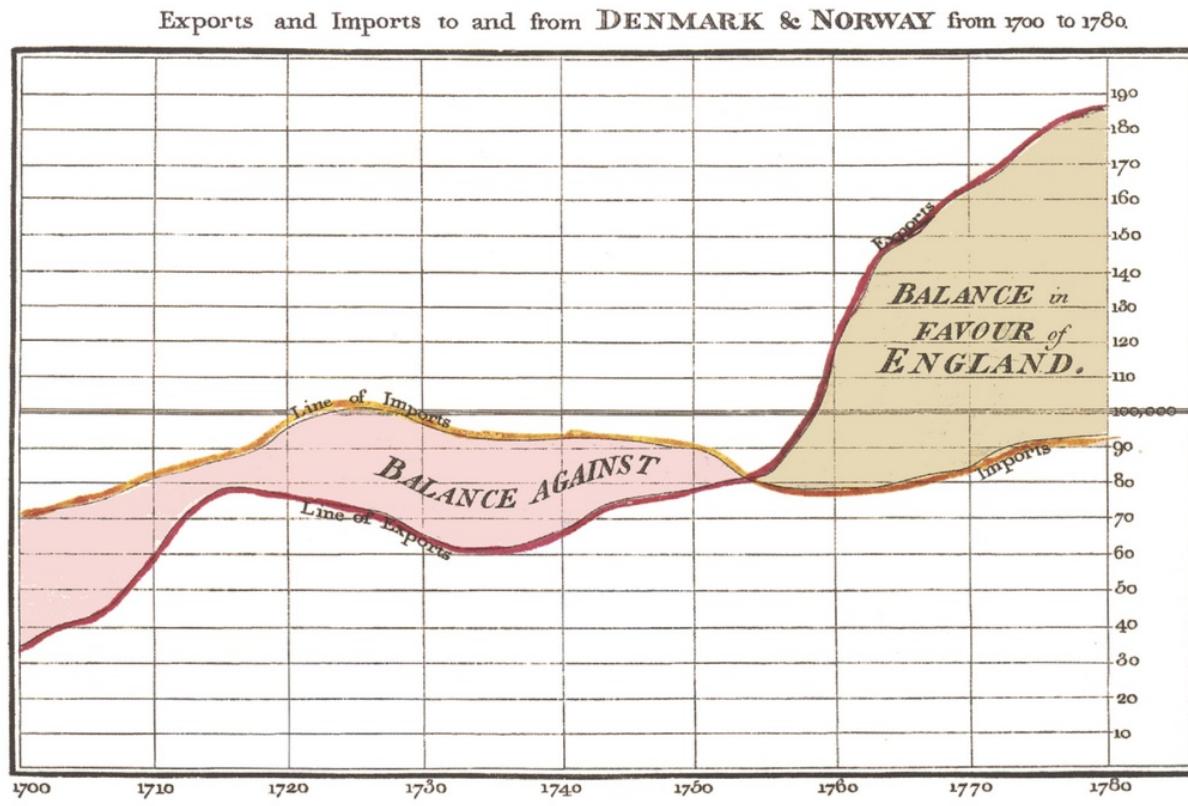
El punto A constituye claramente un punto fuera de rango, y sin embargo puede pasar inadvertido en las correspondientes distribuciones marginales.

# Una representación gráfica debe...

---

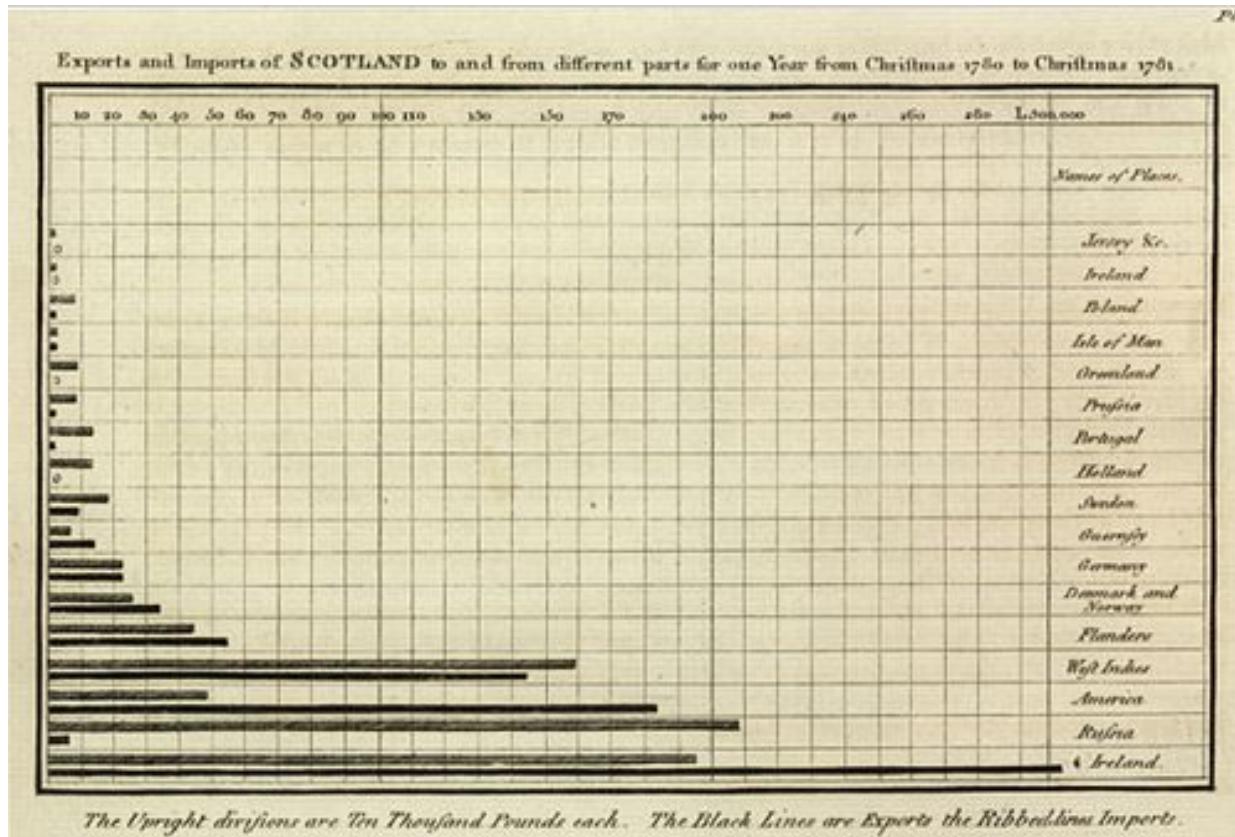
1. Mostrar los datos y ser **fiel** a los mismos.
2. Tener un **objetivo claro**: describir, explorar, comparar, descubrir, etc.
3. Llevar al observador a **razonar sobre lo substancial**, no sobre la metodología de dibujo, o sobre tecnología informática, o cualquier otro asunto.
4. Ser capaz de **sintetizar** 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.

# Desde 1700



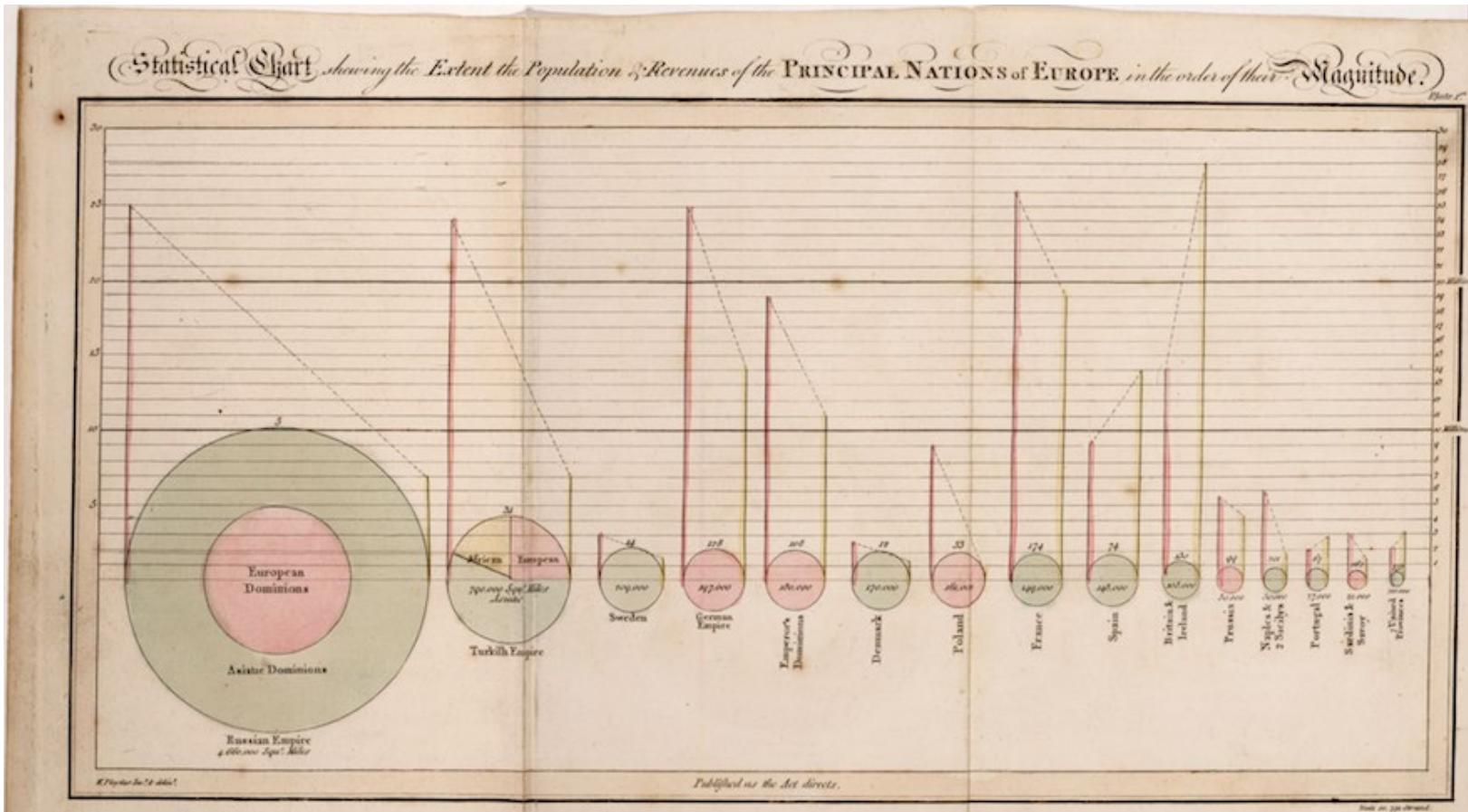
A William Playfair se atribuyen los primeros gráficos de carácter estadístico: gráficos de barras, gráficos de sectores y series temporales.

# Desde 1700



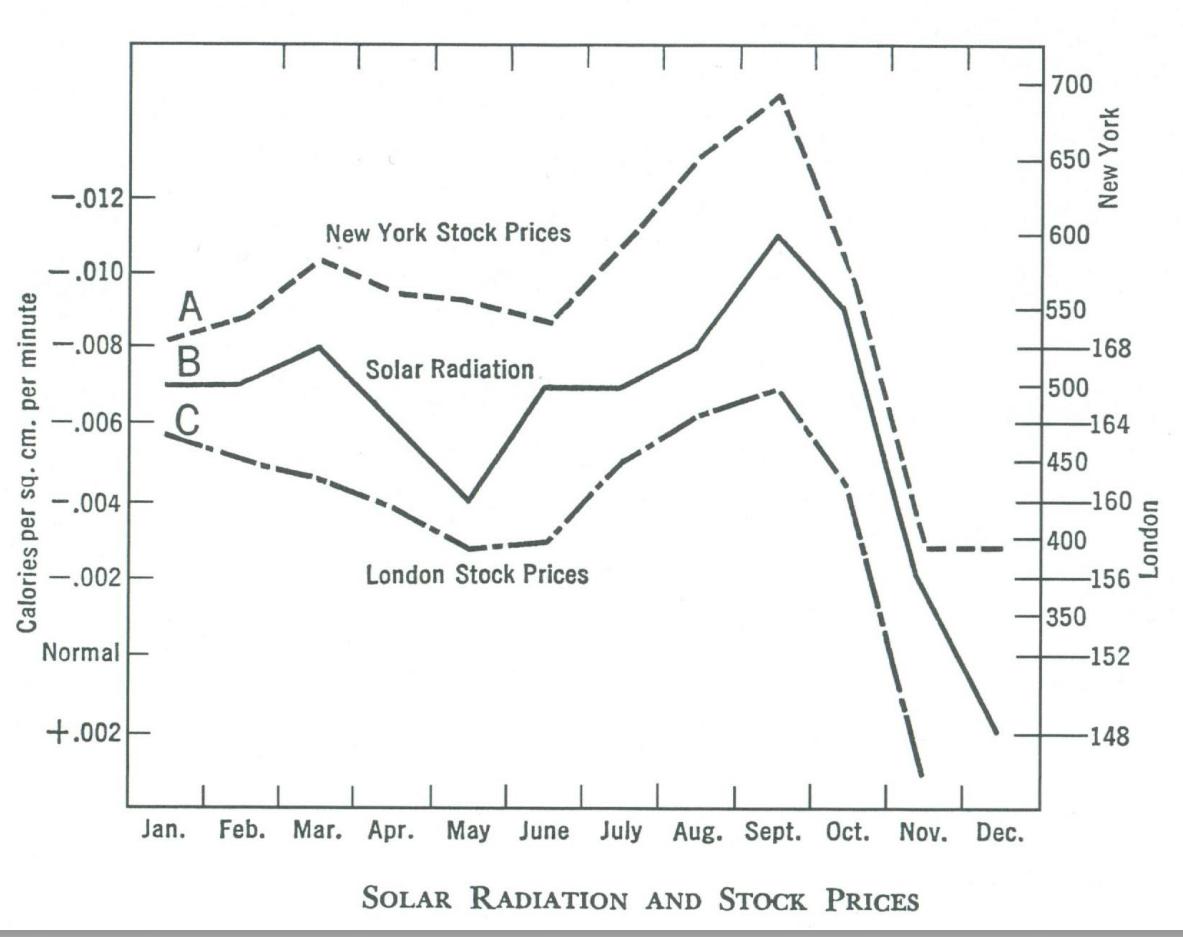
A William Playfair se atribuyen los primeros gráficos de carácter estadístico: gráficos de barras, gráficos de sectores y series temporales.

# Desde 1700



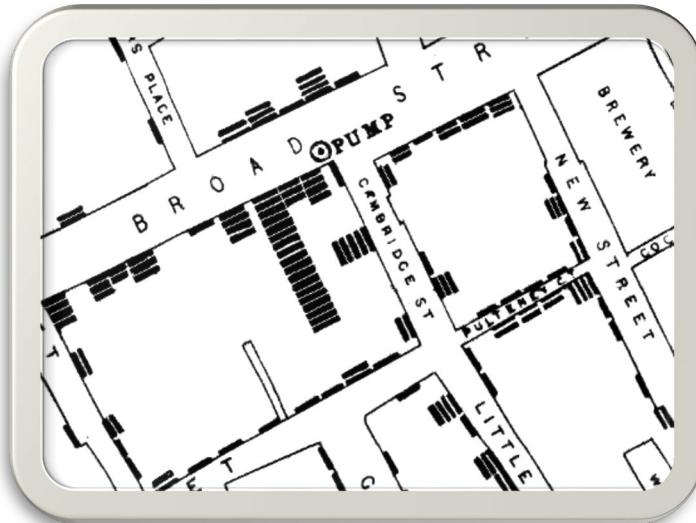
A William Playfair se atribuyen los primeros gráficos de carácter estadístico: gráficos de barras, gráficos de sectores y series temporales.

# Pero si la hipótesis es absurda...



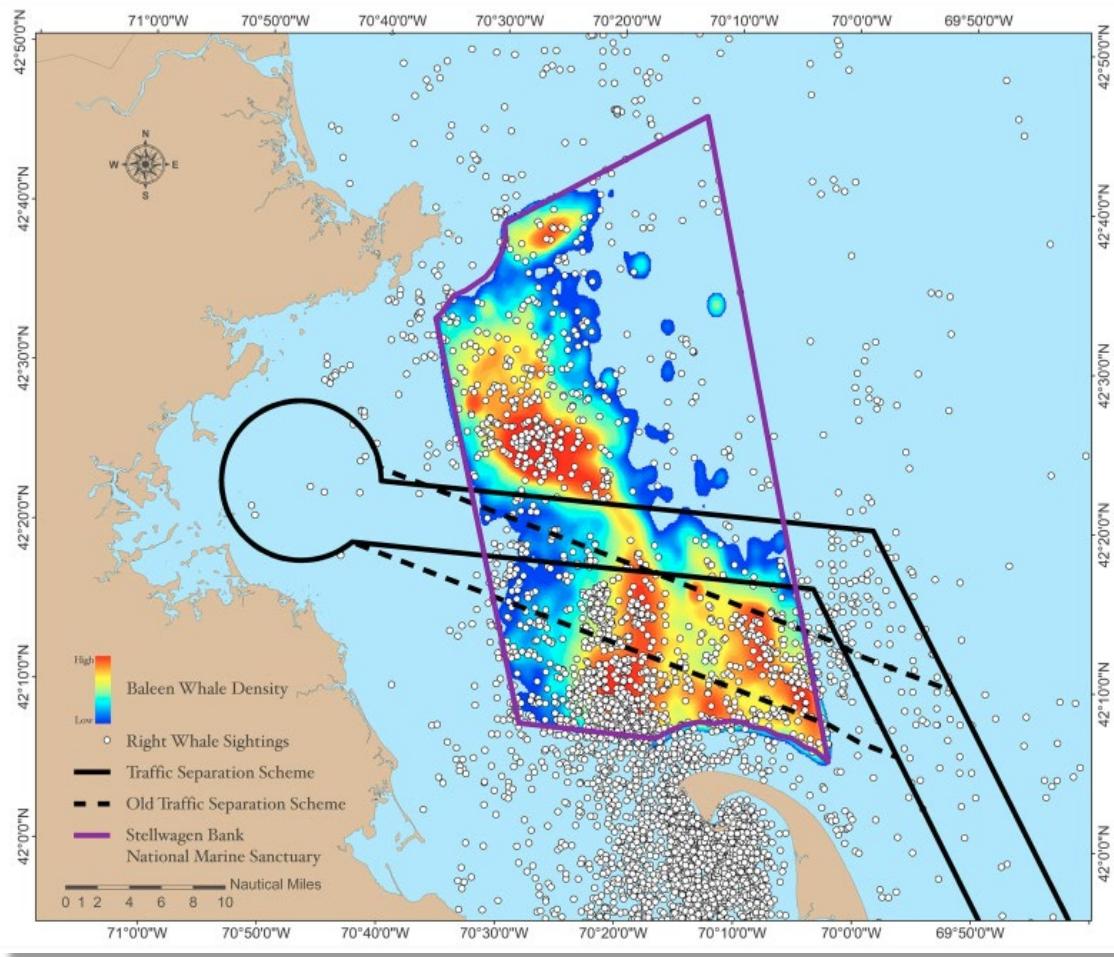
... su representación gráfica también lo es.

# Una nueva forma de razonar



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

# Una nueva forma de razonar



Desplazamiento del tráfico marino de Boston para evitar la colisión con ballenas barbadás.

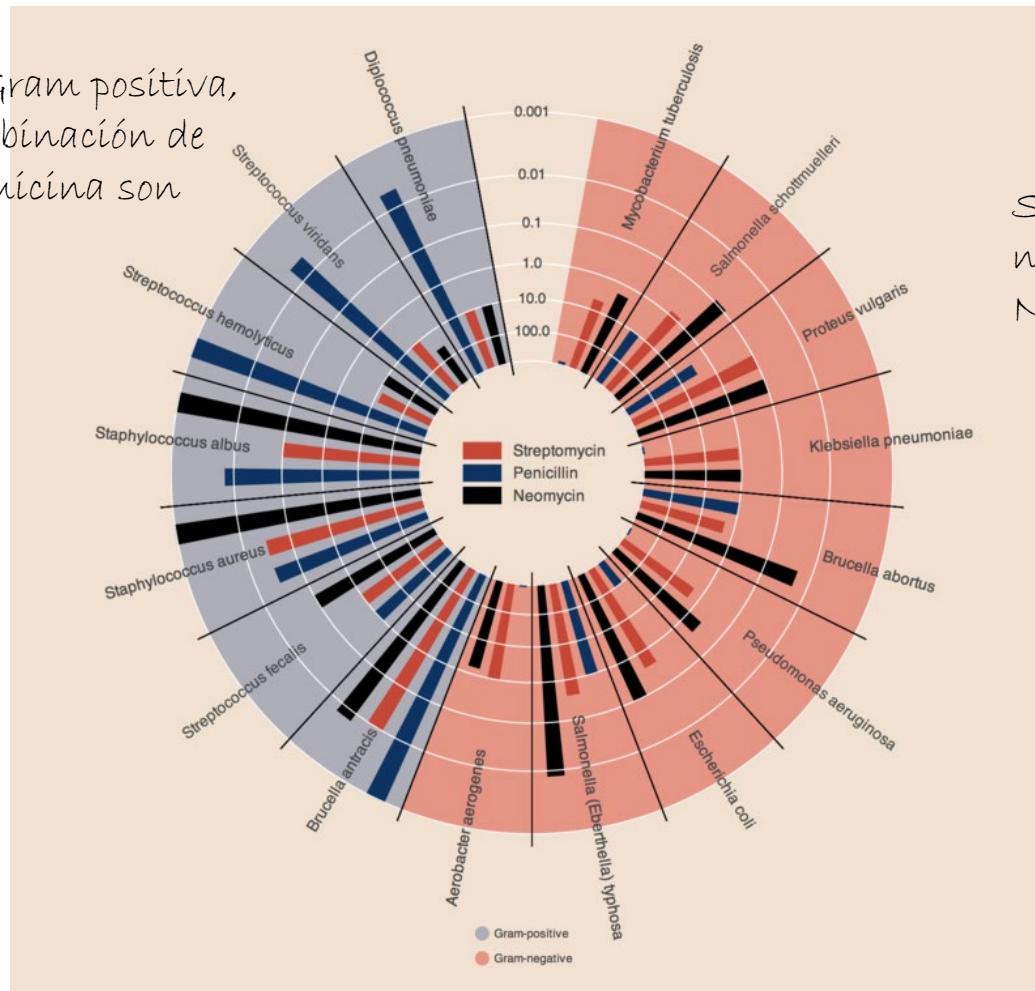
# Una nueva forma de razonar

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	+

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

# Una nueva forma de razonar

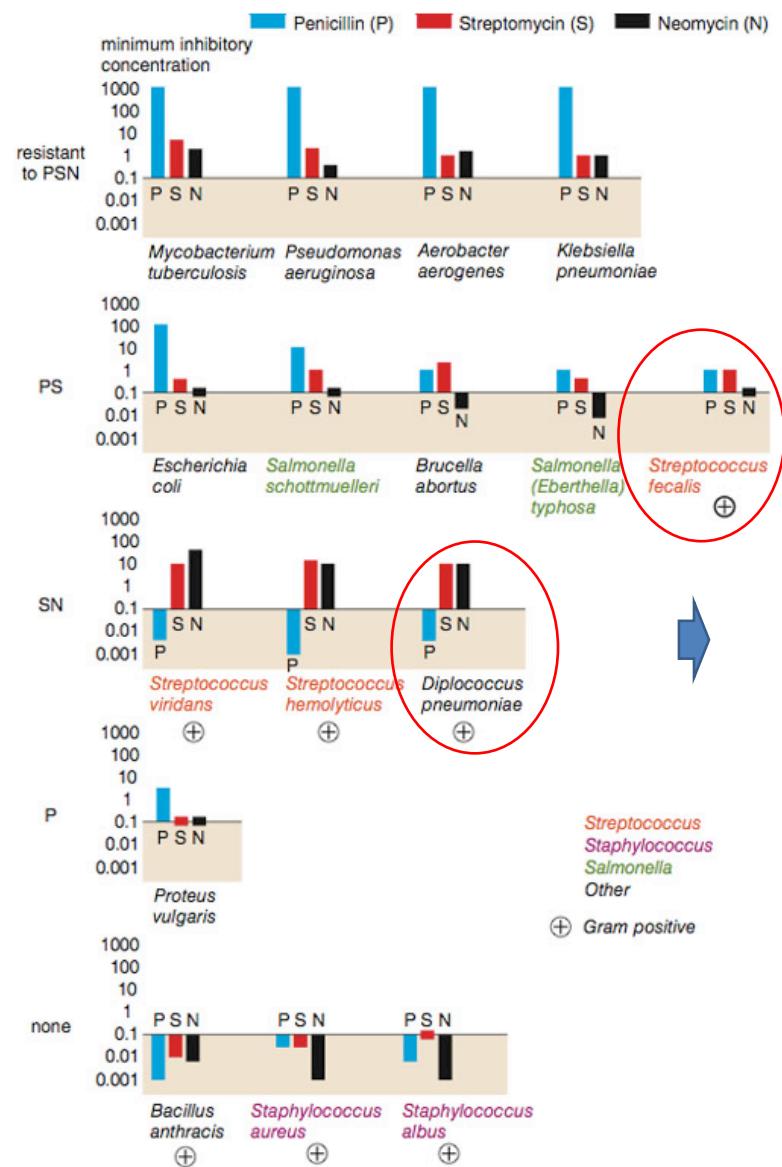
Sí la bacteria es Gram positiva, entonces una combinación de Penicilina y Neomicina son efectivas



Sí la bacteria es Gram negativa, entonces la Neomicina es más efectiva

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

# Hagamos otras preguntas

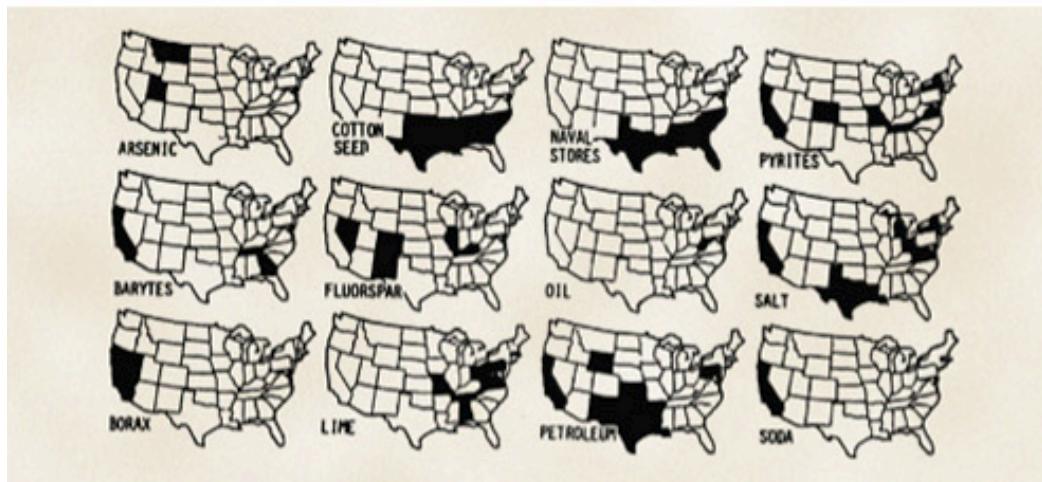
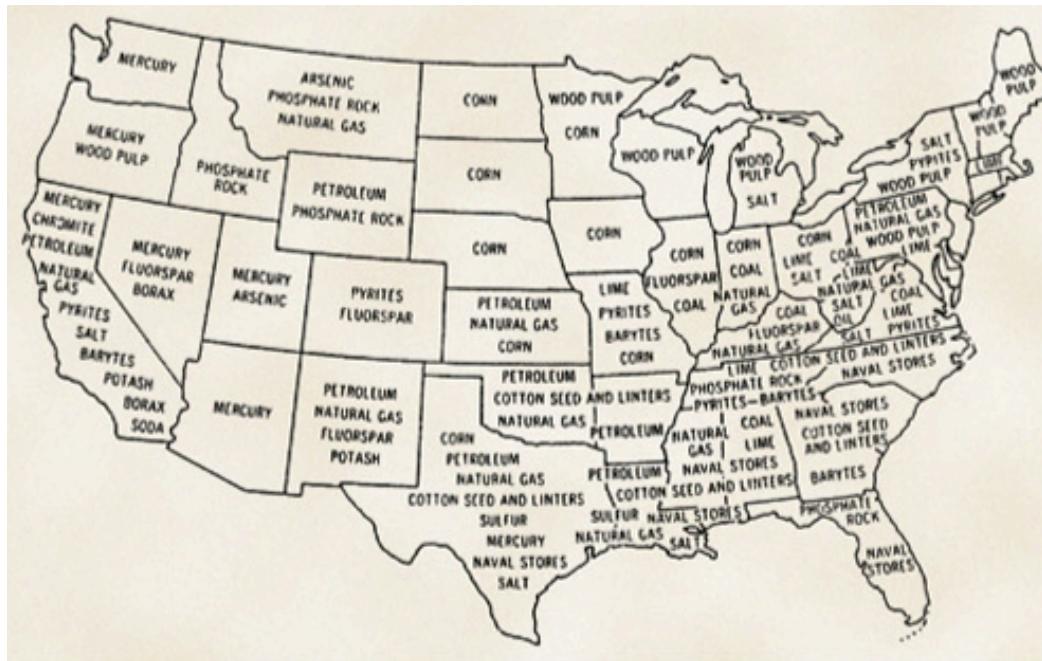


Se observó que el *Streptococcus fecalis* no obedecía a algunas características de los estreptococos, y acabó cambiando su nombre en 1984 por *Enterococcus fecalis*.

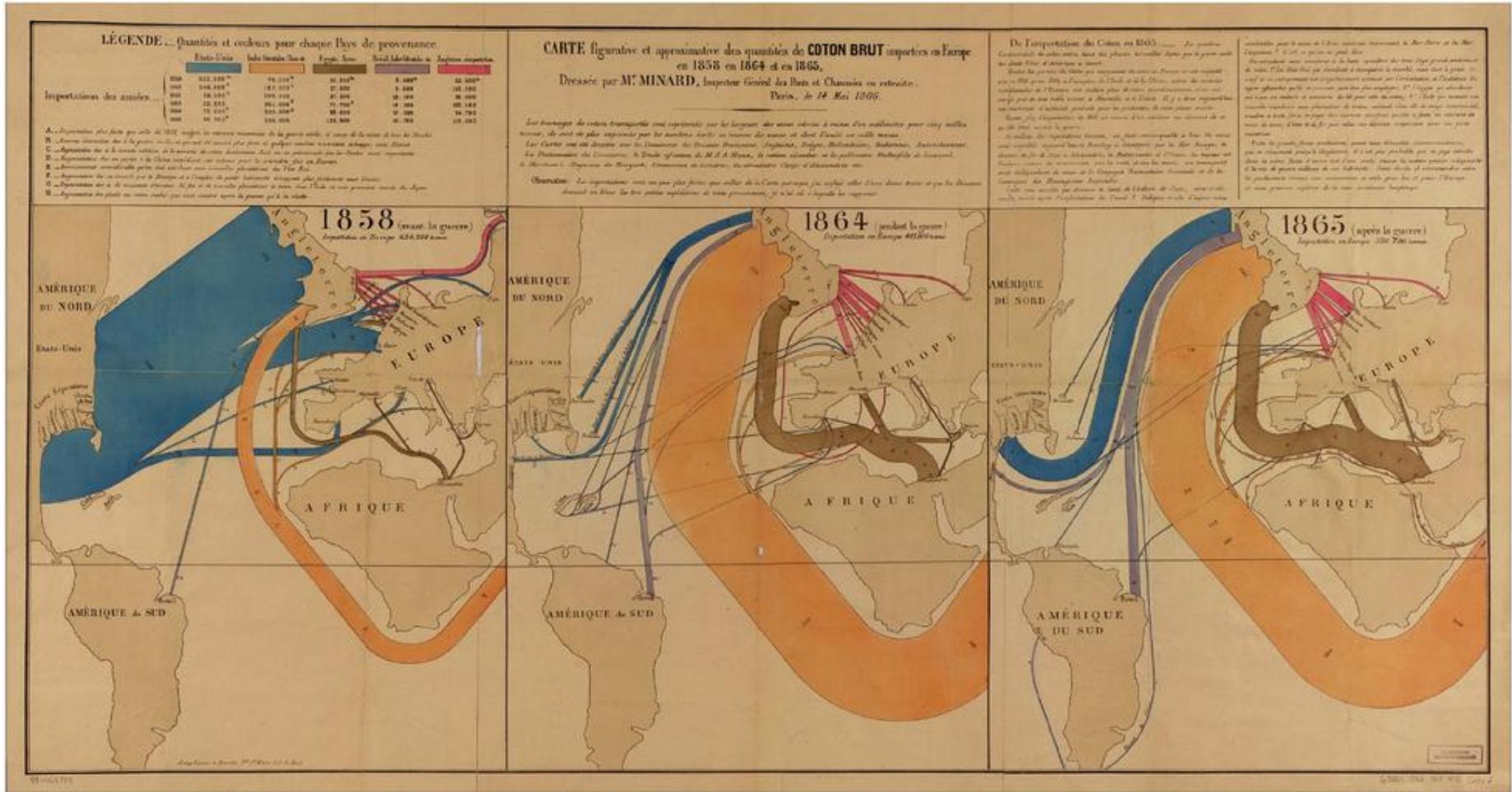
Se observó que el *Diplococcus pneumoniae* se comportaba como un espectrococo, y en 1974 se cambió su nombre por *Streptococcus pneumoniae*.

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

# Un gráfico – una pregunta

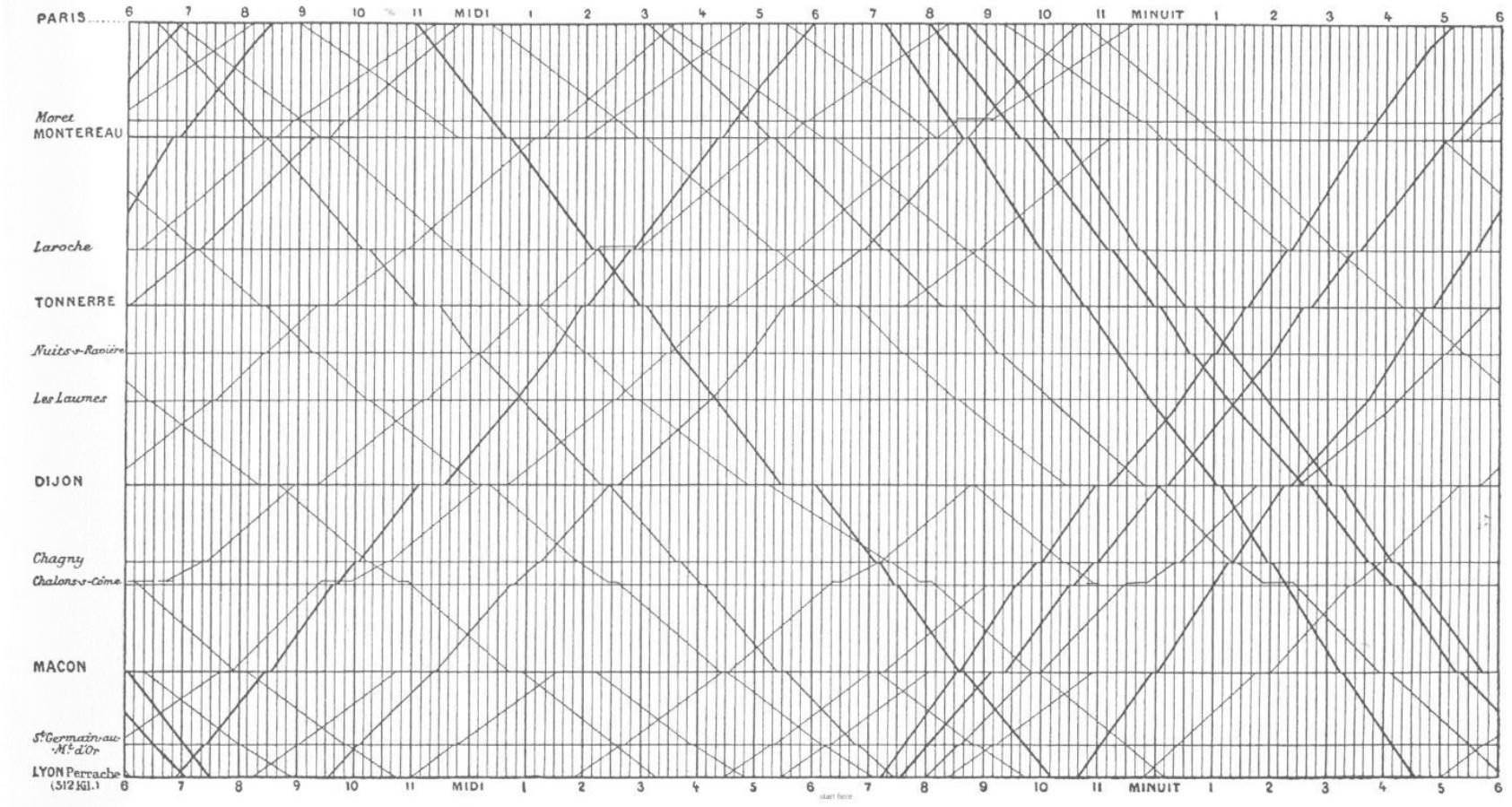


# Metáforas visuales



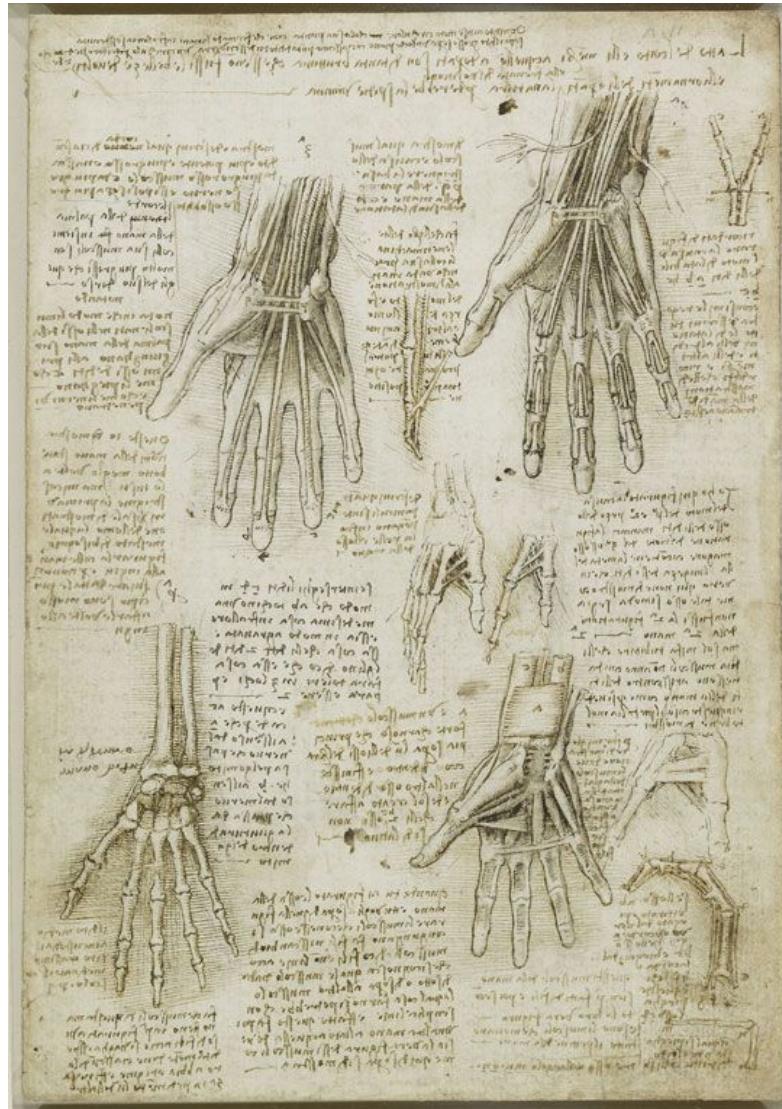
Mapas realizados por el francés C.J. Minard para representar las importaciones de algodón en Europa en tres fechas distintas.

# Máxima sencillez



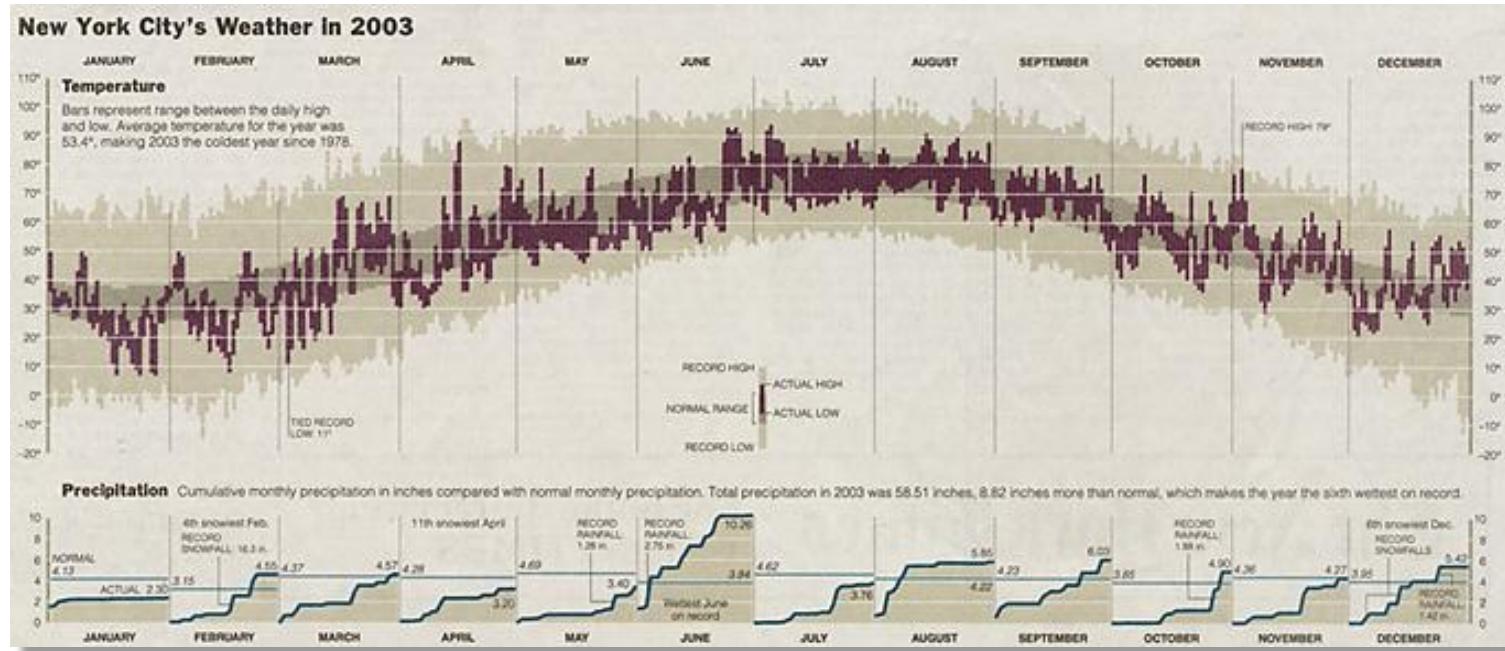
Horario de trenes de París a Lyon de 1880. Publicado por E.J. Marey.

# Visualizar para comprender



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

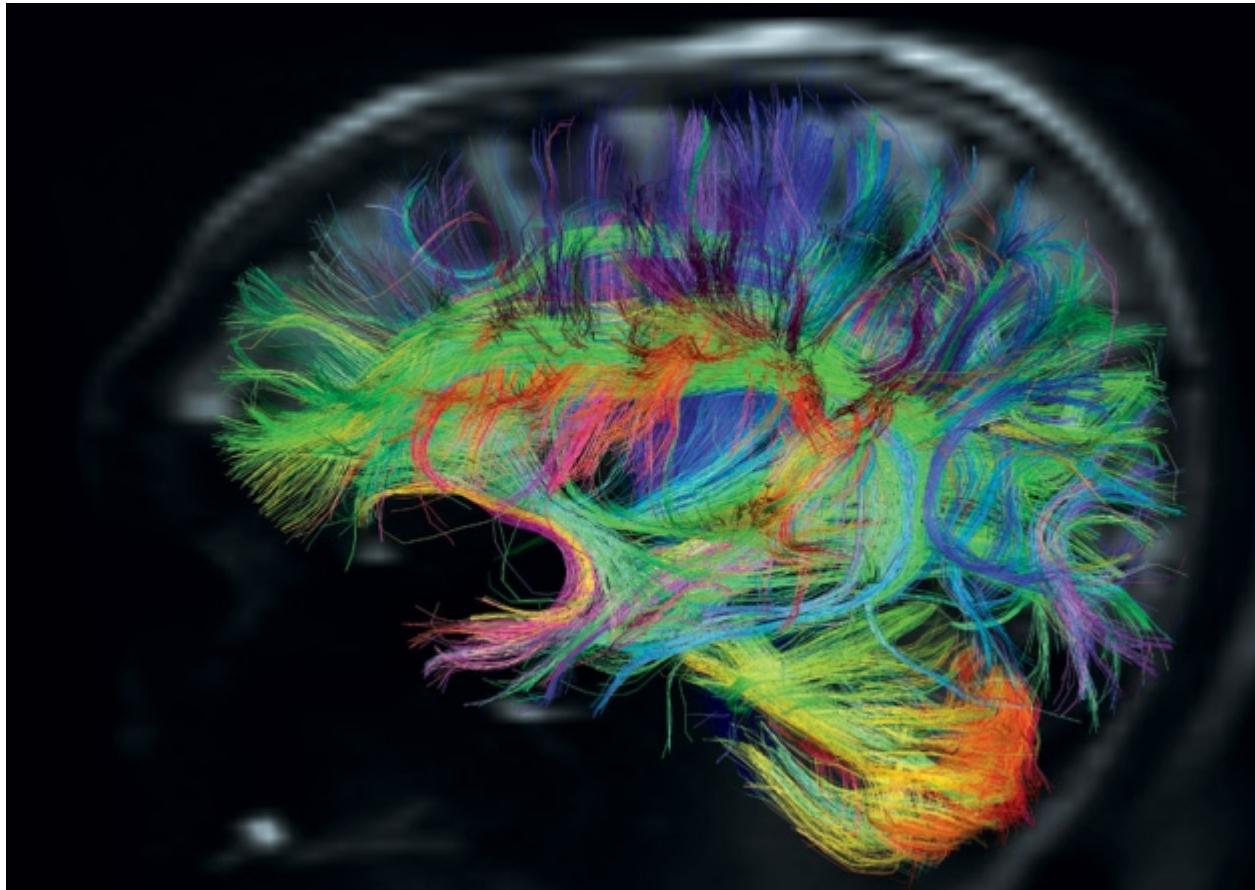
# Representando miles de datos



New York Times, 4 de enero de 2004.

# Representando millones de datos

---



Proyecto conectome: 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.

# Herramienta de comunicación



# Comunicación de resultados

HOME PAGE | MY TIMES | TODAY'S PAPER | VIDEO | MOST POPULAR | TIMES TOPICS | Get Home Delivery | Log In | Register Now

The world awaits.

Holland America Line

# The New York Times

Tuesday, June 3, 2008 Last Update: 5:56 PM ET

Holland America Line

Find your cruise.

Ameriprise Financial

NYT Archive Since 1981 | Search

**JOB**  
REAL ESTATE  
AUTOS  
ALL CLASSIFIEDS

**WORLD**  
**U.S.**  
Politics Washington Education  
**N.Y./REGION**  
**BUSINESS**  
**TECHNOLOGY**  
**SPORTS**  
**OLYMPICS**  
**SCIENCE**  
**HEALTH**  
**OPINION**  
**ARTS**  
Books Movies Music Television Theater  
**STYLE**  
Dining & Wine Fashion & Style Home & Garden Weddings/ Celebrations

**Superdelegate Declarations Push Obama Closer to Goal**  
By JEFF ZELENY and ADAM NAGOURNEY 18 minutes ago  
A parade of superdelegates endorsed Senator Barack Obama, placing him on the cusp of clinching the Democratic nomination as the 16-month primary campaign drew to a close.

**Obama Looks to Recruit Clinton's Top Fund-Raisers**  
By CHRISTOPHER DREW and LESLIE WAYNE 3:44 PM ET  
The Obama camp believes Clinton's team could add a boost to the record amounts it is already raising.

**PHOTOGRAPHS: THE PRIMARIES**

**THE 2008 CAMPAIGN**  
LOOKING BACK: HOW DIFFERENT GROUPS VOTED  
How men in each state voted

← Stronger Clinton support → Stronger Obama support

Men	Under age 30	Under \$15K	No college
Women	Age 30-44	\$15K-30K	Some college
Blacks	Age 45-59	\$30K-50K	College grads
Whites	Age 60+	Over \$50K	Post graduate

Source: Edison/Mitofsky exit polls Shan Carter and Amanda Cox

**THE CAUCUS**  
**The Superdelegate Tally**  
By JULIE BOSMAN 8 minutes ago  
The Obama campaign says it is 16.5 delegates away.

**MOVIES »**  
**A Reverse Rollout**  
Sony Pictures Classics will open its comic thriller in Austin. **Trailer:** 'Baghead'  
**OPINION »**  
**A Reverse Rollout**  
Sony Pictures Classics will open its comic thriller in Austin. **Trailer:** 'Baghead'  
Herbert: Gift to G.O.P.  
Brooks: Calling Dr. Doom  
Freakonomics: Hockey  
Editorial: Immigration 369 Comments

**MARKETS »** At 12:38 PM ET  
**S.&P. 500** 1,952.77 **Dow** 16,371.66 **Nasdaq** 4,799.77  
+0.48 +41.26 +3.52  
+0.02% +0.25% +0.07%

**GET QUOTES** My Portfolios »  
Stocks, ETFs, Funds Go

REAL ESTATE | AUTOS | JOBS | ALL CLASSIFIEDS

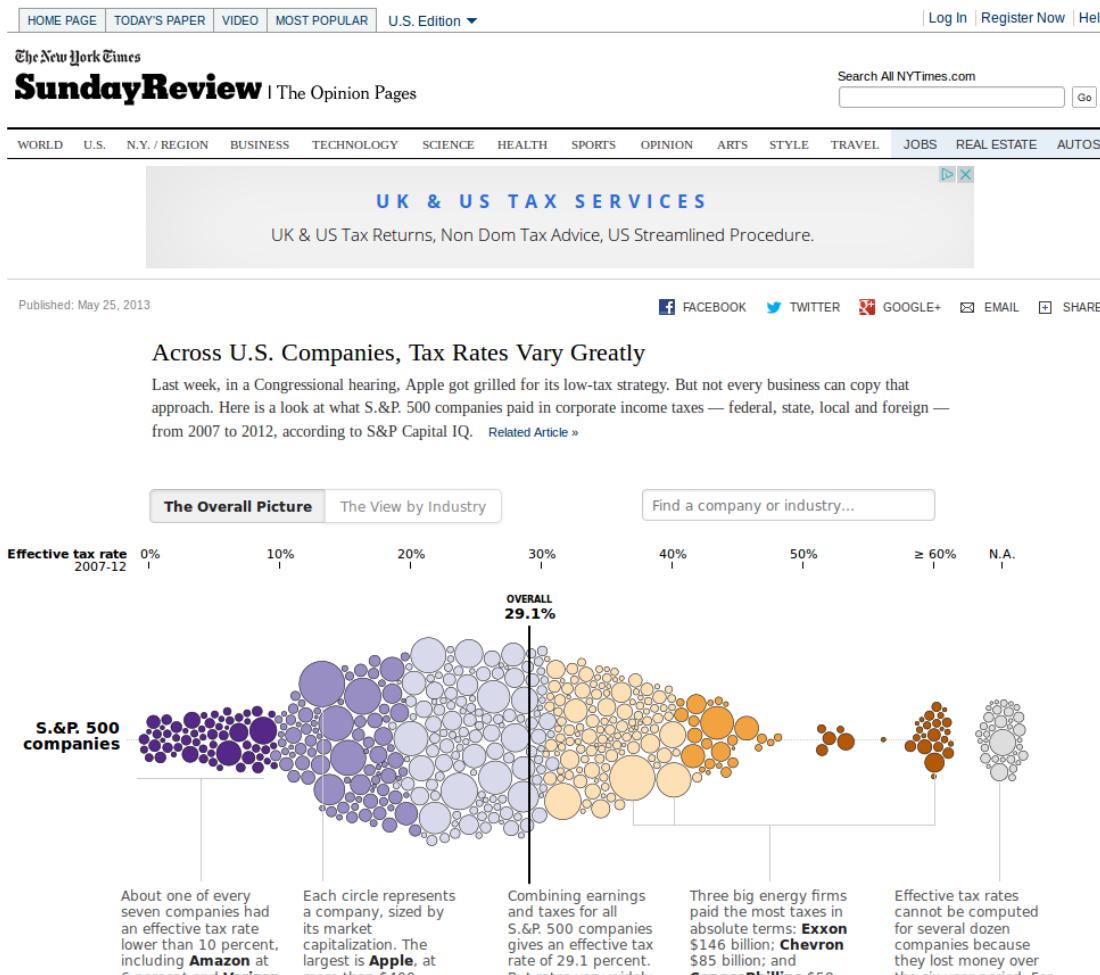
TOWN & COUNTRY AND GRAND CARAVAN SEARCH FOR NEW CARS  
Select Make Go

Redesign Rearranges the Furniture

SEARCH FOR USED

<http://www.nytimes.com/interactive/2012/03/07/us/politics/exit-polls-boxes.html>

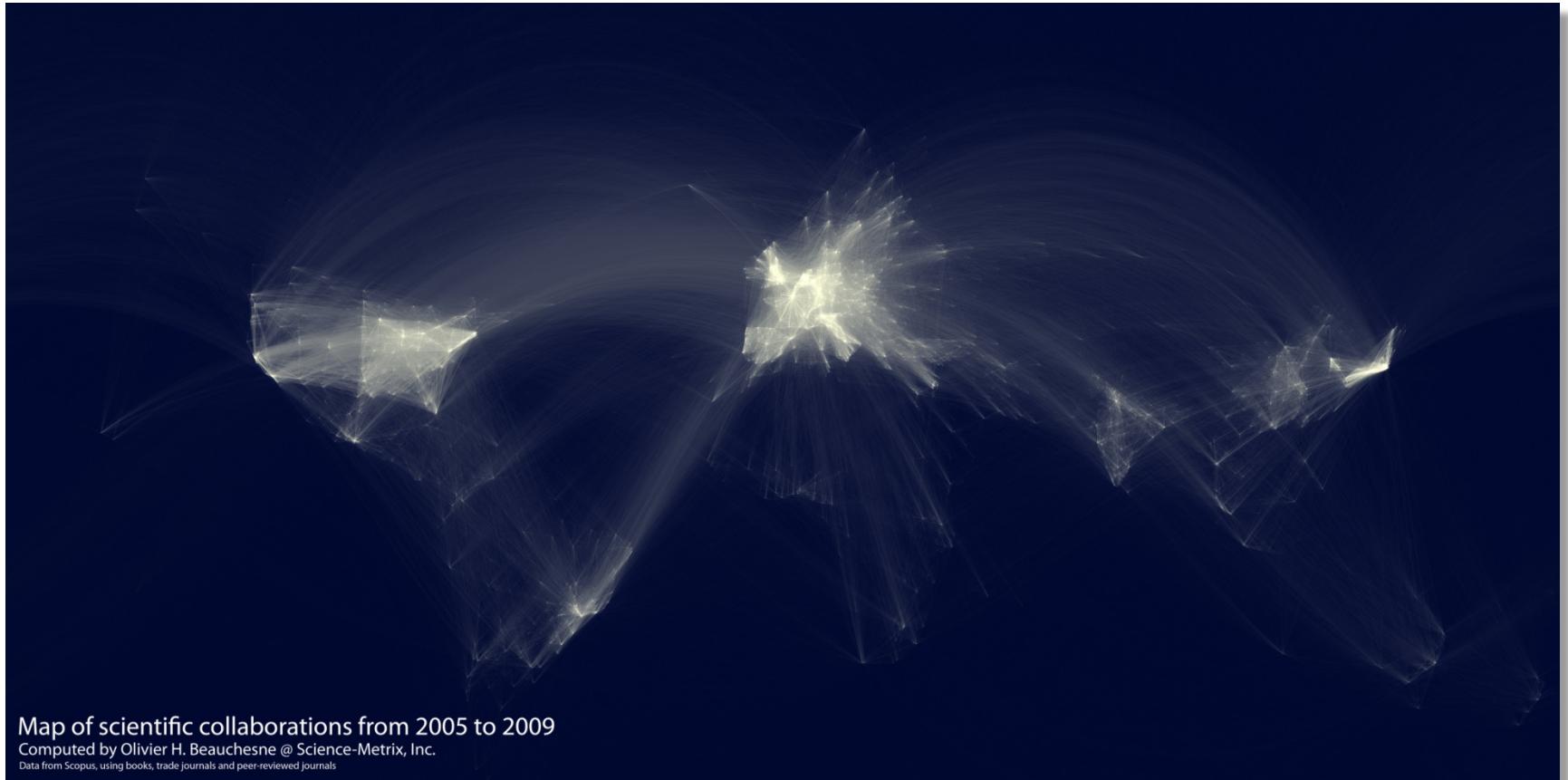
# Comunicación de resultados



[http://www.nytimes.com/interactive/2013/05/25/sunday-review/corporate-taxes.html?\\_r=0](http://www.nytimes.com/interactive/2013/05/25/sunday-review/corporate-taxes.html?_r=0)

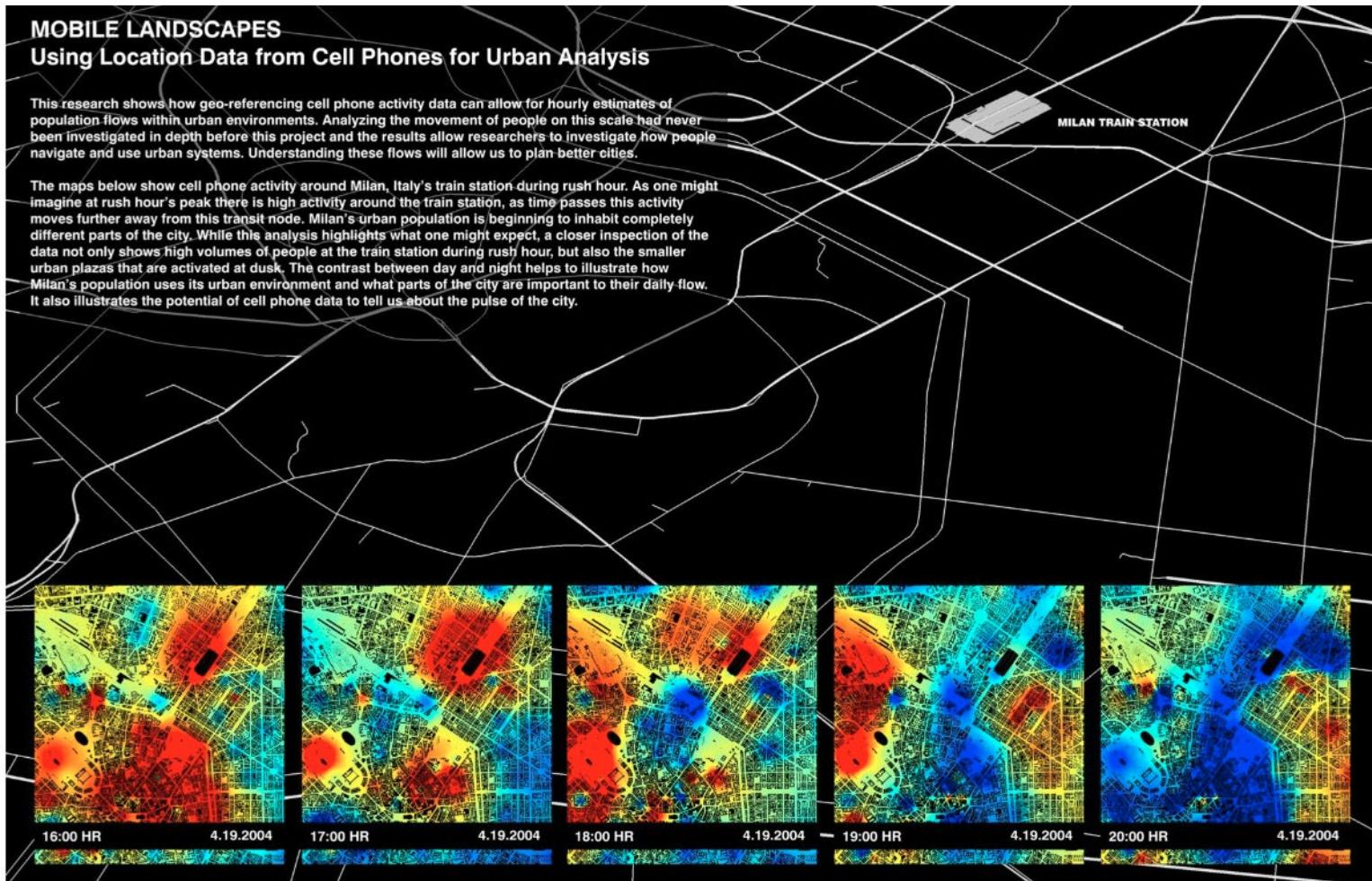
# Diseño orientado a preguntas

¿Dónde?



# Diseño orientado a preguntas

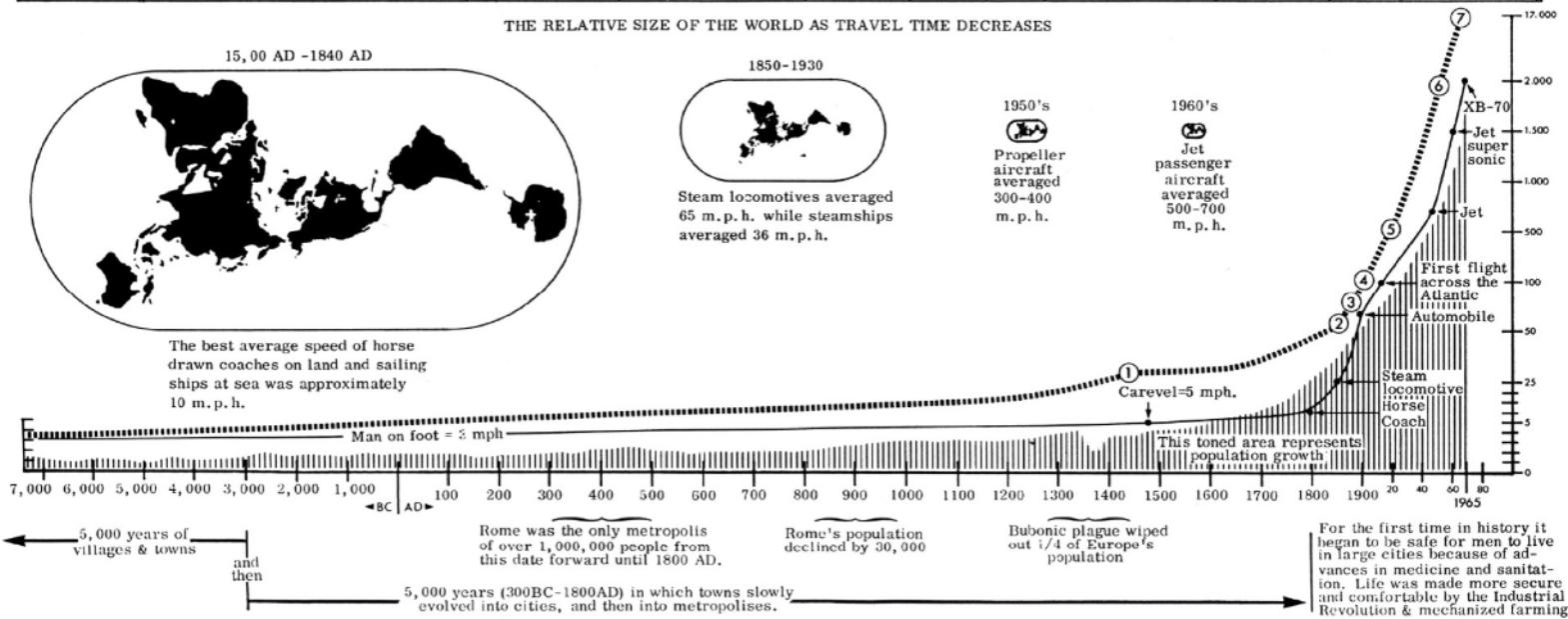
## ¿Dónde?



# Diseño orientado a preguntas

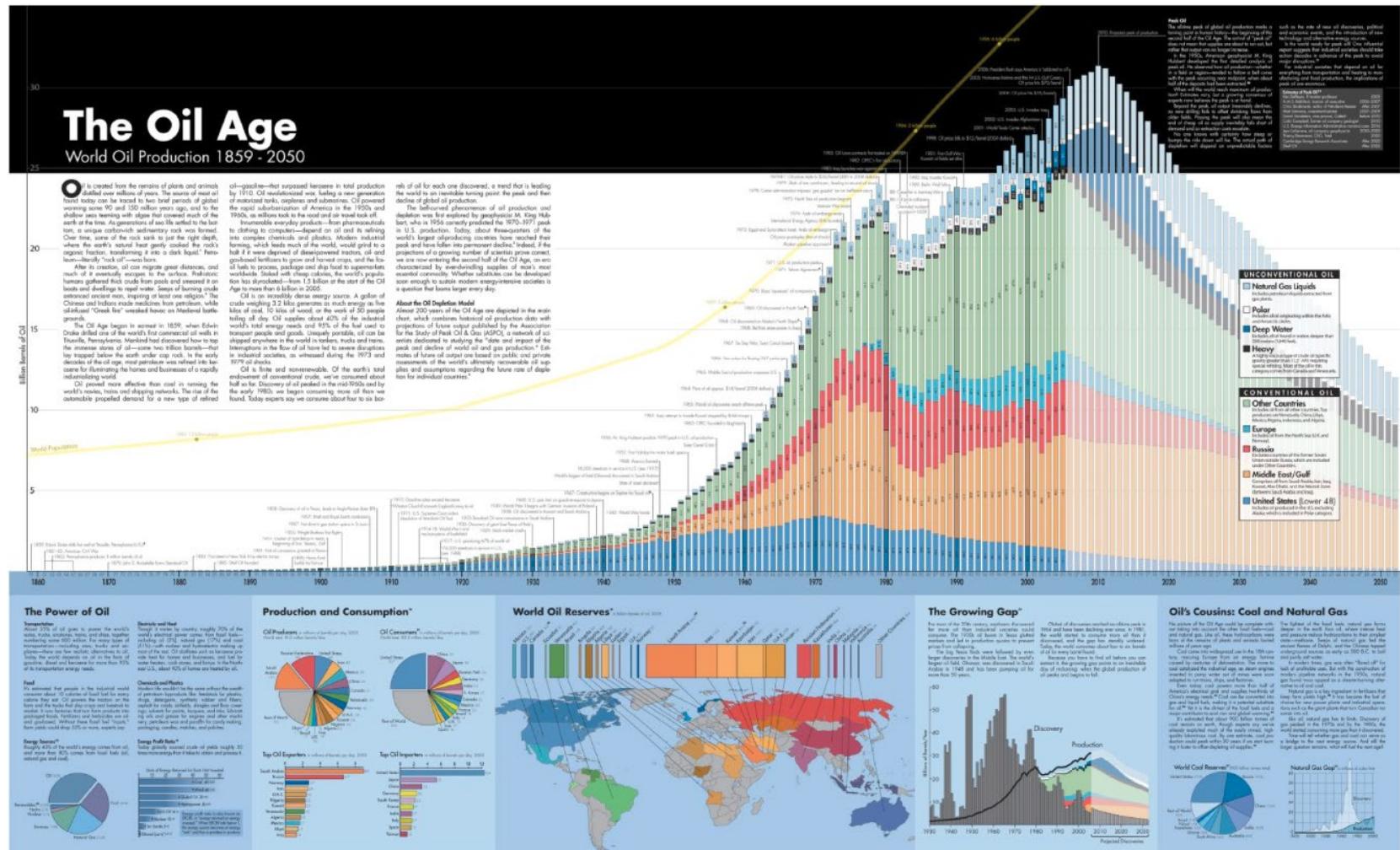
¿Cuándo?

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, railways, 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	Rail 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 printing press	② The rapid print Web newspaper press	③ The Bell telephone	④ The Marconi telegraph	⑤ First commercial radio broadcast	⑥ National television	⑦ Transcontinental T.V. with the introduction of Early Bird satellite	



# Diseño orientado a preguntas

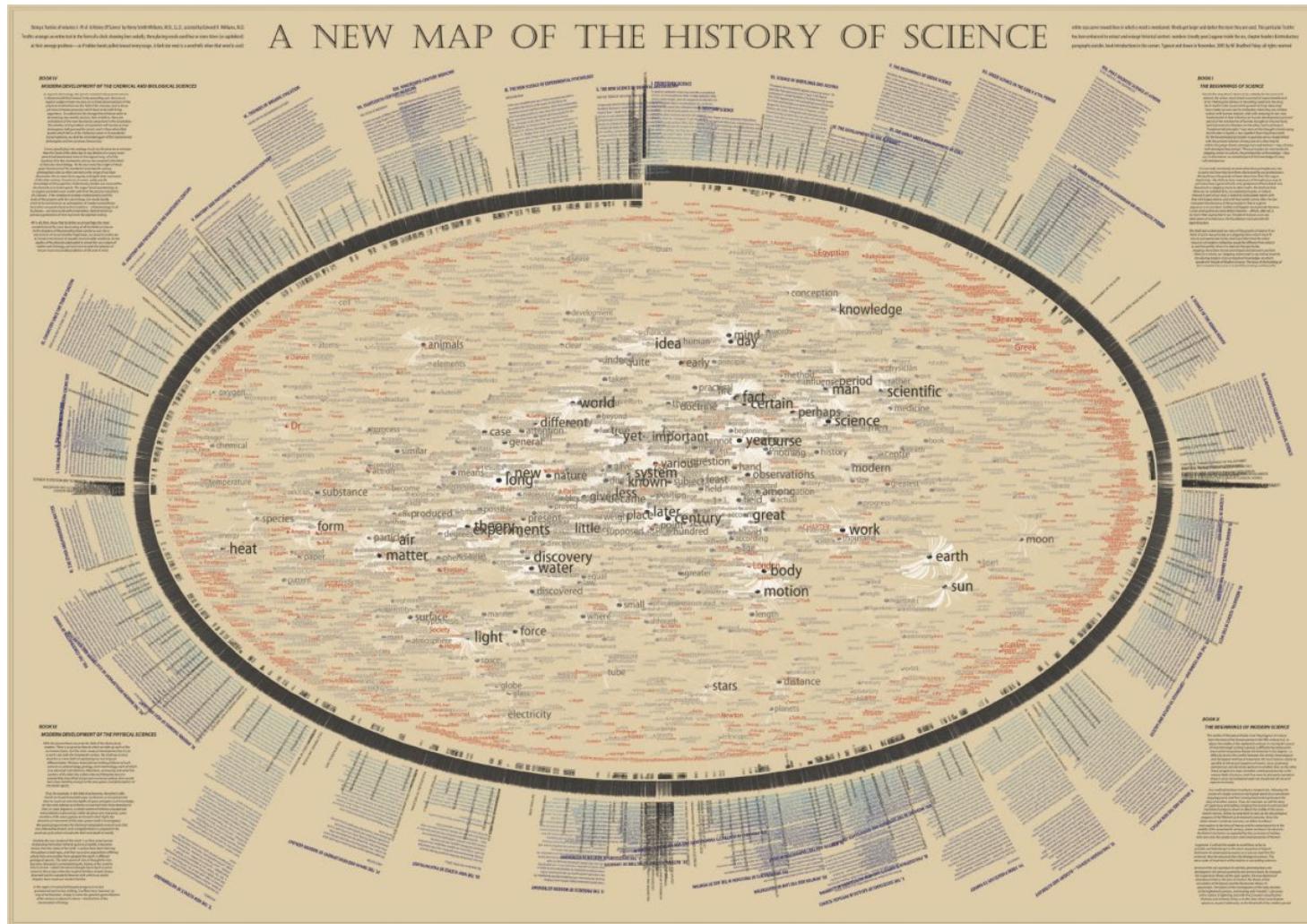
## ¿Cuándo?



# Diseño orientado a preguntas

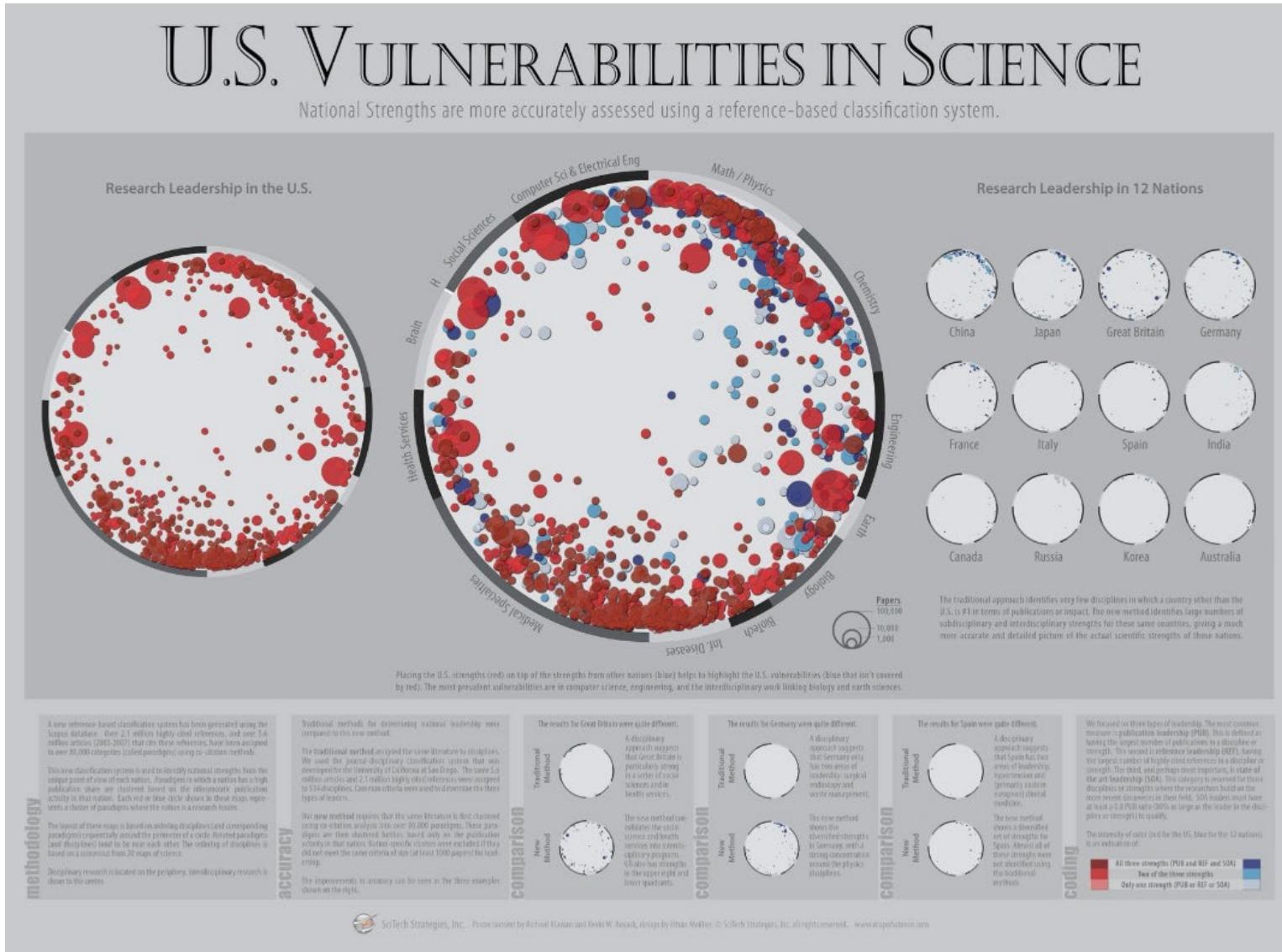
---

¿Qué?



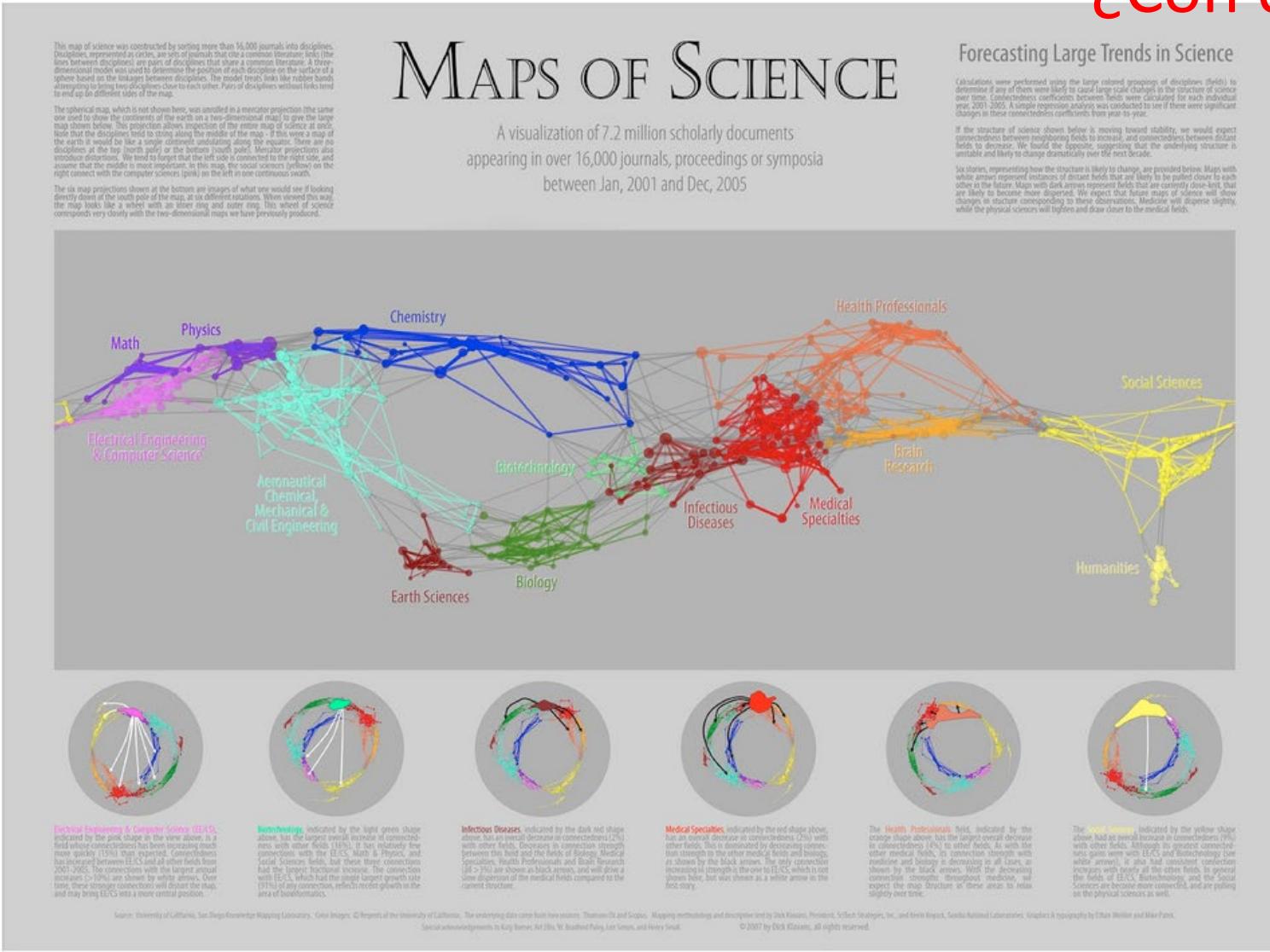
# Diseño orientado a preguntas

## ¿Qué?



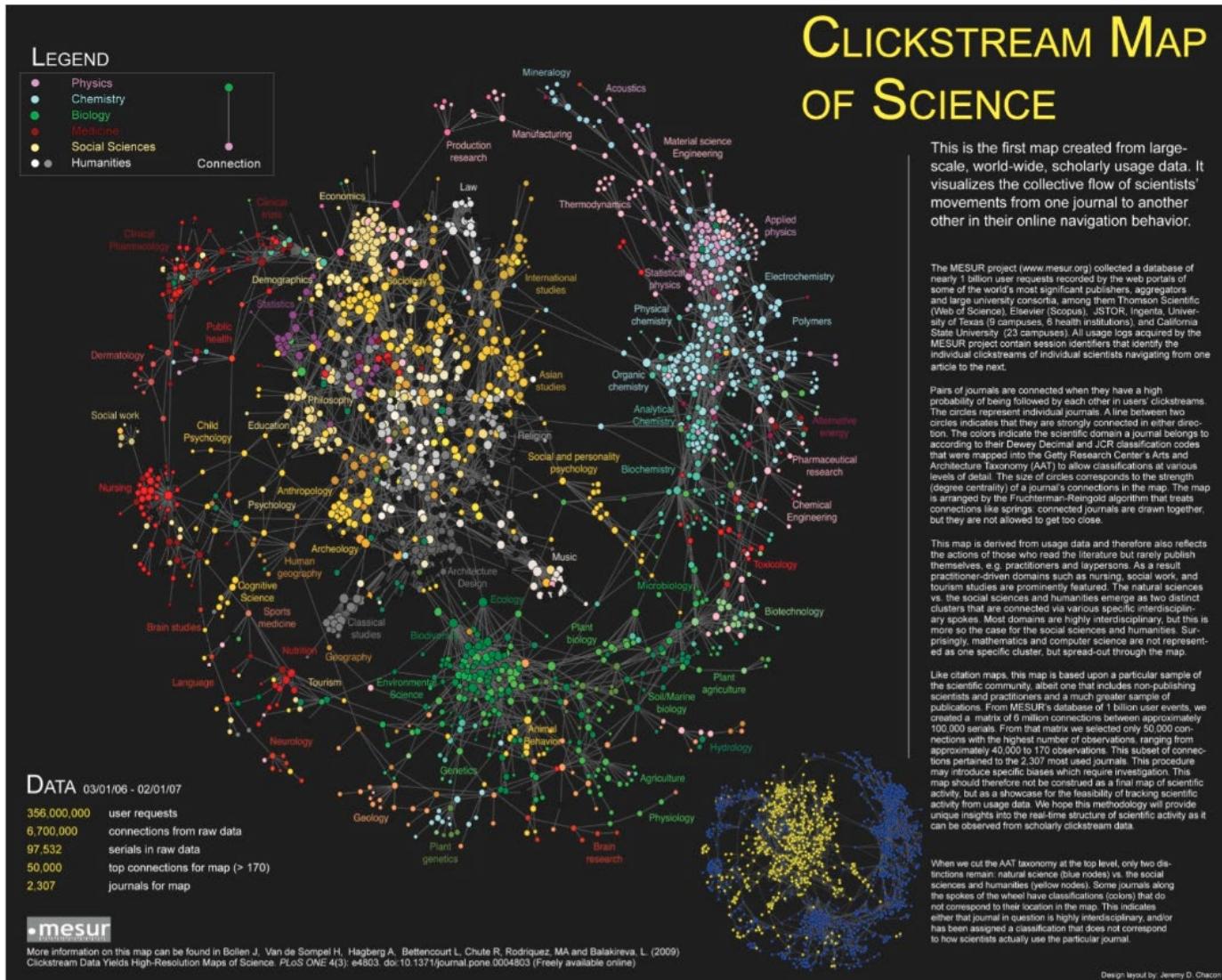
# Diseño orientado a preguntas

¿Con quién?

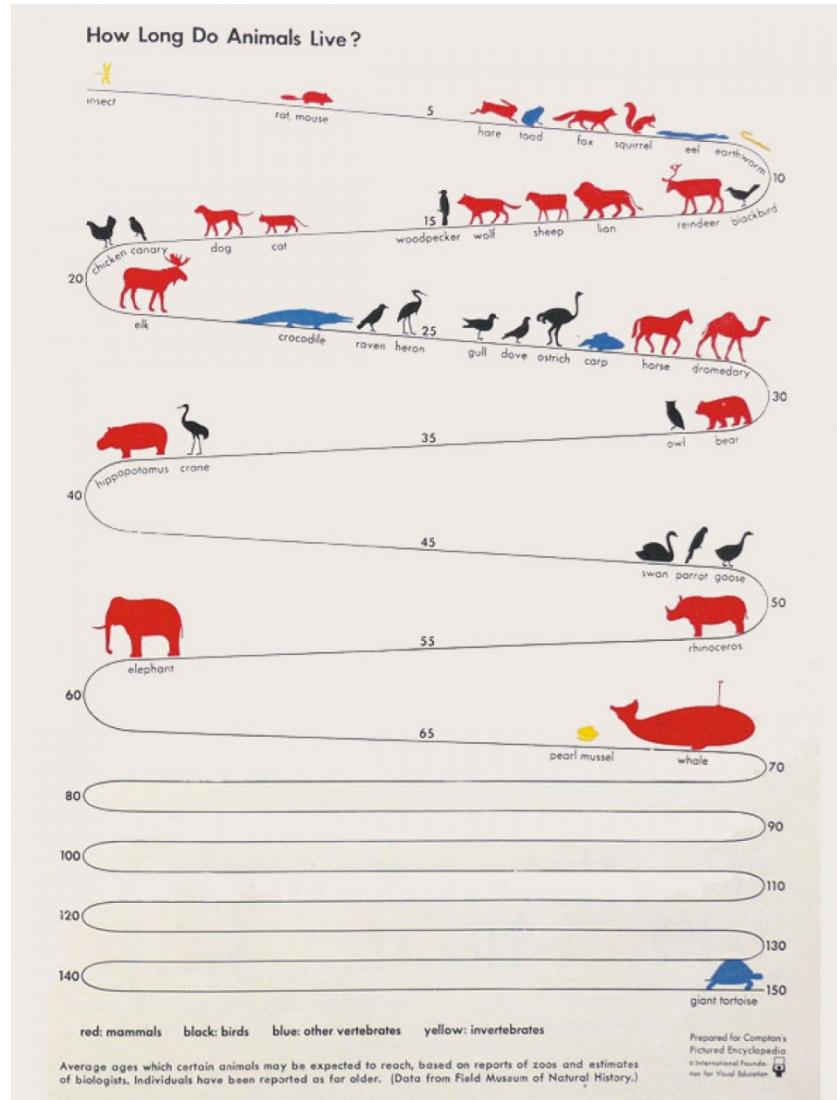


# Diseño orientado a preguntas

¿Con quién?



# La representación exige imaginación



# Ayuda

---

- ▷ Sobre malos ejemplos: [wtfviz.net](http://wtfviz.net)
- ▷ Para pedir directamente ayuda: [policyviz.com](http://policyviz.com)

# Bibliografía

---

- ▷ K. Börner and D.E. Polley. *Visual insights*. The MIT Press, 2014.
- ▷ A. Cairo. *El arte funcional*. Alamut, 2011.
- ▷ S. Kandel, J. Heer, C. Plaisant, et al. Research directions in data wrangling: visualizations and transformations for usable and credible data. *Information visualization*, 10(4):271-288, 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.