

Visualización de datos

Tema 3: Rigor



Paulo Félix Lamas

Área de Ciencias da Computación e Intelixencia Artificial

Departamento de Electrónica e Computación

Una frase

“Contempt for graphics and their audience, along with the lack of quantitative skills among illustrators, has deadly consequences for graphical work: over-decorated and simplistic designs, tiny data sets, and big lies.”

E.R. Tufte

Resumen

- ▷ Este tema aborda el problema del rigor en la visualización, y las diferentes formas de sortearlo diseñando gráficos que conducen a interpretaciones falsas o engañosas.
- ▷ Pero también es un tema sobre la percepción, ya que se ponen de manifiesto las dificultades que tenemos los seres humanos para realizar un análisis cuantitativo a partir de una percepción visual.

Algunas observaciones importantes

Los seres humanos percibimos con facilidad:

1. La posición (x,y) en el plano.
2. Magnitudes como la longitud.

Los seres humanos percibimos con dificultad:

1. Magnitudes como el área, o el ángulo.
2. Un texto rotado.
3. Aquéllo que se solapa o superpone.
4. Diferencias entre más de 7 colores.
5. Diferencias entre más de 5 sombras.
6. La gradación de color.

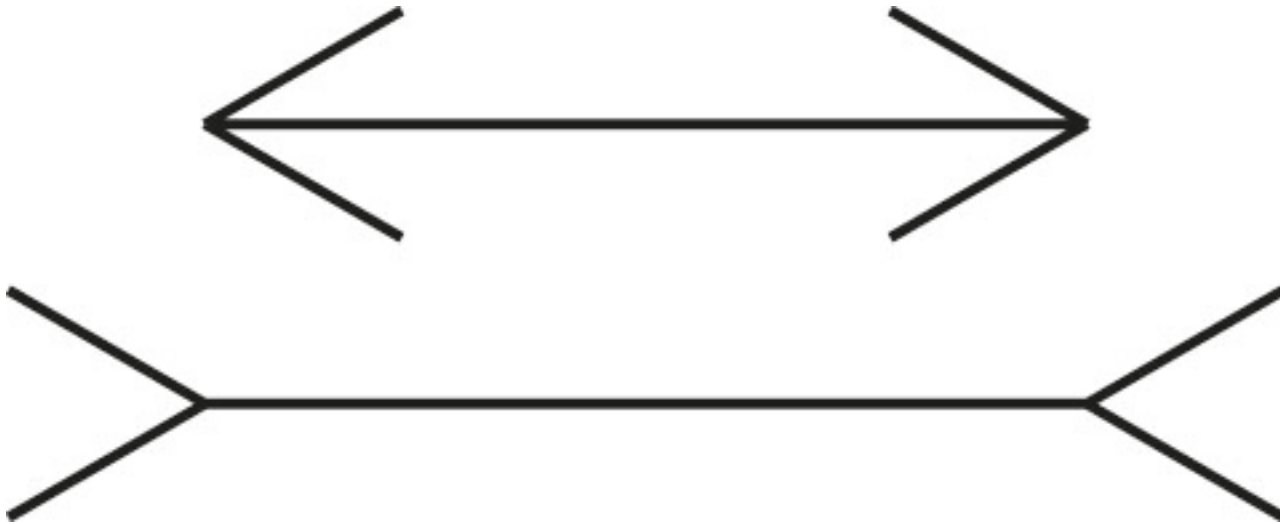
Características mensurables

No todas las características de un gráfico son igualmente discernibles o mensurables. Si las ponemos en orden éstas serían:

1. LONGITUD
2. ÁNGULO
3. ÁREA
4. COLOR
5. SOMBRA
6. VOLUMEN

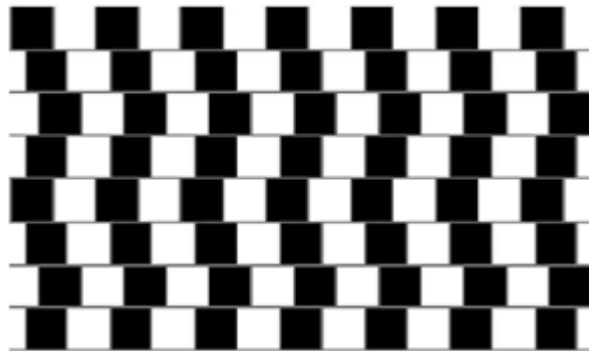
Veremos a continuación algunos problemas ocasionados por un uso deficiente de algunas de estas características.

Dificultades en la percepción visual



¿Cuál de los dos segmentos es el más largo?

Dificultades en la percepción visual



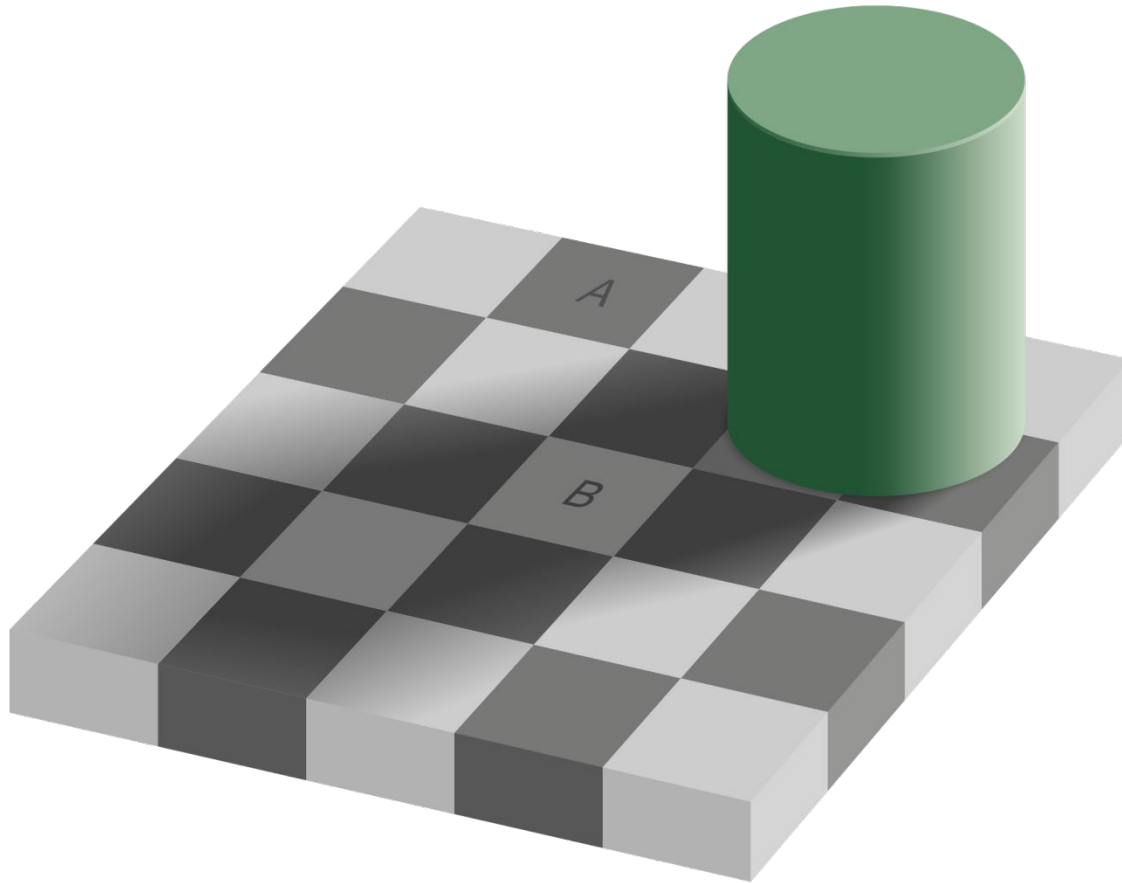
¿En qué sentido parecen inclinarse las series de cuadrados?

Dificultades en la percepción visual



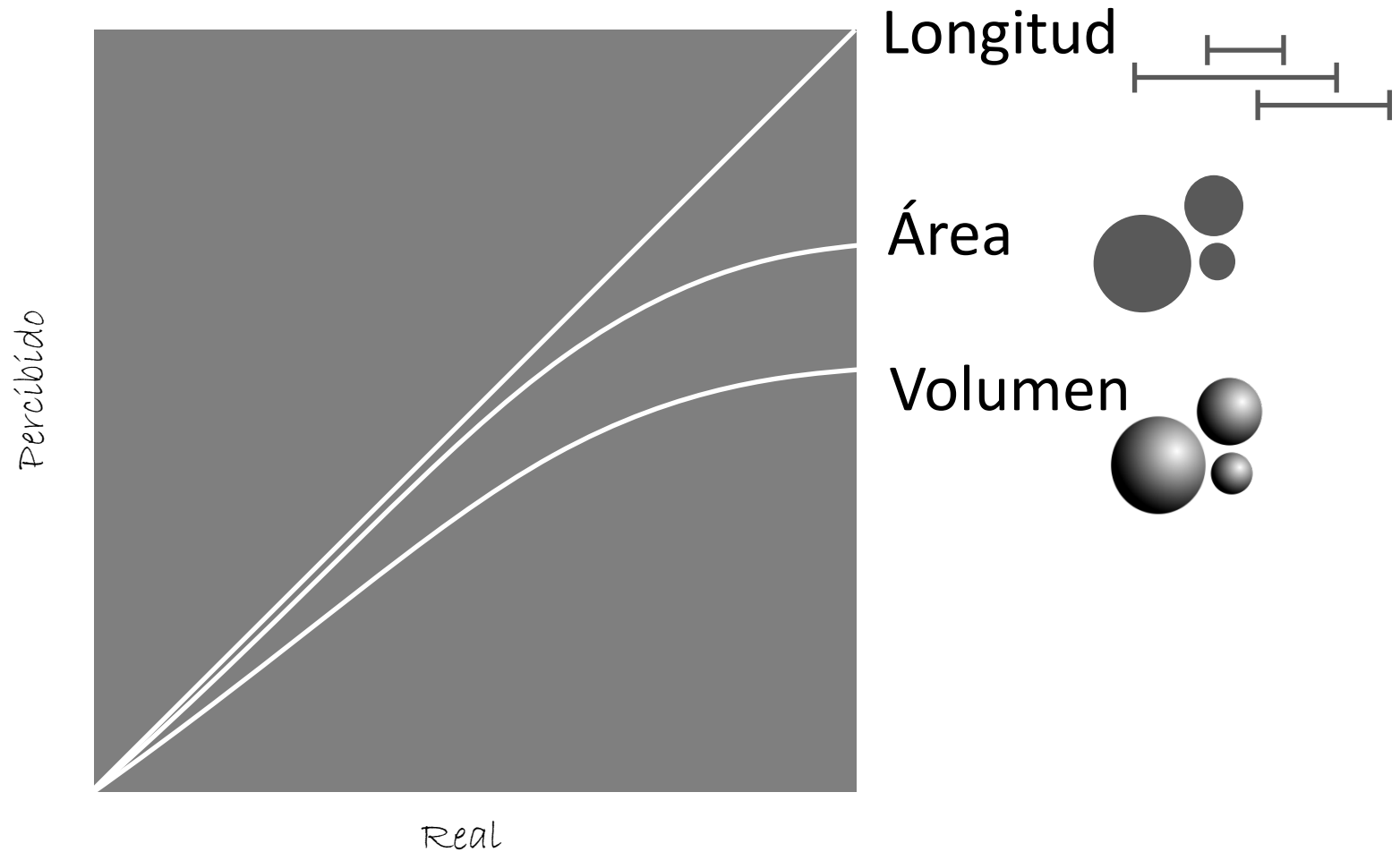
De los dos grupos de círculos, ¿cuál de los dos círculos del medio es más grande?

Dificultades en la percepción visual

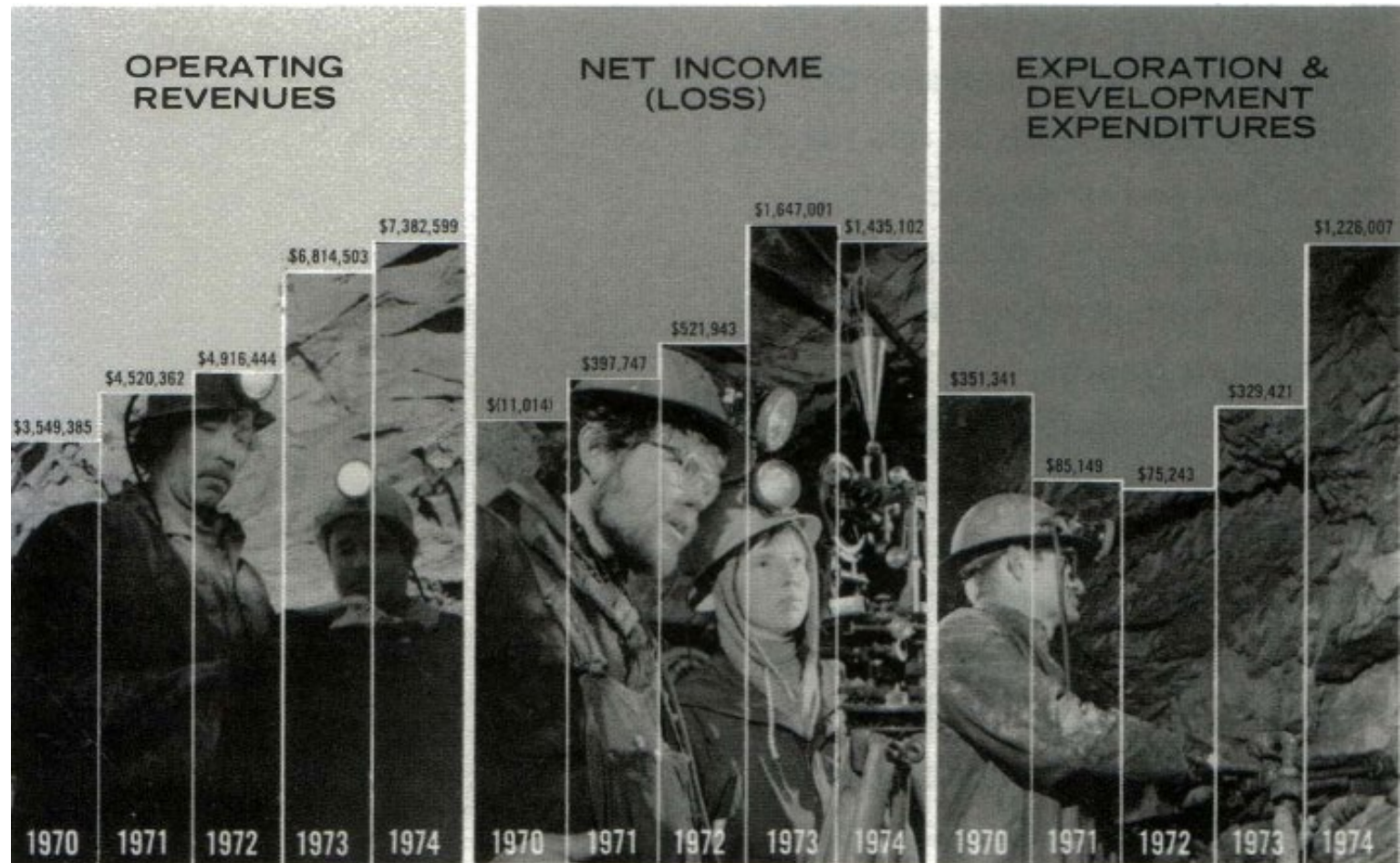


¿Cuál de los dos cuadrados, A o B, es más oscuro?

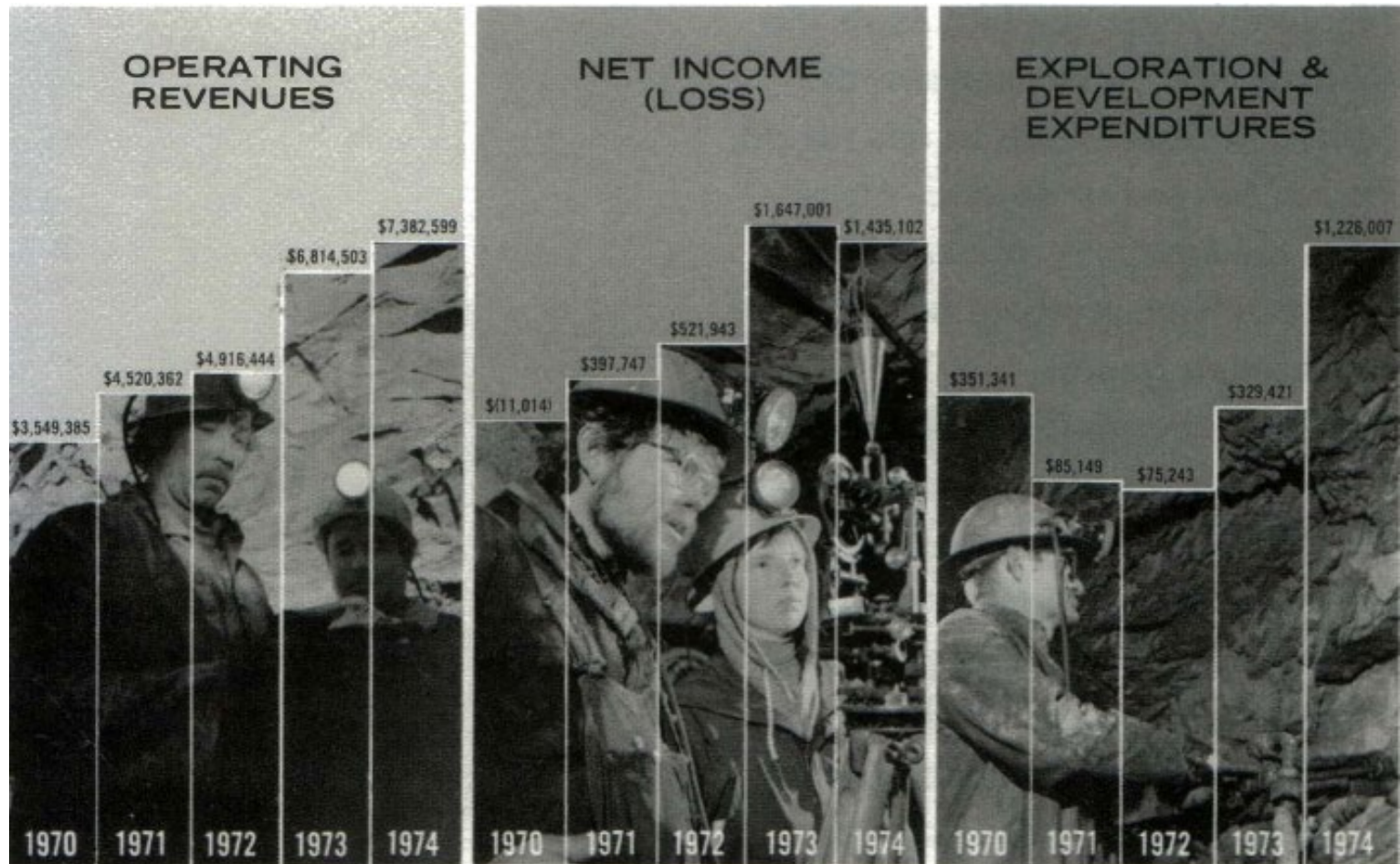
Estimación de la longitud, área y volumen



Falta de rigor

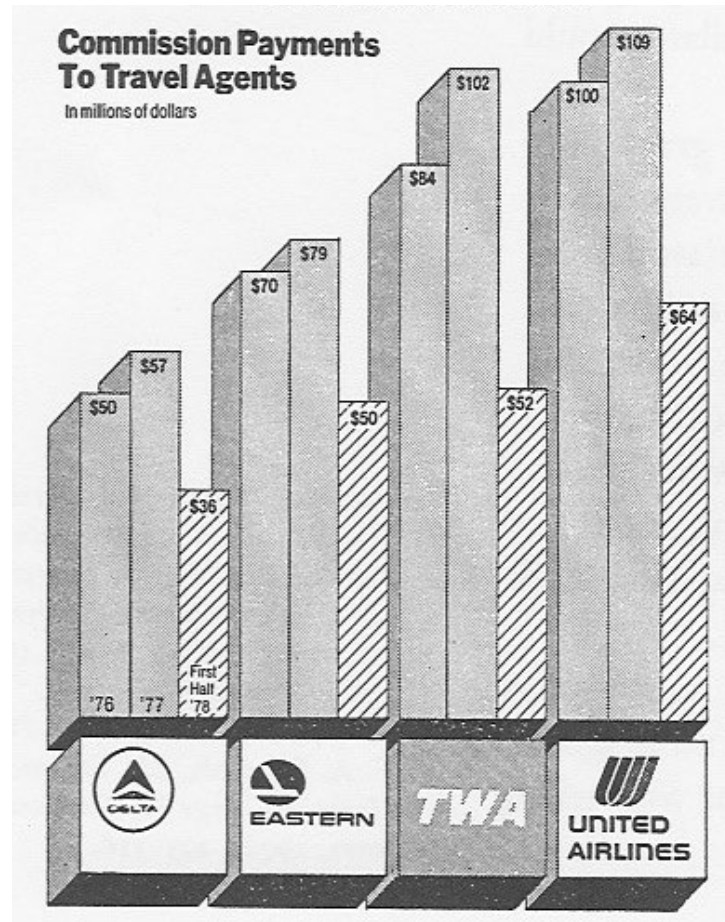


Falta de rigor



Se observa que en la figura central los ingresos de 1970 son negativos, y cómo se ha disfrazado situando el comienzo de las barras en varios millones en negativo.

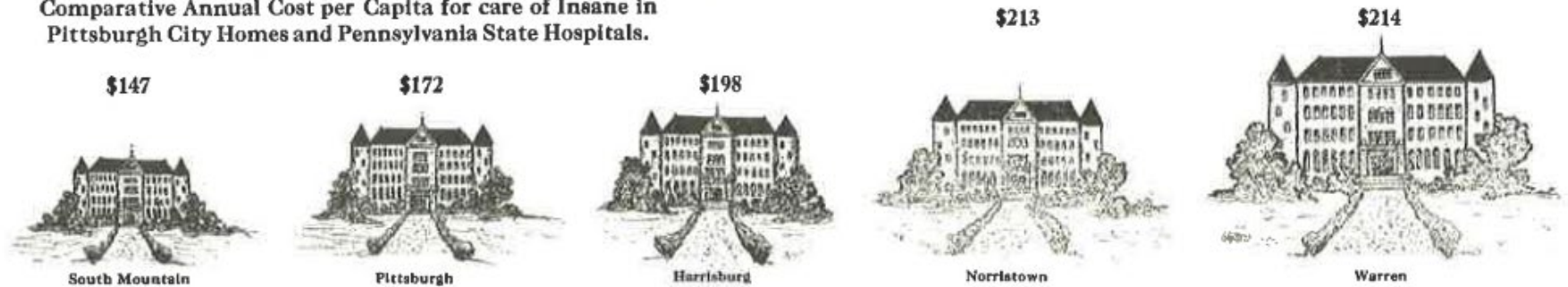
Falta de rigor



Nótese cómo se compara el primer semestre del año 1978 con los datos anuales de 1976 y 1977, induciendo a pensar en una disminución de las comisiones.

Falta de rigor

Comparative Annual Cost per Capita for care of Insane in
Pittsburgh City Homes and Pennsylvania State Hospitals.



Nótese la falta de correlación entre los costes y el tamaño supuestamente proporcional de la figura.

Falta de rigor

Valor de mercado de los principales bancos

En miles de millones de dólares

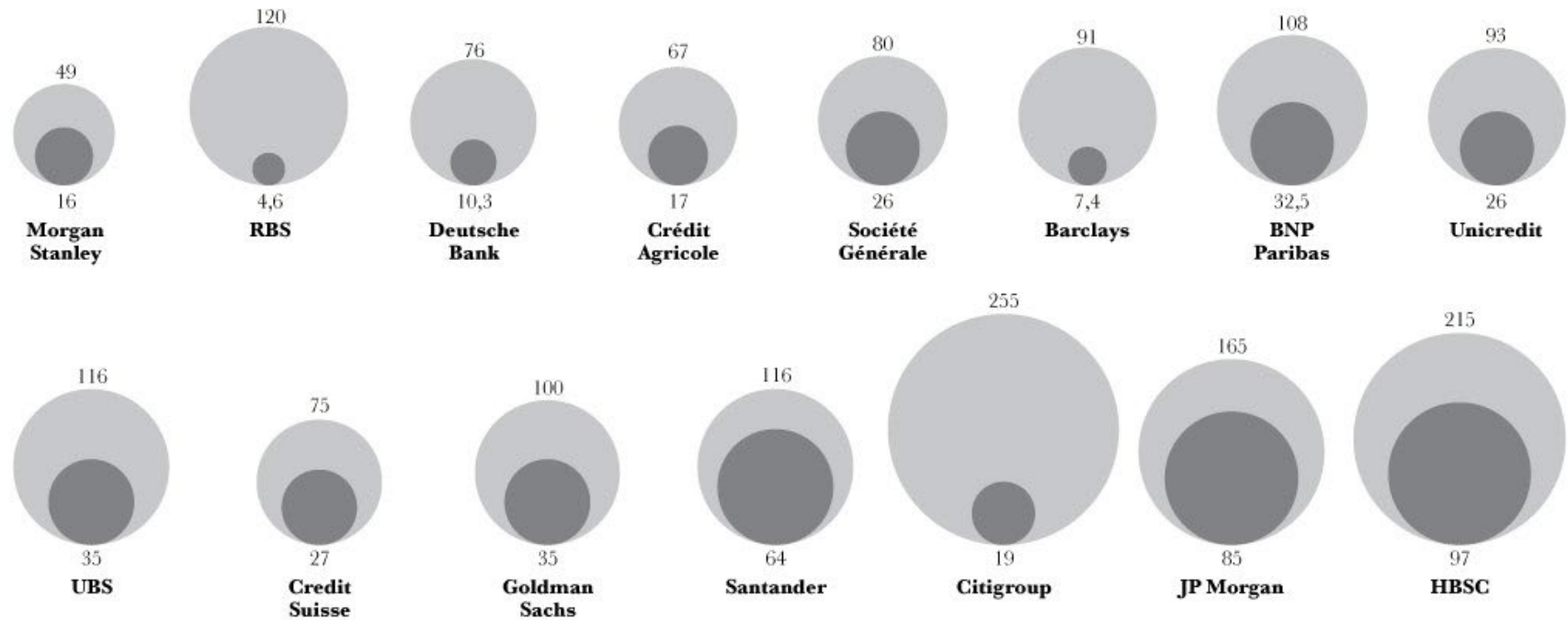


Enero de 2007



Enero de 2009

Fuente: Bloomberg



¿Qué noción transmiten las proporciones?.

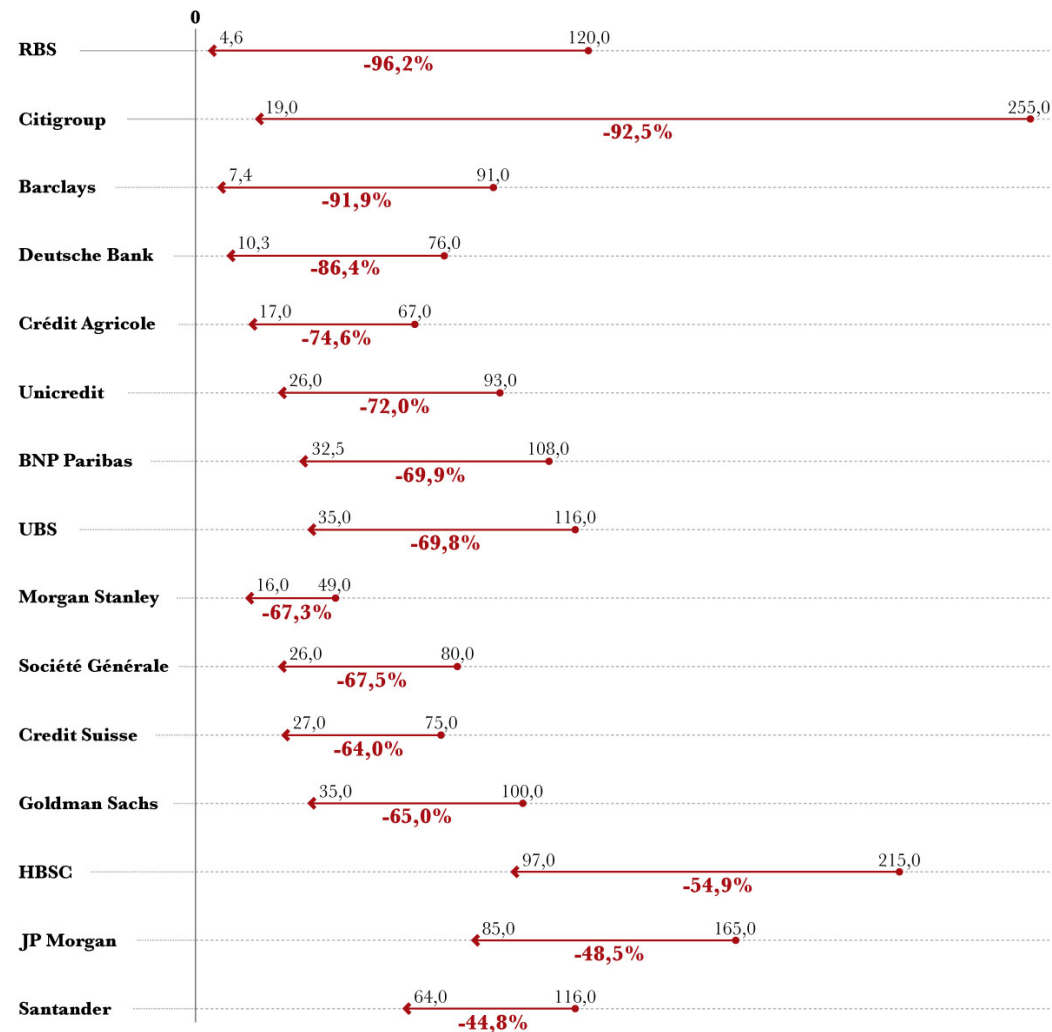
Rigurosamente

Valor de mercado de los principales bancos

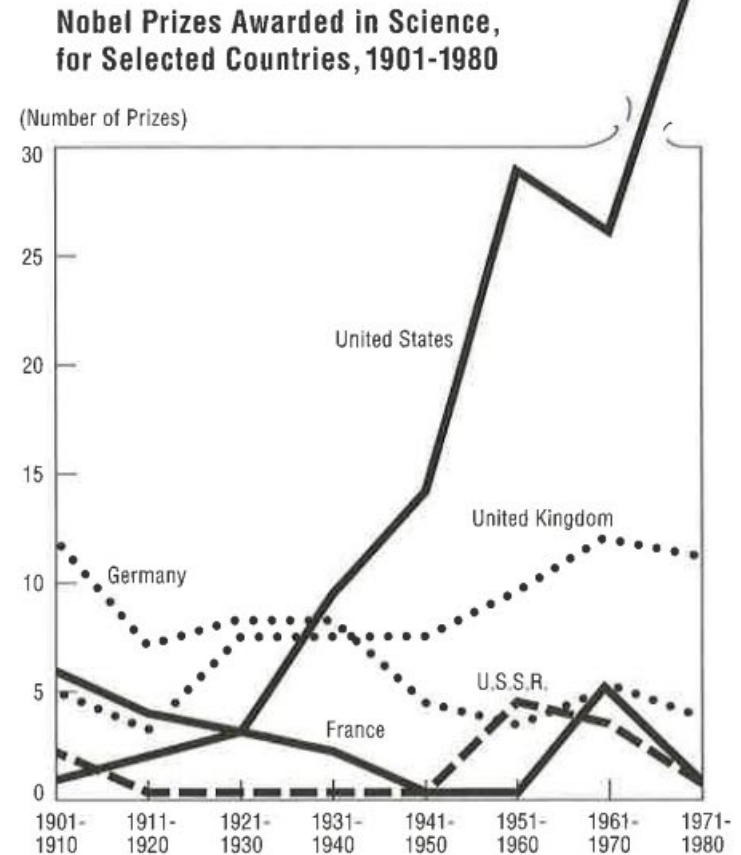
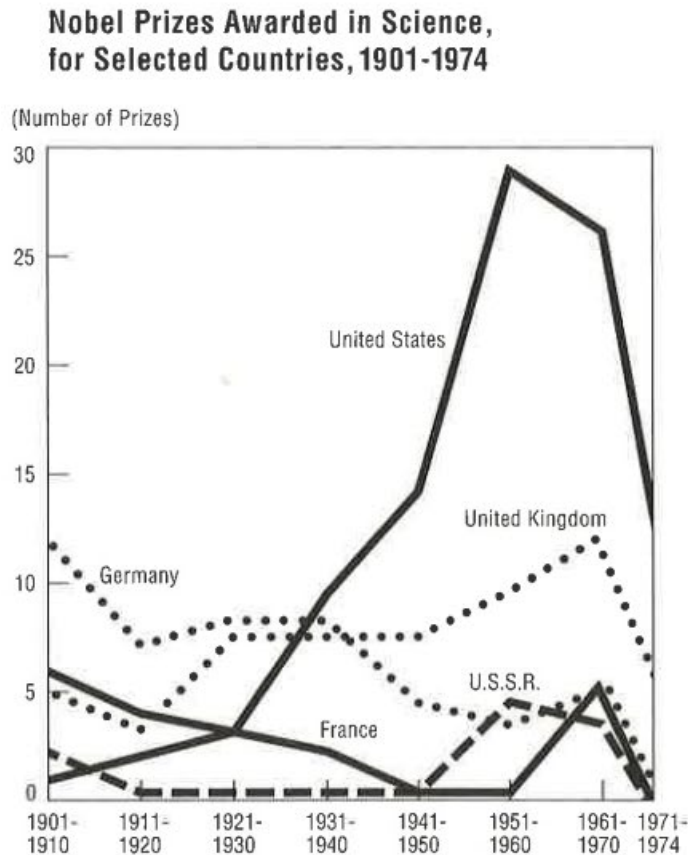
En miles de millones de dólares

Enero de 2009 ← Enero de 2007

Fuente: Bloomberg

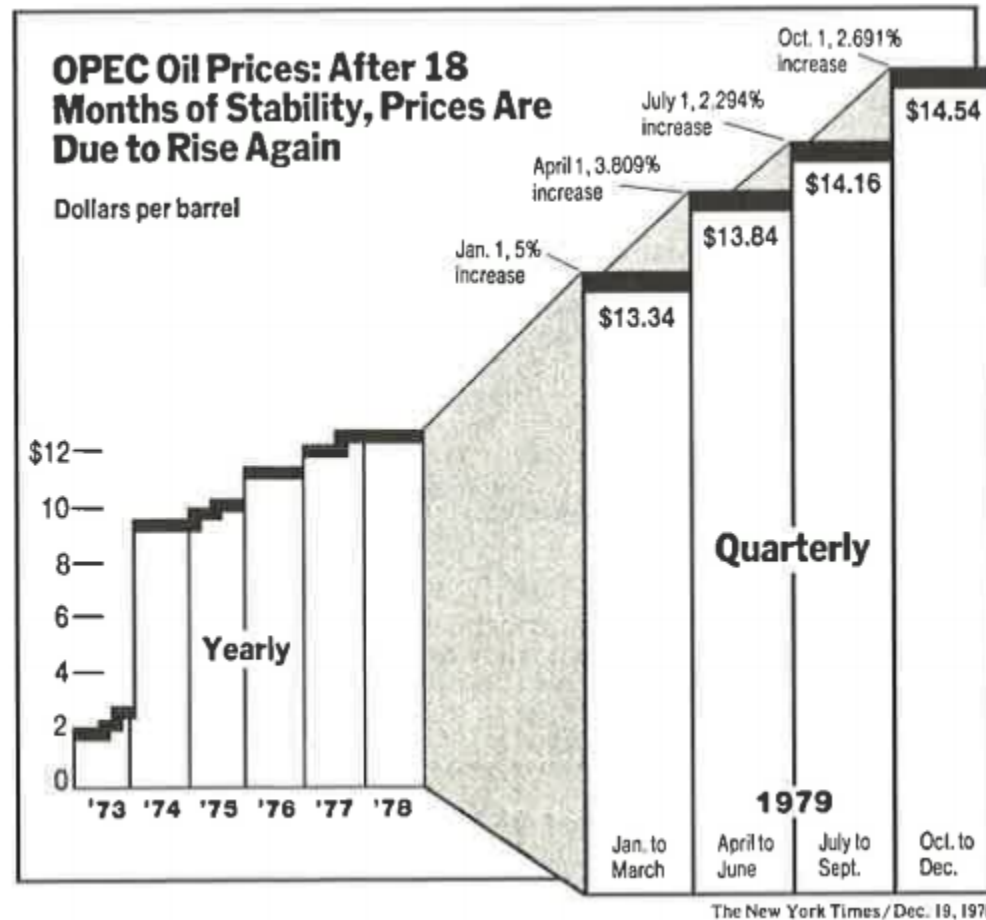


Falta de rigor



un "error" recurrente en la representación de gráficos temporales resulta de comparar periodos de distinta duración

Falta de rigor



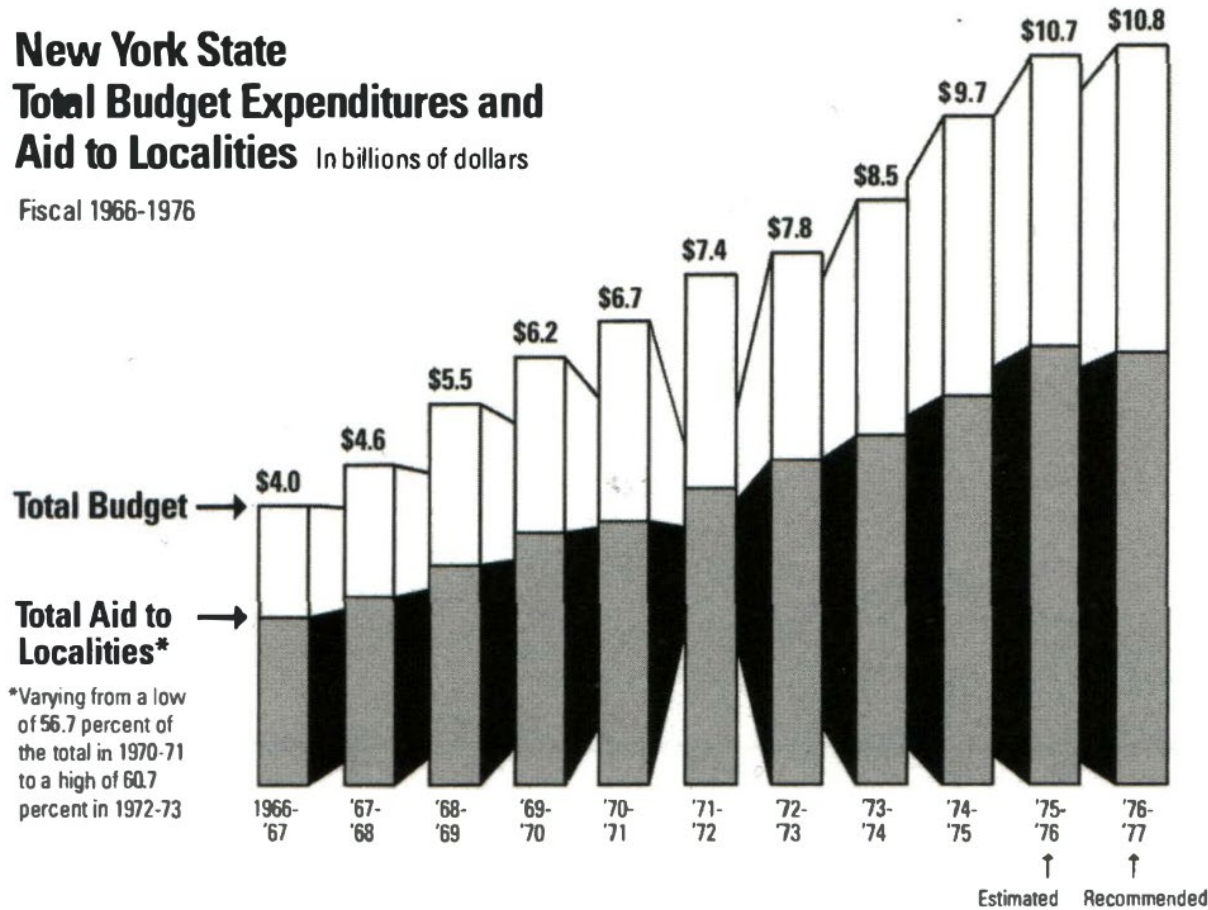
una tentación muy recurrente es la de utilizar un cambio de escala para enfatizar una idea, o un titular...

Falta de rigor

New York State Total Budget Expenditures and Aid to Localities

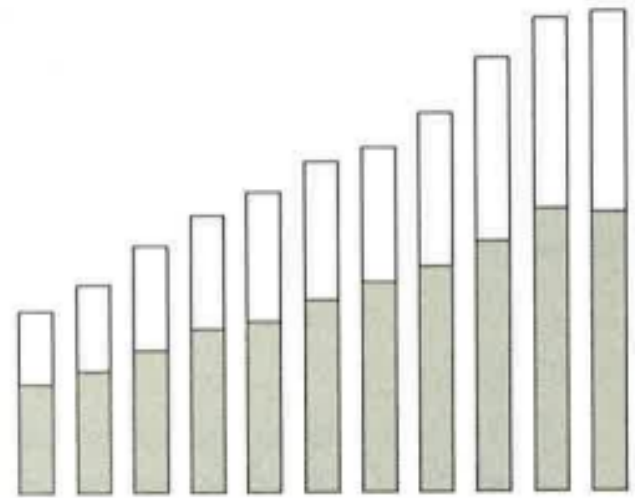
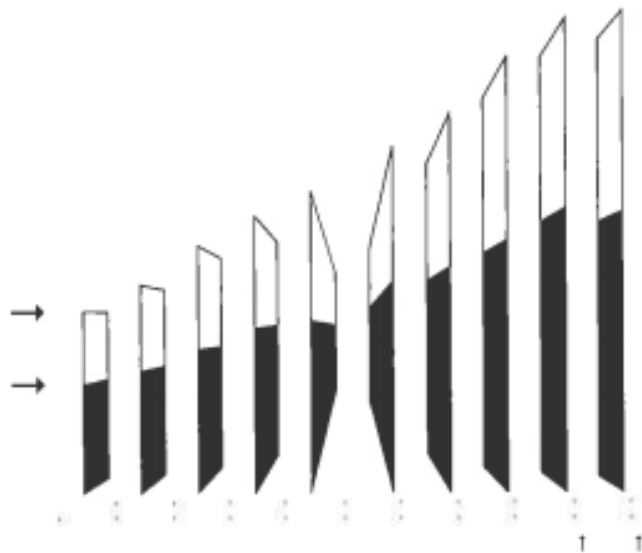
In billions of dollars

Fiscal 1966-1976



Se utiliza la perspectiva para acentuar un efecto y una interpretación interesada de los datos.

Falta de rigor



Nótese la diferencia entre la información contenida en las perspectivas y la información contenida en los niveles.

Falta de rigor



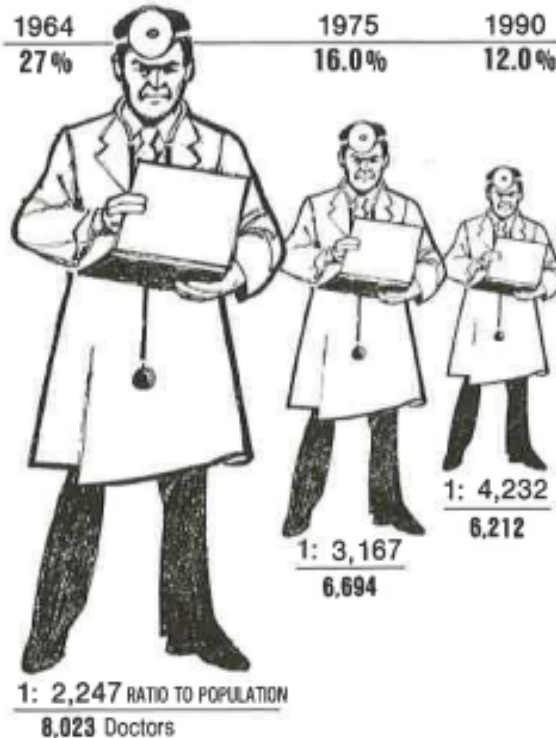
Y ahora los datos en términos relativos a la población y con la actualización de la inflación.

Falta de rigor

THE SHRINKING FAMILY DOCTOR In California

Percentage of Doctors Devoted Solely to Family Practice

1964	1975	1990
27 %	16.0 %	12.0 %



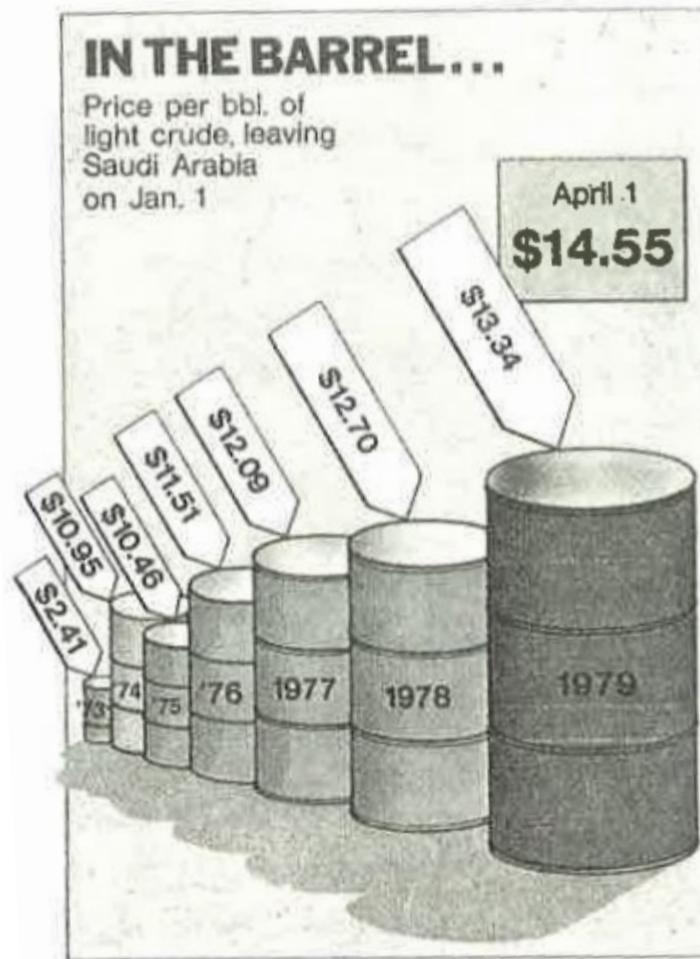
Aquí se utiliza una visualización en dos dimensiones para transmitir una información de una sola dimensión.... y hay otro truco.

Falta de rigor



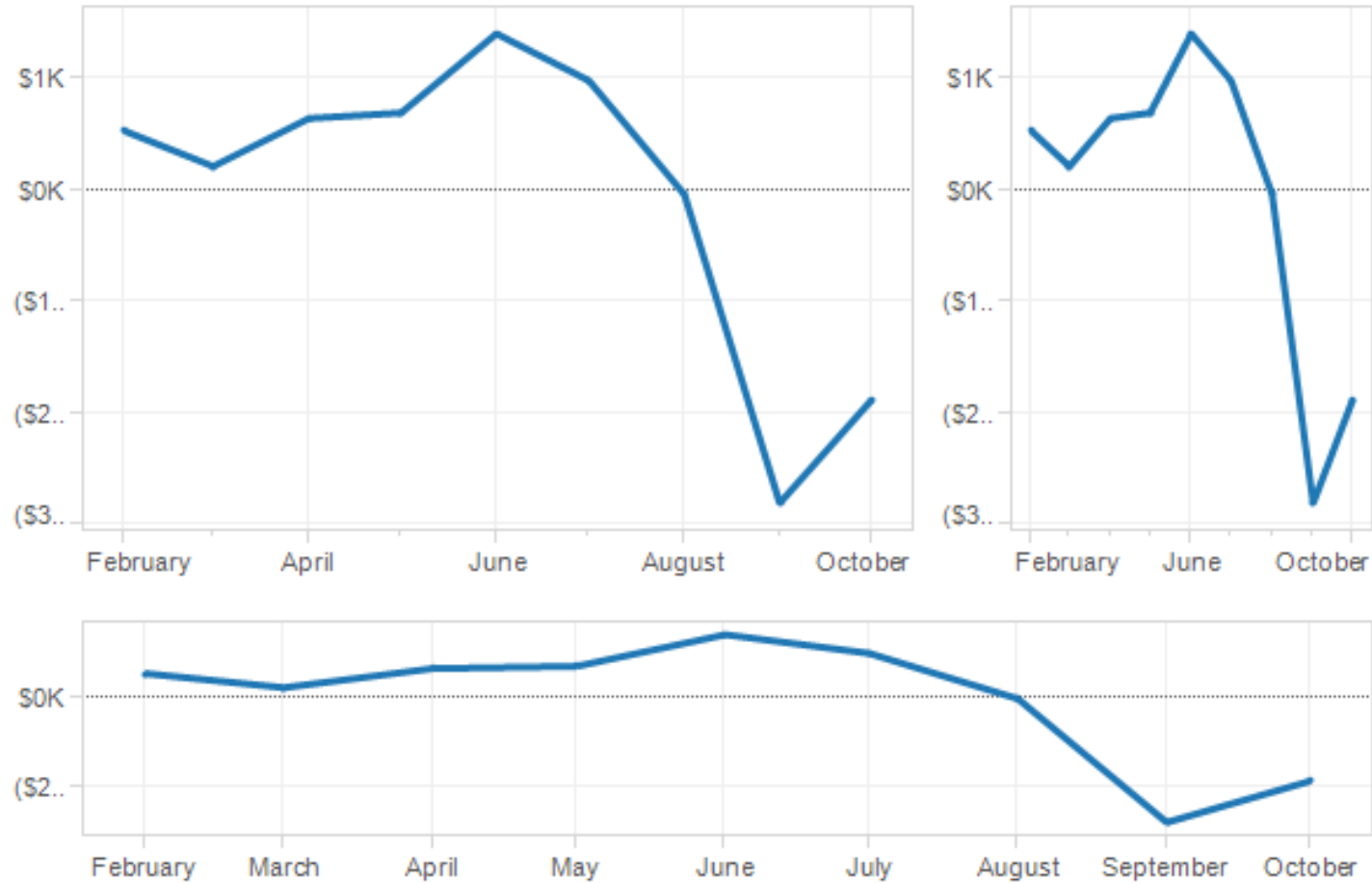
A veces se busca el efecto contrario: minimizar diferencias.

Falta de rigor

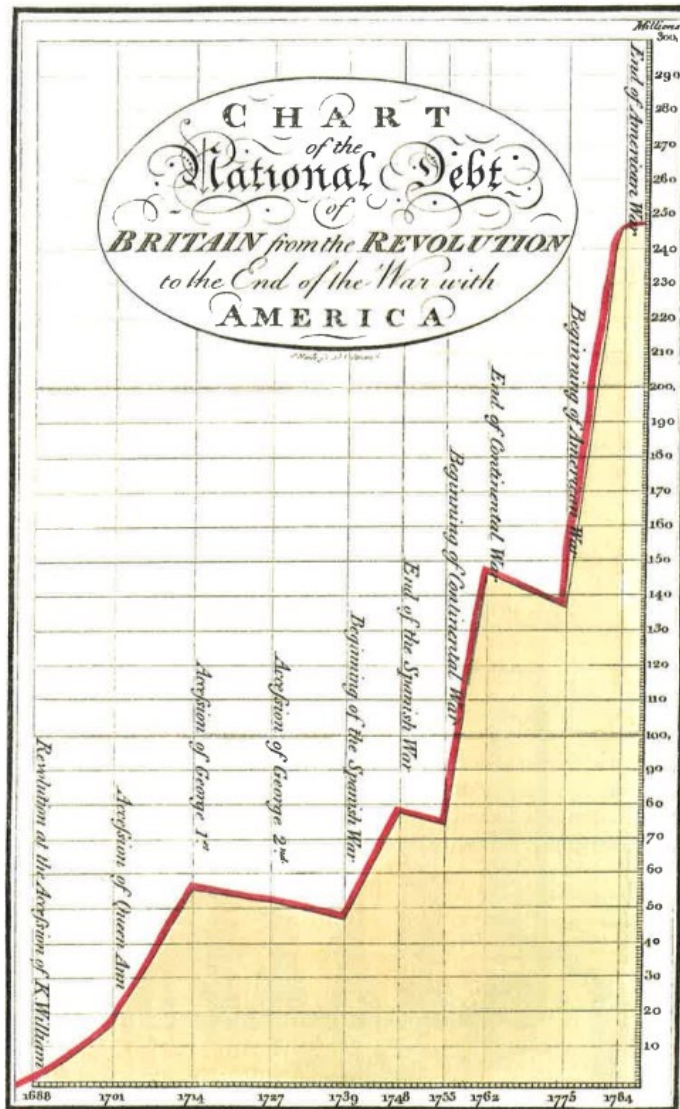


No es raro encontrarse gráficos donde se escala en una sola dimensión y las otras dimensiones multiplican el efecto.

El problema de la escala



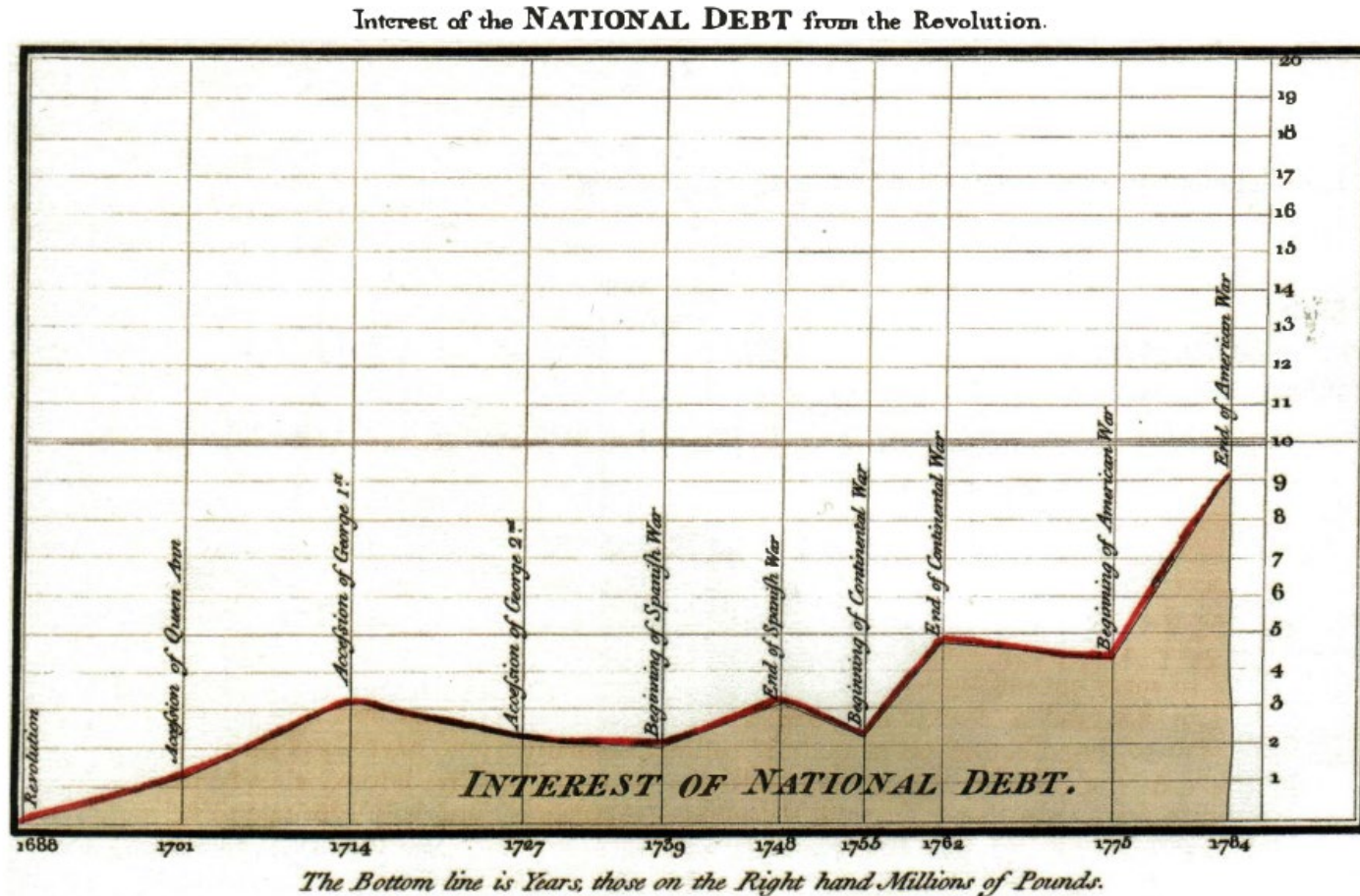
El problema de la escala



The Divisions at the Bottom are Years, & those on the Right hand Money.

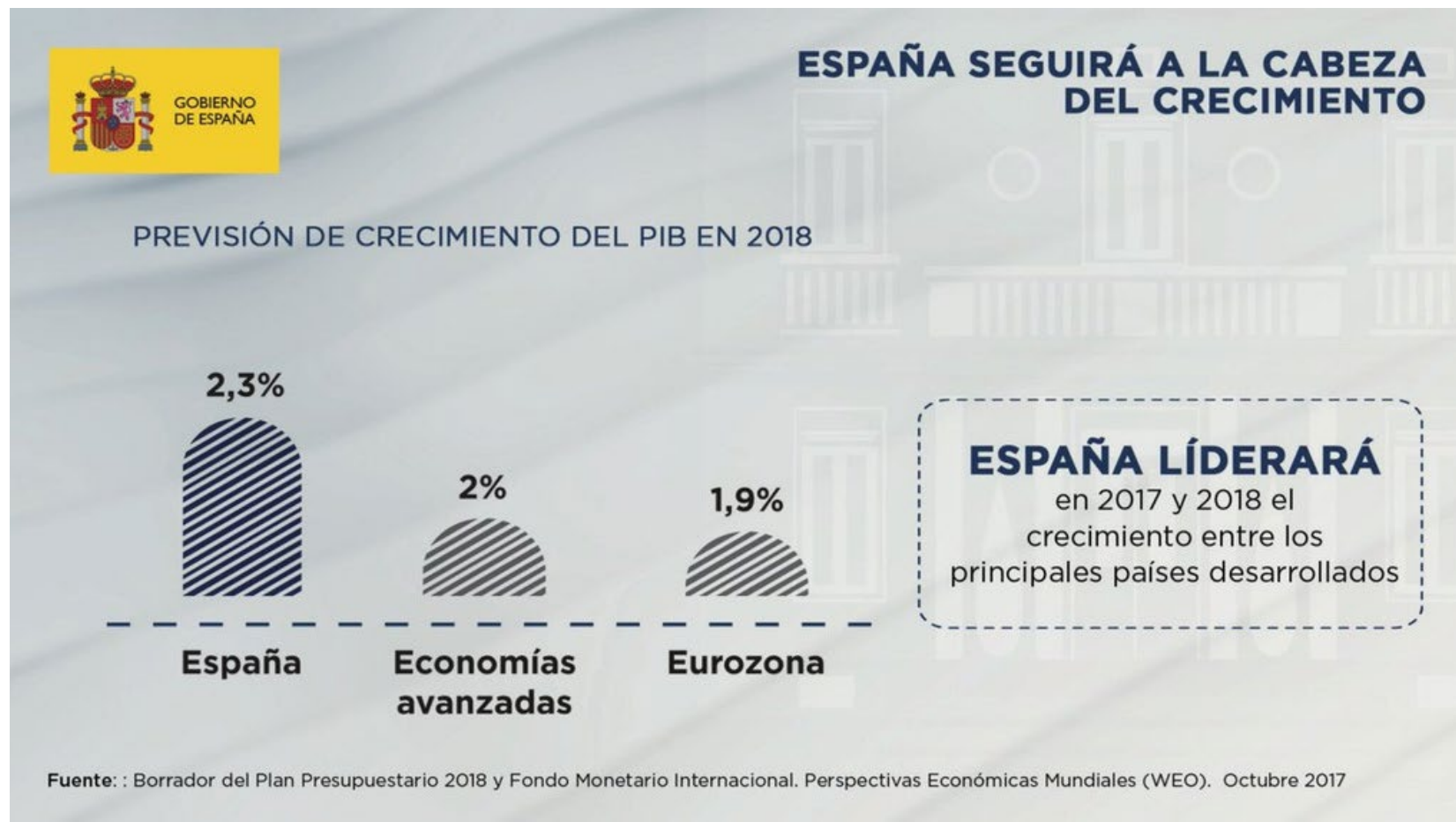
un clásico en los diagramas de costes, gastos, deuda,... sobre todo si es pública. Hacerlo acusadamente vertical, con una clara idea preconcebida.

El problema de la escala

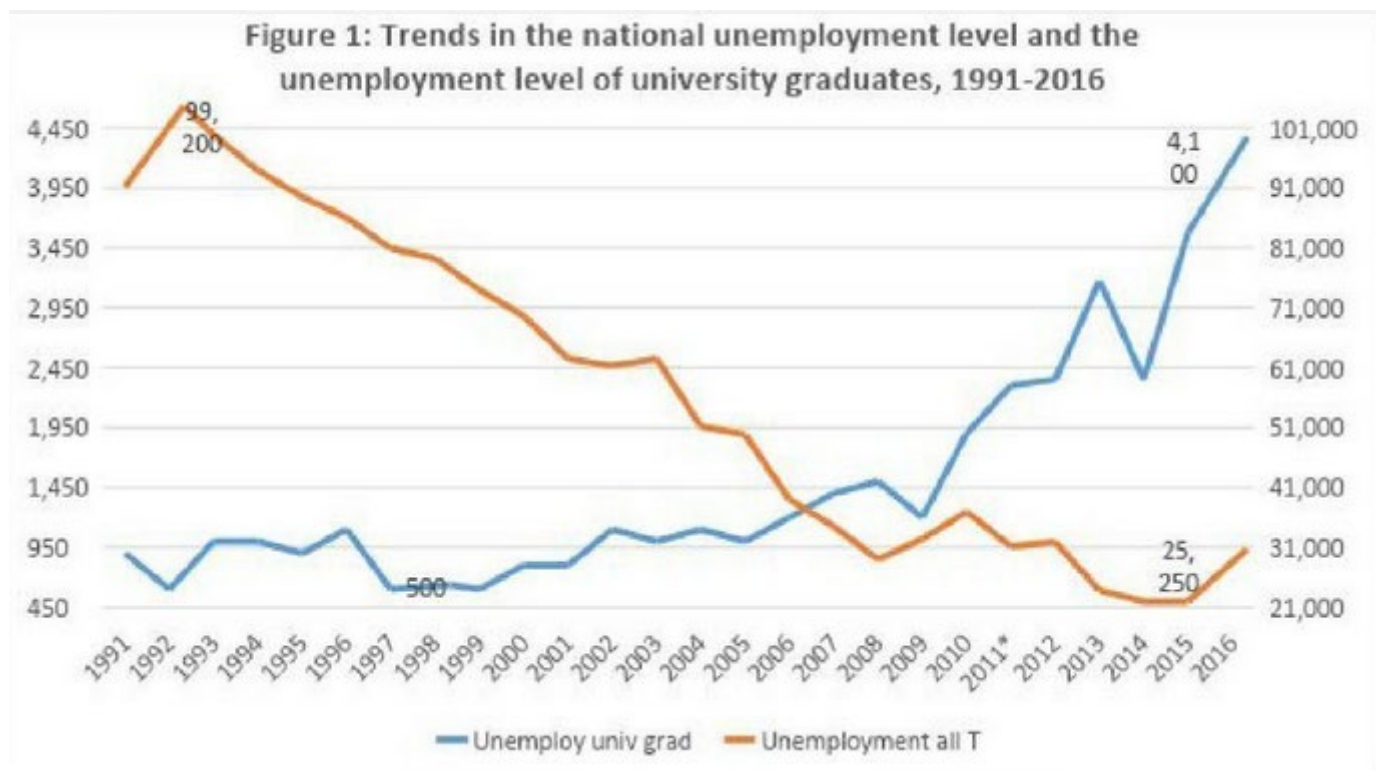


... aunque el mismo Playfair se corrigió después y publicó un gráfico con los mismos datos, pero más veraz

El problema de la escala

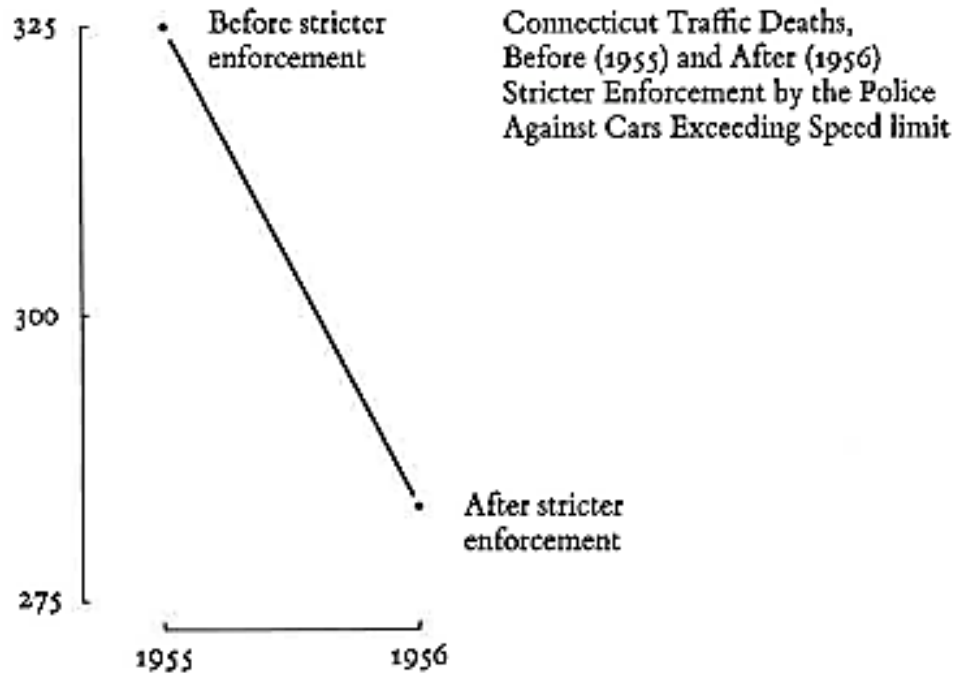


El problema de las escalas

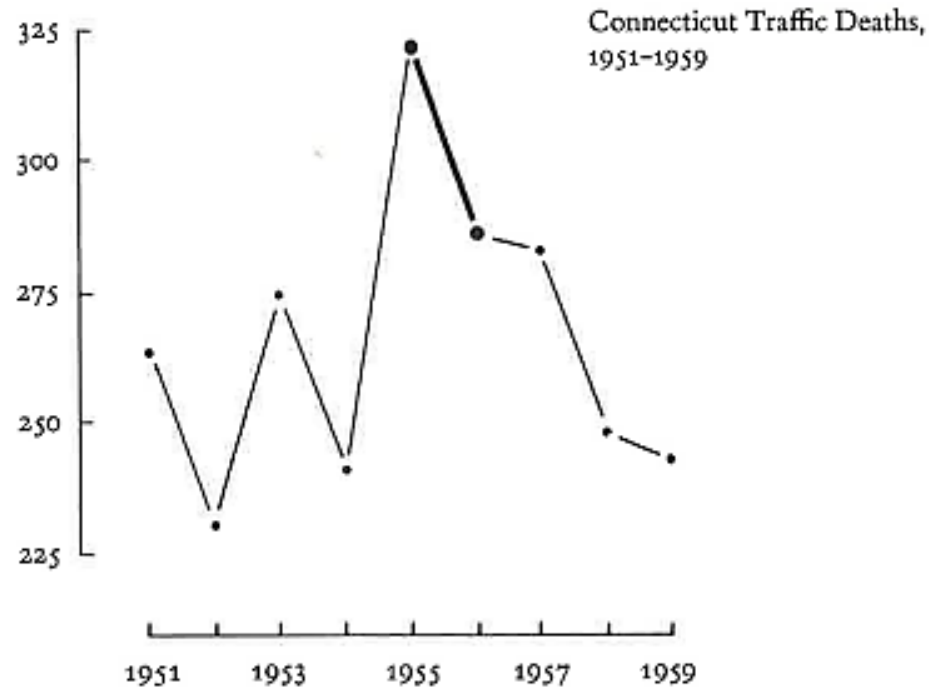


Tendencias en el desempleo de Trinidad y Tobago: población general y titulados universitarios.

Los datos en su contexto

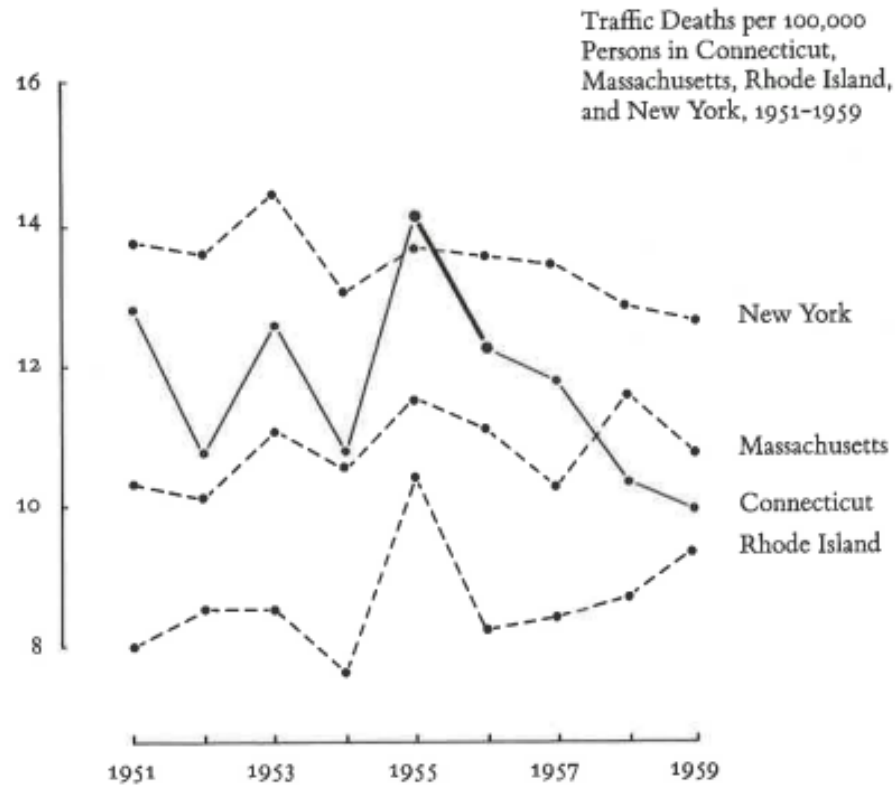


Los datos en su contexto



Sólo una representación eficaz de los datos en su contexto permite una interpretación correcta.

Los datos en su contexto



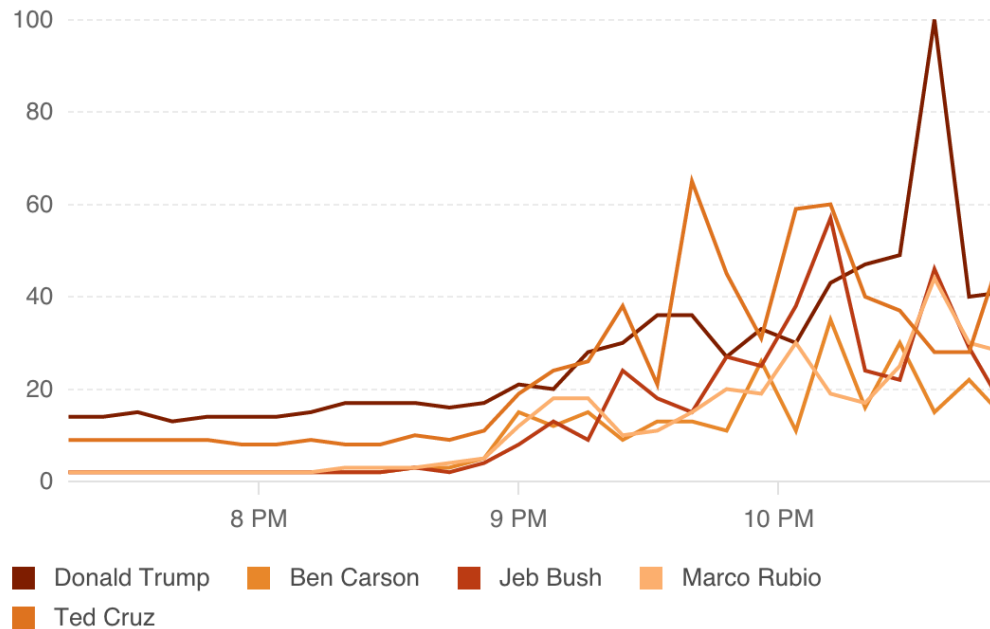
Cuánto más alcance tiene el contexto más alcance tiene la interpretación.

El color

Google Trends During the Ninth Republican Debate



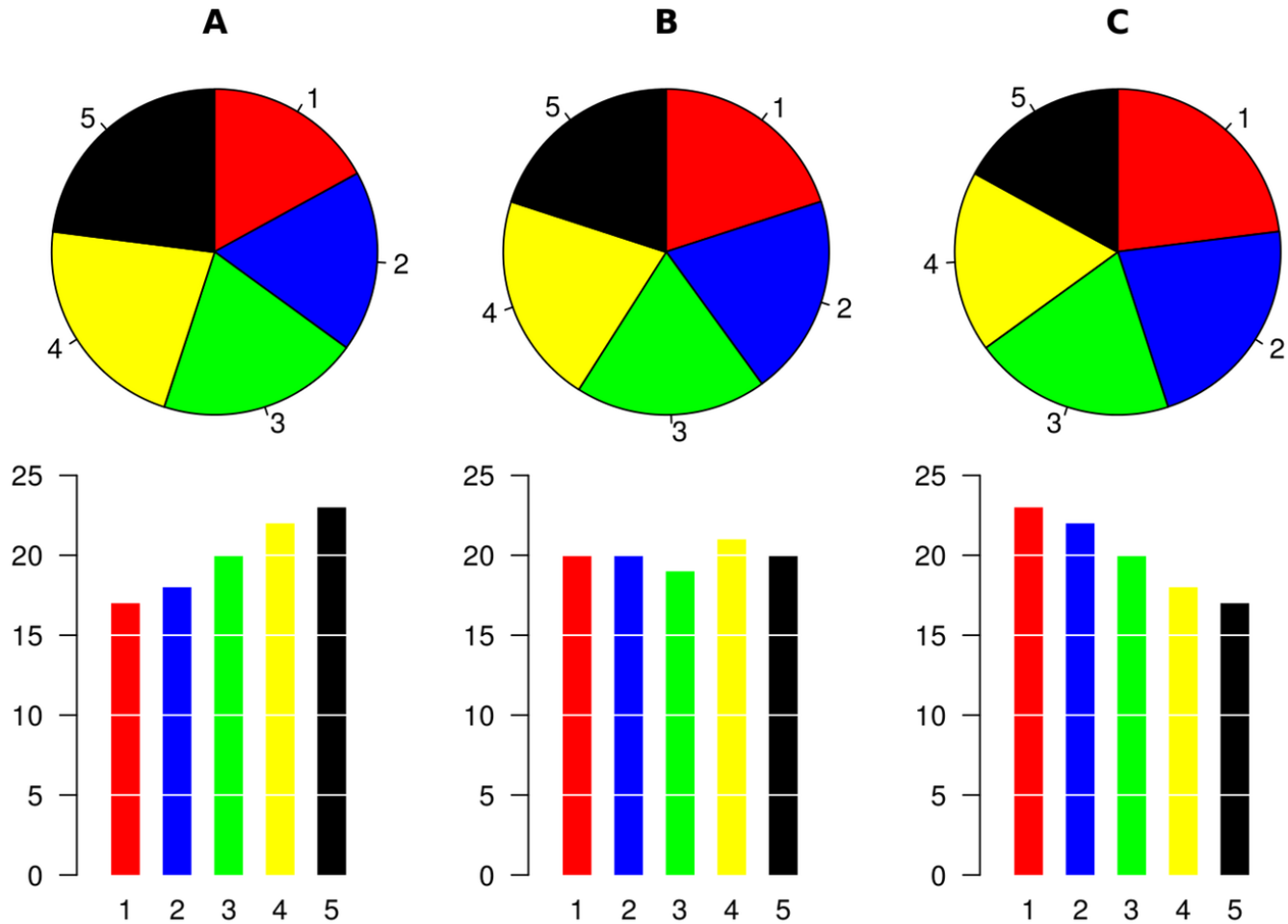
Google Trends tracks and ranks how often a presidential candidate's name is searched leading up to and during the debate, and ranks it on a scale from 1 to 100. The higher the value, the more frequently the name was searched.



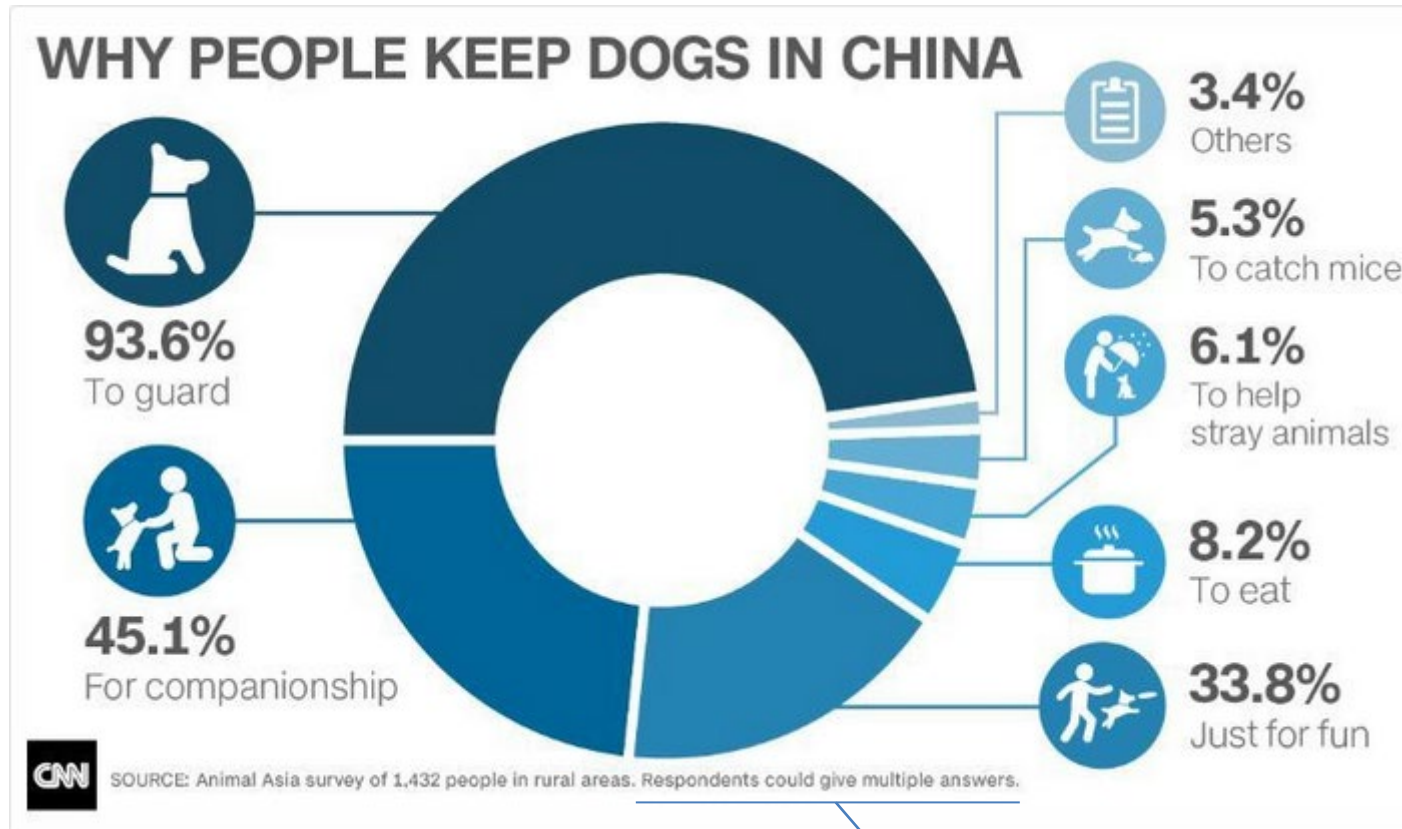
Source: [Google Trends](#). As of February 13, 2016

[See more details >](#)

Los gráficos de sectores

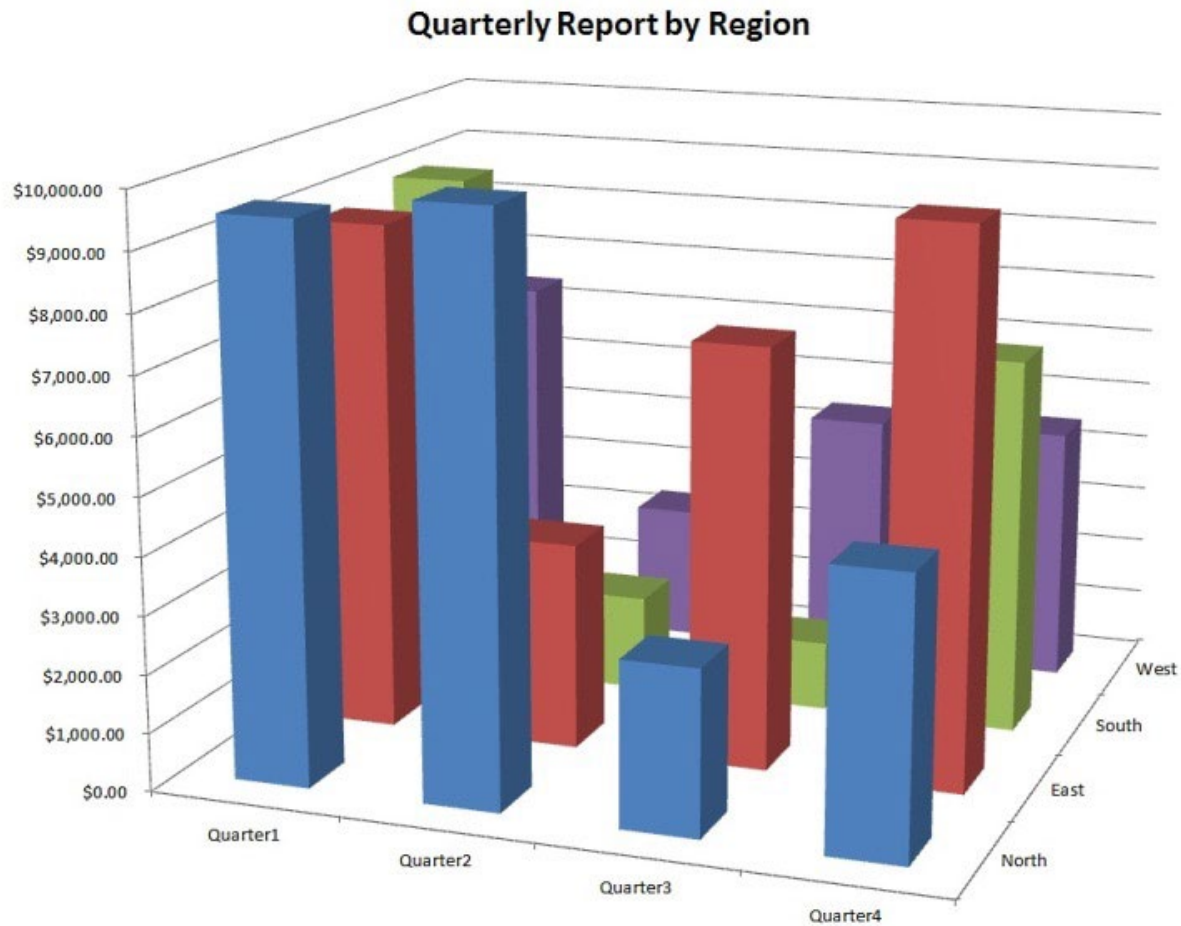


Cuando las cuentas no salen

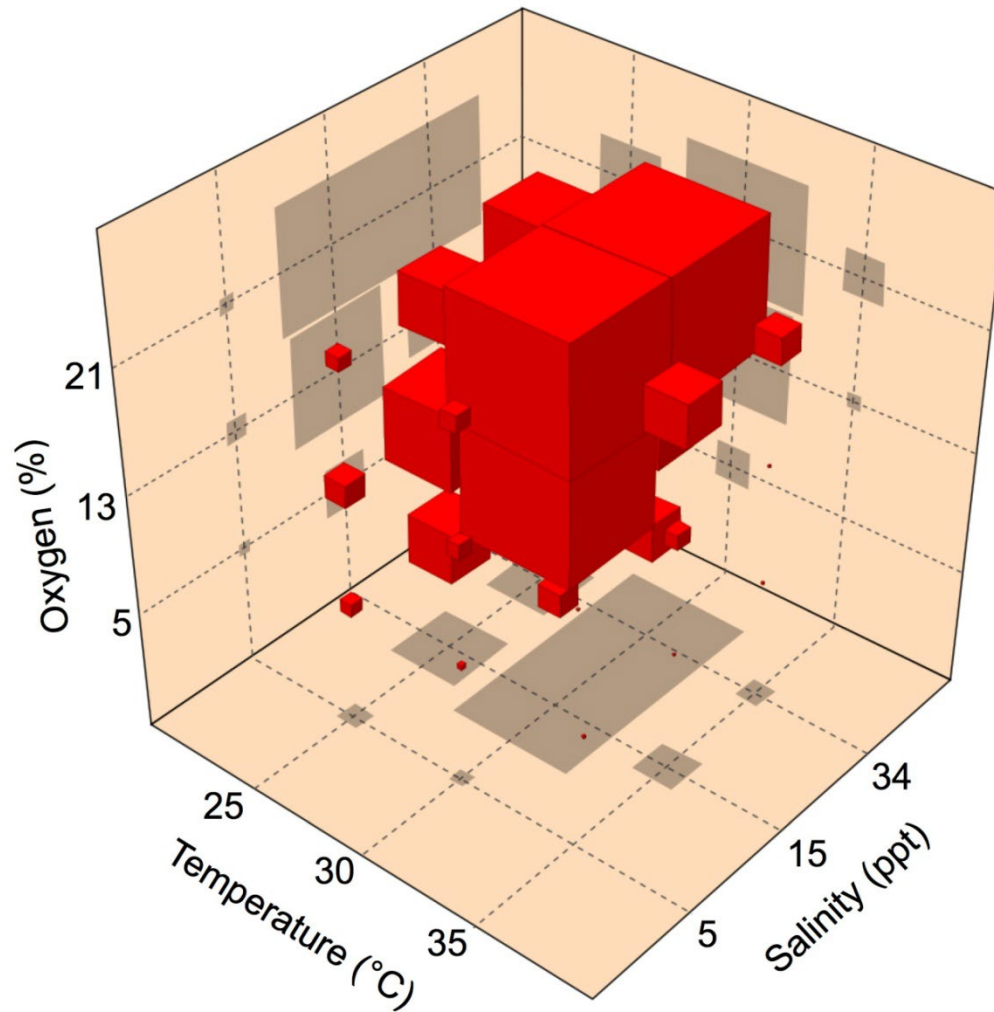


Esta no es una justificación admisible.

¿Por qué no 3D?



¿Por qué no 3D?



Confusión

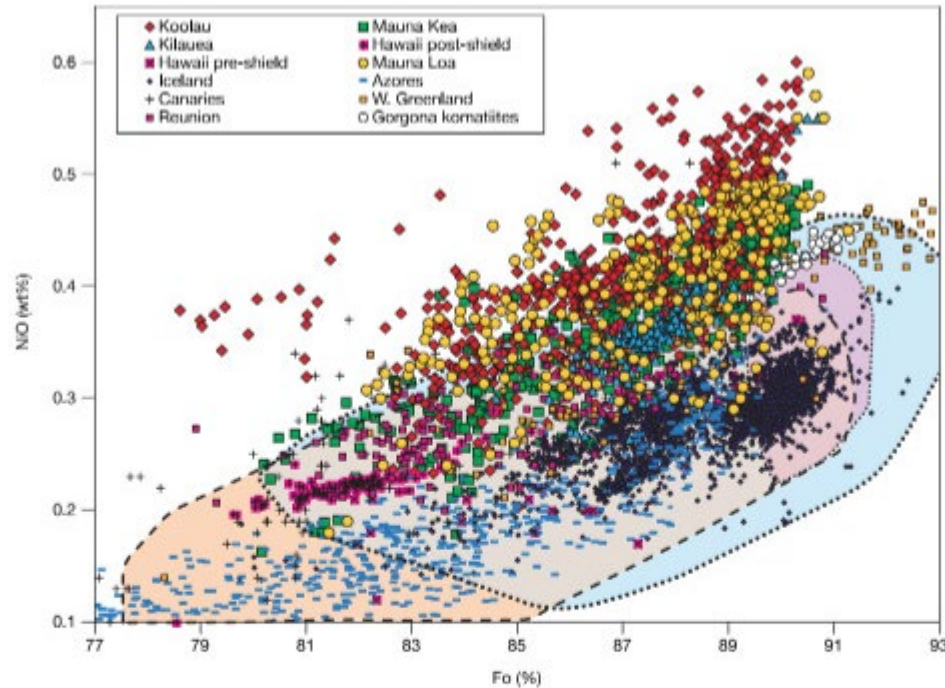
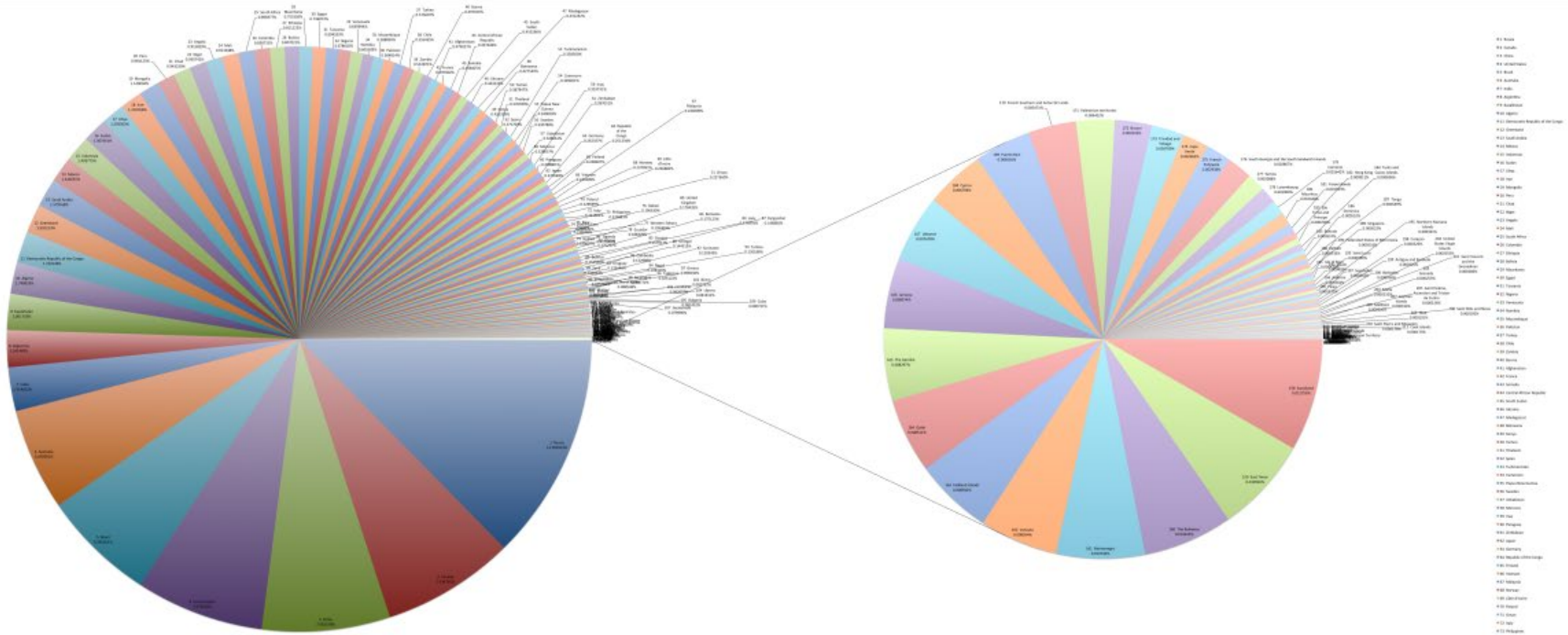


Figure 1 Compositions of olivines from mantle-derived rocks. Blue field, peridotites from mantle xenoliths, orogenic massifs and ophiolites; purple field, oceanic abyssal peridotites; beige field, phenocrysts from mid-ocean-ridge basalts; light green field, overlap between peridotite and phenocryst fields; pink field, overlap between oceanic abyssal peridotites and phenocrysts from mid-ocean-ridge basalts. Most data are from our unpublished database (data of A.V.S. on Hawaii, D. Kuzmin on Iceland, V. Kamenetsky on Gorgona, I. Nikogosian and T. Elliott on the Azores, I. Nikogosian on the Canaries and Reunion and V. Batanova for olivines from mantle peridotites). Olivines of Archaean

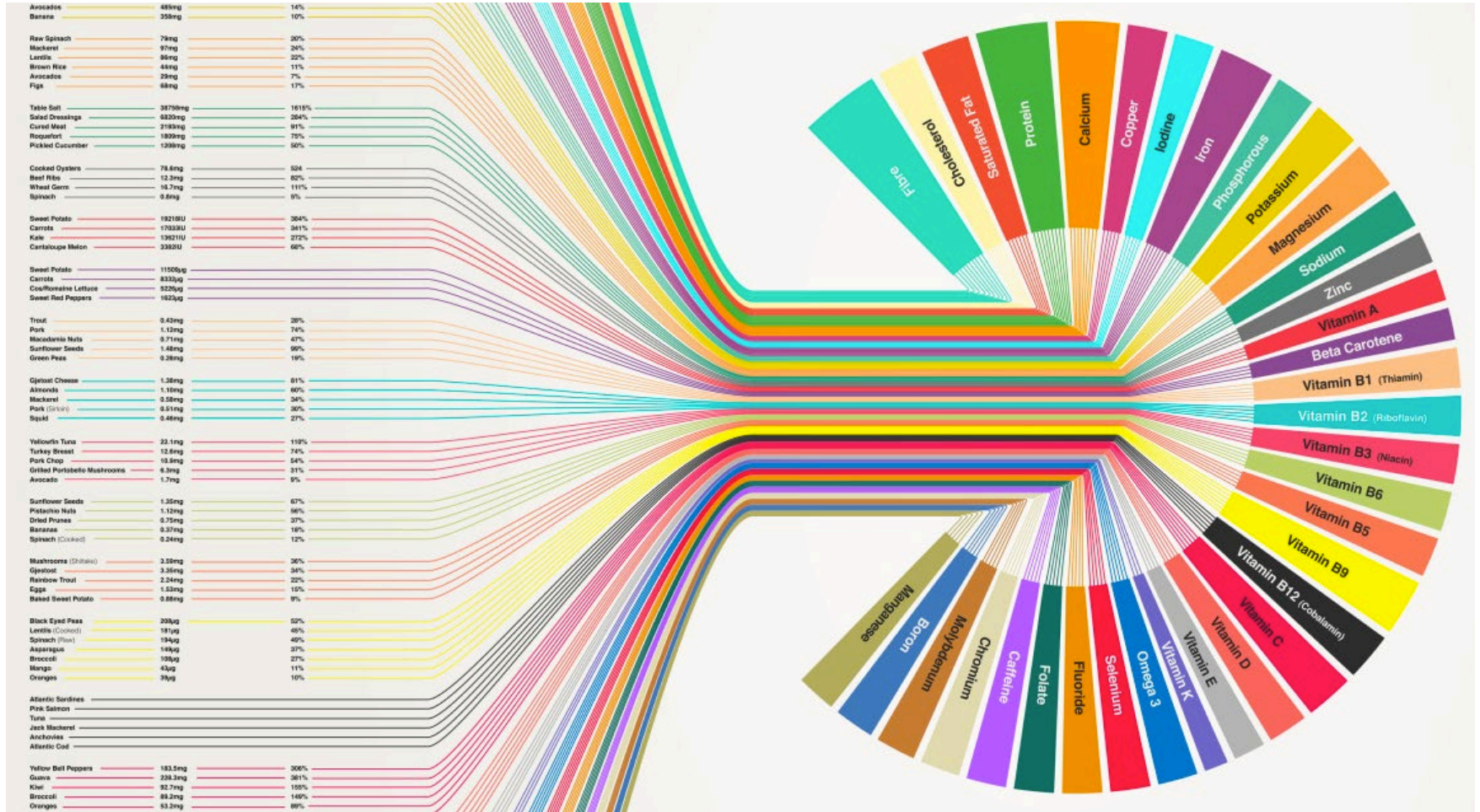
komatiites from Belingwe show NiO contents only 0.02 wt% higher than Gorgona komatiites (L. Danyushevsky, personal communication) and follow the upper boundary of the mantle peridotite field (blue). Additional data are from the GEOROC and PETDB databases⁴⁶ (see Supplementary Information for major references) and from ref. 47. Olivines from shield-stage Hawaiian basalts vary significantly in Ni content at constant Fo, with the majority systematically enriched in Ni compared with olivine from mantle peridotites, komatiites and common basalts. Olivines from post-shield and pre-shield Hawaiian basalts are similar to peridotites and common basalts.

Aparecido en la revista científica Nature:
¿Cuál es exactamente la idea que subyace?

Confusión

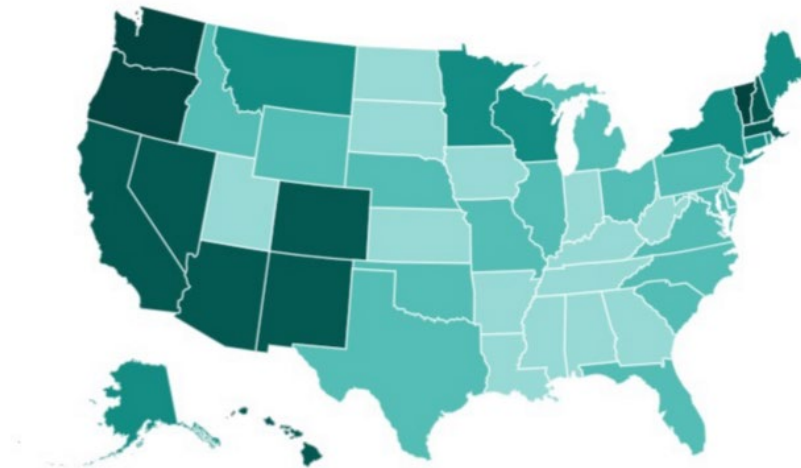


Confusión



Más desastres visuales en wtfviz.net

Estimated Transgender Individuals
per 100,000 People



Low  High

Transgenders (Minimum Estimate)

Source: U.S. Census Bureau.
As of 2010 | Refreshed July 13, 2015

Per 100,000.

Algunos principios:

1. Si se desea trasladar a los números que acompañan a una gráfica la diferencia de magnitud, su tipo debe ser proporcional a las cantidades que representan.
2. Se debe rotular de una manera clara y detallada, a fin de resolver cualquier ambigüedad que pudiera haber en la representación gráfica. Se deben rotular aquellos eventos que se consideran singulares.
3. Se ha de mostrar la variación en los datos, no la variación en el diseño.
4. En las series temporales que muestran cantidades monetarias, se deben utilizar medidas ajustadas a la inflación, en lugar de medidas nominales.
5. El número de dimensiones que muestran algún tipo de información no debe ser mayor que el número de dimensiones de los datos mostrados.
6. Los datos no se deben visualizar fuera de contexto.

Bibliografía

- ▷ A. Cairo, *El arte funcional*, Alamut, 2011.
- ▷ J. Tukey and M.B. Wilk, “Data analysis and statistics: techniques and approaches”, in E.R. Tufte, ed., *The Quantitative Analysis of Social Problems*, Reading, 1970.
- ▷ E.R. Tufte, *The visual display of quantitative information*, Graphic Press, 2001.
- ▷ D.T. Campbell and H.L. Ross, “The Connecticut crackdown on speeding: time series data in quasiexperimental analysis”, in E.R. Tufte, ed., *The Quantitative Analysis of Social Problems*, Reading, 1970.