

Seminarios de **Bioestadística**

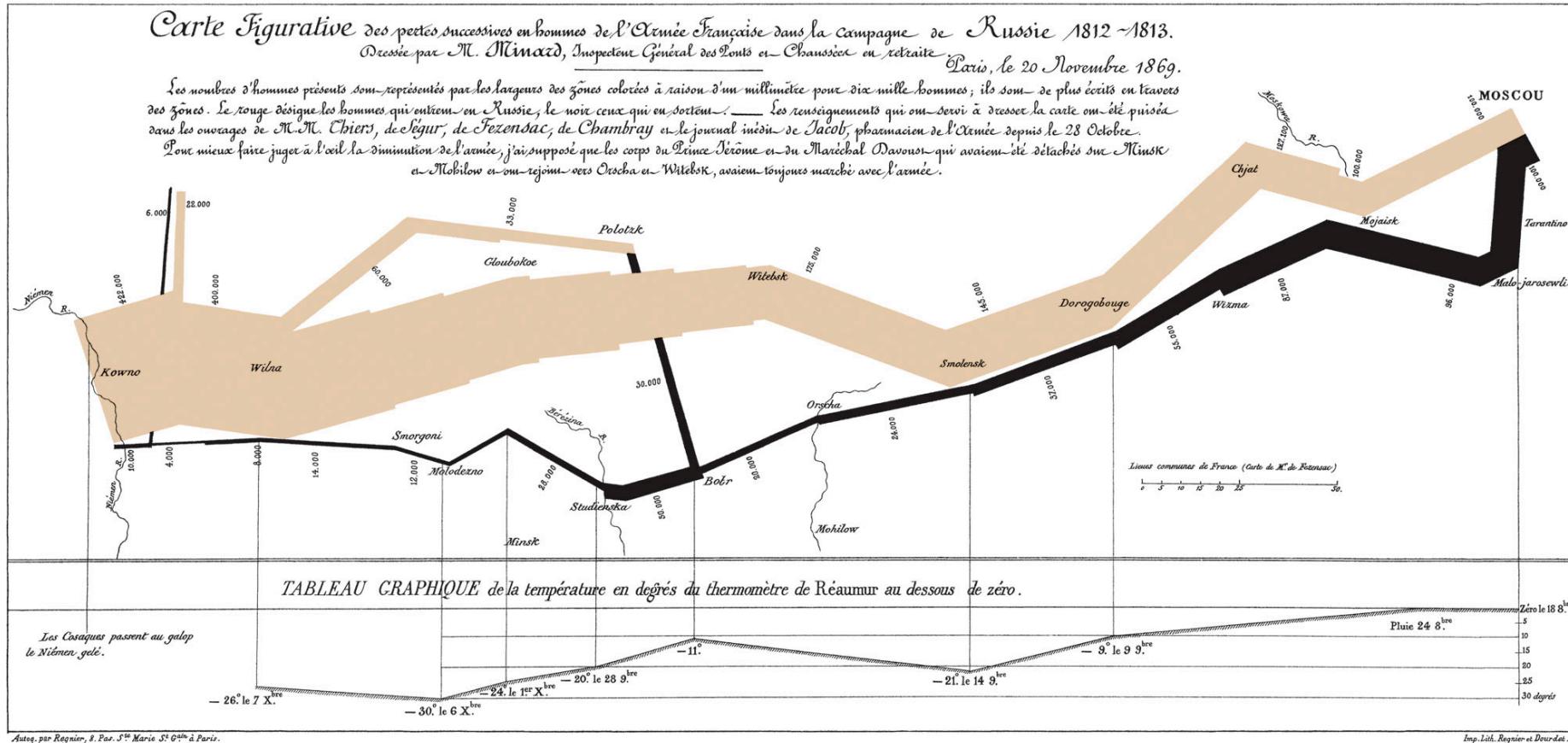
Seminario 1 | La importancia de mostrar los datos

Marta Coronado (marta.coronado@uab.cat)

Grado en Genética | Curso 2025/26



"El mejor gráfico estadístico jamás dibujado" ~Edward Tufte



El gráfico es notable por su representación en dos dimensiones de seis tipos de datos: el número de tropas de Napoleón; la distancia; la temperatura; la latitud y la longitud; la dirección del viaje; y la ubicación en relación con fechas específicas. [Wikipedia](#).

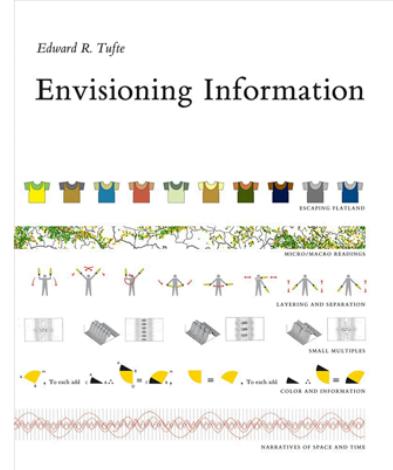
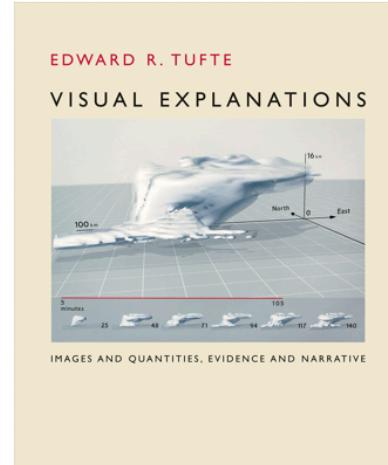
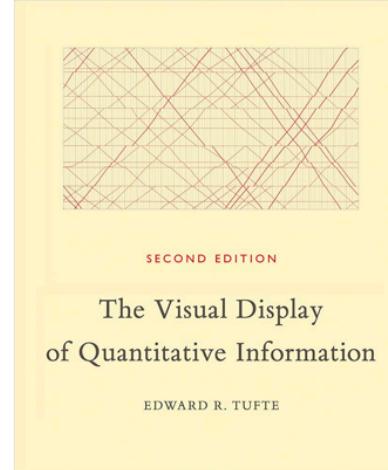
01

Una introducción a la visualización de datos

Integridad de los gráficos – Edward Tufte (1942-)

Seis principios de Tufte:

1. **Comparaciones**: Muestra comparaciones para ilustrar contrastes y diferencias.
2. **Causalidad**: Demuestra cómo una o más variables independientes afectan o influyen en las variables dependientes.
3. **Multivariable**: Combina diferentes tipos de datos.
4. **Integración**: Incorpora diversos modos de información.
5. **Documentación**: Incluye atribuciones, títulos detallados y mediciones (escalas).
6. **Contexto**: Describe el estado antes y después.



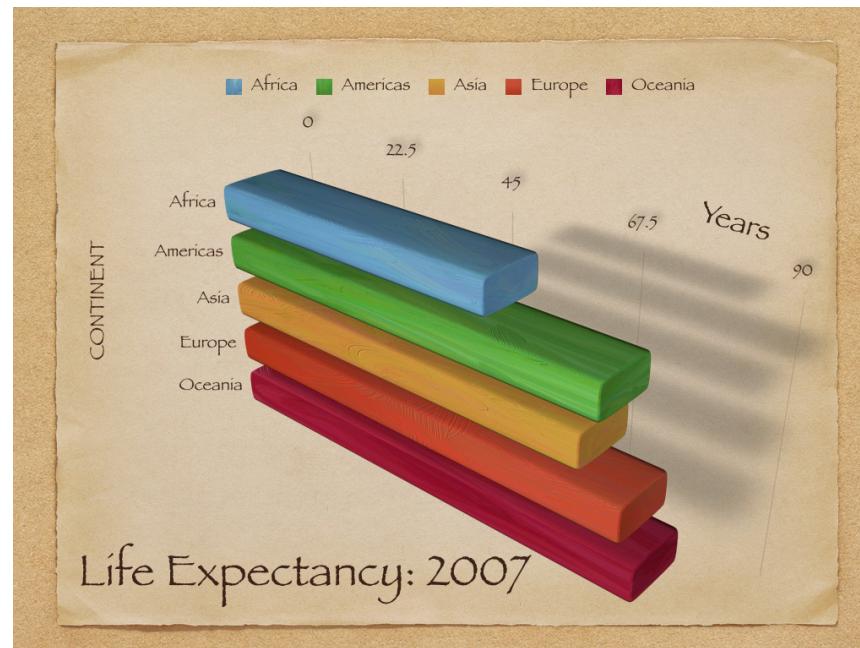
“Above all, show the data”

Chartjunk

Uso excesivo e innecesario de efectos gráficos en los gráficos

Excellence

Comunicación de ideas complejas con claridad, precisión y eficiencia



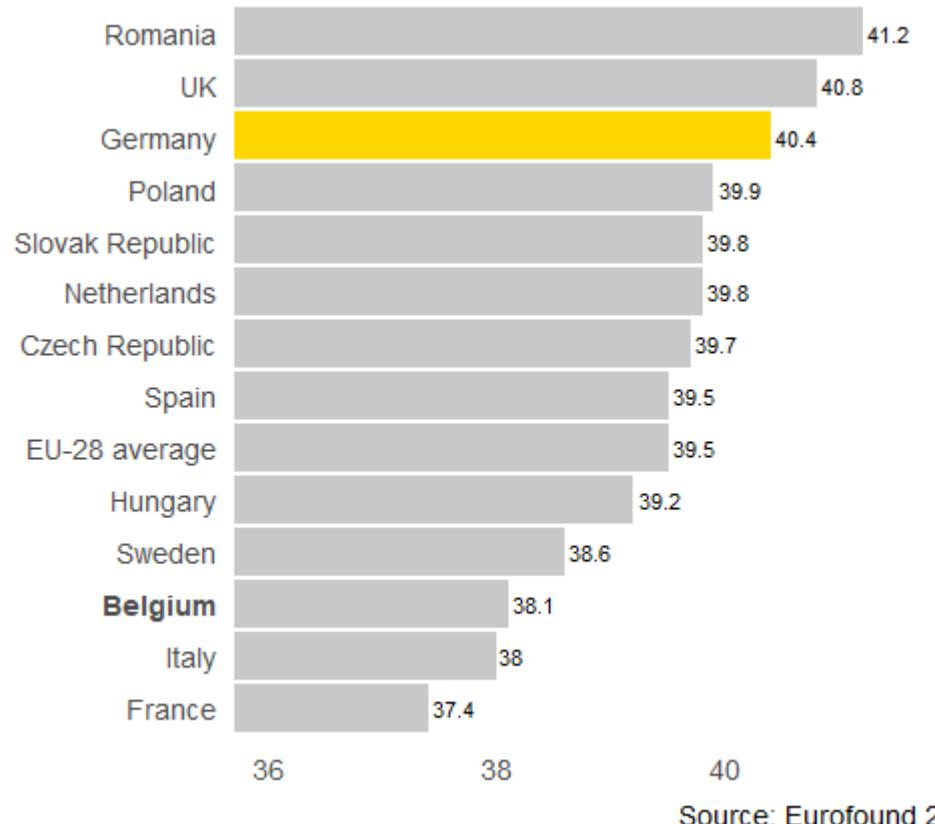
💡 ¿Qué problemas identificas en esta figura?

02

“Above all, show the data”

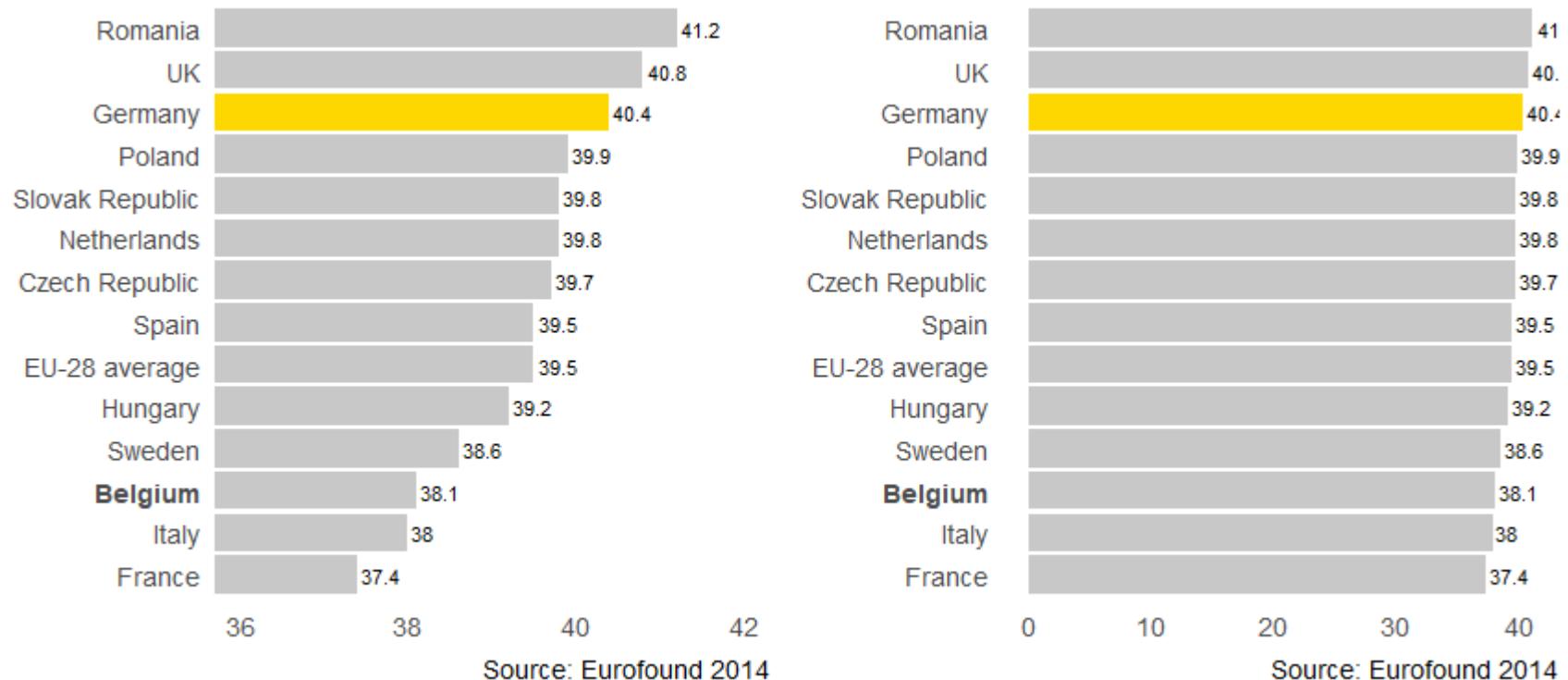
Límites justos de los ejes

Promedio de horas de trabajo/semana (Eurofund 2014). Gráfico creado por la agencia de desarrollo económico alemana GTAI. [Fuente](#).



Límites justos de los ejes

Promedio de horas de trabajo/semana (Eurofund 2014). Gráfico creado por la agencia de desarrollo económico alemana GTAI. [Fuente](#).

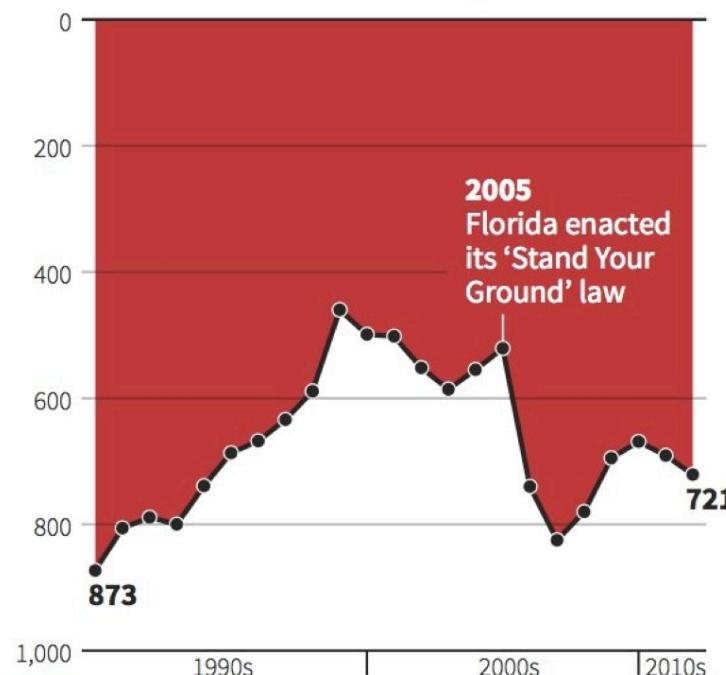


- ❶ Las **barras** en un **gráfico de barras** deben (casi) siempre **extenderse al cero** (magnitud absoluta de los valores). No es necesario en gráficos lineares (cambio en la variable dependiente a medida que cambia la variable independiente).

Límites justos de los ejes

Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

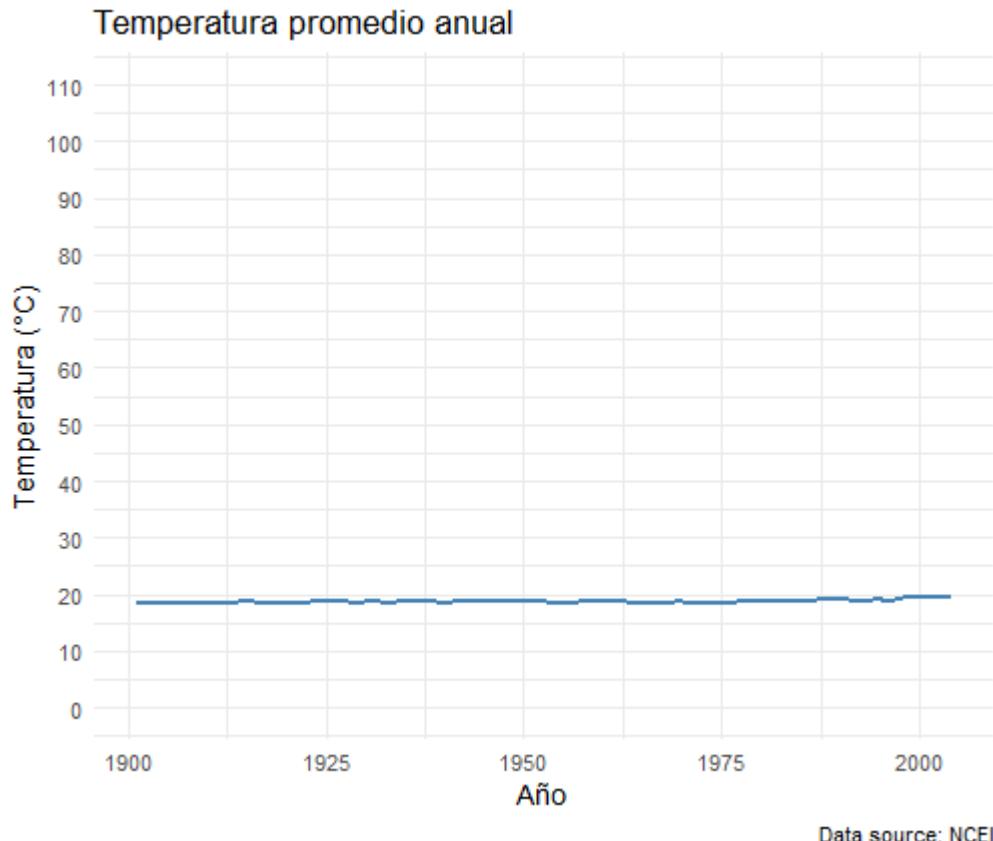
REUTERS

- Impresión visual inmediata de que las muertes por armas de fuego disminuyeron drásticamente tras la aprobación de la ley de defensa propia ("Stand Your Ground") en Florida
- El descenso es una ilusión

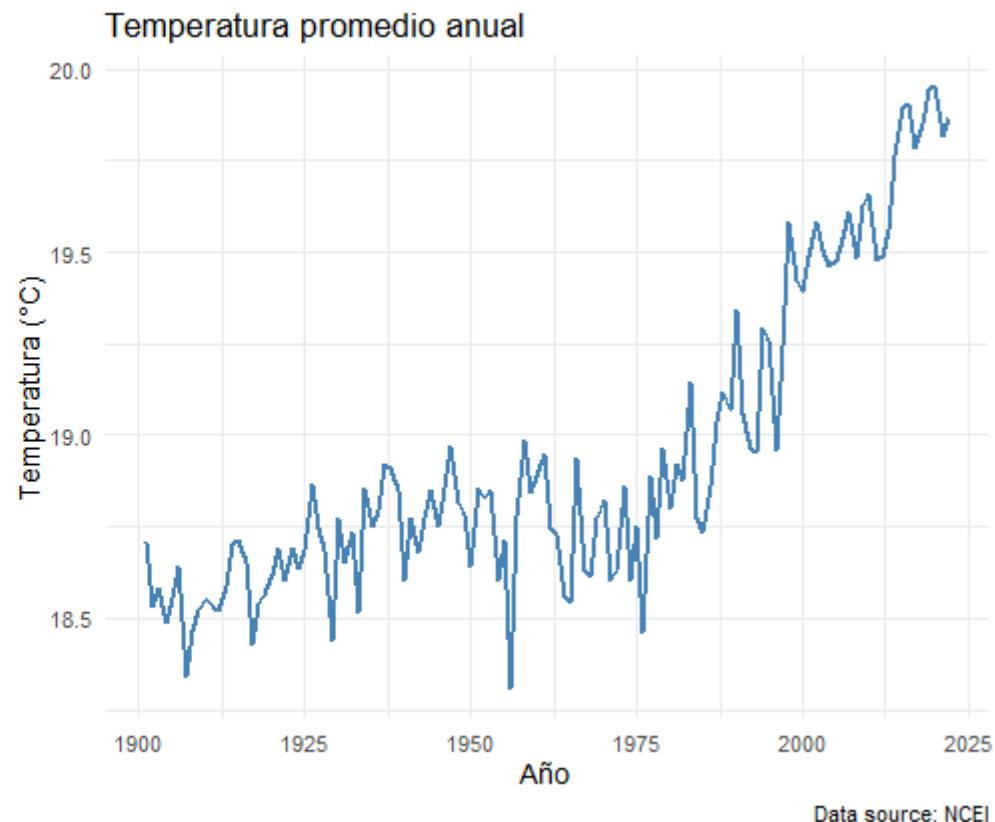
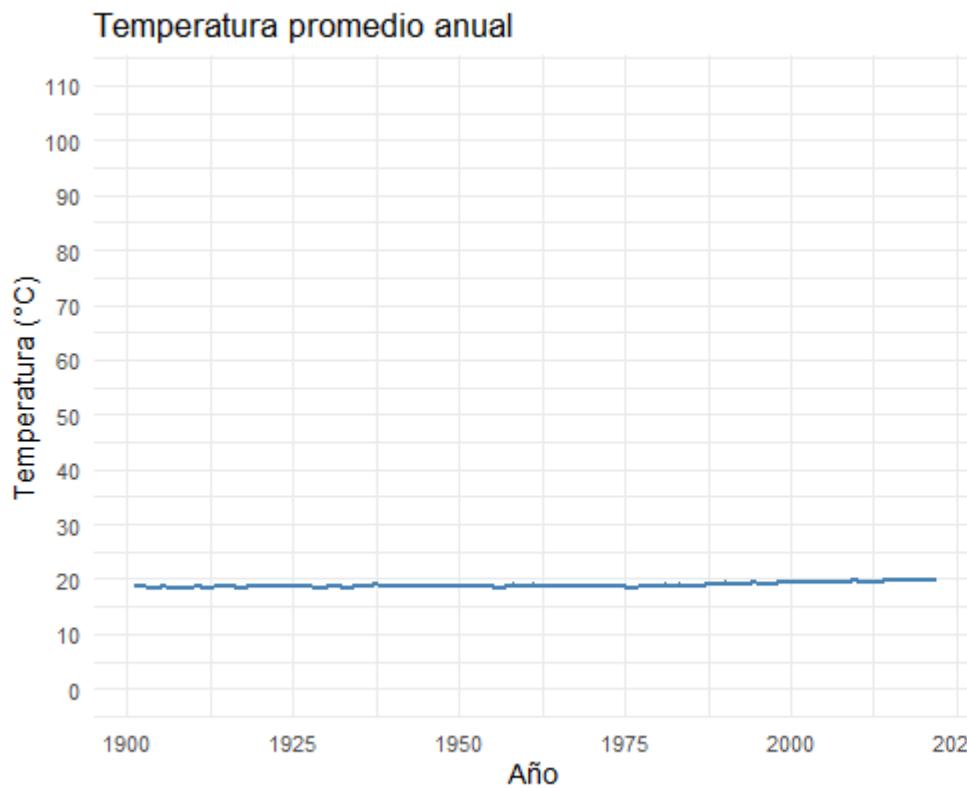
Límites justos de los ejes



Límites justos de los ejes



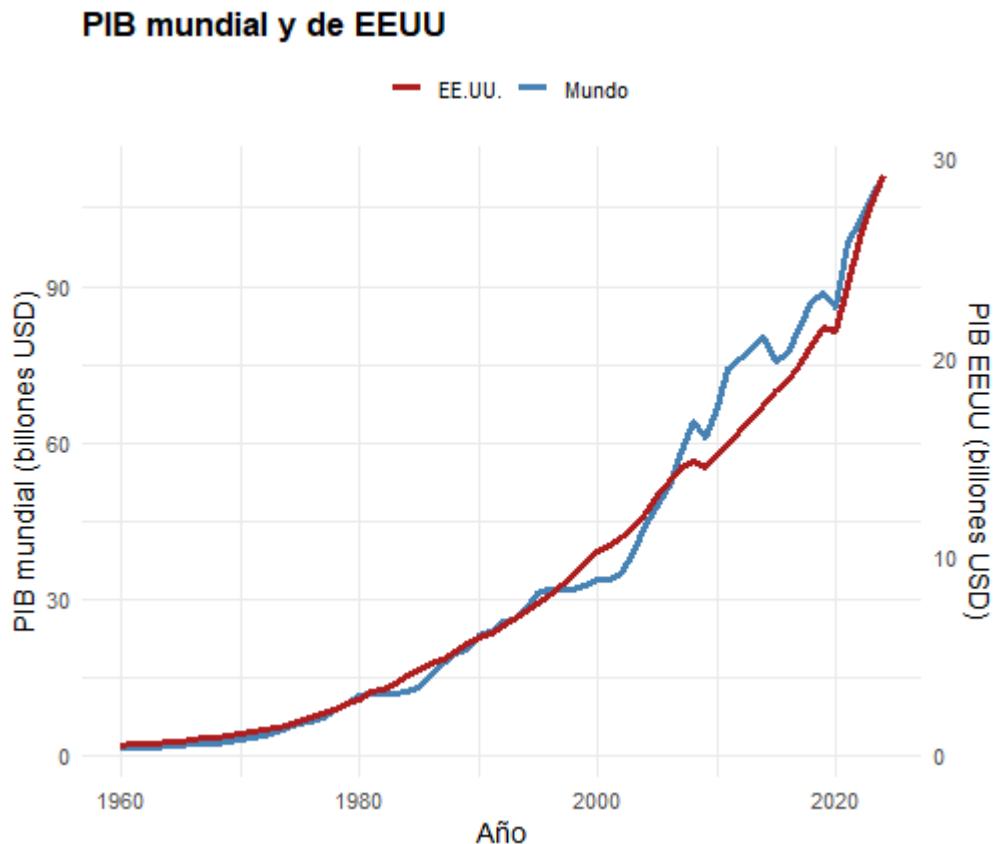
Límites justos de los ejes



Cuando hablamos de temperatura global, un cambio de 1-2°C es un cambio muy significativo.

Armstrong McKay, David I. (2022). "Exceeding 1.5°C global warming could trigger multiple climate tipping points". *Science*. 377 (6611): eabn7950.

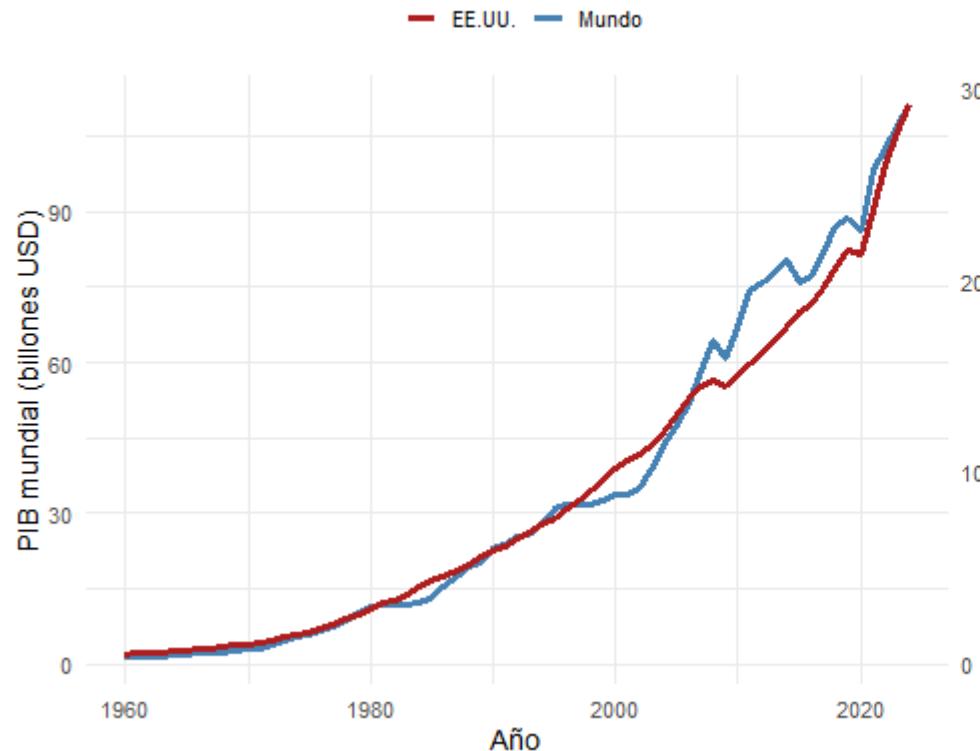
Uso de los dobles ejes



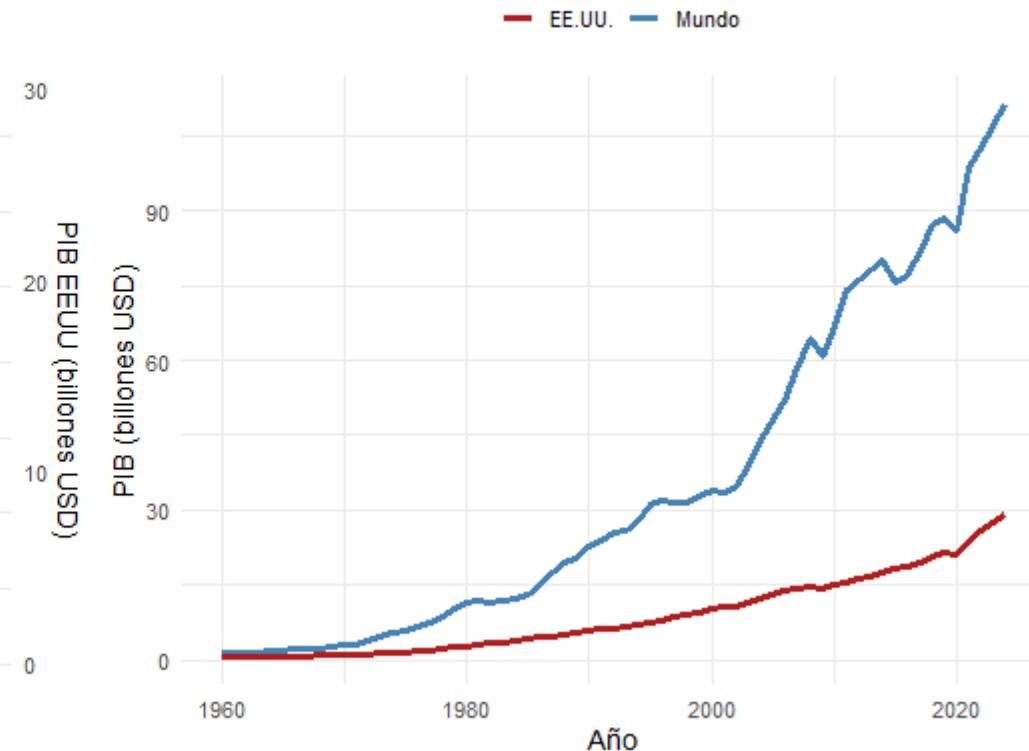
¿El PIB de EEUU es igual al mundial?

Uso de los dobles ejes

PIB mundial y de EEUU

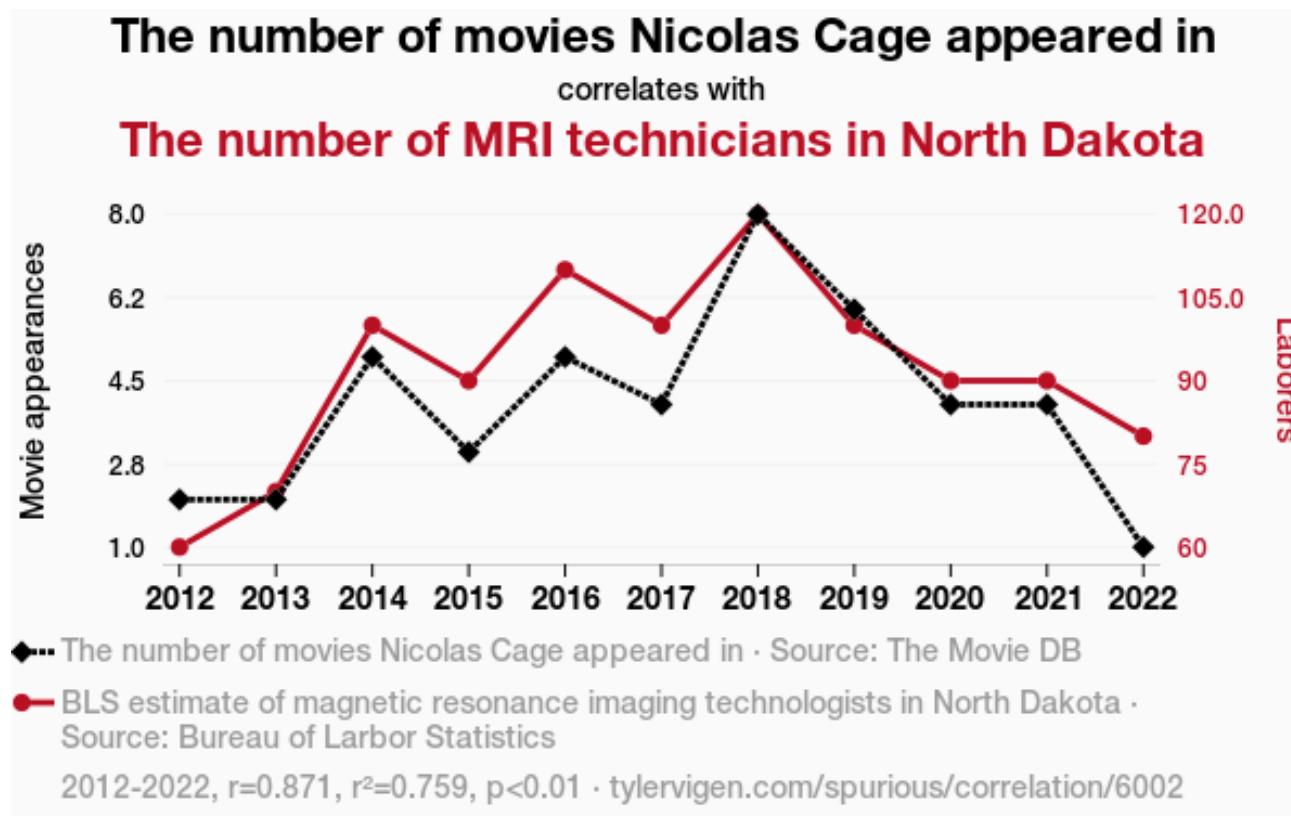


PIB mundial y de EEUU



Los datos con las mismas unidades de medida se muestran mejor usando la **misma escala** en un único eje.

Evitar el uso de dobles-ejes



Correlaciones espúreas ([link](#)).

Cuando se trata de ejes dobles con dos unidades de medida diferentes, en la mayoría de los casos es preferible usar dos gráficos separados en lugar de un gráfico con doble eje.

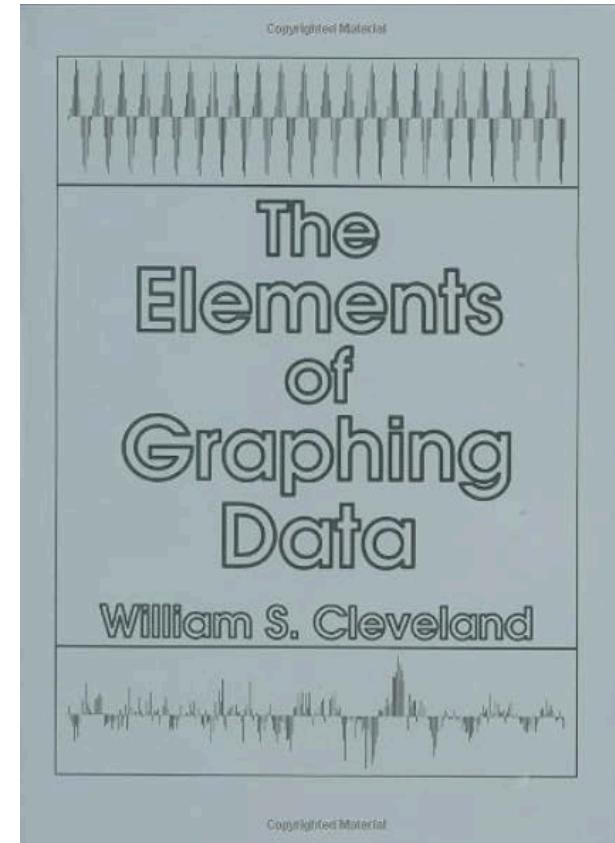
03

How humans see data

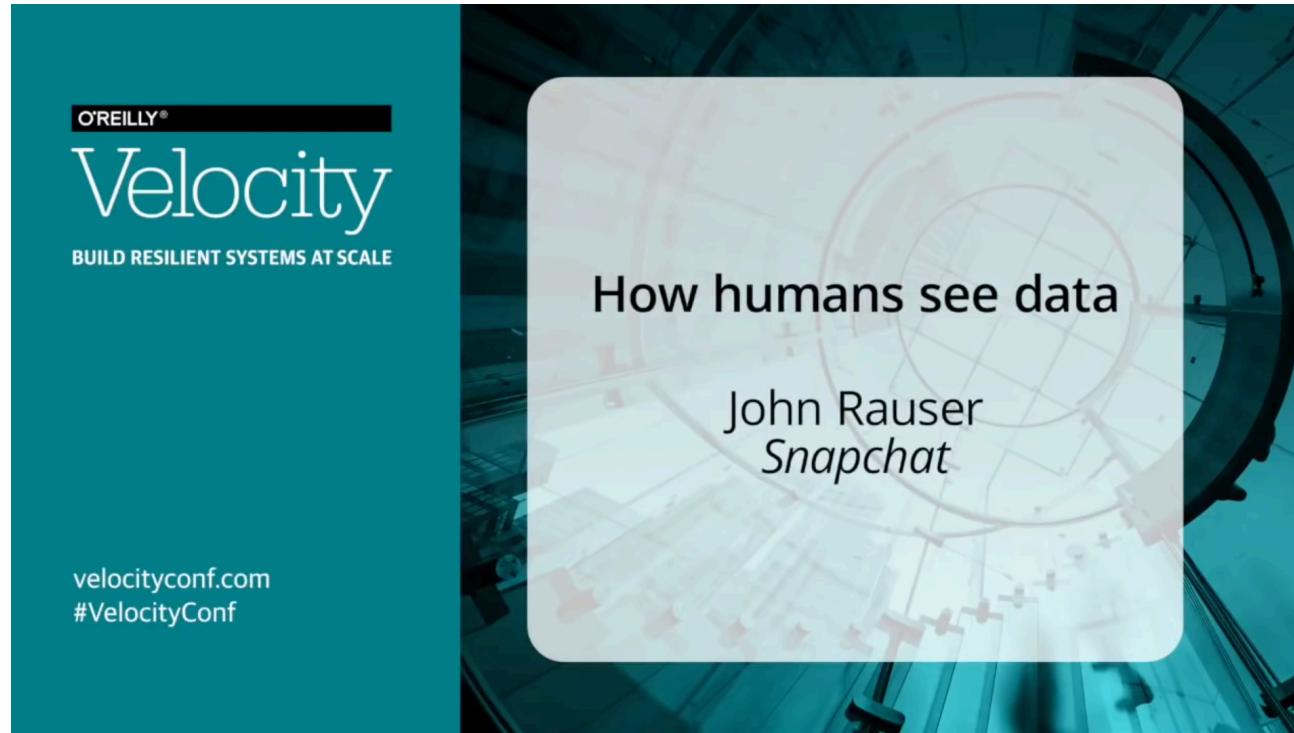
Las tres operaciones visuales de percepción de patrones de Cleveland

1. **Detección**: el reconocimiento visual de que un objeto geométrico codifica un valor físico.
2. **Agrupación**: la organización de los elementos gráficos detectados.
3. **Estimación**: la evaluación visual de la magnitud relativa de dos o más valores físicos cuantitativos.

Crédito: John Rauser



Las tres operaciones visuales de percepción de patrones de Cleveland



<https://www.youtube.com/watch?v=fSgEeI2Xpdc>

Crédito: John Rauser

Niveles de estimación

Crédito: John Rauser

Tres niveles diferentes de estimación:

- | | |
|---|----------|
| 1. Discriminación : dos valores son diferentes | X=Y X!=Y |
| 2. Ranking : un valor es mayor | X>Y X<Y |
| 3. Comparación de razones : un valor es dos veces más grande | X/Y = ? |

Todas estas implican comparación:

La comparación eficiente entre diferentes puntos de datos es casi siempre el objetivo de una visualización.

Ranking de Cleveland

Crédito: John Rauser

■ Cleveland and McGill (1985) Graphical Perception and Graphical Methods for Analyzing Scientific Data. Science 29(4716):828-833

Rango	Aspecto evaluado
1	Posición en una escala común
2	Posición en escalas idénticas pero desalineadas
3	Longitud
4	Ángulo o pendiente
5	Área
6	Volumen, densidad o saturación del color
7	Tonalidad del color

Qué tan precisos somos los humanos al estimar cantidades que se codifican de diferentes maneras.

Siete formas diferentes de codificar un valor cuantitativo, clasificadas de la más efectiva a la menos efectiva.

Ejemplo con el color

Crédito: John Rauser

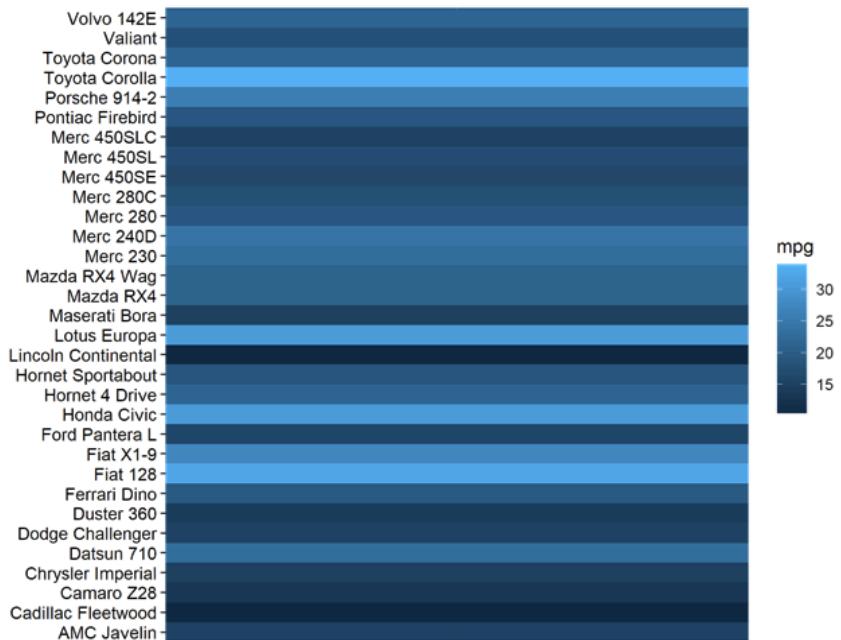
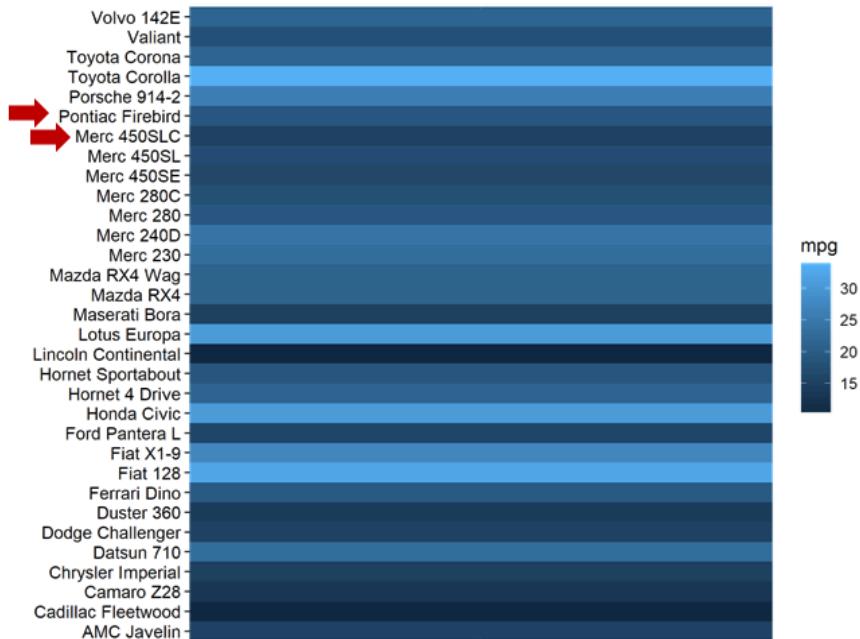


Gráfico que codifica información utilizando el color

- Conjunto de datos de 32 automóviles que fueron evaluados en una edición de 1974 de la revista Motor Trend. Eficiencia de combustible en millas por galón (mpg).

Ejemplo con el color

Crédito: John Rauser



La primera tarea de estimación de Cleveland es **discriminación**:

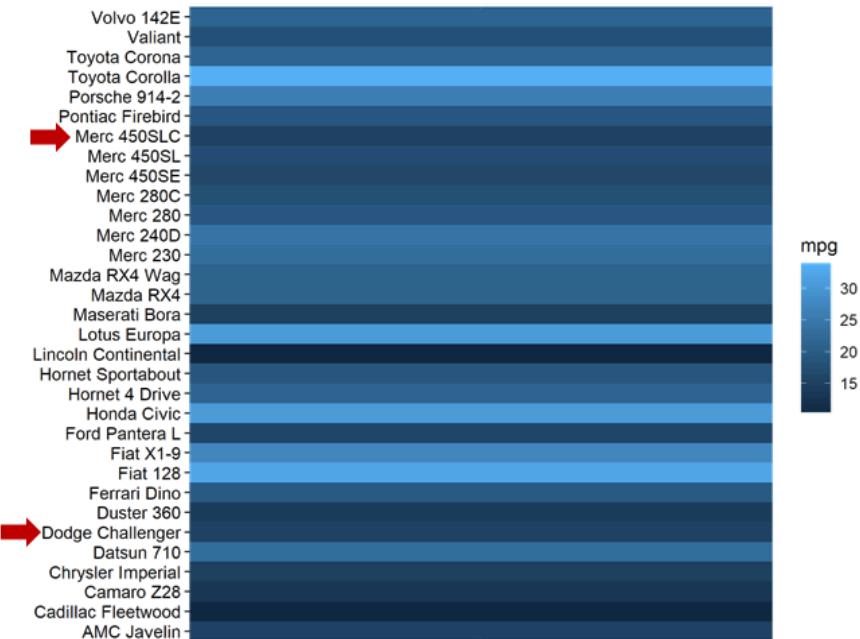
💡 ¿Qué piensas sobre estos dos valores: Pontiac Firebird y Merc450SLC? ¿Son iguales o diferentes?

Gráfico que codifica información utilizando el color

- Conjunto de datos de 32 automóviles que fueron evaluados en una edición de 1974 de la revista Motor Trend. Eficiencia de combustible en millas por galón (mpg).

Ejemplo con el color

Crédito: John Rauser



La primera tarea de estimación de Cleveland es **discriminación**:

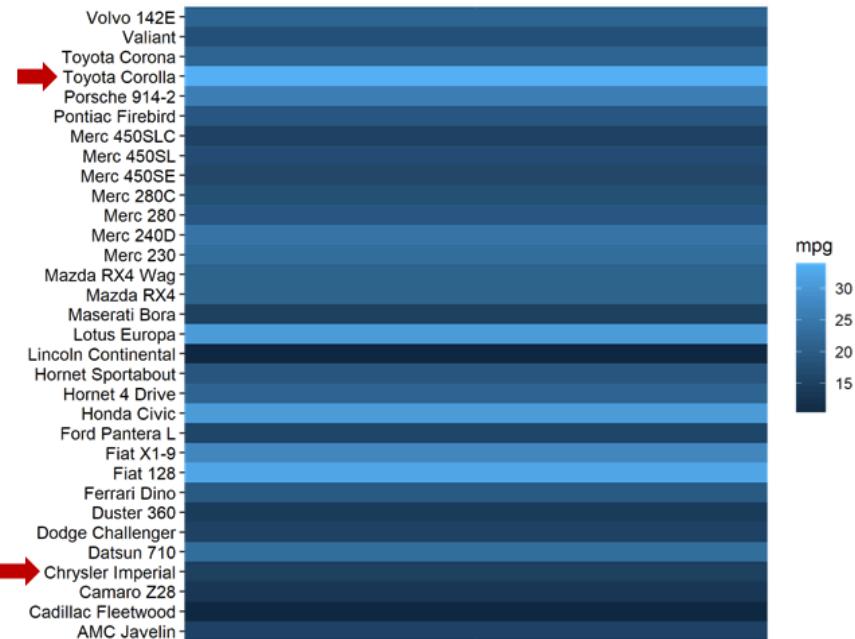
? ¿Qué hay de Merc450SLC frente a Dodge Challenger? ¿Son iguales o diferentes?

Gráfico que codifica información utilizando el color

- Conjunto de datos de 32 automóviles que fueron evaluados en una edición de 1974 de la revista Motor Trend. Eficiencia de combustible en millas por galón (mpg).

Ejemplo con el color

Crédito: John Rauser



La segunda tarea de estimación de Cleveland es **ranking**:

?

¿Qué hay de Toyota Corolla frente a Chrysler Imperial? ¿Cuál tiene mejor eficiencia de combustible?

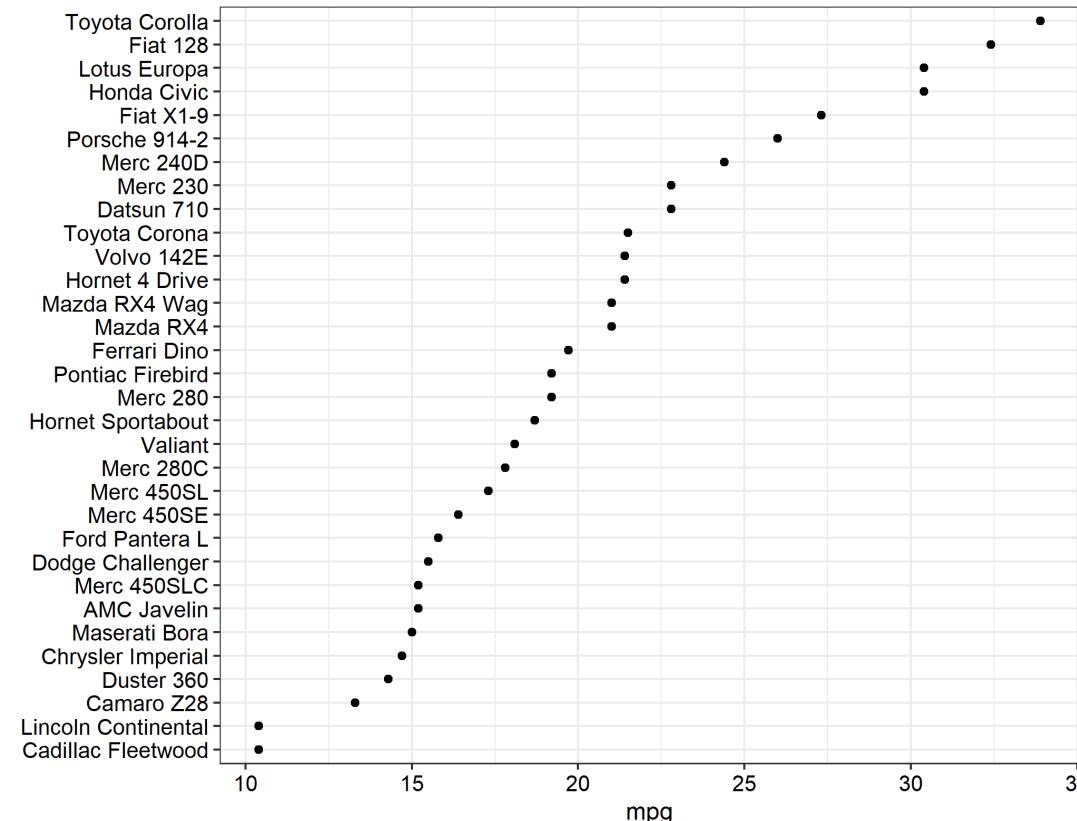
Gráfico que codifica información utilizando el color

- Conjunto de datos de 32 automóviles que fueron evaluados en una edición de 1974 de la revista Motor Trend. Eficiencia de combustible en millas por galón (mpg).

Ejemplo con posición (escala común)

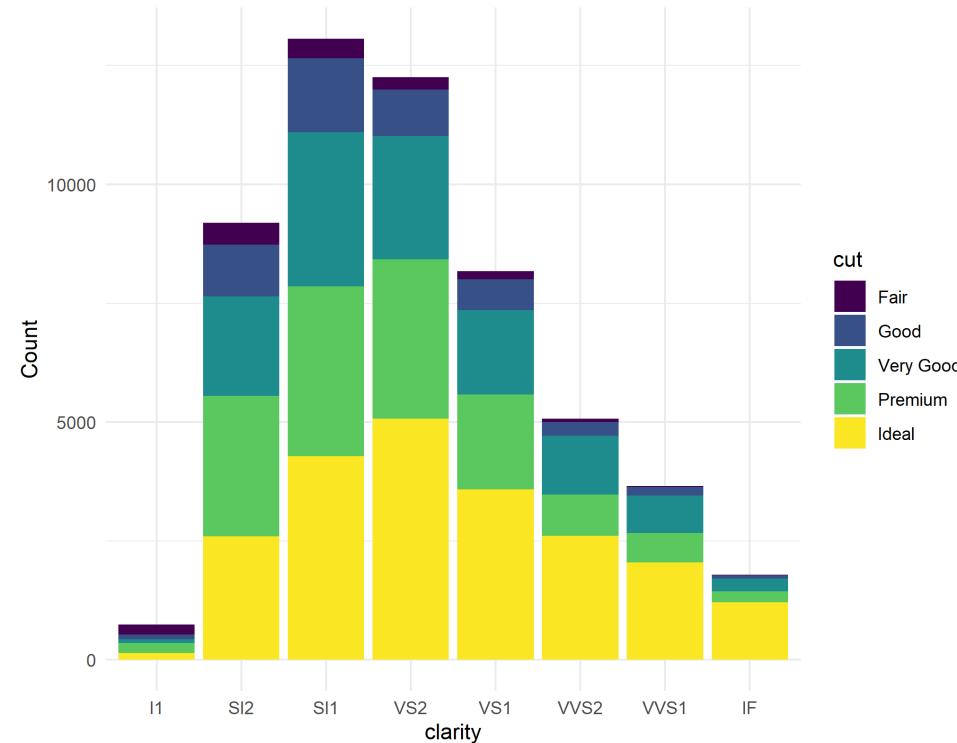
Crédito: John Rauser

Así es como los datos deberían haber sido representados: ¡la discriminación, la comparación de razones y el ranking son triviales!



¿Por qué es malo apilar datos?

Crédito: John Rauser

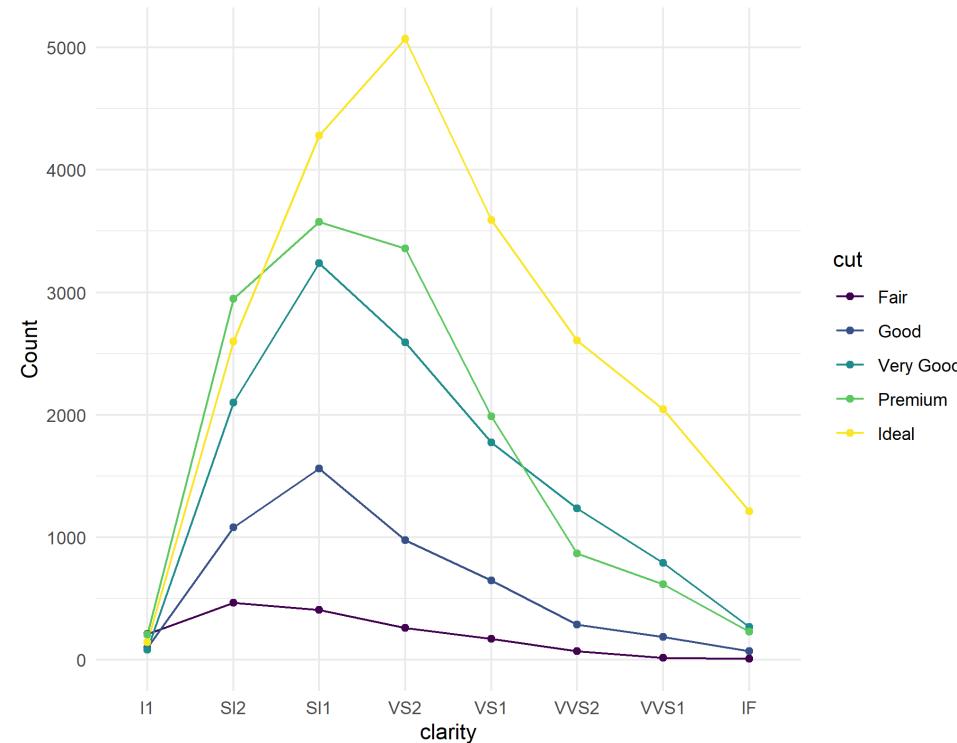


Conjunto de datos de 54,000 diamantes. **Gráfico de barras apiladas:** Conteo de diamantes en cada combinación de corte y claridad.

- ¿Hay más diamantes de corte premium con claridad SI1 o con claridad SI2?

¿Por qué es malo apilar datos?

Crédito: John Rauser



Conjunto de datos de 54,000 diamantes. Gráfico de coordenadas paralelas.

- Si quieras comunicar el conteo en cada combinación de corte y claridad, codifica esa información utilizando la posición en una escala común.

¿Tienen algún problema estas noticias?



Madrid Opina, emitido el 5 de octubre de 2010 en Tele Madrid



3 de septiembre del 2013 en TVE



La Razón, 21 de abril de 2010



Telediario 1, 1 de octubre de 2010

Si encuentras más noticias como está, añádela al foro del CV.