



Introducción al análisis de datos y presentación de resultados



¿Qué es analizar datos?

Significa categorizar, ordenar, manipular y resumir datos.

El propósito del análisis es reducir los datos a una forma entendible e interpretable para que las relaciones de los problemas de investigación puedan ser estudiadas y probadas.

Tipos de Variables

Variables Numéricas ("Cuantitativas")

Se pueden expresar en términos de valores numéricos.

Se pueden medir y definir intervalos.

Variables Categóricas ("Cualitativas")

Se centran en las características o cualidades de los objetos.

En algunos casos, se pueden ordenar en una secuencia, pero no se pueden medir (no se les puede asignar un valor numérico)

El papel de las Gráficas

Las gráficas de datos muestran la relación entre una **variable dependiente (y)** y una **variable independiente (x)**.

Describen los cambios en Y como una función de los cambios en X.

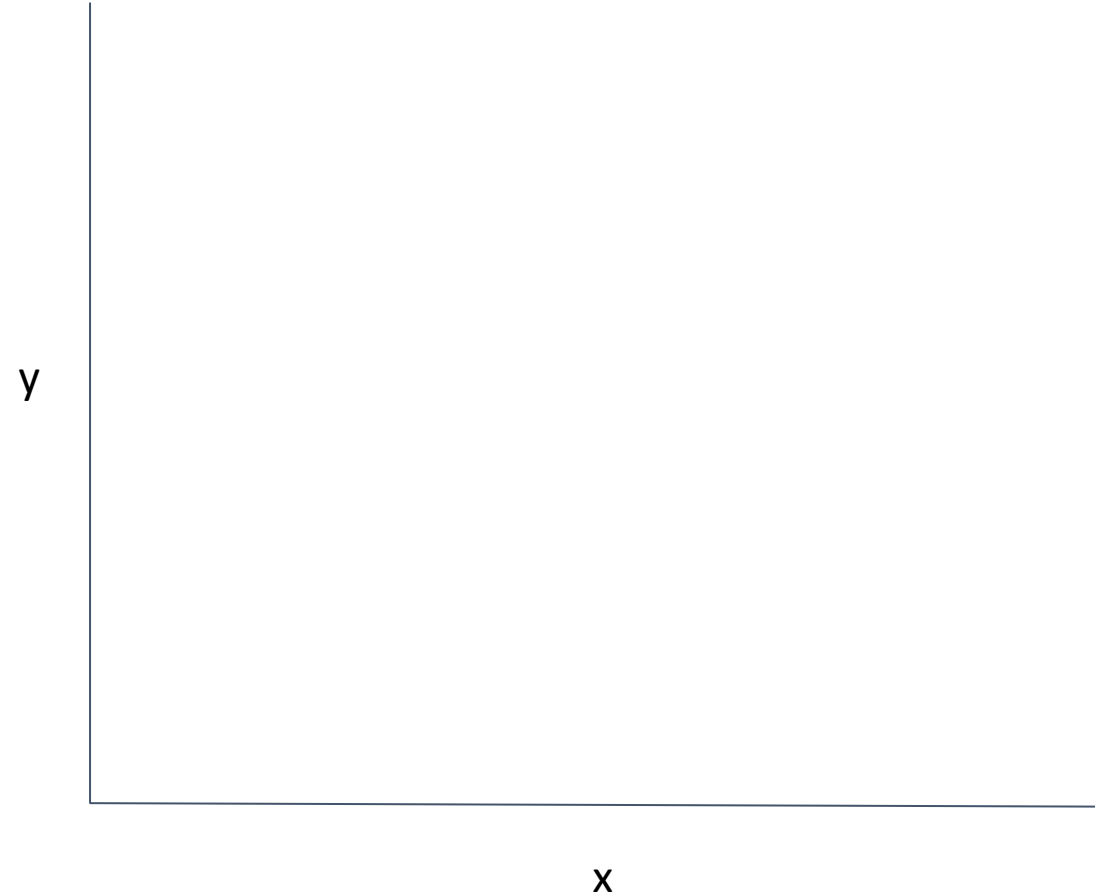


Chart 1: Total Worldwide Cases of Coronavirus

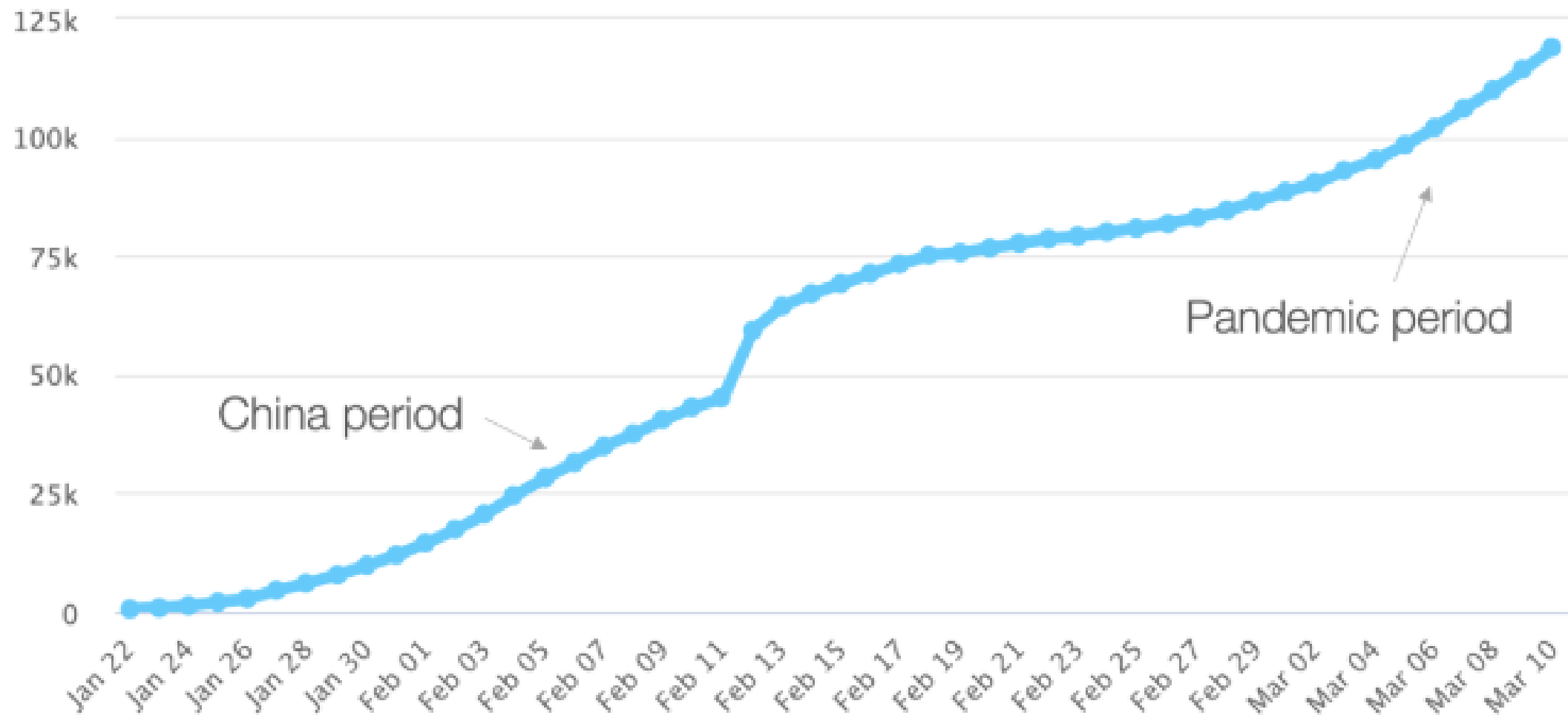


Chart 1: Total Worldwide Cases of Coronavirus

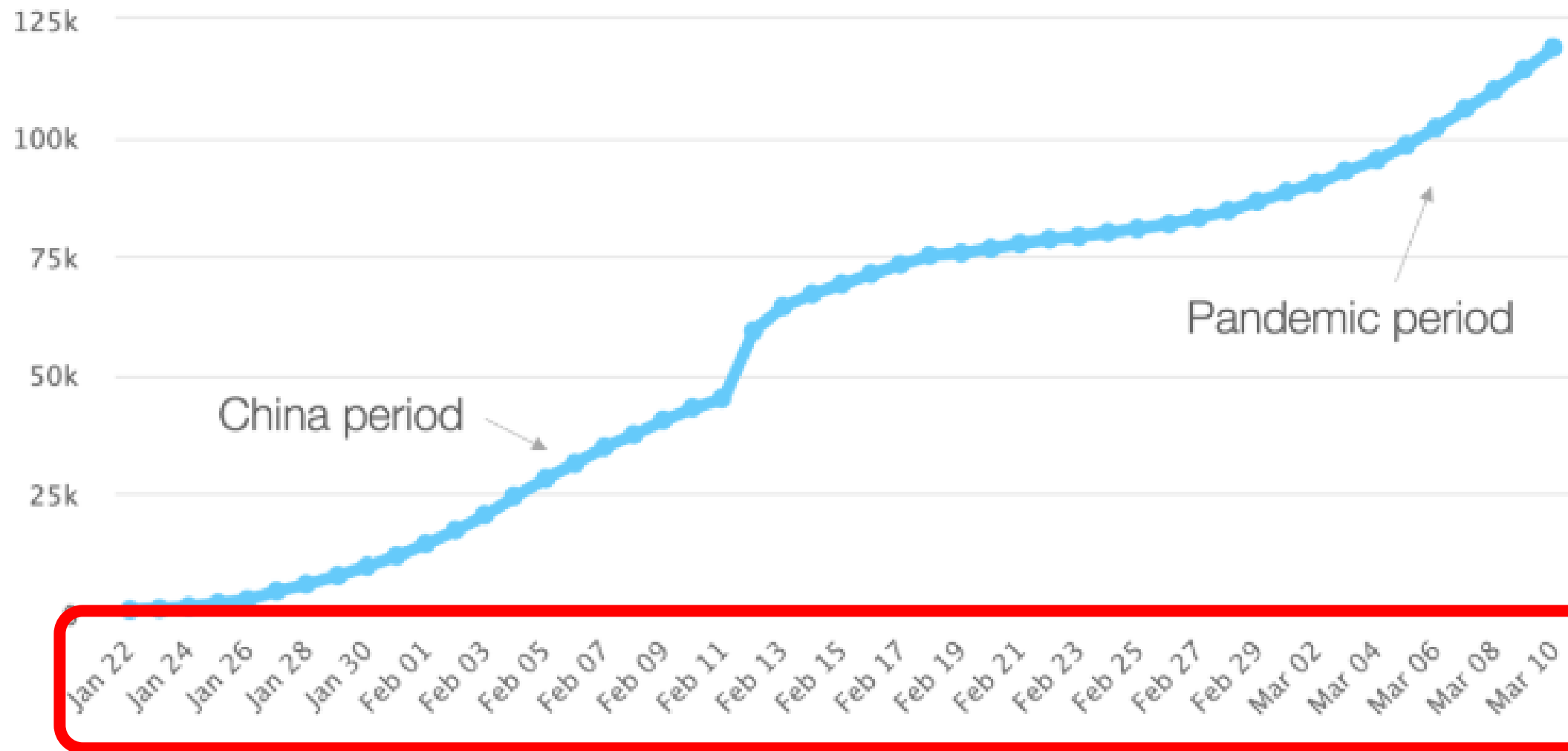


Chart 1: Total Worldwide Cases of Coronavirus

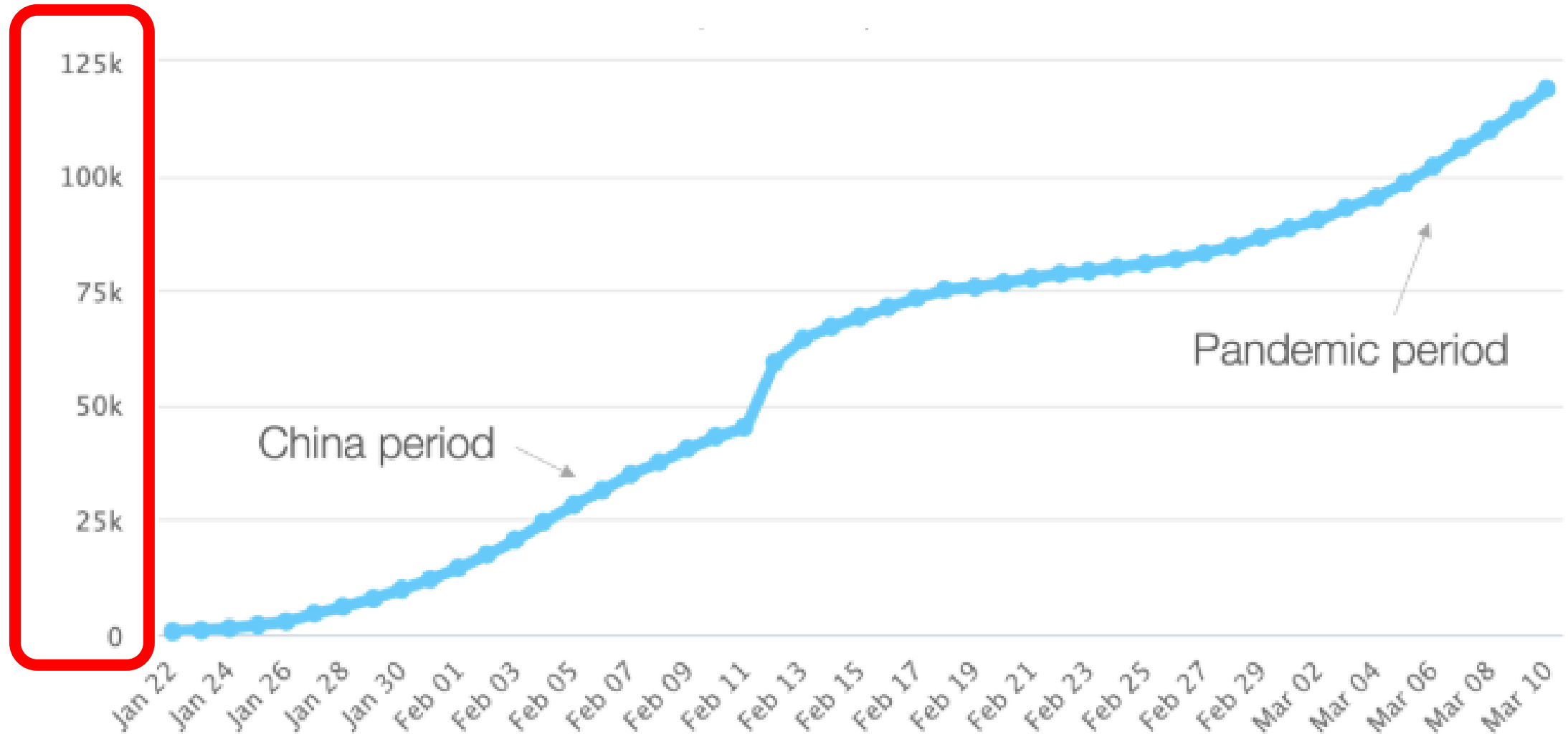


Chart 1: Total Worldwide Cases of Coronavirus

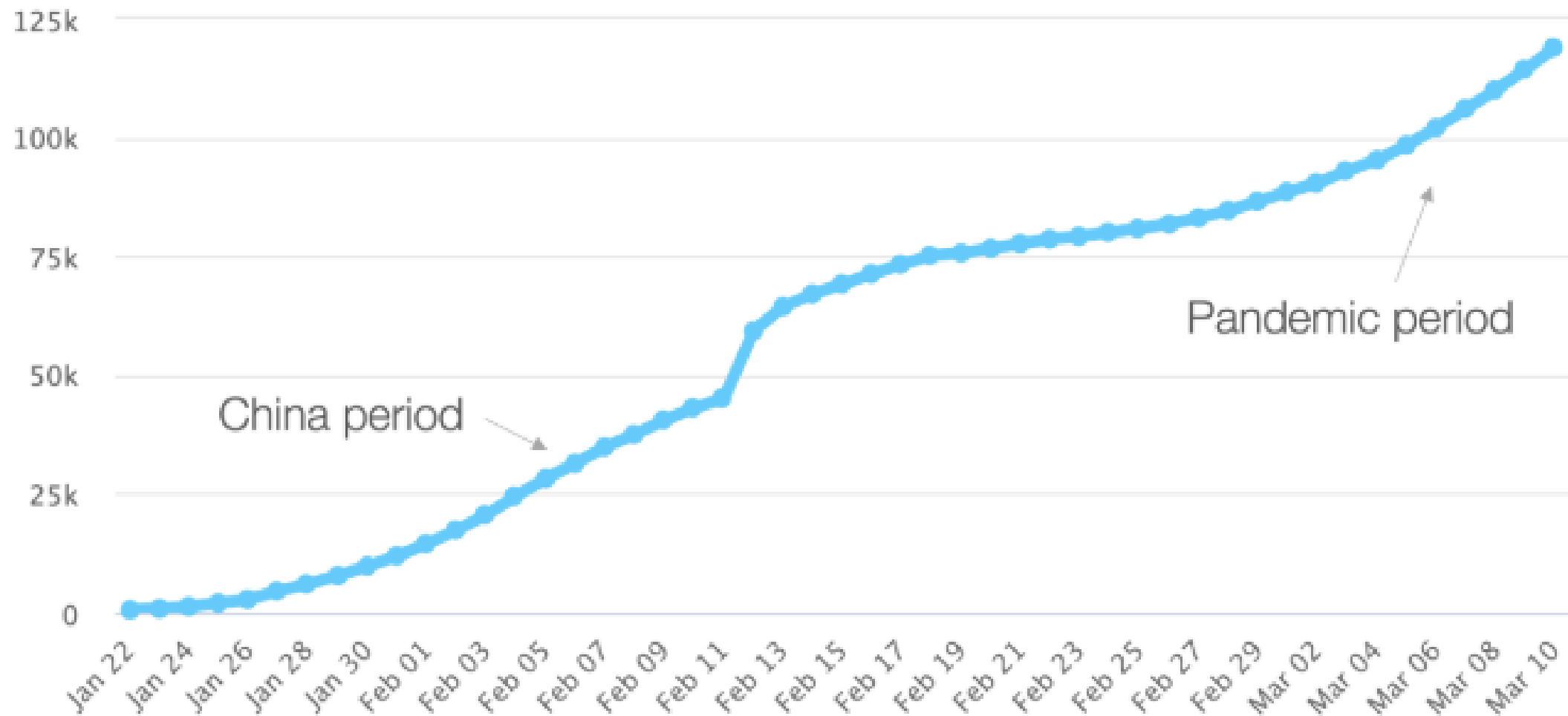
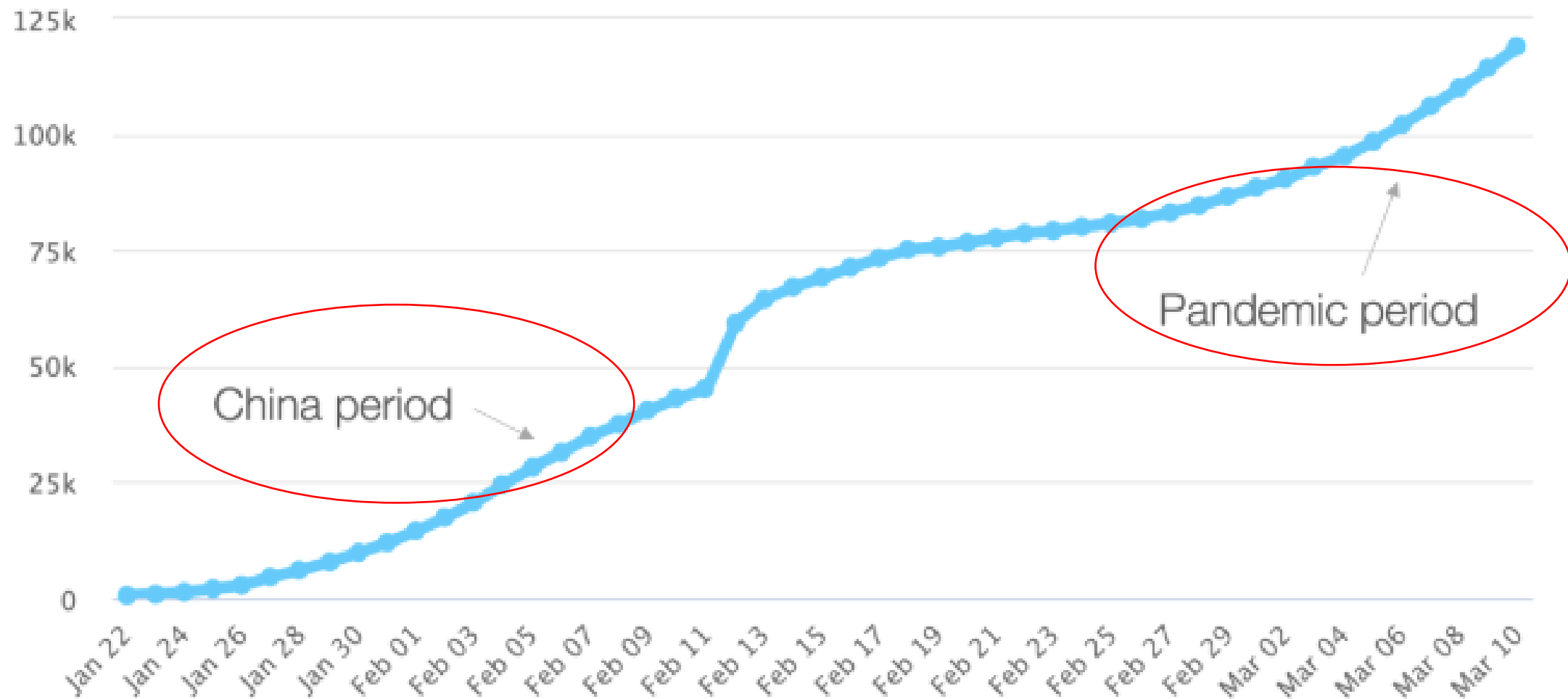


Chart 1: Total Worldwide Cases of Coronavirus



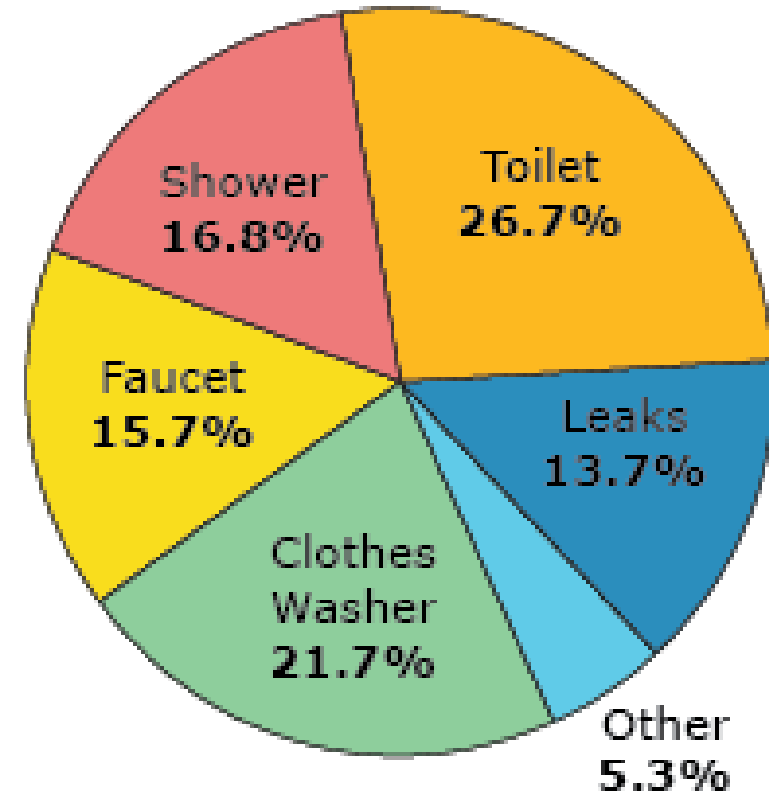
Tipos de Gráficas

- Gráfica de Pastel
- Gráfica de Barras
- Histograma
- Gráficas lineales
- Polígono de frecuencia

Gráficas de Pastel

- Se utilizan para representar **porcentajes (proporciones)**.
- Deben sumar 1.
 - La suma de las partes no puede ser más ni menos que el TOTAL.

How Much Water Do We Use?



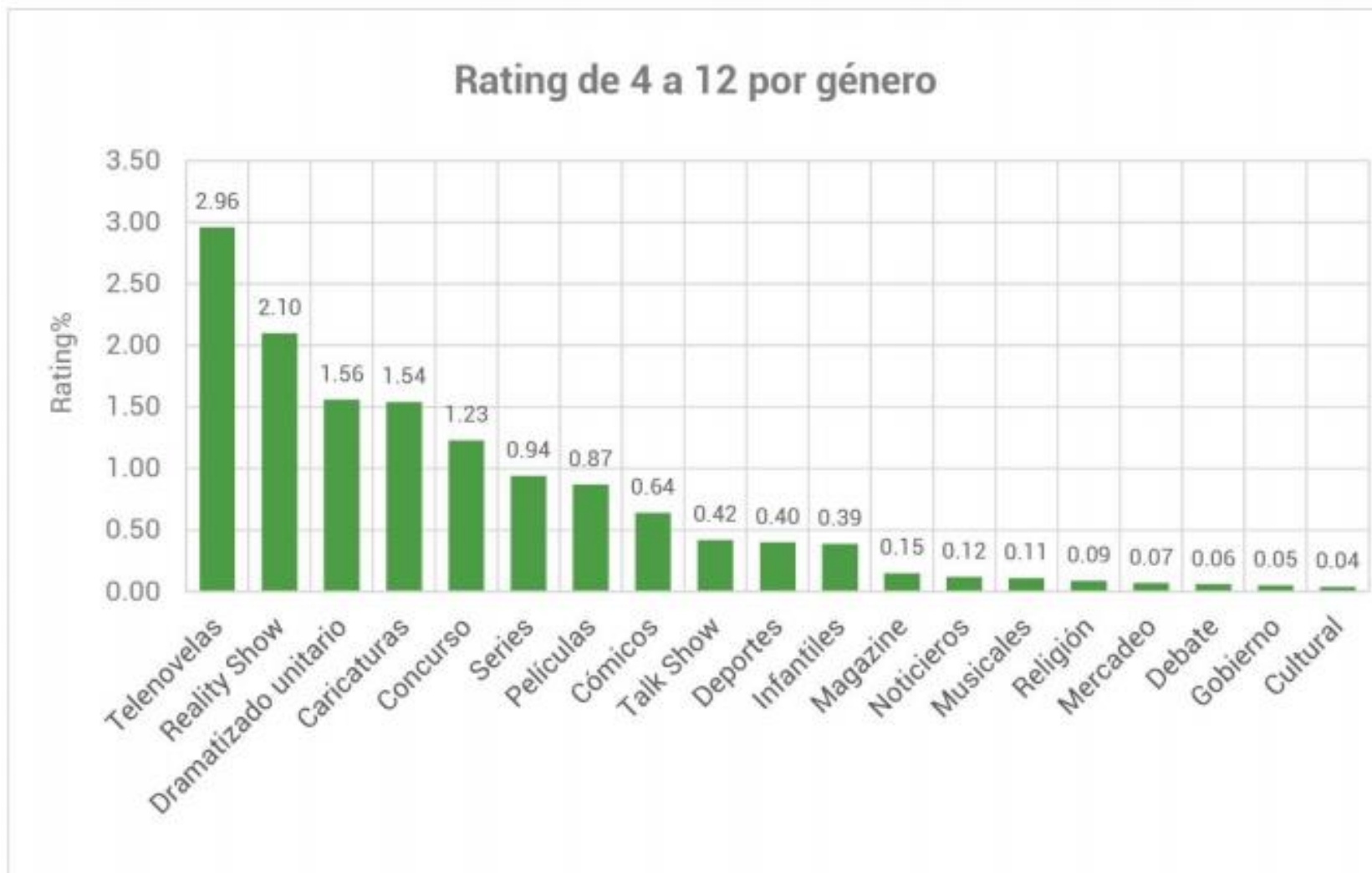
Source: American Water Works Association Research Foundation, "Residential End Uses of Water." 1999

Gráficas de Barras

Presenta los valores absolutos de cierta variable (y) en diferentes grupos o categorías (x).

X = Variable categórica

Y = Variable numérica

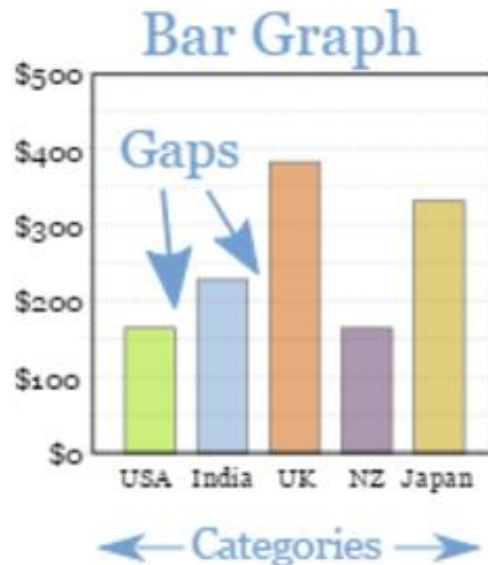


Gráficas de Barras vs Histogramas

Gráfica de Barras

X = Variable categórica

Y = Variable continúa



Histogramas

X = Variable numérica (Continua)

Y = Frecuencias



Gráficas Lineales

Presentan los cambios en la variable dependiente (y) a lo largo de diferentes valores en (x).

X = Variable Numérica

Y = Variable Numérica

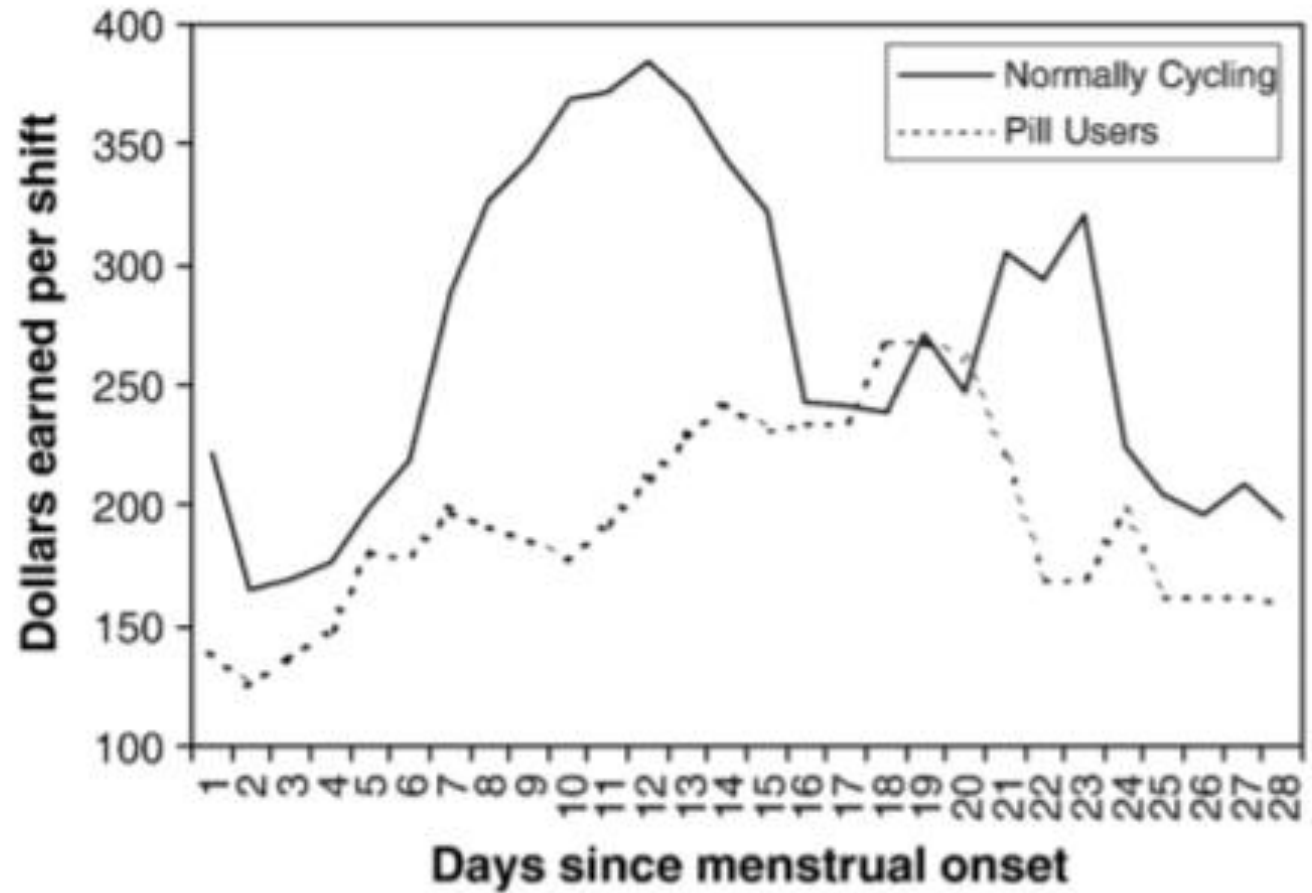


Fig. 1. Effects of ovulatory cycle (Days 1–28) on average tip earnings per shift, for normally cycling women versus women using hormonal contraception (pill users); each data point represents a 3-day average of the indicated day, the previous day, and the following day.

... Pictogramas

Representación de ideas e información por medio de gráficos. Generalmente, se usan para **divulgar información de la manera más sencilla y 'atractiva' posible.**

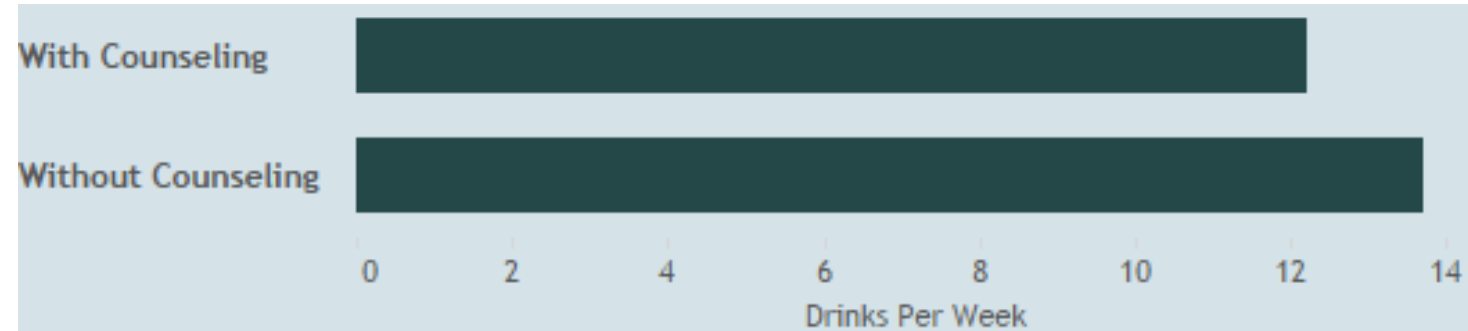


... Pictogramas

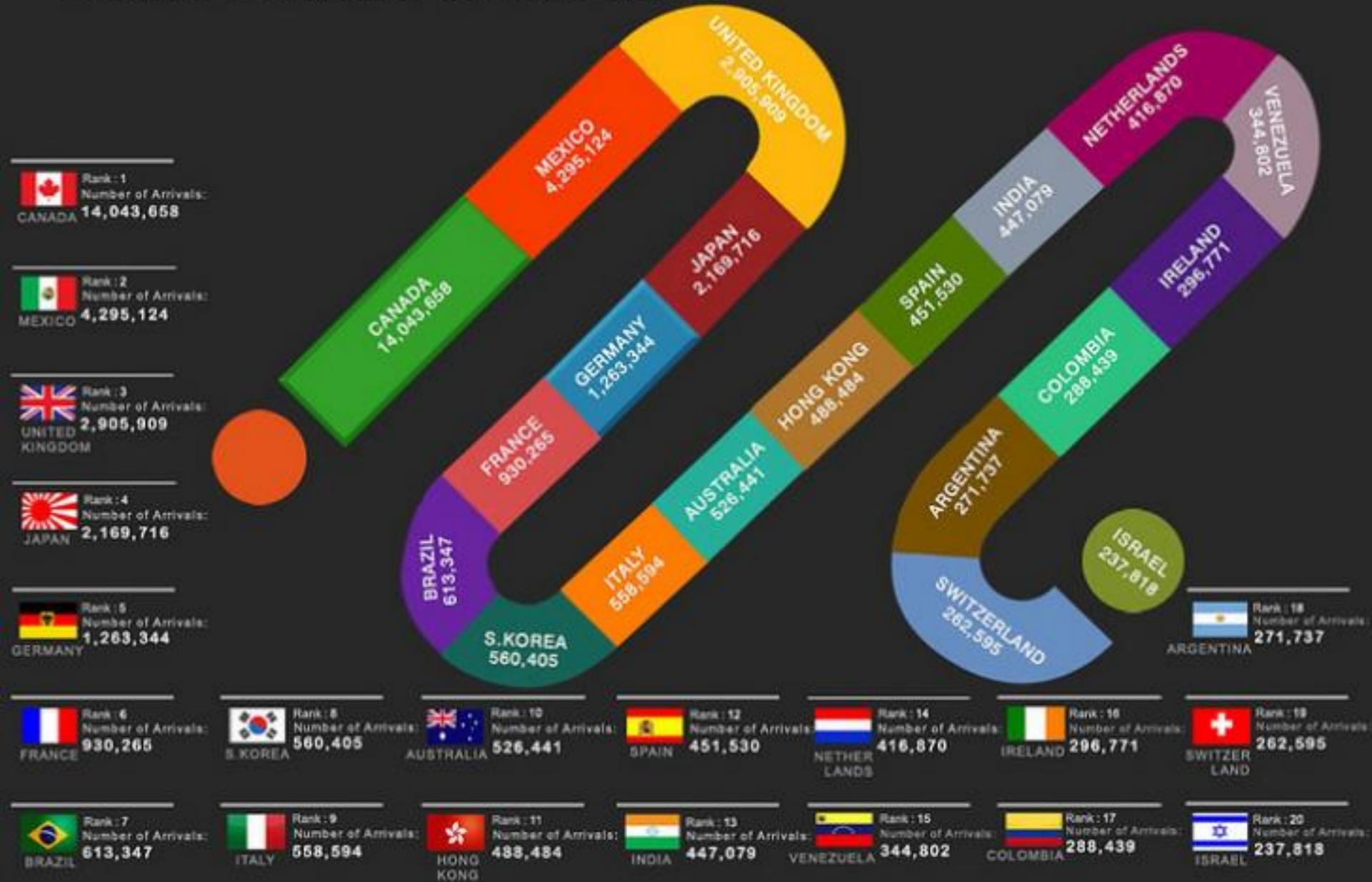
Siguen el mismo principio que las gráficas de barras, **pero** utilizando dibujillos ilustrativos en lugar de barras.

‘Atractivos’, pero poco precisos.

X = Variable Categórica
Y = Variable Numérica (con dibujitos)



Top 20 Tourist Generating Countries To UNITED STATES OF AMERICA

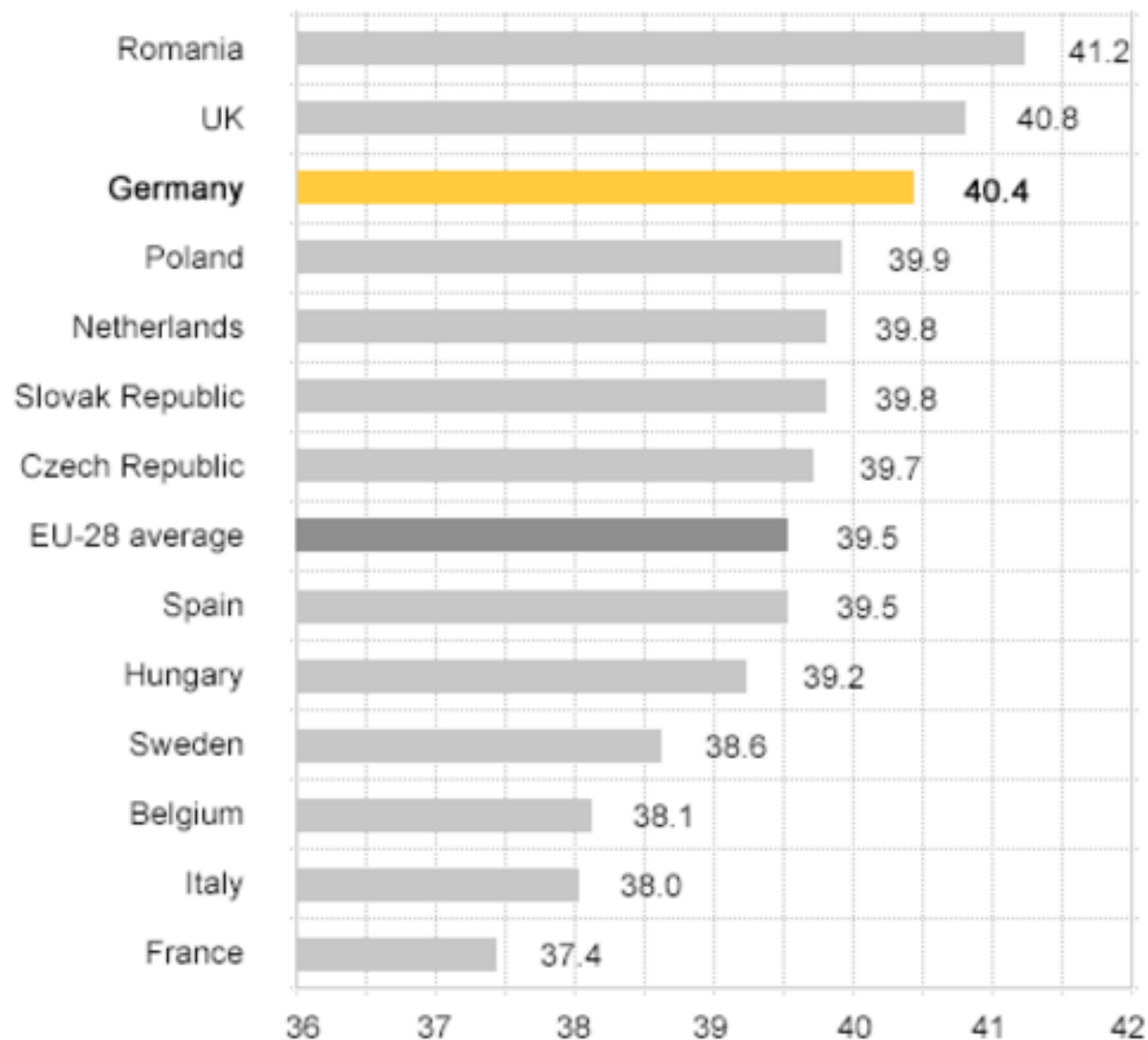


‘Atractivos’...
pero poco
precisos.

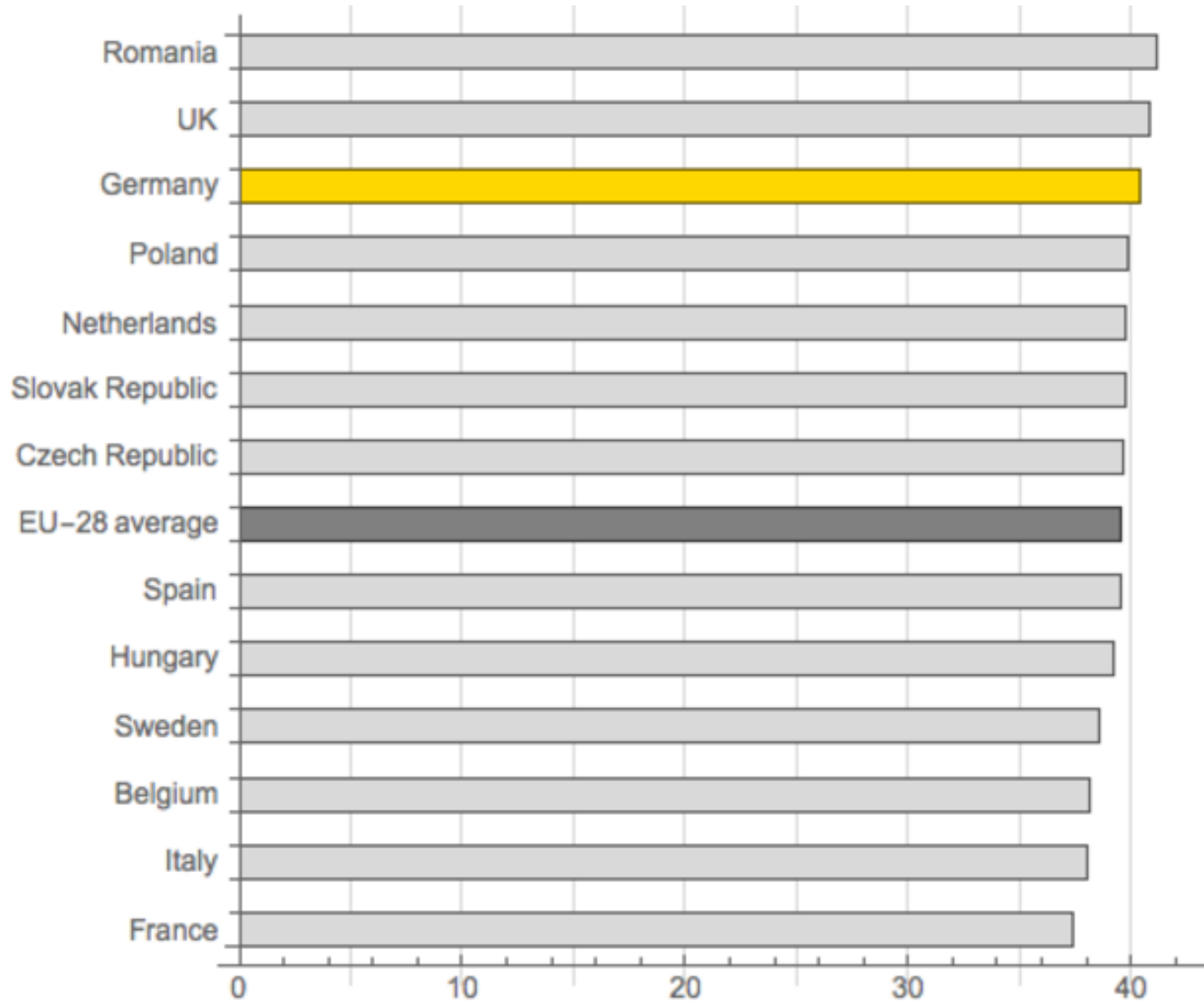
¡Cuidado con las gráficas sugerentes*!

1) Los valores en los Ejes

Ejemplo:



¡Corrijamos!



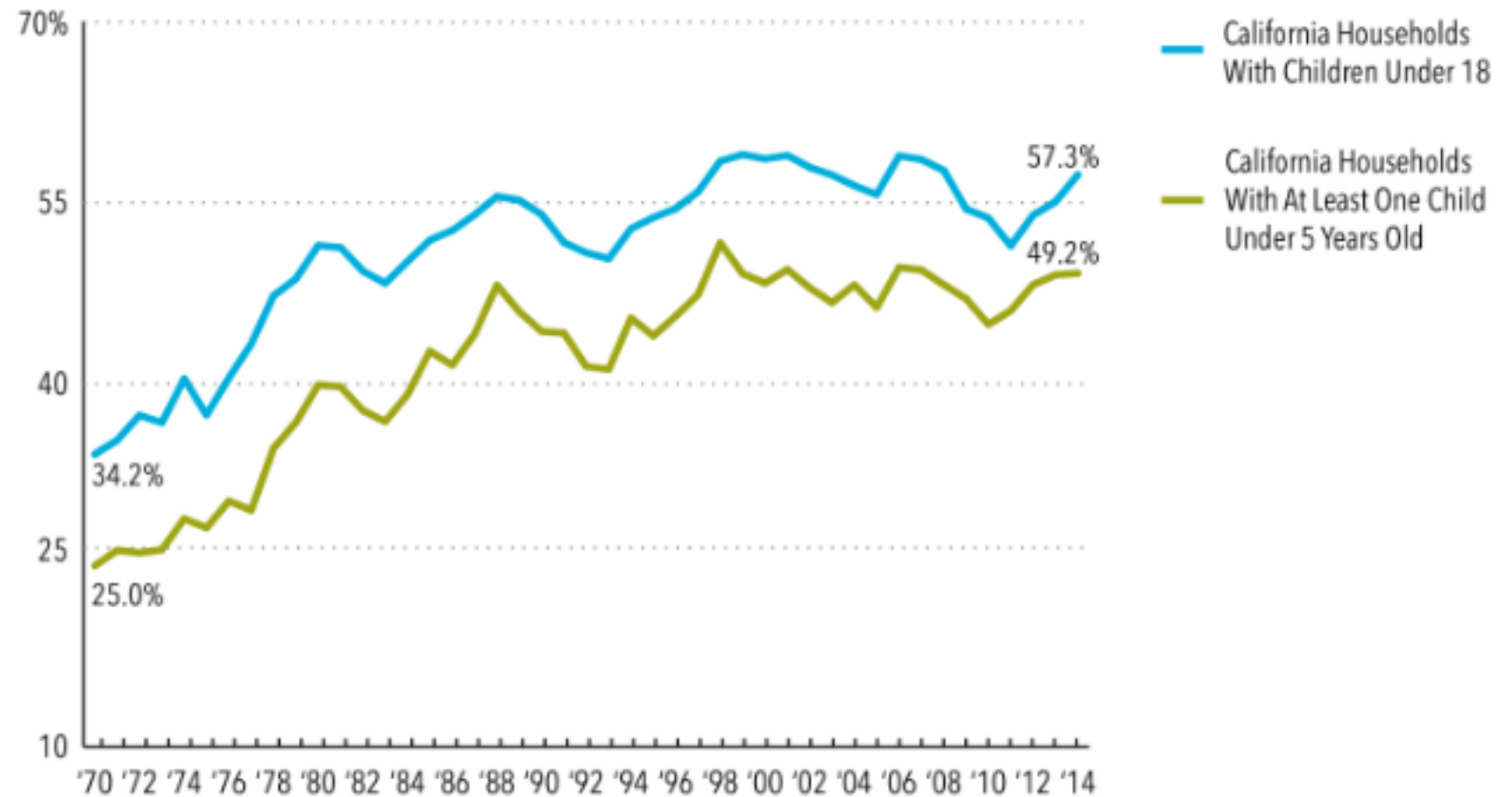
¿“Las gráficas que no parten del valor 0 son *”?

No necesariamente

More California Households Have All Parents Working, Making Access to Child Care an Important Priority

Percentage of California Households Where All Parents Work, 1970 to 2014

Omitir el valor 0 **tiene sentido cuando** lo que nos interesa observar es la **tendencia** de nuestra variable dependiente a cambiar conforme cambia la variable independiente.



Note: A "household where all parents work" includes single-parent households and dual-earner households. Parents include stepparents and adoptive parents.
Source: Budget Center analysis of US Census Bureau data



California Budget
& Policy Center
Independent Analysis. Shared Prosperity.

¿Omitir el valor 0 es bueno o malo?

— — — —
Es **malo** cuando...

Se promueve que los lectores salten a las conclusiones erróneas acentuando injustificadamente la diferencia entre dos grupos a comparar (variables categóricas)

En otras palabras, las Gráficas de Barras e Histogramas siempre deben partir de 0

Ejemplo:

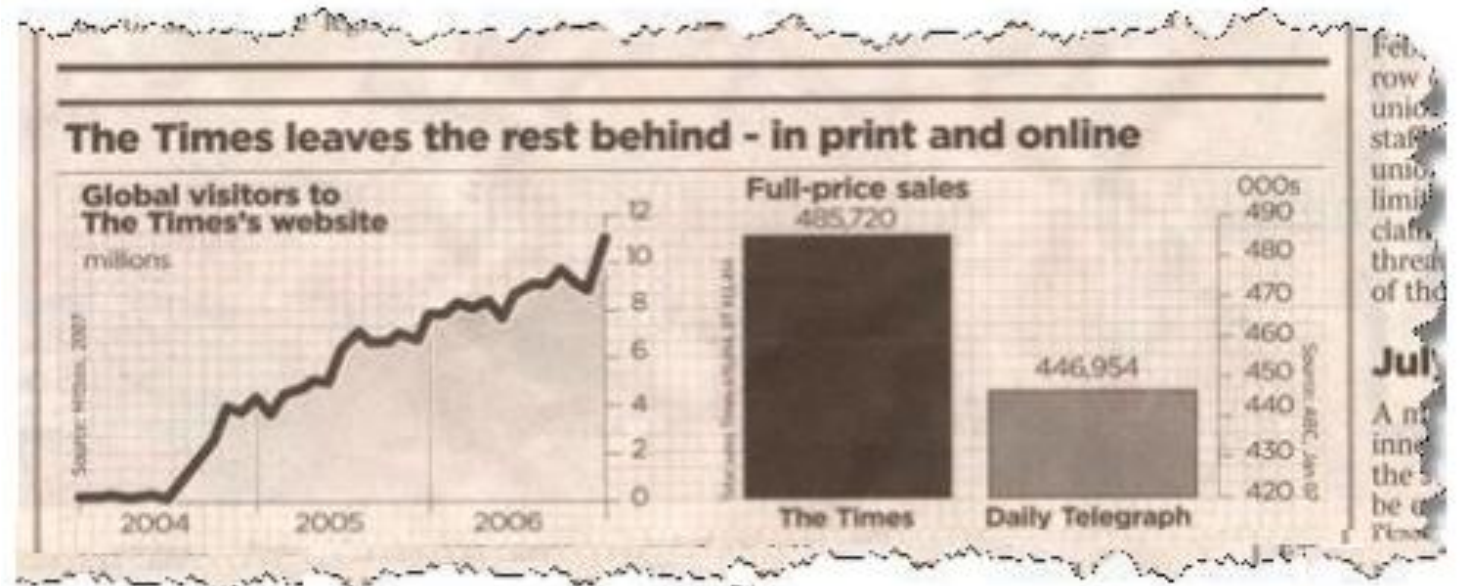


Image: University of Kentucky.

Over 100 Million Now Receiving Federal Welfare

2:40 PM, AUG 8, 2012 • BY DANIEL HALPER



PRINT PAGE



PRINT



LARGER TEXT



SMALLER TEXT



ALERTS

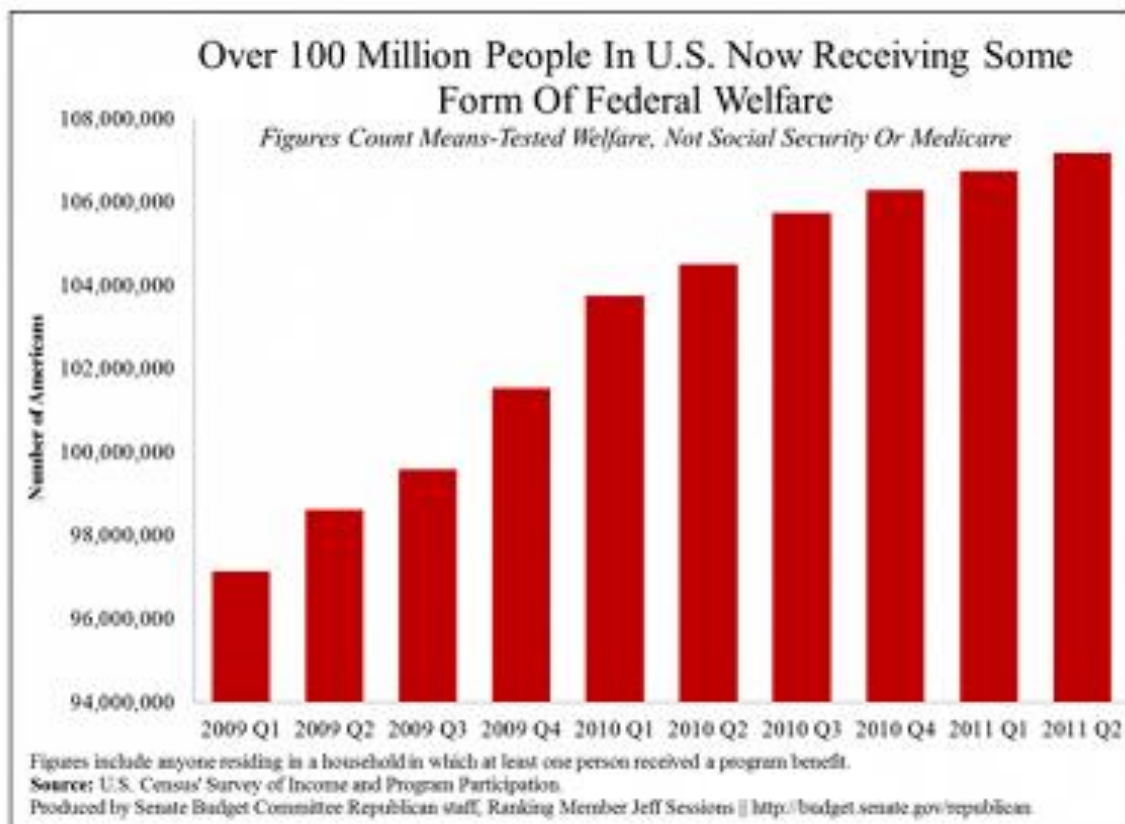


f

t

g+

A new chart set to be released later today by the Republican side of the Senate Budget Committee details a startling statistic: "Over 100 Million People in U.S. Now Receiving Some Form Of Federal Welfare."

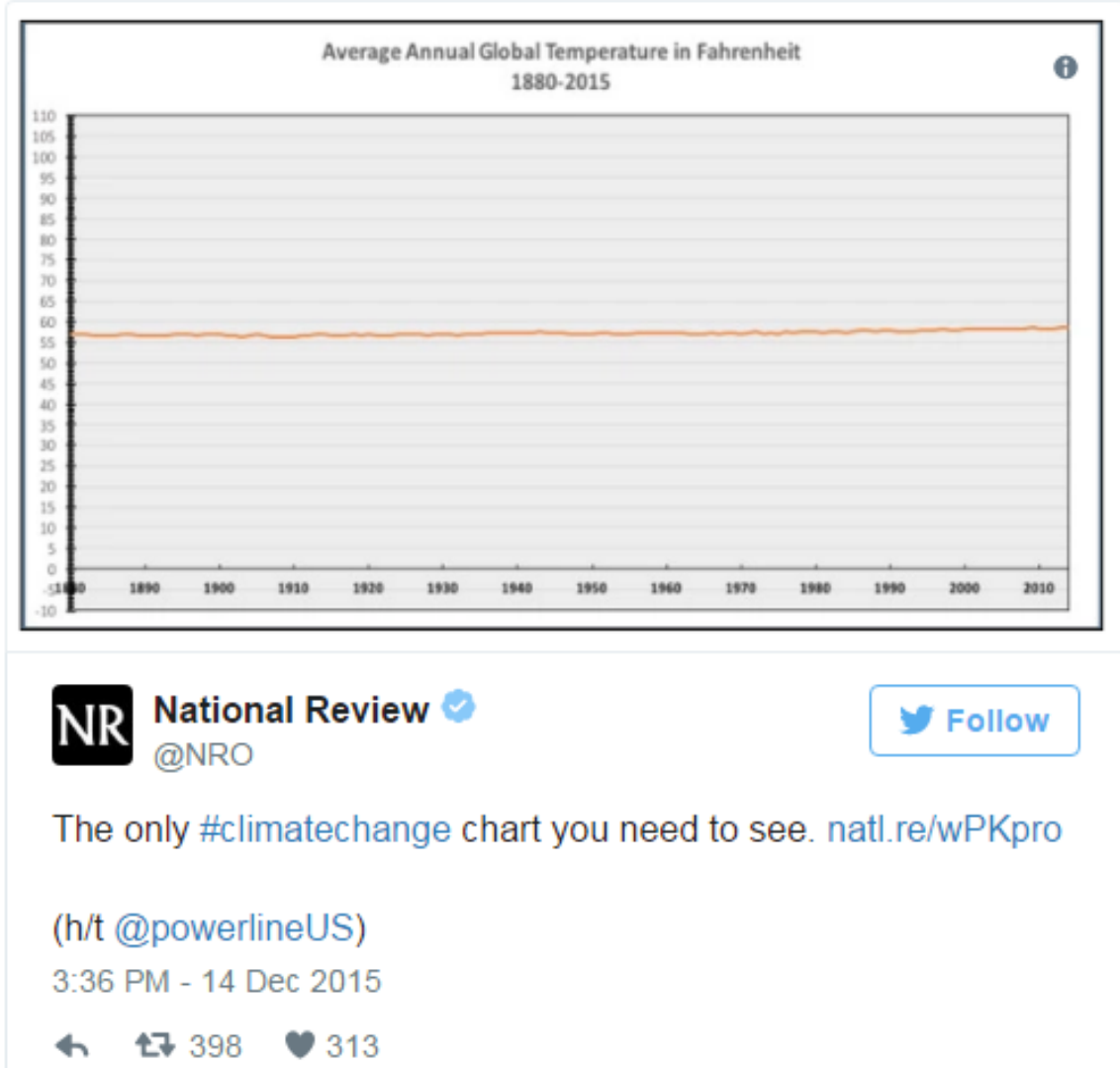


Ejemplo:

Ejemplo:

-No existe el cambio climático

Esta gráfica sí que incluye el valor 0, ¿de qué manera esto repercute las conclusiones?



¿Omitir el valor 0 es bueno o malo?

— — — —
Es **malo** cuando...

Se promueve que los lectores salten a las conclusiones erróneas acentuando injustificadamente la diferencia entre dos grupos a comparar (variables categóricas)

En otras palabras, las Gráficas de Barras e Histogramas siempre deben partir de 0

Tiene sentido si...

1.- Lo que interesa es observar los cambios en Y a través de los cambios en X.

2.- Tenemos razones para asumir que la variable difícilmente va a mostrar el valor 0

- Gráficas Lineales

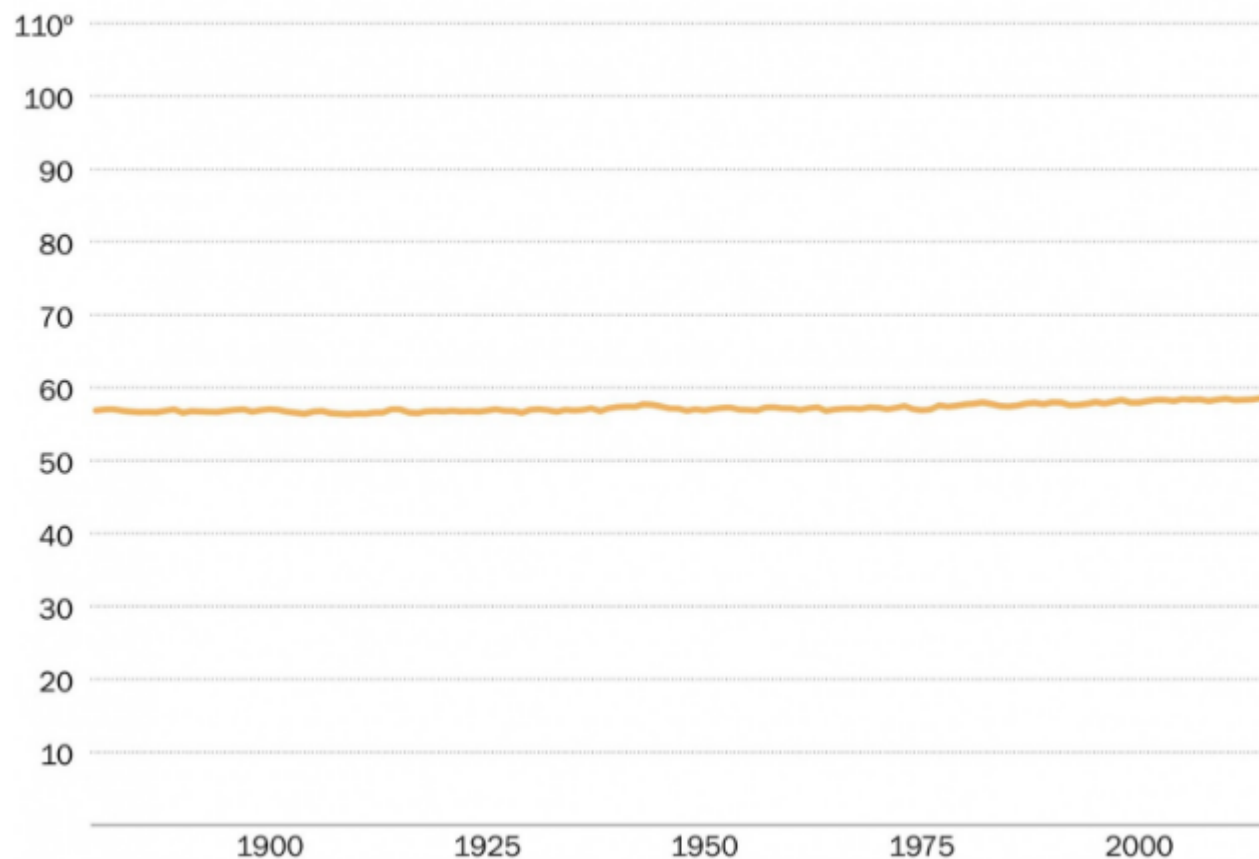
2) La Escala/Intervalo cubierto por los ejes

Average Annual Global Temperature in Fahrenheit
1880-2015



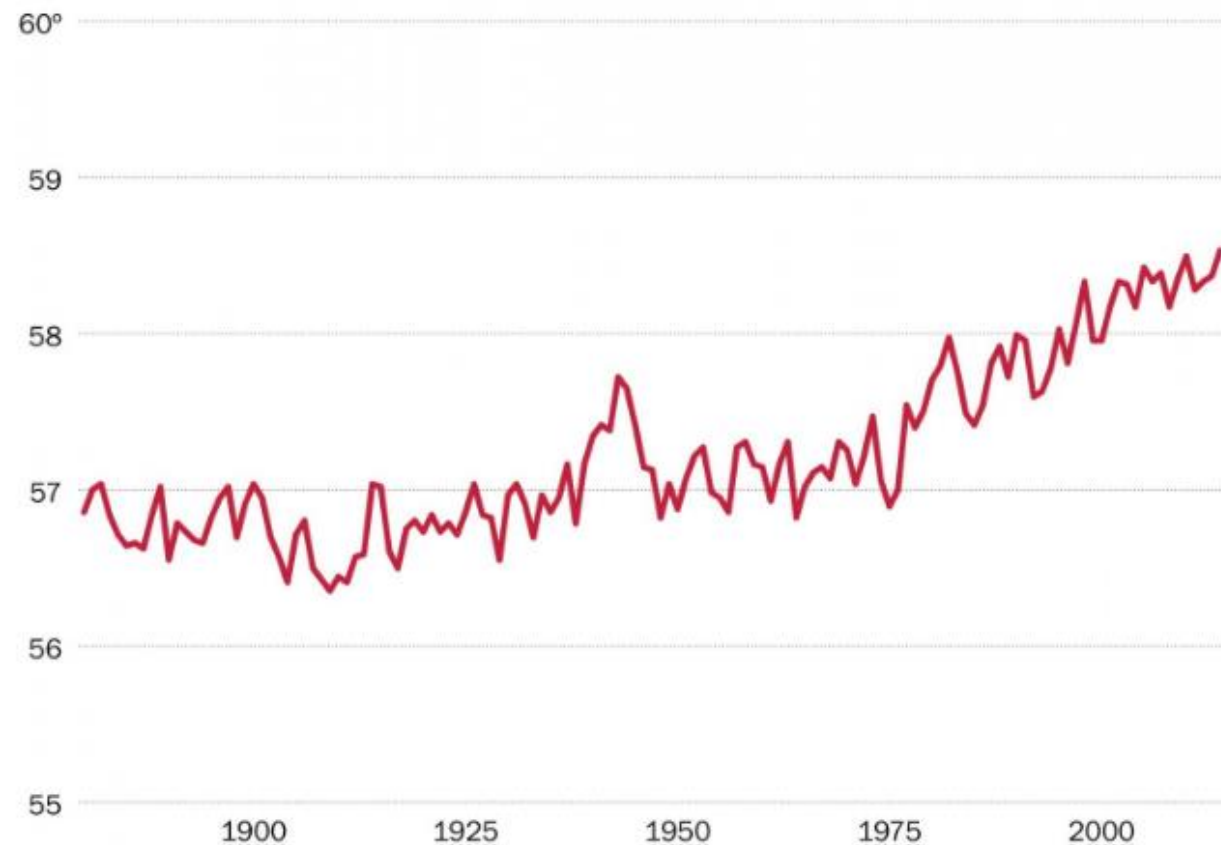
Average global temperature by year, 50x scale

Data from NASA/GISS.



Average global temperature by year

Data from NASA/GISS.



The Washington Post - “Why this National Review Global temperature graphic is so misleading” by Philip Bump (Link en ‘Fuentes’)

3) Proporcionalidad

Ejemplo:

-Los 5 países con más medallas, parecen guardar siempre una distancia proporcional.

Según la representación de medallas ganadas en Alemania, dos figuras valen 500 medallas; pero en Francia se utiliza una figura más para representar una diferencia de 24 medallas.



Figure 1. Source: Erickson Times

Ejemplo:



Naomi Robbins, CONTRIBUTOR

I help people communicate data clearly with graphs. **FULL BIO** ✓

Opinions expressed by Forbes Contributors are their own.

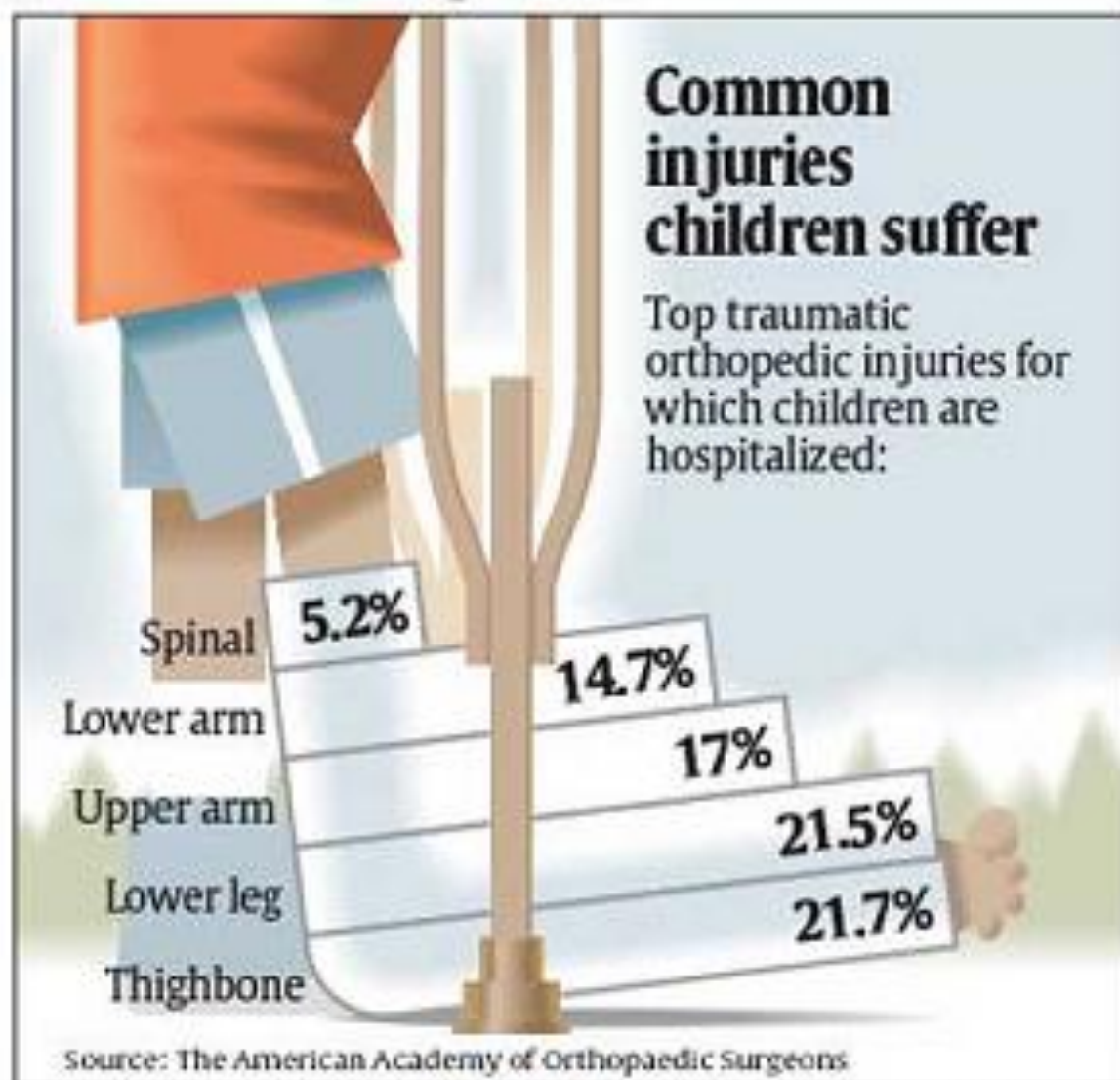
I came across the graph in Figure 1 while making my year end contributions. The source is a non-profit organization that provides invaluable services to our society, so I'd rather not name it here.



4) Títulos y encabezados sugerentes

Ejemplo:

USA TODAY Snapshots™



By Shannon Reilly and Frank Pompa, USA TODAY

5) La información cuando es mucha, se pierde.

20,000 -

Unisex names history

Some names are predominantly male and others female. These 121 baby names were split near evenly during the years they are shown here.

10,000 -

0 -

10,000 -

20,000 -

1930

1940

1950

1960

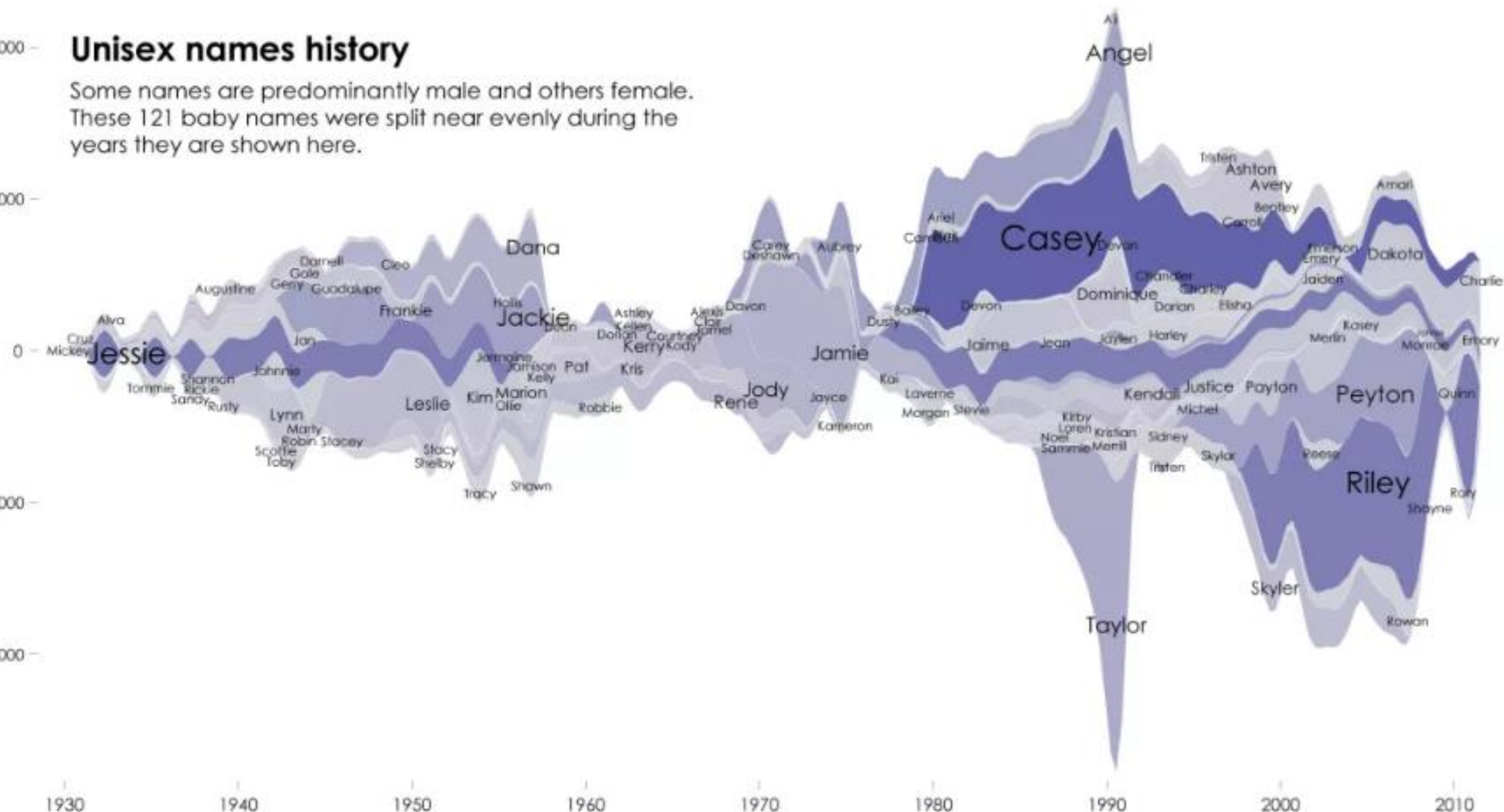
1970

1980

1990

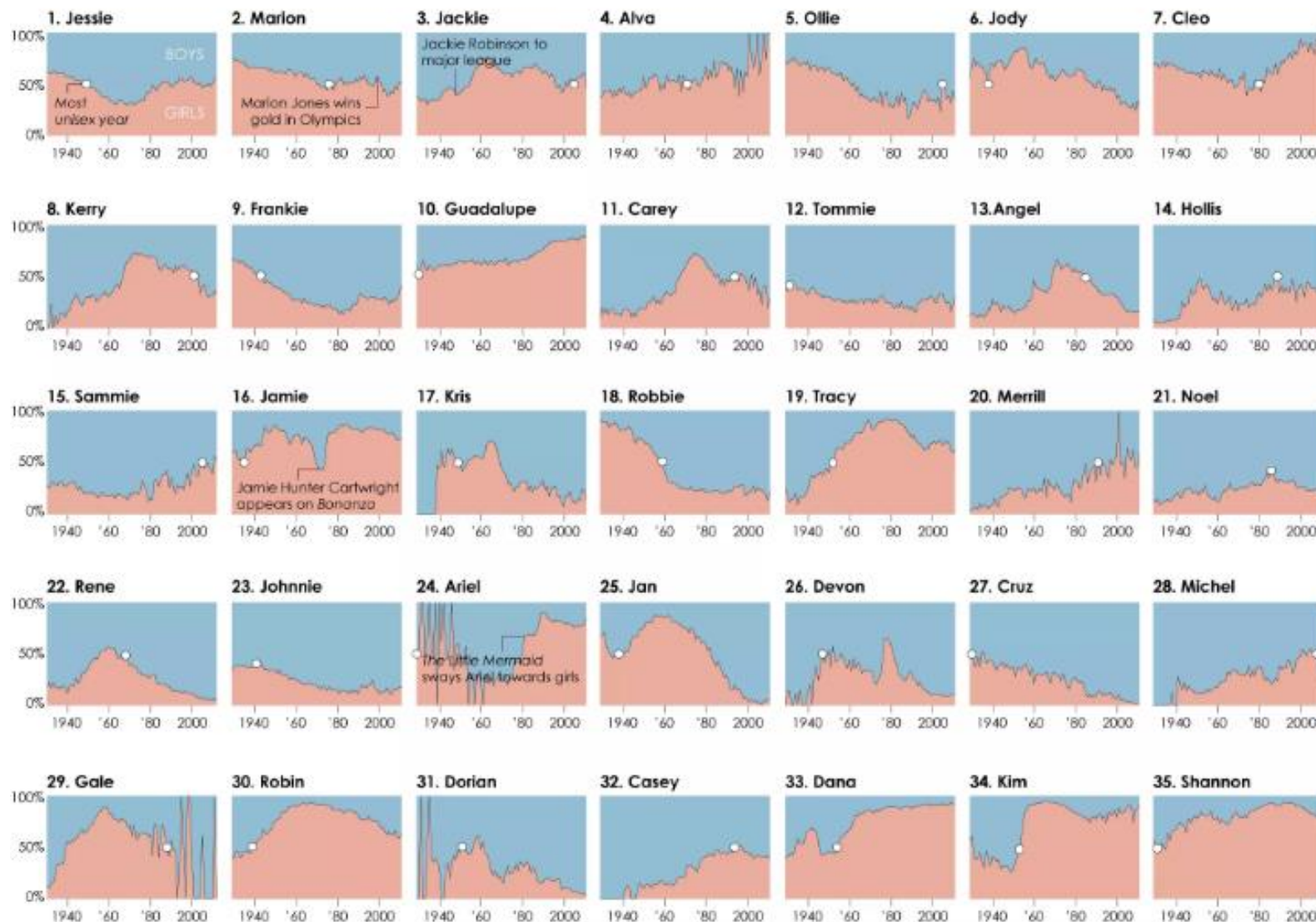
2000

2010



The Most Unisex Names in US History

BY NATHAN YAU / POSTED TO [DATA UNDERLOAD](#) / TAGS: NAMES



Source: Social Security Administration | By: <http://flowingdata.com>

Conclusión

Los gráficos de datos son una ventana a los resultados encontrados e invita al lector a sacar **sus propias** conclusiones.. ¿o no?.

Cambios aparentemente sutiles en la forma de presentar los datos pueden tener un impacto notorio en la forma en que se interpreta.

Como lector... hay que preguntarnos si la gráfica que se nos muestra refleja los datos obtenidos ó si fue diseñada para reflejar las expectativas de lo que se esperaba encontrar.

Como investigador...

— — — —