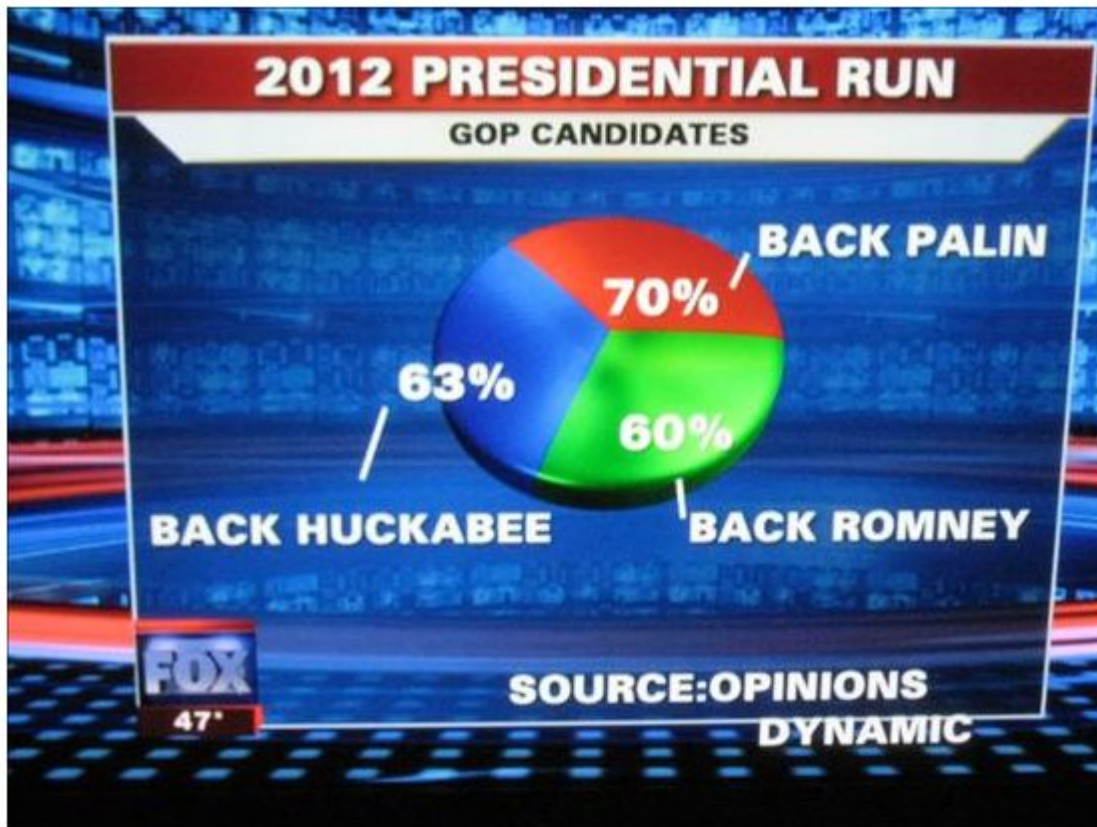


Lectura de Gráficas

La importancia de cómo se presentan los datos



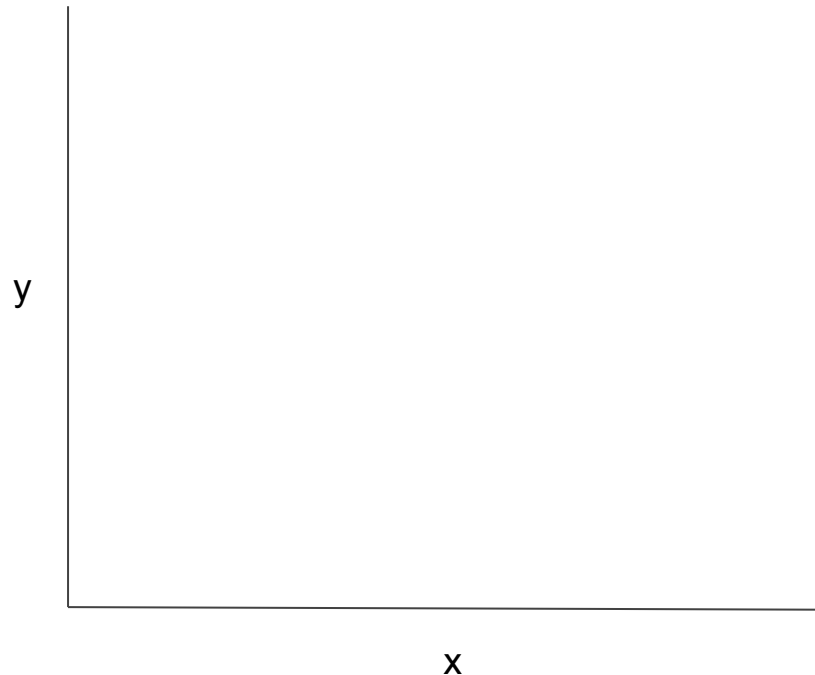
?????

Image: <http://flowingdata.com/2009/11/26/fox-news-makes-the-best-pie-chart-ever/>

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.



Tipos de Gráficas

Tipos de Gráficas

— — —

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

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)

Variables Numéricas



Variables Continuas

Varían de manera continua (se pueden representar con números decimales). Por ejemplo:

- Peso
 - Altura
 - Distancia
-
- **Intervalares:** La distancia entre cada valor es la misma.
 - La diferencia entre 1 y 2 kg, es la misma que entre 37 y 38 kg.
 - **De Razón:** Son variables intervalares que parten de un 0 absoluto.

Variables Discretas

NO varían en un continuo (sólo se pueden representar con números enteros). Por ejemplo:

- Número de casos en una categoría específica

Variables Categóricas

— — —

Nominales

Son aquellas que asignan una etiqueta de pertenencia a una categoría particular (Nombran las cosas)

- Género
- Escolaridad
- Género musical favorito

Dicotómicas

Distinguen entre la presencia y ausencia de una cualidad.

- Sexo
- Lectores - No lectores
- Fumadores - No Fumadores

Ordinales

Sabemos que llevan un orden, pero no conocemos la diferencia entre cada salto.

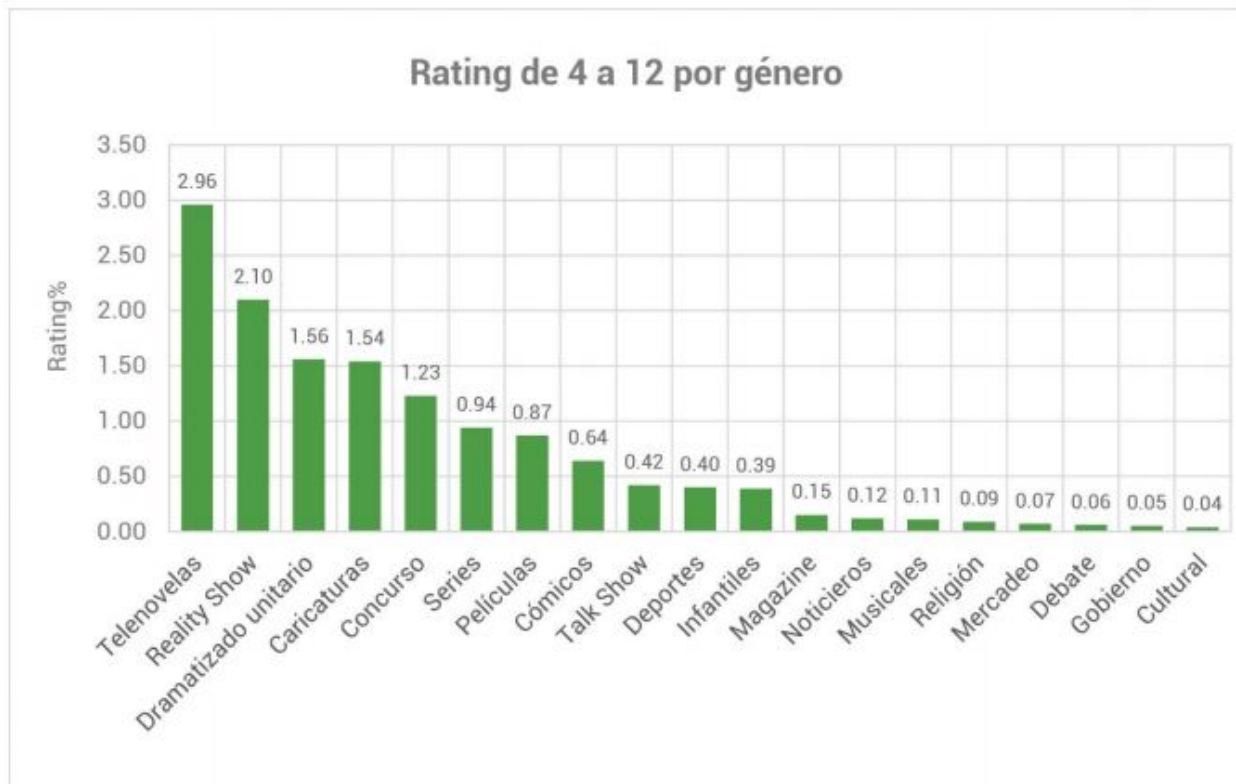
- Escalas Likert
 - 'Poco', 'Medio', 'Mucho'
 - 'Siempre', "A veces", "Nunca"

Gráficas de Barras

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

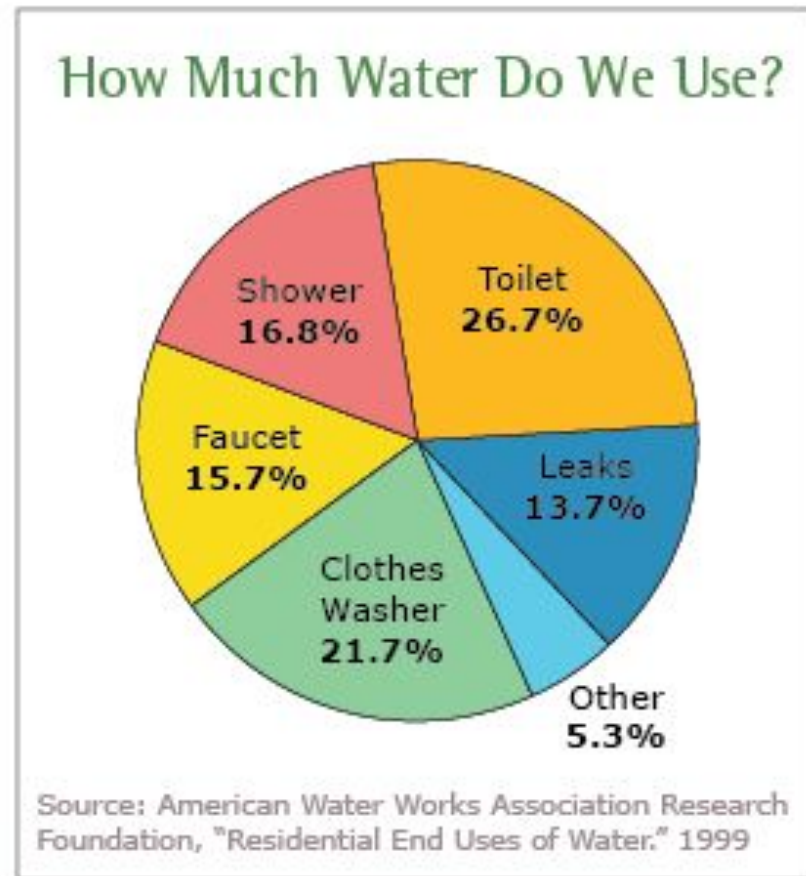
X = Variable categórica

Y = Variable numérica



Gráficas de Pastel

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



Ejemplo: ¿Qué pasó aquí?

— — —
Al reportar los resultados de una encuesta sobre Opinión Política, el presentador de un noticiario televisivo dice

“

“Sarah Palin tiene más soporte entre los candidatos Republicanos para su nominación como presidente en el año 2012.”

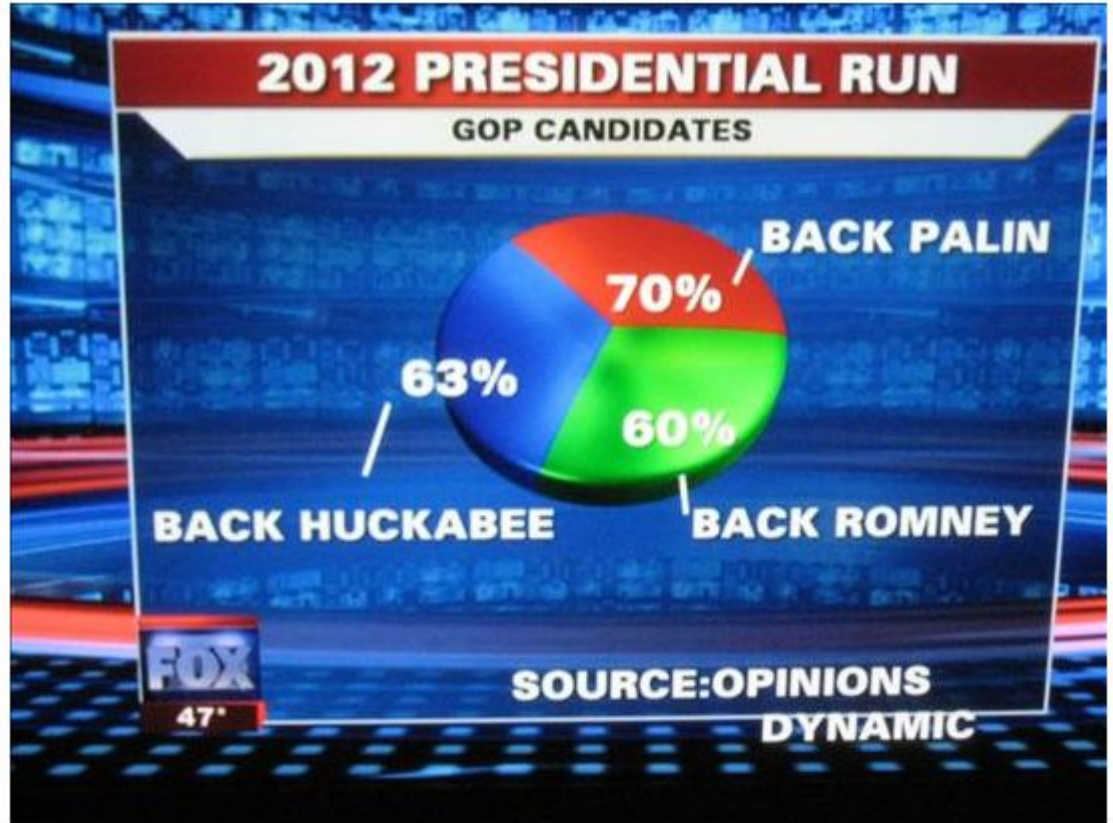


Image: <http://flowingdata.com/2009/11/26/fox-news-makes-the-best-pie-chart-ever/>

“Los datos tienen que estar mal”.

— — —

Los datos provienen de una encuesta telefónica hecha en Noviembre del 2009 a 900 votantes registrados a lo largo de E.E.U.U.

“Voy a leerle una lista con el nombre de varios individuos. Por favor, indíqueme si tiene una opinión favorable o no-favorable sobre cada uno. Si no les reconoce, por favor, hágame saber”

Sarah Palin

	Favorable	Unfavorable	Can't say	Never heard
17-18 Nov 09	47%	42	10	2
Democrats	25%	62	9	4
<u>Republicans</u>	<u>70%</u>	21	8	1
Independents	49%	38	11	1

Mike Huckabee

	Favorable	Unfavorable	Can't say	Never heard
17-18 Nov 09	45%	23	19	13
Democrats	28%	30	21	20
<u>Republicans</u>	<u>63%</u>	15	17	5
Independents	44%	28	17	11

Mitt Romney

	Favorable	Unfavorable	Can't say	Never heard
17-18 Nov 09	38%	27	18	17
Democrats	20%	36	16	29
<u>Republicans</u>	<u>60%</u>	16	16	8
Independents	35%	34	18	12

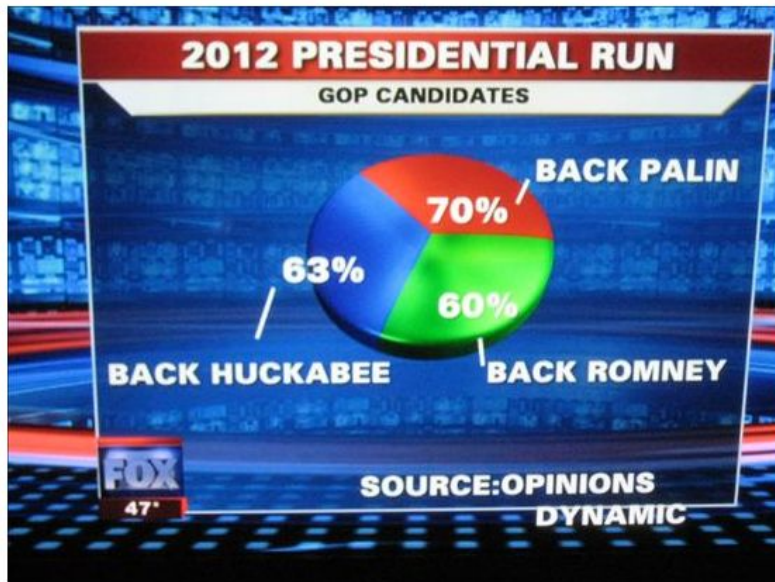
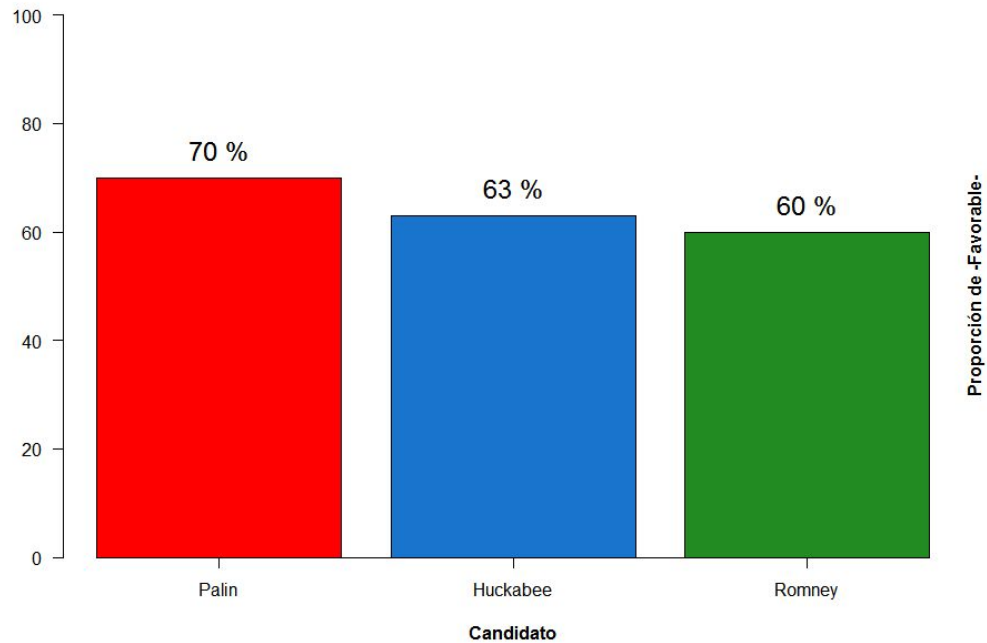


Image: <http://flowingdata.com/2009/11/26/fox-news-makes-the-best-pie-chart-ever/>

2012 Presidential Run: GOP Candidates

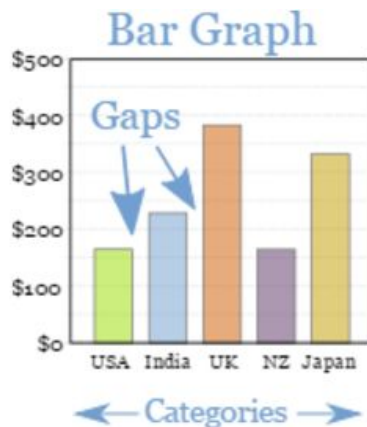


Gráficas de Barras vs Histogramas

Gráfica de Barras

X = Variable categórica

Y = Variable continua



Histogramas

X = Variable numérica (Continua)

Y = Frecuencias



Polígonos de frecuencia

— — —

Representan la misma información que los histogramas, sin la presentación en barras.

Son útiles para entender las distribuciones de probabilidad.

Permiten observar de manera clara distribuciones de frecuencia acumulada.

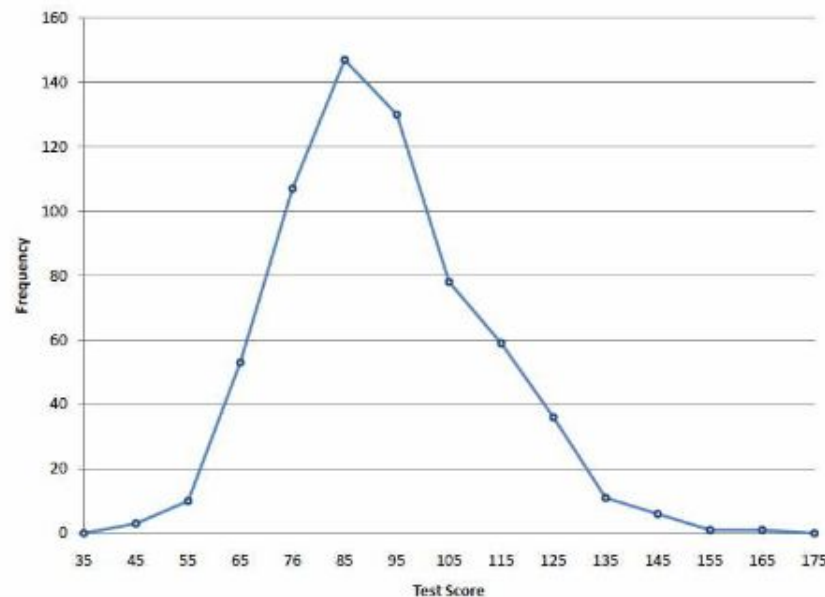


Figure 1. Frequency polygon for the psychology test scores.

Gráficas Lineales

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

X = Variable Numérica

Y = Variable Numérica

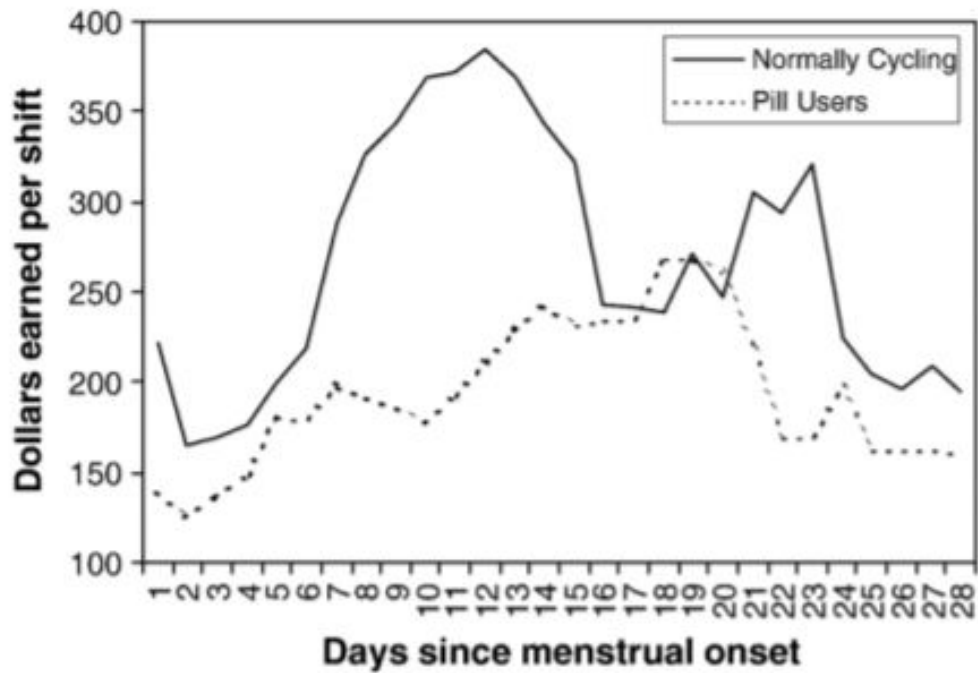


Fig. 1. Effects of ovulatory cycle (Days 1–28) on average tip earnings per shift, for normally cycling women versus women using hormonal contraception (pill users); each data point represents a 3-day average of the indicated day, the previous day, and the following day.

... Pictogramas

— — —

“Se refieren a la representación de ideas e información por medio de gráficos”

Generalmente, se usan para divulgar información de la manera más sencilla y ‘atractiva’ posible.



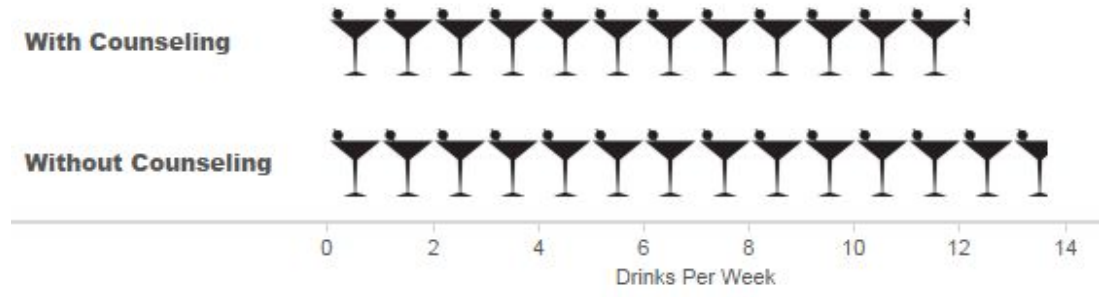
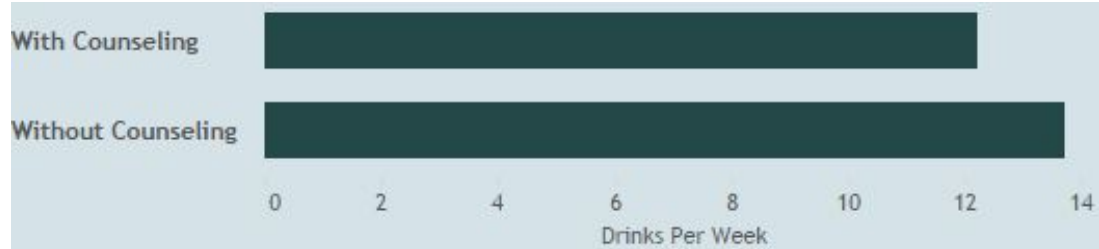
... Pictogramas

— — —

Siguen el mismo principio que las gráficas de barras, **pero** utilizando dibujitos 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

Rank: 1
Number of Arrivals:
CANADA 14,043,658

Rank: 2
Number of Arrivals:
MEXICO 4,295,124

Rank: 3
Number of Arrivals:
UNITED KINGDOM 2,905,909

Rank: 4
Number of Arrivals:
JAPAN 2,169,716

Rank: 5
Number of Arrivals:
GERMANY 1,263,344

Rank: 6
Number of Arrivals:
FRANCE 930,265

Rank: 7
Number of Arrivals:
BRAZIL 613,347



Rank: 18
Number of Arrivals:
ARGENTINA 271,737

Rank: 8
Number of Arrivals:
S. KOREA 560,405

Rank: 10
Number of Arrivals:
AUSTRALIA 526,441

Rank: 12
Number of Arrivals:
SPAIN 451,530

Rank: 14
Number of Arrivals:
NETHERLANDS 416,870

Rank: 16
Number of Arrivals:
IRELAND 296,771

Rank: 19
Number of Arrivals:
SWITZERLAND 262,595

Rank: 9
Number of Arrivals:
ITALY 558,594

Rank: 11
Number of Arrivals:
HONG KONG 488,484

Rank: 13
Number of Arrivals:
INDIA 447,079

Rank: 15
Number of Arrivals:
VENEZUELA 344,802

Rank: 17
Number of Arrivals:
COLOMBIA 288,439

Rank: 20
Number of Arrivals:
ISRAEL 237,818

‘Atractivos’...
pero poco
precisos.

¡Cuidado con las gráficas sugerentes*!

1) Los valores en los Ejes

Ejemplo:

-¡Los impuestos se irán
por los cielos!

— — —

Gráfica presentada en Fox News para ‘ilustrar’ las consecuencias de la cancelación del programa de reducción de impuestos iniciado en el gobierno de George W. Bush.

A primera vista, parece que
los impuestos se
cuadruplicaron... ¿es este el
caso?

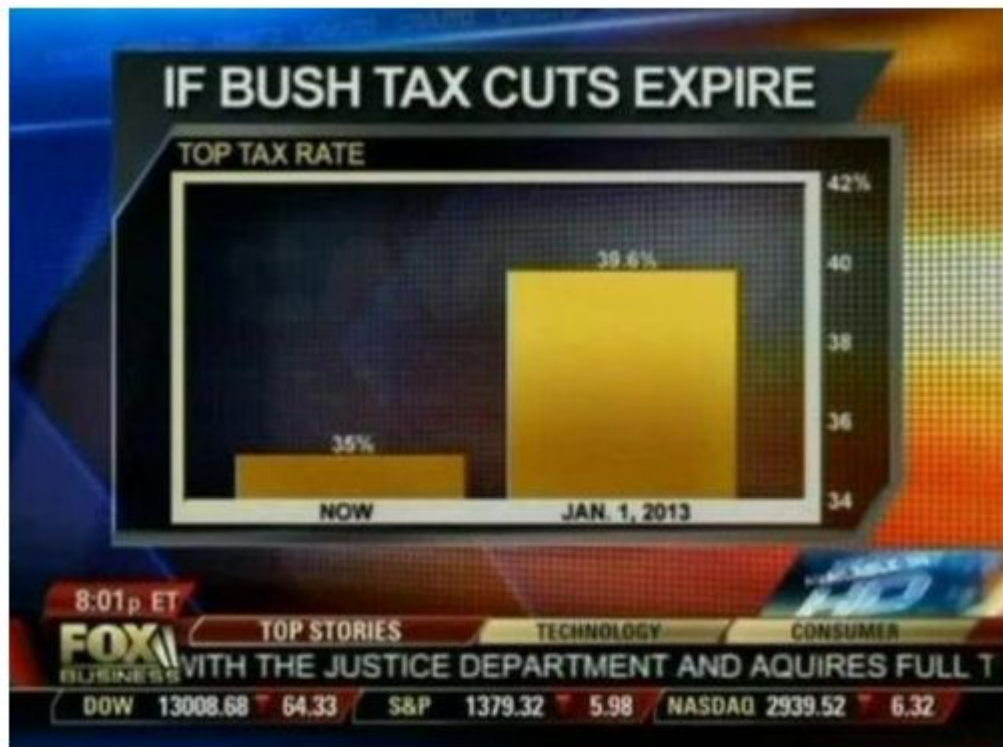


Figure 1. Source: <https://twitter.com/DanaDanger/status/230851016344600576/photo/1/large>

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

— — —

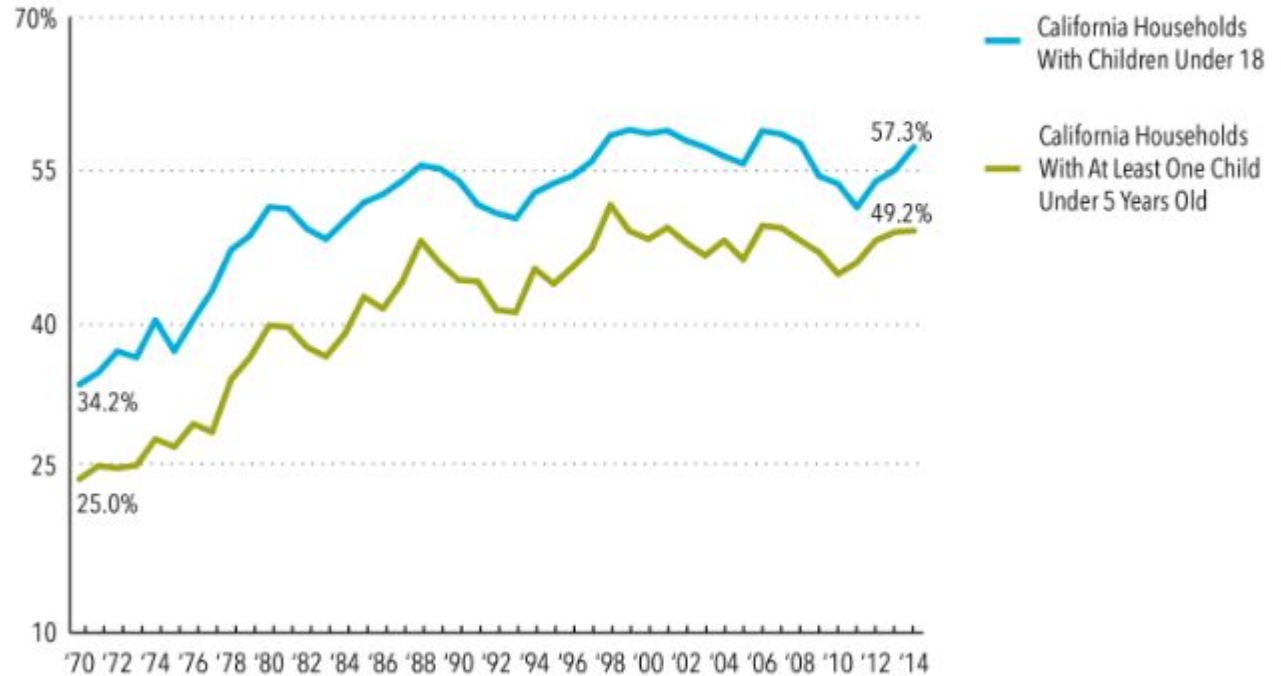
No necesariamente

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:

-Times se vende en más
del doble que el Daily
Telegraph

La gráfica de la derecha
hace parecer que el
periódico Times se vende más
de dos veces más que el
Daily Telegraph, ¿es esto
cierto?

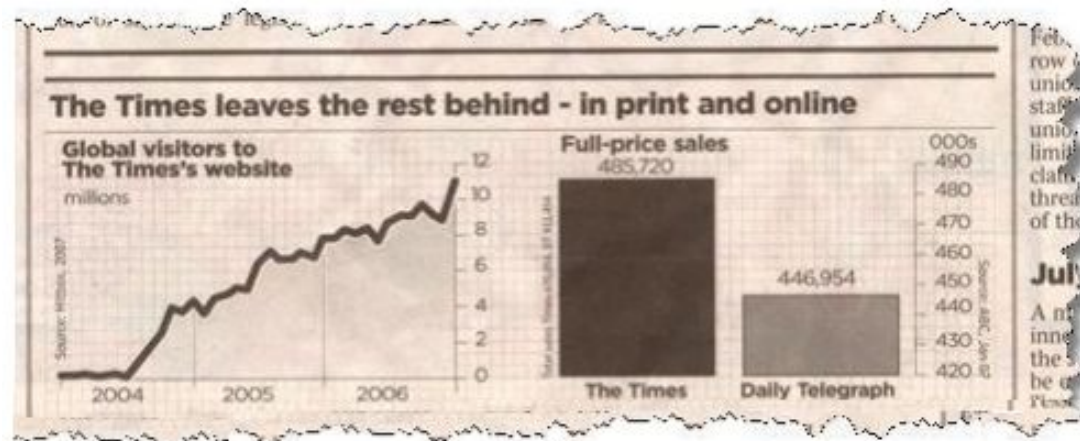


Image: University of Kentucky.

Ejemplo:

- La cantidad de gente
necesitando apoyo
- — —económico crece
fuera de control!

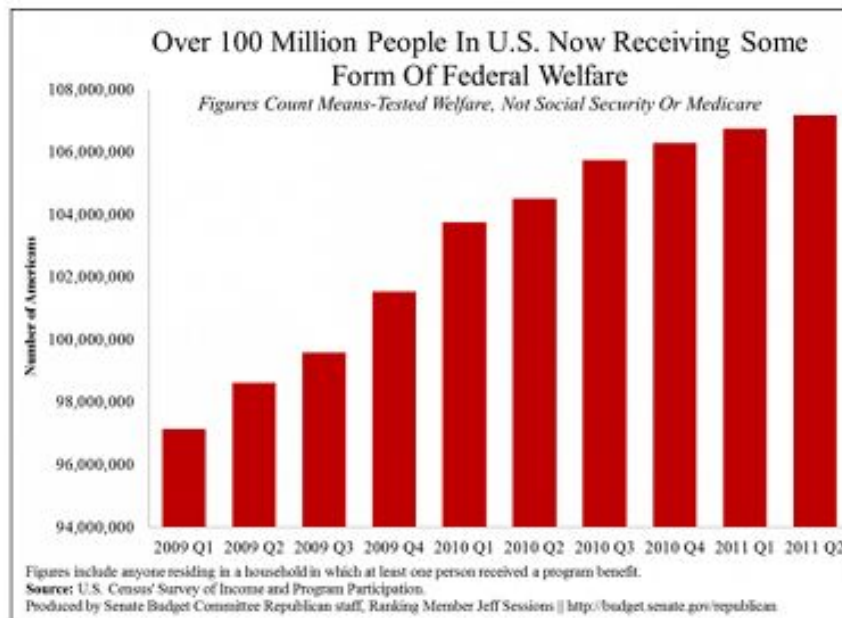
THE BLOG

Over 100 Million Now Receiving Federal Welfare

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



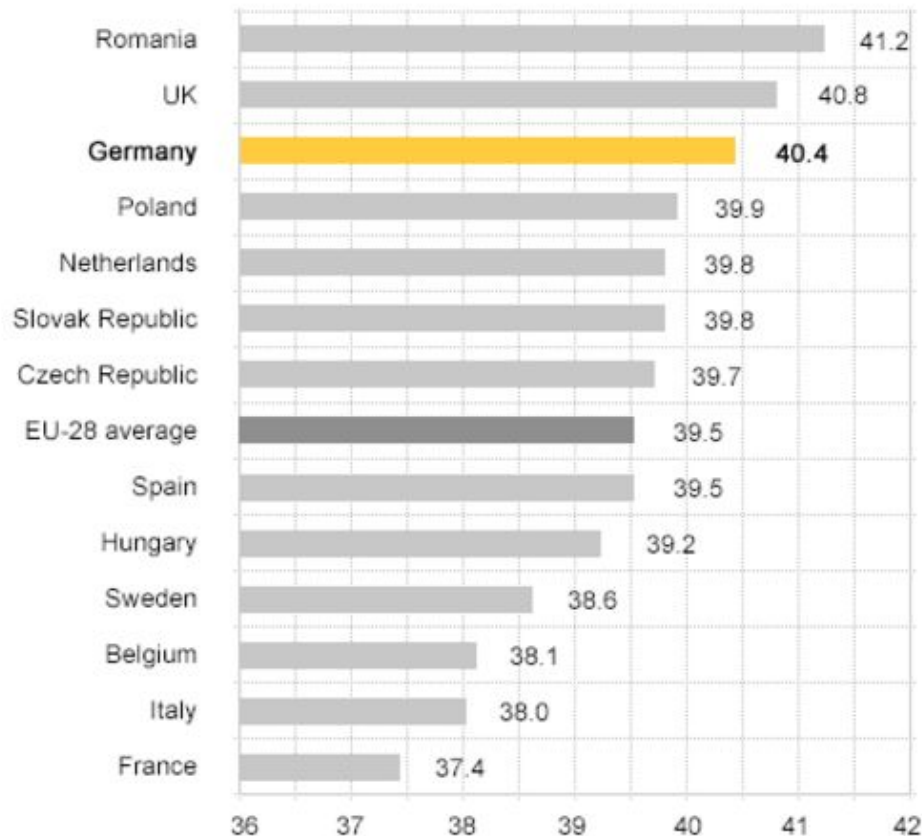
A new chart set to be released later today by the Republican side of the Senate Budget Committee details a startling statistic: "Over 100 Million People in U.S. Now Receiving Some Form Of Federal Welfare."



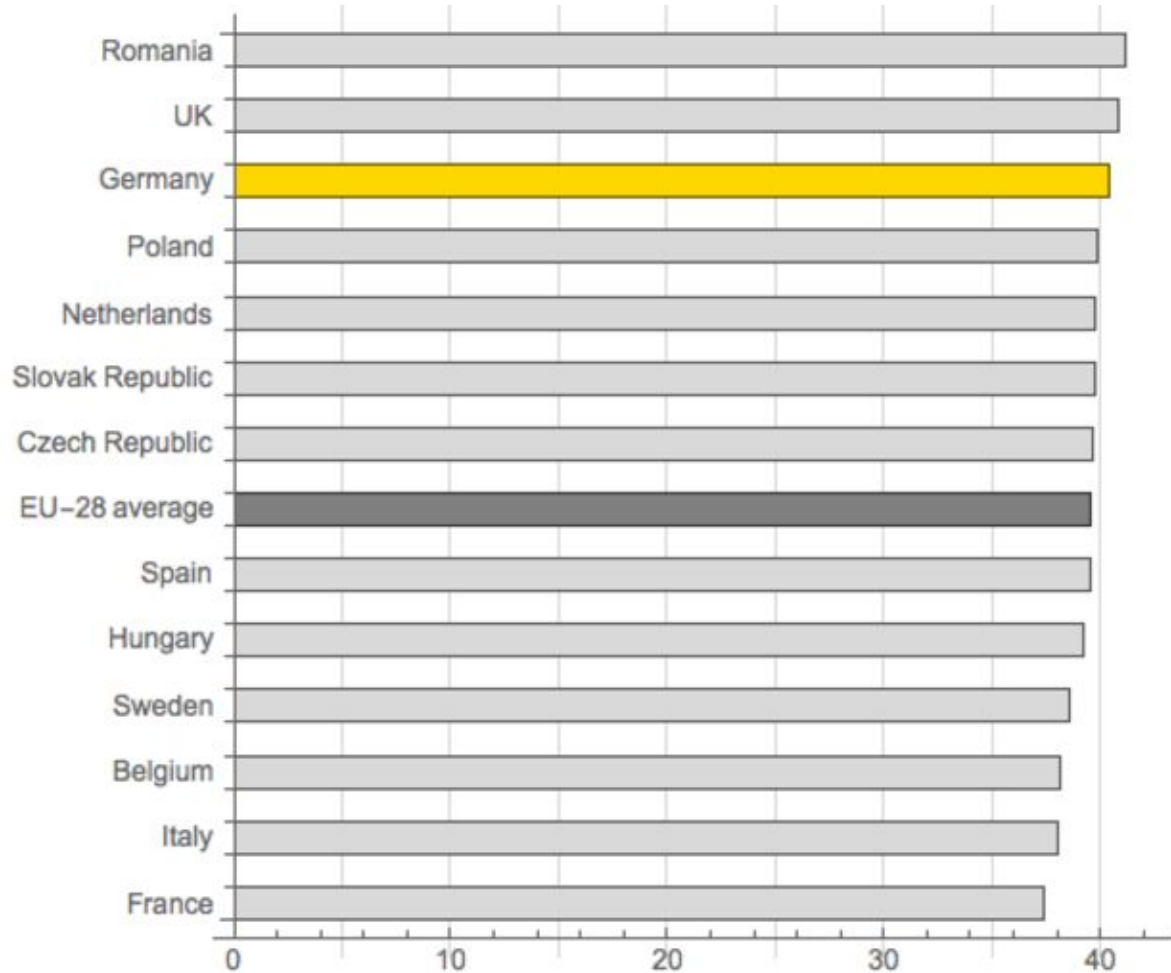
Ejemplo:

-En Rumania se trabaja
3 veces más que en
Francia a la semana

Average number of actual weekly hours of work in main job, full-time employees, 2013



— — —
¡Las conclusiones a las
que se llega están
influenciadas por la
forma en que se
presentan los datos!

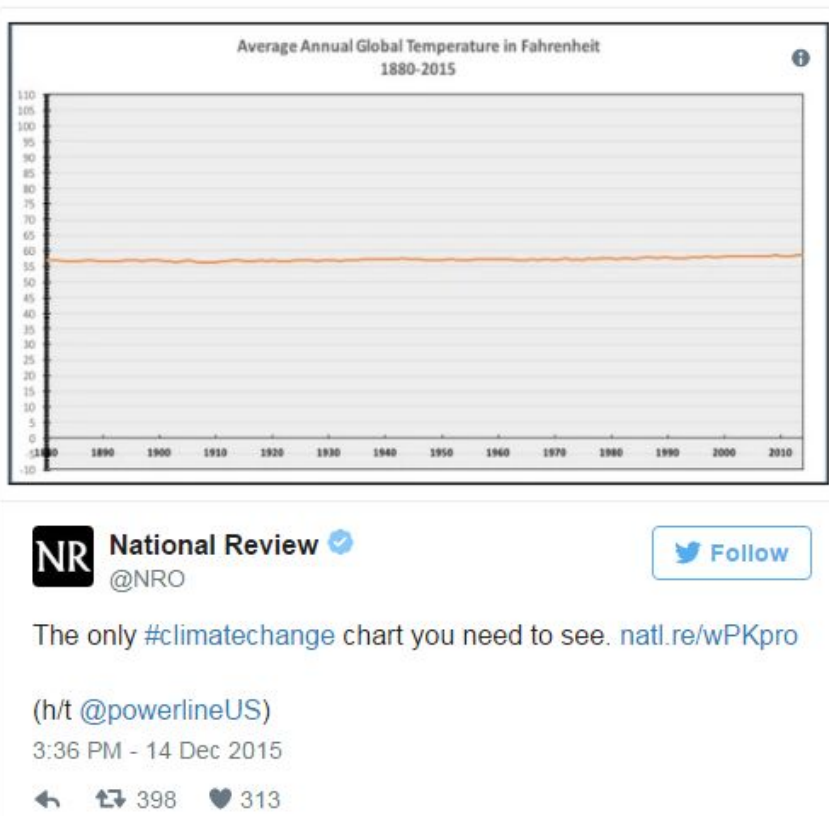


Ejemplo:

-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...

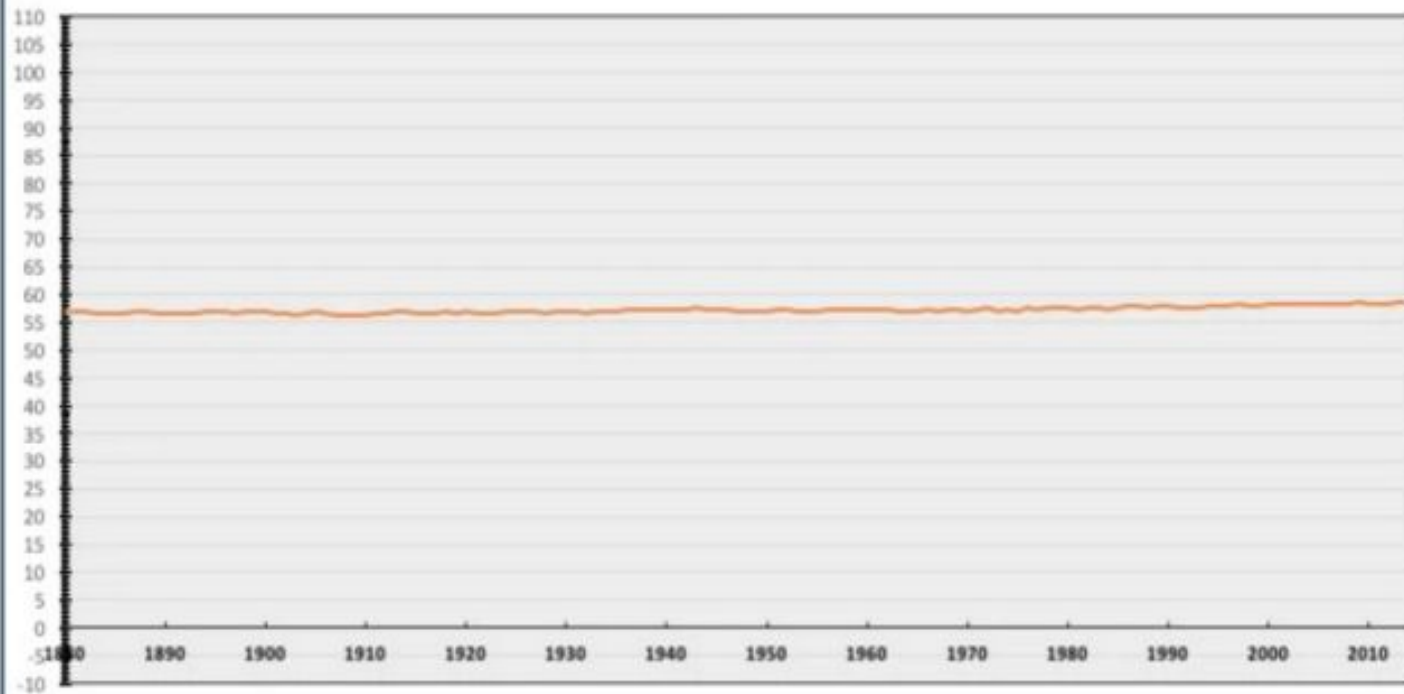
1.- Lo que interesa es observar los cambios en Y a través de los cambios en X.

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

- Gráficas Lineales

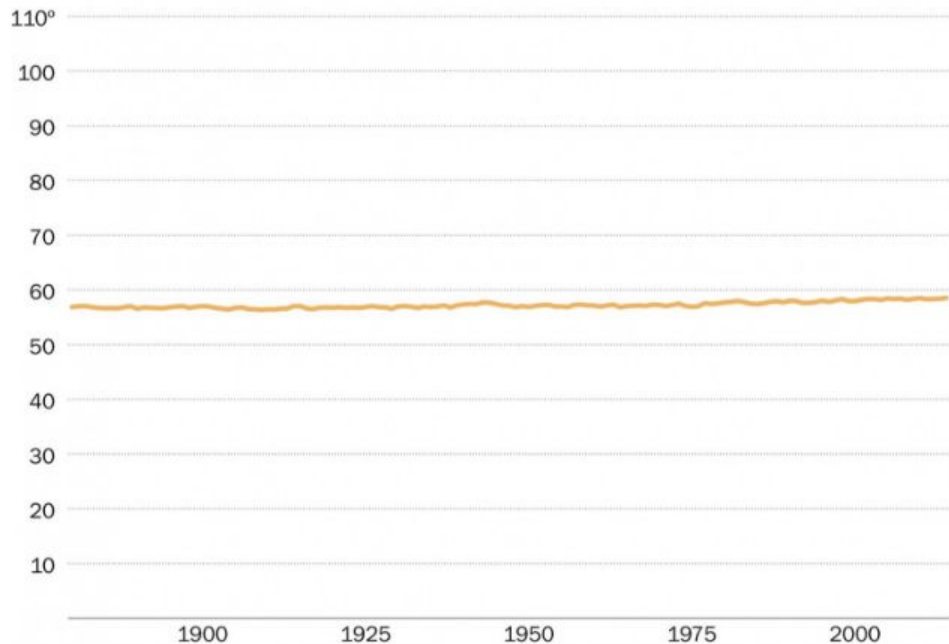
2) La Escala/Intervalo cubierto por los ejes

Average Annual Global Temperature in Fahrenheit
1880-2015



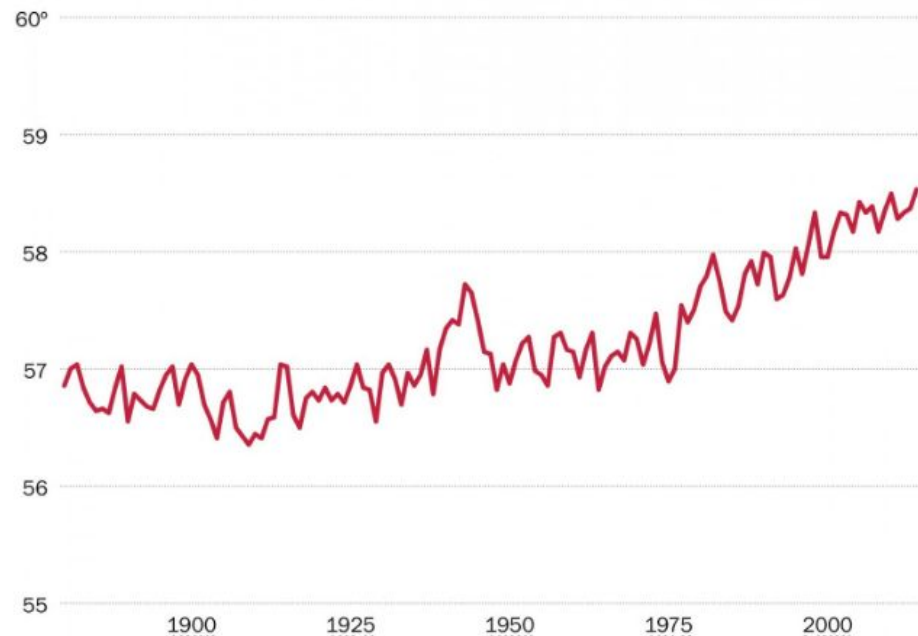
Average global temperature by year, 50x scale

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) Proporción en los ejes

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.

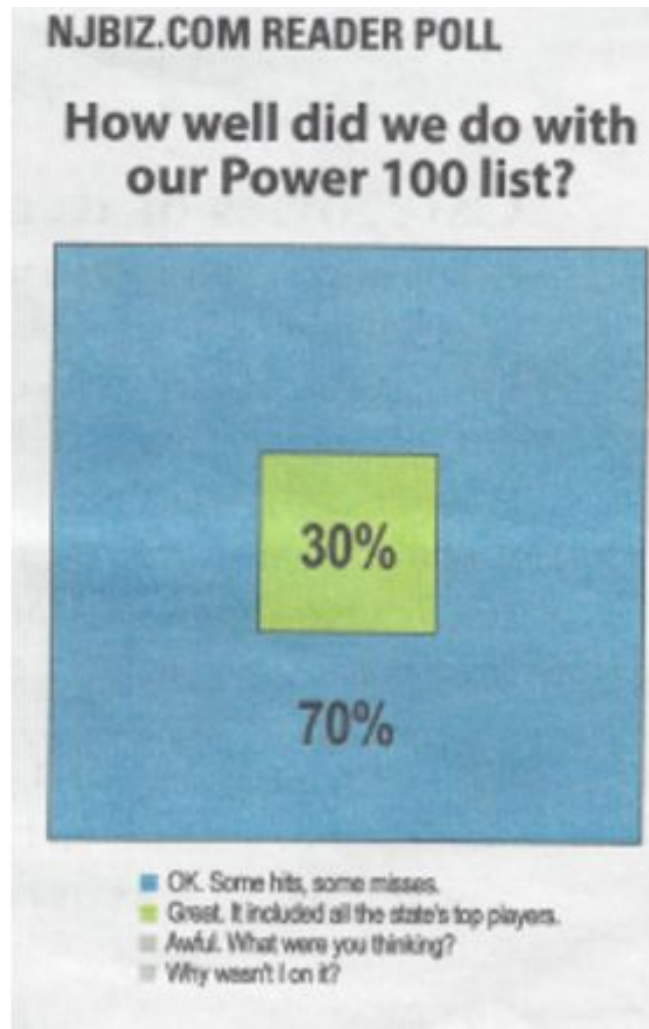
I came across the graph in Figure 1 while making my year end contributions. The source is a non-profit organization that provides invaluable services to our society, so I'd rather not name it here.



Ejemplo:

— — —

Resultados de una encuesta realizada a los lectores del periódico NJBIZ (Nueva Jersey) sobre un listado de las 100 personas más influyentes ('Power 100'), publicado por el mismo periódico.



4) Títulos y encabezados sugerentes

Ejemplo:

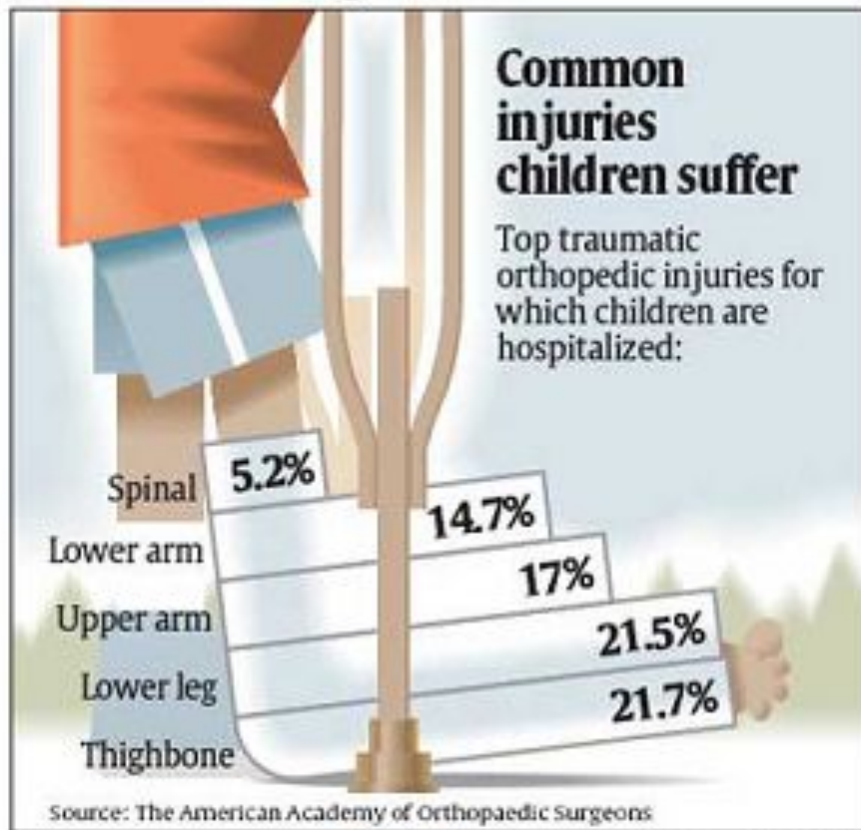
“¿El 5.2% de los niños
sufren lesiones de
columna?!”

— — —

¿El 5.2% de qué grupo? ¿De los niños en general?, ¿o se trata del 5.2% del subconjunto de niños que son hospitalizados?

El título invita a saltar a una conclusión que no coincide con la descripción de lo graficado.

USA TODAY Snapshots™



By Shannon Reilly and Frank Pompa, USA TODAY

Ejemplo:
“Trump y Bush son
claramente más
eficientes al reducir el
desempleo.”

¿Los primeros tres meses de
gestión son reflejo del
trabajo del nuevo
presidente, o de lo que
quedó del presidente
anterior?





Fox News
@FoxNews

Follow



Jobless rate after first 3 months: Trump vs. Obama vs. Bush vs. Clinton.



RETWEETS

564

LIKES

1,187



4:34 PM - 29 Apr 2017



Steven Spencer
@sspencer_smb

Follow

here u go @FoxNews. had a few minutes before dinner to help out. unemployment rates when each President left office. #NFP

4:53 PM - 29 Apr 2017

123 160

— — —
NOTA: No sabemos de dónde
sacó este usuario los datos...
pero el punto se entiende.

Ejemplo:

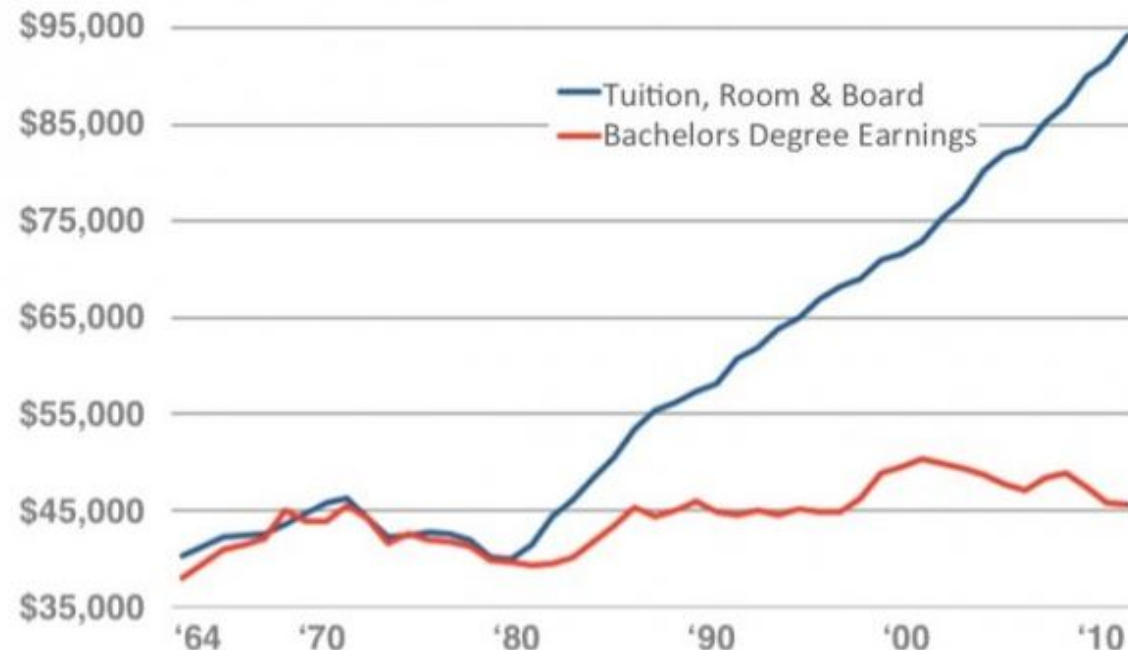
- "Hacer una carrera no es rentable"

— — —

Estudiar una carrera de 4 años cuesta alrededor de 95,000 dólares. Y en promedio, las personas con licenciatura ganan 45,000 dólares al año...¡AL AÑO!

The diminishing financial return of higher education

Costs of 4-yr degree vs. earnings of 4-yr degree



Source: Source; U.S. Census Data & NCES Table 345.

Notes: All figures have been adjusted to 2010 dollars using the Consumer Price Index from the BLS.

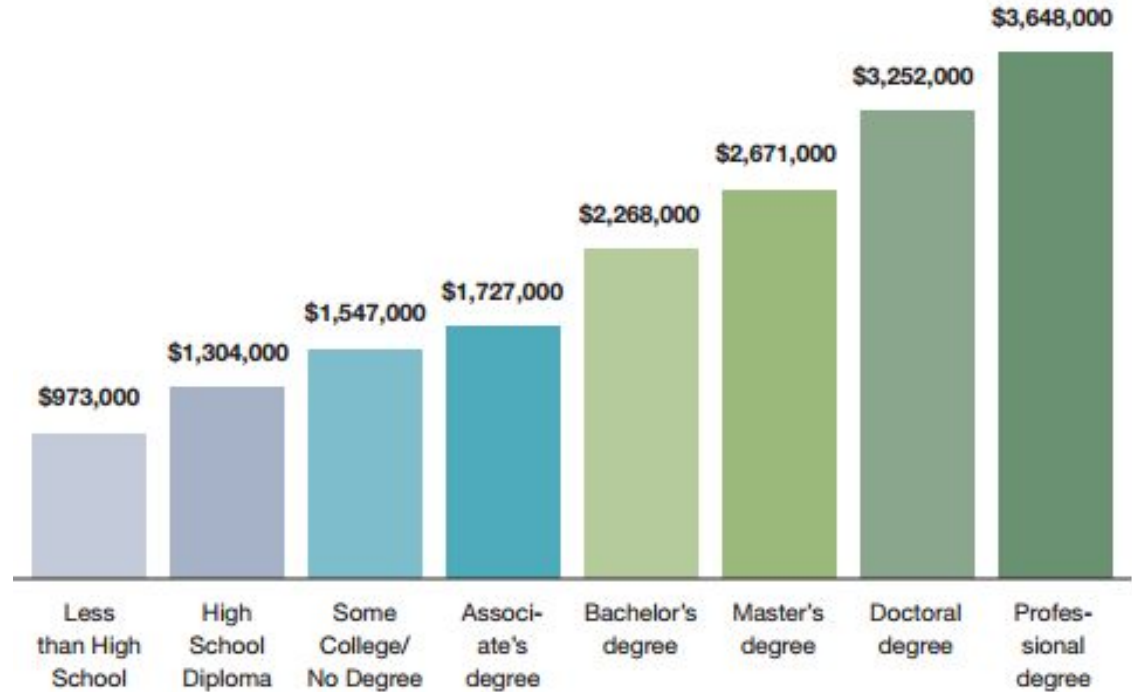
Business Insider - "This two charts prove that a College Education just isn't worth the money anymore"

(<http://www.businessinsider.com/these-two-charts-prove-a-college-education-just-isnt-worth-the-money-anymore-2012-6>)

— — —

Aún suponiendo que las personas con licenciatura hubieran invertido 100,000 dólares en su educación, a lo largo de su vida habrán acumulado una mayor ganancia.

No sólo recuperarán lo invertido, sino que de hecho habrán ganado más de lo que habrían podido ganar sin el título.



5) Gráficas mezcladas: Más de un eje Y

Ejemplo:

“Los problemas de salud crónicos en EU han incrementado y el uso de insecticidas y maíz transgénico también...
¿Coincidencia? ¡No lo creo!”

Journal of Organic Systems, 9(2), 2014

ORIGINAL PAPER

Genetically engineered crops, glyphosate and the deterioration of health in the United States of America

Nancy L. Swanson¹, Andre Leu^{2*}, Jon Abrahamson³ and Bradley Wallet⁴

¹ Abacus Enterprises, Lummi Island, WA, USA

² International Federation of Organic Agricultural Movements, Bonn, Germany

³ Abacus Enterprises, Lummi Island, WA, USA

⁴ Crustal Imaging Facility, Conoco Phillips School of Geology and Geophysics, University of Oklahoma, USA

* Corresponding author: andreleu.al@gmail.com

Abstract

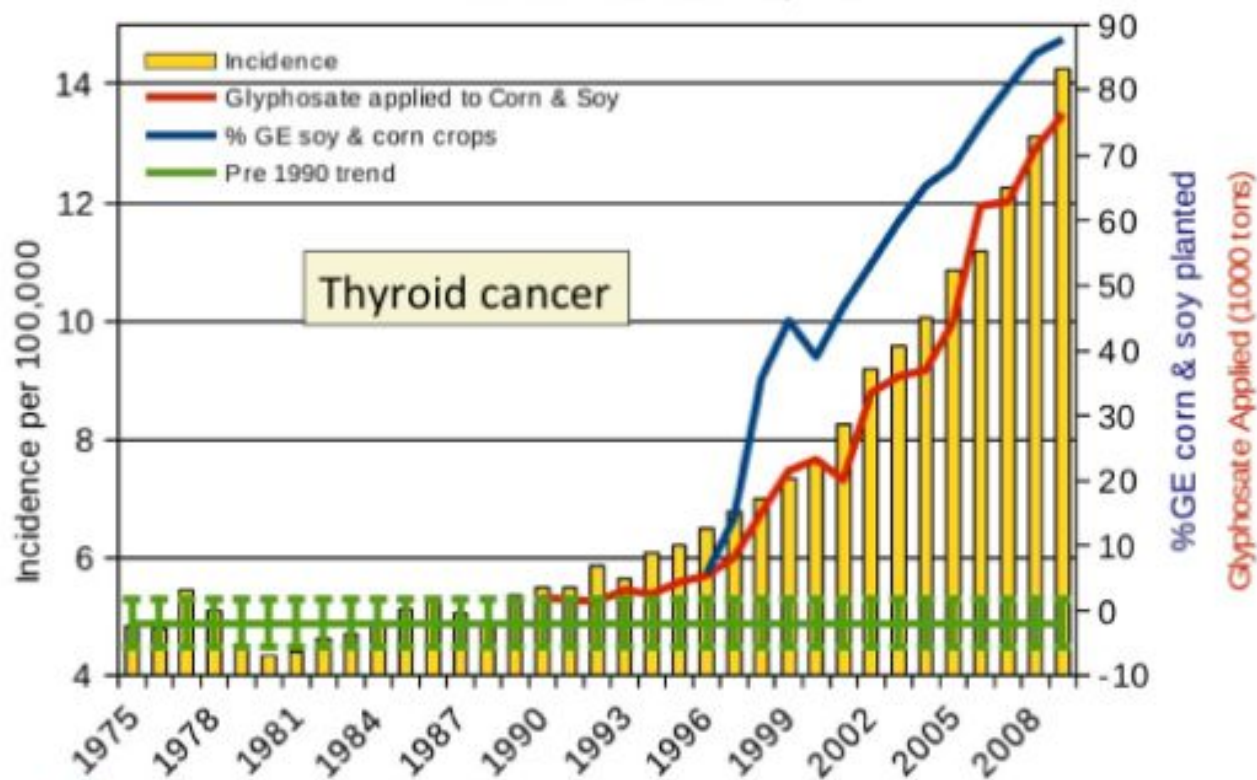
A huge increase in the incidence and prevalence of chronic diseases has been reported in the United States (US) over the last 20 years. Similar increases have been seen globally. The herbicide glyphosate was introduced in 1974 and its use is accelerating with the advent of herbicide-tolerant genetically engineered (GE) crops. Evidence is mounting that glyphosate interferes with many metabolic processes in plants and animals and glyphosate residues have been detected in both. Glyphosate disrupts the endocrine system and the balance of gut bacteria, it damages DNA and is a driver of mutations that lead to cancer.

Thyroid Cancer Incidence Rate (age adjusted)

plotted against glyphosate applied to U.S. corn & soy ($R = 0.988$, $p \leq 7.612e-09$)

along with %GE corn & soy crops $R = 0.9377$, $p \leq 2.152e-05$

sources: USDA:NASS; SEER



* Figure 10, Swanson et al. Journal of Organic Systems 2014; 9(2):6-37.

— — — “If the disease data were linearly increasing prior to the 1990s, a linear trend line was overlaid on the plot in green(...)

(...) In some cases, the axes have been adjusted to better illustrate the correlation; otherwise the data are plotted as is.”

“In all cases, the left vertical axis is the prevalence or the rate of incidence or death from the disease. The right vertical axis is both the percentage of GE corn and soy planted and the amount (in 1,000 tons) of glyphosate applied to the corn and soy crops.”

Liver and Intrahepatic Bile Duct Cancer Incidence (age adjusted)

plotted against glyphosate applied to corn & soy ($R = 0.9596$, $p \leq 4.624e-08$)
along with %GE corn & soy planted in U.S. ($R = 0.9107$, $p \leq 5.402e-05$)
sources: USDA:NAS; SEER

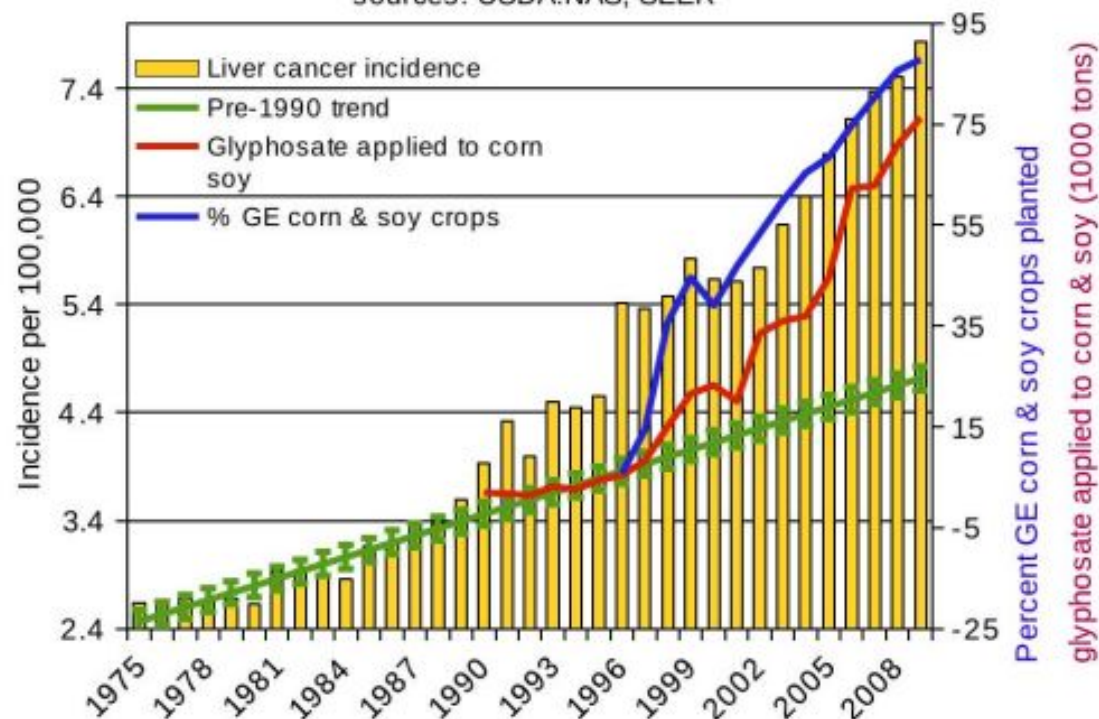
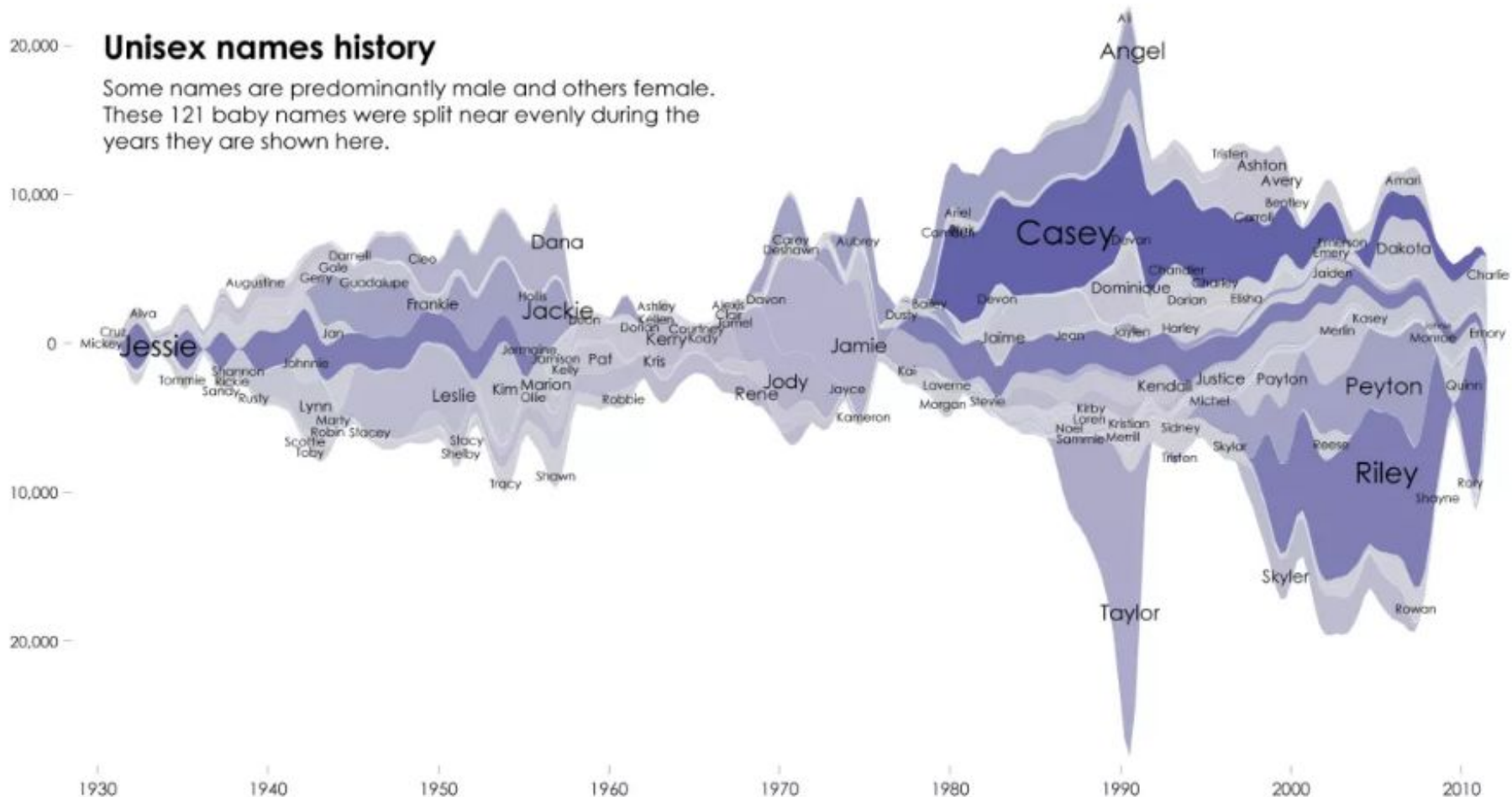


Figure 7. Correlation between age-adjusted liver cancer incidence and glyphosate applications and percentage of US corn and soy crops that are GE.

6) La información cuando es mucha, se pierde.

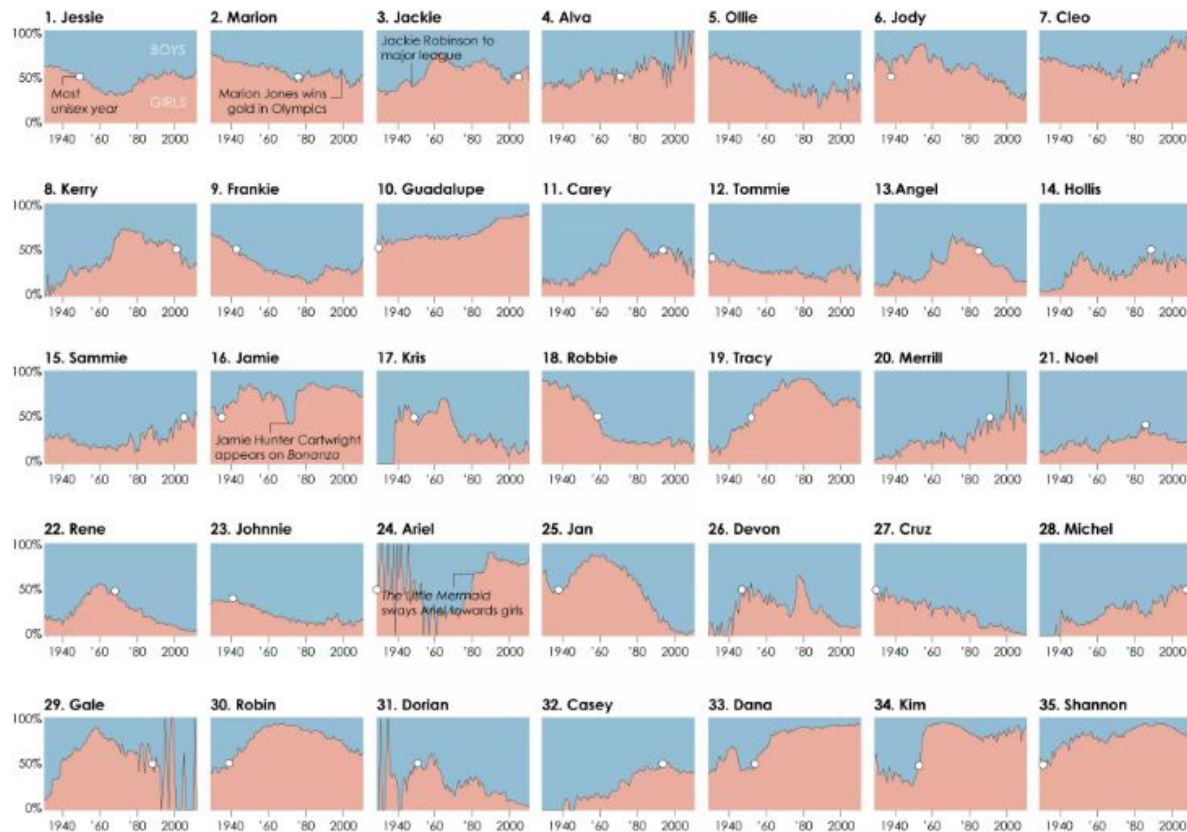
Unisex names history

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



The Most Unisex Names in US History

BY NATHAN YAU / POSTED TO DATA UNDERLOAD / TAGS: NAMES



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

7) OJO: Nunca subestimes el Maquillaje de datos'

Ejemplo:

-Obama y el
desempleo-

— — —

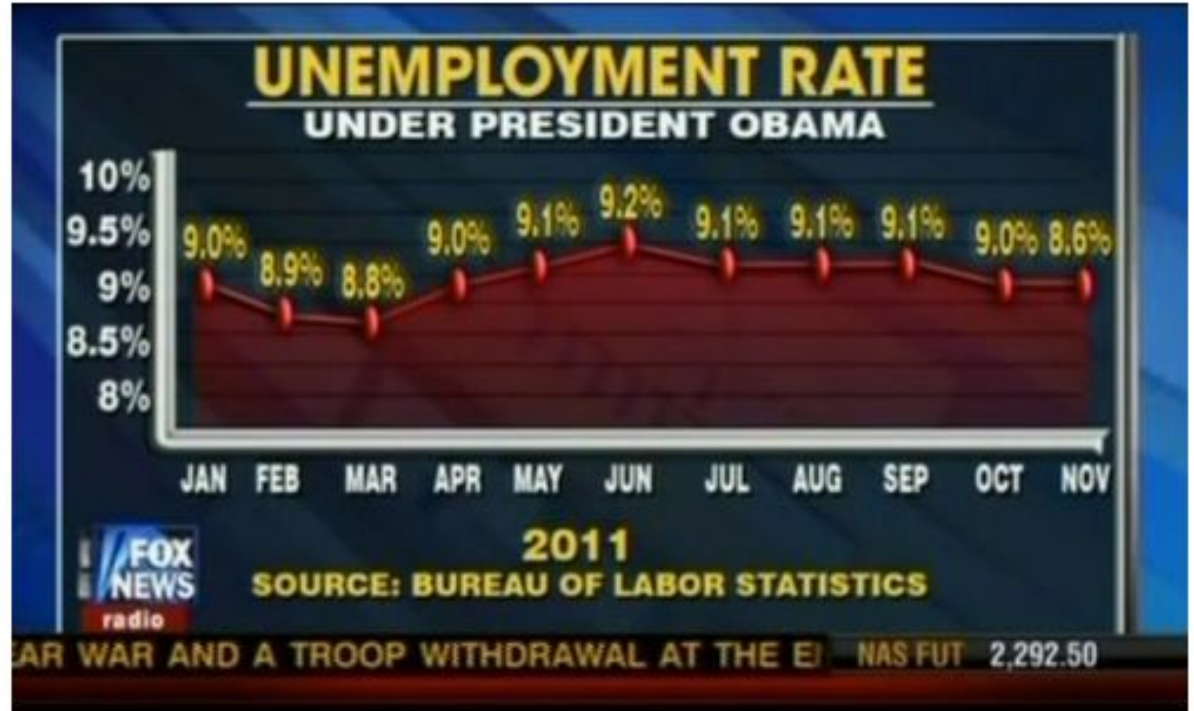


Image Source: <http://cloudfront.mediamatters.org>

The November rate is lower than the March rate of 8.8 percent, but it's shown to be higher in the Fox News chart. Here's what the graph should look like, according to data from the Bureau of Labor Statistics:

Unemployment Rate in 2011



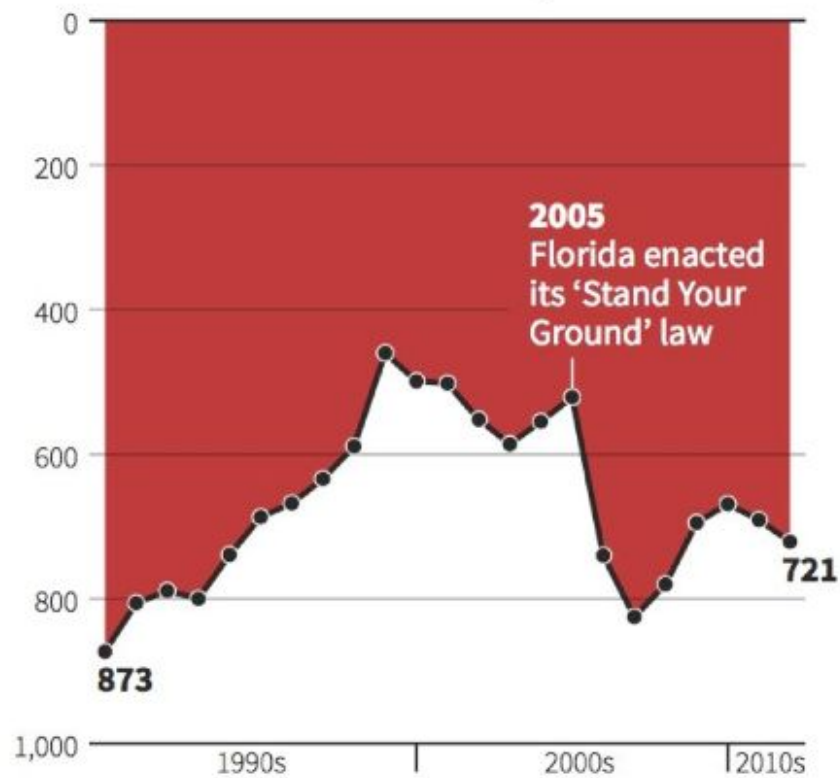
Source: Bureau of Labor Statistics

Ejemplo:

Las muertes por arma de fuego
en Florida han decrecido... ¡un
momento!

Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

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...

— — —

Fuentes con ejemplos (¡Consúltelo usted mismo!)

— — —

- <http://www.statisticshowto.com/misleading-graphs/>
- [http://callingbullshit.org/tools/tools misleading axes.html](http://callingbullshit.org/tools/tools_misleading_axes.html)
- https://www.washingtonpost.com/news/the-fix/wp/2015/12/14/why-the-national-reviews-global-temperature-graph-is-so-misleading/?utm_term=.5bf1d6bbd082
-

¡Hágalo usted mismo! (Enlaces con ejercicios)

— — —

<http://faculty.atu.edu/mfinan/2043/section31.pdf>