Misleading Charts

Data Visualization and Exploration

Ozan Kahramanoğulları

by Matteo Ceccarello based on material from How charts lie – Alberto Cairo

Misleading charts

Collected by Matteo Ceccarello based on material from How charts lie – Alberto Cairo

How Charts Lie

Getting Smarter about Visual Information

Alberto Cairo

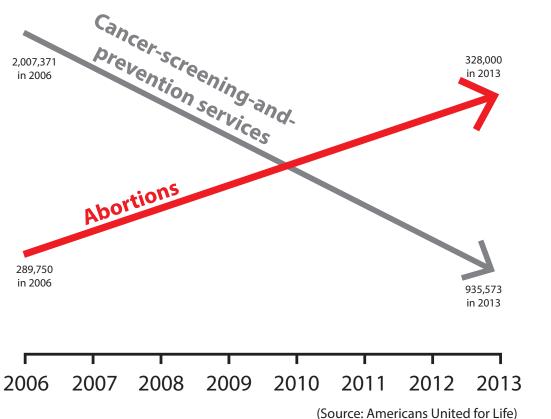


5 Ways to lie with charts

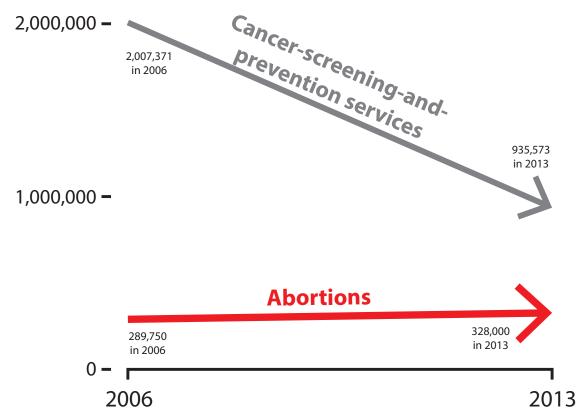
- Bad design choices
- Using dubious data
- Showing insufficient data
- Concealing uncertainty
- Suggesting misleading patterns

Poor design

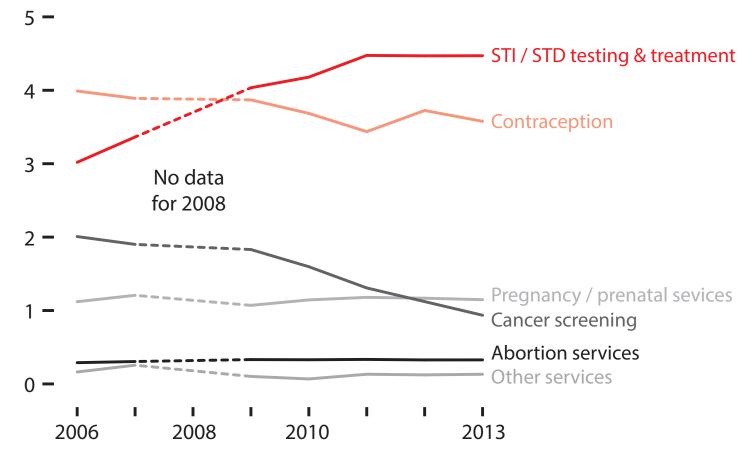
Planned Parenthood Federation of America: Abortions up—life-saving procedures down



Planned Parenthood Federation of America



Poor design, mixing different scales



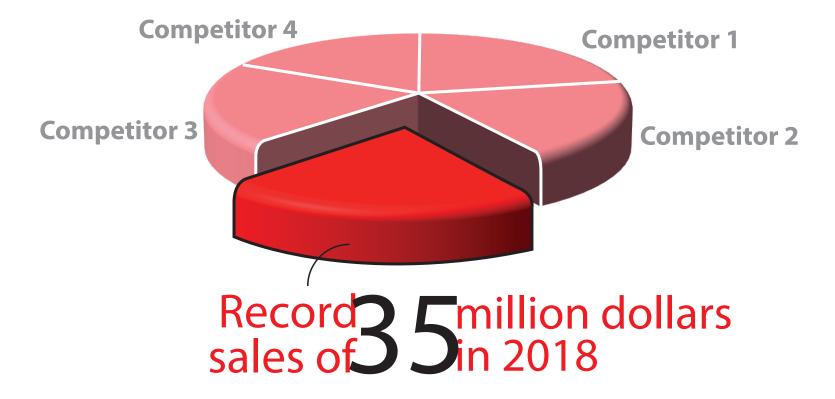
Services provided by Planned Parenthood (millions)

(Source: Emily Schuch and Planned Parenthood Federation of America)

Even better: provide all the context

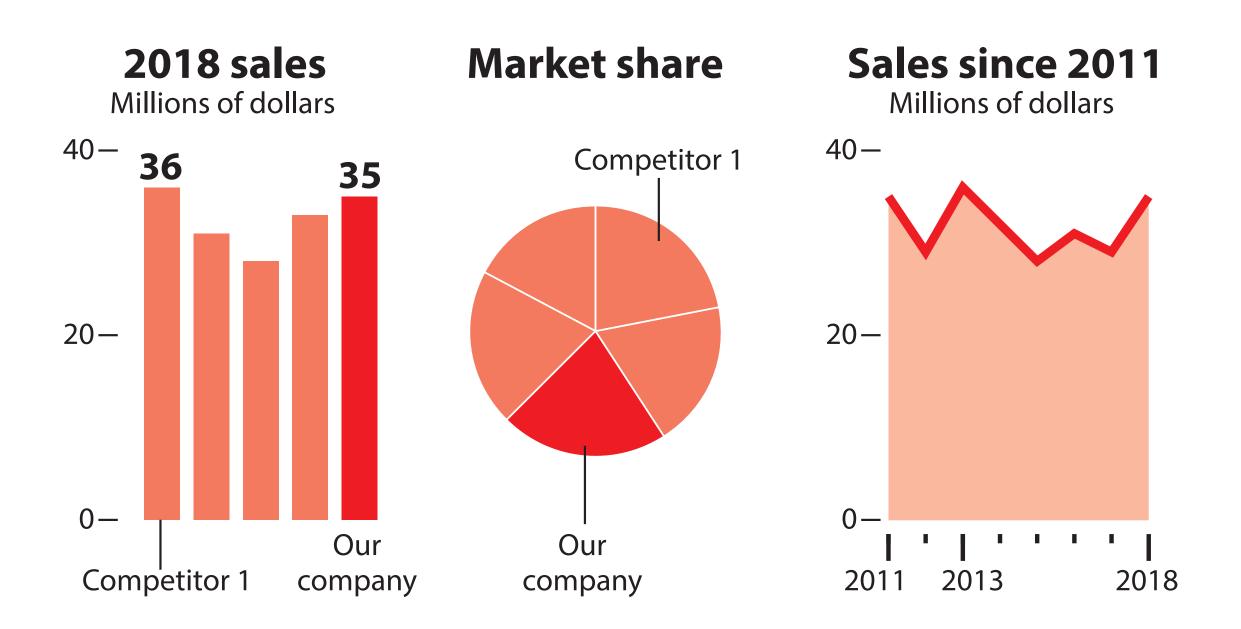


Share of All Companies in Our Market



Our Company's Sales Have Soared!



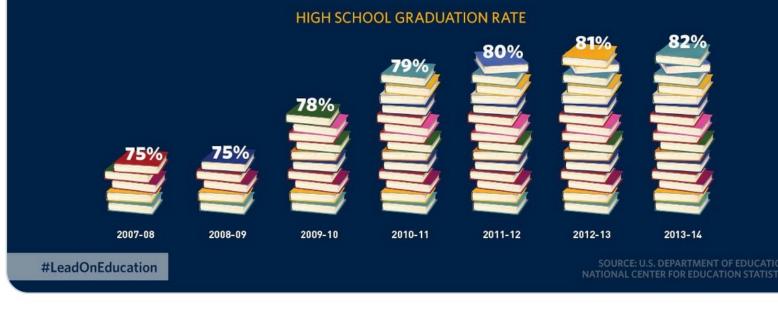




Good news: America's high school graduation rate has increased to an all-time high. The wapo.st/1m40Mei

Traduci il Tweet

UNDER PRESIDENT OBAMA, MORE STUDENTS ARE EARNING THEIR HIGH SCHOOL DIPLOMAS THAN EVER BEFORE

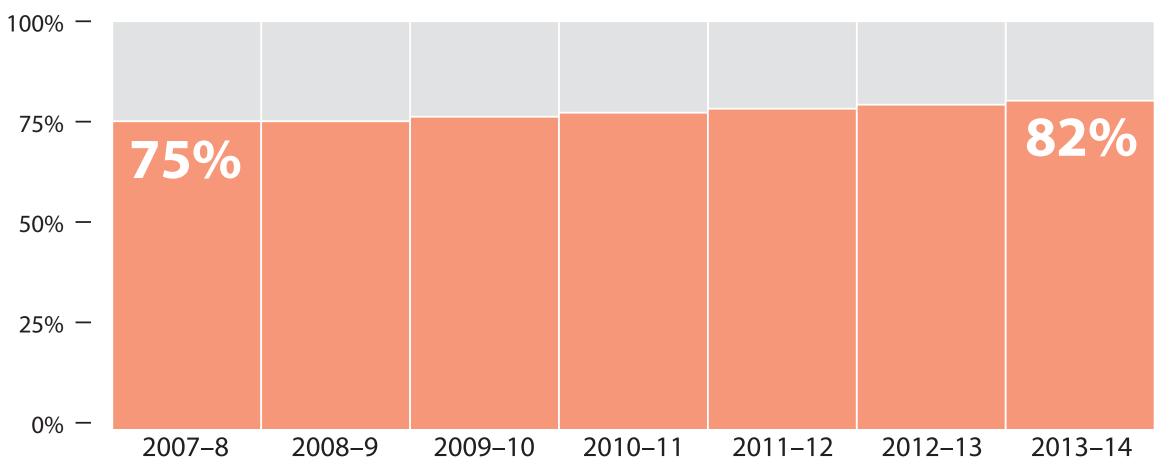


Poor design: non-zero baseline!

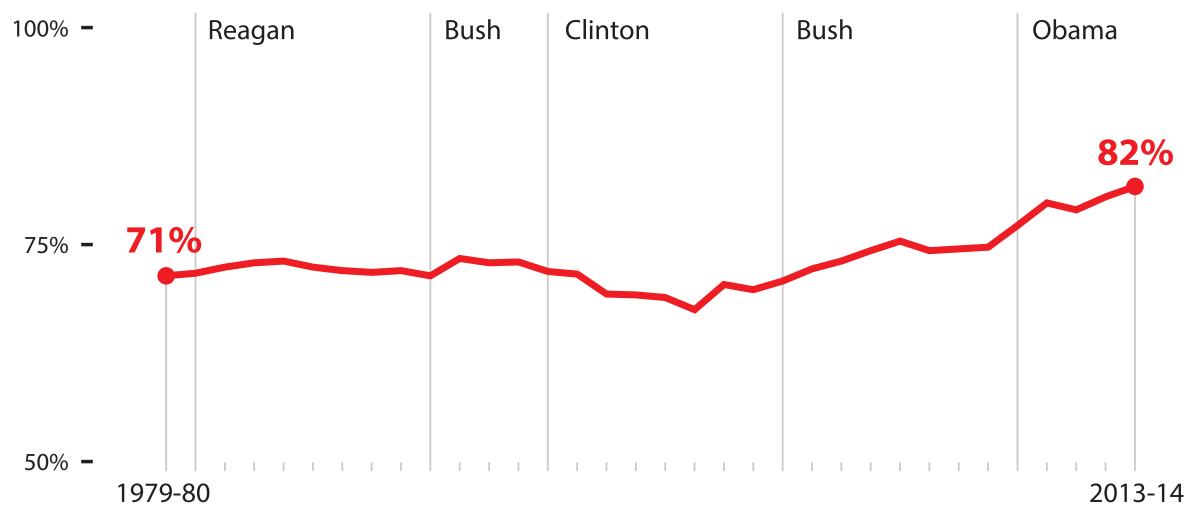
7:11 PM \cdot 16 dic 2015 \cdot TweetDeck

Under President Obama, more students are earning their high school diplomas than ever before

High school graduation rate

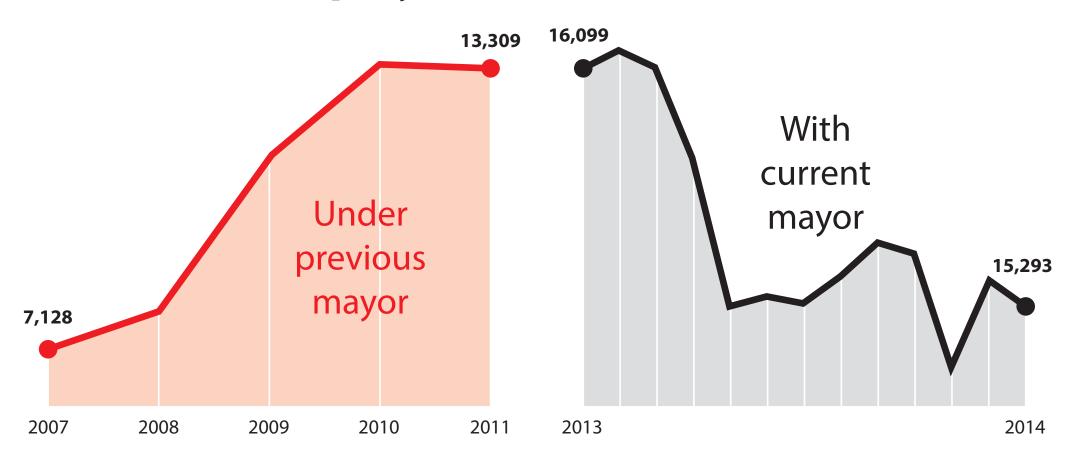


High school graduation rates under each president



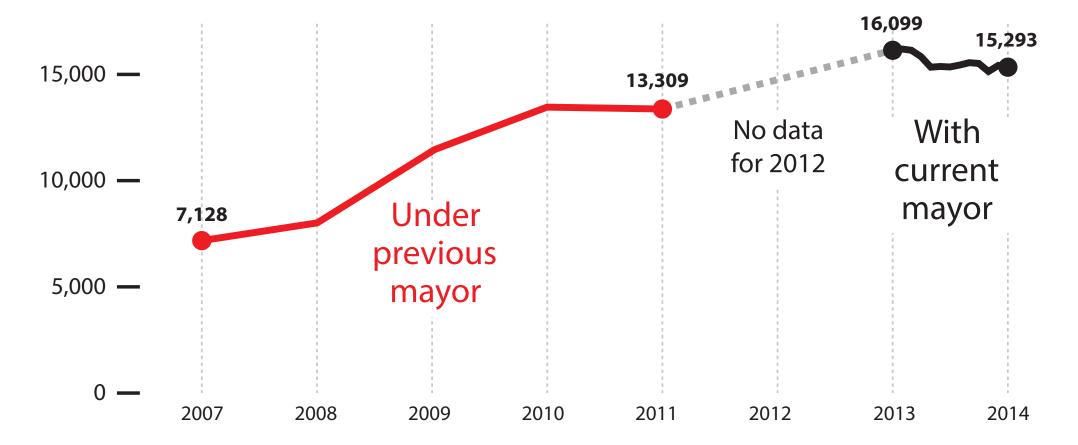
(Source: National Center for Education Statistics)

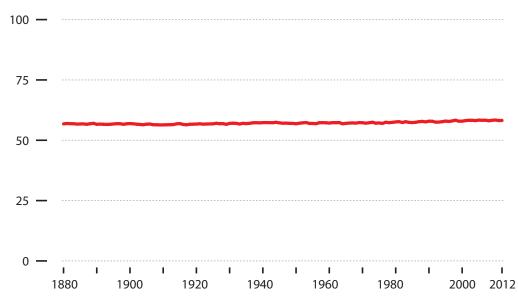
Unemployed adults in Alcorcón



Bad design: different x and y axes

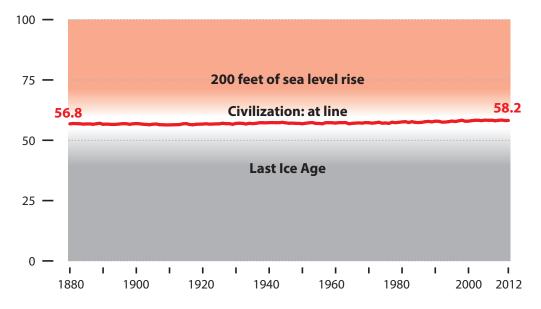
Unemployed adults in Alcorcón



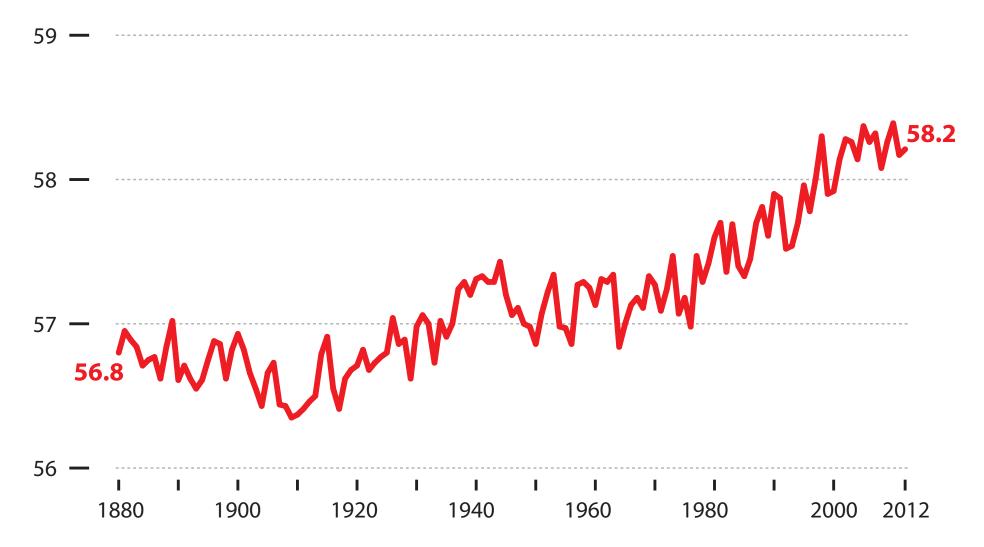


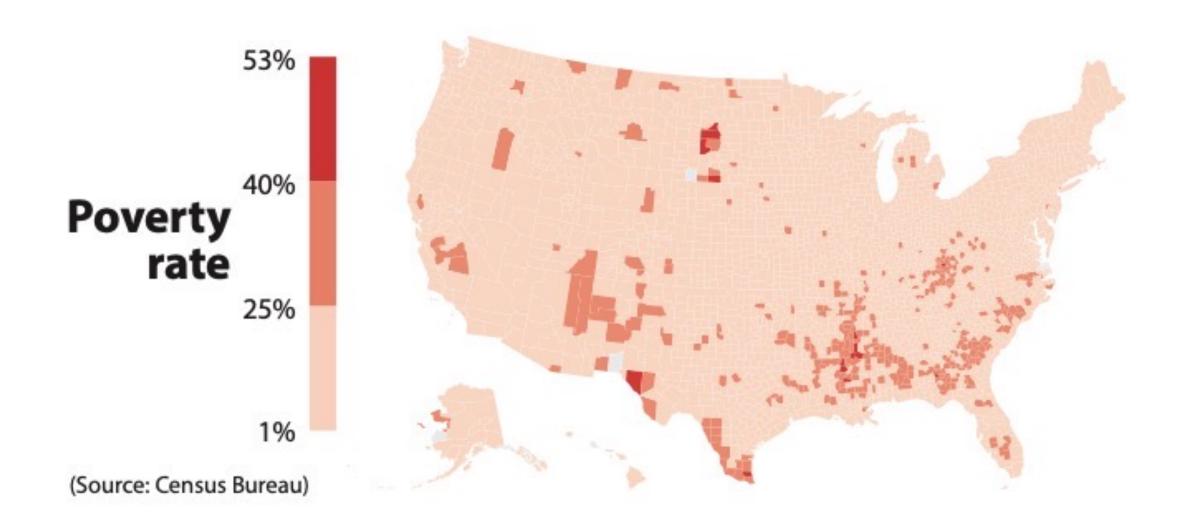
Average annual global temperature in degrees Fahrenheit

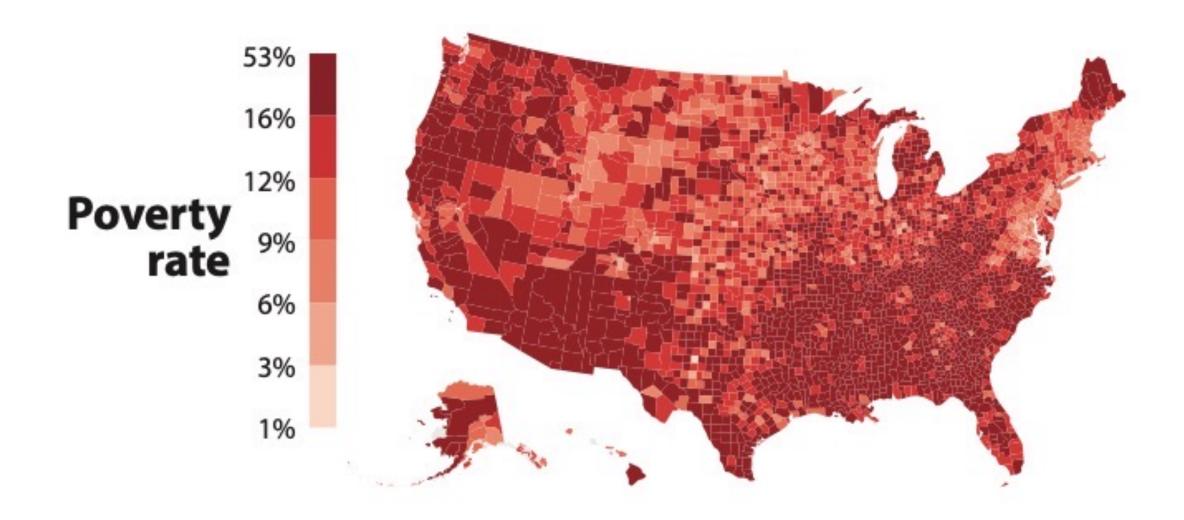
Average annual global temperature in degrees Fahrenheit



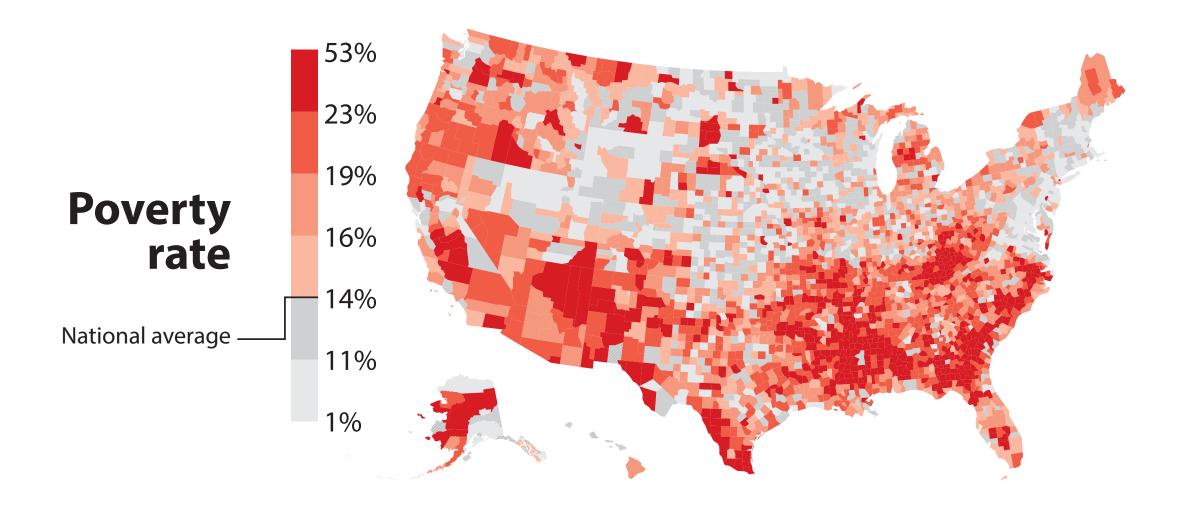
Average annual global temperature in degrees Fahrenheit







Bad design: distorted color scale with uneven bins



Watch out!

- Multiple scales for the same aesthetic
- Uneven scales (including colour)
- Misplaced baselines
- Wrong ranges

Showing dubious data

What and how is being counted



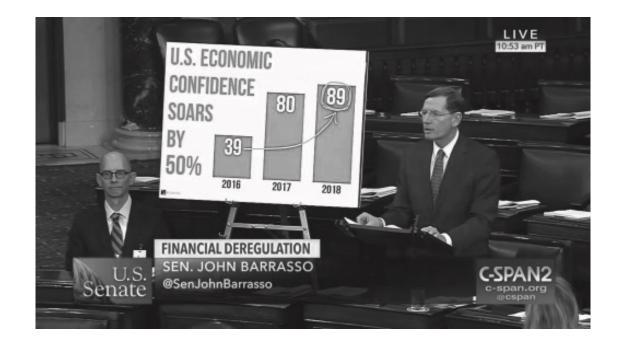
U.S. ECONOMIC CONFIDENCE 80 SOARS BY 39 50% 2016 2017 2018

FINANCIAL DEREGULATION SEN. JOHN BARRASSO @SenJohnBarrasso

U.S.

Senate

c-span.org



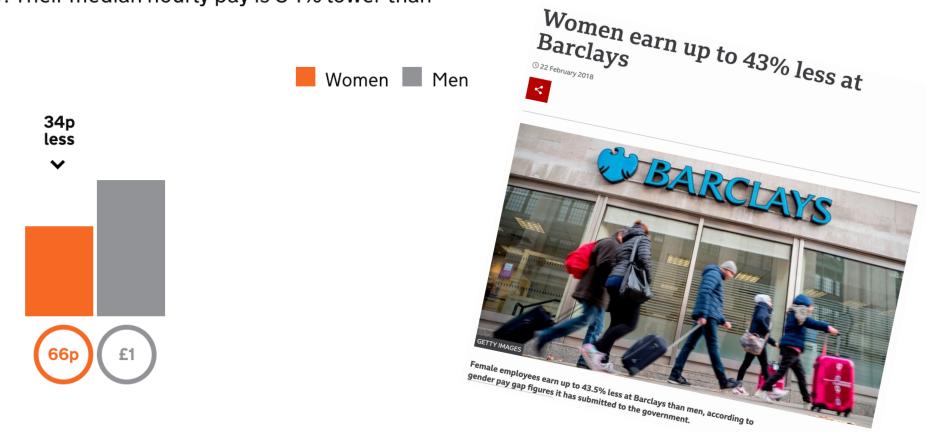
Confusion between
percentage change
$$\left(\frac{89-39}{39} \cdot 100 = 128\%\right)$$

and
percentage point change $(89 - 39 = 50\%)$

Hourly pay gap

About median and mean

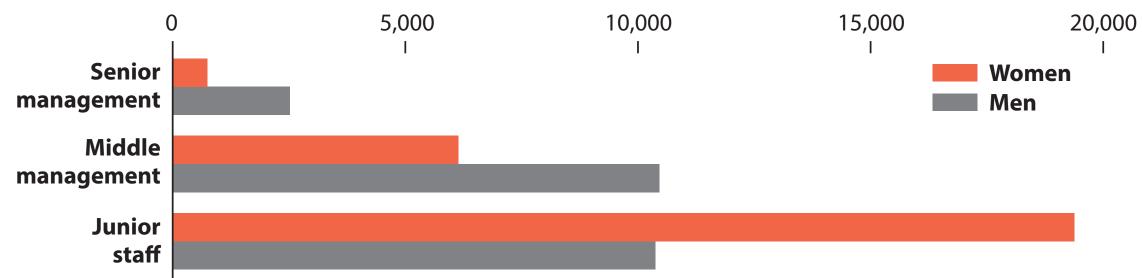
In this organisation, women earn 66p for every £1 that men earn when comparing median hourly pay. Their median hourly pay is 34% lower than men's.



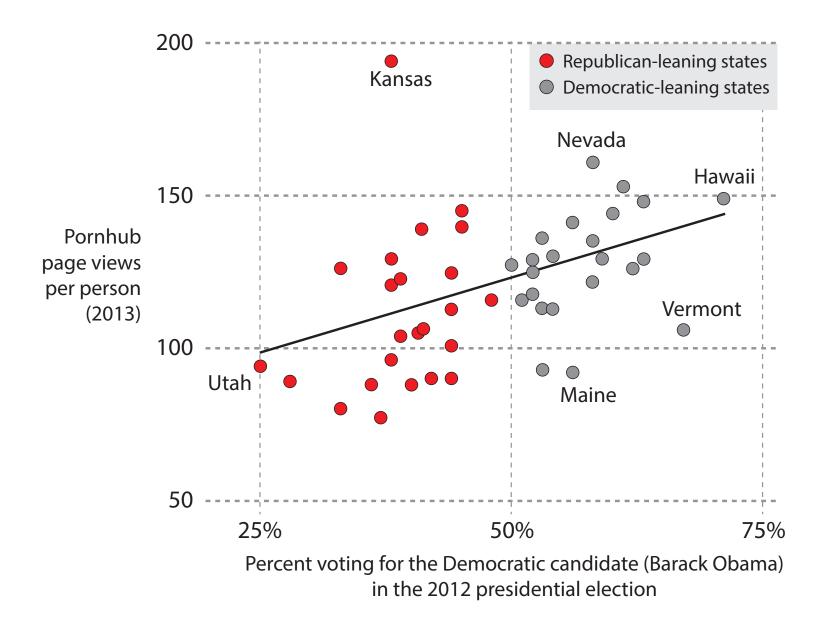
When comparing mean (average) hourly pay, women's mean hourly pay is 42.1% lower than men's.

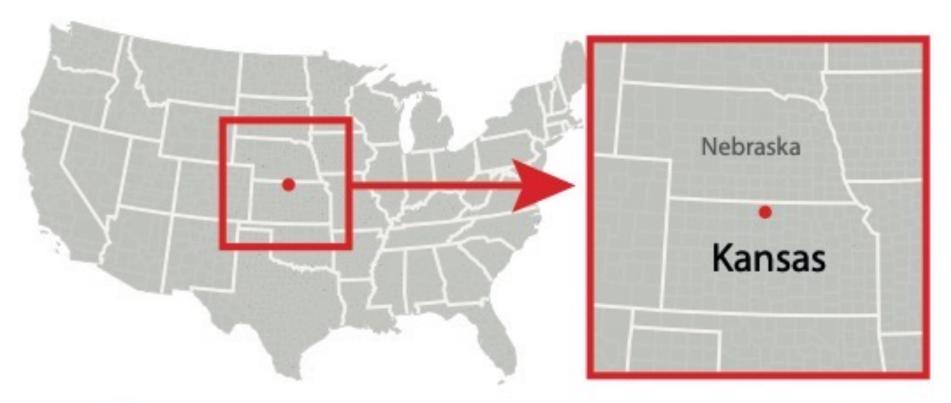
https://gender-pay-gap.service.gov.uk/Employer/MxCVjH2H/2020

Barclays Bank employees in the UK



There is a gender pay-gap problem, but it's due to the fact that managerial roles are mostly occupied by men!





Geographic center of the contiguous United States

The data collection process might distort the data!

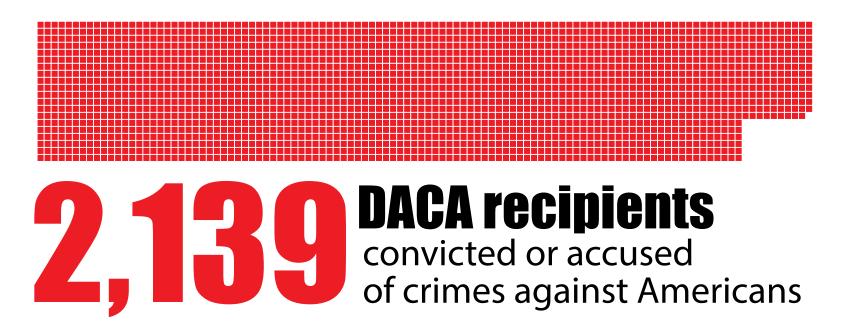
Watch out!

- Misinterpretations of basic concepts
- Issues approached from a single point of view
- Problems in the data collection process

Displaying insufficient data

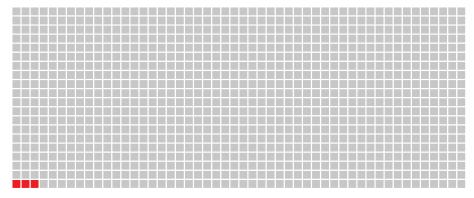
An easy way to distort reality

Graphic after Breitbart news:



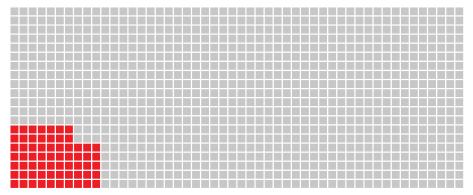
DACA stands for Deferred Action for Childhood Arrivals, and is meant to protect from deportation people who entered the USA illegally when they were children

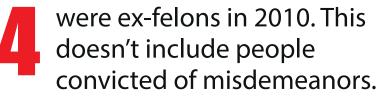
Out of every 1,000 DACA recipients ...



have had their temporary protected status revoked due to "a felony criminal conviction; a significant misdemeanor conviction; multiple misdemeanor convictions; gang affiliation; or arrest for any crime in which there is deemed to be a public safety concern."

Out of every 1,000 voting-age people who live in the U.S....

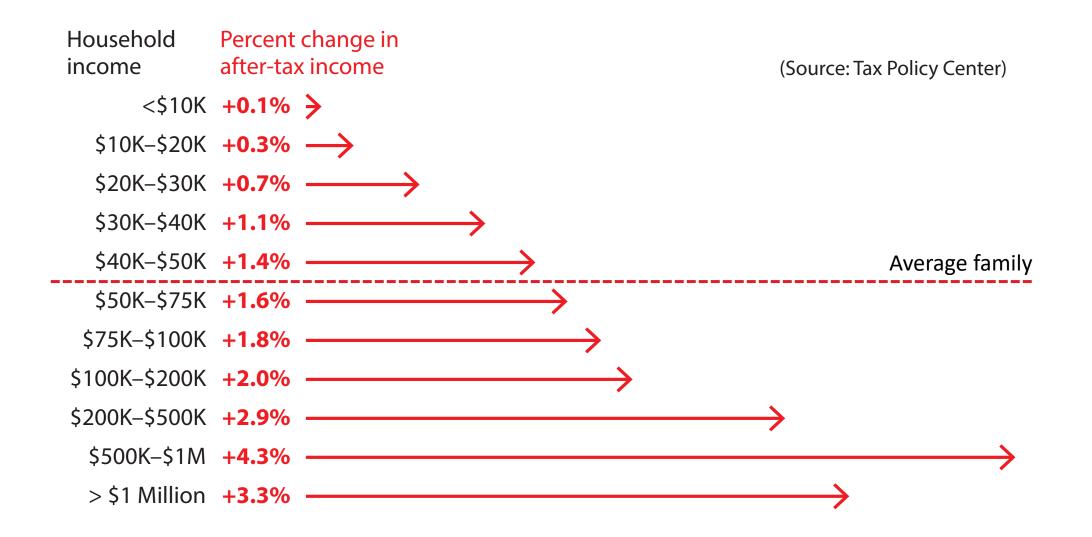




(Source: Shannon, Sarah K.S., et. al. "The growth, scope, and spatial distribution of people with felony records in the United States, 1948 to 2010." Demography 54(5)(2017): 1795-1818)



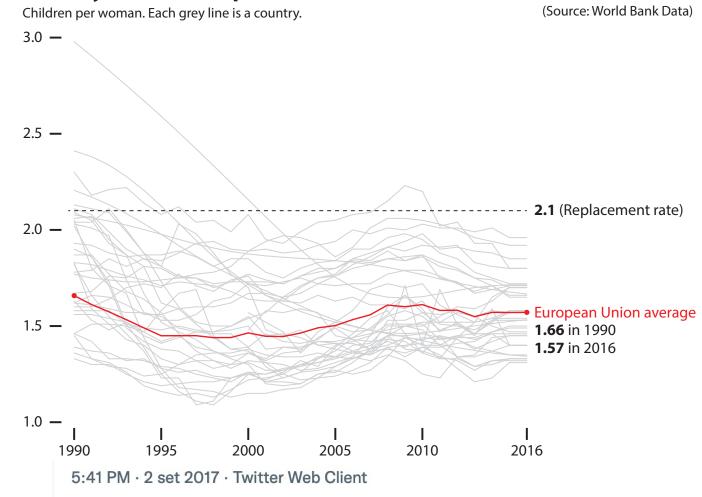






Capitalism+atheism+feminism = sterility = migration. EU birthrate = 1.6. Replacement = 2.1. Merkel, May, Macron, Gentiloni all childless.

Fertility rate in European countries



...

Fertility rate in European countries, 1990–2016

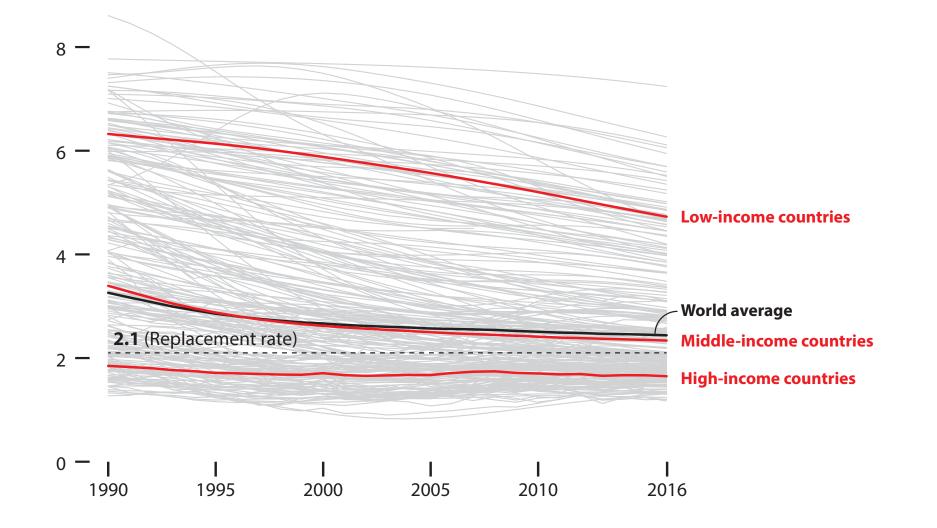
Compared to the replacement rate: 2.1 children per woman Note: not all these countries belong to the European Union

2.1 European Union 2.1	Albania	Austria	Belgium	Bulgaria		Cyprus
0.0 1990 2016						
Czech Rep.	Denmark	Estonia	Finland	France	Germany	Greece
$\overline{}$		\frown	~~~~		· · · · · · · · · · · · · · · · · · ·	~~~
Hungary	Iceland	Ireland	Italy	Latvia	Lithuania	Luxembourg
	Sand Carlos			\sim		~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
Macedonia	Malta	Montenegro	Netherlands	Norway	Poland	Portugal
	~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~			~~~~~		~~~~~
Romania	Slovakia	Slovenia	Spain	Sweden	Switzerland	U. Kingdom
<u> </u>			~~~~		~	~~~~

#### Fertility rate

Children per woman. Each grey line is a country.

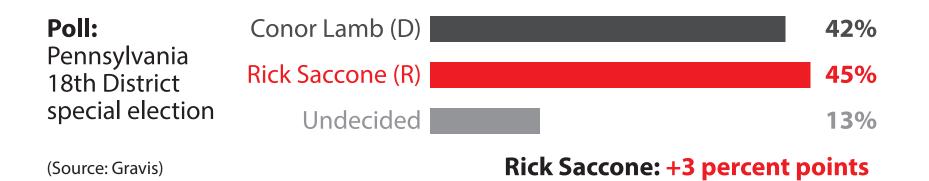
(Source: World Bank Data)



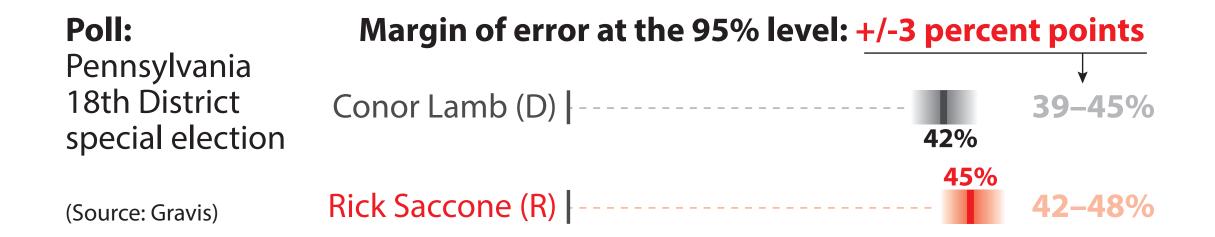
## Watch out!

- Non-normalized data
- Lack of context
- Averages only are misleading
- Too much data is just like too little data

# Concealing uncertainty



Results:	Conor Lamb (D)	49.8%
Pennsylvania 18th District	Rick Saccone (R)	49.6%
special election March 13, 2018		Conor Lamb: +0.2 percent points





### **ORIENTAMENTI DI VOTO: 22 NOVEMBRE 2021**

	Stima 22/11/2021	Stima 15/11/2021	Trend
PARTITO DEMOCRATICO	20,8	20,3	+0,5
FRATELLI D'ITALIA	19,8	20,2	-0,4
LEGA	18,4	18,7	-0,3
MOVIMENTO 5 STELLE	15,7	16,1	-0,4
FORZA ITALIA	6,8	6,9	-0,1
AZIONE	4,0	3,9	+0,1
SINISTRA ITALIANA	2,5	2,5	=

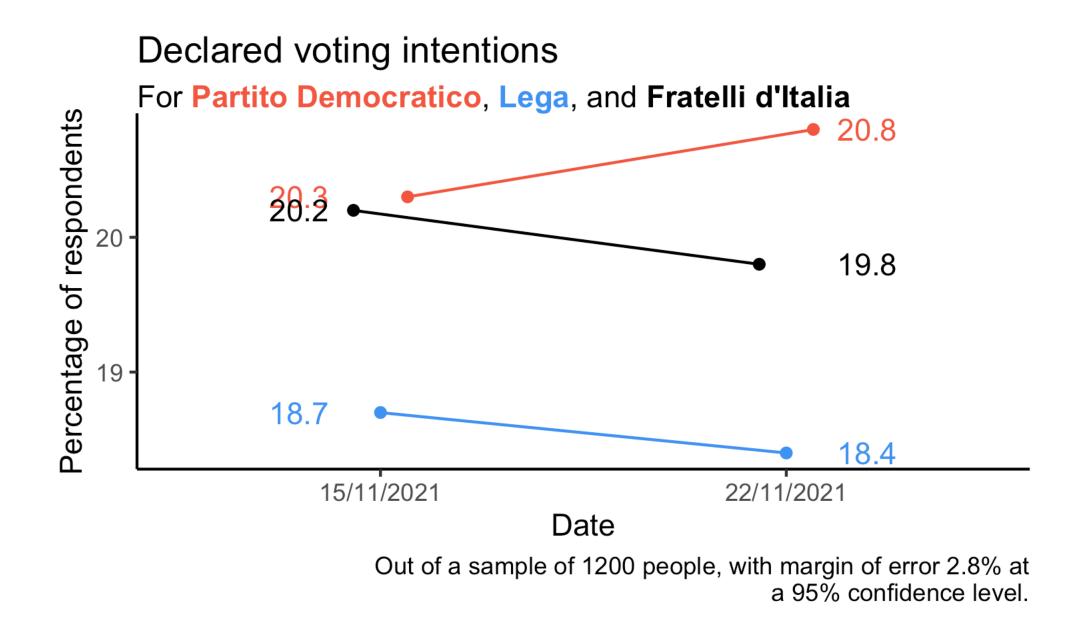


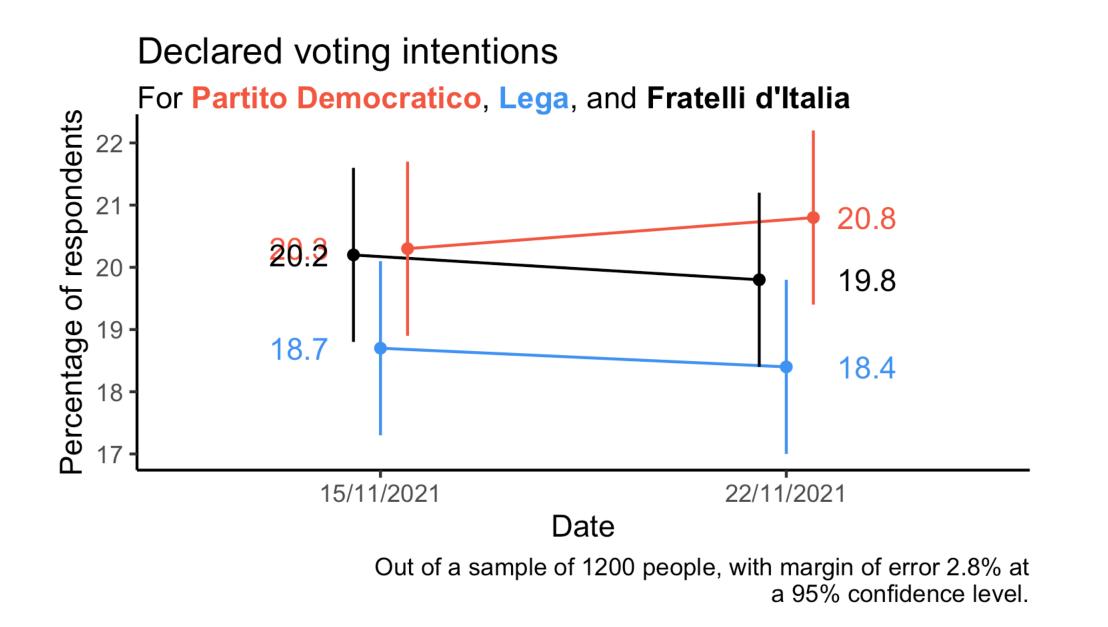
0:39 / 1:37

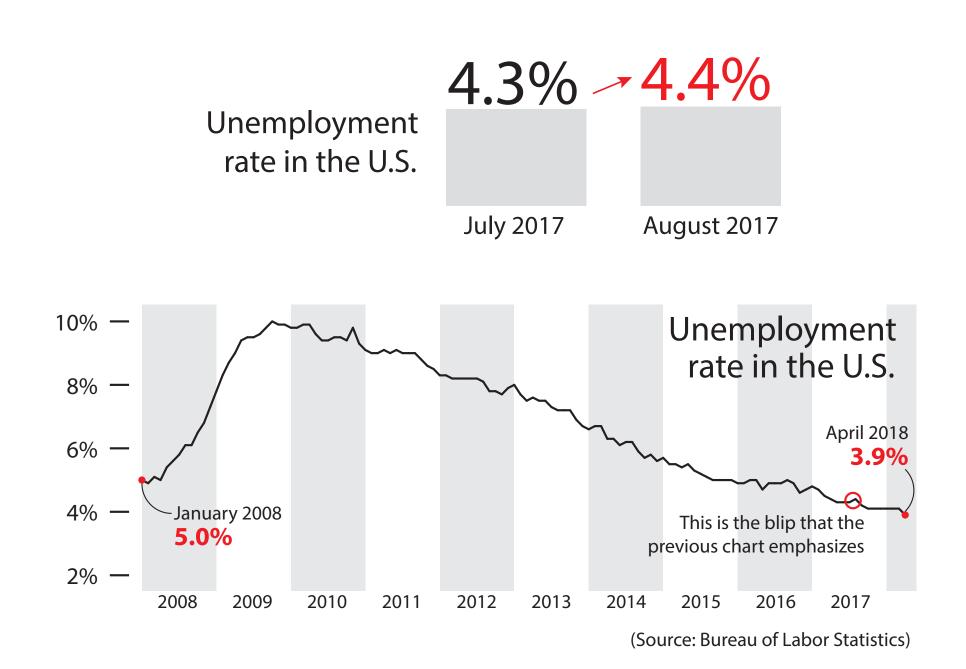
П

Valori %. Sondaggio realizzato da SWG S.p.a. per conto di La7 S.p.A.. Indagine condotta con tecnica mista CATI-CAMI-CAWI su un campione di 1.200 soggetti maggiorenni residenti in Italia (4513 non rispondenti) tra il 17 e il 22 novembre 2021. Il campione è stratificato per zona e prevede quote per età e sesso. Il margine d'errore statistico dei

dati riportati è de 0:34 % a un intervario di confidenza del 95%. Il documento informativo completo del sondaggio sarà disponibile sul sito www.sondaggipoliticoelettorali.it





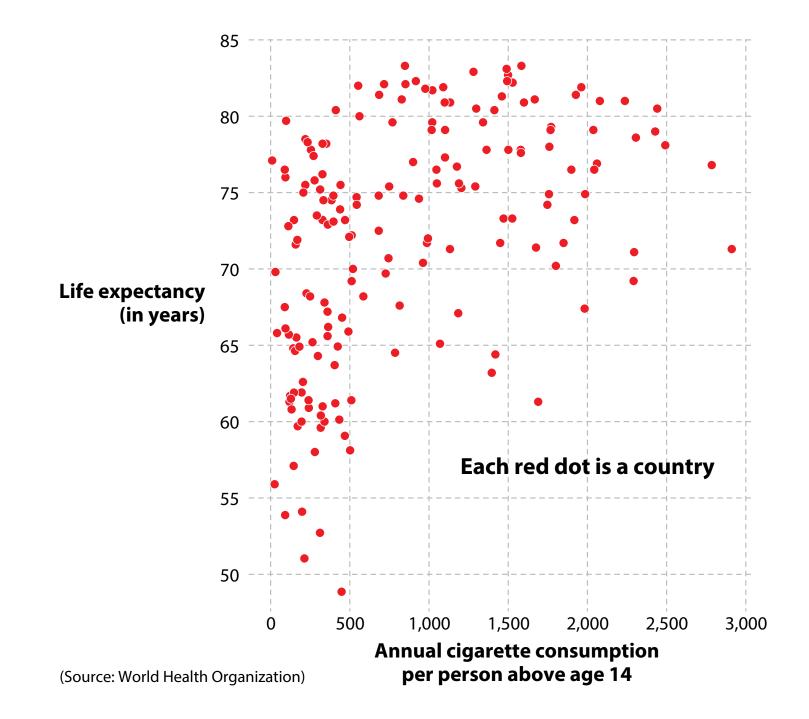


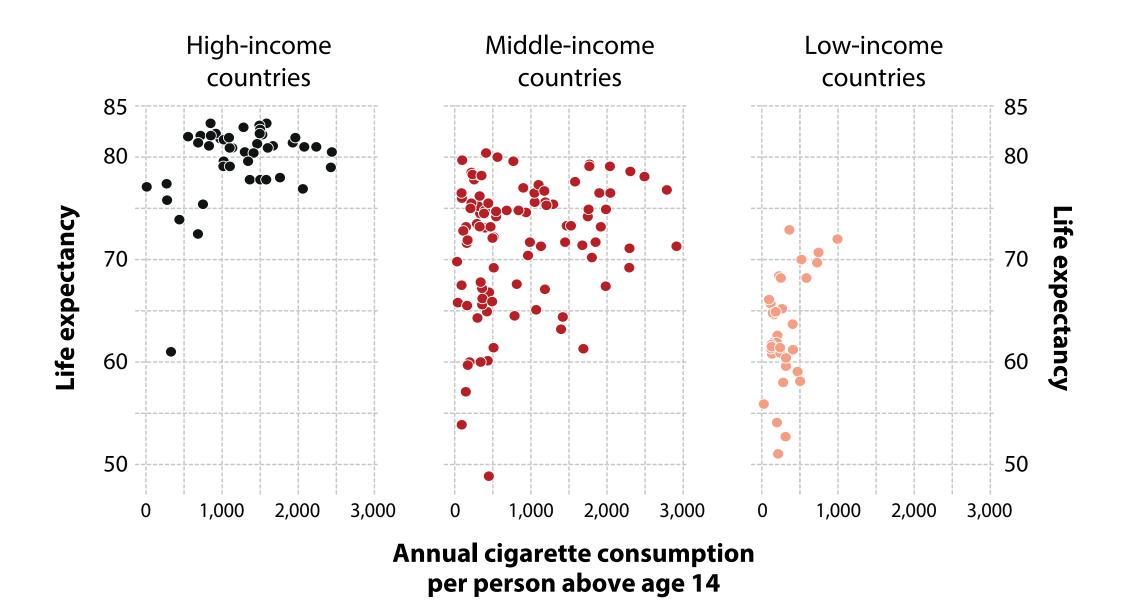
### Watch out

• Always report the uncertainty surrounding your estimates

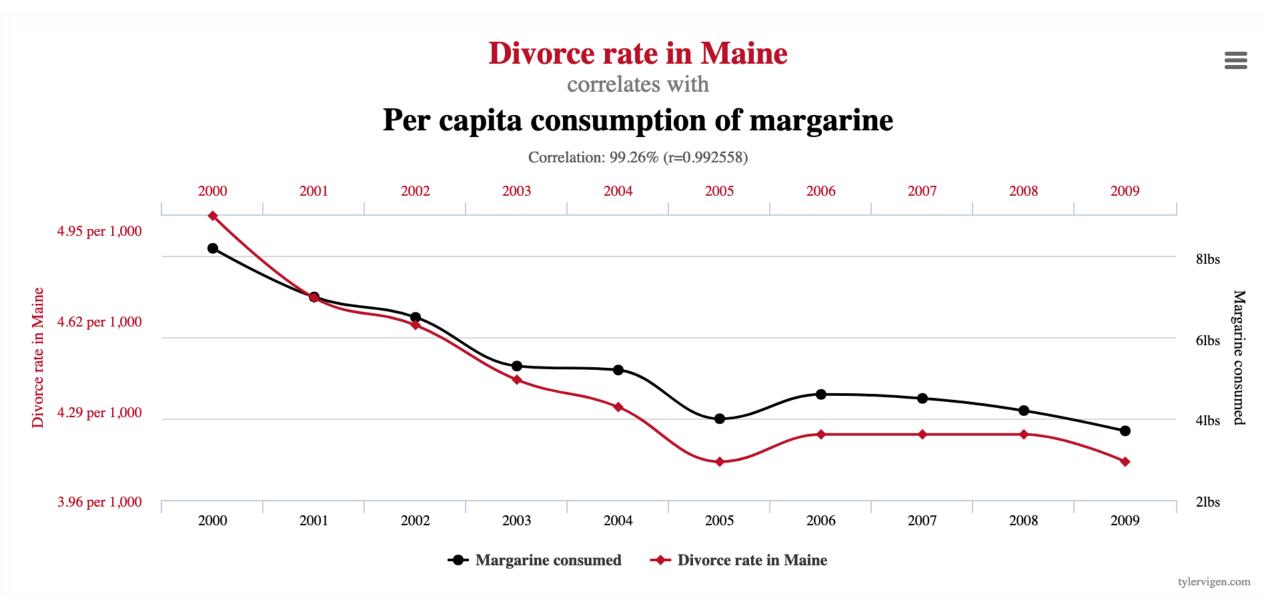
• Estimates without explicit uncertainty are almost useless.

## Suggesting misleading patterns





https://www.tylervigen.com/spurious-correlations



Data sources: National Vital Statistics Reports and U.S. Department of Agriculture

## Watch out!

• Correlation is not causation!