

Misrepresentation

Data Science in a Box
datasciencebox.org



Causality





TIME

Exercise Can Lower Risk of Some Cancers By 20%

People who were more active had on average a 20% lower risk of cancers of the esophagus, lung, kidney, stomach, endometrium and others compared with people who were less active.

Alice Park. Exercise Can Lower Risk of Some Cancers By 20%. Time Magazine. 16 May 2016.



datasciencebox.org



Los Angeles Times

Exercising drives down risk for 13 cancers, research shows

[...] those who got the most moderate to intense exercise reduced their risk of developing seven kinds of cancer by at least 20%.

Melissa Healy. Exercising drives down risk for 13 cancers, research shows .
Los Angeles Times. 16 May 2016.

Original study

Moore, Steven C., et al. "**Association of leisure-time physical activity with risk of 26 types of cancer in 1.44 million adults.**" JAMA internal medicine 176.6 (2016): 816-825.

- **Volunteers** were **asked** about their physical activity level over the preceding year.
- Half exercised less than about 150 minutes per week, half exercised more.
- Compared to the bottom 10% of exercisers, the top 10% had lower rates of esophageal, liver, lung, endometrial, colon, and breast cancer.
- Researchers found no association between exercising and 13 other cancers (e.g. pancreatic, ovarian, and brain).

Carl Bergstrom and Jevin West. Calling Bullshit: The art of skepticism in a data-driven world.
Random House, 2020.

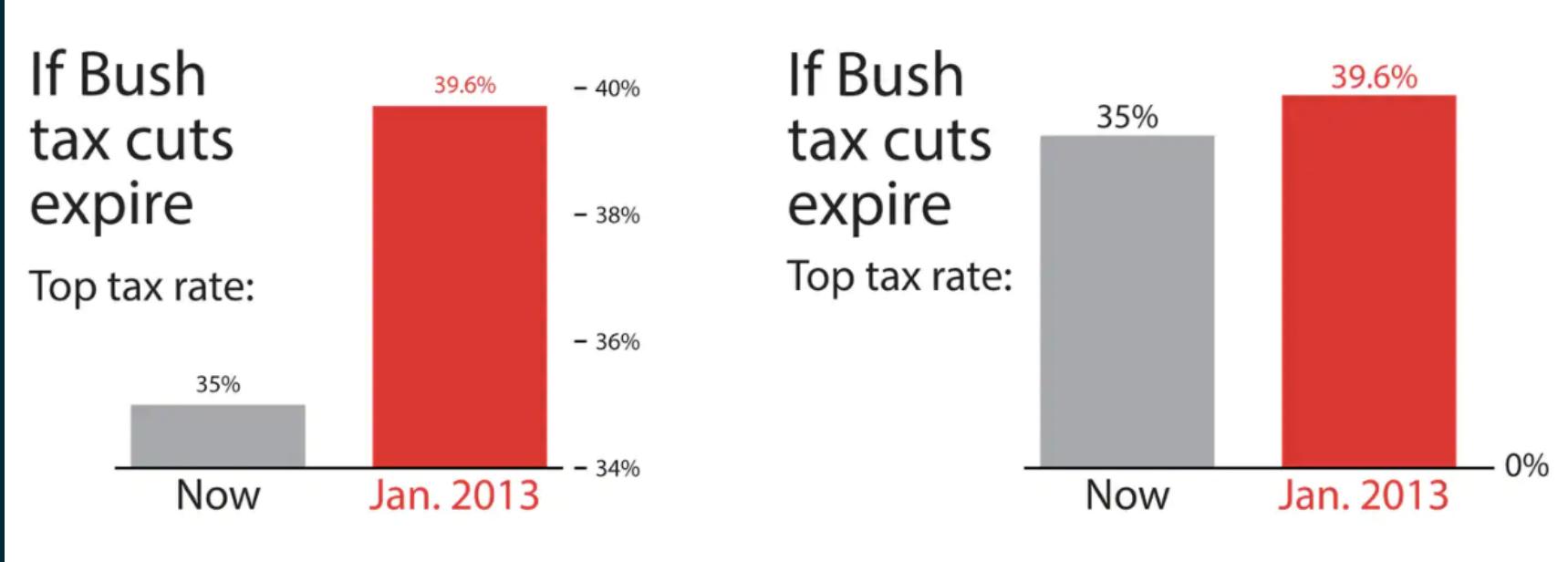
Sharon Begley. "Does exercise prevent cancer?". StatNews. 16 May 2016.



Axes and scale

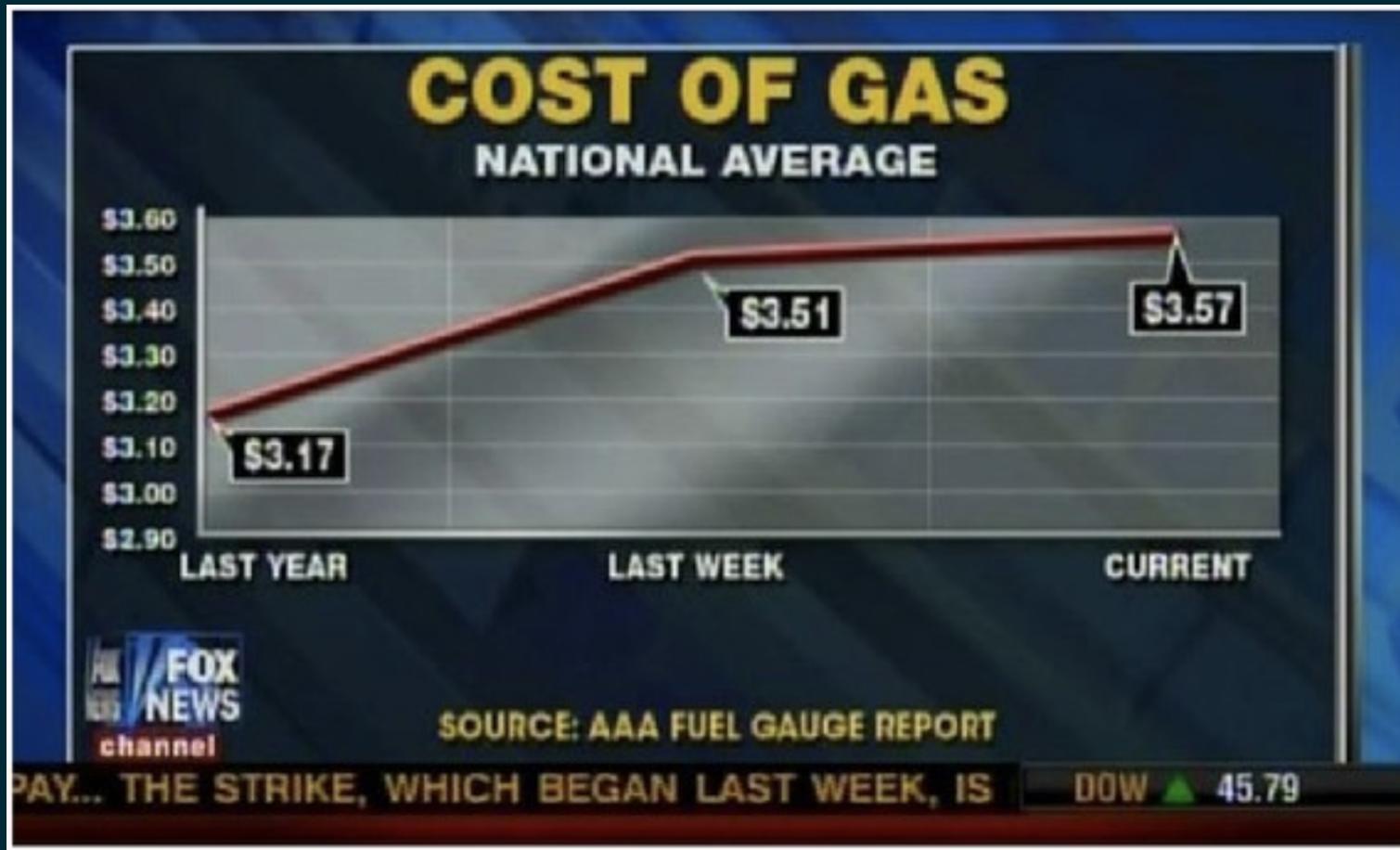


What is the difference between these two pictures? Which presents a better way to represent these data?

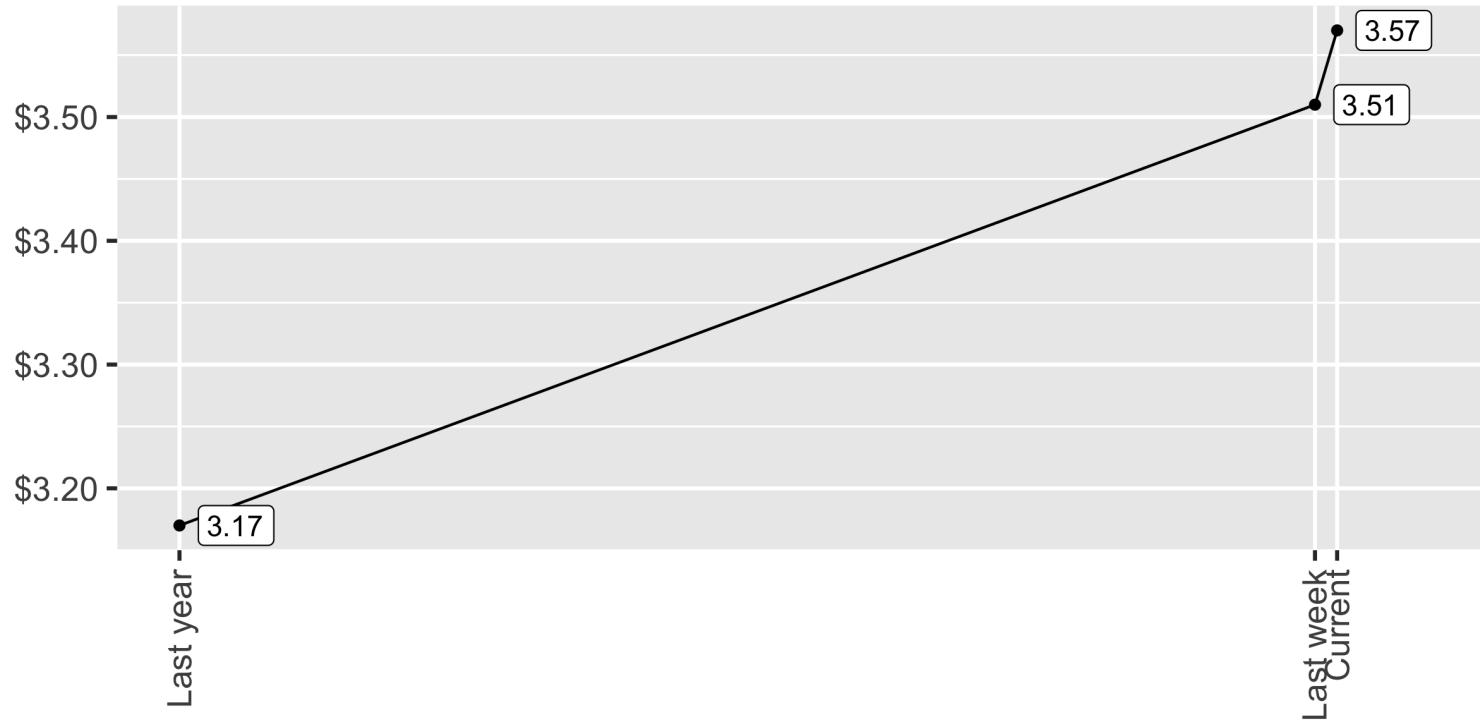


Christopher Ingraham. "You've been reading charts wrong. Here's how a pro does it.". The Washington Post. 14 October 2019.

What is wrong with this picture? How would you correct it?



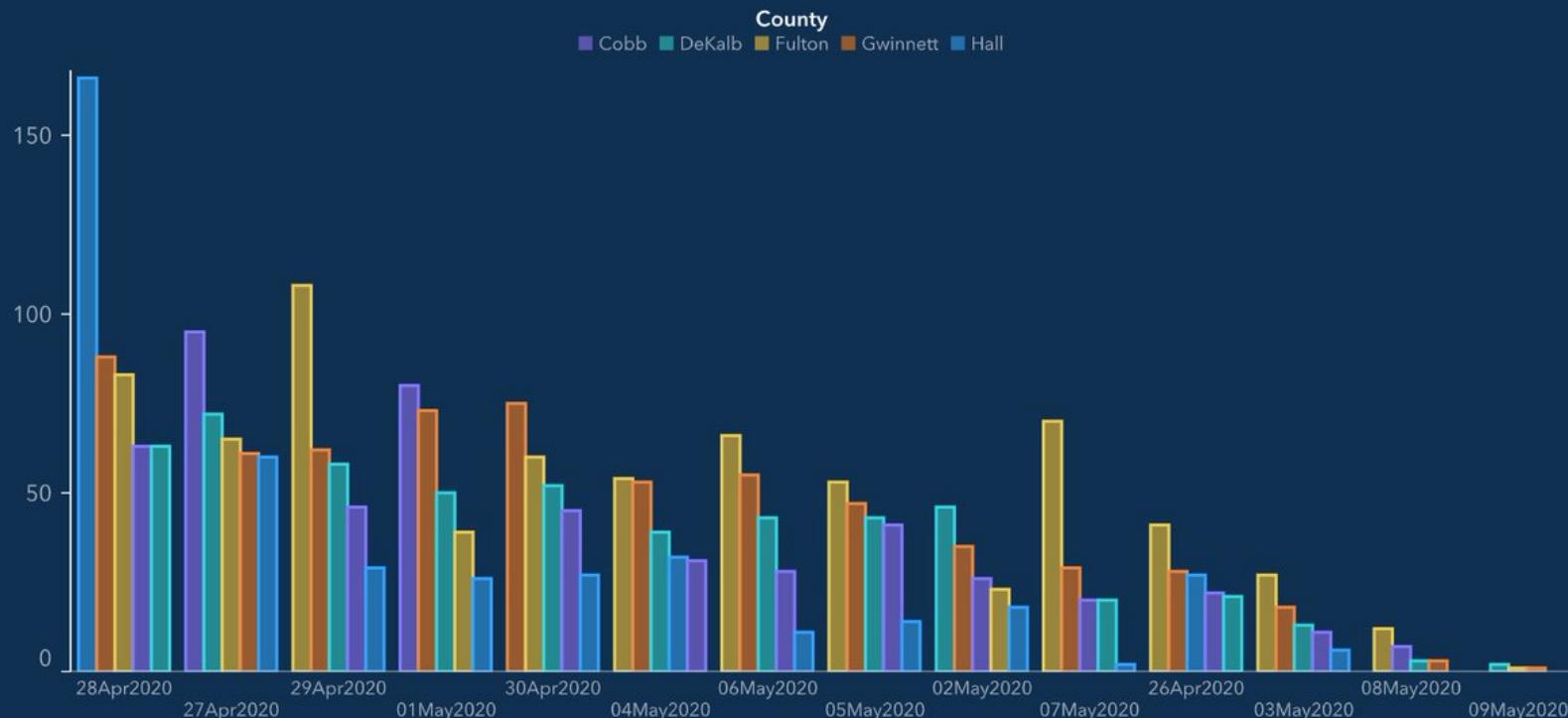
Cost of gas National average



What is wrong with this picture? How would you correct it?

Top 5 Counties with the Greatest Number of Confirmed COVID-19 Cases

The chart below represents the most impacted counties over the past 15 days and the number of cases over time. The table below also represents the number of deaths and hospitalizations in each of those impacted counties.



Graph detective

1



Graph detective (..../../.../2020/05/17/graph-detective/)

 Lucy D'Agostino McGowan

 May 17, 2020  12 min read

 estate / / / / testacorras/testacorras/covid-19 / / / / testacorras/covid-19/data-visualizations

Lucy D'Agostino McGowan. Graph detective. Live Free or Dichotomize. 17 May 2020.



datasciencebox.org

Maps and areas



Do you recognize this map? What does it show?



Lazaro Gamio. "Election maps are telling you big lies about small things".
The Washington Post. 1 Nov 2016.

WRONG

CITIZENS ^{FOR}
TRUMP



THE INSIDE STORY
OF THE PEOPLE'S MOVEMENT
TO TAKE BACK AMERICA



JACK POSOBIEC

RIGHT

COUNTIES ^{FOR}
TRUMP



[Download](#)

THE INSIDE STORY
OF 46% OF VOTERS' MOVEMENT
TO TAKE BACK AMERICA

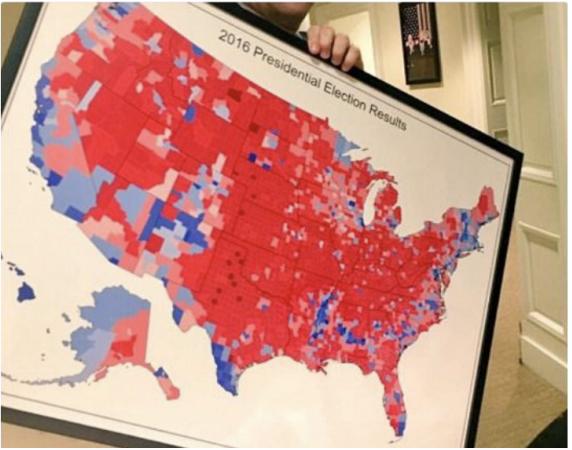


JACK POSOBIEC

Alberto Cairo. Visual Trumpery talk.



datasciencebox.org



Surface on the
county-level map:

Red: 80%

Blue: 20%

SHARE OF THE POPULAR VOTE IN THE 2016 PRESIDENTIAL ELECTION

Donald Trump	46.1%	62,984,825 votes
Hillary Clinton	48.2%	65,853,516 votes
Other candidates	5.7%	

PERCENTAGE OF ELIGIBLE VOTERS

Didn't vote	40.0%
Voted for Donald Trump	27.7%
Voted for Hillary Clinton	28.9%
Voted for other candidates	3.4%

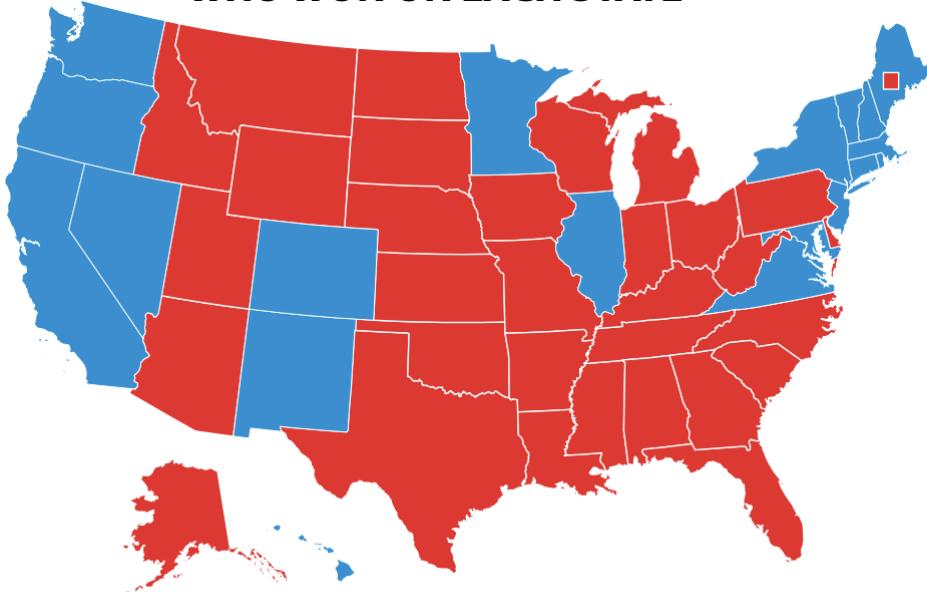
Alberto Cairo. Visual Trumpery talk.

ELECTORAL
VOTES

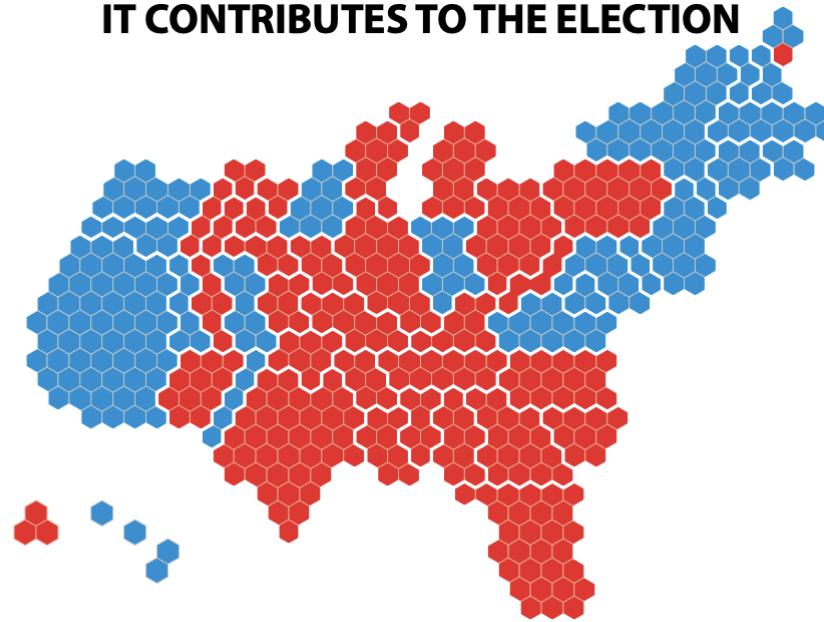
TRUMP
306

CLINTON
232

WHO WON ON EACH STATE



STATE SIZE ADJUSTED BY ELECTORAL VOTES
IT CONTRIBUTES TO THE ELECTION



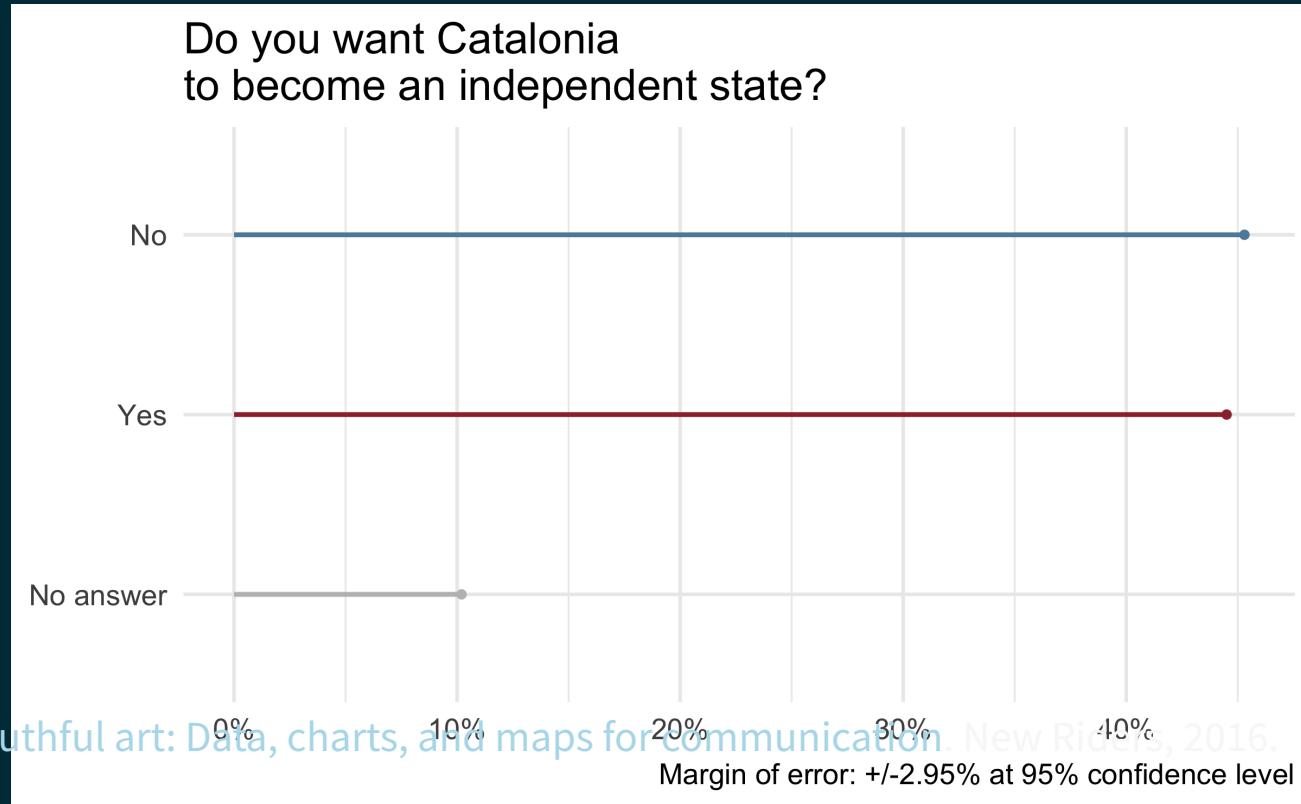
Alberto Cairo. Visual Trumpery talk.

Visualising uncertainty

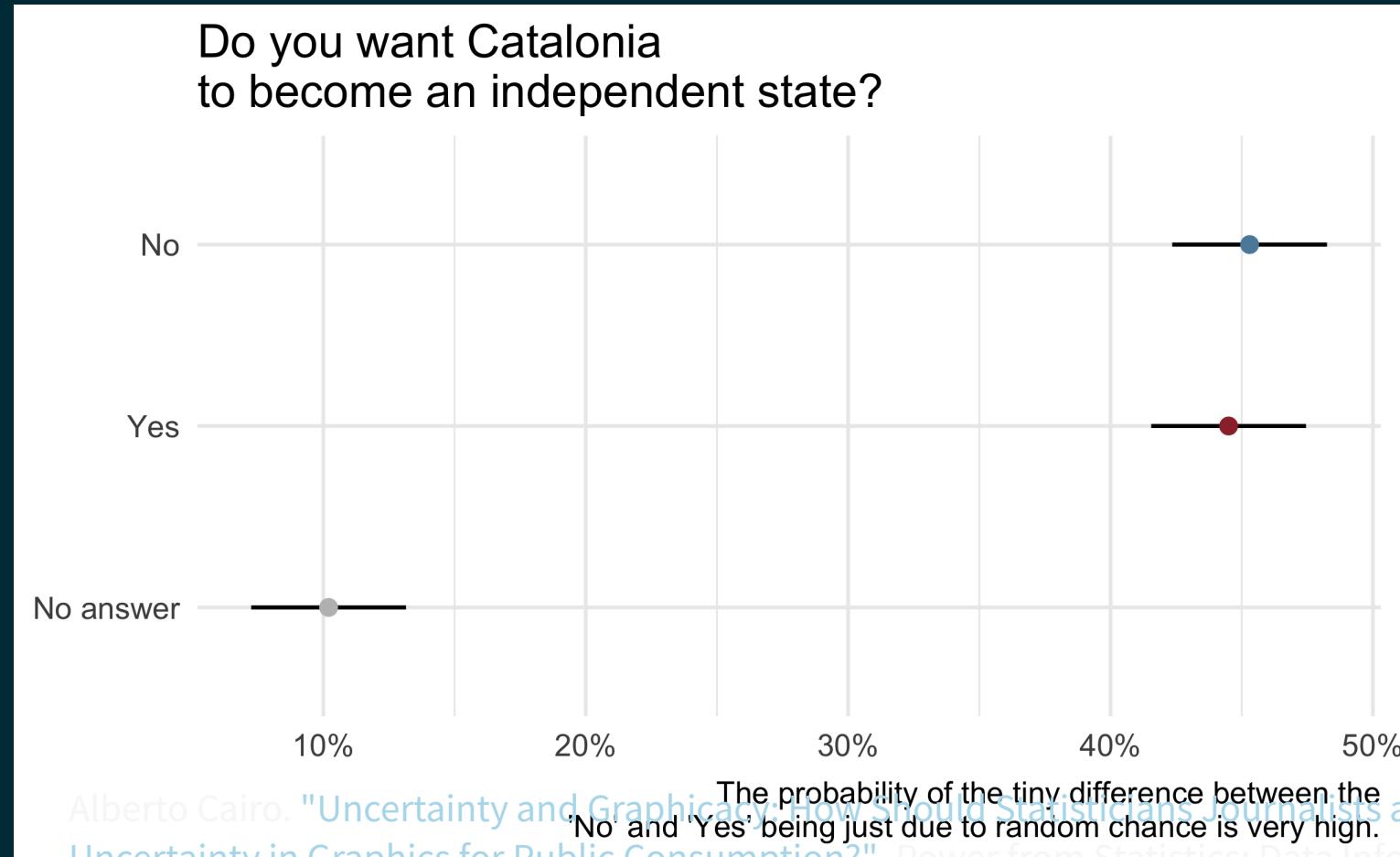


On December 19, 2014, the front page of Spanish national newspaper El País read "*Catalan public opinion swings toward 'no' for independence, says survey*".

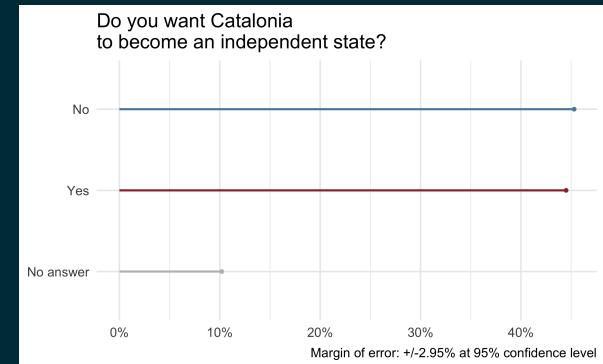
```
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use
## `guides(<scale> = "none")` instead.
```



```
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use  
## `guides(<scale> = "none")` instead.
```



```
## Warning: `guides(<scale> = FALSE)` is deprecated. Please use  
## `guides(<scale> = "none")` instead.
```



Alberto Cairo. "Uncertainty and Graphicacy: How Should Statisticians, Journalists and Designers Reveal Uncertainty in Graphics for Public Consumption?", *Power from Statistics: Data Information and Knowledge*, 2017.

Further reading



How Charts Lie

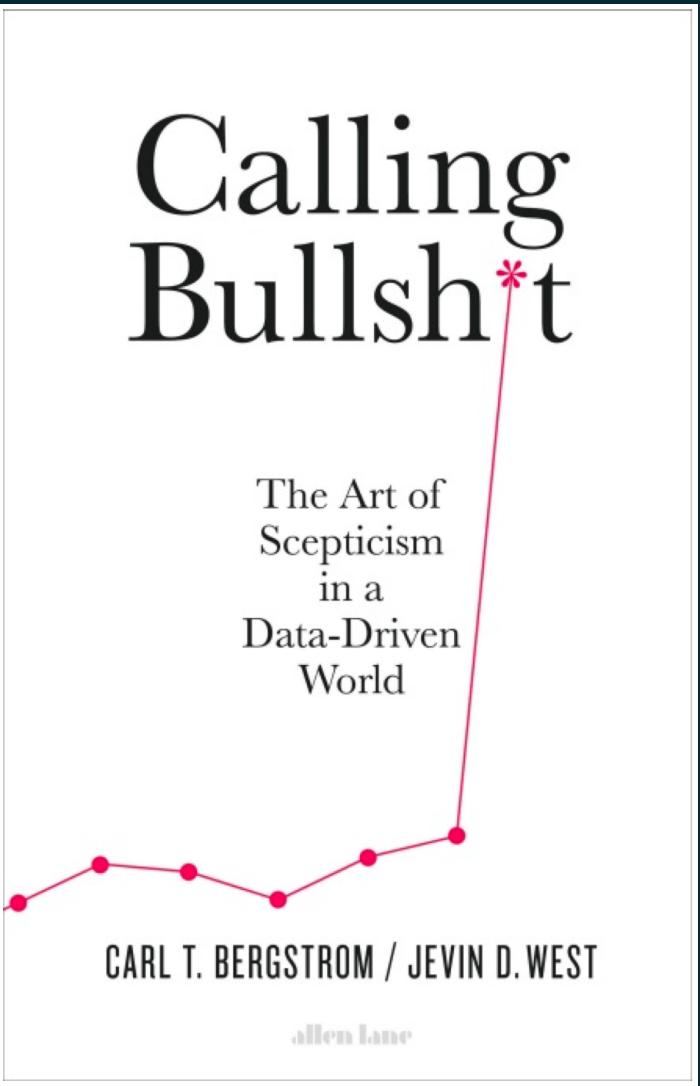


Getting Smarter about
Visual Information

Alberto Cairo

How Charts Lie
Getting Smarter about Visual Information
by Alberto Cairo





Calling Bullshit
The Art of Skepticism in a
Data-Driven World

by Carl Bergstrom and Jevin West



datasciencebox.org