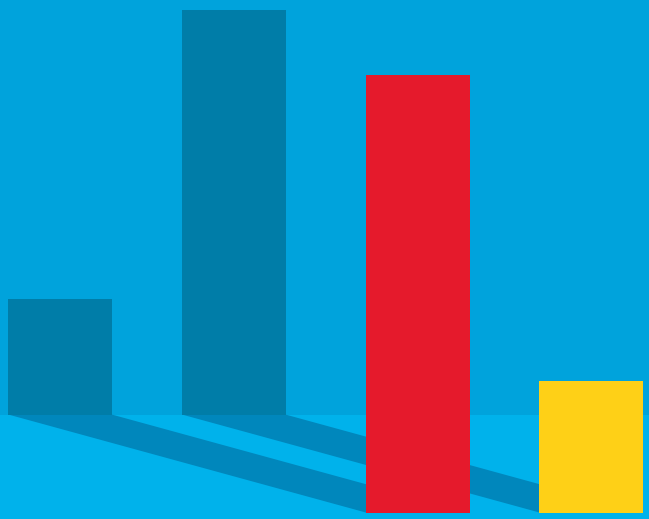
How Charts Lie



Getting Smarter about Visual Information

Alberto Cairo

Praise for

How Charts Lie

"Funny and engaging. A must-read for anyone who wants to stay informed."

—**CATHY O'NEAL**, author of **Weapons of Math Destruction**

"This book offers a succinct, elegant, accessible look at the ways data can be represented or misrepresented and is a perfect primer for anyone who cares about the difference. I loved this book!"

—**CHARLES WHEELAN**, author of **Naked Statistics**

"I wish we lived in a world where you didn't need to read Alberto Cairo's How Charts Lie, a robust guide to self-defense against graphs and figures designed to mislead. But here we are, and yes, you do."

—**JORDAN ELLENBERG**, author of **How Not to Be Wrong**

"Alberto Cairo has written a wise, witty, and utterly beautiful book. You couldn't hope for a better teacher to improve your graphical literacy."

—**TIM HARFORD**, author of **The Undercover Economist** and presenter of **More or Less**

"Alberto Cairo shares great examples of data visualization and storytelling for anyone who wants to dig into their data."

—**DONA WONG**, author of **The Wall Street Journal Guide to Information Graphics**

"A picture may be worth a thousand words, but only if you know how to read it. In this book, Alberto Cairo teaches us how to get smarter about visual information by reading charts with attention and care. I found a lot to steal here, and you will, too."

—**AUSTIN KLEON**, author of **Steal Like an Artist**

"This book will open your eyes to how everyone uses visuals to push agendas. A master visual designer, Alberto Cairo shows you how to read charts and decode design. After this book, you can't look at charts with a straight face!"

—**KAISER FUNG**, author of **Numbers Rule Your World**

This document contains the
Introduction of Alberto Cairo's

How Charts Lie

To purchase the book visit

Indibound.com

Amazon.com

Amazon.co.uk

Barnes&Noble

How Charts Lie

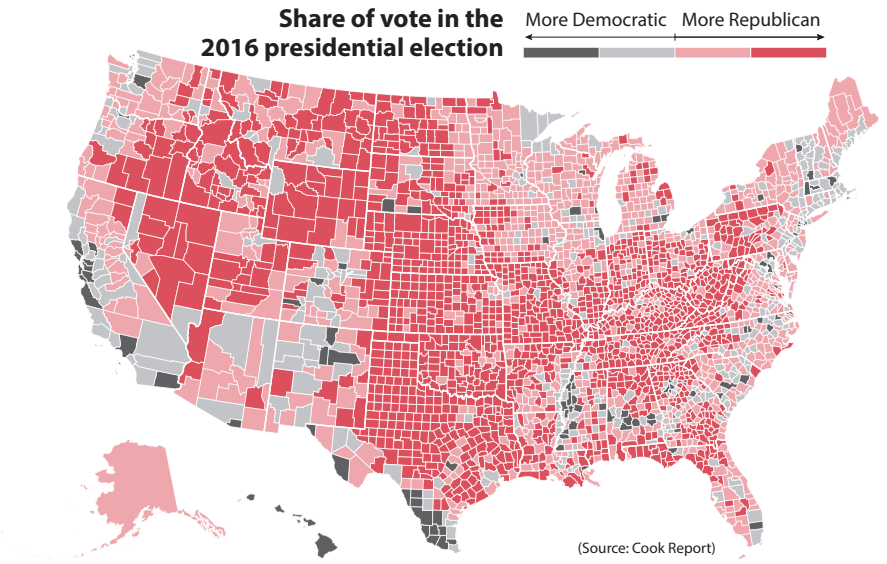
Getting Smarter about Visual Information

Alberto Cairo

W. W. NORTON & COMPANY
Independent Publishers Since 1923

Introduction

On April 27, 2017, President Donald J. Trump sat with Reuters journalists Stephen J. Adler, Jeff Mason, and Steve Holland to discuss his accomplishments in his first 100 days in office. While talking about China and its president, Xi Jinping, Trump paused and handed the three visitors copies of a 2016 electoral map:[1]

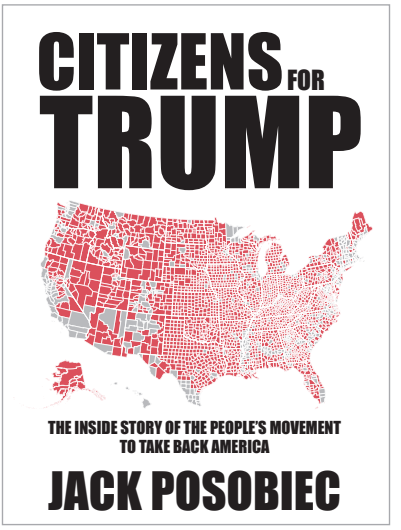


The president then said, "Here, you can take that, that's the final map of the numbers. It's pretty good, right? The red is obviously us."

When I read the interview, I thought that it was understandable President Trump was so fond of that map. He won the 2016 election despite most forecasts, which gave him between 1% and 33% chances of succeeding; a Republican establishment that distrusted him; a bare-bones campaign that was often in disarray; and numerous controversial remarks about women, minorities, the U.S. intelligence services, and even veterans. Many pundits and politicians predicted Trump's demise. They were proved wrong. He seized the presidency against all odds.

However, being victorious isn't an excuse to promote misleading visuals. When presented alone and devoid of context, this map can be misleading.

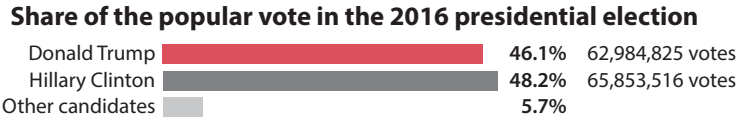
The map appeared in many other places during 2017. According to The Hill,[2] White House staffers had a large, framed copy of it hanging in the West Wing. The map was also regularly touted by conservative media organizations, such as Fox News, Breitbart, and InfoWars, among others. Right-wing social media personality Jack Posobiec put it on the cover of his book, *Citizens for Trump*, which looks similar to this:



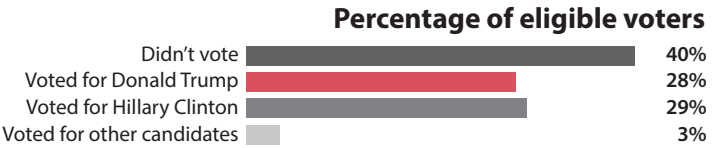
I've spent the last two decades making charts and teaching others how to design them. I'm convinced that anyone—including you, reader—can learn how to read and even *create* good graphics, so I'm usually happy to offer my free and constructive advice to whoever wants to take it. When I saw Posobiec's book on social media, I suggested that he needed to change either the title or the map, as the map doesn't show what the book title says.

The map is misleading because it's being used to represent the *citizens* who voted for each candidate, but it doesn't. Rather, it represents *territory*. I suggested that Posobiec either change the graphic on the cover of his book to better support the title and subtitle, or change the title to *Counties for Trump*, as that is what the map truly shows. He ignored my advice.

Try to estimate the proportion of each color, red (Republican) and grey (Democratic). Roughly, 80% of the map's surface is red and 20% is grey. The map suggests a triumph by a landslide, but Trump's victory wasn't a landslide at all. The popular vote—Posobiec's "citizens"—was split nearly in half:



We could be even pickier and point out that turnout in the election was around 60%;[3] more than 40% of eligible voters didn't show up at the polls. If we do a chart of *all eligible voters*, we'll see that the citizens who voted for each of the major candidates were a bit less than a third of the total:

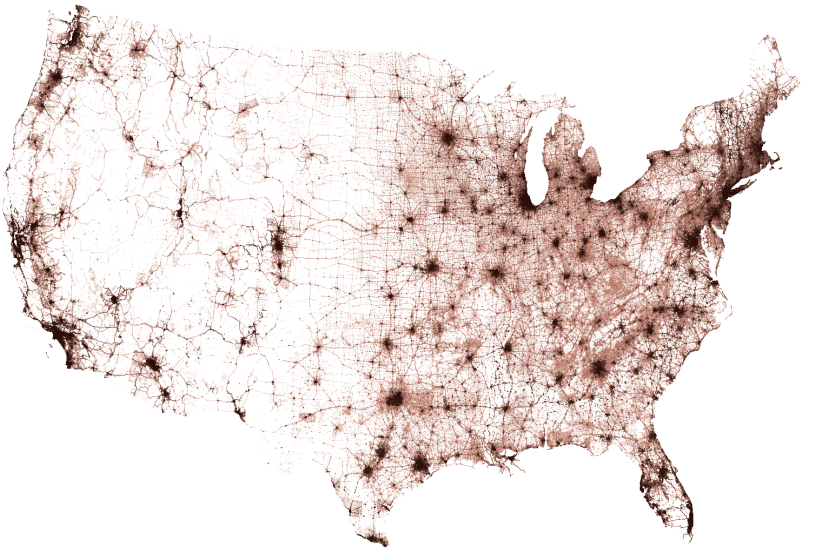


And what if we count *all* citizens? There are 325 million people in the United States. Of those, around 300 million are citizens, according to the

Kaiser Foundation. It turns out that "Citizens for Trump" or "Citizens for Clinton" are just a bit more than one-fifth of all citizens.

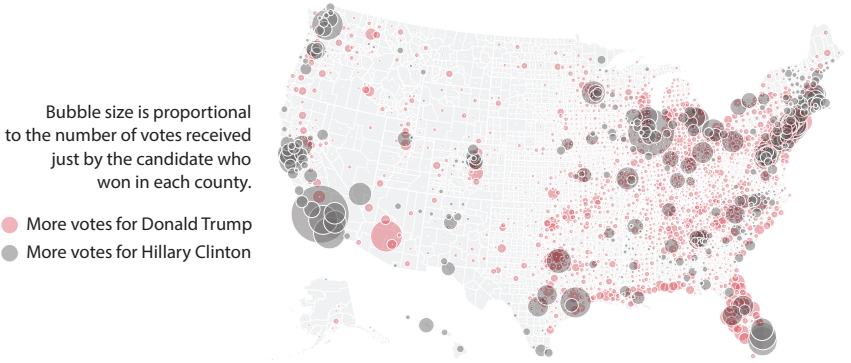
Critics of President Trump were quick to excoriate him for handing out the county-level map to visitors. Why count the square miles and ignore the fact that many counties that went for Trump (2,626)[4] are large in size but sparsely populated, while many of the counties where Clinton won (487) are small, urban, and densely populated?

That reality is revealed in the following map of the continental U.S., designed by cartographer Kenneth Field. Each dot here represents a voter—grey is Democratic and red is Republican—and is positioned approximately—but not exactly—where that person voted. Vast swaths of the U.S. are empty:



As someone who strives to keep a balanced media diet, I follow people and publications from all ideological stripes. What I've seen in recent years makes me worry that the increasing ideological polarization in the U.S. is also leading to a divide on chart preferences. Some conservatives I read love the county-level map President Trump handed out to reporters. They constantly post it on their websites and social media accounts.

Liberals and progressives, on the other hand, prefer a bubble map proposed by *Time* magazine and other publications.[5] In it, bubbles are sized in proportion to the votes received by the winning candidate in each county:



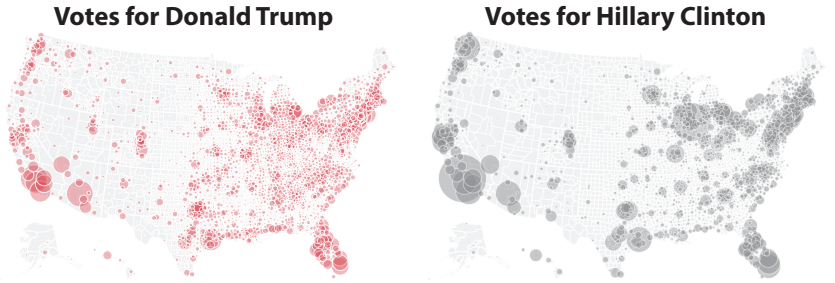
Both conservatives and liberals laugh at the other side's stupidity. "How can you tweet *that* map? Don't you see that it distorts the results of the election?"

This is no laughing matter. Both sides in this debate are throwing different charts at each other because we all often use information to reinforce our beliefs: conservatives love to convince themselves of a crushing victory in the 2016 election; liberals console themselves by emphasizing Hillary Clinton's larger share of the popular vote.

Liberals are correct when they claim that the colored county map isn't an adequate representation of the *number* of votes each candidate received, but the bubble map favored by liberals is also faulty. By showing only the votes for the winning candidate in each county, this chart ignores those received by the *losing* candidate. Plenty of people voted for Hillary Clinton in conservative regions. Many voted for Donald Trump in very progressive ones.

Kenneth Field's map or the pair of maps below may be a better choice if what we care about is the popular vote. There are many more visible red bubbles (votes for Trump) than grey bubbles (votes for Clinton), but the fewer grey ones are often much bigger. When these maps are put side by side, it's easier to see why the election was decided by a relatively small number of

votes in a handful of states; if you add up the area of all red bubbles and the area of all grey bubbles, they are roughly the same:



Bubble size is proportional to the number of votes per county

Having said this, both conservatives and liberals are missing the point. What makes you win a presidential election in the United States is neither the territory you control, nor the number of people you persuade to vote for you *nationally*. It's the Electoral College and its 538 electors. To win, you need the support of at least 270 of electors.

Each state has a number of these folks equal to its congressional representation: two senators plus a number of representatives in the House that varies according to the state's population. If you are a small state with the fixed number of senators (two per state) plus one representative in the House, you are allotted three electors.

Small states often have more electors based on their populations than what pure arithmetic would give them: the minimum is three electors per state, no matter how small the population of that state is.

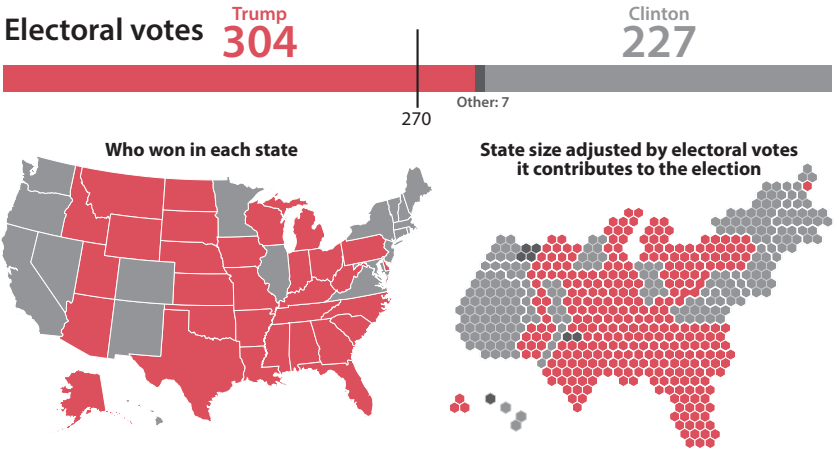
Here's how you receive the support of a state's electors: with the exception of Nebraska and Maine, the candidate who wins even a razor-thin advantage in a state's popular vote over his or her opponents is supposed to receive the support of *all* that state's electors.

In other words, once you've secured at least one more vote than any of your opponents, the rest of the votes you receive in that state are useless. You don't even need a majority, just a *plurality*: if you get 45% of the popular vote

in one state, but your two opponents get 40% and 15%, you'll receive all the electoral votes from that state.

Trump got the support of 304 electors. Clinton, despite winning the national popular vote by a margin of three million and getting tons of support in populous states like California, received only 227. Seven electors went rogue, voting for people who weren't even candidates.

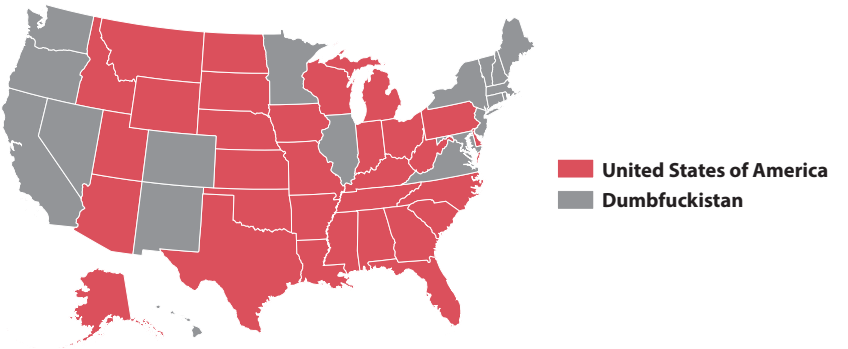
Therefore, if I ever got elected president—which is an impossibility, since I wasn't born in the U.S.—and I wanted to celebrate my victory by printing out some charts, framing them, and hanging them on the walls of my White House, it would be with the ones below. They are focused on the figures that really matter—neither the number of counties, nor the popular vote, but the number of *electoral* votes received by each candidate:



Maps are among the many kinds of charts you'll learn about in this book. Sadly, they are among the most misused. In July of 2017, I read that a popular U.S. singer called Kid Rock was planning to run for the Senate in the 2018 election.[6] He'd later claim that it was all a joke,[7] but it sounded like a serious bid at the time.

I didn't know much about Kid Rock, so I wandered through his social

media accounts and saw some of the merchandise he was selling in his online store, KidRock.com. I love graphs and maps, so one T-shirt with an intriguing map of the results of the 2016 election was irresistible. Its legend indicated that, according to Mr. Rock, the results of the election matched the boundaries of two separate countries:

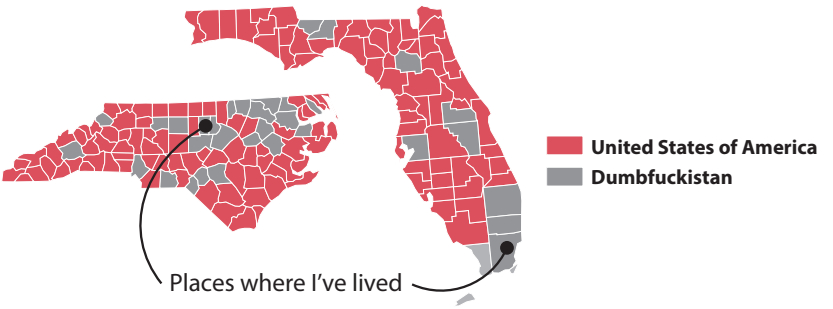


As you might now expect, this map isn't an accurate representation of the borders between the United States (read: Republican America) and Dumbfuckistan (read: Democratic America). An electoral precinct- or county-level map may be much more accurate.

Now, as an aside, I want to point out that I lived in North Carolina between 2005 and 2008. Originally from Spain, I knew little about the Tar Heel State before I arrived, other than that it was often red on the presidential election maps I'd always seen in Spanish newspapers. I was expecting to settle in a conservative place. Fine with me. I'm ideologically moderate. But my expectations were misguided. To my surprise, when I arrived, I didn't land in the United States of America—if we follow Kid Rock's nomenclature—I landed in deep Dumbfuckistan! The Chapel Hill–Carrboro area, in Orange County (North Carolina), where I lived, is quite progressive and liberal, more so than most of the rest of the state.

The city where I am now, Kendall (Florida), part of the greater Miami area, is also quite proud of its Dumbfuckistani heritage. The following maps

reveal what I'd say are the true borders between the two countries Mr. Rock's T-shirt depicts:

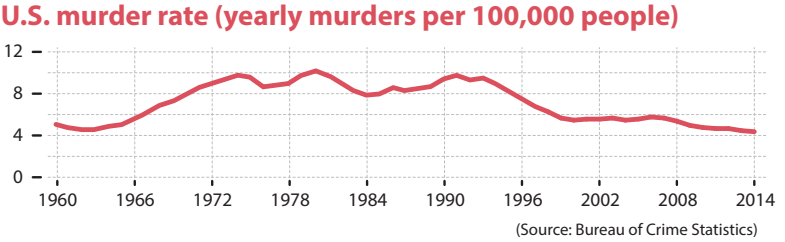


President Donald Trump gave his first State of the Union address on January 30, 2018. Pundits on the right sang praises to his great performance as he read from a teleprompter, and those on the left criticized him. Trump devoted some time to talking about crime and got the attention of economist and Nobel Prize winner Paul Krugman, a columnist for the *New York Times*.

On several occasions during the presidential campaign in 2016, and also during his first year in office, Trump mentioned a supposedly sharp increase of violent crime in the United States, particularly murders. Trump blamed undocumented immigrants for this, an assertion that has been debunked many times over and that Krugman called a "dog whistle" in his column.[8]

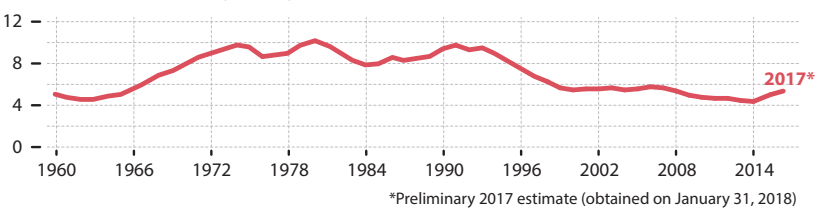
However, Krugman didn't stop there. He added that Trump wasn't "exaggerating a problem, or placing the blame on the wrong people. He was inventing a problem that doesn't exist" as "there is no crime wave—there have been a few recent bobbles, but many of our big cities have seen both a surge in the foreign-born population and a dramatic, indeed almost unbelievable, decline in violent crime."

Here's a chart that Krugman provided as evidence:



It seems that what Krugman said is true: the United States has witnessed a noticeable drop in murders since the peaks in the 1970s, 1980s, and early 1990s. The trend is similar for violent crime in general.

However, isn't it odd that an article published at the beginning of 2018 includes only years up to 2014? While detailed crime statistics are hard to obtain, and it would be impossible to get a good estimate up to the day when Krugman's column was published, the FBI already had solid stats for 2016 and a preliminary estimate for 2017.[9] This is what the chart looks like if we add those years. The murder rate increased in 2015, 2016, and 2017. It doesn't look like a "bobble" at all:



I doubt that someone with Krugman's record would conceal relevant data intentionally. Based on my own experience as a chart designer and journalist who's made plenty of silly mistakes, I've learned to never attribute to malice what could be more easily explained by absentmindedness, rashness, or sloppiness.

It's true, as Krugman wrote, that the murder rate today is way lower

than it was thirty years ago. If you zoom out and take a look at the entire chart, the overall long-term trend is one of decline. Tough-on-crime politicians and pundits often ignore this, quite conveniently, and focus instead on the last few years.

However, the uptick since 2014 is relevant and shouldn't be concealed. How relevant is it, though? That depends on where you live.

This national murder rate chart, as simple and easy to read as it looks, *hides as much as it reveals*. This is a common feature of charts, since they are usually simplifications of very complex phenomena. Murders aren't increasing everywhere in the United States. Most places in the U.S. are pretty safe.

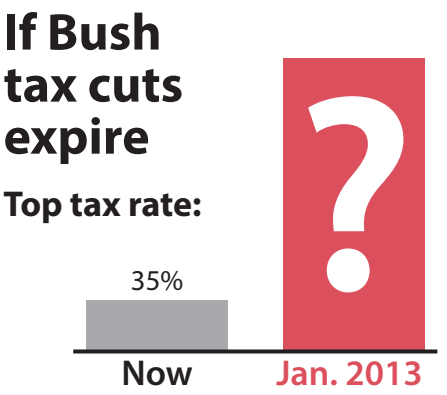
Instead, murder in the U.S. is a localized challenge: some neighborhoods in mid-sized and big cities have become so violent that they distort the national rate.[10] If we could plot those neighborhoods on the chart, they would be way above its upper gridline, perhaps even beyond the top edge of the page! If we took them off the chart, the national-murder-rate line might stay flat or even go down in recent years.

Doing this wouldn't be appropriate, of course: those cold numbers represent people being killed. However, we can and *should* demand that, when discussing data like this, politicians and pundits mention *both* overall rates *and the extreme values—also called "outliers"—that may be distorting those rates*.

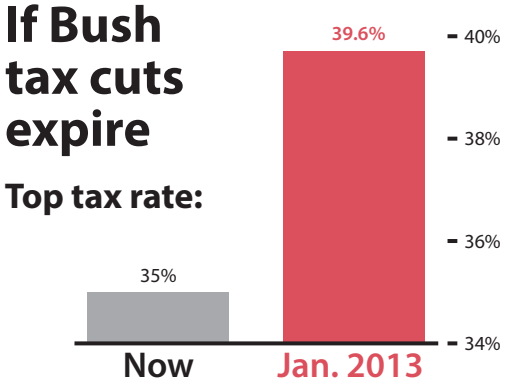
Here's an analogy to convey the statistics and help you grasp the role of outliers: Imagine you're in a bar enjoying a beer. Eight other people are drinking and chatting. None of you has killed anyone in your life. Then, a tenth person comes in, a hitman for the Mob who's dispatched 50 rivals in his career. Suddenly, the average kill count per drinker in the bar jumps to 5! But of course that doesn't automatically make you an assassin.

Charts may lie, then, because they display either the wrong information or too little information. However, a chart can show the right type and amount of information and lie anyway due to poor design or labeling.

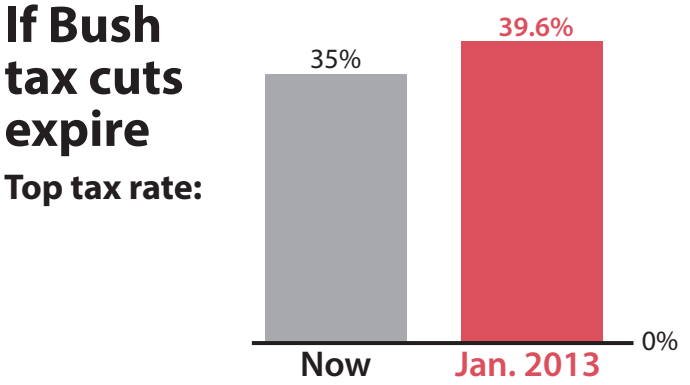
In July 2012, Fox News announced that President Barack Obama was planning to let President George W. Bush's cuts to the top federal tax rate expire by the beginning of 2013. The very wealthy would see their tax bills increase. By how much? Please estimate the height of the second bar in comparison with the first one, which represents the top tax rate under President Bush. It's a massive tax increase!



The chart that Fox displayed for a few seconds contained the figures, but they were quite tiny and hard to read. Notice that the tax increase was roughly five percentage points, but the bars were grossly misshapen to exaggerate it:



I like taxes as little as anyone else, but I dislike arguments defended with dubious charts even more, regardless of their creators' political leanings. Whoever designed this chart broke an elementary principle of chart design: if your numbers are represented by the length or height of objects—bars, in this case—the length or height should be *proportional* to those numbers. Therefore, it's advisable to put the baseline of the chart at zero:



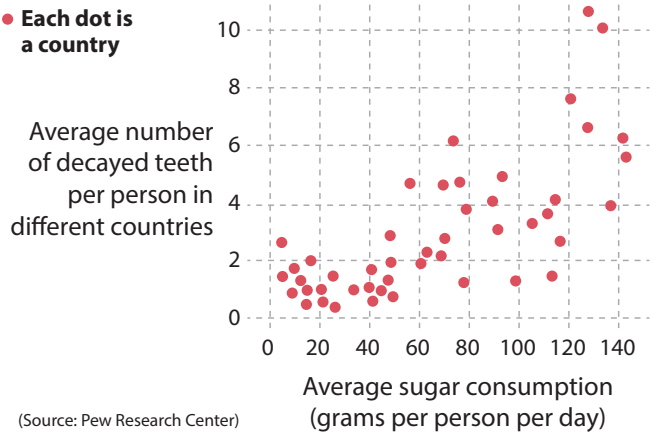
Starting a bar chart at a baseline different from zero is the most conspicuous trick in the book to distort your perception of numbers. But fudging with scales is just one of the many strategies used by swindlers and liars from all ideological denominations. There are many others that are far less easy to spot, as we'll soon see.

Even if a chart is correctly designed, it may still deceive us because we don't know how to read it correctly—we can't grasp its symbols and grammar, so to speak—or we misinterpret its meaning, or both. Contrary to what many people believe, most good charts aren't simple, pretty illustrations that can be understood easily and intuitively.

On September 10, 2015, the Pew Research Center published a survey

testing U.S. citizens' knowledge of basic science.[11] One of the questions asked participants to decode the following chart. Try to read it and don't worry if you get it wrong:

Relationship between sugar consumption per person and average number of decayed teeth



(Source: Pew Research Center)

In case you've never seen a chart like this, it's called a *scatter plot*. Each dot is a country; we don't need to know which one. The position of each dot on the horizontal axis corresponds to the daily sugar consumption per person. In other words, the farther to the right a dot is, the more sugar people consume in that country, on average.

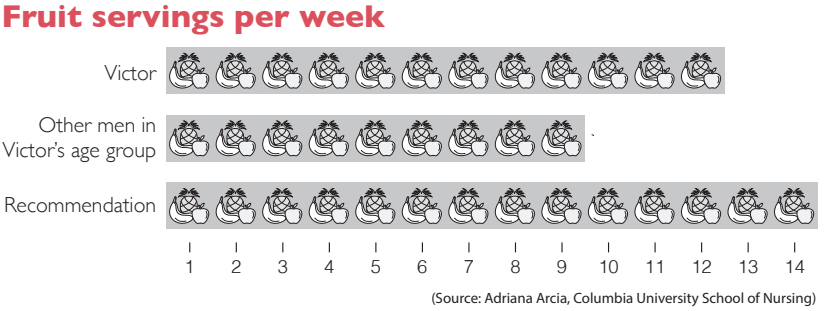
The position of a dot on the vertical axis corresponds to the number of decayed teeth per person. Therefore, the higher up a dot is, the more bad teeth people in that country have, on average.

You've probably detected a pattern: in general, and with some exceptions, the farther to the right a dot is, the higher up it tends to be as well. This is called a positive *correlation* between two metrics: sugar intake is positively correlated with worrisome dental health at the country level. (This chart on its own does not prove that more sugar leads to more decayed teeth, but we'll get to that soon.) Correlations can also be negative; for instance,

the more educated countries are, the smaller the percentage of poor people they usually have.

The scatter plot is a kind of chart that is almost as old as those we all learn to read in elementary school, such as the bar graph, the line graph, and the pie chart. Still, roughly 4 out of 10 people in the survey (37%) couldn't interpret it correctly. This may have to do in part with how the questions in the survey were posed or some other factors, but it still suggests to me that a large portion of the population struggles to read charts that are commonplace in science and that are also becoming common in the news media.

And it's not just scatter plots. It also happens with charts that, at least at first glance, look easy to read. A group of researchers from Columbia University showed the following pictorial chart to more than 100 people:[12]



The chart reveals that "Victor," an imaginary fellow, is consuming more fruit servings per week than other men of his age, but fewer than the recommended 14 servings per week.

What the chart is intended to say is: "Victor is currently eating 12 servings *of any kind of fruit* every week. He's eating more than the average man in his age group, but 12 servings aren't enough. He should be eating 14."

Many participants read the chart too literally. They thought that Victor needed to eat the exact same amount *and kinds* of fruits portrayed in the

chart 14 times every week! A participant even complained, "But a whole pineapple?" The results were similar if the icon used to represent "fruit serving" was a single apple. In that case, one participant complained about the "monotony" of eating the same fruit every day.

Charts are still seductive and persuasive, whether or not many people are able to read them correctly. In 2014, a team of researchers from New York University conducted several experiments to measure how persuasive charts are in comparison with textual information.[13] They wanted to see whether three charts—about the corporate income tax, incarceration rates, and the reasons children play video games—modified people's opinions. For instance, in the case of video games, the goal was to show participants that, contrary to some messages in the media, children don't play video games because they enjoy violence, but because they want to relax, let their imaginations fly, or socialize with friends.

Many participants' minds changed because of the charts, particularly if they didn't have strong preexisting opinions about the charts' topics. The authors conjectured that this happened "partially due to the increased sense of objectivity" that "evidence supported by numbers carries."

Studies like this have limitations, as their authors themselves acknowledged. For instance, it's hard to tell what exactly participants found persuasive: Was it the visual representation of the numbers or the numbers themselves? As the saying goes, more research is needed, but the tentative evidence we have suggests that many of us are cajoled by the mere presence of numbers and charts in the media we consume, no matter whether we can interpret them well.

The persuasiveness of charts has consequences. Very often, charts lie to us because we are prone to lying to ourselves. We humans employ numbers and charts to reinforce our opinions and prejudices, a psychological propensity called the confirmation bias.[14]

Republican congressman Steve King, a strong proponent of strict limits to immigration, posted on Twitter in February 2018:

> Illegal immigrants are doing what Americans are reluctant to do. We import young men from cultures with 16.74 times the violent death rate as the U.S. Congress has to KNOW more Americans will die as a result.[15]

King also added a table. The United States isn't shown, but it's in the 85th position, with a violent death rate of around 6 per 100,000 people:

Violent death rate per 100,000 people

Rank	Country	Rate	Rank	Country	Rate
1	El Salvador	93	11	Panama	34
2	Guatemala	71	12	D.R. Congo	31
3	Venezuela	47	13	Brazil	31
4	Trinidad-Tobago	43	14	South Africa	29
5	Belize	43	15	Mexico	27
6	Lesotho	42	16	Jamaica	27
7	Colombia	37	17	Guyana	26
8	Honduras	36	18	Rwanda	24
9	Swaziland	36	19	Nigeria	21
10	Haiti	35	20	Uganda	20

King was fooled by his own data and chart and, as a result, he likely also fooled some of his constituents and followers. These countries are very violent, yes, but you cannot infer *from the chart alone* that the people moving from them to the United States have violent inclinations. The opposite may be true! It may well be that immigrants and refugees from dangerous countries are the meek and the peaceful, fleeing from societies where they can't work hard and thrive because they're being harassed by criminals.

To give you an anecdotal analogy, an enormous number of Spanish men my age love soccer, bullfighting, Flamenco dance, and the reggaeton song "Despacito." I'm a Spaniard, but I don't like any of those, and neither do any of my closest Spanish friends, who prefer to engage in much dorkier rec-

reations, such as strategic board games and reading comic books, popular-science books, and science fiction. We must always be wary of inferring features of *individuals* based on statistical patterns of *populations*. Scientists call this the ecological fallacy.[16] You'll soon learn more about it.

Charts may lie in multiple ways: by displaying the wrong data, by including an inappropriate amount of data, by being badly designed—or, even if they are professionally done, they end up lying to us because we read too much into them or see in them what we want to believe. At the same time, charts—good and bad—are everywhere, and they can be very persuasive.

This combination of factors may lead to a perfect storm of misinformation and disinformation. We all need to turn into attentive and informed chart readers. We must become more *graphicate*.

Geographer William G. V. Balchin coined the term "graphicacy" in the 1950s. In a 1972 address to the annual conference of the Geographical Association, he explained its meaning. If literacy, said Balchin, is the ability to read and write, articulacy is the ability to speak well, and numeracy the ability to manipulate numerical evidence, then graphicacy is the ability to interpret visuals.[17]

The term "graphicacy" has appeared in numerous publications since then. Two decades ago, cartographer Mark Monmonier, author of the classic book *How to Lie with Maps*, wrote that any educated adult should possess a good level of not just literacy and articulacy but also numeracy and graphicacy.[18]

This is even truer now. Public debates in modern societies are driven by statistics, and by charts, which are the visual depiction of those statistics. To participate in those discussions as informed citizens, we must know how to decode—and use—them. By becoming a better chart reader, you may also become a better chart designer. Making charts isn't magic. You can create them with programs installed on common personal computers or available

on the web, such as Sheets (Google), Excel (Microsoft), Numbers (Apple), open-source alternatives such as LibreOffice, and many others.[19]

By now you've seen that charts can indeed lie. I hope to prove to you, however, that by the end of this book you'll be able to not only spot the lies but also recognize the truths in good charts. Charts, if designed and interpreted properly, can indeed make us smarter and inform conversations. I invite you to open your eyes to their wondrous truths.