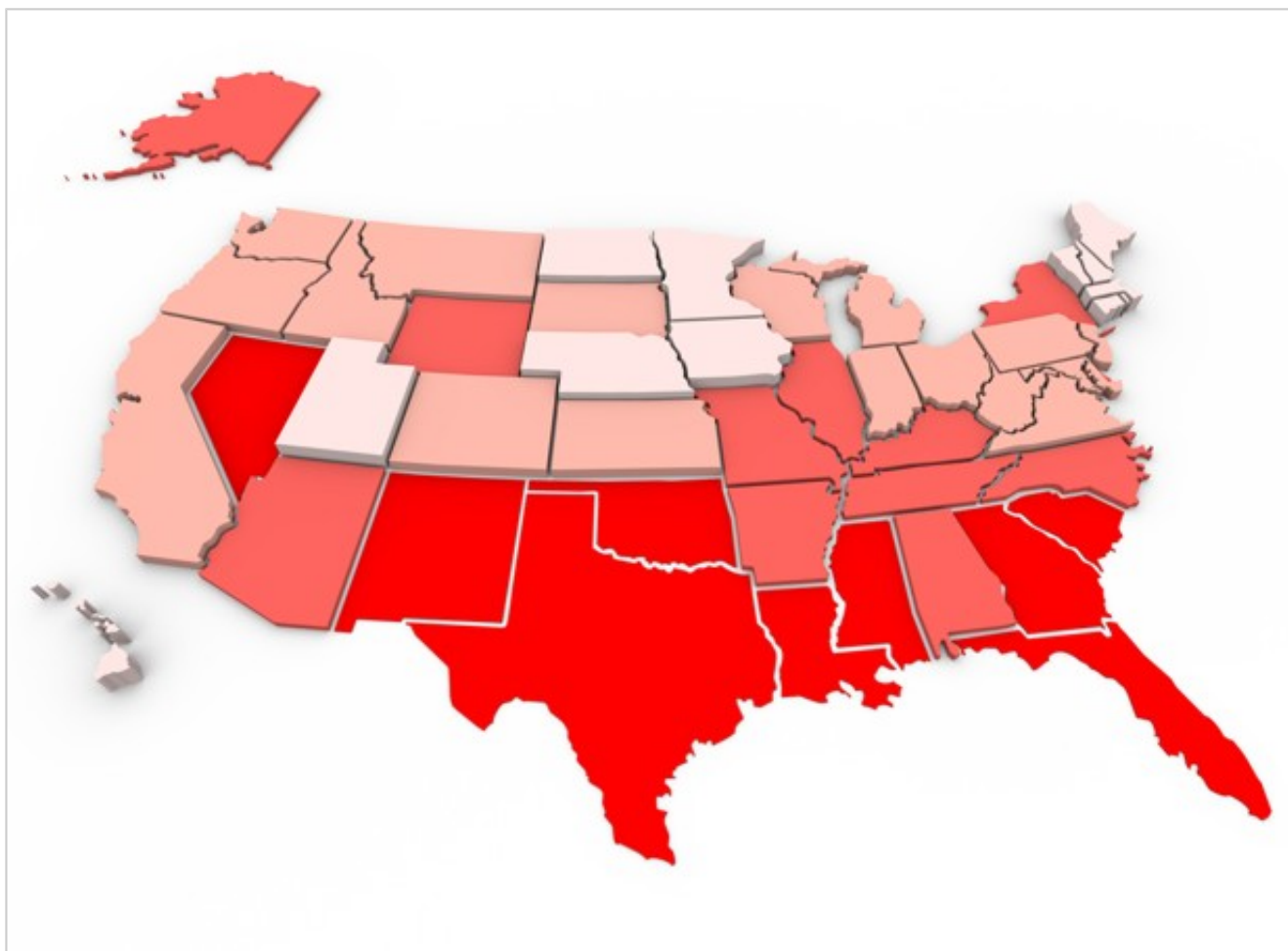From The Atlantic

CITYLAB

MAPS

When Maps Lie

Tips from a geographer on how to avoid being fooled.

ANDREW WISEMAN | 🐦 @wisemana | Jun 25, 2015 | 💬 17 Comments

What does this map even mean? (iQoncept / Shutterstock.com)

Maps are big these days. Blogs and news sites (including this one) frequently post maps and those maps often go viral—40 maps that explain the world, the favorite TV shows of each U.S. state, and so on. They're all over Facebook, Twitter, and Tumblr, and news organizations are understandably capitalizing on the power that maps clearly have in digital space: they can visualize a lot of data quickly and effectively. But they can also visualize a lot of data

inaccurately and misleadingly.

A map is not just a picture—it's also the data behind the map, the methodology used to collect and parse that data, the people doing that work, the choices made in terms of visualization and the software used to make them. A map is also a representation of the world, which in some ways must always be a little inaccurate—most maps, after all, show the roughly spherical world on a flat surface. Certain things are always left off or highlighted while others are altered, as no map can show everything at once. All of those choices and biases, conscious or not, can have important effects on the map itself. We may be looking at something inaccurate, misleading, or incorrect without realizing it.

Americans are taught from an early age to analyze and understand the meaning and manipulation of words, but they are rarely taught the same skills about maps.

As Mark Monmonier writes in the fantastic book *How to Lie With Maps*, Americans are taught from an early age to analyze and understand the meaning and manipulation of words, such as advertising, political campaigns, news and the like (to be "cautious consumers of words" as he puts it) but they are rarely taught the same skills about maps.

Education about using maps (and geography as a whole) is not thorough or common in U.S. schools. The high school Advanced Placement exams for human geography only started being offered in 2001*, for example, and many top private universities do not offer geography as a subject. Harvard dropped it in 1948, which some academics blame for kicking off a decrease in the learning of geography across the country.

Numerous studies report that the vast majority of Americans lack geographic literacy and are unable to find places like Afghanistan or Iraq on a map, let alone understand more complex spatial relationships about them—where are things, why are they there, how does that influence other things? (Harvard, to its credit, formed a Center for Geographic Analysis in 2006.) If they think of it at all, many Americans think geography is just memorizing a list of state capitals or looking at pictures of cool animals in *National Geographic*.

It's no surprise then that people often assume maps are accurate, because it's so often unclear how they are made—maps are "arcane images afforded undue respect and credibility" that are "entrusted to a priesthood of technically competent designers and drafters," as Monmonier puts it. Almost everybody can write, but not everyone can make a map.

At the same time, the use of geographic information systems (GIS) has exploded as computers and software get more powerful and less expensive. New web mapping tools and the availability of data are democratizing cartography, allowing almost anyone to attempt mapmaking—something that was formerly possible only for experts or users of specialized software. That means many more people are creating their own maps, which is surely a good thing, but it also means that there are many more inaccurate, incorrect maps out there—either by design (to push viral or push a viewpoint) or because the creators don't fully understand what they're doing.

Maps are still fun, even the inaccurate ones. But there are a few steps you can take and concepts you can keep in mind to avoid being fooled by a map.

Don't trust a title

A good mapmaker should correctly and succinctly explain what their map actually shows, rather than making grand claims. One recent example is a *New York Times* article and interactive map is called "The Best and Worst Places to Grow Up." That might seem self-explanatory, but the data behind this particular map really only shows that people make more or less money depending on where they grow up. That's still a very interesting finding, but earning more or less money does not necessarily mean that one place is better or worse to grow up in than another, something that is hard to define and would surely

include many variables: maybe a place has a lower cost of living, or better schools, better health care, more recreation options, and so on—and [an accompanying article](#) actually makes that argument.
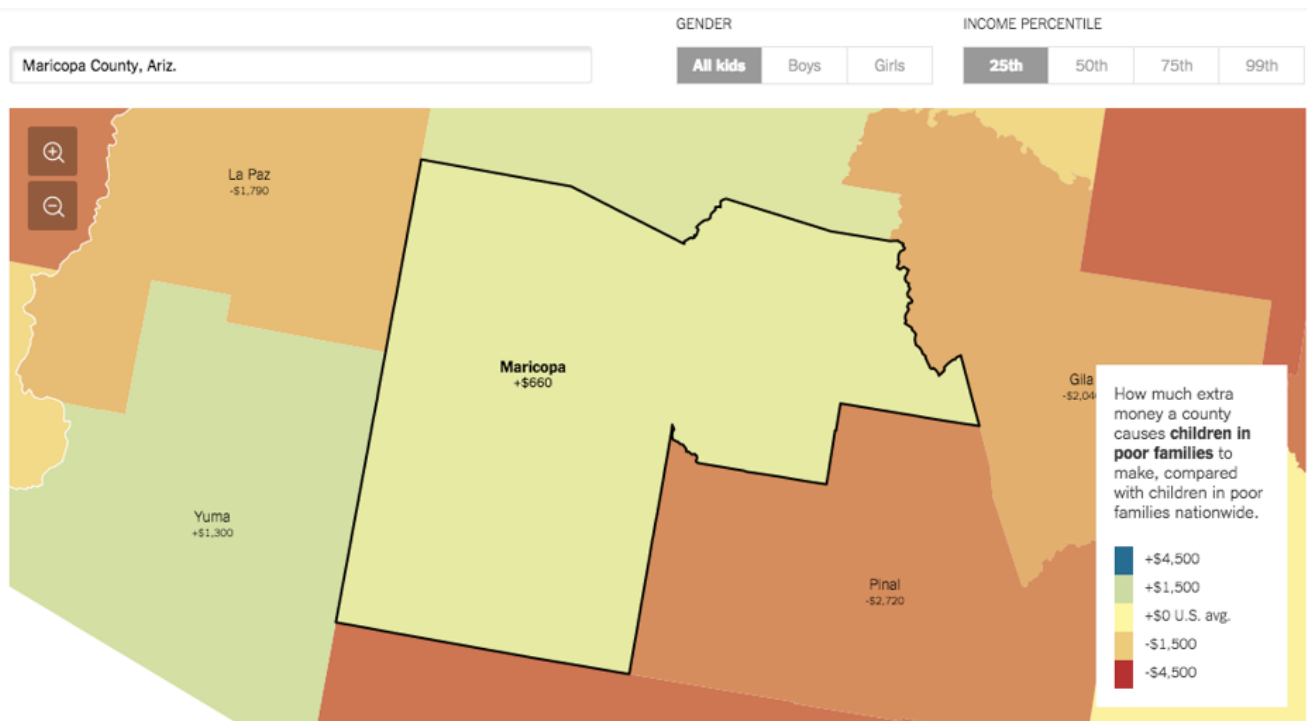
Maricopa County is about average for income mobility for children in poor families. It is better than about 48 percent of counties.

(The New York Times)

The data is very attractively presented, and impressively the map actually changes based on where you are reading from, but the title does not really show what it purports to show. If you just look at the map and think a place is good or bad, you are missing the whole story. Are Atlanta or Denver bad places to grow up? Are the best places Wyoming, Utah, the Dakotas, and rural Minnesota and Iowa? That depends on what you consider best and worst.

A red flag with map titles should also be Upworthy-esque words like amazing, incredible, gorgeous, and so on, or ones that claim to explain everything —"[this incredible map shows half of U.S. output is generated by a few cities,](#)" for example. Looking more into the data and the way the maps are made, those amazing statements often turn out to be pretty boring.

That particular map, which shows gross domestic product (GDP), seems like it suggests that people in cities are more productive than people in rural areas,

but really it shows that 50 percent of the GDP of the country is generated by 50 percent of the population, since far more people live in cities than in rural areas: hardly an amazing story. A more interesting map would be GDP by population—are certain places more productive than others? Why? (I like to tell my students that a good map may raise a question rather than answer one.) A similar example is a map that shows that 92 percent of U.S. AIDS cases happen in 25 counties—but another article notes most of the U.S. population lives in those 25 counties, so that's probably to be expected. What would be more interesting is to map out the variation in counties.

A related issue with titles is how they interpret maps, and speaks to why a good map should include the source of the data on the map itself (more on that shortly). This map, for example, the Corporate States of America, claims to show the "most recognizable" company in each state. But what does most recognizable mean? The author of the map explains, to his credit, that it's just ones he thought best represented the state rather than something with actual data behind it, like percentage of people who are familiar with a company, its market value, or something else. Florida doesn't need to be ashamed that Hooters is its representative on this map, for example. Another popular map of the most distinctive band in a state (i.e., bands that are more popular than in other states), which had actual data behind it, came to be shared as "your state's favorite band"—two different things.

Often these "amazing" maps are just something that someone made up. Of course, that doesn't mean they can't also be nice-looking and fun. (In fact, they should be nice-looking. Don't share ugly maps.)

The source is important

The source of the data on a map is one of the first things you should look for—where did this information come from? Is that a trustworthy source? Is it recent? Can you look at it yourself?

If the source is not listed, be suspicious—but be suspicious even if it is. For example, in a recent viral map of the most hated college basketball teams, the cartographer includes a bit of metadata listing himself as the creator, which is helpful. However, the data itself is from a two-question Google Docs survey

that was posted on Reddit's college basketball board, hardly a representative group of Americans. The survey did require respondents to enter where they are from, but people could just say the wrong thing on purpose—someone who doesn't like Duke could say they were from North Carolina to skew the data.

In this case, a more accurate description would be "Most hated college basketball teams among self-selecting Reddit college basketball board users"—and those kind of explanations should really be on the map, rather than making the reader find them. Then again, that might not have helped it get picked up by newspapers published where the supposedly hated teams are.

Can you trust the data in this map? Look closer. (via Playboy)

This is a silly, innocuous example (who cares if a college basketball hate map is inaccurate?) but the same thing can happen with any kind of data—politics, health, religion and so on.

Source data that's readily available can also let you see if something is biased or even just flat made up. An amusing example of this is "Most Common Cause of Death That THEY Keep Covering Up" by state, including things like trolls, a Russian invasion, lake monsters, and lactose intolerance. This is clearly a parody of viral maps that show things like most popular film or job or food in a state—or weightier subjects like cause of death—which can all be misleading or misinterpreted and can easily hide variations between states.

Viral maps like "signature food from each state" and "most popular television show set in each state" (or is it the most significant show? Again, title is important) also come from Reddit users, which is probably why things like "meth cookies" appear as Arkansas's signature food. Maps like these are fun, to be sure, but drawing conclusions based on them wouldn't be the best idea. Presumably there aren't many restaurants in Little Rock with meth on the menu.

Heat and density maps can confuse

The webcomic XKCD skewers this type amusingly, with maps showing the density of three unrelated things: people who visit a website, people who subscribe to *Martha Stewart Living*, and people who consume furry pornography; and the density of all three is the same.
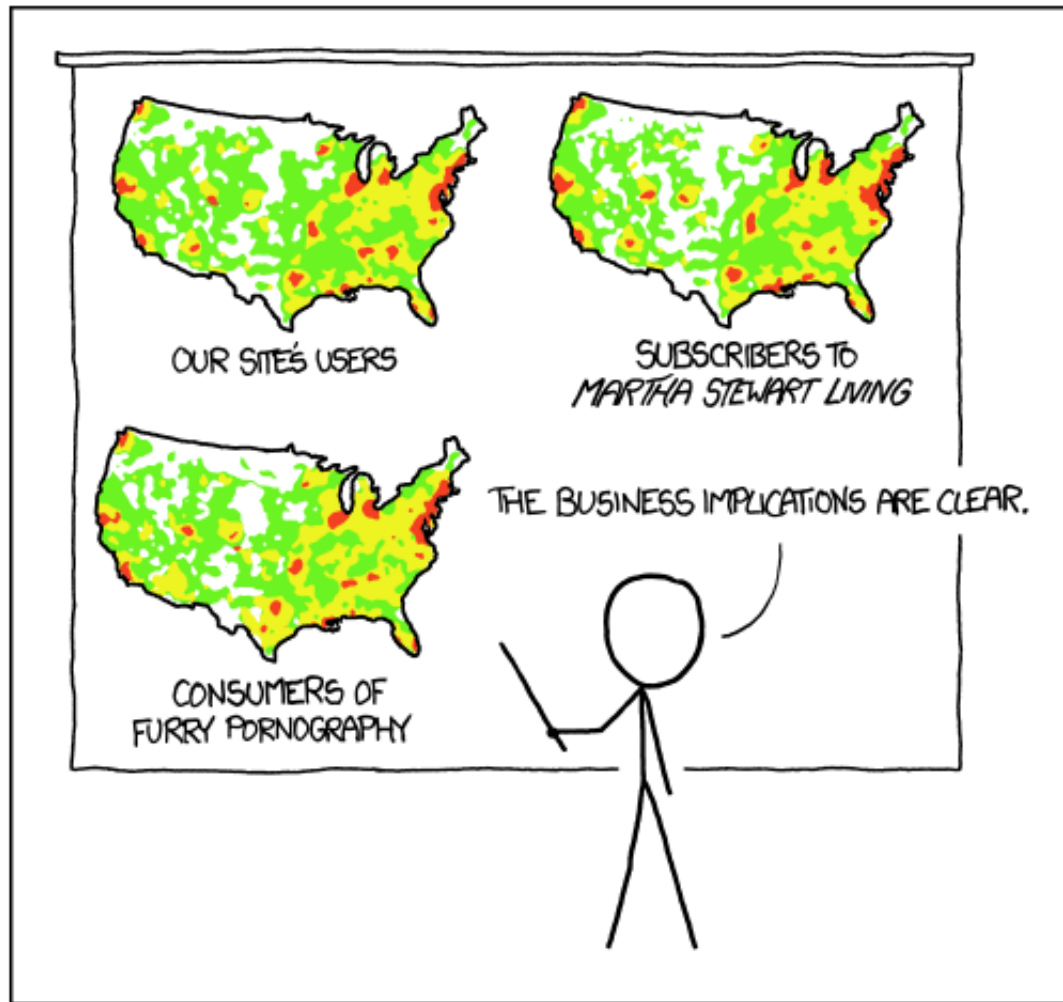
(XKCD)

The joke is that each map is just a population map, and that more people tend to do more things. The GDP map mentioned above is also really just a population map—there is more GDP in places where there are more people.

Another map that says it shows every tweet ever falls into this same category: while a truly beautiful map, it is basically showing that people tweet more where there are more people, as the geographers at the great site Floating Sheep point out. Maybe just showing there is a lot of something in a place is enough, but if you're trying to map something other than just population, or draw useful conclusions, a heat map needs to do something else, like generalize the data by population or some other useful factor.

A map of more or less tweets per capita would be interesting, for example: that way you can see if the thing you are looking at is actually more or less common. (Eric Fischer, who made that tweet map, has some other interesting examples of that, such as looking at where tourists versus locals take geo-tagged pictures, and some that show more tweets in areas that don't necessarily have more people, like east of the Anacostia River in Washington, D.C.)

What is the cartographer trying to show? What are they trying to hide?

Maps are representations of the world where certain things are highlighted and certain things removed. This often is a good thing: if there's a map for a specific purpose, it should only include things that are helpful (road types, names and cities on a road atlas, for example) or maybe for clarity's sake you reduce how jagged or complicated something is, like the lines on a subway map. But it can be hard to tell when something has been manipulated or removed, so it's important to think about the choices the cartographer could have made.

A simple example of this are maps that advertise or try to convince you of something: does that real estate map accurately show distance? Does it leave off the coal power plant or stinky poultry farm nearby? (Think of *The Simpsons* and the monorail—it put Brockway, Ogdenville and North Haverbrook on the map!)

Color and size can also be used to highlight or hide something. If a mapmaker wants to show something is bad, they'll make it red; if a they want to highlight something, they might make it big and brightly colored and make the things they don't want people to notice small and gray (or take them off altogether.)

If they want to show that there is a lot of something (potholes in New Jersey, for example) they might make the symbols big and bold. Those big symbols can hide a lot of variation within the data—are these potholes a foot wide, or only an inch? Are there more potholes in New Jersey than other places? It's impossible to tell much other than "wow, there are a lot of potholes." (And the map doesn't list the source of the data, either.)

This can also be inadvertent: people might assume something is bad if it's colored red or differently, even if that's not the case—and red also tends to draw the eye away from more subtle colors. One confusing example of this is an Associated Press map showing population growth by state:

U.S. population growth slows

The Census Bureau put the nation's population on April 1 at 308,745,538. The growth rate for the past decade was 9.7 percent, a slower pace than the 13.2 percent increase from 1990 to 2000.

Population growth, in thousands

■ <0%   ■ <5%   ■ 5-10   ■ 10-15   ■ 15-20   ■ 20+
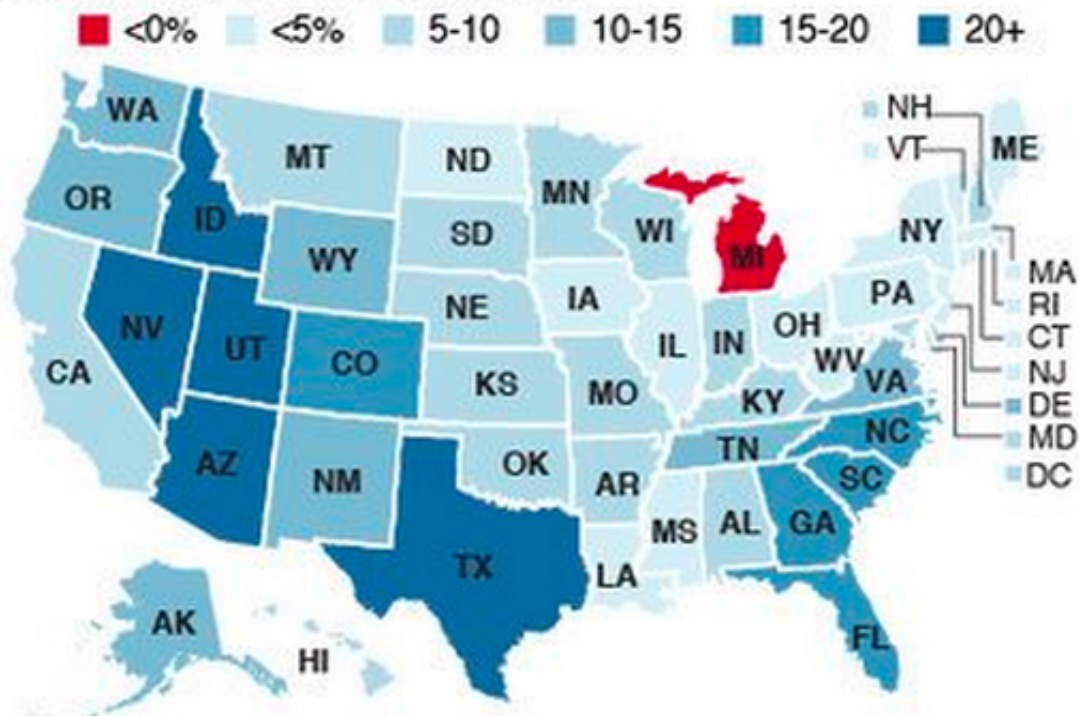
SOURCE: U.S. Census Bureau   AP

Michigan is red, so that means bad, right? But the legend is confusing for many reasons. There are three different kinds of ranges here: two "less than" signs, three numerical ranges like 5-10, and then one "plus" sign. That variety of how the numbers are displayed can make it hard to understand. And there's overlap, too: is 10 percent in the 5-10 range or 10-15? More useful ranges would be something like 0-4.9, 5-9.9, and so on. Less than 5 percent would technically include less than 0 percent, too. And finally, the note says the data is in thousands, but then lists percent. The population of Texas went up a lot, but it didn't increase by 20,000 percent. All of these make a good argument why legends and colors should not be overlooked. If red is going to be bad,
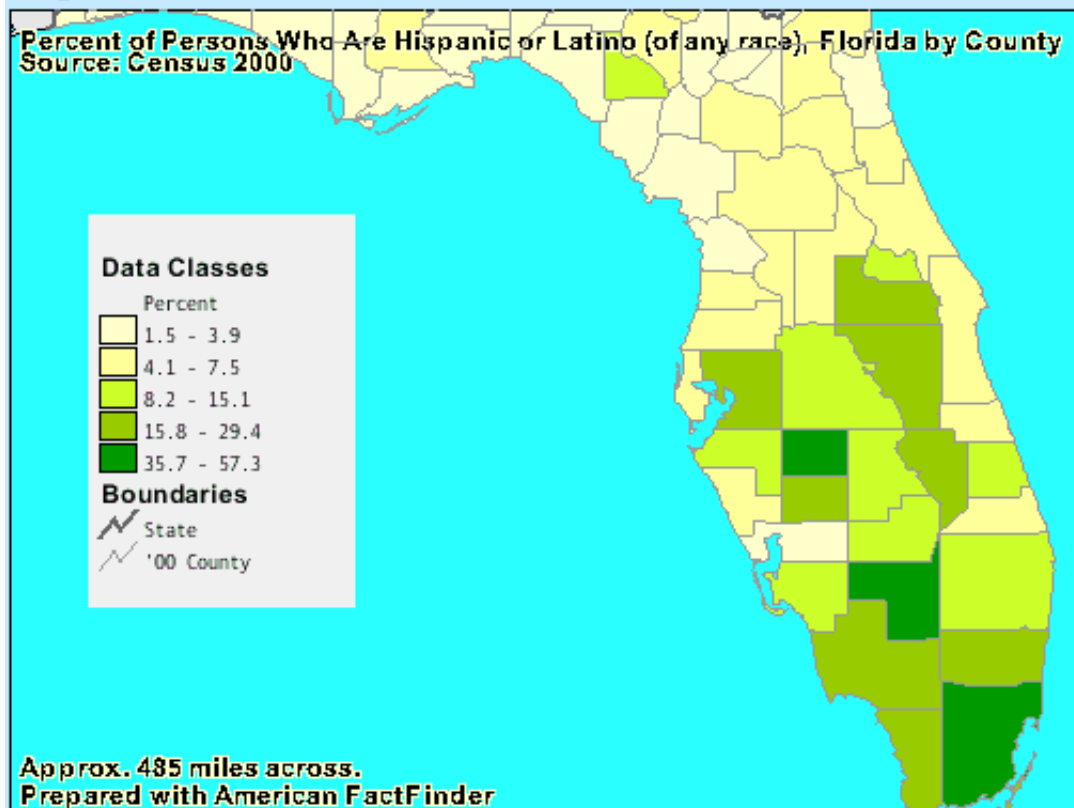
make it clear why.*

A related feature to look out for is how the data is structured—is it the values of data itself, such as total unemployment or unemployment rate, or something else derived from that data? The farther away you get from the actual values, the more suspicious you should be. Are you looking at the rate of change, or even the *decrease* in the rate of change? If someone is highlighting the second derivate, they might be obscuring the fact that things are bad overall. (There is an important corollary to that, however: normalizing data by population, which we'll get to shortly.)
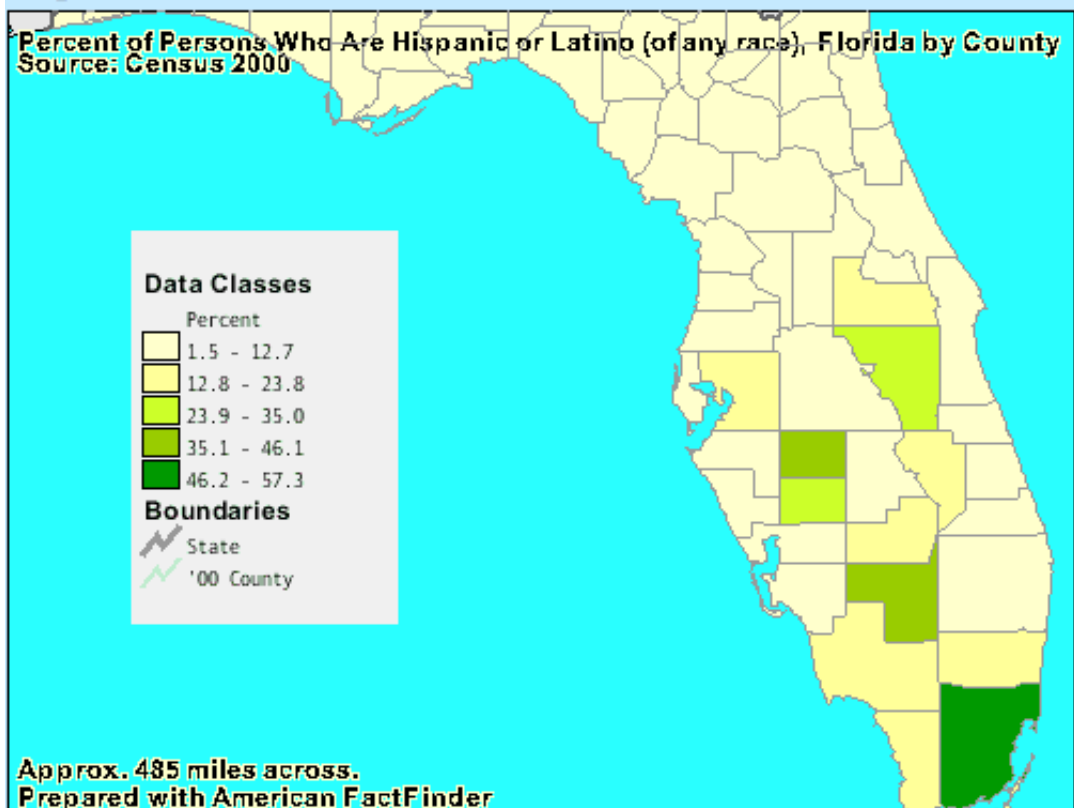
The way the data is distributed is important

If a map is showing the variation of something using different colors, say population by state, where lighter means fewer and darker means more (what's called a choropleth map) take special note of the arrangement of the values into different classes. This distribution can be misleading or obscure reality.

It's often normal to make a map that divides data into equal intervals—population by county from 1-10,000, 10,0001-20,000, 20,001-30,000, and so on. However, sometimes that logical organization isn't the best way to present the information. If the mapmaker wants to highlight or hide something, it's easy to manipulate the class breaks (those dividing lines between groupings) to put all the high or low values into one large class while making all the other values into their own classes. In this way, you are emphasizing part of the data while hiding another part of it. A good example of this are these maps of the Hispanic population in Florida:

Map 1

Percent of Persons Who Are Hispanic or Latino (of any race), Florida by County
Source: Census 2000

Data Classes

Percent

- 1.5 - 3.9
- 4.1 - 7.5
- 8.2 - 15.1
- 15.8 - 29.4
- 35.7 - 57.3

Boundaries

- State
- '00 County

Approx. 485 miles across.
Prepared with American FactFinder

Map 2

Percent of Persons Who Are Hispanic or Latino (of any race), Florida by County
Source: Census 2000

Data Classes

Percent

- 1.5 - 12.7
- 12.8 - 23.8
- 23.9 - 35.0
- 35.1 - 46.1
- 46.2 - 57.3

Boundaries

- State
- '00 County

Approx. 485 miles across.
Prepared with American FactFinder

Map 3

Percent of Persons Who Are Hispanic or Latino (of any race), Florida by County
Source: Census 2000

Data Classes
Percent
1.5 - 2.7
2.8 - 4.9
5.0 - 9.3
9.5 - 57.3
Boundaries
State
'00 County
Features
Major Road
Railroad
Jail

Items in gray text
are not visible
at this zoom level

Approx. 485 miles across.
Prepared with American FactFinder

While they all use the exact same data, the different choices for classes in each map make the number of Hispanic people appear vastly different.

And be very suspicious if the map doesn't actually say what those divisions are, such as if it just includes the colors with "high" and "low," or leaves off the legend altogether. This is another warning that the source might be not trustworthy or just something somebody made up.

The groupings of the data can also be arbitrary: if a mapmaker wants to highlight something they think is positive, such as a decline in unemployment, they might make a map showing counties where unemployment went down and by how much, with a progression of darker colors showing how much it dropped, while counties where unemployment went up were all the same, unobtrusive color (like a light gray.) Maybe the increases in unemployment in those places were actually much greater than those highlighted drops, you wouldn't know that from looking at the map. A map showing the same data with a scale including both increases and decreases would give the reader a much different understanding of the data. (Yet another reason why the source of the data is important.)

Choropleth maps can be tricky

Choropleth maps, which I mentioned earlier, can show data very effectively, but they can cause a lot of problems if used improperly. One of those is what's called the "modifiable areal unit problem," which arises because states, counties, census tracts and so on are not uniform in size or population, and population is not evenly distributed within them. This means that clusters or patterns in the real world can be obscured by the boundaries used to divide the data. You may need to dig into the data to understand those patterns.

This series of choropleth maps of the 2012 U.S. presidential election shows this problem very well. The first map makes it appears that the presidential race was pretty close or that the country is very divided (red states versus blue states.) However, coloring a state red or blue based on who won obscures the margin of victory, the total number of votes, and any variation within that state.

If the cartographer chose to be more granular and show the same data by county, a logical choice, it appears that there was much more support for Romney than Obama, despite the fact that Obama received 5 million more votes and won the Electoral College vote handily. We know this because each county is not the same population—Obama overwhelmingly won the small urban counties with very large populations while Romney won many more sparsely populated rural counties. Even if you show the data by degree of support, from blue to purple to red (the "purple state" maps that were popular after the election) you are still missing that urban counties are often quite small in area but large in population.

The modifiable areal unit problem is also why it can be helpful to map something like density (say, population density) rather than raw numbers, since small counties or census tracts might be very dense.

One solution to this is a cartogram, a graphic where the areal units (in this case, counties) are sized based on their population, so counties with more people appear much bigger.
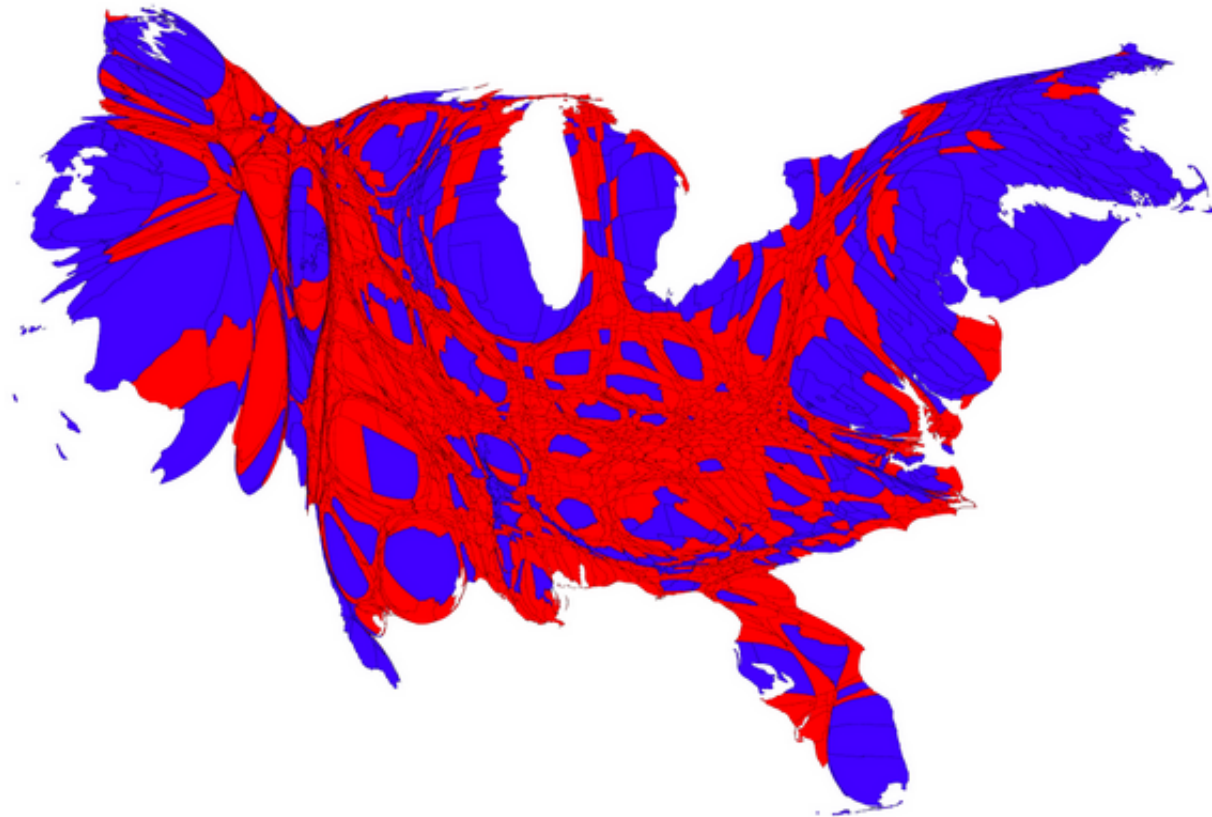
Examples of cartograms of 2012 U.S. presidential elections results, courtesy Mark
Newman at the University of Michigan.

That's somewhat helpful, but can be quite confusing or disorienting, even if we
already know what those states or counties are. Adding in a gradation in color
showing the margin of victory in each county helps even more:
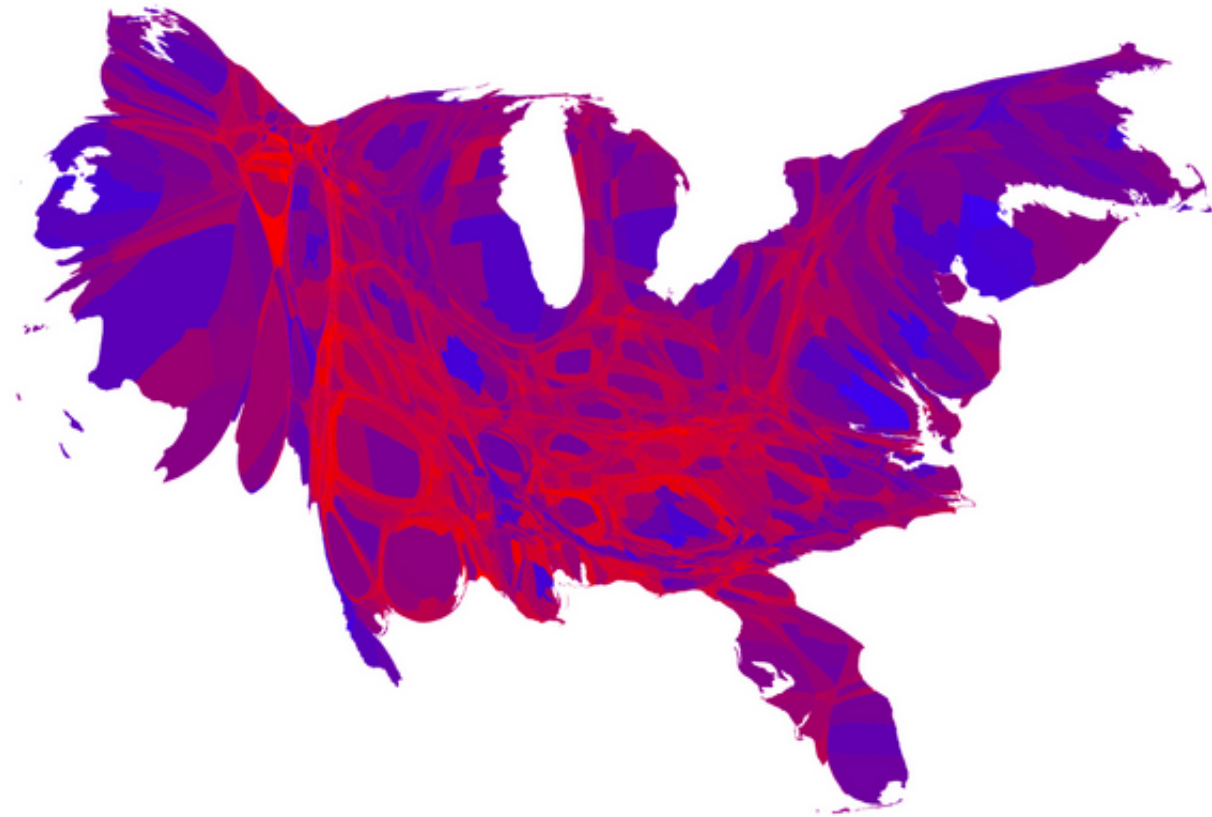
That cartogram shows that most counties with big populations went overwhelmingly for Obama, while the rest of the counties were mostly purple.

This all may seem obvious, but that's because we are generally familiar with the U.S. election results. Information we are not as familiar with can seem to show patterns than don't exist or are incorrect based on how it is displayed or organized. That's why it's important that the cartographer understands their data and what they are mapping, and why the data should be available so others can look into it.

A bad example of that understanding was a now-updated article from FiveThirtyEight on kidnappings in Nigeria. Originally, the author mapped data on reports of kidnappings without fully understanding how it was collected and organized and drew a lot of incorrect conclusions from it, resulting in a lengthy retraction. For example, data that hadn't been mapped to a specific town in Nigeria was placed in the exact center of the country, so a choropleth map made it appear that there was a massive wave of kidnapping in the district that contained that center point. In addition, the article stated it was a map of kidnappings, when in fact it was a map of *reports* of kidnapping, a subtle but

important difference and another example of why an accurate title and description is important.

Base data is important, too

Even the base data used for a map can have important consequences—boundaries, locations and so on. For example, Google Maps changes boundaries based on where you are: the boundaries of China, India and Pakistan are quite different in each country because of each country's conflicting land claims. The company does the same thing for Crimea in Ukraine, Russia and the rest of the world, subtly adapting to (or even changing) perceptions for people in those countries. You might not realize something is off because you've always seen it the same way, while someone somewhere else may have seen different maps their whole life.

Even the way a map is shown can be important. The infamous Mercator projection being a notable example of exaggerating the size of Europe and North America while making Africa and South America appear much smaller than they really are. A funny example of this comes from the *West Wing*.

Base data can have systematic flaws too: the viral map of "every river in the United States" is pretty to look at, but if you look a littler closer, the data has some problems.

A map of every river in the U.S.? Not quite.

For one, there are some straight lines and rectangles in Texas and Oklahoma where the density of rivers changes—surely not a natural feature, but an artifact in the data where for whatever reason, more rivers were mapped in some areas than others. Base data used on a map might have other flaws like this that aren't apparent at first glance.

But maps are still good

None of this means that all maps are bad, or that we should always be suspicious of them, or that only experts should make maps. Maps are inherently interesting and fun (so is geography!), but a little bit of thought and increased awareness of how they can manipulate or obscure is a good thing, too. Just like advertisements and political campaigns, we shouldn't trust maps (or the data behind them) inherently, but they can still be powerful, interesting and amusing.

CORRECTION: This story originally misstated the year that the high school Advanced Placement exams for human geography began. The exams were first offered in 2001, not 2010. Also, a paragraph that describes issues with the legend in an Associated Press population change map has been updated for clarity.

Top image: [iQoncept](#) / [Shutterstock.com](#)

About the Author

Andrew Wiseman is a geographer with the USAID's Office of Transition Initiatives and teaches cartography and geographic information systems at the George Washington University in Washington, D.C.

ALL POSTS | 🐦 @wisemana