



## Data & Infographics: Crash Course Navigating Digital Information #8

Crash Course: Navigating Digital Information

<https://youtube.com/watch?v=OiND50qfCek>

<https://nerdfighteria.info/v/OiND50qfCek>

Hi, I'm John Green. This is Crash Course Navigating Digital Information.

So, what would you say if I told you that 90% of people polled say that they love Crash Course, and think we offer consistently reliable and accurate information on the most important educational topics? You might say, "hold on. I've seen the comments. That can't be true." And, you'd be kind of right, but I would also be kind of right. Because, I did do that survey, and 90% of people did agree with those positive statements about Crash Course. But, I surveyed 10 people who work on Crash Course. It would've been 100%, but Stan said, "It this for a bit? I'm not participating."

Anyway, whether it's 4 out of 5 dentists or 9 out of 10 Crash Course viewers, source and context can make all the difference. We like to think of data as just being cold, hard facts, but, as we've already learned in this series, there is no single magical way to get at the singular truth. We have to place everything in its context, even statistics. In fact, especially statistics.

[Intro]

Ok, so, data is raw quantitative or qualitative information, like facts and figures, survey results, or even conversations. Data can be derived from observation, experimentation, investigation, or all three. It provides detailed and descriptive information about the world around us. The number of teens who use Snapchat, the rate at which millennials move in or out of a neighborhood, the average temperature of your living room, those are all data points.

And, data is a really powerful form of evidence, because it can be absorbed quickly and easily. Like, we often consume it as numbers, like statistics, or as visual representations, like charts and infographics. But, as Mark Twain once famously noted, "There are three kinds of lies: lies, damned lies, and statistics."

Statistics can be extraordinarily helpful for understanding the world around us, but, because statistics can seem neutral and irrefutable, they can be used to profoundly deceive us as well. The truth is, neither data nor interpretations of it are neutral. Humans gather, interpret, and present data, and we are flawed, complex, and decidedly un-neutral.

Unfortunately, we often take data at face value. Just like with photos and videos, we can get stuck in the "seeing is believing" trap, because we don't all have the know-how to critically evaluate statistics and charts. Like, a Stanford History Education Group study from 2015 bears this out. SHEG developed the MediaWise curriculum that this series is based on. And, they asked 201 middle schoolers to look at this comment on a news article. As you can see, the comment includes healthcare statistics, but doesn't say where they came from. It doesn't provide any biographical information on the commenter, either, but 40% of the students indicated they'd use that data in a research paper. In fact, many cited the statistics as the reason they found the comment credible and useful. The sheer existence of quote-unquote data enhanced its credibility despite there being no real reason to trust that data.

So, whenever we come across data in the wild, we should ask ourselves a couple of questions: Does this data actually support the claim being made, and is the source of this data reliable?

Here's an example when it comes to data relevance. At the 2018 U.S. Open, Serena Williams was penalized for yelling at the umpire and smashing her racket during the game. On the court, she argued that men yell far worse things at umpires and physically express their emotions all the time without being penalized, and a few weeks later, journalist, Glenn Greenwald, cited a *New York Times* story in a tweet: "Now, NYT just released a study of the actual data:

contrary to that narrative, male tennis players are punished at far greater rates for misbehavior, especially the ones relevant to that controversy: verbal abuse, obscenity, and unsportsmanlike conduct"

Well, that sound very authoritative. And also, he linked to a table that showed that far more men have been fined for racket throwing and verbal abuse than women during grand slam tournaments. However, as statistician, Nate Silver, helpfully pointed out, this stat only shows that men are punished more, which could be because they misbehave more. So, all these statistics actually show is the raw number of punishments, not the rate of punishment, despite Greenwald's claims.

To get the rate of punishment, we'd have to divide the number of punishments by how many times men and women misbehave, and that data isn't provided here. So, the data, in the end, does not support Greenwald's tweet at all, making his claim that male tennis players are punished more frequently problematic at best. To be fair, Serena Williams claim is also anecdotal, although, you know, she does watch a lot of tennis.

We also need to investigate whether the source providing the data is reliable, and we can do that through lateral reading. That means opening new tabs to learn more from other sources about who commissioned the research behind data, who conducted the research, and why? And, we also need to know if the source of the information is authoritative, or in a good position to gather that data in the first place.

Like, remember in episode 3 of this series when we talked about the claim that Americans use 500 million straws per day? We couldn't confirm how many straws Americans actually use every day, but we did see that sources across the web cited that statistic, even though we found out that it came from a 2011 report written by a then-nine year old child, Milo Cress. To come up with the figure, he called up straw manufacturers to ask how many straws they made. There's no way of knowing if those manufacturers were telling the truth, or if the group he called is representative of the whole industry. He was 9. He was obviously a very bright and industrious 9 year old, but he was 9! Apologies to all the 9 year olds watching. Thank you for being careful in how you navigate digital information, friends. A more reliable source of such far-reaching information might be a nonpartisan research organization like the Pew Research Center. They're known for reliable, large-scale studies on U.S. trends and demographics.

Once we know who is a source of data is, whether they're authoritative, and why they gathered it, we should ask ourselves what perspective that source may have. They could have a vested interest in the results, like the beauty influencer you follow who's always saying 92% of users of this snail slime facial get glowing skin in 10 days. That study may be accurate, but there also may be a #ad in the caption to quietly let you know that the brand in question is paying them. But, forget about snail slime. Have I told you about Squarespace? We have to take into account when people cite data that helps them make money, including me.

Alright, so once we know more about where our data comes from, it's time to analyze how it's presented. Data visualizations, like charts and graphs and infographics, can be amazing ways of displaying information, because, one, they're fun to look at and, two, the best infographics can take complex subjects and abstract ideas and turn them into something that we understand. Like, I love this one that shows how factual movies "based on a true story" really are. Oh, and also this one on cognitive biases, although I might be cognitively biased towards appreciating a graphic about cognitive biases.



## Data & Infographics: Crash Course Navigating Digital Information #8

Crash Course: Navigating Digital Information

<https://youtube.com/watch?v=OiND50qfCek>

<https://nerdfighteria.info/v/OiND50qfCek>

The great things about data visualization is that it's a creative field, limited only by a designer's imagination. But, of course, with artistic license comes the ability to present data in ways that sacrifice accuracy. It's really quite easy to invent a nice-looking graphic that says whatever you want it to say. So, we need to read them carefully, and make sure there's actually data behind a data visualization.

For instance, look at this chart. It makes a claim that, when guns are legal, lives are saved because gun owners prevent deadly crimes- the "good guys with guns" theory. But, if you read the fine print, the chart acknowledges that statistics are not kept on crime prevention or crimes that never happened, so these figures are not based on real data at all. The chart also says that fewer homicides take place when guns are legal than when they're banned. But, what it doesn't say is where this change would supposedly take place and over what span of time. For instance, homicides went down in Australia after strict gun control legislation was passed; on the other hand, they also went down in the United States as gun ownership increased.

What's clear upon closer inspection is that this graphic, which initially appears to have some pretty dramatic estimates about gun control, is, by its own admission, mostly speculation. To trust a data visualizations, we need to make sure that it is based on real data and that the data is presented fairly.

Let's go to the thought bubble.

Here's a graph that was posted to Twitter by The National Review, a conservative site that often denies the effects of climate change. It uses data from NASA on the average global temperature from 1880 to 2015. It looks like a nearly straight line, with only a slight increase at the end, and the tweet, "The only #climatechange chart you need to see." implies that it once and for all shows that the climate isn't really getting warmer. However, the y-axis of this chart shows negative 10 to 110 degrees, which makes the scale of this data very small. One might say that the chart misleads by zooming out too far. If, for instance, the scale was truncated to show just 55 to 60 degrees, as in this Washington Post graphic using the same data, the change over time looks much more dramatic.

And, the original post also leaves out some much needed context: The entire globe shifting its average temperature by even a couple of degrees over the period shown is extremely unusual, and has an outsized impact on how the climate functions. The first chart does not present the change in this data or its significance in good faith.

On the other hand, data visualization can also be very misleading if it zooms in too much. This chart, produced by the administration of President Barack Obama, shows how a truncated y-axis can create manipulation, not solve it. The data behind this chart on graduation rates is reliable, but by zooming in the scale to show from around 70 to 85%, it makes the change throughout Obama's administration look much more dramatic. Here's what it would look like if you could see the entire scale. The increase in graduation rates suddenly look much less significant.

This follows the proportional ink principle of data visualization. The size of a filled in or inked area should be proportional to the data value it represents.

Thanks thought bubble.

So, a few simple tweaks to how data is presented can really make a big difference in how it's interpreted. Whenever we encounter data visualizations, we need to check that the data is accurate and relevant, that its source is reliable, and that the information is being presented in a way that is honest about the conclusions it draws.

Actually, once you get the hang of sorting the useful, well-designed data visualizations from poorly designed ones, the bad ones can be pretty entertaining. If you'd like to see some exceptionally terrible charts, take a spin through [viz.wtf](http://viz.wtf) or the subreddit "data is ugly". I'm especially fond of this completely indecipherable chart about the Now That's What I Call Music CDs, courtesy of the BBC.

The challenge and opportunity of images is that they're so eye-catching that we sometimes forget that they're created by and for humans who have the ability to manipulate them for their own ends: to make our information of lower quality and thereby make our decisions of lower quality. And, the use of infographics and big data have become even more popular as our attention spans have waned. After all, it's much easier to read a pie chart than an essay or an academic report. Plus, it fits into a tweet.

In summary, whether you're encountering raw data on its own or visual representations of it, it's very important to keep a critical eye out for reliability and misrepresentation.

Thank you for spending several minutes of your waning attention with us. We're going to get deeper into that next time. I'll see you then.

[Outro]

Thank you for watching Crash Course, which is filmed here in Indianapolis, Indiana with the help of all these nice people.

For this series Crash Course has teamed up with MediaWise, a project out of the Poynter Institute that was created with support from Google. The Poynter Institute is a non-profit journalism school. The goal of MediaWise is to teach students how to assess the accuracy of information they encounter online.

The MediaWise curriculum was developed by the Stanford History Education Group based on civic online reasoning research they began in 2015. If you're interested in learning more about MediaWise and fact-checking, you can visit [@MediaWise](https://www.mediawise.org) on Instagram.

Thanks again for watching, and thanks to MediaWise and the Stanford History Education Group for working with us on this project.