Module 4 Video 2: Interpretation in Data Journalism

[00:00:00] Hi. OK. For this video, we are going to talk a little bit more about ethical and equitable communication in terms of your data journalism stories. And we're going to start by looking back at the example that we talked about before. Is smoking cigarettes good for your health? And you will remember that we initially interpreted this data to mean that smoking four cigarettes a day would increase your life by 10 years, which we all know is not actually true, but a very, very easy mistake to make. And during the communication and visualization part of writing your data story, it's essential that you pay very close attention to how you are interpreting the data and the analysis that you're working with.

[00:00:55] It's very, very easy to accidentally misinterpret the data, just like we did here, where it looks very easy to say that smoking is good for your health because it is associated with longer life expectancy. We caught that mistake because it goes against what we believe. But a lot of times the mistakes confirm what we believe, using something that is called confirmation bias. So whenever you're using data, you need to be very careful to check what interpretations you're making of the data and of the results and whether or not they're actually accurate and whose worldview do they represent.

[00:01:39] A really good example that's been in the news quite a lot is the Compas Algorithm. So the Compas Algorithm is a mathematical equation that gives each person a score. And the data that's fed into the Compas Algorithm is some data about people who are in contact with the law and the nature of their crimes. Some background information about the individual and their relationship with law enforcement in the justice system. So the Compas Algorithm takes all this data in and then puts out a score. And that score is sometimes or used to be sometimes used by the judicial system to help decide whether or not a person would be granted bail, granted parole, things like that. And very good investigative journalists. Did some data work with this Compass Algorithm and realized that it was extremely problematic and racist in many ways. So they brought this to the attention of the necessary people. And you would have seen it probably in widely reported on in the media. It's a great example of really good investigative data journalism.

[00:03:04] Now, part of why this data journalism was necessary was because the people using the data were misinterpreting the Compass Score. They were saying that the Compass Algorithm gives you a score that predicts how likely it is that this person will reoffend. But that's an interpretation of a number and actually an incorrect interpretation of the number. What the Compass Algorithm actually gives you, is not a score that predicts how likely it is that this person will reoffend. It's a score that predicts how likely it is that this person might be in contact with the police, again, be arrested by those police and not have the money for immediate bail release.

[00:03:48] So you can see that even if the data that was fed into the Compas Algorithm was not already extremely biased, even if we were somehow able to get unbiased data, the score that came out of a Compass Algorithm needs to be interpreted extremely carefully and accurately. And once you interpret it correctly, it's very easy to see the problems that are going to underlie that data point, just as we saw the problems with misinterpreting the smoking data.

[00:04:25] So any time that you as a journalist are taking your own results or someone else's results and moving them over into a story, you're taking numbers and embedding them with meaning and a narrative to explain them. And so it's your responsibility to make sure that you really get those numbers correct and that you only interpret to the extent that you can based on what methodology was used to collect the data and analyze the data. And whose culture and worldview were embedded in the data.

[00:05:00] So, for example can we take a study and say, eighty five percent of our respondents reported feeling happier on the days when they had eaten breakfast. Is that an interpretation? That is acceptable. And how would you know? It's likely that this is a solid interpretation because it only talks about the respondents in the study. Eighty five percent of our respondents. So it doesn't extrapolate from the people who contributed data. It just talks about those people. And it just says what they reported, reported feeling happier. So this interpretation uses words. That probably accurately embed an appropriate interpretation into this headline, Eating Breakfast Will Make You Happier is probably a much more appealing headline or narrative from the perspective

of a journalist than eighty five percent of our respondents reported feeling happier on the days when they'd eat breakfast. But you can see eating breakfast will make you happier is also an interpretation, but it makes causal claims. Instead of saying eighty five percent of our respondents had this experience. It's saying this thing that our respondents did caused something in those respondents. So this is the difference between causation and correlation, which I know you've heard about over and over and over. That correlation is not causation. And it's really, really challenging as a data journalist to deal with because very, very few data projects are actually able to demonstrate causality.

[00:06:52] But when people come to a data story, they're not really interested in hearing what happens. They want to know why things are happening. And when you get into trying to explain why things are happening, that's causal analysis. So if you want to craft a data journalism story that employs causal languaging, like eating breakfast will make you happier rather than our respondents felt happier. The difference between those two sentences, essentially one is talking about an association and one is talking about a causal effect. And in our optional resources section in this week's module, I've given you some links to places where you can look at difference in causal words and association words and also a little quiz about the same thing. And you as a data journalist have to find that sweet spot where you're telling the truth. You're accurately reporting what happened in the data without making too big of a stretch to say that something causes something else, which can be very, very tempting. So I highly recommend you check out the little quizzes that I put in the optional resources.