Module 1 Video 1: Is Data Objective?

[00:00:00] Hi, I'm Heather Krauss, and welcome to our first video in the module where we're going to talk about how to get started embedding ethics and equity into your data journalism projects. And I've been involved in data analysis and data journalism for many, many years. And I got into this field because I was really, really excited about the idea that we could use data, quantitative data specifically, to embed equity and objectivity into our stories, into our narrative. That we could use data to kind of get past politics and religion and color and gender and create a shared understanding of what was going on in the world by using quantitative data, because quantitative data, as I was trained growing up and going through school, was objective. I have advanced degrees in mathematics and computer science and statistics, and I really loved the fact that data was objective.

[00:01:09] And then I started working in the field and very, very quickly realized that data is in fact not objective.

[00:01:18] And the place where I really, really came to understand this was when I was working in rural Bangladesh and we were telling stories about what was happening on the ground in rural Bangladesh and what was helping and what was not helping in the human conditions in rural Bangladesh. And this is where I realized that this is what data looks like on the ground.

[00:01:45] The thing that we called data, that ends up in a spreadsheet or a model or an algorithm that gets put into a data journalism story looks like this in a lot of places on the ground. This is what data starts its life as.

[00:02:04] And you can see that here in rural Bangladesh, we were collecting data using symbols and pictures and things that were meaningful on the ground. And then by the time that ends up in a data story on 538 or New York Times or any of the other fantastic sources of journalism that you might read or watch or consume on a daily basis, the number of transformations, something like this goes through before it ends up in a data story is obviously for many, many sources of of non-objectivity that's being embedded into that data.

[00:02:50] And in order to really honor the place that data comes from and to use data accurately in a data journalism story, it's essential that you learn the techniques for embedding ethics and equity into your data journalism story. And one of the examples that I lived through that really changed to the way that I work with data was we were trying to tell a story about whether one of the projects that a large international NGO was doing on the ground. And we wanted to know whether that project was working. And they sent us their data and we looked and found out that, yes, indeed, the data shows that project participants on the right hand side of this graph had a big increase in income from the start to the end of the project compared to a control group who had a much lower increase in income from the start to the end of the time that the project wasn't in in the field.

[00:03:53] Now, a control group means a group of people who are supposed to be really, really similar to the project participants but aren't in the projects. So control group is supposed to be what would have happened if we didn't have our project. So when we compare these two things, the graph on the left, which is the control group and the graph on the right, which is the project participant group we see, hey, This project looked like it was a success. Let's write a story about this successful project.

[00:04:24] However, we were on the ground in this community and we were really aware that there were a lot of different ethnic groups in the community where the project was happening. And we looked at data on the ethnic groups and found that the project worked really differently for different ethnic groups. So for ethnic group one, which is the red lines in these graphs. The project worked really, really well. So if you see the line on the left, the control group, their income went up a little bit from one twelve to one twenty one is where that red line goes up on the right hand side for the project participants, that ethnic group, that red line goes from one ten to one fifty nine. It goes up really, really high. So you can see that the ethnic group, number one, which is the two red lines. The people involved in the project. Have a lot more income at the end of the

project than people who are similar to them at the end of the same time period. But not being part of the project.

[00:05:42] However, if you look at the yellow line in this chart, you can see something really, really different. These two lines, the yellow line on the left and the yellow line on the right show that those lines are about the same. Ethnic group number three, their income goes from ninety to ninety seven in the control group. And on the right and the project participants, their line goes from 90 to 95. So those are essentially the same. So what that means is that this project worked really, really well for increasing the incomes for the ethnic group, number one. But it didn't work at all for increasing the incomes in ethnic group number three. So if I had gone and written a story that said this project worked great because we have this data that showed that the project worked great. I would have been complicit in hiding the lived experiences of people, an ethnic group number three. And the people in ethnic group number three, probably don't have access to this data. So if I write a story in the news paper that says this project was a great success, we based it on this data and the people in the project had much higher income than the people who were not in the project.

[00:07:03] And I'm an ethnic group, number three, and this is the only data that I have access to. I'm looking around and I'm thinking, well, the project didn't work for me. Is this data wrong or am I doing it wrong?

[00:07:17] So the difference between using general data that you kind of understand or very specific data that you've spent a lot of time to understand makes the difference between an equitable or not equitable data story. And there's one more step to talk about in this in this example is that from the perspective of income inequality, you can see that at the beginning. Of our or not at the beginning. But, yeah, at the beginning of our time period, the red I ethnic group number one is earning significantly more than yellow ethnic group number three. And that's the same in the control group and the project participant group. But by the end, in the control group, by the end of our time period, there's a twenty four dollar gap. The income inequality between ethnic group one in three is a twenty four dollar gap. By the end of the time period for the control group, that's the people who are not in our project. However, on the right hand side, you can see that by the end of our time period, among the project participants, there's a sixty four dollar gap between the highest ethnic group and the lowest ethnic group. So our project on the ground actually increased the income inequality in the community.

[00:08:54] So you or me as a data journalist have three options so we can report on the project using this data and say the average income increased. Looked like the project worked. Or we can say the average income increased for some people, but not for others. Looked like the project worked a little bit. Or we can say the project greatly increased the income inequality gap in our community. And the project didn't work at all. Those are three of many, many possible different ways to use data in a data story. So this these possibilities are my lived example of why it's essential that if you're going to write data journalism stories, you need to have a set of strategic processes and checklists to decide and make transparent what are the ethical and equitable uses of data in your data story?

[00:09:58] And what is equity and data anyways? Equity and data is processes to avoid racism, sexism and homophobia in data and analysis that goes into your story. What is bias? As a journalist you've probably been taught to think that bias is taking value judgment rather than simply trying to me maintain objectivity. In data journalism, bias is systematic error introduced into sampling or testing by selecting or encouraging one outcome over the others. So bias is kind of a data process way to set up the way that you interact with data to encourage a specific result. Often it's done unconsciously and accidentally. Sometimes it's consciously and intentionally, and usually it's a combination of the two.

[00:11:05] So in this course, we're going to walk you through a really specific seven step data equity cycle where you can learn how to embed equity and ethics into your data projects. And the seven step data cycle is a systematic, tool based and checklist based way for you to monitor and think about everything that you're putting into your data project.