Research and Data Literacy
(CSHS Racial Equity Training Session 3)
Job Aid

SHIELD
School Health Institute for Education and Leadership Development

What is data?

1

Data is a collection of discrete values that are recordable and convey information (numbers, categories, text). Data encode aspects of the world we can observe and wish to measure. Some things are easier to encode (i.e., we use degree measurements on a thermometer to encode temperature), while some things are more complex to encode and need multiple specific measurements to fully describe them (i.e., to fully encode a storm, you would need multiple measures, like geographic coordinates, direction, land speed, average temperature, precipitation type and amount, wind speed, etc.).

How is data useful?

2

- Data stores information about the past
- Data can point to possible causes of why things happened in the past
- Data can show a need and raise awareness about a particular issue
- Data enable us to make predictions about the future
- Data may enable us to avoid human biases and see the world as it is

What are some limitations of data?

3

- Data are not magic
- Data sets can be inaccurate, biased, and/or undesirable
- Data sets contain noise (information that is useless for the purposes of the question you wish to answer)
- Data can lead to inaccurate predictions or conclusions
- Data can cause toxic feedback loops
- The data you collect may only reflect certain aspects of what you are trying to capture (not a complete picture)

What are statistical bias and social bias?

4

Statistical bias is when a data set does not accurately represent the larger population you wish to measure. A statistically biased data set can lead to conclusions that are not representative of the general population. Examples of data sets with potential statistical bias include reports with limited survey responses or reports with self-reported illness.

Social bias is when certain arbitrary groups of people are systemically disadvantaged, the effects are reflected in the measurements we take, leading to inaccurate/undesirable conclusions. A socially biased data set accurately reflects existing inequities and unfairness of a society and exposes patterns that are undesirable. Data that may be socially biased could include: bullying, absenteeism, race/gender, transportation resources, caregiver surveys, and communication issues.

Why is the way data is collected important?

5

The way data is collected is very important. Data is generally never inaccurate (unless the measurement tool is broken), however, the conclusions, inferences, and predictions we draw from those data may be inaccurate or incorrect. If the question or process used to collect the data was flawed or contained too many variables to answer the initial question, it may be impossible to make conclusive findings from the data set.

What is an algorithm and how can it become biased?

6

An algorithm is a precise sequence of steps that transforms input data into an output.

Algorithms are useful in aiding decision-making by:
- Scaling decision-making (making decisions more quickly)
- Eliminating/automating decision-making
- Making better-informed decisions
- Eliminating the influence of human bias and fallibility
- Discovering new solutions to existing problems (i.e., in the AI world)

Properties that make algorithms objective include: they are discrete (subdivided into steps), they are explicit (each step clearly and precisely defined), and they are repeatable (does the same thing every time regardless of input). Properties that make algorithms subjective include: the fact that they are written by humans, they have no knowledge of the world except what humans tell them, the results are interpreted and put into practice by humans, and they are susceptible to feedback loops.

Algorithmic bias describes systematic and repeatable errors in a computer system that create unfair outcomes. If there is bias in an algorithm, you'll get biased data out of it. Algorithmic bias is a symptom of social and institutional biases in our society.

How can bias in data lead to toxic feedback loops?

7

Using data from the past to inform decisions about the future can create a toxic feedback loop. If there is bias in the data we collect, and we form interventions around that biased data set, the bias continues. If we acknowledge that there are social biases in the world and are intentional about building equity and fairness into algorithms, we will collect data that is less biased, and the conclusions we draw will become more meaningful.

What is the difference between statistical race correction and statistical race awareness?

8

Statistical adjustment/correction is a statistical method to remove unwanted/unrelated effects from data. It removes some of the "noise" from the information being sought. It is a routine procedure in many analyses but doesn't work for all instances. Statistical race correction can cause harm by erasing the differences between people that are biological variables, changing the interpretation of our results, and making the data set less accurate.

Statistical race awareness can indicate areas where we must do better and where problems actually fall. Looking at variables beyond race to explain health disparities is an example of looking at race responsibly.

How can I look at data in a more responsible way?

9

As you explore data sets, be a detective. Look at the data you already collect in a new way. Make observations about the data – what stands out, what makes you want to learn more, and how you might use this data to advocate for your students. Assess if the data is representative of whom you're trying to help. Look beyond just the data and dive deeper into what other factors might have contributed to the results you see in that data set. That will give you the context you need to form interventions that are more holistic, sustainable, and meaningful.

What are some good health data maps?

10

Although health data maps also have limitations, they can be useful tools to start your data exploration. Here are links to some helpful interactive data maps:
- CDC, PLACES: Local Data for Better Health
- The Opportunity Atlas
- Diversity Data Kids ,Child Opportunity Map
- USDA Economic Research Service, Food Access Research Atlas