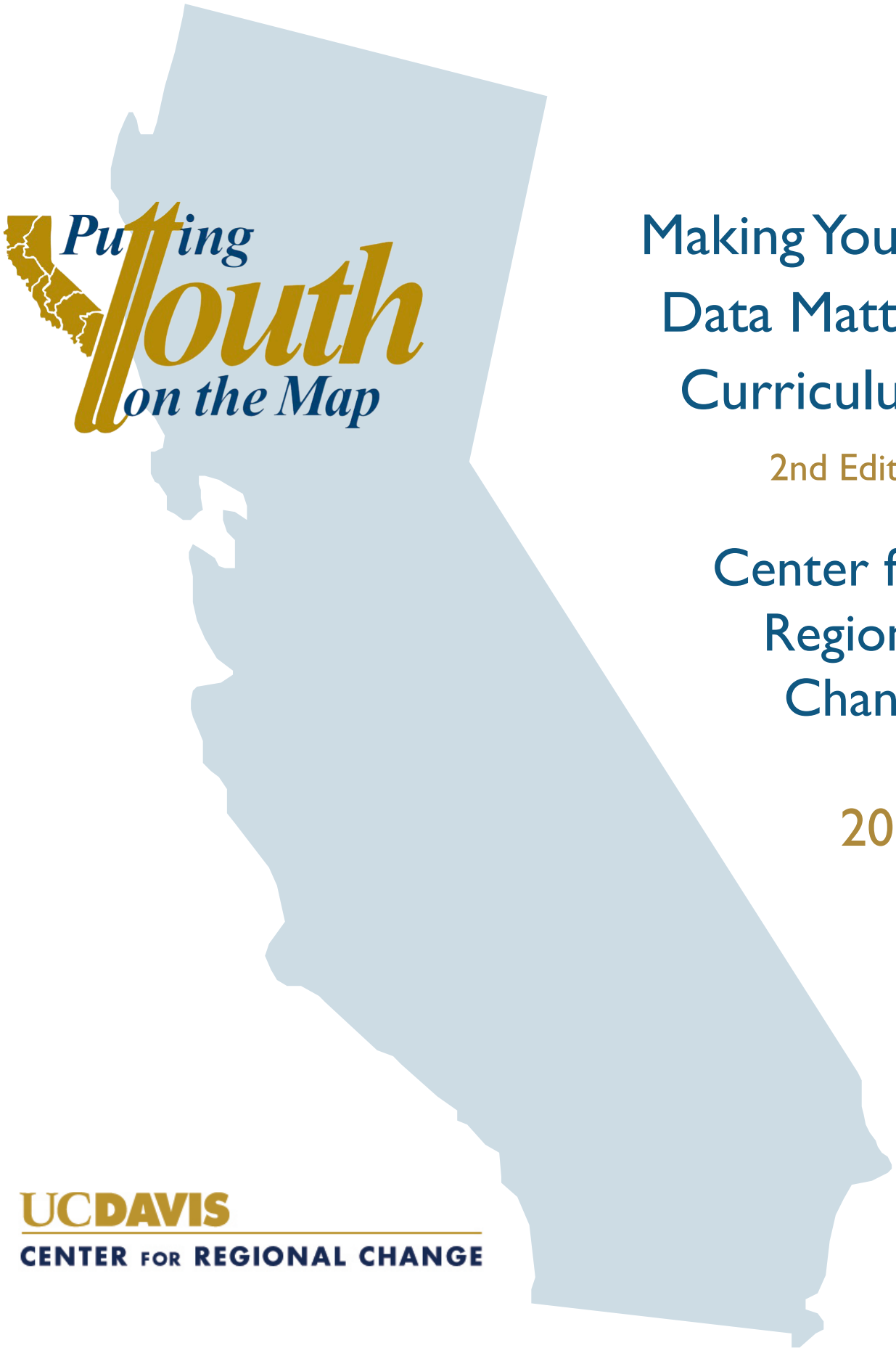

Making Youth Data Matter Curriculum

2nd Edition

Center for Regional Change

2017

UCDAVIS
CENTER FOR REGIONAL CHANGE

http://interact.regionalchange.ucdavis.edu/youth/resources.html#learn

3.6.A The Data-ing Game

Participant Roles and Scripts

Welcome to the Data-ing game.
In each round a Map is going to be looking for the perfect data for their project.
For each Map the ideal dataset is: Trustworthy, Accurate, Representative, Accurately Visualized

Round One

Welcome to the first round — Let's meet Map number 1 who is hoping to help youth advocates research high school dropout rates in California, and identify schools across the state with the highest rates of graduation.

Questions for the Data contestants
Q1: Tell us who collected your data?
Q2: Tell us how old you are?
Q3: Tell us about your survey sample?
Q4: Can you show us some results?

Data #1
A1: I was collected by a company in Michigan that mostly does web design for restaurants.
A2: I am just five years young.
A3: Sure, my results come from an extensive survey of every high school senior from two high schools in Eureka, CA.
A4: Shares colorful, unlabelled charts and banners.

Data #2
A1: I was collected by the California Department of Education or "CDE," the state agency to which all public schools have to report educational data.
A2: I am six months old, but I'll be updated again at the end of the next school year.
A3: I come from reports required of every public school in California.
A4: Shares CDE data via a well-labeled chart.

Data #3
A1: I was collected by a very nice lady named Yulia Vann.
A2: Well, it's been a while since Yulia has had a chance to update me, so I'm about twelve years old.
A3: Well, Yulia actually surveyed a group of grandparents to see how they thought their grandchildren were probably doing in school.
A4: Shares very unclear bar graph.

Round Two

Welcome to the second round of the data-ing game — In this round we have a Map that wants us all to get moving. This Map is going to help our very own city council to improve the sidewalks and bike paths.

Questions for the Data contestants
Q1: Tell us where you're from.
Q2: What's your favorite survey question?
Q3: Who do you spend time with when you are doing a survey?
Q4: Can you show us some data?

Data #1
A1: I was collected by a transportation advocacy group.
A2: My favorite survey question is probably: "How far do you walk in a week?"
A3: I really like the quiet life, so I tend to just stick to the suburbs when I am doing my surveys.
A4: Shows barchart reflecting lack of data for downtown area.

Data #2
A1: I was collected by a small group that organizes an event called "drive to work week."
A2: You know surveys take a lot of time, so usually we just sort of guess what people are thinking.
A3: The great thing about not doing surveys is that it gives me more time to drive around in my car by myself, and not have to talk to anyone.
A4: My results are actually private so I can't show them to you.

Data #3
A1: I was collected by the California Department of Transportation, you can just call them DoT.
A2: I really like asking people if they ride their bike in the city, and if they say no I like to ask why not.
A3: I'm a real people person, so I try to spend time with everyone, usually I'll just go door to door in the whole city when I'm doing a survey.
A4: Shows bar chart of data broken out for males and females without information on numbers of survey participants.

Round Three

Welcome to the last round of the data-ing game — In this round we have a Map that's looking to support environmental justice. This map is going to help a legal advocacy center trying to show the impact of Superfund sites in the Central Valley of California. (Superfund sites are abandoned places with toxic waste that might pollute the environment around them).

Questions for the Data contestants
Q1: Tell us who collected your data?
Q2: How would you describe yourself? What kind of data are you, and where do you come from?
Q3: How do you feel about sharing?
Q4: Will you show us some results?

Data #1
A1: I was collected by a university research team.
A2: I am an index of toxic exposure for all of California.
A3: Well, I love to share my index results from all over the state, but I can't really talk about the different domains that make up the index.
A4: I've created this 300 page report.

Data #2
A1: I was collected by a group sponsored by a chemical company. I can't really tell you which one though.
A2: I am a database of all the Superfund sites in the US including their current conditions.
A3: You can buy all of my data and metadata for what I think is a reasonable price.
A4: I've created this beautiful website to share our research, and it's available to all subscribers.

Data #3
A1: I was collected by a network of youth community organizers.
A2: I am a collection of soil and air samples from neighborhoods in Los Angeles, San Francisco and Sacramento.
A3: I love to share. All of my data is freely available to anyone who wants to see it
A4: Our youth team has created this website to share our methods and research. It's not slick, but it works.

Audience Roles: Your job as the audience is to help the Map select the perfect data match to display. During the game, the Map will ask very important questions of the Data participants. Use these data criteria to help the Map select the perfect data match.

- **Source Trustworthiness**
  - What are the source's biases (what is their agenda)?
  - Does the source have adequate skills/resources to collect/process the data?
- **Data Accuracy**
  - Was the information collected likely correct/accurate (e.g. survey questions are not leading, participants likely provided accurate information)?
  - Was the information likely to be accurately entered into the data system?
  - Are the data up-to-date (or is the most recent data available)?
  - Are you confident that the analyses were done well (e.g. mathematical analyses are likely correct)?
- **Representation**
  - Are the data representative enough of all locations and/or populations of interest (consider margins of error, participation rates, where surveys were/were not administered, whether data collection used strategies to reach people who speak limited/no English, etc.).
- **Data Display**
  - How do map breakpoints affect the "story" told by the map?
  - Do the geographic units used for display visually distort interpretation? If yes, then how?
  - How do colors used for display affect visual interpretation?

3.6.B Trainer Notes for the Data-ing Game

Round 1
In this round there is a clear winner - Data #2.
The source is an organization that has the ability and skills to collect and analyze the data. It is very representative since all high schools are included, is up to date and will be regularly updated. In the data display there is information about the sources, methods, and opportunity to learn more about the data.

For Data #1, the source doesn't seem to have the skills to collect this type of data, and the data is not representative, comes from just two high schools, and is a bit out of date. The data are fun to look at, but doesn't have any information about sources or substance. Data #3 were also collected by someone who doesn't seem to have the skills to collect the data, are not representative, are very out of date, and are not presented well (for example "success" is misspelled).

Round 2
In this round it is a little bit less clear. Data #2 is not a good choice. The organization collecting the data is biased, they are asking the wrong questions to help understand biking and walking, and their data is not representative. Finally, we can't evaluate their data display since they won't share it.

Data #1 and Data #3 are pretty equal. They were both collected by groups that should have the skills to collect and analyze the data. They have both asked good questions, and they display their data in clear informative ways. But Data #1 is not as representative since it ignores the downtown and rural areas.

Round 3
This round is the one that doesn't really have a clear right answer. It shows that sometimes there are no perfect data, but you have to choose the data that work best for your question or project, or even combine data from different sources.

Each source is very different, but has the skill to collect the data and analyze it. All of the data also seem like they could be accurate. Some problems with Data #1 are that the university research group will only share their inished index, so it's hard to evaluate how accurate they actually are. Also their report is not very accessible since they only provide a long, complicated report. Some of the problems with Data #2 are that the group might be biased, (but they are willing to share all of the information you need to evaluate their work) and the data might be expensive. Data #3 is free, and the group that collected it will share all data, but the map is interested in the Central Valley of California, and most of these data comes from outside of the Valley. But maybe we could use their methods to collect our own data.

3.6.E Data Criteria Handout

What to look for when critically looking at data that you are using or that is being used in maps.

- **Source Trustworthiness**
  - What are the source's biases (what is their agenda)?
  - Does the source have adequate skills/resources to collect/process the data?
- **Data Accuracy**
  - Was the information collected likely correct/accurate (e.g. survey questions are not leading, participants likely provided accurate information)?
  - Was the information likely to be accurately entered into the data system?
  - Are the data up-to-date (or is the most recent data available)?
  - Are you confident that the analyses were done well (e.g. mathematical analyses are likely correct)?
- **Representation**
  - Are the data representative enough of all locations and/or populations of interest (consider margins of error, participation rates, where surveys were/were not administered, whether data collection used strategies to reach people who speak limited/no English, etc.).
- **Data Display**
  - How do map breakpoints affect the "story" told by the map?
  - Do the geographic units used for display visually distort interpretation? If yes, then how?
  - How do colors used for display affect visual interpretation?