Unit 5: A Community Data Collection Project

Introduction:

At this point in the sequence, students have developed a variety of understandings that allow them to run a data-based project from start to finish. They have used many of the individual areas of knowledge and understanding already and this is a chance for students to bring them all together. In this unit of lessons, the students will design their own data project as a class, collect and analyze their data (as they have before), and conclude by using their data in a communication to a stakeholder. This is a longer project that may extend over many days. It gives students opportunities to design data collection instruments, collect data, and answer questions of importance to their community.

Note: At the end of this lesson there is a vignette sharing a detailed description of a class going through this project.

Time	Activity	Materials
30 min	**Data Talk + introduction to Citizen Science**	GalaxyZoo graphs • https://blog.galaxyzoo.org/2020/03/20/radio-galaxy-zoo-lofar-the-first-classification-results/ • Citizen Science handout, page 8
25 min	**Brainstorming:** • Share with students that the project will involve collecting data on an issue, analyzing it, and then sharing their conclusions with stakeholders in order to create change. • Ask students to brainstorm topics • Agree on a topic	• Poster paper, white board or google document for collecting ideas
25 min	**Data collection tool design:** • Ask students to research different ways to collect data • Groups of students decide upon and create their data collection tool	• Technology to conduct research

Hours...	**Collecting data:** Ask students to determine the intended participants of their study Ask, "How will you distribute your survey?"	• Chart paper • Graph paper • Colored pens or pencils • Rulers or scissors • The data could be collected using google tools
40+ min	**Analyzing and presenting data:** • Students work in groups... • Analyzing their data • Preparing their presentation report and materials • Groups share their reports with the class	• Technology • CODAP • Google tools • Chart paper • Graph paper • Colored pens or pencils
30+ min	**Communicating with data:** • Ask students to brainstorm and make a list of who should see their data • Students prepare a presentation for the stakeholders	• Poster paper, white board or google document for collecting ideas
5 min	**Reflection/journal entry:** • What other aspects of your life could you change with the power of data?	• Journal

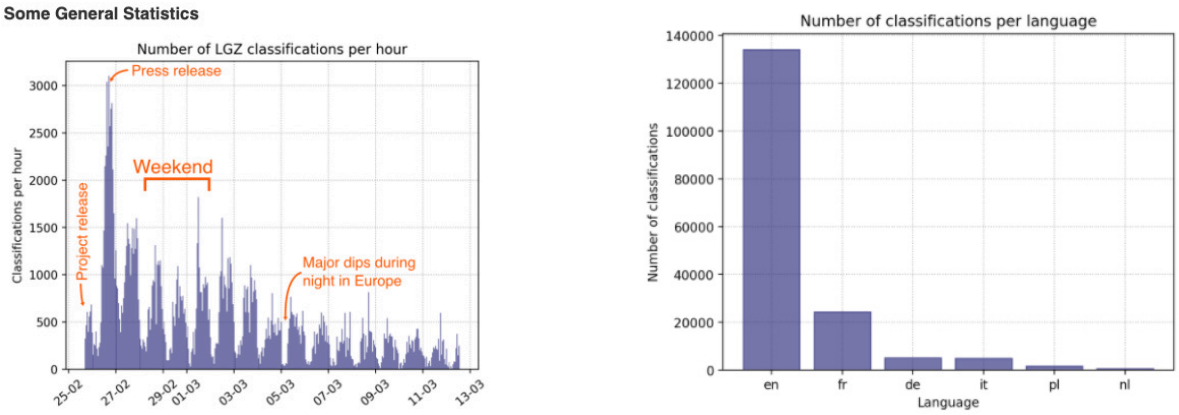
Data Talk + introduction to Citizen Science:

To start, use the Citizen Science handout to share with students what Citizen Science is and how it allows anyone to be involved in projects that further our knowledge. As described in the handout, some projects require more involvement than others. This data talk shares data from a project where scientists asked people to classify pictures taken by a radio telescope according to a few rules. It was important because it's not hard for people to classify images, yet it is still a challenge for computers. The data talk shows a couple of graphs from the first few days after the project was launched.

In the trialing of this Data Talk students shared the following:

- Dips during nighttime in Europe means that most participants are likely in Europe or Africa.
- It's interesting that some people knew about the project and worked on it before the press release.
- English is a really common language on the internet, so I wonder if there were lots of users from England (and other English-speaking countries) or if users are more distributed in Europe/Africa but they prefer English anyway.

Once your class has discussed the graphs, discuss how the project they will be doing as a class is the more involved kind of citizen science, where they will be taking the project from start to finish.



Brainstorming project ideas:

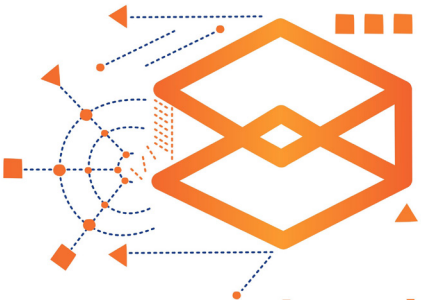
This project can be conducted around any concern or curiosity the students have about their community. Share with students that the project will involve collecting data on the issue, analyzing it, and then sharing their conclusions with stakeholders in order to create change. We encourage you to adapt it to your circumstances, however, we are presenting an example around public health since that is a topic that concerns every community.

Public health might seem like too large a topic to cover, but you can lead the discussion by first asking students "What are practices that contribute to keeping people healthy?". This will likely get answers like "eating healthy food" and "exercising" which are still too broad, but they lead to the more interesting question of "What are features of our community that make it easier or harder to accomplish those?" That is where your brainstorm can focus. In brainstorming, make sure no one is evaluating any suggestions. This is a time to propose any idea, no matter how simple, complicated, or ridiculous it might sound, as it might inspire others. Encourage students to respond to each other's suggestions with "Yes, and..." and get as many ideas on the table as possible. Once you have a rich list of factors and features that make it easier or harder to be healthy, as a class, choose a topic to focus on. Take some time to discuss each option and its feasibility. You might agree by consensus or open it to a vote.

Data collection tool design:

Once a topic has been chosen, each group of 3-5 students selects a method of data collection they would like to use. Encourage them to do some preliminary research on the topic, as they may find data collection tools or methods that have been used before. Regardless of the topic, surveys of stakeholders and collecting data on the availability of the feature or resource are generally good options. It is okay if there is some overlap in what the groups choose to do. For example, all groups might end up making surveys, just make sure their target audiences do not overlap. Remind students this is a whole-class effort and the more diverse forms of data collection they have, the better their data will be at the end of the project. The final product is intended to be a convincing argument that will produce change.

Collecting data:
Preparing their method of data collection will look different for different groups. Some might be using an existing survey while others might be spending more time in this step. This is okay. Different groups will find different parts of their project more time-consuming than others, just encourage students to make the best decisions for their data collection and analysis.



Students will likely need several days for data collection. Ideally, they can do part of it during class time, but they might require some additional time to work on it and visit other locations. Make sure this is not over-burdensome to students. They can adjust their goals as necessary. If possible, partway through the data collection process, have some class time dedicated to progress check-ins where students can share what has been successful and what has been difficult so far. Students might be able to help each other brainstorm solutions to their issues or they might choose to shift their focus if needed. Data collection, while incredibly important, can also be messy and difficult, and recognizing that can be an important part of this step.

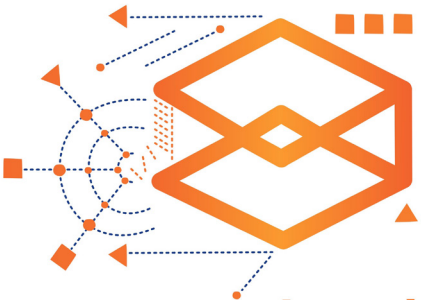
Analyzing and presenting data:
Data analysis will look different for different groups. This is the portion of the project they should have the most experience with by now. They can use the skills they developed earlier in the course to put their data into CODAP and look at it. Some groups might find they have data that cannot be reflected through CODAP's visualization options (for example, geographical data). Remind them of all the different visualizations they have seen throughout the course like maps and timelines and encourage them to be creative. Their visualizations can be hand-drawn if students decide this is the best option for their data.

Students make a poster or a set of slides with their findings to share with the class. They should make sure to include visuals or other ways to justify their conclusions with data.

Communicating with data:
Before the poster presentations, take some time to discuss possible stakeholders who could learn from the project. A short brainstorm can help. The stakeholders can be people in positions of power who could enact change or other people in the community who could be galvanized to make change happen. Each group should pick a stakeholder. A variety of stakeholders would ensure the issue is heard in various circles, but it is ok if several groups pick the same stakeholder, as each group will have a different approach.

Once groups have selected who they would like to share their work with, each group can present their posters or slides. Encourage students to discuss similarities or differences between their group's findings and also to take notes on other groups' findings that they think might be helpful in their communications with their stakeholders.

Finally, students design their communications. These can come in many forms: an article for a local or school newspaper, a letter to the mayor or a school administrator, a video to share with their peers, etc. Students should choose whichever mode they think will work best. They should carefully consider what they want to share: are they bringing an issue to light? Proposing a solution to an existing issue? Making recommendations for the public based on their data? Any of these are powerful communications, they should just make sure to use the data to back up their ideas and use at least one visual to make their data clear.



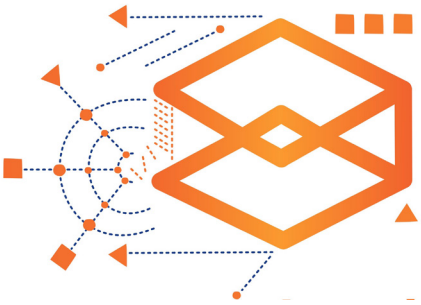
Look-fors:

- **Are the question ideas answerable with data that can be collected during this project?**
  The project ideas brainstorm is all about sharing as many ideas as possible regardless of feasibility, since impossible ideas can evolve into really interesting projects as students inspire each other. However, after the brainstorm, students will need to take some time to bring reality back into the picture and discuss the ideas proposed. Look for students discussing the kind of data they will need to collect in order to answer their questions. They can consider whether the data will be available to them, the data collection timeline, the kinds of places they might need to go, and who they might need to contact. This can help students narrow down their choices as well as being a great exercise in thinking about forms of data collection needed to answer different types of questions.
- **How are students organizing their data?**
  This is a question that will come up in multiple parts of the project, but it will be key in both the data collection and data analysis stages. When students are designing their data collection method, encourage them to think about how they plan on analyzing the data, and whether organizing their data collection in certain ways might make it easier for them to deal with the analysis later. Perhaps it is collecting their data in a table they can later put into CODAP. Encourage this if it makes sense, otherwise discuss with them how to best keep track of what they find. Make sure to remember, however, that just because the data might not fit easily on a spreadsheet, it does not mean it is not valuable data!
- **How are students visualizing their data?**
  The way data is visualized impacts the story it tells. As students analyze their data, encourage them to experiment with different representations to find the ones that highlight the aspects of the data they wish to showcase. When creating their stakeholder communication, students may wish to create a new visual that speaks more directly to the point they want to convey to the stakeholder. Encourage students to constantly be thinking about the story they want to tell about the data, because that should directly inform how it is visualized.

Reflection/journal entry:

What other aspects of your life could you change with the power of data?

Vignette:



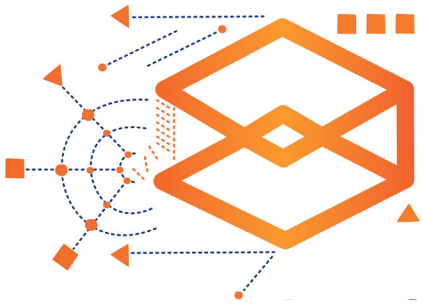
To start, students learn about the concept of citizen science and how people collecting data can make a difference for themselves and their communities. They brainstorm topics they would be interested in contributing to and discuss factors that help or hinder the ability of their community to be healthy. Some of the topics they discuss are: availability of healthy foods in their community, lack of outdoor space to exercise, stress levels in the school community, the walkability of the area around the school.

Students settle on walkability of their neighborhood, since many students wish it was easier to walk to school. They split into groups of 4 or 5 students and each group takes some time to research walkability and how it has been measured in the past. One group (we will call this one group A for this vignette) decides to survey their schoolmates on their experiences, and a second group (group B) discovers a walkability checklist online, which lists features of a space that make it conducive to walking. Group B decides to use that checklist to evaluate a two-block radius around the school. To prepare their methods, Group A comes up with questions and creates a survey using Google Forms. Group B prints out several copies of the checklist they found and looks at a map to assign sections of the neighborhood to the different group members.

Students go off in their data collection journeys. They have a week to collect all the data they need. Group A sends their survey out to a student mailing list requesting that everyone fills it out. Group B fills out a separate walkability checklist for each section of the school's surroundings. Two days into their data collection week, the class has some check-in time to see how everyone's data collection is going. Group A shares they sent the link to the school community, but they are not getting as many responses as they hoped. Other students in the class suggest printing posters with QR codes and putting them on bulletin boards, as well as having each group member personally ask a few people to fill out the survey. Everyone in the class agrees to take it that afternoon as well. Group B shares that they are having trouble keeping track of all the filled out checklists since they are spread out among the group members. Other students in the class suggest inputting the checklist results in a spreadsheet as soon as they are collected. That way, everyone in the group can see the results as they come in, and students don't have to worry about misplacing the paper checklists.

Now back in the classroom with their data, students tackle data analysis. Group A's data looks much like the kind of data they have worked with before, they have experience downloading the data from Google Sheets and uploading it into CODAP. They look at the results of each question as well as how some of the results correlate. They find, for example, that the longer it takes a student to get to school, the busier intersections a student had to cross. They print several of the graphs they find meaningful and use them to make a poster to share with the class. They make sure to give each graph a title and a sentence explaining the conclusion they drew from it.

Since Group B organized their checklist results in a spreadsheet, they can similarly download it and import it into CODAP. There they can compare and visualize how many checks each section of the neighborhood got. They give each section an overall "walkability score" based on how many checklist items it had. After some time, they realize that there is a dimension of their data that CODAP cannot easily account for - geographical location. They decide to draw a map of the area surrounding their school and annotate it with their findings from CODAP. They color a section green if it has an overall "walkability score" above a certain threshold, and they color it red if it scored below a certain threshold.

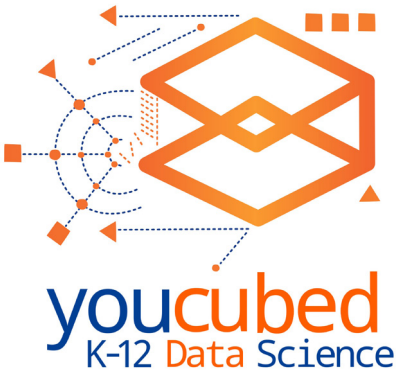


Then, the class takes some time to discuss who the stakeholders are in this situation. They come up with 4 ideas: the students and teachers, the school administration, local government, and other people who live near the school (even if they do not work or go there). Each group picks a group they are particularly interested in. Group A picks the school administration. They learned that the crosswalk in front of the school not always being respected by motorists was the issue most students ranked as their main concern and they are hoping the school administration can do something about it. Group B picks the other people who live near the school as their stakeholder, since they spent time in the surrounding area. The different groups present their findings to the class, and everyone takes notes on what pieces of information might be useful in communicating with their stakeholders.

Group A puts together a set of slides to present to the school administration. They share the most important findings from their survey, as well as some of the data from another group that surveyed parents, and at the end, they recommend assigning a crossing guard at that crosswalk in front of the school at the hours of heaviest foot traffic. They use the visuals of their data to show how important an issue it is.

Group B decides to write an article for the local newspaper about neighborhood walkability. They prominently display their map visualization to draw people in and convey the information in a more digestible way. They also include some findings from Group A's survey, including a statistic about the percentage of students at the school who feel unsafe walking to school. They email their final product to the editor of the local newspaper and explain that they hope to inspire change by spreading this information to the community.

Citizen Science Handout



"Citizen science" is data collection and research that is conducted by regular people instead of professional scientists. Citizen science helps us to make sure that research is connected to the experiences of everyday life. It helps us to understand the world around us in new ways. It also creates new knowledge that can be used to make positive changes. There are three main types: Citizen science FOR the people, WITH the people, and BY the people[1].

Citizen Science FOR the people:

This type of citizen science is the one where regular people are involved the least. Scientists create their projects and ask for other people to give them samples or data about themselves for the scientists to work with. For example, there are some projects where scientists request blood samples from people with a disease to learn more about it.

Citizen Science WITH the people:

In this type of citizen science, regular people help scientists collect their data. Scientists create the projects and ask people for help. Citizen Scientists work to gather data and pass it on to scientists for analysis. For example, FrogWatch USA asks people to report the mating calls of frogs and toads in wetlands near them. This helps scientists learn about frogs' lives!

Citizen Science BY the people:

In this type of citizen science, regular people do all of the work! They follow the same methods scientists do, but applied to their own questions. This is really powerful because citizen scientists can use their own data to improve their communities. For example, a group of older adult citizen scientists gathered information about safe walking in their neighborhood and used their findings to work with their city on sidewalk improvements and community cleanup[2].

If you want to read more about citizen science, you can read about it here: https://www.nationalgeographic.org/encyclopedia/citizen-science/. Don't be afraid to click on some of the projects described there and get involved! Most of these projects are FOR and WITH the people. In this unit, you will engage in the most powerful kind of citizen science by designing and working on a project BY you, the people.

[1] You can learn more about the types of citizen science in this paper: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4978140/

[2] You can read the scientific paper they published here: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4978140/

Data Talk

youcubed
K-12 Data Science

Some General Statistics