# The Earth Scientist

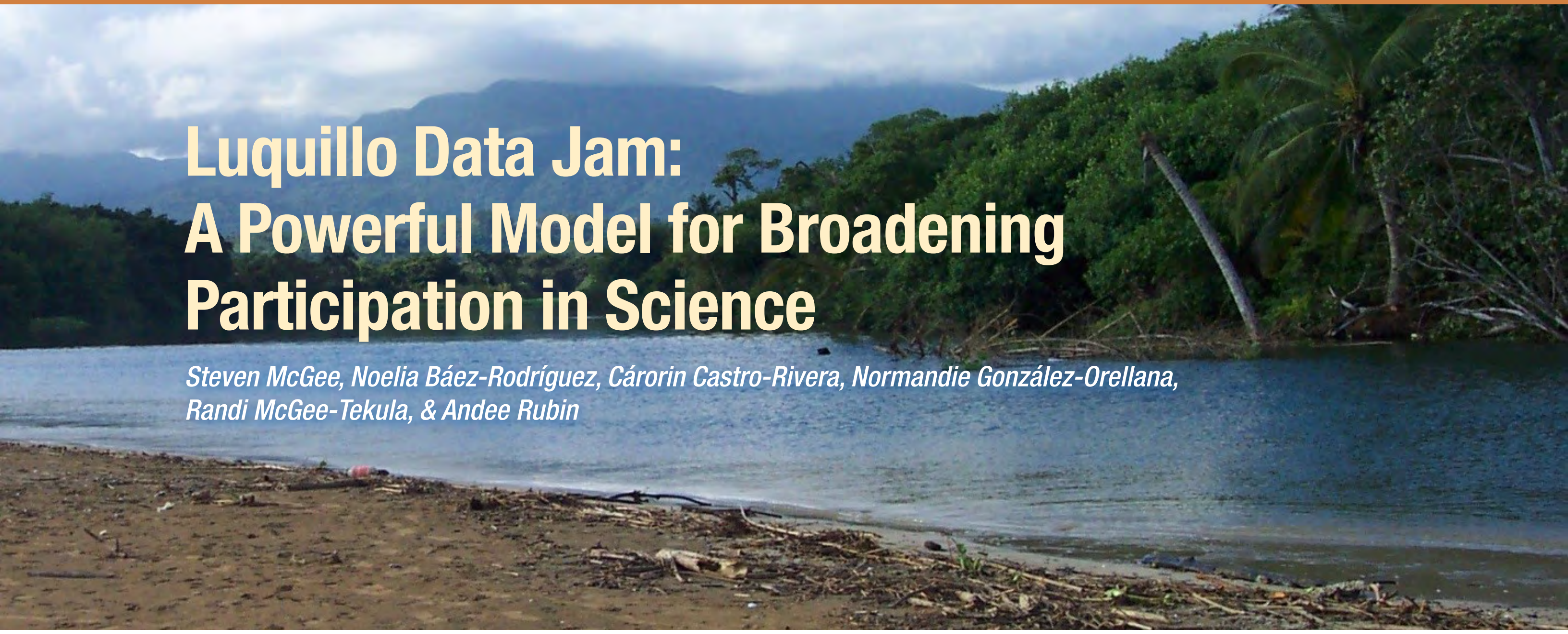
Volume XLI • Issue 2 • Summer 2024 $10.00*

The Vera C. Rubin Observatory on Cerro Pachón during a snowy day.
Photo Credit: NOIRLab/NSF/AURA/C. Corco

## INSIDE THIS ISSUE



*ISSN 1045-4772



# Luquillo Data Jam: A Powerful Model for Broadening Participation in Science

*Steven McGee, Noelia Báez-Rodríguez, Cárorin Castro-Rivera, Normandie González-Orellana, Randi McGee-Tekula, & Andee Rubin*

Pico del Este and the Rio Mameyes, El Yunque National Forest, Puerto Rico. Orographic clouds (as shown) are a significant source of annual precipitation and streamflow. Photo credit Martha Scholl, USGS.

## Abstract

The Luquillo Long-Term Ecological Research (LTER) program in Puerto Rico has implemented a Data Jam initiative that fosters student engagement in ecological data science. In December 2023, over 175 middle and high school students presented 51 data- driven projects at a symposium in El Yunque National Rainforest. The initiative, rooted in the LTER's extensive datasets, allows students to develop research questions, analyze data using the CODAP tool, and present findings. The program includes professional development for teachers, with workshops and Virtual Lab Meetings that equip them to guide students through the scientific process. Teachers collaborate across disciplines, integrating Data Jam into their curricula. The program has positively impacted students, increasing their interest in local environmental issues and confidence in data analysis. The use of publicly available datasets tailored for educational purposes, alongside mentor support from LTER scientists, provides a robust framework for students to engage deeply with ecological data. This initiative not only enhances STEM education but also strengthens students' connection to their local environment and the broader scientific community.

## Introduction

On December 13, 2023, more than 175 middle school and high school students from 19 schools across Puerto Rico presented 51 data science projects based on independent research they conducted related to the Luquillo Long-Term Ecological Research (LTER) program (See Figure 1). This research symposium took place in the visitor center at the El Yunque national rainforest in Puerto Rico. Twenty-one projects were related to factors that affect the abundance of shrimp in El Yunque. Eighteen projects examined the hydrology of rivers in El Yunque. Twelve projects examined factors related to the size and abundance of dragonflies in streams in El Yunque and surrounding urban areas in Puerto Rico. Roughly two dozen undergraduate students, graduate

**Figure 1.** Data Jam students discuss a poster at the symposium.

students, and faculty from the University of Puerto Rico attended the symposium to view the projects and provide feedback.

This symposium is the culminating activity of the Luquillo LTER Schoolyard Data Jam program. The foundation of the Luquillo Data Jam program is the vast amounts of publicly available datasets that are generated as part of the LTER program. The LTER Network was created by the National Science Foundation in 1980 to conduct research on ecological issues that can last decades and span huge geographical areas. As the largest and longest- lived ecological network in the United States, LTER provides the scientific expertise, research platforms, and long-term datasets necessary to document and analyze environmental change. Luquillo joined the LTER network in 1988. An important part of the LTER mission is to provide public access to the long-term datasets produced across the network. The Data Jam model was initially developed at the Jornada Basin LTER in 2012 to help students and teachers access and use these LTER data for educational purposes (Bestelmeyer et al., 2015). The Luquillo LTER in Puerto Rico joined the Data Jam program and hosted its first Luquillo Data Jam in 2015 (McGee & Baez-Rodriguez, 2017).

Within Data Jam, classroom teachers support middle and high school students in developing their own scientific questions that can be answered using LTER and related datasets. The Luquillo Data Jam bilingual website (https://datajam.cloud) provides access to student-friendly versions of long-term environmental data along with data science lesson plans. The projects were all conducted with data collected in Puerto Rico, thus helping students connect with the environment on their island. Each of the datasets includes metadata, which describes the structure of the data and how it was collected. While Data Jam students do not collect their own data, they are expected to become familiar with the metadata and be able to describe the data and consider how the data was collected when drawing conclusions.

Students then use a student-friendly graphing tool called CODAP to analyze the data as evidence and develop a scientific argument that addresses their scientific question. During the 2023-2024 school year, teachers implemented Data Jam in their classrooms with more than 650 students. Data Jam culminates with teachers selecting the best projects from their students to present at the Data Jam symposium. The symposium participants are provided feedback on their presentations by peers, family members, educators, and scientists.

Surveys of Data Jam participants show that students report increased levels of interest in their local environment and confidence in using data to investigate environmental science topics.

Successful Data Jam projects require that students engage in a variety of data practices that underlie the STEM field of data science. Data Jam supports students in developing high quality Data Jam research questions that anticipate statistical variability. To investigate these Data Jam research questions, students describe the distribution of the data in terms of its center, spread, and shape. They use CODAP to develop graphs as a means to display the distribution of the data. Students can compare distributions using box plots and analyze the extent of overlap between the interquartile ranges of the two distributions. Students can analyze the relationship between variables using a regression line on a scatter plot.

To support teachers in the enactment of Data Jam, we have developed a professional development model that engages teachers as apprentices to ecological scientists and provides the teachers with the support they need to engage their students in data-based ecological research. Teachers participate in a weeklong summer professional development program to improve their statistical education skills and prepare them to support students conducting their Data Jam projects. During the school year, teachers participate in Virtual Lab Meetings with their peers and a scientific mentor to examine student progress and adjust the support that they provide leading up students' culminating final project.

## Data Jam Examples

For the fall 2023 Data Jam, students had access to three datasets related to shrimp, Odonata (dragonflies), and hydrology. Here we provide a project example related to each dataset. Most students developed their projects in Spanish and the remainder developed their projects in English.

- **Shrimp:** For more than 30 years, Luquillo LTER scientists have been following shrimp populations in the Quebrada Prieta, a stream located in the El Yunque rainforest. Freshwater shrimp are paramount for the ecosystem, since they help keep the water clean and distribute resources across the food web. The Data Jam version of the shrimp dataset contains weekly data on the amount of shrimp per trap in pool 0 in Quebrada Prieta from 2012-2019, along with environmental and water chemistry data (González-Orellana, 2023c). We provided students with fewer years than are available so the size of the weekly dataset would be manageable. The dataset includes the three most common shrimp species in El Yunque: *Atya lanipes, Xiphocaris elongata* and *Macrobrachium spp*. One group of students who presented at the symposium in English developed a scientific question on the relationship between water temperature and shrimp count: "Is there a relationship between water temperature and the number of shrimp per trap in pool 0?" The students hypothesized that rising temperatures due to global warming may lead to a decrease in the number of shrimp. Figure 2 shows the scatterplot and least squares line that the students produced in CODAP. The students concluded that there is no relationship between temperature and shrimp count due to the low $R_2$ value.

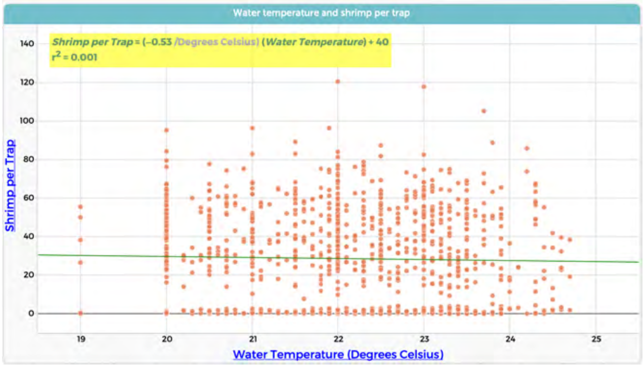



**Figure 2.** Student graph shows the relationship between water temperature and shrimp count.

  As part of their poster, students share what they are still curious about after their analyses. This group of students asked, "How long could shrimp live if the water temperature was higher than the data provided?" The students seem to recognize that although water temperature does not have an impact within the typical range, it is possible that it could impact the shrimp if water temperatures rose beyond this range.

- **Odonata** often serve as bioindicators (a species whose abundance can reveal information about the status of an ecosystem) because of their dependence on fresh water and prey availability. Odonata larvae are aquatic and rely on good water quality for proper development. A significant part of the diet of an Odonata is insects, so they rely on the abundance and availability of other insects to survive. Hence, if there is an abundance of Odonata, it can be concluded that the ecosystem is healthy enough to sustain an abundance of Odonata. The metropolitan area of Puerto Rico is one of the most inhabited and anthropogenically developed areas of the archipelago. This may have detrimental effects on biodiversity. To study the abundance of the Odonata in the streams of the mountains of El Yunque after the impact of the Hurricane Maria, researchers did a survey of Odonata larvae and their size at a stream in the El Yunque National Rainforest and a stream in the metropolitan area of San Juan. This dataset contains species richness and abundance of Odonata (González-Orellana, 2023b). One group of students who presented at the symposium in Spanish developed a scientific question about differences in body length between two species of Odonata: "¿Es diferente el largo del cuerpo

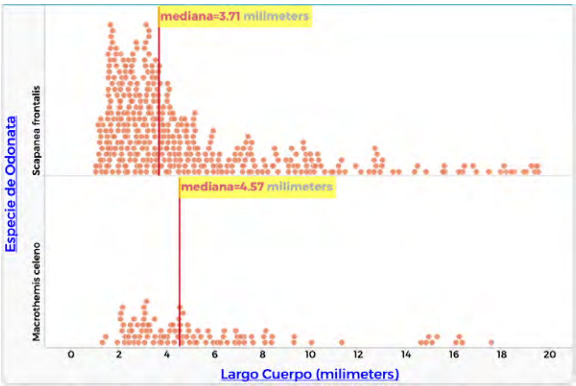



**Figure 3.** Student graph compares the body sizes of two different species of Odonata.

de las larvas del género Odonata según su especie?" *"Is the body length of larvae of the genus Odonata different according to their species?"* As part of their background information, the students reported that Odonata species are differentiated by body size, wings, and eye position. Therefore, they hypothesized that there would be a difference in size between the two species. Figure 3 shows the graphs of the two distributions of body size by species. CODAP displayed a line where the median is for each distribution.

The students concluded that the *M. celero* species are generally larger than *S. frontalis*. These findings led the students to wonder whether the life spans of the two species were different: "¿Las libélulas y las damiselas tienen el mismo largo de vida?" *"Do dragonflies and damselflies have the same life span?"*

- **Hydrology:** The data on the hydrology and meteorology of El Yunque were constructed from data collected by the United States Geological Survey (USGS) in different bodies of water in El Yunque and data collected by scientists at the El Verde Field Station in El Yunque (González-Orellana, 2023a). There are four rivers represented in this dataset. Each river originates in El Yunque. After leaving El Yunque, the Río Mameyes, Río Grande, and Río Espíritu flow through the city of Río Grande to the north of El Yunque. The Río Icacos flows through the city of Naguabo to the southeast of El Yunque. The hydrology dataset includes monthly values for rainfall, water discharge, and water levels at the different rivers from 1990 to 2022. We are able to provide the full dataset since the data are aggregated to monthly values. One group of students who presented at the symposium in Spanish developed a scientific question about the relationship between rainfall and streamflow: "¿Existe una relación entre la descarga y la lluvia en el Río Icacos?" *"Is there a relationship between discharge and rainfall in the Rio Icacos?"* The students developed an interest in studying the river as an outgrowth of the learning activities that involved the hydrology dataset. Figure 4 shows the scatterplot and least squares line that the students produced in CODAP.

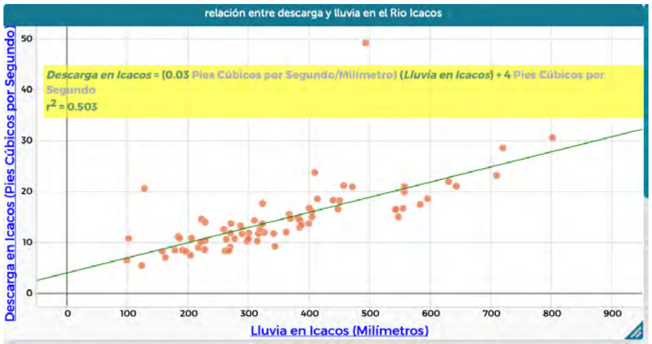



**Figure 4.** Student graph showing the relationship between rainfall and discharge.

The students concluded that there is a moderate relationship between rainfall and discharge based on the R2 value of 0.503. The students pointed out that the outlier value with significantly higher discharge is from September 2017 when Hurricane Maria struck Puerto Rico. These findings led the students to wonder how discharge affects the shrimp in the river: "¿Cómo la descarga del Río Icacos afecta la población de camarones" *"How does the discharge of the Rio Icacos affect the shrimp population?"*

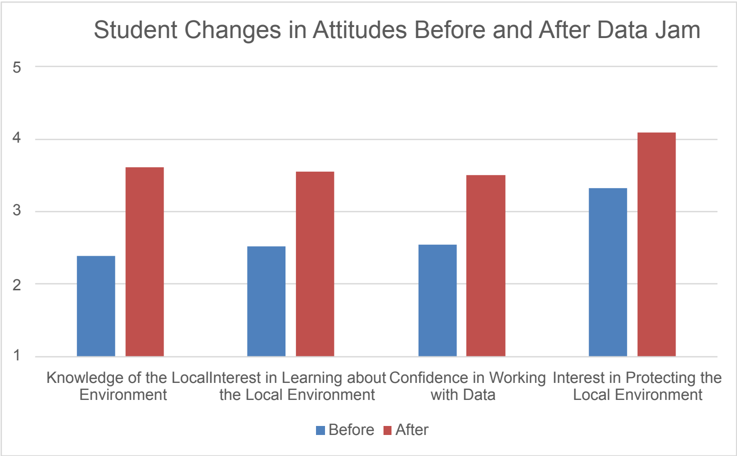



**Figure 5.** Student changes in attitude before and after the Data Jam.

## Student Reactions

All participating students were surveyed about their experiences with Data Jam. Students felt the most confident in working as a team to choose a dataset and create the graphs. The students felt the least confident in analyzing the data and developing their scientific argument. Overall, students reported a significant increase in their knowledge of the local environment, their interest in learning more about the environment, their confidence in working with data, and their interest in protecting the local environment (Figure 5).

Forty percent of the students indicated that participating in Data Jam increased their interest in majoring in a science discipline when they go to college. We interviewed a randomly selected subset of students who participated in the symposium. Many of the comments about their experiences with Data Jam were consistent with the survey results. Below are representative student quotes.

| **What did you like the most about Data Jam?** | |
|---|---|
| *Lo que más me gustó de Data Jam fue graficar, porque ahí es donde podías sentido a lo que estábamos tratando de graficar. Y también el diálogo con los científicos de ayer. Me hicieran sentir exitoso, como si lo que hice valiera la pena y que descubriera algo importante.* | What I liked most about Data Jam was graphing, because that's where you could make sense of what we were trying to graph. And also, the dialogue with the scientists yesterday. They made me feel successful, like what I did was worthwhile and that I discovered something important. |
| *Lo más importante eran los gráficas buscar los puntos para poder conectarlos. Eso es lo que más me gustó.* | The most important thing was the graphs, looking for the dots to be able to connect them. That's what I liked the most. |
| **How has Data Jam changed your view of yourself as a scientist?** | |
| *Siento que encajo, porque al final, son datos que ya existen, pero lo estoy haciendo, y lo analizo con mi forma de pensar y cómo interpreto los datos.* | I feel like I fit, because at the end of the day, it's data that already exists, but I'm doing it, and I analyze it with my way of thinking and how I interpret the data. |
| *Siempre he dicho, que he querido ser más como un médico. Ahora estoy cambiando. Sobre todo mis amigos me dicen "últimamente te he imaginado más como un científico, como si quisieras encontrar una solución a todo". Mis padres se dan cuenta. Ayer llegué con mi afiche y se lo expliqué. Como que ven que esa parte de la ciencia me emociona más.* | I've always said that I wanted to be more like a doctor. Now I'm changing. Especially my friends tell me "lately I've imagined you more like a scientist, like you want to find a solution to everything". My parents notice. Yesterday I came in with my poster and explained it to them. They kind of see that that part of science excites me more. |

## Professional Development and Implementation

Data Jam aims to impact math and science teachers from the public and private education system in Puerto Rico. The professional development model (Delgado-Quiñones et al., 2023) aligns to the five key features of effective professional development outlined in Desimone and Garet (2015)'s empirical framework and includes:

1. Sustained duration of more than 20 hours of professional development,
2. A focus on disciplinary content through investigations of data,
3. Actively engaging teachers in learning with and reflecting on data,
4. Supporting coherence with course goals, and
5. Fostering collaboration among teachers.

The Data Jam professional development model includes a total of 34.5 hours of professional development spread over a weeklong summer workshop (30 hours) and three 90-minute Virtual Lab Meetings. Math and science teachers from the same school are encouraged to participate in Data Jam together. At the summer workshop, teachers actively engage in Data Jam activities using a "student hat" (Lowell & McNeill, 2020) to develop an understanding of the phenomena, the datasets, and CODAP as well as develop a sense of what the student experience is like. Taking on the role of student during the workshop, teachers engage in active learning to develop an understanding of data analysis and interpretation of graphs. The features of CODAP enable teachers and their students to interrogate data by making it easy to organize data, visualize distributions and see relationships in the data. We provide the metadata and background information for each dataset so that the datasets are easier for teachers and students to understand and use in their analyses. Throughout the summer workshop, teachers engage in a sequence of activities related to (a) using box plots to characterize the center, spread, and shape of the distribution of a single variable, (b) using box plots to compare two distributions of data, (c) using percentages to compare categorical data, (d) using scatterplots to examine the relationship between two numerical variables, and (e) using time series graphs to examine changes in a numerical variable over time.

During the workshop, teachers complete three formative challenge activities similar to what the students complete during the Data Jam implementation. The teachers then participate in a collaborative protocol for examining student work (discussed below). The weeklong workshop culminates with teachers completing a mini Data Jam project similar in structure to the students' Data Jam projects (see Figure 6).

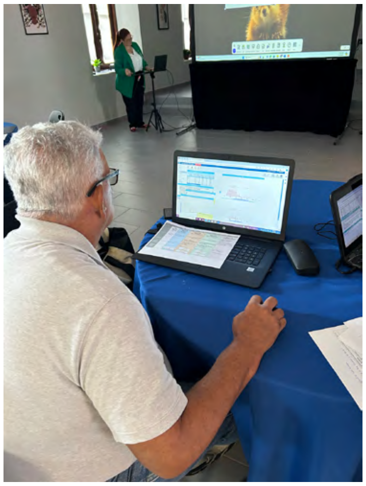


**Figure 6.** A Data Jam teacher completing the mini Data Jam Project.

**Figure 7.** A Luquillo graduate student provides feedback on the mini Data Jam Project.

Teachers are provided with structured planning time to make decisions about where the Data Jam lessons fit within their curriculum. Each year, the Data Jam team recruits LTER scientists who serve as Data Jam mentors during the professional development. Throughout the workshop experiences, the Data Jam mentors give presentations to the teachers about their current research work and their career trajectory that brought them to where they are today. The mentors provide feedback to teachers on their mini Data Jam projects in preparation for supporting students' projects (Figure 7).

## Data Jam Implementation

Teachers implement the sequence of learning activities that they experienced in the workshop during the next fall semester. About once per month from September to November, the students complete a formative challenge task. These challenge tasks are administered online and serve a formative assessment purpose. Each challenge contains an investigation question, brief background information, and a link to the necessary CODAP dataset. Students demonstrate their data science abilities by conducting the necessary analyses and uploading their graph. Students submit a written argument that uses the data as evidence to respond to the investigation question. The expectations of the challenge tasks provide a vision of what skills teachers need to develop with their students prior to the challenge task administration window.

The challenge tasks are also a component of the professional development plan. The responses to the challenge tasks are collected by the Data Jam team and a random sample of the responses is selected for discussion with teachers. The teachers meet in a cross-school Virtual Lab Meeting (VLM) with a small group of their peers to examine the student work using a validated looking at student work protocol (McDonald et al., 2013). The protocol ensures that teachers have equal amounts of time to contribute to the discussion. The VLMs are facilitated by a Data Jam team member with the active participation of a science mentor. The protocol provides a systematic means for teachers to focus on how the work demonstrates what students know, what they still need to learn, and stimulates discussion of how to support student learning in subsequent lessons. At the end of the 45-minute student work protocol, teachers have 45 minutes of unstructured time with the scientific mentors to get feedback about their implementation of Data Jam and ask questions about the phenomena underlying the datasets. This unstructured discussion between the teachers and mentors is designed with the intention of clarifying statistical concepts, data analysis strategies, and interpretations.

**Figure 8.** Science mentors and teachers provide feedback to students at the symposium.

The scientist mentors provide in-person or virtual presentations to the students that are focused on science identity, STEM careers, and each scientist's particular research. These presentations provide Data Jam participants with a vision of what Puerto Rican scientists do and what opportunities are available. The mentors' engagement and familiarity with the students' projects throughout the implementation as well as the students' familiarity with the mentor's research and practice enhances the level of interaction when the students and mentors meet at the symposium (Figure 8).

## Teacher Reactions

We conducted focus groups with new and experienced Data Jam teachers to identify strengths and recommendations following the 2023-2024 implementation of Data Jam. Unlike individual interviews, focus groups allow researchers to explore the diverse individual perspectives of participants, enriched by the interaction among them during the discussion. The questions aimed to explore experiences related to the main components of Data Jam: summer workshop, VLMs, Data Jam activities, and the mentor presentations.

- Teachers reported that during the workshop, the opportunities to collaborate were particularly beneficial, especially when science teachers interacted with mathematics teachers, as this facilitated the sharing of interdisciplinary knowledge.
- Participants noted that the examples of active practices (such as activities, exercises, and the use of the CODAP platform) incorporated into the workshop design were crucial for enhancing their understanding of concepts and facilitating the practical application of the content taught.
- Teachers expressed appreciation for the efforts of the staff and presenters in fostering a collaborative and supportive learning environment.

Participants expressed that the VLMs were highly positive and beneficial, particularly during mentor-teacher interactions. Engaging in this type of exercise helped increase teacher confidence in their teaching practices.

- The teachers felt that the science mentors supported and facilitated the transfer of workshop materials into the classroom and enhanced their statistical knowledge.
- Teachers highlighted the relevance of data in the activities and the CODAP platform. They believe that one of the greatest benefits of the project is that students can interact with data contextualized to Puerto Rico.
- Teachers noted that the CODAP platform has a user-friendly interface for students and is easy to navigate.
- Teachers felt that these activities foster student empowerment, leadership skills, critical thinking, question formulation, effective communication, and research skills and developed their public speaking skills and the ability to defend their research perspectives.
- The teachers reported that the interaction between students and mentors was very positive. Students really enjoyed the presentations and were very excited about having the opportunity to interact with a real scientist. One of the highlights was that teachers believed that the presentations awakened students' curiosity about science.
- The teachers expressed a desire to have the science mentors give direct feedback on students' final projects prior to the symposium, which will be implemented for the 2024-2025 Data Jam.

Teachers reported that some of the primary implementation barriers included (1) schools undergoing reorganization which changed some of the teacher's roles and reduced the class period length from 90 minutes to 50 minutes, (2) slow internet connection at their schools, (3) feeling anxious about not being able to complete the challenge tasks by the due date, (4) interruptions of unplanned school activities, and (5) lack of director support.

## CODAP

A significant part of each student project is an analysis of a complex ecological data set. Understanding the evidence these data provide requires creating, manipulating and interpreting graphical representations. Data Jam uses [CODAP](https://codap.concord.org) ([https://codap.concord.org](https://codap.concord.org)), a free, web-based visualization and analysis tool that has been specifically designed to facilitate this process for students in middle school and high school. CODAP makes it easy to create and modify graphs with a drag-and-drop interface, so that students can quickly engage in interpreting data, rather than laboriously plotting points. CODAP allows students to build graphs one step at a time, rather than having to declare that they want a particular type of graph (e.g. "a bar graph") before they begin. With CODAP, it is simple to create multiple graphs from the same dataset and to see how a case is represented across graphs. For example, Figure 9 displays data about dragonfly (Odonata) larvae found in two different habitats.

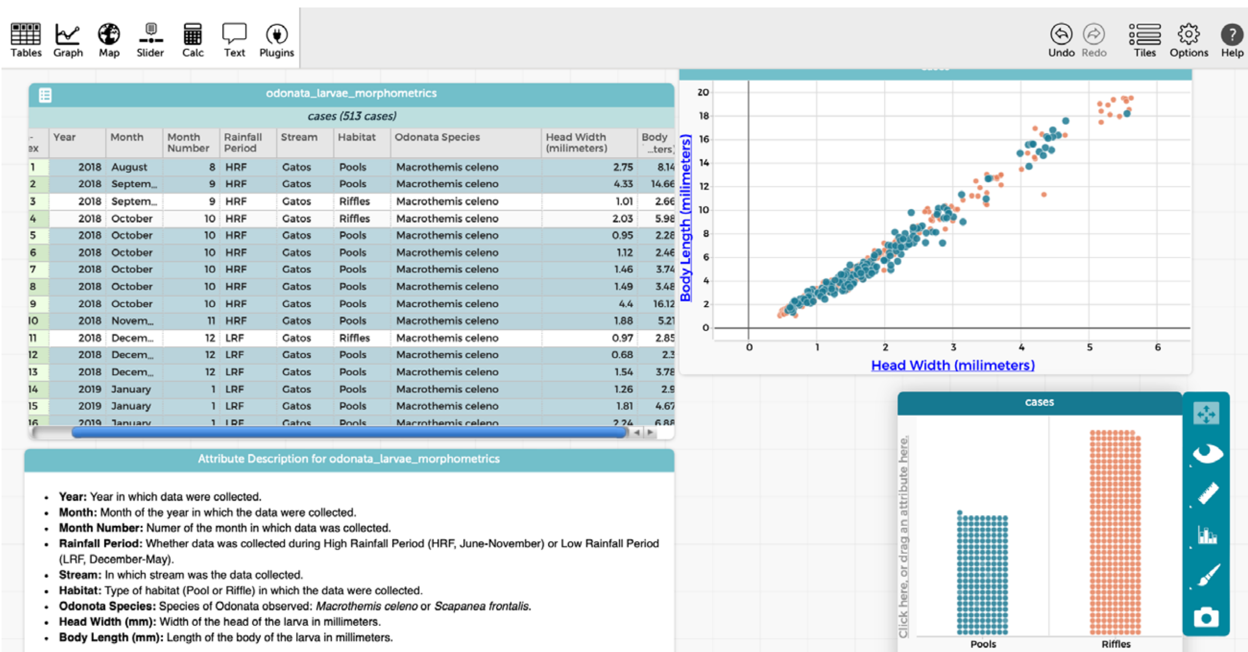



**Figure 9.** CODAP screenshot of multiple graphs related to Odonata larvae.

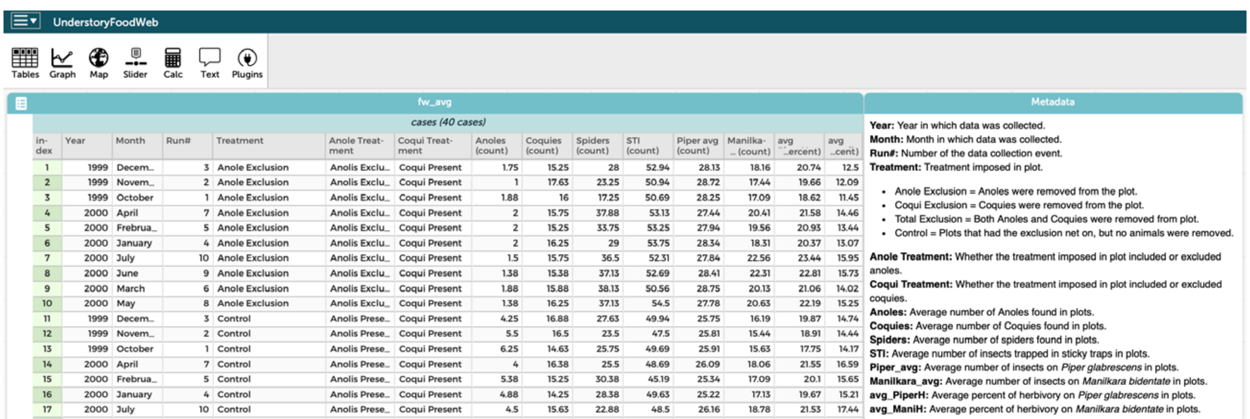
In each of the graphs, a dot represents a larva. The top graph shows the relationship between the size of each larva's head (plotted on the X-axis) and the length of the larva (plotted on the Y-axis). This graph indicates a clear relationship between the two variables; larvae that are longer tend to have larger heads. The bottom graph indicates the habitat where each larva was found, either in pools or in riffles (a shallow place in a river where water flows quickly past rocks); more larvae were found in riffles than in pools. Having both graphs on the screen at once, along with the table and the metadata, allows students to engage more deeply with the meaning of the data. But even more powerful is CODAP's ability to connect graphs together. Note in the bottom graph that the user selected all the points representing the larvae in the pools, which colored the dots blue. The dots representing larvae in the riffles are colored in the default orange. In CODAP, all the representations are linked. The selected dots in the bottom graph are also highlighted in the scatterplot and the data table. The same coloring is used across the representations, that is, all the larvae found in pools are blue and those found in riffles are orange. This allows students to consider the relationship among all three variables: head size, length and habitat. They may notice by examining the scatterplot that a group of the largest larvae are colored orange - that is, they were found in riffles. This observation, in turn, might lead them to directly compare the lengths of larvae found in ponds with those found in riffles with a new graph. This is the kind of data exploration that CODAP supports and which we hope to encourage students to engage in, where one data representation suggests a question that might be pursued with a new data representation, rather than the more common model in which there is a single correct graph to be constructed and a single correct answer.

## Dataset Development

Given the complexity and size of the LTER datasets, it is necessary to create manageable datasets for students. However, an oversimplification of datasets for educational purposes is not necessarily the best practice. Small or short datasets might be perceived as boring, and might limit students on their data inquiry journey, preventing them from addressing their questions and exploring different variables (Rubin, 2021; NASEM, 2023). With the advancement of software and its increasing accessibility, it is possible to engage students with larger and more complicated datasets, since software applications like CODAP permit students to work with such datasets (Rubin 2021; Mokros et al., 2023; NASEM, 2023).

Educators have shied away from usage of the readily available raw datasets, no matter how well curated they are, as it is imperative for students to fully understand the data to come up with appropriate questions and analyses (Rubin, 2021; Mokros et al., 2023). Datasets developed by Data Jam are derived from publicly available datasets collected by scientists in the LUQ-LTER and from other federal and state agencies. We limit the datasets to no more than 5000 cases. Apart from the performance of CODAP being affected after 5000 cases, we also consider that larger datasets might be overwhelming for the students. The number of variables per dataset varies and might depend on the instructional purpose of the dataset. The first step involves creating a master dataset from the scientific dataset and identifying the subset of variables that would be most relevant for students. If the dataset contains more than 5000 cases, we decide whether to aggregate the data or to focus on a specific time period. We also create variable names that will make sense to students and develop student versions of the metadata description. From this master dataset, we can derive multiple versions for different purposes.

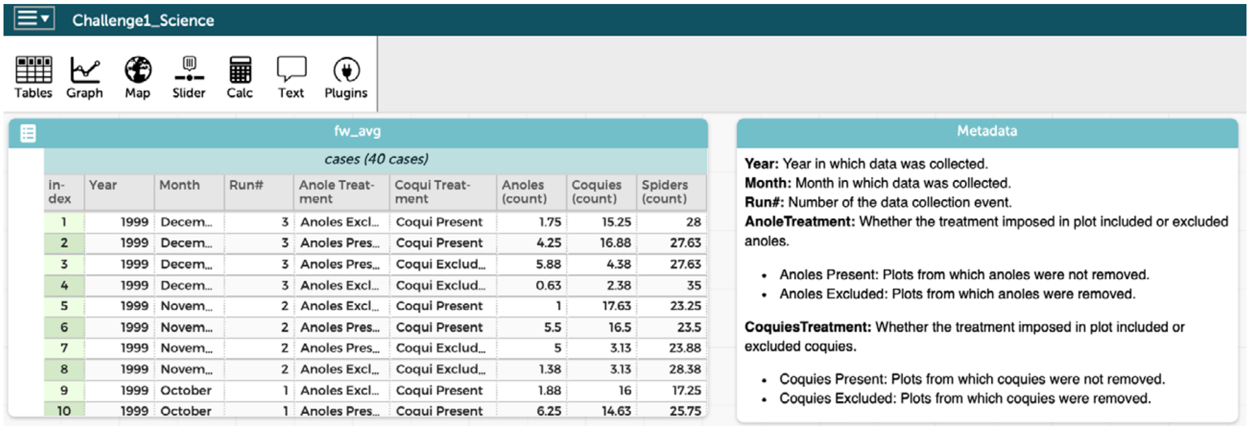
As students work with their teachers on learning and developing their data science skills, they go through worksheets and challenges aimed to polish different data science skills (e.g. single distributions, comparing distributions, categorical comparisons, numerical relationships, and time series). For each of these activities, a dataset might be derived that contains the elements needed for the student to work on a particular skill, without overwhelming them with information not needed at that point of their progress. For example, Figure 10 shows the master dataset for Understory Food Webs in El Yunque, which has 14 attributes.




| index | Year | Month | Run# | Treatment | Anole Treatment | Coqui Treatment | Anoles (count) | Coquies (count) | Spiders (count) | STI (count) | Piper avg (count) | Manilka- ... (count) | avg ...ercent) | avg ...cent) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1999 | Decem... | 3 | Anole Exclusion | Anolis Exclu... | Coqui Present | 1.75 | 15.25 | 28 | 52.94 | 28.13 | 18.16 | 20.74 | 12.5 |
| 2 | 1999 | Novem... | 2 | Anole Exclusion | Anolis Exclu... | Coqui Present | 1 | 17.63 | 23.25 | 50.94 | 28.72 | 17.44 | 19.66 | 12.09 |
| 3 | 1999 | October | 1 | Anole Exclusion | Anolis Exclu... | Coqui Present | 1.88 | 16 | 17.25 | 50.69 | 28.25 | 17.09 | 18.62 | 11.45 |
| 4 | 2000 | April | 7 | Anole Exclusion | Anolis Exclu... | Coqui Present | 2 | 15.75 | 37.88 | 53.13 | 27.44 | 20.41 | 21.58 | 14.46 |
| 5 | 2000 | Frebrua... | 5 | Anole Exclusion | Anolis Exclu... | Coqui Present | 2 | 15.25 | 33.75 | 53.25 | 27.94 | 19.56 | 20.93 | 13.44 |
| 6 | 2000 | January | 4 | Anole Exclusion | Anolis Exclu... | Coqui Present | 2 | 16.25 | 29 | 53.75 | 28.34 | 18.31 | 20.37 | 13.07 |
| 7 | 2000 | July | 10 | Anole Exclusion | Anolis Exclu... | Coqui Present | 1.5 | 15.75 | 36.5 | 52.31 | 27.84 | 22.56 | 23.44 | 15.95 |
| 8 | 2000 | June | 9 | Anole Exclusion | Anolis Exclu... | Coqui Present | 1.38 | 15.38 | 37.13 | 52.69 | 28.41 | 22.31 | 22.81 | 15.73 |
| 9 | 2000 | March | 6 | Anole Exclusion | Anolis Exclu... | Coqui Present | 1.88 | 15.88 | 38.13 | 50.56 | 28.75 | 20.13 | 21.06 | 14.02 |
| 10 | 2000 | May | 8 | Anole Exclusion | Anolis Exclu... | Coqui Present | 1.38 | 16.25 | 37.13 | 54.5 | 27.78 | 20.63 | 22.19 | 15.25 |
| 11 | 1999 | Decem... | 3 | Control | Anolis Prese... | Coqui Present | 4.25 | 16.88 | 27.63 | 49.94 | 25.75 | 16.19 | 19.87 | 14.74 |
| 12 | 1999 | Novem... | 2 | Control | Anolis Prese... | Coqui Present | 5.5 | 16.5 | 23.5 | 47.5 | 25.81 | 15.44 | 18.91 | 14.44 |
| 13 | 1999 | October | 1 | Control | Anolis Prese... | Coqui Present | 6.25 | 14.63 | 25.75 | 49.69 | 25.91 | 15.63 | 17.75 | 14.17 |
| 14 | 2000 | April | 7 | Control | Anolis Prese... | Coqui Present | 4 | 16.38 | 25.5 | 48.69 | 26.09 | 18.06 | 21.55 | 16.59 |
| 15 | 2000 | Frebrua... | 5 | Control | Anolis Prese... | Coqui Present | 5.38 | 15.25 | 30.38 | 45.19 | 25.34 | 17.09 | 20.1 | 15.65 |
| 16 | 2000 | January | 4 | Control | Anolis Prese... | Coqui Present | 4.88 | 14.25 | 28.38 | 49.63 | 25.22 | 17.13 | 19.67 | 15.21 |
| 17 | 2000 | July | 10 | Control | Anolis Prese... | Coqui Present | 4.5 | 15.63 | 22.88 | 48.5 | 26.16 | 18.78 | 21.53 | 17.44 |



**Figure 10.** Understory Food Webs in El Yunque "original" dataset developed for the Data Jam (CODAP File)



Challenge1_Science

Tables Graph Map Slider Calc Text Plugins

fw_avg

cases (40 cases)

| in-dex | Year | Month | Run# | Anole Treat-ment | Coqui Treat-ment | Anoles (count) | Coquies (count) | Spiders (count) |
|---|---|---|---|---|---|---|---|---|
| 1 | 1999 | Decem... | 3 | Anoles Excl... | Coqui Present | 1.75 | 15.25 | 28 |
| 2 | 1999 | Decem... | 3 | Anoles Pres... | Coqui Present | 4.25 | 16.88 | 27.63 |
| 3 | 1999 | Decem... | 3 | Anoles Pres... | Coqui Exclud... | 5.88 | 4.38 | 27.63 |
| 4 | 1999 | Decem... | 3 | Anoles Excl... | Coqui Exclud... | 0.63 | 2.38 | 35 |
| 5 | 1999 | Novem... | 2 | Anoles Excl... | Coqui Present | 1 | 17.63 | 23.25 |
| 6 | 1999 | Novem... | 2 | Anoles Pres... | Coqui Present | 5.5 | 16.5 | 23.5 |
| 7 | 1999 | Novem... | 2 | Anoles Pres... | Coqui Exclud... | 5 | 3.13 | 23.88 |
| 8 | 1999 | Novem... | 2 | Anoles Excl... | Coqui Exclud... | 1.38 | 3.13 | 28.38 |
| 9 | 1999 | October | 1 | Anoles Excl... | Coqui Present | 1.88 | 16 | 17.25 |
| 10 | 1999 | October | 1 | Anoles Pres... | Coqui Present | 6.25 | 14.63 | 25.75 |

Metadata

**Year:** Year in which data was collected.
**Month:** Month in which data was collected.
**Run#:** Number of the data collection event.
**AnoleTreatment:** Whether the treatment imposed in plot included or excluded anoles.

- Anoles Present: Plots from which anoles were not removed.
- Anoles Excluded: Plots from which anoles were removed.

**CoquiesTreatment:** Whether the treatment imposed in plot included or excluded coquies.

- Coquies Present: Plots from which coquies were not removed.
- Coquies Excluded: Plots from which coquies were removed.

**Figure 11.** Dataset derived from the "original" Understory Food Webs (Figure 10) dataset for Challenge 1.

Figure 11 shows a dataset derived from the master Understory Food Webs dataset aimed specifically for one of the challenges where students are prompted to answer the question: Do coquíes eat spiders?

This challenge is the first time students are required to answer a research question on their own using the data, as well as the metadata and background information. Providing them with a master dataset where they would have 14 variables to choose from in their first attempt at choosing the correct variables to answer a research question could be overwhelming. Instead, students are provided with a smaller dataset (Figure 11), where we still can challenge them and evaluate their ability to choose the correct variables, without having too many distractions that could set up the students for failure. When working with publicly available datasets – and not datasets that students collected themselves—researchers have pointed out the importance of students asking "who, how, when, where, and why" to ensure their understanding and proper treatment of the data (Rubin, 2021). In Data Jam, students are provided with datasets that are composed of at least two files: the CODAP file with the data, and a metadata file with an explanation of each of the variables and how they were measured. The metadata is both in the CODAP file and on a separate document, so students can access it however they feel more comfortable. Datasets on scientific topics that tend to be more complicated have a third file included: the background information document. The background information document has information on the topic studied in the dataset, as well as the research project and study design through which the data was collected. Providing teachers and students with these files all together gives them the tools needed for them to deeply explore the datasets, come up with proper questions and analyses and, when answering their questions, develop a reasoning statement that explains their answer. Datasets are created with different levels of complexity (from the derived and specifically aimed datasets, to the master datasets) to guide the students through their data science journey. The final goal is for the students to use the more complex master datasets when they are developing their Data Jam projects. All datasets are organized and publicly available through the Data Jam Webpage under Datasets (https://datajam.cloud/en/datasets).

## Scientific Community of Practice

Data Jam is one of several outreach initiatives in the Luquillo LTER Schoolyard program. Our overarching goal for the Luquillo LTER Schoolyard program is to engage students as junior members of the Luquillo LTER community. Data Jam exemplifies the five hallmarks of this community of practice approach to science outreach (McGee et al., 2024). It supports identity development as a scientist by empowering students to investigate their own island and heritage, which is unusual for most educational programs that Puerto Rican students experience. The goals and research questions that Data Jam students undertake are shared in common with the Luquillo scientific community. The students can see that the scientific mentors are engaged in similar questions and see the pathways that can deepen their connection with the Luquillo LTER community. Students are using age-appropriate versions of the same Luquillo data and are engaging in age-appropriate Luquillo practices for analyzing the data and developing scientific arguments with an age-appropriate version of the scientific tools of the Luquillo community. Graphing tools are a staple of LTER research. Students are able to make connections between the graph representations and the meaning of those graphs for the question they are investigating.

The metadata and background documents provide a means for students to learn the scientific language of the Luquillo community. The development of the Data Jam projects for presentation to members of the Luquillo community at the symposium provides students an opportunity to apply the language they have learned. Through interactions with members of the community they have the opportunity to get feedback and refine their use of scientific language.

Data Jam fits within the overall learning sequence of the Luquillo LTER schoolyard program; Journey to El Yunque, a middle school science unit centered on the El Yunque Rainforest ecosystem, LTER DATA Jam, and Field Work Protocols, which provides hands-on experience with data collection protocols used by scientists at a student study site within the rainforest. In becoming members of a community, apprentices often learn the practices in the opposite order in which they conducted. In that context, students will complete the entire process of an investigation but focus on completing specific steps individually. In the Journey to El Yunque program, which precedes the opportunity to participate in Data Jam, students are provided with an investigation question, Luquillo long-term data and models about population dynamics and they are expected to develop a scientific explanation of how hurricanes directly and indirectly impact population dynamics (Easley et al., 2023; McGee & Zimmerman, 2016; McGee et al., 2018). In the case of Data Jam, students are provided with Luquillo long-term data and are given the opportunity to ask their own research question and analyze the data to develop a scientific explanation. With the Field Work Protocols, which comes after Data Jam, students are able to collect data and contribute to a long-term dataset of forest productivity in a plot near the Luquillo El Verde Field Station. Students can then ask their own research questions about tree biomass and analyze the long-term tree productivity data to develop a scientific explanation. As students develop their abilities with scientific practices, scientific tools, and scientific language, they are able to complete more steps in the inquiry process.

## Conclusion

Providing explicit opportunities for all students to learn data science skills is foundational for broadening participation in STEM. The Data Jam model has been successfully implemented in a variety of classrooms and has great promise for broader application. Over the years, Data Jam has reached thousands of middle and high school students. The participating students have primarily been Puerto Rican from a range of socioeconomic contexts encompassing a broad range of urban, suburban and rural communities throughout the island of Puerto Rico. Students are able to engage in authentic investigations with Luquillo data about El Yunque, which increases their knowledge of and interest in studying the local environment. The combination of studying their local environment and connecting with practicing Luquillo scientists increases students' confidence in using data and increases their identity development and interest in pursuing a science major.

The professional development model is not only useful for supporting Data Jam, but also in any classroom where students are exploring publicly available datasets. Teachers' comments and recommendations highlight the importance of effective professional development and ongoing support throughout the Data Jam project implementation process. The positive feedback on the activities, the use of the CODAP platform, and interactions with mentors emphasizes the significant impact on students' skill development and understanding of the concepts. For the island of Puerto Rico, Data Jam is a powerful opportunity for engaging students in authentic science and inspiring a new generation of scientists.

## Acknowledgment

The authors were supported in part by National Science Foundation grants DRL-2049061 and DRL-1821146 to The Learning Partnership. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NSF.

## References

Bestelmeyer, S. V., Elser, M. M., Spellman, K. V., Sparrow, E. B., Haan-Amato, S. S., & Keener, A. (2015). Collaboration, interdisciplinary thinking, and communication: new approaches to K–12 ecology education. *Frontiers in Ecology and the Environment,* 13(1), 37-43.

Delgado-Quiñones, I., McGee, S., McGee-Tekula, R., Báez Rodríguez, N., & Rubin, A. (2023). Supporting Teacher Learning in Data Jam: A design-based research of Teacher Professional Development Model. Paper Session presented at the Annual Meeting of the American Educational Research Association in Chicago, IL. https://doi.org/10.51420/conf.2023.4

Desimone, L. M., & Garet, M. S. (2015). Best Practices in Teachers' Professional Development in the United States. *Psychology, Society, and Education,* 7(3), 252-263.

Easley, Kathleen M., Zimmerman, Jess K., McGee, Steven, and McGee-Tekula, Randi. (2023). Scientific Communities of Practice: K–12 Outreach Model around Organism Responses to Repeated Hurricane Disturbances. *Ecosphere* 14(7):e4624. https://doi.org/10.1002/ecs2.4624

González-Orellana, N. (2023a). Hydrology of El Yunque. The Learning Partnership. Available online: https://doi.org/10.51420/data.2023.1

González-Orellana, N. (2023b). Odonata Larvae at El Yunque National Rainforest. The Learning Partnership. Available online: https://doi.org/10.51420/data.2023.4

González-Orellana, N. (2023c). Population of 3 Shrimp Species in Pool 0 of Prieta Stream in El Yunque with Environmental Variables (2012-2019). The Learning Partnership. Available online: https://doi.org/10.51420/data.2023.5

Lowell, B. R., & McNeill, K. L. (2020). Using the Student Hat to Push on Multiple Goals in Teacher Professional Learning Paper presented at the International Society of the Learning Sciences, Nashville, TN.

McDonald, J. P., Mohr, N., Dichter, A., & McDonald, E. C. (2013). *The power of protocols: An educator's guide to better practice:* Teachers College Press.

McGee, S. & Zimmerman, J.K. (2016). Taking students on a Journey to El Yunque. *International Journal of Designs for Learning*, 7(1), 86-106. https://doi.org/10.14434/ijdl.v7i1.19429

McGee, S., & Báez-Rodriguez, N. (2017). Getting students jazzed about critical zone: Engaging students in authentic inquiry through Data Jam. *In the Trenches, 7*(4), 5-8. https://nagt.org/nagt/publications/trenches/v7-n4/engaging_studen.html

McGee, S., Durik, A. M., Zimmerman, J. K., McGee-Tekula, R., & Duck, J. (2018). Engaging Middle School Students in Authentic Scientific Practices Can Enhance Their Understanding of Ecosystem Response to Hurricane Disturbance. *Forests* 9(10), 658. https://doi.org/10.3390/f9100658

McGee, S., Zimmerman, J. K., Tekula, R. M., González-Orellana, N., & Báez Rodríguez, N. (2024). Scientific communities of practice: A K12 outreach model that supports authentic scientific inquiry around very long-term studies. Online *Journal of Ecology and Environmental Science*, 2(1). https://irispublishers.com/ojees/pdf/OJEES.MS.ID.000533.pdf

Mokros, J., Rubin, A., Sagrans, J. & Higgins, T. (2023). Curating Datasets to Support Middle School Students Inquiry. In: EM Jones (Ed.), Fostering Learning of Statistics and Data Science Proceedings of the Satellite conference of the International Association for Statistical Education (IASE) (pp. 31-36). IASE. https://iase- web.org/Webinars.php?p=230711_2000

National Academies of Sciences, Engineering, and Medicine. (2023). Foundations of Data Science for Students in Grades K-12: Proceedings of a Workshop. Washington, DC: The National Academies Press. https://doi.org/10.17226/26852.

Rubin, A. (2021). What to consider when we consider data. *Teaching Statistics*, 43, S23-S33.

## About the Authors

**Steven McGee** is president of The Learning Partnership, an independent research organization that partners to inspire a joy of learning for all children. He serves as the outreach director for the NSF-funded Luquillo Long-Term Ecological Research program in Puerto Rico, his maternal family's homeland. Steven has more than three decades of expertise in fostering long-term partnerships between researchers and practitioners. As the first Puerto Rican learning scientist in the nation, he studies how to support learning from an individual cognitive perspective as well as from the perspective of systems and social contexts. He is a co-developer of the Journey to El Yunque curriculum (elyunque.net) and Tree Guardians (treeguardians.net) curricula and co-author of The Mighty Hurricane interactive children's book (themightyhurricane.net). He has formerly served as a Northwestern University faculty member in the School of Education and Social Policy, Director of Technology and Assessment at Loyola University, and Chief R&D Officer for the NASA Classroom of the Future. He can be reached at mcgee@lponline.net

**Noelia Báez Rodríguez** obtained her bachelor's degree (2000) and master's degree (2005) in Geology from the University of Puerto Rico, Mayaguez Campus, where she worked describing the Stratigraphy of the Silurian-Devonian Upper Helderberg Group in the state of West Virginia. In her first professional steps, she participated in several internships with the United States Geological Survey (USGS) in Reston, Virginia and then worked with The Johnson Company of Vermont, as a field technician in watershed restoration and Super fund projects. Upon her return to the island in 2010, she began her participation as coordinator of the Schoolyard program of the LUQ LTER of the Department of Environmental Sciences of the UPR Rio Piedras Campus; which today she represents with great enthusiasm as it aligns many of her professional and personal interests. Since 2020, it has provided educational and coordination services to the Forward Learning company in projects related to education in Data Science and scientific research to students and teachers at the middle and high school levels of public schools and private schools of Puerto Rico. She can be reached at nbaez@ites.upr.edu

**Cárorin Castro-Rivera** earned a Bachelor of Science degree with a concentration in Microbiology and an M.A.Ed. in Educational Management and Leadership. She recently completed a Ph.D. in Teaching with a specialization in Educational Research. During her doctoral studies, she had the opportunity to work as a principal investigator in qualitative, quantitative, and mixed methods research on topics related to Visible Learning, the Relationship between Competencies and Global Mobility, and Blended Learning. She is an educational consultant for public and private entities, a workshop leader, and a specialist in the Blended Learning teaching model. She has 18 years of teaching experience at the school level and, for the last 5 years, has been a professor in the undergraduate and graduate programs at the Inter-American University of Puerto Rico. For 13 years, she has been part of the Forward Learning team, focusing on the development of 21st-century skills, technology integration, and the development of STEM projects in Makerspaces. She is currently working as an Educational Projects Coordinator and Educational Researcher. In 2023, she began her collaboration in the Data Jam project as a coordinator on behalf of Forward Learning and, in 2024, she joined the research team. She can be contacted at ccastro@forwardlearning.com

**Normandie González-Orellana** is the Data Manager for the Data Jam Program. She obtained her B.Sc. in Biology from the University of Puerto Rico in Humacao and her M.Sc. in Biology from the University of Puerto Rico in Río Piedras. She enjoys working with data to come up with new insights. She believes that data is powerful, and that understanding data is imperative for people to navigate this new data driven world. This belief is why she has dedicated the last 3 years of her career to creating datasets that kids can use in school to develop their data skills. She can be reached at gonormandie@gmail.com

**Randi McGee-Tekula** is vice-president at The Learning Partnership, an independent research organization that partners to inspire a joy of learning for all children. Her career began in research support for immunology and molecular biology research at Northwestern and Stanford Universities. Since then, she has gained experience as a high school chemistry teacher, curriculum designer, and as an industrial project manager. Her passions have always directed her back towards education and providing meaningful learning opportunities for students. She partners with teachers, school administrators and scientists to co-design curriculum and professional development experiences. Her research interests include teacher change, measurement of teaching practices, and equity in STEM teaching. She is a co-developer of the Journey to El Yunque (elyunque.net) and Tree Guardians (treeguardians.net) curricula and co-author of The Mighty Hurricane interactive children's book (themightyhurricane.net). She can be reached at rmcgee@lponline.net

**Andee Rubin** is a mathematician, computer scientist, and learning scientist at TERC, an educational research and development non-profit, where she has been studying the growth of students' and teachers' reasoning about data for over 30 years, particularly as it is enabled by research-based tools for data education. She led the ViSOR (Visualizing Statistical Reasoning) project, which studied how middle and high school teachers used data visualization tools with their students. She has developed curriculum materials around data for students as young as 6, both in and out of school, as well as professional development materials for teachers and out-of-school facilitators. She was PI of the Data Clubs project, which developed three data modules for middle school students in informal contexts. She currently leads multiple projects creating materials to facilitate students' learning about data and doing research on students' successes and struggles as they explore large and complex datasets. She is committed to empowering all learners to use data to pursue social justice and serves as the "data diva" for the Data Jam project. She can be reached at andee_rubin@terc.edu.