Title: Supporting Teacher Learning in Data Jam: A Design-Based Research of Teacher Professional Development Model

Authors: Delgado-Quiñones, I., McGee, S., McGee-Tekula, R., Báez-Rodríguez, N. & Rubin, A

Abstract:

This paper discusses the development and research of a new professional development (PD) model to support teachers in Puerto Rico in engaging middle and high school students in independent data analysis research projects, using ecological data, known as Data Jam. The new PD model was designed using the cognitive apprenticeship model (Collins, Brown, & Newman, 1989) as the driving force for inquiry activities around data practices. This article examines how qualitative data analysis (Saldaña, 2021) informed the re-design of the new PD model as part of a design-based research methodology. The results of our study suggest that to support teacher learning of data practices, more conceptual information is needed to be integrated to our PD Model. This paper is part of a more extensive mixed-methods study that includes student data.

Delgado-Quiñones, Isabel – Postdoctoral Researcher at Forward Learning
email: idelgado@forwardlearning.com

McGee, Steven- President at The Learning Partnership
Email: mcgee@lponline.net

McGee-Tekula, Randi, Vice-president at The Learning Partnership
Email: rmcgee@lponline.net

Báez-Rodríguez, Noelia, LUQ-LTER Schoolyard Program Coordinator
Email: nbaez@ites.upr.edu

Rubin, Andee. - Senior Scientist at TERC
Email: andee_rubin@terc.edu

Suggested Citation:
Delgado-Quiñones, I., McGee, S., McGee-Tekula, R., Báez Rodríguez, N., & Rubin, A. (2023, April 13-16). Supporting Teacher Learning in Data Jam: a design-based research of Teacher Professional Development Model [Paper]. Paper Session presented at the Annual Meeting of the American Educational Research Association in Chicago, IL.
https://doi.org/10.51420/conf.2023.4

1. Introduction

Managing data in our everyday lives can be difficult. Yet, data are an essential part of our lives and a foundation for understanding environmental science. The Data Jam model provides a unique platform for improving how we teach data analysis and interpretation by engaging students in authentic environmental investigations using long-term ecological datasets (Bestelmeyer et al., 2015). Prior results of Data Jam implementation in Puerto Rico (2015-2017) indicate that Data Jam students are able to ask their own research questions, select and graph data to address their research questions, and draw conclusions from data (McGee & Báez-Rodríguez, 2017). However, students rarely apply basic data analysis techniques such as making comparisons based on distributional characteristics of the dataset or creating scatterplots, which elucidate relationships between variables. Instead, they rely on visual comparisons of their graphs. These past results point to a need to better support teachers around data practices related to data analysis.

Our project addresses this need by iteratively designing and analyzing a professional development model that engages teachers with the support they need to engage their students in data-based ecological research. Our research is guided by the following question: What kinds of professional development support do teachers need to be able to mentor students in the process of asking important scientific questions and engaging in data analysis to develop evidence to answer their research questions?

In the next sections, we provide context of the Data Jam model as part of the Luquillo-Long Term Ecological Research Program, further referred to as LUQ-LTER, a description of the previous Data Jam PD model (2015-2020), followed by a detailed description of the 2021 Data Jam PD model, as well as the theoretical framework that guided its design. Next, we present our design-based research methods, techniques, data sources and data analysis. Finally, we describe our findings that informed the re-design of the Data Jam PD model for the second year of the project. Future directions of our design-based study are also discussed.

2. Context of the Data Jam Model

The Long-Term Ecological Research Program (LTER) supported by the National Science Foundation is a network of 28 research programs that consist of diverse ecosystems distributed across the continental United States, Alaska, Antarctica and islands in the Pacific and the Caribbean. The LTER network is characterized by a "rich history of ecological inquiry and collaboration" (https://lternet.edu/) (See Map of LTER sites in Appendix A). In 1988, the LUQ-LTER research program was established with the goal of integrating studies of disturbance and forest structure and dynamics. This monitoring and research program was critical in evaluating the immediate and long-term effects of hurricanes in Puerto Rico (https://lternet.edu/site/luquillo-lter/).

Data Jam is part of LUQ-LTER Schoolyard Program educational activities. The professional development model for Data Jam has evolved since it was first implemented in 2015. The 2015-2020 Data Jam model included a professional development for teachers that consisted of a one-day workshop (6 hours), a mid-term workshop with the students (3 hours), to provide feedback to their research questions, and a final Data Jam Symposium, in which students shared their research poster with scientists, peers, and teachers. The length of the workshop experienced was constrained by the availability of funding. This model was implemented in Puerto Rico from 2015 to 2020. Recently, with support from the NSF ITEST program (DRL-2049061), we developed a renewed Data Jam PD model (2021) that consists of five-day professional development workshop (30-hour) for teachers, scientist mentorship for teachers during the process of implementation, scientist talks to students, and the final Data Jam Symposium (See Figure 2). This paper is part of the NSF research grant that allows us to collect data from participating teachers and students to develop and re-design the Data Jam PD model. This effort will help teachers and students from Puerto Rico develop data literacy skills and engage students in Luquillo science as they develop their scientific identity and STEM career readiness.

3. Description of the 2021 Data Jam PD Model

Teaching students how to analyze and interpret data is an essential part of the Data Jam model. Much research has been done over the last 30 years looking at the perceptual and cognitive processes involved in interpreting graphs (e.g., Carpenter & Shah, 1998; Cleveland & McGill, 1984; Freedman & Shah, 2002; Friel, Curcio & Bright, 2001; Lohse, 1993; Mautone & Mayer, 2007). Many have documented the difficulties students have describing relationships in graphs of statistical data (Lehrer & Schauble, 2004; Mayer, 1993; Shah & Carpenter, 1995, Shah & Hoeffner, 2002; Shah, Mayer, & Hegarty, 1999). Students usually face challenges in developing research questions that are robust enough to warrant an analysis of data. Teachers and students don’t often have the expertise in data analysis to make good use of all the available datasets.

Students that participate in Data Jam develop their own research questions that can be answered using the LTER and related datasets. Even though, students do not collect their own data, they are expected to become familiar with the metadata and be able to pose research questions and to answer them with the available datasets through proper data analysis. For Data Jam to be successful, the projects require students to engage in a variety of data practices that underlie the STEM field of data science (See Table 1). The data practices are part of the process of generating successful Data Jam projects.

Table 1- Data practices aligned to the Data Jam model

Data Practice	**Description**
Research Question	Statistical questions are ones that anticipate variability in the data and account for that variability in drawing conclusions. It is important to have context and purpose to the analysis. Distinguish between correlation and causation.
Metadata	Be able to describe the data and consider the structure of data when drawing conclusions. Describe the nature of the attribute under investigation, how it was measured and its units of measurement.
Set up Dataset	Decide what variables to include in their investigation from the breath of datasets available. This involves manipulating the data.
Visualizations	Identifying relationships and communicating results. Deciding which type of graph to use (e.g., scatter plot graphs). Using regression lines to describe the correlation.
Data Trends	Involves the computational practices of looking for patterns or anomalies. Defining rules to categorize data and identifying trends and correlations. This practice relies on organizing and visualizing the data.
Final Product	Summarize results and make a claim about the data analyses to provide evidence to answer the research question.

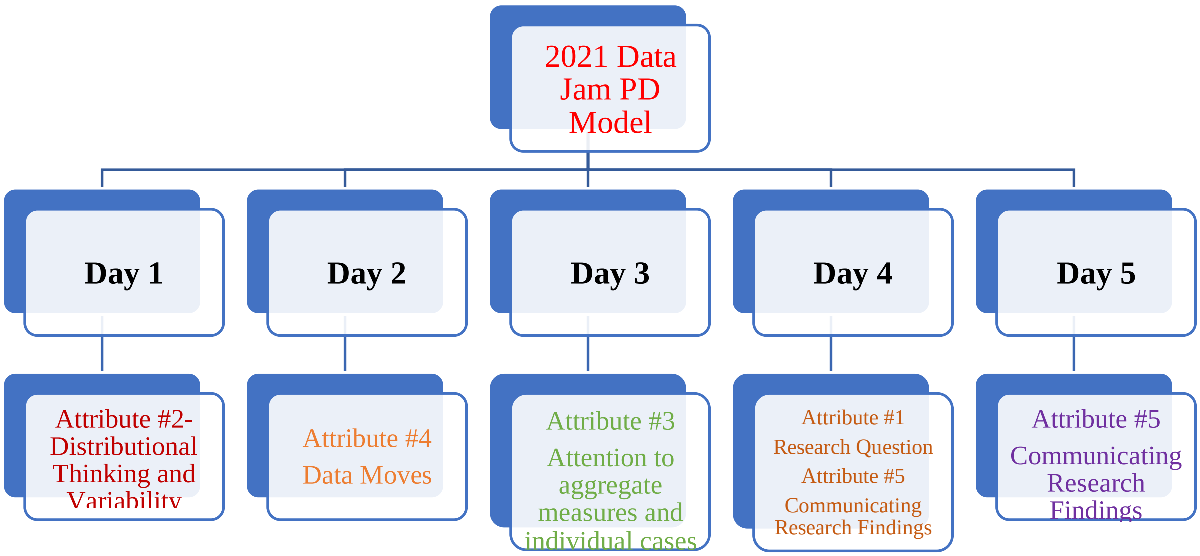
The purpose of the 2021 Data Jam PD Model was to build teacher capacity in these data practices for them to mentor their students in the process of completing the Data Jam posters as a final product. The activities designed for teacher professional development in the Data Jam PD model include (a) the five-day summer workshop and, (b) the three virtual lab meetings in which scientist mentors provide mentorship to teachers during the process of Data Jam implementation in their classrooms. A Data Jam framework (See Appendix B), developed in a previous grant and refined in this one, builds upon these data practices and was used to guide the structure of the daily data activities in the workshop (See Figure 2). For example, the data practice of research question is aligned to attribute #1 of the Data Jam Framework (See Appendix B), while the data practices of data trends and visualizations are aligned to attributes #2 and #3 of the Framework.

The data practices of set up dataset and metadata are aligned to attribute #4- Data moves- and the Final product data practice is aligned with attribute #5- communicating research findings. Each of the attributes in the 2021 Data Jam Framework refers to a critical step in the process of doing research with data:

1. Research question: Coming up with an appropriate research question requires knowledge of both the context of the data (e.g., the ecology of El Yunque, details about the location and elevation of watersheds) and the data themselves.
2. Distributional thinking and variability: One of the defining characteristics of data is variability. Any measurement or other attribute has a range of values that form a distribution. In ecological data, those values are likely to vary over time and place and may indicate weather events or other changes in the environment. Working with data requires understanding and dealing with variability.
3. Attention to aggregate measures and individual cases: Special statistical techniques exist for working with distributions of values. One basic technique is to use aggregate measures such as mean, median, and standard deviation to reflect the central tendency or variability of a distribution. Such measures, however, may hide interesting or anomalous individual cases, so it's important to keep both the actual distribution and individual cases in mind.
4. “Data moves” is a term coined by Erickson et al. (2019) to describe data manipulations such as selecting subsets of data, adding variables to a graph, filtering the dataset, etc. The Common Online Data Analysis Platform (CODAP) is the tool the students use in Data Jam to makes many of these data moves simple.

5. Communicating research findings: Students' final task was to create a poster with their research question, the relevant data, their data analysis, and their findings. Making an effective poster requires considering how best to communicate to an audience.

Figure 1- Summary of the Workshop and the attributes of the Data Jam Framework that guided the activities for each day

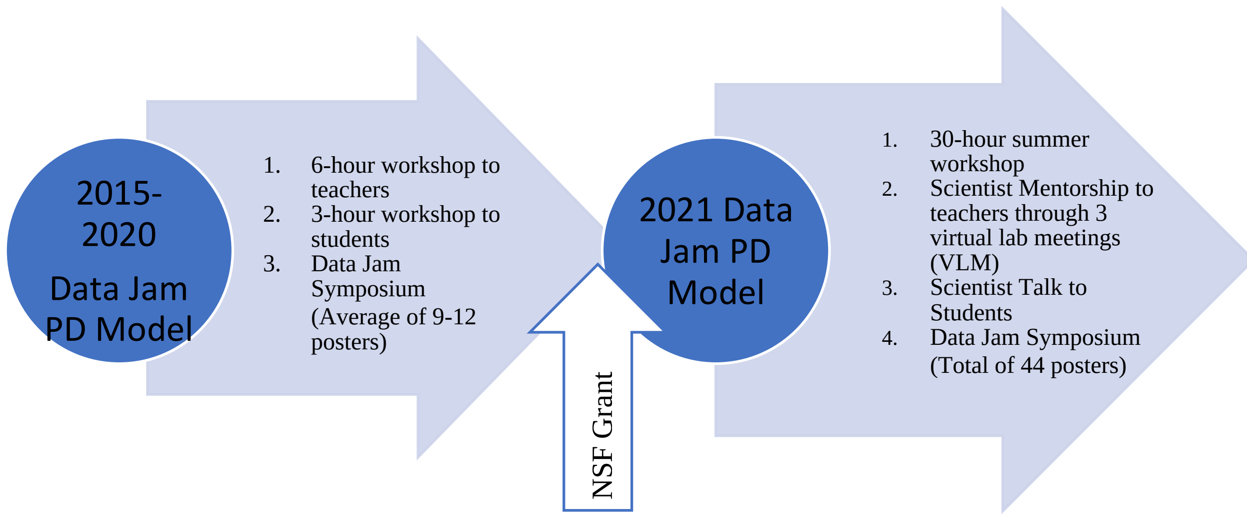


The design of the 2021 Data Jam PD model has been guided by three specific principles of the cognitive apprenticeship theory described by Collins et al. (1989): (a) learning often happens in the opposite order of how tasks are done, (b) making thinking visible and (c) engaging in authentic tasks. This served as a conceptual framework to design and structure the summer workshop activities that are rooted in the data practices necessary to conduct a successful Data Jam project. The cognitive apprenticeship model conceptualizes learning embedded in authentic activities that shapes the interactions between learners working in teams. Also, scaffolded assistance is necessary to support learning in this model (Cheng, 2014).

The elements of the 2021 Data Jam PD model include: 1) 5-day summer workshop, 2) Virtual Lab Meetings, 3) Scientist Talks to Students and 4) Data Jam Symposium (See Figure 2).

In the next sections, we describe the first two elements that involve the teacher professional development directly. The last two elements (Scientist Talks and Data Jam Symposium) are part of the larger study that involve both teachers and students.

Figure 2- Evolution of Data Jam Model- Luquillo Schoolyard Program



3.1- Description of the summer workshop activities of the 2021 Data Jam PD Model

The intent of the workshop was to spotlight data and build data literacy for teachers. By building capacity towards data fluency, the quality of research questions should improve as should the quality and relevance of evidence from analysis of data to answer the research question. We designed the following workshop flow for the learning outcome-oriented activities. The sequence was as follows: (1) given a research question and graph, now interpret graphs, (2) given a research question and data, now create and interpret graphs, (3) given a research question, now carry out data moves, create and interpret graphs, and (4) given three data sets to choose from, now develop an investigation question, carry out data moves, create and interpret graphs.

The activities conducted during the five day workshop utilized three datasets. There were two LUQ-LTER data sets: 1) a hydrology data set and 2) a shrimp data set. A third biological dataset from NEON (National Ecological Observatory Network) was also provided. Each dataset was derived from actual data collected by LUQ-LTER and NEON scientists and their research teams. These data sets were provided to teachers as a link in the CODAP interactive graphing tool (Frischemeier et al., 2021). The features of CODAP enable teachers (and their students) to interrogate data by making it easy to organize, visualize, and see relationships in and between data sets.

The hydrology data set has been used in previous Data Jams (2015-2020). As it had been used in many iterations of Data Jam, it was the most mature set. An initial shrimp data set was prepared specifically for the summer workshop to provide newly available data to returning teachers. The biological dataset from NEON (National Ecological Observatory Network) was made available as it had been used successfully in another research project with teachers in Puerto Rico. However, it was the most limited data set in terms of the number of years it covered. Teachers had opportunities to interact with these data sets as they participated in the different activities designed for each workshop day. Teachers then selected the data set from which to develop their final presentations.

During the workshop, teachers were provided support in the form of scaffolds, teaching strategies and routines to support the teaching and learning process. The use of the CODAP platform was scaffolded with challenge cards, which helped teachers to engage in active recall of CODAP functions for data moves and graph generation. The challenge cards are a deck of cards containing ‘challenges’ at three levels of difficulty. Each card ‘challenges’ the user to do a particular function in CODAP. If they get stuck, the back of the card describes how to carry the function out. The challenge cards contained images as well as words to improve active recall. At key moments throughout the workshop, teachers were challenged to apply what they had previously learned through a ‘see it - now make it activity.’ The facilitators would show one or more graphs on the presentation screen and teachers were asked to recreate the graphs using the skills that they had learned. They were provided hints by indicating which challenge cards were applicable. Andee Rubin (author), a data science educator, created screencasts that were presented at different points during the workshop as demonstrations of using CODAP to investigate research questions using a series of functions from the challenge cards.

The workshop flow imparted a variety of active learning and meaningful learning strategies to engage teachers throughout the week and to model strategies that could be used with their students. In addition to the “see it - now make it” activity described above, participants were continually asked, “what are you curious about?” to model how these situational curiosities can develop into and be refined into robust investigation questions over time. Participants worked independently, in pairs and in small groups around a particular goal. These groupings modeled equity strategies that teachers can use for sharing ideas and discussion. Participants were given an opportunity to participate in a ‘looking at student work’ protocol around prior Data Jam student projects to experience the value of this reflective routine in their practice. Each day one scientist mentor virtually presented their professional career journey and their research conducted in Puerto Rico to describe a variety of ecology career pathways, extend prior science knowledge and initiate community between mentor and participants. On the final workshop day, mentors were present at the workshop location to meet with teachers and watch as teachers presented their ‘mini’ Data Jam posters as their culminating workshop activity and provided feedback.

3.1 Description of Virtual Lab Meetings

The virtual lab meetings (VLM) were a part of the 2021 Data Jam PD model. They consisted t of a one 90-minute meeting (1st) and two 60-minute meetings (2nd and 3rd). The first meeting provided 30 mins of introductions and orientation, 15 minutes for each teacher, and 15 minutes for reflection. Subsequent meetings included 15 minutes for each teacher and 15 minutes for reflection. The information provided in the VLM is one of our data sources, because it was an opportunity for teachers to share the experiences that they are providing their students while implementing Data Jam and to receive support from mentors on how to manage and analyze data for them to guide the students. The VLM provided a platform where teachers could discuss these implementation differences, challenges and support one another. It also continued the opportunity to build community amongst them.

Three rounds of virtual lab meetings were scheduled during the 2021 fall semester as teachers were implementing Data Jam in their classrooms. Each virtual lab meeting group consisted of a mentor and three teachers. A total of five scientist mentors and fifteen teachers participated. A virtual lab meeting protocol inspired by a tuning protocol used by McDonald et al (2013) was designed to structure the VLM interactions to ensure equal opportunity for each teacher to receive mentor and teacher feedback. Aligned to the VLM protocol, a support ticket document (SurveyMonkey form) was completed by teachers one week before coming to the VLM. The support ticket provides valuable information about their status in the process of implementing Data Jam with their students and states what attributes from the Data Jam framework they are working on. Most importantly for the mentor, the support ticket communicates what issues a teacher will be asking for support around during the VLM so they can be better prepared to mentor.

4. Methods

4.1 Design Based Research (DBR)

Our methodological framework draws upon the Cobb et al. (2003) definition of design experiments in educational research, that is, “design experiments are conducted to develop theories” (p.9). Thus, the emphasis on theories is to showcase the view that the “explanations and understandings inherent in them are essential if educational improvement is to be a long-term, generative process” (Cobb et al., 2003, p.9). These authors argue that a design theory aims to explain why “designs work” and how they may be adapted to new contexts or circumstances (Cobb et al., p.9). In the case of Data Jam, we are developing a theory of supporting teacher learning of data practices. Design studies are characterized by two types of analysis: prospective and reflective. On the prospective side, design experiments are implemented with a learning process in mind and the supports needed, in order to expose those details to analysis. They provide the rationale for the design of a program. The reflective side include several levels of analysis that relies on longitudinal data, which in turn reveals its iterative nature (Cobb et al., 2003). In this paper, we focus our analysis on the prospective side as it includes data for one year (2021-2022) of the four-year design research study. The retrospective analysis will be conducted after collecting several years of data that will help us understand growth over time.

Our design-based research (DBR) study includes many of the characteristics of this type of study as described by Campanella and Penuel (2021). These are: future oriented, engaged in building knowledge and theory about learning and design, intended to be useful and iterative. The DBR approach provides a vehicle for combining principles of learning and iterating through multiple versions to refine the implementation to optimize outcomes while at the same time providing empirical evidence about the learning principles. Iteration is an essential characteristic of DBR. It is through this iterative process that “designs are improved, and knowledge and theory are “refined” (Campanella & Penuel, 2021 p. 8). Multiple iterations were conducted by the authors in the process of design and re-design of the Data Jam PD model. We describe these iterations by phases and discuss them in the Data Analysis section.

We followed a systematic qualitative methodology (Creswell, 2012; Merriam & Tisdell, 2016) focused on the inductive process of coding as one way of analyzing qualitative data (Saldaña, 2021). In the context of our first year of the project, we focus on how our main finding informed the changes to our initial design of the 2021 Data Jam PD model and what aspects of the model were effective, guided by our research question: What kinds of professional development support do teachers need to be able to mentor students in the process of asking important scientific questions and engaging in data analysis to develop evidence for their research questions?

4.2 Participants of the Study

The teacher data considered in the analysis comes from 11 participants in the study, 5 experienced Data Jam teachers, (had participated in one or more workshop offered during the previous 2015-2020 Data Jam PD model)- and 6 novice Data Jam teachers (See Table 2).

Table 2- Description of Grade Level and Subject Are for Teacher participants in the study

Subject Matter	**Experienced Data Jam Teacher (N=5)**	**Novice Data Jam Teachers (N=6)**
Science	1 - High School 3 - Middle School 1 - Middle and High School	3 - Middle School 2 - High School
Math		1 - High School

4.3- Data Sources

The following data were collected during the 5-day summer workshop: 1) participant observations of the teachers' interactions and work in collaborative groups, 2) artifacts or products created by teachers during the activities, 3) exit ticket reflections for each day and, 4) audiovisual recordings of the interactions between teachers and scientist mentors on the last day of the workshop. Other sources of data include transcriptions of virtual lab meetings (VLM) in which a scientific mentor met with three teachers to discuss questions about implementation. There were five mentor groups that each met three times (n=15). Lastly, there was a transcription for one focus group in which five teachers participated using purposeful sampling. Criteria to select the teachers were: voluntary participation, balance of experienced and novice teachers and completion of Data Jam with their students.

4.4 Data Analysis

As discussed earlier, our DBR study is iterative and involved two phases of design and re-design. The first phase includes the summer workshop activities, virtual lab meetings and supports as one continuous phase, involving ongoing adjustments. During the summer workshop (2021), we debriefed about our observations of the activities and the exit ticket reflections from teachers each day to adjust the proposed activities for the next day to better build teacher capacity in the data practices. Then, after collecting and analyzing the data from the summer workshop, we as a team agreed to develop and create a new shrimp dataset that was more structured and provided the context of the data that teachers needed to conduct data analysis. A training session on the new shrimp data set for teachers was offered in September 2021. Data collected from the virtual lab meetings (VLM) and Data Jam Symposium are also part of the first phase of design and re-design. The second phase consisted of reviewing the data collected for the whole year to adjust for year 2 of the Data Jam PD model. This phase is what led to the development of a new 2022 PD Data Jam Framework (See Appendix C).

The process of coding was inductive. This means that we enter the data analysis process with an open mind. As Saldaña (2021) states, inductive coding is a “learn as you go process” (p.41). Therefore, we approached the process without any predetermined codes or categories. This is an evolving process that leads to an inductive coding system or Codebook (Saldaña, 2021). First, as suggested by Saldaña (2021, p. 28-33), we conducted the following preliminary coding techniques: pre-coding and preliminary jottings. Preliminary jottings refer to when researchers start coding as you collect the data, you do not wait until all data is collected, and pre-coding is when you write words or phrases that strike you as “codable moments” (Saldaña, 2021 p.30-31). These preliminary techniques were useful to notice an emerging and recurring category throughout all the first VLM, Data Jam implementation. Pre-codes for this category include scaffolds, Data Jam conceptualizations, scientific method (See Appendix D, Codebook of the Data Jam PD Model). As the virtual lab meetings progressed (2nd and 3rd), the differences in the process of Data Jam implementation between the novice and experienced Data Jam teachers were more evident. Still, the category of Data Jam implementation prevailed in the Codebook but with more codes (e.g., supports needed, scaffolds, management of CODAP) that are used to describe this category and the emerging theme of scientist mentorship.

The elemental coding methods of descriptive and concept coding were used to generate the categories and sub-categories (See Appendix D) as part of the first cycle of coding, which means theming of the data “about patterns observed and constructed by the analyst” (Saldaña, 2021 p. 259). Descriptive coding refers to when using a word or phrase to describe the topic of a passage of qualitative data (Saldaña, 2021 p.134). While concept coding applies when the researcher assigns meaning to a broader term. For example, the codes of how, when, or why of data are part of the concept coding process that generates the category of the Data Jam research question (See Appendix D). The column of categories and sub-categories in Appendix D is the result of the first cycle of coding applied to the different sources of data (virtual lab meetings, summer workshop activities, focus group transcript, among others).

A second cycle of coding helped reorganize the data coded in the first cycle and develop the Codebook (See Appendix D) with the final description of the emerging themes from the inductive process of data analysis. Also, analytic memos were then used as a technique to document the researcher's (lead author) reflections throughout the inquiry process as an important analytic tactic. Particularly, research question analytic memos (Saldaña, 2021) were used to guide the process of analysis of the three virtual lab meeting transcripts. For each participating teacher, this type of analytic memo served to trace the changes in each teacher, both experienced and novice, to help understand the process of Data Jam implementation as one of the emerging themes.

5. Findings

In this section we focus our discussion on two topics: 1) how our main finding— the need for conceptual information on data practices— that is organized around the emerging themes, informed the changes for improvement of the Data Jam PD model, and 2) how these themes led us towards the re-design of the 2022 PD Model. We highlight examples from our two cycles of coding that led us to our main finding. It is important to note that the workshop and Virtual Lab Meetings were conducted in Spanish. The quotes in the next sections have been translated to English by the first author, who is Puerto Rican and bilingual in Spanish and English.

5.1 Emerging themes

After conducting the two cycles of coding, the following themes emerged: a) need for context for the shrimp dataset, b) data analysis, c) research question, and d) scientist mentorship to teacher (See Appendix D). The theme descriptions are provided by using statements from the teachers and from different data sources. These emerging themes led to our main finding regarding the teachers' learning needs, and the supports needed. This prospective analysis is what led to the changes made to the 2021 Data Jam PD model and the development of a 2022 Data Jam Framework to be used for the second year of the project.

The active learning activities conducted during the 5-day summer workshop were based on exploring three different datasets in CODAP: the hydrology, the shrimp, and the NEON datasets. The last two datasets were less structured than the hydrology one, which had been used in the previous Data Jam PD Model (2015-2020). Graphs generated from the shrimp dataset were used to discuss concepts of variability and distribution of data. Teachers were asked to interpret the graphs and generate questions about data. These questions were collected in post-it notes that were part of the artifacts analyzed from the summer workshop. The next section, we describe the activities conducted during the workshop around the shrimp dataset and our findings.

a. Active learning activities with the shrimp dataset

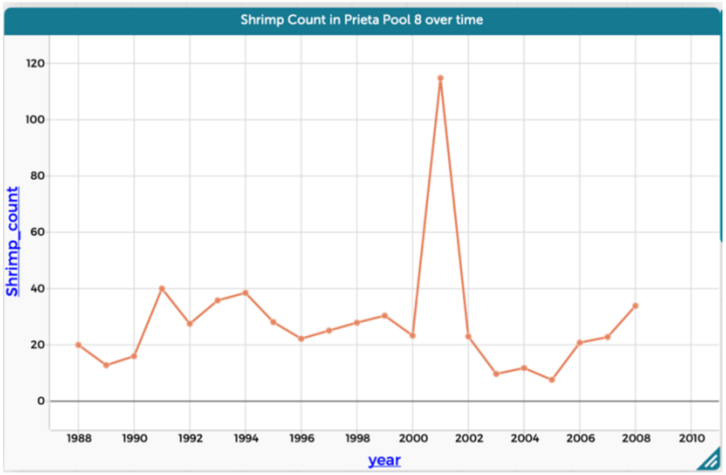
The following graph was presented to the teachers (working in collaborative groups) for them to identify data trends. For this activity, teachers were asked to do the following:

i. Make observations (post-it notes)- What did you observe?

ii. Ask questions (post-it notes) What questions do we have?

iii. Information needed (post-it notes) What information do we need?

This activity provided insights into the **data trends and visualizations practices** and what teachers understood about distributional thinking and variability.



i. Observations

The coding of teachers' observations of the graphs were organized into the three broad categories: 1. Variability, 2. High/Low peaks, and 3. Range. It does stand out that the experienced Data Jam teachers were more specific in their observations when compared to the observations made by the novice teachers. We provide some examples for each category that demonstrate the differences between novice and experienced teachers.

1. Variability:

   "We observe there is an increase in the shrimp population for the year 2001; variability in the populations seems to be normal with a count of 40 shrimps." (Experienced Data Jam Teachers).

   "Great variability in the shrimp population." "The shrimp population increased dramatically at Prieta Pool." (Novice teachers)

2. High/Low peaks

   "There was a high peak in 2001 while the levels in 2000 and 2002 are the same" (Experienced Data Jam Teachers).

"Significant increase in 2001, significant decrease in 2005, then it starts to decrease." (Novice teachers)

3. Range

"The highest count is from 20 to 40" (Experienced Data Jam teachers)

Novice teachers did not make observations regarding the range, only experienced teachers made observations in this category, which stands out as a difference between them.

ii. Questions

The categories that emerged from the questions the teachers posed for the graph presented to them are: 1. Shrimp species, 2. Abiotic factors and 3. Methodology. Most of the teachers asked the same questions in these categories but phrased them differently. Experienced teachers generated more complex questions than the novice teachers did. Some examples of these questions are:

"What kind of shrimp species is this? What functions does this type of shrimp have in this ecosystem? (Novice teachers- category shrimp species)

"What happened in 2001 that caused the increase in the shrimp population?" (Novice teacher- category abiotic factors)

"What events could have caused the variability other than the tendency that seems to be normal for years 2001 and 2005?" (Experienced teachers- category abiotic factors)

"How did the shrimp population change after Hurricane María" (experienced teachers- category abiotic factors")

"Are the sampling dates the same for each year" (Experienced teacher- category methodology)

Novice teachers did not generate questions about the methodology of the data.

iii. Information needed

The categories that emerged in this part of the activity are the same as the previous section, but teachers were more specific in detailing the information they needed to understand the distribution and variability of shrimps in pool 8. In this part of the activity, teachers are more thorough in their questions about the information they need to better understand the distribution of data and its variability. When concluding this activity with the shrimp graph, it is clear that teachers needed the context of the data (how, when, why, and where) to generate possible research questions about the data. The main objective for these active learning activities is to have teachers read and interpret graphs for them to start the process of generating questions about data, which is the first step of the inquiry process of Data Jam.

Overall, experienced Data Jam teachers were more specific in their observations, questions and information needed when looking at different graphs or visualizations of the shrimp data. This can be evidenced through the observations made by the researcher (lead author), exit tickets and the post-it notes generated in activities 2 and 3. In general, teachers struggled with the absence of the context of shrimp data when looking at the graphs presented to them.

b. Data analysis theme

An example of the *data analysis* theme can be seen in the teachers' exit tickets of the summer workshop reflecting the importance of graphs as visualization of data needed to conduct data analysis. As one experienced Data Jam teacher stated in her exit ticket, "Graphing correctly the data is vital for comprehending and analyzing the data." This theme was recurrent in the transcriptions of the virtual lab meetings, when both experienced and novice teachers expressed the importance of knowing how to read graphs for the process of data analysis. This idea is also reflected in the following quote from a novice teacher in the third virtual lab meeting: “I told my students, this is the core, knowing how to do the graph and interpreting what you are seeing, this is the basis of the research, how to interpret the graph.”

Nevertheless, even when teachers acknowledge the importance of graphs and the proper use of CODAP to create them, few of them were able to share preliminary graphs from their students to elaborate on how they were teaching their students to analyze the data. A generalized pattern observed throughout the three virtual lab meetings with the participating teachers was precisely the lack of bringing examples of graphs and data analysis. Only one experienced Data Jam teacher provided concrete examples of their students’ graphs and data analysis to receive feedback from the mentor. Observations gathered from the workshop activities, artifacts created by teachers during the workshop and observations of virtual lab meetings reflect confusion on how to describe the variability and distribution of data to properly conduct data analysis. Our main finding regarding the design of the 2021 Data Jam PD model is the need to explicitly address data practices and statistics concepts and provide more scaffolds for them to mentor their students while implementing Data Jam in their classrooms. Teachers were able to conduct data analysis for their final presentations to the scientist mentors at the summer workshop. However, in the VLM meetings, some teachers indicated that they needed more scaffolds to support students’ data analysis processes while implementing Data Jam. As one experienced Data Jam teacher stated in the 2nd VLM, “I am going to start explaining to my students how to analyze the data, but I need ideas that you would recommend or some exercises for them to conduct data analysis, especially with the shrimp dataset.”

c. Research question theme

Another important theme that emerged from the data analysis was the development of students' research questions. Aligned to the previous theme of data analysis, for many of the teachers (both experienced and novice), looking for patterns and how to apply aggregate measures (mean, median and standard deviation) to support data analysis was a recurring topic in the virtual lab transcripts as something they needed to teach their students to have them develop good research questions. Particularly, how to develop good research questions that can be answered with the data available. For example, one experienced Data Jam teacher stated, “I want my students to look for patterns, if they don’t find the patterns, then they need to change their research question.” This statement made us wonder how teachers define patterns in data as well as how they were applying statistical thinking in the process of data analysis to answer the research question.

d. Scientist mentorship to teachers’ theme

Finally, the theme of mentorship, particularly the scientist mentorship of teachers, emerged in the data analysis as a very valuable aspect of the 2021 Data Jam PD model. While participating in the virtual lab meetings, teachers stated that they were grateful and needed the mentorship from the scientists. The virtual lab meetings provided opportunities for mentors to clarify topics regarding the data practices to the teachers, helping them in the process of constructing and analyzing graphs, as well as to refine and redefine their students’ research questions. Taking as an example of one virtual lab meeting in which, the topic of patterns was discussed, the scientist mentor stated, “I’ve noticed a concern from the three teachers that you do not find a pattern and sometimes, in my opinion, not having a pattern is in itself the answer to the research question.” The mentorship provided by the scientist mentors was highly valued by the teachers as stated in the virtual lab meetings and focus group. For example, participating teachers in the focus group described their interactions with the scientist mentors as: “positive,” “useful,” “excellent,” “accessible,” “motivating,” “sense of companionship,” “gaining scientific knowledge,” and “collaboration.”

5.2 Data Jam Implementation

The 2021 Data Jam PD Model is designed to provide professional development to middle and high school science and math teachers, for them to become mentors of their students in their Data Jam research experience. Therefore, understanding how the PD model design supports teachers in this process is the focus of our study. The process of Data Jam Implementation related to the theme of Scientist mentorship to teachers which is an essential part of the PD model. Regarding this topic, the order of activities of the 2021 Data Jam PD model showed to be something teachers followed in their process of implementation. For example, one experienced Data Jam teacher expressed in VLM, “I have implemented with my students the same activities that we did in the summer workshop and let me tell you, I loved the idea of starting in the opposite order from the minute I saw it in the workshop, it was a success.” “I have to say that with these activities, my students learned what independent and dependent variables are from the graphs, much faster.”

Another aspect that emerged from the data regarding the process of Data Jam implementation is the use of the active learning activities that were designed for the summer workshop. Most of the teachers used the same activities, others adapted them for virtual learning environments and their student needs. For example, a 6th grade novice teacher stated in the second virtual lab meeting that she had used the Challenge Cards for students to manipulate CODAP and then created different performance tasks around CODAP. She stated, “next week my students have a performance test in which students will demonstrate their domain of CODAP." This teacher designed additional curricular materials to support students in topics like statistics (mean, median and standard deviation) before having students construct graphs in CODAP.

Finally, an important finding regarding the process of Data Jam implementation is how CODAP as a platform kept students engaged with the research experience. Since this graphing tool was totally new for students, they felt motivated to work on the project. Comments from teachers, like, "my students are fascinated with CODAP", and "my students played with CODAP for two days, to explore the management of the tool and they loved it" reflect how important this tool is in engaging students in Data Jam. Finally, in the teachers' focus group conducted after the Data Jam Symposium, the participating teachers (both novice and experienced) stated that the Data Jam professional development model was a very enriching experience in which they learned about data from Puerto Rico that could be used in their classrooms. Using the words of one of our novice teachers, "for me it was a great learning experience to be able to work with something so valuable for us Puerto Ricans, that is El Yunque, not from the point of view of visiting the forest, but to use real data from this ecosystem, that accentuates the ecological value of this ecosystem."

5.3 From 2021 to the 2022 Data Jam PD Model

As part of the second phase of our DBR study, some themes suggest what elements of the 2021 Data Jam PD model needed changes and what to keep in the PD model. A summary of these themes and how they informed the redesign of the Data Jam PD model is presented in Table 3. The data analysis and research question themes provided evidence to inform the changes made for the 2022 Data Jam PD model and resulted in the 2022 Data Jam framework

(See Appendix C). This framework was created from the analysis of teacher data sources (described earlier) and of the student products- the Data Jam posters presented at the Virtual Data Jam Symposium in March 2022. In turn, the 2022 Data Jam framework informed the organization of activities for the summer workshop conducted in 2022 to better support teacher learning.

Table 3- Summary of emerging themes to sustain changes to the 2022 Data Jam PD model and effective elements of the PD model

Emerging themes of the DBR study	**Evidence from our data sources that support the changes or what to keep**	**Changes made to the 2022 Data Jam PD Model**
1. Data analysis and research question themes suggest a need for conceptual information on data practices	Mentor virtual lab meetings transcripts Teachers struggled with concepts such as patterns, distributional thinking, and how to read a graph and interpret the data presented in the graph. Teacher focus group Teachers stated that they need more support on data analysis. They consider this as the most difficult part of the process of Data Jam implementation	Integrated explicit talks about data and use of statistical thinking in the summer workshop. The following data talks were included: - Data science talk - What is a case? Focus on graph interpretation - Distribution and variability in data - Correlation vs Causation - Histograms vs Scatterplot graphs - Using aggregate measures to analyze data
2. Context of data: A need for more context about the shrimp dataset to better support teachers and students in data analysis and interpretation	Data analysis from different sources of the summer workshop Data collected from observations, artifacts generated by teachers during the active learning activities, exit tickets and final presentations provided enough evidence to suggest the teachers' need for more context of the shrimp dataset.	A new shrimp dataset was created for teachers and students to use in Data Jam in the fall of 2021 and used as part of the 2022 summer workshop. An additional document was prepared that contained ecological information about the shrimp and their habitat.
Elements that were kept for the 2022 Data Jam PD Model		
3. Mentorship (Scientist mentorship to teachers	Teacher focus group Teachers rated as excellent the daily activities done in the summer workshop. Challenge cards are regarded as highly valuable resource for teachers and students to manipulate CODAP.	Challenge Cards were reviewed and updated for the 2022 summer workshop. The virtual lab meetings (3) were highly valued and were also part of the 2022 PD Model.

The changes incorporated into the 2022 Data Jam PD model shift the focus from the data practices to the elements of the Data Jam Poster as the final product of the Data Jam research experience. The shift in our focus is to make the thinking process more visible to teachers for them to mentor and engage their students in these data practices needed to communicate their research in the final product.

The Data Jam poster elements reflect the process of research using long-term ecological data from the LUQ-LTER. In the final product, students must apply different data practices to complete the task. Therefore, the 2022 Data Jam Framework describes in detail the characteristics for each poster element. These are: 1. Research question, 2. Data/data moves, 3. Graph construction, 4. Graph interpretation and, 5. Answering the research question. Each of these attributes are part of the Data Jam research experience exhibited in the final product (See Appendix C).

6. Discussion of findings

Our findings suggest that the cognitive apprenticeship model described by Collins et al. (1989) supports teacher learning in some, but not all, of the data practices that the PD model intended. The active learning activities conducted during the summer workshop were modeled in the opposite order of how the tasks are done in the Data Jam project but, learning of these practices varied among the teachers as we further noticed in the Virtual Lab Meetings. Thus, the cognitive apprenticeship model used for the design of the 2021 Data Jam PD model was used by some teachers, particularly experienced teachers, as a learning model with their students. Still, some teachers struggled with some data practices. For example, the data practice of visualization and data trends rely on understanding the variability and distribution of data which emerged as one of the topics teachers needed more support. Also, our data shows that teachers needed more support in understanding how to apply aggregate measures, such as mean, median, and standard deviation, to conduct data analysis of their graphs. Therefore, more scaffolds are needed to support teacher learning of these concepts and how to apply them to the Data Jam project.

To better support teachers in the process of Data Jam implementation, we will include more scaffolds to build data literacy that teachers can use with their students. This helps to balance the content of the PD model in both content around data and data practices and the instructional practices that will support their student's learning as well. The Data Jam PD model design and re-design is the focus of our longitudinal research (2021-2025). Providing this opportunity to local math and science teachers certainly helps build capacity for future representation in STEM careers.

7. Future Directions

This research paper includes only one year of data. We will continue the iterative process of design-redesign of the Data Jam PD model. For our first year of the study, we were not able to recruit the number of teachers we set our goal to be, but we hope to have more participants for the next years, thus, increasing a variety of contexts. Also, we will look at the quality of the student posters to reflect upon the success of the 2022 Data Jam PD model and making further changes to the PD model design. Finally, by year 4 of our project we will be able to conduct retrospective analysis as part of our design-based research study. We highlight the fact that our empirical research site is Puerto Rico, which has been an underrepresented population of teachers in educational research.

Acknowledgments

The authors were supported in part by the National Science Foundation grant DRL-2049061 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.

Campanella, M. & Penuel, W. R. (2021). Design-Based Research in Educational Settings. In Philippakos, Z. A., Howell, E. & Pellegrino, A. (Eds.), *Design-Based Research in Education: Theory and Applications* (pp. 3-22). The Guilford Press.

Carpenter, P. A. & Shah, P. (1998). A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied, 4*(2), 75.

Cheng, C.C (2014). Situated learning and professional development: a case study of applying cognitive apprenticeship and community of practices in kindergarten. *Problems of Education in the 21st Century,59*, (91), 15-24.

Cleveland, W. S. & McGill, R. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association, 79*(387), 531-554.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R. & Schauble, L. (2003). Design experiments in Educational Research. *Educational Researcher,32* (1), 9-13.

Collins, A., Brown, J. S., & Newman, S. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. Resnick (Ed.), *Knowing, Learning,* *and Instruction: Essays in Honor of Robert Glaser* (pp. 453- 494). Hillsdale, NJ: Lawrence Erlbaum Associates.

Creswell, J. W. (2012). *Research Design: Planning, Conducting, and evaluating quantitative and qualitative research* (4th ed.). SAGE Publications.

Erickson, T., Wilkerson, M., Finzer, W., & Reichsman, F. (2019). Data Moves. *Technology Innovations in Statistics Education*, 12(1). http://dx.doi.org/10.5070/T5121038001 Retrieved from https://escholarship.org/uc/item/0mg8m7g6

Freedman, E. & Shah, P. (2002). Toward a model of knowledge-based graph comprehension. *Diagrammatic Representation and Inference*, 59-141.

Frischemeier, D., Biehler, R. Podworny, S. & Budde, L. (2021) A first introduction to data science education in secondary schools: Teaching and learning about data exploration with CODAP using survey data. *Teaching Statistics, 43*,182–S189. https:// doi.org/10.1111/test.12283

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, 124- 158.

Lehrer, R. & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal,41* (3),635-679.

Loshe, G.L. (1993). A cognitive model for understanding graphical perception. *Human Computer Interaction,8* (4), 353-388.

Mautone, P. D. 7 Mayer, R. E. (2007). Cognitive aids for guiding graph comprehension. *Journal of Educational Psychology,99* (3), 640.

Mayer, R. E. (1993). Comprehension of graphics in text: An overview. *Learning and instruction,3* (3).

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. & Báez-Rodríguez, N. (2017). Getting students jazzed about Critical Zone: Engaging students in authentic inquiry through Data Jam. *In the Trenches,7(4), 5-8.*

Merriam, S. B. y Tisdell, E. J. (2016). *Qualitative Research: a guide to design and implementation* (4th ed.). Jossey-Bass: A Wiley Brand.

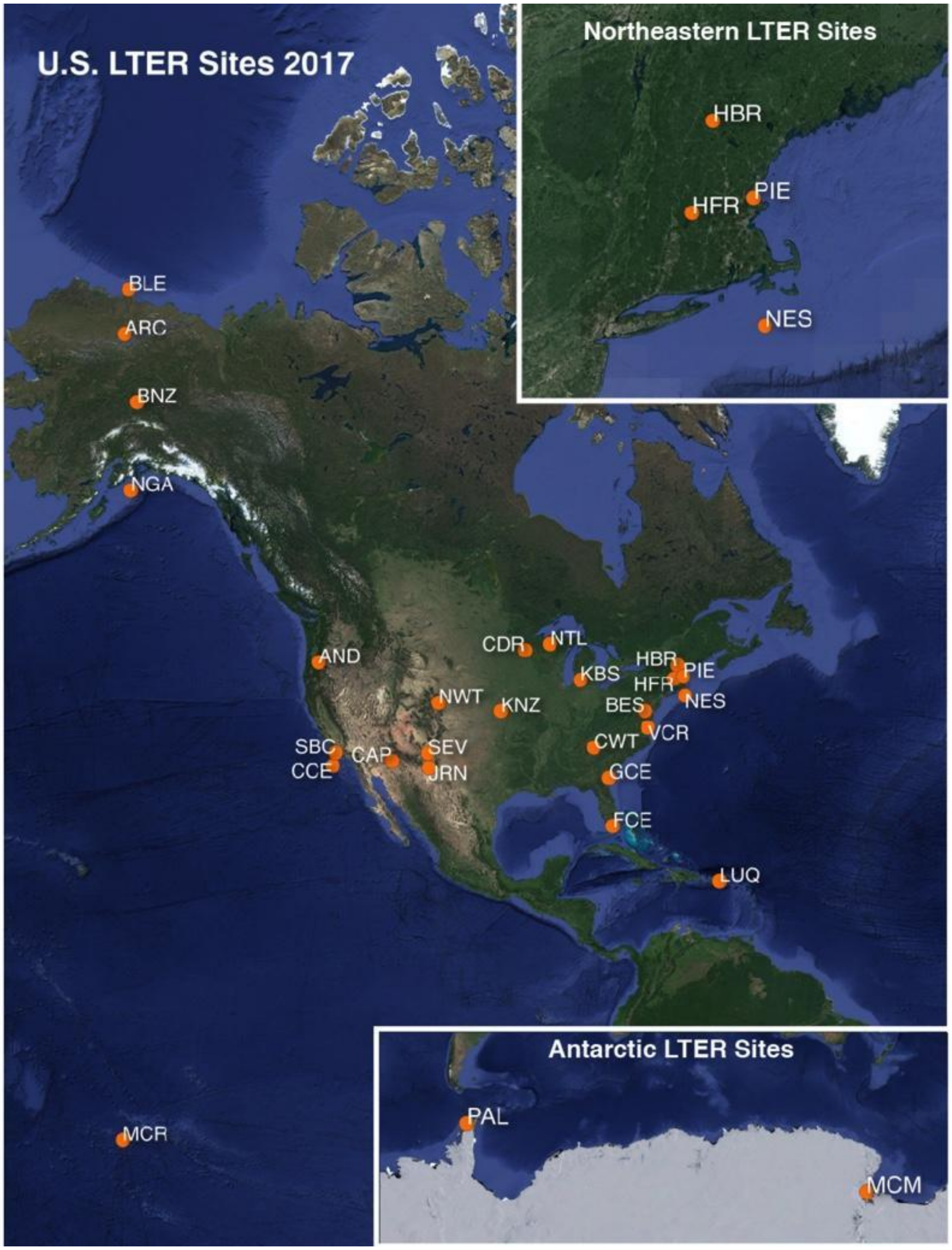
Saldaña, J. (2021). *The Coding Manual for Qualitative Researchers.* SAGE Publications.

Shah, P. & Carpenter, P. A. (1995). Conceptual limitations in comprehending line graphs. *Journal of Experimental Psychology: General*, *124*(1), 43.

Shah, P. & Hoeffner, J. (2002). Review of graph comprehension research: Implications for instruction. *Educational Psychology Review*, *14*(1), 47-69.

Shah, P., Mayer, R. E., & Hegarty, M. (1999). Graphs as aids to knowledge construction: Signaling techniques for guiding the process of graph comprehension. *Journal of educational psychology*, *91*(4), 690.

Appendix A- Map of LTER Sites



Appendix B- 2021 Data Jam Framework

Attribute	Description	Questions to think about when assessing student work
#1 Appropriateness and complexity of research question	*Coming up with an appropriate research question, given a complex dataset, can be tricky.* <u>Good research questions:</u> ● Usually involve finding relationship **between attributes** and/or **a pattern that the data representation reveals** ● Investigating a good research question often **leads to more questions.** <u>Common issues with student research questions</u> 1) Asking **a "trivial" question** ("What was the highest temperature all year?") Investigating a good research question **usually will not end with a clear "yes" or "no."** 2) Asking a complex question that **can't be answered with the data** at hand.	● Does the research question look at a relationship between attributes? ● Does the research question reveal a pattern in the data? ● Does investigating the research question have the potential to lead to more questions? ● Is the research question non-trivial? ● Can the research question be answered yes/no? ● Can the research question be answered using the data?
#2 Distributional thinking and awareness of variability	Variability is one of the hallmarks of data. Understanding the **characteristics of distributions – central tendency, degrees of spread, shape** – is part of sophisticated thinking about data. CODAP makes it possible to visualize distributions in a way that is impossible or much more difficult in most spreadsheet programs, so we expect to see more evidence of distributional thinking when students use CODAP.	● Do student data visualizations show the central tendency of the data? ● Is there evidence that students are thinking about the central tendency of the data? ● Do student data visualizations show the spread and shape of the data? ● Is there evidence that students are thinking about the spread and shape of the data?

#3 Attention to both aggregate values and individual cases	Sophisticated data analysis requires simultaneously **using an aggregate lens** and **noticing individual cases**. **Aggregates** reflect aspects of a distribution that go beyond individual cases. ● The traditional measures: **mean, median** and **mode** ● There are also more informal measures such as "modal clump" (**All the values around the mode)** Sophisticated data analysis requires also noticing **individual cases**, especially ones that might be considered **outliers**.	Does student analysis include aggregate measures: mean, median, mode, modal clump? Does student analysis discuss individual cases, especially outliers?
#4 Data moves	**"Data moves"** are **changes to data structure or values during an analysis.** The notion of "data moves" has been gaining currency in data science education, as data visualization and analysis are seen less as involving a single step and more as being a sequence of steps that build on one another. While we can't necessarily see each of the steps a student took to create a final product, we can **infer some student data moves** by **knowing the underlying dataset** and **analyzing the graph.**	As we consider students representations, and think about the underlying data set, can we tell whether students used multiple data moves to transform the data? ● Did students create new derived variables that were not part of the original data set? ● Did students integrate other data that were not part of the original data set?
#5 Communicating scientific findings	**The final product should clearly communicate the research process and findings.**	● To what extent do the different parts of the research project fit together? (e. g. Do the results address the research question) ● How effectively do student graphs communicate information? (Consider: type of graph, attributes on the graph, representation of attributes on x-axis/y-axis/color legend, annotation of the graph, etc.)

Appendix C- 2022 Data Jam Framework

Data Jam Attribute	**Description**
Research Question	• The research question is clearly stated • The research question references relevant scientific background, i.e. there is a scientific reason why the question is asked. • The research question can be answered with the data provided.
Data/Data moves	• The attributes chosen are aligned with the research question. • The attributes used are described in terms of data collection methodology (including frequency of collection). • If attribute values have been changed (e.g., averages computed), the process is described
Graph construction	• Graphs include the appropriate attributes to explore the research question. • Appropriate graphs are created. • The visualization helps make sense of the data for the person viewing the graph. • Appropriate measures are included on the graphs (points off for too many as well as too few)
Graph interpretation	• Conclusions drawn from the graph match the graph • Graph is annotated to illustrate what aspects of the graph support the conclusions. • Both the trends and the variability are noted in the description of the graph.
Answering the research question	• Here is a clear answer to the research question ("we can't tell from the data" counts as a good answer) • The answer is supported by the graphs. • The connection between the graphs and the answer is described clearly. • What they have learned is effectively communicated to the reader of the poster.

Appendix D- Codebook for the 2021Data Jam PD Model (Examples of pre-codes, codes, and theme descriptions after two cycles of coding)

<table>
<tr><th colspan="2">Examples of codes</th><th>Categories and sub-categories</th><th>Theme Description</th><th>Data sources</th></tr>
<tr><td colspan="2">Shrimp information
Comparison among pools 0, 8 and 9
Climatological events</td><td>Category: Context of Data</td><td>Context of the shrimp dataset
The context of data is essential to generate good research questions.</td><td>Exit Tickets
Observations
Summer workshop artifacts from active learning activities</td></tr>
<tr><td colspan="2">Graphs
Distribution of data
Visualization of data
Variables (independent, dependent)</td><td>Category: Graphs</td><td rowspan="2">Data Analysis
Graphs are essential to conduct data analysis. Graphs help visualize the distribution of data and include the variables under study.

CODAP is an excellent technological tool to create graphs.</td><td rowspan="2">Exit Tickets
Observations
Summer workshop artifacts
Virtual Lab Meeting Transcripts
Focus group</td></tr>
<tr><td colspan="2">Use of CODAP Tool
Integration of Science and Math</td><td>Sub-category: CODAP tool</td></tr>
<tr><td colspan="2">What data
How of data
When of data
Complex
Methodology
Knowledge of central tendency measures</td><td>Data Jam Research question</td><td>Research Question
A good research question should include what, how and why of data. It should be complex in terms of relating variables. Students can develop good research questions.</td><td>Exit Tickets
Audio visual recordings
Summer workshop artifacts
Virtual Lab Meeting Transcripts
Focus group Transcript</td></tr>
<tr><td>Pre-codes
Scaffolds
Data Jam conceptualization
Scientific method</td><td>Examples of codes
Management of CODAP
Support Needed
Scaffolds</td><td>Category: Data Jam Implementation</td><td rowspan="2">Mentorship (Scientist to teacher)
Having scientist mentors is valuable to implement Data Jam.</td><td>Summer workshop artifacts
Virtual Lab Meetings Transcripts</td></tr>
<tr><td colspan="2">Accessible
Motivation
Scientific knowledge</td><td>Sub-category: Scientist Mentors</td><td>Focus group transcript</td></tr>
</table>