

Expanding the Reach of the Data Jam:
*Introducing High School Data Science to
More Diverse Youth, Communities and Regions*

Evaluation and Final Report



Table of Contents

Acknowledgement	3
Introduction	5
<u>Part 1:</u> The Design of the DataJam	4-5
What is the DataJam?	4
How does DataJam work?	4
What resources and assistance are provided to DataJam teams?	4
What is required of a high school DataJam team?	4-5
<u>Part 2:</u> NEBDIH Pilot Project to Expand the DataJam	5-8
Expansion of the DataJam throughout the NE Big Data Innovation Hub Region	5
Project Goals	5-6
Significance of the Pilot Project for Data Science	6
What is the Pilot Project Seeding?	6-7
Intellectual Merit: What is unique and innovative about the project?	7
<u>Part 3:</u> Evaluation of the Pilot Project	8-12
Who were the students participating in the expansion of the DataJam?	8-9
What was the academic background of the students participating in the pilot study?	9-10
What was the student's familiarity with data science and computer science?	10
What was their DataJam experience?	10
Was participation in DataJam associated with a change in how they saw using data science in their future?	11-12
Lessons learned through formal evaluation of the DataJam pilot?	12
Limitations	12
<u>Part 4:</u> Extended Results of the Pilot Project and Future Directions	12-
Impact of the DataJam for Native American students on the Pala Reservation	12-13
News articles about the DataJam	13
Presentation of the DataJam at National Meetings	13
<i>The DataJam Download</i> – A monthly DataJam Newsletter	13-14
Impact of the DataJam for DataJam Team Members and Mentors	14
Teaching the DataJam Mentor Course in a BSLE format to include students and faculty at other universities	14
Awards for the DataJam	14-15
Grants to further expand the DataJam	15
Broader Impacts	15

Acknowledgement:

This project is supported by a seed grant from the Northeast Big Data Innovation Hub through the National Science Foundation (Award #1916585). The opinions, findings, and conclusions, or recommendations expressed in this document are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Introduction

This report represents the results of a small-scale study of the DataJam data science program for teens conducted during the 2021-2022 academic year. It seeks to characterize participants in the program, identify any effects that DataJam had on the participants, and in particular focuses on effectiveness of scaling the DataJam nationally by migrating it to 100% online. We are particularly interested in identifying the feasibility of further scaling to a national level, and the effectiveness of DataJam in providing an equitable and inclusive means of introducing the next generation to the power of data science.

Part 1: The Design of DataJam

What is DataJam?

The DataJam is an academic competition that runs throughout the school year and is designed to introduce high school students to analyze datasets to answer research questions. The DataJam was started in Pittsburgh in 2014, and over the past two years with the Covid-19 pandemic has both embraced remote learning and expanded outside of the Pittsburgh area. The program partners with teachers to bring together students who work in teams of generally 5-7 to formulate a research question, find publicly available data sets, analyze their data, make data visualizations, and present their findings to a panel of judges. Students learn skills pertaining to the scientific method, data analysis, and how to give scientific presentations. Prior to Covid, students were also given the opportunity to visit local companies that use big data to see the practical applications of data science in their community. Teams are mentored by university undergraduate students who take a course on how to serve as effective DataJam mentors, learning about basic statistical approaches, the ethics of collecting and using big data, and how to work with diverse communities. The DataJam is coordinated by Pittsburgh DataWorks, an educational nonprofit, whose primary mission is to coordinate the DataJam. More information and resources for DataJam teams can be found on the DataWorks website: (<https://www.pghdataworks.org/data-jam>)

How does DataJam work?

- High school teams choose a research question they are interested in, find a data set(s) to analyze, learn how to organize and analyze data and make visualizations. Our trained DataJam mentors are available for online mentoring via Zoom and on a DataJam Slack workspace to help teams with all aspects of the project.
- In April teams present their projects as posters and in slide presentations in front of a panel of judges, conducted on Zoom. A DataJam finale is held virtually with teams from across the Northeast and lots of awards are given.

Commented [CB1]: this section weems to have been adapted from wht is sent to teachers. A few items (highlighted in yellow) could be re-worded to make more suitableit a little more formal for this report..

What resources and assistance are provided to DataJam teams?

- The website ([pghdataworks.org](https://www.pghdataworks.org)) has a large amount of information about DataJam, including videos about online resources, what's involved in participating in DataJam, and lots of help finding data sets and doing data analyses.
- University undergraduate students from a wide variety of majors serve as DataJam mentors. They are available to meet with students by Zoom throughout the academic year to help students think about a project they would like to undertake, to help them find data sets to work with, write their proposal, learn data analytical procedures and ways of graphing and visualizing data, to put their final presentations together, and to help them practice their final presentations. DataJam mentor assistance can be scheduled at times that work for students and they are also available on a Slack channel (last year: DataJam2022).

What is required of a high school DataJam team?

November – January:

- Recruit students at high schools and afterschool programs who would be interested in this data science experience as a fun, engaging extracurricular activity. Teams are generally composed of 3-8 students.
- Teams are paired with a DataJam mentor who will help guide them through all steps of their DataJam project. It works best if they find a regular meeting time when both the team and the mentor is regularly available.

- Teams complete a one-page proposal that lists a research question and hypothesis, identifies a data set(s) that can be analyzed to answer the question, and outlines the strategy the team will take to analyze the question. Proposals will be due in early December each year.
- Students join the current year DataJam Slack workspace so they can communicate with DataJam mentors online. This requires a parent signing a permission slip (available from Pittsburgh DataWorks) and student's providing a non-school email address that can be used to access Slack.

February – April:

- Teams analyze data to answer their research question.
- Teams prepare a poster that is due at the end of March and final presentation of their project to be given to judges in mid-April via videoconference.
- Teams attend the DataJam finale, held via Zoom in late April.

Part 2: NEBDIH Pilot Project to Expand the DataJam

Expansion of the Data Jam Throughout the NE Big Data Innovation Hub Region

In this pilot project we expanded participation in the DataJam throughout the region served by the Northeast Big Data Innovation Hub through collaborations with groups well-connected to their local communities and high schools. They included: The University of Pittsburgh, the Massachusetts Green High Performance Computing Center, The Woods Hole Institute, and outside the Northeast Region, under-resourced schools in: Pennsylvania, New Jersey, Massachusetts, and California.

The investigators involved in this pilot project included Sarah Dunton, Catherine Cramer, Stephen Uzzo and Judy Cameron. Sarah Dunton is Director of Outreach for MGHPCC, the Massachusetts Green High Performance Computing Center, a non-profit that is a joint venture of a 5 universities and is dedicated to providing environmentally responsible infrastructure for research computing systems. MGHPCC is located in Holyoke, Massachusetts, a city of need that has one of the highest percentages of economically disadvantaged youth in the state, with 70% of students meeting the criteria of economically disadvantaged and 80% of the students being Hispanic/Latino. Catherine Cramer is the Director of the Woods Hole Institute (WHI), which is a non-profit located in Woods Hole, Massachusetts, home to several large scientific institutions such as the Marine Biological Laboratory, NOAA, the Woods Hole Oceanographic Institution, the US Geological Survey, and the Woodwell Climate Research Center, will include the Woods Hole Science and Technology Education Partnership (WHSTEP). Stephen Uzzo was a Chief Scientist at the New York Hall of Science (NYSCI) at the time this pilot study was conducted. Dr. Uzzo designed, coordinated and analyzed the evaluation data collected in this pilot project. Judy Cameron is Director of Pittsburgh DataWorks, and coordinates all aspects of the DataJam, including recruitment of teams, providing background information to teachers, training of DataJam mentors, connecting mentors to teams, coordinating the production of new resources for all aspects of the DataJam and making them available on the DataWorks website.

Goals: The goals of this pilot project were:

1. To recruit DataJam teams from diverse communities in the NE region, including urban and rural communities, communities with a high percentage of economically disadvantaged students, underserved minority students, native populations and first-generation ELL immigrant populations in order to broaden participation in the DataJam in equitable and inclusive ways, employing constructivist methods to meet the needs of diverse communities.
2. To use participatory design research (PDR) to involve teachers in these communities throughout the year of DataJam participation to understand how the design and resources of the DataJam are working for their student populations, and to work with teachers to design new strategies and resources that will more effectively engage their students.
3. To partner with established professional groups in each region who have a history of partnering with local schools to provide an online program (the DataJam) that has a history of effectively engaging youth in data science education, in a manner that embraces local communities and cultures.

Significance of the Pilot Project for Data Science

In the Pittsburgh region the whole community has benefitted from youth participation in the DataJam. High school students benefit from the education in data analytics, introduced in such a way that it is highly engaging and they are very motivated to learn this skill set. Universities are benefitting as more high school students become familiar with the educational opportunities available with higher education and they have positive experiences with the DataJam mentors who serve as near-peer mentors. The business community is benefitting as more young people become interested in data analytics to provide a future workforce in the region. This broad impact on the community led the DataJam to win the University of Pittsburgh Partnership of Distinction Award in 2021 and the Carnegie Science Center "Best Interdisciplinary Approach to STEM Education" in 2022.

But the overall goals of the DataJam are larger than just to educate more Pittsburgh students about data science. We believe the DataJam is a powerful learning platform that engages a whole community and has the potential to make data science education much more broadly available throughout the U.S. and worldwide. But to do this there is a great need to broaden participation in data science education and meaningfully address the need for educational equity in marginalized communities of students. Economic, cultural and pedagogical barriers have been identified for youth to become data literate. Learners in under-resourced communities are less likely to become data literate to learn the skills needed to gather, analyze, understand and critically think about data and how to apply this knowledge to solving real problems. And we know that low data literacy in high school has long-lasting impact affecting equity in employment and the innovation capacity of a local economy.

We believe that identifying ways to have data science education integrate into local cultures will be key to increasing equity in data science education for this project as well as more generally.

What was the pilot project seeding?

The DataJam has been successful in recruiting an increased number of high school teams in the Pittsburgh area over the past 8 years (from 3 schools in 2014 to over 28 schools), and many resources have been created that are freely available on the Pittsburgh DataWorks website (including a modular curriculum for high school teachers of a variety of disciplines to incorporate more data science into the classroom). The DataJam program has a strong track record of engaging high school students who have gone on to attend college and in many cases to further pursue additional data science training. Not surprisingly, a number of DataJam students have entered college at the University of Pittsburgh and in fact several have gone on to continue being affiliated with the DataJam by serving as DataJam mentors while they were in college. Other past DataJam participants have started Data Jam-like competitions at their colleges or become key participants in college activities that promote data science education.

We believe we are poised to expand this unique collaboration between high schools, universities and companies that use big data much more broadly across the country. However, to support such expansion it is important to assess the effectiveness of the program with much more diverse populations of youth, including youth living in rural communities and urban communities, as well as with native populations and youth from recent immigrant families. Moreover, it is important for us to get feedback from these communities and to work with them to craft resources that will work especially well for their students. The work we proposed in this pilot grant expanded the Data Jam to communities and schools that are affiliated with the entire region of the NE Big Data Innovation Hub and a formal evaluation was made of the with pre- and post-surveys of the youth participating in the pilot study, with feedback collected from teachers.

In August 2022 we submitted an ITEST Developing and Testing Innovations (DTI): *DataJam-Online* grant to continue the national expansion of the DataJam to develop a scalable, adaptive online program aimed at effectively engaging a wide diversity of youth in learning data science skills. This grant is designed to address 3 research questions. RQ1: What cultural assets and funds of knowledge do underrepresented learners from diverse communities bring to forming and sustaining local DataJam-Online nodes? And what effect does that have on student and teacher participation? (ITEST Pillar 3: Strategies for equity in STEM Education); RQ2: What effect does moving mentor training, coursework and mentorship of participating high school students to a collaborative blended synchronous learning environment (BSLE) strategy have on student and teacher participation in DataJam-Online? And how will BSLE mentor

training, online assets and resources affect student skills, agency and the ability to apply knowledge across diverse communities of learners? (ITEST Pillar 1: Innovative use of technologies in learning and teaching); RQ3: What effect will adapting DataJam-Online to a variety of cultural contexts have on building sustainable collaborations with a diversity of businesses and industries, and a diversity of colleges and universities, to support self-efficacy in DataJam-Online participants? (ITEST Pillar 2: Development of key partnerships for career and workforce preparation).

In January 2023 we will be submitting an AISL grant to NSF to support the further expansion of DataJam into out of school learning situations, building on the pilot data collected in this project showing effective engagement of youth at the Learning Center on the Pala Native American reservation in the 2021-22 academic year.

Intellectual Merit: What is unique and innovative about the project?

The pilot project was unique and innovative as it successfully adapted an effective program, the Data Jam, that has been shown to engage high school youth in data science education with a participatory design research strategy to identify barriers, locate and build upon diverse community funds-of-knowledge, and work toward developing culturally sustaining pedagogies in data science. Data science education can be a foundation for more equitable education because of the relevance of data science in careers and computational thinking in STEM, as well as in business, economics, and social sciences. This pilot project worked with the teachers in each community to help them identify locally and culturally relevant research questions and data sets to support a more culturally relevant and sustaining data science pedagogy that is tailored to the community.

The evaluation of this pilot project specifically addressed 4 research questions: (1) What are the barriers or needs of communities of learners in engaging students in the DataJam?, (2) What are the cultural assets or funds of knowledge that underrepresented learners and communities have that are relevant to undertaking a DataJam project which will be engaging to them?, (3) What culturally contextual ways do students use tools and techniques to think with and communicate through data, and are they cross-culturally relevant?, and (4) Can we develop a hybrid model, a framework, for taking a strategy to teach data science to one population and effectively adapt it to work with culturally diverse populations of youth?

A second key to the success of the DataJam has been the use of college students to help guide high school students and serve as near-peer mentors. There is a strong literature documenting the effectiveness of near-peer mentorship as an effective teaching strategy as youth are more likely to relate to near-peers and envision themselves as gaining the skills that are being taught. In the pilot project, we engaged DataJam mentors at one institution of higher education in Pennsylvania as near-peer mentors for all of the teams we were working with. However, the pilot project allowed us to build relationships with institutions of higher education in rural areas of Pennsylvania, Southern California and New Jersey. We have built on those connections and for the 2022-23 academic year we have several universities in these locations training DataJam mentors by offering the DataJam mentor training course by Blended Synchronous Learning Environment (BSLE).

Part 3: Evaluation of the Pilot Project

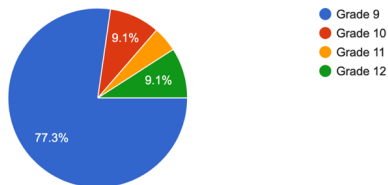
Evaluation was done using 3 instruments: (1) A pre-survey of 24 high school youth who participated in DataJam teams in New Jersey, Massachusetts and Southern California; (2) A post survey of these same students and an additional 36 students who participated in the DataJam from schools in Pennsylvania, where the DataJam has been running since 2014; and (3) a focus group conducted with participating teachers. Data was gathered using Google forms and cleaned and tabulated using Excel. No identifiable information was gathered from the students through these instruments. Focus group data was gathered through transcripts of Zoom meetings with the participating teachers and using narrative analysis to extract general sentiment toward the program and the value that teachers believed their students derived from the program. Below are the results from this process.

Who were the students participating in the expansion of the DataJam?

As shown in the graphs below, the majority of the participants in this expansion of the DataJam were 9th grade students and 45.8% were female. We were very pleased by this percentage of females as data science is dominated by males, yet this DataJam pilot project engaged more females than males. Of the 24 youth in the pilot project, 20 identified as Hispanic, 4 as Black and 5 as American Indianan/Alaska Native. Thus, the vast majority could be characterized as underserved minorities in the American school system.

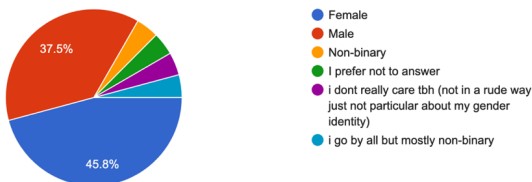
What is your grade?

22 responses



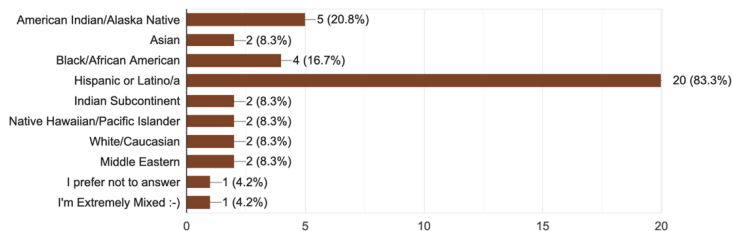
What is your gender?

24 responses



Which of the following best describes you and your family? (Please select all that apply)

24 responses

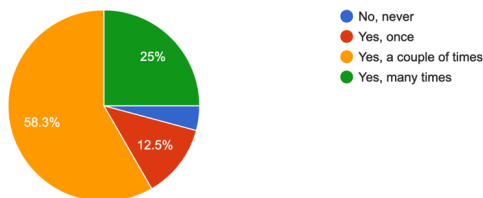


What was the academic background of the students participating in the pilot study?

We were happy to find that the majority of the students participating in the DataJam had previously undertaken a research project, and almost all students had experience working on a team project. When asked about their favorite subject in school over half indicated an interest in science and/or math, which remained consistent across pre and post surveys. In the post survey an additional 7% indicated an interest in computer or data science as a subject in school.

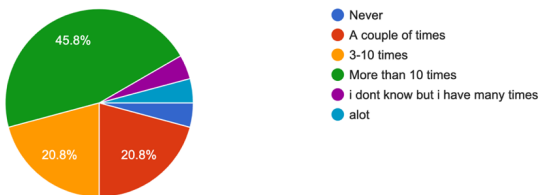
Have you ever done a research project (in/out of school) in science, social studies or other subjects?

24 responses



How many team projects have you worked on before (in/out of school)?

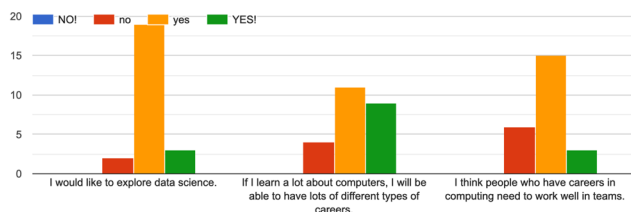
24 responses



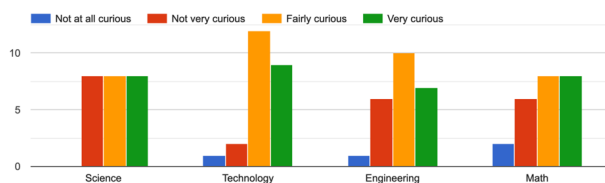
What was the student's familiarity with data science and computer science?

Students had some familiarity with the term "data science" and had heard it in computer science classes primarily. When asked what they thought "data science" involves responses included "something to do with computers", "programming", "coding", and "comparing collected data". Most students showed an interest in exploring data science, believed that if they learned more about computers that they would be able to have a number of different careers and felt that people working in careers with computers needed to work well in teams. Twice as many students were curious about science compared to those not curious about science, and this trend held for interest in engineering and math, while the vast majority of the students were interested in technology. When asked how they used computers to solve problems, the majority of the students were vague and indicated that they would simple solve problems with a few indicating they used Google or a calculator.

Do you agree with the following statement?



Please pick the answer that matches how curious you are about the following things.

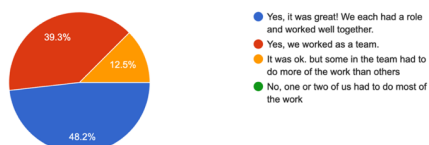


What was their DataJam experience?

Eighty seven percent of the students participating in the DataJam felt they worked together well as a team. When asked how they had used computers in their DataJam project in the post-survey they were able to provide a great deal more specificity than was apparent in the pre-survey. They were able to name programs they used such as Google Sheets, Excel, R-Studio. They indicated they used computers to search for data, clean data, analyze data and make graphs.

Did you feel like your DataJam team worked together successfully?

56 responses

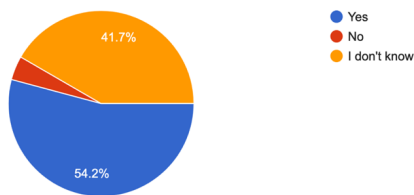


Was participation in DataJam associated with a change in how they saw using data science in their future?

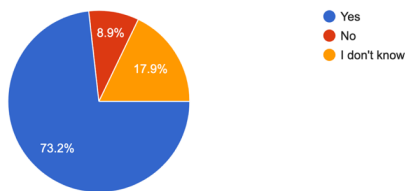
When asked if they knew what type of career they were interested in, in the pre-survey about half replied in the affirmative, with twice as many naming a career utilizing science (i.e., engineering, medicine, psychologist) as those naming a nonscientific career (business, financial consultant). After the DataJam experience, the number of students indicating that they knew what kind of career they were interested in rose by over 20%. And the number of students interested in a STEM career rose by 38%. In particular, there was a 24% increase in the number of students expressing an interest in computer or data science. And virtually all students believed that if they learned about computers many jobs would be available to them, and that people using computers in their jobs need to be able to work well in teams. Students were able to provide much more specificity in their answers when asked what types of problems data science could solve. Answers included things like "Any, honestly. Marketing, supply chain, consumerism, medical,

etc”, and “I think data science can be used to solve many different problems, as long as there is data to be collected and used.” And when asked what data science is, answers were much more comprehensive than in the pre-survey. They included responses like “Data science is collecting information to form an analysis used to come to a conclusion” and “I would say that it is analyzing and forming sets of data in order to see how things affect each other. Then using these results to draw conclusions and find solutions to problems that have been presented.”

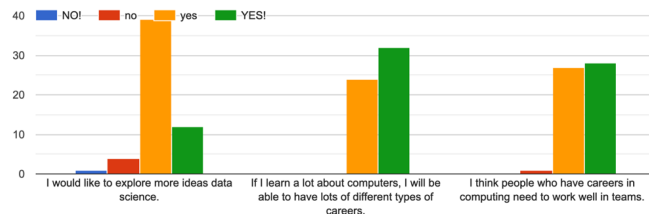
Do you know what type of career(s) you are interested in?
24 responses



Do you know what type of career(s) you are interested in?
56 responses



Do you agree with the following statement?



In the focus group with teachers there was a consensus that they felt both they and the students were adequately supported by the online version of *DataJam*, that the mentors connected well with students, and that the program was a very effective learning experience for the students. Teachers felt the learning goals and resources for the program were clearly articulated and highly available to both the students and teachers. Most of the participating teachers indicated they would participate in *DataJam* again.

Lessons learned through formal evaluation of the DataJam pilot?

The formal evaluation of this pilot project to expand the *DataJam* outside of Pennsylvania, to more diverse populations of youth, showed a lot of promise for national expansion of the *DataJam* as an effective educational program to engage underserved communities in learning about data science. The pilot project was very effective in engaging females and youth from underserved populations. The percentage of students knowing what they wanted to do in their career increased after participation in *DataJam* with 64% being interested in math/science careers. There was greater recognition of what data science was and students were able to give much more detailed answers to what data science was used for. They clearly developed a good understanding that data science is used in virtually all fields of study. Although the pre-survey showed that the majority of students wanted to explore data science there was a marked increase in interested in data science after the *DataJam*. Overall, we believe that the results of the pilot study suggest real promise for the *DataJam* to expand to a wide diversity of communities and a national platform for engaging youth interest in data science.

Limitations

While results from the evaluation indicate substantive change in student attitudes, self-efficacy and interest in STEM and data science, this was a pilot study, and as such a few factors are important to note. The pre-surveys were completed only by the new online cohorts recruited specifically for the pilot study ($n=24$), while the post-surveys were completed by both the pilot study cohort and a subpopulation of students on *DataJam* teams from the Pittsburgh area ($n=24+36$). The 2 survey results were normed in order to compare attitudes before and after participating in *DataJam*. Because no identifiable information was gathered, parsing the two post-survey cohorts was not possible. Also, while pre-surveys were completed early in the program, they were not completed prior to the program commencements and thus some effects of the program might have biased the results and conclusions drawn from these results.

Part 4: Extended Results of the Pilot Project and Future Directions

The NEBDHI pilot project had impact in several areas that were not anticipated at the time that we applied for funding. However, these impacts have been considerable and we feel that they significantly extend the impact of the pilot funding, thus they are briefly described here.

Impact of the DataJam for Native American students on the Pala Reservation

The two students from the Pala Native American reservation outside of San Diego who participated in the 2022 *DataJam*, became very interested in data science through their *DataJam* experience, and they were offered the opportunity to do a summer fellowship with the San Diego Supercomputing Center to allow them to continue learning more about data science. In the summer of 2022 they worked with Kimberly Mann Bruch at the Supercomputing Center, who had served as the advisor for their *DataJam* project at the Pala Learning Center. Their summer internships allowed them to test water around various areas of the reservation – primarily focusing on the youth center and learning center sites. They also worked on learning more about how Google sheets work to conduct simple data science studies; specifically, they were shown how to do simple regression analysis by UC San Diego student Timothy Chu, who was working with Mann Bruch through a National Science Foundation Research Experience for Undergraduates, which was funded by the Big West Data Innovation Hub.

News articles about the DataJam

In the September 12, 2022 issue of NSF News under the section “What’s hot in science and engineering” an article entitled “Pala Students Win ‘Best New Team’ Award at the 2022 DataJam” was posted (https://beta.nsf.gov/science-matters/pala-students-win-best-new-team-award-2022-datajam?utm_medium=email&utm_source=govdelivery). This article described the project that the Pala team undertook for the 2022 *DataJam*. NSF News is a free newsletter widely distributed in the United States and around the world. Ken Chiacchia at the Pittsburgh Supercomputer Center also wrote an article about the national expansion of the *DataJam* that was published in October 2022 (<https://www.psc.edu/student-data-science-competition-2/>), and in UC San Diego newsletter (<https://today.ucsd.edu/story/pala-native-american-youth-receive-datajam-award-with-guidance-from-sdsc>).

Presentation of the DataJam at National Meetings

In July 2022 Kim Bruch, who had served as the advisor for the advisor for the Pala *DataJam* team and Louise Hicks, the *DataJam* mentor for the Pala team, presented a paper and poster at the *Practice & Experience in Advanced Research Computing (PEARC 2022)* meeting, entitled “Developing a Data Science Outreach Program with Rural Native Americans: Southern California Tribal Youth Participate in *DataJam* via San Diego Supercomputer Center.” This brought the *DataJam* national attention and university faculty and high school teachers who attended the conference were able to learn more about the *DataJam* in evening sessions that Louise ran to inform participants about educational opportunities. For this meeting we designed a tri-fold brochure that explains succinctly and with a number of pictures and graphics what the *DataJam* is and how to get a team involved. We have used this brochure extensively in the last several months to provide information about the *DataJam* to all who are interested, and this now can be downloaded from the Pittsburgh DataWorks website.



Kim Bruch from the San Diego Supercomputing Center (left) and Louise Hicks, *DataJam* mentor, presenting a poster about the Pala project at the PEARC 2022 meeting in Boston in July 2022.

The DataJam Download – A monthly DataJam Newsletter

As more teams joined the *DataJam* from across the country, students started asking for updates about what was happening with the *DataJam* and if there were any updates about new resources available for their projects, reminders about dates then things were due, information about the program and mentors. We decided it was time to start publishing a monthly newsletter and distributing it by email and on the Pittsburgh DataWorks website to everyone involved in the *DataJam* and all those who request information about the *DataJam*. The first issue was published in June 2022, and monthly issues are now sent out at the first of every month. They highlight recent news about the *DataJam*, brief articles by *DataJam* Mentors, information about new resources available to *DataJam* teams, and reminders about upcoming due dates for the 2023 *DataJam*. The newsletter has also become a great way to introduce new groups to the *DataJam* and increase the interest in the *DataJam* nationally.

Impact of the DataJam for DataJam Team Members and Mentors

Since 2015 we have many examples of ways that participating in the *DataJam*, either as a high school student or a mentor, has had an important impact on participants. This year we were delighted to receive an unprompted communication from a prior member of the 2021 *DataJam* team that won first place, and had had the opportunity with his teammates to present their *DataJam* project to the Data Analytical team at PPG Industries, an important sponsor of the *DataJam*. The student contacted Devashih Saxena, the Vice President and Chief Digital Officer at PPG, through *Linked In*, and indicated that having the opportunity to present their data at PPG to the global data analytical team had had a big impact on him and he was now in college at UCLA and majoring in data science.

Participating in the PEARC 2022 meeting had a similar large impact on Louise Hick’s career trajectory. Louise mentored not only the Pala team, but two other *DataJam* teams in 2022 and at PEARC 2022 she was introduced to a young entrepreneur with a start-up data science company, who was very interested in hiring her, in part because of her mentoring experience. She is now working at this company in the interim year between her undergraduate education and graduate training she hopes to take. Louise wrote a column for the September 2022 issue of the *DataJam Download* to share her experiences mentoring and its impact on her career trajectory. One of the goals of the way the *DataJam* is designed, with near-peer mentoring by college students for high school teams, and connections with industry data analysts who can connect to both *DataJam* teams and mentors, is to show youth career pathways in data science. Thus, we have instituted “Meet the Mentor” sections into the regular format of our newsletter. In the October 2022 issue Jackson Filosa, a student who has mentored throughout his college years, talks about how his exposure to data science allowed him to get a summer internship at PricewaterhouseCoopers this summer, and in fact they offered him a job this month and he will be working there after he graduates from college in April 2023. Similarly, Anthony Lucchiti who took the *DataJam* mentor training course and mentored several teams last

year, became a data science intern at Capital One's Technology Internship Program, and hopes to go into the data science field when he graduates this spring, and wrote about this in the October newsletter.

Teaching the DataJam Mentor Course in a BSLE format to include students and faculty at other universities

For the past five years a *DataJam* mentor training course (STAT 1050: Using Big Data for Community Good) has been coordinated by Dr. Cameron at the University of Pittsburgh in order to give college students who want to serve as *DataJam* mentors background necessary to be a good mentor in terms of background knowledge in statistics, in the ethics of collecting community data, and in how to work effectively with diverse communities. Over the past year, as we recognized that we were being successful in expanding the reach of the *DataJam* to communities outside of Western Pennsylvania, we realized that we need to expand *DataJam* mentor training to other colleges and universities in a diversity of communities. To do this we decided to use a *Blended Synchronous Learning Environment (BSLE)* format. Using this format, the course taught at the University of Pittsburgh is videoconferenced to other participating universities. We have now established collaborations with Caldwell University in New Jersey, Duquesne University in Pittsburgh, University of Pittsburgh at Bradford in Bradford, Pennsylvania, and UCSD in California. This Fall we now have students at Caldwell University taking the course, and faculty at Duquesne, Bradford and UCSD participating in teaching. We are very excited to continue to expand this format in the future.

Awards for the DataJam

In August 2022 the Carnegie Science Center announced that Pittsburgh DataWorks had won the "Best Interdisciplinary Approach to STEM Education Award" for 2022. This was awarded on October 13, 2022. The published award summary highlighted the expansion effort supported by the pilot project. "Pittsburgh DataWorks has been proactive in developing resources to encourage DataJam participation to a wide diversity of communities. Currently, the University of Pittsburgh offers a university course that attracts students from many majors to be a mentor at Pittsburgh DataWorks. The course focuses on using free software for data visualization and analysis. Expansion of the course, to become a national resource, is planned for later this year using a collaborative blended simultaneous learning environment (BSLE) strategy connecting the opportunity to universities around the nation. The collaborative, interdisciplinary approach will allow mentors to be trained in a variety of locations and receive increased training in how to facilitate high school teams participating in DataJam in under-resourced, urban environments, rural environments, and on Native American reservations."

Grants to further expand the DataJam

To continue the work initiated in the pilot project or expanding the *DataJam* nationally, as well as the *DataJam* mentor training, we have started to apply for grants that would support the expansion efforts. The first such grant we submitted in August 2022 was a proposal to the National Science Foundation, "ITEST Developing and Testing Innovations (DTI): *DataJam-Online*." The goal of the grant was to develop this proposed scalable, adaptive online program that will effectively engage a wide diversity of youth in learning data science skills, as well as to study its effectiveness, we will use a social cognitive career theory (SCCT) lens to address the following research questions: RQ1: What cultural assets and funds of knowledge do underrepresented learners from diverse communities bring to forming and sustaining local DataJam-Online nodes? And what effect does that have on student and teacher participation? (ITEST Pillar 3: Strategies for equity in STEM Education); RQ2: What effect does moving mentor training, coursework and mentorship of participating high school students to a collaborative blended synchronous learning environment (BSLE) strategy have on student and teacher participation in DataJam-Online? And how will BSLE mentor training, online assets and resources affect student skills, agency and the ability to apply knowledge across diverse communities of learners? (ITEST Pillar 1: Innovative use of technologies in learning and teaching); RQ3: What effect will adapting DataJam-Online to a variety of cultural contexts have on building sustainable collaborations with a diversity of businesses and industries, and a diversity of colleges and universities, to support self-efficacy in DataJam-Online participants? (ITEST Pillar 2: Development of key partnerships for career and workforce preparation).

Broader Impacts

This pilot project is serving as both a bridge to the dissemination of adaptive practices for equitable integration of data science across learning settings, and as a first step in forming networks among

communities to increase participation in data literacy and data science skills. All resources and strategies that were developed as part of this pilot project have been made available on the Pittsburgh DataWorks website (<https://www.pghdataworks.org>) and are being shared with the NE Big Data Innovation Hub and made available on their website. We believe the model of participatory design to advance the state of data literacy can be adapted for other regions and demographics, leveraging existing and emerging resources, and allowing other partnerships to run DataJam competitions in their regions.