PITTSBURGH DATAWORKS

The DataJam Download

Official Newsletter of the Pittsburgh DataWorks

See the New Guide on Asking DataJam Questions about Public Health, on the Resource Page of the DataWorks Website, Developed by Yunge Xiao!

Public Health for Social Justice

Social justice refers to the idea that everyone deserves equal economic, political and social **rights** and **opportunities** (San Diego Foundation).

Public health scientists believe all individuals should have equal opportunities to achieve adequate health and well-being.

Public health focuses on which communities are in greatest need and **how** to best address these needs.

A New Resource: How to Ensure Research Data Quality

Each Fall we train a new group of college students who want to become DataJam mentors, how to provide effective mentoring for high school teams participating in the DataJam. This is done by offering a college course that teaches the basics of statistical analysis and how to teach it, introduces students to the issues of how to ensure that DataJam projects use data that represents whole populations (those in urban & rural populations, emigrant communities, Native American reservations, communities where English is the second language), introduces them to how government and businesses use data, and teaches them how to make an effective poster presentation and slide talk. As part of this course, new mentors make resources for DataJam teams.

The course is now offered at several universities by videoconferencing. This year students at Caldwell University in New Jersey joined the DataJam Mentor course for the first time! Two of these students, Anusha Panday and Nishant Pokhrel, made a resource entitled "How to ensure research data quality". You can find this resource on the "Resource" page of the DataWorks (pghdataworks.org) website. If your team follows the advice given in this guide it will be much more likely that you avoid errors in the datasets you use, ensuring that the results you get in your analyses will be meaningful. We strongly suggest that all teams read through this short, easy to use document NOW – when you are just starting your data analyses!

Lots of other resources made by DataJam Mentors are available at pghdataworks.org. Take a look at all of them so you don't miss seeing something that could help with your project.

How to ensure research data quality

Anusha Pandey and Nishant Pokhrel, 2022 Data Jam Mentors

1. Plan properly

An effective way to run a project is by being on top of every single detail.

Everyone in the team should understand the objectives of research questions. This will help team members identify accurate data sources and understand target audience for surveys.

2. Maintain good data storage

It is very easy to lose track of datasets that you come across during research. Make sure your team has a centralized database before starting out so you do not misplace or lose any potential source. You can use Google Sheet/Docs, Noodle tools, Excel Sheets or whatever is convenient for your team.

3. Be consistent in entering datasets

What is your data entry process looking like? Are you copy/pasting your data from sources or are you manually entering values? Maybe you are surveying? Make sure you are consistent while entering your data and that everyone in your team follows the same method.

4. Cite Data Sources Correctly

Data analysis requires citations to help acknowledge original contributors and to help others get to the same resources as you. There are various ways to cite data sources. Make sure you stick with a consistent format throughout your research.

5. Document your data

You might have to revisit data sources in the future. It will be much easier to do so, if you document your rationale for choosing and using particular data sets.
It is also a good idea to make sure your thought processes are captured so that your team members can validate your reasonings or logics behind it as well.

6. Ask for feedback

Your mentors and teachers are always there to help you. If you have questions about any data source, make sure you ask them before analyzing the data; faulty datasets give faulty outcomes. Your data sources should not be biased. They should always be credible.

Meet the Mentors

Yunge Xiao

Hi all! My name is Yunge Xiao and I am a graduating statistics major and sociology minor at the University of Pittsburgh. Beyond the classroom, I love to play ultimate frisbee, pickleball, write songs, and share delicious food with others!

I find great joy in serving my community, I am always looking for creative ways to call upon data to inform efforts that support overall health. Consequently, I enrolled in Dr. Cameron's course at Pitt, DataJam: Using Big Data for Community Good, in hopes of being well-equipped to do so as a mentor. Being a DataJam mentor this past year has been a real treat: there is nothing more exciting than watching my younger peers learn and apply knowledge in meaningful ways.

I have been greatly inspired by the intersections of statistical research methods and community health throughout my college career. My passion for service materialized after interning with a local human rights coalition, where I was able to connect with residents, activists, and organizations to develop solutions for social justice issues. I partnered with a transit equity organization, Pittsburghers for Public Transit (PPT), to help them design and promote a pilot program to alleviate transit fare burden for SNAP (Supplemental Nutritional Assistance Program)-eligible populations. Researching for PPT developed my interest in understanding how having access to the social determinants of health related to community well-being.

Shortly after, I started volunteering for the food pantry on my college campus, Pitt Pantry, where my interactions with shoppers accentuated the gravity of food insecurity for my peers. An executive board opening gave me a chance to apply my statistics coursework to my work at the pantry. Through collection and analysis of pantry survey data, I created reports for community partners and informed better shopper service.

I am constantly reminded of the important and useful role that data plays in our current world as more public datasets and data science tools become increasingly available. I am so grateful that competitions like the Pittsburgh DataJam exist and to be able to play a role in helping high school students cultivate and navigate their passions for problem-solving. This upcoming Fall, I will be pursuing a graduate degree in epidemiology in hopes of becoming a skilled and considerate leader in tackling public health problems. I am so excited to continue playing a role in creating resources and initiatives that empower our communities!

Minchan Kim

Hi all! My name is Minchan Kim, and I am currently a freshman studying Data Science at the University of California – San Diego. I love applying statistics with computing technology and seeing how it can be manipulated and visualized to suit an intended audience. As it is my first year as a DataJam mentor, I am really excited to see how these skills will be used in a multitude of projects on varying different subjects by DataJam teams.

Although I am deeply in love with using Pandas, Numpy, R, Python, and many other ways to "do" data science, I fell in love with the subject with my AP Statistics class in high school. Learning about hypothesis tests, correlation, regression, and much more really put into perspective how numbers are intertwined in our everyday world, besides the other natural sciences. Seeing data presented on the news or online material like YouTube made me appreciate and understand the essence of being a statistician, as their power is in showing data that can be most applicable in context. In fact, my favorite ways to wind down after a hectic day is to be surrounded by the world of statistics in basketball, social media, or even Pokémon.

Self-learning is hard. Having the motivation to keep going without the pressure of grades or hard deadlines can create laziness. However, with a program like DataJam, I cannot wait to help others learn and troubleshoot without them worrying of never finishing what was in the process of being learned. I have always wished that I was exposed to this world when I was younger so I could start learning earlier, and with this program, I hope to introduce and help spread the world of data science as early as possible. In tangent with helping kindle student's inherent passion to learn more about the world that lays beyond the confines of secondary school, my greatest wish is to show students how their outside passions relate to data science so that they understand the new and evolving world that is filling up with more and more data is made for us: data scientists.

As I work with DataJam teams, I hope to enhance my own skills and be able to learn how to mentor effectively so that participants may return for more in the following years. By helping students take their first step, their growth will only be inevitable.
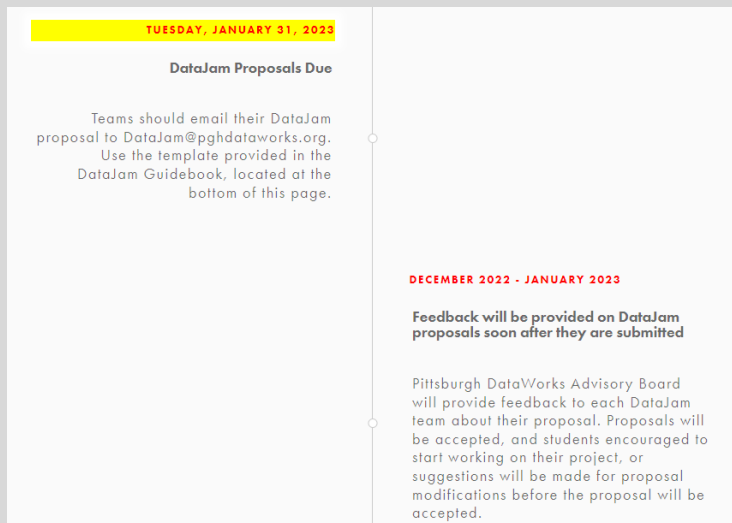
Meet the Data Science Professional

I am Judy Cameron, the Director of Pittsburgh DataWorks, and the person who teams communicate with when emailing datajam@pghdataworks.org. I grew up outside of San Francisco, CA and went to college at the University of California at Berkeley, initially as a Physics major. However, in my senior year in college I took at Physiology course and studied how things you do in everyday life influence the way the brain works. This was a subject I really loved and switched my career plan to a whole different topic. This is the reason I always tell students you should try a variety of courses --- you'll never know you love something if you don't try it!

I'm a professor at the University of Pittsburgh and the research my lab does examines how exercise, nutrition and stress all influence brain function. We use data analysis to determine answers to our research questions. For example, we have shown that exercise protects the brain against age-related decreases in function. You might ask what type of exercise is best? Or how much exercise do you need to do? We are running experiments to answer these questions and data analysis will be required to determine the answers.

I also run science outreach for the university, and it is through that aspect of my job that I first became acquainted with Pittsburgh DataWorks in 2015. At that time the DataJam was small, just 3 schools were involved. I offered to have college students serve as DataJam mentors and by providing this additional help to DataJam teams, the DataJam was able to grow and reach more and more schools every year. At first, we reached just schools in the Pittsburgh area, but then with the Covid-19 pandemic we needed to convert all mentoring to videoconferencing and that allowed DataJam to spread to other states, even as far as California. We are now working to make DataJam a national program, that is available throughout the country, and train DataJam mentors throughout the country. We are also working on developing new ways to get middle school students excited about data science!

DataJam Timeline

TUESDAY, JANUARY 31, 2023

DataJam Proposals Due

Teams should email their DataJam proposal to DataJam@pghdataworks.org. Use the template provided in the DataJam Guidebook, located at the bottom of this page.

DECEMBER 2022 - JANUARY 2023

Feedback will be provided on DataJam proposals soon after they are submitted

Pittsburgh DataWorks Advisory Board will provide feedback to each DataJam team about their proposal. Proposals will be accepted, and students encouraged to start working on their project, or suggestions will be made for proposal modifications before the proposal will be accepted.

*Note – if your team is still working on your 2023 DataJam proposal and need a little extra time, just contact us at datajam@pghdataworks.org and let us know.

We are looking forward to DataJam 2023!
We Hope You Are Too!

Email us at datajam@pghdataworks.org when you are ready to start working with a DataJam Mentor!