The DataJam Download
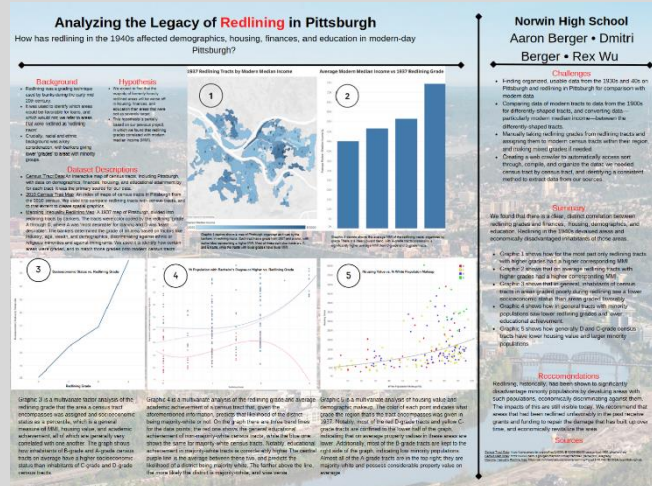
Official Newsletter of the Pittsburgh DataWorks

Pittsburgh DataWorks is the winner of the 2022 Carnegie Science Award for "Best Interdisciplinary Approach to STEM Education"

Pittsburgh DataWorks has been proactive in developing resources to encourage DataJam participation to a wide diversity of communities. Currently, both the DataJam and the university course that trains DataJam mentors, are being expanded nationwide using a collaborative blended simultaneous learning environment (BSLE) strategy. 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, in immigrant communities and on Native American reservations. The awards ceremony will be October 13, 2022.

Norwin High School DataJam Team presented their project "Analyzing the Legacy of Redlining in Pittsburgh" at PPG on 6/16/22

Congratulations to the Norwin High School DataJam Team! As part of the First Place prize for the 2022 DataJam, the team was invited to give their DataJam project presentation to the data analysts attending the **PPG Digital Organization Town Hall** on Thursday, June 16, 2022. Members of the team, Aaron Berger, Dmitri Berger and Rex Wu, accompanied by their teacher, Ciminy St. Clair, presented their project entitled "Analyzing the Legacy of Redlining in Pittsburgh" and fielded questions from members of the PPG data analytical team from around the world. Click here to learn more on our website!

DataJam is presented at a Computing Conference in Boston in July 2022

By Louise Hicks, DataJam Mentor

The world of big data is expanding and so is the DataJam community. This summer I got the opportunity to represent Pittsburgh DataWorks at a conference called Practice and Experience in Advanced Research Computing 2022 or PEARC '22 for short. The conference took place in mid-July in the heart of Boston, Massachusetts and during it I was able to present a poster on some of the amazing work done by DataJam students. The title of the poster was *Developing a Data Science Outreach Program for Rural Native Americans: Southern California Tribal Youth Participate in DataJam via San Diego Supercomputer Center*. The poster detailed the DataJam project completed by a group of students from the Pala Tribe located on a reservation in Southern California, and won the Best New Team award in the 2022 DataJam competition. Additionally, I had the chance to hand out fliers and speak with individuals in greater detail about Pittsburgh DataWorks at a booth dedicated to DataJam. As I described the DataJam competition and the positive impact it has on students, individuals from throughout the computing industry (educators, researchers, industry professionals, etc.) took interest in contributing to the cause of data science education through DataJam. And so, the DataJam students once again have done what they do best -- inspire me and others to imagine a world where big data is used to benefit and better our communities.
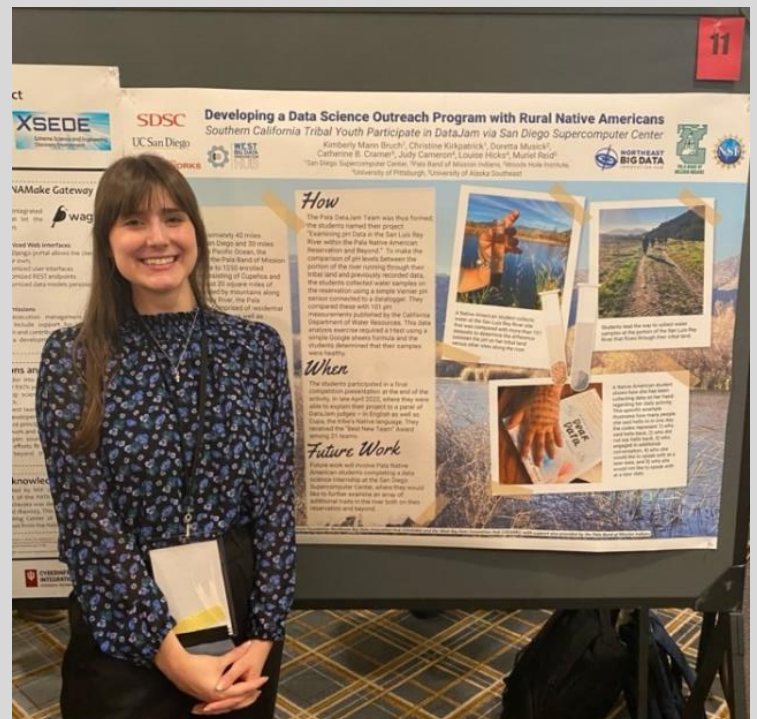
Meet Two DataJam Mentors and Their Resources

Jatin Singh

Hello! My name is Jatin Singh. I'm a junior at Pitt, majoring in emergency medicine, and a second-year mentor for the DataJam. A little about me – I mentored two high school DataJam teams in the Pittsburgh area last year and I'm looking forward to continuing to mentor. I like skateboarding, thrifting, and playing volleyball for Pitt in my free time.

This summer, I developed one of the DataJam's first Dataset Guides. This guide, titled **MedJam**, **serves to help DataJam teams tackle public health and medical topics**. In the guide, I list common diseases and public health challenges that affect the US. I also scoured the internet for health datasets that teams could utilize with ease – both from local and national databases. Data science is heavily utilized in the medical and public health fields to track population health and to improve diagnostic and interventional procedures. We, as the mentoring team, want to encourage teams to explore the intersection of health and statistics. We hope that the guide will be helpful, and I am here as a resource for teams looking at questions pertaining to health. Feel free to reach out to me with any questions.

Lucas Troy

Hi! My name is Lucas Troy. I am a Senior at Pitt studying computer engineering and a DataJam mentor. I work in automation and industrial control systems engineering while I am on co-op, and I hope to work full time in this industry. I enjoy programming, skateboarding, spending time with friends, and solving complex problems. I have mentored some DataJam teams in the past, but I now mostly work on the Pittsburgh DataWorks website, newsletter, and educational resources.

Speaking of educational resources, I was honored to have the opportunity to develop one of the first Dataset Guides! I call my guide **EconomicJam**, as it **covers topics surrounding the US and Global Economies and their possible effects on public health.** Relevant economic principals such as inflation, supply and demand, and recessionary indicators are covered in this guide. To understand each principal, I cover the general definitions, potential research questions, and datasets that explore the questions with how to use them. Economics can be a difficult topic to tackle, especially if this is your first time seeing this type of material, so that's why I developed an Economic Guidebook to go with it. **The Economic Guidebook will explain complex statistics and economic principles in an easier to understand way.** If you decide to use my Dataset Guide, be sure to have the Guidebook open alongside for those concepts that may seem confusing at first. If you have an interest in economics, I hope this guide helps you explore that interest to its fullest potential. I am happy to answer any questions you may have.

Python for the DataJam

By Nate McDowell

Hello! My name is Nate McDowell and I am a senior at the University of Pittsburgh. I was honored to be part of the Pitt Honors 2022 Community Research Cohort this summer. For my project, I created a resource for that will introduce DataJam teams to Python and GitHub. This is not a resource for those just working with statistics for the first time, it is a resource for teams that wish to learn more about computer coding and experiment with more sophisticated analyses in their DataJam project.

Python for the DataJam is a newly created tool that aims to teach DataJam teams how to implement Python and GitHub into their DataJam projects. This resource includes a mock DataJam project that shows students the process of cleaning datasets, creating visualizations, and gathering insights from various statistical tests. It is my hope that teams will now have access to materials that will help grow their budding data science skills. This project was made with help from DataWorks board members Dr. Cameron and Brian MacDonald, as well as Everett Herman of the Pitt Honors college. I hope that this resource pushes the DataJam forward and continues to develop for years to come.

To use these resources, click on the above link. This will take you to the main page of the repository, where there are a few files. Start with the *GitHub_Introduction* read the introduction then complete the *First_Steps* guide. After completing the above steps, you should have a working knowledge of GitHub, a new account, and the GitHub Desktop application. Next, navigate to the *python_tutorial* and work your way through the tutorial (*Data_Cleaning.ipynb, Statistical_Analaysis.ipynb, Visualizations.ipynb*). All the datasets you will need to work your way through the tutorial are available in the datasets folder. Once you have learned to use GitHub and Python we hope these will be valuable tools for you to collaborate with your team as you take on a research question of your own!

See the DataJam Timeline for 2022-2023

On the DataJam page of the website the new 2023 DataJam Timeline has been posted. Click here to see the Timeline.

- **Proposals due Fri., Dec. 2, 2022**
- **Feedback will be received by Fri., Dec. 16, 2022**
- **Posters will be due Fri., March 31, 2023**
- **2023 DataJam Finale will be Thur., April 27, 2023**

MID-AUGUST THROUGH SEPTEMBER

Schedule a DataJam Presentation at Your School

DataJam mentors are available to give a 15-minute presentation at your school to inform students about what the DataJam is, how to form a team, and what is involved in participating in the DataJam. Presentations can be scheduled by requesting a presentation at Pittsburghdatajam@pghdataworks.org. Presentations will be given by zoom. Alternatively, watch the 15-min DataJam Mentor Overview to the DataJam video found at the top of this page.

Register to be a 2023 DataJam Team

It is easy to register your team. Send an email to PittsburghDataJam@pghdataworks.org to let us know you have a team that would like to compete in the 2023 DataJam. Include: (1) the name of your school or group, (2) the name, email address and phone number for the team advisor (a

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

Just email us at pittsburghdatajam@pghdataworks.org when you are ready to start working with a DataJam Mentor!