

The DataJam Download

Official Newsletter of The DataJam



***New Help with
Finding DataSets***

***Three New DataSet
Guides***

***Meet the DataJam
Mentors***

***Meet the Data
Science
Professionals***

***New Section!
DataJam Team
Updates***

***DataJam Timeline
For 2023-2024***

*****See Revised
Proposal Deadline
– January 26, 2024***

New Help with Finding DataSets for Your DataJam Project



Discover the essence of open data with our comprehensive guides, tailored to demystify datasets from diverse sources. Navigate to the Datasets tab on the DataJam website. Customized guides on different topics are listed first and then international, national, and regional dataset links are listed. NEW this month are guides for using some of the regional datasets. You can recognize them as they have the same logo as the dataset but say "Guide To:". Happy exploring and analyzing!

Three New DataSet Guides to Publically Available DataSets

As introduced on page 1 of this newsletter, we are starting to make guides to provide guidance for DataJam teams on how to use large publicly available datasets to answer questions. As with our other DataSet Guides, each of these guides provides some background information about the types of questions that can be answered with the data available in the DataSet, provides directions for how to access the data, and makes suggestions for how the data can be analyzed to answer research questions.

Guide to Informed Decision-Making in the Bay Area

The San Francisco Bay Area, renowned for its diverse communities, thriving tech industry, and picturesque landscapes, boasts one of the world's strongest economies. However, this prosperity has brought forth a myriad of pressing challenges that significantly impact the community's well-being and overall quality of life. These include a housing crisis and wide-spread homelessness, problems with transportation and increased crime rate.

This guide suggests several possible research questions that DataJam teams could ask about these issues. It then provides step-by-step guidance on how to find and download data at <https://datasf.org/opendata/> to answer the questions.

The authors of this guide are DataJam Mentors at San Francisco State University, who were featured in the November 2023 issue of this newsletter.

Guide to the WPRDC

The Western Pennsylvania Regional Data Center (WPRDC) is a fantastic source for open data in Pittsburgh and the surrounding region, with over 330 datasets that are easily accessible, machine-readable, free to use and share. The site also has some very useful tools to look at Economic Indicators, Community Profiles, Greenprint (which allows map making), and Housecat (which provides information about properties).

The guide was developed by Lauren Krest (featured in the December 2023 issue of the newsletter) and Sarah Sirakos (featured in this issue of the newsletter), DataJam Mentors at the University of Pittsburgh.

Guide to Answering Local Questions (Allegheny County)

Allegheny County Analytics is a new open data portal for the county that Pittsburgh, PA is located in. This site was started in December 2023, and will be a place to watch as it grows and has more publicly available data. The goal of the site is to provide an information hub to allow better understanding of issues affecting Allegheny County residents.

The guide was developed by Dan Kun and Daniel Hufnagle (featured in this issue of the newsletter), DataJam Mentors at the University of Pittsburgh.

Guide To:



A DATASET GUIDE TO INFORMED DECISION-MAKING IN THE BAY AREA

BY DONA INAYAH, CARLOS DE LEON, KARTIK PATEL, AKEMI SMART

- I. INTRODUCTION
- II. UNDERSTANDING DATA
- III. IDENTIFYING LOCAL ISSUES IN THE BAY AREA
- IV. EXPLORING RESEARCH QUESTIONS
- V. FINDING DATA
- VI. CHALLENGES YOU MIGHT FACE
- VII. TIPS & ADVICE

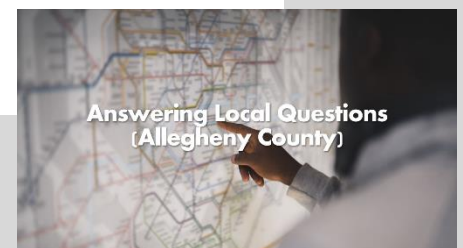
Guide To:



Guide to the WPRDC

Lauren Krest and Sarah Sirakos
University of Pittsburgh

Guide To:



Meet the Mentors

Hi! I'm Sarah Sirakos, a junior Statistics major and Economics minor at Pitt! My love for statistics started a long time ago when I took my first statistics class in high school, and I knew since then that it's what I want to pursue for my career. Statistics and data science is so useful in every single field, and I absolutely love how the opportunities are endless. Spreading the word of how valuable statistics are through a program like DataJam is extremely important for the younger generation!

One of my passions outside of DataJam is helping my community. I am currently the Community Service Chair of Delta Phi Epsilon at Pitt, and I recently hosted a donation drive and collected over \$100 and many essential items to donate to a local women's shelter in Pittsburgh. I've also volunteered for the Cystic Fibrosis Foundation, located here in Pittsburgh, and I had over 200 hours of community service back in high school. I come from a small town called Houston about 40 minutes south of Pittsburgh, so I've always been used to a close-knit community that looks out for each other. DataJam has those same traits, because everyone cares about one another and is looking for ways to get everyone to succeed!

In the future, I hope to use my passion for statistics and become a data analyst. I would love to be able to come up with ways to help our community here in Pittsburgh, or wherever I end up years from now! I won't mind which field I end up in as long as I get to do statistics. Bringing good energy, good ideas, and good actions through data science is my main motivation for my outlook on the future. Right now, I'm trying to figure out how to pursue statistics over this summer in Pittsburgh! I've connected with a few past DataJam mentors, and they've given me the best advice and input. DataJam is really special to have given me that opportunity to speak with people who are just as passionate about data science and statistics as I am!

As for this upcoming semester, I am extremely excited to get to mentor for DataJam. I cannot wait to see what the high school students come up with, and how I can help them on their journey. I hope that I can help all of the DataJam students share that same excitement and passion that I had when first learning statistics!



Hello! I'm Daniel Hufnagle, and I'm a freshman at the University of Pittsburgh from Lancaster, Pennsylvania. Currently, I'm working towards a computer science major with an electrical engineering minor. I'm delighted to be a DataJam mentor for this year. I've found my prior statistics and data science experiences to be both very enjoyable and very meaningful, and I hope that you all will as well while you complete your DataJam projects.

Beyond classes and my work with the DataJam, I'm involved with robotics, where right now I'm helping write the code for a wearable haptic to provide an extra mobility aid for people with blindness and writing more code for an autonomous boat. Most of my robotics work, past and present, has a lot to do with computer vision and training machine learning off large image-based datasets.

Aside from that, I enjoy photography and I spend much of my free time exploring the city, looking for interesting photos. I look forward to helping our teams out with both completing their projects and having fun exploring how data and thoughtful analysis can help solve real problems in our communities. Good luck and have fun with your projects!

Meet the Data Science Professionals

Hello, my name is Luella Fu, and I am an assistant professor of Statistics in the Mathematics Department at San Francisco State University (SFSU). At SFSU, I am involved in our Bachelors degree in Statistics as well as a Master's of Statistical Data Science program. I also teach a lot of computer science students. At SFSU, we strongly emphasize the mathematical backbone of data science so that students better understand when to, when not to, and what to question about methods used to analyze data. We also serve a population of students who mostly want to take their skills into industry jobs after graduation.

Because of these reasons and SFSU's very strong interest in supporting its community, I wanted to involve SFSU in DataJam. It was just what I was looking for to round out the Statistics courses we offer our students! Students could both gain practical experience during their training course and enforce their learning by teaching high school students. Additionally, we would be able to expose local high school students to San Francisco State's statistics programs.

Over the mentor-training term, I've watched my students grow their data science skills and increase their engagement with hands-on data analysis. It has been incredibly rewarding to see students help each other develop a variety of industry skills, such as project management, learning Tableau, and learning Python. They have also been helping each other apply to internships. Carlos, one of our students, has been leading the charge. In an incredibly competitive pool of over 1000 applicants, he made it to the tiny pool of 10 students who were interviewed for a data science job at a start-up in San Francisco. The secret to his success? While I think a lot of this is his persistence, Carlos also mentioned that he puts all his data science projects from his various classes on his resume and his interviewers actually ask about them. I agree that having data science projects does make a resume stronger.

I've also been very proud to witness how my students face technical challenges like how to present virtually when broadcasting from one computer that doesn't have a strong enough microphone to capture all their audio. Their first DataJam project was a little rocky because of IT issues and a persistent echo on Zoom, but they came in with their own equipment to fix the problem for their second project and did wonderfully. They've also implemented algorithms beyond the scope of their mentor training, like the k-nearest-neighbors algorithm to handle missing data. This was no easy feat, as the code was extremely buggy at first, but they persisted and successfully implemented the algorithm.

In general, student engagement has been truly amazing: one of our students, Dona, is even starting a Data Science Club at SFSU, where students can work on data projects like those in DataJam! Our other DataJam students, Akemi, Carlos, and Kartik, are going to be part of the club as well. With all that has already happened in the Fall term, I can't wait to see what happens next term when my students share their skills with high school students!

Hi, my name is Nicholas Horton, Ph.D. (Amherst College), and I'm the editor of the open-access Journal of Statistics and Data Science Education (JSDSE), which is published jointly by the American Statistical Association and Taylor & Francis. The journal was founded 32 years ago as an electronic journal with no author publication fees. I'm admittedly biased, but I believe that the journal is a place where educators and researchers can share their ideas, activities, approaches, and methods to improve data science and statistics education. I encourage you to take full advantage of it!

I came to data science from an interest in study design and data analysis in psychology. [My work](#) with students has been focused on helping them develop data acumen and to make sense of the data that surround them.

I often utilize articles from the journal in my teaching. Examples include the special issue on [computing in the curriculum](#) and activities to allow students to experience the entire [data analysis cycle](#). There have been an increasing number of articles focused on data science at the K-12 level, such as this [one](#), and that [one](#).



Our most recent issue leads off with a paper by Lee Kennedy-Shaffer that recounts the difficult past of statistics, notably the role of eugenics in the development of the field. He argues that we must confront our history in order to move forward and offers three examples of famous statisticians and their work on eugenics to ground the discussion, with [guidance](#) about how to address the troubling history in the classroom.

Another favorite of mine is [Ellis and Slade's paper](#) on considerations for teaching with ChatGPT and other generative AI tools. These tools are likely to be transformative in their ability to introduce new technologies and tools to students (albeit with a related need for them to be able to judge how these omniscient and sometimes overconfident tutors might mislead them).

We welcome readers, reviewers (please reach out to me to be added to the list), and authors (remember: there are no author publication charges). The journal can be found online [here](#)!

New Section! DataJam Team Updates

We are excited to introduce "DataJam Team Updates" starting with the February 2024 issue of the DataJam Download! This section will provide space in the newsletter for DataJam teams to write a short paragraph about the project they are working on and provide one figure. Our goal is for teams across the country to have more communication about their DataJam projects with each other. If your team would like to provide a DataJam Team Update for any issue of the newsletter, please email us at datajam@thedatajam.org

DataJam Timeline for 2023-2024

On the DataJam page of the website the new [2024 DataJam Timeline](#) has been posted. Click [here](#) to see the Timeline.

- **Proposals due Fri., Dec. 8, 2023 with an EXTENDED DEADLINE UNTIL FRI., JAN. 26, 2024**
- **Posters will be due Fri., March 29, 2024**
- **2024 DataJam Finale will be Thur., April 25, 2024**

**FRIDAY, DECEMBER 8, 2023 -
FRIDAY JANUARY 26, 2024**

DataJam Proposals Due

The DataJam will review proposals submitted on a rolling basis, with feedback provided to teams within a couple weeks of their proposal submission. 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.

Make sure the names of the team members (there must be at least 2 team members, but often are 5-8 members), and non-school email addresses are provided (note many schools block emails to student school emails from outside email addresses). Once team members have been identified, you and the students will be invited to join an Edtera Group for DataJam 2024. This is an online platform that every team can access on an internet browser where they can communicate with their team members about their project and also communicate with DataJam mentors to get help with their project.

The instructions for writing the DataJam Proposal are on page 5 of the [DataJam 2024 Guide Book](#), and a template for the one page DataJam Proposal is on page 6. The guidebook can be downloaded from the DataJam page of the Pittsburgh DataWorks website.



DataJam Guidebook - 2024

Seeking inspiration for your DataJam project topic? Explore the [past DataJam projects from 2015 to 2023](#), available at the bottom of the DataJam page, along with links to all the posters. Additionally, check out the [DataSet Guides](#) on the DataSets page for more ideas.

2023

2023 DATAJAM POSTERS

<p>Avonworth Something in the air</p>	<p>Central Dauphin: Team 2 Natural disasters and climate change</p>	<p>North Allegheny: Team 1 Standardized testing: A reflection on intelligence or economic environment?</p>	<p>Passaic Academy for Science & Engineering: Team 1 Virtual vs. reality</p>
<p>Bethel Park: Team 2 Fast food restaurant location density vs. obesity rate</p>	<p>Hampton Improvements to Pittsburgh's city parking lot distribution</p>	<p>North Allegheny: Team 2 Close calls with high cholesterol</p>	<p>Passaic Academy for Science & Engineering: Team 2 Empty streets, cleaner sheets</p>
<p>Bethel Park: Team 4 The effects of social media on teenage sleeping habits</p>	<p>Hempfield Is there a correlation between race, income level, or proximity to fast food restaurants with diabetes rates in Pennsylvania counties?</p>	<p>North Allegheny: Team 4 Diversity: The correlation of diversity and publications</p>	<p>Passaic Academy for Science & Engineering: Team 3 Factors affecting air quality in NY state</p>
<p>Boys & Girls Club of Pittsburgh Pittsburgh regional transit - Vehicle age and why it matters</p>	<p>Keystone Oaks: Team 1 Café GPA</p>	<p>Norwin Redlining: A broader perspective</p>	<p>Passaic Academy for Science & Engineering: Team 4 Correlation between income & crime in NJ counties within minority groups</p>
<p>Carlynton How technology affects teens in car crashes</p>	<p>Keystone Oaks: Team 2 Cows vs. cars</p>	<p>Oakland Catholic Detention or detainment: Does higher enrollment in schools impact crime rates?</p>	<p>South Hunterdon Effects of cancel culture on the attention a celebrity gets on social media</p>
<p>Central Dauphin: Team 1 Improving Central Dauphin's (G)IEP Program</p>	<p>Maple Grove Eurasian Milfoil's abundance in the central & south basins of Chautauqua lake in proportion to other invasive species after being sprayed with herbicides</p>	<p>Pala/Torrey Pines Comparing western bluebird populations throughout Southern California</p>	

We are looking forward to DataJam 2024! We Hope You Are Too!

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