

The DataJam Download

Official Newsletter of The DataJam



**DataJam Posters
Due March 29th**

**DataJam Team
Visits Merck!**

**DataJam Team
Presents DataJam
Project at National
Meeting**

**Meet the Mentors
who Made New
Poster Guides**

**Meet the Mentor
who Made the New
Inclusive Language
Guide**

**Where are they
now? Jackson
Filosa – Past
DataJam Mentor**

**DataJam Team
Updates**

DataJam Posters Due Friday, March 29, 2024



DataJam posters are due Friday, March 29, 2024 from all teams participating in the 2024 DataJam! See the [2024 DataJam Guidebook](#) (page 9) for instructions for how to make the poster. The criteria that will be used to judge the poster are in the Guidebook on pages 10-11.

We have resources on the **Resources** page of the website (thedatajam.org) to guide you in how to make an effective poster. This resource "Making Your DATAJAM POSTER", pictured to the left, was developed by two DataJam mentors, Angelina Jia and Mahitha Ramachandran, who are also featured in this month's Meet the Mentor section of this newsletter!

DataJam Team Visits Merck!

THANK YOU, MERCK! Nothing gets people more excited about a career than meeting outstanding people in the field. Thanks to the invaluable contributions of two key figures in Merck's IT landscape, Terry Rice, VP of IT Risk Management & CISO, and Harjeet Viridi, Director of Merck Animal Health R&D IT. The DataJam's Passaic Academy teams, under the leadership of Richard Chomko and Brandon Linde, alongside DataJam Board of Directors member Beth Bauer, were privileged to witness the dynamic interplay of data and science that are currently propelling the frontiers of biotechnology and cybersecurity at Merck's state-of-the-art Rahway facility.

Our high school teams from New Jersey embarked on a transformative on-site visit to Merck, providing a unique opportunity to engage with real-life data scientists and professionals immersed in the realms of cyber-security and animal health R&D IT. The visit afforded an exclusive (albeit shielded) glimpse into the behind-the-scenes inner workings of threat defense operations at Merck's Cyber Fusion Center, led by Mr. Rice.

Surrounded by cutting-edge technology and guided by brilliant, empathetic leaders, The DataJam's Passaic Academy teams from Passaic, NJ, embraced the realm of possibilities offered by data science during their visit to Merck's headquarters in Rahway, NJ, this past Fall. The experience was not just a journey into Merck's groundbreaking innovations, but a firsthand encounter with the company's century-long commitment to patients, including the "discovery and development of novel medicines and vaccines to fight infectious diseases."

We also had the privilege of interacting with employees, some who shared their experiences, and some who expressed genuine interest in understanding how their own students could found a DataJam team.

We extend our heartfelt gratitude to Terry Rice, Harjeet Viridi, Lorraine Woolever, Jacob Ponulak, Nancy Tamsula, Pippin Carino, Anand Patel, and Lauren Behette for their concerted efforts in ensuring that our students' day was not only informative and engaging but also left a lasting impact. It was a valuable opportunity for everyone to gain insights into the intricacies of working in data and technology, topped off with practical advice to navigate these career paths.

Site visits play a crucial role in raising awareness about the diverse roles in corporations and across industries. Awareness then sparks inspiration in young minds, as they envision themselves within these cutting-



edge facilities and exploring potential opportunities. This is how we fuel the talent pipeline!

Once again, we express our sincere gratitude to Merck for hosting The DataJam's Passaic Academy teams and hope for the privilege of a return visit in the not-too-distant future. Thank you, Merck!

DataJam Team Presents DataJam Project at National Meeting

What do high school students want from their (G)IEP/504 programs? This is the question the Central Dauphin DataJam team lead by Lauryn Chotiner, Connor Grauel, Ethan Erb, Caitlin Wilson & Noah Townsen researched using rigorous, professional standards (exact procedures, random selection, D.O.E., blocking, continuous improvement loops, etc.) while working within a larger research framework with Dr. Sabnis of Miami University. In addition, the team used Six Sigma methodologies to present their research and continuous improvement opportunities to the school district.

On February 15th, upon invitation by NASP (National Association of School Psychologists), Lauryn Chotiner traveled from her current studies, where she is a college student at Georgetown University, to New Orleans with Drs. Sabnis and Whitford of Miami University in Ohio to present the research to a substantial crowd. While there, an Ohio journal began inquiries for a future article about the team's work on this DataJam 2023 project.

Success occurs when preparation and opportunity cross paths. Thanks to the preparation and open-minded support of Central Dauphin School District near Harrisburg, PA and the opportunities afforded by the DataJam, as well as research partners like Dr. Sabnis, our students can conduct and apply real, meaningful research while their brilliant minds are at optimal creativity at the beginning of their professional journeys. At Central Dauphin, we want to best prepare every student.



Lauren Chotiner & Dr. Sabnis by poster at the NASP meeting.

Bob Moreland
Teacher

Linear Algebra, CHS Statistics, Geometry & Six Sigma

Meet the Mentors who Made New Poster Guides

Hi everyone! My name is Angelina Jia. I'm from West Chester, PA, and I am a first-year student majoring in Statistics and Economics at the University of Pittsburgh. In high school, I took AP Statistics. My teacher would assign projects in which we were required to brainstorm an idea, form a hypothesis, collect data, and present the conclusions in a slideshow—very similar to the DataJam. My interest in data science only increased from there. It is a skill that can be applied to every field imaginable, so there are numerous opportunities for people to combine their interests through statistics. Therefore, upon arriving in college, I found that the DataJam course was right up my alley. Not only could I refresh my skills with excel and data cleaning, but I also am now able to encourage younger students into this growing field as well. Through my statistics classes at Pitt, I've gotten to get a feel for R. I aspire to make myself as much help to DataJam teams as possible!



Currently, I am mentoring two teams. Both teams are composed of incredibly bright students who will ask me questions even I am surprised by. I'm grateful to be learning from them. Being able to use statistics as a vessel to learn about numerous topics is very exciting for me, so I am truly enjoying mentoring others in their journeys. I hope to become a data analyst one day and teaching others is an incredible first step.

Outside of school, I love writing, drawing, listening to music, and crocheting. I also love visiting the local museums and gardens here in Pittsburgh. Many of my interests pertain to the arts, which I am always very enthusiastic to talk about! Fashion is another passion of mine. Expressing myself through clothing is one of my favorite things to do.

Last semester, I took part in creating the "Making Your DataJam Poster" guide that is currently on the DataJam website. I tried my best to summarize all the important details I learned from taking part in two DataJam projects into a digestible and easy to read poster for those participating in the competition. As a recent high school graduate, I understand that the workload students receive makes it difficult for them to have time to read through very long guides. I'm hoping that this version will be able to assist many to come in creating some beautiful posters!



Hello! My name is Mahitha Ramachandran. I am a first-year student at the University of Pittsburgh studying Physics & Astronomy, and I am a DataJam mentor. I am so excited to currently be mentoring two DataJam teams as they learn to ask data-centric questions, analyze datasets, practice data ethics, accurately and effectively communicate their results, and discover how applicable the tools of statistics are to a wide variety of fields and topics. Both teams are investigating important, community-oriented questions, and we all learn from each other through this process. So far, it has been an amazing experience; I am looking forward to seeing their final project posters come together soon.

In our world of rapid scientific and technological progress, it is increasingly critical that science education, especially regarding technology and data, is equitable, accessible, and effective. I believe the world would be a better place if everyone had a basic understanding of data ethics and how data can be used (and misused) effectively. It is also incredibly crucial that we have people of diverse backgrounds and identities in data science careers. Thus, programs like DataJam are so important, and I am so honored to be a part of the DataJam mission.

Other than my passions for data science and science education, and my obvious interest in astrophysics, I love music, art, and writing. I play the piano and the cello, and I enjoy many different art forms such as ceramics, poetry, and embroidery. I also love reading, exploring the city, and am passionate about environmental justice.

Last semester, as part of the DataJam mentor course, I worked with fellow mentor Angelina Jia to create a "Making Your DataJam Poster" guide. There are a bunch of great resources and guides on the DataJam website, but we noticed that there was no succinct, approachable guide on making a project poster, which every DataJam team must do. After doing two of our own mock DataJam projects, we felt that we had advice to give in this area and sought to make the at-first daunting task of creating a project poster a little easier for high school students. Our guide is now available on the DataJam website – we hope it is helpful!

Meet the Mentor who Made the New Inclusive Language Guide

Hi, I am Andrew Lindros, and I am a freshman at Pitt and currently undeclared. Outside of school, I enjoy sports, music, reading, and spending time with friends. I am currently on the Club Wrestling team at Pitt, and I love attending many of the sporting events for the school teams. This love of sports inspired my interest in statistics as they are such a fundamental part of all professional athletics. This year, for DataJam teams, I've created the "DataJam Language Guide" and "DataJam Inclusive Language Guide" which I believe can be very helpful. The "DataJam Language Guide" is focused on providing writing and grammar tips for students to use on their presentations/posters. This is important because knowing how to utilize language to properly display your Data is a huge way to set your DataJam project apart. These tips are focused on avoiding mistakes, refining the words used, and clarifying the point that is being made. In the "DataJam Inclusive Language Guide," I pivot into focusing on how students should use the correct and considerate language when working on their DataJam projects. I feel this matters because I don't want anyone to feel left out, mistreated, or improperly represented in the data analysis projects.



Overall, I wanted to create something entirely different for DataJam, as most of the other guides are rightfully focused on the numbers part of data and I wanted to showcase the underrated language component of DataJam. I am very excited for the rest of the DataJam season and want to see all the teams do their absolute best on their projects.



DATAJAM INCLUSIVE LANGUAGE GUIDE

BY: ANDREW LINDROS



Where are they Now? Jackson Filosa – Past DataJam Mentor

With this issue of the DataJam Newsletter we are adding a new regular section to follow up with past students who have participated in the DataJam while in high school and with past mentors who have mentored DataJam teams. If you are using data science in what you do NOW, and you participated in the DataJam in the past --- we WANT TO HEAR FROM YOU! Email us at datajam@thedatajam.org and let us know you would like to write a "Where are they now?" column for the DataJam newsletter! Introducing, Jackson Filosa!

Hello everyone! My name is Jackson Filosa, and I am a Data Science Associate at PricewaterhouseCoopers (PwC). I graduated from the University of Pittsburgh in 2023, where I double majored in statistics and English writing- I love solving



problems with data and communicating the results in a meaningful way. During my time at Pitt, I was a DataJam mentor for three years, and was fortunate to work on several projects developing resources for the DataJam. After graduating, I now work full time as a data scientist!

In my current role, I build digital products that make work easier for the thousands of accountants and consultants working at PwC. For example, if there's a business form that takes 10 hours to manually fill out, my team will build an application that can do it in five minutes!

While the problems that I face and their context may have changed, my day-to-day work is still very similar to that of a DataJam project: identifying a topic to focus on, collecting high quality data, analyzing the data for trends, building models and testing them, and (often most importantly) communicating the impact and solutions from the work.

Just like on a DataJam project, I also work as part of a larger team: from software engineers to user experience designers to project managers, everyone has their own unique role in building a successful product.

Throughout all the statistics and data science projects I have encountered through college and my early career, some of the most creative and impactful ideas have always come from DataJam teams. I am so grateful for everything I have learned from this incredible organization and from you all! Best of luck with your projects this spring!

DataJam Team Updates

North Allegheny Team B

Students: Ashley Jang, Megumi Kawabe, Samuel Xiao, Aditi Motha

Mentor: Angelina Jia, Pitt

Our Data Jam project attempts to analyze the impacts of demographic variables on cancer incidence rates. We are starting with small scale models of single cancer types and income trends over shorter periods of time to establish a reliable analysis procedure and look for overall trends. We are hoping to eventually analyze multiple demographic variables, including race and air pollution, for all Pennsylvania counties over the past decade. We have started sifting through datasets and gathering information that we would like to analyze. Within the next few weeks, we plan to run our first regression on the correlation between breast cancer and incomes within each of Pennsylvania's 67 counties. We look forward to presenting our full results in April!

Avonworth Team 1

Students: Jackson Shieldi, Amelia Hardiman, Brayden Simmons, Madison Hollywood, Jagger Boyd, Sam Cavanaugh

Mentor: Sarah Sirakos, Pitt

We are one of the Data Jam Teams at Avonworth High School. We have three seniors, Amelia, Brayden, and Jackson, one junior, Madison, and two sophomores, Jagger and Samuel. We are investigating the correlation between political party alignment and the number of restaurants and/or bars within a given county per capita. We're finding the density of restaurants per capita to reduce the effect of the confounding variable of population, which we assume affects political alignment. One of our biggest concerns so far has been finding data in a format that we can utilize. Recently, we have discovered some promising sets that we will upload to the computer program R to sort and run tests on. Additionally, we have reached out to the Pennsylvania Liquor Control Board to receive their data in a usable format. We are looking into places that sell liquor to seek clearer relationships between political alignment and types of food establishments. This is to mitigate the risk of confounding variables when using a larger, generalized dataset such as restaurants. Additionally, we are considering stratifying the restaurants, maybe looking at fast food restaurants or just McDonalds. We look forward to seeing what route our research will take!

Keystone Oaks Team 2

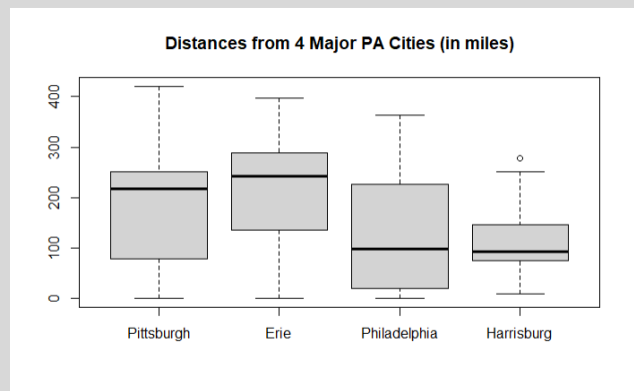
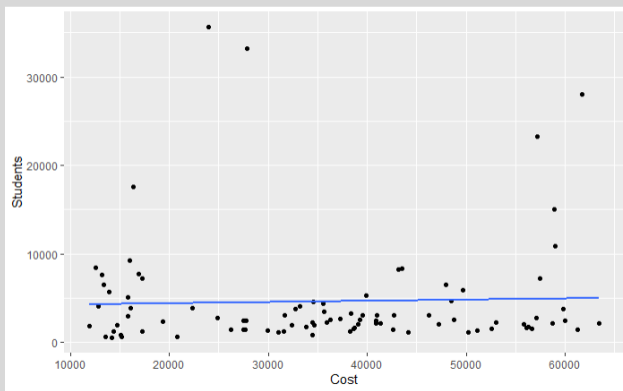
Students: Lia Scott, Alisha Thapa, Madison Williams-Powell, Lael Nowlin, Sierra D’Eramo, Katie Cesario, Anthony Cerminara

Mentor: Angelina Jia, Pitt

There is no shortage of colleges and universities in the state of Pennsylvania. Here, there are some of the best schools for medicine in the form of Pitt or computer science at CMU or any other field of study in any of the 90+ 4-year colleges/universities campuses in Pennsylvania. With so many options to choose from, it is hard to ignore the hefty price tag that comes with choosing where to go to school. Our Data Jam team from Keystone Oaks High School wanted to ask the question: What are the factors that impact college tuition?

We sourced our data manually via publicly available data released by colleges and gathered over 90 viable (meaning they had their information public) colleges. From here, we asked ourselves what we thought would be important factors of tuition. We landed on the following: whether it’s a public or private school, what the top major is, the school’s acceptance rate, its ACT and SAT requirements, the number of students enrolled, its location adjacent to 4 major Pennsylvania cities (Harrisburg, Erie, Pittsburgh, and Philadelphia), and its percentage of out of state students. After finding these specific data points we began the coding process through R.

While continuing to work on our coding process, we have successfully generated a sensible number of valid graphs from our code so far. The two we found most compelling were the following: A correlation graph between tuition and student population, and box plots for distance of colleges from major cities in Pennsylvania.

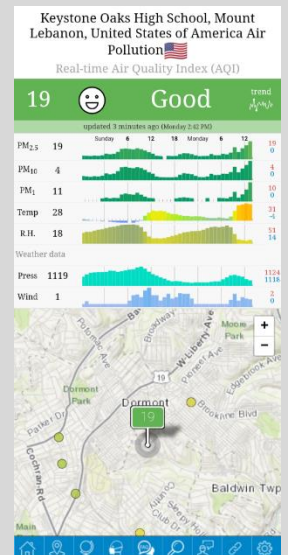
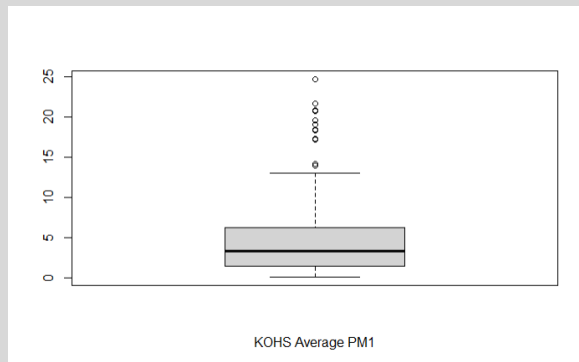


Keystone Oaks Team 3

Students: Alison Leung, Salaha Suleyman, Madeline McDine

Mentor: Bhaskar Chakraborti, Pitt

Our Data Jam project this year is to know what variables (relative humidity, temperature, elevation) have the most effect on pollution. As for now, we are using RStudio to conduct an exploratory data analysis, creating our graphs, and computing summary statistics. Our next steps are to complete EDA and begin creating a multivariate linear regression model.



San Lorenzo DataJam Team

Students: Chaeyi Lee, Mikayla Casey, Hudson McKinney, Kai Wildberger, Jaice Williamson

Mentor: Kartik Patel, SFSU

The question we are asking is “Is there is a significant relationship between wildfire risk and landside risk in California?” We have obtained the original risk value of landslide and original risk value of wildfire of the counties in California that have a wide range of elevations (over 7,000 ft) and the counties that have a narrow range of elevations (under 7,000 ft). We thought that the elevation of each county may be a confounding variable, so we decided to consider the range of elevation of each county. We divided the expected impacted area by the total square mileage, ending up with a proportion of each county expected for wildfires and landslides. We are using Google Colab notebooks for our analyses.

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.

- **Posters will be due Fri., March 29, 2024**
- **2024 DataJam Finale will be Thur., April 25, 2024**

January to Early April

Work on DataJam Projects

Teams will be able to work on their DataJam projects. Send an email to DataJam@thedatajam.org to arrange meetings with DataJam mentors, who are available to help with all aspects of the DataJam projects. Mentors can also be reached directly on the DataJam 2024 Slack workspace.

Friday, March 29, 2024

DataJam Posters Due

Teams should email their DataJam poster to DataJam@thedatajam.org. Instructions for the poster are in the DataJam Guidebook. Posters should be 24"x36" in size and submitted as a PDF.

Teams should email their DataJam poster to DataJam@thedatajam.org. Instructions for the poster are in the DataJam Guidebook. Posters should be 24"x36" in size and submitted as a PDF.



DataJam Guidebook - 2024

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!