

The DataJam Download

Official Newsletter of the Pittsburgh DataWorks



A new resource for considering diversity in your DataJam project

What's new on the Pittsburgh DataWorks website?

Meet the Mentors

Meet the Data Science

Professional

DataJam Timeline for 2023

A New Resource for Considering Diversity in Your DataJam Project



Many DataJam projects are asking research questions about their community, and it is important in all such projects to make sure that you have data available from all members of the community, considering diversity in socioeconomic status, race, ethnicity, language, disabilities and gender. If a portion of the community is left out in your dataset, your project will not be telling you the whole picture about the research question you are asking. This new guide on “Data Science & Diversity”, developed by Srija Konduru, provides helpful pointers on how to make sure you are considering your whole community in the research questions you are asking.

What's New on the Pittsburgh DataWorks Website

DataJam mentors have been hard at work creating new resources for DataJam teams. We have added these new resources to our website (pghdataworks.org), and reformatted parts of the website to make it easier to find resources. New resources and changes are outlined here. **Check them out!**

HOME PAGE

Now at the top of the **HOME** page, see the reformatted video *“A Walkthrough of the Pittsburgh DataWorks Website”*. This is a quick 7-minute tour of the website that provides an easy way to quickly find out what resources we have and where on the website you can locate them.

Near the bottom of the **HOME** page, you can find the latest issue of the *DataJam Newsletter “The DataJam Download”*. The newsletter is published monthly and is a great place to learn about new DataJam resources, upcoming deadlines for DataJam projects, and to learn more about DataJam mentors and Data Science professionals who collaborate to with Pittsburgh DataWorks. If you want to receive an emailed copy of the monthly newsletter sign up and provide us with your email address.

At the bottom of the **HOME** page is the *DataWorks Archive*. The Archive has copies of all past DataJam newsletters, including other news articles about the DataJam.

ABOUT PAGE

Check out the logos of *“Participating High Schools & After School Programs”*. We are constantly adding new team’s logos to this list!

DATAJAM PAGE

We have moved the location of the DataJam Guide Book on the website to near the top of the **DATAJAM** page. Every team should download a copy of the Guide Book. The Guide Book has guidelines for writing a DataJam Project Proposal, a Proposal Template, a Parent Permission Slip for teams that want to join the DataJam 2023 online SLACK workspace --- a great place for teams to work online together on their DataJam 2023 project and to get online assistance from DataJam mentors, guidelines for putting together a DataJam Poster, and the Rubric by which DataJam projects are judged.

On the **DATAJAM** page in the *“DataJam Timeline”* please note that the deadline for submitting DataJam proposals for the 2023 DataJam has been extended to January 31, 2023.

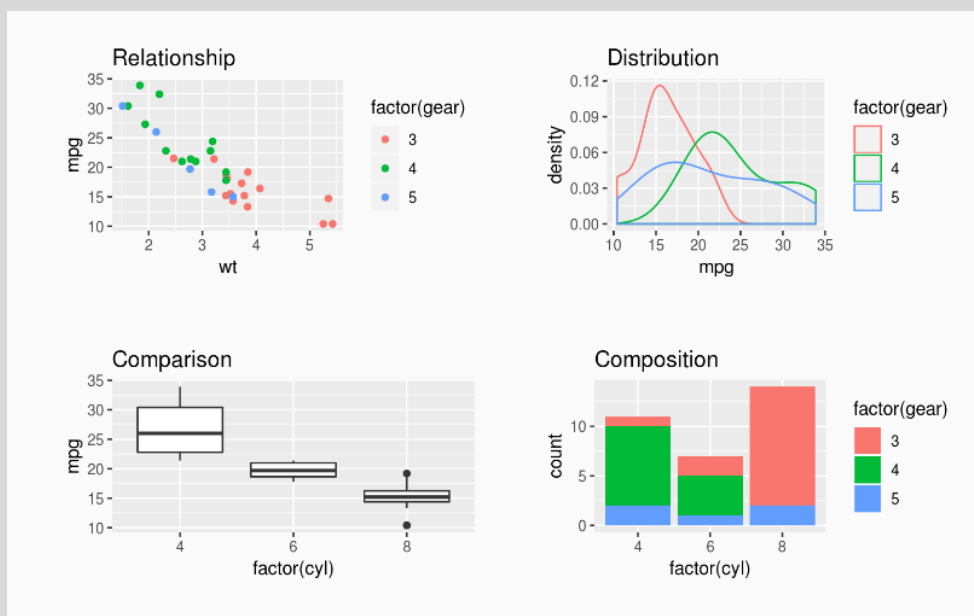
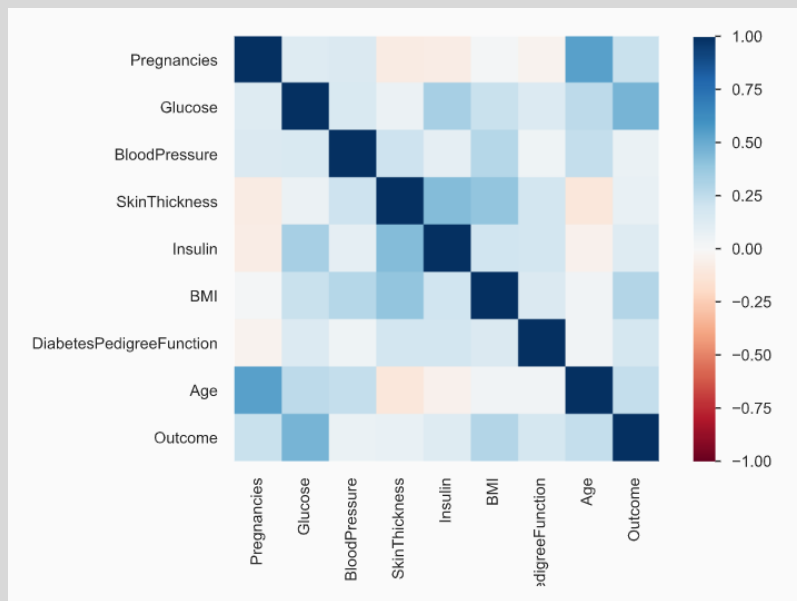
RESOURCES PAGE

A new section has been added to the top of the **RESOURCES** page: *“Game-Changing DataJam Guides”*. DataJam mentors have made a variety of guides to help DataJam teams with each step of the DataJam process. New guides that have just been added in this section include *“Data Science & Diversity”*, *“Public Data For Public Health”*, and *“Ensuring Research Data Quality Guide”*.



Looking for freely available datasets to use in your DataJam project? We have several datasets listed in the **“Show me the data!”** section of the **RESOURCES** page. We would like to draw your attention to the **“Dataset Guides”** in this section. They not only have links to freely available data, but also provide guidance on how to ask research questions on various topics.

Check out two new sections on the **RESOURCES** page: **“Push Forward with Python”** provides instructional tips on various data analytics and visualization tools used in Python. This section also contains a **NEW** guide on using SCIKIT-LEARN for predictive data analysis. You’ll definitely want to check this out if your DataJam project is attempting to predict something from the patterns found in currently available datasets. **“Generating Graphics with GGPlot”** is an open-source data visualization package that offers a lot of flexibility on how you make visualizations of the data you are working with. To help you learn more about GGPlot, DataJam mentors have developed a **“Data Visualization with GGPlot Guide”**! Some examples of what you can make in Python and GGPlot are shown below.



MENTORS PAGE

Learn more about the DataJam mentor you are working with on the Mentors page.

Meet the Mentors

Srija Konduru



Hi everyone! My name is Srija Konduru and I am a senior at the University of Pittsburgh majoring in Microbiology and minoring in Spanish and Chemistry. It is my first year as a DataJam mentor and I am really excited to see all your projects soon! I took the mentoring course at Pitt in the fall of 2022 as an introduction to the DataJam and the basics of data science.

And while I can say that I have an appreciation for math now, I can guarantee that my least favorite subject in grade school was math. Ever since elementary school, I can remember struggling with addition to multiplication, and then even more with the upper-level math classes as I got to high school. But my relationship with math finally turned positive when I took my first statistics class in high school. As opposed to the seemingly random string of problems in all my previous classes, these numbers had meanings: probability, prediction, and values. It was not until I started seeing examples of the applications of data science and data analytics in the real world that I saw how useful and cool (!) math really was.

As I got more involved in scientific research at Pitt, I learned how valuable data analysis is. I worked with a colonial invertebrate animal model in the surgery department at UPMC. We researched the ability for this animal to recognize self vs non-self cells through applied molecular biology and biostatistics. One of the coolest projects I worked on was using a machine learning (ML) program to generate and validate gene models from strings of DNA (the ATGCs that we sequenced) which aided in making the full genome of the animal model.

The data analysis step is the most important and difficult part of research--being able to critically analyze your data and present it in a meaningful way. However, I found the challenge to be the most fun part of research because it forces you to think more creatively and find more interesting problems to either further research or develop solutions for. This is why I got involved in data science and research; these interconnected fields encourage you to keep asking questions. And through mentoring, I found how encouraging students to get interested in learning (and asking questions) is so powerful.

DataJam has been such an amazing experience in connecting me with peers passionate about their fields, opening the vast world of data science, and allowing me to help get students interested in research. I am so excited to share my views on health science, diversity, and data science with a more open perspective from the DataJam as I enter medical school next fall!

Ajeet Subramanian



Hi guys, my name is Ajeet Subramanian, and I am a Junior at the University of Pittsburgh. Currently I am majoring in applied mathematics and minoring in applied statistics. I have a strong interest in data science and using data to solve community-based problems. In my spare time I like to read books and play video games.

I was introduced to the DataJam through the class Stat 1050: "Data Jam: Using Big Data for Community Good." Working with Dr. Cameron and other DataJam mentors in class really gave me a greater understanding of the power of data and how it can be used to help grow communities. Data can be incredibly intimidating, especially when working with it at first. The large amounts of information you work with can be hard to analyze, but with the right learning environment it can become a very fulfilling and doable process. This is, in my opinion, what makes a competition like the DataJam so great. For many high schoolers, something like data analysis seems difficult, but through this competition and the help of us mentors, they can be exposed to it and be a lot more comfortable with the field. Having an opportunity like this for high schoolers is incredible and gives them that experience they need to help them get a better idea of their goals and career prospects when going into college.

Being a mentor for the DataJam has been nothing short of incredible. Introducing high school teams to the world of big data has not only been fulfilling, but it has also helped me improve my knowledge in the field of data analysis. Seeing the creative and innovative ideas that teams have come up with in the past year has been great, which coupled with their passion shows how the data jam can help many students as well as communities.

As I continue to work with DataJam teams, I hope to learn more about the field while also being able to introduce teams to more of the intricacies that are involved with data. Being brand new, the field of big data is growing rapidly, in what feels like new findings daily. Having high school students learn about it this early is a perfect opportunity to not only grow the field more, but also help more people use data to change the world for the better.

Meet the Data Science Professional

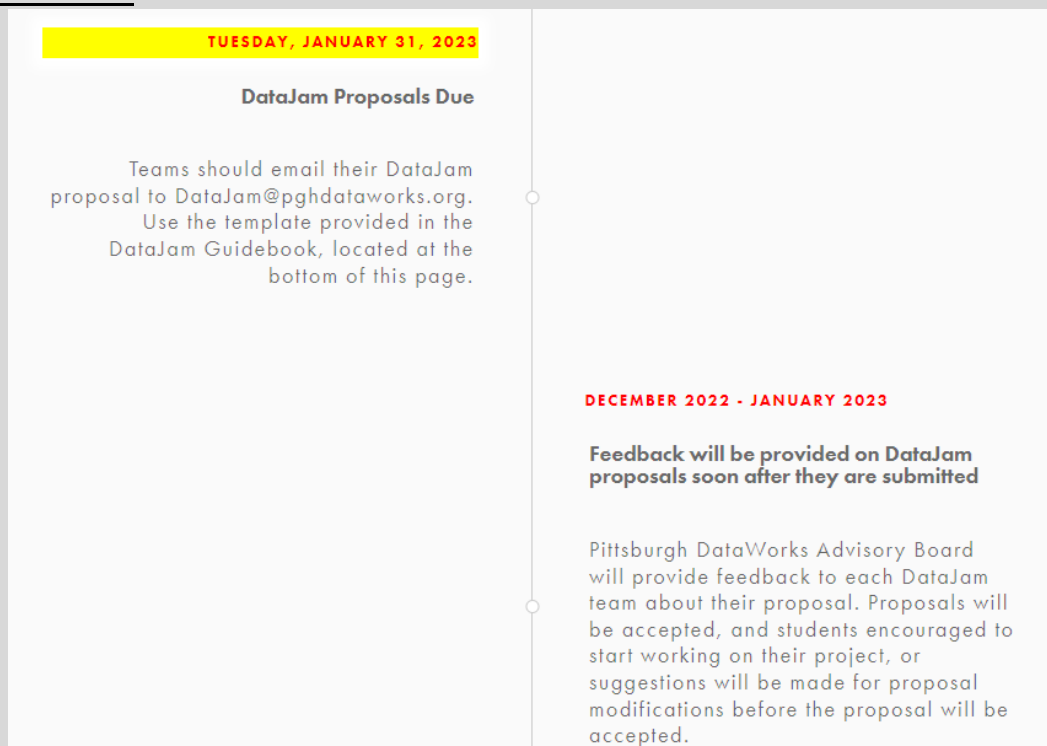
Happy New Year! My name is Sal Ferraro I am an assistant professor at Caldwell University in New Jersey. I started working with the DataJam last year, advising our college students to become mentors within the program. The DataJam allows high school students to understand how to collect, manipulate and analyze data to solve real-world issues. In addition, our college students are mentoring high school students to understand the importance of data and draw conclusions based on data. This allows college and high school students to work together to solve social issues.



For the spring of 2022, Caldwell University has added an Undergraduate degree in Business Analytics. This provides students with the fundamental concepts and tools needed to understand the emerging role of business analytics in organizations and shows students how to apply basic business analytics, preparing them for the workforce. One of the classes which students can take as an elective is the DataJam Mentor training class.

As an advisor to college students who often work as Interns with companies, I really see the importance of understanding data. The ability to talk with employers has become one of the essential skills companies are looking for when they interview students for internships. Businesses are looking for employees that know how to comprehend data. For companies to stay competitive, they must understand the industry's trends and realize the business's direction. Business Analytics is the solution. As technology continues to develop, more data will become available. The companies that leverage this data will be able to enhance sales.

DataJam Timeline



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!