The DataJam Download

Official Newsletter of the Pittsburgh DataWorks

On the Resource Page of the DataWorks Website:
See the New Guide on Data Visualization in R with GGPlot2 by Kevin Xu & Jeremy Tych

DATA VISUALIZATION IN R WITH GGPLOT2

A HIGHLY CUSTOMIZABLE ALTERNATIVE TO EXCEL

KEVIN XU AND JEREMY TYCH

GETTING STARTED WITH R STUDIO

Step 1: Head to this link to download R and RStudio onto your computer.

https://posit.co/download/rstudio-desktop/

Step 2: Open RStudio and go to File > New File > R Markdown to create a new Rmarkdown file. This is where you will write your code!

Step 3: To load the `tidyverse` library, which contains ggplot2, type into your Rmarkdown file:

```r
library(tidyverse)
```

Getting started with R Studio has never been easier! Generate summary statistics to get a glimpse of the data and what it is trying to tell you. Summary statistics can provide great insights that help you visually explore the dataset. Read the guide to learn more!

Pittsburgh DataWorks Joins the Pittsburgh Technology Council

In January 2023 Pittsburgh DataWorks joined the Pittsburgh Technology Council to become better connected with professionals and companies interested in data science. In February, two of the members of the DataWorks Advisory Board (Brian Macdonald with Oracle Co. & Devashish Saxena who has worked with Texas Instruments, Premier Farnell, Rexel and PPG) attended the *Beyond Big Data Learning Summit* in Pittsburgh representing DataWorks. This meeting focused on how data analytics, artificial intelligence, and machine learning are empowering businesses, solving tough challenges, and having the potential to make life easier and more productive for all of us.

2023 Beyond Big Data: AI/Machine Learning Summit

Presented by Pittsburgh Technology Council and CGI

The DataJam is Introduced at the New Jersey STEM Conference for K-12 Teachers

On February 10, 2023, Caldwell University hosted the 2023 New Jersey STEM conference to provide professional development for K-12 teachers. Three sessions at the conference featured speakers who talked about various aspects of the DataJam and encouraged more New Jersey teachers and schools to get involved in the DataJam. Rich Chomko, a teacher at Passaic Academy for Science and Engineering in Passaic, NJ, who has four teams participating in the 2023 DataJam, discussed how he has integrated the DataJam into his "Introduction to Data Science" course. Beth Bauer, CEO of PosiROI, a consulting firm that helps companies use strategic innovations in data, analytics, and tech, as well as a member of the DataWorks advisory board, talked about how companies can adopt a continuous innovation culture. Judy Cameron, Director of DataWorks, talked about what DataJam is and how schools can get involved in having DataJam teams, as well as all the resources available on the DataWorks website, pghdataworks.org.

NJ STEM CONFERENCE FOR K-12 TEACHERS

HOSTED BY CALDWELL UNIVERSITY

Friday, February 10, 2023
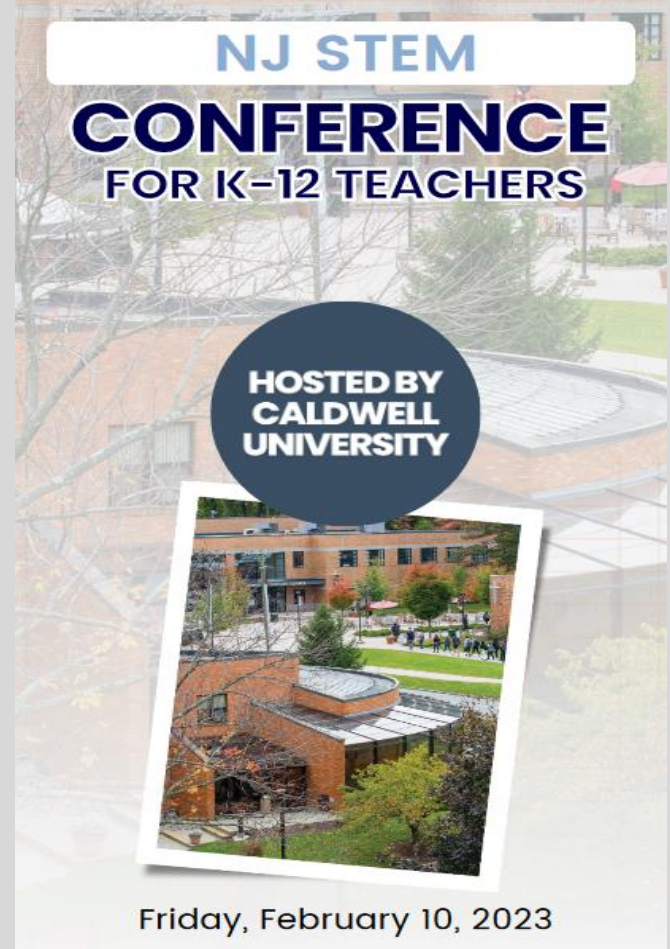
Meet the Mentors

Zhen Wu

Hi everyone! My name is Zhen Wu, and I just graduated with a major in Computer Science and a minor in Math from the University of Pittsburgh in December 2022. In my spare time, I love playing the piano and flute, reading mysteries, singing in the choir, and baking. This is my first time being a DataJam mentor, currently mentoring two teams. I have been excited at the creative and thoughtful project ideas developed by the students. I regard DataJam mentoring as a dual learning process: students learn to build up data-science projects, and I learn from their thoughtful ideas on improving their communities.

My motivation to study Computer Science and Math stemmed from the widespread data science applications, such as the personalized recommendations on Amazon and medical image analysis to predict tumors. The vast amount of user data makes me wonder: How could we improve precision and privacy as new data is emerging? I'm glad that my undergraduate studies have allowed me to develop and apply mathematical and artificial intelligence models to analyze diverse data, spanning from small-group discussions in middle-school science classes, the 3D shapes of robotics twisting under heat, and the interactions between adolescent girls and a social robot. I feel motivated to continue my data science endeavors in my future studies and career.

I was fortunate to get to know DataJam in my last semester at Pitt, by taking the DataJam mentor course. After reading the DataJam website, I was enthusiastic about it – it felt like something I would definitely have gone for in my high school, and teaching is my passion. In the course, I learned more about how big data could help improve our community and how to work with diverse communities. As big data increasingly impacts our lives, I want to introduce data analytic skills and tools to promote technological learning and interests among high schoolers and help them enter the data science world and build meaningful projects for their communities. To me, DataJam is more of an exploration and learning process. Students not only gain data analysis skills, but also think about ongoing issues in their communities, formulate research questions and present their project concisely, and investigate data bias and equity. I'm thrilled to be a small part of this journey.

I am very excited to continue working with the teams and see the progress they have made along the way. I hope the students enjoy joining DataJam and find data science interesting!

Anant Bajwa

Hi all! My name is Anant Bajwa, and I am a senior at the University of Pittsburgh majoring in applied mathematics with a minor in computer science. Outside of the classroom, I enjoy doing triathlons, hiking, and mountaineering! I like traveling to remote destinations, and one of the ways I like to de-stress is by planning my next trip. Before the start of the semester, I was hiking in the Arctic Circle in northern Sweden.

I did not know that DataJam existed until I took a course titled "Data Jam: Using Big Data for Community Good" with Professor Judy Cameron. During the course, we were taught how to effectively mentor students towards a data science related project. It is my second year as a mentor, and it is incredibly rewarding working with high school students, especially in the early stages of brainstorming project ideas, because you get a sense of their interests. I also am an undergraduate teaching assistant for pre-calculus, and algebra, and I find that teaching provides a similar sense of satisfaction. DataJam is incredibly useful to high schoolers since it provides them with the ability to work together in a group over the course of several months, just as is done in college or in the workforce. Teams are able to track their progress and break it down into bite-sized pieces. Every step is responded to by a mentor with helpful feedback to provide the students with a better idea of how to go about their project. High schoolers can learn more about the world of data science without the stress of trying to self-learn statistical techniques themselves. Because the teams are also mentored by college students, the students are provided with an environment in which they can freely ask questions without feeling stressed.

After getting involved with DataJam, I started diving deeper into the world of data science through my own self-research. In the fall, I am headed to the University of British Columbia where I will be using data science techniques to analyze communities most at risk for climate change in the future. Due to climate change, many of the world's glaciers are melting at a faster pace than normal. The freshwater from these glaciers is used by communities surrounding the glaciers, such as in Alberta. I hope to use machine learning to categorize the glacial fresh-water run-off. There are many interesting problems in the world and scientists are just now starting to use data science in their research to try to solve them. The field of data science shows up in almost every aspect of life. You can even work in the field of data science for Disney!

Meet the Data Science Professional

As the Division Director for the Research Data Services at the San Diego Supercomputer Center at UC San Diego, I oversee research IT for the center as well as a special unit, Data Initiatives. This group - made of project managers, informaticists, cloud engineers, science writers, designers, and computer scientists, works at the intersection of data science and research data management – largely for unserved or underserved communities. DataJam is one of our favorite activities and we are so proud to take part again this year. Youth from the Pala Band of Mission Indians formed a team led by Kim Mann Bruch. They were assisted by tribal elders who helped the team translate and present their project in their native Cupeño language. Their project on local water quality demonstrated the strengths of DataJam - selecting a topic related to their community, collecting data, finding existing data, and communicating results of analysis. They won Best New Team and presented at a national computing conference in Pittsburgh. This year we are involving high school students from both the Pala Band of Mission Indians and the Torrey Pines High School. Located less than 50 miles apart, the experiences and resources available to the students are quite different. This interaction not only allows the students to learn more about data science and utilize new skills to compare datasets, but it also provides an opportunity for them to interact with one another and to learn more about their respective cultures.

An additional aspect of this year's SDSC DataJam team effort involves a Research Experience for Undergraduate (REU) participant. Thanks to funding from National Science Foundation, a data science student from UC San Diego is serving as mentor for the Pala/Torrey students. Minchan Kim was featured in the February newsletter.

We are very grateful to the Pittsburgh Dataworks - especially Judy Cameron - and their efforts with DataJam. This program serves its original goals – to inspire students to use data to understand and advocate in their community, as well as to fill the pipeline of a diversified workforce in data. Beyond that, DataJam is impactful on all participants, from the team research leads to the undergraduate mentors and middle and high school mentees. Bonds formed during DataJam activities carry on past the projects and form a network where students see new avenues for their dreams. We are fortunate to have a team again this year focused on the local Western bluebird population. They look forward to participating in the Finale in April.

DataJam Timeline

FEBRUARY TO EARLY APRIL

Work on DataJam Projects

Teams will be able to work on their DataJam projects. Send an email to DataJam@pghdataworks.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 2023 Slack workspace.

FRIDAY, MARCH 31, 2023

DataJam Posters Due

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

DataJam Posters are due March 31, 2023

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!