DataFest for Two-Year Colleges

Rebecca Wong and Rachel Saidi

In 2018, a biostatistics student at Montgomery College asked if he could organize a team to take part in The George Washington University DataFest. He and his teammates from our community college hopped onto the metro each day to travel down to DC to work on the challenge.

In 2019, that same student asked for help to organize a new team. Again, he and his team took the Metro each day to work on that year's DataFest challenge.

In 2020, although the pandemic hit, MC had a relatively thriving data science certificate program with new data science enthusiasts. Students took part in GWU's virtual event that year and won an award for best use of outside data.

The American Mathematics Association of Two-Year Colleges has been working to expand access to data science programs for students at two-year colleges across the country. While students at MC have a data science program and an ASA DataFest event at a nearby four-year college, many two-year colleges are just starting to develop data science programs and do not have local four-year partners. In 2022, we co-chaired the AMATYC Statistics and Data Science Academic Network. The goal of ANet is to support the development and growth of data science programs at two-year colleges, so they decided to pilot a virtual DataFest for students from two-year colleges.

Broadly speaking, community college students tend to have lower incomes, may look for more flexible education options, may have more work and family obligations, and may have more challenges than the traditional four-year student. Although DataFest events take place throughout the world every year, DataFest for two-year colleges is the only event exclusively for community college and two-year college students.

The goal in creating a virtual DataFest was to increase access and level the playing field for this population of students while still providing them with the same level of challenge. According to *Community College Daily,* about four in 10 undergraduates in the US attend a two-year college, so including these students could provide a foundation for increased future engagement.

Prior to 2022, if community college students wanted to take part in DataFest, they could register their team to compete at any of the nearby four-year school hosts, but they would potentially be competing with students with a broader background in data science. The virtual modality of the event also provides access to students who may not have a DataFest at a nearby four-year college.

Rebecca Wong has taught mathematics and statistics to community college students for more than 30 years. She is an emeritus faculty member at West Valley College and adjunct professor at the University of San Francisco. As an active member of the American Mathematical Association of Two-Year Colleges, she has served as chair of the American Statistical Association/AMATYC Joint Committee and chair of the AMATYC Statistics and Data Science Academic Network.

Rachel Saidi is a professor in the math, statistics, and data science department and data science program director at Montgomery College. She is also the American Mathematics Association of Two-Year Colleges DataFest director and a member of the American Statistical Association/AMATYC Joint Committee. She is the recipient of the 2022 Montgomery College Outstanding Full-Time Faculty Award and 2023 NISOD Award for Excellence in Teaching.

Team Regression to the Meme from Skyline College. From left: Yuting Duan, Noel Amankrah-Bonsu, Travis Wellman, and Ekaterina Alekseenko.

Team The Outliers from Montgomery College. From left: Natalia Solomon, Lydia Baick, Mais Alraee, Grace Sampson, and Julia Melo Cavalcante.

In our first year, we advertised heavily to the AMATYC statistics and data science ANet community. At that time, five teams participated from around the country. The pilot was successful enough that the AMATYC executive board made Two-Year College DataFest an officially sanctioned AMATYC annual event and Pearson agreed to sponsor the event, covering the cost of plaques for the winning teams. Each year since, the number of students and schools participating in the virtual DataFest has increased. In 2025, 22 teams took part from multiple states, and we expect interest will continue to grow as more two-year colleges begin to develop data science programs.

It is important to hold this DataFest event specifically for two-year college students for many reasons but, most importantly, the goal is for these students to get access to this rich experience and work collaboratively with large data sets under intense time constraints. DataFest and similar hackathons can help bridge the STEM divide, encourage persistence, and provide opportunities for data-curious students to 'get their hands dirty' in what might be their first time working with such large and unwieldy data sets.

Some of the greatest challenges for our students include the following:

- Balancing participating in this intense weekend-long event with multiple responsibilities, including school, work, and family obligations
- Finding appropriate platforms teams can use to collaborate remotely

Some of the greatest challenges in hosting the virtual DataFest include the following:

- **Navigating deadlines across all the time zones.** How do we ensure a fair start and end time for all teams across the country? For now, we have employed the honor system. We post the material on Friday at 5 p.m. ET. Teams on the East Coast have an end time of Sunday at 5 p.m. Teams in other time zones may access the material at any time starting at 5 p.m. ET and use the honor system to end consistently at that same time on Sunday. We originally set a hard stop of 5 p.m. ET but some students on the West Coast were busy at 2 p.m., so we decided to be more flexible to make it fair for those students.

Montgomery College teams. From left: Alexandra Vereymechik, Hein Htet, Ash Ibasan, Emma Furth, Rebin Muhammad, Eyong Defong, Zhuwan Shwani, and Rachel Saidi.

Team The Insight Insiders from Skyline College. From left: Joyce Tsai and Nicholas Tai.

- **Publicizing the event to two-year students.** Our primary mode of publicizing the event has been through AMATYC channels and membership, but not all faculty teaching at two-year or community colleges are members of AMATYC, so we have also relied on word of mouth.
- **Ensuring all communications and materials reach all participants in a timely fashion.** We make all our materials available on a restricted-access Google drive. We provide access to the team faculty advisers as their point of contact and let them disseminate the materials to students. It is up to each team to decide how they will communicate with each other, work with their faculty adviser, and collaborate with outside resources.
- **Recruiting judges.** As faculty at two-year colleges, the event coordinators tend to have fewer collegial connections to tap for judging help than those who work at larger four-year colleges. Faculty at two-year colleges also do not have graduate students to help.
- **Consistency in judging many teams' submissions.** This year, we had 22 teams and could not ask our volunteer judges to review 22 submissions. We had 17 judges but are still working to develop a system that ensures consistency in judging while not burdening each judge with many entries.

Looking Ahead

The advantage of hosting a national virtual event is that participation is accessible to any student attending a two-year institution and volunteer judges come from across the country. Assuming the event continues to grow, we hope to streamline both the dissemination of materials and the judging system. The annual growth of the virtual DataFest indicates interest in data science programs at two-year colleges is increasing, and we hope the event will contribute to that growth. ■