FEATURING A SPECIAL SECTION ON DATAFEST

September 2025 • Issue #579

AMSTATNEWS

The Membership Magazine of the American Statistical Association • *https://magazine.amstat.org*

15 YEARS OF ASA DataFest

CELEBRATING THE CELEBRATION OF DATA

PLUS:

Recent Challenges Navigating the Employment Market

Meet New Member Lujun Zhang

JSM 2026

BOSTON, MA • AUGUST 1–6 • COMMUNITIES IN ACTION: ADVANCING SOCIETY

MAKE YOUR MARK ON THE PROGRAM

Bring your best ideas to the largest statistical event in North America—develop your session topic, invite dynamic speakers to take part, and send your online proposal today!

PARTICIPATE

July 16 – September 3, 2025
Invited Session Proposal Submission

September 30, 2025
Continuing Education Course Proposal Deadline

November 13 – December 10, 2025
Topic-Contributed Session Proposal Submission

December 2, 2025 – February 2, 2026
General Abstract Submission

January 15, 2026
Computer Technology Workshop Proposal Deadline

January 22 – April 2, 2026
Meeting and Event Request Submission

April 15, 2026
Late-Breaking Session Proposal Deadline

May 31, 2026
Draft Manuscript Deadline

ww2.amstat.org/meetings/JSM/2026/submissions

ATTEND

May 1 (11:00 am ET)
Registration and Housing Open

June 2
Early Registration Deadline

June 3 – 30
Regular Registration

July 1 – August 6
Late Registration

July 2
Housing Deadline

Joint Statistical Meetings

BOSTON, MA • AUGUST 1–6

ww2.amstat.org/meetings/JSM2026

AMSTATNEWS

SEPTEMBER 2025 • ISSUE #579

Amstat News welcomes news items and letters from readers on matters of interest to the association and the profession. Address correspondence to Managing Editor, *Amstat News,* American Statistical Association, 732 North Washington Street, Alexandria, VA 22314-1943 USA, or email *amstat@amstat.org*. Items must be received by the first day of the preceding month to ensure appearance in the next issue (for example, June 1 for the July issue). Material can be sent as a Microsoft Word document, PDF, or within an email. Articles will be edited for space. Accompanying artwork will be accepted in graphics file formats only (.jpg, etc.), minimum 300 dpi.

Amstat News (ISSN 0163-9617) is published eight times a year, February, March, April, June, August, September, November, and December, by the American Statistical Association, 732 North Washington Street, Alexandria, VA 22314-1943 USA. Business and Editorial Offices: 732 North Washington Street, Alexandria, VA 22314-1943. Accounting and Circulation Offices: 732 North Washington Street, Alexandria, VA 22314-1943. Call (888) 231-3473 to subscribe. **Periodicals postage is paid** at Alexandria, VA. POSTMASTER: Send address changes to *Amstat News,* 732 North Washington Street, Alexandria, VA 22314-1943 USA. Send Canadian address changes to APC, PO Box 503, RPO West Beaver Creek, Rich Hill, ON L4B 4R6. *Amstat News* is the member publication of the ASA. For annual membership rates, see *www.amstat.org/join* or contact ASA Member Services at (888) 231-3473.

American Statistical Association
732 North Washington Street
Alexandria, VA 22314–1943 USA
(703) 684–1221
ASA GENERAL: *asainfo@amstat.org*
ADDRESS CHANGES: *addresschange@amstat.org*
AMSTAT EDITORIAL: *amstat@amstat.org*
ADVERTISING: *advertise@amstat.org*
WEBSITE: *https://magazine.amstat.org*

The American Statistical Association is the world's largest community of statisticians. The ASA supports excellence in the development, application, and dissemination of statistical science through meetings, publications, membership services, education, accreditation, and advocacy. Our members serve in industry, government, and academia in more than 90 countries, advancing research and promoting sound statistical practice to inform public policy and improve human welfare.

features

Recent Challenges Navigating the Employment Market

by Jason Brinkley

Page 8

columns

This column is written for those interested in learning about the world of Data for Good, where statistical analysis is dedicated to good causes that benefit our lives, our communities, and our world. If you would like to know more or have ideas for articles, contact David Corliss at *davidjcorliss@peace-work.org.*

CAREER ADVICE FROM THOSE WHO'VE BEEN THERE

Looking for a career change? To mark the launch of the ASA's Career Connect website, the *Practical Significance* team asked **Sastry Pantula, Wendy Martinez,** and **Satrajit Roychoudhury to share the advice they wish they'd had early in their careers.** Visit STAT*tr@k* at *https://stattrak.amstat.org* to read the interview or scan this QR code to go directly to it.

STAT*tr@k*

MORE ONLINE

Read the following full-length articles at https://magazine.amstat.org:

Symposium Honors Legacy of Thomas R. Ten Have with Advances in Statistical Methods for Mental Health

The 12th annual Thomas R. Ten Have Symposium on Statistics in Mental Health brought together an international group of experts this June in Cleveland, Ohio. The event honored Ten Have's legacy while spotlighting innovative **statistical methods to advance mental health research and care**. Learn more about this year's presentations—from SMART designs to modeling childhood trauma.

Scott Vander Wiel of Los Alamos National Laboratory wins the 2025 Gerald J. Hahn Quality & Productivity Achievement Award

With a career spanning 35 years across industry and government, **Vander Wiel pioneered statistical methods that led to major advances in quality improvement, uncertainty quantification, and applied research**—affecting everything from industrial process control to national defense.

He'll deliver the Quality & Productivity plenary address at the **Fall Technical Conference** this October in Houston.

STATISTICIAN'S VIEW: A Call for Thought

In this month's *Statistician's View,* David Banks of Duke University challenges the statistical community to **take a hard look at its publication practices.** From outdated print journals to inefficiencies in peer review, he raises bold questions and calls on the ASA to lead the way in reimagining what **scholarly publishing** could be.

SUBMIT YOUR JSM 2026 TOPIC-CONTRIBUTED SESSION PROPOSAL TODAY

The Survey Research Methods Section is calling on members to submit topic-contributed session proposals for JSM 2026, especially those aligned with the "Communities in Action" theme. The deadline is December 10.

DID YOU KNOW?

ASA Executive Director Ron Wasserstein regularly shares news and updates on social media and the ASA's website. He also responds to **policy decisions** affecting the statistical community.

To read his posts, visit the Communications from the Executive Director page at *https://tinyurl.com/589h8535.*

15 YEARS OF ASA DataFest

CELEBRATING THE CELEBRATION OF DATA

Visit *Amstat News* online.

Reflections on JSM 2025: New Voices, New Energy

Ji-Hyun Lee

I prepared for the Joint Statistical Meetings this year in Nashville, Tennessee, with genuine enthusiasm and excitement to see longtime colleagues and meet new colleagues, but my anticipation was tempered by the sobering realities of the current moment. I knew many colleagues would not be able to participate. In this column, I want to share some reflections from JSM 2025, a gathering that acknowledged both the difficulties we face and our resilience.

During the ASA Board of Directors meeting, we were shaken by the sudden news of the Bureau of Labor Statistics commissioner's firing, which deeply concerned us as statisticians and citizens. It was a reminder that the professional and political landscapes we navigate are connected. Before JSM started, over two full days, the board tackled important decisions for our association—debating, planning, and setting a direction for the year ahead.

One of my favorite moments this year came from the First-Time Attendee Orientation and Reception. The room was buzzing with energy, curiosity, and optimism for the profession's future. After the event, I noticed something new: LinkedIn came alive! Attendees were sharing photos, reflections, and takeaways, creating a wave effect that extended far beyond Nashville.

This year, I had the joy of seeing four University of Florida Health Cancer Center colleagues and a student attend JSM for the first time. I was eager to hear what the experience was like for them, so I asked each to share a short reflection. Their words remind me of the power of community, the value of fresh perspectives, and why JSM continues to matter so much.

Tara Hashemian
Biostatistician, Biostatistics & Computational Biology Shared Resource, University of Florida Health Cancer Center

Attending my first JSM was an experience I will never forget, and I am truly grateful for the opportunity to participate. Being surrounded by thousands of people in the same field was both exciting and energizing, especially since in my everyday life I rarely meet others who have studied statistics.

I attended sessions spanning both soft skills and technical content, covering subjects such as understanding the impacts of the current administration's policies on the statistical profession, building stronger professional networks, enhancing statistical communication for diverse audiences, exploring advanced visualization techniques with ggplot extenders, and leveraging large-scale databases for complex analyses. This breadth of topics will not only strengthen my technical skills but also enhance my ability to collaborate effectively and adapt to evolving professional demands.

Gonghao Liu
Biostatistician, Biostatistics & Computational Biology Shared Resource, University of Florida Health Cancer Center

As a first-time attendee at JSM in Nashville, I began with the first-time orientation, where the friendly guidance and spontaneous conversations immediately made the conference feel accessible. Meeting people in that setting turned my initial nerves into excitement.

The ggplot extender session opened a clear path for contributing to ggplot2, showing how ideas move from a pull request to a polished extension. It demystified the process and left me motivated to join the community and contribute to the future.

I also attended the ASA President's Address and felt proud to see Dr. Lee serve as president this year. Between sessions, I explored Nashville, letting the city's energy and hospitality round out the experience. I left JSM with new connections, practical skills, and a renewed commitment to contribute to both the statistics

community and to the evolving AI–statistics conversation.

Zhongyue Zhang
Biostatistician, Biostatistics & Computational Biology Shared Resource, University of Florida Health Cancer Center

Attending JSM 2025 in Nashville as an early-career statistician was both exciting and humbling. As a first-time attendee, I felt a mix of curiosity, anticipation, and overwhelm. The conference's size and energy were unlike anything I had experienced before.

I began with the First-Time Attendee Orientation and Reception, which provided a welcoming start and helped me navigate the week ahead. From there, I made a deliberate choice to attend the professional skills development session Career Development Panel: Network Like a Pro. As someone who tends to be shy in large gatherings, walking into a room full of strangers was a challenge. The guided format helped break the ice. I quickly discovered that networking doesn't mean forcing a connection. By the end of the session, I had not only made new contacts but also gained a sense of accomplishment in stepping far outside my comfort zone.

One of the most memorable sessions for me was the Communication in Statistics and Data Science panel. Hearing experienced communicators share how they translate complex statistical concepts into messages that resonate with policymakers, journalists, and the public was eye-opening. Their stories reinforced that technical expertise alone is not enough; how we convey our findings can be just as critical in determining their real-world impact. I left the session inspired to strengthen my own communication skills and apply these lessons to my daily work.

For me, JSM was not just about learning from experts, but also about learning more about myself and how I want to grow in this profession.

Ada Wang
Biostatistician, Biostatistics & Computational Biology Shared Resource, University of Florida Health Cancer Center

As a first-time attendee, I was thrilled to be part of JSM, but I'll admit I was a bit overwhelmed by its scale. Fortunately, the orientation for first-time attendees created such a welcoming environment that it encouraged me to reach out to others and start building connections right away.

One of the biggest challenges was deciding which sessions to attend, and there were so many options! Over five days, the talks covered a broad spectrum, from theoretical statistical methods in academia to clinical applications in industry, from high-level policy discussions on the impact of AI to hands-on deep dives into R packages like ggplot2 extensions.

As mentioned in the president's address, these fields aren't silos; they're bridges that connect us all and guide us from the present into the future. Attending JSM has been a major milestone in my professional development, and I look forward to becoming more engaged in this vibrant community.

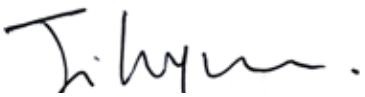
Zhuochao Huang
PhD Student, Department of Statistics, University of Florida

My first impression of JSM was simply its scale. The meeting was packed with a variety of sessions covering the cutting-edge developments in every corner of statistics.

I focused on sessions on causal inference. Hearing different viewpoints, especially on the application of causal inference to real-world problems and what directions are of genuine interest to practitioners broadened my understanding.

At JSM, I met and attended the lectures of prominent figures in statistics such as Tony Cai and Jianqing Fan and it felt just like meeting my idols, while also connecting with fellow PhD students. Some of the most valuable moments happened after sessions ended, when conversations turned into unexpected mentoring opportunities. Those exchanges, both professional and personal, were as enriching as the formal presentations themselves.

Despite the uncertainties that have marked our recent months, the enthusiasm shared by my colleagues who were first-time participants served as a powerful reminder of why the ASA community is important. Moving forward, we carry with us the current realities and renewed commitment that comes from new voices ready to contribute to our profession.



Member Feedback Sought on Proposed Changes to Voting, Fellows Eligibility

Ron Wasserstein, ASA Executive Director

At its August 2025 meeting, the ASA Board recommended two changes to the ASA's governance documents.

The board will vote on these proposed changes in November. Per the rules set forth by our constitution and bylaws, the board seeks comments from ASA members on the proposed changes. Please send comments to ASA Executive Director Ron Wasserstein at *ron@amstat.org* by November 15.

(1) Change to voting process

Constitution *(words in italics are the proposed addition)*

Article IX. METHOD OF SELECTION (first paragraph)

All individual members are eligible to vote for every elected position on the Board of Directors, *subject to additional eligibility requirements set forth in the ASA bylaws.*

Bylaws *(words in italics are the proposed addition)*

Article III. VOTING

1. Quorum. In any vote of the Association's membership, all ballots received within a period set by the Board of Directors will be counted and considered a quorum.

2. Balloting. For all of the Association's elections, the system known as approval voting will be used. Regardless of the number of candidates or the number of places to be filled, the voter may vote for any number of candidates but may not cast more than one vote for a candidate. Winning candidates are those with the highest numbers of votes. Any tie will be broken by random selection; no runoff elections will be held.

   Ballots must not identify the manner by which any candidate was nominated: by the Committee on Nominations; by a Council; or by petition. Names of candidates will appear on the ballot in random order.

   In case of ambiguity or lack of clarity in the election rules, the Executive Committee will determine the procedures.

3. *Eligibility. In any vote of the Association's membership, all individuals whose membership is active (that is, not lapsed as per Article I, Section 4) at the time an election begins are eligible to vote in that election.*

(2) Change to Fellows eligibility

(words struck out are the proposed deletion; words in red are the proposed addition) (Note: This change would take effect for the 2027 Fellows nominations and awards.)

Article I. MEMBERSHIP. Section 6.

6. Fellows. By the honorary title of Fellow, the Association recognizes individual members of established reputation who have made outstanding contributions in some aspect of statistical work. Fellows are selected by the Committee on Fellows.

The number of new Fellows to be selected each year will not exceed one-third of one percent of the individual members. Only a person who has been an individual member of the Association for ~~the prior three years~~ at least five of the last six years will be eligible for selection as a Fellow. In selecting Fellows, the Committee on Fellows will evaluate the impact of the candidate's contributions to the advancement of statistics or areas of application, giving due weight to publications, the position held by the candidate in the organization in which the individual is employed, activities in the Association, and other professional activities. The case for each candidate will be judged individually, with no one of these criteria governing selection to the exclusion of the others. ■

MY ASA STORY

Renée Hanson

MORE ONLINE
View Hanson's entire ASA story online at *https://magazine.amstat.org.*

Getting a great education has always been important to me. I enjoy learning and being abreast of exciting research in the social sciences and science fields. I began my educational career at the University of Maryland, College Park, and earned a bachelor's degree in African American studies. I went on to earn a master's degree in sociology from American University in Washington, DC.

I was always intrigued by the statistical side of my sociology degree. After my master's program, I continued to work; research; and publish in education, health, and history, but really wanted to know more about statistics.

In 2021, I completed an American Educational Research Association professional training seminar analyzing large-scale assessment data using R. Soon after, I earned verified data science certification in R-Basics programming and data visualization certification in R from Harvard X, a Harvard University online learning course.

I enjoyed these courses so much that I became interested in learning more about the American Statistical Association. I began to take part in several virtual workshops, webinars, and courses hosted by the ASA with other academic professional organizations. I participated in the ASA workshop Meeting Within a Meeting and earned a certificate. The meeting provided middle and high school mathematics and science teachers an opportunity to discuss and apply the data analysis, and statistical concepts embodied in the NCTM Catalyzing Change series and the ASA's *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report: A Pre-K–12 Curriculum Framework*.

Another opportunity I took part in was a virtual ASA Philadelphia Chapter traveling course: Data Visualization with R. I enjoyed learning practical skills for R programming. It helped improve the quality of my work with hands-on examples.

I have been in the education field for years. Most recently, I was the associate director of an enrichment program academy for high-achieving students in mathematics, language arts, and science but wanted to take my career a step further by transitioning into a health and statistics field. In recent years, I earned my joint PhD degree in urban systems, specializing in education/global urban studies, health, and environment from Rutgers University Graduate School, Rutgers University Biomedical and Health Sciences School of Nursing, and the New Jersey Institute of Technology.

I was still quite interested in gaining statistical accreditation because I am passionate about data science and statistics, especially in the areas of education, health, and environment. Although I took nontraditional statistics education routes, I was encouraged by my late and great Washington Statistical Society mentor Carol Blumberg to pursue the field by gaining more statistical ability and applying for the ASA statistics certification program. Because I have a sociological and quantitative background, years of professional work experience in the social sciences, and a PhD, I was elated to be accepted as an ASA Graduate Statistician. I am grateful to Carol for steering me in the right direction; I miss her kindness, expertise, and great sense of humor.

I have continued to gain experience in statistics and data science as a peer reviewer, taking part in professional development workshops, seminars, and courses.

I served as one of the judges for the 2024 Undergraduate Statistics Project Competition by the Consortium for the Advancement of Undergraduate Statistics Education and the ASA. I also served as a judge for the 2024 ASA Statistics Project Competition. I continue to be active with ASA volunteer academic services, statistical groups, and ongoing mentorship with a professor and statistician.

The ASA has certainly provided me with countless opportunities, and I want to become a lifelong learner and eventually an expert statistician. I believe ASA membership, accreditation, course participation, and volunteer academic services are excellent to further a person's pathway into a statistics career. I hope to eventually earn full ASA Professional Statistician accreditation. Furthermore, these experiences are teaching me to become more of a leader, team player, and hard worker toward a newfound career. ■

Staff Spotlight: Meet Chance Frye, Staff Accountant

Hello, everyone. My name is Chance Frye, and I work alongside Derek and Adrian in the accounting department at the ASA. I am a native of Northern Virginia and currently reside in my hometown of Ashburn.

Sports have always been a big part of my life, and I would consider myself a pretty big sports fan. I have played throughout my life and ended up having a successful high school football career before going on to accept a football scholarship at The University of Virginia's College at Wise, where I played for two seasons. Multiple injuries ended up derailing my playing career. I was forced to medically retire, and I ultimately transferred to James Madison University, where I finished my degree and graduated in 2019 as an economics major.

My first job out of school was for a consulting firm that worked strictly with nonprofits and associations, and it was something I really enjoyed. Over the last four years or so, I have held positions at for-profit organizations, but the missions, people, and environments within the association world drove me to look for an internal position and I ended up landing at the ASA, which I have thoroughly enjoyed so far!

I am still involved in football and serve as the offensive coordinator at Gainesville High School in Gainesville, Virginia. It is a relatively new school, and we are about to enter our fourth season as a varsity program. Outside of work and sports, I am a big fan of traveling, game shows, and spending time with friends and family. A fun fact about me is I won a primetime jackpot episode of *The Price Is Right!* Some of my prizes were a new Lexus and trips to Mexico, Finland, Singapore, and the Galapagos Islands. If you ever want to know the inside scoop on how one of those shows works, I'd be happy to share.

I am thrilled to be part of the ASA team and greatly enjoyed meeting so many of you at JSM this year! ■

New Member Spotlight: LUJUN ZHANG

This month, we welcome Lujun Zhang, who answered the following questions so we could get to know him better:

How did you become interested in statistics and/or data science?

I initially studied microbiome research, where I realized that cutting-edge discoveries in the field often rely on equally advanced analytical methods. This sparked my interest in statistics and data science as essential tools to drive biomedical innovation.

What do you consider your dream job?

My dream job is to work in a hospital or a pharmaceutical company, where I can contribute to developing real therapies that improve patients' lives.

What do you hope understanding statistics and/or data science helps you accomplish?

I'm especially excited about how recent advances in AI can accelerate drug discovery and optimize treatment strategies. I hope my skills in statistics and data science will allow me to be part of that transformation.

Is there a particular group of statisticians you would like to reach out to you (e.g., from a section, interest group, chapter, committee)?

I'm particularly interested in connecting with biopharmaceutical statisticians and related interest groups.

What is your favorite hobby?

I enjoy sci-fi stories and simulation games—especially those involving space colonization—as well as gardening.

What is something you would like people to know about you that we haven't asked?

I have a green thumb and grow many houseplants, which have been a great source of joy and hope—especially during Minnesota's long, rigid winters.

To view a list of all new members, visit *https://magazine.amstat.org.* If you are a new member interested in being featured, email ASA Communications Manager Megan Murphy at *megan@amstat.org.*

Recent Challenges Navigating the Employment Market

Jason Brinkley

Jason Brinkley is a biostatistician, data scientist, and health researcher who works out of the greater Raleigh, North Carolina, area. A North Carolina State University graduate, Brinkley has worked at East Carolina University, the American Institutes for Research, and Abt Global. He is currently associate professor of biostatistics at Northwell Health and the 2025 chair-elect for the ASA Health Policy Statistics Section. In his spare time, he does his best to not lose at Mario Kart World to his four children.

Major restructuring of the US federal system has led to widescale changes and reductions in force in many agencies. These changes created ripple effects that led to further reductions among federal contractors, technology providers, and adjacent services, including higher education. As a result, the current labor market faces several unique challenges, especially for those who work in the statistical and data sciences. Coupled with major changes in technology due to the rapid adoption of AI and the shifting in remote and hybrid work (dramatically increasing the candidate pools for each job), many are struggling to find or maintain continued employment.

I am no exception. These changes hit me personally as a federal contractor in April because of the reductions in force. I spent a little more than three months searching for my next opportunity. My search was broad and exhaustive, as I applied for positions in health care, technology, education, and professional services. I'm fortunate to have much going for me: a doctorate in statistics (a versatile degree); a large network of colleagues; and nearly two decades' experience doing a wide range of analyses and playing in so many backyards that it would make John Tukey jealous.

I applied for 182 jobs that resulted in 27 interviews (across first, second, and third levels) and two high-quality offers.

Reflecting on this experience, I found myself wanting to summarize some of what I learned, so over the course of 10 days, I wrote a short blog series for my LinkedIn page. This article summarizes the key insights from that series. Although the LinkedIn series touches on several points, I highlight three key themes here: the AI barrier; the importance of networking; and the help of recruiters. While these themes are far from inclusive of all the challenges, they are the ones I see as critical considerations. I highly suggest job seekers join online communities—such as LinkedIn—to discuss pros and cons of specific available tools.

The AI Barrier

Companies are seeing a deluge of high-quality talent applying for every open position, especially those that are remote-eligible. When there is a flood of applications for every opening, tools are needed to help separate those qualified from those who are not. A remote job posting for a data scientist or analyst can receive more than 5,000 applications from all over the world. AI helps employers cut through the static and provide a smaller cohort of applications to be reviewed by humans. Therefore, the job-seeker's goal is to make it into the top 100 applications to cross the AI barrier.

My advice is to expand your résumé as much as possible. Gone are the days of the simple one- or two-page document. Consider the first page of your résumé the main source for AI scanning and matching; focus page one on AI and subsequent pages on human review. Résumés should be tailored to job applications with language that is easy for those systems to absorb. You already understand the importance of having interoperable data; you should think of your résumé as a data point and plan accordingly. Are you finding that when you upload your résumé into a job portal it is always misattributing your education or professional experience? Consider changing your résumé format to make it easier for those systems to read. Professional résumé coaches can be a huge help here for folks with longer résumés.

Your Network Is Everything

The AI barrier creates a homogenous pool of candidates with the exact skills that match the job description, making it even harder for

> Do not wait until you need a network to grow a network.

manual review. Therefore, standing out in the smaller pool of candidates has become an even bigger challenge, which is why networking is more important than ever.

LinkedIn has become the premier platform for social networking among US professionals and was my go-to source for my recent job hunt. My network was instrumental in finding opportunities, gaining referrals so I could get past the AI barrier, and strategizing with others facing similar challenges. If you plan to use LinkedIn as your sole mechanism for job hunting, know it has algorithms that guide the content you see. I highly recommend investing in the premium tier to help you reach new individuals. Personal experience showed my content was different before and after that investment. Similarly, constantly explore your network, decide how much of it is isolated to your current environment, and connect with new groups to expand your opportunities. Do not wait until you need a network to grow a network.

Finally, do not just lean on your network to help find opportunities. Traditionally, folks used their network to help get their foot in the door, and I recommend job seekers expand the use of their network as they are able and feel comfortable. If a friend has helped you get seen by a particular company, ask them for personal insights about some of the people you will be speaking with. Ask peers and mentors—especially those familiar with the types of jobs you are applying to—to do mock interviews (or informational interviews) so you can sharpen your message.

Recruiters Are Your Friends

The rise of AI in employment systems has spurred a parallel rise in companies using recruiters to fill key positions. There are company-specific recruiters and recruiters who work independently or within employment firms. They can be found and networked with on LinkedIn and tend to have an inside track on open positions. When reaching out to a recruiter, be brief and target the opportunity you hope to discuss. Their job is to match people with jobs so make that part easy for them. Respect their time and don't be pesky when doing outreach and follow-up. Likewise, if a recruiter reaches out to you when you're not interested or looking, a simple reply of "I'm not looking at this time" is sufficient. They are reaching out to many others and not looking to have a detailed back-and-forth with each person, especially those who are not interested.

My Recent Job Market Challenges

Read Brinkley's 10-part series on LinkedIn:

Day 1: *https://tinyurl.com/3wn56y75*

Day 2: *https://tinyurl.com/434k5m73*

Day 3: *https://tinyurl.com/4557bd4s*

Day 4: *https://tinyurl.com/4v3spkmd*

Day 5: *https://tinyurl.com/4bmdcn9j*

Day 6: *https://tinyurl.com/h9bdf5st*

Day 7: *https://tinyurl.com/2496p4ay*

Day 8: *https://tinyurl.com/3hasb966*

Day 9: *https://tinyurl.com/3ec7sv55*

Day 10: *https://tinyurl.com/28s9k994*

Final Thoughts

Details, recommendations for specific tools, and stories (including a lot of humor about my misadventures and failures) can be found in my LinkedIn series mentioned in the sidebar. However, I want to underscore a few points here:

- No job is 100% safe. We have seen recently instances in which qualified individuals were cut indiscriminately. In response to my LinkedIn series, I received many responses and private messages that started off as, "I never thought this could happen to me" or "I was completely blindsided." Those of you dealing with this are not alone. Prepare for that potential by keeping your résumé up to date and your professional network strong. Also, give back to the community and recruiters as much as you can.
- The emotions and stress created from a sudden termination are a big deal, and working on maintaining your mental health is key.
- There is no one mechanism that guarantees success in the job hunt. Different people can use the same techniques and methods and be on the job market for three weeks, three months, or three years. Life is more luck than we would like it to be and sometimes success is a convergence of the right person with the right mindset at the right moment. ■

SEPTEMBER Time Series

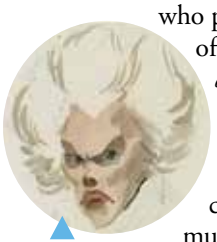
This Month in Statistics History

Penny S. Reynolds, University of Florida College of Medicine

SEPTEMBER BIRTHDAYS

1707 Georges-Louis Leclerc, Comte de Buffon. A naturalist and mathematician, he is of interest to statisticians for his "needle problem," an early example of the use of Monte Carlo simulation principles to estimate *p*.

1752 Adrien-Marie Legendre, who published the method of least squares (*méthode des moindres carrés*) in 1806 in the appendix to his book on comet motion. Gauss later claimed precedence, much to Legendre's annoyance. The caricature (pictured) by Julien-Léopold Boilly is the only known portrait of him.



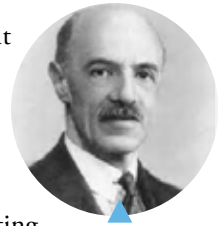
LEGENDRE

1759 William Playfair, a pioneer of statistical graphics and data visualization methods who invented numerous well known graph types, including the pie chart and time series plots. He always maintained that charts communicated better than tables of numbers.

1835 William Stanley Jevons, who pioneered the application of statistical techniques to economics. He was introduced to the use of statistics by the writings of Adolphe Quetelet. He famously distinguished a "mean" as the approximation of a definite existing quantity from the arithmetical average. He also invented an early logic machine.

1860 Clara Collet, one of the first women to be elected Fellow of the Royal Statistical Society (1892) and member of RSS council in 1918. As a civil service statistician, she performed groundbreaking studies of women and the working poor in the London East End during the time of the "Jack the Ripper" murders. P. C. Mahalanobis says she persuaded him to join RSS.

1863 Charles Edward Spearman, known for his eponymous rank correlation statistics and for developing factor analysis. Both Karl Pearson (who considered himself the godfather of correlation) and E. B. Wilson (ASA Fellow 1924, ASA President 1929) attacked his two-factor theory of human intelligence, resulting in a long feud.



SPEARMAN

1876 Edith Abbott, ASA Fellow 1945, was first woman to be an academic dean in the United States (Chicago), pioneered statistical methods for studying social problems, and developed evidence-based policy for crime, welfare, education, and protections for immigrants, working women, and children

1879 Leonard Porter Ayres, ASA Fellow 1917, 21st ASA President 1926. Secretary for the 1915 ASA Joint Committee on Standards for Graphic Presentation chaired by William Brinton. Organized Division of Statistics for the US Army during WWI to track combat casualties. Later involved in education and became head of statistics for the Playground Association of America.

1889 Besse Day (Mauss), ASA Fellow 1951, Intercollegiate Studies Institute Fellow, American Society for Quality Control Fellow, American Association for the Advancement of Science Fellow, pioneered design of experiments applied to engineering. Introduced to Ronald Fisher as a student, she collaborated with him for years. Her 1955 paper, "The Technique of Regression Analysis," won the 1956 American Society for Quality Brumbaugh Award for quality control.

1892 Frank Wilcoxon, ASA Fellow 1955, developed nonparametric rank-sum and signed-rank tests. He attributed his interest in applied inferential statistics through reading Fisher's *Statistical Methods for Research Workers.*

1893 Harold Cramér, ASA Fellow 1950, Guy Gold Medal 1972. "One of the giants of statistical theory," he is best known for developing the Cramér-Rao inequality, Cramér-von Mises statistics, and Cramér-Levey theorem.

1893 Hilda Geiringer-von Mises, Institute of Mathematical Statistics Fellow. A "missed genius," she was an innovative researcher in probability theory and Fourier series, but was denied a permanent faculty position at American research universities because of her gender. She collated and edited the papers of her husband Richard von Mises after his death.

HOTELLING

1895 Harold Hotelling, ASA Fellow 1937, mathematical statistician, and economist. He developed Hotelling's Law, Lemma, T^2 Statistic, and principal component analysis. He was first to recognize the revolutionary importance of Fisher's Statistical Methods for Research Workers.

1907 Maurice Kendall, ASA Fellow 1950. Royal Statistical Society Fellow 1934, RSS President 1966; Guy Silver (1945) and Gold (1968) Medals. Known for his tau rank correlation, he also made notable contributions to experimental design, factor analysis, *k*-statistics, and time series. With Bernard Babington-Smith, he developed one of the first early mechanical devices for generating random digits, and a series of tests for statistical randomness.

1915 Olive Jean Dunn, ASA Fellow 1968; AAAS Fellow 1965; University of California, Los Angeles Woman of Science 1974. Best known for the 1961 method for multiple testing correction for simultaneous confidence intervals that she named after Carlos Bonferroni.

1915 George A. Barnard, ASA Fellow 1961, Royal Statistical Society President 1977, Guy Gold (1975) and Silver (1958) medals. Although he is best known for work on likelihood methods, often overlooked were his contributions to statistical process control, especially development of sequential tests and optional stopping developed independently of Abraham Wald.

1915 Helen Abbey, ASA Fellow 1976, public health statistician named by Johns Hopkins as a "Hero of Public Health" in 1991 for distinguished teaching of statistics (often accompanied by her dog, Peppy) and prolific mentoring. She taught more than 4,000 students and read over 700 doctoral dissertations.

1920 Calyampudi Radhakrishna (C. R.) Rao, ASA Fellow 1972; President ISI, Institue of Mathematical Statistics, and International Biometric Society; National Medal of Science. Developer of the Cramér–Rao bound and Rao–Blackwell theorem, he has been called the "most eminent statistician of our time."

NATRELLA

1922 Mary Gibbons Natrella, ASA Fellow 1981, American Society for Testing and Materials Fellow, Department of Commerce Bronze Medal 1982. A National Bureau of Standards statistician, she developed the *Handbook of Experimental Statistics* (*NBS Handbook 91*), precursor of the NIST/SEMATECH *e-Handbook of Statistical Methods*. Although originally intended for the Army, it is still a landmark publication for statistically based principles for experimental planning and analysis.

1930 Colin Lingwood Mallows, ASA Fellow 1969, fellow of RSS and IMS and codeveloper with Cuthbert Daniel of the Cp statistic for regression model diagnostics. (They called it C after themselves, Cuthbert and Colin; p is the number of variables). He is also one of only two statisticians awarded the Statistics Trifecta: the R. A. Fisher Lectureship (now COPSS award) 1997, Deming Lectureship 2004, and Wilks Memorial Award 2007.

EVENTS IN SEPTEMBER

1853 On September 19, the first International Statistical Congress begins in Brussels. It was organized by Adolph Quetelet. Unfortunately, nothing much was accomplished because of quarrels over agenda and choice of language, and accusations of data misrepresentation. Bismarck pulled the plug on further congresses after 1878 by forbidding Prussian statisticians to participate.

1854 On September 8, John Snow has the Broad Street pump handle removed, pinpointing the source of the 1854 Soho cholera outbreak. He established that cholera was waterborne, not airborne as previously believed. Snow's intervention before establishing definitive cause—*Vibrio cholerae* was not isolated until 1883—was based on extensive ground-truth research, now the model for modern epidemiology.

1885 On September 10, Francis Galton introduces the concept of multivariate regression (the term he substituted for "reversion") to a largely mystified audience at the British Association for the Advancement of Science meeting.

1935 On September 12, the Institute of Mathematical Statistics is officially founded by Harry C. Carver and Henry L. Rietz (first president). Other notable founder members were Walter Shewhart (first vice president), A. T. Craig, B. H. Camp, A. R. Crathorne, and Harold Hotelling.

1937 On September 28, chief Soviet statistician Olimpiy Kvitkin is executed on the orders of Stalin. His census reports showed the catastrophic 1932 famine (Holodomor) caused death tolls in the millions and contradicted Stalin's claim that the population was increasing. Declared part of a "serpent's nest of traitors in the apparatus of Soviet statistics," other statisticians involved in the census are also arrested and shot. ■

JEDI CORNER

STATFEST 2025: Connecting Students with a World of Opportunities

David Corliss

2025 StatFest Schedule

Thursday, September 11: Welcome, career panel, and networking

Thursday, September 18: Keynote, graduate school application process, and graduate student panel

Thursday, September 25: Exhibitor introductions, expo, networking, and closing

StatFest is free to attend, and registration is open on the StatFest website: *https://community.amstat.org/cmis/events/statfest/statfest2025.*

With the change in weather and start of a new academic year, JEDI hearts turn to StatFest. This annual event focuses on students from historically underrepresented groups in statistics and data science, helping them make a strong start to the year by making connections and building their support network. Not to be confused with DataFest, which is a hackathon open to everyone that occurs in the spring, StatFest is held each year in September or October. Hosted or organized by the ASA Committee on Minorities in Statistics, this community event encourages students to consider careers in statistics and data science.

StatFest, which is free and does not require ASA membership to take part in, began in 2001 with a one-day conference organized by Nagambal Shah at Spelman College in Atlanta. Over the years, the event has grown as it has helped grow the community. Many people participating today as speakers and expo presenters once attended as students, attesting to the impact this event has and the lifelong connections it can help make.

Raphiel Murden at the 2023 ASA StatFest

In the past, StatFest was hosted at a particular college. This year, to make it more accessible to a wider audience, it will be virtual and presented in three successive sessions on Thursdays from 7–9 p.m. ET. It will begin with a welcome to students, who will then be introduced to others pursuing careers in statistics and data science and offered advice on career development. The second session will focus on getting ready for graduate school, while the third session will encompass a virtual expo to connect students with opportunities during their college program and beyond.

Every year, StatFest brings together leaders as speakers and expo participants. For students from underrepresented groups, it is a way to see what is possible from people who have done it while getting support for their college journey. Since it is organized by the ASA Committee on Minorities in Statistics, StatFest also allows students to become plugged into everything the committee offers, including their diversity mentoring program.

A critical part of StatFest is how leaders from all fields serve as mentors and role models. If you would like to participate as an exhibitor or sponsor, visit *https://community.amstat.org/cmis/events/statfest/statfest2025* and fill out the form.

StatFest is free to attend, and registration is open on the StatFest website at *https://community.amstat.org/cmis/events/statfest/statfest2025*. Meanwhile, share the word about the event with colleagues and students. As it says on the Committee on Minorities in Statistics website, "Join us to discover where your abilities can take you!" ■

STATS4GOOD

DataFest: Solving Big Problems with Big Data

David Corliss

Getting Involved

In opportunities this month, now is the time to start organizing for DataFest 2026. You can get details about putting together an event at the DataFest hosting page on the ASA website (*ww2.amstat.org/education/datafest/hosting.cfm*). It all starts with a faculty organizer subscribing to the listserv as described there.

The US federal statistical landscape is seeing many changes that could affect the ability of many people and organizations to use federal data for the greater good. In response, we can become familiar with the work of organizations supporting federal statistical agencies, including the Council of Professional Associations on Federal Statistics and the Friends of the Bureau of Labor Statistics.

This month, *Amstat News* is celebrating all things DataFest. This amazing event, now in its 15th year, is known for teams of students collaborating on big data challenges. Perhaps the most important parts of DataFest are what the data is about and what the students do with it. DataFest is a nationwide celebration of students learning to do more than make a grade; they make a real difference!

Since its beginning at the University of California, Los Angeles, in 2011, DataFest has trained students in essential skills needed in Data for Good. DataFest is now managed and coordinated by the American Statistical Association, with thousands of students taking part each year. Students are usually undergraduates, although some groups allow master's degree students to take part. Each participating college hosts its own event over a weekend in the spring, during which teams of two to five students compete against one another over 48 hours to understand, visualize, and interpret a large, complex data set. Mentors are available to assist, but each student team takes the lead in designing, developing, and presenting their own project.

In recent years, solving big problems with big data has become even more of a focus. For example, DataFest 2022 used data from the Play2Prevent Lab at the Yale School of Medicine to identify and analyze patterns of risk-taking behavior in middle school students. Game logs from a gaming community in New Haven, Connecticut—Yale's hometown—provided the unstructured big data used in the student projects.

I saw this happen firsthand at the DataFest event hosted by Purdue's DataMine, during which I participated as an outside statistical expert from industry brought in to help judge the student projects. Participation is a wonderful way to support students and the projects while providing an example of a person in industry volunteering in Data for Good. Officially a judge, the chance to be a role model for the students was the best part!

Another example comes from the 2023 challenge. Using data from the American Bar Association, DataFest participants explored how to best provide legal experts who offer pro bono advice through the Bar Association website. This is a great example of how data experts can partner with legal professionals to make a huge impact on peoples' lives.

While the data used in DataFest projects each year is kept secret until the event begins, the subjects often address critical questions of the day. In 2021, for example, an international survey on prescription drug use with more than 10,000 responses was used to help medical teams identify possible misuse of prescription drugs. Student participants were taught how analysis of big data can literally help save lives and that they can be part of making it happen.

DataFest is not just for students and faculty. Statistical experts from all areas are needed as project mentors and judges. I work in industry and have participated in DataFest multiple times this way. Also, faculty can be involved through other institutions if their college isn't hosting an event that year. Schools that have not hosted a DataFest before can get started by taking part in a DataFest at another college.

Starting out as a multisite hackathon teaching big data skills, DataFest has developed into one of the most important ways students can get their first exposure to Data for Good. By teaching students to use big data to solve society's biggest problems, this single weekend opens the door to a lifetime of high-impact projects helping thousands of people by doing what we love best. If you are part of a DataFest program, let others know the success stories from your college and the impact they are having on people's lives every day. ■

With a PhD in statistical astrophysics, **David Corliss** works as a data scientist in industry. He serves on the ASA Board as a Council of Chapters representative and is the founder of Peace-Work, a data for good nongovernmental organization.

15 YEARS OF ASA DataFest

CELEBRATING THE CELEBRATION OF DATA

Jessica Karch, Jennifer Noll, James K. L. Hammerman, and Traci Higgins

DataFest, a celebration of data that began at the University of California at Los Angeles in 2011, includes 60 sites involving 120 institutions from across the globe. It is a 48-hour-long co-curricular event in which teams of undergraduates come together every spring to tackle an authentic data problem. Corporate and civic sponsors donate data sets to DataFest for students to work on. Data sets can be situated in any authentic context ranging from the play of a national sports team, to online textbook use, to pro-bono legal advice services.

We first became involved with DataFest through research. Jennifer Noll and her colleague from TERC, Andee Rubin, started initial pilot work with Rob Gould in 2019 and 2020. In 2020, they conducted recorded observations of two teams at three points over the 48 hours to better understand the ways teams worked together from start to finish. Building on that pilot work, we ran a three-year, NSF-funded study that took place at six DataFest sites with almost 1,000 students. Our research explored why students take part in DataFest and how teams worked together to navigate the data investigation process within this open-ended context with large, authentic, complex data.

Much like the celebration of data itself, this special section celebrates 15 years of DataFest. It features perspectives from site organizers around the world, who reflect on how DataFest has evolved, what has enabled their events to succeed, how to broaden participation in DataFest and data science, and the unique opportunities and constraints of DataFest in their particular contexts. These include DataFest for two-year colleges, DataFest at international sites, and institutions that travel to host sites. These articles illuminate the overarching similarities across DataFest sites and the variability and flexibility that allow this event to thrive in many locations.

This section also features perspectives that provide insight into how the data for DataFest come to be. Rob Gould and Mine Çetinkaya-Rundel reflect on the affordances and challenges of retrieving and curating a usable data set for DataFest. On the other end of this process, two teams share their experiences as data donors. One donor was a DataFest participant as an undergraduate, and his essay speaks to his

Jessica Karch is a senior researcher at TERC, which uses qualitative and mixed methods to study science learning and learning environments at the undergraduate and graduate levels with a focus on equity.

Jennifer Noll is principal investigator at TERC. Her background focuses on K-12 and undergraduate statistics and data science education through innovative curricula, technology, and teacher professional development.

James K. L. Hammerman co-directs the STEM Education Evaluation Center at TERC. For more than 20 years, he has worked as an evaluator, designer, teacher educator, and adviser for innovative statistics and data science education projects, engaging formal and informal learners of all ages to investigate and make sense of data.

Traci Higgins is a senior researcher in STEM education at TERC. She has more than 20 years' experience conducting research and developing educational materials, processes, and models to support STEM learning and teaching both in and out of school, focusing on K-8 mathematics, data science K-12+, the social sciences, and interdisciplinary thinking.

own experience and his motivation to give back to the community by becoming a data donor. Representatives of the second donor describe their team's prior work as judges at the event, their decision to become a data donor, and what the experience meant to them.

Beyond individual experiences, this collection of essays explores how DataFest has catalyzed meaningful partnerships between universities, industry, and government that extend far beyond the weekend event. These collaborations have profoundly shaped the academic and professional trajectories of participants while also influencing institutional change, from the development of new data science majors to the establishment of centers for data science. The essays and Q&As from mentors, judges, and students capture the immediate excitement and lasting benefits of their DataFest experiences, alongside the unique challenges each group faces.

Finally, we feature findings from our three-year National Science Foundation study that explored why students participate in DataFest and how they make sense of data during the event. The study provides insights into organizer goals and participant expectations and considerations on who DataFest is for.

These articles provide a diversity of perspectives on DataFest and show how the event has evolved and grown over its 15-year lifetime. While some articles explore tensions and challenges, we can see DataFest has evolved in meaningful ways that have broadened participation when taken as a whole.

As the American Statistical Association, universities and two-year colleges, statisticians, data scientists, statistics and data science educators and researchers, industry professionals, and undergraduate and graduate students continue to organize around DataFest and engage in thoughtful dialogue and research, the event will continue to evolve to provide a challenging and important informal learning event for a diverse range of participants.

Through these articles, the American Statistical Association community is invited to reflect on and engage in dialogue about what DataFest is and what role it can play for undergraduate data science and statistics education. ■

Shared Goals and Contrasting Objectives: Understanding DataFest Through Dual Perspectives

Jessica Karch, Jennifer Noll, James K. L. Hammerman, and Traci Higgins

Editor's Note: This material is based on work supported by the National Science Foundation under Grant No. DUE 2216023. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

DataFest has emerged as a prominent informal data science education event, offering participants a unique opportunity to engage with real-world data challenges. This extracurricular event is less structured than a formal university course, takes place over a weekend in the spring, and develops teamwork and communication skills while engaging students in an open-ended challenge with large, complex, authentic data. The core elements of DataFest are similar across sites but the local context and multifaceted motivations, goals, and expectations of both organizers and participants creates opportunities and challenges statistics educators at each site will benefit from understanding.

This work is part of an Improving Undergraduate STEM Education project funded by the National Science Foundation. The research team spent two years at six DataFest sites conducting surveys and interviews with organizers and participants to better understand the goals and motivations for hosting and taking part in DataFest.

Five DataFest organizers, who are primarily statistics professors and lecturers, participated in a focus group interview related to their goals and expectations for DataFest at their sites. Organizers consistently articulated a clear set of two goals for the event. The most fundamental goal organizers described was their aim to provide students with more authentic, out-of-class experiences with real data. This core aim manifests in several key areas. A primary driver for organizers is to facilitate hands-on learning that complements and extends classroom instruction. They envision DataFest as a space for students to apply classroom skills to a real-life data set, learn how to work with messy data, and sharpen presentation skills and soft skills. For many, it's about giving students a tangible sense of what majoring and working in data science feels like.

Organizer Goals

Organizers like John Tukey (a pseudonym) from a private suburban research institution emphasize exposing students to "working with real data outside the classroom / in an unconstrained setting," while Margaret Hamilton (a pseudonym) at a large public research university sees DataFest as a direct application of learned skills. The event is also viewed as a significant curriculum vita item, providing participants with a valuable credential for their résumés and an experience they can discuss in future interviews, thus linking directly to career development and widening perspectives on professional opportunities.

Beyond individual skill enhancement, a second goal for organizers is to foster community. Organizers actively look to bring a sense of community to the stats department by bringing together alumni, faculty, grad students, and industry mentors and even creating regional connections by inviting satellite sites. Hamilton notes the goal of keeping alumni "involved and engaged in an authentic way," while others aim to rebuild communal, collaborative culture lost during COVID. This goal highlights DataFest not just as a learning event, but as crucial for building networking and shared experiences, where students can connect with professionals and peers. At some larger university sites, DataFest is seen as a capstone event for the related majors, while at others (that may not have a data science major) the event is used to get students excited about learning quantitative analysis skills.

Design and Recruitment Impact

Organizer goals directly influence design elements and recruitment strategies that span sites. To support skill development, structures include inviting mentors from industry and, at many sites, offering workshops for hard and soft skills. The 48-hour structure is seen as a way to manage data complexity while providing a challenging experience. The time constraints require participants to prioritize and make decisions quickly.

For community building, organizers aim to involve a wide range of stakeholders, from alumni to local schools. While recruitment strategies vary across sites, they generally target students with some level of ability/comfort with data and often invite students from different majors to encourage interdisciplinary teamwork. Some organizers, like Ada Lovelace (a pseudonym) from a primarily undergraduate institution, also try to "attract newbie students who can get excited about data science," even if the event tends to be dominated by more experienced students.

At other sites, DataFest is seen more as a culminating event for some of their students who are majoring in statistics and data science. In addition, some of these sites have more students who want to participate than the site can accommodate and teams are turned away. Because DataFest tends to attract primarily statistics, data science, and computer science majors, organizers mention the tensions between wanting to broaden participation at the event and the space constraints of their sites.

Student Motivations and Expectations Met

As detailed in "Spending the Weekend with Data: The Appeal of DataFest for Students" (Page 20), most participants are drawn to DataFest for many of the same reasons organizers promote it: skill development and career enhancement. These goals are by and large met during their DataFest experience. Participants also look for real experience to discuss in interviews, a chance to apply classroom skills, and the opportunity to learn to work in [interdisciplinary] teams. The networking aspect, allowing them to connect with industry professionals and meet alumni, faculty, and other students, is also a significant draw. The competitive element, while sometimes a source of stress, also motivates many to sharpen their presentation skills and learn to tell a data story.

Student Expectations Unmet

There were four areas in which we saw variability in terms of participants' expectations being met or unmet based on post survey reflections. First, while some participants expressed great satisfaction with their team experience and learning from one another, teamwork and team dynamics were a challenge for others. Second, while many participants found their mentors a major source of support, a few did not. Third, though many participants found the networking and presence of industry professionals engaging and helpful, a few did not think there were enough structured networking opportunities. Finally, some participants expressed expectations around the judging process and overall desire for more feedback.

Generally, participants who did express an unmet expectation about the judging fell into one of three categories: (1) a lack of transparency in the judging process; (2) no established rubric for judges; and (3) a lack of feedback about the details of their work. A few participants also seemed to expect more structure and guidance throughout the entire event.

In reflecting on our conversations with organizers and participants, we think some of the participant expectations around more feedback, structure, and guidance at DataFest may be grounded in their formal school experiences. There may be an implicit distinction between how organizers and students understand what counts as a learning experience in that students may expect feedback as part of the learning experience of DataFest, while organizers may be more focused on the practice of working through complex, authentic data as the learning experience.

Recommendations

Guided mentorship is crucial, and some sites already have some mentor training. Prior mentor training can equip mentors with strategies for effectively balancing time among groups, proactively engaging with teams, and providing guidance on active listening and constructive criticism approaches. Mentor training also empowers industry professionals who may be shy or less comfortable working with students.

Expecting judges to have familiarized themselves with the data set in advance and provide feedback on the technical aspects of each team's work during each event is a lot to require for judges who are volunteering their time over a weekend. However, each site may be able to provide a general rubric judges would use for all participants at the start of the event to help clarify expectations. In addition, organizers could implement a peer review/judging in which teams in the final round could be judged by participants and there could be a peer award given based on peers' perceptions of the final set of presentations. Engaging in community voting in this way may also keep participants who did not make it to the final round present and engaged. It is also an opportunity for those most familiar with the data (the students) to give their peers constructive feedback on their analyses. Peer feedback has the potential to develop soft skills around giving and receiving feedback, an important skill for working on teams in industry settings.

Finally, team dynamics could be addressed through pre-event meet ups, guided team-building icebreakers, or mechanisms for organizers to help teams address their goals and expectations for DataFest so each person on the team is better prepared to be a contributing member. Setting shared expectations about teamwork ahead of time could help address conflicts before they appear. Some DataFest sites already have a pre-event, and organizers from those sites might have more suggestions for how to make the most of these meet ups.

DataFest is a powerful vehicle for data science education that complements participants' formal educational experiences. By explicitly acknowledging the nuanced motivations and expectations of both organizers and participants and trying to create diverse structures that support participants with varied goals and expectations, statistics educators can ensure DataFest continues to be a transformative and enriching experience. The goal is not to host an event, but to cultivate a vibrant, supportive, and truly educational environment in which every student—regardless of their background, current skill level, or goals—feels empowered to engage with the world of data.

Insights gleaned from organizer and participant expectations and goals offer a valuable roadmap for the broader landscape of informal undergraduate STEM education events. Informal events such as DataFest offer an environment for undergraduate students to gain important experiences working with authentic data in ways many students do not have access to in their formal academic settings. Yet, there can be tensions between student expectations rooted in formal settings. Better understanding these tensions and studying different approaches might allow for best practices for fostering important informal STEM learning experiences. ■

Who Is DataFest For?

Jessica Karch, Jennifer Noll, James K. L. Hammerman, Traci Higgins

Emory College had 11 teams totaling 28 students participate in its first DataFest.

DataFest 2019 participant deep in analysis

During the first year of our National Science Foundation-funded project, we set out to understand what drives students to take part in DataFest, as well as what discourages participation. To do so, we collected multiple streams of data from six DataFest sites: surveys with Likert items and free-response questions implemented with both DataFest participants and nonparticipants before the event; semi-structured interviews with DataFest participants; and surveys about participants' experiences at DataFest implemented after the event. When we ran a logistic regression comparing survey responses from DataFest participants versus nonparticipants, we found one variable that significantly predicted DataFest participation was the extent to which students perceived DataFest as "for me." When we shared this with the organizers, they asked us what "for me" means.

Unpacking what contributes to a sense of belonging is not a trivial task. At an event like DataFest, many personal and professional factors may contribute. Students' majors, personal identities, and how relevant they perceive the DataFest challenge is to their personal and professional goals may all influence whether they see DataFest as an event for someone like them. Belonging at a data science event may also intersect with feelings of belonging in data science as a field more generally.

According to Zippia, approximately 80% of all data scientists in 2021 were men and 64% were white. Prior research suggests that to make data science more inclusive, it's important to challenge some of the exclusionary culture baked into data science and STEM fields more generally. Creating opportunities that are more collaborative than competitive; that affirm students' identities and protect students from racialized and gender-based stereotypes and experiences of microaggressions; and that are intentionally designed for inclusion, social justice, and social impact are all important to welcome students who have been historically marginalized in data and computer science.

Our data could not conclusively answer what mediates a sense of belonging at DataFest. This came down largely to the limitations of our data. Our data set offered rich insight into the experiences of participants at DataFest, with many paired pre/post responses and rich semi-structured interviews. Our data on nonparticipants, however, reflected many of the same limitations organizers faced when trying to recruit students to take part in DataFest. We recruited participants through organizers' existing campus networks, so it is difficult to say who our data set included and who was left out. Although we had similar response rates between participants and nonparticipants overall, nonparticipant response rates varied widely from site-to-site, especially in the first year of our study. Importantly, our data reflected the same gender and racial underrepresentation many organizers saw among their majors and events. To gain insight into the question of "for me," we had to turn away from our quantitative survey data to qualitative survey free responses and interviews with DataFest participants.

A common refrain about DataFest is that it is an event for a bunch of data nerds geeking out. This idea was reflected by both organizers and students. In fact, one DataFest participant in our data set lamented DataFest wasn't geeky enough.

The nature of the 'secret' data set may contribute to this data nerd perception. While other kinds of hackathons and datathons lure students in by inviting them to engage with data problems relevant to their lives or majors, an important feature of DataFest is that the data set remains secret until the Friday evening kickoff. This means a primary draw of DataFest is excitement to work with authentic data, no matter what the data is.

Data nerds were not necessarily data science and statistics majors. A data nerd could be someone who minored in data science, who had an interest in data without formally studying it, or who was interested in learning how data works in ways that could help them in other domains, such as business or the social sciences.

This perception of DataFest being for data nerds meant some students believed advanced data and coding skills were needed to participate in DataFest. However, we also found that the perception of DataFest as being only for experienced data science students could be challenged by the influence of peers and faculty mentors. One participant, for instance, was encouraged to join by a supportive teacher who emphasized there were no expectations and they would do just fine. Several others shared that personal encouragement helped make the idea of taking part in DataFest feel less intimidating.

Other participating students who did not think they had strong data science or computing skills struggled to reframe their own contributions at first, but participating in DataFest helped them shift their perspective on what kind of skills are valuable for data science.

In interviews, we found soft skills were just as important to team success at DataFest as computing or data skills. Being able to design a beautiful presentation, communicate their findings and analysis clearly, and having knowledge or lived experiences that were relevant to the domain of inquiry were all crucial skills to complete the DataFest challenge.

Our advice to organizers: Data nerdom is expansive and multifaceted. Our findings suggest students find many pathways to engage in and become excited about data. Furthermore, a sense of belonging does not happen by accident. As Alex Fisher and Maria Tackett discuss in "A Point of View: Duke University" (Page 42), it takes intentional work to create a more inclusive and successful event. We encourage organizers to carefully consider why they are hosting DataFest and how that goal shapes their messaging about DataFest and who they envision as the ideal DataFest participant.

Some questions for organizers to consider are the following:

- Who counts as a 'data nerd'? Does being a data nerd imply a specific skill set or affinity for specific statistical and computational disciplines? Do social science or other students who work with data count? How does this affect your recruitment?
- Who is your DataFest event currently engaging and how? Who among your students and faculty can act as local DataFest champions, and how can you activate those folks to intentionally and individually invite and encourage students who may not otherwise see themselves as part of DataFest?
- Which students are being systematically excluded from broader data science and statistics programs, and how can you help shift the culture of their programs? If your program has successful inclusion efforts, what has made it successful and how can you help those continue to build and grow?
- Are there opportunities to partner with local community colleges to broaden the impact of their DataFest event?

From our experiences, it is incredibly challenging to understand patterns of participation at an event like DataFest, which recruits broadly and may differ from site to site. Next steps for research can be to work with organizers and students on a case study basis to do a deeper exploration of factors that influence participation at single events.

Some questions for researchers to consider are the following:

- How does culture at a given event, institution, or field impact how marginalized students do or do not see an event as "for me"? How are narratives constructed around what kinds of skills are or are not valuable in a data space?
- How is your data situated within a broader ecosystem? How do you make sense of your phenomenon within that ecosystem?
- Who did you sample from, and how do you know that sample is representative? Whose voices are included, and whose are excluded? How do those patterns of inclusion and exclusion help you gain insight into your phenomena of interest?
- What mediates belonging at DataFest and similar kinds of datathon events?
- What practices and/or structures can lead to a stronger workforce and better outcomes for young people looking to build meaningful and productive careers in data science? ■

Editor's Note: This material is based on work supported by the National Science Foundation under Grant No. DUE 2216023. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

Spending the Weekend with Data: The Appeal of DataFest for Students

Jessica Karch, Jennifer Noll, James K. L. Hammerman, and Traci Higgins

Editor's Note: This material is based on work supported by the National Science Foundation under Grant No. DUE 2216023. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

The American Statistical Association's DataFest has grown into a hallmark event in the academic calendar for many statistics and data science programs at universities across the United States and as far away as the UK, Germany, and South Africa. Far from being just another competition, it has become a vibrant nexus for students with diverse backgrounds who share a common interest in working with real-world data challenges. As statistics and data science educators, understanding why students choose to dedicate an intense weekend to this immersive experience is crucial for maximizing its impact and ensuring its continued success. For the past three years, our Improving Undergraduate STEM Education project, funded by the National Science Foundation, has embarked on a qualitative and quantitative research journey to uncover the driving forces behind student participation in DataFest.

Our investigation spanned two academic years and encompassed six DataFest host sites. We surveyed 892 students and interviewed 56 student participants. Survey questions held a place for open-ended responses as to why students decided to take part in DataFest, as well as choice options with a range of reasons for participating. The semi-structured interviews gave us a chance to more deeply explore what drew students to DataFest.

This blended methodological approach allowed us to corroborate qualitative insights with quantitative trends, painting a holistic picture of student motivations. We share some initial general results from our survey and interview data around student motivations for taking part. The categories described in the table on the following page are the categories developed in the analysis of both survey and interview responses. Note: Values sum to more than 100% because students could select multiple reasons for participation on the survey.

Overall, our findings paint a clear picture: Skill development (both soft and technical skills) stands out as the predominant motivator for students attending DataFest. Specifically, the desire to apply classroom-learned skills to real-world and often messy data sets consistently appeared as a top reason. This was closely followed by a keen interest in developing or refining specific data science and statistical skills. While the development of coding ability was certainly a factor, it was often framed within the broader context of data science applications, rather than as a standalone goal.

Beyond technical prowess, networking with peers and professionals and career development ranked highly, possibly reflecting students looking toward their future career prospects. The set of quotes in the soft skills development category highlights (see table on next page) that some participants recognize the importance of soft skill development in terms of teamwork and communication and that these skills will be important in their future careers. While social reasons—such as connecting with friends or meeting new people—were undeniably important, they tended to be secondary drivers when compared to the tangible benefits of skill acquisition and professional development.

In summary, student participation in DataFest is overwhelmingly driven by a profound desire for practical skill development, particularly in applying statistical and data science methodologies to real-world challenges. Networking opportunities also serve as a strong draw, fostering connections crucial for burgeoning careers. While social engagement and the allure of competition or rewards play supporting roles, they rarely stand as primary motivators. Students' primary motivations in terms of expanding their technical and soft skill sets might point to the limits of classroom work and how DataFest expands opportunities to develop important skills students do not find in a formal classroom environment.

Factors such as the allure of 'free stuff' (e.g., food, swag) or the competitive drive to win were present but generally served as ancillary perks, rather than primary motivators acting more as 'nice-to-haves' for students already drawn by the core experience. However, we would recommend site organizers continue to think about swag (because as one participant noted, it adds to the festive atmosphere) and (allergen safe) free food. Free food is important because it builds a sense of community and keeps participants happy and well fed, especially given that some teams travel to DataFest sites. Not having access to free food could create a deterrent for students working on tight budgets. Some participants appreciated the event's atmosphere, noting the presence of food, advertisements, and other engaging elements that contributed to the overall experience.

Reasons for Participating in DataFest	Percentage of Participants
Technical Skill Development: Apply academic skills to authentic, open-ended complex problem-space; build data science, statistics, or computing skills	33%
Career Development: Networking, something to add to a résumé or generally good opportunity for career, achievement—winning a prize, friendly competition	29%
Soft Skill Development: Teamwork, collaboration, communication, and/or confidence skills	22%
Social Reasons: Friendly—spending time with friends or making new friends	20%
Welcoming Environment: The event is supportive, non-judgmental, low bar to participate, and/or inclusive	14%
Encouraged to participate: Faculty, teacher, mentor, etc. encouraged the participant to sign up	11%

The following table uses the same categories as the one above but shares quotes from interview participants to offer additional insights into what motivated these students.

Reasons for Participating in DataFest	Student Quotes
Social Reasons	• We were already a friend group and one of our friends just presented this to us. • I'm only going here [DataFest], so I can have fun with my friends. Like, so, I can hang out with my friends and stuff.
Career Development	• It's a good résumé builder if you win, or even if you just participate. • Adapt [skills learned] into the future if I want to have a job.
Technical Skill Development	• DataFest was the first time I could really see how a messy data set could be used or analyzed into actual actionable suggestions for a company to move forward. • This topic was what I tackled before, only theoretically. So, this time ... it was really helpful for me to analyze real data and visualize it. • I think I might have learned more in that 24 hours than I did for my whole data literacy class about R. • I learned a lot about the importance of data visualization in helping tell your story.
Soft Skill Development	• I enjoyed the community aspect. I think that there was a big emphasis on collaboration. • Everyone had different skill sets, and we had to kind of make sure that everyone's skills were focused in on what they were good at. • So, I think definitely being able to build off the exist-ing knowledge you had, but also being flexible and understanding. It was great to hear the perspectives of other people.
Welcoming Environment	• I felt like it was an environment that someone, like a group like ours, that was there to mainly learn and have fun, could thrive.

While not often mentioned in surveys or interviews, encouragement from a teacher, mentor, or faculty member appeared as a powerful motivator for some students. One participant described how their data science teacher, who also served as a consultant at DataFest, reassured them they would do fine and there were really no expectations. It seems that for prospective DataFest sites and site organizers, or for groups historically marginalized in STEM, the personal connection faculty can bring by inviting participants and building their confidence may have important implications for broadening participation (see "Who Is DataFest For?" on Page 18).

Finally, our data also hinted at site-specific nuances. For example, participants from one site showed a slightly larger proportion of interest in developing teamwork/collaboration skills, while demonstrating a slightly lower emphasis on real-world application skills compared to the overall average. This site also differed from other sites in that it did not have a data science major, only a data science minor. This suggests the local academic offerings or specific promotional strategies for DataFest at a particular institution might subtly shape participant motivations. This raises an intriguing question for organizers: How might we better tailor DataFest experiences or promotional strategies to align with the unique interests and motivations of students at our host site? ■

DATA: The Engine That Drives DataFest

Robert Gould and Mine Çetinkaya-Rundel

You can't have a DataFest without data. But how does the ASA DataFest find and assemble the data sets thousands of students scrutinize to squeeze out every drop of information?

Rob Gould and Mine Çetinkaya-Rundel have been selecting the data for DataFest from the beginning. Gould founded DataFest at the University of California at Los Angeles in 2011, when Çetinkaya-Rundel was in her final year as a PhD student. They have been planning DataFests ever since.

One driving force behind founding DataFest was to provide undergraduate students with a real-life, data-driven problem. Data sets with large numbers of rows and columns, while common in the workplace or well-funded research labs, are rarely found in the classroom. As a result, students are likely to work with well-explored data that holds little mystery or room for discovery.

In 2011, Gould and Çetinkaya-Rundel's goal was to provide students with a challenge that had never been tackled and a client who was truly interested in what the students had to say. "We thought organizations looking for assistance with their data would be eager to have a group of talented undergraduates look hard at their problems," wrote Gould and Çetinkaya-Rundel. "We imagined these 'data donors' would be mostly nonprofits and civic/government organizations, since these might not have statisticians or data scientists on staff or might have interesting problems they don't have the time or resources to tackle." Instead, Gould and Çetinkaya-Rundel found many organizations are eager to take part.

The first data set was provided by the Los Angeles Police Department. At the time, data-driven policing was relatively new, and the LAPD was eager to use data to reduce crime and increase public accountability. The LAPD had been working with UCLA researchers for a few years after their new publicly available crime map erroneously suggested the most crime-ridden location in Los Angeles was the headquarters of the *Los Angeles Times*. The LAPD was eager to recruit the best young data scientists into the new realm of predictive policing, so they were a natural donor candidate for the first DataFest.

The data often requires heavy preparation, and we often need to provide additional context to help the students have a productive weekend.

The data required a lot of preparation before it was released to the students, and the entire UCLA Department of Statistics pitched in. For example, the location data used a unique origin point and proprietary mapping algorithm and had to be back-engineered to be mapped using standard latitude and longitude. In addition, Gould and Çetinkaya-Rundel decided to supplement the data set with the locations of transitional living facilities.

"Little has changed since then," wrote Gould and Çetinkaya-Rundel. "The data often requires heavy preparation, and we often need to provide additional context to help the students have a productive weekend." One thing that has changed, however, is businesses are more aware of the value of data, so it has become a challenge to get data from prominent corporations.

In the early years, Gould and Çetinkaya-Rundel were able to entice major tech companies into donating data: Edmunds.com; Ticketmaster; e-Harmony; and Expedia. But as companies grew wary of supplying data, students expressed interest in seeing data from beyond the corporate world. In response, data has come from the UCLA Department of Psychology (CourseKata Project), Yale School of Medicine (Play2Prevent Lab), Canadian National Women's Rugby Team, American Bar Association, and Rocky Mountain Poison Control Center.

Gould and Çetinkaya-Rundel find the data sets mostly through word of mouth, though some come from cold calls. As the years go by, more DataFest alumni are in positions to donate data, so Gould and Çetinkaya-Rundel ask former participants, presenters at the Joint Statistical Meetings, presenters at their local departmental seminars, and other faculty. "Ideally, we have about three promising leads by September," wrote Gould and Çetinkaya-Rundel. "We engage in conversations with these leads about what we're looking for and whether we will get their institution's permission to use the data. We ask for sample data and use this to try to flesh out a challenge and assess whether the data will meet our criteria."

Usually, one or two leads drop out as they realize they can't get institutional buy-in or don't have the time to assemble the data. Sometimes, they can donate data the following year, which helps.

Once Gould and Çetinkaya-Rundel have a data donor, they get the full data set and begin cleaning and preparing a codebook. The ASA supports a graduate student to help them with this, and this person usually does the bulk of the testing. Testing includes the usual tasks of data cleaning but also includes simulating how students might approach the data and looking for potential roadblocks they might meet.

Gould and Çetinkaya-Rundel make decisions about which aspects they should clean based on their assessments of how much time it might take the students to fix the data. This is an iterative process, with many queries back to the data donor, and leads to several versions of increasingly refined data.

Preparing a codebook is also a major endeavor, as organizations often have little documentation that can be shared publicly. This might be because they rely on institutional knowledge and practices, the documentation has proprietary information, or the assembled data is entirely new to the organization.

During this process, Gould and Çetinkaya-Rundel write and rewrite the challenge, which is the second-most important aspect of the competition. "We strive to provide a challenge that gives all students a path forward, regardless of their level of data sophistication," they wrote. "A good challenge provides sufficient leeway for students to exercise creativity and 'choose their own adventure.' At the same time, the challenge needs to be precise enough that students are clear about what the data donor is looking for."

The most important component of DataFest is the data, and the primary requirement for the data is that it be rich. For Gould and Çetinkaya-Rundel, "richness" is measured mostly by the number of distinct variables and the possibility of enhancement through external data or feature engineering with the available data. The context is key; it must be accessible to students so they do not waste valuable time learning a new subject but instead have sufficient understanding to think of interesting questions. There is also a novelty aspect; the data should not be publicly available, although parts of it can be. The data donor must also be willing to record a video to put a face with the data set so it is clear real people are listening and curious about what the students might find.

A constraint is that the data be large, but not too large. Gould and Çetinkaya-Rundel limit the data set and documentation to under 3 GB (2GB is ideal), which is about the limit of what can be distributed simultaneously to many students.

Additionally, Gould and Çetinkaya-Rundel cannot ask students to sign nondisclosure agreements, and this makes it difficult for data donors from private institutions. Students are asked to read a statement that informs them that by taking part in DataFest, they cannot use the data for any purpose other than DataFest without permission from the data donor.

Finding data for DataFest has been one of the more rewarding aspects of Gould and Çetinkaya-Rundel's careers, although there have been years of great stress. They hope to enlarge their data-finding committee. If anyone is interested, send an email to *rgould@stat.ucla.edu*. And if you know of anyone with data to share, send it their way. ■

Rob Gould is a teaching professor in the University of California, Los Angeles, Department of Statistics and Data Science. The founder of ASA DataFest in 2011, he is a Fellow of the American Statistical Association and recipient of the ASA Founder's Award. He is also a co-author of the introductory statistics textbook *Exploring the World Through Data*.

Mine Çetinkaya-Rundel is professor of the practice and the director of undergraduate studies in the department of statistical science and the director of first-year experience at Duke University. She is a Fellow of the ASA, recipient of the Waller Education Award, and co-author of *R for Data Science* and *OpenIntro*.

DONOR PERSPECTIVE

CourseKata's Experience as a DataFest Donor

Ji Y. Son, Claudia C. Sutter, and James W. Stigler

Who provides the data behind student projects—and what do they gain in return? In the following pieces, donors share their motivations and takeaways.

University of Toronto students watch the CourseKata introduction video (Jim Stigler and Ji Yun are pictured in the video).

CourseKata is a research and development project devoted to improving how students learn statistics and data science. We design innovative online curricula for use in high school and college classrooms, and we partner with learning scientists to figure out how to improve the teaching and learning experience by improving the curriculum.

We've long argued that education needs an R&D arm because the same kind of data-driven insights that fuel innovation in tech, medicine, and policy could (and should) improve how people learn. But here's the catch: Education data is notoriously hard to access, interpret, and act on. We're trying to change that.

When we first served as judges at the UCLA site of DataFest, we were blown away by the quality of student thinking. These weren't just shallow dashboards or flashy visuals. They were thoughtful, creative, and rigorous analyses of authentic data. Each team approached the same data set from a different angle, like the proverbial blind men encountering different parts of an elephant. Then, in just two slides and five minutes, each team presented their part of the story. And by the end of DataFest, a fuller picture of the data had emerged.

So, in 2024, CourseKata became a DataFest donor.

What We Donated

We shared an anonymized data set of student behavior and learning outcomes from 1,625 college students across 48 courses at 11 institutions, all using our online curriculum in 2023. This data set included engagement logs (e.g., active time on page); responses to formative assessment questions; and measures of psychological constructs related to motivation, mindset, and more. All data was de-identified per institutional review board protocols and publicly documented. (You can now find this data set at *research.coursekata.org*.)

What We Got Back

The student teams brought a range of perspectives and found plenty of surprises. One team investigated cheating and discovered the best source for getting the right answers to CourseKata questions … was the CourseKata textbook, itself.

Pomona College students present their CourseKata findings to the UCLA DataFest attendees.

Other teams tackled the question of whether instructional videos helped or hindered learning. Their answer? Both. Some found watching videos was associated with better performance, while others found not enough people watched the videos to make them worth it.

Several groups applied clustering techniques to explore patterns across motivation, engagement, and learning. These groups found students tended to fall into three clusters: those who engaged a little and got a lot right; those who engaged a lot and got a lot right; and those who engaged a little and got a lot wrong. These distinctions are helping us rethink how we interpret "time on task" and how to support different learner profiles.

One group that especially impressed us, Team Abercrombie, analyzed the behaviors of the highest-performing students and found they were the students most likely to revisit earlier sections of the textbook when working on new problems. Based on this insight, the team designed a system called ABER (Automated Built-in Engine for Review). ABER used vector embeddings to represent all textbook paragraphs and questions, then used similarity matching to recommend the most relevant sections of the text for review after a student answered a question.

We were so intrigued by their work that we invited several members of the team—Justin Gong, Hairan Liang, and Lukas Hager, all from UCLA—to join us as interns. They helped us refine a prototype of ABER, explored ways to use large language models to match new formative assessment questions to our learning goals, and even worked on generating new questions for concepts that were under-measured in our current curriculum.

Other DataFest participants have since joined us as interns, continuing to bring their creativity and curiosity to real-world projects. Over the past year, they've contributed to initiatives such as prototyping researcher dashboards, visualizing motivation and engagement across different textbook versions, and tracking changes in these measures across academic years.

For many of these students, it was their first opportunity to work directly with educational data in a research and development setting. They told us how much they appreciated the chance to work on projects with meaningful (and sometimes real-time) impact.

Their contributions weren't just for a class assignment or a grade, but to shape how future students learn hard subjects. Their reflections confirmed what we've always believed: When students are given access to complex data, support to explore it, and the autonomy to be creative, they generate ideas that can move the field forward.

Possibilities

DataFest isn't just a showcase of student talent. It's a proof point for what's possible when curious minds are unleashed on real data. Because these student researchers are closer in age and experience to the learners in our data set, they brought a much-needed perspective we don't get from seasoned researchers or curriculum developers. For CourseKata, it was a rare opportunity to be both generous and selfish at once: to support the next generation of data scientists while also learning more about how students learn.

If we want education to be truly evidence-based, we need to be thoughtful (and even a little shrewd) about how we use data to drive impact in a world in which data is too often used just to sell ads, products, and services. That means sharing it, contextualizing it, and being open to what others might find. DataFest gave us all that, and it reminded us that sometimes the best analysts are the ones just starting out. ■

DONOR PERSPECTIVE

Savills Workplace Studio

John Rissmiller, Associate Director, Savills Workplace Studio

John Rissmiller is associate director of Savills Workplace Studio.

I took part in DataFest in 2016 as a sophomore at Connecticut College. After that experience, I continued to study statistics at Conn and spent two years as a tutor, but I was unable to finish the statistics minor due to scheduling challenges. However, I was always trying to find ways to use mixed methods to inform my research. I used a statistical analysis in my honors thesis in anthropology and today spend my time drawing on both qualitative and quantitative data analysis to help clients reimagine their workplaces.

In my current role at Savills Workplace Studio, I help build out a database of benchmarking data. I also get data from our clients and analyze it to help them make decisions. The ability to distill lots of data down to a quick and effective story was something I first began practicing at DataFest and is a skill I employ every day. Finding my love for data and complexity at DataFest has driven my career and had a lot to do with why I chose my current organization.

My organization uses data in everything we do, from helping with leases and designing spaces to managing the build-out of spaces. Data and being active in our communities is in our DNA. So, the opportunity to use our data to give back and give opportunity to students around the world was something everyone at Savills was excited to do.

Our data set considered a wide variety of corporate real estate data, from rents and locations to sizes and time. It was critical to provide data that could be looked at in a variety of ways and complexities. This allows students of all skill sets to engage and have fun during what is an intense few days.

As a former participant, it was great to see students engage in our data and think about it from a totally different perspective than I or my team might. More importantly, after a few events, I got LinkedIn messages from students who wanted to tell me about how much fun they had and how interested they are in joining the industry. It's great to see that our work can be inspiring for students today.

I highly recommend any organization interested in being a data donor to do it and go to a competition. It's truly inspiring and a great way to give back to the next generation of budding data specialists.

There is nothing quite like DataFest for students. Even coming from a school that focused on project-based learning, the mix of the time crunch, the large data, working with a team—all of it combines into a really fun weekend in which you learn so much, not only about statistics, but about yourself, your skills, and where you want to grow and develop.

I highly recommend any organization interested in being a data donor to do it and go to a competition. It's truly inspiring and a great way to give back to the next generation of budding data specialists. ■

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. ■

INTERNATIONAL DATAFEST

Teamwork Makes the Dream Work: The Leeds-Pretoria-Wits DataFest Hackathons

Pierre-Philippe Dechant, Rukia Nuermaimaiti, Inger Fabris-Rotelli, Justine Nasejje, Najmeh Nakhaeirad, Raeesa Docrat, and Anna Kaduma Gumbie

Pierre-Philippe Dechant shares the story behind International DataFest—how it came to be and the lessons learned from organizing it. To complement his perspective, we also feature two Q&As with student participants: Anna Kaduma Gumbie of the University of the Witwatersrand and University of Pretoria and Rukia Nuermaimaiti of the University of Leeds.

How did the global collaboration to run DataFest come about?

We had various connections across institutions and research and student education that grew slowly over the years. Connecting the following dots essentially allowed us to team up to deliver the first DataFests in South Africa and England:

- About three years ago, Justine Nasejje at Wits instigated a joint research grant with Dechant at Leeds, which laid the foundations for more collaborative international work.
- Two years ago, Leeds and Pretoria developed a strategic institutional partnership, so Dechant was part of a delegation traveling to Pretoria to talk about data science. Nasejje and Dechant were trying to meet up then but Nasejje was at Wits and not Pretoria after all.
- But Wits and Pretoria have local links, so Nasejje connected Najmeh Nakhaeirad from Pretoria with Dechant. They then secured an International Strategy Fund grant to collaborate on data science education around sustainability, jointly attending CompEd in Botswana in October 2025, just up the road from Wits and Pretoria.
- With this tripartite link established and a focus on student education and research, they connected with other colleagues at the three institutions: Gumbie, Nuermaimaiti, Inger Fabris-Rotelli, and Raeesa Docrat.
- Nuermaimaiti from Leeds had heard from the Edinburgh organizers about DataFest and was keen to bring it to England. Her vision and enthusiasm pulled everyone along.
- In October 2024, Nuermaimaiti, Nakhaeirad, Nasejje, Dechant, and other collaborators ran student collaborative design workshops across Leeds, Wits, Pretoria, and Southwest Jiaotong University: collaborative online international learning, known as COIL. Students from data science–adjacent disciplines formulated a local sustainability challenge and collaborated on ideas for solving these with data science methods.
- Nasejje, Dechant, collaborators, and students submitted a joint research article on data science, health, and sustainability.
- A donation from David Fine created opportunities for international, interdisciplinary collaboration between Leeds and Wits. In January 2025, Nasejje traveled to Leeds, meeting Nuermaimaiti and Dechant and sparking ideas for data science research projects. Nuermaimaiti, Nasejje, Dechant, and collaborators were successful in getting a pump-priming and then a large grant on student education, data science, community engagement, and sustainability in June 2025.

The lessons learned from this? You cannot plan everything, but unexpected opportunities might come your way. Try to be open to these and agile, rather than precisely mapping out what is beyond your control. If you try to connect dots at each step, the network and impact grow and aims align, unlocking synergies between them. We often think we are too time poor to collaborate, but the gains are probably worth the effort in terms of time savings and increased opportunities.

Often there is a divide between research and student education. Using both in tandem, as appropriate, allowed the team to make big strides. And despite the large distance and different contexts, the team realized you can learn a lot from each other and might actually be doing similar things in different global and local contexts.

Global collaboration opportunities between universities are valuable. Universities are a globally networked superpower, with profound local links into communities and global links to each other. And a student education focus has such a huge reach. Your impact is literally through whole generations of students. What can be more satisfying than helping students develop their skills, collaborate, communicate, and connect globally?

Leeds academic staff members and jury panel members George Mbaeyi and Amos Chinomona. Chinomona, from South Africa, was one of the team members supporting the students during the hackathon. Mbaeyi will be one of the organizers of next year's Leeds DataFest.

What were the timelines, and what happened after Datafest?

The ongoing and practical ASA DataFest activities led to the following more sustained collaboration between the three institutions:

- November – February: Peer collaboration and mentoring on how to run the hackathon; sharing insights and experiences across Wits, Pretoria, and Leeds
- March 21–23: Leeds DataFest
- May 2–4: Pretoria and Wits run a joint South Africa DataFest.
- May 6–30: Several South African colleagues visit Leeds, funded by Horizon Platform and the International Strategy Fund.
- May 7: Wits colleague talks about decolonization, sustainability, and community engagement, particularly data science education and hack/maker spaces.
- May 8: COIL-Ed spring conference and workshop, an educator-facing COIL conference building on the October student-facing workshops. Nuermaimaiti, Nakhaeirad, and student attendees present on the Leeds-Pretoria-Wits DataFest experience, while other colleagues and student attendees present on the October COIL workshops as part of an international set of conference presentations including COIL work from Singapore, China, UK, and Ukraine.
- May 8: Inger Fabris-Rotelli and Nakhaeirad give a double statistics seminar in the school of mathematics; discussions about the possibility of joint supervision of PhD/MSc students
- June 23–26: Nuermaimaiti presents on the DataFest experience at the UKCOTS student education conference.
- October 2025: Nakhaeirad, Nasejje, Dechant, and collaborators attend CompEd data science education conference in Gaborone, Botswana, and lead a working group there

For 2026, the team hopes to offer DataFests again and pass the organization of the Leeds DataFest to the next generation of student education colleagues. They also hope to iterate and expand to Imperial College London, maybe even Singapore, and increase numbers overall. The team would like to increase opportunities for students to interact internationally, either through a synchronous approach across all three institutions or a meaningful asynchronous way to connect outside of the DataFest weekend. The team is planning to have joint student education conferences (e.g., COIL-Ed) that give colleagues and students opportunities to share collaborative international experiences, active/experiential learning, or both.

What were the international links for staff and students?

Driven by Nuermaimaiti's enthusiasm for DataFest, the team met online as an international team 8–10 times. They learned from each other about the DataFest format, the requirements for ASA membership and receiving the data set, convening an industry panel, raising funds in both South Africa and Leeds, and prizes and certificates.

For the students, there was no direct contact due to time

Rukia Nuermaimaiti and Pierre-Philippe Dechant

constraints. However, in terms of framing the event, the students were made aware that they were part of a global wave of DataFest hackathons in general and Leeds, Wits, and Pretoria in particular. At Leeds, the team recorded footage and greetings from the Leeds students in late March that they turned into a video for the South African DataFest participants in early May. The South African participants then recorded greetings the team could share on the Leeds DataFest teams space and at the COIL-Ed spring conference.

In the Leeds feedback questionnaire, 11 of 12 respondents reported they liked being part of a global wave of DataFest hackathons and having the link with South Africa. Nine of 12 participants answered the question, "How important do you think it is to include an international or cross-cultural dimension in hackathons like this?" with either "very important" or "somewhat important," with another two neutral and one responding "not very important." Generally, within Leeds, the groups were highly diverse as to disciplines, years of studies, and international background. A common theme from participants was how enriching it was to work in diverse teams, that everybody had something to contribute, that they learned from each other, and that it allowed them to play to their strengths. The industry panel was also diverse and highly international with members hailing from Nigeria, the Democratic Republic of Congo, and the Philippines.

Lessons Learned

The team members learned how similar, despite the different national contexts, their institutions and interactions with students and the local community are, along with how beneficial it is to share experiences and resources.

They originally looked into the possibility of doing a joint hackathon between all three institutions, but soon realized the different times in the northern and southern hemispheres, holidays, and teaching and assessment schedules weren't going to allow it.

One lesson the team learned is to keep it simple. Where there are low-hanging fruit in terms of synergies, go for it. However, most people are time poor, so doing something practical that works simply is key. The team aligned itself with the following approach: try something simple; prototype/pilot; get feedback; expand; revise; iterate.

DataFest is an experiential learning opportunity for students, as well as a learning opportunity for educators and a vehicle for expanding collaborations and fostering community. It was key for the team's evolving collaborations, straddling student education and research, data science, experiential learning, and sustainability. ■

INTERNATIONAL STUDENT PERSPECTIVE

Rukia Nuermaimaiti

University of Leeds

Why did you choose to host DataFest at your institution?

We chose to host the ASA DataFest data science hackathon to bring this dynamic, collaborative event to England and offer our students an invaluable opportunity to engage with real-world data challenges. In recent years, the importance of data literacy, collaboration, communication, and interdisciplinary analytical thinking has grown across nearly every field. Hosting DataFest aligned perfectly with our institutional goals to foster these skills among students, not only in STEM disciplines but across the wider academic community.

By organizing this event, we aimed to bridge the gap between theoretical coursework and practical, industry-oriented experience. Our goal was to give students a platform to test their skills, work with messy, real-life data, form insights, and communicate those insights clearly mirroring what data scientists do in practice. DataFest also enabled us to connect students with industry professionals and academic mentors in a setting that promotes creativity, innovation, and collaboration.

What are some highlights or memorable moments from this year's event?

We opened registration to students across multiple departments, welcoming participants not only from mathematics and

statistics but also from biology, chemistry, medicine, computer science, and more—more than 10 departments in total. This diversity of academic backgrounds led to rich, multidisciplinary teamwork and a range of creative approaches to problem-solving.

One of the standout moments from this year's event was the participation of a visiting team from South Korea. The team comprised one student from the school of mathematics and three from the school of languages. Their unique combination of technical and communication skills impressed the judging panel, earning them the Best Communication & Data Storytelling Prize (a Leeds-specific award). Judges commended their ability to present their findings clearly and concisely, supported by effective visualizations that communicated the key messages of their analysis. Their success underscored the power of interdisciplinary collaboration and demonstrated that effective data storytelling is as important as analysis.

How did students engage with the data or challenge?

Students enthusiastically engaged with the real-life data set and found the process of developing their own research questions to be both a challenge and a valuable learning experience. Many undergraduates noted that, unlike regular coursework where questions are predefined, this open-ended exploration pushed them to think critically and creatively. With guidance from PhD students and academic staff, all teams produced commendable work. The Best Use of External Data Prize especially encouraged students to think outside the box and integrate diverse data sources into their analyses.

Group photo from Leeds ASA DataFest

What lessons did you learn from organizing the event?

One of the key challenges we faced was securing industry professionals to serve on judging panels and offer feedback to students. Initially, it was difficult to find people who were comfortable engaging in an academic hackathon setting. Scheduling the event during weekends and the panel for Sunday afternoon also posed some equity, diversity, and inclusion considerations for both the panels and students. However, once committed, the industry participants were generous with their time and expertise. Their presence added value, giving students direct exposure to how their work would be received in real-world settings.

As organizers, we also gained insights into the kinds of qualities that are most valued in industry today. These include not only technical data-handling abilities, but also skills such as storytelling, communication, collaboration, and visualization. These insights will inform how we teach data-related modules and help us better prepare students for data-driven careers.

How did the event foster community or collaboration?

Some teams registered as preformed groups, while others were matched based on their indicated skills and preferences. We took care during registration to understand students' interests, technical strengths, and desired roles. All participants agreed to a code of conduct beforehand, and the atmosphere throughout the weekend was incredibly respectful, collaborative, and inclusive.

Feedback showed students valued the chance to work with peers from different cultural backgrounds and levels of expertise, articulating how important diversity is for effective teams. Many formed friendships and professional connections during the event. Cross-disciplinary collaboration emerged as one of the most rewarding aspects, both for the students and the mentors who supported them.

What feedback did you receive from students, judges, or faculty?

Feedback from everyone was positive. Students described the event as useful and fun, with many expressing interest in returning. They appreciated the chance to apply their learning in a challenging and supportive environment.

Judges—both academic and industry-based—praised the quality of the student work and were impressed by the participants' ability to synthesise and communicate complex findings under time constraints.

One of our academic judges, George Mbaeyi, has already volunteered to take the lead in organizing next year's event. Additionally, several faculty members have expressed interest in expanding the scope of the event. In particular, colleagues with links to China are exploring the possibility of co-hosting a future DataFest in collaboration with our partner institution, Southwest Jiaotong University. ■

INTERNATIONAL STUDENT PERSPECTIVE

Anna Kaduma Gumbie

University of the Witwatersrand

Anna Kaduma Gumbie

Team Stats Alchemists (from back): Odey Redi Mofokeng (University of the Witwatersrand); Awswell Makhubele (University of the Witwatesrand); and Orilwela Thagwana (University of Pretoria)

Why did you choose to host DataFest at your institution?

We chose to host the ASA DataFest (as a joint event between the University of the Witwatersrand and the University of Pretoria) as a way to increase value in our statistics students' learning experience and foster collaborative work between our two departments. We also saw this as an opportunity to expose our students to the excitement of real-world data analysis with the added advantage of working on a problem being worked on by other institutions in other parts of the world.

What are some highlights or memorable moments from this year's event?

There were many memorable moments, but among those that stand out were the collegiality amongst students; they had a jovial camaraderie. The initial excitement coupled with a bit of anxiety in the faces of the students not knowing what to expect, followed by the spirit of competitiveness, which emerged as the groups were formed and the work was started. The moments of laughter and excitement as they delved deeply into the amounts of data presented to them and the sheer overwhelming feeling at the end of Day One for many of the students. The determination to solve the problem presented, the final day excitement and anxiety waiting to present results to judges, and—of course—a lot of tea and pastries consumed.

How did the students engage with the data or challenge?

Students took the task of answering the questions to heart by remembering it is imperative to explore your data thoroughly—keeping the questions in mind—before breaking down the task at hand into smaller sections. We must note the main reaction, though: OVERWHELMING!

After calming down, students threw themselves into the task of answering/addressing each section they identified as necessary. There was much enthusiasm, and it was fascinating to see how each group (we had five, each comprised of three students) emerged with quite different sections at times. It brought to the fore how important data is in telling a story/making informed decisions and how the decisions are made depending on where the focus is placed.

What lessons did you learn from organizing the event?

Being first-time participants, our lessons are many and varied. Our main lesson, however, has to be from the funding perspective. We learned it is imperative to start early in seeking funding for such an event, especially if we are to follow the same modus operandi we did this year. It is not easy to raise funding for an event such as this in our parts of the world, which involved having to organize funding for accommodation at a common suitable venue, transportation to the venue for some of our participants, and catering. We hosted the event near UP, and Wits students needed to be transported and accommodated at the venue (we are about 37 miles apart).

The event was made possible through funding provided by the DSTI-NRF Centre of Excellence in Mathematical and Statistical Sciences, South Africa, in partnership with the Statistics HUB

Team Data Heads (from back left): Nicolas Andersch (University of the Witwatersrand); Michael Tankle (University of the Witwatersrand); and Nuelle Janse van Vuuren (University of Pretoria); Anna Kaduma Gumbie

Team Number Ninjas_SA (from left): Kopano Letsela (University of the Witwatersrand); Rachel Lock (University of Pretoria); and Pfano Khakhu (University of Pretoria)

Team Greyt (winners, from left): Munyaradzi Ndumeya (University of the Witwatersrand); Itesiwajuayo (Tes) Babalola (University of Pretoria); and Kumaaipurua Rukoro (University of Pretoria)

of the department of statistics at the University of Pretoria.

We are indeed happy that, with the short time we had to prepare (about four months), we were able to host a successful event. We are indeed grateful to our colleagues from both institutions who participated in the event as judges and facilitators for their commitment and self-sponsoring (transport) to and from the venue.

In what ways did the event foster community or collaboration?

Our joint hosting of this event was the first of its kind between our two departments. We were delighted to see our students come together to work on problems in a very collegial manner. Although it was set in a competitive environment, the collaborative mood that existed among them was encouraging and showed how, as future professionals, these students will have little to no challenge in engaging with peers from different walks of life to solve common problems that require their expertise. These students have kept in contact with each other through avenues such as LinkedIn, and it is our hope this will continue as they transition into their professions in the statistics field.

Did you receive any feedback from students, judges, or faculty?

Our students highly appreciated the chance to participate in the ASA DataFest. Feedback from them included statements such as "teamwork makes (for) dream work," "we learned a lot," "exciting," "intense," "a great experience," "got a good insider of what real-world data looks like," "lots of fun," and "interesting to work with different types of people and see how they think." They thought this experience should be a compulsory session and incorporated into their study year as practical experience. They considered this a vital step in helping them build their confidence and preparedness in facing the real world once they complete their studies.

Both students and lecturers were grateful and excited to have partaken in the first DataFest in South Africa and we look forward to hosting more DataFest sessions in the future. ■

Judging DataFest: Insights from the Other Side of the Table

Michael Sarkis, 2nd Order Solutions

We caught up with Michael Sarkis, a data science manager at 2nd Order Solutions and returning DataFest judge, to hear what keeps him coming back year after year. He shares what judges look for, how they weigh technical rigor against storytelling, and what makes a team's work stand out.

Michael Sarkis is a manager of data science at 2nd Order Solutions. He has a wide range of experience within the consumer credit industry and uses various statistical modeling techniques and approaches. Outside of work, he is a big New England sports fan and enjoys using statistics in his fantasy sports leagues.

Tell me a little bit about yourself.

I'm Michael Sarkis, a data science manager at 2nd Order Solutions, a niche consulting firm in Richmond, Virginia, we refer to as 2OS. I currently do data science work in the consumer and small business credit area, working with a wide range of companies—from fintech start-ups to top-10 US banks. I've worked on a broad range of projects, but almost all of them have involved using data to find insights and create solutions through statistical modeling. Before 2OS, I got my master's degree in statistical sciences from Duke University and an undergraduate degree in statistics and economics from Cornell University.

How did you get involved with being a judge at DataFest, and what keeps you coming back?

2OS is a sponsor of the Duke DataFest, and we always try to send a few employees down to Durham to help consult with the teams and be judges. As an alumnus myself, I was happy to go back down to Duke's campus and take part in the event. I also used to be a teaching assistant for Professor Fisher (who has run the Duke DataFest for the past few years), so I had some familiarity with the sponsoring team. What keeps me coming back is that I find it fun and rewarding to judge the competition and provide an industry perspective that can help future data scientists on their career paths.

What are your top five criteria for judging a DataFest project?

1. Business Insight – Is there clear insight being offered that will actively help the "client" who provided the data?
2. Communication – Are the insights and results effectively communicated in both the slides and verbal presentation?
3. Validity of Analysis – Is the analysis correct, and does it support the points presented?
4. Visualizations – Are the visualizations used in the presentation useful and aesthetically pleasing?
5. Statistical Complexity – What statistical methods are being used and how 'difficult' is the analysis the team did?

How do you weigh the merit and depth of an analysis versus the clarity and interest of the data story?

This is a tough one because both are very important in the competition. However, I do weigh the "clarity and interest of the data story" a decent amount more than the "merit and depth of the analysis." Coming from consulting, I am trying to evaluate the team's work from the client's perspective. A lot of the time in industry, the story and the why of the analysis matters a lot more than the actual process or complexity of the analysis. This is a shift from the academic perspective a lot of students are used to, where the merit and complexity of the method can take the forefront. In my experience, not many clients want to know the ins and outs of the statistics used (just that it is correct) and are more concerned about how they can use the results.

What would you tell DataFest participants is the most important thing to consider when they're putting together their DataFest presentation?

Focus on the insights you are providing the business. The key insights and results are what really matter most from a company's perspective. It is still important to show the approach used, but there is a very limited amount of time and space (3–4 slide limit) with which a team can present their work. A lot of the exploratory data analysis and specific details on methods will likely need to be cut from the final presentation to leave room for developing a compelling story with the few good nuggets of information your team discovered.

Specifically on the data visuals, a couple of well-made and clear graphs/plots is a lot better than a bunch of quickly made ones. Also, try to make a visualization your audience will be able to understand without you explaining it.

What's challenging for you about being a judge?

The biggest challenge for me as a judge is the limited amount of time and information we have to evaluate a team. All the student teams that have presented to me have clearly put in a lot of work, but only so many can be passed to the final round and eventually win some of the awards. As a judge, I hear a team present for four minutes, see their four slides, and then likely get to ask one or two questions. Due to the number of teams, judges, and the timelines, it all makes sense, but it is an added challenge, as each team could likely present for at least 30 minutes on their analyses.

As an industry professional, do you see DataFest as an important component of a student's education in terms of supporting future data science career opportunities?

I think DataFest provides students with a unique opportunity that is hard to find in a classroom. The freeform nature of the analysis plan, the usually very messy real-world data, and the communication aspects of the event are things many data scientists meet in their everyday work. It challenges students to use the skills they have learned and apply them in a collaborative setting to reach a solution.

How is DataFest similar to or different from working as a professional data scientist?

DataFest is very similar to my work as a professional data scientist. As a data science consultant, getting new data from clients, analyzing it, and then making a slide deck to share our insights is a very common project structure. The challenges student teams face with messy data, clarity of analysis, or story building are a lot of the same ones I see in my work. However, the most important similarity is probably the fact that the final output is a presentation. As a consultant, communication is incredibly important and that's where I see the DataFest really shining.

In terms of things that are different, DataFest is a lot more condensed and rushed than most of my projects. It happens over only a weekend instead of a couple of weeks to several months. DataFest is also a freeform analysis problem and, while that is not uncommon, there are projects that have very structured analysis plans with well-defined goals. The last difference is that there isn't a senior person on the project advising. A team of new data scientists wouldn't be placed onto a project by themselves to solve an issue. There would be a manager or adviser working closely with the team to make sure everyone was on the right track and issues they encountered were solved. This difference, however, is a great learning opportunity for the student teams to be the main decision-makers and get experience working independently under a tight deadline.

What advice would you give to a DataFester about how to make the most of their experience?

Don't leave your presentation until the last minute! Make sure you put some time and effort into making it a polished product so your team can put its best foot forward for the judges. You and your team will need to prioritize your time so you have a finished product by the end of the weekend. This means you may not be able to dive as deep as you want or you may be limited to only trying a few statistical methods. Prioritization is a very important skill in data science work, and this will be useful experience.

Don't worry about having the most statistically complicated methodology. A linear regression that offers a useful insight is a lot more powerful than haphazardly throwing more complicated methods at the problem. The second piece to this is to make sure you really understand the methods/models your team used, both for your own knowledge and because it's a lot easier to present a topic you know really well.

What's one of the best team names you've ever heard?

There have been a lot of good ones, but a recent one I liked was the Standard Deviants because I love a good stats pun. ■

We invited several DataFest hosts from universities across the country to share their experiences. From how the event got started to how students tackled the challenge, they reflect on memorable moments, lessons learned, and the ways DataFest builds community and sparks collaboration.

A Point of View: Five Colleges

Ben Baumer

By the late 2000s, a small group of isolated statisticians was dining monthly over Chinese food in Amherst, Massachusetts, and discussing their teaching, coordinating course offerings, offering each other mentorship, and strategizing. The group included George Cobb and Janice Gifford of Mount Holyoke College, Amy Wagaman of Amherst College, Katherine Halvorsen and Nicholas Horton of Smith College, and Michael Lavine of the University of Massachusetts, among others.

"At that point, there were between zero and two statisticians at each of the four colleges and all existed within larger structures (typically a mathematics department)," Horton recalled. "To some extent, it provided a 'department' for those faculty."

You would have been hard-pressed to find an official mention of data science anywhere nearby. Fast forward to the present and the University of Massachusetts boasts a Center for Data Science and Artificial Intelligence, Mount Holyoke College offers an interdisciplinary bachelor's degree in data science, Amherst College now has a standalone department of statistics, and Smith College is approaching a decennial review for its undergraduate major in statistics and data science (one of the first at a liberal arts college). With certainty, the Five Colleges have benefited from the rising tide of interest in statistics and data science that spans international waters. But the Five College group was able to exploit another often-overlooked catalyst: The ASA Five College DataFest.

This is the story of how a small data analysis competition for undergraduate students fueled the transformative growth in statistics and data science at one of the nation's premier higher education consortiums. And while sufficient tinder existed (as it did in many places), it was the combination of DataFest and its emergent corporate sponsor, MassMutual, that ignited the fire.

In 2009, the group (shortly thereafter organized officially as the Five College Statistics Program) received a grant from the National Science Foundation to hire a series of Five College postdocs, who would be based at UMass but teach at several of the other colleges over a three-year period. The third, and final, person they hired was Andrew Bray, a freshly minted statistician from the University of California at Los Angeles who was involved in the organization of the first two DataFests as a graduate student. Bray brought the idea of a Five College DataFest to this group.

Bray knew how to organize the event but needed a little bit of money (most of which would buy T-shirts and food) to run it. A cold call from Bray to MassMutual, a Fortune 500 life insurance company with a large office in nearby Springfield but no obvious ties to statistics and data science, proved to be fateful. It turned out a high-ranking executive at MassMutual named Gareth Ross lived in Amherst and had millions of dollars to spend on jump starting data science at the company. He used a strategy that would later be described as "*Moneyball* for insurance." Ross agreed to sponsor the inaugural 2014 ASA Five College DataFest and dispatched his chief data scientist, Sears Merritt, to put eyes on the event. What he saw changed the course of, at a minimum, several people's careers, including my own.

"I remember the energy," recalled Dana Udwin, a Smith student at the time who took part in the event. "The format is conducive to real creativity and discovery. You're not in a classroom and there is no right answer. You're with your buddies handling a mess of data."

Udwin's five-person team won the Best in Show prize at the first Five College DataFest, not by impressing the judges with sophisticated

statistical models, but by using their data wrangling skills to quickly overcome challenges that waylaid their competitors, creating an innovative and comprehensible data graphic, and telling a coherent story. It was these meta data science skills that impressed the judges and Merritt.

"It taught me the value of a 'commit and go' mentality," Sara Stoudt, another member of Udwin's team, reflected. "I definitely had more of a perfectionist tendency in those days, so it was helpful to experience a time-boxed task that was on an even shorter scale than something like a class project."

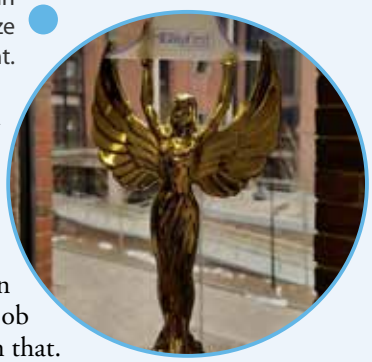
A few months after the first DataFest, Ross and Merritt commissioned Bray and me to develop a prospectus for what would become the MassMutual Data Science Development Program, a three-year training program in which recent graduates would pursue a master's degree at UMass while getting paid to do data science work at MassMutual's new dedicated office in Amherst. Udwin, who was the first person hired into the program, saw the DataFest-to-MassMutual connection become "an incredible pipeline of local talent." Over the next half decade, scores of Five College graduates passed through the program, and MassMutual fleshed out its data science team.

"MassMutual was founded in the 1800s (!!!)," Udwin noted. "They've amassed a lot of domain expertise and a trove of data. Melding that foundation with the fresh, innovative science that is showcased at DataFest was a recipe for success."

But there weren't enough students graduating with these skills to meet MassMutual's needs. They wanted to know what we would need to generate more such graduates. More faculty was the clear response¾we knew what to teach and how to teach it; we just needed more faculty to help us do it. What followed was a series of grants from MassMutual to various members of the Five Colleges to fund faculty positions: a $2 million grant for "Women in Data Science" to Smith and Mount Holyoke; another million-dollar grant to Smith for runway money for two tenure-track positions; and a $15 million grant to UMass for the Center for Data Science.

Enrollments and participation at DataFest continued to swell. After roughly 40 students took part in that first Five College DataFest, registrations grew to more than 200 by 2019, making the Five College event one of the largest in the country.

For all these reasons, the Five Colleges and MassMutual were recognized by the ASA with the 2019 SPAIG Award, which celebrates partnerships between industry, academia, and government. While DataFest certainly played a large role in that nomination letter, its instrumental role in growing the community around data science in the Pioneer Valley has not been celebrated publicly until now.



A large trophy spends the year at the institution whose students win the Best in Show prize at the previous event.

While DataFest can serve as an on-ramp to a job in industry, it's more than that. Both Stoudt and Udwin went on to complete PhDs in (bio)statistics, and DataFest bolstered their experience in graduate school.

"DataFest provides an environment where you learn to borrow strength across a team," Stoudt says, "and I do think that the willingness to lean on team members and not being afraid to fail publicly were both essential ingredients to my time in graduate school."

DataFest continues to be a focus of the Five College Statistics Program and the statistics and data science community in the area. The related Meetup group, founded in 2013, now boasts more than 600 members, many of whom have served as judges or VIP consultants at DataFest. Much to the amusement of students and faculty alike, a large trophy spends the year at the institution whose students won the Best in Show prize at the previous event.

The added faculty lines made the beefed-up presence of statistics and data science across the Five Colleges possible. Yet, while MassMutual and the Meetup group continue to support networking events that involve food and drink, an unfortunate casualty of the growth of statistics and data science in the Five Colleges has been the meals over Chinese food. With so many (now non-isolated) statisticians around, the Five College Statistics Program has moved to a representative governing model, and meetings are held over Zoom for convenience.

For Horton, the meals were "a place to vent frustrations and an opportunity to share ideas and proposals to leverage our connections" that are sorely missed.

No one could have predicted the downstream consequences of the decision to host that first DataFest. And while much has changed in the last decade, Stoudt encourages students to take the plunge.

"Think of it as getting the experience of learning on the job without having to get a foot in the door first," she advises. "Giving up one weekend to go through the full investigative data process on a real-world data set is more than a fair trade. Use that experience to land that next experience!" ■

A Point of View: Purdue University

Mark Daniel Ward

Fulya Gökalp Yavuz

Some of the first STAT-LLC participants from fall 2014 and spring 2015

Purdue University has coordinated a DataFest every year since March of 2015, except for the spring of 2020 due to the COVID-19 pandemic. In March of 2015, we had only five teams competing. The 20 sophomore undergraduate students that year were all participating in the Statistics Living Learning Community, called STAT-LLC. Those students worked on research projects throughout their sophomore year. They took courses together, and they also lived on the same floor of the same residence hall. Their work was supported by the National Science Foundation. Since the students knew each other well, they had a special sense of camaraderie during the event.

The next year, in April of 2016, Purdue had 19 teams of students sign up, and we knew this event would never be the same. The DataFest fever had taken hold! In the early days, we served food for all the students throughout the weekend. Also, many of the students used a computing node on campus, which had a remarkable (at the time) 384 GB of RAM, 50 TB of disk space, and 24 processing cores.

Things have changed a lot at Purdue University. The students now routinely work on a computing cluster with 1,000 computing nodes and 128 processing cores per node. Many students are also involved in a program called The Data Mine, which is on track to have more than 2,000 graduate and undergraduate students participating during the upcoming 2025–2026 academic year.

This growth parallels the broader, nationwide interest in statistics and data science. As programs in data-driven areas have exploded in popularity, the ASA DataFest is now held all over the world. Beyond statistics, students come from broad backgrounds, including engineering, business, liberal arts, and humanities.

One lesson learned is the earlier we advertise the event, and the more broadly we publicize the event on campus, the more students we attract. This may seem obvious, but this small insight can help people organizing an ASA DataFest event for the first time. There are advantages to starting small and growing the event as resources and administrative personnel for the program grow.

ASA DataFest also occurs during the last few weeks of the semester on many campuses, so it is important to plan the dates and locations in such a way that avoids spring break, midterm exams, campus alumni events, and campus visit days for prospective students.

A few more observations:

- Purdue students have used a wide variety of technologies over the years. One team of students made a 3D visualization of the globe, showing how the data affected countries worldwide.
- Graduate students love to serve as judges and interact with the undergraduate students, who take advantage of several tools for their analysis. This year, Purdue had almost 20 judges for our DataFest.
- It can be insightful to visit other colleges and see how one's neighbors and peers are coordinating their ASA DataFest. Colleagues from Miami University in Oxford, Ohio, have invited us to see their event on multiple occasions over the years. In this way, the ASA DataFest can lead to deeper friendships and collaborations across institutional boundaries.

Most of all, ASA DataFest is a joyous weekend for students to work with an intensity that is comparable to real-world, real-time data analysis projects that arise in industry. Students spread their wings, learn teambuilding skills, and hone their communication skills. They learn to focus on the most important, essential aspects of analysis and statistical insight. If you have not (yet) organized an ASA DataFest in your neck of the woods, we highly encourage you to do so! ■

A Point of View: Pomona College

Jo Hardin

From left: Connor Ford, Amy Watt, Madelyn Andersen, Ethan Ashby, and Adam Rees explain how they discovered it took three days for members of a rugby team to fully recover from jetlag.

I first heard about DataFest in the fall of 2011, when Rob Gould shared his inaugural experiences of the first University of California at Los Angeles DataFest during a meeting of the ASA Southern California Chapter. Although Pomona College is 50 miles from UCLA, I immediately asked Gould how I could involve my own students in the UCLA DataFest. Gould was working on growing his local DataFest and welcomed my students to join.

Every year since 2012 (except the pandemic years when DataFest happened virtually), I have brought Pomona students to UCLA to take part in the competition. In some years, my students have done quite well. In other years, we haven't even made it to the final round. Every single year, my students have learned a ton and enjoyed their experience.

The DataFest competition at UCLA happens over 48 hours. The students stay up quite late working together in a single space. They are provided with food and consultants who help them work through any hurdles. Which is to say, being 50 miles from the competition complicates taking part for the Pomona students. So that they can have the full experience, we put them up in a hotel close to the UCLA campus both Friday and Saturday nights. Another complication is transportation to the event. Usually, there are enough students with cars to make it work, but we sometimes must rely on ride-share for transportation.

One of my favorite analyses was by a group of students (Madelyn Andersen, Amy Watt, Connor Ford, Adam Rees, Ethan Ashby) who won the award for Best Use of External Data in 2019. The data set had biometrics information from the Canadian National Women's Rugby Team. The Pomona students were able to look through social media to find the exact date the team flew to an international match and correlate the recovery time to jet lag, recognizing that the players needed three days postflight for a full recovery.

My students regularly tell me that the DataFest line on their CV garners the most attention in interviews for jobs or graduate schools. Even students who didn't win a prize are asked about their experience. They talk about learning how to problem solve in real time and about parsing down a large query into small pieces, something they have typically not done in class assignments.

The last few years have brought large language models into the competition. In some ways, students have become less frustrated because they can easily get help with their coding. In other ways, students struggle more if they haven't developed their own core coding skills, independent of ChatGPT. And in yet a third way, the DataFest challenge is the same because the students need to use creative approaches to tackle the problem—something AI is unable to do.

I am a firm believer in collaboration. Almost all my own research is done in collaboration, and I encourage my students to work in groups on many assignments. DataFest is a perfect venue for students to fully grasp the importance of collaboration to solve real problems. Many perspectives facilitate having a strong approach to problem solving, and having different technical skills allows the group to try a range of analyses. DataFest is fun, low stakes, and yet both collaborative and competitive, giving students a taste of what statistics and data science can be at their best. ■

A Point of View: Duke University

Maria Tackett

Alexander Fisher

Students from across Duke and The University of North Carolina at Chapel Hill take part in Duke DataFest in March of 2024.

Real data is big. Real data is messy. The students in today's classrooms will go on to work with real data daily in their professional and personal lives. Despite this, it is challenging to provide many opportunities for students to work on open-ended applied problems as complex as the ones they will experience outside the classroom. While internships and research assistantships can help fill this void, the stakes are high and the opportunities are few. At Duke DataFest, we strive to engage as many students as possible, from multiple institutions and a wide variety of backgrounds, in real data analysis in a low-stakes environment.

Our task begins with outreach several months before the event, as we send email announcements to students in statistics, computer science, and other STEM programs. We make in-person announcements in the introductory statistics courses to encourage students early in the curriculum to take part. We also send email announcements to a variety of affinity groups across campus to expand our reach to students who may be interested in data-driven work but not currently enrolled in statistics and related courses.

Additionally, we recruit faculty at other local institutions who are interested in being faculty sponsors for Duke DataFest. The faculty sponsors not only play an integral role in recruiting students at their home institutions but also volunteer time at the DataFest event. Our hope is that by being physically present at DataFest, the faculty sponsor evokes an *esprit de corps* among the students who do not attend Duke.

Our main partner institution, The University of North Carolina at Chapel Hill, has also reimbursed local student travel to our campus to encourage participation.

Fundamentally, DataFest is a team event; however, not all students who wish to take part sign up with a team. We give students the option to sign up individually and be assigned a group. We sort students in groups based on a survey that asks students about their anticipated engagement with the event and their coding language of choice. We expect these two criteria will help reduce team conflict and promote collaborative computing. In fact, some of these groups go on to win awards at DataFest.

For students who would like to better prepare for the event, we offer optional workshops on topics such as data visualization, regular expressions, and Tableau in partnership with Duke library staff. Additionally, pre-recorded workshop sessions are hosted on the Duke DataFest website to help students new to data science learn the basics of wrangling data in R and Python.

Once DataFest begins, our goal is to create an environment in which students can focus on the data task, persevere, and present their findings at the conclusion of the event. We do this, in part, by making it possible for students to be physically present the majority of DataFest weekend. We provide free meals to participants, as well as an abundance of snacks, throughout the event. Moreover, we have a variety of activities between meals—group photos, trivia, guided stretching exercises, and small prize giveaways—to help increase participation.

To streamline onboarding into the competition, we partner with NCShare and the Duke Office of Information Technology to provide all participants containerized versions of RStudio and Python notebooks that can be accessed through a web browser. These containerized environments come pre-loaded with the DataFest data set, documentation, and a suite of data wrangling packages. This enables students to avoid debugging local installation of software and simply focus on DataFest. Students who opt to use their local computing environment download the data set and documentation from a private Box link.

In the past three years, we've had 180 teams take part, representing six schools in the area. In 2025, we had 57 teams made up of 161 attendees participate in Duke DataFest. Of these 57 teams, 31 submitted a final presentation. ■

Up the Creek with Many Possible Paddles: How Students Cope with Being Awash in Data at DataFest

Jessica Karch, Jennifer Noll, James K. L. Hammerman, and Traci Higgins

As big data becomes ubiquitous across multiple sectors, it is increasingly important to create opportunities for students to work with large, authentic, complex data—referred to as LAC data here—during their undergraduate training. Working with LAC data is challenging, as students need to extract meaning from the data and work and think like data scientists. "In a data science situation—especially as a beginner—you often don't know what to do. You have way too much data, and the data you have is confusing. Even though you're not literally in a boat, you are awash in data," write Tim Erickson and Ernest Chen in the introduction to "Awash in Data: Introducing Data Science with Data Moves and CODAP."

To cope with being awash in LAC data, students need to manage complexity and navigate many possibilities. This can lead to feeling overwhelmed, especially as students move from formal classroom contexts to navigating real-world data that has not been cleaned, pre-structured, or connected to a chapter teaching a particular technique.

DataFest presents a unique opportunity for students to engage with LAC data. In many ways, the DataFest challenge

mimics real data science work: The data is messy and complex, with many observations and variables. Also, the task is not prescribed but instead must be constructed around a meaningful problem that can be addressed by the data and that yields solutions or insights the client can act upon.

These authentic data investigations involve a complex workflow involving the following six actions, as Hollylynne Lee and her collaborators describe in their 2022 *Statistics Education Research Journal* article, "Investigating Data Like a Data Scientist: Key Practices and Processes":

1. Framing the problem in relation to the real-world phenomena and broader context and pose investigative question(s)
2. Considering and/or gather data, which may involve examining the data at hand and considering data-gathering methods, the measures, the type of data and how it is structured, sample size, and what questions it can address and whether additional data is needed
3. Processing the data by organizing, structuring, cleaning, and possibly transforming the data or computing new metrics
4. Exploring and visualizing to identify patterns and relationships
5. Considering models
6. Communicating and proposing action to stakeholders

This workflow is typically iterative and nonlinear. For example, a data scientist may realize the current structure of the data does not allow them to answer the problem as formulated and engage in a further round of considering the data, identifying different variables of interest, processing the data to arrive at new metrics, or merging in additional data—which in turn leads to a reformulation of the question.

Over the course of two years at six sites, the research team for our Improving Undergraduate STEM Education: Directorate for STEM Education-funded study interviewed participants from 28 teams approximately one week after DataFest about their approach to the DataFest challenge. We used Lee and collaborators' framework for the data investigation workflow and Erickson's concept of authentic data science as producing an experience of feeling "awash" to explore how teams navigated the complexity of working with LAC data.

Our qualitative analysis of the interview data revealed teams often expressed a sense of being awash in the data early in their investigation as they worked on developing a framing. In traditional, hypothesis-driven statistics, problems are well defined and the data collection is designed to flow from the investigative question(s). In the classroom, the data may be given, but it fits within the context of learning new techniques and students are often given the problem they are expected to solve.

At DataFest, teams interact with very large, messy, complex data sets and must consider how to plot a path through that data to some sort of actionable insight or data product the data donors or judges recognize as meaningful and relevant. To narrow down options and give the investigation focus, teams sought a way to frame the problem and generate investigative questions that could

anchor their work. We found the following three common anchoring strategies that teams used to find this framing:

1. Lean in on contextual or domain knowledge to generate questions that feel meaningful or relevant
2. Formulate a problem space that allow for showcasing of certain technical skills or techniques
3. Adopt a data-driven approach using simple exploration of the data to identify where the data seem to clump or patterns that were interesting or unexpected

Strategy 1. Leveraging contextual knowledge

Students used salient domain knowledge to help frame their investigation. In our first year of data collection in 2023, students were asked to make sense of data from the American Bar Association's online tool that connected clients with lawyers for pro bono legal advice. One team had domain knowledge about incarceration in the United States from a previous course. They used this domain knowledge as an initial way to frame the data, interrogating it for data about incarcerated people. Although this strategy did not pan out (the legal issues were civil, not criminal, so there was minimal data about incarceration), this outside knowledge initially provided some traction, giving their work direction. This helped them overcome the feeling of being overwhelmed by the size and complexity of the data sets.

When this framing was discovered to be at odds with the data available, the team shifted momentarily to a data-driven approach, exploring question categories to see which accounted for the bulk of the data, but then again leaned on contextual knowledge as they reformulated their framing of the problem to focus on understanding characteristics of clients posting under this category.

Strategy 2. Showcasing and/or centering technical skills

Teams narrowed in on a way to frame the problem by taking stock of what computational or statistical tools they were familiar with and considering how applying those tools could lead to a particular framing of the problem space. For example, when one team first explored the data, they immediately began to consider what tools they could use to further explore the text data without doing a lot of time intensive processing. Topic modeling fit these criteria, so the team applied this technique in an exploratory way to gain a better understanding of the data. This analysis produced an unexpected finding. Their questions flowed from this finding and led to a productive framing when they began to discover that certain lawyers had an outsized impact on the data—they drew on their knowledge of "whales" from mobile gaming (Strategy 1) to conceptualize this trend and create a productive framing around investigating these lawyers' answers.

Strategy 3. Leveraging the data themselves

Teams often immersed themselves in the data through exploratory data analysis. In this strategy, students used a variety of techniques to familiarize themselves with the data and see what emerged in a very open way. For example, in Year 2, which focused on data from an online statistics textbook, (see "CourseKata's Experience as a DataFest Donor" on Page 24), one team described feeling overwhelmed by the number of variables in the data set. To ground themselves, the team read the descriptions of the variables in the data definition document and began making graphs through a process of trial and error. The team had a vague sense of what their task was—give feedback to CourseKata—and visualizing the relationships between variables helped them see what stood out and what trends merited further investigation and framing.

There are multiple entry points for teams of students to engage with LAC data during authentic data science experiences. When students feel "awash" without a way to frame the investigation, they can bootstrap using knowledge that speaks to the contextual domain, that draws on technical tools to process and work with the data, or that employs exploratory data analysis to seek patterns that demand further explanation.

Just as there's no one best way to approach LAC data, DataFest helps students learn that being a data scientist means being flexible and able to employ multiple strategies and sources of knowledge when navigating the full data investigation process. DataFest provides opportunities to learn how to integrate knowledge from computational, statistical, and domain knowledge sources to deepen their data inquiry process. ■

Editor's Note: This material is based on work supported by the National Science Foundation under Grant No. DUE 2216023. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

Professional Opportunities

Professional Opportunity listings plus equal opportunity information are due the 20th of the month two months prior to when the ad is to be published.

These listings and additional information about these ads can be found at *https://careerconnect.amstat.org/jobs.*

To advertise in *Amstat News,* email *advertise@amstat.org*

To find the latest jobs in statistics and data science, visit ASA Career Connect at *https://careerconnect.amstat.org.*

Call for Case Study: CAS Seeks Forecasting Paper from Statisticians Outside Property-Casualty Insurance

The Casualty Actuarial Society (CAS)—the world's largest international organization of property-casualty actuaries—is offering up to $50,000 for a case study, presented as a paper that forecasts future loss payments on past insurance policies using longitudinal data sets provided by CAS.

This opportunity is specifically for statisticians outside the property-casualty industry to bring fresh perspectives to insurance forecasting for reserves. The submission deadline is **October 27, 2025.**

CAS credentials actuaries specializing in non-life and health insurance across sectors including personal and commercial auto, homeowners and commercial property, workers' compensation, and liability coverages.

Submitted papers must include:

- A rationale for the selected modeling technique and evaluation method.
- Explanation of the results.
- Open-source code (in R or Python) with comments explaining the modeling sequence.
    - o Acceptable packages must be vetted by organizations such as CRAN or PyPI.
- A bibliography that provides technical background for a property-casualty audience

For more information, visit *https://www.casact.org/article/new-research-rfp-seeks-insight-researchers-outside-insurance-industry-forecasting-future* or contact Elizabeth Smith, Director of Publications and Research, at *esmith@casact.org* with the subject line "Longitudinal Reserving RFP Proposal."

Missouri

■The Department of Mathematics and Statistics at Missouri University of Science and Technology invites applications for the Havener Endowed Department Chair position, with an anticipated start date of August 2026. The department seeks an exceptionally qualified scholar to provide leadership and vision as the department chair. Apply at *https://hr.mst.edu/careers* (Position #00095735). The review of applications will begin on 10/3/2025.

The University of Missouri is an Equal Opportunity Employer. *https://www.umsystem.edu/ums/hr/eeo*

To request ADA accommodations, please call the Office of Equity & Title IX at (573) 341-7734.

National University of Singapore

■ Assistant, associate, and full professor positions in the Department of Statistics and Data Science at the National University of Singapore invites applications for tenure track and tenured positions in statistics, data science, and related areas, at the assistant professor, associate professor and professor levels. The anticipated start date of these positions is July 2026. Applicants must possess doctorates in their respective fields by the time of appointment.

The National University of Singapore offers internationally competitive salaries, generous research funding, travel support, relocation assistance and other benefits. The department has nearly 40 faculty members and provides a stimulating research environment.

At the assistant professor position, we are interested in applicants with strong research potential. At the associate and full professor positions, we are interested in applicants with a good track record in research, teaching and leadership.

Submit a cover letter, curriculum vitae, research and teaching statements, and at least three letters of recommendation, uploaded by the letter writers, to *mathjobs.org*.

More information about the university and the department can be found at *www.nus.edu.sg* and *www.stat.nus.edu.sg.* ■

Top Ten Rejected Professional Development Course Ideas

Amstat News continues its entertaining offering by ASA Executive Director Ron Wasserstein, who delivers a special Top 10—one that aired during a recent episode of *Practical Significance*. Wasserstein views the Joint Statistical Meetings as a means to professional growth and says, "JSM is loaded with continuing education and professional development opportunities, and you should engage with them if possible."

Wasserstein

"But, alas, not *all* the ideas for professional development we receive are the best," admits Wasserstein. "So, in our relentless efforts to improve the lives of our podcast listeners, here are the 'Top Ten Rejected Professional Development Course Ideas.' Don't sign up for courses like these!"

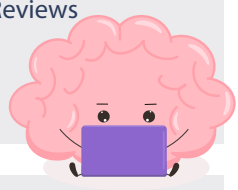
To listen to the *Practical Significance* podcast, visit *https://magazine.amstat.org/podcast-2*.

10

The Lone Wolf Statistician: Collaboration Is for Losers

09

Writing Your Own Yelp Reviews



08

Using Transparencies for Your Presentations (featuring a mini-course on FAX usage)

07

Speaking with Your Back to the Audience, and Other Skills for the Shy Statistician

06

MCMC on Your TI-83: Sure, You Can!

05

Making Your Convenience Samples Even More Convenient

04

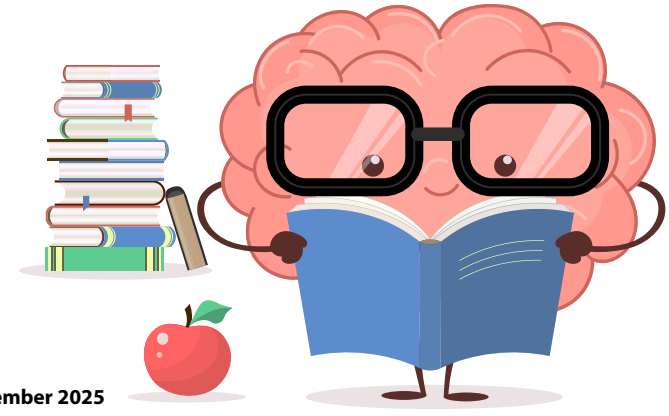
Six Easy Ways to Make Your AI-Generated Work Look Like You Did It

03

How to Have Opinions on Things You Know Nothing About

02

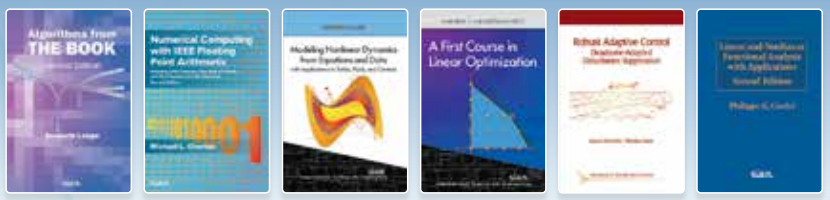
Keeping Those Healthy Impulses in Check



#01

Living the Ron Wasserstein Way!

New books from SIAM



Algorithms from THE BOOK
Second Edition
Kenneth Lange

Most books on algorithms are narrowly focused on a single field of application. This unique book cuts across discipline boundaries, exposing readers to the most successful algorithms from a variety of fields. Since publication of the first edition of *Algorithms from THE BOOK*, the number of new algorithms has swelled exponentially, with the fields of neural net modeling and natural language processing leading the way. These developments warranted the addition of a new chapter on automatic differentiation and its applications to neural net modeling. The second edition also adds worked exercises and introduces new algorithms in existing chapters. In *Algorithms from THE BOOK, Second Edition*, the majority of algorithms are accompanied by Julia code for experimentation, the many classroom-tested exercises at the end of each chapter make the material suitable for use as a textbook, and appendices contain not only background material often missing in undergraduate education but also solutions to selected problems.

2025 • xiv + 343 pages • Softcover • 9781611978384
List $74.00 • SIAM Member $51.80 • OT204

Numerical Computing with IEEE Floating Point Arithmetic
Including One Theorem, One Rule of Thumb, and One Hundred and Six Exercises, Second Edition
Michael L. Overton

This book provides an easily accessible, yet detailed, discussion of computer arithmetic as mandated by the IEEE 754 floating point standard, arguably the most important standard in the computer industry. Although the basic principles of IEEE floating point arithmetic have remained largely unchanged since the first edition of this book was published in 2001, the technology that supports it has changed enormously. Every chapter has been extensively rewritten, and two new chapters have been added: one on computations with higher precision than that mandated by the standard and one on computations with lower precision than was ever contemplated by those who wrote the standard, driven by the massive computational demands of machine learning. It includes many technical details not readily available elsewhere, along with many new exercises.

2025 • xx + 126 pages • Softcover • 9781611978407
List $59.00 • SIAM Member $41.30 • OT205

Nonlinear Spectral Model Reduction for Equations and Data
with Applications to Solids, Fluids, and Controls
George Haller

This concise text presents an introduction to the emerging area of reducing complex nonlinear differential equations or time-resolved data sets to spectral submanifolds (SSMs). SSMs are ubiquitous low-dimensional attracting invariant manifolds that can be constructed systematically, building on the spectral properties of the linear part of a nonlinear system. SSM-based model reduction has a solid mathematical foundation and hence is guaranteed to deliver accurate and predictive reduced-order models under a precise set of assumptions. This book introduces the foundations of SSM theory to the novice reader; reviews recent extensions of classic SSM results for the advanced reader; and illustrates the power of SSM reduction on a large collection of equation- and data-driven applications in fluid mechanics, solid mechanics, and control.

2025 • xii + 151 pages • Hardcover • 9781611978346
List $62.00 • SIAM Member $43.40 • CS34

A First Course in Linear Optimization
Amir Beck and Nili Guttmann-Beck

This self-contained textbook provides the foundations of linear optimization, covering topics in both continuous and discrete linear optimization. It gradually builds the connection between theory, algorithms, and applications so that readers gain a theoretical and algorithmic foundation, familiarity with a variety of applications, and the ability to apply the theory and algorithms to actual problems. To deepen the reader's understanding, the authors provide many applications from diverse areas of applied sciences, such as resource allocation, line fitting, graph coloring, the traveling salesman problem, game theory, and network flows. The book also includes more than 180 exercises, most of them with partial answers and about 70 with complete solutions, as well as a continuous illustration of the theory through examples and exercises. It is intended to be read cover to cover and requires only a first course in linear algebra as a prerequisite.

2025 • x + 370 pages • Softcover • 9781611978292
List $74.00 • SIAM Member $51.80 • CS33

Robust Adaptive Control
Deadzone-Adapted Disturbance Suppression
Iasson Karafyllis and Miroslav Krstic

This book presents a solution to a problem in adaptive control design that had been open for 40 years: robustification to disturbances without compromising asymptotic performance. This original methodology builds on foundational ideas, such as the use of a deadzone in the update law and nonlinear damping in the controller, and advances the tools for and the theory behind designing robust adaptive controllers, thus guaranteeing robustness properties stronger than previously achieved. The authors present all stability notions, old and new, that are useful in adaptive control, provide numerous examples, and contrast their analysis to landmark approaches to robustification of adaptive controllers in prior literature. This book develops the Deadzone-Adapted Disturbance Suppression (DADS) control, illustrates it on the wing rock instability application, and provides ideas for the extension of the control scheme to cases not studied in the book.

2025 • xii + 178 pages • Hardcover • 9781611977421
List $84.00 • SIAM Member $58.80 • DC43

Linear and Nonlinear Functional Analysis with Applications
Second Edition
Philippe G. Ciarlet

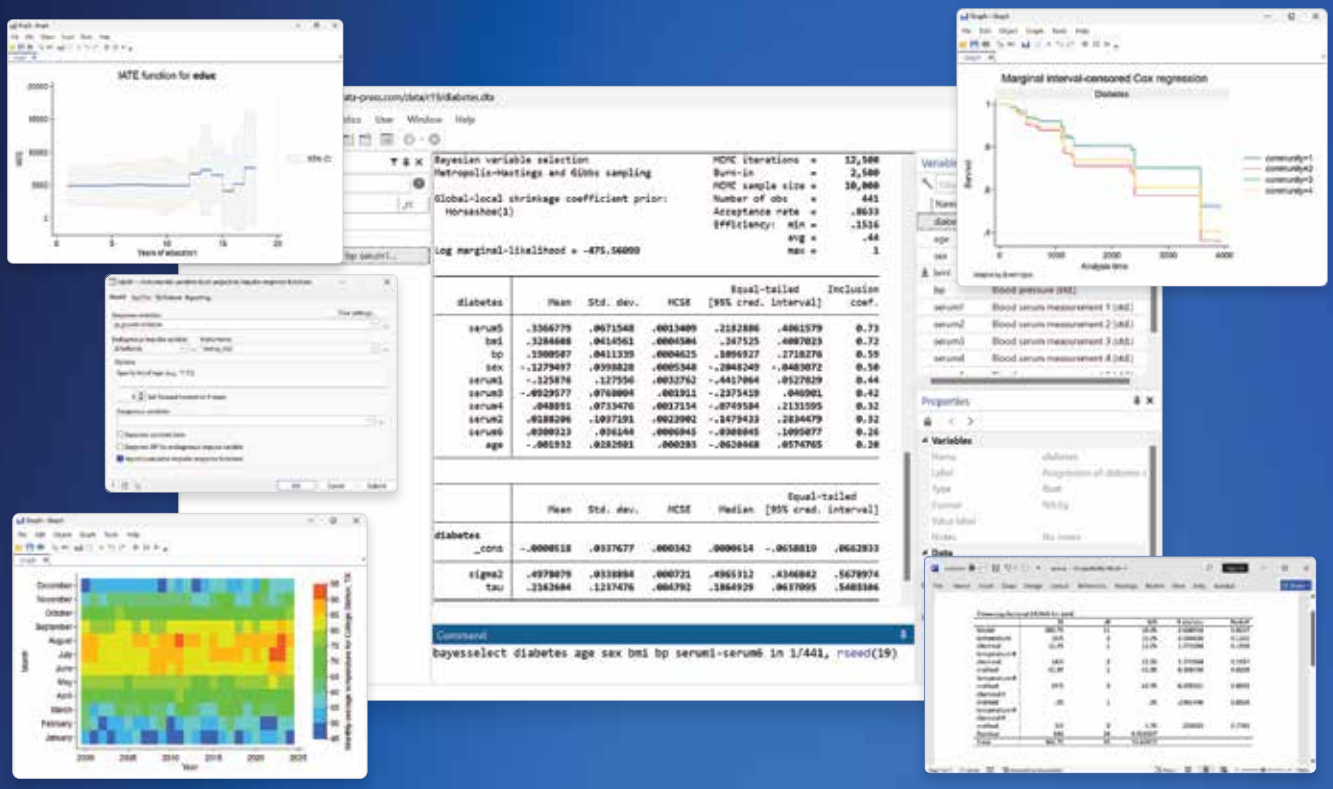
This new, considerably expanded edition covers the fundamentals of linear and nonlinear functional analysis, including distribution theory, harmonic analysis, differential geometry, the calculus of variations, and degree theory. Numerous applications are included, especially to linear and nonlinear partial differential equations and to numerical analysis. All the basic theorems are provided with complete and detailed proofs. The author has added more than 450 pages of new material and more than 210 problems. Two entirely new chapters, one on locally convex spaces and distribution theory and the other on the Fourier transform and Calderón–Zygmund singular integral operators, have also been added. In addition, the chapter on the "great theorems" of nonlinear functional analysis has been enlarged and split into two chapters, one on the calculus of variations and the other on degree theory.

2025 • xviii + 1287 pages • Hardcover • 9781611977230
List $114.00 • SIAM Member $79.80 • OT203

siam | Society for Industrial and Applied Mathematics
BOOKSTORE
TO ORDER, visit bookstore.siam.org

STATA

STATISTICS • VISUALIZATION • DATA MANAGEMENT • REPORTING



Celebrating 40 years of trusted, reproducible statistical analysis

See why researchers worldwide rely on Stata for their most important work.

stata.com/amstat-difference