

ADSA Data Science Community Newsletter

Data Science Community Newsletter features journalism, research papers and tools/software for February 10, 2022.

Please let us ([Micaela Parker](#), [Catherine Cramer](#), [Brad Stenger](#), [Laura Norén](#)) know if you have something to add to next week's newsletter. We are grateful for the generous financial support from the [Academic Data Science Alliance](#).

ANNOUNCEMENTS FROM ADSA Last chance to register for the ADSA Spring Meeting.

Dates Event: March 7-9, in Irvine, CA. Registration deadline: 28 February 2022 [Hyatt Regency Newport Beach](#) has a [room block](#). Deadline: 13 February 2022 Check out [the program](#).

TRAVELING FOR WORK: NOW? NOT YET? NEVER AGAIN? [SURVEY]

Take the 3-question survey now. Results will be posted in the next issue.

[Are you traveling for work again? \[Survey\]](#)

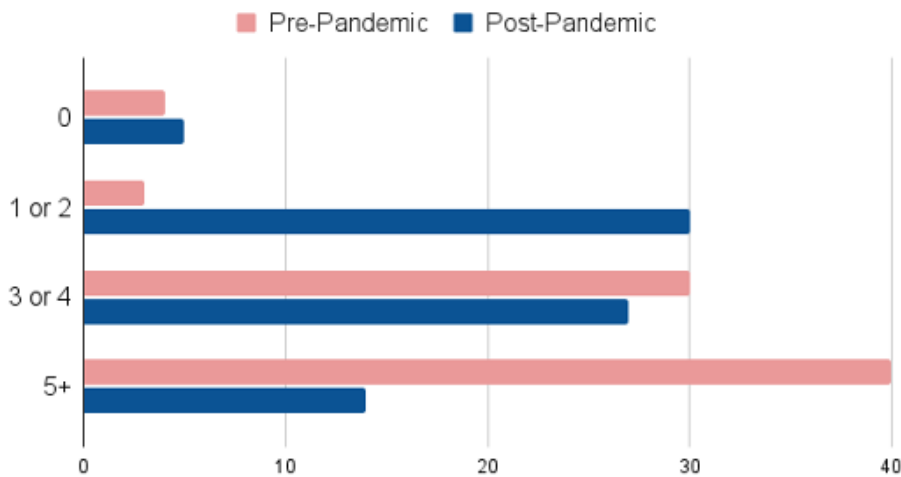
THE LASTING IMPACT OF COVID ON CAMPUS, REAL-TIME SURVEY DATA

We are exploring the lasting impact of the past couple years in a series of surveys.

We never closed [the survey](#) about the number of days people plan to go into campus post-pandemic. It gathered enough new responses that we're running an updated histogram (N=78).

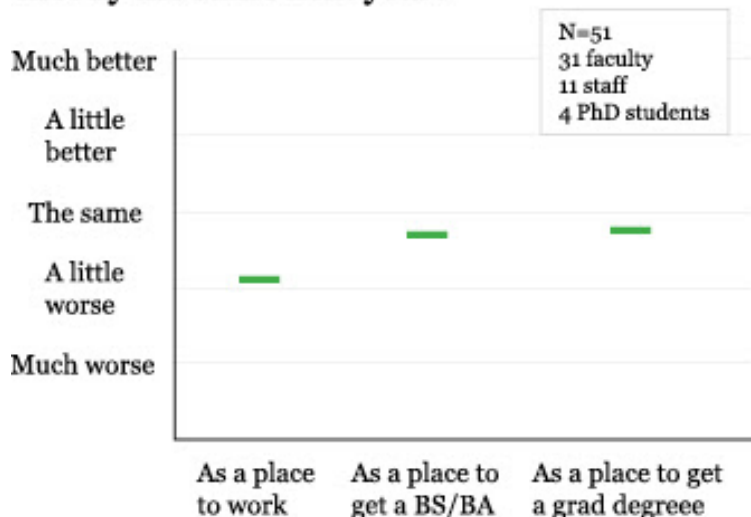
There's an emerging pattern that about one-third of respondents plan to go in 3-4 days per week, another third plan to go in 1-2 days per week, and of the final third, about 2/3rds plan to be in 5 days per week with the remainder not going in at all.

How many days do you plan to spend on campus each week?



In the last issue, we asked about Quality of Campus Life. Here's what we found. Respondents think it's going to be about the same as a place to get a degree, but a little worse as a place to work. Most respondents are faculty and staff who are well positioned to comment on campus as a place to work.

Compared to pre-pandemic do you think your college/university will be better, worse, or mostly the same next year?



Frankly, we need to hear from more students regarding the quality of campus life. Faculty made up more than half of respondents. Staff were reasonably well-represented. Only four students responded, all PhD candidates. We will leave this survey open - please encourage students you know to participate.

[Take the Quality of Campus Life survey](#)

LONG COVID CHARACTERIZED BY DATA GAPS

Much of what we know about long COVID comes from patient-scientists in groups like the [Body](#)

[Politic](#). Their work has been extremely helpful in keeping long COVID part of the conversation, though long COVID and these patient-scientists are far too often marginalized. **Eric Topol** highlights a [new op-ed](#) in *Nature* calling for more research on long COVID in kids and teens. This is a major issue, one that should shape public health policy going forward. Unfortunately, it appears too many of us (especially elected officials) are choosing to let public pandemic fatigue drive policy choices at this point (see/vote in [FT poll](#)).

Featured Events

See the [ADSA Events Page](#) for more details and more opportunities.



[UTSA Draper Data Science Business Plan Competition.](#)

Open to university students globally. Event is virtual & in-person: April 13-15, 2022. Teams can win prizes totaling \$100,000.

Application Deadline: 11 February 2022.

IS DATA SCIENCE A TRADE DEGREE?

"Let's not let Purdue regress into a glorified approximation of a trade school," [wrote](#) professor of Electrical Engineering and Computer Science, **Avi Kak**. Kak points to a pattern we suspect is widespread — computer science and data science are hiring rapidly in academia but humanities departments like Purdue's English department are losing tenure lines and worrying if they can keep their graduate departments alive. We'll get into this in greater detail later this semester, but wanted to foreshadow what will be one of the more important post-pandemic reckonings facing the academy. How important are the arts and humanities to the university as a civic enterprise?

Featured Job



[Postdoc - 2 year](#). NYU Abu Dhabi, Social Research and Public Policy (health). Abu Dhabi, UAE (no remote option).

HARVARD PROFESSORS TAKE OBJECTIONABLE STANCE IN SEXUAL HARASSMENT CASE

In [a case](#) whose particulars cut way too close to home, a new graduate student, **Lilia Kilburn**, eager to work with esteemed anthropology professor **John Comaroff**, felt his behavior — an unwanted kiss on the lips followed by comments about the "corrective rape" she could experience as a lesbian should she venture to Cameroon for fieldwork — was reportable sexual harassment. She reported it. Harvard removed his teaching duties on January 20th. Then...things took a turn (though were quietly deranged before this). Anyway. Thirty-eight highly esteemed senior professors — **Henry Louis Gates, Jr., Sheila Jassanoff, Homi Bhabha, Jill Lepore, Jamaica Kincaid, Randall Kennedy, Stephen Greenblatt, Jennifer Hochschild, Michele Lamont and Paul Farmer** — all signed an open letter supporting Comaroff. Here's a snippet:

HARVARD PROFESSORS TAKE OBJECTIONABLE STANCE IN SEXUAL HARASSMENT CASE

In [a case](#) whose particulars cut way too close to home, a new graduate student, **Lilia Kilburn**, eager to work with esteemed anthropology professor **John Comaroff**, felt his behavior — an unwanted kiss on the lips followed by comments about the "corrective rape" she could experience as a lesbian should she venture to Cameroon for fieldwork — was reportable sexual harassment. She reported it. Harvard removed his teaching duties on January 20th. Then...things took a turn (though were quietly deranged before this). Anyway. Thirty-eight highly esteemed senior professors — **Henry Louis Gates, Jr., Sheila Jassanoff, Homi Bhabha, Jill Lepore, Jamaica Kincaid, Randall Kennedy, Stephen Greenblatt, Jennifer Hochschild, Michele Lamont and Paul Farmer** — all signed an open letter supporting Comaroff. Here's a snippet:

"We the undersigned know John Comaroff to be an excellent colleague, advisor and committed university citizen who has for five decades trained and advised hundreds of Ph.D. students of diverse backgrounds, who have subsequently become leaders in universities across the world. We are dismayed by Harvard's sanctions against him and concerned about its effects on our ability to advise our own students."

This argument — he's such a good guy! What will happen to Harvard if we trust the word of an incoming grad student over our buddy who's been here for so much longer? — is quite common amongst friends and colleagues of those accused of sexual harassment. The accused didn't harass them, so there's a cognitive bias toward assuming the accused couldn't harass anyone. The response to this letter by the Title IX office was that three students, not one, had reported him and that the 38 esteemed signatories to the letter supporting Comaroff didn't have the full story. (Um, right. Of course they didn't. Shouldn't each one of them have known why it's a terrible, terrible idea to victim shame at all, but especially when you don't even know the whole story????)

But let's not be too kind to the Title IX office. The [legal filing](#) revealed a breathtaking violation of Kilburn's privacy, "Harvard obtained Ms. Kilburn's private therapy records without her consent and disclosed them to Professor Comaroff." I'll let that sit there like the fetid putrescence of procedural violence it is.

There is a lesson to be learned here for senior faculty. It goes like this.

Believe the complainants.

Even if the complaint is against someone you really like, someone who is married (lol — since when did being married have anything to do with harassment?), someone who has not had complaints against them before, the complainant needs to be taken seriously. To all the faculty who signed that letter: I'm

sure you know that the odds against complainants being taken seriously have a chilling effect. But...let's be honest...the 38 signatories appear to know this very well. What other explanation can there be for deciding to publicly shame anyone for filing a Title IX complaint in good faith. Vile. Faculty, you need to do better or get out of the way.

Seventy-three professors have now signed [a letter](#) expressing "strong opposition" to the actions of the first 38. Twitter [has been vocal](#).

Featured Job



Neuroscience needs data scientists. [Fellowship](#) at the Interface of Data & Neuroscience (Scientist I position - Allen Institute, Seattle).

SAT GOES ALL-DIGITAL. AND WHO REALLY CARES?

The SAT has announced it will be [an all-digital exam](#) starting in 2023 outside the US and 2024 in the US. It will be shorter — 2 hours instead of 3 — and less relevant as fewer universities require standardized test results. When I lie awake at night between 3 and 5:30, I sometimes wonder if this type of "digital transformation" coming so long after the point where it should have happened is another tally in the "irrelevant" column associated with bachelor's degrees overall. Why can't more students get trained without having to move to expensive cities and pay for room & board? Do people need bachelors degrees to make it in the US? The answer has been a resounding "yes" for decades. The problem now is that many people don't feel they can ever "make it" — home prices are jumping far faster than wages; gig jobs are replacing those that came with steady salaries and modest benefits; climate change is unlikely to make anything cheaper or easier. Reasonable people may prefer not to put themselves into heaps of student loan debt just to tilt at windmills.

Featured Job

See the [ADSA Jobs Page](#) for more opportunities.



[Admin. Associate Director](#). University of Wisconsin-Madison American Family Insurance Data Science Institute. Madison, WI (Some remote work possible).

OPTIMIZATION ALGORITHM FOR BUILDING TECH

Wei Lu, University of Michigan professor of mechanical engineering is the corresponding author of a new *Nature* paper that demonstrates a [building technology algorithm](#). He was joined by colleagues at **Northeastern University**. The algorithm is designed to account for active and passive loads on built structures, material qualities, structural design elements, and make optimizations faster. Using a computationally intensive non-gradient optimization approach is justified here. Why? Because it's being used in even more resource intensive decisions about how to repair or replace bridges, dams, and highways and to optimize manufacturing processes that will impact the production of thousands of parts. I especially like this as an example of excellent data science that is *not* artificial intelligence. Much can be accomplished by an algorithm based on the best human knowledge.

VERMONT COULD INVENTORY ALL STATE GOVERNMENT AI APPLICATIONS

There's [a bill](#) in front of the Vermont Assembly that would create an AI commission. One of its first tasks would be to inventory all applications of AI — from robotic process automation and chatbots to predictive algorithms used in criminal justice sentencing — and then to weigh in on the equity and civic suitability of the use cases. As with all such legislation, it's unclear if it will pass.

There are privacy bills up for debate in many states during this legislative session, but I haven't seen any signs that **Washington State**, **Minnesota**, or any of the other states are going to be able to pass their bills.

IS ACADEMIC FREEDOM THREATENED IN STATE SCHOOLS?

If you have been dismayed at the way professors at the **University of Florida** have [faced repercussions](#) for exercising their academic freedom or autonomy in general, you may be wondering which universities have the [strongest boundaries](#) between university and state.

“**Michigan**, **California** and **Minnesota** continue as the states with the strongest judicial recognition of constitutional autonomy” of the university system from the oversight of state legislators and other authorities. Mask mandates, for instance, can be made at these public schools much the way private schools get to do it: autonomously, without meddling from a governor named DeSantis (for example). FYI: The [University of Minnesota Press](#) has been publishing some great non-fiction work lately.

ELSEVIER FINGERPRINTING ARTICLE PDFS

Ah, academic publishing, what fresh ridiculousness have you set up during the last two years while everyone was legitimately stressed out about ::waves hands::? Turns out **Elsevier** is embedding unique digital fingerprints in the metadata of each downloaded pdf, along with timestamps. This will identify where that particular pdf was originally downloaded (which institution) and when. There's [code presented](#) on Twitter by [@json_dirs](#) that will remove the uniquely identifying metadata, should you be so brash as to want to read and share pdfs that are not trackable.

WE'RE SO BURNED OUT IT'S EVEN EVIDENT IN OUR EMOJIS

Researchers at the **University of Michigan** used **GitHub** data in which ~5% of users include emojis to see if they could [predict worker dropout](#). They attributed the 3x higher dropout rates they found to burnout. (It is possible that dropout could be due to something other than burnout, of course, even at scale.)

CULTURAL KNOWLEDGE CRITICAL FOR LARGE LANGUAGE MODELS

Language and culture are layered, complex, and do not map cleanly onto national borders (which were often determined by colonizers). This is not news. However, this reality seems to have been misunderstood by many of the groups seeking to build large language models. A [new paper](#) by a team at the **Allen Institute** (now [hiring](#)) at the **University of Washington**, the **University of Michigan**, and

the **University of New Mexico** beautifully demonstrates that "there is no such thing as a general purpose corpus" for large language models. Let's get to work adding support for local vernacular languages, slang (old and new), and speech impediments.

PRIVACY IS STILL THE WRONG LENS

In my opinion, privacy has never been the most productive lens for thinking about how to establish a reasonable balance of power between individuals and those who are capable of gleaning insights from data generated by or about them. **Alicia Solow-Niederman** has a [new paper](#) that carefully, coherently explains why rules established to protect a user's privacy by giving them a few controls over certain kinds of data (health data, educational data, etc) are poorly matched to the inference economy.

"Machine learning provides pathways to use data and make probabilistic predictions—inferences—that are inadequately addressed by the current regime. For one, seemingly innocuous or irrelevant data can generate machine learning insights, making it impossible for an individual to anticipate what kinds of data warrant protection. Moreover, it is possible to aggregate myriad individuals' data within machine learning models, identify patterns, and then apply the patterns to make inferences about other people who may or may not be part of the original data set. The inferential pathways created by such models shift away from 'your' data, and towards a new category of 'information that might be about you.' And because our law assumes that privacy is about personal, identifiable information, we miss the privacy interests implicated when aggregated data that is neither personal nor identifiable can be used to make inferences about you, me, and others."

At **University of California, Berkeley**, researchers realized that machine learning models are generally rewarded for overfitting to what is common in ways that may end up disadvantaging underrepresented students. To address this they [forcibly redesigned](#) the school's approach to student performance prediction, weighting less common patterns more heavily. It worked, but had an overall accuracy cost. (That seems right. Humans in college should be exploring new routines, priorities, subjects, activities, foods, music, art, etc. beyond common patterns.)

DOUBLING OF NSF BUDGET? GREEN CARDS FOR STEM PHDS?

The *America COMPETES Act* could possibly [more than double](#) the NSF budget. As written, [the legislation would](#) "effectively stapl[e] a green card to the doctoral diploma of every international student in STEM." This is generally a good thing for America and for foreign STEM PhDs. However, this could have the unintended consequence of keeping graduate student salaries quite low. The labor market will likely "price in" the value of the green card. **Georgia State University** economist **Paula Stephan** argues this is already happening in the postdoc labor market.

AUSTRALIAN UNIVERSITIES: MATCH RESEARCH TO INDUSTRY NEEDS

This feels both inevitable and shocking. "Universities will be directly challenged to respond to research opportunities that are specific, measurable, focused on industry needs and driven by explicit national goals," the [University Research Commercialisation Plan](#), released yesterday, states."

TWO MORE MEMBERS OF GOOGLE'S ETHICAL AI TEAM LEFT LAST WEEK

Alex Hanna and **Dylan Baker** left **Google** to join **Timnit Gebru**, a former colleague of theirs on the Ethical AI team who has started a new AI research non-profit called **DAIR** (Distributed Artificial Intelligence Research). [Hanna](#) wrote a public exit letter: "I am quitting because I'm tired."

ERIC LANDER RESIGNS OVER BULLYING BEHAVIOR

In another departure, Biden's science advisor and head of the Office of Science Technology and

Policy, **Eric Lander**, [resigned](#) over "demeaning" bullying behavior towards staff. He also didn't look like such a great collaborator in the **Walter Isaacson's** 2021 book about the discovery of CRISPR. (See: [The Lessons of Lander](#) by *Science* Editor-in-Chief **Holden Thorp**. Questions were raised before Lander's appointment)

IRS DROPS PLANS TO USE ID.ME FACIAL RECOGNITION

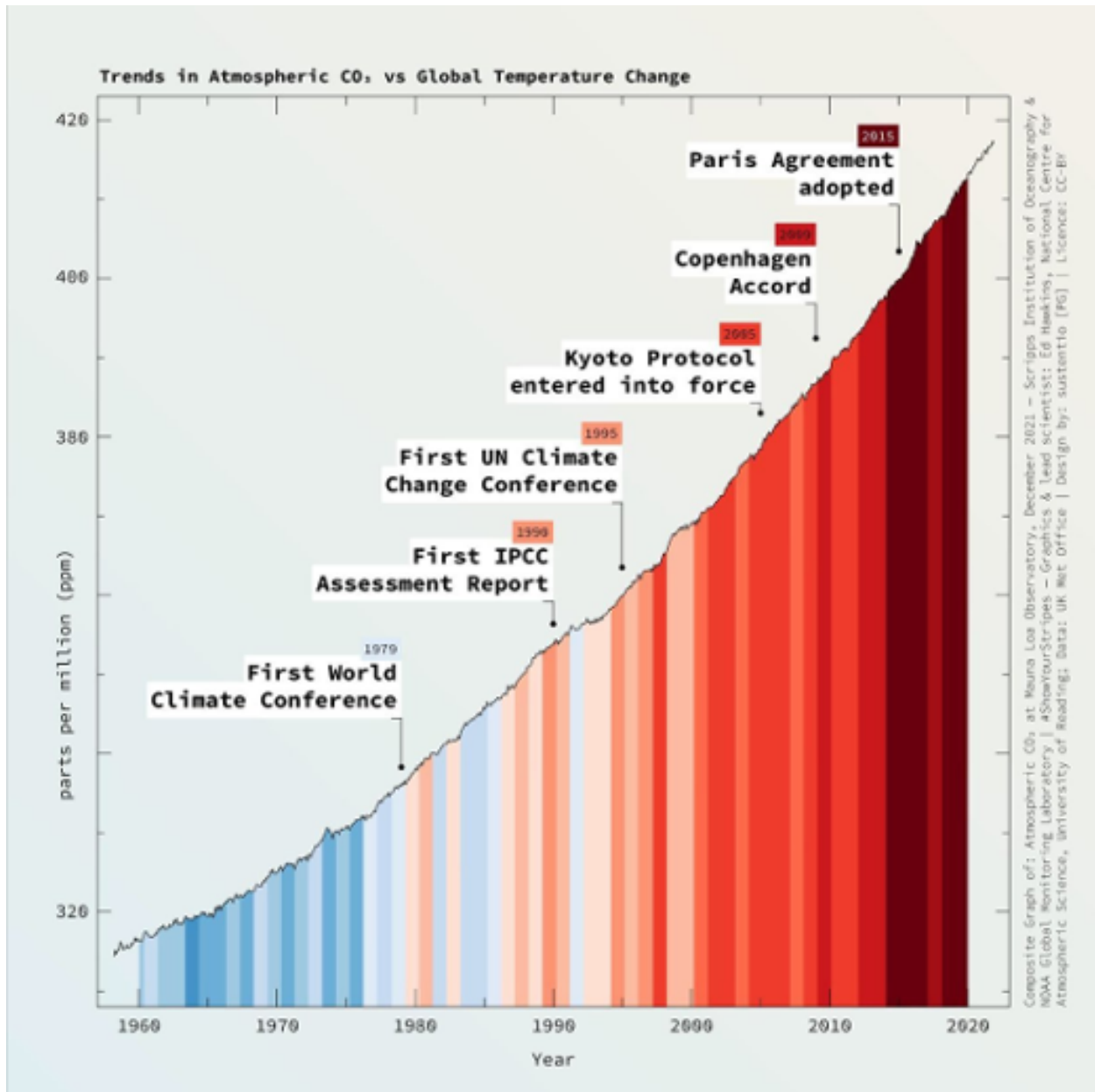
Last week we wrote that the **IRS** would use **ID.me** to help combat fraud during this tax season, which is underway. The agency received swift backlash and has [halted plans](#) to use ID.me. Sadly, this means there is little standing in the way of fraudsters.

NEW PROGRAMS, FOLLOW THE MONEY

Click through to access [a structured spreadsheet](#) of New Programs and money moving around in academic data science.

DATA VISUALIZATION OF THE WEEK

Twitter, Jason Hickel from February 8, 2022



You can experience this graphic as a [3D installation](#) by [Ed Hawkins](#) at the [Spencer Museum of Art](#) in Lawrence, Kansas.

Deadlines

Studies/Surveys

[Want to help the @pulitzercenter understand better how to support journalists in algorithmic accountability reporting?](#)

"Here's a survey they're running" via [Nick Diakopoulos](#).

Conferences

[cfp - 2nd ACM Symposium on Computer Science and Law](#)

"The 2nd ACM Symposium on Computer Science and Law is soliciting submissions of both original research papers and systematizations of knowledge (SoKs). The latter are papers that evaluate, systematize, and contextualize existing knowledge, often providing important new viewpoints, challenging long-held beliefs, or devising useful taxonomies." Deadline for submissions is March 15.

Education Opportunities

[Our 2022 ASAPbio Fellows program is now open for applications!](#)

"A program that will provide you with tools and skills to drive discussions about #preprints & represent #ASAPbio. Applications close March 25."

Tools & Resources

[This Wikipedia of Algebraic Geometry Will Forever Be Incomplete. That's the Point.](#)

Columbia University, Columbia News from February 2, 2022

"**Columbia** professor **Johan de Jong** has spent the last 15 years gathering the foundational theorems of algebraic geometry in one place. His creation, the *Stacks Project*, offers a new model for organizing and visualizing mathematical information."

[Researchers use tiny magnetic swirls to generate true random numbers](#)

Brown University, News from Brown from February 7, 2022

The research "reveals previously unexplored dynamics of single skyrmions, the researchers say. Discovered around a half-decade ago, skyrmions have sparked interest in physics as a path toward next-generation computing devices that take advantage of the magnetic properties of particles — a field known as spintronics."

About Us: The Data Science Community Newsletter was founded in 2015 in the Moore-Sloan Data Science Environment at NYU's Center for Data Science. We continue to be supported by the Gordon and Betty Moore Foundation and the Alfred P. Sloan Foundation through the [Academic Data Science Alliance](#). Our archive of newsletters is at <https://academicdatascience.org/resources/newsletter>. Our mailing address is [1037 NE 65th St #316; Seattle, WA 98115](#).