

ADSA Data Science Community Newsletter

Data Science Community Newsletter features journalism, research papers and tools/software for May 5, 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](#).

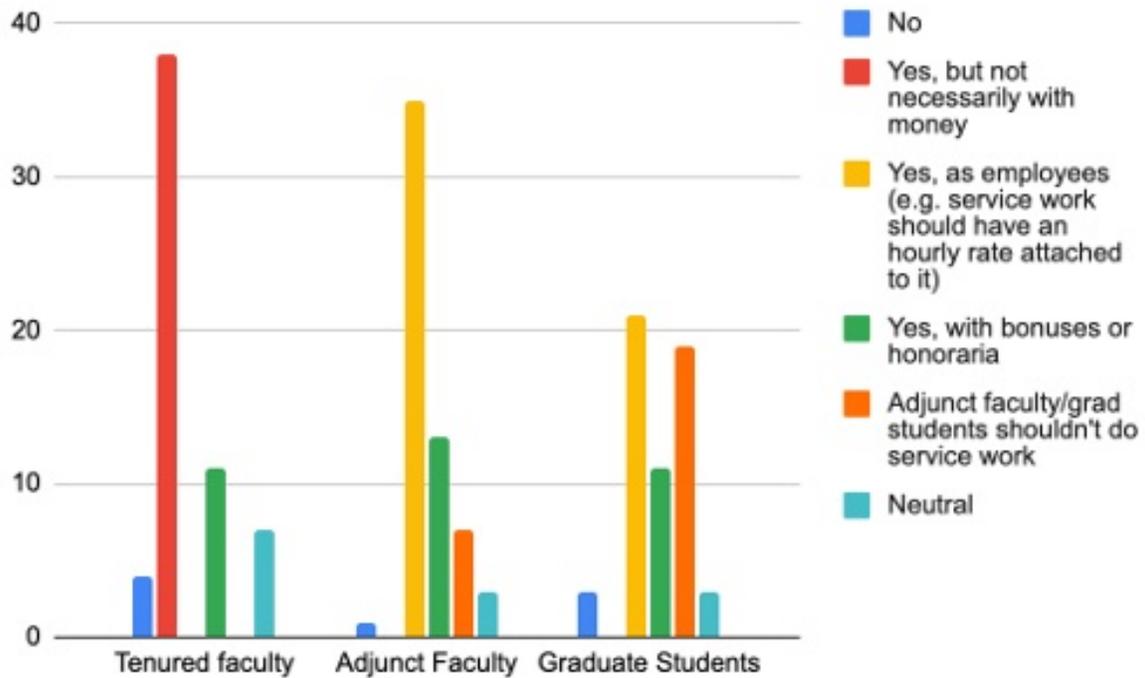
COST OF FREE LABOR? [PULSE POLL RESULTS]

There is general agreement among our 65 respondents that compensation for service work is reasonable.

Tenured faculty were our largest pool of respondents. They believe that the best way to compensate themselves for service work is not with money but with things like reduced teaching loads and more graduate students/postdocs. For tenure track faculty only, though.

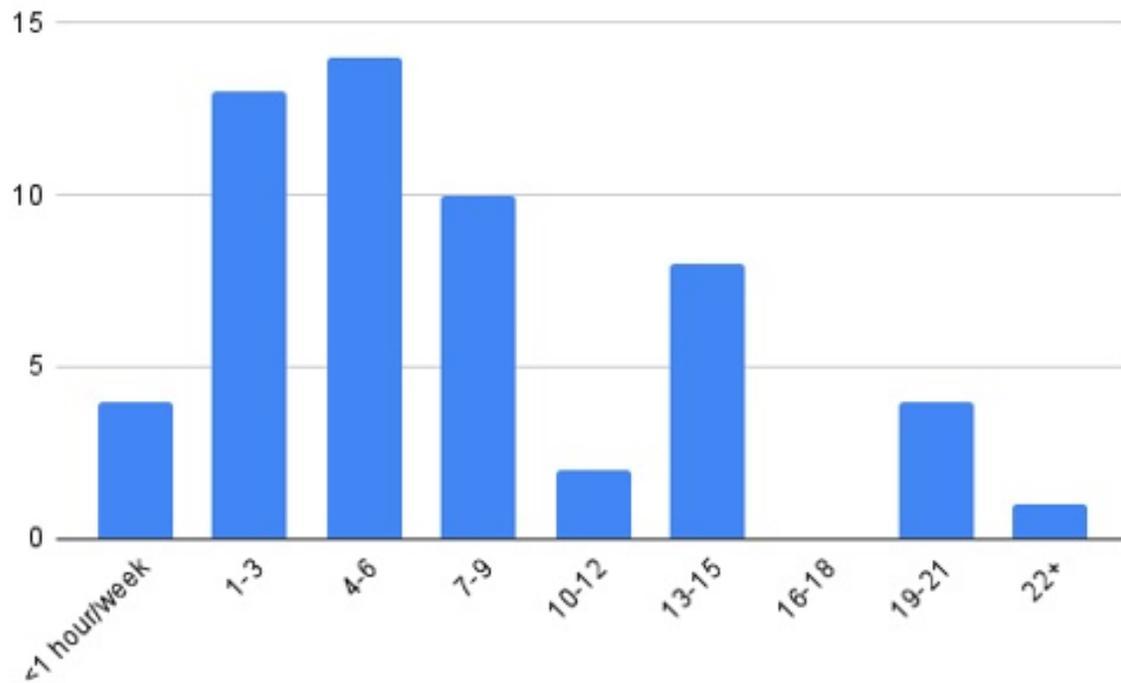
Adjuncts should get hourly rates and grad students either shouldn't do more than 1-2 hours of service work per week, should be compensated on an hourly basis, or maybe given honoraria.

Should academics be compensated for service work?



In terms of how much service work is getting done — there's a wide dispersion from less than an hour per week all the way up to 22+ hours/week. Almost 50% are sitting somewhere between half a day and a day per week.

Hours of service work per week?



There weren't enough adjunct and grad student respondents to get a reliable picture of how much service work they are doing. The vast majority of respondents are currently working at R1 and R2 universities.

SPONSORED CONTENT

A sponsored content banner for the University of Virginia School of Data Science. The banner features a dark blue background with white text. On the left and right sides, there are images of two men sitting at a desk in a meeting room, looking at a laptop. The central text reads: 'JOIN OUR TEAM' followed by a list of roles: 'GENERAL FACULTY', 'TENURE TRACK FACULTY', 'LEADERSHIP FACULTY', 'POST-DOCS', and 'RESEARCH SCIENTISTS'. Below the list is a white button with the text 'LEARN MORE'. At the bottom of the banner, the University of Virginia logo and 'SCHOOL OF DATA SCIENCE' are displayed.

JOIN OUR TEAM

- GENERAL FACULTY
- TENURE TRACK FACULTY
- LEADERSHIP FACULTY
- POST-DOCS
- RESEARCH SCIENTISTS

[LEARN MORE](#)

UNIVERSITY OF VIRGINIA | SCHOOL OF DATA SCIENCE

PHIL BOURNE DESIGNS A DATA SCIENCE INSTITUTE AT UVA

Dean of the **University of Virginia, School of Data Science**, **Phil Bourne**, explained the organizing principles he and his faculty are using to structure their School for Data Science. This kind of public statement of organizational principles is relatively rare and we encourage you to take the 3 minutes to [read the whole post](#).

The TL;DR version goes like this: any data science institute will be interdisciplinary and relatively new. Fights for funding and efforts to generate recognition on campus and beyond are known challenges for any new campus center. This has some similarity to the problem facing a for-profit startup, except that with data science centers, the lean startup or minimum viable product narrow focus is not intellectually or socially inclusive and thus would be disastrous to the communities of thinkers and impact-makers most centers want to foster. Still, putting meaningful boundaries on what these centers are and aren't is necessary. Where *does* interdisciplinarity end? And if literally everyone is welcome, how are all those people brought into common understanding and a sense of purposeful unity?

Bourne sets up two crosscutting organizational structures. The first "define[s] data science as 4 domains – Systems, Design, Analytics and Value, all applied to Practice across many disciplines (+1) – our so-called **4+1 model**" and the second outlines a subset of research areas that UVA SDS will prioritize: "biomedicine, environment, democracy, business, digital humanities, cybersecurity." He is then quick to point out that the research areas should NOT be understood as traditional disciplines because that would undermine the first 4+1 structure.

Because UVA has been one of the largest movers in the US data science space – it's one of the only universities to establish a separate School where others have been more willing to establish centers and institutes – their structural choices will be watched closely. As you can see elsewhere in this newsletter (see "abominations" we like), boundary fights between data science and computer science are getting spicier as more students, researchers, and funders are drawn to them.

Oh! And [UVA School of Data Science is hiring](#).

WHEN A SOCIOLOGIST HEADS WHITE HOUSE OSTP

We have covered the **White House's** personnel churn at the **Office of Science, Technology, and Policy**, which has typically been headed by a physicist though **President Biden** appointed **Eric Lander**, a biologist focused on genetics who had to step down and has been followed by now-Temporary Director, sociologist/anthropologist of science and technology, **Alondra Nelson**. As a sociologist, I see only upsides to the nation stemming from Nelson's intellectual background. **Slate** writer and **Arizona State University** postdoc **Nicholas Weller** [agrees](#). For those of us who care about the near-term future of OSTP leadership, **Nancy Scola** has an [excellent long read](#) on Nelson in **POLITICO**, an outlet that's been doing a lot of excellent reporting these days.

SPONSORED CONTENT



Welcome message delivered by Dr. Elwood L. Robinson with a keynote by Dr. Desmond Upton Patton.

NYU SCHOOL OF MEDICINE ATTEMPTS TO HIRE PROFESSOR RECENTLY DISMISSED FOR MISCONDUCT

New York University's Grossman School of Medicine is trying to hire professor **David Sabatini**, the son of the school's former chair (Dr. Sabatini the elder was chair from 1972 to 2011), was recently dismissed from **MIT**, **Whitehead Cancer Institute**, and **HHMI** for research and sexual misconduct. NYU alumni, postdocs, and some faculty members have staged walkouts, signed letters opposing the hire, and given interviews with numerous news outlets from **Vice** to **Science** to **The New York Times**. Opposition only grew stronger following a closed-door meeting in which Grossman leadership explained that they would like to proceed with hiring Sabatini the younger. They also revealed the identity of the woman and former member of Sabatini's lab whose complaint formed the basis of the case against him, a big no-no.

Yesterday, NYU and Sabatini **agreed** that he should no longer be considered for the position. Sabatini wrote, "I will eventually be vindicated and able to return to my research." Elsewhere, Nobel Prize Winner **Jennifer Doudna** gave **an interview** about how hard it is for women to collaborate in science.

WHOLE FOODS GETS CREEPY

Whole Foods has started live-piloting surveillance systems (palm readers, lots of cameras running facial recognition) that allow shoppers to gather items and be charged remotely without going through a checkout line. **People** on Twitter were seriously creeped out: "it is anathema to every semblance of decency."

COLUMBIA JUKES THE US NEWS STATS – FOLLOW UP

One of our favorite statisticians – **Andrew Gelman** – reviewed/rehashed his **Columbia University** colleague **Michael Thaddeus's** claims that Columbia has been juking the US News and World Report stats for years. Thaddeus originally looked at the main undergrad ranking, which he found to be dubious. More recently, he looked into the ranking for the engineering school and discovered major issues there, especially for Columbia's online Masters program, which has been ranked first for quite some time. Such a **fun read**, unless you are part of Columbia's administration.

Featured Job

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



Research Data Scientist, School of Law, University of Virginia, Arthur J. Morris Law Library, Legal Data Lab. Charlottesville, VA.

THE ACADEMIC "ABOMINATIONS" WE LIKE

UC Berkeley has **rescinded** an enrollment block that only allowed students majoring in EECS or intending to major in EECS to take *Discrete Mathematics and Probability Theory*. The enrollment block specifically caused problems for data science majors and, as a key pre-requisite, effectively prevented anyone outside of CS from taking upper division CS courses. Professor **Satish Rao**, who typically teaches the course, said "It was an abomination to treat these students this way."

Could be an unpopular opinion, but writing about this kind of solve-able academic "abomination" feels like a certain kind of return to normal.

EPIDEMIOLOGISTS' ADVICE ON PAR WITH NUCLEAR WAR

Combining the blissful with the macabre, **John Ioannidis** **agrees** with statistician **David Spiegelhalter** who believes epidemiologists need to be prevented from giving public health advice, "Epidemiologists alone making policy inferences can be a recipe for disaster. We can destroy the world more efficiently than nuclear warfare!" Both Ioannidis

and **Katherine Baicker**, dean of the **Harris School of Public Policy** at the **University of Chicago**, prefer to get their advice from economists. (Here, I assert that sociologists, political scientists, and communications scholars may be equally deserving, but I wouldn't have been so quick to dismiss the epidemiologists.)

OXFORD RESEARCHERS MEASURE OXFORD'S IMPACT ON BIODIVERSITY

Oxford University researchers used "data on purchasing, travel bookings, utility bills and other information from the 2018–19 and 2019–20 academic years...the day-to-day running of buildings and transport services; travel (including flights) for students and researchers; construction of laboratories and other buildings; consumption of food and beverages at restaurants and cafeterias; and use of medical supplies and other materials in research labs" to **compute** the biodiversity loss attributable to Oxford. The point wasn't to point fingers at Oxford, but to demonstrate how large institutions that have made commitments to preserving biodiversity can operationalize their measurements.

They found that Oxford is two orders of magnitude less impactful than the major multinational firm **Microsoft**, but bigger than the **London Stock Exchange** and about the same as the entire island nation of **Saint Lucia**. The main contributors were, "the supply chain for research activities (such as for chemicals, medical products, organic tissue and plastics); the supply chain for the day-to-day running of buildings (for paper, information technology and so on); food consumption; electricity consumption; and the supply chain for construction."

DESANTIS LIMITS TENURE; UF STRUGGLES TO FIND A PRESIDENT

In a speech following the bill signing ceremony, **Florida Governor Ron DeSantis** proclaimed, "We need to make sure the faculty are held accountable and make sure they don't just have tenure forever without having any type of ways to hold them accountable or evaluate what they're doing."

Elsewhere in Florida's higher education **news**: "A federal appeals court has concluded that a **University of Central Florida** policy targeting 'discriminatory harassment' likely violates the 1st Amendment by chilling the speech of students who want to advocate against abortion and affirmative action."

All of this is happening as the state's flagship public university – **University of Florida** (UF) – tries to recruit a new President. (With Presidential openings at **Columbia**, **NYU**, **MIT**, **Howard**, and other major east coast campuses, the competition for talent will be intense.) Finding a new UF President, **according to Gainesville Sun** columnist **Carl Ramey**, will require locating a leader willing to tolerate a legal environment featuring "new laws that: 1) require Florida universities to survey faculty and students to assess "political diversity" (to help counter liberal bias); 2) impose new, campus-wide restrictions on instructional materials relating to race; 3) prescribe how civic education should be taught; 4) mandate online posting of course syllabuses (to better monitor "stale" political "ideolog[ies]"); and 5) subject existing tenure reviews to stricter, state-developed standards (to ensure that faculty are more

"accountable" and less driven by "intellectual orthodoxy")." University leaders are not accustomed to serving state governors' offices quite so directly.

UNLV PRESIDENT COMMISSIONS CHATBOT DOPPELGANGER OF HIMSELF

University of Nevada, Las Vegas president, **Keith Whitfield**, has [bizarrely] **created** a chatbot doppelganger of himself in an attempt to "talk" to each of the 30,000 UNLV students. In particular, Whitfield thinks the chatbot will be "especially beneficial for first-generation and economically disadvantaged students." This does not feel like a great implementation of inclusivity.

COVID SILVER LINING: OPEN SCIENCE GROWS, ACCELERATES

Jacqueline Lane and **Hila Lifshitz-Assaf**, both of **Harvard**, **wrote** an important piece for the **Brookings Institution** identifying a positive contribution of the COVID pandemic – major advances in the use of open science, open publishing, and open data. They highlighted four key impacts:

- + Opening the ivory tower: "During the COVID-19 crisis...almost a third of open access article readers were non-academic"
- + Interdisciplinarity increased: "open access encouraged new collaborations, particularly cross-disciplinary collaborations" between social scientists and medical scientists
- + Rapid sharing: using open data and tools like Google translate, teams around the globe could share information and collaborate faster than ever
- + Social science accelerated adoption of existing open access repositories (e.g. SSRN) and a new social science tracker on GitHub links to "300+ in-progress research projects"

One early career computer scientist at the **University of Colorado Boulder** who happens to be a DSCN reader and supporter of open science, **Daniel Larremore**, **won** the **National Science Foundation's** Alan T. Waterman Award for early-career research, recognition for his extensive work on COVID and computational epidemiology.

DNA SEQUENCING RACES

In what is being called the **sequencing Olympics**, **Stanford** scientists have been light-heartedly competing to sequence genes faster than their colleagues at **Rady Children's Hospital** in San Diego. The current time to beat is less than ten hours. Speed isn't everything, of course, but these teams are getting 100% accuracy over SNPs and about 88% accuracy over insertions and deletions. Pretty damn good.

ALGOS BIASED AGAINST BLACKS, EVEN WITH DECENT TRAINING DATA

The **City of Pittsburgh** has been **using an algorithm** to help screen children and families reported to child welfare services. The problem is, the tool appears to reflect bias against black families: " If the tool had acted on its own to screen in a comparable rate of calls, it would have recommended that two-thirds of Black children be investigated, compared with about half of all other children reported." Human intervention by social workers tends to override these mistakes, but the impact of spending time investigating the wrong cases can mean children in need don't get

attention fast enough (or at all) and can mean some families that wouldn't have attracted attention go through anxiety-provoking reviews.

In [another study](#) German and American researchers looked at racial bias in medical data sets. They found that researchers using fMRI and other patient data specifically trained their models on data only from white people, only from black people, and from a mixed pool of whites and blacks. The data trained on white people was unsurprisingly more accurate when used to make predictions about white people. The data from a mixed pool also did better for whites, following other studies investigating racial bias in medical algorithms. The odd thing is that algos trained on black people and then used on black people were less accurate for black people than the white-only data was for white people.

Mona Sloane, a research sociologist at **New York University** who works on AI fairness, seems to be tiring of flagging the social problems that come with AI applications. [Writing](#) in **Nature**, she proposes three principles that should improve her (and our) situation:

- + "Include lived experience" and, when possible, lean on experts
- + "Shift power" or at least be alert to the power dynamics at work
- + "Check AI's assumptions" and remember that guidelines aren't always useful or practical

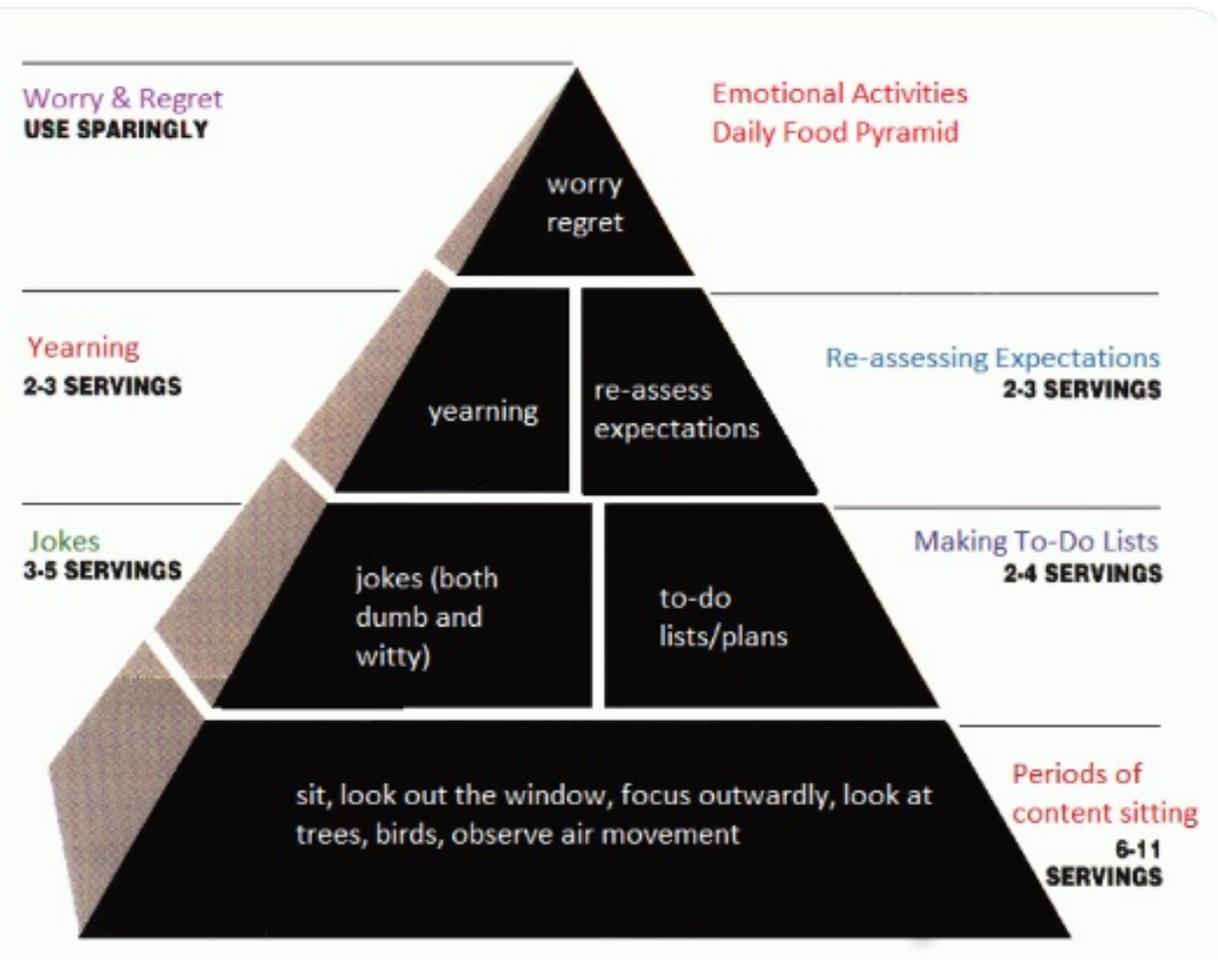
"Vague mandates won't work," according to Sloane, but incorporating frameworks, even relatively simple ones like this, into development will help.

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

Emotional Activities Daily Food Pyramid by [@donmexlar](#) from April 29, 2022



Deadlines

Conferences

The Statistical Inference for Network Models workshop is back at @netsci2022 on July 12 and 13. (Virtual)

"Send your latest work at <https://sinm.network>! Deadline: June 10"

Education Opportunities

We are pleased to announce that the 2022 Diversity Workshop and Mentoring Program will be held during the Joint Statistical Meetings on August 7-10 in Washington, DC.

"The application deadline is May 31."

RFPs

Apply for the #SBDH's #SEEDSgrant program before the June 3 deadline!

"Our mid-range grants (up to \$50K) and large grants (up to \$100K) can support the creation of shareable resources, fund smaller scale research projects + more!"

Tools & Resources

103 Bits of Advice I Wish I Had Known

The Technium blog, Kevin Kelly from April 30, 2022

"About 99% of the time, the right time is right now." And 102 more pearls of wisdom.

PyScript: Python in the browser

Towards Data Science blog, Sophia Yang from May 1, 2022

"Developed by the team from **Anaconda** including **Peter Wang, Fabio Pliger** and **Philipp Rudiger**, PyScript is, as Peter mentioned in his talk, "a system for interleaving Python in HTML (like PHP)." This means you can write and run Python code in HTML, call Javascript libraries in PyScript, and do all your web development in Python."

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](#).

OPT OUT: If you do not want to receive this newsletter, please email brad.stenger@gmail.com with the word 'unsubscribe' in the subject line.

OPT IN: Feel free to forward the Data Science newsletter to colleagues. They can sign up for the newsletter [using this web form](#).

