

ADSA Data Science Community Newsletter

Data Science Community Newsletter features journalism, research papers and tools/software for August 24, 2022.

Please let us ([Micaela Parker](#), [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](#).

ADSA EFFORT — FEEDBACK REQUESTED

With funding from the **National Science Foundation**, ADSA [collaborated with peer organizations](#) to convene a series of small workshops of current and past faculty at US **Tribal Colleges and Universities** (TCUs) and **Historically Black Colleges and Universities** (HBCUs). From these conversations, we generated a draft Exploratory Report to discuss the strategies and resources that might best support student success in Environmental Data Science at TCUs and HBCUs.

We are now [requesting your feedback](#) on the preliminary findings from our conversations *if you have experience at these institutions*.

TREE EXTINCTION

There are 58,497 tree species in existence. Eight hundred eighty-one are native to North America. The endangered species list can and does list plants, but botanists and arborists note that there's a general "plant blindness" bias in which animals — especially mammals and birds — are more likely to make the list than plants. Though scientists haven't overlooked tree species, the studies have been obfuscated by appearing in single articles spread across journals. Until now.

A [new meta-study](#) combed the literature to comprehensively coalesce the situation facing trees. The findings were grim. **Murphy Westwood** of the **Illinois Arboretum** is part of an international team called **Botanic Gardens Conservation International**. She was the lead author in the US, and [she told The Washington Post](#) that "as many as 1 in 6 trees native to the Lower 48 states are in danger of being wiped out." Because trees are the longest lived species and are slated to play a major role in climate change mitigation by absorbing 12% of our carbon emissions, their decline will have an outside impact.

In a [related study](#), **Stephen Polasky** et al. have identified the best ecoregions to target for tree species diversity protection based on the number of species represented and the ease associated with the

conservation efforts: "The largest gains in species conservation come in Melanesia, South and Southeast Asia, the Anatolian peninsula, northern South America and Central America." The US, in fact, has significantly less tree species diversity than these ecoregions.

And on **PBS**, the finale of docu-series *American Outdoors* [visited](#) sustainable farmers **Lise and David Abazs**. Host **Baratunde Thurston** goes to the Abazs' **Round River Farm** along the northern shores of Lake Superior where they grow tree seedlings alongside their rhubarb. The baby trees will likely replace the evergreens that might not survive global warming in the Northwoods forest canopy.

US HAS LOST ITS APPEAL TO CHINESE SCHOLARS

Following the visa restrictions placed on Chinese students before and during COVID, the surveillance of faculty with ties to China that started pre-COVID, and the rise of educational alternatives to US-based options, the number of visas issued to Chinese students has plummeted over the past two years.

The **U.S. State Department** saw a [major decline](#): "in the first six months of 2022, the U.S. issued 31,055 F-1 visas to Chinese nationals, down from 64,261 for the same period in 2019." **BONARD**, a market research firm that tracks international student populations, [says](#) that UK universities have been the biggest beneficiaries.

The long term impact of missing out on attracting foreign students is likely to be fewer entrepreneurs launching careers in the US. According to a new **NBER working paper**, most foreign born entrepreneurs first came to the US on a student visa. The authors also note that immigrant founders account for ~20% of all venture-backed companies over the past 30 years.

SUNY POTSDAM IN DIRE FINANCIAL SITUATION

The **State University of New York at Potsdam** has seen declining enrollment for about a decade, with a major acceleration during COVID. Now it has "put 17 majors, 8 minors, and 2 departments 'on notice' that they could be eliminated", including math, physics, and computer science. In order [to avoid being cut](#), the departments must meet enrollment goals and restructuring requirements (e.g. offering more online courses).

The burnout factor is hitting administrators and faculty extremely hard, "There's a national search for a president, the dean of the arts & sciences school recently resigned, as did the head of the school's diversity and inclusion office, the head of finances, and the provost." While this fate may be limited to a small number of schools, it is important to note that even public schools operate largely under the steam of market forces. Schools that cannot attract a large enough number of students will potentially collapse without significant government intervention.

Other schools, however, are touting [record enrollments](#), including HBCUs and small liberal arts schools.

PEOPLE WHO ARE MOST LIKELY TO DISREGARD SCIENTISTS, THEY THINK THEY KNOW ALL ABOUT SCIENCE

A new study confirms what many of us likely already suspected. People who are most likely to ignore or disregard scientific findings [overestimate their own understanding of science](#). (This is one reason why misinformation is worse than no information. Convincing a person to change their mind is harder than teaching an open mind.)

The study's author **Steve Sloman** recommends asking scientific refuseniks to explain the mechanism they think they know all about in greater detail — *Hey, Bob, how do vaccines work?* This can sometimes lead them to find the gaps in their own knowledge base, though whether they fill those gaps with knowledge derived from the scientific method or they go back to the same poisoned well of knowledge where they got their original poor information was not part of the current study..

FINANCIAL AID OFFICES PERSONALIZE PRICING, RAISING QUESTIONS

A new [article](#) by **Brookings Institution** author **Alex Engler** argues that the most coercive personalized pricing model is not coming from airlines or banks or tech companies. It's coming from colleges and universities as they try to optimize enrollment of high performing students with the least amount of

scholarship support. To some degree, this is old news. Admissions offices have had to predict how many of their admissions offers would be accepted to plan for dorm spaces, classroom staffing, and other basic university operations for decades.

However, "when these algorithms work as intended, they may reduce average per-student scholarship support. Second, it is problematic that these algorithms typically optimize scholarships for yielding students, rather than using scholarships to support student graduation and success. The third concern is the possibility of algorithmic bias, through which subgroups of applicants who appear to the algorithm to be less affected by changes in scholarship funding may be mistreated." In other words, the stated goal of these algorithms is to provide just enough aid to persuade a given student to enroll, and not a penny more. This may mean certain cohorts of students are 'optimized' into large loan packages.

In other news, **President Biden** has **announced** a plan to reduce federal student loan debt by \$10,000 for those making less than \$125,000 with another \$10k knocked off the loan debt of students who received Pell grants. Biden also further extended the loan repayment moratorium until January 2023. For context on the cohort receiving Pell grants, "the **Center for American Progress** [found that] 90% of people who default on student loans had received a Pell grant at some point, suggesting significant financial need in college [...] Given the high percentage of college students who drop out with debt they are unable to pay, it is a matter of national concern that hundreds of higher education institutions may be optimizing scholarships to entice students to attend, rather than succeed or even graduate." President Biden's plan may help some of these students with their debt, but for those that dropped out before completing their degree, the higher earnings associated with getting the credential are likely elusive.

HEALTHY FOOD IS MORE SUSTAINABLE (EXCEPT FOR NUTS, SEAFOOD)

An **epic new study** of **57,000 different food products** confirms what ethical vegans and vegetarians have been saying for decades: foods that are more sustainable are also good for your health. The major exceptions to this correlation are nuts and seafood, both of which are healthy for humans, but are unsustainably produced.

OPEN SOURCE 176 BILLION PARAMETER BLOOM MODEL RELEASED

A new open source AI model similar to GPT-3 (but open source!) called **BLOOM** has been developed and released last month by a team from US-based **HuggingFace**, France-based **GENCI** and France-based **IDRIS**. The model was trained on a larger number of languages than some of its competitors (either **20** or **46?**).

Admittedly, we haven't attempted to use this model yet and are reticent to go overboard with enthusiasm in the absence of evidence. Readers, if you start using the BLOOM model or were involved in developing it, please reach out. We would like to understand how this particular foray into an open source codebase is working.

TOMVIZ

Tomviz is a **more mature open source project**. Tomviz is a software that allows researchers working with electron microscopy at the nanoscale (e.g. proteomics, materials scientists) port images into their standard commercial grade laptops to see and manipulate 3D simulations. This process used to take days, but now it takes minutes. Near real-time interaction on everyday laptops is a major achievement because it enables thousands of researchers. **University of Michigan** corresponding author **Robert Hovden** remarked that with the newest tomviz release, "You can start interpreting and doing science before you're even done with an experiment."

CLIMATE: INSURERS MORE RIGOROUS; MONARCHS MORE SCARCE

As we predicted, insurers are working hard to update their models so they can maintain profitability. **Aon** has partnered with researchers at **UCLA** and **University of California, Merced** to update their current wildfire risk model. Aon representative **Dan Dick** notes that their current wildfire model is more concerned with frequency than severity, which is no longer sufficient to predict losses accurately. This is **everyday basic work** for the insurance industry, and that's the point. The insurance industry will present one of the first major pain points for average citizens as they navigate the climate-changing future.

A [new climate-modeling research study](#) about monarch butterflies is a harbinger of future climate studies. This study fully admitted what previous studies already found: monarch populations are being decimated nearly to the point of collapse. What this new study offers is a map of the small number of counties in which monarchs are predicted to survive (maybe). The hope is that people living there (northern Ohio and southern Michigan) will do all they can to foster monarch-friendly habitat. Plant that milkweed!

IS ART OVER NOW THAT WE HAVE AI?

Two perspectives — [one](#) by commercially successful wedding photographer (7 figures level success) **Pye Jersa** and another by **Alexander Wales** who [focused](#) on the graphic design industry — feel their fields may be threatened by AI. AI is quicker, it's generative, and it's detailed. For many commercial purposes, it can produce what is required, cheaper and faster.

And yet. This particular author disagrees that AI can override the role of artists and art. Creating an aesthetic product or even a generative mimic of a [particular artist's style](#) is not the same as making a piece of art that truly resonates at the social and cultural level the way good art does. There may be fewer wedding photographers. But there will still be art and struggling artists. This shape of controversy — what does it mean to produce an original work of art? — has been cycling through the art world since at least the time of **Rodin**, who was the first sculptor to produce casts that could make the 'same' sculpture twice. A similar panic hit the art world at the advent of photography. [See **Walter Benjamin's** classic essay [The Work of Art in the Age of Mechanical Reproduction](#).]

What I would really like AI to do on the aesthetic front — make all slide decks pixel perfect automagically. How many hours have we all collectively spent polishing pixels in slide decks? Let's automate ourselves away from the drudgery of slide deck polishing, shall we?

UBC CANCELED CS CLASSES 2 WEEKS BEFORE FALL COURSES START

Students at the **University of British Columbia** who had registered for an upper level CS class [received an email](#) 2 weeks before the first class notifying them that their course had been canceled due to lack of faculty. Three other CS classes that would typically have been offered were also unavailable. The root causes are that 1) CS enrollments have increased by 45% since 2018 and 2) the job market for CS PhDs is incredibly competitive. This type of rolling CS brownout is likely to continue, hitting campuses where there's a strong local tech sector particularly hard.

CHEATING HAS NEVER BEEN EASIER FOR CS STUDENTS

As CS classes get bigger and bigger, figuring out how to examine students is getting harder. **University of Massachusetts Amherst** professor **Emery Berger** has [referred](#) to GitHub Co-pilot as "bringing Uzis to a knife fight" when it comes to testing his CS students with code-based exams. Berger notes, "Using Copilot, students can instantly generate code solutions as auto-completions, given just the problem statement or even just the function name." Be careful out there, professors.

BUT, there's another [paper](#) out by **Chad Topaz** et al. that looks at how important getting grades of at least C or higher in intro STEM classes are. The goal of the paper was to assess claims far too often made by faculty and administrators that they have nothing to do with the lower STEM degree completion rates for students from underrepresented backgrounds. The new study found, "The probability of obtaining a STEM degree for a STEM-intending white male student with average academic preparation who receives grades of C or better in all introductory courses is 48%. In contrast, for an otherwise similar URM female student, the probability is merely 35%." The upshot is, yes, current faculty, your teaching and mentoring does matter for under-represented students. But I also thought it was interesting to see how much evidence there is that setting up intro classes to "weed out" students actually works: "If these students receive less than a C in even one introductory STEM course, the probabilities [that they will complete their intended STEM degree] drop to 33% and 21%, respectively." Opinions will differ, but perhaps educators should not be trying to remove "weeds" but instead should be trying to cultivate blossoms.

BACK TO CLASS (MASKLESS)? NEW FINDINGS ON LONG COVID

The illustrious **Eric Topol** put together a neatly summarized **Los Angeles Times op-ed** on what we know about Long COVID. It's not great reading, but it should help you determine where your personal risk threshold is as so many of us (or our kids) head back to class.

- + 6-12.5% of people who have been infected likely have long Covid (for reference, that's at least 10 million Americans)
- + still not sure what causes it
- + don't know how to treat it
- + repeated infections increase risk
- + more likely to hit "30 to 50 years old [who were] previously healthy"
- + "typical symptoms include marked fatigue, exercise intolerance, difficulty breathing, brain fog, muscle pain and weakness, chest pain, headaches and fast heart rate"
- + typical duration is unknown because it lasts so long: a "report from more than 1.25 million people with COVID showed an increased risk of developing brain fog, dementia, seizures and psychosis over *two years*" which is about as long or longer than we've been studying long COVID.

Topol's advice? "Our best chance to prevent long COVID is to not get COVID or to avoid getting it again. That's why it's vital to stay COVID-cautious now, and not capitulate to the notion that we should 'live with COVID.'" Topol explicitly disagrees with the CDC's current guidance.

PSEUDONYMOUS IS NOT PRIVATE

In this week's note about why pseudonymity is not much of a privacy guarantee, two researchers **trained a privacy attack model** to reidentify individuals from deidentified location and behavioral data. They did pretty well, which is sad news for those relying on k-anonymity, especially when k-thresholds are only 10. Authors **Arnaud J. Tournier** and **Yves-Alexandre de Montjoye** were able to reidentify individuals correctly about 80% of the time "in a location dataset of 0.5 million people and 93% within a set of 10 candidates. Similarly, on a grocery shopping dataset of 85,000 individuals, our model correctly identifies 65% and 74% within a set of 10 candidates." Even after they added a bunch of noise to a larger dataset, they were still hitting about 60% reidentification rates.

AUDIO OF THE WEEK [UPCOMING]

If you ever wanted an MST3K version of audio play-by-play regarding over-hyped AI, you may want to hear **Alex Hanna** and **Emily Bender** next week on Wednesday at 1pm Pacific for **Mystery AI Hype Theatre 3000**.

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

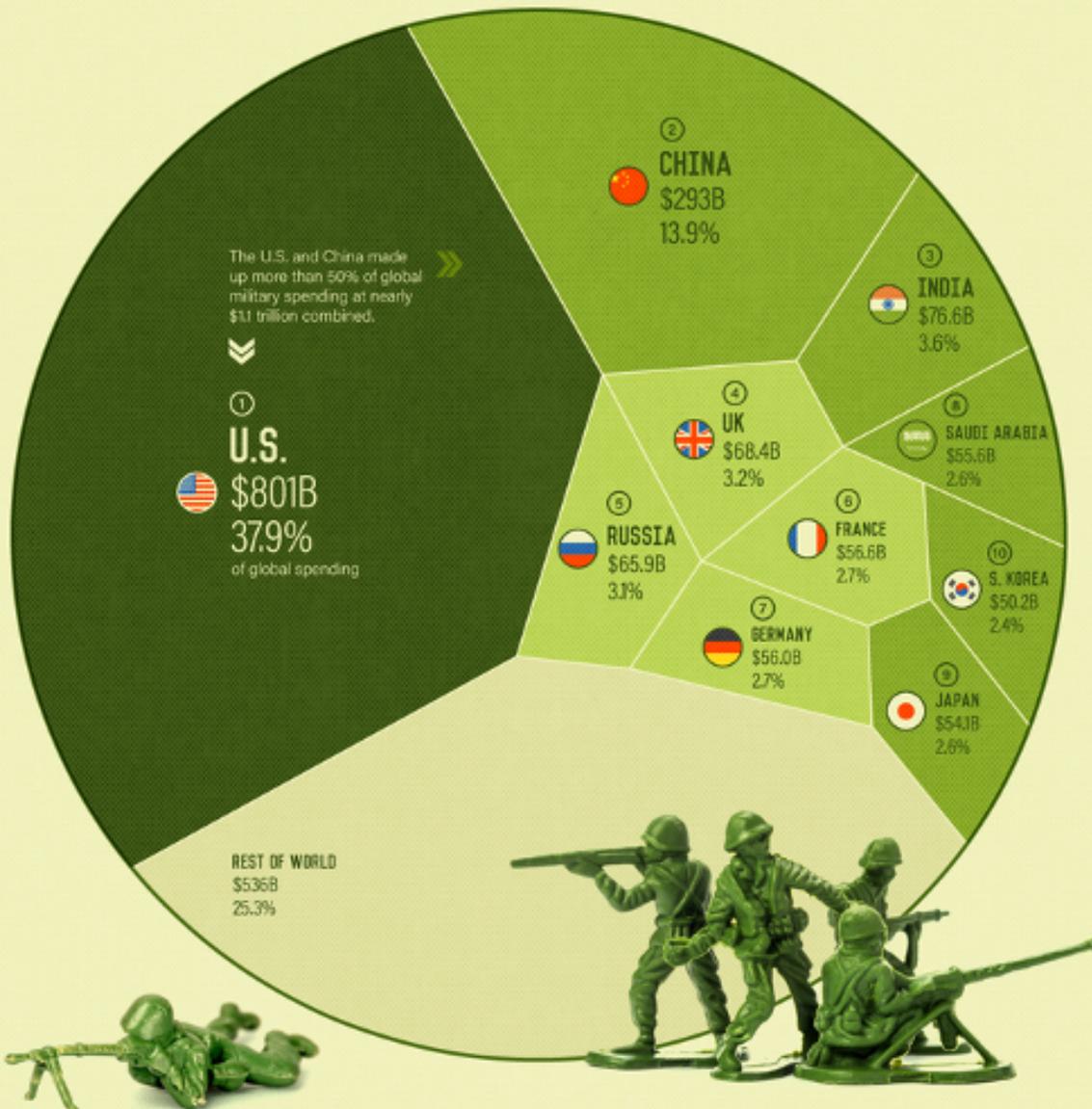
by **Visual Capitalist, Niccolo Conte** from August 18, 2022

THE WORLD'S BIGGEST MILITARY SPENDERS



The world's military expenditure totaled **\$2.1 trillion** in 2021, marking the seventh year in a row of increased global military spending.

WHICH COUNTRIES SPENT THE MOST ON MILITARY IN 2021?



The only country in the top 10 that did not increase its spending in 2021 was Saudi Arabia (-13.9% from 2020).



China's military spending has increased 27 years in a row—by \$148B (100.2%) since 2012.



Russia spent 4.1% of its GDP on its military, with only two other nations in the top 15 having a higher share (Israel 5.2% and Saudi Arabia 6.6%).

Deadlines

Contests/Award

[Ready, set, code! The Open Problems in Single-Cell Analysis competition at #NeurIPS2022 is open](#)
"featuring \$25,000 in prizes & a new multimodal #singlecell timecourse deeply profiling hematopoiesis."

RFPs

[IMLS is accepting FY 2023 proposals for National Leadership Grants for Libraries \(NLG-L\) & the Laura Bush 21st Century Librarian Program \(LB21\).](#)

"Applicants have through September 21, 2022, to submit their two-page preliminary proposals."

Tools & Resources

[@SurrealDB is a document DBMS that just came out of stealth on top of @tikvproject](#)

Twitter, Andy Pavlo from August 22, 2022

"It looks to be written by one guy (@tobiemh) in #rustlang. What I like most about their pretty website are the custom logos for *all* their open-source repos. Impressive."

[ChromeOS 106 gets a Follow site option for desktop RSS \(updated\)](#)

About Chromebooks, Kevin C. Tofel from August 16, 2022

"Earlier this week, I asked the question "Is Google Reader coming back?" That was based on a new "Feed" option in the ChromeOS side panel. I wasn't sure if this upcoming feature was based on RSS or some Google interpretation of it. Now, I'm definitely sure the side panel in Chrome and ChromeOS will use RSS (see update after the post). The latest Dev Channel version of ChromeOS 106 adds a "Follow site" menu option. "

[CSET Launches Tool for Exploration of Global AI Ecosystem](#)

Computing Community Consortium, The CCC Blog from August 18, 2022

"The **Center for Security and Emerging Technology** just launched the Country Activity Tracker (CAT), an interactive tool that portrays countries' level of tech competitiveness and collaboration. Many countries are in the race for Artificial Intelligence (AI), making massive investments in research and infrastructure to remain competitive in the tech sphere."

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](#).
