ADSA Data Science Community Newsletter

Data Science Community Newsletter features journalism, research papers and tools/software for April 7, 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](#).

ADSA FOUNDING MEMBER BOOK NOW AVAILABLE!

GRAD SCHOOL: WHO'S HAPPIEST BY FIELD?

Last week we shared initial findings about whether grad school is worth it for those who hold advanced degrees. To recap, most graduate degree holders are overall satisfied with their decision to get their degrees, feel better off financially, but report that they had to make sacrifices in their personal lives to make space for grad school.

This week we look at which fields are the happiest with their Masters and PhDs. Math and Stats degree holders came out on top. Social scientists were second happiest at both the Masters and

PhD levels. Interestingly, Arts and Humanities degree holders were much happier at the Masters level than the PhD level while natural science degree holders were much happier with PhDs than Masters. Computer science, data science, and engineering degree holders were in the middle of the pack at both the Masters and PhD levels whereas Business/ Finance/ Accounting degree holders were towards the bottom in terms of satisfaction with their degrees in both cases.

Happiness ranking	Masters	PhD
Happiest! 😁	Math/Stats	Math/Stats
2nd happiest 😊	Social Sciences	Social Sciences Natural Sciences (tied for 2nd)
3rd happiest 🙂	CS, Eng, Data Sci Arts & Humanities (tied for 3rd)	
4th happiest		CS, Eng, Data Sci
5th happiest	Business/Finance	Business/Finance
6th happiest	Natural Sciences	Arts & Humanities

We should also note that Masters degree holders reported greater overall satisfaction with their degrees than PhD holders.

We also looked at the correlation between reports of sexual harassment and/or sexual assault during graduate school on the sacrifices reported in personal lives and relationships during grad school. First, overall reports of sexual harassment were lower in Masters degree programs (4.4%) than PhD programs (12.6%) and lower for men (2%) than for women (18%). (There were not enough non-binary people in the sample to report without creating a potential privacy violation.) A chi-square test of independence between sexual harassment and the reported personal life benefits and sacrifices associated with of getting a degree was not signficant, but that's expected with a sample size of N=179.

TRULY COMPLETE, GAPLESS HUMAN GENOME SEQUENCED
When the human genome was "fully" sequenced roughly 20 years ago "fully" meant 92% of the genome. **Now** a full 100% telomere-to-telomere sequence is available. It's possible some of the newly released sequence contains disease-causing genes, but scientists are especially hopeful that the newly available sequences could reveal the secrets of aging. The reason these sections took so long to sequence, according to **National Human Genome Research Institute** director **Eric Green**, was that "there were parts of the human genome that contained DNA that were so repetitive, and so

rugged and so difficult, that any of the methods that were available at that time, they just choked, they couldn't do it".

JOHNS HOPKINS HAS NEW PCR-LIKE COVID SENSOR; SANTA CRUZ'S USHER IS CRITICAL FOR TRACKING GENETIC LINKAGES OF NEW VARIANTS

In an extremely **promising breakthrough** for the COVID-times, **Johns Hopkins University** engineers **Ishan Barman** and **David Gracias** have developed a simple-to-use sensor that can test people for COVID with PCR-like accuracy. The sensor could eventually be part of a wearable that only requires saliva to test for COVID. According to JHU's press office, "During initial testing, the sensor demonstrated 92% accuracy at detecting SARS-COV-2 in saliva samples — comparable to that of PCR tests." Piling on the excitement here, the technology is adaptable to other viral diseases, "The sensor was also highly successful at rapidly determining the presence of other viruses, including H1N1 and Zika." The sensor isn't go-to-market ready yet, but we are hopeful this tech will be cheap enough that daily awareness of viral load could be coming to a toothbrush near you.

Before we're all getting viral load reports as often as we attend to oral hygiene, scientists will be using software **developed** at **University of California, Santa Cruz** to locate new variants on COVID's growing genetic tree. Designed to help local public health agencies, UShER "tells them on which branch of the broader phylogenetic tree their samples fit, therefore telling them which lineages of the virus are present in their community." These are both telling signs of how the next normal is going to go.

There's also a new **database paper** out for the **COVID-19 Data Hub**, a repository of "daily time-series of COVID-19 cases, deaths, recovered people, tests, vaccinations, and hospitalizations, for more than 230 countries, 760 regions, and 12,000 lower-level administrative divisions."

SUPER FISHERY SURVEILLANCE = SURVEYS + SATELLITES

By combining qualitative 'what do humans know?' surveys with quantitative data gathered from satellites, researchers at **Stanford** have begun to identify and track fishing fleets engaging in abusive labor practices and illegal fishing. Surveillance on the open ocean is especially difficult, particularly when there are large monetary incentives for various entities working at fishing ports to sidestep legal compliance that aims to eliminate the market for protected fishes. Led by **Elizabeth Selig**, researchers **combined** "an anonymous survey distributed to experts from seafood companies, research institutions, human rights organizations and governments" to "quantify the degree of certainty around whether particular ports were associated with either labor abuse or IUU fishing." Then, "using machine learning, the team then combined survey responses with satellite-based vessel-tracking data curated by **Global Fishing Watch** to identify higher-risk regions associated with transshipment, where crew and catches are exchanged between vessels, and at sea." The upshot here is that the team combined excellent qualitative research methods with excellent quantitative research to solve an ocean-scale project in an academic lab.

APPRENTICE-LIKE DATA SCIENCE POSITIONS TO UNLOCK EQUITY

Instead of assuming the university's main role in leveling playing fields is by providing courses of study culminating in a degree, **Georgia Tech** has **created apprentice-like positions**. The program, called DataWorks, hires "young adults and train(s) them in entry-level data science skills, such as cleaning and formatting data, using tools ranging from off-the-shelf spreadsheet software to custom scripts in programming languages such as Python." This is a promising new / old approach (apprenticeships predate university-based higher education) that may work better for those unwilling to commit to the financial and opportunity costs of undergraduate degrees.

CREATING "IMMORTAL DATA" FOR SCIENCE

The painstaking process of establishing scientific experimentation protocols to address tightly limited hypotheses ideally results not only in an insightful set of findings, but also in data that might be useful in related studies. The problem is that even with open science data repositories, it still takes time and effort to prepare data so it's useful in an archive. Many scientists don't want to share all of their data while they are still actively using it, but once they have moved on with their investigations there is little incentive to go back, curate it, annotate it, and ensure the metadata is suitable for machines and humans alike. A 2018 project **GO FAIR** based in the Netherlands has been slowly making scientific data 'immortal', joined by efforts at DataCite. If you're **interested in prepping** your lab's data for contribution — perhaps something you can hire a young person to help with in an apprenticeship style model — **Stanford's Center for Expanded Data Annotation and Retrieval** (CEDAR) "runs a platform that generates simplified forms to produce FAIR-compliant metadata."

NIH MANDATES DATA SHARING

After a long period of stakeholder engagement and development, the **National Institutes of Health** has **released** a new **data sharing policy** whose adoption is required for any research funded by the NIH. The policy has two basic stipulations: "First, that researchers share all the scientific data that other teams would need in order to 'validate and replicate' the original research findings. And second, that researchers include a two-page data management plan as part of their application for any NIH funding." Given the depth of NIH pockets and the proliferation of multi-authored projects (*someone's* likely to have NIH funding), this mandate is a significant agent of scientific change.

Many in our readership are already well past the NIH mandate, publishing papers with titles like, **"Share the code, not just the data."**

NSF RELEASES STRATEGIC PLAN

While the NIH is moving on a data sharing plan, the **National Science Foundation** already had such a plan, and made news by releasing its **strategic plan**. Somehow it was released as a pdf, not in a more accessible web-based format. The focus has a strong diversity, equity, and inclusion component. They also want to "Inform applicants of funding decisions in a timely manner." That would be nice.

THREE CREATORS OF AUTHOREA ARE READY FOR ARXIV 2.0

In a **new paper** explaining what's wrong with arXiv — reliance on LaTeX formatting — three creators of the **Authorea** document collaboration tool designed for scientific communities, make an interesting case for what an arXiv 2.0 could look like. Decoupling from LaTeX may also transform arXiv's tech stack into much more than a pdf dump. Clearly, arXiv is already more than that, it is a community of scholarship, but it is provocative to contemplate how many more fields and scholars could be served by the original digital pre-print server if the expected outputs weren't so darn frustrating to format.

WHAT HAPPENS WHEN YOU PAY FOX NEWS VIEWERS TO WATCH CNN

Recall that our survey found that social scientists are pretty happy they got PhDs. Maybe it's because they can do projects that involve paying people to consume (or stop consuming) specific elements of their media diet. Does sound kind of fun.

In the latest example of a social science **media-diet study**, research subjects who reported watching a fair amount of **Fox News** were instead paid to watch **CNN** for 7 hours. After their "treatment" of CNN, **David Broockman** and **Joshua Kalla** found that Fox's coverage omitted certain inconvenient facts in a way that threatens the civil democratic tenets of US political participation. Factually, Fox watchers often had incomplete understandings. For example, "those watching CNN instead of Fox were more likely to know about long COVID & that other countries had controlled the

virus better — info scarcely mentioned on Fox." Without complete knowledge of pandemic facts, voters may not be informed enough to vote in their own best interests.

UC BERKELEY DRILLS BORE HOLE TO EXPLORE GEOTHERMAL HEATING

University of California, Berkeley is **exploring geothermal heat pumps** as part of its clean campus initiative. We are very excited about this energy saving approach that is generally cost effective within ~7 years. In related news, UC Berkeley just **released** a new campus plan without Evans Hall.

UPDATE FROM GEORGIA TECH'S MASSIVE ONLINE CS PROGRAM

One of the earliest experiments in massive online learning at the undergraduate level — **Georgia Tech's** Online Master of Science in Computer Science — has **shared numbers** on enrollments and graduations. Almost 900 students graduated in Fall 2021. That's huge, only slightly smaller than their 12,016 enrollment size would suggest. This strong completion rate for an all-online course hasn't yet represented the wave of the future, but may become more appealing if even a small percentage of students show less willingness to relocate for college in "the next normal."

NEW PAPER ON COMPUTATIONAL ETHICS: ASKS GOOD QUESTIONS

A new paper by authors from 33 institutions in the US and Europe — titled **Computational Ethics** — raises questions in a structured way that could lead to a useful framework in this field. "We propose that understanding the human moral mind using the tools of cognitive science is critical to developing artificial intelligence systems with a human-like moral sense," writes co-author **Sydney Levine** in **a related Twitter thread**. Keep watching this space for updates.

TEXAS GOVERNOR DAN PATRICK CRUSADES AGAINST TENURE

Texas public universities' faculty are on notice. **Governor Dan Patrick** has signaled an interest in exercising control over elements of education. He has **committed [listen at the link]** to pursue "abolishing academic tenure at all Texas public universities" which is part of a larger plan to "extend a ban on teaching what the state defines as critical race theory to institutions of higher education." Please let us know what is happening within your universities, Texas public university readers.

LINKEDIN FULL OF AI-GENERATED FACES (ON FAKE PROFILES)

Not sure why, exactly, but **LinkedIn** has a growing cadre of **fake profiles** crowned by AI-generated profile pics. The go-to guy for sniffing out digital fakes, **Hany Farid**, confirmed what the naked eye generally cannot: their head shots are DeepFakes. Still unclear what the point of a fake LinkedIn profile is.

STATE DEPARTMENT LOOKING TO HIRE A BUNCH OF DATA SCIENTISTS

If you're sniffing around to see what other careers may be out there, the **U.S. State Department** has launched an **initiative** to hire a bunch of data scientists. Unlikely to be dull work at State.

NASA WILL BE IMPLEMENTING DOUBLE-BLIND PEER REVIEW

NSF could learn a little about equity and inclusion from **NASA**. The space agency has moved to implement a double-blind peer review process after **seeing evidence** that it can increase the representation of women PIs among those with accepted proposals.

ALL THE MUSIC - NOT THE GRAMMYS, PROBABLY BETTER

Normally, we would never claim MIDI files are better than the Grammys. In **this case**, 68 billion MIDI files represent all possible songs "with the basic 8 major notes (C4,D4,E4,F4,G4,A4,B4 and C5) with length 12." The MIDI files were (and are) sketches of songs — not all that enjoyable to listen to. But these sketches faithfully represent melodies and have been released with Creative Commons license. This should stop copyright claims on melodies, according to the project's creators **Damien**

Riehl (programmer/copyright attorney) and **Noah Rubin** (programmer).

MIT REINSTATES SAT/ACT ADMISSION REQUIREMENT
Hard to say if **MIT** is bucking a trend or starting a trend. They have **announced** they will reinstate the expectation that undergraduate admissions candidates must present SAT/ACT test results. They turn to the rationale that fueled initial interest in standardized exams — "standardized tests help us better assess the academic preparedness of all applicants, and also help us identify socioeconomically disadvantaged students who lack access to advanced coursework or other enrichment opportunities that would otherwise demonstrate their readiness for MIT. We believe a requirement is more equitable and transparent than a test-optional policy."

BURGERS OR PIZZA
There's **a new meat cloud** at the mall in Jersey City. The unmistakeable aroma of charred beef fills the air around a tall red metal cabinet that houses **Roboburger**, a vending machine with a footprint the size of my desk that serves up freshly made cheeseburgers. And it's the second most impressive fast food robot in today's rundown. **MIT** has **a pizza machine** that has mastered one of soft robotics grand challenges: handling material whose shape cannot be described in equations and navigating a multi-step process. Ingredients also matter. Roboburger touts prime angus from **Pat LaFrieda's** pastures. But **Enersto Di Maio**, a yeast-allergic Italian researcher, created a precise model of pizza dough's material properties to invent **a viable yeast substitute**. No smartphone app yet for Roboburger but customers (so far) don't mind the 6-minute wait for their burger. Other **app-hesitant** drive-through fast food customers are ~~seeing~~ **hearing work-from-home order takers**, provided by **Bite Ninja**, a Memphis startup. Worried that fast food disruption will displace cherished classics? Don't. **Patt Morrison's** **Explaining L.A. column** says that fast food's constants remain constant. "**Carl's Jr.** founder **Carl Karcher** once said, 'Everyone in the fast-food industry envies **In-N-Out**,' founded in 1948. 'We're working on new products every year, and In-N-Out keeps the same menu and knocks 'em dead.'"

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, Refik Anandol from April 5, 2022

Refik Anadol ✓
@refikanadol

Gm! Super excited for last two days for our studio's epic collaboration with @inspiration4x and @SpaceX data to support @StJude hospital! Thanks to @nftcws for aligning the worlds and @manifoldxyz for the web3 development for the drop! important-memory.com 7th of April! :)

0:36 2,960 views
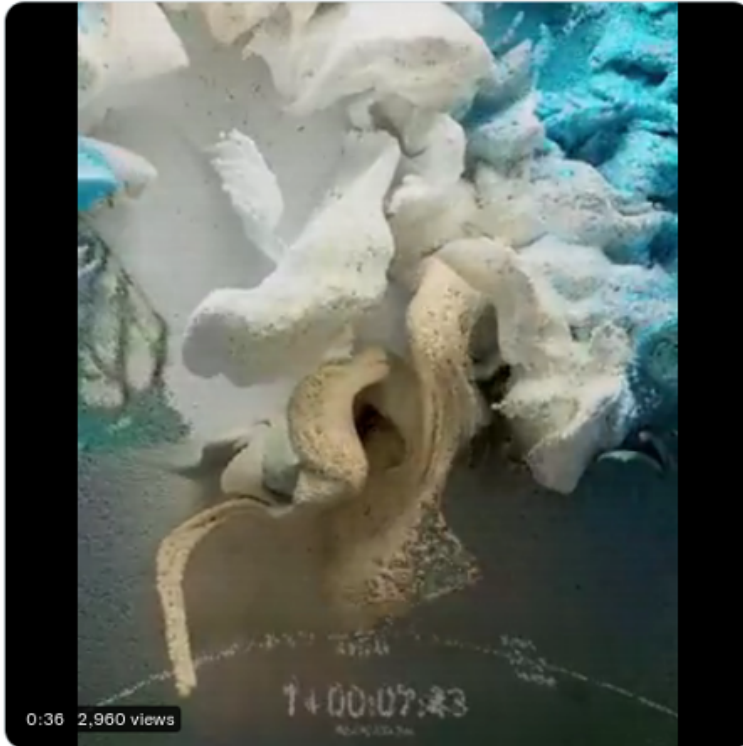
11:20 AM · Apr 5, 2022 · Twitter for iPhone

40 Retweets   8 Quote Tweets   196 Likes

Deadlines
Studies/Surveys
Help governments get smarter about national compute plans by filling out this @OECD survey.
"The goal of this survey is to provide a lot of information about how different people and organizations use compute for AI tasks"

Conferences
Call for papers for the fourth Workshop on Narrative Understanding (WNU, to be held at #NAACL2022) is now available!
"Submission deadline: April 15"

Contests/Award

Allen Institute Launches Updated Embodied AI Challenge

"The challenge requires competitors to design an autonomous agent that can move objects in a virtual room and includes several improvements including a new dataset and faster training using the latest release of the AI2-THOR simulation platform." Deadline for submissions is June 3.

Tools & Resources

Zotero 6: Your research workflow, transformed

Zotero from March 17, 2022

"We're thrilled to announce the release of Zotero 6, the next major version of Zotero and the biggest upgrade in Zotero's history. Zotero 6 introduces an array of new features to transform how you do research, including a completely new way of working with PDFs and notes."

From data to physical artifact: Challenges and opportunities for designing physical data artifacts

ACM Interactions, Kim Sauvé and Steven Houben from March 25, 2022

"In recent years, perspectives such as human-data interaction (HDI), data commons, and data humanism proposed to shift the ownership, actionability, and interaction with data toward people themselves. But democratizing the access and presentation of data in a meaningful context to nonexperts remains a problem, as there is a systemic lack of tools, visualization approaches, and conceptual and interaction models targeted at nonexpert groups."

My book is out today!

Twitter, Cathy O'Neil from March 22, 2022

"My new book, *The Shame Machine*, will be on sale on March 22nd. I can't wait to talk to you about it!"

Featured Events

See the ADSA Events Page for more details and more opportunities.

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.