

Please let us ([Laura Norén](#), [Brad Stenger](#)) know if you have something to add to the newsletter. We are grateful for financial support from the Academic Data Science Alliance.

DATA SCIENCE NEWS

University Data Science News

New evidence from **Elisabeth Bik**, **Leonid Schneider**, and **Smut Clyde** follows other [recent evidence](#) from **Jana Christopher** and suggests that there's an academic paper mill in China that sells academic papers to Chinese medical school students who are required to publish in order to get their degrees. At least one paper about tadpole development is so good that it has been published 400 times in journals run by publishers like **Elsevier** and **Wiley**. The key to finding the papers has been the images. They're all the same. Please, if you are a computer vision expert, I'm sure there are other papers besides the tadpole paper out there in the paper mill. This is a project for you! Go find those other papers and help us all figure out how widespread this practice is. Scientific misinformation undermines the entire scientific community at a time when we're already facing political headwinds.

In another dispatch from the realm of academic fraud, **Justin Pickett**, an associate professor at **SUNY Albany**, [wrote a long read](#) about how he tried and failed to get **Florida State University** to investigate claims that **Eric Stewart** fabricated data underlying five studies. Even though the school declined to investigate, Stewart retracted all five studies *subsequent* to the complaint. Pickett decided to look into other cases of academic fraud and found that universities are likely to drag their feet, ignore their own policies, and generally do what they can to avoid drawing negative attention to themselves or their researchers (and potentially cause them to have to return federal grant money or at least jeopardize the researchers' ability to raise additional funds). Get it together, universities! Tolerating this kind of research fraud is absolutely unacceptable as it undermines the scientific body of knowledge.

The **University of California System** came to [an agreement](#) with open access publisher **Public Library of Science** (PLOS) to make open access publishing easier for authors. Open access journals, including PLOS, ask authors to pay hefty fees for the privilege of publishing. This keeps access free for readers. However, authors are facing thousands or even tens of thousands in publishing fees per year depending on where and how much they publish. To smooth the wrinkles, "UC Libraries will automatically pay the first \$1,000 of the article processing charge (APC) for all UC authors who choose to publish in a PLOS journal." This experiment will run for 2 years. Obviously, PLOS is thrilled.

Facebook has opened [\\$2 million in unrestricted grants](#) to researchers interested in studying "misinformation and polarization on social media." The trick? You have to apply by April 1.

Architectural planning is underway for **Virginia Tech's new Innovation Campus** in Alexandria, Virginia. As previously noted, a country school establishing a city campus (Alexandria is near DC) has precedent. Cornell completed the Cornell Tech campus in Manhattan recently. Medical and law schools have also played the urban outpost role, extending a university's brand and physical presence into new territory. This could dramatically elevate Virginia Tech's status, given the strength of the northern Virginia tech corridor and the absence of any other strong technical university in that area. (To my colleagues at **George Washington University**, I consider you to be geographically out of the northern Virginia scene.)

Yale University, already in the process of renovating a building to house Departments of Astronomy, Mathematics, and Statistics and Data Science as well as the Department of Physics, is now toying with the idea of building a **wholly new building for quantum research**. It's nice to have a giant endowment and billionaire alums.

Billionaire **Dean Takahashi** **started** the **Carbon Containment Lab** at **Yale's School of Forestry and Environmental Studies** last summer. Takahashi is doing what he can to address climate change by hiring a dozen people to work with him. I applaud the effort but find it odd that billionaires can just start labs in universities?

Georgia State is in the process of introducing a **new bachelors degree in data science**.

Middle Tennessee State University will be **offering a bachelors degree in data science** and a graduate certificate aimed at professionals looking to retrain.

Year for year, **women chemists publish just as much as men (1.33-1.32 papers per year)**, but they tend to leave science earlier, lowering their lifetime publication count according to a new bibliographic study conducted by **Junming Huang** et al. of **Northeastern University**. (In case you were wondering, the impact factor is also quite similar across genders.) The paper didn't get into why women may leave, but a rousing, unrelated Twitter conversation suggests that either **being single** or being unable to count on a spouse running the household can maximally stress academics out. We know from years of economic research that women are less likely to be supported by stay at home spouses and that if their stay at home or even lower earning spouse is a man, he may be less likely to share in housework equally. This may be part of the pressure women academics face that some men academics do not face.

R turns 20 on Saturday, February 29th!

Company Data Science News

Securiti.ai is a privacy ops start-up that **won** the **RSA** Innovation Sandbox competition for the most promising *security* start-up. Yes, a privacy company won the security startup competition. Securiti.ai will basically automate a variety of typical privacy related tasks — figuring out where personally identifiable information (PII) is in a company's data stack, performing some aspects of third party vendor assessment for privacy compliance and exposure, and processing data subject requests to

know and to delete. I didn't get a real demo, so I cannot confirm that this tech actually works as described yet. I'm trying to find time to write a longer piece on the benefits and drawbacks of taking a security-style approach to privacy. The TL;DR is that we all benefit when companies automate their privacy ops because it means they are much more likely to perform a thorough, rapid process every time. However, enterprise products for privacy ops serve the needs of companies first and this moment of technologizing solutions will foreclose some of the wider ranging conversations about the privacy principles we *should* have. It's easy to stop talking about something we, as an industry, may feel that we have solved.

Speaking of security and privacy intersecting with one another. **Clearview AI**, the company we wrote about last week for having built an app that let's users input a photo and match it to other images of them from surveillance cameras, social media, and possibly drone footage, **has been hacked**. All the names of the police departments using the app have been exposed, along with the specific usernames, and number of searches run. It's unclear who did it or why, but it emphasizes the importance of maintaining ever better security controls to protect ever more sensitive data. Once those usernames are out there, it's much easier to phish those users to get access to specific searches, etc.

Clarifai is a New York startup that has previously been in the DCSN because of **questionable ethical business choices**. They are participating in Project Maven which is an autonomous weapons research project run out of the **Department of Defense**. The company is again **in the news for ethical lapses** with legal consequences. This time, they used **OKCupid** photos **to train their facial recognition model** which is...being used in autonomous weapons research. The case was brought in Illinois as a violation of the Illinois Biometric Information Privacy Act. Investors in Clarifai reportedly had access to OKCupid's entire database — that's how all the photos were transferred to Clarifai. Does anyone know if OKCupid reported a data breach? I haven't seen anything. They "should" have been forced to report and inform OKCupid account holders by various data breach laws.

The **Partnership on AI** put out a nicely **illustrated explainer paper on facial recognition** systems. It describes how facial verification (is this person who they say they are?) and facial identification (from a group of many faces, can I find a match for just target face?) are distinct. I can't tell if the paper is simply trying to educate or if it's trying to carve out a space for using some kinds of facial recognition — what they would call facial verification — as logical, acceptable and ethical. PAI does mention that facial recognition systems can "amplify civil liberties and human rights concerns" and has written on the **troubling implications of using biased data in predictive criminal justice systems**. I'll give them the benefit of the doubt here.

Twitter has recently changed its acceptable content policy to ban deepfakes and is now **testing several approaches to reduce the influence of misinformation on the platform**. One idea that was leaked last week is to allow those who are sharing useful information to rack up points within their community. This might help us all tell the difference between accounts deemed reliable and those spreading misinformation which are often newly created and thus wouldn't have high community ratings due to their novelty. (New accounts haven't had enough time to get upvotes.) The

other ideas for reducing misinformation were not leaked so we'll just have to wait and see. I am, however, generally appeased by Twitter's consistent tidying efforts around the evolving world of poo posting. Meanwhile, **Facebook** has [released a white paper](#) that sees the future of its content regulation as a problem for regulators outside the company. ::eyebrows-raised::

Google will be [moving its UK users' data out of the GDPR-zone](#) and into the US data jurisdiction as Brexit unfolds. As it stands, the plan for transitioning data stored in the UK to a post-Brexit legal status is still being written. Large companies do not like Brexit-levels of uncertainty.

Two years in the making, **The Cleveland Clinic** has finally gotten approval to build a [new \\$11.3 million biorepository](#) to store thousands of tissue samples.

JP Morgan hired an outside group to investigate the impact of climate change; it could be "catastrophic" and "irreversible." The [report recommended taxing carbon immediately](#) in order to bring carbon emissions to zero by 2050, though they note that is highly unlikely. It's very important to get major, powerful institutions on the correct (albeit grim) side of history on the climate change question because their opinions and reports are often much more immediately impactful than scientific reports written by scientists.

Government Data Science News

In the most exuberantly exciting government news of the week — correction, of the year! — the [Smithsonian has publicly released 2.8 million 2D and 3D images](#) from its collection on an [open access platform](#). This includes "data and material from all 19 Smithsonian museums, nine research centers, libraries, archives and the **National Zoo**." All you creative readers can get out there and remix in any way you see fit. The museums will continue to release images with another 200,000 scheduled to be available in 2020.

And...if you're more interested in public access to clinical trial data than pictures of paintings and zoo animals, please note that a judge in **New York's Southern District** has ruled that [scientists heading up 1,000 non-compliant clinical trials conducted between 2007 and 2017 must release their data](#). The story gets much murkier at this point in the reporting. No deadline for submitting data or punishment for the failure to submit data have been shared (maybe because they haven't been established at this time).

The **EU Commission** has met recently and unveiled [plans](#) for creating a data commons within the EU to help industrial giants to share data with each other and with small- and medium-sized enterprises. Having rightfully earned their status as the world's data privacy referee, they are now getting into the game itself (even though, as one pithy economist, **Guntram Wolff**, noted, ["referees never win"](#)). The new sharing-centric model will first develop guidelines for privacy protection and then make cloud and edge computing resources available to foster the sharing. Since Europe is a major manufacturing center, it's likely that much of the data will be derived from factories, though there is also a fair amount of society-ops data (train schedules, weather data, public health data, traffic data, land use data, etc). I have long been interested in the notion that setting up well-regulated

data commons is more likely to share the benefits of data driven advances while simultaneously allowing various scientific and civil rights watch dogs access to data for purposes of bias testing, reproducibility, and cost-reductions. I find the EU's plans to be truly visionary and likely to address their status as a lagging third behind the US and China in the AI race. The proof, of course, will be in the execution.

The **U.S. Department of Defense** has [released its guidelines for using AI](#). Defense Secretary **Mark Esper** discussed the **Defense Innovation Board's** five key guidelines: "the Pentagon pledges that its AI efforts will be: 1) Responsible, 2) Equitable, 3) Traceable, 4) Reliable and 5) Governable." In particular, the Pentagon seems to want Americans to know the feds have no plans to use cameras and facial recognition to surveil the public....well, except at the airport or any other boarder crossing. That's fair game. The main issue I want the DoD to be thinking about when it comes to AI is whether or not autonomous weapons are ethical and whether or not facial recognition attached to weapons so far away the target can't see them with the naked eye is ethical. I believe those projects are already live in product or in QE testing. Is skipping over questions about autonomous and invisible weapons following the first principle of being responsible? Who gets to determine what responsible means?

The **Pentagon** just announced it will appoint **Alka Patel** to ["lead the implementation](#) of its new ethical principles for using artificial intelligence in warfare." This news is very fresh and I am still exploring her background. So far, she seems junior - no military experience, no policy writing experience, not much published research. I'll update you on this incredibly important appointment as soon as I have more to share.

Extra Extra

There's a [new 6E WiFi technology](#). It will require the purchase of new routers and upgrades to all existing devices' WiFi chips. It's not a bad idea to make sure any new devices — including smart home devices — are 6E compatible to get the service improvement. The improvement is certainly not big enough to warrant rushing out to buy new gear.

I'm not going to comment on [WIRED's recent profile](#) of **Google X** (the commercial computing lab where money flows in swells large enough to propel a hundred surfers at once) except to say that I'm happy to see its Captain of Moonshots, **Eric "Astro" Teller**, helping bring rollerblades back to popularity.

And...**Terry Gross** [interviews Steven Levy](#) about his new book chronicling the history of **Facebook**.

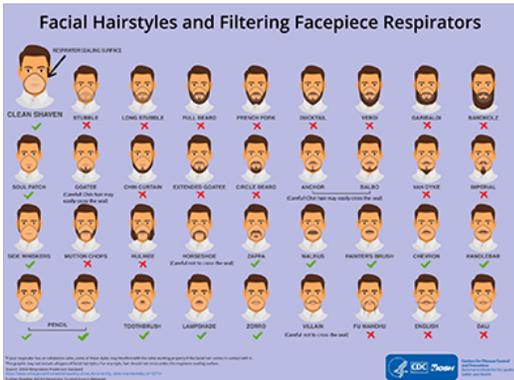
Tweet of the Week

Twitter, Paul Gorski from February 27, 2020



Data Viz of the Week: How to shave for CoVid-19 according to the CDC

U.S. Centers for Disease Control from February 26, 2020



EVENTS

Counter-technologies at the Edgelands Conference

San Diego, CA, and Tijuana, Mexico February 28-29. "In this conference, we bring together activists, organizers, academics and artists to interrogate how border communities are particularly affected by new technologies of militarism, police, and border enforcement in San Diego/Tijuana and globally." [free, registration required]

Women in Data Science (WiDS) 2020 at Stanford University

Stanford, CA March 2, 2020. "This one-day technical conference features amazing thought leaders in data science from academia, industry, non-profits, and government. Topics presented will cover a wide range of domains from data ethics and privacy, healthcare, data visualization, and more." [\$\$\$]

TechCrunch Robotics + AI — March 3 at UC Berkeley

Berkeley, CA March 3, starting at 8:30 a.m., Zellerbach Hall. "In addition to a main stage that includes the likes of **Amazon's Tye Brady**, **UC Berkeley's Stuart Russell**, **Anca Dragan** of **Waymo**, **Claire Delaunay** of **Nvidia**, **James Kuffner** of **Toyota's TRI-AD** and a surprise interview with **Disney Imagineers**, we'll also be offering a more intimate Q&A stage featuring speakers from **SoftBank Robotics**, **Samsung**, **Sony's Innovation Fund**, **Qualcomm**, **Nvidia** and more." [\$\$\$]

Seeing Data: A Conversation on Visualization

New York, NY March 4, starting at 6 p.m., **NYU Arthur L. Carter Journalism Institute** (7th Floor Commons, [20 Cooper Square](#)). "A conversation with **Betsy Mason** and **Enrico Bertini**."

Reality Virtually Medical Hackathon

Boston, MA June 25-28. "All medical professionals and virtual reality enthusiasts are welcome and encouraged to submit an application!" [application required]

2020 MIT Platform Strategy Summit

Cambridge, MA July 8, starting at 8 a.m., **MIT Media Lab**. "This event gathers a global community of executives and academics to explore the economics and management of platform-centered markets and discuss their implications for managers, industry and governmental policy." [\$\$\$]

DEADLINES

Contests/Award

ICWSM2020: The Data Challenge: Safety

"ICWSM 2020 is hosting the first ICWSM data challenge to bring together researchers from across disciplines to solve societally-relevant problems together as a community. This will be enabled by fostering collaboration and exchange of ideas in a structured setting. This year's data challenge theme is Safety. To achieve this, we invite participants to work on two pertinent datasets in the areas of Misinformation and Abusive behavior in social media." Deadline for submissions is April 25.

Conferences

CHI Play 2020

Ottawa, ON, Canada November 1-4. The conference "grew out of the increasing work around games and play emerging from the ACM annual conference on Human Factors in Computing Systems (CHI) as well as smaller conferences such as Fun and Games and Gamification." Deadline for submissions is April 7.

RecSys 2020 – Call for Contributions – RecSys

Rio de Janeiro, Brazil September 22-26. "RecSys also strongly encourages the submission of algorithmic papers that repeat and analyze prior work." ... "Submissions regarding replicability or reproducibility papers are welcome in all areas related to recommender systems (see the main track Call for Papers for a list of topics)." Deadline for submissions is May 4.

NeurIPS 2020 Call for Papers

Vancouver, BC, Canada December 6-12. Deadline for abstracts submissions is May 5.

Education Opportunities

Engaging Geography in the Humanities

Boston, MA July 6-24 at **Northeastern University**. "The Institute will explore the possibilities and productive tensions at the intersection of geography and the humanities. By engaging with readings, lectures, discussions, workshops, and field visits, the Institute will introduce scholars teaching in the humanities (and related disciplines) to concepts and methods from geography, as participants consider how these approaches can enhance their own research and teaching." Deadline to apply is March 1.

Argonne Training Program on Extreme-Scale Computing

St. Charles, IL July 26-August 7. The program "provides intensive, two-week training on the key

skills, approaches, and tools to design, implement, and execute computational science and engineering applications on current high-end computing systems and the leadership-class computing systems of the future." Deadline to apply is March 2.

Call for participation at the 2nd Excited Summer School on Research in Computing Education

Trondheim, Norway June 8-12. "The school is intended for PhD students and post docs. There will be varied sessions led by experts in the computing education field, combining presentations with discussion and hands-on groupwork. Topics range from how to teach beginner-level programming courses to the inclusion of environmental sustainability in IT education." Deadline to apply is March 10.

Technology Fellow

"Fellows work both as a cohort across the foundation and as part of program teams to explore the role and impact of technology with social justice and help develop a critical technology lens throughout the foundation." Deadline to apply is March 11.

Center for Technology and Society Belfer Fellowship

The fellowship "was established to support innovative research and thought-leadership on combatting online hate and harassment for all. Fellows will be drawn from areas like the technology community, academia, and media to push innovation, research and knowledge development. They will embark on projects that align with CTS's mission, like studying new areas of the online hate ecosystem and providing recommendations for industry, government, and civil society to promote respectful and inclusive digital environments and reduce the spread of toxicity online." Deadline to apply is March 31.

Journal/Publishing

Call For Papers: Special Issue on AI Fairness, Trust and Ethics

"This Special Issue on AI Fairness, Trust and Ethics calls for research that can unpack the potential, challenges, impacts, and theoretical implications of AI. We welcome research from different perspectives regardless of the approach or methodology. Submissions with novel theoretical implications that span disciplines are strongly encouraged. We seek submissions that can improve our understanding about the impacts of AI in organizations and our broader society." Deadline for full paper submissions is March 31.

Studies/Surveys

Request for Information: Public Access to Peer-Reviewed Scholarly Publications, Data and Code Resulting From Federally Funded Research

"OSTP and the **National Science and Technology Council's (NSTC) Subcommittee on Open Science (SOS)** are engaged in ongoing efforts to facilitate implementation and compliance with the 2013 memorandum Increasing Access to the Results of Federally Funded Scientific Research [1] and to address recommended actions made by the Government Accountability Office in a November 2019 report.[2] OSTP and the SOS continue to explore opportunities to increase access to unclassified published research, digital scientific data, and code supported by the U.S. Government. This RFI aims to provide all interested individuals and organizations with the opportunity to provide recommendations on approaches for ensuring broad public access to the peer-reviewed scholarly

publications, data, and code that result from federally funded scientific research." Deadline to submit comments is March 16.

RFP

NSF-Simons Research Collaborations on the Mathematical and Scientific Foundations of Deep Learning (MoDL)

The **National Science Foundation** and the **Simons Foundation Division of Mathematics and Physical Sciences** "will jointly sponsor up to two new research collaborations consisting of mathematicians, statisticians, electrical engineers, and theoretical computer scientists. Research activities will be focused on explicit topics involving some of the most challenging questions in the general area of Mathematical and Scientific Foundations of Deep Learning." Deadline for letter of intent is March 20.

TOOLS & RESOURCES

Launching a Jupyter notebook from within a virtual environment

Vicki Boykis from February 18, 2020

"When I'm working with Jupyter notebooks, I often want to work with them from within a virtual environment. The general best practice is that you should always use either virtual environments or Docker containers for working with Python, for reasons outlined in this post, or you're gonna have a bad time. I know I have."

Getting Started with NeurIPS 2020

Medium, Neural Information Processing Systems Conference from February 19, 2020

"This is the first in a series of notes that we will write to shed light on the organization, review process, and program for NeurIPS 2020."

ConvNetJS CIFAR-10 demo

Andrej Karpathy from February 19, 2020

"This demo trains a Convolutional Neural Network on the CIFAR-10 dataset in your browser, with nothing but Javascript. The state of the art on this dataset is about 90% accuracy and human performance is at about 94% (not perfect as the dataset can be a bit ambiguous). I used this python script to parse the original files (python version) into batches of images that can be easily loaded into page DOM with img tags."

In depth guide to running Elasticsearch in production

TechNotes, Mattis Haase from February 22, 2020

"Elasticsearch can be opaque, confusing, and seems to break randomly for no reason. In this post I want to share my experiences and tips on how to set up Elasticsearch correctly and avoid common pitfalls."

Over 100 partners to help you succeed with the GitHub Student Developer Pack

The GitHub Blog, Scott Sanicki from February 25, 2020

"Six years ago, we launched the GitHub Student Developer Pack, providing over 1.8 million students access to the best real-world developer tools and training for free. We're excited to share that we've reached over a hundred companies partnering with GitHub to help student developers get their careers off to the right start."

"This February, we're adding 14 new offers, each providing a free and unique benefit available to all student members of the Pack. Whether you're designing your first game, developing your first desktop application, or publishing your first mobile app—the Pack provides you with the tools you need to be successful."

CAREERS

Tenured and tenure track faculty positions

Faculty position at the rank of Professor (with tenure), Associate Professor (with tenure) or Assistant Professor (tenure-track)

University of Washington, Department of Biostatistics; Seattle, WA

Faculty Positions in Computer Science

University of Rochester, Department of Computer Science; Rochester, NY

Two Open-Rank Faculty Positions in Cognitive Science

Central European University, Department of Cognitive Science; Vienna, Austria

Quantitative Foundation Associate Dean for Academic and Faculty Affairs

University of Virginia, School of Data Science; Charlottesville, VA

Executive Director, Computational Social Systems

University of Pennsylvania, The Wharton School; Philadelphia, PA

Assistant Professor in Interaction Design

KTH Royal Institute of Technology, Media Technology & Interaction Design Division; Stockholm, Sweden

Computing & Data Sciences - All Ranks

Boston University, Faculty of Computing & Data Sciences; Boston, MA

Full-time, non-tenured academic positions

Research Scientist (Open Rank)

Columbia University, Data Science Institute; New York, NY

Data Library Engineer

University of California-San Francisco, Data Science CoLab; San Francisco, CA

Instructor

University of Colorado, Department of Information Science; Boulder, CO

Science Communication Liaison

University of California-Santa Barbara, National Center for Ecological Analysis and Synthesis (NCEAS); Santa Barbara, CA

Senior Research Data Scientist

University of California-Davis, DataLab: Data Science & Informatics; Davis, CA

Senior Data Analyst

University of Minnesota, IPUMS; Minneapolis, MN

Data Analyst

University of Minnesota, IPUMS; Minneapolis, MN

Assistant Professor, Library Chief Data Strategist

University of Rhode Island, University Libraries; Kingston, RI

Sea Grant Marine Extension Associate (Climate Community Resilience)

University of Maine Cooperative Extension, Maine Sea Grant; Orono, ME

Postdocs

Postdoctoral fellowship (-s)

Aarhus University, DATALAB – Center for Digital Social Research; Aarhus, Denmark

CDS-CEAP Faculty Fellow

NYU Center for Data Science, NYU Center for Environmental and Animal Protection; New York, NY

Postdoc - influence of social media on politics

New York University, Center for Social Media and Politics; New York, NY

Postdoctoral Appointee - Machine Learning for Multiphase Flow Modeling

Argonne National Laboratory; Lemont, IL

Postdoctoral Associate

Stony Brook University, Department of Political Science; Stony Brook, NY

Postdoctoral Associate/Senior Researcher

Stony Brook University, Department of Political Science; Stony Brook, NY

Postdoctoral Fellow

Northeastern University, Institute for Experiential AI; Boston, MA

Postdoctoral Research Assistant in High Precision Radial Velocity Searches for Exoplanets (2 posts)

University of Oxford, Department of Physics; Oxford, England

Full-time positions outside academia

Technical Assistant to the CEO

The Allen Institute for Artificial Intelligence; Seattle, WA

Chief Innovation Officer

City Manager of Oklahoma City; Oklahoma City, OK

Research Lead - Personalization

Spotify; London, England

Research Fellow - Physiology & Biomedical Engineering

Mayo Clinic, Multimodal Neuroimaging Lab; Rochester, MN

Human Computation Architect - CTO Data Science

Bloomberg LLP; New York, NY

Critical Designer, Data Genesis

New York University, AI Now Institute; New York, NY

Design Director or Senior Digital Designer

Infographics Group GmbH; Berlin, Germany

Internships and other temporary positions

Bioinformatics Summer Intern

Broad Institute, Bayer Precision Cardiovascular Lab; Cambridge, MA

Data Science Intern (Summer)

AARP; Washington, DC

writer/support intern

Chartable, Datawrapper; Berlin, Germany

Doctoral Studentship in Gender and AI Applications

University of Oxford, Oxford Internet Institute; Oxford, England

Click here to receive the Data Science Community Newsletter and/or to have us follow your twitter feed so that our data science twitter bot can easily grab links from your tweets.

To send us an announcement for the newsletter, please email laura.noren@nyu.edu and brad.stenger@gmail.com. We retain curatorial discretion.

Data Science Community Newsletter Issue 191.