

Data Science Community Newsletter

10 April 2020

Data Science Community Newsletter

Please let us ([Laura Noren](#), [Brad Stenger](#)) know if you have something to add to the newsletter.

We are grateful for financial support from the [Academic Data Science Alliance](#) and to NYU's Center for Data Science.

Data Science News

University Data Science News

Dr. Marie-Louise Landry of **Yale Medical School's Clinical Virology Lab** decided to be the change we are all seeking: she **retooled her lab to test for COVID-19**. Of course, hers isn't the only research lab that reoriented themselves to fight the pandemic in a matter of weeks, but every story like this is a story that we need to hear. Landry notes that supply chain limitations with test reagents continues to be a barrier. She also points out that we all need to realize how immunologically naive we are to this virus. Very few of us will have been exposed after the first wave crests and subsides. Testing capacity looks likely to remain limited until mid-May, if Landry's experience is typical. It will be difficult to go back to normal until we have the testing capacity to quickly recognize and contain any new cases and their recent contacts. That will require lots of tests, enough to test many people who turn out not to have coronavirus.

Following our discussion from last week and the week before about using mobile phone location data to track people's movement during COVID-19, we are **covering a project** led by a team of researchers from **Harvard's TH Chan School of Public Health, Harvard Medical School,** and **Direct Relief** with a coalition of research institutes in the US, UK, and India. They have received aggregate data from **Facebook**, which is installed on many mobile phones. That data can be used to assess the impactfulness of stay-at-home orders. Are people really staying at home? Do people start to stray when the weather gets nice and cabin fever breaks records? I admit that I like the overall goals of the project, but am deeply uncomfortable by the way we're relying on a large, powerful, not-always-trustworthy private company (Facebook) for key public health research infrastructure.

Carnegie Mellon University also has a research partnership with **Facebook** around COVID. In **this project**, **Facebook** will display a survey link to certain users. The survey is not hosted by Facebook, nor does Facebook see the responses or know who took the survey. The survey, like others sponsored by university researchers (including **work** by Friend of the Newsletter **Dan Larremore**), asks about COVID-19 symptoms and monitors social distancing. Lead author **Ryan Tibshirani** claims this type of aggregate symptom tracing cannot be obtained without the Facebook partnership due to lack of testing. While he has a point about the utterly maddening inability of the US to provide enough testing infrastructure, relying on Facebook is just one among many alternatives to

widespread testing.

For instance, **Laura Huckins** at **Mount Sinai School of Medicine** set up a COVID-19 symptom [texting application](#). New Yorkers are urged to sign up and receive automated SMS prompts guiding them to report symptoms and locations so that researchers can understand disease dynamics within NYC. No corporate partnership required.

Perhaps the [most famous COVID-tracking project](#) — the [COVID-19 dashboard](#) at **Johns Hopkins School of Public Health** — is also able to gather global data and reach large audiences without corporate sponsorships. Professor **Lauren Gardner** working closely with **Ensheng Dong**, put the dashboard together quickly, labored through a bunch of manual data entry, and have now had their project forked 20,000 times on **GitHub** while becoming a de facto source of truth on the progression of disease as hotspots shift from region to region.

The **Association for Computing Machinery** (ACM) has "strongly encouraged" all of its affiliated conferences to publish conference proceedings for all of the 2020 ACM conferences, whether they take place in person or not. ACM's Human Robot Interaction conference built an [app](#) that not only pulls in the DOIs of all the papers, but also carries funny videos to try to foster interactions that would have been part of the hallway track.

Additionally, **ACM** has entered into [Open Access Read + Publish](#) agreements with 18 colleges and universities, largely in California ([University of California System news release](#), [Caltech news release](#)) and Ireland. **MIT**, **Carnegie Mellon**, and **Iowa State** were also early adopters of the new model. (Does anyone know the inside story on why this isn't called the Read + Write model? They *must* have debated that.) These agreements provide "unlimited access to the complete contents of the ACM Digital Library and unlimited Open Access publication in ACM's journals" for a single one-time fee (3 year duration). At the moment, this will cause some tension between corresponding authors and their coauthors. Publication will be free and open for corresponding authors with their home base at a participating university. If coauthors are at a participating university, their work will be free for others within their institution to read, but won't be available for free to the entire world. The ethical dilemma between making someone the corresponding author in order to ensure the paper is available to the world even if they aren't the best scientific correspondent and giving corresponding author credit to the most worthy author is sticky. ACM hopes more universities will join, thereby reducing the incidence of this kind of authors' needs vs readers' needs dilemma.

Washington University in St. Louis will be offering a [new environmental analysis major](#) in response to climate change and the political leadership it will demand.

With accreditation in place, **The University of Oregon** is [launching its data science undergraduate major next fall](#). President **Michael Schill** said, "I am proud to note that the UO program will place a strong emphasis on building ethical frameworks for working with and learning from big data." There is a [required](#) Technology Ethics course, but it hasn't been assigned to a quarter yet, which may mean they haven't yet found the perfect instructor. If they're willing to have a

remote lecturer, I'd be honored!

Speaking of ethical challenges, the current [crackdown on scholars moving back and forth between China and the U.S.](#) while working on substantially similar or closely related projects has resulted in at least one **FBI-driven arrest** of **Harvard** Chemistry Department Chair **Charles Lieber** and two Boston-based Chinese nationals.

Now **American Association of Universities** President **Mary Sue Coleman** has weighed in. She [writes](#), "we must secure the research we conduct on our campuses from foreign interference while at the same time protecting the openness and free flow of scientific knowledge". Rather than being a complex discussion of applied ethical philosophy, there was no discussion of what healthy international scientific partnerships could or should look like, as her argument developed, just a lot of fear mongering about liars, spies, and thieves. Coleman discusses the importance of the rights of the taxpayer, cooperation with FBI agents, and adopting stronger cybersecurity practices. As someone who works in cybersecurity, I'm always a fan of good security measures, happy to set up a demo, but only in contexts where security is reasonable and in line with the expectations and practices of those involved. Cybersecurity systems are excellent, efficient, surveillance tools (well, the good ones are excellent and efficient). For those willing to be surveilled for their own protection — which includes many of us! — these systems are fantastic. But for academics who aren't sure how this talk about FBI cooperation and anti-Chinese sentiments will add up, the recommendations for more surveillance and law enforcement in scientific research strikes me as a little sinister. There is bad behavior in science. I'm not defending it. Chinese nation-state attacks and espionage are quite sophisticated and can be highly damaging. But let's think carefully about what we all give up when we surround scientists with surveillance and cops.

Company Data Science News

If you use the **Uber** ride sharing system, please understand that your drivers [cannot get paid time off until they've received a positive COVID test](#). Since testing is still limited, this may mean drivers are working through the early days of the infection during which they are shedding lots of virus. Good alternatives: Stay home. Drive yourself. Ride a bike. If you must be driven, arrange for a stretch limo.

Watching **Zoom** over the past four months is a little like watching a Greek tragedy. First, there was a huge wave of adoption and a new cohort learned how mute works. Everyone was on Zoom, especially professors and students. Generation Z was going to be called the Zoomers! Pet and kid cameos charmed us all. Even **Janet Yellen** did a [Zoom webinar](#), looking stately in a pantsuit and pearls in front of a white built-in bookshelf. Then the Zoombombing turned charm into revulsion. Journalists found the many articles written in places like **Dark Reading** and **ZDNet** outlining Zoom's security weaknesses — hackers had been exploiting Zoom's very aggressive installation hooks to take over cameras and make recordings. Wait, what?? Naturally, the **New York City Dept of Education** [officially switched](#) to **Microsoft Teams**. Some groups have turned to **Cisco's WebEx**. Readers, as my CEO has pointed out, Zoom and WebEx were written by the same guy, billionaire **Eric Yuan**. Mr. Yuan divulged earlier this week in a radio interview that his prioritization schema preferred to limit the number of clicks rather than increase privacy and security protections.

Like a good capitalist, he has promised to patch things up. This being litigious America, he is now [being sued by an investor](#) in a class-action case using CCPA.

Two of the largest location-based advertising companies — **FourSquare** and **Factual** — have [announced a merger](#). The merger will create one of "the largest location data sets in the world, spanning more than 500 million devices, a panel of 25 million opted-in, always-on users and more than 14 billion user-confirmed check-ins. The company will also have data on more than 105 million points of interest across 190 countries and 50 territories." With most consumers staying home and the future of brick-and-mortar retail, restaurant, and event venues in a tenuous position, the merger may be required for viability in the advertising market.

Eventbrite CEO **Julia Hartz** was forced to [layoff 45% of her employees](#) due to the prohibition on gatherings of any size. The layoff was conducted over Google Hangouts. It remains to be seen how much audiences are willing to pay to attend an online event.

In an excerpt from her [new book *No Filter*](#), **Sarah Frier** shares [the inside story](#) on **Instagram** and its tension-filled integration into the **Facebook** social media mafia family. You can thank me for this juicy, non-COVID indulgence of an anti-tech bro story by hitting reply and sharing your favorite part. One lucky reader gets the chance to thank me by buying me my very own copy. ;)

Government Data Science News

The **U.S. National Institutes of Health** has already kicked off a [six-city serological survey](#) to test for the presence of SARS-CoV-2 antibodies in blood donated by healthy donors. Right now, the study includes Seattle, San Francisco, Los Angeles, New York, Boston, and Minneapolis, but similar studies will expand to include others. Results detailing which percentage of New Yorkers and Seattle residents have been exposed (whether or not they were tested or ever developed symptoms) will be available in 2 months. This effort is a critical component of our national decision making apparatus as we face questions about what it means to get past the first big wave of infections. Serological studies will not only shed light on the potential size of the asymptomatic infected population, but also on how long immunity might last. Will those who were infected by immune to reinfection for 1-2 years? 20 years? A lifetime? We will soon understand much better. (Presumably, serological studies in China will be available first, and will answer questions about the temporal extent of acquired SARS-CoV-2 immunity.)

In other testing news, the federal government has announced it will [end funding for community testing sites Friday, April 10th](#). Though testing is still not available to all patients with symptoms in most of the country, a spokesperson for the **U.S. Department of Health and Human Services** described this move as a transition to state-supported, hospital-based testing, not a closure of community testing. Those who have been working at the community-based sites are concerned that hospitals are stretched too thin to accommodate any additional testing demand. To an untrained eye (but also to a trained eye) the closures appears to be an admission of defeat.

In a more blatant admission of defeat, the **E.U.'s** top science advisor, **Mauro Ferrari** has [resigned in](#)

frustration over the **European Research Council's** handling of the coronavirus. Ferrari had only assumed the position in January. He indicated it involved too much politics and palace wars, not enough science.

Add HIPAA to the list of government regulations that will probably be crisis-changed irrevocably. In order to provide agencies like CDC with quick access to patient data the **U.S. Department of Health and Human Services** (always HHS) **won't be enforcing HIPAA violations** of public health projects during the pandemic.

Important non-coronavirus news: A U.S. Federal Judge **ruled in favor** of **University of Michigan** professor **Christian Sandvig** and the **American Civil Liberties Union (ACLU)** in a recent First Amendment case allowing researchers to violate online terms of service in order to investigate potential algorithmic discrimination. The **Electronic Frontier Foundation** **wrote that the decision** "solidifies the trend in courts to reject overbroad and constitutionally troublesome interpretations of the [Computer Fraud and Abuse Act] that have threatened to criminalize innocuous online activity, such as beneficial research and journalism."

Extra Extra

Zeynep Tufekci, writing in **The Atlantic**, **makes the point** that keeping parks open, even if access needs to be rationed, is the right thing in terms of public cost-benefit.

How to connect with students in the COVID years? Play virtual bingo. Or **make a replica of your campus** in **Minecraft**.

Nathan Yau has constructed a **toilet paper calculator** to help assuage your fears that you will run out. It is unclear how he knows exactly how many sheets are on rolls produced by different manufacturers, but he does and it matters. Like all tp calculators I have seen, it is lacking consideration for differences in genitalia.

Tweet of the Week

Twitter, *Carl Bergstrom* from March 31, 2020

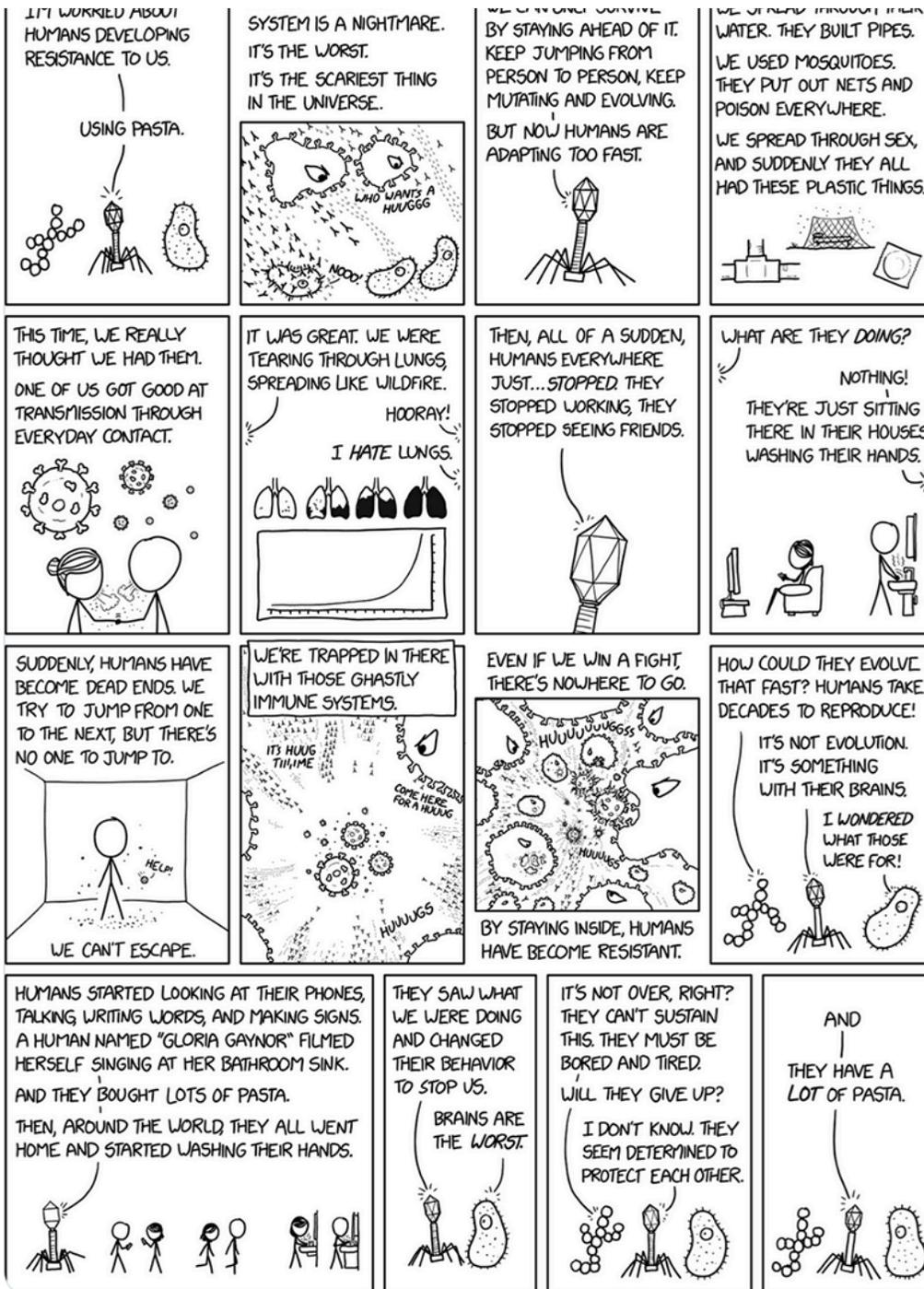


So many times I've laughed aloud at xkcd, but so seldom has it make me feel this sentiment.

Goodnight and be well.

xkcd.com/2287/





Data Visualization of the Week

Twitter, Simon German, National Geographic from March 29, 2018



EVENTS

Bloomberg Quant (BBQ) Seminar Series

Online April 15, starting at 5:30 p.m. EDT. "In this session, chaired by **Bruno Dupire**, **Alexander Lipton** will present his current research, followed by several 'lightning talks' of 5 minutes each in quick succession. This format gives the audience the opportunity to be exposed to a wider variety of topics." [registration required]

Data Visualization Foundations, Stamen Design

Online April 25-26. "If you or your company wants to present data effectively using data visualization, needs in-depth background knowledge about the art and science of data visualization, or needs to acquire data visualization literacy, this event is for you!" [\$\$\$], limited to 30 attendees]

The schedule for How to do Map Stuff: A Live Community Sharing Event

Online April 29-30. "Join us on Apr 29th or 30th (depending on your time zone) and watch a whole bunch of mappers share their knowledge live on the internet!"

!!Con 2020

Online May 9-10. "Two days of ten-minute talks (with lots of breaks, of course!) to celebrate the joyous, exciting, and surprising moments in computing." ... "will be held as an experimental remote conference on the weekend of May 9-10, 2020!! More information following soon." [save the date]

Join us for Data for Black Lives III

Cambridge, MA December 11-13 at **MIT Media Lab**. "Returning to Cambridge will make it possible for us to accomplish all of our goals: expand our programs, continue to build our network, and host the best conference yet." [save the date]

DEADLINES

Conferences

Future of Synthesis in Ecology Workshop

Santa Barbara, CA September 21-22. "Are you interested in shaping the future of synthesis in ecology and environmental science? We are seeking an interdisciplinary, diverse group of 100 participants across all career stages. Approximately 80 participants will be provided with full funding for travel and accommodations for 2 nights in Santa Barbara, CA." Deadline to apply is May 1.

Education Opportunities

2020-2021 Bloomberg Data Science Ph.D. Fellowship

"**Bloomberg** invites exceptional Ph.D. students working in broadly-construed data science, including natural language processing, machine learning, and artificial intelligence to apply for the Bloomberg Data Science Ph.D. Fellowship for the academic year of 2020-2021." Deadline for applications is April 12.

ACM SIGHPC Computational & Data Science Fellowships

"**ACM SIGHPC** has created the Computational and Data Science Fellowships, a continuation of the program started with Intel (see below) to increase the diversity of students pursuing graduate degrees in data science and computational science. Specifically targeted at women or students from racial/ethnic backgrounds that have not traditionally participated in the computing field, the program is open to students pursuing degrees at institutions anywhere in the world." Deadline for nominations is April 30.

Studies/Surveys

Help Predict COVID-19's Spread in Your Community

"Leading epidemiologists are assisting the CDC and other federal and state agencies to develop computational models of how COVID-19's might spread in your community. These models depend on knowing the different social distancing measures in each of the 3000+ counties in the United States. We urgently need volunteers to help us build a dataset of social distancing measures."

"Due to a high rate of response from volunteers, we are taking longer than usual to verify and post data to our site. Thank you for your contributions!"

RFP

Alan Turing Institute: Call for COVID-19 rapid response data science taskforce

"This is a rapid response call for teams of researchers as well as individual researchers to join us in addressing these urgent questions. Teams will rapidly trial tested, robust and reproducible data science and machine learning methods on this new national data resource of COVID 19 data, to create a near real-time, human-in-the-loop analytics platform and information console to support acute care management and clinical decision making. DECOVID has been structured to be nationally scalable." Deadline for talent pool applications is April 10.

Emergent Ventures Fastgrants - Fast Funding for COVID-19 Science

"If you are a scientist at an academic institution currently working on a COVID-19 related project and in need of funding, we invite you to apply for a Fast Grant. Fast Grants are \$10k to \$500k and decisions are made in under 48 hours. If we approve the grant, you'll receive payment as quickly as your university can receive it." Deadline for applications is April 12.

European Space Agency: Submit an innovative proposal for characterising COVID-19 impacts

"How can EO data help with monitoring COVID-19 impacts on the society?" Deadline for proposals is April 17.

Department of Energy Announces \$30 Million for Machine Learning and Artificial Intelligence Research | Department of Energy

The initiative encompasses two separate topic areas. One topic is focused on the development of ML and AI for predictive modeling and simulation focused on research across the physical sciences. ML and AI are thought to offer promising new alternatives to traditional programming methods for computer modeling and simulation.

A second topic is focused on basic ML and AI research for "decision support" in managing complex systems. Potential eventual applications could include cybersecurity, power grid resilience, and other complex processes where ML and AI can make or aid in making decisions in real time. Deadline for pre-submission application is May 1.

TOOLS & RESOURCES

pewmethods

GitHub - *pewresearch* from March 24, 2020

"An R package developed by the **Pew Research Center Methods** team for working with survey data."

COVID-19 Data Science Resources

Academic Data Science Alliance from April 07, 2020

"The **Academic Data Science Alliance** is working with partners to pull together data and data science resources related to the COVID-19 pandemic. This is a living list of resources and we welcome additions, suggestions, and collaborations. Please send additions, corrections, comments, and suggestions to us using this feedback form."

Brain Imaging Data Structure Index | Brain Imaging Data Structure

International Neuroinformatics Coordinating Facility from January 02, 2020

"Neuroimaging experiments result in complicated data that can be arranged in many different ways. So far there is no consensus how to organize and share data obtained in neuroimaging experiments. Even two researchers working in the same lab can opt to arrange their data in a different way. Lack of consensus (or a standard) leads to misunderstandings and time wasted on rearranging data or rewriting scripts expecting certain structure. With the Brain Imaging Data Structure (BIDS), we describe a simple and easy to adopt way of organizing neuroimaging and behavioral data."

Nine simple steps for better-looking python code

Towards Data Science, Vladimir Iglovikov from April 08, 2020

"By simple, I mean easy to implement, less than 5 minutes on each step, and the changes should force you to do things better." ... "1. Use version control to track changes in your code"

CARRERS

Tenured and tenure track faculty positions

Director of the Lucy Institute

University of Notre Dame, Lucy Family Institute for Data & Society; South Bend, IN

Call for Assistant Professor positions – WASP-HS

Umeå University, Department of Computing Science, and Wallenberg Foundations, Wallenberg AI, Autonomous Systems and Software Program on Humanities and Society (WASP-HS); Umeå, Sweden

Open Rank Data Science General Faculty

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

Open Rank Data Science General Faculty: Data Warehousing and Cloud Computing Focus

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

Open Rank Professor of Data Science - Data Analytics and Data Systems Engineering

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

Open Rank-General Faculty in Computer Science & Data Science

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

Full-time, non-tenured academic positions

Center for Data Science Clinical Faculty

New York University, Center for Data Science; New York, NY

Postdocs

NCES Learning Engineering Fellow

University of Pennsylvania, Learning Agency Lab; Philadelphia, PA

Full-time positions outside academia

Customer Relations Specialist

Center for Open Science; Charlottesville, VA

OSF and Registries Product Owner

Center for Open Science; Charlottesville, VA

Junior Data Scientist

Theta Equity Partners; New York or San Francisco

Science Writer

West Big Data Hub; San Diego, CA, or Remote

Chief Data Officer

Department of Health And Human Services, Centers for Disease Control and Prevention; Atlanta, GA

Geographic Information Systems Specialist, Data Operations, Geo

Google; Mountain View, CA

Data Scientist

Thresher; Arlington, VA

Internships and other temporary positions

Urgent: Build, evaluate, and deploy of COVID models for clinical applications

New York University, Langone School of Medicine; New York, NY

One-year visiting professor

MIT Media Lab, Comparative Media Studies; Cambridge, MA

Introducing the SAGE Ocean Fellowship

SAGE Ocean; Remote

Data and Technology Advancement (DATA) National Service Scholar Program

National Institutes of Health; Bethesda and Rockville, MD, or Research Triangle Park, NC

DATA Scholar, BioData Catalyst

National Institutes of Health, National Heart, Lung, and Blood Institute (NHLBI); Bethesda, MD

DATA Scholar, cloud architect/engineer

National Institutes of Health, National Institute of Mental Health (NIMH) and National Institute of Drug Abuse (NIDA); Bethesda, MD

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 197.