



## Data Science Community Newsletter

2 April 2020

Data Science Community Newsletter Issue 196

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

The most interesting story this week is not about COVID-19. It's about [the ethical implications of computing at the brain-computer interface](#), especially as technologies are now good enough to read people's thoughts in a limited way. **The Morningside Group**, composed of 25 global neuroscience scholars, released a proposal for five NeuroRights:

- + **The Right to Personal Identity:** Boundaries must be developed to prohibit technology from disrupting the sense of self or blurring the line between a person's internal processing and external technological inputs;
- + **The Right to Free Will:** People should control their own decision making, without manipulation from external neurotechnologies;
- + **The Right to Mental Privacy:** Data obtained from measuring neural activity should be kept private, and the sale, commercial transfer, and use of neural data should be strictly regulated;
- + **The Right to Equal Access to Mental Augmentation:** Global guidelines should be established to ensure the fair and just access to mental-enhancement neurotechnologies that could further increase societal and economic disparities; and
- + **The Right to Protection from Algorithmic Bias:** Countermeasures to combat bias should be the norm for machine learning when used within neurotechnology devices, and algorithm design should include input from user groups to counter bias.

These are serious, well constructed principles, some of which are more social than neurological, including the right to equal access to mental augmentation. I will certainly continue to pass along updates from this thoughtful, but largely invisible think tank, The Morningside Group. The alteration and augmentation of the brain could catalyze an immense schism between the haves and have nots that would make our current economic inequality appear quaint.

**Matt Salganik** was a co-author on the biggest Kaggle-style competition in social science: the Fragile Families Challenge. The Challenge was a multi-year project that used data provided about 5,000 children born in 20 large American cities between 1998 and 2000 to accurately predict six later life outcomes like GPA, grit, eviction, and three others. [Salganik deflates the AI hype](#) by pointing out, "we have hundreds of participants and a rich data set, and even the best AI results are still not accurate." (Before you get your knickers in a twist, know that Salganik is not alone; he had [111 co-authors](#).) One co-author, **Sara McLanahan** also threw cold water on social science, "Either luck plays a major role in people's lives, or our theories as social scientists are missing some important

variable." The machine learning approaches performed only marginally better than linear and logistic regression.

**MIT** and **The University of Texas at Austin** were **awarded \$10 million over five years** from the **NSF** through its Expeditions in Computing Program. The researchers will work in four domains - organic chemistry, cognitive science, cell biology, and network science - to develop data scientific approaches that can better represent insights. The NSF and **DARPA** have been interested in pursuing more explainable data science and AI for years.

**The University of Montana** received ~\$10 million for **a new population health center**, which isn't necessarily \*data science\* but it is relevant to current events.

**New York University** has also made an announcement about the growing **NYU School of Global Public Health**. The School will get a **large slice of prime Washington Square campus** real estate. Renderings have been released.

**University of Chicago** with some help from colleagues at the **University of Wisconsin-Madison** put together an **interactive map of COVID-19 infections by county in the US**. It's one of the better COVID visualizations to come online this week. The northeast corridor **Amtrak** route is a hot zone, for sure.

The **University of Washington Institute for Health Metrics and Evaluation** **has a useful COVID map** that forecasts regional hospital demands in the face of COVID-19. And the **COVID status map** produced by **Johns Hopkins Center for Systems Science and Engineering** is the go-to numbers resource for network news.

The **IEEE VR** conference will take place as **a remote-attendance event** facilitated by **Mozilla Hubs** while other conferences **try their own all-virtual formats**. **Microsoft** is **rethinking how the organization should do large-scale events**, both external and internal. It remains to be seen how the idea of 'conference' will change, possibly for the better (especially if we can reduce the footprint of these fossil fuel fests). I find it hard to attend virtual events that last more than a couple hours so the idea of webinar upon webinar is gross. Is anyone else with me? Many of us conference organizers are hard at work finding ways to deliver the best virtual events and are watching each other closely to see how experiments in being virtually together go over.

Free tuition programs and beefed up support for student health and safety programs **may be in jeopardy** as the economic downturn damages state's revenue. Today's students may get the austerity college experience even after the COVID crisis has receded, but what about faculty? I'm reminded of a 2017 **qualitative study** conducted by **Ruth Barcan** of academic careerists. She found, "people who truly and deeply loved academic work are much happier and healthier now that they are no longer doing it. Casual staff were demoralised by the precariousness, while tenured staff were disturbed by the increasingly corporatist ethos of the university and found the workload, as one participant puts it, 'close to not-doable'. Another sums it up like this: 'I genuinely have a life now that I

would not have if I had taken the academic path. I have time for friends, for sport, for life!" Barcan's study was completed before coronavirus. But with upheaval and isolation comes introspection and decision points. How many postdocs, adjuncts and research staff will reflect on their position and stop trying to ascend the ivory tower?

**Martin Dameris** of the **German Aerospace Center** is bearing witness to [a giant hole in the ozone over the Arctic](#) caused by a confluence of very cold temperatures and clouds containing chlorine and bromine from refrigerants and other industrial chemicals. It won't harm human health and it will close again, but the size of it is a once-in-a-lifetime event.

Collecting data about student performance and behavior is increasing quietly, occasionally behind the scenes in admissions offices and other administrative departments with an eye on revenue targets. This has raised hackles among a number of faculty and student groups who are [concerned](#) about protecting student's privacy, ensuring equitable data usage, and devising sensitive, productive ways to discuss data-driven insights with students that won't simply alienate them. To be sure, some students relish having quantified, visualized, contextualized, granular, immediate feedback about their performance, especially in large classes where access to the professor is limited. This is a space full of players and is one to watch more closely than I historically have.

**C3.ai**, **Microsoft Corp.**, the **University of Illinois at Urbana-Champaign (UIUC)**, the **University of California, Berkeley**, **Princeton University**, the **University of Chicago**, the **Massachusetts Institute of Technology**, **Carnegie Mellon University**, and the **National Center for Supercomputing Applications at UIUC** announced [two major initiatives](#). The first is run jointly by **UC-Berkeley** and **UIUC** and will initiate "a coordinated effort to advance the digital transformation of business, government, and society." That's quite a broad purview and I haven't been able to unpack it to understand what would be excluded. The other effort is a call for research proposals that, "embrace the challenge of abating COVID-19." Embrace is a strange verb to choose, given the nature of the challenge.

**Cornell** postdoc **Maithra Raghu** and former **Google** CEO **Eric Schmidt** published [a 36-page 'getting started with deep learning' guide](#) on **arXiv.org**. It aims to help those who have basic familiarity with machine learning understand domain-specific algorithms and other application-space challenges.

### **Company Data Science News**

**Quest Diagnostics**, one of America's two large for-profit medical lab companies, promised it would be able to perform far more COVID-19 tests than it has been able to complete. **CNN** is reporting that Quest had a [backlog of 160,000 cases on March 25th](#) which fell to 115,000 as of Wednesday night. Doctors are "burning through PPE on quarantined patients" while waiting up to [7 days to get test results back](#).

**The Broad Institute** situated in Cambridge between **MIT** and **Harvard** has joined other medical institutes by [performing 2000 tests a day](#).

This patchwork of testing is the best we can do, but labs forced to come online faster than they typically would, often without safeguards in place, may be leading to inaccuracies in testing. One group of doctors reported that up to one-third of their negative tests were probably false negatives, based on the way their patients progressed. This corresponds with evidence that genetic tests are [about 75% accurate](#), mostly producing false negatives when they fail.

Also this week in why-is-it-so-hard-to-test-Americans-for-coronavirus, **Abbott Labs** announced it had developed a rather stylish piece of hardware capable of completing a coronavirus test in 5-13 minutes. Those machines are not exactly rolling off the manufacturing line at warp speed. That's just not how research and development works. The Trump administration trumpeted the magnificence of private sector ingenuity while thinking of sending the new machines to rural areas that have few cases, [inciting an uproar](#) in harder hit areas. Detroit mayor **Mike Duggan**, facing a huge uptick in case load, woke outgoing Abbott CEO **Miles White** in the wee hours begging for testing machines and nabbed five. That's how this is going down, readers. Desperate reconfiguration of existing lab infrastructure and when that doesn't work, after hours begging.

**Fred Hutchinson Cancer Research Center** has [laid off 14 data scientists](#) as the medical research center disbands the Hutch Data Commonwealth in favor of the Hutch Data Core. The switch is a move back to the traditional shared core services that large life science research organizations have had in place for decades. Founding Director of the experimental collaboration setup, **Carly Strasser**, is now Director of Alliances and Data Strategy at FHCRC. The report notes that this had nothing to do with COVID-19, but that additional layoffs may be in the works if the country stays at home for more than three months.

**Microsoft's** venture arm **M12** has [divested](#) from **AnyVision**, an Israeli facial recognition start up and will avoid any future investments in facial recognition technology. The announcement came after reports that AnyVision's technology is being used by the Israeli government to profile and surveil residents of the West Bank. An M12 blog post explained, "the audit process reinforced the challenges of being a minority investor in a company that sells sensitive technology, since such investments do not generally allow for the level of oversight or control that Microsoft exercises over the use of its own technology. By making a global change to its investment policies to end minority investments in companies that sell facial recognition technology, Microsoft's focus has shifted." Microsoft has a robust ethics oversight team not replicated at many other companies, particularly not small venture capital firms.

**The Google Faculty Research Award** program will be discontinued. It will be replaced by two new efforts to support faculty, while also advancing Google's interests. The [Research Scholar Program](#) will support early career researchers (PhD +7 years max) "in fields relevant to Google." The second program - Award for Inclusion Research Program - is a pilot designed to support the work of researchers from underrepresented identity backgrounds.

Most of you have probably stumbled across the [data visualization](#) showing where the people who

congregated on a single Miami beach for spring break returned to the following week, put together by **Tectonix GEO**. That visualization proved that the Mississippi is the vacation destination dividing line - those to the east go to Florida - that the partiers could easily have helped distribute COVID-19 to half of America, and that privacy as we know it is in peril. As a data ethics scholar, I have long advocated for a kind of hybridity in ethical approaches: there are some lines we should never cross, but most of the time it makes sense to weigh up the benefits of using a technology before we make a final decision about whether we want to restrict use of data for privacy reasons. Public health is one of the more interesting arenas to have this conversation. Should companies? The government? Be able to surveil each of us everywhere we go? I would say no. (Chinese authorities would disagree.) But what if that kind of tracking slows the spread of a potentially fatal disease? (And, say, just hypothetically, that more reasonable testing and tracking schemas have been as effective as beating a drum with a spaghetti noodle?)

A team including the now semi-famous epidemiologist **Marc Lipsitch** of **Harvard T.H. Chan School of Public Health** has published a [five paragraph paper/essay](#) in **Science** on this exact set of questions. They write, "the protection of personal privacy must be paramount...we do not advocate the use of individual data." Instead, they argue that there are robust de-identification techniques - including using trusted data guardians who serve no other objective but to protect data - which should be used to track the movements of people using phone data. We can fight about whether any de-identification is ever 100% perfect, but the methods they are proposing would offer privacy protection to most while still allowing the public health experts to understand where mitigation efforts are failing. Just Thursday, **Florida Governor Ron DeSantis issued a stay at home order**.

Unnecessary side note: the mitigation efforts appear to be failing in Newport Beach and Laguna Beach. It seems the wealthy set has decided to rent beach houses and weather the stay-at-home order by not staying in their own homes or even in their beach homes. Bike rides and beach tussles for everyone! Per capita cases are [the highest in Orange County](#).

If you're suddenly working from home, please note that cybercrime of every type is way up. My day job is working in privacy for **Obsidian Security**. We were built to 'cover your SaaS' but had no idea how important our technology would become nearly overnight in this work-from-home revolution. Beleaguered IT workers, I hear you. **Allan Liska**, a cybersecurity consultant with **Recorded Future** put together a much more [comprehensive list of exploits and vulnerabilities](#). Protect yourself against ALL the viruses at this time.

## Government Data Science News

The **FBI** raided the home of **Baruch Feldheim** who had [hoarded 80,000 N95 medical masks](#) and was offering to sell them at a markup of 700%.

**Britian** toyed with herd immunity, but **Sweden** is going for it: [unfettered society, fulling cranking economy](#) during COVID. Schools are in session. Gyms are open. The government used a simulation that modeled a higher rate of asymptomatic infected people than the Brits did. We shall see how the Swedish experiment plays out.

**The Netherlands** has also avoided the lockdowns Italy, Spain, Germany, the US, the UK, and China put in place. Their choices are **driven by a disease model** with an R0 value just above 2 and the assumption that a fair percentage of the population will be infected, but asymptomatic. Evidence from cruise ships where all passengers **were tested** regardless of their symptoms does not support the assumption that large numbers of people have mild symptoms or none at all.

On the other end of the spectrum completely, **President Duterte** of the **Philippines** had to walk back a statement that those violating the stay-at-home order could be **fatally shot by local police**. One commentator noted that Amnesty International's outrage was overblown because everyone in the Philippines understood he only meant to target *leftists* violating the stay-at-home order.

And **Germany** is **taking steps towards emulating Singapore with a contact-tracking app** to contain COVID-19 spread. **Heinrich Hertz Institute** (HHI) is developing the app and wants to have it in place within a few weeks. **Justice Minister Christine Lambrecht** says the app will be voluntary and HHI promises 100% compliance with German personal data protection laws.

### **Extra Extra**

There is plenty of misinformation circulating in the wake of COVID-19. The best way to read about it is through satire. You'll appreciate **Noah Haber's** description of Dunning-Kruger-Effect-19 (DKE-19) an infectious malady that causes the infected population "to lack the ability to understand their lack of ability." He finds that it "**primarily affects men ages 24–36 working in tech, for reasons unknown to scientists who are unaccountably also men.**" Click through to enjoy the whole take-down. You probably won't regret it.

PSA: If you are a US citizen planning to use the federal tax filing extension, please note that **your state taxes may be due sooner** than July 15th.

### **Data Visualization of the Week**

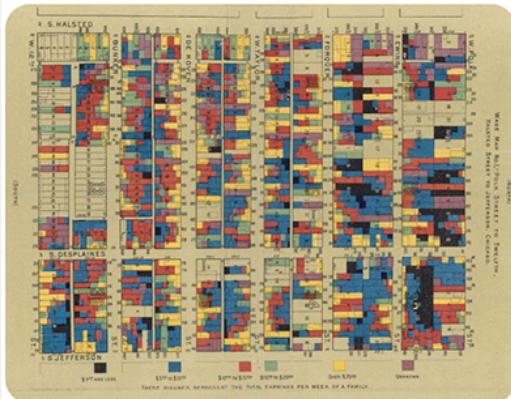
*Twitter, #WOMENSART from March 31, 2020*



#WOMENSART

@womensart1

Florence Kelley, US social/political reformer campaigned for children's/workers' rights, early 20thC, by creating 'slum maps' of Chicago showing demographic info used to campaign for change. Technically brilliant, they helped improve lives #womensart



10:37am · 31 Mar 2020 · Twitter Web App

3 Replies 101 Retweets 340 Likes



...

## Tweet of the Week

Twitter, @PedantDog from March 31, 2020



**Whole Wheat Dog**

@PedantDog

Replies to @vermontgmg

My 90 day trial of 2020 has ended.  
How do I cancel?

5:24pm · 31 Mar 2020 · Twitter for Android

13 Replies 192 Retweets 971 Likes



...

## EVENTS

## **MIT Covid19 Challenge**

**Online** April 3-5. "Build a solution for the COVID-19 crisis. **MIT** is hosting a series of challenges to empower YOU to take action on the COVID-19 crisis." [application required]

## **NYC Media Lab, Combine Virtual Demo Day 2020**

**Online** April 3, starting at 9:30 a.m. "Join us for the big day... **NYC Media Lab's** Combine cohort will pitch their startup concepts virtually! The Combine 2020 startups range from AR/VR to eye-tracking video technology and solar powered chargers. Attend to see how 12 weeks of venture development and market validation shaped each team's ideas." [registration required]

## **Urban Institute's Data@Urban Digital Discussions**

**Online** April 3, starting at 2 p.m. EDT. Participants: **Enrico Bertini** and **Moritz Stefaner**. [registration required]

## **The Web Conference 2020**

**Online** April 20-24. The organizing committee "has been working very hard behind the scenes in making the 2020 edition a safe and successful one in considering of that the COVID-19 is an extraordinary global public health issue." [registration required]

## **IPAM Workshops**

**Online** During April and May. "Transitioning **IPAM's** spring workshops to online means more available spots for participants! We are happy to announce registration is open again for the popular High Dimensional Hamilton-Jacobi PDEs workshops." [registration required]

## **Knowledge Graph Conference**

**Online** May 6-7. "Knowledge Graphs form an organized and curated set of facts that provide support for models to help understand the world. This conference gathers technology leaders, researchers, academics, vendors — and most important, practitioners who know the discipline." [\$\$\$]

## **DEADLINES**

### **Contests/Award**

#### **COVID-19 Open Research Dataset Challenge (CORD-19)**

"A list of our initial key questions can be found under the Tasks section of this dataset. These key scientific questions are drawn from the NASEM's SCIED (**National Academies of Sciences, Engineering, and Medicine's Standing Committee on Emerging Infectious Diseases and 21st Century Health Threats**) research topics." ... "Many of these questions are suitable for text mining, and we encourage researchers to develop text mining tools to provide insights on these questions." Round 1 submission deadline is April 16.

#### **12th IoT/WT Innovation World Cup**

"The world's leading competition for IoT and wearable tech start-ups, scale-ups and SMEs, is welcoming techpreneurs from around the world to make their innovations striking!" Deadline for submissions is September 23.

### **Conferences**

#### **ICML 2020 Workshop on Uncertainty & Robustness in Deep Learning - Call for Papers**

**Online** "There is growing interest in making deep neural networks more reliable." ... "This workshop will bring together researchers and practitioners from the machine learning communities, and highlight

recent work that contributes to addressing these challenges." Deadline for submissions is May 22.

### **Studies/Surveys**

#### **wesurvived.nyc - Dear New Yorkers, Do not let small businesses close down!**

Please help identify neighborhood changes and contribute your memories about beloved places! It will help protect small businesses from massive neighborhood changes and advocate for commercial rent stabilization.

#### **@RockefellerUniv scientists are urgently seeking people who have recovered from #coronavirus infection**

Your cells may hold the key to fighting #COVID19 !!

INFO: <http://clinicalstudies.rucares.org/coronavirus>

### **RFP**

#### **HHMI Opens New Program for Medically Trained Scientists**

The "Medically Trained Scientists Program (MTS) seeks early career scientists with knowledge of human biology to move discovery science forward. MTS will ultimately invest up to \$120 million in up to 40 fellows. MTS fellows are creative and highly committed individuals whose research will impact significant biological problems. Fellows are committed to careers in academic research and will receive long-term support as they transition through mentored postdoctoral training to independent faculty." Deadline for applications is August 11.

### **TOOLS & RESOURCES**

#### **COVID-19 Mobility Data Network**

*Caroline Buckee and Andrew Schroeder from March 23, 2020*

"We are a network of infectious disease epidemiologists at universities around the world working with technology companies to use aggregated mobility data to support the COVID-19 response. Our goal is to provide daily updates to decision-makers at the state and local levels on how well social distancing interventions are working, using anonymized, aggregated data sets from mobile device, and to provide them analytic support for interpretation."

#### **USC Researchers Release Public Coronavirus Twitter Set for Academics**

*University of Southern California, Viterbi School of Engineering from March 25, 2020*

"**Emilio Ferrara** and **Kristina Lerman**, the principal researchers on this project, have a history of studying social media and bots to understand how misinformation, fear and influence spread online."

#### **Fair Use 101: Librarians' Advice for Educators**

*American Library Association, ilovelibraries blog, lsimon from March 23, 2020*

"**Nancy Sims** and other librarians specializing in copyright have put together a guide for instructors about navigating fair use for remote teaching. In copyright law, fair use refers to circumstances where people can legally use copyrighted works without permission. Fair use can apply to many different situations; historically, courts have often allowed people to use copyrighted materials for educational purposes or to serve the public's interest."

#### **Learn D3: Introduction**

*Observable*, Mike Bostock from March 24, 2020

"**Observable** is the ideal environment for learning D3 because it simplifies code with dataflow, like a spreadsheet. As you edit, cells run automatically for rapid feedback. You can add interaction or animation with almost no code! We'll cover Observable's quirks and features as we go along."

### **covid-19-data: An ongoing repository of data on coronavirus cases and deaths in the U.S.**

*GitHub - nytimes* from March 28, 2020

"**The New York Times** is releasing a series of data files with cumulative counts of coronavirus cases in the United States, at the state and county level, over time. We are compiling this time series data from state and local governments and health departments in an attempt to provide a complete record of the ongoing outbreak."

### **Introducing Fiber, a new open source platform for distributed machine learning, especially population-based methods like Enhanced POET.**

*Twitter, Jeff Clune* from March 26, 2020

"Program locally with a standard multiprocessing API then deploy to thousands of workers on any cluster. <https://uber.github.io/fiber/introduction> Led by @\_calio"

### **The COVID-19 Hospital Impact Model for Epidemics (penn-chime.phl.io)**

*GitHub - CodeForPhilly* from March 29, 2020

"The CHIME (COVID-19 Hospital Impact Model for Epidemics) Application is designed to assist hospitals and public health officials with understanding hospital capacity needs as they relate to the COVID pandemic. CHIME enables capacity planning by providing estimates of total daily (i.e. new) and running totals of (i.e. census) inpatient hospitalizations, ICU admissions, and patients requiring ventilation."

### **Open Access to ACM Digital Library During Coronavirus Pandemic**

*ACM, ACM Bulletins* from March 30, 2020

"We believe that **ACM** can help support research, discovery and learning during this time of crisis by opening the ACM Digital Library to all. For the next three months, there will be no fees assessed for accessing or downloading work published by ACM."

### **vega\_vis @vega\_vis To celebrate the Vega-Lite 4.9 release, we now have a new logo!** **(1/6)**

*Twitter, @vega\_vis* from April 01, 2020

"For the new features, Vega-Lite now supports arc marks, which let you create your favorite pie and donut charts (and of course, reproduce our new logo)"

### **biomed-sanity**

*Andrej Karpathy* from April 01, 2020

"Built in spare time by **Andrej Karpathy** following **biorxiv** and my older project arxiv-sanity to help with influx of papers on COVID-19, currently indexing 1076 papers."

### **The A-Z of AI**

*Google, Oxford Internet Institute* from March 25, 2020

"This A-Z guide offers a series of simple, bite-sized explainers to help anyone understand what AI is, how it works and how it's changing the world around us."

## **CAREERS**

### **Tenured and tenure track faculty positions**

#### **Assistant/Associate Professor and Director of the Spatial Analysis & Visualization Initiative**

Pratt Institute, School of Information

#### **Director of the Lucy Institute**

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

## **Postdocs**

### **Postdocs/Research Scientists in Digital and Computational Demography**

Max Planck Institute for Demographic Research, Lab of Digital and Computational Demography;  
Rostock, Germany

#### **Postdoctoral Research Associate**

University of Illinois at Urbana-Champaign; School of Information Sciences; Champaign, IL

#### **NCES Learning Engineering Fellow**

University of Pennsylvania, Learning Agency Lab; Philadelphia, PS

## **Full-time positions outside academia**

#### **Senior UX Designer**

CCC Information Services; Chicago, IL

#### **UX Researcher**

TomTom; Farmington Hills, MI

#### **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

## **Internships and other temporary positions**

### **Internship/Temporary position: Applied Deep Learning/AI (COVID-19)**

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

#### **Visiting Professor (1 year)**

MIT Media Lab, Comparative Media Studies; Cambridge, MA

[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](mailto:laura.noren@nyu.edu) and [brad.stenger@gmail.com](mailto:brad.stenger@gmail.com). We retain curatorial discretion.

**Data Science Community Newsletter Issue 196.**

