

Data Science Community Newsletter

Data Science Community Newsletter features journalism, research papers and tools/software for December 21, 2020

Please let us ([Micaela Parker](#), [Steve Van Tuyl](#), [Laura Noren](#), [Brad Stenger](#)) 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](#) .

Happy Holidays

This is the final newsletter for calendar 2020. We are grateful for your attention and look forward to the year ahead. For these final days of the year, we wish you a fun, safe and socially-distant-but-still-festive holiday season.

Academic Data Science News

The **University of Michigan** has [hired an outside consulting firm, Guidepost Solutions](#), to help overhaul campus culture. This is the "way forward" that was [promised in August](#) when the university President and Board were deliberating on their response to a report documenting decades of sexual misconduct by a former university Provost. Michigan's Computer Science department and Athletic department have also experienced recent sexual misconduct scandals ([link](#), [link](#)). And a 2019 [student survey](#) was a no-confidence statement with respect to the school's response to sexual misconduct complaints. This slow, deliberative path to progress is similar to what the **University of Rochester** went through following that school's 2017 Cognitive Science department [harassment scandal](#), in which after some initial falling on swords, the [momentum for change faded](#) as student attention waned, eventually vanishing as campus populations churn through their education programs. Process and decision-making are expressions of organizational culture. It's emblematic that the **University of Texas at Austin** took seven years to [stop using GRADE](#), a biased machine learning system that was part of Computer Science department admissions. Data Science is central to the evolution of research universities and the motivation for change needs to be intrinsic for lasting change to occur. As Gandhi [sort-of said](#), "Be the change you wish to see."

The state of our water-based resources and all that depends on them is much in the news, from the small - 3D-printed sea turtle "eggs" [equipped with GPS tracking](#) to help reduce poaching – to the larger – data from [a half-century of ocean observation](#) taken at a **University of Alaska** oceanographic monitoring station – to the largest, as **Yale University** launches the [Planetary Solutions](#)

Project. Another thing that's a half-century old? California's groundwater management guidelines, that's what. **McGill University** researchers have analyzed [more than 200,000 groundwater samples](#) in an effort to help the multi-billion-dollar agricultural industry adopt sustainable guidelines for groundwater management. Groundwater dependent ecosystems and wildfires might seem unlikely bedfellows but loss of groundwater creates more tinder for fires, and in turn, fires can pollute groundwater. California has both groundwater issues and devastating wildfires. Fortunately researchers at the **U.S. Department of Energy's Pacific Northwest National Laboratory** are using ML to [better understand the conditions](#) that lead to those devastating fires. To help us keep track of all of this environmental data, **Cambridge University Press** has [launched *Environmental Data Science*](#), focused on using AI to help us monitor changes in the environment.

Reproducibility is a hot topic in several domains, notably [psychology and ecology](#) (the latter's push for more reliable research being inspired by the former's), and now from **University of Michigan** a new effort on [research reproducibility in data science](#). Meanwhile sociologists are looking at AI [from the perspective of sociology](#), and **ICML** has been running [experiments on the review process](#). The **Open Data Inventory** (ODIN) highlights [the state of open data](#) in 187 countries worldwide. Good news: COVID did not slow down the effort to make open data a priority. Not so good news: openly published data on gender and encouraging open data in lower income countries have made less progress.

Universities continue to spawn partnerships to better understand and manage COVID-19. **Carnegie Mellon's Delphi Research Group** and **Change Healthcare** have announced [updates to COVIDcast](#), collecting real-time county-level data on self-reported COVID-19 symptoms nationwide and now adding de-identified COVID-19 claims to its combination of surveying, testing and mobility data. **University of Washington** and **Washington State Department of Health** have announced WANotify, [a state-wide COVID notification app](#) that relies on voluntary reporting of positive test results in order to notify other users if they have been potentially exposed. **Cornell University** physicians and statisticians joined forces to [create a secure database](#) with medical data on more than 4,000 patients who had clinical symptoms of COVID-19 and were seen at one of three **New York-Presbyterian Hospital** campuses. From research using this registry alone, over 20 peer-reviewed papers have been published.

In a previous newsletter we highlighted how universities are analyzing sewage to better monitor COVID spread, and **Stanford** researchers have [published a](#)

[paper](#) on their findings.

Ibram Kendi and **Azer Bestavros** are [founding researchers](#) at the new **Racial Data Lab at Boston University**. The Lab's first project will be the Racial Data Tracker, expected to become the nation's largest public online collection of racial inequity data.

Cornell [will establish](#) the **Ann S. Bowers College of Computing and Information Science** and construct a new building to accommodate faculty hiring and increasing CIS student enrollment.

University of Toronto [launched](#) the **Temerty Centre for AI Research and Education in Medicine**. **Muhammad Mamdani** directs the multi-disciplinary AI and medicine center, er uh, centre.

National Research Council of Canada and **Simon Fraser University** [will partner](#) to increase artificial intelligence research and training capacity on Canada's West Coast.

LaSalle University in Philadelphia [will start](#) a Master of Science in Business Systems and Analytics program with the first students arriving in the Fall.

Editor's Picks

Google's firing of AI bias researcher **Timnit Gebru** remains big news, and internal battles have escalated as more Google AI researchers [demand new policies](#). Opinion seems to be running strongly anti-Google, from pointing out how far they have [strayed from their motto](#) "don't be evil" to the [dark side of corporate research](#) to an unflattering look at their [publishing review process](#). Of course there's always some sun shining in Google-land – their UK-based AI lab **DeepMind's** breakthrough technology **AlphaFold**, which predicts the shape of proteins, raised revenue and convinced Google to [waive a \\$1.5 billion loan](#), and the effort to create [a smarter food system](#), known as Project Delta, has moved to **X**, the "moonshot factory" and Google subsidiary.

Two longform articles argue that the pandemic has been a turning point for the collective scientific enterprise. **Ian Sample**, [writing](#) in **The Guardian**, points to the January 2020 publication of the complete COVID-19 genetic code that was sequenced by **Yong-Zhen Zhang's** team at **Fudan University** as the launch of this "unprecedented effort to test, treat and ultimately vaccinate against COVID-19," estimating that "in the last 11 months, probably 10 years' work have been done." **Ed Yong** has earned praise covering the awful progression of COVID in the United States for **The Atlantic**. Yong [cites](#) bibliometrics – the 74,000 COVID-related papers in **PubMed**, the 30,000 submissions to **New England Journal of Medicine** – as evidence of an unprecedented shift in

scientific priorities. A lens on [corporate research re-prioritizing for coronavirus](#) at **Facebook AI Research** is yet another longform article, this in the online magazine **Protocol**.

The current state of testing and public health participation show that progress continues, but challenges remain. **MIT's Institute for Data, Systems, and Society** debuted [WhenToTest.org](#), an online calculator for [cost analysis](#) of COVID-19 testing tradeoffs. **Texas State University** professor **Rodney Rohde** warns that [burnout is afflicting U.S. testing laboratory workers](#) at alarming rates. **University of California-Merced** economists [surveyed undergraduates](#) to understand who complied with public health orders for COVID-19 and who did not. Risk perception was a critical factor, and media messages that amplify motivations like selfishness among guideline violators, or fear among preventive supporters, don't help.

In a Capitalist society, when do you know that an issue has become mainstream? When Wall Street shows interest. The **Federal Reserve** has formally joined a global group focused on [how climate change could impact the financial world](#) and it has already been noted that **President-Elect Biden's emerging approach to climate policy** will affect global energy markets as well as giving needed support to clean energy businesses and technologies. Water futures and wildfire risk are just two of the many distinct markets that are being addressed. For the first time in U.S. history, water futures began [trading as a commodity](#) – California water futures, indicating increasing fears of water scarcity. This is the state that this year suffered through a devastating wildfire season after an 8-year drought – and where antiquated rules for calculating wildfire risk might be discouraging innovation, [according to predictive modelers](#) in the insurance industry. The hope is using those predictive models would actually encourage a reduction in wildfires.

The **Federal Trade**

Commission ordered **Facebook, Twitter, Amazon, YouTube** and **ByteDance** (parent of **TikTok**) to [provide detailed information](#) "on how they collect and use consumers' personal data and how their practices affect children and teens." The information is due in late January and will be the source for a comprehensive FTC study. Children and teenagers are also [the focus](#) of the federal Artificial Intelligence Education Act, which is being included in the current National Defense Authorization Act. The NDAA, having passed the **U.S. Congress** with [a veto-proof majority](#), is [rife with AI](#), also containing the Artificial Intelligence Initiative Act, the National AI Research Resource Task Force Act, the Artificial Intelligence for the Armed Forces Act, and the Deepfakes Report Act. The **European Union** is [overhauling how it](#)

[regulates](#) Digital Services and Digital Markets with significant implications on content and on consumers' data. **China** [has drafted](#) its Personal Information Protection Law, set to become a centerpiece for data governance in that country. And **Microsoft** will make oversight a two-way street, launching an AI initiative to [uncover government corruption](#).

Astronomy.com [posted a superb history](#) of Project Stargazer, a 1960s-era high-altitude balloon research project. The article concludes with a progress report on 2019's SuperBIT balloon imaging telescope.

Research News

Twenty percent of Americans were wearing fitness trackers or sport watches in 2019, according to **Pew Research data**. That number has surely climbed and will go still higher as studies continue to show the early warning potential that these devices offer, for example [against COVID-19](#). In some cases, like COVID, the personal health data is destiny, but not so in other cases and **University of Virginia Medical Center** researchers are doing extensive work to understand the [effect of exercise on disease](#). **Evidation Health** is a California-based startup with a collaborative platform for bringing individuals into health research studies using personal smart devices and wearables; the company [recently partnered](#) with the **American College of Cardiology** on Achievement for Heart Health, a program for engaging individuals and researchers around cardiovascular health and healthcare. **Google** has also [created a Health Studies platform and app](#) for enrolling large populations in longitudinal health studies, something that **Apple** [has been doing](#) since November 2019. Everything is not rosy, however. A [recent study](#) by **Drexel University** public health researchers found that chronic exposure to stressors, attributable to systemic racism, erodes ongoing health among minorities and immigrants. The 2019 Pew data points out that income is a significant barrier to wearable health technology adoption. The people who stand to benefit most from personal health technologies are being left out.

Just-announced **Pew Research survey results** show that half of those Americans surveyed about COVID-19 contact tracing did not trust the federal government to keep their personal data secure. Only 41% trusted public health organizations and just 19% trusted doctors and health care providers.

2019 Nobel Prize-winner **Frances Arnold** of **Caltech** is [excited about the progress](#) her lab is making using machine learning to guide directed evolution of engineered enzymes.

University of Pittsburgh tissue engineers led by **Mo Ebrahimkhani** [use machine learning algorithms](#) that help to approximate the complexity of

vascular systems in liver organoids, a tissue scaffold that grows into a functional organ after transplant.

The best rundown of **NeurIPS** was [the summary](#) by **Jeremy Kahn** at **Fortune.com**.

MIT CSAIL researchers invented algorithms that [infer human goals and plans](#), including ones that lead to mistakes. And **Garret Kenyon** wrote [an insightful look](#) at how more "sleep" might benefit AIs for **Scientific American**. After a year like 2020, more sleep sounds really, really good.

Tweet of the Week

Twitter, Brian Moore from December 16, 2020

The image shows a screenshot of two tweets from Brian Moore (@lanewinfield). The first tweet, dated Dec 16, features a video of a man sitting at a desk with two monitors, one of which shows a video call. The video has 1.2M views and 53.4K likes. The second tweet, dated Dec 17, is a text-based tweet sharing a GitHub repository named 'lanewinfield/zoomout', which is described as a pull switch for video calls. It has 846 likes and includes a link to the repository on github.com.

Brian Moore @lanewinfield · Dec 16
I always awkwardly struggle to get to the end call button on video calls. So I made this

1.2M views 0:19 / 0:22

537 10.5K 53.4K

Brian Moore @lanewinfield · Dec 17
all files on github, if you're curious and/or want to make one yourself:

lanewinfield/zoomout
a pull switch (or BYO button) that gets you out of video calls, quick - lanewinfield/zoomout
[github.com](#)

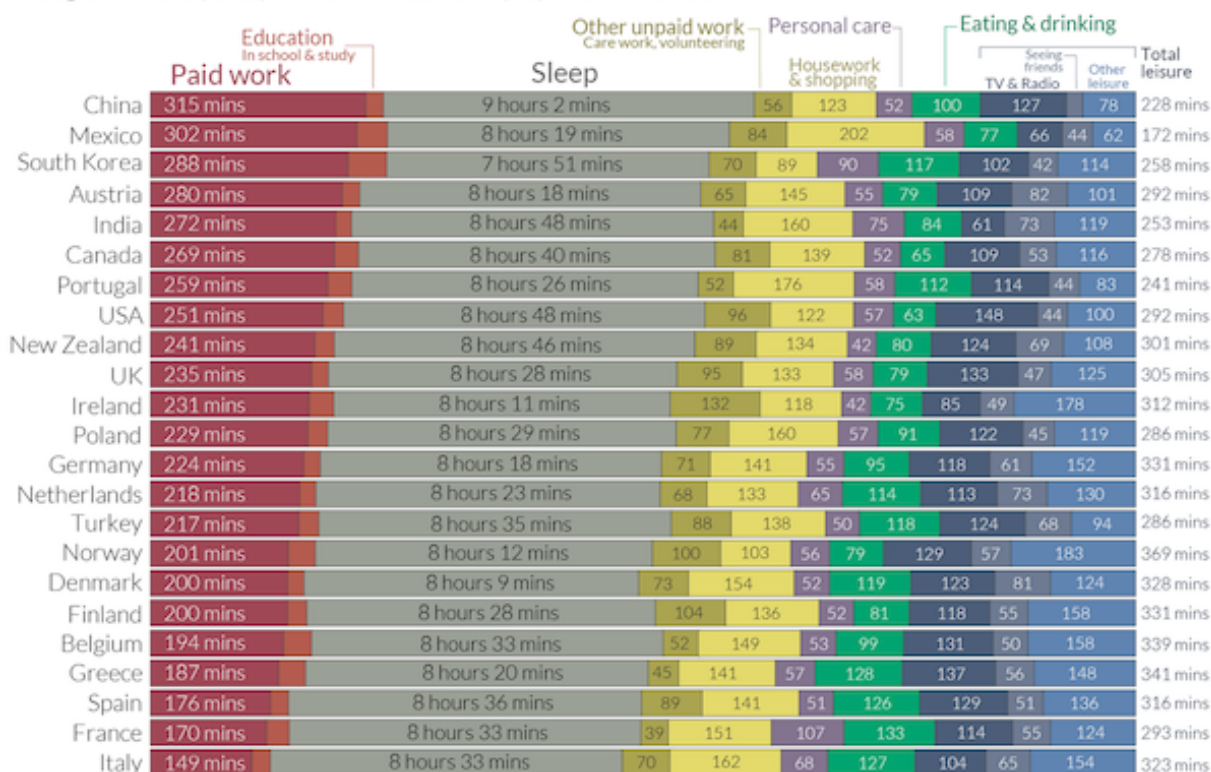
13 57 846

Data Visualization of the Week

Visual Capitalist, Iman Gosh from December 11, 2020

How do people spend their time?

Averages of minutes per day from time-use diaries for people between 15 and 64.



Data source: OECD Time Use Database, Gender Data Portal. For most countries surveys were conducted between 2009 and 2016, but surveys for some countries are older. OurWorldinData.org - Research and data to make progress against the world's largest problems. Licensed under CC-BY by the author Esteban Ortiz-Ospina.

Events

AI DEBATE 2 - Moving AI Forward: An Interdisciplinary Approach

Online December 23, starting at 4 p.m. Eastern time. Confirmed speakers: **Ryan Calo, Yejin Choi, Daniel Kahneman, Celeste Kidd, Christof Koch, Luis Lamb, Fei-Fei Li, Adam Marblestone, Margaret Mitchell, Robert Osazuwa Ness, Judea Pearl, Francesca Rossi, Ken Stanley, Rich Sutton, Doris Tsao, Barbara Tversky** and more TBA. [rsvp required]

Networks 2021: A Joint Sunbelt and NetSci Conference

Washington, DC July 6-11, 2021. "We expect this to be the largest networks conference ever held. It will combine the annual meeting of the **International Network for Social Network Analysis** (Sunbelt XLI), and the annual meeting of the **Network Science Society** (NetSci 2021)." Deadline for abstracts submissions is January 24, 2021.

COVID19 Precision Immunology App-a-thon

"**precisionFDA** calls upon the scientific and analytics communities to develop innovative and user-friendly approaches to improve the understanding of the relationship between personalized immune repertoires and COVID disease-relevant phenomena and their associated external factors by accurate

interpretation of rapidly shared datasets in public databases, such as iRECEPTOR Gateway. This precision immunology app-a-thon phase I will launch on Nov. 30, 2020 and end on Jan. 29, 2021." [sign up required]

Deadlines

Contests/Award

[NOAA Launches Crowdsourcing Competition for Better Forecasting of Magnetic Field with DrivenData and HeroX](#)

"The **National Oceanic and Atmospheric Administration** (NOAA) in partnership with the **NASA Tournament Lab** today announced its crowdsourcing challenge, 'MagNet: Model the Geomagnetic Field.' The challenge, which is being implemented by **DrivenData** and **HeroX**, seeks to mitigate the impact of geomagnetic storms on navigation systems through improved forecasting by increasing the accuracy in real-time magnetic field modeling and reducing errors in the magnetic navigation systems." Deadline for submissions is February 12, 2021.

Conferences

[Workshop on Transparency and Explanations in Smart Systems \(TExSS\)](#)

Online April 13-17, 2021, held in conjunction with **ACM Intelligent User Interfaces** (IUI). Deadline for paper submissions is December 23.

Education Opportunities

[DeepMind Scholarship Program for Underrepresented Students](#)

New York University Center for Data Science "is thrilled to once again partner with **DeepMind** in offering a series of scholarships aimed at increasing representation of underrepresented students in higher education AI programs. Increasing representation in AI is a significant opportunity to bring diverse voices to the conversation, and ultimately shape continuing developments in AI as a technology that benefits everyone."

[Princeton CITP Emerging Scholars in Technology Policy](#)

"The goal of the CITP Emerging Scholars program is to train scholars who have completed their undergraduate studies but require more coursework or research preparation to improve their chances of admission to highly competitive Ph.D. programs or another competitive career path. The program provides intensive research and/or work experience, coursework, and mentoring. The ideal outcome for participants is to either enter a competitive graduate program or to find an impactful placement in government, nonprofits, or the private sector." Deadline for applications is January 10, 2021.

[Carnegie Mellon HCII Seeks Researchers for Summer 2021 Program](#)

"Summer 2020 was an unprecedented year for all. The ongoing pandemic prevented us from welcoming our students to Pittsburgh for our usual in-person program. However, we quickly made a new plan and virtually welcomed nearly 40 students for our first ever remote program! We may not know what the

future holds in regards to whether the program will be held in-person or remotely in summer 2021, but we do know that we will provide you with a high quality summer research experience." Deadline for applications is January 15, 2021.

RFP

[Art + Technology Lab 2021 Call for Proposals](#)

"While there is a preference for projects that explore emerging technology, prior technological experience or knowledge is not required. Artists who have not used advanced technology in their practice are encouraged to consider how technology applications might build upon and expand the trajectory of their work. Recipients need not be located in or near Los Angeles. The Lab welcomes proposals for projects that are presented outside of the bounds of the museum campus, including conceptual projects and projects that unfold in virtual, online, extraterrestrial locations." Deadline for submissions is February 25, 2021.

[I've been there. Fighting stereotypes in the world of science.](#)

"As the Deputy Assistant Director of **NSF's Social, Behavioral and Economic Sciences directorate**, my colleagues and I looked inward as we considered the small number of financial awards that our organization makes each year to minority-serving institutions. As a result, we launched **Build and Broaden**, a funding opportunity inviting proposals aimed at fostering partnerships and kick-starting research collaborations in the social, behavioral and economic sciences among minority-serving institutions while enhancing the research infrastructure and capacity of those institutions." Deadline for proposals is March 5, 2021.

Tools & Resources

[Jennifer Aaker & Naomi Bagdonas, "Harnessing Humor as a Secret Weapon"](#)

YouTube, Stanford Graduate School of Business from December 04, 2020

"In 'A New Type of Leader: Harnessing Humor as a Secret Weapon,'

Professor **Jennifer Aaker** and Lecturer **Naomi Bagdonas** draw from their experience in teaching the GSB course 'Humor: Serious Business' and their latest book *Humor, Seriously: Why Humor Is a Secret Weapon in Business and Life*, in discussing the power of humor and how you can harness it as a leader to disarm, encourage bold actions and give other people courage to do the same."

[R Consortium Providing Financial Support to COVID-19 Data Hub Platform](#)

R-bloggers from December 14, 2020

"The **R Consortium's COVID-19 Working Group** is providing a new home for

the COVID-19 Data Hub Project. The goal of the COVID-19 Data Hub is to provide the worldwide research community with a unified dataset by collecting worldwide fine-grained case data, merged with external variables helpful for a better understanding of COVID-19."

Fire data at your fingertips

Twitter, NIST from December 10, 2020

"**NIST** researchers pulled together a decade's worth of fire experiments into a robust online resource for everyone to access. Explore the Fire Calorimetry Database"

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.

Data Science Community Newsletter Issue 209