NYU Data Science Community Newsletter features journalism, research papers, events, tools/software, and jobs for July 31, 2015.

Please Let us (Laura Noren, Brad Stenger) know if you have something to add to the newsletter. We are grateful for the generous financial support from the Moore-Sloan data Science Environment and NYU’s Center for Data Science.

NYU Data Science Community Newsletter Issue 012.

Data Science News

Let's Make Gender Diversity in Data Science a Priority Right from the Start

_PLOS Biology, Francine Berman_ from July 27, 2015

The emergent field of data science is a critical driver for _innovation_ in all sectors, a focus of tremendous workforce development, and an area of increasing importance within science, technology, engineering, and math (STEM). In all of its aspects, data science has the potential to narrow the gender gap and set a new bar for inclusion. To evolve data science in a way that promotes gender diversity, we must address two challenges: (1) how to increase the number of women acquiring skills and working in data science and (2) how to evolve organizations and professional cultures to better retain and advance women in data science. Everyone can contribute.

Hadley Wickham, the Man Who Revolutionized R

_Pricenomics_ from July 24, 2015

If you don’t spend much of your time coding in the open-source statistical programming language R, his name is likely not familiar to you -- but the statistician _Hadley Wickham_ is, in his own words, “nerd famous.” The kind of famous where people at statistics conferences line up for selfies, ask him for autographs, and are generally in awe of him. “It’s utterly utterly bizarre,” he admits. “To be famous for writing R programs? It’s just crazy.”

Wickham earned his renown as the preeminent developer of packages for R, a programming language developed for data analysis. Packages are programming tools that simplify the code necessary to complete common tasks such as aggregating and plotting data. He has helped millions of people become more efficient at their jobs -- something for which they are often grateful, and sometimes rapturous. The packages he has developed are used by tech behemoths like _Google, Facebook_ and _Twitter_, journalism heavyweights like the _New York Times_ and _FiveThirtyEight_, and government agencies like the _Food and Drug Administration (FDA)_ and _Drug Enforcement Administration (DEA)_.

Truly, he is a giant among data nerds.

GitXiv.com - Collaborative Open Computer Science.

_reddit.com/r/MachineLearning_ from July 28, 2015

Reddit weighs in on _GitXiv_, a new platform for collaborating on Open Source computer science

Big Data & the Risky Eclipse of Statisticians

_Hacker Rank_ blog from July 20, 2015

If statisticians have historically been leaders of data, why was there a need for a brand new breed of data scientists? While the world is exploding with bounties of valuable data, statisticians are strangely working quietly in the shadows. Statistics is the science of learning from data, so why aren’t statisticians reigning as kings of today’s Big Data revolution?
In 2009, when Google was still fine tuning its PageRank algorithm based on the statistical innovation Markov Chain, Google’s Chief Economist Hal Varian declared statistician as the sexiest job of the decade. We’re about halfway through, and it seems that Varian missed the target.

Environmental science: Agree on biodiversity metrics to track from space
Nature News & Comment from July 22, 2015
Global biodiversity loss is intensifying. But it is hard to assess progress towards the Aichi Biodiversity Targets for 2011–20 set by the Convention on Biological Diversity (CBD). Target 5, for instance, aims to halve global deforestation rates by 2020; but reliable indicators for deforestation that can be monitored remotely have not been developed or agreed on. National biodiversity monitoring programmes differ widely, most data sets are inconsistent, and few data are shared openly.

To focus priorities, ecologists have proposed classes of 'essential biodiversity variables' — including species traits and populations, and ecosystem function and structure. But measuring these on the ground is laborious and limited.

US tailored-medicine project aims for ethnic balance
Nature News & Comment from July 21, 2015
The clock is ticking for experts charged with designing a US government programme to collect genetic, physiological and other health data from one million volunteers over the next two decades. The plan for the effort, part of the US$215-million Precision Medicine Initiative (PMI) announced in January, is due in the next few weeks — a daunting deadline, in part because the effort’s priorities include filling racial and socio-economic gaps left by other long-term studies.

Kurtis Heimerl to join UW CSE faculty
UW CSE News from July 27, 2015
We are thrilled to announce that Kurtis Heimerl will be joining the UW CSE faculty in early winter 2016. Kurtis’ research interests span information and communication technologies and development (ICTD), human-computer interaction, and networks and systems.

WPI to offer doctorate in data science
WBJournal from July 24, 2015
Worcester Polytechnic Institute will offer an interdisciplinary doctoral program in data science starting in the fall, a year after it launched a master's program in the growing field of study.

A statement from WPI said the new program will be the first in the United States. It draws on the expertise of WPI faculty in business, computer science and mathematics, and is designed to meet a "growing demand for highly trained scientists with trans-disciplinary technical and scientific expertise who can solve problems at the cutting edge of big-data analytics," the statement read.

Data Science University Programs
Silk from July 27, 2015
A Map of Data Science Degree Programs Around the The World

The Guy Who Taught AI to ‘Remember’ Is Launching a Startup
WIRED, Business from July 28, 2015
... Recurrent neural nets, or RNNs, can not only recognize complex moving images, but
automatically generate detailed captions for online photos and videos, improve online services that translate from one language to another, and more. They’re pushing into companies like Facebook and Baidu as well as Google, and in recent weeks, this burgeoning technology received another shot in the arm with the arrival of a new startup called Nnaisense.

According to the company’s website, Nnaisense was founded by Jürgen Schmidhuber, a key figure in the development of modern RNNs, and four researchers who work alongside him at the Swiss AI lab called IDSIA (Istituto Dalle Molle di Studi sull’Intelligenza Artificiale). The German-born Schmidhuber helped create a breed of recurrent neural net called LSTM, or Long Short Term Memory, and his work has influenced the latest AI research at the likes of Google, Microsoft, IBM, and others.

Mapping Systemic Risk: Critical Degree and Failures Distribution in Financial Networks

PLOS One from July 24, 2015
The financial crisis illustrated the need for a functional understanding of systemic risk in strongly interconnected financial structures. Dynamic processes on complex networks being intrinsically difficult to model analytically, most recent studies of this problem have relied on numerical simulations. Here we report analytical results in a network model of interbank lending based on directly relevant financial parameters, such as interest rates and leverage ratios. We obtain a closed-form formula for the “critical degree” (the number of creditors per bank below which an individual shock can propagate throughout the network), and relate failures distributions to network topologies, in particular scalefree ones. Our criterion for the onset of contagion turns out to be isomorphic to the condition for cooperation to evolve on graphs and social networks, as recently formulated in evolutionary game theory. This remarkable connection supports recent calls for a methodological rapprochement between finance and ecology.

As Tech Booms, Workers Turn to Coding for Career Change

After Paul Minton graduated from college, he worked as a waiter, but always felt he should do more.

So Mr. Minton, a 26-year-old math major, took a three-month course in computer programming and data analysis. As a waiter, he made $20,000 a year. His starting salary last year as a data scientist at a web start-up here was more than $100,000.

Data Science for Social Good 2015, Project blogs

UW eScience Institute, Data Science for Social Good 2015 from July 14, 2015
As part of the DSSG program, participants have been blogging about their experience. In team blogs devoted to the different projects, they have been sharing their thoughts coming into the program, and will continue to share their thoughts about the challenges and victories they encounter along the way. As we enter the second half of the program, a few of the interesting results will also start appearing on these blogs, documenting the progress of the projects.

Events

Accelerating Data-driven Discovery and Innovation: New York Scientific Data Summit

The first-annual New York Scientific Data Summit (NYSDS) is a no-fee meeting that aims to accelerate data-driven discovery and innovation by bringing together researchers,
developers and end-users from academia, industry, utilities and state and federal governments. The Summit is jointly organized by Brookhaven National Laboratory, Stony Brook University and New York University. Sponsors include the New York University Center for Data Science, and the Moore-Sloan Data Science Environment.

The 2015 theme is "Frontiers in Scientific Data." There is no registration fee to attend. Registration includes a sponsored welcome reception, conference dinner and coffee breaks. Note that registration will be limited to 150 participants.

Sunday-Wednesday, August 2-5, at NYU.

Data Science Workshop, 2015
This NSF sponsored 2.5 day workshop will bring together graduate students from diverse domain sciences and engineering with Data Scientists from industry and academia to discuss and collaborate on Big Data / Data Science challenges.

Wednesday-Friday, August 5-7, at University of Washington, Seattle

Dining on Data: The Future of Cooking, Eating & Running a Food Biz
This month, we're thrilled to hear from five companies that are using data to improve everything from cooking to restaurant concept and menu development. Join us for an inside look at these companies' business models, the innovative ways they use data and their lessons learned.

Tuesday, August 11, at 7 p.m., Redwood Studios, 55 9th Street, Brooklyn

PyGotham 2015 Tickets, New York
PyGotham is an eclectic Py-centric conference covering many topics. There's a diverse speaker list, and some things which will be quite different.

Saturday-Sunday, August 15-16, at AMA New York Executive Conference Center, 1601 Broadway, Times Square

NYC Media Lab 2015 Annual Summit
NYC Media Lab's Annual Summit is a snapshot of the best thinking, projects, and talent in digital media from universities in NYC and beyond.

This is an opportunity for media executives, technologists, and decision makers to see more than 100 university prototypes and demonstrations that explore interesting technologies and applications related to the future of media. Attendees will be also invited to roll up their sleeves during a series of interactive workshops led by NYC faculty.

Friday, September 25, at 9 a.m., NYU Skirball Center for the Performing Arts

Deadlines
NYC Media Lab & Bloomberg Data for Good Exchange (DGX) Prize & Impact Grant Program
Calling all students, faculty and university researchers! Are you currently working on data-driven research relating to education, government innovation, public health or climate change? Are you currently collaborating with a nonprofit / NGO on data analysis for a civic cause? Do you believe in the power of data to connect ideas for social impact?
Deadline for Submissions: Saturday, August 1

**Esri Human Health and Climate Change App Challenge**
The White House just released more than 100 health-related open data resources to keep people more informed. To reinforce our commitment and support President Obama's Climate Data Initiative, Esri announces the Esri Human Health and Climate Change App Challenge.

Submission Deadline: Friday, August 14

**UNICEF and Bloomberg Announce Data Science Researcher-in-Residence Program**
The researcher-in-residence, supported by Bloomberg, will help leverage knowledge transfer and capacity building on Data Science for humanitarian development at a global level while helping to design new data science methods or applying existing methods to help solve pressing problems that face large humanitarian organizations such as UNICEF.

Candidates interested in applying for this role can download and submit an application from http://www.bloomberg.com/company/d4gx/#unicef.

Deadline for applications is 21 August 2015.

**ACM SIGSPATIAL 2015, November 3-6 in Seattle**
The ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems 2015 (ACM SIGSPATIAL 2015) is the twenty-third event in a series of symposia and workshops that began in 1993 with the aim of bringing together researchers, developers, users, and practitioners in relation to novel systems based on geo-spatial data and knowledge, and fostering interdisciplinary discussions and research in all aspects of geographic information systems.

Workshops include UrbanGIS 2015 and Geo Privacy, and have paper submission deadlines beginning in mid-August.

**CDS News**

Here’s What Inspired Top Minds in Artificial Intelligence to Get Into the Field

*Bloomberg Business* from July 29, 2015

... We spent the past few months asking some of the field’s most renowned researchers and entrepreneurs what inspired them to pour their intellectual life into something that once seemed so unlikely and ominous.

Yann LeCun, the director of artificial intelligence research at Facebook, remembers watching 2001: A Space Odyssey when he was about 10 years old. He was enthralled by the hyper-intelligent computer within the spacecraft. LeCun is “not interested, particularly, in how humans function” but says his obsession with developing AI stems from a belief that it could lead scientists to develop a theory for how cognition works, whether biological or digital. “The analogy I always use is: Birds fly, and so do airplanes,” LeCun says. “They use very different details of implementation, but the underlying principles of flight are the same, and that’s based on aerodynamics. What’s the equivalent of aerodynamics for intelligence like this? That’s the big question.”

Movement Tracking Technology Sheds Light on Different Speech Disorders in Children
Facial motion capture – the same technology used to develop realistic computer graphics in video games and movies – has been used to identify differences between children with childhood apraxia of speech and those with other types of speech disorders, finds a new study by NYU’s Steinhardt School of Culture, Education, and Human Development.

“In our study, we see evidence of a movement deficit in children with apraxia of speech, but more importantly, aspects of their speech movements look different from children with other speech disorders,” said study author Maria Grigos, associate professor in the Department of Communicative Sciences and Disorders at NYU Steinhardt. The study, coauthored by Aviva Moss and Ying Lu of NYU Steinhardt, is published in the August issue of the Journal of Speech, Language, and Hearing Research.