NYU Data Science Community Newsletter features journalism, research papers, events, tools/software, and jobs for October 9, 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 022.

Data Science News

When a Genetic ID Card Is the Difference Between Life and Death
The Atlantic, Ed Yong from October 05, 2015
A simple genetic test can stop a severe drug reaction that causes people's skin to peel off in sheets. Why isn't it more commonly used?

SMAPP Twitter Toolkit
GitHub, SMAPPNYU from October 07, 2015
This is an user-friendly python package for interfacing with large collections of tweets. Developped at the SMaPP lab at New York University.

Announcing our 2015 Mozilla Fellows for Science!
Mozilla Science Lab from October 05, 2015
It's my pleasure today to announce the first class of Mozilla Fellows for Science. The folks chosen are representative of the change we want to see in the community, championing openness, collaboration, and mentorship in science. Over the next ten months, the Fellows will work on projects to help their local communities engage with open data, open source software and teach forward to their peers. They will also receive training and support from Mozilla to hone their skills around open source, participatory learning, and data sharing.

MS Ventures Accelerator Seattle Opens Applications for ML & Data Science Startups
TechNet Blogs, Machine Learning blog from October 07, 2015
The Microsoft Ventures Accelerator in Seattle has opened applications for its newest batch of startups, and our focus this third time around is on machine learning and data science startups.

New Report Puts Numbers on Data Scientist Trend
Wall Street Journal, The Numbers blog from October 07, 2015
Data scientist—a job that barely existed a decade ago—has become one of the hottest and best-paid professions in the U.S. Companies say they need people who have the skill sets—both business and technical—to analyze the rising tide of data produced by customers and operations.

RJMetrics, a software startup that itself is looking for data scientists to fill open positions, dove into LinkedIn data to gauge the field’s scope.

How scientists fool themselves—and how they can stop
Nature News & Comment from October 07, 2015
In 2013, five years after he co-authored a paper showing that Democratic candidates in the United States could get more votes by moving slightly to the right on economic policy, Andrew Gelman, a statistician at Columbia University in New York City, was chagrined to learn of an error in the data analysis. In trying to replicate the work, an undergraduate student named Yang Yang Hu had discovered that Gelman had got the sign wrong on one of the variables.

Gelman immediately published a three-sentence correction, declaring that everything in the paper’s crucial section should be considered wrong until proved otherwise.

Reflecting today on how it happened, Gelman traces his error back to the natural fallibility of the human brain: “The results seemed perfectly reasonable,” he says. “Lots of times with these kinds of coding errors you get results that are just ridiculous. So you know something’s got to be wrong and you go back and search until you find the problem. If nothing seems wrong, it’s easier to miss it.”

The smartest economist you’ve never heard of
The Washington Post from October 03, 2015
When David Lipton, a promising economist, was finishing his graduate work at Harvard in the early 1980s, he faced one of those potentially life-changing choices. He had one job offer from the International Monetary Fund in Washington, the multinational institution that for 70 years has served as a lender of last resort and dispenser of orthodox economic advice to countries that get into financial trouble. There was also an offer of a teaching job from the University of Virginia. Unsure of which path to take, he turned for advice to an intellectually restless and charismatic assistant professor, a Frenchman named Olivier Blanchard. ... As one colleague put it, Blanchard “changed the way the Fund looked at the world and the way the world looked at the Fund.” In the process, he helped the IMF pull the global economy back from the brink of another Great Depression.

Semantic Sensors
Pete Warden from October 03, 2015
The other day I was catching up with neighborhood news, and saw this article about “people counters” in San Francisco's tourist district. These are cameras watching the sidewalks and totaling up how many pedestrians are walking past. The results weren't earth-shattering, but I was fascinated because I'd never heard of the technology before. Digging in deeper, I discovered there's a whole industry of competing vendors offering similar devices.

Why am I so interested in these? Traditionally we've always thought about cameras as devices to capture pictures for humans to watch. People counters only use images as an intermediate stage in their data pipeline, their real output is just the coordinates of nearby pedestrians. Right now this is a very niche application, because the systems cost $2,100 each. What happens when something similar costs $2, or even 20 cents? And how about combining that price point with rapidly-improving computer vision, allowing far more information to be derived from images?

Those trends are why I think we're going to see a lot of “Semantic Sensors” emerging.

Machine Learning in the World of Sports - Rajiv Maheswaran
YouTube, O'Reilly from October 08, 2015
Second Spectrum CEO Rajiv Maheswaran talks at Strata+Hadoop on the beginnings of machine learning in the world of professional sports. [video autoplays, 9:37]

Services provided by Planned Parenthood, 2006-2013 : dataisbeautiful
reddit.com/r/dataisbeautiful from October 03, 2015
For those who don't know, this post is in direct reply to arguably the worst, most misleading, idiotic chart ever created (which was widely shared by other idiotic people). ... 1388 comments, as of 10/9

NYU-Backed Startup Agrilyst Wins TechCrunch Disrupt Cup
NYU Local from September 29, 2015
When the highly anticipated TechCrunch Disrupt competition drew to a close last week in San Francisco, a New York City-based, NYU-backed startup came out on top.

Agrilyst, an intelligence platform for indoor farms, was named the winner of TechCrunch Disrupt's insanely competitive Startup Battlefield.

STAT 94 Fall 2015, University of California-Berkeley
Ani Adhikari from October 06, 2015

This introductory course in data science is built on three interrelated perspectives: inferential thinking, computational thinking, and real-world relevance. Given data arising from some real-world phenomenon, how does one analyze that data so as to understand that phenomenon? How does one collect data to answer questions that one is interested in? Inferential thinking refers to an ability to connect data to underlying phenomena and to the ability to think critically about the conclusions that are drawn from data analysis. Computational thinking refers to the ability to conceive of the abstractions and processes that allow inferential procedures to be embodied in computer programs, and to ensure that such programs are scalable, robust and understandable. In addition to teaching basic skills in computer programming and statistical inference, the course will also involve the hands-on analysis of a variety of real-world datasets, including economic data, document collections, geographical data and social networks, and it will delve into social and legal issues surrounding data analysis, including issues of privacy and data ownership.

Events

Moore-Sloan Data Science Lunch Seminar Series

Introducing the Divergent Autoencoder Classifier for Machine Learning Applications
by Kenneth J. Kurtz ’72, Binghamton University (SUNY) Department of Psychology

Monday, October 12, 12 noon, at NYU CDS, 726 Broadway 7th Floor. Lunch provided.

NYU music informatics group will host a guest seminar by Mohamed Sordo from the University of Miami.

Title: Exploiting Knowledge Bases for Music Browsing and Discovery

This talk will briefly describe how KBs can be exploited for music browsing and discovery.

Tuesday, October 13, at 10 a.m., 6th floor conference room, 35 W 4th Street

Conference on Responsible Use of Open Data: Government and the Private Sector

The Conference on Responsible Use of Open Data: Government and the Private Sector, to be held at New York University on November 19-20, 2015 is co-organized by BCLT and NYU’s Information Law Institute and Department of Media, Culture and Communication. The event continues and expands the issues addressed by the April 2015 BCLT event Open Data: Addressing Privacy, Security, and Civil Rights Challenges Symposium. We are delighted that Dr. Amen Ra Mashariki, the City of New York Chief Analytics Officer and in charge of the Mayor’s Office of Data Analytics has agreed to open the conference in the afternoon of November 19th. Registration is required.

Thursday-Friday, November 19-20, at NYU School of Law
Deadlines

**Personal Data: Examined Lives**
This shift towards personal use leads to challenging new research questions. This special issue of *Human-Computer Interaction* focuses on emerging research about how people might appropriate and use personal data for personal purposes such as:
- Self-monitoring and self-understanding
- Identity work, self-representation, reminiscing, and legacy
- Behavior change that might promote physical and mental well-being
- Developing and maintaining interpersonal and community relationships
- Monitoring and managing their relationships with organizations and applications

Deadline for Proposals: Thursday, October 15

**IJCAI-2016 Call for Workshop Proposals**
The IJCAI-16 Workshop Program Co-Chairs invite proposals for workshops to be held July 9-11, 2016, immediately prior to the main conference. The aim of the workshop program is to provide a structured setting for the discussion of specialized technical topics; the format of proposed workshops should be designed to promote an active exchange of ideas between attendees. All members of the AI community are invited to submit workshop proposals for review.

Submission Deadline: Tuesday, November 3

**The Allen AI Science Challenge**
The Allen Institute for Artificial Intelligence (AI2) is working to improve humanity through fundamental advances in artificial intelligence. One critical but challenging problem in AI is to demonstrate the ability to consistently understand and correctly answer general questions about the world. ... Using a dataset of multiple choice question and answers from a standardized 8th grade science exam, AI2 is challenging you to create a model that gets to the head of the class.

Started: 4:13 pm, Wednesday 7 October 2015 UTC
Ends: 11:59 pm, Saturday 13 February 2016 UTC (129 total days)

**CDS News**

**Using data science to improve New York — Faculty Profiles: Shivendra Panwar**
CATT is a New York state sponsored program that works to improve existing technology and infrastructure through university and industry partnerships in the fields of economic development, technology transfer, workforce training, entrepreneurial support, and research & development. Data science is one of the three areas we are focusing on, in addition to wireless networking and cyber security. I got involved in the mid-1980s, and I've been the center's director since 1996.

New Study Raises Questions about Training by Repetition for Those With Autism

New York University from October 06, 2015

Training individuals with autism spectrum disorder (ASD) to acquire new information by repeating the information may harm their ability to apply that learned knowledge to other situations, a study by a team of neuroscientists shows.

This finding, published in the journal *Nature Neuroscience*, challenges the popular educational approaches designed for ASD individuals that focus on repetition and drills. It has been thought that because those with ASD sometimes acquire a new behavior or skill only in a specific context, and have difficulty transferring that learned skill or information to a new context, repetition can aid the learning process.

“Our conclusion is quite different: breaks in repetition allow the visual system some time to rest and allow autistic individuals to learn efficiently and to then generalize,” says David Heeger, a professor in NYU’s Center for Neural Science and one of the study’s co-authors.

Hacking the Cosmos: Event Hopes to Solve Complex Data Challenges

Space.com from October 07, 2015

Last week, astronomers, astrophysicists, data scientists and programmers came together at New York University to try to solve some of astronomy's toughest problems — in just five days.

The event, called Astro Hack Week, has only one rule: Everybody has to produce something. It might be a build of an astronomy data search algorithm, a series of programming tutorials or a bot that generates fake (and surprisingly plausible) tweets from one of the event creators. It might be planned from the get-go or something dreamed up based on a morning teaching session.

Click here to receive the NYU Data Science Community Newsletter 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@nyu.edu by 9 pm Eastern time on Thursday evenings for inclusion in Friday’s newsletter. We retain curatorial discretion.

NYU Data Science Community Newsletter Issue 022.

Medium, NYU Center for Data Science from October 07, 2015

... Could you talk about the CATT grant and your involvement?