NYU Data Science Community Newsletter features journalism, research papers, events, tools/software, and jobs for July 8, 2016

Data Science News

Accelerating research into dark energy

Royal Astronomical Society from July 06, 2016
A quick method for making accurate, virtual universes to help understand the effects of dark matter and dark energy has been developed by scientists at **University College London** and the **Aragon Centre for Cosmological Studies** in Spain. Making up 95% of our universe, these substances have profound effects on the birth and lives of galaxies and stars and yet almost nothing is known about their physical nature.

Also in astronomy:
- Modeling confounding by half-sibling regression (July 05, Proceedings of the National Academy of Sciences; Bernhard Schölkopf, David W. Hogg, Dun Wang, Daniel Foreman-Mackey, Dominik Janzing, Carl-Johann Simon-Gabriel and Jonas Peters)

Power to the People: How One Unknown Group of Researchers Holds the Key to Using AI to Solve Real Human Problems

Medium, Greg Borenstein from June 30, 2016
... What's needed for AI's wide adoption is an understanding of how to build interfaces that put the power of these systems in the hands of their human users. What's needed is a new hybrid design discipline, one whose practitioners understand AI systems well enough to know what affordances they offer for interaction and understand humans well enough to know how they might use, misuse, and abuse these affordances.
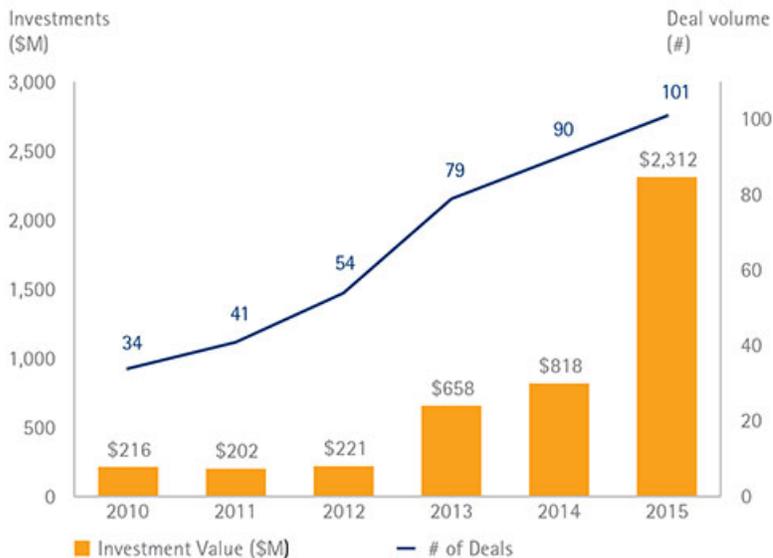
More on data & design:
- A framework for evaluating electronic health record vendor user-centered design and usability testing processes (July 03, Journal of the American Medical Informatics Association; Raj M Ratwani et al)
- Mike Kuniavsky on the mindshift needed to design for ecosystems (July 07, O'Reilly Radar, Mary Treseler and Mike Kuniavsky)
- Ambient Computing by Mike Barlow (July 08, O'Reilly Media, Mike Barlow)
- Data Visualization, Design and Information Munging (November 2015, Martin Krzywinski/ Genome Sciences Center)
- How Fast Food Chains Use Data to Test New Products and Drive Sales (July 07, Eater)

How New York Beat Silicon Valley in Fintech Funding in Q1

Datamation from July 06, 2016

New York Fintech Financing Activity

Investments ($M) / Deal volume (#)

Year	Investment Value ($M)	# of Deals
2010	$216	34
2011	$202	41
2012	$221	54
2013	$658	79
2014	$818	90
2015	$2,312	101

Source: Accenture, Partnership Fund analysis of CB Insights data

It helps that Wall Street calls New York home, but a concerted effort by the area's financial firms to partner with technology startups is cementing the city's reputation as a fintech hub.

SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods

EPJ Data Science from July 07, 2016

There is a strong need to conduct a thorough apple-to-apple comparison of sentiment analysis methods, as they are used in practice, across multiple datasets originated from different data sources. Such a comparison is key for understanding the potential limitations, advantages, and disadvantages of popular methods. This article aims at filling this gap by presenting a benchmark comparison of twenty-four popular sentiment analysis methods. [full text]

Building a Community of Social Scientists with Big Data Skills: The ICOS Big Data Summer Camp

University of Michigan, Michigan Institute for Data Science (MIDAS) from July 01, 2016

As the use of data science techniques continues to grow across disciplines, a group of **University of Michigan** researchers are working to build a community of social scientists with skills in Big Data through a week-long summer camp for faculty and graduate students.

Tesla and Google Take Different Roads to Self-Driving Car

The New York Times from July 04, 2016

In Silicon Valley, where companies big and small are at work on self-driving cars, there have been a variety of approaches, and even some false starts.

The most divergent paths may be the ones taken by **Tesla**, which is already selling

cars that have some rudimentary self-driving functions, and **Google**, which is still very much in experimental mode.

More on cars:
- The autonomous car as a driving partner (July 05, O'Reilly Radar, David Beyer and Daniela Rus)
- Self-Driving Laws (July 05, SSRN; Anthony J. Casey and Anthony Niblett)
- Without legislation, robotic cars can't be tested on Mass. roads (July 04, The Boston Globe)
- As Self-Driving Cars Hit the Road, Innovation Is Outpacing Insurance (July 03, The New York Times)
- BMW strikes autonomous car deal with Intel, Mobileye (July 05, ReadWrite)

How the Computer Beat the Go Player

Scientific American, Christof Koch from July 01, 2016
The victory in March of the computer program AlphaGo over one of the world's top handful of go players marks the highest accomplishment to date for the burgeoning field of machine learning and intelligence. The computer beat **Lee Se-dol** at go, a very old and traditional board game, at a highly publicized tournament in Seoul in a 4–1 rout. With this defeat, computers have bettered people in the last of the classical board games, this one known for its depth and simplicity. An era is over, and a new one has begun. The methods underlying AlphaGo, and its recent victory, have startling implications for the future of machine intelligence.

President Obama: Medicine's next step

The Boston Globe from July 07, 2016
... By bringing together doctors and data like never before, precision medicine aims to deliver the right treatments in the right dosage at the right time — every time. It helps target the causes of a condition rather than just the symptoms. This is one of the greatest opportunities we've ever seen for new medical breakthroughs, but it only works if we collect enough information first.

More on government precision medicine grants:
- Scripps Research gets record $120M to change medicine (July 06, The San Diego Union-Tribune)
- NIH awards $55 million to build million-person precision medicine study (July 06, National Institutes of Health)

Tweet of the Week

Twitter, Sonal Chokshi from July 01, 2016

Sonal Chokshi
@smc90

Follow

here are the most and least metal words (according to this NLP analysis on lyrics corpus bit.ly/29iXZeV)

Most Metal Words				Least Metal Words		
Rank	Word	Metalness		Rank	Word	Metalness
1	burn	3.81		1	particularly	-6.47
2	cries	3.63		2	indicated	-6.32
3	veins	3.59		3	secretary	-6.29
4	eternity	3.56		4	committee	-6.16
5	breathe	3.54		5	university	-6.09
6	beast	3.54		6	relatively	-6.08
7	gonna	3.53		7	noted	-5.85
8	demons	3.53		8	approximately	-5.75

Events

AI Camp

An open source, community-run and not-for-profit conference focused on open source artificial intelligence (AI), deep learning and machine learning technologies. AI Camp will involve a smaller daytime event followed by a large cross-community meetup during the evening.

New York, NY Tuesday, July 12, at United Nations HQ.

Who's speaking at Data Day Seattle 2016?

Data Day Seattle will be held at the newly-renovated Westin Seattle. This year we're taking the entire building! Enough room for 50+ talks, six tracks, workshops, office hours, booksignings and a post-conference happy hour / job fair.
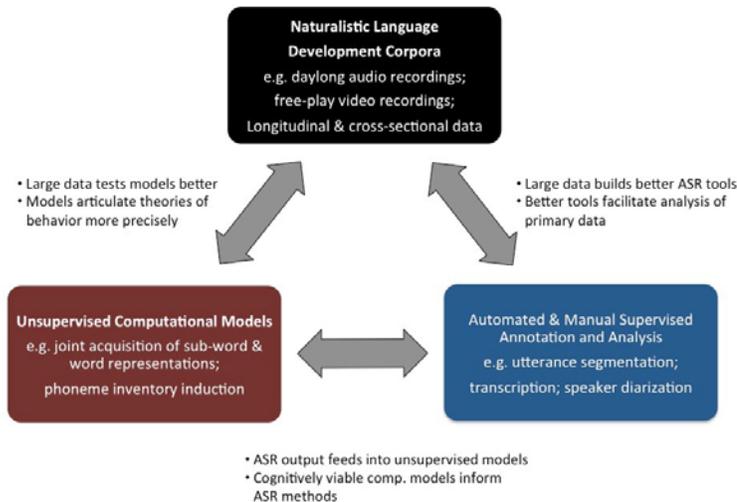
Seattle, WA Saturday, July 23.

Wrangle

Wrangle is a single-day, single-track industry event about the principles, practice, and application of Data Science, across multiple data-rich industries. It includes talks from practitioners at innovative companies about the hardest problems they've faced, and the solutions they found for them.

San Francisco, CA Thursday, July 28, at Broadway Studios (435 Broadway)

CogSci2016: Naturalistic Language Acquisition Data workshop

Philadelphia, PA Wednesday, August 10, part of CogSci 2016, the 38th Annual Meeting of the Cognitive Science Society.

HackOn(Data)

HackOn(Data) is a two-day event that will bring together the Toronto data community to take a closer look at data that touches our daily lives.

Toronto, Canada Saturday-Sunday, September 10-11.

Embedded Vision Alliance hosting training event for deep learning for vision

The **Embedded Vision Alliance** and **Berkeley Design Technology, Inc.** is hosting a one-day tutorial event on convolutional neural networks (CNN) for vision and the Caffe framework for deep learning, presented by the primary Caffe developers from the **Berkeley Vision and Learning Center**.

Cambridge, MA Thursday, September 22.

Deadlines

OpenCon2016 Application Form

deadline: Conference

Washington, DC OpenCon is the conference and community for students and early career professionals interested in advancing Open Access, Open Education and Open Data. OpenCon 2016 will be held on November 12-14 in Washington, DC. Attendance at OpenCon 2016 is by application only.

Deadline for applications is Monday, July 11.

Mozilla Fellowships for Science

deadline: Career Opportunity

We're looking for researchers with a passion for open source and data sharing, already working to shift research practice to be more collaborative, iterative and open. Fellows will spend 10 months starting September 2016 as community catalysts at their institutions, mentoring the next generation of open data practitioners and researchers and building lasting change in the global open science community.

Deadline for applications is Saturday, July 16.

DAPS2016 – Data mining for the Analysis of Performance and Success

deadline: Conference

The First International Workshop on Data mining for the Analysis of Performance and Success (DAPS 2016) will be held in conjunction with The IEEE International Conference on Data Mining series (ICDM) .

Barcelona, Spain Monday-Friday, December 12-16.

Deadline to submit paper proposal is Monday, August 1.

GE Intelligent World Hackathon: Insights from Street Lights

deadline: Contest/Award

In the Intelligent World Hackathon you can be one of the first developers to access **GE** smart LED streetlight network data and build urban apps on Predix, GE's new IIoT data analytics platform.

The hackathon runs until Tuesday, August 2.

CDS News

ODSC East 2016 | Stefan Karpinski - "Solving the Two Language Problem"

YouTube, Open Data Science from May 26, 2016
Stefan Karpinski is one of the co-creators of the Julia programming language and a co-founder of Julia Computing, Inc., which provides support, consulting and training for commercial usage of Julia. He has previously worked as a software engineer and data scientist at Akamai, Citrix, and Etsy. He is currently a Research Engineer at NYU as part of the Moore-Sloan Data Science Initiative. [video autoplays, 11:54]

What Does a Data Scientist Do?

NYU Center for Data Science from July 07, 2016
What is a data scientist? The question is seemingly simple, and yet deceptively complex. A data scientist obviously works with data, but in what capacity? Is a data

scientist more concerned with the collection of data, or data analysis? And what is the end goal for a data scientist? With data available on everything from digital advertisements to online purchases to personal health records, the breadth of fields requiring data scientists makes the definition even more elusive.

Last Tuesday, an organization called the **New York Data Science Study group** coordinated a data science career panel talk, titled, "What Does a Data Scientist Do?"

Tools & Resources

A beginner's guide to setting up a development environment on Mac OS X

GitHub - nicolashery from July 01, 2012
This document describes how I set up my developer environment on a new MacBook or iMac. We will set up Node (JavaScript), Python, and Ruby environments, mainly for JavaScript and Python development. Even if you don't program in all three, it is good to have them as many command-line tools use one of them. As you read and follow these steps, feel free to send me any feedback or comments you may have.

Yelp announces expanding "Yelp Knowledge" social analytics program

Marketing Land from June 29, 2016
Yelp is expanding the group of third party companies that can directly access its complete reviews data and announcing a new program called "Yelp Knowledge." ... The concept is that through the **Yelp** data, multi-location brands and franchises won't simply see reviews (e.g., 3.5 stars) but gain a more holistic sense of how their locations are performing and how customer sentiment is trending.

Why And How I Switched From Python To Erlang

Farsheed Ashouri from July 06, 2016
In this article, I am explaining my trip from Python to Erlang. If you are not a Python developer (+ probably with a deep understanding of Python based web services), or you don't need/want to scale thing to a very huge scale, you won't be able to find this article much useful. If you don't want to develop an infrastructure for your business or If you develop simple blogs, small asset management systems or Hello World-ish websites, This article won't help you at all and If you are about to choose a language to start, please do not decide based on my words. I am going to tell you what problems I encountered using Python and how Erlang is able to solve those specific problems for me.

A curated list of awesome pipeline toolkits inspired by Awesome Sysadmin

GitHub - pditommaso from June 22, 2016
starting with ... Pipeline frameworks & libraries
- ActionChain - A workflow system for simple linear success/failure workflows.
- Airflow - Python-based workflow system created by AirBnb.
- Anduril - Component-based workflow framework for scientific data analysis.

Careers

NLP Post-Doc Job Posting - Center for Data Science and Public Policy

University of Chicago, Center for Data Science and Public Policy from July 05, 2016
Chicago, IL This Post-Doc Fellow will lead the NLP work on our Workforce Science project, a part of the **White House's** Workforce Data Initiative.

Computational Social Sciences Consultant

Duke University from July 05, 2016
Durham, NC The **Duke Population Research Institute** (DuPRI) is seeking a highly motivated Computational Social Sciences Consultant to work closely with DuPRI faculty and researchers in the social and biological sciences.

Senior Data Scientist Job at Akamai Technologies

LinkedIn, Akamai Technologies from July 06, 2016
Cambridge, MA As a Senior Data Scientist, **Akamai** is a great place to apply your unique mix of communication, statistical, and coding skills to interesting and unique data sets. You will be responsible for selecting algorithms best suited to modeling data sets in a wide variety of domains and developing and testing those models in such a way that they can be implemented in a production environment.

Sr. Strategic Analyst (Data Scientist) - Careers at Memorial Sloan Kettering Cancer Center

Memorial Sloan Kettering Cancer Center from July 08, 2016
New York, NY We are seeking a Senior Strategic Analyst who will serve as the lead on crucial projects. In partnership with the Director of Patient Access, the Senior Strategic Analyst will empower data-driven decisions and process improvement to improve patients' outcomes and experience.