

**NYU Data Science Community Newsletter** features journalism, research papers, events, tools/software, and jobs for September 9, 2016

Please let us ([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 Moore-Sloan Data Science Environment and to NYU's Center for Data Science.

### It's been a while, let's catch up

The newsletter kept on chugging over the summer. Thank you for pushing us past the 2,000th recipient mark today!

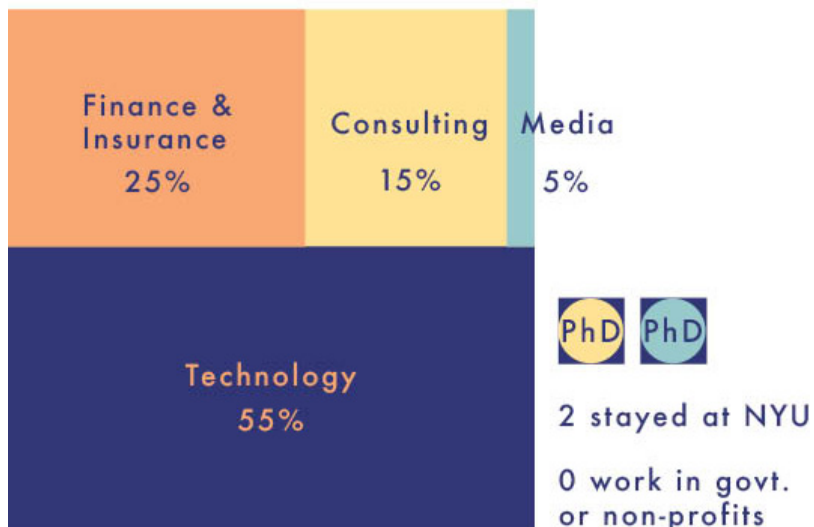
### NYU Data Science Community Newsletter Growth



### What happened to the masters students who graduated last May?

Our masters students have (mostly) started new jobs. Half landed jobs in tech for companies like **Apple**, **Facebook**, **Google**, **Pinterest**, **Twitter**, **Verisk**, and **WeChat**. A quarter took jobs in finance and insurance for companies like **Morgan Stanley**, **BlackRock**, and **AIG**. A few took consulting jobs at **Deloitte**.

## Where are NYU's data science masters students now?



### Notable absences?

Despite all the excitement about data science for social impact, none are working for non-profits or government agencies. This justifies continued foundation and [corporate support](#) for programs like **Eric Schmidt** and **Rayid Ghani's** [Data Science for Social Good](#) and non-profits like [DataKind](#).

Part of the problem is that salaries are too low at non-profits and in government to repay student loan debt. Another part of the problem is that the datasets are richer, deeper, and more exciting in private companies, an issue that also contributes to [academics departing for tech](#).

### Salaries?

O'Reilly published their latest [data science salary survey](#) this past week.

### Data Science News

---

#### [Precision medicine: Power to the patients?](#)

*Nature, Nature Research, Katherine Bourzac* from September 07, 2016

"For precision medicine to live up to its potential, millions of people must share their genomic data, their health records, and their experiences. To researchers, all of it is precious. The richer the databases, the better patient care will become." But who holds power in these systems? The patients may benefit from machine learning medical research, but will they be able to speak up about which questions to ask and how their data are managed?

#### More precision medicine and computational genomics:

- [Introducing Seeq - We want to make genomics inexpensive, interactive, and fun.](#) (September 06, Medium, The Seeq Blog, **Joe Pickrell**)
- [Smart medicine is coming of age, but will doctors bite?](#) (August 31, New Scientist, Technology News, **Aviva Rutkin**)

- [Reproducible Science: Cancer Researchers Embrace Containers in the Cloud](#) (September 06, Google Research Blog, **Nicole Deflaux** and **Jonathan Bingham**)

### [Audio data - Noise pollution research takes off](#)

*National Science Foundation* from September 06, 2016

NSF awarded \$4.6 m over five years to the Sounds of New York City (SONYC) project based at **NYU** and **The Ohio State University**.

#### **More audio data news and tools:**

- [How Can Data Collection Be Used to Map City Sounds?](#) (September 07, NYU Center for Data Science, research by **Tae Hong Park**, **Claudio Silva** and **Juan Bello**)
- [Urban Sound Classification](#) (September 03, a howto by **Aaqib Saeed**)
- [WaveNet: A Generative Model for Raw Audio](#) (September 08, **Google DeepMind**, **Aaron van den Oord**, **Heiga Zen**, **Sander Dieleman**)

### [The rent is too damn high!: Data science and real estate](#)

*Geoff Boeing* from August 24, 2016

**Geoff Boeing** at **UC-Berkeley** used **Craigslist** data from 11 million rental listings all over the US and [demonstrates that](#) in hot markets like New York and San Francisco significant dispersion in rental prices with a long tail at the top end. In cities like Detroit, there's a stronger peak around what appears to be a market clearing price.

#### **Others using real estate data:**

- [Boston officials use Big Data to find deceptive rental ads, unsafe units](#) (September 02, The Boston Globe, **Matt Rocheleau**)
- [Building a Team from the Inside Out: Alok Gupta on the Evolution of Data Science at Airbnb](#) (September 06, Kaggle, No Free Hunch blog, Alok Gupta)
- [Technology is finally changing the apartment rental experience](#) (September 04, TechCrunch, **Omri Barzilay**)

### [How Kalman shaped the world \(obituary\)](#)

*MIT Technology Review* from September 05, 2016

"**Rudolf Kalman**, a Budapest-born engineer and mathematician died on July 2 in Gainesville, Florida, at age 86. His fundamental contribution, an algorithm called the Kalman filter, made possible many essential technological achievements of the last 50 years. ... Someone once described the entire GPS system—an Earth-girdling constellation of satellites, ground stations, and computers as 'one enormous Kalman filter.'"

### [\[1609.01704\] Hierarchical Multiscale Recurrent Neural Networks](#)

*arXiv, Computer Science > Learning; Junyoung Chung, Sungjin Ahn, Yoshua Bengio* from September 06, 2016

"Learning both hierarchical and temporal representation has been among the long-standing challenges of recurrent neural networks. Multiscale recurrent neural networks have been considered as a promising approach to resolve this issue, yet there has been a lack of empirical evidence showing that this type of models can actually capture the temporal dependencies by discovering the latent hierarchical structure of the

sequence. In this paper, we propose a novel multiscale approach, called the hierarchical multiscale recurrent neural networks, which can capture the latent hierarchical structure in the sequence by encoding the temporal dependencies with different timescales using a novel update mechanism."

### [Tweet of the Week](#)

Twitter, Bored Yann LeCun from September 08, 2016



**Bored Yann LeCun**  
@boredyannlecun

DL researchers @ university: "let's be careful not to overstate connections..." @ startup: "harnessing the brain to disrupt mind change"

1:33pm · 8 Sep 2016 · Twitter Web Client

13 RETWEETS 22 LIKES



### Events

---

#### [2016 RecSys – ACM Recommender Systems](#)

**Boston, MA** The ACM Recommender Systems conference (RecSys). 15-19 September 2016.

#### [Symposium on Big Data and Human Development](#)

**Oxford, England** Thursday-Friday, 15-16 September 2016.

#### [Text as Data Speaker Series](#)

**New York, NY David Bamman (UC-Berkeley)** — Beyond Bags of Words: Linguistic Structure in the Analysis of Text as Data — Thursday, September 22, 4-5:30 p.m. in room 217, 19 West 4th St.

#### [#NYCML16 Preview: Data Science](#)

**New York, NY NYC Media Lab's Summit, #NYCML16**, taking place on September 22nd, 2016 at **Columbia University's** Alfred Lerner Hall, will host several discussions, workshops and demos. [\$\$\$]

[Data for Good Exchange 2016](#)

**New York, NY** Theme is “better governance”. Sunday, 25 September 2016 [free]

[IBM DataFirst Launch Event](#)

**New York, NY** Hudson Mercantile Building, Tuesday, 27 September 2016. [free]

[Ethics of Artificial Intelligence](#)

**New York, NY** Friday-Saturday, 14-15 October 2016, at NYU. [free]

[2016 Workshop on Visualization for the Digital Humanities](#)

**Baltimore, MD** Part of IEEE VIS 2016. Monday, October 24

**Deadlines**

---

[Call for Papers - ChASM 2016](#)

*deadline: Conference*

**Bellevue, WA** "The 4th edition of the Workshop on Computational Approaches to Social Modeling (ChASM) will be colocated with SocInfo'16, taking place in Bellevue, WA; 14-17 November 2016."

Deadline for submissions: Monday, 12 September 2016.

[Give feedback on NYC's proposed geospatial #opendata standards](#)

*deadline: Survey*

Public comments will close on Thursday, 15 September 2016.

[Help a colleague: 5 minute online study of visual attributes in networks](#)

*deadline: Survey*

"The goal of this study is to understand which visual representations are most prominent in network visualizations."

Deadline Thursday, 15 September 2016.

[Letters of Interest - Laura and John Arnold Foundation](#)

*deadline: RFP*

The Foundation seeks Letters of Interest for grants in Open Science and Science and Technology Research and Development.

Deadline for submissions is Friday, 16 September 2016.

### [SoCal ML Symposium](#)

*deadline: Conference*

**Pasadena, CA** "The Southern California Machine Learning Symposium brings together students and faculty to promote machine learning in the southern California region. The workshop serves as a forum for researchers from a variety of fields working on machine learning to share and discuss their latest findings." -- Friday, November 18, at **Caltech**.

Deadline for submissions is Tuesday, October 4.

### [Increasing Openness and Transparency in Research](#)

*deadline: Contest/Award*

**RWJF** wants to use these grants to help increase our understanding of reproducible research work flows, and the understanding of the research community, about the potential for these approaches to accelerate the discovery process and expand access to findings.

Deadline for letters of intent is Wednesday, 5 October 2016.

### [R/GA Marketing Tech Venture Studio](#)

*deadline: Contest/Award*

The **R/GA Marketing Tech Venture Studio** with **Interpublic** is designed for emerging companies revolutionizing the future of marketing and digital products by leveraging artificial intelligence and machine learning.

Deadline to apply is Monday, 24 October 2016.

### [Call for Papers - NIPS 2016 Workshop on Machine Learning for Health](#)

*deadline: Conference*

**Barcelona, Spain** In this workshop we will bring together clinicians, health data experts, and machine learning researchers working on healthcare solutions. -- Friday, December 9, part of NIPS 2016.

Deadline for submissions is Friday, October 28.

Other [NIPS 2016](#) workshops with paper submission deadlines (in parentheses):

- [Approximate Inference](#) (Tuesday, November 1)
- [Interpretable ML for Complex Systems](#) (Thursday, October 20)
- [Neural Abstract Machines & Program Induction](#) (Friday, October 14)

### [The Digital Mammography DREAM Challenge](#)

*deadline: Contest/Award*

The Digital Mammography DREAM Challenge will attempt to improve the predictive accuracy of digital mammography for the early detection of breast cancer.

Deadline for entries is Monday, March 13, 2017.

## NYU Center for Data Science News

---

### Summer's notable departures and arrivals at NYU Center for Data Science

- **Foster Provost** stepped down from the Moore-Sloan steering committee and handed over the CDS Interim Directorship to **Claudio Silva**.
- **Roy Lowrance**, CDS Managing Director, moved into industry.
- **Dan Cervone**, Moore-Sloan Fellow, took a job with the LA Dodgers starting in October.
- **Pablo Barberá**, a Moore-Sloan Fellow, started his tenure-track position at USC in July.
- **Andrea Rooy-Jones** joined Moore-Sloan in a research and outreach capacity.
- **Michael Gill** joined us as a Moore-Sloan Fellow.
- Our top-notch administrator **David Clark** started an MBA program this fall; we welcome **Kathryn Angeles** in his position.

### [Master's in Data Science Open House](#)

*NYU Center for Data Science* from September 27, 2016

**New York, NY** If you would like to learn more about the MSDS program and are in the NYC area, please plan to attend our Open House listed below. The Open House provides an overview of the program and gives you a chance to interact with faculty, students, and staff of CDS. Next Open House: Tuesday, September 27, at **NYU Center for Data Science**, (60 5th Avenue, 7th Floor).

### [Why Your Field Needs a Hack Week: Bringing Data Science Into Astronomy](#)

*YouTube, Berkeley Institute for Data Science, Daniela Huppenkothen* from September 02, 2016

[[Astro Hack Week](#)] is one approach to improving data literacy in astronomy while at the same time providing a collaborative venue for researchers to share ideas ... "I will give an overview of the ideas that underlie Astro Hack Week, highlight some results from this week's workshop, and try to convince you that your field needs a hack week too." [video autoplays, 48:18]

Also in hackathons:

- [Data Hackathons Workshop](#) in Denver, part of [International Data Week](#) (September 15)
- [Seattle Sports Tech Hackathon](#) (September 10-11)

## Tools & Resources

---

### [Manage your APIs with Google Cloud Endpoints](#)

*Google Cloud Platform Blog, Dan Ciruli* from September 01, 2016

"We're announcing the open beta release of the newest set of features and open source components in Google Cloud Endpoints, a distributed API management suite that lets you deploy, protect, monitor and manage APIs written in any language and running on Google Cloud Platform (GCP)."

### [The 7 Best Data Science and Machine Learning Podcasts](#)

*iamwire, Matt Fogel* from August 29, 2016

"I've listened to a bunch of machine learning and data science podcasts in the last few months, so I thought I'd share my favorites: 1. *The Data Skeptic*."

### [OSF | Preprints](#)

*Center for Open Science* from September 06, 2016

The open preprint repository network, a singular interface for searching the major preprint archives.

### [Tutorials and resources for machine learning and data analysis enthusiasts](#)

*Laura Harding Graesser, Learning Machine Learning* from September 02, 2016

"My previous post walked through the code for implementing a general feedforward neural network program in Python. This post shows the program in action, applied to two classification problems."

### [Learn Node.js - A Free Course for Node Beginners](#)

*HyperDev* from September 05, 2016

"Through 13 exercises, follow *The Node Beginner Book*, with live-running code examples on HyperDev. By the end, you'll know how to create a complete web app that allows users to view web pages and upload files."

### [Announcing Quill 1.0](#)

*Quill, Jason Chen* from September 06, 2016

Quill is designed as an easy to use editor, to support content creation across the web. It is built on top of consistent and predictable constructs, exposed through a powerful API. With coverage across both ends of the complexity spectrum, Quill aims to be the defacto rich text editor for the web."

### [Awesome Network Analysis](#)

*GitHub - briatte* from September 06, 2016

An awesome list of resources to construct, analyze and visualize network data.



## Careers

---

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

#### **[Assistant Professor, Department of Statistics & Data Sciences](#)**

University of Texas; Austin, TX

#### **[Assistant Professor - Data Science, Human Augmentation, Robotics, or related fields](#)**

University of California-San Diego; San Diego, CA

#### **[Assistant Professor - Data Ethics](#)**

Information School, University of Washington; Seattle, WA

#### **[Assistant Professor - Biostatistics](#)**

School of Public Health at the University of California, Berkeley; Berkeley, CA

#### **[Assistant Professor of Cognitive Psychology](#)**

Stanford University; Stanford, CA

#### **[Assistant Professor - Human-Computer Interaction \(HCI\) and social good](#)**

Information School of the University of Washington; Seattle, WA

#### **[Open Rank Professor - Data Science](#)**

New York University - Shanghai; Shanghai, China

### **Full-time, non-tenured academic positions**

#### **[Scientific Information Manager - Santa Barbara Coastal Long-term Ecological Research \(SBC LTER\) project](#)**

UCSB Marine Sciences Institute; Santa Barbara, CA

#### **[Astronomer - Space Telescope Science Institute](#)**

Johns Hopkins University; Baltimore, MD

#### **[Research Analyst \(accountability\) – apply by Sep 22](#)**

Data & Society; New York, NY

### **Postdocs**

#### **[Postdoc - State of Alaska's Salmon & People Synthesis](#)**

University of California, Santa Barbara's National Center for Ecological Analysis and Synthesis; Santa Barbara, CA

#### **[Postdoc - rOpenSci](#)**

rOpenSci Project at University of California, Berkeley; Berkeley

**Full-time positions outside academia**

**[Allen Institute \(AI2\) is hiring - many roles!](#)**

Allen Institute for Artificial Intelligence; Seattle WA

**[Senior Data Scientist](#)**

Hired; San Francisco, CA

**Internships and other temporary positions**

**[Google Brain Resident, 2017 Start \(Fixed Term Employee\)](#)**

Google; Mountain View, CA

**OPT OUT:** If you do not want to receive this newsletter, please email [brad.stenger@nyu.edu](mailto:brad.stenger@nyu.edu) with the word 'unsubscribe' in the subject line.

**OPT IN:** Feel free to forward the Data Science newsletter to colleagues. They can sign up for the newsletter using [this web form](#).