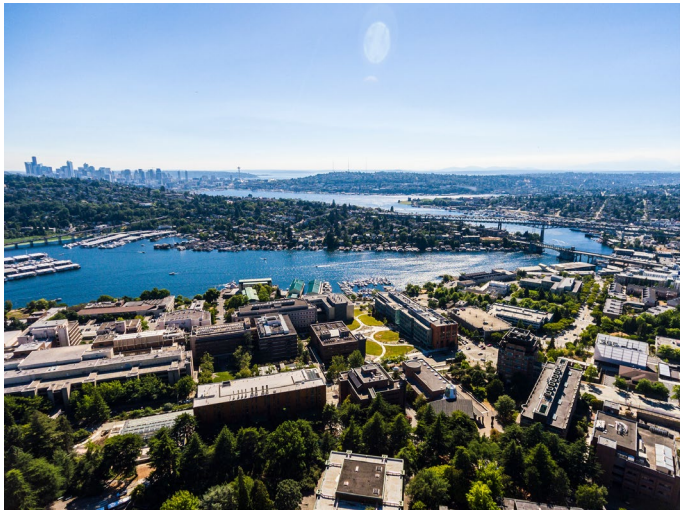


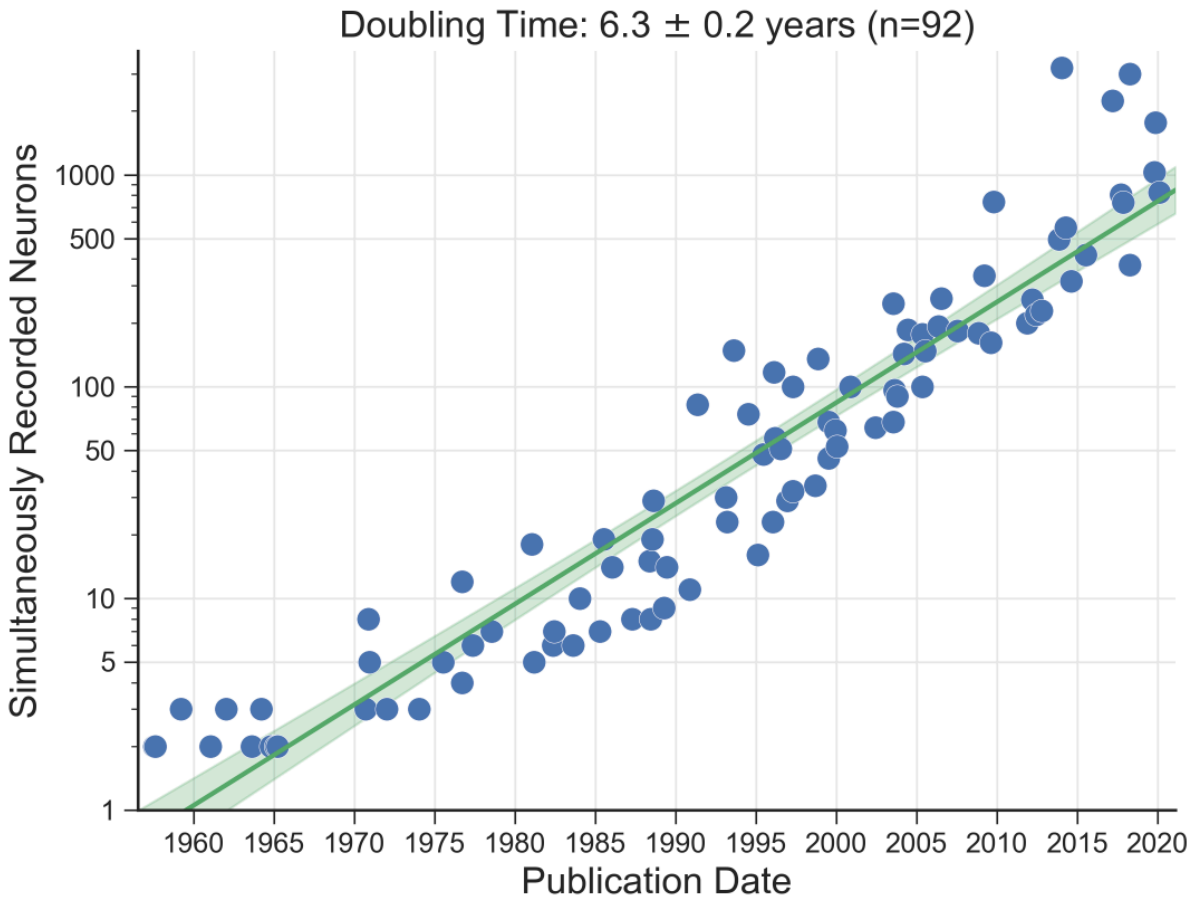
Data science and neuroinformatics

Ariel Rokem
Department of Psychology & eScience Institute
The University of Washington
https://neuroinformatics.uw.edu/

UNIVERSITY of WASHINGTON

Outline

- Why do we need data science?
- What is data science?
- How did we get here? A historical perspective
- Data science in practice
- Intellectual infrastructure
- This session



Stevenson and Kording (2011); Updated July 2020

N=1,200; Completed

Adolescent Brain Cognitive Development Study

N=10,000; Ongoing

CHILD MIND® INSTITUTE
healthy brain network

N=10,000; Ongoing

biobank uk
Imaging study

N=100,000; Ongoing

Data science?

- All of the things that need to be in place for us to be able to analyze and understand all that data
  - Computational infrastructure
  - Statistical analysis techniques
  - Software
  - …

International Statistical Review (2001), **69**, 1, 21–26, Printed in Mexico

Data Science: an Action Plan for Expanding the Technical Areas of the Field of Statistics

William S. Cleveland

Statistics Research, Bell Laboratories, 600 Mountain Avenue, Murray Hill NJ07974, USA
E-mail: wsc@research.bell-labs.com

- **(25%) Multidisciplinary Investigations**: data analysis collaborations in a collection of subject matter areas.
- **(20%) Models and Methods for Data**: statistical models; methods of model building; methods of estimation and distribution based on probabilistic inference.
- **(15%) Computing with Data**: hardware systems; software systems; computational algorithms.
- **(15%) Pedagogy**: curriculum planning and approaches to teaching for elementary school, secondary school, college, graduate school, continuing education, and corporate training.
- **(5%) Tool Evaluation**: surveys of tools in use in practice, surveys of perceived needs for new tools, and studies of the processes for developing new tools.
- **(20%) Theory**: foundations of data science; general approaches to models and methods, computing with data, teaching, and tool evaluation; mathematical investigations of models and methods, computing with data, teaching, and evaluation.

20 years of data science!

JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS
2017, VOL. 26, NO. 4, 745–766
https://doi.org/10.1080/10618600.2017.1384734

50 Years of Data Science

David Donoho

THE FUTURE OF DATA ANALYSIS[1]

BY JOHN W. TUKEY

Princeton University and Bell Telephone Laboratories

1962

Meanwhile, in Silicon Valley (ca. 2010)

"I was at LinkedIn building the data team, and Jeff Hammerbacher [co-founder of Cloudera] was bustling at Facebook's data team, and we would collaborate and compare notes sometimes. One of the things we realized was that we didn't know what to call ourselves, … Do you call yourself an analyst? It feels too Wall Street. A research scientist or statistician? Feels too academic, … But because I was working at LinkedIn, I just tested all the job titles we could think of to see which one would get the most interest from job applicants. Turns out that everybody wanted to be a data scientist, so we're like, OK, that is what we will call ourselves."

DJ Patil (https://observer.com/2019/11/data-scientist-inventor-dj-patil-interview-linkedin-job-market-trend/)

So, what is data science?

> Not really a scientific discipline?

> Not really a profession?

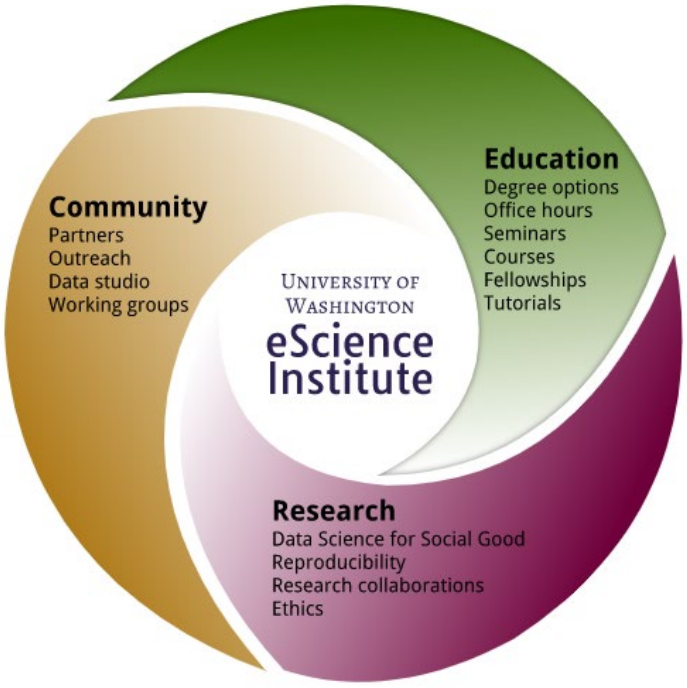
=> An engagement with the practice of data analysis

> Crosses disciplinary boundaries

> Crosses sectors

amazon

Data science in practice



Data science ≠ computer science

Data science ≠ statistics

Brain data science ⊆ Neuroinformatics

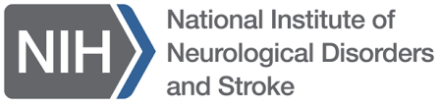
Intellectual infrastructure

- People
- Ideas
- Communities

- Statistics and machine learning
- Software engineering
- Open-source software
- Communities of practice

- Eva Dyer (GA Tech)
- Mackenzie Mathis (EPFL)
- Franco Pestilli (UT Austin)

Thanks!



https://neuroinformatics.uw.edu/