Data scientist: qualifications and skills

Jeff Leek

@jtleek

www.jtleek.com

What does a data scientist do?

- Design experiments
- Pull and clean data
- Analyze data
- Communicate results

https://vwo.com/ab-testing/

What skills do they need?

- Statistics (inference)
- Machine learning (prediction)
- Data analysis
- Data communication

https://www.coursera.org/specializations/jhudatascience

About This Specialization

Courses

Pricing

Creators

FAQs

Data Science Specialization

From $29

Enroll
Starts September 7, 2015

COURSE 1

The Data Scientist's Toolbox

Starts September 7, 2015

Commitment	1-4 hours/week
Duration	4 weeks

About the Course

In this course you will get an introduction to the main tools and ideas in the data scientist's toolbox. The course gives an overview of the data, questions, and tools that data analysts and data scientists work with. There are two components to this course. The first is a conceptual introduction to the ideas behind turning data into actionable knowledge. The second is a practical introduction to the tools that will be used in the program like version control, markdown, git, GitHub, R, and RStudio.

COURSE 2

- Background of data scientists
  - Statistics + application + engineering
  - Quantitative + data science transition
  - Software engineering + statistics

http://www.jhsph.edu/departments/biostatistics/

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

Protecting Health, Saving Lives—*Millions at a Time*

Biostatistics

- ABOUT US
- DIRECTORY
- PROSPECTIVE STUDENTS
- ACADEMICS & STUDENT LIFE
- RESEARCH
- BIOSTATISTICS CONSULTING CENTER
- ALUMNI
- RESOURCES
- CONTACT US
- MAKE A GIFT

Home > Departments > Biostatistics

Biostatistics

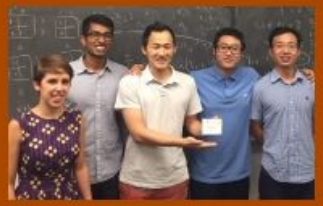
ANNOUNCEMENTS

The Department of Biostatistics seeks outstanding, highly-qualified candidates for tenure track faculty positions.
LEARN MORE

PROSPECTIVE STUDENTS

We appreciate your interest

OUR PEOPLE



Bmore Dream Team

SEMINARS

Take a look at our weekly

Solutions versus software

29
MAY

What statistics should do about big data: problem forward not solution backward

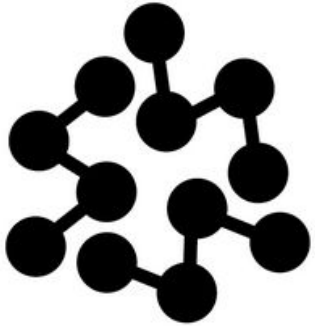
POSTED BY JEFF LEEK / UNCATEGORIZED

126 43 6 36

There has been a lot of discussion among statisticians about big data and what statistics should do to get involved. Recently Steve M. and Larry W. took up the same issue on their blog. I have been thinking about this for a while, since I work in genomics, which almost always comes with "big data". It is also one area of big data where statistics and statisticians have played a huge role.

A question that naturally arises is, "why have statisticians been so successful in genomics?" I think a major reason is the phrase I borrowed from Brian C. (who may have borrowed it from Ron B.)

> *problem first, not solution backward*



Search..

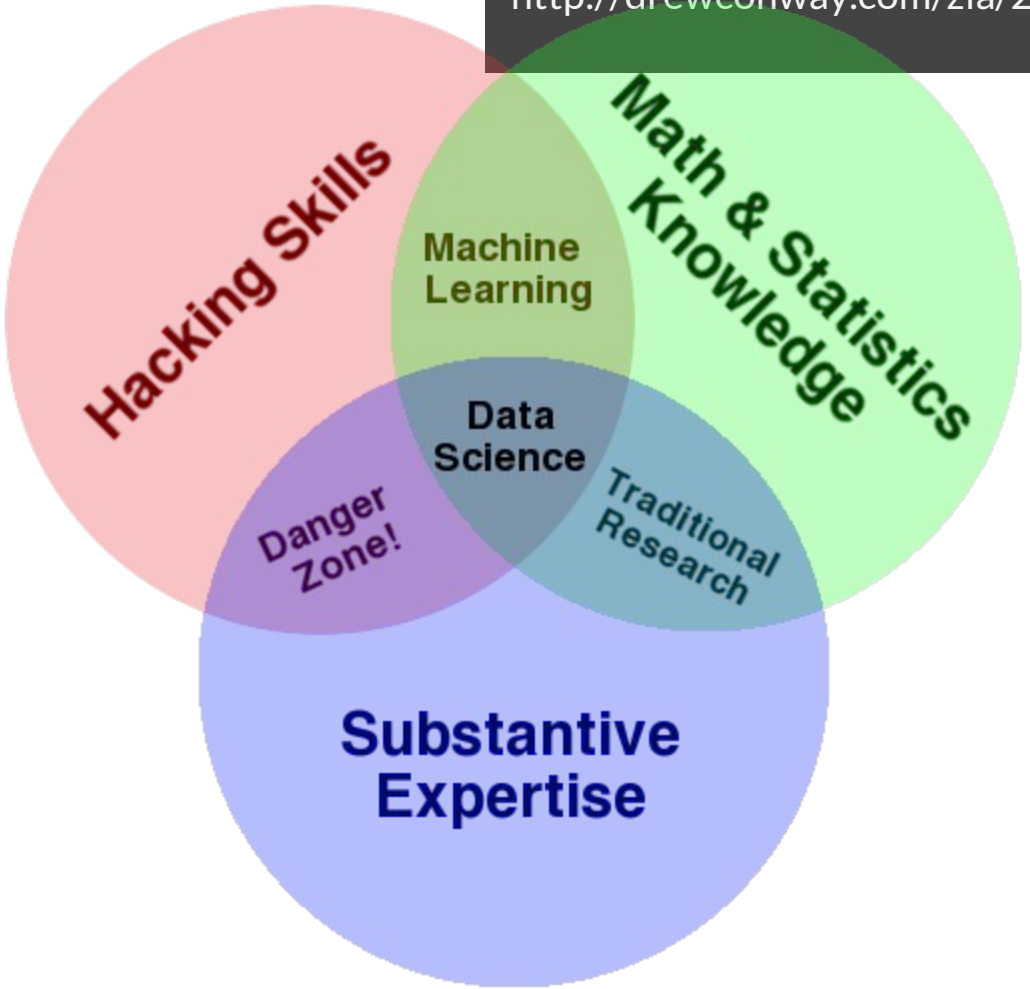
RECENT POSTS

Interview with COPSS award Winner John Storey

The Next National Library of Medicine Director Can Help Define the Future of Data Science

Key characteristics

- Willing to find answers on their own
- Unintimidated by new data
- Willing to say I don’t know
- Friendly but relentless

http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram

friendly! +