The Data Science Track

Jeffrey Leek

May 17, 2016

Why do data science?

"It is not the critic who counts: not the man who points out how the strong man stumbles or where the doer of deeds could have done better. The credit belongs to the man who is actually in the arena, whose face is marred by dust and sweat and blood, who strives valiantly, who errs and comes up short again and again, because there is no effort without error or shortcoming, but who knows the great enthusiasms, the great devotions, who spends himself for a worthy cause; who, at the best, knows, in the end, the triumph of high achievement, and who, at the worst, if he fails, at least he fails while daring greatly, so that his place shall never be with those cold and timid souls who knew neither victory nor defeat."

The key challenge in data science

"Ask yourselves, what problem have you solved, ever, that was worth solving, where you knew all of the given information in advance? Where you didn't have a surplus of information and have to filter it out, or you didn't have insufficient information and have to go find some?"

About us

Data intensive statistics in biology and medicine

- Brian Caffo
- Website http://www.bcaffo.com/
- Twitter [@bcaffo](https://twitter.com/bcaffo)
- Github https://github.com/bcaffo
- Jeff Leek
- Website http://biostat.jhsph.edu/~jleek/, http://simplystatistics.org/
- Twitter [@jtleek](https://twitter.com/jtleek)
- Github https://github.com/jtleek
- Roger Peng
- Website http://www.biostat.jhsph.edu/~rpeng/,http://simplystatistics.org/
- Twitter [@rdpeng](https://twitter.com/rdpeng)
- Github https://github.com/rdpeng

Why data science?

The Economist
Obama the warrior
Misgoverning Argentina
The economic shift from West to East
Genetically modified crops blossom
The right to eat cats and dogs
The data deluge
AND HOW TO HANDLE IT: A 14-PAGE SPECIAL REPORT

Why data science?

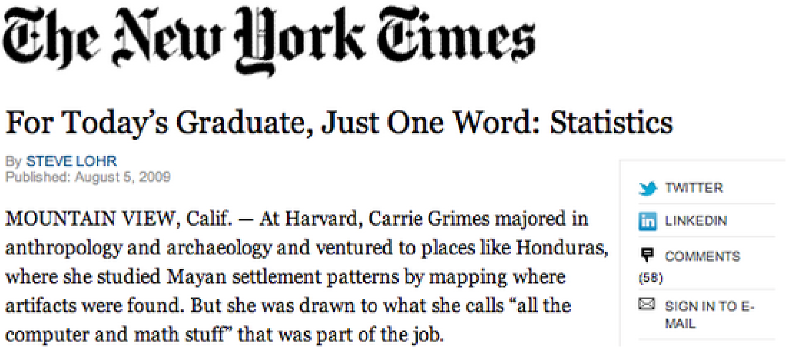
McKinsey Global Institute

June 2011

Big data: The next frontier for innovation, competition, and productivity

http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation

Why statistical data science?



The New York Times

For Today's Graduate, Just One Word: Statistics

By STEVE LOHR
Published: August 5, 2009

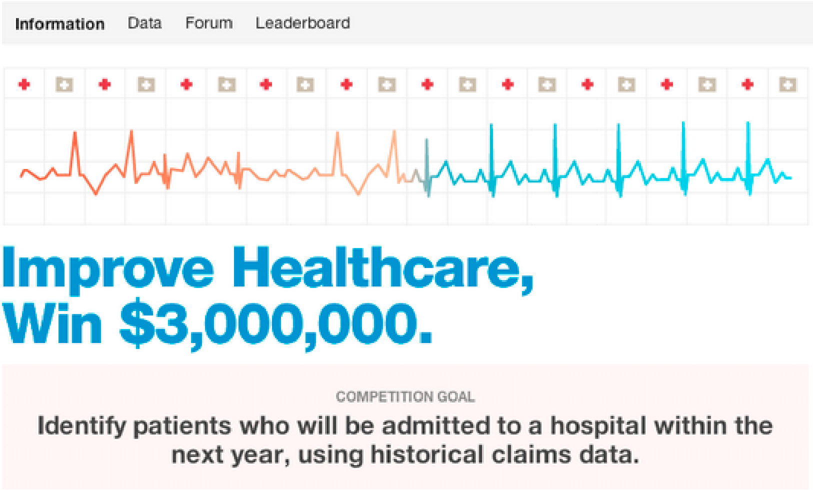
MOUNTAIN VIEW, Calif. — At Harvard, Carrie Grimes majored in anthropology and archaeology and ventured to places like Honduras, where she studied Mayan settlement patterns by mapping where artifacts were found. But she was drawn to what she calls "all the computer and math stuff" that was part of the job.

http://www.nytimes.com/2009/08/06/technology/06stats.html?_r=0

Why are you lucky?

DECEMBER 27, 1999 $3.95
www.time.com
TIME
PERSON OF THE YEAR
AMAZON.COM'S
JEFF BEZOS
E-COMMERCE IS
CHANGING THE WAY
THE WORLD SHOPS

Why are you lucky?



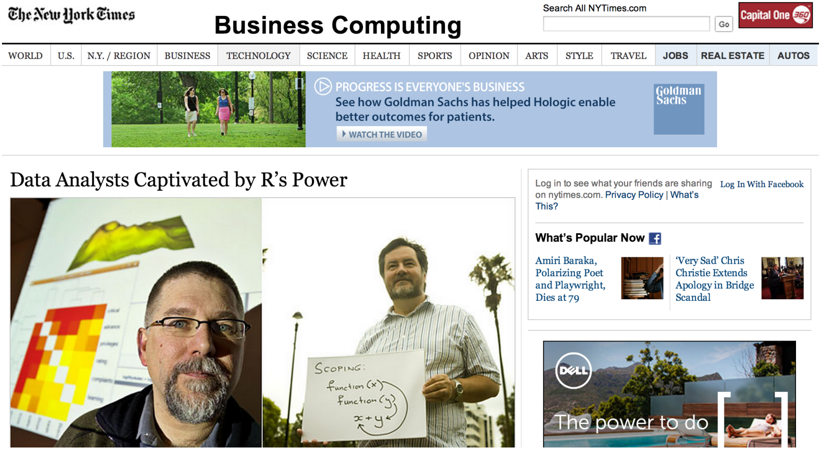
**Improve Healthcare,
Win $3,000,000.**

COMPETITION GOAL

Identify patients who will be admitted to a hospital within the next year, using historical claims data.

Heritage Health Prize

Why R?



Data Analysts Captivated by R's Power

Log in to see what your friends are sharing on nytimes.com. Privacy Policy | What's This?

What's Popular Now

Amiri Baraka, Polarizing Poet and Playwright, Dies at 79

'Very Sad' Chris Christie Extends Apology in Bridge Scandal

http://www.nytimes.com/2009/01/07/technology/business-computing/07program.html?pagewanted=all

Why R?

- It is free
- It has a comprehensive set of packages
- Data access
- Data cleaning
- Analysis
- Data reporting
- It has one of the best development environments - Rstudio `http://www.rstudio.com/`
- It has an amazing ecosystem of developers
- Packages are easy to install and “play nicely together”

Who is a data scientist?

Daryl Morey

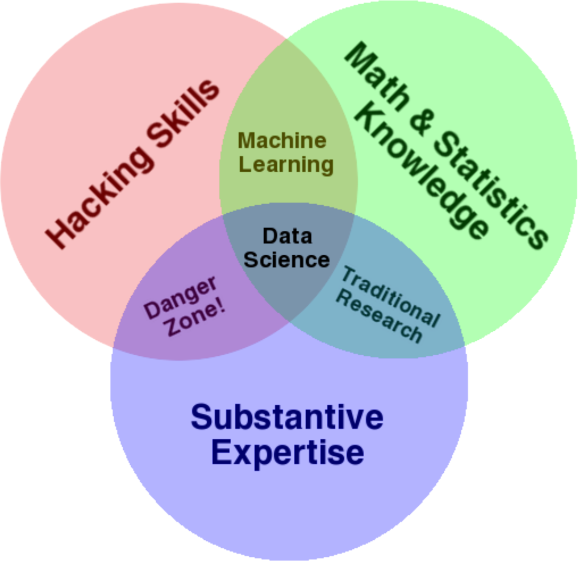
Who is a data scientist?

Hilary Mason

Who is a data scientist?

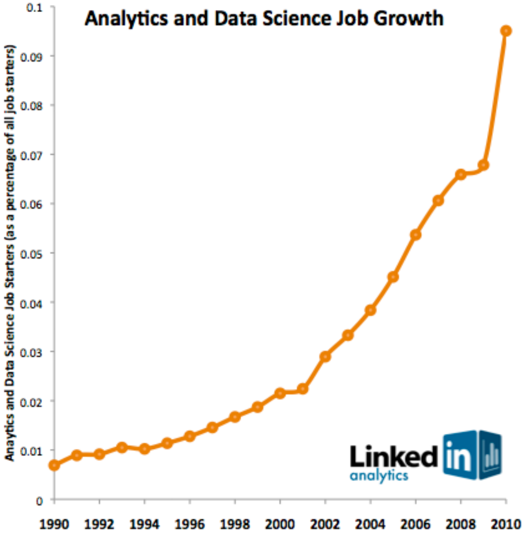
Who is a data scientist?

Our goal



Drew Conway

Plus jobs



http://radar.oreilly.com/2011/09/building-data-science-teams.html

This course

- Introducing you to the track
- Getting tools set up
- Giving you basic background