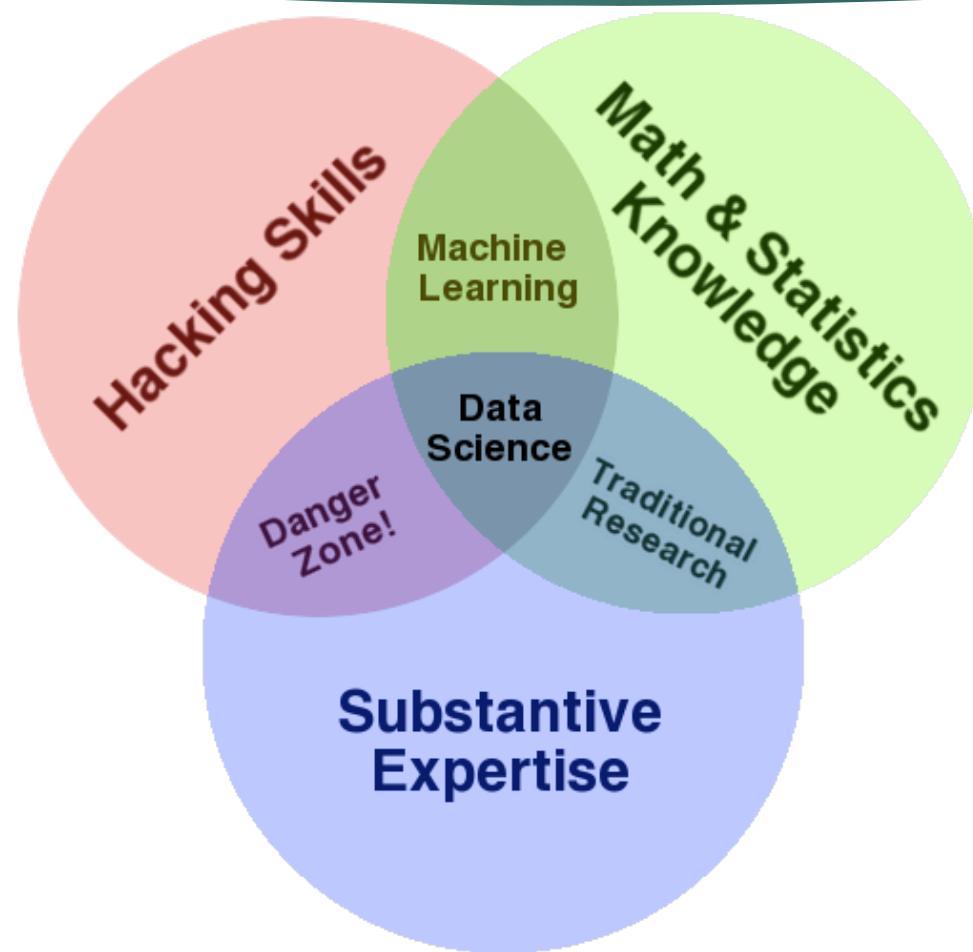


Introduction to Data Science

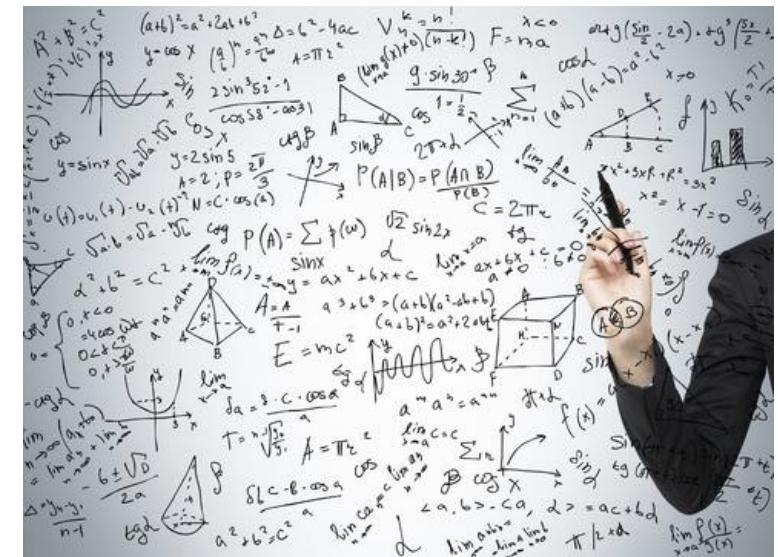
AFRAH SHAFQUAT

What is Data Science?

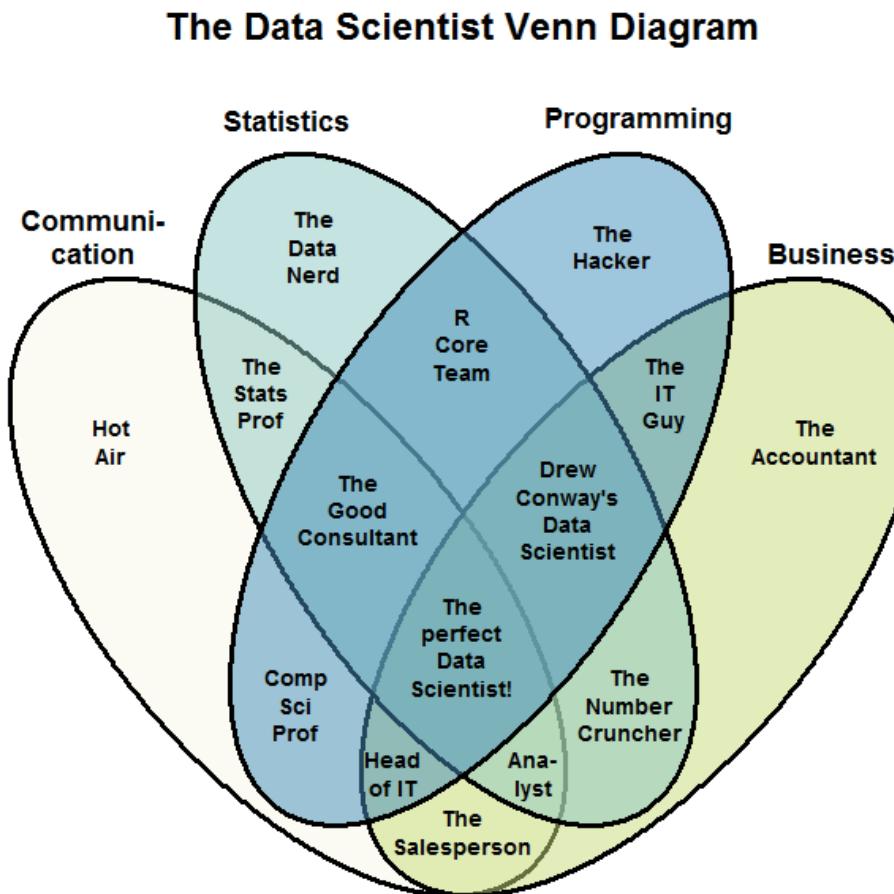


Where have we been and where are we going?

- ▶ Perspective: Human Genome
 - ▶ Then: 10 years, 2883 authors (finished in 2003)
 - ▶ Now: ~26 hours



What it means to be a Data Scientist



Applications

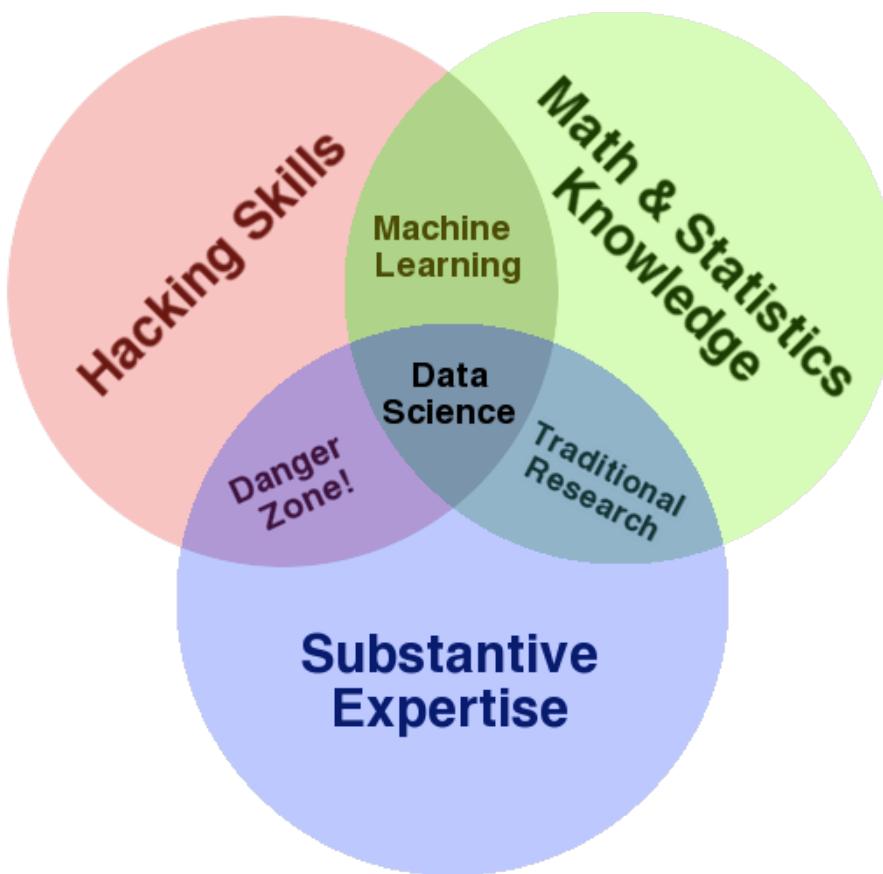
- ▶ Urban Living: Fighting Crime with Data (predictive policing; 200% police departments are interested)
- ▶ Healthcare: Medical Exams by bathroom mirrors (skin temperature, blood pressure)
- ▶ Predictive Diagnostics (90% of MRSA patients are predicted by new model; reducing readmissions)
- ▶ Better Fraud Detection
- ▶ Shopping Recommendations
- ▶ Electric Cars and Data Science (charging locations)

Live Coding

Where do we go from here

- ▶ Motivation
- ▶ MOOCS(Coursera, EdX, Udacity, IBM,)
- ▶ Tutorials
- ▶ Practice, practice and practice!

Targeting the BIG three



Build the foundation: Programming

- ▶ Why are you learning this?
 - ▶ **Jobs:** Targeted learning! Start with Python/R, and then advance towards the targeted working environment (Databases, scripting languages, data structures).
 - ▶ **Personal motivation:** Start with Python!
 - ▶ Highly recommend: Jupyter Notebooks
- ▶ Scientific libraries: numpy, scipy, matplotlib, pandas, seaborn

Foundation: Statistics/Math

- ▶ Search Keywords:
 - ▶ Introduction, Statistics,
 - ▶ Probability, Linear Algebra
- ▶ Sources:
 - ▶ MIT OpenCourseware ([ocw.mit.du](http://ocw.mit.edu))
 - ▶ Edx (edx.org)
 - ▶ Coursera (coursera.org)

- ▶ Other Free data science courses: <http://datascienceacademy.com/free-data-science-courses/>
- ▶ Before advancing on topics: look at pre-requisites!

The screenshot shows the MIT OpenCourseware homepage for the course 'Introduction to Probability and Statistics'. The page features a navigation bar with links for Home, Courses, Mathematics, Find Courses, About, Donate, Featured Sites, and a search bar. A banner at the top right encourages users to 'Subscribe to the OCW Newsletter'. The main content area displays the course title 'Introduction to Probability and Statistics' and several course links: COURSE HOME, SYLLABUS, CALENDAR, THIS COURSE AT MIT, READINGS, CLASS SLIDES, ASSIGNMENTS, and EXAMS. To the right of these links is a graph titled 'Bayesian updating with conjugate normal distributions' showing four overlapping bell curves in blue, cyan, purple, and green. Below the graph, a caption reads: 'Bayesian updating with conjugate normal distributions. (Image by Jerry Orloff and Jonathan Bloom.)'. At the bottom right, there are social sharing icons for Like, Tweet, G+, Share, and a submit button.

Alternative Learning Paths

- ▶ Sources:
 - ▶ Big Data University: <https://bigdatauniversity.com/learn/>
 - ▶ Coursera Data Science Courses: <https://www.coursera.org/browse/data-science?languages=en>
 - ▶ Potential Trajectory: Probability and Statistics > Data Analysis > Machine Learning
 - ▶ EdX: Microsoft Program in Data Science <https://www.edx.org/microsoft-professional-program-certificate-data-science>
 - ▶ Udacity Nanodegree: <https://www.udacity.com/nanodegree> (not free :/)
 - ▶ Others: <http://datasciencemasters.org/>
 - ▶ (P.S. Google is your best friend!)

Advanced topics: Machine Learning

- ▶ Andrew Ng's Machine Learning Course
(<https://www.coursera.org/learn/machine-learning>)
 - ▶ Caution: If you are having trouble understanding this course, you might want to revisit your foundations and take more courses.
- ▶ Other machine learning courses:
 - ▶ Search keywords: Machine Learning, Data Science
 - ▶ Sources:
 - ▶ Big Data University <https://bigdatauniversity.com/>
 - ▶ Coursera
 - ▶ EdX
- ▶ If using python, scikit-learn and its documentation are a great resource as well

Practice, practice, Practice

- ▶ Kaggle (<https://www.kaggle.com/>)
- ▶ Tutorials
- ▶ Public datasets
 - ▶ <https://www.quora.com/Where-can-I-find-large-datasets-open-to-the-public>
 - ▶ <https://github.com/caesar0301/awesome-public-datasets>

Lists of lists of lists

- ▶ Data Science Weekly
- ▶ KD Nuggets (<http://www.kdnuggets.com/>)
- ▶ <http://www.kdnuggets.com/faq/learning-data-mining-data-science.html>
- ▶ <https://github.com/bulutyazilim/awesome-datascience>
- ▶ <https://github.com/josephmisiti/awesome-machine-learning>
- ▶ <http://www.computerworld.com/article/3114136/big-data/get-started-in-data-science-5-steps-you-can-take-online-for-free.html>
- ▶ <https://www.linkedin.com/pulse/how-become-data-scientist-free-zeeshan-ul-hassan-usmani>