Distinguishing Data Science from...

- Business Intelligence
- Statistics
- Data(base) Management
- Visualization
- Machine Learning

Huge number of relevant courses, new and existing.

- Concepts in Computing with Data, Berkeley
- Practical Machine Learning, Berkeley
- Artificial Intelligence, Berkeley
- Visualization, Berkeley
- Data Mining and Analytics in Intelligent Business Services, Berkeley
- Data Science and Analytics: Thought Leaders, Berkeley
- Scalable Machine Learning, Berkeley
- Analyzing Big Data with Twitter, Berkeley
- Machine Learning, Stanford
- Paradigms for Computing with Data, Stanford
- Mining Massive Data Sets, Stanford
- Data Visualization, Stanford
- Algorithms for Massive Data Set Analysis, Stanford
- Research Topics in Interactive Data Analysis, Stanford
- Data Mining, Stanford
- Machine Learning, CMU
- Statistical Computing, CMU
- Machine Learning with Large Datasets, CMU
- Machine Learning, MIT
- Data Mining, MIT
- Statistical Learning Theory and Applications, MIT
- Data Literacy, MIT
- Introduction to Data Mining, UIUC
- Learning from Data, Caltech
- Introduction to Statistics, Harvard
- Data-Intensive Information Processing Applications, University of Maryland
- Statistical Inference, UPenn
- Introduction to Data Science, Columbia
- Dealing with Massive Data, Columbia
- Data-Driven Modeling, Columbia
- Introduction to Data Mining and Analysis, Georgia Tech
- Computational Data Analysis: Foundations of Machine Learning and Data Mining, Georgia Tech
- Applied Statistical Computing, Iowa State
- Data Visualization, Rice
- Data Warehousing and Data Mining, NYU
- Data Mining in Engineering, Toronto
- Machine Learning and Data Mining, UC Irvine
- Knowledge Discovery from Data, Cal Poly
- Large Scale Learning, University of Chicago
- Data Science: Large-scale Advanced Data Analysis, University of Florida
- Strategies for Statistical Data Analysis, Universität Leipzig
- Data Analysis, Johns Hopkins (via Coursera)
- Computing for Data Analysis, Johns Hopkins (via Coursera)

"I worry that the Data Scientist role is like the mythical "webmaster" of the 90s: master of all trades."

-- Aaron Kimball, CTO Wibidata

What "data science" tells me:

- If you're a DBA, you need to learn to deal with unstructured data
- If you're a statistician, you need to learn to deal with data that does not fit in memory
- If you're a software engineer, you need to learn statistical modeling and how to communicate results.
- If you're a business analyst, you need to learn about algorithms and tradeoffs at scale

Breadth

tools ⟷ abstractions

Hadoop	MapReduce
PostgreSQL	Relational Algebra
glm(…) in R	Logistic Regression
Tableau	InfoVis

Depth

Management

Relational Algebra

Standards

Analysis

Linear Algebra

ad hoc files

Scale

main memory

R

local files

distributed

Hadoop

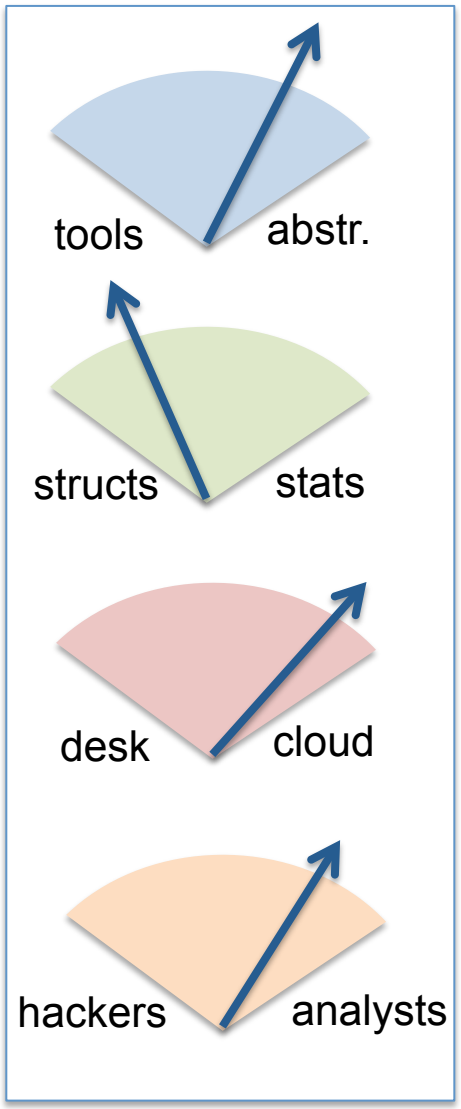
S3, Azure Storage

Target

hackers ⟷ analysts

Assume proficiency in Python, Java, R

Assume little or no programming


tools
abstr.
structs
stats
desk
cloud
hackers
analysts