What is Data Science

Introduction to Data Science Algorithms

Jordan Boyd-Graber and Michael Paul

AUGUST 23, 2016

What you need for this course

- You need to use Python
- Helps to have a laptop to bring to class
- Math background
  - Logarithms, Exponents
  - Take derivatives
  - Algebraic manipulation
- Computer / programming skills
  - Must know how to program
  - Manipulate data (text files)
  - Algorithms relatively simple

Classroom Style

- Hands-on practice
- Lectures: do reading, ask questions
- Labs: you help each other, and we work through examples

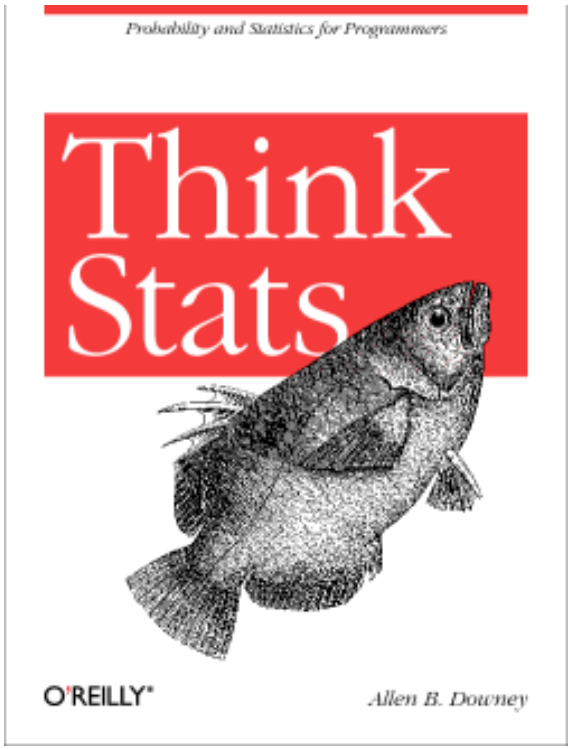
Administrivia

- Keep track of course webpage
- Homeworks: 5 late days
- Midterm
- Final
- Let me know about special needs

Administrivia

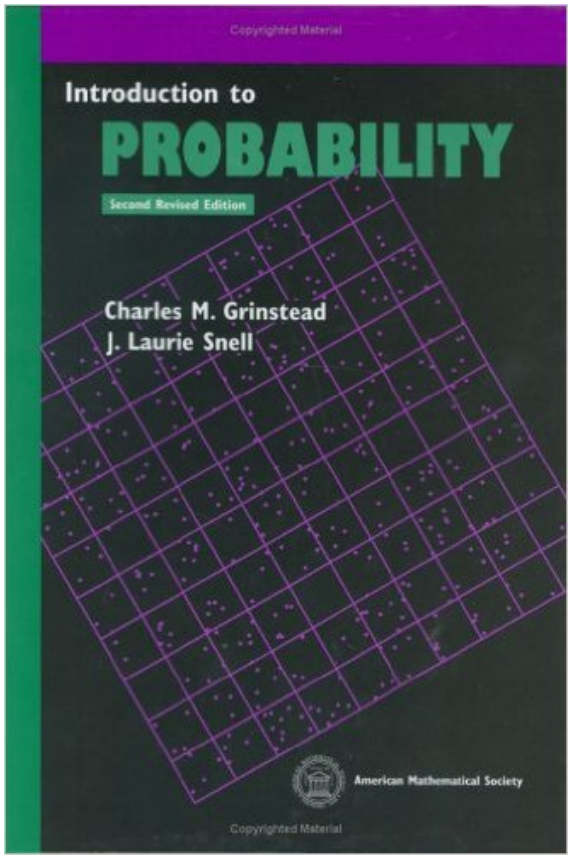
- Keep track of course webpage
- Homeworks: 5 late days
- Midterm
- Final
- Let me know about special needs
- Read the syllabus!
  - Grade breakdown
  - Policies on lateness beyond free late days

Course reading



- We will provide reading materials, mostly from the book.
- Slightly different focus: same concepts, use book as starting point

Course reading



- We will provide reading materials, mostly from the book.
- Slightly different focus: same concepts, use book as starting point
- Statistics will be from suggested book

Communicating with Piazza

We will use Piazza to manage all communication

`http://piazza.com/colorado/fall2016/csci3022`

- Questions answered within 1 day (hopefully sooner)
- Hosts discussions among yourselves
- Use for any kind of technical question
- Use for **most** administrative questions
- Can use to send us private questions too
- Will be a factor in participation

How to ask for help

- Explain what you're trying to do
- Give a minimal example
  - Someone else should be able to replicate the problem easily
  - Shouldn't require any data / information that only you have
- Explain what you **think** should happen
- Explain what you get instead (copy / paste or screenshot if you can)
- Explain what else you've tried

Me

- Seventh year assistant professor
  - Office: 111B ECCS
- Was formerly a professor at University of Maryland
- Research: topic models, question answering, machine translation
- First time teaching **this** class (taught several related courses)
- Born in Colorado (where all my family live)
- Grew up in Iowa (hometown: Keokuk, Iowa)
- Went to high school in Arkansas
- Undergrad in California
- Grad school in New Jersey
- Brief jobs in between:
  - Working on electronic dictionary in Berlin
  - Worked on Google Books in New York
- ying / jbg / jordan / boyd-graber

Research

Research

human	evolution	disease	computer
genome	evolutionary	host	models
dna	species	bacteria	information
genetic	organisms	diseases	data
genes	life	resistance	computers
sequence	origin	bacterial	system
gene	biology	new	network
molecular	groups	strains	systems
sequencing	phylogenetic	control	model
map	living	infectious	parallel
information	diversity	malaria	methods
genetics	group	parasite	networks
mapping	new	parasites	software
project	two	united	new
sequences	common	tuberculosis	simulations

Research

Tea Party in the House

Who's who

- Michael Paul: authoring slides (don't bug him!)
- Pedro Rodriguez: concept help (do bug him!)
- Apoorva Bapat: grader (homework grades only)

Next time

- Data wrangling
- LAB! (Bring laptop)
- Subject of first homework