Conclusion
Life After Data 8

- Develop methodology or apply in specific domain:
  - Most connectors (data.berkeley.edu)
  - Stat 28

- Go deeper into the theory and code:
  - Stat 88, Stat 89A, CS 88

- Go even deeper:
  - Course doesn’t have to say “data science” in title
  - Berkeley has been teaching data science for years
Basic Tools

Programming: CS 61A or Data 8 + CS 88

Linear algebra: Math 54 or EE 16A; Math 110 if you want to understand what’s really going on

Calculus: Math 1A/1B or Math 10; preferably Math 53
Probability and Inference Core

- **Probability:**
  - Here’s the model; what can you say about the sample?
  - Stat 134 or Prob140

- **Inference:**
  - Here’s the sample; what can you say about the model?
  - Stat 135 (requires 134/140)
Deeper into Data and Stochastics

- Visualization, scraping/munging, analysis; R: Stat 133
- Multiple variables, inference, prediction:
  - Stat 151A: linear models, multiple regression, principal components
  - Stat 154: machine learning (or CS 189)
- Random variables indexed by time or space:
  - Stat 150 (stochastic processes; probability)
  - Stat 153 (time series; inference)
Computer Science Depth

- CS 61B
- DS (CS/Stat) 100
- 189 (machine learning)

Any of these:
- 186 (databases)
- 170 (efficient algorithms and intractable problems)
- 169 (software engineering)
- 162 (operating systems and systems programming)
Don’t Be Crazy

- You can’t do all of it
- Just pick bits that you like in each discipline
- And you’ll be a fine data scientist
Plan for RRR Week

- **Mon during lecture**, I will review:
  - Inference methods, then all the “mathy” stuff
- **Tue 12-4**: I will have office hours in 413 Evans
- **Wed during lecture**: I will continue with “mathy” etc
- GSIs will review **Wed and Thurs during lab times**:
  - First hour: review problems on particular topic
  - Second hour: office hour
  - Topics and review leaders TBA; watch Piazza
- **Fri**: Go see a dumb movie or relax in some other way
Final Exam

- Final Exam
  - Monday December 12, 8:00 - 11:00
  - RSF Field House
  - Bring something to write with and something to erase with; but not your breakfast. Water is OK.
  - We will provide a couple of reference sheets, with drafts posted on Piazza during RRR week
  - Exam will look like midterm, with more questions
  - More on the final on Monday of RRR week
Big Picture of Course Contents

1. Python
2. Describing data
3. General concepts of inference
4. Theory of probability and statistics
5. Methods of inference
1. Python

- Textbook sections
  - General features and Table methods: 3.1 - 8.2, 15.3
  - np.median: 9.3
  - proportions_from_distribution: 10.1
  - percentile: 11.1
  - np.mean, np.std: 12.1, 12.2
  - stats.norm.cdf: 12.3
  - minimize: 13.3
2. Describing Data

- **Tables**: Chapter 5
- **Classifying and cross-classifying**: 7.2, 7.3
- **Distributions and visualization**: Chapter 6, 7.5
- **Center and spread**: 12.1-12.3
- **Linear trend and non-linear patterns**: 7.1, Chapter 13
3. General Concepts of Inference

- Study, experiment, treatment, control, confounding, randomization, causation, association: Chapter 2
- Distribution: 6.1, 6.2
- Sampling, probability sample: 8.5
- Probability distribution, empirical distribution, law of averages: 9.1
- Population, sample, parameter, statistic, estimate, bias, variability: 9.3
- Model: 10.2, 14.1, every null and alternative hypothesis