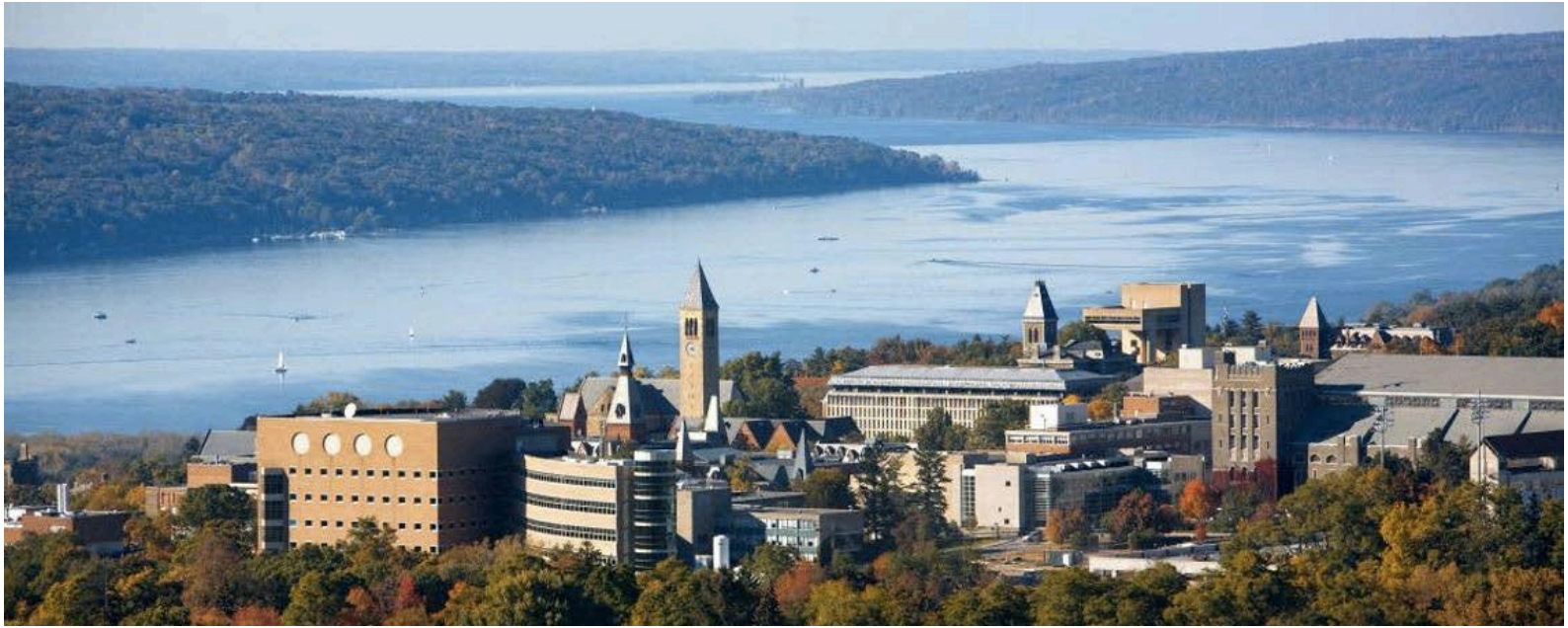
Data Science for All (CS/ORIE/STSCI 1380) and Vocareum

David P. Williamson

Cornell University, School of Operations Research and Information Engineering



A bit about me

- At Cornell since 2004, joint appointment between ORIE and Information Science.
- Previously worked for IBM Research (1995-2003).
- Research interests in discrete/ combinatorial optimization, algorithms.
- Why I'm teaching data science: because the assigned prof left for another university, and I volunteered to cover the course.

The class

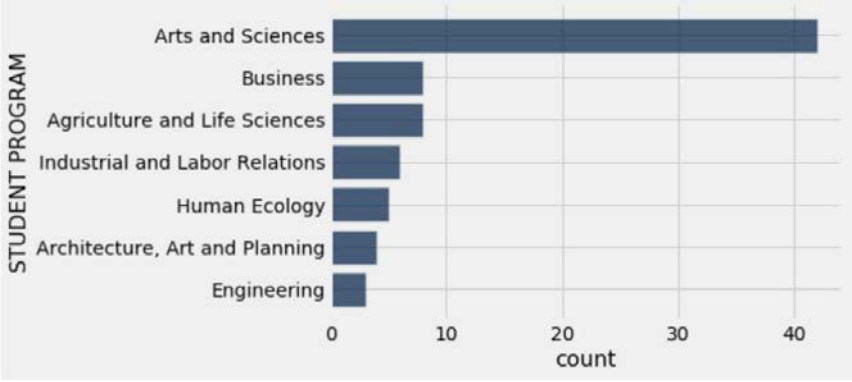
- Imported from Berkeley's Data 8: The Foundations of Data Science, in 2018.
- Taught at Berkeley Fall, Spring, Summer since Fall 2015; very large enrollments.
- Textbook, assignments (from Spring 2017, Fall 2016), lecture videos (from Fall 2016) all available at data8.org; lecture slides available upon request.
- Lecture demos/assignments all use Python in Jupyter notebooks; Python package datascience developed just for the course.
- Cornell site: cornell-dsfa.org.

Ani Adhikari

John DeNero

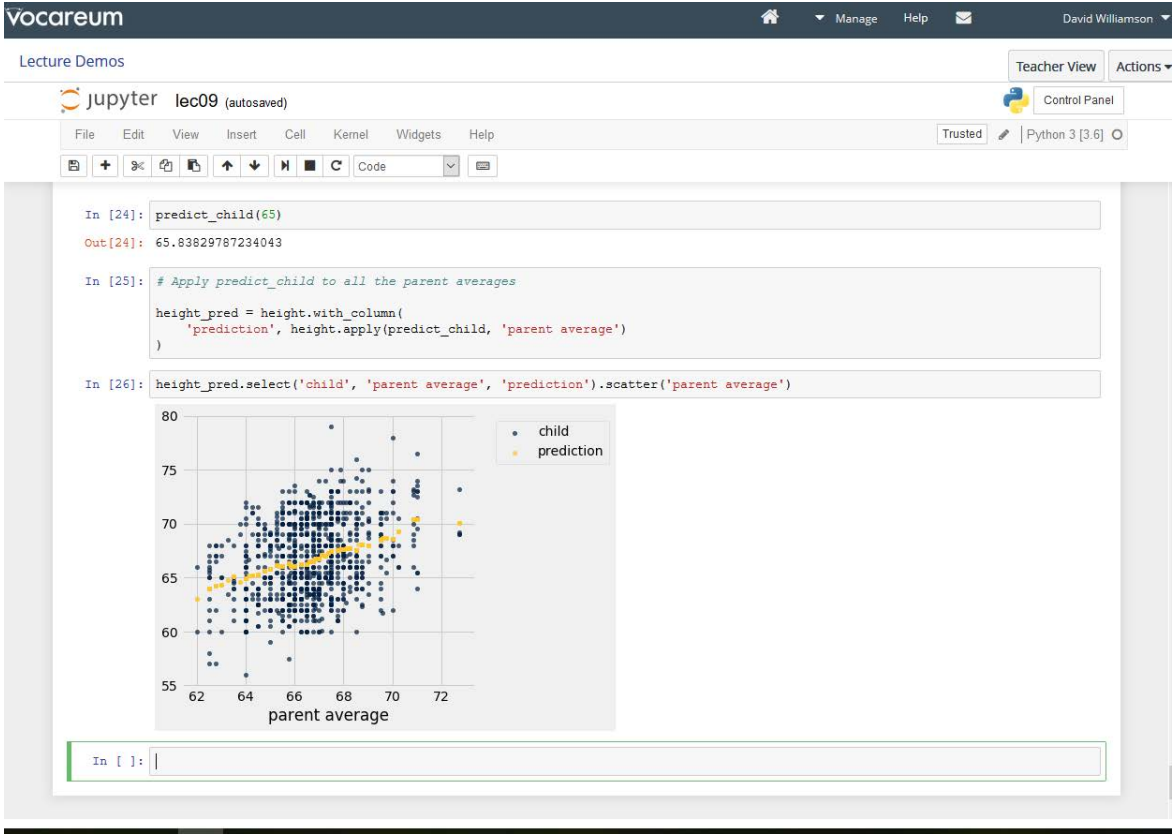
Course goals

- Intro to data science for students who have not had any programming or statistics.
- Common remark: “This is the first college STEM course I’ve taken’’ or “the only one I plan to take.”
- Meets an Arts & Sciences requirement for Math and Qualitative Reasoning.



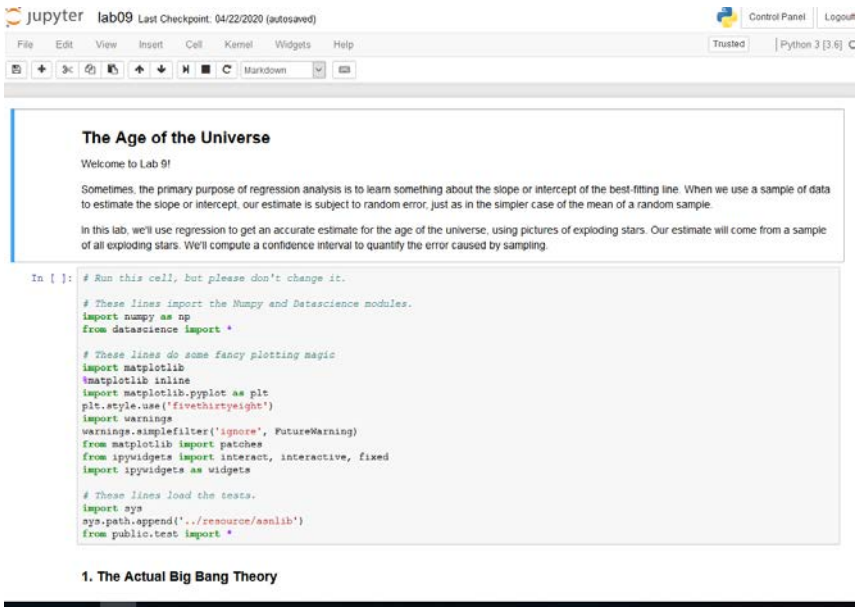
Course coverage

- Intro to Python, with focus on manipulating data in tables, visualizing data.
- Intro to probability and statistics: parameter estimation, hypothesis testing, correlation, regression.
- Intro to machine learning: classification via nearest neighbors.



Types of assignments

- All Jupyter notebook/Python
- Weekly lab sections: TA-led assignment, students work in pairs
- Weekly homework: students work on their own
- Three projects: two-week long homework assignments, students can work in pairs
  - Global population
  - Death penalty
  - Classifying movies into action/ romance based on script



Data Used in Lectures, Assignments

- Text of books
- Movies and actors
- Population (US Census)
- Baby birth weight
- NYC Bikeshare trips
- Chronic kidney disease
- Voter database
- Athlete performance
- Flight delays
- Global poverty
- Death penalty and murder rates
- Movie scripts
- World population
- Farmers markets
- Size and age of universe
- Old Faithful eruptions
- Many, many more!

Vocareum

- All assignments and lecture demos hosted on Vocareum.
- Students pay a nominal fee for using the site for the term.
- Initial port of Data 8 assignments to Vocareum by the 2018 Cornell instructors, including ability to provide student feedback (e.g. is answer close to correct?) and autograde certain questions.

Madeleine Udell

Michael Clarkson

Why Vocareum?

- We assume ability to use web browsers, but no other computer competencies.
- No need for students to install Python/Jupyter on their own machine; no need for staff to manage inevitable installation problems.
- Very easy to use grading capabilities. Autograding also easy to use (once you've set it up).

Why Vocareum?

- Specific features we used:
  - Integration with Canvas (grades, assignment posting)
  - Bonus points for early submissions (can also set penalties for late submissions)
  - Unlimited submissions, ability to see autograder scoring question by question
  - “Slip days” for late submissions
  - Ability to set student-specific submission deadlines

The 2020 update

- In mid-March, Cornell sent all students home, to restart classes online three weeks later.
- Advantage Vocareum:
  - We didn’t need to change what we were doing for distributing/collecting/grading assignments. Unlike many (most?) other classes on campus, this part of the course stayed the same for our students.

Questions?

Berkeley: data8.org

Cornell: cornell-dsfa.org

davidpwilliamson@cornell.edu