Stat 108
Spring 2022

Foundations of Data Science

Module 1 (lec01)

Introduction

1

Welcome to 108: Foundations of Data Science

Yuyu Tang
Instructor

Haoran Hu
TA

2

Data Science

3

What is Data Science?

Drawing useful conclusions from data using computation

- **Exploration**
  - Identifying patterns in information
  - Uses visualizations
- **Inference**
  - Using data to draw reliable conclusions
  - Uses statistics
- **Prediction**
  - Making informed guesses
  - Uses machine learning

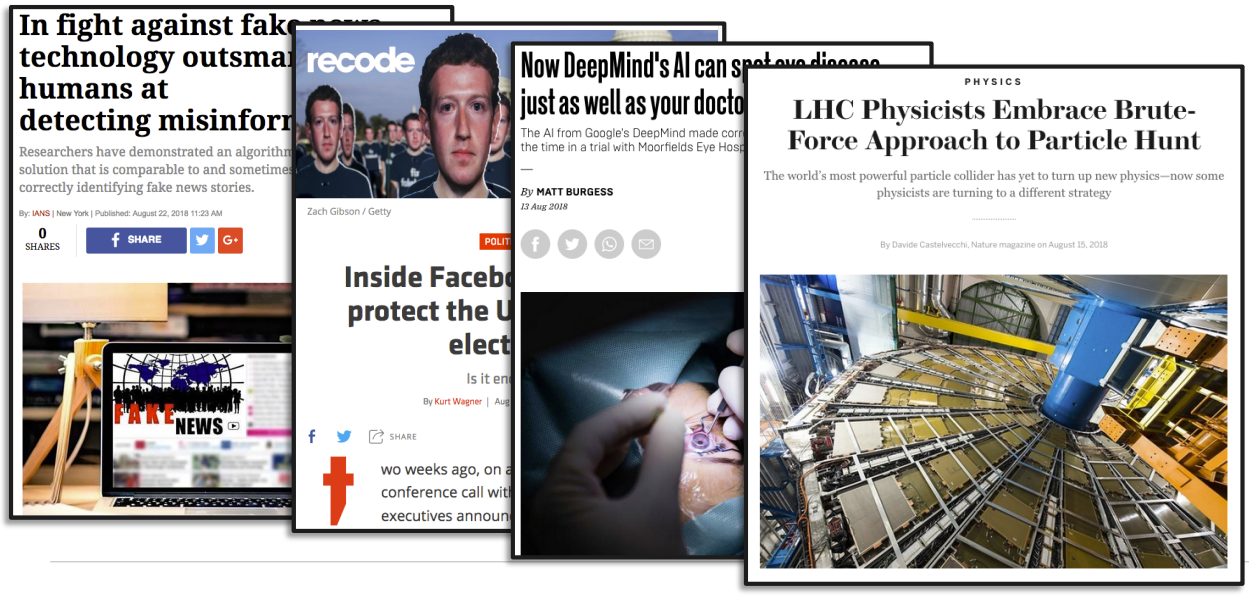
4

Applications

- Data science is driven by applications
- Data analysis is playing an increasingly important role in many fields including Biology, Chemistry, Economics, Earth Systems, Education, Environmental Science, Finance, Geography, Geology, Kinesiology, Linguistics, Management, Political Science, Public Health, Psychology, Sociology, …
- Every data-driven subject brings new challenges

5

Examples



6

Course Structure

7

What does the course cover?

- An introduction to programming in Python with a focus on manipulating, visualizing, and analyzing data.
- An introduction to statistics that is grounded in computer simulations.
- An introduction to predictive modeling and machine learning.

8

Course Components and Grading

- Lectures (often interactive)
- Weekly labs including attendance (pass/fail)
- Weekly graded homework assignments
- Midterm exam & final exam

Homework	35%
Labs	20%
Midterm Exam	20%
Final Exam	25%

9

Course Technology

- **Moodle:** Course website and gradebook (lecture slides, demos(notebook), assignments, labs, etc.)
- **Piazza:** Online discussion forums, course Q&A, announcements, and instructor DM.
- **Github.io:** Free, online textbook
- **DataHub:** Web-based Python compute environment for completing labs and homework assignments.
- **Gradescope:** website for Homework assignments and exams submission and grade.
- **Links to all resources can be found on Moodle.**

10

Course Policies

- Late Homework
- Re-grades
- Academic Honesty
- https://umass-data-science.github.io/190fwebsite/policies/

11

Collaboration Policy

Asking questions is highly encouraged

- You can discuss homework and lab questions with each other
- Do not take notes or pictures out of discussions
- The work you turn in must be your own

The limits of collaboration

- You may work each other on Labs and HW, but you cannot cut and paste code. You must type all the code yourself!
- Copying solutions from any source will be dealt with under UMass' Academic Honesty procedures: https://www.umass.edu/honesty/

12

Getting Help

- The course staff are here to help you be successful in the course!
- When you need help post on the Piazza discussion forums or/and stop by office hours.
- The lab sessions are also a good time to ask questions and get help.

13

Let's Dive In!

14

Demo: 1.1.ipynb and 1.2.ipynb

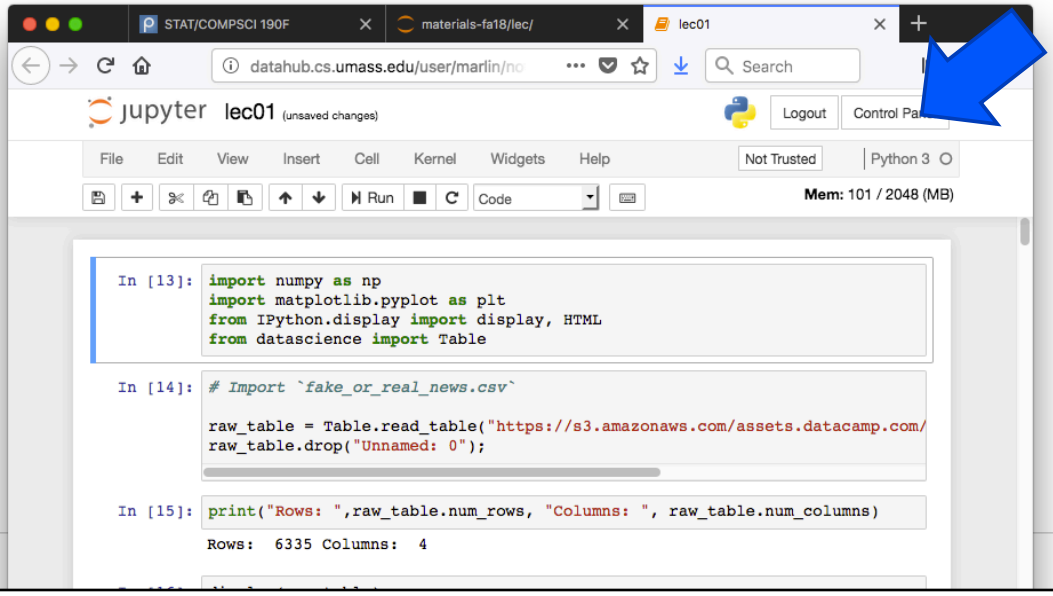
15

Datahub

- If you have an @umass.edu email address, you should now be able to access the course's data hub.
- Datahub uses UMass Google Apps authentication. Use your @umass.edu email address and Spire password to log in. It takes a minute to start up.
- When you're done working with the Datahub, make sure to shut your datahub server down and then log out.

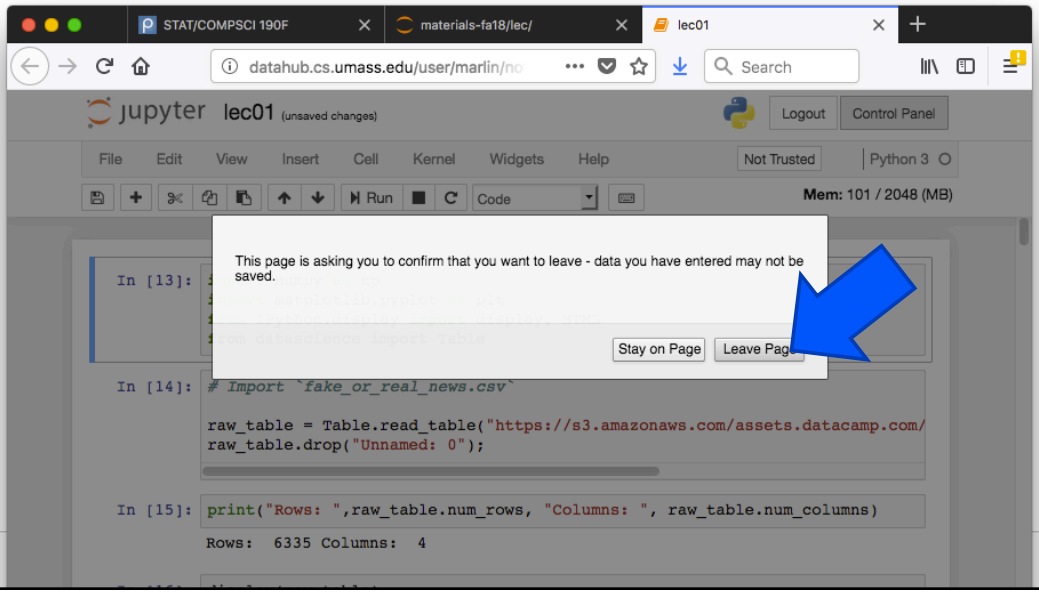
16

Stopping Your Data Hub



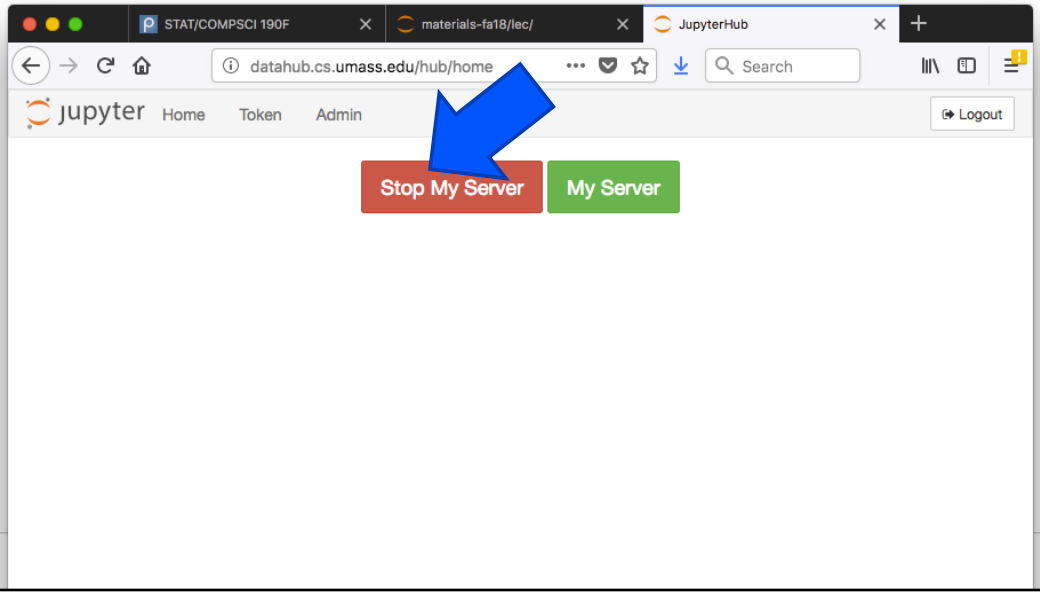
17

Stopping Your Data Hub



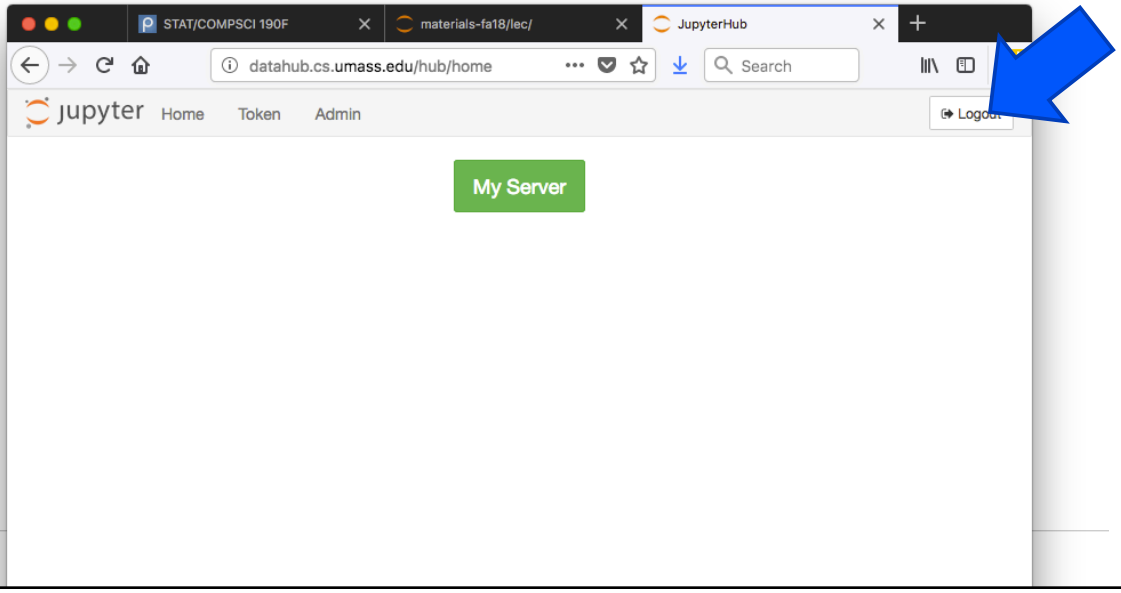
18

Stopping Your Data Hub



19

Stopping Your Data Hub



20