Introduction to Data Science

STAT 2600: Spring 2020

Time & Location:

- Tuesdays & Thursdays from 2:00pm-3:15pm in Engineering Center ECCR 265.
- *Lab Sections:* Wednesdays 2:00pm-2:50PM in Ramaley Biology Bldg N1B31 and 3:00pm-3:50pm in Engineering Center ECCR 118

Professor

- Dr. Maziar Raissi: maziar.raissi@colorado.edu
- Office hours (ECOT 337): Tuesdays & Thursdays from 3:30pm-5:00pm or by appointment

Teaching Assistants

- Nicholas Varberg: nicholas.varberg@colorado.edu
- Ashton Wiens: ashton.wiens@colorado.edu

Course Websites

We will use Canvas and GitHub for grading, readings, and assignments.

Textbooks

- Our textbook for the first half of the class is **R for Data Science** by *Garrett Grolemund* and *Hadley Wickham* available for free at https://r4ds.had.co.nz/ (https://r4ds.had.co.nz/)
- Our textbook for the second half of the class is **Python Data Science Handbook** by *Jake VanderPlas* available for free at https://jakevdp.github.io/PythonDataScienceHandbook/ (https://jakevdp.github.io/PythonDataScienceHandbook/)

Course Description

This course introduces importing, tidying, exploring, visualizing, summarizing, and modeling data and then communicating the results of these analyses to answer relevant questions and make decisions. Students will learn how to program in R and Python using reproducible workflows. During weekly lab assignments students will collaborate with their teammates to pose and answer questions using real-world datasets.

Overall Learning Objective

To develop technical and professional skills necessary to analyze data as a member of a team. This includes:

- Understanding fundamental statistical concepts
- Visualizing and exploring data
- Importing and tidying datasets
- Programming effectively in R & Python
- Building basic statistical models
- Collaborating with teammates to discover and communicate interesting findings and recommendations based on data.
- Mastering reproducible statistical workflows.

In other words, to learn R & Python to do interesting and useful things with data.

Grading

This course consists of 10 modules. At the end of each module, there will be either an indiviual or a group assignment. The overall course grades will be determined by these two components:

- 60% Individual Performance: 6 individual assignments will determine the individual performance grade.
- 40% Team Performance: 4 group assignments will determine the team performance grade.

Honor Code

Submit original work in your own words for all individual assignments. All students enrolled in a CU Boulder course are responsible for knowing and adhering to the Honor Code. Violations of the policy may include: plagiarism, cheating, fabrication, lying, bribery, threat, unauthorized access to academic materials, clicker fraud, submitting the same or similar work in more than one course without permission from all course instructors involved, and aiding academic dishonesty. All incidents of academic misconduct will be reported to the Honor Code Council (honor@colorado.edu; 303-735-2273). Students who are found responsible for violating the academic integrity policy will be subject to nonacademic sanctions from the Honor Code Council as well as academic sanctions from the professor. Additional information regarding the academic integrity policy can be found at the Honor Code Office website (http://www.colorado.edu/honorcode/).

Respect for Other Students

All students must be treated with respect during the class period, during team work outside of class, and during coursework online.

Modules and Chapter Readings

The course content will be divided into the following two parts.

- R for Data Science (https://r4ds.had.co.nz/ (https://r4ds.had.co.nz/)):
  - Module 0 - Workflows: Anaconda; Jupyter Notebook; Hello R; RStudio; RMarkdown; GitHub; Atom (https://atom.io/) (Ch. 1, 26, 27, 30)
  - Module 1 - Data visualization: Using ggplot2 in R to visualize data (Ch. 2-4)
  - Module 2 - Exploratory Data Analysis (EDA): Transforming data and EDA (Ch. 5-8)
  - Module 3 - Importing and tidying data: Using readr; tibbles; tidying data (Ch. 9-12)
  - Module 4 - Working with various types of data: Relational data; strings; factors; dates and times (Ch. 13-16)
  - Module 5 - Programming in R: Pipes; functions; vectors; iteration (Ch. 17-21)
  - Module 6 - Statistical modeling: Building and using models to gain insight into data, answer questions, and solve problems (Ch. 22-25).

- Python Data Science Handbook (https://jakevdp.github.io/PythonDataScienceHandbook/ (https://jakevdp.github.io/PythonDataScienceHandbook/)):
  - Module 7 - Introduction to NumPy (Ch. 1-2)
  - Module 8 - Data Manipulation with Pandas (Ch. 3)
  - Module 9 - Visualization with Matplotlib (Ch. 4)
  - Module 10 - Machine Learning (Ch. 5)

To Get Started with Git

1. Register for an account on github.com (https://github.com/).
2. Apply for student benefits on https://education.github.com/benefits (https://education.github.com/benefits).
3. Download, install and configure git (https://git-scm.com/). Windows users please install Git for Windows (https://gitforwindows.org/).
4. Make sure to set Git with your name and email address using the following commands on the command-line with your name and email address:

   ```
   $ git config --global user.name "FirstName LastName"
   $ git config --global user.email "email@example.com"
   ```

   This is important because Git will use this information when you work on a project.

Accommodations

If you qualify for accommodations because of a disability, please submit your accommodation letter from Disability Services to your faculty member in a timely manner so that your needs can be addressed. Disability Services determines accommodations based on documented disabilities in the academic environment. Information on requesting accommodations is located on the Disability Services website (http://www.colorado.edu/disabilityservices/students). Contact Disability Services at 303-492-8671 or dsinfo@colorado.edu for further assistance. If you have a temporary medical condition (http://www.colorado.edu/disabilityservices/students/temporary-medical-conditions) or injury, see Temporary Medical Conditions under the Students tab on the Disability Services website.

Religious Holidays

Campus policy regarding religious observances requires that faculty make every effort to deal reasonably and fairly with all students who, because of religious obligations, have conflicts with scheduled exams, assignments or required attendance. In this class, please talk to me before being absent for a religious observance. See the campus policy regarding religious observances (http://www.colorado.edu/policies/observance-religious-holidays-and-absences-classes-andor-exams) for full details.

Classroom Behavior

Students and faculty each have responsibility for maintaining an appropriate learning environment. Those who fail to adhere to such behavioral standards may be subject to discipline. Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with race, color, national origin, sex, pregnancy, age, disability, creed, religion, sexual orientation, gender identity, gender expression, veteran status, political affiliation or political philosophy. Class rosters are provided to the instructor with the student's legal name. I will gladly honor your request to address you by an alternate name or gender pronoun. Please advise me of this preference early in the semester so that I may make appropriate changes to my records. For more information, see the policies on classroom behavior (http://www.colorado.edu/policies/student-classroom-and-course-related-behavior) and the Student Code of Conduct (http://www.colorado.edu/osccr/).

Sexual Misconduct, Discrimination, Harassment and/or Related Retaliation

The University of Colorado Boulder (CU Boulder) is committed to fostering a positive and welcoming learning, working, and living environment. CU Boulder will not tolerate acts of sexual misconduct intimate partner abuse (including dating or domestic violence), stalking, protected-class discrimination or harassment by members of our community. Individuals who believe they have been subject to misconduct or retaliatory actions for reporting a concern should contact the Office of Institutional Equity and Compliance (OIEC) at 303-492-2127 or cureport@colorado.edu. Information about the OIEC, university policies, anonymous reporting (https://cuboulder.qualtrics.com/jfe/form/SV_0PnqVK4kkIJIZnf), and the campus resources can be found on the OIEC website (http://www.colorado.edu/institutionalequity/).

Please know that faculty and instructors have a responsibility to inform OIEC when made aware of incidents of sexual misconduct, discrimination, harassment and/or related retaliation, to ensure that individuals impacted receive information about options for reporting and support resources.

Disclaimer

We reserve the right to modify the syllabus as needed during the course of the semester.