Course Description

In this course, you will learn to pose interesting questions and draw meaningful conclusions from data. It is an introductory statistics course, with a focus on data analysis. We will also cover the basic theory behind the methods that we use in the course. Much of the course will focus on analyzing datasets drawn from a number of different fields. We will discuss case studies using these datasets, and you will conduct your own analyses.

In this course you will learn to do data analysis using the **Python** (https://python.org/) programming language. Specifically, you will learn to use a few Python libraries that support data analysis, especially **Pandas** (https://pandas.pydata.org/) . You do not need to have prior experience with computer programming, or with Python. This is not a programming class, and we will only cover the programming topics that we need in order to accomplish our data analysis goals. You will gain some proficiency in Python, but this course will not give you a systematic and thorough introduction to Python or to computer science. Instead, the main goal of the course is to introduce you to the conceptual and practical aspects of rigorous data analysis.

Instructors

- Lead Instructor: Dr. Mark Fredrickson
  - Email: **mfredric@umich.edu** (mailto:mfredric@umich.edu)
  - Zoom Office Hours using **Office Hours Queue** (https://officehours.it.umich.edu/queue/132)
    - Wednesday: 2 - 4pm
    - Thursday: 3 - 5pm
    - Friday: 9 - 11am
  - LEO LECTURERS' UNION AFT-MI LOCAL 6244 THIS COURSE IS TAUGHT BY A LEO LECTURER
- GSIs (you may attend any GSI's office hours):
  - Rachel Himmel, **rhimmel@umich.edu** (mailto:rhimmel@umich.edu) , Office Hours: Tuesday 1:00-2:30 PM, Thursday 8:30-10:00 AM, via **Zoom** (https://umich.zoom.us/j/95669945012)
  - Sang Hoo Kook, **sangkook@umich.edu** (mailto:sangkook@umich.edu) , Office Hours: Tuesday 11:30am-1pm G219 Angell Hall, Thursday 1pm-2:30pm via **Zoom** (https://umich.zoom.us/j/3892103719)
  - Divya Santhanam, **divyasan@umich.edu** (mailto:divyasan@umich.edu) , Office Hours: Monday/Thursday 2:00-3:30pm in G219 Angell Hall

- Eman Wong, emanwong@umich.edu (mailto:emanwong@umich.edu) , Office Hours: Monday 9:30am-12:30pm in G219 Angell Hall (11/6 OH on Zoom (https://umich.zoom.us/j/8271961504))
- Huiqian Wang, huiqianw@umich.edu (mailto:huiqianw@umich.edu) , Office Hours: Monday 4:00-5:30pm, Tuesday 10:00 - 11:30am, via Zoom (https://umich.zoom.us/j/7876865878)
- Xinyu Liang, lixinyu@umich.edu (mailto:lixinyu@umich.edu) , Office Hours: Tuesday 10:00am-11:30am, Thursday 11:00-12:30 AM, in G219 Angell Hall

Meeting Times

- Lectures (please attend the lecture that you are registered for)
  - Section 100: Monday/Wednesday, 10:30a - 11:20am, Hutchins 100
  - Section 200: Tuesday/Thursday, 8:30am - 9:50am, East Hall 1360
- Labs (please attend the lab for which you are registered; if you need to switch lab sections you must register for the lab new lab session)
  - 101, Tues 1:00-2:30pm, B760 EH, Eman Wong (https://rooms.lsa.umich.edu/classrooms/EHB760)
  - 102, Tues 8:30-10:00am, 2244 USB, Xinyu Liang
  - 103, Tues 10:00-11:30am, 2244 USB, Sang Hoo Kook
  - 201, Wed 10:00-11:30am, B760 EH, Divya Santhanam
  - 202, Wed 8:30-10:00am, 2244 USB, Huiqian Wang
  - 203, Wed 11:30am-1:00pm, 2244USB , Rachel Himmel

Software

Link to Great Lakes (http://greatlakes.arc-ts.umich.edu)

To get started you will need to request an account on Great Lakes, install the UM VPN, and install Microsoft One Note (https://umich.instructure.com/courses/615553/pages/software) .

Materials

- Practical Statistics for Data Scientists (https://search.lib.umich.edu/catalog/record/9918734480160638 1) by Peter Bruce, Andrew Bruce, and Peter Gedeck is our main text book.
- A concise introduction to programming in python by Mark J. Johnson (https://search.lib.umich.edu/catalog/record/9918728587300638 1) covers our basic introduction to programming with Python.
- Additional material related to Python and Pandas can also be be found in Python for Data Analysis, 3rd ed. (https://learning.oreilly.com/library/view/python-for-data/9781098104023/) by Wes McKinney.

Course slides will be made available prior to lecture.

The course aims to teach all Python required (in lecture or in the Python focused textbooks). Additional documentation on Python and the main libraries we will use can be found at:

- **Python Standard Library** (https://docs.python.org/3/library/) . Of particular interest:
  - **Built in functions** (https://docs.python.org/3/library/functions.html)
  - **Lists** (https://docs.python.org/3/library/stdtypes.html#sequence-types-list-tuple-range) , **Dictionaries** (https://docs.python.org/3/library/stdtypes.html#mapping-types-dict)
  - **Math** (https://docs.python.org/3/library/math.html) , **Basic statistical summaries** (https://docs.python.org/3/library/statistics.html)
- **Pandas library** (https://pandas.pydata.org/docs/)
  - **Detailed reference for Series** (https://pandas.pydata.org/docs/reference/series.html)
  - **Detailed reference for DataFrame** (https://pandas.pydata.org/docs/reference/frame.html)
- **Seaborn** (https://seaborn.pydata.org/api.html)
- **Numpy** (https://numpy.org/doc/stable/)
  - **Mathematical functions** (https://numpy.org/doc/stable/reference/routines.math.html)
  - **Logic functions** (https://numpy.org/doc/stable/reference/routines.logic.html)
  - **Sorting and Searching** (https://numpy.org/doc/stable/reference/routines.sort.html)

Course Structure and Grading

The course will be based on the following structure:

- Homework Assignments (40%)
- Quizzes (10%)
- Three Exams (15%, each) (announced at least two weeks in advance)
  - Mid-term Exam 1, TBA
  - Mid-term Exam 2, TBA
  - Final Exam, due 12/13
- Lab Attendance/Completion (5%), two allowed absences

Late homework will not be accepted. Please email the professor (**mfredric@umich.edu** (mailto:mfredric@umich.edu)) if you have circumstances requiring an extension. Please contact the professor as soon as possible to discuss any accommodations for exams.

Working with other students on homework assignments is encouraged! All students should contribute to all work. Student groups should not simply copy and paste solutions from one student's assignment to another. We expect solutions to be similar for students working together, but students should make sure that they understand and can competently explain any solution if asked.

Students are not to collaborate on quizzes or exams. Since quizzes and exams are given electronically and students will be completing them at different times, please do not discuss quizzes and exams with other students until after the due date.

Course Policies

Additional course policies can be found on our **Course Policies Page.** **(https://umich.instructure.com/courses/615553/pages/course-policies)**