

Statistics 507, Fall 2021 (./index.html)

Data Science in Python

Syllabus

Contact Information

- Instructor: James Henderson, PhD (jbhender@umich.edu (mailto:jbhender@umich.edu))
- Instructor office hours: Tuesdays 4-6pm (Zoom until further notice).
- GSIs: Cheng Ma (chengmc@umich.edu (mailto:chengmc@umich.edu)), Jing Ouyang (jingoy@umich.edu (mailto:jingoy@umich.edu))
- GSI office hours:
 - Jing - Thursdays 2-5pm (Virtual)
 - Cheng - Mondays 4:30-7:30pm (Virtual).
- See Canvas or Google Calendar for Zoom Links.
- Course website: jbhender.github.io/Stats507/F21/ (<https://jbhender.github.io/Stats507/F21/>)
- Course repository: www.github.com/jbhender/Stats507_F21/ (https://www.github.com/jbhender/Stats507_F21/)

Meetings

This is a lecture format class. We meet Tuesdays and Thursday from 8:30-10am in 2548 NUB (North University Building). We will end by 9:50am.

Key Dates

- No Class:
 - Tu Oct 19 (Fall Study Break)
 - Th Nov 25 (Holiday)
- Final Class: Th Dec 9
- Office of the Registrar (<https://ro.umich.edu/>)
- Anticipated due dates of problem sets:
 - PS1 - September 17
 - PS2 - September 24
 - PS3 - October 1
 - PS4 - October 8
 - PS5 - October 15
 - PS6 - October 29
 - PS7 - November 5
 - PS8 - November 12
 - PS9 - November 19
 - PS10 - December 10

If assignment due dates interfere with religious holidays you observe please let me know prior to October 1 to work out alternative arrangements.

Grading Information

Your grade will be based on 10 problem sets worth 100 points each for 1,000 total points, so that 1 point = 0.1% of the course grade. The course will use the following grading scale:

- A = [1000, 920] points,
- A- = (920, 880] points,
- B+ = (880, 840] points,
- B = (840, 800] points,
- B- = (800, 760] points.

Totals less than 760 points will receive a grade on a case-by-case basis. I reserve the right to modify this scale and other grading policies, but grade thresholds will be no higher than shown here.

Text Books

There are no required texts for this course, but there are required readings from texts and other materials available online or through the library. There are three recommended texts:

- Think Python 2 (<https://greenteapress.com/wp/think-python-2e/>), Allen B. Downey
- Python for Everybody (<https://www.py4e.com/book>), Charles R. Severance
- Python for Data Analysis (<https://www.oreilly.com/library/view/python-for-data/9781449323592/>), Wes McKinney. (Accessible online to UM students.) (<https://search.lib.umich.edu/databases/record/10263>)

Pre-requisites

You should have taken or be currently enrolled in an intermediate applied statistics course such as Stats 500 or equivalent. You should be comfortable programming in at least one programming language or scripting in a statistical software language.

Course Description

The primary goals of Stats 507 are to help you: learn Python from the ground up, become proficient in using Python for data analysis, gain fluency in other data science tools (e.g. git, Linux, SQL), and to make you comfortable reading documentation and independently learning new computing skills.

1. Students will gain fluency with frequently used programming patterns, concepts, and libraries used by data scientists and statistical analysts working in Python.
2. Students will develop good coding, documentation, and work flow habits.
3. Students will improve and gain confidence as independent learners capable of learning new software by reading documentation and working examples.
4. Students will demonstrate growth in presenting and communicating with and about data and data analyses.

The topics below represent an approximate scope for the course. **I reserve the right to modify this scope.** Materials related to each topic will include lectures and readings and, less frequently, examples and/or case studies.

- *Introduction to Python* - the Python ecosystem, programming patterns, data structures, I/O, notebooks, classes and objects, functional programming.
- *Python's core scientific modules* - NumPy, Pandas, and (some) SciPy.
- *Communicating with and about data* - markdown, Latex, minimal HTML, writing “polished” and professional reports.
- *Linux shell skills and git* - using git, the Linux file system, data processing in the Linux shell, ssh.
- *Visualization* - matplotlib, plotnine, and (possibly) seaborn.
- *Statistical models* - linear and generalized linear models in `statsmodels`.
- *Machine learning* - core ML models using `scikit-learn`, specifying, training, and evaluating models, neural network models using `keras` and `tensorflow`.
- *Big data and distributed processing* - parallel and asynchronous computing, SQL, mapReduce, Spark.

Computing Resources

All of the necessary software in this course is freely available (most is also open source) and can be installed on your personal computer.

You should install a Python 3 distribution and necessary modules for using Jupyter notebooks. I recommend installing Python 3.9. You will also need an ssh client, such as Terminal (Mac, *nix), Putty (<https://www.putty.org/>) (Windows) or git for windows (<https://gitforwindows.org/>) (Windows).

If you don't already have one, you should request a Great Lakes (<https://arc.umich.edu/greatlakes/>) user account.

Owing to the large size of this class, neither I nor the GSIs are here to provide tech support in terms of installation. While we may offer generic advice, please work with your peers and an internet search engine to solve installation or other technology issues.

Submitting work late

You are expected to turn in all problem sets by the due dates listed on Canvas. However, to accommodate unexpected circumstances you may utilize “late days” when turning in problem sets. Late days work as follows:

- you have 7 total late days for the semester,
- one late day grants a 24 hour extension for the problem set,
- there is no penalty for using late days,
- you must state how many late days you intend to use prior to the initial due date for a problem set,
- you may not request additional late days after the original due date,

- to use late days, upload a draft solution to Canvas and add a comment saying how many late days you wish to use,
- submit the problem set to be graded by the self-extended due date.

There is no exam for the course, however, late days may not be used to extend any assignment due date past the time of the course's scheduled exam time of Monday December 13 by 3:30pm.

Course Recordings

Course lectures, including slides and classroom audio, may be recorded and made available to students in this course through YouTube, Canvas, or Dropbox. **As part of your participation in this course, you may be recorded.** If you do not wish to be recorded, please contact the instructor by Tuesday September 7 or as soon as you enroll in the course (whichever is latest) to discuss alternative arrangements.

Students are prohibited from distributing recordings of any class activity without prior written permission from the instructor, except as necessary as part of approved accommodations for students with disabilities. Any approved recordings may only be used for the student's own private use.

I intend to provide recordings to all students as a courtesy to students who have requested them. While I will invest reasonable effort to provide recordings, I do not guarantee the quality of the recordings nor that all recordings will be available. Finally, I reserve the right to alter this recording policy at any point.

Accommodations for students with disabilities

If you think you need an accommodation for a disability, please let me know at your earliest convenience. Some aspects of this course, such as assignments, in-class activities, and recorded lectures may be modified to facilitate your participation and progress. As soon as you make me aware of your needs, we can work with the Services for Students with Disabilities (SSD) office to help us determine appropriate academic accommodations. SSD (734-763-3000; <http://ssd.umich.edu>) typically recommends accommodations through a Verified Individualized Services and Accommodations (VISA) form. Any information you provide is private and confidential and will be treated as such.

Academic Integrity

Unless specifically stated otherwise, students are expected to complete problem sets and any other assignments independently without copying code or text from other students in this course. Students are encouraged to discuss assignments and to help one another with concepts and syntax, but such **discussions should not reach the point of representing one student's work as another's.**

Use of materials from other sources should fall within the license for those materials and include proper attribution.

For more on academic integrity, please be advised of Rackham policy:

<http://www.rackham.umich.edu/academic-policies/section8/>

(<http://www.rackham.umich.edu/academic-policies/section8/>).

Mental Health and Wellbeing

The University of Michigan is committed to advancing the mental health and wellbeing of its students. If you or someone you know is feeling overwhelmed, depressed, and/or in need of support, services are available. For help, contact Counseling and Psychological Services (CAPS) at 734.764.8312 and caps.umich.edu during and after hours, on weekends and holidays, or through its counselors physically located in schools on both North and Central Campus. You may also consult University Health Service (UHS) at 734.764.8320 and <https://www.uhs.umich.edu/mentalhealthsvcs> (<https://www.uhs.umich.edu/mentalhealthsvcs>), or for alcohol or drug concerns, see <https://www.uhs.umich.edu/aodresources> (<https://www.uhs.umich.edu/aodresources>).

For a listing of other mental health resources available on and off campus, visit <https://uhs.umich.edu/stressresources> (<https://uhs.umich.edu/stressresources>).

Mandatory Reporting and Sexual Misconduct

Title IX prohibits sex discrimination to include sexual misconduct: harassment, domestic and dating violence, sexual assault, and stalking. If you or someone you know has been harassed or assaulted, you can receive confidential support and academic advocacy at the Sexual Assault Prevention and Awareness Center (SAPAC). SAPAC can be contacted on their 24-hour crisis line, 734-936-3333 and online at <https://sapac.umich.edu> (<https://sapac.umich.edu>). Alleged violations can be reported non-confidentially to the Office for Institutional Equity (OIE) at institutional.equity@umich.edu (<mailto:institutional.equity@umich.edu>). Reports to law enforcement can be made to University of Michigan Police Department at 734-763-3434.

As an instructor, one of my responsibilities is to help create a safe learning environment on our campus. I also have a mandatory reporting responsibility. I am required to share information regarding sexual misconduct or information about a crime that may have occurred on U-M's campus with the University. Students may speak to someone confidentially by contacting SAPAC's Crisis Line at (734) 936-3333.