Stats 131 Course Syllabus

Python and Other Technologies for Data Analysis

Miles Chen, PhD

2019 Spring Quarter

Course Summary

This class covers the use of Python and other technologies for data analysis and data science. The course will focus on programming with Python and a selection of its libraries: NumPy, pandas, matplotlib, and scikit-learn, for the purpose of data processing, data cleaning, data analysis, and machine learning. Other technologies covered will include Jupyter notebook, SQL, and Git.

The goal of the course is to provide students hands-on programming experience with data technologies that are used in industry. Students will be required to complete a group project with real data and to present their analysis and findings.

Students who have taken Stats 21 should not take Stats 131, and students who take Stats 131 should not take Stats 21.

Learning Outcomes

Upon completion of this course, students

- will have created Jupyter Notebooks and will have published them to a Github repository
- will be able to read and write intermediate code in Python
- will have imported, manipulated, and summarized data in Python for exploratory data analysis
- will have created data visualizations in Python
- will have fit and evaluated a simple linear regression model with scikit-learn

Prerequisites:

Stats 102A

Tentative Outline of topics:

1) Week 1: Git, GitHub, Jupyter, Python basics
2) Week 2: Python: Data Types, Dictionaries, Loops
3) Week 3: Python: Functions/Lambdas, List Comprehensions
4) Week 4: Numpy
5) Week 5: Pandas: Importing, Reshaping, and Cleaning data
6) Week 6: Pandas: Wrangling, and Aggregation
7) Week 7: Data Visualization with matplotlib
8) Week 8: SQL
9) Week 9: scikit-learn
10) Week 10: final presentations

Office Hours

Wednesdays 2-2:50 PM, and by appointment

Math Sciences 8105 J

No office hours during finals week.

Course website

All course materials, announcements, and assignments will be posted via CCLE. http://ccle.ucla.edu

Grading

- 20% Homework Assignments
- 30% DataCamp Homework
- 20% Midterm Exam
- 30% Final Project
- 2% Extra credit for attendance
- 1% Extra credit for achieving red-bird status on Campuswire

Letter Grade cutoffs

Letter grades are assigned on a straight scale as follows:

- 59.9 and below: F
- 60.0 - 69.9: D
- 70.0 - 76.9: C, 77.0 - 79.9: C+
- 80.0 - 82.9: B-, 83.0 - 86.9: B, b7.0 - 89.9: B+
- 90.0 - 92.9: A-, 93.0 and up A, top 5% of students: A+

I do not curve grades. I do not assign C-, D+ or D-. The top 5% of students will get an A+. An A+ is the only grade assignment that is assigned based on performance relative to peers.

If you are a Stats major or minor it is important that you take the class for letter grade. Taking the grade as pass/no-pass will not fulfill the requirements for completing the major or minor.

Midterm Exam Date

The midterm exam is a take-home exam.

- Midterm: 7th Week Friday May 17, 2018 **3 - 5PM**

Please note that the midterm is two hours long. Please make sure you are available for the full two hours and have access to a private area where you will not be bothered for the duration of the exam. If you know ahead of time that you are not available during this time, let me know right away.

Software

In this course, we use Anaconda Python, a popular data science platform. Anaconda can be downloaded and installed for free from: https://www.anaconda.com/download/

Textbooks

Students must be connected to the UCLA network to access the textbooks for free. Students who would like to access the textbook off-campus may do so by connecting the the UCLA network via VPN. https://www.it.ucla.edu/bol/services/virtual-private-network-vpn-clients

Python for Data Analysis, 2nd Edition

- By: William Wesley McKinney
- Publisher: O'Reilly Media, Inc.
- Pub. Date: October 10, 2017
- Print ISBN-13: 978-1-4919-5766-0
- URL for free access: http://proquest.safaribooksonline.com/book/programming/python/9781491957653

Python Data Science Handbook

- By: Jake VanderPlas
- Publisher: O'Reilly Media, Inc.
- Pub. Date: November 30, 2016
- Print ISBN-13: 978-1-4919-1205-8
- URL for free access: http://proquest.safaribooksonline.com/book/programming/python/9781491912126

DataCamp Homework

Homework will be assigned via DataCamp.

Students are required to sign up for their free subscription to DataCamp. Students must make sure their DataCamp account has the first and last name as it appears in CCLE so that we can transfer the scores from DataCamp to CCLE.

DataCamp only has three possible statuses for homework assignments:

- completed on time
- completed late
- not completed

We will record grades for the homework assignment 7 days after its due date. Assignments completed on time will receive 100 points. Assignments completed late will receive 70 points. Assignments not completed by the time we record the score in the gradebook will receive 0 points.

Please make sure you complete the assignment in DataCamp well ahead of its deadline to ensure you receive full credit for the assignment.

Unfortunately, the rigidity of the DataCamp assignment interface does not allow me to grant extensions.

Homework

The number of homework assignments is still TBD. Each homework assignment is worth the same amount. Each homework assignment will be scored out of a total of 100 points. None of the homework scores will be dropped. It is your responsibility to verify that your homework assignment successfully uploaded by the deadline.

All homework assignments will be posted as an ipynb file.

You will download the ipynb file and modify it with your answers.

After entering all of your answers, select "Kernel" > "Restart and Run All". Check to make sure all requested answers are clearly visible.

Then choose "File" > "Download As" > "HTML"

Then students will submit the HTML file to CCLE. Do not submit the ipynb file.

Avoiding grading issues with the homework

The graders have to go through many students' submissions. The grader will not spend more than 5-10 minutes on your submission. The grader does not have the luxury of trying to run your code or debug your code, or even see what your individual lines of code can do.

The grader will look for the printed results, so it is **very important** you properly print the requested results in your document. It is your job to make sure that you get the credit you deserve.

Late Policy for Homework

There is a 10 minute grace period. Submissions up to 10 minutes and 59 seconds late will be accepted with no penalty.

Homework assignments submitted 11 minutes late or more will be accepted with penalty. There is a deduction of 5 points for being less than one hour late. An additional 5-point deduction will be taken for each additional hour it is late.

Deductions are suspended between the hours of 12:00 AM and 5:59 PM the following day. The penalties resume at 6 PM.

The maximum late penalty is 60 points.

So, if an assignment is due at 6pm, a homework assignment submitted on CCLE at

- 6:10:59 pm will receive no penalty.
- 6:11:00 pm will receive 5-point penalty.
- 6:59:59 pm will receive 5-point penalty.
- 7:00:00 pm will receive 10-point penalty.
- 7:59:59 pm will receive 10-point penalty.
- 8:00:00 pm will receive 15-point penalty.
- 11:59:59 pm will receive 30-point penalty.
- 12:00:00 am (following day) will receive 30-point penalty.
- 5:59:59 pm (following day) will receive 30-point penalty.
- 6:00:00 pm (following day) will receive 35-point penalty.
- 10:59:59 pm (following day) will receive 55-point penalty.
- 11:00:00 pm (following day) will receive 60-point penalty.
- Anything later will be accepted with a 60-point penalty.

I strongly recommend uploading the homework to CCLE well in advance of the deadline in case there are connectivity problems or server issues.

After 9 calendar days, late homeworks will not be accepted for any reason.

Homework extensions

I will grant a 72-hour reprieve from the late penalty for homework submissions if the homework is accompanied with documentation.

The documentation must be uploaded onto CCLE with the homework submission.

CCLE accepts multiple file submissions for homework assignments. Submit your homework assignment with a PDF or photograph of the documentation showing the **date** and reason for your inability to submit the homework on time.

Do not email the professor or your TA with your documentation.

Extensions beyond 72 hours will be granted in only the most extreme circumstances, and will only be granted in person by the professor.

Homework Regrades

Homework assignments are graded by designated graders (graduate students in the Statistics Department). Occassionally a grader may make a mistake in grading your submission.

Both the TA and the professor have the authority to change the grade of a homework assignment.

All requests for regrading a homework assignment must be done **in person** during office hours.

Do not email the TA or the professor for a request to regrade an assignment.

Please only come to office hours if you are certain that a mistake was made in the grading of your homework.

Final Project

Students will complete final projects in groups of four students.

Groups will be required to identify a data set and perform some exploratory data analysis. Groups will then identify research questions, and appropriate models. Groups will then perform basic statistical modeling.

Groups will submit a report of the analysis in a reproducible Jupyter Notebook published on GitHub.

Groups will schedule a 40-minute final project presentation with the professor during 10th week.

A detailed grading rubric for the final project will be released during the course.

Make up policy for Exam

We will have one exam in Week 7.

If you miss this exam, you will need to provide documentation for your reason for missing the exam. Make-ups of the exam must be taken within 7 days of the actual exam, or a zero will be recorded. (The exam must be made up by week 8 Friday by 5pm.)

Attendance

Attendance is extra credit.

Attendance is generally taken using an online tool, but I may do an actual roll call.

I may even choose to use the online polling tool, followed by an actual roll call. If a student responded as being present via the online tool, but is not present for the roll call (never came to class or left class early), the student will forfeit the entire attendance portion of the course grade.

Campuswire Participation

A Campuswire discussion forum has been created for the class. Students will be graded on their reputation level and/or repuation points as determined by Campuswire.

Campuswire awards reputations points as follows:

- 2pt for each question asked
- 1pt for each like you receive on a question you asked
- 5pt for each question you answer
- 5pt for each upvote you receive on an answer you provided

Campuswire has tiers of reputation levels as follows:

- Level 1: Starter (yellow bird). To level-up to Starter:
    - Answer 1 question on the Class Feed AND
    - Receive 1 upvote from a classmate
- Level 2: Intermediate (red bird). To level-up to Intermediate:
    - Answer 5 questions on the Class Feed AND
    - Receive 10 upvotes from classmates
- Level 3: Advanced (eagle). To level-up to Advanced:
    - Answer 20 questions the Class Feed AND
    - Receive 50 upvotes from classmates

Campuswire Grading:

Participation on Campuswire can contribute 1 extra credit point towards your final grade. I will download Campuswire's reputation report on Week 10 Friday and will enter grades as follows:

- Students who achieve starter level AND earn 50 reputation points will get 0.5 point of extra credit.
- Students who achieve intermediate level (red bird) OR 100 reputation points will get 1 point of extra credit.

Academic Integrity

As a student and member of the University community, you are here to get an education and are, therefore, expected to demonstrate integrity in your academic endeavors. All students must uphold University of California Standards of Student Conduct as administered by the Office of the Dean of Students. Students are subject to disciplinary action for several types of misconduct, including but not limited to: cheating, multiple submissions, plagiarism, prohibited collaboration, facilitating academic dishonesty, or knowingly furnishing false information.

In addition, each student is the sole owner of their own code and work and must NOT:

- Submit work that is not original.
- Publish course material (including homework solutions and exams) online.
- Submit someone else's work, or a modification of that work, with or without that person's knowledge.
- Allow someone else to submit their work, or a modification of their work.
- Contract course work out to others.
- Post the course questions on forums other than the designated course Piazza. This means students cannot post questions on places like stack overflow or other similar places.
- Plan or execute with another student some form of cheating during an exam.
- Make use of unauthorized material during an exam.

For more information about academic integrity and student conduct code, please go to http://www.deanofstudents.ucla.edu/Student-Conduct

Gender Discrimination, Sexual Violence, and Harassment

Title IX prohibits gender discrimination, including sexual harassment, domestic and dating violence, sexual assault, and stalking. http://www.sexualharassment.ucla.edu/

http://www.sexualviolence.ucla.edu/

> The University of California is committed to creating and maintaining a community dedicated to the advancement, application and transmission of knowledge and creative endeavors through academic excellence, where all individuals who participate in University programs and activities can work and learn together in an atmosphere free of harassment, exploitation, or intimidation.

If you, or someone you know, have been a victim of sexual harassment or sexual violence, please seek help.

Any type of harassment of any other student will not be tolerated. Please be aware that your TA and Lecturer are legally compelled to report any information in regards to sexual harassment or sexual violence to the Title IX officer.

Confidential support and advocacy is available for students who have experienced sexual harassment or sexual violence at the CARE Advocacy Office for Sexual and Gender-Based Violence, John Wooden Center West 1st Floor, (310) 206-2465, CAREadvocate@careprogram.ucla.edu. Sexual violence or sexual harassment can also be reported directly to the University's Title IX Coordinator, 2241 Murphy Hall, titleix@conet.ucla.edu, (310) 206-3417.

Reach a confidential UCLA advocate, 24 hours a day, 7 days a week by calling the CAPS 24/7 Crisis Hotline (310) 825-0768

Mental Health and Student Well-Being

> Counseling and Psychological Services (CAPS) is a multidisciplinary student mental health center for the UCLA campus. CAPS supports the academic and student development missions of the University and the Division of Student Affairs. We're here to support your mental health needs as you pursue your academic goals. Our services are designed to foster the development of healthy well-being necessary for success in a complex global environment.

http://www.counseling.ucla.edu

Attending and completing university coursework is difficult and can become burdensome both physically and psychologically. Take care of your mental well-being and seek support from some of the many excellent resources offered at CAPS.

Accommodations for students with disabilities

Students needing academic accommodations based on a disability must contact the Center for Accessible Education (CAE) at (310) 825-1501 or present in person at Murphy Hall A255. As the professionals delegated authority from the campus to determine reasonable disability accommodations, CAE will assess all requested accommodations and communicate appropriately with faculty. In the event that a student has approval for proctoring arrangements during exams, please inform your respective professors and/or Teaching Assistant before date of exams. When possible, students should contact the CAE within the first two weeks of the term as reasonable notice is needed to coordinate accommodations.

For more information visit http://www.cae.ucla.edu

Course Grade Changes

After course grades have been submitted to the registrar, grades are final. Grade changes will only be considered if there has been a clerical or procedural mistake. Students have one quarter to make requests for a grade change. Graded exams and other materials will be kept for one quarter. After one quarter, course grade changes will not be made.

Undergraduate Mentoring

I am a mentor for undergraduates for the UCLA Department of Statistics. This means that you may visit me during office hours (or by appointment) for an informal meeting where you can ask questions unrelated to course content, such as research opportunities, graduate studies, career paths, or any other topic pertinent to your education. You are welcome to visit and meet even after the course ends.

Requests for Letters of Recommendation

I am generally only willing to write letters of recommendation for students who I know and recognize. If you think you might request a letter of recommendation in the future, please make an effort to make yourself known to me. I am more likely to recognize students who sit mear the front of class and who actively participate. I am also more likely to recognize students who make several visits to office hours during the quarter.

For a full guideline, please see this page on my personal website: https://smileschen.com/lor.html