

ECE 20875: Python for Data Science

Fall 2020
Course Syllabus

Due to the complications surrounding COVID-19, to accommodate everyone to the largest extent possible, this course will be offered in the **flipped classroom** format. This means that lectures will take place online outside of the set class time, and the class time will instead be used as an (optional) recitation period led by the instructors. There will also be separate office hours offered online each day by the teaching assistants.

Instructors

Prof. Chris Brinton (cgb@purdue.edu)
Office: MSEE 342

Prof. Qiang Qiu (qqiu@purdue.edu)
Office: MSEE 358

Graduate TAs

Somosmita Mitra [Lead] (mitra26@purdue.edu)
Jiaqui Guo (guo498@purdue.edu)
Serena Nicoll (snicoll@purdue.edu)

Undergraduate TAs

Julia Taylor [Lead] (taylo773@purdue.edu)
Aagam Shah (shah297@purdue.edu)
Aathavan Thevasenapathy (bthevase@purdue.edu)
Adam Popper (apopper@purdue.edu)
Erik Wilson (wilso969@purdue.edu)
Ethan Glasser (glasere@purdue.edu)
Harsh Ajwani (hajwani@purdue.edu)
Jhen Ruei Chen (chen2774@purdue.edu)
Joseph Robert Bushagour (jbushago@purdue.edu)
Kevin Chen (chen2776@purdue.edu)
Kevin Kwon (kwon128@purdue.edu)
Kyle Wolf (wolf103@purdue.edu)
Lohith Roy Chittineni (lchittin@purdue.edu)
Marvin H Mui (mui1@purdue.edu)
Minjun Zhang (zhan3624@purdue.edu)

Rufat Imanov (rimanov@purdue.edu)
Runjia Shen (shen403@purdue.edu)
Shan Huang (huan1372@purdue.edu)
Sneha Mahapatra (mahapat0@purdue.edu)
Vikram Srivastava (srivas65@purdue.edu)

Sections

There are three sections of the course: 001, 002, and 003 (distance). All three sections of the course will be managed very similarly. They will have the same homeworks, the same discussion forum, the same announcements, and, as much as possible, the same lectures.

The exception to this is the in-person recitation period component, where the instructor assigned to each section will manage their corresponding recitation, except for 003 which does not have an assigned in-person recitation.

Communication

We will use Piazza as the primary form of announcements and communication between the teaching staff and the students. **If you have any questions about the course, homeworks, or projects, you should post them on Piazza.** It's a shared discussion forum, where your question can be answered by the instructors, the TAs or your fellow students!

Find our Piazza page here:

piazza.com/purdue/fall2020/ece20875/home

Class Website

To maintain consistency across the sections, we will manage a separate website for the course. This website will have links to the course material, recorded lectures, assignments, practice exams, and other information.

Find our course website here:

<http://www.cbrinton.net/ECE20875-2020-Fall.html>

Brightspace will be used *solely* to enter and track your grades. All other matters for the course will be administered through Piazza and the website. The Brightspace page is available to you through Purdue's website.

Lectures

Lectures will be live-streamed via Zoom two nights a week: **Monday and Wednesday, 7:00-8:15pm ET**. This timeframe is chosen to not interfere with anyone's daily

schedules. Either way, the live stream will also be recorded and put on YouTube, so if you cannot attend, you are welcome to watch the lecture on your own time.

Why attend the live stream if the videos are recorded?

- It gives you an opportunity to ask questions in real time. You will be able to virtually “raise your hand” in Zoom, and the instructor will be able to unmute you so you can ask your question live. You can also use the chat window to ask questions.
- If not enough people are attending, the instructor reserves the right to stop offering the live stream, and instead may just post the videos directly as recordings. In this case, the recordings will not have the added benefit of live Q&A embedded in them.
- When the instructor writes on the board behind him, he will have to remember to stop sharing screen in order for it to be captured on the recording. It is possible, though not likely, that there will be times where he forgets, and in this case only students participating will be able to see the writing on the board. That being said, the lecture slides are pretty self-contained.

To emphasize, however, attending the live stream is **optional**. Attendance will not be required, expected, or rewarded in any way this semester.

Here is the link to access the live lecture streams:

<https://purdue-edu.zoom.us/j/95196547573?pwd=MIpnUFpRRmFpa3dXOWFwZm90VG5HUT09>

Here is the link to access the recorded lectures:

https://www.youtube.com/playlist?list=PLCXMxrU-4FxF_5ELcsFFX01-RwUsvjNOp

Classroom

The recitation periods will take place during the class times listed for the course. They will be broadcasted live via Zoom for everyone, but **only those students enrolled in 001 and 002 are permitted to attend them in person.**

Moreover, to attend the recitations in person, **you must follow the groupings we have set up** to adhere to social distancing guidelines. **Pay careful attention to which section and which group you have been assigned to.** Each group is assigned to one live recitation a week, according to the schedule below:

Section 001

- Instructor: Prof. Chris Brinton
- Room: EE 170
- Schedule
 - Tuesday 10:30-11:45 (Group 1 of 001 permitted to attend in person)
 - Thursday 10:30-11:45 (Group 2 of 001 permitted to attend in person)
- Zoom link (for anyone, including those in 003, to participate remotely): <https://purdue-edu.zoom.us/j/92753891103?pwd=Qnd5bEFqSWWhBaXF1WkVnalRRROW5uZz09>

Section 002

- Instructor: Prof. Qiang Qiu
- Room: WTHR 320
- Schedule
 - Monday 11:30-12:20 (Group 1 of 002 permitted to attend in person)
 - Wednesday 11:30-12:20 (Group 2 of 002 permitted to attend in person)
 - Friday 11:30-12:20 (Group 3 of 002 permitted to attend in person)
- Zoom link (for anyone, including those in 003, to participate remotely): <https://purdue-edu.zoom.us/j/97582249113?pwd=NVVBSmZmYXE2QkMvQmQzdmM5SURidz09>

Section 003

- Virtual, no in-person component
- Feel free to join the Zoom links above during the allotted times

The instructors will treat these sessions as their office hours. In order to promote a setting in which students are comfortable asking their questions, these recitations will not be recorded. Also, if no students are present, the instructors reserve the right to leave.

Classroom Guidance Regarding Protect Purdue

The Protect Purdue Plan, which includes the Protect Purdue Pledge, is campus policy and as such all members of the Purdue community must comply with the required health and safety guidelines. Required behaviors in this class include: staying home and contacting the Protect Purdue Health Center (496-INFO) if you feel ill or know you have been exposed to the virus, properly wearing a mask in classrooms and campus building, at all times (e.g., mask covers nose and mouth, no eating/drinking in the classroom), disinfecting desk/workspace prior to and after use, maintaining appropriate social distancing with peers and instructors (including when entering/exiting classrooms), refraining from moving furniture, avoiding shared use of personal items, maintaining robust hygiene (e.g., handwashing, disposal of tissues) prior to, during and after class, and following all safety directions from the instructor.

Students who are not engaging in these behaviors (e.g., wearing a mask) will be offered the opportunity to comply. If non-compliance continues, possible results include instructors asking the student to leave class and instructors dismissing the whole class. Students who do not comply with the required health behaviors are violating the University Code of Conduct and will be reported to the Dean of Students Office with sanctions ranging from educational requirements to dismissal from the university.

Any student who has substantial reason to believe that another person in a campus room (e.g., classroom) is threatening the safety of others by not complying (e.g., not wearing a mask) may leave the room without consequence. The student is encouraged to report the behavior to and discuss next steps with their instructor. Students also have

the option of reporting the behavior to the Office of the Student Rights and Responsibilities. See also Purdue University Bill of Student Rights.

If you feel ill, have any symptoms associated with COVID-19, or suspect you have been exposed to the virus, you should stay home and contact the Protect Purdue Health Center (496-INFO). In the current context of COVID-19, in-person attendance will not be a factor in the final grades, but the student still needs to inform the instructor of any conflict that can be anticipated and will affect the submission of an assignment or the ability to take an exam. When conflicts can be anticipated, such as for many University-sponsored activities and religious observations, the student should inform the instructor of the situation as far in advance as possible. For unanticipated or emergency conflict, when advance notification to an instructor is not possible, the student should contact the instructor as soon as possible by email. When the student is unable to make direct contact with the instructor and is unable to leave word with the instructor's department because of circumstances beyond the student's control, and in cases of bereavement, quarantine, or isolation, the student or the student's representative should contact the Office of the Dean of Students via email or phone at 765-494-1747. Our course Brightspace includes a link on Attendance and Grief Absence policies under the University Policies menu.

Office Hours

We are fortunate to have many wonderful graduate and undergraduate TAs on the teaching staff this semester. Many of the TAs have taken the class before and/or previously served as a TA. With their help, we are able to offer the following office hours schedule:

- "Domestic" [many more TAs available here, please give priority to attending these]
 - 1-2pm ET and 5-8pm ET
 - Every day Monday through Friday
- "International" [few TAs here, please try to reserve these for international students]
 - 1-2am ET Monday through Friday
 - 8-10am ET Monday through Wednesday
 - 12-1pm ET on Friday

Please take advantage of these sessions to get one-on-one assistance, particularly for help on the homework assignments.

Due to COVID-19, the university has recommended that we conduct our office hours online. **As a result, office hours will be held using the Discord platform.** The protocol for using Discord is as follows:

- During lab hours, each TA will join one of the open voice channels in Discord. Students who have questions can join these voice channels to talk with a TA one-on-one and share screens if needed.
- After discussion, the student should leave the TA voice channel and go back to the general channel to allow other students to talk to the available TAs.

Here is the link to the Discord channel:

<https://discord.gg/t9DQxey>

Lectures vs. Recitations vs. Office Hours

We will expect that you have been watching the lectures regularly and reviewing the notes before asking questions in recitation and office hours, i.e., these venues must be used as a supplement, not an alternate, to the lectures. During the recitations, we will give preference to conceptual or clarification questions over debugging questions because debugging questions could be answered during the daily TA lab hours.

Email

Technical questions should be posted to Piazza, raised during the recitation period (primarily questions on the lecture), or raised during office hours (primarily questions on the homeworks/project). This allows other students who might have similar questions to benefit from the answers. As a result, the professor and TAs will not typically answer programming questions via email. Of course, if you have questions of a personal or confidential nature, we welcome your email.

Course Outcomes

A student who successfully fulfills the course requirements will have demonstrated:

1. An understanding of regular expressions [1, 2]
2. An ability to use Python to write data analyses [1, 2, 6]
3. An ability to explain when particular data analyses are appropriate [1, 2, 3, 6]
4. An ability to explain the results of data analyses [2, 3, 5]
5. An ability to incorporate classes in their Python programs [1, 2, 6]
6. An ability to incorporate associative arrays in their programs [1, 2, 6]
7. An ability to work with a partner to choose appropriate analyses to solve a problem, perform those analyses, and interpret the results of those analyses [1, 2, 3, 5, 6]

These outcomes are extremely high level. In more detail, after taking this course you will be able to:

- Write programs in Python that incorporate: basic control structures; functions; data structures such as lists, tuples and associative arrays; classes and objects; iterators and generators; higher order functions; and regular expressions.
- Perform basic scripting tasks in Bash and Python (e.g., redirecting the output of one program to the input of another, setting and reading environment variables).
- Understand the basics of sampling, estimation, and hypothesis testing.
- Write data analyses that perform: visualization; textual analysis; regression; classification; and clustering.
- Perform more complex analyses using approaches like neural networks.

Course assessment: The achievement of the course objectives will be through programming assignments (covering outcomes 2, 4, 5, and 6), exams (covering outcomes 1, 3, and 4), and a project (covering outcome 7).

Course Grading

Grades will be assigned as follows:

- 50% — Three exams (all equally weighted)
- 40% — 10-12 programming assignments (all equally weighted)
- 10% — Mini-project

Due to COVID-19, unlike previous semesters, **class participation and attendance is completely optional**. However, the students who are the most active (and helpful) in answering questions on Piazza may receive bonus points.

Programming Assignments

Programming assignments will be due every 1-2 weeks, except for the weeks of the exams and during the final two weeks (which is set aside for the mini-project). They will test the concepts covered in class, both programming and statistical. The following rules apply to *all* programming assignments:

1. All programs should run correctly in the versions of Python available on the Scholar cluster. You will have a Scholar account created for you at the beginning of the semester.
2. Some assignments will include a separate writeup as well as a code submission that will produce the required output when run.
3. Unless otherwise specified, assignments are due at 11:59 PM on the deadline.

Programming assignments will be submitted via GitHub Classroom (<https://classroom.github.com>). As such, *you are required to have a GitHub account*. These can be obtained for free at <https://github.com>.

Please fill out the following form **as soon as the semester starts** to provide your GitHub account information:

<https://docs.google.com/forms/d/e/1FAIpQLSdWehPu1zesbzYubdwUecsXaXEfGelQuYzUI29DG0LiNQwa1g/viewform>

Without your GitHub username, we cannot put your grades on Brightspace!

Assignment late submission policy: Except for medical and family emergencies (accompanied by verification), there will be *no individual extensions* granted for programming assignments or the mini-project. Late submissions will be scaled according to lateness, docking 10% from your score per day late, up to a maximum of 50%. Submissions more than 5 days late will be assigned a score of 0.

Exams

The first two exams will take place roughly during the 6th and 11th weeks of the semester, and the third exam will take place during finals period. Exams are equally weighted and are not cumulative.

Due to COVID-19, exams will be administered online. To accommodate multiple timezones, each exam will have a window of 24 hrs in which you can take it. **During this 24 hr period, it is absolutely no contact with anyone in the course, including the teaching staff, is permitted.** Once you open the exam, you will have 1-2 hrs to complete it.

Missing an exam will result in a grade of 0 for that exam. Exam reschedules will only be granted if you demonstrate that it is absolutely necessary (accompanied by verification) and obtain written approval from one of the instructors in advance of the exam.

We will be using Gradescope (www.gradescope.com) to administer the exams. Once you open the exam, you will print it out, and then scan it back to Gradescope (using Gradescope's app) when you are finished. Thus, **you must have access to a printer and a cell phone.**

Here is the link to our course's Gradescope page:
<https://www.gradescope.com/courses/166890>

Regrade Requests

If you believe that we have made an error in grading, you may submit a regrade request with an explanation through the graduate TA who is designated for that homework. **All such requests must be made within one week of when the grade was posted.** Please note that during the regrading process, we reserve the right to regrade any portion of your submission.

Course Schedule

Below is a rough schedule of when and what will be covered in class:

Week	Topic(s) covered
1 8/24-8/30	Introduction Python basics (variables, control structures, functions) Git basics
2 8/31-9/6	Python basics (continued) Python data structures (lists, tuples, associative arrays) Histograms

Week	Topic(s) covered
3 9/7-9/13	Histograms (continued) Probability distributions Random variables
4 9/14-9/20	Filters Map/reduce Higher-order functions
5 9/21-9/27	Sampling and estimation Hypothesis testing Confidence intervals
6 9/28-10/4	Hypothesis testing (continued) Confidence intervals (continued) Exam 1: Wednesday September 30th
7 10/5-10/11	Regular expressions Basic text processing Bash scripting
8 10/12-10/18	Introduction to supervised learning Matrix algebra Regression and least squares
9 10/19-10/25	Regression (continued) Regularization Cross validation
10 10/26-11/1	Regression (continued) n-grams Natural language processing
11 11/2-11/8	Objects and Classes Introduction to unsupervised learning Clustering: K-Means
12 11/9-11/15	Exam 2: Monday November 9th Clustering: GMMs Inheritance
13 11/16-11/22	Iterators and Generators Classification: Naive Bayes Classification: kNN
14 11/23-11/29	Classification: Logistic regression Thanksgiving Break: No Class
15 11/30-12/6	Introduction to deep learning Neural network architectures Conclusion
Finals	Exam 3

Academic Honesty

Unless expressly allowed, you are expected to complete all assignments by yourself. However, you are allowed to discuss general issues with other students (programming techniques, clearing up confusion about requirements, etc.). You may discuss particular algorithmic issues on Piazza (but do not copy code!).

We will be using software designed to catch plagiarism in programming assignments, and copying on exams. A student is considered in violation of the academic honesty policy regardless of whether they are the one “copying” or the one “being copied from.”

Punishments for academic dishonesty are severe, including receiving an F in the course or being expelled from the University. By departmental rules, all instances of cheating will be reported to the Dean of Students. On the first instance of cheating, students will receive a 0 on the assignment or exam; the second instance of cheating will result in a failure of the course.

Campus Interruptions

In the event of a major campus emergency, course requirements, deadlines and grading percentages are subject to changes that may be necessitated by a revised semester calendar or other circumstances beyond the instructor’s control. In such an event, information will be provided through the course website, Piazza, and email.

Nondiscrimination Statement

Purdue University is committed to maintaining a community which recognizes and values the inherent worth and dignity of every person; fosters tolerance, sensitivity, understanding, and mutual respect among its members; and encourages each individual to strive to reach his or her own potential. In pursuit of its goal of academic excellence, the University seeks to develop and nurture diversity. The University believes that diversity among its many members strengthens the institution, stimulates creativity, promotes the exchange of ideas, and enriches campus life. More details are available on our course Brightspace table of contents, under University Policies.

Accessibility

Purdue University strives to make learning experiences as accessible as possible. If you anticipate or experience physical or academic barriers based on disability, you are welcome to let me know so that we can discuss options. You are also encouraged to contact the Disability Resource Center at: drc@purdue.edu or by phone: 765-494-1247. More details are available on our course Brightspace under Accessibility Information.

Mental Health Statement

If you find yourself beginning to feel some stress, anxiety and/or feeling slightly overwhelmed, try WellTrack. Sign in and find information and tools at your fingertips, available to you at any time.

If you need support and information about options and resources, please contact or see the Office of the Dean of Students. Call 765-494-1747. Hours of operation are M-F, 8 am- 5 pm.

If you find yourself struggling to find a healthy balance between academics, social life, stress, etc. sign up for free one-on-one virtual or in-person sessions with a Purdue Wellness Coach at RecWell. Student coaches can help you navigate through barriers and challenges toward your goals throughout the semester. Sign up is completely free and can be done on BoilerConnect. If you have any questions, please contact Purdue Wellness at evans240@purdue.edu.

If you're struggling and need mental health services: Purdue University is committed to advancing the mental health and well-being of its students. If you or someone you know is feeling overwhelmed, depressed, and/or in need of mental health support, services are available. For help, such individuals should contact Counseling and Psychological Services (CAPS) at 765-494-6995 during and after hours, on weekends and holidays, or by going to the CAPS office of the second floor of the Purdue University Student Health Center (PUSH) during business hours.