

BDA 572: Python Scripting for Social Science

Spring 2023

Schedule # 12873

Course Information

Class Days: Wed

Class Times: 4:00-6:40

Class Location: SSW 2512

Mode: lecture, face-to-face

Platform: Canvas

Jean Mark Gawron

Phone: 619-594-5268

Email: gawron@sdsu.edu

Office location: SHW 238

Zoom

<https://SDSU.zoom.us/j/3549775500>

Office hours: Tu Th 12:30-1:30

Course Description

The primary goal of the course is to acquaint students with the basics of the Python programming language and to introduce them to some of the many tools it provides for data collection and data analysis. A secondary goal is to promote a computational view of data collection and data analysis. In addition to providing the raw computing power that makes many data cleanup and data sorting tasks feasible, tools such as computer visualizations and machine learning can provide real help in understanding a whole host of statistical and structural concepts in data analysis, while amplifying their power in practical settings. For example, many of the machine learning packages in Python's `scikit_learn` module teach fundamental statistical concepts and are furnished with examples that apply them to large-scale data sets. Topics covered in the course include Python types, Python functions and control structure, computing with matrix data representations (`numpy`), data aggregation tools (`pandas`), social networks (`networkx`), and machine learning (`scikit_learn`, `pytorch`). There are no course pre-requisites. No knowledge of programming will be assumed. Students need upper division standing and some openness to acquiring computational skills.

Face-to-Face Instruction

Students who registered for face-to-face classes were expected to attend as indicated in the course schedule. Faculty teaching face-to-face courses are not required to create a new, alternative on-line class as an accommodation for any student. Students should prepare for class by doing reading and assignments, and they should attend with the intention of participating. That means participating in discussion, asking and answering questions, being respectful to your peers and actively contributing to the ongoing process of collaborative problem-solving that will take up much of our time.

Course materials

Materials	Required or optional	Where and how it can be obtained
A Whirlwind Tour of Python. Jake Van der Plas. 2016. O'Reilly	Optional	Whirlwind Tour Of Python Github Page
Python for Social Science. Jean Mark Gawron. Online textbook, distributed free.	Required	Book Draft
Numerous online Python notebooks for inclass work and assignments	Required	

Course Design: Major Assignments and Assessments

The course has one required textbook, [Python for Social Science](#) (Jean Mark Gawron) and one recommended [A Whirlwind Tour Of Python](#) (Jake Vanderplas), as well as numerous Python notebooks. All are available online. There will be assignments, a midterm, and a final project.

The class will use Canvas. All assignments will submitted on Canvas. In addition, some additional class materials, including the online textbook, will be on [the class web site](#). There are also a number of materials (Python notebooks) hosted on [my github repository](#). Here is a week by week outline, which may undergo slight changes as the semester progresses.

Course Schedule		
------------------------	--	--

Date	Activity	Assignment
Wed Jan. 18	Introduction	Running Python assignment
	Python types notebook	Python Types assignment
Wed Jan 25	Loops, conditionals, programming examples	Functions/loops assignment
Wed Feb 1	Functions, Classes, importing and namespaces, programming examples	
Wed Feb. 8	Introduction to numpy . Functions assignment debriefing. Boolean notebook.	Numpy assignment
Wed Feb 15	Introduction to numpy . Functions assignment debriefing. Boolean notebook.	
Wed Feb 22	Numpy and linear algebra, transforming data in matrix form. Pandas intro	Pandas assignment A
Wed Mar 1	Pandas ctd	Pandas assignment B
Wed Mar 8	Midterm. Pandas II	Midterm
Wed Mar 15	Pandas II ctd.	
Wed Mar 22	Simple regression, Linear classifiers	Regression and Classification assignment
Wed Mar 29	Spring break	
Wed Apr. 5	Issues in classifying text. Introduction to scikit_learn	Text Classification assignment
We Apr. 12	Text classifiers concluded; Introduction to Social Networks	
Wed Apr. 19	Social Networks continued. Introduction to Networkx .	Social networks assignment
Wed Apr. 26	Final unveiled	
Wed. May 3	Last day of class. Data bias.	Final
Wed May 10	Finals week	

Grading Policies

Grading will be based on exercises, quizzes, and a final project.

- Exercises: 50%
- Extended Midterm Assignment 25%
- Final Assignment 25%

Assignments, except for the first one, will generally be worth 6 points. Here is a rough rubric

6 (100%)	Every problem attempted, effort on all problems, commented code. Even if the answers to some are wrong or give incorrect results, effort has been made and code has been tested. If something doesn't work, comments explain what happens.
4,5 (67%-83%)	Very little or no effort made for at least one problem, code has very obviously not been tested, or code is uncommented.
2 (33%)	No effort made on at least half the problem set. Little on the rest/

Assignments are worth 50 raw points; there are 8 assignments worth 6 points and one, the first, worth 2, for a total of 50 assignment points. To get a B in this class, you must earn an average score of 4 on your assignments and do comparably on your midterm and final assignment. Late assignments will be graded according to the lateness policy. Note that a perfect assignment with a score of 6 earns 3 points if it's late, while an assignment worth 3 points on time is worth 1.5 points late.

Most of the assignments in this course will be given to students in the form of a Python notebook, often one that follows lecture material quite closely. Every notebook assignment will be posted online in an assignment notebook and students must turn in an edited copy of that notebook on Canvas to receive credit for the assignment

The general structure of the course is not well-suited to late assignments. Some assignment solutions will be discussed in detail and posted on the day they are turned in, and thus students

who turn assignments in late will be at a nominal advantage. However, problems requiring explanations must still be answered in the student's own words, solution code must not simply be copied, and late assignments must include all problems for which solutions have not been posted in order to receive any credit at all.

Here is the lateness policy:

1. Up to two weeks late: 50% credit for assignment.
2. More than two weeks late: not accepted.
- 3 There is no grace period for the extended midterm assignment and the final project, but students with special circumstances may communicate with the instructor to request an extension.

Collaboration is allowed on the assignments (including the Running Python assignment). It is not allowed on the Midterm and Final. The class assignments are learning activities that earn you points. The Midterm and Final are out and out evaluations of what you've learned. Collaboration on the Midterm or Final is considered **cheating**.

Group work

Collaboration is allowed on the assignments (including the Running Python assignment). It is not allowed on the Midterm and Final. The class assignments are learning activities that earn you points. The Midterm and Final are out and out evaluations of what you've learned.

To put it differently, group work is encouraged on the assignments. The Midterm and Final must be completed without any help or collaboration.. You should not seek outside help for these parts or the class, nor should you offer your classmates any help on extended midterm assignment. Collaboration on the Midterm or Final is considered **cheating**.

Student Learning Outcomes

- Read a simple python program and then describe what it does, and what resources are required to do it;
- Write simple scripts executing basic data analysis tasks for social science data, especially aggregation operations,
- Apply some native Python data analysis tools to data, including pandas and numpy, and understand the basic data structures on which those tools operate, especially arrays and data frames.
- Transform data of various kinds (especially text data) into a form in which useful statistical analysis and classification can be applied;
- Apply simple Python visualization tools to gain insights into variable dependencies.

Communication

My preferred gender pronouns are he/him/his. My official class roster only has your legal names. I will gladly honor your request to address you by an alternate name and/or gender pronoun. Please advise me of this early in the semester so that I may make appropriate changes to my records.

Questions during lecture will be encouraged, as we will be interacting live with Python and computational and conceptual issues can often be resolved simply. You can also email me (gawron@sdsu.edu) with questions about concepts and assignments or to schedule meetings not during my office hours. I prefer that we meet during office hours, but if my office hours conflict with your schedule, I will try to accommodate you.

Technology/Materials access

Canvas will be used for submitting assignments. Most of the online material will be available through links provided on Canvas.

Medical-related absences

Students with medical conditions that would present a COVID-related risk in a face-to-face instructional setting should contact the Student Ability Success Center (<https://sdsu.edu/sasc>) to begin the process of getting support. Please do not come to campus if you do not feel well. Remain home and monitor your symptoms and seek medical attention as needed.

- University policy instructs students to contact their professor/instructor/coach in the event they need to miss class due to an illness, injury, or emergency. All decisions about the impact of an absence, as well as any arrangements for making up work, rest with the instructors.
- If a student misses class because of COVID-19, either because they have been diagnosed and are quarantined or are required to isolate and would like to request a class excuse letter, the student should send an email to vpsafrontdesk@sdsu.edu to notify the university. Student Affairs and Campus Diversity will initiate the process for absent letters to be sent to course instructors, Assistant Deans, and the Provost. Medical documentation may be required prior to the letter being issued.
- [Student Health Services](#) (SHS) does not provide medical excuses for short-term absences due to illness or injury. When a medical-related absence persists beyond five days, SHS will work with students to provide appropriate documentation.
- When a student is hospitalized or has a serious, ongoing illness or injury, SHS will, at the student's request and with the student's consent, communicate with the student's instructors via the Vice President for Student Affairs and Campus Diversity and may communicate with the student's Assistant Dean and/or the [Student Ability Success Center](#).

Finding Help on Campus (& Academic Support Services)

Need help finding an advisor, tutor, counselor, or require emergency economic assistance? The [SDSU Student Success Help Desk](#) is here for you. Student assistants are available via Zoom Monday through Friday, 9:00 AM to 4:30 PM to help you find the office or service that can best assist with your particular questions or concerns.

Also consider the CAL Student Success Center: <https://cal.sdsu.edu/student-resources/student-success>.

A complete list of all academic support services—including the [Writing Center](#) and [Math Learning Center](#)—is available on the Student Affairs' [Academic Success](#) website. Counseling & Psychological Services (619-594-5220, sdsu.edu/cps) offers a range of psychological services for students. Emergency support is available after hours at the same phone number. The San Diego Access and Crisis Line can also be accessed 24 hours/day (1-888-724-7240).

Academic Honesty

The University adheres to a strict policy prohibiting cheating and plagiarism. Examples of academic dishonesty include but are not limited to:

- Copying, in part or in whole, from another's test or other examination;
- Obtaining copies of a test, an examination, or other course material without the permission of the instructor;
- Collaborating with another or others in coursework without the permission of the instructor;
- Falsifying records, laboratory work, or other course data;
- Submitting work previously presented in another course, if contrary to the policies of the course;
- Altering or interfering with grading procedures;
- Assisting another student in any of the above;
- Using sources verbatim or paraphrasing without giving proper attribution (this can include phrases, sentences, paragraphs and/or pages of work);
- Copying and pasting work from an online or offline source directly and calling it one's own;
- Using information found from an online or offline source without giving the author credit;
- Replacing words or phrases from another source and inserting one's own words or phrases.

Unauthorized recording or dissemination of virtual course instruction or materials by students, especially with the intent to disrupt normal university operations or facilitate academic dishonesty, is a violation of the Student Conduct Code. This includes posting of exam problems or questions to on-line platforms. Violators may be subject to discipline.

The California State University system requires instructors to report all instances of academic misconduct to the Center for Student Rights and Responsibilities. Academic dishonesty will result in disciplinary review by the University and may lead to probation, suspension, or expulsion. Instructors may also, at their discretion, penalize student grades on any assignment or assessment discovered to have been produced in an academically dishonest manner.

Classroom Conduct Standards

SDSU students are expected to abide by the terms of the [Student Conduct Code](#) in classrooms and other instructional settings. Violation of these standards will result in referral to appropriate campus authorities. Prohibited conduct includes:

- Willful, material, and substantial disruption or obstruction of a University-related activity, or any on-campus activity.
- Participating in an activity that substantially and materially disrupts the normal operations of the University or infringes on the rights of members of the University community.
- Unauthorized recording, dissemination, or publication (including on websites or social media) of lectures or other course materials.
- Conduct that threatens or endangers the health or safety of any person within or related to the University community, including:
 1. Physical abuse, threats, intimidation, or harassment.
 2. Sexual misconduct

Zoom Etiquette

Since we are scheduled for face-to-face class sessions most of this semester, classes will only be on Zoom if university policy changes or there is an emergency. But for those classes or any others we may have to conduct on Zoom, here are some guidelines. Students should treat these sessions as much as possible like real classes. In particular, they should prepare for class by doing reading and assignments, and they should attend with the intention of participating. That means participating in discussion, asking and answering questions, being respectful to your peers and actively contributing to the ongoing process of collaborative problem-solving that will take up much of our time.

Students will not be required to turn on video for class, but they are encouraged to do so. When attending a class virtually, students should dress appropriately as they would for a live class. Most class sessions will be recorded. Students will be informed when a class is being recorded. Students who wish to remain off camera will be allowed to do so.

A lot of research has shown that social presence is a factor in learning, and you can contribute to the sense of community and the learning experience of your peers, by being onscreen and actively participating.

Accommodations

SDSU via the [Student Ability Success Center](#) (SASC) provides accommodations for students with documented disabilities or medical conditions covered under the Americans with Disabilities Act (ADA). In keeping with current public health guidance, I cannot provide arrangements to students without an ADA-qualified disability or medical condition.

If you are a student with a disability and are in need of accommodations for this class, please contact the Student Ability Success Center at sascinfo@sdsu.edu (or go to sdsu.edu/sasc) as soon as possible. Please know accommodations are not retroactive, and I cannot provide accommodations based upon disability until I have received an accommodation letter from the Student Ability Success Center. SASC registration and accommodation approvals may take up to 10-14 business days, so please plan accordingly.

Student Privacy and Intellectual Property

The [Family Educational Rights and Privacy Act](#) (FERPA) mandates the protection of student information, including contact information, grades, and graded assignments. I will use [Canvas / email] to communicate with you, and I will not post grades or leave graded assignments in public places. Students will be notified at the time of an assignment if copies of student work will be retained beyond the end of the semester or used as examples for future students or the wider public. Students maintain intellectual property rights to work products they create as part of this course unless they are formally notified otherwise.

Religious Observances

According to the University Policy File, students should notify instructors of planned absences for religious observances by the end of the second week of classes.

Sexual violence / Title IX mandated reporting

As an instructor, one of my responsibilities is to help create a safe learning environment on our campus. I am a mandated reporter in my role as an SDSU employee. It is my goal that you feel able to share information related to your life experiences in classroom discussions, in your written work, and in our one-on-one meetings. I will seek to keep the information you share private to the greatest extent possible. However, I am required to share information regarding sexual violence on SDSU's campus with the Title IX coordinator, Gail Mendez (619-594-6464). She (or her designee) will contact you to let you know about accommodations and support services at SDSU and possibilities for holding accountable the person who harmed you. Know that you will not be forced to share information you do not wish to disclose and your level of involvement will be your choice. If you do not want the Title IX Officer notified, instead of disclosing this information to your instructor, you can speak confidentially with the following people on campus and in the community. They can connect you with support services and discuss options for pursuing a University or criminal investigation. Sexual Violence Victim Advocate (619-594-0210) or Counseling and Psychological Services (619-594-5220, psycserv@sdsu.edu). For more information regarding your university rights and options as a survivor of sexual misconduct or sexual violence, please visit titleix.sdsu.edu.

SDSU Economic Crisis Response Team

If you or a friend are experiencing food or housing insecurity, technology concerns, or any unforeseen financial crisis, it is easy to get help! Visit sdsu.edu/ecrt for more information or to submit a request for assistance. SDSU's Economic Crisis Response Team (ECRT) aims to bridge the gap in resources for students experiencing immediate food, housing, or unforeseen financial crises that impact student success. Using a holistic approach to well-being, ECRT supports students through crisis by leveraging a campus-wide collaboration that utilizes on- and off-campus partnerships and provides direct referrals based on each student's unique circumstances. ECRT empowers students to identify and access long-term, sustainable solutions in an effort to successfully graduate from SDSU. Within 24 to 72 hours of submitting a referral, students are contacted by a member of ECRT and are quickly connected to the appropriate resources and services.

For students who need assistance accessing technology for their classes, visit our ECRT website (sdsu.edu/ecrt) to be connected with the SDSU library's technology checkout program. The technology checkout program is available to both SDSU and Imperial Valley students.

Land Acknowledgement

For millennia, the Kumeyaay people have been a part of this land. This land has nourished, healed, protected and embraced them for many generations in a relationship of balance and harmony. As members of the San Diego State University community, we acknowledge this legacy. We promote this balance and harmony. We find inspiration from this land, the land of the Kumeyaay.

Diversity and Inclusion

To begin with, let's discuss the classroom experience: Here are some ideas intended to help create a learning environment that supports diversity of thought and honors your identities (including race, gender, class, sexuality, religion, and ability).

- If you have a name and/or set of pronouns that differ from those that appear in your official student records, let me know!
- I (like many people) am still in the process of learning about diverse perspectives and identities. If you feel that something that was said or done in class (by anyone) affected your performance (for example, your comfort level in speaking in class or your ability to concentrate or to learn), please talk to me about it. I would prefer that you come to office hours or email me, but if you feel you need to communicate anonymously, you might consider getting a "burner" email account not linked to your real name. One possibility is [GoogleGmail](https://www.google.com/gmail/about/), which is currently asking for a verifiable cellphone number to create an email account, but that information will not be attached to your emails, and if you opt for anonymity, I promise not to make any efforts to try to get it.
- If you feel that your performance in the class is being impacted by interactions outside of class, again, don't hesitate to come and talk with me. If you wish, you can also email an anonymous description of the issue (which may lead to me making a general

announcement to the class, to address your concerns). If you prefer to speak with someone outside of the course, Luke Wood, who runs the [Division of Student Affairs and Campus Diversity](#), is an excellent resource.

- Finding your purpose and your passions is important. One of the core values of diversity-based education is that exposure to a diverse set of ideas enriches all of us, making it that much more likely that we will find the things that are the most meaningful to us. An important part of this is that everyone try to speak out. To help provide you with opportunities, I will try to call on as many of you as I can throughout the semester. I also welcome your visits during office hours or schedule appointments to discuss diversity-related issues that have arisen in class, or how to pursue your professional and academic growth in ways that may not relate directly to material on the syllabus.

Finally, it may not seem that scientific data analysis, in particular, text analysis with machine learning, interacts much with issues of diversity, but in fact, it has been shown that machine learning algorithms incorporate bias induced both by the data and the algorithm designers. Examples include automatic speech recognition systems that perform poorly on underrepresented dialects, automatic interview systems that build in the biases of their training data, text generation systems trained on "standard" English documents sounding always sounding like public service announcements, even when sharing a recipe. As we mature as data analysts trying to build faithful pictures of the world, we will pay more and more attention in class to issues of training data and algorithmic bias, and how you as data scientists can contribute to unbiasing your data and analysis tools.