

Python for Social Science BDA/Ling 572

Jean Mark Gawron
San Diego State University
gawron@sdsu.edu
http://gawron.sdsu.edu/python_for_ss

2019 Aug 27

1 Course Information

Sched #	22201	email	gawron@sdsu.edu
Class days	Wed	Office	SHW 238
Class Times	16:00-18:40	Office hours	W 11:00-12:00 & Th 16:00-17:00
Class location	Online	Office Hours Zoom link	https://SDSU.zoom.us/j/3549775500

2 Goals

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 available for data collection and data analysis.

3 Practice

The course has one required textbook, *Python Data Science Handbook* (Jake VanderPlas). It will also rely on online materials designed both by the instructor and others. There will be weekly assignments, a midterm, and a final project.

The class will not be using Blackboard, except for assignment submission. All class materials will be on [The class webpage](#).

4 Learning Outcomes

1. Read a simple python program and then describe what it does, and what resources are required to do it;
2. Write simple scripts executing basic data analysis tasks for social science data, especially aggregation operations,
3. 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.
4. Transform data of various kinds (especially text data) into a form in which useful statistical analysis and classification can be applied;
5. Apply simple Python visualization tools to gain insights into variable dependencies.

5 Pre-requisites

No course pre-requisites. No knowledge of programming will be assumed. Upper division standing. Some openness to acquiring computational skills. Some knowledge of what counts as interesting data in your own Social Science.

6 Grading

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

- Exercises: 50%
- Extended Midterm Assignment 20%
- Final project 30%

7 Grading Assignments

Grading of problem sets is as follows:

- Plus (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.
- Check (75%) Very little or no effort made for at least one problem, code has very obviously not been tested, or code is uncommented.
- Minus (50%) No effort made on at least half the problem set.

To get a C in this class, you must earn a Plus on the majority of the assignments. Three Minus scores are a cause for concern and will require a meeting with the instructor. Late assignments will be graded according to the lateness policy, as solutions will be posted immediately.

8 Late Assignments

The general structure of the course is not well-suited to late assignments or missed quizzed. Assignment solutions will be discussed in detail on the day they are turned in, and thus students who turn assignments in late will be at an advantage. However, to allow for some flexibility, late assignments will receive partial credit. Here is the lateness policy:

1. Up to two weeks late: 50% credit for assignment (this basically turns a Plus into a Check, and a Check into a Minus). Late assignments must include all problems for which solutions have not been posted in order to receive any credit at all.
2. More than two weeks late: not accepted
3. There is no grace period for the extended midterm assignment and the final, but students with special circumstances may communicate with the instructor to request an extension.

9 Group Work

Group work is encouraged on the assignments. The extended midterm assignment and the final should be completed without any help or collaboration.

To be clear on this, collaboration or group work on the midterms and finals will be considered cheating. You should not seek outside help for these parts or the class, nor should you offer your classmates any help on midterm or final problems.

When turning in collaborative assignments, your collaborators should be identified on your paper.

One option to explore with respect to group work:

You may wish to change groups in mid-semester. This is a good idea. I encourage existing groups to admit new members. Moving around will help you get to know more people and you will probably learn more. On the other hand, if you find a group that is a good fit, there is no requirement that you leave it.

10 Attendance

Attendance is not a formal part of your grade. In the time of this global pandemic, flexibility about how you attend class is going to be essential.

However, be aware that assignments are part of your grade, and class time will mostly be devoted to doing worksheets (which are available on Blackboard) that contain practice problems very like the homework problems. During the inclass work sessions, hints on how to solve problems on the assignments are handed out liberally.

11 Zoom Etiquette

Since we are not having face-to-face class sessions this semester, classes will be conducted in Zoom. 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, 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.

12 Cheating: Academic Dishonesty

Academic misconduct includes the following examples as well as any other similar conduct which is aimed at falsely representing a student's academic performance: cheating, plagiarizing, unauthorized collaborating on course work, stealing course examinations or materials, falsifying records or data, or intentionally assisting another individual in any of the above.

Students who cheat will receive an "F" for the course grade and the Linguistics Department's Undergraduate Advisor and SDSU's Judicial Coordinator will be notified. As explained in the Group Work Section, group work and collaboration on assignments is allowed, but **any** collaboration on the Midterm or Final is cheating. These are tests intended to evaluate your individual level of achievement.

For more information on SDSU's policies and procedures regarding academic misconduct visit the following site: [Academic Dishonesty](#).

13 Student Accommodation

If you are a student with a disability and believe you will need accommodations for this class, it is your responsibility to contact Student Ability Success Center at (619) 594-6473. To avoid any delay in the receipt of your accommodations, you should contact Student Ability Success Center as soon as possible. Please note that accommodations are not retroactive, and that I cannot provide accommodations based upon disability until I have received an accommodation letter from Student Ability Success Center. Your cooperation is appreciated.

14 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 Blackboard 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.

15 Religious observances

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

16 Special Note

We are still in the middle of a devastating pandemic that has affected every aspect of our lives. There may be times during this semester when things become overwhelming. I understand that, and I want you to know that I will do everything I can to help you be successful in this class. You never have to give me personal information that you feel uncomfortable disclosing — just let me know what I can do to help. If I cannot help you, I can put you contact with someone on campus who can.

17 Office Hours

W 11:00-12:00
W 2:00-3:00
Th 16:00-17:00
by appointment

18 Weekly Syllabus

Week One	Running & Installing Python; basic features of the language & community
Week Two	Python Types: Containers
Week Three	Control Structures: If-then, Loops, Functions
Week Four	Putting together larger functions, classes
Week Five	Numerical tools, tabular data (arrays)
Week Six	Machine learning tools, text classification (sklearn)
Week Seven	Text processing basics (nltk), regular expressions (re)
Week Eight	Networks, social and otherwise (networkx)
Week Nine	R-like Dataframe tools (pandas)
Week Ten	Tools for simple data analysis, regression (sklearn)
Week Eleven	Visualization I: Telling a story (bokeh), parallel coordinates, color maps, correlation heat maps
Week Twelve	Visualization II: Maps, Kernel Density estimates of distribution (sklearn)
Week Thirteen	Visualization III: Normal distributions, boxplots, violin plots
Week Fourteen	Web crawling basics
Week Fifteen	Project construction: From Data study