

Python for Social Science

Ling 572

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

2019 Aug 27

1 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.

2 Practice

The course has one required textbook, *Automate the Boring Stuff with Python* (Al Sweigart). It will also rely on online materials designed both by the instructor and others. There will be weekly assignments and several inclass quizzes, a midterm, and a final project.

3 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;

3. Apply some native Python data analysis tools to data;
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.

4 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.

5 Grading

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

- Exercises: 40%
- Quizzes: 20%
- Midterm 20%
- Final project 20%

6 Grading Assignments

Grading of problem sets is as follows:

- Plus 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 Very little or no effort made for at least one problem, code has very obviously not been tested, or code is uncommented.
- Minus No effort made on at least half the problem set, or problem set not turned in.

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.

7 Late Assignments

The general structure of the course is not well-suited to late assignments or missed quizzes. 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. Quizzes are designed to test understanding of foundation needed for further work, and without those foundations, progress will be slowed. However, to allow for some flexibility, late assignments will receive partial credit. Here is the lateness policy:

1. Up to one week 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 one week late: not accepted

8 Attendance

Attendance is not a formal part of your grade.

However, be aware that assignments are, and extensive amounts of class time will be devoted to working through exercises like those on the assignments. Similarly, hints on how to solve problems on the assignments and the midterms are handed out liberally in class. These hints will not be posted on the web pages.

9 Group Work

Group work is encouraged on the assignments. The midterm and final project should be completed without any help. To be clear on this, collaboration or group work on the midterms and finals will be considered cheating.

When turning in collaborative assignments, your collaborators should be identified on your paper. The code you write on your group assignments should be your own.

10 Office Hours

Tu,Th 12:45-1:45
Wed by appointment
1:30-2:30 (choose a appointment in this interval)

11 Mailing address

Dept: Department of Linguistics and Oriental Languages
Uni: San Diego State University
Address: 5500 Campanile Drive
City: San Diego, CA 92182-7727
Telephone: (619) 594-0252
Office: SHW 238
Email: gawron@mail.sdsu.edu

12 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