

Python for Public Policy and Management

PPD 599 | Fall 2022

Draft syllabus; subject to change

Thursday – 6PM to 9:20 PM Pacific Time
Location: VPD 106
4 Units

Instructor: Nicolas Duquette
Office: 234 Lewis Hall
Office Hours: 10 am to noon on Tuesdays, or by appointment.
Email: nduquett@usc.edu

Course Assistant: TBD
Office Hours: TBD
Email: TBD

Course Description

This course will provide you with the skills required to conduct quantitative public policy analysis using the Python programming language. We will begin with an introduction to the Python programming language in general and its packages for cleaning and processing data sets in particular. We will then learn commonly used tools for data analysis, including linear regression, logistic regression, regularized regression, random forest classification, and text analysis methods.

Prerequisite: None, but prior experience with basic concepts of probability and statistics (such as PPD 504, PPD 525, or similar) is recommended.

Learning Objectives

This course will focus on training students to be capable users of a modern computing platform for quantitative analysis. The class will emphasize both learning to use the Python programming language for data science and analytics, as well as fundamental concepts of statistical reasoning, predictive analysis, and data visualization that allow for thoughtful use of data science in the service of understanding public policies and social data.

Course Assignments and Grading Policies

Grades will be based on a mixture of exams, problem sets, and mini-tasks. Attendance and participation in in-class activities are extremely important for mastery of the course material, but are not formally captured in your grade.

Problem Sets

There will be seven problem sets in this class. Each is worth ten percent of your total grade, except for your lowest problem set grade, which is dropped. The six undropped problem sets therefore account for 60% of your final grade.

Students may work on the problem sets in groups of three or fewer and submit as a group.

Problem sets are due at the start of class on the due date listed here. Problem sets submitted that day after the start of class will lose 10% of possible points. Problem sets submitted the day after class lose 20% of the possible points. Problem sets submitted two days after class lose 30% of possible points. After that, I will post problem set solutions, and problem sets still not submitted will not be accepted for grading and will receive a zero grade.

Exams

We will have two exams: one midterm and a final. Exams count for 30% of your grade in total. The exams will either count equally, or the final will be worth five times more than the midterm, whichever results in the higher grade.

Mini-tasks

There will be six mini-tasks you must complete by specific deadlines. Each is worth between 1 and 3 points and graded complete/incomplete. Mini-tasks cannot be submitted late for partial credit.

The first mini-task is a syllabus quiz. It is due by the start of class in the second week of the semester. Take your syllabus quiz at [this link](#).¹

The other mini-tasks are provided as PDFs in the mini-task folder of the class web page.

¹<https://forms.gle/eww1rnMmqfxPHbQ1A>

Graded items by date and weight

Assignment	Due date	Weight
<i>Problem sets</i>		60*
1 - Python basics	September 8	10
2 - Working with pandas	September 15	10
3 - Description and plotting	September 29	10
4 - Ordinary regression	October 13	10
5 - Regularized prediction	November 10	10
6 - Trees and forests	November 17	10
7 - Text analysis	December 1	10
<i>Exams</i>		30**
Midterm	October 20	15 or 5
Final exam	December 8	15 or 25
<i>Mini-tasks</i>		10
1 - Syllabus quiz	September 1	1
2 - Join Cocalc	September 1	1
3 - FRED API key	September 8	1
4 - Data Privacy Attestation	October 6	1
5 - iMDB API key	October 20	1
6 - Film scores	November 10	2
7 - Text categorization	November 17	3
<i>Total weight</i>		100

* – The lowest problem set score will be dropped, so total points on problem sets sum to 60, not 70.

** – Exams sum to 30, but weights on individual exams will be either equal or highly weighted toward the final, whichever leads to a higher score.

The individual graded items will be converted into percentage points of the final course grade and then a letter grade for the course assigned as follows:

Letter grade assignment	
Raw grade percentage	Letter grade
≥ 93 out of 100	A
90-92.99	A-
87-89.99	B+
83-86.99	B
80-82.99	B-
77-79.99	C+
73-76.99	C
70-72.99	C-
<70	$\leq D+$

Technology

You must attend class with a laptop computer able to connect to the Internet and perform basic browsing tasks, and you must buy a semester license to use [Cocalc](#). These licenses cost \$14. Cocalc is a browser-based data science platform, like Google Docs for scientific computing, which will make it easy to work with other students on problem sets and to work with Python without doing software installation on your personal device.

You will also need to buy one month of a paid data service to complete our in-class exercises. That license will cost about \$15. There are no required textbooks or other required costs for this course, although you may decide to buy books if you find them useful.

USC technology rental program

If you need resources to successfully participate in your classes, such as a laptop or internet hotspot, you may be eligible for the university's equipment rental program. To apply, please submit an application at [this link](#). The Student Basic Needs team will contact all applicants in early August and distribute equipment to eligible applicants prior to the start of the fall semester.

Health

I do not expect any further classroom changes or disruptions from the COVID-19 pandemic. However, surprises happen. If there is an unexpected disruption to public health, such as from

a resurgent COVID-19, a different epidemic disease, a natural disaster, or some other cause, I will at my sole discretion change this course to adapt to circumstances. This may include remote learning, a change in grading, or other changes.

For now, please use common sense to protect your health and others':

1. If you feel sick, please either *stay home*, or if you do come to class, please *wear a covering over your mouth and nose*.
2. You do *not* need to provide a doctor's note or other documentation of your illness. Please do not send me photographs of bodily disease or traumatic injury; I believe you.
3. Your grading incentive to attend and participate in class is the opportunity to learn from in-class exercises. You can miss class with no direct grading penalty.
4. If you happen to be ill on an exam date, please contact me as soon as possible so I can determine the best course of action for your particular situation.

Readings

Each class meeting has an associated list of suggested readings, listed below. Where possible, I suggest readings that can be accessed in an electronic format through the USC library, the course Google Drive, or another legitimate online source at no cost. Most of these also have paper or e-book versions which may be easier to read but not free of charge.

For technical topics like programming and statistics, I often find it useful to read more than one text on the same topic. In case your brain works the same way, I try to provide a variety of readings for each topic that approach concepts from different perspectives. I do not expect you to read all provided sources for all topics. Students should choose readings based on their own tastes, career goals, and learning objectives.

Symbol key:

★ – A core reading. If you look at one thing, look at this.

📄 – Available as a PDF in the class Google Drive

Course Schedule: A Weekly Breakdown

Below is a list of topics by weeks and associated methodological readings.

1. August 25, 2022

What do we want to learn from data? / Introduction to Python

Downey, A. (2015). *Think Python: How to Think Like a Computer Scientist*. Green Tea Press, second edition. Chapter 1. An introduction to Python and to computer programming for people who know very little about either. Available at no cost from <https://greenteapress.com/wp/think-python-2e/>.

2. September 1, 2022

Numerical Python programming

Downey. Chapters 2–3.

- ★ McKinney, W. (2018). *Python for Data Analysis: Data wrangling with Pandas, NumPy, and IPython*. O’Reilly Media, Inc., second edition. Chapters 2–3. Available for free with a USC email address at <https://learning.oreilly.com/home/>.

Adhikari, A. and DeNero, J. (2020). *Computational and Inferential Thinking: The Foundations of Data Science*. <https://inferentialthinking.com/>. Chapters 1, 3 and 4. A free online textbook on data science in Python.

3. September 8, 2022

Importing and cleaning data with Python

- ★ McKinney, chapters 6–8

4. September 15, 2022

Descriptive charts and tables

- ★ McKinney, chapter 9

Michael Waskom, Seaborn tutorial at <https://seaborn.pydata.org/tutorial.html>.

McKinney, chapter 10.

Heiss, F. and Brunner, D. (2020). *Using Python for Introductory Econometrics*. <http://upfie.net/>. Chapter 1. Available for purchase or as a free online version. Covers core ideas from regression analysis and econometrics with Python implementations of examples from Jeffrey Wooldridge’s very popular textbook.²

Bruce et al., chapter 3

Taboga, chapters 22, 26, 50–54

5. September 22, 2022

Web scraping

- ★ Mitchell, R. (2018). *Web Scraping with Python*. O’Reilly Media, Inc, second edition, chapters 1 and 2. Available for free with a USC email address at <https://learning.oreilly.com/home/>.

6. September 29, 2022

Ordinary least squares / univariate regression

- ★ McKinney, chapter 5
- Grus, J. (2019). *Data science from Scratch : first principles with Python*. O’Reilly Media, Inc., second edition, chapters 5–6. Short and to the point, with Python examples. Available for free with a USC email address at <https://learning.oreilly.com/home/>.
- Bruce, P., Bruce, A., and Gedeck, P. (2020). *Practical statistics for data scientists : 50+ essential concepts using R and Python*. O’Reilly Media, Inc., second edition, chapter 2. A very visual introduction to applied probability and statistics, with a mix of R and Python code. Available for free with a USC email address at <https://learning.oreilly.com/home/>.
- Cunningham, S. (2021). *Causal Inference: The Mixtape*. Yale University Press, New Haven. Chapter 2.1–2.10 “Probability and Regression Review.” Approachable overview of causal inference with workhorse estimators in econometrics. The print edition is great, but you’ll likely want to focus on the free online version, which includes Python code for every exercise: <https://mixtape.scunning.com/>.
- Taboga, M. (2017). *Lectuers on Probability Theory and Mathematical Statistics*. StatLect.com, third edition. Chapters 10–21. Recommended for PhD students and others looking for more detail on , this textbook presents mathematical probability and statistics with short, clear chapters and generous solved exercises. I recommend the very affordable paperback version [available on Amazon](#) for serious self-study. The book is a structured version of the free web version available at StatLect.com.

7. October 6, 2022

Multivariate regression estimation

- ★ Cunningham, “Probability and Regression Review” 2.13–2.25
- Bruce et al., chapter 4
- Heiss and Brunner, chapters 2–5.
- Taboga, Online appendix, “Linear regression models.” [[Hyperlink](#)]

8. October 13, 2022

Principles of machine learning

- ★ Burkov, A. (2019). *The Hundred-page Machine Learning Book*. <http://themlbook.com/>. Available as a pay-what-you-want e-book. Short but clear.
- Raschka, S. and Mirjalili, V. (2019). *Python machine learning machine learning and deep learning with Python, scikit-learn, and TensorFlow*. Packt Publishing, third edition, chapters 3 and 6. Available at no cost via the USC libraries [[hyperlink](#)].
- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly Media, second edition. Ch. 4 to p. 134. Available for free with a USC email address at <https://learning.oreilly.com/home/>.
- [Bruce et al.](#), chapters 5 and 6.
- [Taboga](#), Online appendix, “Logistic classification model.” [[Hyperlink](#)]

October 20, 2022

Midterm Exam

9. October 27, 2022

Regularized predictive regression

- ★ [Raschka and Mirjalili](#), chapter 10.
- [Burkov](#)
- [Taboga](#), Online appendix, “Ridge regression.” [[Hyperlink](#)]

10. November 3, 2022

Machine classification

[Burkov](#)

11. November 10, 2022

Decision trees and random forests

★ [Burkov](#)

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and regression trees*. CRC press. The decision tree OG, I haven't found a clearer exposition of the math behind the algorithm (in greater detail than [Burkov](#)) than this. Available as a free PDF through the USC libraries [[hyperlink](#)].

[Raschka and Mirjalili](#), chapters 3 and 10

12. November 17, 2022

Text-as-data and document classification

Gentzkow, M., Kelly, B., and Taddy, M. (2019). Text as data. *Journal of Economic Literature*, 57(3):535–74

Grimmer, J., Roberts, M. E., and Stewardt, B. M. (2022). *Text as Data*. Princeton University Press

Bird, S., Klein, E., and Loper, E. (2019). *Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit*. <https://www.nltk.org/book/>. The online version of the book is free and updated to Python 3; don't buy the paper version from O'Reilly, which is out of date now.

November 24, 2022

Thanksgiving recess—no class

13. December 1, 2022

Advanced natural language processing methods / Review and conclusion

[Raschka and Mirjalili](#), chapter 8.

December 8, 7–9 p.m.

Final Exam

Last updated: August 11, 2022

Statement on Academic Conduct and Support Systems

Academic Conduct

Plagiarism – presenting someone else’s ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in SCampus in Part B, Section 11, “Behavior Violating University Standards” <http://policy.usc.edu/scampus--part--b>. Other forms of academic dishonesty are equally unacceptable. See additional information in SCampus and university policies on scientific misconduct, <http://policy.usc.edu/scientific--misconduct>.

Students and Disability Accommodations

USC welcomes students with disabilities into all of the University’s educational programs. The Office of Student Accessibility Services (OSAS) is responsible for the determination of appropriate accommodations for students who encounter disability-related barriers. Once a student has completed the OSAS process (registration, initial appointment, and submitted documentation) and accommodations are determined to be reasonable and appropriate, a Letter of Accommodation (LOA) will be available to generate for each course. The LOA must be given to each course instructor by the student and followed up with a discussion. This should be done as early in the semester as possible as accommodations are not retroactive. More information can be found at osas.usc.edu. You may contact OSAS at (213) 740-0776 or via email at osasfrontdesk@usc.edu.

Support Systems

Counseling and Mental Health - (213) 740-9355 – 24/7 on call

<http://studenthealth.usc.edu/counseling>

Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

National Suicide Prevention Lifeline – 1 (800) 273-8255 – 24/7 on call

<http://suicidepreventionlifeline.org>

Free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week.

Relationship and Sexual Violence Prevention Services (RSVP) - (213) 740-9355(WELL), press “0” after hours – 24/7 on call

<http://studenthealth.usc.edu/sexual-assault>

Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

Office for Equity, Equal Opportunity, and Title IX (EEO-TIX) - (213) 740-5086

<http://eeotix.usc.edu>

Information about how to get help or help someone affected by harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants.

Reporting Incidents of Bias or Harassment - (213) 740-5086 or (213) 821-8298

http://usc-advocate.symplcity.com/care_report

Avenue to report incidents of bias, hate crimes, and microaggressions to the Office for Equity, Equal Opportunity, and Title for appropriate investigation, supportive measures, and response.

USC Campus Support and Intervention - (213) 821-4710

<http://campussupport.usc.edu>

Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

Diversity, Equity and Inclusion - (213) 740-2101

<http://diversity.usc.edu>

Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

USC Emergency - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call

<http://dps.usc.edu>, emergency.usc.edu

Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

USC Department of Public Safety - UPC: (213) 740-6000, HSC: (323) 442-120 – 24/7 on call

<http://dps.usc.edu>

Non-emergency assistance or information.

Office of the Ombuds - (213) 821-9556 (UPC) / (323-442-0382 (HSC)

<http://ombuds.usc.edu>

A safe and confidential place to share your USC-related issues with a University Ombuds who will work with you to explore options or paths to manage your concern.

Occupational Therapy Faculty Practice - (323) 442-3340 or otfp@med.usc.edu

<http://chan.usc.edu/otfp>

Confidential Lifestyle Redesign services for USC students to support health promoting habits and routines that enhance quality of life and academic performance.