Advanced Geospatial Data Analysis (in Python)
GEOG-G 489/589

Department of Geography
Indiana University
Fall 2019 Syllabus

Lecture: Monday & Wednesday 1:00 PM to 2:15 PM
Lecture Location: Student Building 221

Instructor: Dr. Natasha MacBean
Office: Student Building 204
Email: nmacbean@indiana.edu
Office hours: Monday 12:00 PM to 1:00 PM and Wednesday 3:15 PM to 4:15 PM.

Course description

This course introduces students to the Python programming language and the Python skills needed to perform complex data analysis and data visualization with large spatiotemporal datasets (for example model simulations and remote sensing data).

In the course we will learn how to install python and how to download the various packages we will use, as well as how to download the remote sensing, climate, and modeling datasets we may analyze. In the first part of the course we will review fundamental concepts in Python (e.g. data types, data structures, indexing, subsetting, looping over data) that will be useful for handling large volumes of data. Following that, we will cover the following data handling topics: reading and manipulating large, multi-layer spatio-temporal datasets, vector and raster data manipulation, map reprojection, area selection based on geographic coordinates, masking data, data visualization (including plotting maps). Finally, throughout the course we will cover the following data analyses on large geospatial datasets, including: applying basic statistics to large datasets (sub, mean, max, min), interpolation and smoothing, regression, function fitting for extracting information from time series data, change detection, image classification and spatial clustering, dimensionality reduction.

Throughout, we will focus on developing the most fundamental and useful data science skills – such as cleaning and tidying data in preparation for data analysis, and elegant and efficient code writing – so that your code writing skills result in useful scripts that can be used in the future to repeat similar analyses on new scientific questions.

Students will become familiar with the following Python packages: NumPy (Numerical Python), GDAL (Geospatial Data Abstraction Library), SciPy (Scientific Python), (Geo)Pandas (Python Data Analysis Library – also relies on packages Shapely and Fiona), CartoPy, PySAL (Python Spatial Analysis Library), scikit-learn, Matplotlib and Seaborn. There will be an emphasis on remote sensing (satellite) data, climate reanalysis and modeling datasets.

We will meet in the Student Building Room 221 computing cluster and all classes will be lab focused where topics will be introduced with practical exercises. No previous programming

experience is required, but it would be helpful (for example GEOG-G 250/577, GEOG-G 488/588, SICE I210).

Course Goals and Learning Outcomes

Course goals:
Students who engage in this class will be equipped to use Python to develop their own scientific code to analyses large spatiotemporal data analyses that will be useful in their own research and careers in geospatial data analysis. Students will be given a grounding in the basic principles of algorithm development and the construction of scientific programs for analyzing data that can be transferred and widely applied outside the topics covered in the class.

Learning outcomes:
At the end of this course, students should:

- Understand the basics of scientific computing
- Have a good working knowledge of the Python programming language
- Know how to install commonly available scientific software packages
- Be able to develop code and construct scientific programs Python to perform data manipulation tasks with large (high volume) geospatial datasets
- Be able to write efficient, well-structured and commented Python scripts for a range of data analysis tasks.
- Be able to carry out a variety of analytical tasks on large, multi-layer spatiotemporal datasets
- Understand technical issues with data visualization and geospatial data analysis and approaches to solve those issues.

Python
Python is a high-level programming language that is freely available, relatively easy to learn and portable across different computing systems. In Python, you can rapidly develop code/programs to perform the analyses you may need to do both in your research and future career. Code written in Python is also easy to maintain, is (or should be) self-documented, and can easily be linked to code written in other languages.
Advantages of Python include:

- it is automatically compiled and executed
- code is portable provided you have the appropriate Python modules installed.
- there are many well-documented existing packages you can employ for you data analysis needs
- there is an active user and development community, which means that there is a lot of useful advice and documentation online. Within this community, new capabilities appear over time and there are many existing extensions and enhancements easily available to you.

I suggest you use the anaconda python distribution if you are using your own computer, but we can discuss options for installing on your own computer during class.

Optional texts

There are no required texts for this course. All the relevant material will be provided in the assignments on Canvas. When learning how to program there are often a lot of useful tutorials and websites with advice online. This is especially true for a freely available programming language such as Python. However, students may find the following textbooks useful. I will bring copies of these textbooks to class in the first few weeks so students can have a look and decide if they are worth buying themselves.

- McKinney, W. (2012). *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. O'Reilly Media, Inc.
- Scopatz, A., & Huff, K. D. (2015). *Effective computation in physics: Field guide to research with python*. O'Reilly Media, Inc.
- Lawhead, J. (2015). *Learning Geospatial Analysis with Python Second Edition*. Packt Publishing Ltd.
- Grus, J. (2015). *Data science from scratch: first principles with python*. O'Reilly Media, Inc.
- Garrard, C. (2016). *Geoprocessing with Python*. Manning Publications.

The course page will also be updated with useful websites and articles specific to each topic that will help us during each class.

Course Format and Grading

In-class exercises: At the beginning of each class we will review the concepts that we will be covering during that class. This will include developing a logical argument for the algorithm we develop during the class. The algorithm is essentially the method, or set of methods, we will use to complete our programming task/data analysis. We will then discuss together the Python code needed to put that algorithm into an executable Python script. Following this group discussion, you will each get time to write the script in a Jupyter Notebook (a commonly used open-source application for writing and executing code). In-class exercise assignments will be posted on Canvas. The assignment will be due by the beginning of the next class. Assignments (in the form of your Jupyter Notebook) will be uploaded to Canvas. Homework in the form of DataCamp.com tutorials may be set before each class to give an introductory overview of a given Python topic before covering that topic in more detail during the class. The homework will not be graded.

In-class assessments: There will be three in-class assessments to test your knowledge of the material learned so far in the course during the 2nd class in the 4th, 8th and 12th week of the semester. These will be silent but open-book assessments in which you can also search online for relevant python syntax, but you will need to complete the assessment during the 1h 15 minutes of the class and you will not receive any help from the instructor or other students on the course.

Final Project: The final project will be designed and carried out in Python by each student on a topic of their own choosing. This project will be completed in the final 4 weeks of the course. The student will use the data downloading, manipulation and analysis methods learned in the class to answer a scientific question pertaining to their own data analysis and research interests. More information about the final project will be provided at the start of the semester. *Additional requirement for G589 students: the final project will be more extensive and based on your research interests. It will make up a greater percentage of your overall grade. More information will be given in class*.
Please note: The final project deadline will be 11:59pm Sunday 15th December.

Grading (489): In-class exercises – 45%
In-class assessments – 20%
Final project – 30%
Participation – 5%

Grading (589): In-class exercises – 35%
In-class assessments – 20%
Final project – 40%
Participation – 5%

Grading Scale

97% - 100%	A+	77% - < 80%	C+
93% - < 97%	A	73% - < 77%	C
90% - < 93%	A-	70% - < 73%	C-
87% - < 90%	B+	67% - < 70%	D+
83% - < 87%	B	63% - < 67%	D
80% - < 83%	B-	60% - < 63%	D-
		< 60%	F

Grade Dissemination:
All grades will be posted on Canvas. Please make sure to track your own grades, as mistakes can occasionally occur. If you have received a grade by mistake, please see me for a correction in Canvas.

Course Policies & Services

Assignments and Late Work: You may work with others on the exercises, but you may not copy directly from anyone. Your work must be your own. *Late exercises will have 10% deducted per day, but exercises will not be accepted once graded copies are returned to the class.* The only allowable exceptions pertain to the IU policy on religious observances, military

duty and family emergencies, illness with a doctor's note, or prior permission from me. If you have an excused absence for any of these reasons, make arrangements in advance or ASAP to makeup missed activities. *Plagiarism and cheating in any form will not be tolerated.*

Attendance: Attendance is mandatory in order to receive full credit for in-class activities. Your active participation in this class (aside from excused absences) makes up 5% of your grade.

Civility: I expect students to support one another in learning this material.

Academic Integrity: As a student at IU, you are expected to adhere to the standards detailed in the *Code of Student Rights, Responsibilities, and Conduct* (*Code*). Academic misconduct is defined as any activity that tends to undermine the academic integrity of the institution. Violations include: cheating, fabrication, plagiarism, interference, violation of course rules, and facilitating academic dishonesty. When you submit an assignment with your name on it, you are signifying that the work contained therein is yours, unless otherwise cited or referenced. Any ideas or materials taken from another source for either written or oral use must be fully acknowledged. All suspected violations of the *Code* will be reported to the Dean of Students and handled according to University policies. Sanctions for academic misconduct may include a failing grade on the assignment, reduction in your final course grade, and a failing grade in the course, among other possibilities. If you are unsure about the expectations for completing an assignment or taking a test or exam, be sure to seek clarification from your instructor in advance. Please also see this website for the IU College of Arts and Science's policies on Academic Integrity.

Plagiarism and Academic Dishonesty: Plagiarism is the act of taking someone else's work and presenting it as your own. Plagiarism can occur in several forms, but whether the action is intentional or not, it is in violation of the IU Student Code of Conduct. Copying and pasting text off of the Internet or any other source is NOT acceptable, as this is stealing someone else's work. If you use an author's materials verbatim, you must place these words in quotation marks, and must correctly reference them. Proper references must be placed in the text as well as in the bibliography. Be aware that when citing sources, you must do so not only when a passage is a direct quotation, but also when paraphrasing. Failure to do so will result in a "0" for the project. If you have questions or concerns regarding how to properly cite your resources, please ask me or contact Campus Writing Services. Additionally, if you are new to this university, are unclear about what plagiarism is, or would like a brief review of IU's standards, please look here and here.

Note Selling: Several commercial services have approached students regarding selling class notes/study guides to their classmates. Selling the instructor's notes/study guides in this course is not permitted. Violations of this policy will be reported to the Dean of Students as academic misconduct (violation of course rules). Sanctions for academic misconduct may include a failing grade on the assignment for which the notes/study guides are being sold, a reduction in your final course grade, or a failing grade in the course, among other possibilities. Additionally, you should know that selling a faculty member's notes/study guides individually or on behalf of one of these services using IU email, or via Canvas may also constitute a violation of IU information technology and IU intellectual property policies; additional consequences may result.

Students with Disabilities: The Americans with Disabilities Act (ADA), the Indiana Civil Rights Act, and Indiana University policy prohibit discrimination in educational programs against students with disabilities. Disabilities may include medical, auditory, visual, learning, psychological, mobility, or neurological problems. It is the policy of Indiana University to provide reasonable accommodations in a timely manner and on an individualized basis while maintaining institutional standards of performance. These accommodations are designed to counter the effects of disabilities where they may pose a barrier to the education process; they will not give the student an easy grade or an advantage over other students. See the Office of Disability Services for Students for accommodation and documentation.

Religious Observances: See here from more information on religious accommodation.

Sexual Harassment: As your instructor, one of my responsibilities is to help create a safe learning environment on our campus. Title IX and our own Sexual Misconduct policy prohibit sexual misconduct. If you have experienced sexual misconduct, or know someone who has, the University can help.

If you are seeking help and would like to speak to someone confidentially, you can make an appointment with:

The Sexual Assault Crisis Service (SACS) at 812-855-8900

Counseling and Psychological Services (CAPS) at 812-855-5711

Confidential Victim Advocates (CVA) at 812-856-2469 IU Health Center at 812-855-4011.

More information about available resources can be found here: http://stopsexualviolence.iu.edu/help/index.html.

It is also important that you know that federal regulations and University policy require me to promptly convey any information about potential sexual misconduct known to me to our campus' Deputy Title IX Coordinator or IU's Title IX Coordinator. In that event, they will work with a small number of others on campus to ensure that appropriate measures are taken and resources are made available to the student who may have been harmed. Protecting a student's privacy is of utmost concern, and all involved will only share information with those that need to know to ensure the University can respond and assist.

I encourage you to visit http://stopsexualviolence.iu.edu to learn more.

Support services for Students

Knowledge base and UITS support center:

For any technical support, see the Knowledge Base or go to the UITS Support Center website.

Schedule

Please see the following page.

Please note that this tentative syllabus/schedule may change without notice in order to reflect the needs of our classroom. See the course webpage on Canvas for updates.

Week	Date	Topic
1	8/26/19	Course Introduction – What is Python and geospatial data analysis?
1	8/28/19	Introduction to Python and Python Lists
2	9/2/19	**Labor Day – No Class**
2	9/4/19	Numerical Python (NumPy) and Introduction to simple functions for data analysis
3	9/9/19	Introduction to Matplotlib (data visualization in Python)
3	9/11/19	Introduction to Pandas (reading, creating, accessing and manipulating tabular data from multiple .txt, .dat or .csv files)
4	9/16/19	Fundamental Data Science Skills: Loops, Logic, Control Flow and Filtering
4	9/18/19	**In-class assessment**
5	9/23/19	In-depth data visualization with Matplotlib and Seaborn
5	9/25/19	Image (2D array) plotting with Matplotlib
6	9/30/19	NetCDF files for large geospatial climate data and NumPy masked arrays
6	10/2/19	Plotting geographic maps and more on dealing with missing values in an exploratory data analysis
7	10/7/19	Introduction to GDAL for reading and manipulating raster image data
7	10/9/19	Geospatial data change detection
8	10/14/19	Review: tidying, processing large volume geospatial data
8	10/16/19	**In-class assessment**
9	10/21/19	Statistical analyses of time series and image data with Seaborn
9	10/23/19	Statistical analyses of time series and image data with scikit-learn and Scipy
10	10/28/19	Geospatial pattern recognition: supervised learning/classification
10	10/30/19	Geospatial pattern recognition: supervised learning/classification
11	11/4/19	Clustering for dataset exploration
11	11/6/19	Dimension reduction for dataset exploration
12	11/11/19	Review statistical analyses on geospatial data and Final Project Planning
12	11/13/19	**In-class assessment**
13	11/18/19	Final Project
13	11/20/19	Final Project
14	11/25/19	Final Project *(Thanksgiving Break – No Class)*
14	11/27/19	Final Project *(Thanksgiving Break – No Class)*
15	12/2/19	Final Project
15	12/4/19	Final Project
16	12/9/19	Final Project
16	12/11/19	Final Project