

Python for Environmental Analysis and Visualization
Short Title: Big Data with Python
Syllabus
BC EESC 3050
Milbank 222
Monday, Wednesday: 10:10-11:25

Brian Mailloux

edblogs <http://edblogs.columbia.edu/eescx3050-001-2015-3/>

Office hours Sunday, Tuesday 8:30-9:15 PM (Might change). Via Zoom.

[Office Hour Link \(Personal Zoom Meeting Room\)](#)

appointment. <http://tinyurl.com/mailloux-calendar>

bmaillou@barnard.edu

212 854 7956

COURSE OBJECTIVES

Short Description for Catalog

Big Data is changing how we interact with and understand the environment. Yet analyzing Big Data requires new tools and methods. Students will learn to use the Jupyter Notebook programming environment with Python to analyze and visualize large environmental data sets. This will include both time series and spatial analyses.

Course description:

Big Data is changing how we interact with and understand the environment. Yet to analyze Big Data requires new tools and methods. Microsoft Excel is not sufficient to analyze big data and quickly becomes frustrating. We are going to analyze environmental data using Python in order to better understand the earth's system. The class will be a hands-on flipped classroom with programming occurring interactively during class and assignments designed to strengthen the methods and results of class.

The class will begin by analyzing climate data from Central Park and comparing it to different cities around the world. This analysis will enable us to better learn about climate and comparative statistics while at the same time learning the programming language Python.

The class will then utilize Python to analyze bacterial transport distances. We will learn how computer programs can quickly compute and store hundreds to millions of simulations to find optimal fitting parameters.

The class will analyze 50,000 arsenic measurements from Bangladesh. We will spatially analyze the data and learn how to sub-select large data sets and perform statistics while mapping the results.

The class will then visualize and analyze global scale data sets. We will download global scale data sets from the internet and use the data in our own analyses.

The class will then be introduced to time series data and learn how to organize a time series and then analyze a time series for trends.

Finally, the students will individually analyze a data set and present the results of the analysis. Students will learn how to clean and organize data and independently perform an analysis on the data set with the goal of testing a hypothesis or answering a specific question.

Student Learning Goals and Outcomes:

- To be able to organize and analyze large datasets.
- To be able to perform statistical analyses on large data sets
- To be able to write code in Python.
- To be able to download and analyze datasets from the internet
- To be able to plot and map large data sets.
- To develop a hypothesis and then be able to take any data set no matter how large or complicated and analyze and visualize it in order to test that hypothesis.

Lecture topics:

Week #	Date	Lecture #	Topic	Due
1	1/17	1	Class Introduction Central Park Example	
2	1/22	2	Print, Markdown, directories, lists	Code.org 1 hour Frozen Install Anaconda
	1/24	3	Make a list and plot it	Notebook 1
3	1/29	4	More lists, Start Strings	Notebook 2
	1/31	5	For loops	Notebook 3
4	2/5	6	If/elif/else Plot LGA data	All code.org due Notebook 4
	2/7	7	Fit a line. Fit to LGA and JFK data	Notebook 5
5	2/12	8	Numpy, fit a line. GDP data	Notebook 6
	2/14	9	Arrays and Colormaps	Notebook 7
6	2/19	10	Pandas (the Python library) and Arsenic data	Notebook 8. Read arsenic paper. Listen to Podcast
	2/21	11	More Pandas, Multiple plots	Notebook 9 All Sololearn Due
7	2/26	12	Correlations in Pandas	Notebook 10
	2/28	13	Senior thesis data? p-hacking	
8	3/4/24	14	Sediment Core Data on your own. Test a hypothesis	Read Chillrud Paper. Listen to Podcast
	3/6	15	Sediment Core Data on your own	Notebook 11
			Spring break	
9	3/18	16	CO ₂ Time Series	Turn in Final Topic Meet with Fellow I
	3/20	17	CO ₂ Time Series Part 2	Notebook 12
10	3/25	18	Mapping part 1	Turn in final topic Data
	3/27	19	Work on Data Set part 1	Notebook 13d

Week #	Date	Lecture #	Topic	Due
11	4/1	20	Mapping part 2	
	4/3	21	netCDF part 1	
12	4/8	22	netCDF part 2	
	4/10	23	Animated GIF part 1	Notebook 14d
13	4/15	24	Animated GIF part 2	
	4/17	25	Work on Data Set part 2	Notebook 15d
14	4/22	26	Work on Data Set part 3	
	4/24	27	Work on Data Set part 4	
15	4/29	28	Final Presentations	Final Projects Due

Website

We will be using both courseworks and edblogs <http://edblogs.columbia.edu/eescx3050-001-2015-3/>. Be prepared to use both. Generally we do

- Courseworks is for handing in homework.
- Edblogs is for all course information.

Class Structure: All classes will be in person in Milbank 222 unless things change. Students will come on time and take a packet and get started. After 10 minutes I will give an introduction and show some examples on the computer. Students will work through the packet and I walk around helping students as issues arise. You are encouraged to talk to your neighbor and work with them. I will walk around class and help people as needed. Sometimes I will give an answer, sometimes I might point something out, and sometimes I will ignore people and let them figure it out on their own. With about 5 minutes left in class we will ask 2 students to summarize what they learned. Once you are home I encourage you to do the packet a second time. This will save you time for the homework as all questions build off of the class packet.

Attendance: I will have a class sign in sheet to sign everyday. All students are expected to attend class during the class time period. You need to attend class to learn the material. However, we all need to stay safe. I trust my students. If you are healthy come to class. If you are sick stay home. You have 2 absences no questions asked. After that you need a reason from your Dean.

Late Work Policy: I am using a new late policy. Make sure you understand it and the implications. You have 4 days of late passes. You can use the 4 days as you see fit but you need to use a whole late day. Each fraction of a day gets rounded up to the next whole day. For example, 1 hour late is 1 day; 2.1 days late is 3. You can use your 4 days on any assignment and you can use multiple days on one assignment. But once an assignment is late the maximum score you can get on it is a 40. The last day you can hand in any assignment is the Tuesday following the last class. This semester it is Tuesday April 30 at 11:59pm. After that all missing assignments get a zero. Up to that time missing assignments can get a 40. To use a late pass put a comment in courseworks with the assignment. If you do not put in a comment I will assume you are not using a late pass. I will then make an assignment in courseworks where I track the number of late passes you have used. Courseworks labels how late an assignment is so it is easy

to track. If you are a minute late, courseworks will call it late, and I will count it late. So make sure to get your work in on time.

Class Seating: Talking with classmates can really help you learn. After the first day or two most people find a subgroup in the class to work with.

Videos

Required

Before many classes I will post a video on the blog and youtube. You are responsible for watching the videos before coming to class. In the videos I will say something that you will then need to answer on a courseworks quiz to prove you watched it.

Podcasts

Required

I have started making podcasts where I talk to authors about their work. There are two you need to listen to over the course of the semester.

Readings

Required

Each day you will download a packet on the day's work. These packets should be put into a binder and brought to every class. Or save them electronically in an organized fashion. They will become your book. You should add notes to them and use them. I use them all the time as I program. Many times I write in a hint or say you will need this point later. You will!!!!

Supplemental

Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython Paperback – October 29, 2012 - by Wes McKinney (Author)

Introduction to Computer Science Using Python: A Computational Problem-Solving Focus Paperback – December 25, 2012, by Charles Dierbach (Author)

Grading:

Online Tutorials 5%

Python Notebooks for each Section 70% (1 day notebook ~3% and a 2 day notebook is ~6%)

Final Presentation 15%

Class Attendance 5%

Quizzes 5%

CLASS PREMISE: This is a flipped classroom. You will learn some programming from code.org and sololearn. During class time we will work on in class projects that are designed to teach about different components of Environmental Science while at the same time teaching you how to analyze and visualize large data sets and how to program. At the end of each class or section you will have an iPython/Jupyter Notebook that teaches you multiple concepts. Then at the end of each class you will use this knowledge for your nightly homework. Many times it is quicker to review the day's notes before you begin the homework. This is worth repeating. Going through the class notes a second time before starting the homework usually saves you time! It is easy for me to tell when you didn't work through a notebook. All homeworks try to

build on concepts from that day. The Notebooks and scripts will contain all your notes, figures, results, etc. By running and examining the Notebooks and results we will be able to quickly determine how much of the in-class material you completed. You cannot rush through the class work because taking your time and learning it you learn the concepts needed for the homework and the next class. I am repeating this also. Do not rush through the class notebooks. Take your time and lean them. It all builds on itself.

CLASS PROJECTS:

WEB BASED TUTORIALS: In order to help you learn programming you will do web based tutorials during the first half of class. These tutorials don't always match class but will help you learn how to program. Code.org is designed for younger kids but really teaches for loops and if statements so it is really helpful. Students are always thankful at the end of the semester that I made them do all the tutorials. For the web based tutorials you will join the class group at each website and your progress will be tracked online by the professor. You have three due dates for the web based tutorials. Your grade will be the percent you complete by the due date. If you complete 100% of the tutorials you get 100%. So make sure you give yourself enough time and complete the tutorials. In the past some people have done the wrong tutorial. Make sure you are doing the correct tutorial and sign up for the correct class. **YOU WILL NOT GET CREDIT FOR DOING THE WRONG ONE.** Your three tutorials are

1. Code.org 1 hour with Frozen
2. Code.org 20 hour "Accelerated Course". (You do not need to do the offline parts).
3. Sololearn.

NIGHTLY PYTHON NOTEBOOKS: There will be a nightly homework based on each class. The homeworks will be posted on edblogs. If you diligently worked through the work in class you should be able to do the nightly homework. This makes coming to class a necessity as if you fall behind it will be difficult to accomplish all of the steps independently. The notebooks will consist of code, annotation, and results. The figures are presented inline in the notebooks. You need to add comments and markdown to your notebooks describing what you did. You are expected to have a figure caption and a description of the figure. Homework is due at the beginning of class time. I will use the time stamp on courseworks dropbox to determine if the homework is on time. I will post grading rubrics on courseworks/edblogs. You can see exactly what I am looking for. For the notebooks, if the notebook takes two classes it will be worth twice as much in the grading. If you do not add Markdown and notes you will lose points and the amount of points will get larger with each homework. For many homeworks I will make the final 10% of the grade harder to achieve and take more time. For example you can choose to not work on the last question and get a 90. This will be clearly stated on each homework.

COMMENTS: One of your homeworks will be to leave feedback on three different class days. You will get this homework later in the semester. I take the feedback very seriously and try to incorporate what you suggest. So please provide lots of comments!

STACKOVERFLOW: Part of learning to program is learning to find answers on your own. This might be through trial and error, asking a friend or looking online. When learning something new I will watch a youtube video or read a tutorial. But for programming questions

stackoverflow is a commonly used website. People post questions and answers. It can be really helpful for figuring out bugs and issues and doing something new. Unfortunately people can also be mean. You will need to use it over the course of the semester. I can't answer all your questions and you need to learn how to learn on your own. Make a login and start liking your favorite answers!

FINAL POSTER SESSION: Each student is going to create a poster presentation from a data set. Each student will analyze their own data set. The student will develop questions or hypotheses about the data set and will use what we learned throughout the semester to tackle their problem. We will have 3-4 days of classes near the end of the semester devoted to the final project. Students will present their results during the finals time in a virtual poster session. You can analyze a dataset that you are interested in and find or I can supply you with a dataset. You cannot just repeat an analysis you did for another class. Many people have used this section to analyze their senior thesis data more in depth. Final Posters will be due before the final class. There will be a separate page describing the process and the due dates. These due dates are **very strict** as everyone needs to present during the final session. The three due dates on the syllabus are:

1. Your poster topic.
2. Your poster data
3. Your final poster.

HANDING IN HOMEWORK: In order to do well in class you need to hand in each homework and get them in on time. When handing in homework you will post your notebook on dropbox in courseworks. Courseworks automatically names the files so that is not a worry. But you need to hand in the python notebook. Not an image and not a pdf. After the first homework handing in the wrong thing will make it count as late. For some assignments you also need to hand in extra files you create. Do not get behind on homeworks. There are a lot and it is hard to catch up.

Class Notes:

- No headphones in class. Make sure you are listening to the class and questions. It will help you in the long run as people ask questions. You might even know the answer!
- You may be tempted to come late or leave early. Do the notebooks. Students who take their time and do the notebooks do really well.
- You need to be in class coding in order to learn.
- Zoom video and backgrounds: When you come to office hours it is tempting to mute zoom and to not show the video. We are going to share screens and chat. If you come to office hours participate!
- Technical glitches: Something is going to go wrong. My internet can get slow and go out. My computer might crash. We will make it work.

OFFICE HOURS: I do office hours via zoom the night before classes. This has seemed to work well as it is when students are working on the homework. Students can jump on and off zoom when working on homework and share their screen. This way students can hear all the conversations. Plus other students can also jump in and help. Screen sharing is our friend.

Collaboration and Cheating:

When working on coding and projects it is easy to run into roadblocks and stumble. This is when talking to a classmate is very helpful. You can ask how you tackled a problem. How did you iterate over the data? What type of if statements did you use? I strongly urge you to work with and talk to your classmates. The “classroom” should have people speaking. **However, you cannot copy a classmate’s code.** Since everyone’s comments and codes are slightly different it is easy to observe when someone has copied and pasted and this is not allowed and will be considered plagiarism. If I notice two identical notebooks you both get a zero the first time. If it happens a second time you will be reported to the appropriate Dean. So talk and help each other but do your own work.

ChatGPT: The goal of this class is to learn how to code. If you use ChatGPT you will never learn.

Email:

I try to answer emails as fast as possible. But do not expect rapid replies late at night or on weekends. Also, if a person emails me a good question, I usually do not reply directly but send an email to the whole class through Courseworks with the question and answer.

Courseworks

Courseworks will be utilized to post all files and readings from class.

Course Requirements

If you miss something you need a note from your class Dean in order to reschedule the exam.

Definition of Grades: All grades will be based on a scale of 100 with

A+		Rare performance. Reserved for exceptional achievement
A	93.00-100	Excellent work. Outstanding achievement.
A-	90.00-92.99	Excellent work that exceeds course expectations.
B+	87.00-89.99	Very good work. Solid achievement (expected of Barnard/Columbia undergraduates) that meets all course expectations.
B	83.00-86.99	Good work. Acceptable achievement that meets almost all course expectations.
B-	80.00-82.99	Satisfactory work. Acceptable achievement that meets major course expectations.
C+	77.00-79.99	Fair achievement just above that which is minimally acceptable
C	73.00-76.99	Fair achievement but only minimally acceptable.
C-	70.00-72.99	Barely acceptable achievement.
D	60.00-69.99	Very low performance. Unsatisfactory work. Lowest achievement to still allow for a passing grade. This grade may not be counted toward the major or minor option

F	59.99 or less	Failure.
---	---------------	----------

Consult the Barnard or Columbia College Catalogs or the Registrar's Office for other information about grading, including: the definitions of other letter grades, pass/D/fail option, incompletes, and calculation of GPA.

The Honor Code and Academic Integrity

The Barnard Honor Code (established in 1912) reads:

We, the students of Barnard College, do hereby resolve to uphold the honor of the College by refraining from every form of dishonesty in our academic life. We consider it dishonest to ask for, give, or receive help in examinations, quizzes, or to use in them any papers or books in any manner not authorized by the instructor, or to present oral or written work that is not entirely our own, except in such way as may be approved by the instructor. We pledge to do all that is in our power to create a spirit of honesty and honor for its own sake.

The Honor Code governs all aspects of academic work. If a violation should arise, it will be reported to the Dean of Studies for appropriate action. Honor Board Guidelines on the procedures for implementing the Honor System and acting on charges of dishonesty can be found in the Student Handbook. Remember that Barnard students reaffirm their acceptance of the Honor Code by signing their registration form. Columbia students commit themselves to the Honor Code upon registering for a Barnard course. Because the Honor Code is not entirely specific and contains qualifications and exceptions, such as "authorized by the instructor" or "approved by the instructor", please read the attached summary of "What behaviors constitute academic dishonesty?"

Students with disabilities

Students who may need disability-related classroom accommodations are encouraged to make an appointment to see me as soon as possible. Disabled students who need test or classroom accommodations must be registered in advance with the Office of Disability Services (ODS) in 105 Hewitt. Thank you.

OFFICE HOURS ZOOM

Brian Mailloux is inviting you to a scheduled Zoom meeting.

Topic: Brian Mailloux's Personal Meeting Room

Join Zoom Meeting

<https://columbiauniversity.zoom.us/j/6132274079?pwd=MnBMWUs0bCsvVVNCNGIPa3Y0a2U0Zz09>

Meeting ID: 613 227 4079

Passcode: 251126

One tap mobile

+16468769923,,6132274079#,,,,*251126# US (New York)

+13126266799,,6132274079#,,,,*251126# US (Chicago)

Dial by your location

+1 646 876 9923 US (New York)

+1 312 626 6799 US (Chicago)

+1 301 715 8592 US (Washington DC)

+1 346 248 7799 US (Houston)

+1 669 900 6833 US (San Jose)

+1 253 215 8782 US (Tacoma)

Meeting ID: 613 227 4079

Passcode: 251126

Find your local number: <https://columbiauniversity.zoom.us/j/6132274079?pwd=MnBMWUs0bCsvVVNCNGIPa3Y0a2U0Zz09>

Join by SIP

6132274079@zoomcrc.com

Join by H.323

162.255.37.11 (US West)

162.255.36.11 (US East)

221.122.88.195 (China)

115.114.131.7 (India Mumbai)

115.114.115.7 (India Hyderabad)

213.19.144.110 (Amsterdam Netherlands)

213.244.140.110 (Germany)

103.122.166.55 (Australia Sydney)

103.122.167.55 (Australia Melbourne)

209.9.211.110 (Hong Kong SAR)

64.211.144.160 (Brazil)

69.174.57.160 (Canada Toronto)

65.39.152.160 (Canada Vancouver)

207.226.132.110 (Japan Tokyo)

149.137.24.110 (Japan Osaka)

Meeting ID: 613 227 4079
Passcode: 251126