# OCN 760: Data Analysis with Python

Prof. Brian Powell

powellb@hawaii.edu

Class Time: MW 10:30–11:45a

## COVID-19

Due to the ongoing pandemic, the course this semester will be entirely online. No student will need to be physically present for this course during the semester. You will be expected to attend each online lecture and download and submit any required assignments via laulima.

## Introduction

The purpose of this course is to introduce you to the concepts of scientific programming and data analysis using the Python programming language. This is a very broad topic that cannot be covered adequately in a single semester course, but we will cover what I hope will the material you need to start you on the way to becoming a proficient scientist and working with large data sets.

We will be learning and using the Python language. It is widely used in almost all applications of computing (*e.g.*, Facebook, Google, Dropbox, DoorDash, etc. are major users of Python), and it is now the standard language in data science, machine learning, physical sciences (high-energy physics, chemistry, atmospheric, geology, and ocean physics). High profile journals, such as *Nature* encourage papers to include Python Notebooks with their submissions for both reviewers and readers alike to verify your work. Do you want to verify the LIGO gravitational wave discovery? It's all here in Python. Python, is often ranked one or two the "most popular" programming language in use. As such, Python has an vast suite of libraries available for whatever your needs.

Python is *libre* software meaning that it is free (as in no-cost) and it is also free (as in you can do whatever you like to do with the source code). This is key to science. The other major language used primarily in pure statistics and biology is the R programming language, which is began as a *libre* clone of the commercial S programming language that is part of the SASS product. We will not cover any aspects of that language; however, with high-level languages, the concepts are consistent across them, primarily the syntax varies.

This course provides an introduction to a variety of concepts and techniques used in interpreting observations and numerical model output, with an emphasis on time series. In most cases, applying these techniques requires computing; hence, at least minimal programming.

If you will need to do some programming as a student and/or as a professional, a moderate amount of time spent learning good tools and practices at an early stage will pay off ever after. By working in

Python with both simulated and real data sets, we will try to develop computing skill along with an understanding of basic statistics and data analysis techniques. Along the way we will introduce tools of working with your computer, including: command line interfaces, tools such as git, a Distributed Version Control Systems, and more.

## Administrative

The class meets MW 10:30–11:45am HST in Zoom.

### Online Tools and Communication

As mentioned, lectures and recitations will be conducted via Zoom. The nature of this course is to be open and facilitate discussions, so please leave your mic and video active so we can discuss while the class is going along. All work will be submitted electronically and there is a forum in laulima to post messages to the class.

## Textbook and Resources

There are two free books that we will use material from during the semester:

- Automate the Boring Stuff with Python
  This book provides a nice level introduction to the Python language with many examples of practical things that you can program.
- Python Data Science Handbook
  This book is for those with Python familiarity and introduces various topics in data science and how to accomplish them.

## Grading

**Participation 10%** This class is rather interactive with discussions and contributes from every student. Although it will be less social due to being held online, we will try to have a lively group.

**Assignments 40%** Assignments will be given from time to time with exercises for you to work on. We will discuss each assignment when they are given.

**Project 50%** Most of your grade will be based on some analysis you choose to do on a dataset that is of interest to you. As most of you are graduate students, then you should have a data set or subject that is already of interest to you. We will discuss the project throughout the semester.

## Assignments

A number of assignments will be given this semester via laulima. To submit your solution, you will be able to upload your file submission via the Assignment tab in laulima.

## Calendar

Class will not be held on the following days:
18 Jan, 15 Feb, 15 and 17 March.

## Computing

As this is a computing class, it is expected that each student will have a modern computer for their use that they can install software and interact with. If anyone does not have a computer, please notify me, and we can try to find an alternative.

## Course Topics

The following topics will be covered this semester. Due to the online nature of this semester, we don't know the date of each lecture and exam, we will see how fast we get through the material.

1. Introduction and Review of Computing
    - Command-Line: file structure; commands
    - Computer Systems: packaging and installing software
    - Collaborating: git, bundles
    - Installing Python: native, anaconda, miniconda
    - Computing: how do computers work, memory, numerics, files, types
2. Python Topics
    - Python Packages: pip, conda, git
    - Python Interaction: REPL, IPython, Jupyter, Scripts
    - Programming in Python
        - Data Types: numbers, strings, tuples, lists, dictionaries
        - Variables
        - Conditionals
        - Flow Control: loops, exceptions
        - Functions
        - Classes
        - Scientific Computing: numpy, matplotlib, sympy, scipy, pandas
        - Loading Data
3. Data Analysis
    - Time-Series analysis
        - Basic stats: probabilities, correlation, spectral

    - Regression: least-squares, multi-variate
  - Fields: empirical orthogonal functions
  - Advanced Methods:
    - Bayesian Analysis
    - Machine Learning: simple neural network estimators

## Student/Instructor Contract

By remaining enrolled in this online class, you agree to:

- Be responsible for keeping on-time with the course and attending lectures;
- Do all assignments (unless stated otherwise) on your own to be prepared for the exams;
- Finish all lessons (including the assignments and quizzes) by the deadline;
- Not request time extensions except in cases of emergency;
- Abide by the rules for each exam;
- Check the announcements in laulima to stay up-to-date on the class and expectations;
- Abide by the University of Hawai'i Student Conduct Code.

The instructor agrees to:

- Respond within a day to inquires;
- Be excited and passionate about the course materials;
- Grade exercises and exams in a timely manner;
- Adjust course materials and schedules as needed by the COVID-19 and online situation;
- Be open to feedback on the course delivery and content.

## University Policies and Procedures

The University of Hawai'i is an equal opportunity/affirmative action institution. It is committed to a policy of nondiscrimination on the basis of race, sex, victims of domestic or sexual violence, gender identity and expression, age, religion, color, national origin, ancestry, citizenship, disability, genetic information, marital status, breastfeeding, income assignment for child support, arrest and court record (except as permissible under State law), sexual orientation, national guard absence, or status as a covered veteran. For additional details, visit the UH Systemwide Policies and Procedures Information System (PPIS) site.

## Student Conduct

It is a privilege to be a member of the University of Hawai'i at Mānoa community. This privilege provides students with the opportunity to learn and to participate in the many programs that are offered on campus. Along with this privilege, students are expected to be responsible in relationships with others and to respect the interests of the institution. These interests are fully set forth in the UH Systemwide Student Conduct Code (EP 7.208).

Faculty members are encouraged to respond to behaviors which are disruptive to the academic environment. Students may be referred to the Office of Judicial Affairs for possible disciplinary action including suspension, dismissal or expulsion and/or the Department of Public Safety summoned in serious cases of disruptive behavior.

## Academic Honesty

Cheating, plagiarism, or other forms of academic dishonesty are not permitted within this course and are prohibited within the Systemwide Student Conduct Code (EP 7.208). Examples include: fabrication, facilitation, cheating, plagiarism, and use of improper materials. Any incident of suspected academic dishonesty will be reported to the Office of Judicial Affairs for review and possible adjudication. Additionally, the instructor may take action in regards to the grade for the deliverable or course as they see fit.

Acts of dishonesty, including but not limited to the following:

1. Cheating is an act of academic dishonesty and includes, but is not limited to:
    - use of any unauthorized assistance in taking quizzes, tests, or examinations;
    - use of sources beyond those authorized by the instructor in writing papers, preparing reports, solving problems, or carrying out other assignments;
    - the acquisition, without permission, of tests or other academic material belonging to a member of the UH faculty, staff or student body; and
    - engaging in any behavior specifically prohibited by a faculty member in the course syllabus or class discussion.
2. Plagiarism is also an act of academic dishonesty and includes, but is not limited to:
    - the use, by paraphrase or direct quotation, of the published or unpublished work of another person without full and clear acknowledgement.
    - It also includes the unacknowledged use of materials prepared by another person or agency engaged in the selling of term papers or other academic materials.
    - Furnishing false information to any UH official, faculty member, or office.
3. Forgery, alteration, or misuse of any UH document, record, or form of identification.

## UH Title IX

The University of Hawai'i is committed to providing a learning, working and living environment that promotes personal integrity, civility, and mutual respect and is free of all forms of sex discrimination and gender-based violence, including sexual assault, sexual harassment, gender-based harassment, domestic violence, dating violence, and stalking (UH Title IX). If you or someone you know is experiencing any of these, the University has staff and resources on your campus to support and assist you.

If you would like to report incidents of sex discrimination or gender based violence, contact your campus Title IX Coordinator or submit the online reporting form. If you wish to remain ANONYMOUS, speak with someone CONFIDENTIALLY, or would like to receive information and support in a CONFIDENTIAL setting, contact your campus' confidential resource.

## Accommodation Statement

The University of Hawai'i is committed to a barrier-free campus and provides accommodations to ensure students with disabilities equal access to education. We agree to make academic adjustments to ensure non-discrimination of students with disabilities. This commitment is in accordance with applicable state and federal laws, including the Americans with Disabilities Act, and Sections 504 and 508 of the Rehabilitation Act.

Under the Americans with Disabilities Act (Title II) and the Rehabilitation act of 1973-section 504 and 508, individuals with disabilities have protections against discrimination and are assured access to programs, services and activities. For more information see "Americans with Disabilities Act" and "Rehabilitation Act of 1973 – Section 504 or Section 508".

Students must self-identify to UH Mānoa KOKUA and complete the intake process before receiving reasonable accommodations. To ensure the prompt and effective provision of accommodations, students should contact the Disability Services Office as early as possible. KOKUA can also be reached at (808) 956-7511 or (808) 956-7612, email: kokua@hawaii.edu.

## Student Support

Academic Advising Bookstore Career Services Counseling Services Registration

- Academic Advising
- Bookstore
- Career Services
- Counseling Services
- Registration

## Financial Aid Statement

If students do not begin attendance in a course or stop participating in a course, Title IV funds must be returned according to Federal Return of Title IV funds regulations (34 CFR 668.21(a)). This means you may be required to return some (or all) of the financial aid you have received. It is very important

to remember that colleges are required to take steps necessary to ensure that students are academically engaged in order to justify the disbursement of Federal Title IV student aid funds. If at any time your plans change and you no longer plan to participate in the courses in which you enrolled, you must contact the financial aid office to minimize any possible negative financial impact.

For more information on financial assistance for your education, please contact Manoa Financial Aid Office. Financial assistance may include grants, scholarships, and other resources to help you pay for the cost of college. A financial aid adviser will be able to help you navigate this process to determine your eligibility for these funds.