Syllabus

BIOF309 - Introduction to Python Programming

Fall 2017

Section 1: Monday 5:30PM - 7:30PM

Section 2: Monday 7:30PM - 9:30PM

This document is subject to revision!

Changes are tracked using the git version control system.

Instructors:

- Martin Skarzynski - marskar at gmail, and also GitHub, Slack, DataCamp etc.
- Benjamin Cohen - benjapaulcohen at gmail

Communication

Before contacting us, please check to see if your question has already been answered:

https://stackoverflow.com/help/how-to-ask

In general, please use the class Slack team to communicate with classmates and instructors. Please use email only in case of personal and/or private question/concerns. If you have a course-relevant question or something to share, Slack is simply the better way to communicate.

First class: September 11, 2017

Final class: December 18, 2017

Course Description

This course is designed for non-programmers, biologists, or those without specific knowledge of Python to learn how to program. Week by week we will slowly build up your skills and understanding of programming and the Python language. There will be in class demonstrations, using Jupyter Notebooks, activities to be completed outside of class, mostly using DataCamp, and homework for you to practice and learn at your own pace.

Learning Objectives

By the end of this course you should be able to:

1. Look at a task and determine if you can or should automate it
2. Create working Python programs
3. Understand the difference between Python object types (i.e. lists, dicts)
4. Perform data analysis and visualization with Python
5. Use git for version control and collaboration
6. Demonstrate your Python skills with a project

Logistics

This is a one-semester course starting on the 11th of September 2017 and finishing on 18th of December 2017.

Class Location: Rathskeller, Building 60, NIH Bethesda campus

Attendance in class is strongly recommended; however, we realize other commitments will occasionally prevent attendance. If you miss a class, please review the materials available at the course github repository and keep up with activities and homework.

You may only attend the section for which you signed up. Do **NOT** attend other sections **UNLESS** given permission to do so by the instructor.

Most classes will have hands-on tutorials and assignments. Both practice and graded assignments will generally be provided. Graded assignments should be submitted prior to the following class. So that you can follow along during class, bringing a laptop to each class is strongly encouraged.

Important dates:

- September 11 : Class starts
- September 11 - September 29 : Late Registration ($10.00 late registration fee per course applies)
- October 6 : Last day to drop/withdraw
- November 10 : Last day to change status (credit or audit)
- December 18 : Class ends

Required Materials

Each student is encouraged to bring their own laptop to each class.

Programing without a computer would be an exceptional feat.

For the course, we will use Python 3. Any Python installation should work, but you must be able to install packages. The Anaconda Scientific Python Distribution from Continuum Analytics will likely be the easiest approach to configuring Python, if you do not already have Python installed. The Anaconda installer will automatically install many of the packages we will use during the course, including Jupyter Notebooks. For more information, see the guidelines on the The Hitchhiker's Guide to Python.

In addition, we will use [Git](#) and [GitHub](#) for version control and collaboration.

All of the course materials are available on GitHub. Before accessing the [course materials repo](#), you should that

1. it is likely to be be under constant development throughout the semester and
2. you are not expected to work through every single Jupyter Notebook contained therein!

We will discuss the most interesting examples during class and point out the others to be reviewed outside of class.

Most homework assignments will be managed by [GitHub Classroom](#). The list of assignments is as follows:

1. [Hello World!](#)
2. [GC Percent](#)
3. [Loop Example](#)
4. [Conditionals Example](#)
5. [Cubed Program](#)
6. [NumPy Problem Set](#)
7. [Pandas Problem Set](#)
8. [Make A Plot](#)
9. [Machine Learning Example](#)
10. [Final Project](#)

Last but certainly not least, the very nice folks at [DataCamp](#) have given us access to their awesome teaching materials.

The DataCamp assignments are:

1. [Python Basics](#)
2. [Manipulating Files and Directories](#)
3. [Python Lists](#)
4. [Loops](#)
5. [Logic, Control Flow and Filtering](#)
6. [Functions and Packages](#)

Optional Materials

[GitHub](#) is offering some free awesome resources to students, that might be of interest to you, depending on your background: [GitHub student pack](#).

In particular, the [Atom](#) text editor is noteworthy because of its [integration with Git and GitHub](#), and ability to offer [Jupyter Notebook](#)-style functionality through packages like [Hydrogen](#) and [jupyter-notebook](#).

You may, however, use other another text editor, in you prefer.

Recommended Books

There is no required textbook for this course.

We do, however, highly recommend [Python for Data Science](#) and its companion text [A Whirlwind Tour of Python](#). Both books are available free on GitHub in Jupyter Notebook form. For maximum enjoyment, consider reading the relevant chapters before coming to class.

We will link to relevant online resources throughout the course.

If you would like additional material on the basics, the following resources may be useful:

- [Python for Biologists](#) by Martin Jones; an archived PDF may be found in this repo in the extras folder under the name p4b.pdf.
- [Learn Python the Hard Way (ebook freely available from the author)](#) by Zed A. Shaw; a [video course](#) is also available.
- [Think Python (ebook freely available from the author)](#) by Allen B. Downey.
- [Python for Everybody: Exploring Data in Python 3 (ebook freely available from the author)](#) by Charles Severance
- [Python Cookbook](#) by David Beazy For more information about Python, please see the official [Python Software Foundation website](#).

Assignments and Grading

The emphasis of the course is on learning and mastering the skills covered. We hope that everyone will be able to complete the assignments and project. If some of the material appears unclear, please ask for clarification.

Assignments will be uploaded to the GitHub classroom setup for the course.

Grading **assignments** will be done using the following rubric:

- Program runs, produces correct result, contains useful comments, meaningful variable names, follows coding conventions: A+
- Program runs, produces correct result, contains useful comments: A
- Program runs, produces something close to the correct result: B
- Program runs, does not produce correct result: C
- Program does not run: Incomplete (I)

Grading the **final project** will be done using the following rubric:

- Project description / Specification

  - Goals unclear, difficulty demonstrating functionality (1-3)
  - Goals for the project and functionality are discussed but difficult to follow (4-6)
  - Goals for the project and functionality are discussed (7-9)

- Goals for the project and functionality are logically presented and clearly communicated (10-12)

- Documentation

  - Only comments embedded in the code (1-3)
  - Objects and methods have docstrings (4-6)
  - Objects and methods have docstrings, additional standalone documentation (7-9)
  - Objects and methods have docstrings, extensive standalone documentation with example usage (10-12)

- Readability

  - The code is poorly organized and very difficult to read (1-3)
  - The code is readable, but challenging to understand (4-6)
  - The code is fairly easy to read (7-9)
  - The code is well organized and very easy to read (10-12)

- Reusability

  - The code is not organized for reusability (1-3)
  - Some parts of the code could be reused (4-6)
  - Most of the code could be reused (7-9)
  - Each part of the code, and the whole, could be reused (10-12)

- Performance

  - Program does not run (1-6)
  - Program runs, but does not produce correct output (7-12)
  - Program runs, produces correct output under most conditions (13-18)
  - Program runs, produces correct output with robust error checking (19-24)

Course Materials

Course materials are available from the course [GitHub repository](#).

Schedule

Week 1 (Sept 11 2017):

- [Intro survey](#)
- Course overview
- An introduction to programming
- Why Python?
- What can you do with Python?
- Troubleshooting software installation
- Introduction to Jupyter Notebook
- Reading: Chapters 01-06 in [Whirlwind Tour of Python](#)
- Reading: Chapters 01.01-01.04 in [Python Data Science Handbook](#)
- Activity: [DataCamp - Python Basics](#)
- Homework:
    - Find a cool notebook to share with the class!
    - Add a link to the notebook to the "cool-notebooks" channel on our [class Slack team](#)
    - Prepare a 1-minute presentation (no powerpoint!) to introduce yourself and your chosen notebook next class

Week 2 (Sept 18 2017):

- Introduction to UNIX/bash/shell
- IPython shell and shell-related commands
- Review Homework: Cool notebook presentations (1 minute each)
- Discuss the "Hello world" interactive program
- Reading: Chapter 01.05 in [Python Data Science Handbook](#)
- Reading: Lessons 01-03 in [Software Carpentry - Shell](#)
- Activity: [DataCamp - Managing Files and Directories](#)

- Homework: Prepare the "Hello world" interactive program

Week 3 (Sept 25 2017):

- Review homework, debugging
- Version control with Git
- Reading: Lessons 01-07 in Software Carpentry - Git
- Activity: Try Git: Git Tutorial
- Homework:
    - Create a GitHub repo
    - Push your "Hello world" program to this repo

Week 4 (Oct 2 2017):

- Intro and review
- Review Homework
- Data Structures: Lists, Sets, and Tuples
- System arguments and how to use sys.argv
- Reading: Chapter 07 in Whirlwind Tour of Python
- Activity: DataCamp - Python Lists
- Homework:
    - Write a program to calculate GC percentage
    - Push this program to GitHub

Week 5 (Oct 9 2017):

Columbus day - No class

We will make up this class on December 19th.

Week 6 (Oct 16 2017):

- Intro and review
- Review Homework
- Loops
- Reading: Chapter 08 in [Whirlwind Tour of Python](#)
- Activity: [DataCamp - Python Loops](#)
- Homework: Prepare an example of how you might use a loop

Week 7 (Oct 23 2017):

- Intro and review
- Review Homework
- Conditional tests
- Reading: Chapter 08 in [Whirlwind Tour of Python](#)
- Activity: [DataCamp - Python Logic and Control Flow](#)
- Homework: Prepare an example of how you might use conditionals

Week 8 (Oct 30 2017):

- Intro and review
- Review Homework
- Functions
- Reading: Chapter 09 in [Whirlwind Tour of Python](#)
- Activity: [DataCamp - Functions and Packages](#)
- Homework:
    - Write the "Cubed" Program and push it to GitHub
    - Submit your project proposal via GitHub

Week 9 (Nov 6 2017):

- Intro and review
- Review Homework

- NumPy
- Reading: Chapter 02 in [Python Data Science Handbook](#)
- Activity: [DataCamp - NumPy](#)
- Homework: Numpy practice set

Week 10 (Nov 13 2017):

- Intro and review
- Review Homework
- Data analysis with pandas
- Reading: Chapter 03 in [Python Data Science Handbook](#)
- Activity: [DataCamp - Dictionaries and Pandas](#)
- Homework: Pandas practice set

Week 11 (Nov 20 2017):

- Intro and review
- Review Homework
- Data Visualization
- Reading: Chapter 04 in [Python Data Science Handbook](#)
- Activity: [DataCamp - Matplotlib](#)
- Homework: Prepare a plot using Matplotlib to share with the class

Week 12 (Nov 27 2017):

- Intro and review
- Review Homework
- Machine Learning
- Reading: Chapter 05 in [Python Data Science Handbook](#)
- Activity: [DataCamp - Titanic 1](#)
- Homework:

- Submit project milestone report
  - Prepare an example of how you might use machine learning

Week 13 (Dec 4 2017):

- Intro and review
- Review Homework
- Biopython
- Student-selected topics
- Reading: Chapter 00-02 in [Biopython-Notebook](#)
- Activity: [DataCamp - Titanic 2](#)
- Homework: Work on Final Project/Presentation

Week 14 (Dec 11 2017):

- Project presentations (Presentation slots will be randomly assigned)
- Activity: [DataCamp - Titanic 3](#)
- Homework: Work on Final Project/Presentation

Week 15 (Dec 18 2017):

- Project presentations (Presentations slots will be randomly assigned)