

CSB1021HS, LEC0140

Introduction to Python

Offered by the *Centre for the Analysis of Genome Evolution & Function (CAGEF)*

Instructors:

Dr. David S. Guttman, CSB, CAGEF
Dr. Calvin Mok, CAGEF Bioinformatics

david.guttman@utoronto.ca
calvin.mok@utoronto.ca

Date

January 6 – Feb 17 (7 weeks)
Tuesdays, 10:00am-1:00pm
ESC3087

Enrollment:

16 graduate students
Auditor spaces based on availability

Course Objectives

This is a beginner's introduction to Python for data science applications. The course is intended for students with no computer science background who want to develop the skills needed to analyze their own data. Students who complete this course will be able to:

- Perform data analysis in Python using the Jupyter Notebook environment.
- Understand Python data structures and data types.
- Manipulate Python objects such as lists, data frames, and dictionaries.
- Import data into Python and transform 'messy' datasets into 'tidy' datasets.
- Use flow control to develop branching code.
- Use regular expression and string manipulation to explore and clean data.
- Make exploratory plots.

Throughout the course we'll work with a set of data that takes us through the various steps of analysis from importing to data wrangling to visualization. Each class will consist of a short introductory section followed by 'code-along' hands-on learning that will gradually build up the lecture's topic(s). Students are expected to have access to a computer during class and are encouraged to ask questions while coding along with the instructor. A homework assessment will be assigned after each class to reinforce the skills learned and a final project will test overall knowledge and application. The course will be provided through Quercus and lectures will be held in-person.

Course Availability

This course will be held in-person (unless otherwise determined) and will be available to graduate students in CSB and EEB. Auditor spaces will be based upon available space to postdocs, staff, and faculty, although only registered students will be evaluated. The course will count as a single module (0.25 credits) for CSB graduate students. All graduate students interested in taking the course for credit should enroll through ACORN.

Anyone wishing to audit the course should fill out the request form at

<https://forms.gle/XWnGtgbwvJFdwQ5RA>

Evaluation

Item	Note	% Mark
Completed Jupyter Notebook	7 lectures x 2% each*	14%
Homework Assignments	7 weekly assignments x ~5% each	36%
Term project	Due 2 weeks after the end of the course	50%

* a 3.5% bonus (0.5% per lecture) will be awarded for submitting notebooks within 24 hours of lecture completion.

Pre-requisites: Access to a computer and internet. No prior programming experience needed.

Reference Material: 2016. Severance, Charles. Python for Everybody: Exploring Data Using Python 3. http://do1.dr-chuck.com/pythonlearn/EN_us/pythonlearn.pdf

Course Tools: University of Toronto Jupyter Hub, DataCamp, Zoom.

Syllabus

Class	Topic
1	Intro to Python and Jupyter Notebooks: Basics about Python, using Jupyter Notebooks, how to run Python code, as well as an introduction to Python variables, functions, modules, best coding practices, data types, missing data, code debugging and getting help.
2	Python data structures, Numpy and Pandas: List, Dictionaries, Tuples, Sets, Series, mathematical operations with Python objects, Introduction to NumPy and Pandas.
3	How to Read, Write, and Manipulate Your Data: The wide and long formats, reading in data, data wrangling with Pandas, and writing data.
4	Data visualization with seaborn: The grammar of graphics; scatter, line, box, bar, and density plots, among other types of graphics.
5	Flow control: Flow control, for loops, Conditionals
6	Regular Expressions: Classes, quantifiers, operators, pattern-matching, String manipulation.
7	User-defined functions: Defining a function, best practices in user-defined functions, and web scraping

Subject to change