Tools For Analytics
Instructor: Paul Logston paul.logston@columbia.edu
Office Hours: TBA, Mudd 318
CAs:TBA
Office Hours: TBA

Description
The goal of this course is to introduce students to the basics of programming in Python and the tools within the programmer’s ecosystem. By the end of the course, students will have a working knowledge of the following:

- Python
- Data Analysis tools in Python (numpy, pandas, bokeh)
- Git
- Bash
- SQL
- VIM
- Linux / Debian
- SSH

Prerequisites
There are no prerequisites for this class.

Computers in Class
Computers are a requirement for this course and you are expected to bring one for every class. We’ll do a lot of programming and the best way to learn is to see something in action. Python is an especially good language for making things happen! Make sure that your laptops have sufficient charge for the class!

Students will be given access to a virtual machine (VM) on which to do all their programming. Once they receive their coupon code, students should follow the setup instructions found here: https://github.com/logston/py-for-or/ (Links to an external site.)

Texts + Other Resources
There is NO textbook for this class. The following will be helpful if you want to go above and beyond the material covered in the course:

- Google’s Python Class: https://developers.google.com/edu/python/?csw=1 (Links to an external site.)
- Python Essential Reference by David M Beazley
Evaluation and Learning Components
(25%) We will have three quizzes, all in class.
(40%) Homework. There will a homework set after each class session except for the last. The homework is due before the following session.
(35%) There will be one final exam. The exam will include topics from the entire course.

** Student's must bring grading discrepancies to a CA or the instructor within 1 week of getting their grade and the solution set for an assignment **

Rough Topic Schedule - Weeks & Topics:
1. Intro, Linux, Bash, & VIM
2. More Bash, SSH, GIT, & GitHub
3. SQL
4. Basics Syntax and Keywords of Python
5. Data Structures, Control Flow, & Functions
6. Errors, Generators, & OOP
7. Web Pages, Web Requests, & Parsing
8. APIs, JSON, XML, & Storing Results
9. SQL + Python
10. Numpy & Pandas
11. Graphing and Publishing to GitHub Pages
12. Basic Machine Learning With scikit-learn
13. TBA
14. Final