Course Description

The goal of this course is to introduce you to the basics of programming in Python and to give you a working knowledge of how to use Python to extract knowledge and information from data. At the end of this course you can expect to reasonably competent in Python, be familiar with working with relational databases such as MySql and how to integrate Python with MySql, and be aware of the Python libraries that you can use to work with and analyze data from databases as well as from the web.

This is an intensive hands on course so be prepared for a lot of work and a significant time commitment. But your reward - proficiency in data management skills - will be substantial.

Prerequisites

There is no prerequisite for this class. Prior exposure to some programming language is helpful but not necessary. I encourage you to explore online Python programming courses before the start of the semester (bearing in mind that we’ll be working with Python 3.5 and not 2.7).

Topics:

1. Programming in Python (approximately 4 weeks)
2. Getting data from the Internet (APIs/Web Scraping)
3. Database basics/MySQL
4. Data analysis and data visualization
5. Real time reporting (Web Apps)
6. Text mining
7. Analyzing social networks
8. Machine learning libraries in Python
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 - the best way to learn is to see something in action and Python is an especially good language for making things happen. Make sure that your laptops have sufficient charge for the class!

**Mac vs Windows**: Either is fine but, if you have the choice, then please use a Mac. It is much easier to install needed libraries on a Mac than it is on a Windows machine. In particular, if you have a Mac and are using some sort of Windows emulator then please use Mac OS-X and not the Windows emulator. The double redirection will make everything a lot slower. But, either Mac or Windows will work so don’t worry if you’re a Windows user.

**Python2 versus. Python3**. Unfortunately, there are two versions of Python and they aren’t compatible. We’ll be using Python3 (current version: 3.5) mainly because Python2 is slated to go away (though, in the open source world, that doesn’t mean a whole lot!) but also because it is a better language.

Texts

There is no text for this class. The following will be helpful if you want to go above and beyond the material covered in the course:


Online resources

Python documentation: [http://docs.python.org/3.4/index.html](http://docs.python.org/3.4/index.html)
Python tutorial: [https://docs.python.org/3.4/tutorial/](https://docs.python.org/3.4/tutorial/)
Python Regular Expressions [https://docs.python.org/2/library/re.html](https://docs.python.org/2/library/re.html)
Web crawling (urllib): [https://docs.python.org/3.4/howto/urllib2.html](https://docs.python.org/3.4/howto/urllib2.html)
Evaluation and learning components

Mini Quizzes (10%): We’ll have a few, very short, quizzes mainly to reinforce points made in class and also to help you get your hands dirty. Most quizzes will be online ‘do whenever you have the time’, though some may be in-class. All quizzes are open book and you’re welcome to check your solutions on your computers. Quizzes will be lightly graded so make a good faith effort and you’ll do fine.

Home assignments (20%): We’ll have a few home assignments as well. Like the quizzes, assignments are not meant to be diagnostic but rather to help you practice and learn so they will be very lightly graded. You can consult with others, ask me questions, google for help, but do try them on your own first. Because it is important that you do the assignments, I will accept late assignments. Assignments submitted within a week of the due date will be penalized 25% and assignments submitted more than a week late will be penalized 50%.

Project (30%): You will be required to choose a data set (you need to find your own!) and analyze it. You will need to Final submission will include an in-class “speed-date” presentation and demonstration.

Participation (10%): Demonstrate engagement in the course by asking questions. I’ll respond to every question, either online or, if the response is of general interest, in the classroom.

Exams (30%): One or two in-class, closed book, exams