CSCI 1360

Foundations for Informatics and Analytics

Location TBD

Dr. Shannon Quinn

Email: squinn@cs.uga.edu
Website: http://cs.uga.edu/~squinn/courses/fa16/csci1360/
Office Location: Boyd GSRC 638A
Office Hours: TBD

The course syllabus is a general plan; when (not if) deviations arise, they will be announced.

Course Description: Informatics, or "data science," are rapidly becoming essential skills for scientists across fields; in addition to field-specific specializations, researchers require knowledge of and experience with quantitative analytical techniques for extracting knowledge from raw data.

This course aims to provide an introduction to concepts in scientific programming and data science using the Python language. Students are given hands-on opportunities to learn techniques applicable to quantitative analyses across a broad range of fields. These core techniques involve formulating solutions in terms of their inputs and outputs (functional programming), repeated operations (loops), branching operations (conditionals), different methods of organizing data (data structures), how to implement an optimal problem-solving strategy (algorithm design), and methods for visualizing and interpreting results.

Prerequisites: MATH 1113 Precalculus.

Credit Hours: 4

No required textbooks! Recommended texts include:

- Shaw, Zed. *Learn Python the Hard Way* (3^{rd} ed., 2013) ISBN-13: 978-0321884916.
- Grus, Joel. *Data Science from Scratch: First Principles with Python* (1^{st} ed., 2015) ISBN-13: 978-1491901427.
- McKinney, Wes. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython* (1^{st} ed., 2012) ISBN-13: 978-1449319793.

Topical Course Outline

1. Introduction to informatics
2. `hello_world.py` and Python variable types
3. Loops, conditionals, and control flow
4. Data structures: lists, arrays, dictionaries, sets, matrices
5. Functions and functional programming
6. Vectorized programming
7. Data formats, transformations, and preprocessing
8. Classification, regression, and clustering
9. Data visualization
10. Extending the Python ecosystem

Grade Distribution:

Participation	5%
Assignments	60%
Midterm Exam	15%
Final Exam	20%

There will be 10 relatively brief programming assignments, each worth 6% of your total grade. These are intended to give you hands-on experience with the concepts being taught in the class and to familiarize you with the Python language and its ecosystem. There will be midterm and final exams, and a participation component. This can take several forms: asking questions from lecture, attending office hours, or submitting or answering questions on the mailing list or bulletin boards.

Course Policies

- **Attendance**
  - Come to lecture. If you're not in lecture to answer every question I pose to the class, and you have a bad day on the midterm, I won't have any way to know that you actually understand the material. This course has an enrollment limit of 15; I'll know when someone is missing.
  - If you cannot attend lecture, let me know ahead of time and we'll work something out.
  - Try to keep in-lecture hacking on your laptop to a minimum. I certainly understand testing out methods that we discuss in class, but I may interleave critical exam hints into the lecture if I detect a lack of attentiveness; I cannot be held responsible if these hints fall on distracted ears.
- **Assignments**
  - Assignments are due by 11:59:59pm on the noted date. Assignments turned in after this deadline will lose 25/100 points for every subsequent 24 hour-period they are late.
  - *The presence or absence of any form of help or collaboration, whether given or received, must be explicitly stated and disclosed in full by all involved, on the first page of their assignment* ("I did not give or receive any help on this assignment" or "I helped [person] with [specific task]."). Collaboration without full disclosure will be handled severely; except in usual extenuating circumstances, my policy is to fail the student(s) for the entire course.

  - **DO NOT COPY CODE.** I cannot stress this enough. Coding is a lot like writing: everyone has their own style that is very recognizable. It's not difficult to tell when students share their code. Don't do it.

- **Exams**
  - **Any** material covered in lecture or homework assignments is considered fair game for both exams.
  - Both exams will be cumulative. I may even be lazy / clever and copy-paste midterm questions into the final.
  - The exact format of both exams will be variable; previous years' exams have included a mixture of multiple choice, matching, true/false, and hand-coding. These will likely follow the same pattern.
  - Time permitting, "practice" versions of the exams will be released. I will make every effort, but it will depend entirely on my schedule when the time comes; as such, do not assume it will happen.

Academic Honesty

As a University of Georgia student, you have agreed to abide by the University's academic honesty policy, "A Culture of Honesty," and the Student Honor Code. All academic work must meet the standards described in "A Culture of Honesty" found at: `https://ovpi.uga.edu/academic-honesty/academic-honesty-policy`. Lack of knowledge of the academic honesty policy is not a reasonable explanation for a violation. Questions related to course assignments and the academic honesty policy should be directed to the instructor.

- Read "A Culture of Honesty," the UGA academic honesty policies and CS Academic Integrity policies.
- You must not allow others to copy or look at your work.
- You must not give/share your lab/project assignment work to a fellow student.
- Copying significant portions of code from a fellow student or any other source (including internet) is plagiarism and will be dealt with as such.
- If you have questions about an assignment or if you run into problems, contact your instructor/lab instructor.
- During exams, no assistance and no additional materials are allowed.
- All of your coursework must meet the aforementioned policies and rules. Students that violate any of these rules or the UGA Academic Honesty policies will be liable to a penalty. The instructor will strictly enforce Academic Honesty policies and report any violation of the aforementioned policies and rules.

****Highly Tentative** Course Outline**
Subject to change.

Week	**Content**
Week 1	• Introduction to informatics and data science
Week 2	• Loops and conditionals • *HW1 released*
Week 3	• Data Structures • *HW1 due; HW2 released*
Week 4	• Data Structures • *HW2 due; HW3 released*
Week 5	• Vectorized programming and linear algebra • *HW3 due; HW4 released*
Week 6	• Functions • *HW4 due; HW5 released*
Week 7	• Functions • *HW5 due*
Week 8	• Review • *Midterm*
Week 9	• File I/O • *HW6 released*
Week 10	• Functional programming • *HW6 due; HW7 released*
Week 11	• Data analytics: classification • *HW7 due; HW8 released*
Week 12	• Data analytics: regression • *HW8 due; HW9 released*
Week 13	• Data analytics: clustering • *HW9 due; HW10 released*
Week 14	• Visualizing data • *HW10 due*
Week 15	• Packaging Python modules; Review • *Review for final*
Week 16	• TBD; Review • *Review for final*