Stat 650/750 Course Description and Syllabus Fall 2018

Vital Info

Class Schedule	TR 12–1:30PM, BH 136A
Instructors	Christopher Genovese 232E Baker Hall (x8-7836) genovese@cmu.edu Alex Reinhart 232K Baker Hall areinhar@stat.cmu.edu (Please put "[650]" or "[750]" in all email subject lines.)
Office Hours	To be determined empirically and by appointment
TAs	Taylor Pospisil, Nic Dalmasso
Web Page	https://github.com/36-750 (See below for more on GitHub.)
Text	Class notes and handouts

Course Objectives and Scope

Computing is an essential—and increasingly important—part of statistical practice and methodology, but computing's role in the typical graduate Statistics curriculum (including ours) has not been commensurate with its importance. For most students, the primary opportunity to develop computing skills comes during research projects, but this tends to put completion above learning and tends to cover ideas that are focused on the particular needs of the project. While students may be able to take a computing-related course here or there, it is rare for such a course to cover the fundamental concepts in computing and develop critical skills in a way that will pay off during later research, work experience, and beyond. This course, 36-650/750, aims to fill that gap.

The premise of this course is that building a broad and solid foundation in computing will pay significant dividends throughout a student's research career. We will focus on four main themes: (a) effective programming practices; (b) fundamental principles of software design; (c) important algorithms, data structures, and representations; and (d) essential tools and methods. Along the way, we will also consider a variety of applications and techniques that are important for statistical practice. However, the focus of the course is not on specific statistical techniques per se. Similarly, although

you will practice new tools and approaches to programming, the purpose of this course is *not* teaching you to program.

A second premise of this course is that practice is the key to developing strong computing skills. To this end, the work will consist of many programming tasks of various sizes, both between and during class. You will have access to a repository of exercises varying in complexity so that you can target your work to your experience level, and you will have an opportunity to work interactively with your peers. Prior programming experience is not a requirement, but if you have not programmed before (or much), you will be expected to work toward learning a chosen language. This course is, for the most part, language agnostic, and indeed, you will be asked to do several (simpler) tasks in a language outside your comfort zone to gain the valuable perspective that this offers.

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

- develop correct, well-structured, and readable code;
- design useful tests at all stages of development;
- effectively use development tools such as editors/IDEs, debuggers, profilers, testing frameworks, and a version control system;
- build a small-to-medium scale software system that is well-designed and that facilitates code reuse and generalization;
- select algorithms and data structures for several common families of statistical and other problems;
- write small programs in a language new to you.

Class Mechanics

Classes will feature a combination of lectures, interactive discussions, and (single and group) programming activities. We will often discuss, edit, and run code from a variety of sources, and we will do a fair amount of real-time programming. Hence, ***you should bring your laptop to every class***.

We will have mechanisms for sharing code snippets interactively and turning in homeworks interactively. Further details are described below in Assignments. For this purpose, sign up for a (free) account at `github.com` and then visit `https://classroom.github.com/a/MMfm3qKz` to get your working repositories for the class. (You do not need to sign up for the educational discount, since it's all under our organization.)

Participation and attendance are important parts of the class and as such will constitute a nontrivial portion of your final grade.

Resources

Class notes will be made available on GitHub shortly after class in the `Lectures` subdirectory of the `documents` repository. In addition, there will be occasional handouts and readings, all of which will be available within the same repository.

We will also publish the lecture notes and related documents on a more comprehensive and user-friendly site, at `https://36-750.github.io`.

There is no primary text for the class, but several useful (though not required) books are available for your use from Professor Genovese. This includes several general books on coding and algorithms along with some language-specific guides for particularly recommended languages. A list of these books along with other online resources will be available in the `documents` repository.

The instructors will hold regular office hours at times to be announced in class. The TAs will also be available to answer questions.

Assignments

The main work in this course consists of programming and related exercises. There will be three main types of exercises:

- Stand-alone. These are short, self-contained tasks that illustrate or build upon an idea we have discussed in class. Many of these involve writing a short program, but others will involve some related task (e.g., debugging, reasoning).
- Vignettes. These are groups of exercises around a central theme, data set, or idea. Working on these exercises in succession is intended to help you understand the central idea more deeply. Within a vignette, it is recommended that you do all the exercises that are not marked optional, though this usually takes place over several assignments.
- Challenge. The Challenge project will be a larger programming task integrating several different ideas and skills. You'll be able to choose between several Challenges at the beginning of the semester, and will then complete your chosen Challenge in four parts submitted every three weeks. Deadlines are shown in Table 1.

All of these exercises will be available in the course repository on GitHub.

Short assignments will be graded on a simple rubric as either "Mastered" or "Not yet mastered". Challenges can also be graded as "Sophisticated". The criteria for these levels will be specified in rubrics that will be posted in the `documents` repository.

Part	Deadline
1	October 2
2	October 23
3	November 13
4	December 4
Final revisions	December 13

Table 1. Challenge submission deadlines

Using the facilities on GitHub, the TAs will be able to review your submissions and provide feedback, with comments both in general and on specific lines or sections of your code. Because revision is an important skill for you to practice, you will have the opportunity to revise your assignments to address issues found during code review. Each assignment and project may be revised at most twice. We will describe in detail the procedure for requesting a review by the TAs of an original or revised submission. Note, however, that the review is contingent on meeting the basic requirements of the submission. For example, if required tests or scripts are missing or if there are significant deficiencies with coding style and organization, the review will stop at noting that issue and you will have used a submission/revision for little gain.

We will give you as much latitude as possible in selecting which exercises to do from the repository. The exercises in the repository cover a wide range, both in topic and complexity, and we want you to challenge yourself in selecting exercises that expand your knowledge and skill. At times, however, specific exercises will be assigned or exercises will be constrained to be selected from a particular subset of those available to fit with work we are doing in class.

Finally, this course is (nearly) language agnostic: you can use as your primary programming language any reasonable one that the instructors or TAs have worked with, which covers a very wide range. (If you have any questions as to suitability, ask Chris or Alex.) In addition, we believe strongly that getting experience with multiple languages helps broaden your perspective on thinking about problems, so we recommend that you do *at least two* assignments (simpler ones, if desired) in a language that is new to you. We will offer recommendations in class.

Grade	Assignment points
R	Fail to meet requirements for D
D	15
C	20
B	25
A	30
A+	35

Table 2. How the base grade is calculated

Grading

Course grading is based on the rubric mentioned above. We will use a simple set of requirements to set your grade. The semester will have one Challenge project, submitted in four pieces with specific deadlines given in Table 1; stand-alone assignments can be submitted ***at most twice per week***, including revised assignments, so do not procrastinate. (There are fifteen weeks, so you have 28 opportunities to submit, not including the first week.) The TAs will grade the first two assignments or revisions submitted each week, leaving any excess submissions to be graded in subsequent weeks. Revisions for each Challenge part must be submitted ***at least one week*** before the deadline for the next part.

To pass the course with at least a D, ***you must Master all four parts of the Challenge***. Additionally, your grade is determined by the number of assignments you Master and whether your final Challenge submission is Sophisticated. Each Mastered assignment is worth from 1 to 3 points, based on its difficulty, and so your base grade is determined based on the number of assignment points. See Table 2.

The base grade can be adjusted in three ways:

1. The grade is adjusted upward one step if you earn a Sophisticated on the Challenge.

2. The grade may be adjusted downward one step if Challenge parts are submitted late or if you do not turn in at least 10 assignment points by October 19th, when mid-semester grades are assigned. (Not all these submissions must yet be graded.) Mid-semester grades will be based on Challenge parts and assignment points submitted by this date.

3. The grade may be adjusted upwards a step for students who demonstrate outstanding participation in class; it may be adjusted downwards up to two steps for poor participation or attendance.

Policies

EMAIL. When sending an email, please put either "[650]" or "[750]" at the beginning of the subject line, so that we can easily identify the message. Also, please be advised that merely sending email ***does not eliminate your responsibility for completing assignments on time***.

COLLABORATION, CHEATING, AND PLAGIARISM. Discussing assignments with your fellow students is allowed and encouraged, but it is important that every student get practice working on these problems. This means that ***all the work you turn in must be your own***. You must devise and write your own solutions and carry out your own tests. The general policy on homework collaboration is:

1. You must first make a serious effort to solve the problem.
2. If you are stuck after doing so, you may ask for assistance from another student. You may discuss strategies to solve the problem, but you ***may not look at their code***, nor may they spell out the solution to you step-by-step.
3. Once you have done so, you must write your own solution individually. ***You must disclose***, in your GitHub pull request, the names of anyone you got help from.

This also applies in reverse: if someone approaches you for help, you must not provide it unless they have already attempted to solve the problem, and you ***may not share your code***.

You may also use external sources (books, websites, papers, ...) to

- Look up programming language documentation, find useful packages, find explanations for error messages, or remind yourself about the syntax for some feature,
- Read about general approaches to solving specific problems (e.g. a guide to dynamic programming or an introduction to unit testing in your programming language), or
- Clarify material from the course notes or assignments.

But external sources must be used to *support* your solution, not to *obtain* your solution. You may ***not*** use them to

- Find solutions to the specific problems assigned as homework (in words or in code)—you must independently solve the problem assigned, *not* translate a solution presented online or elsewhere.

- Find course materials or solutions from this or similar courses from previous years, or
- Copy code snippets to use in your submissions without attribution.

If you do use code from online or other sources, you ***must*** include code comments identifying the source.

Please talk to us if you have any questions about this policy. Any form of cheating is grounds for sanctions to be determined by the instructors, including grade penalties or course failure. Students taking the course pass/fail may have this status revoked. We are also obliged in these situations to report the incident to your academic program and the appropriate University authorities. Please refer to the University Policy on Academic Integrity (link).

Disability Resources. If you have a disability and have an accommodations letter from the Disability Resources office, we encourage you to discuss your accommodations and needs with us as early in the semester as possible. We will work with you to ensure that accommodations are provided as appropriate. If you suspect that you may have a disability and would benefit from accommodations but are not yet registered with the Office of Disability Resources, we encourage you to contact them at access@andrew.cmu.edu.

Wellness. Course work at this level can be intense, and we encourage you to take care of yourself. Do your best to maintain a healthy lifestyle this semester by eating well, exercising, avoiding drugs and alcohol, getting enough sleep and taking some time to relax. This will help you achieve your goals and cope with stress.

All of us benefit from support during times of struggle. If you are having any problems or concerns, do not hesitate to come speak with either of us. There are also many resources available on campus that can provide help and support. Asking for support sooner rather than later is almost always a good idea.

If you or anyone you know experiences any academic stress, difficult life events, or feelings like anxiety or depression, we strongly encourage you to seek support. Counseling and Psychological Services (CaPS) is here to help: call 412-268-2922 or visit their website at http://www.cmu.edu/counseling/. Consider also reaching out to a friend, faculty member, or family member you trust to help get you the support you need.

Policy Updates. Updates to policies and course information will be posted in updated versions of this syllabus and announced via email.