Syllabus

Statistical Computing, 36-350

Fall 2022

General information

- Instructor: Alessandro Rinaldo (arinaldo at cmu dot edu)
- Class times: Tuesdays and Thursdays, 10:10-11:30am, in POS 152
- Lectures will take place in the first half of Tuesday's class. The rest of the class time will consist of supervised work on the weekly labs, via Zoom (see Canvas for the passcode).
- Head TA: [SSH] Shamindra Shrotriya (shamindra at cmu dot edu)
- Grad TAs (with andrew IDs): [JP James Pak (jpak1) and [XS] Xiaochen Sun (xsun3)
- Undegrad TAs (with andrew IDs): [LC] Linpeng Chen (linpeng), [XG] Xiaofeng Gan (xgan), [AG] Aaron Gong (yingyaog), [LL] Li Li (lili2), [PM] Parth Maheshwari (parthmah), [SS] Shannon Sun (sysun) and [ST] Steven Tang (sjtang)
- Office hours will be on Zoom (see Canvas for the passcode):
  - Tuesday: 9:00am-10:00am [SSH]
  - Wednesday: 2:00pm-3:00pm [PM/SS]
  - Thursday: 2:00pm-7:00pm [LC/AG/LL/PM/SS/JP/XS]
  - Friday: 1:00pm-6:00pm [LC/XG/AG/LL/PM/SS/JP/XS/ST]
- Announcements: Piazza group (access also via Canvas). Make sure to check pinned posts for updates regularly.

- Labs: Submit completed lab work using Gradescope (access also via Canvas). The solutions will be posted on Canvas: see Files > solutions > labs.
- The course will be based on the R language, a free software for statistical computing and graphics. We will use R Studio, an integrated development environment for R and Python and R Markdown, a markup language for combining R code with text and graphics.

Course overview

Computational data analysis is an essential part of modern statistics. Competent statisticians must not just be able to run existing programs, but to understand the principles on which they work. They must also be able to read, modify, and write code, so that they can assemble the computational tools needed to solve their data analysis problems, rather than distorting problems to fit tools provided by others. This class is an introduction to statistically-oriented programming, targeted at statistics majors, without assuming extensive programming background.

Students will learn the core of ideas of programming—data structures, functions, iteration, debugging, logical design, and abstraction—through writing code to assist in statistical analyses. Students will learn how to write maintainable code, as well as debug and test code for correctness. They will learn how to set up and run stochastic simulations, how to fit basic statistical models and assess the results, and how to work with and filter large data sets. Since code is an important form of communication among scientists, students will also learn how to comment and organize code.

The class will be taught entirely in the R programming language.

Course website

The course website is https://www.stat.cmu.edu/~arinaldo/Teaching/36350/F22/. The course schedule, lecture notes, labs, supplementary materials, etc., will be posted there.

Prerequisites

This is an introduction to programming for statistics students. Prior exposure to statistical thinking, to data analysis, and to basic probability concepts is essential. Previous programming experience is not assumed. Formally, the prerequisites are "Computing at Carnegie Mellon", 36-202 or 36-208, and 36-225.

Course mechanics

This class will be run in a **flipped format**. Instead of having regular lectures during our scheduled class times, the week will be structured as follows.

- Each **Monday morning**, the **lecture notes** for the week will be posted on the course website. During the first 40 minutes of **Tuesday's class period**, the Professor will walk through the lecture material. These lectures will be delivered **in person.** Attendance is optional, and if you choose not to attend, it is up to you to digest the lecture material.
- Each **Monday morning**, the **lab** for the week will also be posted on the course website. It will be due **9pm on Friday** night, and submission will be done via Gradescope.
- The **second half of Tuesday class period (after the first 40-45 minutes of lecturing by the Prof.) and Thursday class period** are open lab work time. The Professor and TAs will be available **virtually via Zoom** to answer any questions you may have. You can also attend one of the weekly office hours (details given on the course website) held by the TAs; or you can do the lab completely on your own.
- Lastly, there will be a short **quiz**, posted on Gradescope, consisting of true/false and multiple choice questions on the lecture material. It will be due **9pm on Wednesday** night. This quiz is optional and will only count as extra credit towards your weekly lab score.

Should there arises the need to switch to an online/remote modality during the semester, this change can be implemented in a straightforward manner and with minimal disruption: the lectures (first half of Tuesday's class periods) will be broadcast on Zoom using the Zoom lectures room (see Canvas for the passcode) and the recordings will be made available on Canvas.

Grading

Grades will be calculated as follows:

- Labs: 80%
- Exam: 20%
- Quizzes: extra credit towards labs

Here are the cutoffs for letter grades, based on total percentages:

- A: 90% or higher
- B: 80% to 89%
- C: 70% to 79%
- D: 60% to 69%
- R: 59% or lower, on a case by case basis

The Professor may adjust these cutoffs, but only in the direction that favors the students. For example, the cutoff for an "A" may end up being adjusted to be lower than 90%, but not higher.

R and RStudio

R is a free, open-source programming language for statistical computing. All of our work in this class will be done using R. You will need regular, reliable access to a computer running an up-to-date version of R. If this is a problem, then let the Professor or TAs know right away.

RStudio is a free, open-source R programming environment. It contains a built-in code editor, many features to make working with R easier, and works the same way across different operating systems. Most importantly it integrates R Markdown seamlessly. You will use RStudio for the labs and final.

Getting help

Labs

Attending labs are often the best way to get help. You will be able to ask questions of the Professor and TAs for the entire time (aside from the lecture recitation period, the first 40 minutes of Tuesday).

Office hours

Office hours will be held by the TAs, and the times will be spread out over the week. The office hours are **only conducted via Zoom**. The timings and Zoom details are found on the course website.

In addition to the weekly lecture labs, the office hours (not Piazza) are your main source of help with the course materials. They are offered from Tue-Fri at various times for the convenience of the students. If you can't attend any of the TA office hours, then you are advised to attend the lecture labs.

Note that some of these office hour timings may change slightly over the semester. Please check in regularly on the course website for the most up to date details.

Piazza

IMPORTANT: In this course we will **not** use Piazza as a discussion forum for weekly labs or course materials. Such course related materials are meant to be discussed during office hours and lecture labs.

Instead Piazza will only be used by the course staff to make important timely announcements about the weekly lab and logistics. These announcements include:

- A pinned FAQ post for that weeks lab, e.g. *Lab Z FAQs* which will clarify common student questions for Lab Z from office hours.
- Pinned posts on common `R` installation issues and troubleshooting.
- Any additional useful tips and tricks to help students with labs.

It is your responsibility to regularly check Piazza for any posts and updates. We will in general not answer any questions on Piazza regarding specific lab questions, which are meant for office hours.

You are welcome to ask questions on Piazza about **course logistics**, e.g., if there are any issues with zoom links. You can also raise any bugs, typos, or errors in the lab materials you can also flag them on Piazza and we will update the lab FAQs accordingly.

If you post anything on Piazza, please keep in mind that content deemed inappropriate will be taken down by the course TAs, and reported to the course Professor.

Email

Email will be used for questions on class administration (class policies, exceptional circumstances, etc.), rather than class contents. Please direct such inquiries to the Head TA. The subject line of all emails should begin with "[36-350]". The Professor will be available for issues that cannot be resolved first with the Head TA.

Evaluation

Quizzes

Quizzes will be short (about 8 questions), and consist of true/false and multiple choice questions. They are supposed to be an easy recap of the material covered in the week's lecture materials. Quizzes will be **turned in via Gradescope**, due **9pm on Wednesday** each week. They are optional and will only count as extra credit towards your weekly lab score.

Labs

Labs will be completed in R Markdown format (file extension Rmd). They will involve writing a combination of code and written prose, and the R Markdown format is crucial since it allows for a combination of the two. Labs will be **turned in via Gradescope**, due **9pm on Friday** each week, and they **must be submitted only in PDF format**, the result of calling "Knit PDF" from RStudio on your R Markdown document. Be careful that you do this, because work submitted in any other format will receive a grade of 0, without exception.

Note also: all code used to produce your results must be shown in your PDF file (e.g., do not use `echo=FALSE` or `include=FALSE` as options anywhere).

Students may choose to collaborate with friends on the labs, but must indicate with whom they collaborated. Also, be sure to carefully read the collaboration policy below.

Regrades

Any regrade requests for labs and quizzes must be **submitted on Gradescope within 24 hours** of grade release. We will shut down the regrade release after 24 hours, no exceptions.

Note that we prepare detailed and consistent rubrics, so we expect minimal such requests. If we find you are over-submitting regrade requests despite us providing good feedback, your case will be flagged for the Professor to review.

Exam

There will be a final exam. It will be entirely similar in format to the quizzes (true/false and multiple choice questions), and will be comprehensive.

Late work

For **quizzes** no late days will be accepted.

For **labs** note the following late submission rules:

- Labs submitted *between* 9:01pm-11:59pm (inclusive) on Fridays get a **20% late penalty**
- Labs submitted *after* 23:59 Fridays without valid extension get a **100% late penalty (0 grade)**
- Only your **most recent** Gradescope submission will be considered for grading and late penalty purposes.
- Your lowest lab score will be dropped at the of the semester.

In case of truly exceptional situations—such as illness or other emergencies—the Head TA can make exceptions and allow late work (labs or quizzes). Note that knitting issues or technical difficulties with submission on your end will **not** receive extensions.

To submit an emergency extension request you must send an email to the **Head TA** with the subject `[36-350] - Extension for Lab Z`, where Z is the relevant lab number. This must be provided **by 3pm Friday** on the due date of the lab. Please provide a brief reason for your request. Any extension requests after 3pm Friday for that lab will not be considered. Please note that any suspicious requests will be flagged with the Professor for review.

Collaboration, copying, and plagiarism

You are encouraged to discuss course material with your classmates. All work you turn in, however, must be your own. This includes both written explanations, and code. Copying from other students, books, websites, or solutions from previous versions of the class, (1) does nothing to help you learn how to program, (2) is easy for us to detect, and (3) has serious negative consequences for you, as outlined in the university's policy on cheating and plagiarism. If, after reading the policy, you are unclear on what is acceptable, please ask the Professor.

Accommodations for students with disabilities

If you have a disability and are registered with the Office of Disability Resources, please use their online system to notify us of your accommodations and discuss with us your needs 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, consider contacting them at access@andrew.cmu.edu.

Take care of yourself

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. You are not alone. Asking for support sooner rather than later is often helpful. Check out the CMU Community Health and Well-Being website for a wealth of information and resources about wellness and mental health.

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 and visit their website at http://www.cmu.edu/counseling/. Consider reaching out to a friend, faculty or family member you trust for help getting connected to the support that can help.

If you or someone you know is feeling suicidal or in danger of self-harm, call someone immediately, day or night:

- CaPS: 412-268-2922
- Resolve Crisis Network: 888-796-8226

If the situation is life threatening, call the police:

- On campus: CMU Police: 412-268-2323
- Off campus: 911

If you have questions about this, then please let us know.

COVID-19

Please refer to the CMU COVID-19 update webpage for COVID-related information and resources for the CMU community.