# 36-350: Statistical computing
# Course policies and syllabus, Summer 2019 Syllabus

Benjamin LeRoy

August 7, 2019

## 1 Overview

**Instructor:** Benjamin (Ben) LeRoy, Email: bpleroy@stat.cmu.edu
Office Hours: Tuesday 9:30 - 10:30 am & Friday 4:30 - 5:30 pm (Wean 3715)
*(Email the TA for any course logistics)*

**Teaching assistant:** Tudor Manole, Email: tmanole@stat.cmu.edu
Office Hours: Thursday 2:30 - 3:30 pm (Wean 3715)
*(Allow 24 hours for email responses. Put "[36-350]" in the subject of the email.)*

**Lectures:** MTWThF, 12:00PM to 1:20PM, Scaife Hall 208

**Course Website:** https://canvas.cmu.edu/courses/10334

## 2 Course overview and objectives

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, input and output, 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 in the R programming language.

### 2.1 Canvas

All course material (slides, labs, homework, and announcements) and submissions will be done through Canvas (URL: https://canvas.cmu.edu/courses/10334). The syllabus and contacts will also be on Canvas. Announcements on occasion will be posted on Piazza but **check Canvas to be up-to-date with all announcements**.

Do not send emails to the Instructor or TA through Canvas.

### 2.2 Pre-requisities

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, but familiarity with the computing system is. Formally, the pre-requisites are "Computing at Carnegie Mellon", 36-202 or 36-208, and 36-225.

### 2.3 Course mechanics

Each week, lecture will be given for roughly the first 50 minutes of class on Monday, Wednesday and Friday. The rest of class on Monday, Wednesday, and Friday and the whole class period on Tuesday and Thursday will be lab sessions, in which students work through a set of exercises. The lab from each week is to due 10pm on Tuesday and Thursday, on Canvas. There will also be a homework each week, due 10pm on Monday, on Canvas. Lastly, there will be a final take-home exam.

### 2.4 Grading

Grades will be calculated as follows:

- Labs: 30%
- Homework: 50%
- Final exam: 20%

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 Instructor 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.

### 2.5 R and R Studio

R is a free, open-source programming language for statistical computing. You can download R at https://cran.r-project.org/. Almost 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 Instructor or TA know right away.

R Studio is a free, open-source R programming environment. You can download R Studio at https://www.rstudio.com/products/rstudio/download/#download. 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 R Studio for the labs, homework, and final.

### 2.6 Laptop and cellphone policy

**No laptops may be used during lectures on Monday, Wednesday, and Friday** (see Section 4.4 for the structure of the week). Since lectures only occur on Monday, Wednesday, and Friday, it is important for you to be attentive during those times. The instructor would kindly tell you close your laptop if you are seen using it during lecture. During the lab periods, you will be using your laptops.

If you do not have a laptop, please contact the instructor. We will be able to make arrangements to lend you a laptop for the summer session.

## 3 Getting help

### 3.1 Office hours

There will be 2 office hours per week, each for an hour and a half long. Please check the Canvas website and Piazza for the specific time and locations, as they are subject to change.

### 3.2 Piazza

Piazza will be used for questions and discussion on the class contents. It is the Instructor's intent for Piazza **to be a student-driven forum**, where the students answer each other's

questions and the Instructor and TA moderate. It is **not intended to be a question-answer hotline**. Class announcements will also be made through Piazza (URL: https://piazza.com/cmu/summer2019/36350u/).

Piazza can be a very successful medium for helpful, class-wide discussions, but without rules, discussions can also quickly get out of hand. Here are the rules for our Piazza group:

1. Be considerate to others (respectful language, no sarcasm).
2. Before posting a question, check that it (or a related question) has a not already been posted. If it has, then use the existing thread for further questions or discussion.
3. For questions about the labs or homework, "What is wrong with this code?" is not an acceptable question. Code that is part of your lab or homework solution **cannot be posted to Piazza**.
4. Along with your posted question, explain step-by-step what you've tried to answer your own question (without posting your solution code).
5. Avoid private questions on Piazza (an option for questions that only Instructor and TA can see), since they might not be answered in a reliable/timely manner.

Content deemed inappropriate – by the above rules and otherwise – will be taken down by the Instructor or TA. If students are unsure about whether a specific question is allowable, you can always ask the Instructor or TA during lab.

### 3.3 Email

Email will be used for questions on class administration (class policies, exceptional circumstances, etc.). This primarily includes: special circumstances, attendance, regrades and extensions. Please direct such inquiries to Tudor Manole (TA). The subject line of all emails should begin with "[36-350]". The Instructor (Ben LeRoy) is available but only for issues that cannot be resolved first with the TA.

**Do not email the Instructor or TA about coding questions related to the labs/homeworks or other course content.** These should be posted on Piazza instead.

## 4 Assignments

### 4.1 Submissions

All assignments (labs, homework, final take-home exam) must be turned in electronically, through Canvas.

All assignments must be completed in R Markdown format (file extension Rmd). Since assignments will involve writing a combination of code and written prose, the R Markdown

format is crucial since it allows for a combination of the two. **All assignments must be submitted only in HTML format**, the result of calling "Knit HTML" from R Studio 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 HTML file (e.g., do not use `echo=FALSE` or `include=FALSE` as options anywhere).

### 4.2 Labs

Each lab will occupy about 110 minutes of class time each week. The lab is your opportunity to get direct feedback from both the instructor and the TA without needing to go to office hours. The grading breakdown for labs: 40% for attendance and 60% for completion of the lab questions. To obtain a full score for completion, each question must be faithfully attempted with relevant code, but is not required to be correct. Attendance will be checked by random sampling on each day.

The lab each week is due 10pm on Tuesday and Thursday, on Canvas. Labs submitted after that Tuesday and Thursday at 10pm will be automatically considered late, using up the 5 late days (see below). For labs due on Tuesday, the 40% attendance is split 13% for Monday and 14% for Tuesday and 13% for the Friday before. For labs due on Thursday, the 40% attendance is split 20% for Wednesday and 20% for Thursday. Students may choose to work with friends on the lab, but read carefully the collaboration policy below.

### 4.3 Homework

There will be a homework assignment every week, due at 10pm on Monday, submitted on Canvas. Students may choose to collaborate with friends on the homework, but read carefully the collaboration policy below. Homeworks submitted after that Monday at 10pm will be automatically considered late, using up the 5 late days (see below).

### 4.4 Structure of a week

The following is a broad summary of what to expect each week of the course. Take note of the outline on Wednesday. *The first week is a slightly different schedule - see section 6.*

- **Monday**: Lecture in the first 50 minutes, lab for the remaining 30 minutes. Attendance is taken. Homework from the previous week is due at 10pm.
- **Tuesday**: Lab for the entire 80 minutes. Attendance is taken. Lab is due at 10pm.
- **Wednesday**: Review of common confusions on homework/lab for up to the first 10 minutes. Lecture for 40 minutes. Lab for the remaining 30 minutes.
- **Thursday**: Lab for the entire 80 minutes. Lab is due at 10pm

- **Friday**: Lecture in the first 50 minutes, lab for the remaining 30 minutes. Attendance is taken.

### 4.5 Final exam

In place of an in-class final exam, there will be a take-home exam. It will be essentially like a homework, but no collaboration with peers is allowed.

### 4.6 Late work

You have a total of 5 late days in the semester, to use between the labs and the homework. Up to a maximum of 3 days can be used at 1 time. After these 5 late days are used up, no late work will be accepted (i.e., they will receive a 0).

You do not need to notify the TA beforehand if you are planning to use a late day(s). The late day(s) are counted by the TA based on the time of the latest submission of the lab or homework.

In case of truly exceptional situations – such as family emergencies or illness – the Instructor can make exceptions and allow late work. If you think your situation is truly exceptional but is not an emergency, then you must notify the TA of your situation at least 2 full days before the particular assigment (lab or homework) is due.

You may **NOT** use late days for the final take-home exam. Any lateness (if not excused beforehand) on the final take-home exam will immediately receive a 0.

### 4.7 Collaboration, copying, and plagiarism

You are encouraged to discuss course material – especially lab work, but also including homework assignments – 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 Instructor.

### 4.8 Disability resources

If you have a disability and require accommodations, please contact Catherine Getchell, Director of Disability Resources, 412-268-6121. If you have an accommodations letter from the Disability Resources office, we encourage you to discuss your accommodations and needs with the instructor as early in the semester as possible. We will work with you to ensure that accommodations are provided as appropriate.

## 5 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. There are many helpful resources available on campus and an important part of the college experience is learning how to ask for help. Asking for support sooner rather than later is often helpful.

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 Re:solve 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 the Professor know.

## 6 Course schedule

The following is the tentative schedule for the course.

| Day | Date | Topic | Due dates (10 pm) |
|---|---|---|---|
| Monday | July 1 | Syllabus, *Lecture:* R Basics | |
| Tuesday | July 2 | *Lecture:* Index & Iteration | |
| Wednesday | July 3 | *Lab* | |
| Thursday | July 4 | *Holiday* | |
| Friday | July 5 | *Lecture:* Text Manipulation | Lab 1 |
| Monday | July 8 | *Lecture:* Data Frames and `apply` | Homework 1 |
| Tuesday | July 9 | *Lab:* | Lab 2 |
| Wednesday | July 10 | *Hw Review* & *Lecture:* Visualization (`ggplot`) | |
| Thursday | July 11 | *Lab* | Lab 3 |
| Friday | July 12 | *Lecture:* Visualization Theory & Coding Style | |
| Monday | July 15 | *Lecture:* Functions | Homework 2 |
| Tuesday | July 16 | *Lab* | Lab 4 |
| Wednesday | July 17 | *Hw Review & Lecture:* Tidyverse's `dplyr` | |
| Thursday | July 18 | Lab | Lab 5 |
| Friday | July 19 | *Lecture*: Tidyverse's `tidyr` | |
| Monday | July 22 | *Lecture:* Simulations | Homework 3 |
| Tuesday | July 23 | *Lab* | Lab 6 |
| Wednesday | July 24 | *Hw Review* & *Lecture:* Fitting Models to Data | |
| Thursday | July 25 | *Lab* | Lab 7 |
| Friday | July 26 | *Lecture:* OOP in R | |
| Monday | July 29 | *Lecture:* Debugging | Homework 4 |
| Tuesday | July 30 | *Lab* | Lab 8 |
| Wednesday | July 31 | *Hw Review* & *Lecture*: Github | |
| Thursday | August 1 | *Lab* | Lab 9 |
| Friday | August 2 | *Lecture:* R packaging | |
| Monday | August 5 | *Lecture:* Prediction & CV | Homework 5 |
| Tuesday | August 6 | *Lab* | Lab 10 |
| Wednesday | August 7 | *Lecture:* Advanced Computing: Split-Apply-Combine, Parallelization, Deep Learning | |
| Thursday | August 8 | Final project work time | |
| Friday | August 9 | Statistical Computing Graduate Student Panel | Final exam |