# Special Topics in Statistics: Computing Tools for Data Analytics
## STAT 679 / STAT 992
## Fall 2020

Instructional Modality: online only
Number of credits: 3

The credit standard for this course is met by an expectation of a total of 135 hours of student engagement with the courses learning activities (45 hours per credit), which include regularly scheduled instructor: reading, writing, videos, problem sets, coding, projects, peer reviews, etc.

## INSTRUCTOR AND TEACHING ASSISTANT

**Instructor**: Cécile Ané, Professor. Preferred contact is email: cecile.ane@wisc.edu.

**Teaching Assistant**: Tzu Hsiang (Sky) Hung. Preferred contact is by email: thung6@wisc.edu.

**Office hours** will be conducted remotely on Blackboard Collaborate Ultra (BBCU):

| | |
|---|---|
| Cecile Ane | Mon & Wed 9-10:30 am |
| Tzu Hsiang (Sky) Hung | Fridays 2-3pm |

Communication will also occur via Canvas announcements, Piazza / Canvas discussions and virtual meetings (on BBCU, zoom or other platform) by appointment.

## OFFICIAL COURSE DESCRIPTION

Special topics in statistics at the master's level. Subject matter varies.
Fall 2020: Computing Tools for Data Analytics

Tools and topics:

- the shell: power of command lines to interact with a machine
- version control and collaboration: git and GitHub
- best practices for scientific computing
- scripting: bash/zsh, Python and Julia

**Requisites**:
Graduate/professional standing or member of the Statistics Visiting International program

## COURSE WEBSITE, LEARNING MANAGEMENT SYSTEM and INSTRUCTIONAL TOOLS

Canvas: https://canvas.wisc.edu/courses/217124 and http://cecileane.github.io/computingtools. For communication, we will use Blackboard Collaborate (integrated in Canvas) and GitHub.

## COURSE LEARNING OUTCOMES

Acquire a strong basis for carrying out computations for graduate research or on the job. More specifically, students will learn to:

1. Do computations efficiently: possibly handling very large data sets
2. Do computations automatically: using scripts to repeat tasks (and avoid manual errors)
3. Build and manage a computing project reproducibly: by oneself or by others
4. Build and manage a computing project collaboratively: share work with a version control system
5. Demonstrate the ability to adapt to change in computing platforms

## REQUIRED TEXTBOOK, SOFTWARE & OTHER COURSE MATERIALS

There is no required textbook. Students need their own computer, running either Mac or Linux, and a reliable internet connection. Campus provides students with technology guidelines and recommendations for instruction. Students should consult these resources prior to the start of the semester.

This "Mac or Linux" requirement reflects current practice in statistics, computational biology and many other scientific research fields. PhD students in biology experience this requirement first-hand when working with large molecular sequence data, to run available analysis pipelines. Computing servers in Statistics are all Linux servers. These servers are accessible remotely, however.

Students without a computer have the following options, none of which are recommended nor supported:

- buy an Intel-based chromebook running Linux (starting around $180)
- if you have a 64-bit machine running Windows 10, you may try the "Windows Subsystem for Linux".

## GRADING

Each item will be graded with a point system. Points will be averaged with the following weights:

- [20%] exercises and participation
- [40%] homework problems
- [20%] peer reviews
- [20%] final project

Final letter scores will be curved.

## HOMEWORK & OTHER ASSIGNMENTS

Homework problems, exercises, projects and peer reviews will be submitted online, on Canvas or via GitHub. Short exercises, quizzes and graded discussions will be assigned on a weekly basis. Short assignments will be **due twice a week**, to provide a regular weekly schedule, during the first half of the semester. During the second half of the semester, problem sets will grow in scope and will be assigned at lower frequency, typically due on Mondays 11:59 pm.

For "exercises and participation" assignments: **Late submissions** will receive **no credit**. The 2 assignments with your lowest scores will be dropped to calculate your overall score in the "exercises and participation" category.

For homework problems, peer reviews and the final project, **late submissions** will receive a maximum of 50% of the total number of points, if submitted less than a week after the deadline. Work submitted later than a week after the deadline will not receive any credit. Deadline extensions will be granted under extenuating circumstances. Vacation or travel does not constitute extenuating circumstances.

Students may discuss assignments and homework problems with others including peers, the TA and instructor. But each student should write their own computer code and project documentation, and should obtain their computer output independently. For the final project, students may *not* discuss the specifics of their work with anyone outside of their group. They should discuss all the details of their final project with group members. Students may discuss high-level topics with anyone.

## EXAMS, QUIZZES, PAPERS & OTHER MAJOR GRADED WORK

Because one cannot master computational tools without practice or without trouble-shooting errors, the course will be primarily project-based. All the topics will be taught in the context of practical examples, and case studies in biology.

During the second third of the course, problem sets will increase in scope. The last assignment before the final project will be **due on Monday 11/9**. This due date of 11/9 is a **hard deadline**, because peer-review will start on Tuesday 11/10. Peer reviews will be due on Monday 11/16.

The final project will be collaborative in groups of 3 or 4 students. The final project will be assigned on Monday 11/16, and **due on Monday 12/7**. This due date of 12/7 is a **hard deadline**, because peer-review will start on Tuesday 12/8. Peer reviews will be due on Monday 12/14.

If a student does not turn in an assignment that is followed by peer review, the student's work will not be peer reviewed, but it will be graded. However, it may not be possible to include the student in the review process, so the student's peer-review grade may be 0 for this assignment's peer review. If a student is worried about a late assignment, they should contact the instructor as soon as possible to discuss the issue. For example, it is best to submit an assignment in which the code is known to contain a bug, than to fail to submit the assignment. Everyone learns during the peer-review process, especially in challenging situations.

## EXPECTATIONS FOR ONLINE INTERACTIONS

**Netiquette:**

See https://kb.wisc.edu/50548 for a general netiquette (e.g. CAPITAL is interpreted as shouting). Any comment or answers must be on topic, polite, and respectful of others. Students must not post answers to homework problems. Students should not send questions to the TA or instructor via email, unless for personal or sensitive issues. Questions should be sent via Piazza or Canvas discussions. Students should not expect an immediate answer to a question posted late at night before an assignment due date.

**Expectations on webconferencing use**:

Learning is facilitated by interactions with instructors, TAs and other learners. Community is an asset for learning, but is made difficult in an online format. Multiple group activities will target the learning objective to "Build and manage a computing project collaboratively". We need to build trust with other learners in the group. Yet we don't trust people we don't know, and we don't know people we don't see or talk to, to put it bluntly.

To build community, all course participants are expected to **turn on their video** during webconferecing (e.g. during office hours, group work and other synchronous activities). We understand that everyone is joining from a different place, such as our homes. We expect a variety of working spaces, including bedrooms, kitchens and public spaces. Housemates (possibly including kids) in the background are expected. We value each learner as a person first, and we acknowledge that each student learns in a different context.

During webconferencing, ways to engage will include: a chat box, breakout rooms for small groups, and audio questions. Microphones should be muted by default. Participants should raise their hands in BBCU and wait for the presenter to invite them to speak, before turning their mic on.

## HOW TO SUCCEED IN THIS COURSE

It is recommended that students set a **weekly schedule** for their own course study plan, to make regular progress throughout the semester. Frequent short assessments will be provided on a regular schedule to help students succeed.

Students are encouraged to take full advantage of **office hours**. Using screensharing, the instructor, TA and other students can help each other with code, debugging, and troubleshooting technical issues. Running into bugs is a required part of learning computing skills. To succeed, students must spend time to encounter and overcome these "bug" challenges. The instructor and TA will help students go through this learning process, with guidance and encouragement.

Students are encouraged to use small groups created for assignments as opportunities for growing their **support network**, and forming learning groups to help each other. One goal for these group assignments is precisely to help students benefit from interacting with peers and be part of a community despite the online nature of the course.

Other campus-wider resourses and services include:

- University Health Services
- Undergraduate Academic Advising and Career Services
- Office of the Registrar
- Office of Student Financial Aid
- Dean of Student Office

## PRIVACY OF STUDENT RECORDS and the USAGE of AUDIO RECORDED LECTURES

See information about privacy of student records and the usage of audio-recorded lectures.

Lecture materials and recordings for STAT 679 / 992 are protected intellectual property at UW-Madison. Students in this course may use the materials and recordings for their personal use related to participation in this class. Students may also take notes solely for their personal use. If a lecture is not already recorded, you are not authorized to record my lectures or office hours without my permission unless you are considered by the university to be a qualified student with a disability requiring accommodation [Regent Policy Document 4-1]. Students may not copy or hare lecture materials and recordings outside of class, including posting on internet sites or selling commercial entities. Students are also prohibited from providing or selling their personal notes to anyone else or being paid for taking notes by any person or commercial firm without the instructor's express written permission. Unauthorized use of these copyrighted lecture materials and recordings constitutes copyright infringement and may be addressed under the university's policies, UWS Chapters 14 and 17, governing student academic and non-academic misconduct.

## STUDENTS' RULES, RIGHTS & RESPONSIBILITIES

During the global COVID-10 pandemic, we must prioritize our collective health and safety to keep ourselves, our campus, and our community safe. As a university community, we must work together to prevent the spread of the virus and to promote the collective health and welfare of our campus and surrounding community.

### UW-Madison Badger Pledge

### UW-Madison Face Covering Guidelines

This class is held online. Students may *not* meet in person, even for group work. Webconferencing must be used for group work, with screensharing for pair-programming. Still, students are reminded that while on campus (such as to attend other courses or for their own research), they are required to wear appropriate and properly fitting face coverings while present in any campus building unless working alone in a laboratory or office space.

### Quarantine or Isolation Due to COVID-19

Student should continually monitor themselves for COVID-19 symptoms and get tested for the virus if they have symptoms or have been in close contact with someone with COVID-19. Student should reach out to instructors as soon as possible if they become ill, in order to make alternate plans for how to proceed with the course. Students are strongly encouraged to communicate with their instructor concerning their illness and the anticipated extent of their absence from the course. The instructor will work with the student to provide alternative ways to complete the course work.

### Course Evaluations

UW-Madison now uses an online course evaluation survey tool, AEFIS. You will receive an official email two weeks prior to the end of the semester when your course evaluation is available. You will receive a link to log into the course evaluation with your NetID where you can complete the evaluation and submit it, anonymously. Your participation is an integral component of this course, and your feedback is important to me. I strongly encourage you to participate in the course evaluation.

## ACADEMIC CALENDAR & RELIGIOUS OBSERVANCES

See: https://secfac.wisc.edu/academic-calendar/#religious-observances

## ACADEMIC INTEGRITY & DATA ETHICS

By virtue of enrollment, each student agrees to uphold the high academic standards of the University of Wisconsin-Madison; academic misconduct is behavior that negatively impacts the integrity of the institution. Cheating, fabrication, plagiarism, unauthorized collaboration, and helping others commit these previously listed acts are examples of misconduct which may result in disciplinary action. Examples of disciplinary action include, but is not limited to, failure on the assignment/course, written reprimand, disciplinary probation, suspension, or expulsion.

The members of the faculty of the Department of Statistics at UW-Madison uphold the highest ethical standards of teaching, data, and research. They expect their students to uphold the same standards of ethical conduct. Standards of ethical conduct in data analysis and data privacy are detailed on the ASA website, and include:

- Use methodology and data that are relevant and appropriate; without favoritism or prejudice; and in a manner intended to produce valid, interpretable, and reproducible results.
- Be candid about any known or suspected limitations, defects, or biases in the data that may affect the integrity or reliability of the analysis. Obviously, never modify or falsify data.
- Protect the privacy and confidentiality of research subjects and data concerning them, whether obtained from the subjects directly, other persons, or existing records.

By registering for this course, you are implicitly agreeing to conduct yourself with the utmost integrity throughout the semester.

## ACCOMMODATIONS FOR STUDENTS WITH DISABILITIES STATEMENT

The University of Wisconsin-Madison supports the right of all enrolled students to a full and equal educational opportunity. The Americans with Disabilities Act (ADA), Wisconsin State Statute (36.12), and UW-Madison policy (Faculty Document 1071) require that students with disabilities be reasonably accommodated in instruction and campus life. Reasonable accommodations for students with disabilities is a shared faculty and student responsibility. Students are expected to inform faculty [me] of their need for instructional accommodations by the end of the third week of the semester, or as soon as possible after a disability has been incurred or recognized. Faculty [I], will work either directly with the student [you] or in coordination with the McBurney Center to identify and provide reasonable instructional accommodations. Disability information, including instructional accommodations as part of a student's educational record, is confidential and protected under FERPA. (See: McBurney Disability Resource Center)

## DIVERSITY & INCLUSION STATEMENT

Diversity is a source of strength, creativity, and innovation for UW-Madison. We value the contributions of each person and respect the profound ways their identity, culture, background, experience, status, abilities, and opinion enrich the university community. We commit ourselves to the pursuit of excellence in teaching, research, outreach, and diversity as inextricably linked goals.

The University of Wisconsin-Madison fulfills its public mission by creating a welcoming and inclusive community for people from every background – people who as students, faculty, and staff serve Wisconsin and the world.

## COMPLAINTS

If you have a complaint about a TA or course instructor, you should feel free to discuss the matter directly with the TA or instructor. If the complaint is about the TA and you do not feel comfortable discussing it with him or her, you should discuss it with the course instructor. Complaints about mistakes in grading should be resolved with the instructor in the great majority of cases. If the complaint is about the instructor (other than ordinary grading questions) and you do not feel comfortable discussing it with him or her, contact the Director for Graduate Studies, Professor Bret Larget (bret.larget@wisc.edu).

If your complaint concerns sexual harassment, please see campus resources listed at https://compliance.wisc.edu/titleix/resources/. In particular, there are a number of options to speak to someone confidentially.

If you have concerns about climate or bias in this class, or if you wish to report an incident of bias or hate that has occurred in class, you may contact the Chair of the Statistics Department Climate & Diversity Committee, Professor Po-Ling Loh (ploh@stat.wisc.edu). You may also use the University's bias incident reporting system, which you can reach at https://doso.students.wisc.edu/services/bias-reporting-process/.