

# CSE 8803 EPI, Fall 2025

## Data Science for Epidemiology

---



### Grading

- Homework: 30%
- Scribe Notes: 10%
- Project: 50%
- Class Participation: 10%

### Late grading, questions and requests

- Clarification questions about assignments should go on Piazza.
- Regrading requests for any assignments should be done in writing, via e-mail to the instructor, at the latest 1 week after graded assignments are distributed. We reserve the right to re-grade the entire assignment though.
- The assignment due date is the one posted on the web. All assignments (including Project deadlines) are due at the beginning of the lecture, on the due date.
- **Slip days:** To accommodate for coinciding deadlines you may have from other courses, or personal unforeseen events such as sickness, each person has an automatic extension of 4 calendar days---no questions asked. You can use the extension on any assignment(s) remaining during the semester (\*except\* for the project deliverables, all of which will have a **firm** deadline). For instance, you can hand in one assignment 4 days late, or each of three assignments 1 day late.
  - **Late** days are rounded up to the nearest integer. For example, a submission that is 4 hours late will count as one day late.
  - When you hand in a late assignment, you must identify at the top of the assignment, (i) how late this assignment is, and (ii) how much of the total slip time you have left.

- After you have used up your slip time, any assignment handed in late will be marked off 20% per day. That is, after 5 days, the grade will be zero.
- **Extreme circumstances, like medical emergencies, etc.:** In such cases, additional, no-penalty extensions will be granted. Contact the instructor, with some written documentation (like doctor's note).

## Honor Code

This is a graduate level class---hence I expect students to want to learn and not google for answers. The purpose of problem sets in this class is to help you think about the material, not just give the right answers. As we often reuse problem set questions from previous years, covered by papers and webpages, I expect the students not to copy, refer to, or look at the solutions in preparing their answers.

That said, Georgia Tech aims to cultivate a community based on trust, academic integrity, and honor. Students are expected to act according to the highest ethical standards. For information on Georgia Tech's Academic Honor Code, please visit <http://www.catalog.gatech.edu/policies/honor-code/> or <http://www.catalog.gatech.edu/rules/18/>. Any student suspected of cheating or plagiarizing on a quiz, exam, or assignment will be reported to the Office of Student Integrity, who will investigate the incident and identify the appropriate penalty for violations.

Homework assignments are strictly individual efforts, while projects can be done by small groups (3 people or less) or individuals. You are expected to do your own work. No one may give you answers to homeworks. In other words, students are encouraged to communicate about general principles of the course, but all assigned homework must be done on an individual basis. You may not exchange any code or solutions, either in pieces or in entirety, by any electronic means or hard copy.

## Attendance and/or participation

Participation is important for this course and learning in general. We use posting on piazza (piazza discussion links are listed on the home page) are as the proxy to measure the participation level.

## Student-Faculty Expectations

At Georgia Tech we believe that it is important to continually strive for an atmosphere of mutual respect, acknowledgement, and responsibility between faculty members and the student body. See <http://www.catalog.gatech.edu/rules/22/> for an articulation of some basic expectations - that you can have of me, and that I have of you. In the end, simple respect for knowledge, hard work, and cordial interactions will help build the environment we seek. Therefore, I encourage you to remain committed to the ideals of Georgia Tech, while in this class.

# Accommodations for Students with Disabilities

If you are a student with learning needs that require special accommodation, contact the Office of Disability Services at (404)894-2563 or <http://disabilityservices.gatech.edu/>, as soon as possible, to make an appointment to discuss your special needs and to obtain an accommodations letter. Please also e-mail me as soon as possible in order to set up a time to discuss your learning needs. ~

## Scribe Notes

Each student will need to scribe 1 lecture in LaTeX. As there is no established textbook for this course, we will post these notes on the course website for the benefit of all students.

1. Please take careful notes as you hear the lecture and any relevant in-class discussions. Also read the suggested reading associated with the lecture.
2. Prepare your notes into a document that is written out in complete prose sentences that would be understandable even by a student who might have missed class, not merely a transcription of the telegraphic notes that you jotted down during class. You should write the notes yourself in your own words. It is okay to refer to the readings; however, directly copying from these or any other sources is a form of plagiarism and should be avoided. Do not include administrative information (such as due dates for homeworks) in your scribe notes. Feel free to contact me if you have any questions about anything covered.
3. Scribe notes should reflect a good understanding of the material covered in class, should be clear, precise, and well-formatted, should be free of grammatical errors and typos, and follow standard stylistic conventions for technical writing.
4. Use the following LaTeX template: **SOURCE** (`./template.tgz`) (and how the **PDF** (`./template/template.pdf`) will look).
5. Email the instructor and the TAs the scribe notes (i.e. the latex source, pdf and any needed bib and figure files). Please email us this within 7 days of your assigned lecture e.g if you are the assigned scribe for Tuesday we should get your scribe notes before start of lecture time next Tuesday. We will review it and you should get back to us the revised notes within 2-3 days. Hence we expect that on day 10/11 of the lecture the scribe notes will be done and posted on the website.

## Project

Since project is an important part for this course, it has to be substantial. It can be done in groups of 2-3. The project can be:

- Theoretical: considers a model (can be novel), or an algorithm, or a metric and derives a rigorous result about it (i.e. tighter bounds, surprising properties etc.).
- Experimental: evaluation of different algorithms and models on an interesting dataset(s)

- Extension: an extension or improvement of a method or model covered in class to a different or more general setting (eg. time-varying, distributed, any-time, scalable, etc.) and experiments that justify the new proposal.

The project can not be only a survey. And we will follow the no 'double-dipping' policy.

## Deliverables

The deliverables include:

1. Project Proposal: 15%, Due Date: TBD
2. Project Milestone Report: 10%, Due Date: TBD.
3. Final Report: 20%, Due Date: TBD
4. Final Presentation in class: 5% on TBD

All the write-ups should be in the **ACM SIG format** (<https://www.acm.org/publications/proceedings-template>), in either the Latex (preferred) **one** (<https://www.overleaf.com/latex/templates/association-for-computing-machinery-acm-sig-proceedings-template/bmvfhcdnxfty>) or Word **one** ([https://www.acm.org/binaries/content/assets/publications/word\\_style/interim-template-style/interim-layout.docx](https://www.acm.org/binaries/content/assets/publications/word_style/interim-template-style/interim-layout.docx)).

## Project Proposal

The proposal should contain a detailed survey of the related work (at least 6-8 papers, outside of the required class reading list) and identify what are strengths and weaknesses of the papers and how they may be addressed. You should be thinking how these papers are interrelated and at the same time different from each other. You should not just copy the abstract of any paper: that would be plagiarism.

The proposal should then focus on describing the proposed research directions and questions. How precisely do you plan to pursue them? What methods/data do you plan to use? A useful guide as to what a proposal should answer, see **Heilmeier's Catechism** ([http://en.wikipedia.org/wiki/George\\_H.\\_Heilmeier#Heilmeier.27s\\_Catechism](http://en.wikipedia.org/wiki/George_H._Heilmeier#Heilmeier.27s_Catechism)). In addition to the survey, the proposal should contain at least some amount of each of the following types of content:

- What is the problem you are solving? Give the formal problem definition (in addition to a lay-person version).
- Which algorithms/techniques/models you plan to use/develop? Be as specific as you can!
- How will you evaluate your method? How will you test it? How will you measure success?
- What data will you use (how will you get it)? Give data specifics (eg. size, format, etc.).
- What do you expect to accomplish by the end of the semester? (eg. novel algorithm, parallel implementation, etc.)

- You must describe what portion of the project each team member will be expected to do. Include an expected time-line of activities.

Your proposal should be self-contained. For example, don't just say: "We plan to implement John Doe's Foo-graph algorithm [Doe2001], and we will study its performance with our approach." Instead, you should briefly review the key ideas in the references, and describe clearly the alternatives that you will be examining.

The proposal should be **3-4** pages in the given format, with pictures if they seem useful (more than 4 pages won't be read). Check the grammar and syntax (there will be a small penalty for each typo/grammar error---please **do not** submit without a **spell check**). Include the names and email-addresses of the group members. Submit the PDF on canvas via one of the team member's account.

### Project Milestone Report

Think of this as a draft of your final report but without your major results. We expect that you have completed 30% of the project. Provide a complete picture of your project even if certain key parts have not yet been implemented/solved. Include the parts of your project which have been completed so far, such as:

- Thorough introduction of your problem
- Review of the relevant prior work---please feel free to add more references and related work you may have read.
- Description of the data collection process.
- Description of any initial findings or summary statistics from your datasets.
- Description of any mathematical background necessary for your problem.
- Formal description of any important algorithms used.
- Description of general difficulties with your problem which bear elaboration.
- Make sure to at least outline the parts which have not yet been completed so that it is clear specifically what you plan to do for the final version.
- **Important:** Respond to the comments made in the graded proposal. Include a section after the introduction which addresses each comment and mentions what changes have been made in the milestone.

The milestone report should be **4-5** pages in the **given format**, with pictures if useful. Submit the PDF on canvas via one of the team member's account.

## Project Final Report

This should be a detailed description of what you did, your results and what have you learned and/or conclude from your work.  
COMPONENTS:

1. **Write-up:** A minimum of 5 pages, a maximum (hard-limit) of 8-pages in the given format.
  1. [2%] Introduction/Motivation.
  2. [3%] Problem Definition.
  3. [5%] Related work and Survey.
  4. Proposed Method
    1. [10%] Intuition: why should it be better than the state-of-art?
    2. [25%] Description: your approach, algorithms, models. Be as clear as you can (as otherwise we won't understand what you are trying to do).
  5. Experiments/Results
    1. [5%] List of questions your experiments are designed to answer, description of your testbed
    2. [30%] Details of your experiments, observations, findings. Make sure you also **interpret** and **explain** your observations.
  6. [5%] Conclusion and Discussion. Feel free to mention any avenues of future work here.
2. **Software:** [10%] packaging, documentation, and portability. The goal is to provide enough material, so that other people can use it and continue your work. Create a tar.gz file which contains:
  1. A concise, short README.txt file, corresponding to the "user's manual". This file should describe the package in a few paragraphs, how to install it, how to use it, and how to run a demo.
  2. A DOC directory, with your writeup, your presentation slides. All your code should be in a SRC directory.
  3. Make sure that your package includes only the **absolutely necessary** set of files! Do NOT just make a 20MB core-dump of all your files and submit!
3. **Web-page:** [5%] create a webpage which should contain the title of the project, the names of the members (with portrait pictures preferably), and links to a PDF version of your write-up, your software tar-ball files, and your presentation slides. Also have a short summary of your project and its results on the webpage.

HOW TO SUBMIT:

- Submit the PDF on canvas via one of the team member's account. Important: Please keep the graded phase 1 and 2, and make sure to attach a copy (of the graded proposal and milestone) to the PDF of your final report.
- Upload your presentations to Canvas as well.
- Finally, upload your software tar-balls on your webpage.

Please note the times, as all are **hard deadlines**, with **no extensions or slip-days** allowed.

## Project Final Presentation

The project presentations will be on Dec 2. The presentation order will be posted on Piazza. 12mins per team---time limits will be strictly enforced. Plan for 10mins for presentation, and 2 mins for questions and transition to the next team. **Practice** your timing and delivery---giving a good talk is hard! Grading scheme:

- [20%] Motivation/Introduction and Problem Definitions: remind us what you're doing, why it's important and why we should care.
- [20%] Your proposed approaches (algorithms, models, analysis): what it is, its intuition, why does it work, etc. Have enough technical details so we can understand what you are doing.
- [10%] What's your data: where you got it, what's its characteristics (e.g., size on disk, # of records, temporal or not, etc.)
- [20%] Experiment and results: how did you evaluate your approaches? What are the results? How do your methods compare to other methods (if any)?
- [10%] Conclusions (and optionally future work/discussion)
- [20%] Presentation and Delivery (e.g., did you practice? good slides with pictures?)

All students are expected to attend the lectures, and be prepared to ask (tough!) questions to other project groups.

Last updated by B. Aditya Prakash on: Sun Aug 24 2025 23:12:51 GMT-0400 (Eastern Daylight Time)