

Video 1: Welcome to Computational Statistics with R

Stats 102A

Miles Chen, PhD

Highlights from the Syllabus

Welcome!

- My name is Miles Chen
- You can call me: Miles, Professor Chen, or Dr. Chen, whatever name you are comfortable with. I don't like being called by my last name only (e.g. "Hi Chen!")

Grade Breakdown

- 15% Video Quizzes
- 36% Homework (Up to 6 assignments, none are dropped)
- 4% Campuswire Participation
- 20% Midterm Exam
- 25% Final Exam

Grading

Letter grades are assigned on a straight scale as follows:

- 59.9 and below: F
- 60.0 - 69.9: D
- 70.0 - 76.9: C, 77.0 - 79.9: C+
- 80.0 - 82.9: B-, 83.0 - 86.9: B, 87.0 - 89.9: B+
- 90.0 - 92.9: A-, 93.0 and up A, top 5% of students: A+

I do not curve grades. Do not ask me to round your grade. Those emails will be ignored.

Video Quizzes

Videos will be accompanied by short quizzes. Please be sure to take the quizzes.

Homework

- 10 minute grace period.
- 5 point deduction for each hour assignment is late, but no deductions after midnight. Deductions resume 6pm following day. More details in syllabus.
- A 72-hour extension granted if documentation is submitted with the homework. **No need to contact the professor if you will include documentation with the homework submission.**
- If you need an even longer extension, please visit professor in office hours.

Office hours:

Office hours are my preferred method of contact.

Questions and issues are generally resolved much more quickly via office hours.

When you come to office hours, please **introduce yourself**. Say “Hi Miles, I’m Joe Bruin.” Do this **every** time you visit me until I start calling you by your name.

I like when students come to office hours with questions about material. I love to explain things and to help students understand.

I like when students come to office hours to tell me more about themselves and to seek counsel about classes to take or next steps. I am happy to make accommodations for students who face difficult circumstances and may need extensions for assignment deadlines. Please do not hesitate to visit office hours.

I am happy to correct grading mistakes. I do not want to get in arguments with students over points. I do not like arguing whether a particular mistake should be a 5 or 10 point deduction.

Campuswire for questions. Office hours for requests.

Post your questions on Campuswire. You will likely get a quick response from classmates.

If it is a question you don't want public, you can DM me on Campuswire. You are likely to get a faster response via Campuswire than email. Keep your messages short.

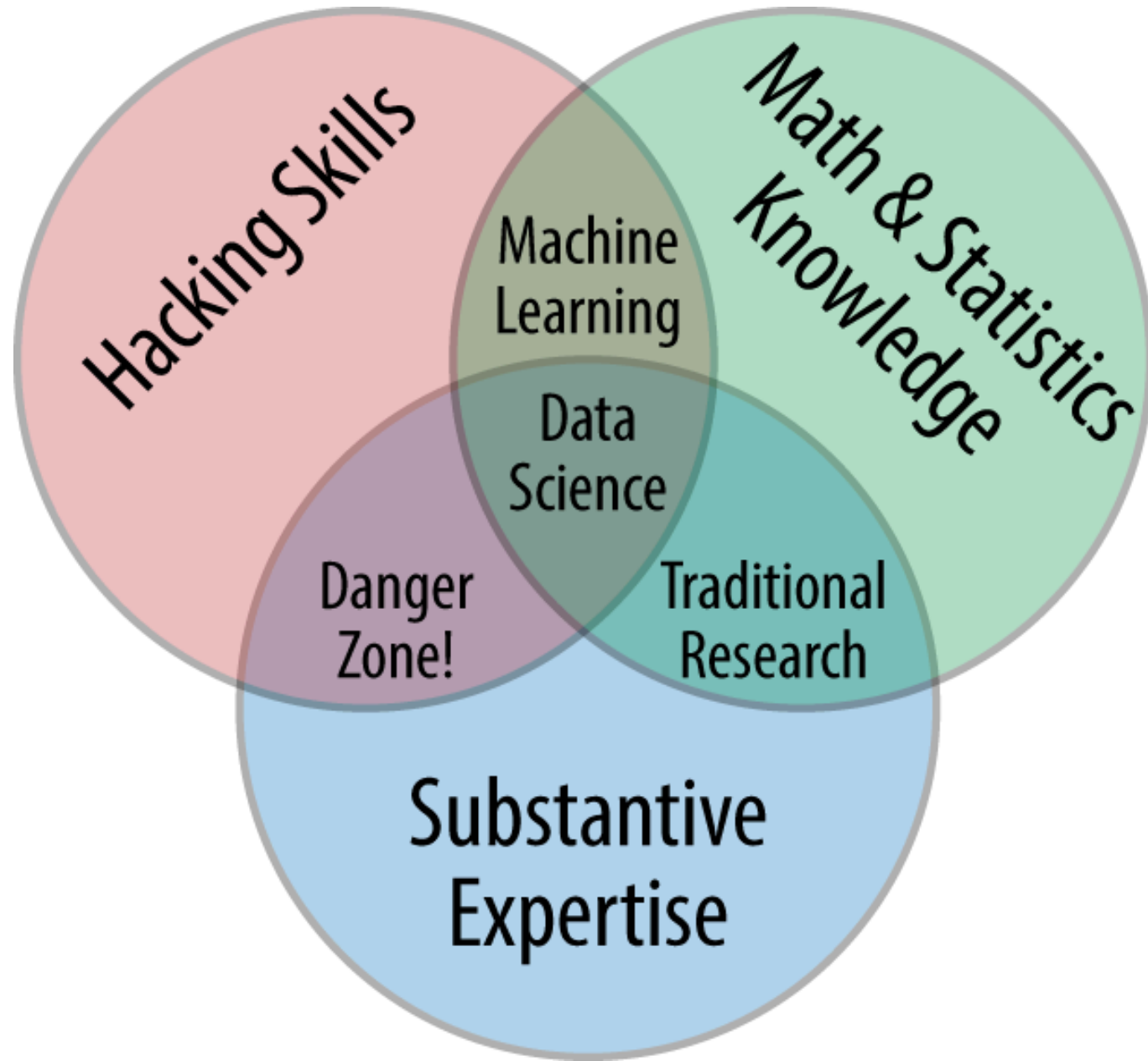
If you need me to do something - e.g. grant an extension, schedule a make up, change a grade, etc. - please come to office hours.

What is this class about?

Foundations for Data Science

Drew Conway's Data Science Venn Diagram consists of

- Coding Skills
- Math & Stats Knowledge
- Domain Expertise



Drew Conway's Data Science Venn diagram

This course

- Stats 102A is an introduction to the field of computational statistics and data science.
- This course aims to prepare you for data science by further developing your coding skills in R and applying those skills to some statistics applications
- The course can be thematically split into two parts.
 - Advancing your programming and coding skills
 - Introducing the methods of computational statistics
 - Some parts of the course have heavy overlap with Math 151A. Those weeks are like an “light” version of Numeric Methods that are useful for Stats majors (who aren’t required to take Math 151A)

What we'll cover in lecture:

Programming and Coding

- R Data structures
- subsetting
- Flow control
- writing functions
- scope, environments
- Web scraping
- tidyverse: tidyr and dplyr
- regular expressions
- object oriented programming
- ggplot

What we'll cover in lecture: Methods of Computational Statistics

- floating point arithmetic
- Root finding methods
- Numerical optimization methods
- Randomization tests
- Bootstrapping

What you will (need to) develop outside of lecture

I can't "teach" you the following skills. You develop them with practice.

- reading / writing code
- debugging code
- problem solving (decomposing a problem and thinking of a solution using the available coding tools)
- reading documentation

The homework assignments are intended to be structured exercises that will help you develop these skills. Approach the homework intentionally with the goal of developing these skills, rather than 'just trying to get it done.'

Think of this as a “Coding II” class

- To graduate from UCLA, you have to take a Writing II course. When you take a Writing II class, you can expect to write a lot of essays and papers. This is what UCLA says about the writing requirement

Writing is essential to thinking and learning. Clear, concise writing is a key strength for an academic career. Students who write well can lead expressive lives and create powerful communications about their ideas and opinions.

- You can think of this class as something like a “Coding II” class. You can expect to write a lot of code for your homework assignments. This is my modification for why coding is important

Coding is essential to interacting with and learning from data. The ability to write clear, readable code is a key strength for a data-related career. Students who code well can create powerful tools to implement their ideas and solutions.

Academic Integrity and Plagiarism

Let's talk about Plagiarism

Some truths:

- There is a lot of high quality code that does exactly what you need available out on the Internet. Some of it is available in ready-to-install packages and some are available as solutions on places like stackexchange and github.
- There are AI tools that can help solve programming tasks. The AI tools are incredibly advanced and can solve many of the homework prompts.
- If the goal is to accomplish a task, you can use the readily available packages, code solutions, and AI tools to get the job done.
- However, the goal of the homework is not to accomplish some task. The goal of the homework is to give you an opportunity to practice and develop the skills needed to write code.

No Pain, No Gain

Think of the gym. The goal of lifting weights at the gym is not that the weights get picked up. Lifting weights is a means to the real goal of gaining strength.

“No Pain, No Gain”: if your weight training does not result in some muscle soreness, you probably did not exert enough effort to expect muscle gain. Experiencing muscle soreness is a sign that your muscles will go through repairs and get stronger.

Your brain is similar: if your brain does not struggle when writing code, then it has no reason to create additional neuron connections that will improve your abilities as a coder. On the other hand, if your brain struggles with writing code, then your brain will try to create new connections between neurons so the next time will not be as hard. And thus you become a better coder.

No Pain, No Gain

Plagiarizing code or asking an AI to solve a difficult problem is like having a stronger person lift the weights that are too heavy for you.

This would be a good solution if the goal of lifting weights were to lift the weights. But it does not help achieve the goal of gaining strength.

Copying, pasting, and modifying someone else's (or an AI's) code works if the goal is to accomplish a coding task. It does not help towards the goal of creating neuron connections in the brain that will make you a better coder.

Course Goals

I believe students resort to plagiarism because they have confused the goals of the course.

Students who plagiarize believe the goal for them is to get the homework done for the sake of finishing it (and avoid a bad grade). For these students, the goal of learning is secondary to the goal of avoiding bad grades.

But this is wrong! The goal of the course is your learning.

I will admit, a major conflicting issue here is that I am not able to create individualized grading schemas that evaluate exactly how much each student learned over the course. All students are graded on the same criteria and evaluated on what they turn in for the assignments.

That said, I hope you can judge your performance in class based on what you learned and not your letter grade.

My expectations

When you face a challenging homework assignment, I expect

- you put forth a good faith effort
- you not to seek out solutions online, from an AI, or view another (current or former) student's code
- if you are not able to complete everything required by the assignment by the deadline
 - you submit what you have and accept a grade that is less than 100%
 - you view this not as a failure of your coding abilities, but as a indication of areas for growth and improvement

I (and the statistics department) take issues of plagiarism seriously and will escalate cases to the Dean of Students. Full details regarding academic integrity are in the syllabus.

Collaboration Policy

Read the course collaboration policy and be sure you understand it.

For all homework assignments, verbal collaboration but no code sharing.

You are allowed to collaborate verbally with other students but you are not allowed to look at or show someone else the code you are writing. This applies to discussions on Campuswire.

Allowed vs Not allowed

Question: “How did you approach problem 2?”

Allowed response: “I created a for loop and within each iteration I subset the x vector to the desired values and then used the sort function on the result. Be sure to assign the results to update the output object.” “Thank you! I’ll be sure to note your help.”

Policy Violating response: “Let me show you what I did... [proceeds to show screen with code, or actually shares code via email on Campuswire]”

Allowed vs Not allowed

Question: “Can I see how you did problem 2 to double check my work?”

Allowed response: “I can’t show you my code but I can tell you what I did.”

Policy Violating response: “Let me show you what I did...
[proceeds to show screen with code, or actually shares code via email or Campuswire]”

Allowed vs Not allowed

Question: “Can you help me find what I’m doing wrong? I’ve got a bug but I can’t figure out where. It keeps saying ‘missing value where TRUE/FALSE needed’”

Allowed response: “Did you check to make sure?”

Allowed response: “That error message often pops up if you have an NA inside an if statement.”

Policy Violating response: “Let me see your code... [proceeds to look at code]”

You are encouraged to discuss code that is not part of an assignment!

This is a coding class! As long as the code is not part of a homework assignment, you can post and discuss code.

You can always post and discuss code that appears in lecture.

You are encouraged to modify the examples that appear in lecture and discuss the effect of each change you make.

You can post and discuss code that is for the purpose of learning a particular concept or how a function works.