UC San Diego

BILD 5
Data Analysis and Design for Biologists
Winter 2026

Instructor: Mingyu Yang, Ph.D.
ymy@ucsd.edu
Lecture: MWF 9:00-9:50am
Catalyst 0125
Office hours: See Canvas homepage

Introduction

Welcome!
Welcome to BILD 5! As your professor, I care deeply about fostering a joyful, inclusive classroom climate. Our classroom should be a place where everyone is respected, feels safe to make mistakes, and lifts one another up. Meeting with students is my favorite part of the job, and I look forward to getting to know you all - you can reach me through discussion sections, office hours, and Piazza. College is a time of tremendous growth, both inside and outside the classroom, and I'm excited to support and celebrate your growth this quarter!

Course description
Introduction to information literacy, experimental design, and data analysis for biologists, including coding, data management, and quantitative analysis. Practical skills related to effectively asking and answering biological questions with data.

Course learning objectives
1. Create testable hypotheses addressing valid biological questions.
2. Evaluate the credibility and value of relevant scientific information.
3. Design experiments that effectively test hypotheses.
4. Construct figures that effectively communicate data.
5. Perform appropriate quantitative and statistical analyses on experimental data.
6. Interpret the results of quantitative statistical models and associated analyses.
7. Utilize the R programming language for scientific data analysis and graphing.
8. Combine the elements of a complete investigative cycle in a student-designed project.
9. Explore the modern intersection between different subfields of biology, technology, and data science.
10. Examine the ethical responsibilities of scientists when creating and communicating scientific evidence.

Course components

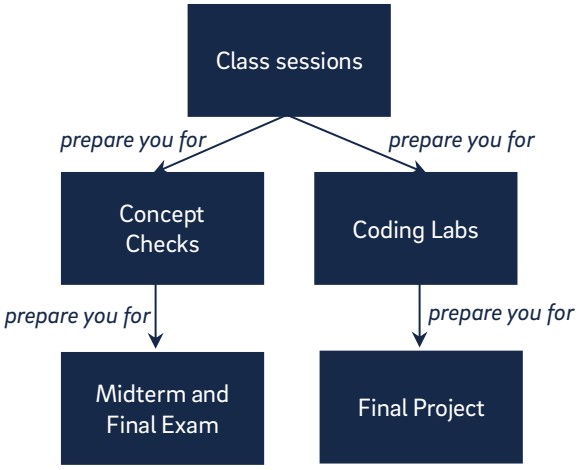
Class sessions
Class time will be a combination of lecture and individual/group problem solving. To participate in class polling, you'll need either a physical iClicker or an electronic device registered with iClicker Cloud. For every class in which you actively participate (responding to >75% iClicker questions), you will receive 1 course point, up to a maximum of 20 points. This allows you to miss approximately two weeks of class without penalty. iClicker points are graded for participation and not accuracy. Each class session will be podcasted and posted on the Canvas Media Gallery.

Concept Checks

There will be an online Concept Check due most Mondays at 9:00am. Some of the questions will be fairly simple to make sure you understood the basic ideas from class, and others will be more difficult exam-style questions. All Concept Checks are **open-book**, **open-collaboration**, and have no time limit – you can and should work together with your peers! You will only have one attempt per Concept Check, but at the end of the quarter we will **drop your lowest-scoring Concept Check** when calculating your final course grade.

Coding Labs

There will be a **Coding Lab** due most Fridays at 9:00am. In Week 5, you'll have an opportunity to resubmit a Coding Lab from a previous week to improve its grade. Additionally, at the end of the quarter we will **drop your lowest-scoring Coding Lab**. During discussion section, you'll work through the Coding Lab with the support of the instructional team. Because of space restrictions, you should only attend the discussion section for which you are registered.



Final Project

BILD 5 culminates in the **Final Project**, where you'll go through an entire investigative cycle on your own, from conceptualizing a research question to analyzing and communicating the data. Think of the Final Project as a 'mega lab', where you'll bring together all the skills you've learned to a dataset that is uniquely yours. The Final Project is due on Friday 9:00am of Week 10 (taking the place of the Coding Lab that week). There will also be Final Project Checkpoints to give you feedback throughout the process.

Exams

There will be one Midterm Exam and one Final Exam. The exams will assess your *conceptual understanding* of the material, not your ability to write code from scratch (although some of the questions will ask you to *read* code and/or *interpret* code output). The exams are **in-person** and **closed-collaboration**; you may bring **one single-sided** 8.5 x 11" aid sheet for the Midterm, and **one double-sided** 8.5 x 11" aid sheet for the Final. We will collect your aid sheets after each exam, so you should scan them if you'd like to keep them. Exceeding the page limit or attempting to attach additional paper/flaps onto your sheet will result in a 50% deduction from your exam score. If your percentage score on the Final is higher than that of your Midterm, then **your score on the Final will replace your Midterm score**. Because of this policy, we will not offer a make-up Midterm. This policy is also meant to accommodate for missing the Midterm due to illness – the last thing I want is for you to feel pressure to study while you're ill!

Office hours

Office hours are an opportunity to clarify course content and also to chat more broadly about anything you want. You do *not* need to come with any questions prepared; in fact, listening to everybody else's questions can often help to resurface questions that you didn't even realize you had. Most office hours will be held in a conference room or classroom, where you can pop in or out at any time. Right before assignments are due, office hours tend to become quite busy – should you find yourself waiting to receive help, I encourage you to **help one another**! I take notice of and remember the students who support each other – this really means a lot to me. Outside of synchronous office hours, you can also ask questions on the course **Piazza** forum.

Assessments and grading

The course is graded out of 320 points total. Your total score will be converted to a letter grade based on the following absolute scale. This class is **not curved**; therefore, you are not in competition with each other, and in theory everyone can earn an A grade. Grades will not be rounded; instead, I'll provide everyone a few opportunities for extra credit that *may* 'round you up' if you're close to a grade boundary. I request that you don't ask for additional grade bumps or extra credit opportunities beyond

those that are already provided for the entire class – it wouldn't be equitable if I give you an extra credit opportunity that isn't available to the rest of the class.

A+ ≥ 314	A ≥ 298	A- ≥ 288
B+ ≥ 279	B ≥ 266	B- ≥ 256
C+ ≥ 247	C ≥ 234	C- ≥ 224
	D ≥ 192	
	F < 192	

Grade breakdown

Class participation	(1 pt each, 26 total, drop 6)	20
Concept Checks	(10 pts each, 9 total, drop 1)	80
Coding Labs	(10 pts each, 8 total, drop 1)	70
Final Project	(Includes two checkpoints)	30
Midterm Exam		40
Final Exam		80
Total		**320**

Final Exam grade replacement policies

The goals of the following policies are twofold: to (a) incentivize sustained effort throughout the quarter and (b) reward improvement on the Final. As such, there are two mechanisms for your Final Exam score to replace other components of your grade. These work independently of each other, i.e., you might end up taking advantage of one, both, or neither. These policies are intended as safety nets and should not replace your efforts to engage consistently with all components of the course.

1. If your percentage score on the Final Exam is higher than your Midterm score, then your Midterm score will be replaced by your percentage on the Final. For example, if your Midterm score is 20/40 (= **50%**), and your Final score is 72/80 (= **90%**), then your Midterm score will be replaced with 36/40 (= **90%**).

2. If your percentage score on the Final Exam is higher than your participation score, then your participation score will also be replaced by your percentage on the Final. That said, you should still make every effort to attend class, and treat class participation as 'free points' that can reduce the weight of the Final.

Course policies

Academic accommodations

If you need academic accommodations due to a disability, please send Ming a current Authorization for Accommodation (AFA) letter issued by the Office for Students with Disabilities (osd.ucsd.edu). We'll schedule accommodations for the Midterm and Final exams directly with the Triton Testing Center.

Late policy

You will be allotted four "late tokens" to use throughout the quarter. Each late token extends the deadline of any assignment (Concept Check or Coding Lab) by 24 hours. **You can only use one late token per assignment, and you cannot use a late token on the Final Project**. Late tokens are also intended to cover any technical mishaps you might encounter; please don't leave any assignments to the last minute, because even a 9:01am submission will incur a late token. The reasons for this strict cut-off are two-fold: (1) we need to be fair to all students, therefore we have to draw the line somewhere (2) we want to disincentivize missing class to work on an assignment right before/after the deadline.

We'll automatically apply late tokens at the end of the quarter, and **you do not need to ask to use one**. It's your responsibility to keep track of how many late tokens you have left. If you run out of tokens and submit an assignment late, a grade will still

appear on Canvas. However, at the end of the quarter, tokens will be allocated to the assignments that benefit your overall score the most. Then, any remaining late assignments will have their grade replaced with a zero.

Academic integrity and artificial intelligence

By upholding high standards of integrity in this class, you can be confident that every student is fairly assigned the grade that they earned. This means you should never misrepresent somebody else's work as your own; for example, during class time, you should only click in for yourself and not your peers. In addition, although Concept Checks are collaborative and open-resource (including artificial intelligence tools), your submission should ultimately represent your own effort. You should be able to defend all your answers, so please do not blindly copy from your peers or from the Internet. Getting into bad habits early on will prepare you poorly for exams, which do not permit collaboration or use of the Internet.

For all Coding Labs, you are welcome to **discuss and share R code with your classmates**, and you can **freely use artificial intelligence tools**. However, we'll also ask you to annotate your code to describe what it's doing; even if you copied code from a different source, your **explanation of that code must be your own**. Just blindly copying code from other sources will prepare you poorly for life beyond BILD 5, in addition to severely disadvantaging you on the Final Project, where you'll receive a unique dataset that you'll analyze differently from everybody else in the class. Ultimately, we trust you to succeed with integrity, and we have no interest in policing your behavior. Please return our trust by behaving with honesty.

Anonymous feedback

The class Canvas page links to an anonymous feedback form that will be open throughout the quarter. Only Ming has access to this form, and there will be no way to trace the response back to you. We especially encourage you to complete the form if you had an experience that was demeaning, intimidating, offensive, or threatening in any way. If I said anything in class that made you feel this way, this is also a good place to tell me (although I welcome non-anonymous feedback too!). Remember, my email inbox and 1:1 office hours are always open to you, if you want me to follow-up directly with you.

Extenuating circumstances

I sincerely believe that everyone can succeed in this class. However, extenuating circumstances may arise that negatively impact your ability to participate in the course (for example, needing to miss more classes than the drop policy accommodates for). If you find yourself needing additional accommodations or support, please reach out to your college's Dean of Student Life, and I will work together with both you and your Dean to plan a path forward. There are two reasons for this policy: (1) If you find yourself needing additional accommodations because of an extenuating circumstance, this is likely affecting your participation in all your classes, not just this one. Your Dean of Student Life can help coordinate with *all* your course instructors, so you won't have to figure it all out alone. (2) As your professor, I don't think it's right for me to be the arbiter of who does or does not 'deserve' additional accommodations – I think this responsibility should be for someone whose full-time job (and professional training) is in student support. I will commit to putting my best-faith effort to working with your Dean to ensure you receive the support and flexibility you need.

Parting remarks

The main goals of this course are for you to learn a lot, build a class community, and have a wonderful experience while doing so. Welcome to BILD 5!

Schedule of topics

(subject to change)

Week	Date	Class	Assignments
01	Jan-05	01 Scientific Reasoning	**Coding Lab 01** *(due Fri 9:00am)*
	Jan-07	02 Getting Started in R	
	Jan-09	03 Data Frames in R	
02	Jan-12	04 Data Types and Distributions	**Concept Check 01**: Classes 01-03 *(due Mon 9:00am)*
	Jan-14	05 Graphing in R	**Coding Lab 02** *(due Fri 9:00am)*
	Jan-16	06 Tidying Data in R	
03	Jan-19	**No class: Martin Luther King Jr. Holiday**	**Concept Check 02**: Classes 04-06 *(**Wed** Mon 9:00am)*
	Jan-21	07 Center, Spread, Normality	**Coding Lab 03** *(due Fri 9:00am)*
	Jan-23	08 Statistics and Sampling	
04	Jan-26	09 Confidence Intervals	**Concept Check 03**: Classes 07-08 *(due Mon 9:00am)*
	Jan-28	10 Data Transformations	**Coding Lab 04** *(due Fri 9:00am)*
	Jan-30	11 Biological to Statistical Hypotheses	
05	Feb-02	12 Comparison of One Mean	**Concept Check 04**: Classes 09-11 *(due Mon 9:00am)*
	Feb-04	13 Comparison of Two Means	**Coding Lab Resubmission** *(due Fri 9:00am)*
	Feb-06	14 P-Values, Effect Size, Error	**Concept Check 05**: Classes 12-13 *(due **Fri** 9:00am)*
06	Feb-09	**Midterm Exam: Classes 01-13**	**Coding Lab 05** *(due Fri 9:00am)*
	Feb-11	15 Power Analysis	
	Feb-13	16 Multiple Comparisons	
07	Feb-16	**No class: Presidents' Day Holiday**	**Concept Check 06**: Classes 14-16 *(due **Wed** 9:00am)*
	Feb-18	17 Analysis of Variance	**Final Project Checkpoint 1** *(due **Wed** 9:00am)*
	Feb-20	18 Data Transformations, Revisited	**Coding Lab 06** *(due Fri 9:00am)*
08	Feb-23	19 Correlation	**Concept Check 07**: Classes 17-18 *(due Mon 9:00am)*
	Feb-25	20 Simple Linear Regression	**Coding Lab 07** *(due Fri 9:00am)*
	Feb-27	21 Simple Linear Regression, Continued	
09	Mar-02	22 Data Transformations, Concluded	**Concept Check 08**: Classes 19-21 *(due Mon 9:00am)*
	Mar-04	23 Comparison of Categorical Data	**Final Project Checkpoint 2** *(due Mon 9:00am)*
	Mar-06	24 Experimental Design	**Coding Lab 08** *(due Fri 9:00am)*
10	Mar-09	25 Replication and Sample Size	**Concept Check 09**: Classes 22-24 *(due Mon 9:00am)*
	Mar-11	26 Measurement Validity and Controls	**Final Project** *(due Fri 9:00am)*
	Mar-13	27 Ethics and Your Future	
11	Mar-18	**Final Exam: Comprehensive**	