

ECOL 596W Section 001
Practical and Reproducible Data Science for EEB
Tuesdays and Thursdays 9:30 – 10:45 AM
Koeffler 209
Fall 2024

Professor: Dr. Sabrina McNew (she/her/ella)

Email: mcnew@arizona.edu

Office: BSW 326

Office Hours: By appointment

Course Information and Communication:

1. The course website is https://github.com/smcnew/ECOL_596W_2024
2. If you have questions or concerns feel free to email the instructor (mcnew@arizona.edu).
3. Tuesdays will be mostly lecture, Thursdays will be mostly in-class practice. Bring a computer with R and R studio to Thursdays, and some Tuesdays TBD.

Course Description

Data management and analysis skills are essential for graduate students in the biological sciences. However, these skills can be challenging to learn because they sit at the intersection of three fields: biology, math, and computer science. Here we will pull from each of these disciplines to build your toolkit as a researcher. In this course, we will develop general best analytical practices to ensure that your data are safely stored, that your results are reproducible, and that your research leads to meaningful insights. We will not cover every statistical approach that you may need to complete your graduate degree, and we will not deeply delve into the math behind most analyses. However, by the time you leave the course you should have a foundation that will help you tackle many common types of data analyses, and you should have the confidence to gain new skills to answer your specific research questions.

General Learning Outcomes:

By the end of the course you should be able to:

- *Manage your data:* You will learn how to “wrangle” your data into an analysis-ready format. You will learn best practices for storing and archiving your data to ensure reproducibility and easy collaboration with colleagues and future you.
- *Visualize your results:* Plot your data early and often. Visualizing your data is an important first step when organizing your analysis and sharing results with collaborators.

- *Use common statistical approaches*: Know what kind of data you have and what family of statistical approaches to use. You will gain a strong foundation in linear regression and its variations. You will also become acquainted with other techniques including PCA, Bayesian analysis, phylogenetic comparative methods, and more.
- *Learn to google*: Professional scientists face new data challenges with every project. Your goal is not to know every statistical approach but rather how to gain skills to tackle new problems.

Classroom Philosophy:

I'm excited you want to learn data science! This class is here to serve you as you work towards your PhD. As a group, our goal is to ensure that everyone feels included and free to engage fully with the content. We welcome:

- *Diverse identities and backgrounds* – all members of this class deserve to be here. We recognize and welcome the fact that our identities and experiences shape our perspectives and what we contribute to the classroom.
- *Respectful interactions* – that show value for others' opinions, identities, and time.
- *Commitment to shared learning* – recognizing that we each have different ways of demonstrating what we learn.

Prerequisites:

This course will be taught in R with R Studio. Prior knowledge of R is not a prerequisite; however, there is a learning curve to the language and the more comfortable you are with the code the easier this class will be.

Course texts/sources:

Irizarry, [Introduction to Data Science: Data Analysis and Prediction Algorithms with R](#)
Whitlock and Schluter, [The Analysis of Biological Data](#)
McElreath, [Statistical Rethinking](#)

Prior to the first week of class:

1. Download R and R Studio. If you need detailed instructions click [here](#).
1A. *Already downloaded but it's been a while? Time to update R and R studio*
<https://www.r-bloggers.com/2022/01/how-to-install-and-update-r-and-rstudio/>
2. If you are totally new to the R language please try one or more of these tutorials to get acquainted with how it works:
<https://swirlstats.com/>
<https://stat545.com/r-basics.html>
<https://datacarpentry.org/R-ecology-lesson/>

Week	Day	Topic	Class Activities/Problem Sets	Readings
Week 1 26-Aug	Tuesday	Class overview, philosophies of collaborative and reproducible science		
	Thursday	Intro to R: coding basics, dplyr	Irizarry Ch. 2	
Week 2 2-Sep	Tuesday	Principles of data management		Zuur et al. 2009
	Thursday	Data wrangling, summarizing	Irizarry Ch 4	
Week 3 9-Sep	Tuesday	Github 1	Before class: https://happygitwithr.com/ Chapters 4 - 6	
	Thursday	Github 2		
Week 4 16-Sep	Tuesday	Principles of Data Visualization	Before class: Send Sabrina an example of a good or a bad figure	Irizarry Ch. 9
	Thursday	Plotting in R	Irizarry Ch. 8	Irizarry Ch. 9
Week 5 23-Sep	Tuesday	Group Project 1		
	Thursday	Group Project 1		
Week 6 30-Sep	Tuesday	First looks at your data/Review		
	*Thursday	R Language Quiz		
Week 7 7-Oct	*Tuesday	Samples and uncertainty		
	Thursday	t-tests and variance		
Week 8 14-Oct	Tuesday	Bit o Bayes 1		Ioannidis 2005
	Thursday	Bit o Bayes 2		McElreath Statistical Rethinking Ch. 1-2
	Tuesday	Regression/Linear Models	Lms problem set	

Week 9 21-Oct	Thursday	lm()		
Week 10 28-Oct	Tuesday	Generalized Linear Models	GLMs/LMMs problem sets	Buckley 2015 Harrison et al. 2018
	Thursday	glm() lmer()		
Week 11 4-Nov	Tuesday	Model Selection Final Project Intro		Burnham and Andersen 2014 Valpine 2014 Murtaugh 2014
	Thursday	Causal Inference		
Week 12 11-Nov	Tuesday	No class		
	Thursday	Group Project 2		
Week 13 18-Nov	Tuesday	Group Project 2		Gould et al. 2023
	Thursday	Many Analysts Discussion		
Week 14 25-Nov	Tuesday	PCAs/ordination		
	Thursday	No Class: Thanksgiving		
Week 15 2-Dec	Tuesday	Present a Package		
	Thursday	Present a Package		
Week 16 9-Dec	Tuesday	Special Topics		
17-Dec		Final Project Due		

Assessment Breakdown

	Percent of final grade
Participation	20%
Group Projects	40%
Present a Package	20%
Quiz	5%
Final Project Peer Review	5%
Final Project	10%

Additional Resources:

Wickham and Golemund, R for Data Science. Guide focused mostly on coding in R.

Whitlock and Schluter, The Analysis of Biological Data. Good classic biostatistics reference, particularly for learning the math behind basic statistical tests.

McElreath, Statistical Rethinking. Definitive book on Bayesian statistics, great lectures available for free. Advanced.

General Information:

Classroom Behavior Policy: To foster a positive learning environment, students and instructors have a shared responsibility. We want a safe, welcoming, and inclusive environment where all of us feel comfortable with each other and where we can challenge ourselves to succeed. To that end, our focus is on the tasks at hand and not on extraneous activities (e.g., texting, chatting, reading a newspaper, making phone calls, web surfing, etc.).

Additional resources for students: UA Academic policies and procedures are available at <http://catalog.arizona.edu/policies>

Campus Health

<http://www.health.arizona.edu/>

Campus Health provides quality medical and mental health care services through virtual and in-person care.

Phone: 520-621-9202

Counseling and Psych Services (CAPS)

<https://health.arizona.edu/counseling-psych-services>

CAPS provides mental health care, including short-term counseling services.

Phone: 520-621-3334

The Dean of Students Office's Student Assistance Program

<http://deanofstudents.arizona.edu/student-assistance/>

Student Assistance helps students manage crises, life traumas, and other barriers that impede success. The staff addresses the needs of students who experience issues related to social adjustment, academic challenges, psychological health, physical health, victimization, and relationship issues, through a variety of interventions, referrals, and follow up services.

Email: DOS-deanofstudents@email.arizona.edu

Phone: 520-621-7057

Survivor Advocacy Program

<https://survivoradvocacy.arizona.edu/>

The Survivor Advocacy Program provides confidential support and advocacy services to student survivors of sexual and gender-based violence. The Program can also advise students about relevant non-UA resources available within the local community for support.

Email: survivoradvocacy@email.arizona.edu

Phone: 520-621-5767

Confidentiality of Student Records: Your educational record, including grades and other aspects of classroom performance are considered confidential and yours alone to disclose. More details below:

<http://www.registrar.arizona.edu/personal-information/family-educational-rights-and-privacy-act-1974-ferpa?topic=ferpa>

University-wide Policies link

Links to the following UA policies are provided here,

<https://academicaffairs.arizona.edu/syllabus-policies>

- Absence and Class Participation Policies
- Threatening Behavior Policy
- Accessibility and Accommodations Policy

- Code of Academic Integrity
- Nondiscrimination and Anti-Harassment Policy
- Subject to Change Statement