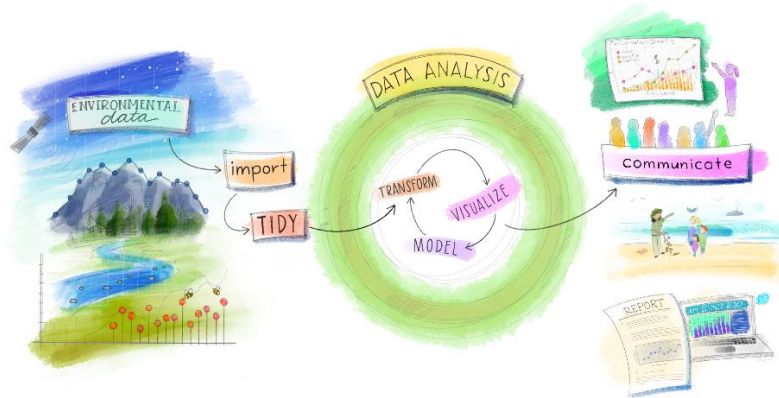


WILD 6900 – Reproducible Data Science

Syllabus



Artwork by Allison Horst

Credits: 3

Semester: Spring 2021

Schedule: Tuesday and Thursday 9 AM – 10:15 AM

Delivery format: Online, synchronous (on Zoom)

Instructor: Dr. Simona Picardi

Contact: simona.picardi@usu.edu

Office hours: Friday 11 AM – 12 PM or by appointment (on Zoom)

Course Description

Reproducibility of results is a pillar of the scientific method. In our day and age, scientists rely on software and code as fundamental tools to carry out their science and ensure it is reproducible. Being proficient in the use of programming tools and effectively apply them to store, process, manage, analyze, and visualize data have become must-have skills to take part in the scientific discourse and to be successful on the job market, academic or not.

In this course, students will learn best practices in data management and processing for reproducible science. Topics covered will encompass computational tools and techniques to effectively manage data throughout its life cycle, from the moment they get entered into a computer to the moment they are used in a published document. Throughout the course, the students will use a variety of software tools and programming languages to tackle different aspects of data management: spreadsheets for data entry, SQL relational databases for data management, R and the tidyverse for data cleaning, processing, analysis, and visualization, Git for version control of code scripts and data files, and the GitHub platform for code sharing and efficient collaboration.

Learning Objectives

The goal of this course is to help students establish their own workflow for reproducible data management that they can apply in their day-to-day research endeavors, by building long-lasting habits in efficient project organization, data management, and processing. Specific learning objectives include:

- Design functional project structures;
- Organize data in tidy format;
- Design databases that reflect real structure in the data;
- Write code to process data from raw to analysis-ready;
- Maintain code and data integrity with version control;
- Perform automated tasks on data.

Prerequisites

This course has no formal prerequisites. Prior experience with R is beneficial but not required. No experience required for any of the other tools and programming languages.

Teaching Philosophy

One of the major obstacles that students encounter when learning to program is that computational tools are taught incidentally within the context of a course that is centered on a different topic, often statistics. This means students have to split their cognitive load between learning the subject matter and learning the tools to practice it, which is a struggle. This course is designed as an opportunity for students to focus undisturbed on learning the tools. Another common misconception about learning computational tools is that the best way to learn is to be left alone to confront the code. But without prior instructions and explanations of how things work, being thrown to the wolves to tackle a programming language is nearly guaranteed to be a frustrating experience. I strive to apply evidence-based pedagogy techniques that are shown to be effective at facilitating learning. These include, among others:

- Setting explicit, measurable goals for each learning activity to facilitate self-assessment;
- Helping students build mental models of each topic before challenging them with practical problem-solving;
- Providing multiple exposure to each topic by combining traditional lectures, live-coding demonstrations, hands-on exercises, and group activities.

Instructional Methods

The course will be delivered entirely online. The format will include a combination of synchronous and asynchronous learning. Following a flipped-classroom approach, students will be expected to read the material on each week's topic on their own and then join a virtual class focused on exercises, live coding demonstrations, and group discussions. As a final outcome of the course, students will complete an individual project related to their own research (or any topic of their choice) that implements the techniques learned in class. Weekly lectures are organized chronologically in the order students are going to need them to work on their project, and each lecture will build off of previous ones.

Course Resources

All software used in this course is open source and freely available online. There is no required textbook. All reading material will be provided in the form of a digital book available on the course website.

Course Policies

Attendance

I will not be keeping track of attendance, but students are encouraged to regularly participate in class activities for best learning outcomes. Because the content of each class iteratively builds off of the previous ones, missing classes can set you back. At the same time, I understand the need for flexible schedules, especially during these times. The flipped classroom approach is meant to allow some flexibility and independent learning. However, a significant part of learning happens during live exercises, which I recommend not missing out on.

Assignments

The course will include weekly assignments and a final project to be handed in at the end of the semester. Many of the weekly assignments will be part of your final project, so working on them each week will make sure you are not left with a pile of work to do at the end of the semester. Because each assignment builds off of previous ones, it's best not to fall too far behind or it will be difficult to catch up. However, I will adopt flexible deadlines, with an automatic 48-hour extension to anyone who needs it (no need to request it or justify.) If you need more time on top of the 48-hour extension to complete an assignment, just let me know. You will not need to disclose personal information to get an extension.

Grading

Assignments will count as 50% of the final grade, active participation as 20%, and the final project as 30%. Final grades will be in accordance with the following standards:

Grade	Range
A	100 % to 93.0%
A-	< 93.0 % to 90.0%
B+	< 90.0 % to 87.0%
B	< 87.0 % to 83.0%
B-	< 83.0 % to 80.0%
C+	< 80.0 % to 77.0%
C	< 77.0 % to 73.0%
C-	< 73.0 % to 70.0%
D+	< 70.0 % to 67.0%
D	< 67.0 % to 60.0%
F	< 59.0 % to 0.0%

Zoom etiquette

During synchronous lectures on Zoom, kindly keep your microphone muted when you are not speaking. You are encouraged to turn on your video (it's always nice to see faces and get feedback from body language instead of teaching into the void), but you are not required to do so if you are not comfortable with it. Participation is welcome and encouraged at all times, either verbally or in the chat. To protect privacy of participants, live sessions will not be recorded.

Course Schedule

Week 1

- Introduction: what is reproducibility and why should we care?
- Project organization

Week 2

- Version control with Git
- Collaborative science with GitHub

Week 3

- Best practices in the use of spreadsheets
- Relational databases, pt. 1: Designing database structures

Week 4

- Fundamentals of SQL language
- Relational databases, pt. 2: Querying data across multiple tables

Week 5

- Relational databases, pt. 3: Building the database
- Interfacing SQL databases with R

Week 6

- Dynamic documents with Rmarkdown
- Automatically synchronized websites with GitHub pages

Week 7

- R programming, pt. 1: Review of base R
- R programming, pt. 2: Automating repeated tasks

Week 8

- R programming, pt. 3: Troubleshooting code
- R programming, pt. 4: Dependency management

Week 9

- R programming, pt. 5: Data wrangling with tidyverse

Week 10

- R programming, pt. 5 (continued): Data wrangling with tidyverse

Week 11

- R programming, pt. 6: Publication-quality figures with ggplot2

Week 12

- R programming, pt. 7: Introduction to geospatial data in R
- Problem decomposition

Week 13

- Mini-hackathon

Week 14

- Recap + AMA + Final project consultation

USU Policies

Academic Freedom and Professional Responsibilities

Academic freedom is the right to teach, study, discuss, investigate, discover, create, and publish freely. Academic freedom protects the rights of faculty members in teaching and of students in learning. Freedom in research is fundamental to the advancement of truth. Faculty members are entitled to full freedom in teaching, research, and creative activities, subject to the limitations imposed by professional responsibility. [Faculty Code Policy #403](#) further defines academic freedom and professional responsibilities.

Academic Integrity – "The Honor System"

Each student has the right and duty to pursue his or her academic experience free of dishonesty. To enhance the learning environment at Utah State University and to develop student academic integrity, each student agrees to the following Honor Pledge:

"I pledge, on my honor, to conduct myself with the foremost level of academic integrity."

A student who lives by the Honor Pledge is a student who does more than not cheat, falsify, or plagiarize. A student who lives by the Honor Pledge:

- Espouses academic integrity as an underlying and essential principle of the Utah State University community;
- Understands that each act of academic dishonesty devalues every degree that is awarded by this institution; and
- Is a welcomed and valued member of Utah State University.

Academic Dishonesty

The instructor of this course will take appropriate actions in response to Academic Dishonesty, as defined the University's Student Code. Acts of academic dishonesty include but are not limited to:

- Cheating: using, attempting to use, or providing others with any unauthorized assistance in taking quizzes, tests, examinations, or in any other academic exercise or activity. Unauthorized assistance includes:
 - Working in a group when the instructor has designated that the quiz, test, examination, or any other academic exercise or activity be done “individually;”
 - Depending on the aid of sources beyond those authorized by the instructor in writing papers, preparing reports, solving problems, or carrying out other assignments;
 - Substituting for another student, or permitting another student to substitute for oneself, in taking an examination or preparing academic work;
 - Acquiring tests or other academic material belonging to a faculty member, staff member, or another student without express permission;
 - Continuing to write after time has been called on a quiz, test, examination, or any other academic exercise or activity;
 - Submitting substantially the same work for credit in more than one class, except with prior approval of the instructor; or engaging in any form of research fraud.
- Falsification: altering or fabricating any information or citation in an academic exercise or activity.
- Plagiarism: representing, by paraphrase or direct quotation, the published or unpublished work of another person as one’s own in any academic exercise or activity without full and clear acknowledgment. It also includes using materials prepared by another person or by an agency engaged in the sale of term papers or other academic materials.

For additional information go to: [ARTICLE VI. University Regulations Regarding Academic Integrity](#)

Sexual Harassment/Title IX

Utah State University is committed to creating and maintaining an environment free from acts of sexual misconduct and discrimination and to fostering respect and dignity for all members of the USU community. Title IX and [USU Policy 339](#) address sexual harassment in the workplace and academic setting. The university responds promptly upon learning of any form of possible discrimination or sexual misconduct. Any individual may contact USU’s [Affirmative Action/Equal Opportunity \(AA/EO\) Office](#) for available options and resources or clarification. The university has established a complaint procedure to handle all types of discrimination complaints, including sexual harassment ([USU Policy 305](#)), and has designated the AA/EO Director/Title IX Coordinator as the official responsible for receiving and investigating complaints of sexual harassment.

Withdrawal Policy and "I" Grade Policy

Students are required to complete all courses for which they are registered by the end of the semester. In some cases, a student may be unable to complete all of the coursework because of extenuating circumstances, but not due to poor performance or to retain financial aid. The term ‘extenuating’ circumstances includes: (1) incapacitating illness which prevents a student from attending classes for a minimum period of two weeks, (2) a death in the immediate family, (3) financial responsibilities requiring a student to alter a work schedule to secure employment, (4) change in work schedule as required by an employer, or (5) other emergencies deemed appropriate by the instructor.

Students with Disabilities

USU welcomes students with disabilities. If you have, or suspect you may have, a physical, mental health, or learning disability that may require accommodations in this course, please contact the [Disability Resource Center \(DRC\)](#) as early in the semester as possible (University Inn # 101, (435)

797-2444, drc@usu.edu). All disability related accommodations must be approved by the DRC. Once approved, the DRC will coordinate with faculty to provide accommodations.

Diversity Statement

Regardless of intent, careless or ill-informed remarks can be offensive and hurtful to others and detract from the learning climate. If you feel uncomfortable in a classroom due to offensive language or actions by an instructor or student(s) regarding ethnicity, gender, or sexual orientation, contact:

- Division of Student Affairs: <https://studentaffairs.usu.edu>, (435) 797-1712, studentservices@usu.edu, TSC 220
- Student Legal Services: <https://ususa.usu.edu/student-association/student-advocacy/legal-services>, (435) 797-2912, TSC 326,
- Access and Diversity: <http://accesscenter.usu.edu>, (435) 797-1728, access@usu.edu; TSC 315
- Multicultural Programs: <http://accesscenter.usu.edu/multiculture>, (435) 797-1728, TSC 315
- LGBTQA Programs: <http://accesscenter.usu.edu/lgbtqa>, (435) 797-1728, TSC 3145
- Provost's Office Diversity Resources: <https://www.usu.edu/provost/diversity>, (435) 797-8176

You can learn about your student rights by visiting:

The Code of Policies and Procedures for Students at Utah State University: <https://studentconduct.usu.edu/studentcode>

Grievance Process

Students who feel they have been unfairly treated may file a grievance through the channels and procedures described in the Student Code: [Article VII](#).

Full details for USU Academic Policies and Procedures can be found at:

- [Student Conduct](#)
- [Student Code](#)
- [Academic Integrity](#)
- [USU Selected Academic Policies and Procedures](#)
- [USU Academic Policies and Procedures](#)
- [Academic Freedom and Professional Responsibility Policy](#)