

ESRM 6990V: Advanced Seminar - Intro to R

Spring 2025, Thursday, 5:00-7:45PM, Classroom GRAD 229,
2025/01/13 - 2025/05/01

Jihong Zhang

2025-01-13



1 General Information

- Course Code: ESRM 6990V
- Credits: 3 CH
- Course time and location: Wed 17:00-19:45; GRAD 210
- Instructor: Jihong Zhang, Dr.
- Contact Information: jzhang@uark.edu
- Personal Website: <http://jihongzhang.org>
- Office Location: GRAD 133B
- Office Hours: Tu 1:30-4:30PM
- Office Phone +1 479-575-5235
- Semester: Spring 2025

1.1 Course Objectives, Materials, and Pre-Requisites:

The online syllabus will always contain the most up-to-date information. This course provides a comprehensive introduction to R programming with an emphasis on data analysis and visualization in environmental and natural resource management contexts. Students will learn through hands-on coding exercises, practical examples, and real-world applications.

This course has three main objectives:

1. **Develop proficiency in R programming fundamentals**

- Master R syntax, data structures, and basic operations
- Learn to write efficient, reproducible code
- Understand R's package ecosystem and workspace management

2. **Build practical data analysis skills**

- Import, clean, and manipulate various data formats
- Create publication-quality visualizations using ggplot2
- Perform basic statistical analyses in R
- Apply R to environmental and natural resource datasets

3. **Establish foundation for advanced R applications**

- Develop problem-solving strategies for programming challenges
- Learn best practices for documentation and version control
- Prepare for advanced R applications in environmental science

Class time will be devoted to interactive lectures, live coding demonstrations, and hands-on exercises. All lecture materials, including R scripts and datasets, will be available in .html and .R formats on the course website prior to each class. Weekly coding assignments will reinforce concepts covered in lectures. While recorded lectures will be available upon request, active participation in class is essential for developing programming skills.

- Required software:
 - R (latest stable version)
 - RStudio Desktop
 - Git (for version control)

No prior programming experience is required, but basic statistical knowledge and familiarity with data analysis concepts is recommended. Students should have access to a laptop computer capable of running R and RStudio.

1.2 **How to Be Successful in This Class**

- Come to class ready to learn and come with your laptop;
- Complete in-class exercises;
- If you become confused or don't fully grasp a concept, ask for help from your instructor
- Know what is going on: keep up with email, course announcements, and the course schedule.
- Try to apply the information to your area of interest — if you have a cool research idea, come talk to me!

1.3 **Software**

- R and R packages
- Quarto
- RStudio (Posit)

2 **Course Materials**

2.1 Web Books

1. Introduction to Data Science Part I - Data Wrangling and Visualization with R by *Rafael A. Irizarry*
2. Introduction to Data Science Part II - Statistics and Prediction Algorithms Through Case Studies by *Rafael A. Irizarry*
3. R for Data Science - Visualize, Model, Transform, Tidy and Import Data by *Hadley Wickham & Garret Grolemund*
4. Yet Again: R + Data Science by *Albert Rapp*
5. Happy Git and GitHub for the useR by *Jennifer Bryan*
6. R Programming for Data Science by *Roger D. Peng*

2.2 R Community

1. RWeekly.org

2.3 Publications

2.4 Outline

This 17-week course is structured in four main modules that progressively build students' R programming and data science skills:

1. **Foundations of R Programming** (Weeks 1-3)
 - Introduction to R basics and environment
 - Core programming concepts
 - Fundamental R syntax and operations
 2. **Data Management and Processing** (Weeks 4-8)
 - Academic reporting with R projects and Quarto markdown
 - Data manipulation using tidyverse ecosystem:
 - Data summarization
 - Data cleaning
 - Data transformation
 - Data importing techniques
 3. **Data Visualization** (Weeks 9-11)
 - Comprehensive coverage of ggplot2:
 - Basic plotting principles
 - Advanced visualization techniques
 - Complex visualization applications
 4. **Professional Tools and Project Management** (Weeks 13-17)
 - Command-line operations (Terminal/PowerShell)
 - Version control with Git and GitHub
 - Final project implementation and presentation
- Key Features:
 - Hands-on learning approach with weekly lectures
 - Focus on tidyverse ecosystem for modern R programming

- Integration of professional tools (Quarto, Git)
- Culminates in a final project
- Two scheduled breaks: MLK Holiday (Week 2) and Spring Break (Week 12)

Learning Progression: The course is designed to progress from basic R fundamentals to advanced data manipulation and visualization techniques, concluding with professional development tools and a practical final project that synthesizes all learned concepts.

2.5 Why ITDS?

This book is free and comprehensive for a broad R tasks of data analysis. For example, the book covers R programming (**basics of R**), data wrangling with **dplyr**, data visualization with **ggplot2**, file organization with **Unix/Linux shell**, version control with **Git/Github**, and reproducible document preparation with *Quarto* and *knitr*.

3 Final Project Requirements

3.1 Overview

The final project is an opportunity to demonstrate your proficiency in R programming and data analysis by conducting an end-to-end data science project. You will apply the skills learned throughout the course to analyze a real-world environmental or natural resource dataset of your choice.

3.2 Project Objectives

1. Demonstrate proficiency in R programming fundamentals
2. Apply data manipulation and visualization techniques
3. Create reproducible research documentation
4. Practice version control and project management
5. Present findings effectively using R-based tools

3.3 Project Components and Weights

- **Project Proposal (10%)** - Due Week 14
 - Dataset description and source
 - Research questions/objectives
 - Proposed analysis methods
 - Timeline for completion
- **GitHub Repository (20%)**
 - Well-organized project structure
 - Clear documentation
 - Regular commits showing progress
 - README.md file with project description and instructions
- **R Code and Analysis (40%)**
 - Clean, well-commented R scripts
 - Efficient use of tidyverse functions
 - At least three different types of data visualizations using ggplot2

- Appropriate data cleaning and transformation steps
- Error handling and code optimization
- **Final Report (20%)**
 - Created using Quarto markdown
 - Professional formatting
 - Clear methodology description
 - Effective visualization presentation
 - Discussion of findings and limitations
 - 2,500-3,000 words (excluding code)
- **Presentation (10%) - Week 17**
 - 10-minute presentation
 - 5-minute Q&A
 - Clear communication of findings
 - Demonstration of key code features

3.4 Requirements

3.4.a Dataset

- Minimum 150 observations
- At least 5 variables
- Must be related to environmental science or natural resources
- Can be public or private data (with appropriate permissions)
- Must require significant cleaning or transformation

3.4.b Required Technical Elements

1. Data Processing

- Data import from at least two different file formats
- Data cleaning and transformation using tidyverse
- Creation of derived variables

2. Data Visualization

- Minimum three different types of plots using ggplot2
- At least one complex visualization (multiple layers/facets)
- Professional formatting and theming

3. Programming

- Use of functions
- Proper error handling
- Efficient code structure
- Well-documented code

4. Version Control

- Minimum 15 meaningful GitHub commits
- Clear commit messages
- Proper .gitignore file
- Well-structured repository

3.5 Deliverables Timeline

- Week 14: Project proposal due
- Week 15: Progress check-in
- Week 16: Code and repository review
- Week 17: Final presentation and submission of all materials

3.6 Submission Requirements

1. GitHub repository link containing:
 - All R scripts
 - Raw and processed data (or data access instructions)
 - Quarto markdown files
 - README.md with setup instructions
2. Final report in PDF format generated from Quarto
3. Presentation slides in PDF format

3.7 Evaluation Criteria

- Code quality and efficiency (25%)
- Data analysis depth and appropriateness (25%)
- Visualization effectiveness (20%)
- Documentation clarity (15%)
- Presentation quality (15%)

3.8 Tips for Success

- Start early and commit code regularly
- Choose a dataset that interests you
- Focus on clean, readable code
- Document your process thoroughly
- Test your analysis with different approaches
- Seek feedback during office hours
- Practice your presentation

3.9 Academic Integrity

- All code must be original or properly cited
- Data sources must be properly referenced
- Collaboration is encouraged for brainstorming, but all submitted work must be individual

This project is designed to be challenging but achievable with the skills learned in the course. It provides flexibility in topic selection while ensuring rigorous application of R programming concepts.

4 Academic Policies

4.1 AI Statement

Specific permissions will be provided to students regarding the use of generative artificial intelligence tools on certain graded activities in this course. In these instances, I will communicate explicit permission as well as expectations and any pertinent limitations for use and attribution. Without this permission, the use of generative artificial intelligence tools in any capacity while completing academic work submitted for credit, independently or collaboratively, will be considered academic dishonesty and reported to the Office of Academic Initiatives and Integrity.

4.2 Academic Integrity

As a core part of its mission, the University of Arkansas provides students with the opportunity to further their educational goals through programs of study and research in an environment that promotes freedom of inquiry and academic responsibility. Accomplishing this mission is only possible when intellectual honesty and individual integrity prevail.

Each University of Arkansas student is required to be familiar with and abide by the University's **Academic Integrity Policy** at honesty.uark.edu/policy. Students with questions about how these policies apply to a particular course or assignment should immediately contact their instructor.

4.3 Emergency Preparedness

The University of Arkansas is committed to providing a safe and healthy environment for study and work. In that regard, the university has developed a campus safety plan and an emergency preparedness plan to respond to a variety of emergency situations. The emergency preparedness plan can be found at emergency.uark.edu. Additionally, the university uses a campus-wide emergency notification system, UARKAlert, to communicate important emergency information via email and text messaging. To learn more and to sign up: <http://safety.uark.edu/emergency-preparedness/emergency-notification-system/>

4.4 Inclement Weather

If you have any questions about whether or not class will be canceled due to inclement weather, please contact me. If I cancel class, I will notify you via email and/or Blackboard. In general, students need to know how and when they will be notified in the event that class is cancelled for weather-related reasons. Please see here for more information.

4.5 Access and Accommodations

Your experience in this class is important to me. University of Arkansas Academic Policy Series 1520.10 requires that students with disabilities are provided reasonable accommodations to ensure their equal access to course content. If you have already established accommodations with the Center for Educational Access (CEA), please request your accommodations letter early in the semester and contact me privately, so that we have adequate time to arrange your approved academic accommodations.

If you have **not** yet established services through CEA, but have a documented disability and require accommodations (*conditions include but not limited to: mental health, attention-related,*

learning, vision, hearing, physical, health or temporary impacts), contact CEA directly to set up an Access Plan. CEA facilitates the interactive process that establishes reasonable accommodations. For more information on CEA registration procedures contact 479-575-3104, ada@uark.edu or visit cea.uark.edu.

4.6 Academic Support

A complete list and brief description of academic support programs can be found on the University's Academic Support site, along with links to the specific services, hours, and locations. Faculty are encouraged to be familiar with these programs and to assist students with finding an using the support services that will help them be successful. Please see here for more information.

4.7 Religious Holidays

The university does not observe religious holidays; however Campus Council has passed the following resolution concerning individual observance of religious holidays and class attendance:

When members of any religion seek to be excused from class for religious reasons, they are expected to provide their instructors with a schedule of religious holidays that they intend to observe, in writing, before the completion of the first week of classes.

5 Schedule

Following materials are only allowed for previewing for students registered in ESRM 64503. DO NOT DISTRIBUTE THEM on the internet. They will be removed after the course ended.

Week	Date	Topic	Reading	HW	Code/Data
1	01/13	Lecture 1: Basics of R	ITDS Ch.1		
2	01/20 (No Class)	(No Martin Luther King Holiday)			
3	01/27	Lecture 2: Program- ming Basics			
4	02/03	Lecture 3: Write academic reports with R projects and Quarto markdown			
5	02/10	Lecture 4: Data summa-			

Week	Date	Topic	Reading	HW	Code/Data
		rization with <i>tidyverse</i>			
6	02/17	Lecture 5: Data cleaning with <i>tidyverse</i>			
7	02/24	Lecture 6: Data transformation with <i>tidyverse</i>			
8	03/03	Lecture 7: Data Importing			
9	03/10	Lecture 8: Data Visualization I with <i>ggplot2</i>			
10	03/17	Lecture 9: Data Visualization II with <i>ggplot2</i>			
11	03/24 (No class)	(No Lecture 10: Data Visualization III with <i>ggplot2</i>)			
12	03/31	Spring Break			
13	04/07	Lecture 11: File Organizing with Terminal / Powershell			
14	04/14	Lecture 12: Git and GitHub	Pro Git Book		
15	04/21	Final Project			
16	04/28	Final Project			

Week	Date	Topic	Reading	HW	Code/Data
17	05/05	Final Project			