EVR 528 / 628: Introduction to Data Management and Visualization for Environmental Scientists

Email: jc_villasenor@miami.edu

Office hours: 10:15-12:00

Office: MSC 326

Course materials:
jcvdav.github.io/EVR_628

Schedule: Tu/Th: 9:00 – 10:15

Location: SLAB 120

Course description for UM Bulletin

"*Introduction to Data Management and Visualization for Environmental Scientists*" is designed for students looking to gain basic data management and visualization skills relevant to careers in environmental science and policy. The course provides an introduction to using R and RStudio to interact with environmental data; **No coding experience is required.** Students will learn highly marketable skills like visualizing tabular and geospatial data, data management, and reproducibility. All concepts will be introduced using real-world environmental datasets and questions. In-class exercises and homework assignments will mimic the types of tasks and questions that students will encounter in the workforce. By the end of the course, students will be comfortable working in R.

Course description

We now produce as much data in one year as has been collected in the past century[1]. Environmental professionals working in **marine conservation, coastal zone management, climate and society, and exploration sciences** are increasingly tasked with using data to extract insights and inform decisions that shape the world around us. Existing courses already teach students how to collect or analyze data, but no course teaches students the intermediate steps of data management, data visualization, and reproducibility. These these skills can be learned through an *introductory* class on data science, here understood as using computer code to interact with and extract insights from data.

"*Introduction to Data Management and Visualization for Environmental Scientists*" is designed for students looking to gain basic data science skills relevant to careers in environmental science, policy, and management. The course provides a gentle introduction to using R and RStudio –the most common coding language used in the Environmental Sciences– to interact with environmental data. Students will learn principles of responsible data visualization and communication, data management and reproducibility, and will do so using datasets relevant to the careers in **marine conservation, coastal zone management, climate and society, and exploration sciences**. For example, in-class exercises and homework assignments will mimic the types of tasks and questions that students will encounter in the workforce. Additionally, all concepts will be introduced using real-world environmental datasets, questions, and challenges based on research performed by EVR faculty and students. Some examples include:

---

[1] Tanhua, T. et al. Front. Mar. Sci. 6, 440 (2019)

- **Field-collected data** such as underwater ecological surveys in marine protected areas and fished areas or stomach content and biometrics of invasive lionfish.
- **Administrative data** like fisheries catch and effort from NOAA or health and income metrics from the US Census Bureau.
- **Geospatial data**, including animal telemetry data, vessel tracking data, hurricane tracks, political boundaries, remotely-sensed environmental measures (e.g. Sea Surface Temperature), and species distribution model outputs.

In parallel with providing students with technical skills, the course will also cover principles of responsible data utilization. For example, we will consider how data collection protocols and mandates shape what and who is represented in the data, and how this drives conclusions made with these data. We will also examine how different ways of visualizing the same data can sometimes result in opposing conclusions. By the end of the course, students will have developed the critical thinking necessary to work with data produced by others.

The course is designed to help students gain hands-on experience and transferable skills that will make them more competitive in the job market. Throughout the course, students will build a "data science portfolio" that showcases their coding skills and that can be included in job applications as proof of proficiency. Additionally, like many other skills, coding is one that atrophies if not practiced frequently. The portfolio will also serve as a valuable legacy notebook that students can consult in the future.

Undergraduate students welcome
Some seats are reserved for undergraduate students who have some relevant experience and permission. The grading expectations for undergraduates are more lenient. All other expectations remain the same.

Prerequisites

This is an introductory course with no prerequisites.

Course objectives

- To become familiar and comfortable with using R and RStudio to work with environmental datasets.
- To learn modern tools, techniques, and standards used in data visualization, data management, and reproducibility.
- To develop critical thinking necessary to ingest and produce data products.
- To develop a data science portfolio to be used in internship and job applications.

Reading resources

- R for data science by Hadley Wickham et al
- Visual Style Guide by the IPCC
- The Visual Display of Quantitative Information by Edward Tufte
- Software documentation and vignettes (examples below):

- ○ Wikham et al., Welcome to the Tidyverse. CRAN (2019).
  https://cran.r-project.org/web/packages/tidyverse/vignettes/paper.html
  - ○ Rudis, Ross, & Garnier. Introduction to the viridis colormaps. CRAN (2024).
  https://cran.r-project.org/web/packages/viridis/vignettes/intro-to-viridis.html
- ● *Ad hoc* peer review literature and blog posts (examples below)
  - ○ Brett et al., "Ocean data need a sea change to help navigate the warming world."
  Nature 582, no. 7811 (2020): 181-183.
  - ○ Englander. "Timeless principles for better figure design". The World Bank. (2024).
  https://blogs.worldbank.org/en/impactevaluations/timeless-principles-for-better-figure-design

Course content and schedule

Students must complete any assigned readings and software tests **before** class. Note that content, and timing of the content, is subject to change. Any changes will be clearly communicated via Blackboard. (***** Week contains holiday)

Week 1, Aug 18 - Aug 22: Introduction to DataScience and RStudio IDE
- Software installation and testing
- Overview of R Studio IDE
- R Studio Projects and project organization
- R scripts
- R packages: *tidyverse, custom, data*

Week 2, Aug 25 - Aug29: Data visualization
- Types of visualization
- Tufte's visualization principles
- Colorblindness, IPCC's visual style guide, and *viridis*
- The grammar of graphics and *ggplot*

*Week3, Sept 1 - Sept 5: Keeping track of your code with Git + GitHub
- Reproducible research
- Introduction to file structure
- Version control systems
- Building your first repository (hello world!)

First assignment: Setting up your portfolio in GitHub

Week 4, Sept 8 - Sept 12: Good coding principles
- Code style and documentation
- File structure and organization
- Classes, objects, variables, values
- Indexing and subsetting vectors and data frames
- Useful functions in *base* and *stats*

Week 5, Sept 15 - Sept 19: Scaling up your code and visualizations
- Layers, geometries, and aesthetics in *ggplot2*
- Themes with *ggplot2*
- Creating documents and presentations with Quarto

Second assignment: Data visualization

Week 6, Sept 22-26: Data management
- Reading and writing tabular data with *here*
- Metadata and documentation
- Retrieving environmental datasets from the Internet
- Raw data vs processed data

Week 7, Sept 29- Oct 3: Data transformation
- data.frames and tibbles in the *tidyverse*
- Rows (filter, arrange, distinct) in *dplyr*
- Columns (select, rename) in *dplyr*
- The pipe (native and *magrittr*)
- Grouping and summarizing data

Week 8, Oct 6 - Oct 10: Data tidying and wrangling
- Principles of tidy data
- Lengthening data with *tidyr*
- Widening data with *tidyr*
- Combining multiple sources of data (**_join*s)

Week 9, Oct 13 - Oct 17: Dealing with data
- Regular expressions with *stringr*
- Managing dates and times with *lubridate*
- Ordering factors with *forcats*

Third assignment: Data wrangling

Week 10, Oct 20-24: Working with spatial data in R
- Vector data and *sf*
- Raster data and *terra*
- Exploratory visualizations with *plot* and *mapview*

Week 11, Oct 27-31: Visualizing spatial data
- Attribute operations
- Building maps with *ggplot2* and *tmap*

Fourth assignment: Visualizing spatial data

Week 12, Nov 3 - 7: Extensions
- Connecting to external databases with *DBI*
- Animating ggplots with *gganimate*
- Interactive maps with *leaflet*

Week 13, Nov 10-14: Programming
- User-defined functions
- Iteration with loops
- Functional programming with *purrr*
- Background jobs
- Standardizing the environment with *.Rprofile*

Week 14, Nov 17 - Nov 21: Shiny Apps Framework
- Reactive programming
- UI / UX
- Front-end vs back-end

*Week 15, Nov 24 - Nov 28: Thanksgiving recess
Week 16: Dec 4 - Dec 10: Final presentations (egg timers)

Course Policies

Grading policy

- Assistance and participation (50%)
- Assignments (40%; 10% each, graded as completed / not completed)
- Final presentation (10%)

Grade:	Range:	
A	100 %	to 90.0%
B	< 90.0 %	to 80.0%
C	< 80.0 %	to 70.0%
D	< 70.0 %	to 60.0%
F	< 60.0 %	to 0.0%

Students can enroll as pass/no pass or for a letter grade. There are four assignments to track important milestones: 1) setting up your portfolio, 2) data visualization and communication, 3) data wrangling, and 4) spatial data.

Attendance policy

Students are expected to abide by the University of Miami's attendance policy. Other absence reasons should be discussed with me in advance. Unjustified absences will result in final grade Penalization. Classes may be recorded and made available to students upon reasonable request and justification.

Communication policy

All communication should be carried out via email. Please make sure to include the code code (e.g. EVR528 / 628) in the subject line.

Policy on the use of large language models (e.g. ChatGPT)

The use of large language models to debug and test your code is permitted.

Honor Policy

All students are expected to abide by the University of Miami's honor code. Anyone caught cheating on exams, improperly referencing published written or electronic material, or submitting work that is not their own will fail the course.

Policy on the use of technology during lectures

The use of technology for any other purpose that is not taking notes or working on the class content is highly discouraged. Repeated violations to this policy will result in final grade penalization.