Environmental Data Analysis in R

GEOG 41195 | 51195 | 71195 (3 credits)

Fall 2021 | Online (remote, synchronous) – Zoom Room | W 2:15-3:55pm

Instructor

Dr. Timothy Assal (he/him)
Office: 437 McGilvrey Hall
Email: tassal@kent.edu (*best way to contact me)*
Phone: 330-672-2046

Student (office) Hours

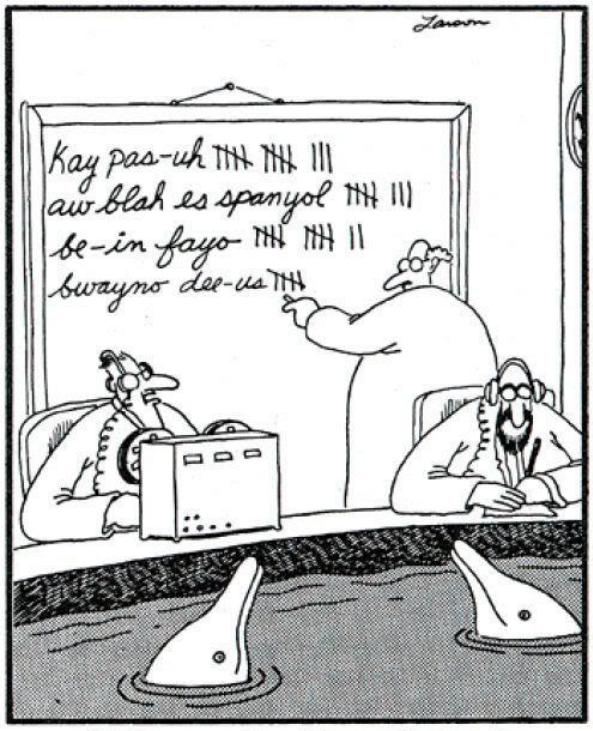
M/W 9:30 am – 10:45 am
M 1 – 3 pm; W 4-5 (in office or in my Zoom Room)
If you are unable to meet during these times, please contact me for an appt.

Course Description

The aim of the course is to teach students basic concepts, skills, and tools for working with data in R. R is an open source (free!) programming language for statistical computing and graphics that is widely used by statisticians, scientists, data miners, and geographers (https://www.r-project.org/). The number of R users grows exponentially each year, contributing to a massive online community and countless free tutorials. This is a hands-on course designed for students with no programming experience. However, this course does not teach statistics: understanding of basic statistics and common statistical tests is assumed and prior coursework in statistics is encouraged. With the knowledge gained in this course, students will be ready to undertake their own data analysis in R and increase their efficiency through reproducible workflows.

Course Learning Outcomes and Goals

- This course will teach you how to use the computer programming language R.
- The course assumes no prior knowledge of R, programming, or how we will interact with your computer via the command line interface.
- By the end of the course you will be able to: import and export data, produce publication-quality graphics, analyze data and write up results correctly, be confident in continuing to learn R, articulate the principles of best practices in data management, data analysis, graphics, workflow, and statistical approaches: And use them!
- Get comfortable with spatial data handling, visualization, and analysis
- Programming, writing code, command line interface, and collection, storage, analysis and display of data are all transferable skills that will be useful whatever software you end up using or work you end up doing.
- Please note, this is *not* a statistics course. Links to background material will be provided, but it is not intended to teach you statistics.



"Matthews ... we're getting another one of those strange 'aw blah es span yol' sounds."

COVID-19 Information: This course is being offered as a remote, synchronous class and we will meet during the allotted class period via Zoom. Please let me know if you become ill during the semester and we will make arrangements where appropriate. I will do the same if I become ill. Please follow the Flashes Safety Principles, take care of yourself and each other and I am confident we will have a great semester.

Course Materials*

Software

- The R statistical program (available from https://cran.r-project.org/)
- R Studio - an Integrated Development Environment (IDE) for R (available from https://www.rstudio.com/products/rstudio/download/)

Reference Books – (all are provided on Blackboard)

- R Programming for Data Science by Roger Peng (available at: https://leanpub.com/rprogramming)
- R for Data Science by Hadley Wickham and Garrett Grolemund (I will post a pdf version on Blackboard; however, the latest bookdown version is available at: https://r4ds.had.co.nz/)
- The R Book by Michael J. Crawley (available at: Kent State Library (ebook); I will also provide pdfs of the chapters).

**Both the software and reference books are open source and available free of charge (except Crawley). We will work with the open source ethic throughout this course.*

Course Structure

This course will provide an overview and introduction to the statistical software R. Class time will primarily be used for working through examples and problems as a class or individually. The best way to improve and feel comfortable in R is to use it frequently and regularly. ***There is no secret to learning R - the best way to approach the language is learn by doing (and re-doing…). Course mantra: work hard, learn a game-changing tool for your research, and have an incredibly in-demand skill to help land you a job!***

There will be five sections of the course outlined at the end of the syllabus. We will spend two to four weeks on each section, although the pace will depend on the comfort of the class with each subject. I will assign a **reading assignment** each week (in class, on Blackboard, or via email) to be completed before the next class meeting. Each class meeting will contain a **lecture** on the scheduled topic(s), followed by a **lesson** to work though the days material (individually or as a group). I will assign a **lab** each week (due before the next class meeting; ~ 1 week; approx. 10 total) designed to reinforce the material and develop coding skills in R. Labs are open book, can be completed with classmates, but you are expected to write your own code. ***Labs turned in by the deadline will be eligible for feedback and can be resubmitted within 4 days to potentially obtain full credit.***

Lab assignments will be graded as follows:

- Produces the correct answer using the requested approach: 100%
- Generally uses the right approach, but a minor mistake results in an incorrect answer: 90%
- Attempts to solve the problem and makes some progress using the core concept: 50%
- Answer demonstrates a lack of understanding of the core concept: 0%

Attendance/Participation
I strongly encourage attendance to ensure you do not fall behind and learn something about R. Please let me know if you will miss class ahead of time if possible. The class only meets once per week to maximize time dedicated to the material. Please note, this semester the course will run as a remote synchronous class. **Although we are online, we will meet during the allotted class period via Zoom.** I will set up a discussion board on Blackboard so we can all engage in discussion and/or questions between class periods. Each student has a 48-hour grace period to turn in lab assignments (life happens), but please do not take advantage of this policy as it will be detrimental to staying on top of the material as the semester rolls on (and I likely will not be able to give you a chance to resubmit).

Course Assessment*

Category	Points per item	Total Points in Category	Perc. of class grade
Lab Assignments	10	100	~ 49%
Project Proposal	10	10	5%
Final Project	75	75	37%
Attendance/Participation	N/A	20	10%
Total		**205**	**100%**

*There will be **no examinations** in this course.

Final Grade Scale

A 92.5 - 100	B+ 86.5-89.4	C+ 76.5 - 79.4	D+ 66.5 - 69.4
A- 89.5 – 92.4	B 82.5 - 86.4	C 72.5 - 76.4	D 59.5 - 66.4
	B- 79.5 - 82.4	C- 69.5 - 72.4	F < 59.4

Final Project
The goal of this project is to advance your workflow on how to tackle an analysis to answer a question. It is my hope this project will serve as a blueprint for your future analyses or better yet, a chance to dive into your own data. If you do not have your own data, you may use an existing dataset found online or provided by another researcher (faculty advisor or collaborator) or one of my datasets. Identical questions should not have been addressed with the specific data before (i.e. it must be a novel question on an existing dataset). The project will proceed in two stages:

A. Proposal (due mid-semester)
Include the following sections (pdf format, 1 page document)

1. *Introduction* A short introductory paragraph describing the scientific rationale for the question.
2. *Question* The specific question(s) or hypothesis(es) you will address
3. *Dataset* A brief description of the data (e.g. unit of observation, number of observations, summary of covariates and response variables) and either a link to the dataset or the dataset itself (if not online and data-sharing is agreed with data owner).

I'll check that the questions and data are feasible and appropriate. You will not receive credit until all necessary information is provided (to make sure you're on track!)

B. Project (due at end of semester in lieu of final exam)
The project will include complete exploration of the data, at least one statistical test to address the question, at least one publication quality graphic of the results, some spatial representation of the data (e.g. field location, a map of the area, etc.), and a final report that explains all of it.

Students will submit a zip/compressed file containing five documents:

1. The raw data set or link to it (see me if there are issues with data sharing)
2. A file of R code which documents the entire process from data entry to publication quality graphics (length will depend on the analysis and should be formatted appropriately (submit as .R file).
3. The zip file should also contain the .Rproj file so that the workflow is quickly reproducible.
4. A summary document written in style of an academic paper, containing the question, methods and results section pertinent to the question. This document should be submitted as a PDF file in the style of an academic journal (Introduction, Methods, Results, Discussion, Literature Cited) and will include in-text citations to references and figures.
5. A graphical abstract (~200-250 word abstract) with one figure to be posted on Blackboard (not to exceed 1 page).

Grades will be based on:

- Comprehensive exploration of the data (plotting, testing of model assumptions, etc.)
- Sufficient commenting of code to ensure understanding by other readers
- Correct statistical analysis of the data
- The summary document sufficient to replicate the analysis and that correctly reports and describes the results
- Publication-quality graphics demonstrating best practice

Contribution to a peer-reviewed publication is a possibility (the sky is the limit!).

University Policies

Important Dates

The official registration deadline for this course is **September 1st**. University policy requires all students to be officially registered in each class they are attending. Students who are not officially registered for a course by published deadlines should not be attending classes and will not receive credit or a grade for the course. Each student must confirm enrollment by checking his/her class schedule (using Student Tools in FlashLine) prior to the deadline indicated. Registration errors must be corrected prior to the deadline. The course withdrawal deadline is **November 3rd**.

Academic Dishonesty

University policy 3-01.8 deals with the problem of academic dishonesty, cheating, and plagiarism. None of these will be tolerated in this class. The sanctions provided in this policy will be used to deal with any violations. If you have any questions, please read the policy at http://www.kent.edu/policyreg/administrative-policy-regarding-student-cheating-and-plagiarism

Accommodations & Accessibility
University policy 3-01.3 requires that students with disabilities be provided reasonable accommodations to ensure their equal access to course content. If you have a documented disability and require accommodations, please contact the instructor at the beginning of the semester to make arrangements for necessary classroom adjustments. Please note, you must first verify your eligibility for these through Student Accessibility Services (contact 330-672-3391 or visit www.kent.edu/sas for more information on registration procedures).

Survey of Instructor
The Student Survey of Instruction (SSI) is now online. We will dedicate a portion of a class period later in the semester for students to complete this survey.

Course Policies

Respect for Diversity, Equity, and Inclusion
In this class, we are seriously committed to supporting diversity and inclusion among all classroom community members (our university is too!). We treat one another fairly and honor each other's experiences, beliefs, perspectives, abilities, and backgrounds, regardless of race, religion, language, immigration status, sexual orientation, gender identification, ability status, socio-economic status, national identity, or any other identity markers. It is my intent that students from all diverse backgrounds and perspectives be well served by this course, that students' learning needs be addressed both in and out of class, and that the diversity that students bring to this class be viewed as a resource, strength, and benefit. Your suggestions are encouraged and appreciated. Please let me know ways to improve the effectiveness of the course for you personally or for other students or student groups. In addition, if any of our class meetings conflict with your religious events, please let me know so that we can make arrangements for you.

Land Acknowledgment
We acknowledge that the territory on which Kent State University stands is that of The Kaskaskia and The Erie People. This statement is one small step in acknowledging the history that brought us to reside on the land, and to help us seek understanding of our place within that history. For more information, please visit the Native American Indian Center of Central Ohio.

Mental Health
College life can be incredibly stressful, and you may experience a range of issues that can cause barriers to learning and your well-being. Learn more about university sponsored resources to help.

Course Schedule

Week	Date	Topic	Reading	Lab Due
Section 1: Introduction and Basic Building Blocks of R				
1	1-Sep	Course Overview, Introduction to R & Rstudio		
2	8-Sep	Nuts and bolts of R, data structures	P: Ch. 3; NY Times article; BBC article	Lab 1
3	15-Sep	Data structures (cont'd), subsetting R objects, loading packages, base graphics	P: Ch. 5	Lab 2
4	22-Sep	Getting data in/out of R, data wrangling, creating workflows through scripts	P: Ch. 6; WG: Preface; Reference - WG: Ch. 3 & 9; P: Ch. 13 (Peng uses dplyr outside the Tidyverse)	Lab 3
Section 2: Exploring and Visualizing Data				
5	29-Sep	The beauty of ggplot2 for publishable graphics	WG: Ch. 1	Lab 4
6	6-Oct	Exploratory Data Analysis (aka EDA!)	WG. Ch. 5	Lab 5
Section 3: Statistical Analysis of Data				
7	13-Oct	Classical Tests (Chi-square, t-test, ANOVA, etc.) ***PROJECT PROPOSALS DUE***	Reference - The R Book: Ch. 8 & 11	
8	20-Oct	Regression Analysis (simple & multiple linear; logistic regression)	Reference - The R Book: Ch. 8 & 10; Linear Regression Primer	Lab 7
9	27-Oct	Generalized Linear Models	Reference - The R Book: Ch. 13	Lab 8
Section 4: Working with Spatial Data				
10	3-Nov	Spatial Data Theory (vector/raster, projections, importing spatial data)	sf package vignette	Lab 9
11	10-Nov	Spatial Data Visualization and Manipulation	Reference - CRS primer	
12	17-Nov	Using R as a GIS (raster analysis, raster viz)	raster package vignette	Lab 11
Section 5: Programming in R				
13	24-Nov	**NO CLASS MEETING – Thanksgiving Break** (work on Final Project)		
14	1-Dec	Control Structures (The apply family, if-else, for, while, repeat, etc.)	P: Ch. 14; WG: Ch. 17	Lab 12
15	8-Dec	Creating custom functions	P: Ch. 15; WG: Ch. 15	
	15-Dec	**Final Project Due at 5 pm**		

***Please note:** there are 10 labs, but they are not successive in numbering. The lab number refers to the week the material was covered.