Geospatial data analysis in R
Week 1 - Introduction

Josh Merfeld

KDI School

2024-09-08

Introductions

- Let's start with a little introduction

- Name, year, program, research interests, etc.
  - Why are you taking this class?

Course Overview

- Geospatial data analysis in R

- Major themes:
  - Geospatial data types
    - Shapefiles (vector files), rasters, etc.
  - Data visualization
    - ggplot, tidyterraarch
  - You will be doing assignments in R Markdown

Course Overview

- Today will just be a short introduction
- For next class, please come with R and R Studio installed on your computer
  - You can find instructions on the syllabus
  - You **must** bring a laptop to class. If you cannot do this, please speak with me.
- Course website: https://github.com/JoshMerfeld/geospatialdataR
  - You can find slides, assignments, and other materials here
  - It will be updated as we go throughout the semester
    - (I am still making slides!)

Course Overview

- This is a brand new class, so I will likely be making changes as we go
- Please check the course website regularly for updates

Detailed outline (tentative)

1. Linear regression (week 2)
   - Inference (confidence intervals, hypothesis testing, bootstrapping, etc.)
2. Maximum likelihood estimation (week 3)
   - Discrete choice (logit, probit, multinomial logit, etc.)
3. Introduction to causality (week 4)
   - Potential outcomes framework
   - Problems with simple regression
   - Why randomization works

Detailed outline (tentative)

4. Differences-in-differences (weeks 5 and 6)
   - Fixed effects, including two-way fixed effects
   - Event studies
   - Synthetic control

5. Instrumental variables (weeks 7 and 8)
   - Assumptions
   - IVs in RCTs (LATE)
   - Some examples
   - Weak instruments
   - Bartik (shift-share) instruments

Detailed outline (tentative)

6. Regression discontinuity (week 9)
   - Canonical regression discontinuity
   - Parametric vs. non-parametric

7. Machine learning in economics (week 10)
   - ML for prediction (lasso, ridge, elastic net)
   - Cross validation
   - Heterogeneous treatment effects
   - Brief introductionto other supervised ML (time dependent)

1. Homework - coding tasks (55%)

- The homeworks form the main grading component of the course
- The goal is to get you comfortable with coding *and writing* in R
  - I will also ask you to interpret things to make sure you understand what you are doing statistically
- I expect you to do your homeworks in R Markdown and turn in the code along with a pdf output[1]]
- I expect you will have four or five homeworks throughout the semester
- For those of you without a background in R, the first few weeks will take a bit of effort. It will get easier, I promise.

[1] Note: If you have a strong preference for using a different language (e.g. Python), please let me know and we can discuss it. However, you **must** be able to produce a pdf output with your code and results. Using Word is a no-go.

Grading

1. Final exam (35%)
   - This will be a take-home exam with a mix of theory and coding.
2. Participation (10%)
   - I expect everyone to participate in class. That means asking questions, answering questions, and participating in discussions.

TA sections

- The goal of TA sections is to help you with R and R Markdown
- For help with the actual material, please come to my office hours

Questions?

- Any questions about the course?

Next class

- We need to have R and RStudio installed for what's next
  - Another code editor is also acceptable: VS Code, for example

- Course website: https://github.com/JoshMerfeld/geospatialdataR