Sabanci University

Econ-494 Spatial Data Science, Fall-2022 (online delivery)

Contact Information

Instructor: Ivan G. Lopez Cruz
Office: FASS-2120
Office Hours: Friday, 5 am-6 am through Zoom (See the Zoom invitation on the syllabus' last page).
Email: ilopezcr@sabanciuniv.edu

Description

This course's central goal is to introduce the student to the analysis and employment of spatial datasets in the social sciences. It begins with a thorough description of R's tools and methods to manipulate and visualize geographic data. After becoming acquainted with the construction of spatial variables, the student learns how economists exploit the latter to uncover the causal mechanisms determining the link between historical developments (e.g., the colonization of America) and today's regional development levels. The course also deepens into various statistical models that incorporate parameters governing a given phenomenon's spatial diffusion, thereby tackling questions such as: how intense is the dissemination of violence across space following the outbreak of civil conflict? Will one municipalities' improvements in educational levels spill to adjacent localities? A discussion on estimation techniques, hypothesis testing, and an introduction to spatial unsupervised learning methods mark the course's end.

Class Information

- This is course's delivery method will be exclusively through the internet.
- All lectures will use the same Zoom meeting information (see the invitation at the end of the syllabus). *I will not allow Non-Sabanci accounts into our meetings.* The passcode of the conference will change after the second week of classes.
- I will use Zoom's video recorder feature for every lecture.
- Our 1-hour lecture's delivery method will be asynchronous, e.g., through a pre-recorded video. Occasionally, I will give the 1-hour lecture via Zoom, and will let you know about it ahead of time
- The class website is SUCOURSE+, where you will access diverse materials, including the recorded lectures, homework sets, and the like. Moreover, I will post the course's announcements on SUCOURSE+ as well.

Evaluation:

This class comprehends the following grading components:

- Three homework sets: 50 % of your final grade. Each of these activities will require you to deliver several pieces of code to SUCOURSE. More specifically, as time progresses, I will gradually make available a sequence of R scripts within the right folder (HMW1, HMW2, HMW3) with the practice questions. This scheme aims to allow you work on the exercises ahead of time. By the deadline, you should upload to the corresponding assignment all the R scripts with your answers. Make sure you use the package "HERE" so that I can easily run your

work on my computer. Sometimes, you will have to accompany your code with a pdf file displaying the outputs of your programs and your interpretation of results. Uploading any material after the corresponding due dates will result in a zero. I will announce the deadlines later.

- One (Guided) Project + Short Presentation: 40 % of your final grade. You will complete this task in steps:
  1. By the end of the third week of classes, I will show you how to handle the contents of the ACLED's website. Moreover, I will assign to you a specific project whose outputs you will need to reproduce (graphs, tables, and maps).
  2. Alongside with your second homework assignment, you must turn in a script reproducing most, if not all, the project's outputs.
  3. The final outputs of this endeavor will be a report in which you will reproduce the entire project I assigned to you in the beginning; furthermore, you will add a short section in which you will perform your own analysis that must comprise at least one map, one graph, and one table of your own. Please, interpret these results as well and upload the report and the code to SUCOURSE. (30 % of your grade)
  4. By the end of the semester, we will hold a series of seminars in which each student will discuss her project in a 10-minutes presentation (10% of your grade).
- Attendance-Participation: 10 % of your final grade. I will keep record of your attendance and participation in class.

Failing to present your project:

- If missing your presentation results from illness, you must hand in a formal doctor's note from a hospital. (A doctor's note taken from a doctor's private practice does not count as formal.) The university health service must verify the doctor's prescription in all instances. Therefore, you should submit the doctor's note to the Health Center timely. Failing to present your work without any valid documented reason will result in a zero for it.
- Suppose you experience any technical problem with your internet connection during the project presentation. In that case, you must send me ASAP evidence showing the technical problem (a screenshot displaying what the problem was).
- Suppose you miss your presentation altogether. However, you presented a valid documented reason (as per above). In that case, you can recuperate the lost points via an **IN PERSON**, comprehensive, 1.5 hours long **makeup exam** on **TBA**. This Exam's weight will match the grade percentage you have missed. Finally, note that it will evaluate all the topics we cover before **TBA**, and, therefore, it will be very hard.

Further details

- **Textbook And Readings**: We do not have a textbook for this class since the array of topics we will cover are scattered among several sources, including my own research. However, the following online books may serve as occasional references:
  - https://rspatial.org/index.html
  - https://keen-swartz-3146c4.netlify.app/
  - http://132.72.155.230:3838/r/
  - https://mgimond.github.io/Spatial/index.html

  The only required reading is "Spatial Analysis in the Social Sciences and Humanities" by Hering et al. It may give you some insights for your term paper. You can find this piece on SUCOURSE.
- **R Packages: "**TIDYVERSE" "TMAP" "SF" "SPDEP"
- **Cheat sheets**: I have made available a few cheat sheets concerned with the packages of interest; you must become acquainted with them for they will provide crucial guidance.

- **Class Materials**: I will post diverse materials to complement the textbook, such as videos and presentations.
- **Prerequisites:** Multivariate Differential Calculus and Econometrics. It is your responsibility to review these tools.
- **Individualized Extra-Credit:** Under no circumstances will I grant extra-credit work to individual students.
- **GRADING SCALE: You will have weighted scores at the end of the semester, adding up to 100 points or less.**
- **Words of advice:** The key to obtaining a good grade is to keep up. The material builds on itself as the semester progresses.

IMPORTANT: During the first two weeks of classes, you will be responsible for reviewing the basics of R. To do so, you must go through the following free tutorials:

https://www.youtube.com/watch?v=_V8eKsto3Ug : thorough introduction to R. Watch everything up to "Entering Data"

https://www.youtube.com/watch?v=jWjqLW-u3hc : Introduction to the package "DPLYR" for a data manipulation.

https://mgimond.github.io/ES218/Week03a.html : I strongly encourage you to review Manuel Gimond's material on DPLYR.

Course Contents

PART I: Handling Spatial Data

1. Introduction to the main concepts in SDS. (Week 1)
    a. What is SDS?
    b. Spatial Data Analysis
    c. Practice: DPLYR
2. Handling Spatial Data in R. (Week 2)
    a. Vector Data
    b. Raster Data
    c. Coordinate Systems
    d. Some Famous Data Sets (useful for your term project)
3. Spatial Data Frames. (Week 3)
    a. The SF package.
    b. Vector sub-setting, Aggregation, and Joins.
    c. Raster sub-setting and Summaries.
4. Binary Operations on spatial objects. (Week 4)
    a. Vector Spatial Operations.
    b. Distances.
    c. Raster Spatial Operations.
5. Geometry Operations (Week 5)
    a. Altering vectors.
    b. Altering Raster Data.
6. Raster-Vector Interactions (Weeks 6-7)
    a. Raster cropping, raster extraction, rasterization.
    b. Vectorization of spatial data.
    c. Spatial-temporal datasets.

HOMEWORK 1: Will encompass the first five chapters and will be due one week or so after we finish covering “Geometry Operations.”

PART II: Analysis and Modelling

7. Exploratory Analysis. (Weeks 8-9)
    a. Good vs. Bad Maps.
    b. Statistical Maps.
    c. Useful plots and Exploratory Analysis.
8. Spatial Autocorrelation. (Week 10-11)
    a. Local vs global statistics.
    b. Visualizing SA.
    c. LISA and Local Moran.
    d. LISA: inference and interpretation.

HOMEWORK 2: Will encompass chapters 6-8, and will be due one week or so after we finish covering “The use of frontiers and distances to identify causal effects.”

9. Using Spatial Variables in Econometric Models. (Weeks 13-14)
    a. Brief introduction to causality.
    b. Regression overview.
    c. The use of frontiers and distances to identify causal effects.
    d. Application: the persistence of socioeconomic traits through history.

HOMEWORK 3: Will encompass chapter 9, and will be due sometime after the last day of classes.

BONUS: I may have time to go over some of the below topics (Maybe through pre-recorded videos)

10. Spatial Statistical Models.
    a. Weight Matrices.
    b. Spatial Autocorrelation.
    c. Spatial Regressions.
11. Unattended spatial learning
    a. Clustering of spatial points with k-means, PAM and CLARA algorithms
    b. Clustering with the DBSCAN algorithm
    c. Spatial Principal Component Analysis
    d. Spatial Drift
    e. Spatial hierarchical clustering
    f. Spatial oblique decision tree

Letter Grades:

A	A-	B+	B	B-	C+	C	C-	D+	D	F
90-100	89-85	84-80	79-75	74-70	69-65	64-60	59-55	54-50	49-45	0-39

NOTE: I will try my best to go over the entire above list. However, the chances are that we will fall a bit short, in which case I will have to redesign the contents of the homework sets on the go.