

Econ-494 Spatial Data Science, Spring-2023

Contact Information

Instructor: Ivan G. Lopez Cruz
Office: FASS-2120
Office Hours: Friday, 9 am- 11 am.
Email: ilopezcr@sabanciuniv.edu

"We may have to revise the course plan according to the reassessment to be made country-wide, regarding higher education, at the beginning of April. The content to be delivered is certain but the method of course delivery, the dates appearing in this syllabus, and some other details are subject to change."

Description

This course introduces students to analyzing and employing spatial datasets in the social sciences. It describes 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

- **IMPORTANT: You must bring your laptop to every class.** This request is because we will spend the lectures' time coding together and discussing a few theoretical concepts. Thus, please download the lecture's data and R script before each class.
- The class website is SUCOURSE+, where you will access diverse materials, including lecture slides, 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:

- Two homework sets: 50 % of your final grade. Each activity 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 correct folder (HMW1 and HMW2) with the practice questions. This scheme aims to allow you to work on the exercises beforehand. You should upload all the R scripts with your answers to the corresponding assignment by the deadline. Make sure you use the package "HERE" so I can efficiently 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 the following steps:

1. By the end of the third week of classes, I will show you how to handle the contents of ACLED's website. Moreover, I will assign you a specific project whose outputs you must reproduce (graphs, tables, and maps).
2. Alongside your second homework assignment, you must submit a script reproducing most, if not all, of the project's outputs.
3. The final output of this endeavor will be a report in which you will reproduce your assigned project. Furthermore, you will add a short section in which you will perform your 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 seminars in which each student will discuss their project in a 10-minutes presentation (10% of your grade).

NOTE: If the number of enrolled students is high, we will form teams for the ACLED projects.

- Attendance-Participation: 10 % of your final grade. I will keep a 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 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.
- If we decide to conduct the final presentation through Zoom, the following applies: 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 issue (a screenshot displaying the problem).
- 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**; therefore, it will be very hard.

Further details

- **Textbook And Readings:** We do not have a textbook for this class since the topics we will cover are scattered among several sources, including my 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. However, it may give you some insights for your term project. You can find this piece on SUCOURSE.
- **Main R Packages:** "TIDYVERSE," "TMAP," "SF," "SPDEP," and "PURRR"
- **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 a basic command of statistics and probability notions. 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> : a thorough introduction to R. Watch everything up to "Entering Data"

<https://www.youtube.com/watch?v=jWjqLW-u3hc> : Introduction to the package "DPLYR" for 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. R overview. (Week 1 and 2)
 - a. Programming with DPLYR.
 - b. Functional Programming with PURRR.
 - c. Reshaping Datasets.
2. Introduction to the main concepts in SDS. (Week 3)
 - a. What is SDS?
 - b. Spatial Data Analysis
 - c. ACLED data
3. Handling Spatial Data in R. (Week 4)
 - a. Vector Data
 - b. Raster Data
 - c. Coordinate Systems
 - d. Some Popular Data Sources.
4. Spatial Data Frames. (Week 5)
 - a. The SF package.
 - b. Vector sub-setting, Aggregation, and Joins.
 - c. Raster sub-setting and Summaries.
5. Binary Operations on spatial objects. (Week 6)
 - a. Vector Spatial Operations.
 - b. Distances.
 - c. Raster Spatial Operations.
6. Geometry Operations (Week 7)
 - a. Altering vectors.
 - b. Altering Raster Data.
7. Raster-Vector Interactions (Weeks 8)
 - a. Raster cropping, raster extraction, and rasterization.
 - b. Vectorization of spatial data.
 - c. Spatial-temporal datasets.

HOMEWORK 1: This will encompass the first seven chapters and will be due one week or so after we finish covering "Geometry Operations."

PART II: Analysis and Modelling

8. Exploratory Analysis. (Weeks 8-9)
 - a. Good vs. Bad Maps.
 - b. Statistical Maps.
 - c. Useful plots and Exploratory Analysis.
9. Spatial Autocorrelation. (Week 10-11)
 - a. Local vs. global statistics.
 - b. Visualizing SA.
 - c. LISA and Local Moran.
 - d. LISA: inference and interpretation.
10. Using Spatial Variables in Quantitative Models. (Weeks 13-14)
 - a. A brief introduction to causality.
 - b. Gentle Introduction to Regression.
 - c. RDD designs: classic and spatial
 - d. The use of frontiers and distances to identify causal effects.
 - e. Application: the persistence of socioeconomic traits through history.

BONUS: I may have time to go over some of the below topics:¹

11. Spatial Statistical Models.
 - a. Weight Matrices.
 - b. Spatial Autocorrelation.
 - c. Spatial Regressions.
12. 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

HOMEWORK 2: This will encompass every topic in PART II and will be due by the last day of classes.

Highest	Lowest	Letter
100.00%	93.00%	A
92.99%	90.00%	A-
89.99%	87.00%	B+
86.99%	83.00%	B
82.99%	80.00%	B-
79.99%	77.00%	C+
76.99%	73.00%	C
72.99%	70.00%	C-
69.99%	67.00%	D+
66.99%	60.00%	D

59.99%---0.00% results in F

¹ **NOTE:** I will try my best to review the list above. 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.

Forums and Technical consultations

I noticed that several students get honestly terrified after an error message prompts in R. The thing is that when one writes code 80% of the time, it is necessary to amend small mistakes. Thus, eluding error messages is impossible. On the bright side, others have solved your problems in 99% of instances and made their solution publicly available on the web. As such, you only need to copy-paste the annoying error message tormenting you on your browsers, and you will see many curated solution methods.

Regarding the above, I will open a forum (anonymity does not work on SUCOURSE, I am sorry but you will have to be brave and post your comments here):

- Coding/R forum: You can post issues like "I get the following error message: error in select(x,y)... I consulted a few websites and still cannot figure it out!"

I usually let the class find joint solutions and occasionally intervene if the forum remains unpopulated with quick fixes. Of course, you are always welcome to visit me in my office on Fridays between 9 and 11 am. You can also come outside such a schedule, but you may not find me.

Attendance Policy

Although I will be physically in the classroom, you all can attend the course through Zoom (at least until April). However, I will post no recordings of the lectures on SUCOURSE. Additionally, I will record your attendance via the usual Zoom meetings capabilities and in written format for those who decide to follow the class in person.