

Econ 494: Spatial Data Science

Ivan Lopez Cruz



Contact

email: ilopezcr@sabanciuniv.edu

web: <https://ivanlopezcruz.wordpress.com/>

office hours: F 09:00-11:30 a.m.

Course Description (Spring 2024)

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. There are two reasons for this request. First, we will spend the lectures' time coding together. Second, we will have frequent pop-out quizzes that you will work out on your computers.

- If you miss any number of Quizzes for whatever reason, you will have a chance to regain the missed points through a comprehensive in-person exam by the end of the semester (more below)
- 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.
- Please download the lecture’s data and R script before each class as per our schedule on SUCOURSE.

Required Texts

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. Furthermore, it may give you some insights for your term project. You can find this piece on SUCOURSE.

Other required materials

- You will need access to a computer that can run R and RStudio. These programs are free and can run on every major operating system.
- **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://dcl-wrangle.stanford.edu/> : I strongly encourage you to review this book’s first two chapters:
 - * Chapter 1: Basics
 - * Chapter 2: Other single-table verbs

NOTE: it would be most efficient if you could review these materials before the start of the fall term.

Course Policies

Please use the information below as a reference for how this class will be conducted. I ask that you review this information before contacting me with any questions.

Grading Policy

Pop-out quizzes

- **Several pop-out quizzes will determine 40% of your grade** during regular class hours. These activities will require you to use your laptops in the classroom. Thus, you must bring your computers to every class.

Two homework sets

20% of your grade will come from two homework sets (10% each). 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 correct folder (HMW1 and HMW2) with the practice questions. This scheme aims to allow you to work on the exercises ahead of time. 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 **teams**. It comprises 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).¹
2. The final output of this endeavor will be a report in which you will replicate your assigned project. **Furthermore, you will add a section discussing your own ideas and insights.** Such a section 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. (20 % of your grade)

¹The [Armed Conflict Location & Event Data Project](#) (ACLED) is a disaggregated data collection, analysis, and crisis mapping project.

3. By the end of the semester, we will hold seminars in which each team will discuss their project in a 15-minutes presentation (20% of your grade).

Forums and Technical consultations

I noticed that several students get honestly terrified after an error message prompts in R. Indeed, when one writes code, 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 a.m. You can also come outside such a schedule, but you may not find me.

Email Policy

I welcome emails from students regarding class policies and assignments. I generally respond to all emails within 24 hours; however, it may take longer on weekends. In addition, it is unlikely that I will be able to react very quickly before deadlines, such as exams and papers, so please provide at least one day for a response if you need an issue clarified about an assignment.

I'm happy to answer any questions you may have about the class discussions through email. However, I prefer to have a more in-depth and meaningful conversation with you in person during my office hours. I am looking forward to seeing you there!

Makeup Exam Policy

- If having missed a Quiz/Quizzes and/or your presentation **results from illness**, the student 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 a valid documented reason will result in a zero for the corresponding grade item.²

²In case 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 a grade item/items. 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 semester's topics; therefore, it will be very hard.*

TENTATIVE Class Schedule

I expect students to have downloaded the corresponding R scripts and data from SUCOURSE before every class.

PART I: Handling Spatial Data.

1. R overview. (Week 1 and 2)
Programming with DPLYR.
Functional Programming with PURRR.
Reshaping Datasets.
2. Introduction to the main concepts in SDS. (Week 3)
What is SDS?
Spatial Data Analysis
ACLED data
3. Handling Spatial Data in R. (Week 4)
Vector Data
Raster Data
Coordinate Systems
Some Popular Data Sources.
4. Spatial Data Frames. (Week 5)
The SF package.
Vector sub-setting, Aggregation, and Joins.
Raster sub-setting and Summaries.
5. Binary Operations on spatial objects. (Week 6)
Vector Spatial Operations.
Distances.
Raster Spatial Operations.

6. Geometry Operations (Week 7)

Altering vectors.

Altering Raster Data.

7. Raster-Vector Interactions (Weeks 8)

Raster cropping, raster extraction, and rasterization.

Vectorization of spatial data.

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)

Good vs. Bad Maps.

Statistical Maps.

Useful plots and Exploratory Analysis.

9. Spatial Autocorrelation. (Week 10-11)

Local vs. global statistics.

Visualizing SA.

LISA and Local Moran.

LISA: inference and interpretation.

10. Using Spatial Variables in Quantitative Models. (Weeks 13-14)

A brief introduction to causality.

Gentle Introduction to Regression.

RDD designs: classic and spatial

The use of frontiers and distances to identify causal effects.

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.

Weight Matrices.

Spatial Autocorrelation.

Spatial Regressions.

12. Unattended spatial learning

Clustering of spatial points with k-means, PAM, and CLARA algorithms

Clustering with the DBSCAN algorithm

Spatial Principal Component Analysis

Spatial Drift

Spatial hierarchical clustering

Spatial oblique decision tree

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

Grading Table

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

Any grade below 60% will produce a grade mark of **F**.