# GEOG 215: Introduction to Spatial Data Science

*Spring 2020*

*MW 3:35 pm - 4:50 pm*

**Location:** 204 Carolina (Hurston) Hall
**Time:** MW 3:35pm - 4:50pm
**Course Website:** geog215-spds.rbind.io
**Class Discussion Website:** https://piazza.com/unc/spring2020/geog215/home

**Instructor:** Varun Goel | varung@live.unc.edu
Office Hours: *223 Carolina (Hurston) Hall | Tue & Thu 2-4 pm, or by appointment*

**Teaching Assistant:** Prabisha Shreshta | prabisha@live.unc.edu
Office Hours: *317 Carolina (Hurston) Hall | TBD*

**Graduate Research Consultant:** Kate Brandt | kebrandt@live.unc.edu

---

## COURSE DESCRIPTION

This course will introduce you to data science with a focus on spatial (geographic) data, which are data that are referenced to a particular location on Earth's surface. You will learn concepts, techniques, and tools you need to apply various facets of data science practice, including data collection, management, and integration, descriptive modeling, exploratory spatial data analysis, and effective communication via data visualization and mapping. Real world examples and datasets spanning physical, social, and health sciences will be used throughout the course in an effort to promote contextual learning.

## COURSE OBJECTIVES

The overall objective of this course is to introduce you to the core concepts, tools and intuition required to be a succesful spatial data scientist. This course is much more about "learning to learn" and problem solving rather than acquiring specific programming tricks or stats wizardry.

By the end of the semester you should be able to:

1. Pull down publicly available datasets from non-conventional data sources within R (no more pointing and clicking on the numerous weblinks)
2. Quickly and efficiently organize, clean, and visualize spatial datasets programmatically using tidydata principles
3. Describe and interpret interesting spatial patterns in data
4. Critically evaluate and choose appropriate Exploratory Spatial Data Analysis to analyze data
5. Write and communicate your results in a scientific and reproducible manner using a literate programming language such as Mardown

6. Answer useful research questions by applying the tools and principles of spatial data science to a real world dataset with an inter-disciplinary team

## COURSE STRUCTURE

To help you achieve the course objectives, we will primarily engage in active learning through numerous in-class activites and labs. You will be provided with readings and interactive modules to complete *before class* so that most of the class time can be used for discussion and hands-on activites instead of passive lecturing. Similarly, the homeworks, exams and the final project are aimed to help you solidify your understanding of the material and apply tools and principles learned in class. The main purpose of having these numerous *checks* is not to test you on what is right or wrong, but to help you think creatively and provide you with a platform to express your creativity. Since you will spend majority of the class programming in R, I hope that you consider programming as a medium (and an important skill for the job market) that enables you to be creative, but in a scientifically sound, transparent and meaningful manner.

Programming is a skill that is best learnt by doing, and programming languages, like any other language, are best learnt through practice, practice and practice. As such, the best way to learn R or any language is to fully commit to the language and do everything you can to practice it most days of the week for at least half an hour, and ideally more. Other than the materials covered in class, you can hone your R skills by writing other class homeworks in RMarkdown, creating a personal website in R, making presentations in R, or even sending emails through R. The main idea is to immerse yourself into the language, learn how to ask for help, and continuously practice what you learn. Although coding, just like any other language can seem frustrating at the beginning, I promise that this immersive approach will help in the short and long run. As your instructor, I will facilitate your immersion as much as possible through a variety of class activites and assignments. While I will introduce you to new tools and techniques, it is up to you to take them and make use of them.

The course is divided into the following *five* components:

### Classroom participation and discussion *(15%)*

Class sessions will include a mixture of lecture, discussion, and applied activities (conducted both individually and in small groups). All students are expected to not only attend class, but also be active participants. In addition to the classroom, we will be using **Piazza** for class discussion. Piazza is a widely used class discussion platform to assist you with getting help efficiently from (and providing help to) your classmates, the teaching assistant and myself. Rather than emailing questions to me, I encourage you to first post your questions on Piazza. Find our class page at: https://piazza.com/unc/spring2020/geog215/home Class participation and discussion will be graded on *both presence and active participation in the class and on piazza*.

### Labs *(20%)*

The objective of the labs is to give you hands on experience with data analysis using modern statistical software. The labs will also provide you with tools that you will need to complete the project successfully. All labs must be submitted electronically through *Sakai* by the **start** of class on the due date as indicated in the schedule below. The lab with lowest score for each student will be dropped.

### Homeworks *(15%)*

Beyond the in-class activities and labs, you will be given two larger data analysis tasks during the semester. The purpose of these homeworks is to combine different skills introduced in the labs to conduct an integrated data analysis. All assignments must be submitted electronically through *Sakai* by the **start** of class on the due date as indicated in the schedule below.

### Exams *(30%)*

There will be 2 exams during the semester. Each exam will require you to complete a number of small programming and/or analysis tasks related to the material presented in the class. The exams will be take home, and each exam will be written to take between 3-5 hours. The exams are to be completed individually and you are not allowed to talk about exam content to other students in the class. The exact structure and content of the exams will be discussed in more detail before they are assigned.

### Final project *(20%)*

You will complete a final project for this course, the goal of which is to tackle an interesting problem using the tools and techniques covered in this class. This project will be completed in teams assigned by me to ensure a good balance of skills. Additional details on the project will be provided as the course progresses. To assist you with your project, you will work with a Graduate Research Consultant, Kate Brandt. The GRC Program is sponsored by the Office for Undergraduate Research (http://our.unc.edu), and you may be able to use this research-exposure course to meet a requirement of the Carolina Research Scholars Program (http://our.unc.edu/students/crsp). I encourage you to visit the OUR website to learn about how you might engage in research, scholarship and creative performance while you are at Carolina.

### Grading Criteria

- Class participation: **15%**
- Labs: **20%**
- Homeworks: **15**
- Exams: **30%**
- Final project: **20%**

Grades will be based on the following low cutoff values, although I reserve the right to modify the exact values: **A** *(93%)*, **A-** *(90%)*, **B+** *(87%)*, **B** *(83%)*, **B-** *(80%)*, **C+** *(77%)*, **C** *(73%)*, **C-** *(70%)*, **D** *(60%)*, anything below *60%* is an **F**

### Required Text

The course has no required text. We will read chapters and articles that are freely available online. There will be roughly 1-3 readings per week, which will be posted on the course website and in Sakai. The expectation for readings is that you will be able to summarize and discuss the main points/topics/ideas, as well as to identify interesting or important aspects of each as they pertain to your personal interests or the broader course material.

### Prerequisites

There is no prerequisite for this course. However, some familiarity with spatial data, basic statistics, computation, and/or coding is recommended.

### Class Resources

All lectures, notes, readings and other resources will be posted both on the course website and on Sakai. We will use sakai for assignment submissions and grading.

### Computers and Software

In the course, we will be working with the programming language, R. R is an open-source programming language and software environment widely used in both industry (high employable) and academia. It can run on both Windows and Mac OS. A software document will be provided to assist you with getting set up with R. Computers should be brought to class *everyday* to complete in-class activities and demonstrations.

## OUTLINE AND SCHEDULE

The class schedule is subject to change as needed. ***(Deliverables are due before class on the mentioned date)***

| Week | Day | Date | Topic | Deliverables |
|---|---|---|---|---|
| **1** | Wed | 1/8 | Intro to Spatial Data Science, spatial data, & geographic information | |
| **2** | Mon | 1/13 | Intro to R and modern computing environments | |
| | Wed | 1/15 | R basics | |
| **3** | ***Mon*** | ***1/20*** | ***NO CLASS (MLK day)*** | ***Lab 1*** |
| | Wed | 1/22 | Tidy data: Intro to Tidyverse | |
| **4** | Mon | 1/27 | Representing location and space | **Lab 2** |
| | Wed | 1/29 | Collecting spatial data | |
| **5** | Mon | 2/3 | Acquiring spatial data | **Lab 3** |
| | Wed | 2/5 | Handling spatial data | |
| **6** | Mon | 2/10 | Principles of reproducible research | **Lab 4** |
| | Wed | 2/12 | Data wrangling I | |
| **7** | Mon | 2/17 | Data wrangling II | **Lab 5** |
| | Wed | 2/19 | Data visualization I | |
| **8** | Mon | 2/24 | Data visualization II | **Lab 6, HW1** |
| | Wed | 2/26 | Review for Exam 1 | |
| | ***Sun*** | ***3/1*** | *****EXAM 1 due at 5 pm***** | |
| **9** | Mon | 3/2 | Intro to exploratory spatial data analysis (ESDA) | |
| | Wed | 3/4 | Mapping spatial data | |
| **10** | ***Mon*** | ***3/9*** | ***NO CLASS (Spring Break)*** | |
| | Wed | 3/11 | NO CLASS (Spring Break) | |
| **11** | Mon | 3/16 | Descriptive spatial statistics | **Lab 7, Project Proposal** |
| | Wed | 3/18 | Interpreting spatial patterns | |

*(continued)*

| Week | Day | Date | Topic | Deliverables |
|---|---|---|---|---|
| **12** | Mon | 3/23 | Wrangling with spatial data | **Lab 8** |
| | Wed | 3/25 | Spatial weights | |
| **13** | Mon | 3/30 | Spatial autocorrelation I | **Lab 9, HW2** |
| | Wed | 4/1 | Spatial autocorrelation II | |
| **14** | Mon | 4/6 | Spatial clustering | **Lab 10** |
| | Wed | 4/8 | Review for Exam 2 | |
| | ***Sun*** | ***4/12*** | *****EXAM 2 due at 5 pm***** | |
| **15** | Mon | 4/13 | Advance topics in spatial data science | |
| | Wed | 4/15 | Data science in action (guest lecture) | |
| **16** | Mon | 4/20 | Project peer review | **Draft Project Report** |
| | Wed | 4/22 | Course wrap up and summary | |
| | ***Mon*** | ***4/27*** | *****FINAL Project Report due at 5 pm***** | |
| **18** | Mon | 5/4 | Final Exam Period (Starts at 4:00 pm) | **Presentations** |

# CLASS POLICIES

## Classroom Expectations

I encourage healthy, constructive discussion of the topics and course material. I operate under the assumption and requirement that everyone will be respectful of their fellow classmates in this endeavor. Unless your absence is university approved, I expect you to attend and actively participate in class; frequent absences or tardiness will be considered a legitimate cause for grade reduction. All of you bring your own unique skills and viewpoints to this class. I expect you to both learn from, and help your classmates by being active participants in the classroom and on piazza.

**Note on class attendence policy:** In general, you are not permitted to be absent from any class except for these University Approved Absences: - Authorized University activities - Disability/religious observance/pregnancy, as required by law and approved by Accessibility Resources and Service and/or the Equal Opportunity and Compliance Office (EOC) - Significant health condition and/or personal/family emergency as approved by the Office of the Dean of Students, Gender Violence Service Coordinators, and/or the Equal Opportunity and Compliance Office (EOC).

For situations when an absence is not University approved or fall under the above mentioned criteria (e.g., a job interview or club activity), please **communicate with me early** about potential absences. Please be aware that you are bound by the Honor Code when making a request for a University approved absence.

## Diversity Statement

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 the students bring to this class be viewed as a resource, strength and benefit. It is my intent

to present materials and activities that are respectful of diversity: race, gender identity, national origin, ethnicity, religion, social class, age, sexual orientation, political background, and physical and learning ability. 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.

Furthermore, I would like to create a learning environment for my students that supports a diversity of thoughts, perspectives and experiences, and honors your identities (including gender identity, sexuality, disability, age, socioeconomic status, ethnicity, race, nationality, religion, and culture.) To help accomplish this:

– If you have a name and/or set of pronouns that differ from those that appear in your official university records, please let me know.

– If you feel like your performance in class is impacted by your experiences outside of class, please don't hesitate to talk with me. I want to be a resource for you. I have completed both the Safe Zone and Mental Health First Aid trainings, and while I may not be able to directly help you, I can help you find and reach those that can help.

– If you prefer to speak with someone outside of the course, your academic dean is an excellent resource.

– I (like many people) am still in the process of learning about diverse perspectives and identities. If something is said in class (by anyone) that makes you feel uncomfortable, please talk to me about it.

### Late submissions

All assignments (labs,homeworks,exams,projects) will be posted on the course website and on Sakai with detailed instructions. All assignments must be submitted electronically prior to the due date/time. Late submissions will be penalized 20% for each day late. Late submissions will only go unpenalized for documented medical reasons or by previous agreement with the instructor.

### Changing Grades

Out of fairness, grades will not be changed unless I have made a data entry or arithmetic error.If I have made an error, please send me an email with an explanation of the error within 24 hours of grades being posted and I will check.

### Academic Integrity

No grade is important enough to justify academic misconduct. UNC-CH has an Honor Code with clear guidelines regarding academic integrity. Three fundamental and rather simple principles to follow at all times are that: (1) all work submitted be your own; (2) when using the work or ideas of others, including fellow students, give full credit through accurate citations; and (3) if you are uncertain about the rules on a particular assignment, ask for clarification.

**A note on sharing / reusing code:** I am well aware that a huge volume of code is available on the web to solve any number of problems. Unless I explicitly tell you not to use something the course's policy is that you may make use of any online resources (e.g. StackOverflow) but you must explicitly cite where you obtained any code you directly use (or use as inspiration). Any recycled code that is discovered and is not explicitly cited will be treated as plagiarism. On individual

assignments you may not directly share code with another student in this class. Except for the take home exams, you are welcome to discuss the problems together and ask for advice, but you may not send or make use of code from another team. On the take home exams all communication with classmates is explicitly forbidden.

### Professional Conduct

Students are expected to treat one another with the utmost respect at all times. Disrespectful behavior toward other students or instructors/TAs will not be tolerated and will result in disciplinary action and significant grade penalties of at least 2/3 of a letter grade off the student's final course grade. (If final course grades have already been submitted to the Registrar, any grade penalties associated with disrespectful of unprofessional conduct will be counted as arithmetic errors that occurred during the semester and submitted as amendments to the student's transcript.)

### UNC-CH Resources

- The Writing Center, http://writingcenter.unc.edu/
- Research Hub, http://library.unc.edu/hub/
- University Libraries, http://library.unc.edu/
- Counseling and Psychological Services, https://campushealth.unc.edu/

### Accessibility

UNC at Chapel Hill facilitates the implementation of reasonable accommodations, including resources and services, for students with disabilities, chronic medical conditions, a temporary disability or pregnancy complications resulting in difficulties with accessing learning opportunities. All accommodations are coordinated through the Accessibility Resources and Service Office. See the ARS Website for contact information: accessibillity.unc.edu