

GTECH 38520 / 78520
Data Analysis and Visualization with R

Spring 2023, Tuesday 11:30 AM – 2:20 PM, HN1090B-1
Instructors: Allan Frei, Peter Marcotullio, Shipeng Sun

Instructor Information

Allan Frei (he/him/his)

Office: 1005, Hunter North
Email: afrei@hunter.cuny.edu

Office hours: By appointment

Peter J. Marcotullio (he/him/his)

Office: 1003e, Hunter North
Email: pmarcotu@hunter.cuny.edu

Office hours: See Blackboard
Phone: 212-772-5264

Shipeng Sun (he/him/his)

Office: 1043, Hunter North
Email: shipeng.sun@hunter.cuny.edu

Office hours: M, Th 4-5:15pm
Phone: 212-396-6039

Course Description and Objectives

This course introduces the basics of R and the practical knowledge of data cleaning, re-organization, modeling, statistics, and analysis for research and visualization, particularly in geospatial fields. The class meets once a week for 3 hours per session with hands-on practices. Three different members of the Department of Geography and Environmental Science, as identified above, lead the class.

The goal of the course is to introduce students to the use of R programming for univariate and multivariate analysis and visualization, mapping and spatial analysis.

Learning Outcomes

At the end of the semester, students will be able to

- 1) Use RStudio to perform basic data analysis functions including Input/Output, basic statistical analyses, Exploratory Data Analysis (EDA), and graphical output.
- 2) Use RStudio to perform regression analysis, and to develop, test, and execute R script.
- 3) Use advanced R programming to import, clean, transform, and summarize data
- 4) Use *ggplot2* to visualize data in points, lines, area charts and smoothed curves
- 5) Import and map spatial data using R *sf* and *ggplot2* packages
- 6) Conduct basic cluster analysis and regression using spatial data

Pre-requisite

Required: STAT 113 or 213

Suggested but not required: GTECH 201, GTECH 301, or GTECH 702

A computer language background is helpful, but not required.

Course Materials

Required Textbook: None

Recommended Books:

- *ggplot2, Elegant Graphics for Data Analysis* (2nd Edition), by Hadley Wickham, Springer, (2016);
- *R for Data Science, Import, Tidy, Transform, Visualize and Model Data*, (1st Edition) by Hadely Wickham and Garrett Golemund, O'Reilly (2016).
- *Geocomputation with R* by Robin Lovelace, Jakob Nowosad, Jannes Muenchow (2023). Available at <https://geocompr.robinlovelace.net/>.
- *Spatial Data Science with R* by Robert J. Hijmans (2021). Available at <https://rspatial.org/raster/index.html>

Schedule

The course has 6 learning sections, one midterm, and one final exam. Each professor teaches two of the following sections.

1. Introduction to R, getting started, Exploratory Data Analysis (EDA)
2. Exploration and Manipulation of Large Data Sets, <https://www.analyticsvidhya.com/blog/2015/12/faster-data-manipulation-7-packages/>
3. Basic statistics in one dimension (not spatial), R Programming
4. Visualization of non-spatial data with *ggplot2* package
5. Spatial data organization, I/O, format conversion, and spatial reference systems
6. Visualization of spatial data, mapping, and basic spatial analysis

Software

- R 4.0 or above. Free at www.r-project.org. Must be installed prior to installing RStudio.
- RStudio Desktop (Open-Source License, not the Commercial License) 2021 or later. Download Free at <https://posit.co/products/open-source/rstudio/>.

Grading Scale

Academic performance evaluation criteria include lab assignments, exams and class participation with breakdown as follows.

| Components | Percentage |
|-----------------|------------|
| Lab Assignments | 60% |
| Exams | 30% |
| Participation | 10% |

Course Calendar & Content

This schedule is subject to change.

| Week | Session | Date | Topic | Instructor |
|------|---------|-----------|--|--------------------|
| 1 | 1 | 1/31/2023 | Introduction to this course; Introduction to R and RStudio | Frei |
| 2 | 2 | 2/7/2023 | Univariate EDA | Frei |
| 3 | 3 | 2/14/2023 | Bivariate EDA | Frei |
| 4 | | 2/21/2023 | No Class, Monday Schedule | |
| 5 | 4 | 2/28/2023 | Enhanced Graphics | Frei |
| 6 | 5 | 3/7/2023 | Advanced importation, cleaning, transformation, summarizing and saving data, <i>tidyverse</i> , Part I | Marcotullio |
| 7 | 6 | 3/14/2023 | <i>tidyverse</i> , Part II | Marcotullio |
| 8 | 7 | 3/21/2023 | Visualization of non-spatial data with <i>ggplot2</i> package, Part I | Marcotullio |
| 9 | 8 | 3/28/2023 | Midterm Review | Frei & Marcotullio |
| 9 | | | Midterm available online at 5 pm. | |
| 10 | | 4/3/2023 | Midterm due at 11:59pm | |
| 10 | 9 | 4/4/2023 | <i>ggplot2</i> , Part II. | Marcotullio |
| 11 | | 4/11/2023 | Spring Recess | |
| 12 | 10 | 4/18/2023 | GIS Basics, <i>sf</i> package, spatial data types, spatial data I/O | Sun |
| 13 | 11 | 4/25/2023 | Re-projecting and mapping spatial data with <i>ggplot2</i> | Sun |
| 14 | 12 | 5/2/2023 | Spatial Join and Basic Spatial Analysis | Sun |
| 15 | 13 | 5/9/2023 | Spatial Autocorrelation and Regression | Sun |
| 16 | 14 | 5/16/2023 | Application of R-Spatial | Sun |
| 17 | 15 | 5/23/2023 | Final Exam | |

Syllabus Change Policy

Except for changes that substantially affect implementation of the evaluation (grading) statement, this syllabus is a guide for the course and is subject to change with advance notice. Changes will be announced in class and on Blackboard, which students are expected to check regularly during the semester.

Policies

Preferred Names and Pronouns

All people have the right to be addressed and referred to in accordance with their personal identity. In this class, we will have the chance to indicate the name that we prefer to be called and, if we choose, to identify pronouns with which we would like to be addressed. The instructors will do their best to address and refer to all students accordingly and support classmates in doing so as well.

Attendance

Attendance is crucial. Given the active class-learning environment, meaning that the basis of most your performance is practical assignments rather than tests, adherence to protocols and the course timetables is very important. The instructors and all students will be on time. If you are late, you will miss important information. Your punctuality is a common courtesy and prevents repetitive instructions. Instructors will not tolerate consistent tardiness for either lectures or laboratory/discussion sections. We evaluate involvement in the course through the quality of your assignments. Learning a computer language requires hours of practice. In so doing, the tools gained become a means to an end, rather than the end themselves. For example, in this class you will perform statistical analyses, manipulate data and create many maps. Each of these tasks requires creativity and results in a degree of satisfaction. Remember, however, the results, data frames and maps are for particular scientific purposes.

A significant part of your grade depends upon class participation. Class participation includes attendance at laboratory sessions, participation in organized class discussions, submission assignments on time, and participation in the map poster display (if this is a part of the course this semester). We expect respectfully behavior towards instructors and the other students, through non-threatening language during discussions, and by allowing others to speak, especially if those others are shy and their voices are weak.

Electronics in the classroom

We forbid electronic recording device use during lectures as are personal electronics including cell and smart phones. Computers may be used for taking notes and doing the lab exercises, and if you use them for activities not related to classroom content (personal e-mails, Facebook chats, surfing the Internet for fun ...), you will be asked gently, but firmly, to turn them off.

Course Website

Web-enhancement in the context of this course means that everything pertaining to this course will be communicated through BlackBoard. You are required to check the BlackBoard course site on a daily basis. All changes to the syllabus will be announced on the course home page. All lecture and lab materials are accessible through BlackBoard, and Blackboard is also the place where you upload your assignments. Your exams and lab assignments will be graded based on what you have uploaded to BlackBoard and that is where you will find your grades and may access course statistics that help you to assess your standing at any given time.

Communication

All email messages about this course should be signed with your full name as it appears in CUNYfirst. [Professionalism](#) and “[netiquette](#)” are expected in the communication through emails (check out those links). If your emails are not replied to in a timely fashion, please consider rewriting your emails in a better way.

General Lab Policies

Lab policies are described in detail in

<http://www.geo.hunter.cuny.edu/techsupport/rules.html>.

Hunter College Policy on Academic Integrity

Hunter College regards acts of academic dishonesty (e.g., plagiarism, cheating on examinations, obtaining unfair advantage, and falsification of records and official documents) as serious offenses against the values of intellectual honesty. The College is committed to enforcing the CUNY Policy on Academic Integrity and will pursue cases of academic dishonesty according to the Hunter College Academic Integrity Procedures.

ADA Policy

In compliance with the American Disability Act of 1990 (ADA) and with Section 504 of the Rehabilitation Act of 1973, Hunter College is committed to ensuring educational parity and accommodations for all students with documented disabilities and/or medical conditions. It is recommended that all students with documented disabilities (Emotional, Medical, Physical, and/or Learning) consult the Office of AccessABILITY, located in Room E1214B, to secure necessary academic accommodations. For further information and assistance, please call: (212) 772-4857 or (212) 650-3230.

Hunter College Policy on Sexual Misconduct

In compliance with the CUNY Policy on Sexual Misconduct, Hunter College reaffirms the prohibition of any sexual misconduct, which includes sexual violence, sexual harassment, and gender-based harassment retaliation against students, employees, or visitors, as well as certain intimate relationships. Students who have experienced any form of sexual violence on or off campus (including CUNY-sponsored trips and events) are entitled to the rights outlined in the Bill of Rights for Hunter College.

a. Sexual Violence: Students are strongly encouraged to immediately report the incident by calling 911, contacting NYPD Special Victims Division Hotline (646-610-7272) or their local police precinct, or contacting the College's Public Safety Office (212-772-4444).

b. All Other Forms of Sexual Misconduct: Students are also encouraged to contact the College's Title IX Campus Coordinator, Dean John Rose (jtrose@hunter.cuny.edu or 212-650-3262) or Colleen Barry (colleen.barry@hunter.cuny.edu or 212-772-4534) and seek complimentary services through the Counseling and Wellness Services Office, Hunter East 1123.

CUNY Policy on Sexual Misconduct Link:

<http://www.cuny.edu/about!administrationoffices/Ja/Policy-on-SexualMisconduct-12-1-14-with-links.pdf>