SOC 591: R for Data Analytics and Visualization

Yongjun Zhang

2023-01-20

Course Logistics

- Term: Spring 2023
- Time & Location: Wed 9:15AM - 12:05PM; SBS-N403
- Instructor: Dr. Yongjun Zhang
- Email: Yongjun.Zhang@stonybrook.edu
- Office Hours: Appointments as needed

Course Description

A multidisciplinary introduction to data analytics and visualization using R, emphasizing how social scientists can use open-source software to understand and analyze social behavior in the digital era. Topics include how to use R to collect, process, analyze, and visualize data from the real world to address social problems. This course also introduces state-of-the-art tools for data collection, data wrangling, data modeling, and data visualization.

Course Learning Objectives

This course offers students a set of data analytic and visualization toolkit to acquire the knowledge or skills necessary to achieve the following learning outcomes:

- Understand how to use R to process, model, visualize, and communicate data.
- Understand the methods of inquiry used by social scientists to explore social and behavioral phenomena.
- Skillfully interpret and form educated opinions on social science issues.
- Master the ability to apply computational tools and knowledge to problem-solving.
- Design and build computational systems to explore and analyze some aspects of the human world.

Textbooks

Required:

- Hadley Wickham and Garrett Grolemund. R for Data Science. O'Reilly Media. https://r4ds.had.co.nz/
- Kieran Healy. Data Visualization. https://socviz.co/

Optional:

- Hadley Wickham. Advanced R. https://adv-r.hadley.nz/

Course Requirements and Evaluation

The course will be broadly divided into four modules: R Basics, Data Wrangling, Data Modeling, and Data Visualization. Each module will introduce the basic and latest methods as well as relevant social research using associated methods. Students can choose the specific module that is particularly helpful in their research to develop their final research proposal. For each meeting, it is a mix of mini-lecture, student mini-presentation, instructor or student-led discussions, and lab training. In the mini-lecture, the instructor will give an overview of the corresponding methods. In the mini-lab, the instructor will use R to walk through each method and prepare students with the necessary computational skills for their research. Students will be evaluated based on the following aspects:

Research Proposal. Students are required to develop a short research proposal integrating at least one of the methods introduced in the course. Particularly students are required to use large-scale administrative and digital trace data hosted by Google BigQuery and other platforms or scraped by themselves.

Research Presentation. Students are required to present the research proposal and final project in the class. Students are expected to use at least one of the CSS methods learned in the class.

Crowd-source Participation. Students are required to present in the classroom about some R packages at least twice. This is part of the crowd-sourced efforts to learn more about how to use R efficiently.

Class Participation. Students are required to attend every session and prepare the assigned readings before each session. Students are also required to participate in the class discussion. Students will receive no credits if they have over 2 times of absence.

Course Policies

Make-up exams are not allowed without prior arrangements and documentation of extenuating circumstances. Please speak with the instructor regarding any known absences or emergencies ASAP to avoid any issues regarding assignment days. Late assignments are not accepted without prior arrangements with the instructor.

We all share responsibility for maintaining an appropriate learning environment. For this reason, please mute yourself if attending zoom meetings or when others are speaking so that your peers are not distracted. Finally, all offline or online classroom behavior and discourse should reflect the values of respect and civility.

Composition of Final Grades

- Research Paper x 1 = 30
- Research Presentation x 1 = 30
- Crowd-source Participation x 1 = 20
- Class Participation = 20
- Total 100 points

Grade Scale:

- 95-122 = A
- 90-94 = A-
- 85-89 = B+
- 80-84 = B

- 75-79 = B-
- 70-74 = C+
- 65-69 = C
- 60-64 = C-
- 0-59 = F

Course Schedule

Note: The instructor reserves the right to modify the schedule as deemed necessary. Flexibility throughout the semester will allow us to incorporate the latest computational social science methods into the course.

WEEK 1 (01-25-2023) Welcome and A General Introduction

Assigned Readings:

1. David M. J. Lazer et al. 2020. "Computational social science: Obstacles and opportunities." Science, 369, 6507, Pp. 1060-1062. Publisher's Version Copy at https://j.mp/2YIuWdh
2. Edelman et al. 2020. "Computational Social Science and Sociology." Annual Review of Sociology.https://doi.org/10.1146/annurev-soc-121919-054621
3. Healy, K. and Moody, J., 2014. Data visualization in sociology. Annual review of sociology, 40, pp.105-128.

Lab:

1. Install all necessary software, including R and Rstudio. see https://cran.r-project.org/ and https://posit.co/downloads/
2. Learn how to install and manage packages using `install.packages`, `devtools`, and `pacman`.
3. Understand how to use the command line, like how to run R using terminal.
4. Learn how to efficiently use Rstudio (e.g., create project) and Rmarkdown (e.g., write your code)
5. Github and version control; understand basic git commands, like git clone, git fetch, git pull, and git push; See here for more details: https://help.github.com/en/github/getting-started-with-github/set-up-git

WEEK 2 (02-01-2023) R Basics

Assigned Readings:

1. Data type and structure: Advanced R (Foundations) http://adv-r.had.co.nz/Data-structures.html
2. More on data frame in R, see https://tibble.tidyverse.org/.
3. Data import and export with `tidyverse`,`fread`,`haven`, and `readxl`.
4. Workflow and Pipe coding with `magrittr`: https://magrittr.tidyverse.org/
5. Functional programming: http://adv-r.had.co.nz/Functional-programming.html and https://purrr.tidyverse.org/
6. Regular expression with `stringr`: https://stringr.tidyverse.org/
7. Good coding style: https://style.tidyverse.org/index.html

Lab:

1. Download gss data
2. Use learned skills to create variables and save them into a csv file or rdata file.

WEEK 3 (02-08-2023) Data Preparation

Assigned Readings:

1. Working with Census Data using `tidycensus`:https://walker-data.com/tidycensus/ and Chapter 1-3: https://walker-data.com/census-r/
2. Working with Twitter data using Twitter Academic API `academictwitteR`: https://developer.twitter.com/en/products/twitter-api/academic-research
3. Webscraping using `Rselenium`:https://docs.ropensci.org/RSelenium/

Lab:

1. Understand `Rselenium` and write a rscript using `Rselenium` to scrape a webpage.
2. Use `tidycensus` to create a county-level dataset containing unique GEOID, total population, Blacks, Whites, Hispanics, and Asians.

WEEK 4 (02-15-2024) Data Wrangling with `Tidyverse` (1)

Assigned Readings:

1. Tidy data: https://www.jstatsoft.org/article/view/v059i10
2. Read textbook R for data science about data wrangling module.

Lab:

1. Learn basics on data transformation
2. Learn basics on how to join tables and transform dataset structure

WEEK 5 (03-01-2023) Data Wrangling with `Tidyverse` (2)

1. Processing date and time with `lubridate`
2. Working with text data with `readtext`, `quanteda`, and `tidytext`
3. Basic textual analysis using R

WEEK 6 (03-08-2023) Data Modeling with `tidymodels` and `modelr`

Assigned Readings

1. Model part in R for Data Science: https://r4ds.had.co.nz/model-intro.html
2. Tidymodels:https://www.tidymodels.org/packages/
3. Statistical analysis uing `lmer4`,`fixest`, etc.

Lab:

1. Use tidycensus to compile county-level dataset
2. Run regression with fixed effects models

WEEK 7 (03-15-2022) Spring Break, no class

WEEK 8 (03-22-2023) Data Modeling and Summary with `modelsummary`,`stargazer`,`coefplot`,`effects`

Assigned Readings

1. Modelsummary: https://vincentarelbundock.github.io/modelsummary/articles/modelsummary.html
2. Stargazer: https://www.jakeruss.com/cheatsheets/stargazer/

Lab:

1. Load gss data
2. Run regression
3. Present your regressin tables and figures.

WEEK 9 (03-29-2023) Supervised Machine Learning with `caret`

Assigned Readings:

1. Grimmer, Justin, Margaret E. Roberts, and Brandon M. Stewart. "Machine Learning for Social Science: An Agnostic Approach." Annual Review of Political Science 24 (2021): 395-419.
2. Molina, Mario, and Filiz Garip. "Machine learning for sociology." Annual Review of Sociology 45 (2019): 27-45.
3. Athey, Susan, and Guido W. Imbens. "Machine learning methods that economists should know about." Annual Review of Economics 11 (2019): 685-725.
4. Max Kuhn and Kjell Johnson. Applied Predictive Modeling.
5. R caret Package: https://topepo.github.io/caret/

Lab:

1. Using R caret package to do some basic supervised machine learning

WEEK 10 (04-5-2022) Data Visualization Basics with `ggplot2`, `plotly`, `shiny`

Assigned Readings:

1. Healy, K., 2018. Data visualization: a practical introduction. Princeton University Press.
2. Wickham, Hadley. "Programming with ggplot2."

Lab:

1. Use ggplot2 to make your first graph
2. Use plotly to make your graph interactive

WEEK 11 (04-12-2023) Geospatial Mapping with `sf`,`tigris`,`usmap`,`ggmap`

Assigned Readings:

1. Analyze census data using tigris: https://walker-data.com/census-r/census-geographic-data-and-applications-in-r.html
2. Mapping census data with R: https://walker-data.com/census-r/mapping-census-data-with-r.html
3. Spatial analysis with census data using R https://walker-data.com/census-r/spatial-analysis-with-us-census-data.html

Lab:

1. Use tidycensus or tigris to download U.S. county level base map
2. Visualize county-level population size.

WEEK 12 (04-19-2023) Social Network Analysis and visualization with `igraph`,`stanet`,`ggnet`, `visNetwork`, and `networkD3`

Assigned Readings:

1. Igraph tutorial: https://igraph.org/r/
2. https://kateto.net/netscix2016.html
3. Statnet workshop: https://statnet.org/workshop-intro-sna-tools/

Lab:

1. Use igraph, ggenet, ggplot, or network D3 to visualize Twitter network data set
2. Get data here: https://snap.stanford.edu/data/twitter-2010.html

WEEK 13 (04-26-2022) Research Paper Presentation

WEEK 14 (05-03-2023) Last day of class, work on your paper (due on midnight)

Technical and Software Requirements

Students need a stable internet environment to access the Brightspace and Zoom. If you have any questions or difficulties, please let me know. You also need to set up R and Rstudio. Students will have instructions to install those free software and associated packages or modules.

If you need technical assistance at any time during the course or to report a problem with Brightspace you can: - Phone: 631-632-9800 (client support, Wi-Fi, software and hardware) - Submit a help request ticket: https://it.stonybrook.edu/services/itsm - If you are on campus, visit the Walk-Up Tech Support Station in the Educational Communications Center (ECC) building.

For laptop loans: https://www.stonybrook.edu/commcms/studentaffairs/studentsupport/

For IT support: https://it.stonybrook.edu/services/itsm

Student Accessibility Support Center Statement

If you have a physical, psychological, medical or learning disability that may impact your course work, please contact the Student Accessibility Support Center, ECC (Educational Communications Center) Building, Room 128, (631)632-6748. They will determine with you what accommodations, if any, are necessary and appropriate. All information and documentation are confidential.

Academic Integrity Statement

Each student must pursue his or her academic goals honestly and be personally accountable for all submitted work. Representing another person's work as your own is always wrong. Faculty are required to report any suspected instances of academic dishonesty to the Academic Judiciary. Faculty in the Health Sciences Center (School of Health Technology & Management, Nursing, Social Welfare, Dental Medicine) and School of Medicine are required to follow their school-specific procedures. For more comprehensive information on academic integrity, including categories of academic dishonesty please refer to the academic judiciary website at http://www.stonybrook.edu/commcms/academic_integrity/index.html

Critical Incident Management

Stony Brook University expects students to respect the rights, privileges, and property of other people. Faculty are required to report to the Office of Student Conduct and Community Standards any disruptive behavior that interrupts their ability to teach, compromises the safety of the learning environment, or inhibits students' ability to learn. Until/unless the latest COVID guidance is explicitly amended by SBU, during Spring 2022 "disruptive behavior" will include refusal to wear a mask during classes. For the latest COVID guidance, please refer to: https://www.stonybrook.edu/commcms/strongertogether/latest.php

Copyright Notice

Other Useful Materials

Student Success Resources: A helpful resource is the "For Students" section linked from the Stony Brook homepage: http://www.stonybrook.edu/for-students/ as well as the Division of Undergraduate Education website: http://www.stonybrook.edu/due.

Academic Success and Tutoring Center: This important program opened in September 2013. Please be sure that your students are aware of the available services. Information can be found at: http://www.stonybrook.edu/commcms/academic_success/

Support for Online Learning: https://www.stonybrook.edu/online/

Writing Center: Students are able to schedule face-to-face and online appointments. https://www.stonybrook.edu/writingcenter/

Career Center: The Career Center's mission is to support the academic mission of Stony Brook University by educating students about the career decision-making process, helping them plan and attain their career goals, and assisting with their smooth transition to the workplace or further education. Phone: 631-632-6810; email: sbucareercenter@stonybrook.edu; website: http://www.stonybrook.edu/career-center/

Counseling and Psychological Services: CAPS staff are available by phone, day or night. http://studentaffairs.stonybrook.edu/caps/