Data Science and Statistics for Social Sciences I
CS&SS/SOC/STAT 321
Winter 2024
Section Syllabus

Ramses Llobet
Ph.D. Student
Department of Political Science, UW

Section	Office Hours
AC :: MW 08:30 – 09:20 am (LOW 102)	In zoom, by appointment.
AD :: MW 09:30 – 10:20 am (LOW 102)	Email: *rllobet@uw.edu*

Overview. Sections are designed to complement lectures by reviewing theoretical concepts and learn computational skills in R. We meet twice a week, on Monday and Wednesday. The section contents are divided into modules dedicated to "best practices" in R programming, theory review, data wrangling, visualization, and statistical analyses in R, consolidating techniques learned in lectures and QSS tutorials while introducing new skills relevant to the course contents. All lab materials can be found in my course website, including Zoom recordings of each lab section.

Office hours. I will offer office hours by appointment in Zoom. Please note that it *may* take me up to 24 hours to respond to a student's email, so it is a good idea to plan ahead and email me in advance. When you email me, please include (1) the topic you would like to discuss and (2) your time availability for scheduling a meeting.

In addition, I can also offer spontaneous assistance in Slack, but please email me if you see that I take more than 24 hours in replying you in Slack. I may not answer emails sent past regular working hours, and I recommend you email me rather

than send me messages on the Canvas messaging system.

Participation through Slack. A portion of your final grade depends on section participation, which I will monitor through Slack. Slack, designed for team communication, collaboration, and project management, organizes communication into channels, facilitating real-time messaging.

In some quiz sections, we will use Slack will for sharing `R` code answers and collaborative problem-solving during in-section data analysis exercises. Participation is also considered when addressing questions or bug errors shared in Slack, either by you or your peers (see blow). For final projects, students are encouraged to create private Slack channels to share files, documents, and `R` code.

Programming Assistance. Slack is the most preferred communication channel, which allows you to insert code block in your messages. It has the added benefit of facilitating knowledge spillover through peer discussion and mutual assistance. Please post your questions on **Slack** related to **R** programming, graphic packages, or debugging. When you post a question, the best practice is to create a "minimal, reproducible example," instead of taking a screenshot of a code snippet (see here and here).

Alternatively, please feel free to set an appointment for office hours for further consultation, or email/'Slack' me your questions. Please note that if you want me (or someone else) to debug your code, you should first create a minimal reproducible example of the error(s) along with the required data file for code execution.

Homework Submission. Please submit your homework in PDF. You must use **RMarkdown** to integrate plain text, graphic outputs, and code chunks which can then be rendered ("knitted") into a single PDF output. You will have to submit your homework PDF on the Canvas course website (in assignments).

Section schedule. Below is the tentative schedule of sections and the associated topics to cover during lab sections, which are subject to adjustment depending

on our progress and learning needs:

Week	Monday	Wednesday
1	*No section*	Introduction; R setup
2	Project and file management	Rmarkdown and knitting PDF
3	Getting help and debugging	logical relations and subsetting
4	quantiles and NA data	Base R graphics
5	pivoting and merging data	Intro to ggplot2
6	Inference: samples and populations	correlation vs. causality
7	Bivariate regression	Prediction error and non-linearity
8	What is a standard error?	Uncertainty and hypothesis test
9	Multivariate regression	Selection bias and missing data
10	Good vs. Bad controls	Within-group regression

Module Outline

Week 1 - 3, Module 1: Getting started with R/RStudio. Provides an overview of the sections and logistics. How to install R/RStudio, how to install packages, how to create R projects and manage working directories, introduction to R Markdown, and how to create and share minimal, reproducible examples for programming assistance through Slack.

Week 3 - 5, Module 2: Data management and exploratory visual analysis. Provides a boot camp in Base R, some tidyverse functions, and an introduction to visualization with ggplot2. This module equips students with essential computing skills for data science to successfully complete course assignments and their final projects.

Week 6 - 8, Module 3: Introduction to inference and linear models. Introduction to two causation paradigms in science: randomization and causal modeling. This module familiarizes students with statistical and scientific (causal) inferences. Topics include the distinctions between estimand, estimator, and estimate, the

bias–variance trade-off, standard errors, and the interpretation of confidence intervals. While theory will be reviewed, the approach will be predominantly computational and visual.

Week 9 - 10, Module 4: Multivariate regression and causal modeling. Extension of the bivariate case to the multivariate specification, emphasizing DAGs and causal modeling for observational research designs. Topics include the variance-covariance matrix, control specification selection via minimum adjustment criteria, and linear regression with group-specific intercepts.

Additional Resources. If you're looking for a dataset for your project, make use of research data repository such as Harvard Dataverse, or generic search engine such as Google Dataset Search. In addition, TidyTuesday project offers many interesting datasets with coding examples by community members, and is being updated weekly.