Course Syllabus

Meeting Location: Armitage 219

Class Meeting Time: Mondays, 2-3, p.m.

Hello everyone! Welcome to Applied Statistical Programming!

This class will introduce you to the logic of data analysis, with a focus statistical programming using ChatGPT, the R language, and the R Studio IDE. For some of you, the previous sentence will sound ferocious and daunting. But don't worry--you can do this, and once you get started, you'll find that learning some statistical programming will allow you to answer all sorts of questions.

The goals of this course include:

1. Introducing AI into the workflow of prevention science;

2. Gaining familiarity with different types of data;

3. Learning the techniques of reproducible data analysis;

4. Acquiring skills in cleaning and importing data;

5. Learning how to transform variables;

6. Developing basic proficiency in visual displays;

7. Using these skills to prepare a professional quality poster.

There's much more to come in this syllabus, but I wanted to provide enough information for you to get started so that you're ready for our first class.

Please sign up for ChatGPT Plus before the first class ([https://chat.openai.com)](https://chat.openai.com)

[Links to an external site..](#) This will cost you $20 a month. We will use the data analysis capabilities of ChatGPT Plus for the first class.

Please add Rutgers' VPN software to your computer.

https://it.rutgers.edu/guides/remote-access-with-anyconnect-virtual-private-network/

[Links to an external site.](#)

The VPN will permit you to log into the virtual computer labs, which we will be using for class.

https://it.rutgers.edu/virtual-computer-labs/

Please sign up for a Box account. https://it.rutgers.edu/box/

Install the box drive client on your computer: https://www.box.com/resources/downloads

The book required for this class is *Data Analysis for Social Science*, by Elena Llaudet and Kosuke Imai, and it is available from Amazon (https://www.amazon.com/Data-Analysis-Social-Science-Introduction/dp/0691199434/ref=sr_1_1?crid=2VMPQO0IFZSYL&dib=eyJ2IjoiMSJ9.W_LZ09dqDPFakLXDZmAJgeBuLhLABR-7t1G_-vevqorGimEVOZt8QesHbbtZf6pJx1ycuGHqR2rkRl9THvSsw6SPtXFLIzAR_A-NWliTlIoigtGHfyzQhbzPJduxqeYcb2W1aoyghYI2f4hMW7PtQfbUkIItQFf7reNiaDyyHc0T13c5JBGuXtjQQKLje4ei5MhfiSGX5ifKBK1Cu1HYm8Z69Ny30JgjW1vK3gdECZw.E0tST_qwCJWOt_FFBkHAfQw2qDYioOiVVUVqP0hwX84&dib_tag=se&keywords=data+analysis+for+social+science&qid=1724269280&sprefix=data+analysis+fo%2Caps%2C103&sr=8-1

).

The key to success in this course is to do the work and learn some statistical programming and data analysis.

Your grade in this course will be based on in-class weekly skill checks and your final poster. I expect everyone to do well in this class. You need all the skills we will cover in this class to succeed as a prevention scientist.

As some of you know, the advent of ChatGPT is potentially revolutionary for higher education and computer programming. It's difficult to anticipate, right now, what the future holds in terms of AI-informed work.

One consequence is that this year's version of this course is radically different than last year's. This year, we will rely much more on understanding data analysis and less on learning programming. Another consequence is that it's very difficult to anticipate how quickly we can move through the course.

Your final project will involve using one kind of social capital data from the social capital atlas (https://socialcapital.org

at one level (zip code level, county level, high school, college) and one of the following (gun violence: https://www.gunviolencearchive.org/ , new jersey educational data: https://www.nj.gov/education/doedata/

, or health data: https://www.cdc.gov/places/index.html

). Your broad interests can be incorporated with some combination of these data sets.

Your final project, again, will be a poster reporting original hypotheses, graphs, and statistical analyses suitable for submission and presentation at a conference.

Here's the tentative schedule.

Month	Day	Topic	Overview	Skills to be tested in class
september	9	Introduction	Overview of R, R studio, Box, Chatgpt for data analysis	
	16	Chapter 1	Use chatgpt for data analysis with STAR data set	load STAR into chatgpt, answer some questions about the data.
	23	Chapter 2	Causal effects	Answer simple questions about the STAR study , potential outcome variables, outcome variables, fundamental problem of causal infer causal effect
	30			Difference in means estimator; create new variables; limitations of effect
october	7	Chapter 3	Survey/questionnaire data	Answer questions about samples (representative samples, random frequency tables; Proportions/percentages by row, column, total;
	14		Missing data	cross-tabulations; histograms; skills; density histograms; descripti SD); scatter plots; correlations
	21	Chapter 4	Linear regression	Answer questions about line of best fit, prediction, intercept, slope
	28			Predicting values, natural log transformations, model fit
November	4	Chapter 5	Causal Inference with Observational Data	Answer questions about confounders and importance for causal in
	11	EPA idea workshop		Plausible idea using social capital data and data from CDC

	18	EPA analyses	Initial analyses appropriate for a poster
December	25	EPA poster	Draft of poster
December	1	EPA deadline	Submission
December	2	Chapter 6	Contrast Frequentist/Bayesian perspectives
	9	End of the year discussion	