R Programming for Data Analysis

Instructor: Fred LaPolla, MLS

fred.lapolla@nyulangone.org
212.263.8535
577 First Ave, Room 209,
New York, NY 10016

Teaching assistant: Sana Badri, PhD

Sana.badri@nyulangone.org

Dates: Mondays and Thursday 10:00-11:30am, July 8 - August 16, 2021

Location:
https://nyulangone.zoom.us/meeting/register/tJ0sdu2prTMrG9zM1G8r3yLWnKZLbWf--Xtg

Office Hours: Tuesdays 9:00-10:00am,
https://nyulangone.zoom.us/j/92383292678?pwd=bU9KWUR6SXpJZHBFU0NRSGZOcURFdz09

Thursday 3:00-4:00pm
https://nyulangone.zoom.us/j/96832448857?pwd=Smx2aGFwcnBwMk1sUU4zQUNnYURWZz09

Course Site: https://github.com/fredlapolla/RVilcekMasters2021

Overview:

Understanding code-based approaches to data analysis provide researchers with career opportunities, as well as the ability to conduct reproducible, rigorous research without having to invest in expensive software. R is a widely used language for data science and statistical analysis and provides researchers with tools to unlock insights from their data. This class will introduce students to R using the RStudio development environment, walk them through features and functions of the program to complete original analysis on a dataset of their choosing.

Objectives:

Students will be able to:

- Perform statistical analyses in R using R Studio
- Choose appropriate statistical tests to uncover insights in a dataset of their choosing.
- Use R Markdown to generate a reproducible report suitable for presenting their findings to others.

Class Policies:

If you have any reason for needing accommodation of these policies, please inform the instructor or have the appropriate department contact him to inform of arrangements. Homework must be submitted on time or you will lose points on your assignment. If there is a valid reason it must be late, you need to discuss the possibility of delay **before** the assignment is due to see if an extension can be made.

Homework should be submitted on Brightspace to the class session's dropbox.

You are expected to do your own work and site sources for materials referenced/used.

Schedule (topics and dates subject to change)

Date	Topic	Progress toward final project
July 8, 2021	Syllabus review, discussion of final project, discussion of how to identify data sources	Explanation of final project and data finding
July 12, 2021	Introduction to R and the RStudio Environment, troubleshooting, packages	
July 15, 2021	Indexing and Data Types and Structures in R	Exploration of a dataset
July 19, 2021	Functions and Packages	Students should have identified a data set to work with and explore
July 22, 2021	Apply, For Loops and If	

	statements	
July 26, 2021	Data Cleaning and Dealing with Missing Values	Students should begin to identify a research question of interest to address from their dataset
July 29, 2021	Data Visualization for exploratory data analysis	
August 2, 2021	Data Visualization for publication	Students should be identifying variables to use to answer their research question
August 5, 2021	Hypothesis testing and analysis in R	
August 9, 2021	Hypothesis tests continued, linear regression	Students should have their data cleaned and usable and be ready to run analyses and write up results.
August 12, 2021	Class presentations	
August 16, 2021	Class presentations	

Grading

Homework completion: 65%
Final Project: 35%

Final Project Guidelines: Students will find and choose a dataset from the biomedical life-sciences field. Using what we learn in class, they will identify an original research question to answer based on the data and use R to clean the data, transform it into a usable format and analyze it. Students will present on their results and methodology in a final 10 minute presentation.

Recommended Reading:

1. R for Data Science by Garrett Grolemund & Hadley Wickham (available here)
2. Discovering Statistics Using R by Andy Field, Jeremey Miles and Zoe Field: https://us.sagepub.com/en-us/nam/discovering-statistics-using-r/book236067