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 assistants: Sana Badri, PhD and Stephanie Tuminello

Sana.badri@nyulangone.org
stephanie.tuminello@nyulangone.org

Dates: Mondays and Thursday 10:00-11:30am, July 6 - August 14, 2023

Location:
https://nyulangone.zoom.us/j/93327044883?pwd=dWtCVTlPQktpZCtuZWRkbmFxVnFKQT09

Office Hours: By appointment

Course Site: https://github.com/fredlapolla/RVilcekMaster23

Overview:

Understanding code-based approaches to data analysis provide researchers with 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:

- Upload, clean and transform data to make it suitable for statistical analysis and visualization.

- Utilize R syntax to analyze their data.
- Troubleshoot R challenges to enable them future success in R coding and analysis.
- 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 GitHub Classes 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
Thursday, July 6, 2022	Syllabus review, discussion of final project, discussion of how to identify data sources	Explanation of final project and data finding
Monday, July 10, 2022	Introduction to R and the RStudio Environment, troubleshooting, packages	
Thursday, July 13, 2022	Indexing and Data Types and Structures in R	Exploration of a dataset
Monday, July 17, 2022	Functions and Packages	Students should have identified a data set to work with and explore
Thursday, July 20, 2022	Apply, For Loops and If statements	
Monday, July 24, 2022	Data Cleaning and Dealing with Missing Values	Students should begin to identify a research question of interest to address from their dataset

Thursday, July 27, 2022	Data Visualization for exploratory data analysis	
Monday, July 31, 2022	Data Visualization for publication	Students should be identifying variables to use to answer their research question
Thursday, August 3, 2022	Hypothesis testing and analysis in R	
Monday, August 7, 2022	Hypothesis tests continued, linear regression	Students should have their data cleaned and usable and be ready to run analyses and write up results.
Thursday, August 10, 2022	Class presentations	
Monday, August 14, 2022	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.

Suggested Further Reading:

This class will not use a textbook, and learning will primarily be done by doing coding assignments following each lecture.

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