SIOB 296: Introduction to Programming with R
January 5 - March 11, 2021
Tuesday/Thursday 10:30 – 12:00 PST

Eric Archer
eric.archer@noaa.gov

Course Description

This course will focus on establishing a solid fundamental understanding of data manipulation and analysis with R. No prior programming experience is necessary, and students from all backgrounds are welcome. We will start with an introduction to the R command line, followed by a detailed description of R data structures and how to manipulate them. The course will continue by covering commonly used R functions and students will then learn how to write their own functions and scripts. For the final project, students will write an R script and functions to analyze their own data.

Syllabus

Week	Date	Topic
1	Jan 5	R console, RStudio,
	Jan 7	Data structures, numeric vectors, scripts, numeric indexing
2	Jan 12	Review, character & logical vectors, character & logical indexing
	Jan 14	NULL and NA, coercion, vectorization, sample
3	Jan 19	Matrices, lists, data frames
	Jan 21	Lists, data frames
4	Jan 26	Review, Reading and writing data, file and folder management
	Jan 28	Character and string manipulation
5	Feb 2	Review
	Feb 4	Common functions for data summary and selection
6	Feb 9	Common functions for data summary and selection
	Feb 11	Review
7	Feb 16	Writing functions
	Feb 18	Flow control in functions
8	Feb 23	Review
	Feb 25	Iterating to collect values: lapply, sapply, apply
9	Mar 2	Iterating to collect values: tapply, mapply
	Mar 4	Review
10	Mar 9	Visualizations: ggplot
	Mar 11	tidyverse (magrittr, dplyr, tidyr)

We will be using a Google Drive folder to disseminate notes, homework, and other course materials. The URL for the drive is:

Please download this folder to your local computer. If it downloads as a zip file, unzip it and place it anywhere you choose in the file structure of your system. Do not alter the folder structure of these folders. Throughout the course, we will be adding to these folders and assuming that we can find certain files in the folder structure.

The homework is for your practice purposes. **The homework should not be turned in and will not be graded**. However, I recommend that you take time to try to do the homework as the practice will be important for your understanding of the concepts. We will go over homework on the review day that follows each section.

Zoom

The Zoom link for this course is   Please do not share this link with anyone outside the course. Zoom lectures will be recorded and placed in the folder for the day in the Google Drive.

Software

Students should come on the first day with the latest versions of R and RStudio loaded on their laptops. They can be obtained at:

R: https://www.r-project.org
RStudio: https://www.rstudio.com

Final Project

Students who have officially enrolled in the class will be required to submit a script utilizing a function that they have written. The function should serve some useful purpose for your work by doing some sort of data manipulation, analysis, or visualization using techniques learned in the class. Try to keep the function less than 100 lines long and comment it well. The function should be accompanied by another script that gives an example of its execution. If running the function requires a special data set, send it either as a .csv file or as an R object. If you can't get the function to work, please send what you have at that point along with comments of what is working and what is not working. You will get full credit for the attempt and documentation.

- All files that are part of the project should be sent to me compressed in a .zip file named, "Lastname_F_SIO296.zip", where "F" is the first initial (e.g., "Archer_E_SIO296.zip").
- The function should be in its own .r script file.
- The script that executes an example of the function should be in another .r script file.
- Be sure that the zip file is self-contained, meaning that it contains all necessary data files in its own folder and I will be able to execute it.
- The script should not manipulate files or folders in folders other than its own.
- Please limit the number of external packages it requires and load them at the beginning so I can make sure I have everything necessary installed on my machine.

- If your script needs external data, please include a minimal representation of this data. Subset only what is necessary to demonstrate the function. Do not send me your entire dataset.

The final project is due no later than **March 20, 2021 @ 2359 PDT.**

Suggested Texts and Resources

- Davies, T. 2016. The Book of R: A First Course in Programming and Statistics. No Starch Press. 832pp ISBN 978-1593276515
- Matloff, N. 2011. The Art of R Programming: A Tour of Statistical Software Design. No Starch Press. 400pp ISBN 978-1593273842
- Wickham, H. and Grolemund, G. 2017. R for Data Science. O'Reilly Media. 522pp ISBN 978-1491910399
- R Bloggers: https://www.r-bloggers.com
- Base R Cheat Sheet: https://www.rstudio.com/wp-content/uploads/2016/10/r-cheat-sheet-3.pdf
- R Reference Card: https://cran.r-project.org/doc/contrib/Short-refcard.pdf
- Other Documentation: http://cran.stat.ucla.edu/other-docs.html

Assistance

Sam Woodman (sam.woodman@noaa.gov) has kindly offered his time as a TA for the course. If you have any questions during the lectures, please feel free to shoot a message to Sam. You can also email him if you have questions with homework or other issues that come up. Of course, feel free to email me as well. However, Sam may be able to get back to you during the week faster than I can.