Harmless Data Science Programming Bootcamp for Beginners

January 2021

Instructor:	Dr. Howard Liu	**Time:**	F 14:00 – 18:00 (GMT)
Email:	howard.liu@essex.ac.uk	**Place:**	Zoom

Objectives: This bootcamp is designed to prepare students for ~~their~~ future courses ~~in the data science MSc~~ at Essex when they have no previous programing experience in R. The instructor will teach you how to run R in RStudio, how to work with basic objects and run operations, and also how to create/merge/subset a dataframe. We will also cover some basic data visualization techniques ~~and the ways to report your data analysis result~~. The bootcamp has two components daily. It will start with a programming lecture session and then a practice session where students' questions can be answered interactively and in real-time.

Slack: Using Slack, a online chat platform, to teach programming has been proven to be very successful both in my own experiences and the others. After having your email addresses, I will enroll you onto our Slack workplace where you can ask questions by pasting your scripts as well as error messages. The instructor and a teaching assistant would be able to help out in real-time.

Office Hours: Right after the lecture. The instructor will be available on Zoom as well as on Slack

Course content

Day 1 (Jan. 10th): Introduction and R basics

- Get familiar with RStudio Interface
- Basic Mathematical Operations
- R data types
- Object and variable assignments
- Sequences
- Testing for NA and NaN

Day 2 (Jan. 11th): R Fundamentals

- if-else and else-if statements
- ifelse() function
- for Loop
- while Loop
- Break & next statement
- Creating functions
- the apply functions

Day 3 (Jan. 12th): Vector, Matrix, and List

- Creating Vector
- Vector Operations
- Creating Matrix
- Matrix operations
- Creating Lists
- Subsetting or slicing List

Day 4 (Jan. 13th): Data frame

- Import and export data frame
- Working with data frame
- Subsetting
- merging
- dplyr function
- Pipeline Operator

Day 5 (Jan. 14th): Data Visualization and Beyond

- ggplot function
- Scatter plots
- Histograms
- Visualizing time-series data
- Visualizing cross-sectional data
- Using github to save your code
- Using Rmarkdown to report analysis