**Description**:  This course will teach you how to do data science with R: You’ll learn how to get your data into R, get it into the most useful structure, transform it, and visualize it. You’ll learn how to clean data, draw plots, and how to find help when you’re stuck. These skills are vital for scientific research to happen, and we will go over best practices for doing each of these things with R. You’ll learn how to use the grammar of graphics, literate programming, and reproducible research to save time. Moreover, you’ll learn how to manage cognitive resources to facilitate discoveries when wrangling, visualizing, and exploring data.

**Outline:**

* Lecture 1: Introduction to R
  + What is programming?
  + Why is programming useful for data science?
  + What can you do in R?
  + Setting up RStudio
  + Project oriented workflow with RStudio project files
  + Installing and loading packages
  + RMarkdown
* Lecture 2: Data visualization
  + Overview of the grammar of graphics philosophy and its implementation in ggplot2
  + Bar charts, pie charts, line plots, etc
  + Mapping variables to colors/sizes/shapes
  + Using different coordinate systems
  + Themes and how to modify them
    - Making a chart look like the economist/wsj/fivethirtyeight
  + Facetting plots
* Lecture 3: Data cleaning
  + Readable code
  + Using the magrittr pipe to chain commands
  + Dplyr verbs:
    - Mutate ()
    - Filter()
    - Select()
    - Arrange()
    - Summarise()
    - Groupby()
  + Renaming columns
  + Logical operators and their use in filtering statements
  + Relational data
    - Mutating vs filtering joins
  + Some other useful functions
* Lecture 4: Tidy data
  + A standard way of storing data in a consistent format
    - Each column forms a variable
    - Each row is an observation
    - Each cell is a single measurement
  + Easier for automation and iteration
  + Wide and long representations of data
  + Getting data into tidy format
    - Pivoting
    - Splitting/combining
    - Nesting/unnesting
    - Filling missing values
  + Examples of messy vs tidy data
* Lecture 5: Don’t repeat yourself (DRY)
  + Motivating example:
    - Batch download and process COVID-19 data
    - Identify repeated patterns urls and automate:
      * Reading each csv
      * Standarise column names
  + How to write your own functions and why to do it
  + For loops
  + “Functional” programming for iteration
    - The map family of functions
  + Using dplyr’s across() function to perform the same operations across a selection of columns
* Lecture 6: Factors and dates
  + Factors:
    - What are factors
    - How to create them
    - Improving plots by reordering factor levels
    - Lumping factors with small number of observations into “other”
    - Recoding factor levels
  + Dates
    - Dates vs date-times
    - Creating and parsing them
    - Extracting components of a data/datetime (month, year, weekday, minute)
    - Format date times