

CRASH COURSE R FOR DATA SCIENCE - ALL MATERIALS LIVE HERE!

University of Glasgow - Glasgow International College

GIC PM-617 Module: Practical Training for Science and Engineering Research

This is a repository for materials used for a module on quantitative research resources, taught at Glasgow International College at The University of Glasgow.

WHAT IS R?

R is a very popular programming language for the world of data analysis, data science and statistical computing. The journey of the R language from a rudimentary text editor to the interactive R Studio has engaged many scientists across the world.

This was possible only because of generous contributions by R users globally. The inclusion of powerful packages in R has made it more and more powerful with time. Packages such as `dplyr`, `tidyr`, `readr`, `data.table`, `SparkR`, `ggplot2` have made data manipulation, visualization and computation much faster.

PREPARATION BEFORE THE SEMINAR

To follow the Python tutorial during our seminar, you will need to have a few things installed.

The reason we do this **before** the seminar is because it takes time. And we only have two hours. So I would much rather we spend those hours actually learning Python and its application in Data Science instead of troubleshooting a Python installation gone wrong!

That's why I'm asking you to go through this initial steps **before** the seminar. So we can get started right away!

To teach this course, we will use an interactive package called `swirl`. The `swirl` R package makes it fun and easy to learn R programming and data science.

But first things first.

Get R

In order to run `swirl`, you must have R 3.1.0 or later installed on your computer. If you are on a Linux operating system, please visit [Installing swirl on Linux page](#).

For help installing R, check out one of the following videos (courtesy of Roger Peng at Johns Hopkins Biostatistics):

Installing R on Windows Installing R on Mac

Get R Studio

In addition to R, it's highly recommended that you install RStudio, which will make your experience with R much more enjoyable.

If you need to install RStudio, you can do so [here](#). Select the appropriate installer for your operating system.

Once you have finished installing VS Code, you can continue with these instructions.

Install Swirl (this step can be done during the seminar, if you have seminar hours.)

Open RStudio (or just plain R if you don't have RStudio) and type the following into the console:

```
> install.packages("swirl")
```

Finally, start Swirl.

This is the only step that you will repeat every time you want to run swirl. First, you will load the package using the `library()` function.

Install an interactive course.

The first time you start swirl, you'll be prompted to install a course.

For this workshop, we will be learning the following modules:

- R Programming: The basics of programming in R
- Getting and Cleaning Data
- Exploratory Data Analysis
- Regression Models
- Statistical Inference

During the seminar we will install the courses with the following commands:

```
>library(swirl)
>install_course_github("swirldev", "R_Programming_E")
>install_course("Exploratory Data Analysis")
>install_course("Getting and Cleaning Data")
>install_course("Statistical Inference")
```

Now you are ready for the seminar.

Happy learning!

WHAT TO DO AFTER THE SEMINAR

1. Complete the interactive modules. Do as many as you can
2. E-portfolio Evidence: Just focus on going through the interactive modules and finish as many as you possibly can. Then, write a short 500-1000 words reflection about the course and what you learned. If you finished all exercises, how was it? If you didn't, why? And this, accompanied by screenshots of your local setup with the exercises.

Course Instructor: [Graciela Carrillo](#)

Module Coordinator: [Dr. Molly Huq](#)

Other Resources

[R for Data Science](#)

[R Programming Tutorial - Learn the Basics of Statistical Computing](#)

[R Programming for Beginners | Complete Tutorial | R & RStudio](#)