

# **How To Learn to Code for Data Science in R**

Austin Daigle      John Patrick Flores      Madeline Gillman  
Brian Gural      Lorrie He      Justin Landis      Teresa McGee  
Sarah Mae Parker      Matthew Sutcliffe

5/22/24

## **Table of contents**

# Preface

Welcome to How to Learn to Code!!

We are an organization that hopes to make learning to program approachable, accessible, and effective. We want to improve rigor and reproducibility in science by providing programming resources and experiences to scientists and professionals in all levels of their careers. Our classes are small-group based courses with a teacher:student ratio that allows the students to learn dynamically and independently. During classes, students are able to follow along with the teacher leading the instruction, or work with one of our floating teachers to troubleshoot or to better understand their own code.

This is our curriculum for learning R programming in the context of data analysis. Our curriculum development team has worked tirelessly to develop this new curriculum for the Summer of 2024. We are constantly improving and updating our curricula, so if you're interested in contributing or have suggestions, please visit <https://howtolearntocode.web.unc.edu/> for our most up-to-date contact information. If you have gotten to our Class 7 over Github, or are proficient in Github yourself, feel free to submit an issue or pull request at <https://github.com/How-to-Learn-to-Code/Rclass-DataScience>.

Table 1: Table of Contents

Class Day	Topic	Link
0	Getting Ready for Class	<a href="#">Introduction</a>
1		<a href="#">Introduction</a>
2	1	<a href="#">Introduction</a>
3	1	<a href="#">Data Visualization 1</a>
4	1	<a href="#">Data Vizualization 2</a>
5	1	<a href="#">Data Wrangling 1</a>
6	1	<a href="#">Data Wrangling 2</a>
7	1	<a href="#">Project 1</a>
8	1	<a href="#">Project 2</a>

# 1 Welcome to How to Learn to Code!

This page will walk you through setting up access to UNC's computing cluster and introduce you a bit to R and R Studio so we can hit the ground running in the first class. To ensure you have access to the UNC cluster (and thus able to participate in class), **please review this document in full at least 24 hours in advance of the first class**—Research IT will need time to approve your account request.

## 1.1 Class 0 Objectives

- Request a Longleaf account
- Launch an R Studio session on OnDemand
- Know what each of the four panels in R Studio show

## 1.2 R vs. R Studio

In this class, you'll hear these two terms a lot. They sound similar, but they are actually very different! **R** is the programming language we will be learning in this class. **R Studio** is a user-friendly interface (or **IDE**, integrated development environment) we will be using to write scripts in R and interact with R software.

## 1.3 Longleaf

“Longleaf” is the name for UNC's computing system. Researchers in all departments across UNC use it to run analyses, store data, and use programs that require GPUs. Whenever someone says they are “on Longleaf” or “running code on Longleaf” it means their personal computer is connected to the cluster and they are either actively interacting with a program running on the cluster (we will be doing this with R Studio!) or writing code that tells the cluster to perform certain tasks whenever it has the memory availability.

**Before the first class, you will need to request access to Longleaf.** Follow the instructions on the [Research IT website](#). In addition to your onyen and email address, you'll need the following information:

- Preferred shell: bash
- Faculty sponsor name and onyen: You can put your PI here, or if you do not have a PI, leave blank.
- Type of subscription: Longleaf
- Description of work you will do on the cluster: How to Learn to Code R class

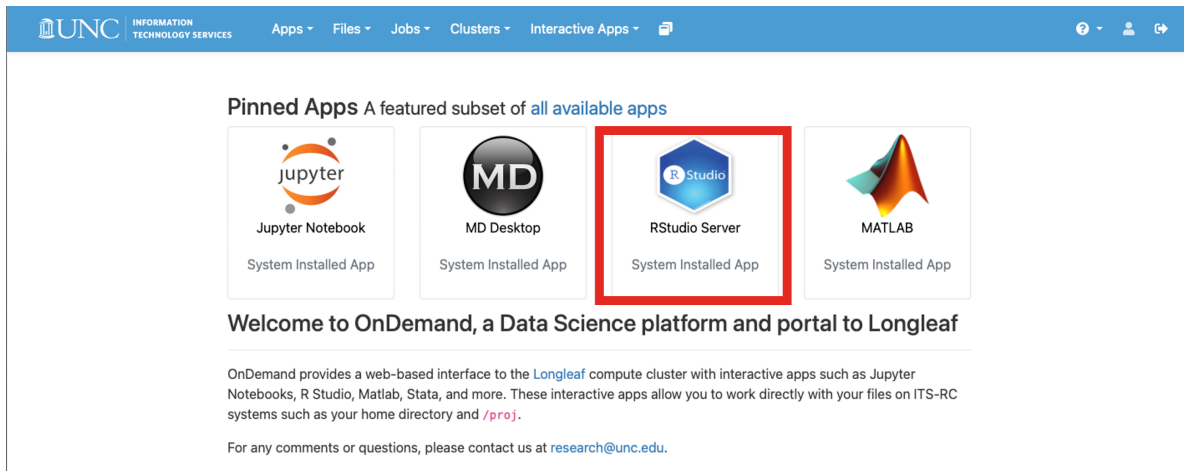
It may take ~24 hours before your account is approved.

## 1.4 OnDemand














OnDemand is a web portal that allows you to access **Longleaf**. We will be using **OnDemand** to launch **R Studio** and run **R** code. You will need to have your Longleaf account approved before accessing OnDemand.

To launch OnDemand, navigate to this site in a browser of your choice: <https://ondemand.rc.unc.edu> (you may want to bookmark this site, you'll be accessing it for each class).

Once you've logged in, you'll see a page like this. Click on the **RStudio Server** tile.



This will take you to a page where you can fill out some parameters for your R Studio Server session. The only one you'll need to adjust is "Number of hours" where you should put "2".

Interactive Apps	
Desktops	
	Longleaf Desktop
	Longleaf Desktop (GPU-FULL)
	Longleaf Desktop (GPU)
	MD Desktop
GUIs	
	3D Slicer
	COMSOL
	FSL
	Firefox
	Freeview
	GView
	GaussView
	ITK-SNAP
	MATLAB

## RStudio Server

This app will launch an RStudio server on a cluster node.

R version

R 4.2.2

Starting from 2/23 this environment is running R 4.1.0 on RHEL8. Packages previously installed and their dependencies may need to be re-installed. Please also ensure your R session is not using a library path containing the previously installed R packages.

Additional Job Submission Arguments

Number of hours

2

Specify **two** hours if this session is for use during a class meeting time.

Number of CPUs

1

Unless you are certain you are running parallel code you should specify **one** cpu.

☐ I would like to receive an email when the session starts

Launch

\* The RStudio Server session data for this session can be accessed under the [data root directory](#).

### Note

You can request up to 10 hours, but it's good practice to only request the amount of time you'll need (Longleaf is a shared resource!). Since each class is 90 minutes, you'll likely only need to request 2 hours for each class. Under "Additional job submission arguments" you adjust the amount of memory requested. This won't be needed for How to Learn to Code classes, but may be needed when you are running your own analyses on large datasets in the future.


After you've filled out the appropriate information, click Launch. This will take you to the "My Interactive Sessions" page. Your session request may be queued for a minute while space on the cluster is being allocated for your session. Once it's ready, click "Connect to R Studio Server". This will launch R Studio in a new tab.


Session was successfully created.


Home / My Interactive Sessions


Interactive Apps

Desktops

 Longleaf Desktop

 Longleaf Desktop (GPU-FULL)

 Longleaf Desktop (GPU)

 MD Desktop

GUIs

RStudio Server (33839599) Queued

Created at: 2024-03-16 12:16:35 EDT

Time Requested: 8 hours

Session ID: 199b0876-f593-4559-b4bd-3cf1312fa1ae

Please be patient as your job currently sits in queue. The wait time depends on the resources requested (e.g. number of CPUs, GPU, time).


Delete


Session was successfully created.


Home / My Interactive Sessions


Interactive Apps

Desktops

 Longleaf Desktop

 Longleaf Desktop (GPU-FULL)

 Longleaf Desktop (GPU)

 MD Desktop

GUIs

RStudio Server (33839599) 1 node | 1 core | Running

Host: >\_c0300ood19.ii.unc.edu

Created at: 2024-03-16 12:16:35 EDT

Time Remaining: 7 hours and 59 minutes

Session ID: 199b0876-f593-4559-b4bd-3cf1312fa1ae

Connect to RStudio Server

Delete

7