

Gentle Intro to R/RStudio

ENT 198 Winter 2019

Michael Culshaw-Maurer and Hanna Kahl

2019-01-03

Course Description

Data management and analysis are crucial aspects of the modern scientific process, but they can seem daunting at first. This course will serve as a gentle introduction to data management, analysis, and visualization using the R programming language and the RStudio environment. The course will consist of live-coding sessions, where instructors write code along with students, and hands-on computing exercises. It will cover a large portion of the data analysis pipeline, from data entry and management to analysis and visualization. This is **not** a statistics course, though the basic framework for conducting statistical analyses in R will be covered. Finally, the last few classes will consist of work sessions, where students can work on their own data with the instructors available to help.

Instructor Contacts

The best way to reach the instructors is via email. Michael can be reached at mjculshawmaurer@ucdavis.edu and Hanna at hkahl@ucdavis.edu. We will use email and [the Canvas site](#) to distribute course information.

Course Prerequisites

No experience with R, RStudio, or data analysis is required. The following programs must be installed by the first class meeting: **the latest versions of R and RStudio, and a spreadsheet management software (Excel or something similar).**

Course Purpose

The goal of this course is not for students to master all aspects of the data science pipeline, but to help students to become more comfortable with the process and more confident in their abilities to learn more.

Students will learn to manage projects and data, read and write their data in R, manipulate and clean up data, and create high-quality visualizations.

Course Objectives

Students completing this course should learn to:

- Properly format and manage raw data
- Create clean and readable project directories
- Import and export data using R
- Clean, summarize, and prepare data for analysis
- Create high-quality visualizations
- Work with the R language
- Seek further knowledge about R and data analysis
- Apply these skills to their own data

Schedule

Week 1

- Spreadsheets and project management
- Project management
- Benefits of using R
- R: Setting working directory and saving scripts

Week 2

- Intro to R and RStudio
- Script writing best practices
- Vectors and data types
- Troubleshooting: understanding errors and warning

Week 3

- Data import/export
- Data frames and subsetting

Week 4

- Summarizing data
- `ggplot2` (short intro and demo)

Week 5

- `dplyr`: Data Summarization and Export
- `ggplot2`: Hands on with Ecology Data

Week 6

- `ggplot2`: Finishing Up
- Data Import/Export Workflow

Week 7

- `ggplot2` (visualization)

Week 8

- Work on your own data!

Week 9

- Work on your own data!

Week 10

- Work on your own data!

Computer Setup

Welcome! There's a few things we'd like everyone to try and get sorted out *before* our first class. Please plan on bringing your laptop (and a charger if your battery isn't going to hold for 2 hrs) to each class.

Please complete these 3 main tasks before our first class:

1. Install *current* **R** on your laptop
2. Install *current* **RStudio** on your laptop
3. Make sure you have a spreadsheet program installed on your computer (Microsoft Office, Libre Office, etc.)

Read below for more detailed information on installing the pieces of software we'll need, and platform specific instructions.

R & RStudio

R and **RStudio** are separate downloads and installations. **R** is the underlying statistical computing environment. **RStudio** is a graphical integrated development environment (IDE) that makes using R much easier and more interactive. You need to install R before you install **RStudio**. Download and install both of these but in this order:

1. **R**: Get the most current version appropriate for your machine. It's free.
2. **RStudio** is a great platform to work with R (note you need R before you can use RStudio). Please install the most recent version. It's free. It does lots of cool things. We'll talk more about it in class.

Windows

If you already have R and RStudio installed

- Open RStudio, and click on *Help > Check for updates*. If a new version is available, quit RStudio, and download the latest version for RStudio.
- To check which version of R you are using, start RStudio and the first thing that appears in the console indicates the version of R you are running. Alternatively, you can type `sessionInfo()`, which will also display which version of R you are running. Go on the [CRAN website](#) and check whether a more recent version is available. If so, please download and install it. You can [check here](#) for more information on how to remove old versions from your system if you wish to do so.

If you don't have R and RStudio installed

- Download R from the [CRAN website](#).
- Run the .exe file that was just downloaded
- Go to the [RStudio download page](#)
- Under *Installers* select **RStudio x.yy.zzz - Windows XP/Vista/7/8** (where x, y, and z represent version numbers)
- Double click the file to install it
- Once it's installed, open RStudio to make sure it works and you don't get any error messages.

macOS

If you already have R and RStudio installed

- Open RStudio, and click on "Help" > "Check for updates". If a new version is available, quit RStudio, and download the latest version for RStudio.
- To check the version of R you are using, start RStudio and the first thing that appears on the terminal indicates the version of R you are running. Alternatively, you can type `sessionInfo()`, which will also display which version of R you are running. Go on the [CRAN website](#) and check whether a more recent version is available. If so, please download and install it.

If you don't have R and RStudio installed

- Download R from the [CRAN website](#).
- Select the .pkg file for the latest R version
- Double click on the downloaded file to install R
- It is also a good idea to install [XQuartz](#) (needed by some packages)
- Go to the [RStudio download page](#)
- Under *Installers* select **RStudio x.yy.zzz - Mac OS X 10.6+ (64-bit)** (where x, y, and z represent version numbers)
- Double click the file to install RStudio
- Once it's installed, open RStudio to make sure it works and you don't get any error messages.

Linux

- Follow the instructions for your distribution from [CRAN](#), they provide information to get the most recent version of R for common distributions. For most distributions, you could use your package

manager (e.g., for Debian/Ubuntu run `sudo apt-get install r-base`, and for Fedora `sudo yum install R`), but we don't recommend this approach as the versions provided by this are usually out of date. In any case, make sure you have at least R 3.4.3

- Go to the [RStudio download page](#)
- Under *Installers* select the version that matches your distribution, and install it with your preferred method (e.g., with Debian/Ubuntu `sudo dpkg -i rstudio-x.yy.zzz-amd64.deb` at the terminal).
- Once it's installed, open RStudio to make sure it works and you don't get any error messages.

Spreadsheet program

To interact with spreadsheets, we can use LibreOffice, Microsoft Excel, Gnumeric, OpenOffice.org, or other programs. Commands may differ a bit between programs, but general ideas for thinking about spreadsheets are the same.

If you don't have a spreadsheet program already, you can use LibreOffice. It's a free, open source spreadsheet program. Alternatively, as a UC Davis student, you are eligible for a free copy of Microsoft Office 365. See [here for instructions](#), make sure to login with your UC Davis email address.

Windows

- Download the Installer
 - Install LibreOffice by going to [the installation page](#). The version for Windows should automatically be selected. Click Download Version X.X.X (whichever is the most recent version). You will go to a page that asks about a donation, but you don't need to make one. Your download should begin automatically.
- Install LibreOffice
- Once the installer is downloaded, double click on it and LibreOffice should install.

Mac OS X

- Download the Installer
 - Install LibreOffice by going to [the installation page](#). The version for Mac should automatically be selected. Click Download Version X.X.X (whichever is the most recent version). You will go to a page that asks about a donation, but you don't need to make one. Your download should begin automatically.

- Install LibreOffice
- Once the installer is downloaded, double click on it and LibreOffice should install.

Linux

- Download the Installer
 - Install LibreOffice by going to [the installation page](#). The version for Linux should automatically be selected. Click Download Version X.X.X (whichever is the most recent version). You will go to a page that asks about a donation, but you don't need to make one. Your download should begin automatically.
- Install LibreOffice
- Once the installer is downloaded, double click on it and LibreOffice should install.

These setup instructions are adapted from those written for [Data Carpentry: R for Data Analysis and Visualization of Ecological Data](#) and [Data Carpentry: Data Organization in Spreadsheets](#), and the [R-DAVIS graduate course](#).