R for Environmental Chemists

David Hall, Steven Kutarna, Kristen Yeh, Hui Peng and Jessica D'eon

2021-06-04

Contents

Howdy

Howdy,

This website is more-or-less the living results of a collaborative project between the four of us. Our ultimate goal is not to be an exhaustive resource for all environmental chemist. Rather, we're focused on developing broadly applicable data science course content (tutorials and recipes) based in `R` for undergraduate environmental chemistry courses. Note that none of this has been reviewed yet and is not implemented in any capacity in any curriculum.

Authors

If you have any questions/comments/suggestions/concerns please email:

- Dave at davidross.hall@mail.utoronto.ca
- Steven at steven.kutarna@mail.utoronto.ca
- Dr. D'eon at jessica.deon@utoronto.ca

Chapter 1

Introduction

- What you'll learn (and won't learn)

    - learn = basic understanding of R
    - won't learn = *L33T hax0r Skillz*

- How book is organized

    - code is covered in chunks or monospaced (`like this`)

- Prerequisite software
- getting help

1.1 Prerequisite software

In order to use this book you will first need to download and install **R** on your computer. The latest build as of March 2021 is v4.0.5. Download the appropriate version of R for your Operating System here.

We recommend using the default settings for installation (i.e. just keep clicking "Next"). *Note that you will not need shortcuts for launching R, as you will always be using Rstudio to run provided code.*

Once R is installed on your computer, you will need to download and install RStudio. Download the appropriate version of RStudio for your Operating System here

Again, just use the default installation settings.

7

1.2   Where to get help

While it's often tempting to contact your TA/Prof at the sign of first trouble, it's often better to try and resolve your issues on your own, especially if they're related to technical issues in R. Given the popularity of R, if you've run into an issue, someone else has... and they've complained about it and someone else has solved it! An oft unappreciated aspect of coding/data science is knowing how to get help, how to search for it, and how to translate someone's solutions to your unique situation.

Places to get help include:

- Stack overflow
- Using built-in documentation (`?help`)
- reference book such as the invaluable (*R for Data Science*)[https://r4ds. had.co.nz/index.html], which inspired this entire project.
- All else fails,holler at your TA/profs.

Getting Setup in R

Chapter 2

Running R Code

- Where to run code in RStudio (script vs. console)
- Basic coding building blocks
- Variable assignment (howdy <- "Howdy world")
- Importing packages

- Calling functions
- R Studio basics

    - Environment windows to inspect workspace data/variables (in conjunction with above); this needs to be explicitly shown as kids never realize they can inspect their data à la Excel in RStudio
    - Plot window
    - Files window (expanded on in R Workflow chapter)
    - Tearable tabs; again, kids don't realize you can tear off a tab into a new window so you can see your data and code side-by-side (click and drag tab to tear off).

    - Customization (i.e. themes for dark-mode)

- Brief style guide

    - Packages loaded at the top.
    - How to comment code (#I'm cold and there are wolves after me) -Suggestion for variable names (good_name vs. thisNameWillWork_butIsAwfulToType); including forbidden names -Script headers (i.e. # heading1, ## subheading); These are picked-up by RStudio and displayed in the Document Outline box allowing easy navigation of long scripts.

Chapter 3

R Workflows

Just like there's a common workflow in any chemistry lab (pre-lab, collect reagents, conduct experiment, etc.) there's a workflow when working with R. This is by no means the only way to work, but it's tried and true and will serve you well as you tackle your coursework.

3.1 Creating an RStudio Project

When you open RStudio for the first time, this is what you should see:

The version of R you just installed should appear in the main window here:

(If this message does not appear, go to *Tools->Global Options* and make sure that the "R version" box is set to the correct folder.)

Before doing anything else, let's create an **R Project**. This will establish a default folder for RStudio and bundle together all your code files in one place. Go to *File->New Project.* Click *New Directory*, then *New Project.* Next, you'll be asked to choose a subdirectory name and location. The name can be whatever you want, but we highly recommend that you create the directory in the same parent folder as your data. Click *Create Project*, and you should now see your chosen file path displayed in the bottom-right window:

You are now ready to use R! If you're in a hurry, you can skip straight to Chapter 2: Organizing Your Data, but we recommend you finish this chapter to familiarize yourself with RStudio's layout.
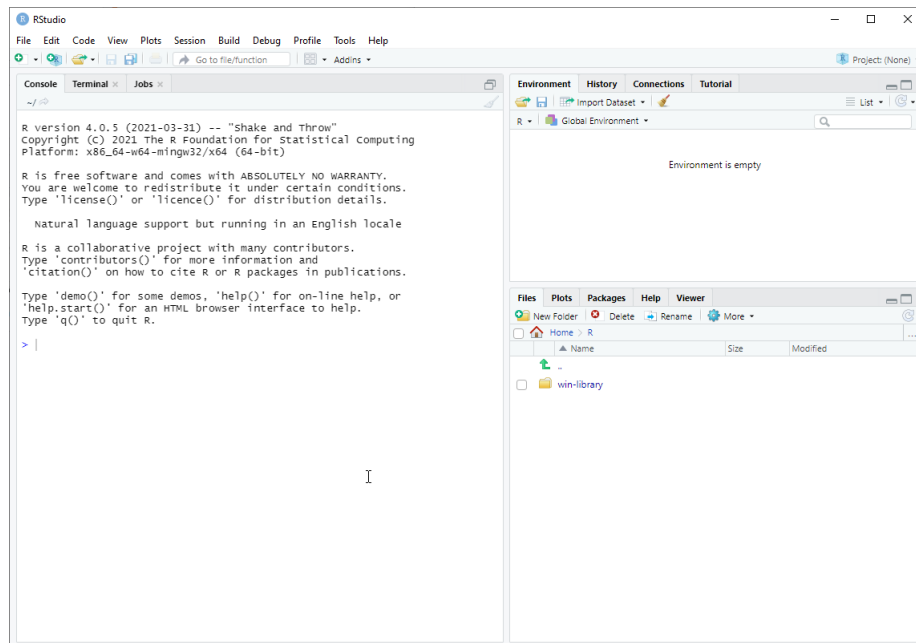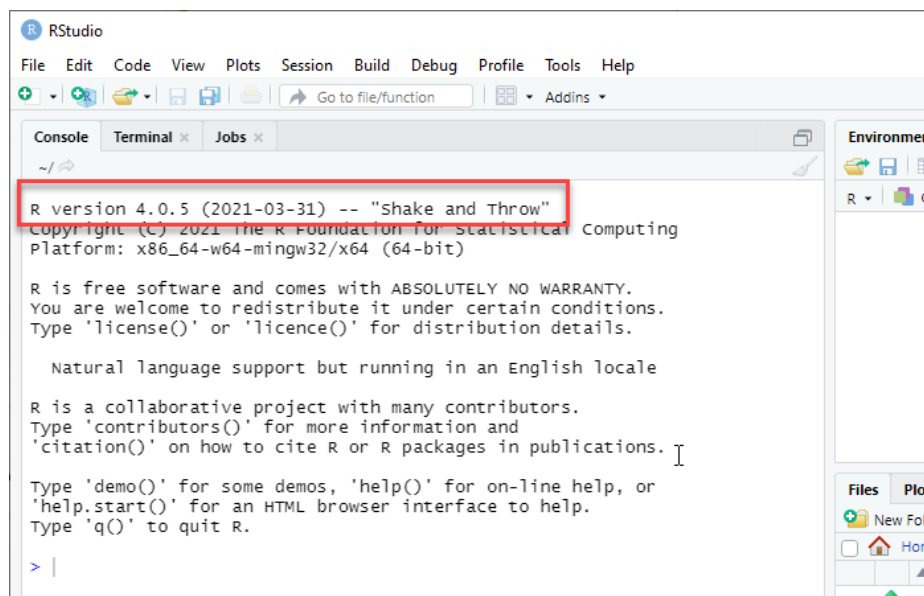
Figure 3.1: RStudio Default View
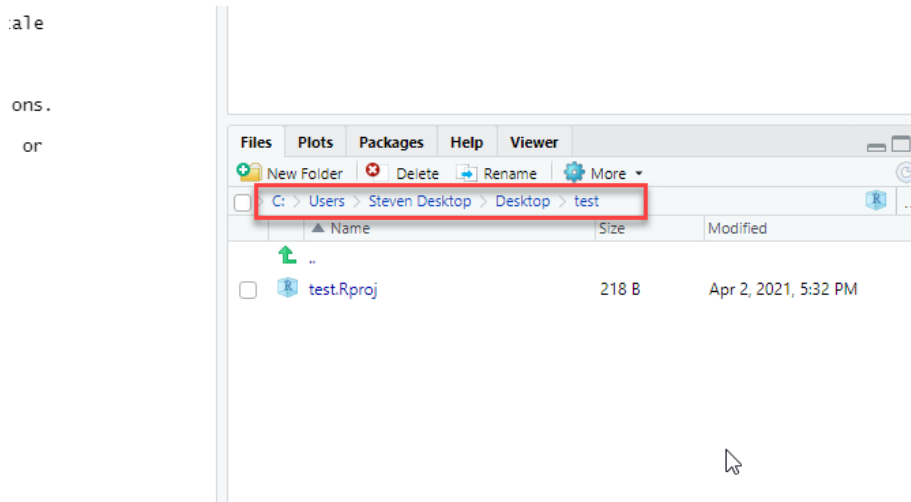

Figure 3.2: RStudio - R Version

Figure 3.3: RStudio Project Folder

3.2   Navigating RStudio

RStudio has 4 main windows, 3 of which should currently be visible. To see the fourth window, go to *File->New File->R Script*. You should now be able to see all 4 main windows:

3.2.1   R Environment

This window is the least important for our purposes. In brief, it will list all variables, packages, and functions which you have run since opening RStudio.

3.2.2   Viewer

The Viewer window has a couple of useful tabs. We will mainly use it to export plots (more on that in later chapters), but you can also open code files from the "Files" tab without having to leave RStudio.

3.2.3   Console

This is where you can type and run code directly, but you will mainly be using the Scripts window to run code (see below). Your primary use of the console will be for installing packages (more on that later). Another thing that the Console window is useful for is quick arithmetic. Try typing "2+2" in the console, then hit Enter.
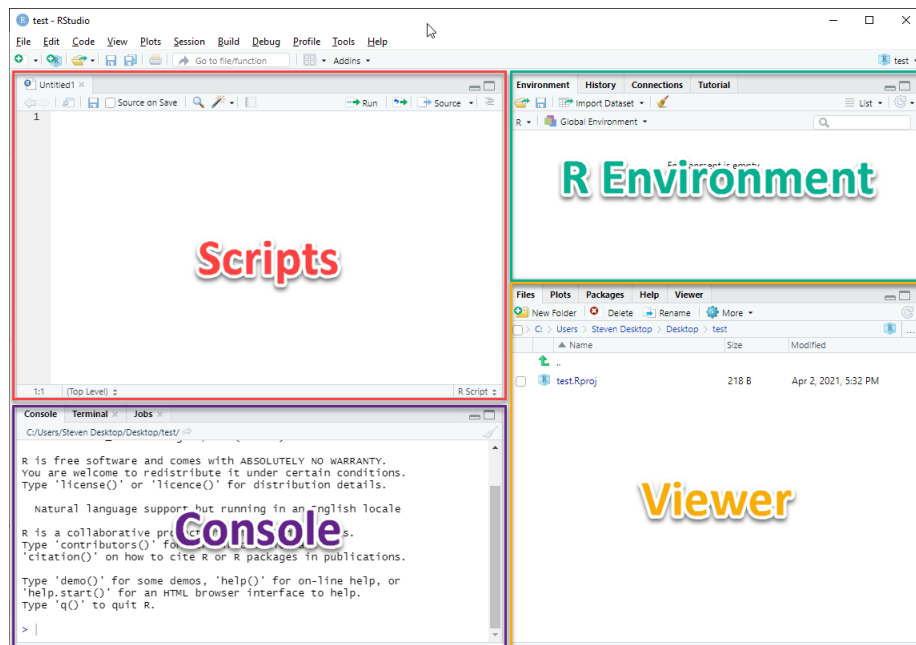
Figure 3.4: RStudio Quadrants

```
2+2
```

```
## [1] 4
```

3.2.4   Scripts

R code files are called scripts, and are saved as *.R files. Whenever you copy code blocks from this website, paste them into the Scripts window. You can then run specific lines by highlighting them and pressing Ctrl+Enter (Cmd+Enter on Mac), or clicking the "Run" button in the top right of this window.

3.3   Dark Mode

Lastly, some of you may be interested in how to set RStudio to Dark Mode. Simply go to *Tools->Global Options*, then click "Appearance" on the left. Then select your preferred Editor Theme from the list. (My personal preference is "Tomorrow Night Bright")
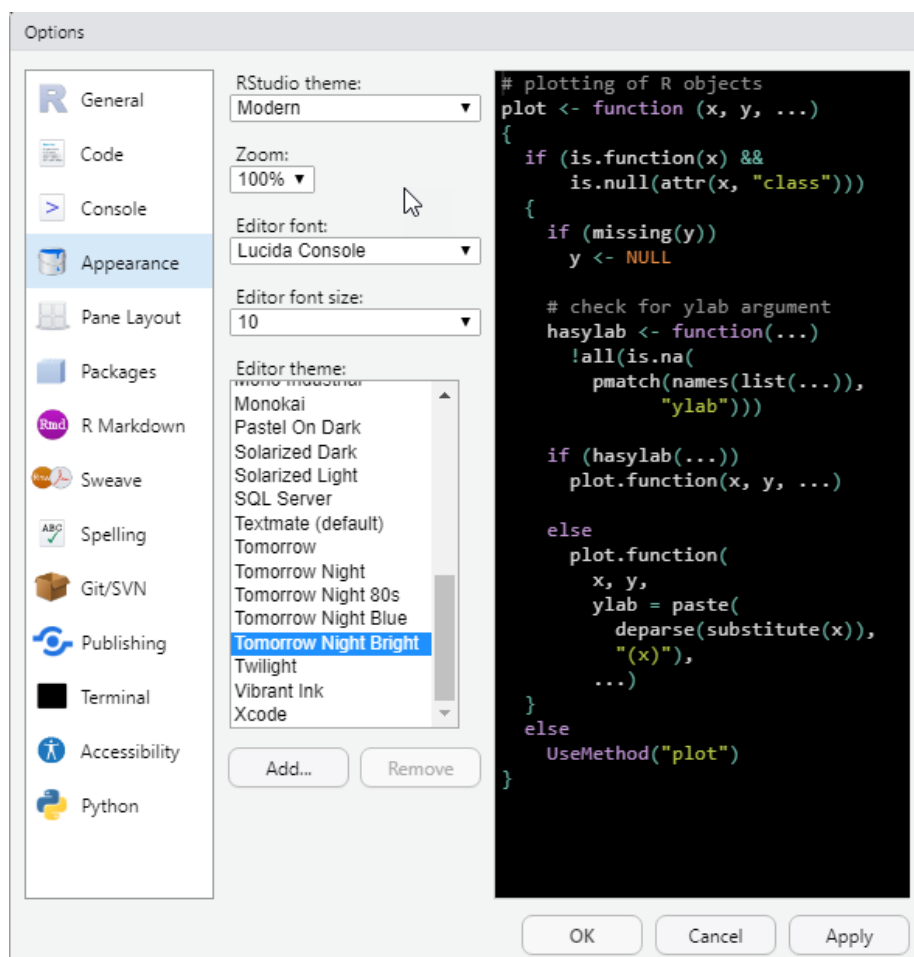
Figure 3.5: RStudio Dark Themes

Data Analysis in R

Chapter 4

Intro to Data Analysis

- introducing the import -> tidy -> transform <-> visualize <-> model <-> communicate workflow
- map where they'll learn these in the upcoming sections
- emphasis this is how they should tackle every problem from now on.