R packages

Now that we’ve installed R and RStudio and have a basic understanding of how they work together, we can get at what makes R so special: packages.

What is an R package?

So far, anything we’ve played around with in R uses the “base” R system. Base R, or everything included in R when you download it, has rather basic functionality for statistics and plotting but it can sometimes be limiting. To expand upon R’s basic functionality, people have developed **packages.** A package is a collection of functions, data, and code conveniently provided in a nice, complete format for you. At the time of writing, there are just over 14,300 packages available to download - each with their own specialized functions and code, all for some different purpose. For a really in depth look at R Packages (what they are, how to develop them), check out Hadley Wickham’s book from O’Reilly, [“R Packages.”](http://r-pkgs.had.co.nz/)

*Side note:* A package is not to be confused with a **library** (these two terms are often conflated in colloquial speech about R). A library is the place where the package is located on your computer. To think of an analogy, a library is, well, a library… and a package is a book within the library. The library is where the books/packages are located.

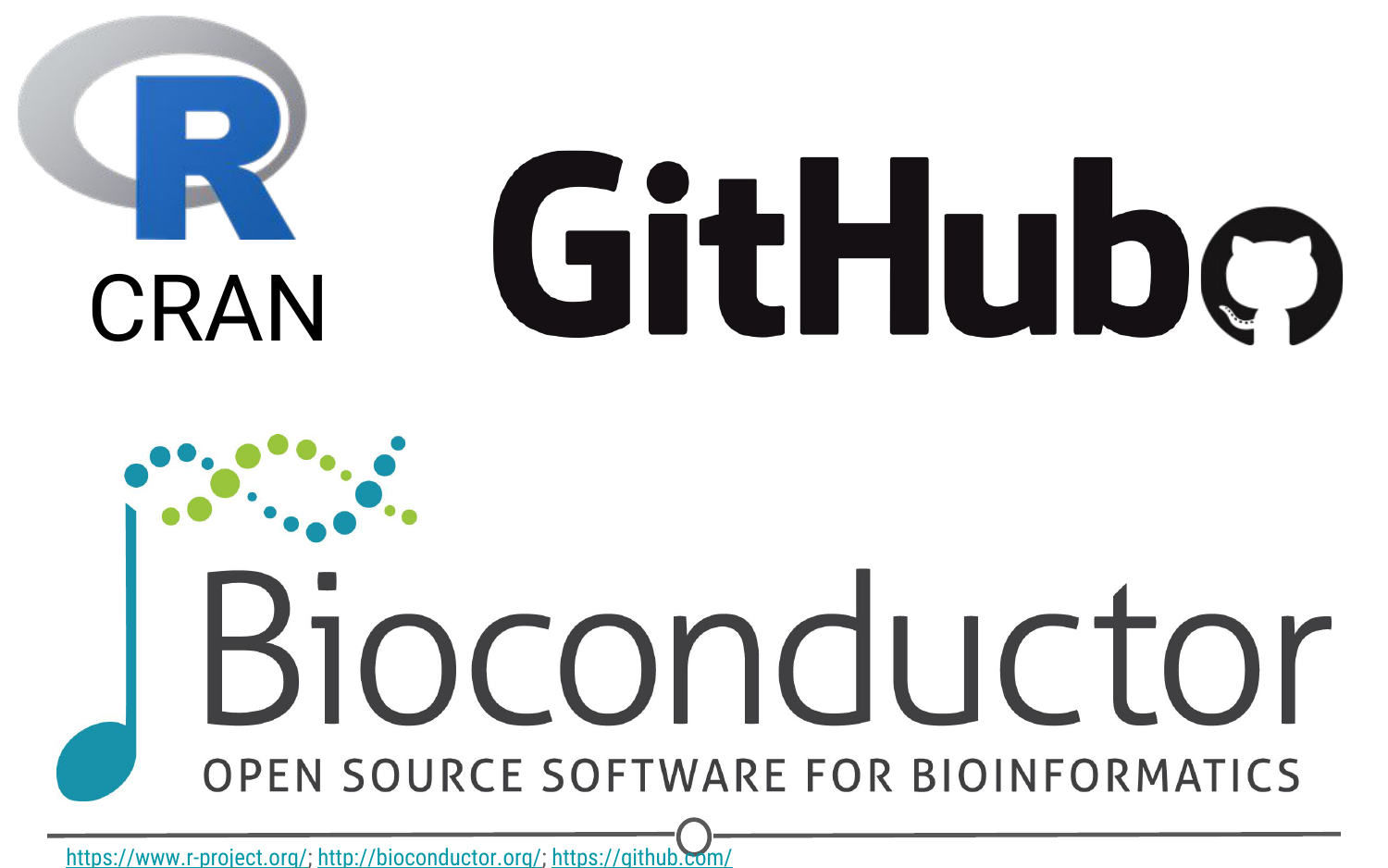
Packages are what make R so unique. Not only does base R have some great functionality but these packages greatly expand its functionality. And perhaps most special of all, each package is developed and published by the R community at large and deposited in **repositories.**

What are repositories?

A repository is a central location where many developed packages are located and available for download.

There are three big repositories:  
1. [**CRAN (Comprehensive R Archive Network):**](https://cran.r-project.org/web/packages/) R’s main repository (>12,100 packages available!)  
2. [**BioConductor:**](https://bioconductor.org/packages/release/BiocViews.html#___Software) A repository mainly for bioinformatic-focused packages  
3. [**GitHub:**](https://github.com/collections) A very popular, open source repository (not R specific!)

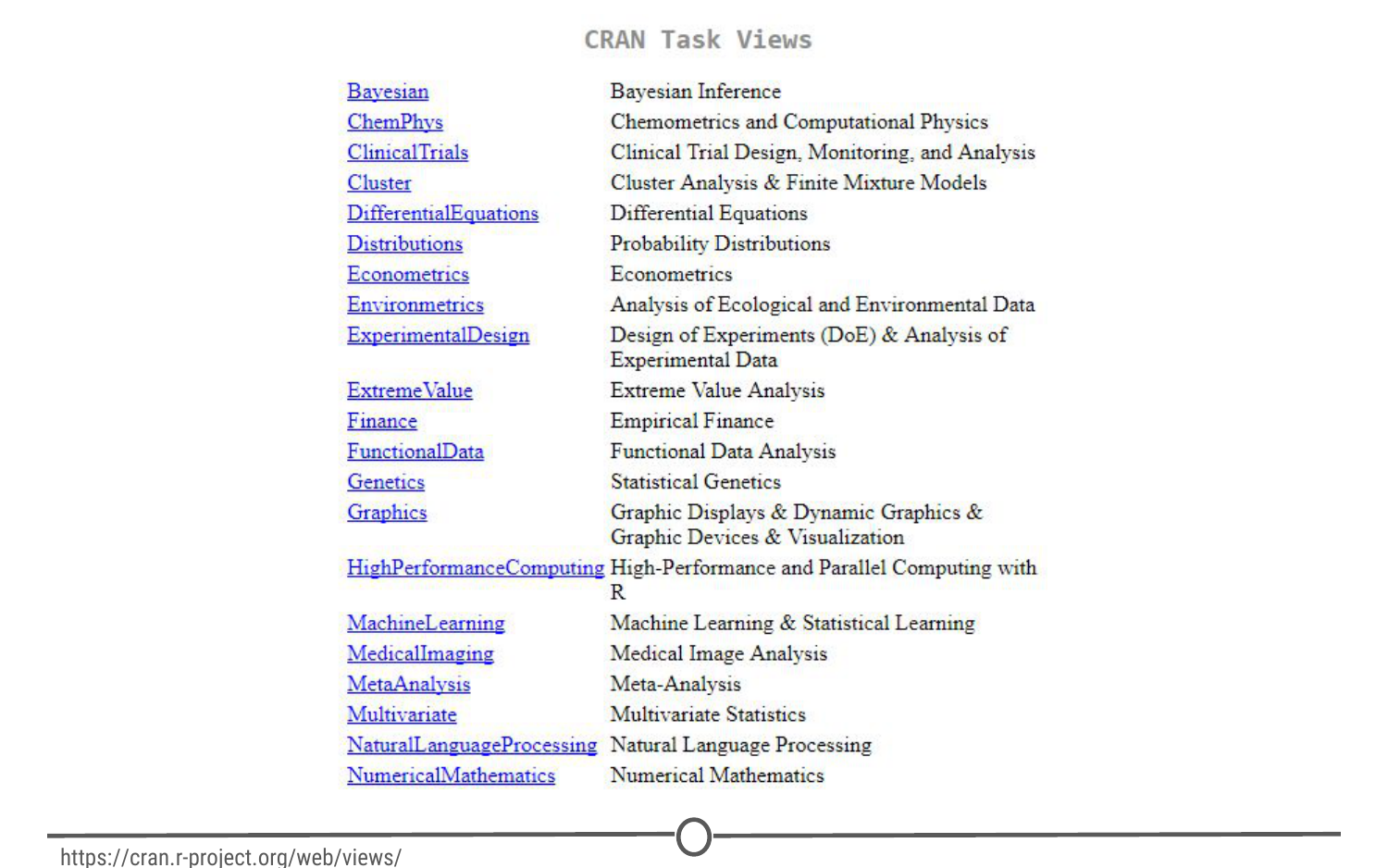
Take a second to explore the links above and check out the various packages that are out there!

**The big three repositories for R packages**

How do you know what package is right for you?

So you know where to find packages… but there are so many of them, how can you find a package that will do what you are trying to do in R? There are a few different avenues for exploring packages.

First, CRAN groups all of its packages by their functionality/topic into 35 “themes.” It calls this its [“Task view.”](https://cran.r-project.org/web/views/) This at least allows you to narrow the packages you can look through to a topic relevant to your interests.

**CRAN’s “Task View” that groups packages into 35 topics**

Second, there is a great website, [**RDocumentation,**](https://www.rdocumentation.org/) which is a search engine for packages and functions from CRAN, BioConductor, and GitHub (ie: the big three repositories). If you have a task in mind, this is a great way to search for specific packages to help you accomplish that task! It also has a [“task” view](https://www.rdocumentation.org/taskviews) like CRAN, that allows you to browse themes.

More often, if you have a specific task in mind, Googling that task followed by “R package” is a great place to start! From there, looking at tutorials, vignettes, and forums for people already doing what you want to do is a great way to find relevant packages.

How do you install packages?

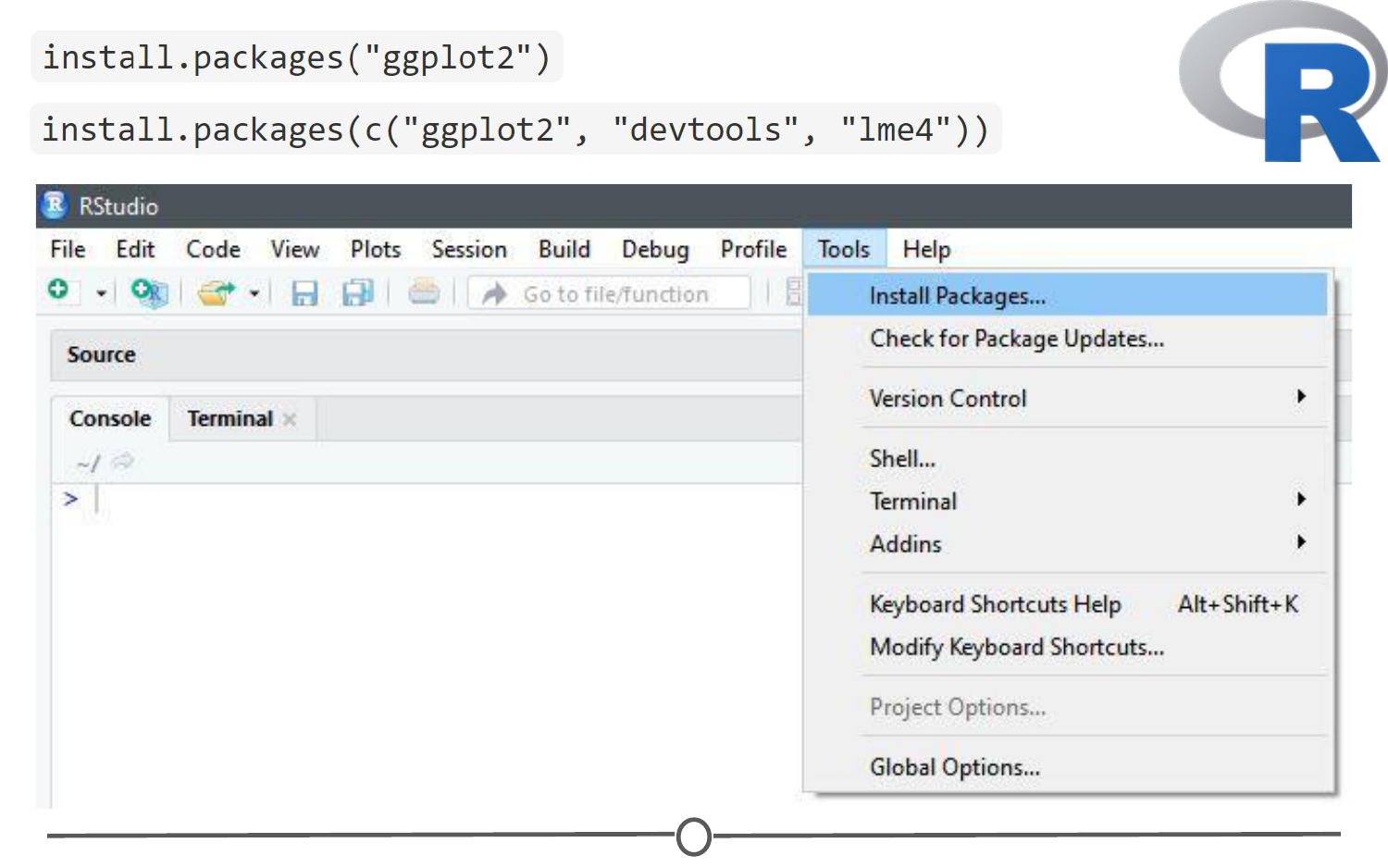
Great! You’ve found a package you want… How do you install it?

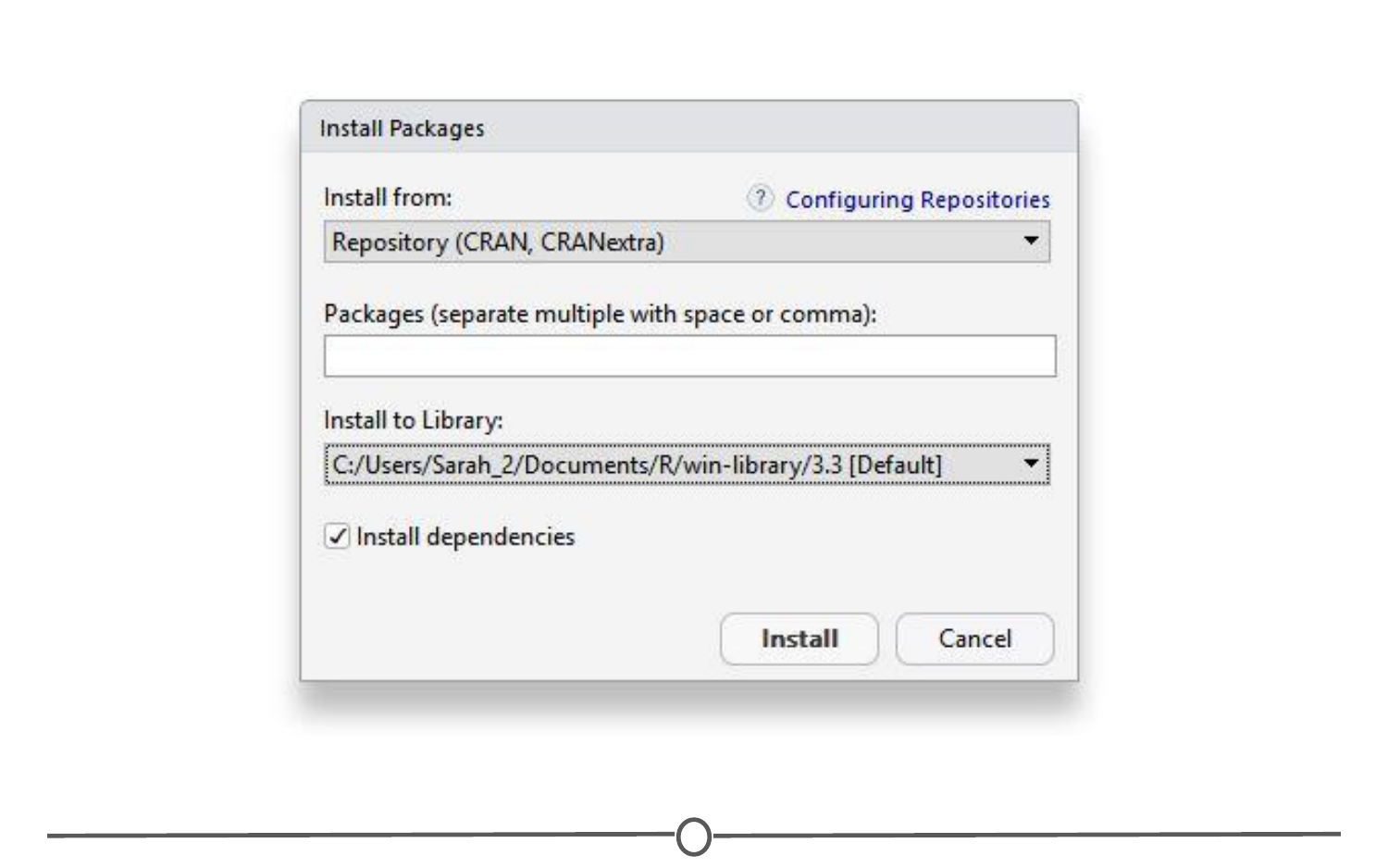
**Installing from CRAN**  
If you are installing from the CRAN repository, use the install.packages() function, with the name of the package you want to install in quotes between the parentheses (note: you can use either single or double quotes). For example, if you want to install the package “ggplot2”, you would use: install.packages("ggplot2")

Try doing so in your R console! This command downloads the “ggplot2” package from CRAN and installs it onto your computer.

If you want to install multiple packages at once, you can do so by using a character vector, like: install.packages(c("ggplot2", "devtools", "lme4"))

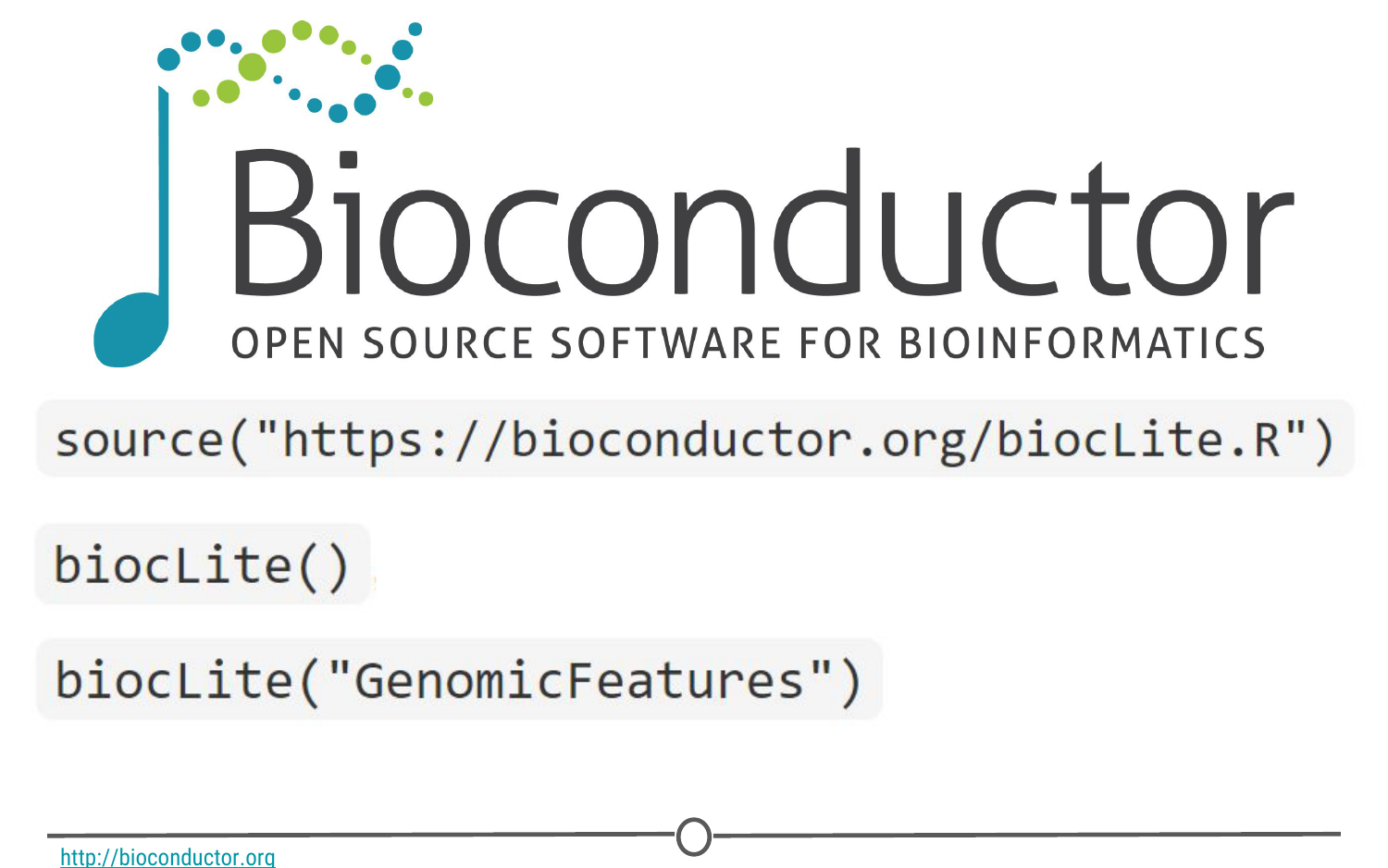
If you want to use RStudio’s graphical interface to install packages, go to the Tools menu, and the first option should be “Install packages…” If installing from CRAN, select it as the repository and type the desired packages in the appropriate box.

**Various methods to install packages within R/RStudio**

**Installing packages from CRAN through R/RStudio**

**Installing from Bioconductor**  
The BioConductor repository uses their own method to [install packages](https://www.bioconductor.org/install/). First, to get the basic functions required to install through BioConductor, use: source("https://bioconductor.org/biocLite.R")

This makes the main install function of BioConductor, biocLite(), available to you. Following this, you call the package you want to install in quotes, between the parentheses of the biocLite command, like so: biocLite("GenomicFeatures")

**Installing packages with BioConductor**

**Installing from GitHub**  
This is a more specific case that you probably won’t run into too often. In the event you want to do this, you first must find the package you want on GitHub and take note of both the package name AND the author of the package. Check out [this guide](http://kbroman.org/pkg_primer/pages/github.html) for installing from GitHub, but the general workflow is:

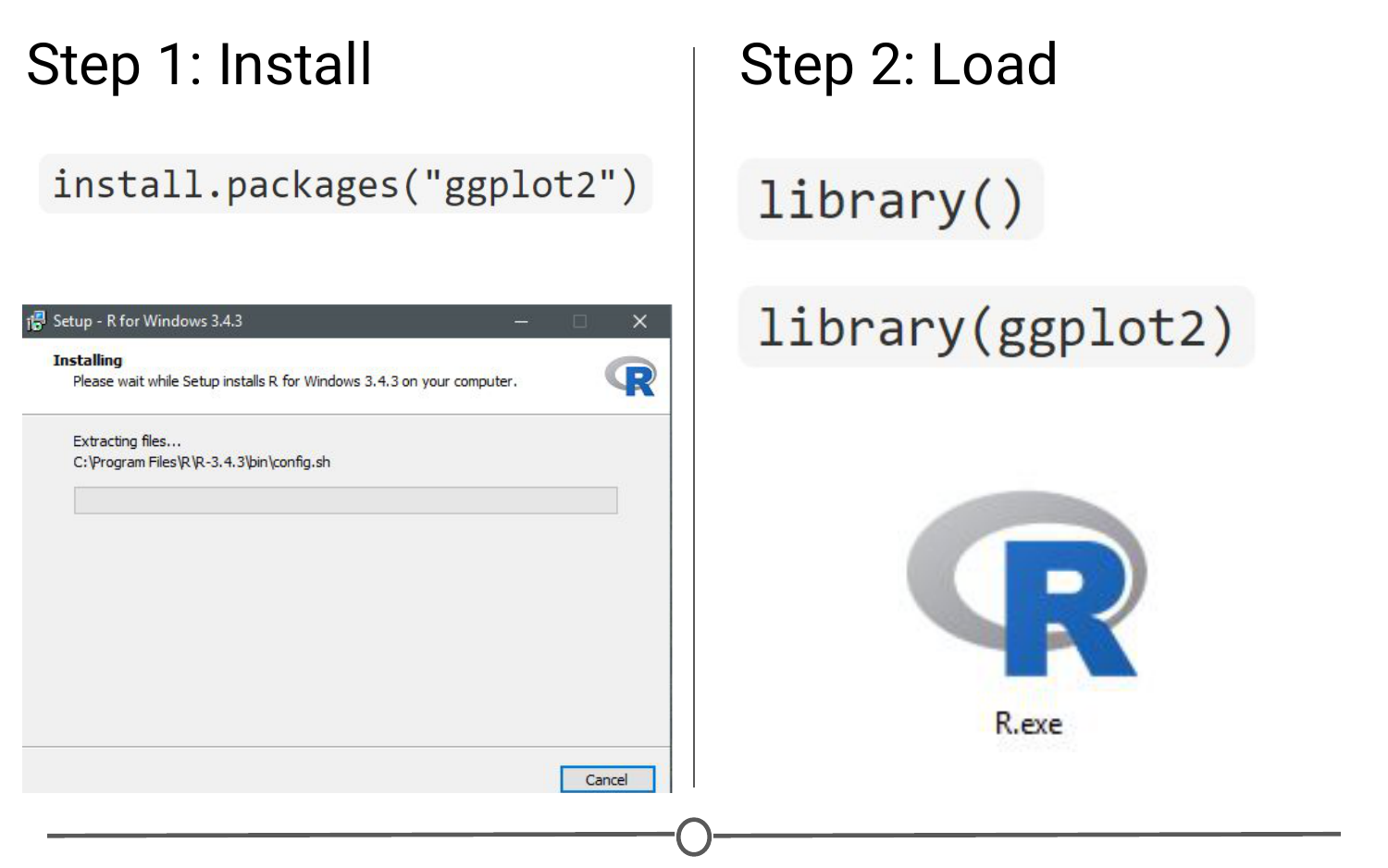
1. install.packages("devtools") - only run this if you don’t already have devtools installed. If you’ve been following along with this lesson, you may have installed it when we were practicing installations using the R console
2. library(devtools) - more on what this command is doing immediately below this
3. install\_github("author/package") replacing “author” and “package” with their GitHub username and the name of the package.

**Installing packages from GitHub**

Loading packages

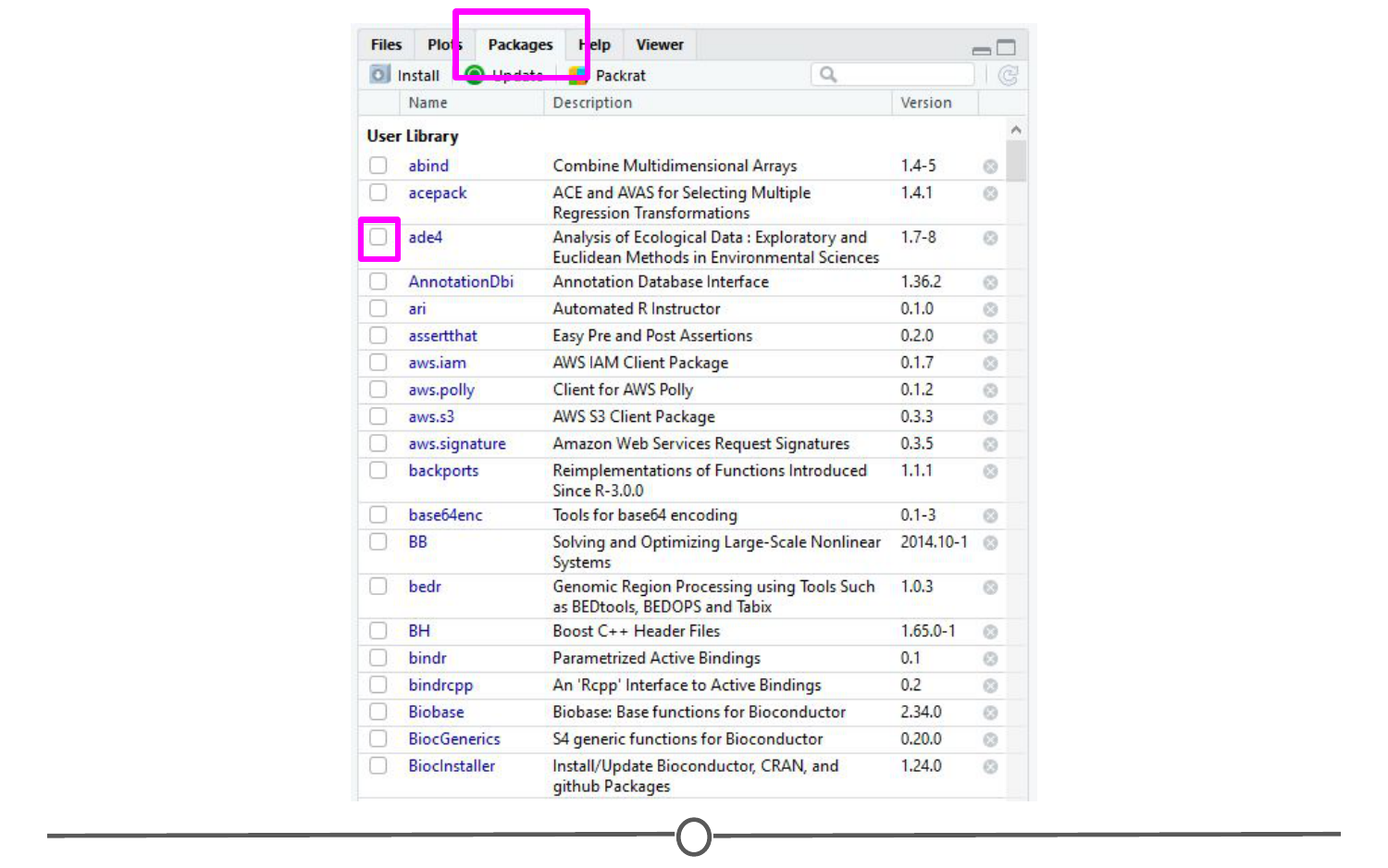
Installing a package does not make its functions immediately available to you. First you must **load** the package into R; to do so, use the library() function. Think of this like any other software you install on your computer. Just because you’ve *installed* a program, doesn’t mean it’s automatically running - you have to open the program. Same with R. You’ve installed it, but now you have to “open” it. For example, to “open” the “ggplot2” package, you would run:library(ggplot2)

**NOTE:** Do **not** put the package name in quotes! Unlike when you are installing the packages, the library() command does not accept package names in quotes!

**Step one of getting a package is installing it, but to use it, you must load it using library(); similar to installing R and then loading it by opening the .exe file**

There is an order to loading packages - some packages require other packages to be loaded first (**dependencies**). That package’s manual/help pages will help you out in finding that order, if they are picky.

If you want to load a package using the RStudio interface, in the lower right quadrant there is a tab called “Packages” that lists out all of the packages and a brief description, as well as the version number, of all of the packages you have installed. To load a package just click on the checkbox beside the package name

**Using the RStudio interface to load a package**

Updating, removing, unloading packages

Once you’ve got a package, there are a few things you might need to know how to do:

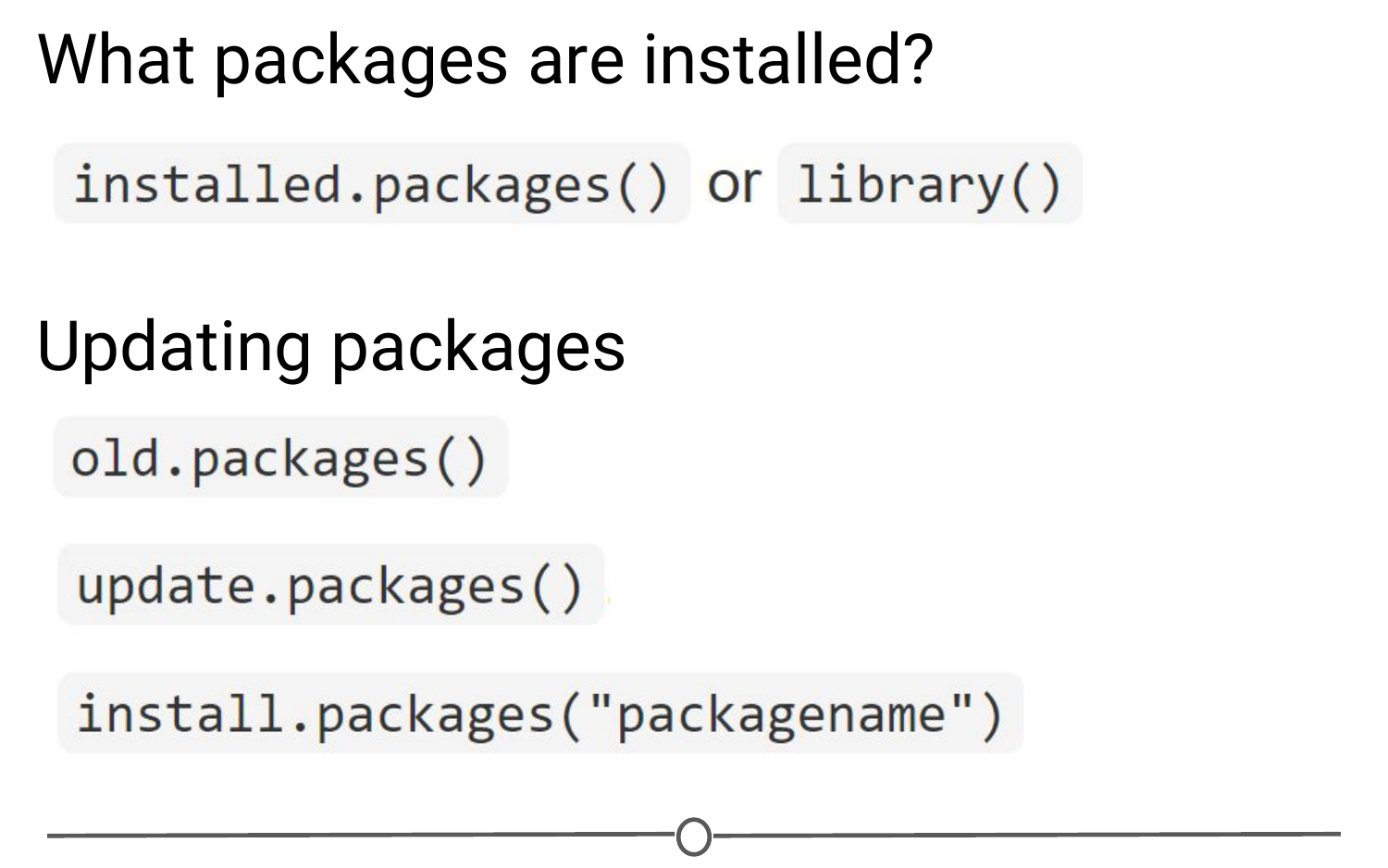
**Checking what packages you have installed**  
If you aren’t sure if you’ve already installed a package, or want to check what packages are installed, you can use either of: installed.packages() or library() with nothing between the parentheses to check!

In RStudio, that package tab introduced earlier is another way to look at all of the packages you have installed.

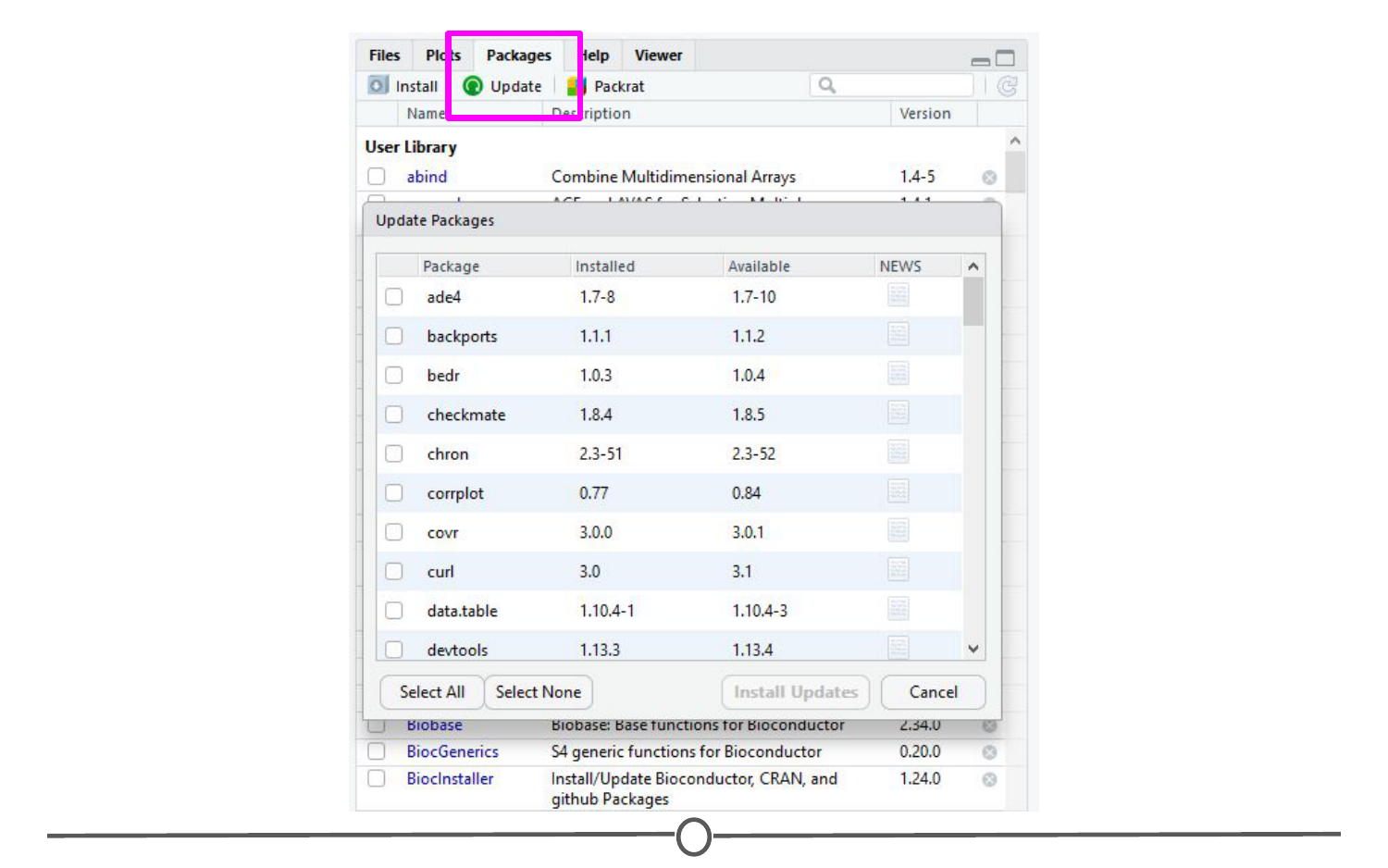
**Updating packages**

You can check what packages need an update with a call to the function old.packages() This will identify all packages that have been updated since you installed them/last updated them.

To update all packages, use update.packages(). If you only want to update a specific package, just use once again install.packages("packagename")

**Functions used to see what packages are installed and update them**

Within the RStudio interface, still in that Packages tab, you can click “Update,” which will list all of the packages that are not up to date. It gives you the option to update all of your packages, or allows you to select specific packages.

**Using the RStudio interface to update your packages**

You will want to periodically check in on your packages and check if you’ve fallen out of date - be careful though! Sometimes an update can change the functionality of certain functions, so if you re-run some old code, the command may be changed or perhaps even outright gone and you will need to update your code too!

**Unloading packages**

Sometimes you want to unload a package in the middle of a script - the package you have loaded may not play nicely with another package you want to use.

To unload a given package you can use the detach() function. For example, detach("package:ggplot2", unload=TRUE) would unload the ggplot2 package (that we loaded earlier). Within the RStudio interface, in the Packages tab, you can simply unload a package by unchecking the box beside the package name.