

Introduction to Genomic Data Visualization in R



Ambu Vijayan

Bioinformatician

BioLit, Thiruvananthapuram

Course Purpose

The purpose of this course is to provide students with a comprehensive introduction to the field of genomic data visualization using the R programming language.

Genomics is a rapidly evolving and data-rich field, and the ability to effectively visualize and interpret genomic data is crucial for researchers and professionals working in genomics and bioinformatics.

This course is designed to bridge the gap between genomics and data visualization, equipping students with the skills and knowledge necessary to extract meaningful insights from biological data through the creation and interpretation of various types of plots and graphs.

By the end of this course, students will be able to:

- Develop a strong foundation in R programming
- Create and customize a wide range of **data visualizations**
- Proficiently use **Multidimensional Scaling (MDS)** for visualizing high-dimensional genomic data
- Generate and interpret **volcano plots** and **MA plots**
- Prepare and visualize **Gene Ontology (GO)** data
- Create customized **heat maps** in R

Introduction to R for Bioinformatics

This course will emphasize the practical application of R rather than delving into its theoretical underpinnings. Nevertheless, a solid grasp of programming fundamentals is essential for gaining a deeper insight into R.

R

R is a language and environment for statistical computing and graphics developed in the early 1990s. It provides a wide variety of statistical and graphical techniques, including linear and nonlinear modeling, statistical tests, time series analysis, classification, and clustering, among many others.

R:

- Is free!
- Is open source and highly extensible, meaning that the user community can (and does) write new R tools
- Makes publication quality figures, including mathematical symbols and formulae
- Compiles and runs on Windows, MacOS, and a wide variety of UNIX and Linux systems
- Has a large and active user community

There are many ways to use R: from the command line, from a script, or in a graphical interface, like RStudio.

RStudio

RStudio is an integrated development environment (IDE), which offers:

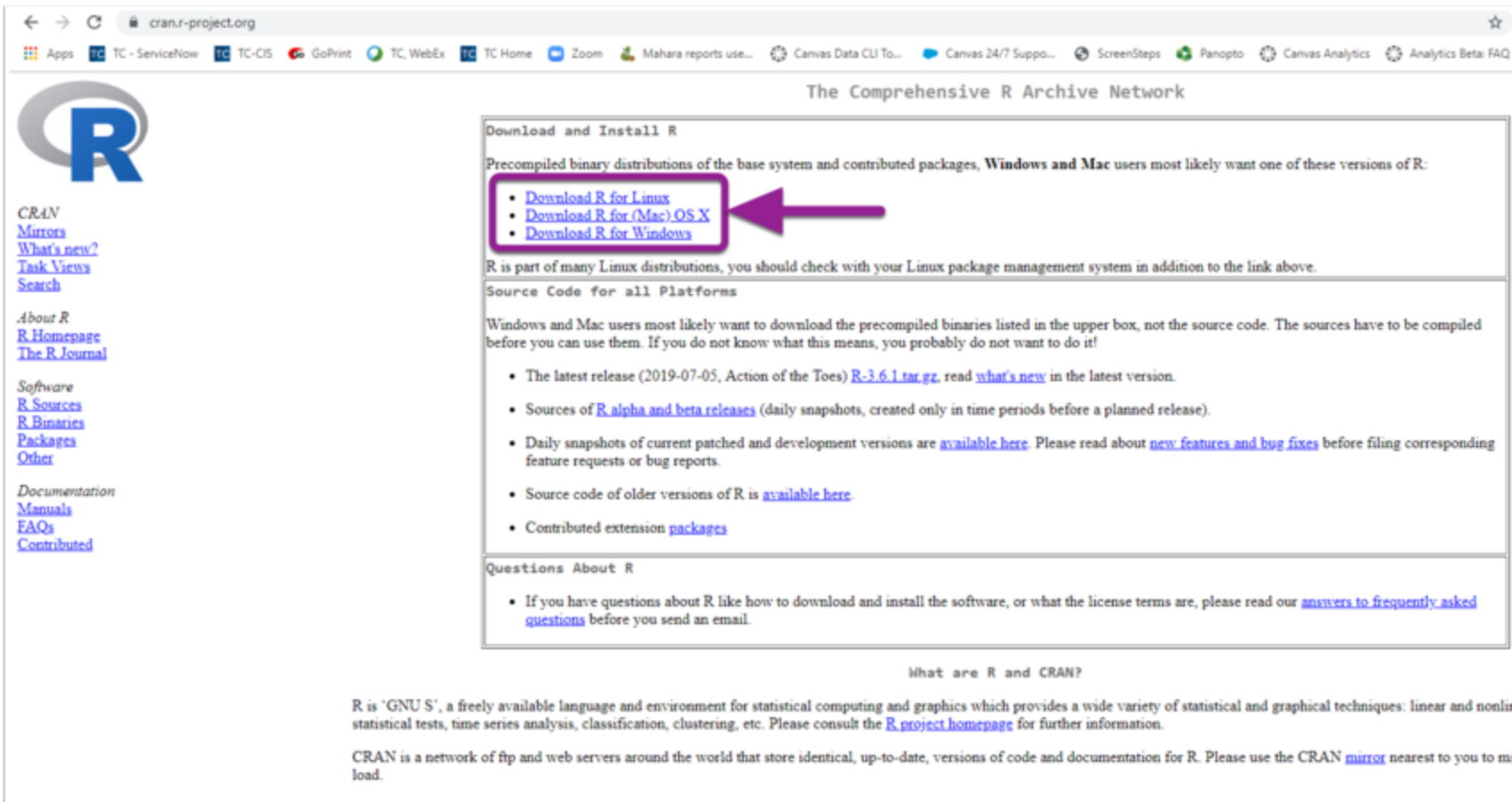
- Syntax highlighting
- Code completion
- Smart indentation
- Workspace browser
- Data viewer
- Embedded plots
- Notebooks that generate PDF or HTML results
- Package management

R Team Packages

The team behind RStudio are also the authors of a suite of R packages for data science and visualization collectively known as the “tidyverse.” We will be using their extremely popular plotting package, `ggplot2`, as well as a few other packages from the tidyverse suite later in this course.

RStudio and R notebooks

To install R, go to <https://cran.icts.res.in/>



The screenshot shows the CRAN (Comprehensive R Archive Network) website. The browser address bar displays 'cran.r-project.org'. The page features the R logo on the left and a navigation menu with links like 'CRAN', 'Mirrors', 'What's new?', 'Task Views', 'Search', 'About R', 'R Homepage', 'The R Journal', 'Software', 'R Sources', 'R Binaries', 'Packages', 'Other', 'Documentation', 'Manuals', 'FAQs', and 'Contributed'. The main content area is titled 'The Comprehensive R Archive Network' and contains a section 'Download and Install R'. This section states: 'Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:'. Below this, there is a list of three links: 'Download R for Linux', 'Download R for (Mac) OS X', and 'Download R for Windows'. A purple box highlights these three links, and a purple arrow points to the 'Download R for (Mac) OS X' link. Below the list, it says 'R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.' The next section is 'Source Code for all Platforms', which explains that Windows and Mac users should download precompiled binaries instead of source code. It lists several options: the latest release (R-3.6.1.tar.gz), sources for alpha and beta releases, daily snapshots of current patched and development versions, source code of older versions, and contributed extension packages. The final section is 'Questions About R', which suggests reading 'answers to frequently asked questions' before sending an email. At the bottom, there is a section titled 'What are R and CRAN?' which describes R as 'GNU S' and CRAN as a network of servers.

cran.r-project.org

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (2019-07-05, Action of the Toes) [R-3.6.1.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

Questions About R

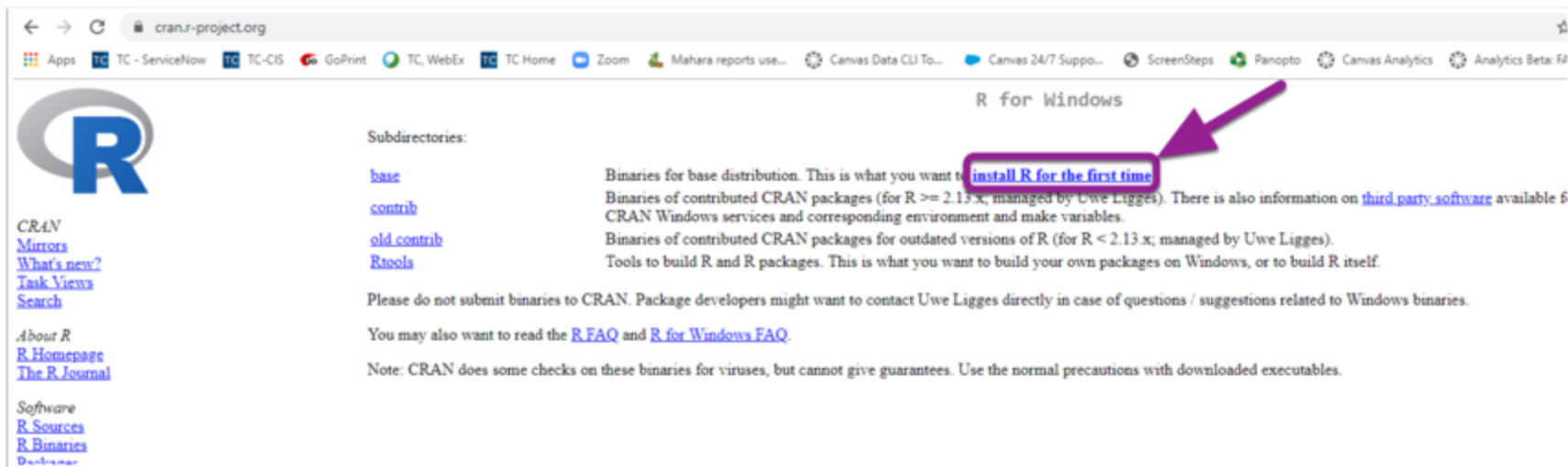
- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

What are R and CRAN?

R is 'GNU S', a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear statistical tests, time series analysis, classification, clustering, etc. Please consult the [R project homepage](#) for further information.

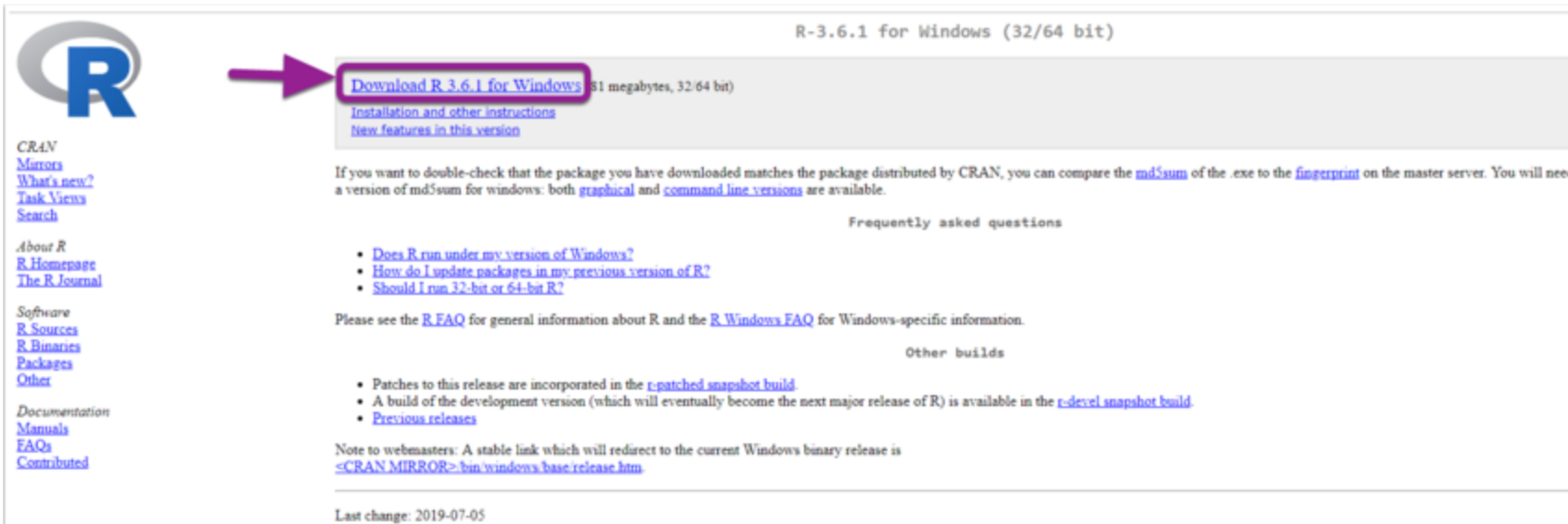
CRAN is a network of ftp and web servers around the world that store identical, up-to-date, versions of code and documentation for R. Please use the CRAN [mirror](#) nearest to you to min load.

Click on install R for the first time.



The screenshot shows a web browser window with the address bar displaying `cran.r-project.org`. The browser's tab bar contains several open tabs, including 'Apps', 'TC - ServiceNow', 'TC - CIS', 'GoPrint', 'TC, WebEx', 'TC Home', 'Zoom', 'Mahara reports use...', 'Canvas Data CU To...', 'Canvas 24/7 Suppo...', 'ScreenSteps', 'Panopto', 'Canvas Analytics', and 'Analytics Beta: F8'. The main content area of the browser shows the 'R for Windows' page. On the left side, there is a large blue 'R' logo and a list of links: 'CRAN', 'Mirrors', 'What's new?', 'Task Views', 'Search', 'About R', 'R Homepage', 'The R Journal', 'Software', 'R Sources', 'R Binaries', and 'Desktop'. The main content area is titled 'R for Windows' and contains a section labeled 'Subdirectories:'. This section lists four subdirectories: 'base', 'contrib', 'old contrib', and 'Rtools'. The 'base' subdirectory is highlighted with a purple box, and a purple arrow points to the link 'install R for the first time' within the description for 'base'. The description for 'base' states: 'Binaries for base distribution. This is what you want to install R for the first time'. The description for 'contrib' states: 'Binaries of contributed CRAN packages (for R >= 2.13.x; managed by Uwe Ligges). There is also information on third party software available 6'. The description for 'old contrib' states: 'Binaries of contributed CRAN packages for outdated versions of R (for R < 2.13.x; managed by Uwe Ligges)'. The description for 'Rtools' states: 'Tools to build R and R packages. This is what you want to build your own packages on Windows, or to build R itself.' Below the subdirectories section, there is a paragraph that reads: 'Please do not submit binaries to CRAN. Package developers might want to contact Uwe Ligges directly in case of questions / suggestions related to Windows binaries.' followed by another paragraph: 'You may also want to read the R FAQ and R for Windows FAQ.' and a final note: 'Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.'

Click Download R for Windows. Open the downloaded file.



The screenshot shows the CRAN website for R-3.6.1 for Windows (32/64 bit). The page layout includes a sidebar with navigation links, a main content area with a download button, and a footer with additional information.

CRAN
[Mirrors](#)
[What's new?](#)
[Task Views](#)
[Search](#)

About R
[R Homepage](#)
[The R Journal](#)

Software
[R Sources](#)
[R Binaries](#)
[Packages](#)
[Other](#)

Documentation
[Manuals](#)
[FAQs](#)
[Contributed](#)

R-3.6.1 for Windows (32/64 bit)

[Download R 3.6.1 for Windows](#) (81 megabytes, 32/64 bit)

[Installation and other instructions](#)
[New features in this version](#)

If you want to double-check that the package you have downloaded matches the package distributed by CRAN, you can compare the [md5sum](#) of the .exe to the [fingerprint](#) on the master server. You will need a version of md5sum for windows: both [graphical](#) and [command line versions](#) are available.

Frequently asked questions

- [Does R run under my version of Windows?](#)
- [How do I update packages in my previous version of R?](#)
- [Should I run 32-bit or 64-bit R?](#)

Please see the [R FAQ](#) for general information about R and the [R Windows FAQ](#) for Windows-specific information.

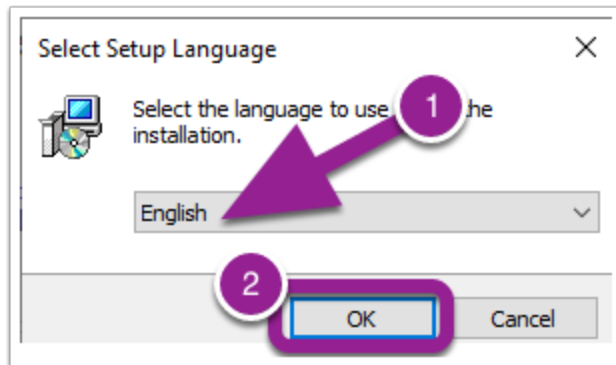
Other builds

- Patches to this release are incorporated in the [r-patched snapshot build](#).
- A build of the development version (which will eventually become the next major release of R) is available in the [r-devel snapshot build](#).
- [Previous releases](#)

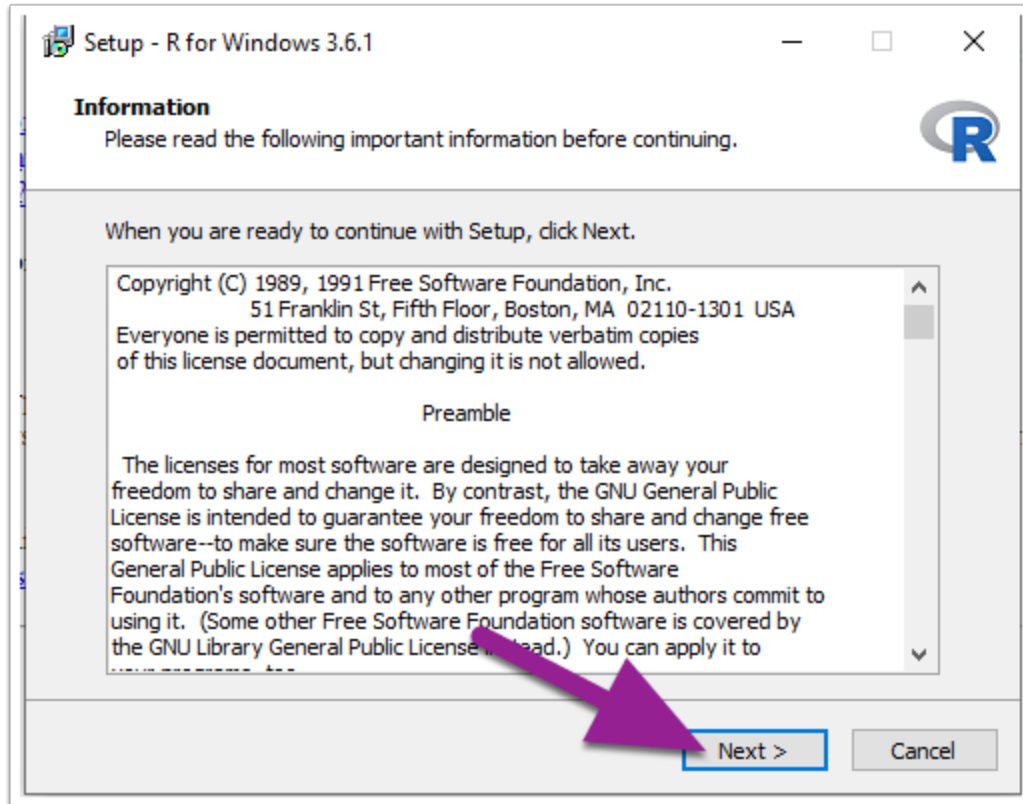
Note to webmasters: A stable link which will redirect to the current Windows binary release is [<CRAN.MIRROR>bin/windows/base/release.htm](#).

Last change: 2019-07-05

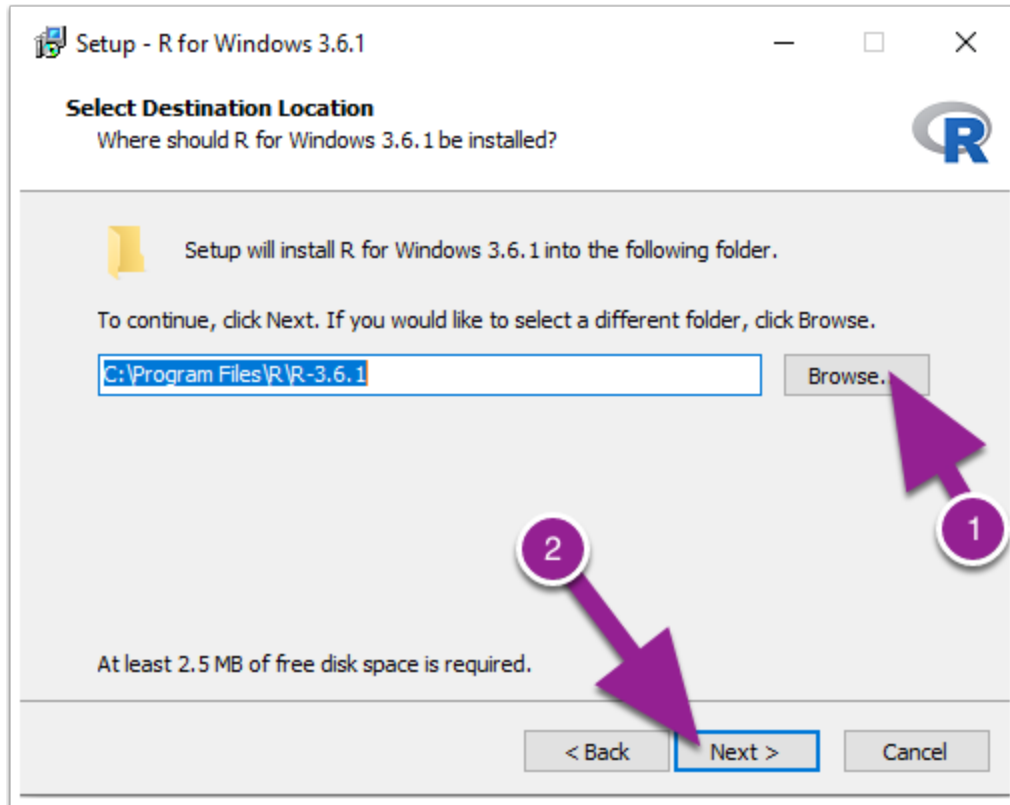
Select the language you would like to use during the installation.
Then click OK.



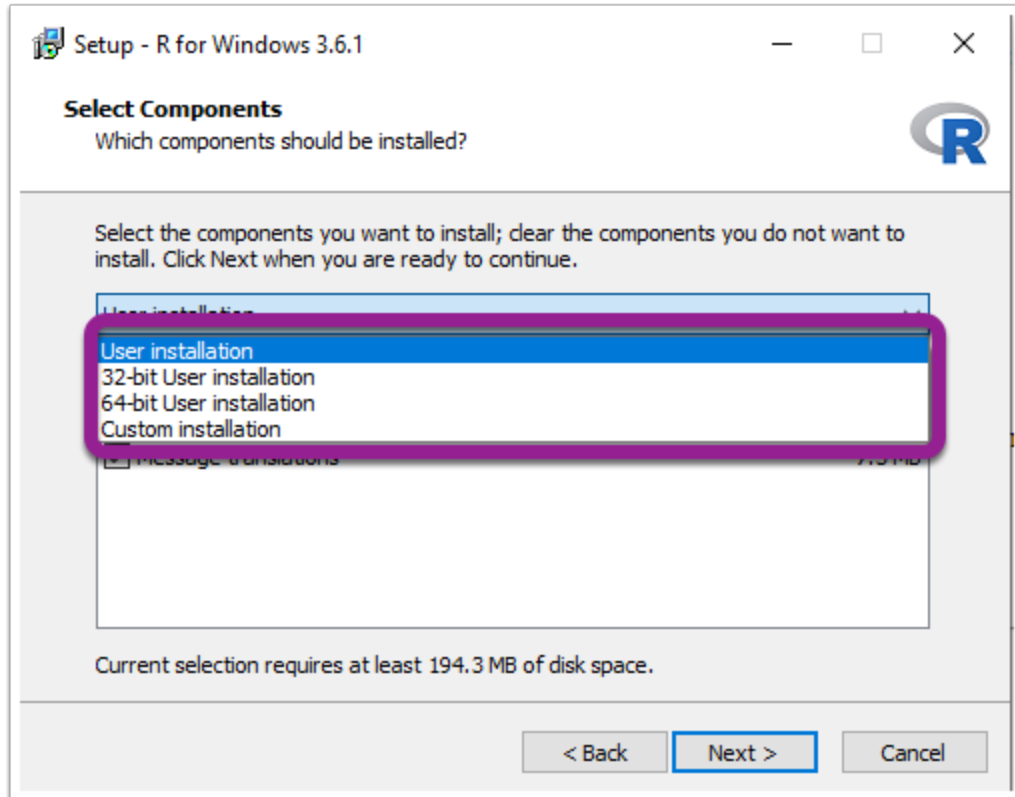
Click Next.



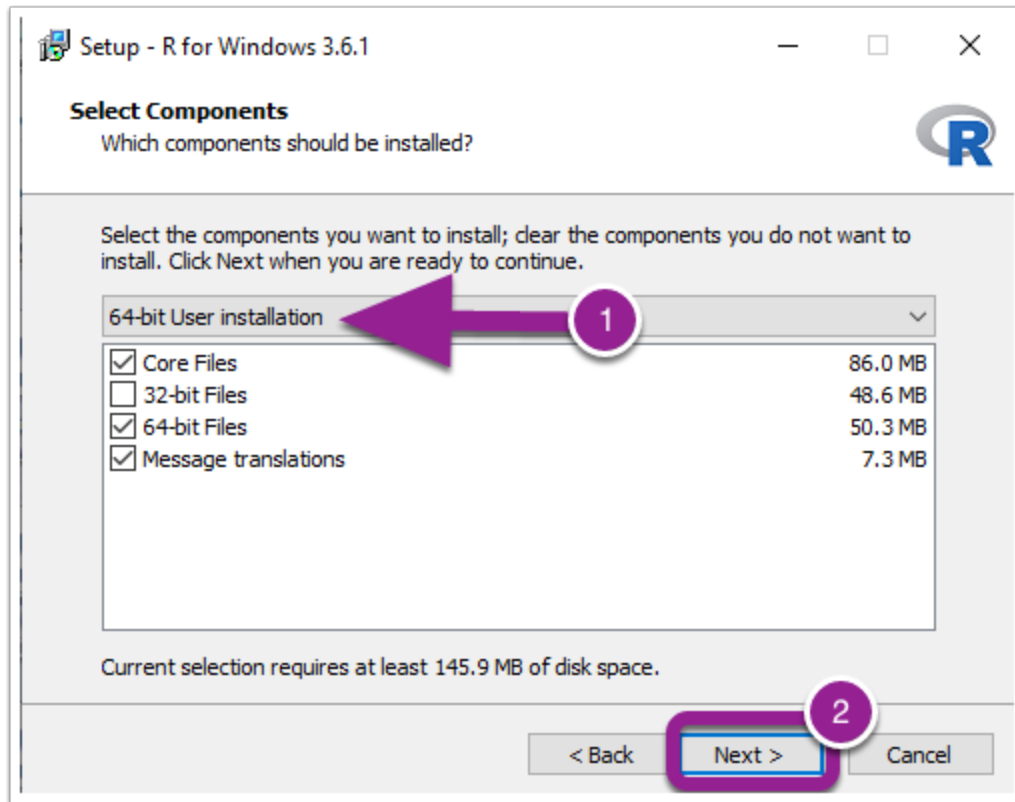
Select where you would like R to be installed. It will default to your Program Files on your C Drive. Click Next.



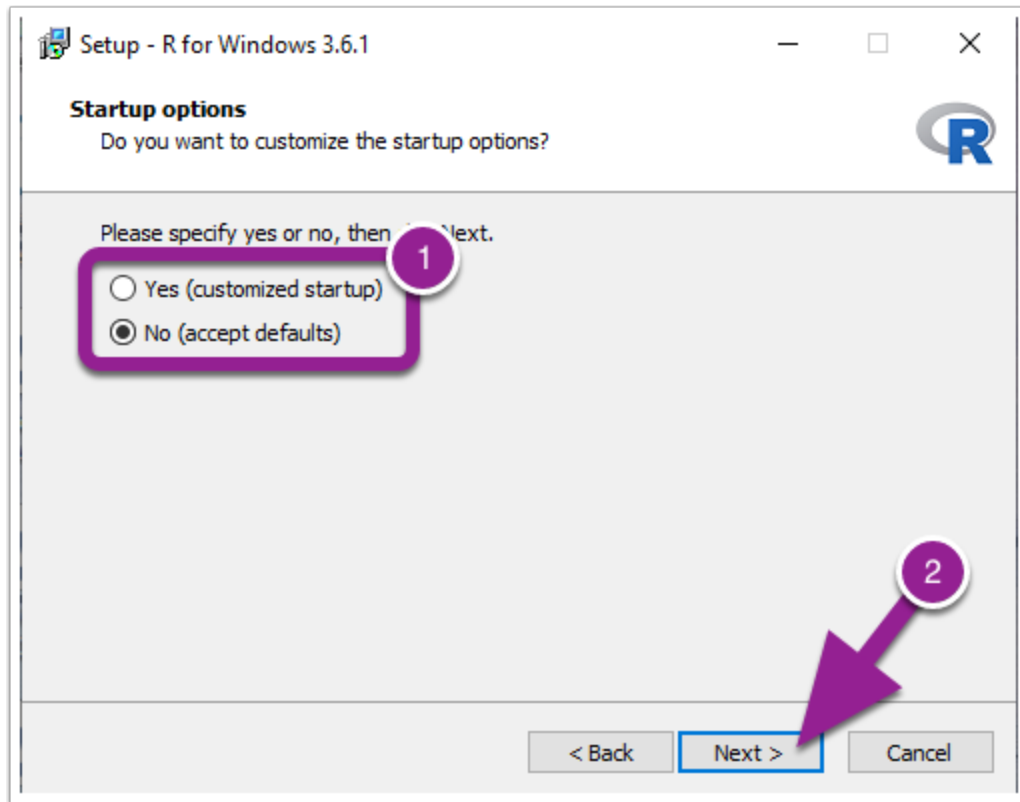
You can then choose which installation you would like.



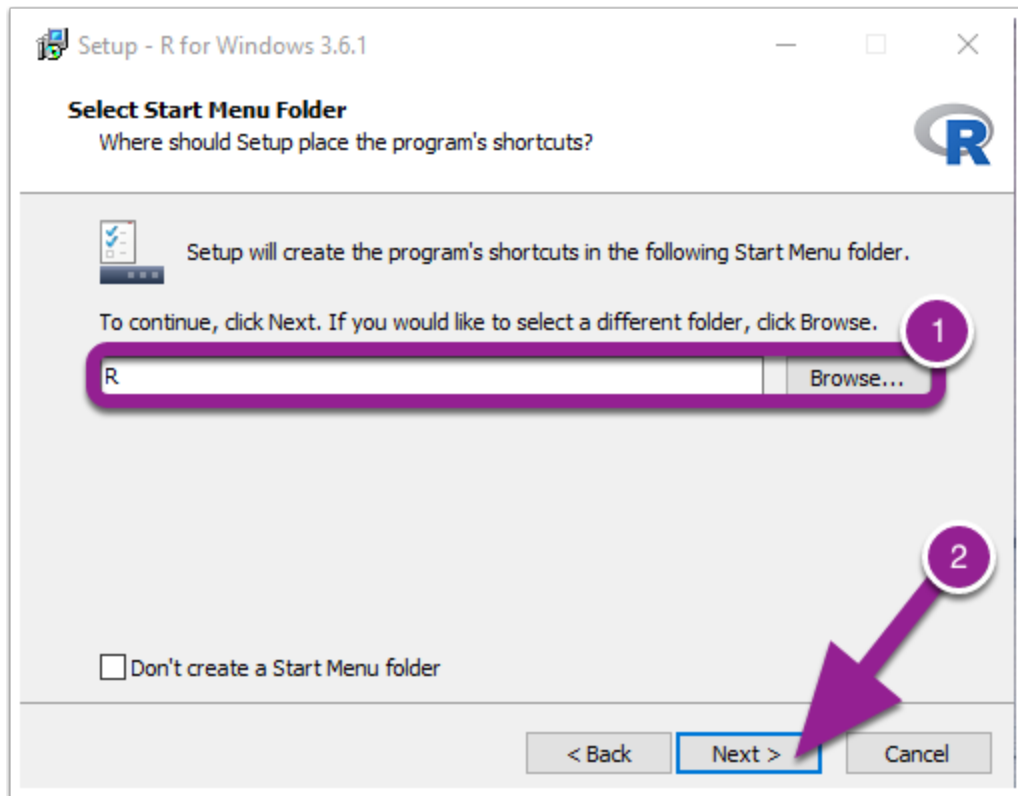
(Optional) If your computer is a 64-bit, you can choose the 64-bit User Installation. Then click Next.



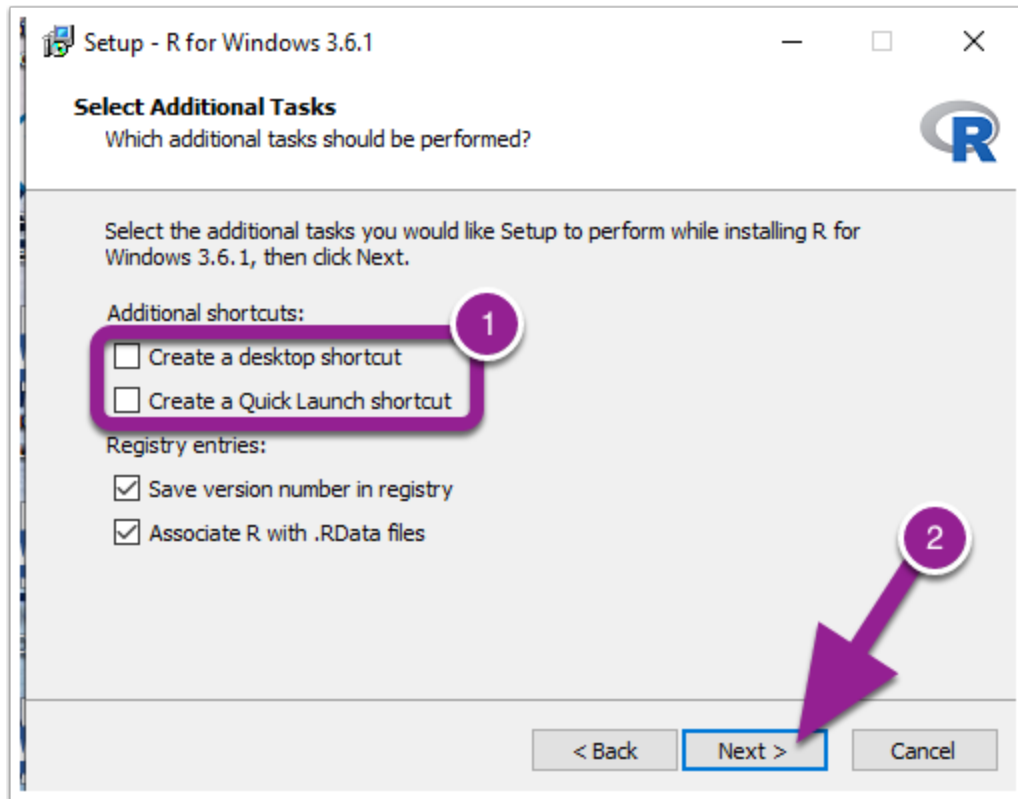
Then specify if you want to customized your startup or just use the defaults. Then click Next.



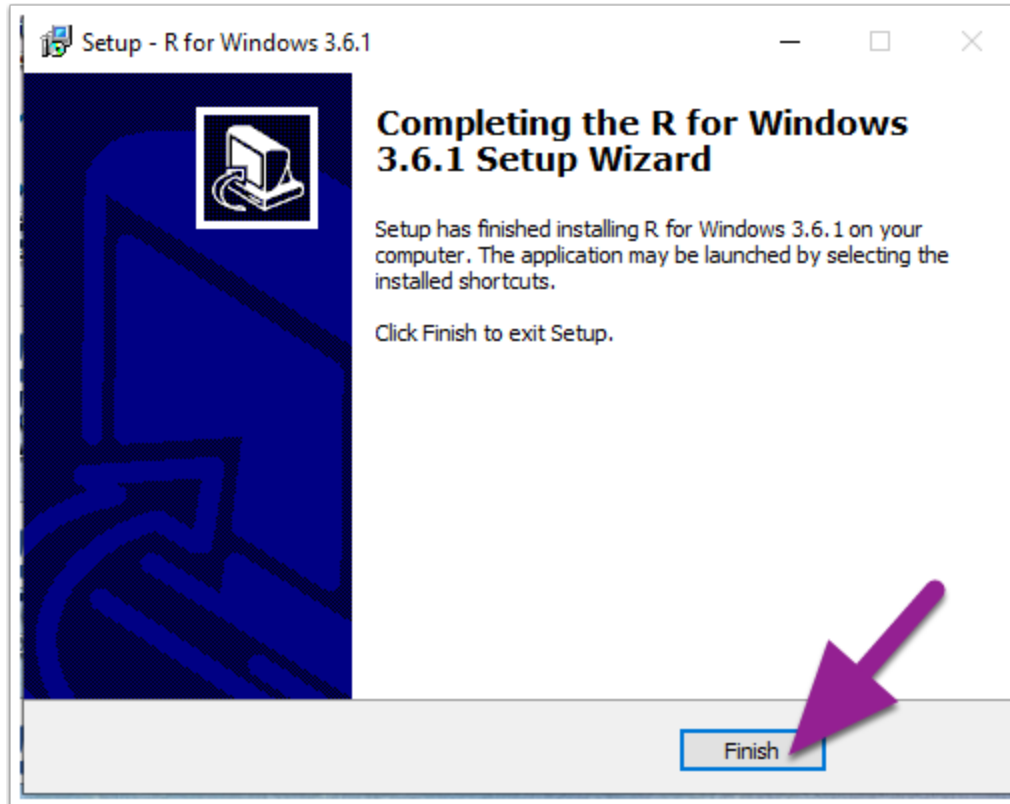
Then you can choose the folder that you want R to be saved within or the default if the R folder that was created. Once you have finished, click Next.



You can then select additional shortcuts if you would like. Click Next.



Click Finish.



Download RStudio. Go to <https://posit.co/>

[PRODUCTS ▾](#)[SOLUTIONS ▾](#)[LEARN & SUPPORT ▾](#)[EXPLORE MORE ▾](#)[PRICING](#)

1: Install R

RStudio requires R 3.3.0+. Choose a version of R that matches your computer's operating system.

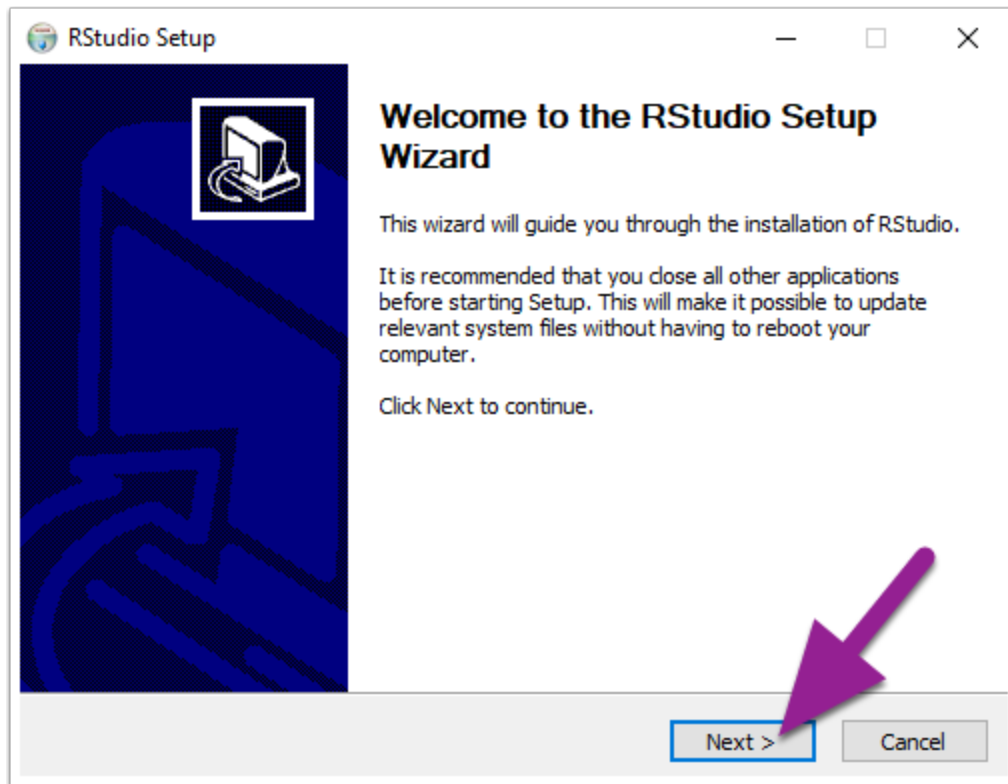
[DOWNLOAD AND INSTALL R](#)

2: Install RStudio

[DOWNLOAD RSTUDIO DESKTOP FOR WINDOWS](#)

Size: 214.34 MB | [SHA-256: FE62B784](#) | Version: 2023.09.1+494 |
Released: 2023-10-17

The RStudio installation wizard will pop-up. Click Next and go through the installation steps.

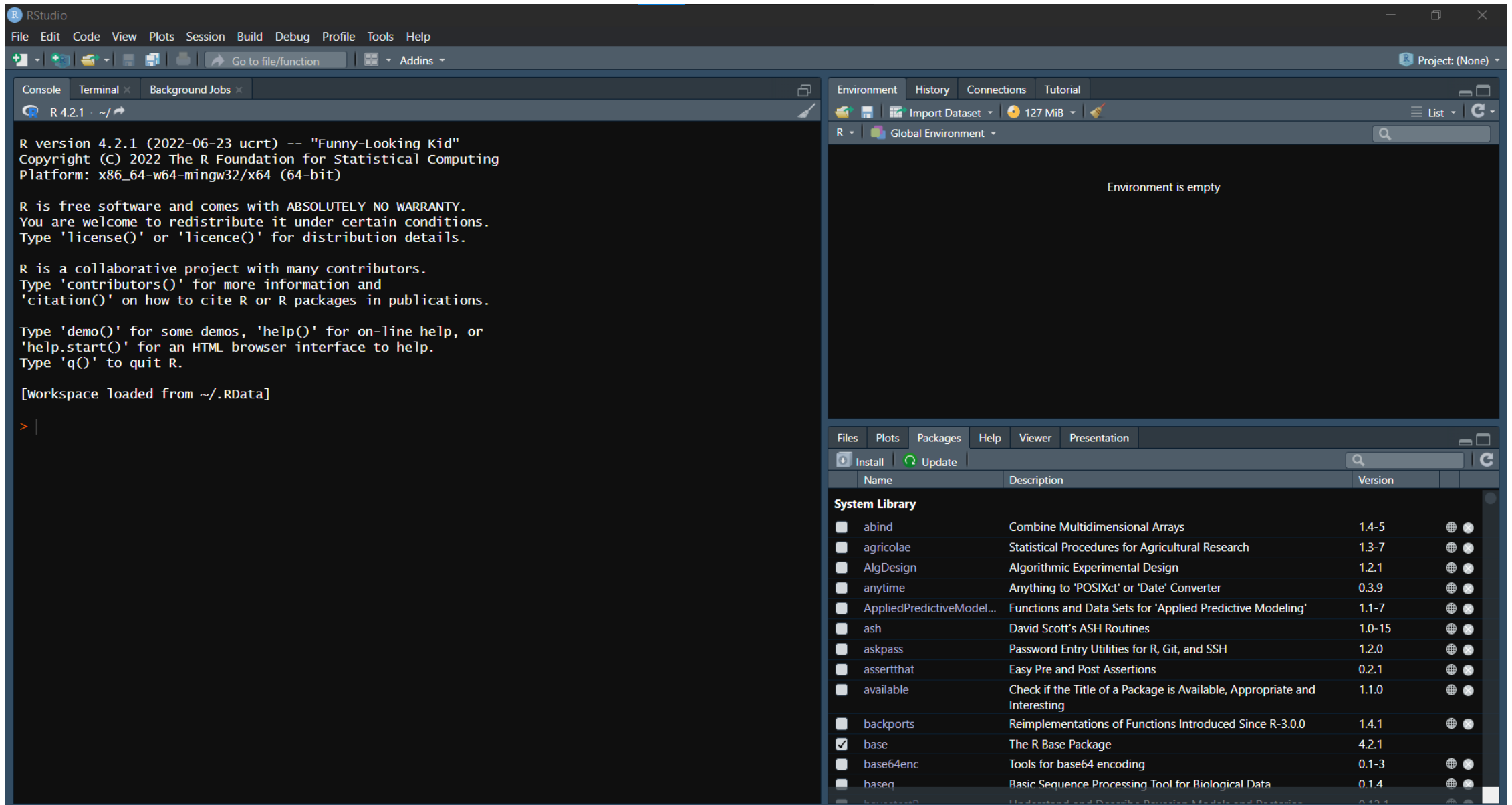


Launch RStudio

1.1 RStudio

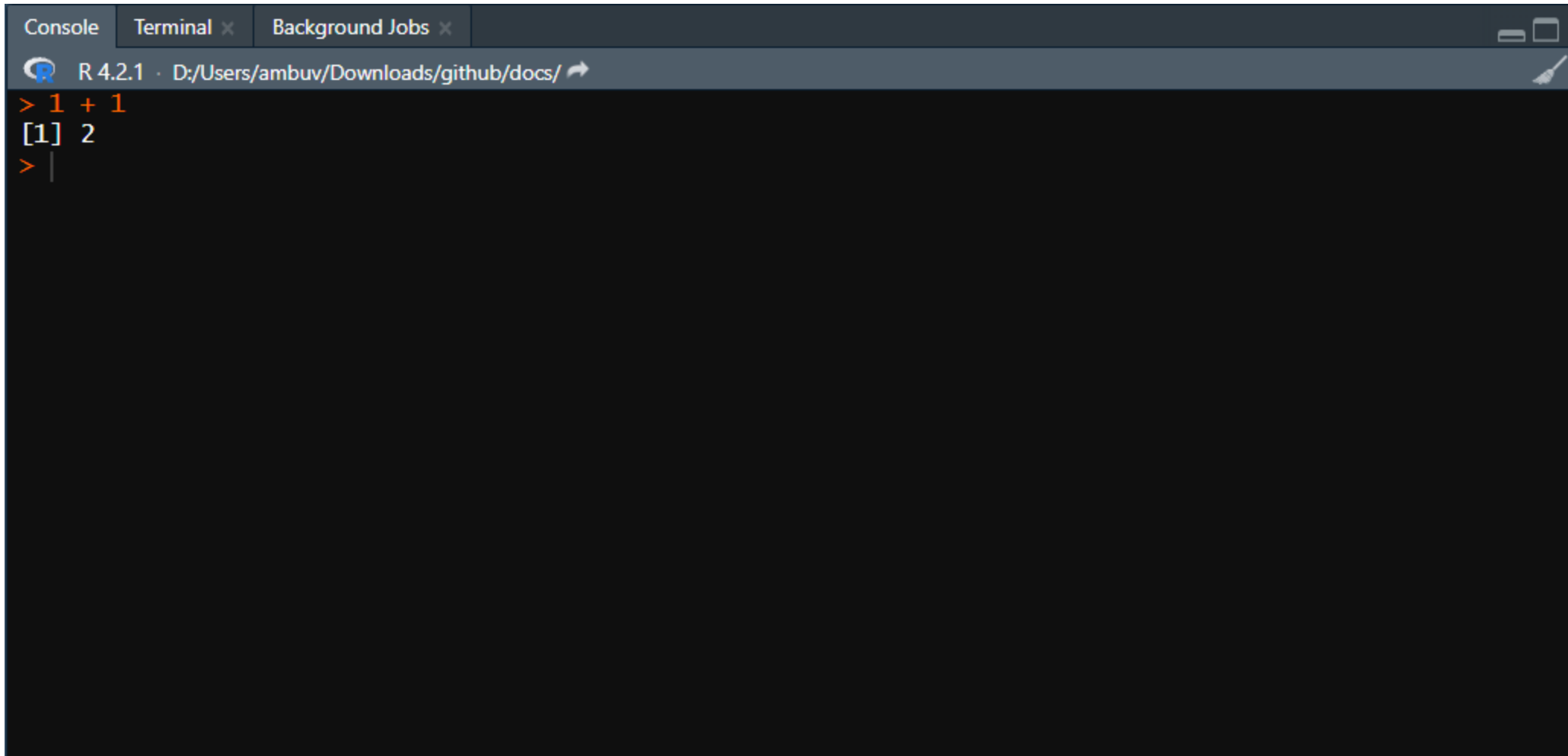
As mentioned above, RStudio is a very nice, but optional, IDE for R. All of the code for this course will work just as well on the command line, but RStudio provides a number of features that improve the experience of learning to the basics of R.

Your RStudio window should look something like this:



1.1.1 The console

On the left hand side is the console. This is a command-line interface for R; you type a command, press enter to run it, and the result will appear below.



The screenshot shows the R console window with the following content:

```
Console Terminal x Background Jobs x  
R 4.2.1 · D:/Users/ambuv/Downloads/github/docs/  
> 1 + 1  
[1] 2  
> |
```

Running a simple command

In R the prompt is a `>` character. When you see this character at the beginning of the line in your console, it means that R is waiting for you to type your next command.

In this course, code is going to appear in gray boxes, like this:

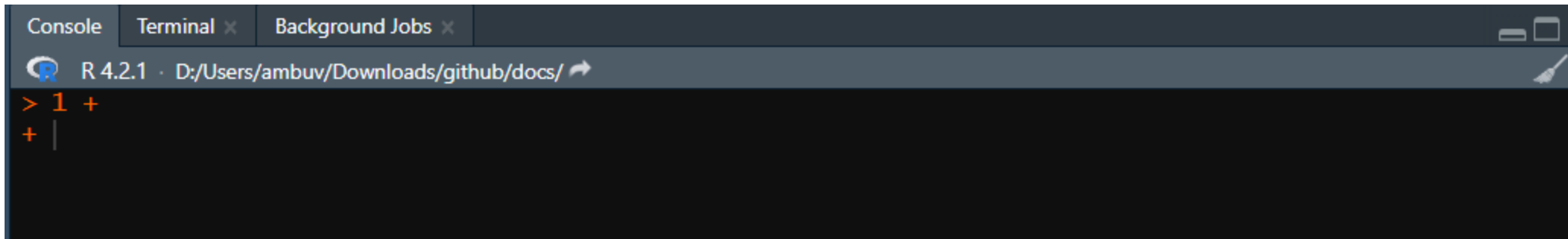
```
1 + 1
```

Whenever you see one of these boxes, try running the code yourself.

What happens if you press enter before you meant to?

```
1 +
```

Running incomplete command



The screenshot shows an R console window with the title bar 'R 4.2.1 · D:/Users/ambuv/Downloads/github/docs/'. The console has tabs for 'Console', 'Terminal', and 'Background Jobs'. The prompt is '>'. The user has entered '1 +', and the next line shows a '+' character followed by a vertical bar, indicating that R is expecting more input.

```
> 1 +  
+ |
```

If you press enter before finishing a command, the next line will begin with a + character. This lets you know that R is expecting more input.

1.1.2 The workspace browser

In the upper right panel is the workspace (or environment) browser. This part of the window shows the objects present in the environment. Currently there are none.

The easiest way to explain an object is to run some code.

```
a <- 1 + 1
```

Unlike the first time we ran `1 + 1`, nothing is printed to the console.

Instead, a new value appears in the workspace browser.

Console Terminal x Background Jobs x

R 4.2.1 · D:/Users/ambuv/Downloads/github/docs/

```
> a <- 1+1  
> |
```

Environment History Connections Tutorial

Import Dataset 127 MiB

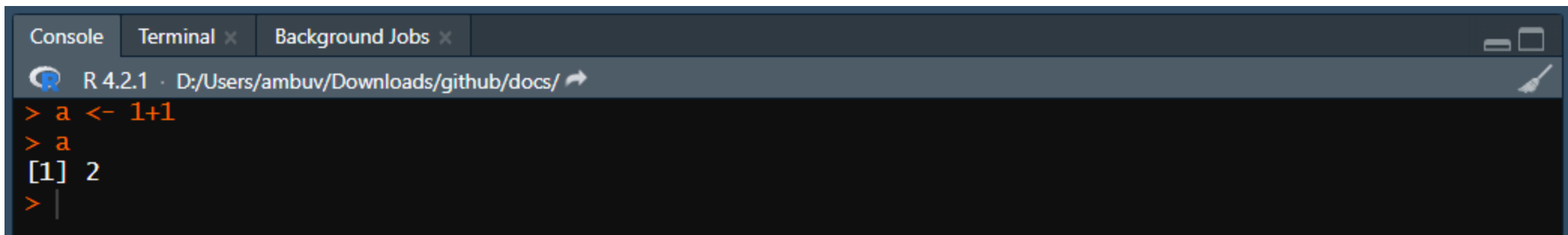
R Global Environment

Values

a	2
---	---

The `<-` is a special pair of characters called the assignment operator that stores the result of the addition in the object referred to on the left-hand side of the "arrow."

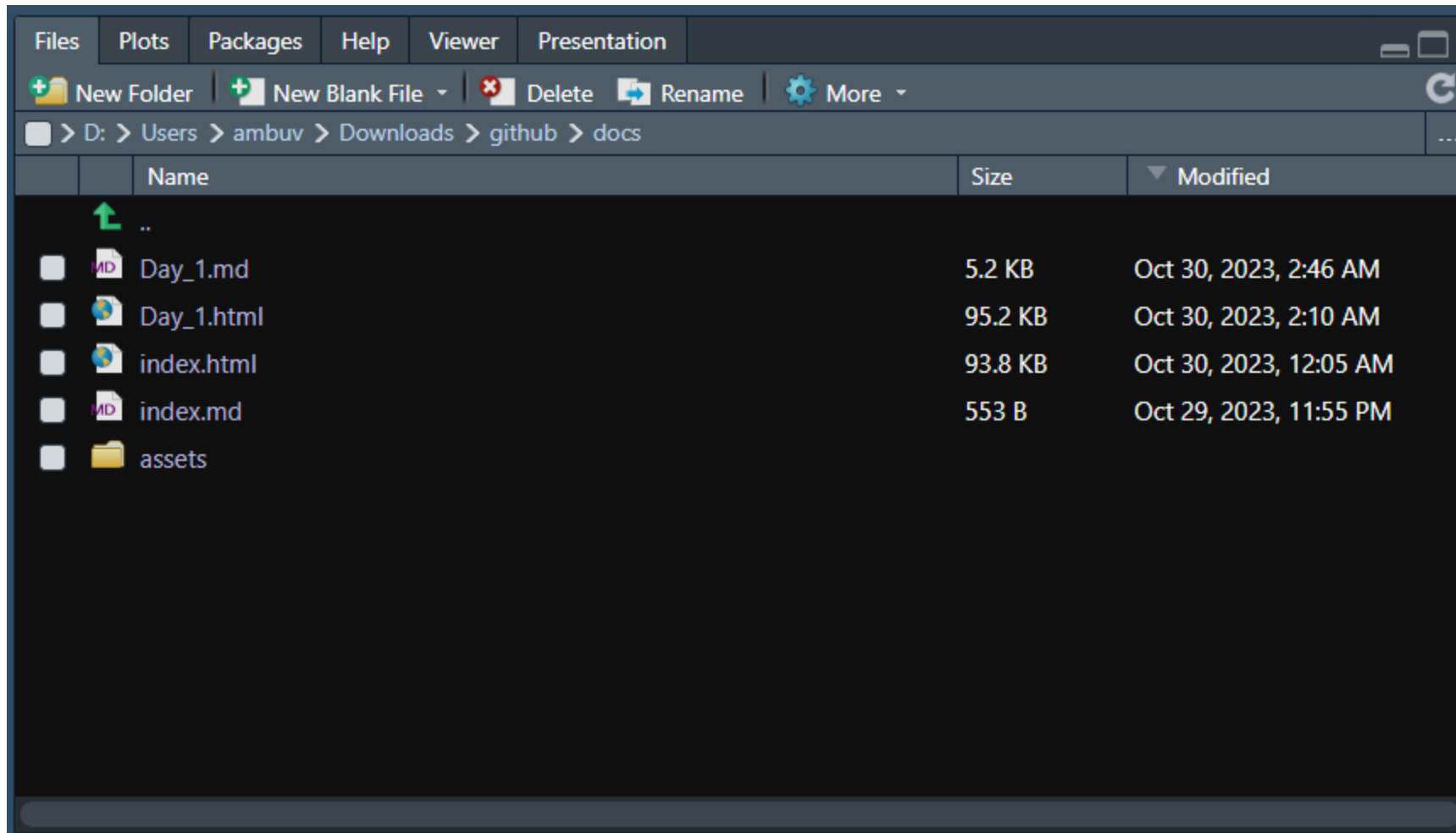
Since no object called "a" existed in R's environment, R created a new object "a" to hold the result of the addition operation $1 + 1$.



```
R 4.2.1 · D:/Users/ambuv/Downloads/github/docs/
> a <- 1+1
> a
[1] 2
> |
```

1.1.3 Help documentation

The lower right pane contains the file browser, plot viewer, and help documentation, which we will be using frequently.



2 Importing and exporting data

For the remainder of this course, we will be working with a single data set to answer a number of simple biological questions about the effects of maternal cigarette use during pregnancy.

Begin by setting up a working directory for this course.

To make things easier, we'll set R's working directory to the directory named "workshop". To do this, select the "Session" menu in the toolbar, then navigate to "Set Working Directory," and click on "Choose Directory." Choose the "worksop" directory.

When you do so, you should see that a line of code ran in your console:

```
setwd("D:/Users/ambuv/Desktop/workshop")
```

file names change based on individual systems

Download the data.

```
download.file("https://raw.githubusercontent.com/ucdavis-bioinformatics-  
training/2022_February_Introduction_to_R_for_Bioinformatics/main/birthweight.csv",  
"birthweight.csv")
```

2.1 Import data using read.csv()

R has a number of functions for reading data in a variety of formats. Let's use the `read.csv()` function to read in a spreadsheet containing data from an experiment.

```
birthweight <- read.csv("birthweight.csv")
```

CSV stands for "comma separated value," and the CSV file is simply a text file where each row in the file represents a row in the data table, and the columns are separated by commas. The contents of the CSV file are now stored in the variable "birthweight."

2.2 Export data using write.csv()

To write the contents of the birthweight object to a new CSV, we can use the write.csv() function.

```
write.csv(birthweight, file = "new_birthweight.csv")
```

The similar read.delim() and write.delim() can be used to read and write tab-delimited files, where columns are separated by tab characters rather than commas.