

Software installation Guide

Introduction to R for applied epidemiology

Date revised: 2022-03-15

Table of Contents

Overview	2
Preparation.....	2
What to install:	2
If you are using a personal computer.....	3
Notes about permissions and network drives.....	4
Installation order:	Error! Bookmark not defined.
Installing R without administrator rights:	4
Installing RStudio without administrator rights:	4
Caveats:	5
What if I can't even download R or RStudio?	5
RStudio on a server:	6
Detailed installation instructions	7
R.....	7
Downloading R:.....	7
Installing R:.....	9
RStudio.....	17
Downloading RStudio:.....	17
Installing RStudio:.....	17
Testing the installations:	20
R Packages	25
Rtools:.....	25
Git.....	30
Wrap-up.....	31

Overview

In this guide, you will find instructions to help you install R and the other software required for this course. The approach we have suggested below should not require administrator rights, as we understand that many people will be using a computer provided by their place of work, which may have restrictions on installing new software.

We recommend that you read this guide and try to complete the installations well before the start of the course; this will allow time for troubleshooting and contacting your local IT department if necessary.

Preparation

What to install:

There are five bundles of software that you will need to install. The first four are essential for this course, while the last one will become essential R companions once you start working with the program regularly:

Essential for the Introduction to R course:

- R (this is the R program itself)
- RStudio (this is a user-friendly interface for R - an *integrated development environment* (IDE))
- RTools (compiles R packages that have not yet been converted to windows binaries)
- R packages (packages contain sets of functions that you will use in R to do tasks)

Recommended for long-term R use:

- Git (a program that interfaces with Github and Gitlab version control systems)

A quick note on why each of these elements is needed:

The R program itself has a very basic user interface, in which you can write and store code as plain text files. Most R users (from beginner to experienced alike) prefer to open and use R within RStudio, which is another program that provides a much more user-friendly interface with many powerful features to help limit errors in your code and run automated reports, to name a few. You can read more about R and RStudio in the [R basics chapter of the Epidemiologist R handbook](#).

Over time, thousands of R users have created new R functions; these are normally grouped together and made available to other users in downloadable *packages*. Each package contains a set of functions to deal with a specific task or type of analysis. This makes R highly customisable, because you can choose the packages that are relevant to your work and the type of analysis that you are doing. You can read more about R packages [here](#).

If you are using a personal computer with administrator rights

If you will be using a personal computer for the course for which you have administrator rights, the steps may be more simple than if you are using a “workplace” computer that has restrictions on what you can easily download.

If using a personal computer for which you have administrator rights, simply install the latest versions of R ([from here](#)) and RStudio ([from here](#)) and RTools ([from here, only if on Windows machine](#)) for your machine as you would normally install software, and into the default locations.

You can see detailed instructions below but realize that they contain steps that written primarily to assist people using workplace computers who may encounter restrictions with downloading files. These steps might include manually creating a folder called “R” and directing the installations there, instead of the default “Program files” folder.

After installing these three software, R, RStudio, and RTools (if on a Windows machine), follow the instructions in the “R Packages” section below. Within RStudio, you will run an R script (a file with code) provided by us that will automatically install the necessary R packages for the course.

If your IT staff is helping you install

Together, you can install R, RStudio, and RTools... but pay attention to the following details:

1. It is best to remove any old R installations on the computer. When you do this, do not accidentally delete old work files of the course participant – only the R installation and R package library. If you are unsure, please check with the course participant, or with the Applied Epi instructor.
2. If there is a syncing program on the computer such as OneDrive, install R and RStudio to the My Documents folder (ideally outside the reach of OneDrive). If the R installation lives in a synced folder this can often cause problems running R.

Notes about permissions and network drives

Many workplaces have restrictions on computers and other devices, to prevent installation of malicious or virus-contaminated software. Unfortunately, this can often place restrictions on the download and installation of work-related software too.

At the most basic level of restriction, installing a program from an executable file (.exe) in the Program Files folder on a PC may require that you log in as an administrator before proceeding.

Installing without administrator rights:

If you do not have administrator rights and cannot otherwise consult with your IT to help you install, you can still install everything in the following order:

1. Create a folder called R in the My documents folder on your computer
2. Install the R program in this folder
3. Unzip RStudio in the same folder
4. Create a short-cut to RStudio.exe in the task bar or on your desktop
5. Open RStudio and follow the instructions below to install the necessary packages.

Installing R without administrator rights:

As many R users face this problem, it is possible to install R in the My documents folder on your computer instead. This is a folder that you have personal access to and does not require administrator rights. The file path to this folder on a PC will be something like this:

```
C:\Users\your_username\Documents
```

We suggest you create a new folder called R in this location and install R, RStudio and any other R-related software there. Installing them in the same folder will ensure that RStudio is able to 'find' R (it will not work without it).

Installing RStudio without administrator rights:

RStudio requires a different approach as the default setting for the installer is to install it in the Program files folder. Fortunately, as again this is a problem many people face, RStudio is available to download in several different formats. The format which will work best if you don't have administrator rights is the .zip file. This is actually a pre-compiled, compressed version of RStudio which is ready to use once unzipped. You can use 7-zip or any other unzip program to unzip it.

Caveats:

Installing programs outside the Program files folder on a PC means that certain features which would automatically be activated need to be performed manually. The two main ones are listed below:

- Putting a link to RStudio in the start menu, task bar, or desktop
- Associating file endings with RStudio

The instructions below explain how to adjust these settings manually.

What if I can't even download R or RStudio?

Some IT restrictions are even more limiting in that they will detect attempts to download .exe and .zip files from your web browser and block them. If this happens to you, in the first instance try a different browser (e.g. try Google Chrome instead of Microsoft Edge). If that doesn't help, it is advisable to contact your IT department and ask them to help you install R and RStudio. If possible, ask a colleague who already uses R in your institution if there are any specific requirements that need to accompany your IT request.

It is worth noting that when this level of restriction is applied, large institutions often have a **software download center** where you can download and install popular work software that has been pre-approved. R and RStudio are increasingly being included in these download centers, but be aware that they may not have the latest version. R and RStudio are regularly being updated and improved, and new versions with major revisions come out every 6 months or so. For this reason, it is advisable to contact your IT department and ensure that the latest versions are in the software download center before installing. This is a consideration for updating R too.

If you have difficulties accessing IT support, as an interim measure you could download R and RStudio on a non-work computer and transfer them to your work computer on a USB. Be aware, however, that some systems may still stop R from installing with this method.

RStudio on a server:

Some institutions may make R and RStudio available to you on a central server. The idea behind this is that R and RStudio are then available to everyone. There are some caveats to this however:

- If you are offline due to poor or no internet connection, you will not be able to use R
- The R and RStudio versions available on the server may not be the latest ones
- You may not have rights (or have limited rights) to install R packages on the server

If this is your situation, it is advisable to contact your IT department and ask them to help you install R and RStudio on your own PC. This will give you more flexibility and you can work off-line. However if in the future you plan to analyse very large data sets, R on a server may be more appropriate as extra memory can be allocated.

Detailed installation instructions

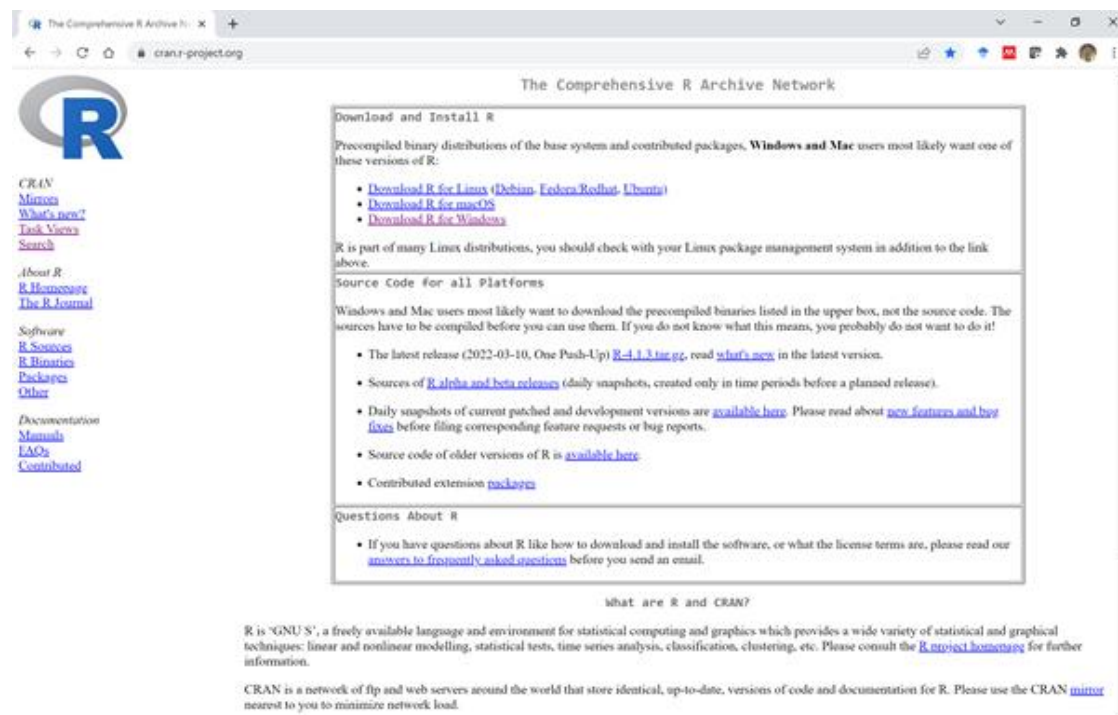
Step-by-step installation instructions have been provided below, along with some screenshots.

R

Downloading R:

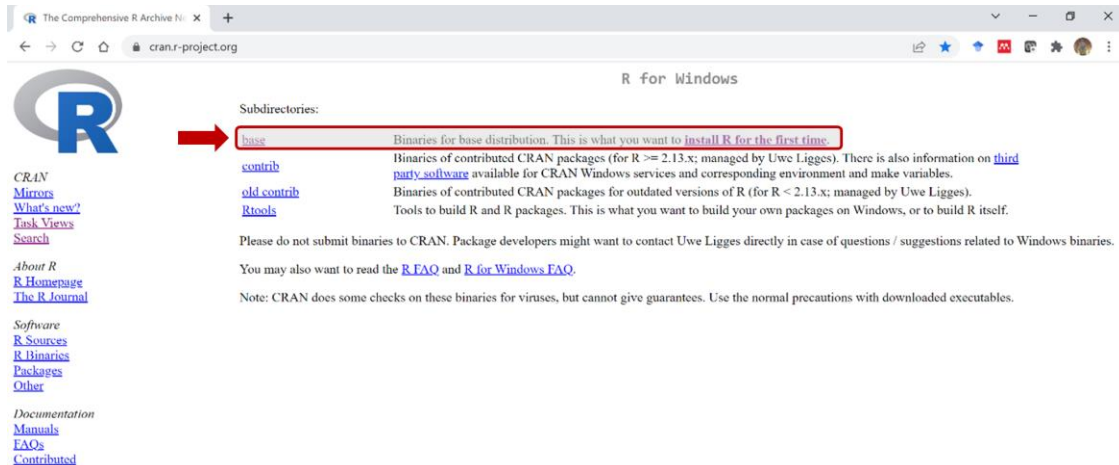
The latest version of R can be downloaded from the [Comprehensive R Archive Network \(CRAN\)](https://cran.r-project.org/) website. Here you can choose the version of R appropriate for your computer's operating system (Windows, MAC or Linux):

Note: If you already had R installed on your computer from some time ago, to update the version you must *re-install* R from the website below. The old version of R can remain on your computer, but the newer version will be used by RStudio from now on. Later in the document we describe steps to take if RStudio does not automatically recognize the new R install.



CRAN website front page

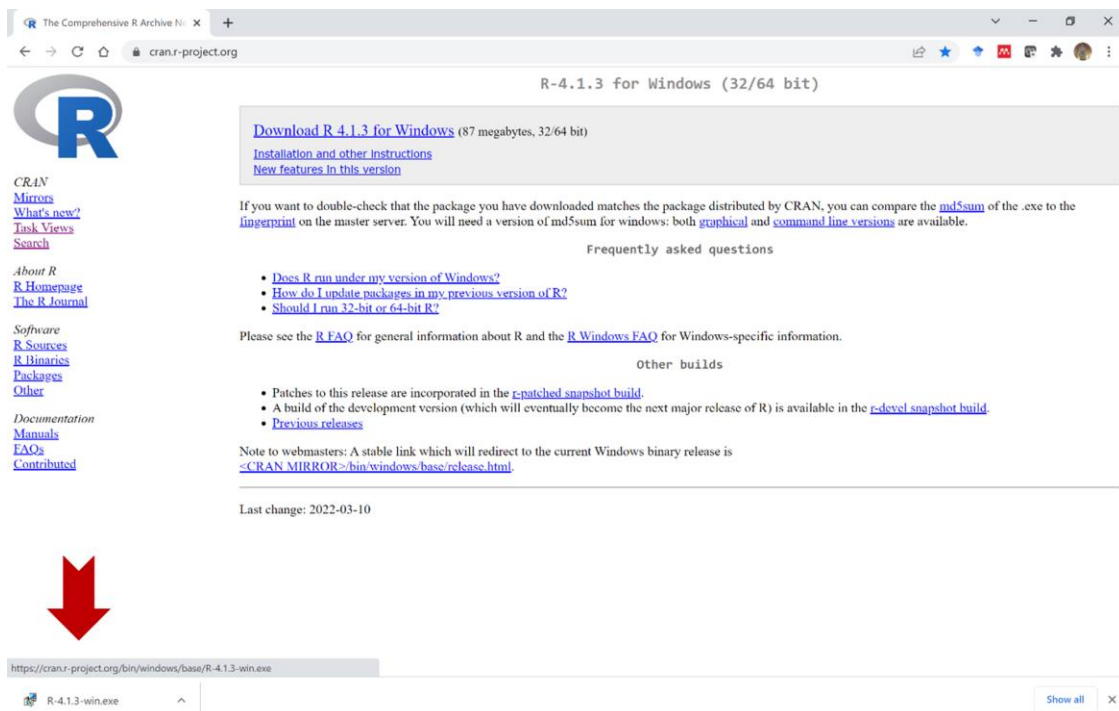
Once you have selected the appropriate version of R for your operating system, clicking on the link will bring you to the next page, where you can choose the suite of packages that will be installed along with R. Normally choosing base will be sufficient (highlighted in red in the image below); other packages can be installed later.



The screenshot shows the CRAN website for Windows. On the left is a sidebar 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 'R for Windows'. Under 'Subdirectories:', there are links for 'base', 'contrib', 'old contrib', and 'Rtools'. The 'base' link is highlighted with a red box and a red arrow pointing to it from the left. Below the links, there is text explaining that the 'base' binaries are for the first-time installation of R, and that users should not submit binaries to CRAN. There are also links to 'R FAQ' and 'R for Windows FAQ'.

CRAN choose R installation

Next, click on the base link to start downloading R. If successful, you should see the downloaded R installer executable (.exe) file in your browser window:



The screenshot shows the CRAN website for R 4.1.3 for Windows. The page title is 'R-4.1.3 for Windows (32/64 bit)'. There is a 'Download R 4.1.3 for Windows (87 megabytes, 32/64 bit)' link. Below this, there are links for 'Installation and other instructions' and 'New features in this version'. A paragraph explains how to verify the download using md5sum. There is a 'Frequently asked questions' section with links to 'Does R run under my version of Windows?', 'How do I update packages in my previous version of R?', and 'Should I run 32-bit or 64-bit R?'. There is also a 'Please see the R FAQ for general information about R and the R Windows FAQ for Windows-specific information.' section. Below that is an 'Other builds' section with links to 'Patches to this release are incorporated in the 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.', and 'Previous releases'. At the bottom, there is a note to webmasters about a stable link that will redirect to the current Windows binary release, and a 'Last change: 2022-03-10' timestamp. A large red arrow points down from the 'Download R 4.1.3 for Windows' link to the browser's address bar, which shows the URL 'https://cran.r-project.org/bin/windows/base/R-4.1.3-win.exe'. The browser's file manager shows the downloaded file 'R-4.1.3-win.exe'.

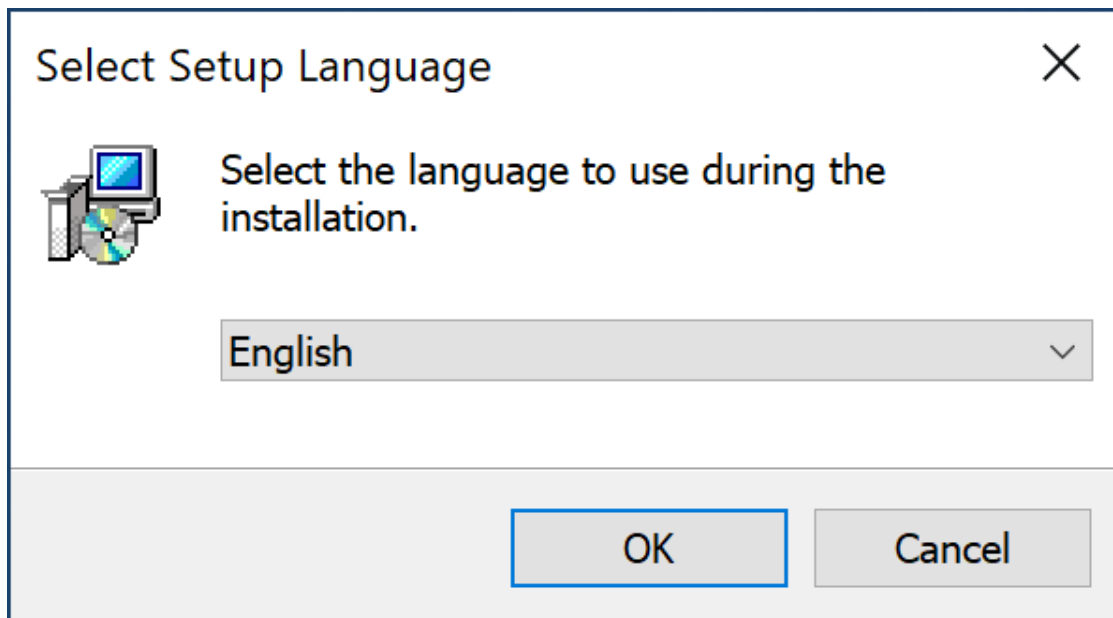
CRAN download R

Note: if the download fails, this may be because executable files (ending in .exe) have been blocked by your institute's firewall. In this case, ask your IT department to download the latest version of R for you and proceed to the next section.

Installing R:

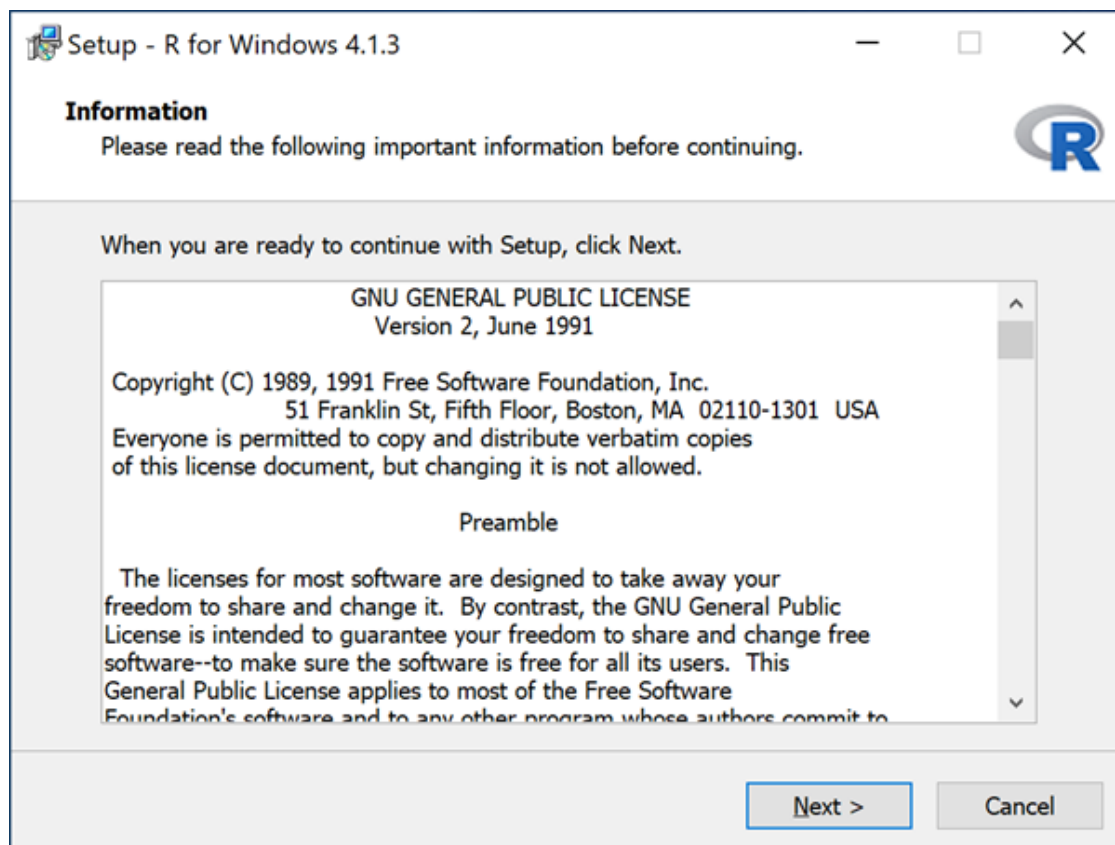
To begin, double-click on the executable file that you downloaded in the last section. At this stage, depending on the settings on your computer, you may be asked to log in as an administrator before proceeding. If this is the case for you, ask your IT department for assistance. They may be able to log in remotely and allow you to complete the install yourself, or alternatively they may prefer to complete the installs on your behalf.

If you see a message asking if you would like to make changes to your computer, click ok. This will bring up the first dialogue window in the installer, which allows you to select the language from the dropdown menu. In this example we will leave it on the default setting (English):



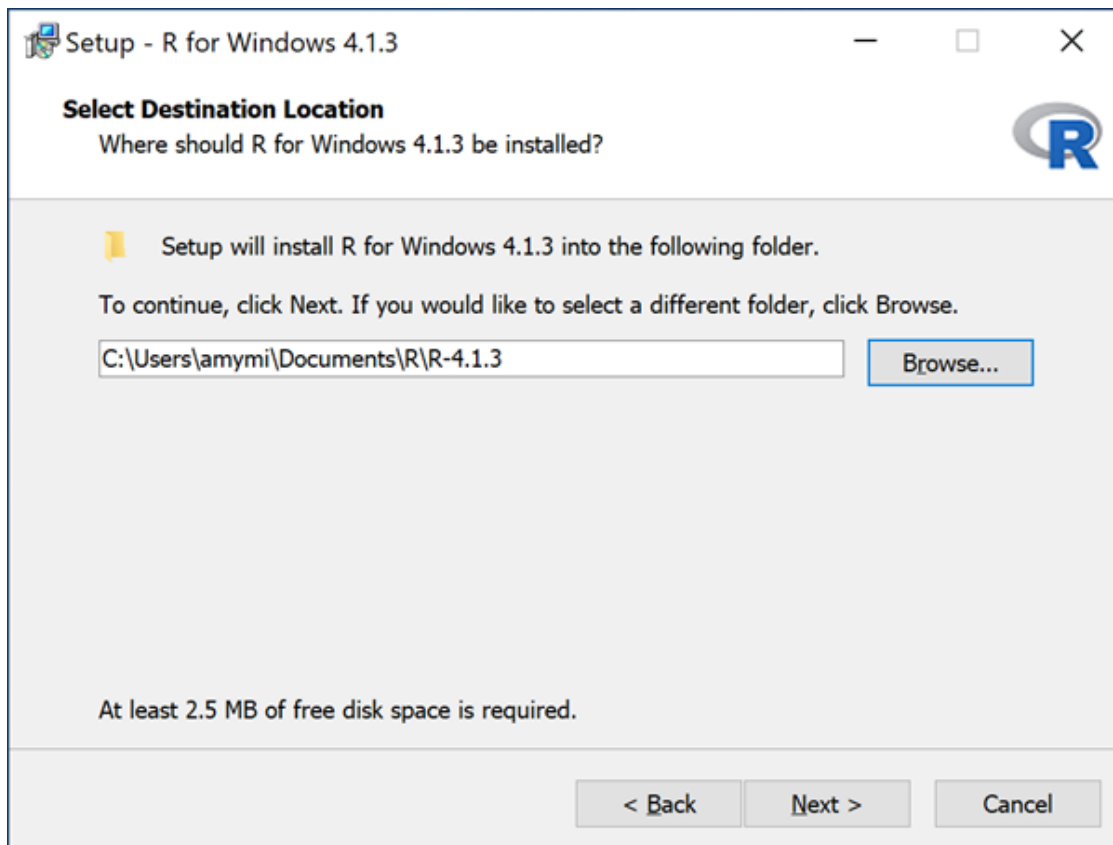
Installing R: select language

R is distributed with a GNU general public licence. In the next dialogue box you will be invited to read and accept the licence conditions (click next when done):



Installing R: accept licence

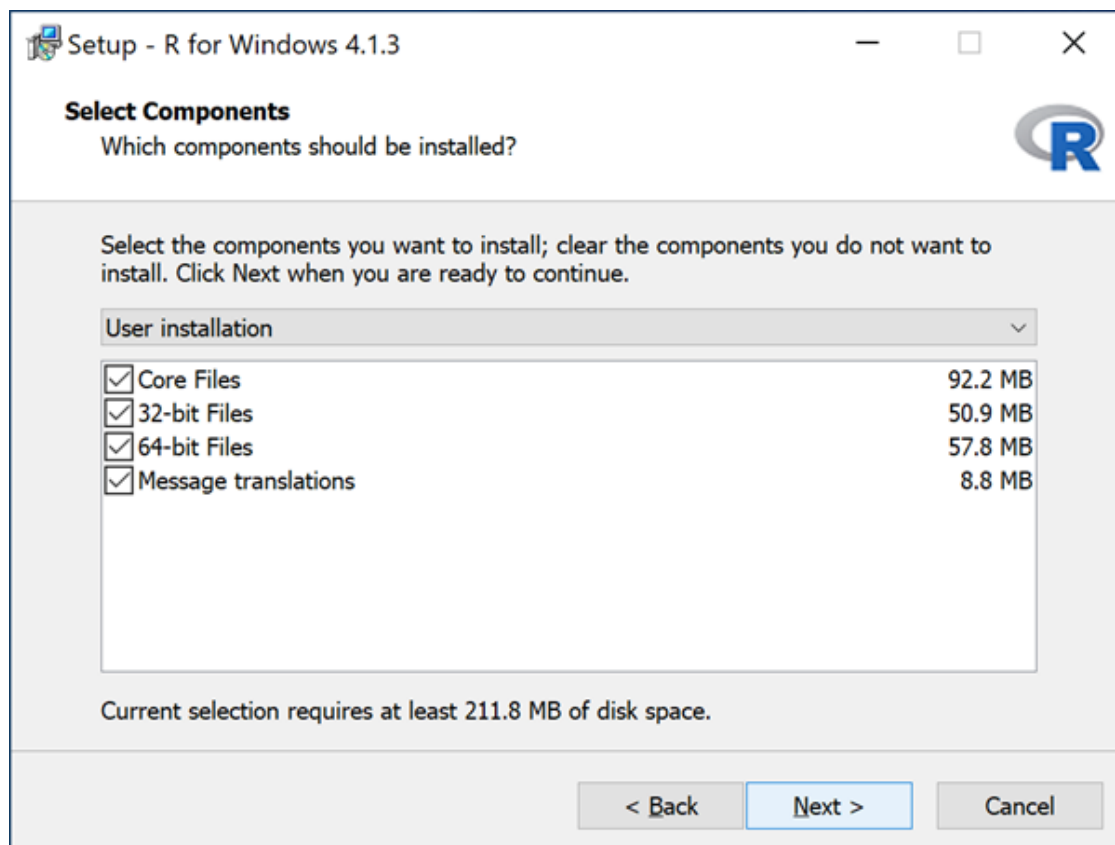
In the next dialogue box, you will be asked to choose where to install R on your computer. The default location is in the Program Files folder; however IF you do not have administrator rights on your computer you will not be able to install R there. If this is the case, we suggest that you create a new folder called R in your My Documents folder and install R there instead. You do not need administrator rights to install software in this folder.



Installing R: choose install location

Once the destination folder is selected, the installer will automatically create a new sub-folder with the name of the R version to install and append this to the destination file path. Click next when done.

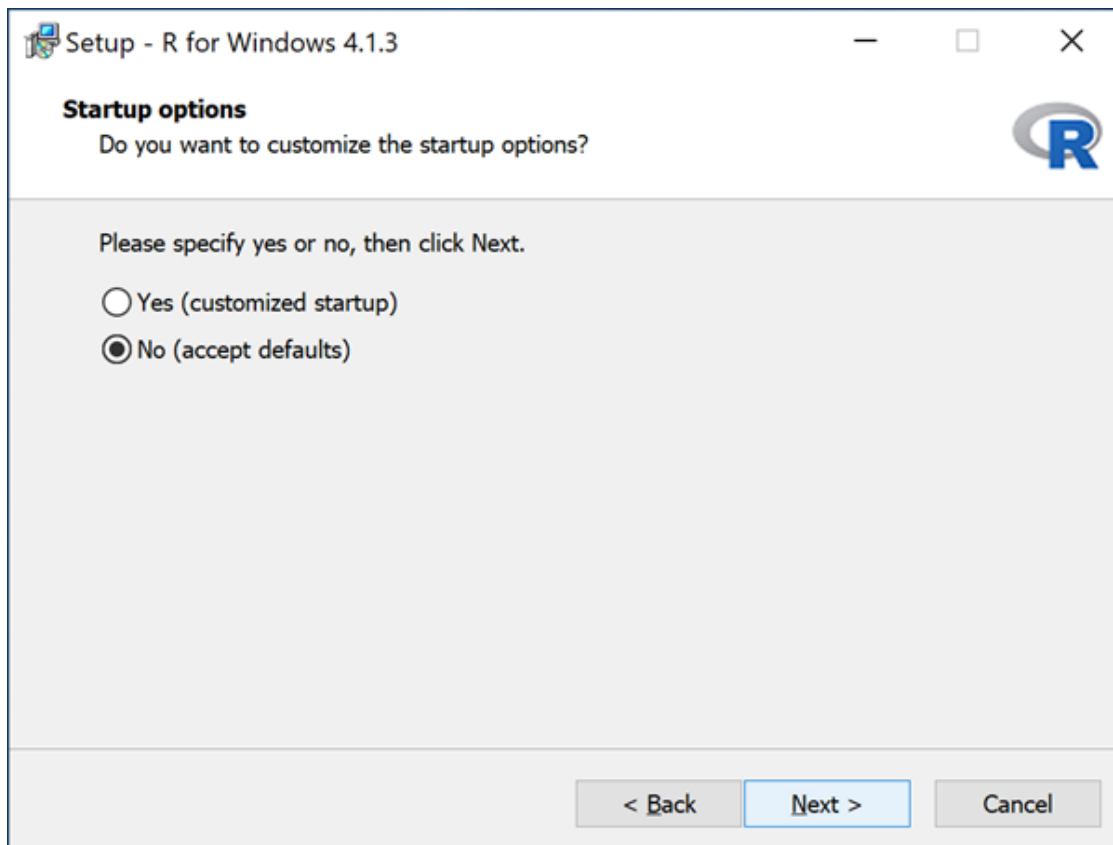
In the next dialogue box you can choose which elements of R you want to install. We suggest leaving this on the default setting (which will install everything):



Installing R: choose what to install

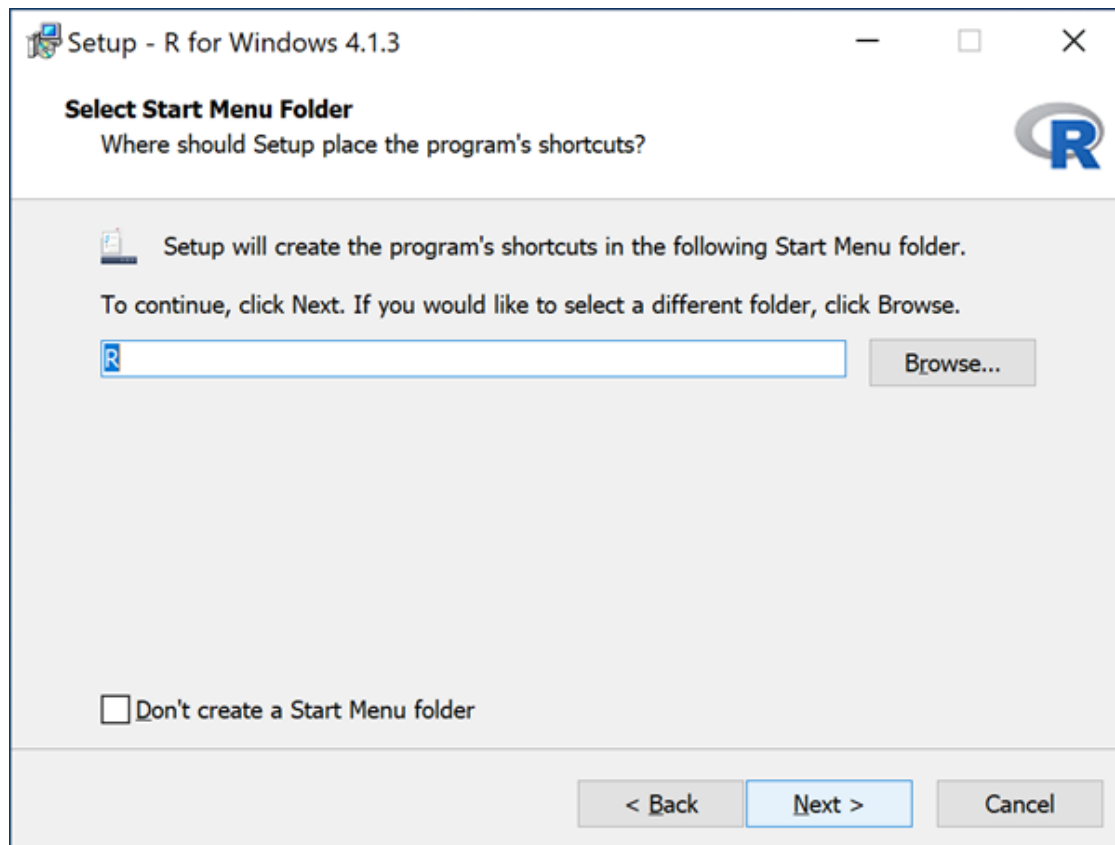
Note that this will install both *32-bit* and *64-bit* versions of R on your system. This is useful because many Microsoft Windows installations come with 32-bit Office programs as default. If you want to query or import data from a 32-bit Microsoft Access database, you will need to do this from 32-bit R. For everything else, 64-bit R is faster and has better memory allocation.

Next, you can choose whether to accept the default startup options, or customize them. Unless you are an advanced user and/or have some specific requirements, we recommend that you select the default setting:



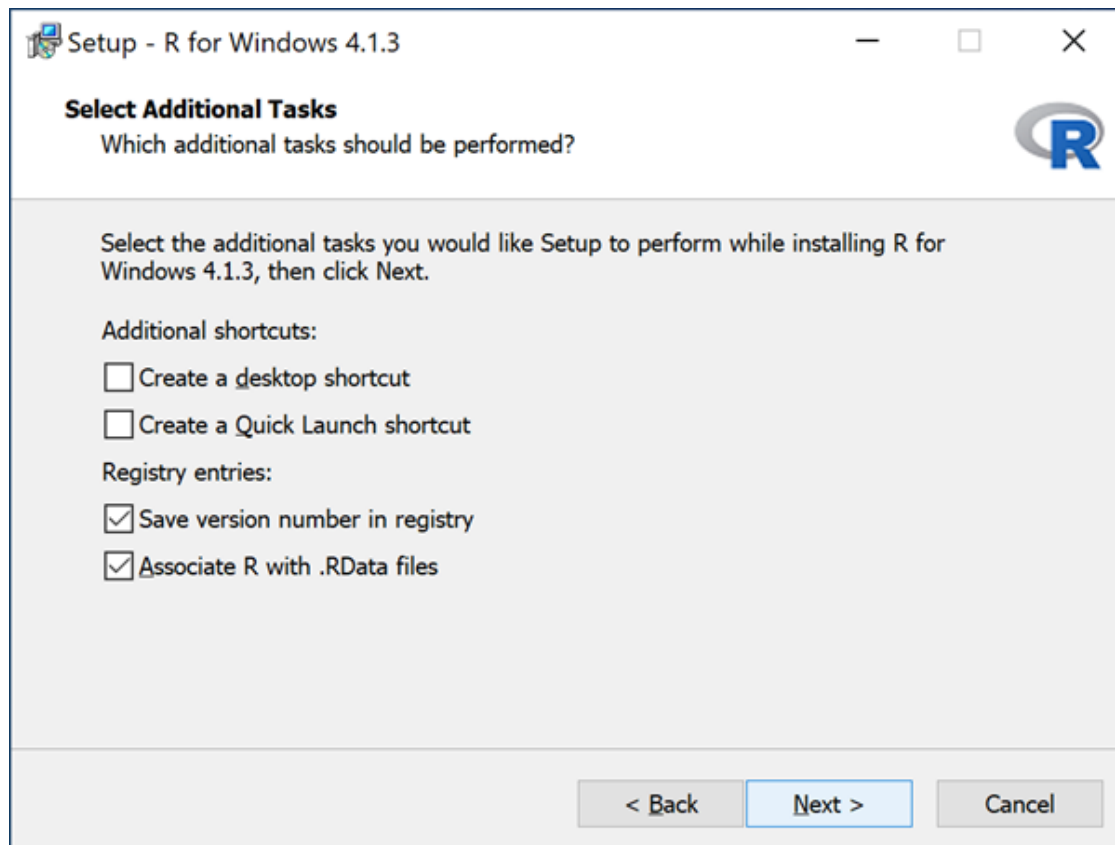
Installing R: choose startup

Next, choose whether you would like to add a shortcut to R in your computer's start menu (note: while we have left this on the default setting - to create a shortcut - this is actually not necessary since most of the time you will be opening R files and interacting with R through RStudio).



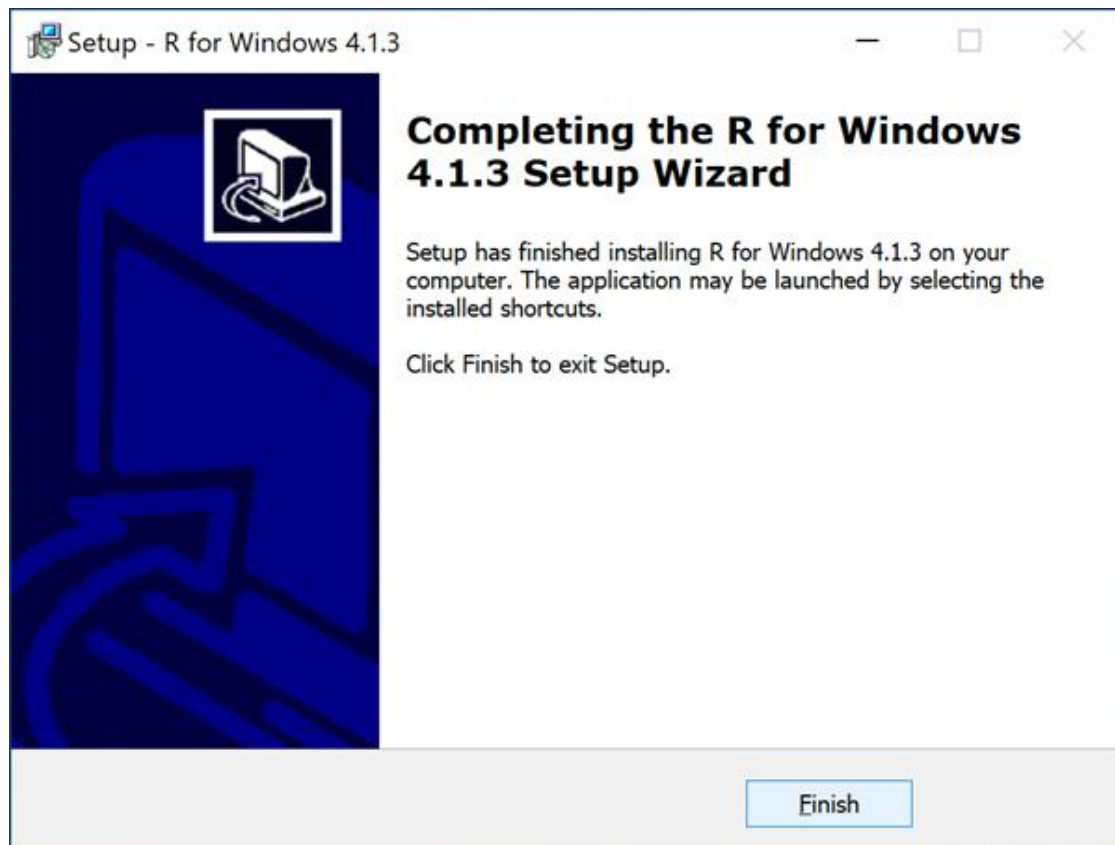
Installing R: add shortcut

Next, you can choose whether to create other shortcuts to R - as above, bear in mind that most interactions with R will be through RStudio, so a shortcut to the R program itself is generally not necessary. This dialogue box will also ask you if you want to save the version number of R in the registry (we recommend you do in case you want to run more than one version of R on the same computer - sometimes this is necessary when upgrading R to check that code still works). We also suggest you associate R with .RData files (these are the files created when you save your work space in R).



Installing R: associate file endings

Click next when done - if the installation has been succesful, this should bring you to the installation complete screen:



Installing R: install complete

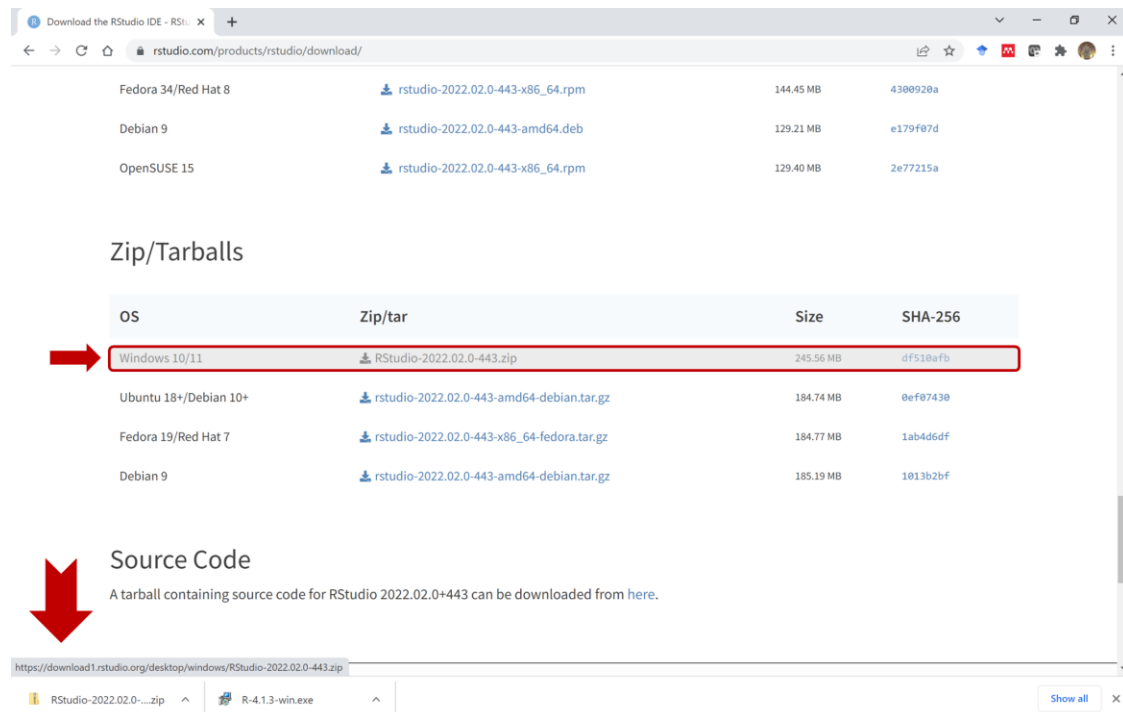
RStudio

Once you have installed R, you will need to download and install RStudio.

Downloading RStudio:

The latest version of RStudio can be downloaded from the [RStudio website](https://www.rstudio.com/). You will note there are several different versions; the free versions appear under the RStudio desktop header.

If you don't have administrator rights to your computer, to avoid difficulties with installation we recommend downloading the pre-compiled zip or tarball versions, which you can find at the bottom of this web page. The relevant version for Windows 10/11 is highlighted in the red box in the figure below:



The screenshot shows the RStudio download page. At the top, there is a table of pre-compiled binaries for various operating systems. Below this, there is a section titled 'Zip/Tarballs' which contains a table of source code packages. The 'Windows 10/11' row in the 'Zip/Tarballs' table is highlighted with a red box, and a red arrow points to it. Below the 'Zip/Tarballs' section, there is a 'Source Code' section with a red arrow pointing to it. The 'Source Code' section contains a link to download a tarball containing source code for RStudio 2022.02.0+443.

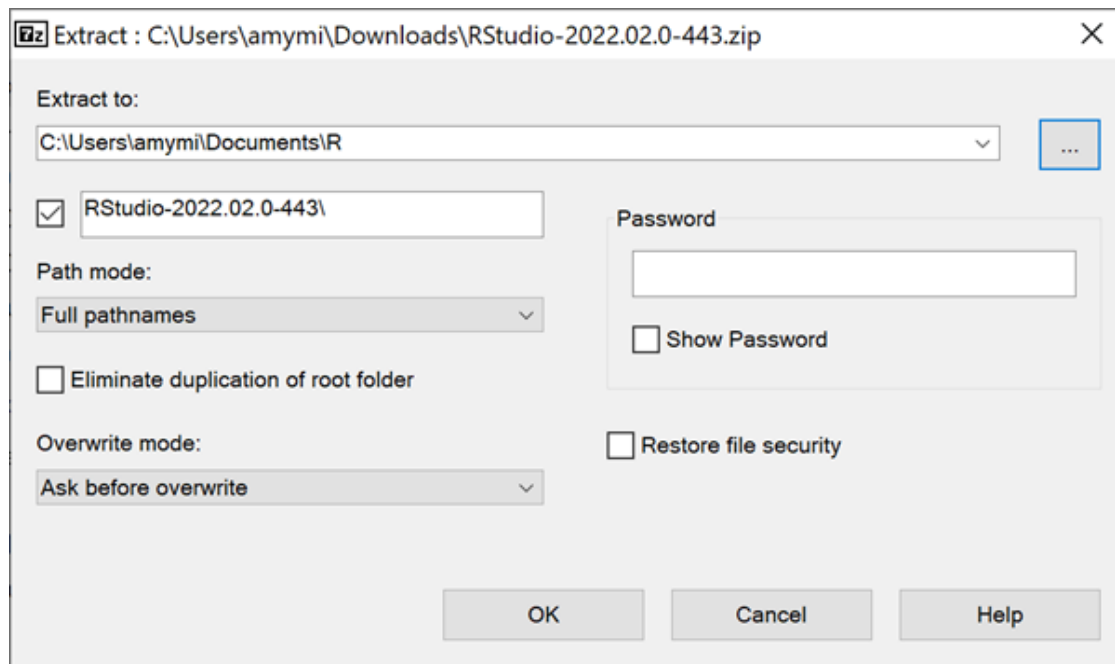
OS	Zip/tar	Size	SHA-256
Windows 10/11	RStudio-2022.02.0-443.zip	245.56 MB	df518afb
Ubuntu 18+/Debian 10+	rstudio-2022.02.0-443-amd64-debian.tar.gz	184.74 MB	0ef07430
Fedora 19/Red Hat 7	rstudio-2022.02.0-443-x86_64-fedora.tar.gz	184.77 MB	1ab4d6df
Debian 9	rstudio-2022.02.0-443-amd64-debian.tar.gz	185.19 MB	1013b2bf

Source Code
A tarball containing source code for RStudio 2022.02.0+443 can be downloaded from [here](#).

Installing RStudio: download

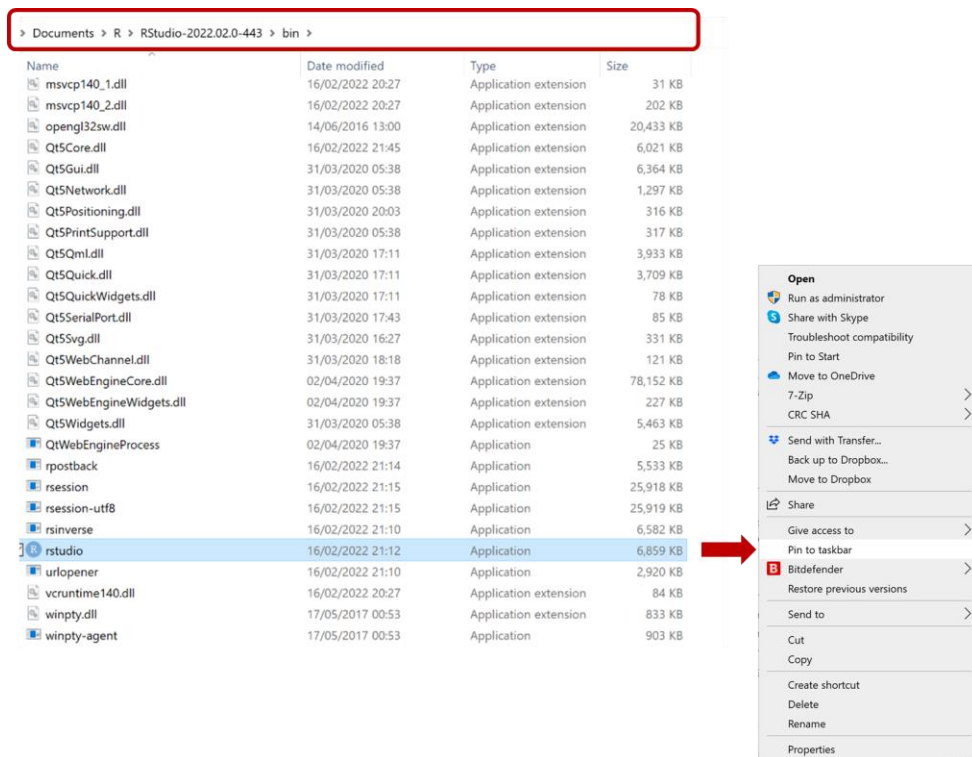
Installing RStudio:

If you are downloading the zip version because you do not have administrator rights, once downloaded you will need to unzip the file, selecting the R folder in My Documents that you created in the previous section as the destination (because you do not have administrator rights). It is important that both R and RStudio are in the same folder so that RStudio can automatically locate R on start-up. RStudio will be unzipped to a folder with the version number in the title, as shown below:



Installing RStudio: unzip

Next, you will need to locate the RStudio executable and create a shortcut to it on your task bar or on the desktop; note that this does not happen automatically when using the zipped version as opposed to an installer. `RStudio.exe` should be found in a sub-folder called `bin` as shown in the figure below:



Installing RStudio: create shortcut

Testing the installations:

Now that both R and RStudio have been installed, you may wish to open RStudio and check that:

1. The shortcut points to the correct (latest) version of RStudio (in case of previous installs)
2. R is located and successfully opened on start-up of RStudio;
3. The correct (latest) version of R is opened (in case of previous installs)

The first and third checks are particularly important if you already have previous versions of R or RStudio on your computer.

At the time of writing (18 March 2022) the latest versions are:

- **R:** v 4.1.3 (22-03-10) “One Push-up”
- **RStudio:** 2022.02.0 Build 443

Open RStudio

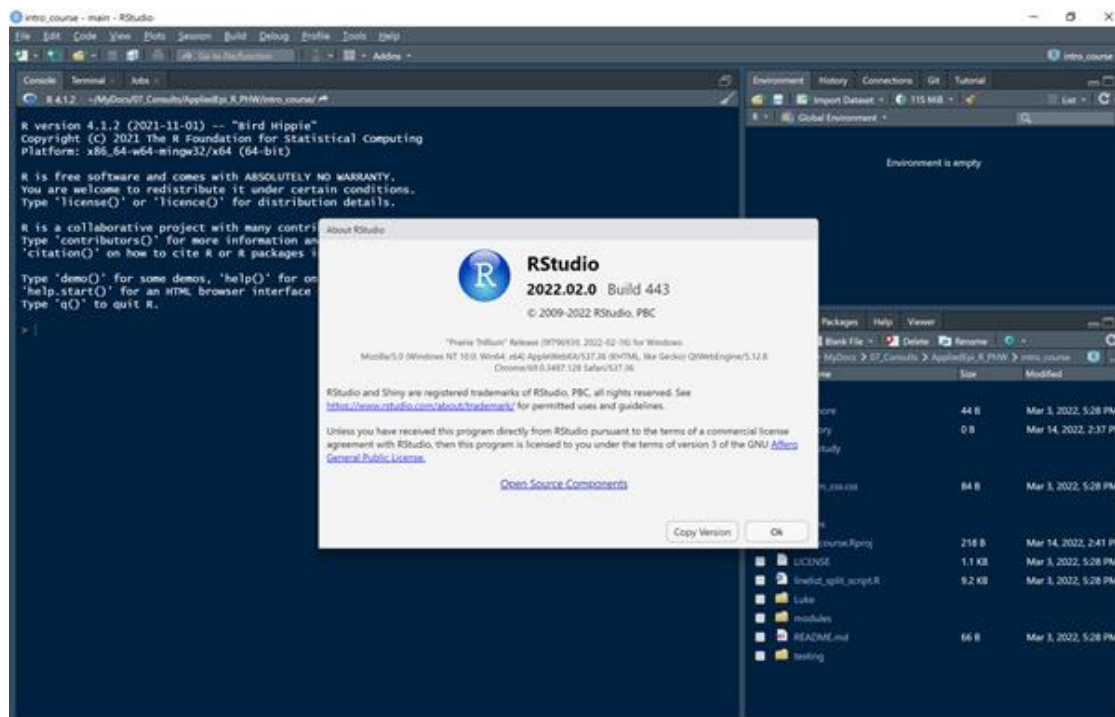
Open RStudio. Note that you must open RStudio and not R. If the setup has gone correctly, opening RStudio will automatically activate R. RStudio is a more friendly user interface for R.

Check the R version

Look at the right side of RStudio, at the R “Console” - the R version will be printed in here start-up.

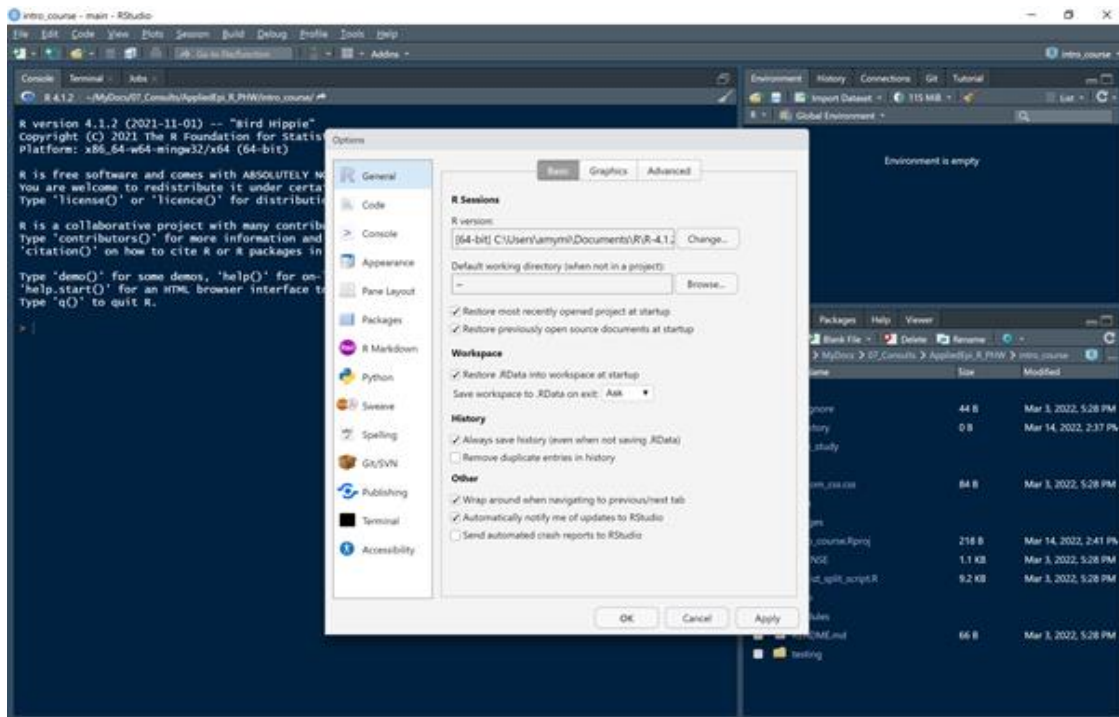
Check the RStudio version

To check the RStudio version you will need to click on the Help menu at the top of the RStudio window and then select About RStudio. This will bring up a display box with full details of the RStudio version as shown below:



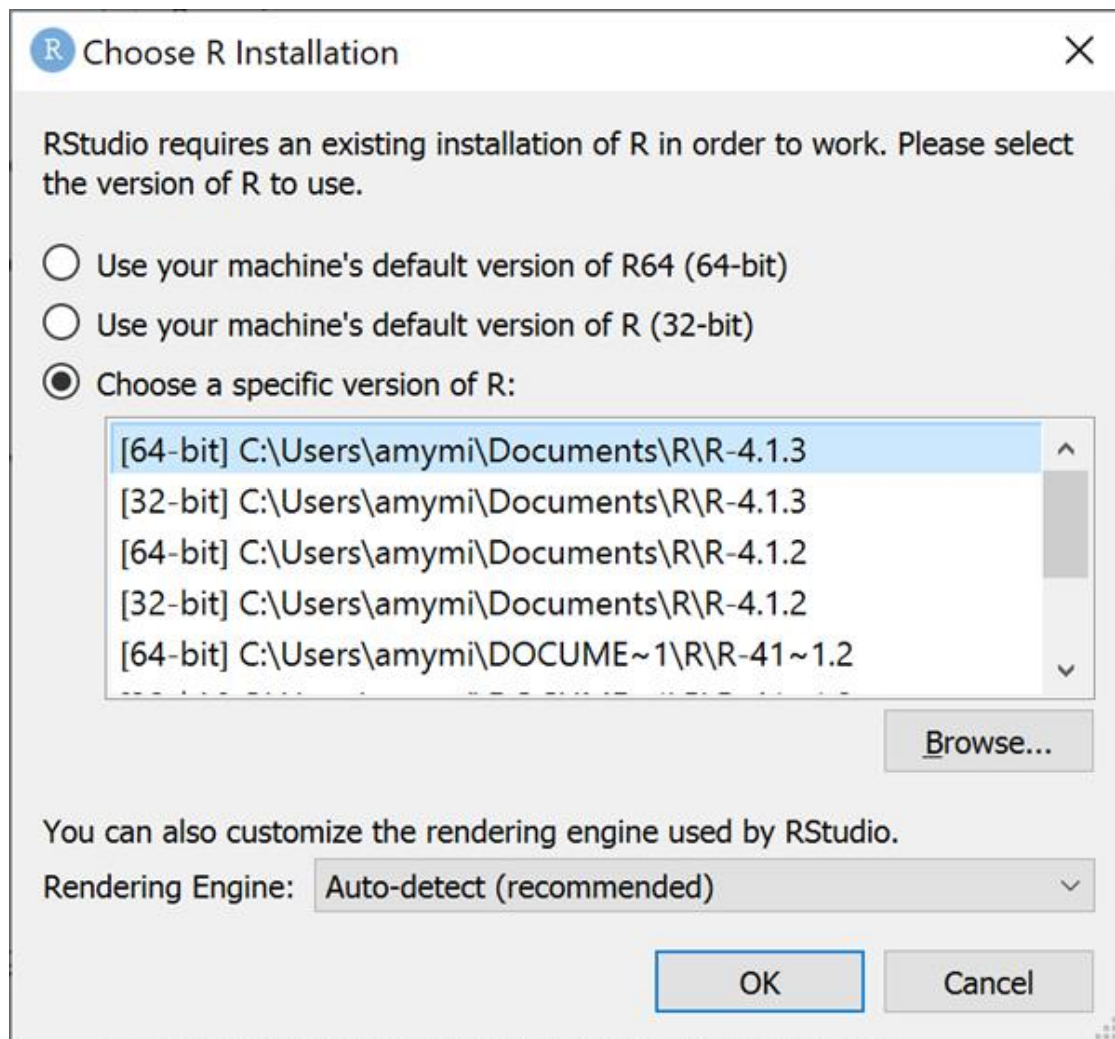
Installing RStudio: check versions

IF you find that RStudio is connected to an older version of R, you can change this by going to the Tools menu at the top of the RStudio window and clicking on Global options (at the bottom). This will bring up a dialogue box where you can change different features of RStudio. At the top of the General tab in this dialogue box, you will see the R version, as shown in the image below:



Installing RStudio: change R version

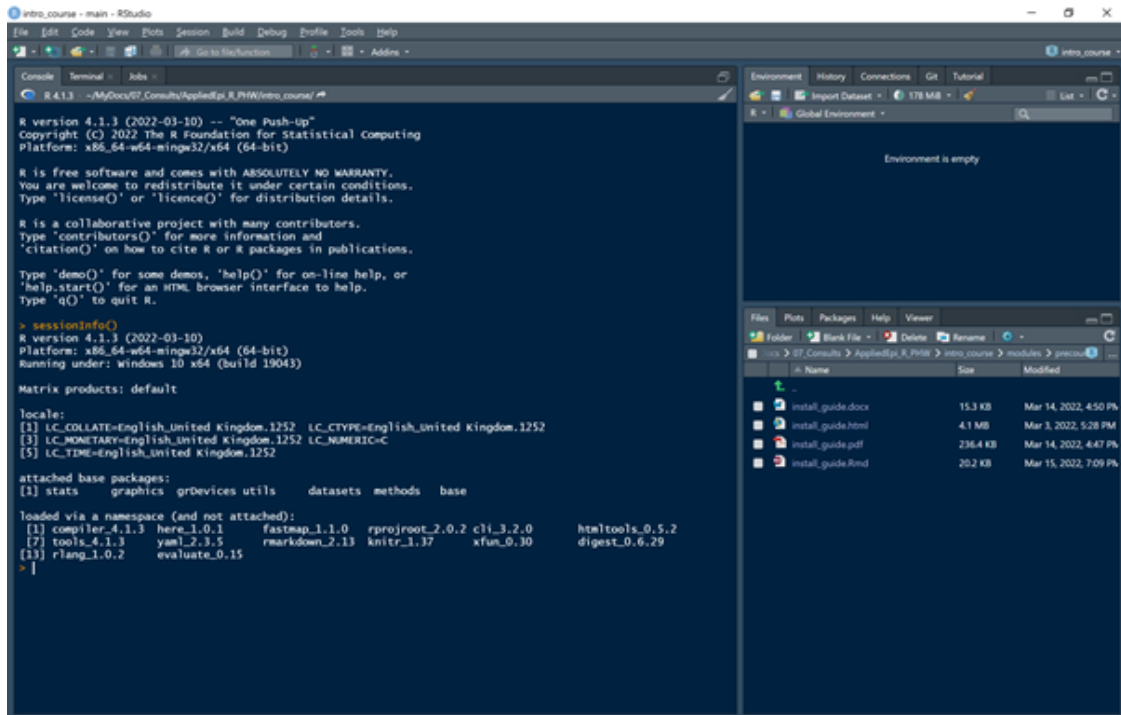
Clicking on the change button next to this will open another dialogue box that lists all the R versions that RStudio has been able to locate. If you don't see the desired R version in the list, this may mean it was installed in a different location to RStudio; in this case you can click on the browse button to manually locate the RStudio.exe file.



Installing RStudio: choose other R version

You will also note that for Windows operating systems, both 32-bit and 64-bit versions of R are listed; we would recommend you choose the 64-bit version and only change to the 32-bit version if you need to interact with 32-bit Microsoft Access databases.

Finally, another way of checking your version of R, the operating system, the language locale and any R packages that you have loaded is to run the `sessionInfo()` command in the RStudio console and look at the output:



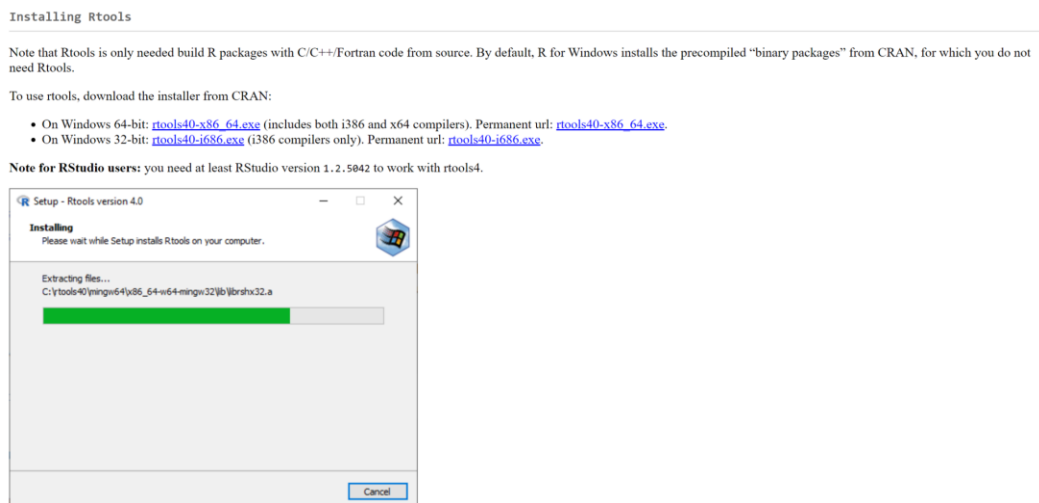
Installing RStudio: sessionInfo

Rtools:

Rtools is a separate piece of software for Windows machines which allows users to create and compile packages. It is useful to have installed even as a beginner, because sometimes the latest versions of packages on CRAN have not yet been bundled into a format that can be easily installed on a windows operating system. Sometimes R users will create packages and put them on the public repository GitHub - again these are unlikely to have been converted to a windows-friendly format, so Rtools will be needed to install them.

Once installed in the correct location, Rtools is automatically called by R when needed, so you do not need to interact with it directly.

Rtools can be downloaded from the [CRAN website](#) by clicking on the Rtools link on this web page:



Installing Rtools

Double-click on the installer to run it and as before, we recommend you choose the R folder in My Documents (where you also installed R and RStudio) as the install location, when prompted.

R Packages

As mentioned in the introduction, the R software that you just installed comes with some basic packages containing functions that nearly everyone needs to use. However, there are many more packages available in public repositories such as the Comprehensive R Archive Network (CRAN), Bioconductor and GitHub.

During this course, we will be using some of these additional packages, which have functions specific to data cleaning, analysis or imaging tasks.

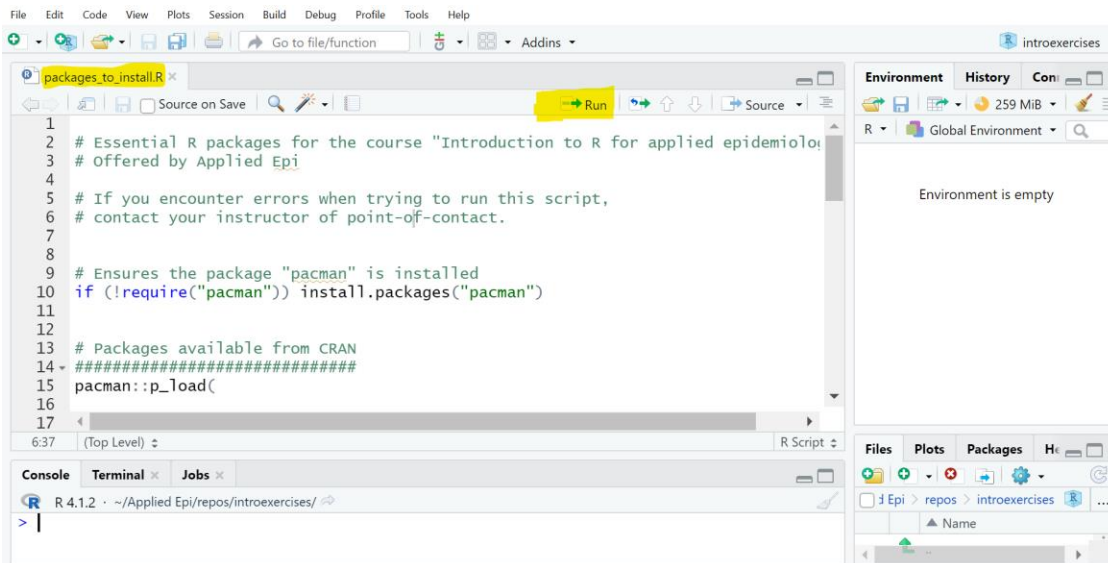
The packages can be installed from within RStudio. For convenience, we have created an R script which will install all the necessary packages when you run it.

First download the script from the email in which you received it. Save it to an easily findable location, such as the “R” folder, or your desktop.

To run the script, you will first need to open RStudio. Within RStudio, click on **File** in the menu at the top of RStudio and select **Open File . . .** This will bring up a dialogue box where you can browse and select that R script from where you saved it, entitled `packages_to_install.R`.

This file will open in a window within RStudio (normally located on the top left). You will see a number of buttons along the top of this new tab; on the top right-hand side there is a button called **Run** with a downward-facing arrow next to it. Clicking on the arrow will reveal a drop-down menu; if you select the last option on this menu (**Run all**) it will run the whole script and install all the necessary packages.

If you do not see the drop-down menu and “Run All”, or need an alternative, you can also run the script by highlighting **ALL** the text in the script, then either click “Run” or press **Ctrl + enter** (at the same time) on your keyboard. If you press **Ctrl** and **Enter** keys at the same time, take care not to delete or replace the text in the script!.



To run this script and install the packages successfully, you will need a good internet connection. It will take several minutes and during this process you will see a lot of status messages in the R console (the bottom-left part of RStudio). If you see any prompts in this Console, select options to continue (you cannot cause any damage to R with any of the prompt options). If you see any error messages or encounter other difficulties, please contact the course organizer soon, who will organize a 1-to-1 session to help you complete the package installations.

For convenience, we have also included a copy of the package installation script below. If you are comfortable with installing packages, you can also copy and paste all the text in the script below directly into the R Console in RStudio (the lower-left pane) and press Enter to run it.

```

#####
# Essential R packages: Intro to R for Applied Epidemiology #
#####

# This script uses the p_load() function from pacman R package,
# which installs if package is absent, and loads for use if already installed

# If you encounter errors when trying to run this script,
# contact your instructor of point-of-contact.

# Ensures the package "pacman" is installed
if (!require("pacman")) install.packages("pacman")

# Packages available from CRAN
#####
pacman::p_load(

# Learning R

```

```

#####
learnr, # interactive tutorials

# project and file management
#####
here, # file paths relative to R project root folder
rio, # import/export of many types of data

# package install and management
#####
remotes, # install from github

# General data management
#####
tidyverse, # includes many packages for tidy data wrangling and
presentation
epitrix, # Useful epi functions

# statistics
#####
janitor, # tables and data cleaning
gtsummary, # making descriptive and statistical tables

# plots - general
#####
#ggplot2, # included in tidyverse
cowplot, # combining plots
RColorBrewer, # color scales
incidence2, # for epidemic curves
i2extras, # for epidemic curves
gghighlight, # highlight a subset
ggrepel, # smart labels
ggExtra, # fancy plots
scales, # helper functions
apyrmaid, # age and sex pyramids

# routine reports
#####
rmarkdown, # produce PDFs, Word Documents, Powerpoints, and HTML
files

# tables for presentation
#####
knitr, # R Markdown report generation and html tables
flextable) # HTML tables

#####
# Packages required to install external dependencies:

```

```
#####
```

```
# tinytex (required for printing R markdown reports to .pdf):
```

```
install.packages("tinytex")
```

```
tinytex::install_tinytex()
```

```
# webshot (required for printing HTML images to Microsoft Word or .pdf):
```

```
install.packages("webshot")
```

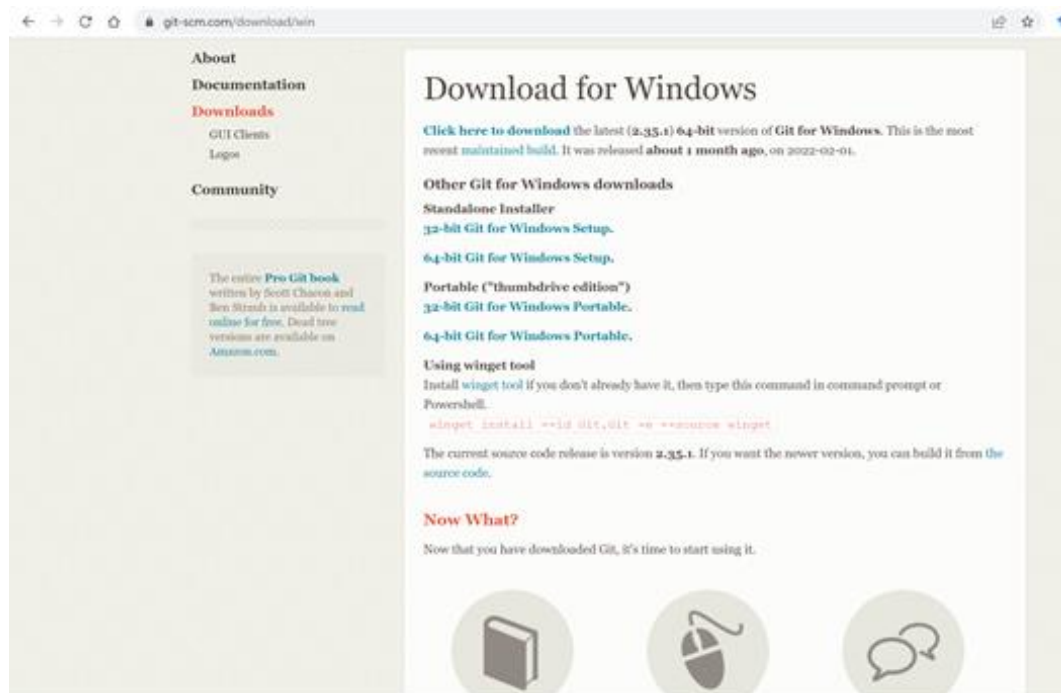
```
webshot::install_phantomjs()
```

```
#####
```

Git

Git and Github Desktop are not necessary for the Intro course. They are version control systems and public (or institutional) repositories hosted on the web where you can store and share R code that you have created. To upload and download code to either of these repositories and make full use of the version control functions, you will need to install the companion software, Git. Although we won't cover the use of Git version control systems in this course, you may wish to use this later on and you can read more about it in the [version control chapter](#) of the Epidemiologist R handbook.

Git can be downloaded from the Git website [here](#). Note there are two options, either an installer (which will require administrator rights) or the thumb-drive (portable) edition. We recommend that if you don't have administrator rights to your computer, you download the thumb-drive edition. As with RStudio, this is a zip file that is ready to use once unzipped.



Installing Rtools

Note that Git needs to interact with any file on your computer that contains code that you want to upload to a repository; for this reason we recommend checking that it is set up correctly with an experienced user before trying it.

Wrap-up

If you have successfully installed all the software in this guide, you are ready to begin and we wish you well on your R learning journey! If you have any difficulties, please don't hesitate to contact the course organizer or designated point of contact and we will organize a session to help you. We do recommend that you attempt installations sufficiently in time before the course, so that we can help troubleshoot if needs be.