R in the Department for Education

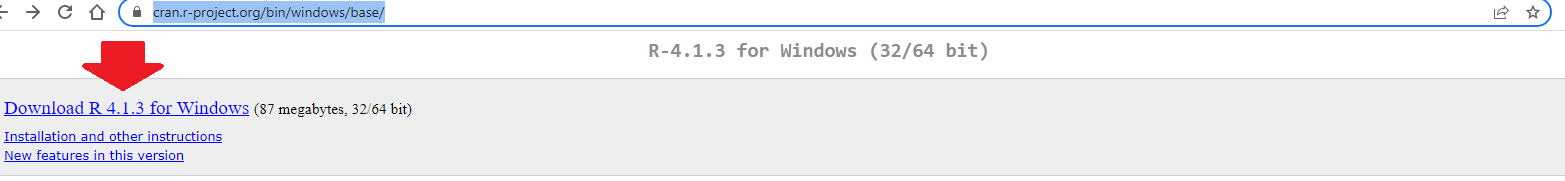
The following can be done *without* admin rights to your computer.

## Get the software

There are two pieces of software needed to use R efficiently: R and RStudio. R is the software that does all the calculating and RStudio is the nice interface through which we use R.

### Download R

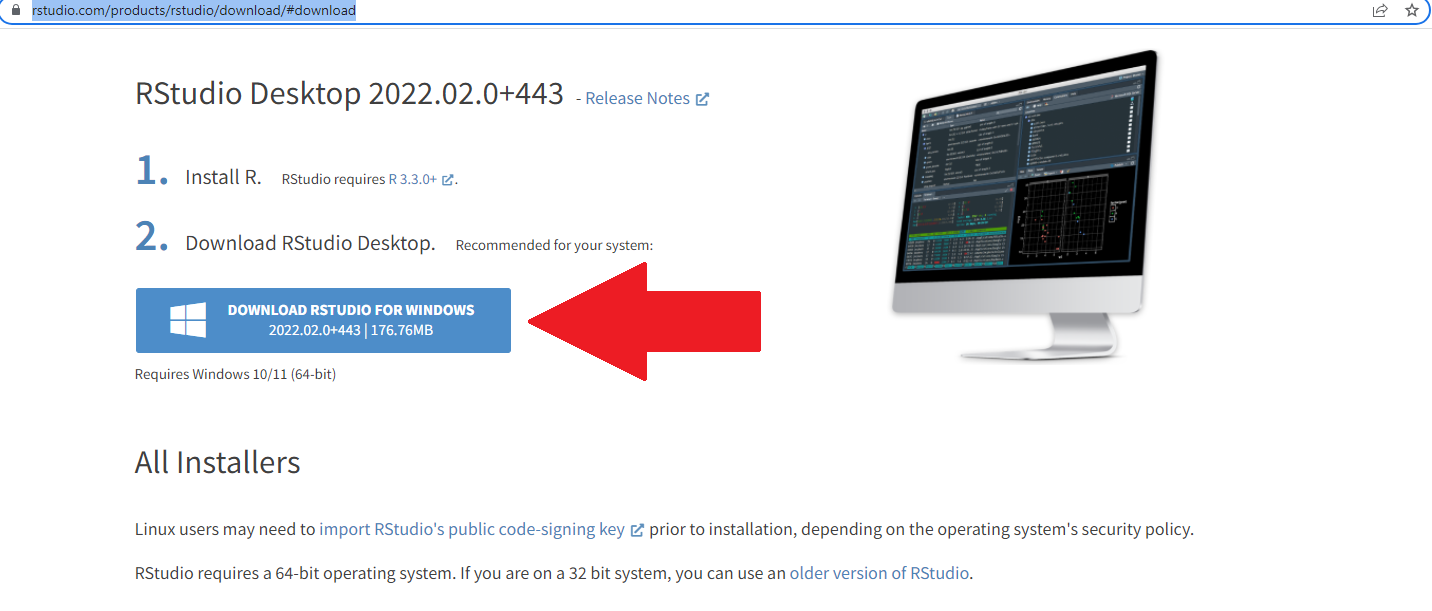
Go to: <https://cran.r-project.org/bin/windows/base/> and click on “Download R for Windows”. Follow the prompts along the way. You won’t be able to install the software unless you already have admin rights but that’s OK for now.



how to install R

### Download RStudio

Go to: <https://www.rstudio.com/products/rstudio/download/#download> and click on “Download RStudio for Windows”. Again, follow the prompts along the way. You won’t be able to install the software unless you already have admin rights but again, that’s OK for now.



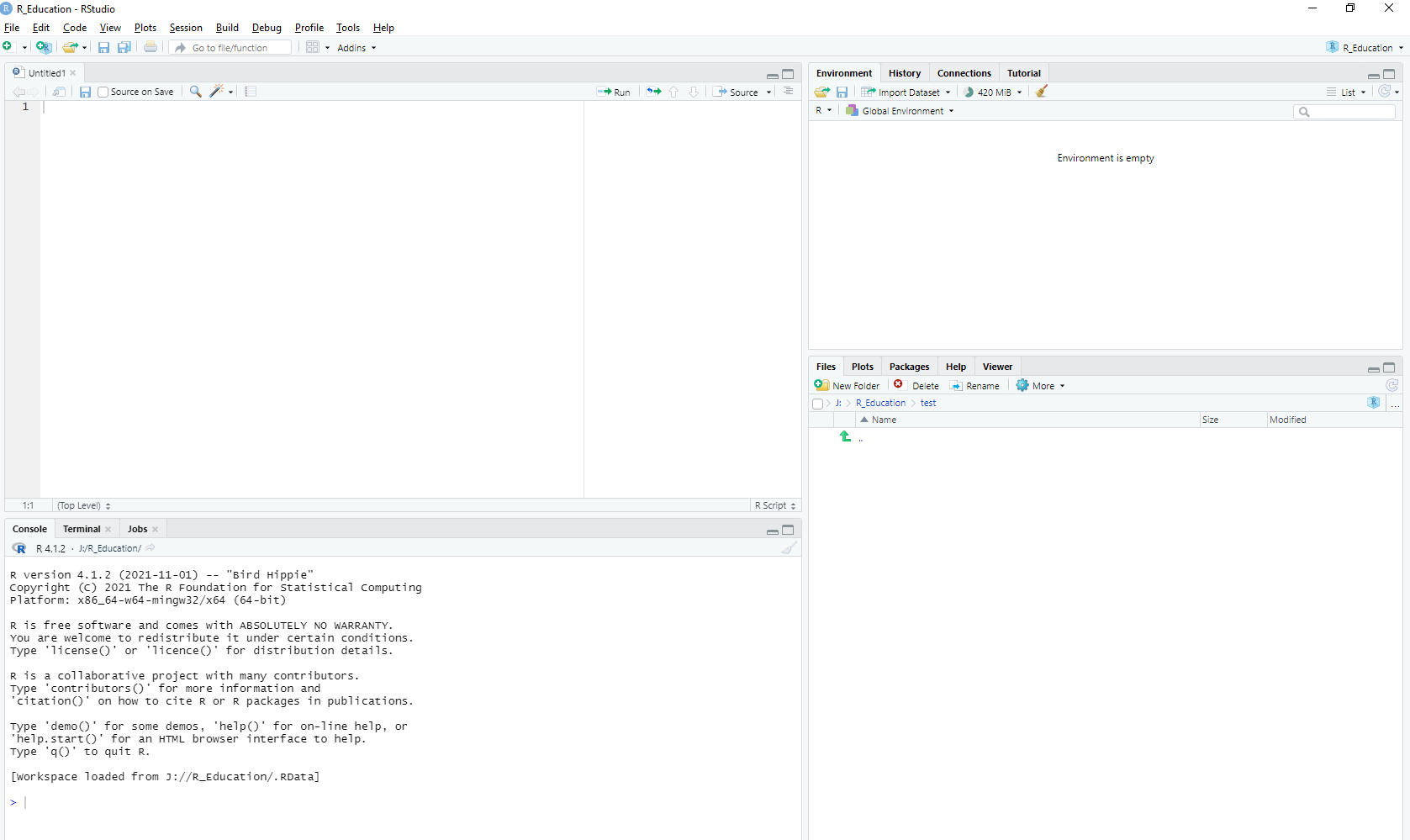
### Ask IT to install R and RStudio

Go to: <https://selfservice.education.sa.gov.au/edit/> and click on “Order something”, “Software”, “Other” then “Non-standard Software”. Fill in the form asking for R and RStudio to be installed.

## Get to work

You’ve just installed two pieces of software but from now on you’ll only interact directly with one: RStudio. So, open RStudio…

### What you see in RStudio



There are four panels. In the bottom-left is a blinking cursor. If you type in a mathematical string and press “Enter” you will get the answer. Give it a try:

1 + 1

## [1] 2

### Get the data

Opening Microsoft SQL Server Management Studio to extract and save data then opening RStudio to analyse it is an inefficient process and makes it harder to keep the documentation of what we did up to date. It makes everything more efficient if we access the databases from within RStudio.

The code we will write is still SQL just like we use in Microsoft SQL Server Management Studio but running it in RStudio has so many advantages.

The way to set this up that I used is:

#### Define a DSN

1. Click the Windows “Start” button and open “Control Panel.”
2. Search for and click on “Administrative Tools” in the list of utilities.
3. Double-click the icon labeled “ODBC Data Sources (64 bit).”
4. Click on the “User DSN” tab and click add.
5. Set the Driver to be “ODBC Driver for SQL Server” with the highest number you have (13, 17 and 18 are all common on DfE machines)
6. Give it the name and description DRADBPROD (actually you can call it anything you want but it will help if we all call it the same thing then our code will be consistent.
7. In the Server section type: DRADBPROD
8. Click “Next”
9. Select that you want it with active directory authentication.
10. Click “Next”
11. Click “Next”
12. Put a tick in the box next to “Trust server certificate.”
13. Click “Finish”
14. Click “Test Data Source”. You should get a message ending with “TESTS COMPLETED SUCCESSFULLY!”

#### Establish a connection in RStudio

Back in RStudio need a few packages. Pasting these two lines into the bottom panel of RStudio where the blinking cursor is will install them.

install.packages("DBI")  
install.packages("odbc")

Then in the top left panel we can create a connection to the server by pasting in this line, highlighting it and clicking the “Run” button in the ribbon at the top of that top left panel:

con <- DBI::dbConnect(odbc::odbc(), "DRADBPROD", timeout = 10)

In the top right panel we can click the “Connections” tab and see an expandable set of all the databases with their tables and views just like we see in Microsoft SQL Server Management Studio.

#### Get the data we want into an R dataframe

Using that connection called con we can then get data into R from any of the databases on that server. On the next line down in the top left panel you can paste the following code, highlight it and click the “Run” button again.

schools\_data <- DBI::dbGetQuery(con,   
 "SELECT TOP 10   
 \*  
 FROM bi.dwh.dim\_site")

The bit inside the quote marks after dbGetQuery(con, is just SQL such as we would use in Microsoft SQL Server Management Studio. But the advantage of running it inside RStudio is that now we have a dataframe with all the data in R. We can then access this dataframe to conduct analyses and visulaisations in R.

In the top right panel we can click the “Environment” tab and see the object schools\_data. This will launch a view of the dataframe in the top left panel.