

Introduction to Biostatistics Class Guidance

Hiroyoshi Iwata, hiroiwata@g.ecc.u-tokyo.ac.jp

2020/4/9

Contents of the lecture

In this course, students will learn the basics of statistics by using R and RStudio.

The content covered in the lecture is as follows.

1. visualizing data with R (April 17)
2. regression and analysis of variance (April 24)
3. principal component analysis, multidimensional scaling constructs (May 1)
4. hierarchical and non-hierarchical cluster analysis (May 8)

Lecture format

Please download required files before the class

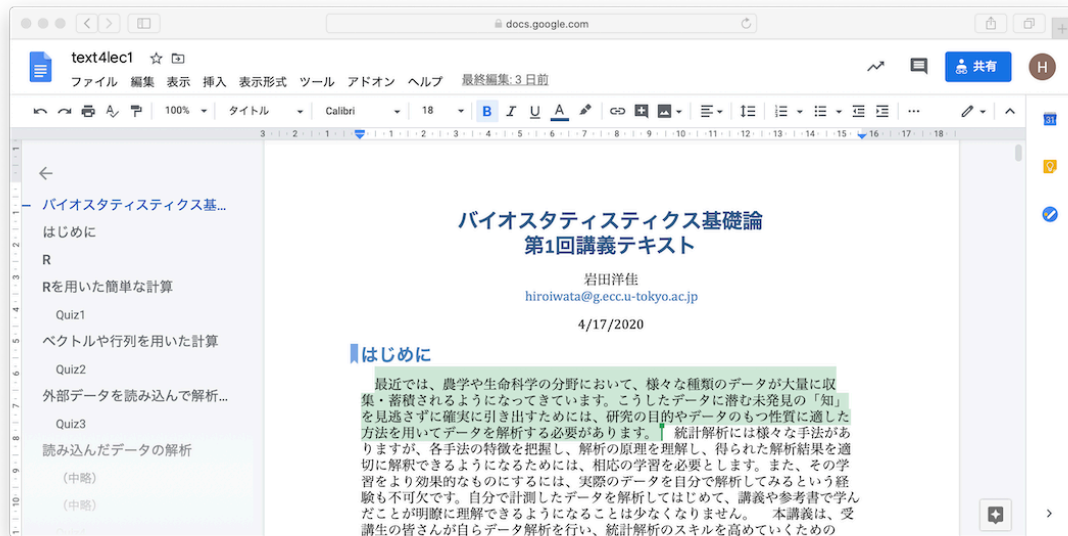
Each lecture is taught in a hands-on format using text (PDF, HTML, Google Docs), R code (R), and data (CSV) files. All files should be available for download from the AgriBio website and ITC-LMS at least one day before the lecture. Please make sure to download the book before the lecture.

The lecture will be provided without video.

In order to reduce the amount of data traffic, this lecture will be provided without video. Because zoom has a chat function, I would like to use you to communicate with me and other students via chat in the lecture. It would be ideal if I could talk like a DJ on a radio show. In addition, it is sometimes difficult for me to have a lecture to you without seeing your faces. Your questions and comments in the chat would be appreciated.

Note that the contents are all the same in all formats: PDF, HTML, and Google Docs. Google Docs has an advantage over the others, since you can see which part I am talking about. Google docs of the next week's lecture (in preparation) can be viewed by clicking the following link. <https://docs.google.com/document/d/1Ozj5aN-AyqXizBKHPSQIDM4--pX61jGemoQw2L6eK0M>.

In Google Docs, you can see which parts I am focusing on (or selecting), as shown below.

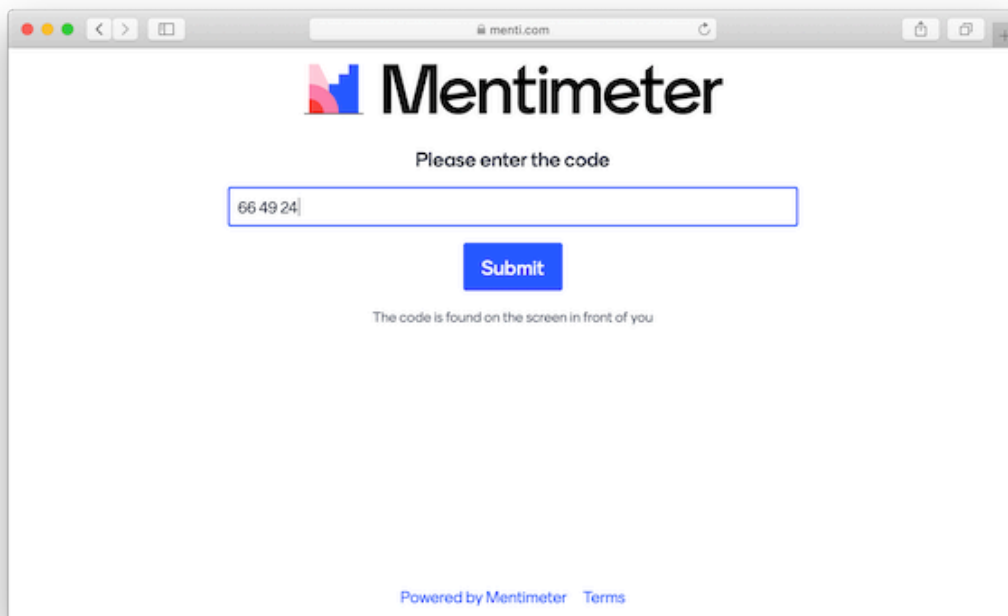


In the lecture, I will focus on the specific place where I am talking about. You can use the R-code in the text and follow my instructions to execute the R-code. In addition, I will explain the meaning of codes and its statistical implications.

There will be quizzes sometimes in the lecture

In order to add a kind of game-like play in the lecture, we will have quizzes in the lecture. Let's do a little practice here.

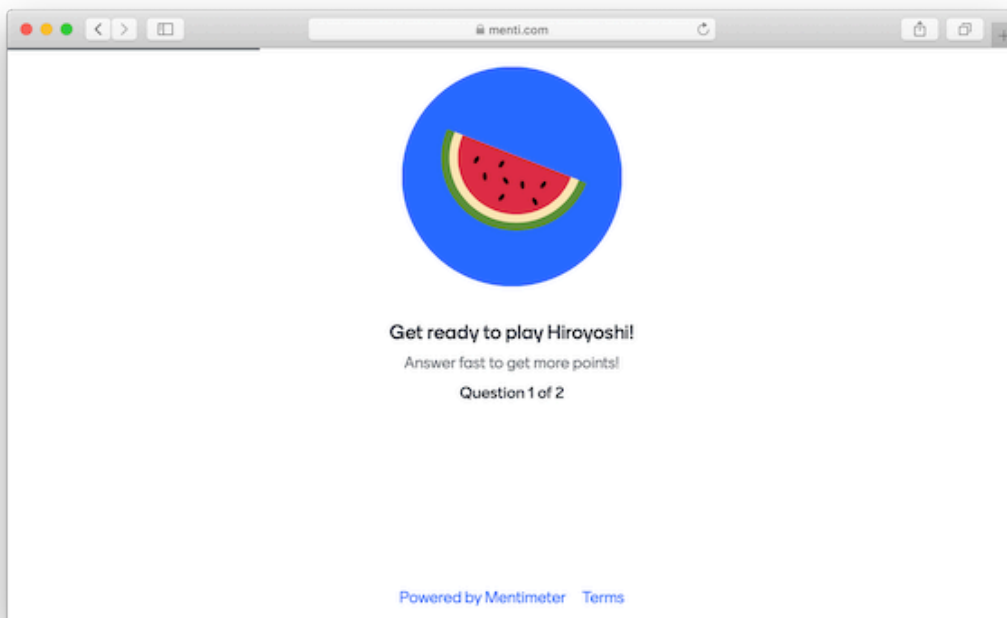
1. First, go to <http://www.menti.com/> and enter the code "66 49 24".



2. Next, enter your nickname and click on "Join quiz".



3. Then it goes into a standby state. After a while, the quiz will start. I wish you all the best in your endeavors.



Please note that these quizzes are just a kind of game and do not affect your grades. Have fun and work on it.

Grades

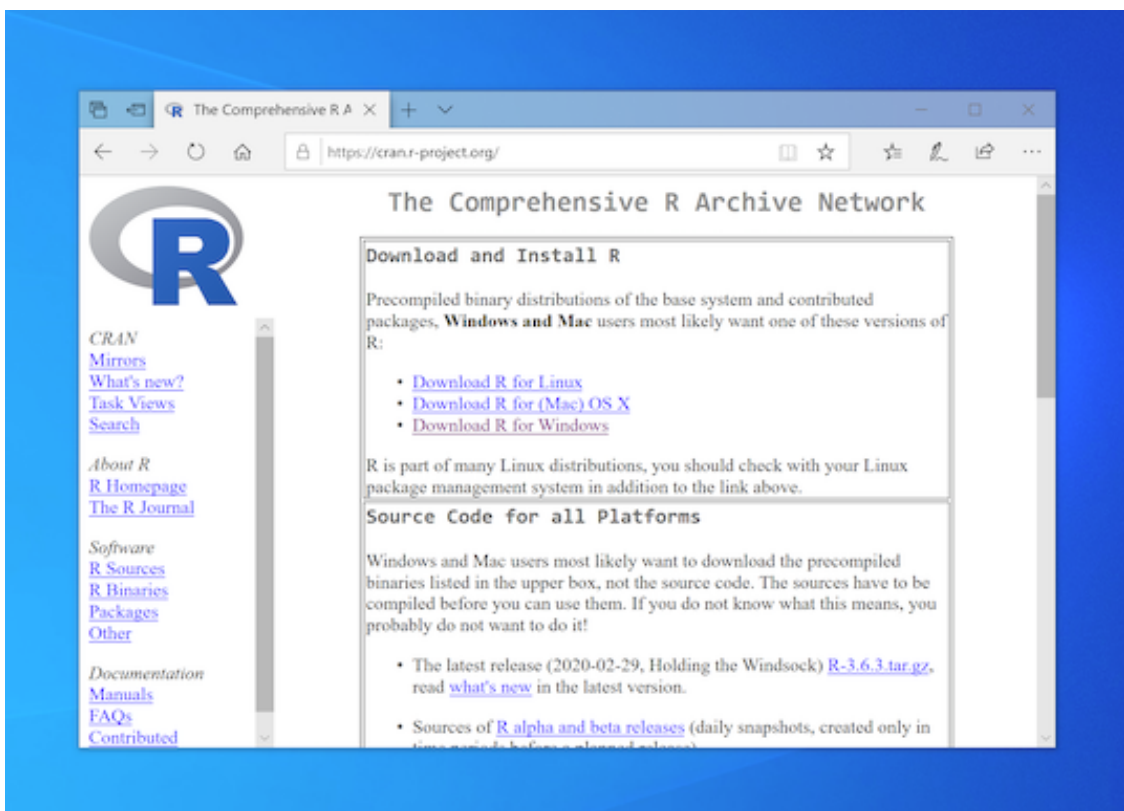
Grades will be based on attendance (which will be taken using the ITC-LMS) and reports (which will be submitted using the ITC-LMS).

Installing R and RStudio

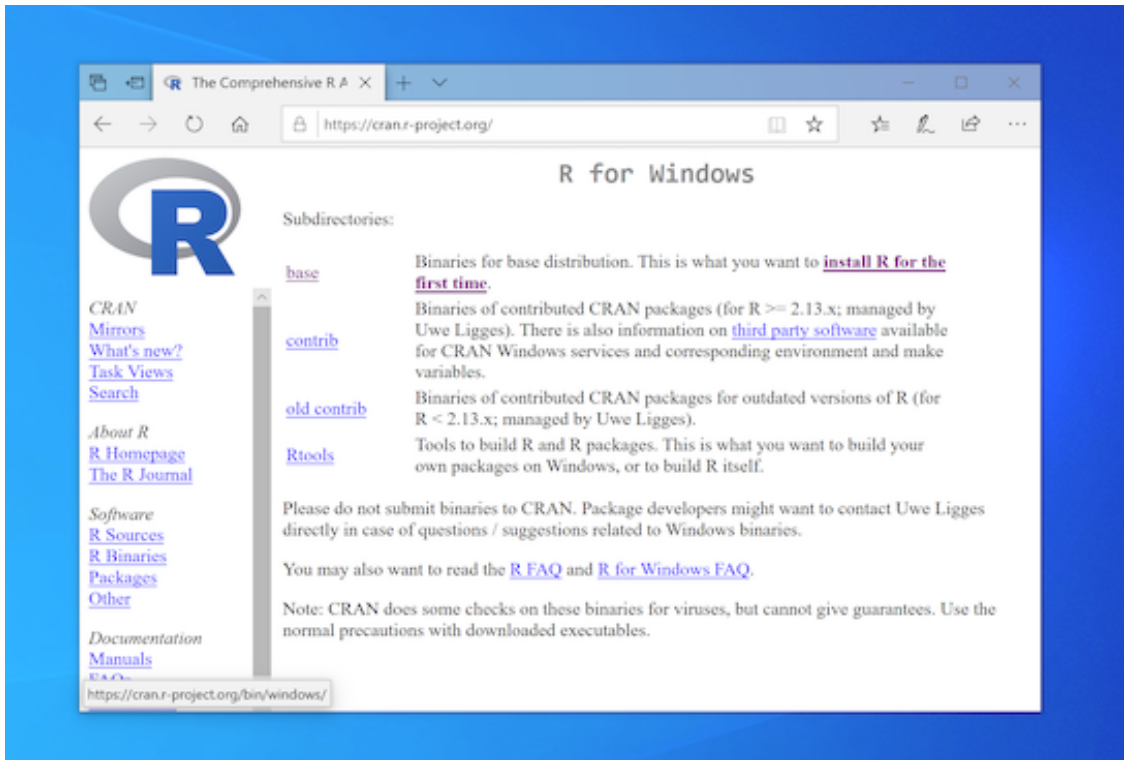
In the case of Windows

Installing R

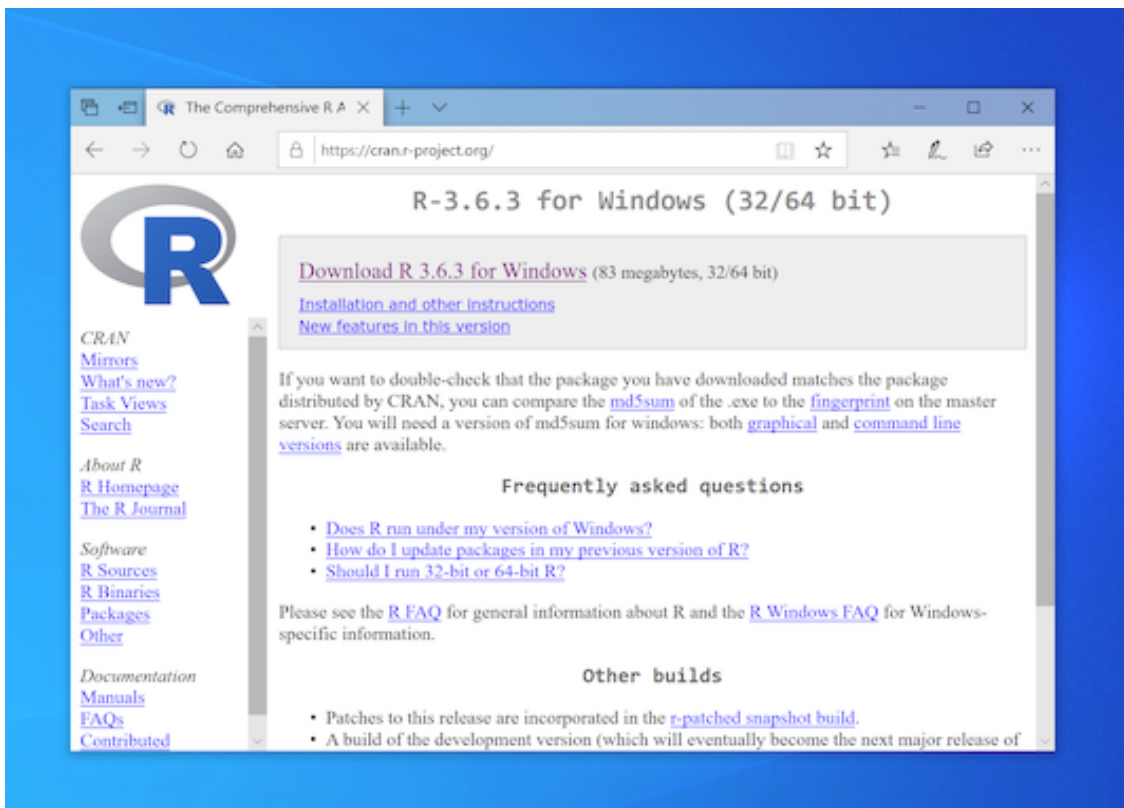
1. First, go to the CRAN site <https://cran.ism.ac.jp/>.
2. Click "Download R for Windows".



3. Select "base".

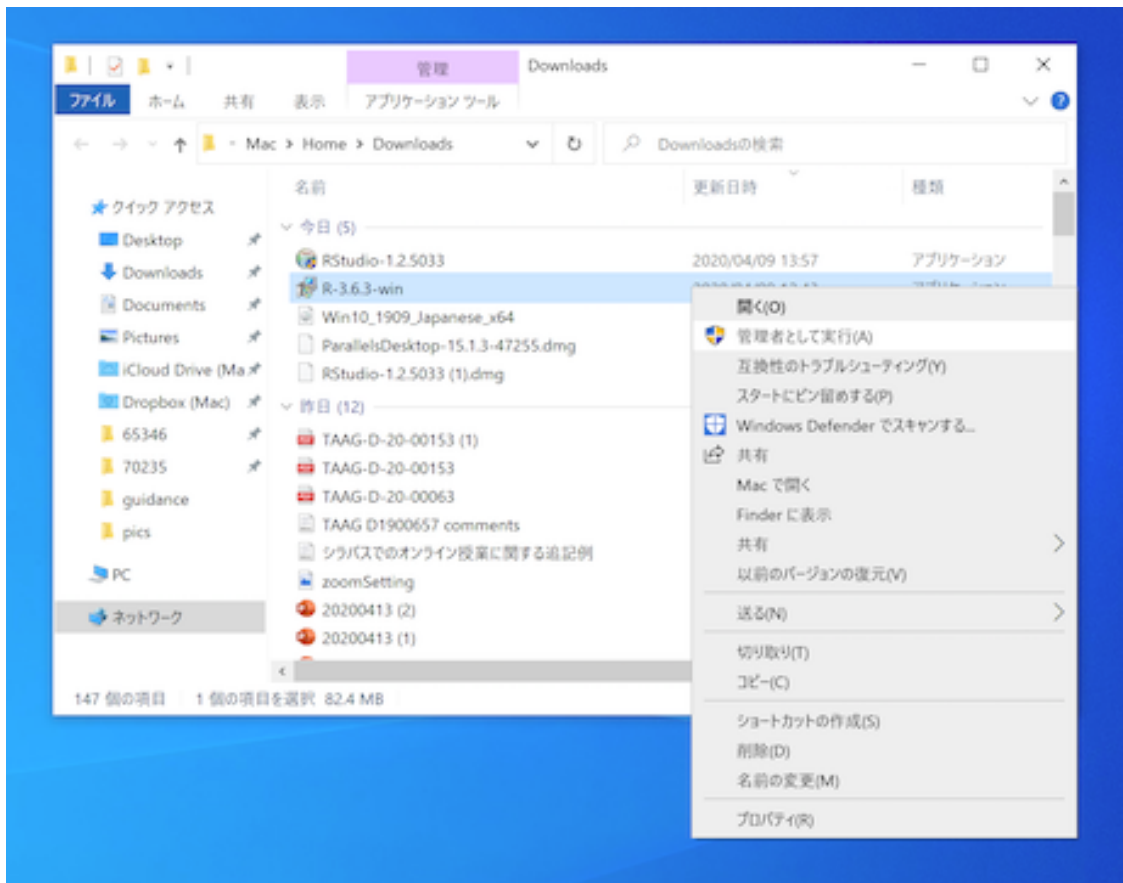


4. Click "Download R 3.6.3 for Windows". The latest version will vary depending on the time.



5. When you download, please "save" instead of "run".

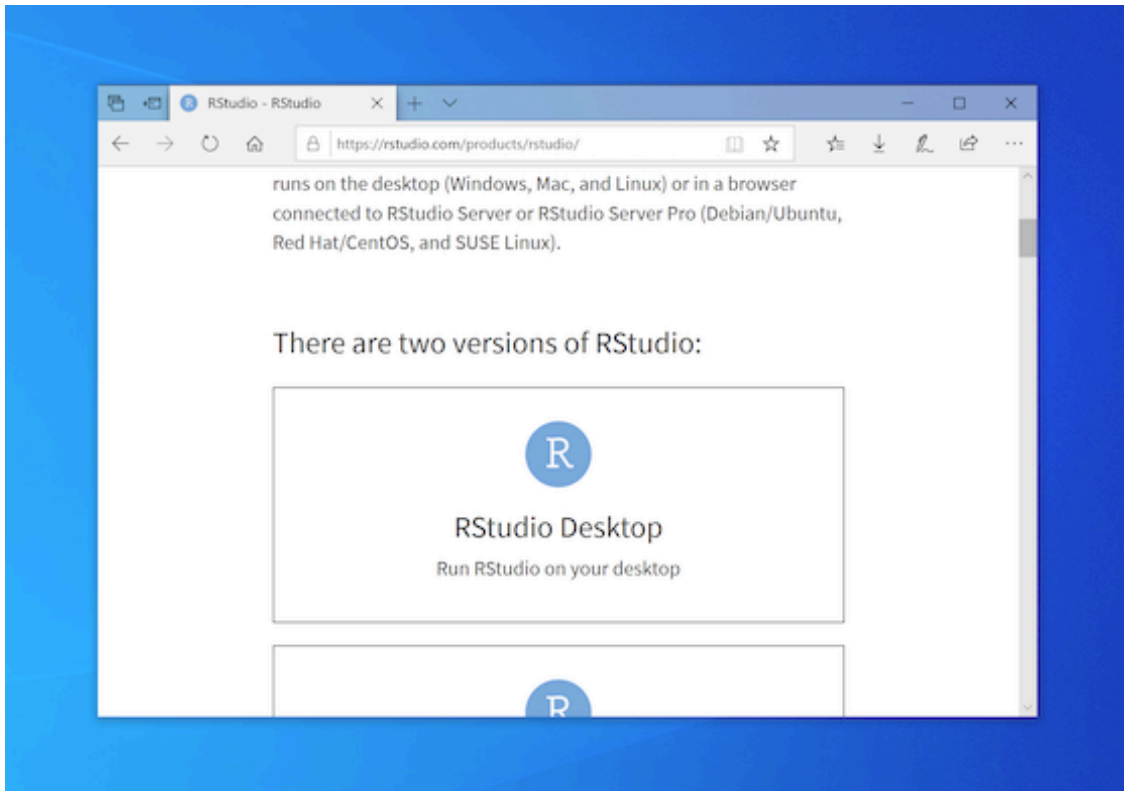
6. Click on the downloaded and file to install R. Please be sure to “Run as an administrator” (install with administrative privileges).



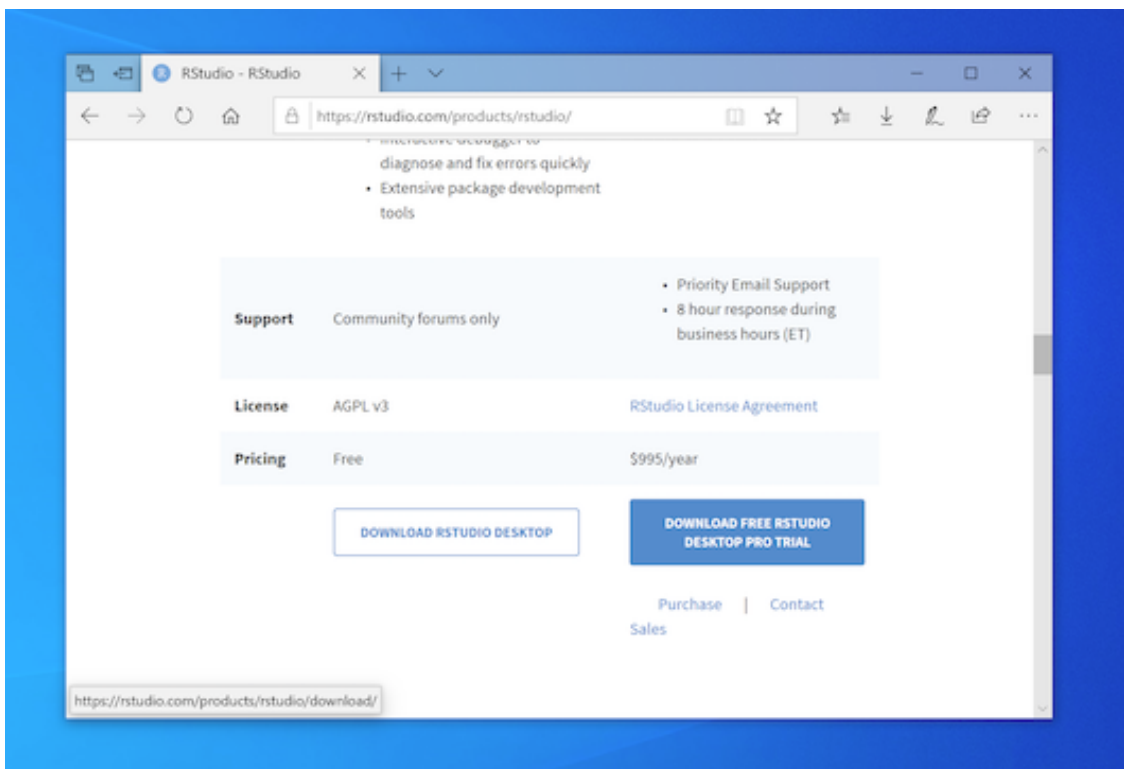
7. You do not need to change any of the installation options. You can proceed with the installation with the default settings.

Installing R Studio

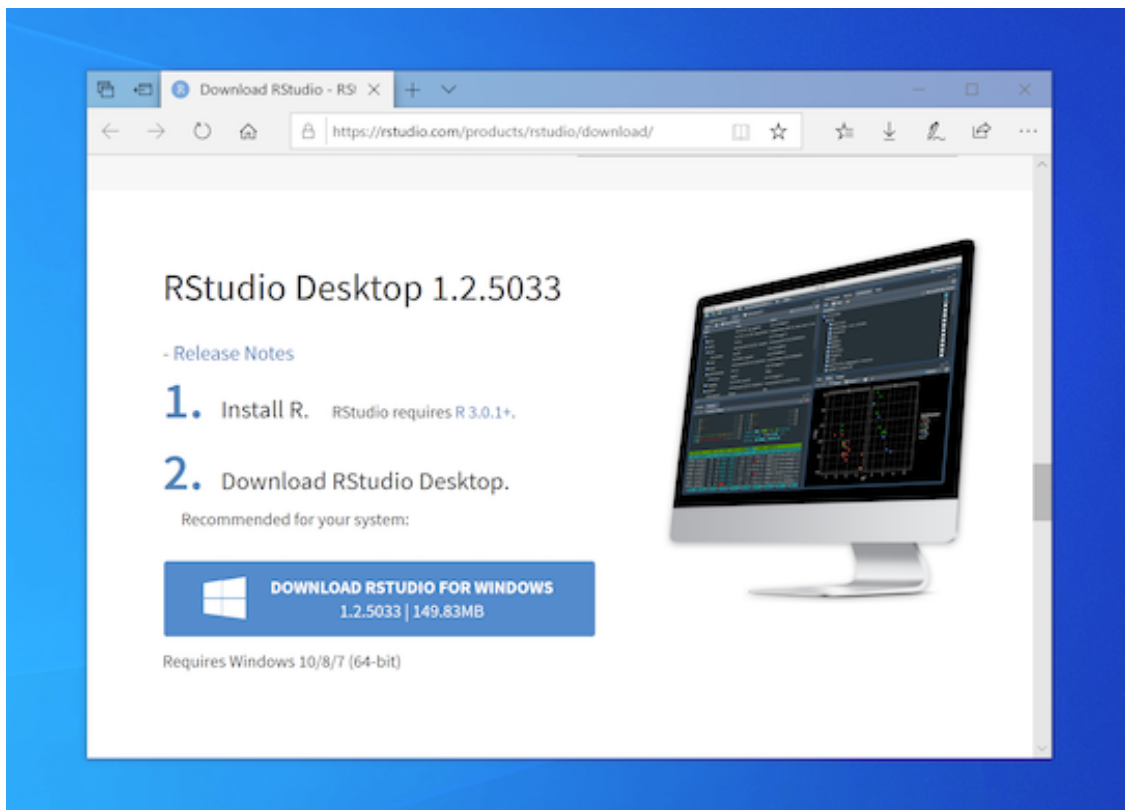
1. Go to the RStudio download site: <https://rstudio.com/products/rstudio/>.
2. Next, click “RStudio Desktop”.



3. Click the "DOWNLOAD RSTUDIO DESKTOP".

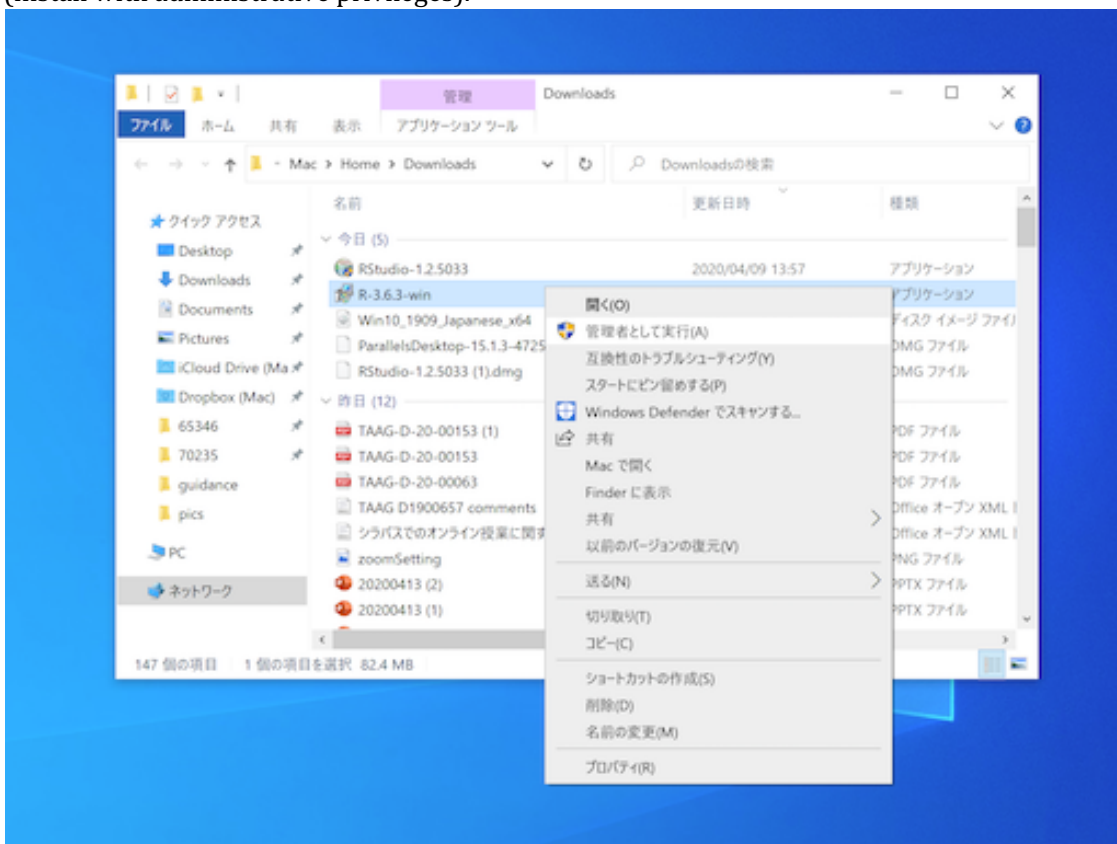


4. Click the "DOWNLOAD" button on the left side (free version of RStudio Desktop).



5. When you download, please "save" instead of "run".

6. Click on the downloaded file to install RStudio. Please be sure to “Run as an administrator” (install with administrative privileges).



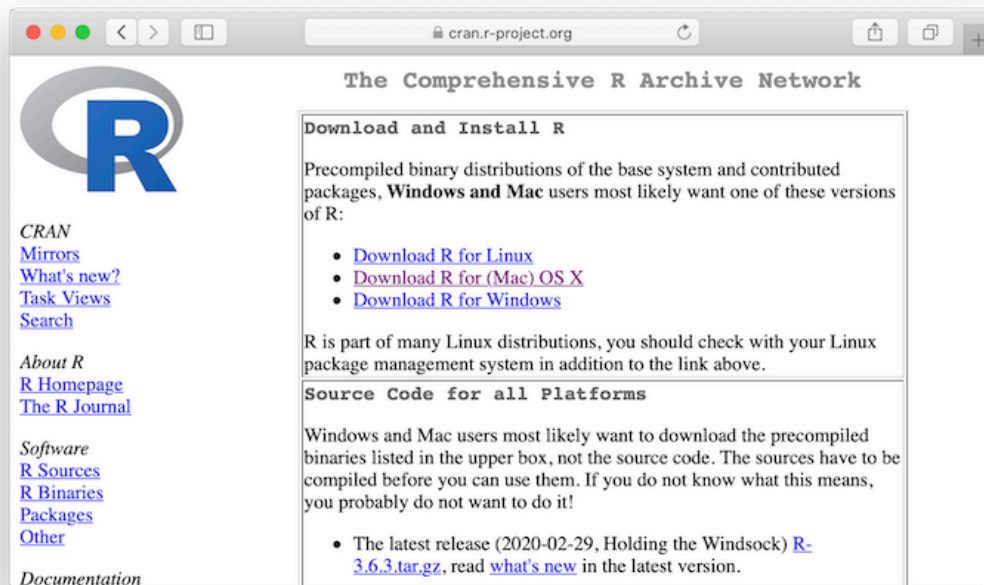
7. You do not need to change any of the installation options. You can proceed with the installation with the default settings.

In the case of Mac OS

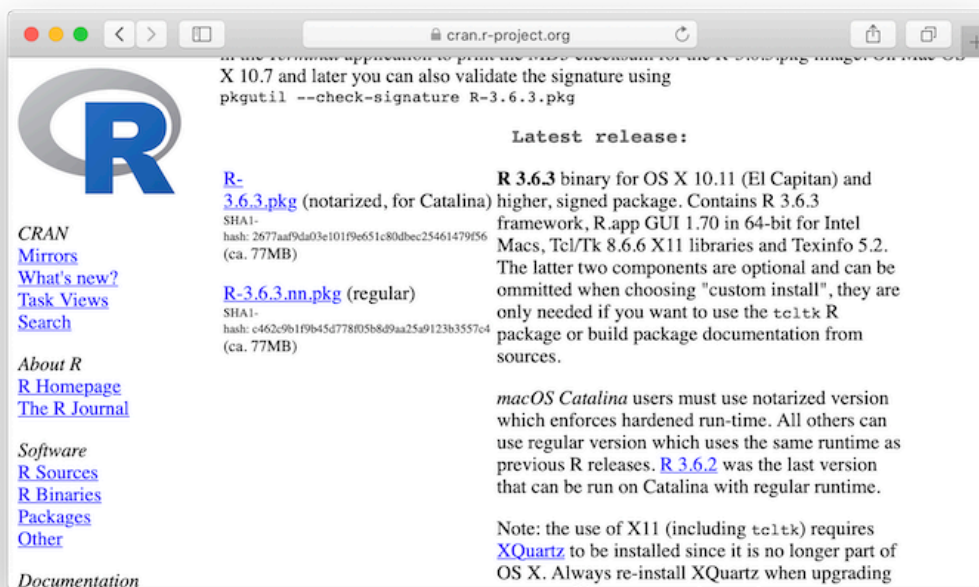
Installing R

1. First, go to the CRAN site <https://cran.ism.ac.jp/>.

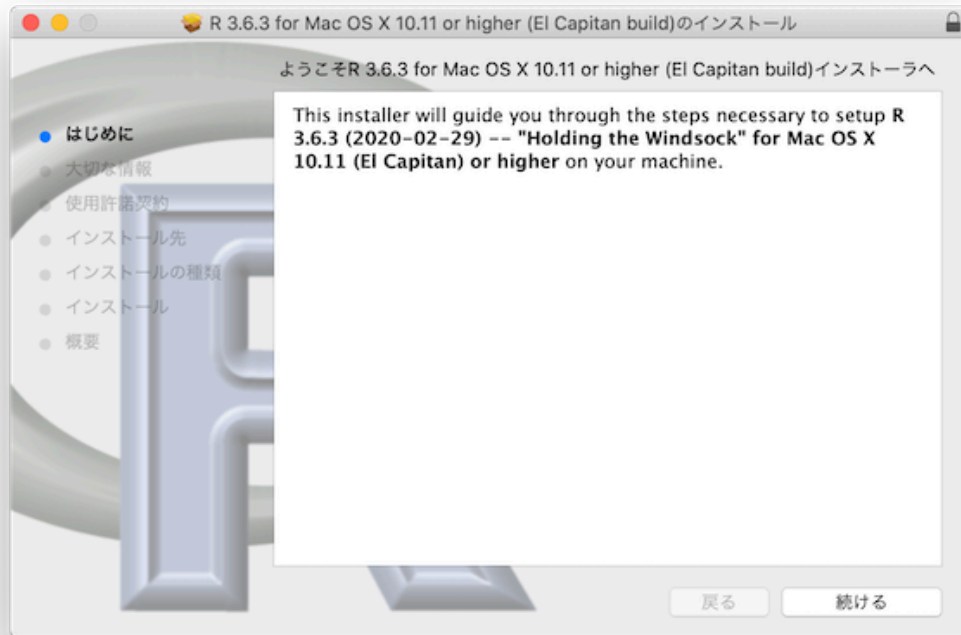
2. Click "Download R for (MAC) OS X".



3. Click "R-3.6.3.pkg". The latest version will vary depending on the time.



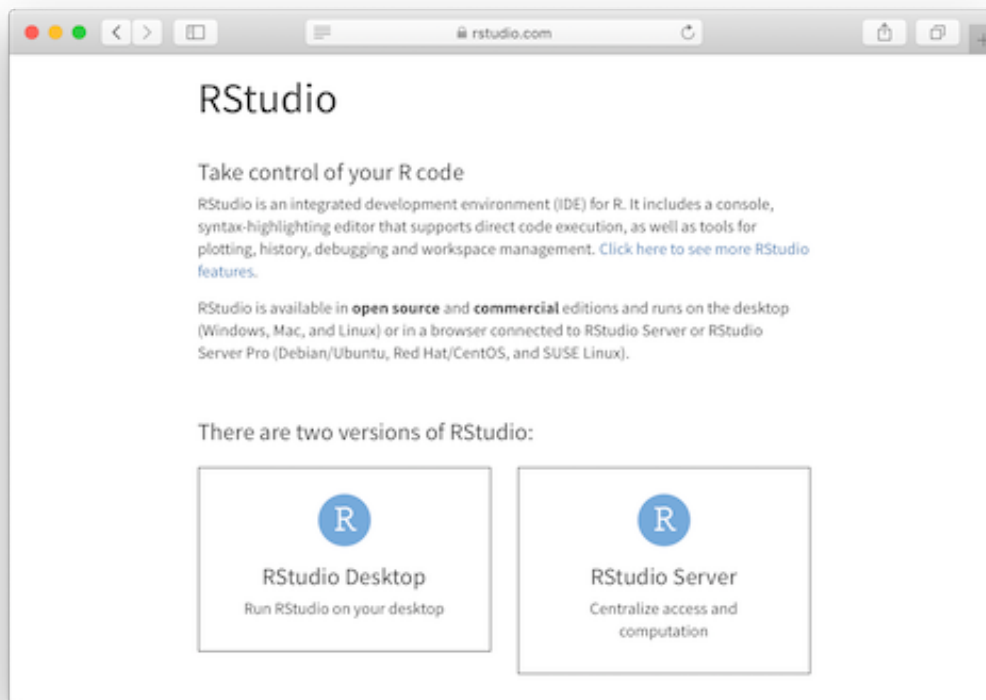
4. Click on the downloaded file to install R.




5. You do not need to change any of the installation options. You can proceed with the installation with the default settings.

Installing R Studio

1. Go to the RStudio download site: <https://rstudio.com/products/rstudio/>.
2. Next, click "RStudio Desktop".

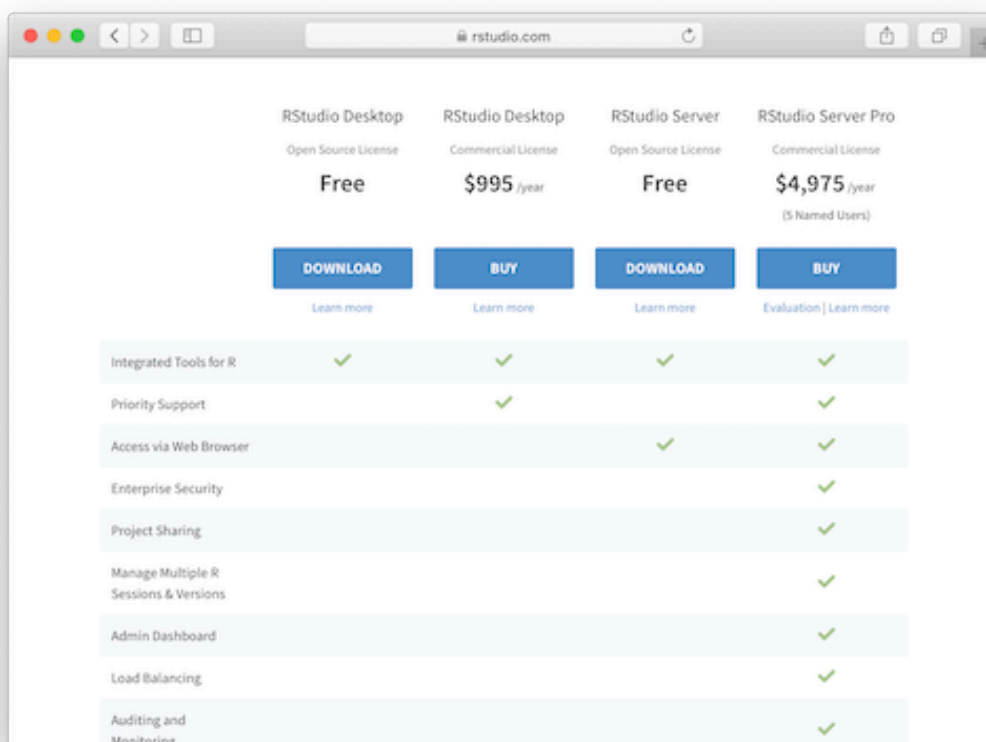


3. Click the "DOWNLOAD RSTUDIO DESKTOP".

 **Studio Desktop**

	Open Source Edition	RStudio Desktop Pro
Overview	<ul style="list-style-type: none"> • Access RStudio locally • Syntax highlighting, code completion, and smart indentation • Execute R code directly from the source editor • Quickly jump to function definitions • Easily manage multiple working directories using projects • Integrated R help and documentation • Interactive debugger to diagnose and fix errors quickly • Extensive package development tools 	All of the features of open source; plus: <ul style="list-style-type: none"> • A commercial license for organizations not able to use AGPL software • Access to priority support • RStudio Professional Drivers
Support	Community forums only	<ul style="list-style-type: none"> • Priority Email Support • 8 hour response during business hours (ET)
License	AGPL v3	RStudio License Agreement
Pricing	Free	\$995/year
	DOWNLOAD RSTUDIO DESKTOP	DOWNLOAD FREE RSTUDIO DESKTOP PRO TRIAL

- Click the “DOWNLOAD” button on the left side (free version of RStudio Desktop).



- When you click on the downloaded file, a window will be displayed as follows. All that's left to do is to drag RStudio and drop it on top of Applications.

!

Trying to use R and RStudio

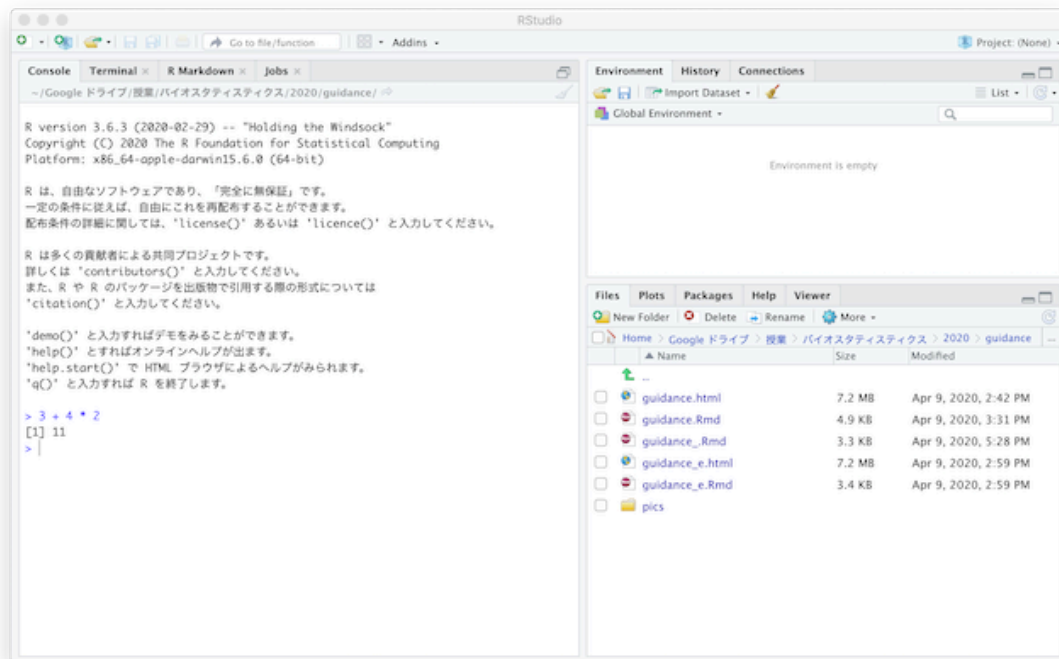
Make sure that R and RStudio are working

From here, I will explain how to use R and RStudio for Windows and Mac OS users together, because RStudio has the same interface in Windows and Mac OS versions.

Let's start by typing the following into the “Console” on the left of the window.

```
3 + 4 * 2
```

```
## [1] 11
```



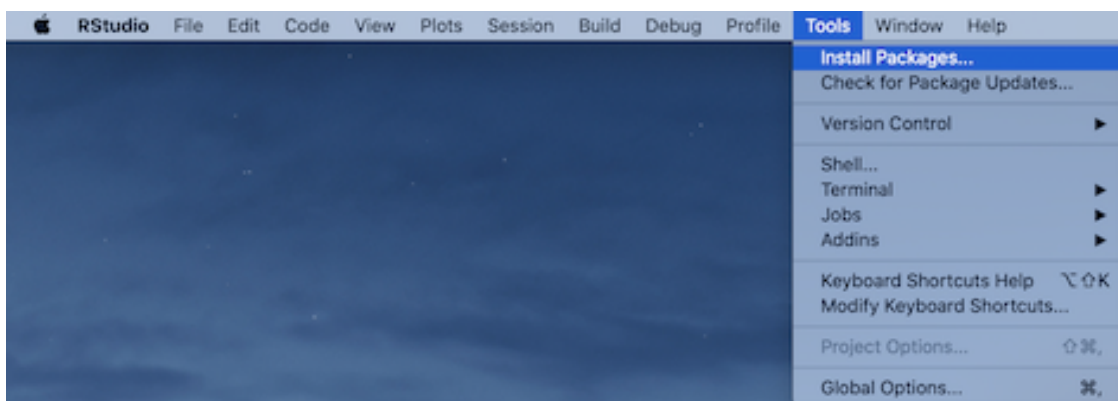
The calculation is performed and the result is shown. I will explain how to use R in detail in next week's lecture.

Preparing to use R and RStudio for the lecture

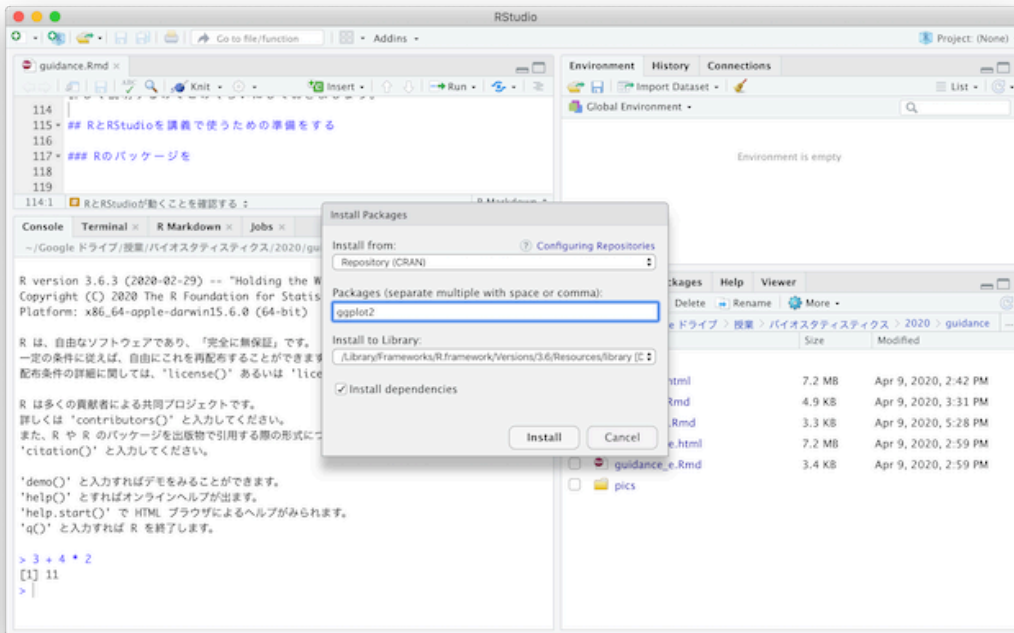
Install R packages.

R can perform a variety of statistical analyses even as it is, but if you install an “extended set” called a package, the number of available analytical methods can be largely expanded. This section explains how to install a package. Please note that your computer must be connected to the Internet in order to perform the installation of a package.

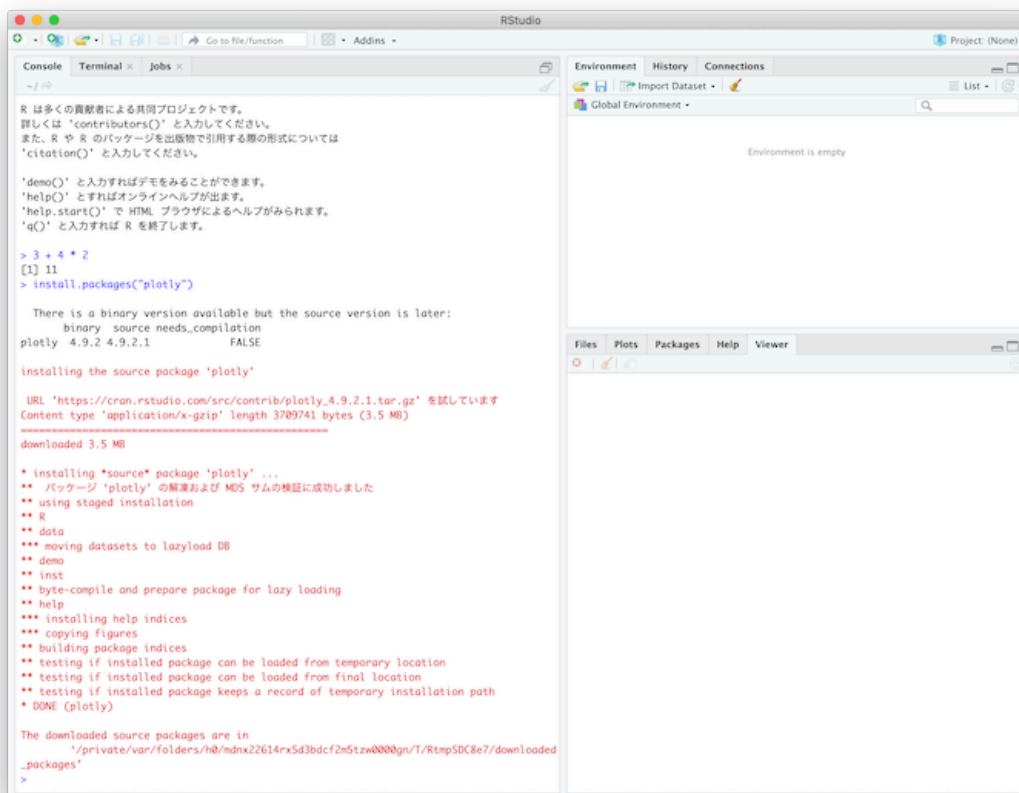
1. Select “Install Packages” from the Tools menu in RStudio.



2. Enter the name of the package you want to install in the dialog box shown in the middle of the figure below. In this section, we will install the package called plotly.



3. The plotly package will be downloaded and installed.

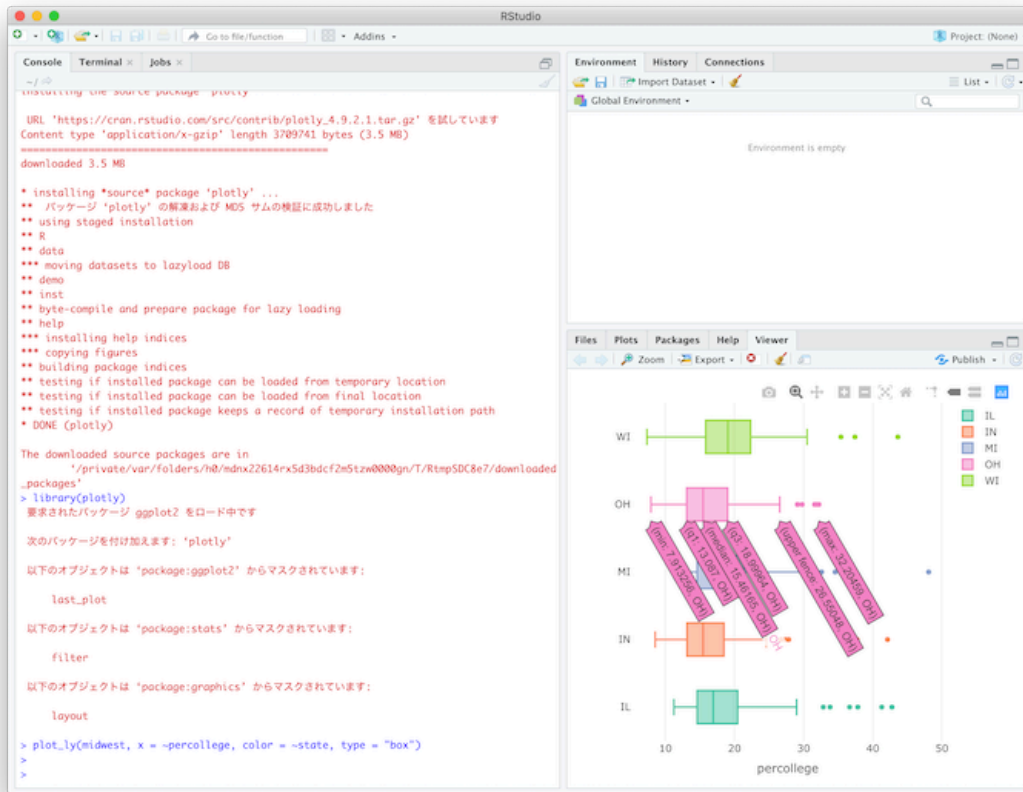


4. Type the following commands to the “Console” to draw an interactive figure.


```
library(plotly)
```

```
plot_ly(midwest, x = ~percollege, color = ~state, type = "box")
```

It will be shown in RStudio as follow.



Configuring the R Working Directory

In R, it is often necessary to specify a “working directory” to run R. This enables R find R codes and data to use. Here, let’s use RStudio to specify a working directory.

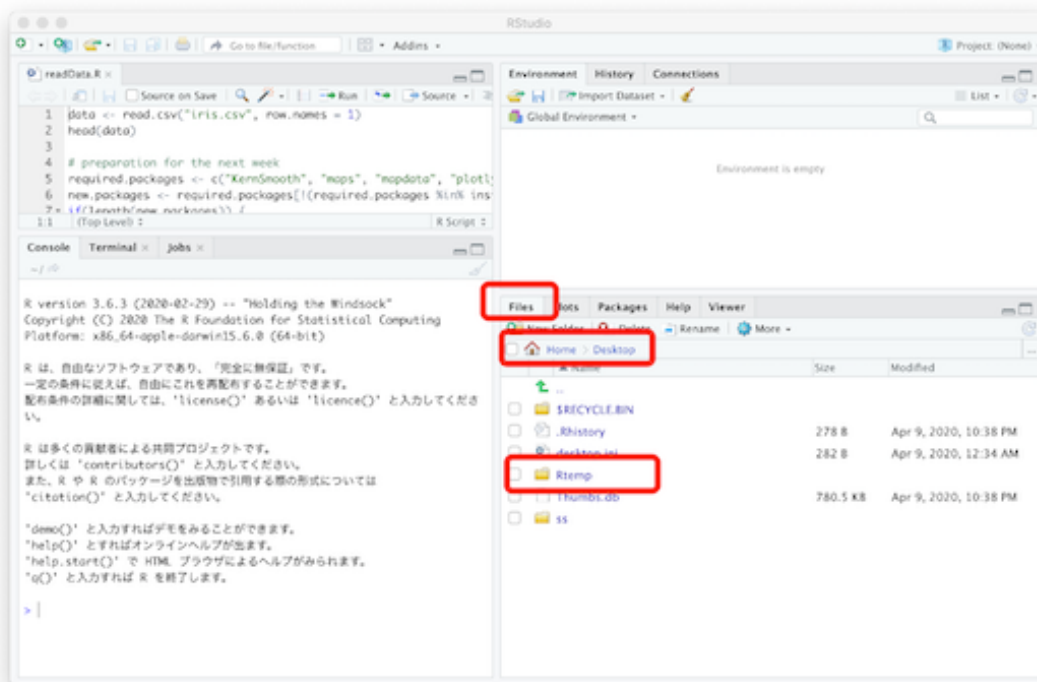
First, download the R code (readData.R) and data (iris.csv) from Google Drive below.

<https://drive.google.com/drive/u/2/folders/160TTfRf0Lz9qxm8vLdciaYZmSOZr9BL>

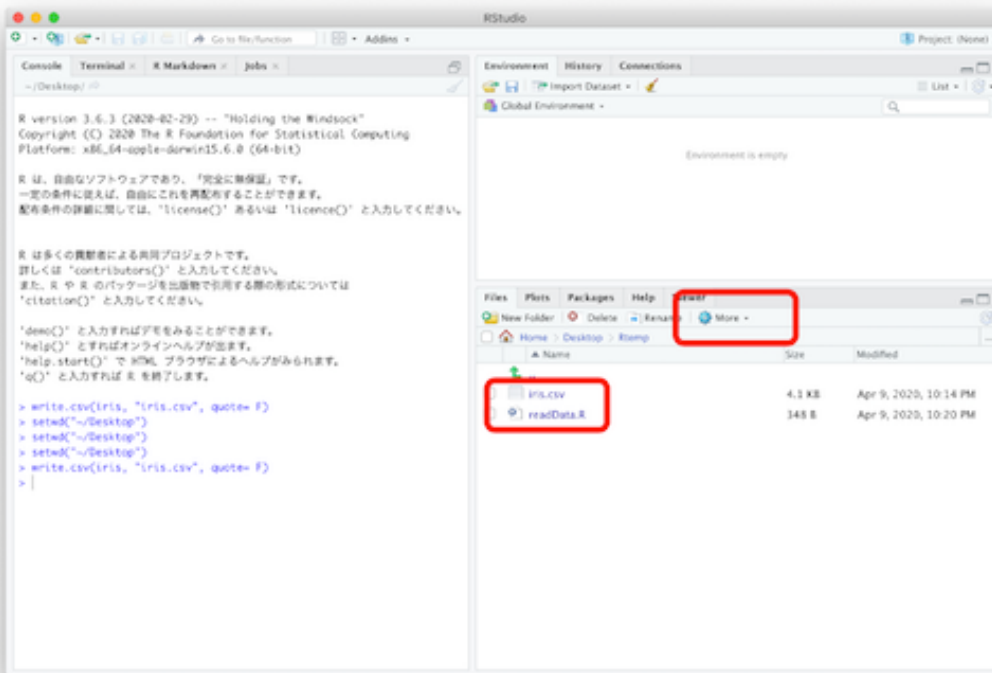
Next, move these files somewhere in a folder (directory) that is easy to work with for you. Here, I assume that you create a folder called Rtemp on your desktop and saved the files in the folder.

You can designate this folder as a “Working Directory” in RStudio as follows:

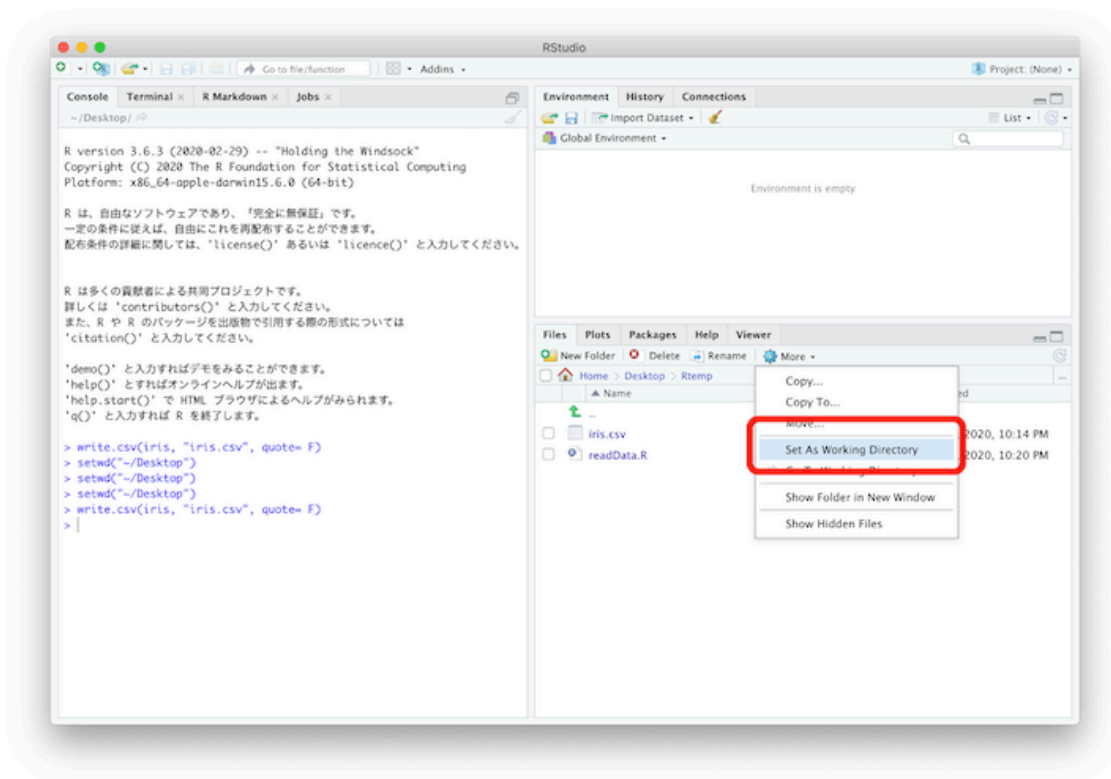
1. Click on the tab called “Files” in the lower right-hand corner of RStudio.



2. You can navigate to the various folders on your computer by clicking on a file path (here it is shown as “Home > Desktop”) or folders shown in the “Files” window.
3. Here we want to move to “Rtemp”, so click on “Rtemp”. You can see that there is the files of R-code and data that you have saved there.
4. Once you have moved to the location where the R-code or data exist, click on “More” (next to the gear-shaped icon).

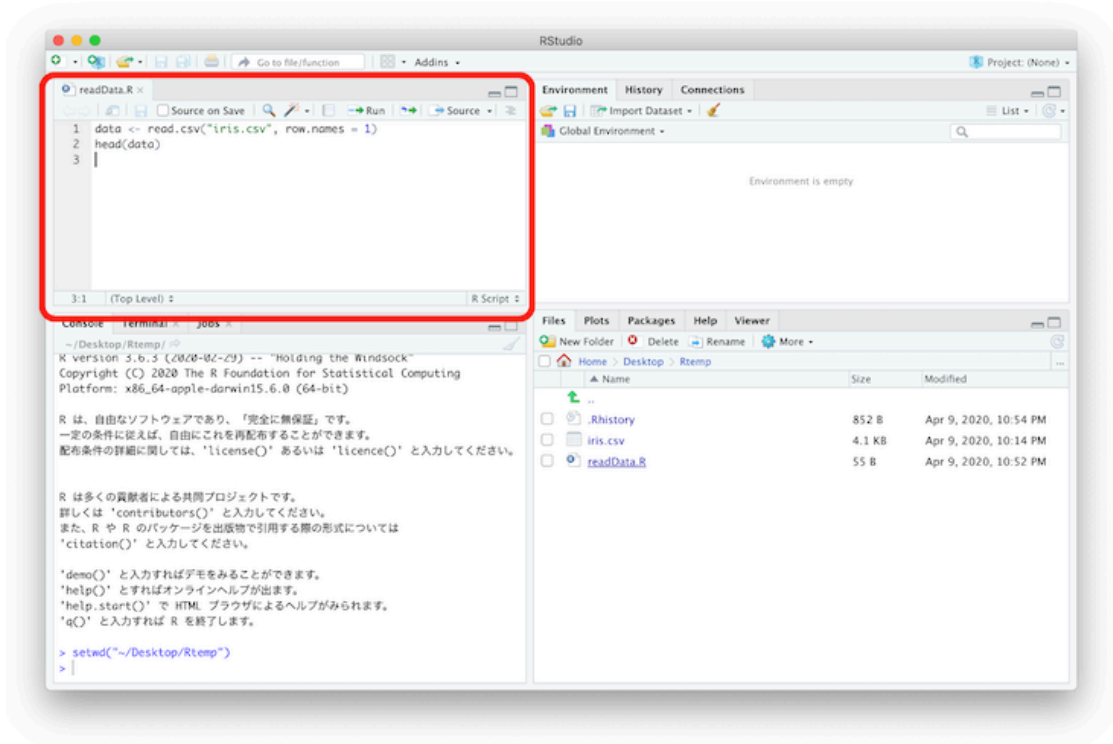


5. A drop down menu will then appear and select 'Set As Working Directory'.

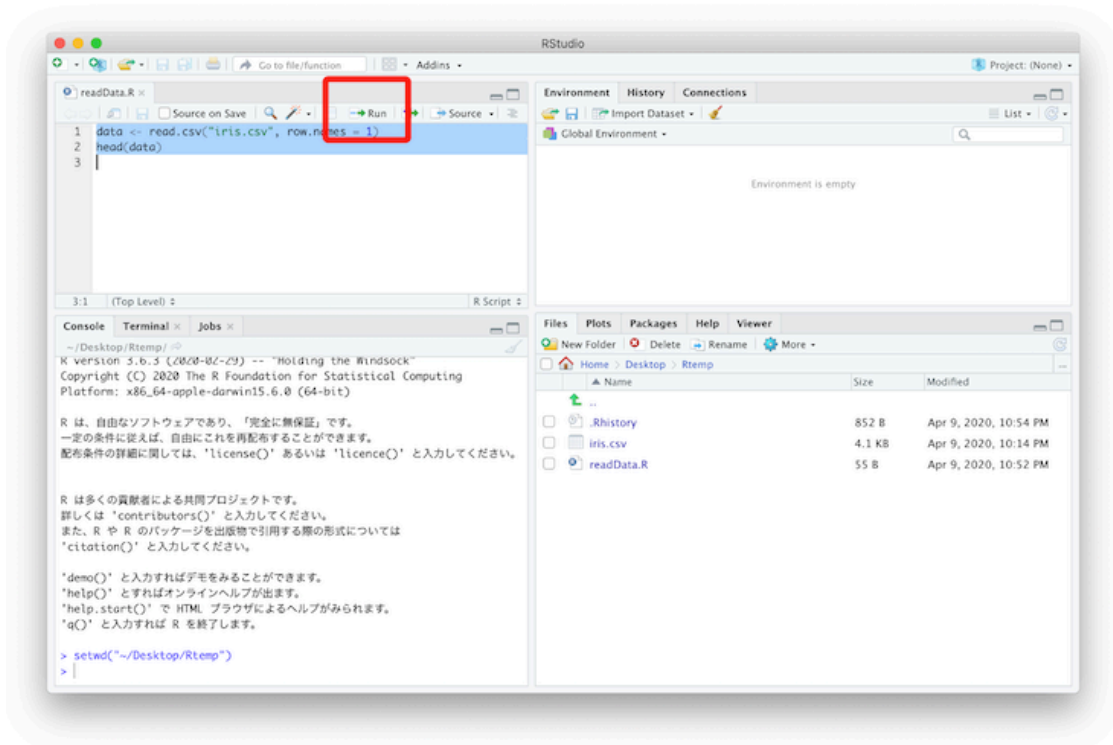


6. Now you designate a "Working Directory" to the "Rtemp" folder on your desktop. This makes it easy to load data files to R.

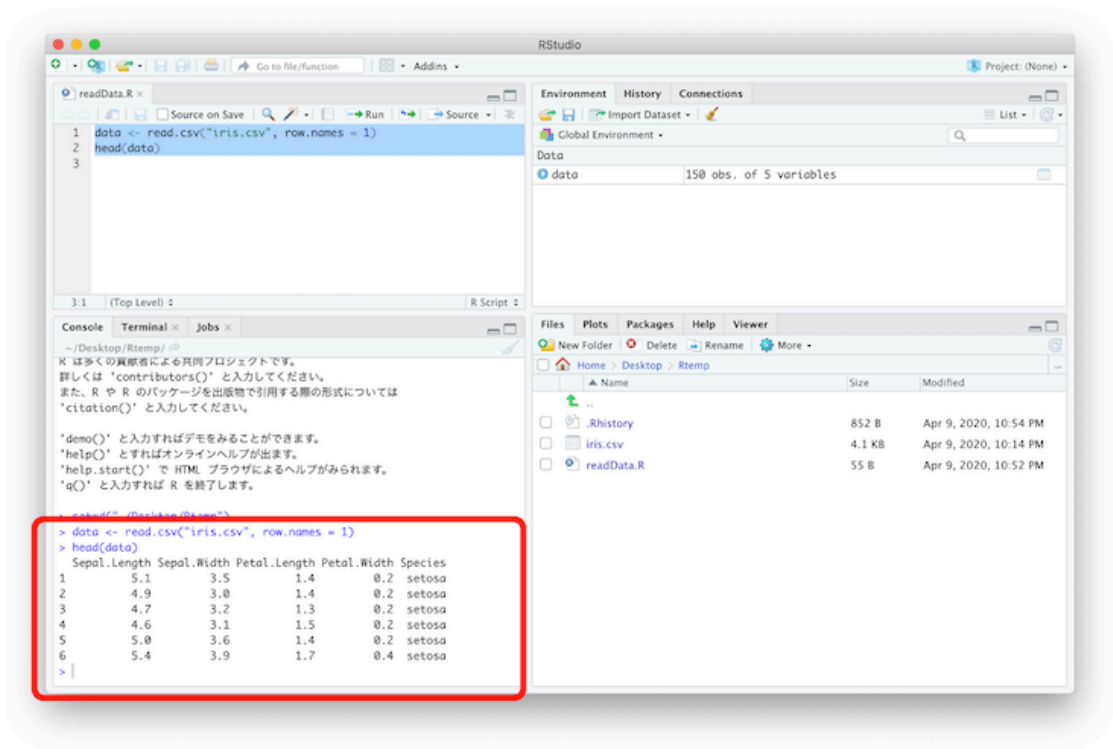
First, let's open the R code. If you click on the file "readData.R" in the "Files" window, you will see the code in the top left window.



Let's select the first two lines of this and run them. To run, click on "Run". Alternatively, you can use "command + return (enter)" on Mac OS and "ctrl + enter (return)" on Windows.



If it runs successfully, you'll see the following result.



If you can do this far, there will be few issues in next week's class.