



# Applied Statistics

---

## Getting Started in R

# What are R and RStudio?

R IS A PROGRAMMING LANGUAGE USED FOR STATISTICAL COMPUTING.

---

**YOU NEED TO INSTALL R FIRST.**

RSTUDIO IS BUILT ON R AND PROVIDES A MORE USER FRIENDLY INTERFACE FOR USING R. **THIS IS WHAT WE WILL USE IN CLASS.**

**YOU NEED TO INSTALL RSTUDIO AFTER YOU INSTALL R.**

**The instructions are found here:**

[https://lhartigan15.github.io/LLO8200\\_summer2021/Installing%20R%20and%20RStudio.pdf](https://lhartigan15.github.io/LLO8200_summer2021/Installing%20R%20and%20RStudio.pdf)

Note: It is important to note the differences between R and RStudio. R is a programming language used for statistical computing while RStudio uses the R language to develop statistical programs. In R, you can write a program and run the code independently of any other computer program. RStudio however, must be used alongside R in order to properly function.

R and RStudio are not separate versions of the same program and cannot be substituted for one another. R may be used without RStudio, but RStudio may not be used without R.

# PC vs Mac

---

The commands for R are the same for both a PC and a Mac environment.

There are some differences in the user interface (particularly in the initial set up when trying to knit to pdf). Let me know if you are on a Mac and unable to knit to pdf (after following all instructions in this slide deck).

## R Markdown files - .Rmd files

---

Each week you will receive access to a .Rmd file. This means that the file extension will be .Rmd (other file extensions that you are probably familiar with are .docx and .xlsx and .txt). We will use the .Rmd files to communicate with R.

You will NOT need to create .Rmd files from scratch but rather simply edit the ones you receive from the faculty in order to complete your assignments. **IMPORTANT: Set up a separate folder for our course** Good file management will save you a lot of time and headaches. You should download the .Rmd file to a folder for our course. **DO NOT LET the .Rmd file LIVE in your Download Folder.** R does not like to read files from the Download Folder on your computer.

# Opening the .Rmd file

---

Once your .Rmd file is downloaded you can open the file one of two ways:


- 1) Double click on the .Rmd file directly from the folder where you saved it . This should cause Rstudio to open and display the file in the upper left panel of the program.

OR

- 2) Click on the Rstudio icon



Then Use File> Open File from the menu bar and browse to find the folder where you have saved the .Rmd file. Click on the file name to open it and it should display the file in the upper left panel

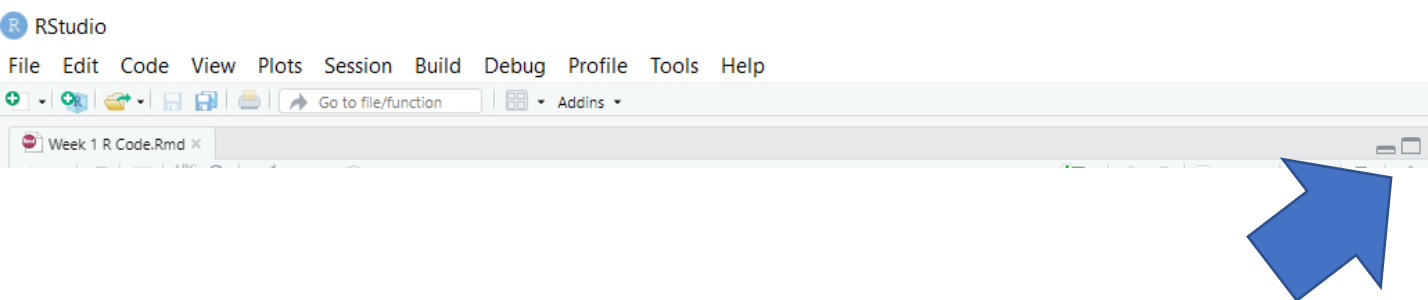
 RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help



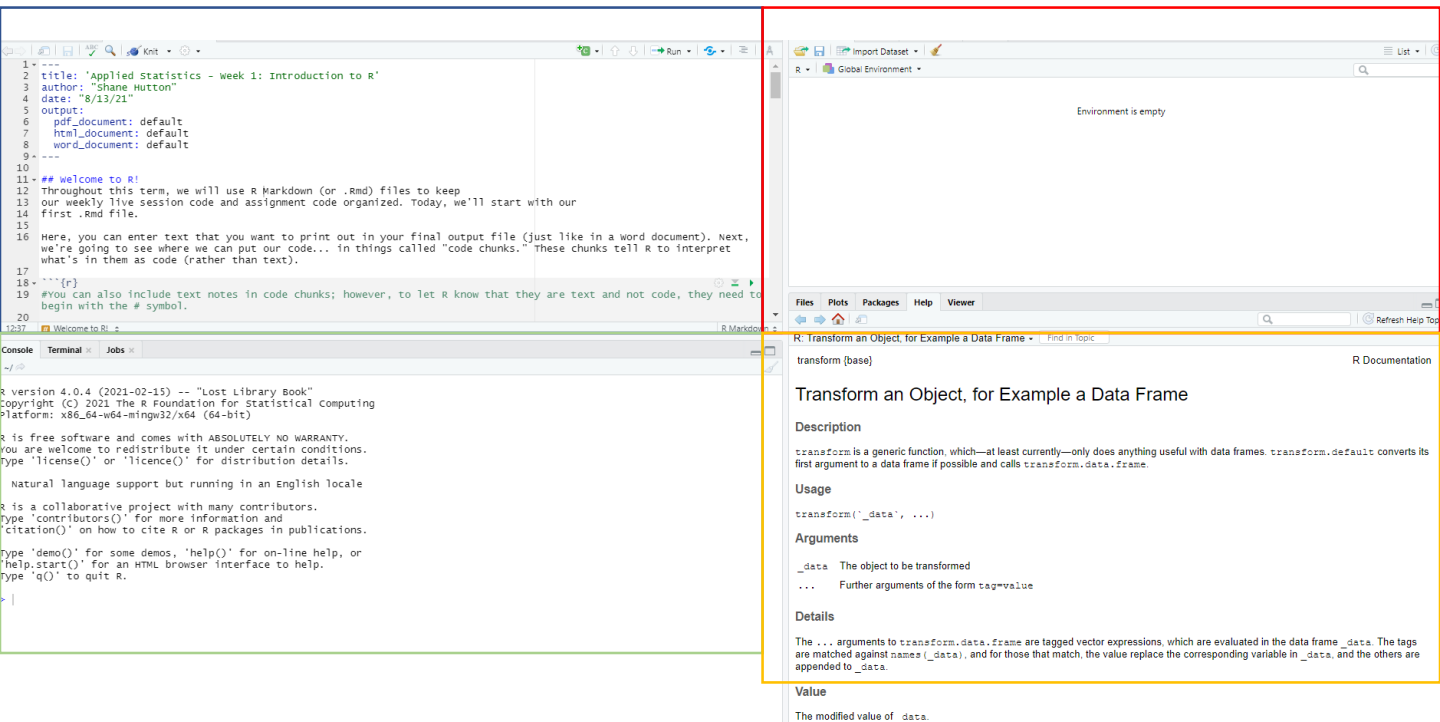
You Should See 4 Panels in your .Rmd file.

If you do not see 4 panels, use the min/max controls to get them all to display



Note: You can customize the colors and fonts of the panels using Tools> Global Options from the menu bar.

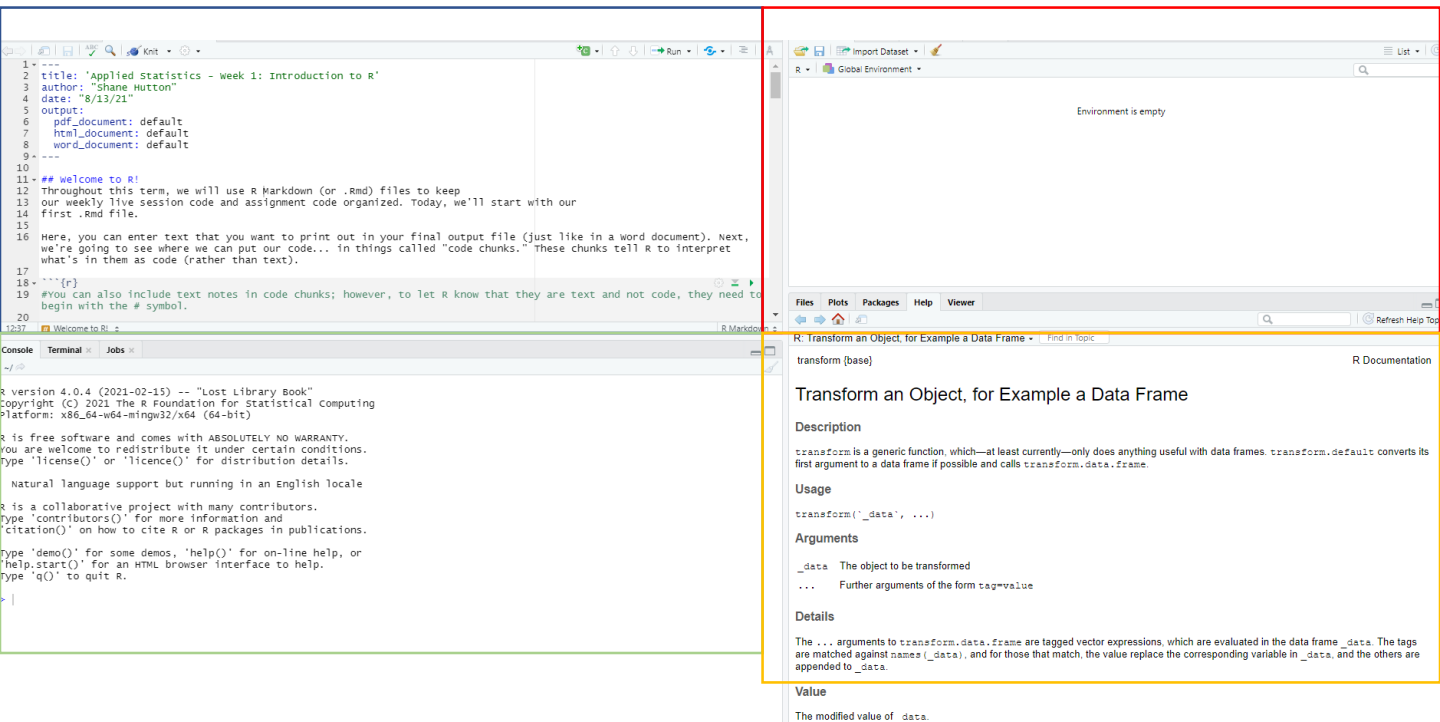
# Overview of each of the 4 Panels in your .Rmd file



**Upper Left panel** is where your source code and text will go. This is where the .Rmd file will be displayed when you open it. This is saved when you save and close the file.

**Lower Left Panel:** this is called the console. When you run a “chunk of code” it will be displayed here (except some plots). This will be used to test your code. This panel is not saved when you save and close the file.

# Overview of each of the 4 Panels in your .Rmd file



**Upper Right panel :** This shows what is in R's brain right now. It is called the Global Environment and will be used to confirm that your data has been properly read in.

**Lower Right Panel:** this area has a variety of tabs. We will use the Packages tab to install Packages. You will see Plots displayed here and you can use the Help feature from here.



# A closer look at the .Rmd file – The YAML header

---

Look in the upper left-hand panel.

Notice that the .Rmd file has a header starting on line 1.

```
1 ---  
2 title: 'Applied Statistics - week 1: Introduction to R'  
3 author: "Shane Hutton"  
4 date: "8/13/21"  
5 output:  
6   pdf_document: default  
7   html_document: default  
8   word_document: default  
9 ---  
10
```

This is called a YAML header – “Yet another Markdown language.”

Your .Rmd file MUST have a header in exactly this format which will be in the .Rmd file that you download.

It MUST begin on line 1 and it MUST begin and end with the 3 dashes.

You can edit the title and author and date. The output shows the options for “knitting” the file. Knitting does what it sound like – it knits together the text and the output and the code. This is what you will submit for your assignments. More on this later.

# A closer look at the .Rmd file - Text

---

Look in the upper left-hand panel.

Notice that the .Rmd file has text starting on line 11

```
11 ## welcome to R!  
12 Throughout this term, we will use R Markdown (or .Rmd) files to keep  
13 our weekly live session code and assignment code organized. Today, we'll start with our  
14 first .Rmd file.  
15  
16 Here, you can enter text that you want to print out in your final output file (just like in a word document). Next,  
we're going to see where we can put our code... in things called "code chunks." These chunks tell R to interpret  
what's in them as code (rather than text).
```

One of the great features of R is that you can just type text as though you were using Word. The text will appear as part of your output file when you knit everything (last step).

This allows you to interpret the output right in the file. You will answer the assignment questions in this manner.

NOTE: on line 11 you see `##` symbols. This makes the text that follows into a heading in the output file. For now, you can just use what we give you. If you want to adjust the way the output looks you certainly can learn about headings. Typically, text needs no symbols to differentiate it from code UNLESS the text is in a “code chunk.”

# A closer look at the .Rmd file - Code Chunks

---

Notice that lines 18 to 24 are gray

```
18 ~~~{r}
19 #You can also include text notes in code chunks; however, to let R know that they are text and not code, they need to
   begin with the # symbol.
20
21 #First we will see how R can perform simple calculations
22 #Suppose we want to calculate 2+2
23 2+2 #you can add notes anywhere
24 ~~~
```

This is called a code chunk in R.

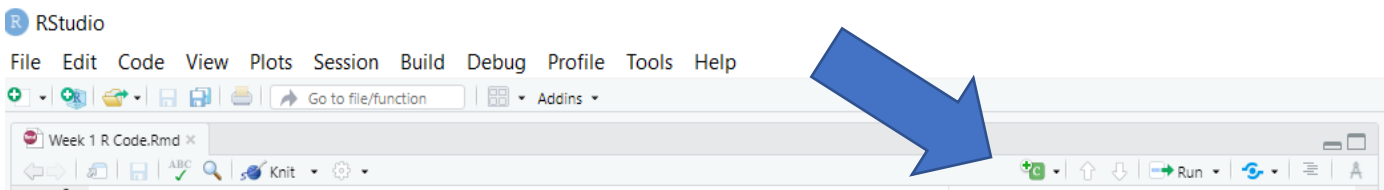
Notice that it **MUST** begin and end with the appropriate symbols to tell R that it is a code chunk.

The code chunks that you need will be in the .Rmd file that you receive. You will just need to edit it to complete your assignments.

# Adding code chunks

---

You may want to add a new code chunk sometime. You can use the dropdown arrow in the menu bar to insert a code chunk.



OR

You can just type the opening and closing lines to delineate code.

BUT be VERY careful that you type it just exactly as it is shown – curly brackets are necessary!

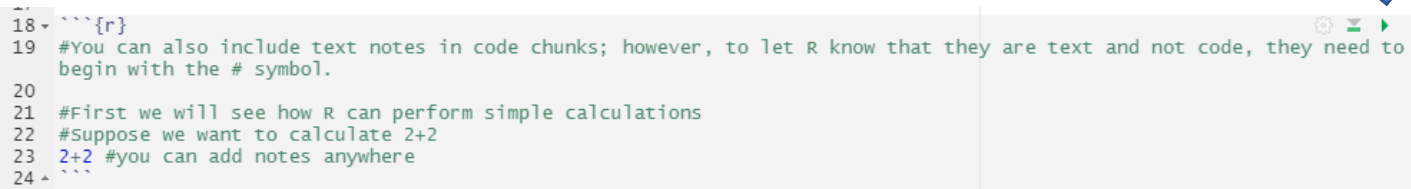
Mac users can add a code chunk by pressing

`⌘ + ⌘ + i`

# Running a Code Chunk

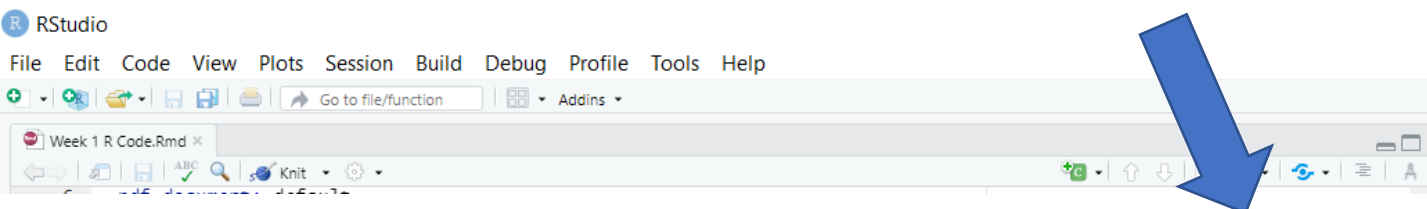
In order to run a code chunk you can:

- 1) Use the “Play” button at the top right-hand corner of the code chunk



```
18 ```{r}
19 #You can also include text notes in code chunks; however, to let R know that they are text and not code, they need to
   begin with the # symbol.
20
21 #First we will see how R can perform simple calculations
22 #Suppose we want to calculate 2+2
23 2+2 #you can add notes anywhere
24 ```
```

- 2) Highlight the code and use the drop-down arrow next to the Run button from the main menu at the top of the upper left panel:



The output will show up in the console and in the upper left panel just below the code chunk that you executed.

Mac Users can run a code chunk by pressing

Shift + ⌘ + Enter

# Adding text within a Code Chunk

---

You can easily add text within a code chunk by preceding the text with a # symbol.

```
18 ~~~{r}
19 #You can also include text notes in code chunks; however, to let R know that they are text and not code, they need to
   begin with the # symbol.
20
21 #First we will see how R can perform simple calculations
22 #Suppose we want to calculate 2+2
23 2+2 #you can add notes anywhere
24 ~~~
```

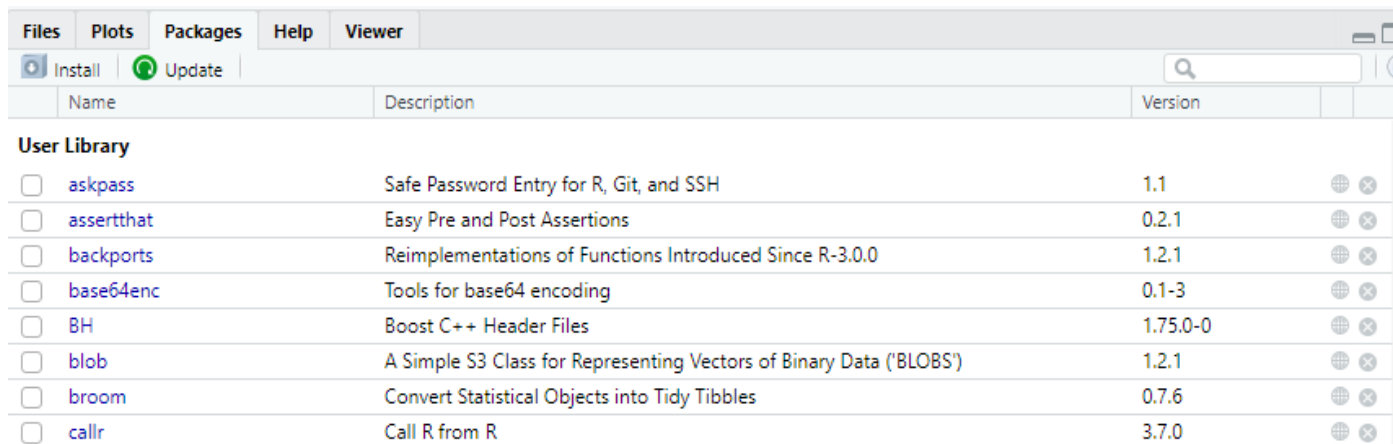
This allows you to put some notes/documentation into the code chunk for future reference.

This also allows you to tell R to ignore the code chunk for the moment. Often you might think you don't need a code chunk and so you delete it. Later, you discover you need it. So, instead of deleting it, you can make R ignore it for the time being by putting a # symbol in front of each line in the code chunk.

# Installing Packages

Tons of free upgrades to the basic R functionality are available. These are called packages.

In order to use a package, you must **first install it**. You only have to install a package once. Then it will show up on your list of packages in the lower right-hand panel when you click on the Packages tab. Installing a package makes it available for use but you still must load it each time you want to use it.



The screenshot shows the RStudio interface with the 'Packages' tab selected. The 'User Library' section lists several installed packages. Each row includes a checkbox, the package name, a description, the version number, and icons for refreshing and removing the package.

|                          | Name       | Description   | Version  |  |
|--------------------------|------------|---|----------|--|
| <input type="checkbox"/> | askpass    | Safe Password Entry for R, Git, and SSH                             | 1.1      |  |
| <input type="checkbox"/> | assertthat | Easy Pre and Post Assertions  | 0.2.1    |  |
| <input type="checkbox"/> | backports  | Reimplementations of Functions Introduced Since R-3.0.0             | 1.2.1    |  |
| <input type="checkbox"/> | base64enc  | Tools for base64 encoding   | 0.1-3    |  |
| <input type="checkbox"/> | BH         | Boost C++ Header Files  | 1.75.0-0 |  |
| <input type="checkbox"/> | blob       | A Simple S3 Class for Representing Vectors of Binary Data ('BLOBS') | 1.2.1    |  |
| <input type="checkbox"/> | broom      | Convert Statistical Objects into Tidy Tibbles                       | 0.7.6    |  |
| <input type="checkbox"/> | callr      | Call R from R   | 3.7.0    |  |

In the console (lower left panel) type the following:

```
install.packages("tidyverse")
```

This will install tidyverse. Notice the **stop sign** in the lower left-hand panel when R is doing the installation. When you see the stop sign, it means R is working on it. Wait until R is finished to do anything else.

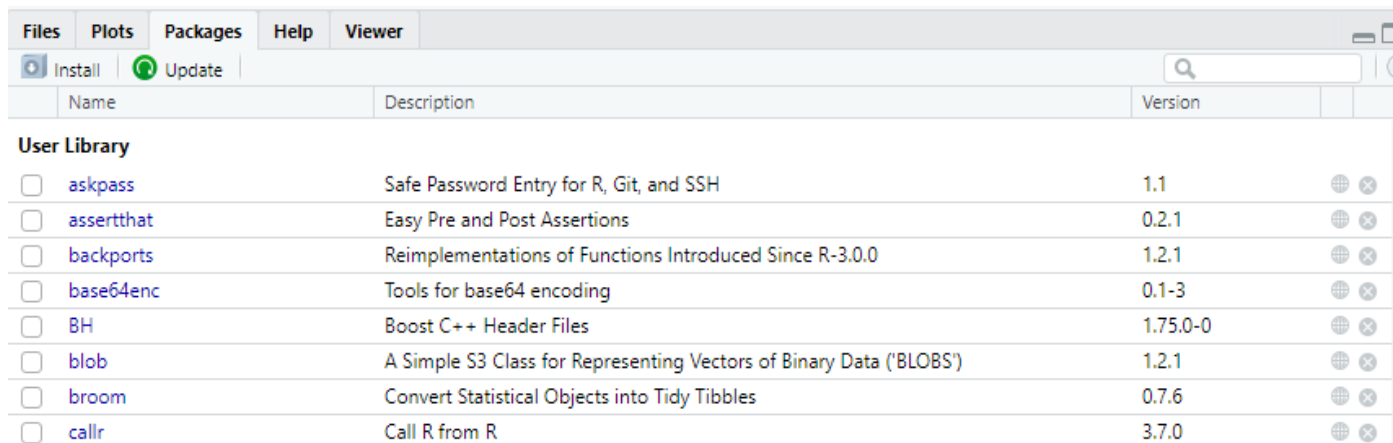
# Using Packages



















After you have installed a package, you **must load it in order to use it**. You must load the package each time you want to use it, so it is best to do that in the upper left-hand panel in an R code chunk.

Use the library function in a code chunk in order to load a package. Type this in a code chunk.

```
library(ggplot2)
```

When you run that code chunk, the package called ggplot will be moved from the “bookshelf” of packages and be available to be used. It will then appear in the Package list with a check mark.



| Files   | Plots   | Packages             | Help  | Viewer  |
|---|---|----------------------|---|---|
|  |  | <input type="text"/> |   |   |
| Name  | Description   | Version              |   |   |
| <b>User Library</b>   |   |                      |   |   |
| <input type="checkbox"/> askpass  | Safe Password Entry for R, Git, and SSH   | 1.1                  |  |  |
| <input type="checkbox"/> assertthat   | Easy Pre and Post Assertions  | 0.2.1                |  |  |
| <input type="checkbox"/> backports  | Reimplementations of Functions Introduced Since R-3.0.0                             | 1.2.1                |  |  |
| <input type="checkbox"/> base64enc  | Tools for base64 encoding   | 0.1-3                |  |  |
| <input type="checkbox"/> BH   | Boost C++ Header Files  | 1.75.0-0             |  |  |
| <input type="checkbox"/> blob   | A Simple S3 Class for Representing Vectors of Binary Data ('BLOBS')                 | 1.2.1                |  |  |
| <input type="checkbox"/> broom  | Convert Statistical Objects into Tidy Tibbles                                       | 0.7.6                |  |  |
| <input type="checkbox"/> callr  | Call R from R   | 3.7.0                |  |  |

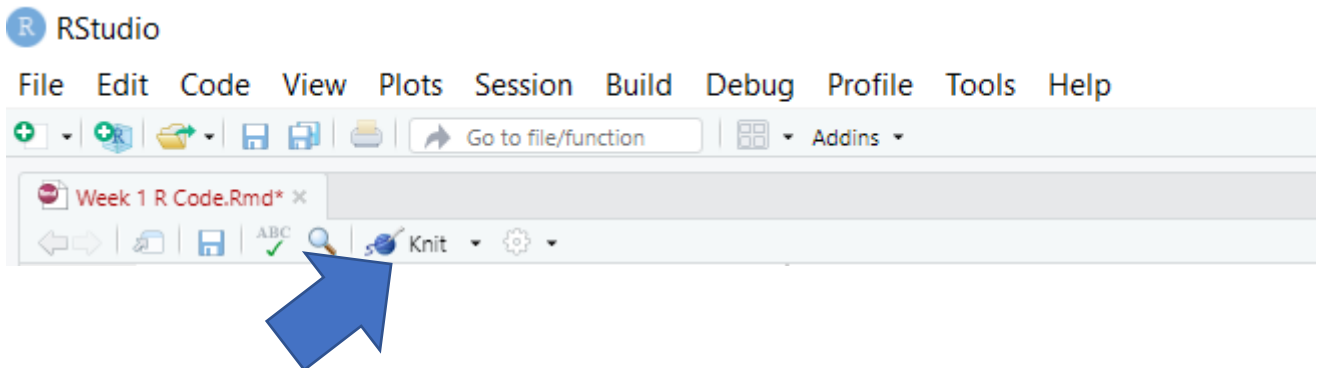


# Knitting to HTML

---

Once you have all your code chunks working and you have added all the text that you wish to see in your report (answered all the questions in the homework assignment) then you are ready to Knit the file.

To do this, choose drop down arrow next to the “ball of yarn” on the main menu in the upper left-hand panel. It says “Knit”. Try knitting to HTML first.



The knitted file should open for you on your screen and get saved to the same folder where you saved the .Rmd file that you are running. This is the file you will submit to me through the Assignments tab in the LMS!

NOTE: If you do not see the knit button then you have probably opened a .R file and not an .Rmd file. You can only knit .Rmd files.

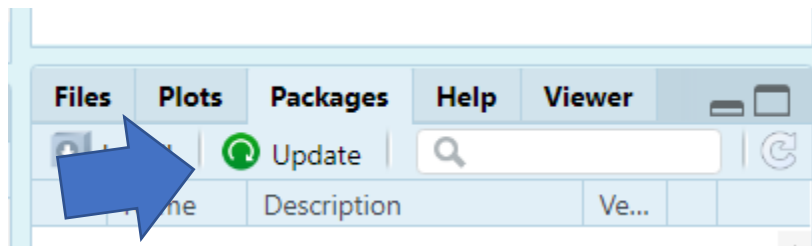
# Knitting to pdf – Troubleshooting on a PC

---

Once you have successfully knit to html, go ahead and try to knit to pdf. If you encounter no errors, fantastic! If you do encounter an error, try the following steps.

## 1. Update all out-of-date packages.

- In the lower-right pane, you'll see a "Packages" tab. Click the green "Update" button on that tab. This will pull up a list of everything that is out of date.

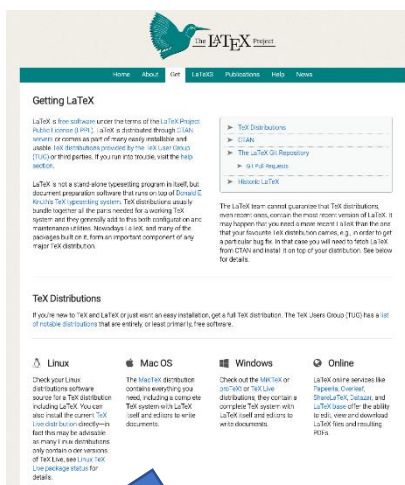


2. Install tinytex by typing the following code into your console (and hitting enter):  

```
tinytex::install_tinytex()
```
3. After that is finished installing, go to FILE > QUIT SESSION
4. Re-open the .Rmd file/RStudio and attempt to knit to pdf again.

# Knitting to pdf – Troubleshooting on a Mac

Mac Users will need to download MacTeX. Go to <https://www.latex-project.org/get/>



Click here

## The MacTeX-2021 Distribution

The current distribution is MacTeX-2021  
This distribution requires Mac OS 10.14, Mojave, or higher and runs natively on Intel and Arm processors.

To download, click [MacTeX Download](#).

You can also install TeX Live 2021 using the TeX Live Unix Install Script.  
This method supports MacOS 10.6, Snow Leopard, or higher and runs on Intel and Arm processors.

To download the smaller distribution, click [Smaller Download](#).

To Obtain Older Versions of MacTeX If You Are Running Mac OS 10.3 through 10.13, [click here](#)

To download the smaller distribution, click [Smaller Download](#).

For suggestions on keeping up to date, go to [Update Schedule](#).

The link below leads to optional download packages:  
[MacTeX and optional pieces](#)

Then here

This is a big file and will take some time to download and install.

# Knitting to pdf – Troubleshooting on a Mac

---

Sometimes Mac users experience problems when knitting files that do not occur when running the code chunks individually. This is often a result of ‘knitr’ restoring the working directory to the original directory.

You can fix this by setting the root directory in knitr in your first code chunk by including the following code:

```
opts_knit$set(root.dir = "  ")
```

# CONGRATS

---

If you have successfully knitted to a pdf and/or HTML file, then you are well on your way!

See you in class!

Each week we will go over the R code needed to implement the weeks stat tools in R.