

Introductions, Defining Data Science and Ethics, Intro to R and R Studio

SSEP 2022 Morning Day 1

Dr. Ab Mosca (they/them)

Introductions

Course Staff

Dr. Ab Mosca (they/them)

- BA in Mathematics, Smith College 2014
- PhD in Computer Science, Tufts University 2021
- Assistant Teaching Professor, Northeastern University
- Research Area: Data Visualization
- Website:
<https://www.khoury.northeastern.edu/home/amosca/>

Esa Schenck (they/them)

- Smith College Class of 2023
- Major: Data Science

Classmates

- Ice Breaker #1
 - Break into groups of 3
 - Introduce yourself
 - What is one piece of media that you consumed over the summer that you enjoyed?
 - Movie, tv show, book, etc.
 - Find one common attribute between all of you.
 - Do you all like the same food? Do you all enjoy the same music? Do you all live in the same place?
 - Add your answer to this Jamboard to share with the class: [Ice Breaker Jamboard](#)

Classmates

- Ice Breaker #2
 - Break into new groups of 3
 - Introduce yourself
 - What is one thing you hope to learn in this class?
 - Add your answers to this Jamboard to share with the class: [Ice Breaker Jamboard](#)

Classmates

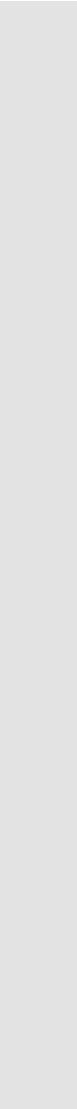
- Ice Breaker #3
 - Break into new groups of 3
 - Introduce yourself
 - What is one thing you are nervous about for this class?
 - Add your answers to this Jamboard to share with the class: [Ice Breaker Jamboard](#)

Ice Breaker Recap

- Let's look at the outcome of our ice breaker questions
- [Ice Breaker Jamboard](#)



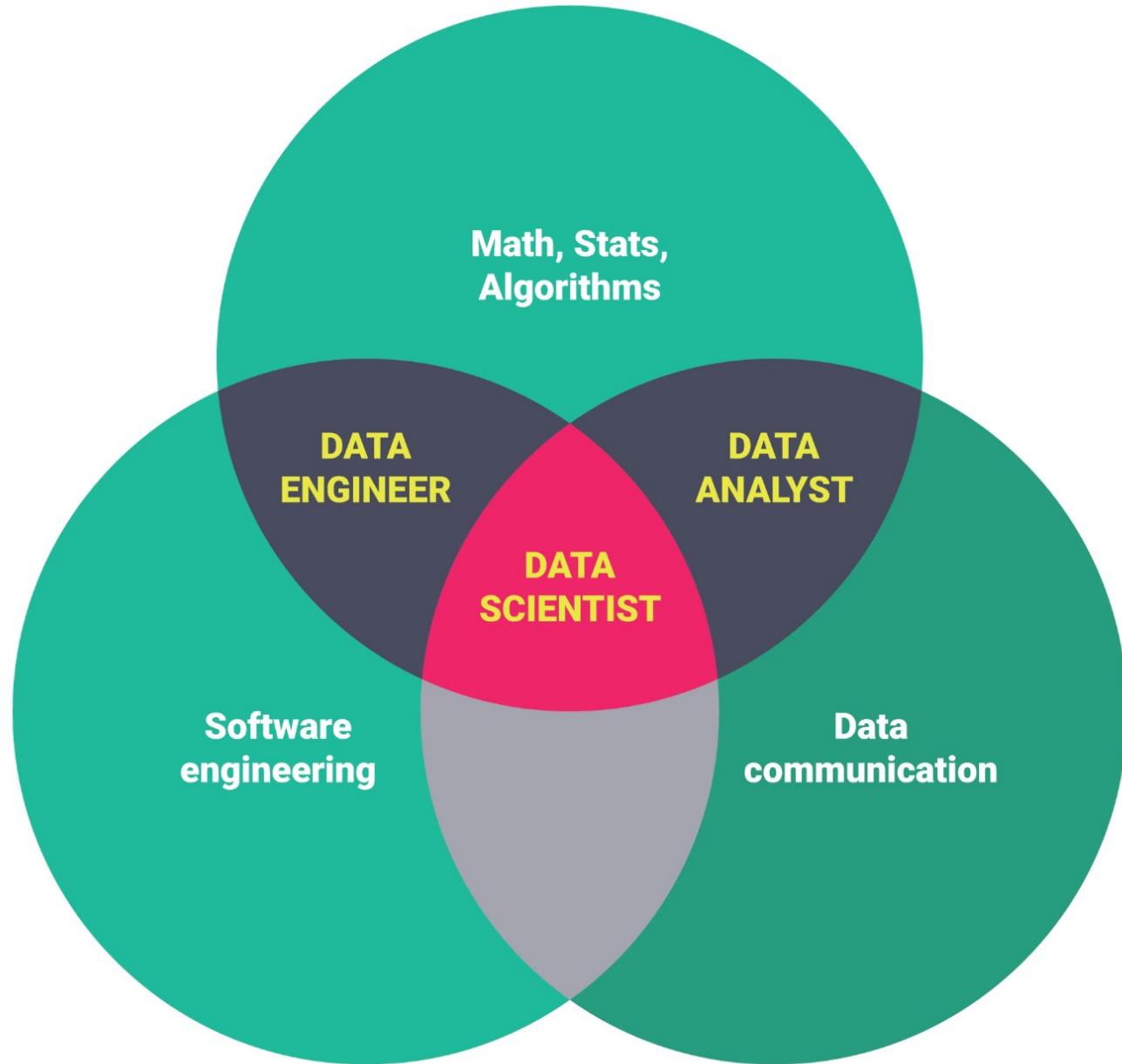
Defining Data Science and Ethics



Defining Data Science

- How would you **define data science?**
- Talk it out with the person sitting next to you then add your definition to this Jamboard:
[Definitions Jamboard](#)

Definition: Data Science



Defining Ethics

- How would you **define ethics?**
- Talk it out with the person sitting next to you then add your definition to this Jamboard:
[Definitions Jamboard](#)

Definition: Ethics

Definition of *ethic*

- 1 **a** : a set of moral principles : a theory or system of moral values
// the present-day materialistic ethic
// an old-fashioned work ethic
—often used in plural but singular or plural in construction
// an elaborate ethics
// Christian ethics
- b** **ethics** \ 'e-thiks \ *plural in form but singular or plural in construction* : the principles of conduct governing an individual or a group
// professional ethics
- c** : a consciousness of moral importance
// forge a conservation ethic
- d** : a guiding philosophy

<https://www.merriam-webster.com/dictionary/ethic>

Defining Ethical Data Science

- How would you **define ethical data science?**
- Talk it out with the person sitting next to you then add your definition to this Jamboard:
[Definitions Jamboard](#)

Definition: Ethical Data Science

- Applying theories and systems of moral values (ethics) to our data science work

Ethical Data Science

- Applying theories and systems of moral values (ethics) to our data science work
- Why is this important?

Ethical Data Science

- Applying theories and systems of moral values (ethics) to our data science work
 - Why is this important?
-



Ethical Data Science

- Applying theories and systems of moral values (ethics) to our data science work
- Why is this important?

Can you think of other examples of things that can go wrong with data science?

Ethical Data Science

- Applying theories and systems of moral values (ethics) to our data science work
- Why is this important?
 - To combat algorithmic bias
 - To protect personal, identifying information
 - To increase reproducibility and replicability
- Systems of power are woven into our society, and so are misunderstandings of what data science is and how it works

Take a break – stand up, chat with
neighbors, use the restroom. We'll
start again in 10 minutes ☺



Intro to R and R Studio

What are R and R Studio?

R

- Statistical Programming Language



cloud.r-project.org

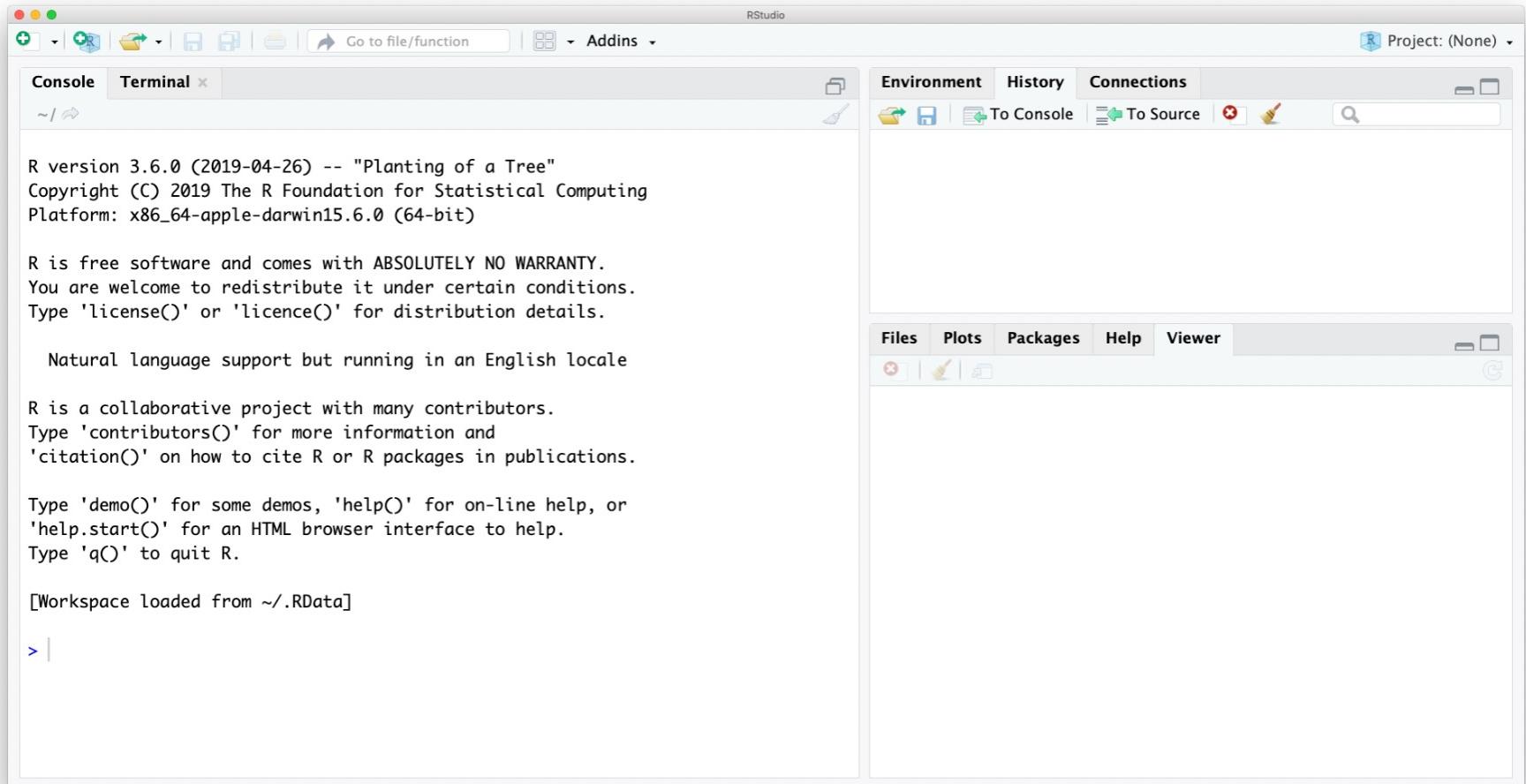
R Studio

- An Interactive Development Environment (IDE) for R

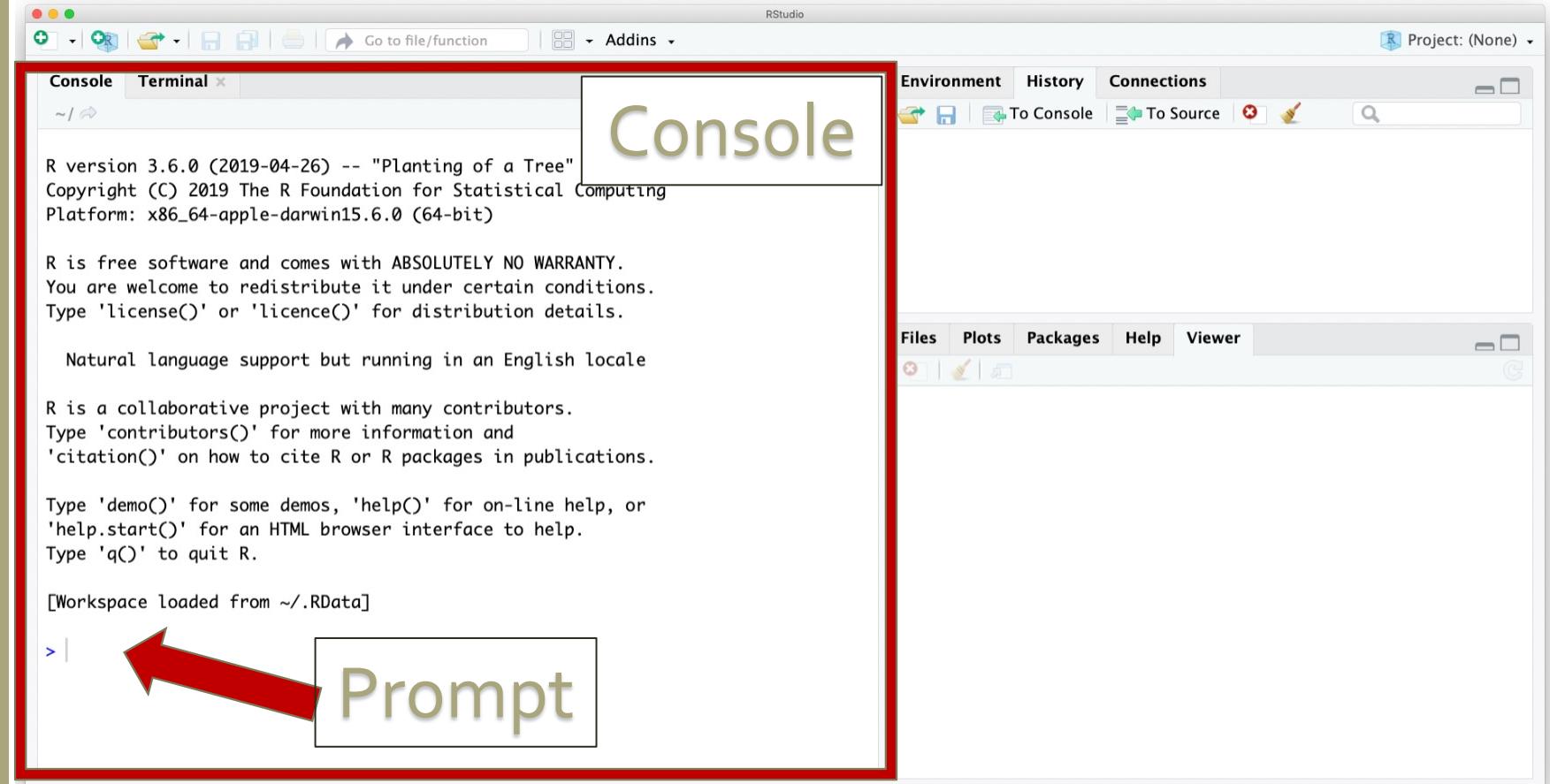


www.rstudio.com

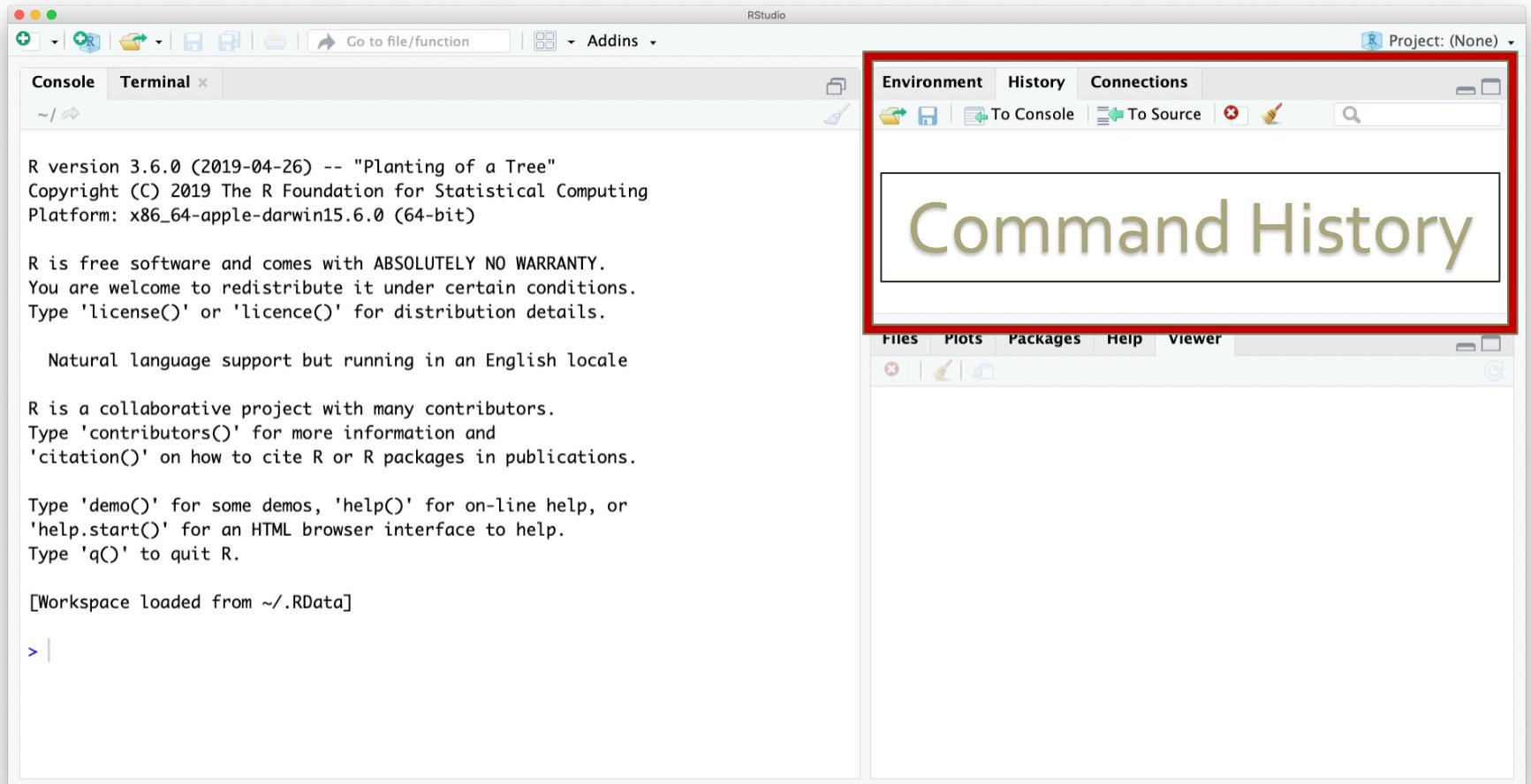
R Studio



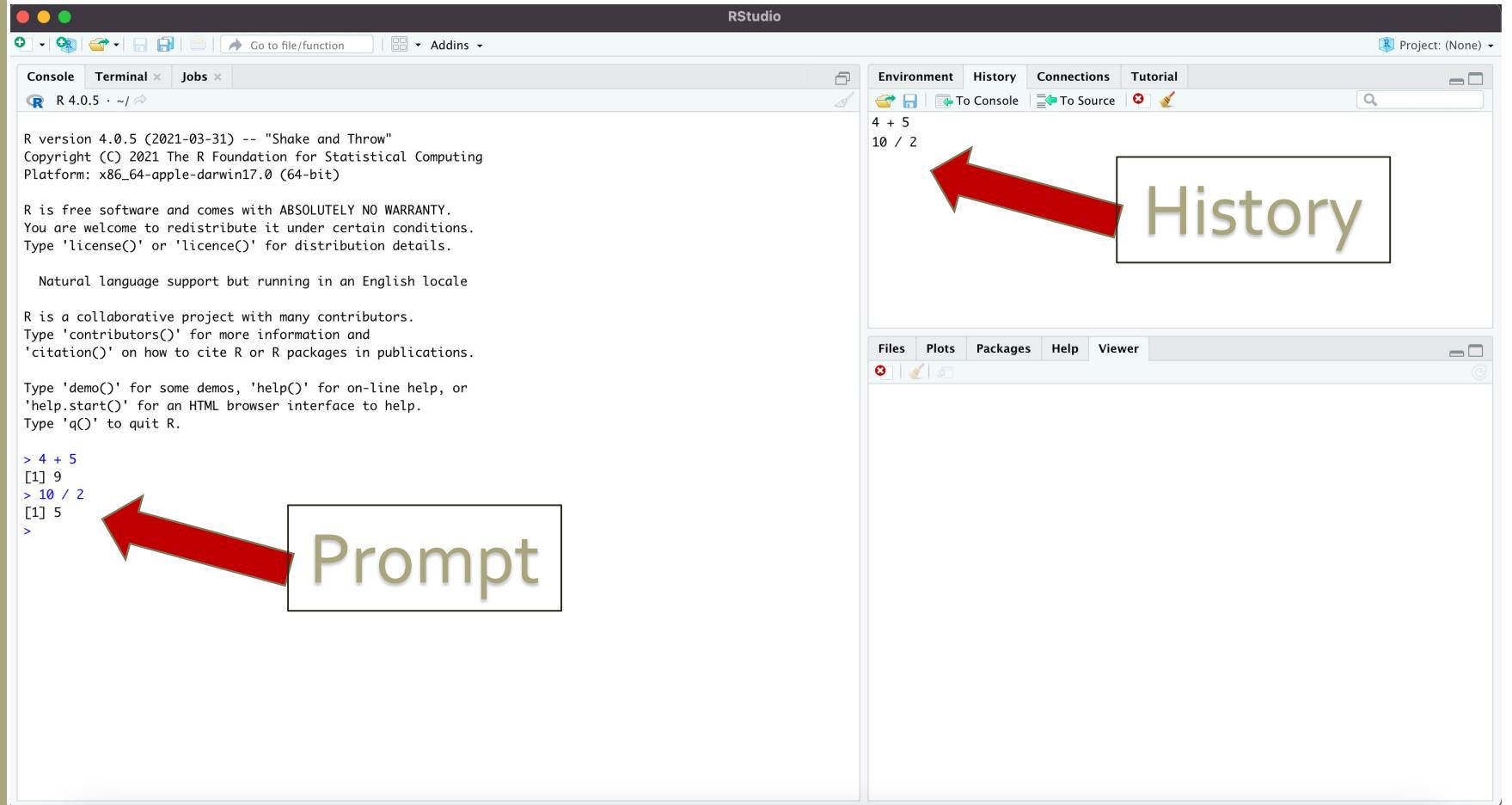
R Studio



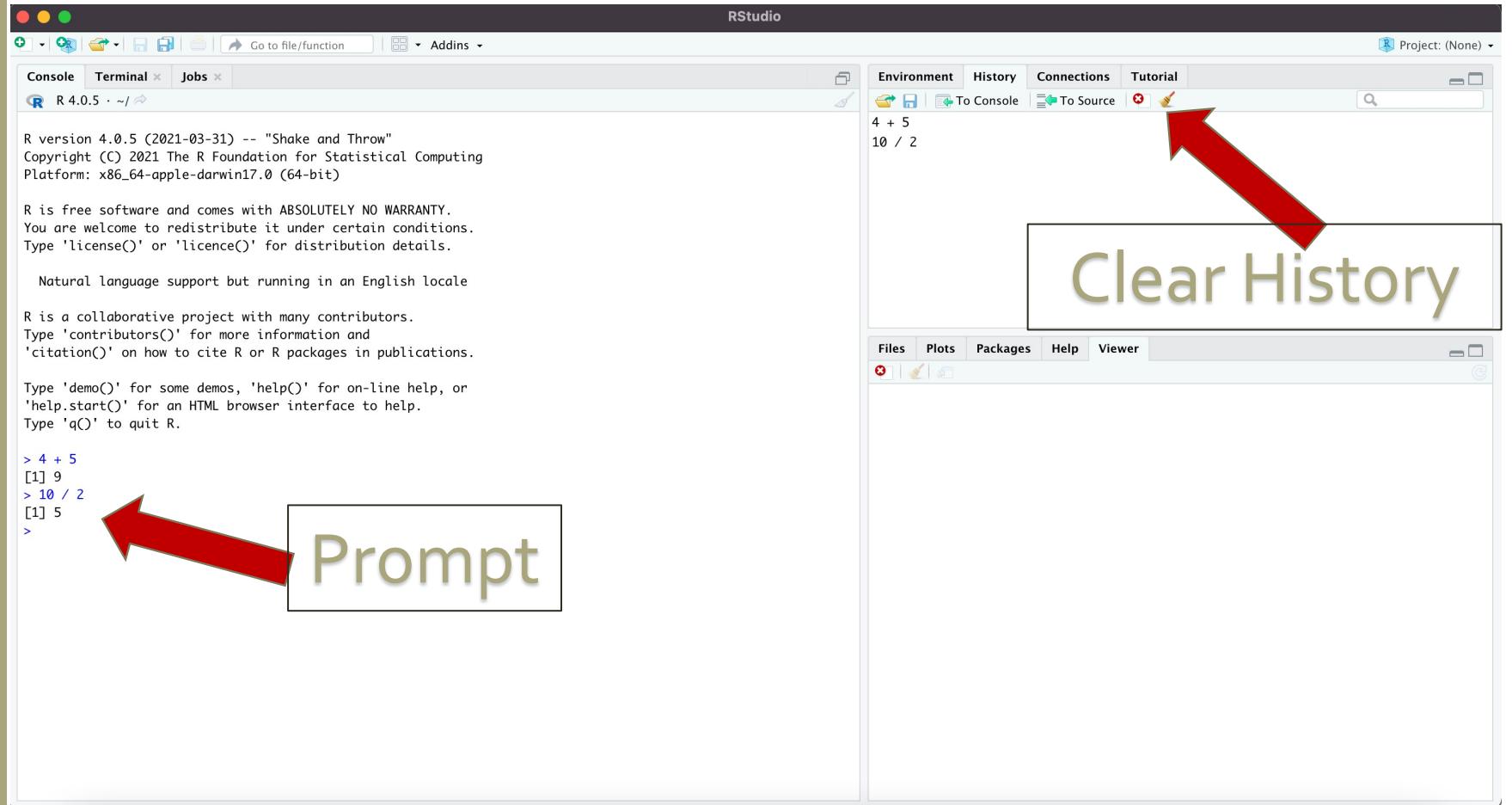
R Studio



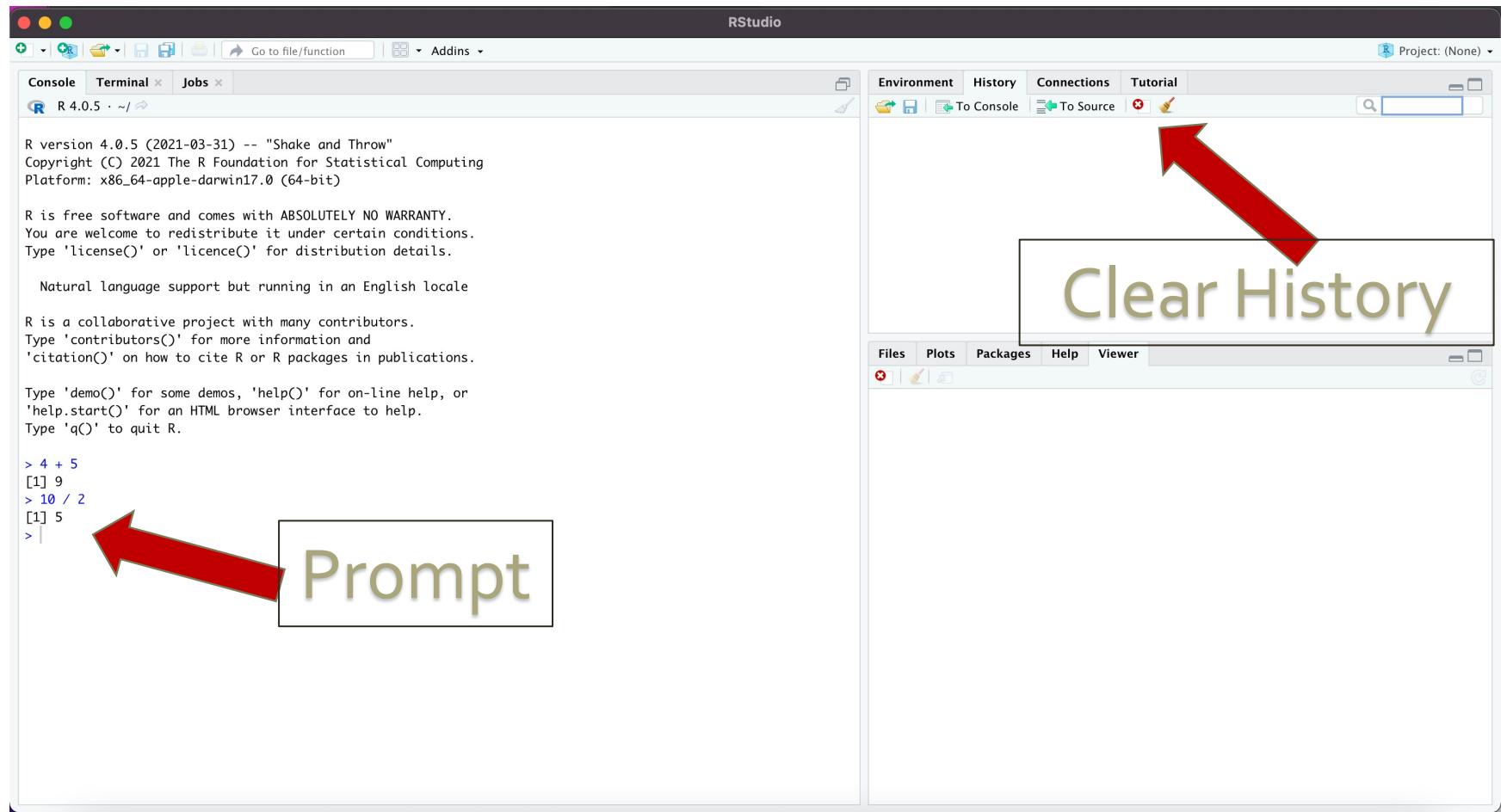
R Studio



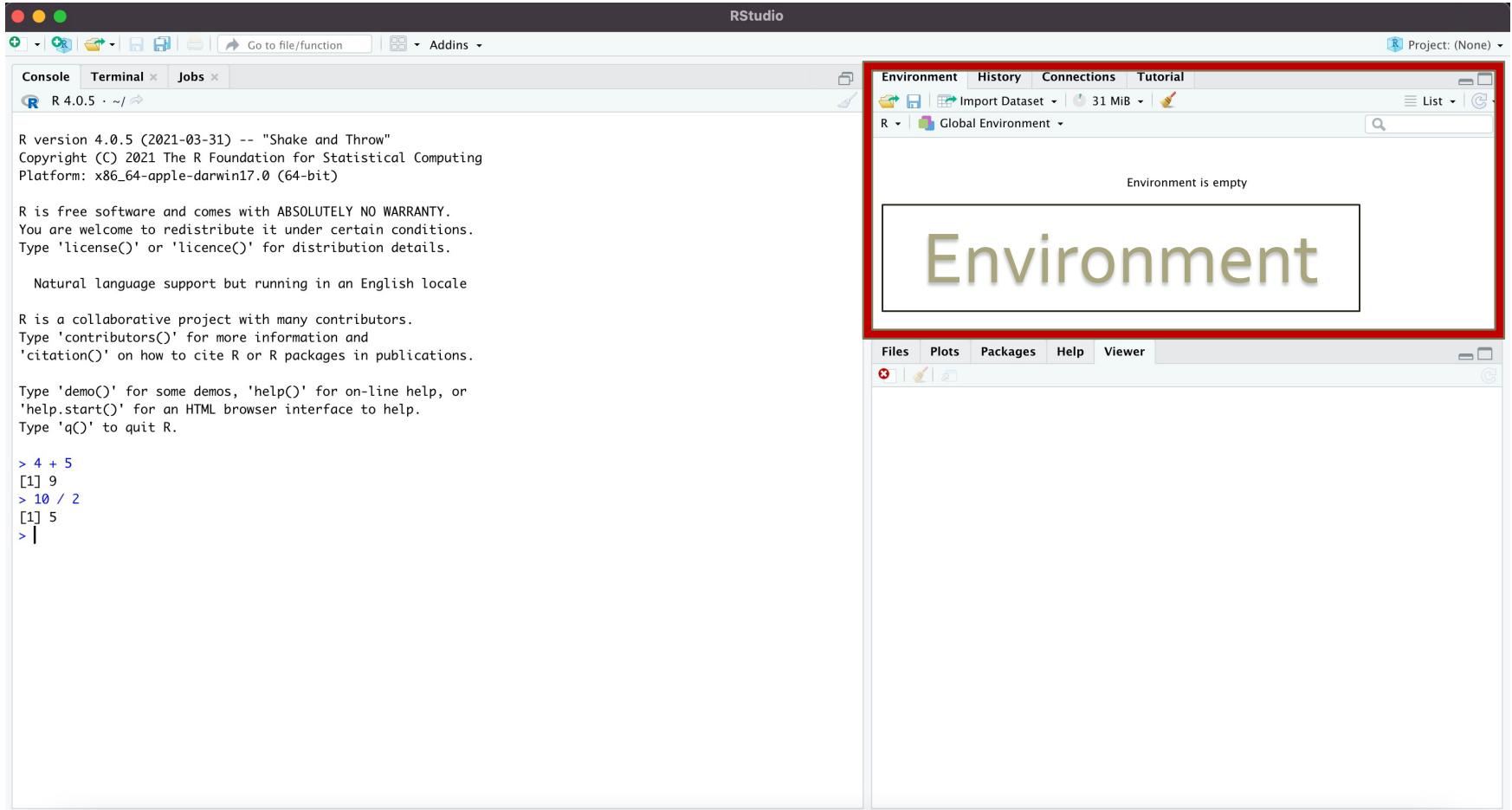
R Studio



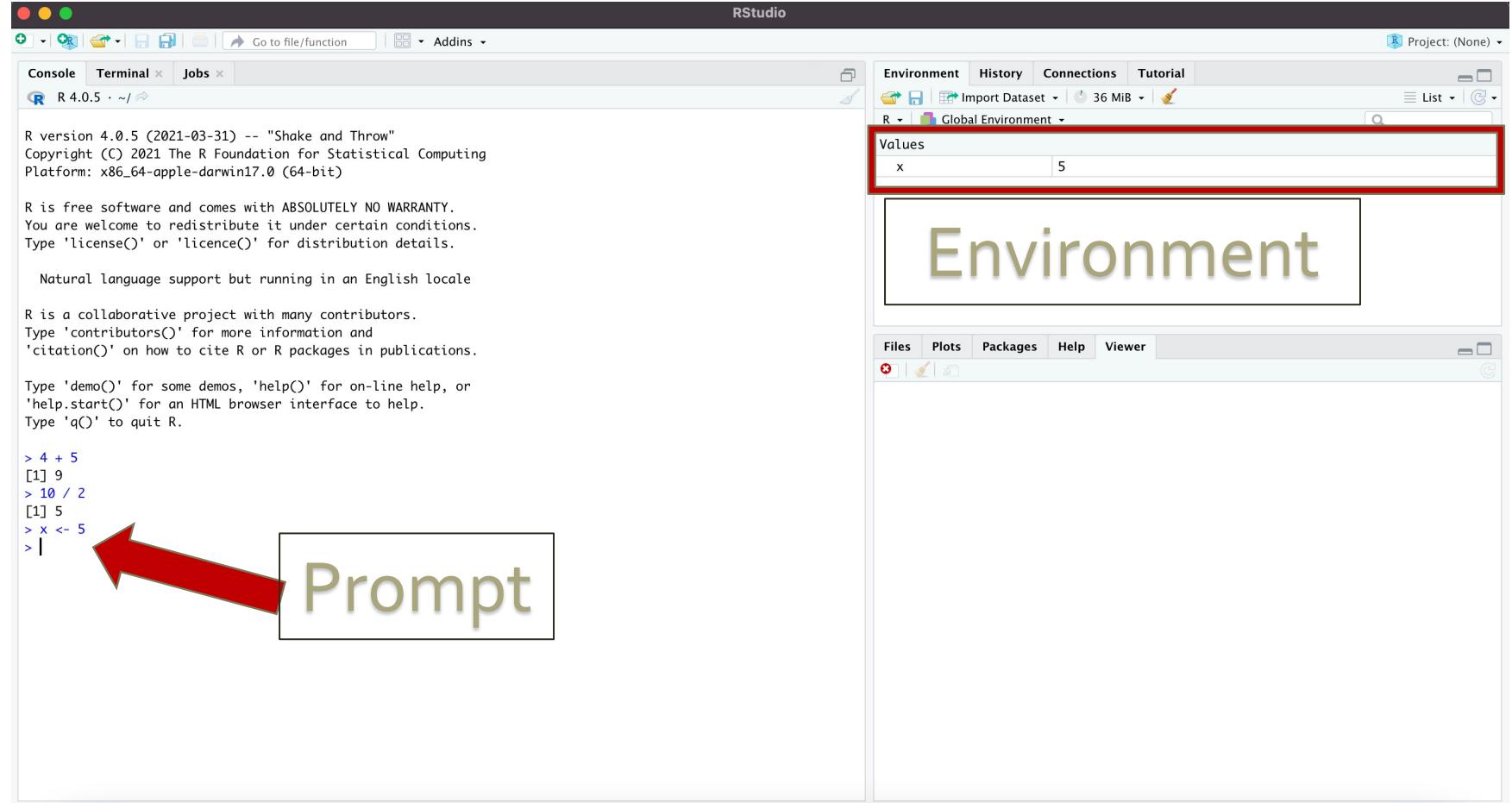
R Studio



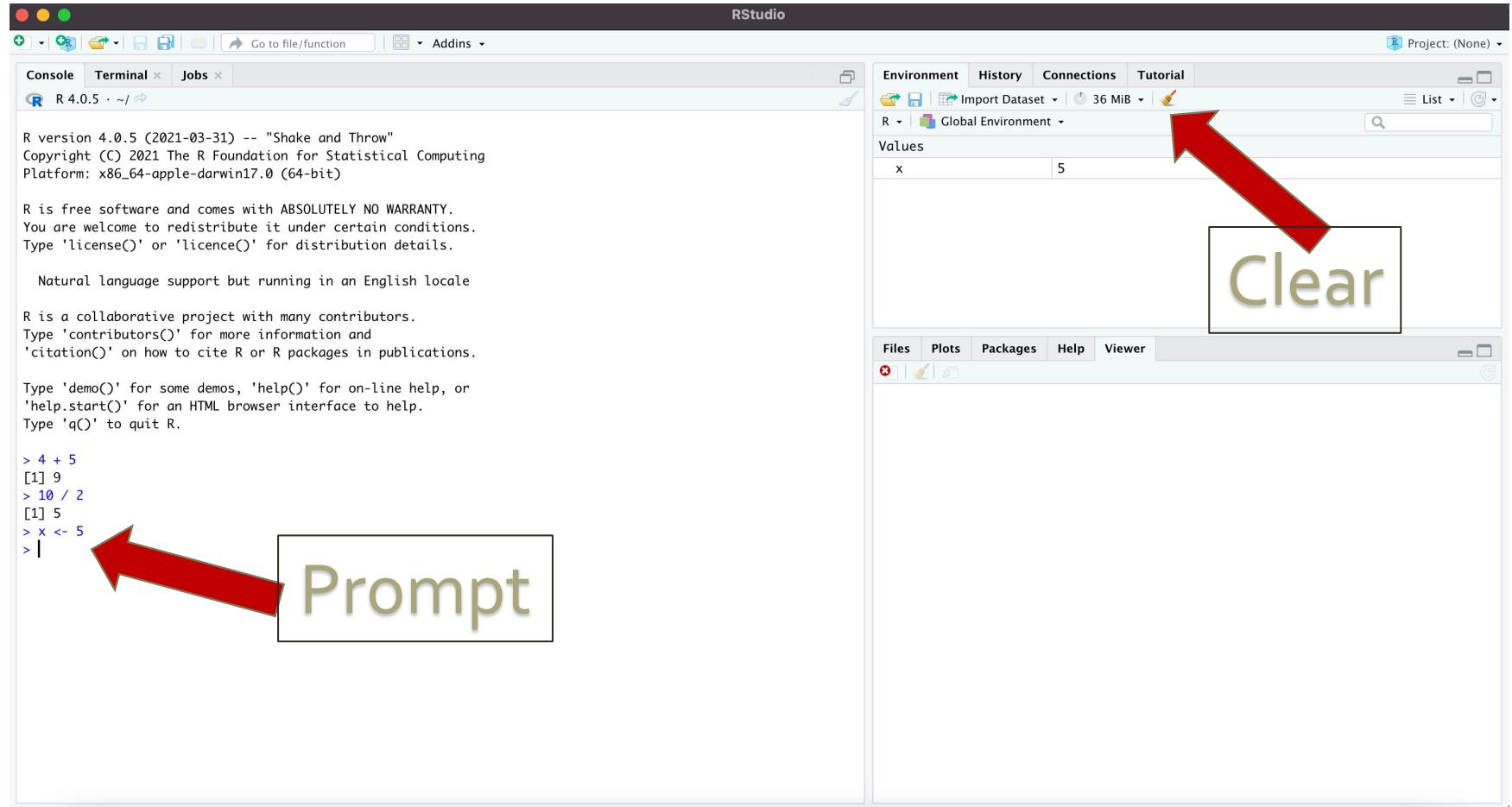
R Studio



R Studio



R Studio



R Studio

The screenshot shows the RStudio interface. On the left, there's a large green sidebar with the text "R Studio". The main window has several tabs at the top: Console, Terminal (selected), and Jobs. The Console tab displays the R startup message and some basic calculations:

```
R version 4.0.5 (2021-03-31) -- "Shake and Throw"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> 4 + 5
[1] 9
> 10 / 2
[1] 5
> x <- 5
> |
```

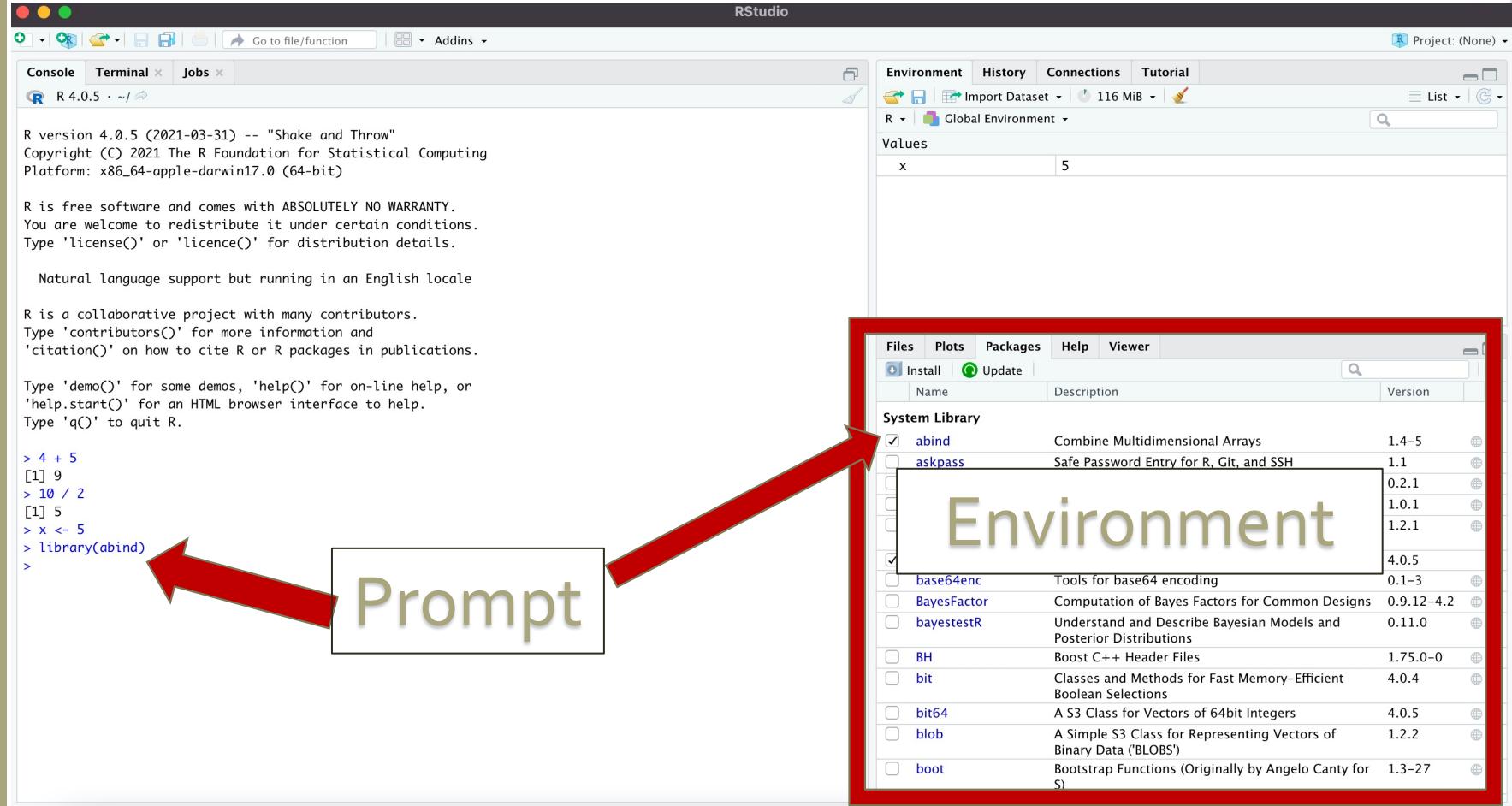
The Terminal tab shows the command "R 4.0.5". The Environment tab in the top navigation bar is selected. In the Environment pane, there's a table with one row:

Values	x	5
--------	---	---

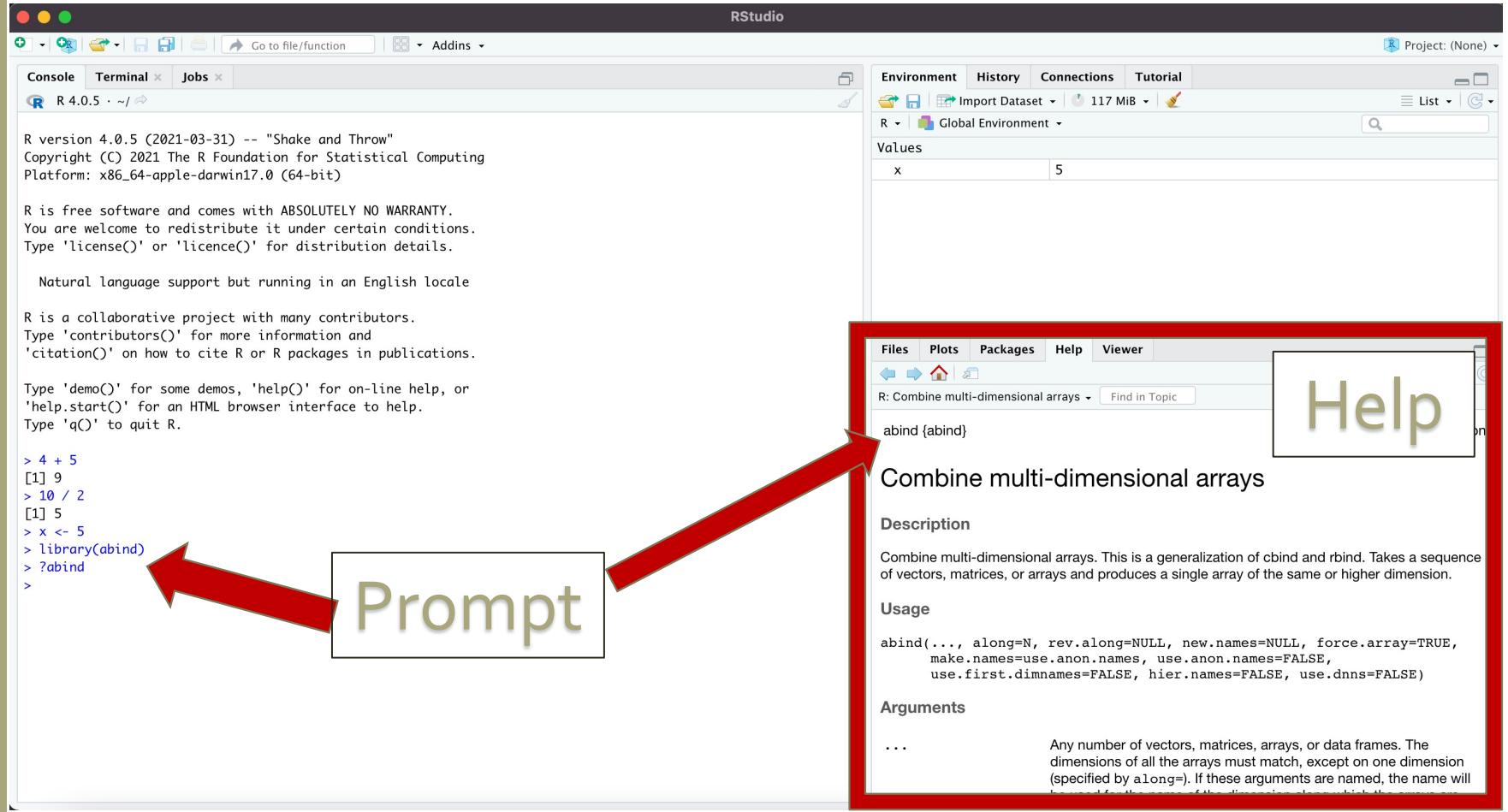
The Packages pane is open and highlighted with a red box. It lists the "System Library" with various packages and their versions. The package "base" is checked.

Name	Description	Version
abind	Combine Multidimensional Arrays	1.4-5
askpass	Safe Password Entry for R, Git, and SSH	1.1
base	(Selected)	0.5
base64enc	Tools for base64 encoding	0.1-3
BayesFactor	Computation of Bayes Factors for Common Designs	0.9.12-4.2
bayestestR	Understand and Describe Bayesian Models and Posterior Distributions	0.11.0
BH	Boost C++ Header Files	1.75.0-0
bit	Classes and Methods for Fast Memory-Efficient Boolean Selections	4.0.4
bit64	A S3 Class for Vectors of 64bit Integers	4.0.5
blob	A Simple S3 Class for Representing Vectors of Binary Data ("BLOBS")	1.2.2
boot	Bootstrap Functions (Originally by Angelo Canty for S)	1.3-27

R Studio



R Studio



R Markdown

The screenshot shows the RStudio interface. On the left, the 'Files' tab of the project browser is selected, showing a list of files including 'R Script', 'R Notebook', 'R Markdown...', 'Shiny Web App...', 'Plumber API...', 'Text File', 'C++ File', 'Python Script', 'SQL Script', 'Stan File', 'D3 Script', 'R Sweave', 'R HTML', and 'R Documentation...'. A file named 'abind.Rd' is currently open in the main editor area, displaying the R documentation for the 'abind' function. The documentation includes the function's purpose, usage, arguments, and examples. On the right, the 'Environment' tab of the tools panel is active, showing the variable 'x' with the value '5'. Below the environment, the 'Help' tab is active, showing the 'abind' documentation page. The page title is 'Combine multi-dimensional arrays', and it includes sections for 'Description', 'Usage', and 'Arguments'.

03-31) -- "Shake and Throw"
R Foundation for Statistical Computing
-darwin17.0 (64-bit)

comes with ABSOLUTELY NO WARRANTY.
distribute it under certain conditions.
licence()' for distribution details.

port but running in an English locale

project with many contributors.
for more information and
cite R or R packages in publications.

demos, 'help()' for on-line help, or
HTML browser interface to help.

```
> 4 + 5
[1] 9
> 10 / 2
[1] 5
> x <- 5
> library(abind)
> ?abind
> |
```

Environment History Connections Tutorial

Import Dataset 26 MB Global Environment

Values x 5

Files Plots Packages Help Viewer

R: Combine multi-dimensional arrays Find in Topic

abind {abind} R Documentation

Combine multi-dimensional arrays

Description

Combine multi-dimensional arrays. This is a generalization of cbind and rbind. Takes a sequence of vectors, matrices, or arrays and produces a single array of the same or higher dimension.

Usage

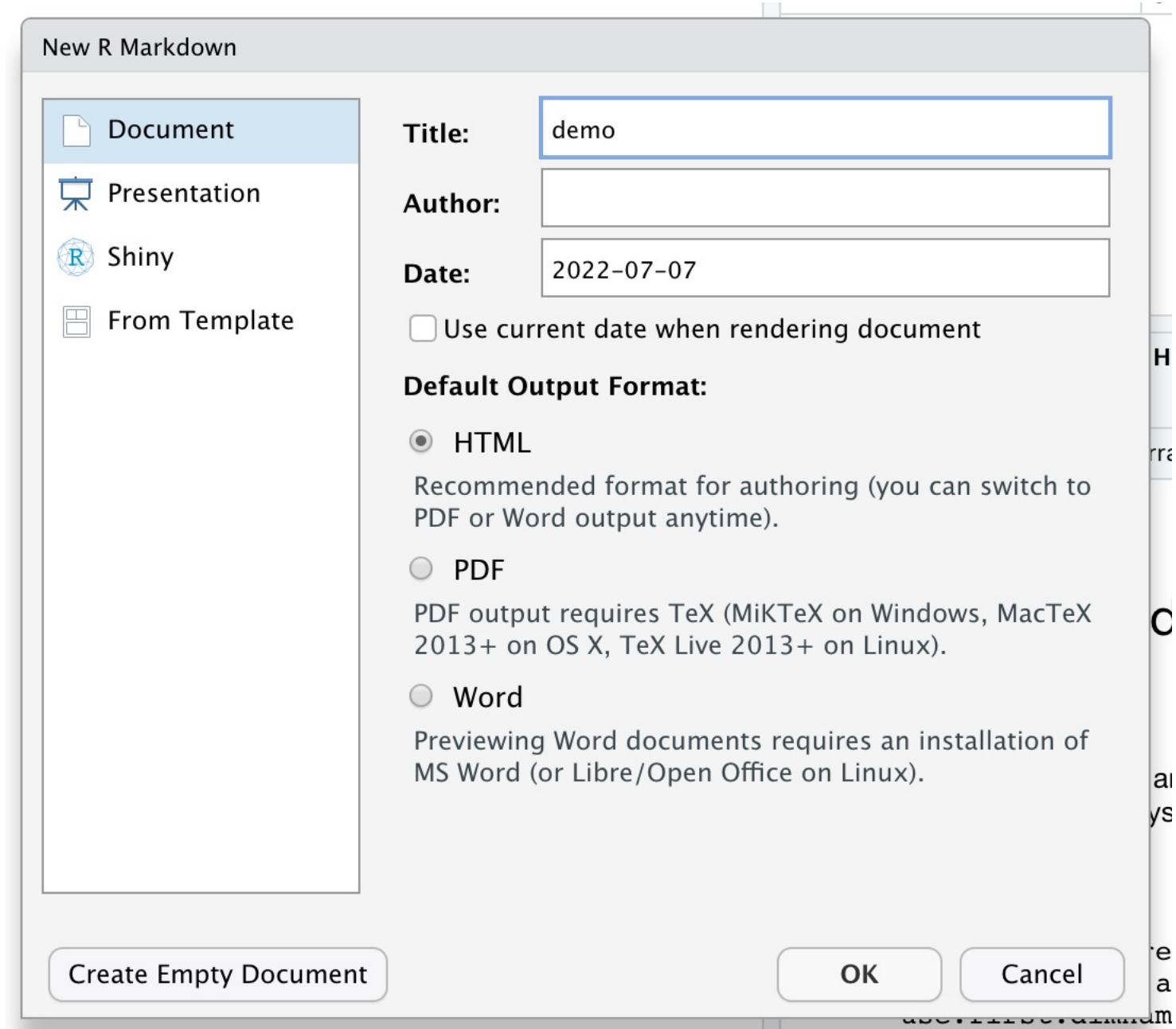
```
abind(..., along=N, rev.along=NULL, new.names=NULL, force.array=TRUE,  
      make.names=use.anon.names, use.anon.names=FALSE,  
      use.first.dimnames=FALSE, hier.names=FALSE, use.dnns=FALSE)
```

Arguments

...

Any number of vectors, matrices, arrays, or data frames. The dimensions of all the arrays must match, except on one dimension (specified by `along`). If these arguments are named, the name will be used for the name of the dimension along which the arrays are

R Markdown



R Markdown

The screenshot shows the RStudio interface. On the left, the 'Source' tab of the 'Untitled1' document is visible, containing R code for generating a Markdown file. A large preview window in the center displays the generated 'Markdown File' with the title 'Markdown File'. On the right, the 'Environment' pane shows a variable 'x' with the value 5. Below it, the 'Help' pane is open to the 'Combine multi-dimensional arrays' documentation, which includes sections for 'Description', 'Usage', and 'Arguments'.

```
1 ---  
2 title: "demo"  
3 output: html_document  
4 date: '2022-07-07'  
5 ---  
6  
7 ```{r setup, include=FALSE}  
8 knitr::opts_chunk$set(echo = TRUE)  
9 ```  
10  
11 ## R Markdown  
12  
13 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.  
14  
15 When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:  
16  
17 ```{r cars}  
18 demo  
19
```

R version 4.0.5 (2021-03-31) -- "Shake and Throw"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Markdown File

Environment History Connections Tutorial

Import Dataset 96 MiB

Global Environment

Values x 5

Files Plots Packages Help Viewer

R: Combine multi-dimensional arrays Find in Topic

abind (abind)

Combine multi-dimensional arrays

Description

Combine multi-dimensional arrays. This is a generalization of cbind and rbind. Takes a sequence of vectors, matrices, or arrays and produces a single array of the same or higher dimension.

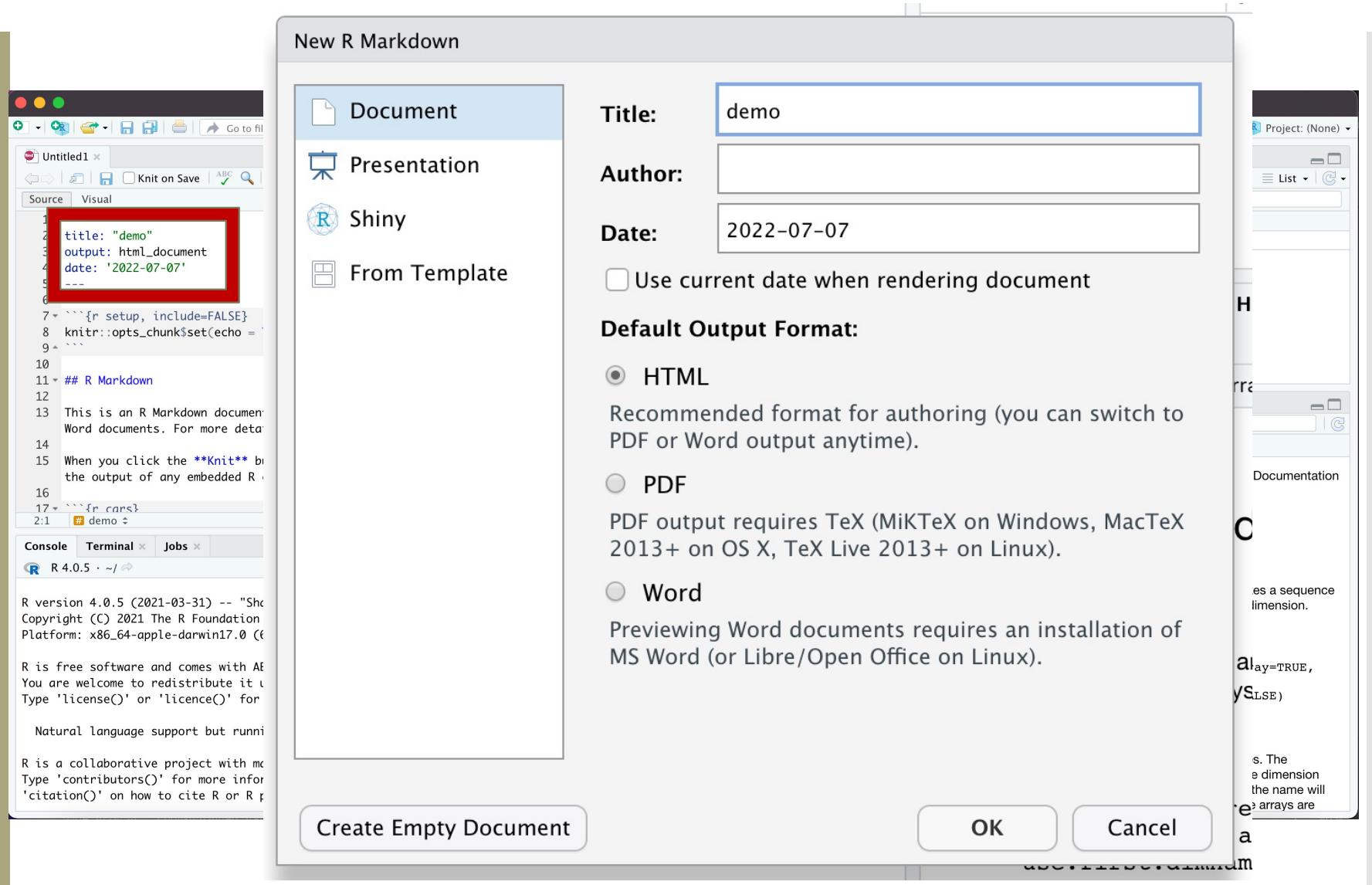
Usage

```
abind(..., along=N, rev.along=NULL, new.names=NULL, force.array=TRUE,  
make.names=use.anon.names, use.anon.names=FALSE,  
use.first.dimnames=FALSE, hier.names=FALSE, use.dnns=FALSE)
```

Arguments

... Any number of vectors, matrices, arrays, or data frames. The dimensions of all the arrays must match, except on one dimension (specified by along=). If these arguments are named, the name will be used for the name of the dimension along which the arrays are

R Markdown



R Markdown

The screenshot shows the RStudio interface with an R Markdown document titled "demo". The document contains two chunks: a "Code Chunk" and a "Text Chunk". The "Code Chunk" is enclosed in a red box and contains R code for setting options and loading the "cars" dataset. The "Text Chunk" is also enclosed in a red box and contains explanatory text about R Markdown and its capabilities. The RStudio environment includes a Source tab, a Visual tab, a Console tab showing R version 4.0.5 output, and a Help section for the "abind" function.

```
1 ---  
2 title: "demo"  
3 output: html_document  
4 date: '2022-07-07'  
5  
6 ```{r setup, include=FALSE}  
7 knitr::opts_chunk$set(echo = TRUE)  
8 ```  
9  
10  
11  
12  
13 This is an R Markdown document. Markdown is a simple way to authoring HTML, PDF, and MS Word documents. For more details on using R Markdown, see studio.com.  
14 When you click the **Knit** button a document will be generated that includes both content as well as  
15 the output of any embedded R code chunks within the document. You can embed an R code chunk like this:  
16  
17 ```{r cars}  
18 abind(demo$dist, demo$speed)  
19  
20
```

Code Chunk

Text Chunk

R version 4.0.5 (2021-03-31) -- "Shake and Throw"
Copyright (C) 2021 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Environment History Connections Tutorial

Import Dataset 96 MiB

Global Environment

Values

x 5

Files Plots Packages Help Viewer

R: Combine multi-dimensional arrays Find in Topic

abind (abind)

Combine multi-dimensional arrays

Description

Combine multi-dimensional arrays. This is a generalization of cbind and rbind. Takes a sequence of vectors, matrices, or arrays and produces a single array of the same or higher dimension.

Usage

abind(..., along=N, rev.along=NULL, new.names=NULL, force.array=TRUE,
make.names=use.anon.names, use.anon.names=FALSE,
use.first.dimnames=FALSE, hier.names=FALSE, use.dnns=FALSE)

Arguments

... Any number of vectors, matrices, arrays, or data frames. The dimensions of all the arrays must match, except on one dimension (specified by along=). If these arguments are named, the name will be used for the name of the dimension along which the arrays are

R Markdown

The screenshot shows the RStudio interface. On the left, the Source tab of an R Markdown document is visible, containing the following code:

```
R: <code>knitr::opts_chunk$set(echo = TRUE)</code>
```
R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:</pre>
```

A red box highlights the first code chunk, and a red arrow points from the text "Run Code Chunk" to the "Run" button in the toolbar above the code editor.

The right side of the interface shows the Environment pane with a variable `x` set to 5, and the Global Environment pane. Below the environment panes, the Help pane displays the documentation for the `abind` function, specifically the section "Combine multi-dimensional arrays".

**Run Code Chunk**

Environment History Connections Tutorial

Values

x 5

File Plots Packages Help Viewer

R: Combine multi-dimensional arrays Find in Topic

abind {abind}

Combine multi-dimensional arrays

Description

Combine multi-dimensional arrays. This is a generalization of cbind and rbind. Takes a sequence of vectors, matrices, or arrays and produces a single array of the same or higher dimension.

Usage

```
abind(..., along=N, rev.along=NULL, new.names=NULL, force.array=TRUE,
make.names=use.anon.names, use.anon.names=FALSE,
use.first.dimnames=FALSE, hier.names=FALSE, use.dnns=FALSE)
```

Arguments

... Any number of vectors, matrices, arrays, or data frames. The dimensions of all the arrays must match, except on one dimension (specified by `along`). If these arguments are named, the name will be used for the name of the dimension along which the arrays are

# R Markdown

The screenshot shows the RStudio interface with the following components:

- Source Editor:** Displays the R Markdown document. A red box highlights the R code chunk starting at line 18:7. Red arrows point from the bottom of the code chunk to the R console and the Global Environment pane.
- Console:** Shows the command `> y <- 6` and its result.
- Global Environment:** Shows variables `x` and `y` with values 5 and 6 respectively.
- Help:** The `abind` function is being documented. A red arrow points from the help text to the `abind` function definition in the code chunk.

```
source('Untitled1.Rmd')
#>
#> ## R Markdown
#>
#> This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.
#>
#> When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:
#>
#> {r}
#> y <- 6
#>
#> ## Including Plots
#>
#> You can also embed plots, for example:
#>
```

`R: Combine multi-dimensional arrays`   `Find in Topic`

`abind {abind}`   `R Documentation`

### Combine multi-dimensional arrays

**Description**

Combine multi-dimensional arrays. This is a generalization of cbind and rbind. Takes a sequence of vectors, matrices, or arrays and produces a single array of the same or higher dimension.

**Usage**

```
abind(..., along=N, rev.along=NULL, new.names=NULL, force.array=TRUE,
 make.names=use.anon.names, use.anon.names=FALSE,
 use.first.dimnames=FALSE, hier.names=FALSE, use.dnns=FALSE)
```

**Arguments**

... Any number of vectors, matrices, arrays, or data frames. The dimensions of all the arrays must match, except on one dimension (specified by `along`). If these arguments are named, the name will be used for the name of the dimension along which the arrays are

# R Markdown

The screenshot shows the RStudio interface with an R Markdown document open. The document contains the following code:

```
9 ---
10
11 ## R Markdown
12
13 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.
14
15 When you click the **Knit** button a document will be generated that includes both content as well as
the output of any embedded R code chunks within the document. You can embed an R code chunk like this:
16
17 ```{r}
18 y <- 6
19 ---
20
21 ## Including Plots
22
23 You can also embed plots, for example:
24
```

The "Knit" button in the toolbar is highlighted with a red box. The R console at the bottom shows the command `> y <- 6`.

The right side of the interface displays the "Environment" tab of the Project Explorer, showing variables `x` and `y` with values 5 and 6 respectively. Below the environment, the "Documentation" pane is open, showing the "abind" function documentation. The title is "Combine multi-dimensional arrays". The "Description" section states: "Combine multi-dimensional arrays. This is a generalization of cbind and rbind. Takes a sequence of vectors, matrices, or arrays and produces a single array of the same or higher dimension." The "Usage" section shows the function signature: `abind(..., along=N, rev.along=NULL, new.names=NULL, force.array=TRUE, make.names=use.anon.names, use.anon.names=FALSE, use.first.dimnames=FALSE, hier.names=FALSE, use.dnns=FALSE)`. The "Arguments" section indicates that the ellipsis (...) represents "Any number of vectors, matrices, arrays, or data frames. The dimensions of all the arrays must match, except on one dimension (specified by `along`). If these arguments are named, the name will be used for the name of the dimension along which the arrays are combined."

# R Markdown



The screenshot shows the RStudio interface with a dark theme. A sidebar on the left contains the title "R Markdown". The main workspace displays a preview of an R Markdown document titled "demo" last modified on "2022-07-07". The document content includes a section titled "R Markdown" which describes what it is and how to use it. It also shows a code chunk with the line "y <- 6" and a section titled "Including Plots" with a note about embedding plots.

demo

2022-07-07

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

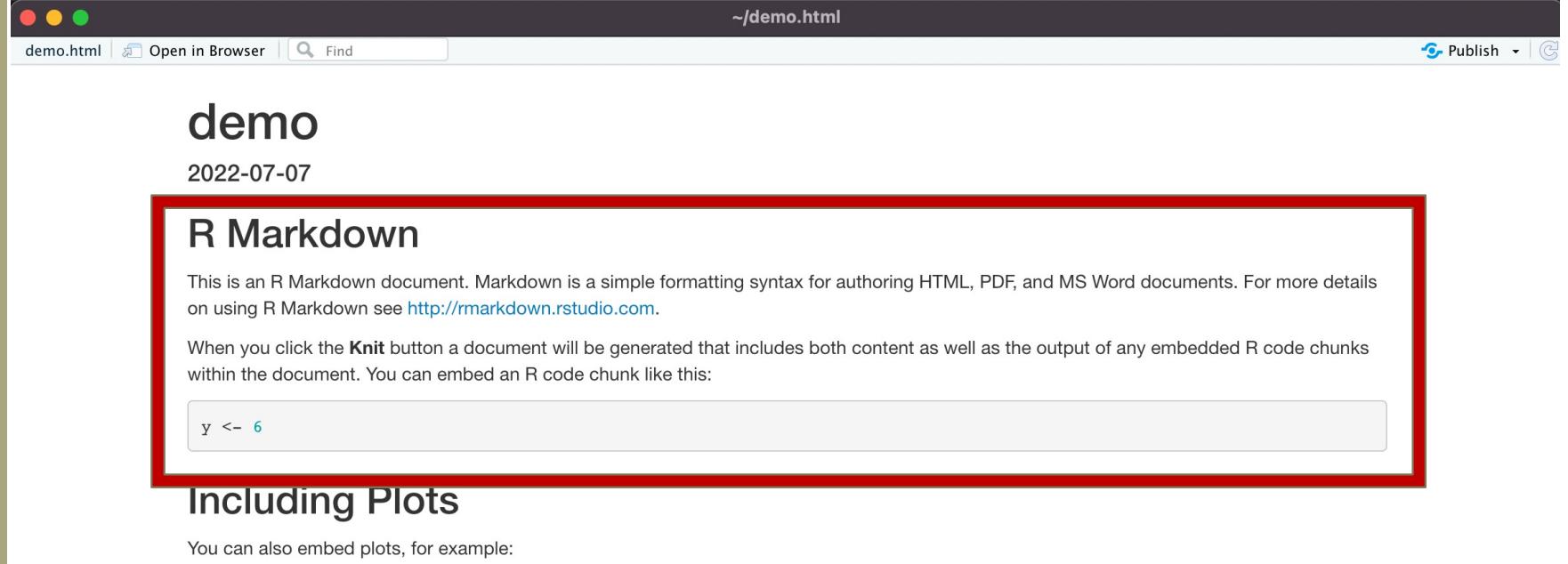
When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
y <- 6
```

## Including Plots

You can also embed plots, for example:

# R Markdown



The screenshot shows a web browser window displaying an R Markdown document titled "demo". The title "demo" is displayed prominently at the top, followed by the date "2022-07-07". A red rectangular box highlights a section of the content. Inside this box, the heading "R Markdown" is shown, followed by a paragraph explaining what R Markdown is and how to use it. Below this, there is a code block containing the R command "y <- 6". The entire content area is enclosed in a red border.

~/demo.html

demo.html | Open in Browser | Find

demo

2022-07-07

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
y <- 6
```

## Including Plots

You can also embed plots, for example:

# R Markdown

The screenshot illustrates the R Markdown workflow within RStudio. On the left, the RStudio interface shows an R Markdown source file named 'Untitled1.Rmd'. The code chunk at line 1 contains a simple assignment: `y <- 6`. The generated HTML output on the right is displayed in a browser window titled 'demo'.

**Generated HTML Output:**

# demo

2022-07-07

## R Markdown

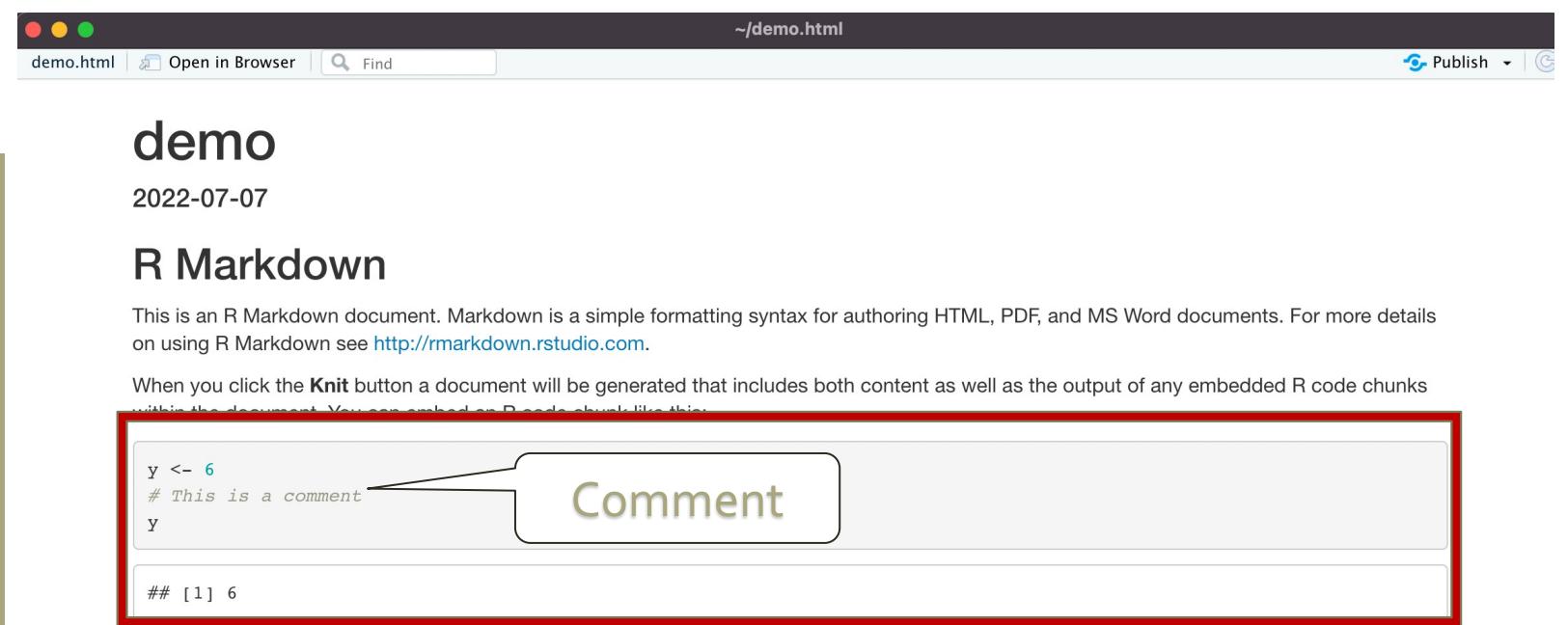
This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
y <- 6
```

**Including Plots**

The RStudio interface also shows the R console at the bottom, where the command `> y <- 6` has been run, and the Global Environment pane on the right, which displays the variable `y` with the value `6`.



# R Markdown

The screenshot shows the RStudio interface with the title bar "RStudio" and the project "Project: (None)". The left pane displays the R Markdown file "demo.Rmd" with the following content:10  
11 ## R Markdown  
12  
13 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <<http://rmarkdown.rstudio.com>>.  
14  
15 When you click the \*\*Knit\*\* button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:  
16  
17 ``{r}  
18 y <- 6  
19 # This is a comment  
20 y  
21 ``A red box highlights the code chunk and its output. A callout bubble labeled "Comment" points to the first line of the code chunk. The right pane shows the "Environment" tab with a variable "y" set to 6.

# R Markdown

- For quick tips and reference, try:
  - <https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf>