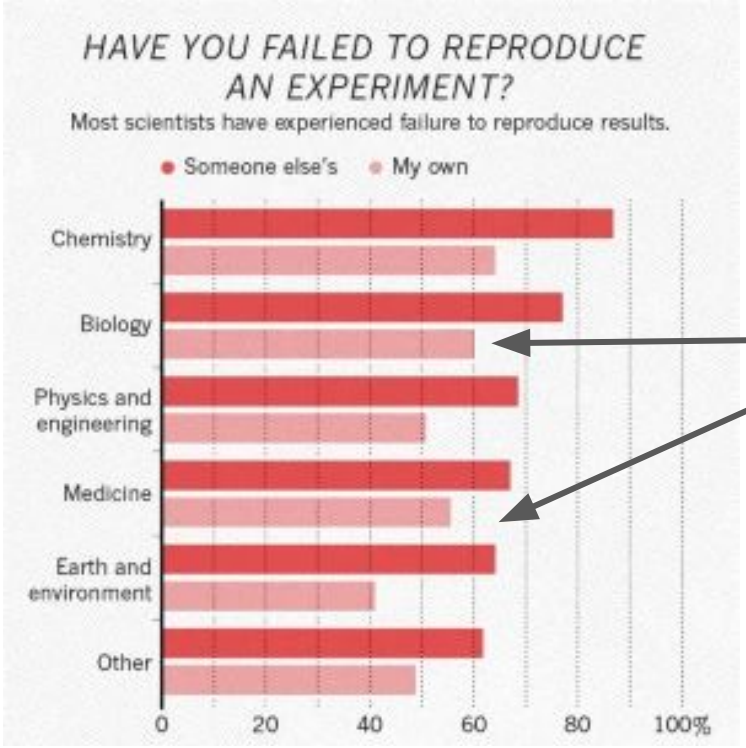
Introduction to R, RStudio, and RStudio Server

The CCDL

Who's been here before?

Reproducibility in 2016



55% and 60% of biologists and clinicians, respectively, could not reproduce *<u>their own</u>* results.

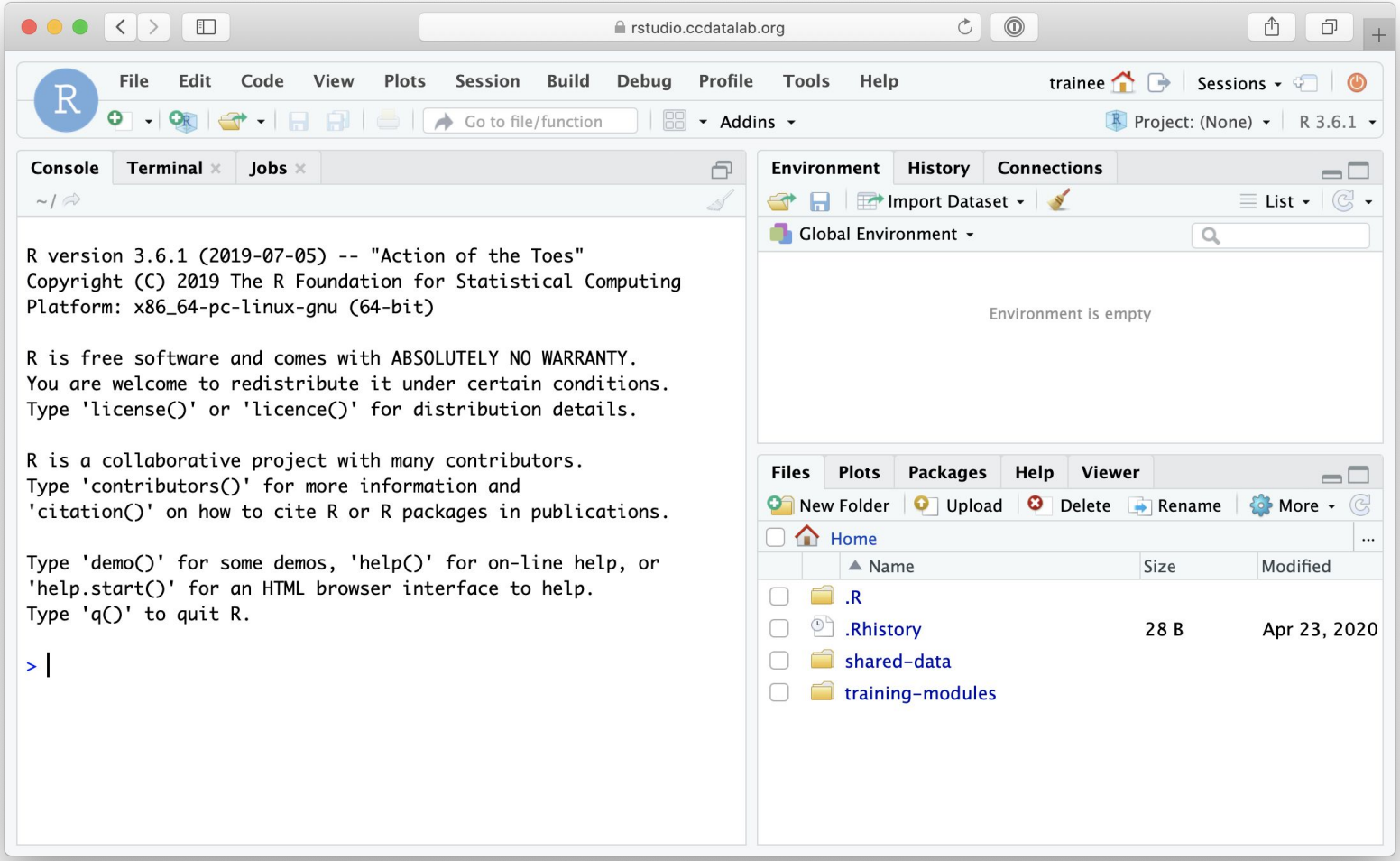
Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* 533, 452–454 (2016). https://doi.org/10.1038/533452a

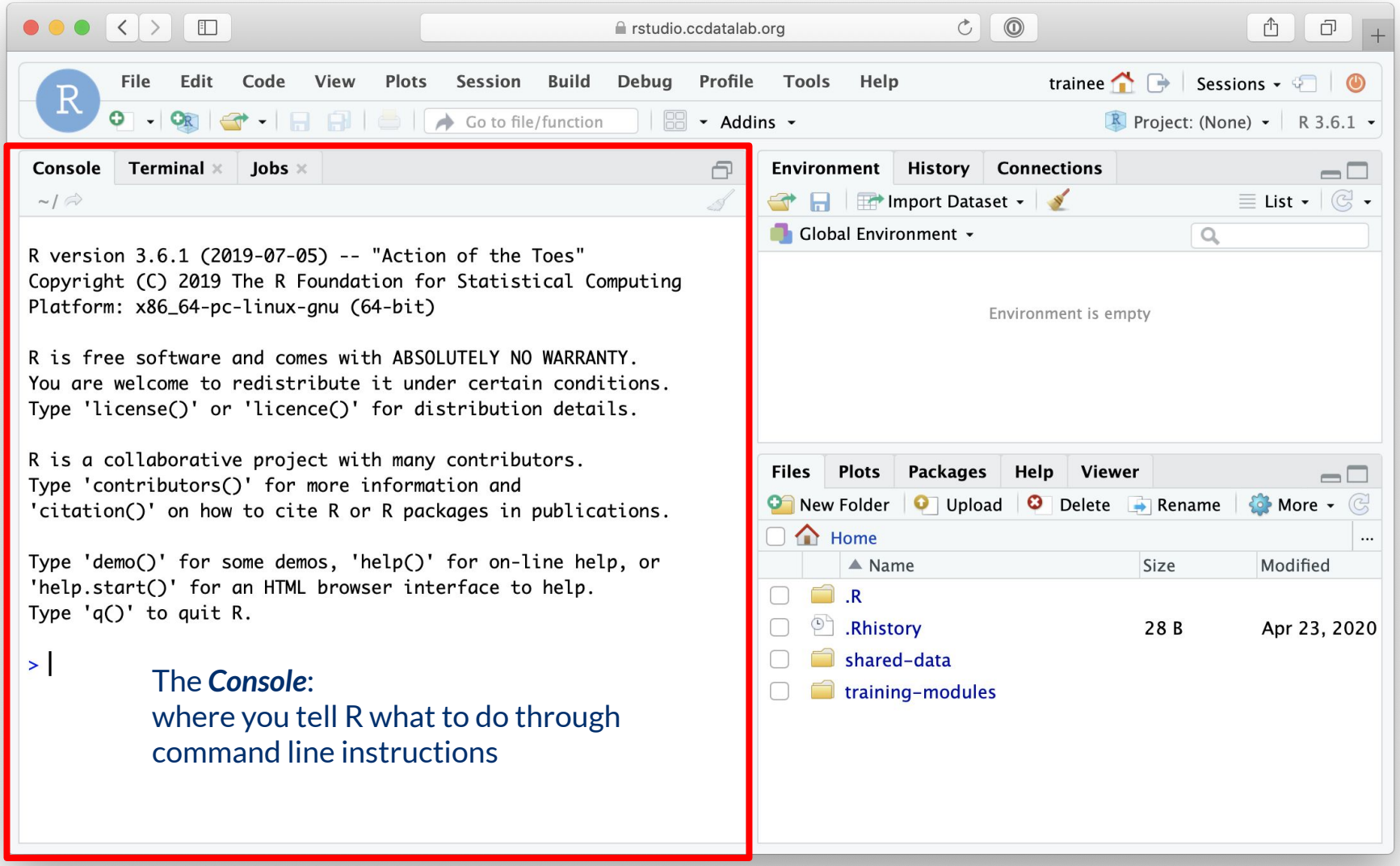
Command line vs GUI (graphics user interface)

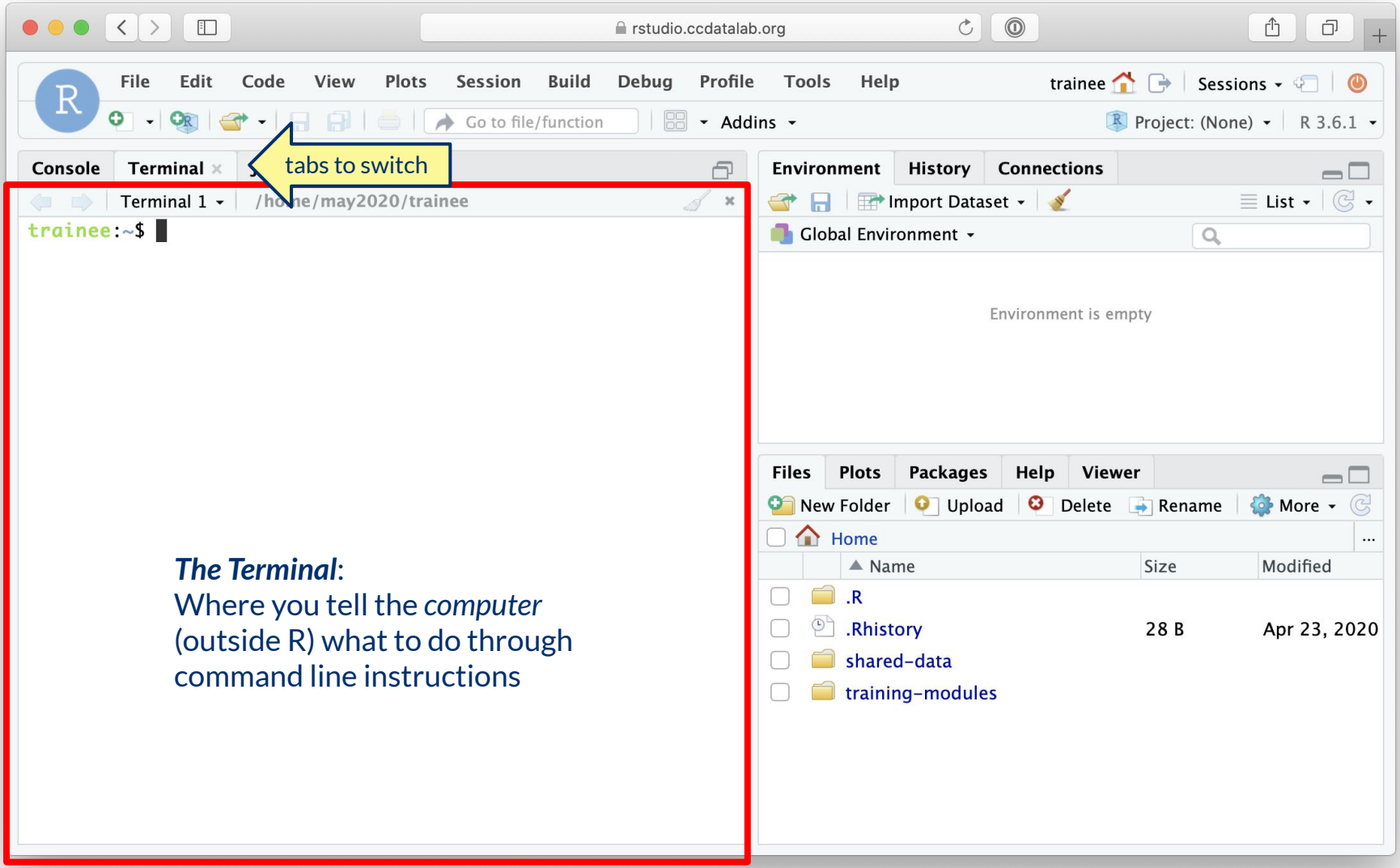
- An interface is how you interact with a program

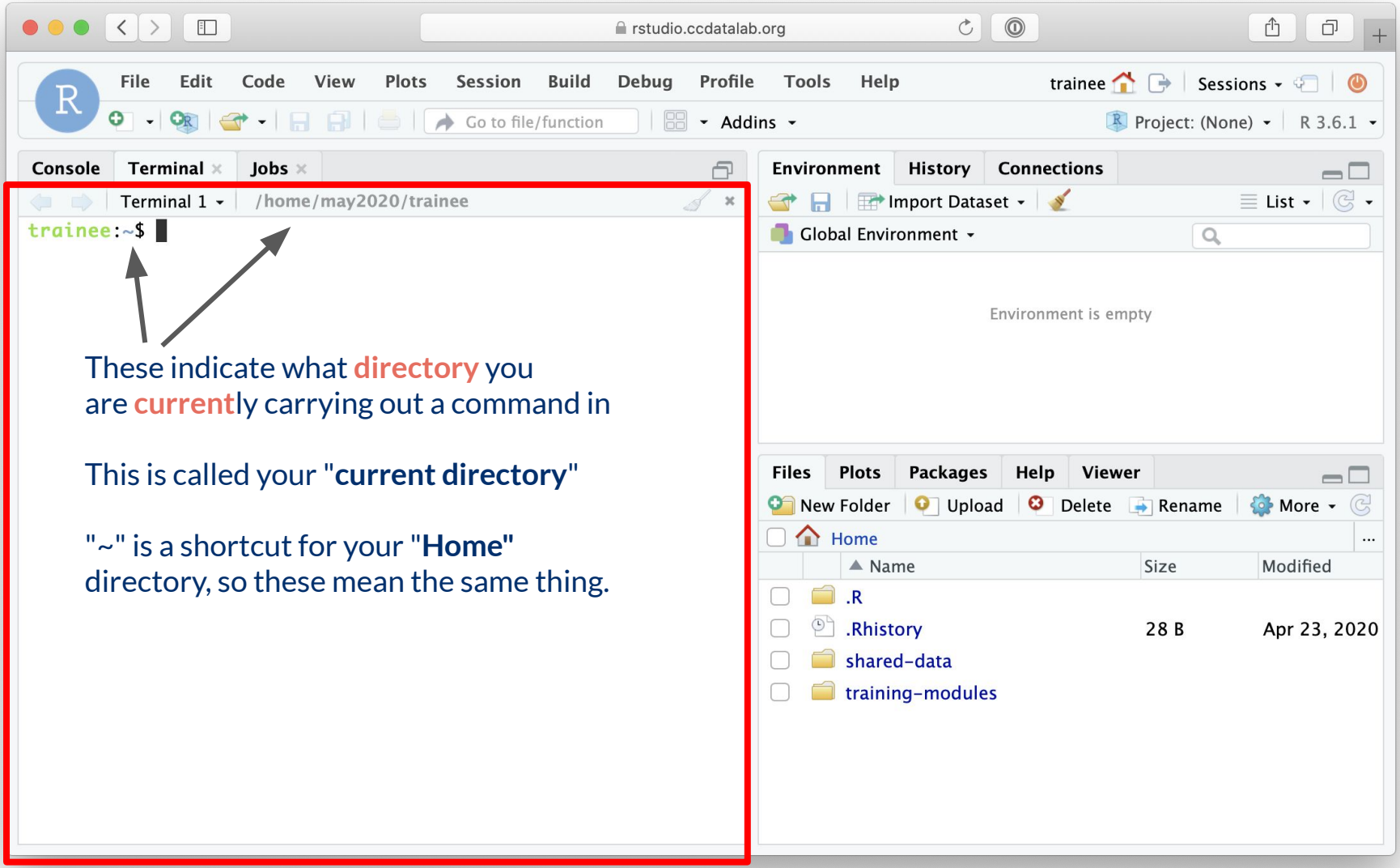
- GUI's have buttons you can *click* to do things, but…
- Command-line interfaces (CLI) have you *type* out things to do them

RStudio Server: A basic guide



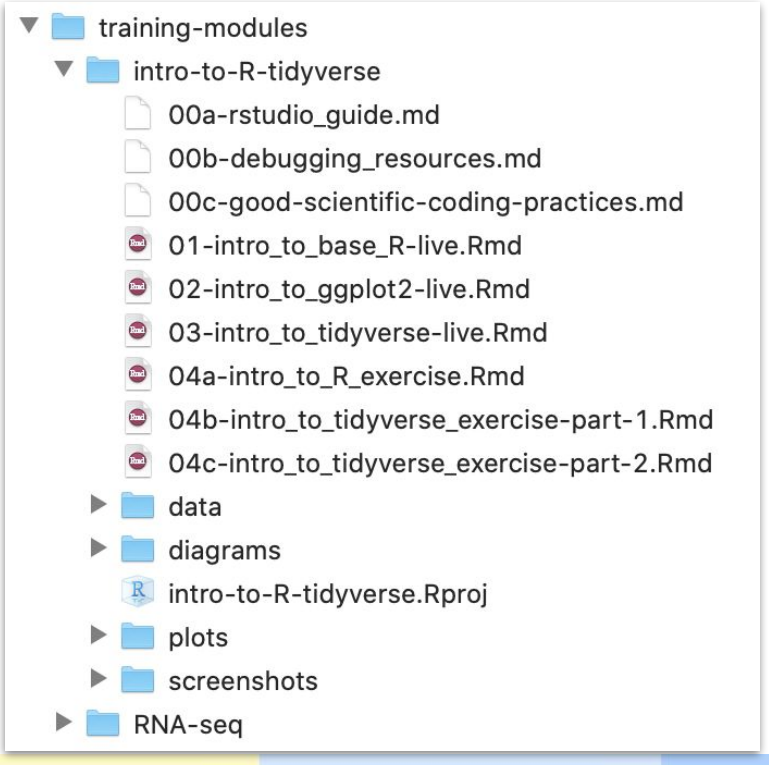

rstudio.ccdatalab.org
File Edit Code View Plots Session Build Debug Profile Tools Help
trainee
Sessions
Go to file/function
Addins
Project: (None)
R 3.6.1
Console
Terminal
Jobs
~/
R version 3.6.1 (2019-07-05) -- "Action of the Toes"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (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.
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.
>
The Console:
where you tell R what to do through
command line instructions
Environment
History
Connections
Import Dataset
List
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
New Folder
Upload
Delete
Rename
More
Home
Name
Size
Modified
.R
.Rhistory
28 B
Apr 23, 2020
shared-data
training-modules

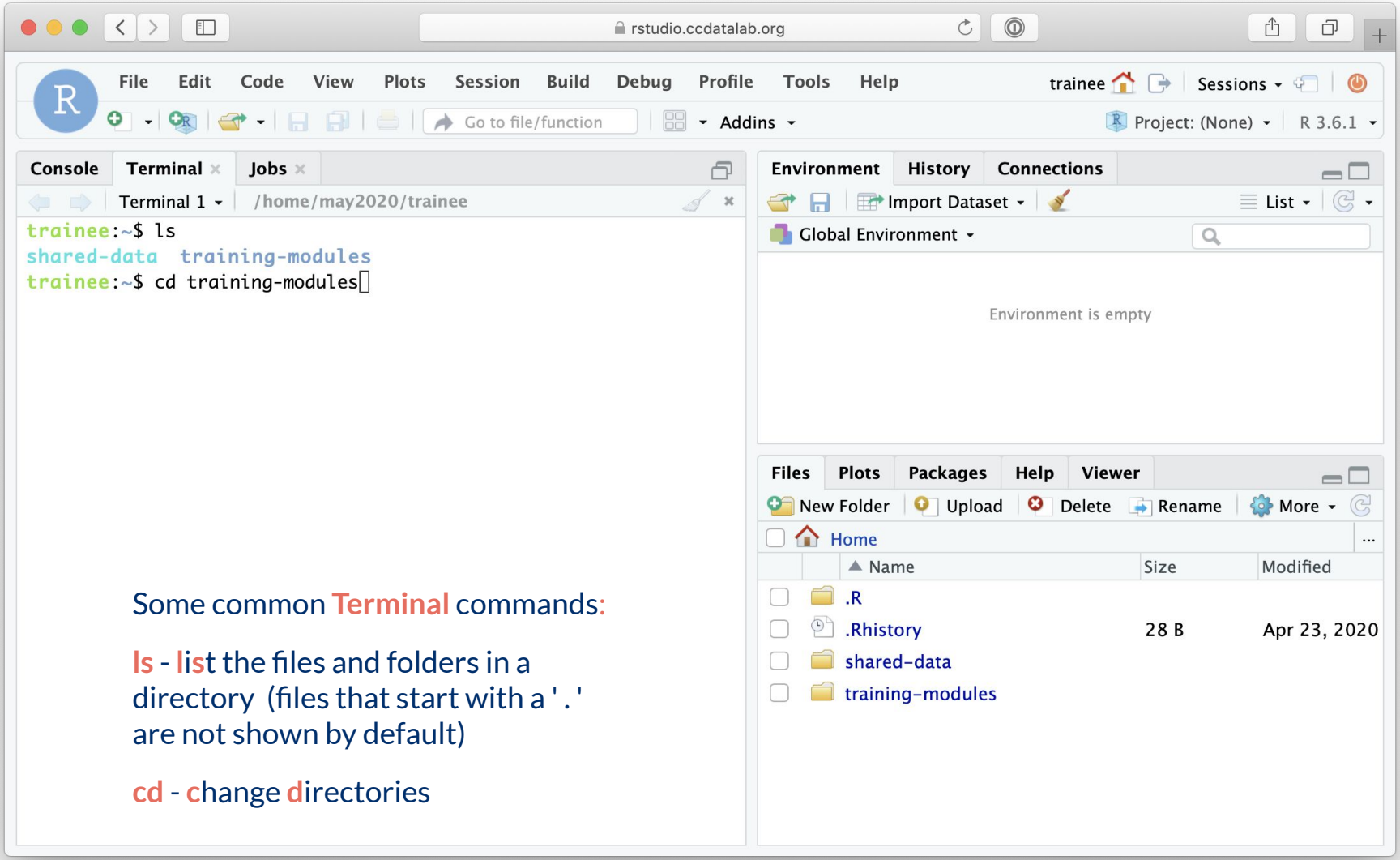

rstudio.ccdatalab.org
File Edit Code View Plots Session Build Debug Profile Tools Help
trainee
Sessions
Go to file/function
Addins
Project: (None)
R 3.6.1
Console
Terminal
tabs to switch
Terminal 1
/home/may2020/trainee
trainee:~$
The Terminal:
Where you tell the computer
(outside R) what to do through
command line instructions
Environment
History
Connections
Import Dataset
List
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
New Folder
Upload
Delete
Rename
More
Home
Name
Size
Modified
.R
.Rhistory
28 B
Apr 23, 2020
shared-data
training-modules

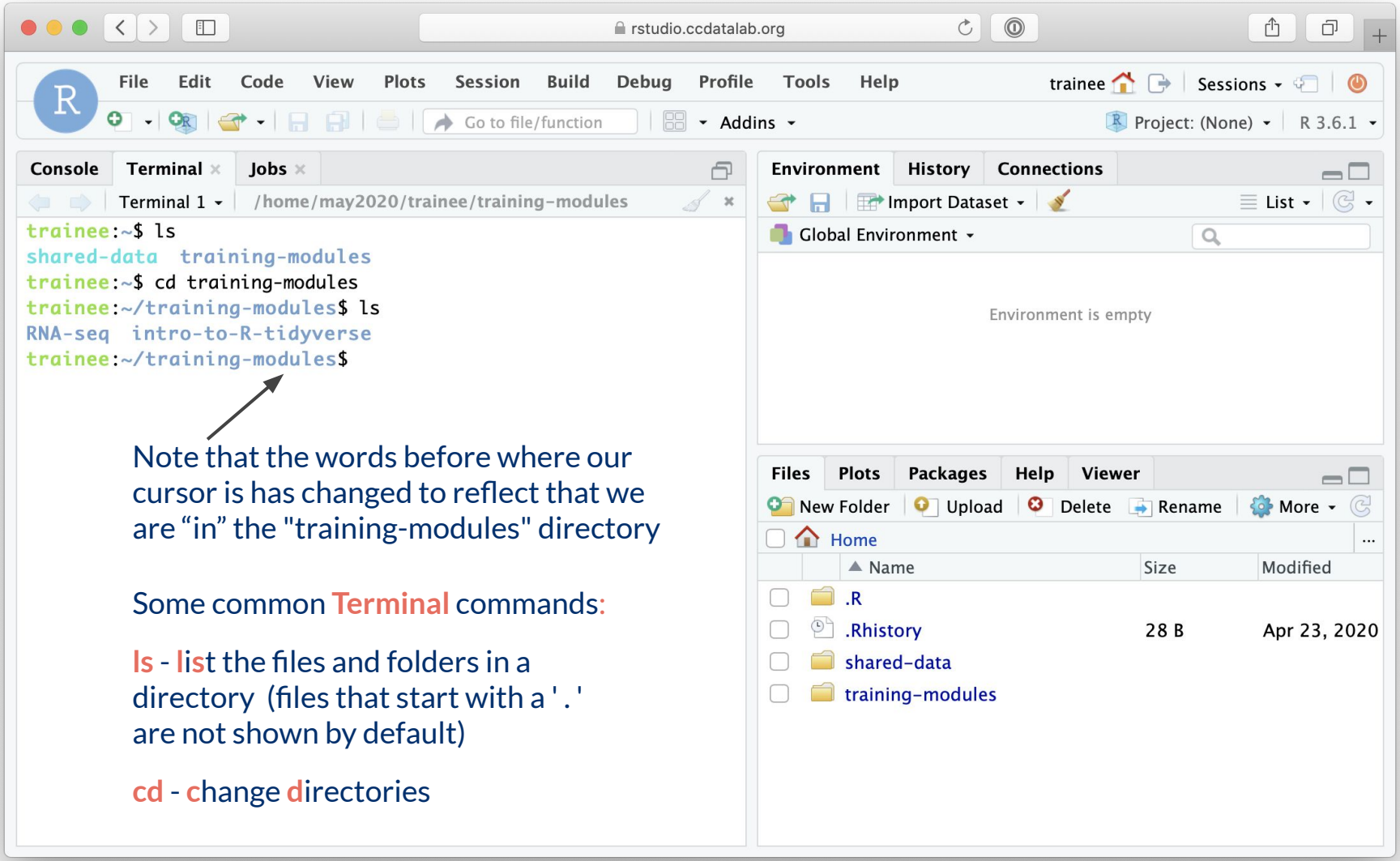

rstudio.ccdatalab.org
File Edit Code View Plots Session Build Debug Profile Tools Help
trainee
Sessions
Go to file/function
Addins
Project: (None)
R 3.6.1
Console
Terminal
Jobs
Terminal 1
/home/may2020/trainee
trainee:~$
These indicate what directory you are currently carrying out a command in
This is called your "current directory"
"~" is a shortcut for your "Home" directory, so these mean the same thing.
Environment
History
Connections
Import Dataset
List
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
New Folder
Upload
Delete
Rename
More
Home
Name
Size
Modified
.R
.Rhistory
28 B
Apr 23, 2020
shared-data
training-modules

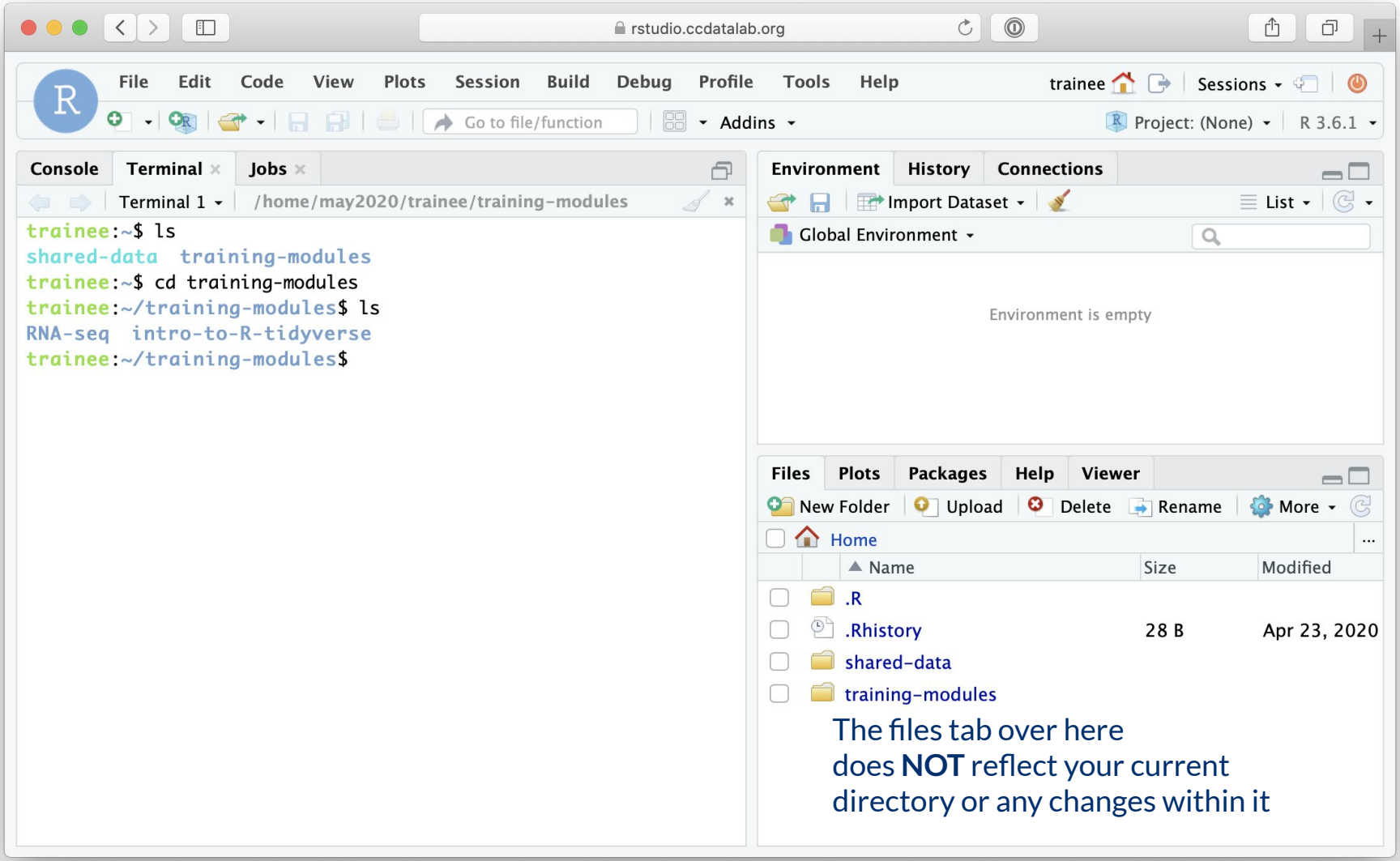
Directories = Folders

When we are working on the command line, we have to keep track of where the files we are using are being kept.



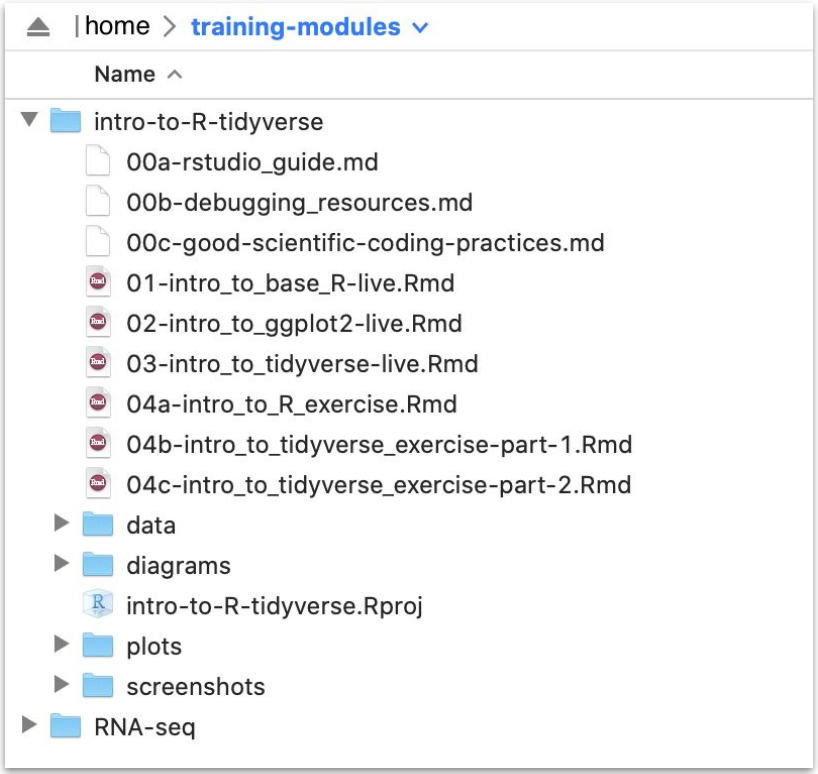

rstudio.ccdatalab.org
File Edit Code View Plots Session Build Debug Profile Tools Help
trainee
Sessions
Go to file/function
Addins
Project: (None)
R 3.6.1
Console
Terminal
Jobs
Terminal 1
/home/may2020/trainee
trainee:~$ ls
shared-data training-modules
trainee:~$ cd training-modules
Environment
History
Connections
Import Dataset
List
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
New Folder
Upload
Delete
Rename
More
Home
Name
Size
Modified
.R
.Rhistory
28 B
Apr 23, 2020
shared-data
training-modules
Some common Terminal commands:
ls - list the files and folders in a directory (files that start with a '.' are not shown by default)
cd - change directories


rstudio.ccdatalab.org
File Edit Code View Plots Session Build Debug Profile Tools Help
trainee
Sessions
Go to file/function
Addins
Project: (None)
R 3.6.1
Console
Terminal
Jobs
Terminal 1
/home/may2020/trainee/training-modules
trainee:~$ ls
shared-data training-modules
trainee:~$ cd training-modules
trainee:~/training-modules$ ls
RNA-seq intro-to-R-tidyverse
trainee:~/training-modules$
Environment
History
Connections
Import Dataset
List
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
New Folder
Upload
Delete
Rename
More
Home
Name
Size
Modified
.R
.Rhistory
28 B
Apr 23, 2020
shared-data
training-modules
Note that the words before where our cursor is has changed to reflect that we are “in” the "training-modules" directory
Some common Terminal commands:
ls - list the files and folders in a directory (files that start with a '.' are not shown by default)
cd - change directories


rstudio.ccdatalab.org
File Edit Code View Plots Session Build Debug Profile Tools Help
trainee
Sessions
Go to file/function
Addins
Project: (None)
R 3.6.1
Console
Terminal
Jobs
Terminal 1
/home/may2020/trainee/training-modules
trainee:~$ ls
shared-data training-modules
trainee:~$ cd training-modules
trainee:~/training-modules$ ls
RNA-seq intro-to-R-tidyverse
trainee:~/training-modules$
Environment
History
Connections
Import Dataset
List
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
New Folder
Upload
Delete
Rename
More
Home
Name
Size
Modified
.R
.Rhistory
28 B
Apr 23, 2020
shared-data
training-modules
The files tab over here
does NOT reflect your current
directory or any changes within it

File paths: Directions to a file or folder

Let's say we want access to "01-intro_to_base_R-live.Rmd"

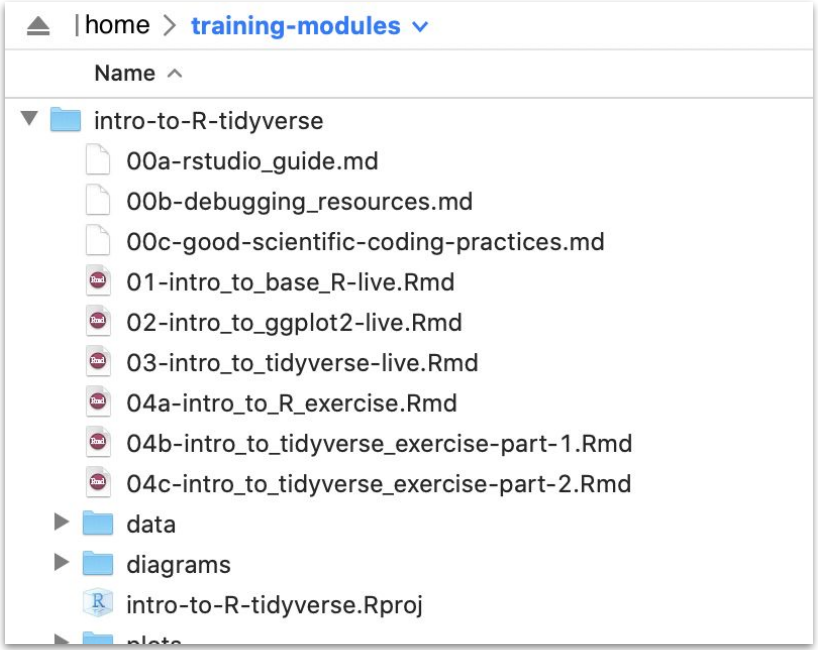


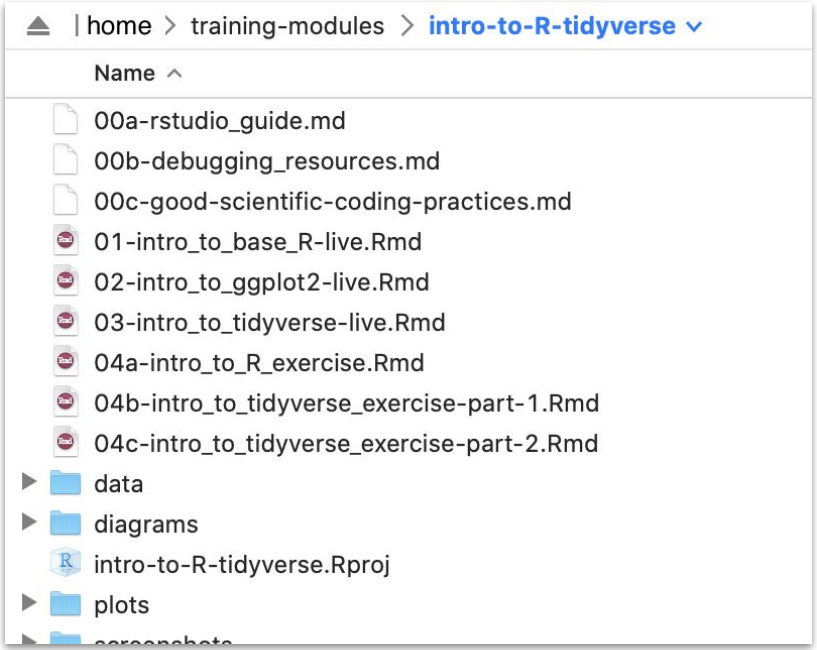
Current directory = "training-modules"

File path = "intro-to-R-tidyverse/01-intro_to_base_R-live.Rmd"

Relative file paths

Let's say we want access to "01-intro_to_base_R-live.Rmd"





Current directory = "training-modules"

Relative file path = "intro-to-R-tidyverse/01-intro_to_base_R-live.Rmd"

Current directory = "training-modules/intro-to-R-tidyverse"

Relative file path = "01-intro_to_base_R-live.Rmd"

Introduction to R

The CCDL

R programming

Programming: making executable scripts for accomplishing a task
(in this case, data analysis is our task)

Scripts allow others to see, step-by-step, what you did.

Why we use R:

- It's free and open-source
- People make cool packages that do stuff for us
- Many researchers in genomics use it (as well as Python)

One in five genetics papers contains errors thanks to Microsoft Excel

By Jessica Boddy | Aug. 29, 2016 , 1:45 PM

What you type	What you see	How Excel stores it
MARCH1	1-MAR	42430
SEPT2	2-SEP	42615

https://www.sciencemag.org/news/2016/08/one-five-genetics-papers-contains-errors-thanks-microsoft-excel
Ziemann et al. Genome Biology (2016) 17:177 DOI 10.1186/s13059-016-1044-7

R, RStudio, and RStudio Server

R is a statistical programming language.

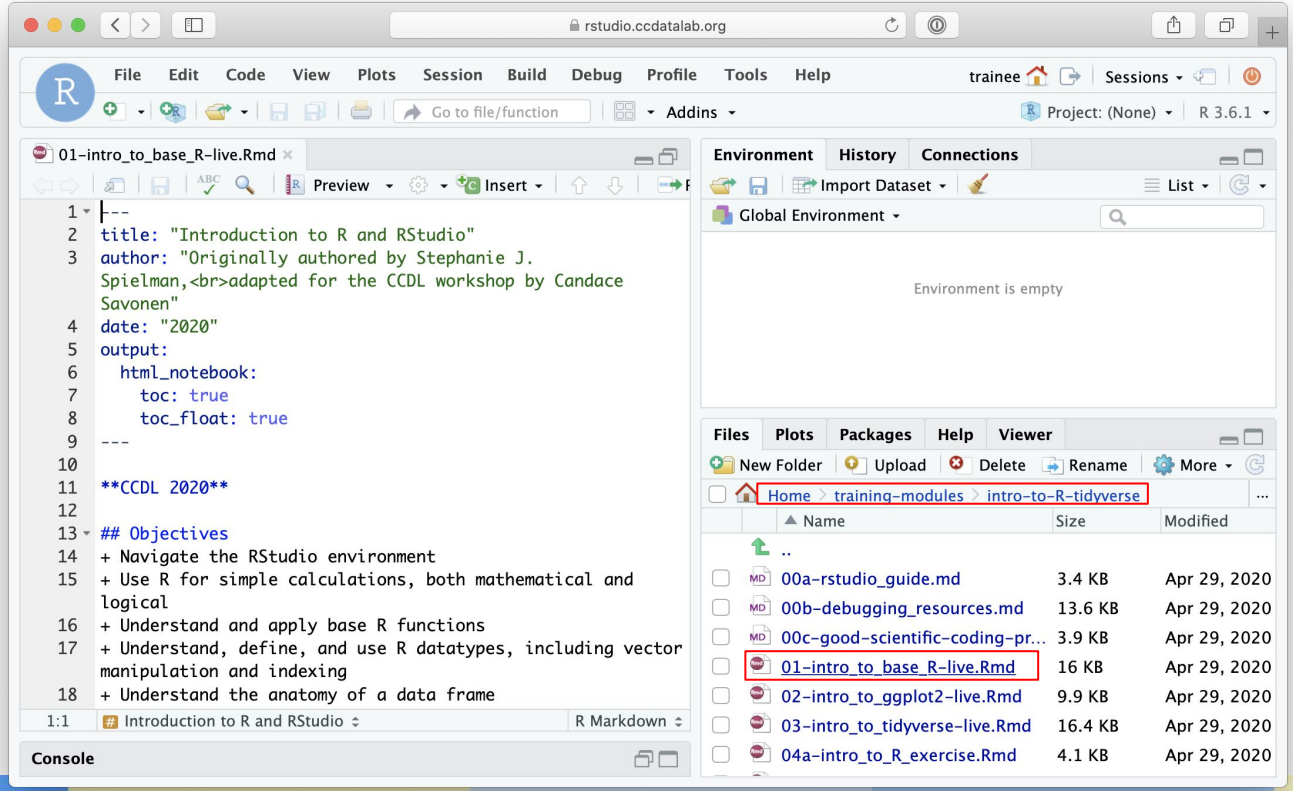
RStudio is an IDE for working in R

- IDE: Integrated Development Environment
- We write R code using the (free!) RStudio IDE

RStudio Server allows us to run the **RStudio IDE** from a browser

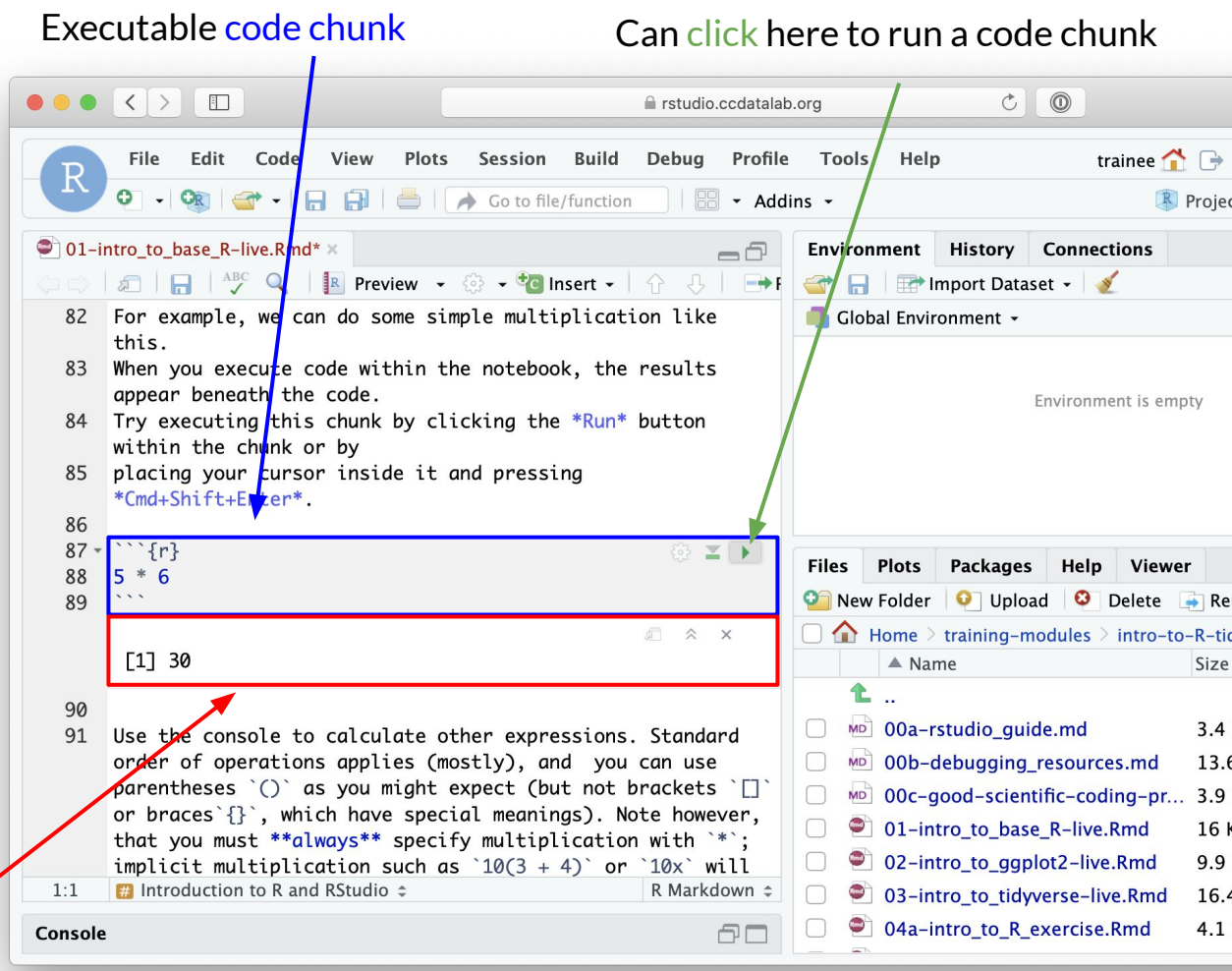
R Notebooks

Use the "**Files**" tab to open: training-modules/intro-to-R-tidyverse/01-intro_to_base_R-live.Rmd



R Notebooks

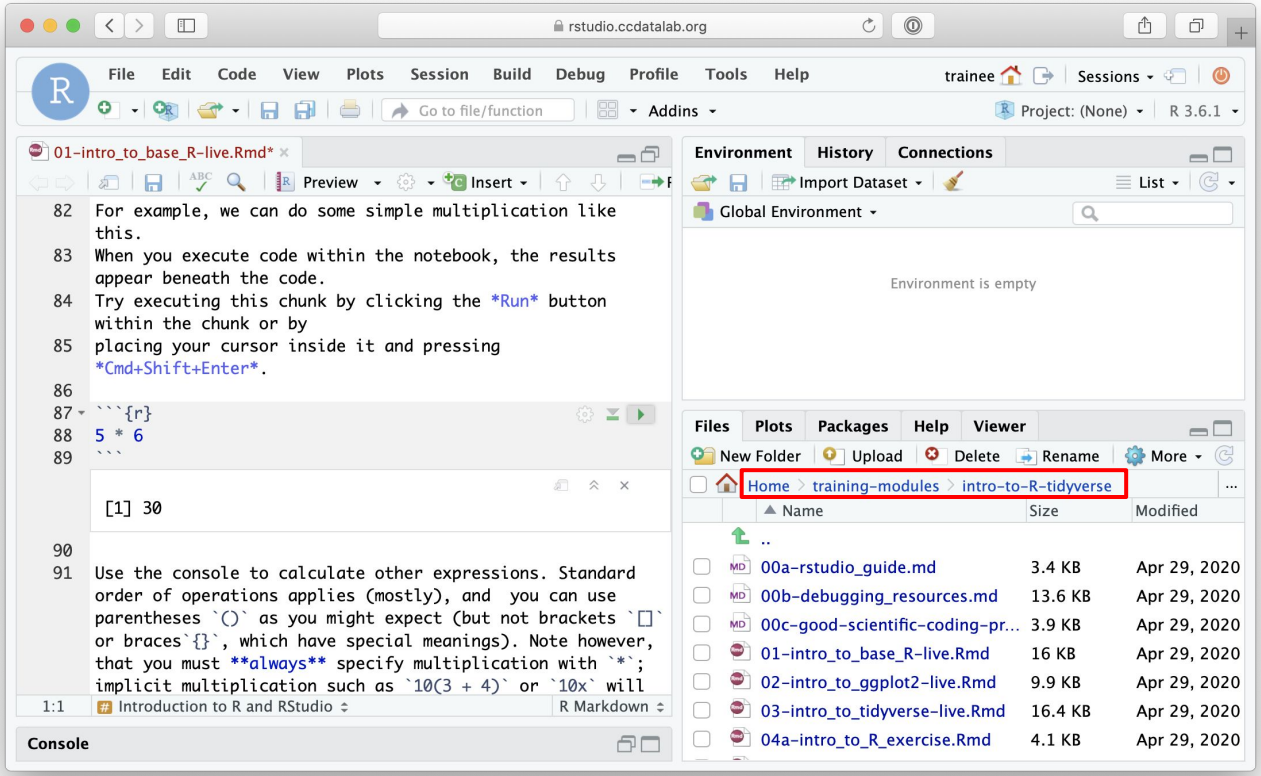
- R Notebooks allow you to have files that show both your code and results

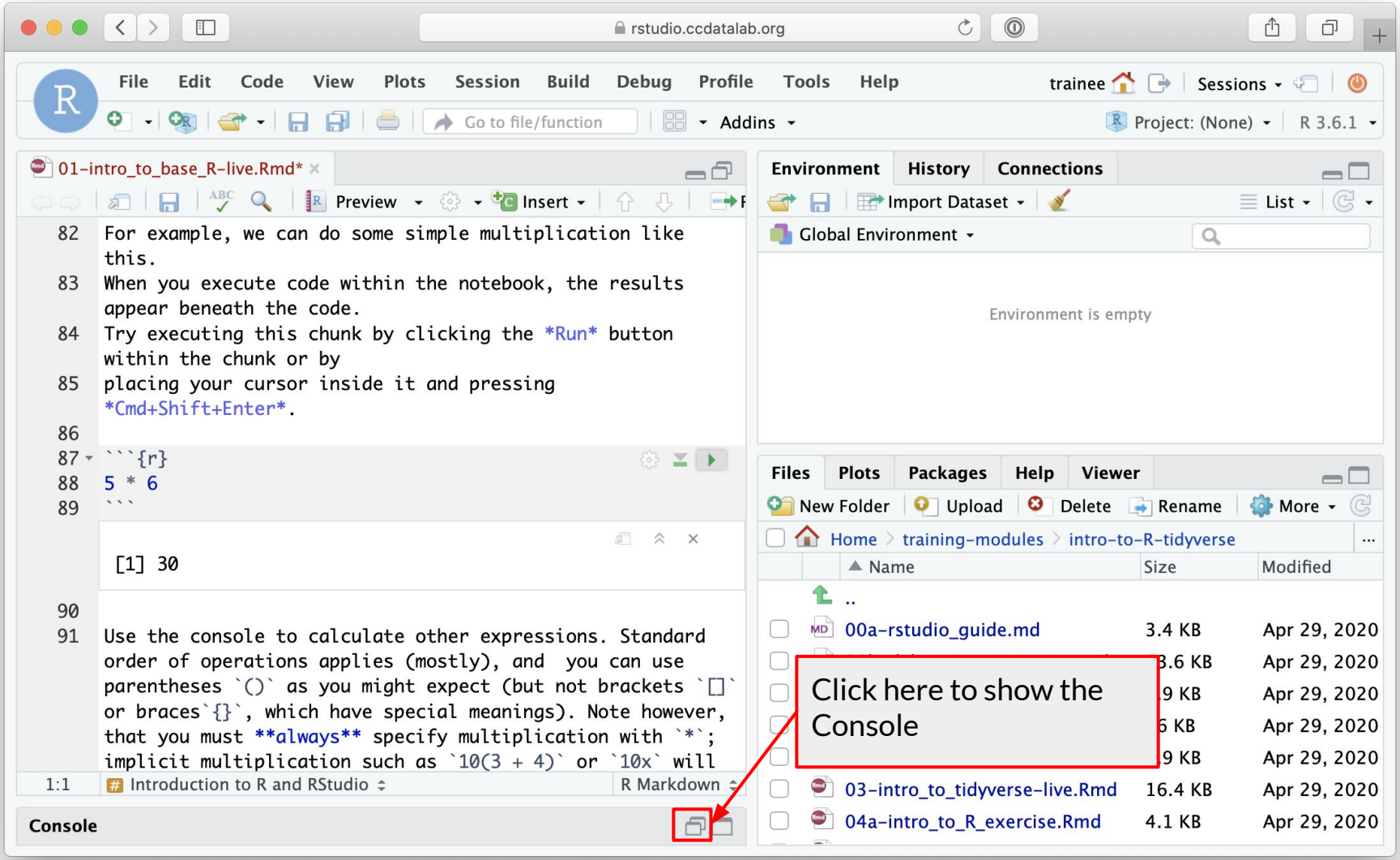


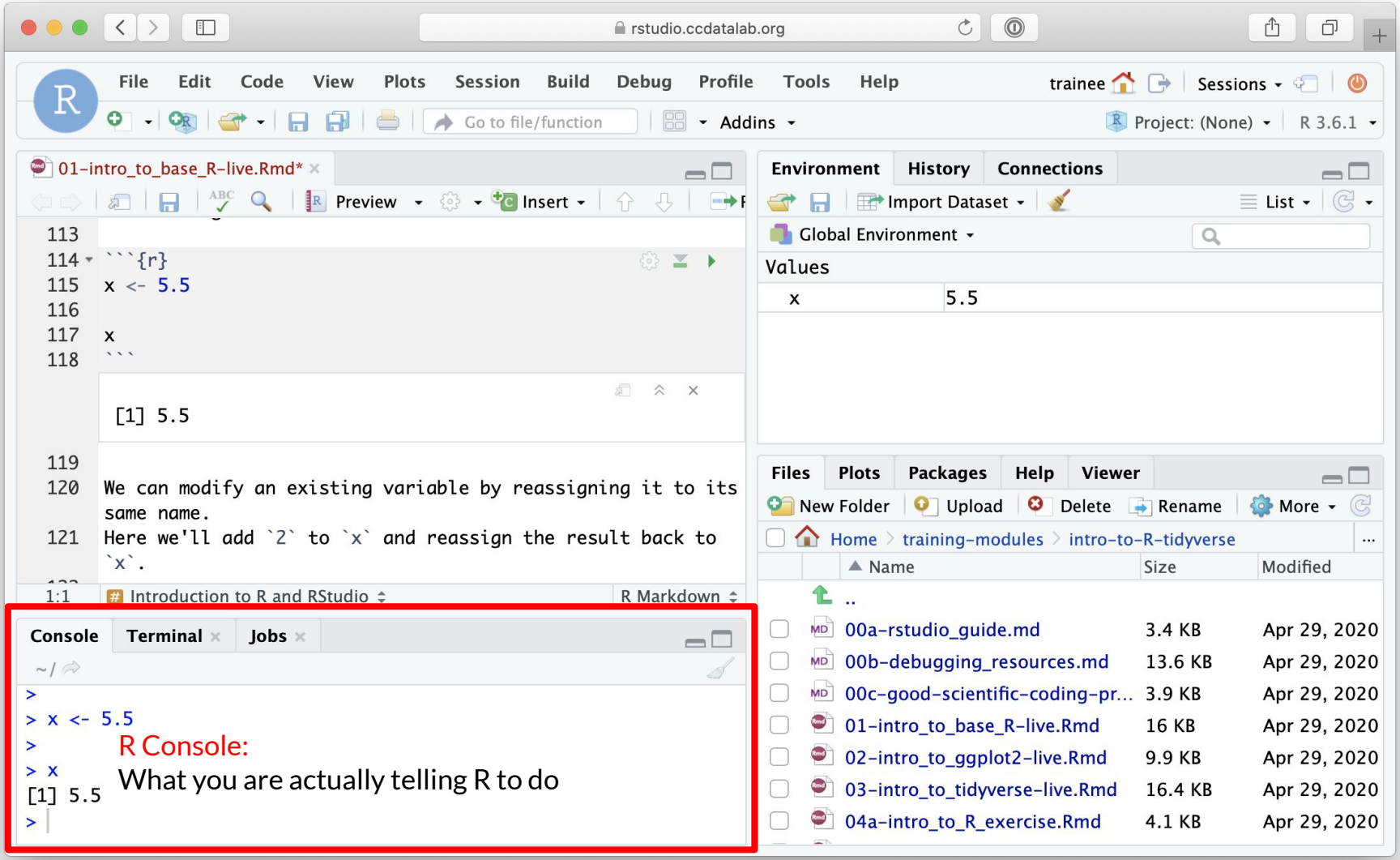
R Notebooks

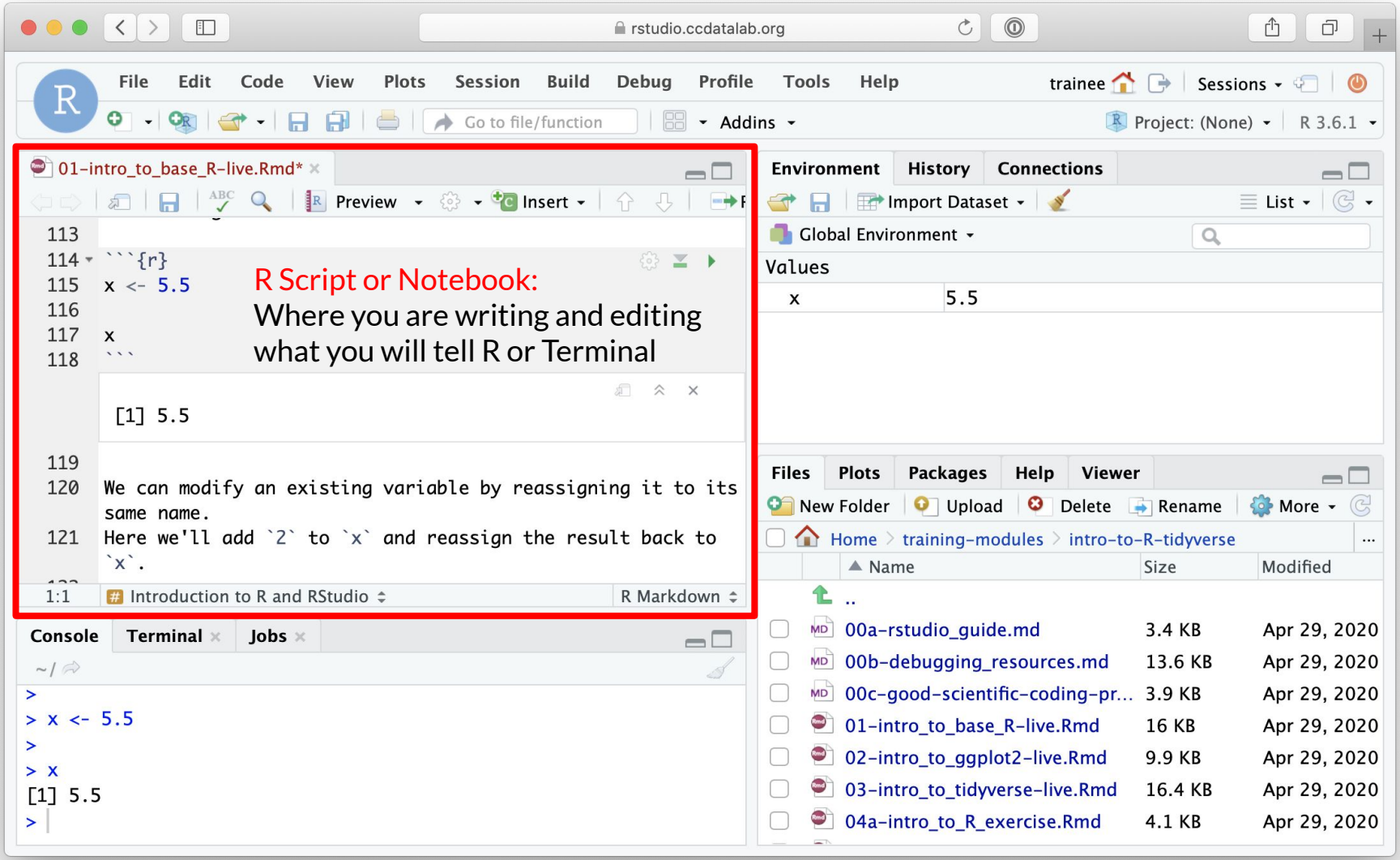
- Code that runs in R Notebooks uses <u>wherever the file is saved</u> as its *current directory*

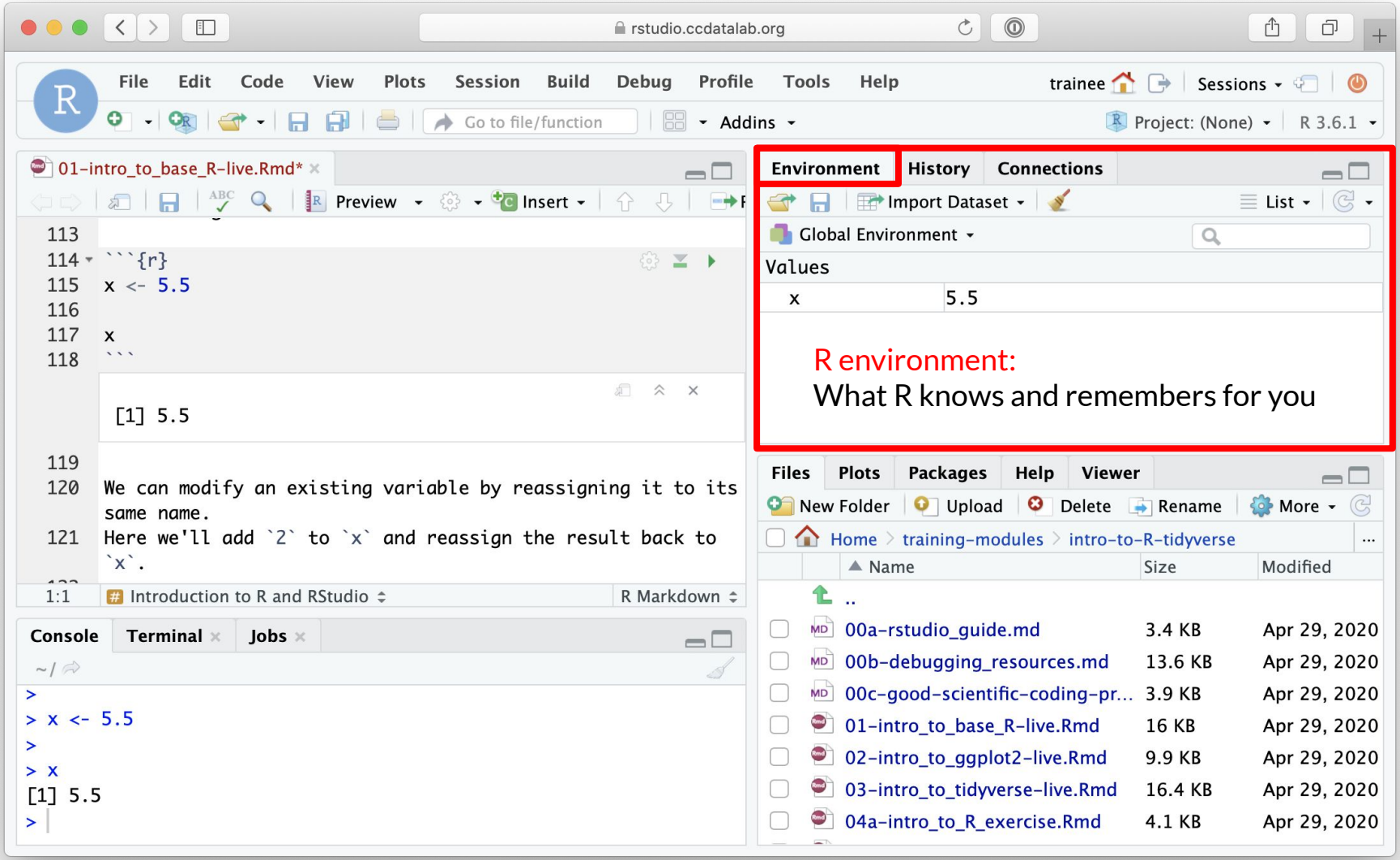
- Warning! That may not be the directory shown in the files pane *or* the console!

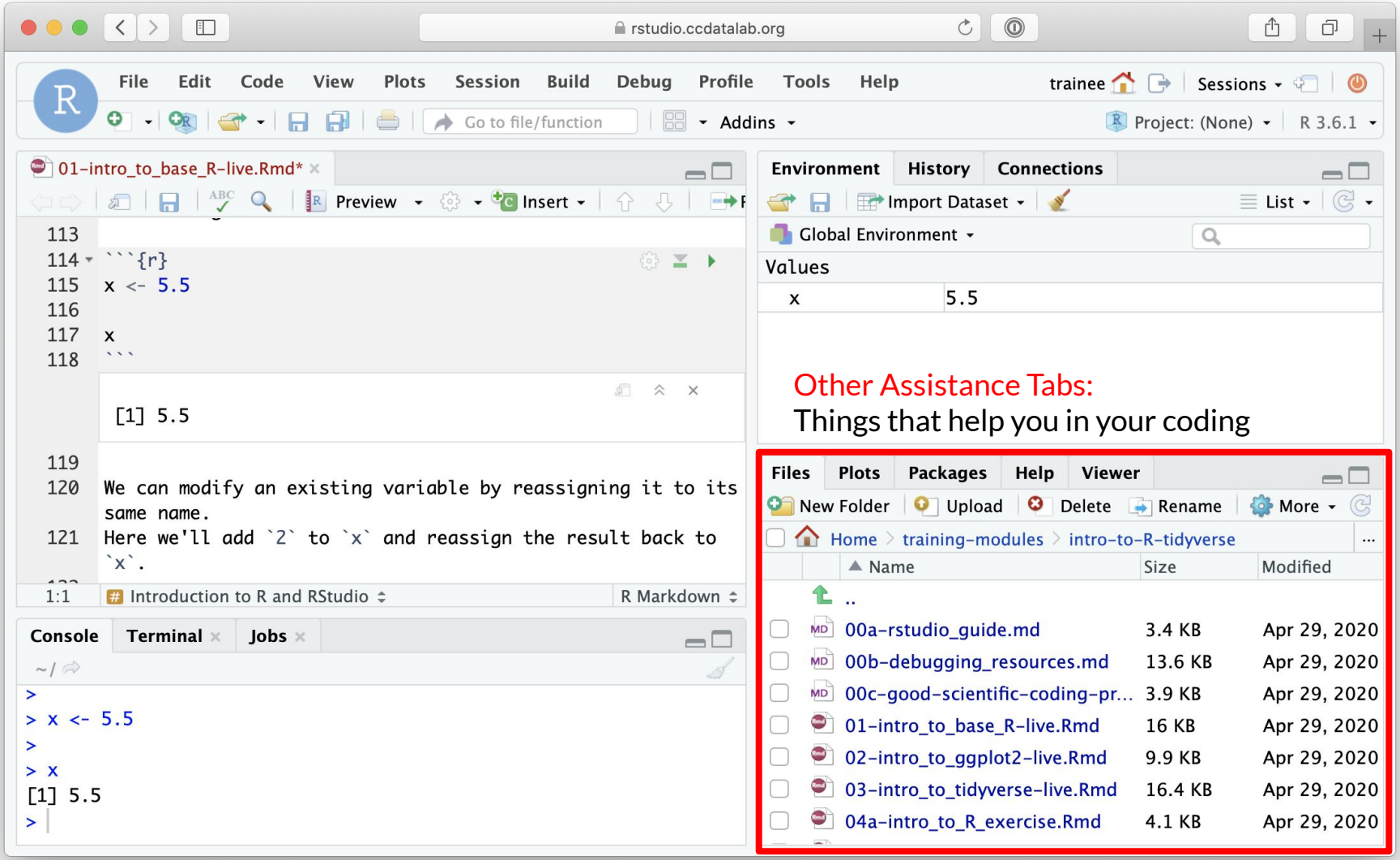



rstudio.ccdatalab.org
File Edit Code View Plots Session Build Debug Profile Tools Help
trainee
Sessions
Go to file/function
Addins
Project: (None)
R 3.6.1
01-intro_to_base_R-live.Rmd*
Preview
Insert
82 For example, we can do some simple multiplication like this.
83 When you execute code within the notebook, the results appear beneath the code.
84 Try executing this chunk by clicking the *Run* button within the chunk or by
85 placing your cursor inside it and pressing *Cmd+Shift+Enter*.
86
87 ```{r}
88 5 * 6
89 ```
[1] 30
90
91 Use the console to calculate other expressions. Standard order of operations applies (mostly), and you can use parentheses `()` as you might expect (but not brackets `[]` or braces`{}`, which have special meanings). Note however, that you must **always** specify multiplication with `*`; implicit multiplication such as `10(3 + 4)` or `10x` will
1:1
Introduction to R and RStudio
R Markdown
Console
Environment
History
Connections
Import Dataset
List
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
New Folder
Upload
Delete
Rename
More
Home
training-modules
intro-to-R-tidyverse
Name
Size
Modified
..
00a-rstudio_guide.md 3.4 KB Apr 29, 2020
Apr 29, 2020
Apr 29, 2020
Apr 29, 2020
Apr 29, 2020
03-intro_to_tidyverse-live.Rmd 16.4 KB Apr 29, 2020
04a-intro_to_R_exercise.Rmd 4.1 KB Apr 29, 2020
Click here to show the Console


rstudio.ccdatalab.org
File Edit Code View Plots Session Build Debug Profile Tools Help
trainee
Sessions
Go to file/function
Addins
Project: (None)
R 3.6.1
01-intro_to_base_R-live.Rmd*
Preview
Insert
113
114 ```{r}
115 x <- 5.5
116
117 x
118 ```
[1] 5.5
119
120 We can modify an existing variable by reassigning it to its same name.
121 Here we'll add `2` to `x` and reassign the result back to `x`.
1:1
Introduction to R and RStudio
R Markdown
Environment
History
Connections
Import Dataset
List
Global Environment
Values
x 5.5
Files
Plots
Packages
Help
Viewer
New Folder
Upload
Delete
Rename
More
Home
training-modules
intro-to-R-tidyverse
Name
Size
Modified
..
00a-rstudio_guide.md 3.4 KB Apr 29, 2020
00b-debugging_resources.md 13.6 KB Apr 29, 2020
00c-good-scientific-coding-pr... 3.9 KB Apr 29, 2020
01-intro_to_base_R-live.Rmd 16 KB Apr 29, 2020
02-intro_to_ggplot2-live.Rmd 9.9 KB Apr 29, 2020
03-intro_to_tidyverse-live.Rmd 16.4 KB Apr 29, 2020
04a-intro_to_R_exercise.Rmd 4.1 KB Apr 29, 2020
Console
Terminal
Jobs
~/
> x <- 5.5
> x
[1] 5.5
R Console:
What you are actually telling R to do


rstudio.ccdatalab.org
File Edit Code View Plots Session Build Debug Profile Tools Help
trainee
Sessions
Go to file/function
Addins
Project: (None)
R 3.6.1
01-intro_to_base_R-live.Rmd*
Preview
Insert
113
114 ```{r}
115 x <- 5.5
116
117 x
118 ```
[1] 5.5
119
120 We can modify an existing variable by reassigning it to its same name.
121 Here we'll add `2` to `x` and reassign the result back to `x`.
R Script or Notebook:
Where you are writing and editing what you will tell R or Terminal
1:1
Introduction to R and RStudio
R Markdown
Environment History Connections
Import Dataset
List
Global Environment
Values
x 5.5
Files Plots Packages Help Viewer
New Folder Upload Delete Rename More
Home training-modules intro-to-R-tidyverse
Name Size Modified
.. 00a-rstudio_guide.md 3.4 KB Apr 29, 2020
00b-debugging_resources.md 13.6 KB Apr 29, 2020
00c-good-scientific-coding-pr... 3.9 KB Apr 29, 2020
01-intro_to_base_R-live.Rmd 16 KB Apr 29, 2020
02-intro_to_ggplot2-live.Rmd 9.9 KB Apr 29, 2020
03-intro_to_tidyverse-live.Rmd 16.4 KB Apr 29, 2020
04a-intro_to_R_exercise.Rmd 4.1 KB Apr 29, 2020
Console Terminal Jobs
~/
> x <- 5.5
> x
[1] 5.5


rstudio.ccdatalab.org
File Edit Code View Plots Session Build Debug Profile Tools Help
trainee
Sessions
Go to file/function
Addins
Project: (None)
R 3.6.1
01-intro_to_base_R-live.Rmd*
Preview
Insert
113
114 ```{r}
115 x <- 5.5
116
117 x
118 ```
[1] 5.5
119
120 We can modify an existing variable by reassigning it to its same name.
121 Here we'll add `2` to `x` and reassign the result back to `x`.
1:1
Introduction to R and RStudio
R Markdown
Console
Terminal
Jobs
~/
> x <- 5.5
> x
[1] 5.5
Environment
History
Connections
Import Dataset
List
Global Environment
Values
x 5.5
R environment:
What R knows and remembers for you
Files
Plots
Packages
Help
Viewer
New Folder
Upload
Delete
Rename
More
Home
training-modules
intro-to-R-tidyverse
Name
Size
Modified
..
00a-rstudio_guide.md 3.4 KB Apr 29, 2020
00b-debugging_resources.md 13.6 KB Apr 29, 2020
00c-good-scientific-coding-pr... 3.9 KB Apr 29, 2020
01-intro_to_base_R-live.Rmd 16 KB Apr 29, 2020
02-intro_to_ggplot2-live.Rmd 9.9 KB Apr 29, 2020
03-intro_to_tidyverse-live.Rmd 16.4 KB Apr 29, 2020
04a-intro_to_R_exercise.Rmd 4.1 KB Apr 29, 2020


rstudio.ccdatalab.org
File Edit Code View Plots Session Build Debug Profile Tools Help
trainee
Sessions
Go to file/function
Addins
Project: (None)
R 3.6.1
01-intro_to_base_R-live.Rmd*
Preview
Insert
113
114 ```{r}
115 x <- 5.5
116
117 x
118 ```
[1] 5.5
119
120 We can modify an existing variable by reassigning it to its same name.
121 Here we'll add `2` to `x` and reassign the result back to `x`.
1:1
Introduction to R and RStudio
R Markdown
Console
Terminal
Jobs
~/
> x <- 5.5
> x
[1] 5.5
Environment
History
Connections
Import Dataset
List
Global Environment
Values
x 5.5
Other Assistance Tabs:
Things that help you in your coding
Files
Plots
Packages
Help
Viewer
New Folder
Upload
Delete
Rename
More
Home
training-modules
intro-to-R-tidyverse
Name
Size
Modified
..
00a-rstudio_guide.md 3.4 KB Apr 29, 2020
00b-debugging_resources.md 13.6 KB Apr 29, 2020
00c-good-scientific-coding-pr... 3.9 KB Apr 29, 2020
01-intro_to_base_R-live.Rmd 16 KB Apr 29, 2020
02-intro_to_ggplot2-live.Rmd 9.9 KB Apr 29, 2020
03-intro_to_tidyverse-live.Rmd 16.4 KB Apr 29, 2020
04a-intro_to_R_exercise.Rmd 4.1 KB Apr 29, 2020