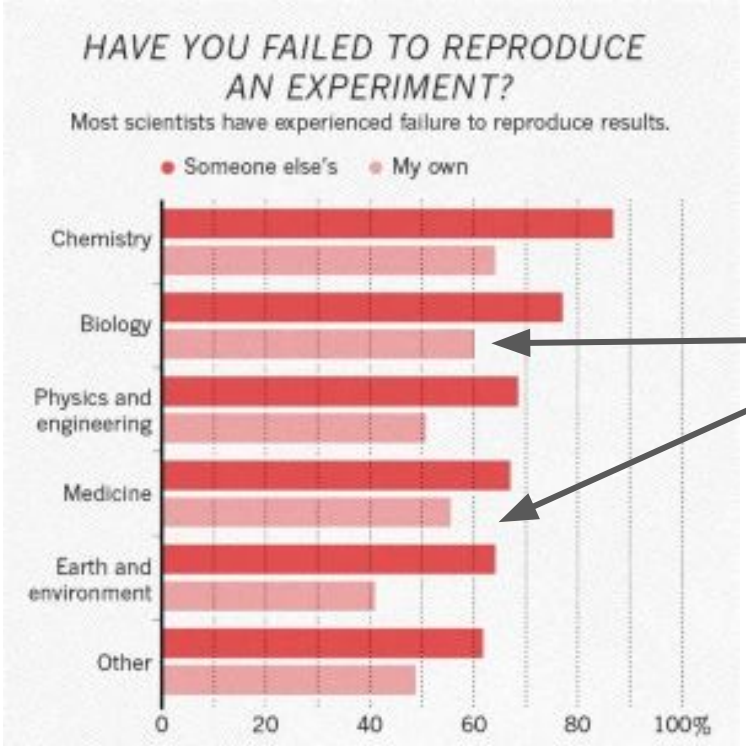
Introduction to R, RStudio, and RStudio Server

The Data Lab

Powered by Alex's Lemonade Stand Foundation

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

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

The problem continues…

NEWS | 13 August 2021 | Correction 25 August 2021

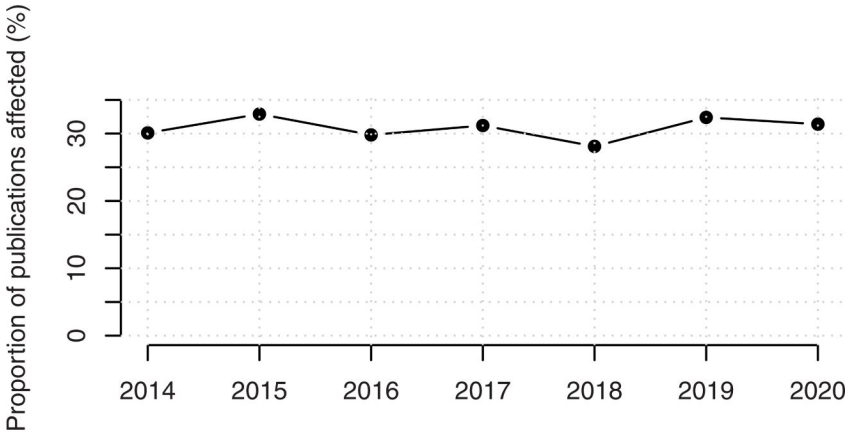
Autocorrect errors in Excel still creating genomics headache

Despite geneticists being warned about spreadsheet problems, 30% of published papers contain mangled gene names in supplementary data.

Dyani Lewis

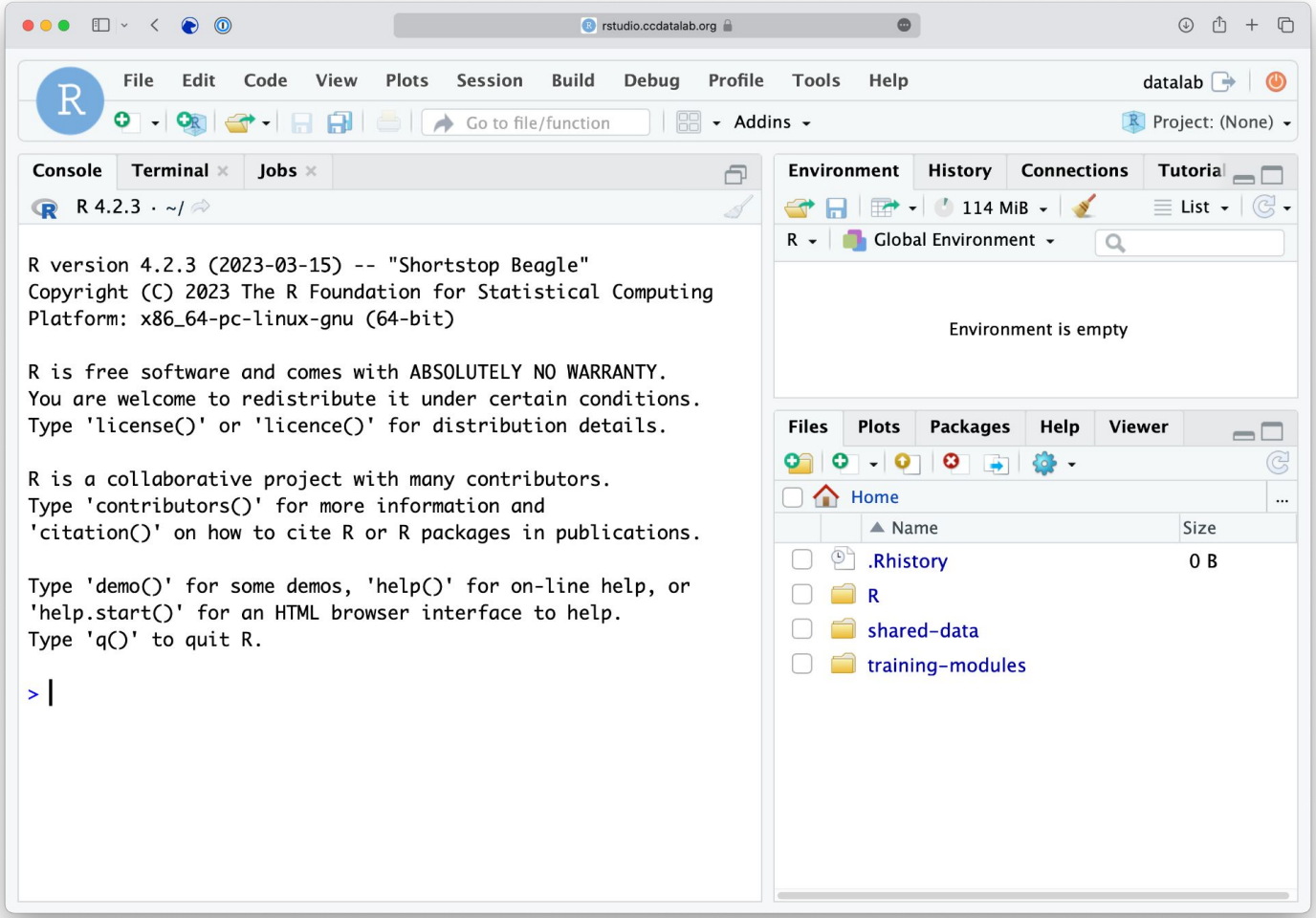
Embarrassing autocorrect mistakes are common fodder for Internet listicles and Twitter threads. But they are also the bane of geneticists using spreadsheet programs such as Microsoft Excel. Five years after a study showed that autocorrect problems were widespread, the academic literature is still littered with error-riddled spreadsheets, according to an

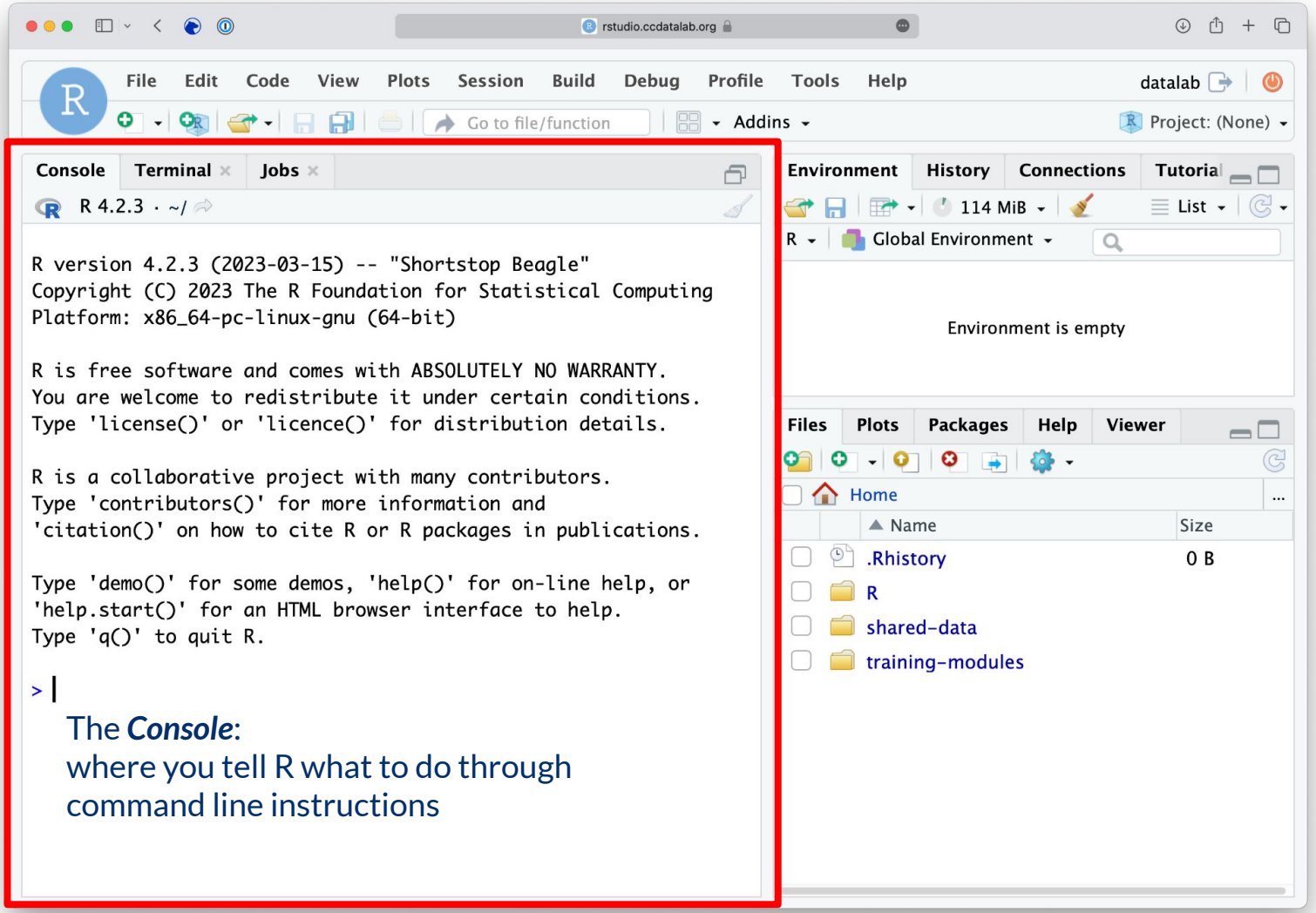
https://www.nature.com/articles/d41586-021-02211-4

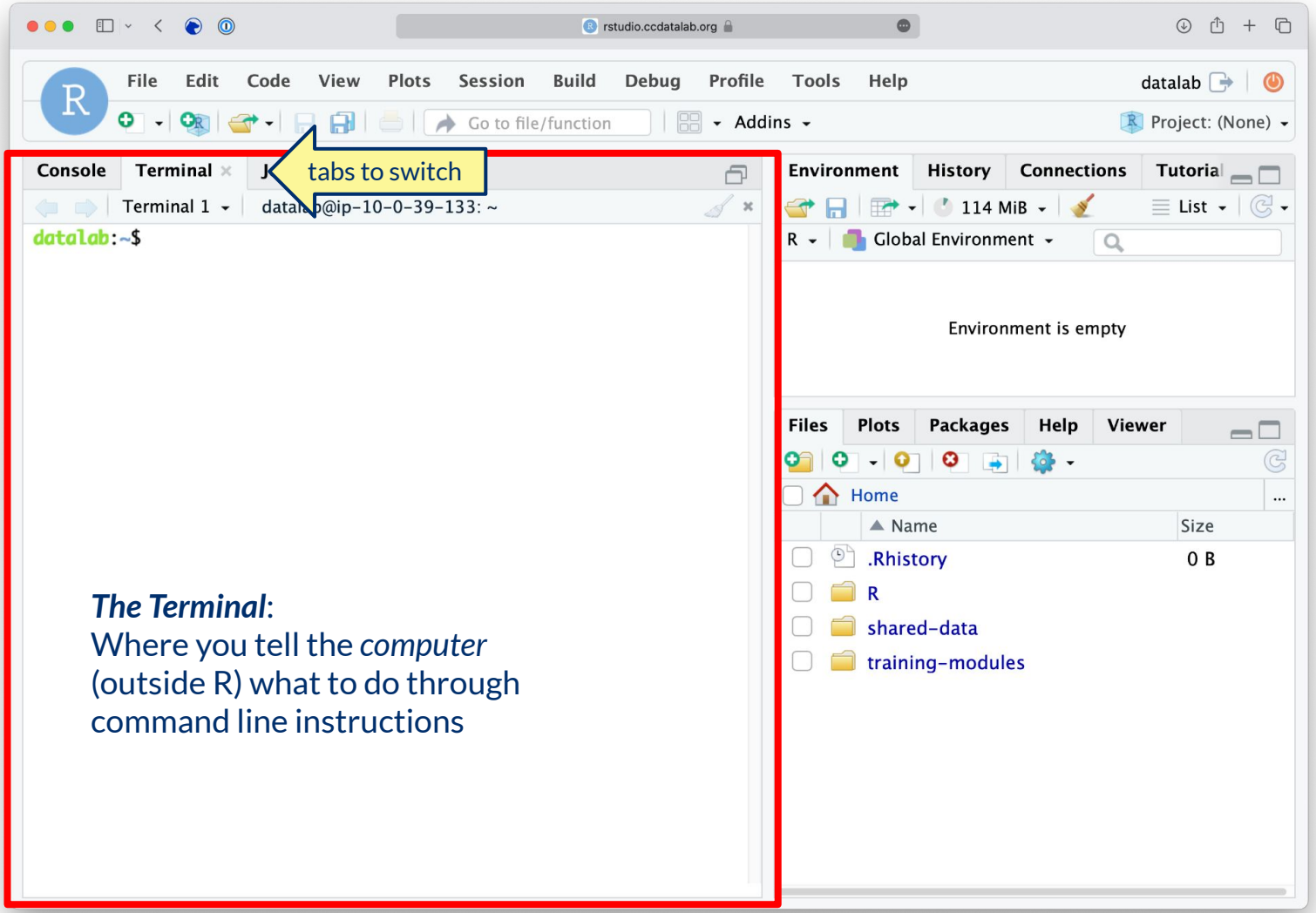


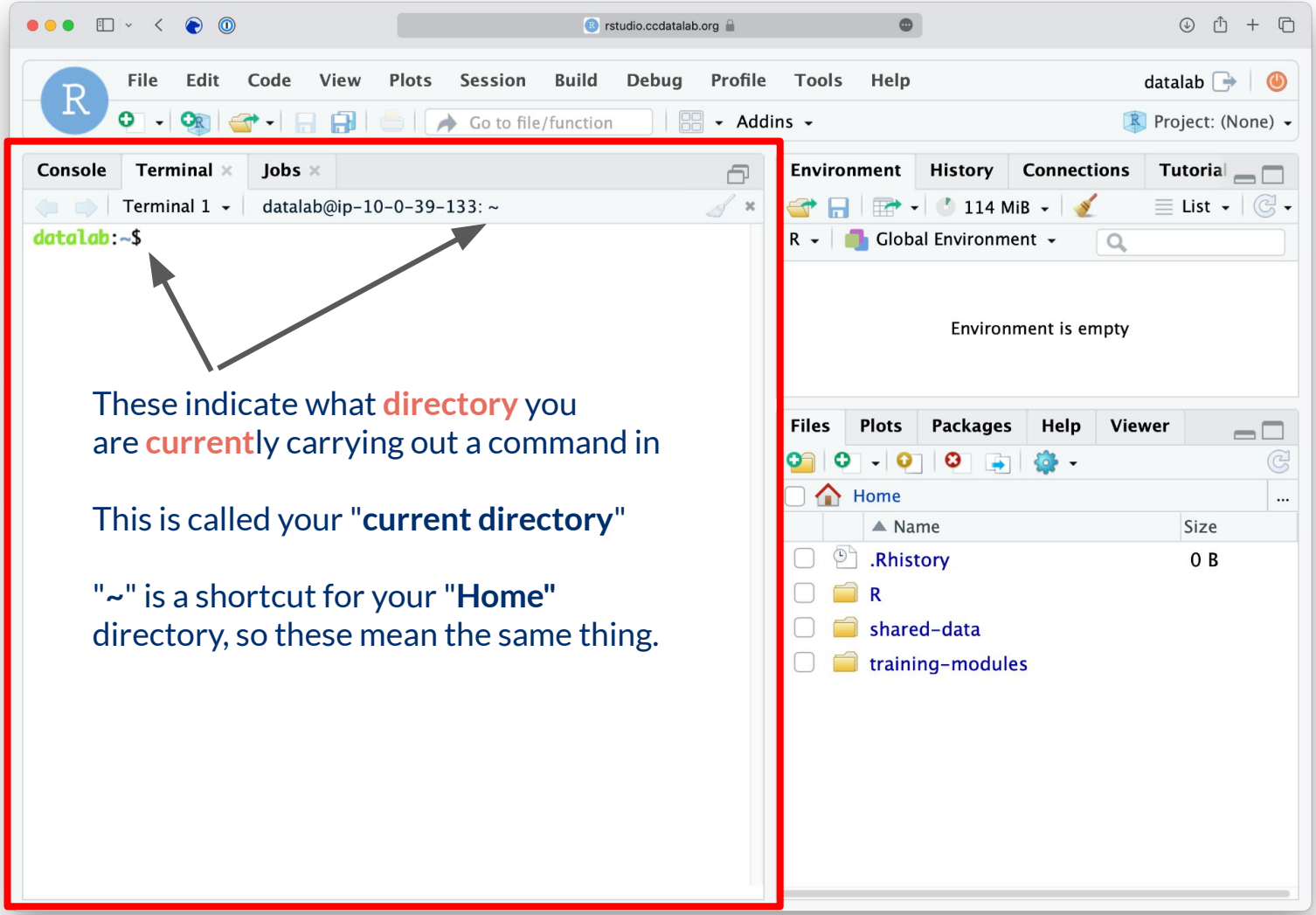
https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008984

RStudio Server: A basic guide

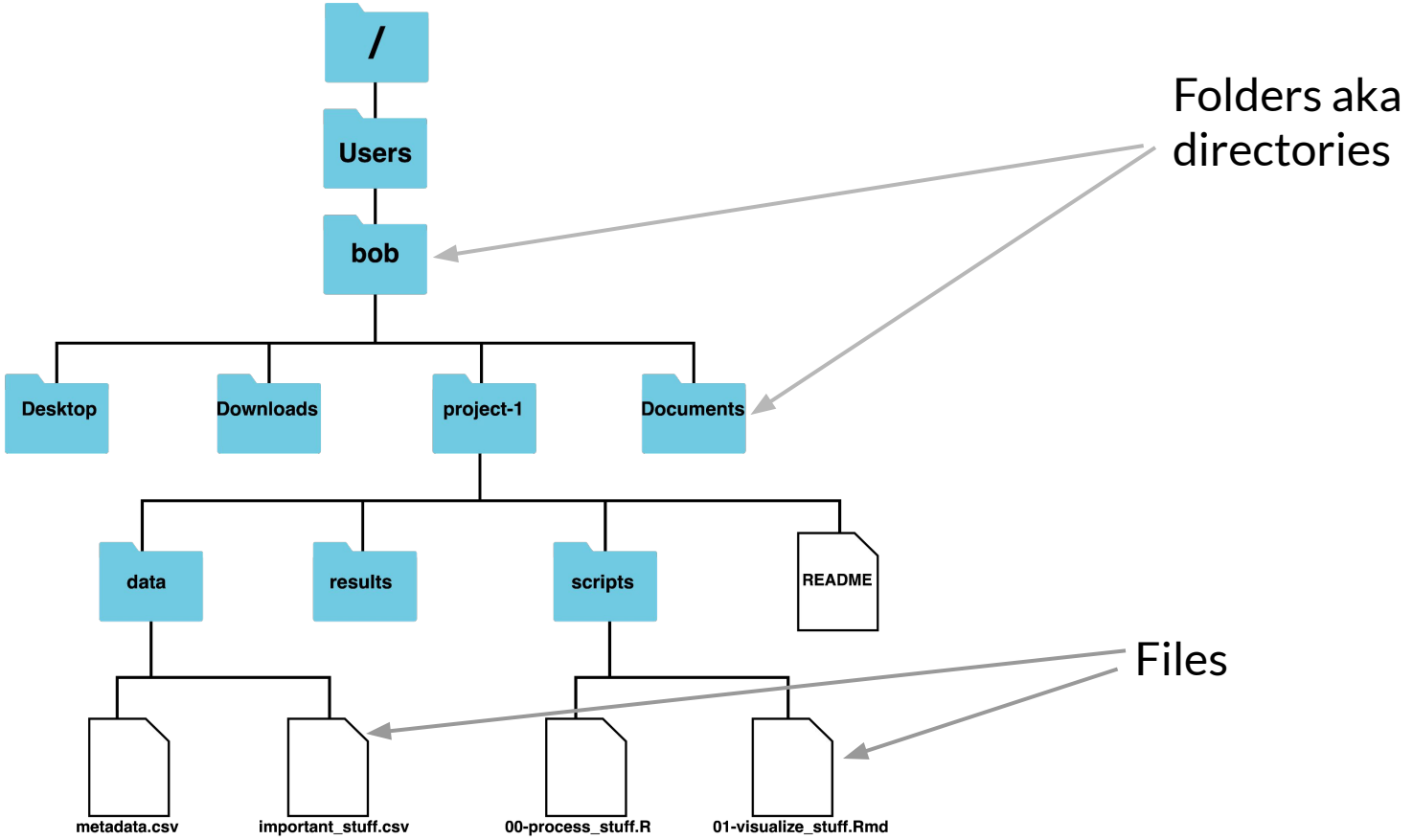



rstudio.ccdatalab.org
File Edit Code View Plots Session Build Debug Profile Tools Help
datalab
Go to file/function
Addins
Project: (None)
Console
Terminal
Jobs
R 4.2.3 · ~/
R version 4.2.3 (2023-03-15) -- "Shortstop Beagle"
Copyright (C) 2023 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (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
Tutorial
114 MiB
List
R
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
Home
Name
Size
.Rhistory
0 B
R
shared-data
training-modules


rstudio.ccdatalab.org
File Edit Code View Plots Session Build Debug Profile Tools Help
datalab
Go to file/function
Addins
Project: (None)
tabs to switch
Console
Terminal
Terminal 1
datalab@ip-10-0-39-133: ~
datalab:~$
Environment
History
Connections
Tutorial
114 MiB
List
R
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
Home
Name
Size
.Rhistory
0 B
R
shared-data
training-modules
The Terminal:
Where you tell the computer
(outside R) what to do through
command line instructions

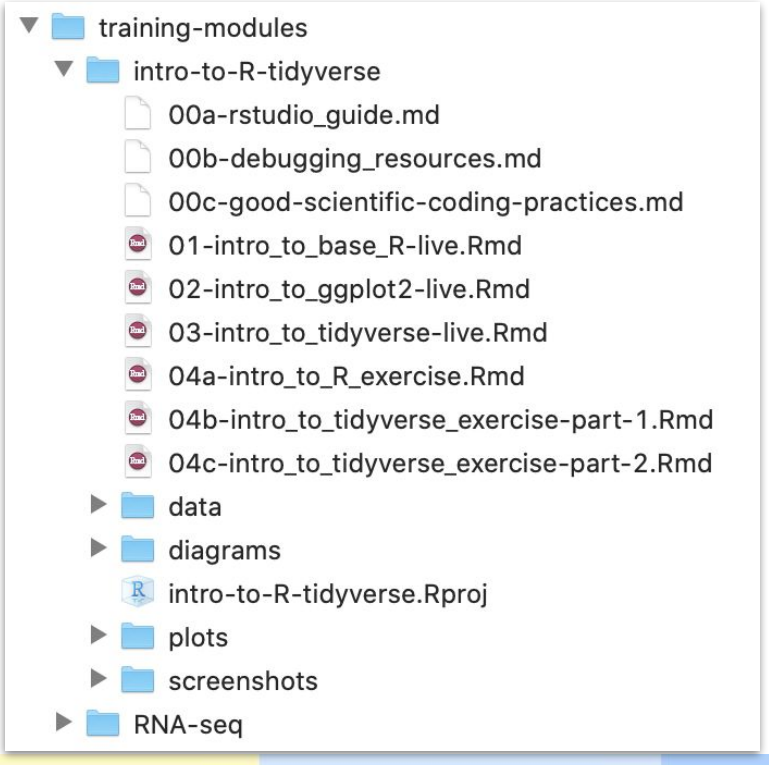

rstudio.ccdatalab.org
File Edit Code View Plots Session Build Debug Profile Tools Help
datalab
Go to file/function
Addins
Project: (None)
Console
Terminal
Jobs
Terminal 1
datalab@ip-10-0-39-133: ~
datalab:~$
These indicate what directory you are currently carrying out a command in
This is called your "current directory"
"~" is a shortcut for your "Home" directory, so these mean the same thing.
Environment
History
Connections
Tutorial
114 MiB
List
R
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
Home
Name
Size
.Rhistory
0 B
R
shared-data
training-modules

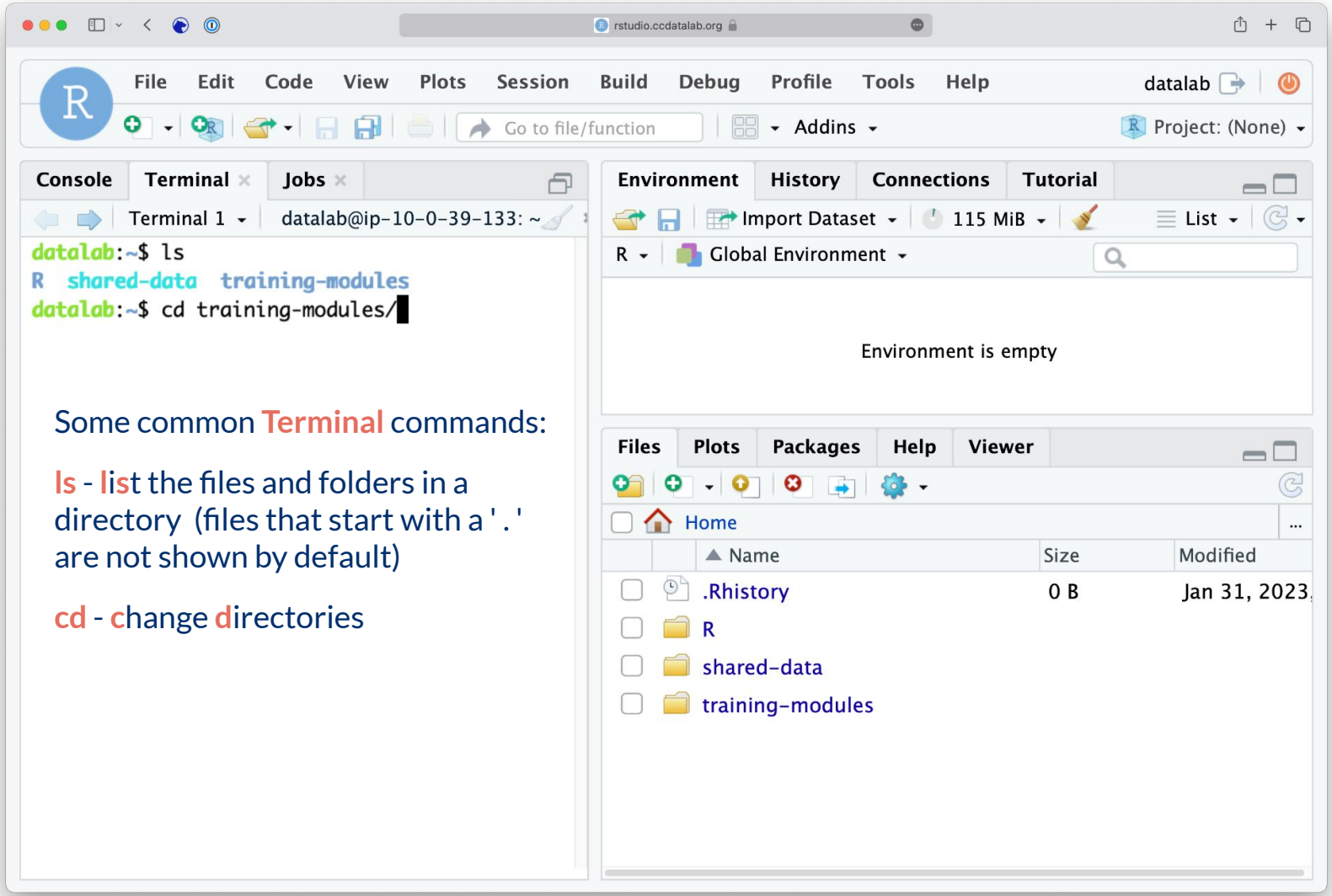
Example of a filesystem hierarchy

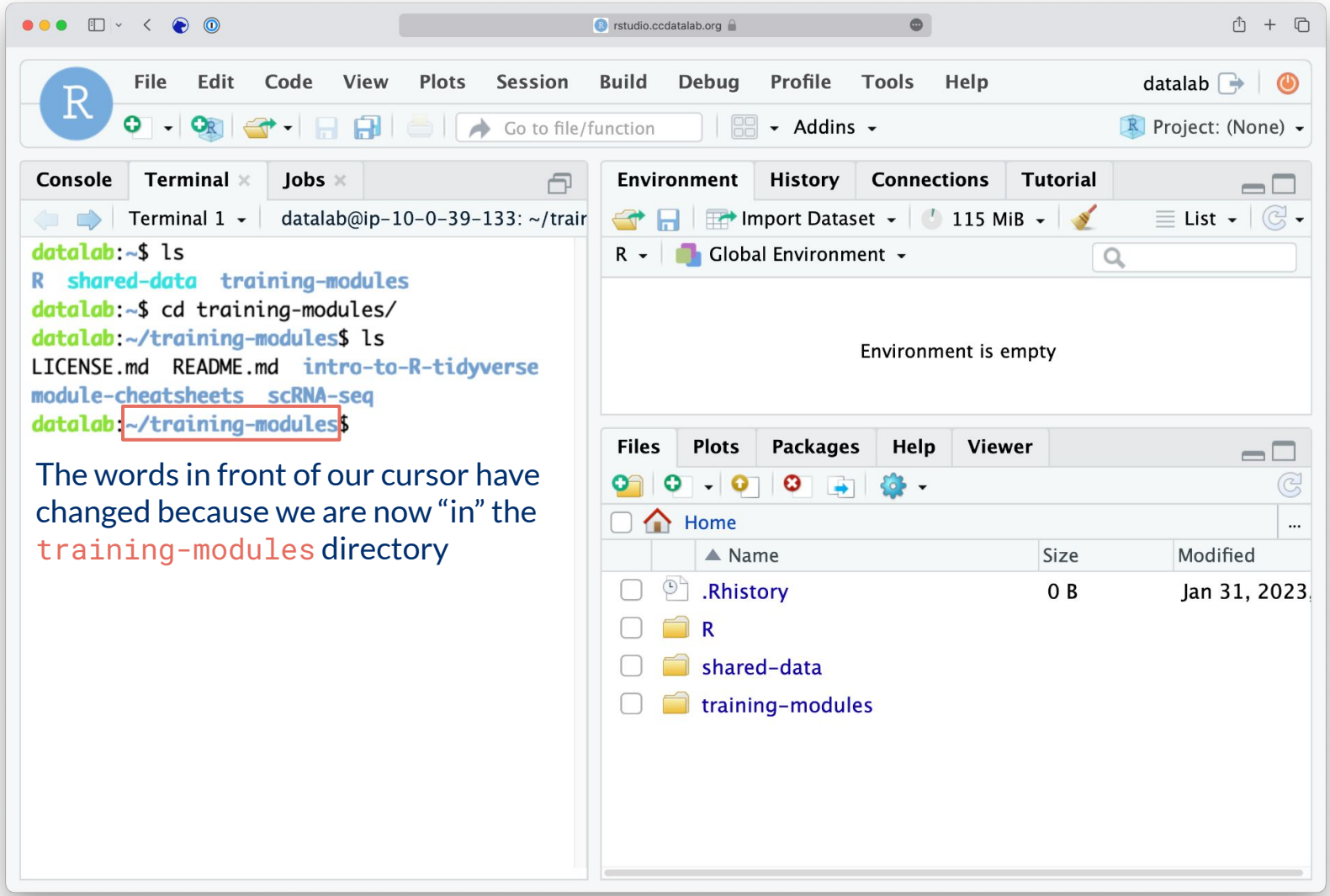


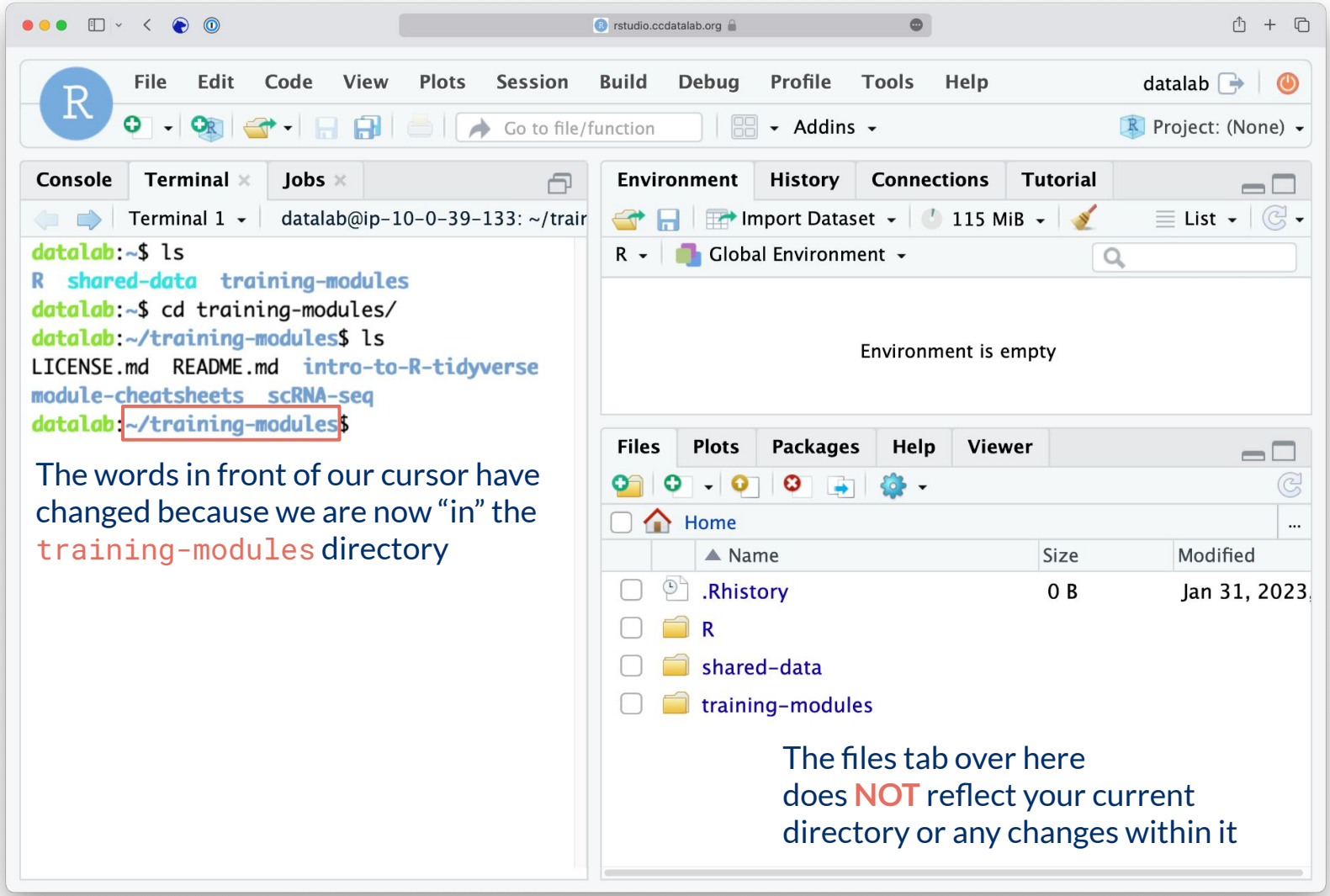
Our "current directory" for today's *Intro to R* module

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

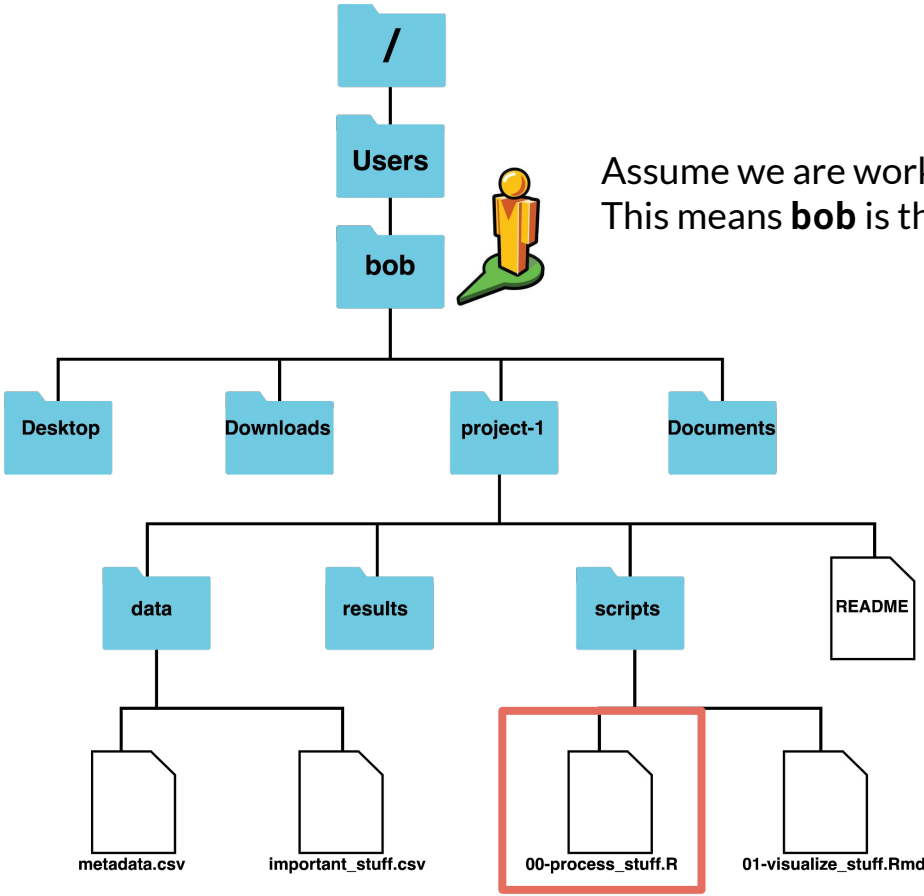



rstudio.ccdatalab.org
File Edit Code View Plots Session Build Debug Profile Tools Help
datalab
Go to file/function
Addins
Project: (None)
Console
Terminal
Jobs
Terminal 1
datalab@ip-10-0-39-133: ~
datalab:~$ ls
R shared-data training-modules
datalab:~$ cd training-modules/
Environment
History
Connections
Tutorial
Import Dataset
115 MiB
List
R
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
Home
Name
Size
Modified
.Rhistory
0 B
Jan 31, 2023
R
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
datalab
Go to file/function
Addins
Project: (None)
Console
Terminal
Jobs
Terminal 1
datalab@ip-10-0-39-133: ~/trair
datalab:~$ ls
R shared-data training-modules
datalab:~$ cd training-modules/
datalab:~/training-modules$ ls
LICENSE.md README.md intro-to-R-tidyverse
module-cheatsheets scRNA-seq
datalab:~/training-modules$
The words in front of our cursor have changed because we are now "in" the training-modules directory
Environment
History
Connections
Tutorial
Import Dataset
115 MiB
List
R
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
Home
Name
Size
Modified
.Rhistory
0 B
Jan 31, 2023
R
shared-data
training-modules


rstudio.ccdatalab.org
File Edit Code View Plots Session Build Debug Profile Tools Help
datalab
Go to file/function
Addins
Project: (None)
Console
Terminal
Jobs
Terminal 1
datalab@ip-10-0-39-133: ~/trair
datalab:~$ ls
R shared-data training-modules
datalab:~$ cd training-modules/
datalab:~/training-modules$ ls
LICENSE.md README.md intro-to-R-tidyverse
module-cheatsheets scRNA-seq
datalab:~/training-modules$
The words in front of our cursor have changed because we are now "in" the training-modules directory
Environment
History
Connections
Tutorial
Import Dataset
115 MiB
List
R
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
Home
Name
Size
Modified
.Rhistory
0 B
Jan 31, 2023
R
shared-data
training-modules
The files tab over here does NOT reflect your current directory or any changes within it

We are always working somewhere!



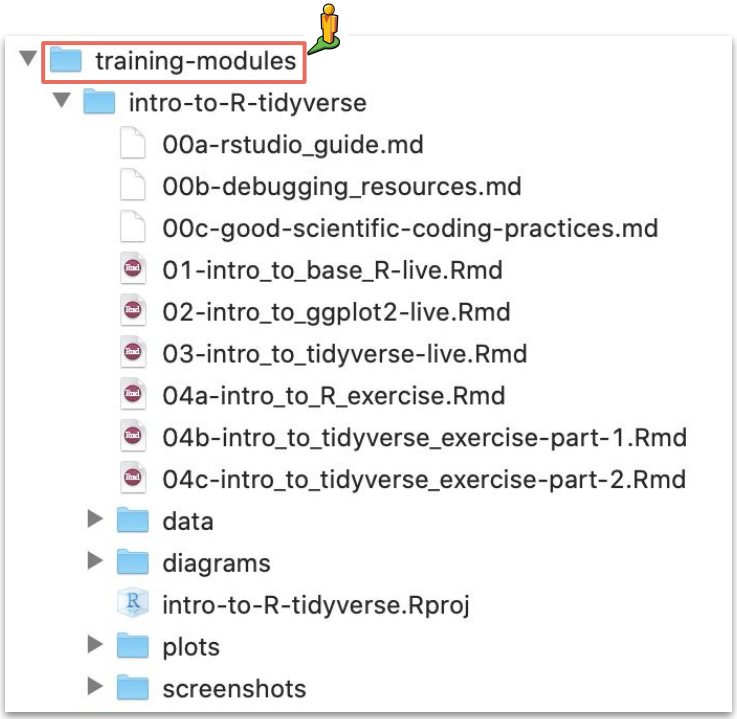
Assume we are working "from" the **bob** directory..
This means **bob** is the ***current (working) directory***

The file we are working on

Relative path: `project-1/scripts/00-process_stuff.R`

File paths: Directions to a file or folder

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

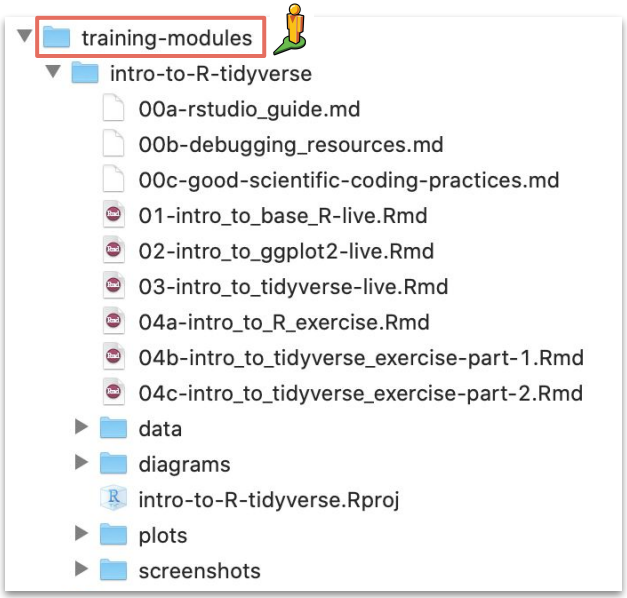


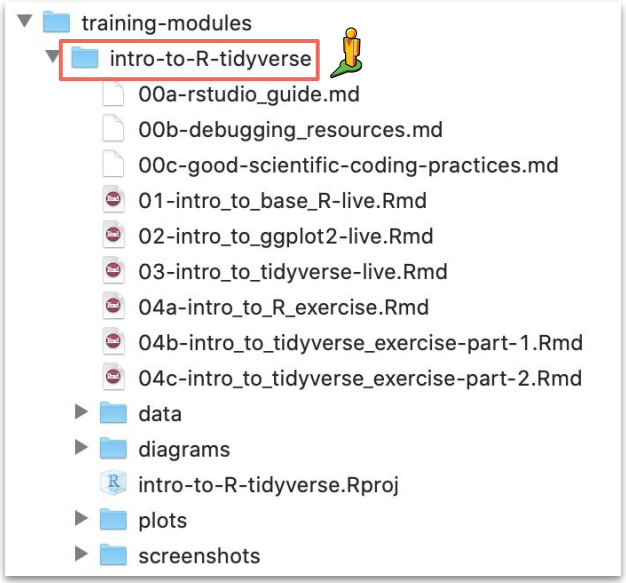
Current directory = `training-modules`

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

Relative file paths

Let's say we want to work with `01-intro_to_base_R-live.Rmd`





`training-modules`

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

`training-modules/intro-to-R-tidyverse`

Relative file path = **`01-intro_to_base_R-live.Rmd`**

Introduction to R

The Data Lab

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.

<u>Why we use R:</u>

- 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)

R, RStudio, and RStudio Server

R is a statistical programming language.

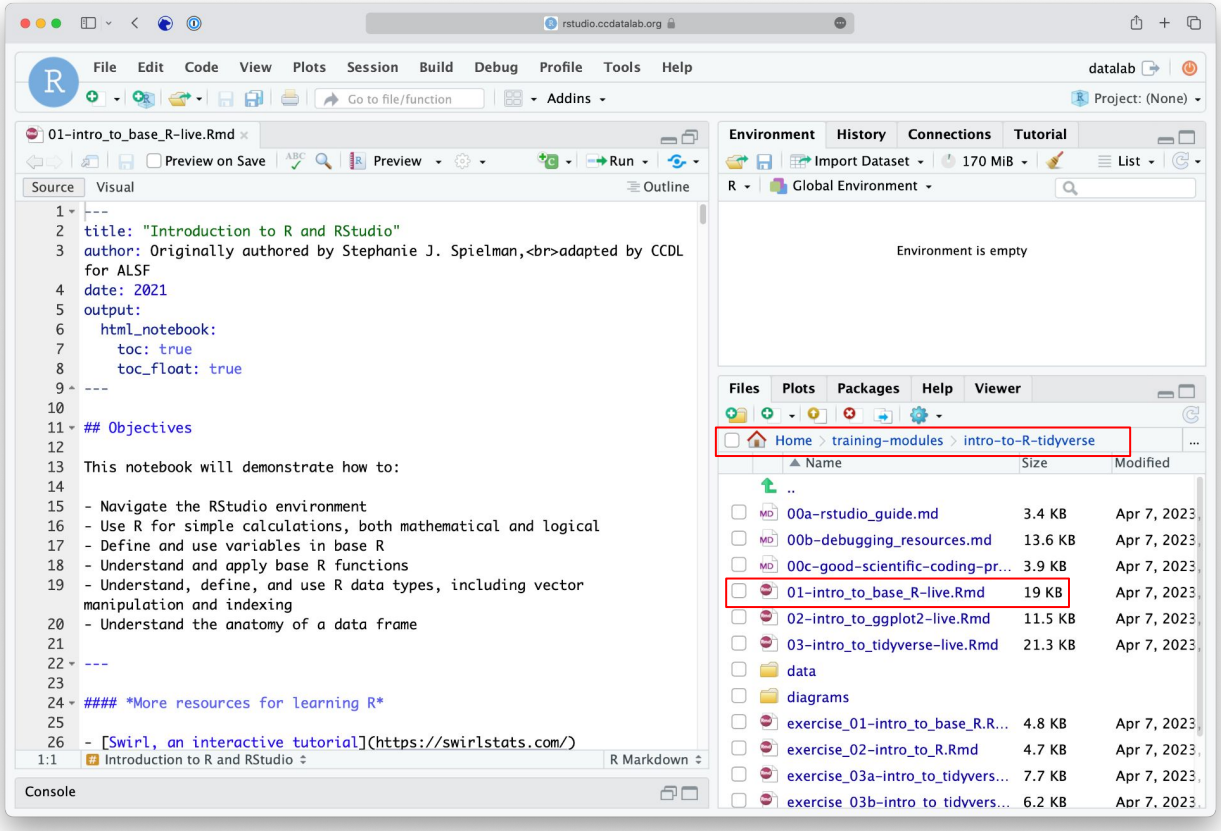
RStudio is an IDE for working in R

- 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

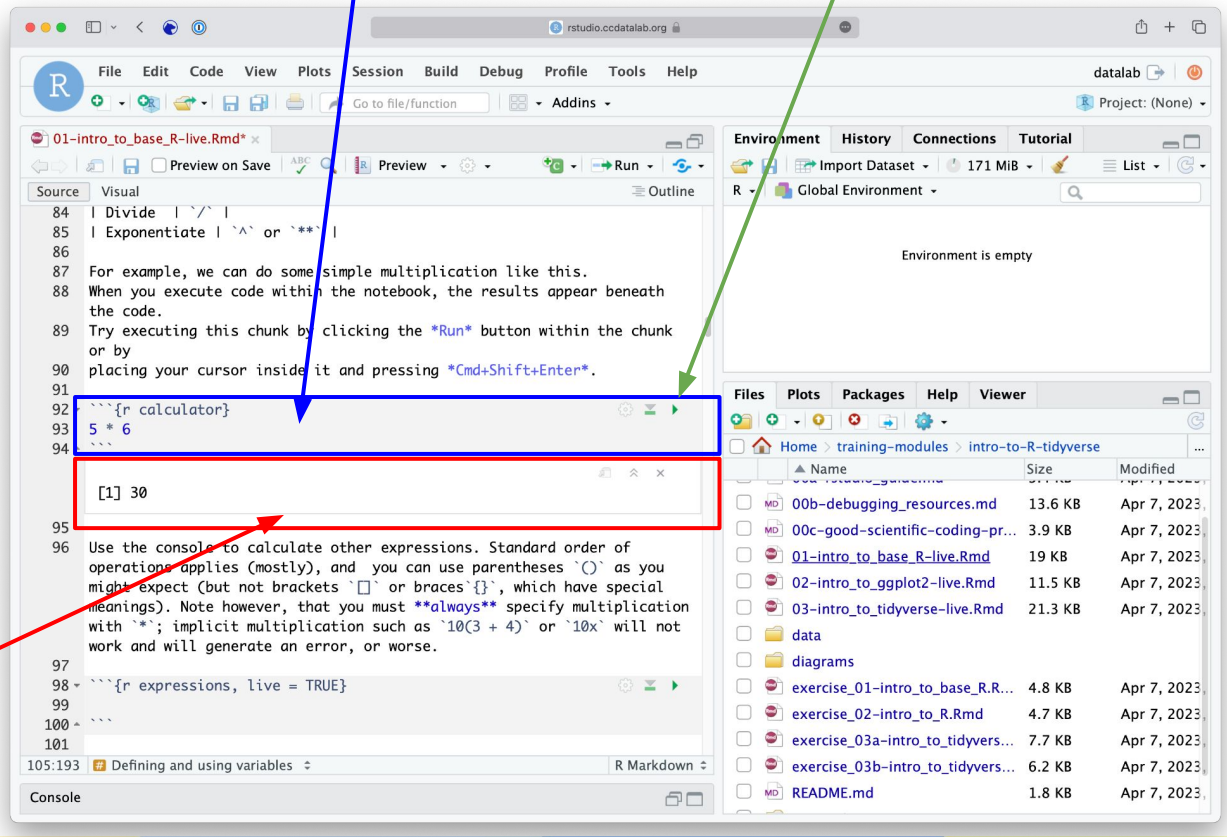
R Notebooks

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



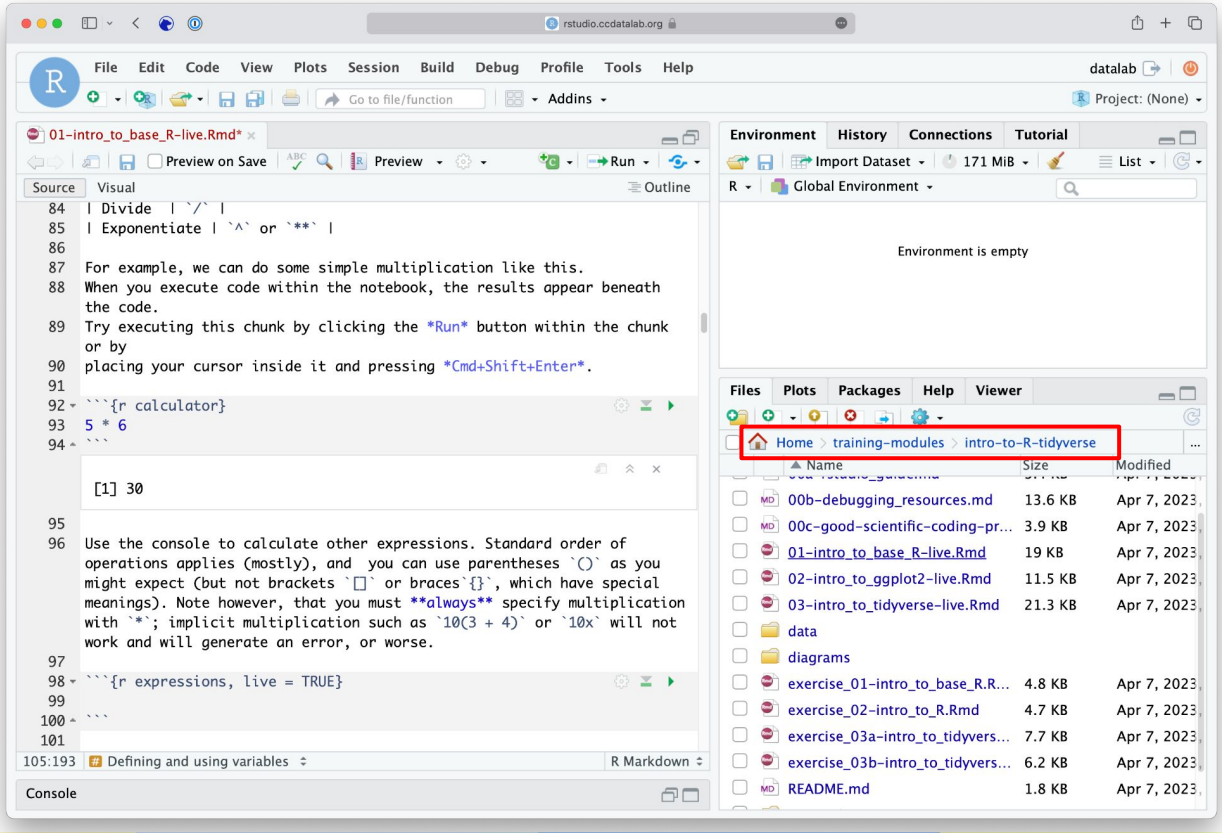
R Notebooks

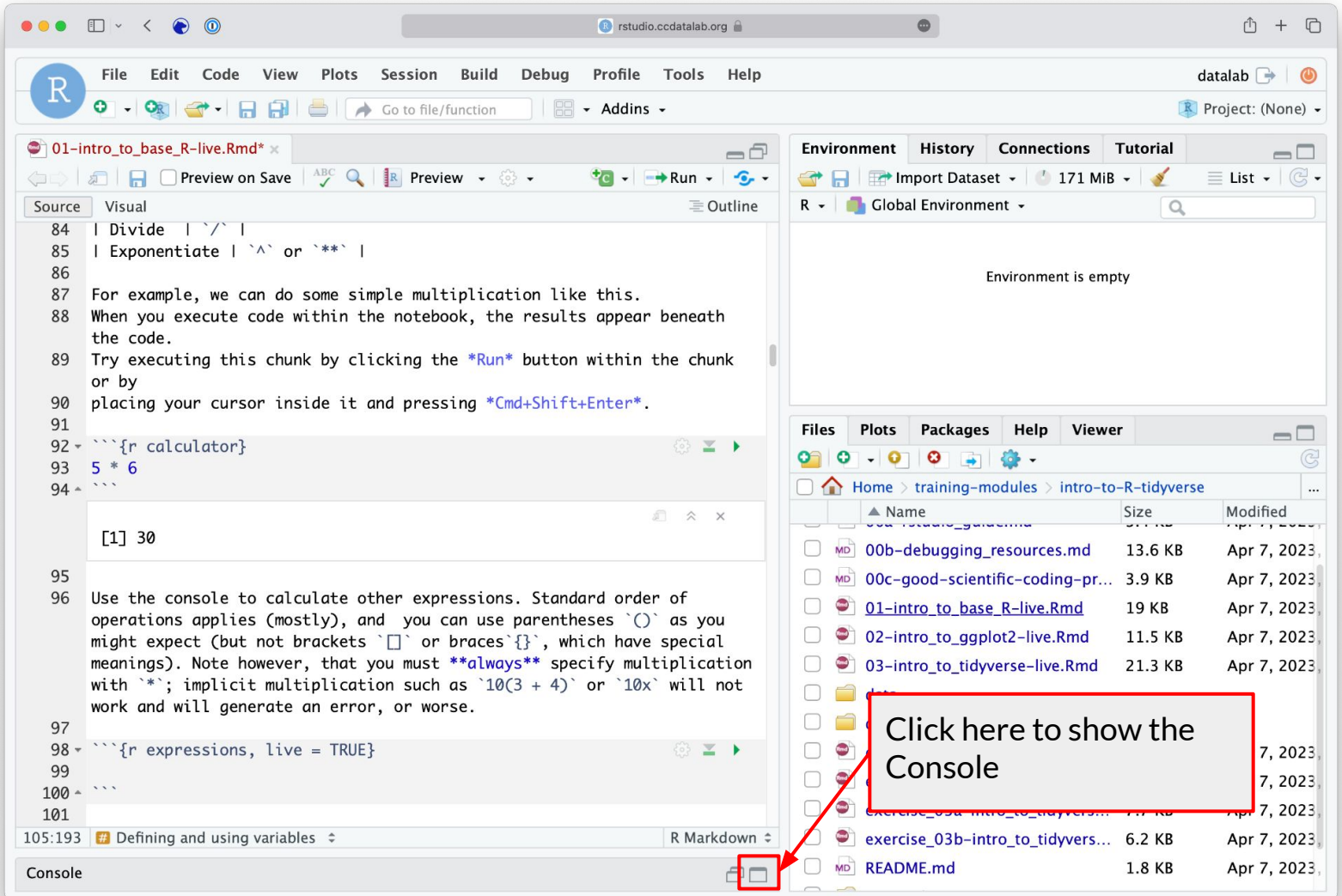
R Notebooks allow you to have files that show both your code and results

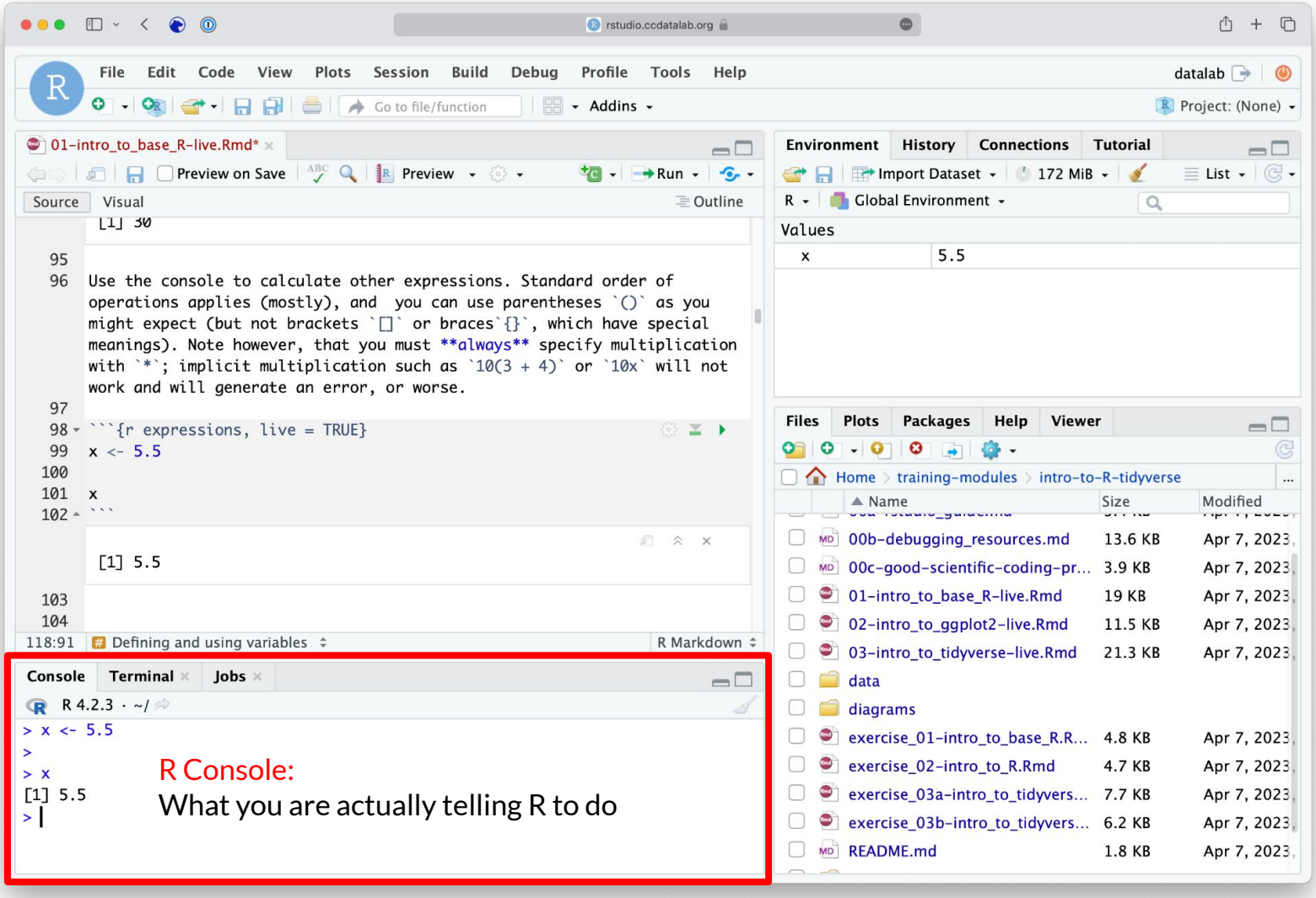


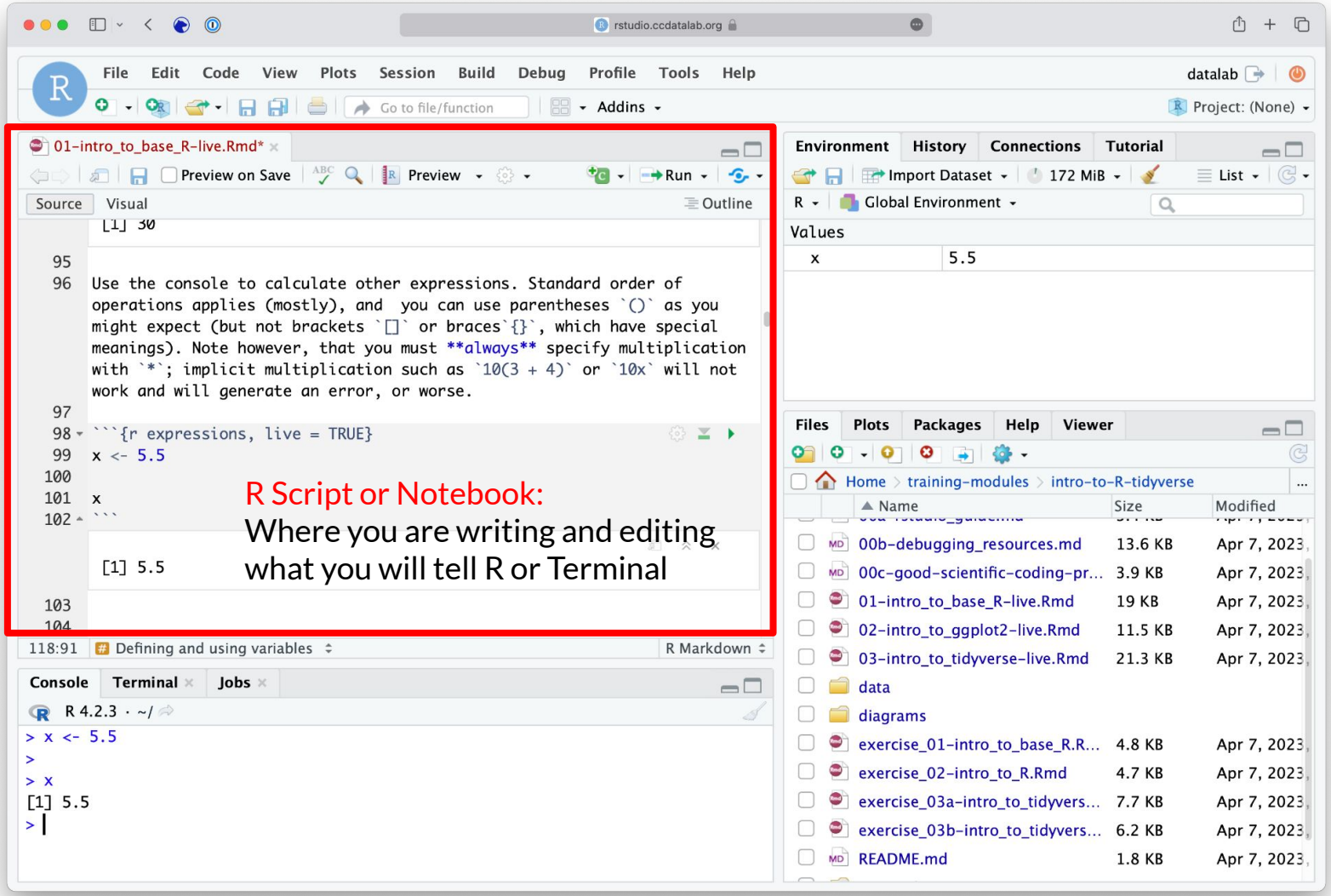
R Notebooks

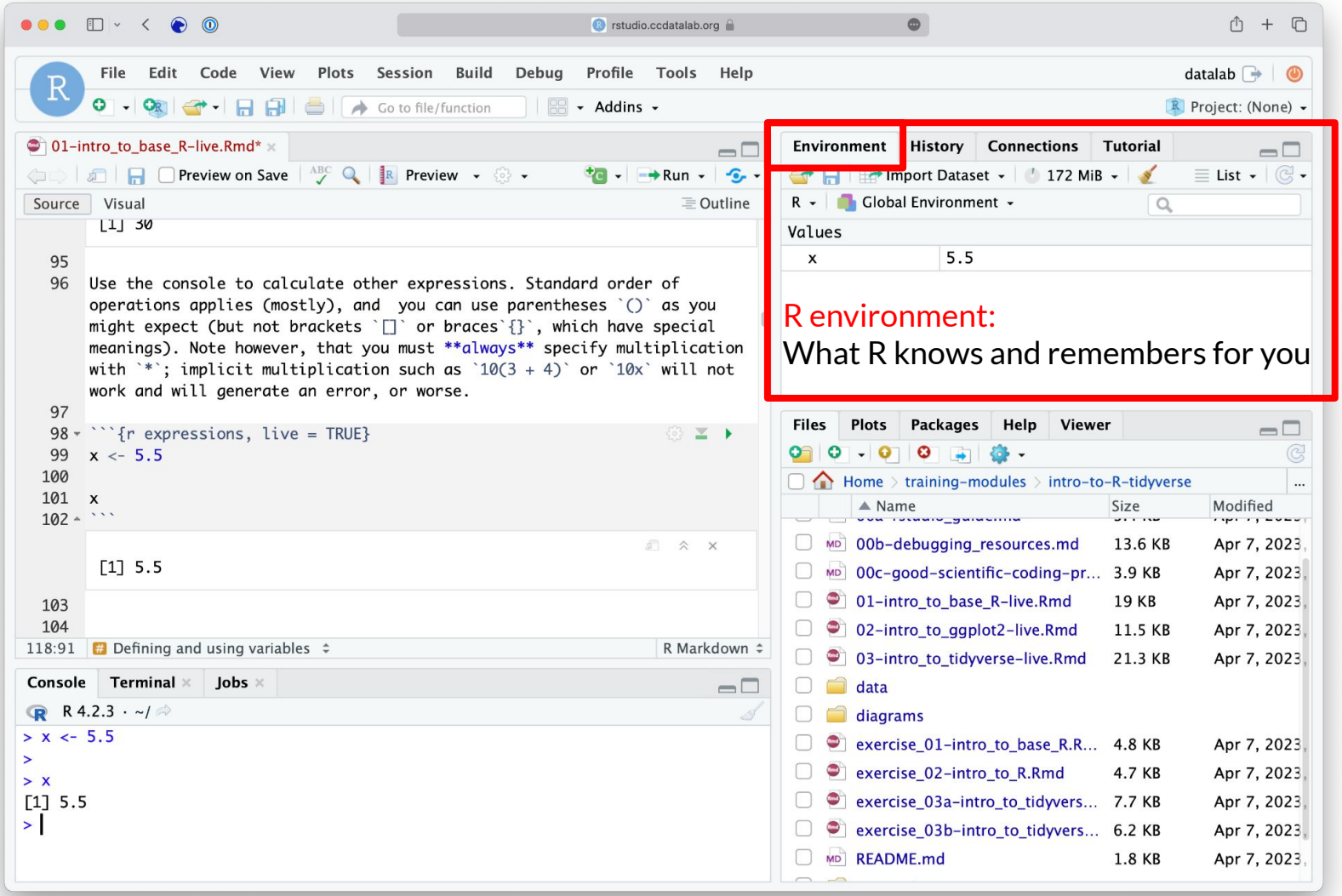
- Code that runs in R Notebooks uses <u>wherever the file is saved</u> as its *current directory*
- Warning! That may not be the directory shown in the files pane *or* the console!

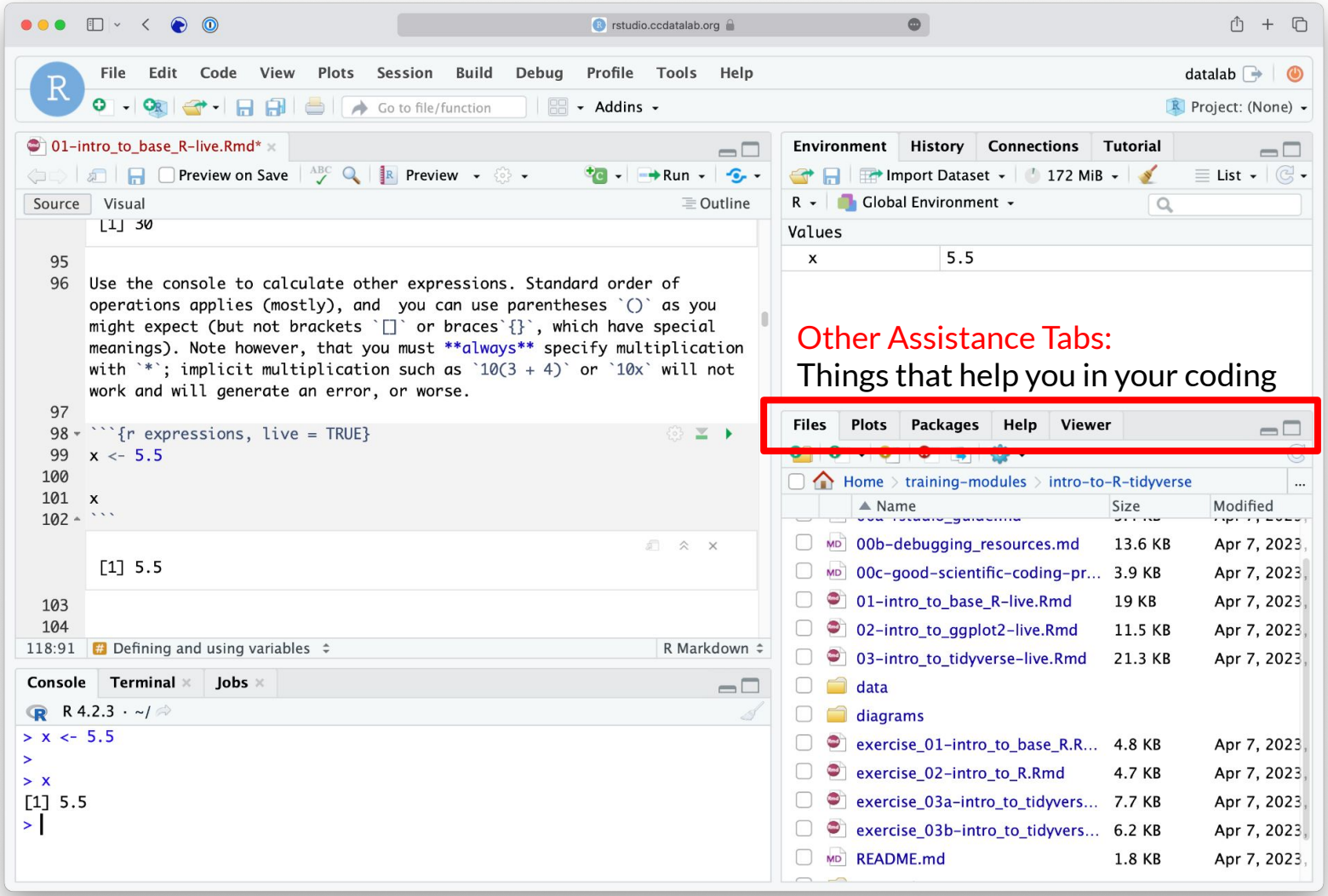



rstudio.ccdatalab.org
File Edit Code View Plots Session Build Debug Profile Tools Help
datalab
Go to file/function
Addins
Project: (None)
01-intro_to_base_R-live.Rmd*
Preview on Save
Preview
Run
Source
Visual
Outline
84 | Divide | `/` |
85 | Exponentiate | `^` or `**` |
86
87 For example, we can do some simple multiplication like this.
88 When you execute code within the notebook, the results appear beneath the code.
89 Try executing this chunk by clicking the *Run* button within the chunk or by
90 placing your cursor inside it and pressing *Cmd+Shift+Enter*.
91
92 ```{r calculator}
93 5 * 6
94 ```
[1] 30
95
96 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 not work and will generate an error, or worse.
97
98 ```{r expressions, live = TRUE}
99
100 ```
101
105:193
Defining and using variables
R Markdown
Console
Environment History Connections Tutorial
Import Dataset
171 MiB
List
R
Global Environment
Environment is empty
Files Plots Packages Help Viewer
Home > training-modules > intro-to-R-tidyverse
Name Size Modified
00b-debugging_resources.md 13.6 KB Apr 7, 2023
00c-good-scientific-coding-pr... 3.9 KB Apr 7, 2023
01-intro_to_base_R-live.Rmd 19 KB Apr 7, 2023
02-intro_to_ggplot2-live.Rmd 11.5 KB Apr 7, 2023
03-intro_to_tidyverse-live.Rmd 21.3 KB Apr 7, 2023
exercise_03b-intro_to_tidyvers... 6.2 KB Apr 7, 2023
README.md 1.8 KB Apr 7, 2023
Click here to show the Console


rstudio.ccdatalab.org
File Edit Code View Plots Session Build Debug Profile Tools Help
datalab
Go to file/function
Addins
Project: (None)
01-intro_to_base_R-live.Rmd*
Preview on Save
Preview
Run
Source
Visual
Outline
[1] 30
95
96 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 not work and will generate an error, or worse.
97
98 ```{r expressions, live = TRUE}
99 x <- 5.5
100
101 x
102 ```
[1] 5.5
103
104
118:91
Defining and using variables
R Markdown
Console
Terminal
Jobs
R 4.2.3 · ~/
> x <- 5.5
>
> x
[1] 5.5
>
R Console:
What you are actually telling R to do
Environment
History
Connections
Tutorial
Import Dataset
172 MiB
List
R
Global Environment
Values
x
5.5
Files
Plots
Packages
Help
Viewer
Home
training-modules
intro-to-R-tidyverse
Name
Size
Modified
00b-debugging_resources.md 13.6 KB Apr 7, 2023
00c-good-scientific-coding-pr... 3.9 KB Apr 7, 2023
01-intro_to_base_R-live.Rmd 19 KB Apr 7, 2023
02-intro_to_ggplot2-live.Rmd 11.5 KB Apr 7, 2023
03-intro_to_tidyverse-live.Rmd 21.3 KB Apr 7, 2023
data
diagrams
exercise_01-intro_to_base_R.R... 4.8 KB Apr 7, 2023
exercise_02-intro_to_R.Rmd 4.7 KB Apr 7, 2023
exercise_03a-intro_to_tidyvers... 7.7 KB Apr 7, 2023
exercise_03b-intro_to_tidyvers... 6.2 KB Apr 7, 2023
README.md 1.8 KB Apr 7, 2023


rstudio.ccdatalab.org
File Edit Code View Plots Session Build Debug Profile Tools Help
datalab
Go to file/function
Addins
Project: (None)
01-intro_to_base_R-live.Rmd*
Preview on Save
Preview
Run
Source
Visual
Outline
[1] 30
95
96 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 not work and will generate an error, or worse.
97
98 ```{r expressions, live = TRUE}
99 x <- 5.5
100
101 x
102 ```
[1] 5.5
103
104
R Script or Notebook:
Where you are writing and editing what you will tell R or Terminal
118:91
Defining and using variables
R Markdown
Console
Terminal
Jobs
R 4.2.3 · ~/
> x <- 5.5
>
> x
[1] 5.5
>
Environment
History
Connections
Tutorial
Import Dataset
172 MiB
List
R
Global Environment
Values
x
5.5
Files
Plots
Packages
Help
Viewer
Home
training-modules
intro-to-R-tidyverse
Name
Size
Modified
00b-debugging_resources.md 13.6 KB Apr 7, 2023
00c-good-scientific-coding-pr... 3.9 KB Apr 7, 2023
01-intro_to_base_R-live.Rmd 19 KB Apr 7, 2023
02-intro_to_ggplot2-live.Rmd 11.5 KB Apr 7, 2023
03-intro_to_tidyverse-live.Rmd 21.3 KB Apr 7, 2023
data
diagrams
exercise_01-intro_to_base_R.R... 4.8 KB Apr 7, 2023
exercise_02-intro_to_R.Rmd 4.7 KB Apr 7, 2023
exercise_03a-intro_to_tidyvers... 7.7 KB Apr 7, 2023
exercise_03b-intro_to_tidyvers... 6.2 KB Apr 7, 2023
README.md 1.8 KB Apr 7, 2023


rstudio.ccdatalab.org
File Edit Code View Plots Session Build Debug Profile Tools Help
datalab
Go to file/function
Addins
Project: (None)
01-intro_to_base_R-live.Rmd*
Preview on Save
Preview
Run
Source
Visual
Outline
[1] 30
95
96 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 not work and will generate an error, or worse.
97
98 ```{r expressions, live = TRUE}
99 x <- 5.5
100
101 x
102 ```
[1] 5.5
103
104
118:91
Defining and using variables
R Markdown
Console
Terminal
Jobs
R 4.2.3 · ~/
> x <- 5.5
>
> x
[1] 5.5
>
Environment
History
Connections
Tutorial
Import Dataset
172 MiB
List
R
Global Environment
Values
x
5.5
R environment:
What R knows and remembers for you
Files
Plots
Packages
Help
Viewer
Home
training-modules
intro-to-R-tidyverse
Name
Size
Modified
00b-debugging_resources.md 13.6 KB Apr 7, 2023
00c-good-scientific-coding-pr... 3.9 KB Apr 7, 2023
01-intro_to_base_R-live.Rmd 19 KB Apr 7, 2023
02-intro_to_ggplot2-live.Rmd 11.5 KB Apr 7, 2023
03-intro_to_tidyverse-live.Rmd 21.3 KB Apr 7, 2023
data
diagrams
exercise_01-intro_to_base_R.R... 4.8 KB Apr 7, 2023
exercise_02-intro_to_R.Rmd 4.7 KB Apr 7, 2023
exercise_03a-intro_to_tidyvers... 7.7 KB Apr 7, 2023
exercise_03b-intro_to_tidyvers... 6.2 KB Apr 7, 2023
README.md 1.8 KB Apr 7, 2023


rstudio.ccdatalab.org
File Edit Code View Plots Session Build Debug Profile Tools Help
datalab
Go to file/function
Addins
Project: (None)
01-intro_to_base_R-live.Rmd*
Preview on Save
Preview
Run
Source
Visual
Outline
[1] 30
95
96 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 not work and will generate an error, or worse.
97
98 ```{r expressions, live = TRUE}
99 x <- 5.5
100
101 x
102 ```
[1] 5.5
103
104
118:91
Defining and using variables
R Markdown
Console
Terminal
Jobs
R 4.2.3 · ~/
> x <- 5.5
>
> x
[1] 5.5
>
Environment
History
Connections
Tutorial
Import Dataset
172 MiB
List
R
Global Environment
Values
x
5.5
Other Assistance Tabs:
Things that help you in your coding
Files
Plots
Packages
Help
Viewer
Home
training-modules
intro-to-R-tidyverse
Name
Size
Modified
00b-debugging_resources.md 13.6 KB Apr 7, 2023
00c-good-scientific-coding-pr... 3.9 KB Apr 7, 2023
01-intro_to_base_R-live.Rmd 19 KB Apr 7, 2023
02-intro_to_ggplot2-live.Rmd 11.5 KB Apr 7, 2023
03-intro_to_tidyverse-live.Rmd 21.3 KB Apr 7, 2023
data
diagrams
exercise_01-intro_to_base_R.R... 4.8 KB Apr 7, 2023
exercise_02-intro_to_R.Rmd 4.7 KB Apr 7, 2023
exercise_03a-intro_to_tidyvers... 7.7 KB Apr 7, 2023
exercise_03b-intro_to_tidyvers... 6.2 KB Apr 7, 2023
README.md 1.8 KB Apr 7, 2023

RStudio Sessions

- On the server, R is running many times at once
  - Each user has their own "**Session**" running, with its own memory and processes
- We will usually want to start new sessions between notebooks to keep the environment clean