

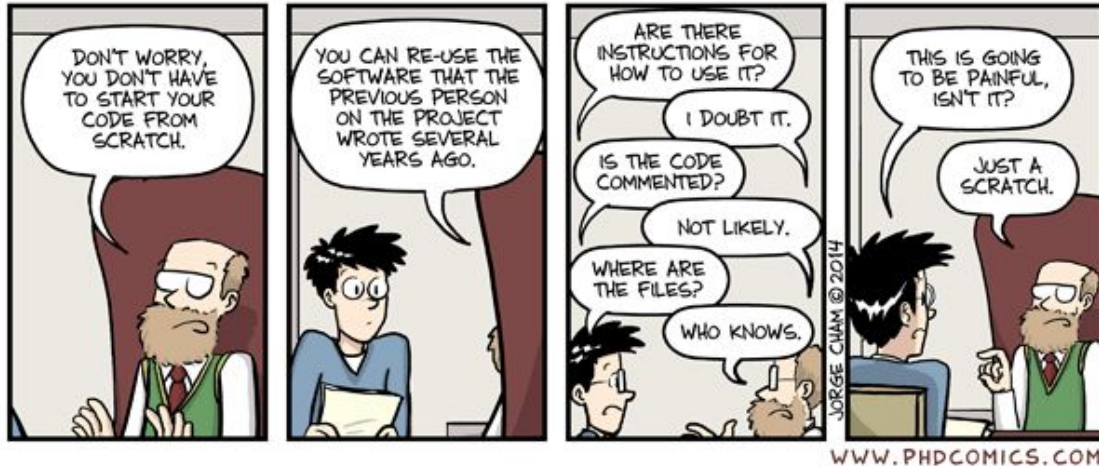


Introduction to R, RStudio, and RStudio Server

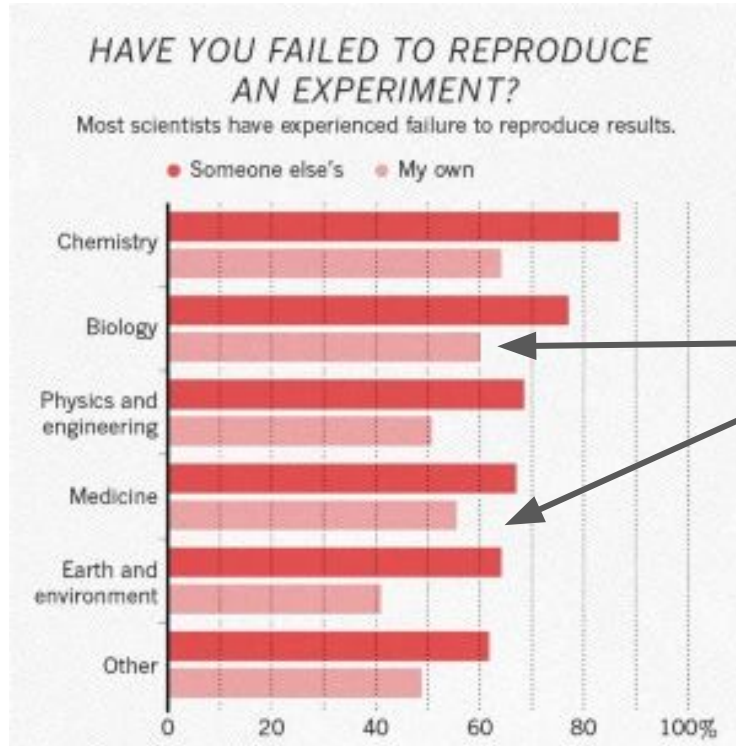
The Data Lab

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 their own 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

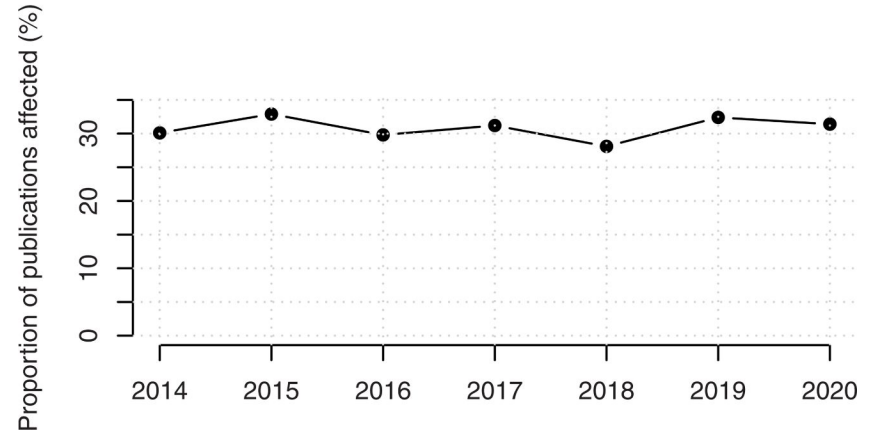
Despite geneticists being warned about spreadsheet problems, 30% of published papers contain mangled gene names in supplementary data.

[Dyani Lewis](#)



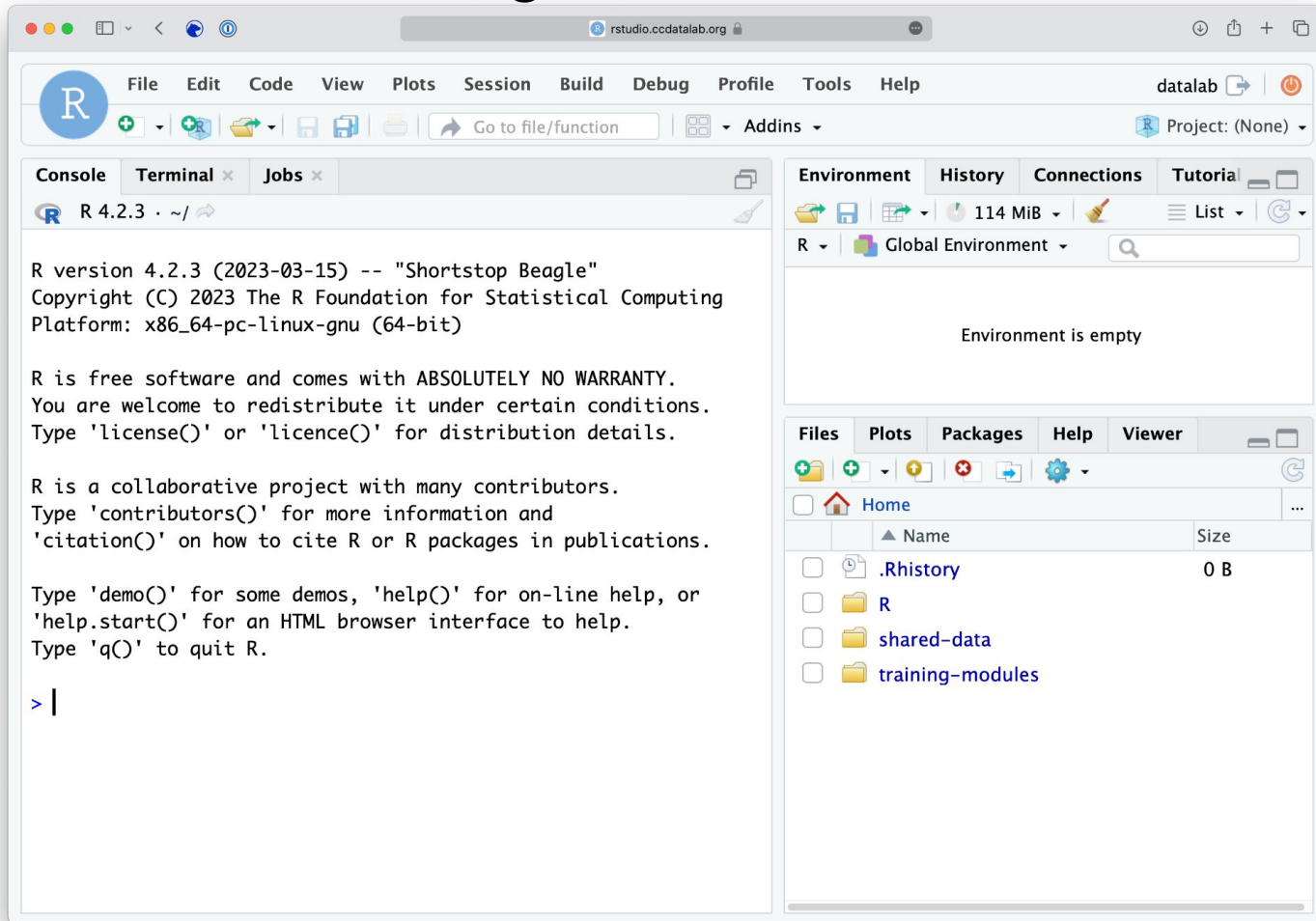
Embarrassing autocorrect mistakes are common fodder for Internet listicles and Twitter threads. But they are also the bane of geneticists using spreadsheet programs such as Microsoft Excel. Five years after a study showed that [autocorrect problems](#) were widespread, the academic literature is still 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



The screenshot displays the RStudio Server web interface. The top navigation bar includes menus for File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The console pane on the left shows the R version 4.2.3 (2023-03-15) and the license information. The environment pane on the right shows the Global Environment, which is currently empty. The file browser at the bottom right shows the Home directory with files and folders: .Rhistory, R, shared-data, and training-modules.

```
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.

```
> |
```

Environment History Connections Tutorial

114 MiB

Global Environment

Environment is empty

Files Plots Packages Help Viewer

Home

	Name	Size
<input type="checkbox"/>	.Rhistory	0 B
<input type="checkbox"/>	R	
<input type="checkbox"/>	shared-data	
<input type="checkbox"/>	training-modules	

The image shows a screenshot of the RStudio web interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The top right corner shows 'datalab' and a power icon. Below the menu bar is a toolbar with various icons and a search bar labeled 'Go to file/function'. The main workspace is divided into several panes. The left pane is the 'Console', which is highlighted with a red border. It shows the R version 4.2.3 (2023-03-15) -- "Shortstop Beagle" and copyright information. The right pane is the 'Environment' pane, which shows 'Global Environment' and 'Environment is empty'. Below the Environment pane is the 'Files' pane, which shows a file browser view with a table of files and folders.

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

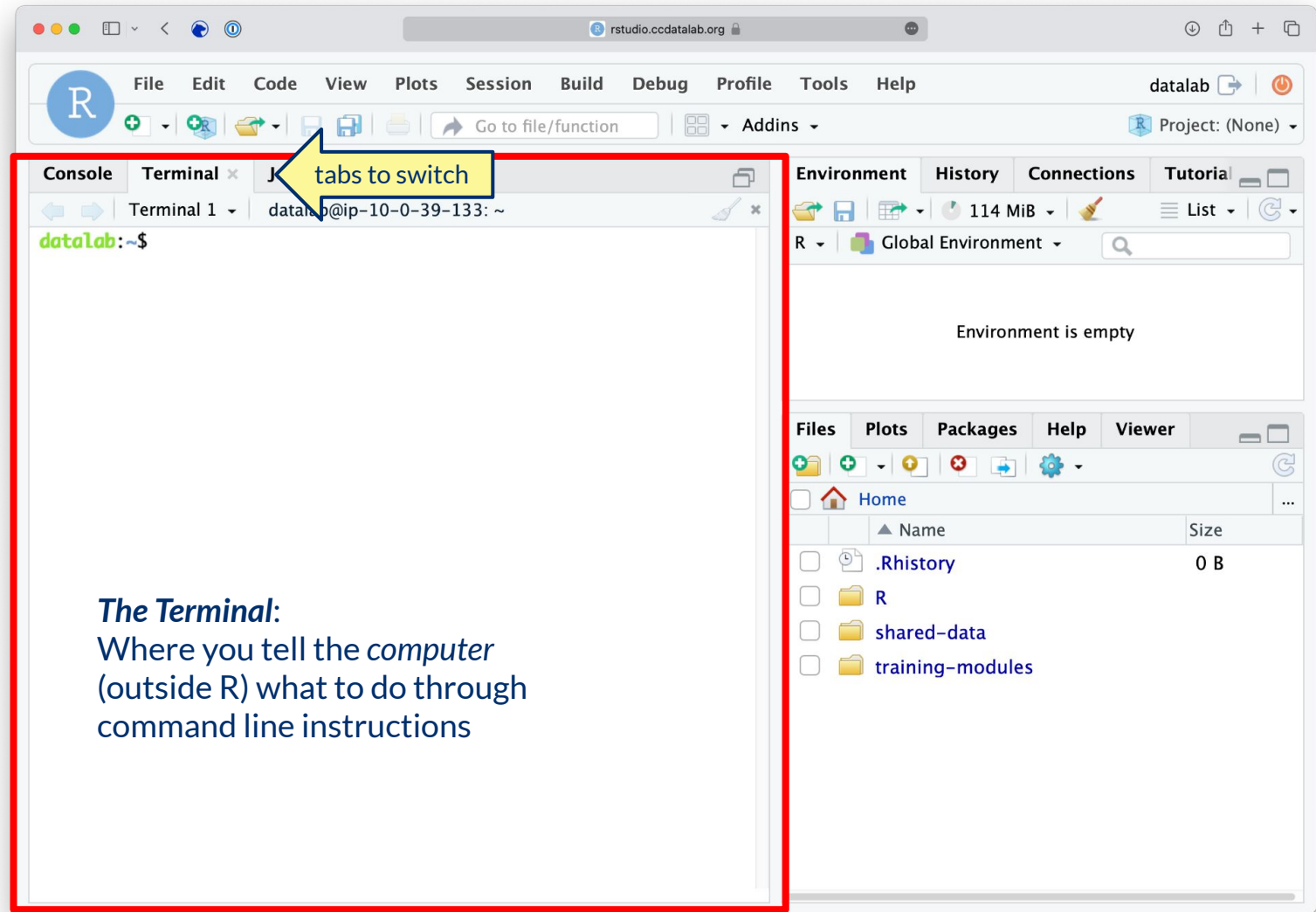
Global Environment

Environment is empty

Files Plots Packages Help Viewer

Home

	Name	Size
<input type="checkbox"/>	.Rhistory	0 B
<input type="checkbox"/>	R	
<input type="checkbox"/>	shared-data	
<input type="checkbox"/>	training-modules	



The Terminal:

Where you tell the *computer* (outside R) what to do through command line instructions

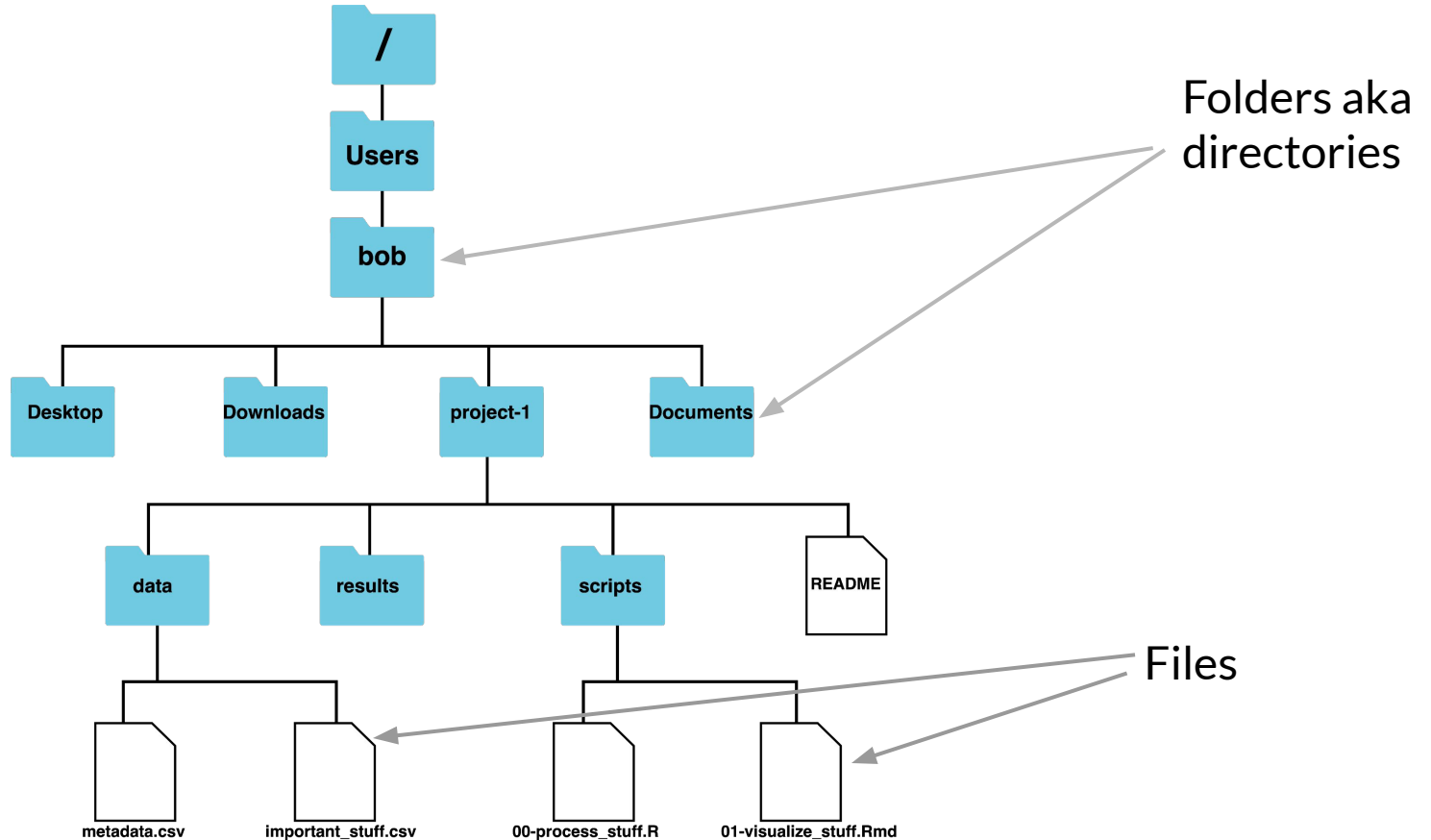
These indicate what **directory** you are **currently** carrying out a command in

This is called your "**current directory**"

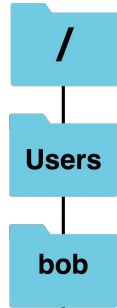
"~" is a shortcut for your "**Home**" directory, so these mean the same thing.

The screenshot shows the RStudio interface with a terminal window highlighted in red. The terminal window displays the prompt `dataLab:~$`. An arrow points from the tilde character (~) to the text below. The RStudio interface includes a menu bar (File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help), a toolbar, and a console area. The Environment pane shows 'Global Environment' and 'Environment is empty'. The Files pane shows a list of files and folders: Home, .Rhistory (0 B), R, shared-data, and training-modules.

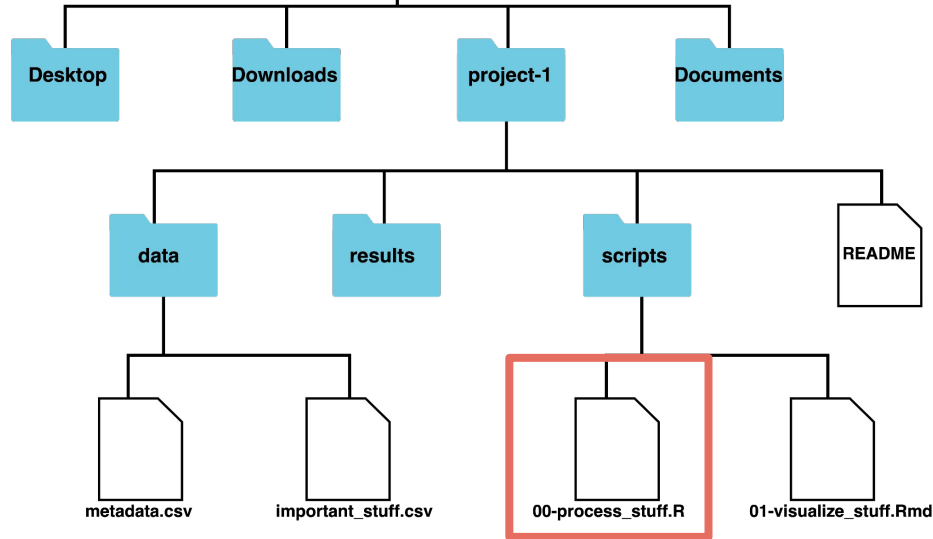
Example of a filesystem hierarchy



We are always working somewhere!



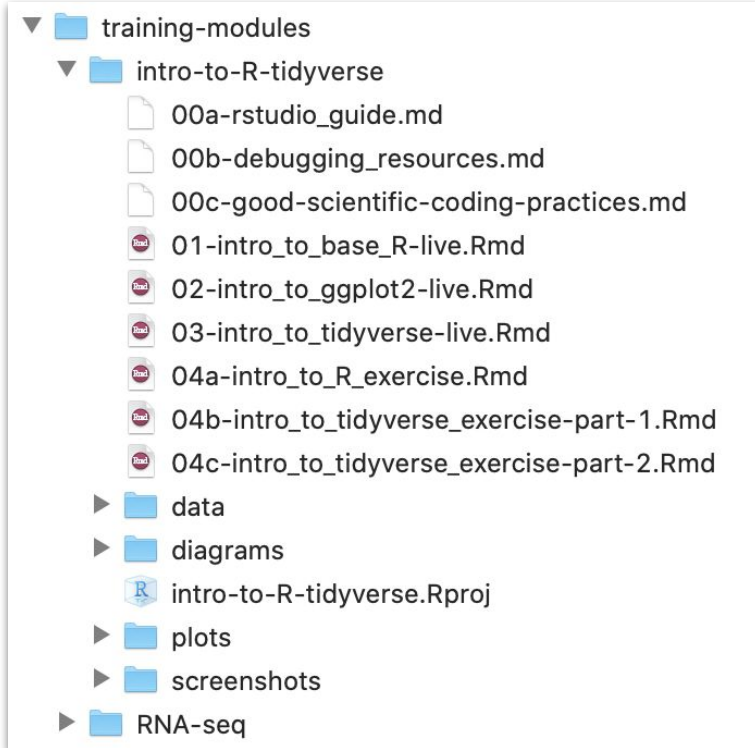
Assume we are working “from” the **bob** directory.
This means **bob** is the *current (working) directory*



The file we are working on
Relative path: **project-1/scripts/00-process_stuff.R**

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.



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.

The screenshot shows the RStudio interface with the following components:

- Terminal:** Shows the prompt `dataLab:~$`. The `~` indicates the current directory is the user's home directory.
- Environment:** Shows "Environment is empty".
- Files:** Lists the following files and folders in the home directory:

Name	Size
.Rhistory	0 B
R	
shared-data	
training-modules	

The image shows the RStudio interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The main workspace is divided into several panes:

- Console/Terminal:** Shows the following commands and output:

```
dataLab:~$ ls
R shared-data training-modules
dataLab:~$ cd training-modules/
```
- Environment:** Shows the Global Environment, which is currently empty.
- Files:** Shows the file browser for the Home directory, listing the following files and folders:

	Name	Size	Modified
<input type="checkbox"/>	.Rhistory	0 B	Jan 31, 2023,
<input type="checkbox"/>	R		
<input type="checkbox"/>	shared-data		
<input type="checkbox"/>	training-modules		

The screenshot shows the RStudio interface with the following components:

- Terminal:** Shows a sequence of commands: `ls`, `cd training-modules/`, and `ls`. The output of the second `ls` command is: `LICENSE.md README.md intro-to-R-tidyverse module-cheatsheets scrna-seq`. The current prompt is `~/training-modules$`.
- Environment Panel:** Shows "Global Environment" and "Environment is empty".
- Files Panel:** Shows the file explorer for the `~/training-modules` directory. The files listed are:

	Name	Size	Modified
<input type="checkbox"/>	.Rhistory	0 B	Jan 31, 2023
<input type="checkbox"/>	R		
<input type="checkbox"/>	shared-data		
<input type="checkbox"/>	training-modules		

The words `training-modules` in the terminal prompt and the file explorer are highlighted in red.

The words in front of our cursor have changed because we are now "in" the `training-modules` directory

The screenshot shows the RStudio interface with the following components:

- Terminal Pane:** Shows the execution of commands: `ls`, `cd training-modules/`, and `ls`. The output of the second `ls` command is: `LICENSE.md README.md intro-to-R-tidyverse module-cheatsheets scrna-seq`. The current directory path `~/training-modules` is highlighted with a red box.
- Environment Pane:** Shows "Global Environment" and "Environment is empty".
- Files Pane:** Shows a file browser view of the `Home` directory with the following table:

	Name	Size	Modified
<input type="checkbox"/>	.Rhistory	0 B	Jan 31, 2023
<input type="checkbox"/>	R		
<input type="checkbox"/>	shared-data		
<input type="checkbox"/>	training-modules		

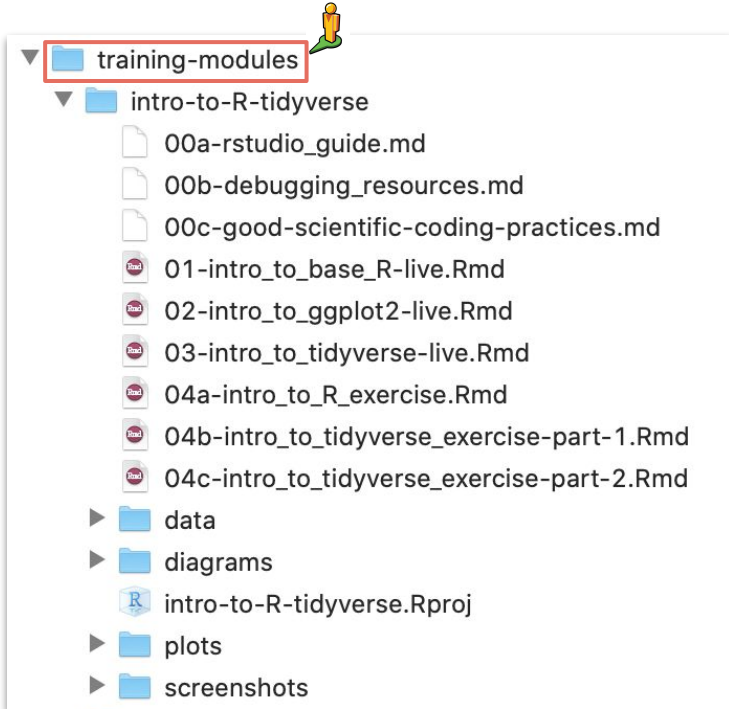
The files pane shows the `Home` directory with a list of files and folders. The `training-modules` folder is visible, but the contents of the terminal are not reflected in this pane.

The words in front of our cursor have changed because we are now "in" the `training-modules` directory

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`

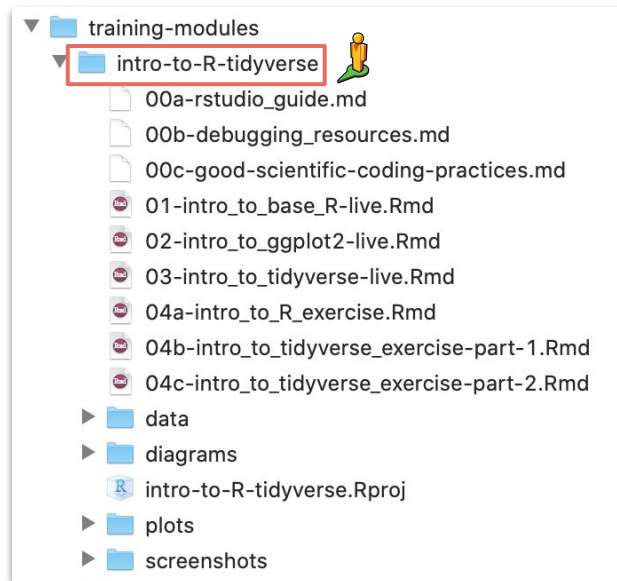
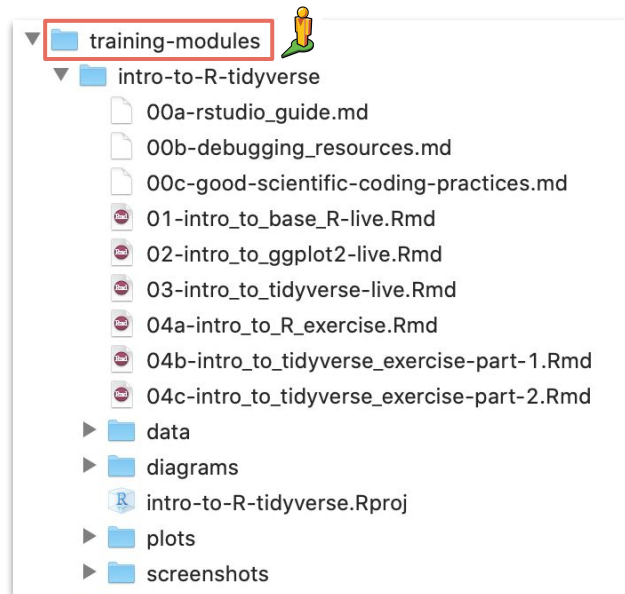


Current directory = `training-modules` 

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

Relative file paths

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



training-modules



training-modules/intro-to-R-tidyverse

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

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.

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)

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`

The screenshot displays the RStudio interface. The main editor shows the source code of an R Notebook. The code includes a title, author information, and a list of objectives for the notebook. The right-hand pane shows the 'Files' tab, which is currently displaying the file browser for the directory `training-modules/intro-to-R-tidyverse`. The file `01-intro_to_base_R-live.Rmd` is highlighted in the file list.

```
1 ---
2 title: "Introduction to R and RStudio"
3 author: Originally authored by Stephanie J. Spielman, <br> adapted by CC DL
  for ALSF
4 date: 2021
5 output:
6   html_notebook:
7     toc: true
8     toc_float: true
9 ---
10
11 ## Objectives
12
13 This notebook will demonstrate how to:
14
15 - Navigate the RStudio environment
16 - Use R for simple calculations, both mathematical and logical
17 - Define and use variables in base R
18 - Understand and apply base R functions
19 - Understand, define, and use R data types, including vector
  manipulation and indexing
20 - Understand the anatomy of a data frame
21
22 ---
23
24 #### *More resources for learning R*
25
26 - [Swirl, an interactive tutorial](https://swirlstats.com/)
```

1:1 Introduction to R and RStudio R Markdown

Environment: Global Environment (Empty)

Files: Home > training-modules > intro-to-R-tidyverse

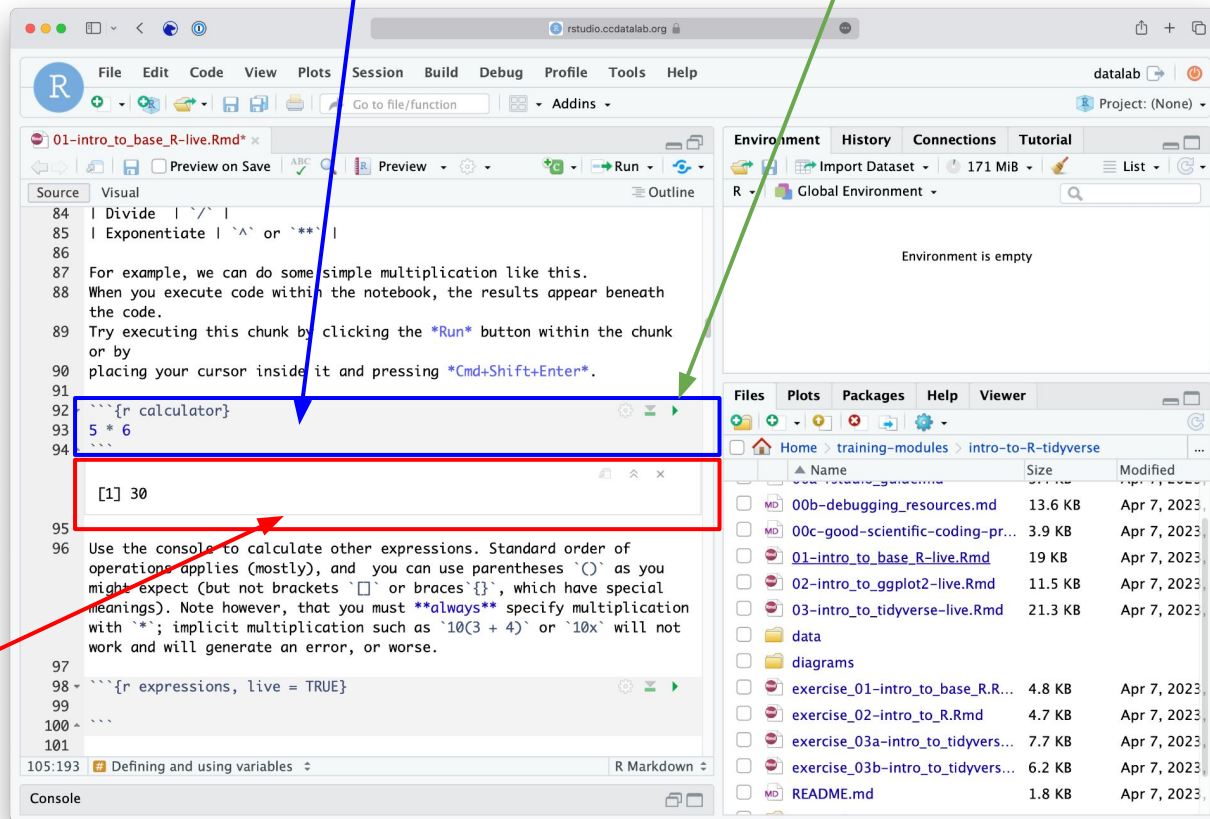
Name	Size	Modified
..		
00a-rstudio_guide.md	3.4 KB	Apr 7, 2023
00b-debugging_resources.md	13.6 KB	Apr 7, 2023
00c-good-scientific-coding-pr...	3.9 KB	Apr 7, 2023
01-intro_to_base_R-live.Rmd	19 KB	Apr 7, 2023
02-intro_to_ggplot2-live.Rmd	11.5 KB	Apr 7, 2023
03-intro_to_tidyverse-live.Rmd	21.3 KB	Apr 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_tidvers...	7.7 KB	Apr 7, 2023
exercise_03b-intro_to tidvers...	6.2 KB	Apr 7, 2023

R Notebooks

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

Executable code chunk

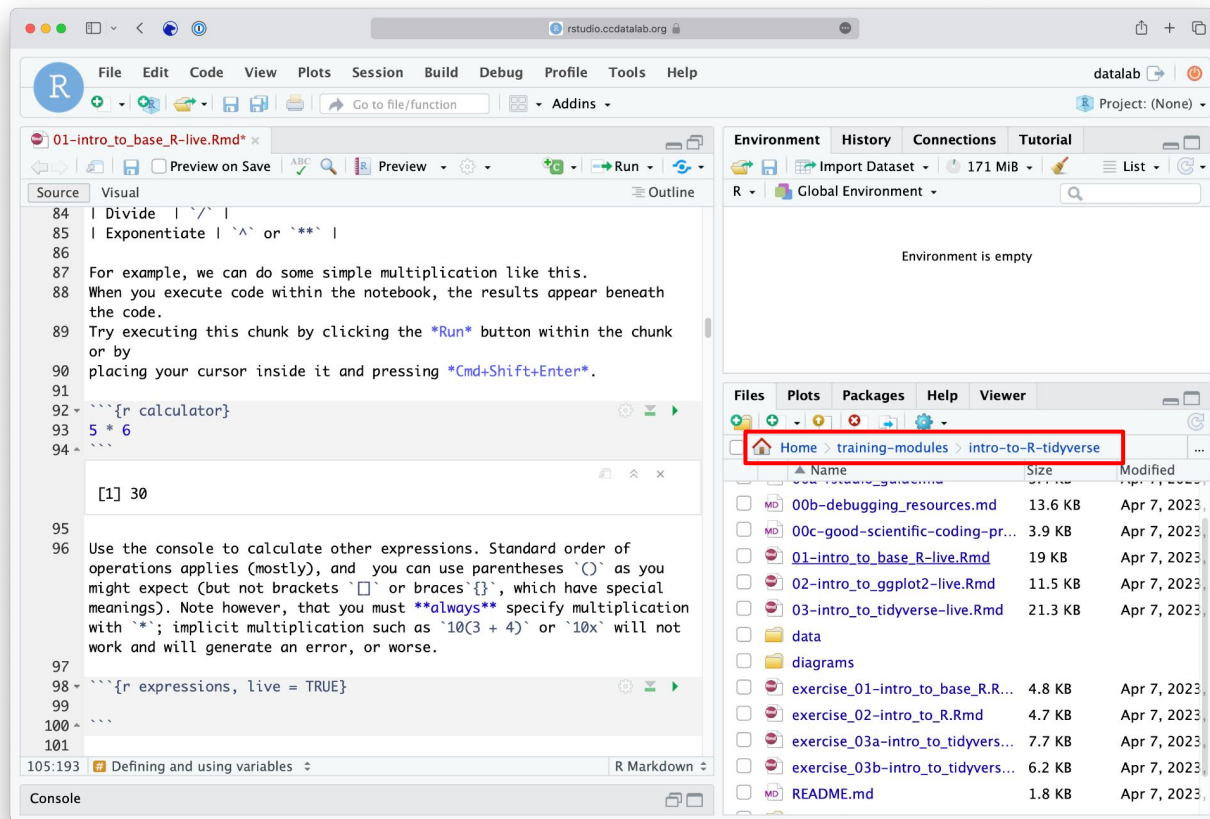
Can click here to run a code chunk



Output from above code chunk

R Notebooks

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





File Edit Code View Plots Session Build Debug Profile Tools Help

datalab

01-intro_to_base_R-live.Rmd*

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
89 the code.
90 Try executing this chunk by clicking the *Run* button within the chunk
91 or by
92 placing your cursor inside it and pressing *Cmd+Shift+Enter*.
93
94 ```{r calculator}
95 5 * 6
96 ```
```

[1] 30

```
97
98 Use the console to calculate other expressions. Standard order of
99 operations applies (mostly), and you can use parentheses `()` as you
100 might expect (but not brackets `[]` or braces `{}`, which have special
101 meanings). Note however, that you must **always** specify multiplication
102 with `*`; implicit multiplication such as `10(3 + 4)` or `10x` will not
103 work and will generate an error, or worse.
104
105 ```{r expressions, live = TRUE}
106
107 ```
```

105:193 # Defining and using variables R Markdown

Console

Environment History Connections Tutorial

Import Dataset 171 MiB

R Global Environment

Environment is empty

Files Plots Packages Help Viewer

Home > training-modules > intro-to-R-tidyverse

Name	Size	Modified
00a-intro_to_tidyverse.md	13.6 KB	Apr 7, 2023
00b-debugging_resources.md	3.9 KB	Apr 7, 2023
00c-good-scientific-coding-pr...	19 KB	Apr 7, 2023
01-intro_to_base_R-live.Rmd	11.5 KB	Apr 7, 2023
02-intro_to_ggplot2-live.Rmd	21.3 KB	Apr 7, 2023
03-intro_to_tidyverse-live.Rmd		
exercise_01a-intro_to_tidyvers...	6.2 KB	Apr 7, 2023
exercise_02a-intro_to_tidyvers...	1.8 KB	Apr 7, 2023
exercise_03b-intro_to_tidyvers...		
README.md		

Click here to show the Console



The screenshot displays the RStudio interface. The main editor window shows a script titled "01-intro_to_base_R-live.Rmd" with the following code:

```
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 ```
```

The Environment pane on the right shows the Global Environment with a variable `x` having a value of `5.5`.

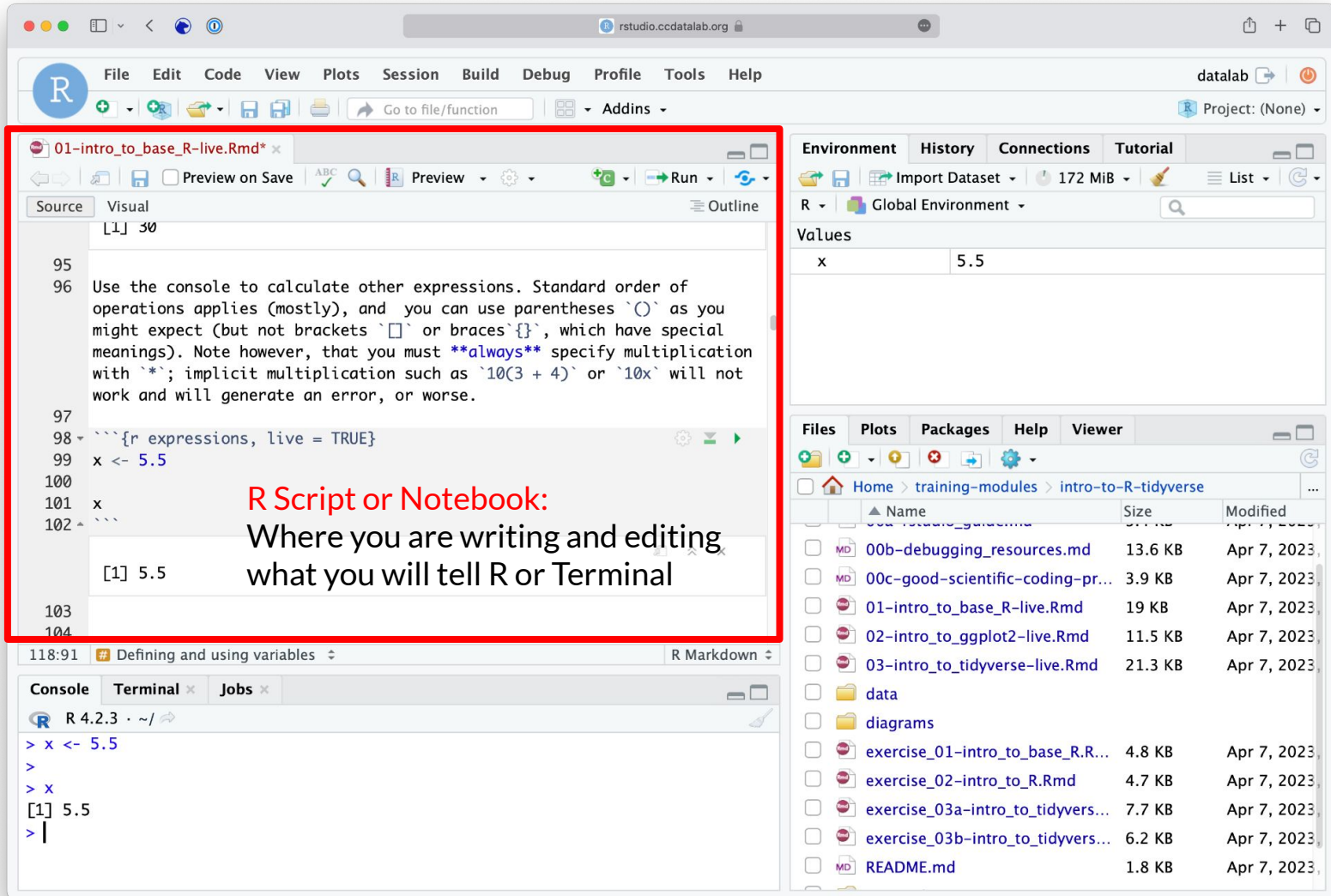
The Files pane on the right shows a directory structure for `training-modules > intro-to-R-tidyverse` with the following files and sizes:

Name	Size	Modified
00a-creating_variables.md	13.6 KB	Apr 7, 2023
00b-debugging_resources.md	3.9 KB	Apr 7, 2023
00c-good-scientific-coding-pr...	19 KB	Apr 7, 2023
01-intro_to_base_R-live.Rmd	11.5 KB	Apr 7, 2023
02-intro_to_ggplot2-live.Rmd	21.3 KB	Apr 7, 2023
03-intro_to_tidyverse-live.Rmd		
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

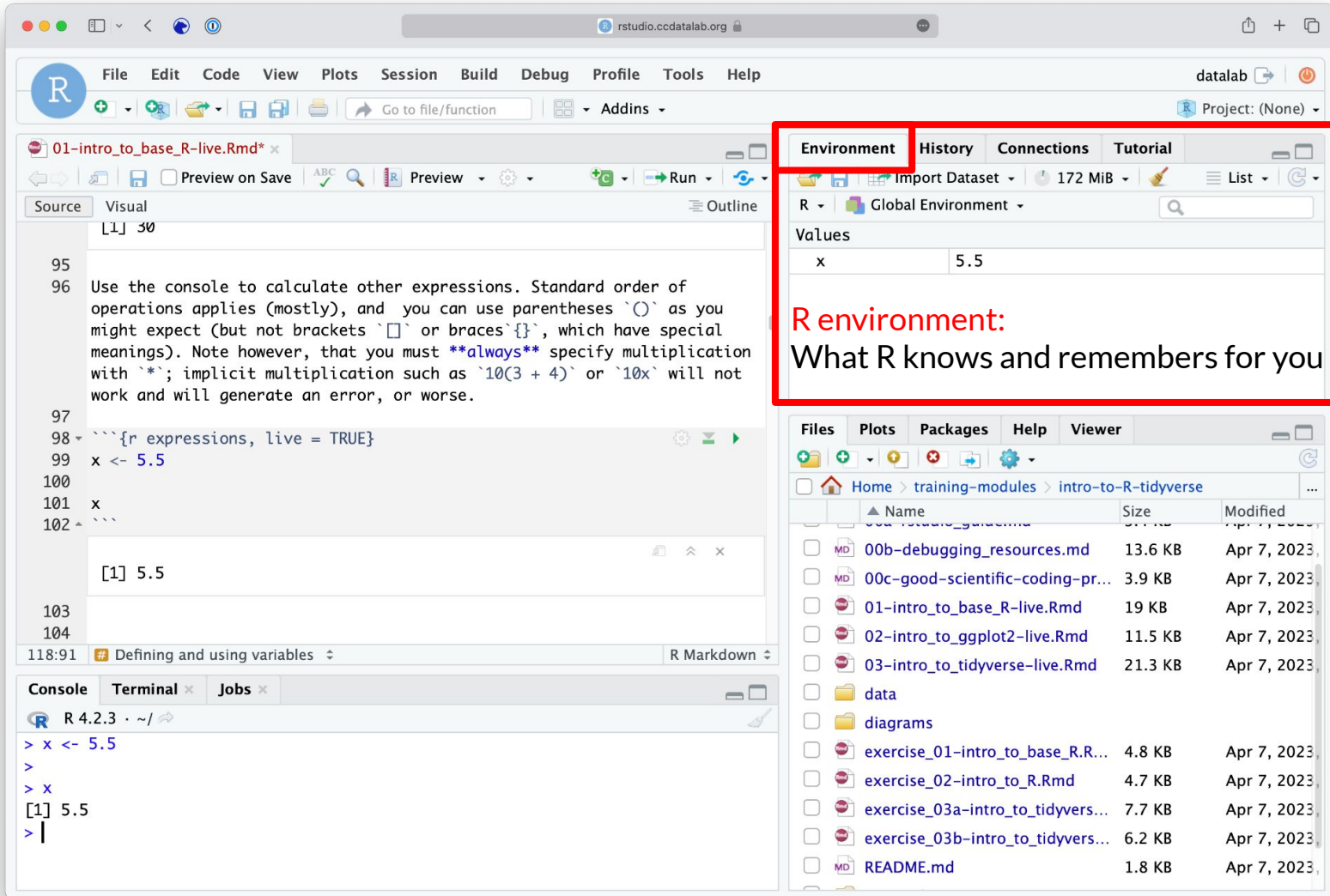
The Console pane at the bottom, highlighted with a red border, shows the R session output:

```
R 4.2.3 · ~/ /  
> x <- 5.5  
>  
> x  
[1] 5.5  
> |
```

R Console:
What you are actually telling R to do



R Script or Notebook:
Where you are writing and editing
what you will tell R or Terminal



File Edit Code View Plots Session Build Debug Profile Tools Help

datalab

01-intro_to_base_R-live.Rmd*

Source Visual

```
[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

Console Terminal Jobs

```
R 4.2.3 · ~/
> x <- 5.5
>
> x
[1] 5.5
> |
```

Environment History Connections Tutorial

Import Dataset 172 MiB

R Global Environment

Values

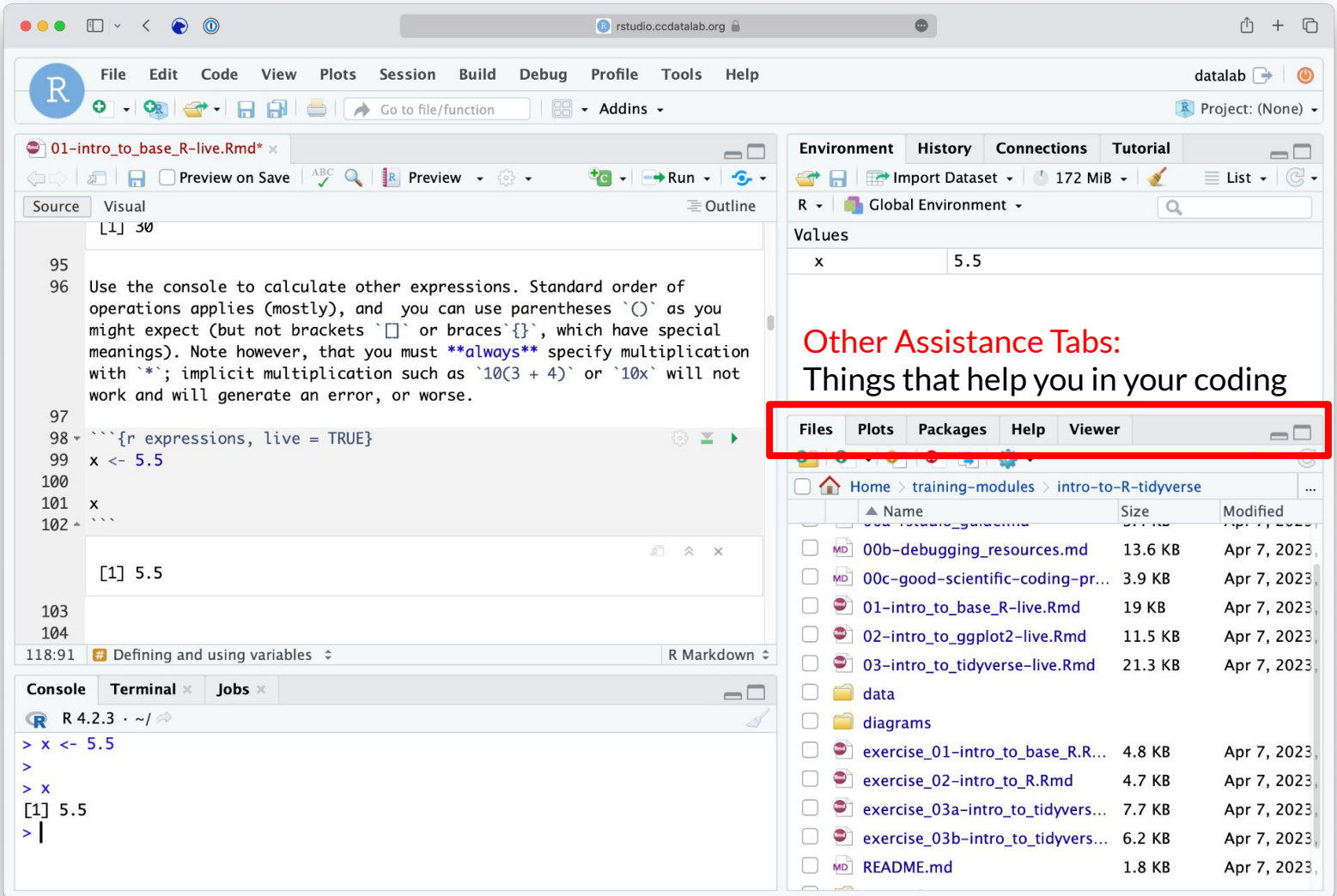
x	5.5
---	-----

R environment:
What R knows and remembers for you

Files Plots Packages Help Viewer

Home > training-modules > intro-to-R-tidyverse

Name	Size	Modified
00a-creating_a_garbage		
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



Other Assistance Tabs:
Things that help you in your coding



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

Log out of website



End the current session and start new session