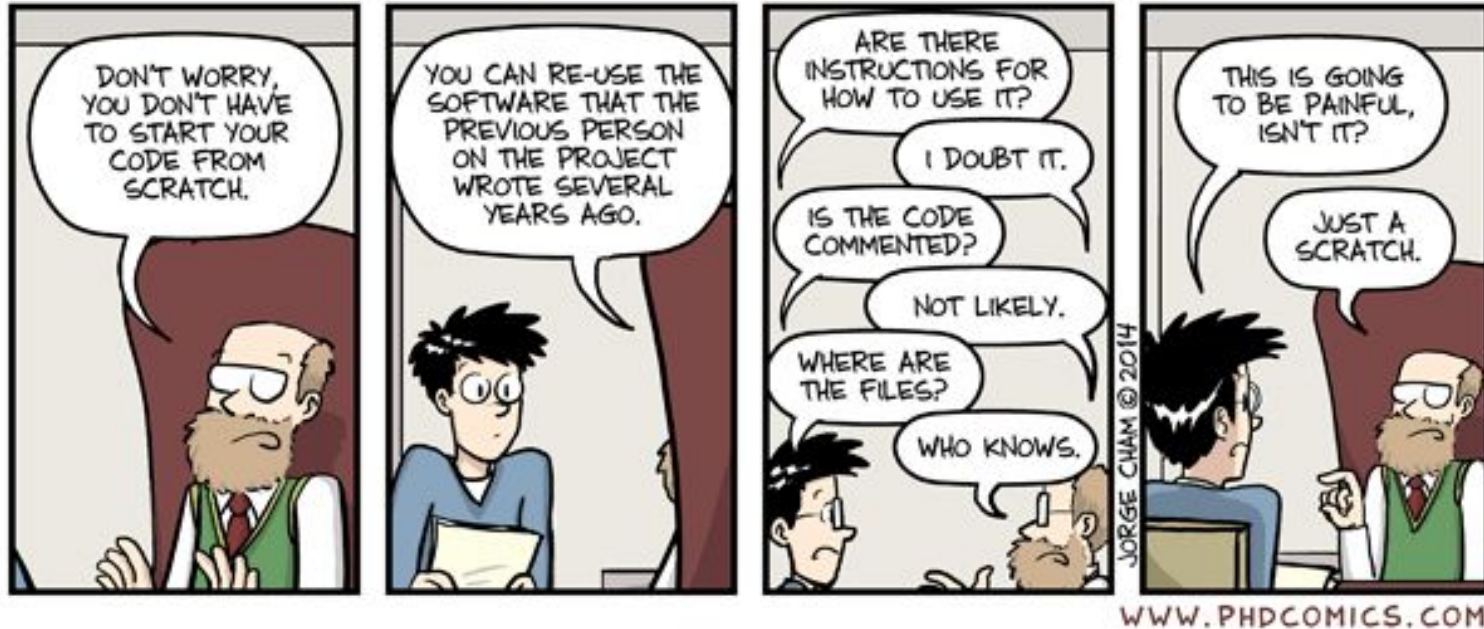




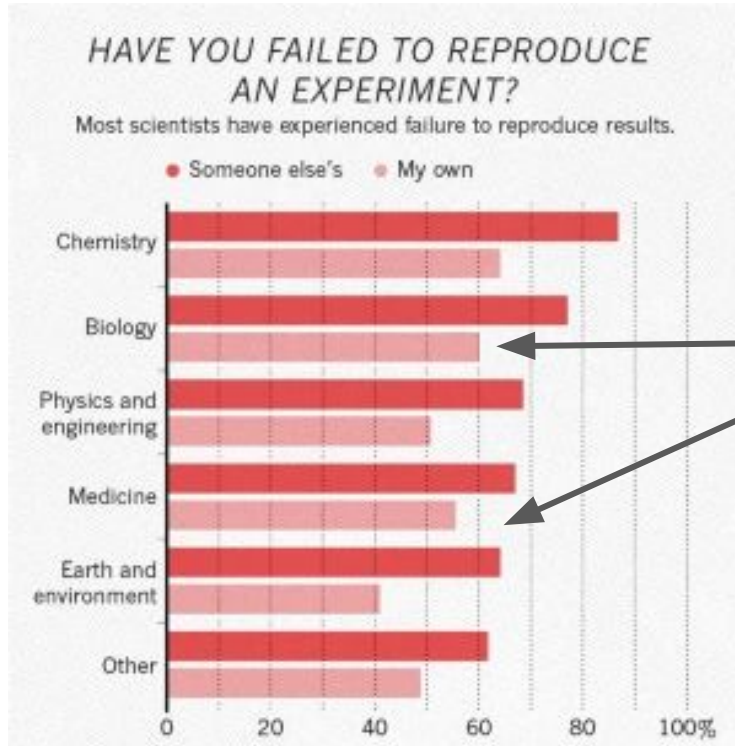
Introduction to R, RStudio, and RStudio Server

The CCDL

Who's been here before?



Reproducibility in 2016



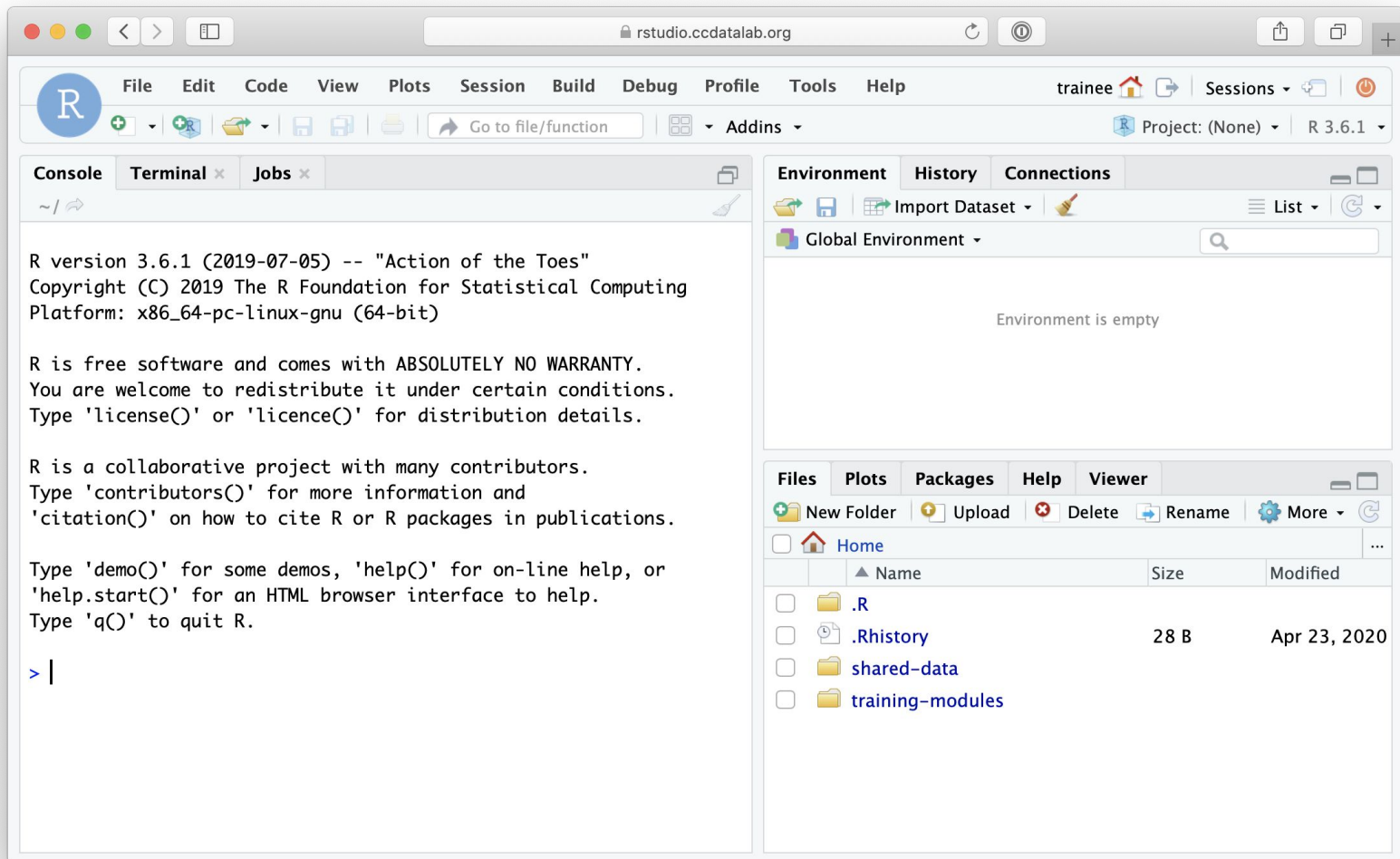
55% and 60% of biologists and clinicians, respectively, could not reproduce their own results.

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

RStudio Server: A basic guide



The screenshot displays the RStudio Server web interface. At the top, the browser address bar shows `rstudio.ccdatalab.org`. The main menu includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The user is logged in as 'trainee' and the R version is 3.6.1. The interface is divided into several panels:

- Console:** Shows the R startup message and license information. The text reads: "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 prompt `> |` is visible at the bottom.
- Environment:** Shows the current environment is empty.
- Files:** A file browser showing the current directory structure. The files listed are:

	Name	Size	Modified
<input type="checkbox"/>	.R		
<input type="checkbox"/>	.Rhistory	28 B	Apr 23, 2020
<input type="checkbox"/>	shared-data		
<input type="checkbox"/>	training-modules		



Console Terminal x Jobs x

~/

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

Global Environment

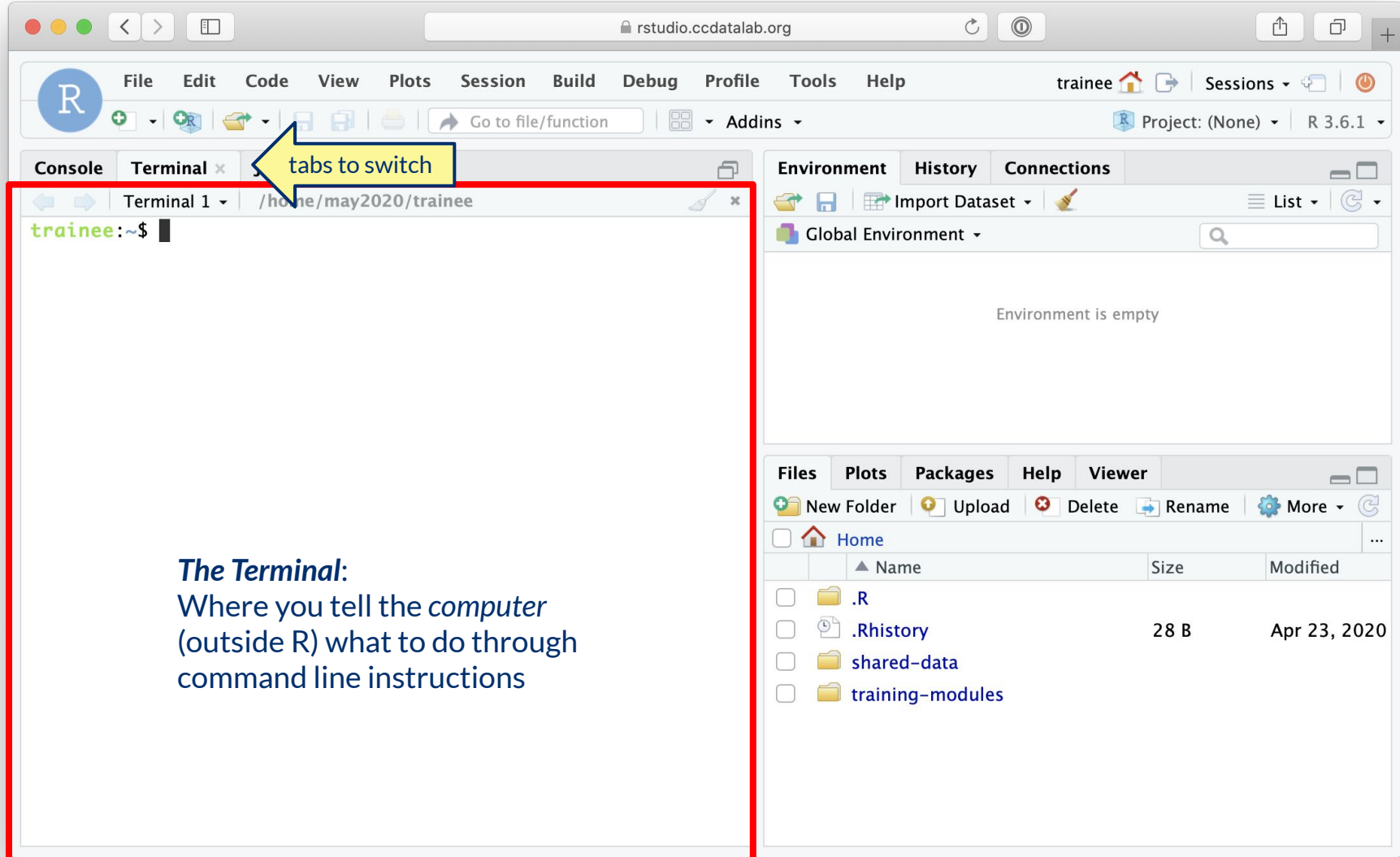
Environment is empty

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

Home

	Name	Size	Modified
<input type="checkbox"/>	.R		
<input type="checkbox"/>	.Rhistory	28 B	Apr 23, 2020
<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

The image shows a screenshot of the RStudio application interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The top right shows the user 'trainee', a home icon, and 'Sessions'. Below the menu bar is a toolbar with icons for file operations and a search bar. The main workspace is divided into several panes. On the left, a terminal window titled 'Terminal 1' is open, showing the prompt 'trainee:~\$' and a cursor. Two arrows point from the text below to the '~' and '\$' characters in the prompt. To the right of the terminal is the 'Environment' pane, which is currently empty. Below the environment pane is the 'Files' pane, showing a file explorer view of the home directory with folders like '.R', '.Rhistory', 'shared-data', and 'training-modules'. The bottom right pane is the 'Viewer' pane, which is currently empty.

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

Global Environment

Environment is empty

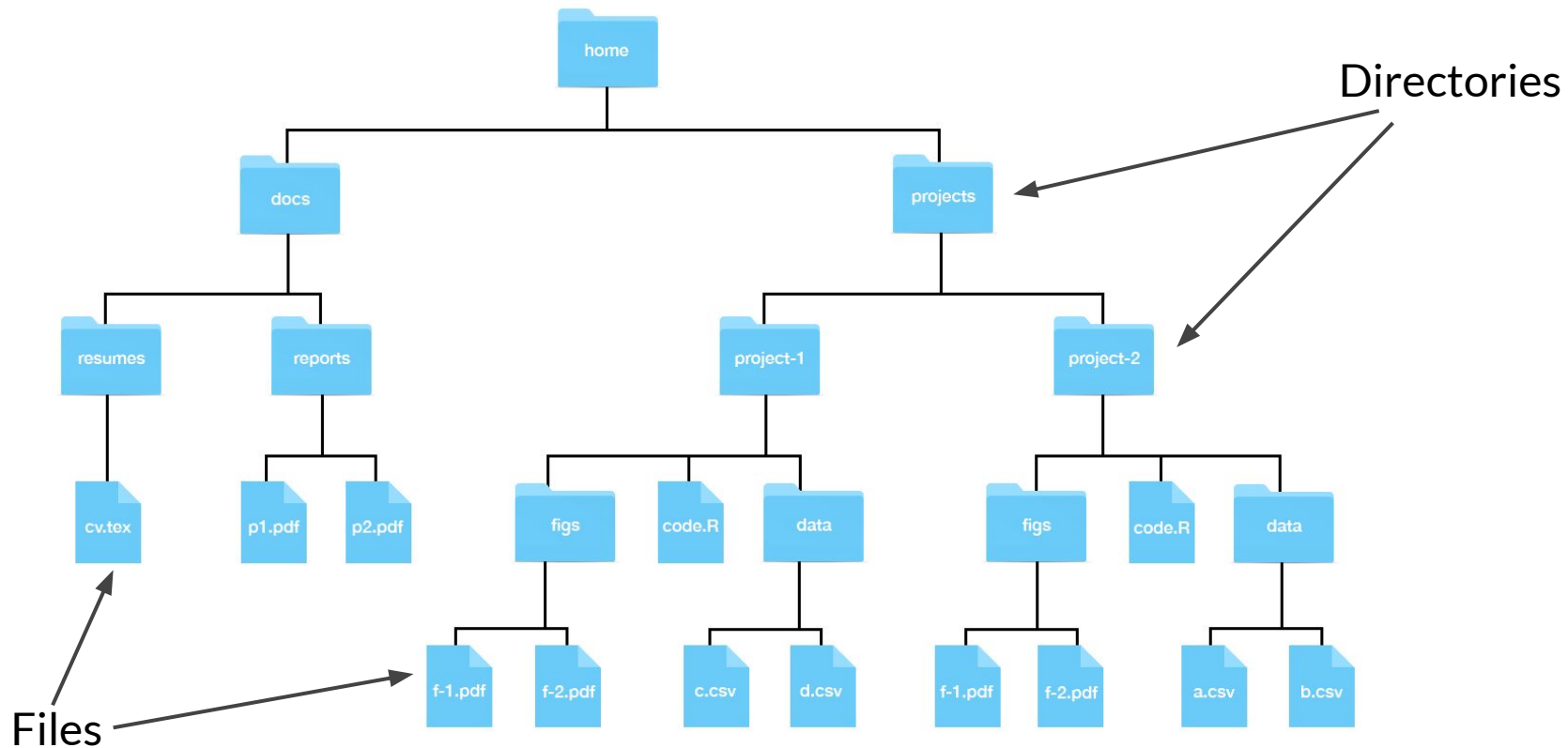
Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

Home

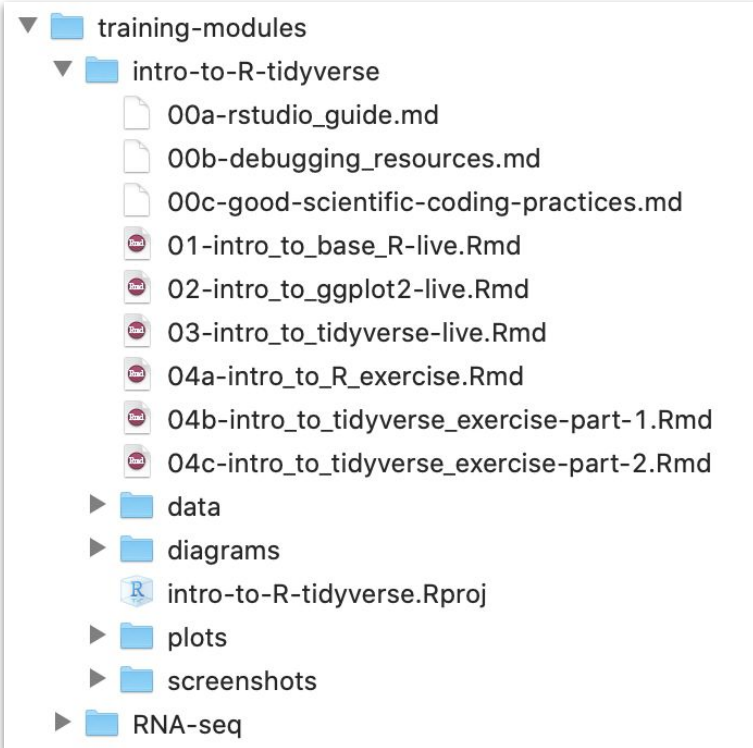
	Name	Size	Modified
<input type="checkbox"/>	.R		
<input type="checkbox"/>	.Rhistory	28 B	Apr 23, 2020
<input type="checkbox"/>	shared-data		
<input type="checkbox"/>	training-modules		

Example of a filesystem “hierarchy”



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.



The image shows the RStudio interface with the following components:

- Terminal:** Shows the execution of `ls` and `cd training-modules` commands in the `~/home/may2020/trainee` directory.
- Environment:** Shows the 'Global Environment' is currently empty.
- Files:** A file explorer view showing the contents of the `~/home` 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

	Name	Size	Modified
<input type="checkbox"/>	.R		
<input type="checkbox"/>	.Rhistory	28 B	Apr 23, 2020
<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 and their output:

```
trainee:~$ ls
shared-data training-modules
trainee:~$ cd training-modules
trainee:~/training-modules$ ls
RNA-seq intro-to-R-tidyverse
trainee:~/training-modules$
```

An arrow points from the text below to the prompt `trainee:~/training-modules$`.
- Environment:** Shows "Global Environment" with the message "Environment is empty".
- Files:** Shows a file explorer view of the current directory with the following items:

	Name	Size	Modified
<input type="checkbox"/>	.R		
<input type="checkbox"/>	.Rhistory	28 B	Apr 23, 2020
<input type="checkbox"/>	shared-data		
<input type="checkbox"/>	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



Console Terminal x Jobs x

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

Global Environment

Environment is empty

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

Home

	Name	Size	Modified
<input type="checkbox"/>	.R		
<input type="checkbox"/>	.Rhistory	28 B	Apr 23, 2020
<input type="checkbox"/>	shared-data		
<input type="checkbox"/>	training-modules		

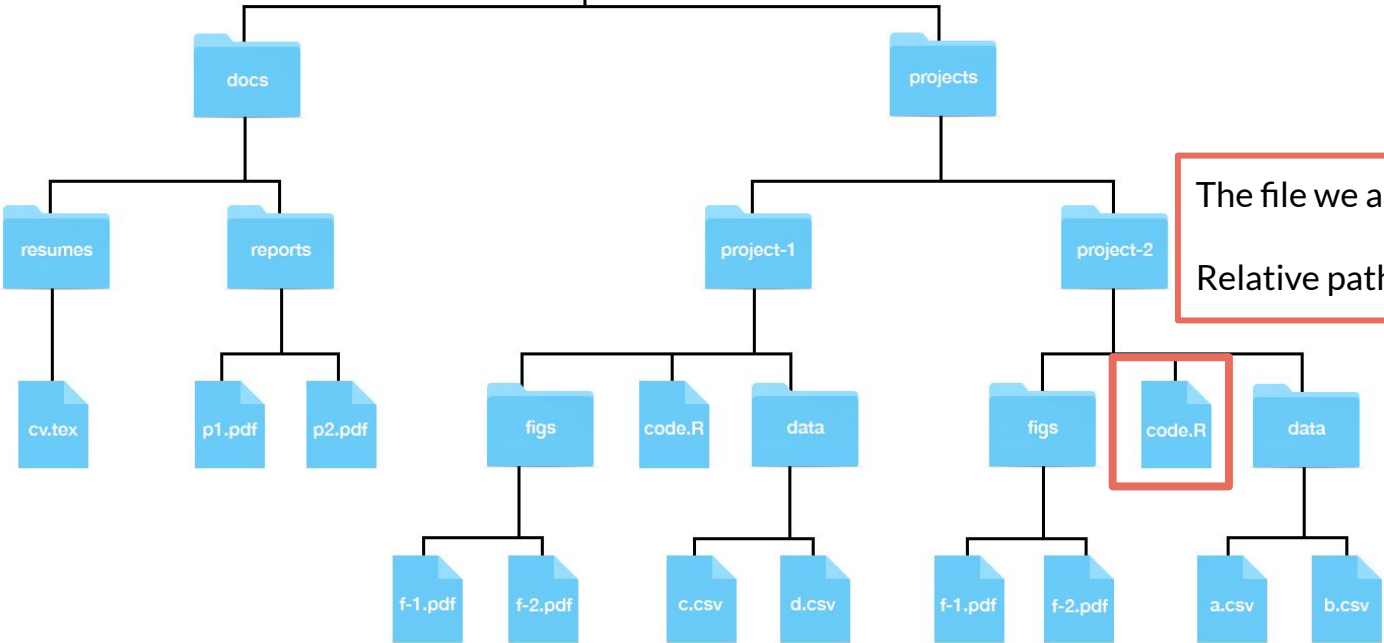
The files tab over here
does **NOT** reflect your current
directory or any changes within it

We are always working somewhere!



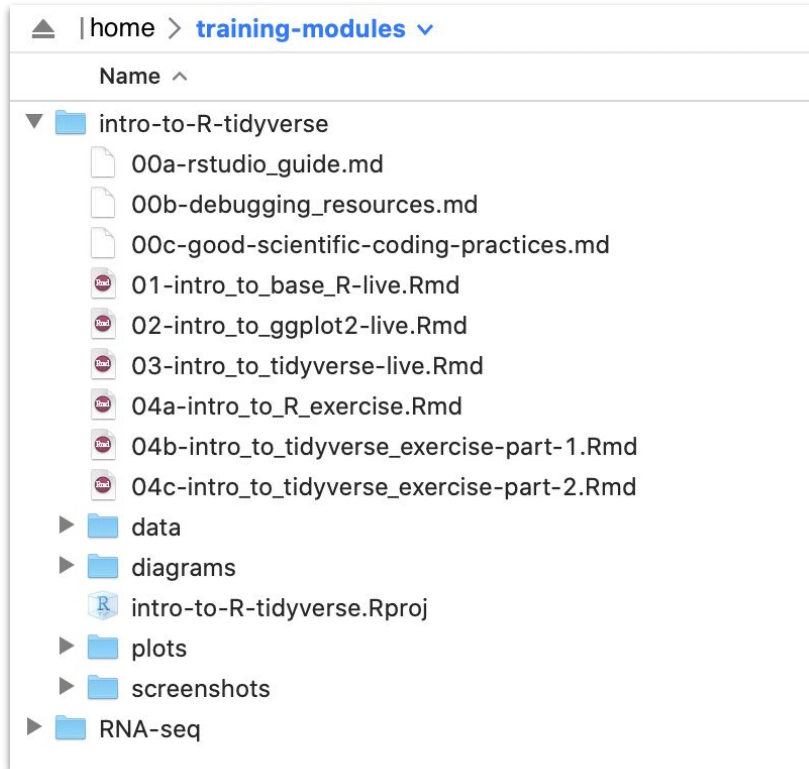
Assume we are working “from” the **home** directory..
This means home is the *current directory*

The file we are working on
Relative path: **projects/project-2/code.R**



File paths: Directions to a file or folder

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

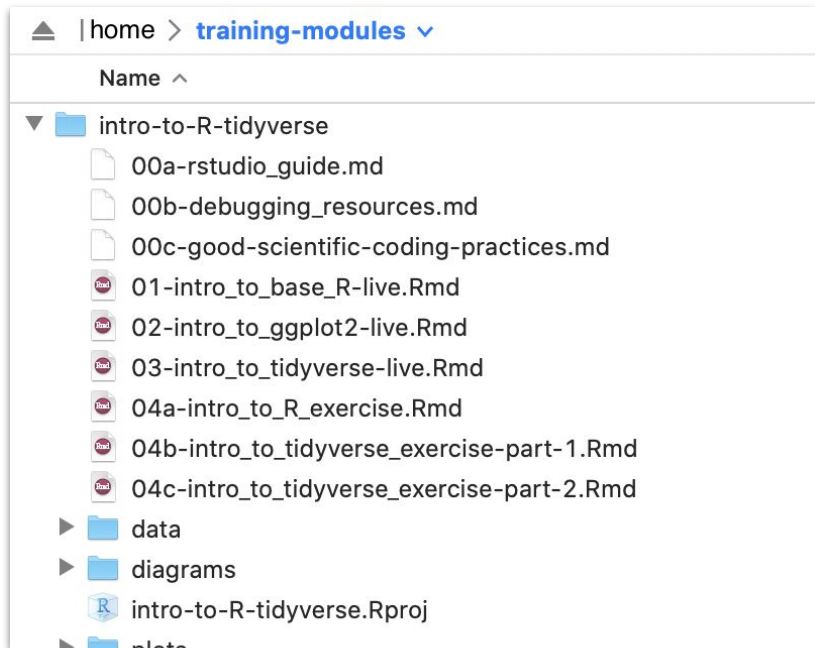


Current directory = "training-modules"

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

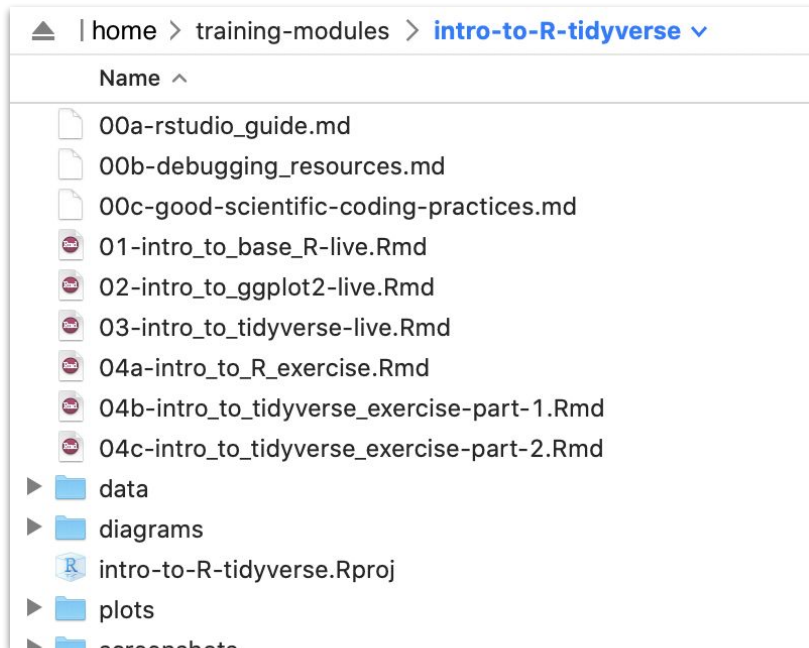
File Paths can be *relative*

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 a company that makes the RStudio IDE

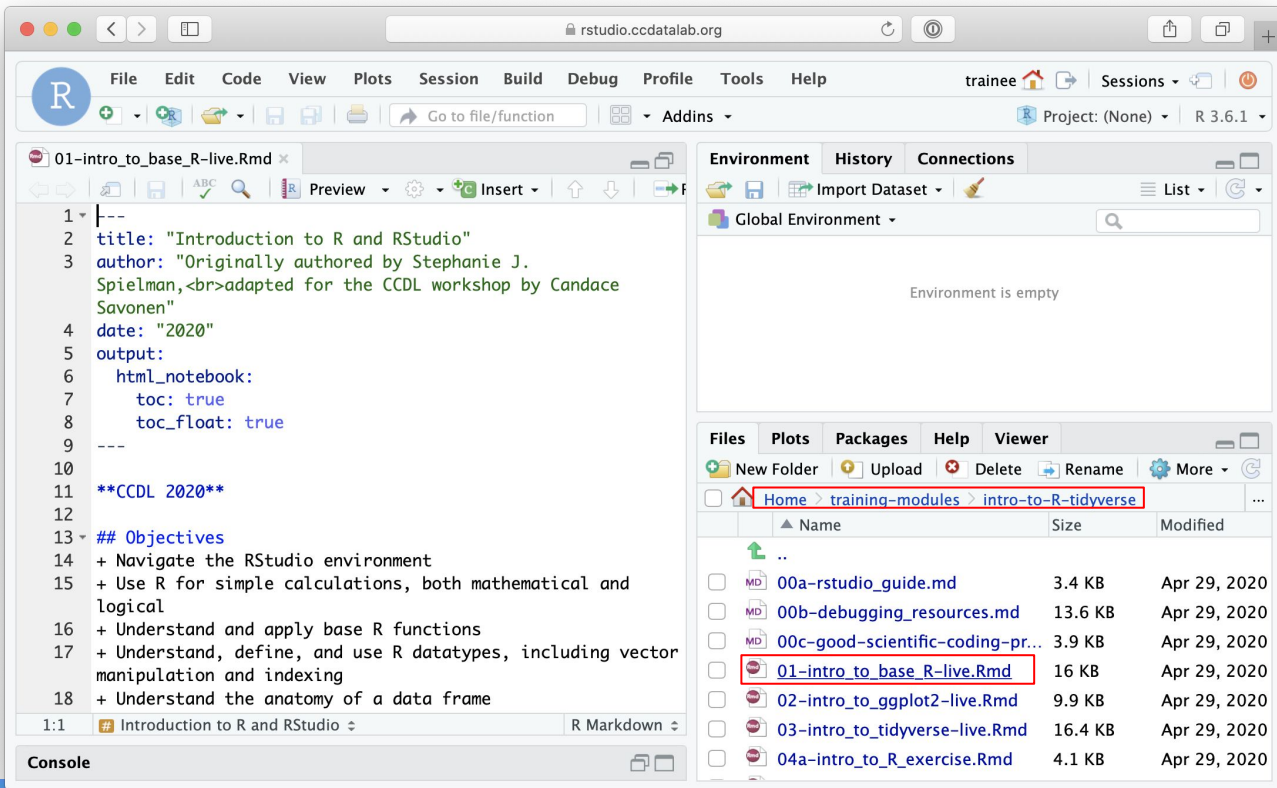
- IDE: Integrated Development Environment
- We write R code using the (free!) RStudio IDE



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 web interface in a browser window. The main editor shows the content of the R Markdown file '01-intro_to_base_R-live.Rmd'. The file content includes a title, author information, date, output settings, and a list of objectives. The 'Files' pane on the right shows the directory structure, with the file '01-intro_to_base_R-live.Rmd' highlighted. The 'Environment' pane is empty, and the 'Console' pane is at the bottom.

```
1 |---
2 |title: "Introduction to R and RStudio"
3 |author: "Originally authored by Stephanie J.
  |Spielman,<br>adapted for the CCDL workshop by Candace
  |Savonen"
4 |date: "2020"
5 |output:
6 |  html_notebook:
7 |    toc: true
8 |    toc_float: true
9 |---
10 |
11 |**CCDL 2020**
12 |
13 |## Objectives
14 |+ Navigate the RStudio environment
15 |+ Use R for simple calculations, both mathematical and
  |logical
16 |+ Understand and apply base R functions
17 |+ Understand, define, and use R datatypes, including vector
  |manipulation and indexing
18 |+ Understand the anatomy of a data frame
```

Files pane path: 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

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

The screenshot shows the RStudio interface. The main editor window displays a code chunk with the following text:

```
82 For example, we can do some simple multiplication like  
83 this.  
84 When you execute code within the notebook, the results  
85 appear beneath the code.  
86 Try executing this chunk by clicking the *Run* button  
87 within the chunk or by  
88 placing your cursor inside it and pressing  
89 *Cmd+Shift+Enter*.  
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
```

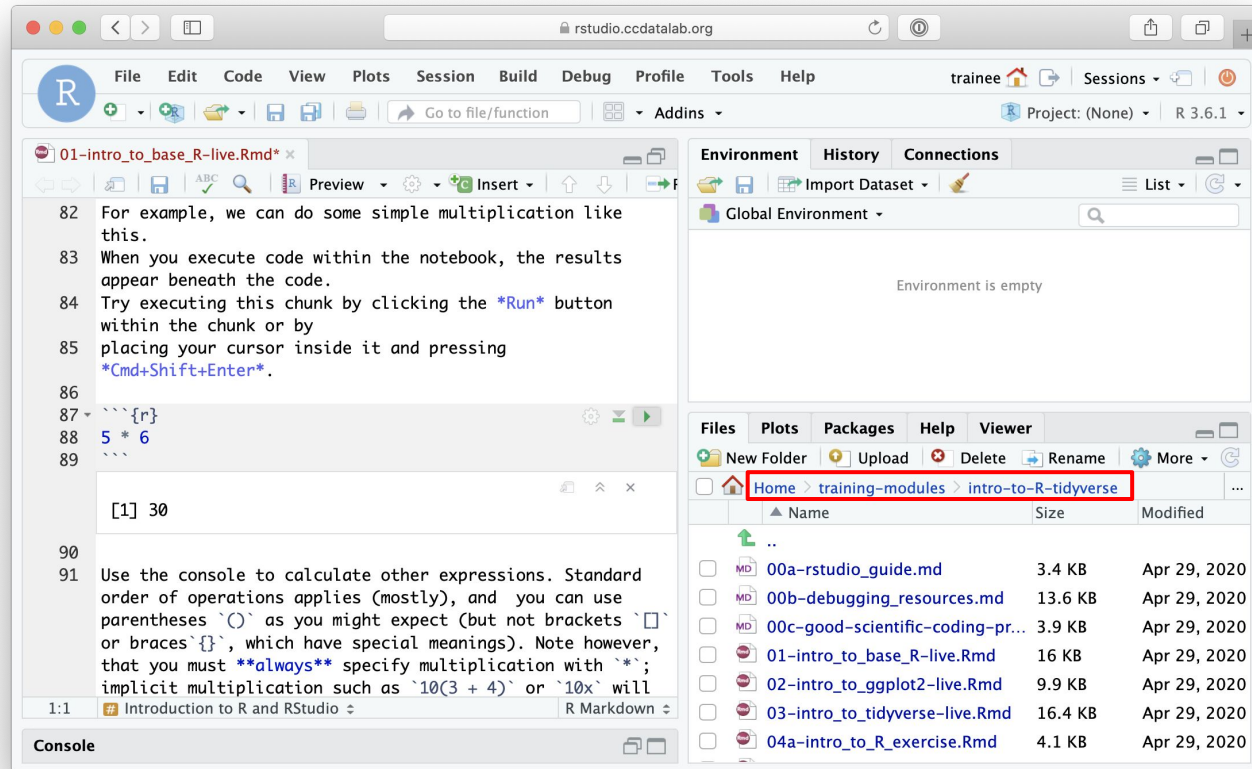
Below the code, a red-bordered box highlights the output: `[1] 30`. A blue arrow points from the text "Executable code chunk" to the code chunk area. A green arrow points from the text "Can click here to run a code chunk" to the green run button in the code chunk toolbar. A red arrow points from the text "Output from above code chunk" to the output box.

The RStudio interface includes a menu bar (File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help), a toolbar with various icons, and a sidebar on the right with tabs for Environment, History, and Connections. The Environment tab shows "Global Environment" with "Environment is empty". The Files tab shows a file explorer view of the current project.

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!





01-intro_to_base_R-live.Rmd*

```
82 For example, we can do some simple multiplication like
83 this.
84 When you execute code within the notebook, the results
85 appear beneath the code.
86 Try executing this chunk by clicking the *Run* button
87 within the chunk or by
88 placing your cursor inside it and pressing
89 *Cmd+Shift+Enter*.
```

```
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

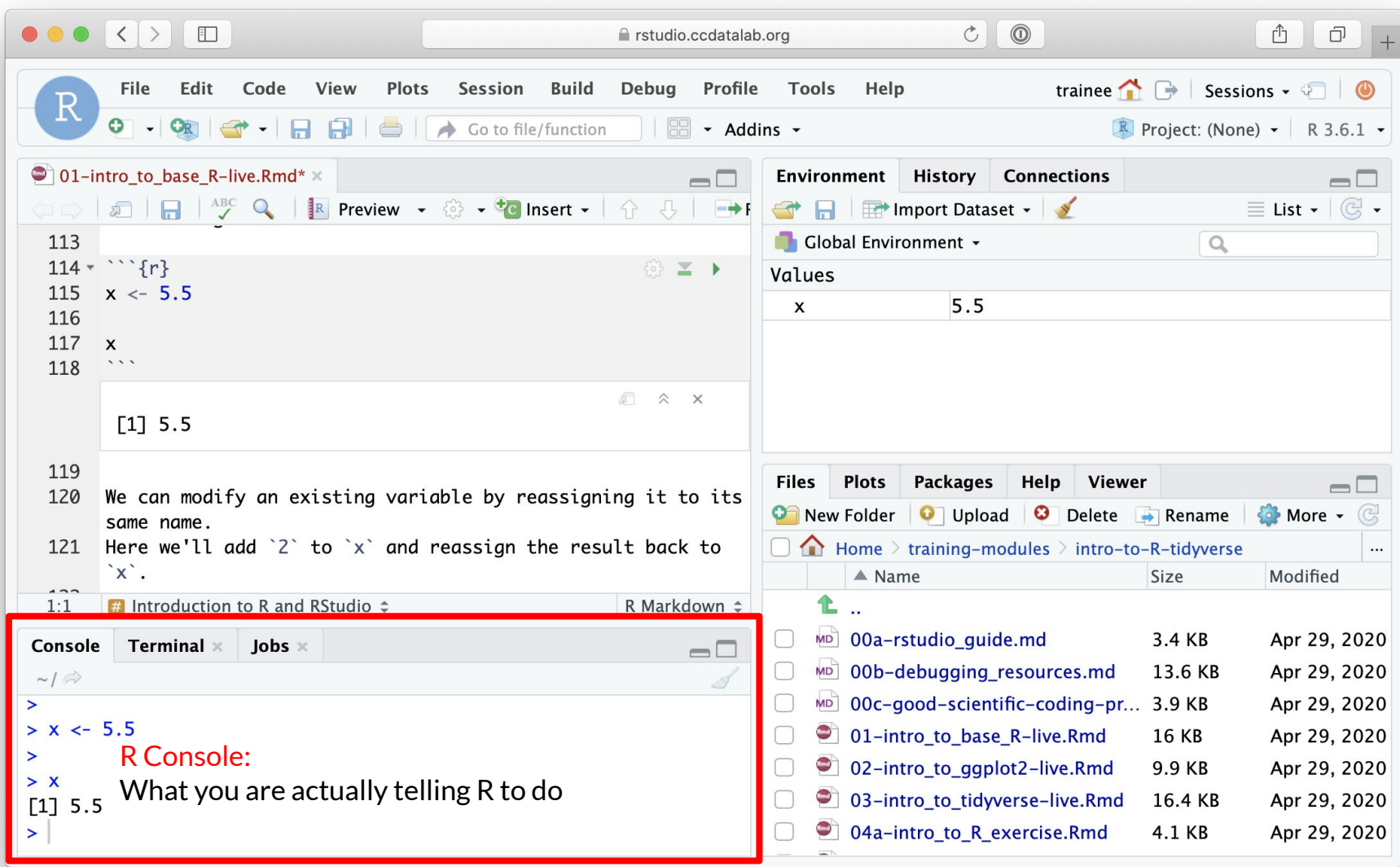
Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

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

	Name	Size	Modified
	..		
<input type="checkbox"/>	00a-rstudio_guide.md	3.4 KB	Apr 29, 2020
<input type="checkbox"/>		3.6 KB	Apr 29, 2020
<input type="checkbox"/>		9 KB	Apr 29, 2020
<input type="checkbox"/>		6 KB	Apr 29, 2020
<input type="checkbox"/>		9 KB	Apr 29, 2020
<input type="checkbox"/>	03-intro_to_tidyverse-live.Rmd	16.4 KB	Apr 29, 2020
<input type="checkbox"/>	04a-intro_to_R_exercise.Rmd	4.1 KB	Apr 29, 2020

Click here to show the Console



R Console:
What you are actually telling R to do



01-intro_to_base_R-live.Rmd*

Preview Insert

```
113
114 ` ``{r}
115 x <- 5.5
116
117 x
118 ` ``
```

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

```
[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`.
122
```

1:1 Introduction to R and RStudio

R Markdown

Console Terminal Jobs

```
>
> x <- 5.5
>
> x
[1] 5.5
> |
```

Environment History Connections

Import Dataset

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

The image shows the RStudio interface with the following components:

- Code Editor:** Contains R code for a live session:

```
113  
114 {r}  
115 x <- 5.5  
116  
117 x  
118  
119  
120 We can modify an existing variable by reassigning it to its  
121 same name.  
122 Here we'll add `2` to `x` and reassign the result back to  
123 `x`.
```
- Environment Pane:** Shows the current R environment with a variable `x` assigned the value `5.5`. It includes a search bar and a "List" button.
- Console:** Shows the execution of the code:

```
>  
> x <- 5.5  
>  
> x  
[1] 5.5  
> |
```
- Files Pane:** Displays a file browser for the directory `training-modules/intro-to-R-tidyverse`. It lists several files with their sizes and modification dates (all from April 29, 2020):

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

R environment:
What R knows and remembers for you

The image shows the RStudio interface with the following components:

- Code Editor:** Contains R code for assigning the value 5.5 to variable x. The output shows [1] 5.5.
- Environment Pane:** Shows the variable x with the value 5.5.
- Files Pane:** A red box highlights the Files pane, which shows a directory listing of files in the 'intro-to-R-tidyverse' folder.
- Console:** Shows the execution of the R code and the resulting output.

Other Assistance Tabs:
Things that help you in your coding

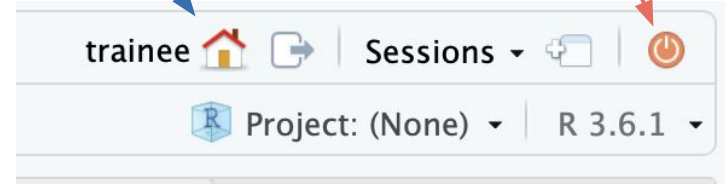
Name	Size	Modified
..		
00a-rstudio_guide.md	3.4 KB	Apr 29, 2020
00b-debugging_resources.md	13.6 KB	Apr 29, 2020
00c-good-scientific-coding-pr...	3.9 KB	Apr 29, 2020
01-intro_to_base_R-live.Rmd	16 KB	Apr 29, 2020
02-intro_to_ggplot2-live.Rmd	9.9 KB	Apr 29, 2020
03-intro_to_tidyverse-live.Rmd	16.4 KB	Apr 29, 2020
04a-intro_to_R_exercise.Rmd	4.1 KB	Apr 29, 2020

RStudio Sessions

- On the server, R is running many times at once
 - Each user has their own “**Session**” running, with its own memory and processes
 - It is possible for a user to have more than one session at a time
- We will usually want to start new sessions between notebooks to keep the environment clean

Go to the sessions page

End the current session



Session Page

