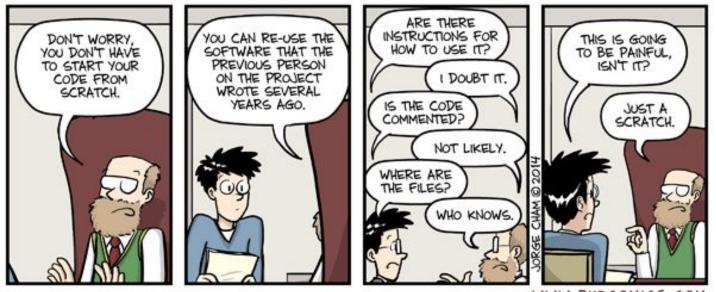
# Introduction to R, RStudio, and RStudio Server

The Data Lab Powered by Alex's Lemonade Stand Foundation

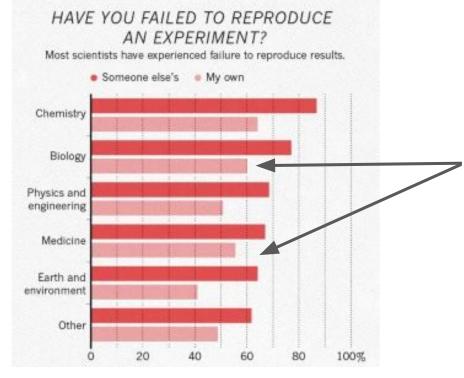
#### Who's been here before?



WWW. PHDCOMICS. COM

"Piled Higher and Deeper" by Jorge Cham www.phdcomics.com Used here with permission.

### 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

### 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

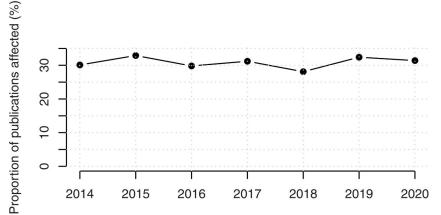
Despite geneticists being warned about spreadsheet problems, 30% of published papers contain mangled gene names in supplementary data.

#### Dyani Lewis

🖌 (f) 🗖

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 <u>autocorrect problems</u> were widespread, the academic literature is still littered with error-riddled spreadsheets, according to an

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



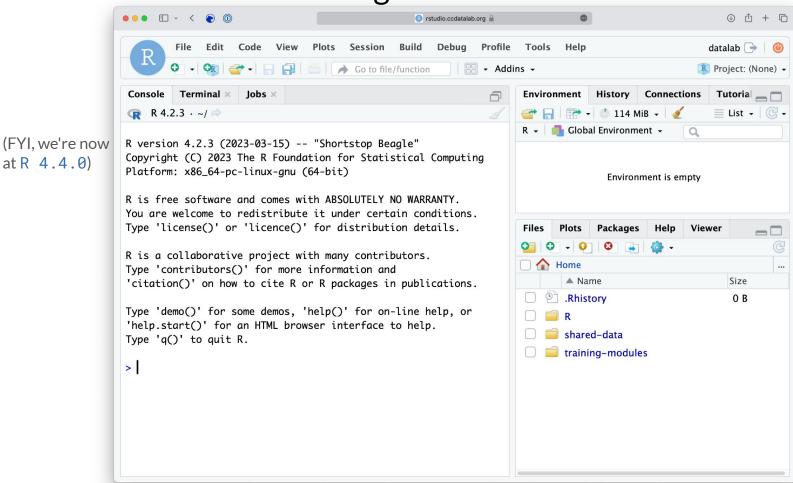
https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008984

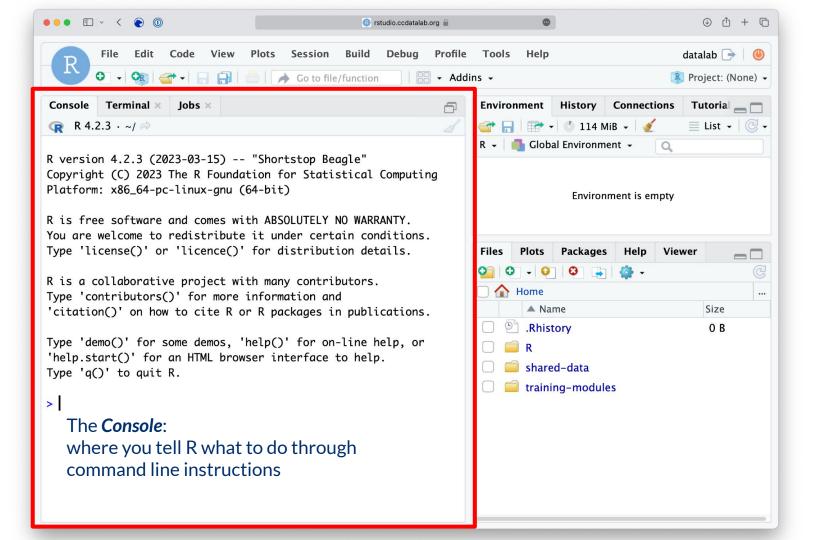
#### Command line vs GUI (graphics user interface)

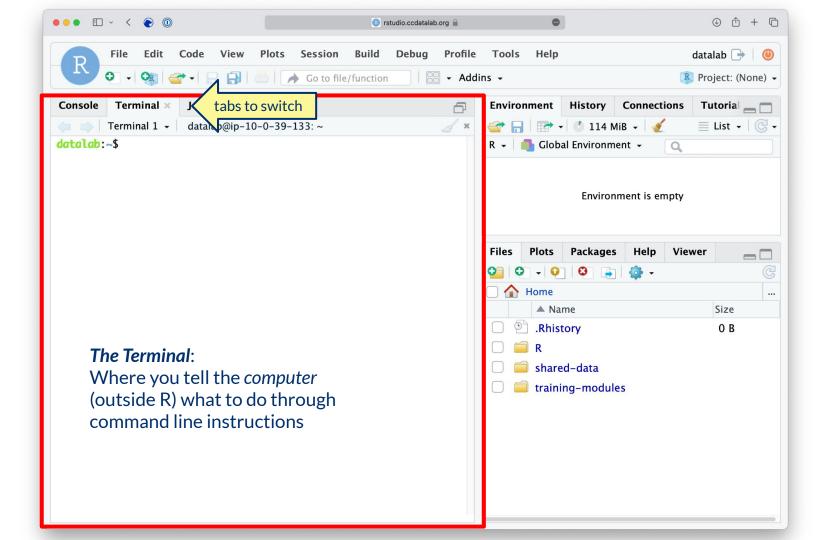
• An interface is how you interact with a program

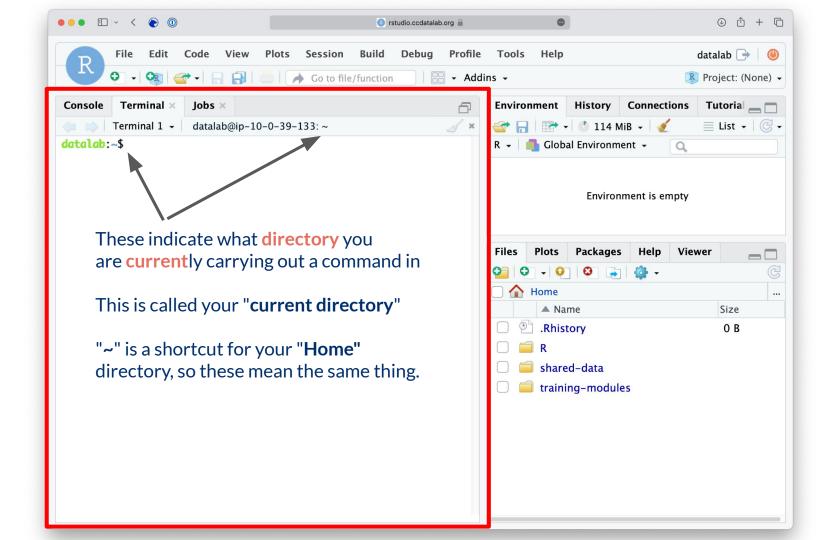
- GUI's have buttons you can *click* to do things, but...
- Command-line interfaces (CLI) have you *type* out things to do them

#### RStudio Server: A basic guide

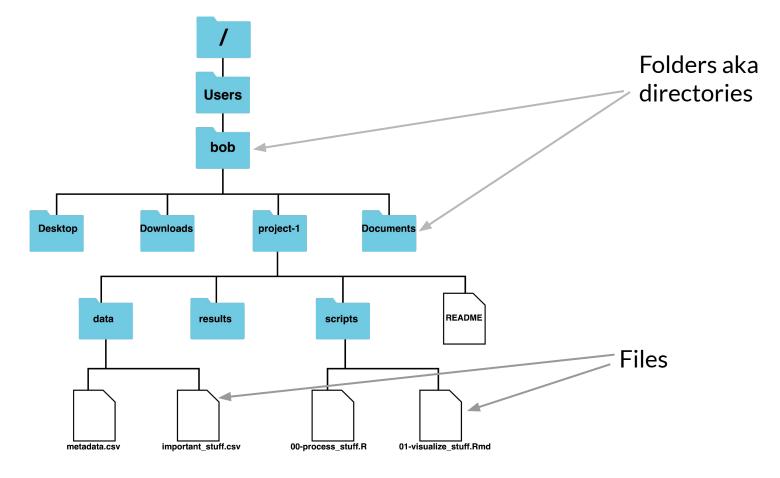




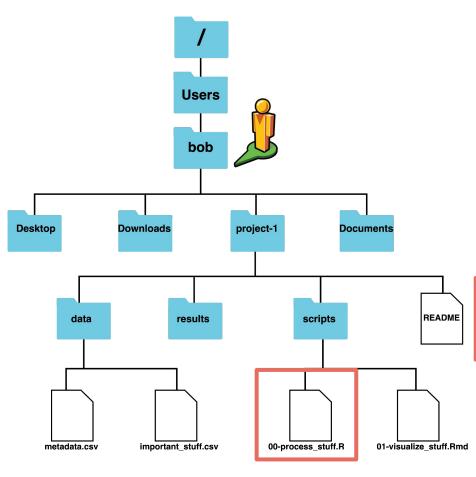




#### Example of a filesystem hierarchy



#### We are always working somewhere!



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

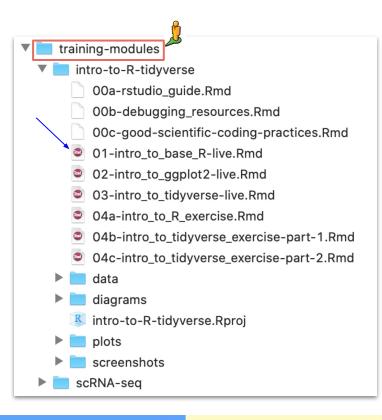
We therefore need to know the **paths** to files we are working with, relative to our working directory, to be able to use those files in our code.

The file we are working on

Relative path: project-1/scripts/00-process\_stuff.R

#### Let's look at our workshop files

Let's say we want access to 01-intro\_to\_base\_R-live.Rmd

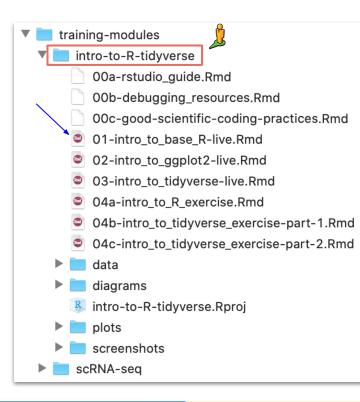


Current/working directory = training-modules

File path = intro-to-R-tidyverse/01-intro\_to\_base\_R-live.Rmd

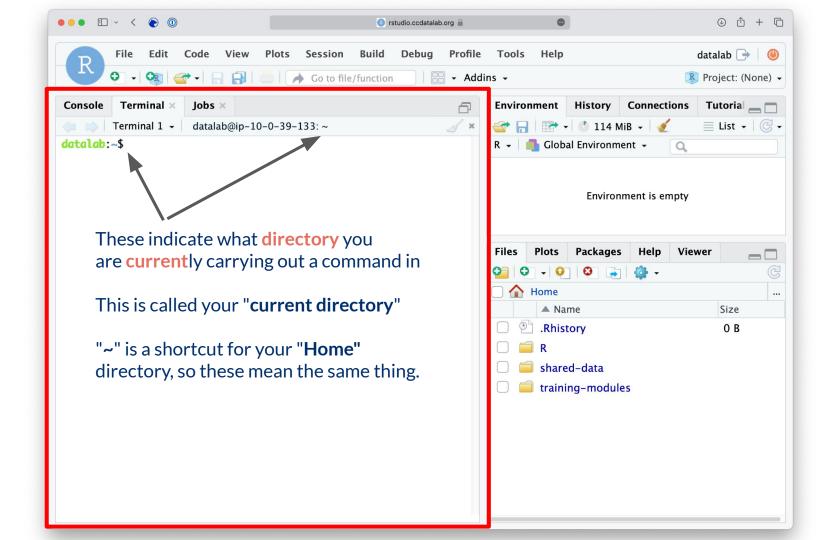
#### Relative paths depend on your working directory

Let's say we want access to 01-intro\_to\_base\_R-live.Rmd



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

File path = intro-to-R-tidyverse/01-intro\_to\_base\_R-live.Rmd



••• • • • • • •	🚯 rstudio.ccdatalab.org 🔒 💿	ů + c					
R File Edit Code View Plots Session	Build Debug Profile Tools Help	datalab 🕞 🛛 🕘					
Image: Control of the state of the sta							
Console Terminal × Jobs ×	Environment History Connections Tutorial						
🗼 🗼 Terminal 1 🗸 datalab@ip-10-0-39-133: ~ 🖉 💈	🐨 📊 🖙 Import Dataset 🗸 🔮 115 MiB 🖌 🔮	≣ List - 🛛 😋 -					
datalab:~\$ ls	R 🗸 📑 Global Environment 🗸 🚺	2					
R shared-data training-modules datalab:~\$ cd training-modules/							
actives of the charming modules	Environment is empty						
	Environment is empty						
Some common Terminal commands:							
Some common rennar commands.	Files Plots Packages Help Viewer						
ls - list the files and folders in a	O · O O ⊕ 🔅 ·	C					
directory (files that start with a '.'	□ <b>☆</b> Home						
are not shown by default)	▲ Name Size	Modified					
ed change divertaries	Image: OrganizationImage: Organization </td <td>Jan 31, 2023</td>	Jan 31, 2023					
<mark>cd</mark> - change directories	🗆 🧰 R						
	🗌 🛑 shared-data						
	🗌 🛑 training-modules						

	🚯 rstudio.ccdatalab.org 🔒	•		ů + C
File Edit Code View Plots Session	Build Debug	Profile Tools H	lelp	datalab 🕞 🕘
😯 🗣 📲 🚍 🚽 📄 📑 🖌 🖓 Go to file/f	unction	- Addins -		Project: (None) 🗸
Console Terminal × Jobs ×	Environment	History Connecti	ions Tutorial	
Terminal 1 - datalab@ip-10-0-39-133: ~/trair	🚰 🔒 🖙 Im	port Dataset 👻 🔮	115 MiB 🖌 💉	≣ List • 🛛 🕃 •
datalab:~\$ ls R shared-data training-modules	R 🖌 🛑 Global	l Environment 👻		٩
<pre>datalab:~/training-modules\$ ls LICENSE.md README.md intro-to-R-tidyverse module-cheatsheets scRNA-seq datalab:~/training-modules\$</pre>	Environment is empty Files Plots Packages Help Viewer			
The words in front of our cursor have		<b>O O O O O O O O O O</b>		e e e e e e e e e e e e e e e e e e e
changed because we are now "in" the	Home			
training-modules directory	A Nam	ıe	Size	Modified
	🗌 🕑 .Rhisto	ory	0 B	Jan 31, 2023
	🗆 🧰 R			
	🗌 🧰 shared	l-data		
	🗌 🧰 training-modules			

	s rstudio.ccdatalab.org					
File Edit Code View Plots Session	Build Debug Profile Tools Help datalab 🕞 🅚					
R       Image: Solution in the solutin the solutin the solution in the solutin the solution in the so						
Console Terminal × Jobs ×	Environment History Connections Tutorial					
🗼 🗼 Terminal 1 🗸 datalab@ip-10-0-39-133: ~/trair	🚰 🕞 🖙 Import Dataset 🗸 🖉 115 MiB 🗸 🔮 List 🖌 🎯 🗸					
datalab:~\$ ls R shared-data training-modules	R 🗸 📑 Global Environment 🗸 🔍					
<pre>datalab:~\$ cd training-modules/ datalab:~/training-modules\$ ls LICENSE.md README.md intro-to-R-tidyverse module-cheatsheets scRNA-seq datalab:~/training-modules\$</pre>	Environment is empty					
	Files Plots Packages Help Viewer					
The words in front of our cursor have						
changed because we are now "in" the	□ <b>1</b> Home					
training-modules directory	▲ Name Size Modified					
	Image: Provide state         O B         Jan 31, 2023,					
	🗌 🧰 shared-data					
	🗌 🧰 training-modules					
	The files tab over here does NOT reflect your current directory or any changes within it					

## 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.

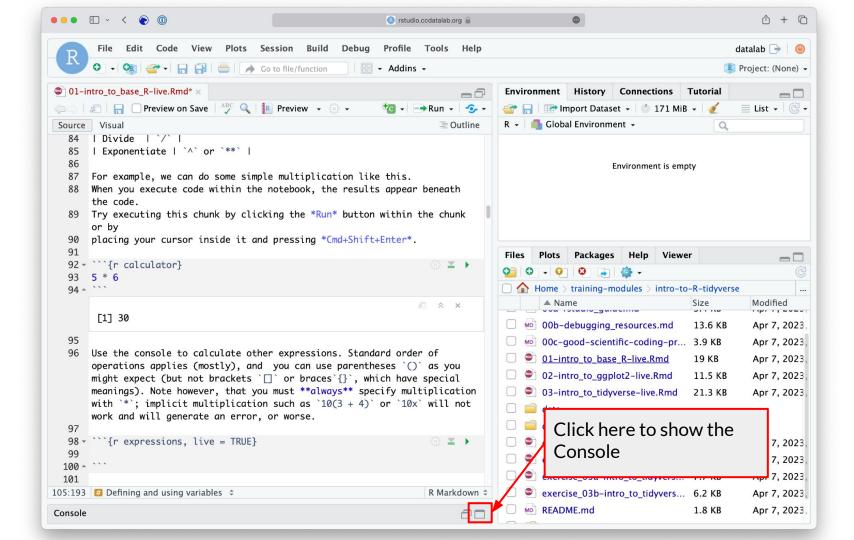


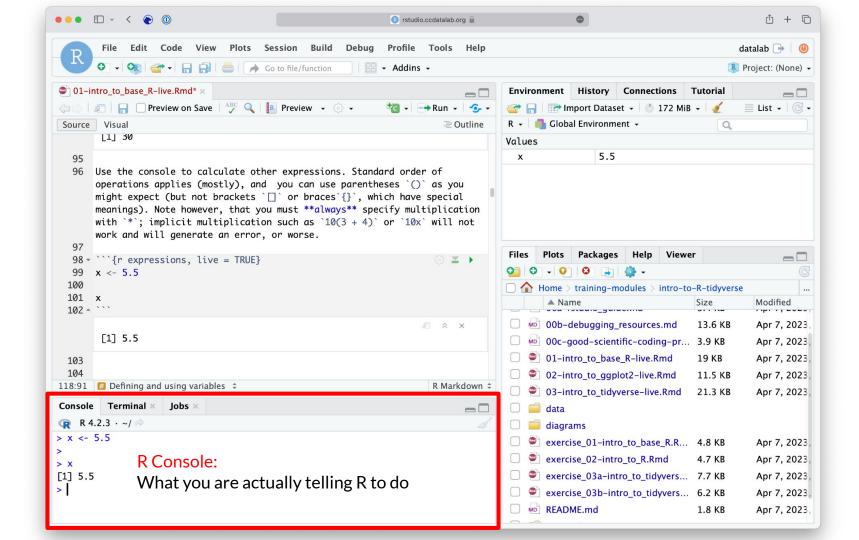
**RStudio** is an IDE for working in R

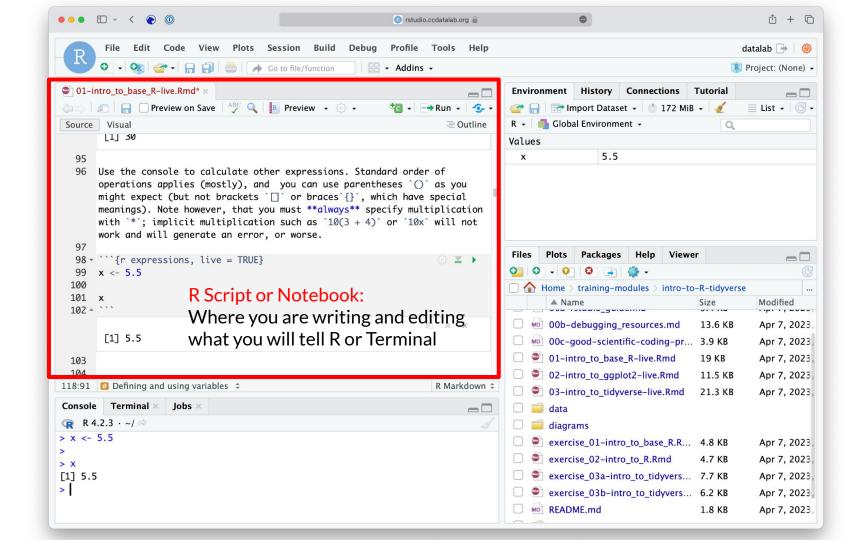
- IDE: Integrated Development Environment
- We write R code using the (free!) RStudio IDE

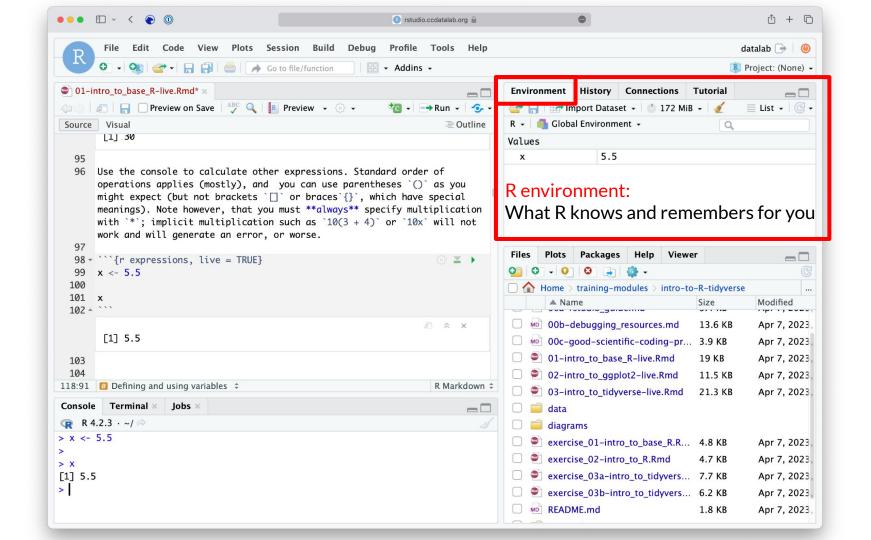


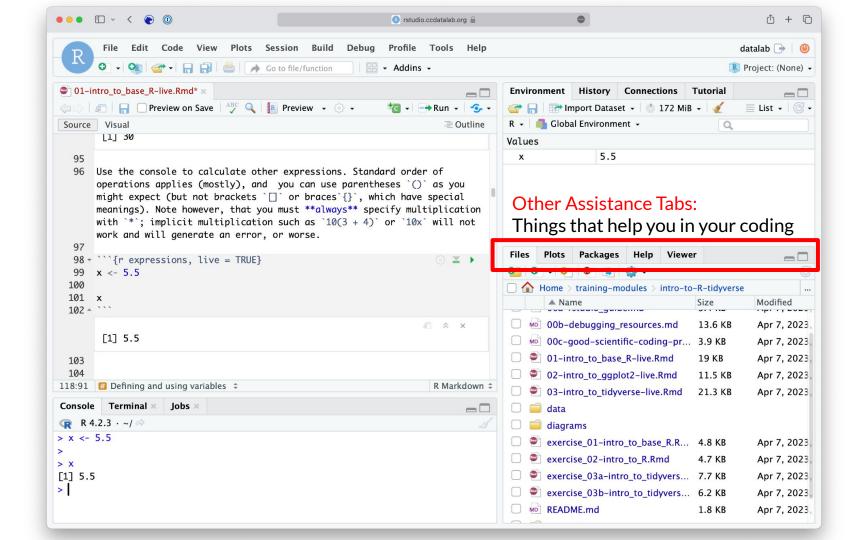
RStudio Server allows us to run the RStudio IDE from a browser











#### **R** Notebooks

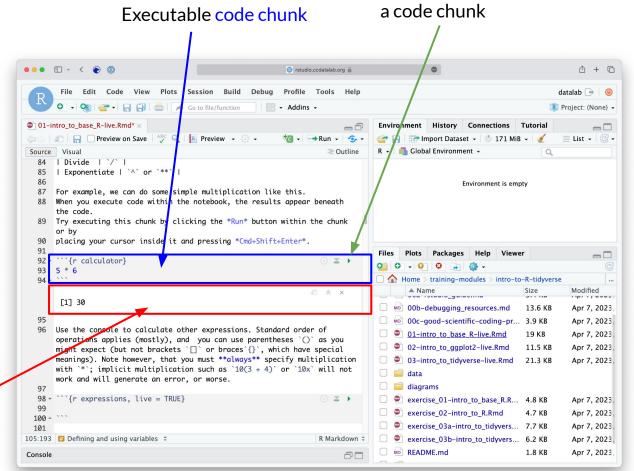
Use the "Files" tab to open: training-modules/intro-to-R-tidyverse/01-intro\_to\_base\_R-live.Rmd

••• • • • • • • • • • • • • • • • • • •	rstudio.ccdatalab.org 🔒	۲	ů + ©
File Edit Code View Plots Session Build Debug Prof	file Tools Help		datalab 🕞 🛛 🎱
🔨 🔹 🔹 🖓 🧁 🚽 🔒 🔒 🍌 Go to file/function	dins 🗸		Project: (None) 👻
01-intro_to_base_R-live.Rmd ×	-6	Environment History Connections Tutorial	
(==) = Preview on Save MC Q R Preview - O - C	• 🛶 Run • 🧐 •	😅 🕞 🖙 Import Dataset 🗸 🔮 170 MiB 🗸 💰	📃 List 🖌 🖂 -
Source Visual	≡ Outline	R 🗸 💼 Global Environment 🗸 🔍 🔍	
1			
2 title: "Introduction to R and RStudio" 3 author: Originally authored by Stephanie J. Spielman, a for ALSF	adapted by CCDL	Environment is empty	
4 date: 2021			
5 output:			
<pre>6 html_notebook:</pre>			
7 toc: true			
<pre>8 toc_float: true</pre>			
9 *		Files Plots Packages Help Viewer	
10		O - O O - O O - 🙆 -	
11 - ## Objectives		☐	se
12		▲ Name Size	Modified
13 This notebook will demonstrate how to: 14		1	mounieu
15 - Navigate the RStudio environment			
16 - Use R for simple calculations, both mathematical and loa	ui col	00a-rstudio_guide.md 3.4 KB	Apr 7, 2023
17 - Define and use variables in base R	,	00b-debugging_resources.md 13.6 KB	Apr 7, 2023,
18 - Understand and apply base R functions		00c-good-scientific-coding-pr 3.9 KB	Apr 7, 2023.
19 - Understand, define, and use R data types, including vect	or	01-intro_to_base_R-live.Rmd 19 KB	Apr 7, 2023,
manipulation and indexing		02-intro_to_ggplot2-live.Rmd 11.5 KB	Apr 7, 2023
20 - Understand the anatomy of a data frame			
21 22		□ • 03-intro_to_tidyverse-live.Rmd 21.3 KB	Apr 7, 2023,
22 *		🗌 💭 🧰 data	
23 24 - #### *More resources for learning R*		🗌 🧰 diagrams	
25		exercise_01-intro_to_base_R.R 4.8 KB	Apr 7, 2023
26 - [Swirl, an interactive tutorial](https://swirlstats.com/	()		
1:1 Introduction to R and RStudio \$	R Markdown ‡	exercise_02-intro_to_R.Rmd 4.7 KB	Apr 7, 2023,
Consult.		exercise_03a-intro_to_tidyvers 7.7 KB	Apr 7, 2023,
Console	60	exercise 03b-intro to tidyyers 6.2 KB	Apr 7 2023

#### **R** Notebooks

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

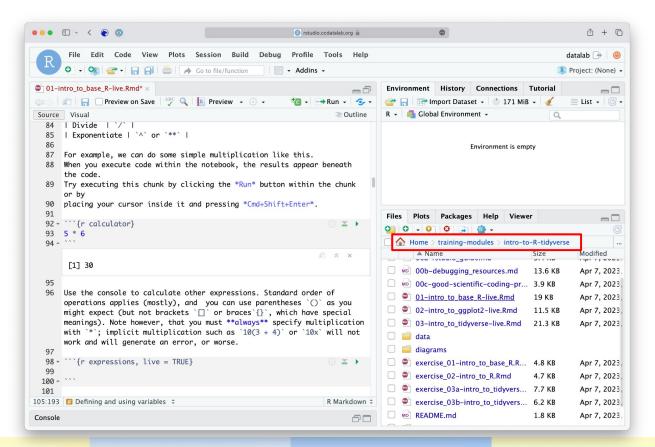
Output from above code chunk



Can click here to run

#### **R** Notebooks

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



#### **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

