

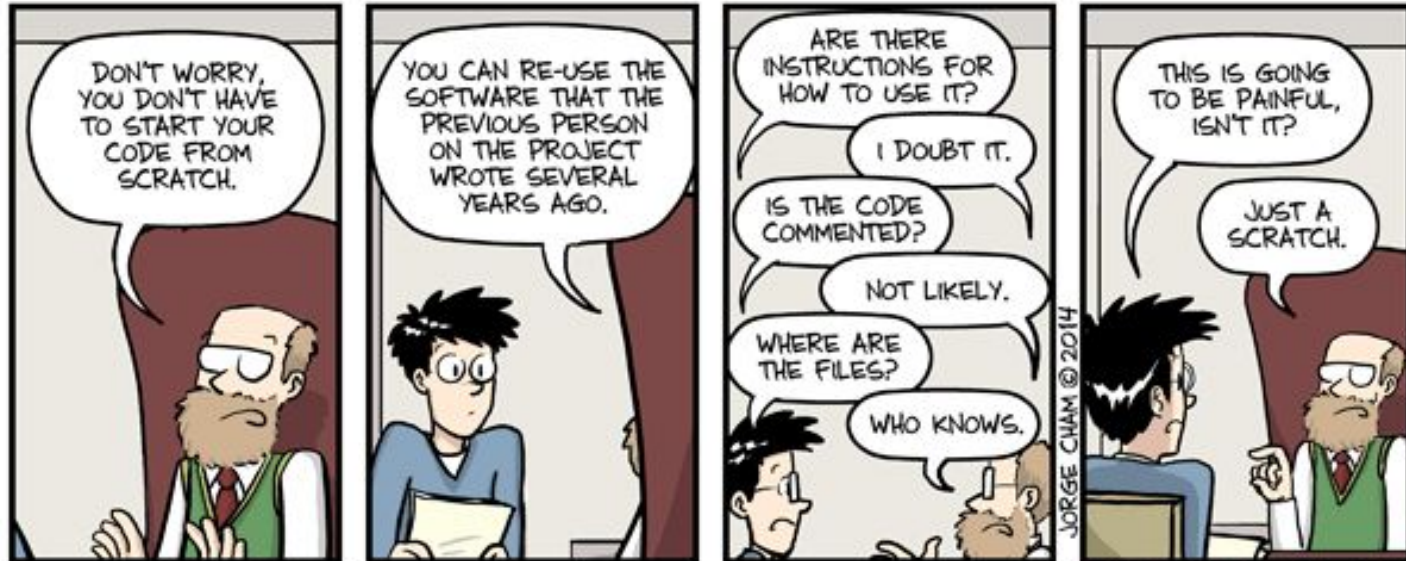


Introduction to R, RStudio, and RStudio Server

The Data Lab

Powered by Alex's Lemonade Stand Foundation

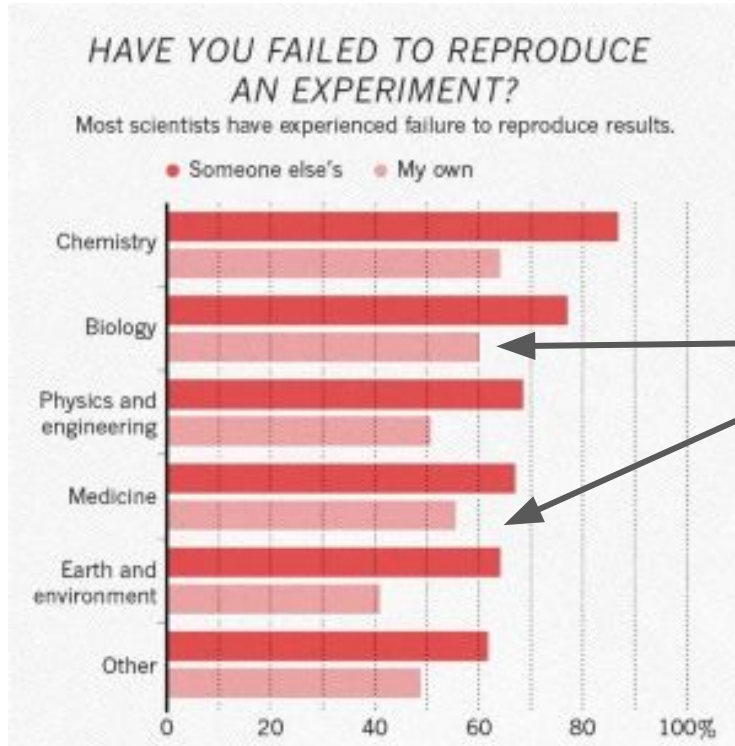
Who's been here before?



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 their own 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

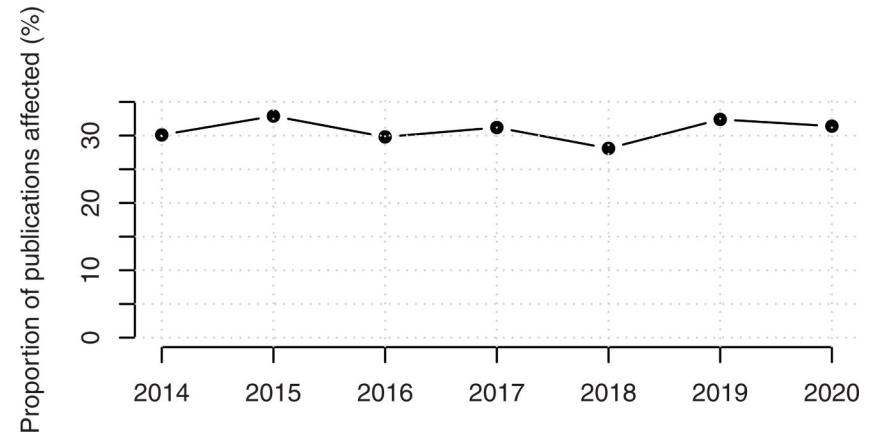
Despite geneticists being warned about spreadsheet problems, 30% of published papers contain mangled gene names in supplementary data.

[Dyani Lewis](#)



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>

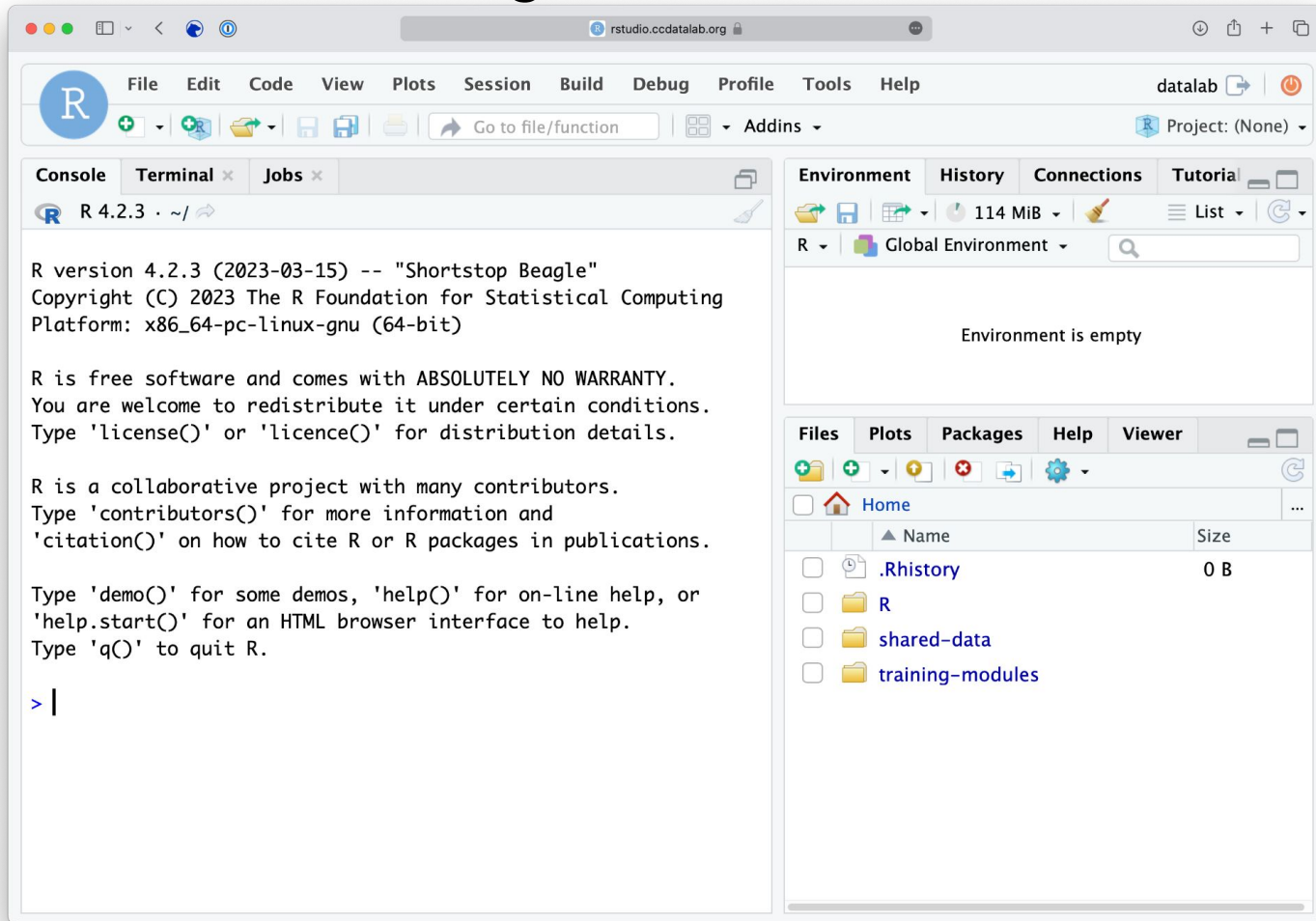


<https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008984>

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. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The console pane shows the R version 4.2.3 (2023-03-15) and the license information. The environment pane shows 'Global Environment' and 'Environment is empty'. The file pane shows the home directory with files like '.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.

```
> |
```

(FYI, we're now
at R 4.4.0)

The screenshot shows the RStudio interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The top right 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 explorer view with a table of files and folders.

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.

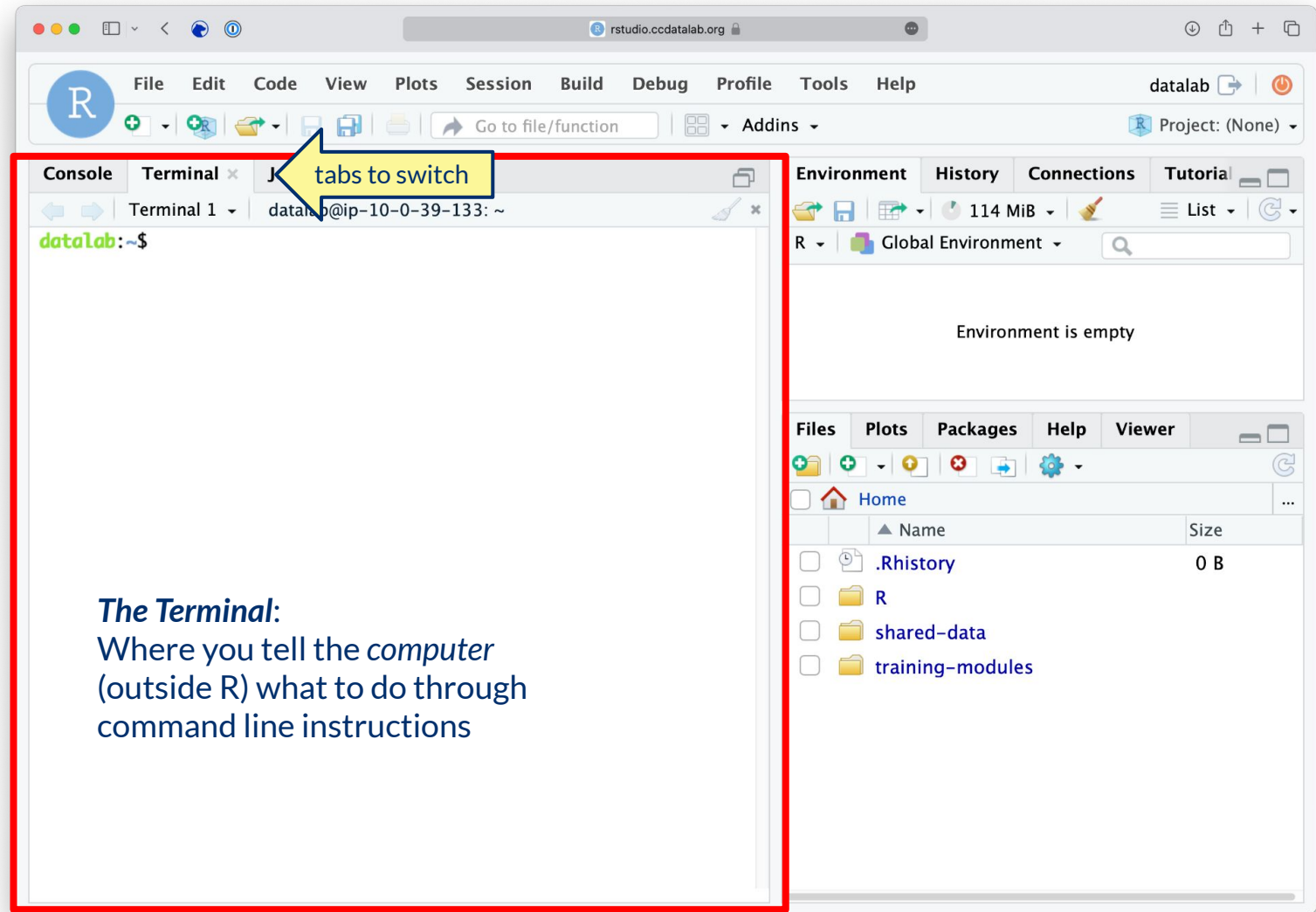
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

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

	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

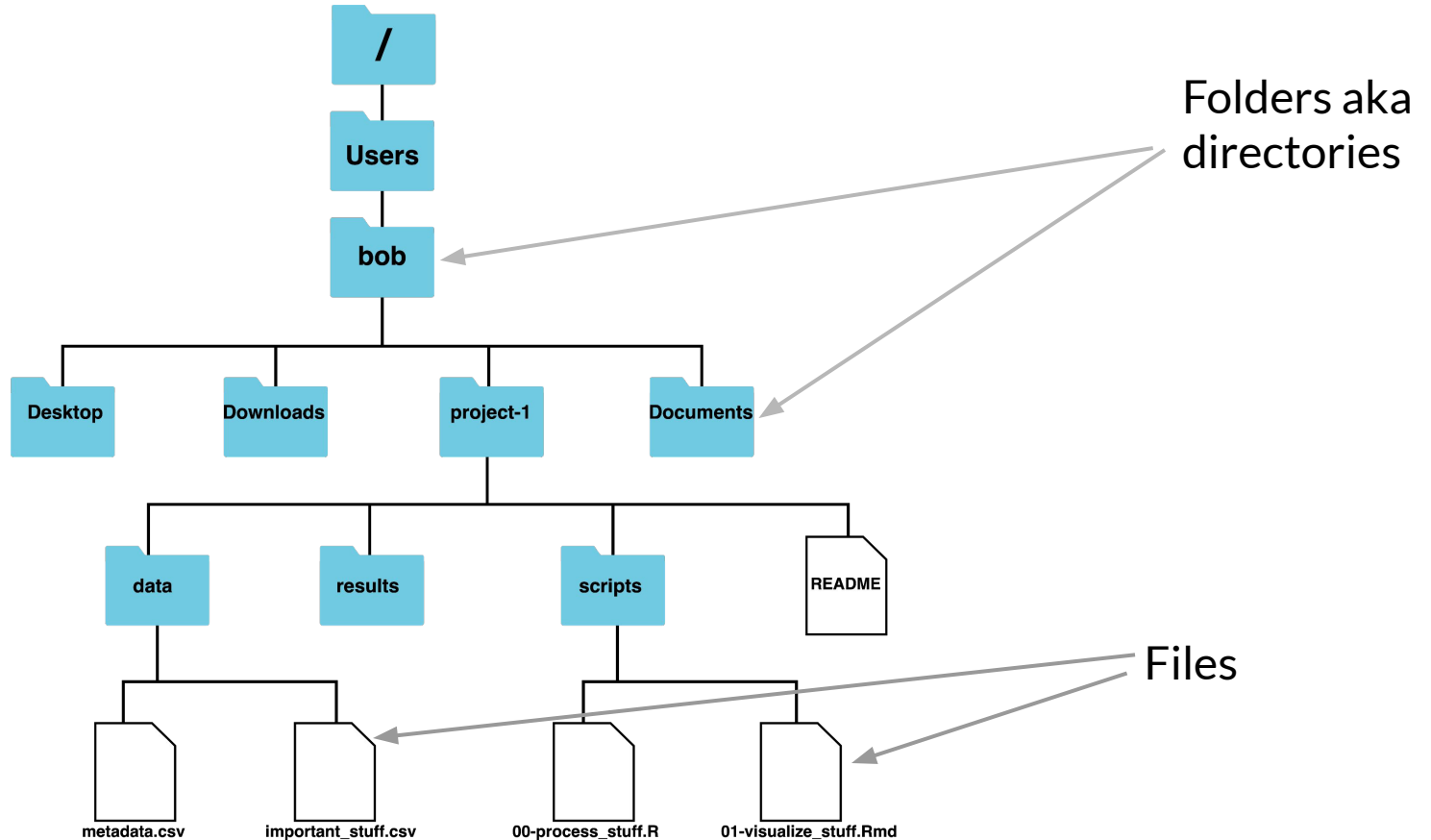
These indicate what **directory** you are **currently** carrying out a command in

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

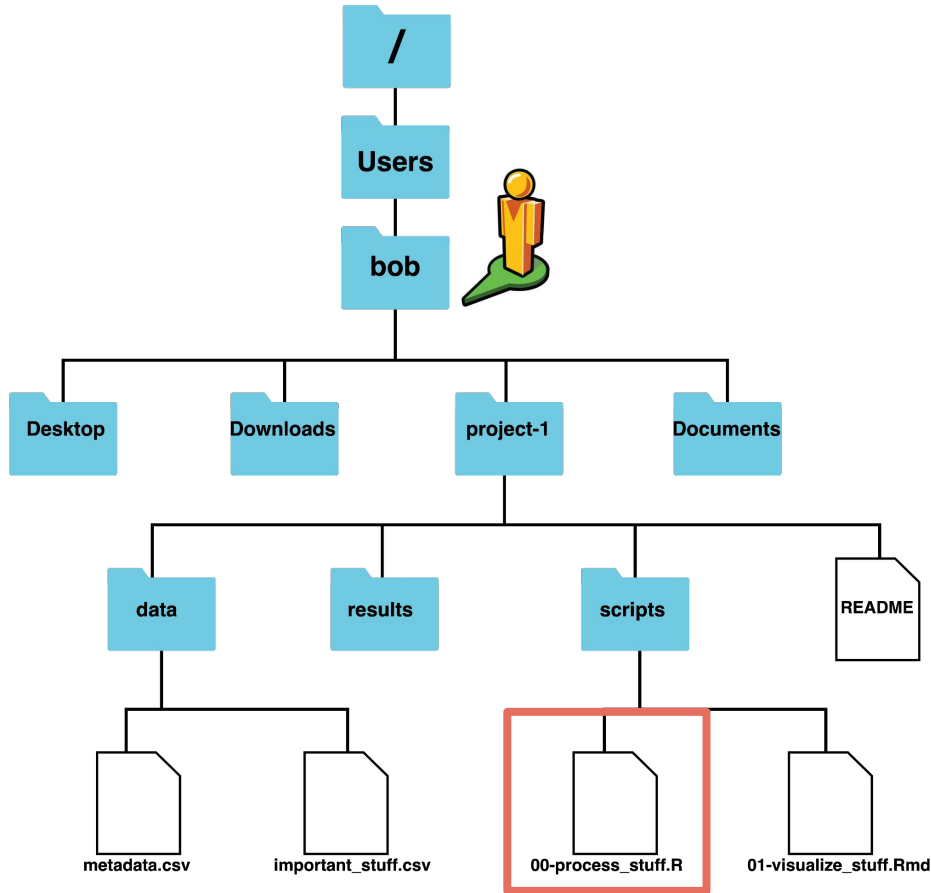
"~" is a shortcut for your "**Home**" directory, so these mean the same thing.

The screenshot shows the RStudio interface with the terminal window highlighted in red. The terminal prompt is `dataLab:~$`. An arrow points from the text to the tilde character (~) in the prompt. The background shows the RStudio environment with a file browser on the right displaying the home directory contents: `.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*

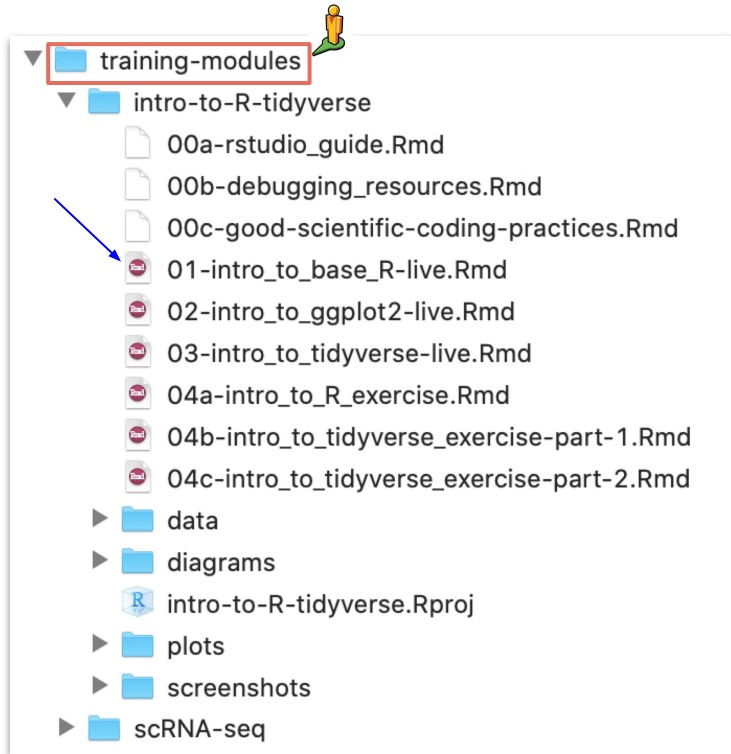
We therefore need to know the **paths** to files we are working with, relative to our working directory, to be able to use those files in our code.

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`

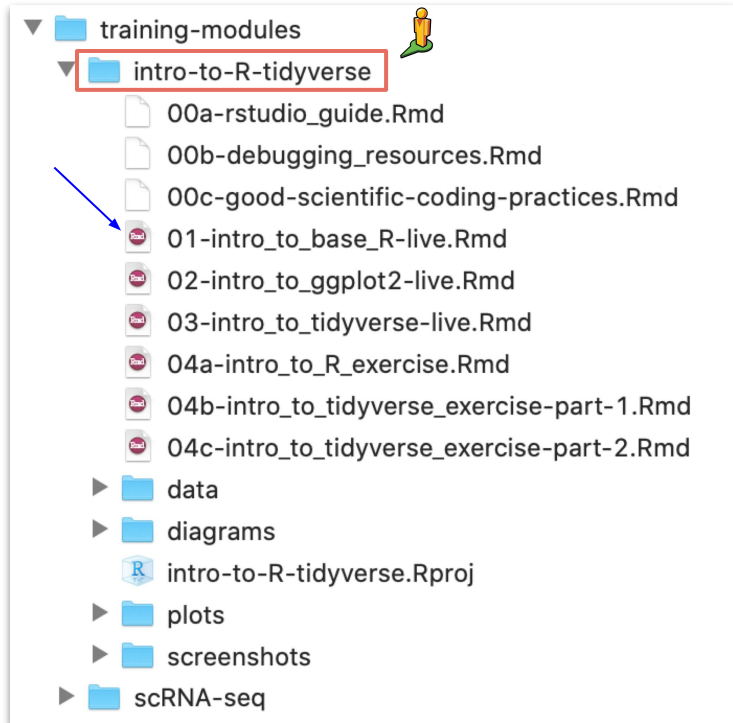


Current/working directory = `training-modules`

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 = `01-intro_to_base_R-live.Rmd`

dataLab:~\$

Environment is empty

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

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

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

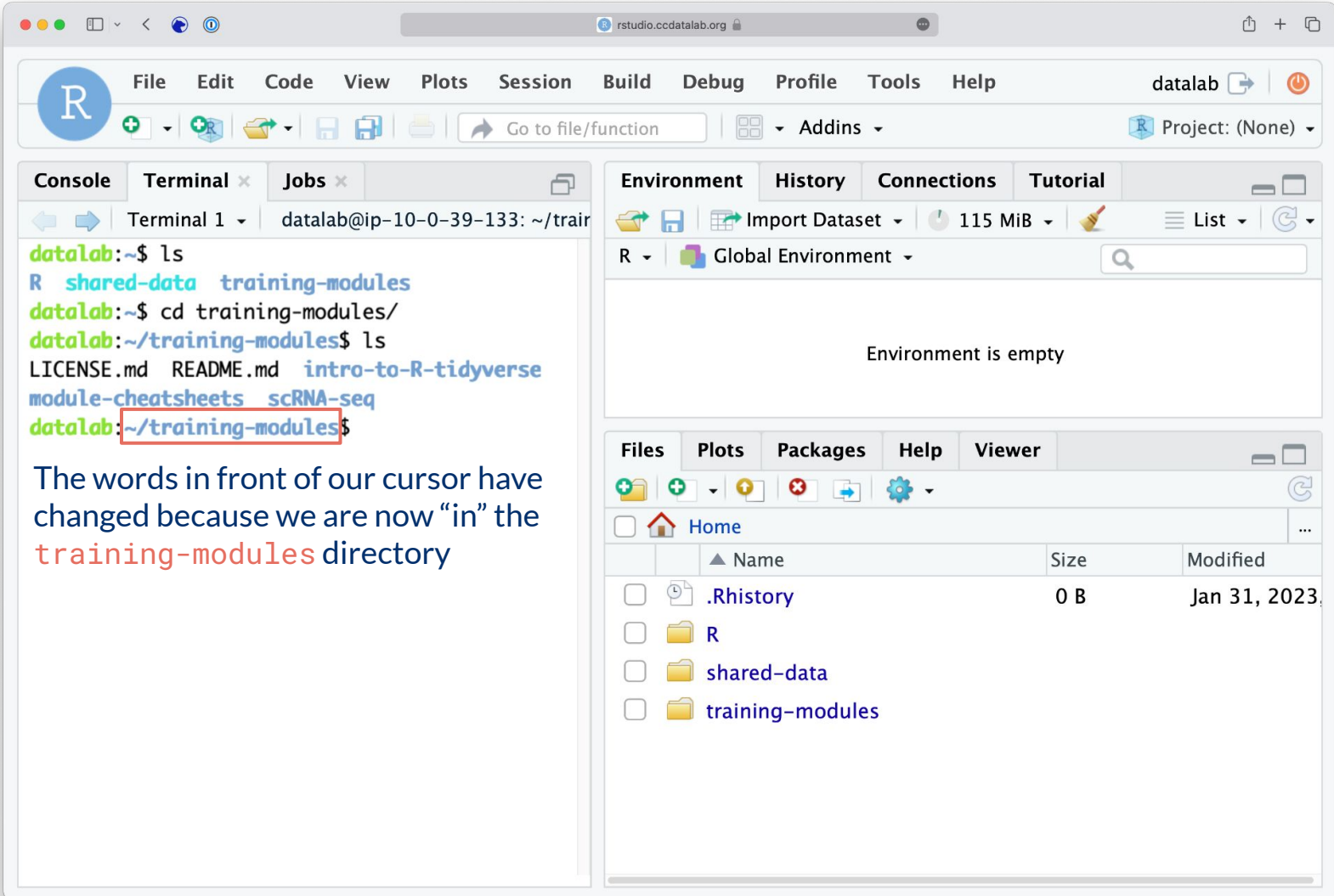
"~" is a shortcut for your "**Home**" directory, so these mean the same thing.

The image shows a screenshot of 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:** Shows terminal output:

```
dataLab:~$ ls
R shared-data training-modules
dataLab:~$ cd training-modules/
```
- Environment:** Shows the current environment is empty.
- Files:** Shows a file browser view of the Home directory with 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 Pane:** Shows a shell prompt where the user has navigated to the `training-modules` directory. The command `ls` has been executed, listing files: `LICENSE.md`, `README.md`, `intro-to-R-tidyverse`, `module-cheatsheets`, and `scrna-seq`. The current directory path `~/training-modules` is highlighted with a red box.
- Environment Pane:** Shows the current environment is empty, with the text "Environment is empty".
- Files Pane:** Shows the file explorer for the `Home` directory. The files listed are `.Rhistory`, `R`, `shared-data`, and `training-modules`. The `training-modules` directory is highlighted.

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



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

The screenshot shows the RStudio interface with a notebook titled "01-intro_to_base_R-live.Rmd". The code in the notebook includes instructions on how to run code chunks and a code chunk that calculates $5 * 6$. The console output shows the result $[1] 30$. The Environment pane is empty, and the Files pane shows a directory listing of files in the "intro-to-R-tidyverse" folder. A red box highlights the "Console" icon in the bottom right corner of the RStudio interface, with a red arrow pointing to it and a text box that says "Click here to show the Console".

```
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 ```
97
98 [1] 30
99
100 Use the console to calculate other expressions. Standard order of
101 operations applies (mostly), and you can use parentheses `()` as you
102 might expect (but not brackets `[]` or braces `{}`, which have special
103 meanings). Note however, that you must **always** specify multiplication
104 with `*`; implicit multiplication such as `10(3 + 4)` or `10x` will not
105 work and will generate an error, or worse.
106
107 ```{r expressions, live = TRUE}
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
```

The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains R code for defining a variable `x` with the value 5.5. The code is as follows:

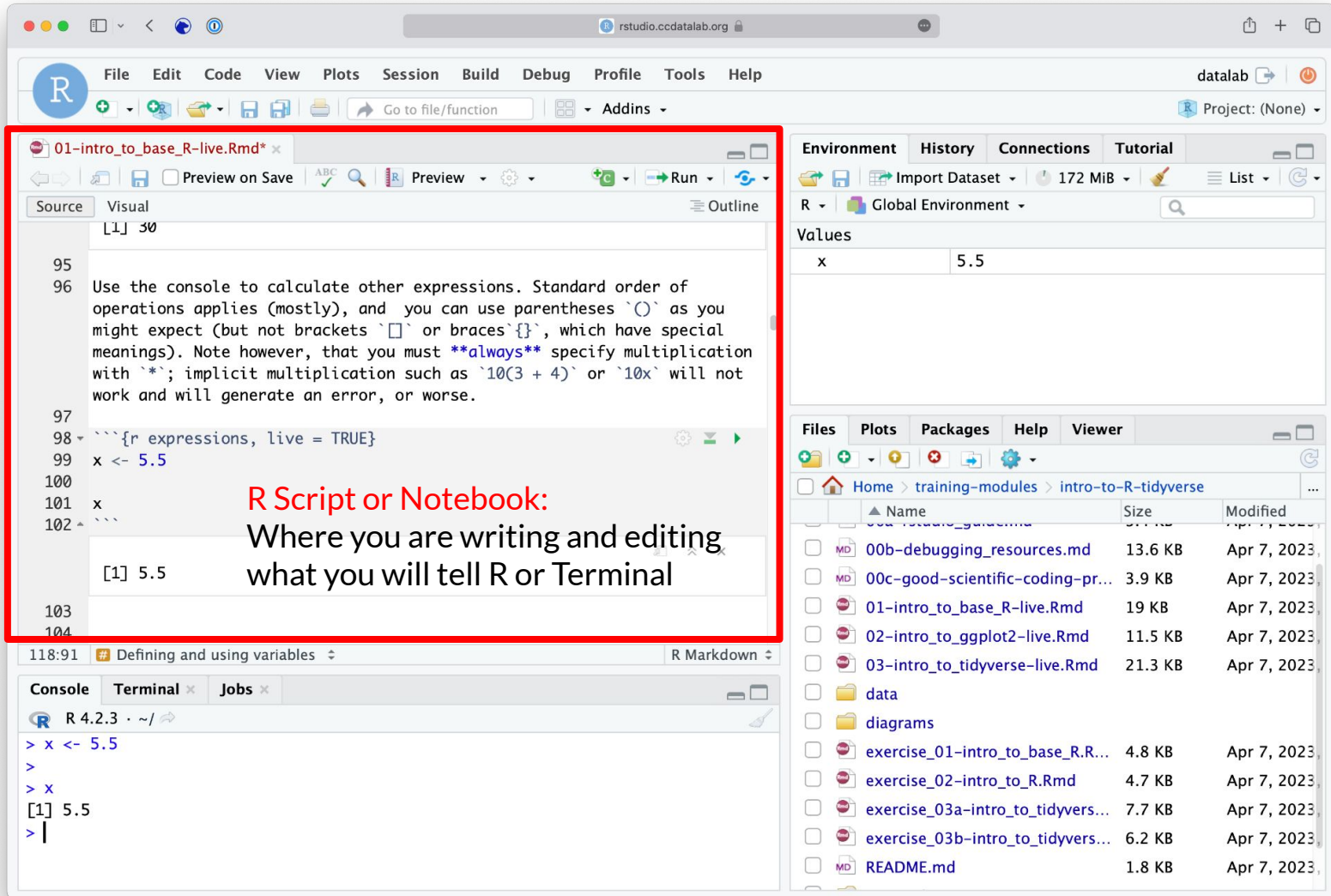
```
95  
96 Use the console to calculate other expressions. Standard order of  
operations applies (mostly), and you can use parentheses `()` as you  
might expect (but not brackets `[]` or braces `{}`, which have special  
meanings). Note however, that you must **always** specify multiplication  
with `*`; 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 ```
```
- Environment Pane:** Shows the variable `x` with the value 5.5.

Variable	Value
x	5.5
- Files Pane:** Shows a file explorer view of the current project directory, listing various R Markdown files and folders.

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
- Console:** Shows the execution of the R code from the source editor. The output is:

```
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

```
01-intro_to_base_R-live.Rmd* x
[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 4.2.3 ~ /
> x <- 5.5
>
> x
[1] 5.5
> |
```

Environment History Connections Tutorial

Import Dataset 172 MiB

R Global Environment

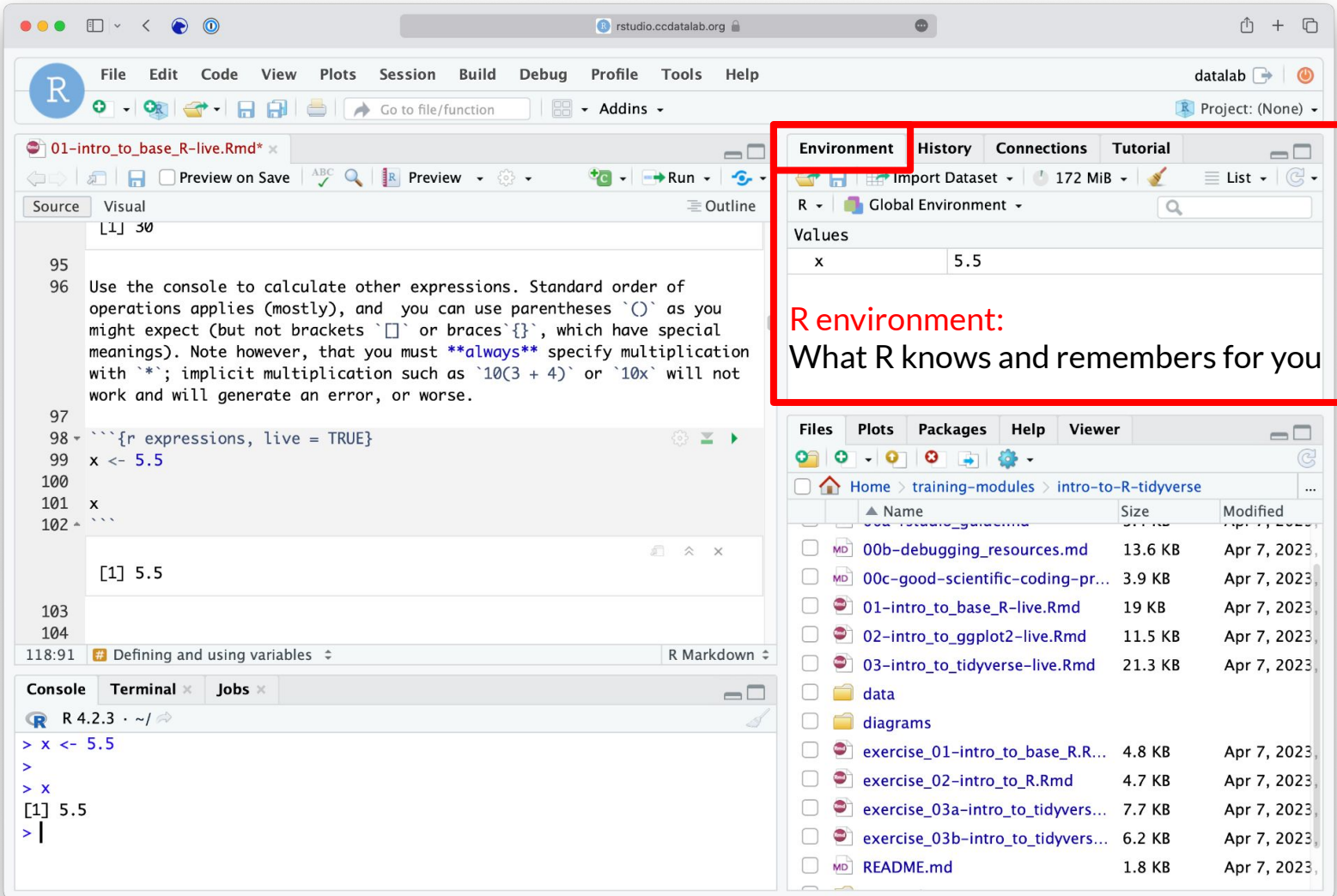
Values

x	5.5
---	-----

Files Plots Packages Help Viewer

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

Name	Size	Modified
00a-tidyverse_guidance		
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



```
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 ```
```

Console Terminal Jobs

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

Environment History Connections Tutorial

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-intro_to_base_R-live.Rmd	4.8 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_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 screenshot displays the RStudio interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The toolbar contains icons for saving, previewing, and running code. The main editor window shows an R script with the following content:

```
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 console at the bottom shows the execution of the code:

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

The Environment pane on the right shows the Global Environment with a variable x containing the value 5.5.

Overlaid on the right side of the interface is a red-bordered box containing the text:

Other Assistance Tabs:
Things that help you in your coding

The tabs highlighted in the red box are Files, Plots, Packages, Help, and Viewer.

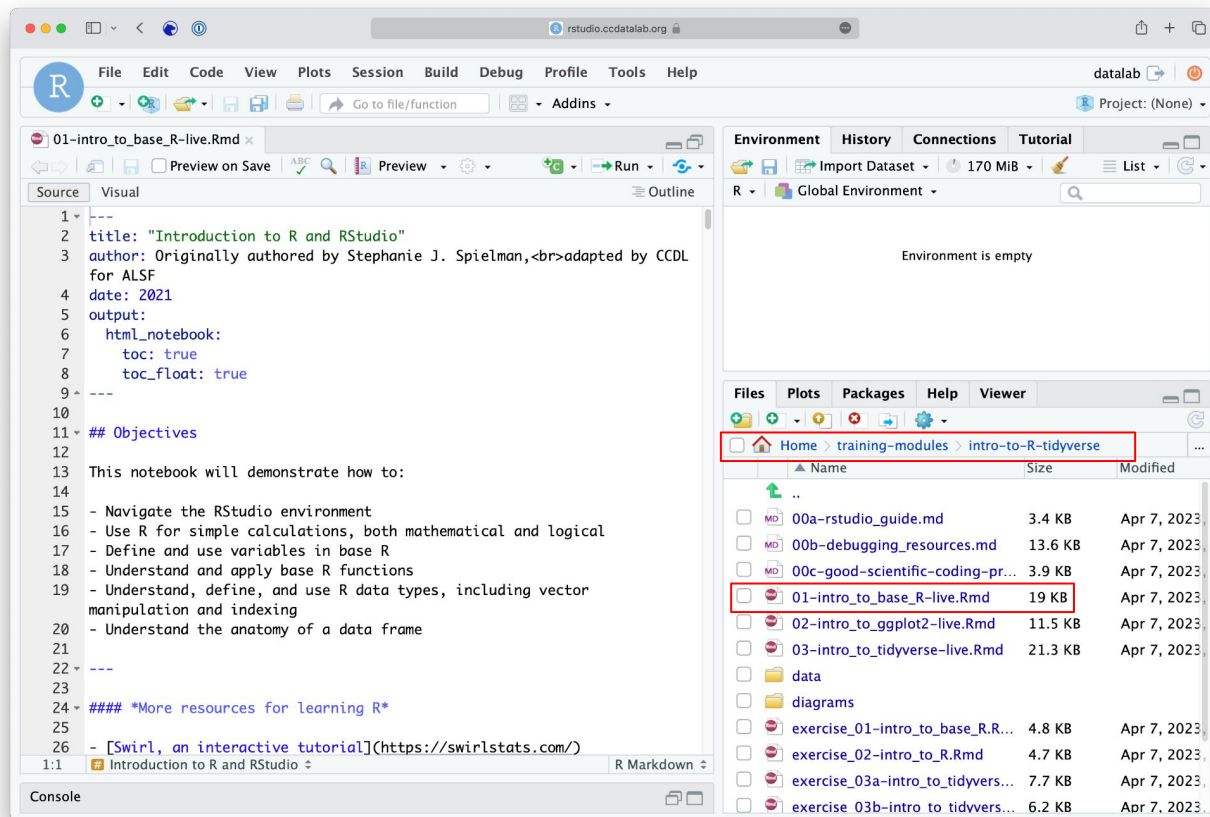
The Files pane on the right shows a file browser view of the directory structure:

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

Name	Size	Modified
00a-intro_to_tidyverse-live.Rmd	19.0 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_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

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 R Notebook content, which includes a title, author information, 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.

```
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/)
```

The 'Files' pane shows the following directory structure:

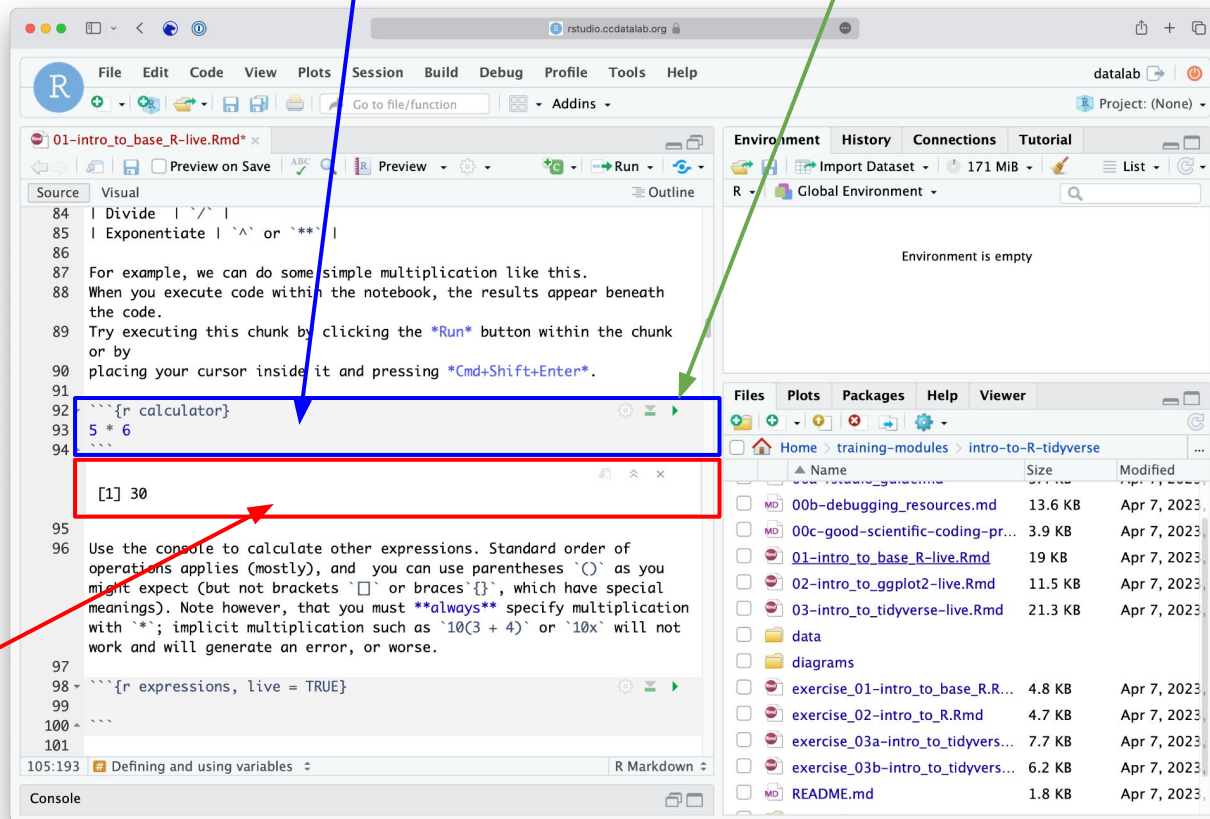
- Home > training-modules > intro-to-R-tidyverse
- ..
- 00a-rstudio_guide.md (3.4 KB, Apr 7, 2023)
- 00b-debugging_resources.md (13.6 KB, Apr 7, 2023)
- 00c-good-scientific-coding-pr... (3.9 KB, Apr 7, 2023)
- 01-intro_to_base_R-live.Rmd (19 KB, Apr 7, 2023)**
- 02-intro_to_ggplot2-live.Rmd (11.5 KB, 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

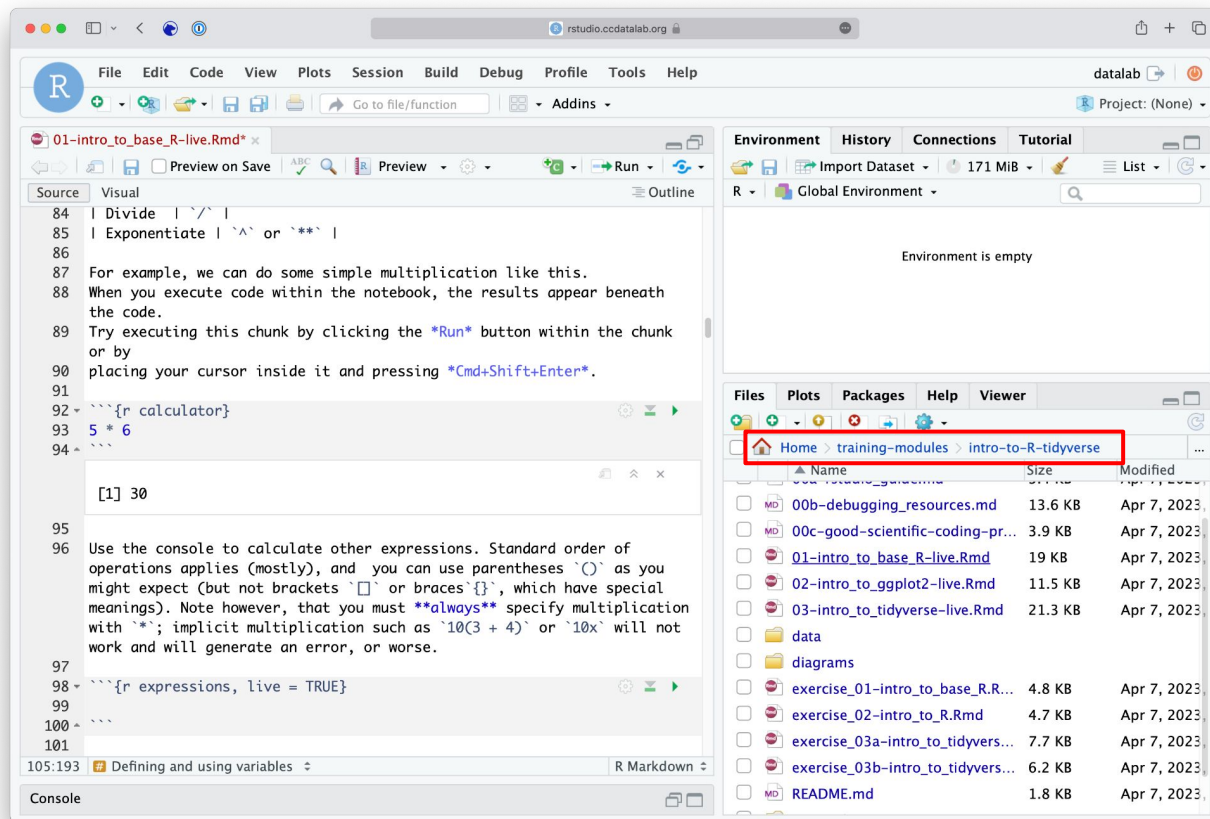
Can click here to run a code chunk



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!



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