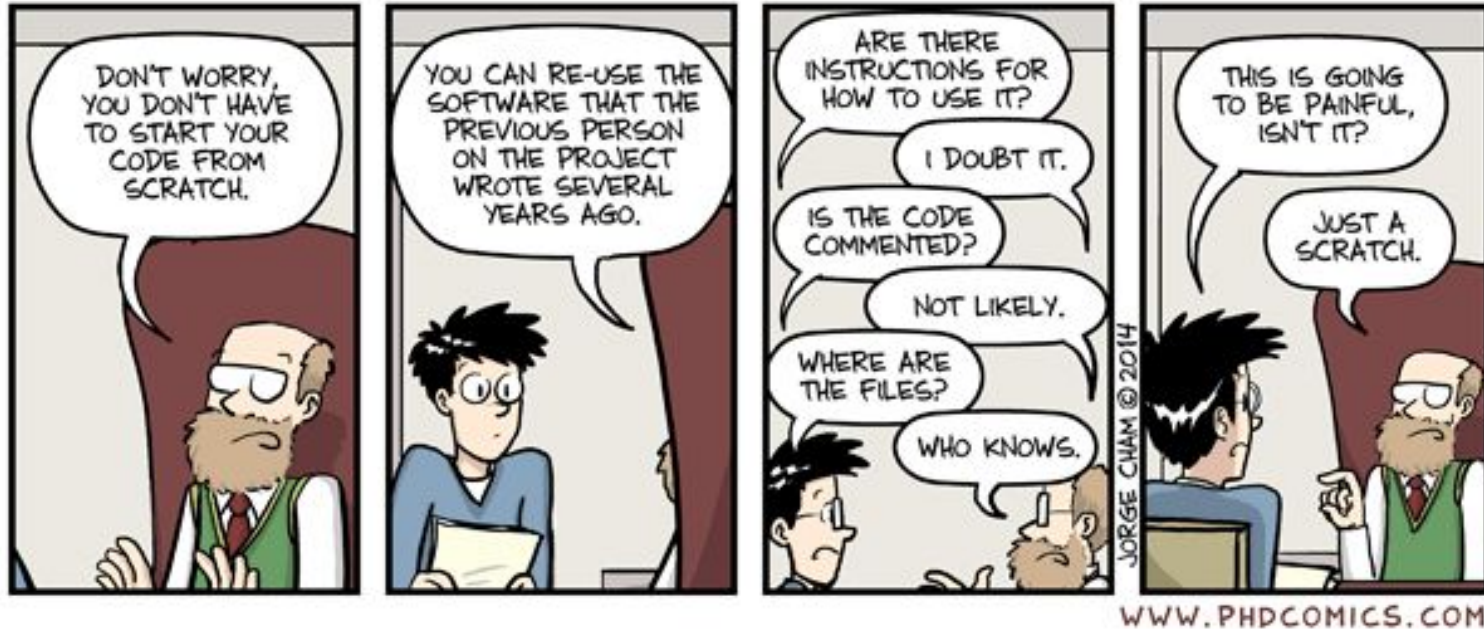




Introduction to R, RStudio, and RStudio Server

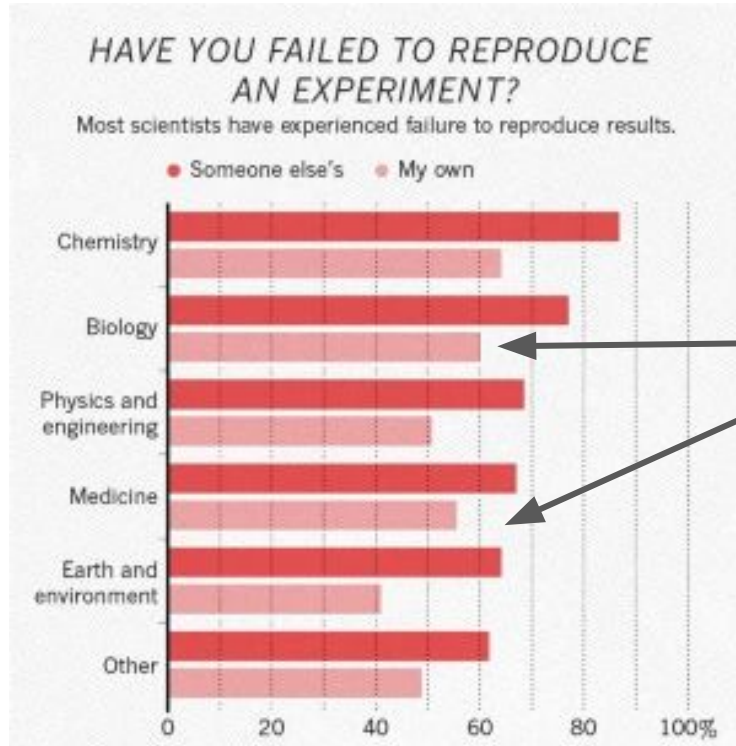
The CCDL

Who's been here before?



"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.

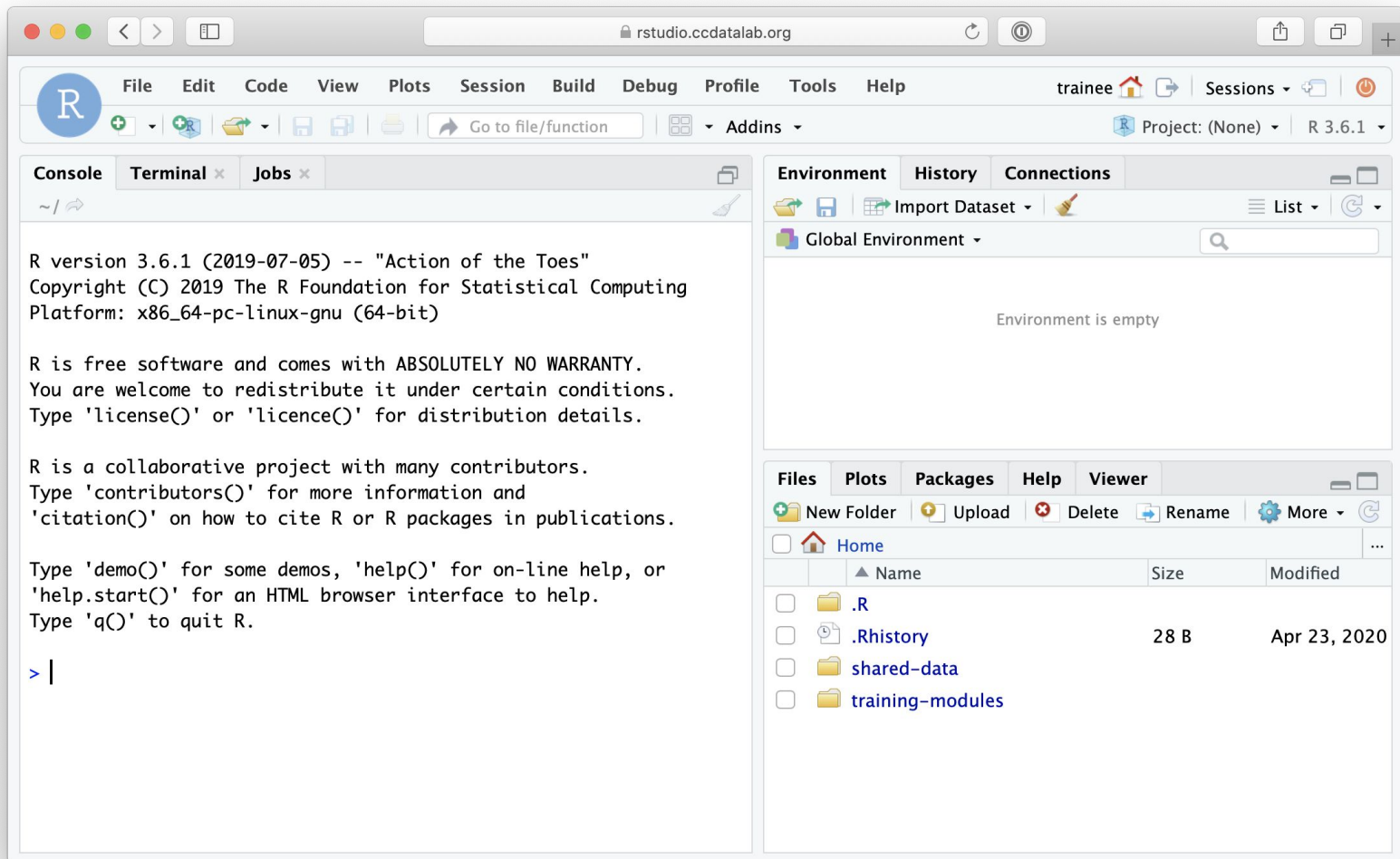
Baker, M. 1,500 scientists lift the lid on reproducibility.
Nature 533, 452–454 (2016).
<https://doi.org/10.1038/533452a>

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



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

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 screenshot shows the RStudio application window. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The top toolbar contains icons for file operations and a search bar. The main workspace is divided into four panes: Console, Terminal, Environment, and Files. The Terminal pane is highlighted with a red border and contains the text `trainee:~$`. A yellow arrow points to the tab bar above the Terminal pane, with the text "tabs to switch" inside it. The Environment pane shows "Global Environment" and "Environment is empty". The Files pane shows a list of files and folders: `.R`, `.Rhistory`, `shared-data`, and `training-modules`.

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

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

The screenshot shows the RStudio interface with the following components:

- Top Bar:** Includes the R logo, menu items (File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help), and session information (trainee, Sessions, Project: (None), R 3.6.1).
- Console/Terminal/Jobs:** The Terminal tab is active, showing a prompt `trainee:~$` and the current directory `/home/may2020/trainee`.
- Environment/History/Connections:** The Environment tab is active, showing "Global Environment" and "Environment is empty".
- Files/Plots/Packages/Help/Viewer:** The Files tab is active, showing a file explorer with the following contents:

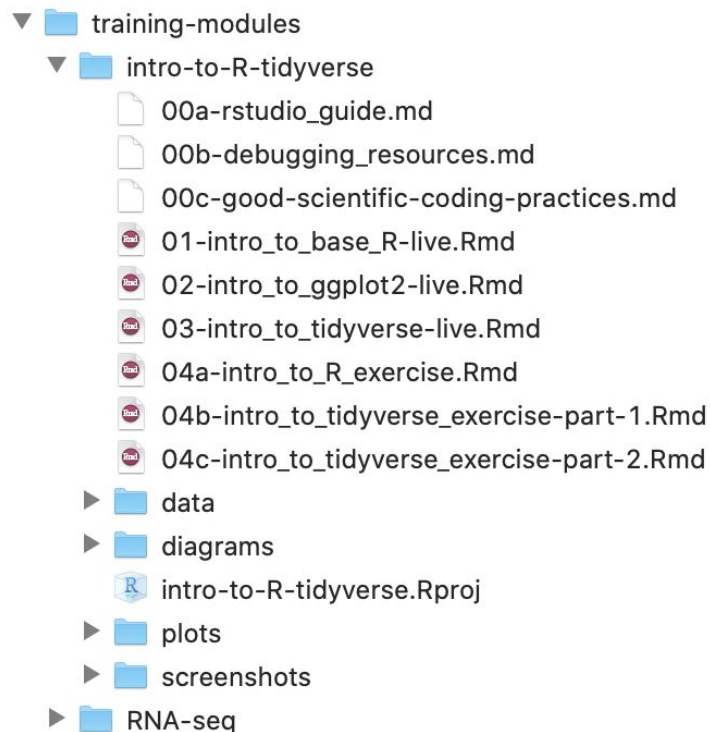
	Name	Size	Modified
<input type="checkbox"/>	.R		
<input type="checkbox"/>	.Rhistory	28 B	Apr 23, 2020
<input type="checkbox"/>	shared-data		
<input type="checkbox"/>	training-modules		

Annotations in the terminal window:

- Two arrows point from the text "These indicate what **directory** you are **currently** carrying out a command in" to the `~` and `$` in the prompt `trainee:~$`.
- The text "This is called your **'current directory'**" is positioned below the first annotation.
- The text "`~` is a shortcut for your **'Home'** directory, so these mean the same thing." is positioned below the second annotation.

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 screenshot shows the RStudio IDE interface. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The top toolbar contains icons for file operations and a search bar. The main workspace is divided into four panels: Console, Terminal, Jobs, and Environment. The Console panel shows the output of the `ls` command in the `training-modules` directory. The Terminal panel shows the execution of the `ls` and `cd` commands. The Environment panel shows the current environment is empty. The Files panel shows the file structure of the project, including the `shared-data` and `training-modules` directories.

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

The screenshot shows the RStudio interface with the following components:

- Top Bar:** Includes the R logo, menu items (File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help), and session information (trainee, Sessions, Project: (None), R 3.6.1).
- Console/Terminal:** The 'Terminal' tab is active, showing the command prompt at `/home/may2020/trainee/training-modules`. The terminal output shows the execution of `ls` and `cd training-modules`, resulting in the current directory being `~/training-modules`. The prompt now shows the contents of the directory: `shared-data training-modules RNA-seq intro-to-R-tidyverse`.
- Environment:** The 'Environment' tab is active, showing an empty environment.
- Files:** The 'Files' tab is active, showing a file explorer view of the `Home` directory. The file list includes `.R`, `.Rhistory`, `shared-data`, and `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

<

>

rstudio.ccdatalab.org

R

File

Edit

Code

View

Plots

Session

Build

Debug

Profile

Tools

Help

trainee

Sessions

Go to file/function

Addins

Project: (None)

R 3.6.1

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

List

Global Environment

Environment is empty

Files

Plots

Packages

Help

Viewer

New Folder

Upload

Delete

Rename

More

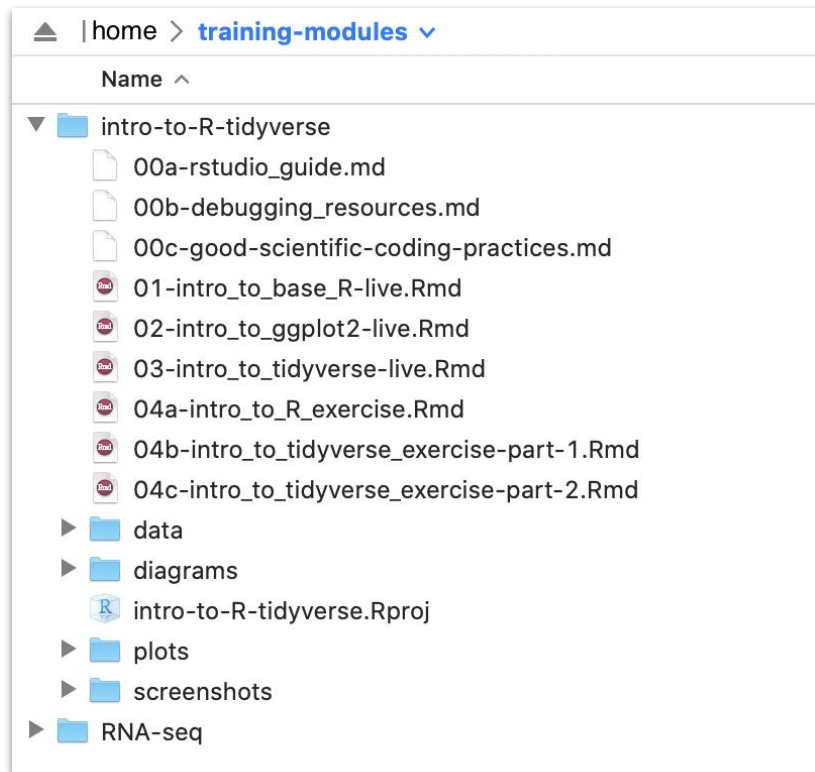
Home

	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

File paths: Directions to a file or folder

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

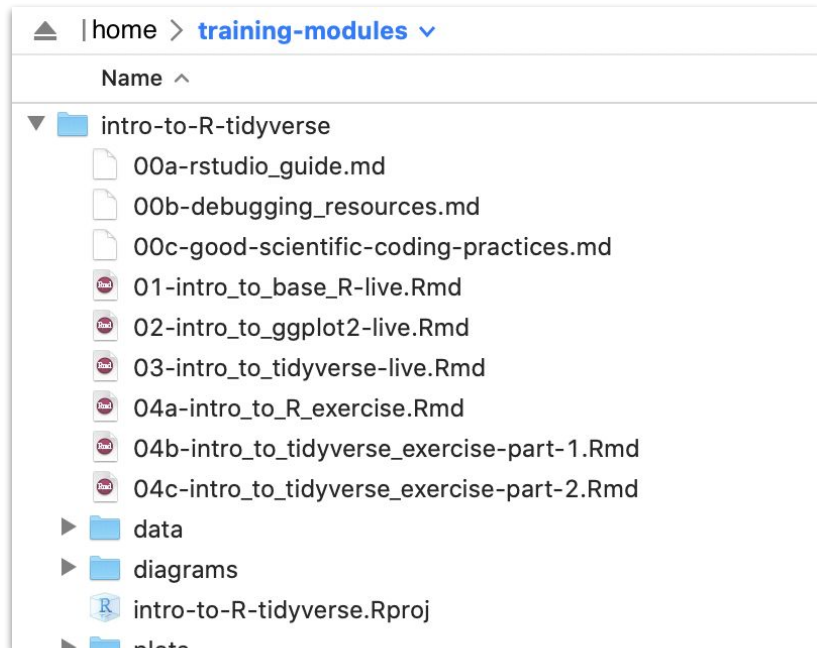


Current directory = “training-modules”

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

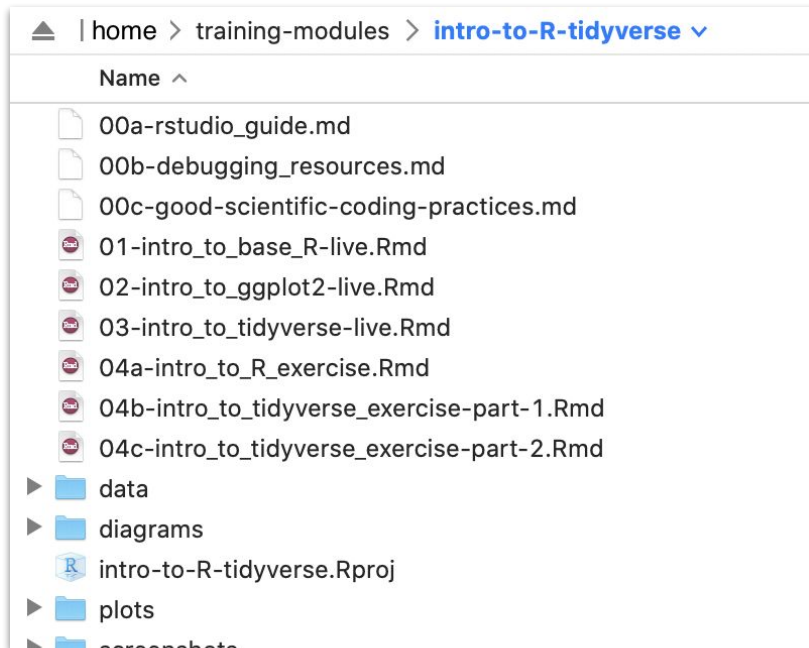
Relative file paths

Let's say we want 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 an IDE for working in R

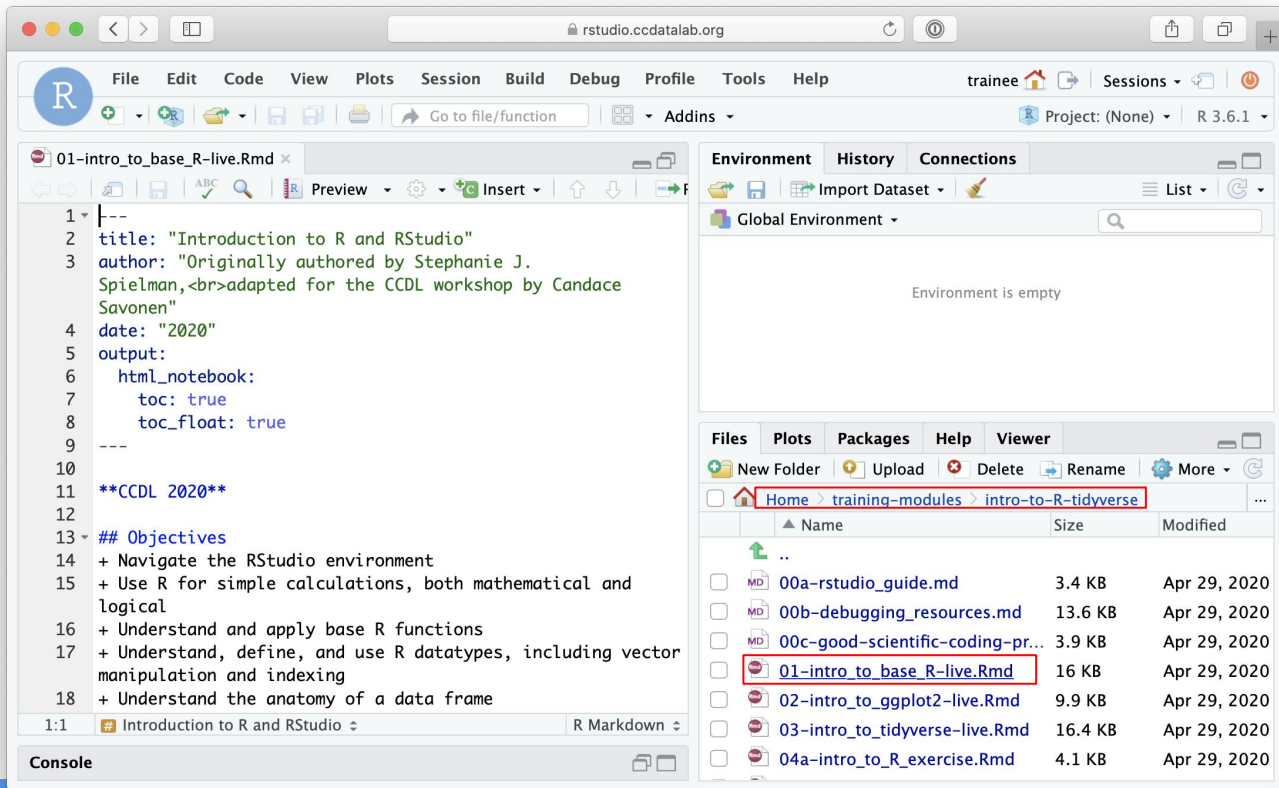
- IDE: Integrated Development Environment
- We write R code using the (free!) RStudio IDE



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

Executable **code chunk**

Can **click** here to run a code chunk

The screenshot shows the RStudio interface with the following components:

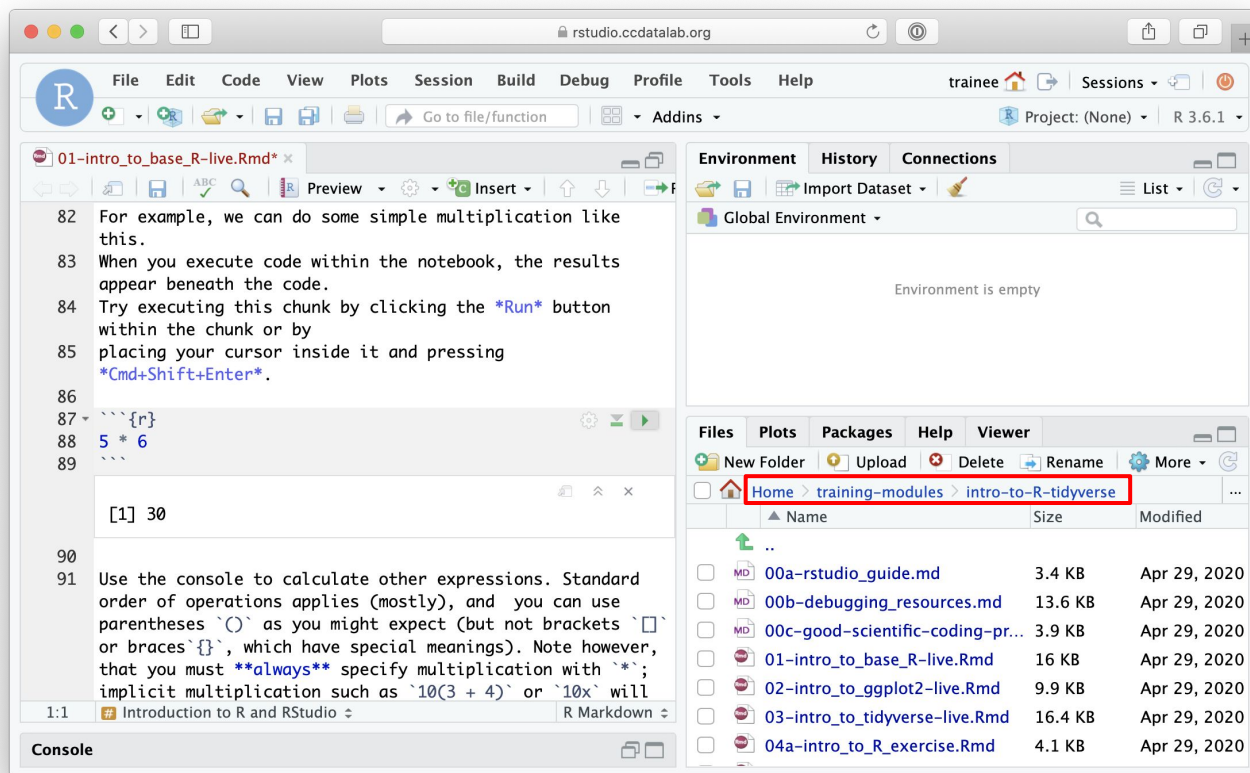
- Source Editor:** Contains R code and text. Lines 82-85 explain how to execute a code chunk. Line 87 contains a code chunk header ````{r}`. Lines 88-89 contain the code `5 * 6`. Line 90 shows the output `[1] 30`. Line 91 explains the use of the console for calculations.
- Environment Panel:** Shows the Global Environment, which is empty.
- Files Panel:** Shows a list of files in the project directory, including `01-intro_to_base_R-live.Rmd`.
- Console:** Shows the command prompt and the output of the code chunk.

Annotations in the image:

- A blue arrow points from the text "Executable **code chunk**" to the code chunk header ````{r}`.
- A green arrow points from the text "Can **click** here to run a code chunk" to the green run button in the top right corner of the code chunk.
- A red arrow points from the text "Output from above code chunk" to the output `[1] 30`.

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.ccdatalab.org

File Edit Code View Plots Session Build Debug Profile Tools Help

trainee Sessions Project: (None) R 3.6.1

01-intro_to_base_R-live.Rmd*

82 For example, we can do some simple multiplication like this.

83 When you execute code within the notebook, the results appear beneath the code.

84 Try executing this chunk by clicking the ***Run*** button within the chunk or by

85 placing your cursor inside it and pressing ***Cmd+Shift+Enter***.

86

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

Environment History Connections

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
	..		
	00a-rstudio_guide.md	3.4 KB	Apr 29, 2020
		3.6 KB	Apr 29, 2020
		9 KB	Apr 29, 2020
		6 KB	Apr 29, 2020
		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

Click here to show the Console

Console

rstudio.ccdatalab.org

File Edit Code View Plots Session Build Debug Profile Tools Help

trainee Sessions Project: (None) R 3.6.1

01-intro_to_base_R-live.Rmd*

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

[1] 5.5

```
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`.
124
125 # Introduction to R and RStudio
```

R Markdown

Environment History Connections

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

Console Terminal Jobs

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

R Console:
What you are actually telling R to do

 File Edit Code View Plots Session Build Debug Profile Tools Help trainee Sessions Project: (None) R 3.6.1

01-intro_to_base_R-live.Rmd*

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`.

1:1 Introduction to R and RStudio R Markdown

Console Terminal Jobs

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

Environment History Connections

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 screenshot displays the RStudio application window. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, and Help. The top toolbar contains icons for file operations and a search bar. The main editor window shows a script file named '01-intro_to_base_R-live.Rmd' with R code and a preview of the rendered document. The Environment pane on the right is highlighted with a red box and contains the text 'R environment: What R knows and remembers for you'. The console at the bottom shows the execution of the code.

Environment History Connections

Global Environment

Values

x	5.5
---	-----

R environment:
What R knows and remembers for you

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

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

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

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

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

rstudio.ccdatalab.org

File Edit Code View Plots Session Build Debug Profile Tools Help

trainee Sessions Project: (None) R 3.6.1

01-intro_to_base_R-live.Rmd*

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

[1] 5.5

```
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`.
124
125 1:1 Introduction to R and RStudio R Markdown
```

Console Terminal Jobs

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

Environment History Connections

Global Environment

Values

x	5.5
---	-----

Other Assistance Tabs:
Things that help you in your coding

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"/>	00b-debugging_resources.md	13.6 KB	Apr 29, 2020
<input type="checkbox"/>	00c-good-scientific-coding-pr...	3.9 KB	Apr 29, 2020
<input type="checkbox"/>	01-intro_to_base_R-live.Rmd	16 KB	Apr 29, 2020
<input type="checkbox"/>	02-intro_to_ggplot2-live.Rmd	9.9 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