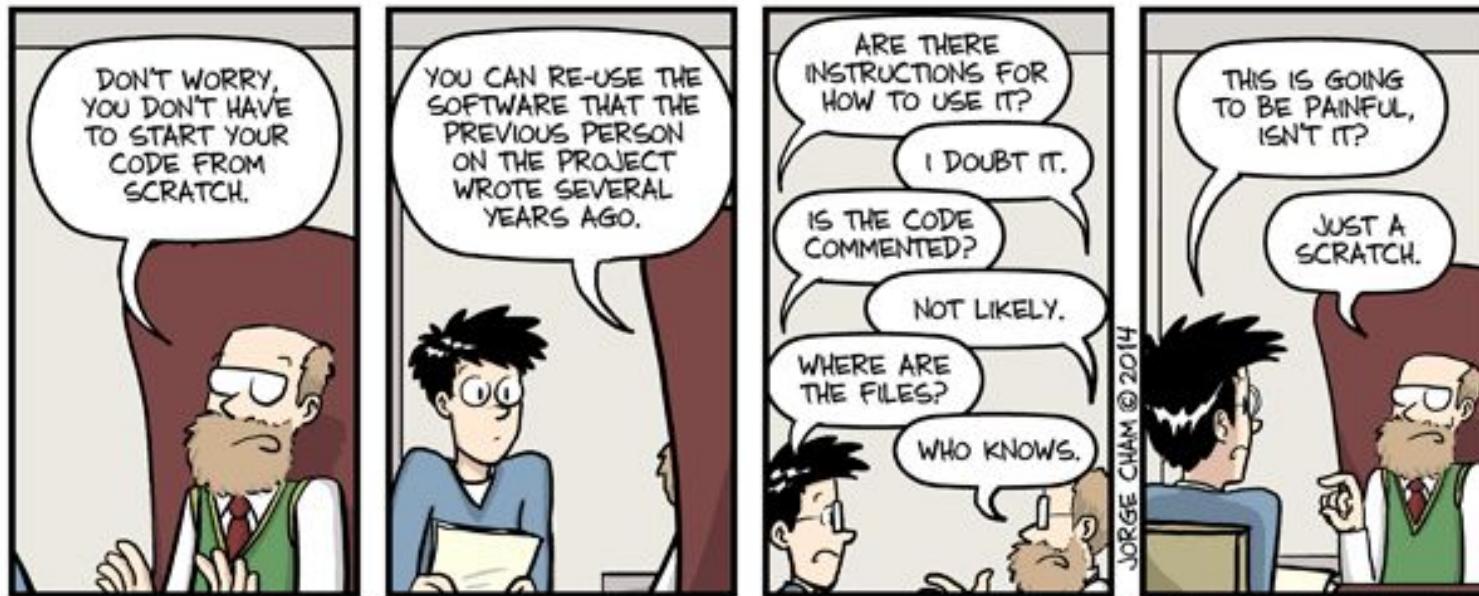


Introduction to R, RStudio, and RStudio Server

The CCDL

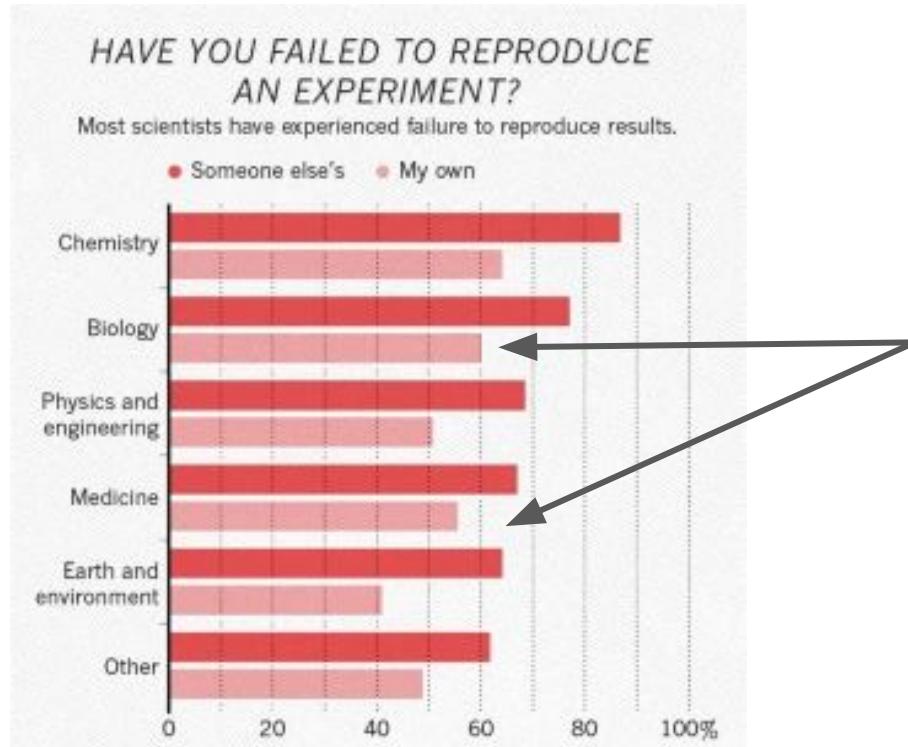
Who's been here before?



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>

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 shows the RStudio Server interface running in a web browser. The top navigation bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help, and Sessions. The user is identified as "trainee". The Project dropdown shows "(None)" and the R version is 3.6.1.

Console Tab:

```
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.
```

Environment Tab:

Global Environment

Environment is empty

Files Tab:

Name	Size	Modified
.R	28 B	Apr 23, 2020
.Rhistory		
shared-data		
training-modules		

Terminal Tab:

```
> |
```

The screenshot shows the RStudio interface running on a Mac OS X system. The title bar indicates the session is named "trainee" and the version is "R 3.6.1".

Console Pane: This pane is highlighted with a red border. It displays the standard R startup message, including the version number (3.6.1), copyright information, and license details. It also shows the R environment as empty.

```
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.
```

Environment Pane: This pane shows the Global Environment, which is currently empty.

Files Pane: This pane shows the directory structure under "Home".

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

The screenshot shows the RStudio interface with several panes open:

- Top Bar:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help, trainee, Sessions, Project: (None), R 3.6.1.
- Toolbar:** Includes tabs for R, Go to file/function, Addins, and Project: (None).
- Console Tab:** Shows "trainee:~\$".
- Terminal Tab:** Shows "Terminal 1 /home/may2020/trainee". A yellow box labeled "tabs to switch" points to the terminal tab.
- Environment Tab:** Shows "Global Environment" with "Environment is empty".
- Files Tab:** Shows a file tree 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		

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

The screenshot shows the RStudio interface with a red box highlighting the Terminal and Environment panes.

Terminal Pane:

- Shows the command `trainee:~$`.
- Two arrows point from the text "These indicate what **directory** you are **currently** carrying out a command in" to the tilde (~) and the dollar sign (\$).
- The path `/home/may2020/trainee` is displayed above the prompt.

Environment Pane:

- Shows the "Global Environment" tab.
- Text: "Environment is empty".

Files Pane:

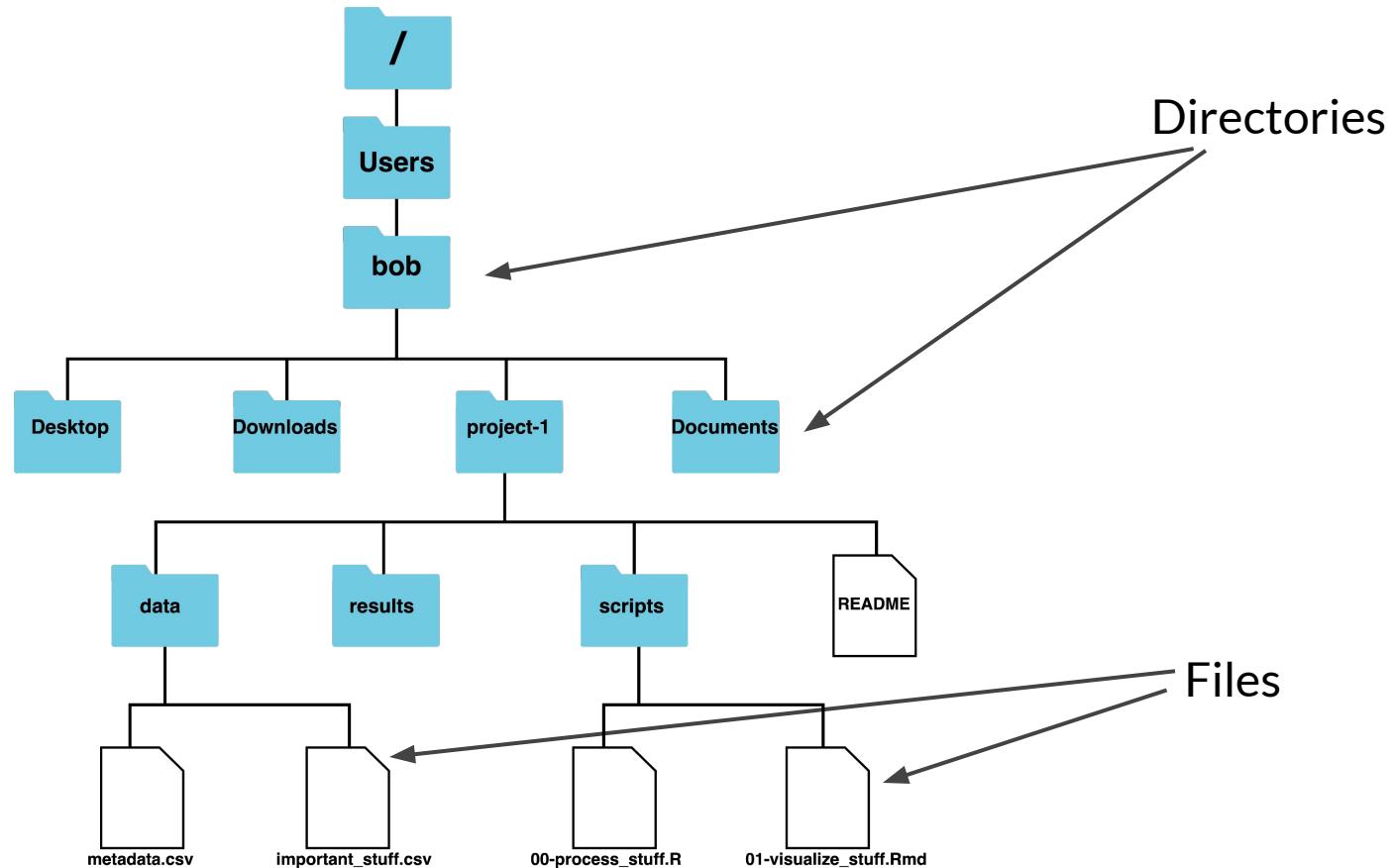
Name	Size	Modified
.R	28 B	Apr 23, 2020
.Rhistory		
shared-data		
training-modules		

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

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

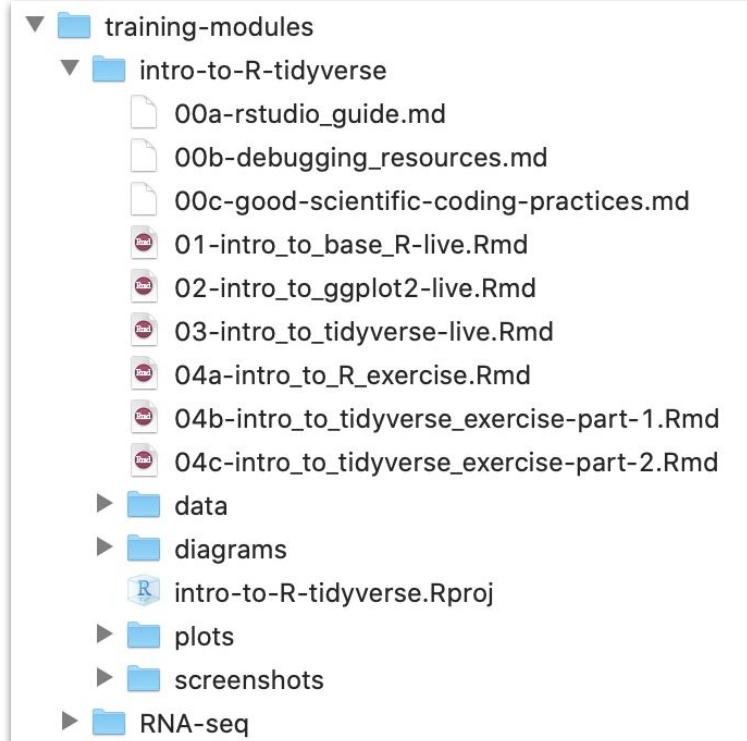
"**~**" is a shortcut for your "**Home**"
directory, so these mean the same thing.

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 screenshot shows the RStudio interface on a Mac OS X system. The top bar includes the title bar with the URL 'rstudio.ccdatalab.org', a refresh button, and a help icon. The menu bar has options: File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help, and Sessions. The top right corner shows 'trainee' with a house icon, a 'Project: (None)' dropdown, and 'R 3.6.1'. The left pane contains a 'Console' tab, a 'Terminal' tab (active), and a 'Jobs' tab. The terminal window shows the command-line history:

```
trainee:~$ ls
shared-data  training-modules
trainee:~$ cd training-modules
```

The right pane is divided into three sections: 'Environment', 'History', and 'Connections'. The 'Environment' section shows the 'Global Environment' with a message 'Environment is empty'. The bottom right pane is the 'Files' viewer, showing the contents of the 'training-modules' directory:

	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		

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 details:

- Top Bar:** Contains the RStudio logo, file menu (File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help), user information (trainee), and session management (Sessions).
- Toolbar:** Includes icons for file operations like Open, Save, and Print.
- Console Tab:** Active, showing a terminal session. The command `ls` is run, listing files in the current directory: `shared-data`, `training-modules`.
- Terminal Tab:** Shows the same terminal session.
- Jobs Tab:** Shows a list of jobs.
- Environment Tab:** Shows the Global Environment, which is currently empty.
- History Tab:** Shows a history of commands.
- Connections Tab:** Shows a list of connections.

```
trainee:~$ ls
shared-data training-modules
trainee:~$ cd training-modules
trainee:~/training-modules$ ls
RNA-seq intro-to-R-tidyverse
trainee:~/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 - **l**ist the files and folders in a directory (files that start with a ‘.’ are not shown by default)

cd - **c**hange **d**irectories

The screenshot shows the RStudio interface with the following details:

- Top Bar:** Contains the RStudio logo, file menu (File, Plots, Packages, Help, Viewer), user information (trainee), and session management (Sessions).
- Toolbar:** Includes icons for New Folder, Upload, Delete, Rename, and More.
- Files Tab:** Active, showing the contents of the current directory:
 - New Folder
 - Upload
 - Delete
 - Rename
 - More

	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 details:

- Top Bar:** Contains the RStudio logo, file menu (File, Plots, Packages, Help, Viewer), user information (trainee), and session management (Sessions).
- Toolbar:** Includes icons for New Folder, Upload, Delete, Rename, and More.
- Files Tab:** Active, showing the contents of the current directory:
 - New Folder
 - Upload
 - Delete
 - Rename
 - More

	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 on a Mac OS X system. The top menu bar includes File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help, and a user-specific section for 'trainee'. The top right corner shows session details ('trainee' at home), a refresh icon, and a plus sign for new sessions.

The main workspace is divided into several panes:

- Console:** Displays a terminal session:

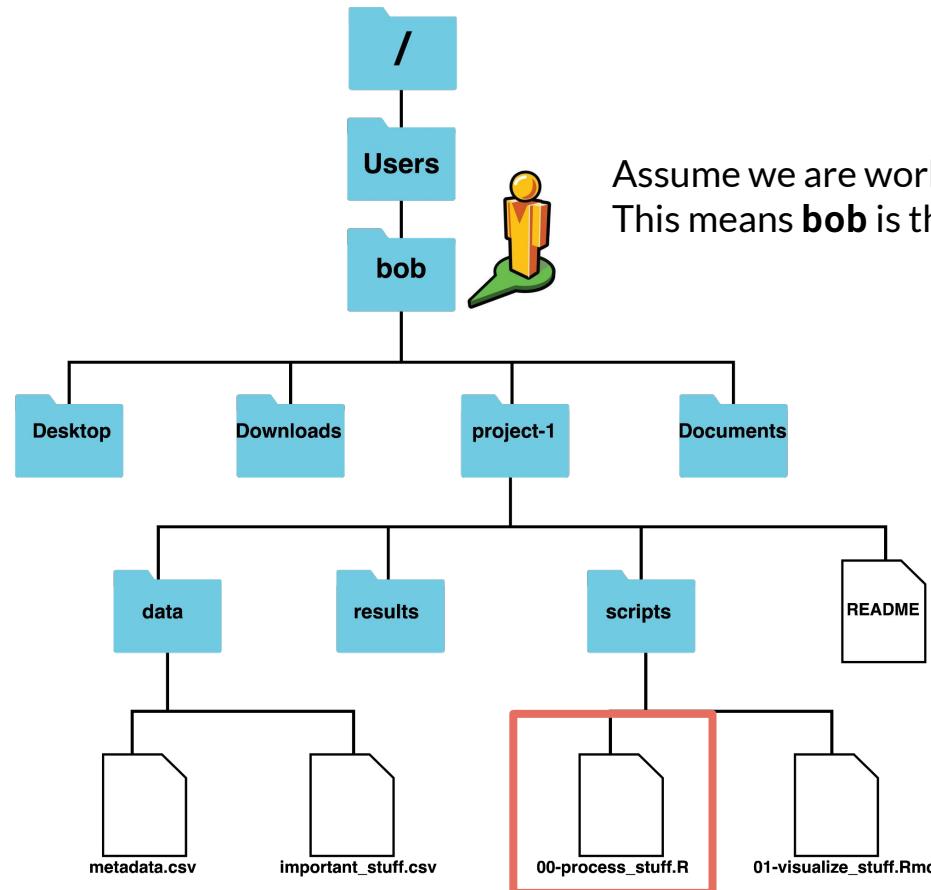
```
trainee:~$ ls
shared-data training-modules
trainee:~$ cd training-modules
trainee:~/training-modules$ ls
RNA-seq intro-to-R-tidyverse
trainee:~/training-modules$
```
- Environment:** Shows the global environment, which is currently empty.
- Files:** Shows the contents of the current directory:

	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		

A large blue text overlay in the bottom right corner reads:

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

We are always working somewhere!



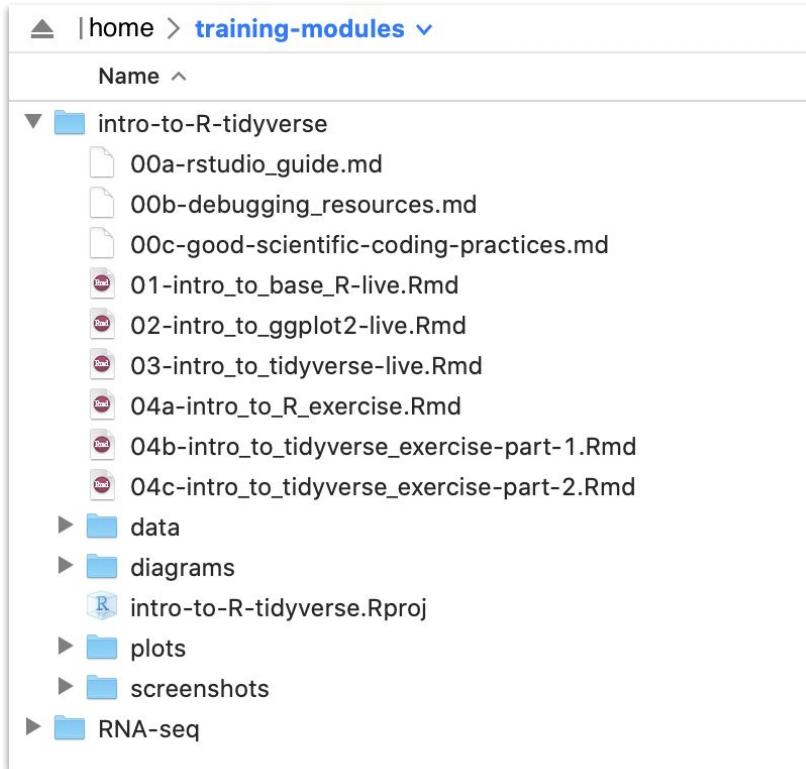
Assume we are working “from” the **bob** directory..
This means **bob** is the **current directory**

The file we are working on

Relative path: **project-1/scripts/00-process_stuff.R**

File paths: Directions to a file or folder

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

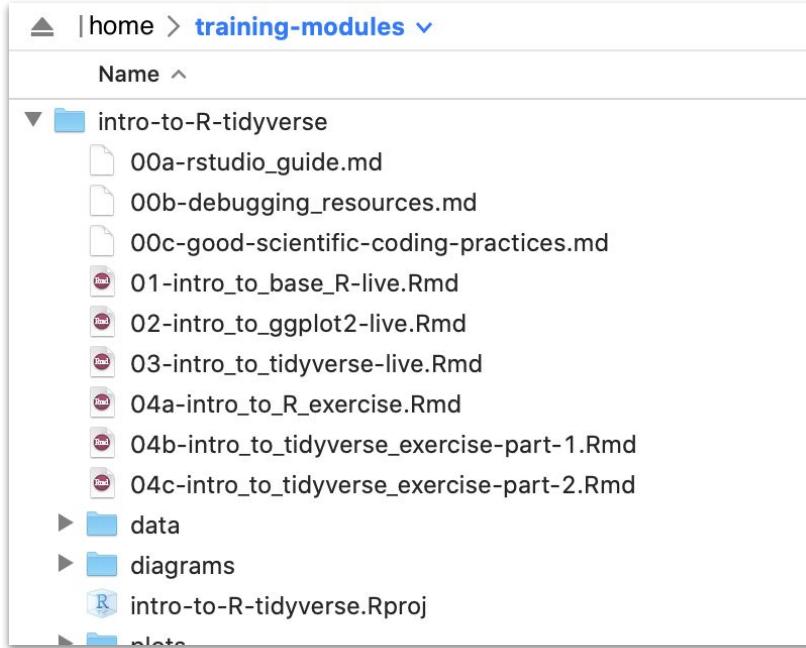


Current directory = “**training-modules**”

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

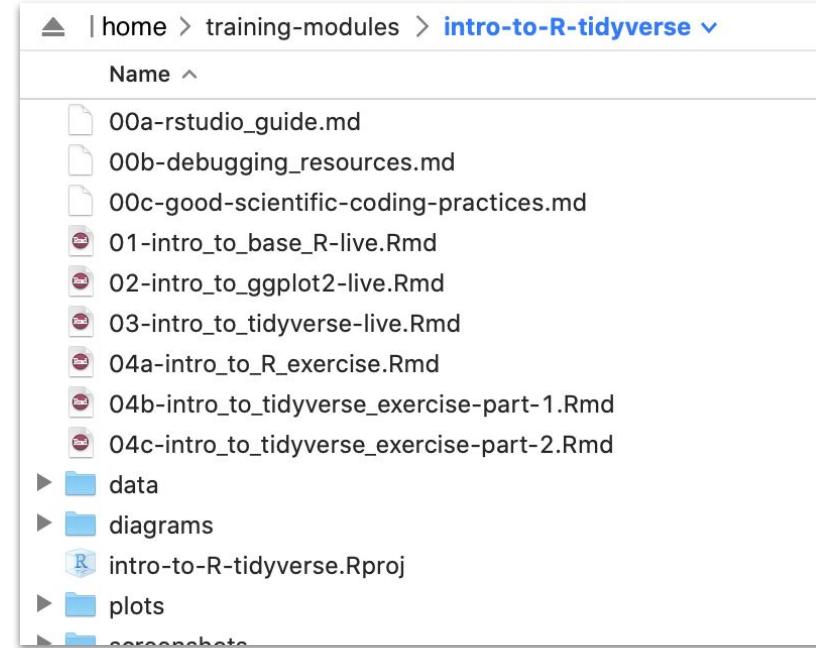
Relative file paths

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



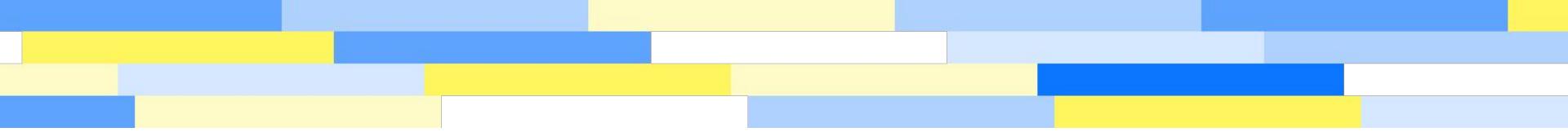
Current directory = “**training-modules**”

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](#)

The screenshot shows the RStudio interface with the following details:

- File Tab:** Shows the file path: `01-intro_to_base_R-live.Rmd`.
- Code Editor:** Displays the R Markdown code, including the YAML front matter and the content of the first code chunk.
- Environment Tab:** Shows the Global Environment, which is currently empty.
- Files Tab:** Selected tab, showing the project structure. The file `01-intro_to_base_R-live.Rmd` is highlighted with a red border.
- Table:** Shows the list of files in the project directory.

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... 01-intro_to_base_R-live.Rmd	3.9 KB 16 KB	Apr 29, 2020 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

Output from above code chunk

Executable code chunk

Can click here to run a code chunk

The screenshot shows the RStudio interface with an R Notebook open. The notebook contains the following text:

```
82 For example, we can do some simple multiplication like
83 When you execute code within the notebook, the results
84 appear beneath the code.
85 Try executing this chunk by clicking the *Run* button
86 within the chunk or by
87 placing your cursor inside it and pressing
88 *Cmd+Shift+Enter*.
89
90 ````{r}
91 5 * 6
92 ````
```

The code chunk starting at line 87 is highlighted with a red box, and its output, "[1] 30", is also highlighted with a red box. A blue arrow points from the text "Executable code chunk" to the start of the code chunk. A green arrow points from the text "Can click here to run a code chunk" to the "Run" button in the toolbar. A red arrow points from the text "Output from above code chunk" to the output "[1] 30".

The RStudio environment pane shows the Global Environment is empty. The file browser pane shows the current project structure:

Name	Size
..	3.4
00a-rstudio_guide.md	13.6
00b-debugging_resources.md	3.9
00c-good-scientific-coding-pr...	16.1
01-intro_to_base_R-live.Rmd	9.9
02-intro_to_ggplot2-live.Rmd	16.4
03-intro_to_tidyverse-live.Rmd	4.1
04a-intro_to_R_exercise.Rmd	

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!

The screenshot shows the RStudio interface with an R Markdown file open. The code chunk at line 88 contains the expression `5 * 6`, which is evaluated in the console at line 89 to produce the output `[1] 30`. The file tree in the bottom right corner shows a folder structure under "Home":

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

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 ▾

Console

Environment History Connections

Import Dataset

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
03-intro_to_tidyverse-live.Rmd	16.4 KB	Apr 29, 2020
04a-intro_to_R_exercise.Rmd	4.1 KB	Apr 29, 2020
..		
3.6 KB		
9 KB		
6 KB		
9 KB		

Click here to show the Console

The screenshot shows the RStudio interface with the following components:

- Top Bar:** Contains the RStudio logo, a search bar with the URL `rstudio.ccdatalab.org`, and various application icons.
- File Menu:** File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help.
- User Information:** trainee, Sessions, Project: (None), R 3.6.1.
- Left Panel:** Shows the code editor for a file named `01-intro_to_base_R-live.Rmd*`. The code includes:

```
113
114 ``{r}
115 x <- 5.5
116
117 x
118```
[1] 5.5
```

Text below the code:
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
- Environment Tab:** Shows the global environment with a value for `x`:

Values
x 5.5
- Files Tab:** Shows the directory structure:

Files	Plots	Packages	Help	Viewer
New Folder	Upload	Delete	Rename	More
Home > training-modules > intro-to-R-tidyverse				

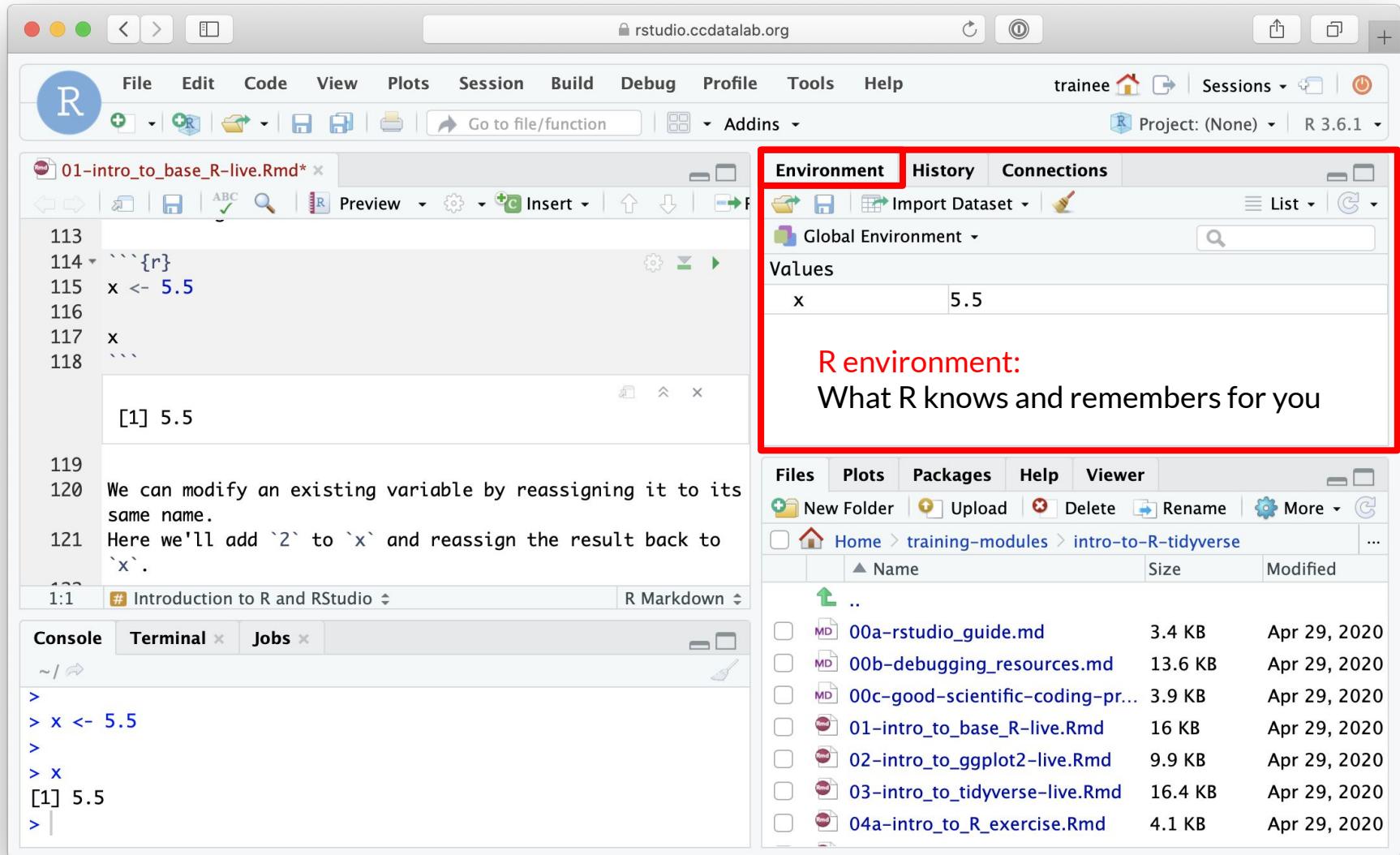
Name	Size	Modified
00a-rstudio_guide.md	3.4 KB	Apr 29, 2020
00b-debugging_resources.md	13.6 KB	Apr 29, 2020
00c-good-scientific-coding-pr... 01-intro_to_base_R-live.Rmd	3.9 KB	Apr 29, 2020
02-intro_to_ggplot2-live.Rmd	16 KB	Apr 29, 2020
03-intro_to_tidyverse-live.Rmd	9.9 KB	Apr 29, 2020
04a-intro_to_R_exercise.Rmd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

```
Console Terminal Jobs
>
> x <- 5.5
> R Console:
> x
What you are actually telling R to do
[1] 5.5
```

The screenshot shows the RStudio interface with several panes:

- Code/Script Editor:** The main editor window contains an R Markdown file titled "01-intro_to_base_R-live.Rmd". It includes code (e.g., `x <- 5.5`), RStudio-specific code (e.g., ````{r}``), and explanatory text. A red box highlights the first few lines of the script.
- Environment:** The top-right pane displays the current environment variables. It shows `x` is assigned the value `5.5`.
- Files:** The bottom-right pane lists files in the current directory: "training-modules/intro-to-R-tidyverse". The files listed are:

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... 01-intro_to_base_R-live.Rmd	3.9 KB	Apr 29, 2020
02-intro_to_ggplot2-live.Rmd	16 KB	Apr 29, 2020
03-intro_to_tidyverse-live.Rmd	9.9 KB	Apr 29, 2020
04a-intro_to_R_exercise.Rmd	16.4 KB	Apr 29, 2020
04b-intro_to_tidyverse_exercise.Rmd	4.1 KB	Apr 29, 2020
- Console:** The bottom-left pane shows the R console history, including the assignment of `x` to `5.5` and its subsequent modification.



The screenshot shows the RStudio interface with the following components:

- Top Bar:** Contains the RStudio logo, file menu (File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help), user information (trainee), session management (Sessions), and project status (Project: (None) R 3.6.1).
- Code Editor:** Displays the file `01-intro_to_base_R-live.Rmd*` containing R code and comments. The code includes:

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

[1] 5.5

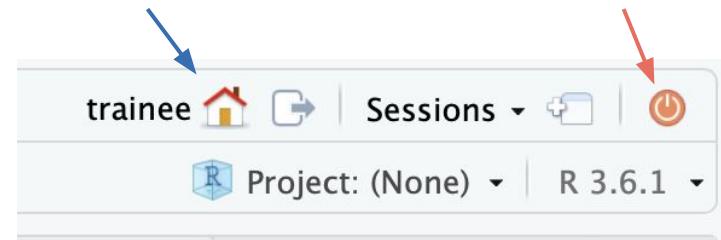
119
120 We can modify an existing variable by reassigning it to its same name.
121 Here we'll add `2` to `x` and reassign the result back to `x`.
- Console:** Shows the command `> x <- 5.5` and its output [1] 5.5.
- Environment Tab:** Shows the variable `x` with value `5.5`.
- Files Tab:** A red box highlights this tab, which displays a file tree and a list of files in the current directory:

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... 01-intro_to_base_R-live.Rmd	3.9 KB	Apr 29, 2020
02-intro_to_ggplot2-live.Rmd	16 KB	Apr 29, 2020
03-intro_to_tidyverse-live.Rmd	9.9 KB	Apr 29, 2020
04a-intro_to_R_exercise.Rmd	16.4 KB	Apr 29, 2020
	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

The 'Session Page' interface for RStudio Server Pro. At the top, it says 'R Studio Server Pro' and shows 'trainee Logout'. Below that is a navigation bar with 'Sessions' (selected), '+ New Session', '(Home)', 'Projects', '+ New Project', and 'Open Project'. The main area is titled 'Currently running sessions' and shows a single session entry:

	(Home)	• IDLE	R 3.6.1
		(Home)	CREATED: 12:59:54 PM
		LAST USED: 12:59:55 PM	

Below this, there are 'Info', 'Suspend', and 'Quit' buttons.