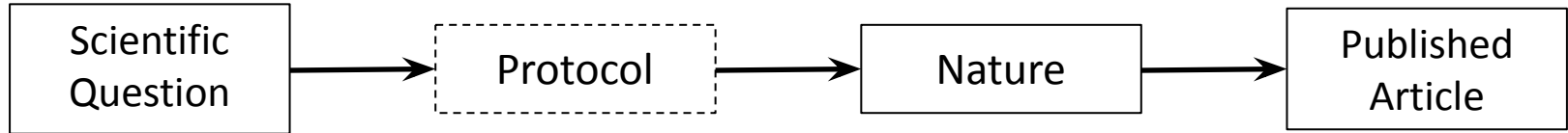


Data Wrangling in R

Reproducible Research

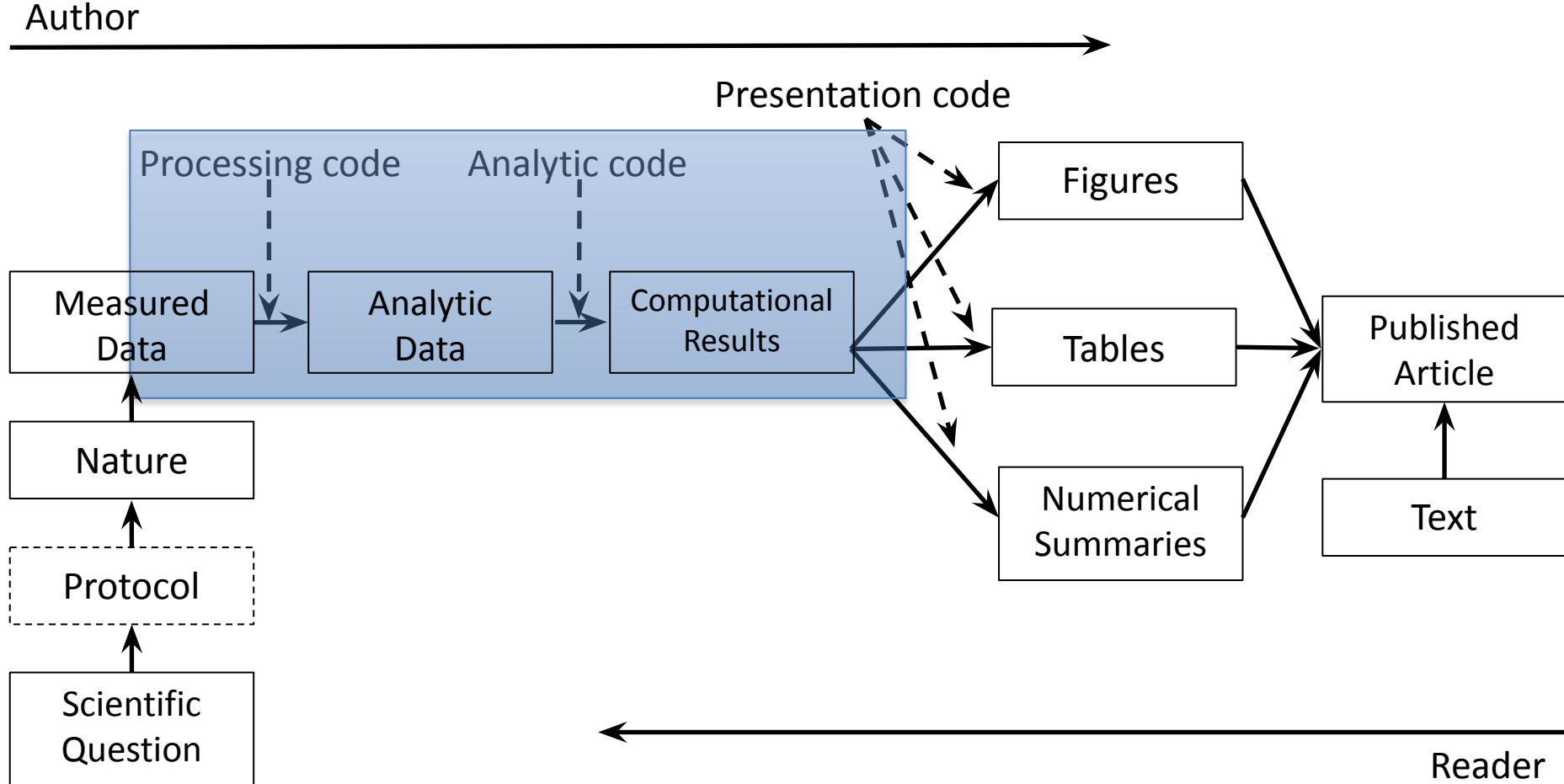
What is Reproducible Research?

Author



Reader

What is Reproducible Research?



Why is this important?

Research Article

Cancer Biology, Computational and Systems Biology

Investigating the replicability of preclinical cancer biology

Timothy M Errington , Maya Mathur, Courtney K Soderberg, Alexandria Denis, Nicole Perfito, Elizabeth Iorns, Brian A Nosek

Center for Open Science, United States; Quantitative Sciences Unit, Stanford University, United States; Science Exchange, United States; University of Virginia, United States

Dec 7, 2021 • <https://doi.org/10.7554/eLife.71601>  



Reproducibility in Cancer Biology: Challenges for assessing replicability in preclinical cancer biology

Timothy M Errington , Alexandria Denis, Nicole Perfito, Elizabeth Iorns, Brian A Nosek

Center for Open Science, United States; Science Exchange, United States;
University of Virginia, United States

Dec 7, 2021 • <https://doi.org/10.7554/eLife.67995>  

“We report the challenges confronted during a large-scale effort to replicate findings in cancer biology, and describe how improving **transparency and sharing** can make it easier to assess rigor and replicability and, therefore, to **increase research efficiency.**”

Your closest collaborator is
you six months ago, but you
don't reply to emails

- Karl Broman

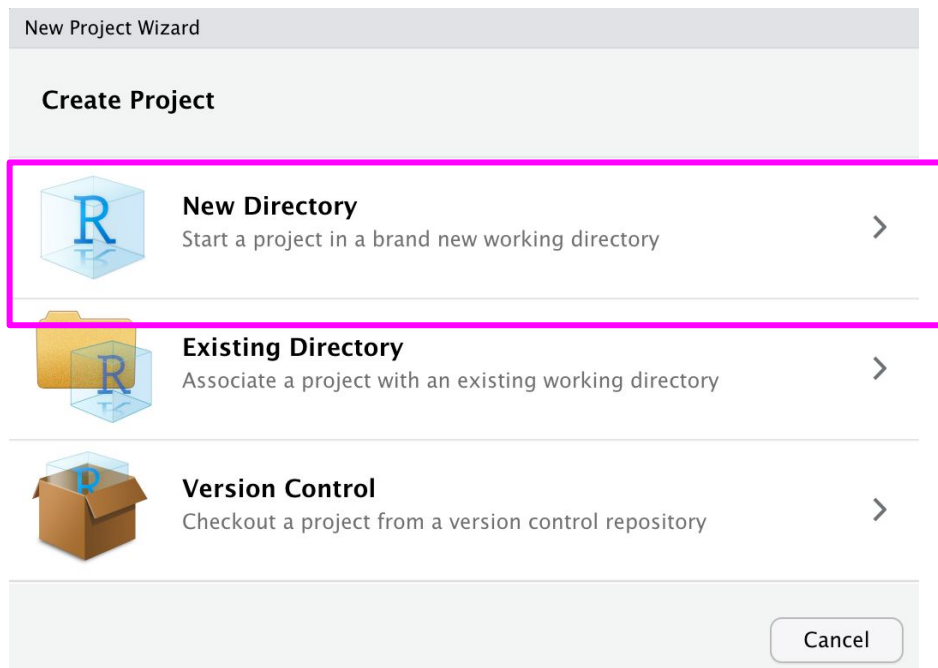
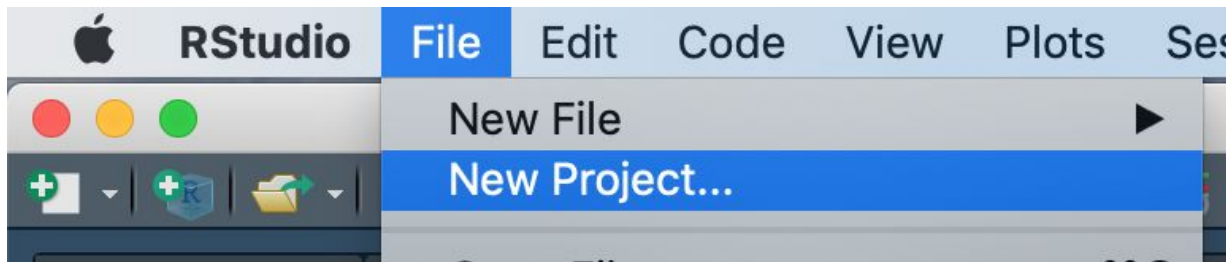
(http://kbroman.org/Tools4RR/assets/lectures/06_org_eda.pdf)

1. Code + documentation
2. Versions of software
3. Data provenance



<https://tenor.com/view/struggle-cant-move-over-it-hard-no-gif-4734482>

RStudio Projects



New Project Wizard

Back

Project Type



New Project



New Project Wizard

Back

Create New Project



Directory name:

SISBID

Create project as subdirectory of:

~/Desktop

Browse...

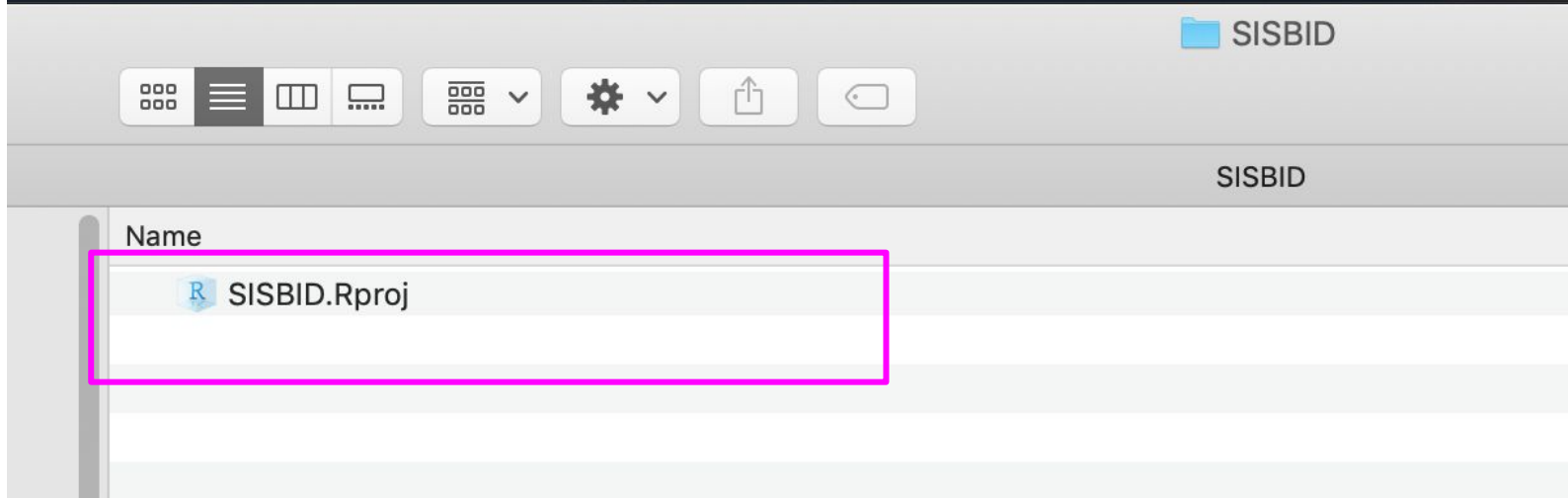
☒ Create a git repository

☐ Use renv with this project

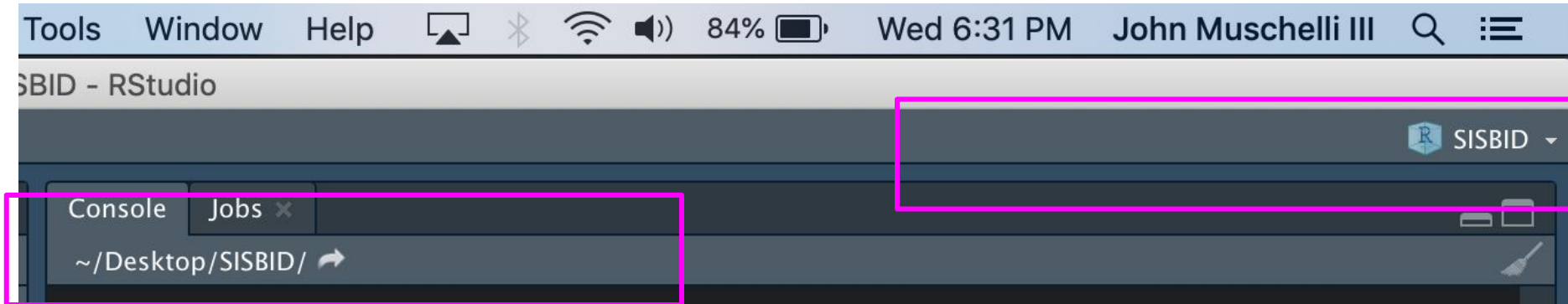
☐ Open in new session

Create Project

Cancel



Double click on the Rproj file - opens RStudio



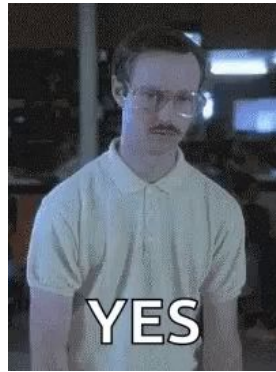
RStudio Projects

- Rproj - Opens to directory of Project
- Makes the working directory where the .Rproj file
- Zip up whole folder and send to someone else

```
read_csv("C:/terrible/path/john/blah.csv")
```



```
read_csv("data/blah.csv")
```



The *here* package

If you have subdirectories or other issues the *here* package makes R projects even better!

```
here("data", "raw_data.csv")
```

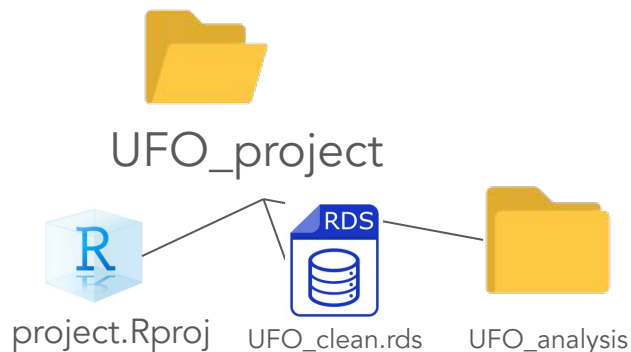
Returns:

Mac: `"/Users/yourname/projects/my_project/data/raw_data.csv"` or

Windows: `"C:\Users\yourname\projects\my_project\data\raw_data.csv"`

- Makes the path work for any OS (Mac, Windows and Linux - different slashes)
- Helps with nested directories (no need to specify going out of a subdirectory)
- Works for a greater variety of circumstances (Rmarkdown/quarto vs. console)

Subdirectory situation - why here helps!



```
```{r}
library(tidyverse)
library(here)
#UFO <- read_rds("UFO_clean.rds") # although this is in the dir with the Rproj file, we are in a
subdirectory
UFO <- read_rds("../UFO_clean.rds") # this works
UFO <- read_rds(here("UFO_clean.rds"))# this works
```
```

here() starts at /Users/carriewright/Documents/GitHub/ /UFO_project

Different contexts with *here*

Assuming UFO_clean.rds is in the directory with the .Rproj file:

UFO <- read_rds("../UFO_clean.rds") Works in an Rmd file

Doesn't work in a script or the console

UFO <- read_rds(here("UFO_clean.rds")) Works for either!

Note - you need *here* package installed and loaded (using *library(here)*), or you could do without loading:

UFO <- read_rds(here::here("UFO_clean.rds"))

R Markdown lets you test your work

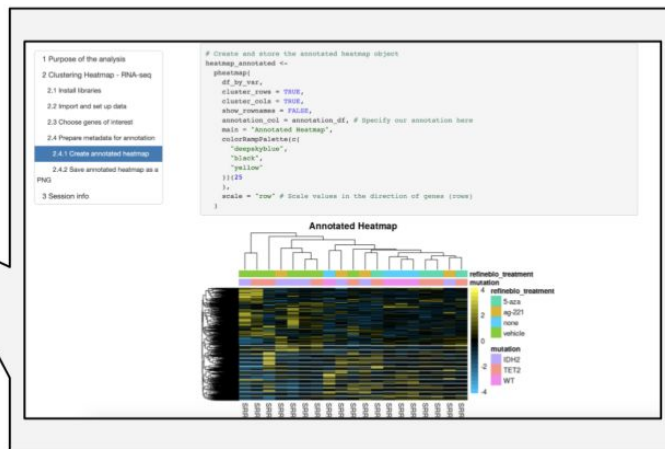
Working from this notebook allows me to interactively develop on my data analysis and write down my thoughts about the process all in one place!



Ruby the
Researcher

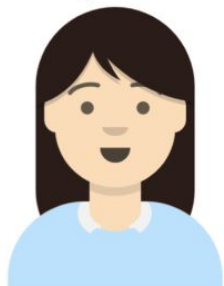


**RMarkdown is conducive to
interactive development!**



R Markdown allows you to more clearly show what you did

Avi, here's some output from this scientific notebook I've been developing from!



Ruby the
Researcher



This is so easy to follow and read, even though I didn't write the code. Thanks for sharing your exciting results!



Avi the
Associate

**RMarkdown creates easily
shareable output!**



Image created by Candace Savonen using Avataars.

R Markdown makes it easier to update code and see results

Yay! I just got the data for 5 more samples. Because of my handy notebook set up, I can easily call one command and re-run the analysis so it is updated with the new samples included!



Ruby the
Researcher



**RMarkdown is handy for
creating updateable reports!**

The magic of Markdown

- bullets
- **bold**
- *italics*
- [links](https://google.com)
- or run inline `r code`



- bullets
- **bold**
- *italics*
- links
- or run inline `r code`

https://rmarkdown.rstudio.com/authoring_basics.html

Introduction

Here is some background you need to know:

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nullam iaculis enim ut enim viverra molestie. In lacinia aliquet urna, nec vulputate quam congue et. Maecenas porta mauris sem, nec laoreet sapien tincidunt non. Integer sit amet consequat neque, non iaculis ligula.

Hypothesis

Pellentesque molestie erat nec elit efficitur, sit amet sodales erat viverra. Mauris sed commodo eros, ac volutpat sem. Morbi convallis leo et dui cursus, eu suscipit turpis efficitur.

Section 1 code and results

First I will run this.

```
##{r}
print("Hello world")
print("Yup, this is important")
```

The output of which is consistent with my hypothesis.

Conclusion

I can move on to the next part of my project

.Rmd
document

Introduction

Here is some background you need to know:

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nullam iaculis enim ut enim viverra molestie. In lacinia aliquet urna, nec vulputate quam congue et. Maecenas porta mauris sem, nec laoreet sapien tincidunt non. Integer sit amet consequat neque, non iaculis ligula.

Hypothesis

Pellentesque molestie erat nec elit efficitur, sit amet sodales erat viverra. Mauris sed commodo eros, ac volutpat sem. Morbi convallis leo et dui cursus, eu suscipit turpis efficitur.

Section 1 code and results

First I will run this.

```
print("Hello world")
```

```
## [1] "Hello world"
```

```
print("Yup, this is important")
```

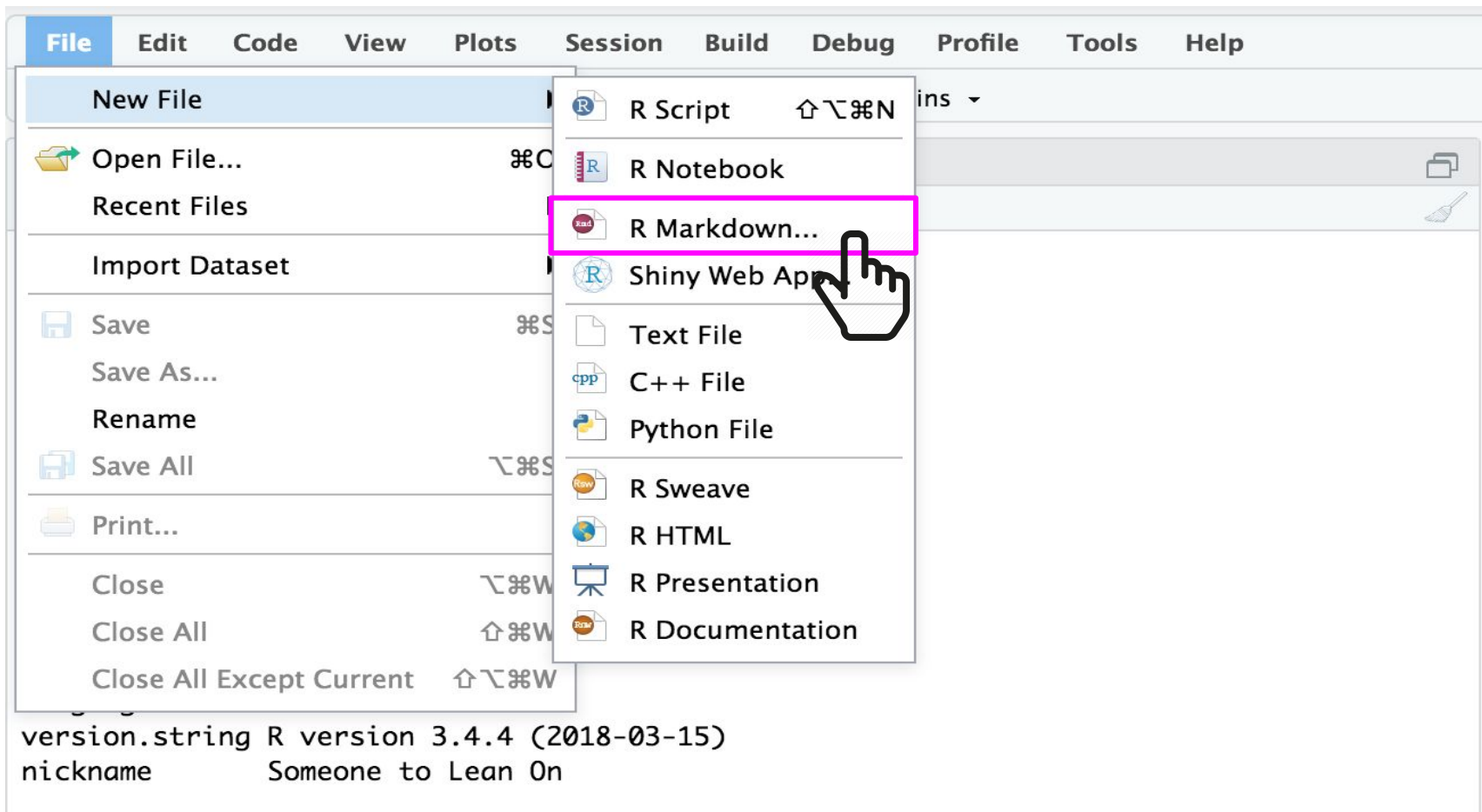
```
## [1] "Yup, this is important"
```

The output of which is consistent with my hypothesis.

Conclusion

I can move on to the next part of my project

PDF, HTML or Word
document



New R Markdown



Document



Presentation



Shiny



From Template

Title:

Untitled

Author:

Default Output Format:

☒ HTML

Recommended format for authoring (you can switch to PDF or Word output anytime).

☐ PDF

PDF output requires TeX (MiKTeX on Windows, MacTeX 2013+ on OS X, TeX Live 2013+ on Linux).

☐ Word

Previewing Word documents requires an installation of MS Word (or Libre/Open Office on Linux).

OK

Cancel

New R Markdown

 Document

 Presentation

 Shiny

 From Template

Title:

Author:

Default Output Format:

☒ **HTML**
Recommended format for authoring (you can switch to PDF or Word output anytime).

☐ **PDF**
PDF output requires TeX (MiKTeX on Windows, MacTeX 2013+ on OS X, TeX Live 2013+ on Linux).

☐ **Word**
Previewing Word documents requires an installation of MS Word (or Libre/Open Office on Linux).

YAML - not markdown,
Spaces matter, usually goes
option: value

TEXT

CODE CHUNK

```
1 ---
2 title: "My First R Markdown Document!"
3 author: "Jane Doe"
4 date: "5/31/2018"
5 output: html_document
6 ---
7
8 ```{r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10 ```
11
12 ## R Markdown
13
14 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word
15 documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.
16
17 When you click the Knit button a document will be generated that includes both content as well as the
18 output of any embedded R code chunks within the document. You can embed an R code chunk like this:
19
20 ```{r cars}
21 summary(cars)
22 ```
23
24 ## Including Plots
25
26 You can also embed plots, for example:
27
28 ```{r pressure, echo=FALSE}
29 plot(pressure)
30 ```
31
32 2:1 # My First R Markdown Document! R Markdown
```

Untitled1 x



1 ---

2 title: "My First R Markdown Document!"

3 author: "Jane Doe"

4 date: "5/31/2018"

5 output: html_document

6 ---

7

8 ```{r setup, include=FALSE}

9 knitr::opts_chunk\$set(echo = TRUE)

10 ```

11

12 ## R Markdown

13

14 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

15

16 When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

17

18 ```{r cars}

19 summary(cars)

20 ```

21

22 ## Including Plots

23

24 You can also embed plots, for example:

25

26 ```{r pressure, echo=FALSE}

27 plot(pressure)

28 ```

29

2:1 # My First R Markdown Document! ↕

R Markdown ↕

Save File – Untitled1

File name:

> / > cloud > project



..



.Rhistory

0 B

May 30, 2018, 12:34 PM



project.Rproj

205 B

May 31, 2018, 3:35 PM

New Folder

Save

Cancel



Header rendered as the title

My First R Markdown Document!

Jane Doe

5/31/2018

R Markdown

Text section rendered as formatted text

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

Code rendered as the input code *AND* the output of running the code chunk

```
##      speed      dist
## Min.   : 4.0    Min.   : 2.00
## 1st Qu.:12.0    1st Qu.: 26.00
## Median :15.0    Median : 36.00
## Mean   :15.4    Mean    : 42.98
## 3rd Qu.:19.0    3rd Qu.: 56.00
## Max.   :25.0    Max.    :120.00
```

Including Plots

You can also embed plots, for example:

test_document.Rmd x

Insert

Run

Knit

```
1 ---
2 title: "My First R Markdown Document!"
3 author: "Jane Doe"
4 date: "5/31/2018"
5 output: html_document
6 ---
7
8 ```{r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10 ```
11
12 ## R Markdown
13
14 This is an R Markdown document. Markdown is a simple formatting syntax for
15 authoring HTML, PDF, and MS Word documents. For more details on using R Markdown
16 see <http://rmarkdown.rstudio.com>.
17:1 R Markdown
```

Console Terminal x

/cloud/project/

> |

Environment History Connections

Files Plots Packages Help Viewer

New Folder Upload Delete Rename More

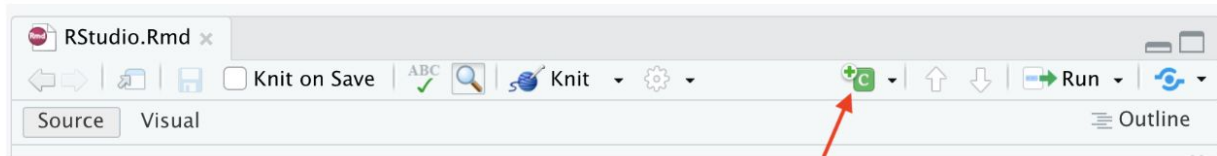
/ > / > cloud > project

| | Name | Size | Modified |
|--------------------------|--------------------|----------|------------------------|
| | .. | | |
| ☐ | .Rhistory | 0 B | May 30, 2018, 12:34 PM |
| ☐ | project.Rproj | 205 B | Jun 1, 2018, 8:33 AM |
| ☐ | test_document.html | 741.2 KB | May 31, 2018, 3:57 PM |
| ☐ | test_document.Rmd | 854 B | May 31, 2018, 3:57 PM |

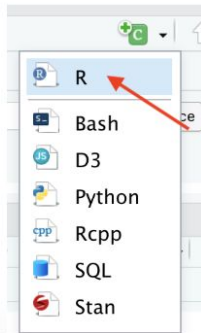
Create Chunks

To create a new R code chunk:

- Use the insert code chunk button at the top of RStudio.

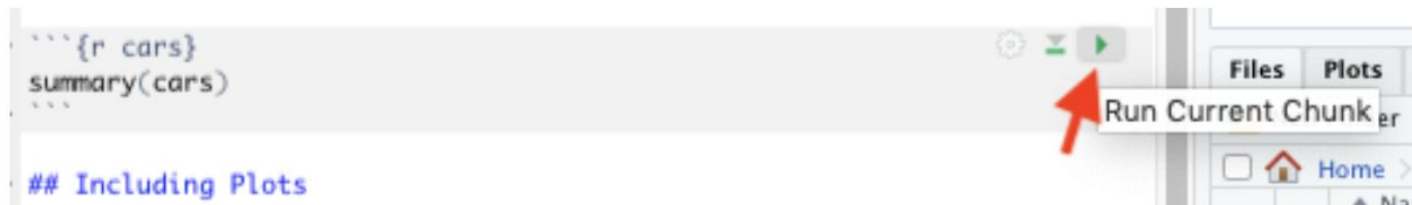


- Select R (default) as the language:



Run code in a chunk

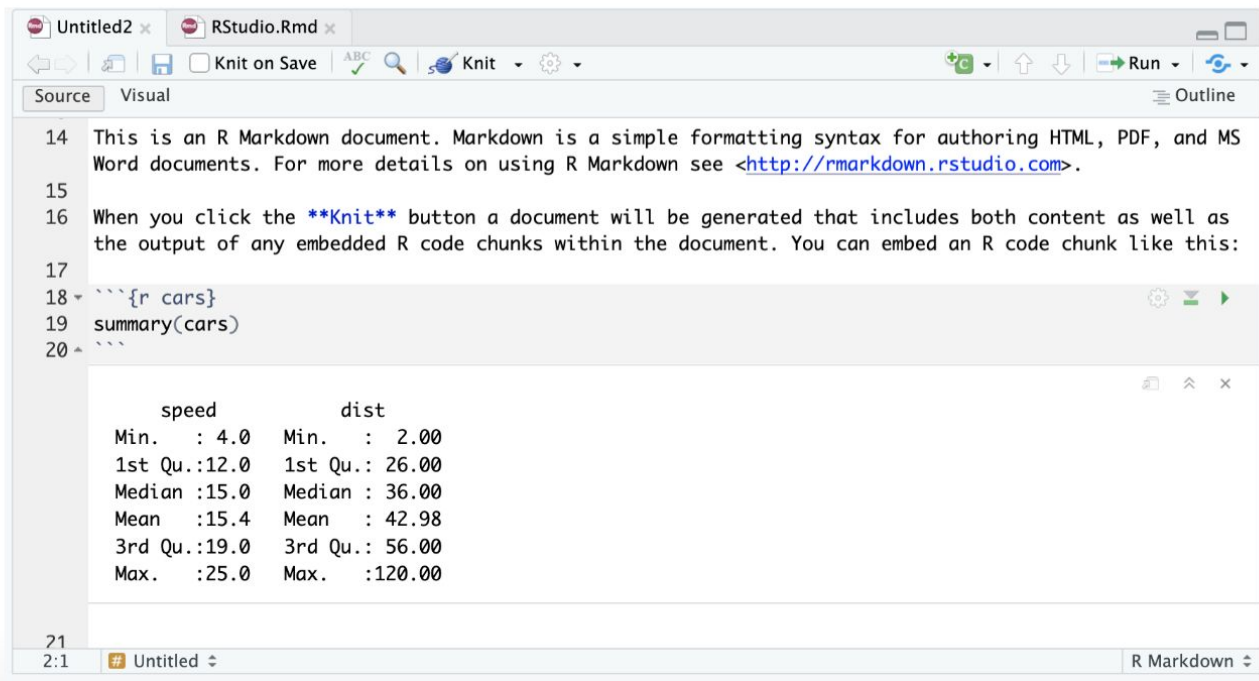
Clicking the run (play) button runs the code in the chunk.



Ctrl + Enter on Windows or Command + Enter on Mac in your script evaluates that line of code

Running a chunk executes the code

- generally see a preview of the output of the code just below the chunk
- see the code in the console



The screenshot shows the RStudio interface with a file named 'RStudio.Rmd' open. The 'Source' pane displays an R Markdown document. The document contains a paragraph of text, a code chunk, and its output. The code chunk is enclosed in triple backticks and contains the R code `summary(cars)`. The output of the code is displayed in a separate pane below the code, showing a summary of the 'cars' dataset. The summary includes statistics for 'speed' and 'dist' variables, such as minimum, 1st quartile, median, mean, 3rd quartile, and maximum values.

```
14 This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS
15 Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.
16 When you click the **Knit** button a document will be generated that includes both content as well as
17 the output of any embedded R code chunks within the document. You can embed an R code chunk like this:
18 ```{r cars}
19 summary(cars)
20 ```
```

| speed | | dist | |
|---------|-------|---------|---------|
| Min. | : 4.0 | Min. | : 2.00 |
| 1st Qu. | :12.0 | 1st Qu. | : 26.00 |
| Median | :15.0 | Median | : 36.00 |
| Mean | :15.4 | Mean | : 42.98 |
| 3rd Qu. | :19.0 | 3rd Qu. | : 56.00 |
| Max. | :25.0 | Max. | :120.00 |

Run previous chunks button

You can run all chunks above a specific chunk using this button:

```
```{r, out.width = "80%", echo = FALSE, fig.align='center'}  
knitr::include_graphics("images/chunk.png")
```
```

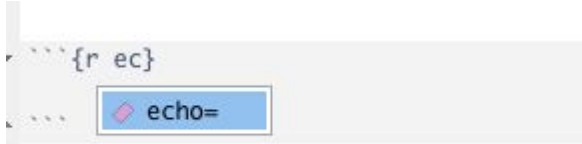


```
{r eval}
```



eval = FALSE will show but not run the code

```
{r ec}
```



echo = FALSE will not show but run the code

Using both set to FALSE will hide the code and will not run the code!

The default is TRUE - do not need to write it out if you want code to be shown and run.

Biiconductor_intro.Rmd x Data_Subsetting_Part1.R x Data_Cleaning.Rmd x

Knit on Save ABC Knit Run

Source Visual Outline

```
60
61 ```{r}
62 p = problems(ufo)
63 p
64 ```
65
66 ## Let's just drop those problematic rows for now.
67
68 Though you would usually want to check them!
69
70 ```{r}
71 ufo = ufo[-p$row,] # brackets can also be used for subsetting
72 ```
```

Environment History Connection

R Global Environment 99 MiB

Chunk Name: Unnamed chunk

Output: (Use document default)

☐ Show warnings

☐ Show messages

☐ Use paged tables

☐ Use custom figure size

[? Chunk options](#)

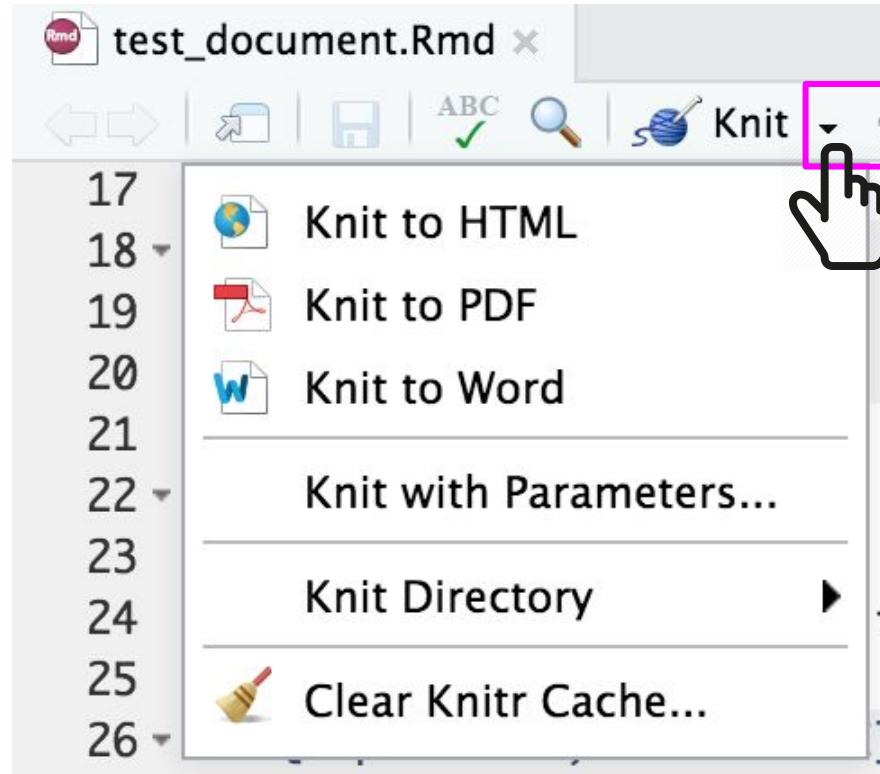
Revert Apply

Can Also influence how a plot is displayed with a chunk!

```
fig.align = "center",  
fig.height = 3, fig.width = 6
```

```
```{r pressure, echo=FALSE, fig.width=6, fig.height=3}  
plot(pressure)
```
```





Gut Check

Where does code go typically in an Rmd file?

A

```
```{r}
```

**B**

```
```
```

C

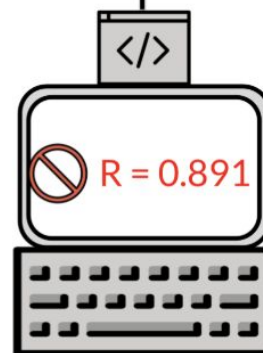
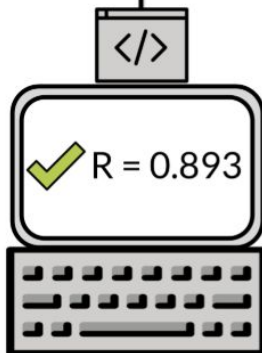
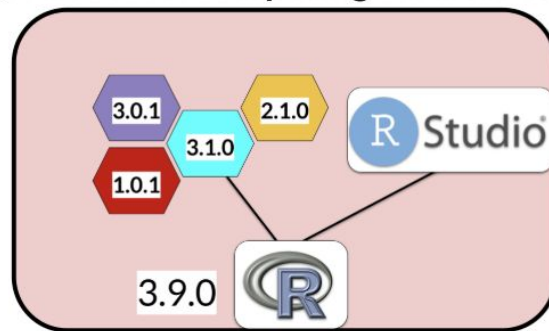
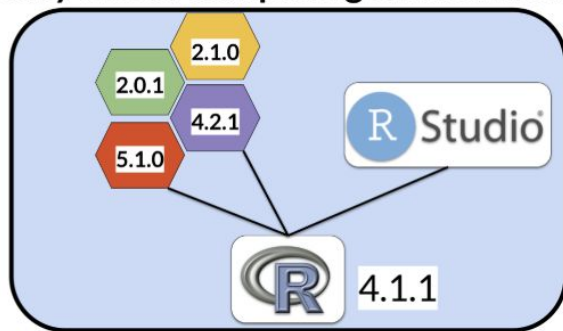
<https://support.rstudio.com/hc/en-us/articles/200711853-Keybaord-Shortcuts>

Rendering in R

```
library(rmarkdown)  
render("Untitled.Rmd")
```

Versions matter

Ruby's local computing environment Avi's local computing environment



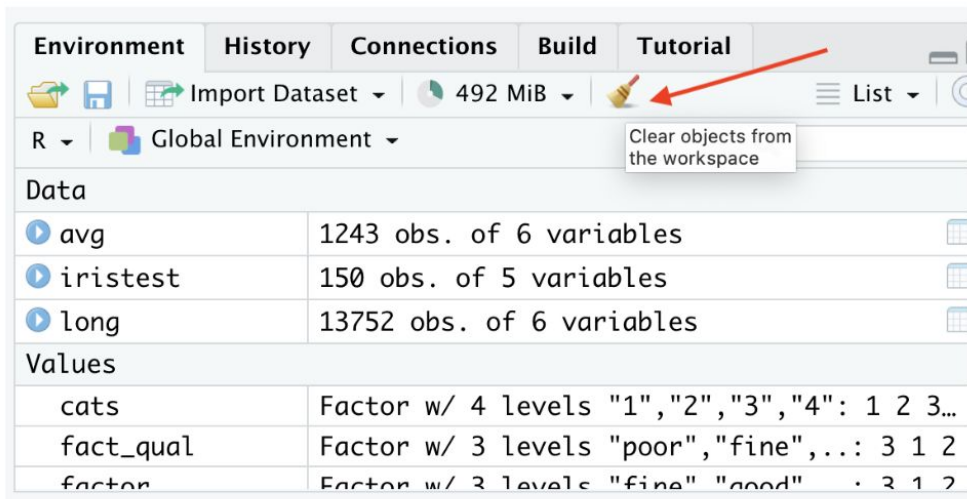
Session information - what's loaded?

```
install.packages("devtools")  
library(devtools)  
session_info()
```

Clean your environment

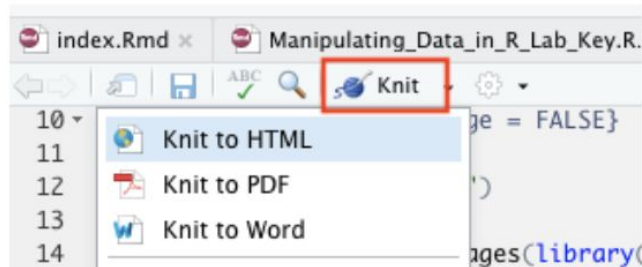
Regularly cleaning your environment and trying your code again, can help ensure that your code is running as expected.

Occasionally we might forget to save a step of our code in our R Markdown file that we ran only in the console. This will help us figure that out.



Check if your file knits regularly

Regularly checking if your file knits will help you spot a missing step or error earlier when you have less code to try to identify where your code might have gone wrong.



GUT CHECK

Why is reproducibility so important?

- A. It helps to ensure that your code is working consistently and it helps others understand what you did
- B. It ensures that your code is correct

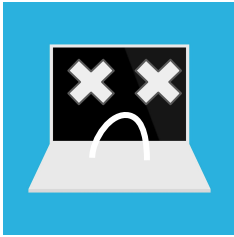
GUT CHECK

What is NOT a practice to improve the reproducibility of our work?

- A. Using R Markdown files to describe what your code is doing
- B. Using scripts instead of R Markdown files
- C. Testing your code with R Markdown files or the run previous button
- D. Regularly cleaning the environment

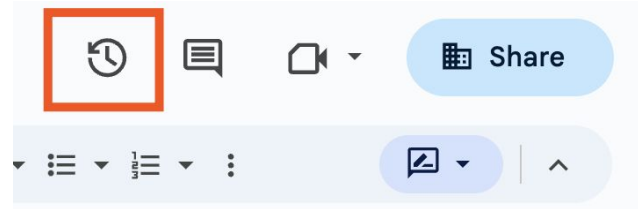
Version Control!

Saving you from not knowing what or where your files are!



Saving you from losing your files when your computer dies or is stolen etc.!

Allowing you to track changes overtime and revisit them!



Like Google Doc history but better.





ARTICLE.dox

Can find specific changes and revert to them if we wish

Change Logs

Commit **cea4b13**

 Browse files

 **carriewright11** authored now · 0 / 1 · Verified

Updating the readme

adding some info about what the course covers.

 **gh-pages**

1 parent **c6d21f7** commit **cea4b13** 

1 file changed

+3 -1 



Search within code



▼ README.md  

+3 -1  <>  ...

@@ -4,4 +4,6 @@

4
5 For more details please visit the html page automatically generated by
GitHub: <http://sisbid.github.io/Data-Wrangling/>

6
7 - See [contribute.md](https://github.com/SISBID/Data-Wrangling/blob/gh-
pages/contribute.md) for details on how to set up the course.

4
5 For more details please visit the html page automatically generated by
GitHub: <http://sisbid.github.io/Data-Wrangling/>

6
7 + See [contribute.md](https://github.com/SISBID/Data-Wrangling/blob/gh-
pages/contribute.md) for details on how to set up the course.

8 +
9 + This course covers major topics in data wrangling and reproducibility
practices with R and Git/GitHub.

GitHub and Git

GitHub

Online place to share code!



Can start here!

Git

Software on your computer to track changes over time.



More advanced but very helpful!

It's like a shared Dropbox folder that you have to sync yourself

Why not just use Dropbox?



GitHub is an online platform that allows for:

- Version control
- Sharing code easily!
- Better record keeping!
- Collaboratively working!
- Extra copy to keep your work safe!

pull

add

clone

push



remote

branch

commit



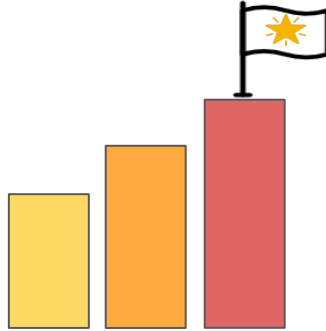
DO NOT PUT PROTECTED
DATA ON GITHUB

Even in a private repository!!!

LET'S PRACTICE!



Next Steps



Step 0: Accounts

Create an account at github.com

Optionally Install git <https://git-scm.com/downloads>

Restart RStudio



Step 1: New repo on GitHub

A new “folder” on github.com that can store files for a project



avahoffman ▾

Repositories

Find a repository...

-  [SISBID/Data-Wrangling](#)
-  [jhudsl/AnVIL_Book_Getting_Started](#)
-  [avahoffman/blue-grama-gene-expression](#)
-  [jhudsl/intro_to_r](#)
-  [avahoffman/DESeq2-viz-demo](#)
-  [jhudsl/jhudsl.github.io](#)
-  [avahoffman/CV-and-resumes](#)

Show more

Recent activity

 Add Instructor checklist as...

 New

All activity



carriewright11 pushed to [SISBID/Data-Wrangling](#) 5 hours ago

6 commits to [gh-pages](#)

 [98387ee](#) fixing links for subsetting lab from 2 to 1

 [faef211](#) Merge pull request [#35](#) from SISBID/subsetting_part2

4 more commits »



carriewright11 pushed to [SISBID/Data-Wrangling](#) 2 days ago

1 commit to [subsetting_part2](#)

 [98387ee](#) fixing links for subsetting lab from 2 to 1



KatherineCox pushed to [jhudsl/AnVIL_Book_Getting_Started](#) 2 days ago

1 commit to [main](#)

 [00d7503](#) Render bookdown

<https://github.com/>

The image shows the GitHub homepage. At the top, there is a dark navigation bar with the GitHub logo, a search bar labeled "Search or jump to...", and links for "Pulls", "Issues", "Marketplace", and "Explore". On the right side of the navigation bar, there is a bell icon, a plus icon in a square (highlighted with a pink box), and a user profile icon. A dropdown menu is open from the plus icon, showing options: "New repository" (highlighted with a pink box), "Import repository", "New gist", "New organization", and "New project". A pink arrow points from the "New repository" option to the "Import repository" option.

Below the navigation bar, the left sidebar shows the user profile for "avahoffman" and a list of repositories: "SISBID/Data-Wrangling", "jhudsl/AnVIL_Book_Getting_Started", "avahoffman/blue-grama-gene-expression", "jhudsl/intro_to_r", "avahoffman/DESeq2-viz-demo", "jhudsl/jhudsl.github.io", and "avahoffman/CV-and-resumes". There is a "New" button and a search bar for repositories.

The main content area shows "All activity" with a list of recent pushes. The first push is by "carriewright11" to "SISBID/Data-Wrangling" 5 hours ago, with 6 commits to "gh-pages". The second push is also by "carriewright11" to "SISBID/Data-Wrangling" 2 days ago, with 1 commit to "subsetting_part2". The third push is by "KatherineCox" to "jhudsl/AnVIL_Book_Getting_Started" 2 days ago, with 1 commit to "main".

<https://github.com/>

Create a new repository

A repository contains all project files, including the revision history. Already have a project repository elsewhere? [Import a repository](#).

Required fields are marked with an asterisk (*).

Repository template

No template ▾

Start your repository with a template repository's contents.

Owner *

 avahoffman ▾

Repository name *

test_project

✔ test_project is available.

Great repository names are short and memorable. Need inspiration? How about [legendary-garban](#)

Description (optional)

Testing out github



Public

Anyone on the internet can see this repository. You choose who can commit.



Private

You choose who can see and commit to this repository.

Initialize this repository with:



Add a README file

This is where you can write a long description for your project. [Learn more about READMEs](#).

Add .gitignore

.gitignore template: None ▾

Choose which files not to track from a list of templates. [Learn more about ignoring files](#).

Choose a license

License: None ▾

A license tells others what they can and can't do with your code. [Learn more about licenses](#).

This will set `main` as the default branch. Change the default name in your [settings](#).



You are creating a public repository in your personal account.

<https://github.com/new>

Create repository

 test_project Private

 Unwatch 1 

 Fork 0 

 Star 0 

 main   1 branch  0 tags

Go to file

Add file 

 Code 

About



Testing out github

 Readme

 Activity

 0 stars

 1 watching

 0 forks

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)

 avahoffman Initial commit

d8bf495 1 minute ago  1 commit

 README.md

Initial commit

1 minute ago

README.md



test_project

Testing out github



© 2023 GitHub, Inc.

[Terms](#)

[Privacy](#)

[Security](#)

[Status](#)

[Docs](#)

[Contact GitHub](#)

[Pricing](#)

[API](#)

[Training](#)

[Blog](#)

[About](#)

Step 1: New repo on GitHub

1. Click "New"
2. Name the repo
3. Select "Public"
4. Add a README
5. Click "Create Repository"

Step 2: Upload a File to GitHub

Add an .Rmd file to the project



test_project Public

Pin Unwatch Fork 0 Star 0

main 1 Branch 0 Tags

Go to file

Add file Code

avahoffman Initial commit 172e618 · 1 minute ago 1 Commit

README.md Initial commit 1 minute ago

README

test_project

Testing out github

About

Testing out github

- Readme
- Activity
- 0 stars
- 1 watching
- 0 forks

Releases

No releases published

[Create a new release](#)

Packages

No packages published

[Publish your first package](#)



Pin



Unwatch

1



Fork

Go to file

t

Add file

<> Code

About

+ Create new file

Upload files

Testing out gi

Readme

Activity

0 stars

1 watching

0 forks

172e61

2 minutes ago

mit



test_project /



Drag additional files here to add them to your repository

Or [choose your files](#)

 My_analysis.Rmd



Commit changes

Add files via upload

Add an optional extended description...

☒ Commit directly to the `main` branch.

☐ Create a new branch for this commit and start a pull request. [Learn more about pull requests.](#)

Commit changes

Cancel



test_project

Public



Pin



Unwatch

1



main



1 Branch



0 Tags



Go to file



Add file

<> Code



avahoffman Add files via upload

566157e · now



2 Commits



My_analysis.Rmd

Add files via upload

now



README.md

Initial commit

17 minutes ago



README



test_project

Testing out github

Step 2: Upload a File to GitHub

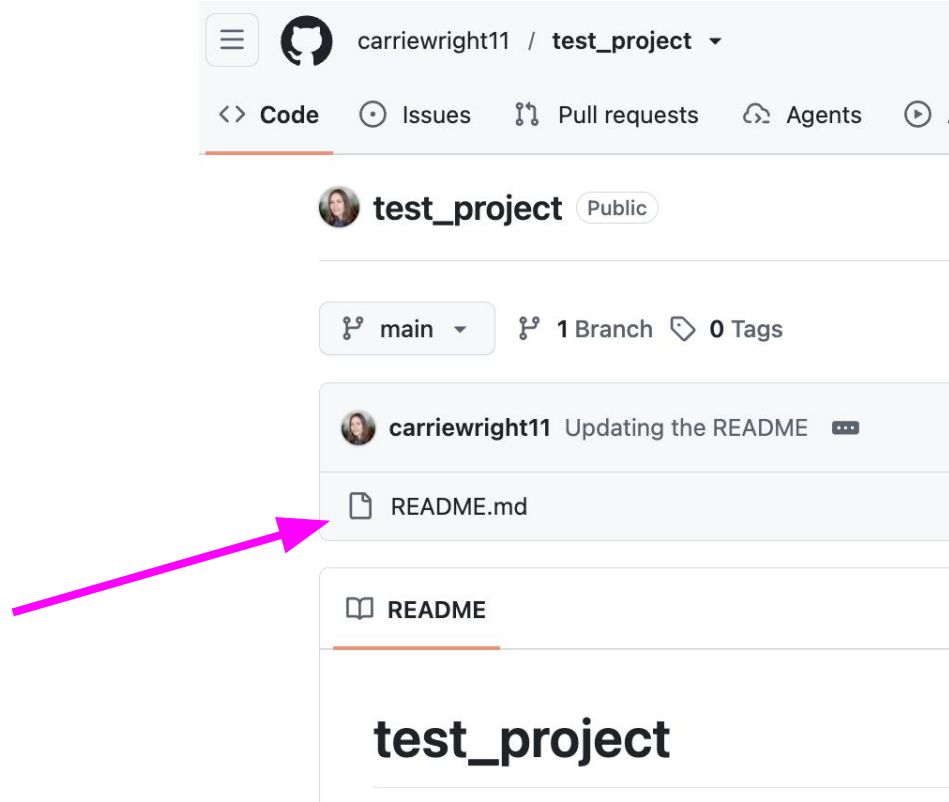
1. "Add File" (or simply a plus if your window is compressed)
2. "Upload Files"
3. Drag and drop / add files
4. Add commit message
5. Click "Commit changes"

Step 3: Upload a Changed File to GitHub

Update the .Rmd file and Upload



Click on the Readme file



Click on the Download Button

test_project / README.md



carriewright11 Updating the README

3888eeb · 10 minutes ago History

Preview

Code

Blame

3 lines (2 loc) · 95 Bytes



Raw



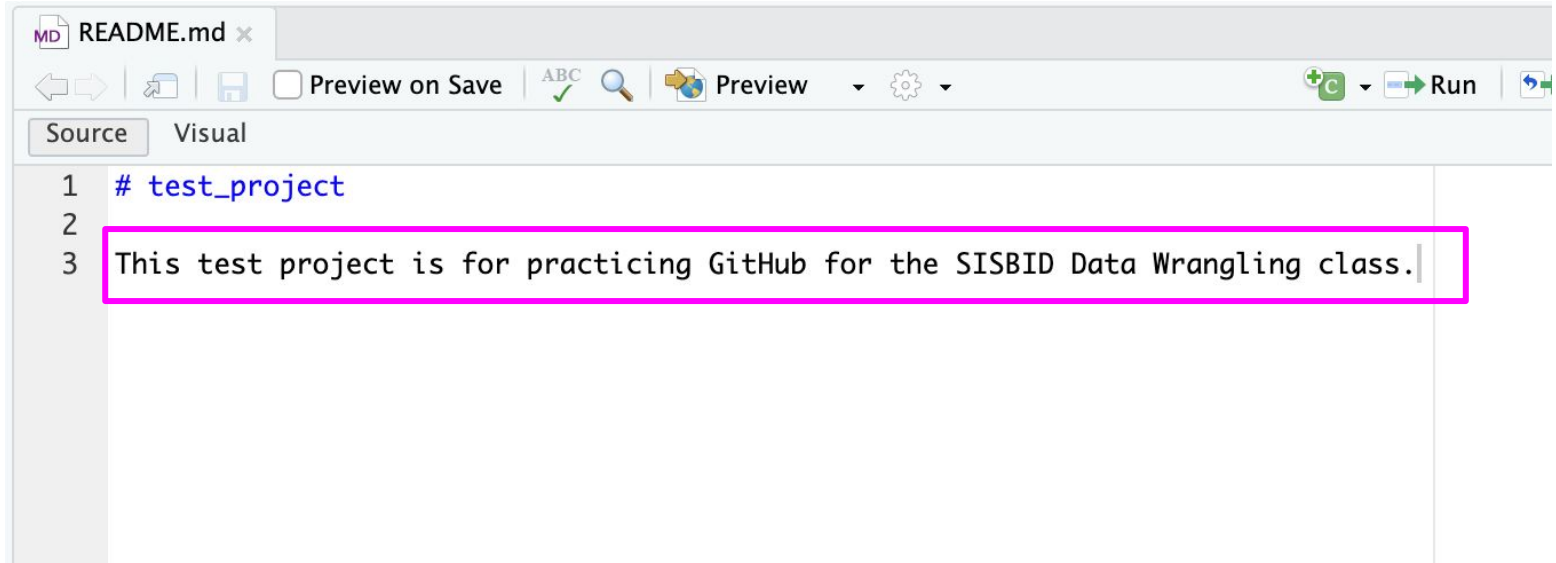
Download raw file

test_project

This test project is for practicing GitHub for the SISBID Data Wrangling class.

Make a change to the readme file

- 1) Open the file in RStudio
- 2) Add a line to the file using RStudio
- 3) Save the file!



The screenshot shows the RStudio interface with the 'README.md' file open in the 'Source' editor. The file content is as follows:

```
1 # test_project
2
3 This test project is for practicing GitHub for the SISBID Data Wrangling class.
```

The third line of code is highlighted with a red rectangular box.



carriewright11 / test_project ▾

🔍 Type to search

Code

Issues

Pull requests

Agents

Actions

Projects

Wiki

Security and quality

Insights

Settings



test_project

Public



Pin



Watch 0

🔗 main ▾

🔗 1 Branch 🏷️ 0 Tags

🔍 Go to file

T

Add file ▾

<> Code ▾

If your
window
is narrow

Code

Issues

Pull requests

Agents

Actions

Projects

Wiki

Security and



test_project

Public



Pin



Watch 0 ▾



🔗 main ▾



Go to file

+

<> Code ▾



Drag additional files here to add them to your repository

Or [choose your files](#)

README.md



Commit changes

Updating the README

Adding some info about what the project is for

☒ Commit directly to the `main` branch.

☐ Create a new branch for this commit and start a pull request. [Learn more about pull requests.](#)

Commit changes

Cancel

Log what the change is so you can find it later!

Click on the Commit Message

The screenshot shows the GitHub interface for a repository named 'test_project' owned by 'carriewright11'. The repository is public. The top navigation bar includes links for Code, Issues, Pull requests, Agents, Actions, Projects, Wiki, Security and quality, and Insights. Below the repository name, there are buttons for Pin and Watch (0). The main content area shows the file history for the 'main' branch. A table lists the files and their commit messages. The first entry is 'README.md' with the commit message 'Updating the README'. A pink arrow points to this commit message. Below the table, the 'README' file is displayed, showing the title 'test_project' and the content 'This test project is for practicing GitHub for the SISBID Data Wrangling class.'

carriewright11 / test_project

<> Code Issues Pull requests Agents Actions Projects Wiki Security and quality Insights

test_project Public

main 1 Branch 0 Tags

Go to file T Add file <> Code

carriewright11 Updating the README 3888eeb · now 6 Commits

| File | Commit Message | Time |
|-----------|---------------------|------|
| README.md | Updating the README | now |

README

test_project

This test project is for practicing GitHub for the SISBID Data Wrangling class.

See the change you made!

Commit 3888eeb

 Browse files

 **carriewright11** authored 15 minutes ago Verified

Updating the README

Adding some info about what the project is for

 **main**

1 parent [43fc695](#) commit 3888eeb 

1 file changed

+1 



Search within code



▼ README.md 

+1 



... @@ -1,2 +1,3 @@

1 # test_project

2

1 # test_project

2

3 + This test project is for practicing GitHub for the SISBID Data Wrangling class.

Comments 0

 Lock conversation

Old version!

New version!

Red means removed, Green means added

Commit 1ed4dd3

 Browse files



carriewright11 authored now

Verified

Rename project and update description

Updated project name and added a date reference.



main

1 parent [3888eeb](#) commit 1ed4dd3 

1 file changed

+2 -2  



Search within code



▼ README.md 

+2 -2  



... @@ -1,3 +1,3 @@

1 - # test_project

2

3 - This test project is for practicing GitHub for the SISBID Data Wrangling class.

1 + # My New Project

2

3 + This test project is for practicing GitHub for the SISBID Data Wrangling class
in July of 2026.

Old version!

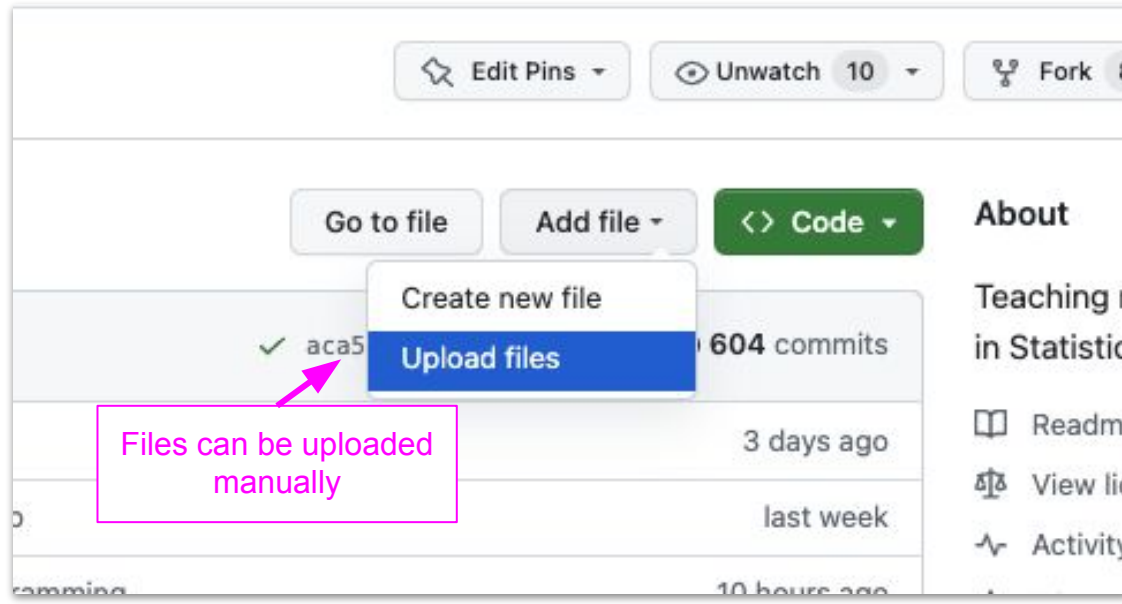
New version!

Step 3: Upload Changed File to GitHub

1. Make a change to the README file
2. "Add File"
3. "Upload Files"
4. Drag and drop / add files
5. Add commit message
6. Click "Commit changes"
7. Click on the commit message to see the change log

If nothing else...

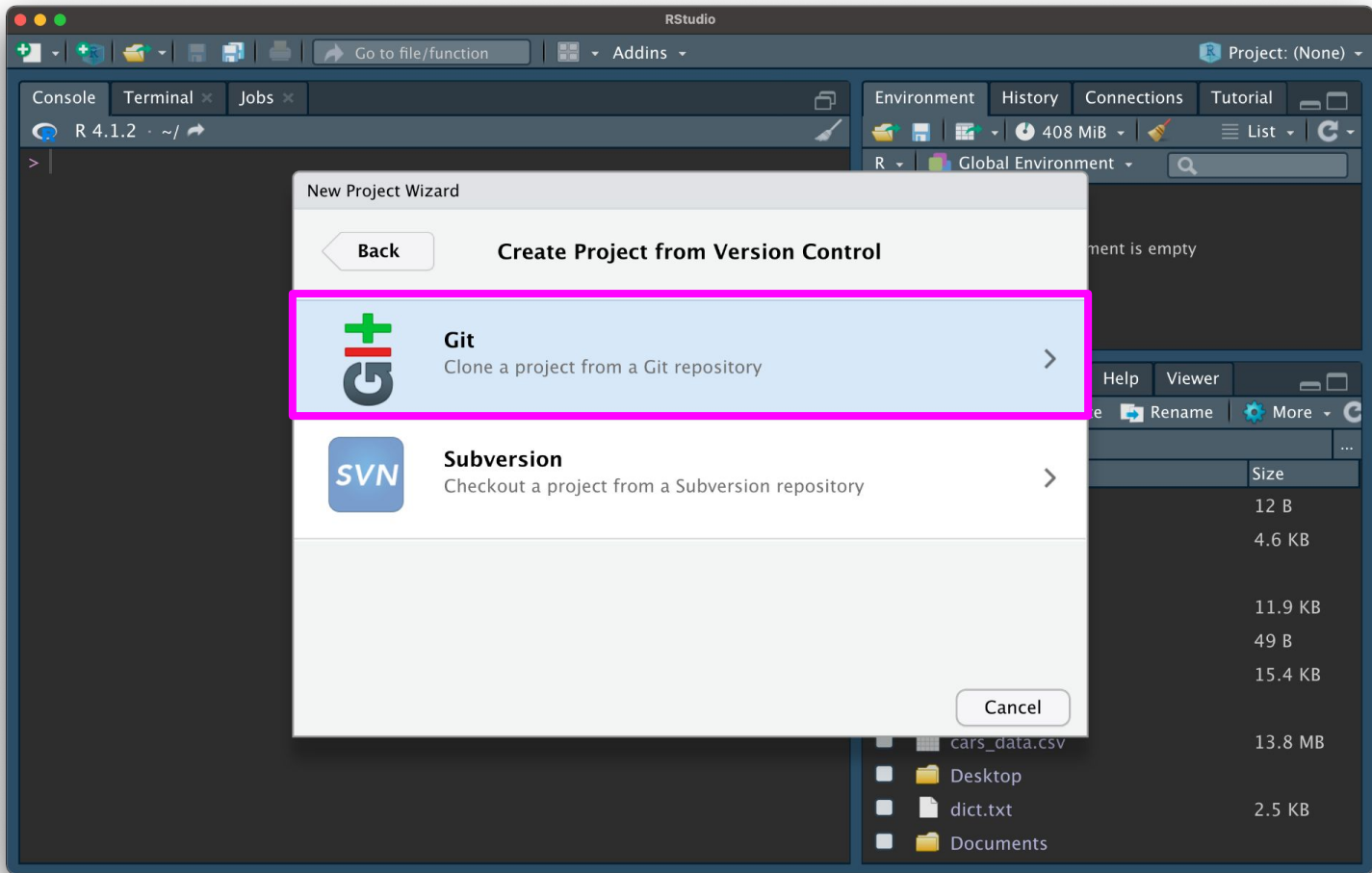
Use github as a place to store and share your code! You can add new versions of the same file name and add commit messages about what has changed - GitHub will show you the changes.

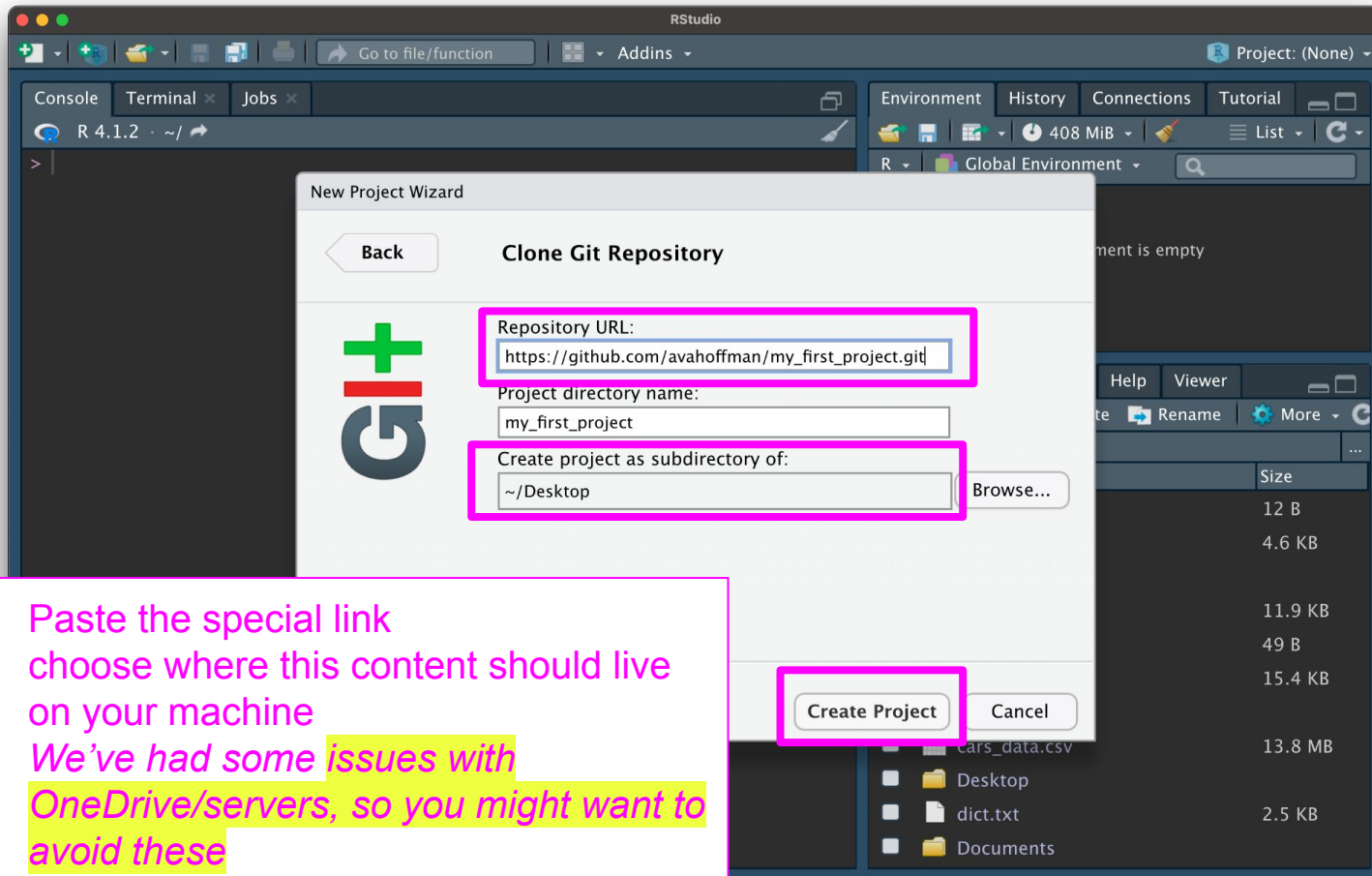


Local git usage

(optional and requires Git to be installed)







- Paste the special link
- choose where this content should live on your machine
- We've had some issues with OneDrive/servers, so you might want to avoid these

Make some changes...

Click Commit

Can commit
and push
changes to
GitHub
Directly!

The screenshot shows the RStudio IDE interface for a project named 'my_first_project'. The top toolbar includes a 'Commit' button, which is highlighted with a pink box. The left pane displays the 'README.md' file with the following content:

```
1 # my_first_project
2 My first project files
3
4 Our project members are: ...
```

The bottom pane shows the R console with the following text:

```
R 4.1.2 ~ /Desktop/my_first_project/
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 right pane shows the 'Git' tab with a table of staged files:

| Staged | Status | Path |
|-------------------------------------|--------|------------------------|
| ☐ | ? | my_first_project.Rproj |
| ☒ | M | README.md |
| ☐ | ? | .gitignore |

Below the table, the file explorer shows the project structure:

| Name | Size | M |
|------------------------|-------|---|
| .. | | |
| .gitignore | 40 B | F |
| my_first_project.Rproj | 205 B | F |
| README.md | 71 B | F |

Summary

To help make your work more reproducible:

- Use R Projects
- RMarkdown
- Clean your environment regularly
- Check the knit of your RMarkdown regularly
- Tell your future self and others what you did!
- Print session info
- Save your code on GitHub tracking changes - consider more doing it directly from RStudio!

Where to learn more...

Resources!

<https://guides.github.com/activities/hello-world/>

<https://git-school.github.io/visualizing-git/>

<https://rogerdudler.github.io/git-guide/>

<https://www.katacoda.com/courses/git>

<https://github.com/Meghansaha/please-let-me-merge?tab=readme-ov-file#resources-for-improving-git-knowledge>

More resources

These are just some quick tips, for more information:

- [Reproducibility in Cancer Informatics course](#)
- [Advanced Reproducibility in Cancer Informatics course](#)
- [The RMarkdown book](#)
- [Jenny Bryan's organizational strategies](#)
- [Write efficient R code for science](#)
- [Tools for reproducible workflows in R](#)

Rmarkdown lab

Download the file from

<https://github.com/SISBID/data-wrangling/raw/gh-pages/labs/rmarkdown-lab.Rmd>

More Advanced GitHub

Step 4: Create “local” link

You can work on files directly and then sync. We will show how to do this with RStudio, but you can also do this with Github desktop or command line.



Search or jump to...



Pulls

Issues

Marketplace

Explore



avahoffman / my_first_project

Unwatch

1

Star

0

Fork

0

Code

Issues

Pull requests

Actions

Projects

Wiki

Security

Copy the special git URL

main

Go to file

Add file

Code



avahoffman Initial commit



README.md

Initial commit

README.md

my_first_project

My first project files

Clone

HTTPS SSH GitHub CLI

https://github.com/avahoffman/my_first



Use Git or checkout with SVN using the web URL.



Open with GitHub Desktop



Download ZIP

About

My first project files

Readme

Releases

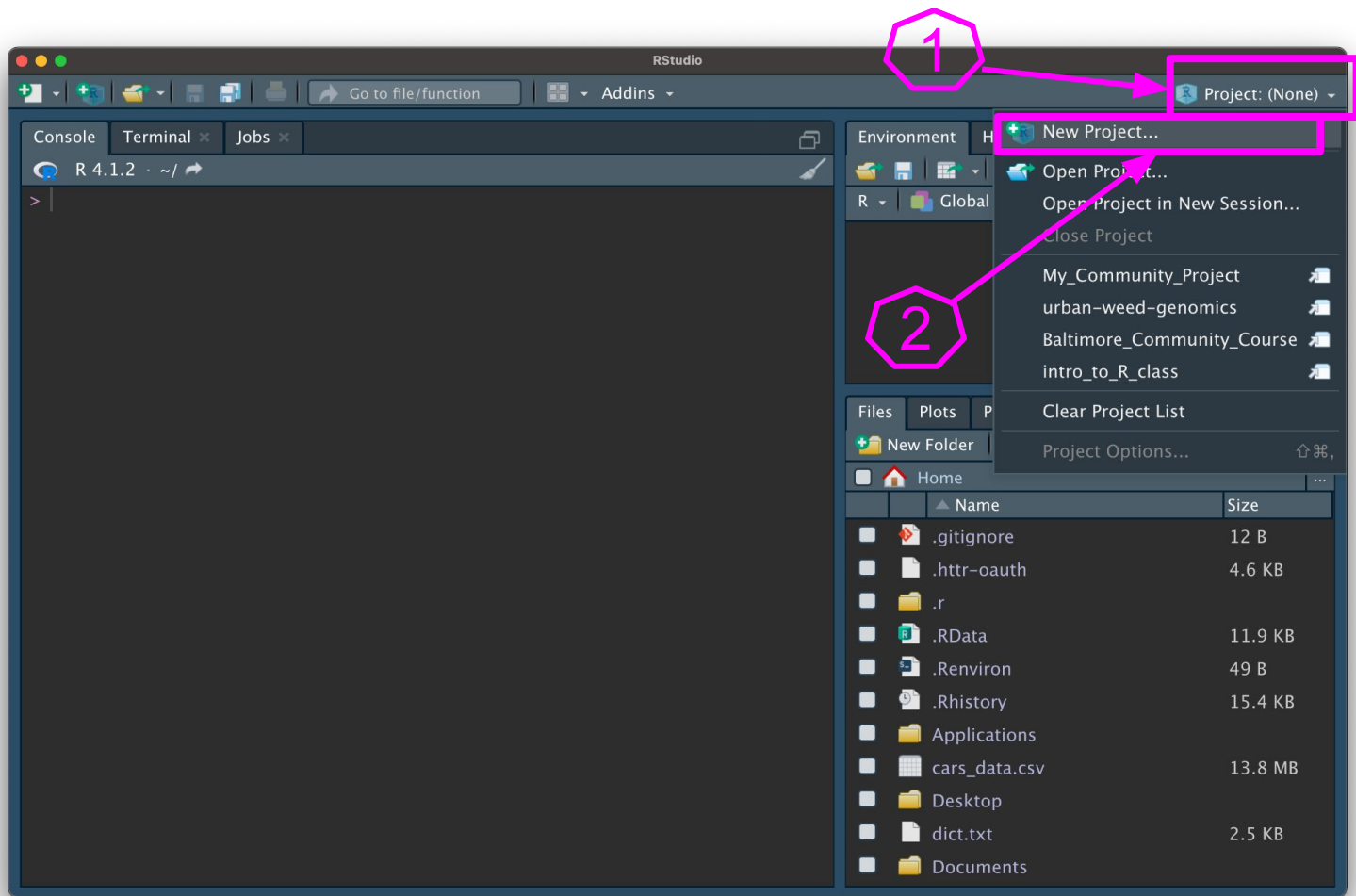
No releases published

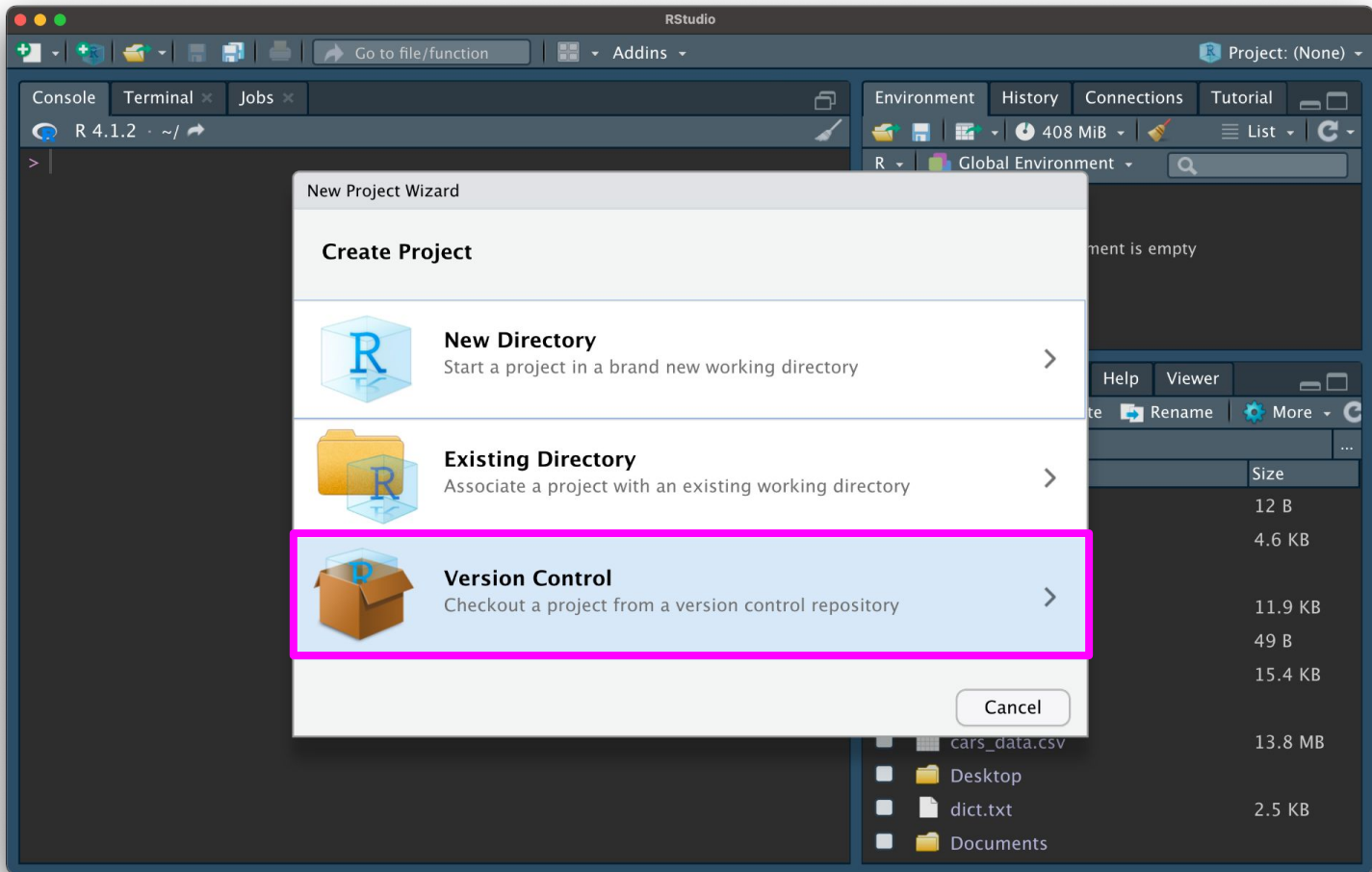
[Create a new release](#)

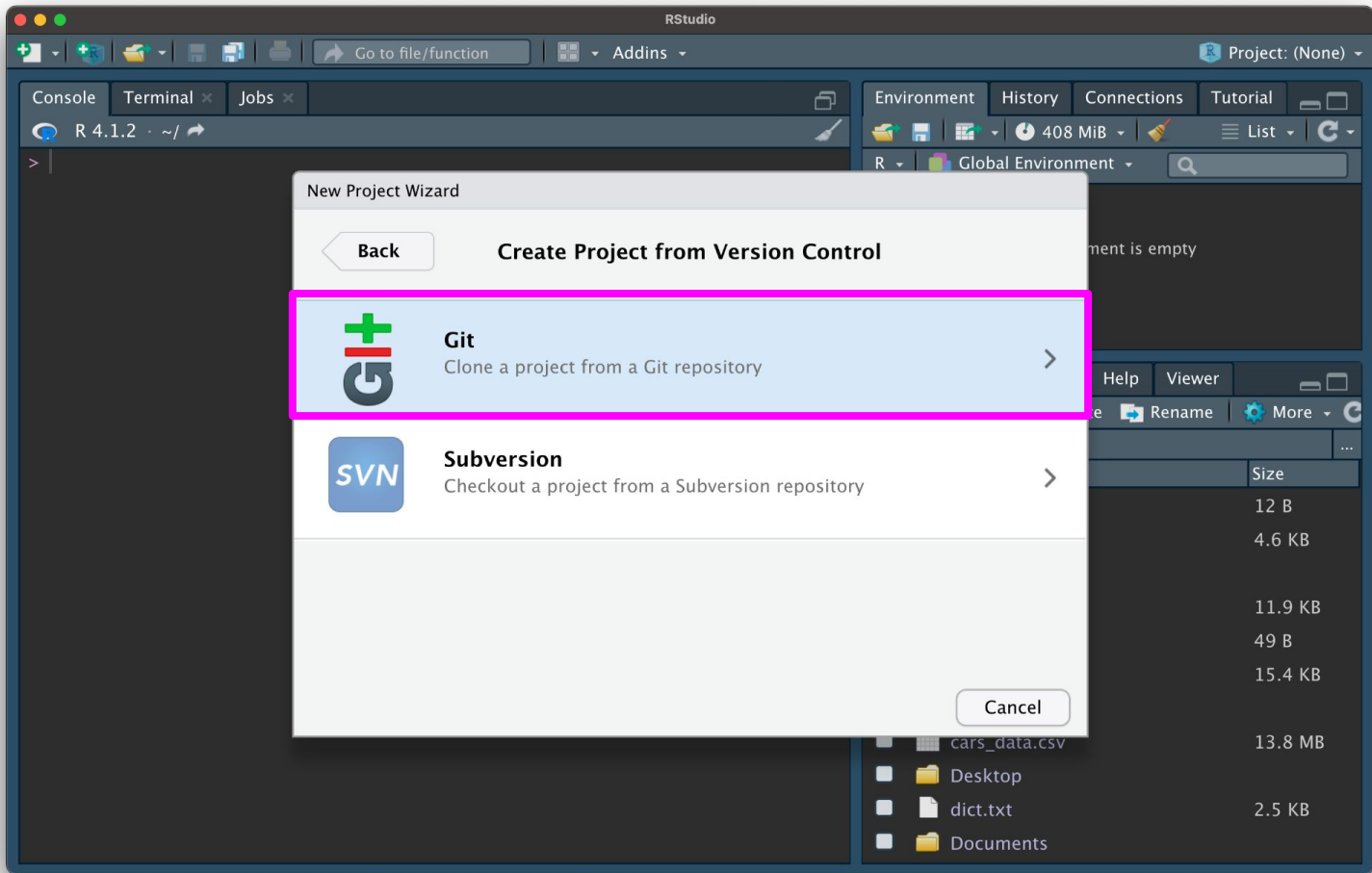
Packages

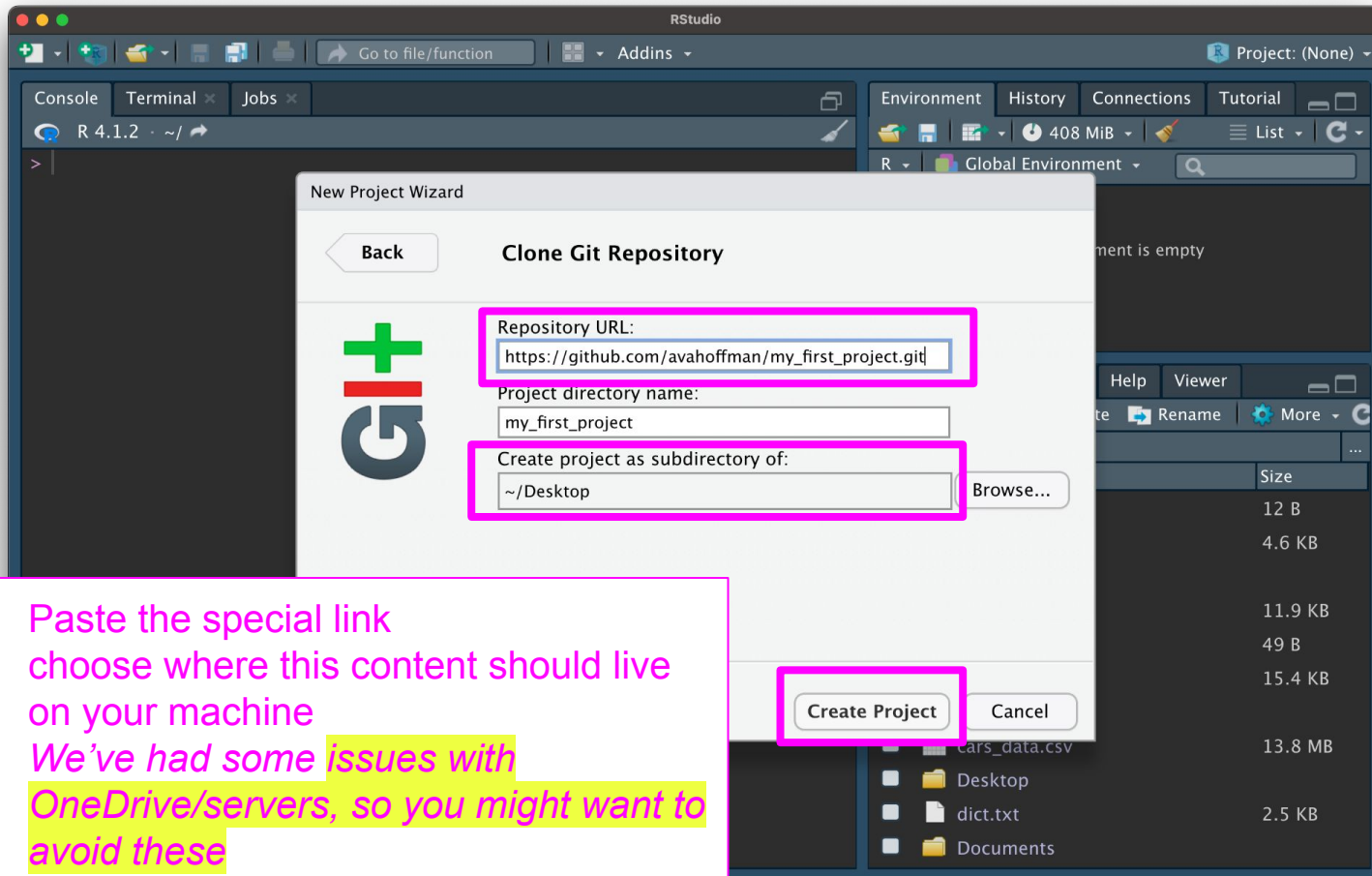
No packages published

[Publish your first package](#)





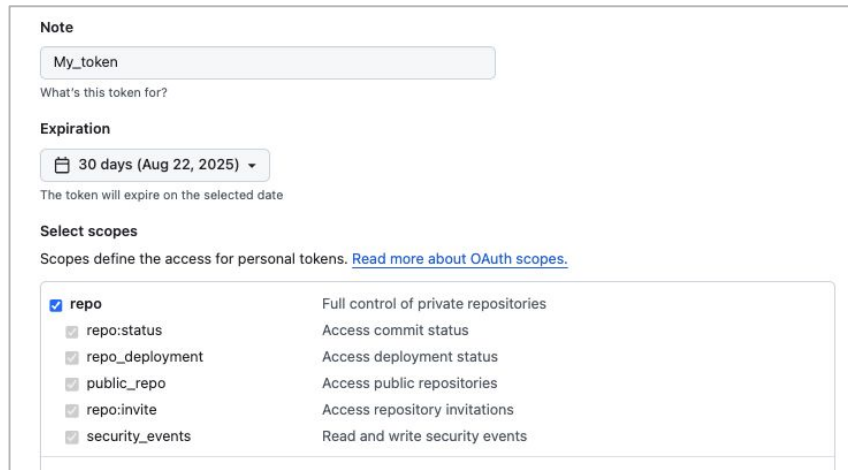




- Paste the special link
- choose where this content should live on your machine
- We've had some issues with OneDrive/servers, so you might want to avoid these

You might need to set a personal access token

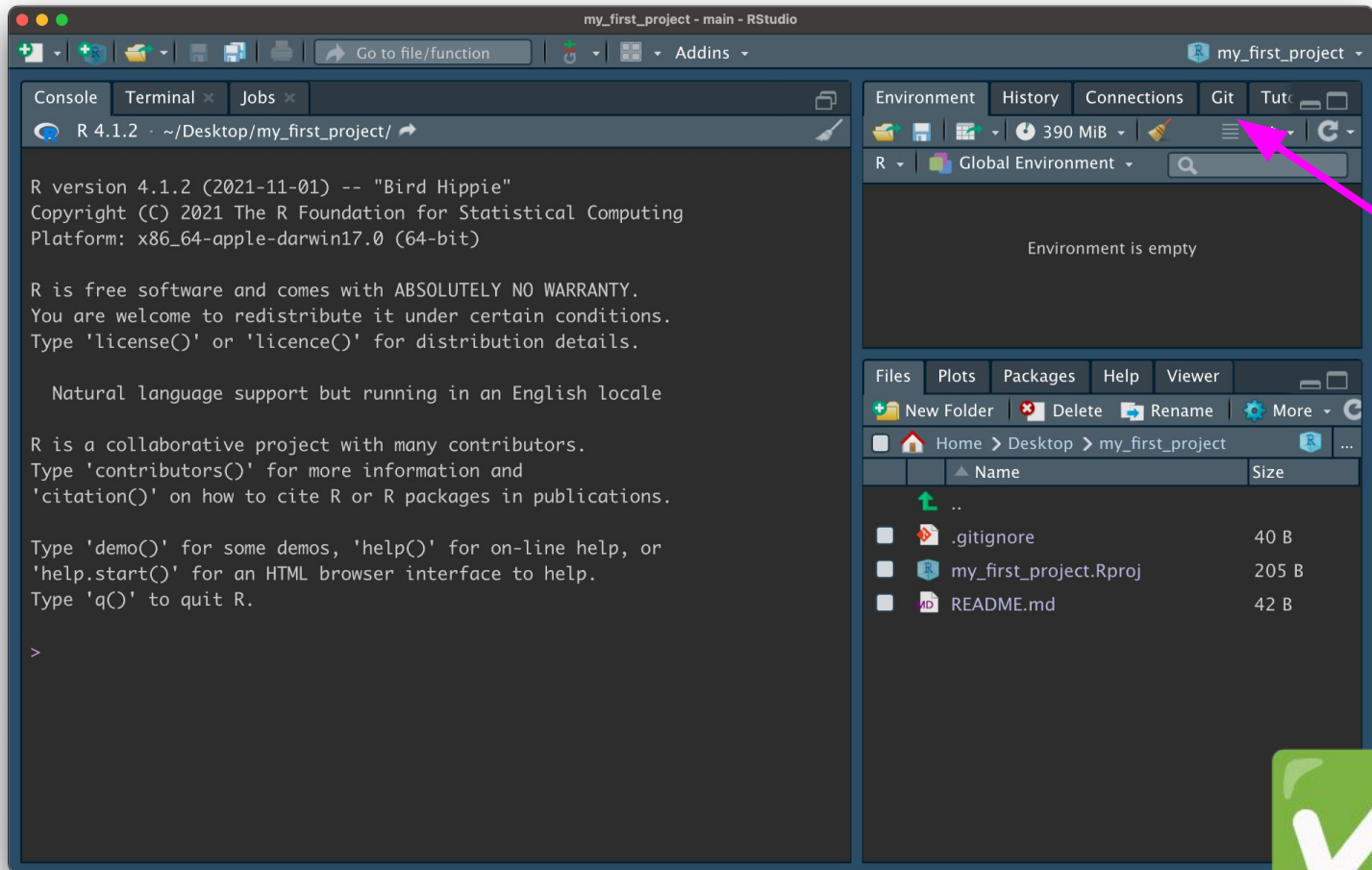
- Like a very strong password!
- Steps:
 - On Github, click your icon > Settings
 - Scroll down and click “Developer settings”
 - Click Personal access tokens > Tokens (classic)
 - Click Generate new token > Generate new token (classic)
 - Name the token under “Note”
 - Select the checkbox that says “repo”
 - Click “Generate token” at the bottom
 - Copy the token and save it somewhere safe
 - E.g., a password manager
 - Will start with “ghp_”
 - Paste the token In RStudio if prompted for GitHub password



The screenshot shows the GitHub 'Generate new token' interface. At the top, there's a 'Note' section with a text input field containing 'My_token'. Below this is a question 'What's this token for?'. The 'Expiration' section shows a date picker set to '30 days (Aug 22, 2025)'. A note states 'The token will expire on the selected date'. The 'Select scopes' section includes a link 'Read more about OAuth scopes.' and a list of scopes with checkboxes: 'repo' (checked), 'repo:status', 'repo_deployment', 'public_repo', 'repo:invite', and 'security_events'. Each scope has a corresponding description of its permissions.

| Scope | Description |
|--|--------------------------------------|
| ☒ repo | Full control of private repositories |
| ☐ repo:status | Access commit status |
| ☐ repo_deployment | Access deployment status |
| ☐ public_repo | Access public repositories |
| ☐ repo:invite | Access repository invitations |
| ☐ security_events | Read and write security events |

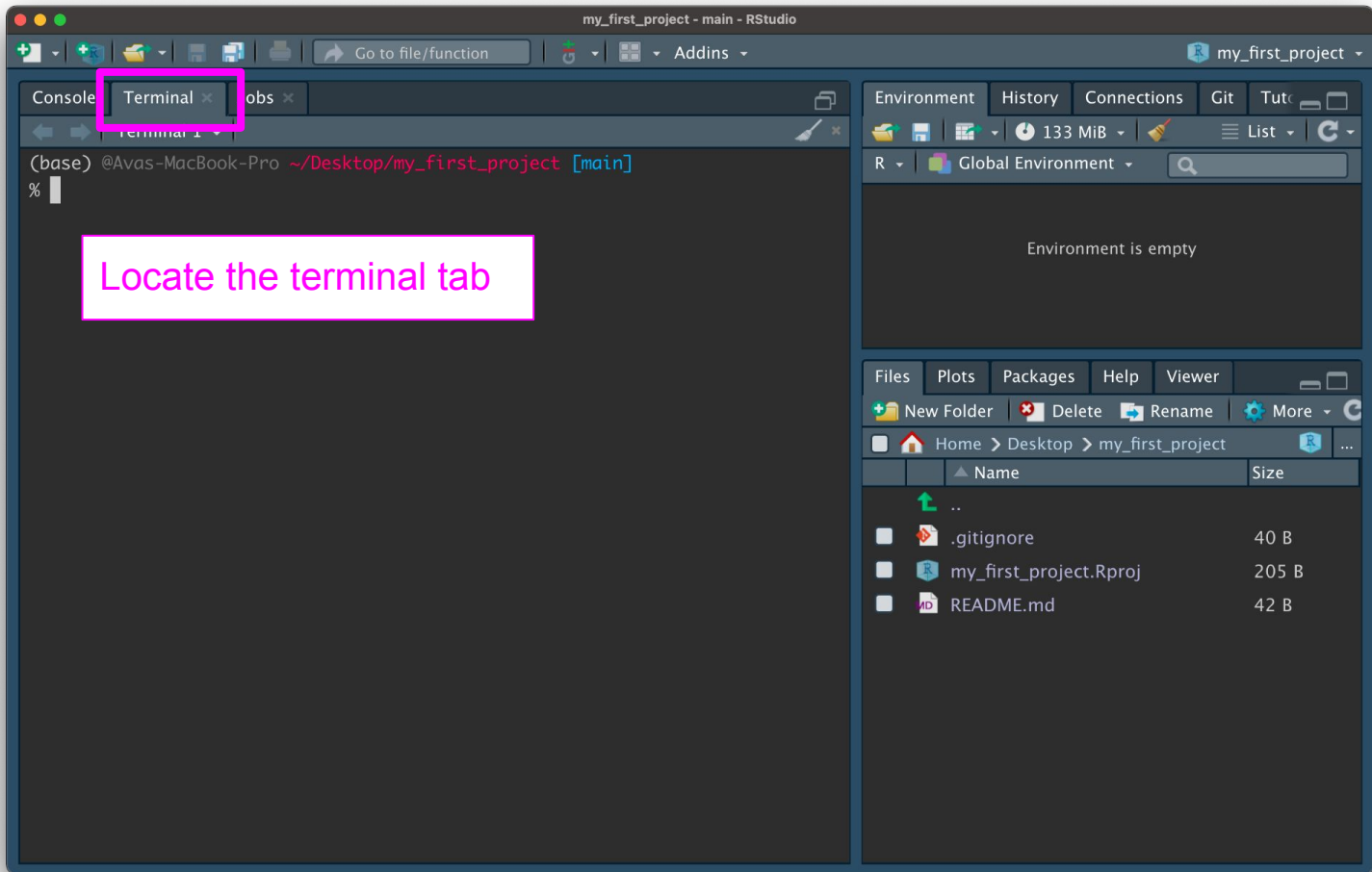
Can also set up your password for RStudio with the [usethis](#) package

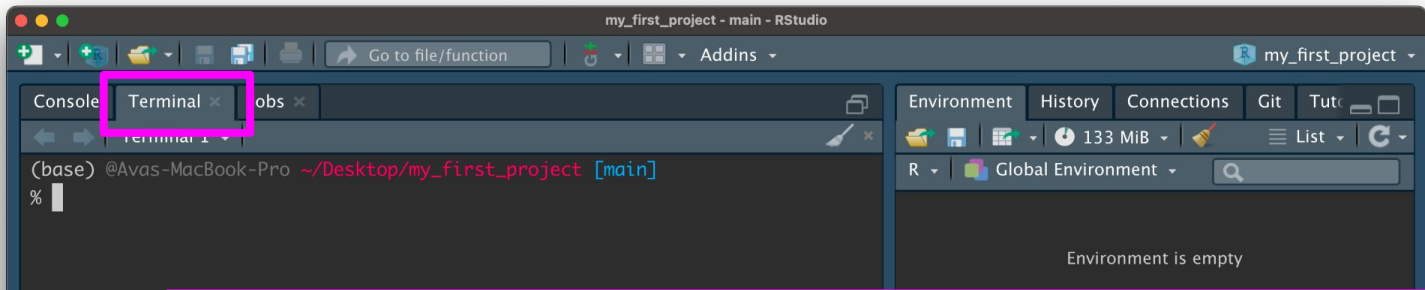


Step 5: Configure git

You should only need to do this once unless you want to use a different GitHub account.







Type the following..

For Mac:

```
git config --global credential.helper osxkeychain
git config --global user.name johndoe123
git config --global user.email johndoe@example.com
```

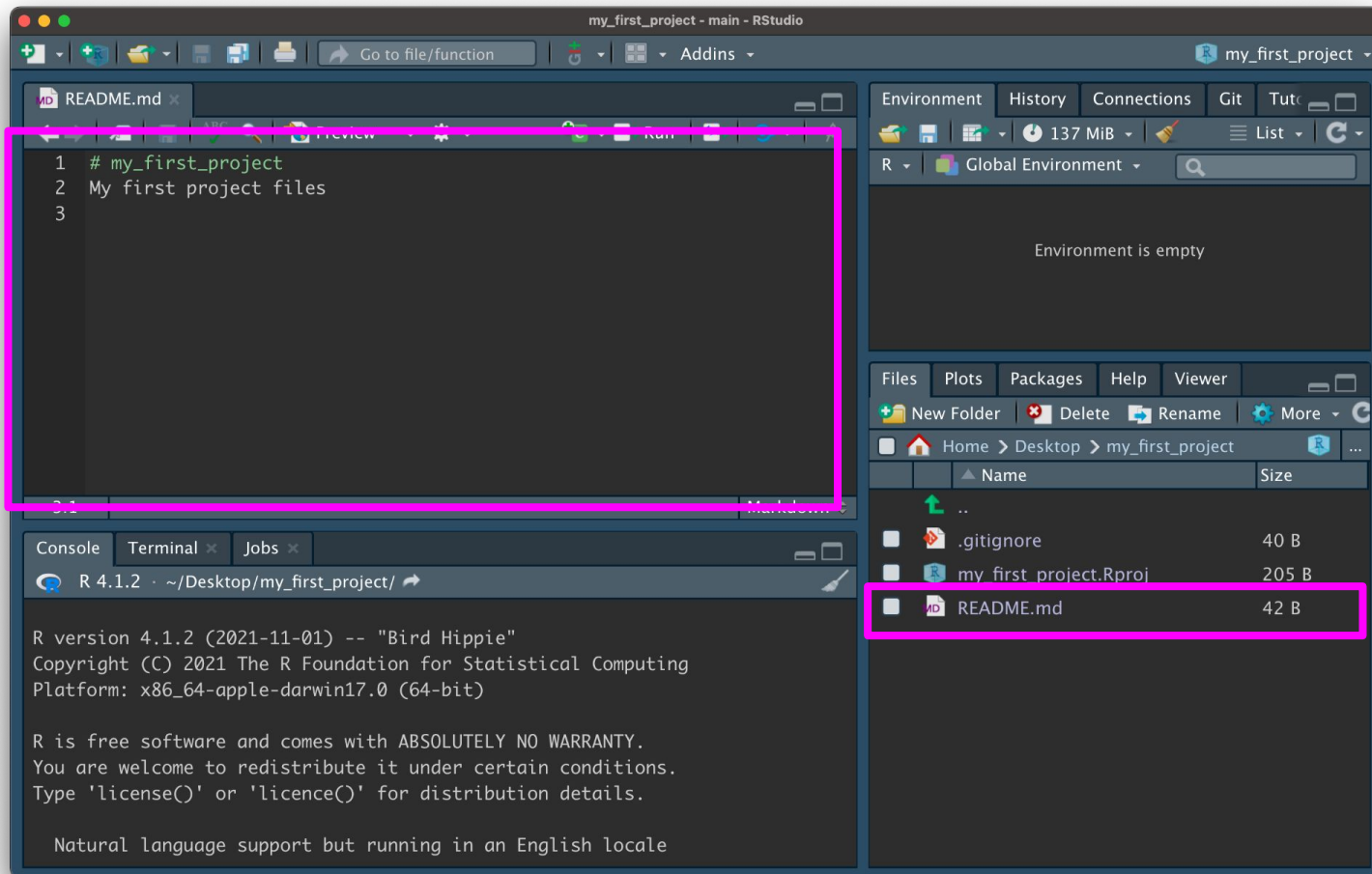
For Windows:

```
git config --global credential.helper manager-core
git config --global user.name johndoe123
git config --global user.email johndoe@example.com
```

Step 6: Use Git to sync changes from RStudio

Must have Git installed too!





Make some changes...

The screenshot shows the RStudio IDE interface for a project named "my_first_project". The main editor window displays the "README.md" file with the following content:

```
1 # my_first_project
2 My first project files
3
4 Our project members are: ...|
```

The status bar at the bottom of the editor shows the time "4:29" and the "Markdown" mode.

The right-hand pane contains two sections:

- Git** section: Shows the "Staged" status of files. The table below represents the data shown:

| Staged | Status | Path |
|-------------------------------------|--------|------------------------|
| ☐ | ? | my_first_project.Rproj |
| ☒ | M | README.md |
| ☐ | ? | .gitignore |

- Files** section: Shows a file explorer view of the project directory. The table below represents the data shown:

| Name | Size | M |
|------------------------|-------|---|
| .. | | |
| .gitignore | 40 B | F |
| my_first_project.Rproj | 205 B | F |
| README.md | 71 B | F |

The bottom pane shows the "Console" tab with the R prompt "R 4.1.2 · ~/Desktop/my_first_project/" and the following text:

```
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publicat
ions.

Type 'demo()' for some demos, 'help()' for on-line hel
p, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
```

Make some changes...

The screenshot shows the RStudio IDE interface for a project named 'my_first_project'. The top pane displays the 'README.md' file with the following content:

```
1 # my_first_project
2 My first project files
3
4 Our project members are: ...|
```

The bottom pane shows the 'Console' tab with the following output:

```
R 4.1.2 ~ /Desktop/my_first_project/
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 right-hand pane shows the 'Git' tab, which is highlighted. It displays a table of staged changes:

| Staged | Status | Path |
|-------------------------------------|--------|------------------|
| ☐ | ? | my_first_project |
| ☒ | M | README.md |
| ☐ | ? | .gitignore |

A callout box with the text 'Locate the Git tab' points to the 'Git' tab in the top right pane.

Make some changes...

Click Commit

The screenshot shows the RStudio IDE interface for a project named 'my_first_project'. The main editor window displays the 'README.md' file with the following content:

```
1 # my_first_project
2 My first project files
3
4 Our project members are: ...|
```

The bottom pane is split into a 'Console' and a 'Terminal' tab. The 'Console' tab is active, showing the R prompt and the following text:

```
R 4.1.2 ~ /Desktop/my_first_project/
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 right-hand pane is split into a 'Git' tab and a 'Files' tab. The 'Git' tab is active, showing the 'Commit' button highlighted with a red box. Below the 'Commit' button is a table of staged files:

| Staged | Status | Path |
|-------------------------------------|--------|------------------------|
| ☐ | ? | my_first_project.Rproj |
| ☒ | M | README.md |
| ☐ | ? | .gitignore |

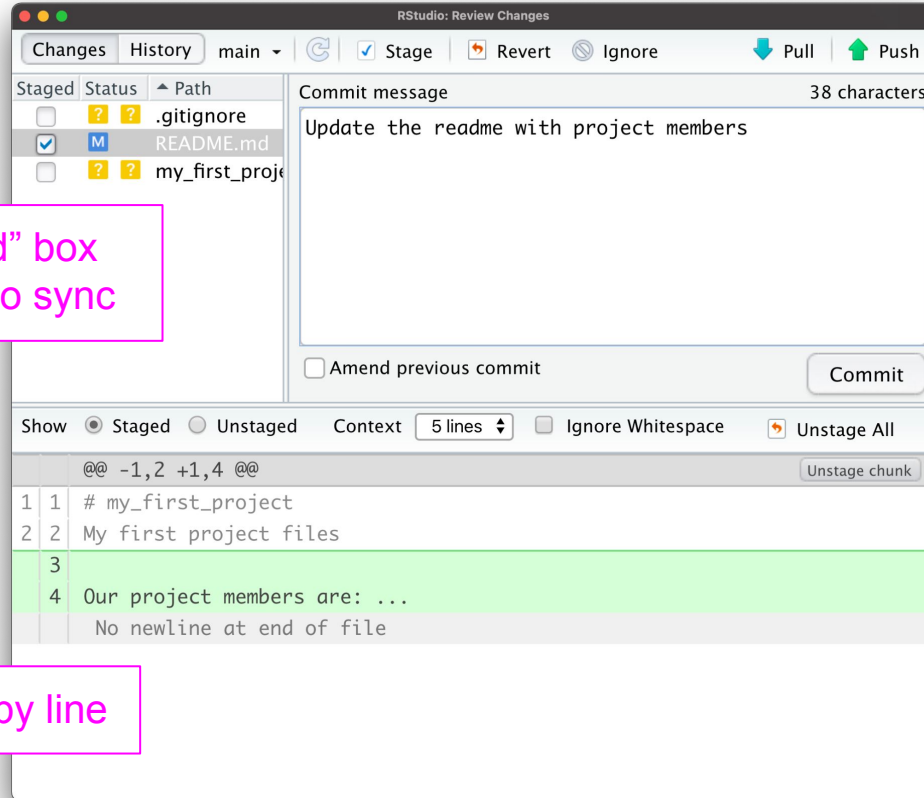
The 'Files' tab is also visible, showing a list of files in the project directory:

| Name | Size | M |
|------------------------|-------|---|
| .. | | |
| .gitignore | 40 B | F |
| my_first_project.Rproj | 205 B | F |
| README.md | 71 B | F |

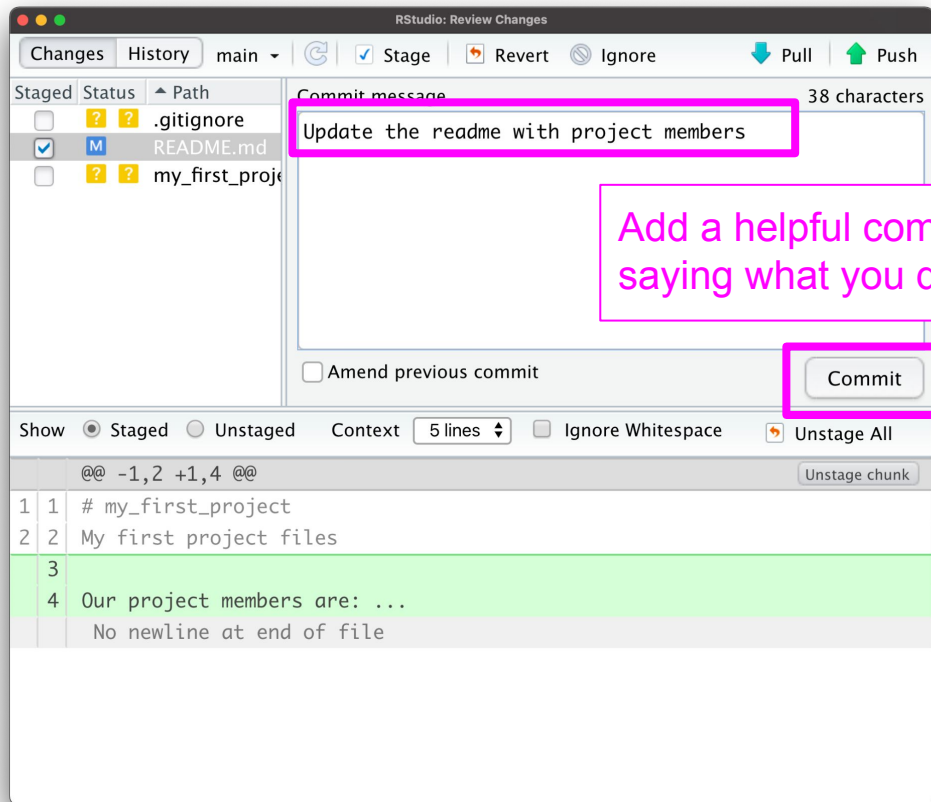
Make some changes...

Check the “Staged” box
for files you want to sync

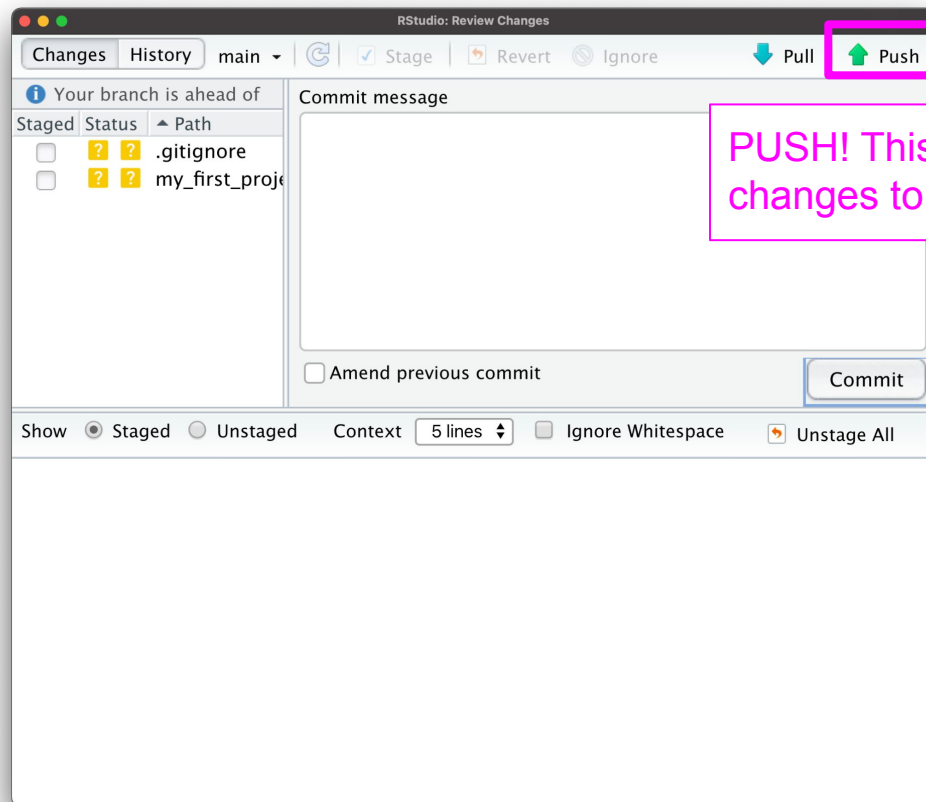
Observe changes by line



Make some changes...

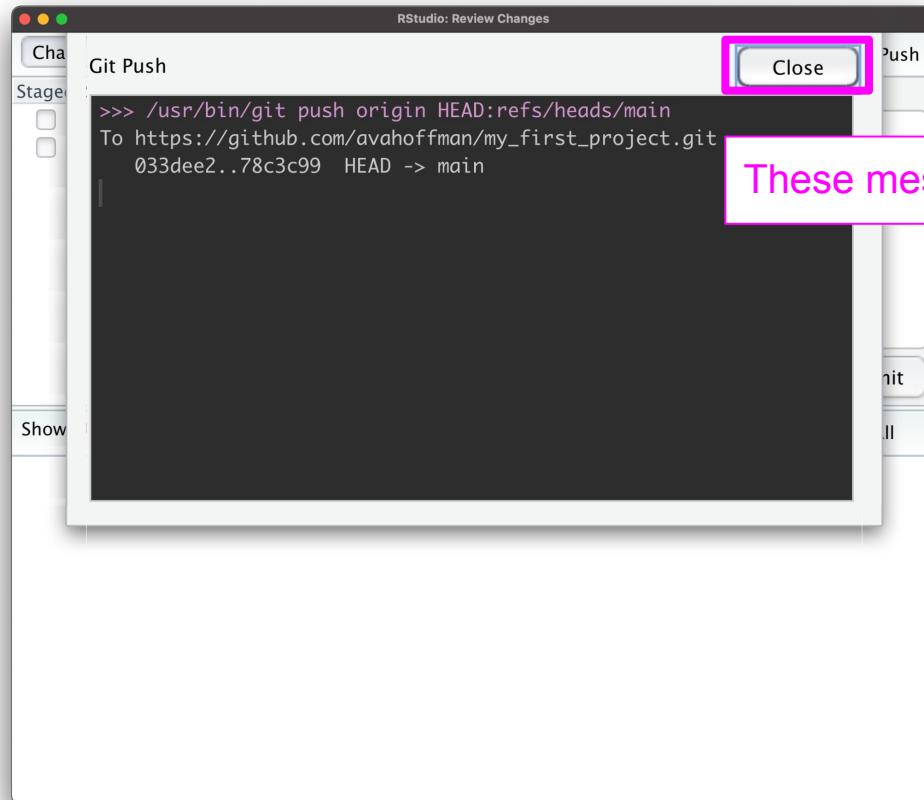


Make some changes...



PUSH! This sends your changes to GitHub

Make some changes...



RStudio: Review Changes

Git Push

```
>>> /usr/bin/git push origin HEAD:refs/heads/main  
To https://github.com/avahoffman/my_first_project.git  
033dee2..78c3c99 HEAD -> main
```

Close

These messages are FYI

You might be prompted for more information

```
/cloud/project$ git commit -m "Update Readme file"
```

```
*** Please tell me who you are.
```

Run

```
git config --global user.email "you@example.com"  
git config --global user.name "Your Name"
```

to set your account's default identity.

Omit `--global` to set the identity only in this repository.

```
fatal: unable to auto-detect email address (got 'rstudio-user@application-4375491-deployment-10694007-tknc5.(none)')  
/cloud/project$ git config user.email "avamariehoffman@gmail.com"  
/cloud/project$ git config user.name "Ava Hoffman"
```

Use your GitHub email and
display name / username

🔗 main ▾

🔗 1 branch

🏷 0 tags

Go to file

Add file ▾

Code ▾

About



My first project files

📖 Readme

★ 0 stars

👁 1 watching

🍴 0 forks

Releases

No releases published

[Create a new release](#)

Packages

No packages published



avahoffman Update the readme with project members

78c3c99 14 minutes ago

🕒 2 commits



README.md

Update the readme with project members

14 minutes ago

README.md



my_first_project

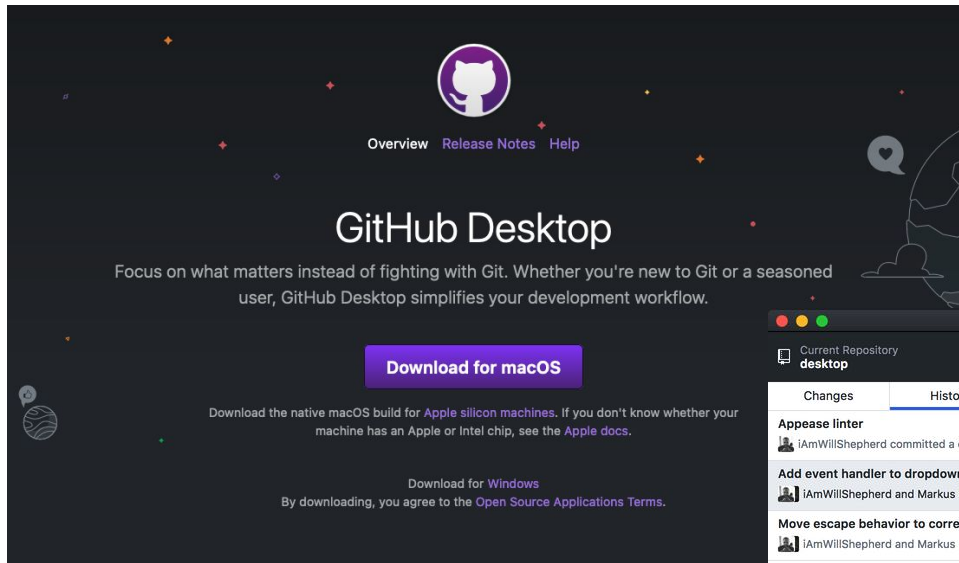
My first project files

Our project members are: ...

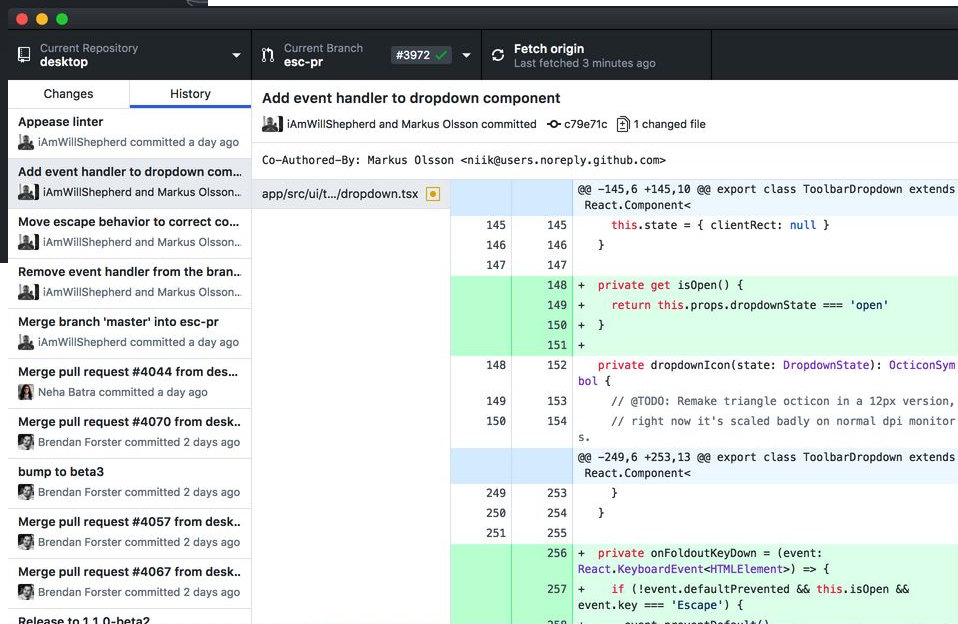
Refresh the webpage to see
the changes you made

Review

- You add/edit files on your personal computer (your local repository).
- You publish/share these changes to a remote repository (GitHub) so that others may see them/incorporate them into their own work.
- Steps:
 - Make changes - be sure to SAVE any files!
 - Stage changes
 - Commit
 - Push



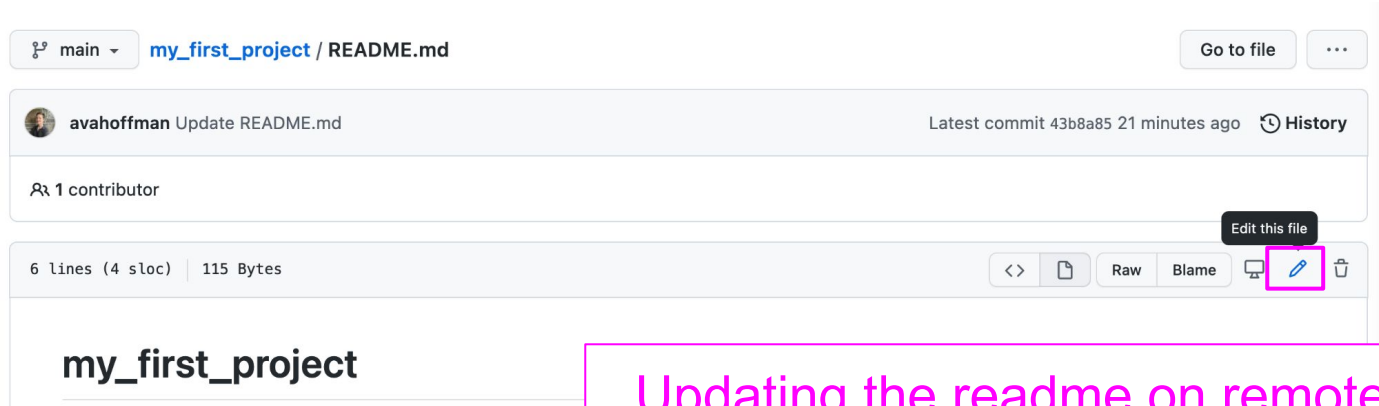
Other tools that can help



Step 7: Incorporate other's changes

If there are any :)





Updating the readme on remote
(on GitHub) will look a lot like
changes made by others

```
1 # my_first_project
2 My first project files
3
4 Our project members are: ...
5
6
7
```

Attach files by dragging & dropping, selecting or pasting them.



Commit changes

Update README.md

Add an optional extended description...

☒ Commit directly to the main branch.

☐ Create a new branch for this commit and start a pull request. [Learn more about pull requests.](#)

Commit changes

Cancel

Make changes, then scroll down and commit.

Meanwhile back in RStudio...

The screenshot shows the RStudio interface for a project named 'my_first_project'. The top-left pane displays a README.md file with the following content:

```
1 # my_first_project
2 My first project files
3
4 Our project members are: ...|
```

The top-right pane shows the Git interface. The 'Pull' button (a green downward arrow) is highlighted with a pink box. A pink callout box with the text 'Click Pull' points to this button. The 'Staged' column shows two files with yellow question marks.

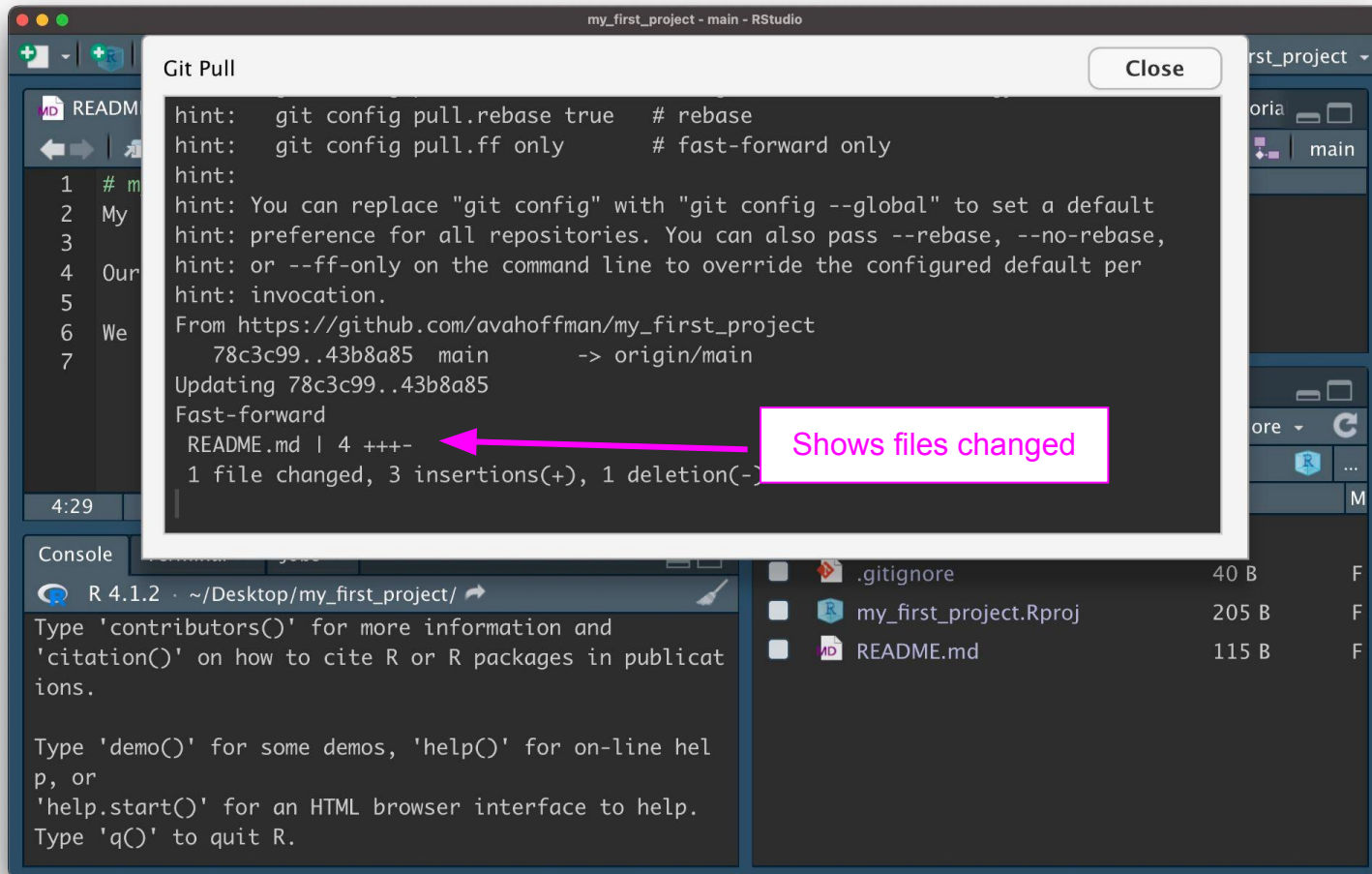
The bottom-left pane shows the Console with the following text:

```
R 4.1.2 ~ ~/Desktop/my_first_project/
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 bottom-right pane shows the Files view of the project directory. The table below lists the files and their sizes:

| Name | Size | M |
|------------------------|-------|---|
| .. | | |
| .gitignore | 40 B | F |
| my_first_project.Rproj | 205 B | F |
| README.md | 71 B | F |



my_first_project - main - RStudio

Go to file/function Addins my_first_project

README.md

```
1 # my_first_project
2 My first project files
3
4 Our project members are: ...|
5
6 We should add a project description here.
7
```

4:29 Markdown

Environment History Connections Git Tutoria

Diff Commit

| Staged | Status | Path |
|--------------------------|--------|------------------------|
| ☐ | ? | my_first_project.Rproj |
| ☐ | ? | .gitignore |

Changes are immediately visible

Files Plots Packag

New Folder Delete Rename More

Home > Desktop > my_first_project

| | Name | Size | M |
|--------------------------|------------------------|-------|---|
| | .. | | |
| ☐ | .gitignore | 40 B | F |
| ☐ | my_first_project.Rproj | 205 B | F |
| ☐ | README.md | 115 B | F |

Console Terminal Jobs

R 4.1.2 · ~/Desktop/my_first_project/

Type 'contributors()' for more information and 'citation()' on how to cite R or R packages in publications.

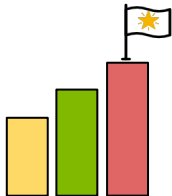
Type 'demo()' for some demos, 'help()' for on-line help, or 'help.start()' for an HTML browser interface to help. Type 'q()' to quit R.

Pulling

- Others have pushed changes to the remote repository.
- You want to incorporate those changes into your local repository.
- It's always a good idea to pull before you start working!
- Steps:
 - `Pull`

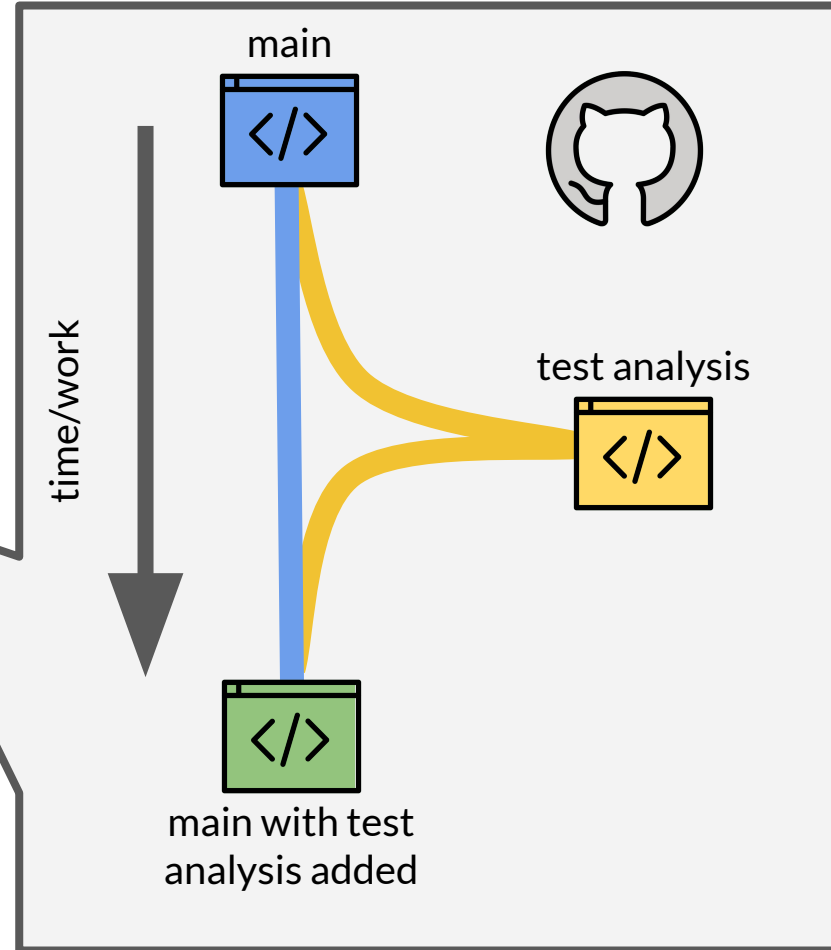
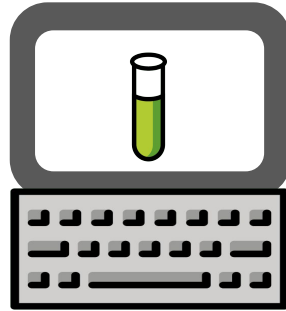
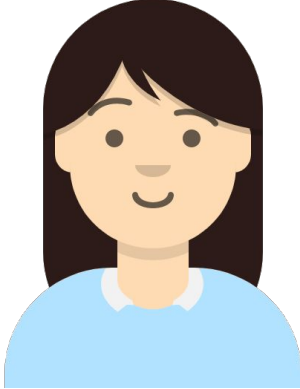
Step 8: Branches

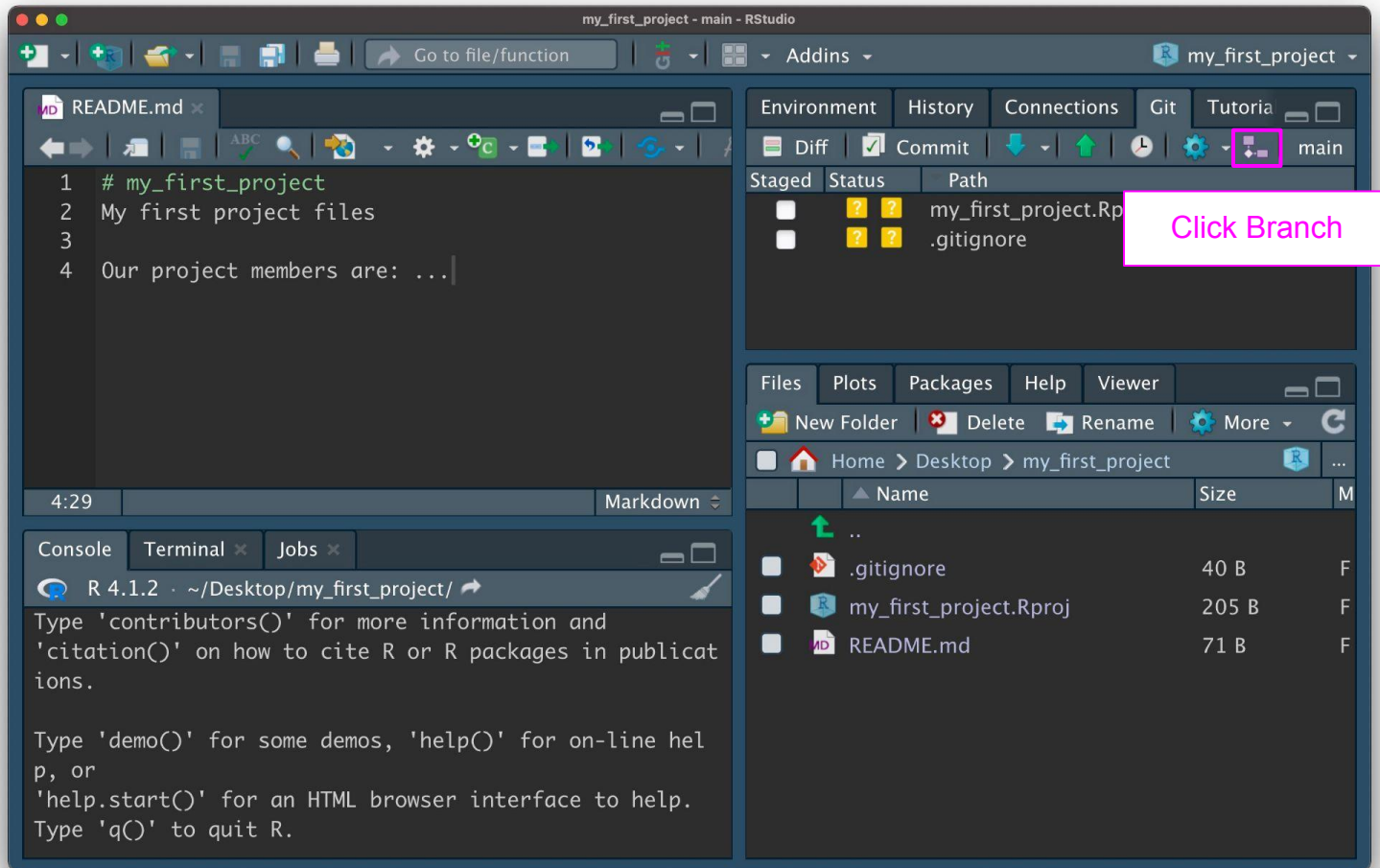
Develop new ideas safely on a branch

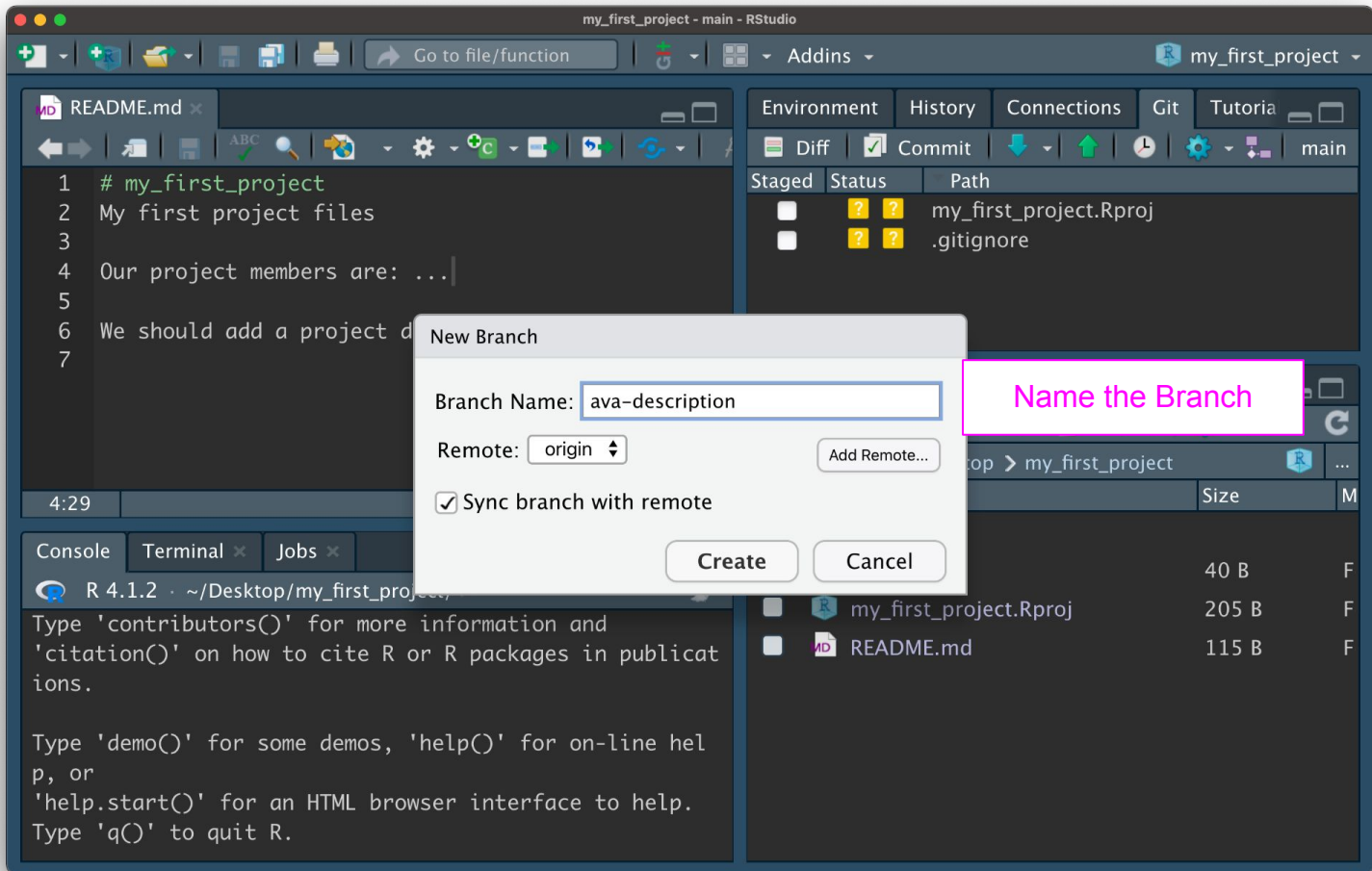


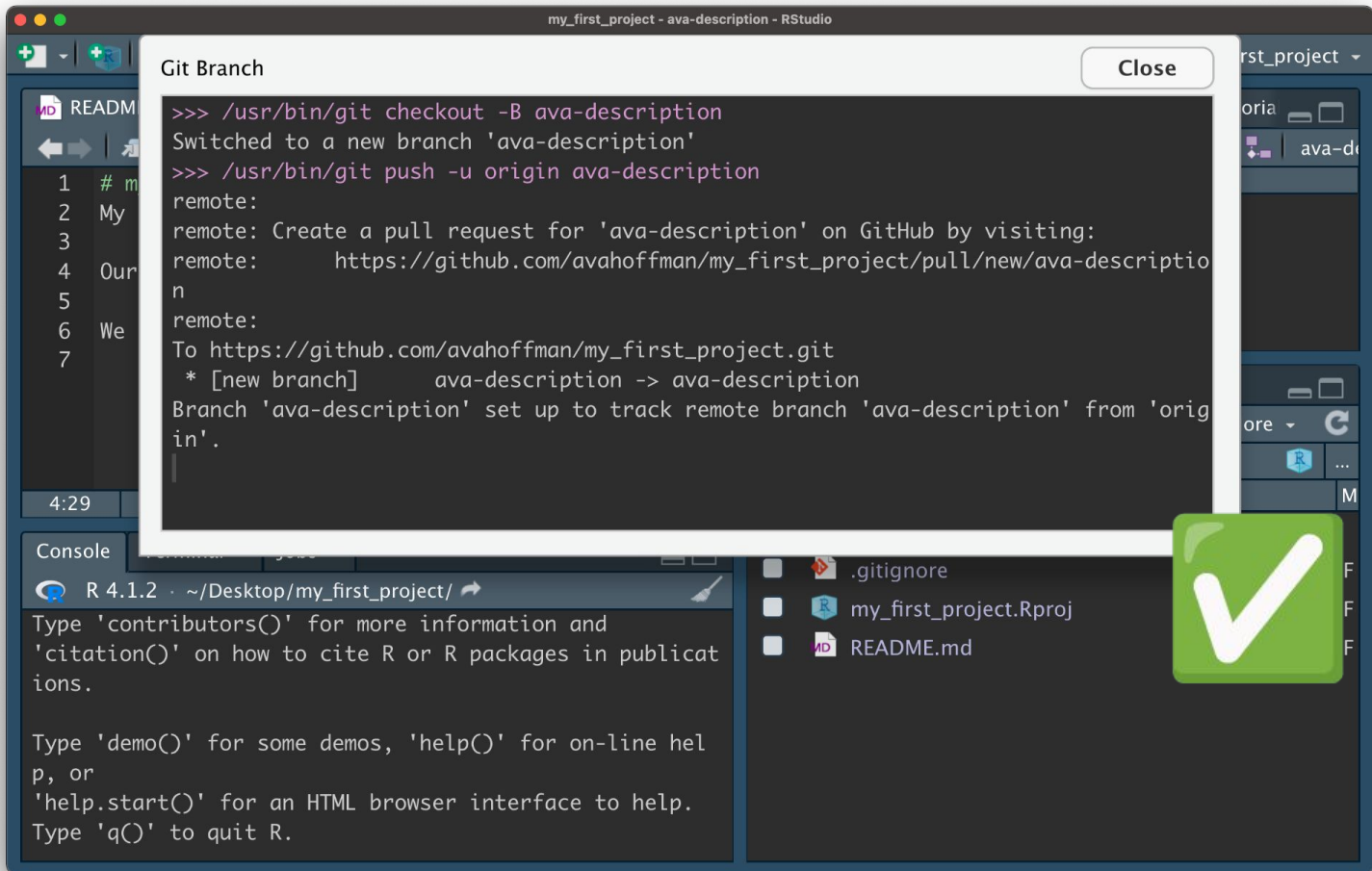
**Isolate what you are working on
from the “perfected” version!**

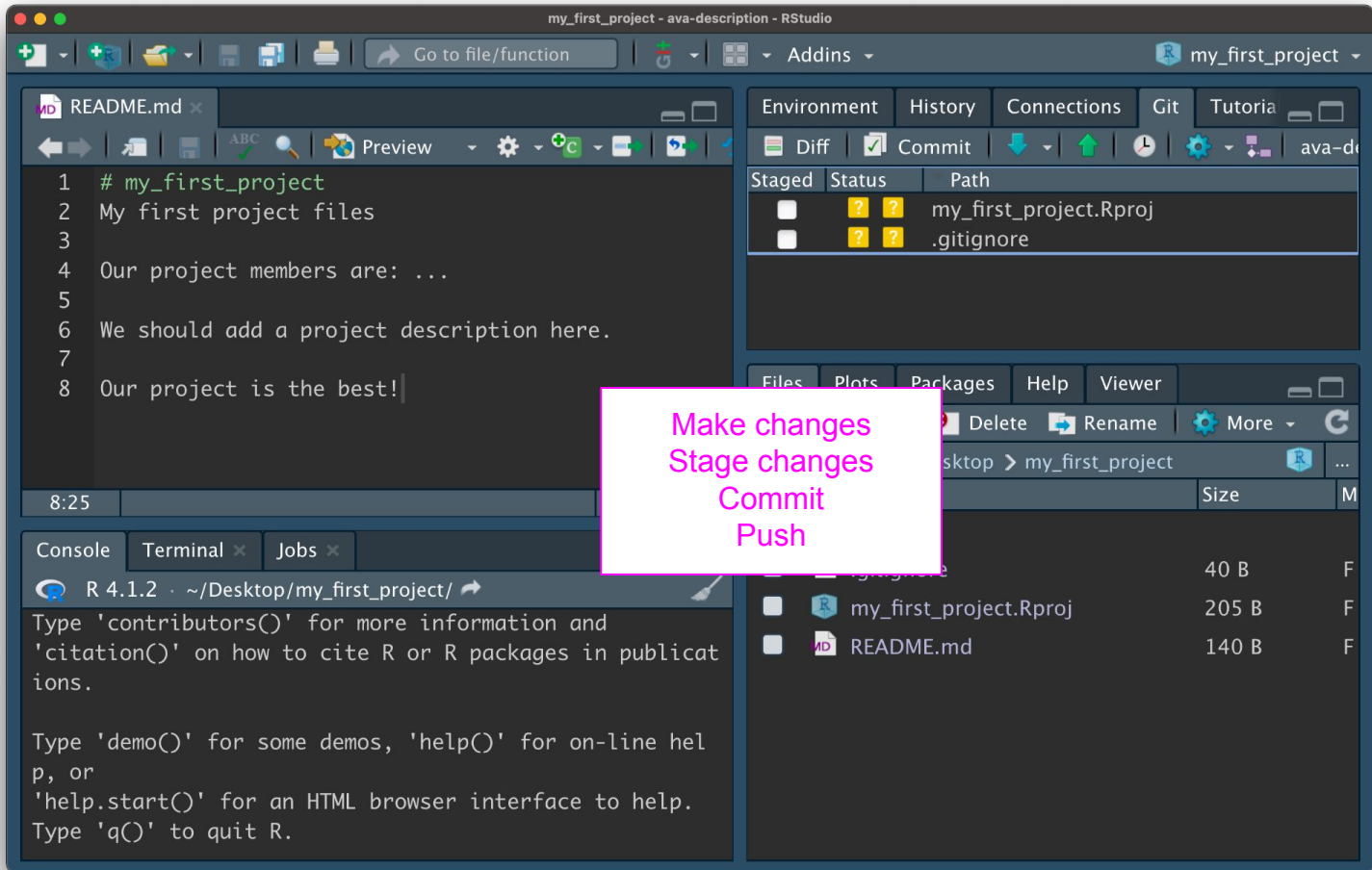
Image created by Candace Savonen using Avataars.











 **ava-description** had recent pushes 4 minutes ago [Compare & pull request](#)

 **main**  2 branches  0 tags [Go to file](#) [Add file](#) [Code](#)

 **avahoffman** Update README.md 43b8a85 34 minutes ago ⌚ 3 commits

 README.md Update README.md 34 minutes ago

README.md 

my_first_project

My first project files

Our project members are: ...

We should add a project description here.

 **ava-description** had recent pushes 5 minutes ago [Compare & pull request](#)

 **ava-description**  2 branches  0 tags [Go to file](#) [Add file](#) [Code](#)

This branch is 1 commit ahead of main. [Contribute](#)

 **avahoffman** ava's change 44d699f 5 minutes ago ⌚ 4 commits

 README.md ava's change 5 minutes ago

[README.md](#) 

my_first_project

My first project files

Our project members are: ...

We should add a project description here.

Our project is the best!

git cheat sheet

learn more about git the simple way at rogerdudler.github.io/git-guide/
cheat sheet created by Nara, jeschke of nimgit.com

create & clone

create new repository `git init`
clone local repository `git clone /path/to/repository`
clone remote repository `git clone username@host:/path/to/repository`

add & remove

add changes to INDEX `git add <filename>`
add all changes to INDEX `git add *`
remove/delete `git rm <filename>`

commit & synchronize

commit changes `git commit -m "Commit message"`
push changes to remote repository `git push origin master`
connect local repository to remote repository `git remote add origin <server>`
update local repository with remote changes `git pull`

branches

create new branch `git checkout -b <branch>`
e.g. `git checkout -b feature_x`
switch to master branch `git checkout master`
delete branch `git branch -d <branch>`
push branch to remote repository `git push origin <branch>`

merge

merge changes from another branch `git merge <branch>`
view changes between two branches `git diff <source_branch> <target_branch>`
e.g. `git diff feature_x feature_y`

tagging

create tag `git tag <tag> <commit ID>`
e.g. `git tag 1.0.0 1b2e1d63ff`
get commit IDs `git log`

restore

replace working copy with latest from HEAD `git checkout -- <filename>`



Tip

Want a simple but powerful
git-client for your mac?
Try Tower: www.git-tower.com/

https://rogerdudler.github.io/git-guide/files/git_cheat_sheet.pdf

Basic use

- ▶ Change some files
- ▶ See what you've changed

```
git status  
git diff  
git log
```

- ▶ Indicate what changes to save

```
git add
```

- ▶ Commit to those changes

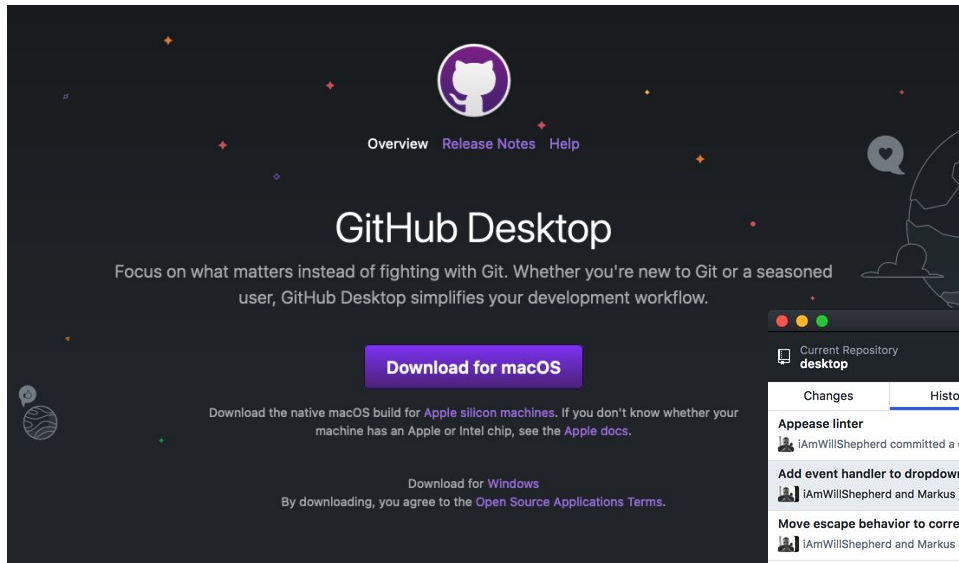
```
git commit
```

- ▶ Push the changes to GitHub

```
git push
```

- ▶ Pull changes from your collaborator

```
git pull
```



<https://desktop.github.com/>

Current Repository
desktop

Current Branch
esc-pr #3972 ✓

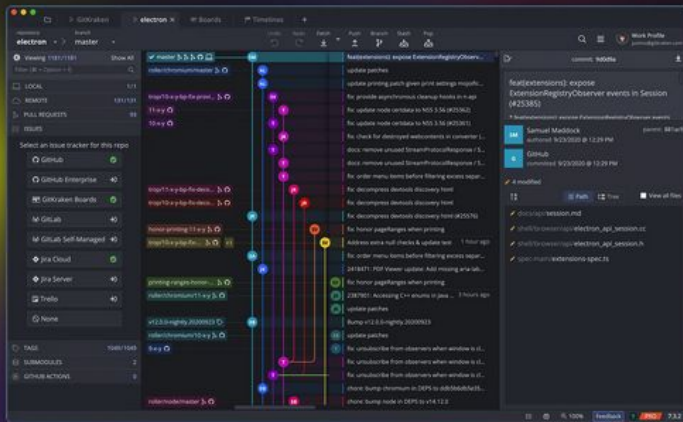
Fetch origin
Last fetched 3 minutes ago

| Changes | History |
|--|--|
| Appease linter
iAmWillShepherd committed a day ago | |
| Add event handler to dropdown component
iAmWillShepherd and Markus Olsson committed 1 changed file | |
| Co-Authored-By: Markus Olsson <nirik@users.noreply.github.com> | |
| app/src/ui/t.../dropdown.tsx | <pre>@@ -145,6 +145,10 @@ export class ToolbarDropdown extends React.Component< 145 this.state = { clientRect: null } 146 } 147 148 + private get isOpen() { 149 + return this.props.dropdownState === 'open' 150 + } 151 + 152 private dropdownIcon(state: DropdownState): OcticonSym 153 bol { 154 // @TODO: Remake triangle octicon in a 12px version, 155 // right now it's scaled badly on normal dpi monitor 156 s. 157 158 @@ -249,6 +253,13 @@ export class ToolbarDropdown extends React.Component< 249 } 250 } 251 252 + private onFoldoutKeyDown = (event: 253 React.KeyboardEvent<HTMLElement>) => { 254 + if (!event.defaultPrevented && this.isOpen && 255 event.key === 'Escape') { 256 event.preventDefault() 257 } 258 }</pre> |
| Move escape behavior to correct co...
iAmWillShepherd and Markus Olsson... | |
| Remove event handler from the bran...
iAmWillShepherd and Markus Olsson... | |
| Merge branch 'master' into esc-pr
iAmWillShepherd committed a day ago | |
| Merge pull request #4044 from des...
Neha Batra committed a day ago | |
| Merge pull request #4070 from desk..
Brendan Forster committed 2 days ago | |
| bump to beta3
Brendan Forster committed 2 days ago | |
| Merge pull request #4057 from desk..
Brendan Forster committed 2 days ago | |
| Merge pull request #4067 from desk..
Brendan Forster committed 2 days ago | |
| Release to 1.1.0-beta2 | |



GitKraken

Legendary Git GUI client for Windows, Mac & Linux



 Download the #1 Free Git GUI

Other Platforms

See Features →

<https://www.gitkraken.com/>