

# Lecture 2: Introduction to R Markdown and GitHub

**Heather Mattie, PhD**

**September 4, 2025**



# Recipe of the Day!

---



## Apple cheddar soup



# Office Hours and Lab

---

## Office Hours

Day	Time	Staff	Location
Monday	10-11am	Sajia	FXB G10
Tuesday	1-2pm	Heather	Building 1, 4 <sup>th</sup> floor, Room 421A
Wednesday	2:30-3:30pm	Carmen	Kresge 201 except on 10/1 where it will be held in Kresge 502
Thursday	12-1pm	Claire	Kresge 205

## Lab

Day	Time	Location
Friday	TBD	Zoom

**\*Lab will not be held every week**

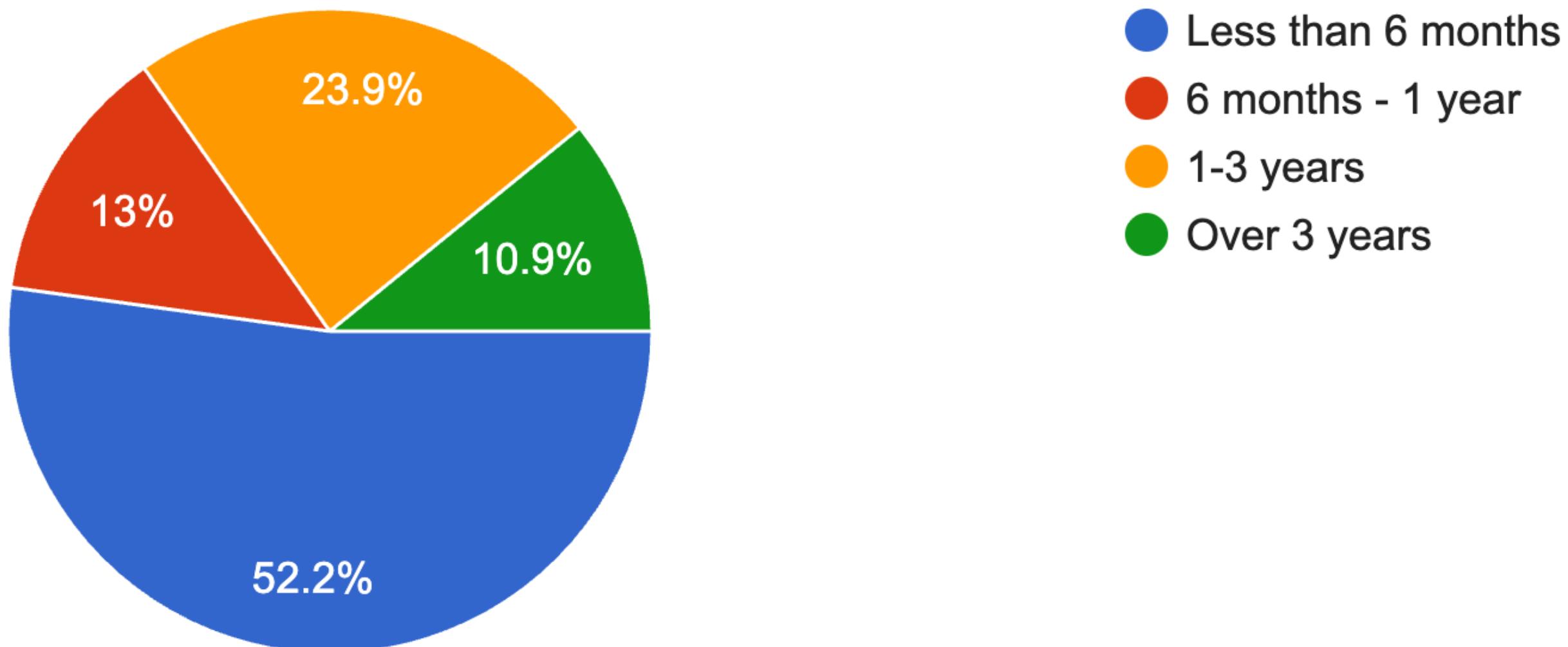


# Preliminary survey results

---

How long have you been programming?

46 responses

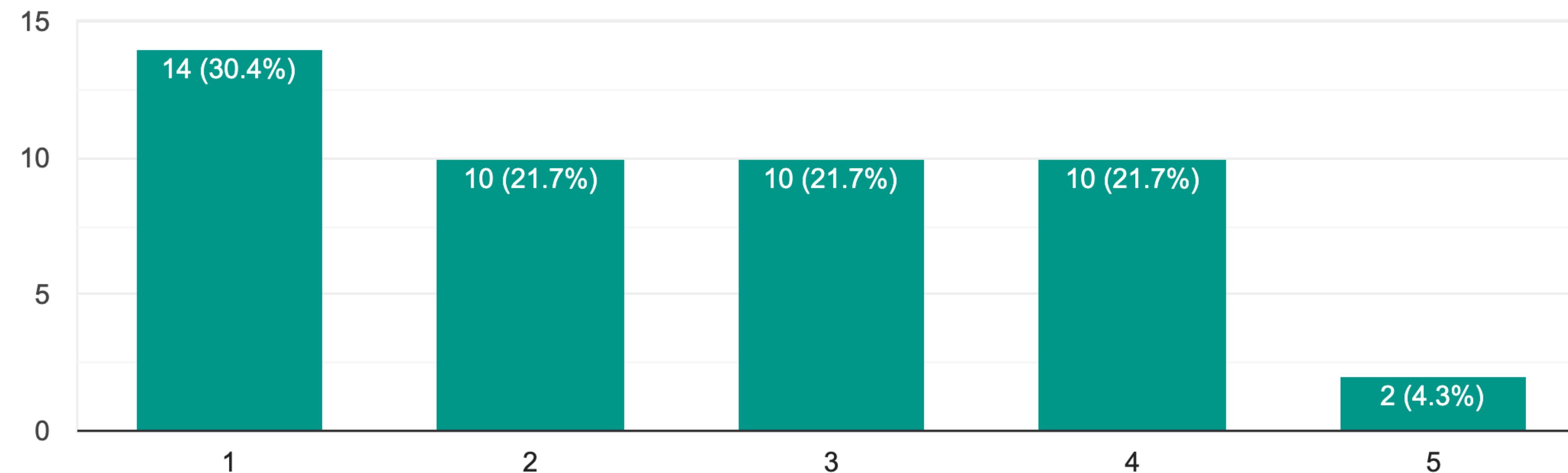


# Preliminary survey results

---

Overall, how comfortable are you with programming?

46 responses



# Preliminary survey results

---

What are you hoping to get out of this course?

- To begin to **demystify** coding/build a foundation of knowledge
- Learning R and becoming **comfortable** coding it
- **Visualization** of data
- **Machine learning**
- Skills in **reproducibility** and **collaborating** on data science projects

# Preliminary survey results

---

What are you hoping to get out of this course?

- I am hoping to be more **comfortable** with programming in general. Right now it **scares** me a lot.
- Gain a better **foundation** in applying coding to different public health questions and being able to **translate those skills to my thesis research**
- A better understanding of how to interact with data science from a **leadership** capacity. I may not be the coding, but I will need to know who and how to partner with the right folks to make the **data come alive**.

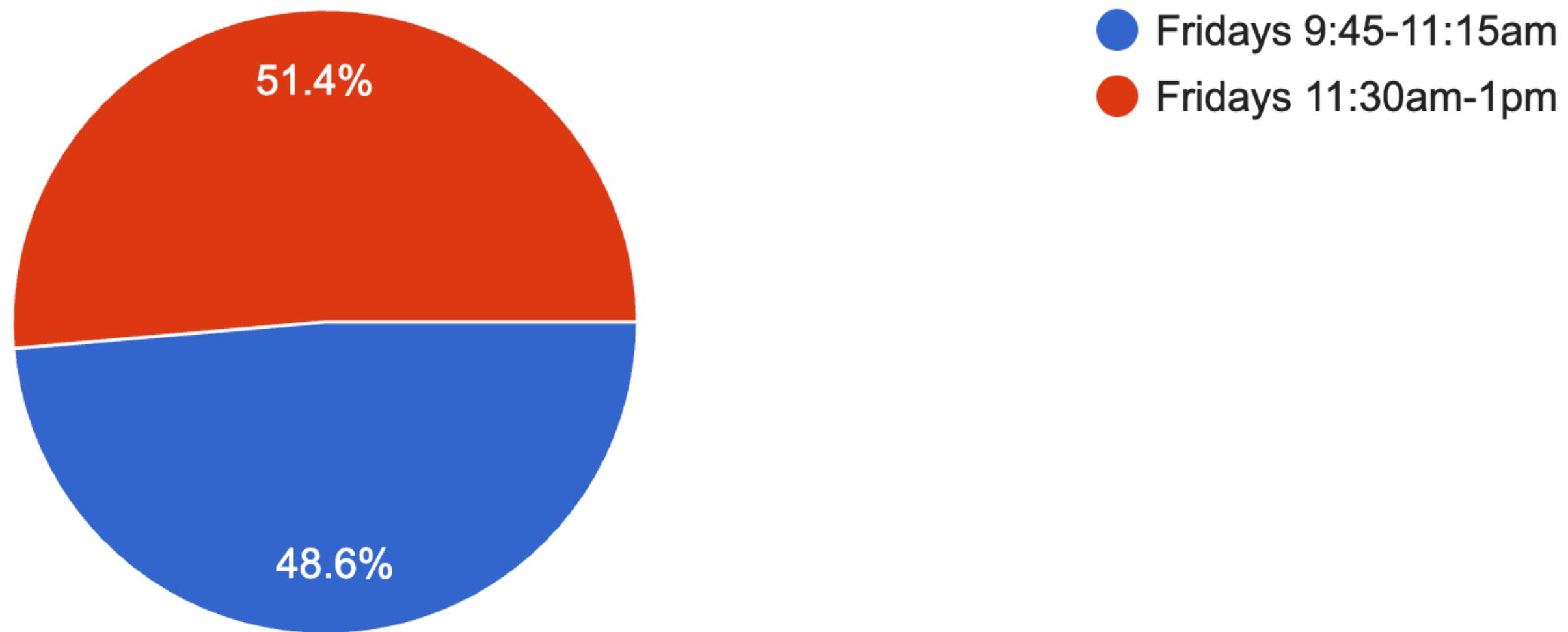


# Preliminary survey results

---

Which time would you prefer for the lab session? Lab will be on Zoom and will be recorded.

37 responses



# Introduction to R Markdown



# What is R Markdown?

- **R Markdown** is a tool for creating dynamic documents that **combine** text, code, and the output of that code into a **single** file



The screenshot illustrates the R Markdown workflow within RStudio:

- SOURCE EDITOR:** Shows the R Markdown file (`report.Rmd`) with the following content:

```
1 ---  
2 title: "Document Title"  
3 author: "Author Name"  
4 output:  
5   html_document:  
6     toc: TRUE  
7 ---  
8  
9 ``{r setup, include=FALSE}  
10 knitr::opts_chunk$set(echo = TRUE)  
11 ``  
12  
13 ## R Markdown  
14  
15 This is an R Markdown document.  
16 Markdown is a simple formatting  
17 syntax for authoring HTML, PDF,  
18 and MS Word documents.  
19  
20 ``{r cars}  
21 summary(cars)  
22 ``
```
- Toolbar Actions:** Numbered callouts indicate the following steps:
  1. New File
  2. Embed Code
  3. Write Text
  4. Set Output Format(s) and Options
  5. Save and Render
  6. Share
- RENDERED OUTPUT:** Shows the generated HTML file (`report.html`) with the following content:

file path to output document  
~/Desktop/report/report.html

report.html | Open in Browser | Find

Document Title

Author Name

  - R Markdown
  - Including Plots

R Markdown

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>.

```
summary(cars)
```

	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

report.html 802.9 KB Jul 9, 2021, 4:35 PM

# What is Markdown?

---

Markdown is a lightweight markup language that uses simple text formatting to structure and style text.

- Plain text formatting
- Easy formatting
- Readability
- Versatile



# Headings

---

To denote a **heading**, add a **number sign** (hashtag) before it

- The more number signs, the smaller the text

```
43  
44  
45  
46 There are six types of headers, in decreasing sizes:  
47  
48 # Header one  
49  
50 ## Header two  
51  
52 ### Header three  
53  
54 #### Header four  
55  
56 ##### Header five  
57  
58 ##### Header six  
59  
60
```

The screenshot shows the RStudio interface with the 'Presentation' tab selected in the top menu bar. The code editor contains the following R code:

```
There are six types of headers, in decreasing sizes:  
# Header one  
## Header two  
### Header three  
#### Header four  
##### Header five  
##### Header six
```

The rendered output on the right side of the interface displays the six types of headers as follows:

- Header one
- Header two
- Header three
- Header four
- Header five
- Header six

# Bold and Italics

---

To make a word or string of words **italic**, surround the word or words with 1 asterisk (\*) or 1 underscore (\_)

```
29  ## Italics and Bold  
30  
31 _This_ is one word in italics.  
32  
33 _This is a string of words in italics._  
34  
35 **This** is one word in bold.  
36  
37 **This is a string of words in bold.**  
38  
39 _**This is a few words**_ in **bold** and  
_italics_.
```

To make a word or string of words **bold**, surround the word or words with 2 asterisks (\*\*) or 2 underscores (\_)

## Italics and Bold

*This* is one word in italics.  
*This is a string of words in italics.*  
**This** is one word in bold.  
**This is a string of words in bold.**  
***This is a few words*** in bold and *italics*.



# Lists

---

You can create **unordered** and **ordered** lists in Markdown.

- To create an **unordered** list, preface each item with an asterisk (\*) and give each list item its own line.
- To create an **ordered** list, preface each item with a number and give each item its own line.

139 To create an **unordered** list, you'll want to preface each item in the  
list with an asterisk (\*). Each list item also gets its own line. For  
example, a grocery list in Markdown might look like this:

140  
141 \* Milk  
142 \* Eggs  
143 \* Salmon  
144 \* Butter  
145

146 An ordered list is prefaced with numbers, instead of asterisks. For  
example:

147  
148 1. Crack three eggs over a bowl  
149 2. Pour a gallon of milk into the bowl  
150 3. Rub the salmon vigorously with butter  
151 4. Drop the salmon into the egg-milk bowl

To create an unordered list, you'll want to preface each item in the list with an asterisk (\*).  
Each list item also gets its own line. For example, a grocery list in Markdown might look like  
this:

- Milk
- Eggs
- Salmon
- Butter

An ordered list is prefaced with numbers, instead of asterisks. For example:

1. Crack three eggs over a bowl
2. Pour a gallon of milk into the bowl
3. Rub the salmon vigorously with butter
4. Drop the salmon into the egg-milk bowl



# Hyperlinks

---

There are two link types in Markdown:

1. To create an **inline link**, wrap the link text in brackets ([] ) and the link in parentheses () .
2. A **reference link** is used when you want to reference the same link multiple times in your document. A specific name is selected as the reference, put in square brackets, and then a colon and the link are placed next to it. To use the reference link, wrap the link text in brackets and the reference link name in brackets.

```
90 -  ## Links  
91  
92 -  ##### Inline links  
93  
94 Visit [this website](https://rmarkdown.rstudio.com/lesson-1.html) for more details about R Markdown!  
95  
96  
97 -  ##### Reference links  
98  
99 Here's [a link to something else][another place].  
100 Here's [yet another link][another link].  
101 And now back to [the first link][another place].  
102  
103 [another place]: https://rstudio.github.io/cheatsheets/  
104 [another link]: https://www.hsph.harvard.edu/
```

## Links

### Inline links

Visit [this website](https://rmarkdown.rstudio.com/lesson-1.html) for more details about R Markdown!

### Reference links

Here's [a link to something else](#).

Here's [yet another link](#).

And now back to [the first link](#).



# Images

---

There are two image types in Markdown:

1. To create an **inline image**, use the exact same syntax as an inline hyperlink, but with an exclamation point (!) in the beginning.
2. A **reference image** also uses nearly the same syntax as a reference hyperlink, but with an exclamation point (!) in the beginning.

# Images

---

There are two image types in Markdown:

1. To create an **inline image**, use the exact same syntax as an inline hyperlink, but with an exclamation point (!) in the beginning.
2. A **reference image** also uses nearly the same syntax as a reference hyperlink, but with an exclamation point (!) in the beginning.

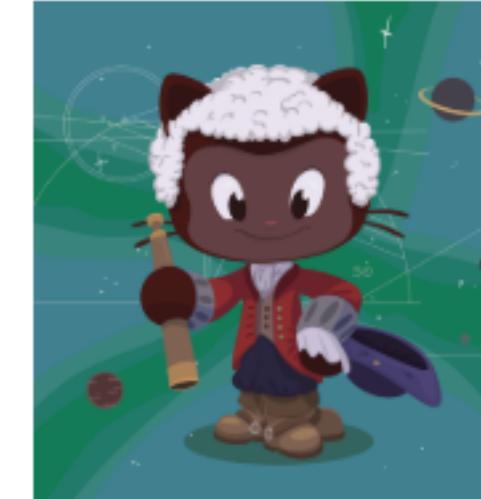
```
125 - ### Images
126
127 - #### Inline images
128
129 For example, to create an inline image link to
https://octodex.github.com/images/bannekat.png, with an alt text
that says, Benjamin Bannekat, you'd write this in Markdown:
130
131 ![Benjamin Bannekat](https://octodex.github.com/images/bannekat.pn
g){ width=25% }

132
133
134
135
```

**Images**

### Inline images

For example, to create an inline image link to <https://octodex.github.com/images/bannekat.png>, with an alt text that says, Benjamin Bannekat, you'd write this in Markdown:



Benjamin Bannekat

# Images

---

There are two image types in Markdown:

1. To create an **inline image**, use the exact same syntax as an inline hyperlink, but with an exclamation point (!) in the beginning.
2. A **reference image** also uses nearly the same syntax as a reference hyperlink, but with an exclamation point (!) in the beginning.

```
136 - ##### Reference images
137
138 ![The first father][First Father]
139
140 !["The second first father"][Second Father]
141
142 [First Father]: http://octodex.github.com/images/founding-father.jpg { width=50% }
143
144 [Second Father]:http://octodex.github.com/images/foundingfather\_v2.png { width=50% }
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
```

Reference images



The first father



The second first father

# Embedding code

There are three ways to create a code chunk:

1. Typing ````{r}` to start a code chunk and then ````` to end it
2. Clicking the **Add Chunk** button in the tool bar
3. The keyboard shortcut

**Ctrl + Alt + I**

(Mac: **Cmd + Option + I**)

The screenshot shows the RStudio interface with an R Markdown file named "00-intro-to-rmarkdown.Rmd". The "Source" tab is selected, displaying the following code:

```
240
249  ```{r}
250  head(pressure)
251  summary(pressure)
252  ...
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
```

The "Knit" button in the toolbar is highlighted. To the right, the "Knit" pane displays the results of the code chunks:

- head(pressure)**

```
##   temperature pressure
## 1          0  0.0002
## 2         20  0.0012
## 3         40  0.0060
## 4         60  0.0300
## 5         80  0.0900
## 6        100  0.2700
```
- summary(pressure)**

```
##   temperature      pressure
## Min.   : 0   Min.   : 0.0002
## 1st Qu.: 90  1st Qu.: 0.1800
## Median :180  Median : 8.8000
## Mean   :180  Mean   :124.3367
## 3rd Qu.:270  3rd Qu.:126.5000
## Max.   :360  Max.   :806.0000
```

# Embedding code

There are different display options for code and code output

- Adding a comma and **echo = FALSE** after the **r** displays the code output but not the code itself
  - Adding a comma and **include = FALSE** prevents code and output from appearing in the finished file

```
249 - ````{r, echo = FALSE}
250 head(pressure)
251 summary(pressure)
252 ````
```

## temperature pressure

## 1 0 0.0002

## 2 20 0.0012

## 3 40 0.0060

## 4 60 0.0300

## 5 80 0.0900

## 6 100 0.2700

---

## temperature pressure

## Min. : 0 Min. : 0.0002

## 1st Qu.: 90 1st Qu.: 0.1800

## Median :180 Median : 8.8000

## Mean :180 Mean :124.3367

## 3rd Qu.:270 3rd Qu.:126.5000

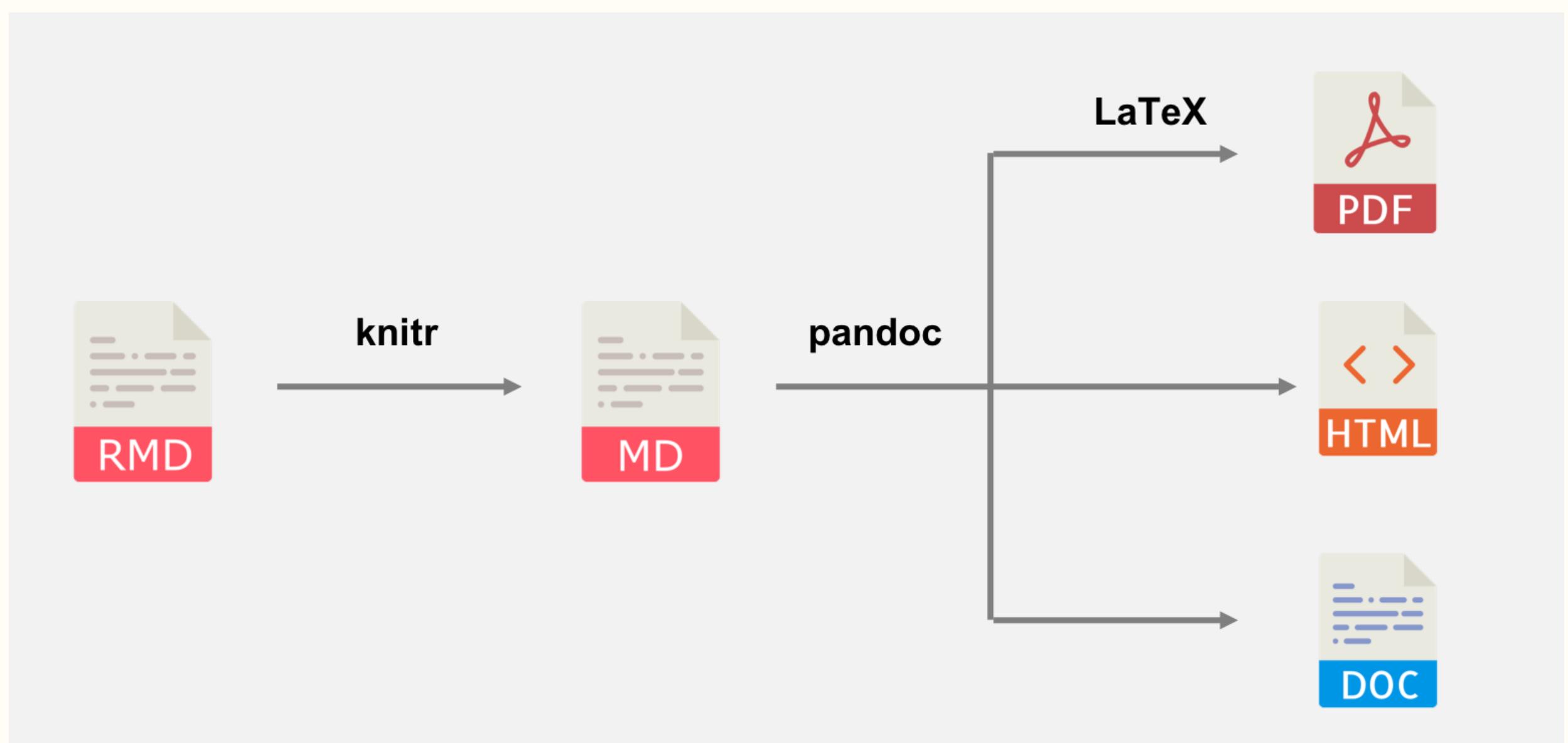
## Max. :360 Max. :806.0000

# Compiling R Markdown files

---

How does an R Markdown file (.Rmd file) compile?

- The **rmarkdown** package renders an R Markdown file by executing other packages that knit the text and code chunks together into an output file
- The output file can be a Word doc, PDF, or HTML file
- The file will not knit (compile) if there is a bug in the code

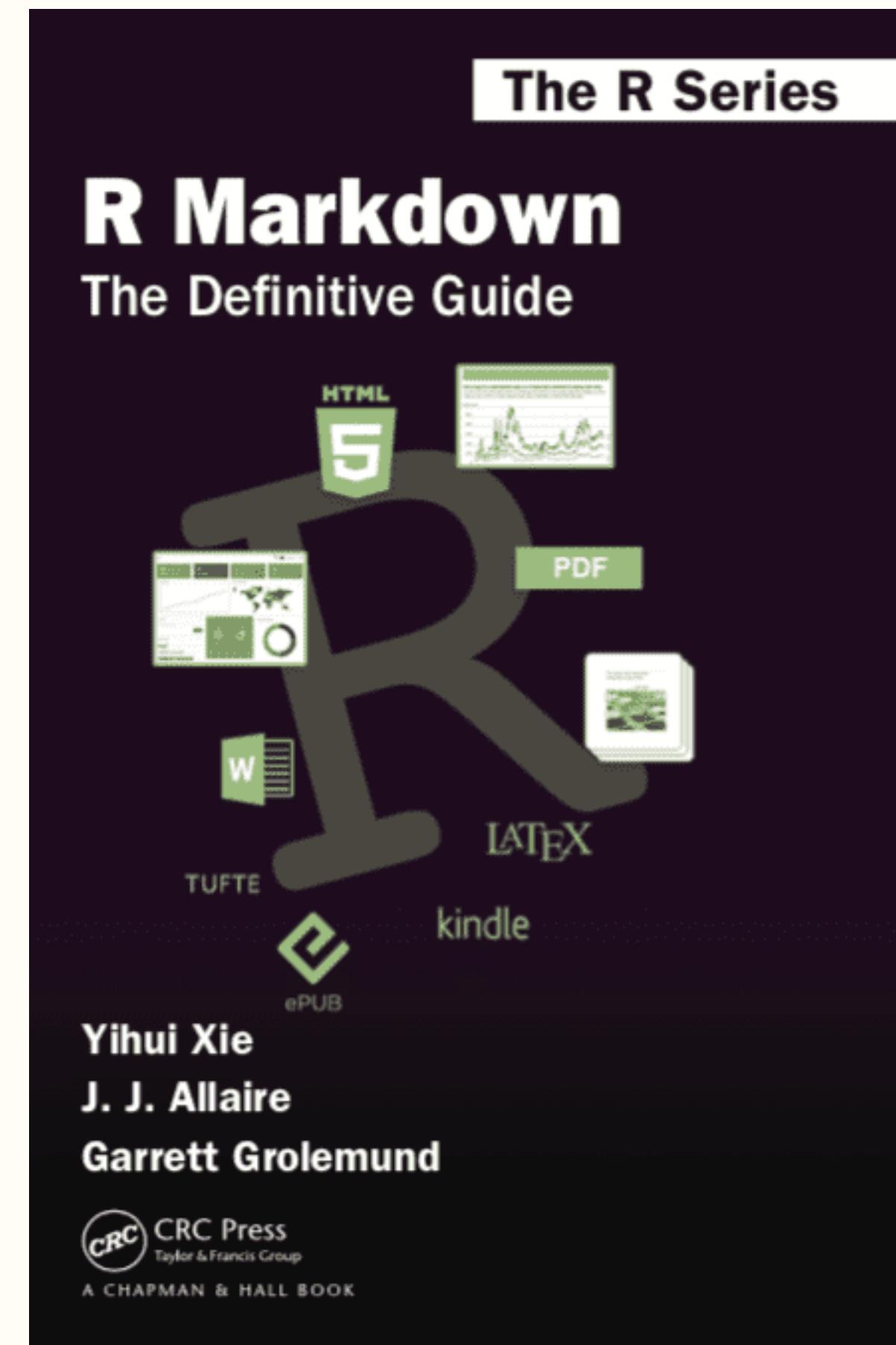


# Other document types

---

R Markdown can also be used to produce

- PDF presentations with beamer
- HTML presentations with ioslides and slidy
- Full academic manuscripts
- Dashboards
- Websites
- Interactive documents



# Summary

---

- R Markdown is a tool that can create documents, slides, websites, etc. that combine text and code into one file
- Markdown is the text formatting markup language used in R Markdown files

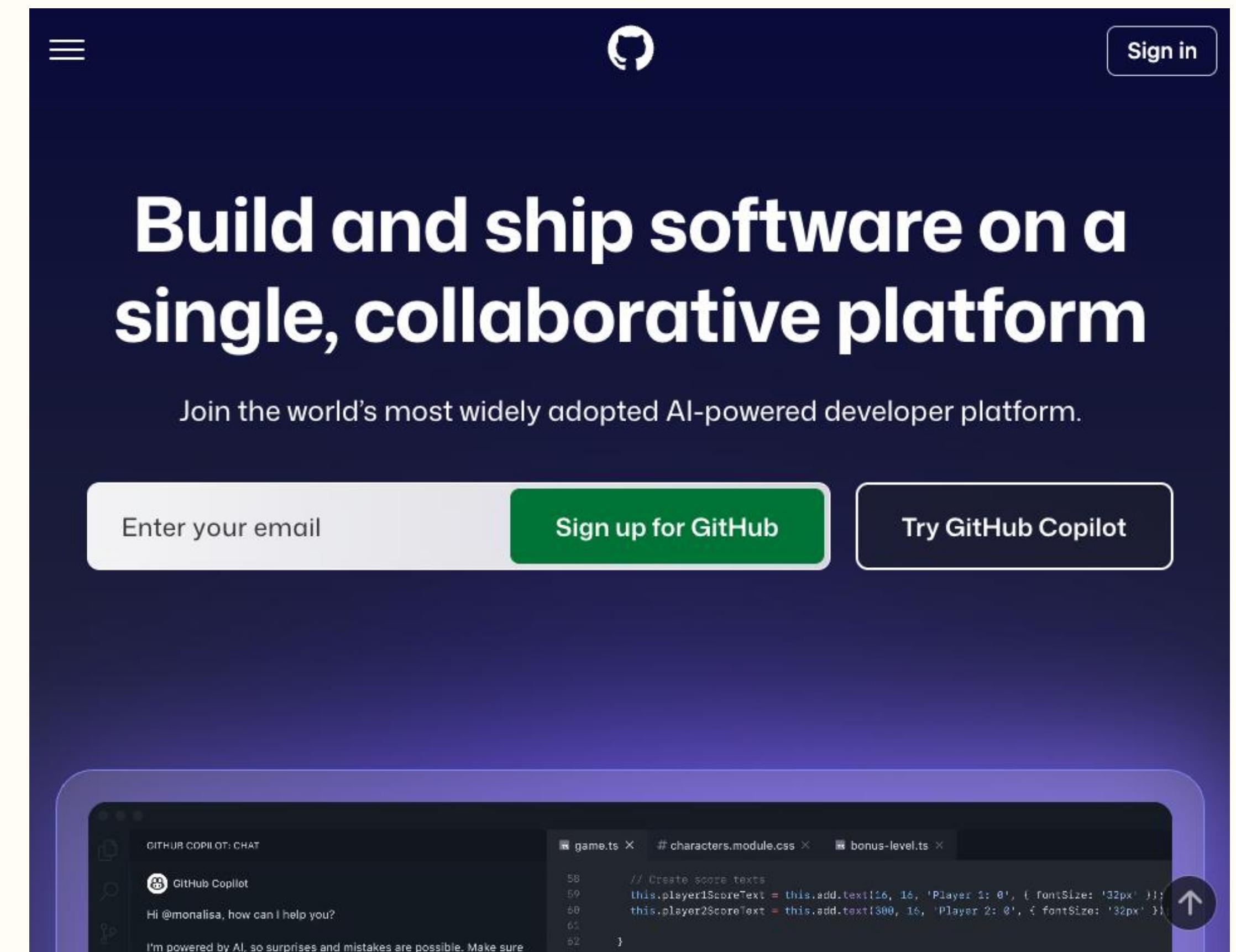
# Introduction to GitHub



# GitHub

**GitHub** is a cloud-based platform for software developers and others to build, store, manage, and share code and projects using Git's version control system. It acts as a central hub for collaboration, allowing teams to

- Track changes
- Work on separate versions at the same time
- Suggest and review contributions
- Manage project tasks



# GitHub

## What will we use GitHub for?

- All course materials
- All of your assignments

The screenshot shows a GitHub repository page for 'BST219\_2025'. The repository is public and owned by 'coredatasience'. The main branch is 'main' with 1 branch and 0 tags. The repository was last updated yesterday. The commit history shows several files being added via upload, including '00\_course\_introduction', '.gitignore', 'LICENSE', and 'README.md'. The 'About' section provides a brief description of the repository as a companion for BST 219 at Harvard T.H. Chan School of Public Health. It also lists standard repository metrics: 1 star, 0 watching, and 0 forks. A 'Welcome to BST 219: Core Principles of Data Science!' message is displayed in the README.

coredatasience / BST219\_2025

Type / to search

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

BST219\_2025 Public

Edit Pins Watch 0 Fork 0 Star 1

main 1 Branch 0 Tags Go to file Add file Code

hmatte Add files via upload 47ee320 · yesterday 5 Commits

00\_course\_introduction Add files via upload yesterday

.gitignore Create .gitignore 2 days ago

LICENSE Initial commit 2 days ago

README.md Update README.md 2 days ago

README MIT license

Welcome to BST 219: Core Principles of Data Science!

Course materials for Fall 2025 can be found here.

About

Companion GitHub repository for BST 219: Core Principles of Data Science at the Harvard T.H. Chan School of Public Health.

Readme

MIT license

Activity

Custom properties

1 star

0 watching

0 forks

Report repository

Releases

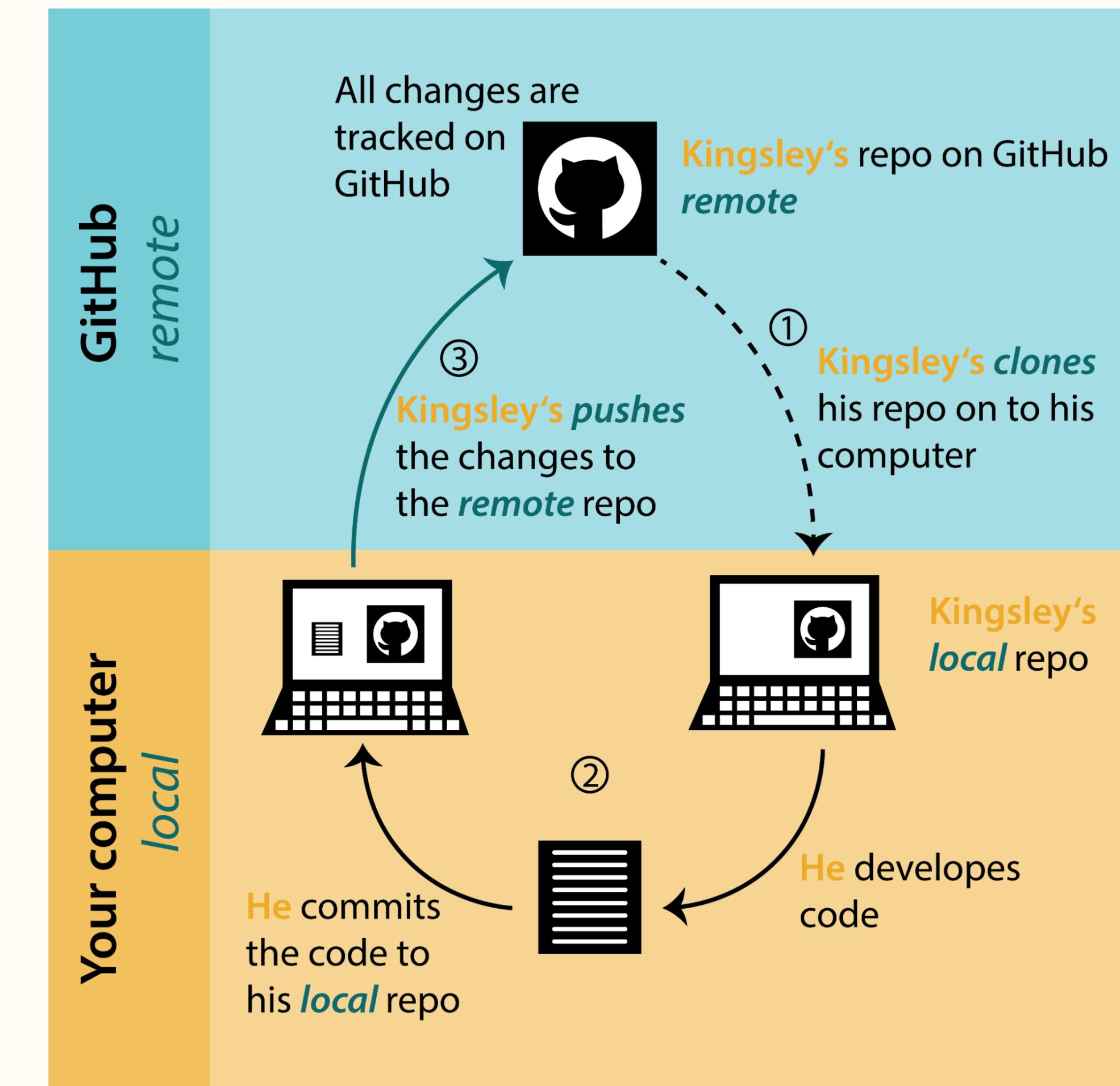
No releases published

Create a new release

# GitHub

## The main reasons we will be using GitHub

- Version control so you don't lose any work
- Always good to be exposed to a tool commonly used in data science



# GitHub

---

- GitHub has specific terminology it would be useful to remember.
- Different commands are used for different steps of the version control process. We will only be using a few of them in this course.



## Definitions

- **repo** - short for repository
- **repository** - the directory or folder that is under version control
- **local** - the repository on your computer
- **remote** - the repository on GitHub
- **commit** - take a snapshot of one or more files in the repository
- **push** - send commits from the local repo to the remote repo
- **pull** - retrieve commits from the remote repo to the local repo
- **.gitignore** a file that tells git which files or types of files you don't want to commit

# Accessing GitHub in RStudio

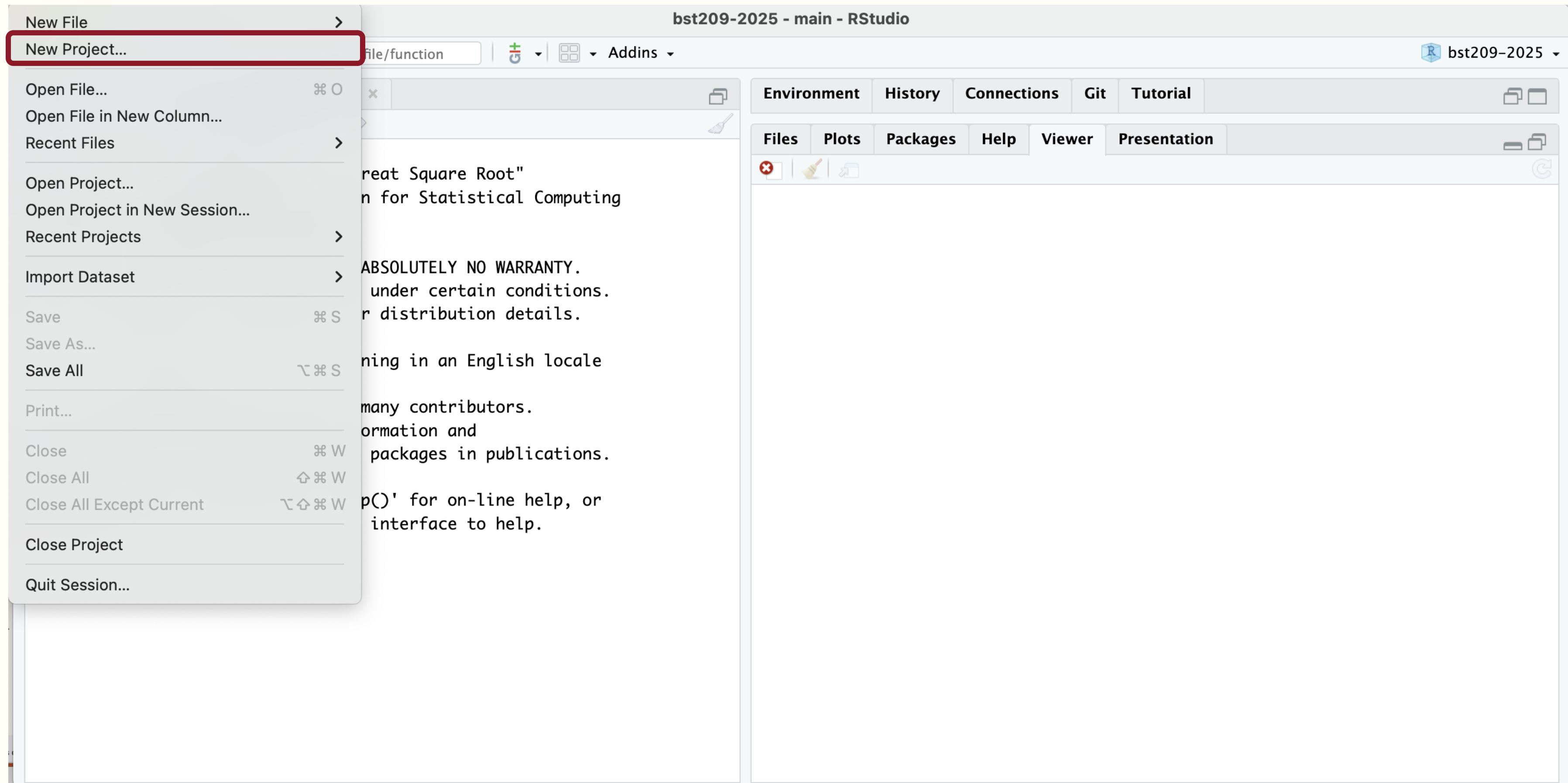
We won't be using the Terminal or Command Line to work with GitHub - we'll be working inside RStudio.

**Step 1:** Clone the repository

The screenshot shows a GitHub repository page for 'coredatasience / BST219\_2025'. The 'Code' tab is selected. A tooltip is overlaid on the 'Code' button, which is highlighted with a red border. The tooltip contains three numbered steps: 1. 'Clone' (highlighted with a red border), 2. 'HTTPS' (highlighted with a red border), and 3. 'https://github.com/coredatasience/BST219\_2025'. The main repository page displays files like '00\_course\_introduction', '.gitignore', 'LICENSE', and 'README.md'. A welcome message at the bottom says 'Welcome to BST 219: Core Principles of Data Science!'.

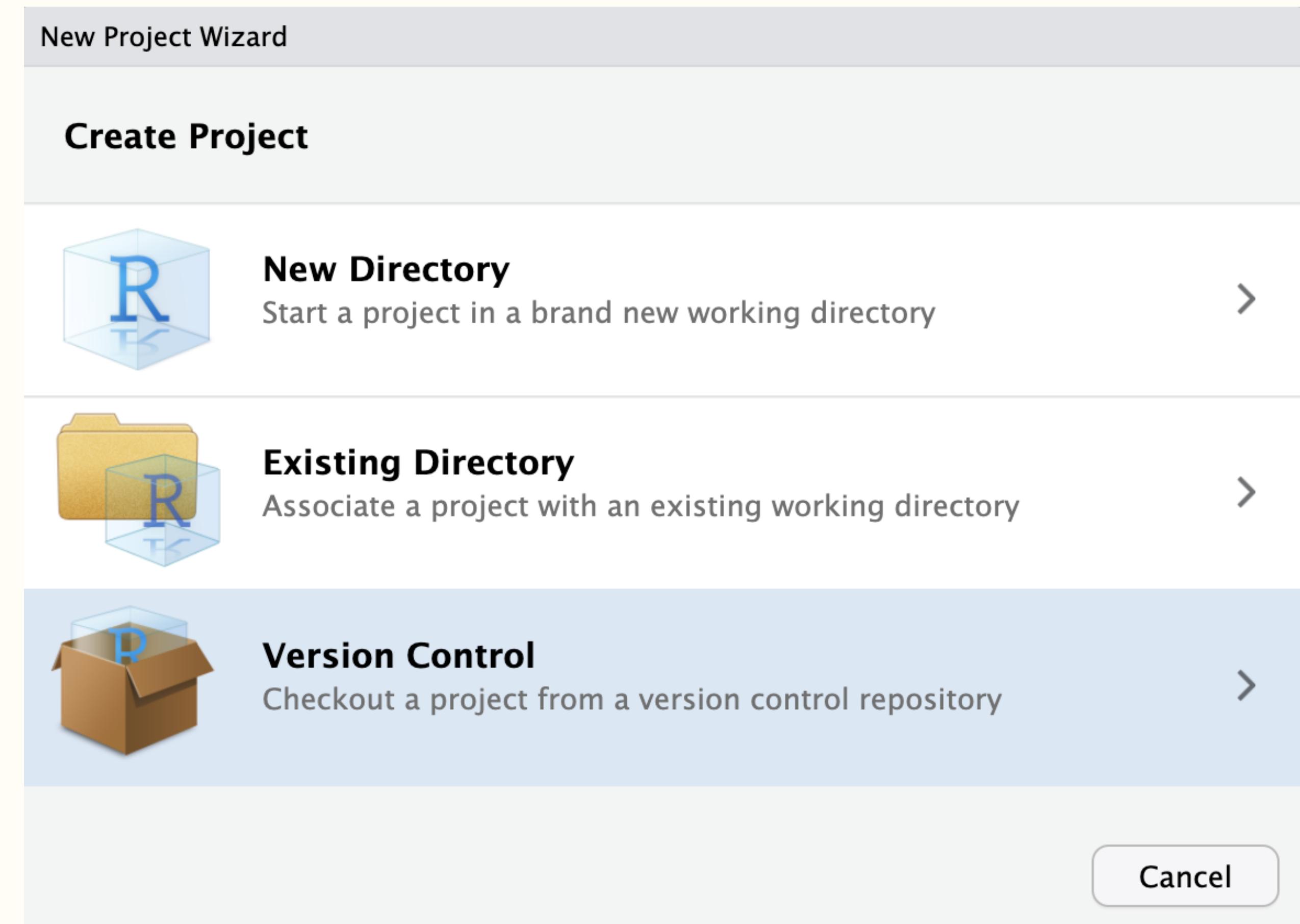
# Accessing GitHub in RStudio

Step 2: Open RStudio and select New Project



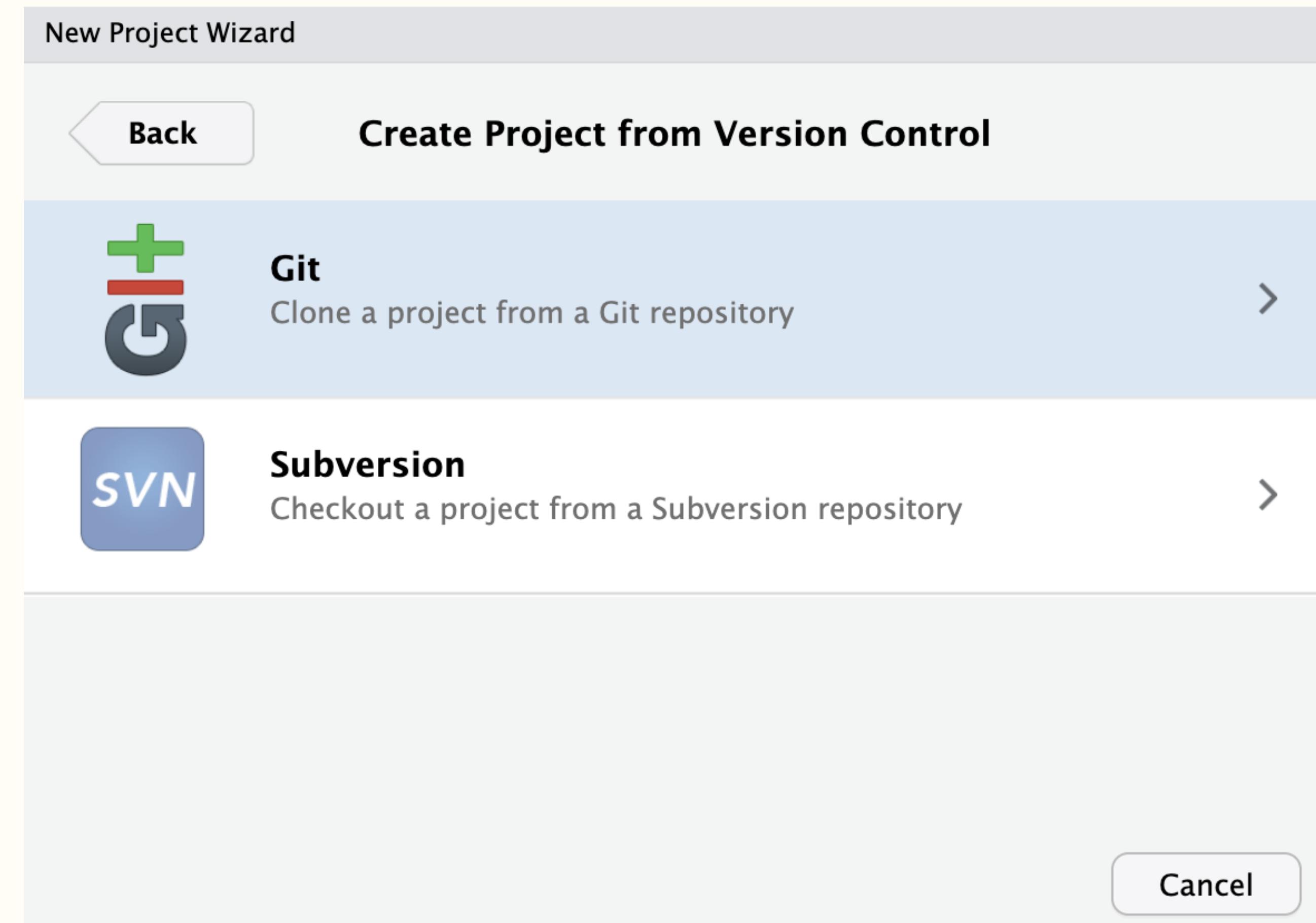
# Accessing GitHub in RStudio

## Step 3: Select Version Control



# Accessing GitHub in RStudio

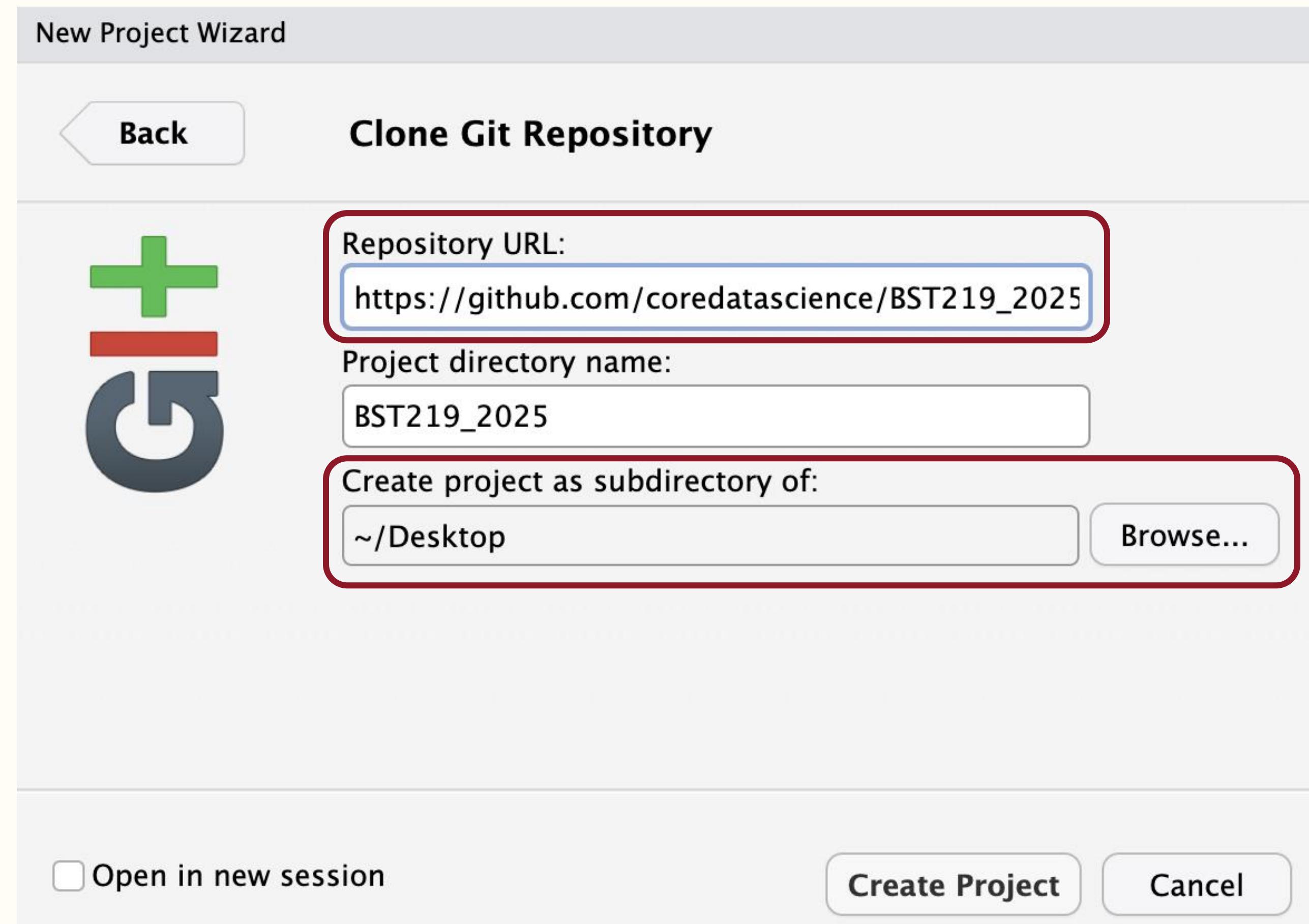
## Step 4: Select Git



# Accessing GitHub in RStudio

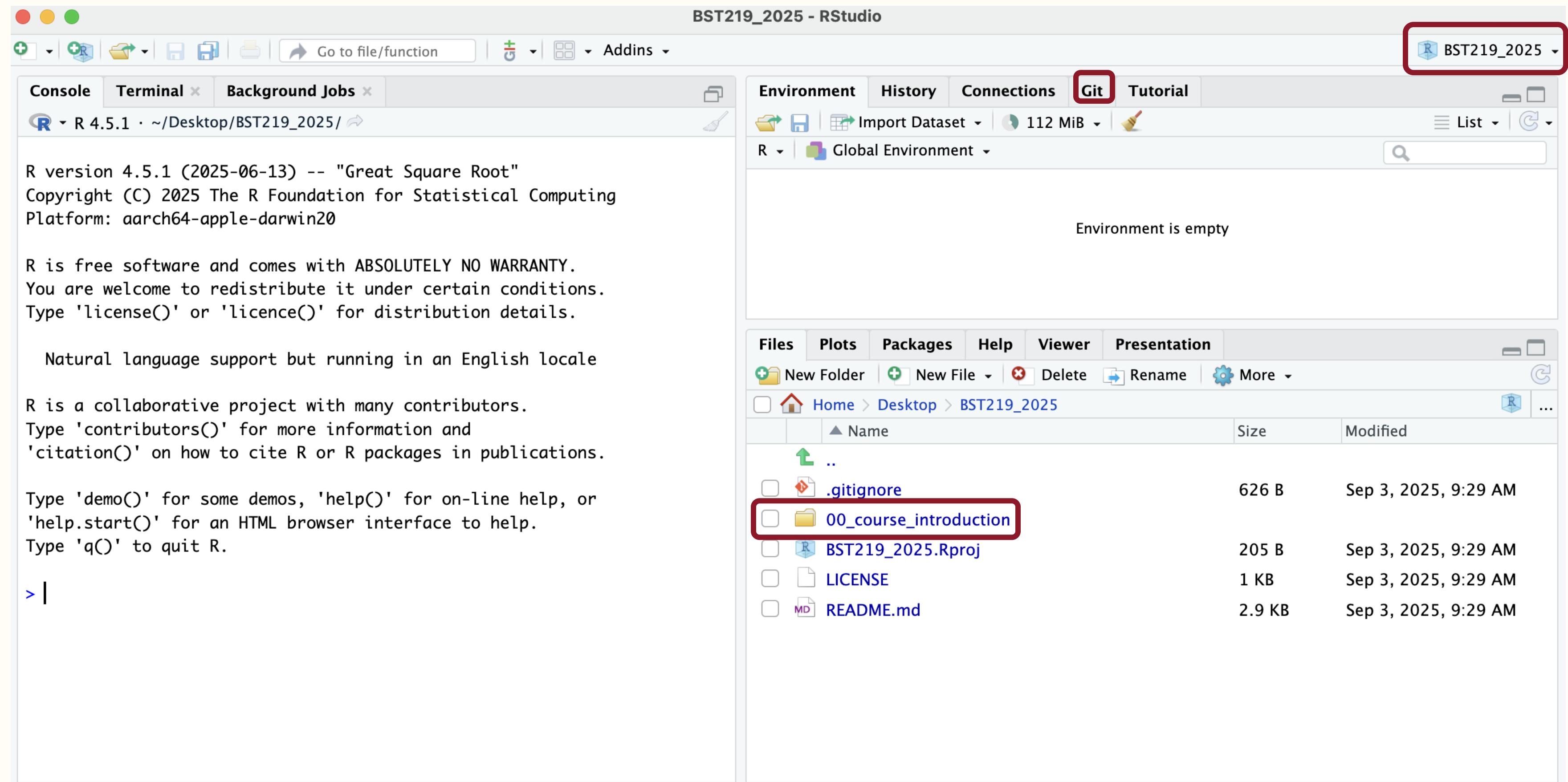
**Step 5:** Paste the repository link you copied on GitHub

This will become a folder on your computer - take note of where you are saving it!



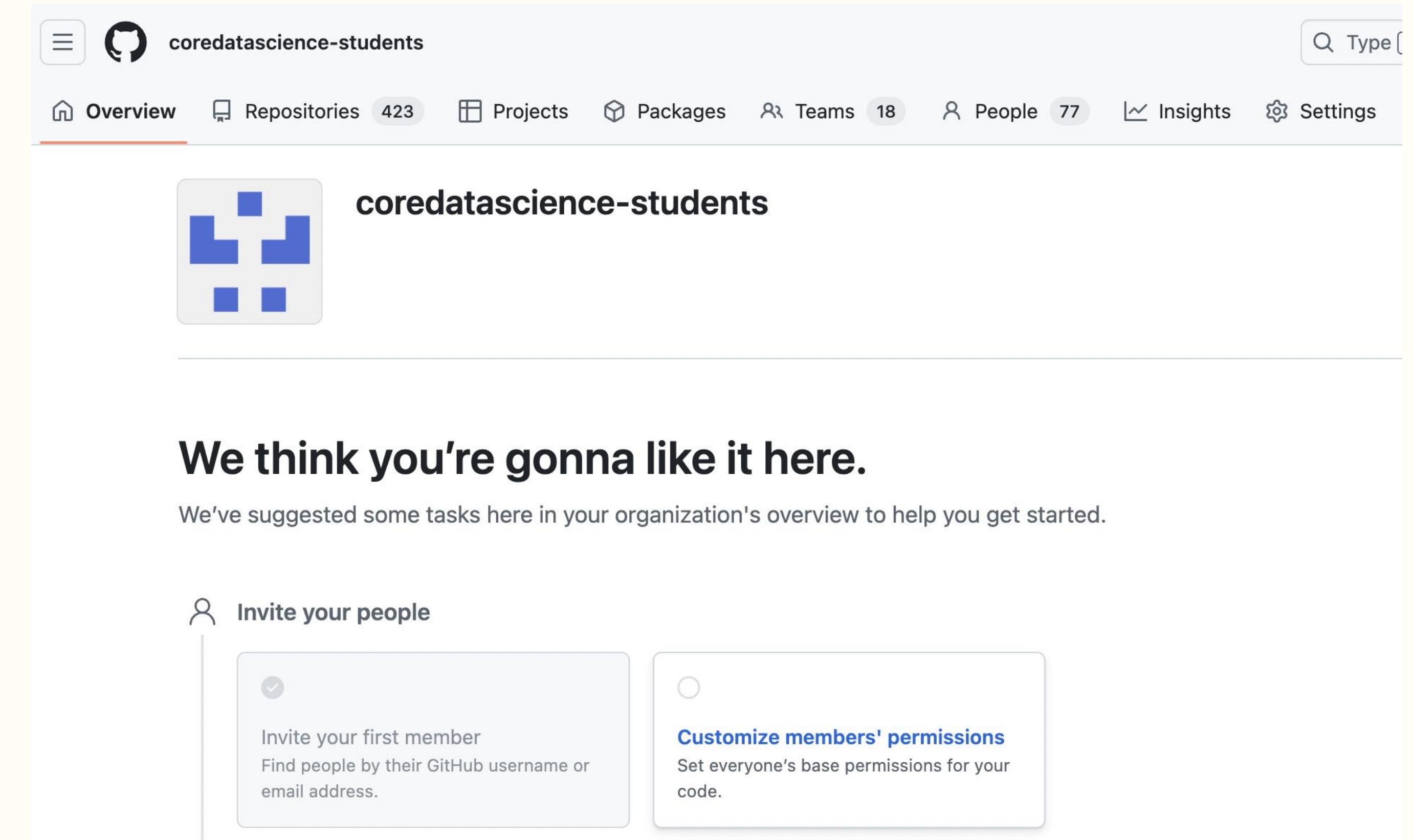
# Accessing GitHub in RStudio

The project should open and you should see the project (repo) name and all folders and files in the repo.



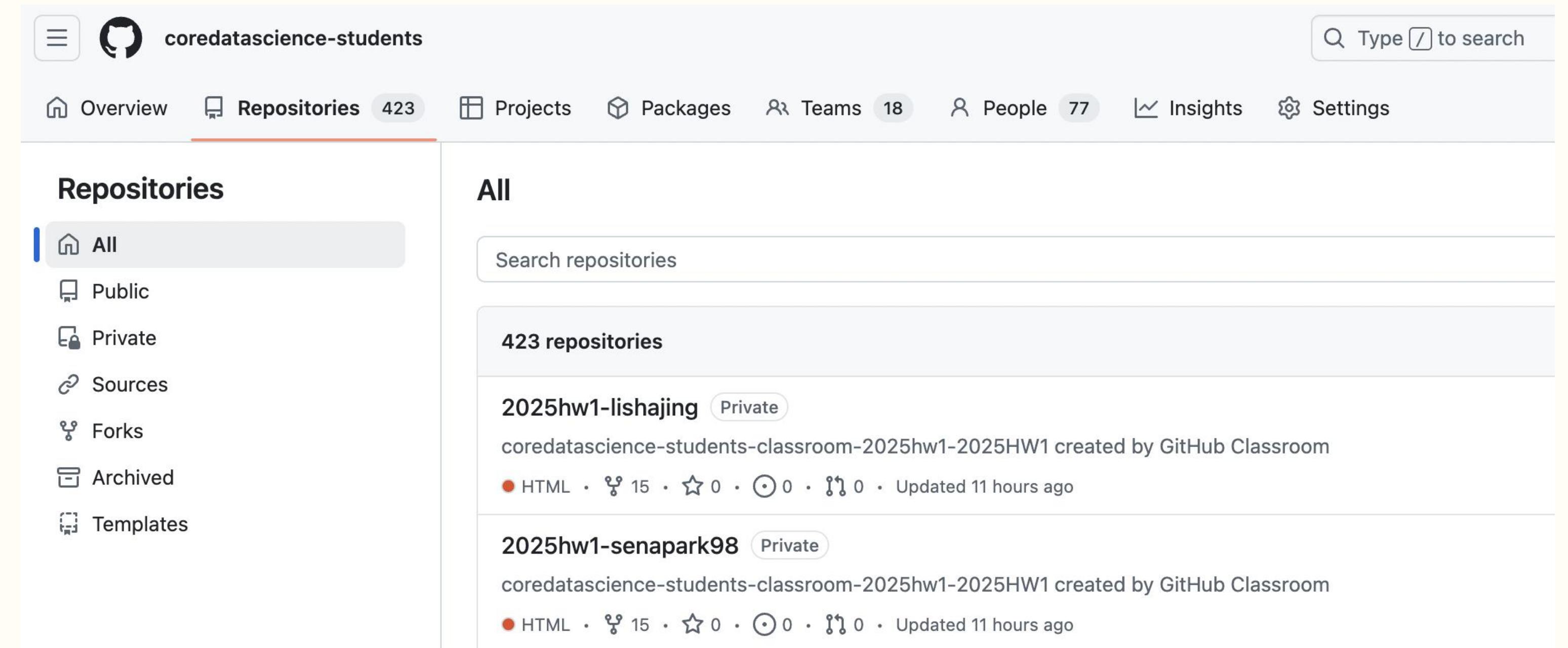
# Working on and submitting homework

The steps to clone your homework repositories are exactly the same as the course materials repository, except you will go to the [student repositories page](#).



# Working on and submitting homework

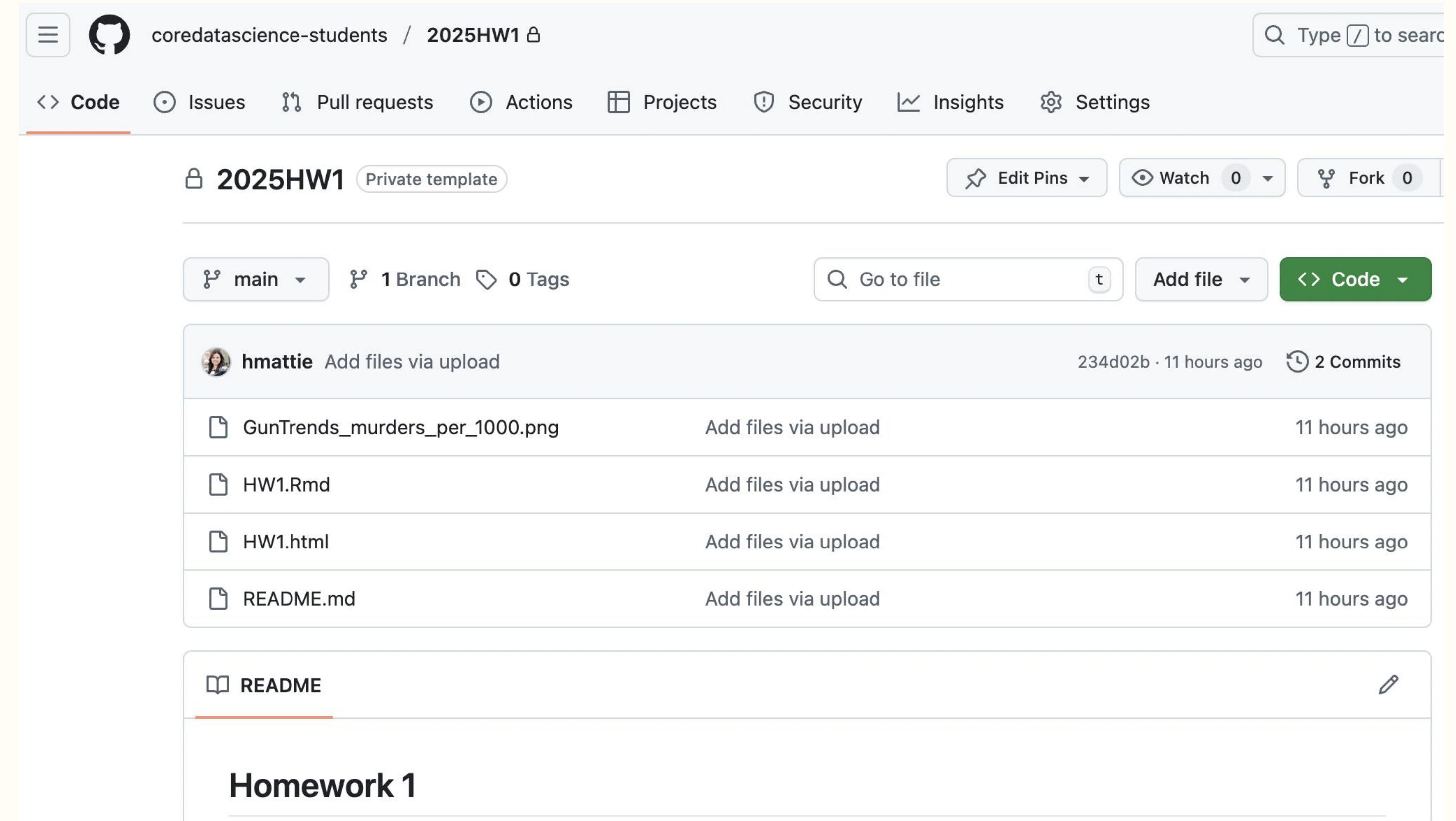
The steps to clone your homework repositories are exactly the same as the course materials repository, except you will go to the [student repositories page](#).



A screenshot of the GitHub interface for the repository 'coredatascience-students'. The top navigation bar shows 'Overview', 'Repositories' (423), 'Projects', 'Packages', 'Teams' (18), 'People' (77), 'Insights', and 'Settings'. The 'Repositories' tab is selected. On the left, a sidebar lists 'All', 'Public', 'Private', 'Sources', 'Forks', 'Archived', and 'Templates'. The main area is titled 'All' and shows '423 repositories'. Two repositories are listed: '2025hw1-lishajing' (Private) and '2025hw1-senapark98' (Private). Both are described as 'coredatascience-students-classroom-2025hw1-2025HW1 created by GitHub Classroom'. Each repository has a link to 'HTML', a star icon with '15', a circular icon with '0', a circular icon with '0', and a circular icon with '0'. The text 'Updated 11 hours ago' is also present.

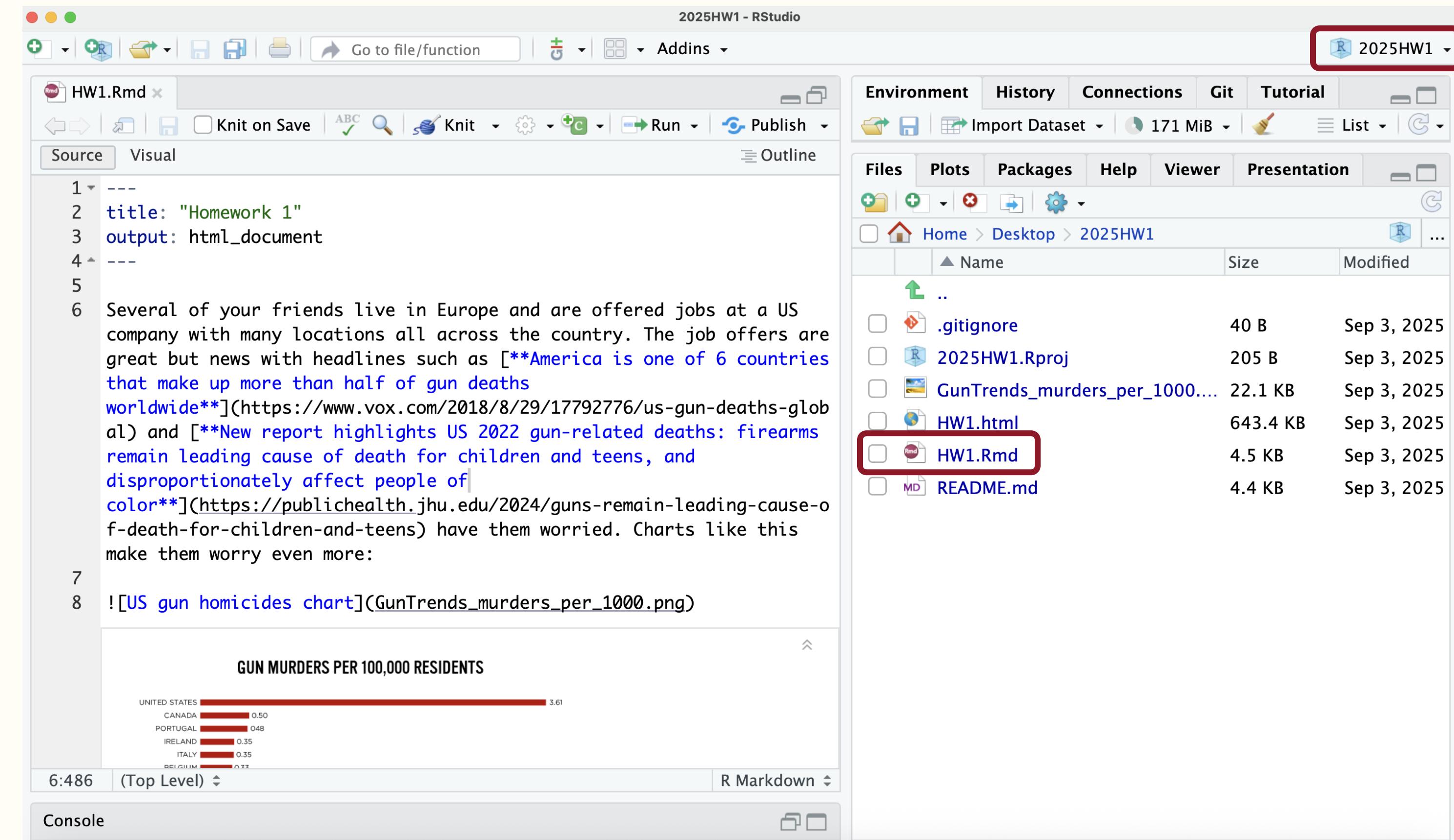
# Working on and submitting homework

The steps to clone your homework repositories are exactly the same as the course materials repository, except you will go to the [student repositories page](#).



# Working on and submitting homework

Once you have cloned your homework repository, click on the .Rmd file to work on your assignment.



The screenshot shows the RStudio interface with the following details:

- Title Bar:** 2025HW1 - RStudio
- File Tab:** HW1.Rmd
- Toolbar:** Go to file/function, Addins, Knit on Save, Knit, Run, Publish.
- Source View:** Contains R Markdown code:

```
1 ---  
2 title: "Homework 1"  
3 output: html_document  
4 ---  
5  
6 Several of your friends live in Europe and are offered jobs at a US  
company with many locations all across the country. The job offers are  
great but news with headlines such as [**America is one of 6 countries  
that make up more than half of gun deaths  
worldwide**](https://www.vox.com/2018/8/29/17792776/us-gun-deaths-global) and [**New report highlights US 2022 gun-related deaths: firearms  
remain leading cause of death for children and teens, and  
disproportionately affect people of  
color**](https://publichealth.jhu.edu/2024/guns-remain-leading-cause-of-death-for-children-and-teens) have them worried. Charts like this  
make them worry even more:  
7  
8 ![US gun homicides chart](GunTrends_murders_per_1000.png)
```
- Plot View:** Displays a horizontal bar chart titled "GUN MURDERS PER 100,000 RESIDENTS". The chart compares murder rates across several countries, with the United States having the highest rate.

Country	Murder Rate (per 100,000 residents)
UNITED STATES	3.61
CANADA	0.50
PORTUGAL	0.48
IRELAND	0.35
ITALY	0.35
BELGIUM	0.22
- File Browser:** Shows the project structure under "2025HW1": .gitignore, 2025HW1.Rproj, GunTrends\_murders\_per\_1000..., HW1.html, HW1.Rmd (highlighted with a red box), and README.md.

# Working on and submitting homework

When you are done working on the assignment and want to save your progress, or submit the final version, knit the .Rmd file and then click on the **Git** tab. Then click the boxes next to the .Rmd and .html file names, and press the **Commit** button.

The screenshot shows the RStudio interface with the following details:

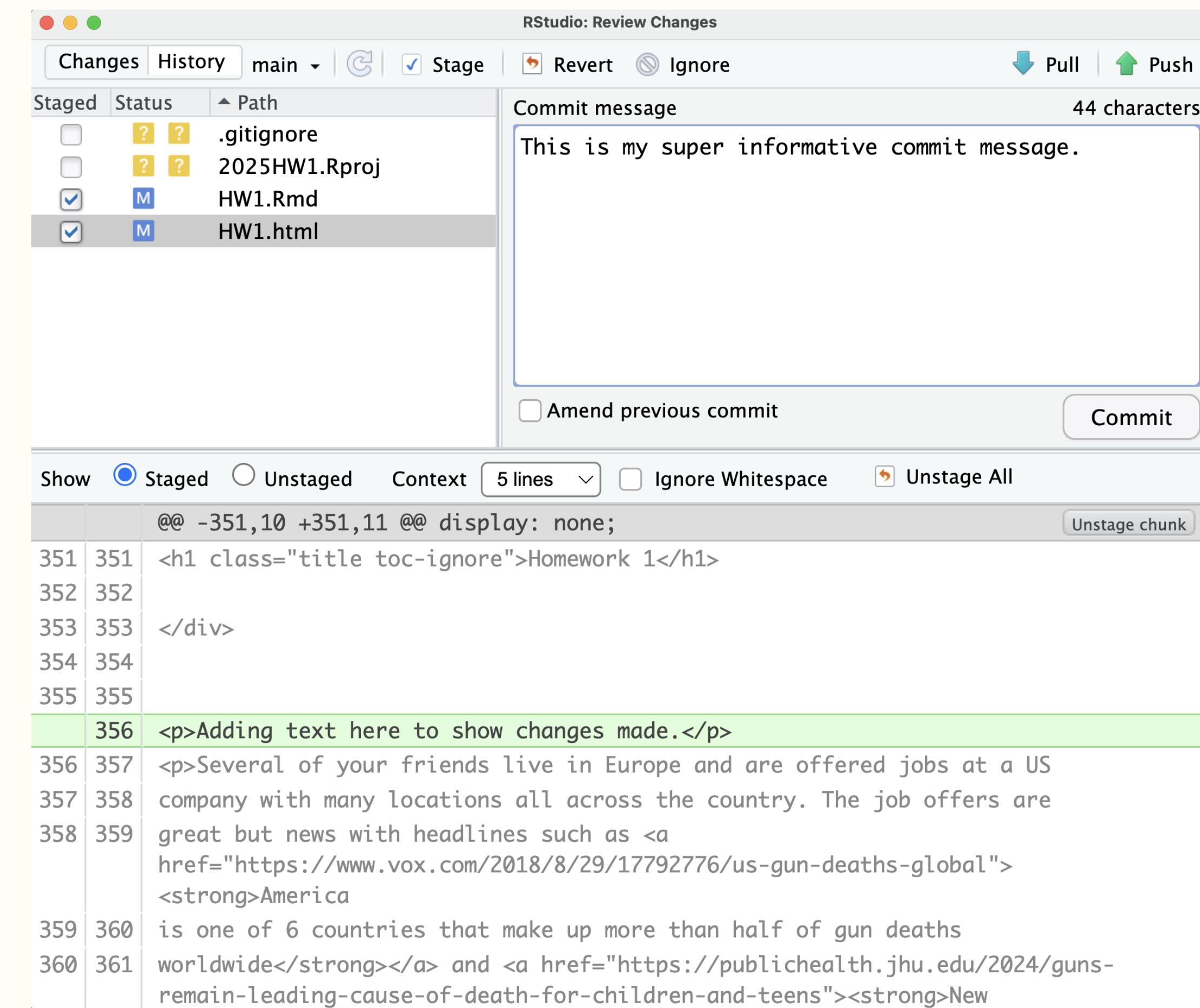
- Source Tab:** Displays the Rmd code for "HW1.Rmd".

```
1 ---  
2 title: "Homework 1"  
3 output: html_document  
4 ---  
5  
6 Adding text here to show changes made.  
7  
8 Several of your friends live in Europe and are offered jobs at a US company with many locations all across the country. The job offers are great but news with headlines such as \[\*\*America is one of 6 countries that make up more than half of gun deaths worldwide\*\*\](https://www.vox.com/2018/8/29/17792776/us-gun-deaths-global) and \[\*\*New report highlights US 2022 gun-related deaths: firearms
```
- Console Tab:** Shows the R console output:

```
R 4.5.1 · ~/Desktop/2025HW1/  
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.
```
- Git Tab:** Shows the Staged and Status panes. A red box highlights the checkboxes next to "HW1.Rmd" and "HW1.html" in the Staged pane.
- Preview Area:** Shows the rendered HTML output for "HW1.Rmd".

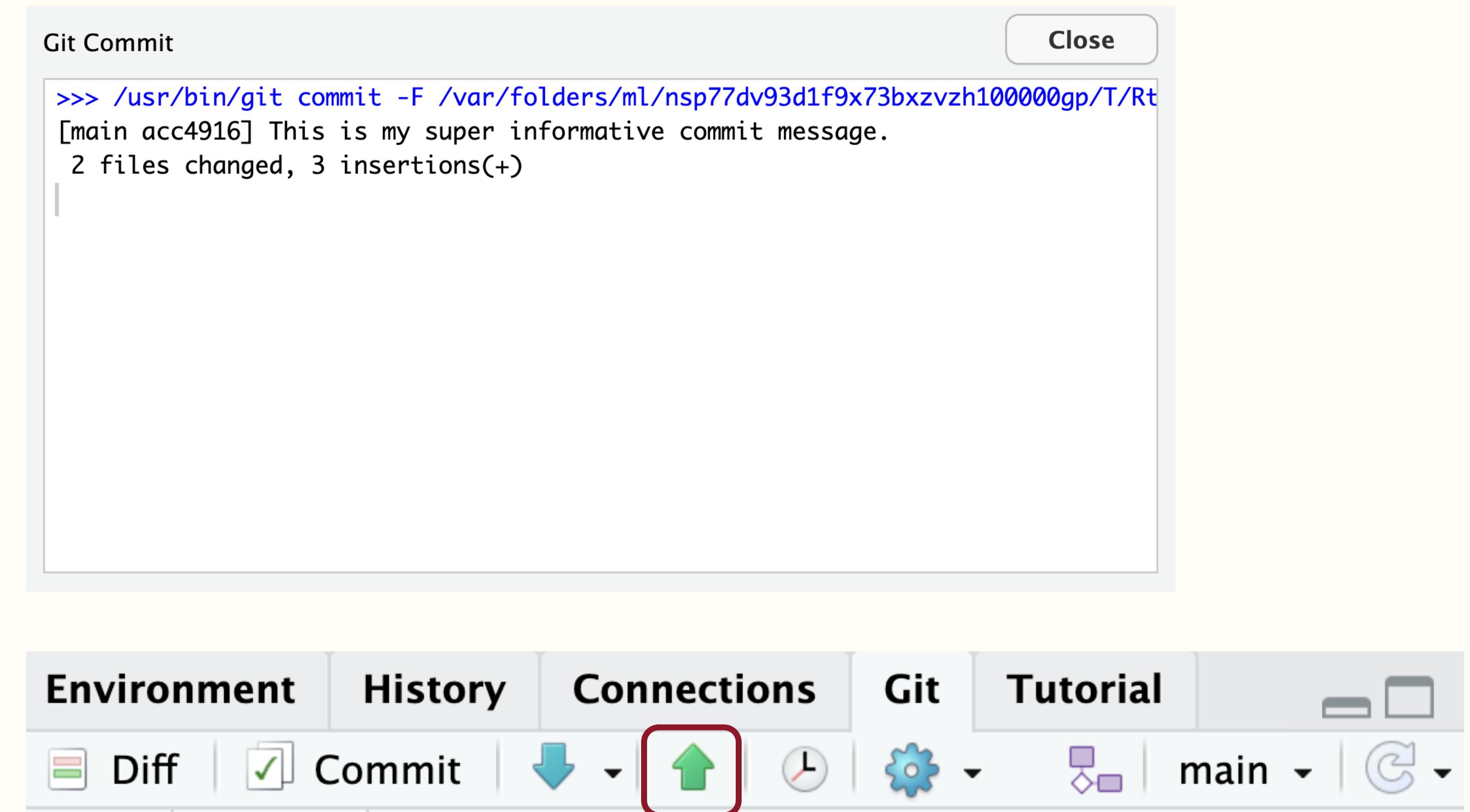
# Working on and submitting homework

A new window will pop up. Write a commit message, e.g., “Question 1 answered” or “Final version” and press **Commit**.



# Working on and submitting homework

A new window will pop up – but you aren't done! Close this window and then click the **Push** green button.



# Working on and submitting homework

---

A new window will pop up that looks something like this. Now you are done!

Git Push Close

```
>>> /usr/bin/git push origin HEAD:refs/heads/main
To https://github.com/coredatascience-students/2025HW1.git
 234d02b..acc4916  HEAD -> main
```

# Working on and submitting homework

You can check that you pushed successfully by going back to your repository on GitHub, refreshing the page, and looking to see your Commit message shows up.

The screenshot shows a GitHub repository named "2025HW1" which is a "Private template". The repository has one branch ("main") and no tags. A search bar at the top right includes options for "Go to file", "Add file", and "Code". The main area displays a commit from user "hmattie" with the message "This is my super informative commit message." and timestamp "acc4916 · now". Below the commit, there are four files listed: "GunTrends\_murders\_per\_1000.png", "HW1.Rmd", "HW1.html", and "README.md". To the right of these files is a button labeled "Add files via upload". Further down, there is a section for the "README" file, which contains the text "Homework 1". A red box highlights the commit message "This is my super informative commit message." from the first commit.

# Working on and submitting homework

We encourage you to knit, commit, and push often to save your progress!

The screenshot shows a GitHub repository named "2025HW1" which is a "Private template". The repository has one branch ("main") and no tags. A search bar at the top right includes filters for "Go to file", "t", "Add file", and "Code". The commit history shows three commits from "hmattie":

- 1. "This is my super informative commit message." (acc4916 · now) - This commit message is highlighted with a red box.
- 2. "This is my super informative commit message." (now) - Another commit message is highlighted with a red box.
- 3. "This is my super informative commit message." (now)

Below the commits, there are sections for "Add files via upload" next to "GunTrends\_murders\_per\_1000.png", "HW1.Rmd", "HW1.html", and "README.md". A "Homework 1" section is also present.