The Git Stage/Commit/Push workflow

Childhood Cancer Data Lab

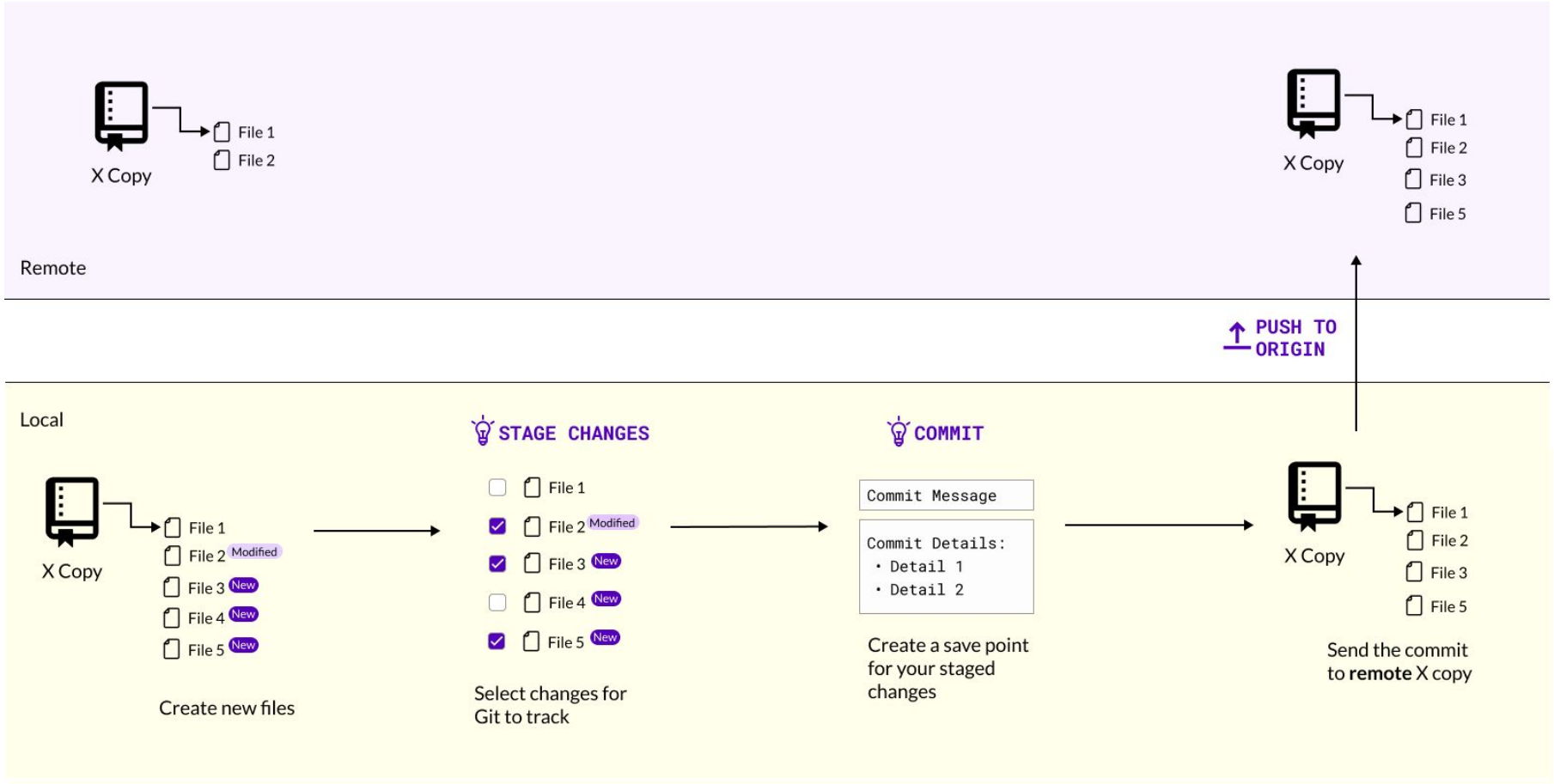
The "stage/commit/push" flow

1. Make code changes

2. **Stage** the changes you want to be part of the next commit
   a. Best practice to stage *one changed file at a time* to avoid problems
   b. Note: you may also hear this step referred to as "add"

3. **Commit** your code changes with an informative message

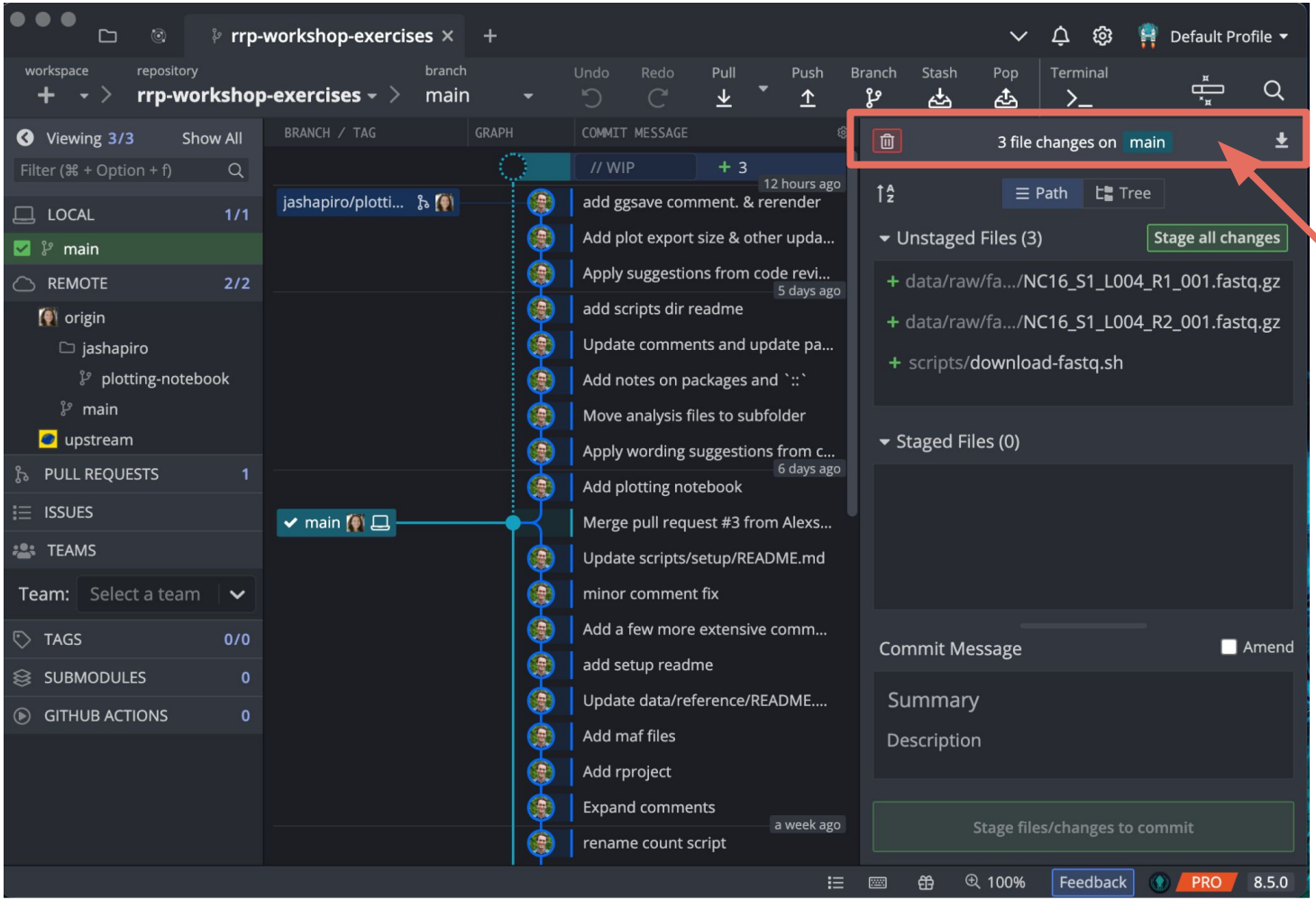
4. **Push** your local commit(s) to the **remote repository** you cloned on GitHub.com

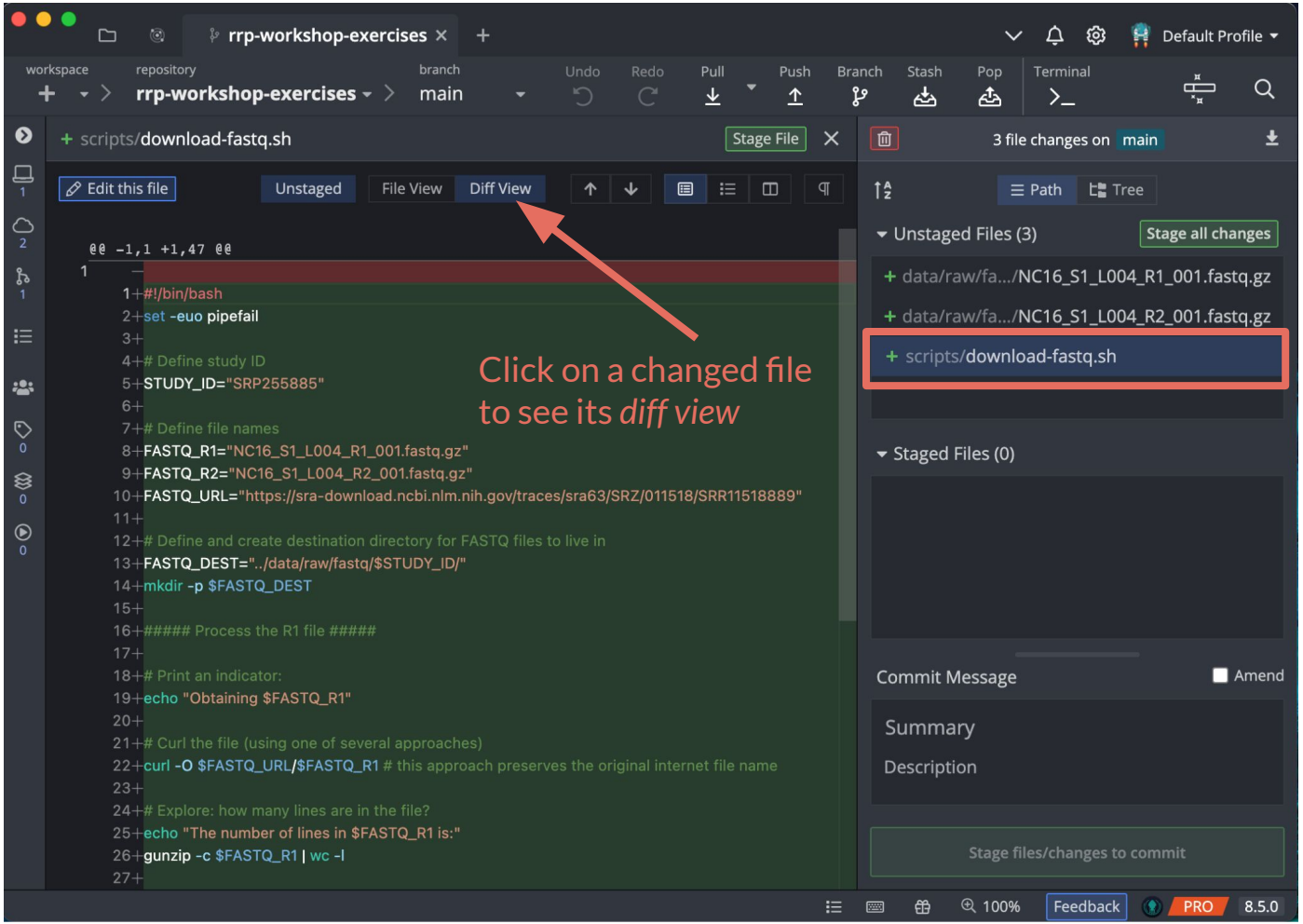

X Copy
File 1
File 2
X Copy
File 1
File 2
File 3
File 5
Remote
PUSH TO ORIGIN
Local
X Copy
File 1
File 2 Modified
File 3 New
File 4 New
File 5 New
Create new files
STAGE CHANGES
File 1
File 2 Modified
File 3 New
File 4 New
File 5 New
Select changes for Git to track
COMMIT
Commit Message
Commit Details:
• Detail 1
• Detail 2
Create a save point for your staged changes
X Copy
File 1
File 2
File 3
File 5
Send the commit to remote X copy

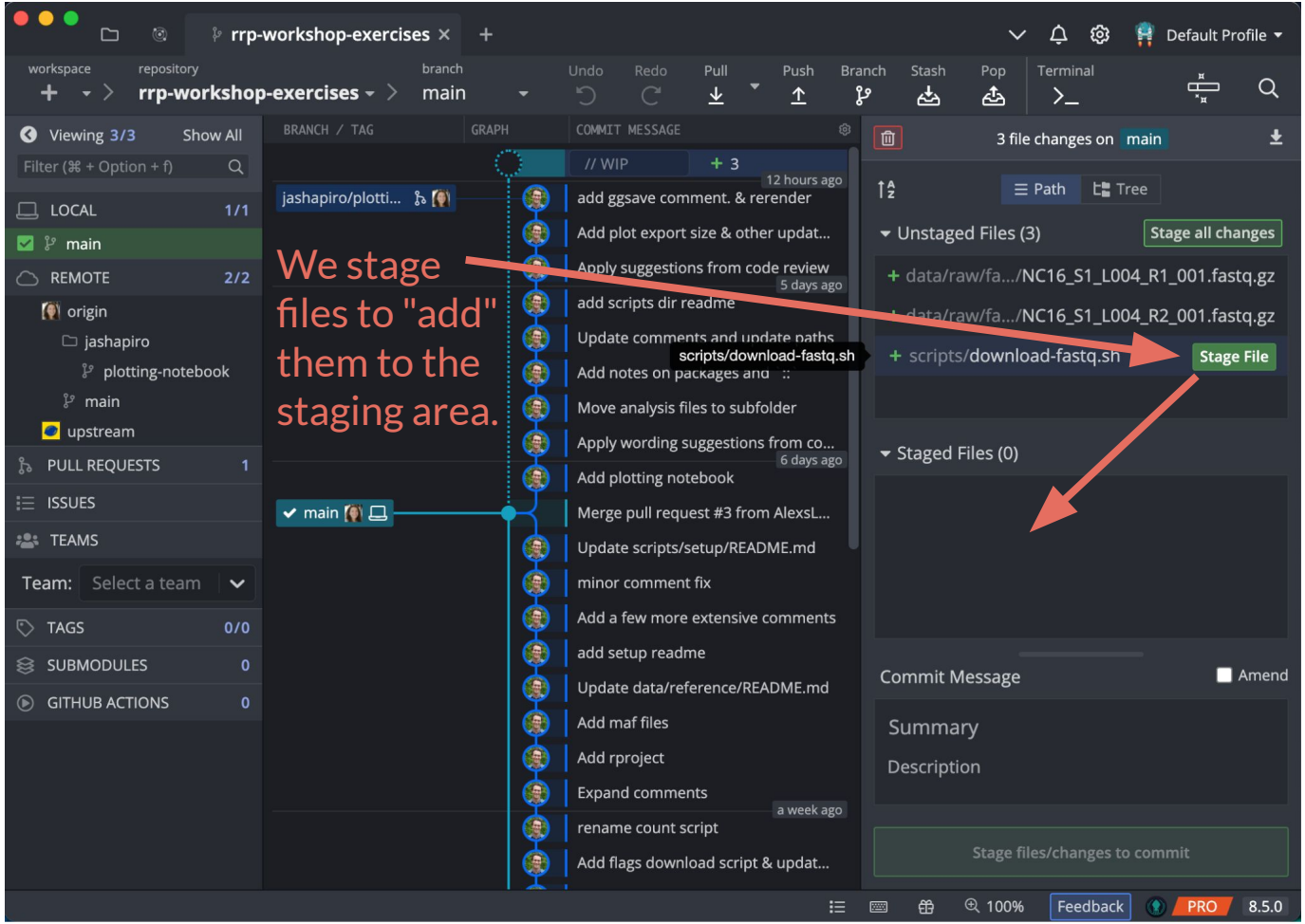
How much should you stage for a commit?

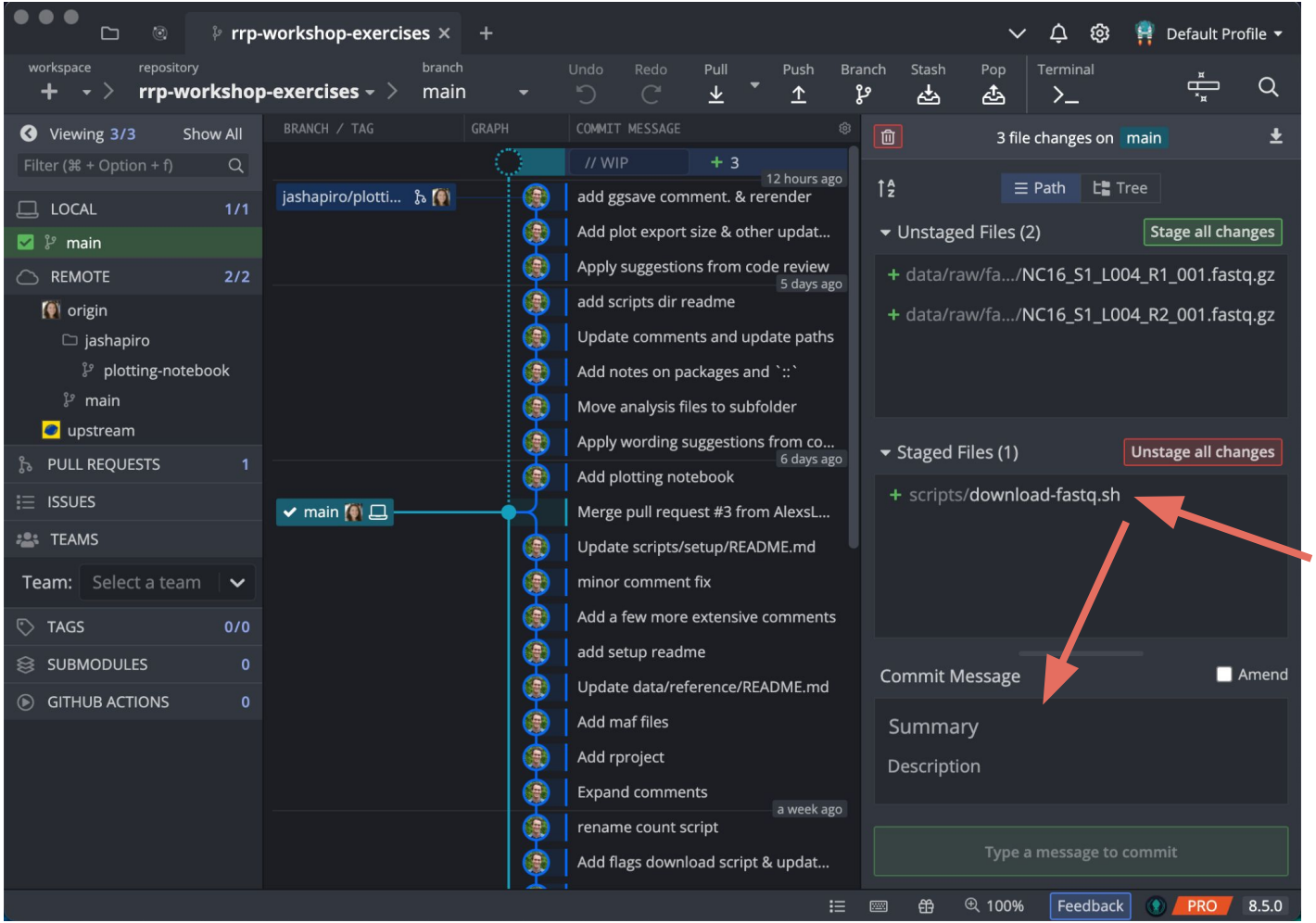
- Smaller units of work are generally better *for you!*
- Imagine you are looking in your code history. Which commit message is more helpful?
  - `All the work I did on Tuesday`
  - `Modified notebook to change plot size`
- Which scenario is easier to look through?
  - 1-3 changed files in a commit
  - 25 changed files in a commit

https://xkcd.com/1296/

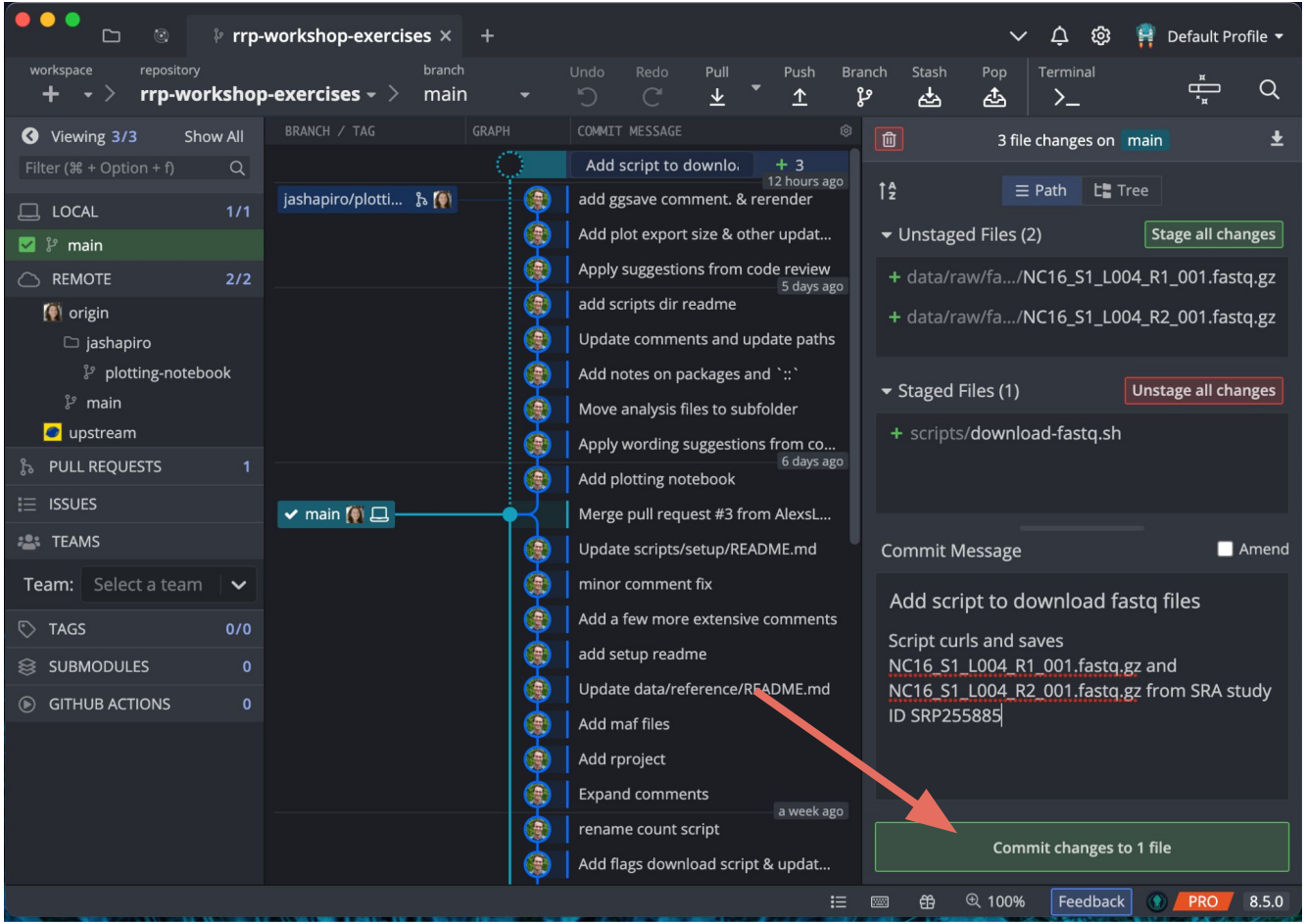

rrp-workshop-exercises
Default Profile
workspace
repository
rrp-workshop-exercises
branch
main
Undo
Redo
Pull
Push
Branch
Stash
Pop
Terminal
Viewing 3/3
Show All
Filter (⌘ + Option + f)
LOCAL 1/1
main
REMOTE 2/2
origin
jashapiro
plotting-notebook
main
upstream
PULL REQUESTS 1
ISSUES
TEAMS
Team: Select a team
TAGS 0/0
SUBMODULES 0
GITHUB ACTIONS 0
BRANCH / TAG
GRAPH
COMMIT MESSAGE
// WIP + 3
12 hours ago
jashapiro/plotti...
add ggsave comment. & rerender
Add plot export size & other upda...
Apply suggestions from code revi...
5 days ago
add scripts dir readme
Update comments and update pa...
Add notes on packages and `::`
Move analysis files to subfolder
Apply wording suggestions from c...
6 days ago
Add plotting notebook
main
Merge pull request #3 from Alexs...
Update scripts/setup/README.md
minor comment fix
Add a few more extensive comm...
add setup readme
Update data/reference/README....
Add maf files
Add rproject
Expand comments
a week ago
rename count script
3 file changes on main
Path
Tree
Unstaged Files (3)
Stage all changes
+ data/raw/fa.../NC16_S1_L004_R1_001.fastq.gz
+ data/raw/fa.../NC16_S1_L004_R2_001.fastq.gz
+ scripts/download-fastq.sh
Staged Files (0)
Commit Message
Amend
Summary
Description
Stage files/changes to commit
100%
Feedback
PRO
8.5.0

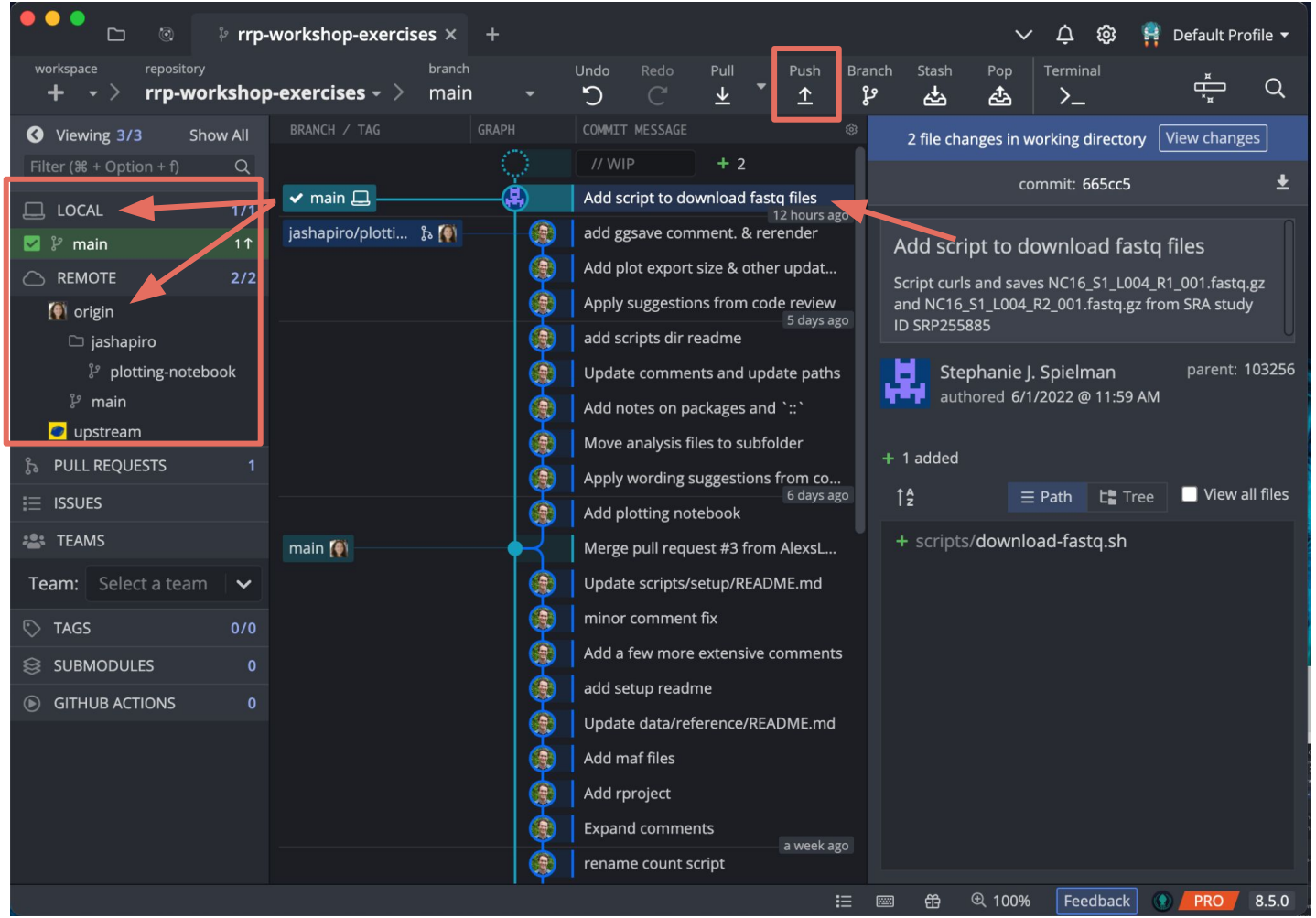

rrp-workshop-exercises
Default Profile
workspace
repository
rrp-workshop-exercises
branch
main
Undo
Redo
Pull
Push
Branch
Stash
Pop
Terminal
+ scripts/download-fastq.sh
Stage File
Edit this file
Unstaged
File View
Diff View
@@ -1,1 +1,47 @@
#!/bin/bash
set -euo pipefail
# Define study ID
STUDY_ID="SRP255885"
# Define file names
FASTQ_R1="NC16_S1_L004_R1_001.fastq.gz"
FASTQ_R2="NC16_S1_L004_R2_001.fastq.gz"
FASTQ_URL="https://sra-download.ncbi.nlm.nih.gov/traces/sra63/SRZ/011518/SRR11518889"
# Define and create destination directory for FASTQ files to live in
FASTQ_DEST="../data/raw/fastq/$STUDY_ID/"
mkdir -p $FASTQ_DEST
Process the R1 file
Print an indicator:
echo "Obtaining $FASTQ_R1"
Curl the file (using one of several approaches)
curl -O $FASTQ_URL/$FASTQ_R1 # this approach preserves the original internet file name
Explore: how many lines are in the file?
echo "The number of lines in $FASTQ_R1 is:"
gunzip -c $FASTQ_R1 | wc -l
Click on a changed file to see its diff view
3 file changes on main
Path
Tree
Unstaged Files (3)
Stage all changes
+ data/raw/fa.../NC16_S1_L004_R1_001.fastq.gz
+ data/raw/fa.../NC16_S1_L004_R2_001.fastq.gz
+ scripts/download-fastq.sh
Staged Files (0)
Commit Message
Amend
Summary
Description
Stage files/changes to commit
100%
Feedback
PRO
8.5.0

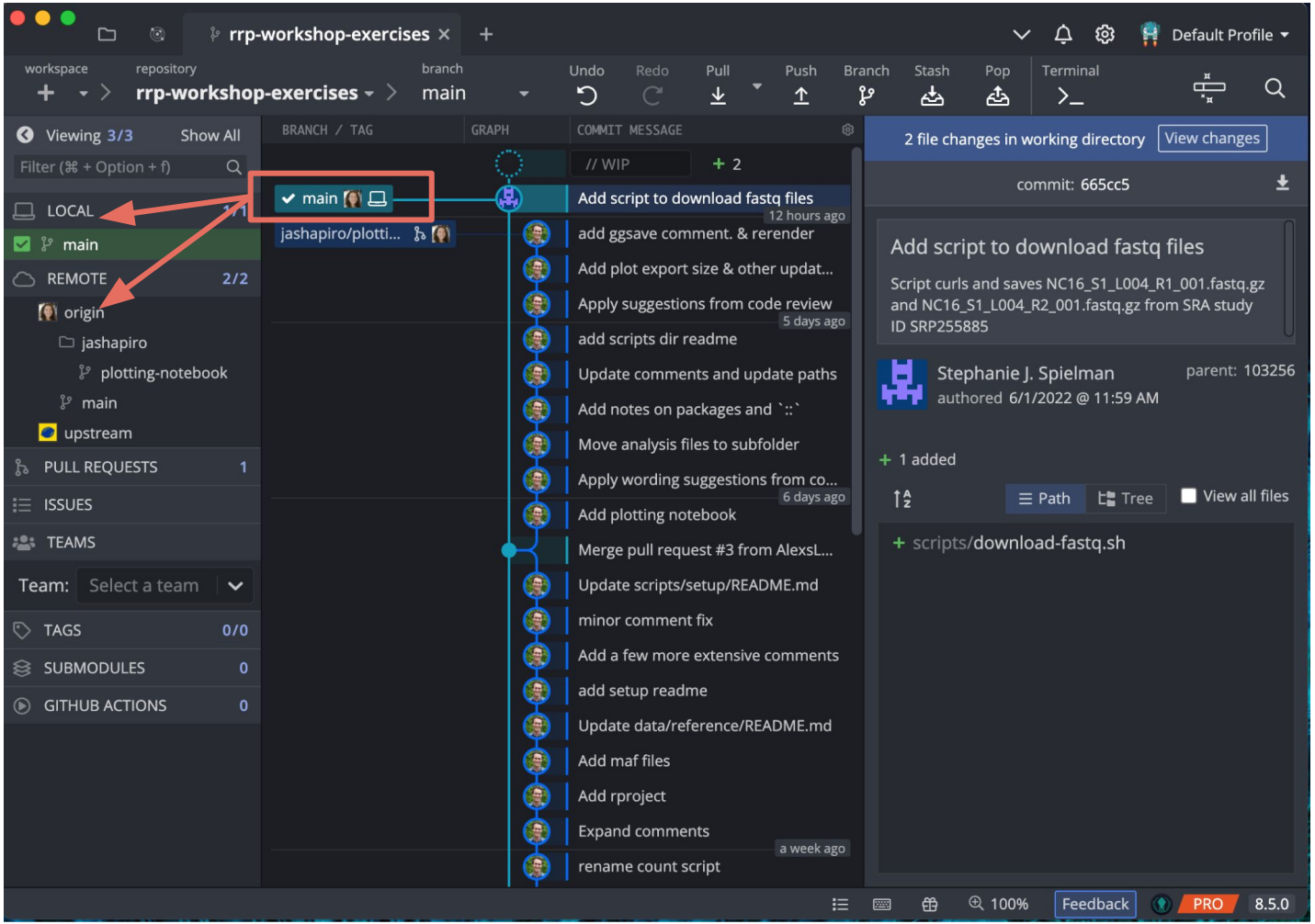

rrp-workshop-exercises
Default Profile
workspace
repository
rrp-workshop-exercises
branch
main
Undo
Redo
Pull
Push
Branch
Stash
Pop
Terminal
Viewing 3/3
Show All
Filter (⌘ + Option + f)
LOCAL 1/1
main
REMOTE 2/2
origin
jashapiro
plotting-notebook
main
upstream
PULL REQUESTS 1
ISSUES
TEAMS
Team: Select a team
TAGS 0/0
SUBMODULES 0
GITHUB ACTIONS 0
BRANCH / TAG
GRAPH
COMMIT MESSAGE
// WIP + 3
12 hours ago
jashapiro/plotti...
add ggsave comment. & rerender
Add plot export size & other updat...
Apply suggestions from code review
5 days ago
add scripts dir readme
Update comments and update paths
Add notes on packages and ::
Move analysis files to subfolder
Apply wording suggestions from co...
6 days ago
Add plotting notebook
main
Merge pull request #3 from AlexsL...
Update scripts/setup/README.md
minor comment fix
Add a few more extensive comments
add setup readme
Update data/reference/README.md
Add maf files
Add rproject
Expand comments
a week ago
rename count script
Add flags download script & updat...
We stage files to "add" them to the staging area.
3 file changes on main
Path
Tree
Unstaged Files (3)
Stage all changes
+ data/raw/fa.../NC16_S1_L004_R1_001.fastq.gz
+ data/raw/fa.../NC16_S1_L004_R2_001.fastq.gz
scripts/download-fastq.sh
+ scripts/download-fastq.sh
Stage File
Staged Files (0)
Commit Message
Amend
Summary
Description
Stage files/changes to commit
100%
Feedback
PRO
8.5.0



Now we are ready to *commit* these staged changes.


rrp-workshop-exercises
Default Profile
workspace
repository
rrp-workshop-exercises
branch
main
Undo
Redo
Pull
Push
Branch
Stash
Pop
Terminal
Viewing 3/3
Show All
Filter (⌘ + Option + f)
LOCAL 1/1
main
REMOTE 2/2
origin
jashapiro
plotting-notebook
main
upstream
PULL REQUESTS 1
ISSUES
TEAMS
Team: Select a team
TAGS 0/0
SUBMODULES 0
GITHUB ACTIONS 0
BRANCH / TAG
GRAPH
COMMIT MESSAGE
Add script to downlo
+ 3
12 hours ago
jashapiro/plotti...
add ggsave comment. & rerender
Add plot export size & other updat...
Apply suggestions from code review
5 days ago
add scripts dir readme
Update comments and update paths
Add notes on packages and `::`
Move analysis files to subfolder
Apply wording suggestions from co...
6 days ago
Add plotting notebook
main
Merge pull request #3 from AlexsL...
Update scripts/setup/README.md
minor comment fix
Add a few more extensive comments
add setup readme
Update data/reference/README.md
Add maf files
Add rproject
Expand comments
a week ago
rename count script
Add flags download script & updat...
3 file changes on main
Path
Tree
Unstaged Files (2)
Stage all changes
+ data/raw/fa.../NC16_S1_L004_R1_001.fastq.gz
+ data/raw/fa.../NC16_S1_L004_R2_001.fastq.gz
Staged Files (1)
Unstage all changes
+ scripts/download-fastq.sh
Commit Message
Amend
Add script to download fastq files
Script curls and saves NC16_S1_L004_R1_001.fastq.gz and NC16_S1_L004_R2_001.fastq.gz from SRA study ID SRP255885
Commit changes to 1 file
100%
Feedback
PRO
8.5.0


rrp-workshop-exercises
Default Profile
workspace
repository
rrp-workshop-exercises
branch
main
Undo
Redo
Pull
Push
Branch
Stash
Pop
Terminal
Viewing 3/3
Show All
Filter (⌘ + Option + f)
LOCAL 1/1
main 1↑
REMOTE 2/2
origin
jashapiro
plotting-notebook
main
upstream
PULL REQUESTS 1
ISSUES
TEAMS
Team: Select a team
TAGS 0/0
SUBMODULES 0
GITHUB ACTIONS 0
BRANCH / TAG
GRAPH
COMMIT MESSAGE
// WIP + 2
main
Add script to download fastq files
12 hours ago
jashapiro/plotti...
add ggsave comment. & rerender
Add plot export size & other updat...
Apply suggestions from code review
5 days ago
add scripts dir readme
Update comments and update paths
Add notes on packages and `::`
Move analysis files to subfolder
Apply wording suggestions from co...
6 days ago
Add plotting notebook
main
Merge pull request #3 from AlexsL...
Update scripts/setup/README.md
minor comment fix
Add a few more extensive comments
add setup readme
Update data/reference/README.md
Add maf files
Add rproject
Expand comments
a week ago
rename count script
2 file changes in working directory
View changes
commit: 665cc5
Add script to download fastq files
Script curls and saves NC16_S1_L004_R1_001.fastq.gz and NC16_S1_L004_R2_001.fastq.gz from SRA study ID SRP255885
Stephanie J. Spielman
parent: 103256
authored 6/1/2022 @ 11:59 AM
+ 1 added
Path
Tree
View all files
+ scripts/download-fastq.sh
100%
Feedback
PRO
8.5.0


rrp-workshop-exercises
Default Profile
workspace
repository
rrp-workshop-exercises
branch
main
Undo
Redo
Pull
Push
Branch
Stash
Pop
Terminal
Viewing 3/3
Show All
Filter (⌘ + Option + f)
LOCAL
main
REMOTE 2/2
origin
jashapiro
plotting-notebook
main
upstream
PULL REQUESTS 1
ISSUES
TEAMS
Team: Select a team
TAGS 0/0
SUBMODULES 0
GITHUB ACTIONS 0
BRANCH / TAG
GRAPH
COMMIT MESSAGE
// WIP + 2
main
jashapiro/plotti...
Add script to download fastq files
12 hours ago
add ggsave comment. & rerender
Add plot export size & other updat...
Apply suggestions from code review
5 days ago
add scripts dir readme
Update comments and update paths
Add notes on packages and `::`
Move analysis files to subfolder
Apply wording suggestions from co...
6 days ago
Add plotting notebook
Merge pull request #3 from AlexsL...
Update scripts/setup/README.md
minor comment fix
Add a few more extensive comments
add setup readme
Update data/reference/README.md
Add maf files
Add rproject
Expand comments
a week ago
rename count script
2 file changes in working directory
View changes
commit: 665cc5
Add script to download fastq files
Script curls and saves NC16_S1_L004_R1_001.fastq.gz and NC16_S1_L004_R2_001.fastq.gz from SRA study ID SRP255885
Stephanie J. Spielman
authored 6/1/2022 @ 11:59 AM
parent: 103256
+ 1 added
Path
Tree
View all files
+ scripts/download-fastq.sh
100%
Feedback
PRO
8.5.0

We'll see how to do this from the command line soon!

Not all files are meant to live under version control

- Sometimes we have files inside the repository directory that we very much *do not want* to be part of version control
  - Large files; GitHub cannot generally handle files >= 100 MB
  - Unreleased/private data files, including PII/PHI data - *even in a private repository, don't do it!*
  - Other sneaky culprits like `.DS_Store` files on MacOS

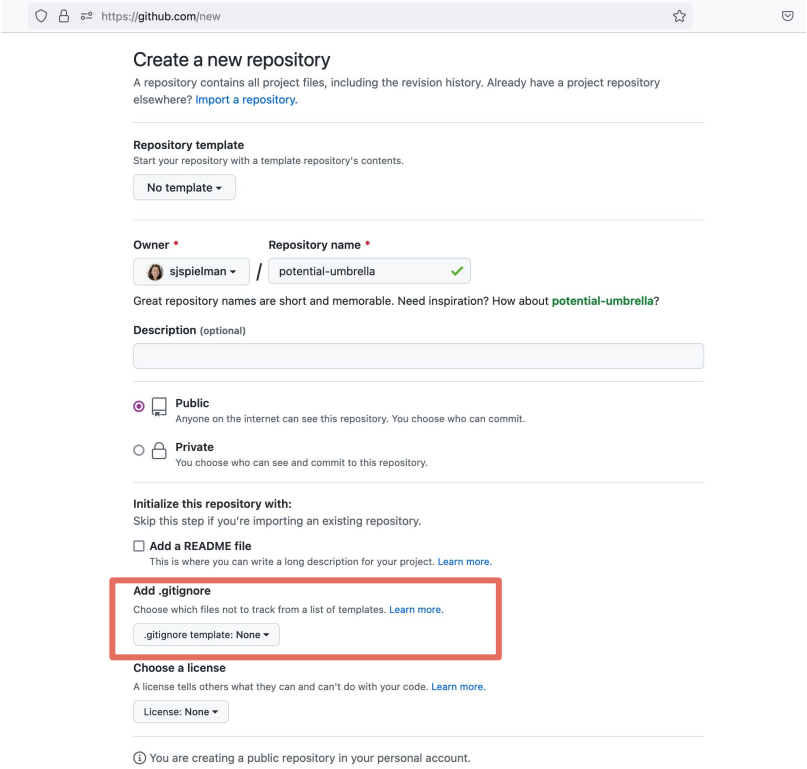
- It's pretty easy to mess up and accidentally stage/commit these files, and Git loves to complain about unstaged files

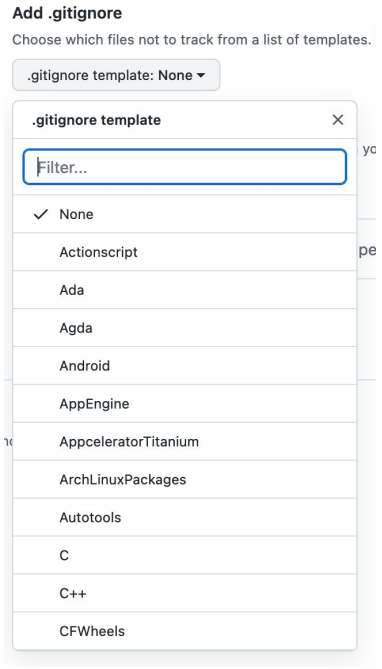
Using `.gitignore` files

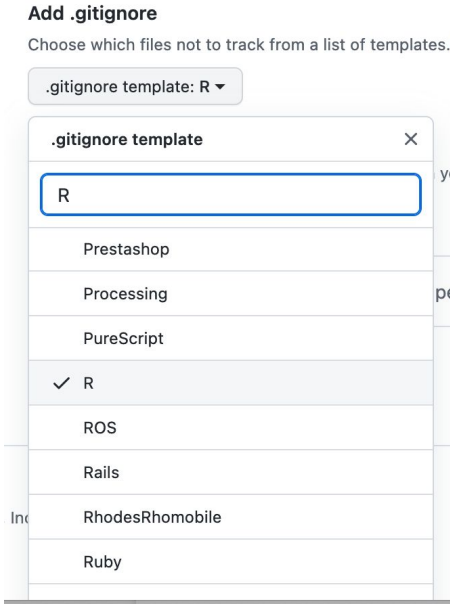
- `.gitignore` files are *hidden files* that tell `git` to ignore certain files
  - `.gitignore` will make Git stop telling you about untracked changes
  - `.gitignore` will prevent you from staging/committing these files by accident
    - 🚨Word to the wise:🚨 Include files ASAP in your .gitignore before you accidentally commit them! (Trust us, we speak from experience….)

- You can have many `.gitignore` files in a repo (in different directories) and/or a single `.gitignore` at the top of your repo that all subdirectories "inherit"
  - You can also have a *global* file in `~/.gitignore` that will apply to all repos on your computer (Requires a little more setup, though)!
  - But caution: Your collaborators probably don't have that file.

github.com has some useful templates







Now, we'll make a `.gitignore` file to also ignore `data/raw/fastq`