



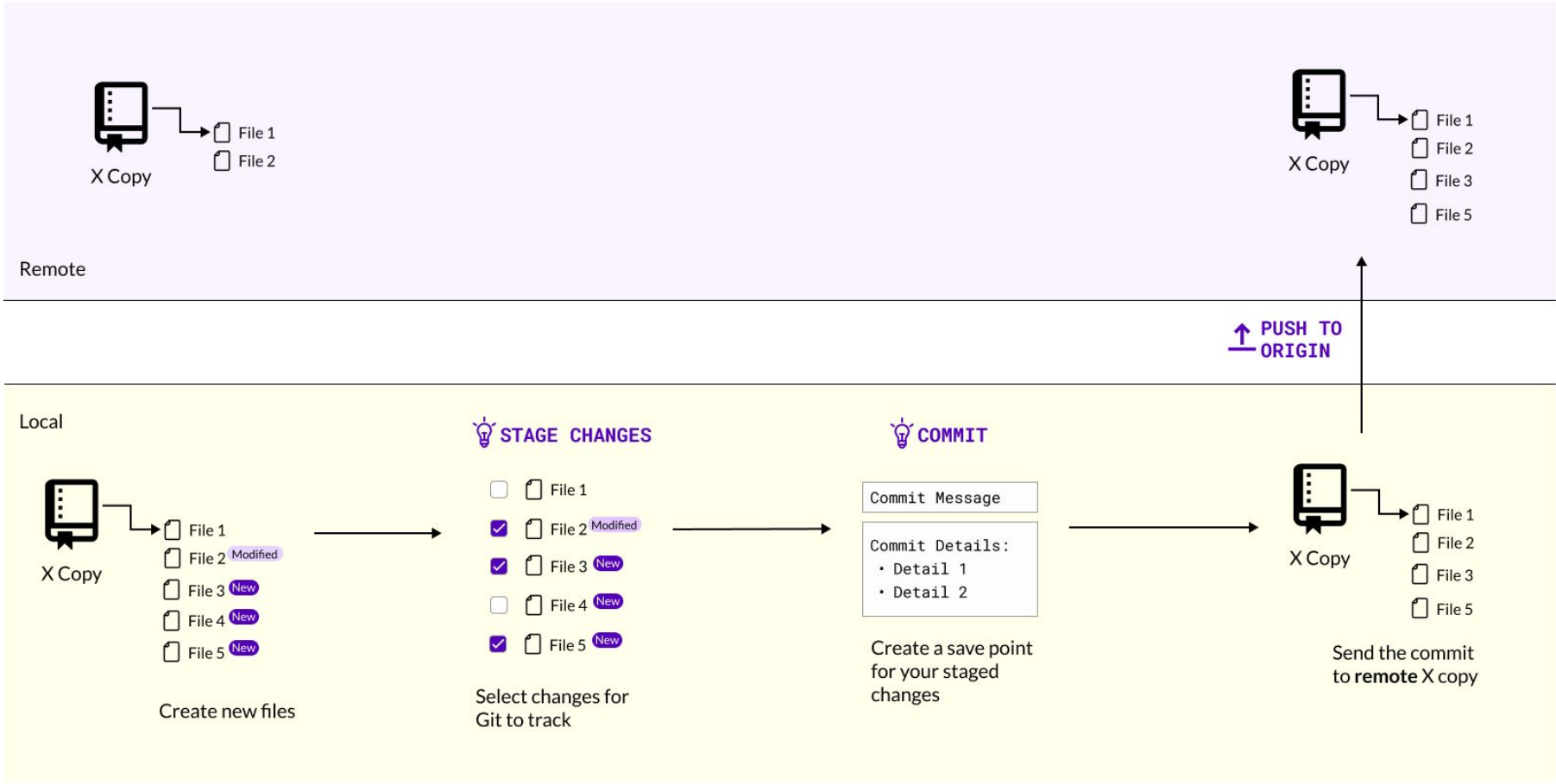
The Git Stage/Commit/Push workflow

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





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



	COMMENT	DATE
○	CREATED MAIN LOOP & TIMING CONTROL	14 HOURS AGO
○	ENABLED CONFIG FILE PARSING	9 HOURS AGO
○	MISC BUGFIXES	5 HOURS AGO
○	CODE ADDITIONS/EDITS	4 HOURS AGO
○	MORE CODE	4 HOURS AGO
○	HERE HAVE CODE	4 HOURS AGO
○	AAAAAAA	3 HOURS AGO
○	ADKFJSLKDFJSOKLFJ	3 HOURS AGO
○	MY HANDS ARE TYPING WORDS	2 HOURS AGO
○	HAAAAAAAANDS	2 HOURS AGO

AS A PROJECT DRAGS ON, MY GIT COMMIT MESSAGES GET LESS AND LESS INFORMATIVE.

The screenshot displays the GitHub Desktop application interface for the repository `rrp-workshop-exercises`. The main window shows a commit history for the `main` branch. The current commit is `// WIP` with 3 changes, committed 12 hours ago by `jashapiro/plotti...`. The commit message is `add ggsave comment. & re-render`. The diff view on the right shows 3 unstage files: `data/raw/fa.../NC16_S1_L004_R1_001.fastq.gz`, `data/raw/fa.../NC16_S1_L004_R2_001.fastq.gz`, and `scripts/download-fastq.sh`. The interface includes a sidebar with repository navigation, a top toolbar with actions like Pull, Push, and Branch, and a bottom status bar with system icons and version information (8.5.0).

Repository: `rrp-workshop-exercises` (main branch)

Commit: `// WIP` (+ 3 changes) - 12 hours ago

Commit Message: `add ggsave comment. & re-render`

Unstaged Files (3):

- `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

rrp-workshop-exercises x +

workspace repository branch Undo Redo Pull Push Branch Stash Pop Terminal

rrp-workshop-exercises main

+ scripts/download-fastq.sh Stage File x

3 file changes on main

Edit this file Unstaged File View Diff View

```
@@ -1,1 +1,47 @@
1
1+#!/bin/bash
2+set -euo pipefail
3+
4+# Define study ID
5+STUDY_ID="SRP255885"
6+
7+# Define file names
8+FASTQ_R1="NC16_S1_L004_R1_001.fastq.gz"
9+FASTQ_R2="NC16_S1_L004_R2_001.fastq.gz"
10+FASTQ_URL="https://sra-download.ncbi.nlm.nih.gov/traces/sra63/SRZ/011518/SRR11518889"
11+
12+# Define and create destination directory for FASTQ files to live in
13+FASTQ_DEST="./data/raw/fastq/$STUDY_ID/"
14+mkdir -p $FASTQ_DEST
15+
16+#### Process the R1 file ####
17+
18+# Print an indicator:
19+echo "Obtaining $FASTQ_R1"
20+
21+# Curl the file (using one of several approaches)
22+curl -O $FASTQ_URL/$FASTQ_R1 # this approach preserves the original internet file name
23+
24+# Explore: how many lines are in the file?
25+echo "The number of lines in $FASTQ_R1 is:"
26+gunzip -c $FASTQ_R1 | wc -l
27+
```

Click on a changed file to see its diff view

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

Feedback PRO 8.5.0

The screenshot displays the GitHub Desktop application interface. On the left, a sidebar shows the repository structure with the 'main' branch selected. The central area features a commit history graph with a vertical timeline of commits. A red text overlay reads: "We stage files to 'add' them to the staging area." A red arrow points from this text to the 'scripts/download-fastq.sh' file in the 'Unstaged Files' list on the right. Another red arrow points from the 'Stage File' button to the 'Commit Message' field. The 'Commit Message' field contains a 'Summary' and a 'Description' section. The bottom status bar shows '100%' zoom, a 'Feedback' button, and 'PRO 8.5.0' version information.

workspace repository rrp-workshop-exercises +

workspac repository rrp-workshop-exercises > 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... scripts/download-fastq.sh
- Add ggsave comment. & rerender
- Add plot export size & other updat...
- Apply suggestions from code review
- add scripts dir readme
- Update comments and update paths
- Add notes on packages and ...
- Move analysis files to subfolder
- Apply wording suggestions from co...
- 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
- rename count script
- Add flags download script & updat...

3 file changes on main

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 Stage File

Staged Files (0)

Commit Message Amend

Summary

Description

Stage files/changes to commit

100% Feedback PRO 8.5.0

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

- add ggsave comment. & rerender
- Add plot export size & other updat...
- Apply suggestions from code review
- add scripts dir readme
- Update comments and update paths
- Add notes on packages and `::`
- Move analysis files to subfolder
- Apply wording suggestions from co...
- 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
- rename count script
- Add flags download script & updat...

3 file changes on **main**

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

Summary Description

Type a message to commit

Feedback PRO 8.5.0

Now we are ready to commit these staged changes.

The screenshot shows the GitHub Desktop application interface. The top bar displays the repository name "rrp-workshop-exercises" and the current branch "main". The left sidebar shows the repository structure, including local and remote branches, and a list of pull requests and issues. The main area displays a commit history graph with a list of commit messages. The right sidebar shows the "3 file changes on main" section, listing unstaged and staged files. The commit message editor is open, showing the message "Add script to download fastq files" and a detailed description of the changes. A red arrow points to the "Commit changes to 1 file" button at the bottom of the commit message editor.

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

- 3 file changes on **main**
- 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

Feedback PRO 8.5.0

The screenshot displays the GitHub Desktop application interface. At the top, the repository is identified as `rrp-workshop-exercises` on the `main` branch. The `Push` button is highlighted with a red box. The left sidebar shows the local repository structure, with the `main` branch selected. The central area shows a commit graph with the current commit highlighted. The right sidebar shows the commit details, including the commit message `Add script to download fastq files` and the author `Stephanie J. Spielman`. The commit message content is visible in the right sidebar, showing the script's purpose and the files added.

workspace repository `rrp-workshop-exercises` branch `main` Undo Redo Pull Push Branch Stash Pop Terminal

Viewing 3/3 Show All BRANCH / TAG GRAPH COMMIT MESSAGE

Filter (% + Option + f) // WIP + 2

LOCAL 1/1
main 1 ↑
REMOTE 2/2
origin
jashapiro
plotting-notebook
main
upstream

main
Add script to download fastq files 12 hours ago
add ggsave comment. & re-render
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 parent: 103256
authored 6/1/2022 @ 11:59 AM

+ 1 added

scripts/download-fastq.sh

Feedback PRO 8.5.0

rrp-workshop-exercises

workspace repository branch
+ > rrp-workshop-exercises > main

Undo Redo Pull Push Branch Stash Pop Terminal

Viewing 3/3 Show All BRANCH / TAG GRAPH COMMIT MESSAGE

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

// WIP + 2

- Add script to download fastq files 12 hours ago
- add ggsave comment. & re-render
- 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 parent: 103256
authored 6/1/2022 @ 11:59 AM

+ 1 added

Path Tree View all files

+ scripts/download-fastq.sh

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

https://github.com/new

Create a new repository

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

Repository template

Start your repository with a template repository's contents.

No template

Owner

Repository name

 sjspielman /

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

Description (optional)

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:

Skip this step if you're importing an existing repository.

Add a README file

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

Add .gitignore

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

.gitignore template: None

Choose a license

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

License: None

 You are creating a public repository in your personal account.

Add .gitignore

Choose which files not to track from a list of templates.

.gitignore template: None

.gitignore template

Filter...

None

Actionscript

Ada

Agda

Android

AppEngine

AppceleratorTitanium

ArchLinuxPackages

Autotools

C

C++

CFWheels

Add .gitignore

Choose which files not to track from a list of templates.

.gitignore template: R

.gitignore template

R

Prestashop

Processing

PureScript

R

ROS

Rails

RhodesRhomobile

Ruby

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

