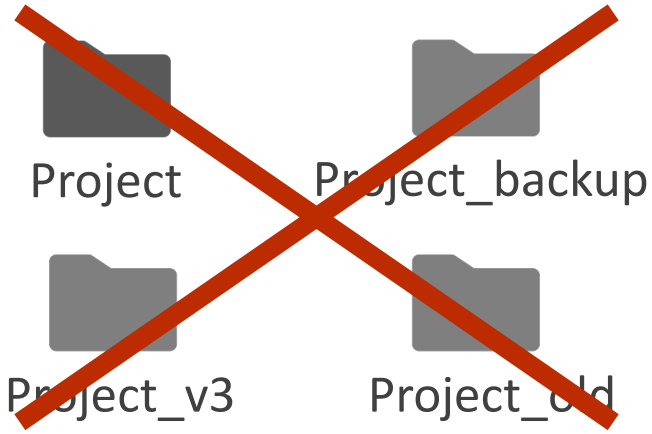
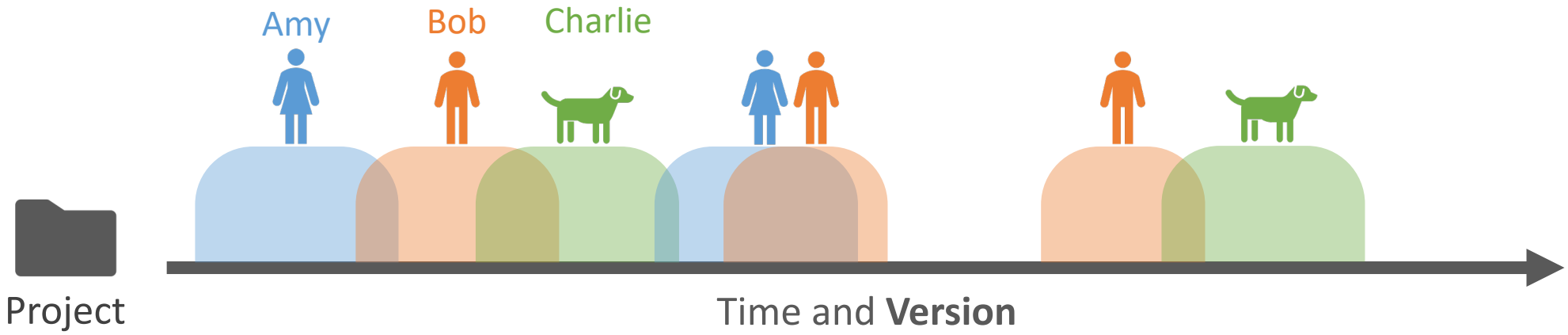
20 April 2022

Erasmus MC
University Medical Center Rotterdam

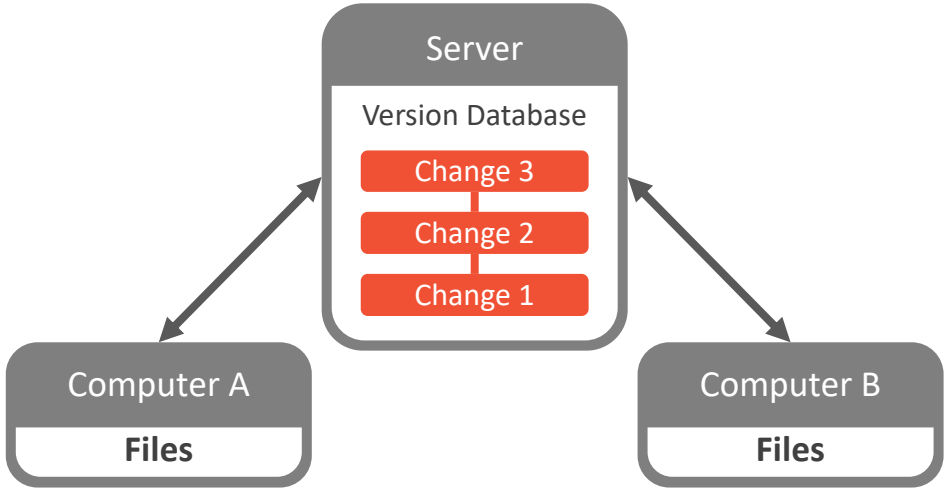
Version Control

- Tracks **changes** in computer files
- Efficient **collaboration** with colleagues



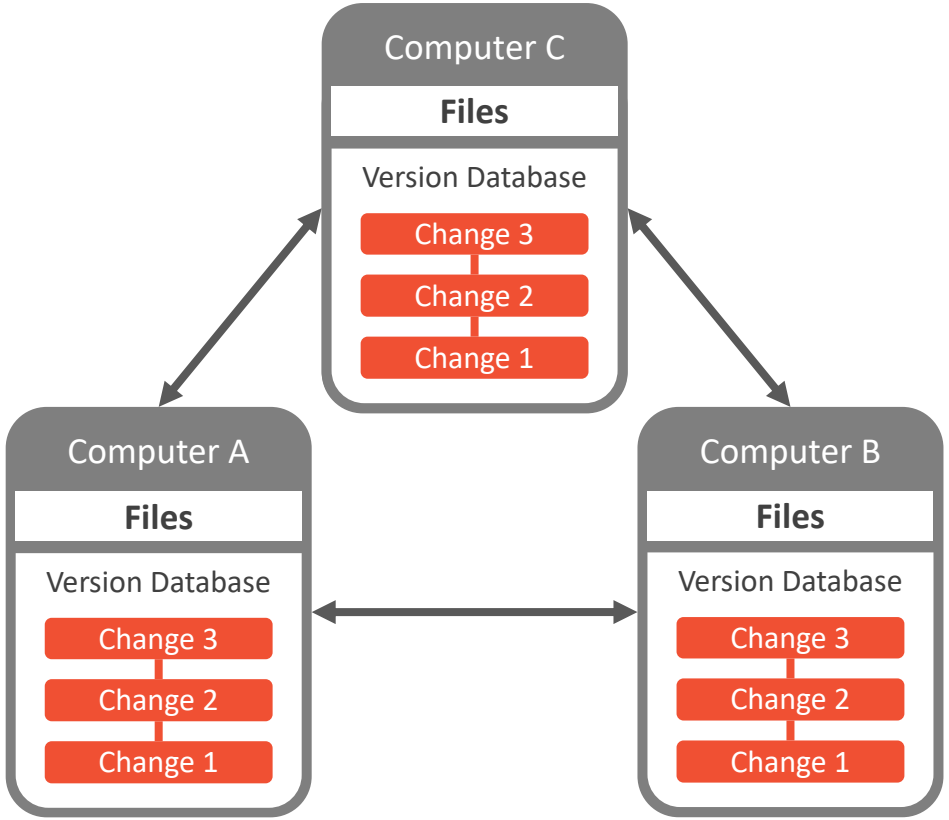


Version Control Systems



Centralized SVN, CVS, Perforce

+ manage access rights



Distributed Git, Mercurial

+ local version database

+ backup on collaborator machines

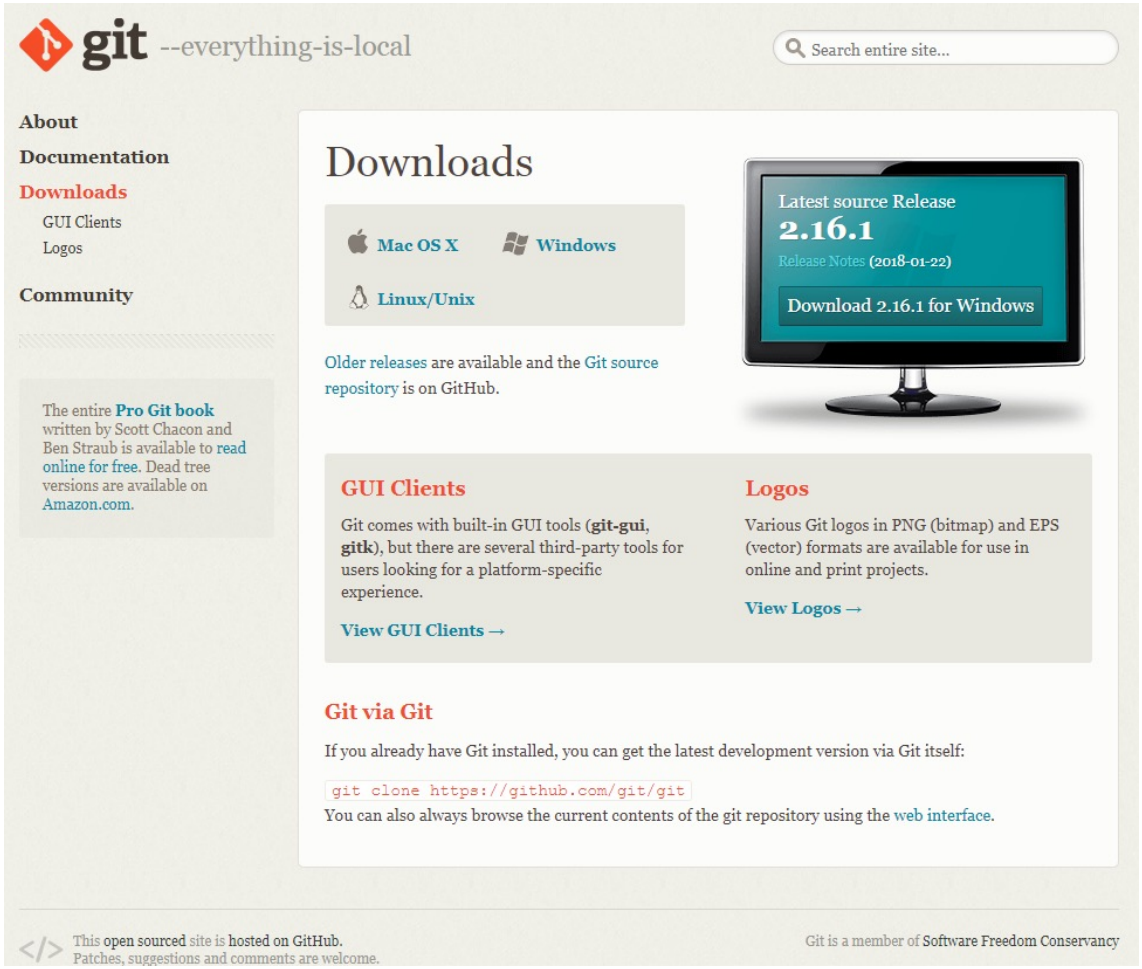
About Git

- Git is a **distributed** version control system for local and remote repositories

- Git is advertised with the following features
  - Offers **review** functionality
  - Work **offline** anywhere
  - **Fast** and lightweight
  - **Journal** of changes rather than a backup
  - Serves the needs of **beginners** and **advanced** users equally well

Installation

- **Download** at https://git-scm.com/downloads
- Accept **default** settings during installation



6

Interfaces

Command Line Interface (CLI)

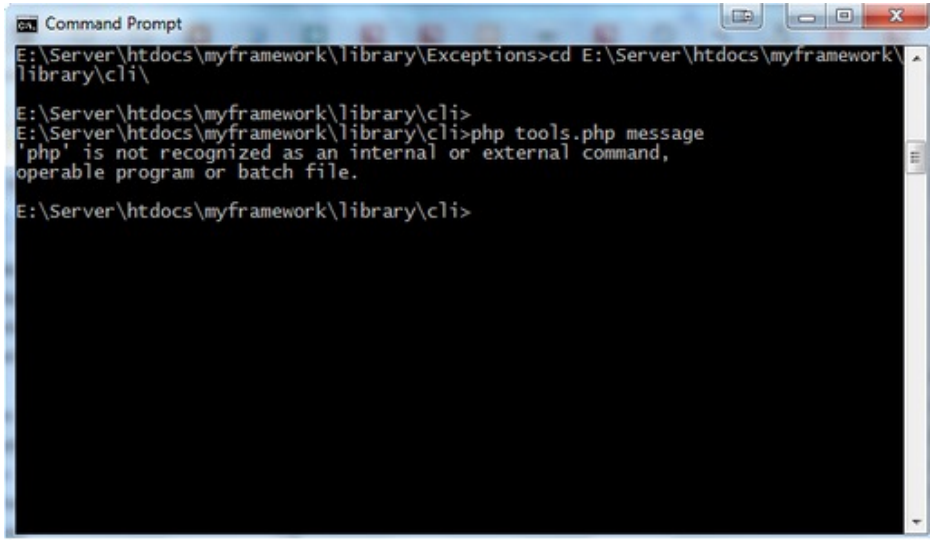
- Syntax: `git [verb] [options]`
- Allows for **customized** commands
- Commands are **platform-independent**
- **Fast**

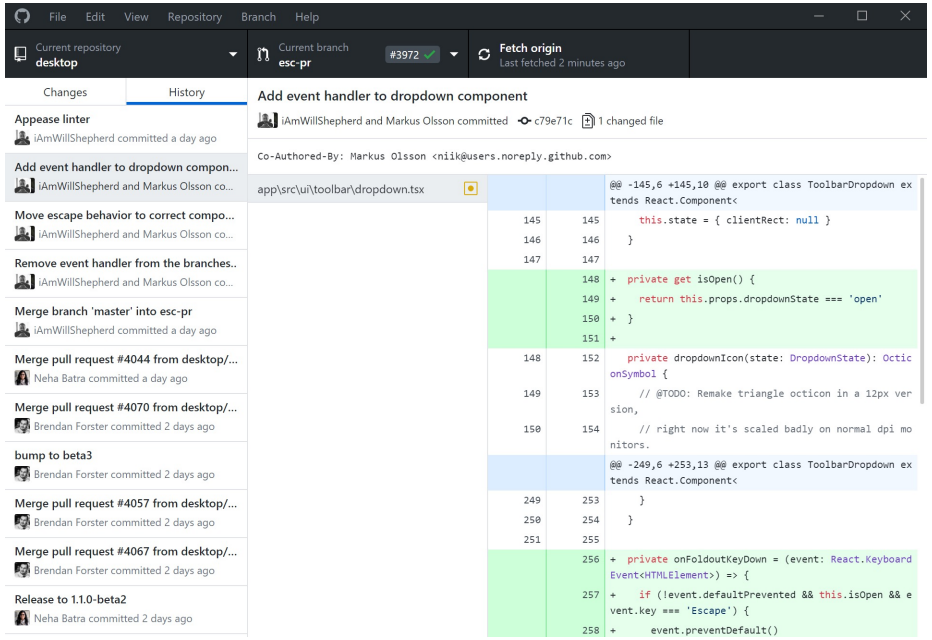
Graphical User Interface (GUI)

- Saves **typing**
- Lower **barrier-of-entry**
- (we recommend GitHub Desktop or Sourcetree)

Integrated Development Environment (IDE)

- Integrates version control with programming **projects** e.g., generates .gitignore.





7

Configure Git – CLI

Verify installed git **version**

```
git --version
```

Configure git on a `--system`, `--global` or `--local` level

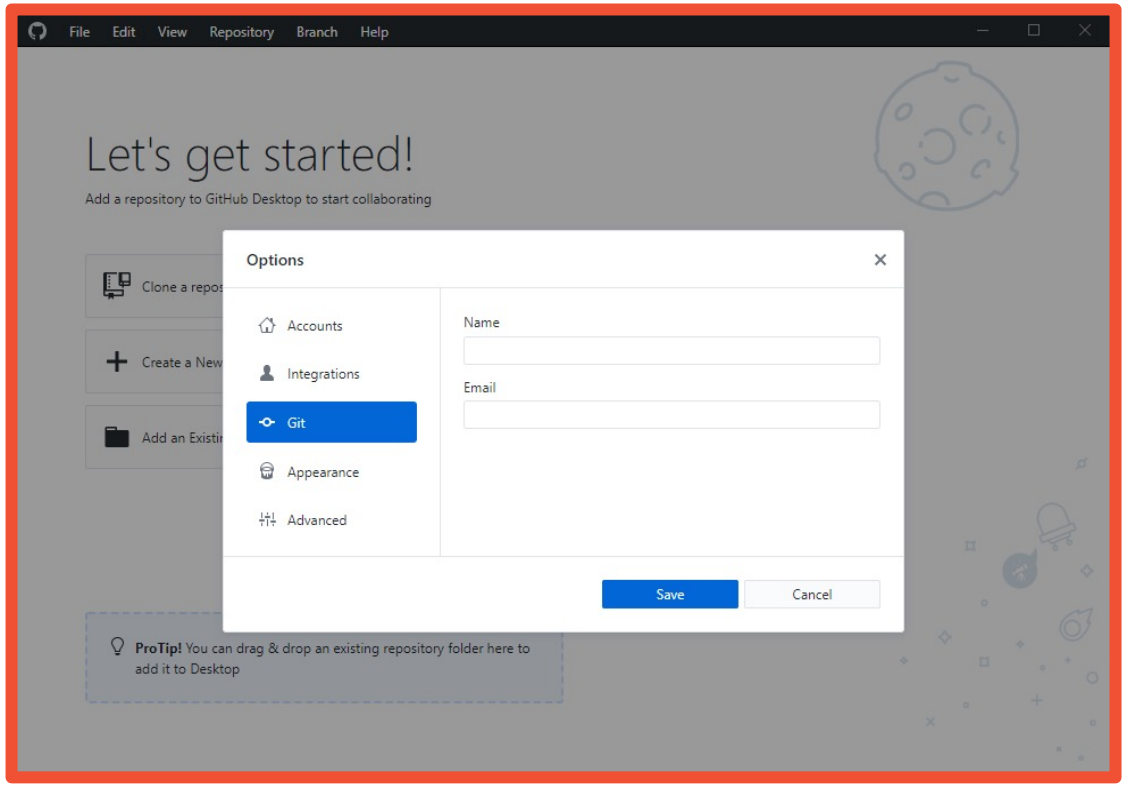
```
git config --global user.name "your name"
git config --global user.email "my@mail.nl"
git config --global --list
```

Open **documentation** of verb (e.g. `config`)

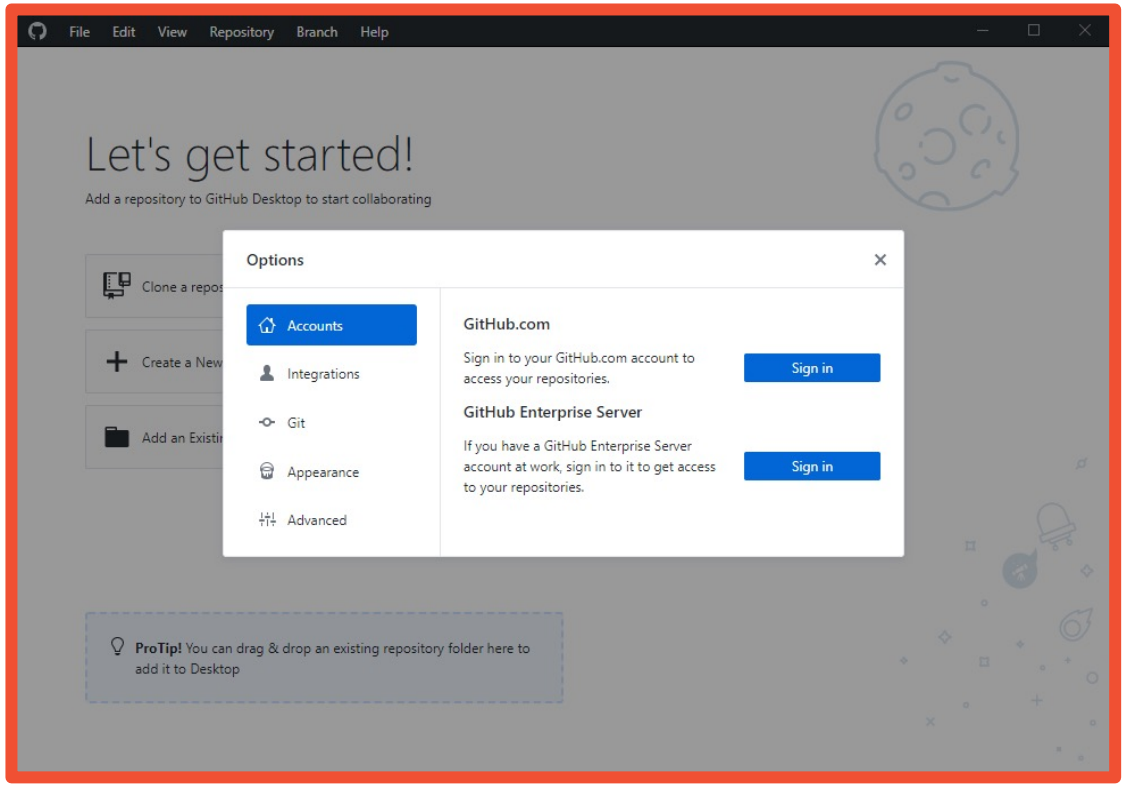
```
git help [verb]
git [verb] -help
git config --global --unset user.name
```

Configure Git – GUI (GitHub Desktop)

File → Options → Git



File → Options → Accounts



Local Repositories – CLI

The local repository is the project folder under version control. It contains the .git subfolder (hidden by default) and all project related files and subfolders.

Initialize local **repository**

```
git init [project-name]
```

Check s**tatus** of local repository

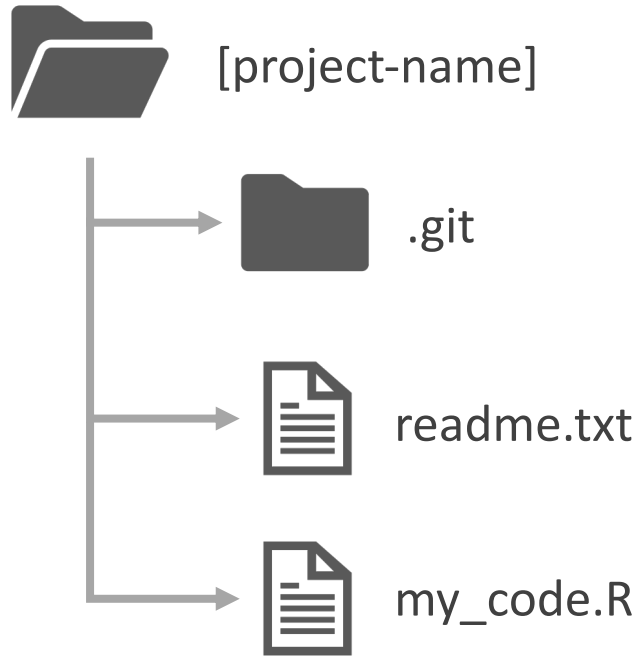
```
git status
```

Create empty file on Windows

```
copy NUL [file-name]
```

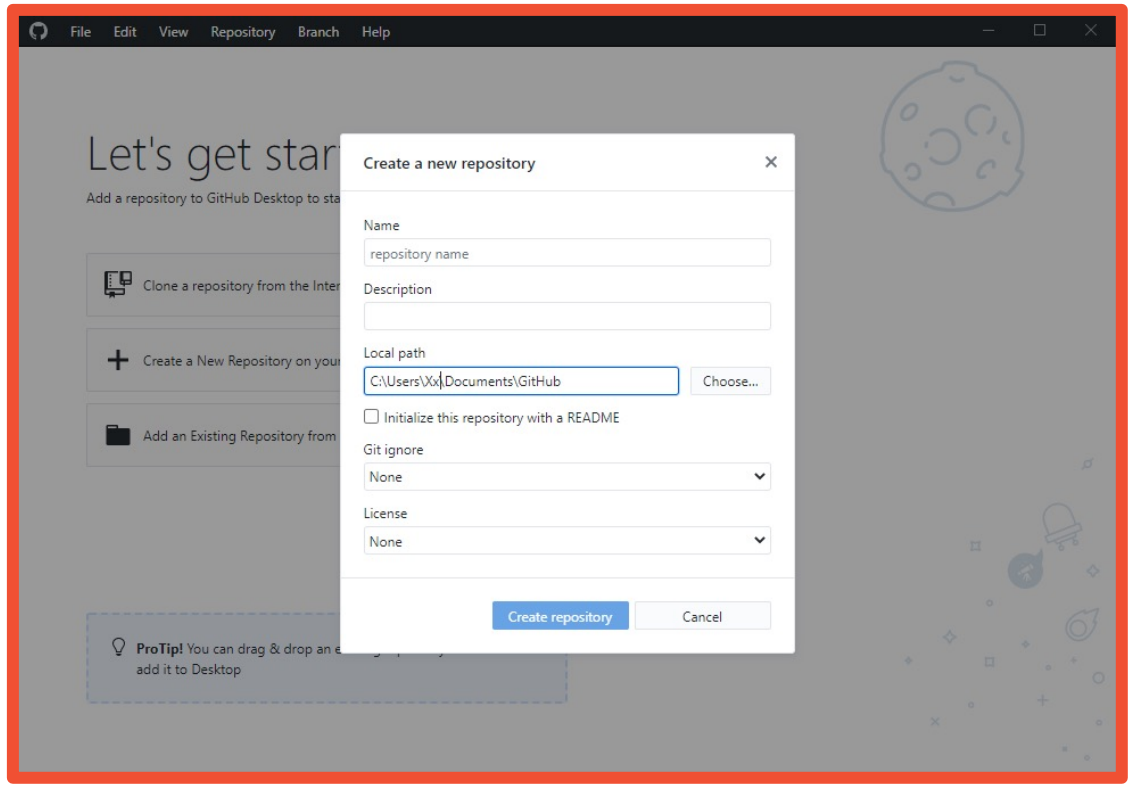
Create empty file on macOS, Linux

```
touch [file-name]
```

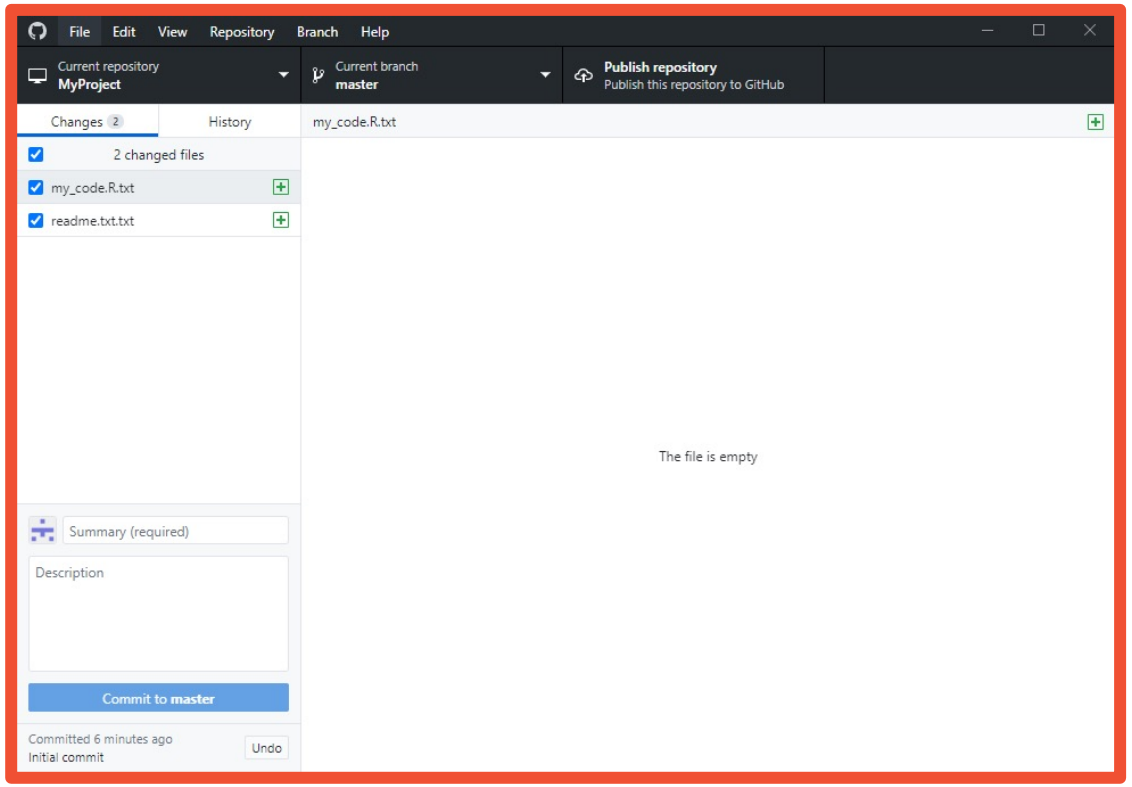


Local Repositories – GUI (GitHub Desktop)

File → New repository



Changes are detected with a background call to `git status`



File Status Lifecycle

Stage untracked/modified file

```
git add [file-name]
```

Commit staged files

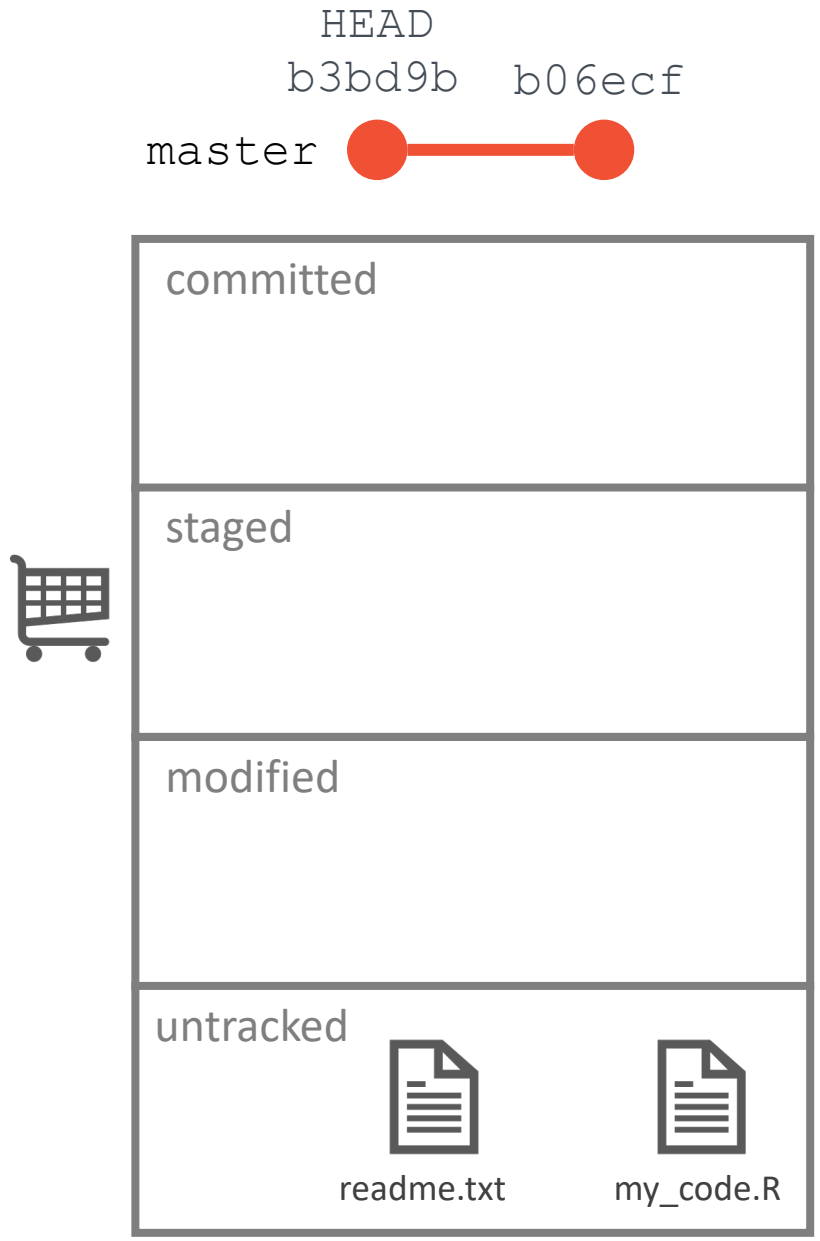
```
git commit -m "[message]"
```

Show **modification**

```
git diff HEAD
git diff [commit] [commit]
```

Stage all files

```
git add -A
```



Log – CLI

Show **logged** commits

```
git log
```

Show command **options**

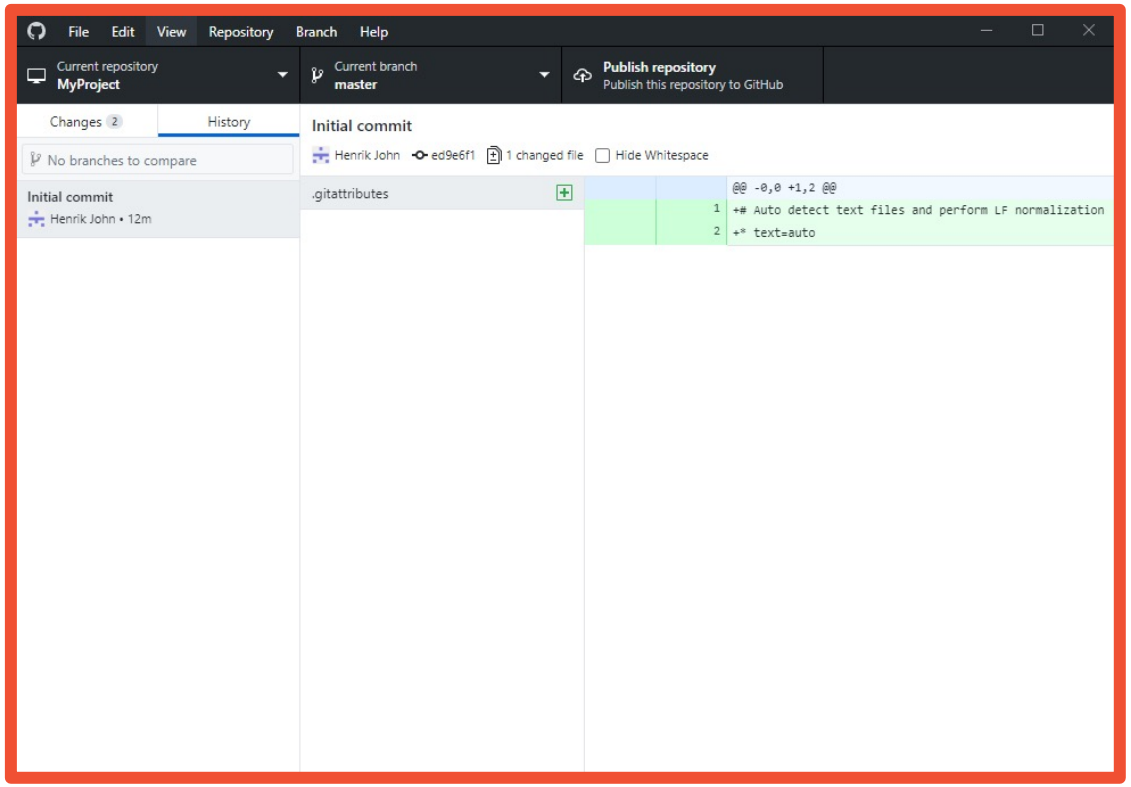
```
git help log
```

Show **formatted** log message and **graph.** Also, check out https://tinyurl.com/pretty-log.

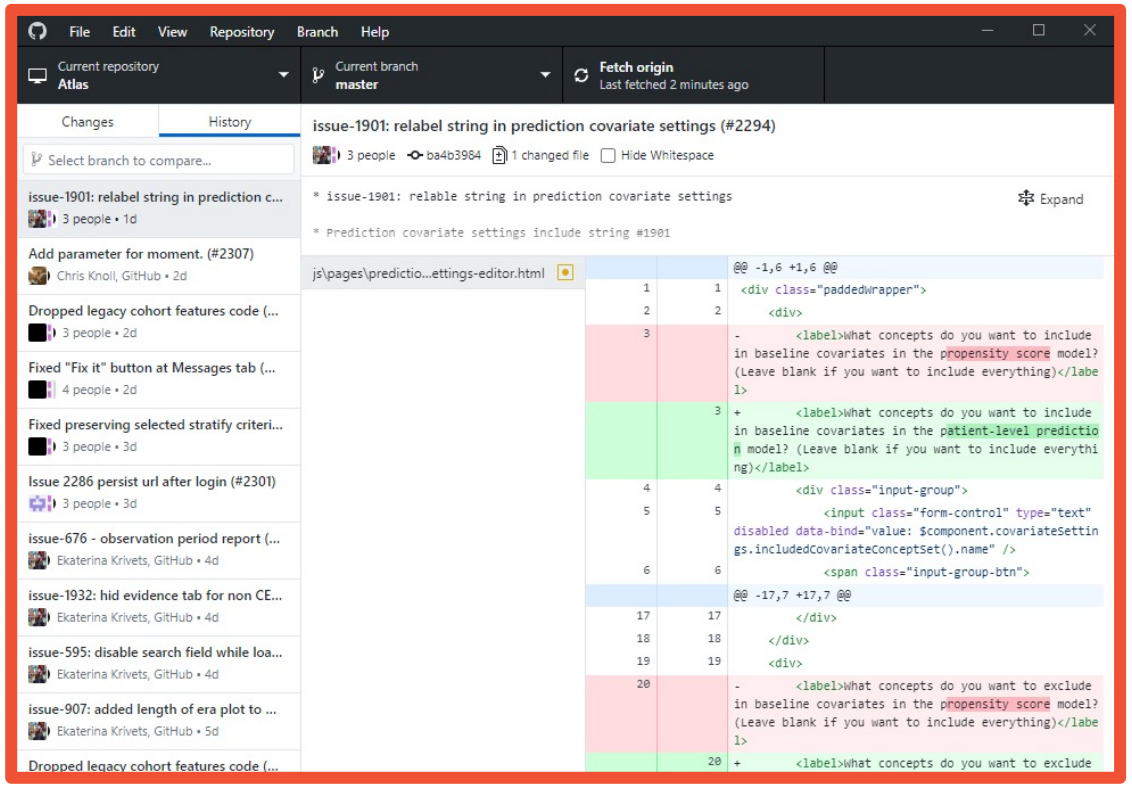
```
git log --graph --oneline
```

Log – GUI (GitHub Desktop)

View → History



History of the OHDSI Atlas repository



Commit message

- Use imperative mode for summary line. Start with "Fix", "Add", instead of "Fixed", "Added".
- Leave the second line blank
- Line break the commit message (at about 72 characters)
- Don't end the summary line with a period, as it is a title

```
Short (50 chars or less) summary of changes

More detailed explanatory text, if necessary. Wrap it to
about 72 characters or so. In some contexts, the first
line is treated as the subject of an email and the rest
of the text as the body. The blank line separating the
summary from the body is critical (unless you omit the
body entirely); tools like rebase can get confused if you
run the two together.

Further paragraphs come after blank lines.

- Bullet points are okay, too

- Typically a hyphen or asterisk is used for the bullet,
  preceded by a single space, with blank lines in
  between, but conventions vary here
```

Custom editor and tools

Git ships with **editors** and **tools** that are mostly command line based. While these tools are fast, they may not offer functionality or usability that suits all users and can, therefore, be changed.

Default core editor is `vim` , but can be changed, e.g., to *Atom*.

```
git config --global core.editor "atom -w"
```

Default `difftool` can be set, e.g., to *Meld* or *Beyond Compare*.

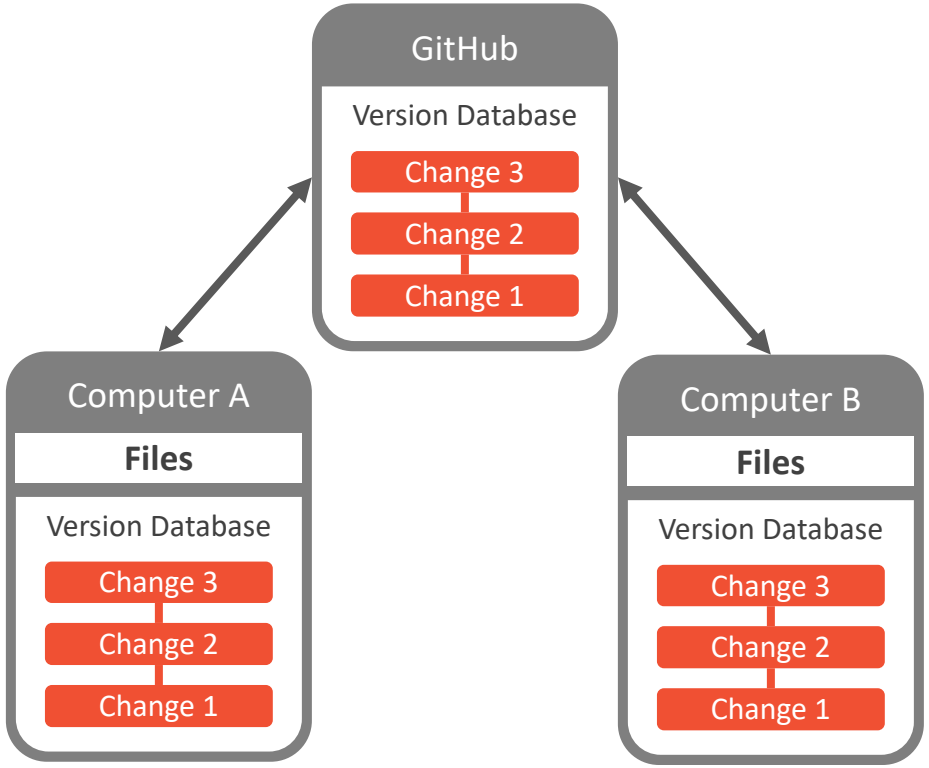
```
git config --global diff.tool "meld”
git difftool
```

Default `mergetool` can be set, e.g., to *Meld* or *Beyond Compare*.

```
git config --global merge.tool "meld”
```

Remote Repositories

- **Bare repository** only contains version database (e.g., GitHub)
- **Working repository** generates working directory from version database



Distributed Version Control System

18

Remote Repositories – CLI

A project copy, which is hosted on a remote server, e.g., **https://github.com/mi-erasmusmc/GitTutorial**

Clone a remote repository

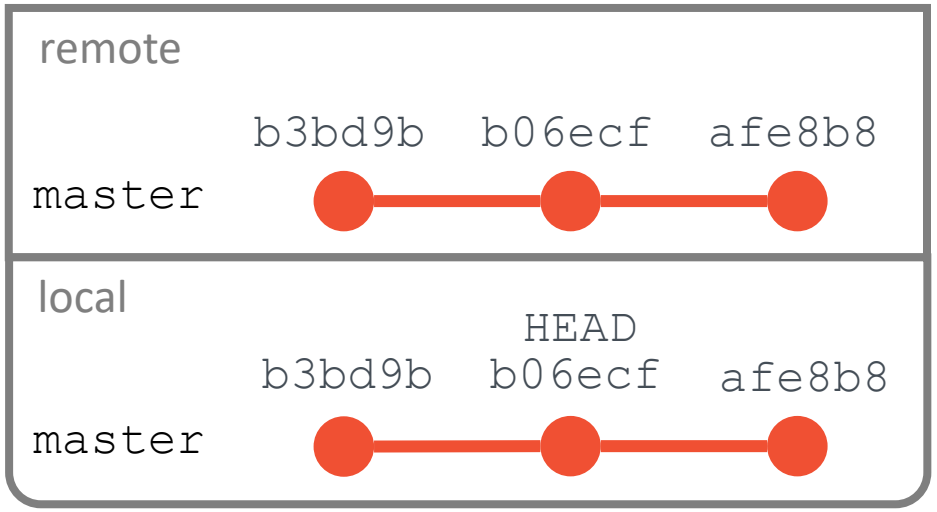
```
git clone [url]
```

Lists remote repository I**nformation**

```
git remote -v
git branch -a
```

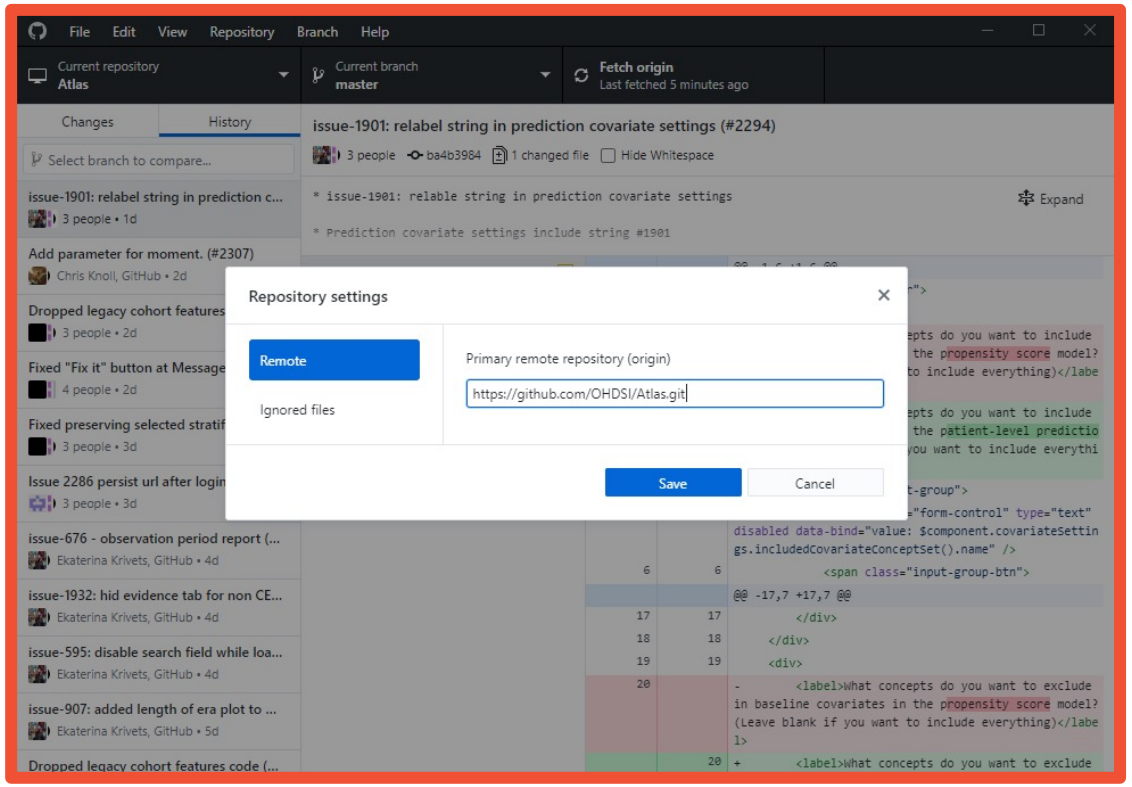
Push changes to remote repository

```
git pull
git push
```



Remote Repositories – GUI (GitHub Desktop)

Repository → Repository settings
→ Remote



Branches – CLI

Diverge from the main line of development to work independently.

List local branches

```
git branch
```

Create a branch

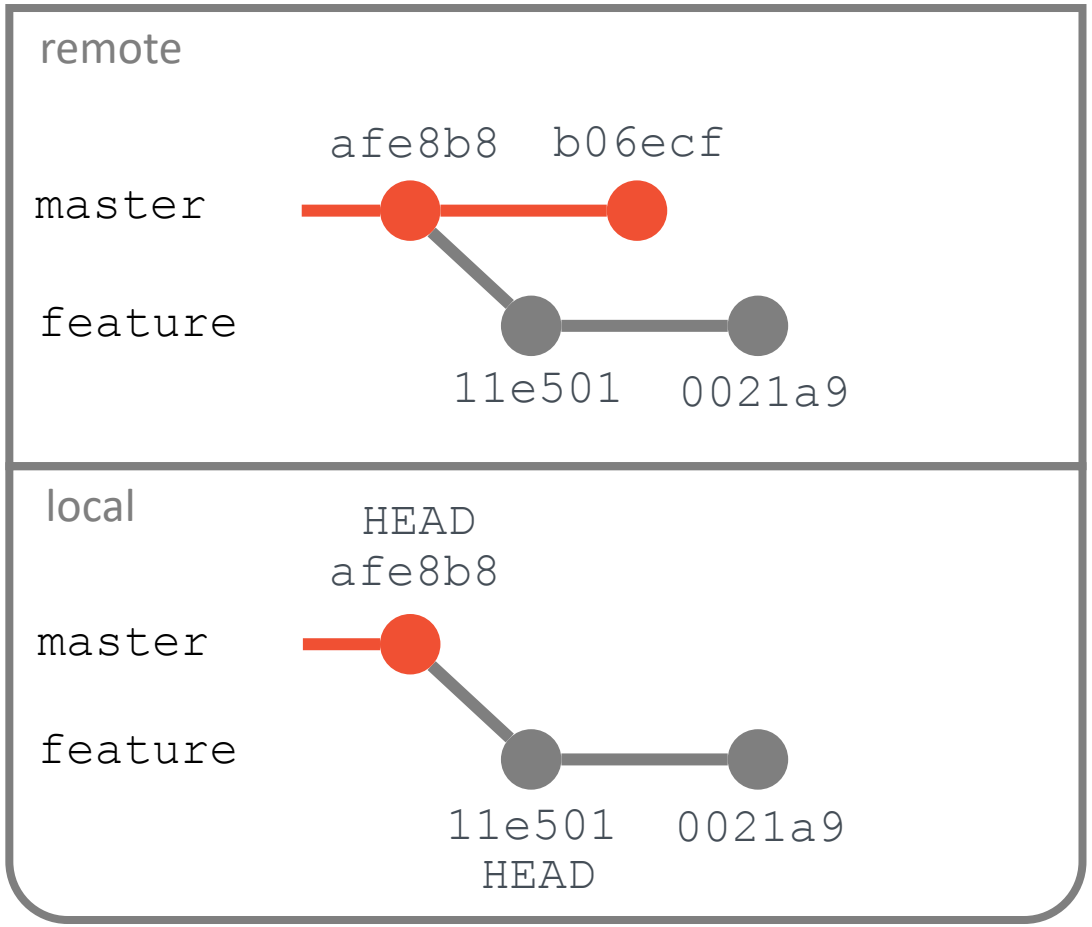
```
git branch [branch-name]
```

Enter a branch

```
git checkout [branch-name]
```

Push branch to remote repository

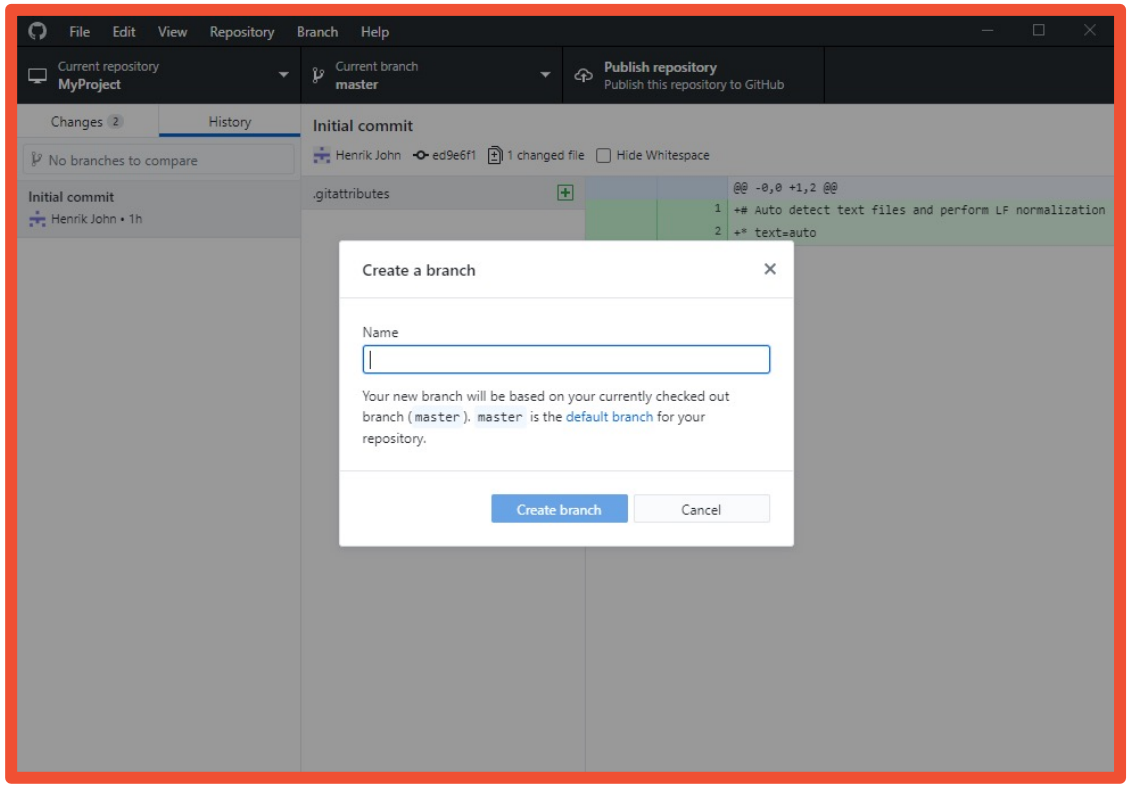
```
git push -u origin [branch-name]
```



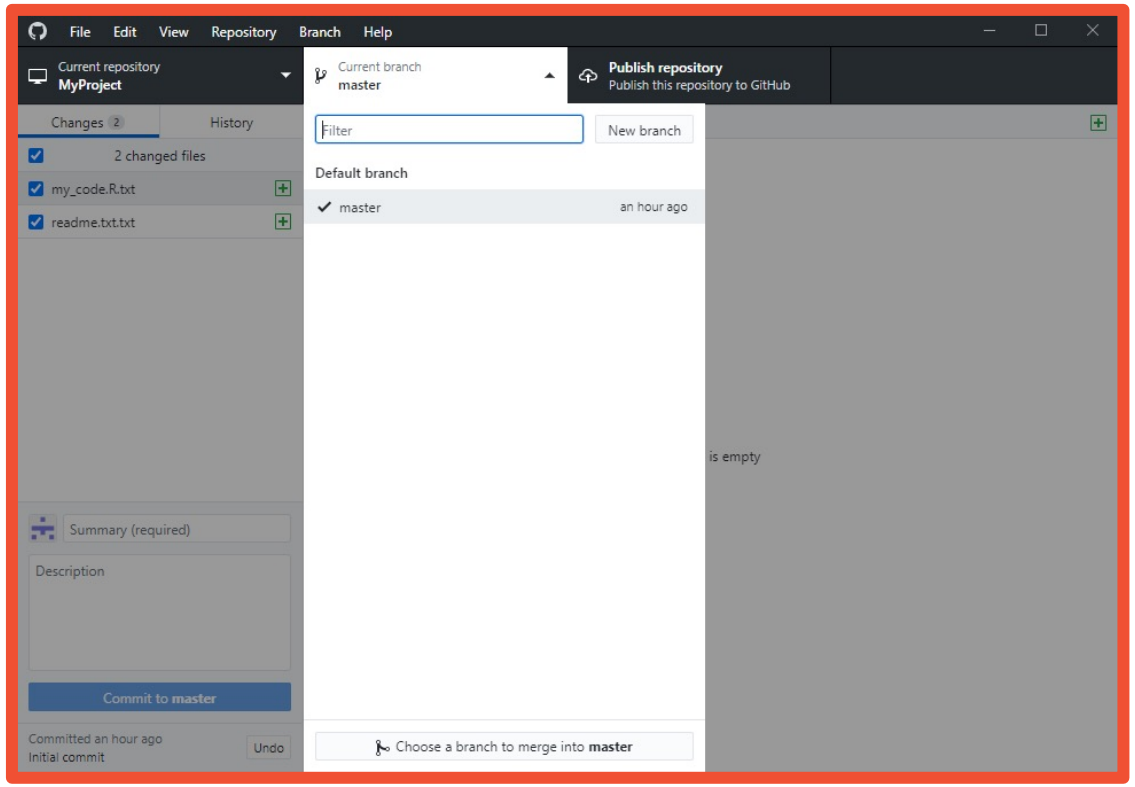
21

Branches – GUI (GitHub Desktop)

Branch → New branch



Checkout branches from the main window



22

Merging – CLI

Combine two lines of development after independent work is completed.

Pull **Master** branch changes

```
git checkout master
git pull
```

Show **Merged** branches

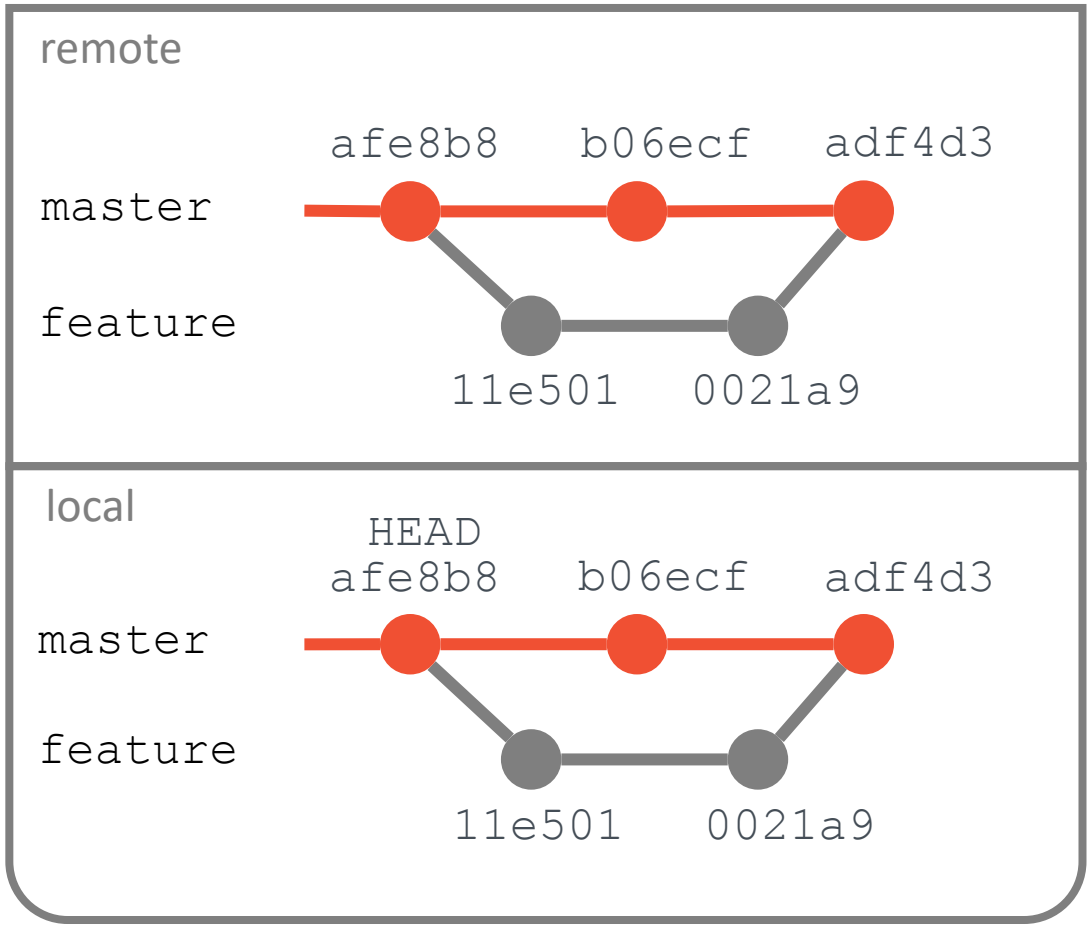
```
git branch --merged
```

Merge feature branch into master

```
git merge [branch-name]
```

Push merge to remote repository

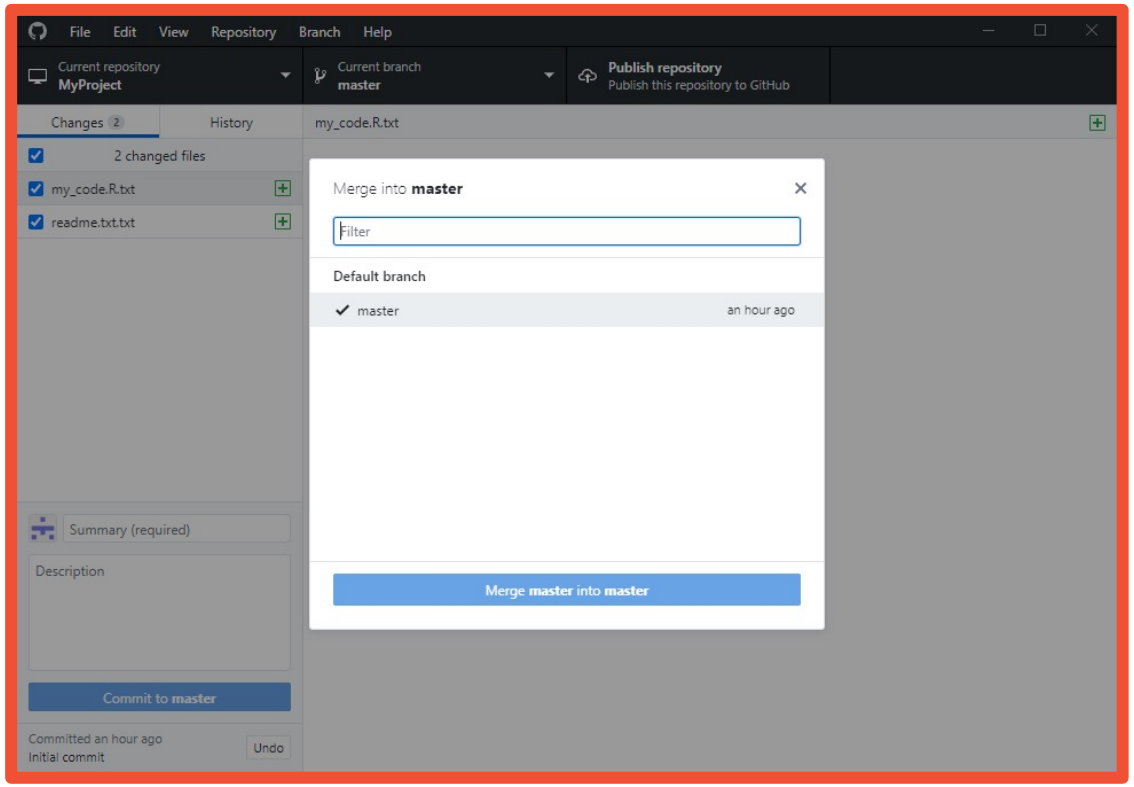
```
git push
```



23

Merging – GUI (GitHub Desktop)

Branch → Merge into current branch



24

Merge Conflicts

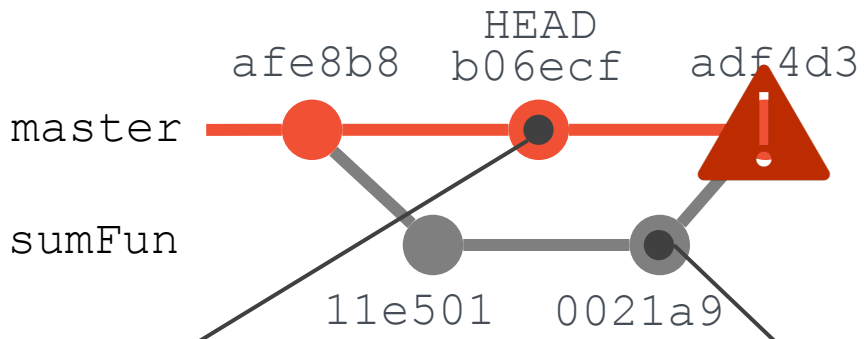
```
# Take sum of vector
sum <- function(x){
  sum <- 0
```

```
# Take sum of vector
sum <- function(x){
<<<<<<< HEAD
  stop ("x has NA values")
=======
  if (any(is.na(x)) {
    stop ("x has NA values")
>>>>>>> sumFun
  }

  for (val in x) {
    sum <- sum + val
  }
  sum
}
```

Amy

sum.R




```
git merge sumFun
```

```
Auto-merging sum.R

CONFLICT (content): merge
conflict in sum.R

Automatic merge failed; fix
conflicts and then commit the
result.
```

```
git mergetool
```

```
# Take sum of vector
sum <- function(x){
  sum <- 0

  if (any(is.na(x)) {
    stop ("x contains NA")
  }

  for (val in x) {
    sum <- sum + val
  }
  sum
}
```

Bob

sum.R

Removing Branch

Show **Merged** branches

```
git branch --merged
```

Remove branch

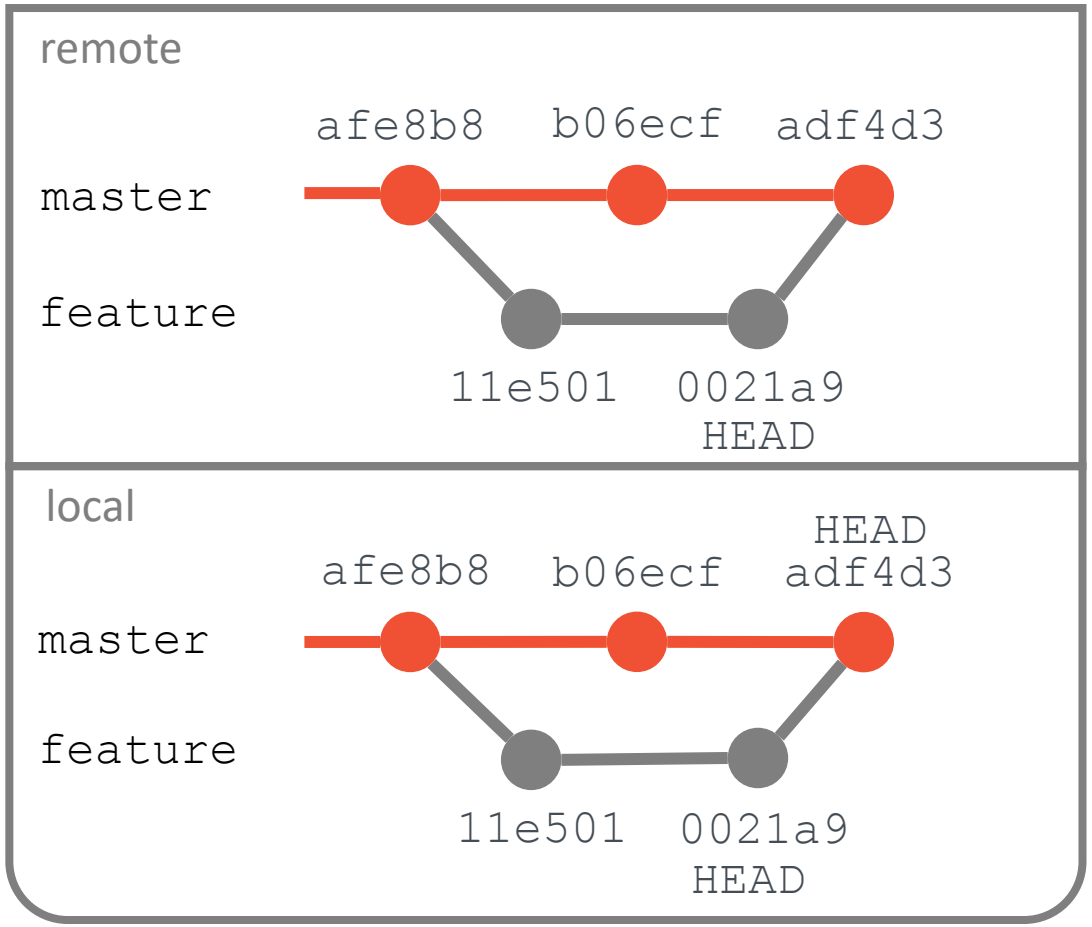
```
git branch -d [branch-name]
```

List **remote** branches

```
git branch -a
```

Remove branch from **remote** repository

```
git push origin -delete [branch-name]
```



26

Rebase – CLI

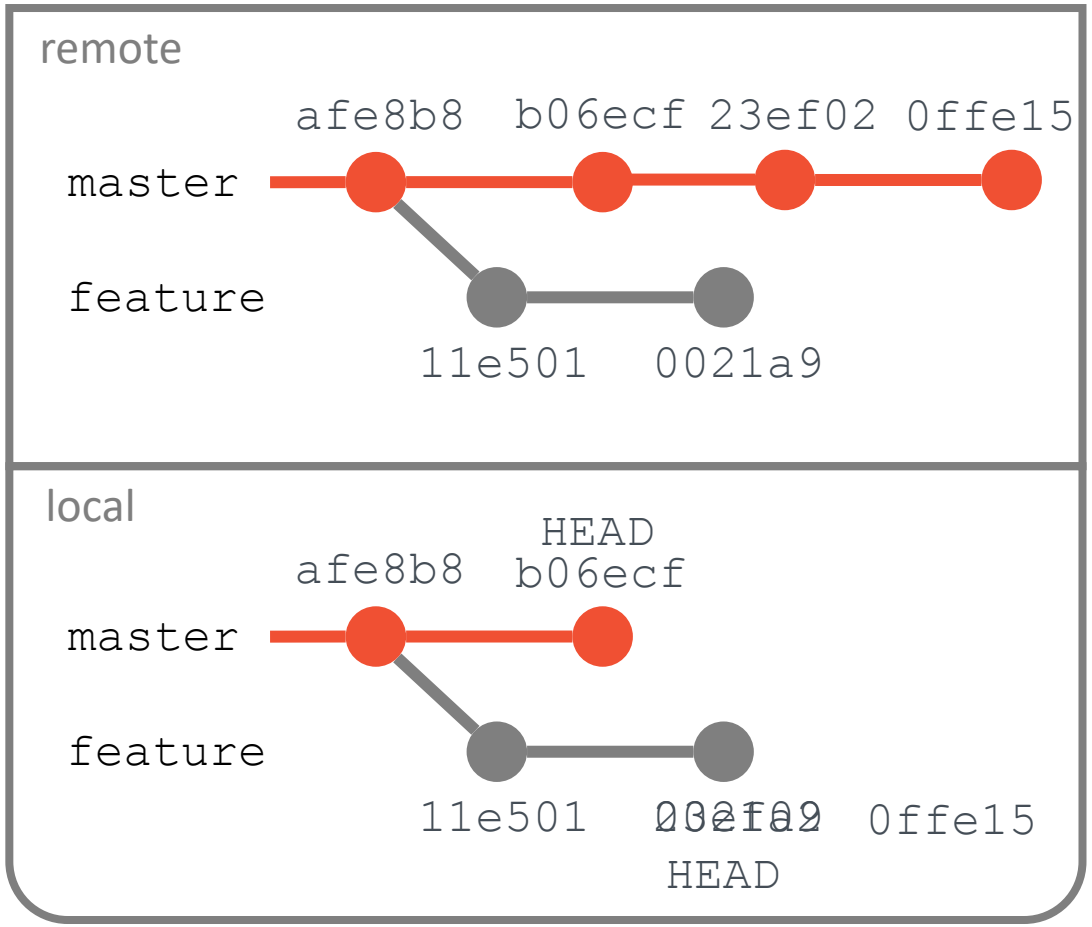
Changes base of a branch from one commit to another. *Moved* branch is composed of entirely new commits.

Rebase branch

```
git checkout [branch-name]
git rebase master
```

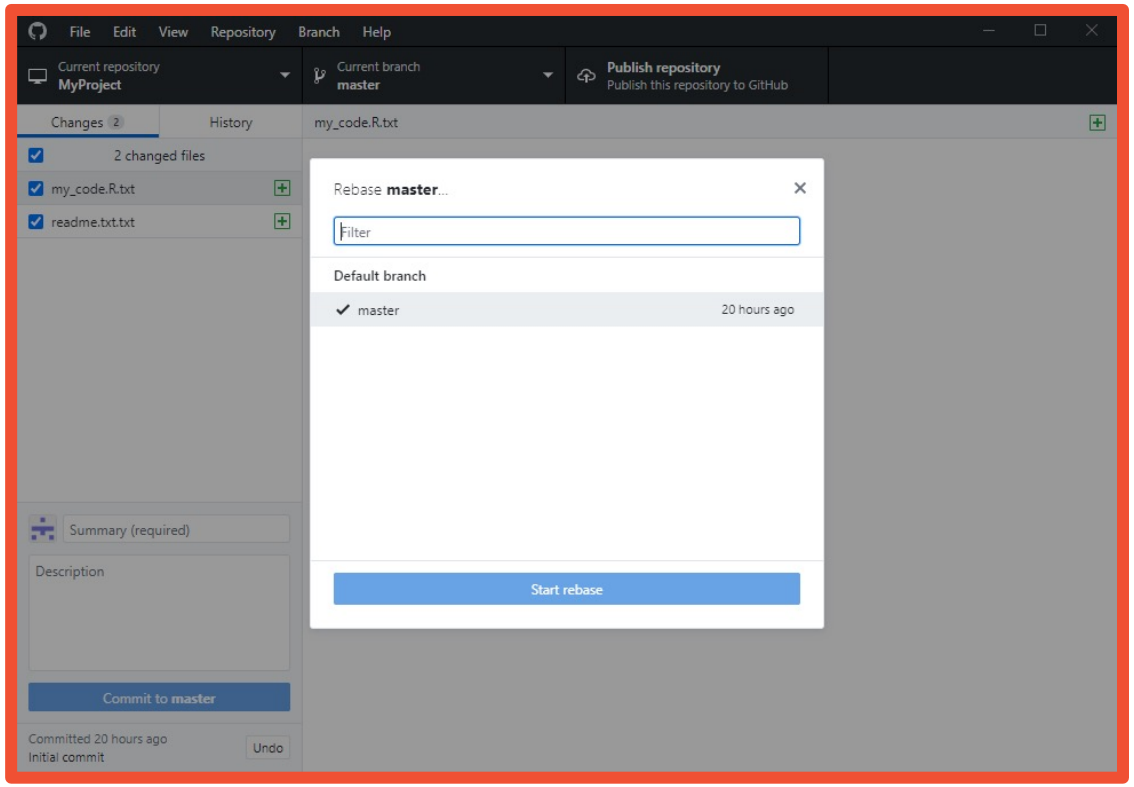
Fast-Forward merge

```
git checkout master
git merge [branch-name]
```



Rebase – GUI

Branch → Rebase current branch

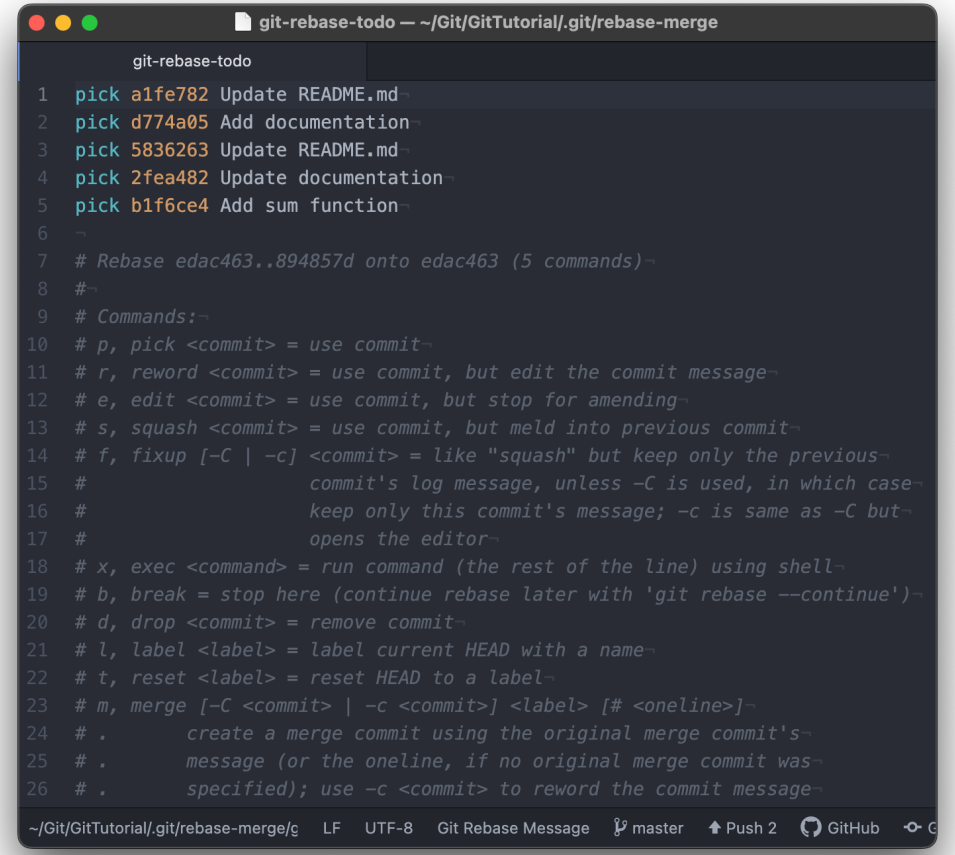


Interactive Rebase – CLI

Rewrite history using the powerful interactive rebase functionality. Combine multiple existing commits into one.

Rewrite last three commits

```
git rebase -i HEAD~3
```



Useful

Stage and commit all files

```
git commit -am "[message]"
```

Tag a commit

```
git tag
```

Stash changes in branch and apply them elsewhere

```
git stash
git stash apply
```

.gitignore

Health Data Science Workflow