

# Intro to git

Bertelsmeier Lab Meeting  
2021 Spring



# Version Control

- History of your files & changes
- Why use version control?
- Version control you probably use:
  - Rename revisions (versions or dates)
  - Sync to dropbox or google drive
  - Periodically copy folder to external drive

## A STORY TOLD IN FILE NAMES:

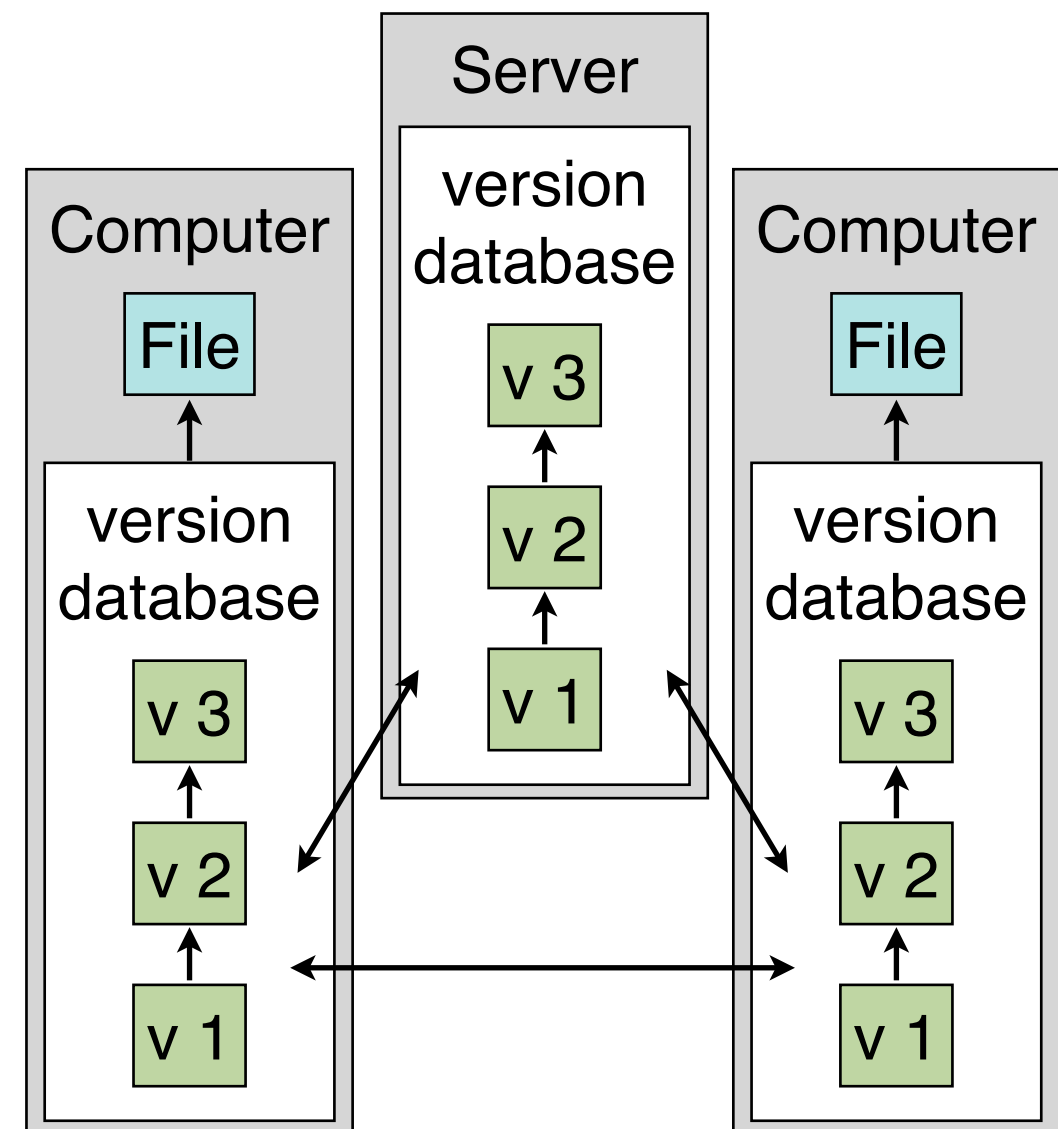
Location: C:\user\research\data

Filename	Date Modified	Size	Type
data_2010.05.28_test.dat	3:37 PM 5/28/2010	420 KB	DAT file
data_2010.05.28_re-test.dat	4:29 PM 5/28/2010	421 KB	DAT file
data_2010.05.28_re-re-test.dat	5:43 PM 5/28/2010	420 KB	DAT file
data_2010.05.28_calibrate.dat	7:17 PM 5/28/2010	1,256 KB	DAT file
data_2010.05.28_huh??.dat	7:20 PM 5/28/2010	30 KB	DAT file
data_2010.05.28_WTF.dat	9:58 PM 5/28/2010	30 KB	DAT file
data_2010.05.29_aaarrgh.dat	12:37 AM 5/29/2010	30 KB	DAT file
data_2010.05.29_#\$_*!&.dat	2:40 AM 5/29/2010	0 KB	DAT file
data_2010.05.29_crap.dat	3:22 AM 5/29/2010	437 KB	DAT file
data_2010.05.29_notbad.dat	4:16 AM 5/29/2010	670 KB	DAT file
data_2010.05.29_woohoo!.dat	4:47 AM 5/29/2010	1,349 KB	DAT file
data_2010.05.29_USETHISONE.dat	5:08 AM 5/29/2010	2,894 KB	DAT file
analysis_graphs.xls	7:13 AM 5/29/2010	455 KB	XLS file
ThesisOutline!.doc	7:26 AM 5/29/2010	38 KB	DOC file
Notes_Meeting_with_ProfSmith.txt	11:38 AM 5/29/2010	1,673 KB	TXT file
JUNK...	2:45 PM 5/29/2010		Folder
data_2010.05.30_startingover.dat	8:37 AM 5/30/2010	420 KB	DAT file

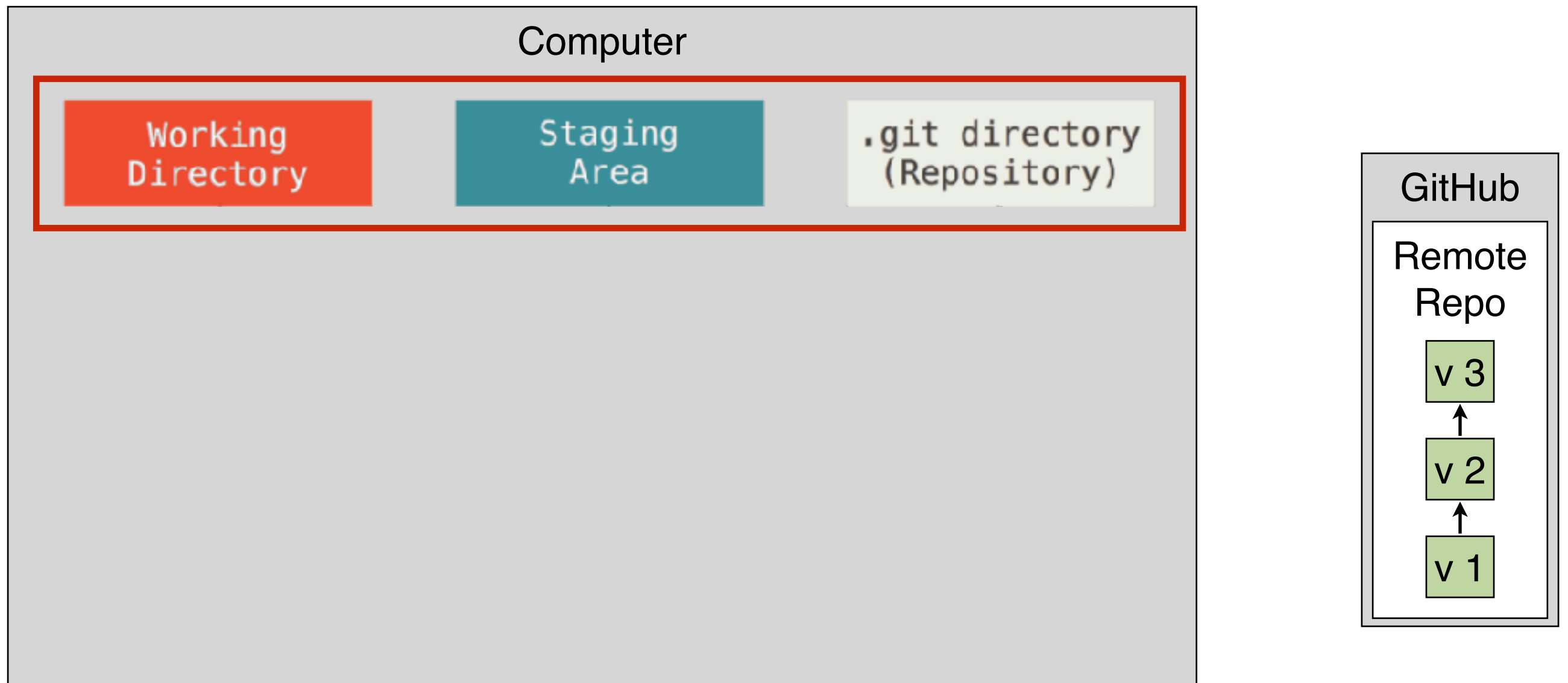
# Git

- Distributed version control system
- Full database on every computer
- Linus Torvalds (2005)
- Speed, data integrity, collaboration
- Stores snapshots of *all* tracked files in directory
- Store locally, push to remote server (e.g., GitHub)

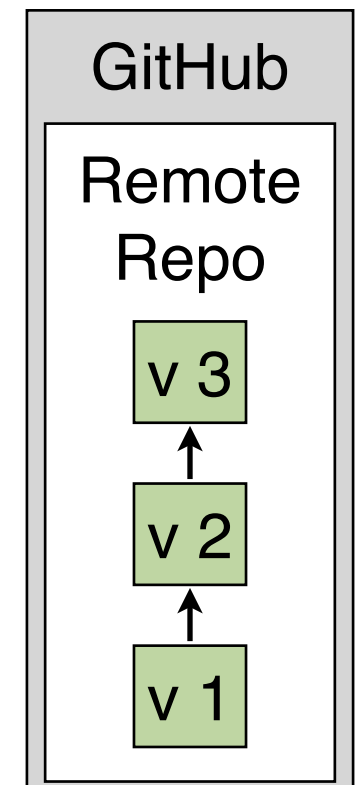
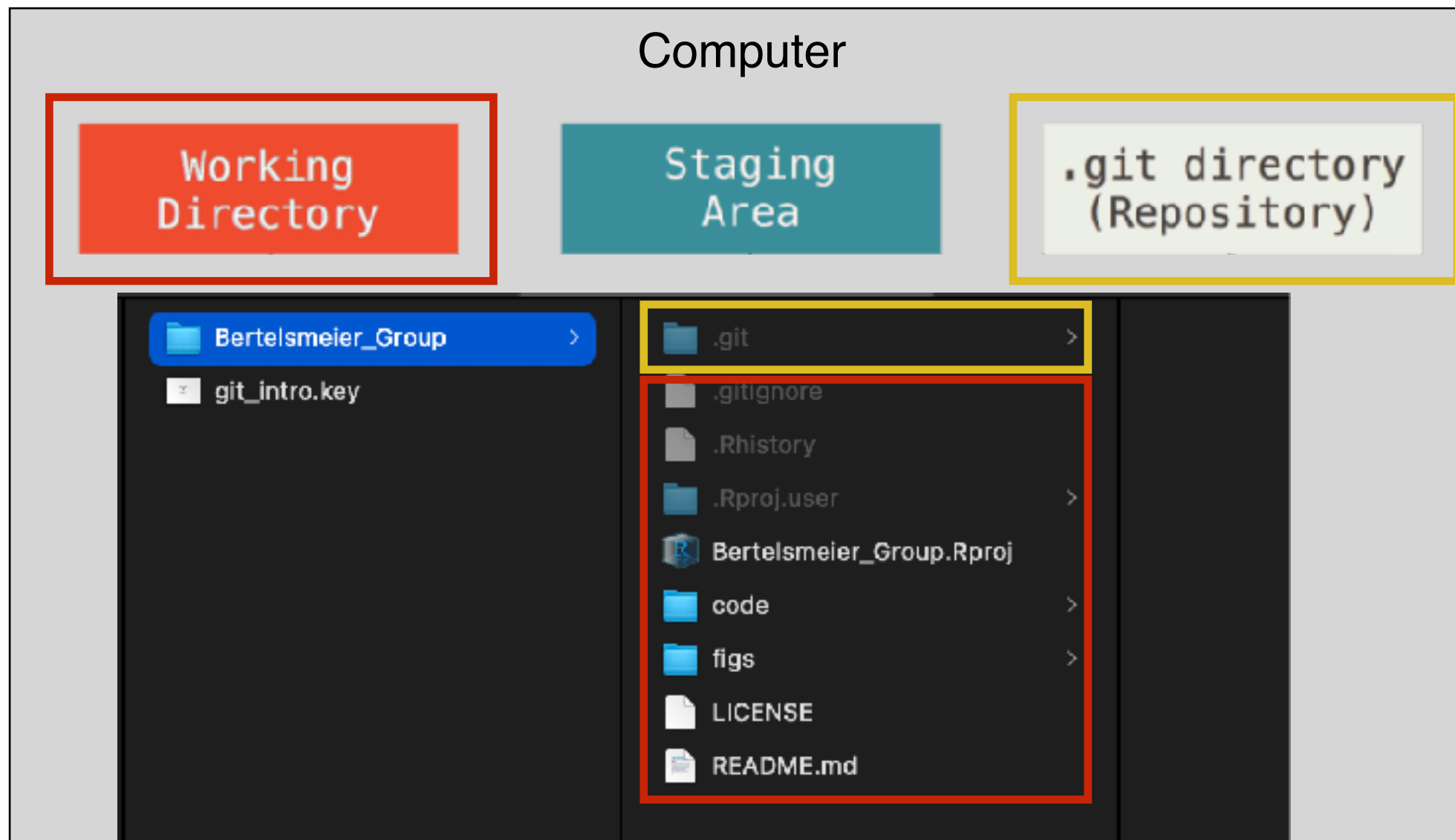
## Distributed VCS



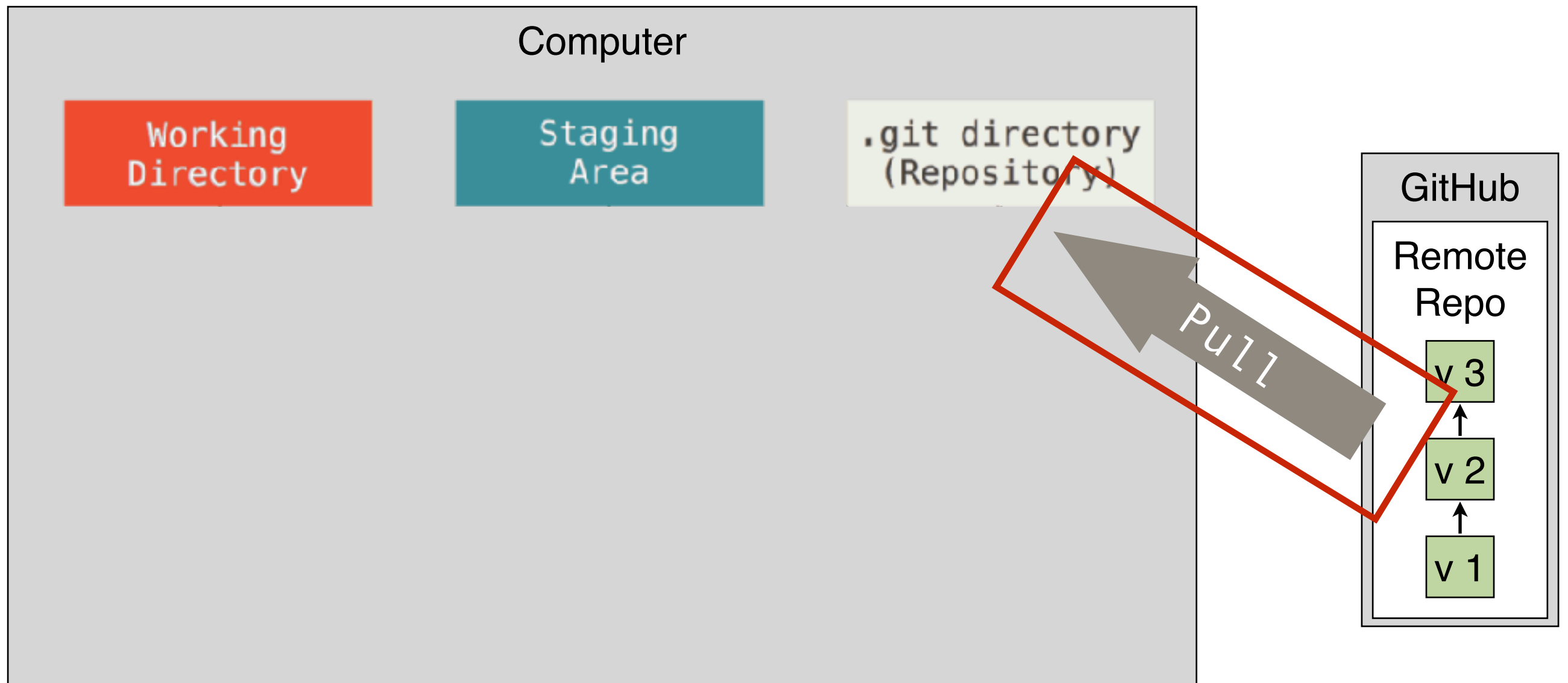
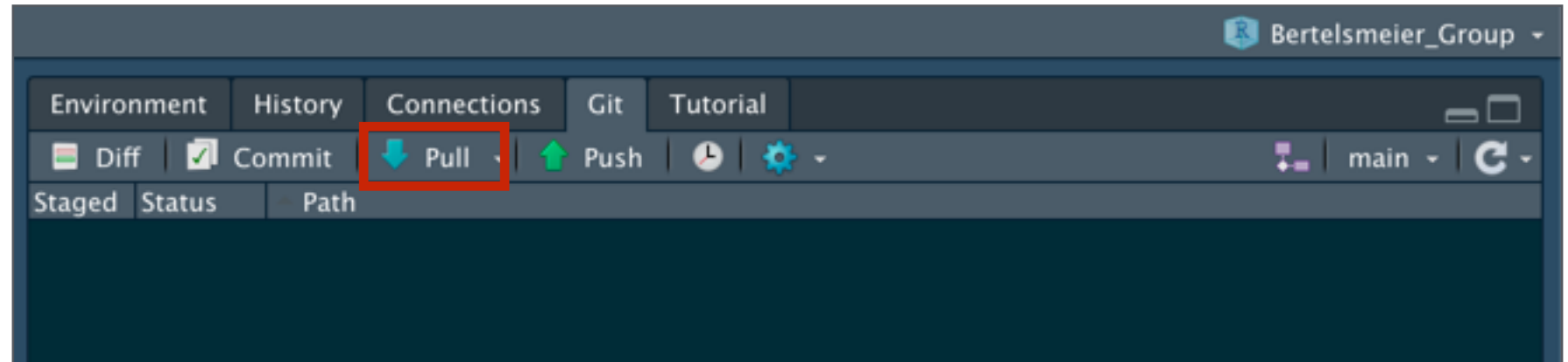
# Git workflow



# Git workflow

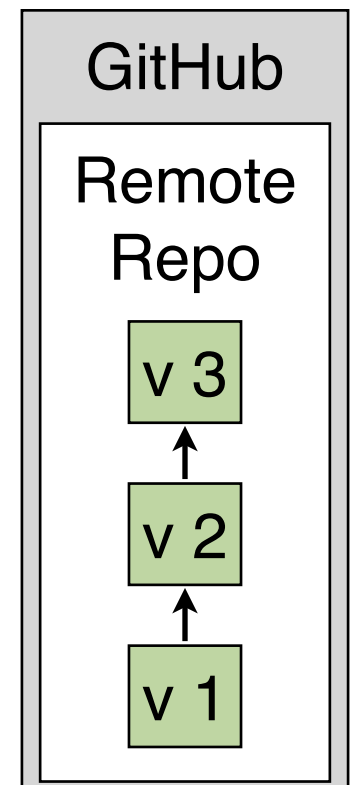
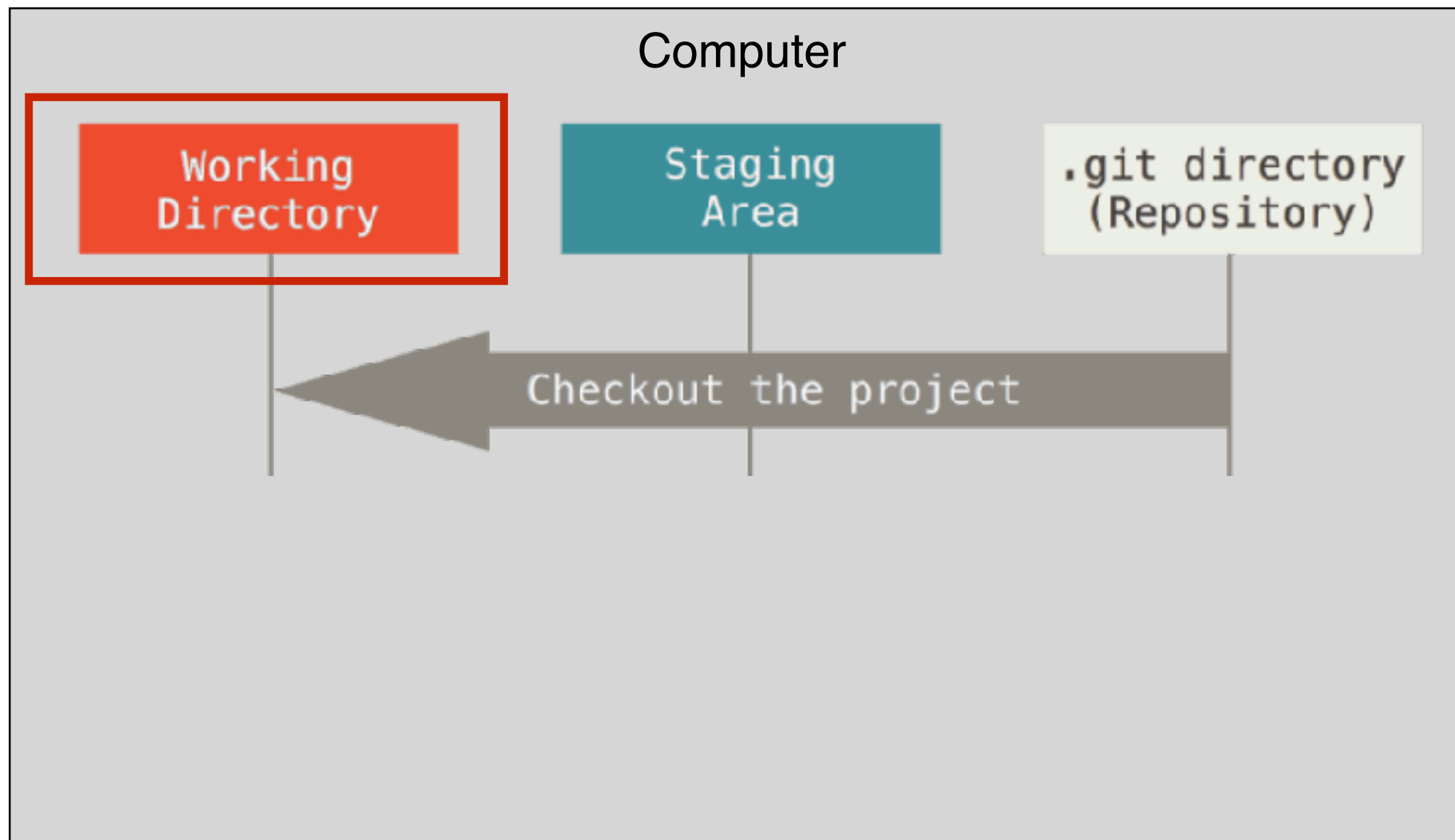
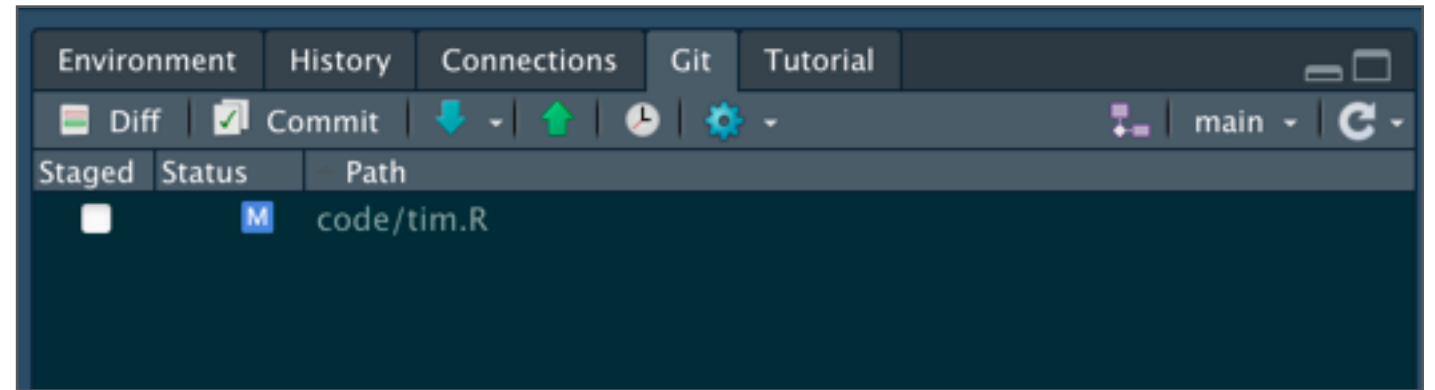


# Git workflow



# Git workflow

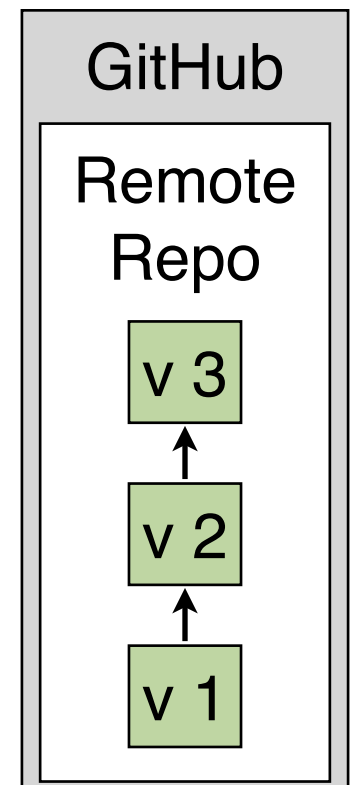
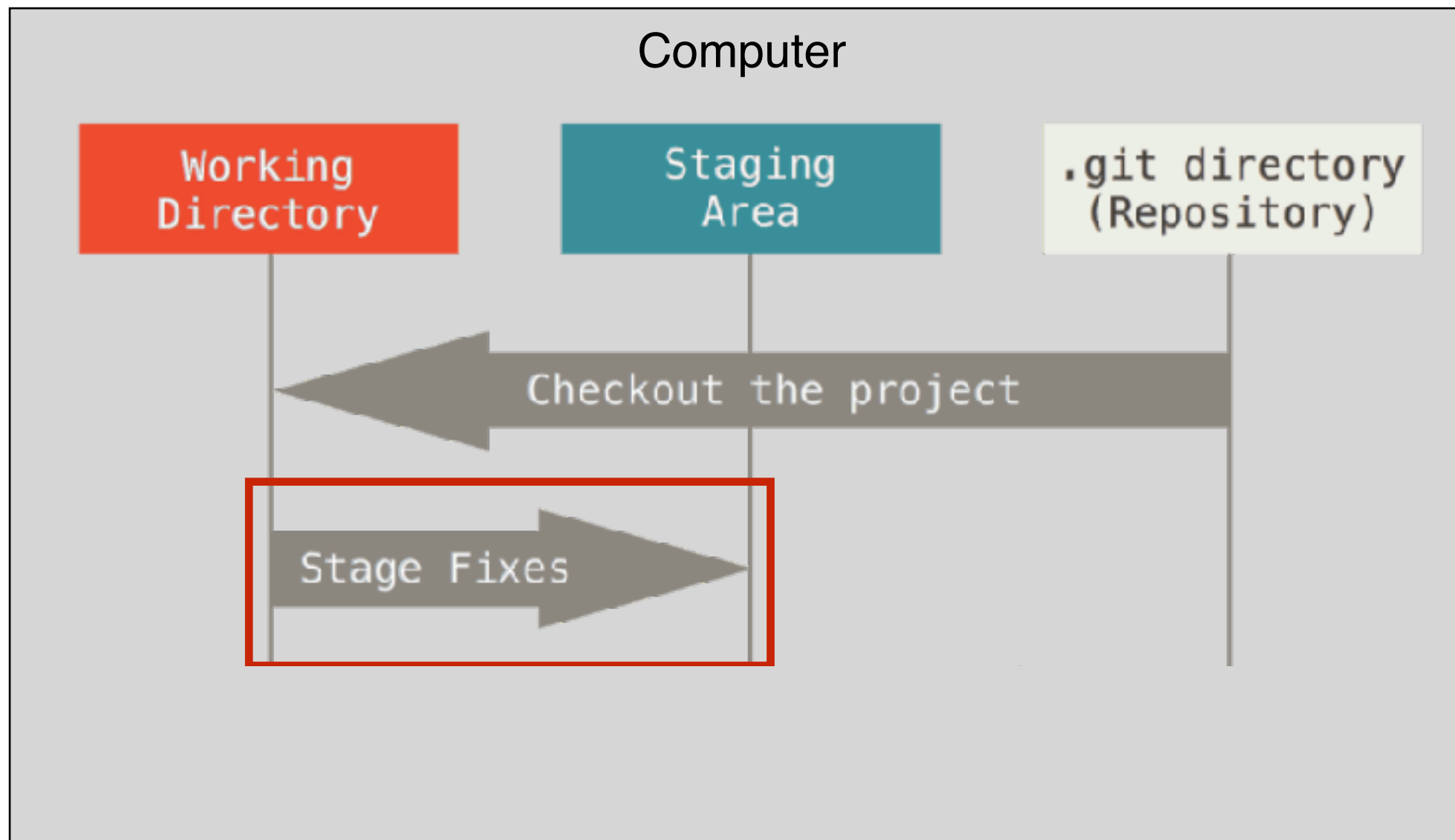
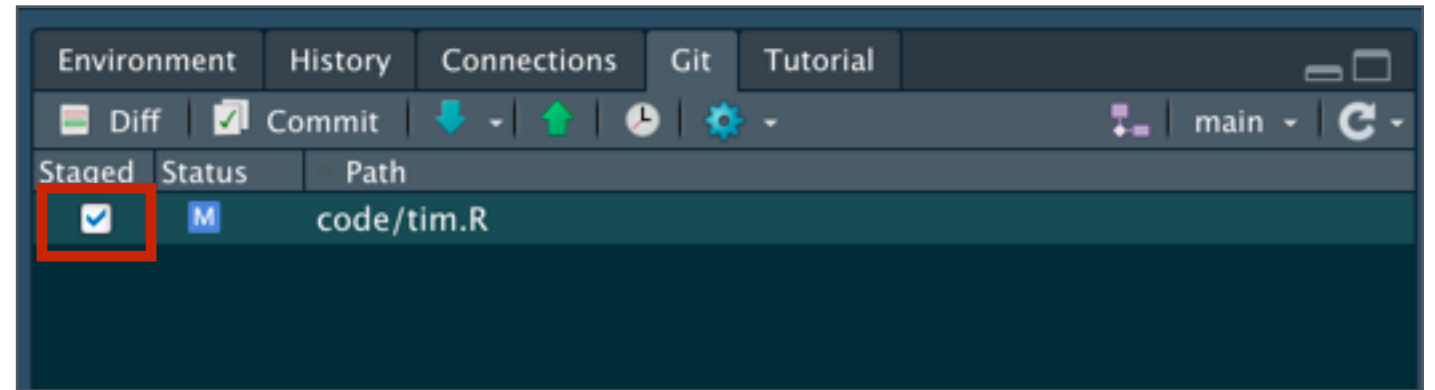
## 1. Work





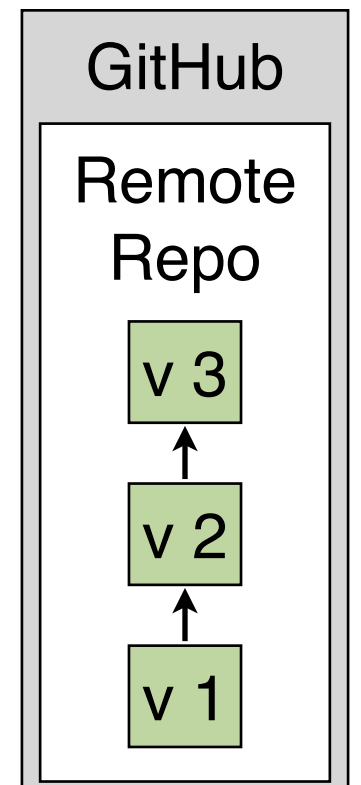
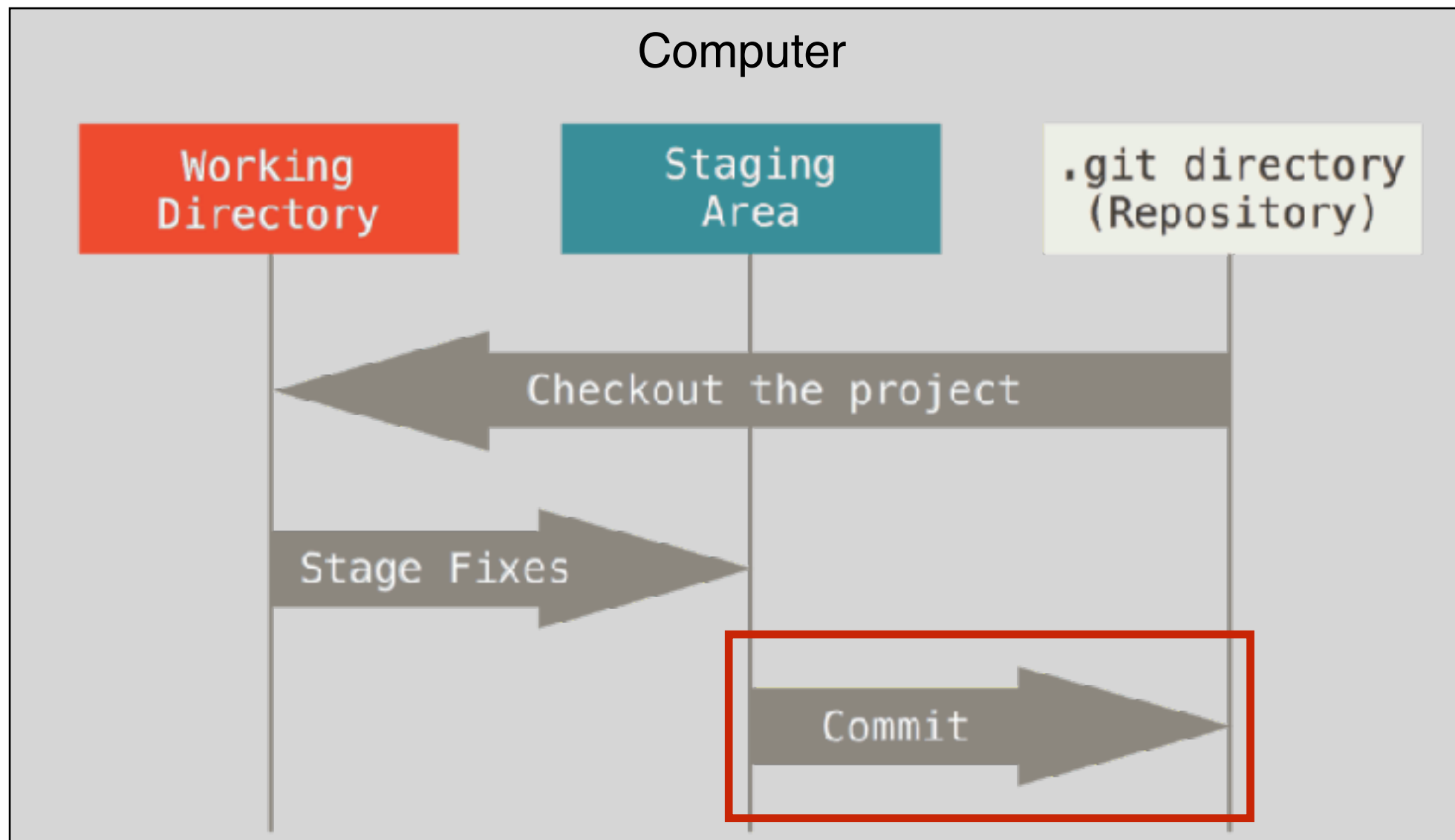
# Git workflow

1. Work
2. Stage changes



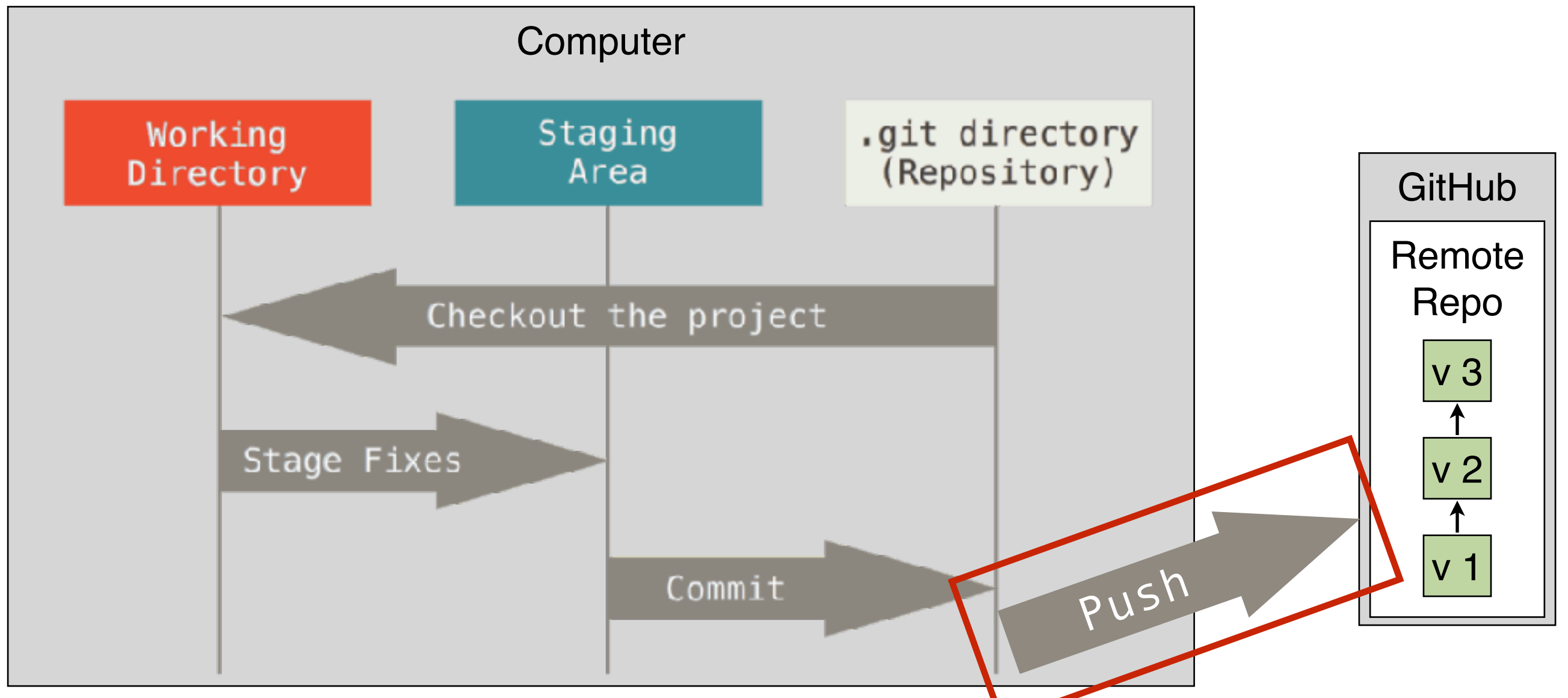
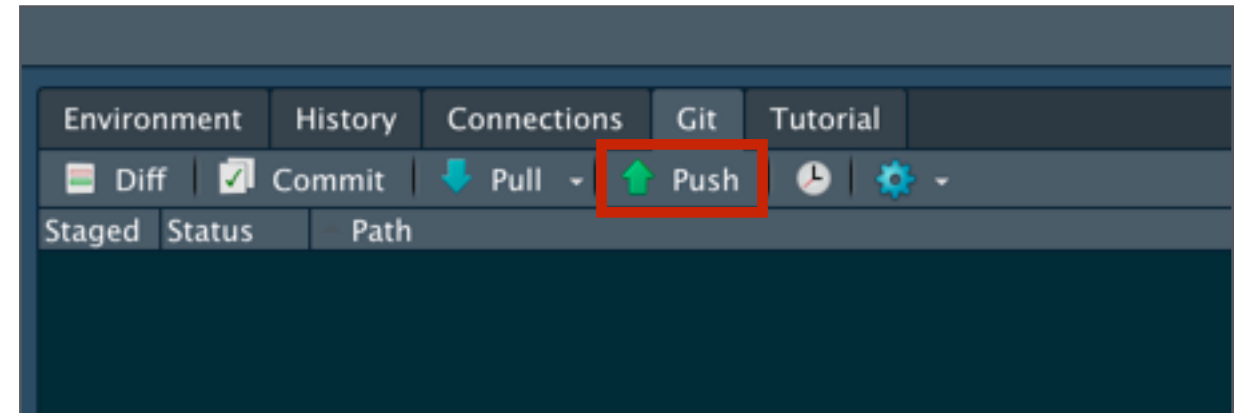
# Git workflow

1. Work
2. Stage changes
3. Commit staged changes



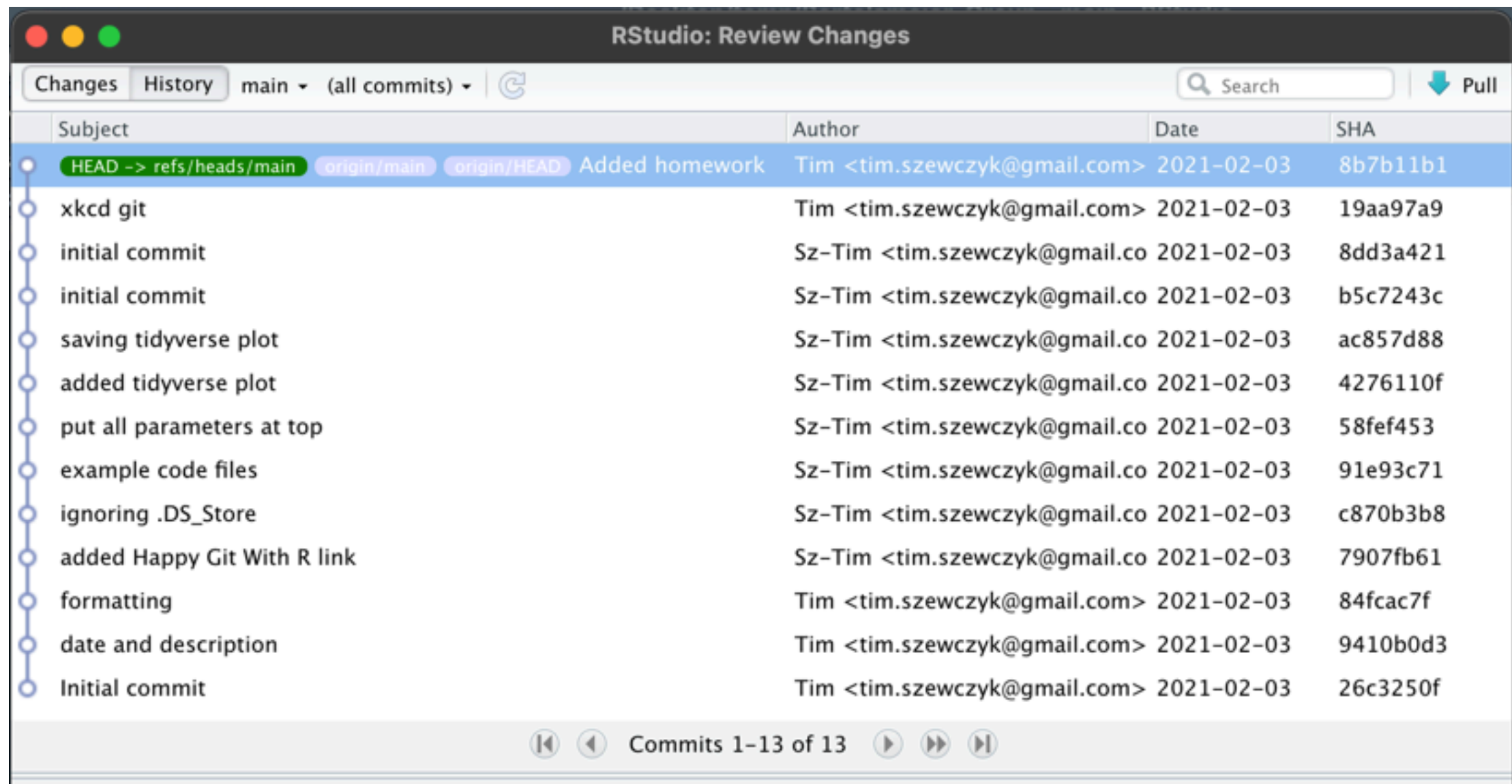
# Git workflow

1. Work
2. Stage changes
3. Commit staged changes
4. Push commits to GitHub



# Commit history

Each commit has a message, author, date/time, and SHA



The image shows the 'RStudio: Review Changes' window. At the top, there are tabs for 'Changes' and 'History', with 'History' selected. Below the tabs, there are dropdown menus for 'main' and '(all commits)', a search bar, and a 'Pull' button. The main area displays a list of commits with columns for Subject, Author, Date, and SHA. The commits are listed in reverse chronological order, with the most recent at the top. The first commit is 'Added homework' by Tim <tim.szewczyk@gmail.com> on 2021-02-03 with SHA 8b7b11b1. The last commit is 'Initial commit' by Tim <tim.szewczyk@gmail.com> on 2021-02-03 with SHA 26c3250f. At the bottom, there are navigation icons and a status bar showing 'Commits 1-13 of 13'.

Subject	Author	Date	SHA
HEAD -> refs/heads/main origin/main origin/HEAD Added homework	Tim <tim.szewczyk@gmail.com>	2021-02-03	8b7b11b1
xkcd git	Tim <tim.szewczyk@gmail.com>	2021-02-03	19aa97a9
initial commit	Sz-Tim <tim.szewczyk@gmail.co	2021-02-03	8dd3a421
initial commit	Sz-Tim <tim.szewczyk@gmail.co	2021-02-03	b5c7243c
saving tidyverse plot	Sz-Tim <tim.szewczyk@gmail.co	2021-02-03	ac857d88
added tidyverse plot	Sz-Tim <tim.szewczyk@gmail.co	2021-02-03	4276110f
put all parameters at top	Sz-Tim <tim.szewczyk@gmail.co	2021-02-03	58fef453
example code files	Sz-Tim <tim.szewczyk@gmail.co	2021-02-03	91e93c71
ignoring .DS_Store	Sz-Tim <tim.szewczyk@gmail.co	2021-02-03	c870b3b8
added Happy Git With R link	Sz-Tim <tim.szewczyk@gmail.co	2021-02-03	7907fb61
formatting	Tim <tim.szewczyk@gmail.com>	2021-02-03	84fcac7f
date and description	Tim <tim.szewczyk@gmail.com>	2021-02-03	9410b0d3
Initial commit	Tim <tim.szewczyk@gmail.com>	2021-02-03	26c3250f

# Commit history

Each commit has a message, author, date/time, and SHA

SHA 58fef453  
Author Sz-Tim <tim.szewczyk@gmail.com>  
Date 2021-02-03 09:32  
Subject put all parameters at top  
Parent 91e93c71

 [code/tim.R](#)

 **code/tim.R**

[View file @ 58fef453](#)

```
@@ -4,9 +4,9 @@
4 4
5 5 # Let's make up some data
6 6 n <- 100
7 x <- seq(-5, 5, length.out=n)
8 7 b <- c(-2, 1.5)
9 8 sigma <- 1.5
9 x <- seq(-5, 5, length.out=n)
10 10 y <- rnorm(n, b[1] + b[2]*x, sigma)
11 11
12 12 plot(x, y)
```

# Benefits of Git

- Screwing up is ok
- Remote, up-to-date backup of entire project
- Collaboration
  - Merge work, branches, user history
- Integrated with RStudio
- Organization, reproducibility, transparency
- All filetypes

# git is as good as you are

	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
○	AAAAAAAAA	3 HOURS AGO
○	ADKFJSLKDFJSDKLFJ	3 HOURS AGO
○	MY HANDS ARE TYPING WORDS	2 HOURS AGO
○	HAAAAAAAAAANDS	2 HOURS AGO

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

# Homework

- [https://github.com/Sz-Tim/Bertelsmeier\\_Group](https://github.com/Sz-Tim/Bertelsmeier_Group)
- Create GitHub account, send to Tim
- Install RStudio, git (if needed)
- Enable git within RStudio
- Tell git your user info