Version control with git and GitHub for scientists

Max Joseph

July 27, 2016

Discuss

What is your current version control system?

What is git

Version control system

- manage different versions of files
- collaborate with yourself
- collaborate with other people

Why use git

“Always remember your first collaborator is your future self, and your past self doesn’t answer emails”

- Christie Bahlai

What is git good for?

- backup
- reproducibility
- collaboration

What you get

Tour of a git repository

Overview

1. Git on the command line
2. Git in RStudio
3. Github vs. GitLab vs. Bitbucket for remote mirroring

Command line git

Make a directory with a file

```
mkdir test
cd test
nano sim.R
```

Then write a short simulation, e.g.

Initializing a repository

Prerequisites:

- git installed
- git configured

```
git config --global user.name "Vlad Dracula"
git config --global user.email "vlad@tran.sylvan.ia"
git config --global color.ui "auto"
git config --global core.editor "nano"
```

Initializing a repository

```
git init
```

Notice the `.git/` directory

Checking repository status

```
git status
```

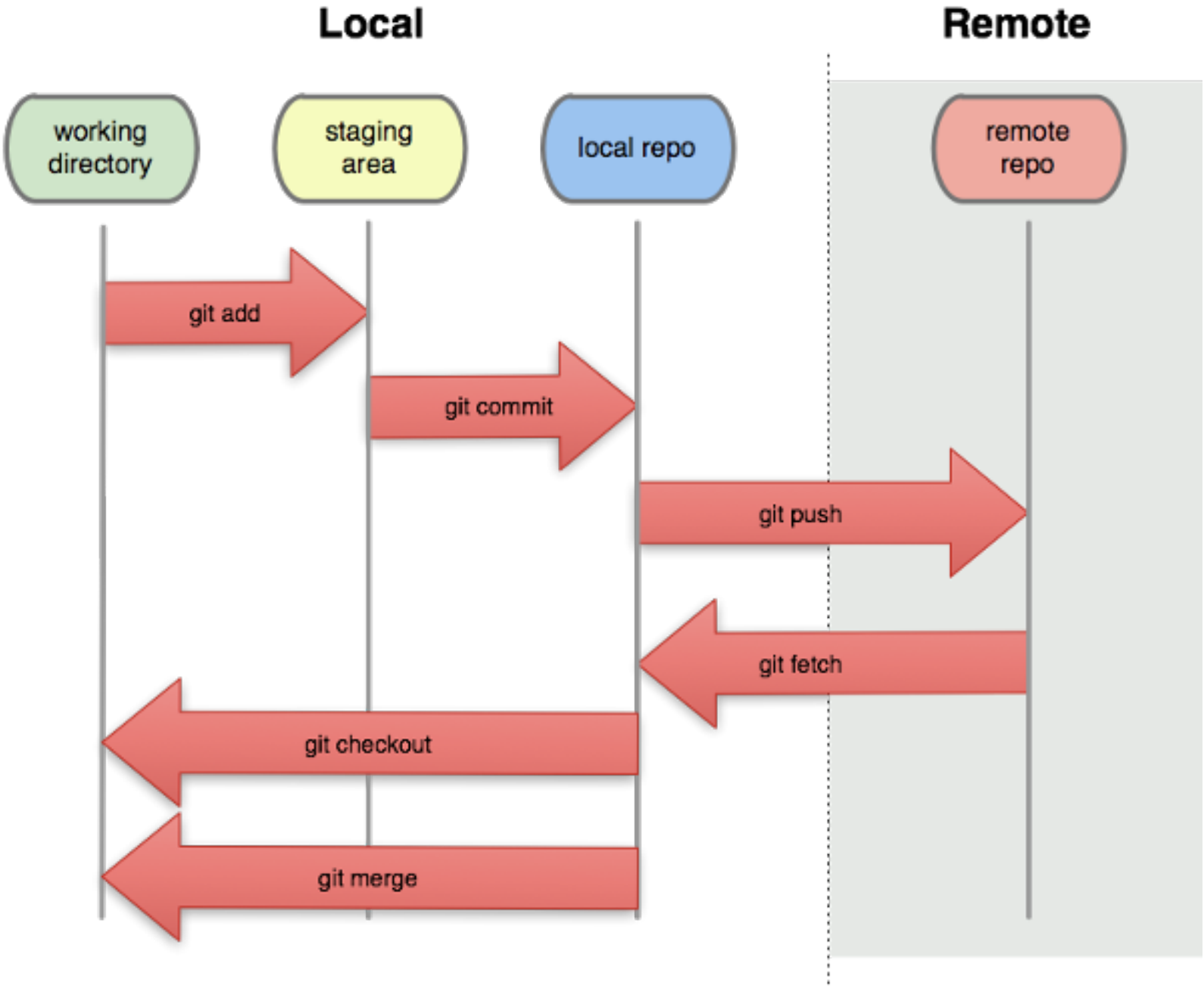
Adding your file

```
git add your_filename.R
```

or, to add everything

```
git add --all
```

Your changes are now staged



Committing

Changes aren’t final until they’re committed

```
git status
```

Committing

Once you're sure that you're changes are worth saving
(THIS WILL GO ON YOUR PERMANENT RECORD)

```
git commit -m 'changed x, y, and z'
```

Commit messages

Describe why

What did we do?

```
git status
git log
```

Make another change

1. Change file
2. Add changes
3. Commit changes
4. View updated log

Now, do something really stupid

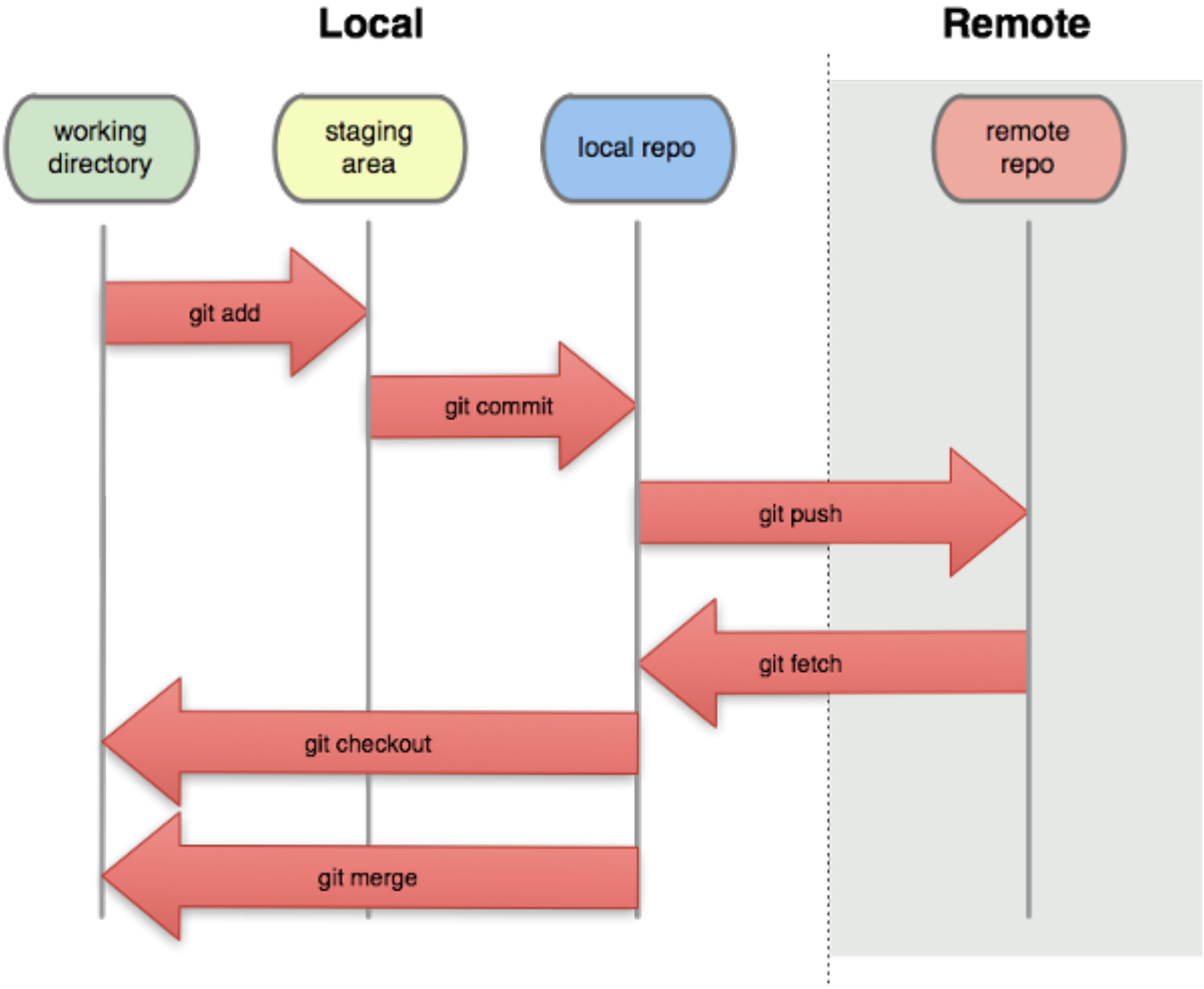
“Accidentally” introduce some errors to your file

Woops

Not that this ever happens…

```
git diff
git checkout HEAD your_file.R
```

What happened?



Wait, what does HEAD refer to?

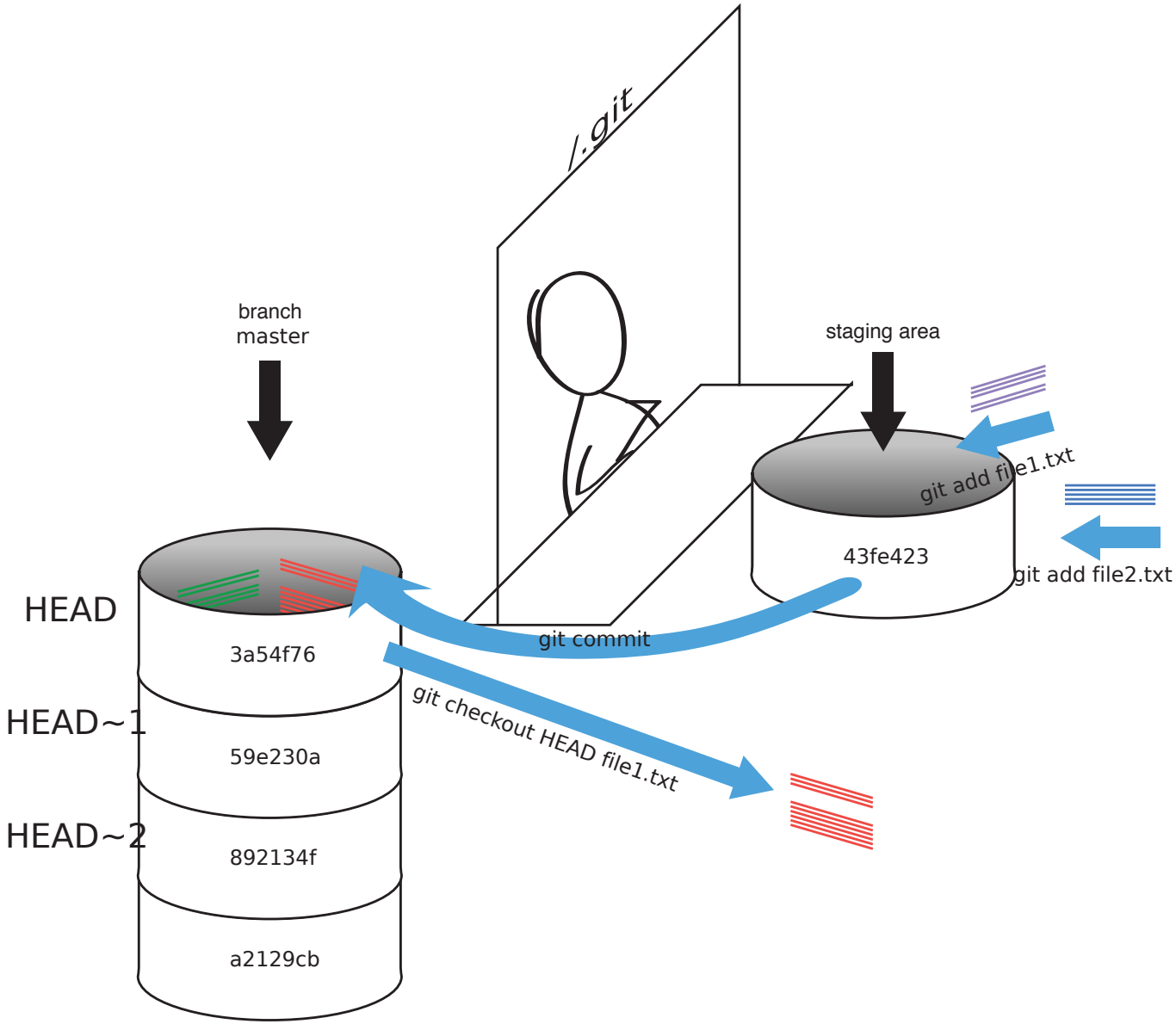


Figure 3: Commits ≈ a stack of heads

Remote repositories

1. Create repository on Github, GitLab, or Bitbucket
2. Add that as a remote

```
git remote add origin https://www.github.com/user/test.git
```

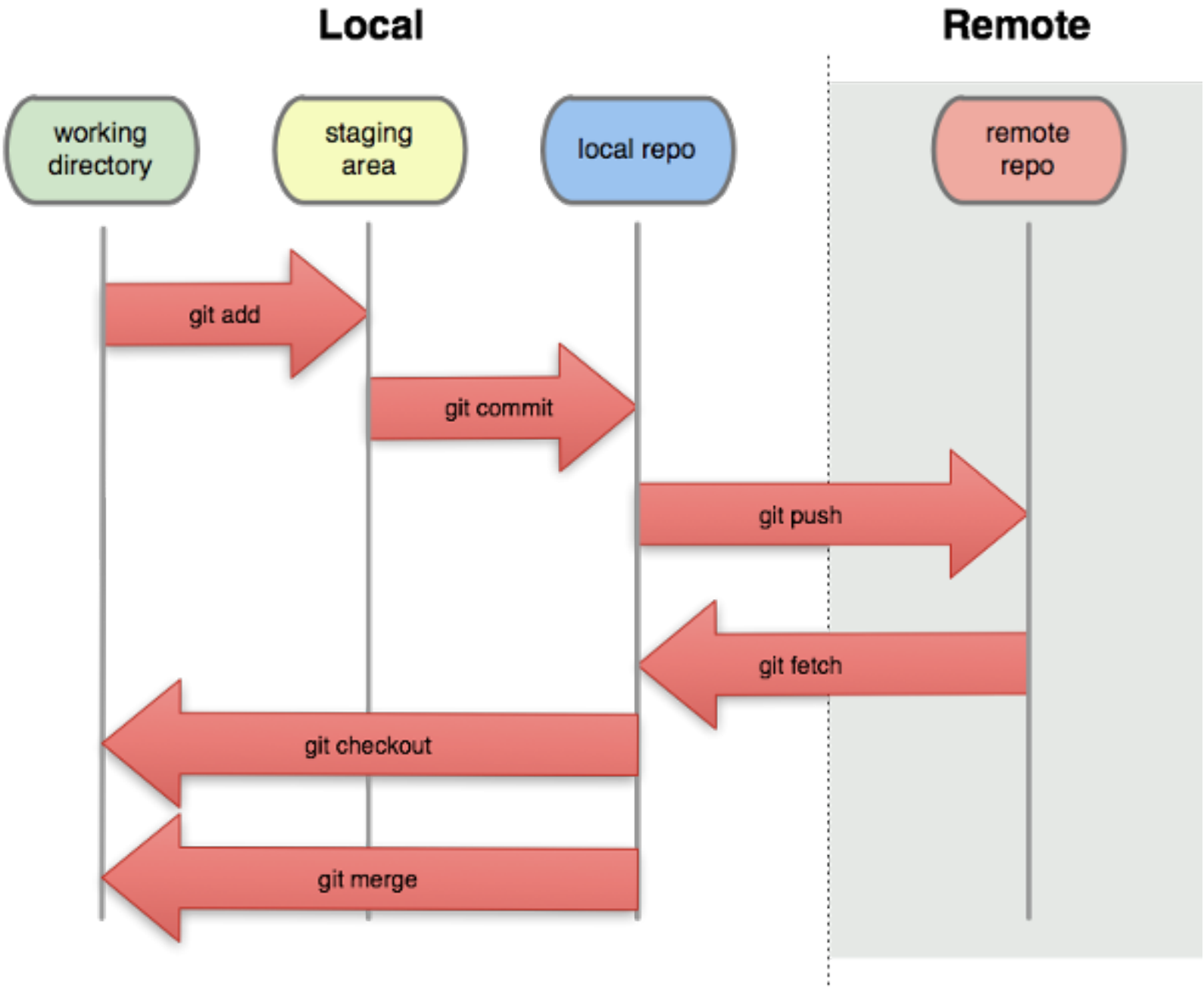
Once your repository has been linked to remote

Push your changes

```
git push -u origin master
```

Check the remote (Github or Bitbucket) to see new changes

Overview


Local
Remote
working directory
staging area
local repo
remote repo
git add
git commit
git push
git fetch
git checkout
git merge

RStudio's interface

Start a new project

Check it out from your remote git repository using ssh or html

(ssh is better, html may require additional config w/ RStudio)

Demo of reverting, adding, committing, pushing

Github vs. GitLab vs. Bitbucket

Private repos:

- free on Bitbucket (w/ < 6 collaborators)
- not free on Github

Github vs. GitLab vs. Bitbucket

Popularity & user base (4 vs. 1 million)

Support

Bottom line

You can use both if you want

Bitbucket wins w/ private repos

Github wins for open projects & big collaborations

Continuing education & advanced stuff

Pro Git

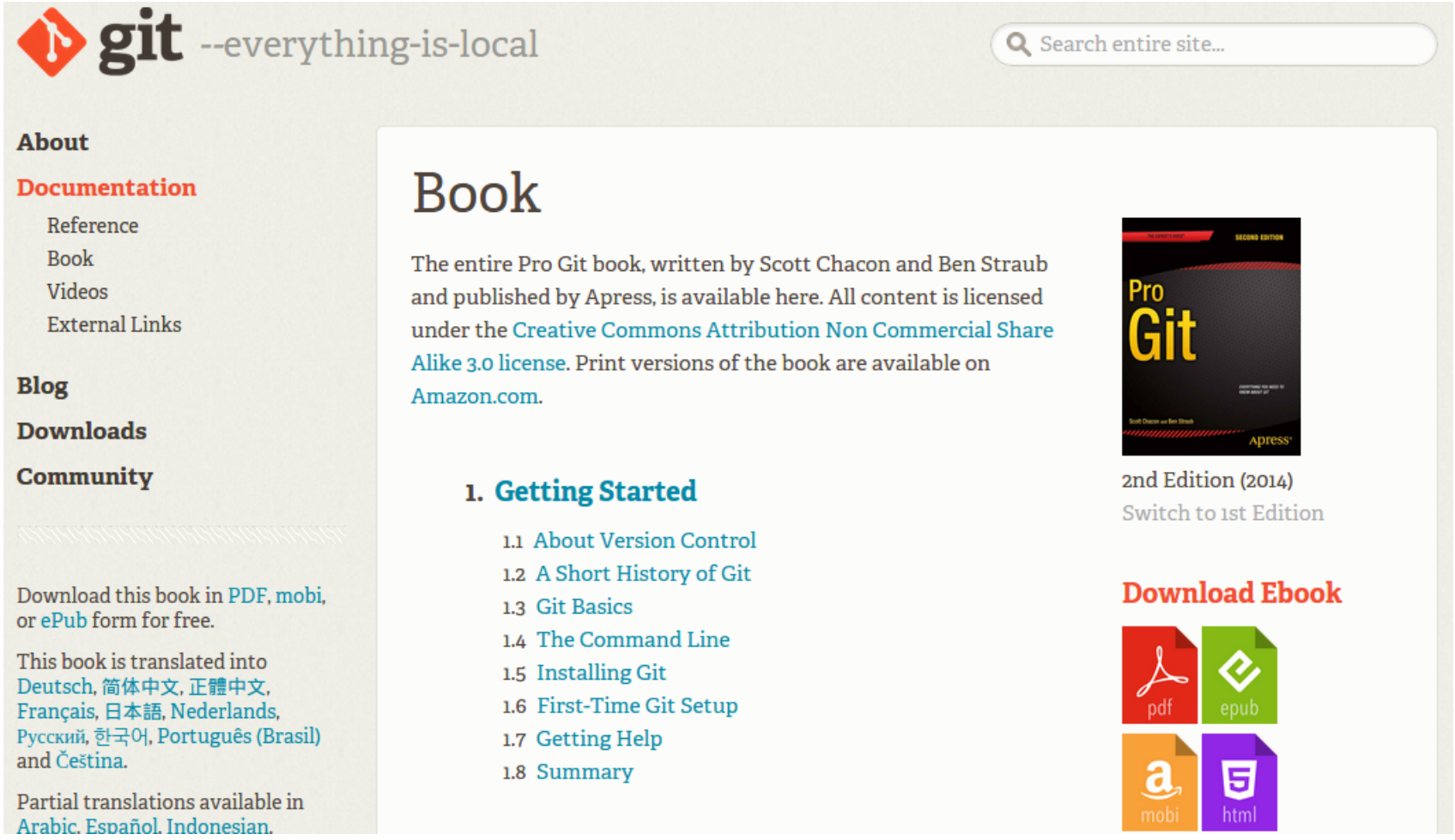


Figure 5: