



Workflows & Git/GitHub Basics

Justin Post

What do we want to be able to do?

Data Science!

- Read in raw data and manipulate it
- Combine data sources
- Summarize data to glean insights
- Apply common analysis methods
- Communicate Effectively

Workflow

Important considerations for data analysis workflow:

- **Reproducibility**
- **Version control**
- **Collaboration**

What are git and github?

- **Git** is a version control software
- **Github** is an online hosting service for Git-based projects

What are git and github?

- **Git** is a version control software
- **Github** is an online hosting service for Git-based projects

Git Tracking

- You associate git with a folder (repo)
- Git keeps track of all files in the folder (repo)
- If you want to keep changes you've made, you **commit** and **push** the changes to the folder (repo)

Github

- Github allows you to have a remote file repository (folder) tracked by git
 - Let's create a repository on github.com
 - **Add** some files and **commit** to the changes
 - Modify some files on github
 - Investigate the version control!

Local vs Remote Work

Mostly you'll want to work on your local computer. Install `git` on your computer!

Local vs Remote Work

Mostly you'll want to work on your local computer. Install `git` on your computer!

Workflow

1. (Initially) **clone** the repo locally. (Later) **pull** to get most recent versions of files
2. Work and make changes
3. **add** and **commit** to changes you like
4. **push** changes to remote repo (on github)

Let's clone our repo and work on it locally!

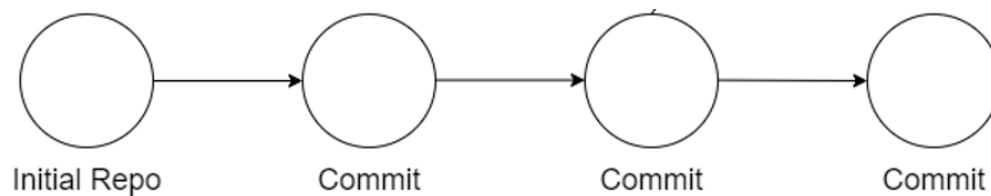
Git & RStudio

Git and RStudio work great together!

- Works through **R Projects**
- Start a new project from git repo
 - Update with command line or git menu!

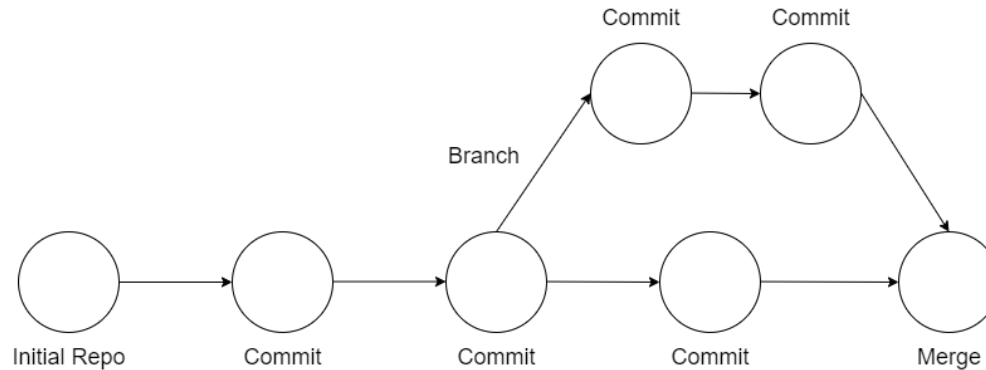
Collaboration Idea

- Everyone can work on the same **branch** and just take turns working



Working on Branches

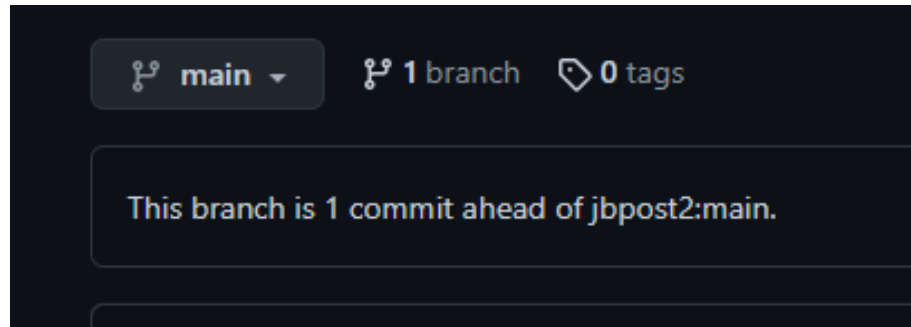
- Alternatively, you can have separate **branches** of the repo



- Work on a branch is similar to working on the main branch
- Can merge when happy!

Forking

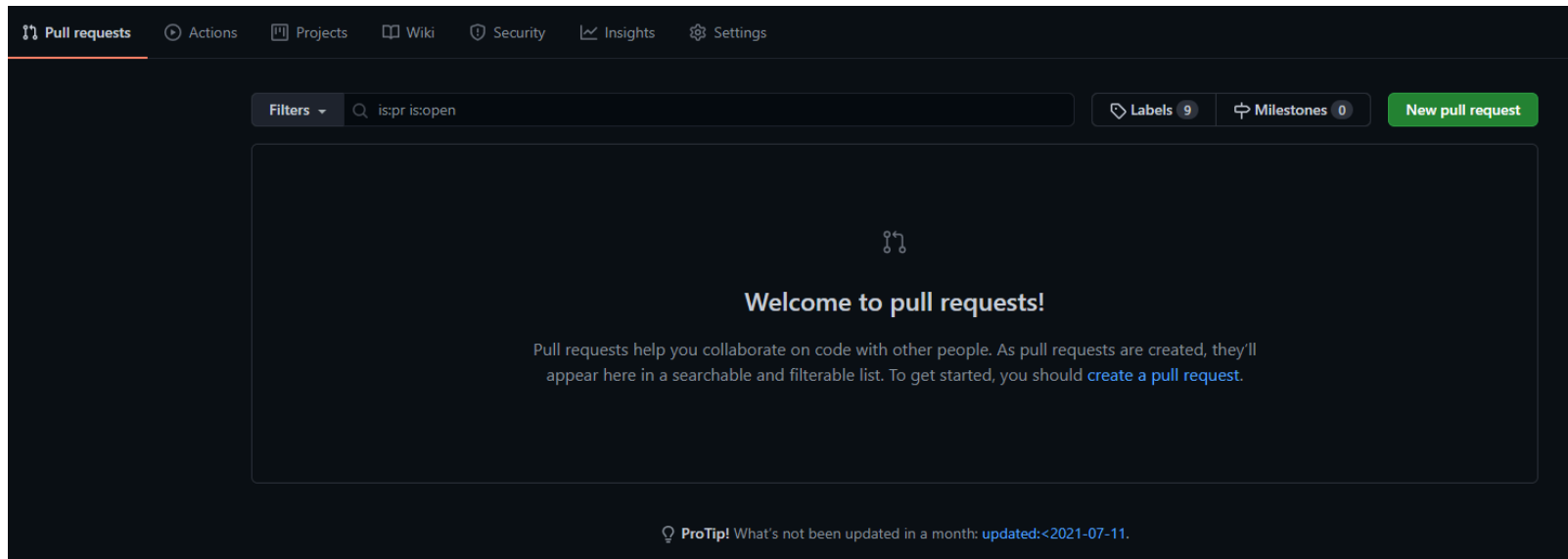
- People often `fork` the repo
- This creates a copy of the repo on your account
- You can then work as normal
- If you do a commit on your branch, you may notice something like this



Merging Branches

Suppose you like your commit and you think the original owner will too!

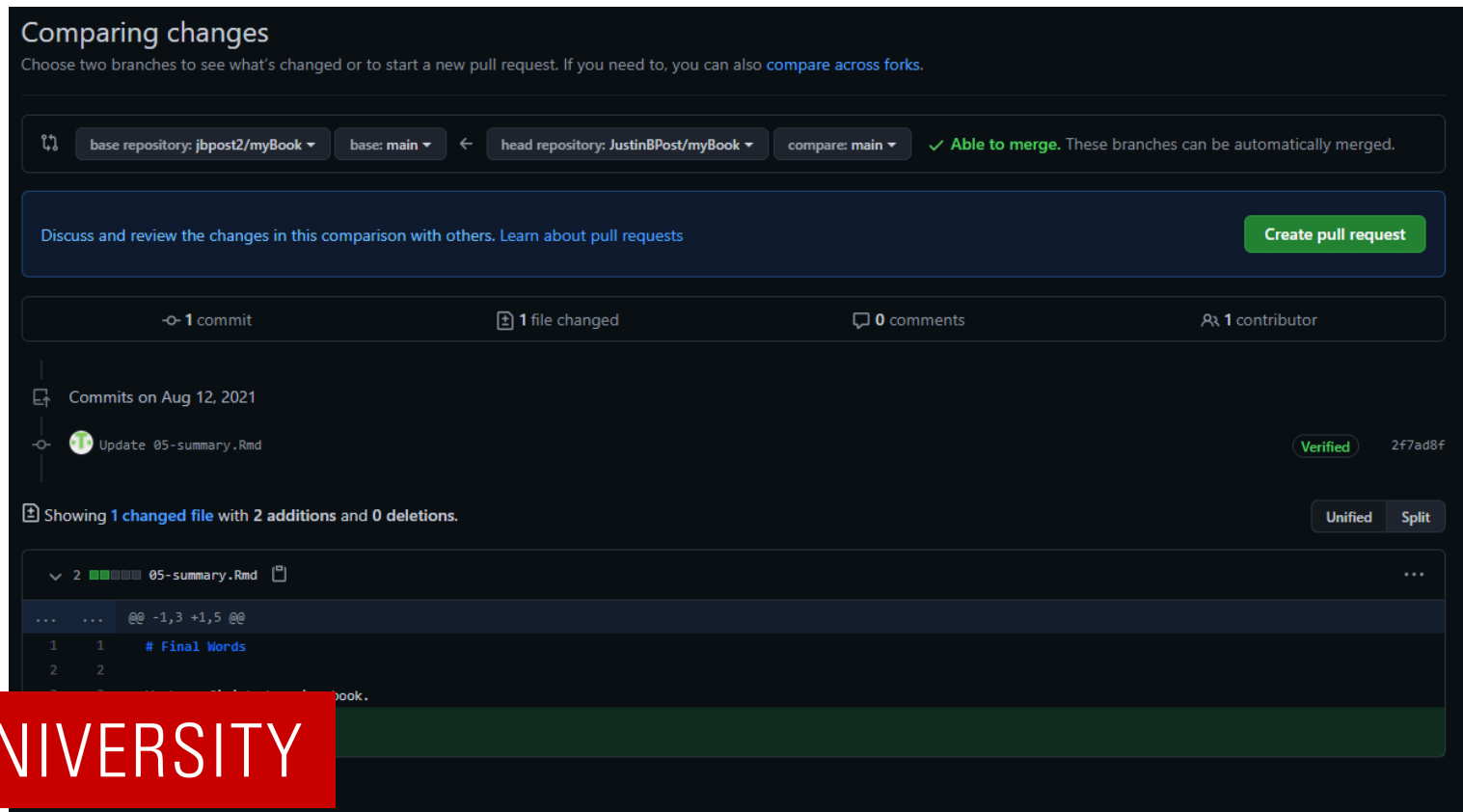
- You can do a **pull** request



Merging Branches

Suppose you like your commit and you think the original owner will too!

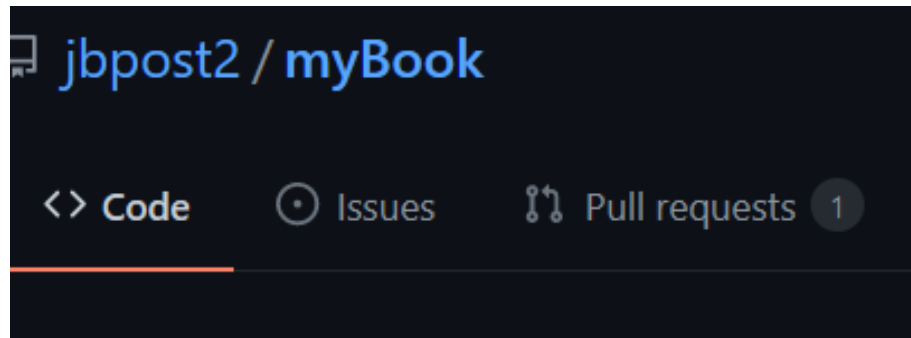
- You can do a **pull** request



Merging Branches

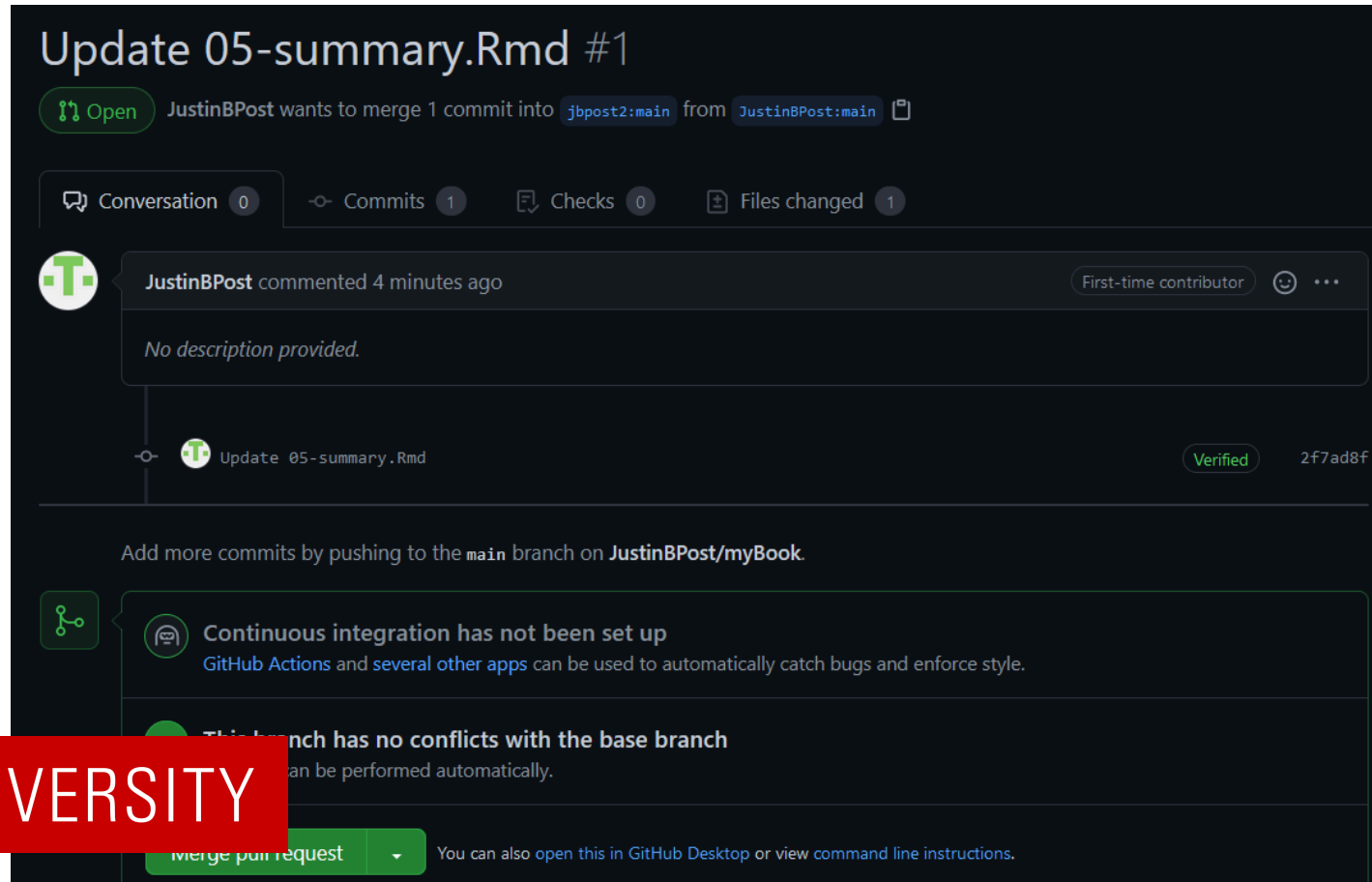
If you are lucky, there won't be any merge conflicts.

- Allows the owner of the original repo to accept the pull request without needing to modify things
- The owner will get a notification that a pull request has been made



Merging Branches

Owner can then investigate the request and choose whether or not to accept it or they can ask for more details



Dealing with conflicts

- Sometimes changes requested conflict with changes already made

The screenshot shows the GitHub 'Comparing changes' interface. At the top, it says 'Comparing changes' and 'Choose two branches to see what's changed or to start a new pull request. If you need to, you can also [compare across forks](#).' Below this, there are dropdowns for 'base repository: jbpost2/myBook', 'base: main', 'head repository: JustinBPost/myBook', and 'compare: main'. A red error message states: 'X Can't automatically merge. Don't worry, you can still create the pull request.' Below the error message is a blue bar with the text 'Discuss and review the changes in this comparison with others. [Learn about pull requests](#)' and a green 'Create pull request' button. Below this bar, there are statistics: '1 commit', '1 file changed', '0 comments', and '1 contributor'. Below the statistics, there is a section for 'Commits on Aug 12, 2021' showing a commit 'Update 02-literature.Rmd' by a user with a green 'Verified' badge. Below the commit, it says 'Showing 1 changed file with 1 addition and 0 deletions.' and has 'Unified' and 'Split' tabs. The file '02-literature.Rmd' is expanded, showing a diff. The diff shows a new line added at line 4: '+ This line is better than yours, approve this change.'

Comparing changes

Choose two branches to see what's changed or to start a new pull request. If you need to, you can also [compare across forks](#).

base repository: jbpost2/myBook base: main ← head repository: JustinBPost/myBook compare: main

✖ Can't automatically merge. Don't worry, you can still create the pull request.

Discuss and review the changes in this comparison with others. [Learn about pull requests](#) [Create pull request](#)

1 commit 1 file changed 0 comments 1 contributor

Commits on Aug 12, 2021

Update 02-literature.Rmd [Verified](#) 366353a

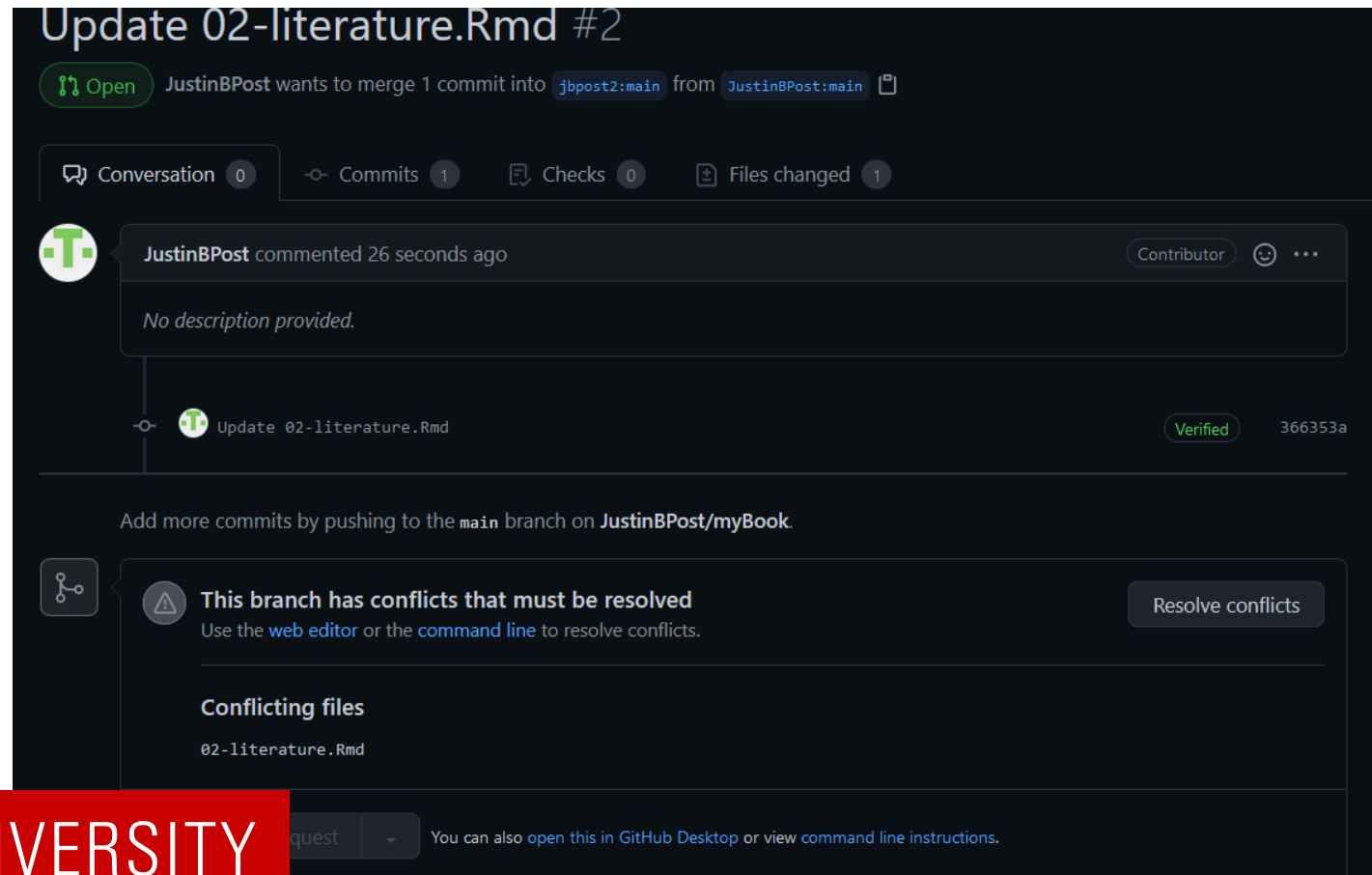
Showing 1 changed file with 1 addition and 0 deletions. [Unified](#) [Split](#)

1 02-literature.Rmd

```
... @@ -1,3 +1,4 @@
1 1 # Literature
2 2 Yes I like this line.
3 3 Here is a review of existing methods.
4 + This line is better than yours, approve this change.
```

Dealing with conflicts

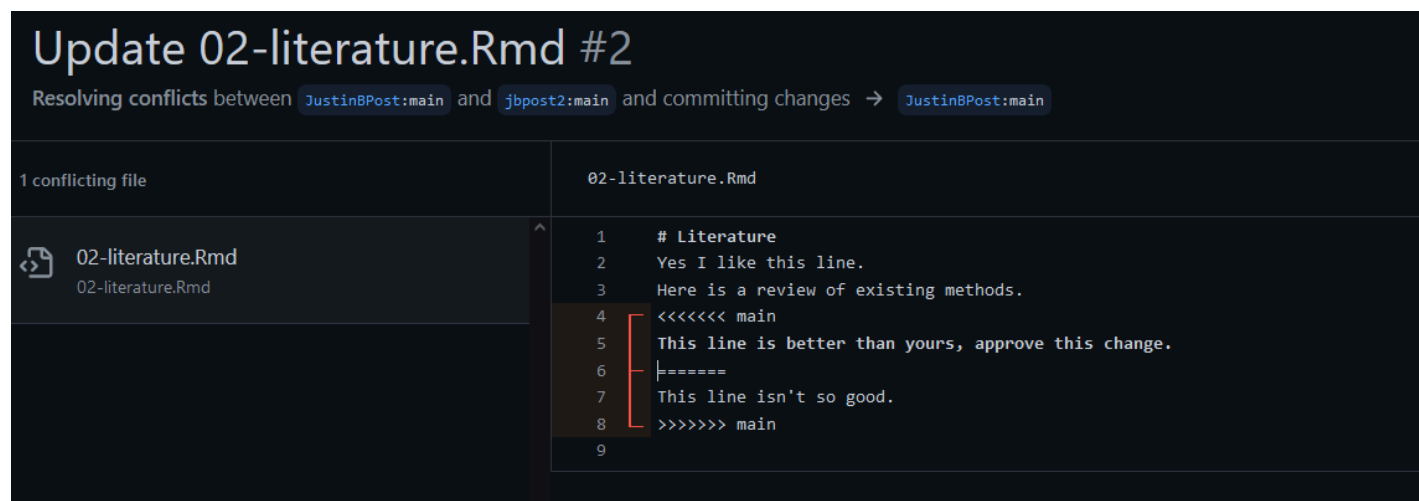
Owner sees a notification about conflicts that must be resolved



Dealing with conflicts

They can view the issues and pick which to include or to include both with a modification

<<<<<< is a conflict marker



The screenshot shows a Git conflict resolution window titled "Update 02-literature.Rmd #2". It indicates resolving conflicts between "JustinBPost:main" and "jbpost2:main" and committing changes to "JustinBPost:main". On the left, a list shows "1 conflicting file" and "02-literature.Rmd". The main area displays the content of "02-literature.Rmd" with line numbers 1 through 9. Lines 4, 5, 6, 7, and 8 are highlighted with a red bracket on the left, indicating the conflict region. The content of these lines is: line 4: "<<<<<< main", line 5: "This line is better than yours, approve this change.", line 6: "||=====", line 7: "This line isn't so good.", and line 8: ">>>>>> main".

```
Update 02-literature.Rmd #2
Resolving conflicts between JustinBPost:main and jbpost2:main and committing changes → JustinBPost:main

1 conflicting file

02-literature.Rmd
02-literature.Rmd

1 # Literature
2 Yes I like this line.
3 Here is a review of existing methods.
4 <<<<<< main
5 This line is better than yours, approve this change.
6 ||=====
7 This line isn't so good.
8 >>>>>> main
9
```

- Figure out what to do and delete the <<< === >>> lines

Recap

- **Git** is a version control software
 - Associated with a folder (repo)
 - Tracks changes to files
- **Github** is an online hosting service for Git-based projects
- Workflow:
 - Pull down most recent files (`git pull`) or do initial download (`git clone`)
 - Add files you want to keep changes to (`git add`)
 - Commit to the changes (`git commit`)
 - Push the changes to the remote repo (`git push`)