Lab 04: Scaffolding reproducible research

Sicheng Wang

Table of contents

1 Reproducible research

1.1 What is reproducible research?

- **Reproducible Research** In academic setting, reproducible research refers to work that can be replicated by other scholars following the steps outlined in the `Methodology` part, resulting in the same results as the original author. As outlined in Recipe #4, the concept of reproducible research specifically refers to reproducible projects, which aims to mitigate the impact of varying computing environments by documenting every step, typically through files like a `README` file outlining the project's objectives and production steps, as well as a `DESCRIPTION` file detailing the computing environment and packages utilized. This approach is particularly valuable for collaborative work, as it allows others to test the reproducibility and applicability of the work across different settings.

- **Computing Environment and Project Structure** The computing environment includes hardware (see Figure 1), the operating system (see Figure 2), and apps (see Figure 3), whereas the project structure refers to the files and folders contained in our project.

- **Git and GitHub** Git is an app which is downloaded on an individual's computer which could be used to track changes in your local repositories while Github is a web which allows individuals to share repositories. When connecting them by Rstudio, the repository one made could upload to one's Github account to let others get it by cloning

Figure 1: Computer Hardware from javaTpoint

Figure 2: Operating System from Welcome to CS

Figure 3: Computer Software from CODERUS

to their own computers or forking to their Github account. And people who clone or fork others' repositories could also edit these repositories on their own computer or in their Github account which would not influence the original author's reporitory. They could use Git app to upload (push) to their own forking repository or replacing the original one if the original author permit that through pull request. Besides, the tracking function of Git makes the changes available to individuals by using the commit button. Under the repository in Github, one could also report issues which shows the interactive function of Github. Though the above functions, Github and Git help our project be more reproducible.

1.2 Why it is important for modern science?

- **In academic**, research is interconnected; scholars build upon previous studies. Moreover, interactions among scholars worldwide are becoming increasingly common. Reproducible studies facilitate the review of others' work and enable scholars to offer more informed suggestions by validating previous research. Additionally, reproducible research can help prevent academic misconduct. For scholars collaborating with peers, reproducible research aids in tracking each others' edits and contributions.

- **For the general public involved in projects**, reproducible research simplifies collaboration, particularly in light of recent significant shifts in work patterns where remote work has become prevalent. With reproducible research, individuals can collaborate more efficiently regardless of their location, enabling them to work remotely with ease.

2 Project management (Taking Lab-04 Repository as an Example)

2.1 Forking

In Github, when you want to copy others' repositories while do not want to download in your own computer, you use fork which is located at below, see Figure 4:

Figure 4: Fork

2.2 Cloning

When you decide to download the repositories you found in Github which may be the one owned by you or shared by others, you could use clone option, as seen in below Figure 5:
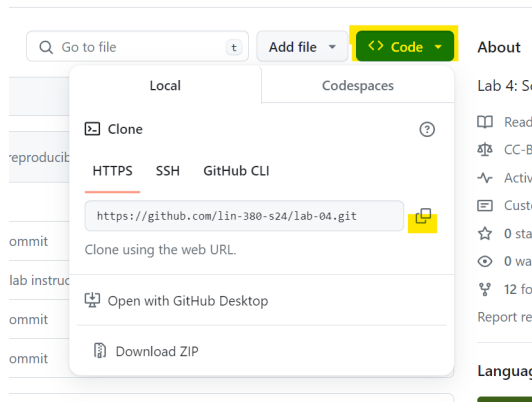

Figure 5: Clone

2.3 Editing

Then, we could edit the repository we clone to our computer, like adjusting files' content, adding or deleting files, and so on based on the purpose of your project.

2.4 Committing

After you made some adjustments to your local repository, you may want to save it. You could `Render on Save`, however, when you are running a long-term project, you may want to revert to the version you done earlier. In that case, you may need to use commit which helps you to track the changes especially when you are collaborating with others. This function is done by Git, here is what you could do when you dealing with the project in Rstudio, see Figure 6:
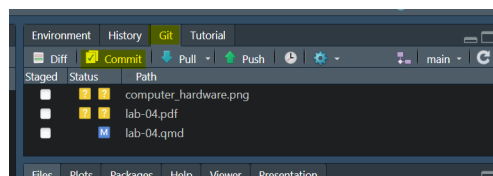

Figure 6: Commit

You could also leave a commit message to remind yourself the changes you made in this version, see Figure 7:
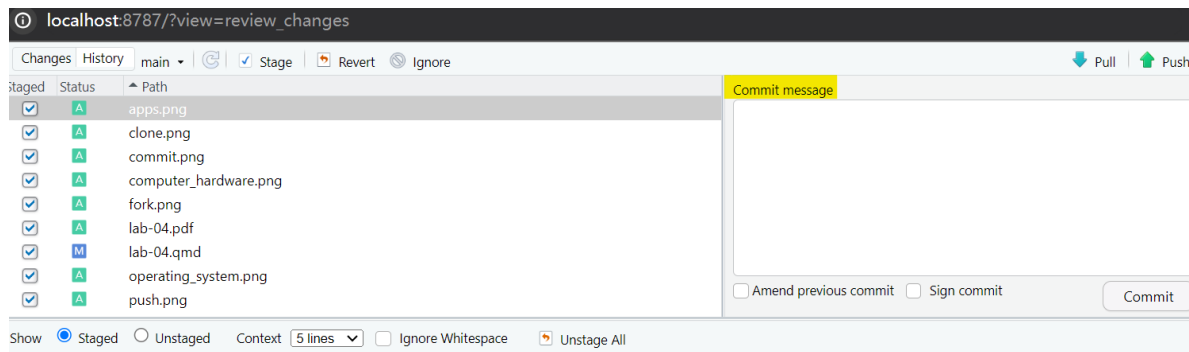
4

Figure 7: Commit message

2.5 Pushing

This step is for individual who wants to refresh the original repository in Github. This is especially useful when you want to keep the repository you shared with others updated or when you are working together with others. Push button helps you to upload the revised local repository to your Github account. In Rstudio's Git pane, your could find the push button, see Figure 8:
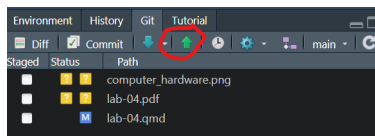

Figure 8: Push

3 Assessment

3.1 What did you learn?

3.2 What did you find most/ least challenging?

3.3 What resources did you consult?

Instructor? R or Quarto documentation, Websites (provide links)?

3.4 What more would you like to know about reproducible research and/ or Git and GitHub?

Find potential resources you might consult to continue your learning. Provide links and a brief description of the resource.