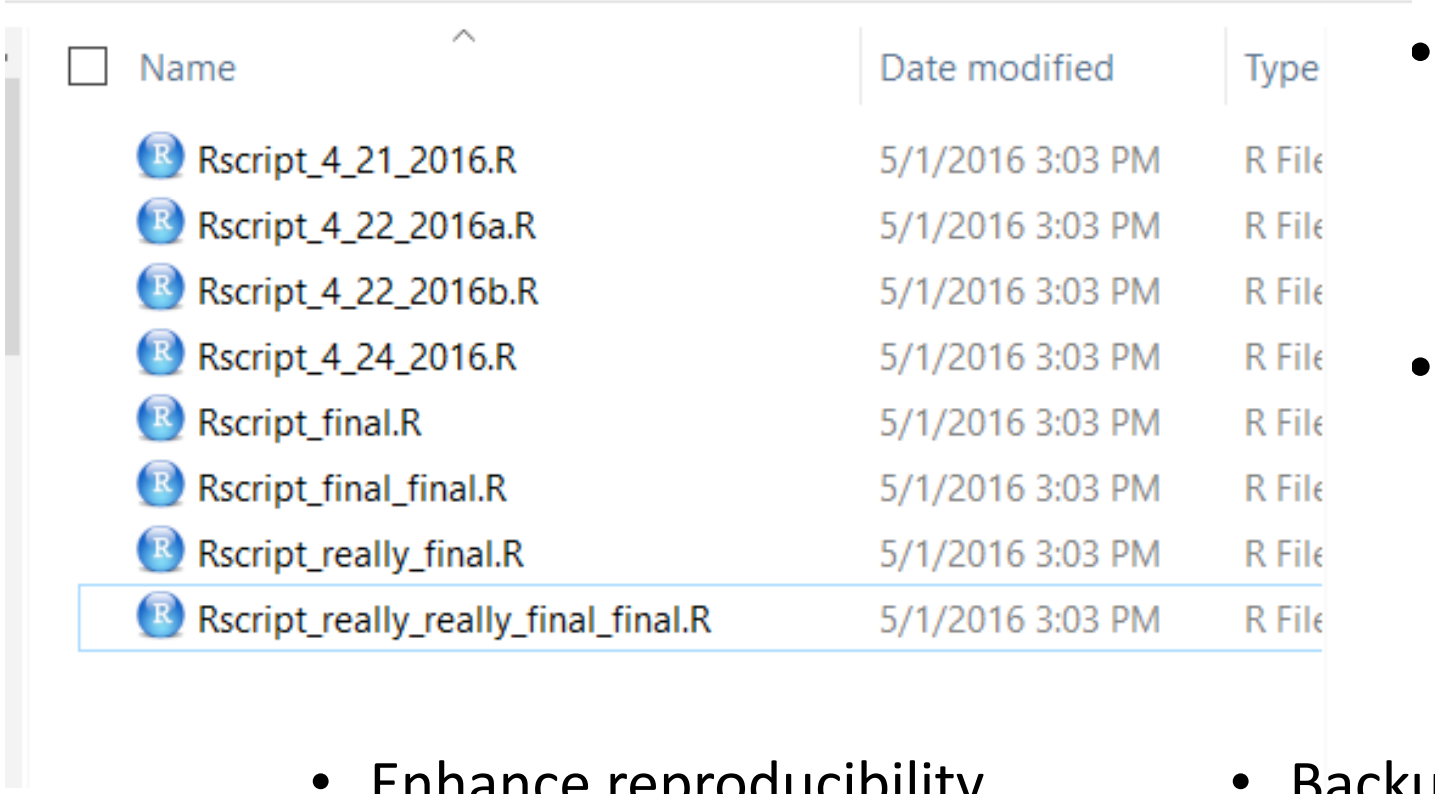
Version Control with Git

Oregon Data Science Collaborative

Spring 2022

What is version control?



- Version control is an organized way of maintaining a record of changes
- Git is a system for distributed version control – not the only one, but popular among scientists

- Enhance reproducibility
- Fix mistakes by reverting to earlier versions
- Improve project structure

- Backup versions in remote repositories**
- Facilitate collaboration**

The mechanics of version control

Single user:

Multiple users:

Image credit: Software Carpentries

Some Git vocabulary

Repository/repo – the collection of files and directories associated with a project and tracked with version control

Commit – a snapshot of a repository's history that is recorded by Git

Diff – Changes in the repository's content associated with the commit

Branches – Concurrent work (changes to file content) can occur in parallel branches, so that you can focus on developing one aspect of the repository/project independently

The basic Git workflow

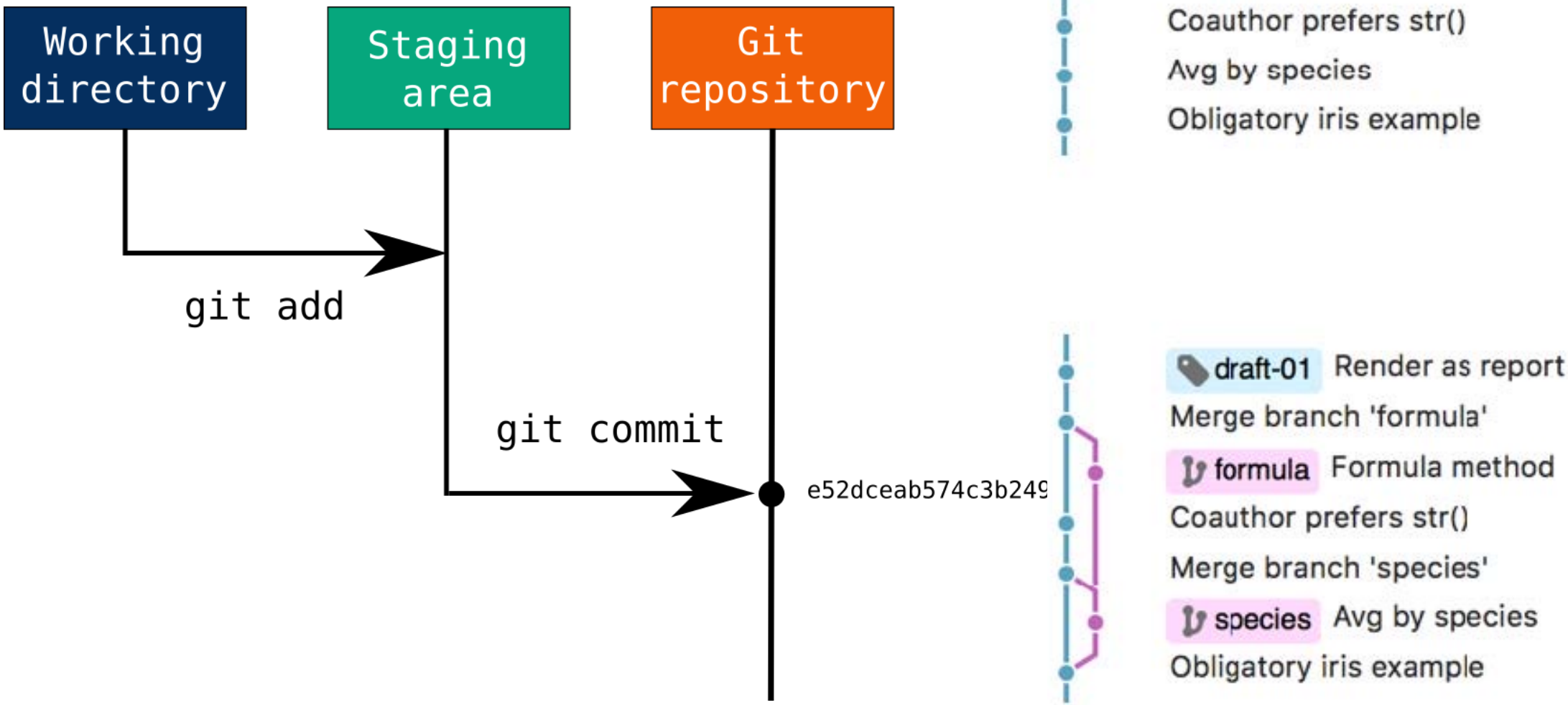


Image credit: J. Bryan; M. Joseph

Use commits to anchor your code

- if you make a mistake, you can’t fall past the previous commit.
- Use more commits when you’re in uncertain or dangerous territory.
- Commits are also helpful to others, because they show your journey, not just the destination.

Hadley Wickham, R Packages

Command line vs. Git client