Git and GitHub

Announcements

- Problem set 1 due this Wednesday at 10pm (submit through GitHub!)
- Future classes will be available via Zoom
- Reminder about hacky hour Friday 3-5

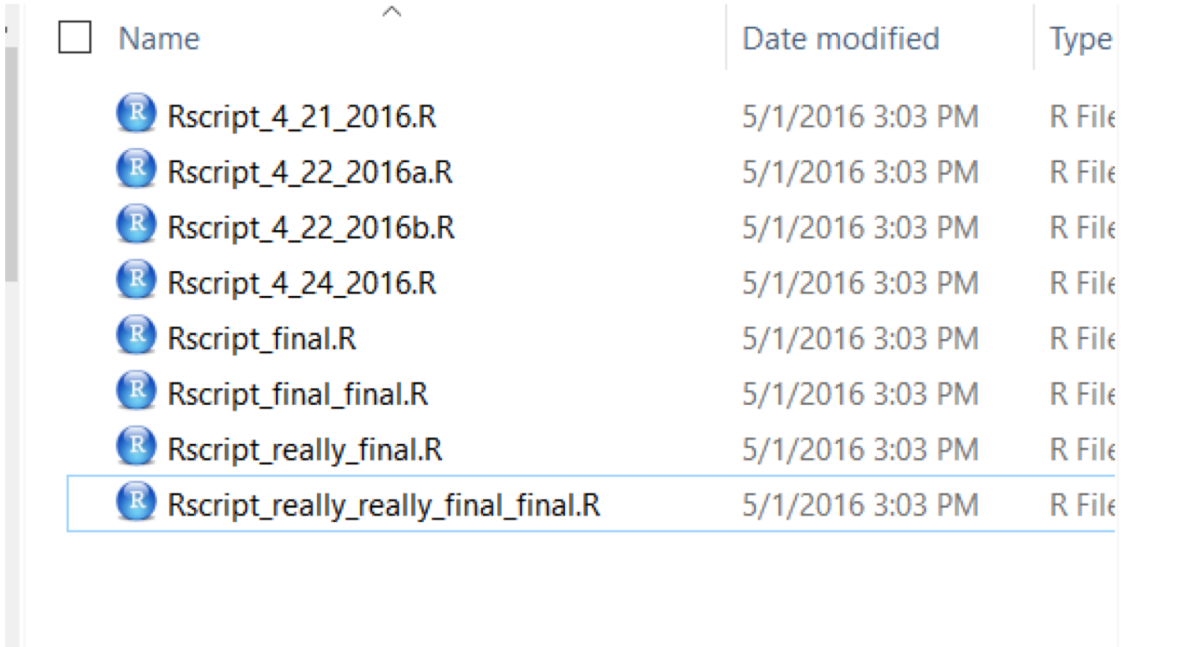
Learning objectives

- Locate your personal GitHub repo through which you'll be submitting homework and exercises for this course
- Create and edit plain text files on GitHub
- Navigate the commit history of a repository and a file on GitHub
- Clone a repository locally using RStudio
- Sync local changes to a file back to remote (and GitHub) with pull, stage, commit, push
- Describe the advantages of a project-oriented workflow in RStudio and set up a version-controlled project directory on your computer

Why Git and GitHub?

Why Git and GitHub?

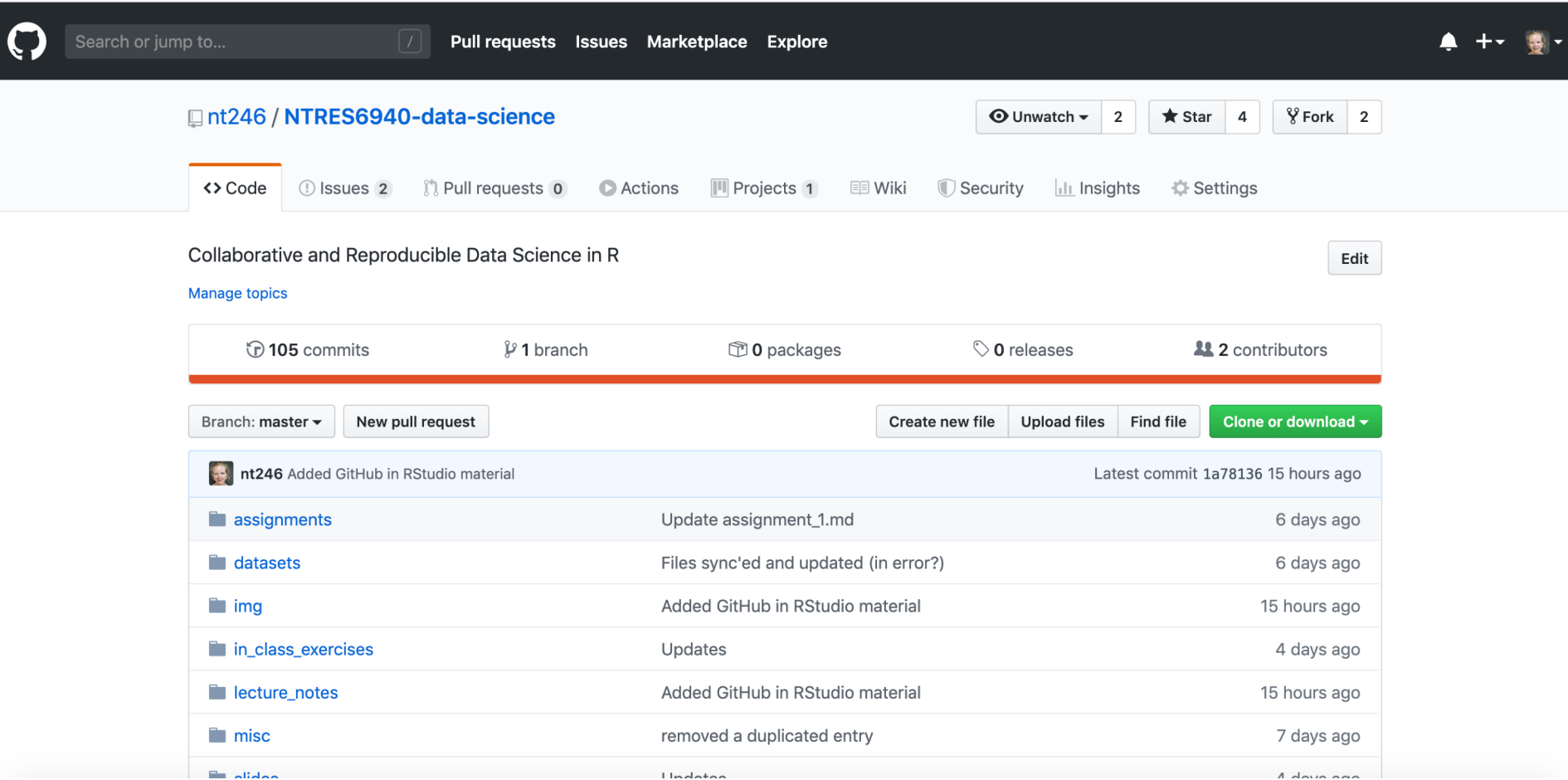
Avoiding this:

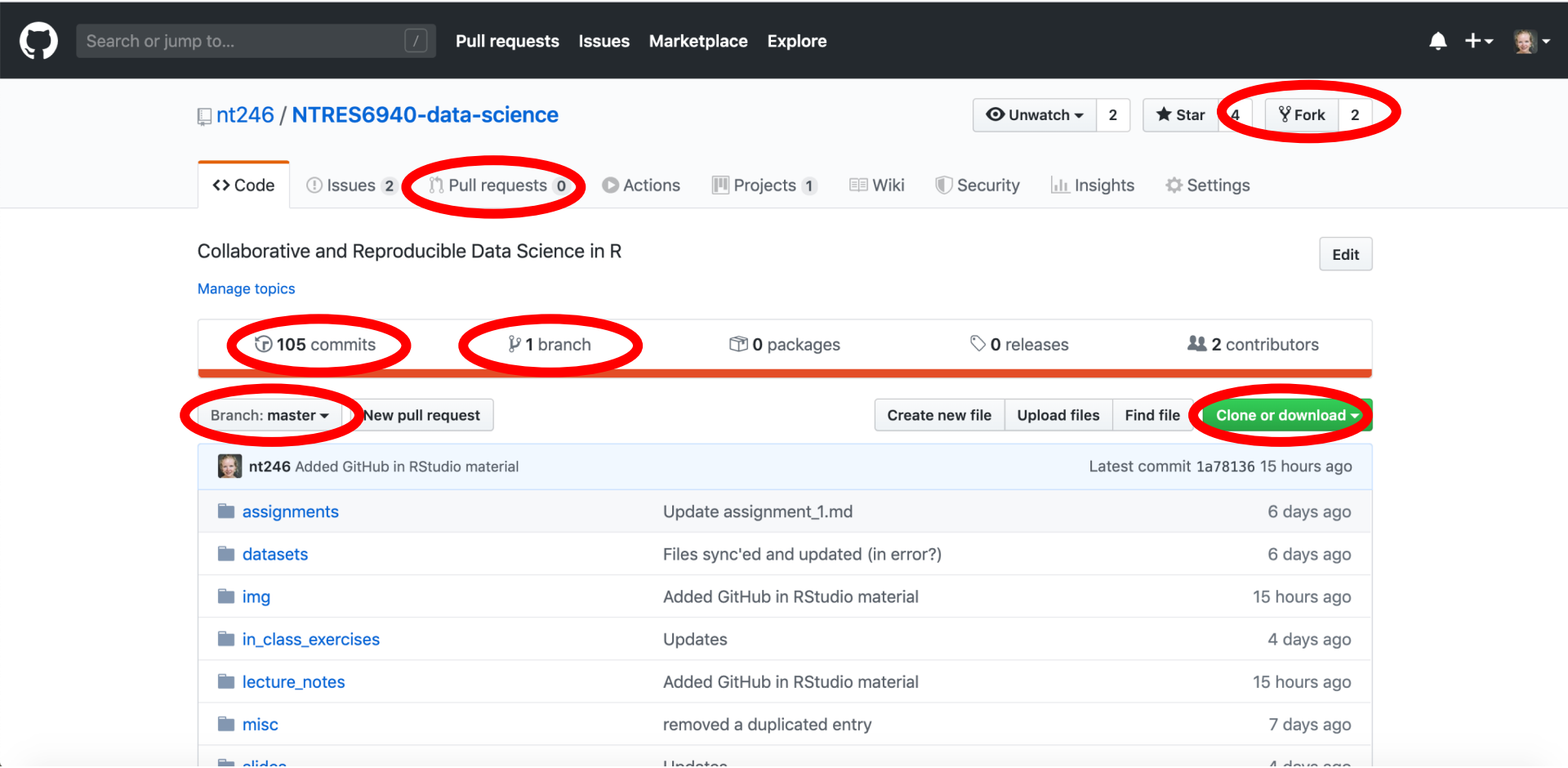


Name	Date modified	Type
Rscript_4_21_2016.R	5/1/2016 3:03 PM	R File
Rscript_4_22_2016a.R	5/1/2016 3:03 PM	R File
Rscript_4_22_2016b.R	5/1/2016 3:03 PM	R File
Rscript_4_24_2016.R	5/1/2016 3:03 PM	R File
Rscript_final.R	5/1/2016 3:03 PM	R File
Rscript_final_final.R	5/1/2016 3:03 PM	R File
Rscript_really_final.R	5/1/2016 3:03 PM	R File
Rscript_really_really_final_final.R	5/1/2016 3:03 PM	R File

Git is the software you will use locally to record changes to a set of files

GitHub is a hosting service that provides a Git-aware home for such projects on the internet
(essentially a social media platform for your Git-versioned files)


Search or jump to...
Pull requests
Issues
Marketplace
Explore
nt246 / NTRES6940-data-science
Unwatch 2
Star 4
Fork 2
Code
Issues 2
Pull requests 0
Actions
Projects 1
Wiki
Security
Insights
Settings
Collaborative and Reproducible Data Science in R
Edit
Manage topics
105 commits
1 branch
0 packages
0 releases
2 contributors
Branch: master
New pull request
Create new file
Upload files
Find file
Clone or download
nt246 Added GitHub in RStudio material
Latest commit 1a78136 15 hours ago
assignments Update assignment_1.md 6 days ago
datasets Files sync'ed and updated (in error?) 6 days ago
img Added GitHub in RStudio material 15 hours ago
in_class_exercises Updates 4 days ago
lecture_notes Added GitHub in RStudio material 15 hours ago
misc removed a duplicated entry 7 days ago


Search or jump to...
Pull requests
Issues
Marketplace
Explore
nt246 / NTRES6940-data-science
Unwatch 2
Star 4
Fork 2
Code
Issues 2
Pull requests 0
Actions
Projects 1
Wiki
Security
Insights
Settings
Collaborative and Reproducible Data Science in R
Edit
Manage topics
105 commits
1 branch
0 packages
0 releases
2 contributors
Branch: master
New pull request
Create new file
Upload files
Find file
Clone or download
nt246 Added GitHub in RStudio material
Latest commit 1a78136 15 hours ago
assignments Update assignment_1.md 6 days ago
datasets Files sync'ed and updated (in error?) 6 days ago
img Added GitHub in RStudio material 15 hours ago
in_class_exercises Updates 4 days ago
lecture_notes Added GitHub in RStudio material 15 hours ago
misc removed a duplicated entry 7 days ago

GitHub

Branching

Forking

Push

Pull request

Cloning

Blame

Commit

Merge

We'll get to all of these….
But one thing at a time!

Branching

Forking

Push

Pull request

Cloning

Blame

Commit

Merge

Key terms

Repository or repo

A directory of files that Git manages holistically

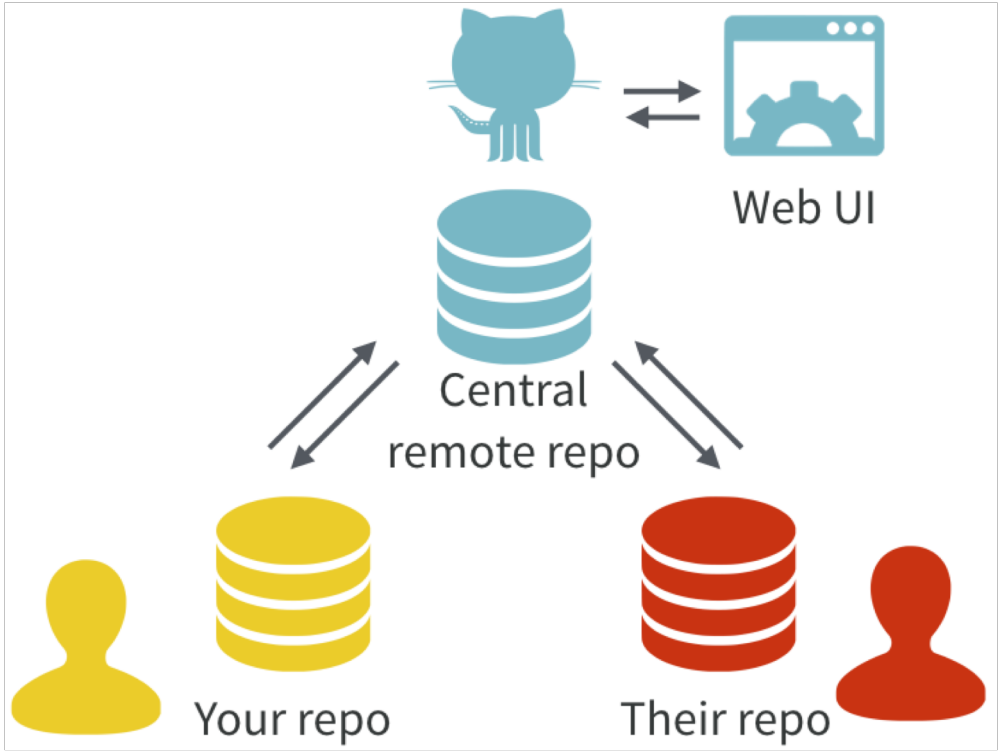
A commit

A snapshot of all the files in the repo, at a specific moment

Diff

Difference between two versions of a file or set of files. Associated with a commit message (summarizes and conveys the motivation for the change)

Local



Local

From “Excuse me, do you have a moment to talk about version control?” by Jennifer Bryan (PeerJ Preprints 5:e3159v2)

Account types

- Repositories (aka “repos”): All files must be organized into repositories. Think of these as self-contained projects. These can either be public or private.
- User accounts vs. organization accounts (aka “Org”): All repositories belong to an account:
  - A user account is the account you just made, and typically holds repositories related to your own work
  - An organization account can be owned by multiple people, and typically holds repositories relevant to a group (like therkildsen-class)

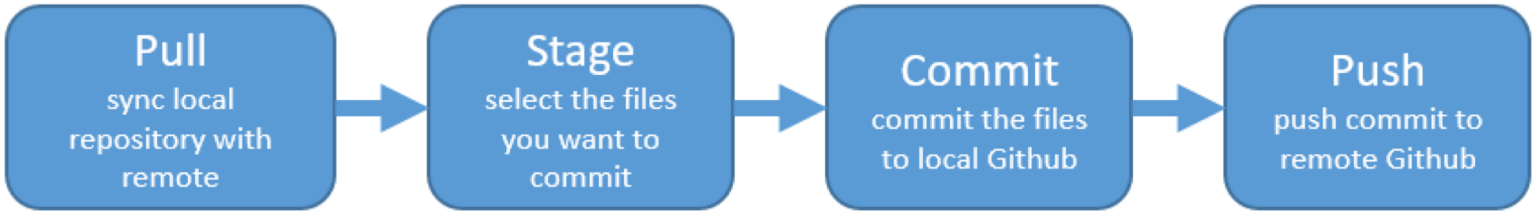
Exercise - editing navigating_github.md

- Add the URL's (with commentary) to
  - your personal class repo
  - your user account page

- Commit the changes

- Next, create a new subdirectory for course notes (remember to not use spaces in names, so you can call it something like course-notes)

Sync from RStudio (local) to GitHub (remote)



File changes

- A Added
- D Deleted
- M Modified
- R Renamed
- ? Untracked

Exercise

Edit your README either directly on GitHub or in RStudio and practice syncing (pull, stage, commit, push). For example,

- Indicate whether you're taking the course for credit
- Add a fun fact about yourself
- Add another line of text with some styling (bolds, italics, headers, bullets etc.)
- Explore your Commit History, and discuss with your neighbor.