git and GitHub for use with R

Tools for versioning and sharing research

Ernest Guevarra

2025-02-10

Until the lions have their historians, tales of the hunt shall always glorify the hunter.

-- African proverb

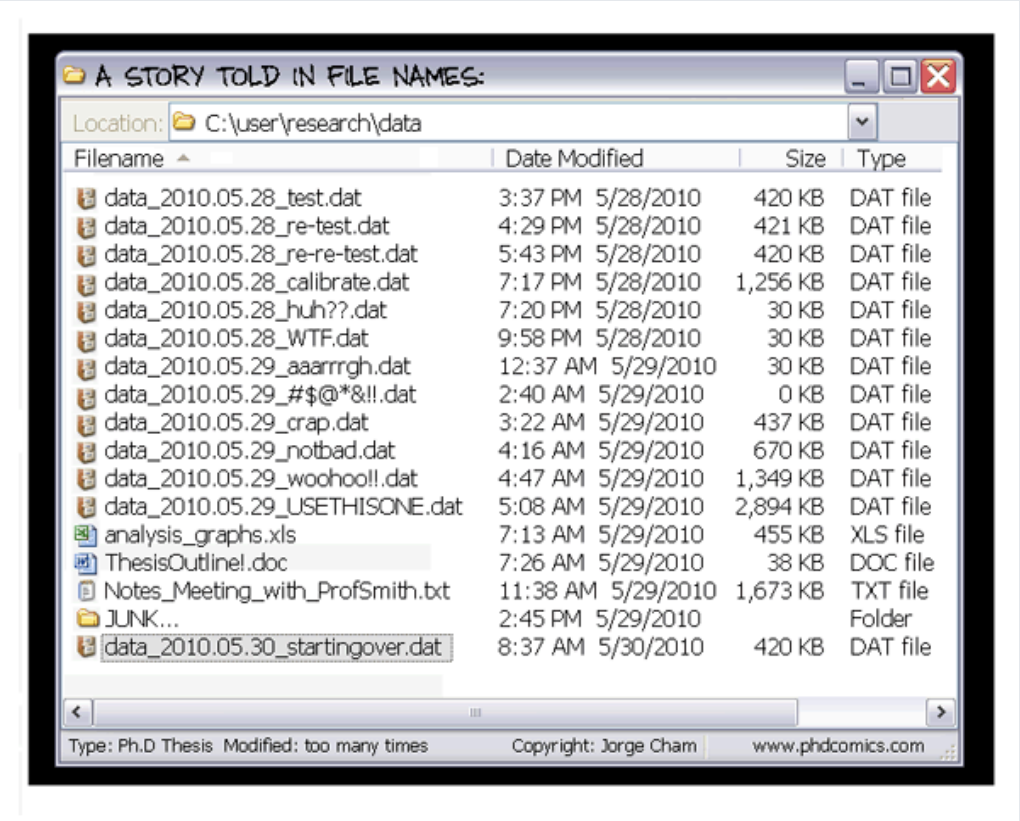
Outline

1. What is git? Why use git?
2. What is GitHub? Why use GitHub?
3. git and GitHub
4. git integration with RStudio
5. Practial session

What is git?

- Free and open source distributed **version control system**
- Built for software development for a group of developers to work collaboratively and to manage the evolution of a set of files
    - like *"Track Changes"* in Microsoft Word on steriods!
- Has been re-purposed to manage a collection of files that make up a typical data analytical project that consists of data, figures, reports, and source code

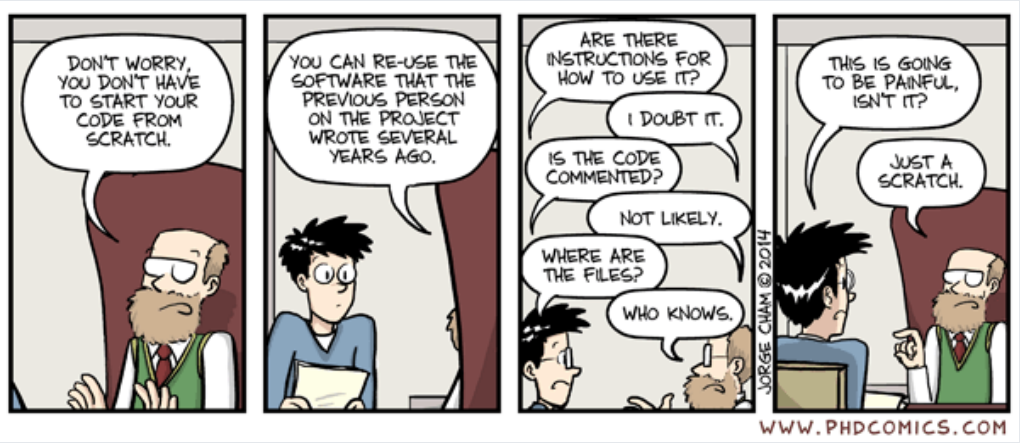
Why use git?



Version control

- Is the only reasonable and sane way to keep track changes in source code, manuscripts, presentations, and data analysis projects
- Documentation of differences between versions
- Exploration of differences between versions

Why use git?



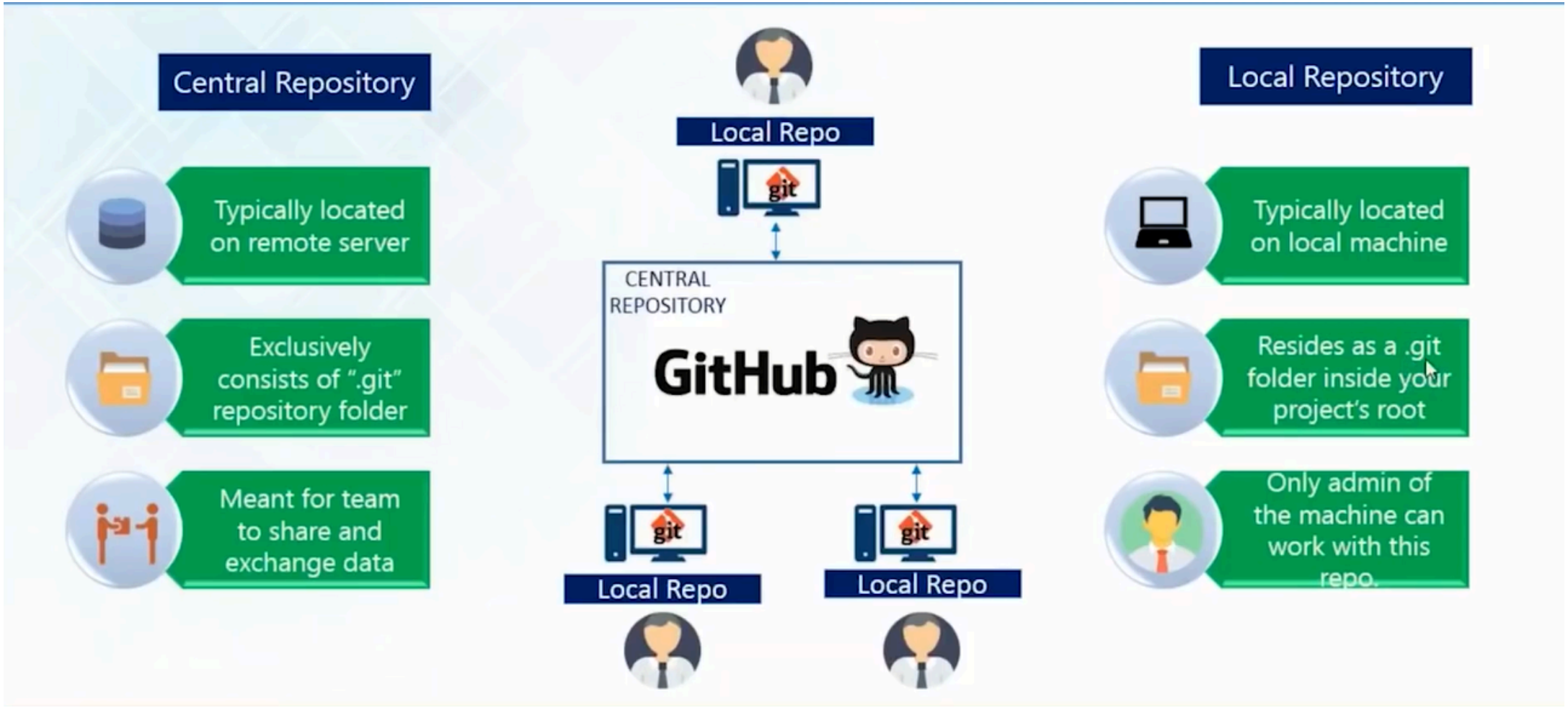
Communication and collaboration

- **Communicating** one's research project with other people is part of the scientific process - not just results but the whole process
- **Collaborating** with others on each other's research project allows us to build on each other's past work, using them for a different context/problem, or re-purposing them to come up with a new approach/solution
- Communication and collaboration on various aspects of the scientific process is faciliated by using git

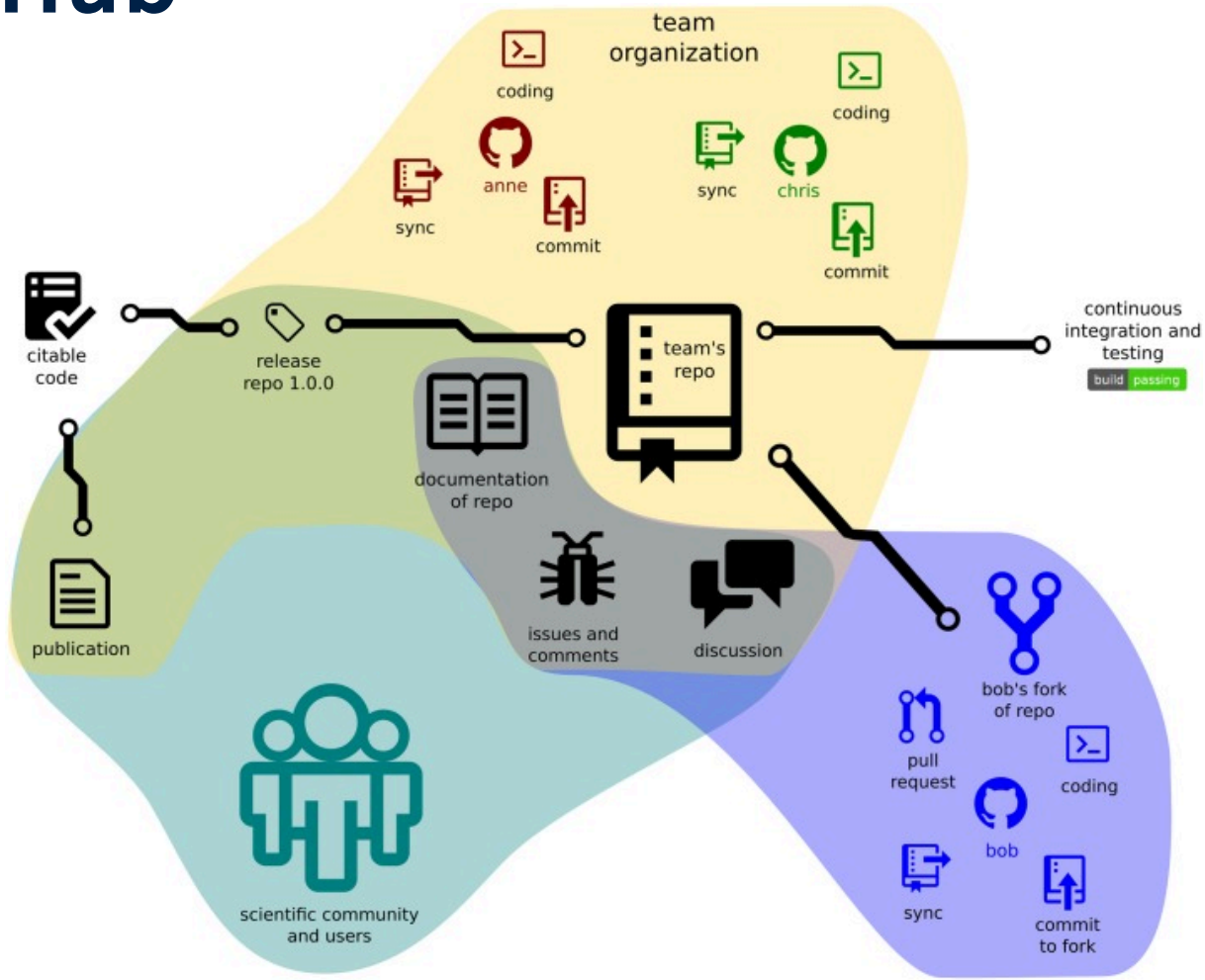
What is GitHub and Why use GitHub?

- Service provider of hosting for software development and version control using git
- Offers distributed version control and source code management functionality of git, plus its own features such as bug tracking, feature request, task management, continuous integration and wikis for every project
- Like *facebook* but for programmers
- Facilitates *"openness"* of **Open Source**
- Lowers the barriers to collaboration

git and GitHub



git and GitHub



Taken from Perez-Riverol, Y., Gatto, L., Wang, R., Sachsenberg, T., Uszkoreit, J., Leprevost, F., Fufezan, C., Ternent, T., Eglen, S. J., Katz, D. S., Pollard, T. J., Konovalov, A., Flight, R. M., Blin, K., & Vizcaíno, J. A. (2016). Ten Simple Rules for Taking Advantage of Git and GitHub. PLoS computational biology, 12(7), e1004947. https://doi.org/10.1371/journal.pcbi.1004947

git integration with RStudio

- RStudio is a popular integrated development environment (IDE) for R
- RStudio has built-in facilities for git and GitHub
- Within RStudio, one can create an RStudio project (a directory with some special files to describe specific RStudio options) which becomes your git repository
- One can easily turn a current git repository into an RStudio project.

Questions?

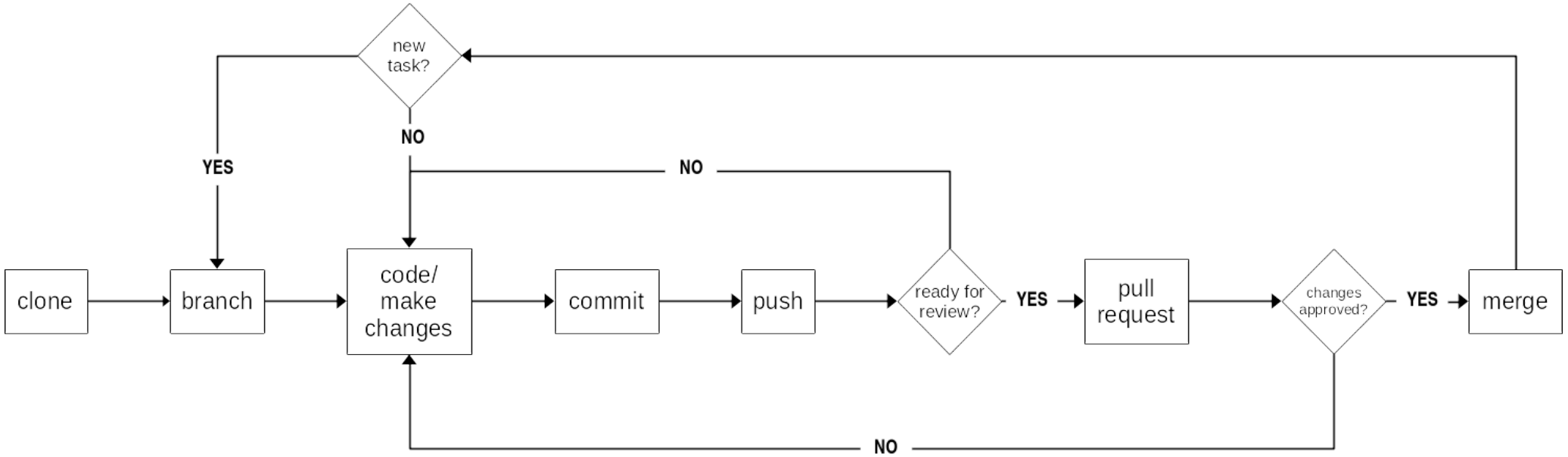
Practical session

- Register a GitHub account
- Install or upgrade R and RStudio
- Install git
- Introduce yourself to git
- Personal access token for HTTPS
- Connect RStudio to git and GitHub

Practical topics

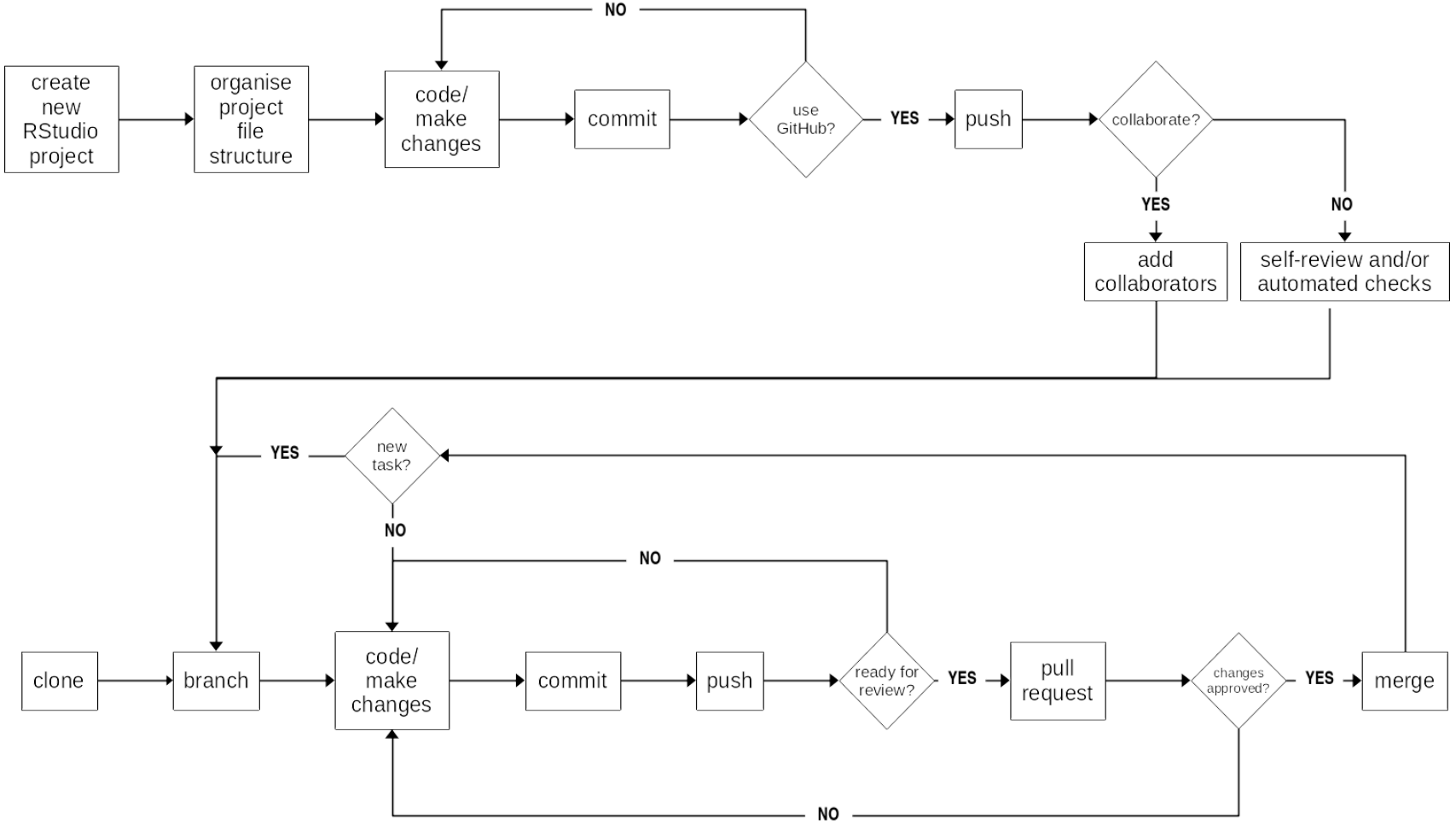
- RStudio, git, and GitHub process for participating in an R-based scientific project/workflow
- RStudio, git, and GitHub process for initiating your own R-based scientific project/workflow

Participating in an R-based scientific project/workflow



see details of this process in this Chapter of the IHTM handbook - https://oxford-ihtm.io/ihtm-handbook/participate-projects.html

Initiating your own R-based scientific project/workflow



see details of this process in this Chapter of the IHTM handbook - https://oxford-ihtm.io/ihtm-handbook/initiate-projects.html

Thank you!

Slides can be viewed at https://oxford-ihtm.io/open-reproducible-science/session7.html

PDF version of slides can be downloaded at https://oxford-ihtm.io/open-reproducible-science/pdf/session7-git-and-github-with-r.pdf

R scripts for slides available here