git and GitHub for use with R Tools for versioning and sharing research

Ernest Guevarra

2024-01-29

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



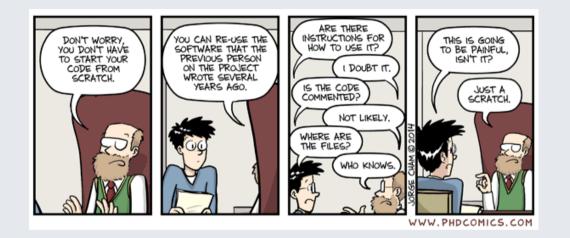
Why use git?

A STORY TOLD IN FILE NAMES			
Location: 😂 C:\user\research\data			*
Filename 🔺	Date Modified	Size	Туре
🚦 data_2010.05.28_test.dat	3:37 PM 5/28/2010	420 KB	DAT file
🚦 data_2010.05.28_re-test.dat	4:29 PM 5/28/2010	421 KB	DAT file
🚦 data_2010.05.28_re-re-test.dat	5:43 PM 5/28/2010	420 KB	DAT file
🚦 data_2010.05.28_calibrate.dat	7:17 PM 5/28/2010	1,256 KB	DAT file
🚦 data_2010.05.28_huh??.dat	7:20 PM 5/28/2010	30 KB	DAT file
🚦 data_2010.05.28_WTF.dat	9:58 PM 5/28/2010	30 KB	DAT file
🔞 data_2010.05.29_aaarrrgh.dat	12:37 AM 5/29/2010	30 KB	DAT file
😝 data_2010.05.29_#\$@*&!!.dat	2:40 AM 5/29/2010	0 KB	DAT file
👸 data_2010.05.29_crap.dat	3:22 AM 5/29/2010	437 KB	DAT file
関 data_2010.05.29_notbad.dat	4:16 AM 5/29/2010	670 KB	DAT file
👸 data_2010.05.29_woohoo!!.dat	4:47 AM 5/29/2010	1,349 KB	DAT file
関 data_2010.05.29_USETHISONE.dat	5:08 AM 5/29/2010	2,894 KB	DAT file
🕙 analysis_graphs.xls	7:13 AM 5/29/2010	455 KB	XLS file
ThesisOutline!.doc	7:26 AM 5/29/2010	38 KB	DOC file
🗈 Notes_Meeting_with_ProfSmith.txt	11:38 AM 5/29/2010	1,673 KB	TXT file
🗀 JUNK	2:45 PM 5/29/2010		Folder
😺 data_2010.05.30_startingover.dat	8:37 AM 5/30/2010	420 KB	DAT file
<]			>
Type: Ph.D Thesis Modified: too many times	Copyright: Jorge Cham	www.phdo	comics.com 🛒

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

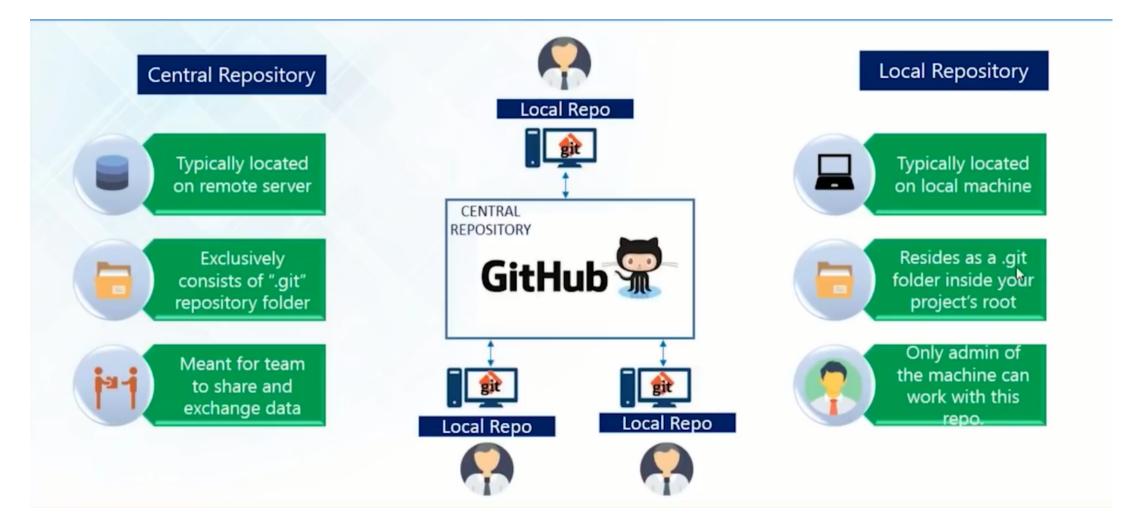
- **Communicating** one's research project with other people is part of the scientific process not just results but the whole process
- **Collaborating** with others on each other's research project allows us to build on each other's past work, using them for a different context/problem, or re-purposing 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?

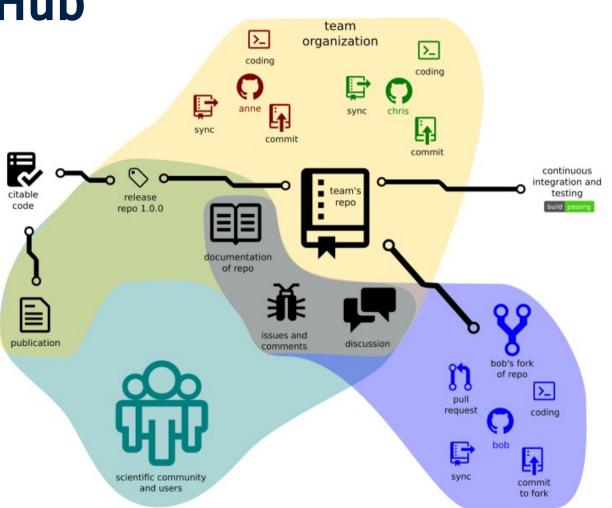


- Service provider of hosting for software development and version control using git
- Offers distributed version control and source code management functionality of git, plus its own features such as bug tracking, feature request, task management, continuous integration and wikis for every project
- 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.

R Studio®

Questions?

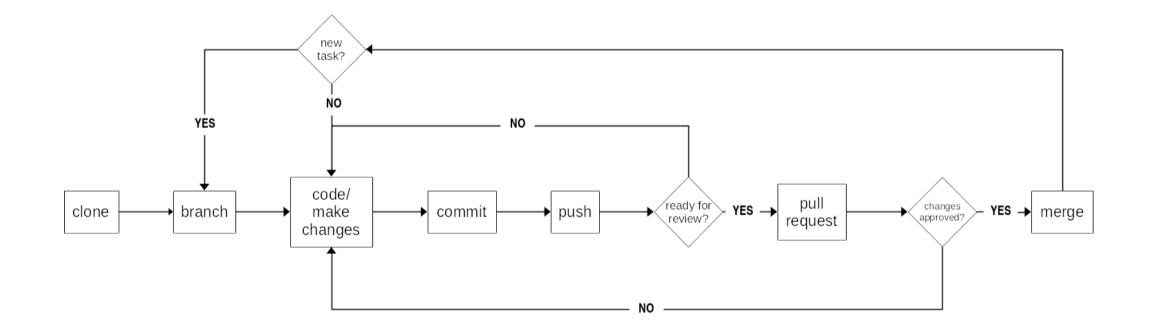
Practical session

- 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

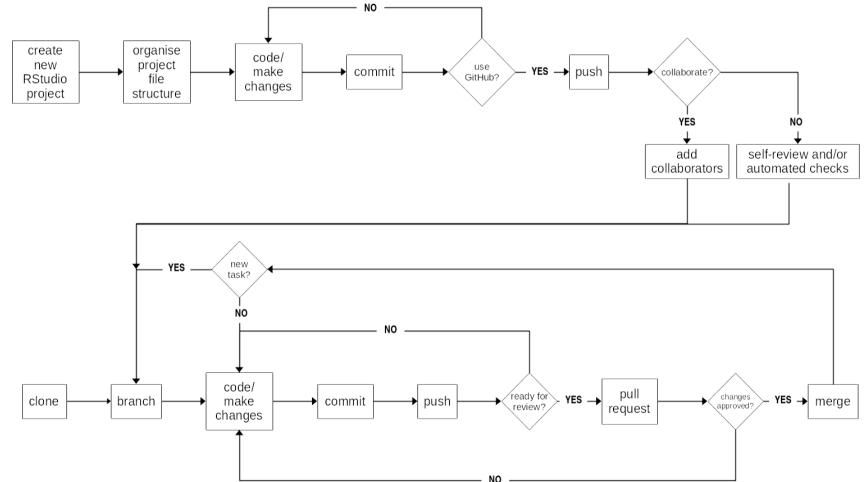
- RStudio, git, and GitHub process for participating in an R-based scientific project/workflow
- RStudio, git, and GitHub process for initiating your own R-based scientific project/workflow

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/session7git-and-github-with-r.pdf

R scripts for slides available her