Getting the right tools for the job

R, RStudio, git, and GitHub

Ernest Guevarra

11 October 2024

Outline

1. What is R?
2. Why use R?
3. What is RStudio?
4. Why use RStudio?
5. What is Git and GitHub?
6. Why use Git and GitHub?

What is R?

- R is a simple but powerful *programming language*
- R is a system for *data manipulation*, *calculation*, and *graphics*. It provides:
    - Facilities/functions for data handling and storage;
    - A large collection of tools for data analysis; and,
    - Graphical facilities for data analysis and display.
- R is a programming language that provides statistical functions as part of a broader programming tool.

Why use R?

- It is an ***open source system*** and is available for ***free***. Even though free, R is ***more powerful than most commercial packages***.
- Considerably ***more flexible than statistical packages*** that relies on menus, buttons, and boxes.
- Every stage of your data management and analysis can be recorded and edited and re-run at a later date.
- huge user and developer community.
- has a robust set of user- and community-developed packages that support statistical methods and techniques.

What is RStudio?

- An ***integrated development environment (IDE)*** for R. RStudio is not R. RStudio is a tool for interfacing with R.
- Includes a ***console***, ***syntax-highlighting editor*** that supports direct code execution, as well as ***tools for plotting, history, debugging and workspace management***.

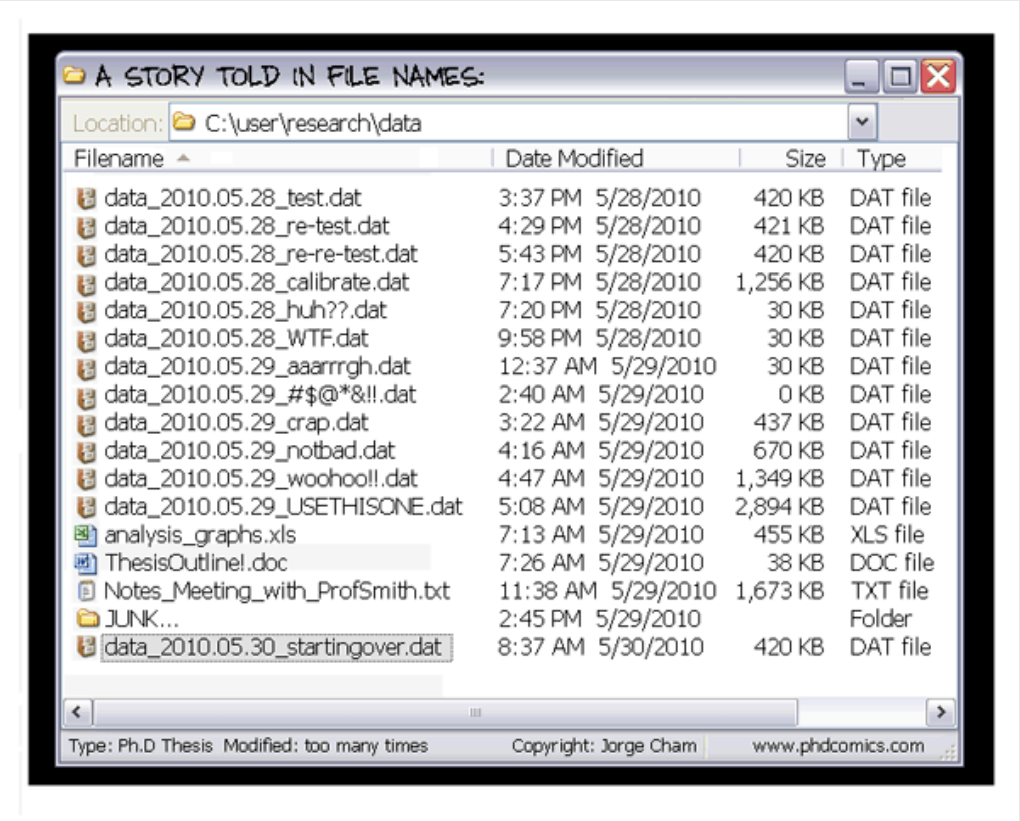
Why use RStudio?

- RStudio is designed to make it easier to work with R
- RStudio facilitates creation of project-orientated workflows
- RStudio makes it convenient to view and interact with the objects in your environment

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 steroids!
- 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

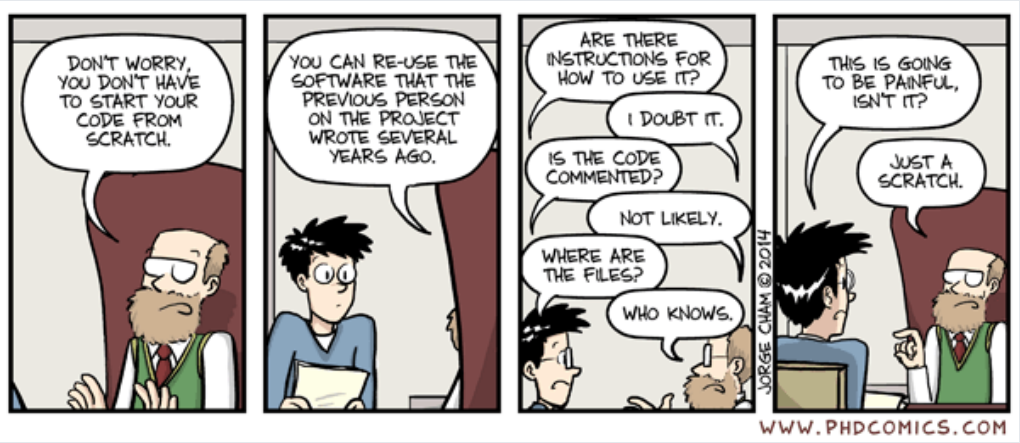
Why use git?



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?



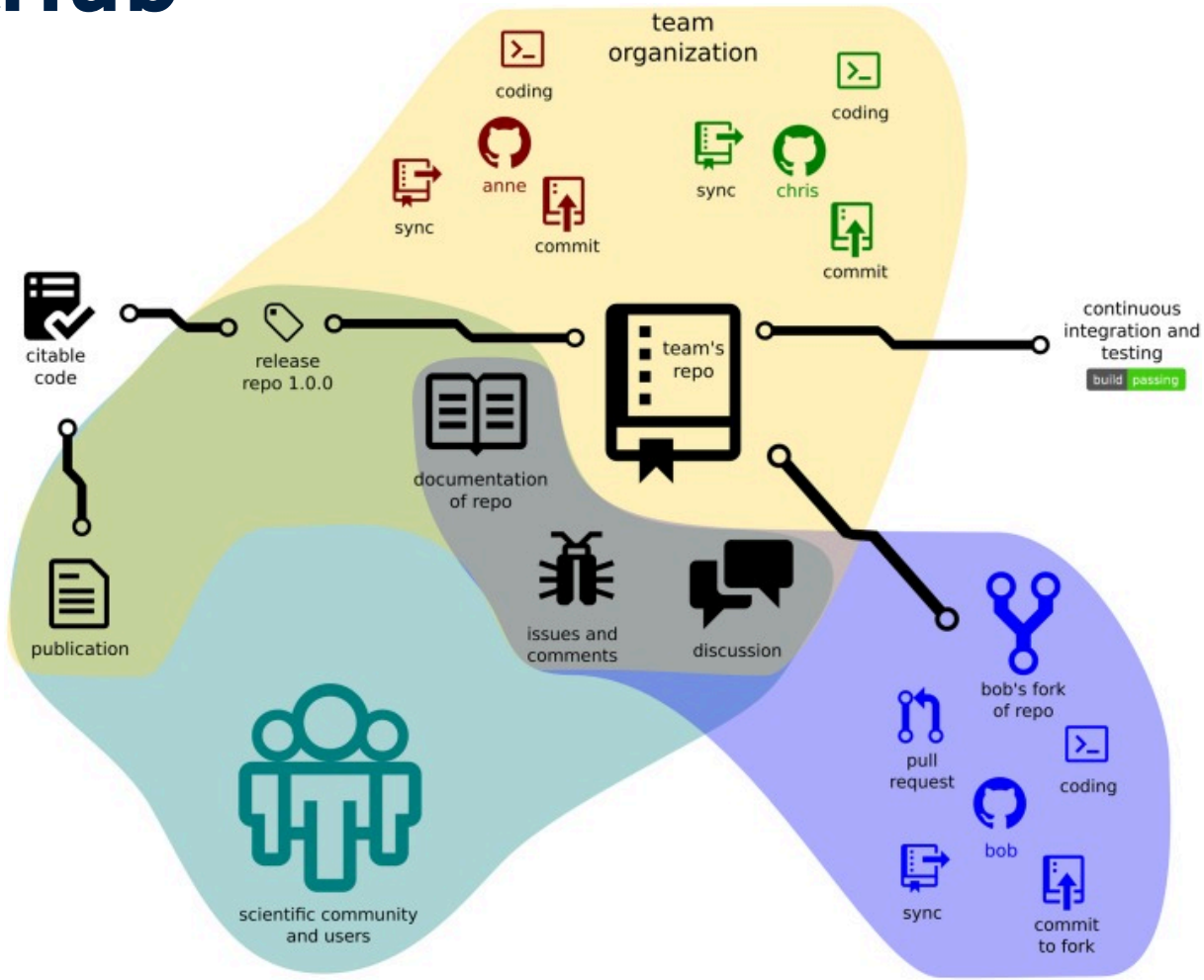
Communication and collaboration

- **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 facilitated 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



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

Questions?

Practical session

Basic git operations with RStudio for retrieving and submitting assignments via GitHub Classroom

Thank you!

Slides can be viewed at https://oxford-ihtm.io/open-reproducible-science/session1.html

PDF version of slides can be downloaded at https://oxford-ihtm.io/open-reproducible-science/pdf/session1-getting-the-right-tools.pdf

R scripts for slides available here