Intro to GitHub for code development and data management

Ian Hussey

What is version control?

- Some parallels with track changes in Microsoft Word
  - Who changed what, where, when, why
  - But better: Can integrate changes made by multiple people, even with conflicts
- Some parallels with syncing/backup software like Google docs/dropbox/sciebo
  - Offsite backup, syncing between multiple team members and/or machines
  - But better: Keep working versions of projects and not just individual files
- Some parallels & integrations with data sharing & preregistration services like OSF

There is definitely a learning curve

But it’s worth it

Nomenclature

- Repo
- Commit
- Push
- Diff
- Fetch
- Pull

- Branch
- Pull request
- Fork

- README
- .gitignore
- .gitattributes

- Integration with .Rmd / Rprojects
- Integration with psychDS data standard

Nomenclature

Core functionality

- Repo - a unit of project organisation
- Commit - local snapshot of files, changes, and annotations
- Push - server/cloud snapshot of commits
- Diff - line-by-line differences between the last and current commit
- Fetch - list all commits on server/cloud that differ from local
- Pull - retrieve the fetched list of commits

Nomenclature

More advanced functionality

- Pull request - request to move commits from one branch to another or one fork into another, e.g. to integrate development work into the main project
- Branch - parallel non-identical versions of a repo allowing for development *within* that repo. E.g., dev by same team with higher tempo of pull requests.
- Fork - parallel non-identical versions of a repo allowing for development *between* repos. E.g., dev by a separate team with a lower (or zero) tempo of pull requests.