Version control for reproducible science

https://stir.ac.uk

Brad Duthie

16 January 2020

Rough outline of version control workshop

1. **14:00** What is version control, and why use it?

2. **14:20** Getting set up – good file management

3. **14:30** The GitKraken interface and simple commits

4. **15:00** Setting up GitHub, pushing and pulling

5. **15:30** Branching using GitKraken

6. **16:00** Merging and merge conflicts

7. **16:30** Independent work using version control

What is version control, and why use it?
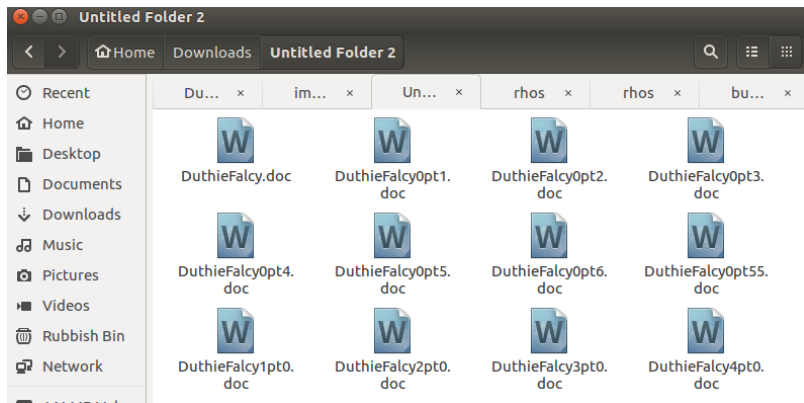
# What version control software does

- Software that records changes you make to files over time
  - Manage different *versions* of files (no need to 'Save As...')
  - Recover old files, keep track of file changes
  - Collaborate with others on shared files

What version control software does

- Software that records changes you make to files over time
  - Manage different *versions* of files (no need to 'Save As...')
  - Recover old files, keep track of file changes
  - Collaborate with others on shared files

- **Put more intuitively**, version control takes a snapshot in time (called a '**commit**') of all the files in one of your folders (called '**repositories**')
  - Visualise changes to your files over time
  - Look at the differences between file versions
  - Record who changed files, and what they changed

Inside of a project on version control
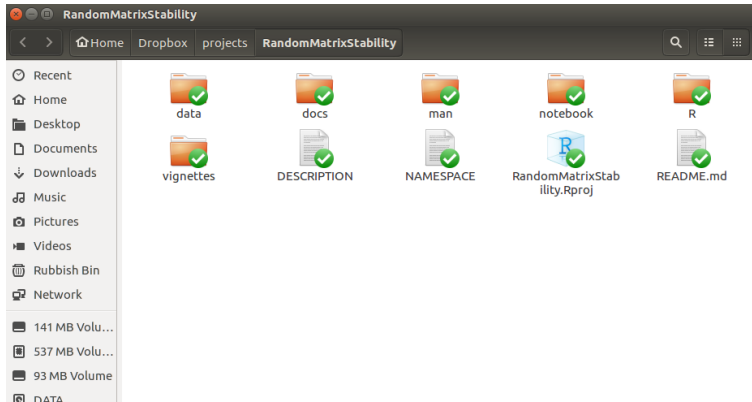
Folders (a.k.a, 'repositories') include all data files, R code, notes, manuscript drafts, etc.

Full annotated timeline of folder changes (GitHub)

Full annotated timeline of folder changes (GitKraken)

Parallel versions ('branches') of a folder (GitKraken)

Collaborative history or a shared folder (GitKraken)
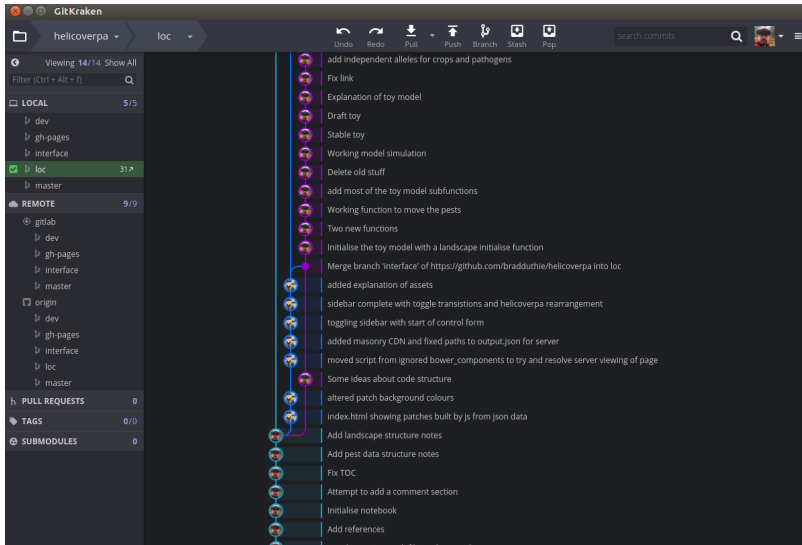
Clear breakdown of what has changed (GitKraken)

Version control makes science easier

- **Organises files** by avoiding 'save as' multiple versions
  - analysis_1.R
  - analysis_2.R
  - analysis_FINAL.R
  - analysis_FINAL_no_really_this_time.R

Version control makes science easier

- **Organises files** by avoiding 'save as' multiple versions

  - analysis_1.R

  - analysis_2.R

  - analysis_FINAL.R

  - analysis_FINAL_no_really_this_time.R

- **Provides a clear history** of what you have done, when, and why (through commit comments)

Version control makes science easier

- **Organises files** by avoiding 'save as' multiple versions
  - analysis_1.R
  - analysis_2.R
  - analysis_FINAL.R
  - analysis_FINAL_no_really_this_time.R

- **Provides a clear history** of what you have done, when, and why (through commit comments)

- **Saves time** by avoiding loss of data, analysis, or writing when integrating with GitHub

Version control makes science easier

- **Organises files** by avoiding 'save as' multiple versions

  - analysis_1.R

  - analysis_2.R

  - analysis_FINAL.R

  - analysis_FINAL_no_really_this_time.R

- **Provides a clear history** of what you have done, when, and why (through commit comments)

- **Saves time** by avoiding loss of data, analysis, or writing when integrating with GitHub

- **Gives peace of mind** to experiment by removing any fear of breaking something that you know works

Version control can help open science
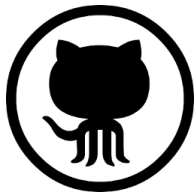
- ▶ Transparent record of data collection, analysis, and writing

- ▶ Record publicly available on GitHub, Bitbucket, or GitLab

- ▶ GitHub repository can be copied, reproduced, and discussed

- ▶ git and GitHub can track individual contributions to a project

Most researchers use git (and GitHub)

- ▶ Free and open-source

- ▶ Separate from GitHub

Most researchers use git (and GitHub)

- Free and open-source

- Separate from GitHub

- Works across platforms
  - Windows
  - Linux
  - Mac

- Invented by Linus Torvalds

Why focus on using GitKraken?

- Free to download and use

- Easy GitHub integration

- Graphical user interface

- Visualisation of repository

Accompanying notes to these slides are available in the version_control repository, and include instructions for using the command line interface, and for editing directly in GitHub.

Objectives: using version control

Guided walkthrough of setting up a project in GitHub and GitKraken to manage a project with version control.

Objectives: using version control

Guided walkthrough of setting up a project in GitHub and GitKraken to manage a project with version control.

Slides:
https://bradduthie.github.com/version_control/vc_slides.pdf

Notes:
https://bradduthie.github.com/version_control/vc_notes.html

Discuss, share, and get additional help by raising an issue in the version_control repository on the Stirling Coding Club.