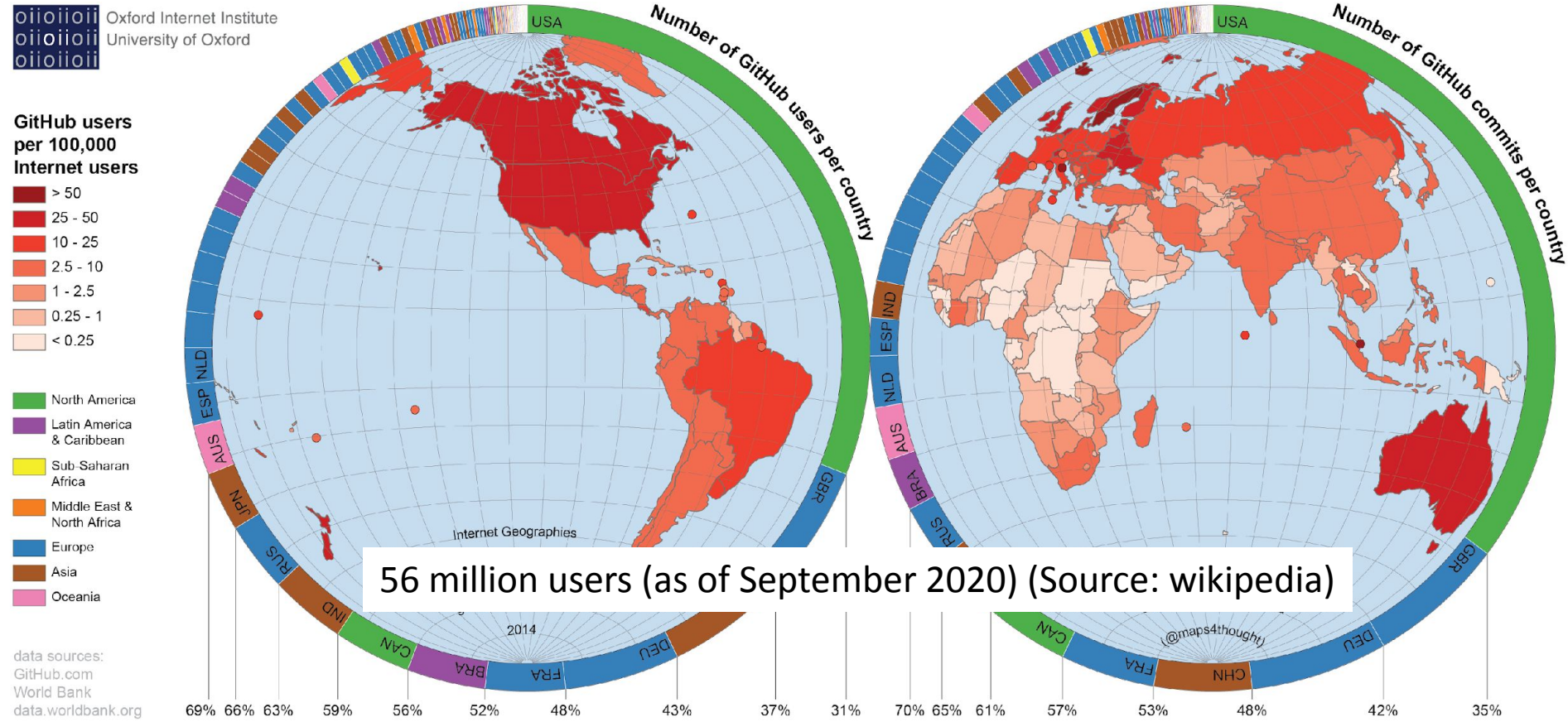
Intro to GitHub

Rosalie Bruel
rosaliebruel@gmail.com
@RosalieBruel
rosalieb

Club stats iEES-Paris, 27/10/2021



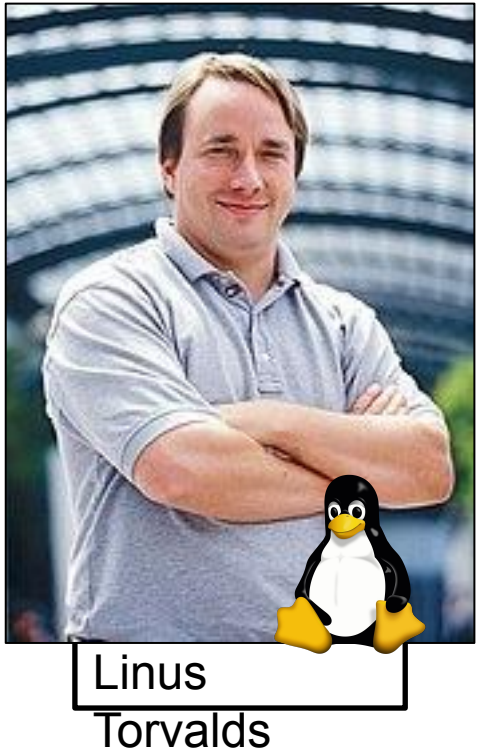
56 million users (as of September 2020) (Source: wikipedia)

GitHub | Mapping collaborative software

What is GitHub?

Git Version control software created by Linus Torvalds, father of Linux (2005)

GitHub A website service that uses Git (so basically a website where you can do version control) (2008)



GitHub is to Git

what Rstudio is to R

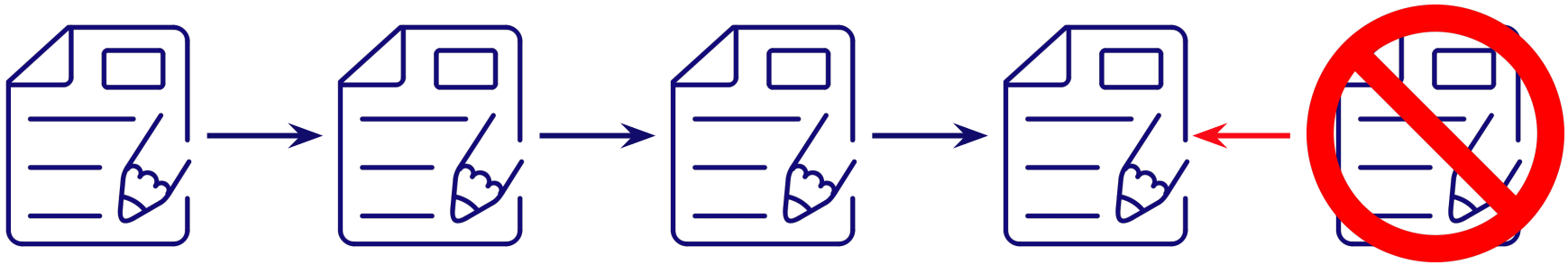
What is version control?

List of content

Version control systems keep track of project updates, providing a detailed history

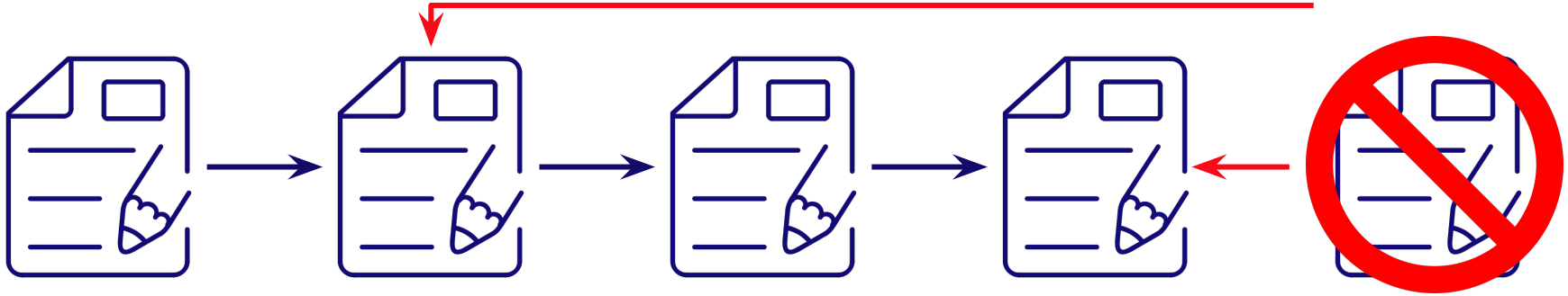
List of content

You can easily revert to a previous version if something breaks



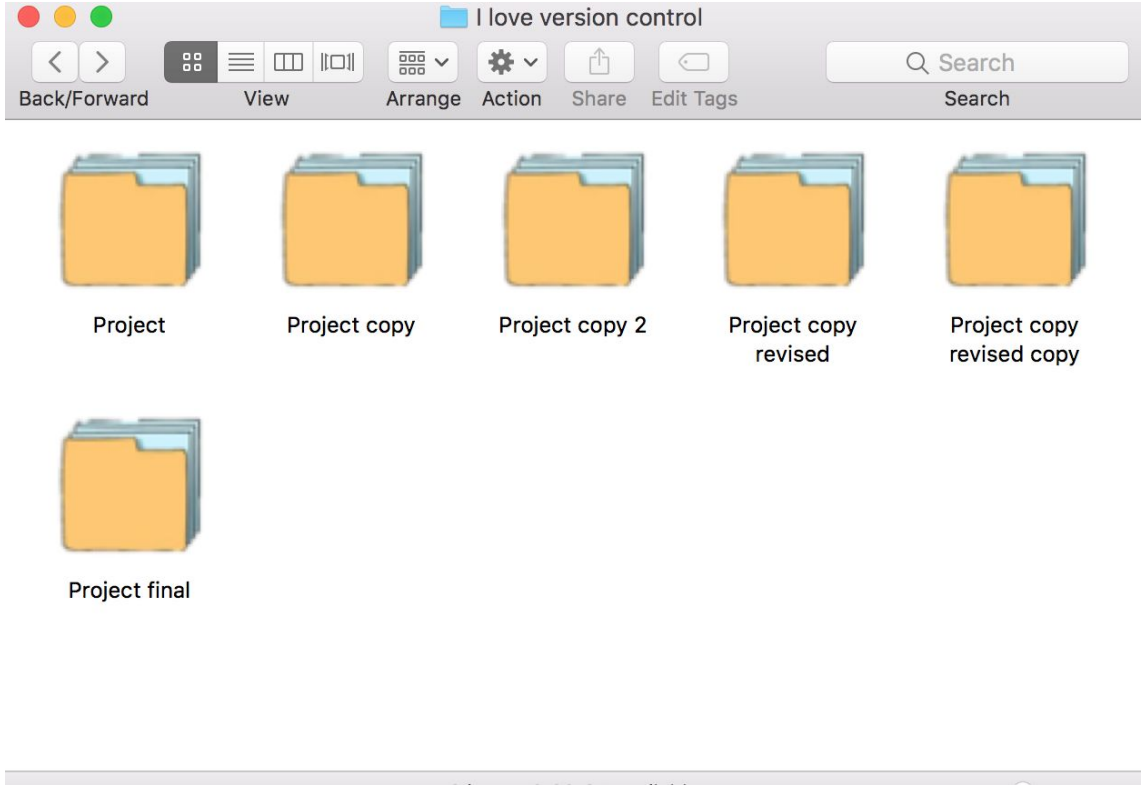
List of content

With version control, you can checkout any commit from the past. If you have multiple branches (alternate code development), you can switch between them



List of content

Why would you even need a version control system?



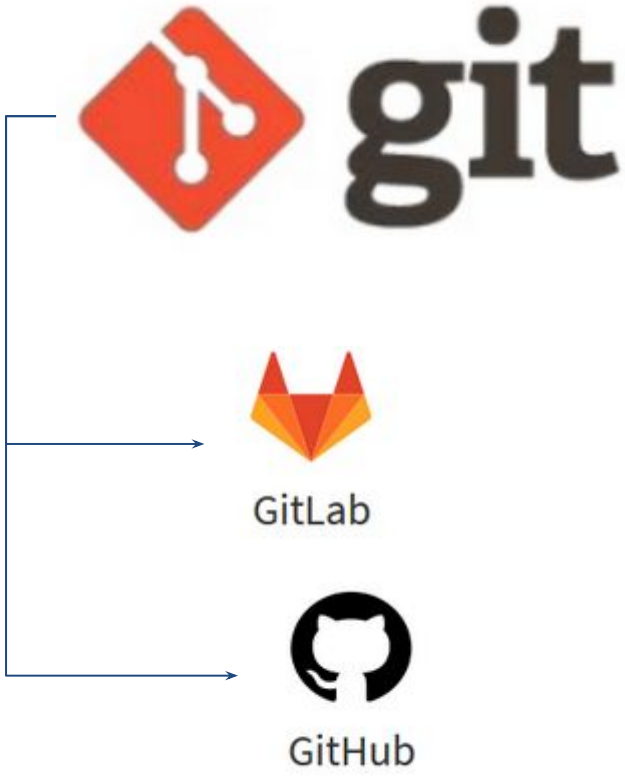
List of content

FYI!

Just so you know… There are many version control systems, each with their strengths and weaknesses. Choose the one that works best for you!
Git is good for codes.

List of content

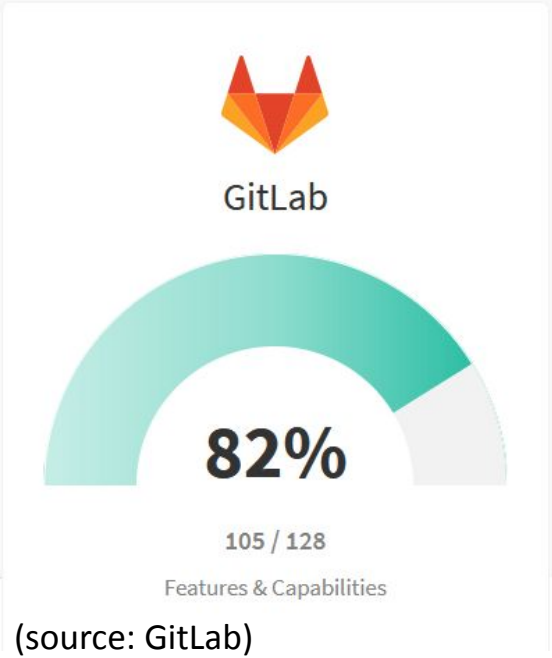
(Better?) alternatives to GitHub exist too.

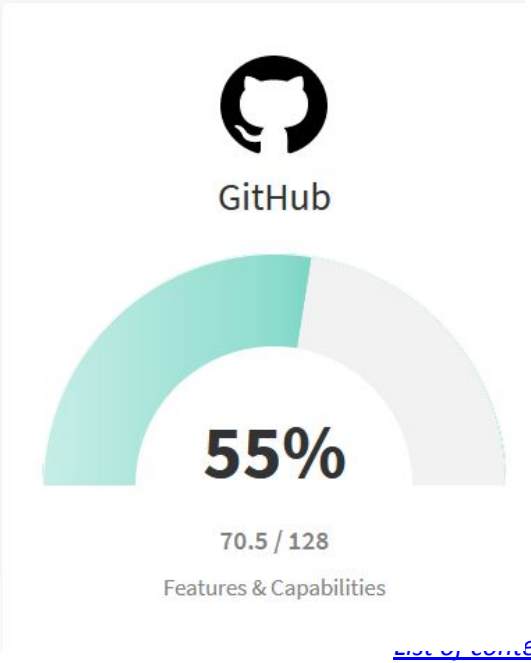


GitLab vs GitHub

Feature maturity comparison against set of industry standards

Learn more about our scoring methodology →

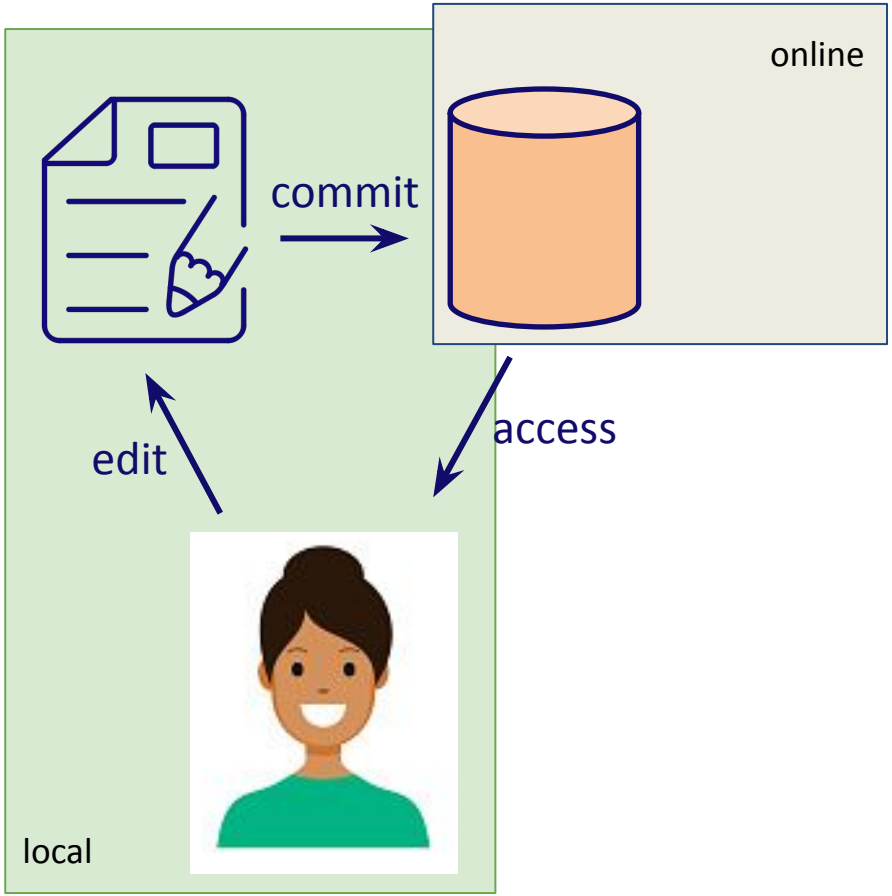




(source: GitLab)

List of content

How does Git work?

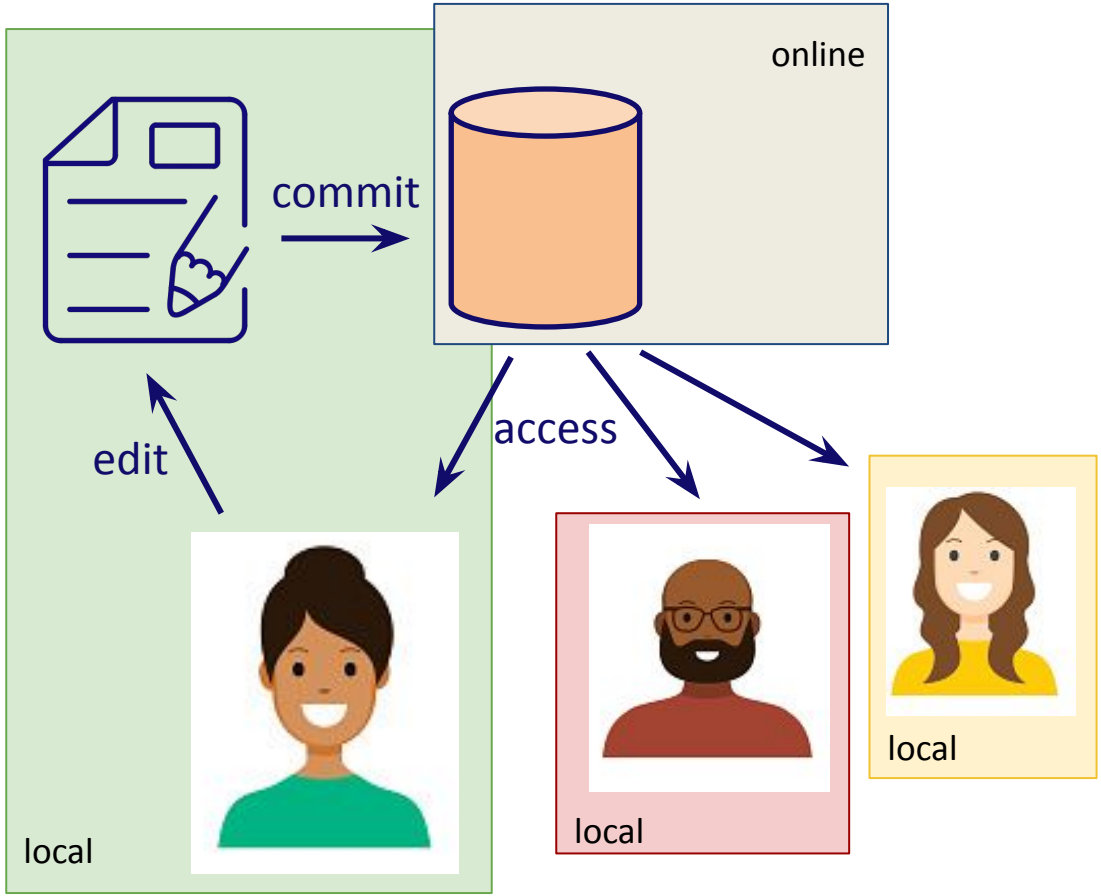


Terminology:

- Repository
- Files
- User
- Commit

List of content

How does Git work?



Terminology:

- Repository
- Files
- Users
- Commit

List of content

Option 1 – stick to the basics

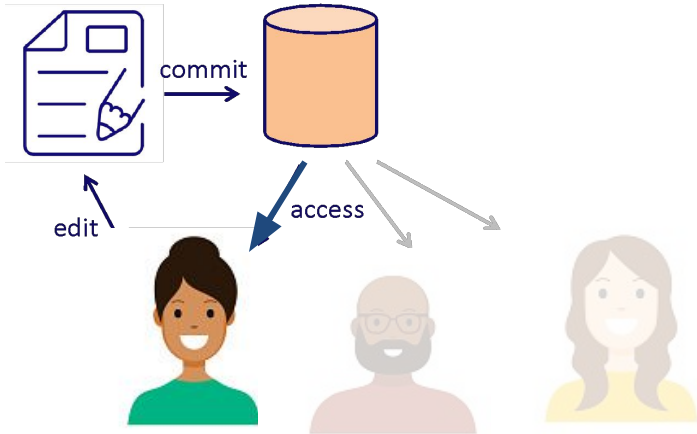
List of content

It’s possible to use GitHub solely for:

- Backups
- Make your codes available after paper publication
- Access codes from publications
- Access codes not implemented in packages

Terminology:

- Repository
- Files
- Users
- Commit



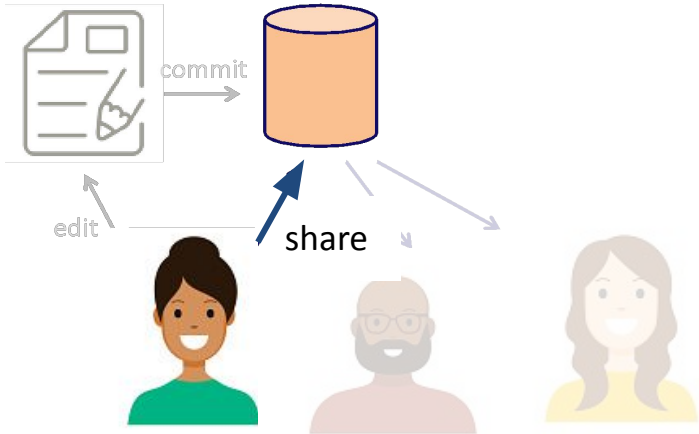
List of content

It’s possible to use GitHub solely for:

- Backups
- Make your codes available after paper publication
- Access codes from publications
- Access codes not implemented in packages

Terminology:

- Repository
- Files
- Users
- Commit



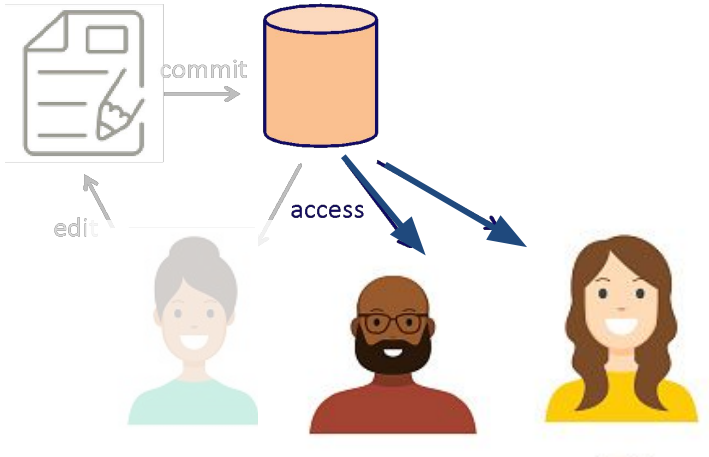
...t of content

It’s possible to use GitHub solely for:

- Backups
- Make your codes available after paper publication
- Access codes from publications
- Access codes not implemented in packages

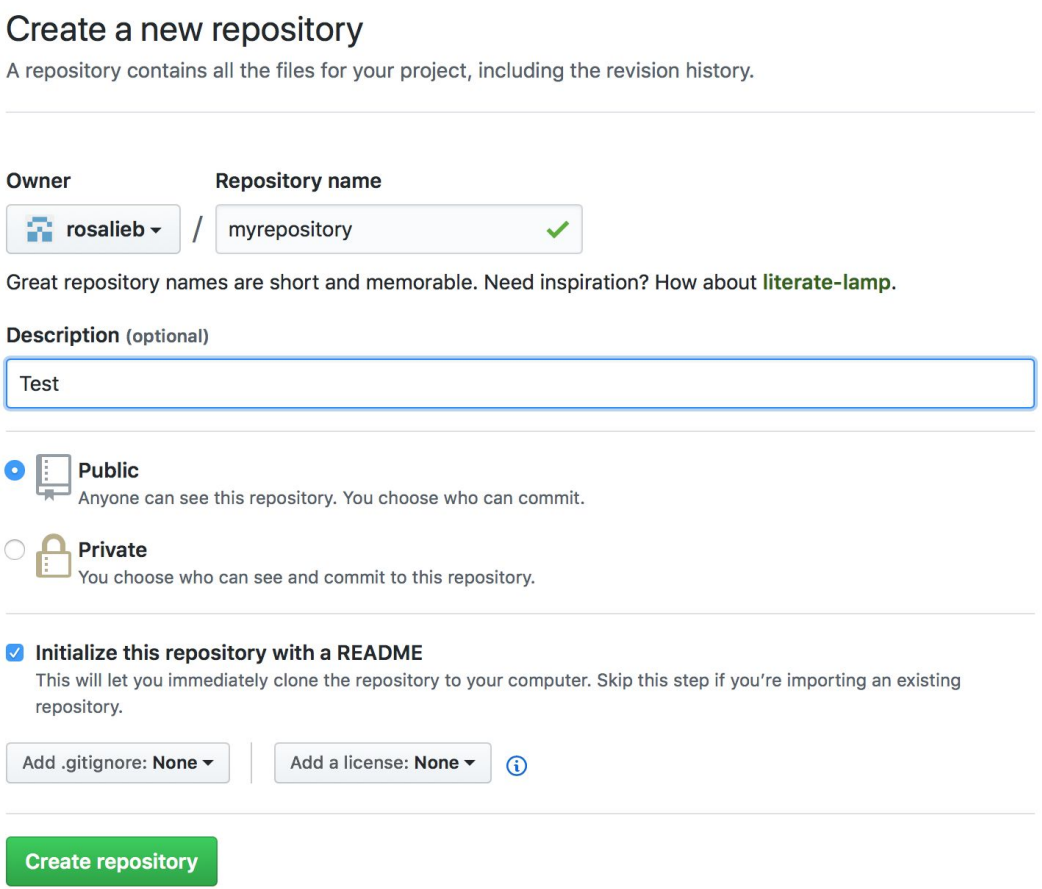
Terminology:

- Repository
- Files
- Users
- Commit



...t of content

Backups/make your codes available: create a repository



Access codes from publications / blogs / other: visit author's repository

Unless stated otherwise, all analyses were conducted using R (version 3.1.3, R Core Team, 2015) with several additional packages: vegan (version 2.2-1, Oksanen et al., 2014) was used to fit the ordinations, analog (version 0.16-0, Simpson, 2007; Simpson and Oksanen, 2014) was used to fit the principal curves and time tracks, and additive models were fitted using the mgcv package (version 1.8.4, Wood, 2004, 2006, 2011). Additional R functions written by GLS (based on suggestions by Simon Wood, pers. comm.) were used to evaluate the first derivative of the trend splines and form the pointwise confidence interval. R scripts implementing the analyses and reproducing the figures are available online from https://github.com/gavinsimpson/bennion-frontiers-2015.

Bennion *et al.*, 2015

We solved the boundary-value problem (5) with the bvpSolve() package in R using function bvptwp() (Mazzia et al. 2014). An R script to illustrate the method using a simple ecological model is found at **https://github.com/SRCarpen/Exit_Time_R**

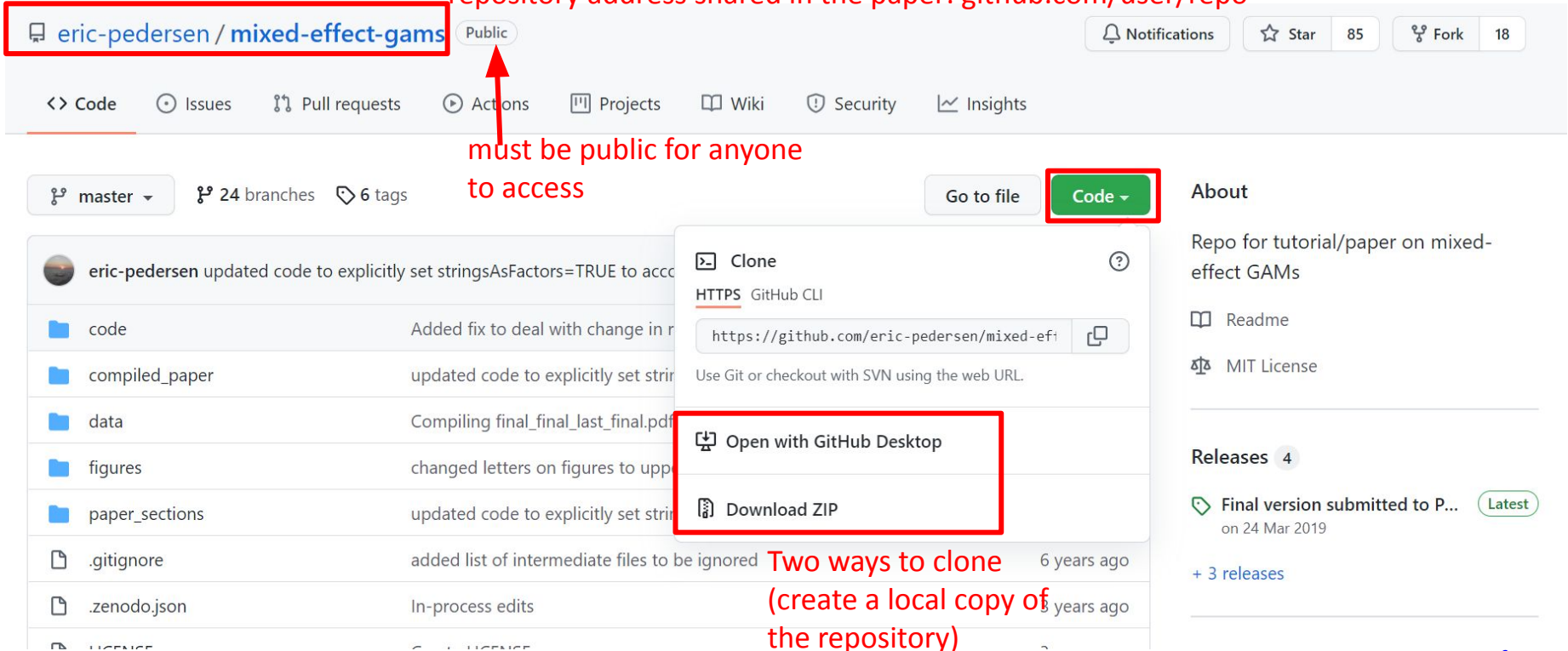
Carpenter *et al.* (2021)

fitting these models, and demonstrate how to fit HGAMs on example data. All code and data used to generate this paper are available at: github.com/eric-pedersen/mixed-effect-gams.

Pedersen *et al.* (2019)

List of content

Access codes from publications / blogs / other: visit author's repository



List of content

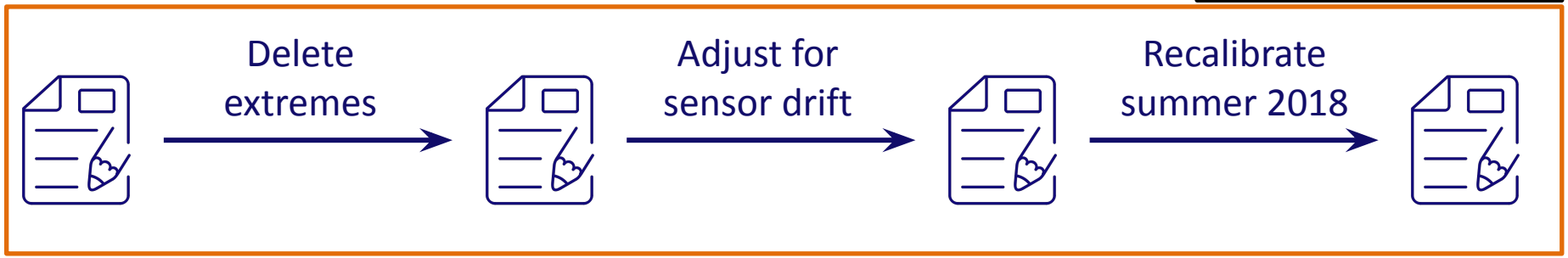
Option 2 – use the version control capacities

List of content

Git for version control

Terminology:

- Repository 
- Files 
- Users 
- Commit
- Master branch



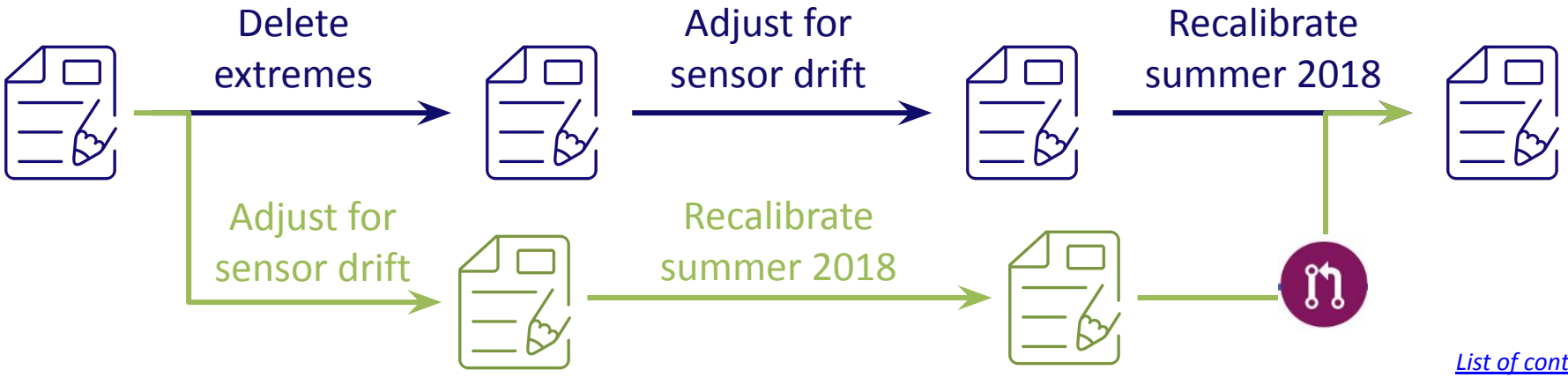
List of content

Git for version control

Featured branches are ‘alternative realities’ that allow you to test scripts

Terminology:

- Repository
- Files
- Users
- Commit
- Master branch
- Feature branch
- Pull request



List of content

Git for version control

Terminology:
- Repository 
- Files 
- Users 

fixed nls() error, added transformations

master

Browse files

taylorstewart authored and **taylorstewart** committed on Feb 2, 2015

1 parent b47351d commit a826adaf3f4d0ce33ff9d1ab6921b7d0be09cfea

Showing **3 changed files** with **32 additions** and **28 deletions**.

Unified | Split

41 server.R

View

```
@@ -196,7 +196,7 @@ shinyServer(function(input, output, session) {
196 196         m <- m[m$season == input$season,]
197 197       }
198 198       # Optional: filter by serial
199     -     if (!is.null(input$serial) && input$serial != "All") {
    199 +     if (!is.null(input$serial) && input$serial != "") {
200 200         m <- m[m$serial == input$serial,]
201 201       }
```

Change at line 199: “All” was deleted

Autumn 2016 Merge #3

Merged **taylorstewart** merged 3 commits into `taylorstewart:master` from `lake-erie-biological-station:master`

Conversation 0 | Commits 3 | Checks 0 | Files changed 38

taylorstewart commented on Dec 4, 2016 — Owner

No description provided.

Unique ID per action

lake-erie-biological-station added some commits on Jul 19, 2016

- `Merge pull request #4 from taylorstewart/master` ··· — `69c2649`

  `significant "house cleaning" of repository`

- `data quality review` ··· — `bd126dd`

  `This corrects and issue with the wingspread measurements and derived values per swept area: also two large lake sturgeon weights were added based up on published length weight regressions.`

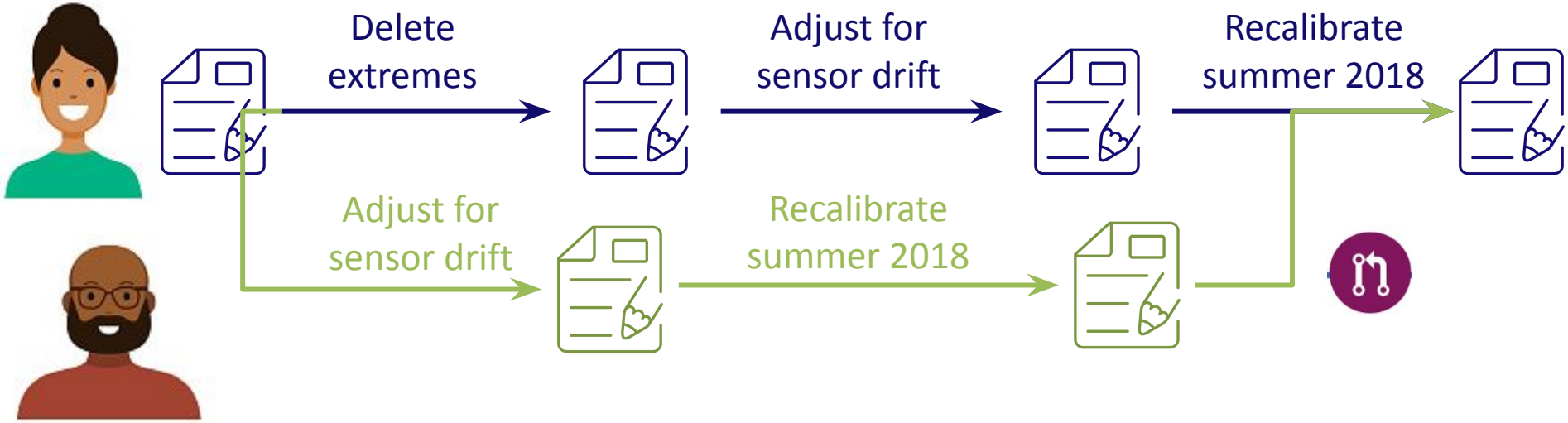
- `UPDATED with Autumn 2016 data` ··· — `bbbe03e`

  `still needs some small revisions to fix MathJax and review the water quality table for functionality`

 **taylorstewart** merged commit **8e14865** into `taylorstewart:master` on Dec 4, 2016 — Revert

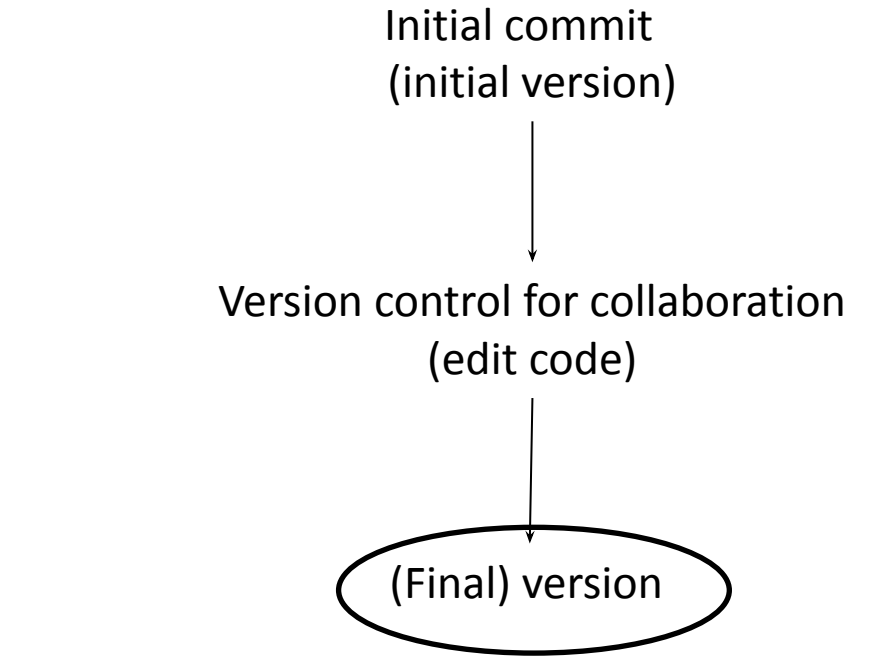
List of content

Git track changes between local and remote repositories, and allows clones (if different users are working on the same file)



List of content

Version control summary



List of content

Option 3 – Types of output

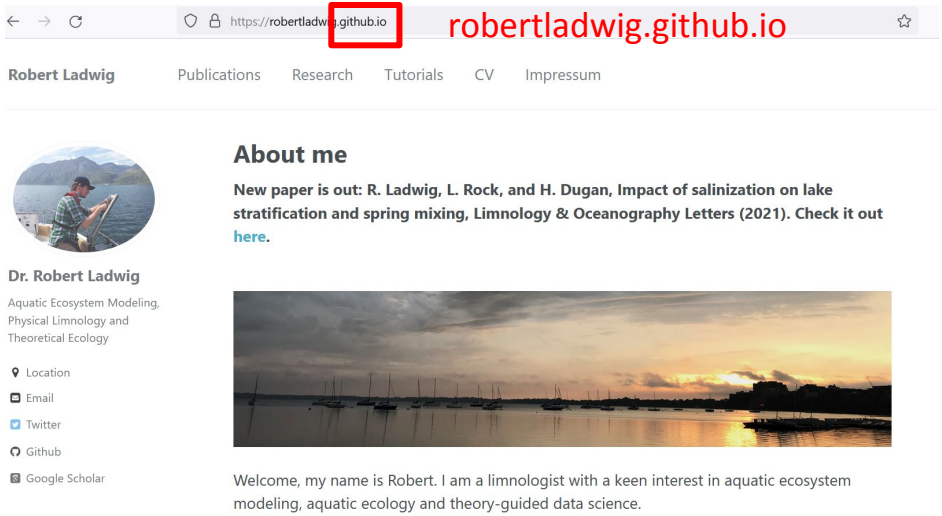
List of content

Types of output

- Local R folder (back-up, code-sharing, etc.)
- R-package (developing version of a CRAN R package, or R package existing only on GitHub). You will need the library devtools in R to access a repository.

```
install.packages("devtools")
devtools::install_github("rosalieb/serac", build_vignettes = TRUE)
library(serac)
```

- Website



List of content

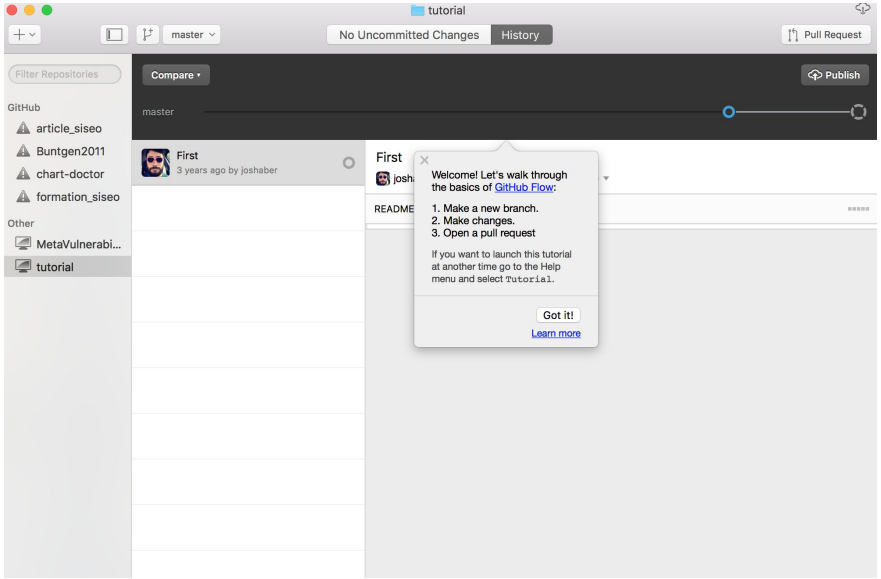
Other resources

Desktop versions

mac.github.com

dekstop.github.com

Linux version still in development?



FYI – Possible to write from Terminal on your dekstop, and then a few command:

cd = change directory
cd .. = up in directory
cd Dekstop/work

ls = tells you which document are in the repository

git clone (and paste the repository from github online).

git status= gives you which file are present on your computer

git add= if you want to commit your script.

git commit

git push
git pull

esc:wq

List of content

Why ECR may be interested in this tool?
(Why you should encourage your students to use Git.)

- **Alternative to academia: data scientist jobs.** Companies will expect you to know how to use version control.


NATUREJOBS | NATUREJOBS BLOG

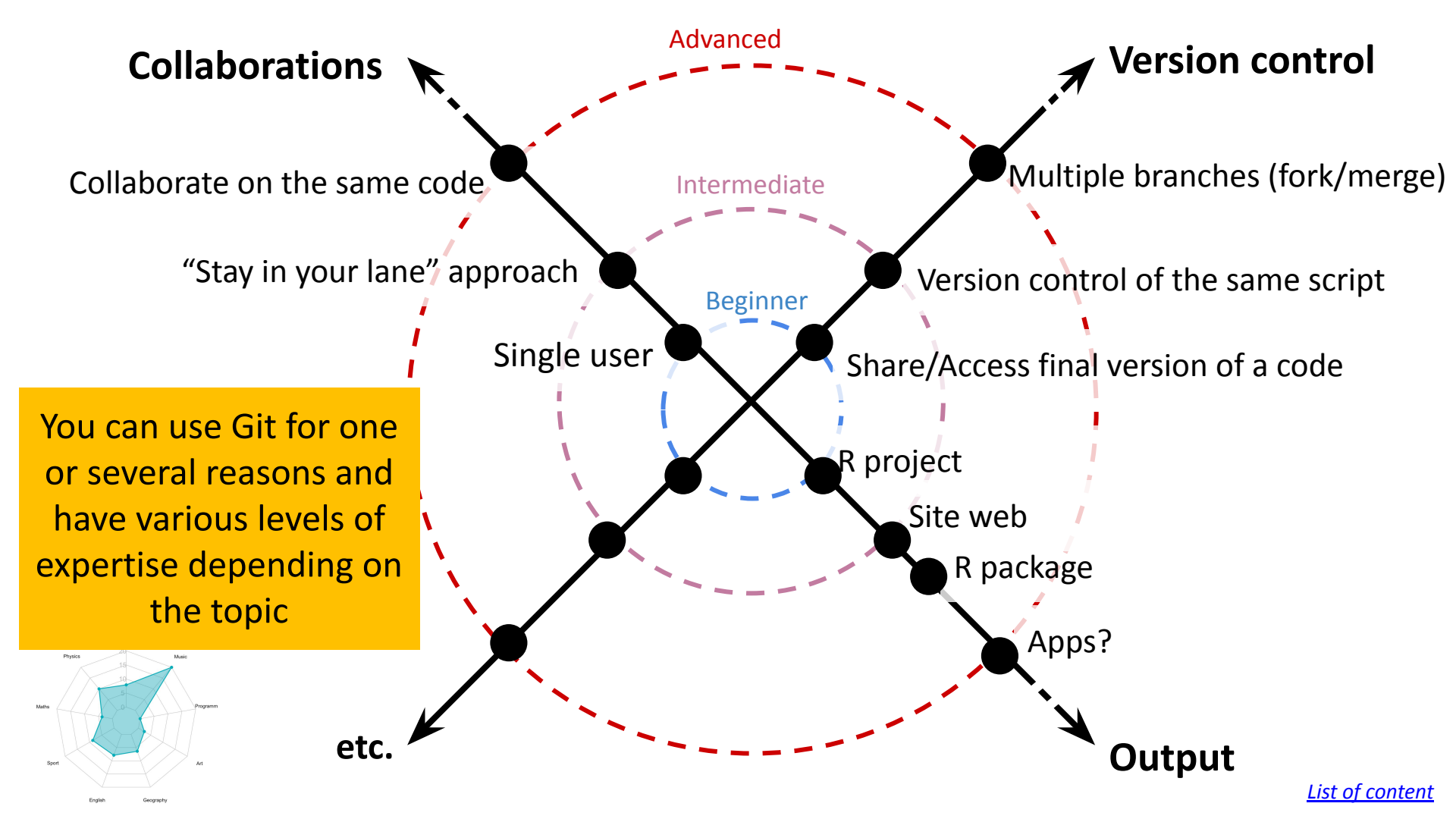
From Doctorate to Data Science: A very short guide

21 Jul 2017 | 13:00 GMT | Posted by Jack Leeming | Category: Academia, Blog, Career paths, Collaboration, Communication, Data, Industry, Mobility, PhD, Postgraduate, Research, Social media, Technology

- **Open science**: sharing codes

- **Ageless reason:** auditability and reproducibility!

List of content



You can use Git for one or several reasons and have various levels of expertise depending on the topic

List of content

Summary:

- Version control is your friend.
- Sharing is caring.
- Depending on your use, you may be better off with systems like ‘track changes’ in Word, or Dropbox...
- ...but for codes, specific systems exist.
- The ball is in your court if you ever need to collaborate on codes!

List of content

Links to specific sections of the presentation

Additional topics:

List of content

Questions?

Since 2019:
GitHub Education offers GitHub Pro for free to students. Just go here and use: https://education.github.com/pack#offers
For faculty and staff, you can try this link but you have to justify in which context you'll be using GitHub with students: https://education.github.com/teachers

Sources:

- Benjamin Wilcox https://www.slideshare.net/BenjaminWilcox/version-control-is-your-friend
- YouTube channel GitHub, the coding train
- Check out: https://guides.github.com/activities/hello-world/
- Wilson *et al.* (2017) Good enough practices in scientific computing, *Plos Computational Biology*

Rosalie Bruel
rosaliebruel@gmail.com
@RosalieBruel
rosalieb

List of content

Component of a “good” repository

- readme file (.md) *(see following slide for example)*
  - Documentation: what is inside the repository, how to cite, prefered ways of reporting an issue, etc.
- License
  - If your repository is public, anyone using your code is free to do so. You can choose among licenses templates to limit the terms of use.
- gitignore file (.txt) *(see following slide for example)*
  - By definition, Git will keep in memory ALL of what you uploaded and whatnot. For example, if every day you upload your progress and your current .Rdata file, it will use a big amount of storage that you don’t really need to be using (if you save the code, you can recreate the objects later).
  - The gitignore list extensions or folders to not commit (e.g., png, jpeg, pdf, csv, xlsx, Rdata, etc.)

List of content

README.md

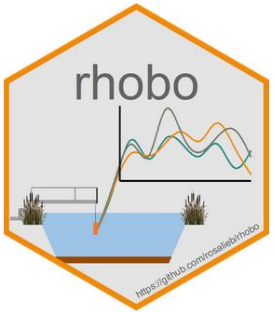
.md language → pretty!

Edit file | Preview

```
1
2  <!-- README.md is generated from README.Rmd. Please edit that file -->
3  <!-- badges: start -->
4  <!-- badges: end -->
5
6  # rhobo <a href="https://github.com/rosalieb/rhobo" target="_blank"><img src="man/figures/hex-
   rhobo.png" align="right" height="220" width="190" ></a>
7
8  The goal of rhobo is to read in the raw HOBO files downloaded through
9  the HOBO software and compute the dissolved oxygen correction factors.
10 Functions are also included to help append new data to previous data
11 files.
12
13 The steps match the routine for data from the artificial lakes at the
14 [CEREEP-Ecotron](https://www.cereep.bio.ens.psl.eu/spip.php?article45),
15 and may need to be adapted to other datasets format to be widely usable.
16
17 **Authors**: Rosalie Bruel (iEES-Paris, CNRS, France) and Sophie Guillon
18 (Mines ParisTech, France)
19
20 ## Installation
21
22 You can install the development version from
23 [GitHub](https://github.com/) with:
24
25     # install.packages("devtools")
26     devtools::install_github("rosalieb/rhobo")
27     library(rhobo)
28
```

Edit file | Preview

rhobo



The goal of rhobo is to read in the raw HOBO files downloaded through the HOBO software and compute the dissolved oxygen correction factors. Functions are also included to help append new data to previous data files.

The steps match the routine for data from the artificial lakes at the CEREEP-Ecotron, and may need to be adapted to other datasets format to be widely usable.

Authors: Rosalie Bruel (iEES-Paris, CNRS, France) and Sophie Guillon (Mines ParisTech, France)

Installation

You can install the development version from GitHub with:

```
# install.packages("devtools")
devtools::install_github("rosalieb/rhobo")
library(rhobo)
```

List of content

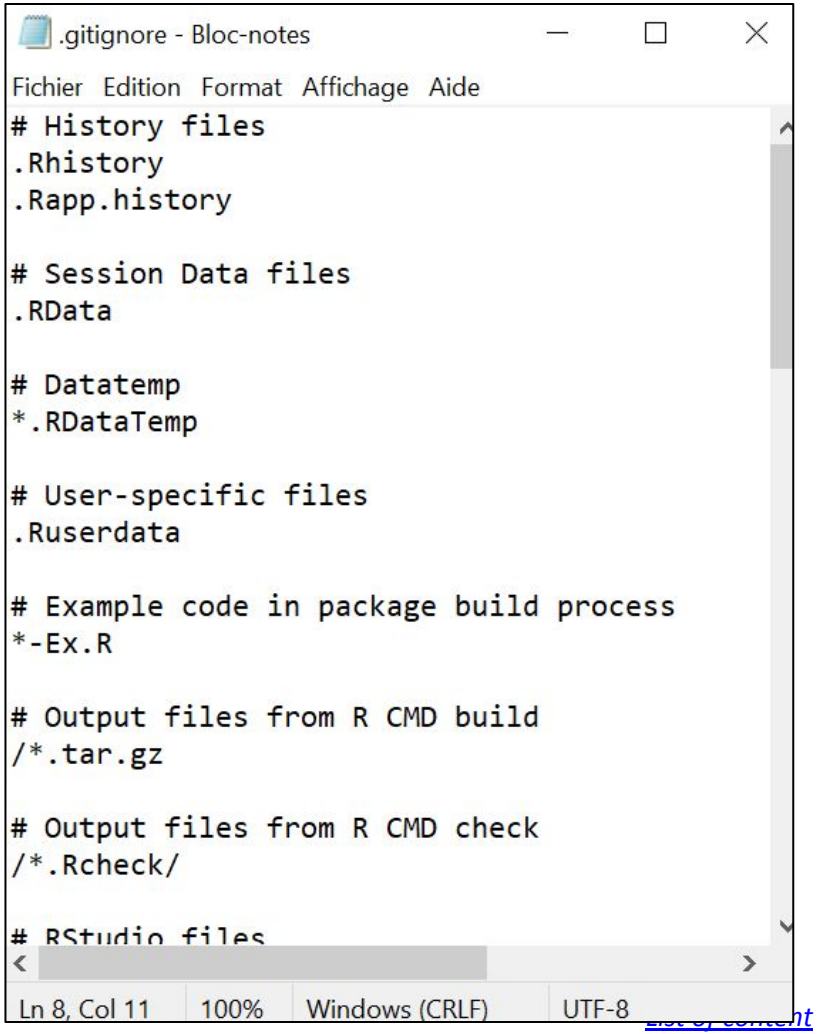
gitignore.txt

A text file, with the list of folders and/or extensions that Git need to omit (i.e., not push).

Can support complex syntax:

```
/Output/Figures/*
!/Output/Figures/toshare/
!Output/Figures/toshare/*
```

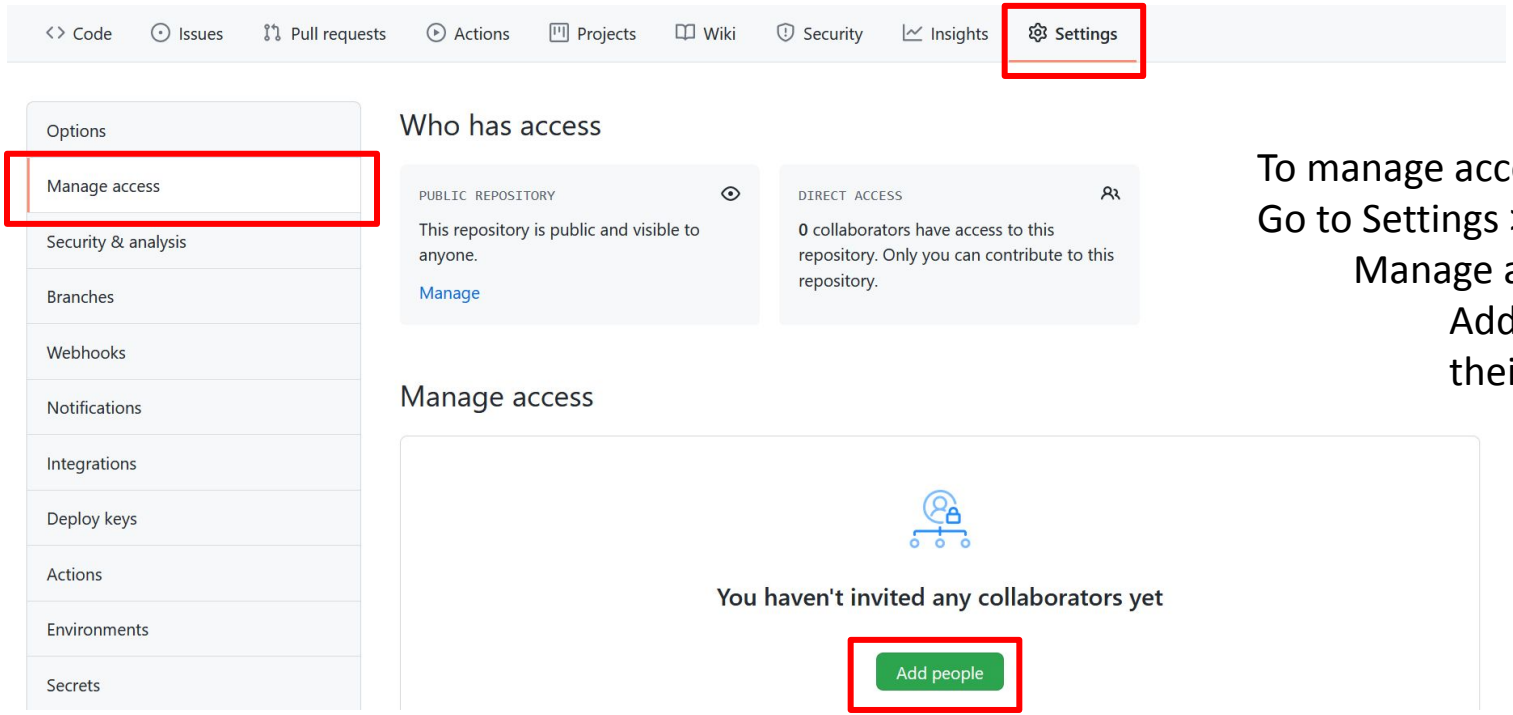
This tell us that all the Files in Output/Figures should NOT be pushed, except if they are in Output/Figures/toshare



List of content

Access to a repository

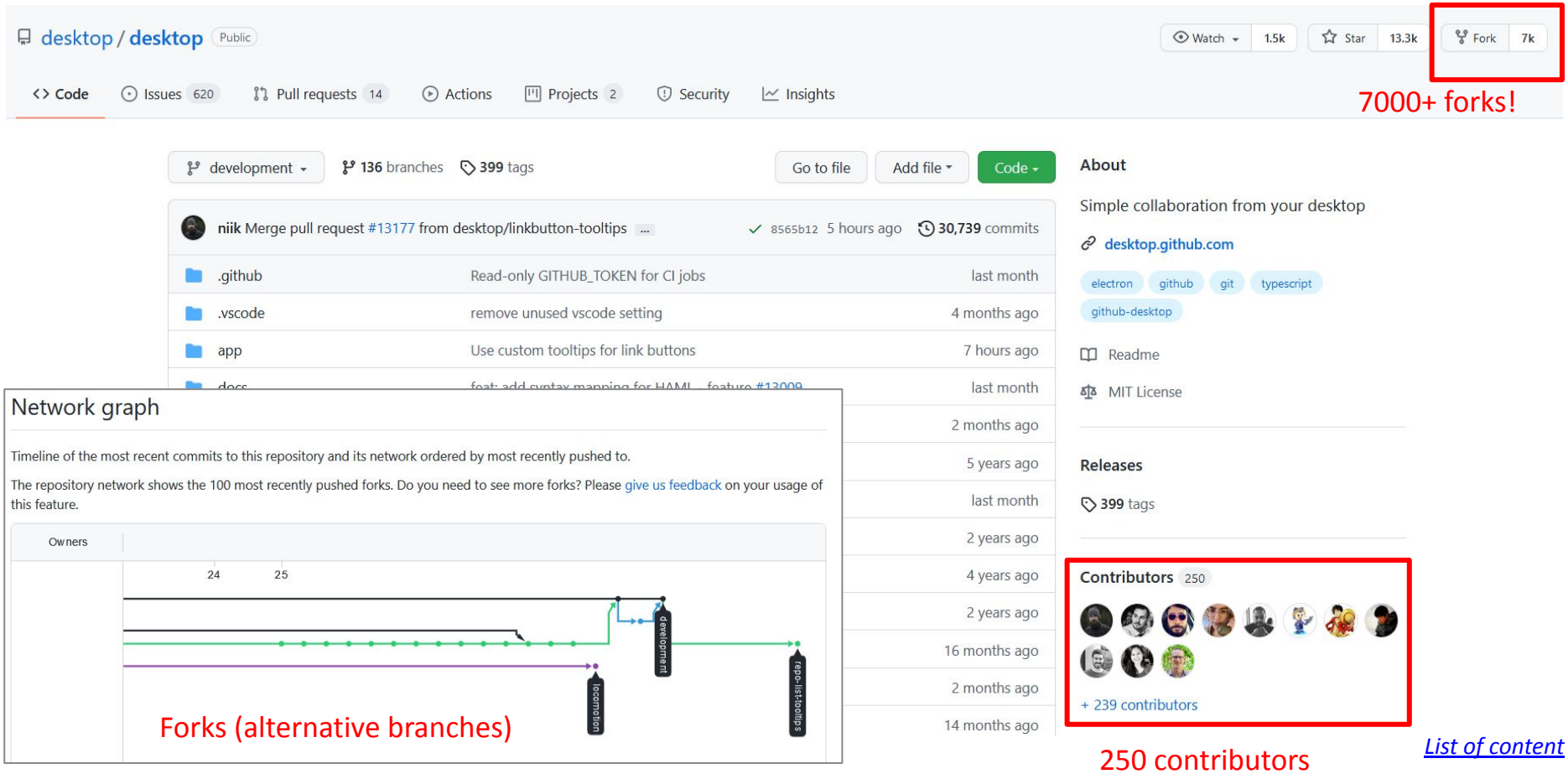
- A **public repository** can be cloned and forked by anyone.
- Only collaborators can commit onto a public repository.
- You have to be a collaborator to clone, fork, commit, etc. a **private repository**



To manage access:
Go to Settings >
Manage access >
Add people (by their username)

List of content

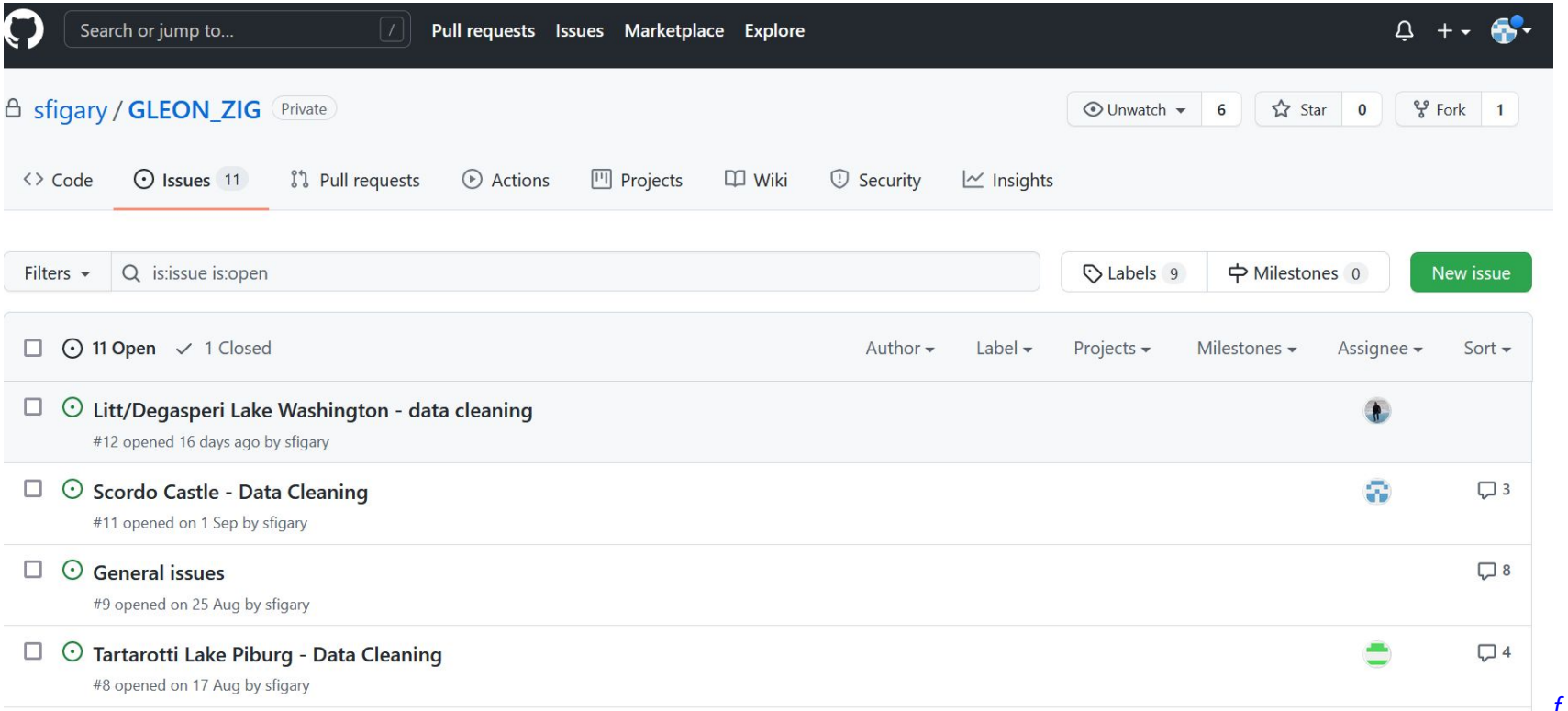
Limiting who can contribute to the master branch avoids breaking the code (otherwise you and I could just go and break up code). Someone must review your pull-request (merge-request).



List of content

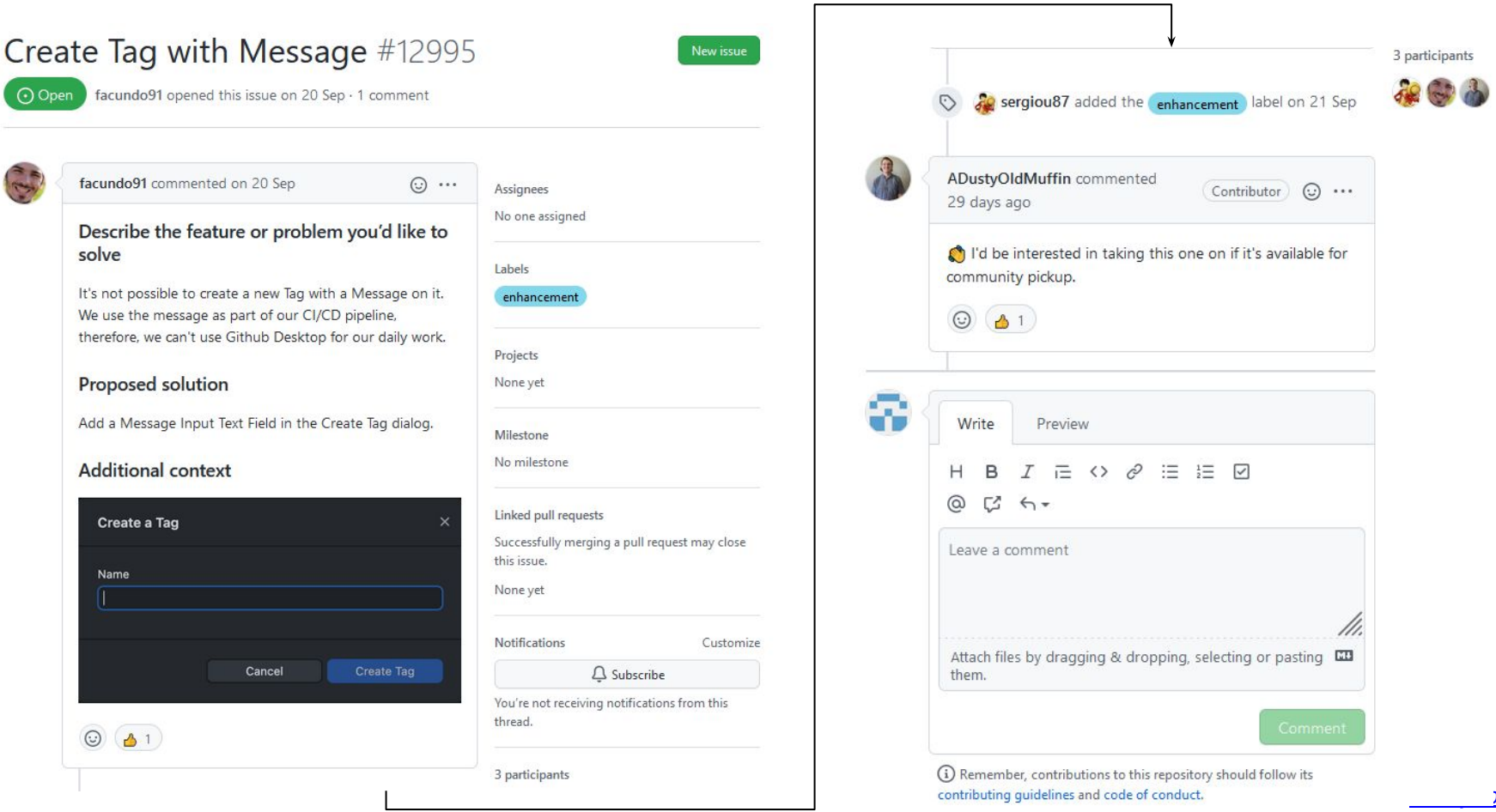
Issues (1/2)

- bring to the attention of the owner(s) of the repository somme issues with the development. Can have assignees, tags, images can be uploaded, etc.
- Can be useful to assign tasks



List of content

Issues (2/2) - example of an issue in github.com/desktop/desktop



ontent

Other resources:

Jenny Bryan's Happy Git with R: Chapter 1 introduces Git and its benefits

Git in R: Lecture notes from Dr. Nina Overgaard Therkildsen's Collaborative and Reproducible Data Science in R course at Cornell University. Lessons 2-5 provide step-by-step instructions for setting up Github and pairing it with Rstudio. These links are shared with permission from Dr. Nina Overgaard Therkildsen.

GLEON's GitHub workshop by Joe Stachelek

Build a R package and share it on GitHub https://sahirbhatnagar.com/rpkg/#building-an-r-package

Create your website in Rmarkdown and publish it via GitHub:
https://www.emilyzabor.com/tutorials/rmarkdown_websites_tutorial.html

Github project management:
https://openscapes.github.io/series/github-issues.html

List of content