Version control & RMarkdown

EcoDataSci-TLV

2018-11-27

✔ Check-in

To follow along, you should have:

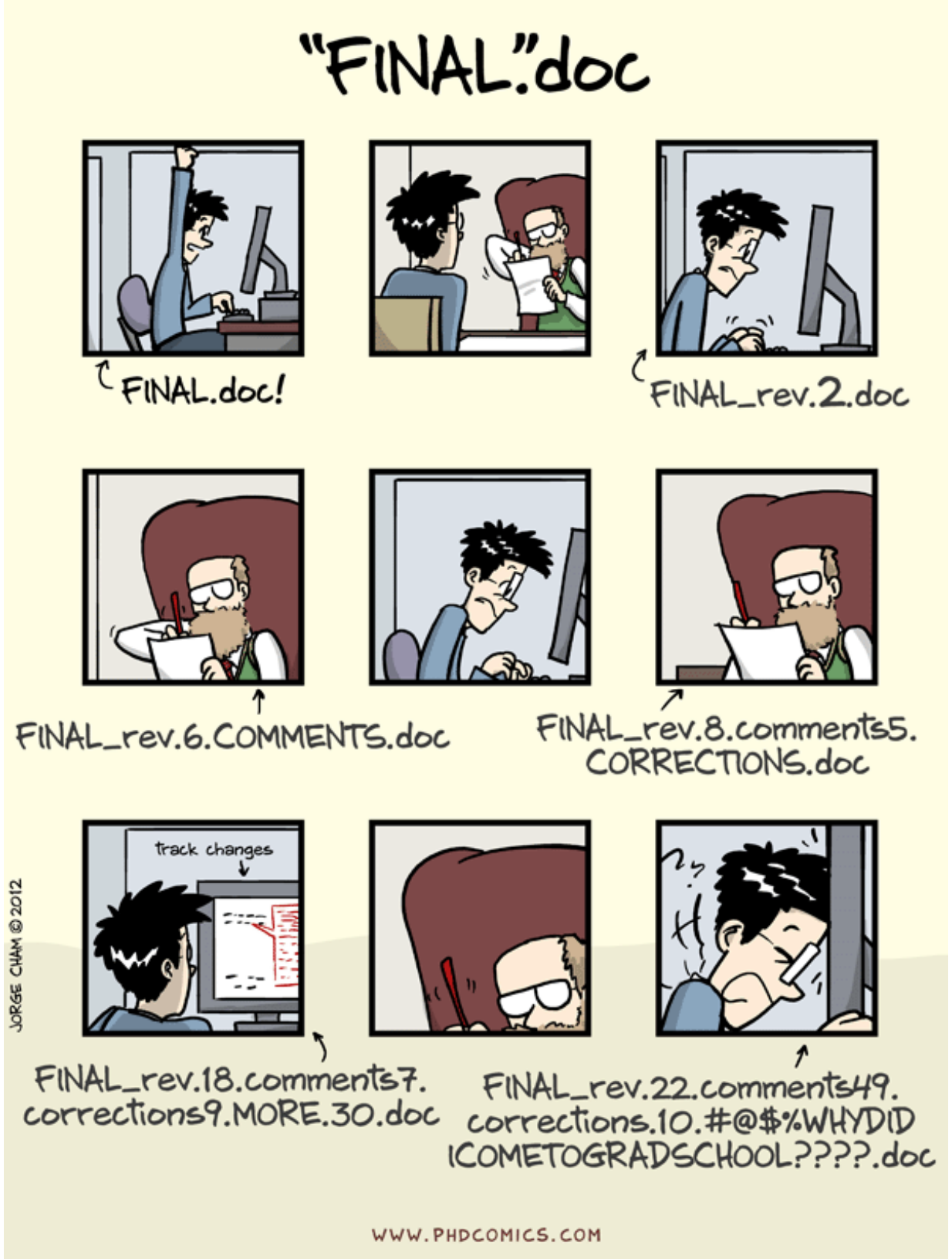
- a GitHub account - create one now if you don't have one already!
- git installed and connected to RStudio (detailed installation instructions here)

Ask questions and share resources on the EcoDataSci-TLV Slack

This presentation is based on:

http://ohi-science.org/data-science-training/github.html

& https://nceas.github.io/training-git-intro/getting-started-with-git-rstudio.html


"FINAL".doc
FINAL.doc!
FINAL_rev.2.doc
FINAL_rev.6.COMMENTS.doc
FINAL_rev.8.comments5.
CORRECTIONS.doc
track changes
FINAL_rev.18.comments7.
corrections9.MORE.30.doc
FINAL_rev.22.comments49.
corrections.10.#@$%WHYDID
ICOMETOGRADSCHOOL????.doc
JORGE CHAM © 2012
WWW.PHDCOMICS.COM

Why learn and use git/GitHub?

- Version control
- Easy to share/distribute files, especially code/analyses
- Files accessible from anywhere with an internet connection
- Improves collaboration

Why learn and use git/GitHub?

- Version control
- Easy to share/distribute files, especially code/analyses
- Files accessible from anywhere with an internet connection
- Improves collaboration

Git enables time travel & alternate realities!

git

Version control system that lives on your computer. Think "track changes" for files

git

Version control system that lives on your computer. Think "track changes" for files

git is not

- a full backup
- meant for data, images, etc

git works best with *text-based* files!

GitHub

"Dropbox" for git-based projects on the internet.

- share/store analyses and functions
- browse past versions of code
- browse source code (CRAN, tidyverse, rOpenSci)
- use packages not on CRAN (`devtools::install_github("account/repo")`)
- host web pages

GitHub

"Dropbox" for git-based projects on the internet.

- share/store analyses and functions
- browse past versions of code
- browse source code (CRAN, tidyverse, rOpenSci)
- use packages not on CRAN (`devtools::install_github("account/repo")`)
- host web pages

You get unlimited private GitHub repositories for *free* (normally $7/month) while you are a student! Sign up here.

OHI-Science / ohiprep

Unwatch 20 Star 4 Fork 2

Code Issues 3 Pull requests 0 Boards Burndown Wiki Pulse Graphs Settings

debugging the species for 2016... I think it's actually OK...

master

Browse files

= committed a day ago 1 parent 7a8bc55 commit b21fe2946589b4b2ed7351bc68c154651aab7953

Showing 88 changed files with 2,432,873 additions and 426,265 deletions.

Unified Split

Sorry, we could not display the entire diff because it was too big.

77 globalprep/spp_ico/v2016/data_prep_SPP.Rmd

View

Code in red was deleted

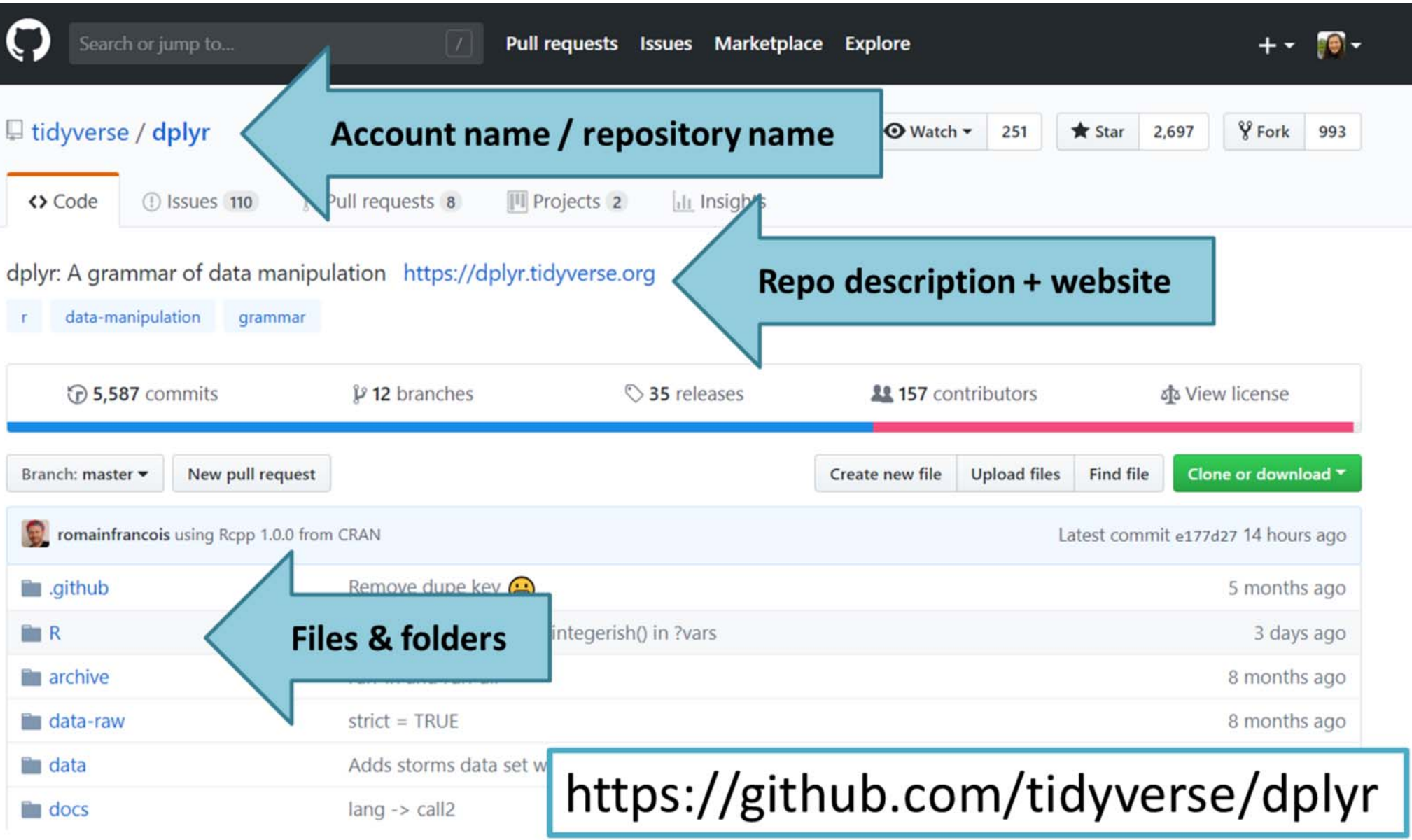
Code in green was added

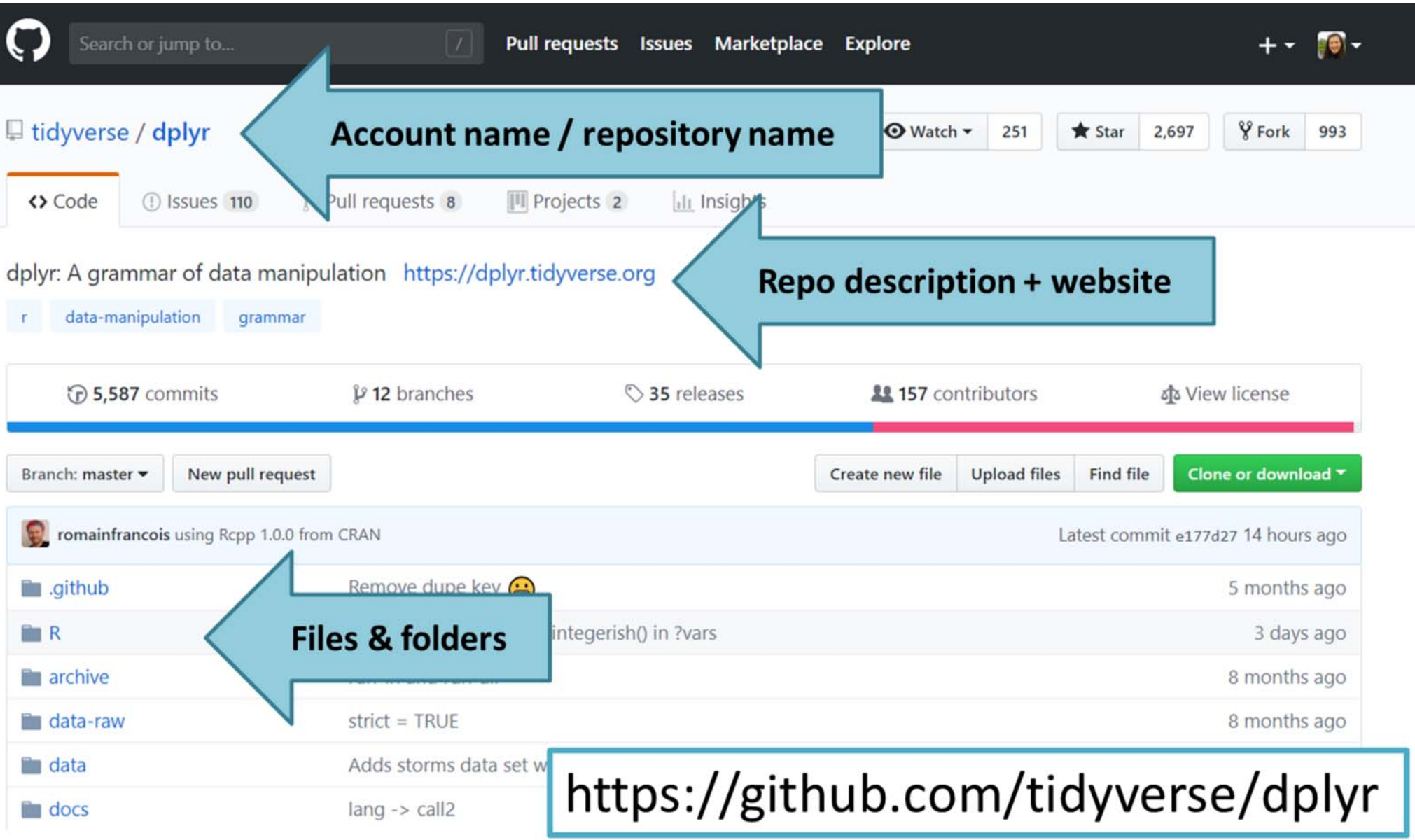
```
@@ -22,18 +22,18 @@ library(foreign)
 library(data.table)
 library(sp)
 library(rgdal)
-library(raster)
-library(maptools)
+# library(raster)
+# library(maptools)
 library(readr)

 source('~/github/ohiprep/src/R/common.R')

 goal     <- 'globalprep/spp_ico'
 scenario <- 'v2016'
 dir_anx  <- file.path(dir_M, 'git-annex', goal)
-dir_data_am    <- file.path(dir_M, 'git-annex/globalprep/_raw_data', 'aquamaps', str_replace(scenario, 'v', 'd'))
-dir_data_iucn  <- file.path(dir_M, 'git-annex/globalprep/_raw_data', 'iucn_spp')
-dir_data_bird  <- file.path(dir_M, 'git-annex/globalprep/_raw_data', 'birdlife_intl')
+dir_data_am    <- file.path(dir_M, 'git-annex/globalprep/_raw_data', 'aquamaps', 'd2015')
+dir_data_iucn  <- file.path(dir_M, 'git-annex/globalprep/_raw_data', 'iucn_spp', 'd2015')
+dir_data_bird  <- file.path(dir_M, 'git-annex/globalprep/_raw_data', 'birdlife_intl', 'd2015')
 dir_git  <- file.path('~/github/ohiprep', goal)

 source(file.path(dir_git, 'spp_fxn.R'))
```

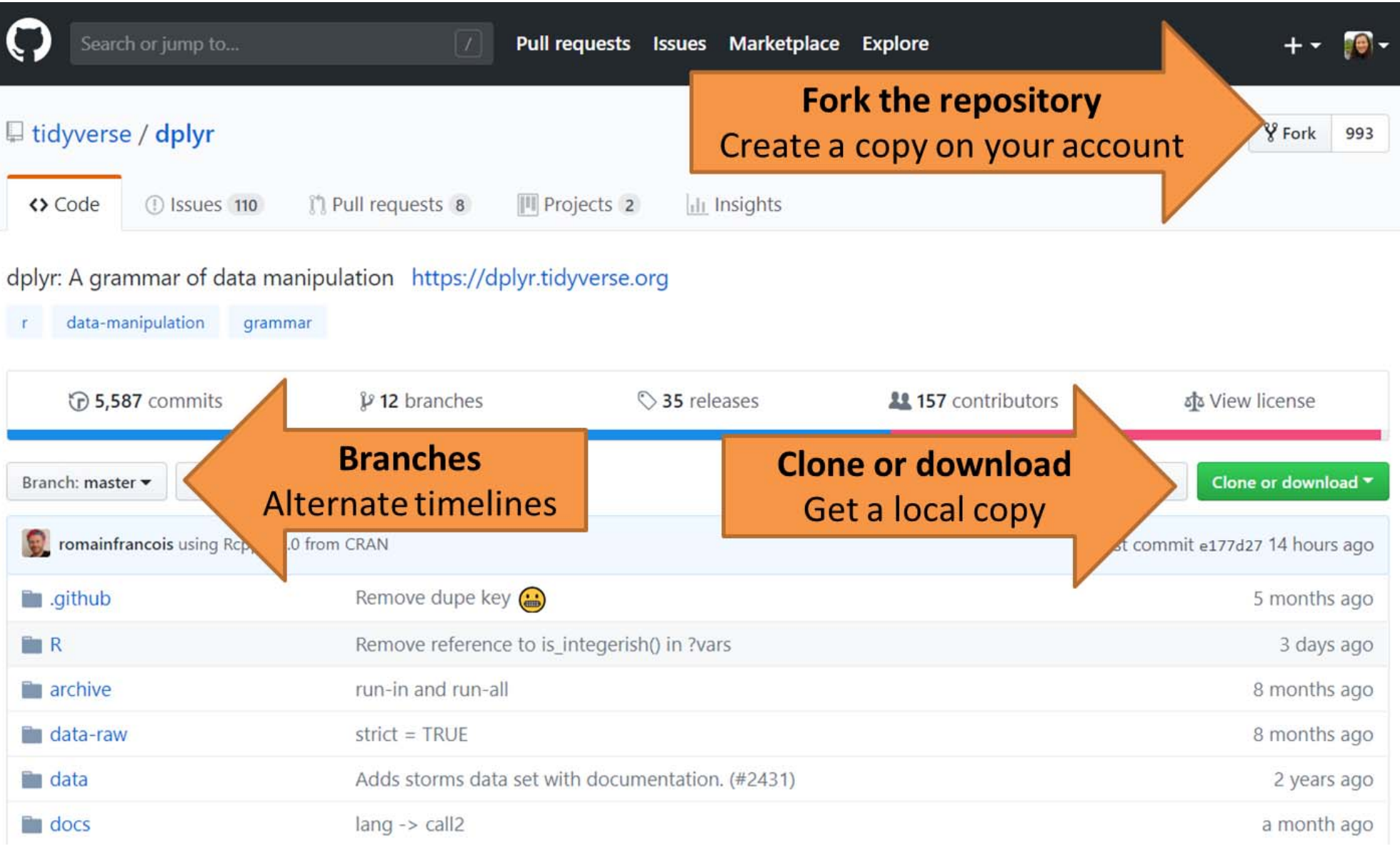
GitHub terminology

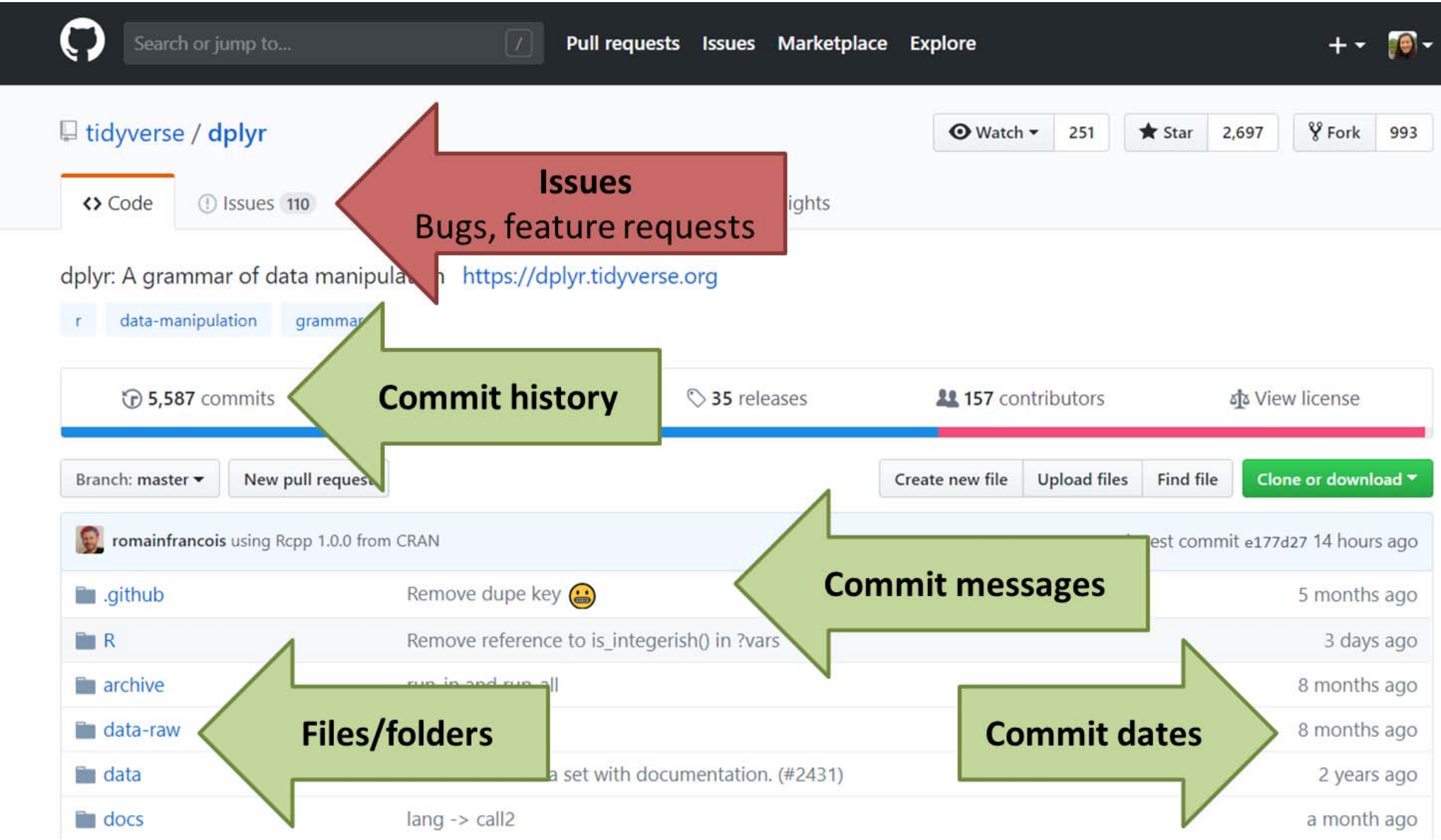
- **repository** or repo - project/folder
- **local** - files stored on your computer
- **remote** - files on github.com

GitHub: let's get oriented!


Search or jump to...
Pull requests
Issues
Marketplace
Explore
tidyverse / dplyr
Account name / repository name
Watch 251
Star 2,697
Fork 993
Code
Issues 110
Pull requests 8
Projects 2
Insights
dplyr: A grammar of data manipulation https://dplyr.tidyverse.org
Repo description + website
r
data-manipulation
grammar
5,587 commits
12 branches
35 releases
157 contributors
View license
Branch: master
New pull request
Create new file
Upload files
Find file
Clone or download
romainfrancois using Rcpp 1.0.0 from CRAN
Latest commit e177d27 14 hours ago
.github
Remove dupe key
5 months ago
R
integerish() in ?vars
3 days ago
archive
8 months ago
data-raw
strict = TRUE
8 months ago
data
Adds storms data set w
docs
lang -> call2
Files & folders
https://github.com/tidyverse/dplyr


Search or jump to...
Pull requests
Issues
Marketplace
Explore
tidyverse / dplyr
Account name / repository name
Watch 251
Star 2,697
Fork 993
Code
Issues 110
Pull requests 8
Projects 2
Insights
dplyr: A grammar of data manipulation https://dplyr.tidyverse.org
Repo description + website
r
data-manipulation
grammar
5,587 commits
12 branches
35 releases
157 contributors
View license
Branch: master
New pull request
Create new file
Upload files
Find file
Clone or download
romainfrancois using Rcpp 1.0.0 from CRAN
Latest commit e177d27 14 hours ago
.github
Remove dupe key
5 months ago
R
integerish() in ?vars
3 days ago
archive
8 months ago
data-raw
strict = TRUE
8 months ago
data
Adds storms data set w
docs
lang -> call2
Files & folders
https://github.com/tidyverse/dplyr

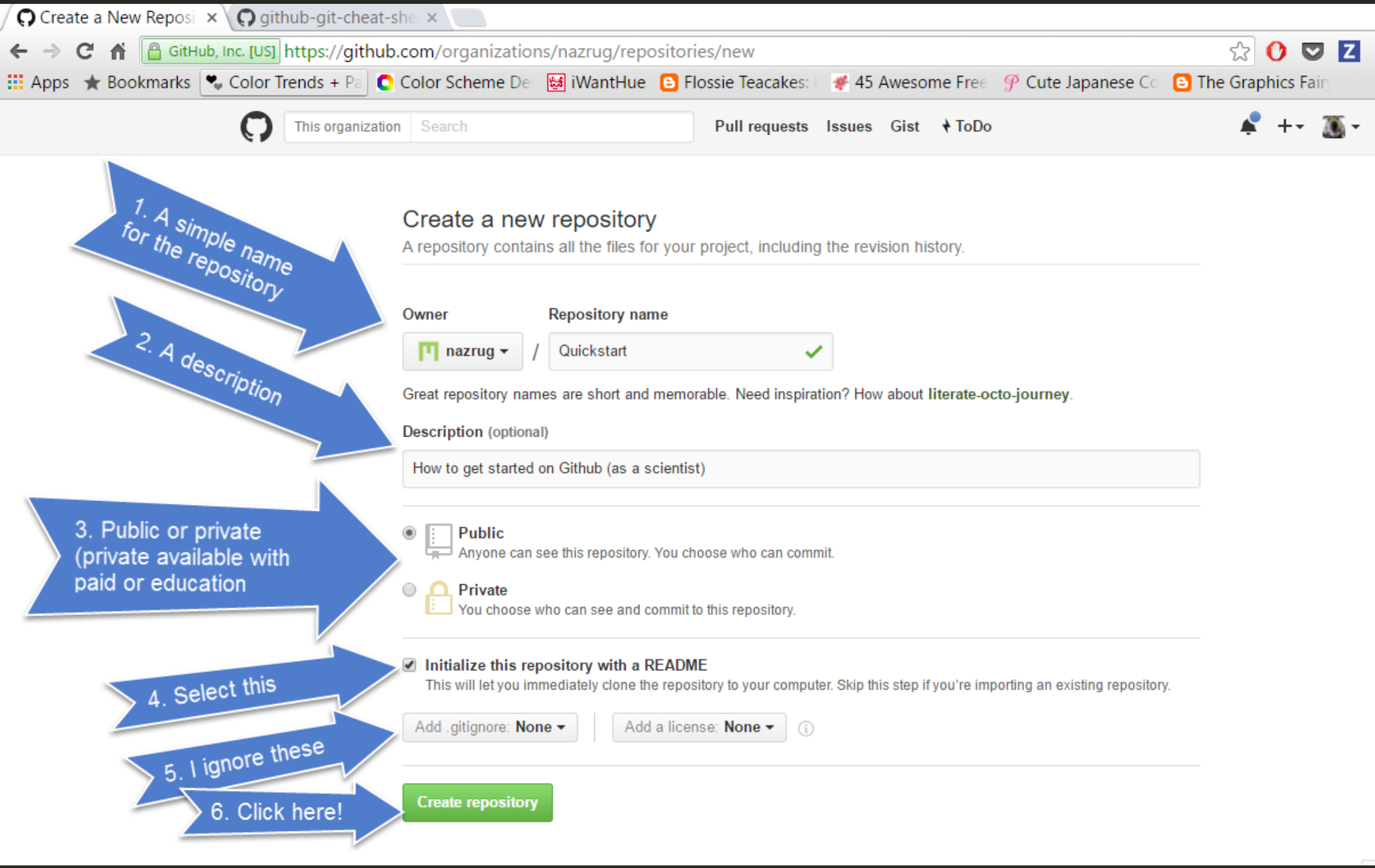

Search or jump to...
Pull requests
Issues
Marketplace
Explore
Fork the repository
Create a copy on your account
tidyverse / dplyr
Fork 993
Code
Issues 110
Pull requests 8
Projects 2
Insights
dplyr: A grammar of data manipulation https://dplyr.tidyverse.org
r
data-manipulation
grammar
5,587 commits
12 branches
35 releases
157 contributors
View license
Branch: master
Clone or download
Branches
Alternate timelines
Clone or download
Get a local copy
romainfrancois using Rcp .0 from CRAN
st commit e177d27 14 hours ago
.github Remove dupe key 5 months ago
R Remove reference to is_integerish() in ?vars 3 days ago
archive run-in and run-all 8 months ago
data-raw strict = TRUE 8 months ago
data Adds storms data set with documentation. (#2431) 2 years ago
docs lang -> call2 a month ago


Search or jump to...
Pull requests
Issues
Marketplace
Explore
tidyverse / dplyr
Watch 251
Star 2,697
Fork 993
Code
Issues 110
Issues
Bugs, feature requests
dplyr: A grammar of data manipula
https://dplyr.tidyverse.org
r
data-manipulation
5,587 commits
Commit history
35 releases
157 contributors
View license
Branch: master
Create new file
Upload files
Find file
Clone or download
romainfrancois using Rcpp 1.0.0 from CRAN
Commit messages
.github
Remove dupe key
5 months ago
R
Remove reference to is_integerish() in ?vars
3 days ago
archive
8 months ago
data-raw
Files/folders
8 months ago
data
a set with documentation. (#2431)
2 years ago
docs
lang -> call2
a month ago
Commit dates

Let's create a repo from scratch...

(choose a unique name!)

- don't use special characters
- name it `something_meaningful` and not `too-long`
- note for advanced users: R packages cannot take underscores or dashes


Create a New Reposi
github-git-cheat-she
GitHub, Inc. [US] https://github.com/organizations/nazrug/repositories/new
Apps
Bookmarks
Color Trends + Pa
Color Scheme De
iWantHue
Flossie Teacakes:
45 Awesome Free
Cute Japanese Co
The Graphics Fair
This organization
Search
Pull requests
Issues
Gist
ToDo
1. A simple name for the repository
2. A description
3. Public or private (private available with paid or education
4. Select this
5. I ignore these
6. Click here!
Create a new repository
A repository contains all the files for your project, including the revision history.
Owner
Repository name
nazrug
/
Quickstart
Great repository names are short and memorable. Need inspiration? How about literate-octo-journey.
Description (optional)
How to get started on Github (as a scientist)
Public
Anyone can see this repository. You choose who can commit.
Private
You choose who can see and commit to this repository.
Initialize this repository with a README
This will let you immediately clone the repository to your computer. Skip this step if you're importing an existing repository.
Add .gitignore: None
Add a license: None
Create repository

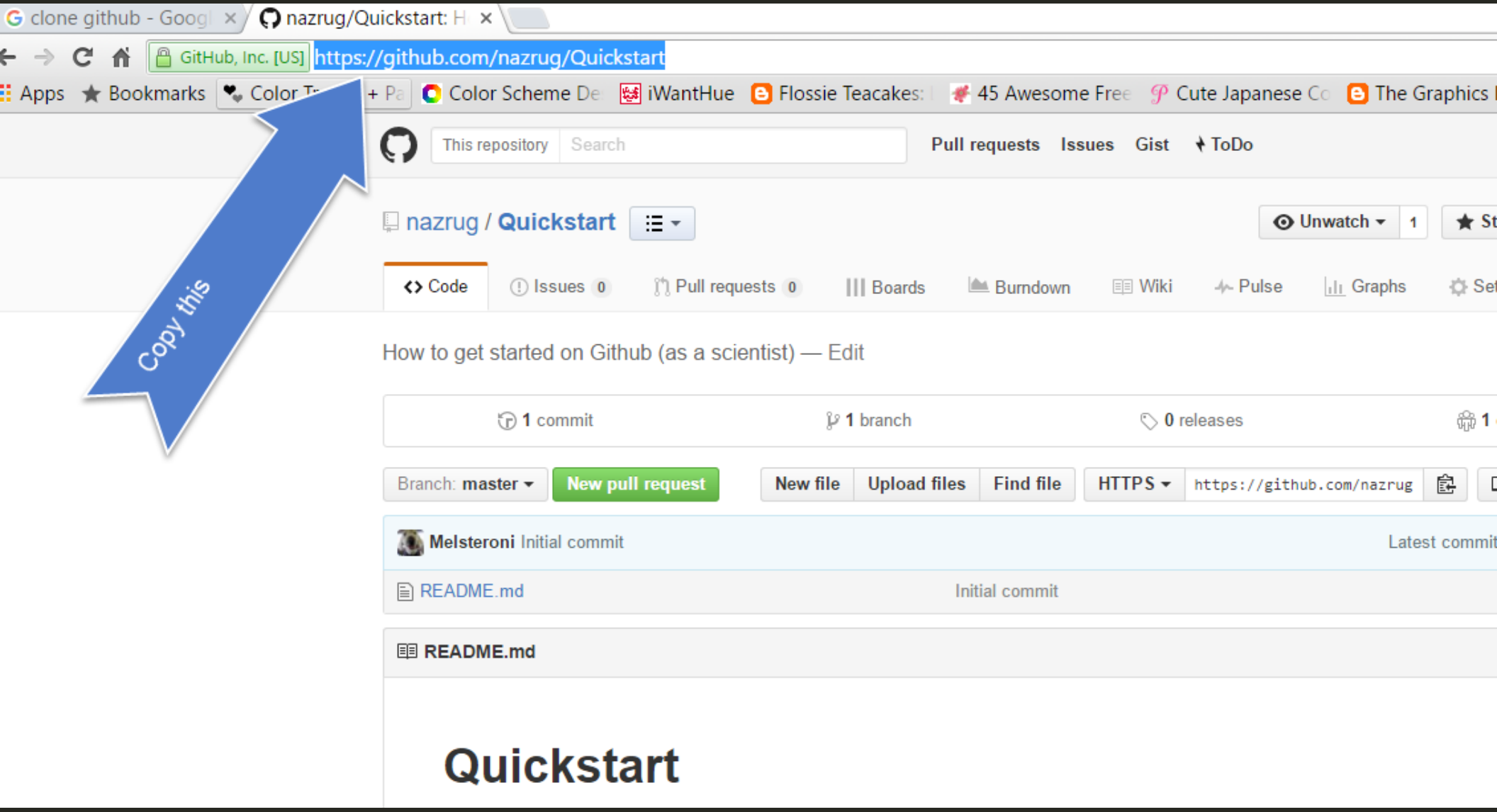
Alternative workflow

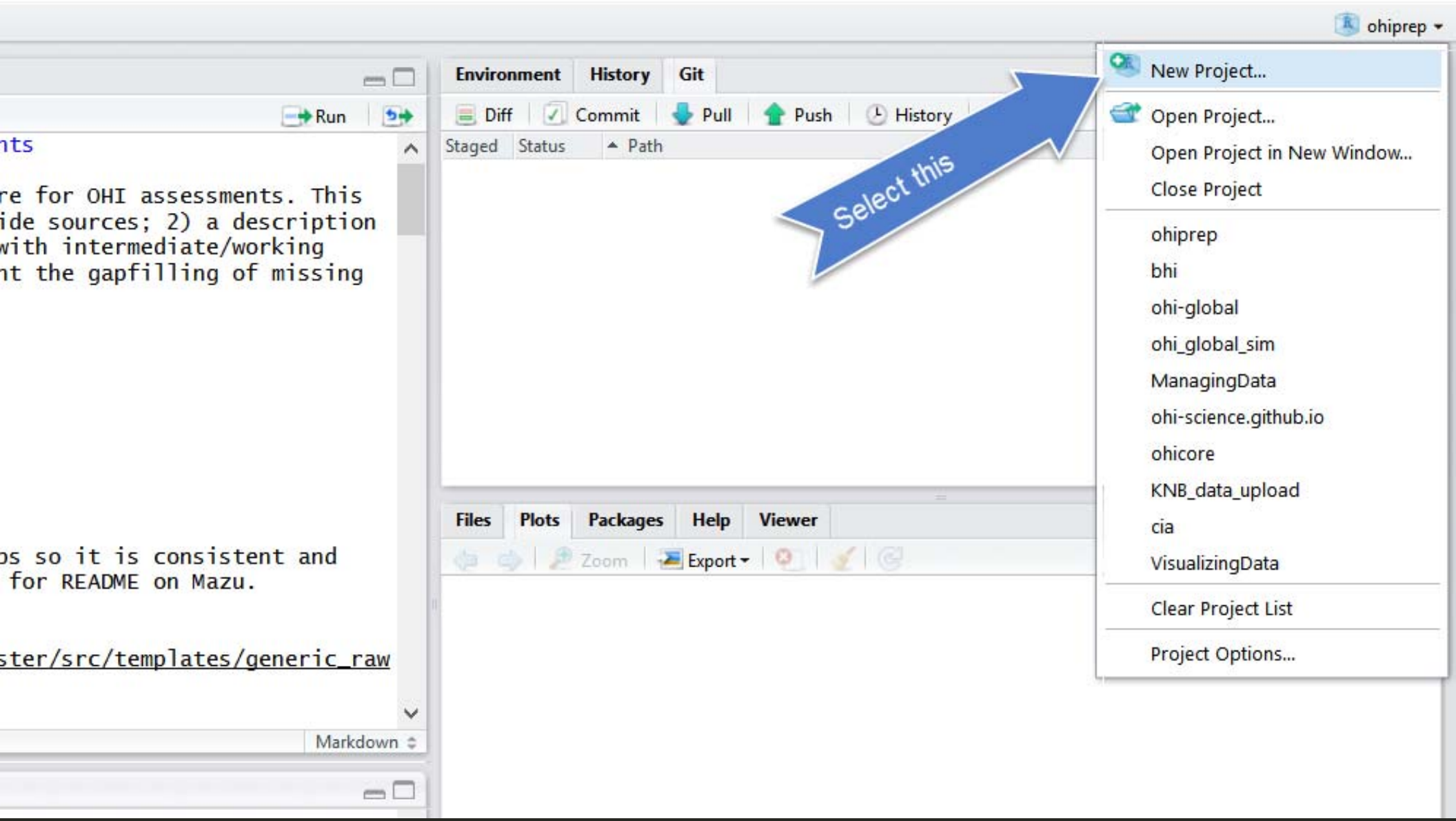
If you want to work off of an existing repository on GitHub...

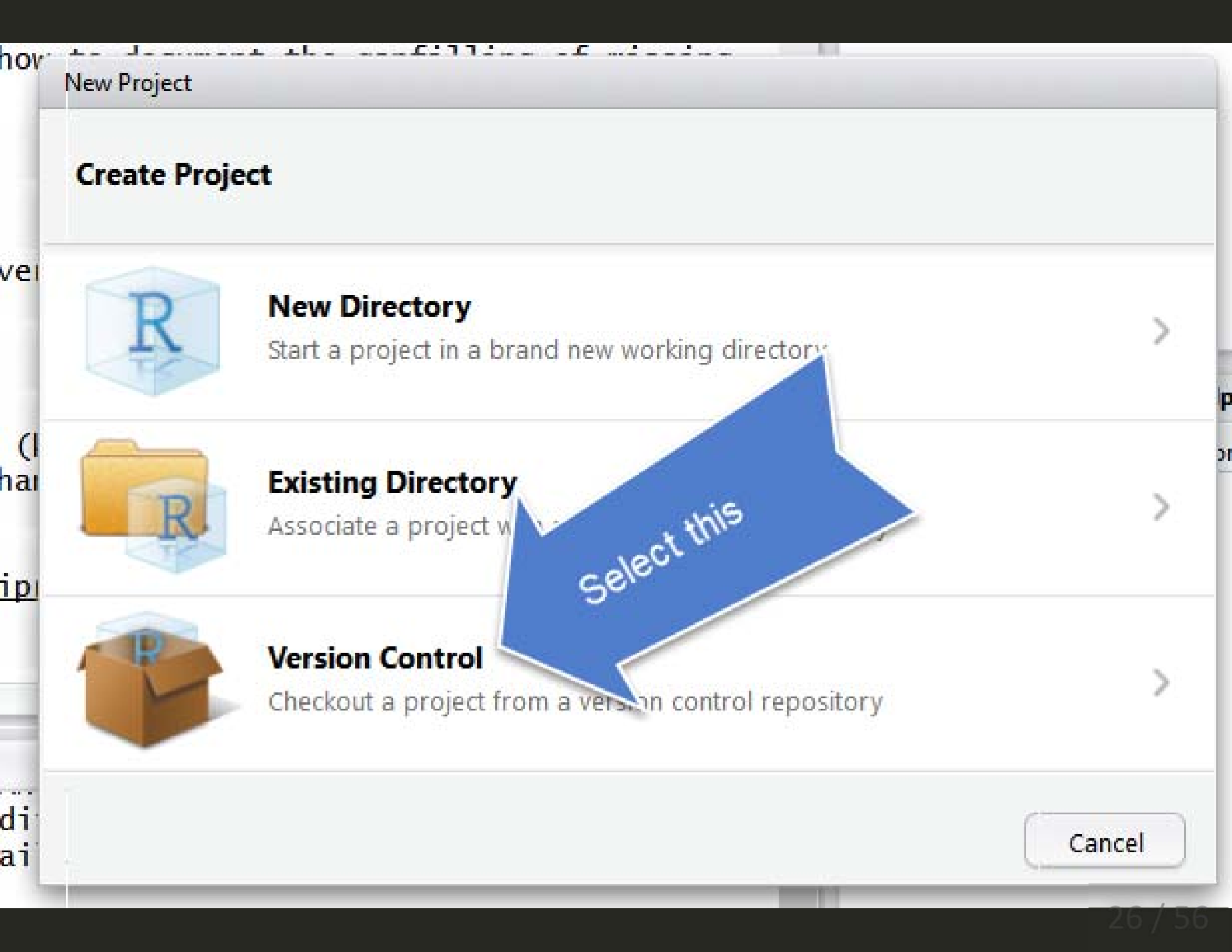
1. **Fork a repo** = make a copy of the repo on your account
2. **Make a branch** = start an alternate timeline
3. **Make a change**
4. **Examine the differences**

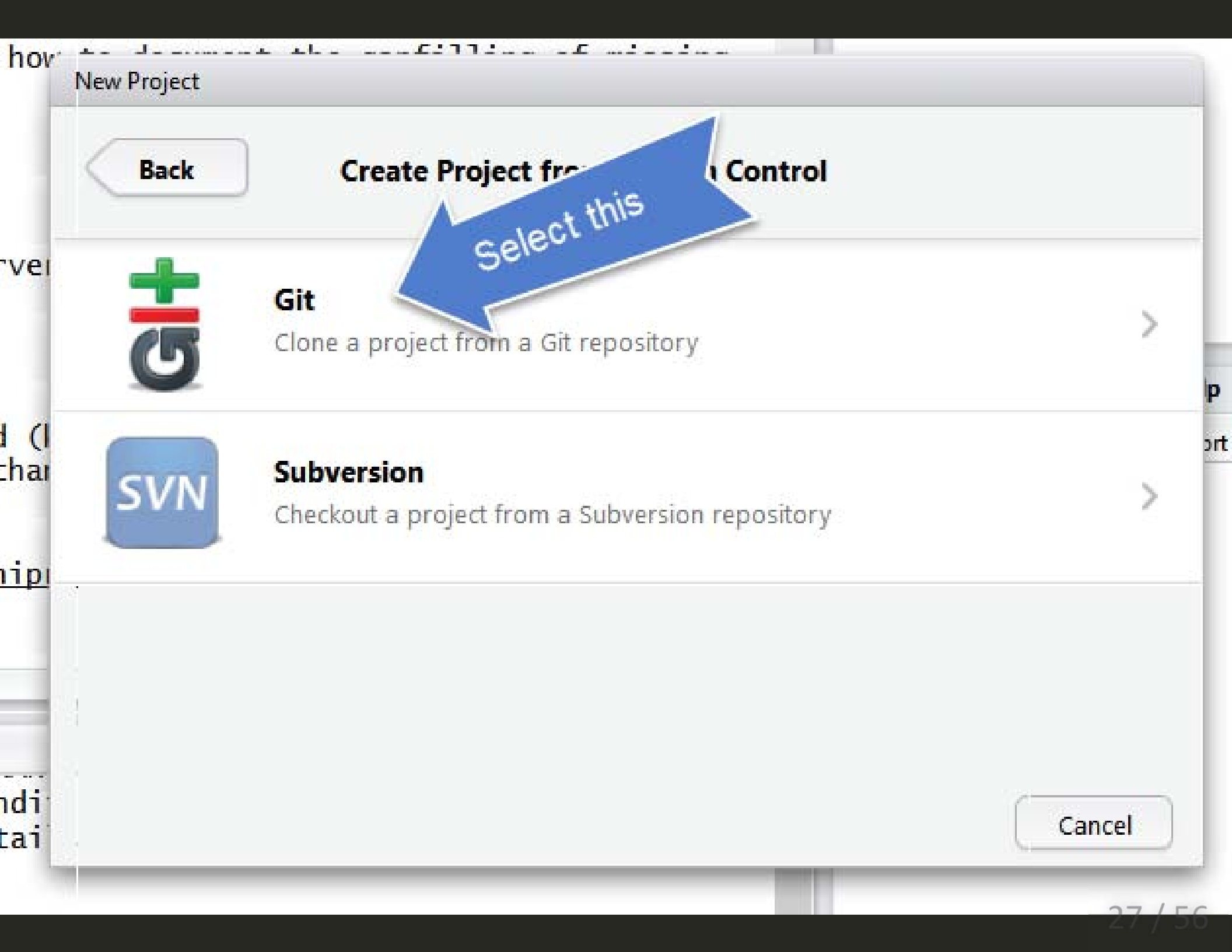
It's good practice to keep your "master" branch clean so that you can keep it synchronized with the original repository

Let's connect to RStudio!


clone github - Googl
nazrug/Quickstart: H
GitHub, Inc. [US]
https://github.com/nazrug/Quickstart
Apps
Bookmarks
Color Scheme De
iWantHue
Flossie Teacakes:
45 Awesome Free
Cute Japanese Co
The Graphics
Copy this
This repository
Search
Pull requests
Issues
Gist
ToDo
nazrug / Quickstart
Unwatch
1
Code
Issues 0
Pull requests 0
Boards
Burndown
Wiki
Pulse
Graphs
How to get started on Github (as a scientist) — Edit
1 commit
1 branch
0 releases
Branch: master
New pull request
New file
Upload files
Find file
HTTPS
https://github.com/nazrug
Melsteroni Initial commit
README.md
Initial commit
README.md
Quickstart


ohiprep
Environment
History
Git
Diff
Commit
Pull
Push
History
Staged
Status
Path
Run
nts
re for OHI assessments. This
ide sources; 2) a description
with intermediate/working
nt the gapfilling of missing
os so it is consistent and
for README on Mazu.
ster/src/templates/generic_raw
Markdown
Files
Plots
Packages
Help
Viewer
Zoom
Export
Select this
New Project...
Open Project...
Open Project in New Window...
Close Project
ohiprep
bhi
ohi-global
ohi_global_sim
ManagingData
ohi-science.github.io
ohicore
KNB_data_upload
cia
VisualizingData
Clear Project List
Project Options...


New Project
Create Project
New Directory
Start a project in a brand new working directory
Existing Directory
Associate a project w
Select this
Version Control
Checkout a project from a version control repository
Cancel


New Project
Back
Create Project fr Control
Select this
Git
Clone a project from a Git repository
SVN
Subversion
Checkout a project from a Subversion repository
Cancel

TAR
WARS

New Project

Back

Clone Git Repository

Repository URL:

https://github.com/nazrug/Quickstart

Project directory name:

Quickstart

Create project as subdirectory of:

C:/Users/Melanie/Github

Browse...

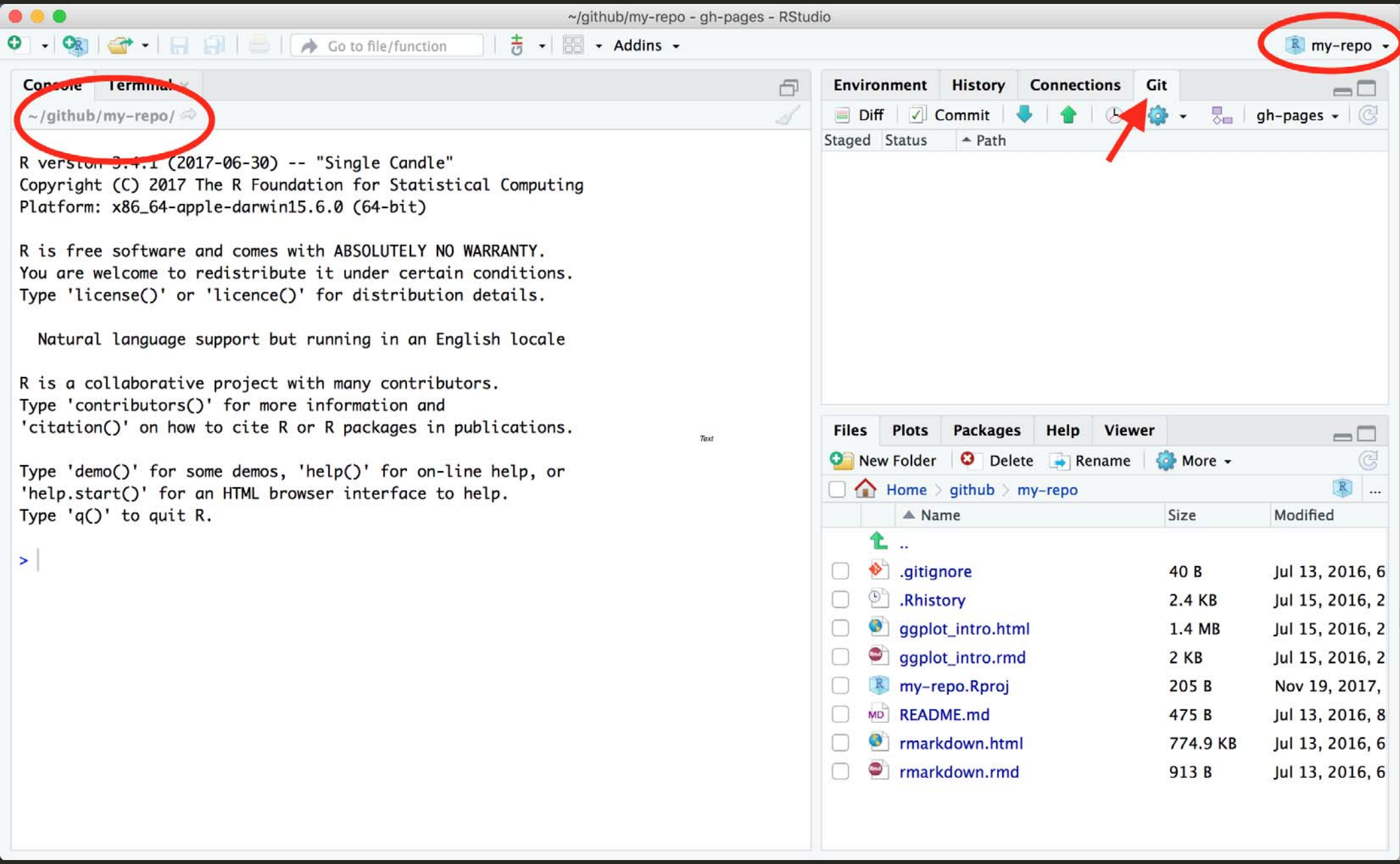
Paste repository name here (the rest should auto-fill). Then select "Create Project" below.

Open in new window

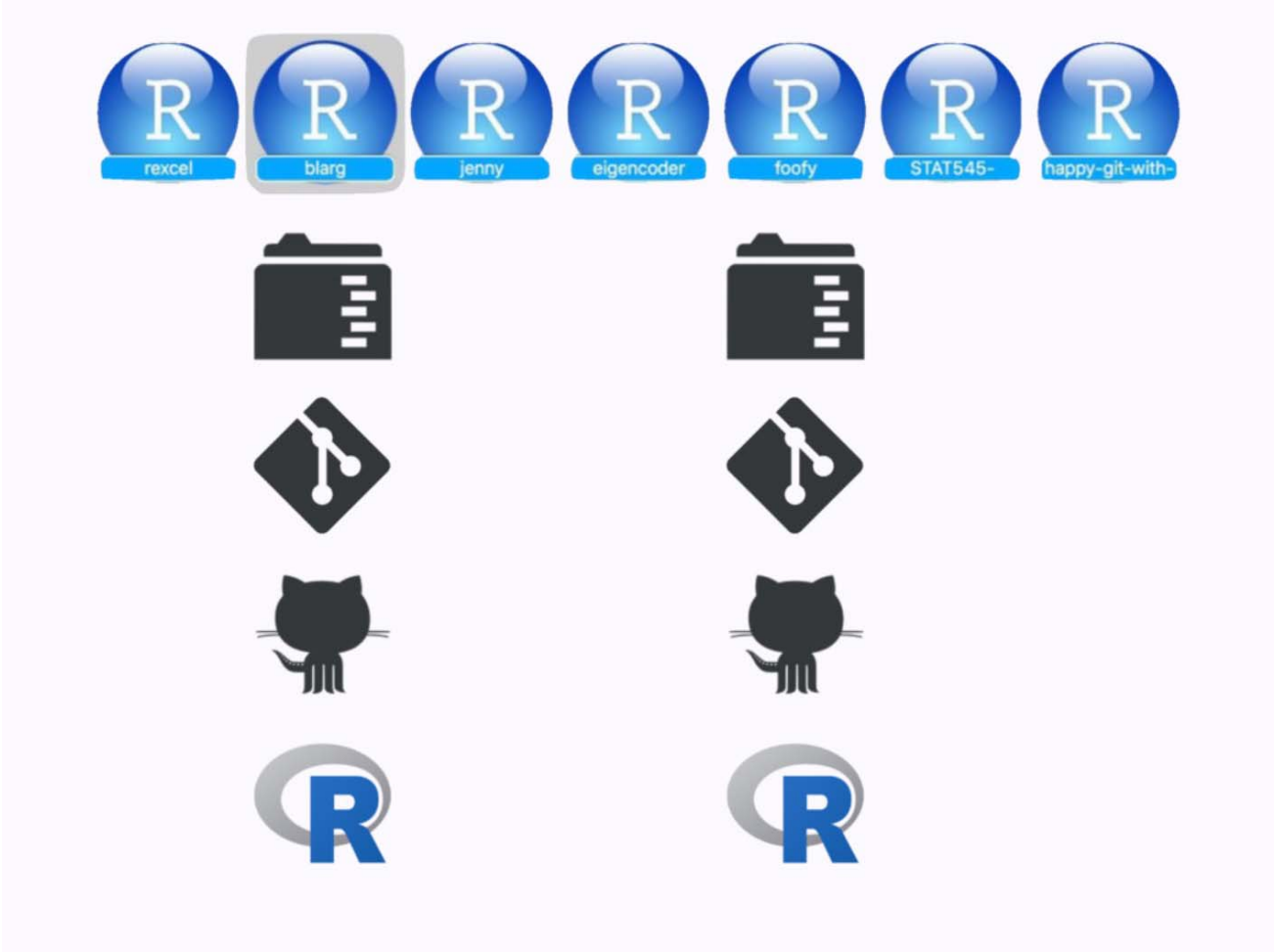
Create Project

Ca

Inspect your repository


~/github/my-repo - gh-pages - RStudio
Go to file/function
Addins
my-repo
Console
Terminal
~/github/my-repo/
R version 3.4.1 (2017-06-30) -- "Single Candle"
Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
>
Environment
History
Connections
Git
Diff
Commit
gh-pages
Staged
Status
Path
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
Home
github
my-repo
Name
Size
Modified
..
.gitignore
40 B
Jul 13, 2016, 6
.Rhistory
2.4 KB
Jul 15, 2016, 2
ggplot_intro.html
1.4 MB
Jul 15, 2016, 2
ggplot_intro.rmd
2 KB
Jul 15, 2016, 2
my-repo.Rproj
205 B
Nov 19, 2017,
README.md
475 B
Jul 13, 2016, 8
rmarkdown.html
774.9 KB
Jul 13, 2016, 6
rmarkdown.rmd
913 B
Jul 13, 2016, 6

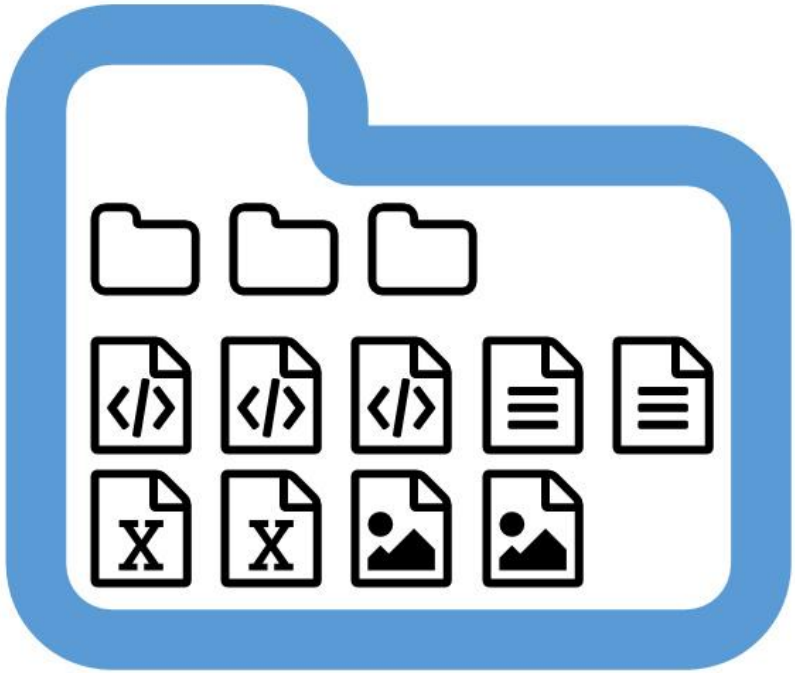
You now have a GitHub PROJECT!



https://speakerdeck.com/jennybc/workflow-you-should-have-one

What's the deal with projects?

Each project is a self-contained set of files. Projects make it **easy to transfer** files to another computer because its **boundaries are clearly defined**



Use relative paths inside projects

- looks cleaner
- works for other people/computers

```
read.csv("data/eilat_survey_2017.csv")
#rather than
read.csv("C:/חזי/phd/thesis/reef_surveys/data/eilat_surv
```

You never have to use `setwd()` again!

RMarkdown Demo & Exercise

Resources

RMarkdown cheatsheet
RMarkdown: the Definitive Guide

Basic git flow ~ life cycle of a relationship

First, you **pull** them in

Then, you **stage** & **commit**

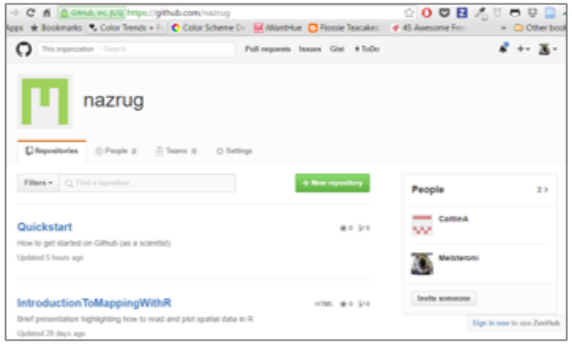
Finally, you **push** out babies

Finally, you **push** out babies

(or **push** them away...)

REMOTE

(aka Github website)



Clone (i.e., copy) repository to your computer (a one time event)

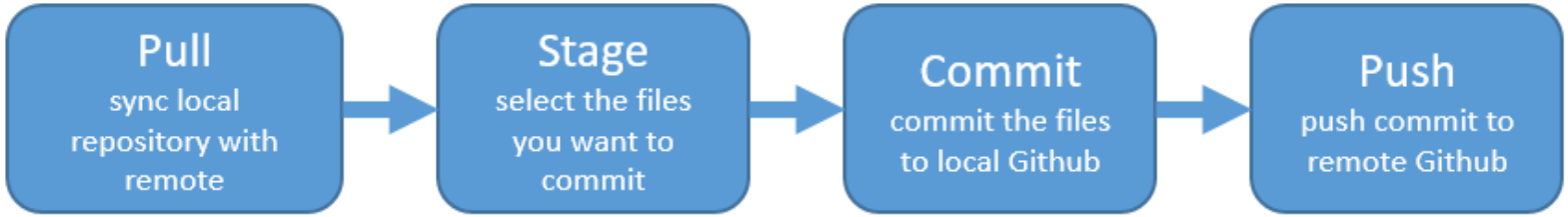
Pull remote changes

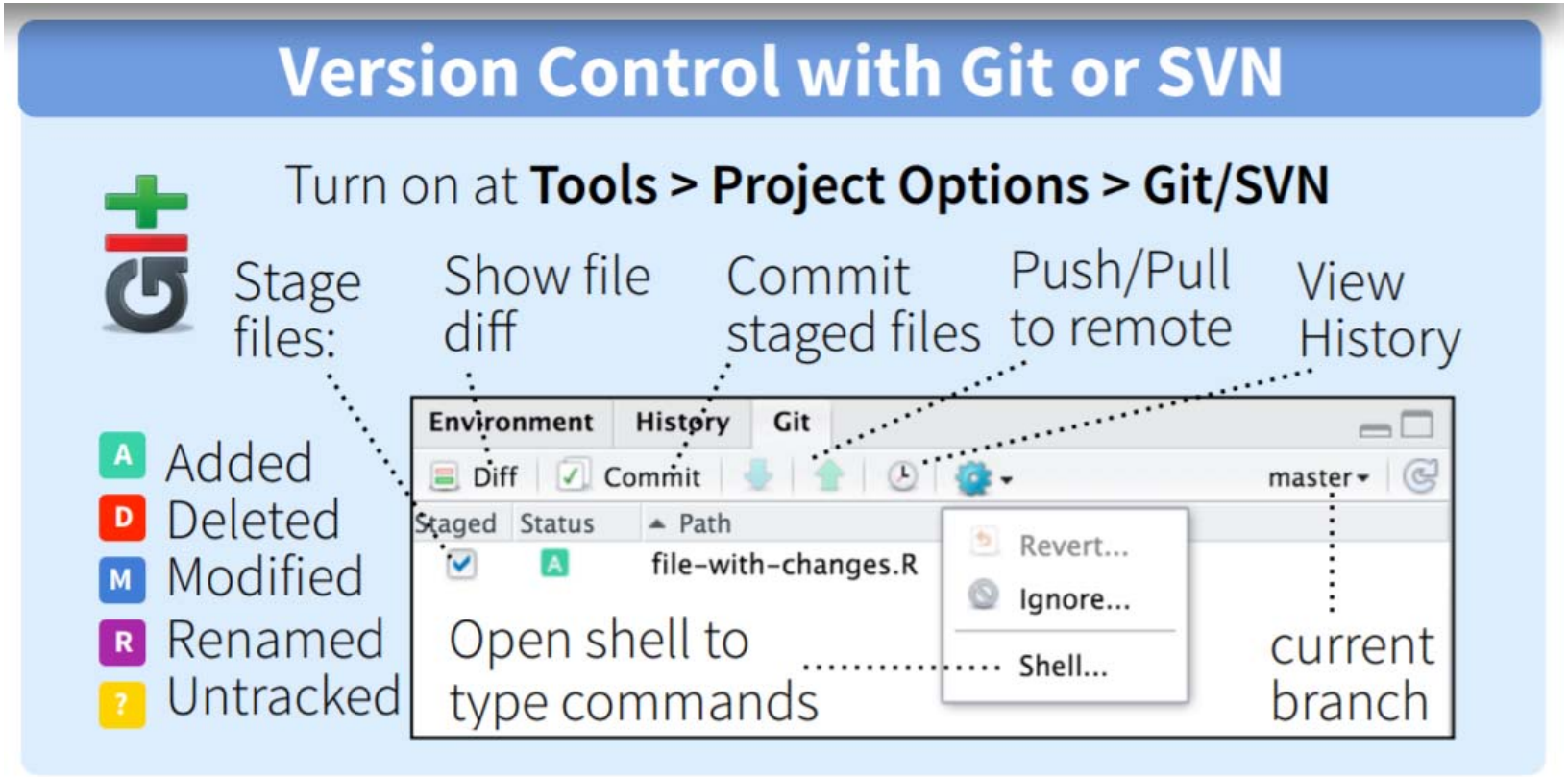
Push local changes

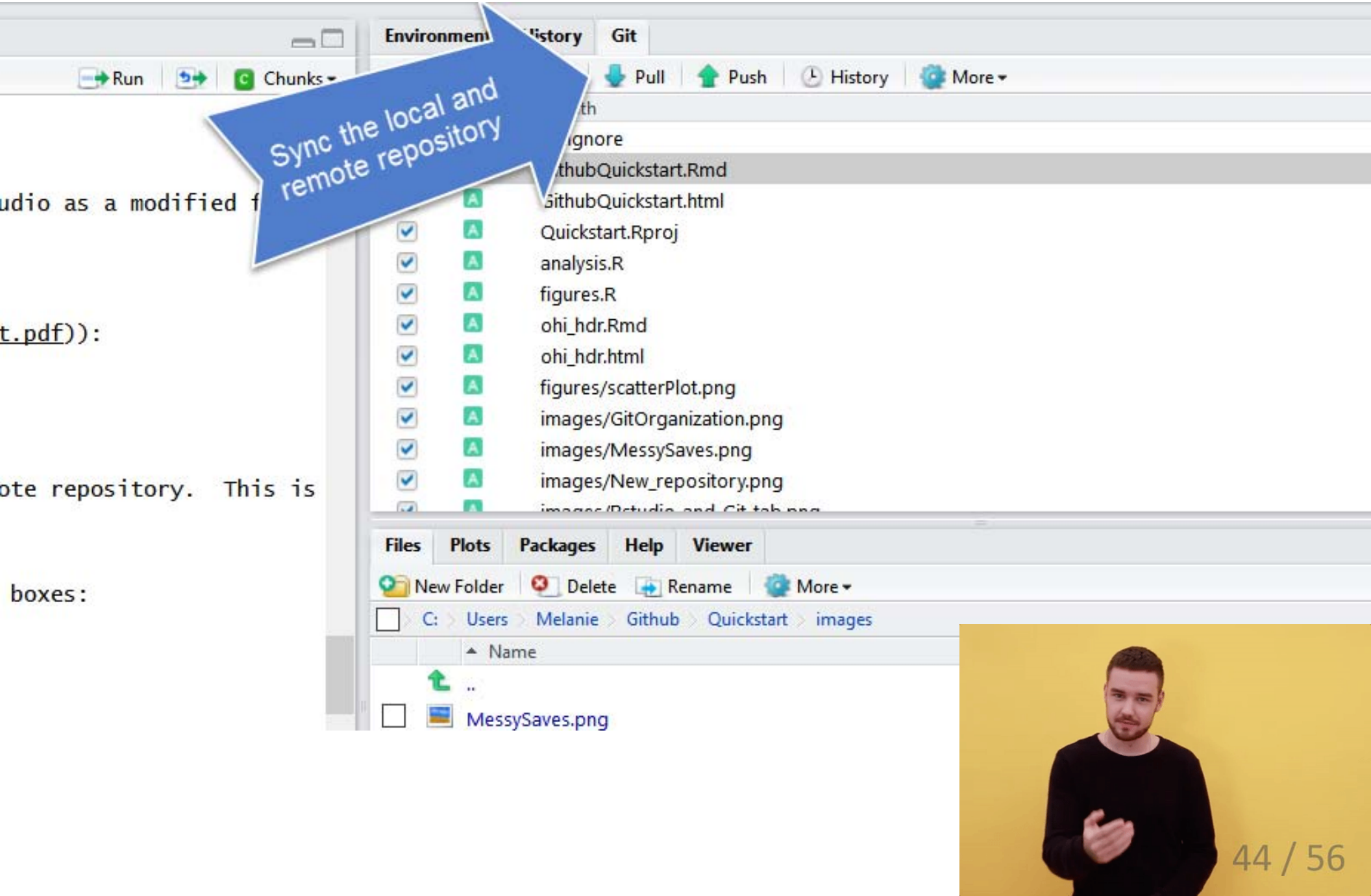
LOCAL

(aka your computer)

Sync from RStudio to GitHub






Run
Chunks
udio as a modified f
t.pdf)):
ote repository. This is
boxes:
Environment
History
Git
Pull
Push
History
More
Sync the local and remote repository
ignore
thubQuickstart.Rmd
GithubQuickstart.html
Quickstart.Rproj
analysis.R
figures.R
ohi_hdr.Rmd
ohi_hdr.html
figures/scatterPlot.png
images/GitOrganization.png
images/MessySaves.png
images/New_repository.png
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
C: Users Melanie Github Quickstart images
Name
..
MessySaves.png

Chunks

Environment | History | Git

Diff | Commit | Pull | Push | History | More

Staged	Status	Path
☑	A	.gitignore
☑	A	GithubQuickstart.Rmd
☑	A	GithubQuickstart.html
☑	A	Quickstart.Rproj
☑	A	analysis.R
☑	A	figures.R
☑	A	ohi_hdr.Rmd
☑	A	ohi_hdr.html
☑	A	figures/scatterPlot.png
☑	A	images/GitOrganization.png
☑	A	images/MessySaves.png
☑	A	images/New_repository.png

These files are "staged"

modified file:

Files | Plots | Packages | Help | Viewer

1. On Git tab select "Commit"

History | Git

Commit | Pull | Push | History | More

Status | Path

.gitignore
GithubQuickstart.Rmd
GithubQuickstart.html
Quickstart.Rproj
analysis.R
figures.R

RStudio: Review Changes

Changes | History | master | Stage | Revert | Ignore | Refresh | Pull

Staged | Status | Path

.gitignore
GithubQuickstart.Rmd
GithubQuickstart.html
Quickstart.Rproj
analysis.R
figures.R
ohi_hdr.Rmd
ohi_hdr.html
figures/scatterPlot.png

Commit message

Adding documents to Quickstart repo

Amend previous commit | Commit

3. Add a commit description

2. Can select files and see changes

4. Click the "Commit" button

Context | 5 line | Unstage All | Unstage chunk

```
1  ---
2  title: "Github Quickstart for Scientists (e.g., not social coders)"
3  author: "*Compiled on `r date()` by `r Sys.info()['user']`*"
4  output:
5    html_document:
6      toc: true
7      number_sections: true
8      theme: cerulean
9      highlight: haddock
10     includes:
11       in_header: ohi_hdr.html
12   pdf_document:
13     toc: true
14 ---
15
16 # What is Git/Github?
```

31 KB	May 1, 2016, 3:06
24.3 KB	May 1, 2016, 3:55
178.2 KB	May 1, 2016, 9:38
71.2 KB	May 1, 2016, 10:1
97.6 KB	May 1, 2016, 10:1
181.8 KB	May 1, 2016, 10:1
105.5 KB	May 1, 2016, 10:1
103.2 KB	May 1, 2016, 10:1
67.8 KB	May 1, 2016, 10:2

modified.png

anges

er ▾ | Stage | Revert | Ignore | Refresh | Push

origin/master' by 1 commit.

commit.png
pull.png
staged.png

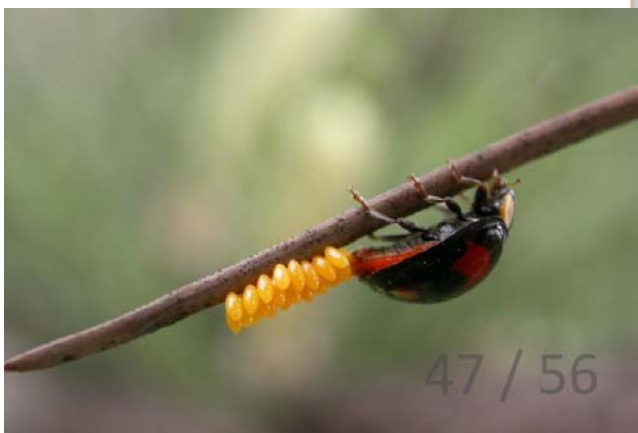
Commit message

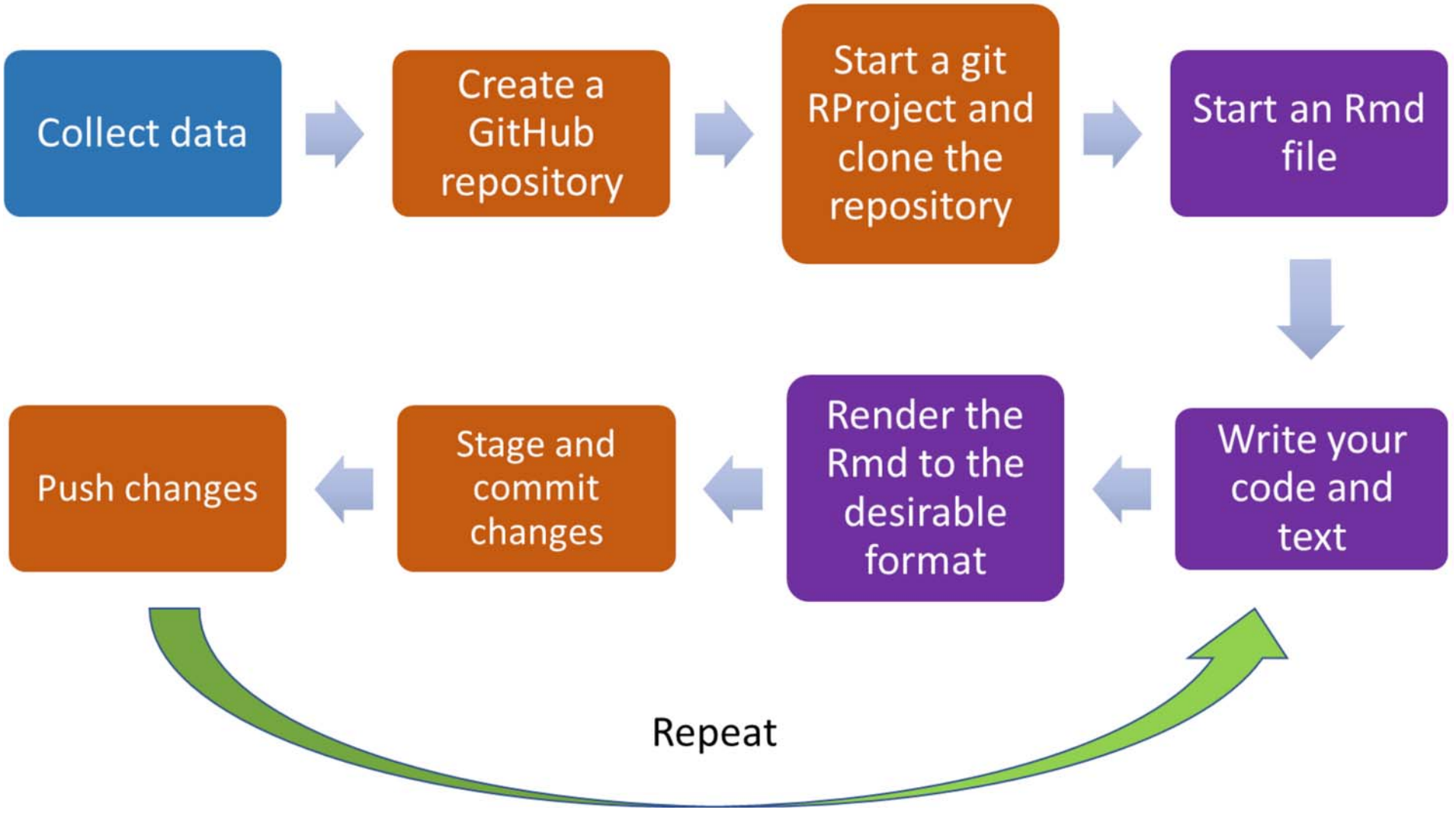
To save local changes to the remote repository

Amend previous commit

Commit

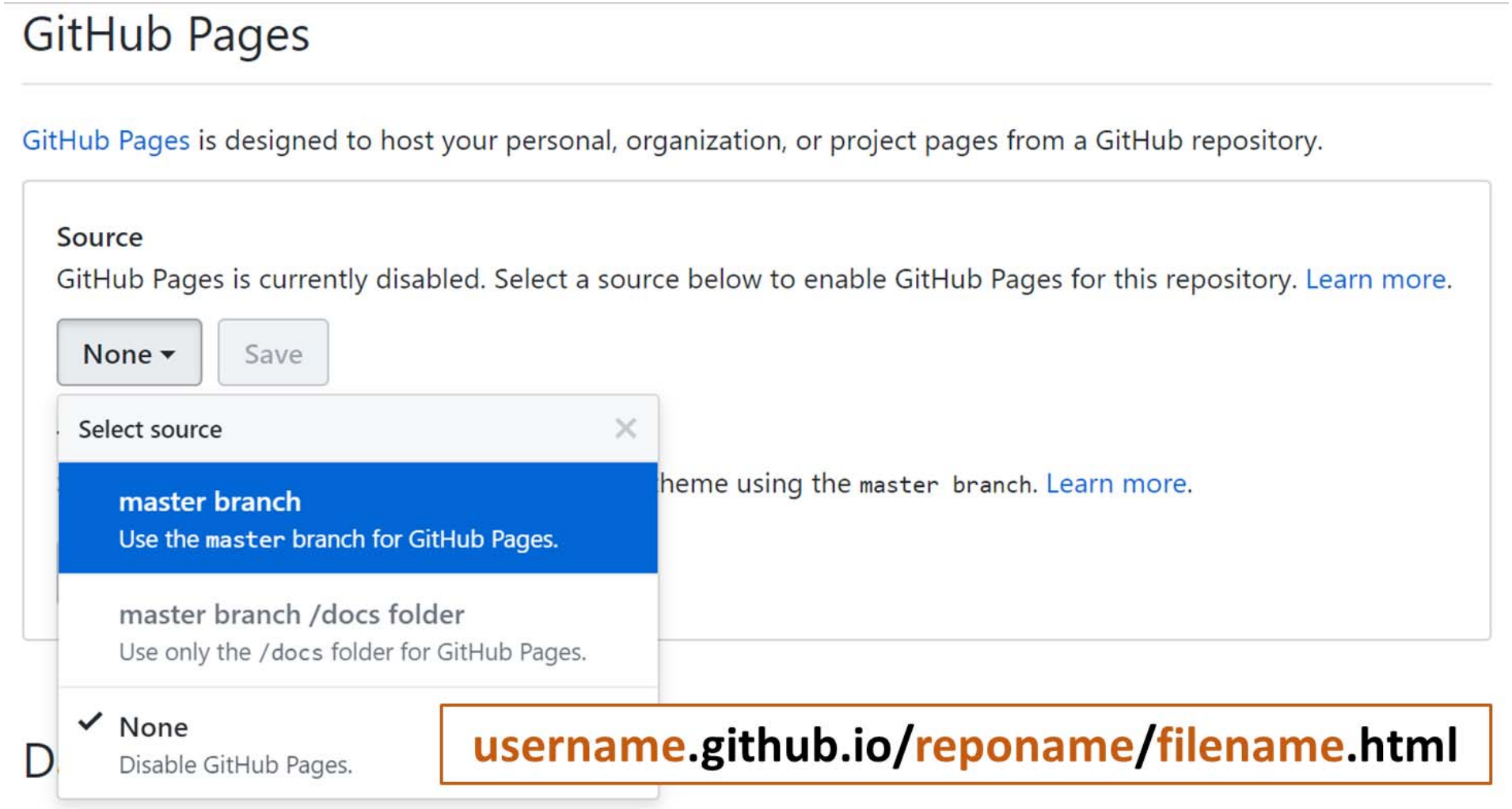
ged | Context | 5 line ▾ | Unstage All




Collect data
Create a GitHub repository
Start a git RProject and clone the repository
Start an Rmd file
Write your code and text
Render the Rmd to the desirable format
Stage and commit changes
Push changes
Repeat

Host Rmd's online with GitHub Pages

In your repo settings (scroll down):



A few tricks

- Use `index.html` to set a home page for your repo: username.github.io/reponame = username.github.io/reponame/index.html
- use `toc_float` to add a floating table of contents to your RMarkdown webpage (more details here)

```
---
output:
  html_document:
    toc: true
    toc_float: true
---
```

Let's collaborate!

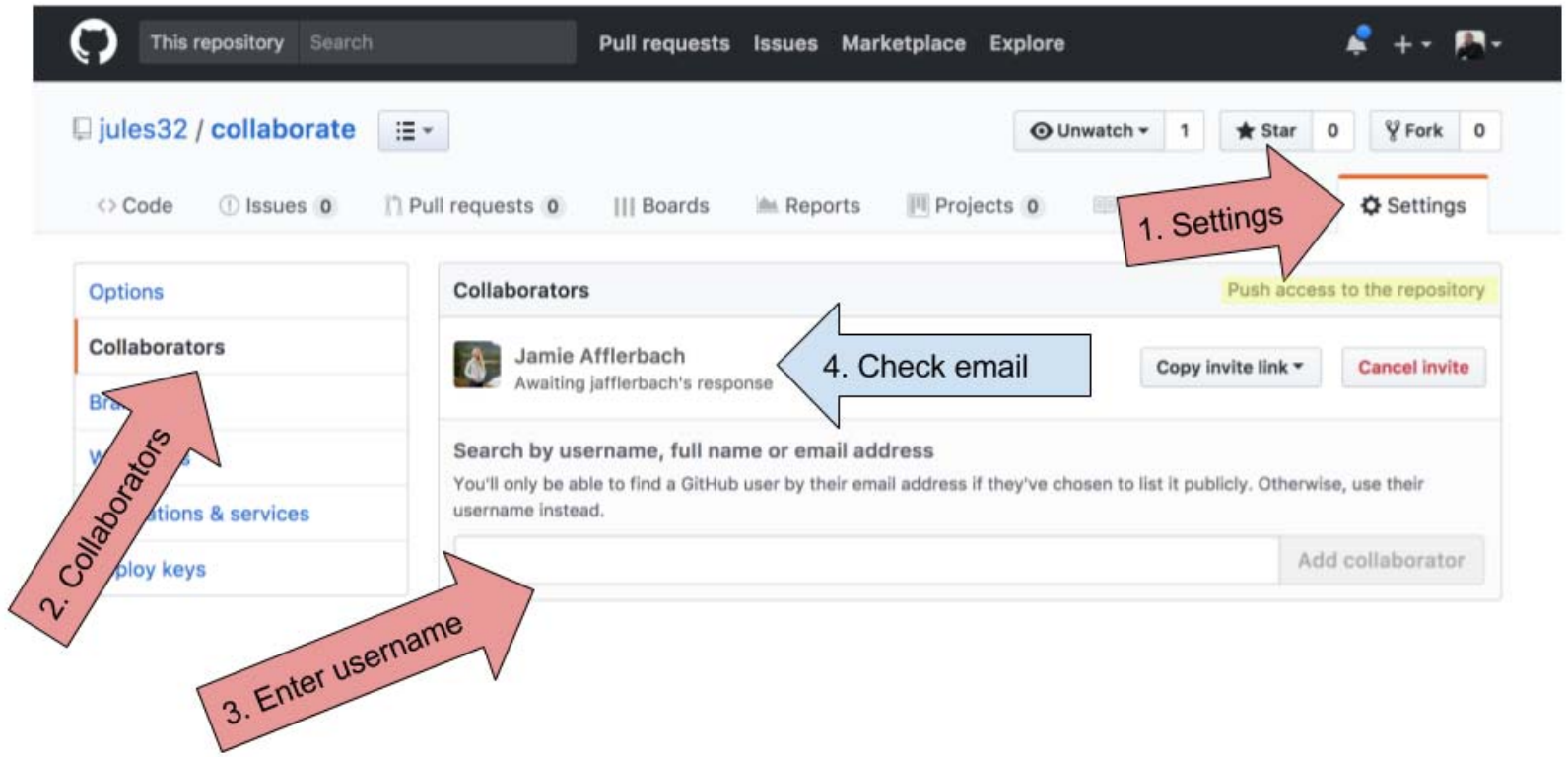
✔ Check-in

You should have:

- a repository on GitHub
- an RProject linked to the GitHub repository

Try this!

In your GitHub repo, go into `Settings` --> `Collaborators`. Add your neighbor as a collaborator



Try this!

Switch over to your *neighbor's* repo:

1. Accept your neighbor's invitation
2. Create a GitHub RStudio project
3. Make a change
4. Pull + stage/commit + push

There's more!

...but we'll cover these next time. In the mean time, check out these resources for step-by-step guides for:

- using `gitignore` to tell git what NOT to track
- dealing with merge conflicts
- collaborating using forks and branches
- burning it all down & starting afresh

Happy Git with R
NCEAS GitHub crash course
Ocean Health Index GitHub training