

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!



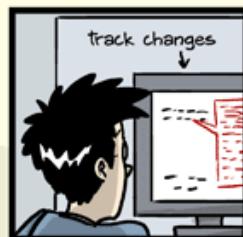
FINAL_rev.2.doc



FINAL_rev.6.COMMENTS.doc



FINAL_rev.8.comments5.
CORRECTIONS.doc



FINAL_rev.18.comments7.
corrections9.MORE.30.doc



FINAL_rev.22.comments49.
corrections.10.#@\$%WHYDID
ICOMETOGRADSSCHOOL????.doc

JORGE CHAM © 2012

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!



"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



"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](#).



This repository

Search

Pull requests Issues Gist ToDo



OHI-Science / ohiprep



Unwatch 20

Star 4

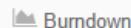
Fork 2



Issues 3



Pull requests 0



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

Browse files

p master

= 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



22 22 @@ -22,18 +22,18 @@ library(foreign)

23 23 library(data.table)

24 24 library(sp)

25 25 library(rgdal)

26 26 -library(raster)

27 27 -library(maptools)

Code in red was deleted
Code in green was added

28 28 +library(raster)

29 29 +library(maptools)

30 30 library(readr)

31 31 source '~/github/ohiprep/src/R/common.R'

32 32 scenario <- 'v2016'

33 33 dir_anx <- file.path(dir_M, 'git-annex', goal)

34 34 -dir_data_am <- file.path(dir_M, 'git-annex/globalprep/_raw_data', 'aquamaps', str_replace(scenario, 'v', 'd'))

35 35 -dir_data_iucn <- file.path(dir_M, 'git-annex/globalprep/_raw_data', 'iucn_spp')

36 36 -dir_data_bird <- file.path(dir_M, 'git-annex/globalprep/_raw_data', 'birdlife_intl')

37 37 +dir_data_am <- file.path(dir_M, 'git-annex/globalprep/_raw_data', 'aquamaps', 'd2015')

38 38 +dir_data_iucn <- file.path(dir_M, 'git-annex/globalprep/_raw_data', 'iucn_spp', 'd2015')

39 39 +dir_data_bird <- file.path(dir_M, 'git-annex/globalprep/_raw_data', 'birdlife_intl', 'd2015')

40 40 dir_git <- file.path('~/github/ohiprep', goal)

41 41 source(file.path(dir_git, scenario, 'cpp_fnv.R'))

GitHub terminology

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

GitHub: let's get oriented!

Account name / repository name

Repo description + website

Files & folders

<https://github.com/tidyverse/dplyr>

tidyverse / dplyr

Pull requests Issues Marketplace Explore

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>

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 dune 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 8 months ago

docs lang -> call2

Account name / repository name

Repo description + website

Files & folders

<https://github.com/tidyverse/dplyr>

tidyverse / dplyr

Pull requests Issues Marketplace Explore

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>

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 dune 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 8 months ago

docs lang -> call2



Search or jump to...



Pull requests Issues Marketplace Explore



tidyverse / dplyr

Code

Issues 110

Pull requests 8

Projects 2

Insights

Fork the repository
Create a copy on your account



993

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

Branches
Alternate timelines

Clone or download
Get a local copy

Clone or download ▾

romainfrancois using Rcpp_1.0.0 from CRAN

last 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

Code

Issues 110

Watch 251

Star 2,697

Fork 993

Issues

Bugs, feature requests

dplyr: A grammar of data manipulation <https://dplyr.tidyverse.org>

r data-manipulation grammar

5,587 commits

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

.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

a set with documentation. (#2431)

8 months ago

data

lang -> call2

2 years ago

docs

Commit dates

Files/folders

Commit messages

Commit history

a month ago

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



This organization

Search

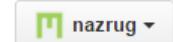
Pull requests Issues Gist ToDo



Create a new repository

A repository contains all the files for your project, including the revision history.

Owner



Repository name

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**

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!

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 - Google Chrome x nazrug/Quickstart: x

<https://github.com/nazrug/Quickstart>

GitHub, Inc. [US] Apps Bookmarks Color Themes + Pa Color Scheme De iWantHue Flossie Teacakes: 45 Awesome Free Cute Japanese Colors The Graphics Co

This repository Search Pull requests Issues Gist ToDo

nazrug / Quickstart Unwatch 1 ★

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

How to get started on GitHub (as a scientist) — Edit

1 commit 1 branch 0 releases 1

Branch: master New pull request New file Upload files Find file HTTPS https://github.com/nazrug/

Melsteroni Initial commit Latest commit

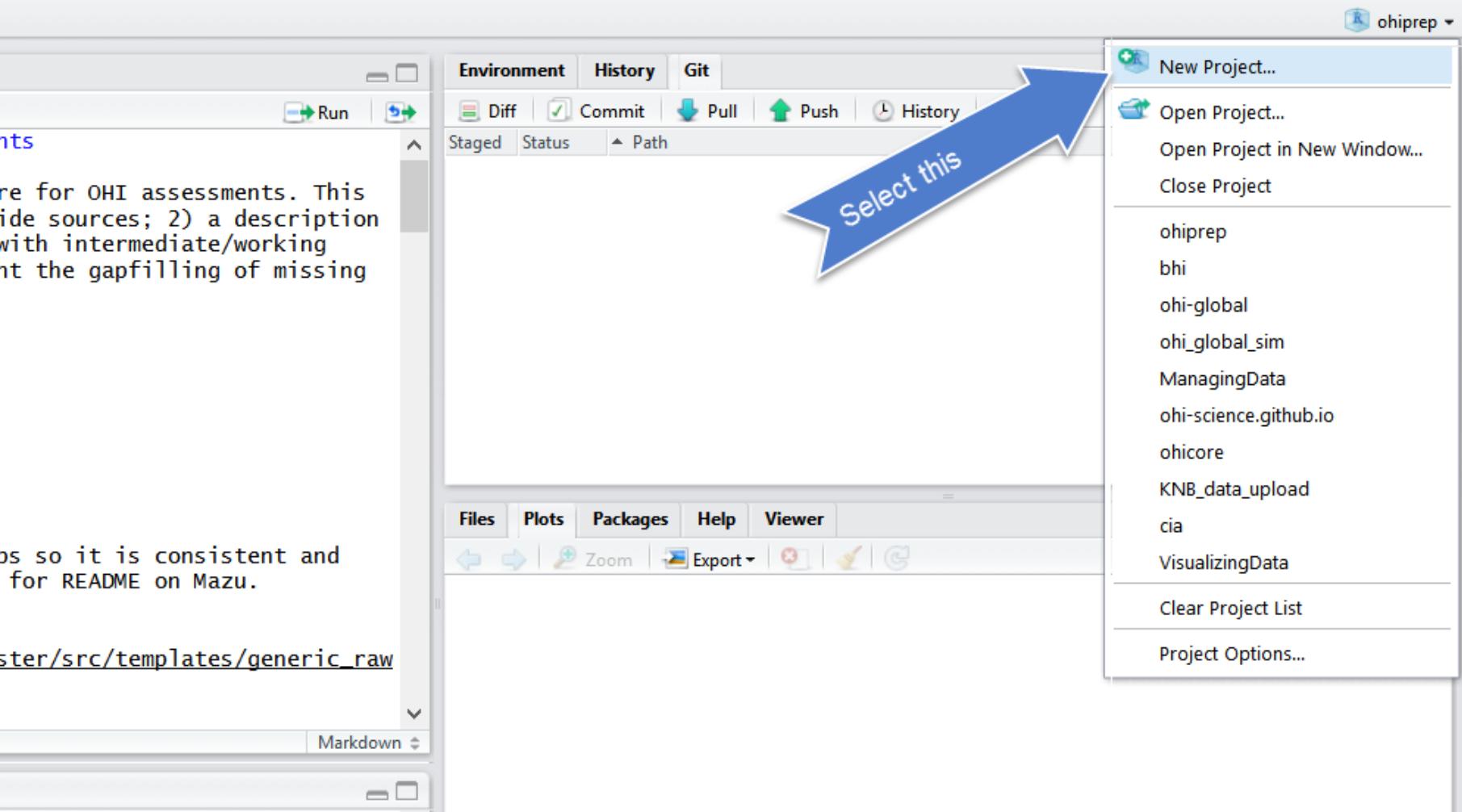
README.md Initial commit

README.md

Quickstart



Copy this



New Project

Create Project



New Directory

Start a project in a brand new working directory



Existing Directory

Associate a project with an existing directory



Version Control

Checkout a project from a version control repository



Select this

Cancel

how

New Project

Back

Create Project from Control

Select this



Git

Clone a project from a Git repository



Subversion

Checkout a project from a Subversion repository



Cancel



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

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

Browse...

Open in new window

Create Project

Cancel



Inspect your repository

~/github/my-repo - gh-pages - RStudio

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 Staged Status Path

my-repo

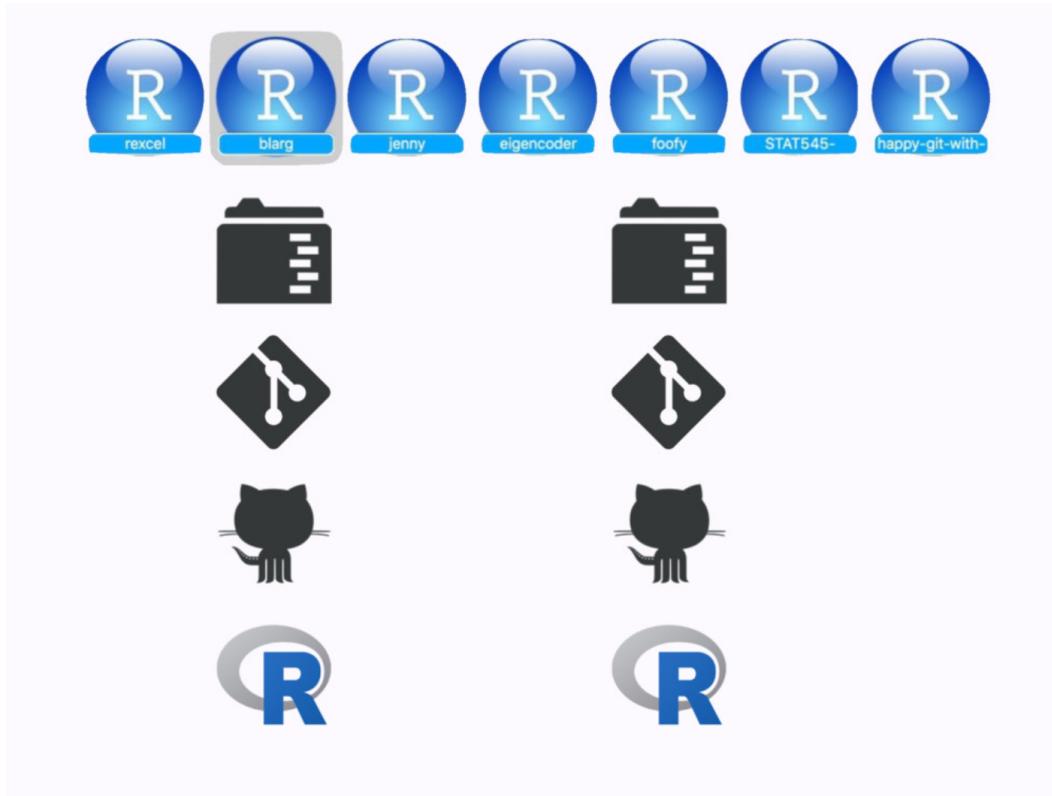
Files Plots Packages Help Viewer

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

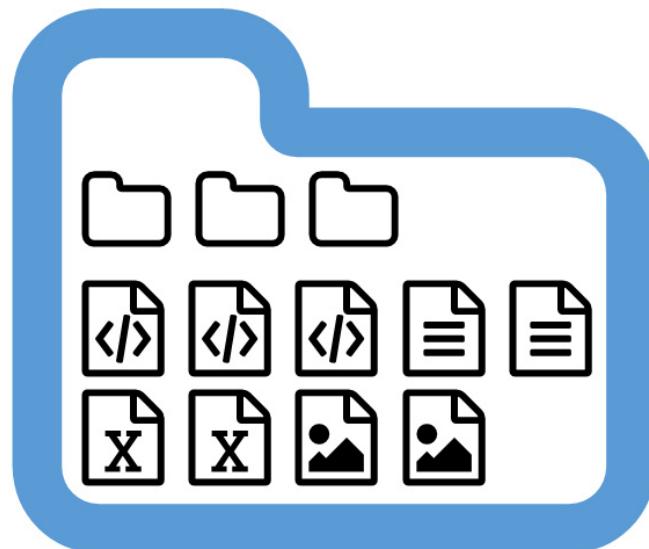
You now have a GitHub PROJECT!



<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:/יִתְ/phd/thesis/reef_surveys/data,
```

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

RMarkdown Demo & Exercise

This presentation is based on:

<https://rmarkdown.rstudio.com/lesson-1.html>

<https://bookdown.org/yihui/rmarkdown/>

<https://github.com/adam-p/markdown-here/wiki/Markdown-Cheatsheet>

<https://ourcodingclub.github.io/2016/11/24/rmarkdown-1.html>

What is markdown?

A lightweight text editor originally built to be presented as html



makeameme.org

What is Rmarkdown?

A **one-stop-shop** for writing code and sharing your work with various formats



Type of outputs for Rmarkdown

- Research notebook ([R Notebook](#))
- Reports ([html](#), [pdf](#), or [word](#))
- Websites ([Rmarkdown wbsite](#)\ [blogdown](#))
- Books ([bookdown](#))
- Presentations ([xaringan](#)\ [ioslides](#))

The different parts of Rmarkdown

YAML Title

Plain markdown text

Code chunk

Magma colors

The screenshot shows the RStudio interface with an Rmarkdown file open. The YAML header at the top left contains the title and output type. A callout points to the first few lines of the code as 'YAML Title'. The main content area contains plain markdown text describing the demo and two code chunks. One code chunk uses the viridis package to create a contour map, and the other uses the magma package. Arrows point from the labels to their respective sections: 'Plain markdown text' to the descriptive text, 'Code chunk' to the first code block, and 'Magma colors' to the resulting heatmap plot.

```
1 ---  
2 title: "Viridis Demo"  
3 output: html_document  
4 ---  
5  
6 ```{r include = FALSE}  
7 library(viridis)  
8 ...  
9  
10 The code below demonstrates two color palettes in the  
11 [viridis](https://github.com/sjmgarnier/viridis) package. Each  
12 plot displays a contour map of the Maunga Whau volcano in  
13 Auckland, New Zealand.  
14 ## Viridis colors  
15  
16 ```{r}  
17 image(volcano, col = viridis(200))  
18 ## Magma colors  
19  
20 ```{r}  
21 image(volcano, col = viridis(200), option = "A")  
22 ...  
23
```

Viridis Demo

The code below demonstrates two color palettes in the `viridis` package. Each plot displays a contour map of the Maunga Whau volcano in Auckland, New Zealand.

Viridis colors

```
image(volcano, col = viridis(200))
```

Magma colors

A heatmap visualization showing the contour map of the Maunga Whau volcano in Auckland, New Zealand. The plot uses the magma color palette, transitioning from dark purple to bright yellow. The x-axis ranges from 0.0 to 1.0, and the y-axis ranges from 0.0 to 1.0. The volcano's shape is clearly visible as a bright yellow-green center surrounded by darker green and blue areas.

YAML header

- Must be enclosed in ---
- output section changes based on the type of document

```
---
title: "Edinburgh Biodiversity"
author: John Doe
date: 22/Oct/2016
output: html_document
---
```

Code chunk

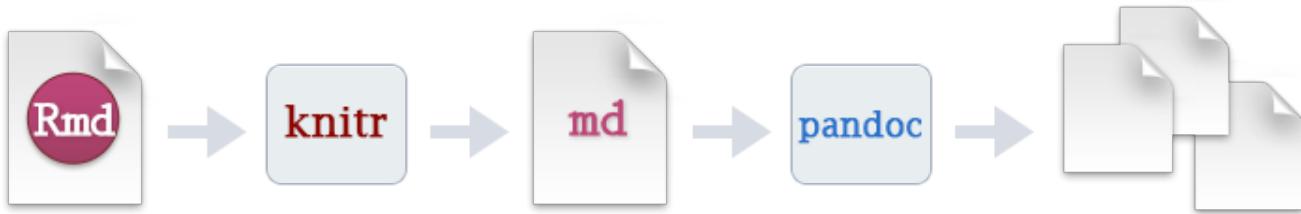
- The chunk must be enclosed between ``` and then {r} at the top
- keyboard shortcut: ctrl + Alt + I (Windows) or cmd + Alt + I (MacOS)

```
```{r}
A <- c("a", "a", "b", "b")
B <- c(5, 10, 15, 20)
dataframe <- data.frame(A, B)
plot(dataframe)
```
```

Plain text

Here you write everything that isn't a code by using markdown related symbols to create headers, bold text , etc.

Workflow of Rmd file



You create an Rmd file, then you knit it with knitr package, it is passed to one of the rendering packages like pandoc or LaTex, and the final format is created

Practice time

1. Open an Rmd file in the project you created before
2. Choose the type of document you want to create
3. Write some plain text and some analyses in the code chunk
4. Click on the knit button at the top bar (or ctrl/cmd + shift + K)

Congratulations!!! 🎉🎊🥳 😊 You created your first Rmd file!! 😊



Rmd plain text syntax



Headers

```
# Header 1  
## Header 2  
### Header 3  
#### Header 4  
##### Header 5  
##### Header 6
```

Header 1

Header 2

Header 3

Header 4

Header 5

Header 6

Alternatively, for H1 and H2, an underline-ish style:

Alt-H1

=====

Alt-H2

Alt-H1

Alt-H2

Highlight text

Bold text __Bold__ or **Bold**

Italics _Italics_ or *Italics*

~~strikethrough~~ ~~strikethrough~~

Use these two links to learn how to adapt the text and the code to your needs:

<https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf>

<https://ourcodingclub.github.io/2016/11/24/rmarkdown-l.html>

Other formats of Rmarkdown_

The YAML header for website

Here is an example from the website used in our last session

```
---
```

```
title: "Visualizations in R"
author: "Alex Slavenko and Maria Novosolov"
output:
  html_document:
    toc: true
    toc_float: true
    collapsed: true
    smooth_scroll: true
    toc_depth: 3
    runtime: shiny_prerendered
    # number_sections: true
---
```

Books

- Written similar to a webslite but has a different YAML header
- Uses bookdown package to generate

More information and **websites** and guidlines on how to build them can be found [here](#)

More information about how to build **books** can be found [here](#)

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

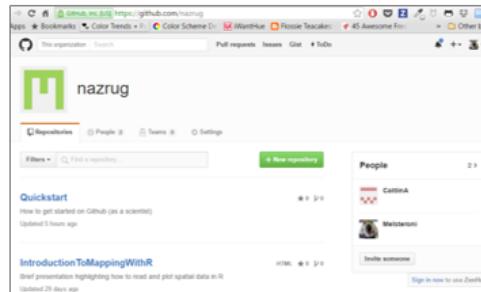


(or push them away...)



REMOTE

(aka Github website)



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

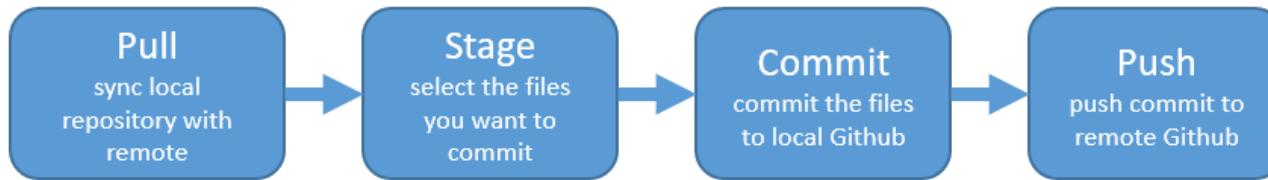
Pull remote changes

Push local changes



(aka your computer)

Sync from RStudio to GitHub



Version Control with Git or SVN

Turn on at **Tools > Project Options > Git/SVN**

The screenshot shows the RStudio interface with the Git tab selected in the top navigation bar. The status bar at the bottom right indicates "master" and "current branch". A context menu is open over a file named "file-with-changes.R", showing options like "Revert...", "Ignore...", and "Shell...".

Stage files:

- A** Added
- D** Deleted
- M** Modified
- R** Renamed
- ?** Untracked

Show file diff

Commit staged files

Push/Pull to remote

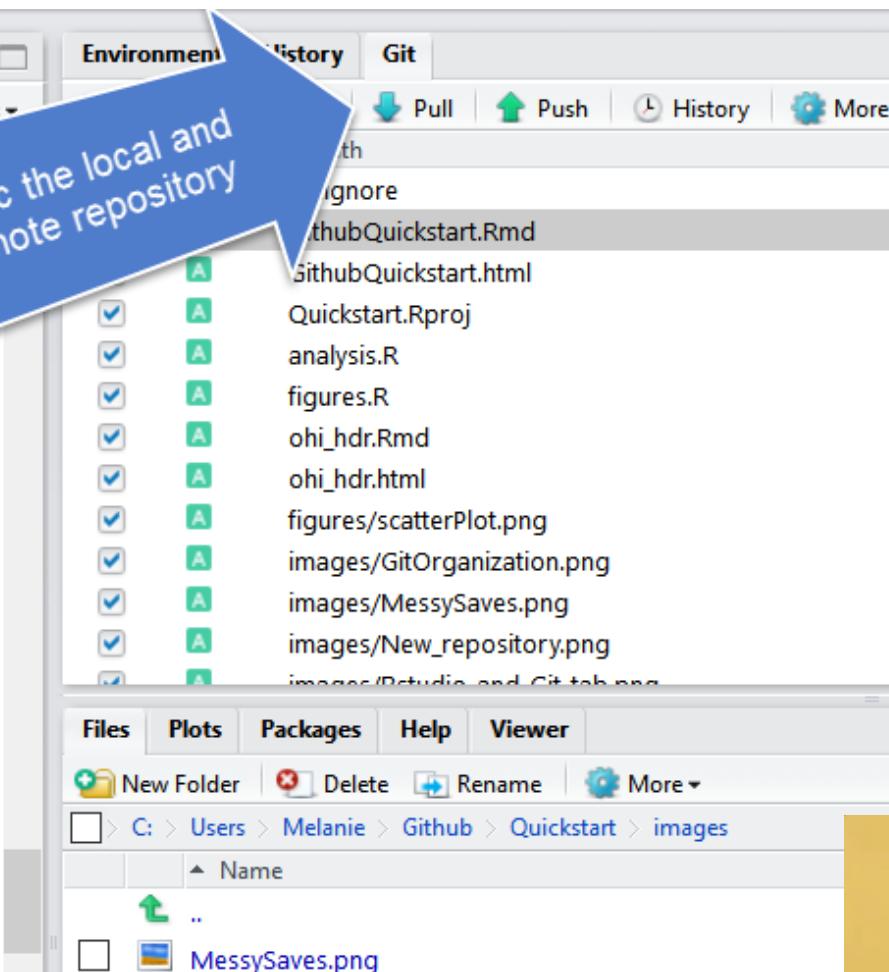
View History

Open shell to type commands

udio as a modified file.
t.pdf)):
ote repository. This is

boxes:

Sync the local and
remote repository



The screenshot shows the RStudio interface with the Git tab selected. A blue callout box points to the 'Sync' button in the top navigation bar. The main pane displays a list of files in a GitHub repository, including Rmd and Rproj files, analysis scripts, figures, and images. Below the list is a file browser showing the local directory structure on disk.

| Name |
|----------------|
| .. |
| MessySaves.png |



These files are "staged"

modified file:

Environment History Git

Diff Commit Pull Push History More

| Staged | Status | Path |
|-------------------------------------|--------|--------------------------------|
| <input checked="" type="checkbox"/> | A | .gitignore |
| <input checked="" type="checkbox"/> | A | GithubQuickstart.Rmd |
| <input checked="" type="checkbox"/> | A | GithubQuickstart.html |
| <input checked="" type="checkbox"/> | A | Quickstart.Rproj |
| <input checked="" type="checkbox"/> | A | analysis.R |
| <input checked="" type="checkbox"/> | A | figures.R |
| <input checked="" type="checkbox"/> | A | ohi_hdr.Rmd |
| <input checked="" type="checkbox"/> | A | ohi_hdr.html |
| <input checked="" type="checkbox"/> | A | figures/scatterPlot.png |
| <input checked="" type="checkbox"/> | A | images/GitOrganization.png |
| <input checked="" type="checkbox"/> | A | images/MessySaves.png |
| <input checked="" type="checkbox"/> | A | images/New_repository.png |
| <input checked="" type="checkbox"/> | A | images/Rstudio_and_Git_tab.png |

Files Plots Packages Help Viewer



The RStudio IDE-cheatsheet.pdf)):

1. On Git tab select "Commit"

2. Can select files and see changes

3. Add a commit description

4. Click the "Commit" button

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

Commit message

Adding documents to Quickstart repo

Amend previous commit Commit

Unstage All

Unstage chunk

modified.png

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, 10:11
97.6 KB May 1, 2016, 10:11
181.8 KB May 1, 2016, 10:11
105.5 KB May 1, 2016, 10:11
103.2 KB May 1, 2016, 10:11
67.8 KB May 1, 2016, 10:21



66 / 76

anges

Stage Revert Ignore

origin/master' by 1 commit.

commit.png
pull.png
staged.png

Commit message

Refresh

Push

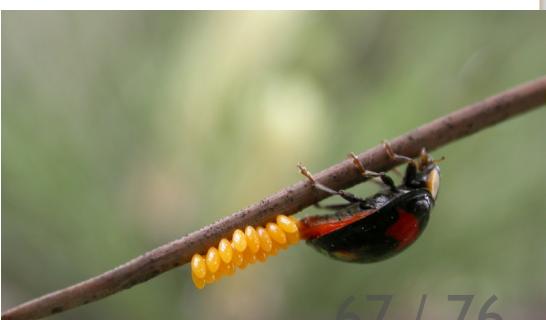
To save local
changes to the
remote repository

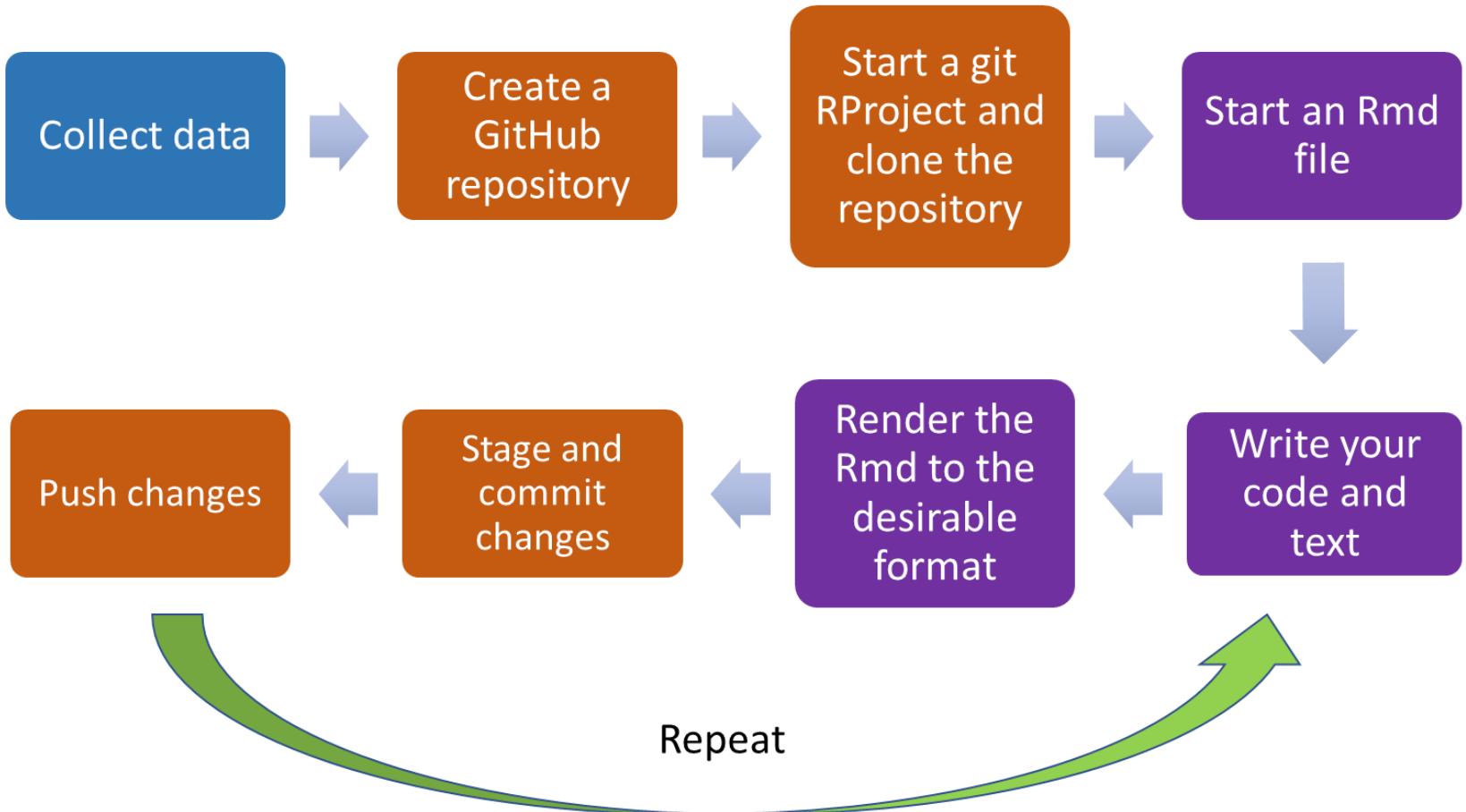
Amend previous commit

Commit

ed Context 5 line ▾

 Unstage All





Host Rmd's online with GitHub Pages

In your repo settings (scroll down):

GitHub Pages

GitHub Pages is designed to host your personal, organization, or project pages from a GitHub repository.

Source

GitHub Pages is currently disabled. Select a source below to enable GitHub Pages for this repository. [Learn more](#).

Select source X

- master branch**
Use the master branch for GitHub Pages.
- master branch /docs folder
Use only the /docs folder for GitHub Pages.
- None
Disable GitHub Pages.

username.github.io/reponame/filename.html

D

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!

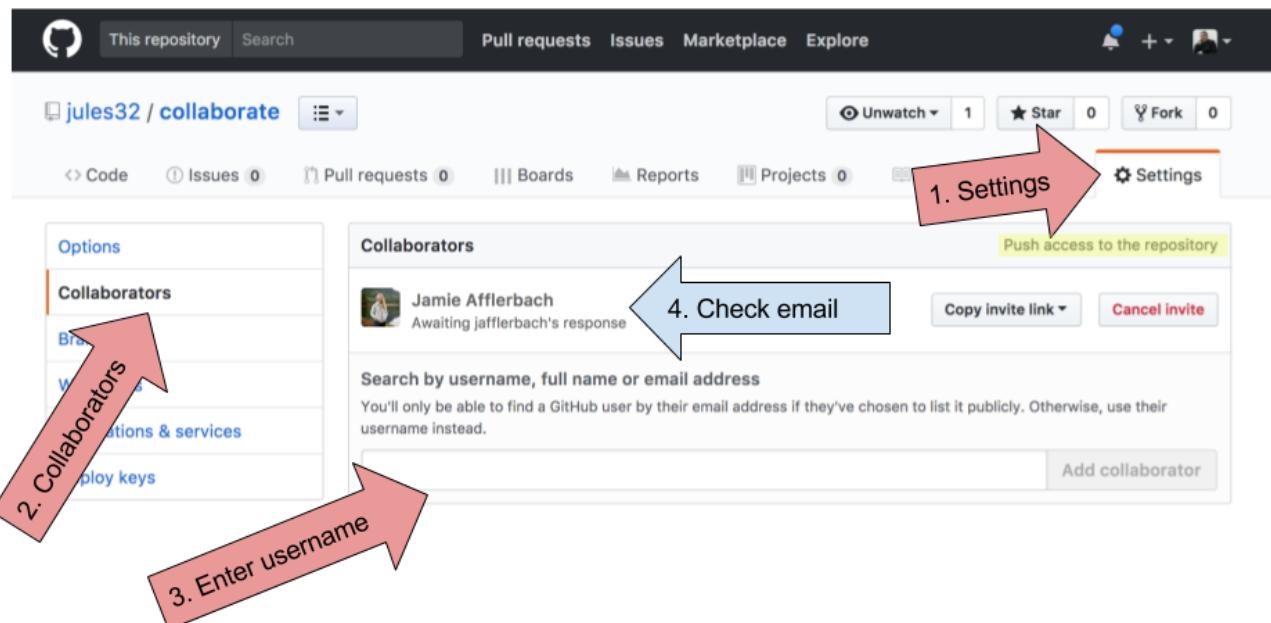
Check-in

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](#)