

GitHub for Fisheries Scientists

AFS-AK Chapter Workshop

February 23, 2022

12:00–1:30 PM

Instructors: Justin Priest and Sara Miller

https://github.com/justinpriest/github_demo_adfg





Welcome! Your Instructors:



Sara Miller

ADF&G Biometrician 3

Justin Priest

ADF&G Fisheries Biologist 3





Overview

- 1 – What is GitHub
- 2 – Why use GitHub?
- 3 – How to use GitHub
- 4 – Show me the demo!
- 5 – Concluding Thoughts /
Best Practices / Resources



1 – What is GitHub?



1 – What is GitHub?



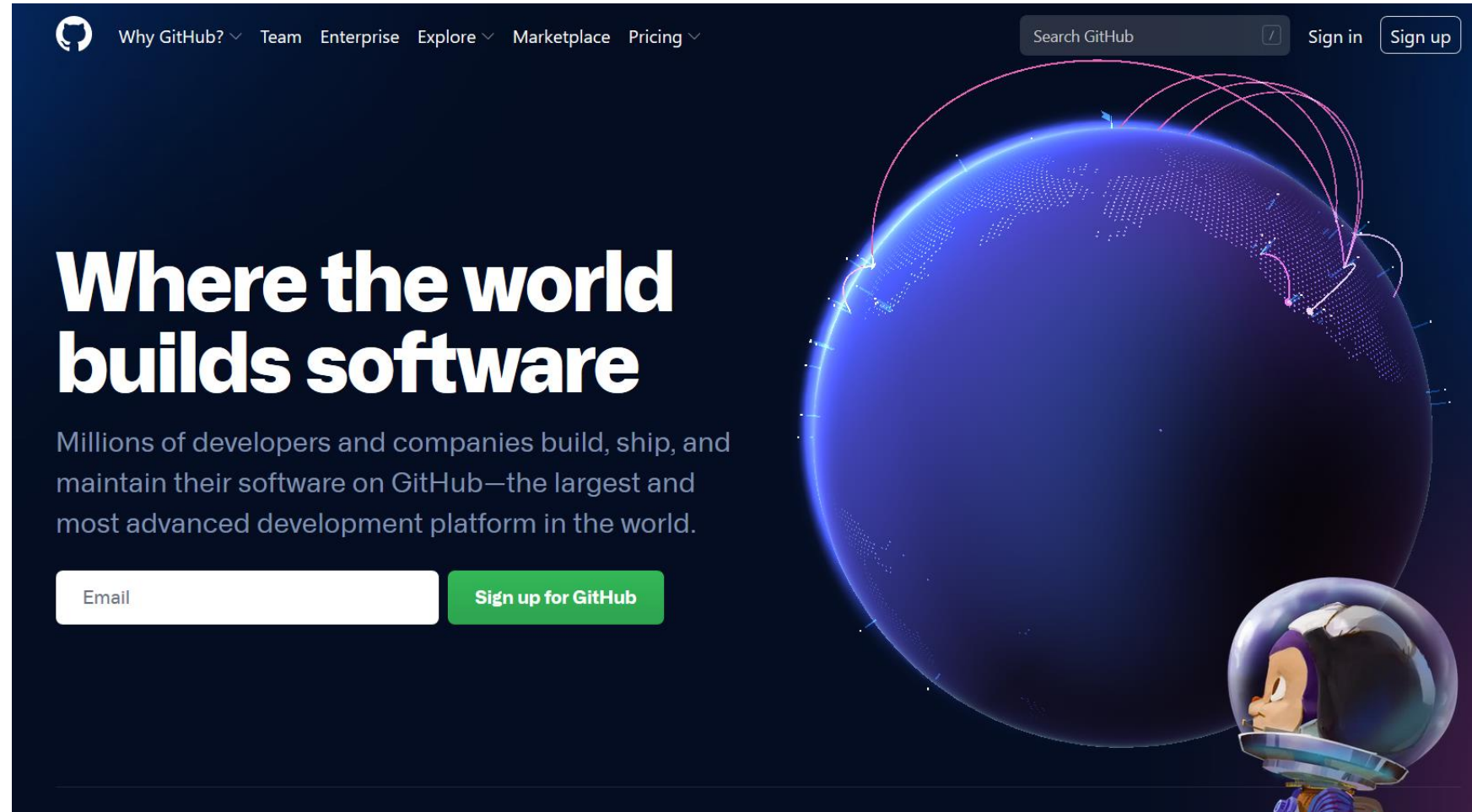
GitHub

- Git is the underlying tech
 - Integrates a version control program into other programs
- Backend (behind the scenes)
- GitHub is a specific site
 - Other sites too (e.g., GitLab and BitBucket)
- Front end (website interface)



1 – What is GitHub?

GitHub is a website that syncs with software on your computer to upload your files to github.com





1 – What is GitHub?

- Git is versioning control software
- This means we can keep track of new additions (in green) and deletions (in red)
- Later, we'll cover more about benefits

Created first draft of the trigger assessment RMD

[Browse files](#)

master



justinpriest committed on Dec 6, 2020 1 parent d592030 commit 14fb5a26a6c3063885f786a7a73380be35c8cdfc

Showing 2 changed files with 574 additions and 2 deletions.

Unified

Split

code/3_US_figures.R

```
@@ -14,6 +14,7 @@ source(here::here("code/1_US_data_import.R"))
14      troll_cpue <- read_csv(here::here("data/SEAK_Coho_TrollFPD_1981-2019.csv"),
15                             guess_max = 84000) %>% #increased guess b/c of many blanks
16      rename("Gear" = `Gear Code`,
17             "TripNum" = `Trip No`,
18             "SellDate" = `Sell Date`,
19             "StatWeek" = `Stat Week`,
20             "TrollArea" = `Troll Area`,
21             @@ -25,8 +26,10 @@ troll_cpue <- read_csv(here::here("data/SEAK_Coho_TrollFPD_1981-2019.csv"),
22             mutate(SellDate = as_date(as.POSIXct(SellDate, format = "%m/%d/%Y", tz = "US/Alaska")),
23                    District = as.factor(District),
24                    Effort_boatdays = DaysFished * HoursPerDay / 13, # Effort is standardized to a 13 hour boat day
25                    CohoCPUE = CohoCatch / Effort_boatdays) %>%
26                    dplyr::select(Year, SellDate, StatWeek, TrollArea, District, StatArea, CohoCatch, Effort_boatdays, CohoCPUE)
27                    CohoCPUE = CohoCatch / Effort_boatdays,
28                    TripNumber = paste0(Year, "-", TripNum)) %>%
29                    dplyr::select(Year, TripNumber, SellDate, StatWeek, TrollArea, District,
30                                StatArea, CohoCatch, Effort_boatdays, CohoCPUE)
```



2 – Why Use GitHub??



2 – Why use GitHub: Version Control

Versioning: The
bane of digital file
storage

Enter Version
Control!

Desktop > ADFG Local Repos > PSC_CohoReport_2020 > code

Name	Date modified	Type	Size	Size
example_code	5/10/2021 9:03 AM	File folder		
functions	5/17/2021 1:48 PM	R File	6 KB	10 KB
pinkforecast2022	5/17/2021 7:13 PM	R File	16 KB	1 KB
pinkforecast2022_5_17_2021_v1	5/17/2021 7:13 PM	R File	16 KB	30 KB
pinkforecast2022_5_17_2021_v2	5/17/2021 7:13 PM	R File	16 KB	1 KB
pinkforecast2022_5_18_2021_v3	5/17/2021 7:13 PM	R File	16 KB	35 KB
pinkforecast2022_5_18_2021_v4	5/17/2021 7:13 PM	R File	16 KB	4 KB
pinkforecast2022_near_actually_the_final	5/17/2021 7:13 PM	R File	16 KB	14 KB
pinkforecast2022_near_final	5/17/2021 7:13 PM	R File	16 KB	11 KB
pinkforecast2022_near_final_final	5/17/2021 7:13 PM	R File	16 KB	8 KB
satellite_data_monthly	5/10/2021 9:04 AM	R File	30 KB	

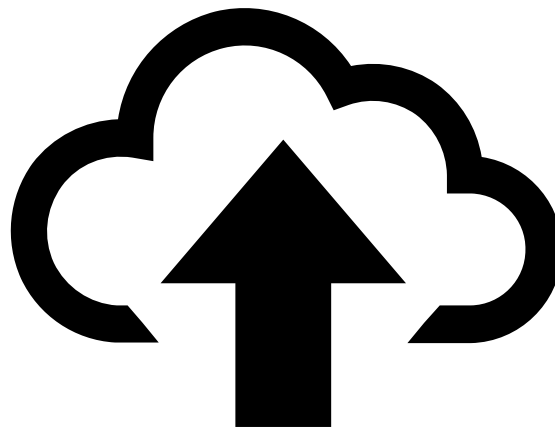
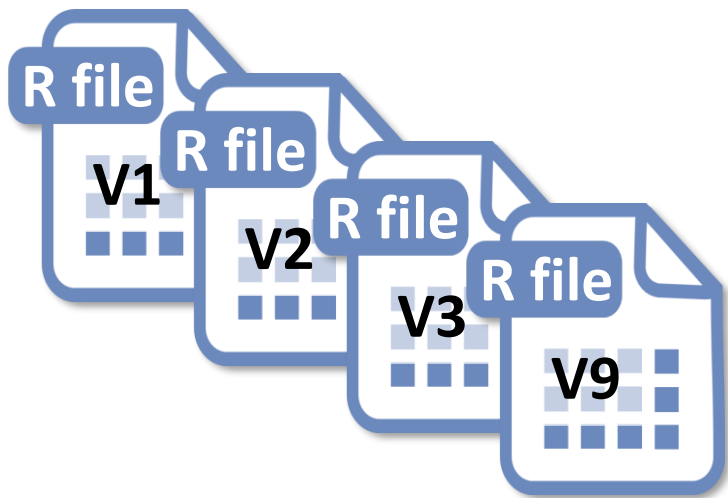
credit: Julie Lowndes & Allison Horst



2 – Why use GitHub?

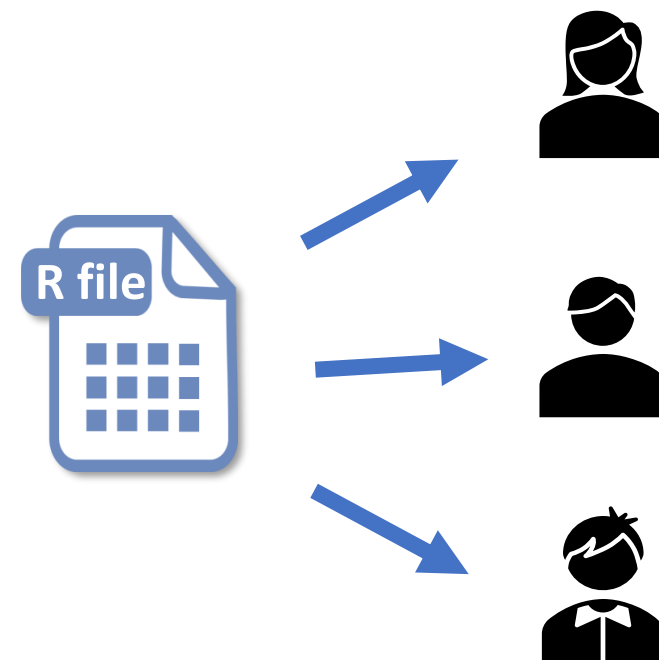
GitHub helps scientists with:

Version control helper



Project & data
analysis storage

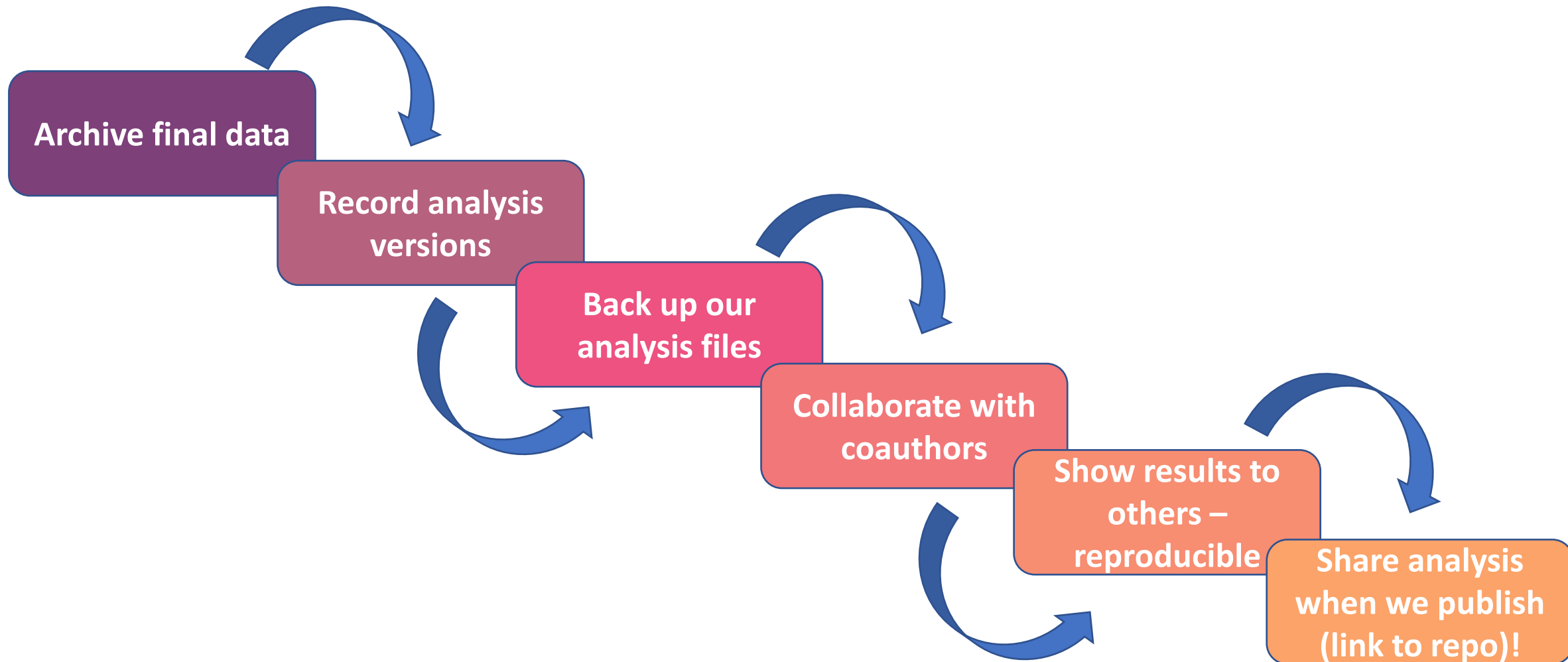
Code sharing





2 – Why use GitHub?

As scientists, we can use GitHub over a project's lifecycle:



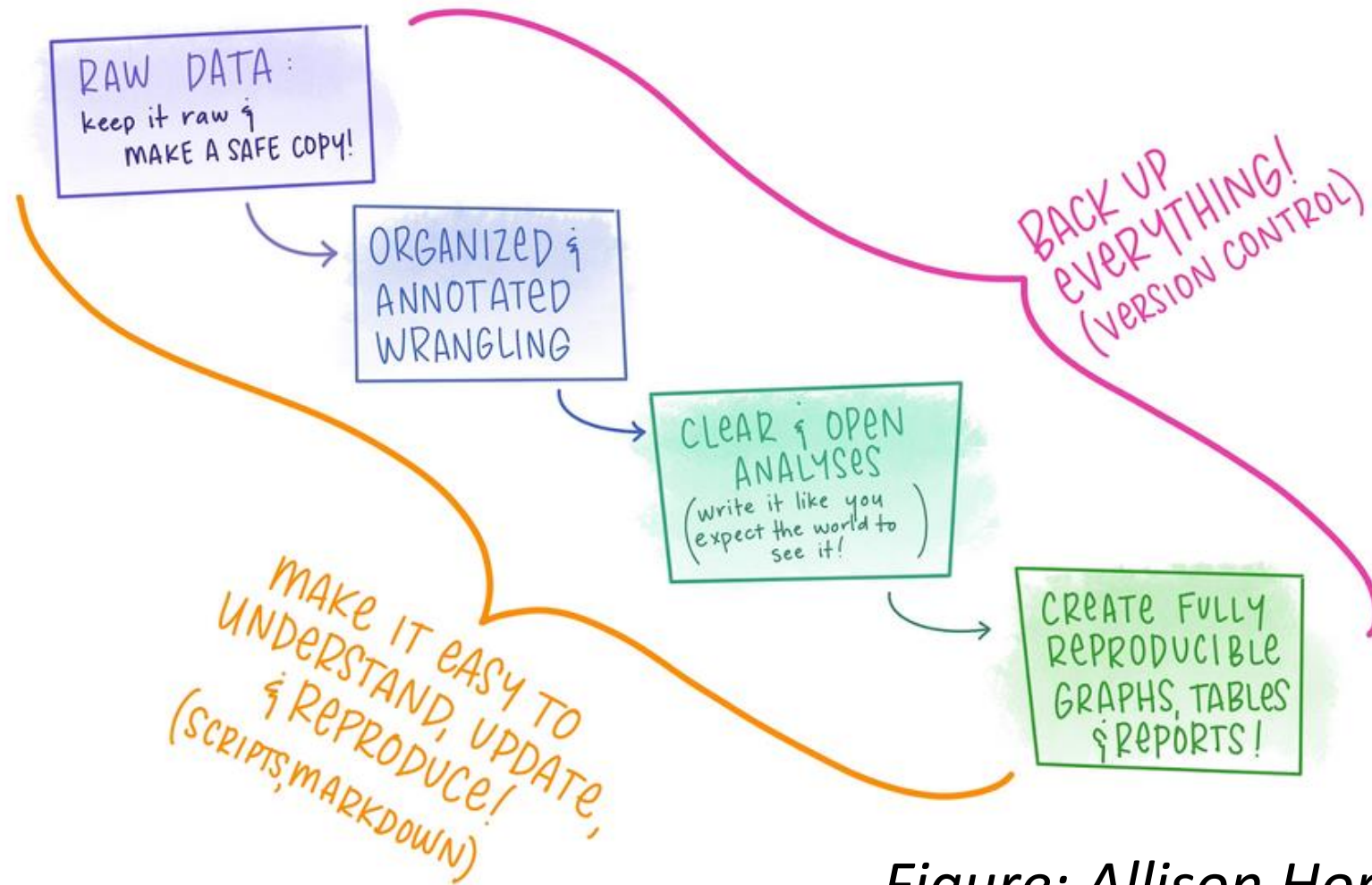


Figure: Allison Horst



3 – How To Use GitHub

One time setup & daily usage



3 – A bit of Jargon

- repo (repository) = a “folder” containing an individual project
- remote = the repo on github.com
- local = the repo on your computer
- clone = copy the remote repo to your computer
- pull = fetching in changes or merging them
- push = send changes to a remote repo
- commit = individual change to a file
- fork = copy of a repo (allows you to freely experiment with changes without affecting the original project)



3 – A bit of Jargon





3 – How to use GitHub: Setup

One Time Setup, part 1

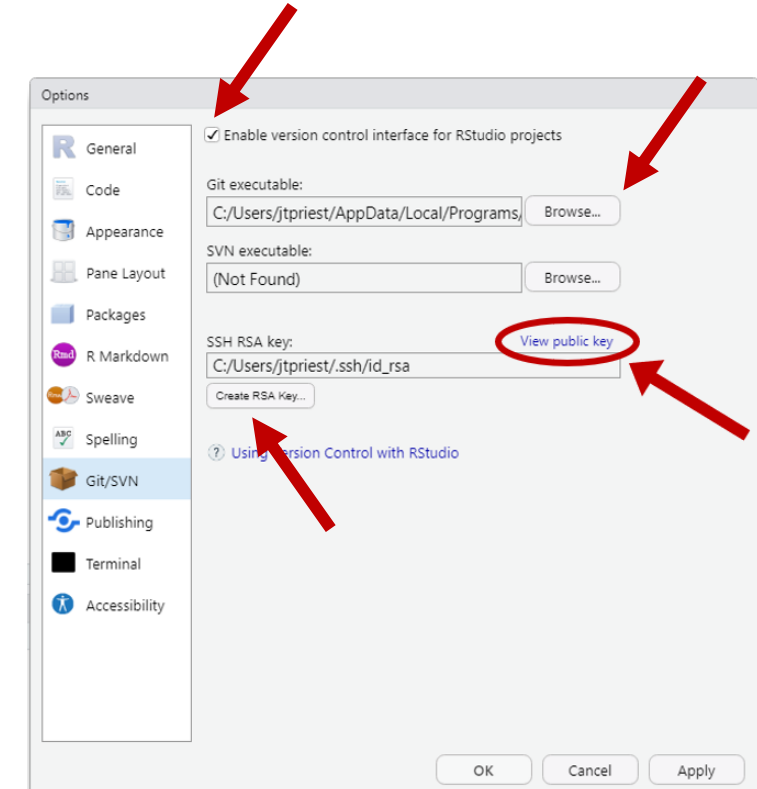
- Get a free account from github.com
- [Download](#) & install the program “git”
 - Note the folder where this installs. In Windows common places are
C:\Users\jtpriest\AppData\Local\Programs\Git *(change to your username)*
C:\Program Files\Git
 - Accept all default settings and click through until installed.



3 – How to use GitHub: Setup

One Time Setup, part 2

- In RStudio, under Tools, -> Global Options -> “Git/SVN”
 - In the window that pops up, select “Enable version control interface for RStudio projects”.
 - Under “Git executable” click browse and select the correct folder noted before. Select the git.exe file, typically found in folder “bin”.
 - Next click “Create RSA Key”. Note folder location. Give it a memorable password, then click Create. Once it is done, you’ll see some random art and file info. Click Close.
 - Now you’ll want to open up and view this key. Click the small blue “View public key”. At this point you’ll see some random text that starts with “ssh-rsa”. Copy this!
- Open up github.com and go to your [profile settings](#)
 - Click on “SSH and GPG keys” on the left. Click the green “New SSH key” button. Paste the key from RStudio in the key section. Give this key a unique title describing which computer you’re using.





3 – How to use GitHub: Setup

One Time Setup, part 3

- In RStudio, install and load package “usethis”. Run this code

```
library(usethis)
```

```
use_git_config(user.name = “Your Name”, user.email =  
“your.email@alaska.gov”)
```

```
git_sitrep() # This shows current name, directory, and other settings
```

Yes, you do have to change these!



3 – How to use GitHub: Setup

Setup complete!

Only a little painful and you'll never have to do that again!



3 – How to use GitHub: New proj

How to set up a new project to work in

Step 1 – Create/choose repo on github.com

- Project already exists on GitHub (“clone” it to your computer)
 - Go to repo webpage on GitHub and copy entire URL at the top (e.g., https://github.com/commfish/Chilkat_Sockeye); or
 - Click on green code button in the repo (<https://...>) OR
- Create a new project
 - While logged in, click green “New” repo button or go to <https://github.com/new>
 - Give it a name, chose public/private, add readme, (optional license & gitignore)
 - Copy entire URL at top

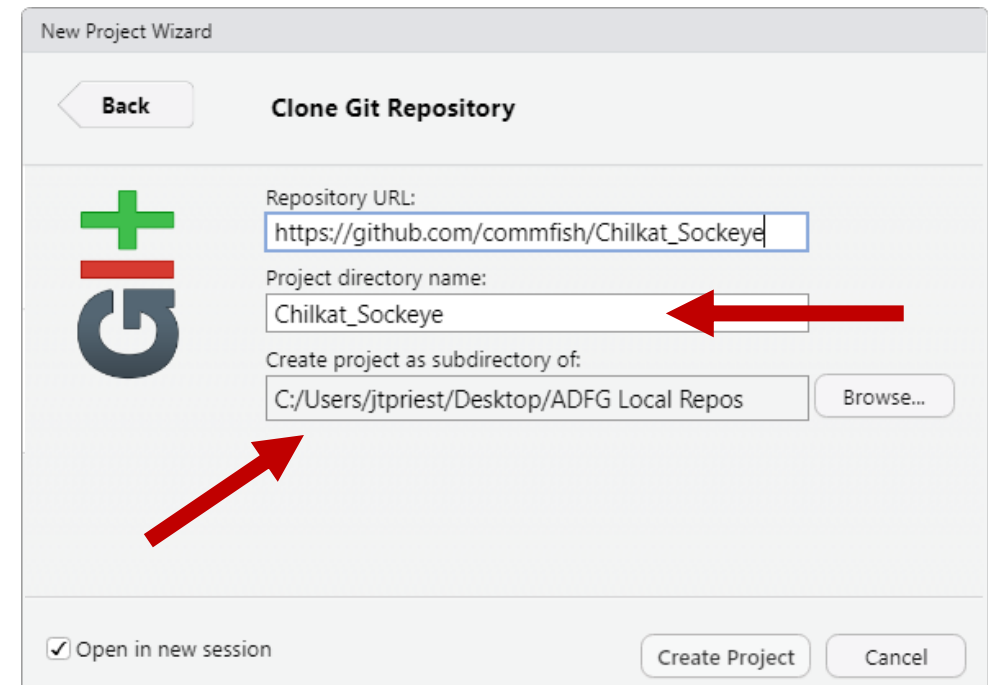


3 – How to use GitHub: New proj

How to set up a new project to work in (cloned project)

Step 2 – Connect RStudio Rproject to the repo

- Open RStudio. Click File -> New Project -> Version Control -> Git. Paste in the github.com address into the Repository URL (https://github.com/commfish/Chilkat_Sockeye)
- It auto fills Project Directory Name (this will be local folder name; change if you like). Set the directory where this folder will be. Open in a new RStudio session (or not).
- Click Create Project and all files on GitHub repo will sync. If large project, might take a few min.





3 – How to use GitHub: Review

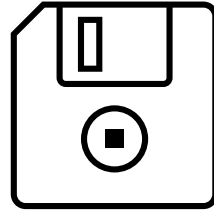
Review. We just:

- Created a GitHub account
- Connected our RStudio to this GitHub account
- Synced an RProject with a github.com repo

This. Is. huge! Now onto daily usage



3 – A bit (more) Jargon



Commit

Prepping your
changes for upload



Pull

Download changes
from github.com



Push

Upload commits
to github.com

Perform them in this order!

Untitled1 x

Source on Save

Run

Source

1

1:1 (Top Level)

R Script

Console Terminal x Jobs x

C:/Users/jtpriest/Desktop/ADFG Local Repos/github_demo_afs/

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

Tutorial

Diff

Commit

main

Staged

Status

Path

☒

A

GitHub Tutorial AFS Student.pptx

☒

A

course_description.docx

☐

?

?

github_demo_afs.Rproj

☐

?

?

~\$GitHub Tutorial AFS Student.pptx

Files

Plots

Packages

Help

Viewer

New Folder

Delete

Rename

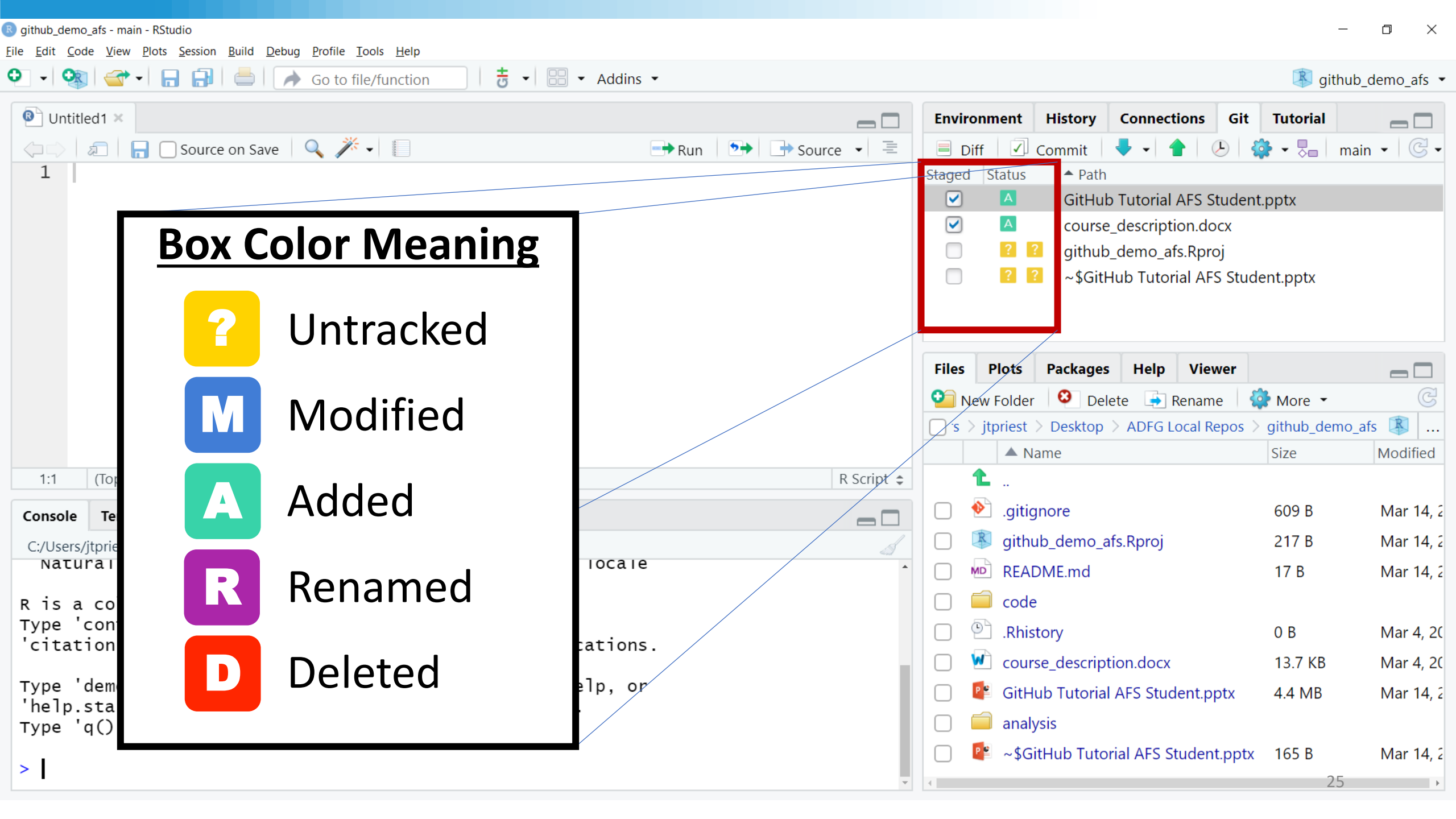
More

s > jtpriest > Desktop > ADFG Local Repos > github_demo_afs

Name Size Modified

..

☐ .gitignore 609 B Mar 14, 2☐ github_demo_afs.Rproj 217 B Mar 14, 2☐ README.md 17 B Mar 14, 2☐ code☐ .Rhistory 0 B Mar 4, 20☐ course_description.docx 13.7 KB Mar 4, 20☐ GitHub Tutorial AFS Student.pptx 4.4 MB Mar 14, 2☐ analysis☐ ~\$GitHub Tutorial AFS Student.pptx 165 B Mar 14, 2



Box Color Meaning



Untracked



Modified



Added



Renamed



Deleted

Environment

History

Connections

Git

Tutorial

Diff

Commit

main

Staged

Status

A

A

?

?

?

?

Path

GitHub Tutorial AFS Student.pptx

course_description.docx

github_demo_afs.Rproj

~\$GitHub Tutorial AFS Student.pptx

Files		Plots	Packages	Help	Viewer
New Folder		Delete	Rename	More	
s > jtpriest > Desktop > ADFG Local Repos > github_demo_afs					
		Name		Size	Modified
		..			
<input type="checkbox"/>		.gitignore		609 B	Mar 14, 2
<input type="checkbox"/>		github_demo_afs.Rproj		217 B	Mar 14, 2
<input type="checkbox"/>		README.md		17 B	Mar 14, 2
<input type="checkbox"/>		code			
<input type="checkbox"/>		.Rhistory		0 B	Mar 4, 20
<input type="checkbox"/>		course_description.docx		13.7 KB	Mar 4, 20
<input type="checkbox"/>		GitHub Tutorial AFS Student.pptx		4.4 MB	Mar 14, 2
<input type="checkbox"/>		analysis			
<input type="checkbox"/>		~\$GitHub Tutorial AFS Student.pptx		165 B	Mar 14, 2

Untitled1 x

Source on Save

Run

Source

Environment

History

Connections

Git

Tutorial

Diff

Commit

Path

Staged

Status

☒

A

A

GitHub Tutorial AFS Student.pptx

☒

A

A

course_description.docx

☐

?

?

github_demo_afs.Rproj

☐

?

?

~\$GitHub Tutorial AFS Student.pptx

Pull First!

Git Pull

Close

```
>>> C:/Users/jtpriest/AppData/Local/Programs/Git/cmd/git.exe pull
Already up to date.
```

1:1

Console

C:/Users/jt

NATUR

R is a

Type 'c

'citati

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.

> |

Files

Plots

Packages

Help

Viewer

New Folder

Delete

Rename

More

s > jtpriest > Desktop > ADFG Local Repos > github_demo_afs

Name

Size

Modified

☐

..

☐

.gitignore

609 B

Mar 14, 2

☐

github_demo_afs.Rproj

217 B

Mar 14, 2

☐

README.md

17 B

Mar 14, 2

☐

code

☐

.Rhistory

0 B

Mar 4, 20

☐

course_description.docx

13.7 KB

Mar 4, 20

☐

GitHub Tutorial AFS Student.pptx

4.4 MB

Mar 14, 2

☐

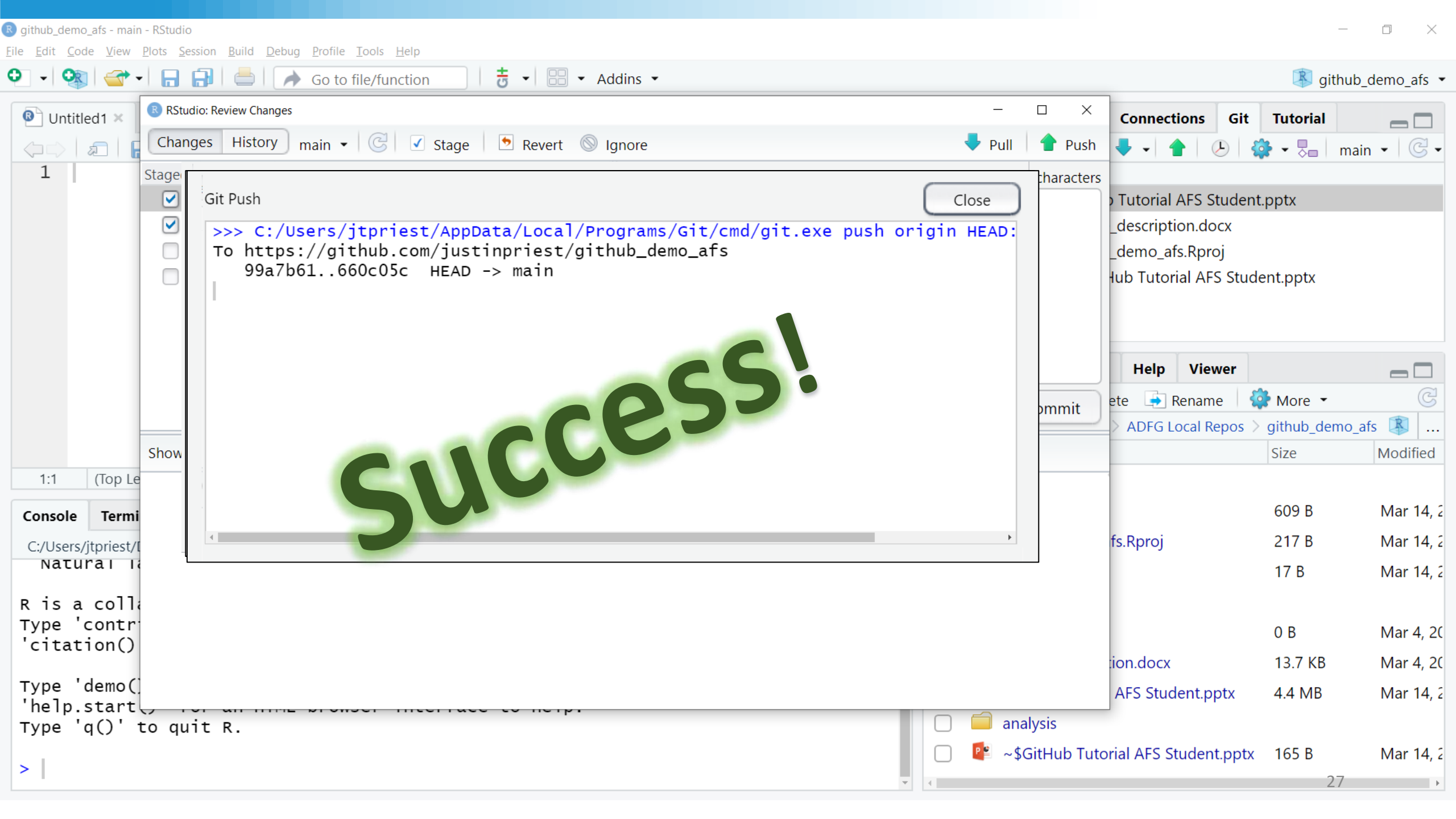
analysis

☐

~\$GitHub Tutorial AFS Student.pptx

165 B

Mar 14, 2



Success!

Git Push

Close

```
>>> C:/Users/jtpriest/AppData/Local/Programs/Git/cmd/git.exe push origin HEAD:
To https://github.com/justinpriest/github_demo_afs
99a7b61..660c05c HEAD -> main
```

Connections

Git

Tutorial

Pull

Push

main

Help

Viewer

Commit

Rename

More

ADFG Local Repos

github_demo_afs

Size

Modified

609 B

Mar 14, 2

217 B

Mar 14, 2

17 B

Mar 14, 2

0 B

Mar 4, 20

13.7 KB

Mar 4, 20

4.4 MB

Mar 14, 2

analysis

~\$GitHub Tutorial AFS Student.pptx

165 B

Mar 14, 2

27



3 – How to use GitHub: Commits

Created first draft of the trigger assessment RMD
master

justinpriest committed on Dec 6, 2020 1 parent d592030 commit 14fb5a26a6c30638

Browse files

Commit message

Showing 2 changed files with 574 additions and 2 deletions.

code/3_US_figures.R

```
@@ -14,6 +14,7 @@ source(here::here("code/1_US_data_import.R"))
14 14 troll_cpue <- read_csv(here::here("data/SEAK_Coho_TrollFPD_1981-2019.csv"),
15 15                       guess_max = 84000) %>% #increased guess
16 16   rename("Gear" = `Gear Code`,
17 17 +     "TripNum" = `Trip No`,
17 18     "SellDate" = `Sell Date`,
18 19     "StatWeek" = `Stat Week`,
19 20     "TrollArea" = `Troll Area`,
@@ -25,8 +26,10 @@ troll_cpue <- read_csv(here::here("data/SEAK_Coho_TrollFPD_1981-2019.csv"),
25 26   mutate(SellDate = as_date(as.POSIXct(SellDate, format = "%m/%d/%Y", tz = "US/Alaska")),
26 27     District = as.factor(District),
27 28     Effort_boatdays = DaysFished * HoursPerDay / 13, # Effort is standardized to a 13 hour
28 -     CohoCPUE = CohoCatch / Effort_boatdays) %>%
29 -   dplyr::select(Year, SellDate, StatWeek, TrollArea, District, StatArea, CohoCatch, Effort_boatdays)
29 +     CohoCPUE = CohoCatch / Effort_boatdays,
30 +     TripNumber = paste0(Year, "-", TripNum)) %>%
31 +   dplyr::select(Year, TripNumber, SellDate, StatWeek, TrollArea, District,
32 +     StatArea, CohoCatch, Effort_boatdays, CohoCPUE)
```

Added text (green)

Deleted text (red)



3 – How to use GitHub: Private data

How to deal with confidential data

- Private repository (only invited individuals can see it)
- Summarize data
- Add to gitignore file
 - Don't upload and add a readme file to the data folder ("Confidential data is not stored publicly. Contact someone@alaska.gov for access")

```
preseason_SEAK_pink_salmon_forecast - master - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
ist2022.R functions.R .gitignore .gitignore .gitignore 2_age12_models.R pinkforecast202

1 # specific pdf and csv file
2 Taku*
3 figs
4 *.csv
5 *.pdf
6
7 # History files
8 .Rhistory
9 .Rapp.history
10
11 # Session Data files
12 .RData
13
14 # User-specific files
15 .Ruserdata
16
17 # Example code in package build process
18 *-Ex.R
19
20 # Output files from R CMD build
21 /*.tar.gz
22
23 # Output files from R CMD check
24 /*.Rcheck/
25
26 # RStudio files
27 .Rproj.user/
28 *.Rproj
```




4 – Show Me the Demo!



4 – Demo 1: Creating a repo

1

Follow along as I share my screen



4 – Demo 2: Cloning an existing repo

2

Follow along as I share my screen



4 – Demo 3: Collaboration!

3

Follow along as I share my screen



4 – Demo 4: Issues

The screenshot shows the GitHub repository page for `commfish/seak_sablefish`. The repository has 4 watches, 6 stars, and 3 forks. The main content area displays a file tree with the following structure:

File/Folder	Description	Last Updated
data	final updates to 2020 forecast	9 months ago
figures	final updates to 2020 forecast	9 months ago
groundfish_statistical_areas_2001	groundfish stat area shapefiles	4 years ago
output	final updates to 2020 forecast	9 months ago
r	Update 0_queryndean_data.R	2 months ago
text	Was behind on track changes, big update for 2020 forecast	10 months ago
tmb	Update folder structure	10 months ago
.gitignore	update gitignore	15 months ago
README.md	Update README.md	8 months ago

The README.md file is displayed below the file tree. It contains the following text:

Northern Southeast Inside Waters (NSEI) sablefish (*Anoplopoma fimbria*) stock assessment

Please direct any questions to: Jane Sullivan (jane.sullivan1@alaska.gov, ummjane@gmail.com) or Rhea Ehresmann (rhea.ehresmann@alaska.gov)

Last updated: June 2020

The right sidebar shows the repository's metadata:

- About:** NSEI sablefish stock assessment
- Releases:** 2 releases, including "2020 forecast" (Latest) on Jun 22, 2020.
- Packages:** No packages published.
- Contributors:** 2 contributors: jysullivan (Jane Sullivan) and ben-williams (Ben Williams).
- Languages:** R (88.7%) and C++ (11.3%).

https://github.com/commfish/seak_sablefish



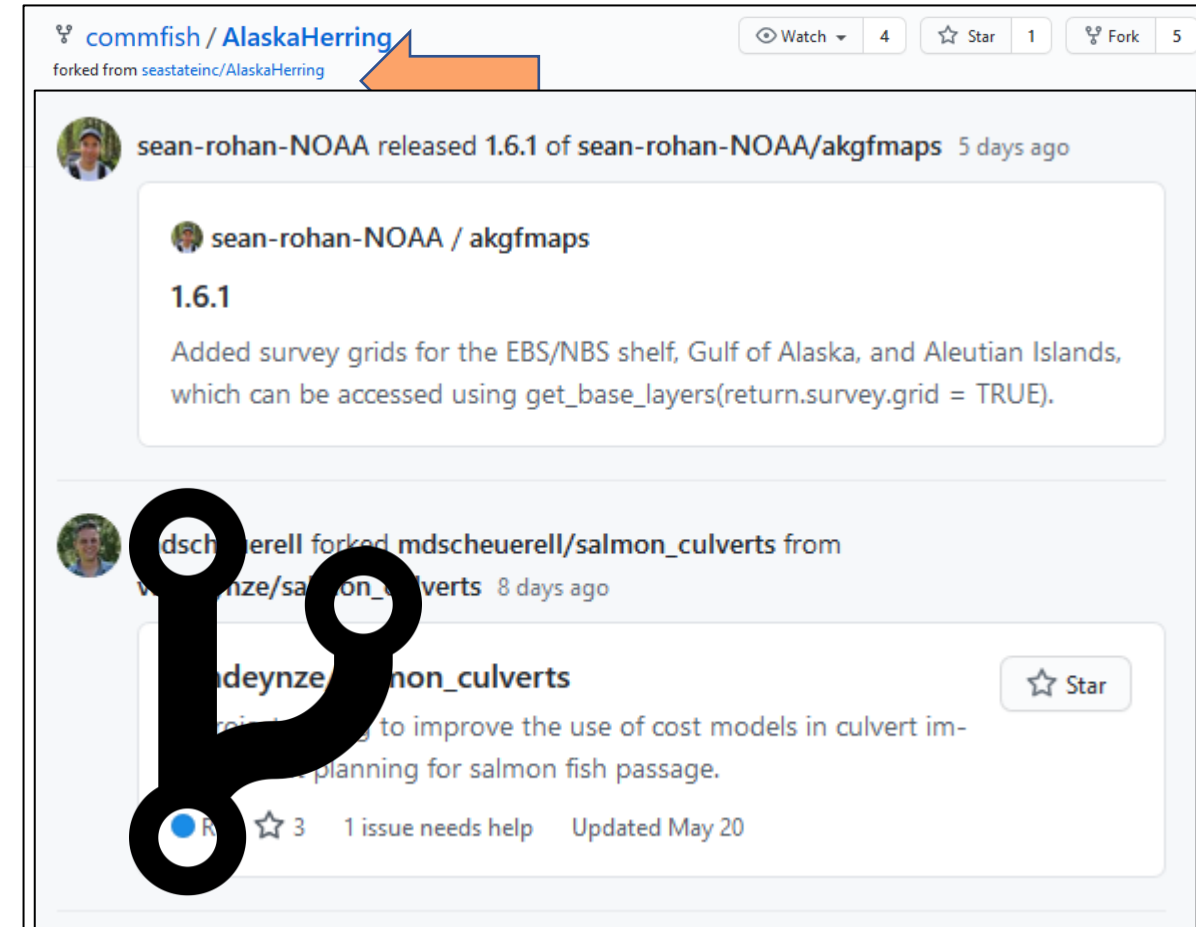
5 – Concluding Thoughts / Best Practices



5 – What else can GitHub do?

We've barely shown the capabilities of GitHub

- Issues can be assigned to people
- “Fork” an existing repo to modify it to your purposes
- Create a project website
- Build and store an R package
- Follow / learn others' code
- Add a “branch” to explore test features without modifying the main analysis





5 – Best Practices: CSVs

- Git works best with “raw” or script-based files: R scripts, RMD presentations, CSV files, etc.
- It works with MS Word, PowerPoint, .XLSX files, but difficult to compare changes
- Takeaway: If you can, use CSV files!



VS





5 – Best Practices: Reading Data

It's common to manipulate your data in Excel, save this file, then import/analyze in R



It's MUCH better to do all cleanup in R

- Saves steps for cleanup that are performed each time





5 – Best Practices: Naming

- If possible, don't use spaces, mixed cases, or periods in filenames or directories
- Name data and R files descriptively
 - If you are **changing** files, then adding a date can be beneficial
 - Be consistent across projects
- Use variables / column names that are:
 - Easy to understand
 - Short yet descriptive

Example:

X10_20.16_T_AND_Y410.csv
versus
taku_sockeye_age_comp.csv

Example:

sablefish_chathamsurvey_bio_1988_2016.csv
secm_survey_cpue_1988_2016.csv

Example:

year, catch instead of c
cohocatch2016

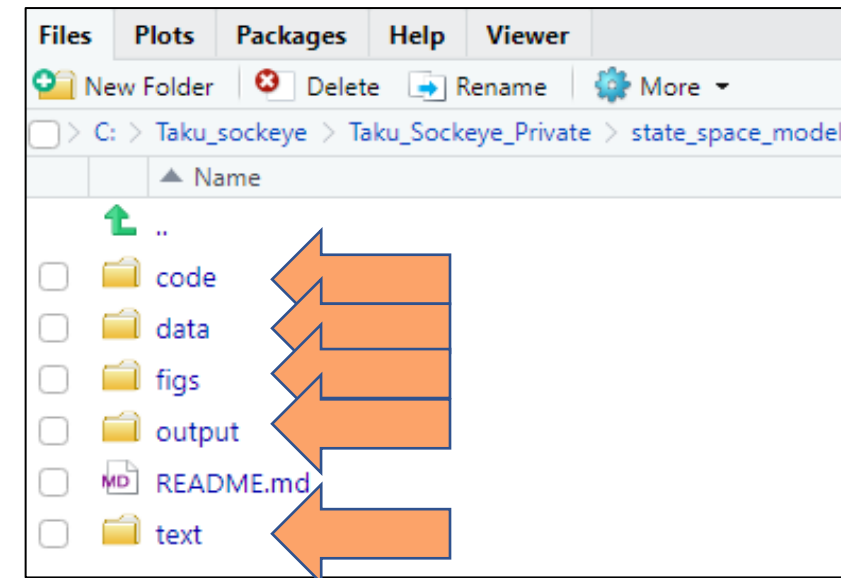
Source: [ADF&G Biometrics Best Practices](#)



5 – Best Practices: Folders

Folder structure

- code
 - All scripts go here (use separate folders for R code, admb, etc.)
- data
 - Use raw data treat as a read only file; never adjust the contents of this file, access/modify via code only
 - Create a metadata text file describing all fields within the data files
- figs
 - Output charts as .png here
- output
 - results of analyses
- text
 - if writing a document it is easier to keep separate



Source: [ADF&G Biometrics Best Practices](#)



5 – Best Practices: Other

- Avoid working in the same file as collaborators!
- Update R versions and packages often
 - Unless you're about to publish
- Break up data import/cleanup & analysis scripts
 - Use several scripts that “link” together rather than a long script
- Comment often and be VERY thorough
- Follow people on GitHub / Twitter/Facebook to read other code and learn better coding practices
- Read in data using `read.csv()` or `read_csv()`. Never use `file.choose()`, never use the “Import Dataset” wizard in RStudio. Keep file paths relative and not static

Name	Date modified
R 1_setup_data	3/16/2021 2:35 PM
R 2_age12_models	3/18/2021 6:49 PM
R 3_age13_models	3/18/2021 6:53 PM
R 4_age03_models	3/18/2021 8:34 PM
R 5_age02_models	3/18/2021 6:48 PM
R 6_summarize_data	3/18/2021 7:45 PM
R functions	3/17/2021 8:12 PM



5 – Recent Update

If you are already a GitHub user and got this email:



- In RStudio: Update packages “usethis”, “gh”, and “gitcreds”
- In Email: Click “Regenerate your personal access token”
- On GitHub.com: Click Regenerate token, copy text on next screen! (starts with “ghp_”)
- In RStudio: Run “gitcreds::gitcreds_set()”, select “2: Replace these credentials”, paste in new token text
- In RStudio: Run “usethis::gh_token_help()” to double check that everything looks good

Hi @justinpriest,

We noticed your personal access token, git: <https://github.com/> on DFGDOUDCF115007 at 15-Aug-2019 12:24, has an outdated format and was used to access the GitHub API on May 23rd, 2021 at 00:35 (UTC) with a user-agent header of git-credential-manager (Microsoft Windows NT 6.2.9200.0; Win32NT; x64) CLR/4.0.30319 git-tools/[1.19.0](#).

We recently updated the format of our API authentication tokens, providing additional security benefits to all our customers.

In order to benefit from this new format, please regenerate your personal access token, git: <https://github.com/> on DFGDOUDCF115007 at 15-Aug-2019 12:24, using the button below.

Regenerate your personal access token



Remember!

Pull -> Do your work -> Pull -> Commit -> Push



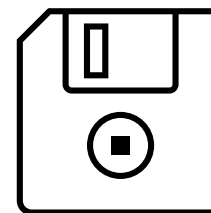
Pull

Work



Pull

Commit



Push



Additional Resources

- [GitHub course for new users](#)
- [Library “Usethis” Config and Setup for GitHub](#)
- [Happy Git with R](#)
- [GitHub for Advanced Ecological Data Analysis](#)
- [R for Excel Users – Github brief intro](#)



Troubleshooting

What if I have an existing RProject that I want to push to GitHub?

- Technically it's possible to push an existing RProject to GitHub, but it gets tricky (requiring command line git). It's simplest to create a new RProject, copy over files, and delete the old project.

I can't push to GitHub!

- Pull first. A collaborator probably made edits to the remote (GitHub.com) version and you need to grab that before you make any edits

How do I add a collaborator?

- Go to the repo page on github.com and click settings -> Manage Access. Now they can push code to this same repo! Remember to work in the file separately and push/pull often!!



Questions? Contact us!



sara.miller@alaska.gov

justin.priest@alaska.gov