

Version control using git in RStudio

Tobin Magle, PhD

Lead data management specialist
Research Computing and Data Services

Today we're going to cover:

- What is **version control**?
- What is **git**?
- Version control using **RStudio**
- Backing up your data to **GitHub**



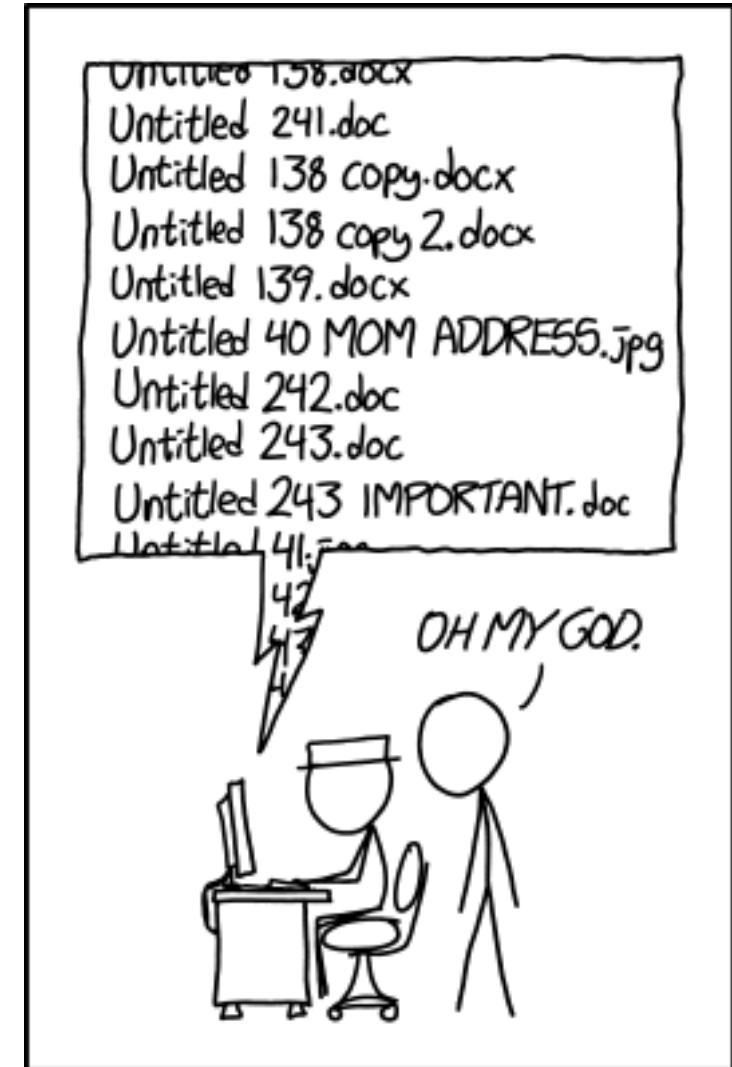
Pre-workshop Setup

- Install R installed
- Install RStudio
- Install git
- Created a GitHub account

Manual Version Control

“Renaming files”

- Creates many files
- Relies on your ability to name files consistently
- Not designed for collaboration



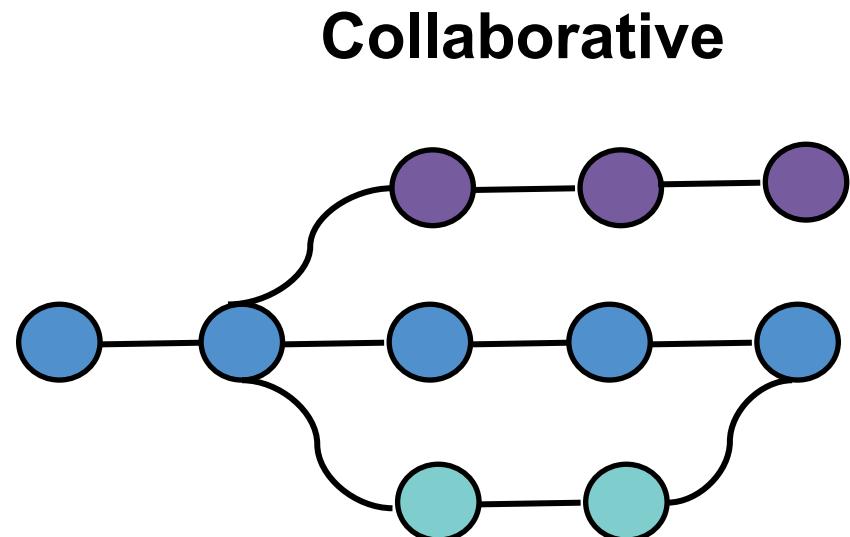
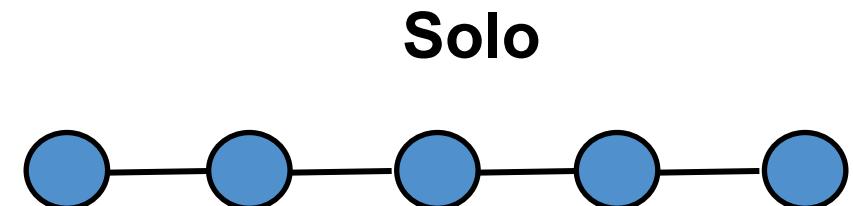
PROTIP: NEVER LOOK IN SOMEONE ELSE'S DOCUMENTS FOLDER.

<https://xkcd.com/1459/>

Version Control Systems

“Unlimited undo”

- One file, many versions
- Records who made what change when
- Good for **collaboration**: identifies and helps resolve conflicting changes



Today we're using git

- Built for large software development projects
- Can get complicated, but the basics are simple
- Free, open source
- Many online resources/tutorials

THIS IS GIT. IT TRACKS COLLABORATIVE WORK ON PROJECTS THROUGH A BEAUTIFUL DISTRIBUTED GRAPH THEORY TREE MODEL.

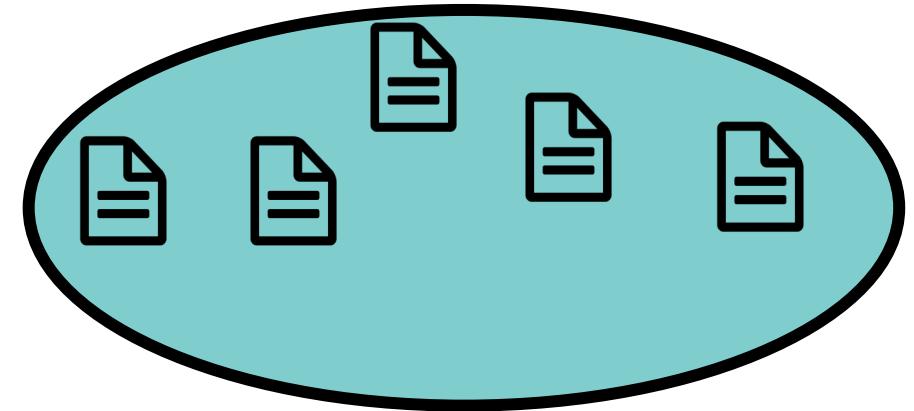
COOL. HOW DO WE USE IT?

NO IDEA. JUST MEMORIZIZE THESE SHELL COMMANDS AND TYPE THEM TO SYNC UP. IF YOU GET ERRORS, SAVE YOUR WORK ELSEWHERE, DELETE THE PROJECT, AND DOWNLOAD A FRESH COPY.



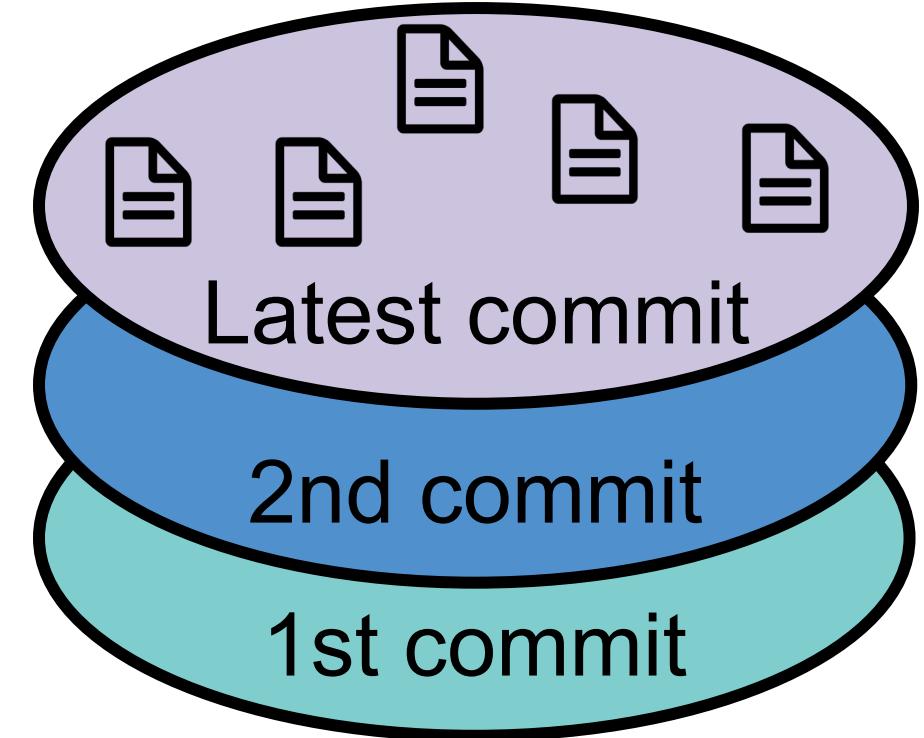
Using git creates a repository

- **Repository** – A folder that is tracked by git



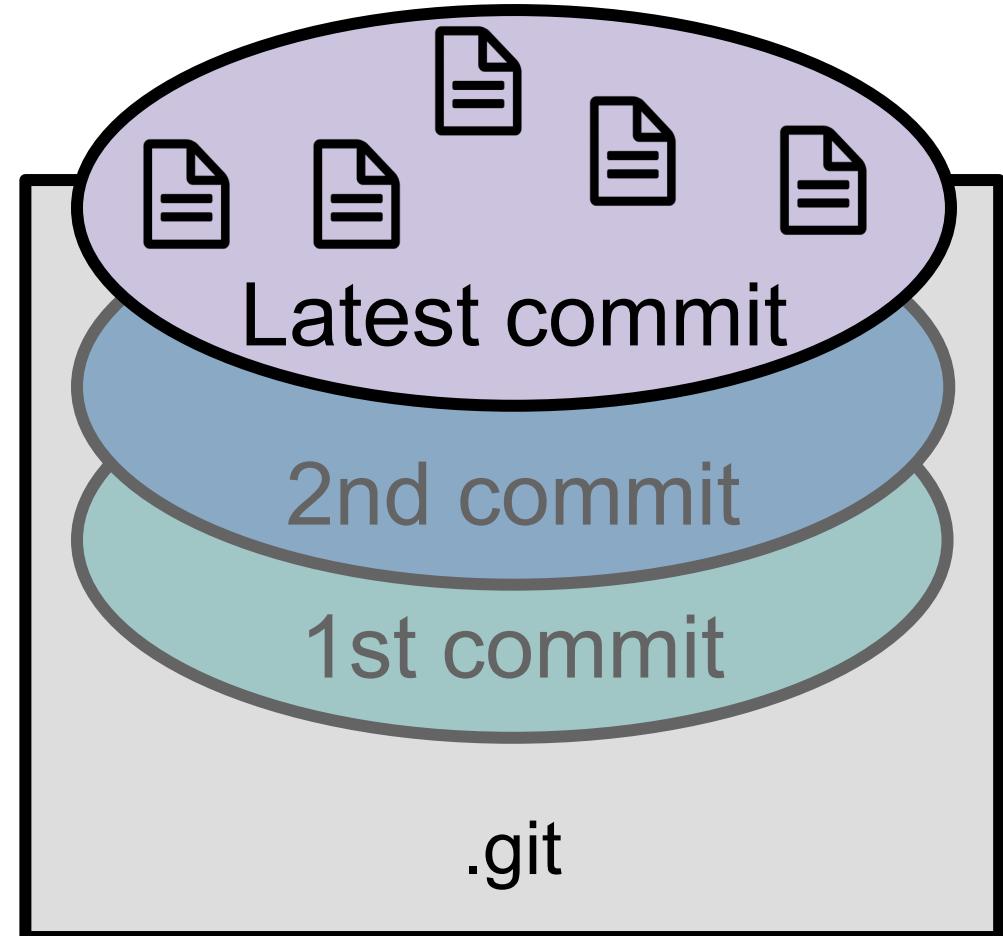
Repositories are made of commits

- **Repository** – A folder that is tracked by git
- **Commit** – a version of what the folder looks like when it was created



You only see your latest commit

- **Repository** – A folder that is tracked by git
- **Commit** – a version of what the folder looks like when it was created
- **Structure is hidden** (.git), you see the most recent version



Many ways to use Git

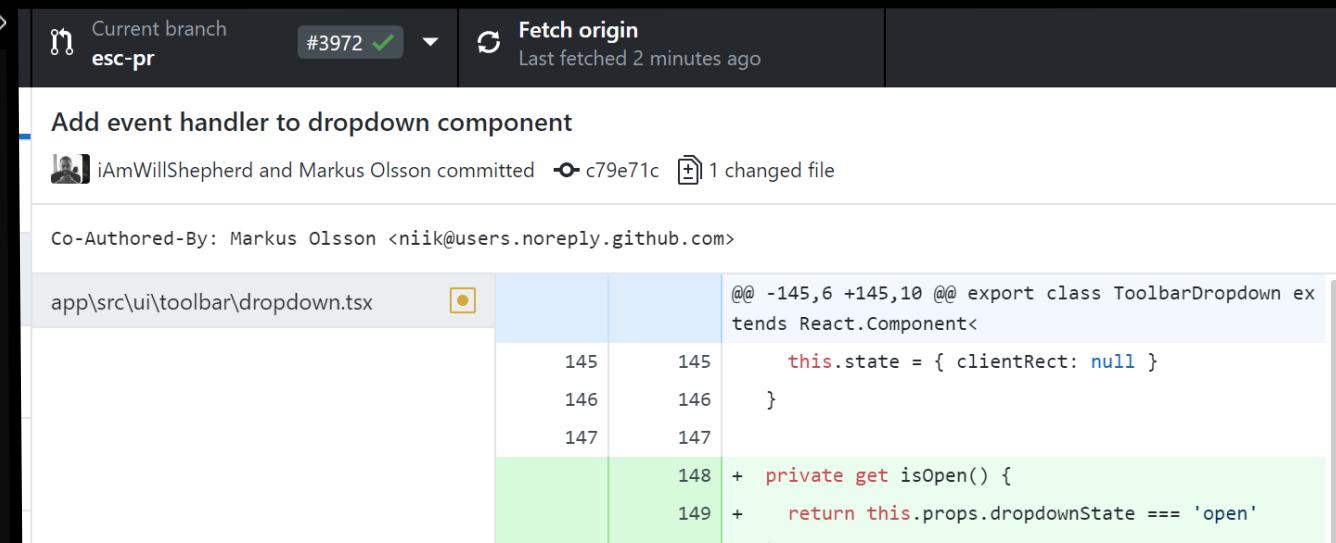
Command line

```
C:\Windows\System32\cmd.exe
D:\GFG>git status
On branch master
Initial commit

Changes to be committed:
(use "git rm --cached <file>..." to unstage)

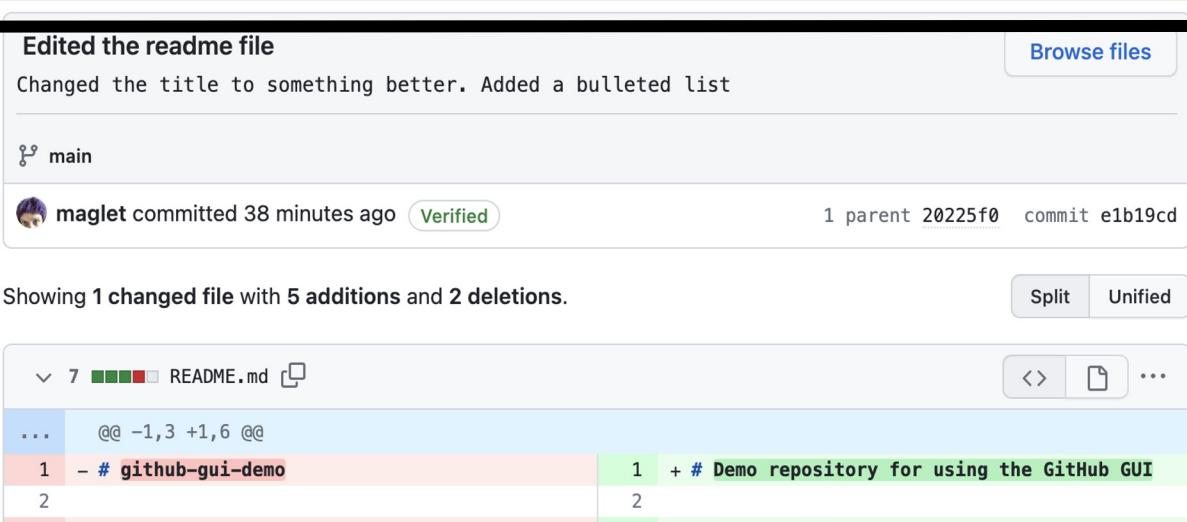
    new file: Basic Code/C++ Project.cbproj
    new file: Basic Code/README.md
    new file: Basic Code/bin/Debug/C++ Project.exe
    new file: Basic Code/data.in
    new file: Basic Code/data.out
    new file: Basic Code/main.cpp
    new file: Basic Code/obj/Debug/main.o
    new file: DS/stackPar.cpp
    new file: DS/stackPar.exe
    new file: DS/stackPar.o
    new file: DS/stlsort.cpp
```

Desktop Clients



The screenshot shows a GitHub desktop application interface. At the top, it displays the current branch as 'esc-pr' and the fact that 'Fetch origin' was last fetched 2 minutes ago. Below this, a pull request titled 'Add event handler to dropdown component' is shown. It includes a commit by 'iAmWillShepherd and Markus Olsson' with a commit hash 'c79e71c'. The commit message says 'Co-Authored-By: Markus Olsson <niik@users.noreply.github.com>'. A detailed code diff is visible, comparing two versions of the file 'app\src\ui\toolbar\dropdown.tsx'. The diff highlights changes from line 145 to 149, with some lines being added (green) and others deleted (red). The code itself is written in TypeScript.

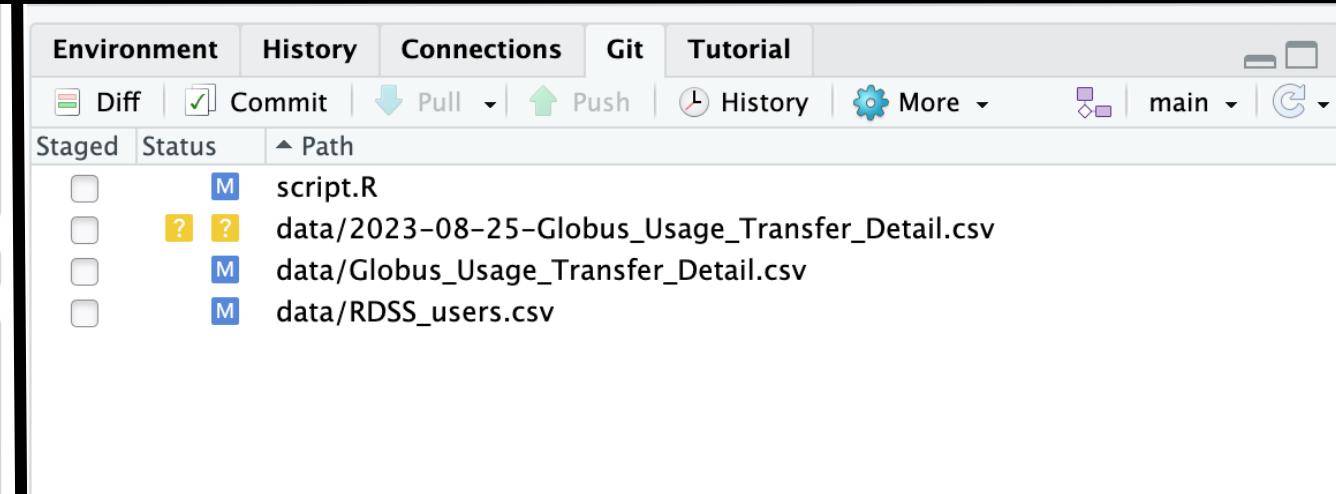
GitHub Website



The screenshot shows a GitHub repository page. At the top, it says 'Edited the readme file' and 'Changed the title to something better. Added a bulleted list'. Below this, a commit by 'maglet' is shown, which was committed 38 minutes ago and is verified. The commit hash is 'e1b19cd'. The commit message indicates there were 5 additions and 2 deletions. A detailed code diff is provided for the file 'README.md', showing the following changes:

```
@@ -1,3 +1,6 @@
1 - # github-gui-demo
2 + # Demo repository for using the GitHub GUI
```

RStudio IDE



The screenshot shows the RStudio IDE with the 'Git' tab selected in the top navigation bar. The 'Environment', 'History', 'Connections', 'Git', and 'Tutorial' tabs are visible. Below the tabs, there are buttons for 'Diff', 'Commit', 'Pull', 'Push', 'History', and 'More'. The 'Staged' tab is selected, showing a list of files with their status (M for modified):

- M script.R
- ?? data/2023-08-25-Globus_Usage_Transfer_Detail.csv
- M data/Globus_Usage_Transfer_Detail.csv
- M data/RDSS_users.csv

Many ways to use Git

Command line

```
C:\Windows\System32\cmd.exe
D:\GEG>git status
On branch master
Initial commit

Changes to be committed:
(use "git rm --cached <file>..." to unstage)

    new file:  Basic Code/C++ Project.cbproj
    new file:  Basic Code/README.md
    new file:  Basic Code/bin/Debug/C++ Project.exe
    new file:  Basic Code/data.in
    new file:  Basic Code/data.out
    new file:  Basic Code/main.cpp
    new file:  Basic Code/obj/Debug/main.o
    new file:  DS/stackPar.cpp
    new file:  DS/stackPar.exe
    new file:  DS/stackPar.o
    new file:  DS/stlsort.cpp
```

Desktop Clients

```
Current branch esc-pr #3972 ✓
Fetch origin
Last fetched 2 minutes ago

Add event handler to dropdown component
iAmWillShepherd and Markus Olsson committed c79e71c 1 changed file
Co-Authored-By: Markus Olsson <niik@users.noreply.github.com>

app\src\ui\toolbar\dropdown.tsx
@@ -145,6 +145,10 @@ export class ToolbarDropdown extends React.Component<
    this.state = { clientRect: null }
}

145 145
146 146
147 147
148 + private get isOpen() {
149 +     return this.props.dropdownState === 'open'
```

GitHub Website

Edited the readme file

Browse files

Changed the title to something better. Added a bulleted list

main

maglet committed 38 minutes ago (Verified)

1 parent 20225f0 commit e1b19cd

Showing 1 changed file with 5 additions and 2 deletions.

Split Unified

README.md

@@ -1,3 +1,6 @@
1 - # github-gui-demo
2 + # Demo repository for using the GitHub GUI

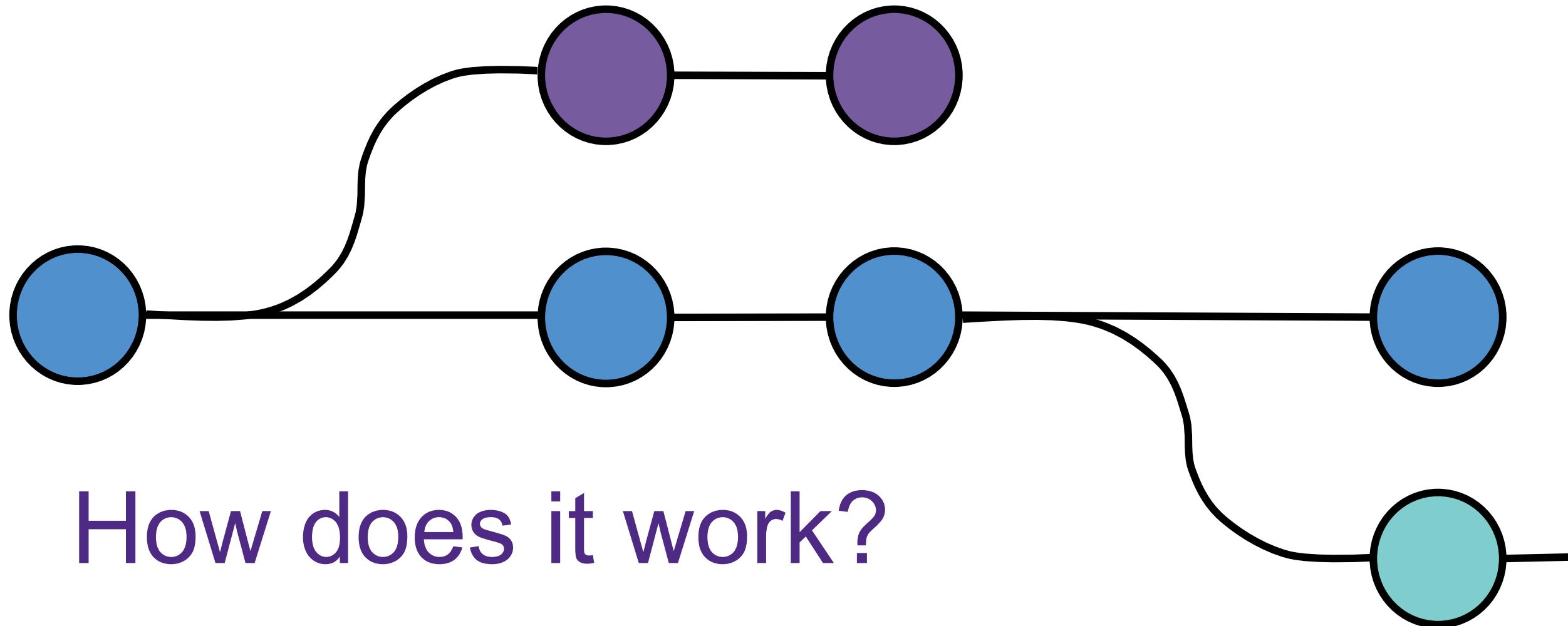
RStudio IDE

Environment History Connections Git Tutorial

Diff Commit Pull Push History More main C

Staged Status Path

- M script.R
- ?? data/2023-08-25-Globus_Usage_Transfer_Detail.csv
- M data/Globus_Usage_Transfer_Detail.csv
- M data/RDSS_users.csv



How does it work?

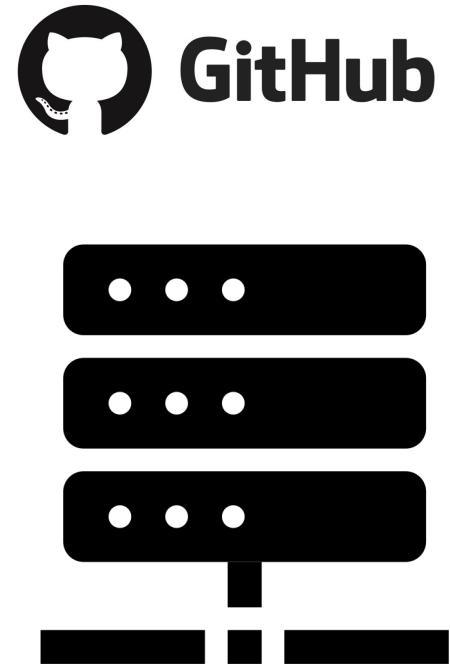
Coding happens on your computer



- Create and edit files in RStudio
- Git is also on your computer
- Use git version control in the RStudio interface
- Creates a local repository

GitHub stores code in the cloud

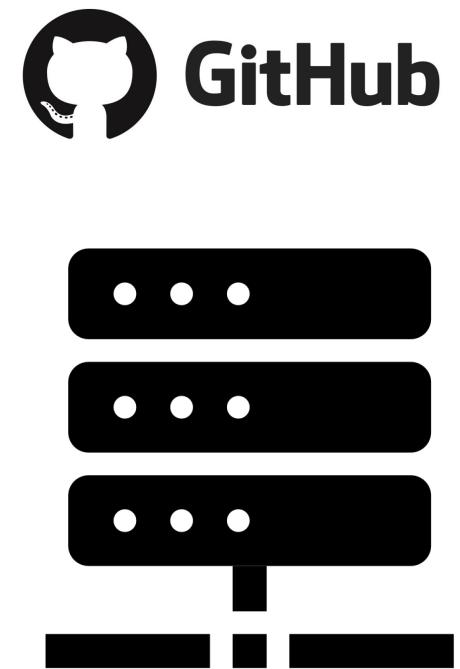
- Back up your work
- Collaborate with colleagues
- Share your code



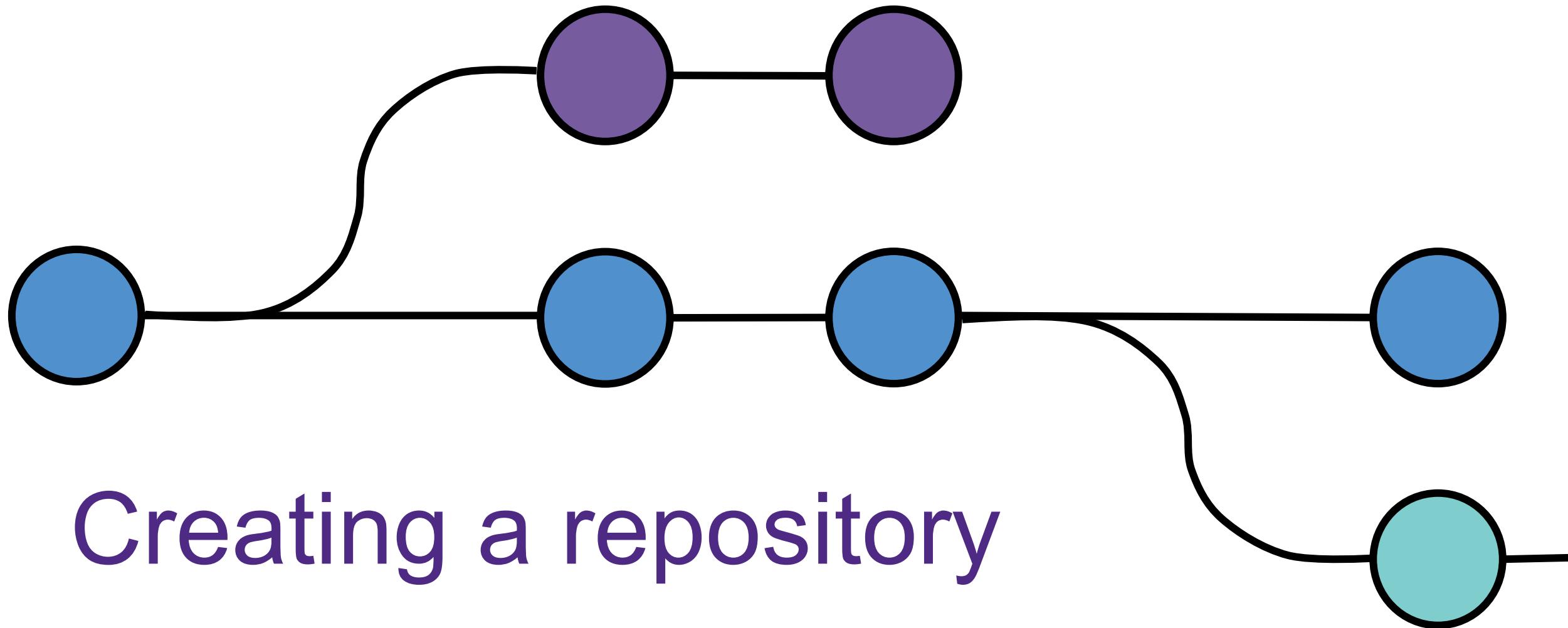
Overall workflow



Local repository
(on your computer)

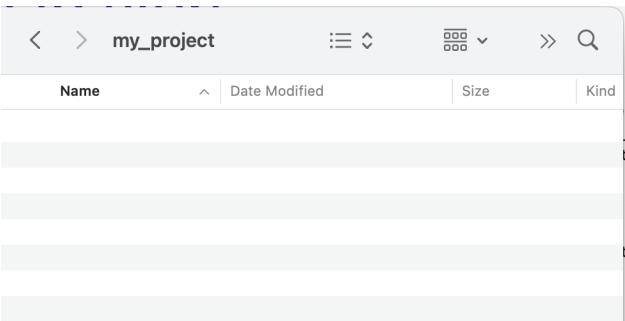


Remote repository
(on the web)



A few ways to start

From nothing



From a local repo

```
(base) EDMRCSLFYLQNWQ0KJ:globus-stats tobinmagle$ ls -a
.
..
.DS_Store
.RData
.Rhistory
.Rproj.user
.git
.gitignore
(base) EDMRCSLFYLQNWQ0KJ:globus-stats tobinmagle$
```

From a GitHub repo

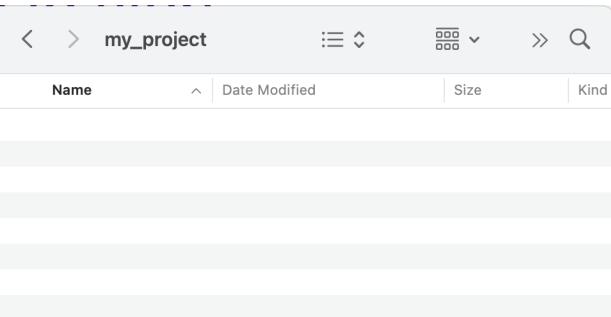
A screenshot of a GitHub repository page for 'github-gui-demo'. The page shows basic repository statistics (1 branch, 0 tags), a code editor, and a list of recent commits. One commit by 'maglet' is highlighted, showing they added an 'edited slide pdf'.

From files outside a repo

| Name | Date Modified | Size | Kind |
|--------------------------------|-------------------------|-----------|-----------------|
| 2023-01-10-merge-quota-used.R | Jan 10, 2023 at 2:04 PM | 4 KB | R Source File |
| data-cleaning.R | Jun 2, 2023 at 5:22 PM | 4 KB | R Source File |
| Lifesciences-usage-report.pdf | Aug 4, 2023 at 10:10 AM | 108 KB | PDF Document |
| Lifesciences-usage-report.Rmd | Jun 9, 2023 at 3:27 PM | 3 KB | R Markdown File |
| number-RDSS-users.R | Today at 12:11 PM | 114 bytes | R Source File |
| RDSS-annual-billing-audit.html | Jun 9, 2023 at 3:55 PM | 625 KB | HTML text |
| RDSS-annual-billing-audit.Rmd | Jun 9, 2023 at 3:55 PM | 2 KB | R Markdown File |
| RDSS-quarterly-audit.html | Jun 9, 2023 at 3:50 PM | 625 KB | HTML text |
| RDSS-quarterly-audit.Rmd | Jun 9, 2023 at 3:50 PM | 2 KB | R Markdown File |
| RDSS-SESP-usage-report.pdf | Aug 4, 2023 at 10:08 AM | 109 KB | PDF Document |
| RDSS-SESP-usage-report.Rmd | Aug 4, 2023 at 1:19 PM | 2 KB | R Markdown File |
| RDSS-usage-report.Rmd | Jun 2, 2023 at 5:22 PM | 9 KB | R Markdown File |
| render_reports.R | Jun 9, 2023 at 3:11 PM | 494 bytes | R Source File |
| Weinberg-usage-report.Rmd | Mar 7, 2023 at 9:55 PM | 2 KB | R Markdown File |

A few ways to start

From nothing



From a local repo

```
(base) EDMRCSLFYLQNWQ0KJ:globus-stats tobinmagle$ ls -a
.
..
.DS_Store
.RData
.Rhistory
.Rproj.user
.git
.gitignore
(base) EDMRCSLFYLQNWQ0KJ:globus-stats tobinmagle$
```

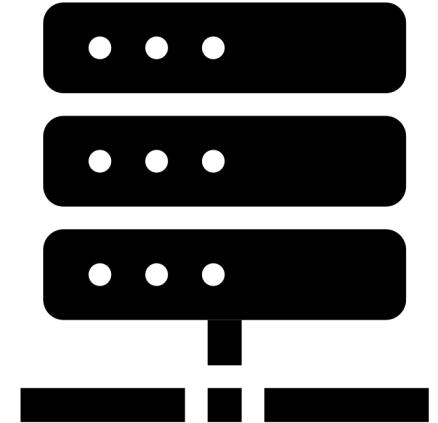
From a GitHub repo

A screenshot of a GitHub repository page for 'github-gui-demo'. The page shows basic repository statistics (1 branch, 0 tags), a code tab, and a list of recent commits. One commit by 'maglet' is highlighted, showing they added an 'edited slide pdf'.

From files outside a repo

| Name | Date Modified | Size | Kind |
|--------------------------------|-------------------------|-----------|-----------------|
| 2023-01-10-merge-quota-used.R | Jan 10, 2023 at 2:04 PM | 4 KB | R Source File |
| data-cleaning.R | Jun 2, 2023 at 5:22 PM | 4 KB | R Source File |
| Lifesciences-usage-report.pdf | Aug 4, 2023 at 10:10 AM | 108 KB | PDF Document |
| Lifesciences-usage-report.Rmd | Jun 9, 2023 at 3:27 PM | 3 KB | R Markdown File |
| number-RDSS-users.R | Today at 12:11 PM | 114 bytes | R Source File |
| RDSS-annual-billing-audit.html | Jun 9, 2023 at 3:55 PM | 625 KB | HTML text |
| RDSS-annual-billing-audit.Rmd | Jun 9, 2023 at 3:55 PM | 2 KB | R Markdown File |
| RDSS-quarterly-audit.html | Jun 9, 2023 at 3:50 PM | 625 KB | HTML text |
| RDSS-quarterly-audit.Rmd | Jun 9, 2023 at 3:50 PM | 2 KB | R Markdown File |
| RDSS-SESP-usage-report.pdf | Aug 4, 2023 at 10:08 AM | 109 KB | PDF Document |
| RDSS-SESP-usage-report.Rmd | Aug 4, 2023 at 1:19 PM | 2 KB | R Markdown File |
| RDSS-usage-report.Rmd | Jun 2, 2023 at 5:22 PM | 9 KB | R Markdown File |
| render_reports.R | Jun 9, 2023 at 3:11 PM | 494 bytes | R Source File |
| Weinberg-usage-report.Rmd | Mar 7, 2023 at 9:55 PM | 2 KB | R Markdown File |

Start in GitHub



Remote

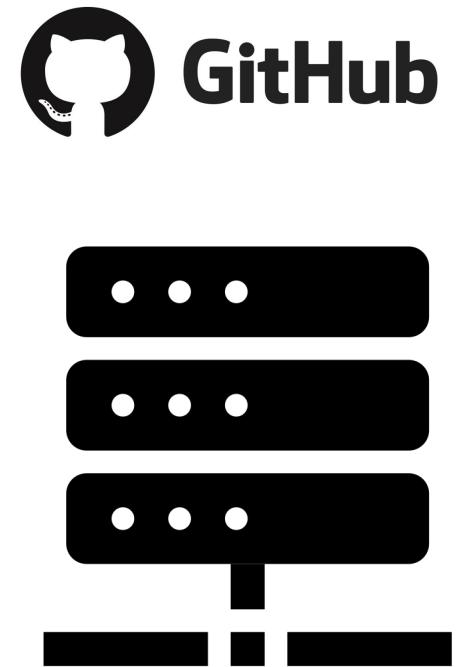
Workflow



Local

Clone

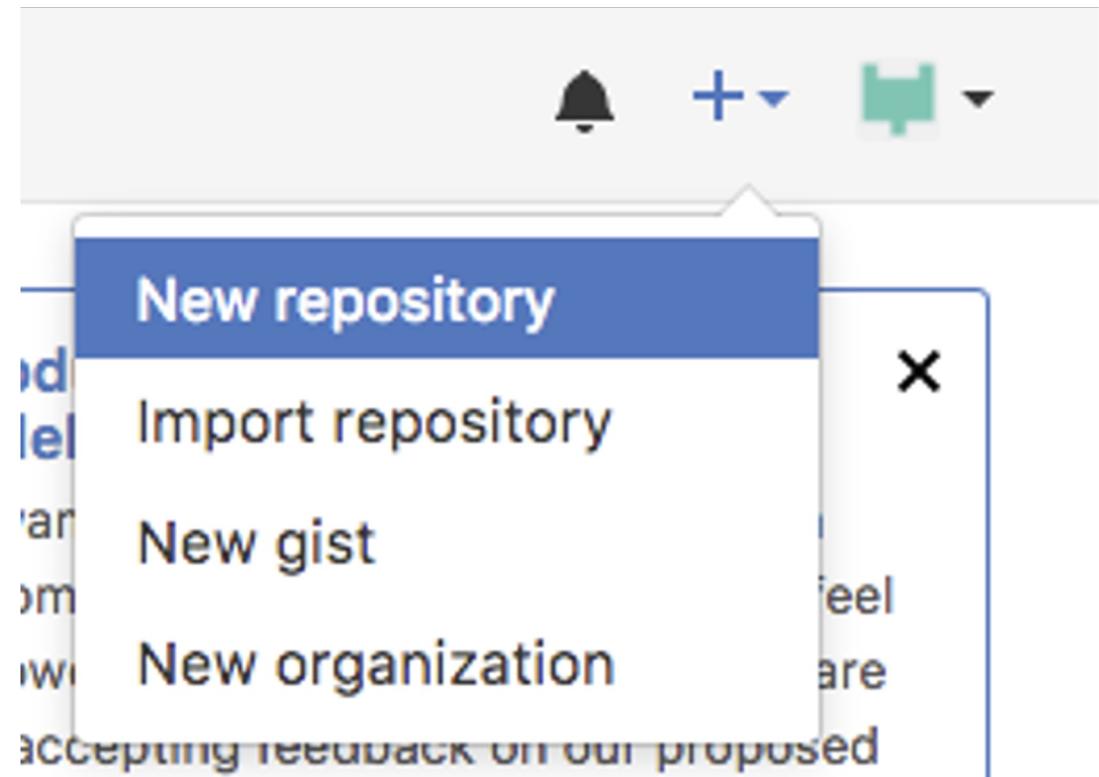
- Establish a connection
- Download remote to local



Remote

GitHub: Make an empty GitHub repo

- Go to <http://github.com>
- Click the plus sign on the upper right-hand side of the screen
- Select New repository



Github: Make an empty repository

- Name your repository
- Provide a description
- Keep it public
- Initialize with a readme

Create a new repository
A repository contains all project files, including the revision history.

Owner /
Great repository names are short and memorable. Need inspiration? How about [miniature-octo-broccoli](#)?

Description (optional)

 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 ▾ 

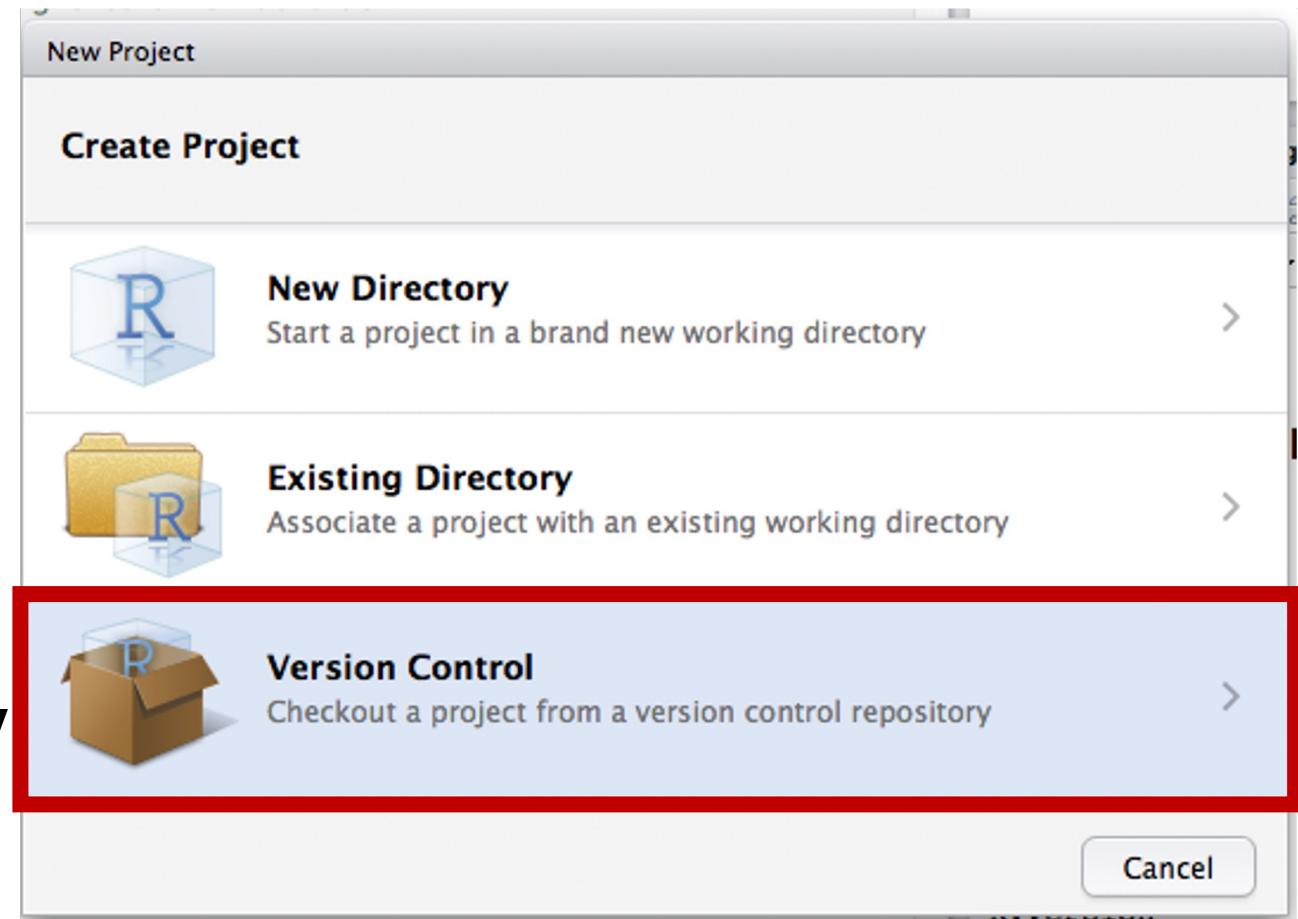
Create repository

RStudio: Make a new project

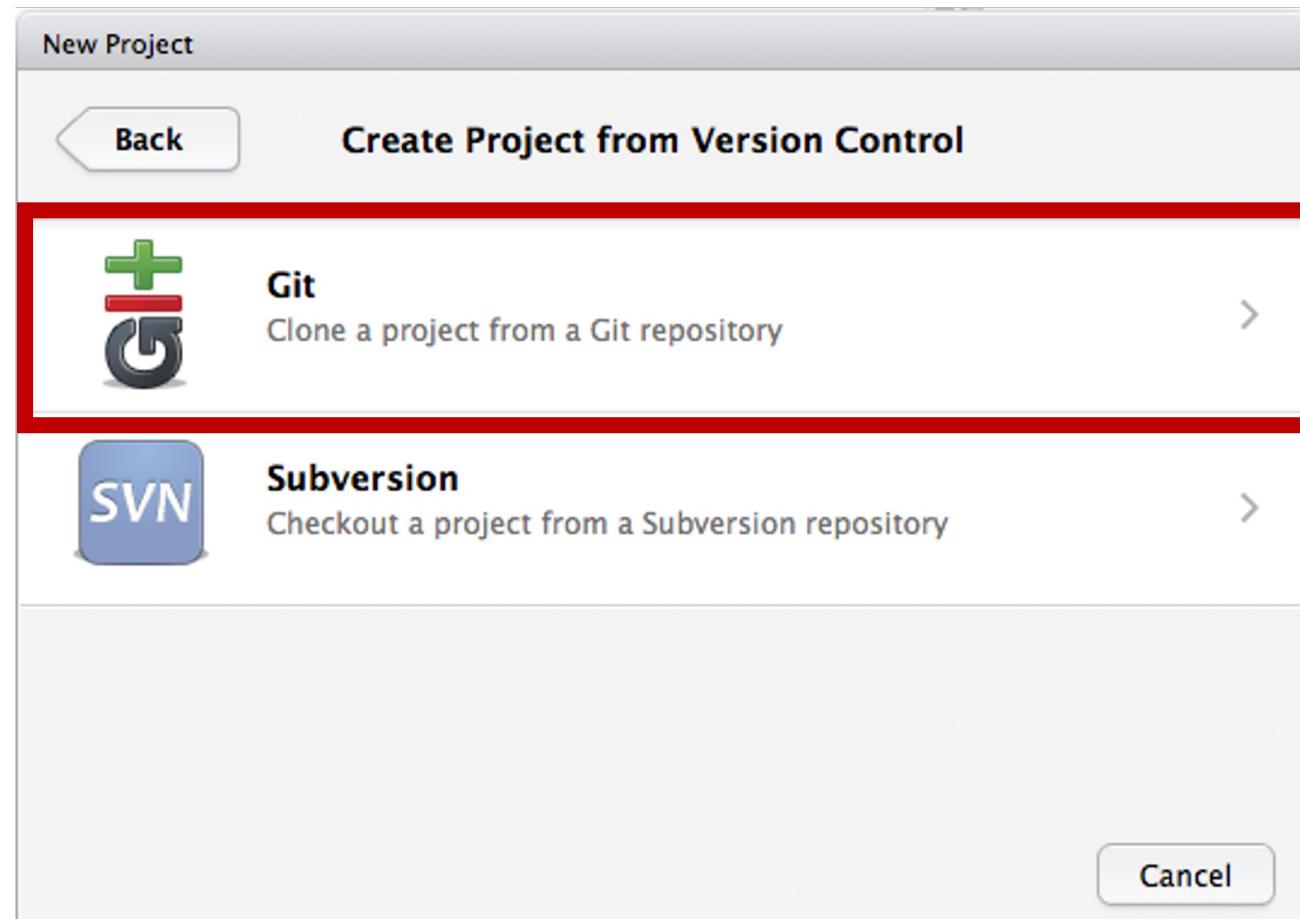
File > New Project



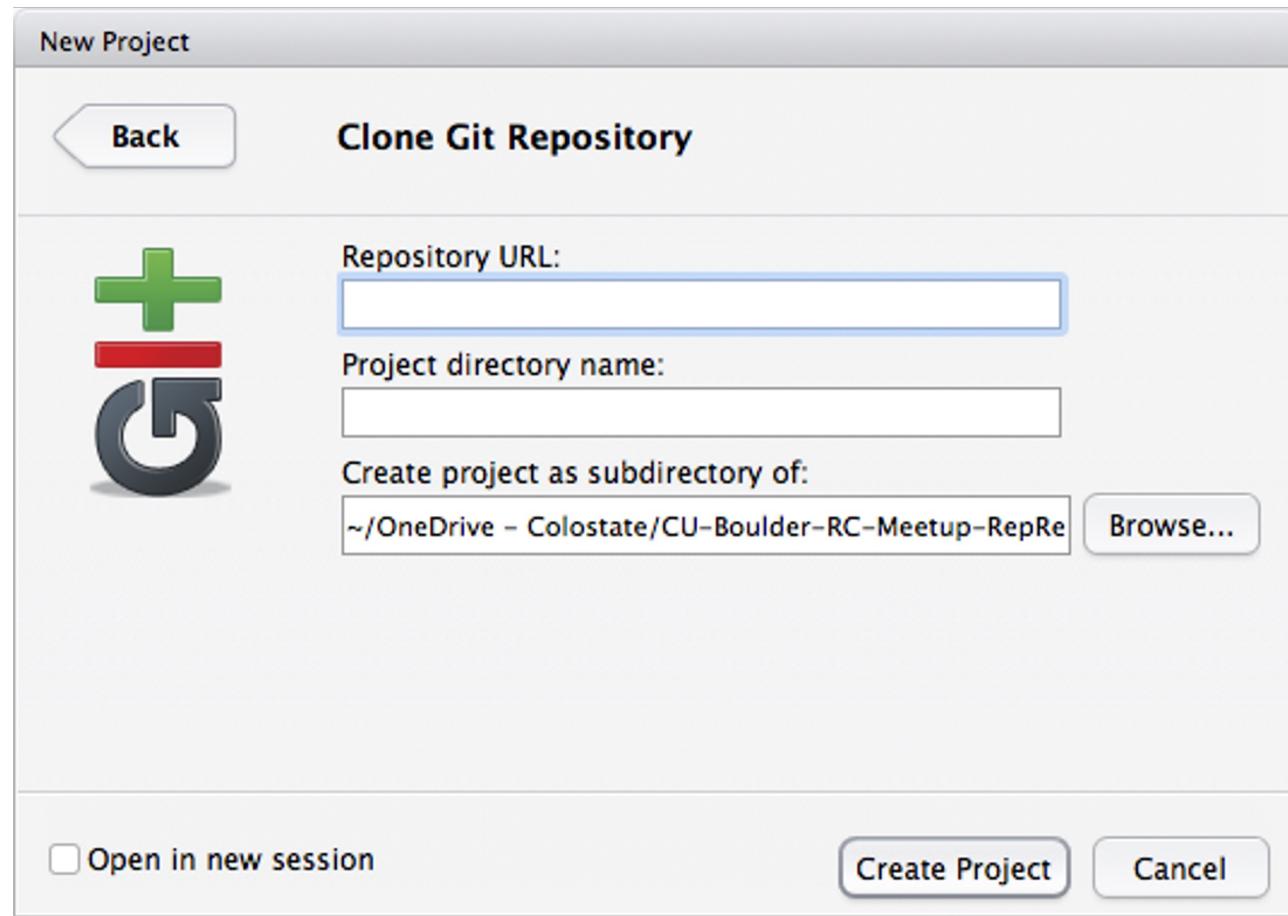
- no version control
- auto-detects existing version control
- **creates a version-controlled directory**



RStudio: Select git

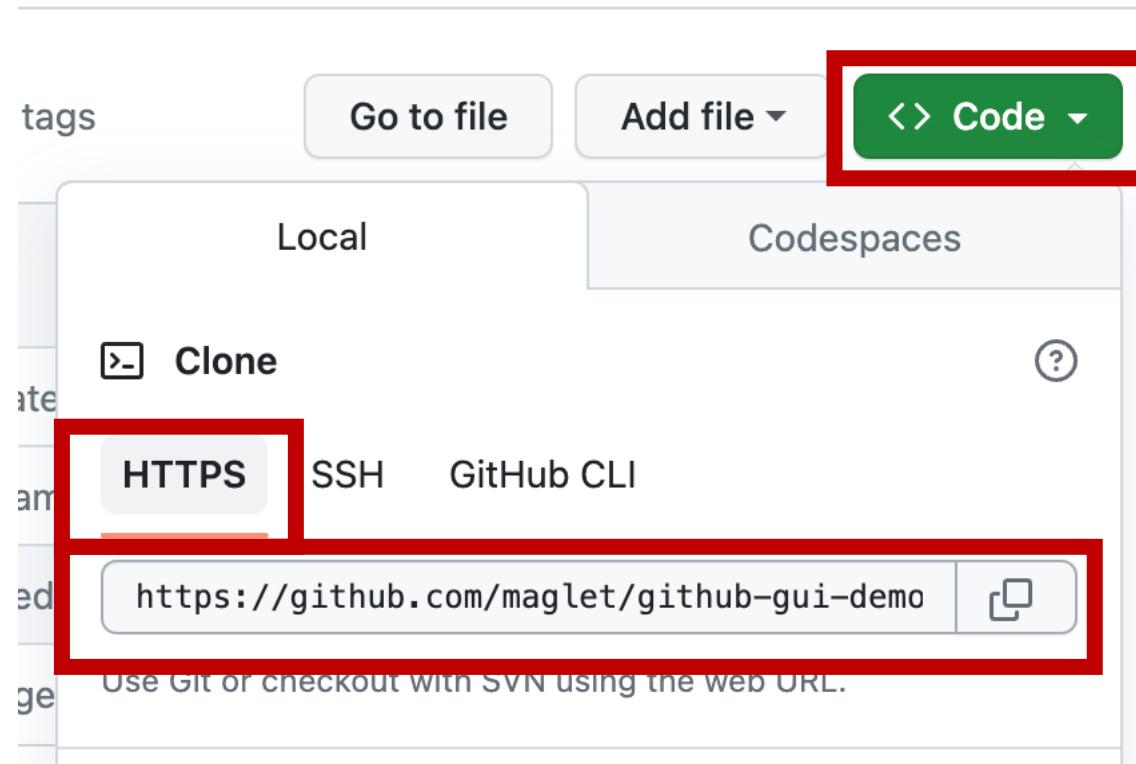


RStudio: Link to GitHub



- URL from github
- New dir name
- Parent directory

GitHub: Get link

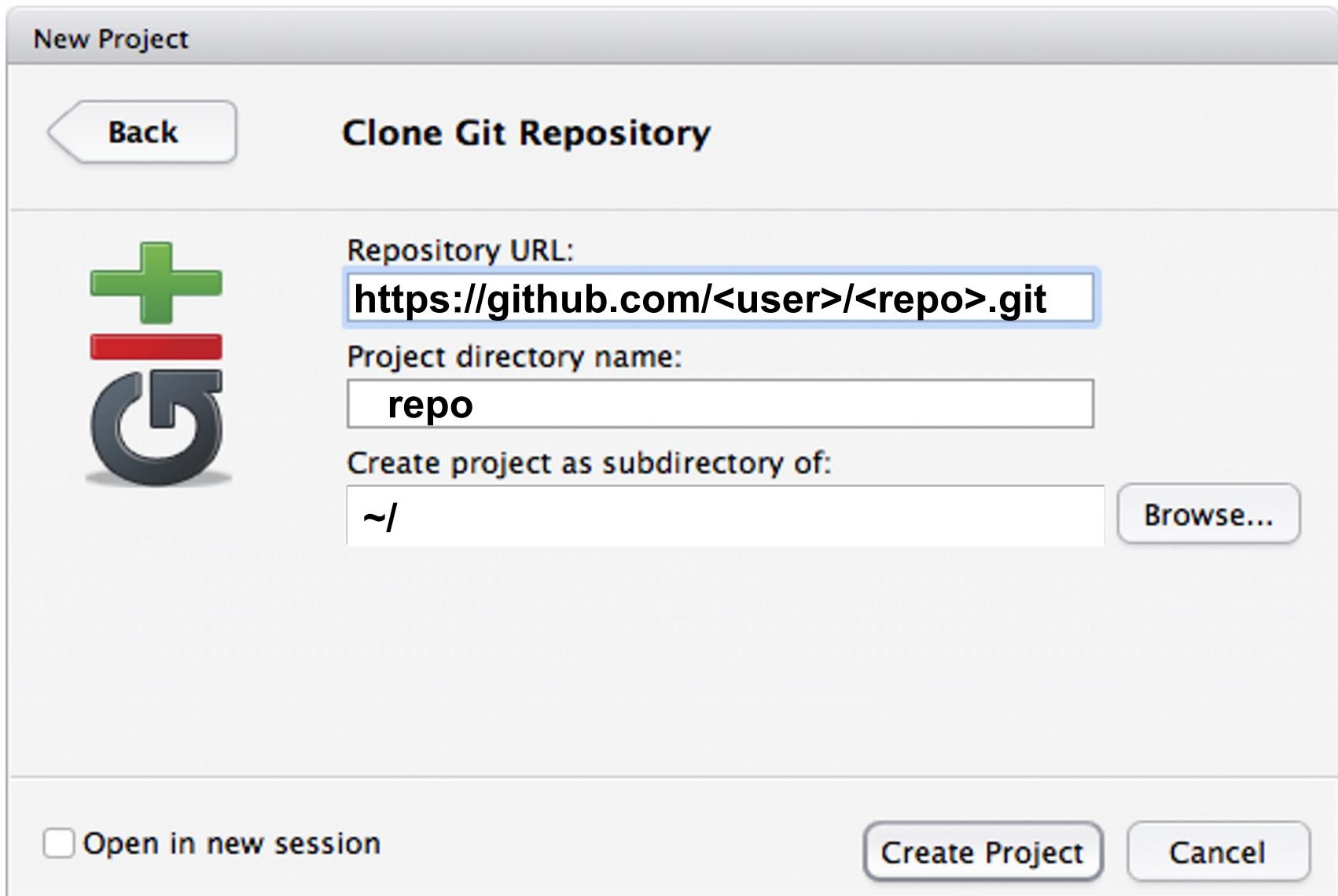


- This method

- Click

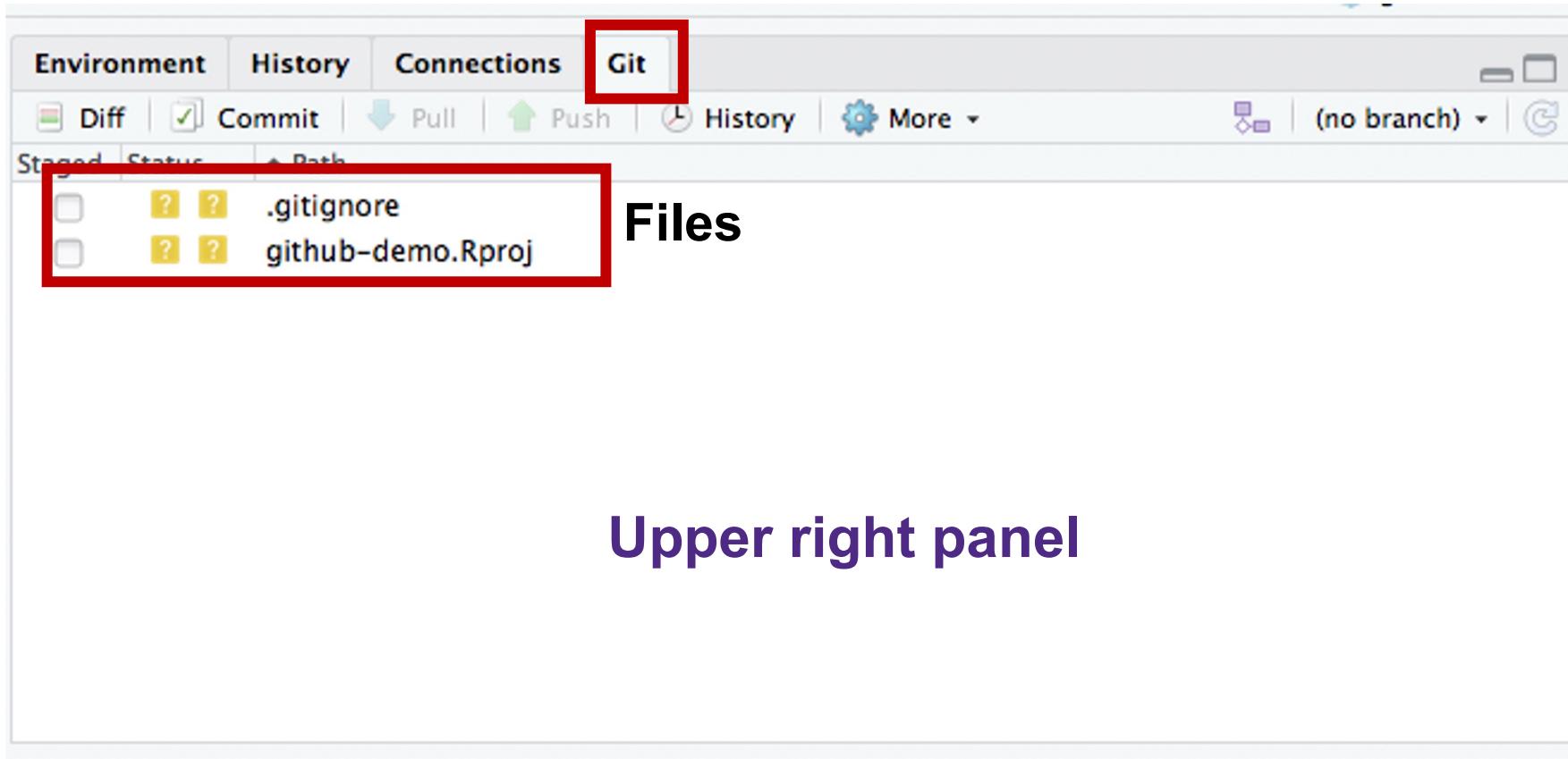
- Copy

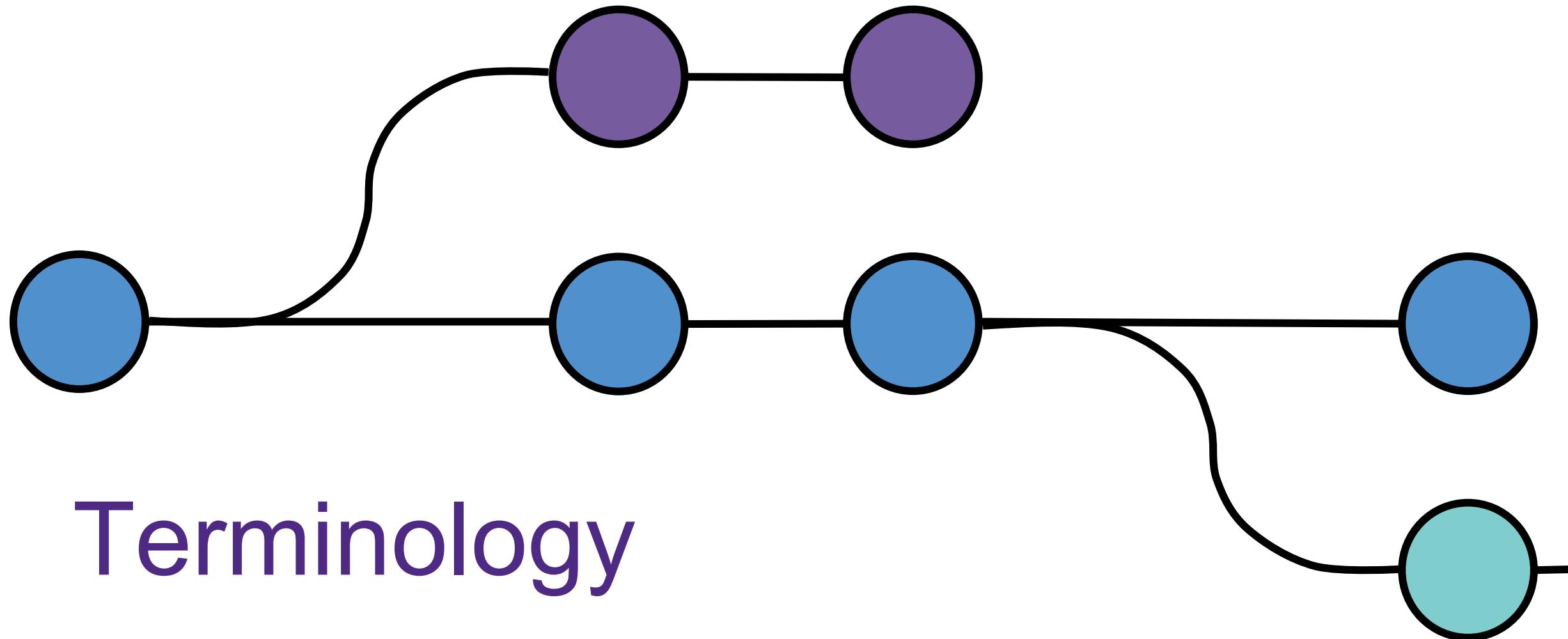
RStudio: Paste GitHub link



- Paste
- Match git repo name
- Pick

RStudio: Check for git tab





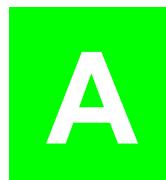
Untracked

Edited files recognized by git,
not acted upon



Added

Edited files to be committed to
your repository



Ignored

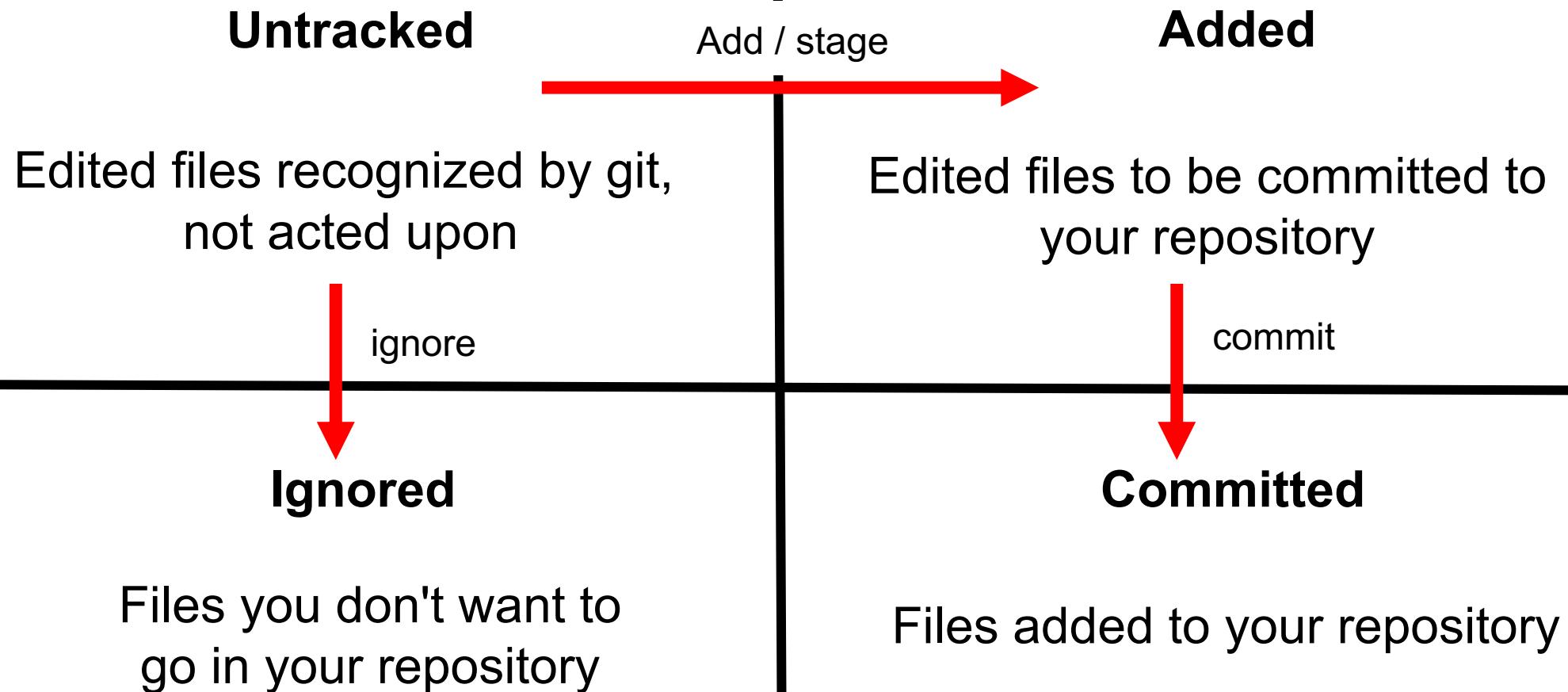
Files you don't want to
go in your repository

(Not seen in git tab)

Committed

Files added to your repository

(Not seen in git tab)



Untracked



.Rproj



.gitignore

(created by RStudio)



Added



Ignored

Committed

README.md

Untracked



.Rproj .git

(created by)



Added

Ignored

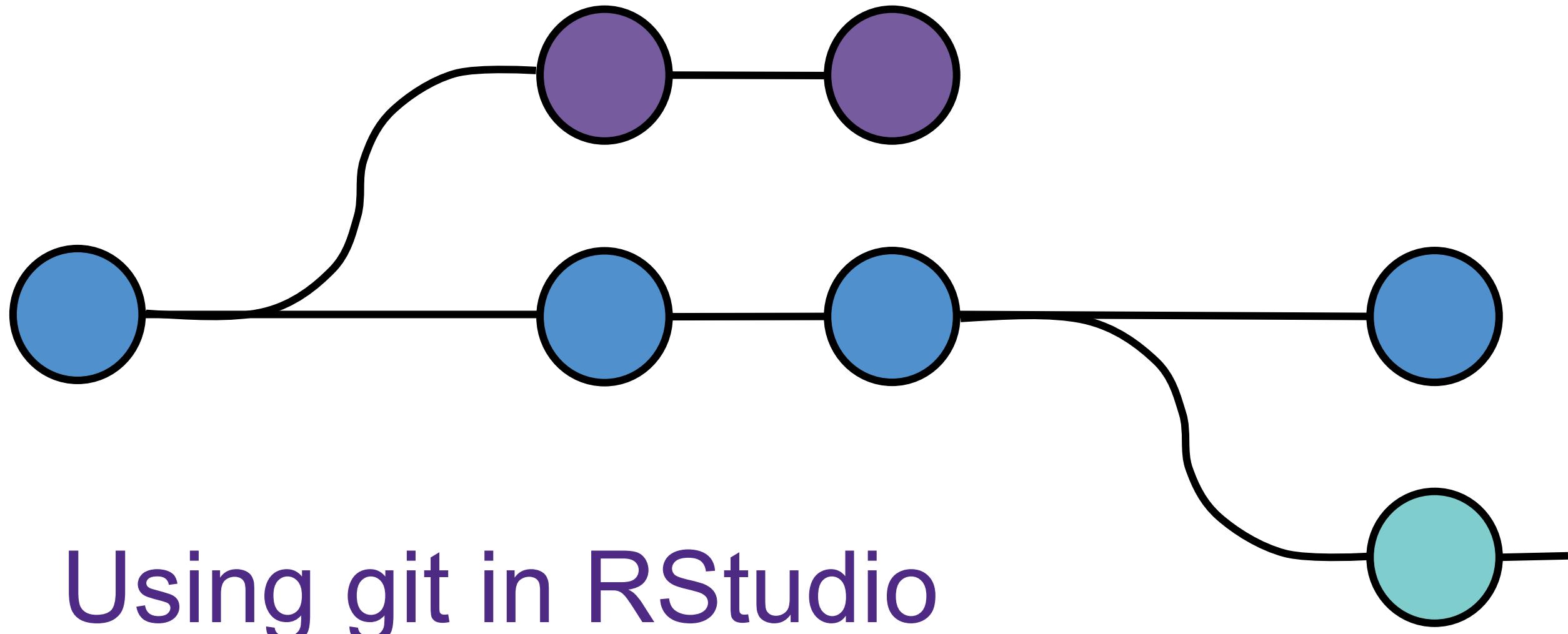
Goal

Commit all changes to the repository
(See nothing in the git tab)

Ignored



README.md

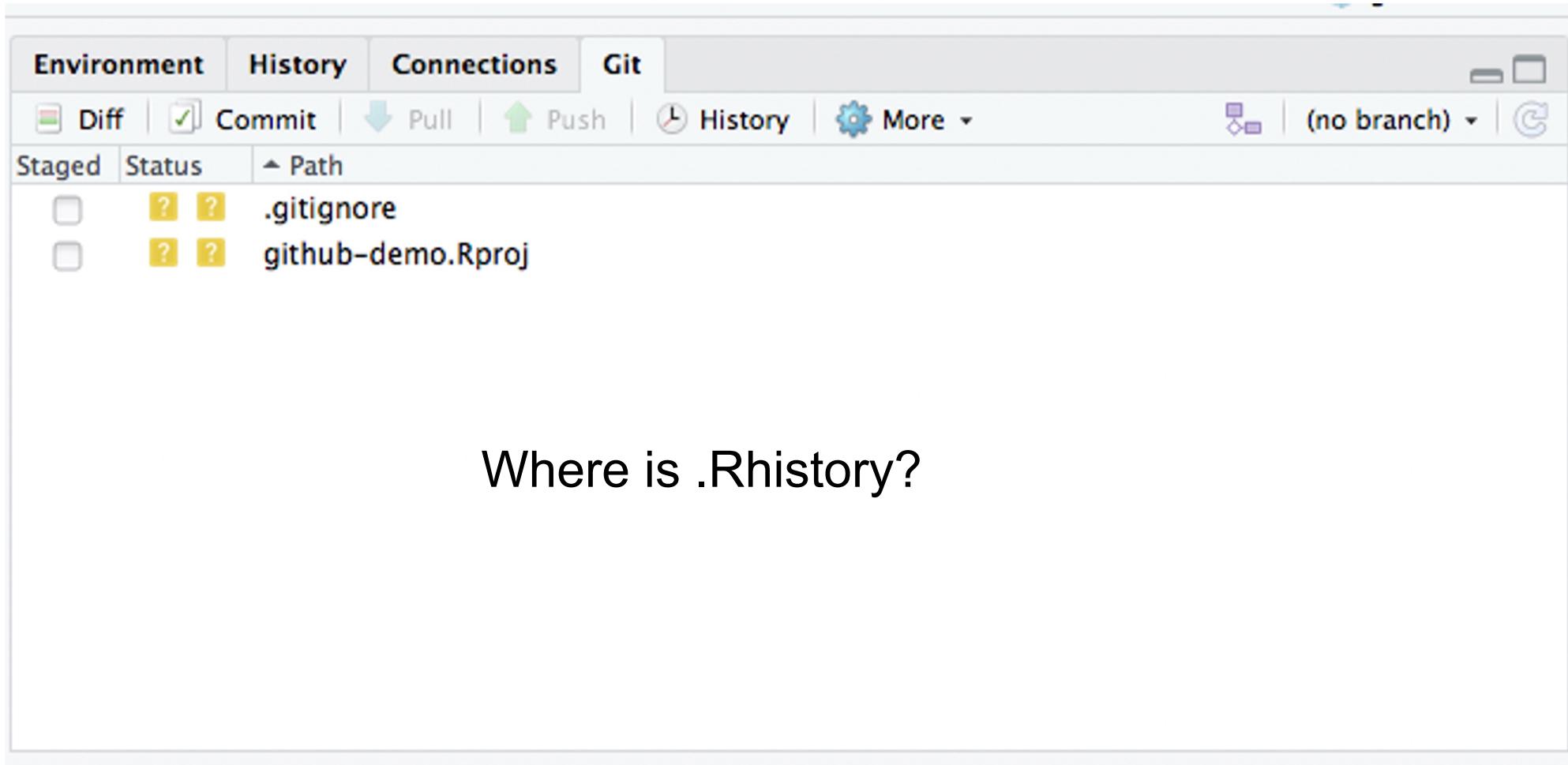


Files in repository folder

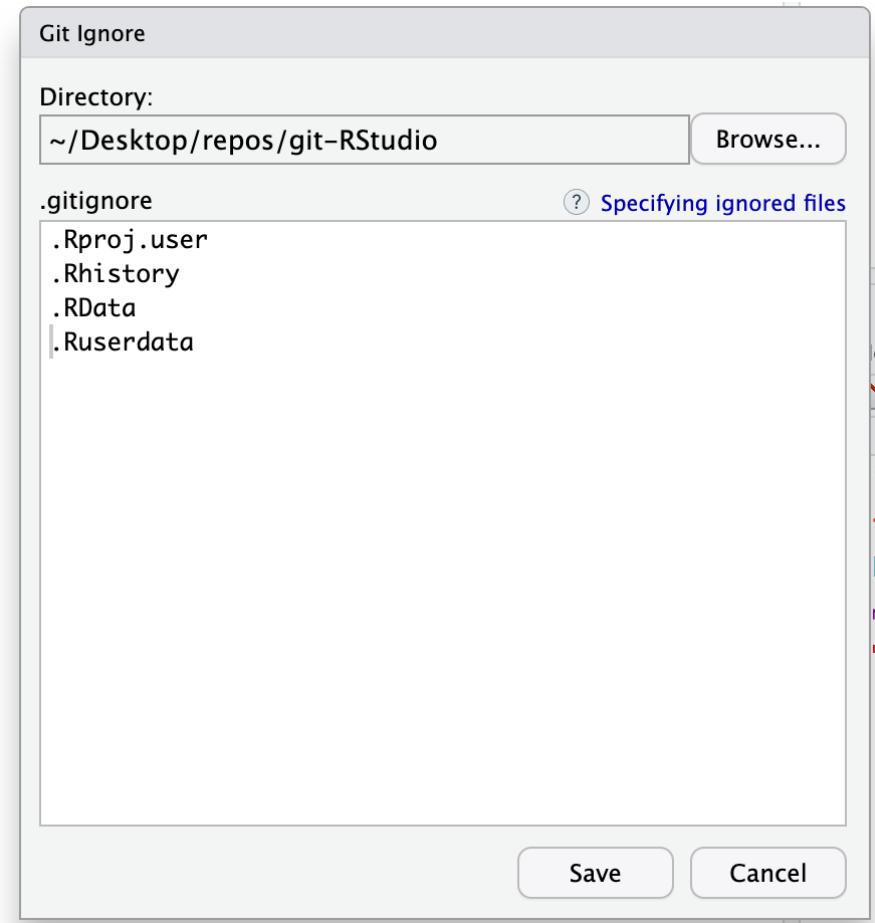
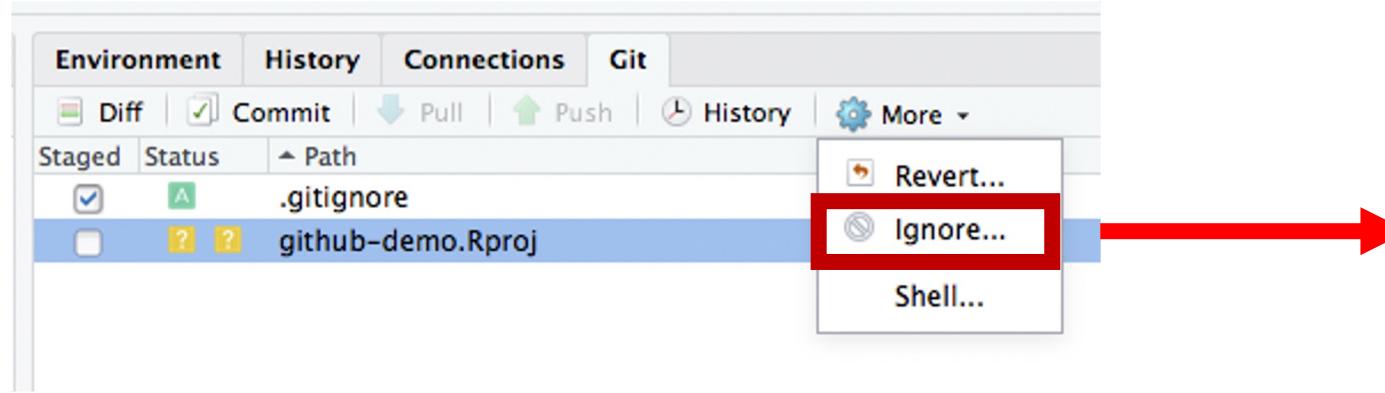
The screenshot shows the RStudio interface with the 'Files' tab selected. The title bar includes 'Files', 'Plots', 'Packages', 'Help', and 'Viewer'. Below the tabs are buttons for 'New Folder', 'Delete', 'Rename', and 'More'. The file path is displayed as 'C: > Users > cmtobin > Desktop > test'. The main area is a table listing files:

| | Name | Size | Modified |
|--------------------------|------------|----------------------|-----------------------------|
| | .. | | |
| <input type="checkbox"/> | .gitignore | From RStudio project | 44 B Mar 22, 2019, 12:34 PM |
| <input type="checkbox"/> | .Rhistory | From RStudio project | 0 B Mar 22, 2019, 1:55 PM |
| <input type="checkbox"/> | README.md | From GitHub | 6 B Mar 22, 2019, 12:34 PM |
| <input type="checkbox"/> | test.Rproj | From RStudio project | 218 B Mar 22, 2019, 2:21 PM |

Files with uncommitted changes



Ignored files



Add a file's name name to the .gitignore file to stop git from tracking it

Untracked



.Rproj



.gitignore

Added



Ignored



.Rhistory

Committed

README.md

Types of files git works best with

“Good” formats (text based)

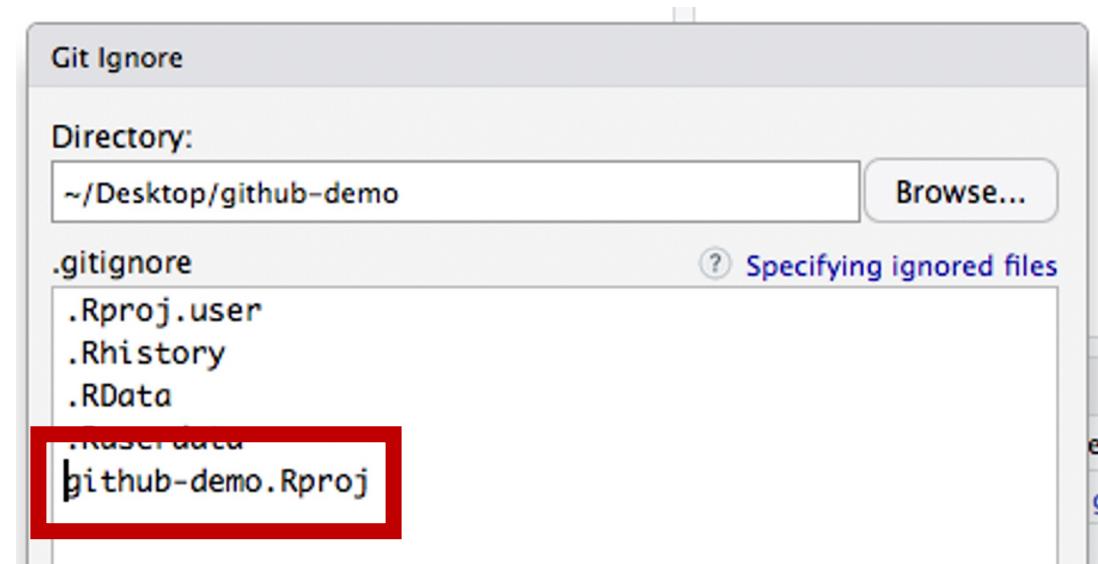
- Documents (.txt, .tex, .rtf, .Rmd)
- Tabular Data (.csv)
- Source code (.R, .py, .c, .sh)

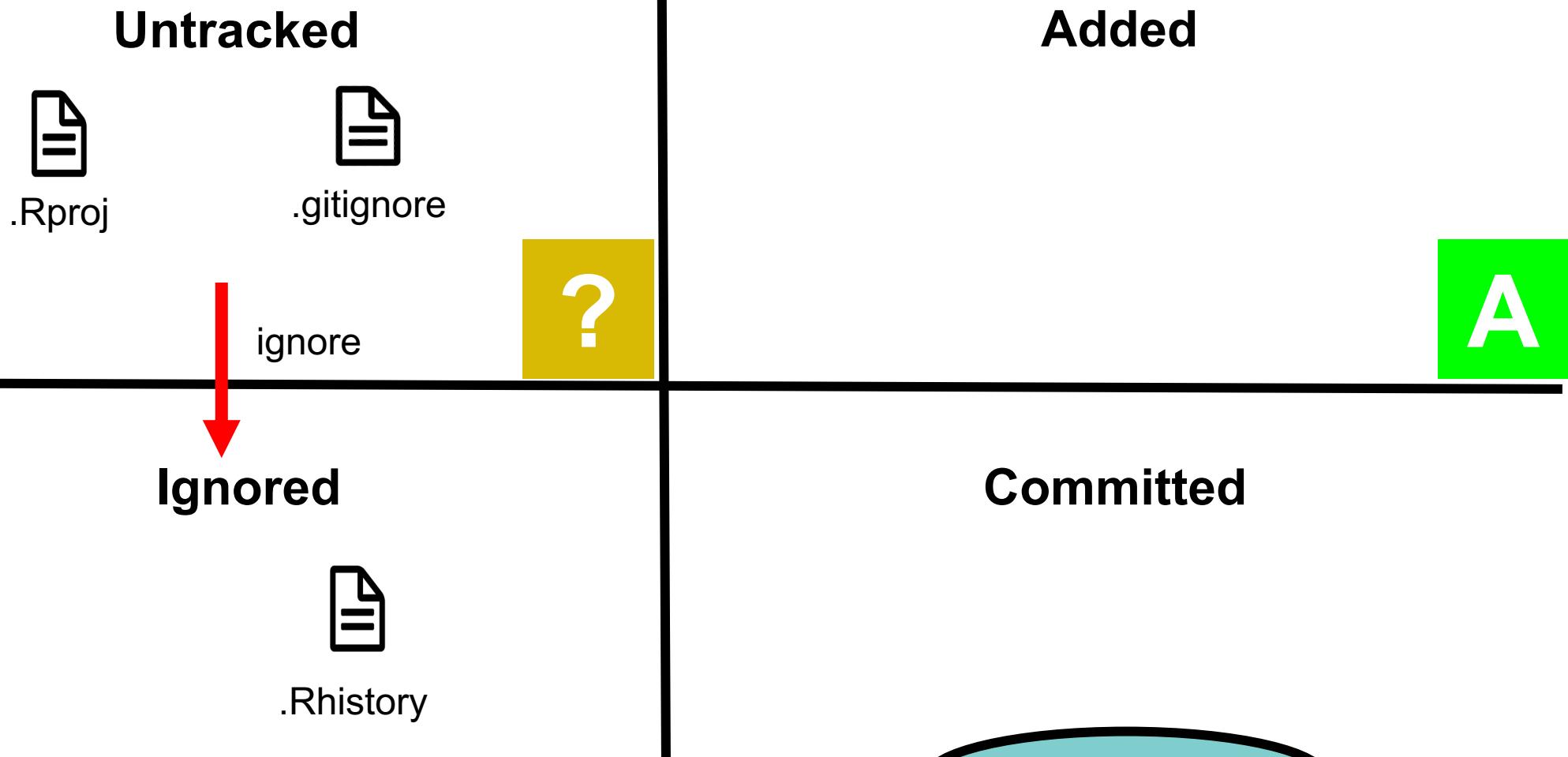
“Bad” formats (binary files)

- Documents (.docx, .pdf, .ppt)
- Excel spreadsheets (.xlsx)
- Media (.jpg, .mp3, .mp4)
- Databases (.mdb, .sqlite)

Ignoring files

- Added the R.proj file to the ignore list





Untracked



.gitignore

Added



Ignored



.Rproj



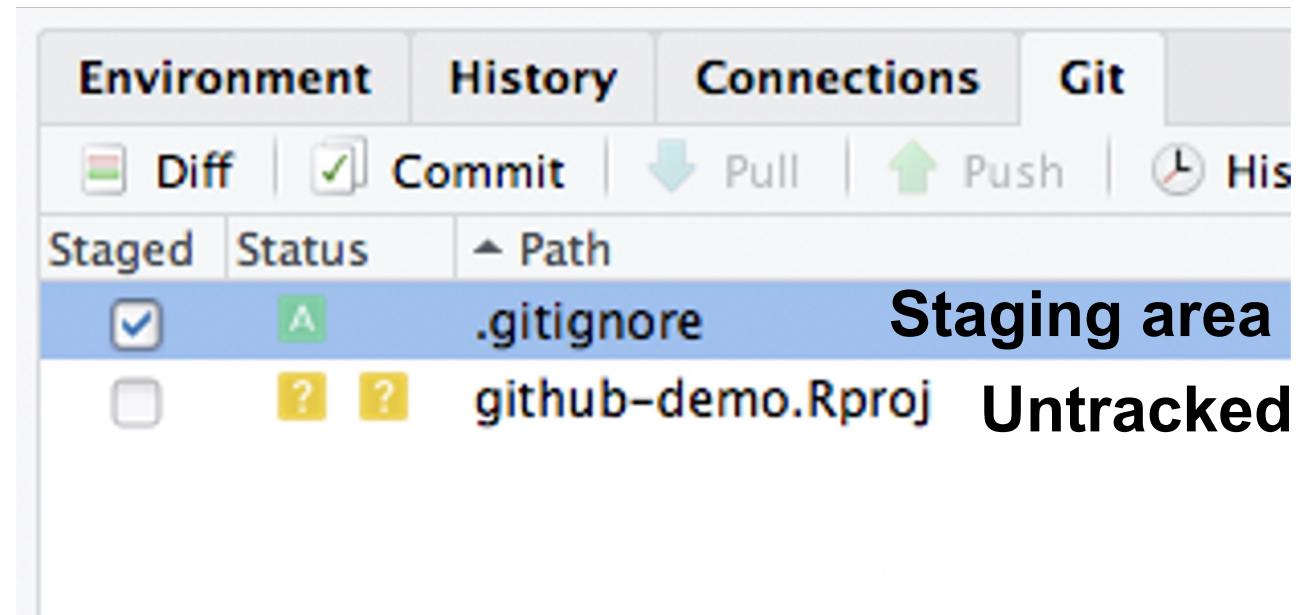
.Rhistory

Committed

README.md

Stage (Add) files

- Check boxes to add files to the “staging area”
- **Status:** Staged
- Staging area compiles changes to be committed to the repository



Untracked



.gitignore

Add / stage



Added



Ignored



.Rproj



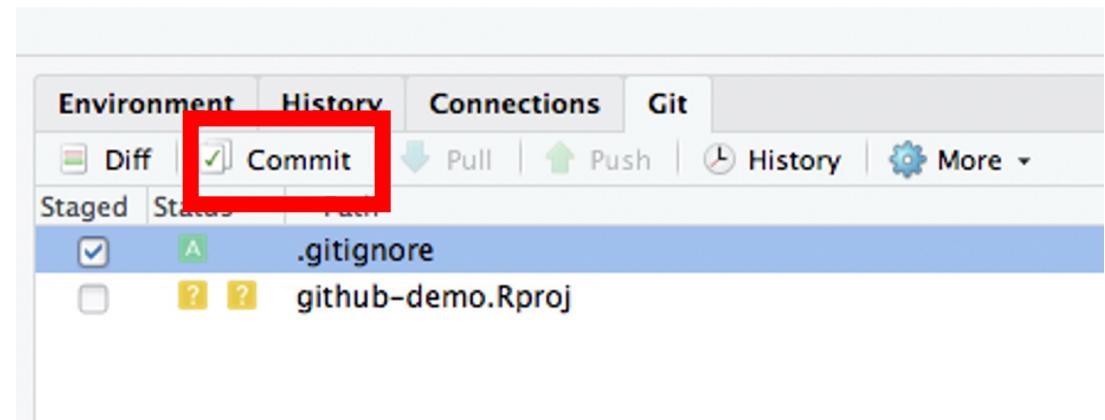
.Rhistory

Committed

README.md

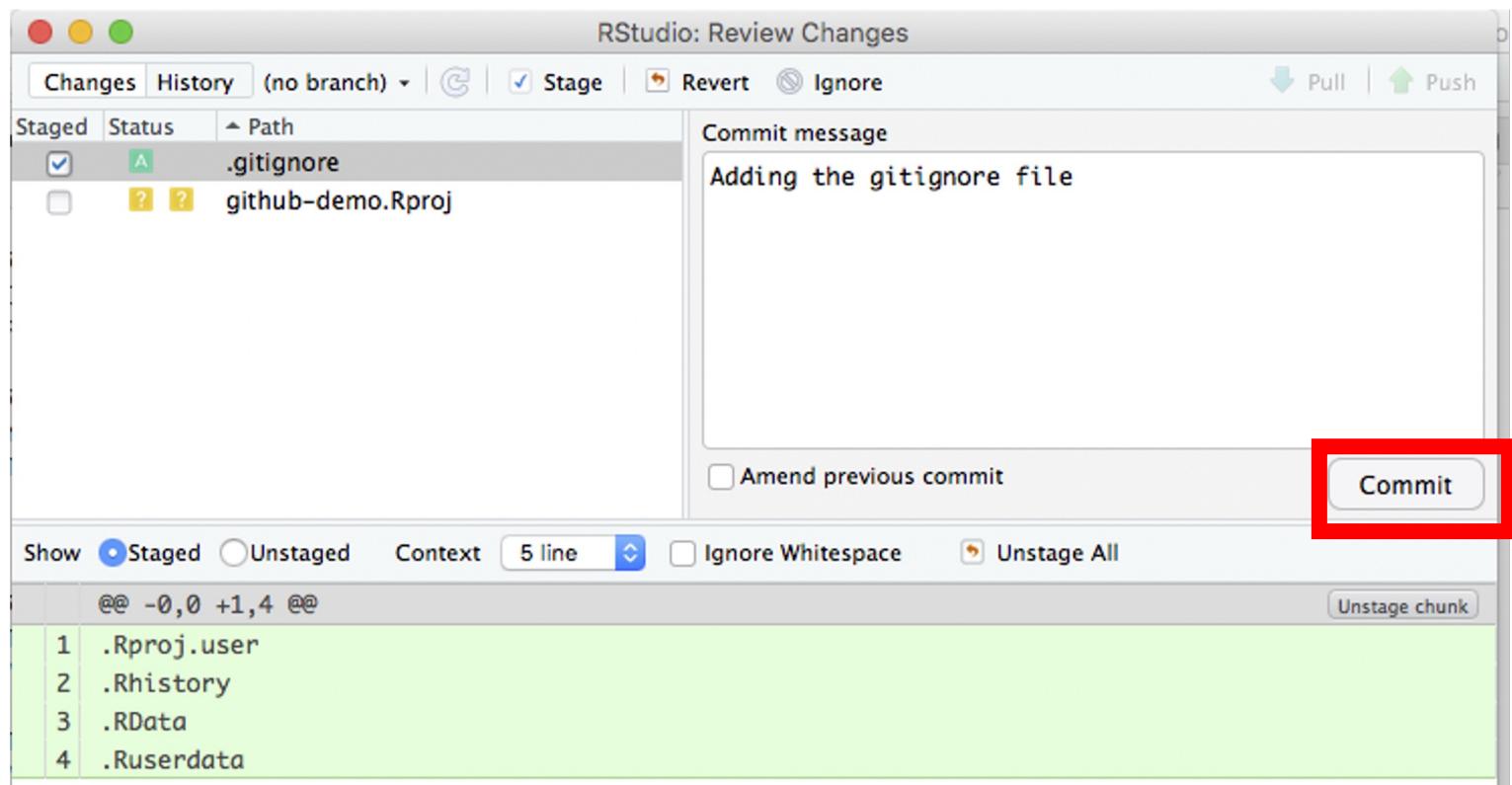
Commit files

- Clicking “commit” adds all files from the staging area to the repository
- Opens a new window



Commit message

- Add a message to say what changed in this commit
- Can see the contents of the file and what changed
 - Green = added
 - Red = removed



Untracked

Added

Ignored

Committed

 .Rhistory

 .Rproj



.gitignore

commit




README.md

Untracked

Added

?

A

Ignored



.Rhistory



.Rproj

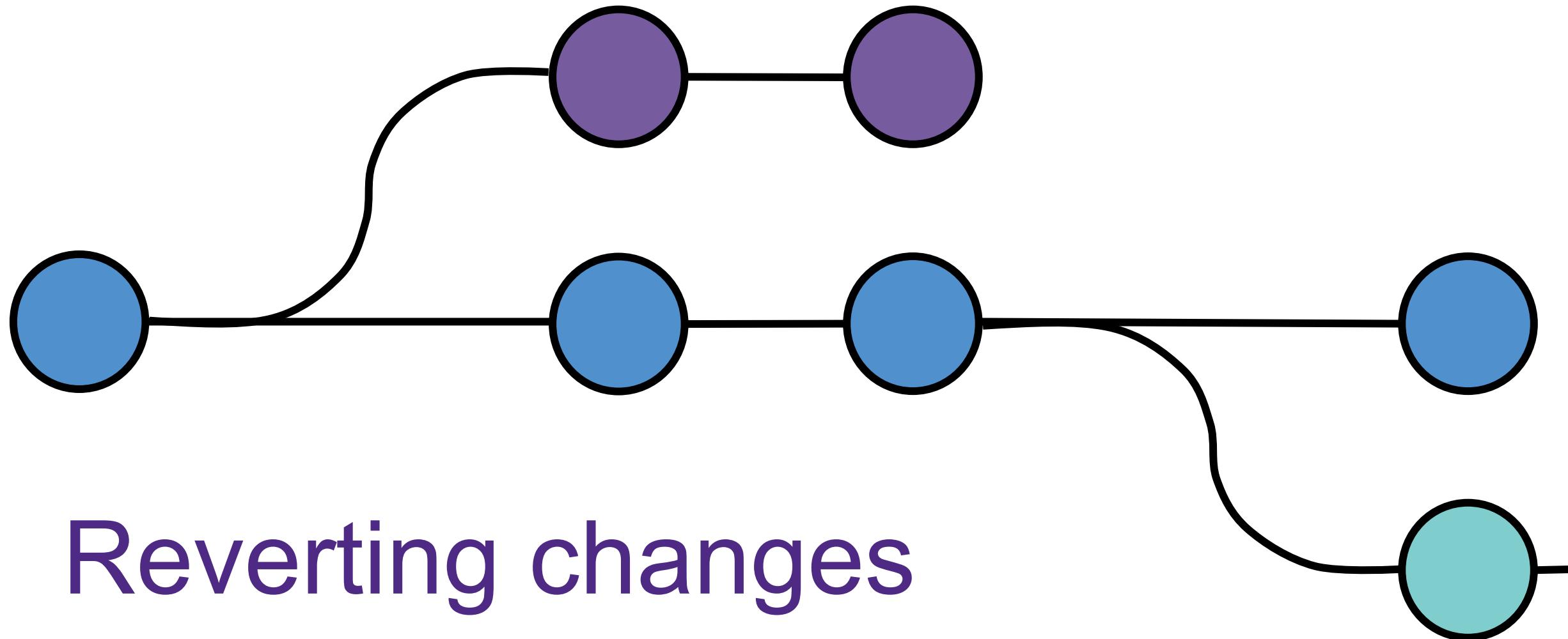
Committed

.gitignore

README.md

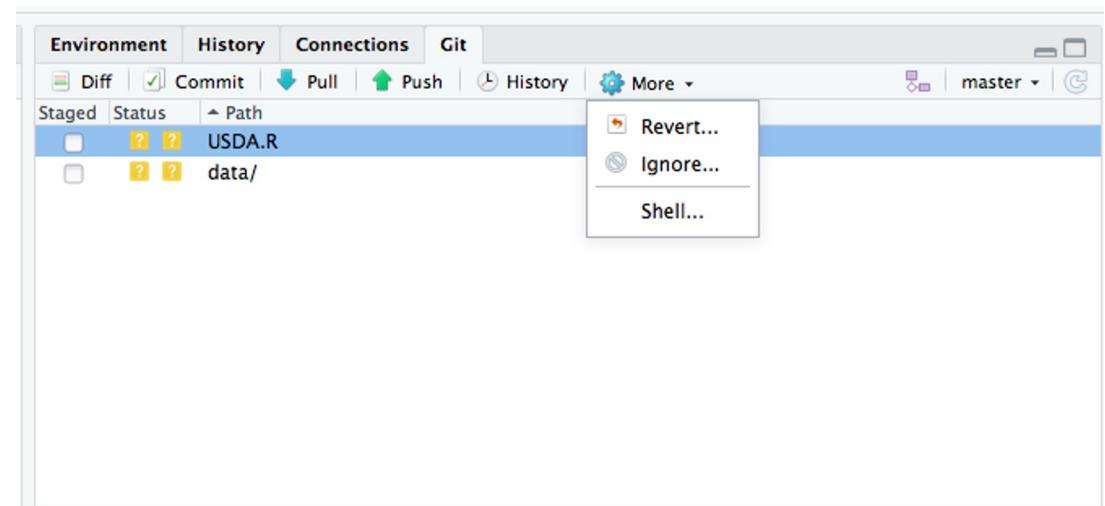
Exercise:

- Edit your README.md file
- **Stage changes: Add the changes file to the staging area**
- **Commit your staging area to the repository**



Revert

- Removes all uncommitted changes
- Includes staged changes
- Can't revert to committed changes from R Studio



https://intro2r.com/use_git.html#undo

History

The screenshot shows the RStudio: Review Changes interface. The top navigation bar includes tabs for 'Changes' (selected), 'History', 'master' (branch dropdown), '(all commits)' (dropdown), a search bar, and a 'Pull' button. The main area displays a table of three commits:

| | Subject | Author | Date | SHA |
|---------------------------|----------------------------|--------------------------------|------------|----------|
| HEAD -> refs/heads/master | ignored .Rproj | Tobin Magle <cmtobin@wisc.edu> | 2019-03-25 | c4b0d865 |
| origin/master | Adding the .gitignore file | Tobin Magle <cmtobin@wisc.edu> | 2019-03-22 | 8c56ce95 |
| origin/HEAD | Initial commit | maglet <ctobmag@gmail.com> | 2019-03-22 | ec9ad922 |

Below the table, commit details for the first entry are shown:

SHA c4b0d865
Author Tobin Magle <cmtobin@wisc.edu>
Date 2019-03-25 22:03
Subject ignored .Rproj
Parent 8c56ce95

A preview of the '.gitignore' file content is displayed:

```
@@ -2,3 +2,4 @@
 2 2 .Rhistory
 3 3 .RData
 4 4 .Ruserdata
 5 test.Rproj
```

A 'View file @ c4b0d865' link is located at the bottom right of the preview area.

Go back to previous commit

- Need to use the terminal
- git checkout <commit_id> <filename>
- File will show as modified, add and commit like normal

| Find Commit ID | | |
|---------------------------|---------------------------|----------|
| here | Tobin Magle <c 2019-03-26 | 3576c8e9 |
| origin/master origin/HEAD | Tobin Magle <c 2019-03-26 | 5a0df057 |
| readmey edits | Tobin Magle <c 2019-03-26 | a08c46b2 |

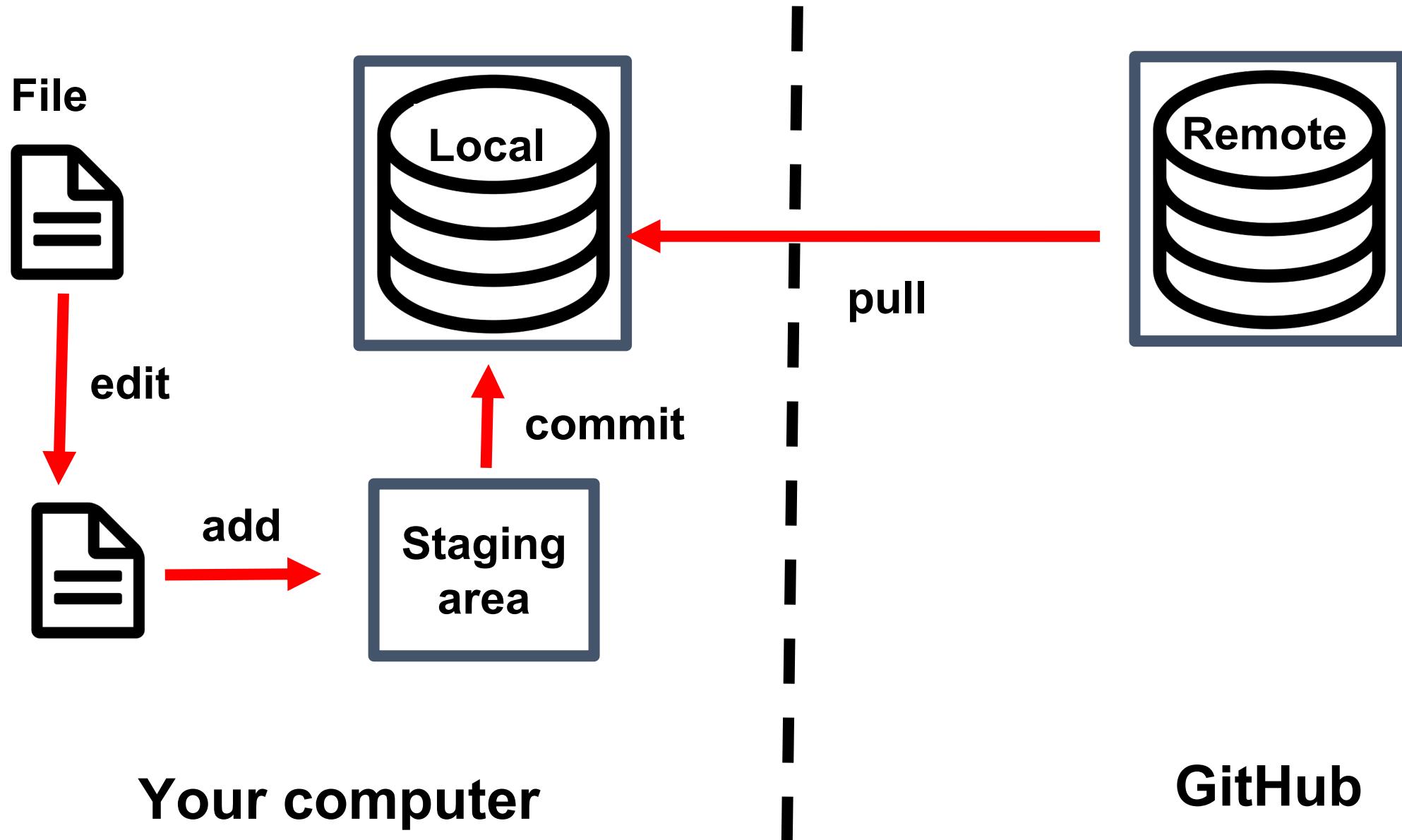
Run the checkout command in the terminal

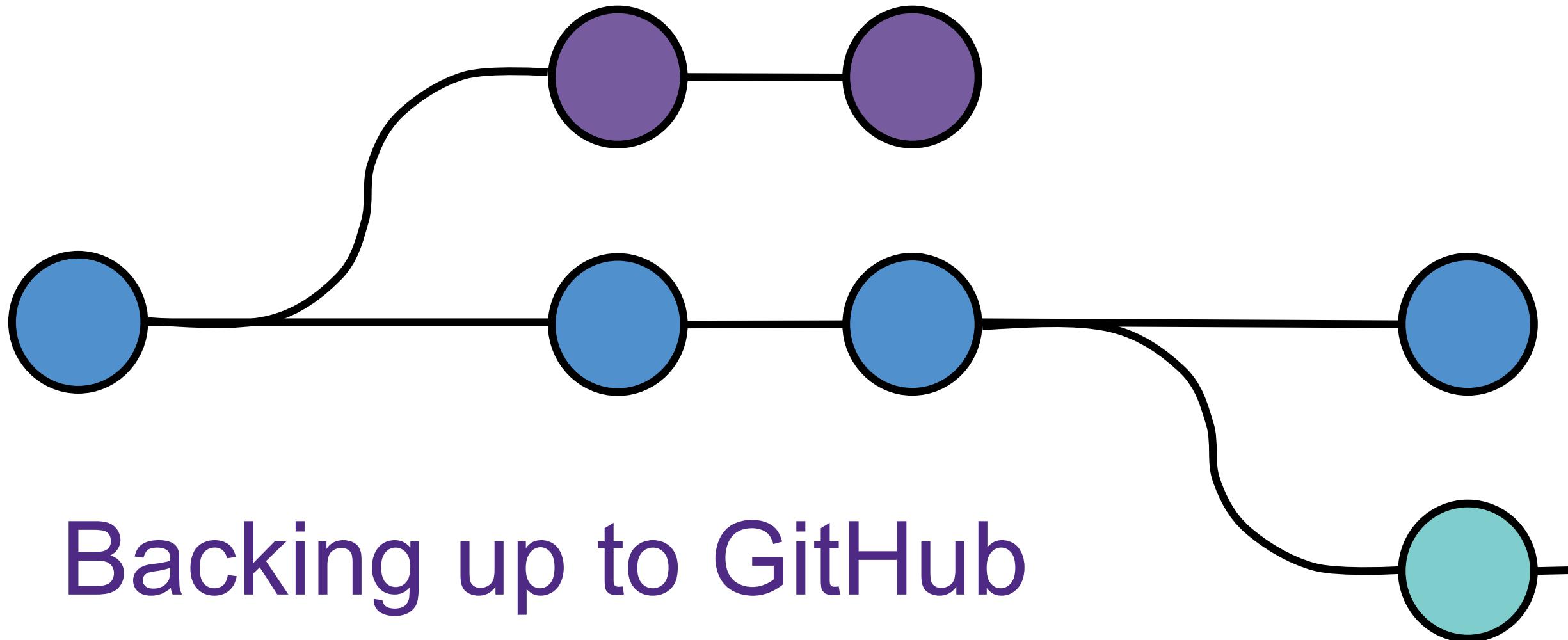
```
cmtobin@GLS-LAP8228 ~/Desktop/test2 (master)
$ git checkout 5a0df057 README.md
```

file has been modified

| Staged | Status | Path |
|-------------------------------------|--------|-----------|
| <input checked="" type="checkbox"/> | M | README.md |

Git workflow

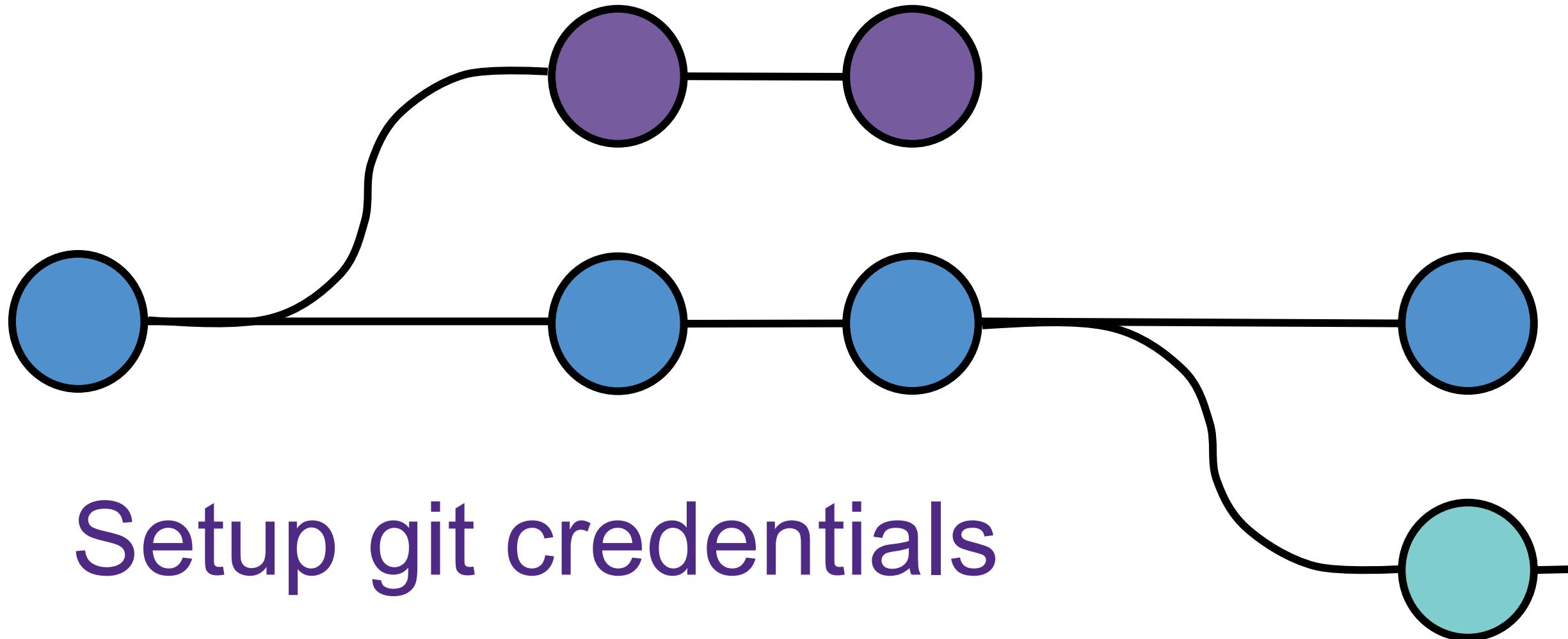




GitHub best practices

- Store **code and documentation**, not data
- Never store **protected or personal data** even in a private repo
- Store **credentials** (private keys or tokens) in a separate file outside of your repository
 - ie, not in your code
- Add a **personal email** to your GitHub account, so you can use it when you leave Northwestern





Setup

During the workshop

- GitHub: Create a personal access token
- RStudio:
 - Install usethis package
 - **Configure** your git user
 - Set up **git authentication**

Configure git on your computer

1. Install and load usethis
2. Set git username and email
 - Commit history shows username
 - **Use your GitHub account email**
3. Set GitHub authentication

In RStudio console

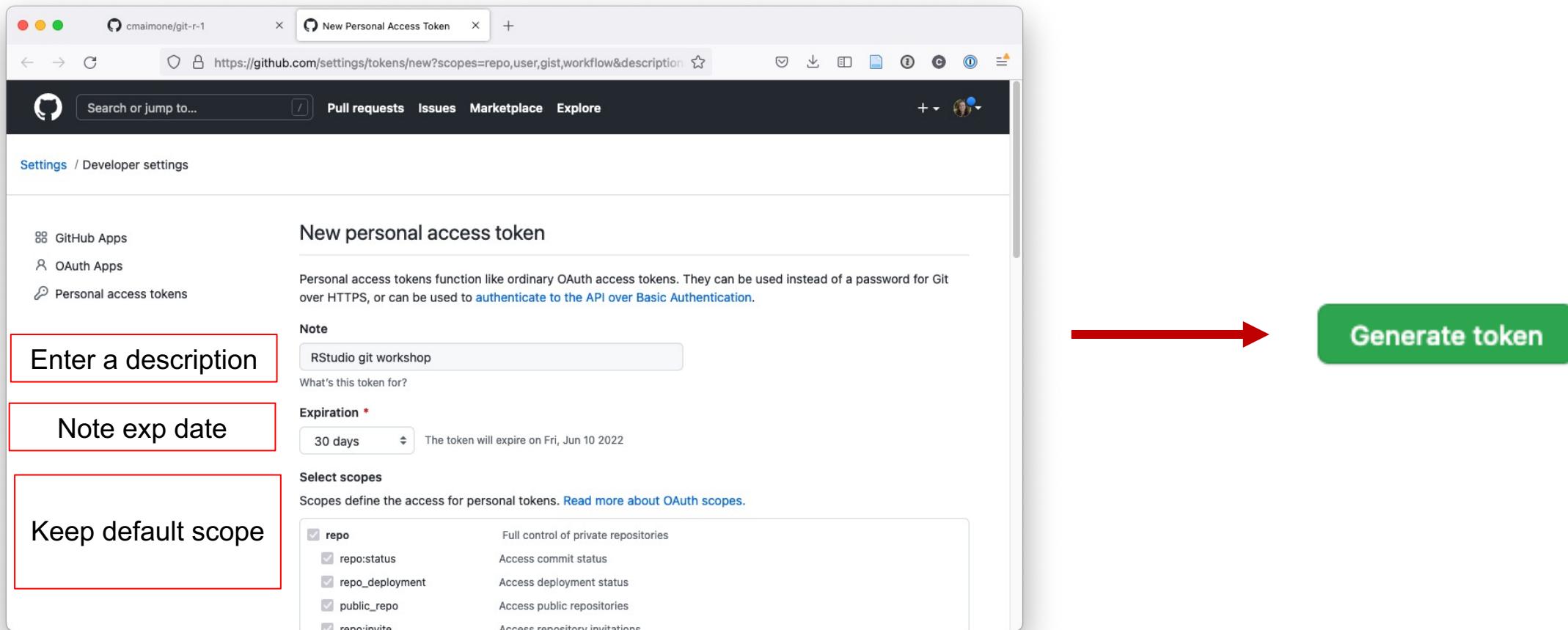
```
install.packages("usethis")
library(usethis)
```

```
usethis::use_git_config(user.name="Jane Doe",
user.email="jane@example.org")
```

```
usethis::create_github_token()
```

Create GitHub Token

use this::create_github_token()
(This page will open)



More about scopes: <https://docs.github.com/en/apps/oauth-apps/building-oauth-apps/scopes-for-oauth-apps>

GitHub Personal Access Token

Copy token (it won't be visible once you leave the page)

Personal access tokens

Generate new token Revoke all

Tokens you have generated that can be used to access the [GitHub API](#).

Make sure to copy your personal access token now. You won't be able to see it again!

| | |
|--|------------------------------------|
| ✓ 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  | Delete |
| downloading individual files programatically from metrics db — repo, workflow | Last used within the last 5 months |
| ⚠ This token has no expiration date. | |

Personal access tokens function like ordinary OAuth access tokens. They can be used instead of a password for Git over HTTPS, or can be used to [authenticate to the API over Basic Authentication](#).

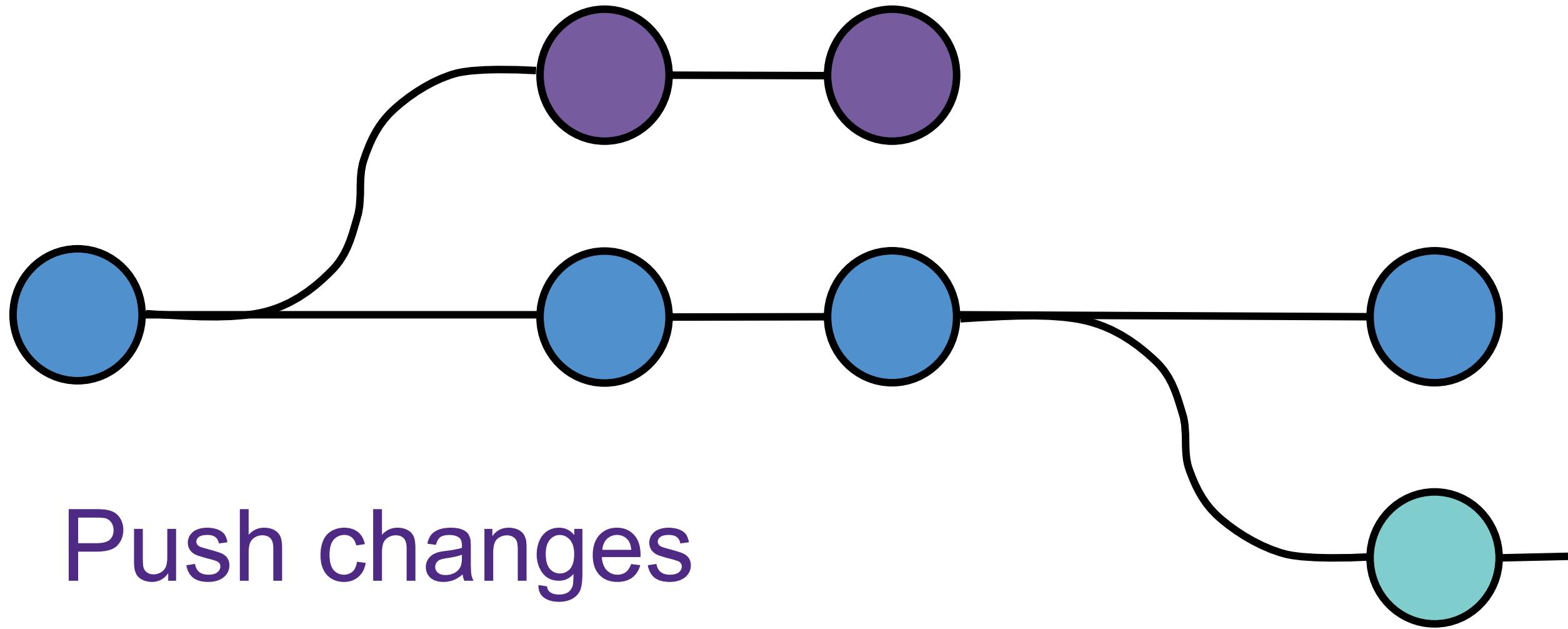
Set GitHub Credentials

In R Console:

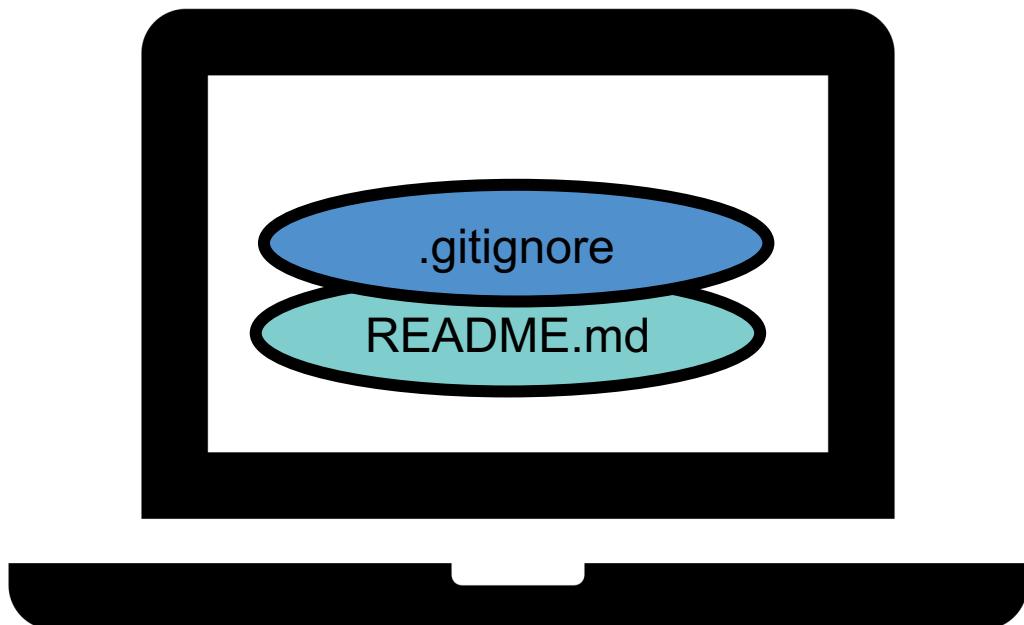
```
gitcreds::gitcreds_set()
```

Follow the prompts

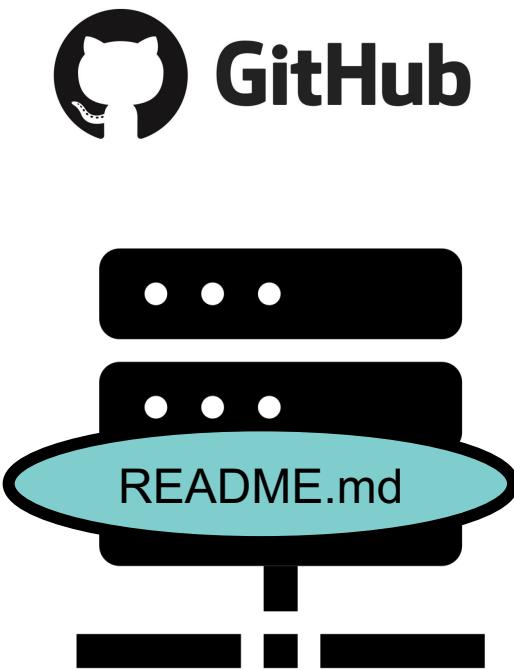
- Paste in the token when asked
- If you had existing credentials (including username/password which are no longer allowed), you may need to choose option 2: Replace these credentials
- git saves these credentials for you



Local changes not backed up

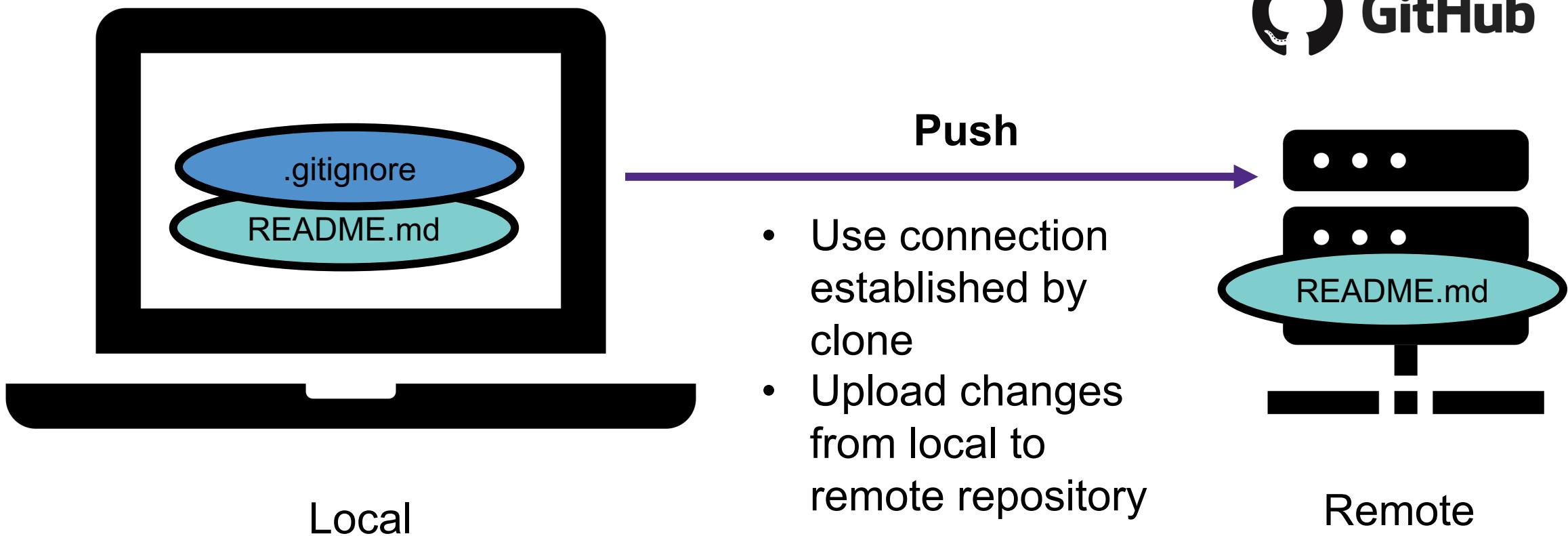


Local



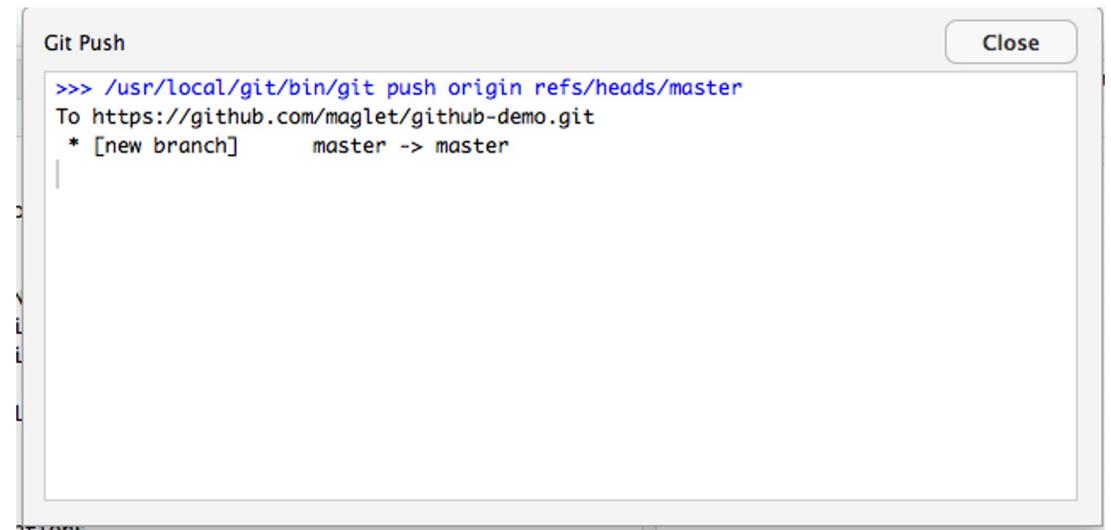
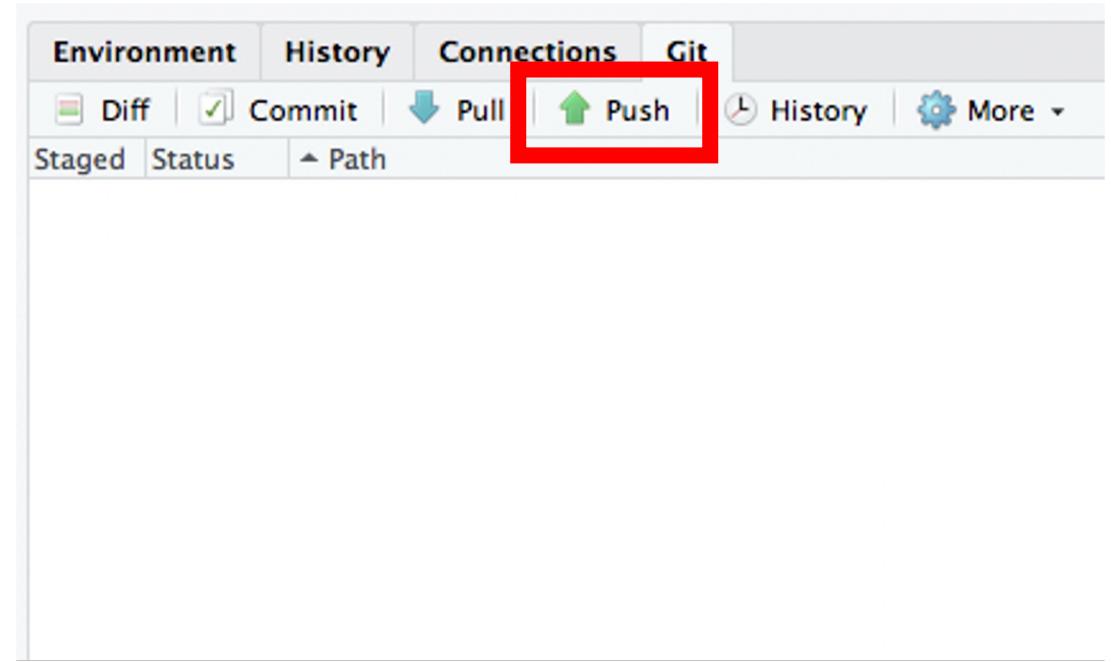
Remote

Push to remote

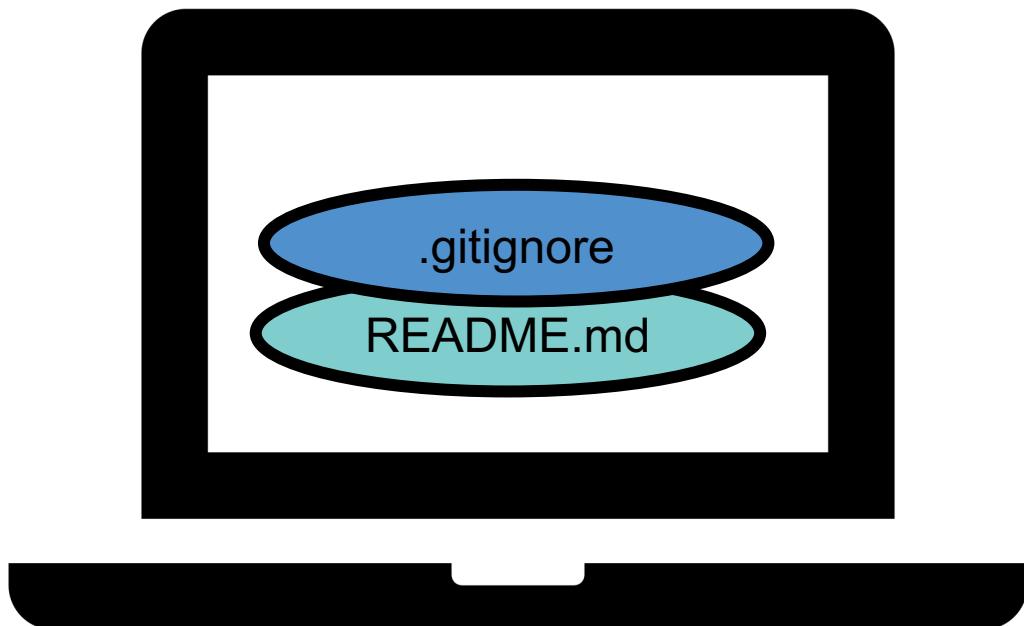


Push

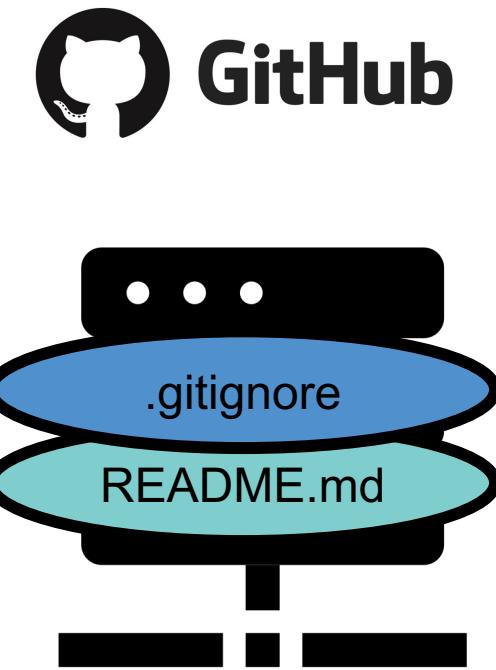
- Save local changes to remote repository
- Changes must be committed before pushing
- Requires that you have authentication set up on your machine



Changes pushed



Local



Remote

Files in remote

maglet / **github-demo**

Code Issues 0 Pull requests 0 Projects 0 Wiki Insights Settings

No description, website, or topics provided. Edit

Add topics

2 commits 1 branch 0 releases 0 contributors

Branch: master ▾ New pull request Create new file Upload files Find file Clone or download ▾

Tohin Magle Added .Rproj to the .gitignore Latest commit 4138947 10 minutes ago

.gitignore Added .Rproj to the .gitignore 10 minutes ago

Help people interested in this repository understand your project by adding a README. Add a README

Practice:

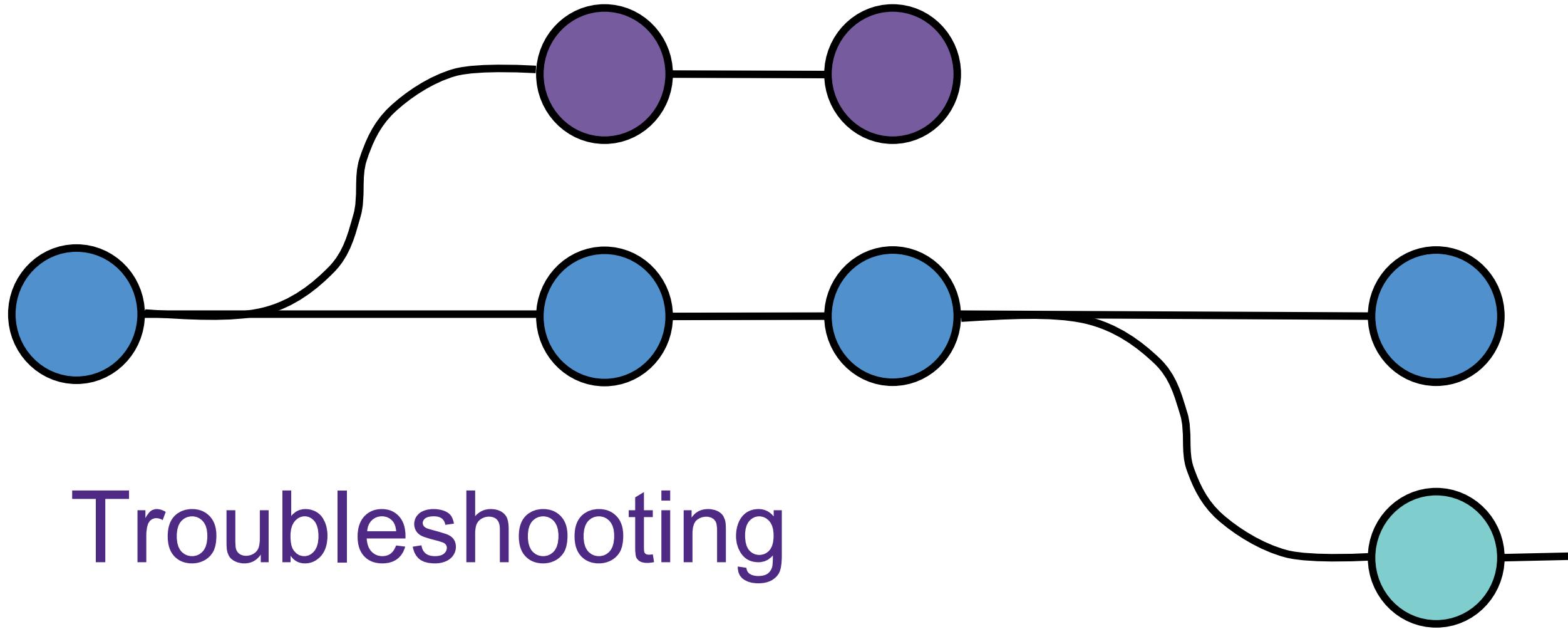
- Make a new R script
- Save it to the R project directory
- Add some R code
- Do the steps necessary to get this file into your GitHub repo

Need help?

Workshop materials: <https://github.com/nuitrcs/git-RStudio>

- Request a consult: bit.ly/rcdsconsult
- Other Resources
 - [RCDS: Written Setup Instructions](#)
 - [RCDS: Other ways to set up Git repos in R](#)
 - What git is and how it works: [Git for Humans](#)
 - Why you should use git: [Academic benefits of using git](#)
 - How to use git with R and RStudio: [Happy git with R](#)

Troubleshooting



Q: I don't see the git tab

A: Check that RStudio sees git

Mac: RStudio > Preferences

Windows: Tools > Global Options

Mac: type
which git
in the Terminal to get the path

