VERSION
CONTROL
Using Git, GitHub and RStudio

"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

Learning objectives

USE VERSION CONTROL TO TRACK CHANGES IN WORK

CREATE, SHARE AND MERGE CONTENT WITH GITHUB

INTEGRATE GITHUB WITH RSTUDIO

VERSION CONTROL

What is version control and what does it do?

Version control

- **Think about a recent group assignments that you have done prior to today …**
  - How does your group keep track of who has written / edited what?
  - How does your group keep track of which is the most up to date version?
  - How does your group manage with accepting/rejecting then merging changes made by different people?
  - What happens if the group loses access to the server?

Version control

- **Now imagine if you had to manage a large data science or epidemiology research project, with multiple team members and large amounts of data …**
  - The data, the analysis and the models might all be changing over time
  - Not just one document but multiple documents and very many code fragments
  - Complex and critical dependencies between functions, packages, etc
  - Your client demands all work to be fully auditable and always verifiable

What is version control?

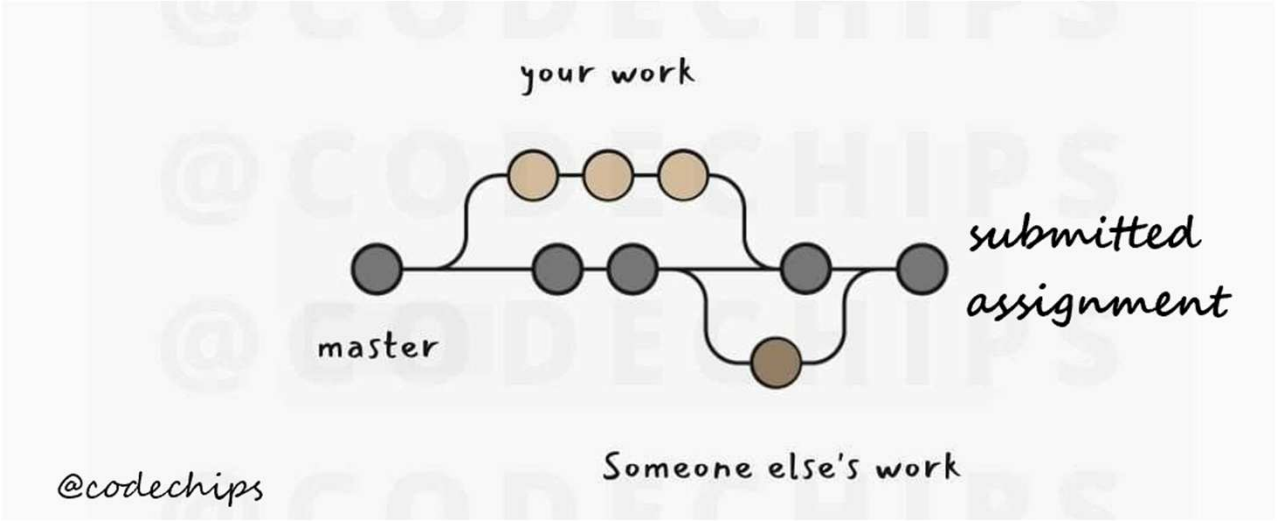
- **Version control is a class of "quality management" system that is responsible for documenting all changes to computer code, documents, web sites, …**
- **i.e. any form of information collection (data version control), AND**
- **any means of processing that information collection in order to yield new information (analysis version control).**

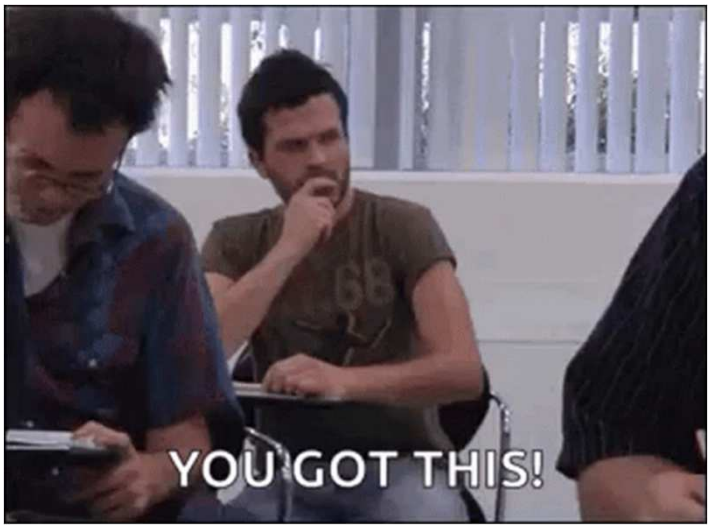
- Originally developed for teams of software developers but it is now widely used in data science.

- A good version control system must enable many people to simultaneously collaborate on the same piece of analysis.

No version control for me - I work alone!

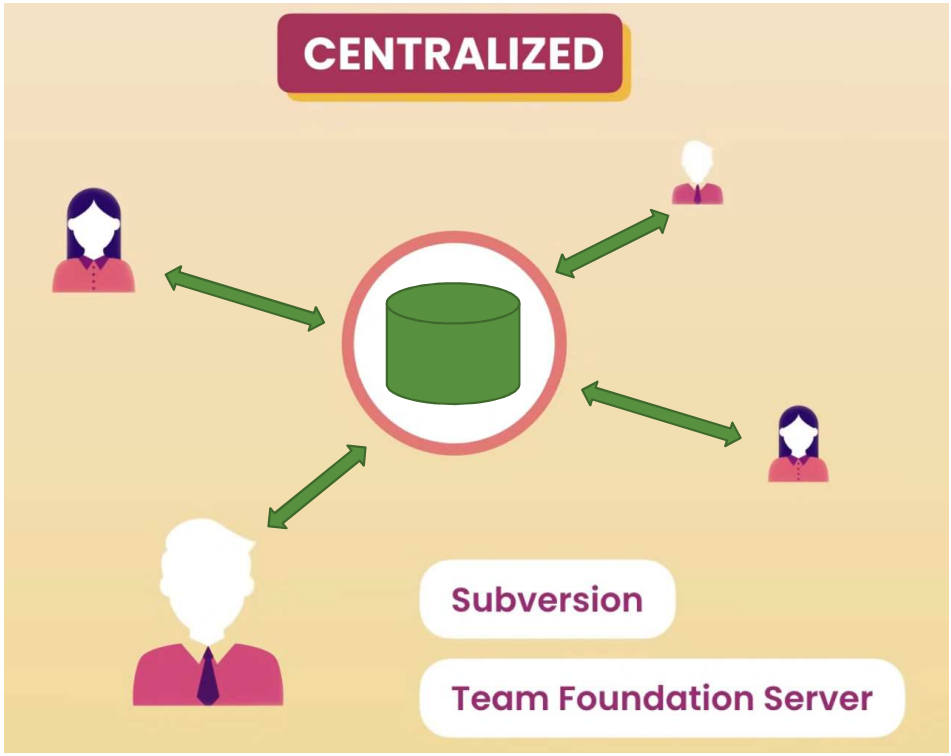
Yes, this is in the exam!

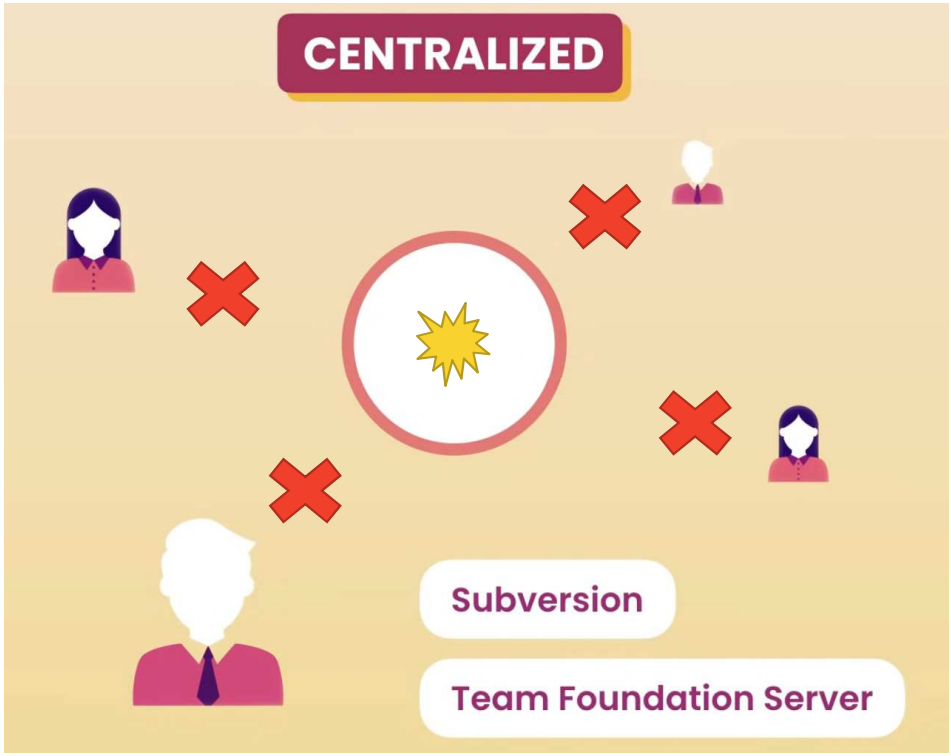


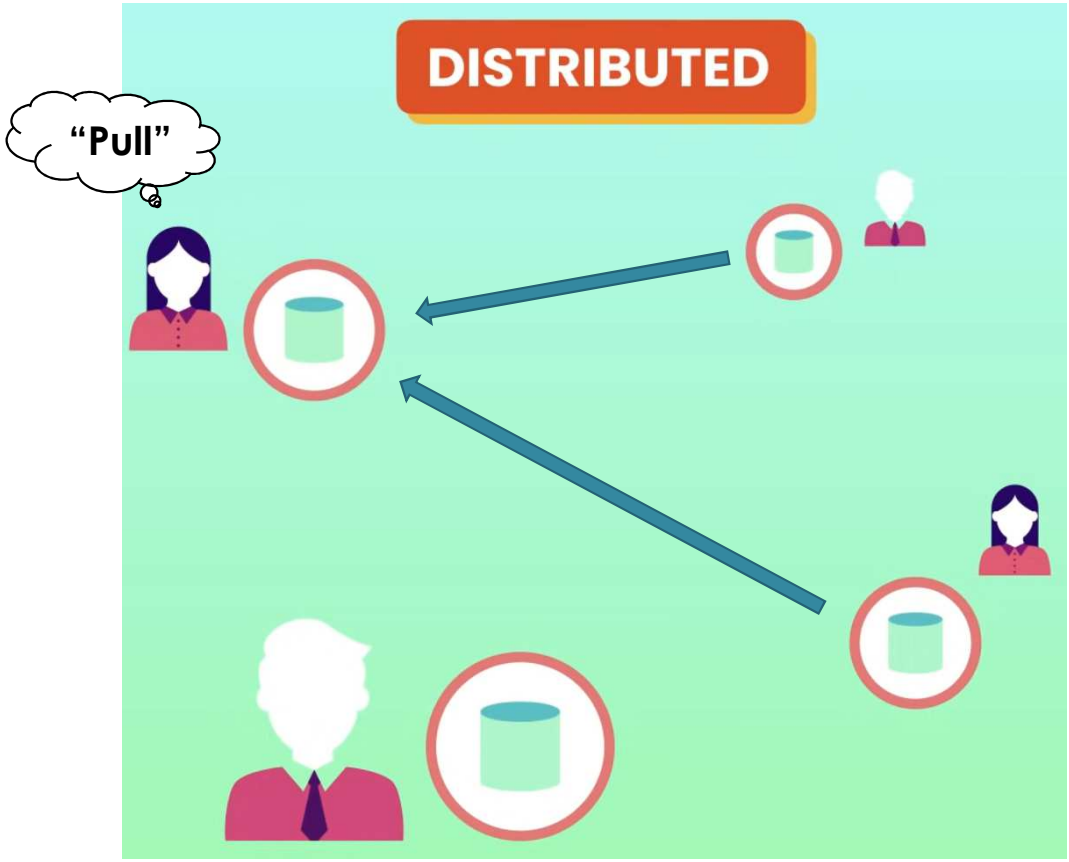


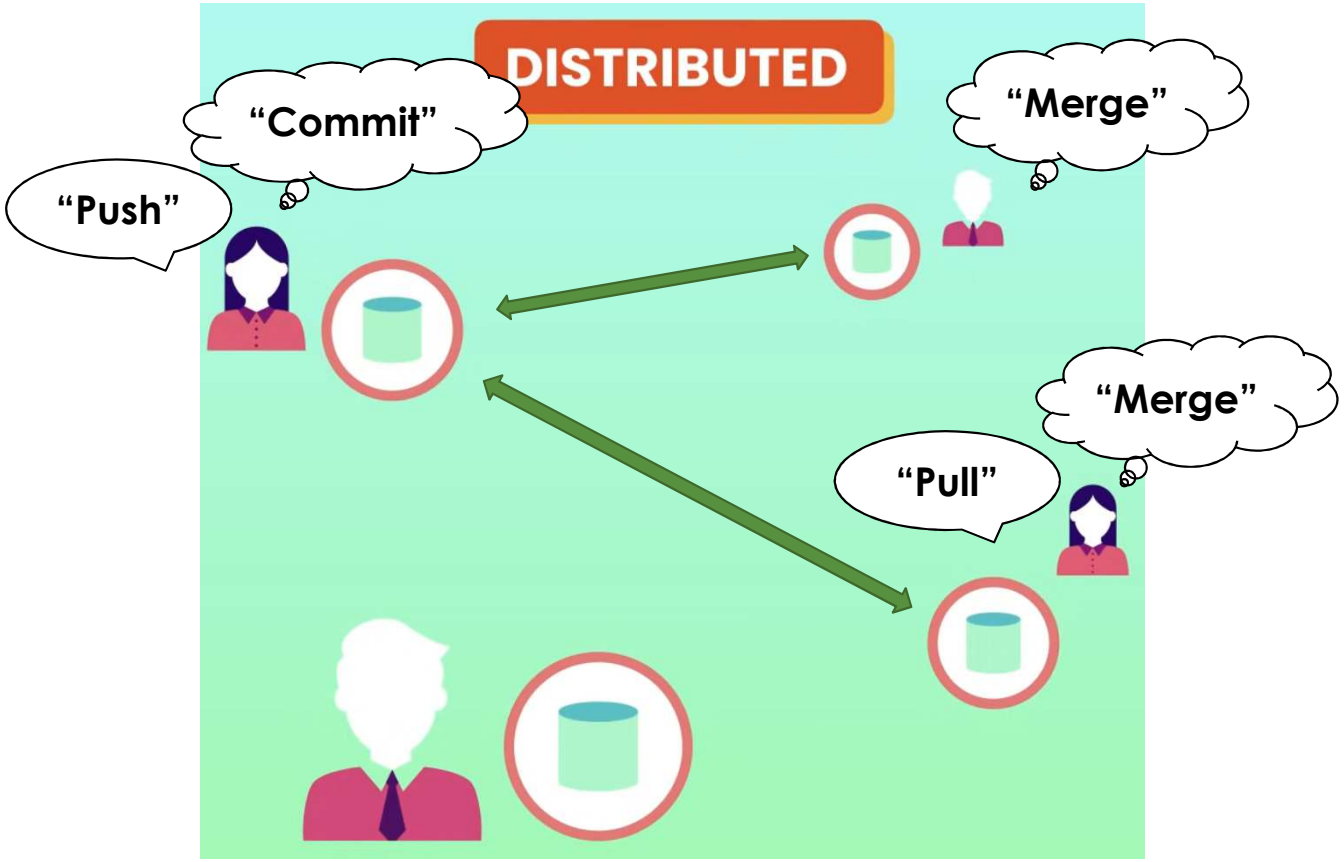
GIT & GITHUB

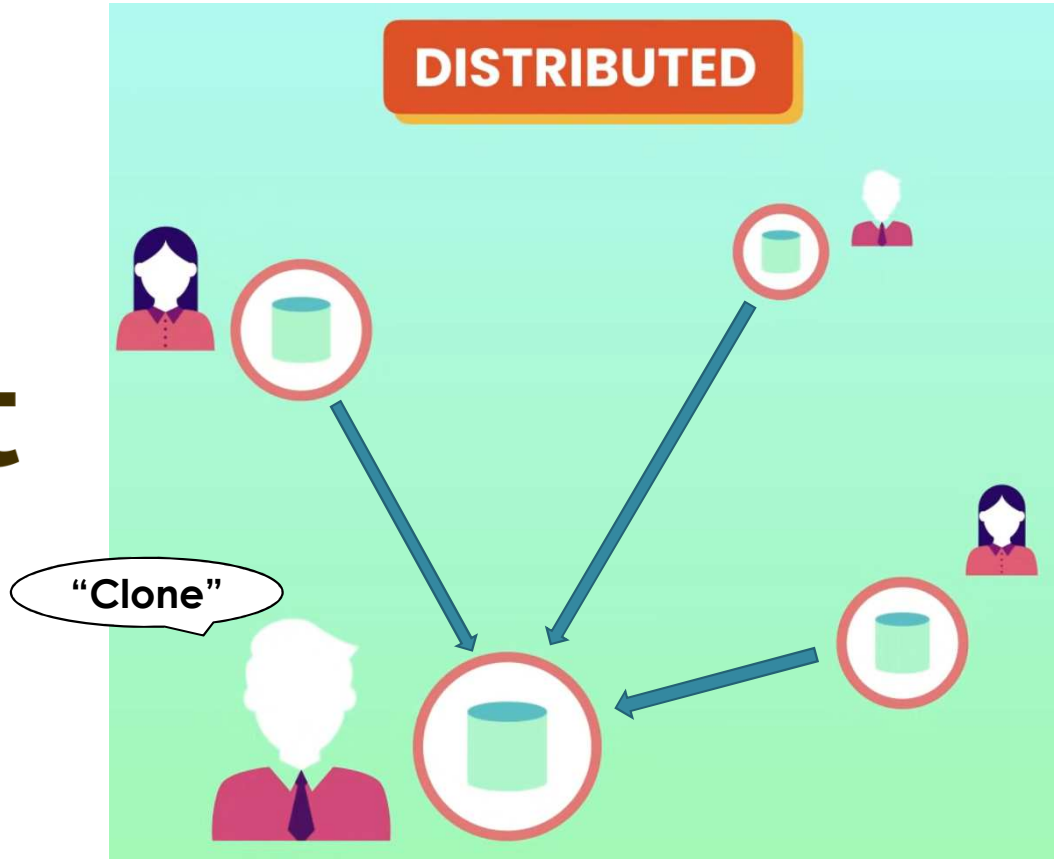
What are these and what is the difference?


CENTRALIZED
Subversion
Team Foundation Server


CENTRALIZED
Subversion
Team Foundation Server


DISTRIBUTED
“Pull”


DISTRIBUTED
“Commit”
“Push”
“Merge”
“Merge”
“Pull”



Trick question:

Where is "the project" being saved?

- **Free**
- **Open source**
- **Fast**
- **Cheap to locally merge/branch**
- Scalable
- Distributed
- Complex workflows allowed
- **Extremely popular**

- **Web-based hosting** platform (mirror) for files you choose to share from your local device
- **Packages all of git's version control features** with graphical user interface, user account levels etc
- Adds **extra project and team management features**, such as issues board, notifications, etc etc
- Opportunities for forming groups, networking and social coding.

- Point your favourite web browser to GitHub : **https://github.com/join**
- Set up a FREE and personal account in your own name
- Make note of your GitHub username or "handle"
- **Tell the teaching team your Github username so we can assign you to the EPI4932 repository.**

- Find a release for your personal device OS : https://git-scm.com/download/
- **Follow the installer instructions.**
- **[MacOS test if works] Terminal window, type : git version**
- **[Windows test if works] Command line or powershell window, type : git version**

CONNECTING RSTUDIO TO GITHUB

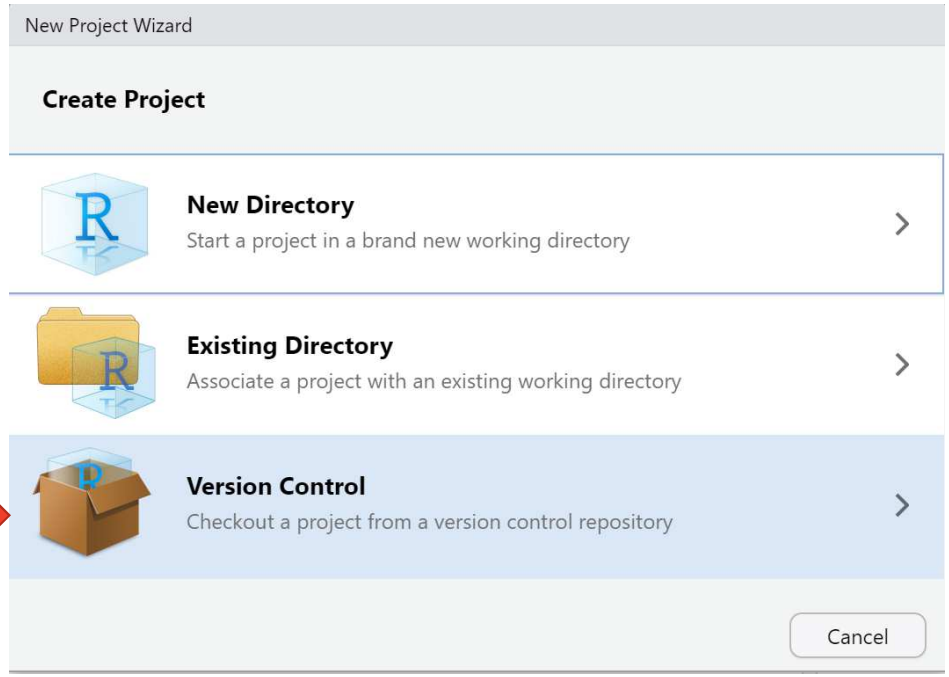
How to work with GitHub

- **First pull from a main version** or clone a project
- **Checkout a branch** of the main repo to do some edits
- Make your edits
- **Commit and push to the branch** version
- Review **changes**, **pull request** and **merge** back into main version**.**
- **Close down all the merged branches**
- **Pull from the main repo again** before starting a new cycle of work

RStudio®

ACTION TIME!

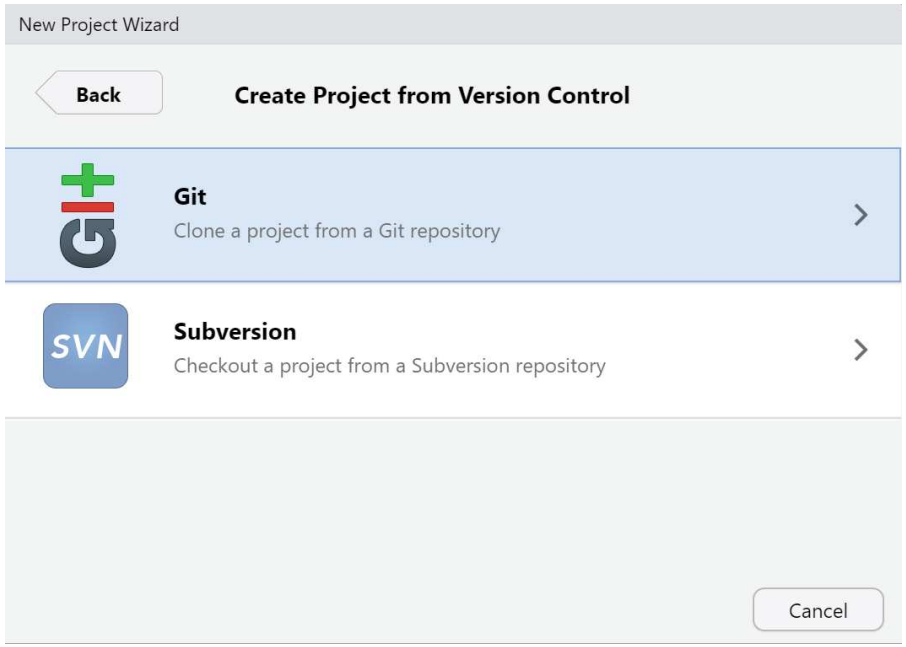
- Start up your local Rstudio app
- **File >> New project**
- **Create project using Version Control option**



RStudio®

ACTION TIME!

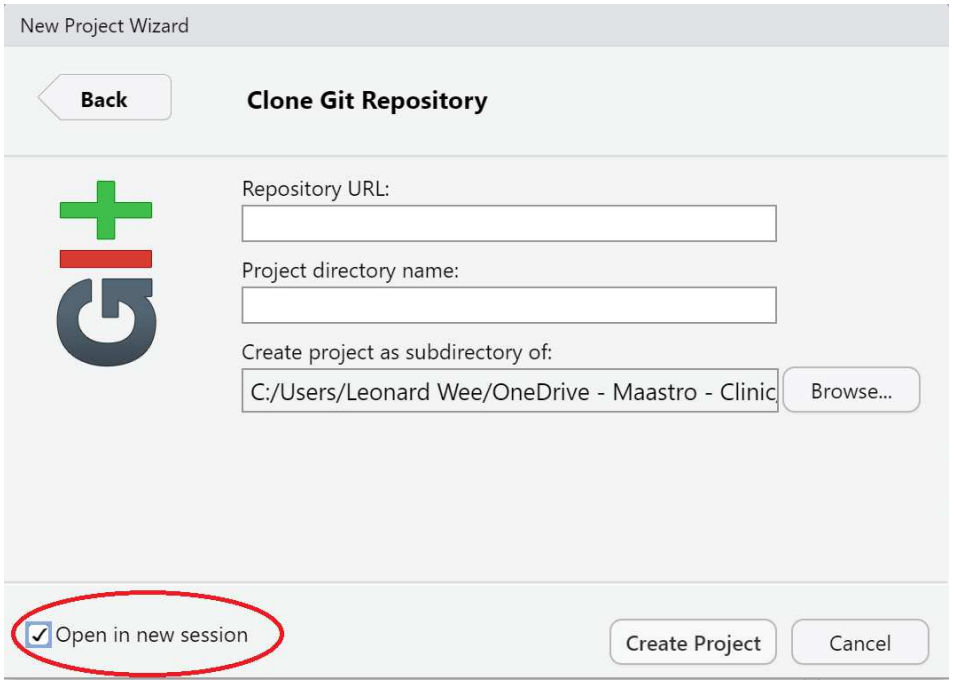
- **Clone an existing project from a Git repository.**



RStudio®

ACTION TIME!

- **Enter the github web link that your teaching team provides.**
- **Click the "Browse" button, please note well where you save this project.**
- **Check Open in "New Session"**
- **RStudio will ask for your Github username and password***
  - *If you set **multi-factor authentication** in GitHub, all next steps will ask for a Personal Access Token instead of password – ask the teaching team for help with this!



Personal access token setup (!)

- Read the docs
- https://docs.github.com/en/authentication/keeping-your-account-and-data-secure/creating-a-personal-access-token
- Personal settings -> Dev settings
- Personal Access Token (PAT) -> make a “classic” token
- Save the PAT as a text object somewhere safe – you’ll need to copy-paste it!
- Github username as usual
- Paste in the PAT in place of your password!

How to work with GitHub

✓**First pull from a main version** or clone a project

❑**Checkout a branch** of the main repo to do some edits

❑Make your edits

❑**Commit and push to the branch** version

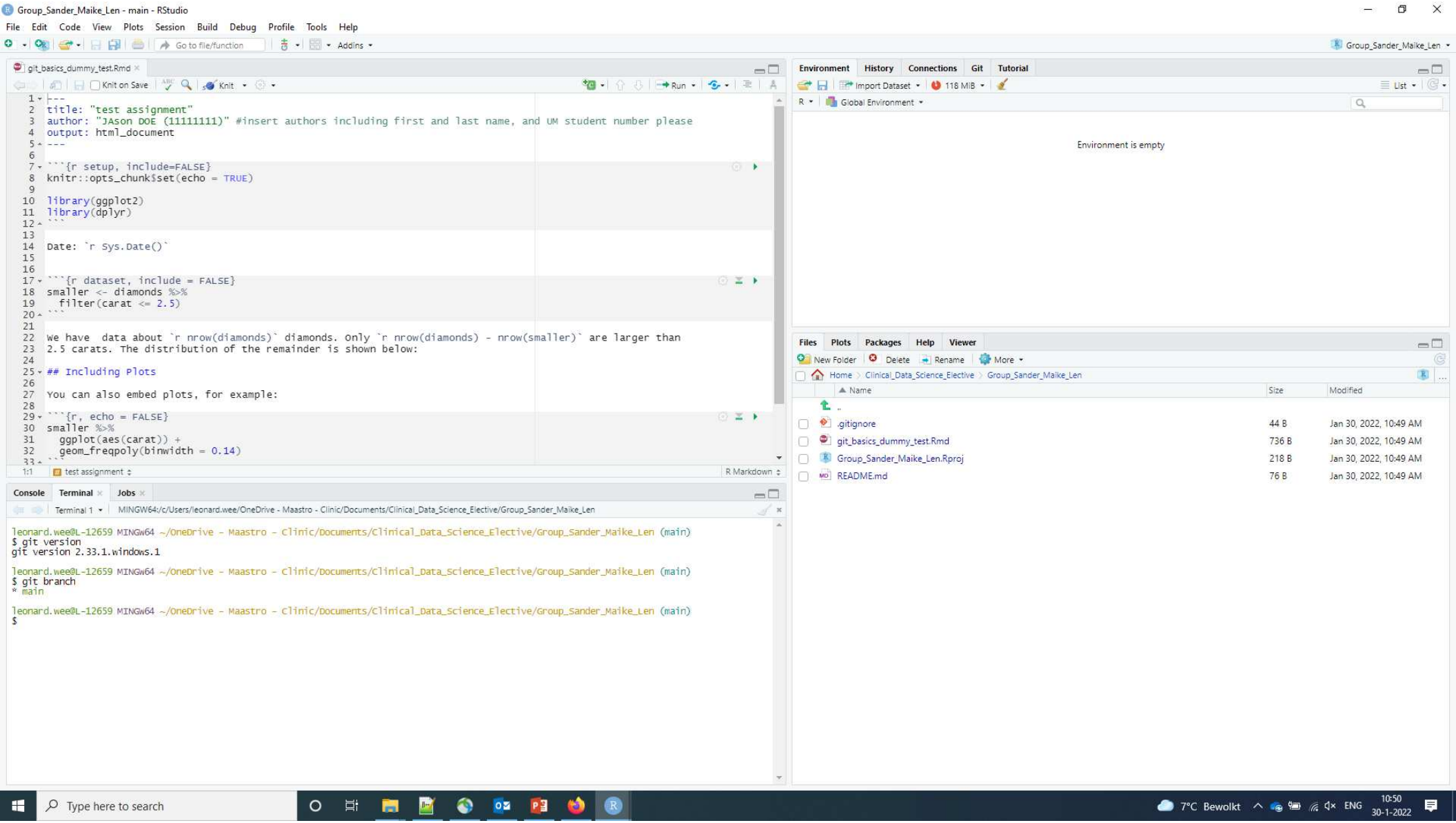
❑Review **changes**, **pull request** and **merge** back into main version**.**

❑**Close down all the merged branches**

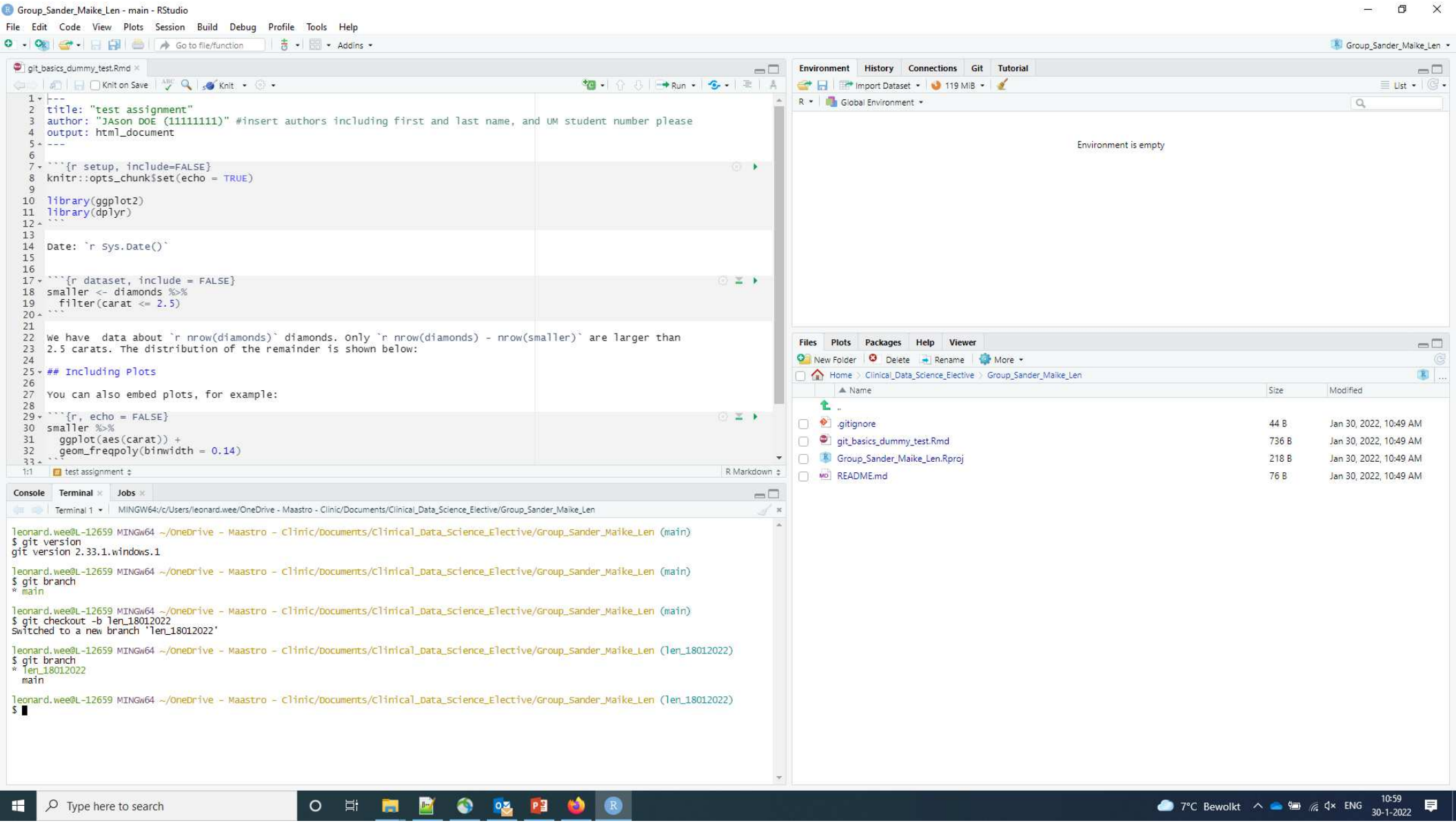
❑**Pull from the main repo again** before starting a new cycle of work

Setting up a branch

- **Terminal tab in RStudio**
- **Can you see that you are working in "main" branch at the moment?**
- **git checkout –b [choose_name_for_new_branch] (eg Len_18012022)**
- **Note that you just switched to your "branch" of the repo**
- **git branch**
- **Warning! I do not trust Rstudio entirely to tell me which branch I am on … I always use "git branch" to double check before I start working.**


Group_Sander_Maike_Len - main - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Group_Sander_Maike_Len
git_basics_dummy_test.Rmd
Knit on Save
Knit
Run
---
title: "test assignment"
author: "JAson DOE (11111111)" #insert authors including first and last name, and UM student number please
output: html_document

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
library(ggplot2)
library(dplyr)
```
Date: `r Sys.Date()`
```{r dataset, include = FALSE}
smaller <- diamonds %>%
filter(carat <= 2.5)
```
We have data about `r nrow(diamonds)` diamonds. Only `r nrow(diamonds) - nrow(smaller)` are larger than
2.5 carats. The distribution of the remainder is shown below:
Including Plots
You can also embed plots, for example:
```{r, echo = FALSE}
smaller %>%
ggplot(aes(carat)) +
geom_freqpoly(binwidth = 0.14)
```
1:1
test assignment
R Markdown
Environment History Connections Git Tutorial
Import Dataset
118 MiB
List
R
Global Environment
Environment is empty
Files Plots Packages Help Viewer
New Folder
Delete
Rename
More
Home
Clinical_Data_Science_Elective
Group_Sander_Maike_Len
Name
Size
Modified
.gitignore
44 B
Jan 30, 2022, 10:49 AM
git_basics_dummy_test.Rmd
736 B
Jan 30, 2022, 10:49 AM
Group_Sander_Maike_Len.Rproj
218 B
Jan 30, 2022, 10:49 AM
README.md
76 B
Jan 30, 2022, 10:49 AM
Console
Terminal
Jobs
Terminal 1
MINGW64:/c/Users/leonard.wee/OneDrive - Maastro - Clinic/Documents/Clinical_Data_Science_Elective/Group_Sander_Maike_Len
leonard.wee@L-12659 MINGW64 ~/OneDrive - Maastro - Clinic/Documents/Clinical_Data_Science_Elective/Group_Sander_Maike_Len (main)
$ git version
git version 2.33.1.windows.1
leonard.wee@L-12659 MINGW64 ~/OneDrive - Maastro - Clinic/Documents/Clinical_Data_Science_Elective/Group_Sander_Maike_Len (main)
$ git branch
* main
leonard.wee@L-12659 MINGW64 ~/OneDrive - Maastro - Clinic/Documents/Clinical_Data_Science_Elective/Group_Sander_Maike_Len (main)
$
Type here to search
7°C Bewolkt
ENG
10:50
30-1-2022

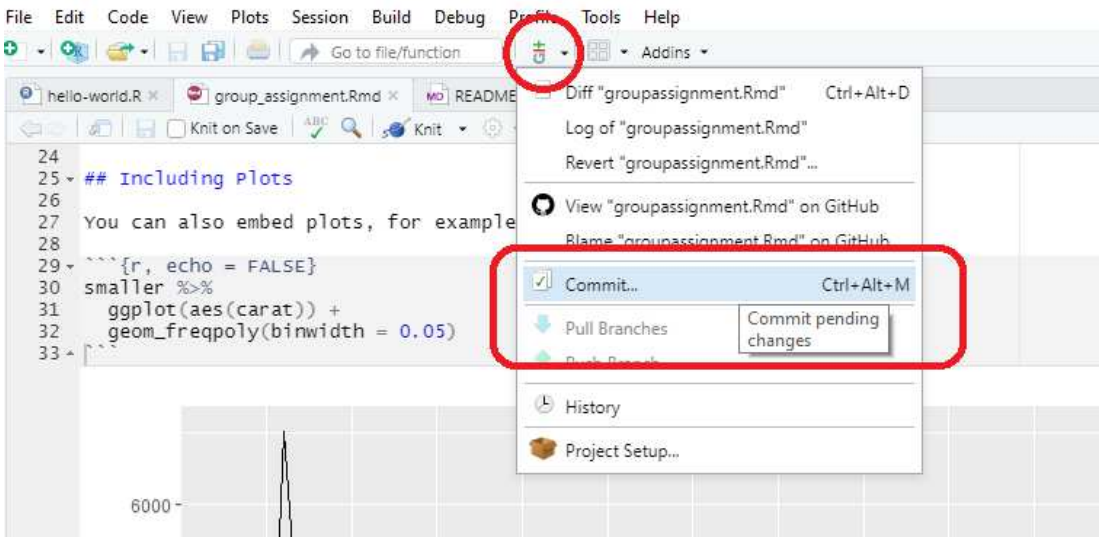

Group_Sander_Maike_Len - main - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Group_Sander_Maike_Len
git_basics_dummy_test.Rmd
Knit on Save
Knit
Run
---
title: "test assignment"
author: "JAson DOE (11111111)" #insert authors including first and last name, and UM student number please
output: html_document

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
library(ggplot2)
library(dplyr)
```
Date: `r Sys.Date()`
```{r dataset, include = FALSE}
smaller <- diamonds %>%
filter(carat <= 2.5)
```
We have data about `r nrow(diamonds)` diamonds. Only `r nrow(diamonds) - nrow(smaller)` are larger than
2.5 carats. The distribution of the remainder is shown below:
Including Plots
You can also embed plots, for example:
```{r, echo = FALSE}
smaller %>%
ggplot(aes(carat)) +
geom_freqpoly(binwidth = 0.14)
```
1:1
test assignment
R Markdown
Environment History Connections Git Tutorial
Import Dataset
119 MiB
List
R
Global Environment
Environment is empty
Files Plots Packages Help Viewer
New Folder
Delete
Rename
More
Home
Clinical_Data_Science_Elective
Group_Sander_Maike_Len
Name
Size
Modified
..
.gitignore
44 B
Jan 30, 2022, 10:49 AM
git_basics_dummy_test.Rmd
736 B
Jan 30, 2022, 10:49 AM
Group_Sander_Maike_Len.Rproj
218 B
Jan 30, 2022, 10:49 AM
README.md
76 B
Jan 30, 2022, 10:49 AM
Console
Terminal
Jobs
Terminal 1
MINGW64:/c/Users/leonard.wee/OneDrive - Maastro - Clinic/Documents/Clinical_Data_Science_Elective/Group_Sander_Maike_Len
leonard.wee@L-12659 MINGW64 ~/OneDrive - Maastro - Clinic/Documents/Clinical_Data_Science_Elective/Group_Sander_Maike_Len (main)
$ git version
git version 2.33.1.windows.1
leonard.wee@L-12659 MINGW64 ~/OneDrive - Maastro - Clinic/Documents/Clinical_Data_Science_Elective/Group_Sander_Maike_Len (main)
$ git branch
* main
leonard.wee@L-12659 MINGW64 ~/OneDrive - Maastro - Clinic/Documents/Clinical_Data_Science_Elective/Group_Sander_Maike_Len (main)
$ git checkout -b len_18012022
Switched to a new branch 'len_18012022'
leonard.wee@L-12659 MINGW64 ~/OneDrive - Maastro - Clinic/Documents/Clinical_Data_Science_Elective/Group_Sander_Maike_Len (len_18012022)
$ git branch
* len_18012022
main
leonard.wee@L-12659 MINGW64 ~/OneDrive - Maastro - Clinic/Documents/Clinical_Data_Science_Elective/Group_Sander_Maike_Len (len_18012022)
$
Type here to search
7°C Bewolkt
ENG
10:59
30-1-2022

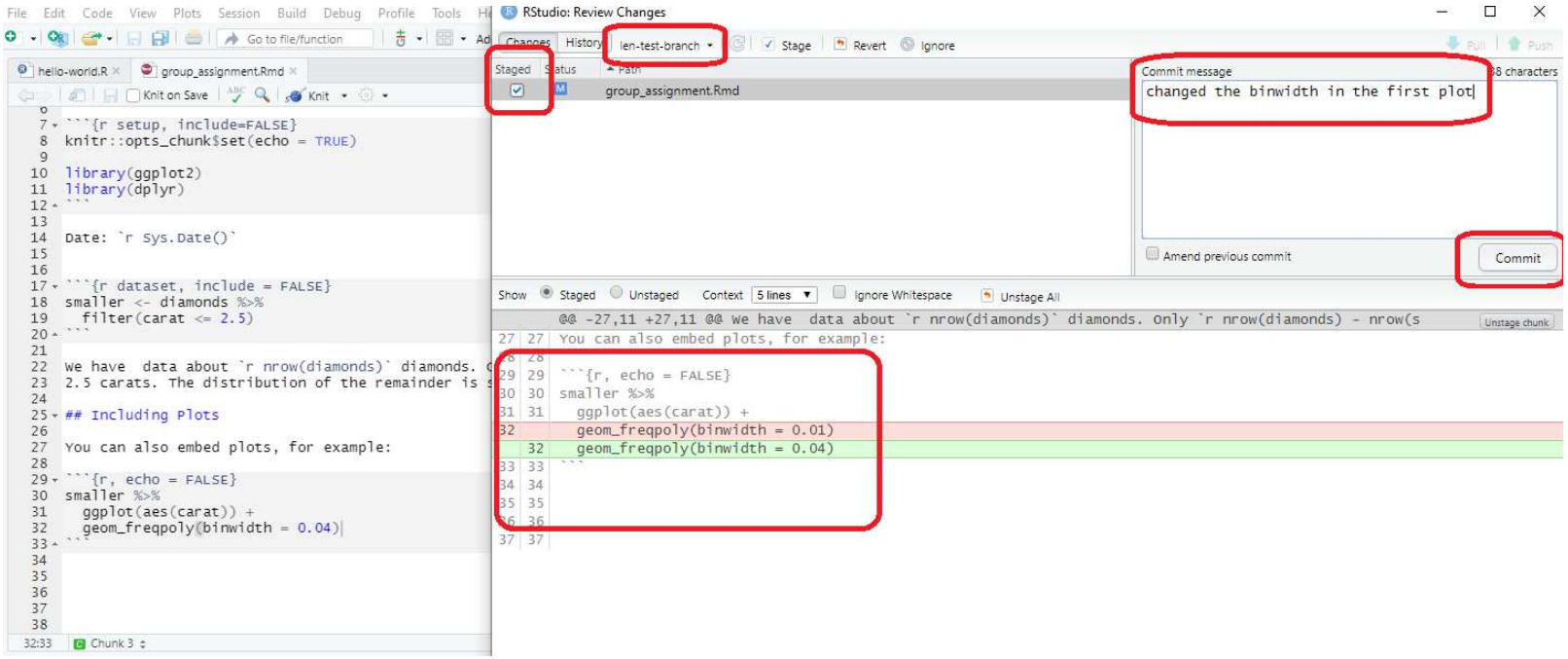
RStudio®

ACTION TIME!

- Make some edits in <u>your branch</u> of git_basics_dummy_test.Rmd
  - Eg Change the name at the top
- Committing your changes :
  - Stage a file for commit (checkbox)
  - Incl a decent comment on edits
  - Click “commit” below
  - git push --set-upstream origin [the_name_of_your_branch]



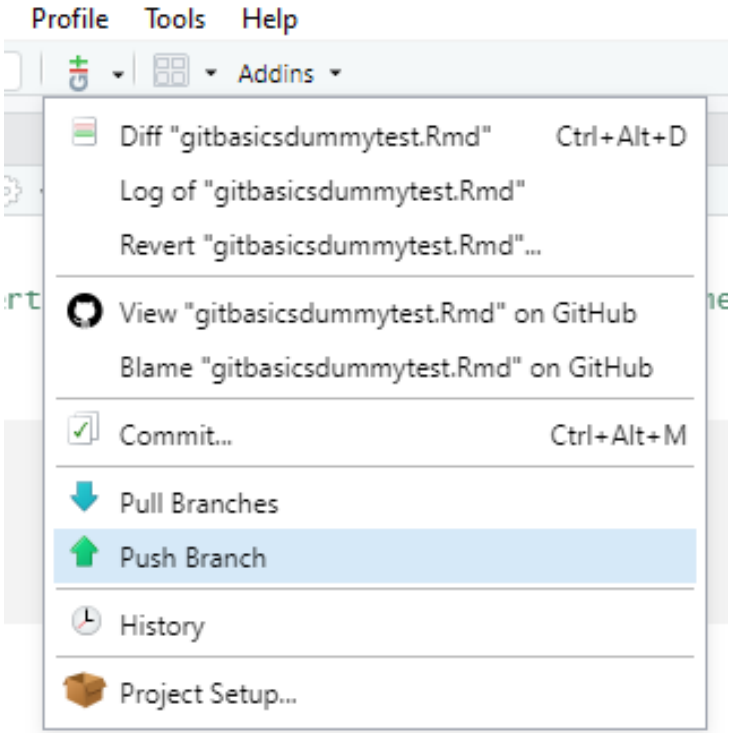
R Studio®

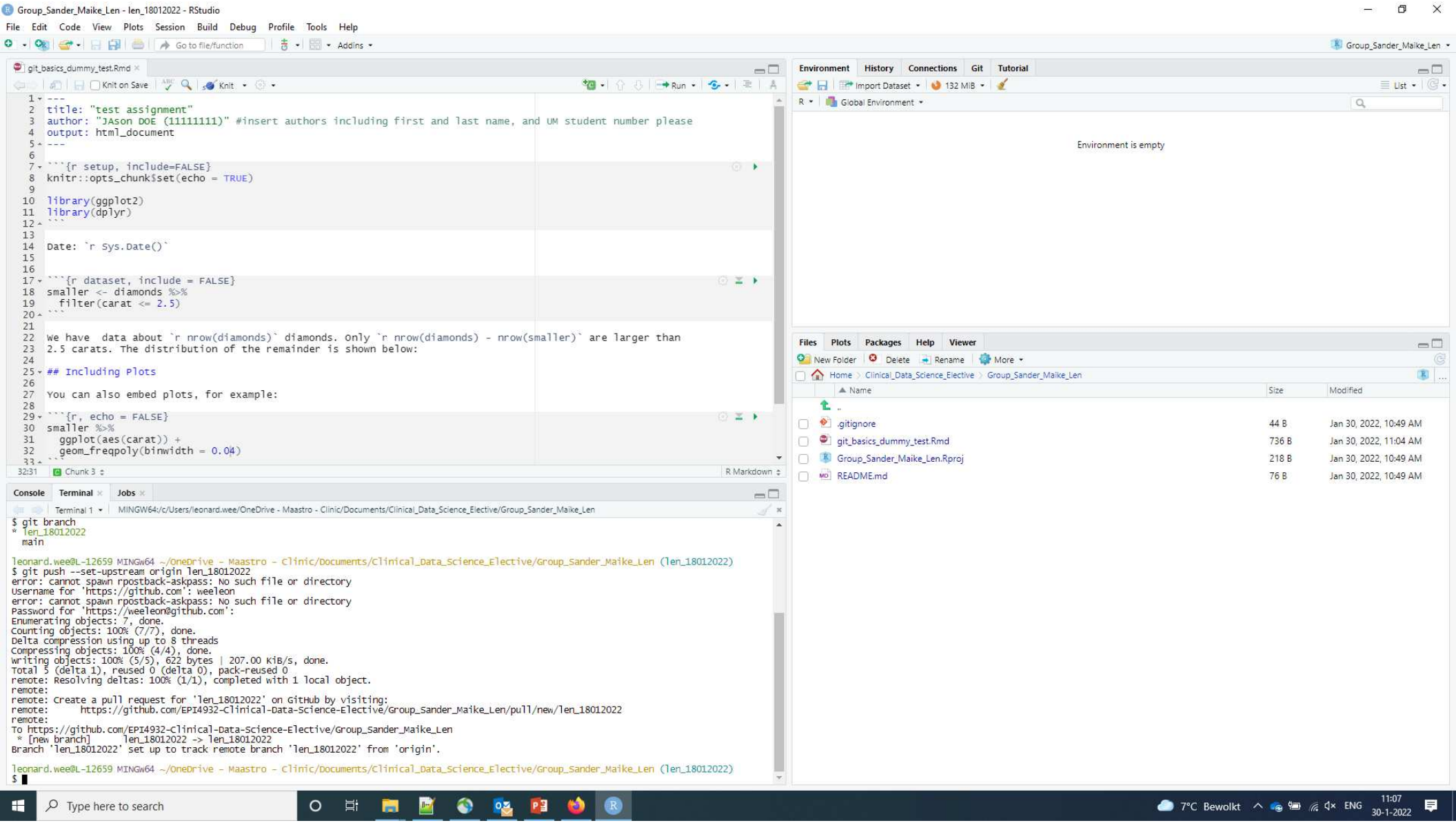

RStudio: Review Changes
Changes
History
len-test-branch
Stage
Revert
Ignore
Staged
Status
Path
group_assignment.Rmd
Commit message
changed the binwidth in the first plot
Amend previous commit
Commit
Show
Staged
Unstaged
Context
5 lines
Ignore Whitespace
Unstage All
@@ -27,11 +27,11 @@ We have data about `r nrow(diamonds)` diamonds. Only `r nrow(diamonds) - nrow(s
You can also embed plots, for example:
```{r, echo = FALSE}
smaller %>%
ggplot(aes(carat)) +
geom_freqpoly(binwidth = 0.01)
geom_freqpoly(binwidth = 0.04)
```
hello-world.R
group_assignment.Rmd
Knit on Save
Knit
Date: `r Sys.Date()`
```{r dataset, include = FALSE}
smaller <- diamonds %>%
filter(carat <= 2.5)
```
We have data about `r nrow(diamonds)` diamonds.
2.5 carats. The distribution of the remainder is
Including Plots
You can also embed plots, for example:
```{r, echo = FALSE}
smaller %>%
ggplot(aes(carat)) +
geom_freqpoly(binwidth = 0.04)
```
32:33
Chunk 3

RStudio®

ACTION TIME!

- **Make MORE edits in <u>your branch</u> of the file : group_assignment.Rmd**
  - **Change the bin width in the plot**
- **Or add a new script file, or anything!**
- **Committing and push your changes :**
  - **Stage a file for commit (checkbox)**
  - **Incl a comment on edits**
  - **Click commit**
  - **(<u>shortcut now available</u>) click green up-arrow for "PUSH"**




Group_Sander_Maike_Len - len_18012022 - RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function
Addins
Group_Sander_Maike_Len
git_basics_dummy_test.Rmd
Knit on Save
Knit
Run
1 ---
2 title: "test assignment"
3 author: "JAson DOE (11111111)" #insert authors including first and last name, and UM student number please
4 output: html_document
5 ---
6
7 ```{r setup, include=FALSE}
8 knitr::opts_chunk$set(echo = TRUE)
9
10 library(ggplot2)
11 library(dplyr)
12 ```
13
14 Date: `r Sys.Date()`
15
16
17 ```{r dataset, include = FALSE}
18 smaller <- diamonds %>%
19 filter(carat <= 2.5)
20 ```
21
22 We have data about `r nrow(diamonds)` diamonds. Only `r nrow(diamonds) - nrow(smaller)` are larger than
23 2.5 carats. The distribution of the remainder is shown below:
24
25 ## Including Plots
26
27 You can also embed plots, for example:
28
29 ```{r, echo = FALSE}
30 smaller %>%
31 ggplot(aes(carat)) +
32 geom_freqpoly(binwidth = 0.04)
33 ```
32:31
Chunk 3
R Markdown
Console
Terminal
Jobs
Terminal 1
MINGW64:/c/Users/leonard.wee/OneDrive - Maastro - Clinic/Documents/Clinical_Data_Science_Elective/Group_Sander_Maike_Len
$ git branch
* len_18012022
main
leonard.wee@L-12659 MINGW64 ~/OneDrive - Maastro - Clinic/Documents/Clinical_Data_Science_Elective/Group_Sander_Maike_Len (len_18012022)
$ git push --set-upstream origin len_18012022
error: cannot spawn rpostback-askpass: No such file or directory
Username for 'https://github.com': weeleon
error: cannot spawn rpostback-askpass: No such file or directory
Password for 'https://weeleon@github.com':
Enumerating objects: 7, done.
Counting objects: 100% (7/7), done.
Delta compression using up to 8 threads
Compressing objects: 100% (4/4), done.
Writing objects: 100% (5/5), 622 bytes | 207.00 KiB/s, done.
Total 5 (delta 1), reused 0 (delta 0), pack-reused 0
remote: Resolving deltas: 100% (1/1), completed with 1 local object.
remote:
remote: Create a pull request for 'len_18012022' on GitHub by visiting:
remote: https://github.com/EPI4932-Clinical-Data-Science-Elective/Group_Sander_Maike_Len/pull/new/len_18012022
remote:
To https://github.com/EPI4932-Clinical-Data-Science-Elective/Group_Sander_Maike_Len
* [new branch] len_18012022 -> len_18012022
Branch 'len_18012022' set up to track remote branch 'len_18012022' from 'origin'.
leonard.wee@L-12659 MINGW64 ~/OneDrive - Maastro - Clinic/Documents/Clinical_Data_Science_Elective/Group_Sander_Maike_Len (len_18012022)
$
Environment
History
Connections
Git
Tutorial
Import Dataset
132 MiB
List
R
Global Environment
Environment is empty
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
Home
Clinical_Data_Science_Elective
Group_Sander_Maike_Len
Name
Size
Modified
.gitignore 44 B Jan 30, 2022, 10:49 AM
git_basics_dummy_test.Rmd 736 B Jan 30, 2022, 11:04 AM
Group_Sander_Maike_Len.Rproj 218 B Jan 30, 2022, 10:49 AM
README.md 76 B Jan 30, 2022, 10:49 AM
Type here to search
7°C Bewolkt
ENG
11:07
30-1-2022

How to work with GitHub

✓**First pull from a main version** or clone a project

✓**Checkout a branch** of the main repo to do some edits

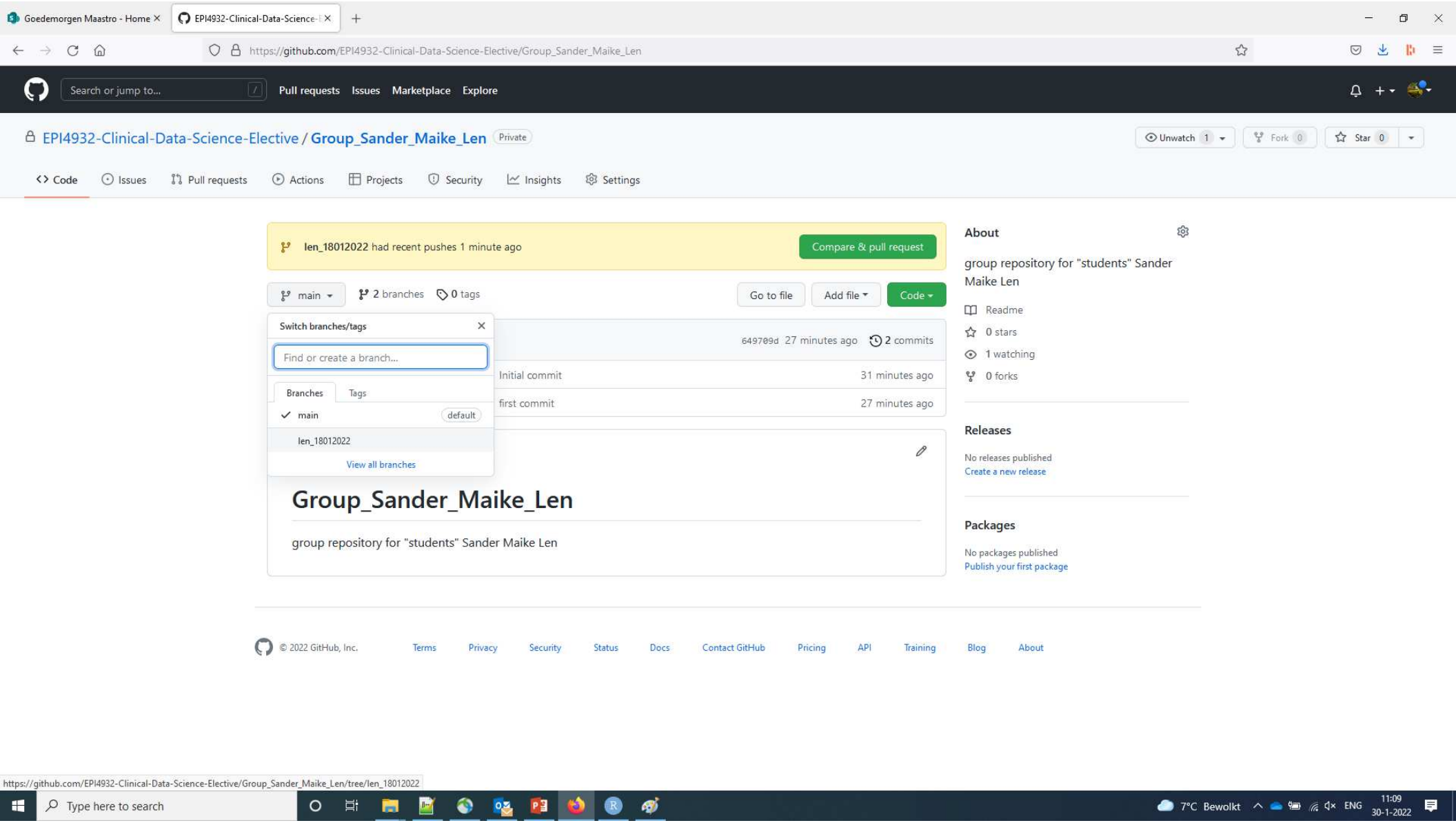
✓Make your edits

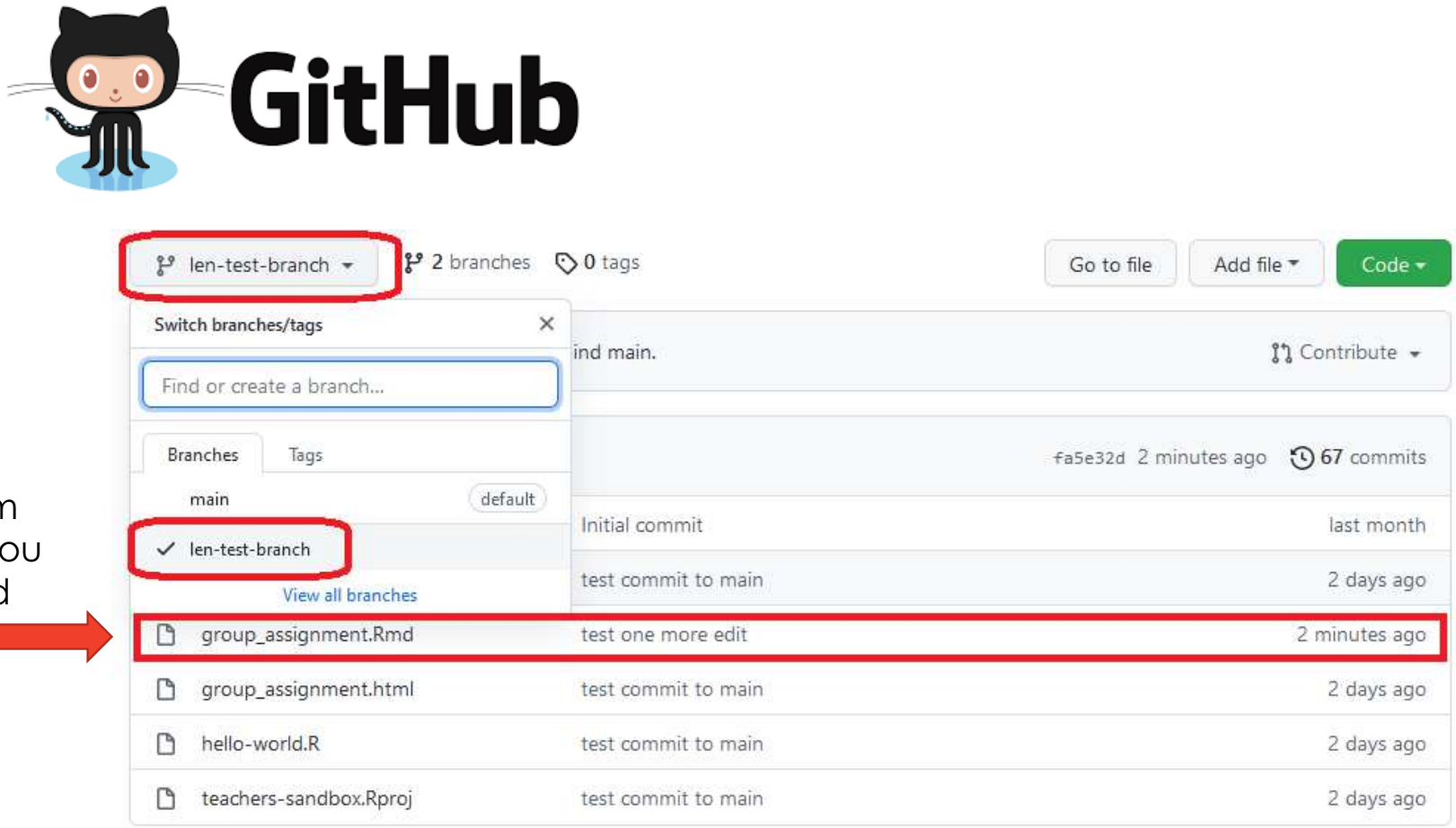
✓**Commit and push to the branch** version

❑Review **changes**, **pull request** and **merge** back into main version**.**

❑**Close down all the merged branches**

❑**Pull from the main repo again** before starting a new cycle of work


Goedemorgen Maastro - Home
EPI4932-Clinical-Data-Science-E
https://github.com/EPI4932-Clinical-Data-Science-Elective/Group_Sander_Maike_Len
Search or jump to...
Pull requests
Issues
Marketplace
Explore
EPI4932-Clinical-Data-Science-Elective / Group_Sander_Maike_Len
Private
Unwatch 1
Fork 0
Star 0
Code
Issues
Pull requests
Actions
Projects
Security
Insights
Settings
len_18012022 had recent pushes 1 minute ago
Compare & pull request
main
2 branches
0 tags
Go to file
Add file
Code
Switch branches/tags
Find or create a branch...
Branches
Tags
main
default
len_18012022
View all branches
649709d 27 minutes ago
2 commits
Initial commit
31 minutes ago
first commit
27 minutes ago
Group_Sander_Maike_Len
group repository for "students" Sander Maike Len
About
group repository for "students" Sander Maike Len
Readme
0 stars
1 watching
0 forks
Releases
No releases published
Create a new release
Packages
No packages published
Publish your first package
© 2022 GitHub, Inc.
Terms
Privacy
Security
Status
Docs
Contact GitHub
Pricing
API
Training
Blog
About
https://github.com/EPI4932-Clinical-Data-Science-Elective/Group_Sander_Maike_Len/tree/len_18012022
Type here to search
7°C Bewolkt
ENG
11:09
30-1-2022



Confirm what you pushed here

main 2 branches 0 tags

Switch branches/tags

Find or create a branch...

Branches Tags

main default

len-test-branch

View all branches

4932-Clinical-Data-Science-El

	Initial commit
	test commit to main
	test one more edit
	test commit to main
hello-world.R	test commit to main
teachers-sandbox.Rproj	test commit to main

- **View all branches**
- **Review changes you made**
- **Create a pull request to merge these changes to MAIN.**

Open a pull request

Create a new pull request by comparing changes across two branches. If you need to, you can also compare across forks.

base: main ← compare: len-test-branch ✓ Able to merge. These branches can be automatically merged.

1 commit | 1 file changed | 1 contributor

Commits on Jan 22, 2022

blah

weeleon committed 3 minutes ago

e53de39

Showing 1 changed file with **1 addition** and **1 deletion**.

Split | Unified

2 group_assignment.Rmd

```
@@ -29,7 +29,7 @@ You can also embed plots, for example:
29 29   ```{r, echo = FALSE}
30 30   smaller %>%
31 31     ggplot(aes(carat)) +
32    -   geom_freqpoly(binwidth = 0.04)
   32 +   geom_freqpoly(binwidth = 0.03)
33 33   ```
34 34
35 35
```

Open a pull request

Create a new pull request by comparing changes across two branches. If you need to, you can also compare across forks.

base: main ← compare: len_18012022 ✓ Able to merge. These branches can be automatically merged.

Len 18012022

Write Preview

I changed the autthors name and edited the plot bin width

Attach files by dragging & dropping, selecting or pasting them.

Create pull request

Remember, contributions to this repository should follow our GitHub Community Guidelines.

Reviewers
No reviews

Assignees
No one—assign yourself

Labels
None yet

Projects
None yet

Milestone
No milestone

Linked issues
Use Closing keywords in the description to automatically close issues

Len 18012022 #1

Open **weeleon** wants to merge 2 commits into `main` from `len_18012022`

Conversation 0 | Commits 2 | Checks 0 | Files changed 3 | +19 −2

weeleon commented now

I changed the autthors name and edited the plot bin width

weeleon added 2 commits 17 minutes ago

`Len changed the bin width` e3ae04f

`Name change` 41e48dd

Add more commits by pushing to the `len_18012022` branch on **EPI4932-Clinical-Data-Science-Elective/Group_Sander_Maike_Len**.

Continuous integration has not been set up

GitHub Actions and several other apps can be used to automatically catch bugs and enforce style.

This branch has no conflicts with the base branch

Merging can be performed automatically.

Merge pull request | You can also open this in GitHub Desktop or view command line instructions.

Reviewers

No reviews

Assignees

No one—assign yourself

Labels

None yet

Projects

None yet

Milestone

No milestone

Linked issues

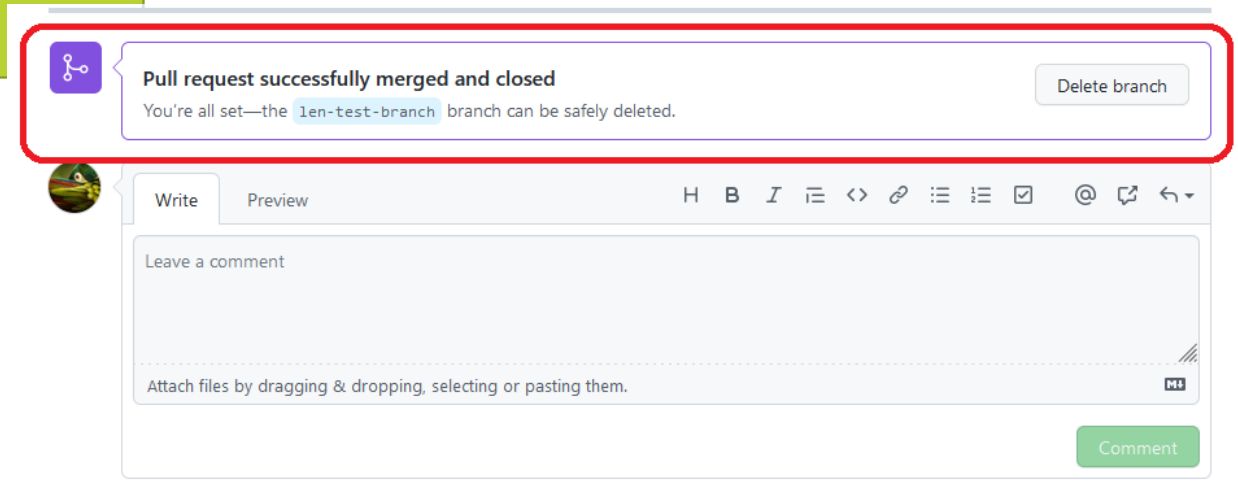
Successfully merging this pull request may close these issues.

None yet

GitHub

ACTION TIME!

- Check and resolve differences
- Merge branches into main
- DELETE the merged branches from GitHub!



How to work with GitHub

✓**First pull from a main version** or clone a project

✓**Checkout a branch** of the main repo to do some edits

✓Make your edits

✓**Commit and push to the branch** version

✓Review **changes**, **pull request** and **merge** back into main version

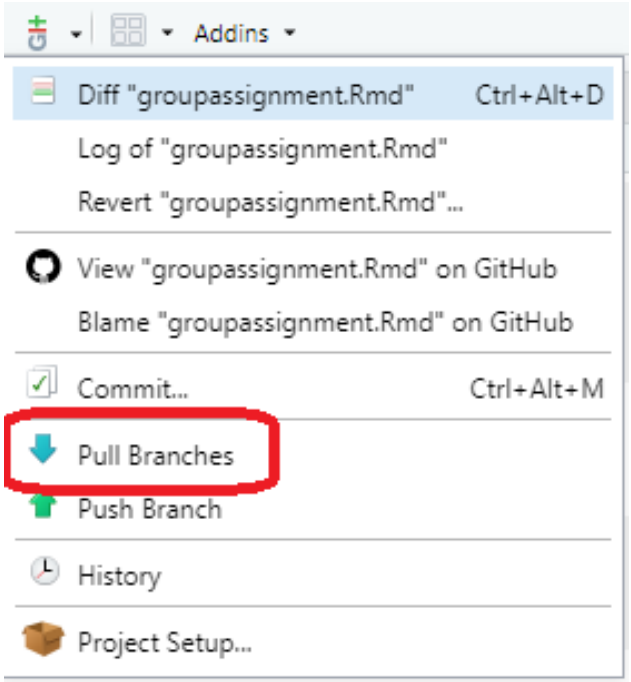
✓**Delete obsolete branches in GitHub and on RStudio**

❑**Pull from the main repo again** before starting a new cycle of work

RStudio®

ACTION TIME!

- git checkout main
- Git button -> blue arrow
- PULL the recently solved main
- git branch –d [your_obsolete_branch]
- git checkout –b [a_new_branch]
- Make changes to your project!
- Rinse repeat etc.



How to work with GitHub

- ✓**First pull from a main version** or clone a project
- ✓**Checkout a branch** of the main repo to do some edits
- ✓Make your edits
- ✓**Commit and push to the branch** version
- ✓Review **changes**, **pull request** and **merge** back into main version
- ✓**Delete obsolete branches in GitHub and on Rstudio**
- ✓**Pull from updated MAIN again** when starting a new cycle of work

Remember the git cycle

- **1. RStudio : git checkout main (if first time on a project, get new project from GitHub)**
- **2. RStudio : First PULL (blue down arrow) from the MAIN branch, unless cloning new project**
- **3. RStudio : git branch –d [a_old_merged_branch] (delete obsolete old branch, if any)**
- **4. RStudio : git checkout –b [the_new_branch] (makes a new branch for current round of edits)**
- **5. RStudio : Make whatever changes as you wish, commit and push (green up arrow)**
  - **git push --set-upstream origin [the_new_branch] (first time push of new branch, the green arrow may be greyed out)**
- **6. GitHub : Create a pull request to merge your changes into main.**
- **7. GitHub : Discuss and review changes in each branch then merge into the main.**
- **8. GitHub : Delete all branches from GitHub leaving ONLY THE MAIN.**
- **9. Repeat again from Step 1.**

Some DO's and DON'Ts

- **DO use branches**
  - **It seems like a chore until something goes wrong and then you wished you used branches**
- **DON'T edit in main branch directly!**
  - **It seems like a pain until something goes wrong and you badly mess up your colleagues**
- **DO develop the habit of working in RStudio with GitHub version control**
  - **RStudio integration with GitHub seems a little weird at times, but it is improving**
  - **Group Git repo MUST SHOW AN AUDIT TRAIL of individual contribution(s) plus group abstract.**
- **DO keep things relatively simple for now**
  - **We have not touched on reverts and many other (quite complicated) git actions.**
- **DO look up the internet and ask for help**
  - **GitHub and R are very popular, so someone somewhere knows how to do what you want**

KNOWLEDGE CHECK

(1)

USE VERSION CONTROL TO TRACK CHANGES IN WORK

CREATE, SHARE AND MERGE CONTENT WITH GITHUB

INTEGRATE GITHUB WITH RSTUDIO

(2)

USE VERSION CONTROL TO TRACK CHANGES IN WORK

CREATE, SHARE AND MERGE CONTENT WITH GITHUB

INTEGRATE GITHUB WITH RSTUDIO

(3)

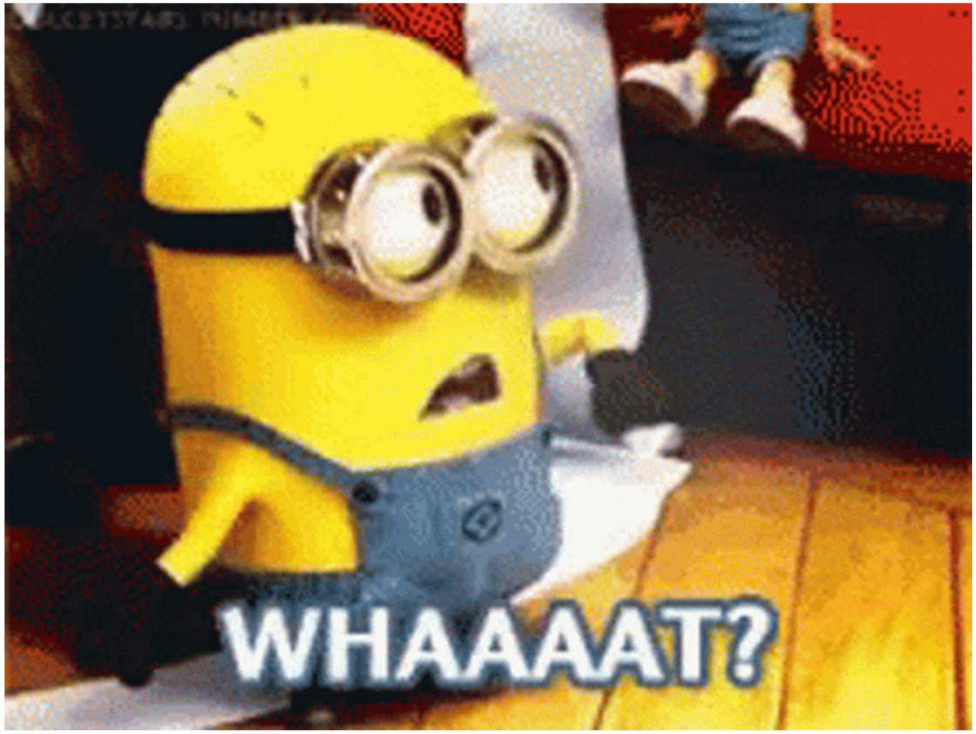
USE VERSION CONTROL TO TRACK CHANGES IN WORK

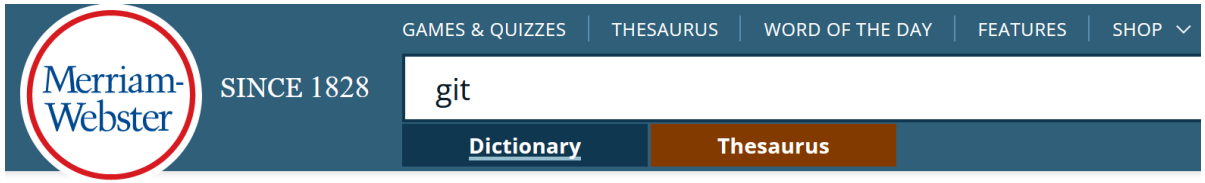
CREATE, SHARE AND MERGE CONTENT WITH GITHUB

INTEGRATE GITHUB WITH RSTUDIO

Questions??

You will be using GitHub and RStudio to complete your group assignment.





git noun

Save Word

\ ˈgit \

Definition of *git* (Entry 1 of 2)

British

: a foolish or worthless person

git

Definition of *git* (Entry 2 of 2)

dialectal variant of GET

Synonyms for *git*

Synonyms: Noun

berk [*British*], booby, charlie (*also* charley) [*British*], cuckoo, ding-a-ling, ding-dong, dingbat, dipstick, doofus [*slang*], featherhead, fool, goose, half-wit, jackass, lunatic, mooncalf, nincompoop, ninny, ninnyhammer, nit [*chiefly British*], nitwit, nut, nutcase, simp, simpleton, turkey, yo-yo

Extra lesson:

English slang for non-native speakers

Just for fun only …

Thank you for your attention today!