

Cleaning up your code and providing it in github - for theses

1. Prepare a clean final version of your code

1.1 R code

- Take a new empty Rmarkdown file.
- Include in the name of the markdown file
 - a. In the beginning the date when you finished your thesis/project in a year-month-date format
 - b. Your full name
 - c. A short version of your thesis/ project title
 - d. eg. 2022-11-11_HansMustermensch_Thesisbla.Rmd
- If it was a long project (e.g. Masterthesis) it can also be several Rmarkdown files, then name them in the order of their execution. (eg. 2022-11-11_HansMustermensch_Thesisbla_1.Rmd, 2022-11-11_HansMustermensch_Thesisbla_2.Rmd)

In the header include:

- the full title of your thesis/project
- yourself as the author
- your supervisor
- the beginning and end date of your project
- a short abstract of your project (2-3 sentences)

```
1 ---
2 title: "My very cool thesis"
3 output: html_document
4 author:
5   - name: "me myself and I"
6   - name: "Supervision: MySupervisor"
7 date: "2022-04-20 - 2022-11-11"
8 abstract: "Blablabla"
9 ---
```

Libraries

- The first code chunk after the setup chunk should be a chunk called libraries. Here you specify all libraries used in your project.
- The very last code chunk of the Rmarkdown should perform sessionInfo() to state the versions of each library that was used.

```
# Session Information

```{r, session_info, include=TRUE, echo=TRUE, results='markup'}
sessionInfo()
```
```

Input files

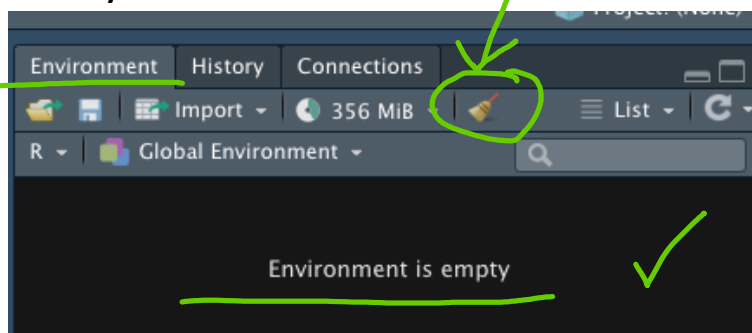
- The next chunk should be loading all needed input files.
- If you did preprocessing in bash or galaxy, load preprocessed files here and look in the sections below, about bash and galaxy

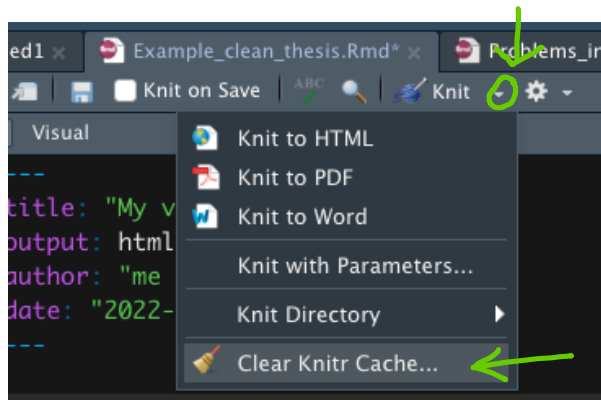
Your code

- Now add all the code necessary to produce all your results (as shown in the results part of your thesis) including all plots and tables starting from the input files that you received or downloaded initially.
- Look at these coding style rules and follow them:
<https://style.tidyverse.org/files.html>
- Comment you code in a way that other people can understand what you did.
- Structure your code in reasonable code chunks. Each chunk does one specific analysis.
- Remove all unnecessary code.
- Remove all out commented code.

Your markdown

- You should knit your markdown to html or pdf in the end.
- **Before you knit clear the knit cache and the environment!!!**





- Use headers (# First big topic) and subheaders (## small part of first topic) to structure your markdown.
- Write quick explanations of why you are doing what you are doing in front of the chunks where you do it (e.g. PCA plots are made to look at variability between samples)

1.2 bash scripts

If you used bash scripts in your thesis prior to further analysis in R include all your .sh files in the rmarkdown as not executed bash chunk.

- Although the bash chunks are not executed keep them in the order of execution.
- Explain the input and output files either in comments in the bash script or between the chunks in the R markdown.

```
```{bash, eval=F}
|
```
```

- Mention the versions of all tools you use in comments.

1.3 galaxy steps

If you preprocessed your input files with galaxy. Copy the description of all steps performed in galaxy form the methods part of your thesis and put it in the beginning of your markdown script.

1.4 other languages

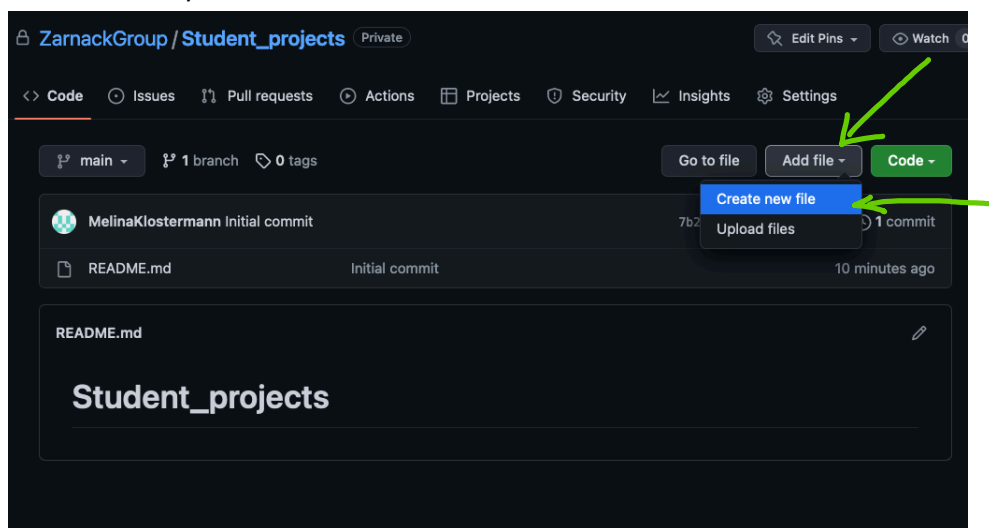
Talk to your supervisor!

2. Upload the finished markdown in github

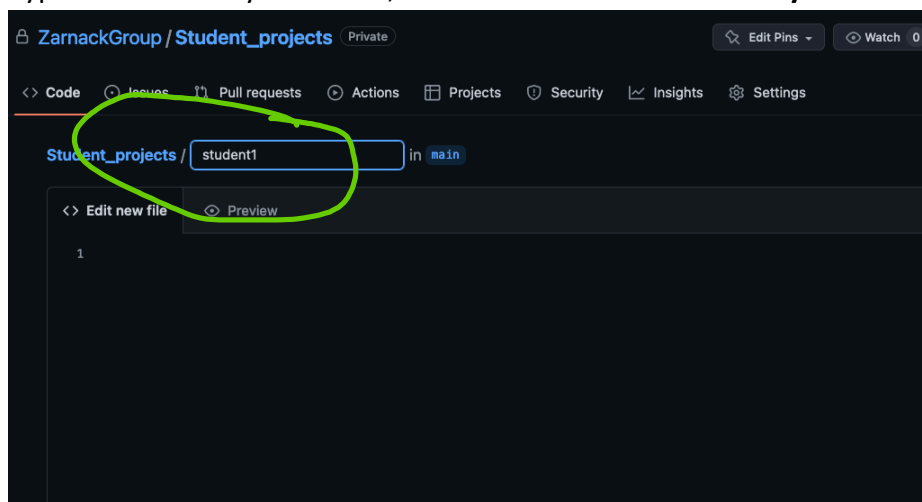
- If you haven't already, create a github account for yourself at <https://github.com/>
- Ask your supervisor to add you as a collaborator to the Zarnack_group/Student projects repository.
- When you are added go to https://github.com/ZarnackGroup/Student_projects/

2.1 make a folder in github repro

- First you have to make a new folder it should have the same title as you Rmarkdown file.
- Go to Add file/ Create new file

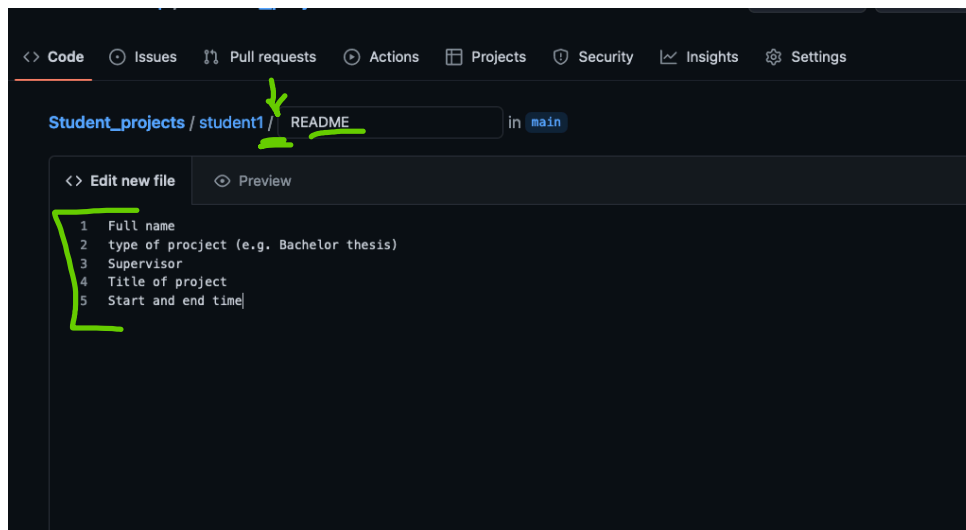


- Type the name of your folder, which should be **the same as your markdown**.

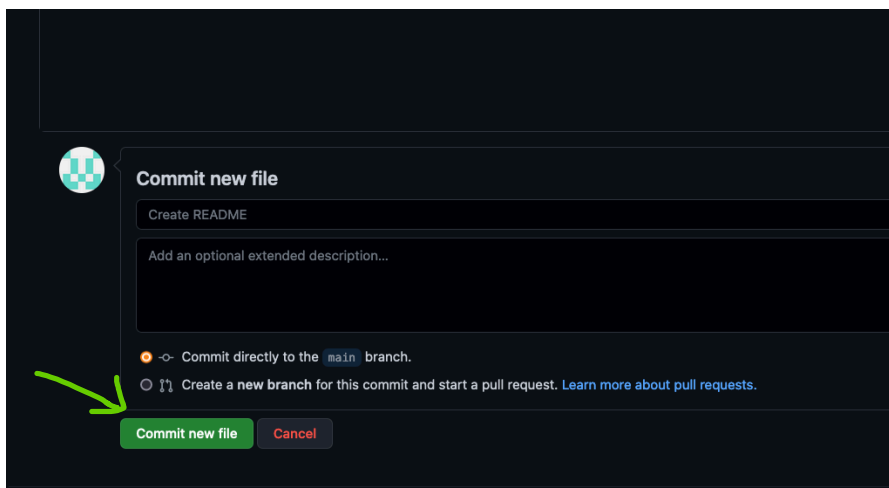


- **Then add a / after the name.** This will create a folder.
- In this folder make a README file that includes
 - Your full name
 - type of project (e.g. Bachelor thesis)

- o Supervisor
- o Title of project
- o Start and end time

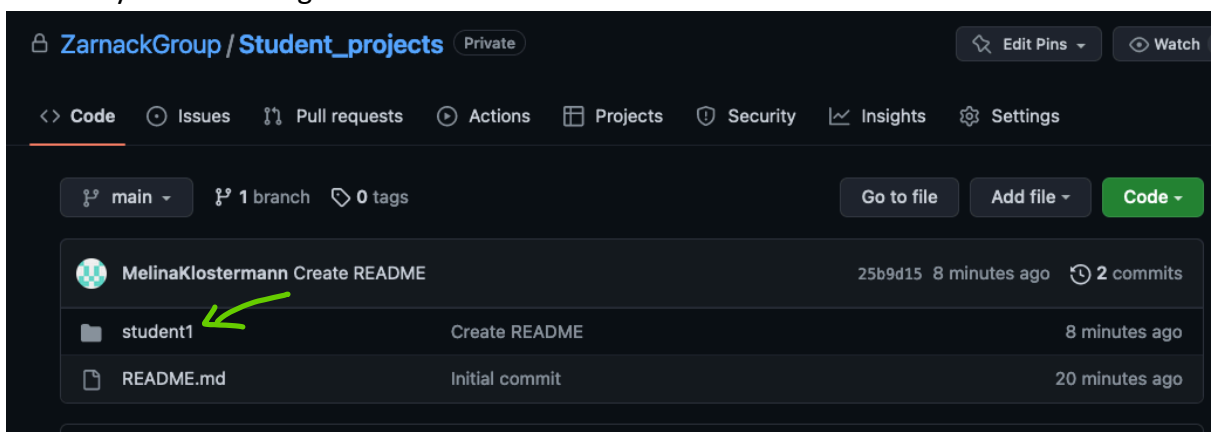


- Commit the folder with the README file.

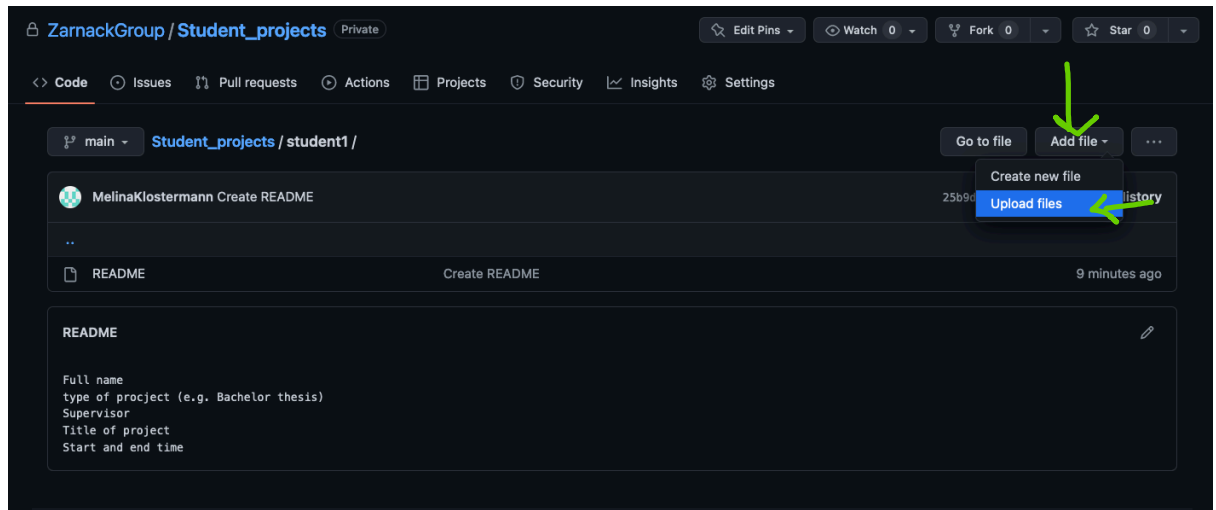


3.2 upload your clean Rmarkdown file

- Click on your folder in github.



- And add your Rmarkdown file inside the folder, by clicking Add file/upload file.



3. Upload your thesis, the knitted pdf/html and all input files on the ebersberger cluster

In addition, you should provide several input and output files on the ebersberger cluster.

3.1 What you should provide:

Make a folder that contains

1. A pdf version of your finished Thesis
2. A pdf or html version of your knitted final Rmarkdown
3. All raw data that you received from collaborators (also if you do not show anything of them in the thesis/your project in the end). Write a file called `raw_files_used.log` and add it to the folder. In this file for each sample specify
 - a. how you call it in your thesis
 - b. how you call it in your Rmarkdown
 - c. what the original name of the file was (e.g. SRA236807585)
 - d. What celline and condition the sample is from
 - e. What experiment type the sample is from (e.g. RNAseq)
4. If you downloaded the data from a public source don't put it in the folder, write a file called `raw_files_used.log` and add it to the folder. In this file for each sample specify
 - a. how you call it in your thesis
 - b. how you call it in your Rmarkdown
 - c. what the original name of the file was (e.g. SRA236807585)

- d. What celline and condition the sample is from
 - e. the link where you downloaded the file
5. If you did preprocessing in bash or galaxy, include the description of the steps in the Rmarkdown file as described above and provide the output files of the preprocessing, so the files that were read into R. For example:
 - a. the htseq count table
 - b. majiq dpsl table + splicegraph + modulaise output
 - c. pureclip bed files
6. Provide output files from your R analysis as .rds files. Examples are:
 - a. The (unfiltered) results table of a DeSeq analysis.
 - b. The defined binding sites granges object.
 - c. A table with binarized LSVs.

→ Ask your supervisor which output files are needed

3.2 Where to put it?

- Give the folder the same name as your Rmarkdown file and the folder on github
- zip the complete folder.
- If you do not have access to the ebersberger cluster and send the zip file to your supervisor.
- Else log into the ebersberger cluster and upload the folder in the **xxx** directory.