R code: DIY

어렵지 않아요

Understanding the process

Workflow

1. You need to **locate** where the **.txt** files are stored and where your **R scripts** are saved.

[Sync]hronize OneDrive folder

Extract info about which .txt files to quality check using Reach_Coding_120524.xlsx

Used R script: fetch_ids.R

2. fetch_ids.R should run successfully and generate **reference.tsv**

3. perform_qc.R should run successfully

Use the information and locate a .txt file

Used R script: perform_qc.R

Perform Quality Check

Used R script: qc_functions.R (loaded inside perform_qc.R)

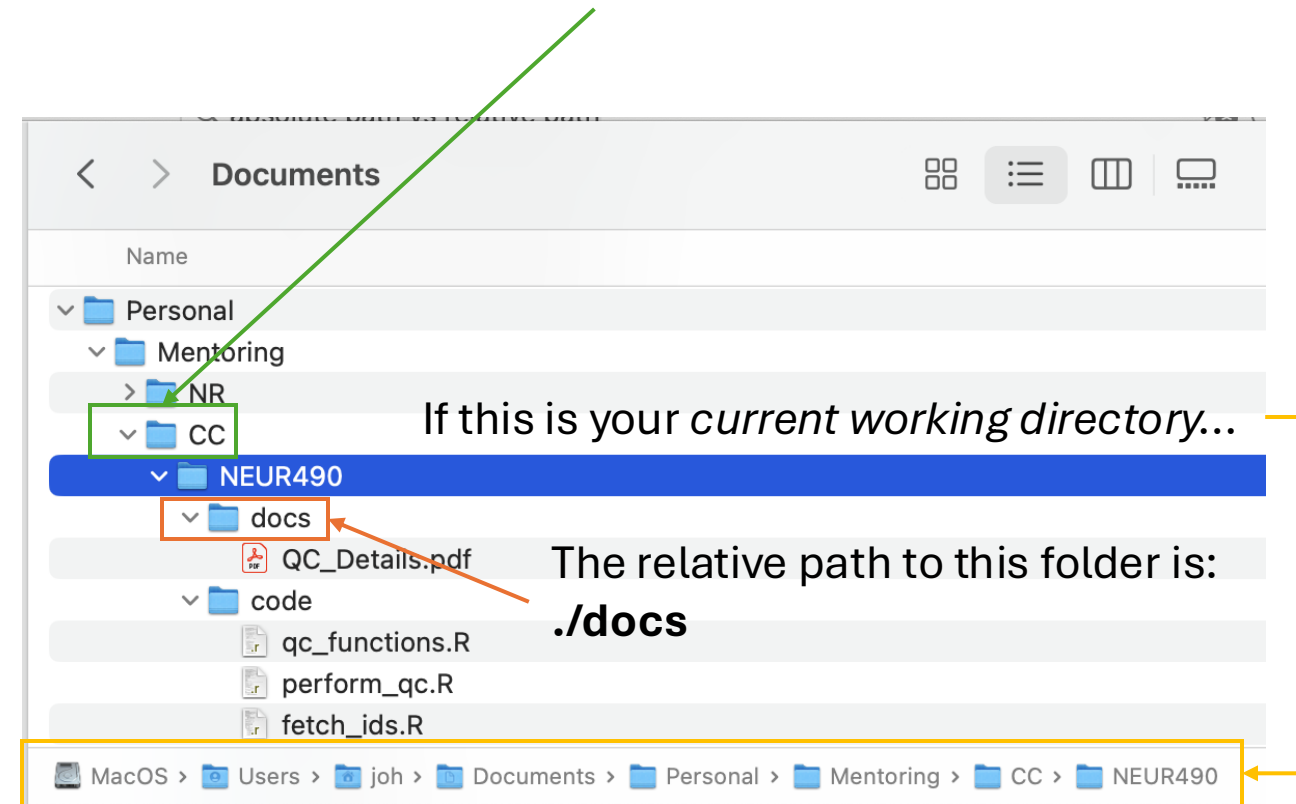Proceed when the loop is over

Save a log

Understanding `path`

- There are two types of path

  - An `absolute path` specifies the full location of a file or directory starting from the root of the file system (**C:** in Windows, **/** in macOS)

  - A `relative path` specifies a location relative to the current working directory
    - A **single dot** (.) means the current working directory
    - A **double dot** (..) means one directory up (parent directory)

The absolute path to this folder is:
/Users/joh/documents/Personal/Mentoring/CC
The relative path is: **../CC**

If this is your *current working directory...*

The relative path to this folder is:
./docs

The absolute path to the current working directory

Change `path`: fetch_ids.R

- You need to change the path to the **excel_file**.

```
27  # path shall be updated – to reflect the actual path
28  HOME = fs::path_home()
29  Mac_OneDrive = 'Library/CloudStorage/OneDrive-SharedLibraries-ChildrensHospitalLosAngeles/Smith, Beth - Reach & Grasp'
30  excel_file = 'Reach_Coding_120524.xlsx'
```

Lines 28-30 of **fetch_ids.R** assumes the following folder structure:

```
/Users
|--- joh*                                          This is the value of HOME
    |--- Library
        |--- CloudStorage
            |--- OneDrive-SharedLibraries-ChildrensHospitalLosAngeles     This is the value of
                |--- Smith, Beth - Reach & Grasp                          Mac_OneDrive
                    |--- Reach_Coding_120524.xlsx
```

* Your *username* will be different

Change `path`: fetch_ids.R

- Let's suppose that you created a folder named **NEUR490** inside your **Downloads** folder and saved **Reach_Coding_120524.xlsx** inside the folder. If so, **what do you need to change**?

Lines 28-30 of **fetch_ids.R** will newly assume the following folder structure:

```
/Users
|--- joh*          This is the value of HOME
    |--- Downloads
        |--- NEUR490     This is the value of ...
            |--- Reach_Coding_120524.xlsx
```

* Your *username* will be different

```
27  # path shall be updated - to reflect the actual path
28  HOME = fs::path_home()
29  Mac_OneDrive = 'Library/CloudStorage/OneDrive-SharedLibraries-ChildrensHospitalLosAngeles/Smith, Beth - Reach & Grasp'
30  excel_file = 'Reach_Coding_120524.xlsx'
```
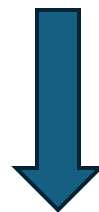
Change `path`: fetch_ids.R

```
32  onedrive_path = file.path(HOME, Mac_OneDrive, excel_file)
33
34  # sheet='CC'
35  if (!file.exists(onedrive_path)){
36      stop("Error: Excel file not found at: ", onedrive_path)
37  }
38  record = read_excel(onedrive_path, sheet='CC', skip=1)
```

fetch_ids.R

If you miss, you will see this error message

If you change things correctly, this line will run

```
97  # save the reference to a tab separated file
98  write.table(tab, file='../processed/reference.tsv', sep= "\t",
99              row.names=F, col.names=T,
100             quote=F)
```

Relative path symbol here: a double dot (..)
You can check your *current working directory*
using the command: **getwd()**. Here's an example:

```
[> getwd()
[1] "/Users/joh/Documents/Personal/Mentoring/CC/NEUR490/code"
```

From this example, the parent directory will be:
/Users/joh/Documents/Personal/Mentoring/CC/NEUR490
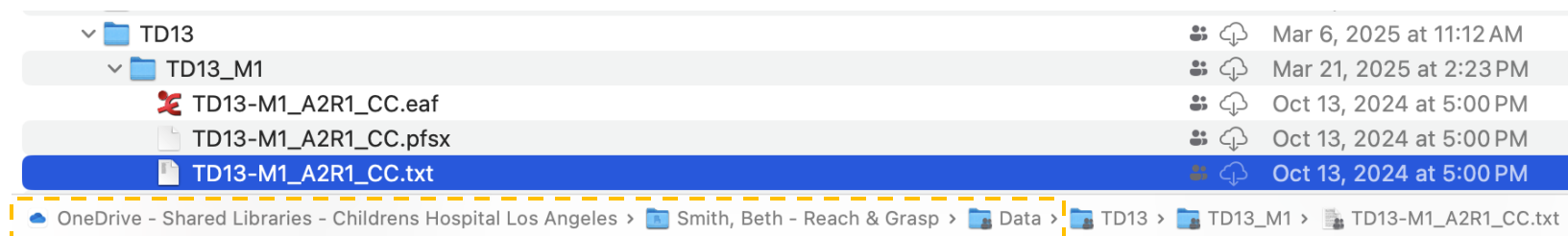Then **reference.tsv** will be saved in:
/Users/joh/Documents/Personal/Mentoring/CC/NEUR490/processed

Change another `path`: perform_qc.R

- Make sure you know the path to **reference.tsv** file, and put in the path to **here**
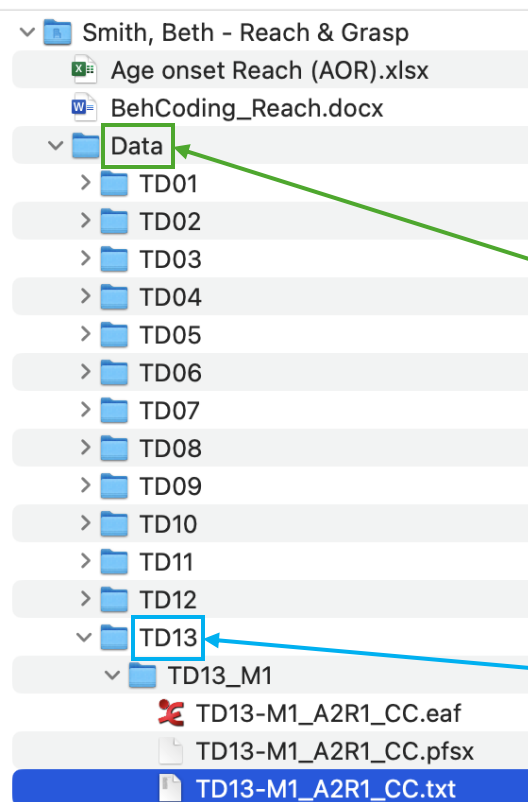
```
18 # This will return error if you did not complete 2) above.
19 references = read.csv('../processed/reference.tsv', sep='\t')
```

- Then you need to know the *path* to your data files (.txt)

Change another `path`: perform_qc.R

- So when I prepared **perform_qc.R**,
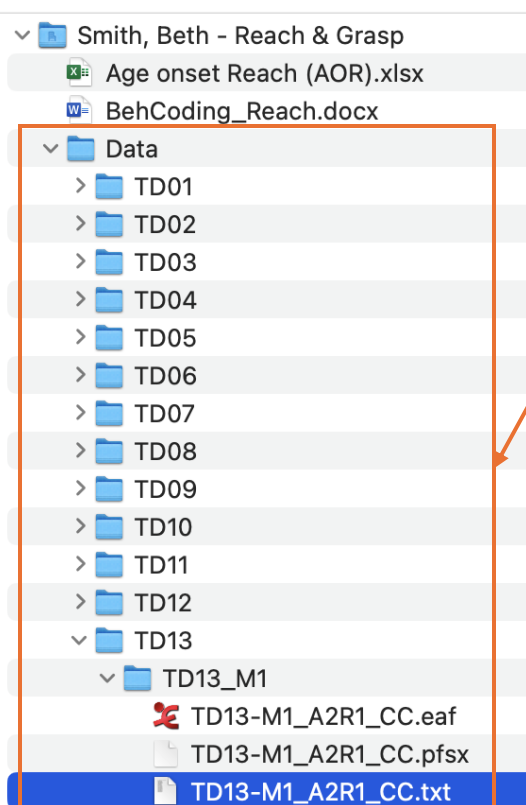 I assumed the following hierarchy:

So I prepared **references.tsv** in the following way. The column **path** saves portions of paths to .txt files, starting from the **Data** folder

	A	B	C	D	E
	subj	month	act	prefix	path
	TD13	M3	A2	TD13-M3_A2	Data/TD13/TD13M3/TD13M3A2
	TD13	M3	A3	TD13-M3_A3	Data/TD13/TD13M3/TD13M3A3
	TD13	M3	A4	TD13-M3_A4	Data/TD13/TD13M3/TD13M3A4
	TD13	M3	A5	TD13-M3_A5	Data/TD13/TD13M3/TD13M3A5
	TD13	M3	A6	TD13-M3_A6	Data/TD13/TD13M3/TD13M3A6
	TD13	M3	A7	TD13-M3_A7	Data/TD13/TD13M3/TD13M3A7
	TD13	M4	A2	TD13-M4_A2	Data/TD13/TD13M4/TD13M4A2
	TD13	M4	A3	TD13-M4_A3	Data/TD13/TD13M4/TD13M4A3
	TD13	M4	A4	TD13-M4_A4	Data/TD13/TD13M4/TD13M4A4

Change another `path`: perform_qc.R

- To change the original code minimally, I ask you to at least maintain the structure up to this level:


```
18  # This will return error if you did not complete 2) above.
19  references = read.csv('../processed/reference.tsv', sep='\t')
20  # Some problematic folders rejected for now (March 5, 2025)
21  subdirs_temp = references$path
22  idx_spare = !grepl("Data/TD27/TD27M3", subdirs_temp)
23  subdirs = subdirs_temp[idx_spare]
24  prefixes_temp = references$prefix
```
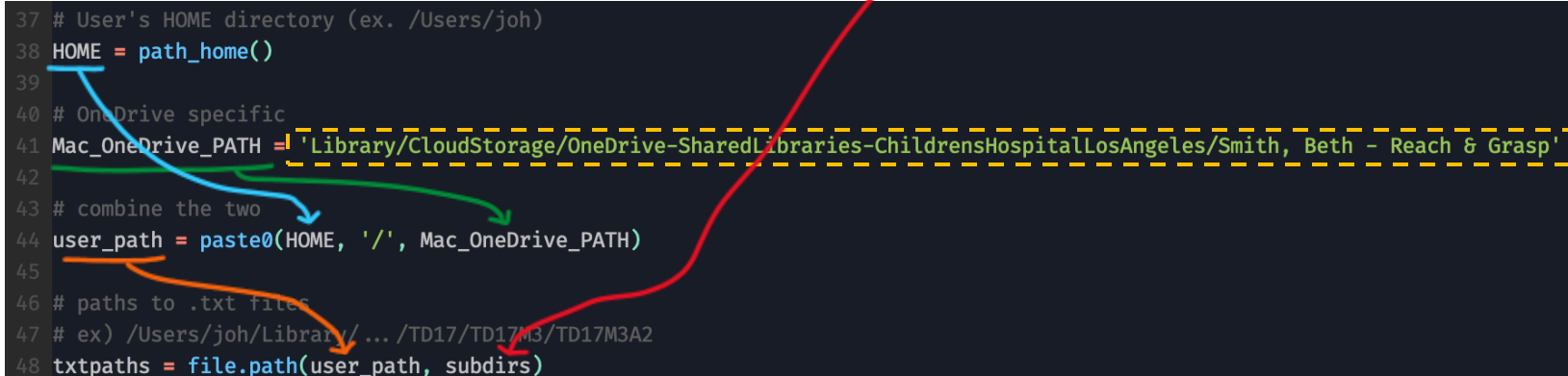
	A	B	C	D	E
	subj	month	act	prefix	path
	TD13	M3	A2	TD13-M3_A2	Data/TD13/TD13M3/TD13M3A2
	TD13	M3	A3	TD13-M3_A3	Data/TD13/TD13M3/TD13M3A3
	TD13	M3	A4	TD13-M3_A4	Data/TD13/TD13M3/TD13M3A4
	TD13	M3	A5	TD13-M3_A5	Data/TD13/TD13M3/TD13M3A5
	TD13	M3	A6	TD13-M3_A6	Data/TD13/TD13M3/TD13M3A6
	TD13	M3	A7	TD13-M3_A7	Data/TD13/TD13M3/TD13M3A7
	TD13	M4	A2	TD13-M4_A2	Data/TD13/TD13M4/TD13M4A2
	TD13	M4	A3	TD13-M4_A3	Data/TD13/TD13M4/TD13M4A3
	TD13	M4	A4	TD13-M4_A4	Data/TD13/TD13M4/TD13M4A4

Change another `path`: perform_qc.R

- Then if you change the **path** correctly, things will run smoothly.
  (ex. if you downloaded **Data** folder and its subfolders inside **Downloads/NEUR490** folder, what should be your **path**?

```
18  # This will return error if you did not complete 2) above.
19  references = read.csv('../processed/reference.tsv', sep='\t')
20  # Some problematic folders rejected for now (March 5, 2025)
21  subdirs_temp = references$path
22  idx_spare = !grepl("Data/TD27/TD27M3", subdirs_temp)
23  subdirs = subdirs_temp[idx_spare]
24  prefixes_temp = references$prefix
25  prefixes = prefixes_temp[idx_spare]
```

```
37  # User's HOME directory (ex. /Users/joh)
38  HOME = path_home()
39
40  # OneDrive specific
41  Mac_OneDrive_PATH = 'Library/CloudStorage/OneDrive-SharedLibraries-ChildrensHospitalLosAngeles/Smith, Beth - Reach & Grasp'
42
43  # combine the two
44  user_path = paste0(HOME, '/', Mac_OneDrive_PATH)
45
46  # paths to .txt files
47  # ex) /Users/joh/Library/ ... /TD17/TD17M3/TD17M3A2
48  txtpaths = file.path(user_path, subdirs)
```

Additional note

- If you see errors related to TD27M3 or TD31M3A2, comment out lines 73~81