Git and GitHub

General Applications

Garrett Speed
Data Steward – Geo Data Team

Git

What is Git?

- A versioning program intended for tracking changes in code
- Allows many people to make and track changes independently, and then merge those changes together.
- Created by the creator of Linux, Linus Torvalds, in 2005 for managing development of the Linux kernel

Tracking Changes in Code

- A versioning program intended for tracking changes in code

Tracking Changes in Code

- A versioning program intended for tracking changes in code
- Actually tracks changes in text-based files
  - Text-based like .txt
  - Not text like Word .docx files... unless you save your .docx in a certain way

Tracking Changes in Code

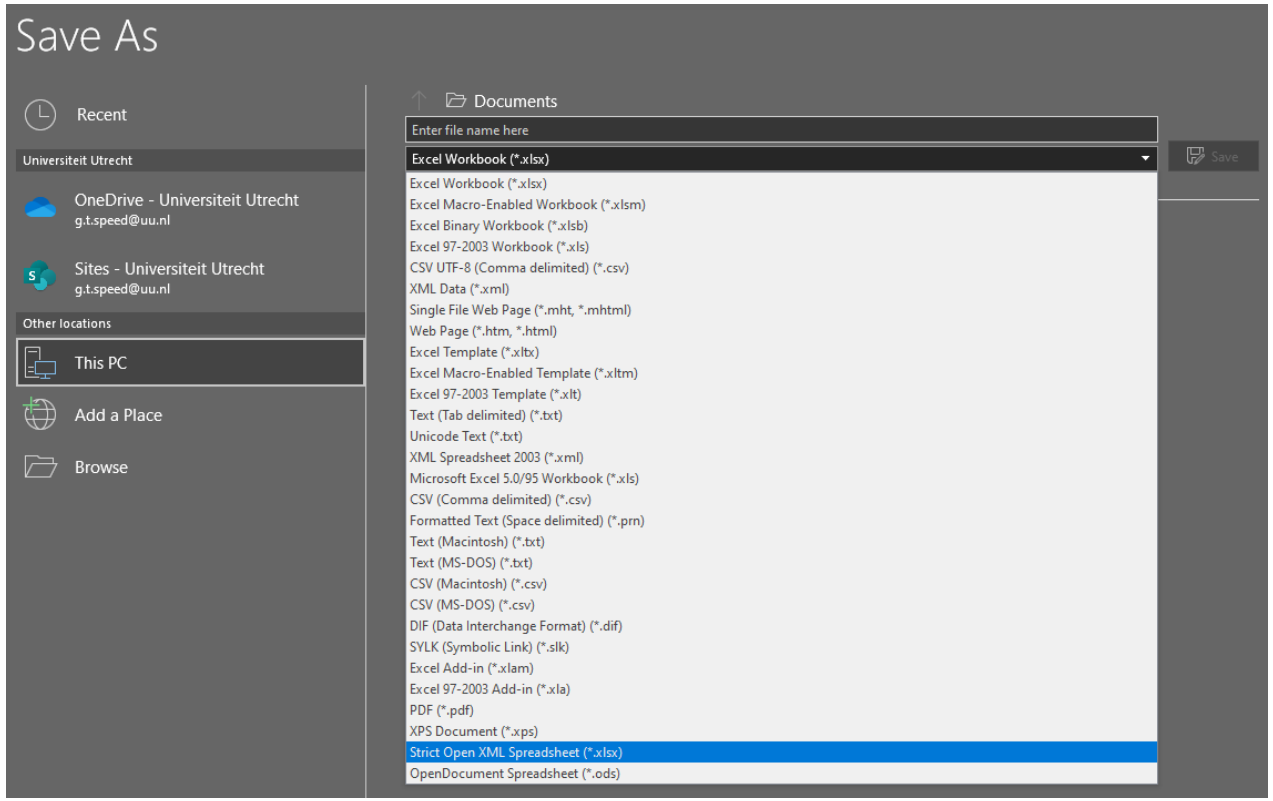
- You can use Git to track changes in
  - Program code
    - Javascript (JS), Python, R, C, C++, C#, Java, Julia, MatLab, Rust, CSS, HTML, etc...
    - Jupyter Lab Notebooks
  - Documents
    - XML-based Word Documents (.docx)
    - Open Document Foundation documents (.odt)
    - Markdown
    - LaTeX

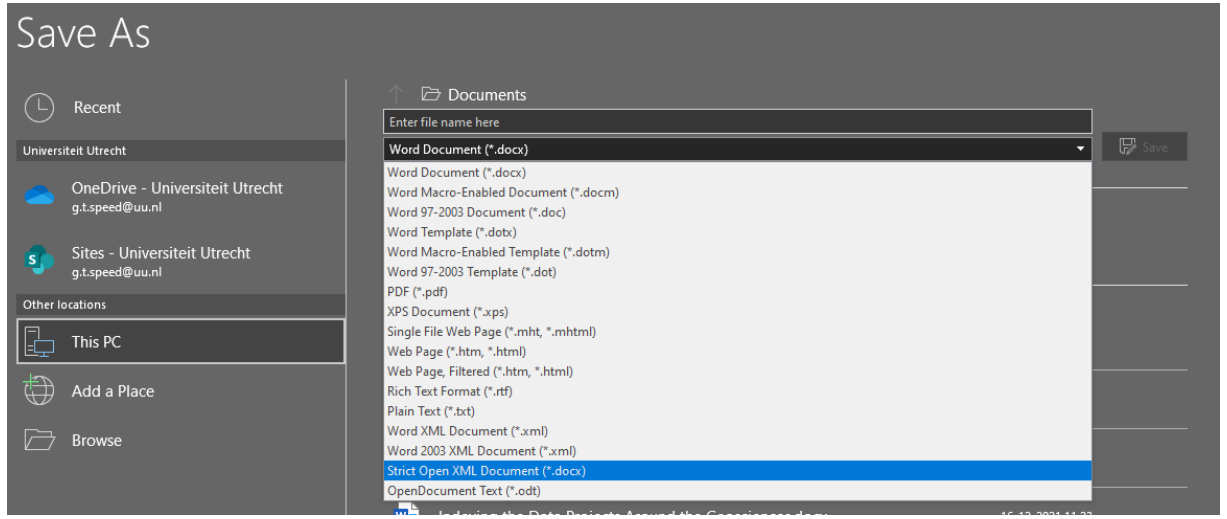
Tracking Changes in Code

- You can use Git to track changes in
  - Spreadsheets/Tables
    - Comma-separated Values (CSVs, TSVs)
    - XML-based Excel files (.xlsx)
    - Open Document Foundation Spreadsheets (.ods)
    - ⚠This can create large Git repositories⚠ changing content on a single line adds an entry for the entire line

XML-Based docx and xlsx

- When saving your files in Word and Excel, select the XML option in the Save As dialogue.



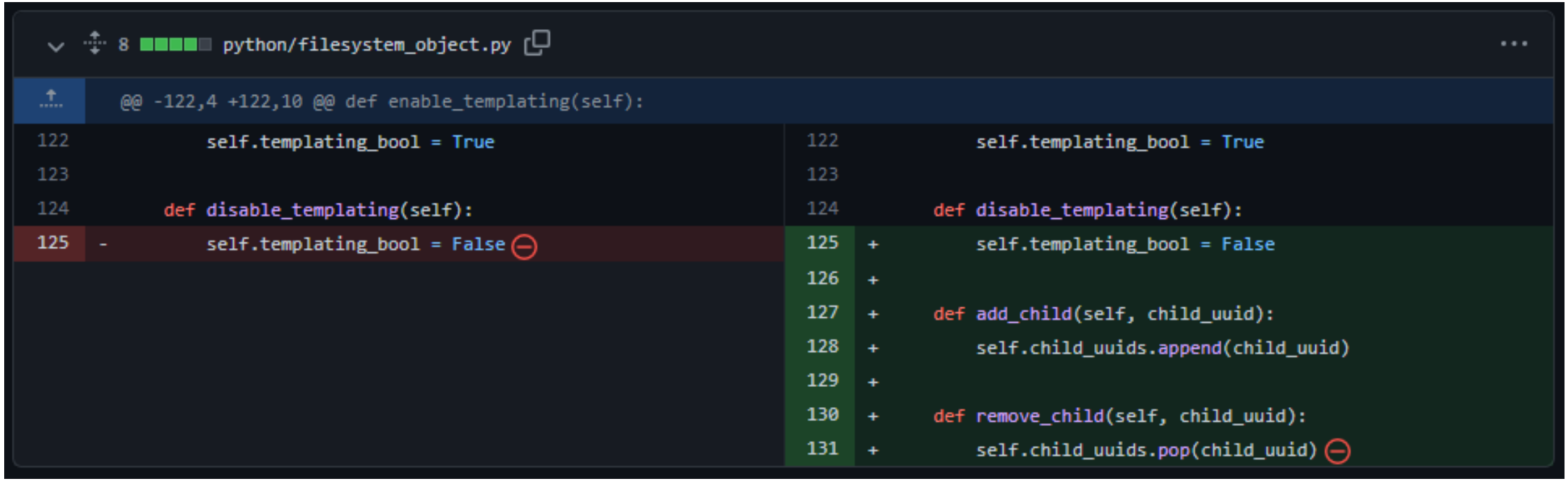


How does it track changes?

- When you initially start tracking a file with Git, it saves the state of the file at that time.
- When you update changes with Git, like changing text on a line, it will record what line number was changed, and whether it was altered, added, or deleted.
- Also able to ignore a situation where unchanged lines are now on different line numbers

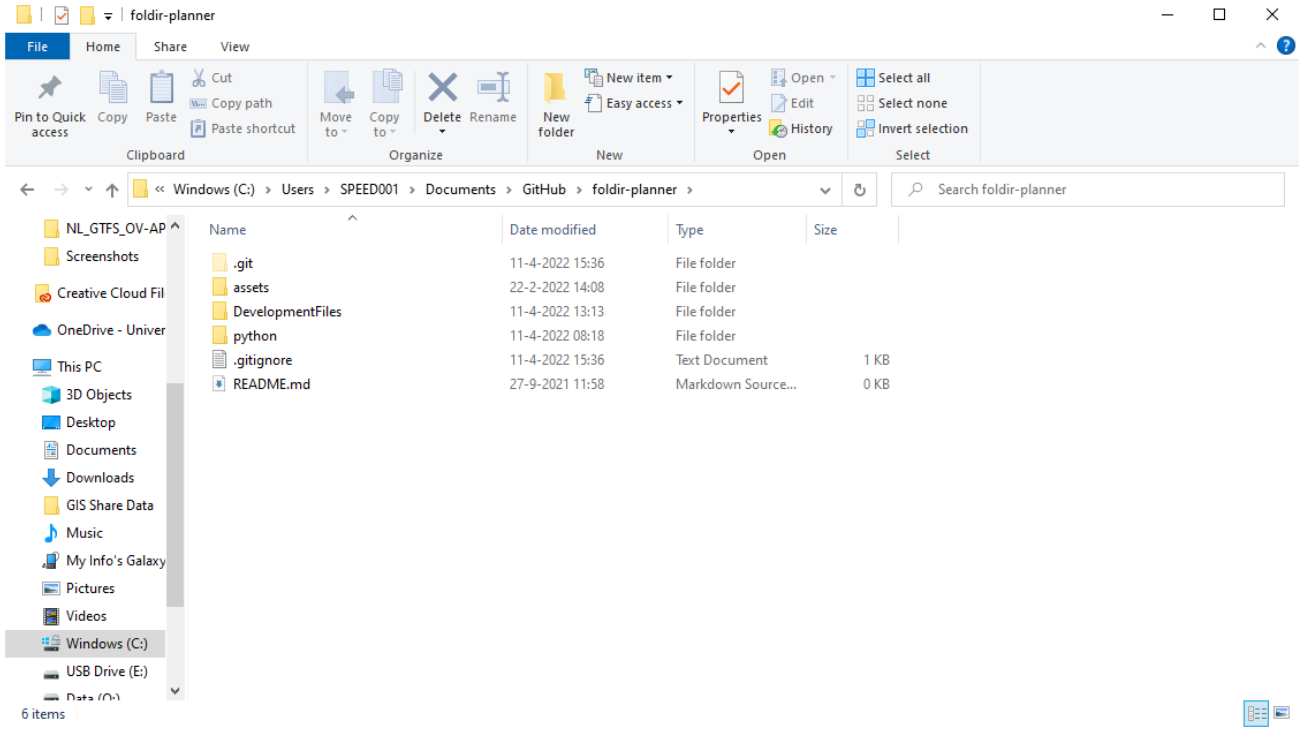
How does it track changes?

- Example of adding code lines after a line
  - Technically in this case, line 125 was changed because an “invisible” new line character was inserted at the end.
  - Therefore in Git, old line 125 was deleted, and a new line 125 was added



Repositories

- A repository is a project folder that will have the changes in it tracked.
- On your computer, it is a folder with all of the project files in it
- On a Git server is a clone of your local folder



Git Commands

- git init
- git add
- git commit
- git clone
- git checkout
- git pull
- git merge
- git push

https://edu.nl/mfnxu

https://www.gitkraken.com/learn/git/commands

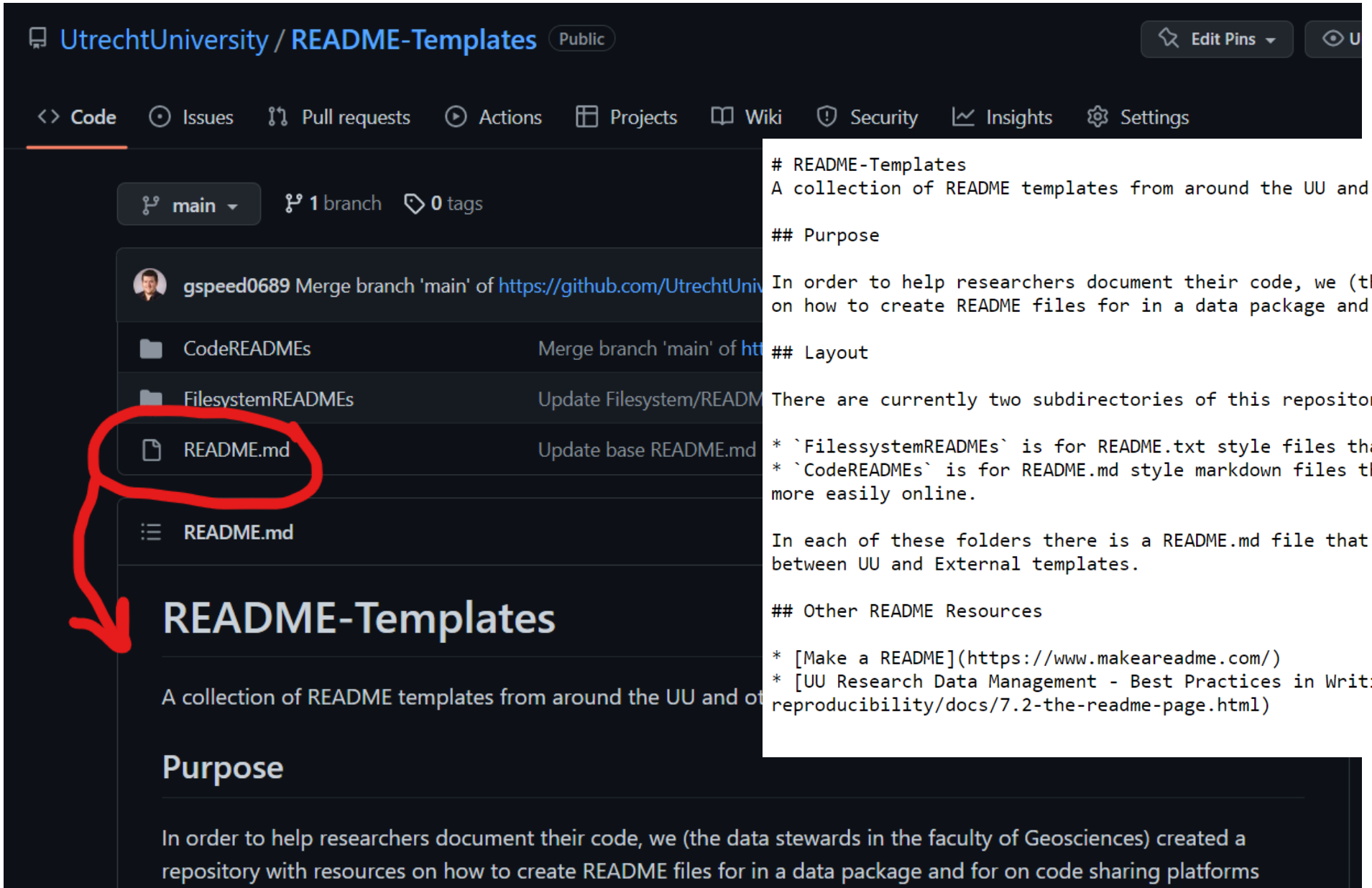
Creating a Repository

- Navigate to your parent folder
- Use the git commands
  - `git init`
  - `git add`
  - `git commit –m {short text about what changes you made}`

README.md

- README markdown documentation file
- GitHub, GitLab, Gitea all recognize README.md filename as the documentation for a folder in a Git repository, and will display the stylized Markdown
- Place a README.md file on most folder levels.
- Find README templates on out GitHub: UtrechtUniversity/README-Templates

README.md




```
# README-Templates
A collection of README templates from around the UU and other institutions.

## Purpose

In order to help researchers document their code, we (the data stewards in the faculty of Geo
on how to create README files for in a data package and for on code sharing platforms such as

## Layout

There are currently two subdirectories of this repository, FilesystemREADMEs and CodeREADMEs.

* `FilesystemREADMEs` is for README.txt style files that should accompany data packages on t
* `CodeREADMEs` is for README.md style markdown files that should be included with code in Gi
more easily online.

In each of these folders there is a README.md file that will explain more about each folder,
between UU and External templates.

## Other README Resources

* [Make a README](https://www.makeareadme.com/)
* [UU Research Data Management - Best Practices in Writing Reproducible Code](https://utrecht
reproducibility/docs/7.2-the-readme-page.html)
```

Markdown

- Text based document with basic formatting
- Popular in programming communities
- How to write documentation for GitHub and GitLab
- Reddit comments use Markdown
- https://edu.nl/cer6v
- https://www.markdownguide.org/cheat-sheet

.gitignore

- Sometimes in your folder you have things you don't want tracked:
  - SSH Keys, Passwords, Config Files, AWS Cloud Keys
  - Data, especially personal data
- .gitignore is basic text document of files you don't want to be tracked, or templates of files you don't want tracked.
- github/gitignore: A collection of useful .gitignore templates

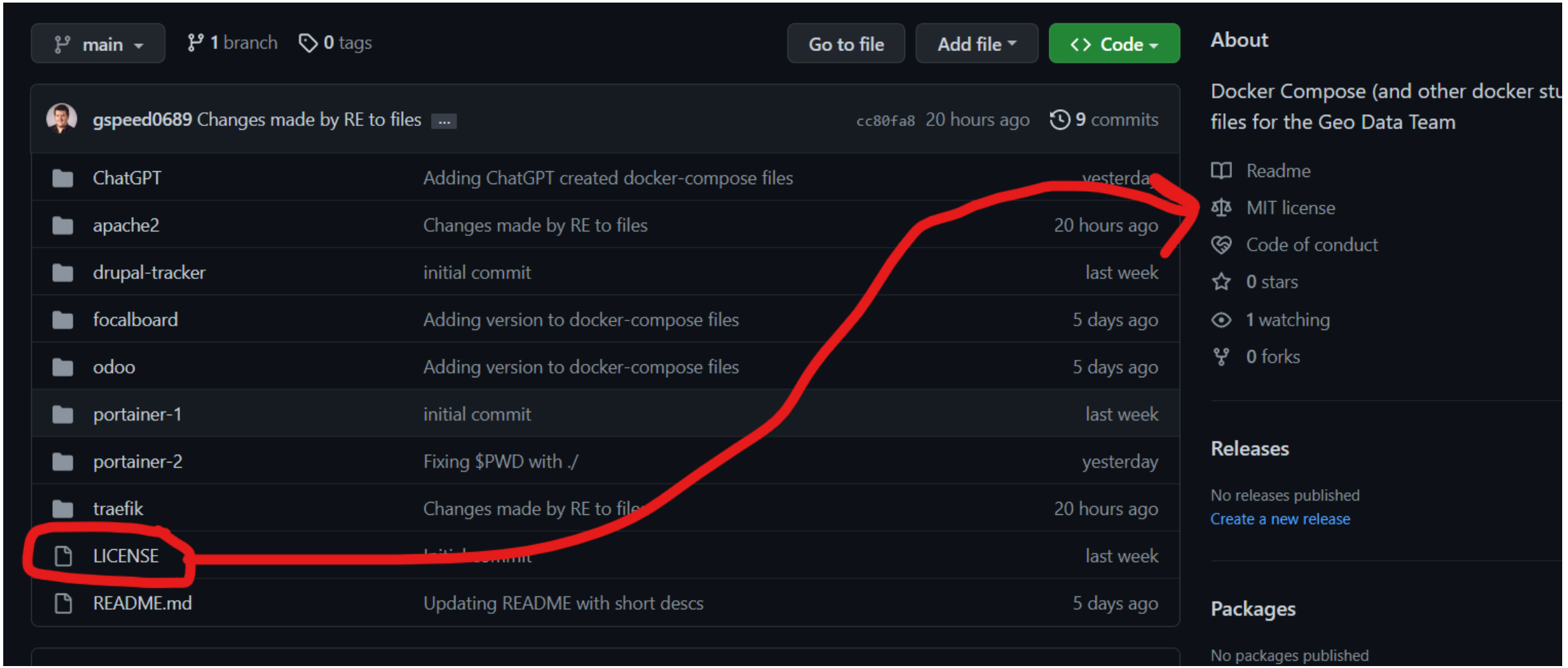
.gitignore

```
local_config.conf
*.ai
*checkpoint.ipynb
*.pyc
*/fconfig.py
PythonServer/fconfig.py
structure_versions/*
current_files/*
test_results/GeoweteschappenSurvey.foldir
PythonServer/IPYNBs/Untitled.ipynb
*.pyc
PythonServer/__pycache__/*
PythonServer/IPYNBs/.ipynb_checkpoints/*
DevelopmentFiles/*
python/generic_function_testing.py
rest-web
```

Code Licenses

- A file named LICENSE should be placed in the repository, GitHub will recognize standard licenses.
- Find out more on our website: https://geo-data-support.sites.uu.nl/open-science-open-data/code-licenses/
- Our recommendations:
  - MIT
  - GNU

Code Licenses



GitHub

What is GitHub

- An online Git server with a friendly user interface for seeing, downloading, and managing Git projects.
- Now owned by Microsoft
- Most popular code repository

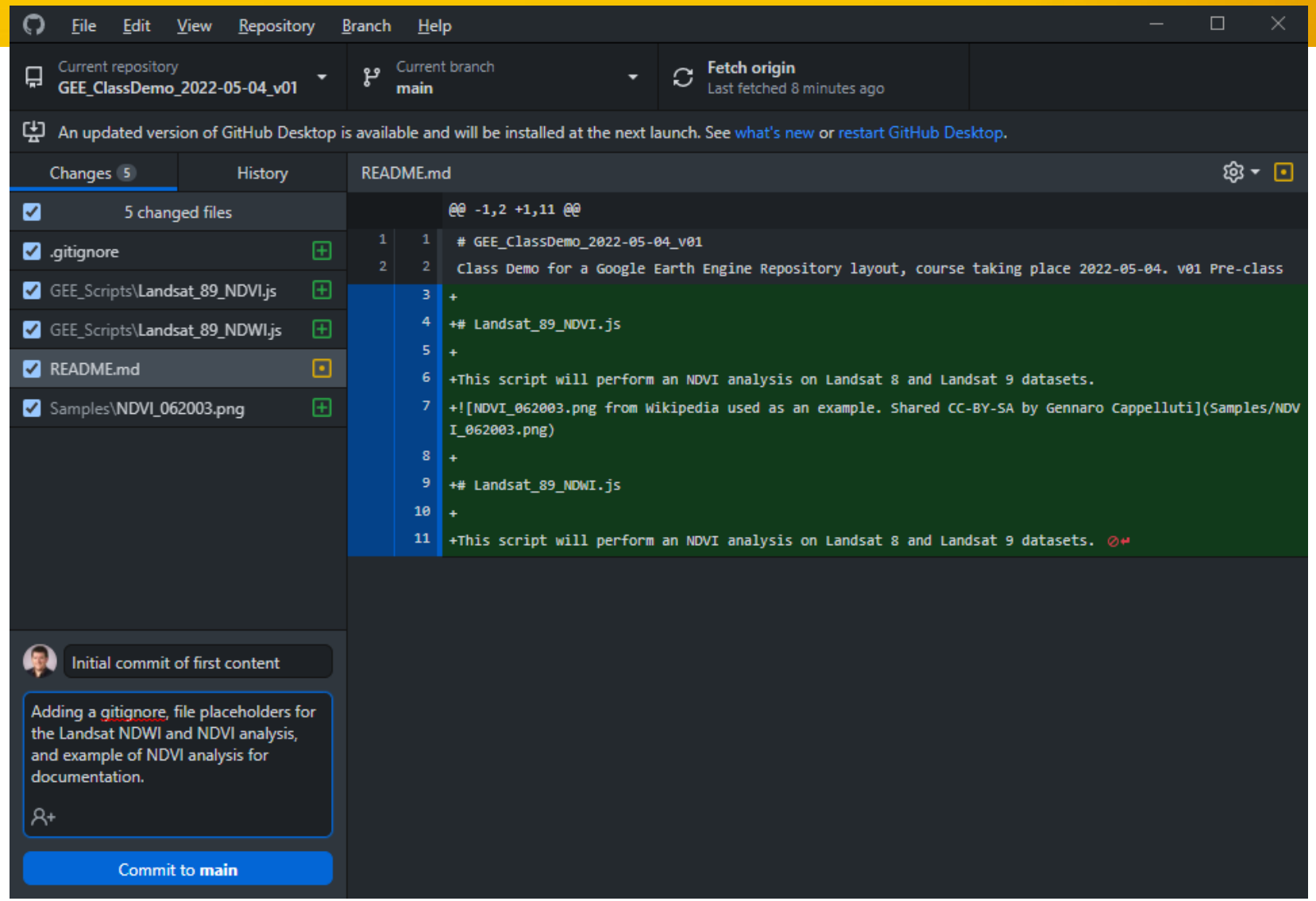
Creating a Repository

- You can create a new empty repository on the GitHub website
- You will need a name for your repository, and it will be addressable as {your_username}/{repository_name}
- Options to add a default license
- Options to add a ReadMe
- Options to add a .gitignore file

GitHub Desktop

- A GUI for managing Git projects.
  - Integrated with GitHub
  - Not required to use a GitHub hosted Git repository
- Today, we're going to use GitHub Desktop, for ease
  - The command line is easier to use on Mac and Linux as those operating systems are Unix based
- No need for SSH keys if working with GitHub

GitHub Desktop



UU GitHub Organization

- Utrecht University has its own GitHub organization, allowing for collaboration officially as the university
- Unlocks more tools and storage for projects
- Requires Multi-Factor Authentication activated on your GitHub account
- Sign-in at https://github.com/UtrechtUniversity

GitHub vs GitLab

- GitLab is an open-source Git management site that hosts repositories, or allows you to host their system yourself
- GitLab has more code deployment tools built in
- GitLab is not owned by Microsoft

Let’s git practical

g.t.speed@uu.nl
datateam.geo@uu.nl

Tasks

- Log into GitHub
- Create a new repository
- Create a Markdown Readme
- Create folders
  - Data
  - Code
- Create a .gitignore file
- Commit changes to code and push to GitHub