

# 1<sup>st</sup> CCU Workshop on Git and VS Code Integration



# Today's Schedule

Git Fundamentals

VS Code Integration with Git

Break – Questions time

Hands on Lab

Conflicts

# You might remember me from ...



# You might remember me from ...

Data analysis meetings began in 2022 with Jacob Davidson and myself as organizers



Jacob Davidson



Daniel S. Calovi

IMPRS – Introduction to Scientific Coding workshop



# The Data Science Consultancy

Join our  
CASCB Slack  
channel



Schedule a  
meeting with me



# Why are we offering this workshop?

Documentation/Versioning - Git is the industry/academia standard for documentation and versioning nowadays

- *“Almost 95% of developers reported using Git as their primary version control system”*

Code sharing – When working on large projects, sharing your code, tool or analysis becomes extremely likely, Git will help with synchronization while working with others

- Portfolio/Publishing – Even when not collaborating, a GitHub account will serve as your portfolio or location to publish algorithms that you publish.

Data Management Plan – One can think of Git as the first step to organizing your project and building a data management plan to have everything securely backed up

# Why are we offering this workshop?

Documentation/Versioning - Git is the industry/academia standard for documentation and versioning nowadays

- *“Almost 95% of developers reported using Git as their primary version control system”*

Code sharing – When working on large projects, sharing your code, tool or analysis becomes extremely likely, Git will help with synchronization while working with others

- Portfolio/Publishing – Even when not collaborating, a GitHub account will serve as your portfolio or location to publish algorithms that you publish.

Data Management Plan – One can think of Git as the first step to organizing your project and building a data management plan to have everything securely backed up

Contact Ilja Werner for creating your DMP



Don't forget to create a  
Data Management Plan



# What is Git?

```
GIT(1)                               Git Manual                               GIT(1)
NAME
  git - the stupid content tracker
```

*“Git is a distributed version control system that tracks versions of files. It is often used to control source code by programmers who are developing software collaboratively.”*

Git was created by Linus Torvalds to be used in the development of Linux (which he is also the creator)

Unlike other Version control systems, Git uses snapshots instead of just difference files

Git is local, you don't actually need to host any of your Git projects online

Git generally only adds Data (and it's difficult to really delete things once published)

- **Never upload a password or sensitive information to to a repository**



# Git Terminology

**Add** / stage – Adding a file to be tracked

- turning track changes on

**Commit** – Creating the snapshot of your project's current state

- Track changes via snapshots, not in real time

**Branch** – A branch allows you to develop new features or test changes independently from the last stable version

**Fork** – Related to a **branch**, a fork is a personal copy of another user's repository that resides in your GitHub account.

**Merge** – Merging combines the changes from one **branch** into another, synchronizing them

**Clone** – Cloning creates a local copy of a remote repository on your machine, including its entire history

- Differs from **Fork** since it does it locally and not in your GitHub account

# Git Terminology

**Workspace** – current working directory

**Local Repository** – The commits that are in your local machine, but not yet pushed to a Remote Repository

**Remote/Remote Repository** – A repository hosted online (e.g. GitHub)

**Master/Main** – Default branch

**Origin** – Name given to a Remote repository that was cloned

# Git basic commands

**Git init** – Initializes a new Git repository in the current folder, creating a .git subfolder

**Git add** – Adds (Stages) files to be tracked (new, modified, or deleted files)

**Git commit** – Creates a snapshot of the tracked (staged) files along with a log/description given by the user

**Git status** – Shows the current status of the working folder, listing staged changes, files not being tracked, etc

**Git log** – Displays a log of commits, showing the commit history along with details like author, date, and commit message.

**Git clone** – Creates a local copy of an existing remote repository on your machine.

**Git branch** – Manages branches; use it to list, create, or delete branches in your repository.

**Git checkout** – Switches between branches or restores files to a previous state.

**Git merge** – Integrates changes from one branch into another.

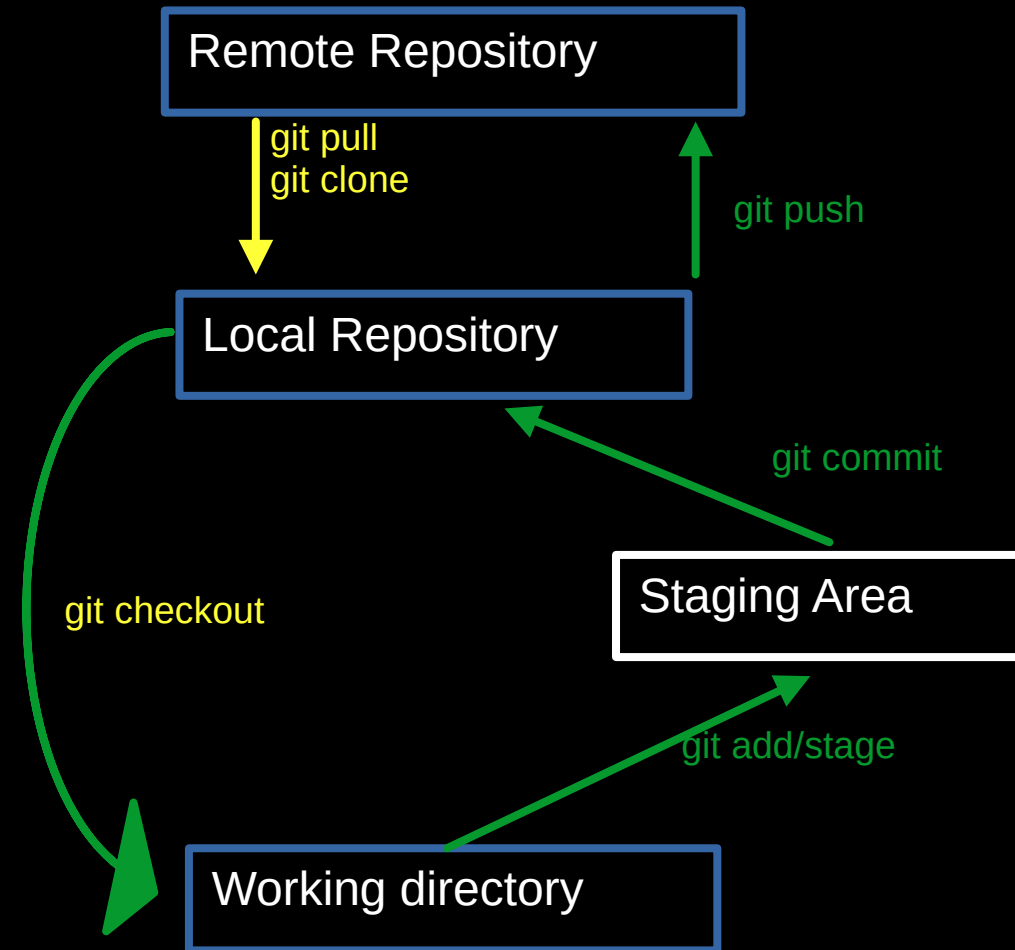
**Git push** – Uploads your local commits to a remote repository.

# Git Logic

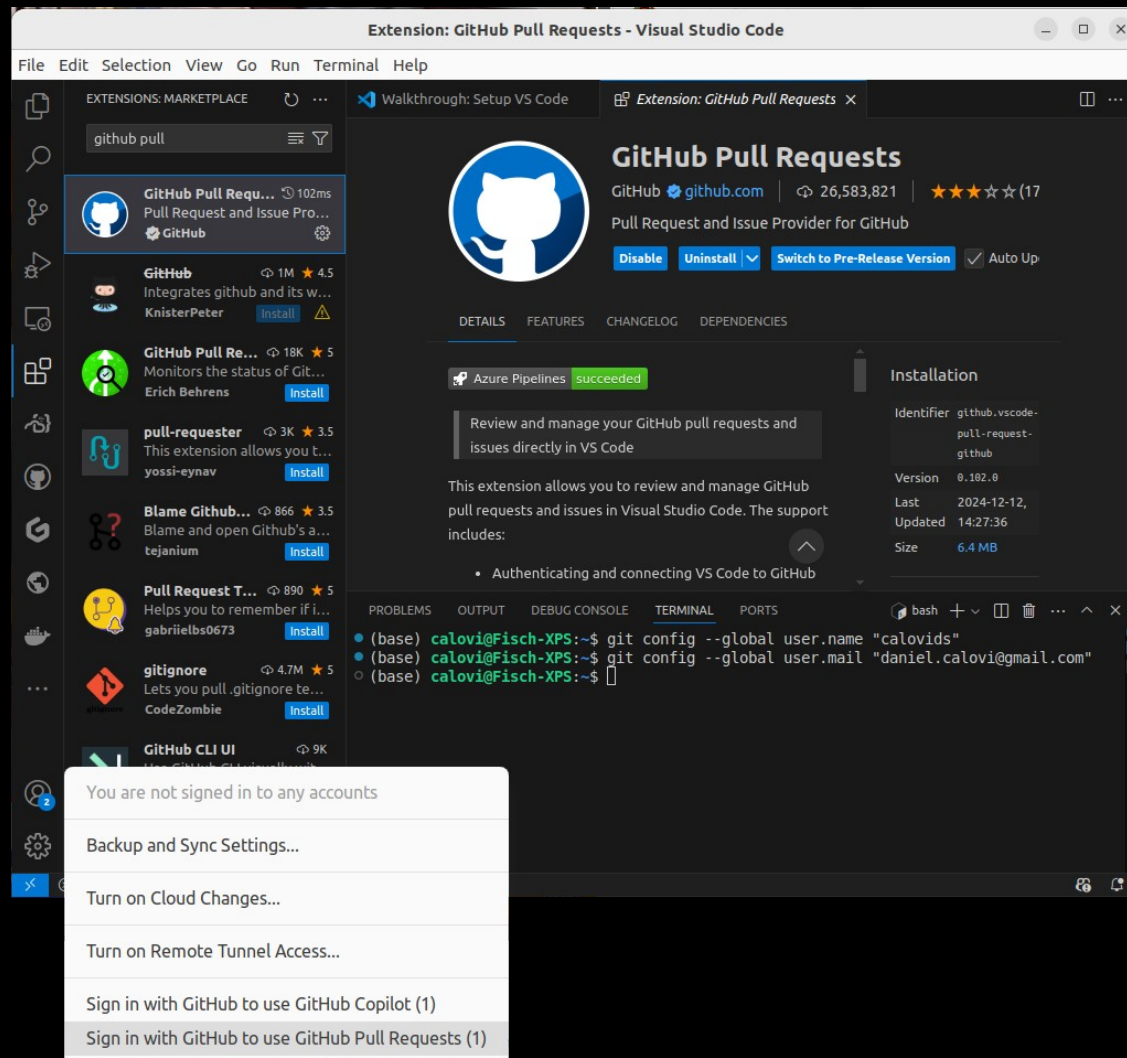
One can start their logic at any blue box

Green commands and arrows are the standard flow of actions

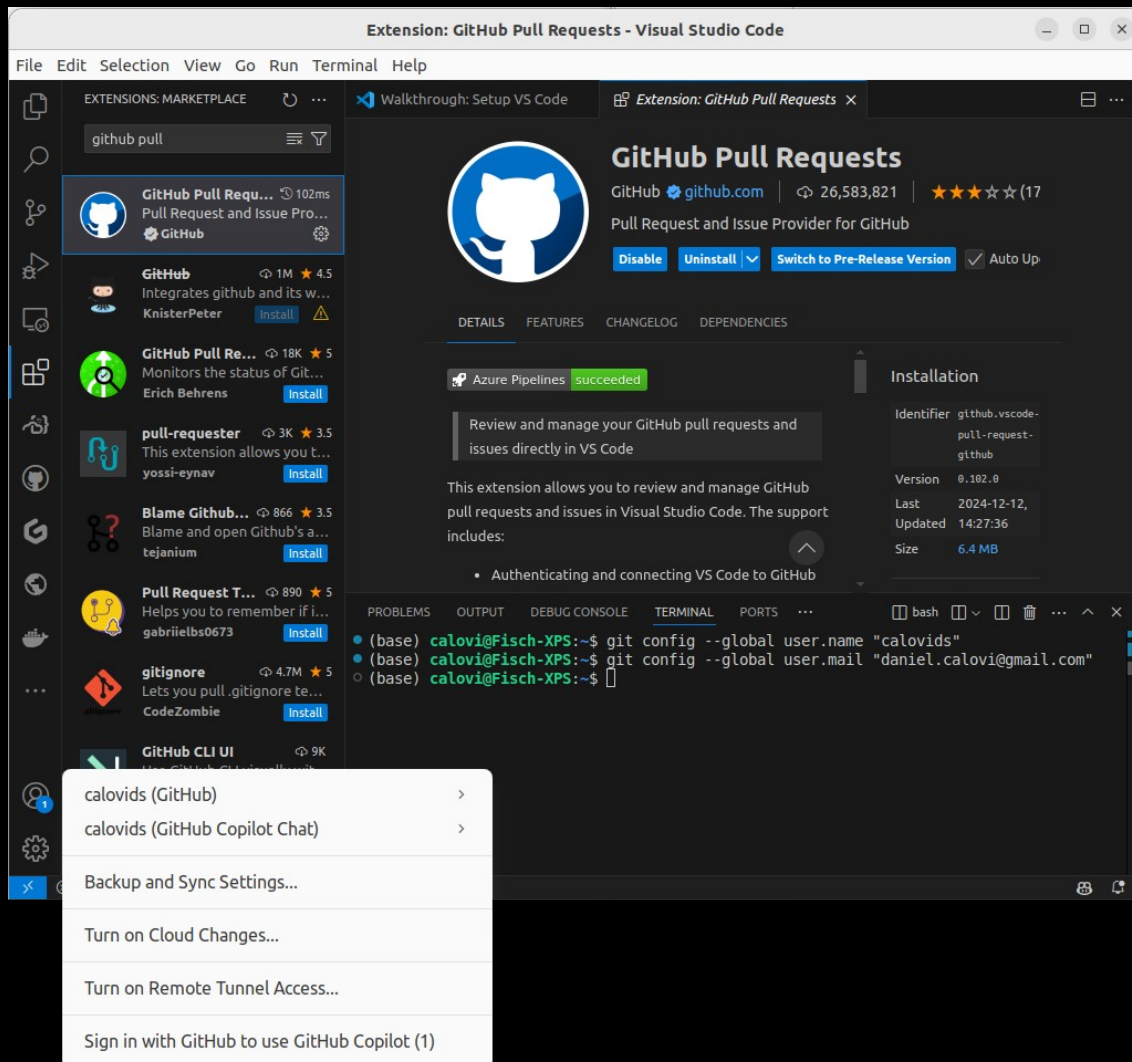
Yellow arrows and commands are possible (but not essential) paths



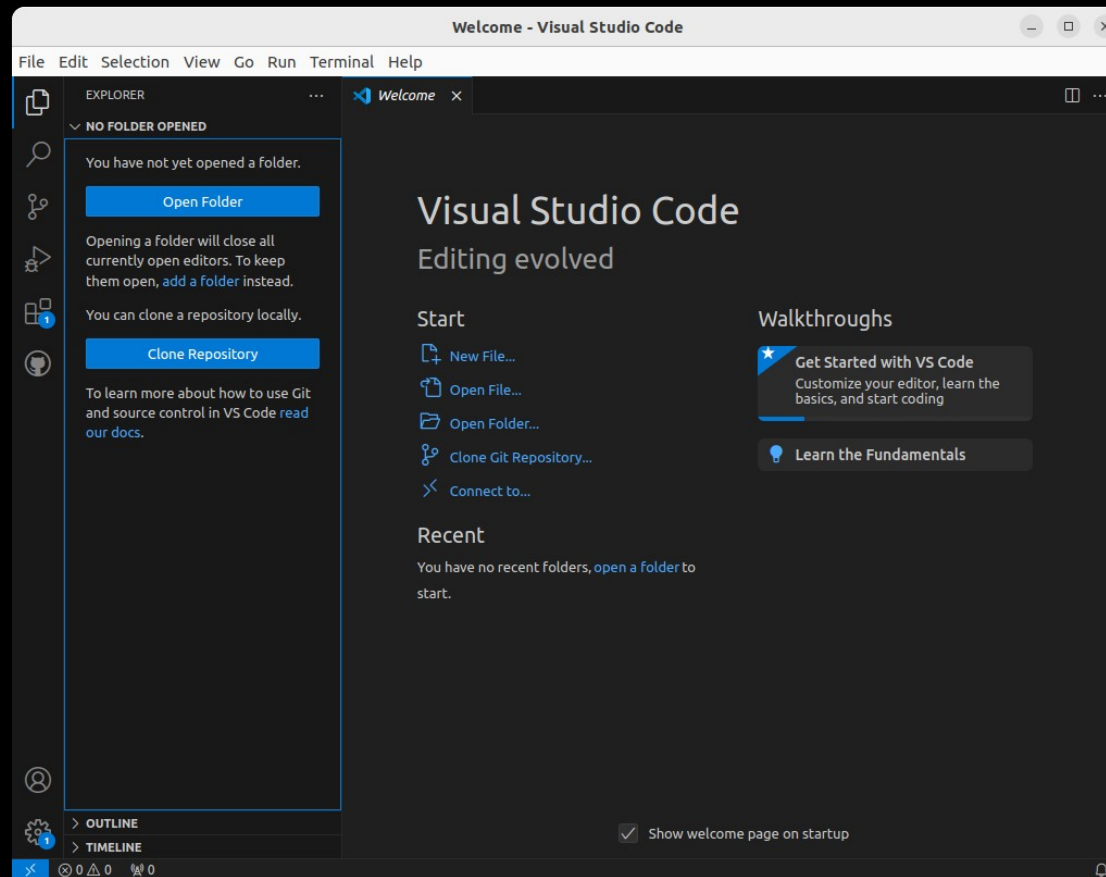
# VS Code Extensions



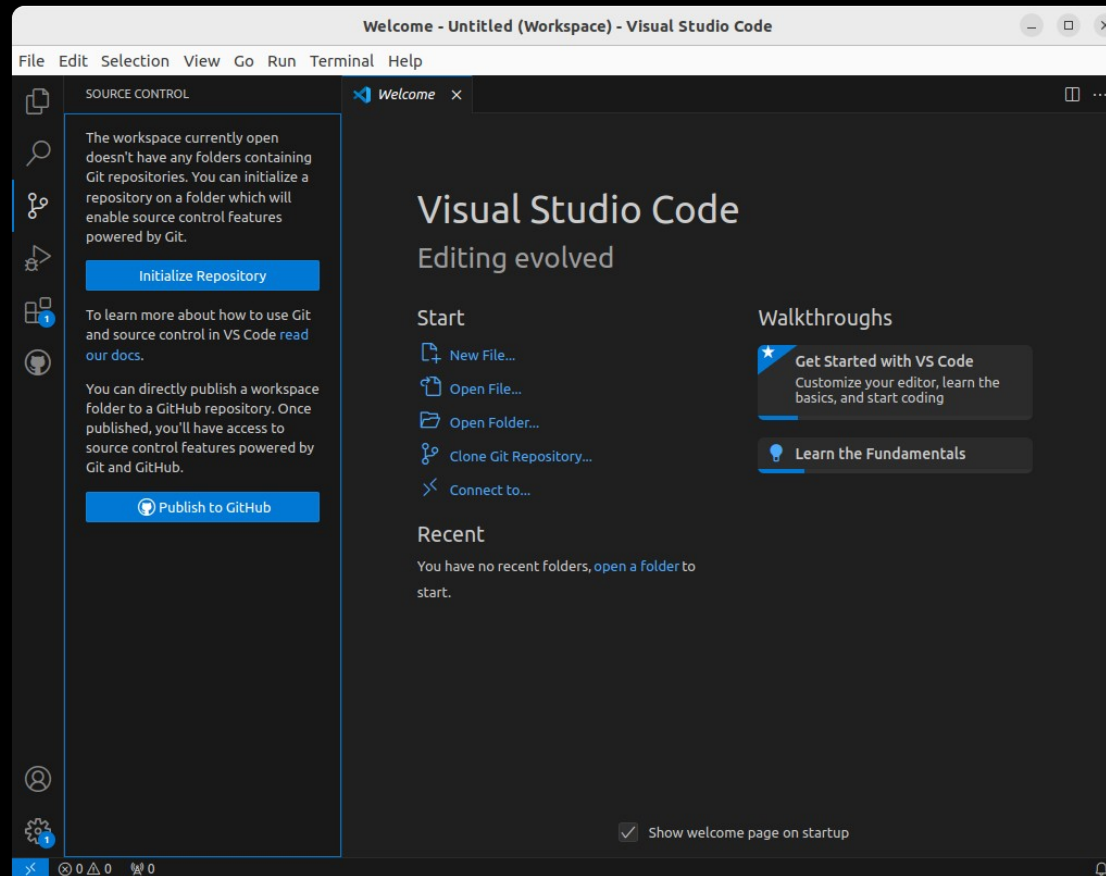
# Setting up Git in VS Code



# Creating a new project (create/open a folder)



# Initialize Repository (git init)





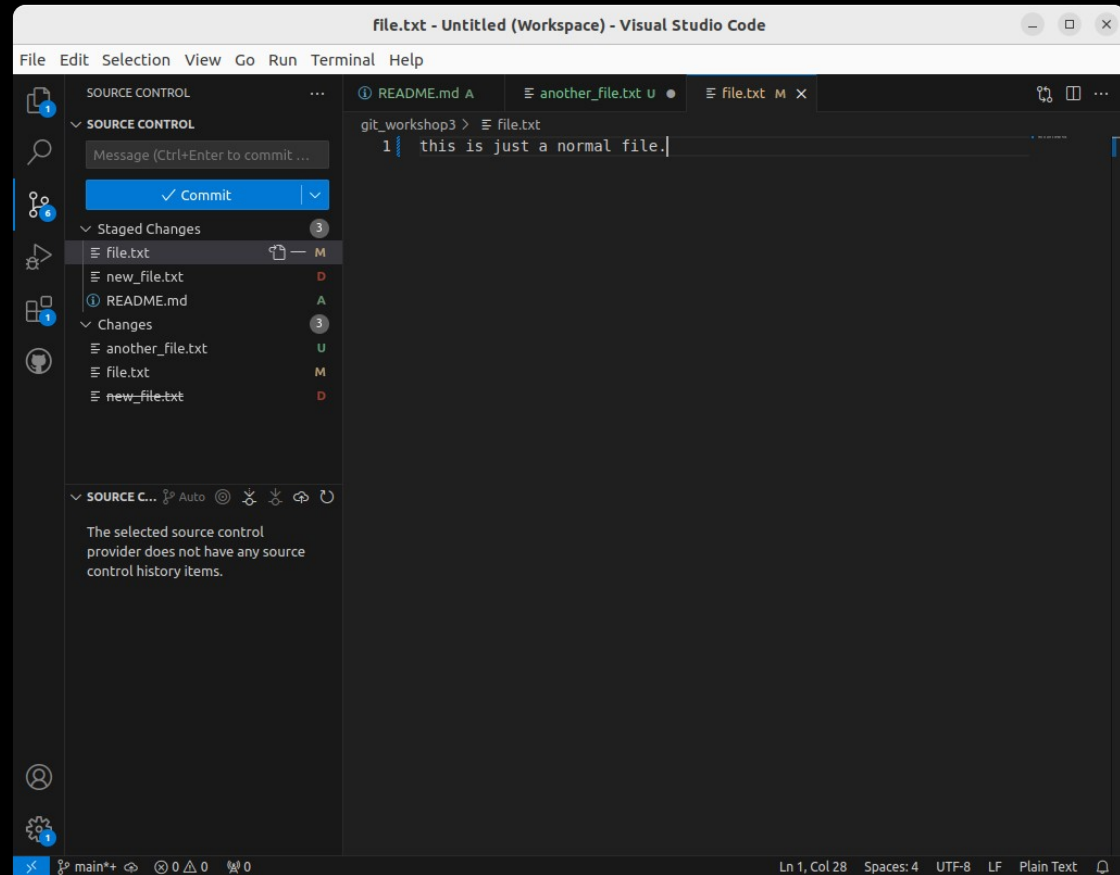
# Staging and Source Control Panel

U for untracked

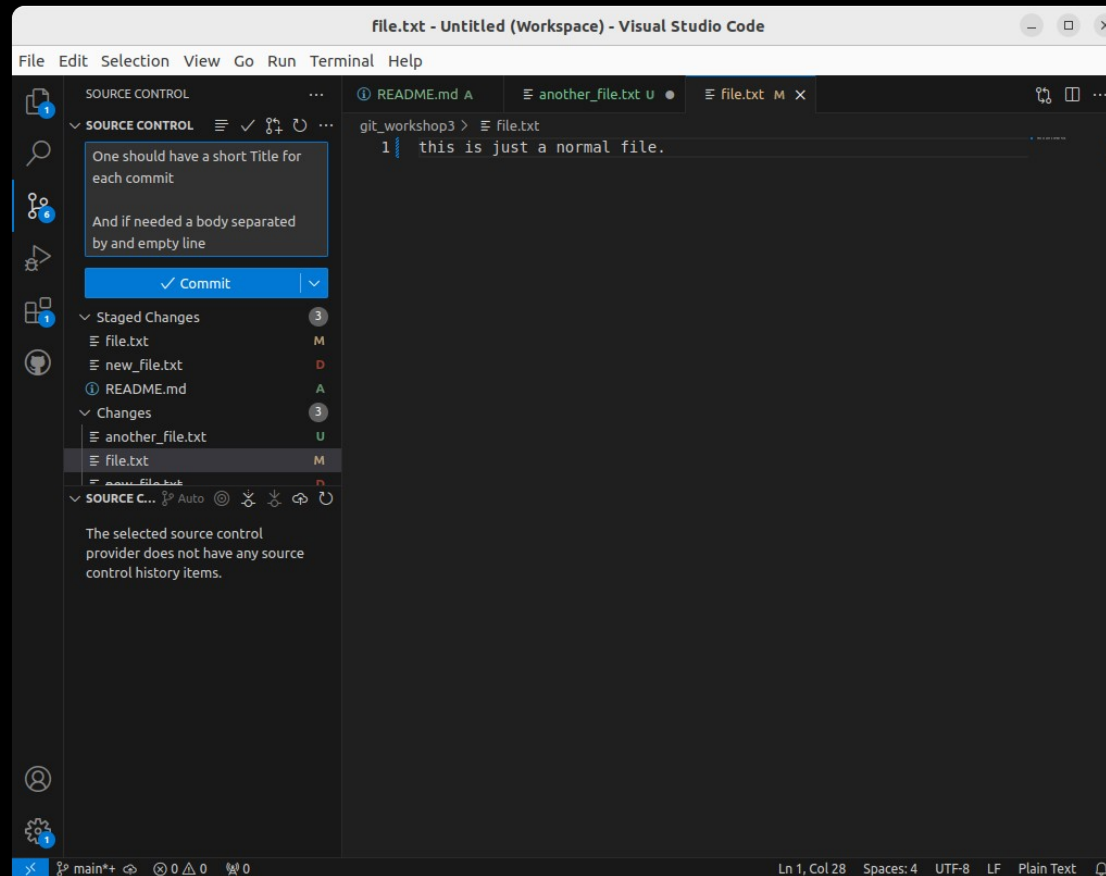
M for modified

A for added

D for deleted



# Commit



# Effective use of git commit

Commit often – Regular commits help reducing the risk of conflicts

(Try to) Commit one feature at a time – Group single logical changes to a single commit

Avoid committing incomplete work – Only commit once you finished a feature

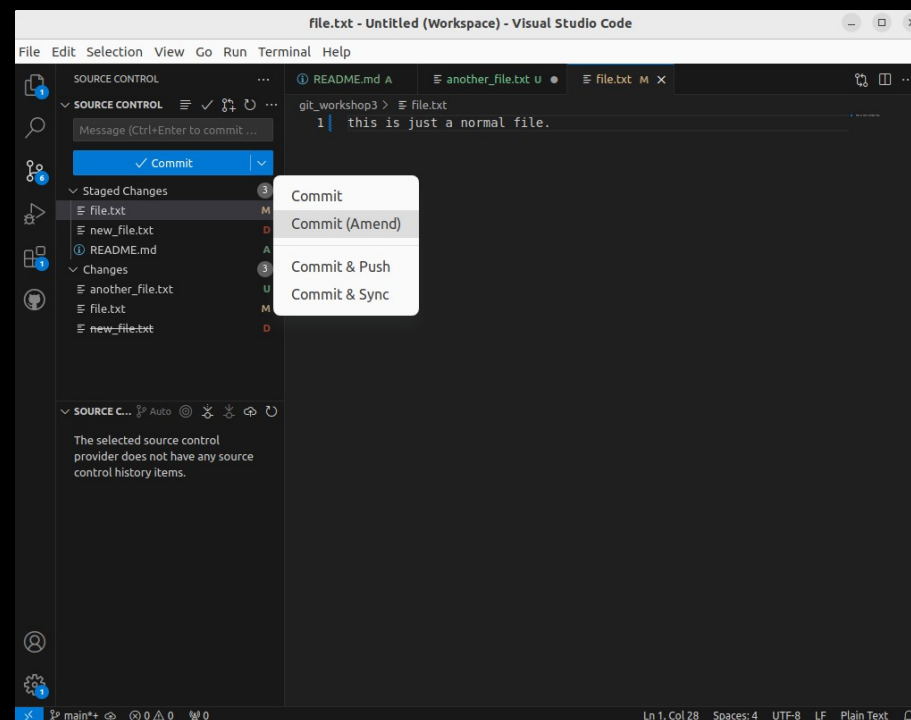
- Use stashing to avoid losing progress without committing
- Use Branches if a new feature is needed while another one is still in development

Write clear commit Messages

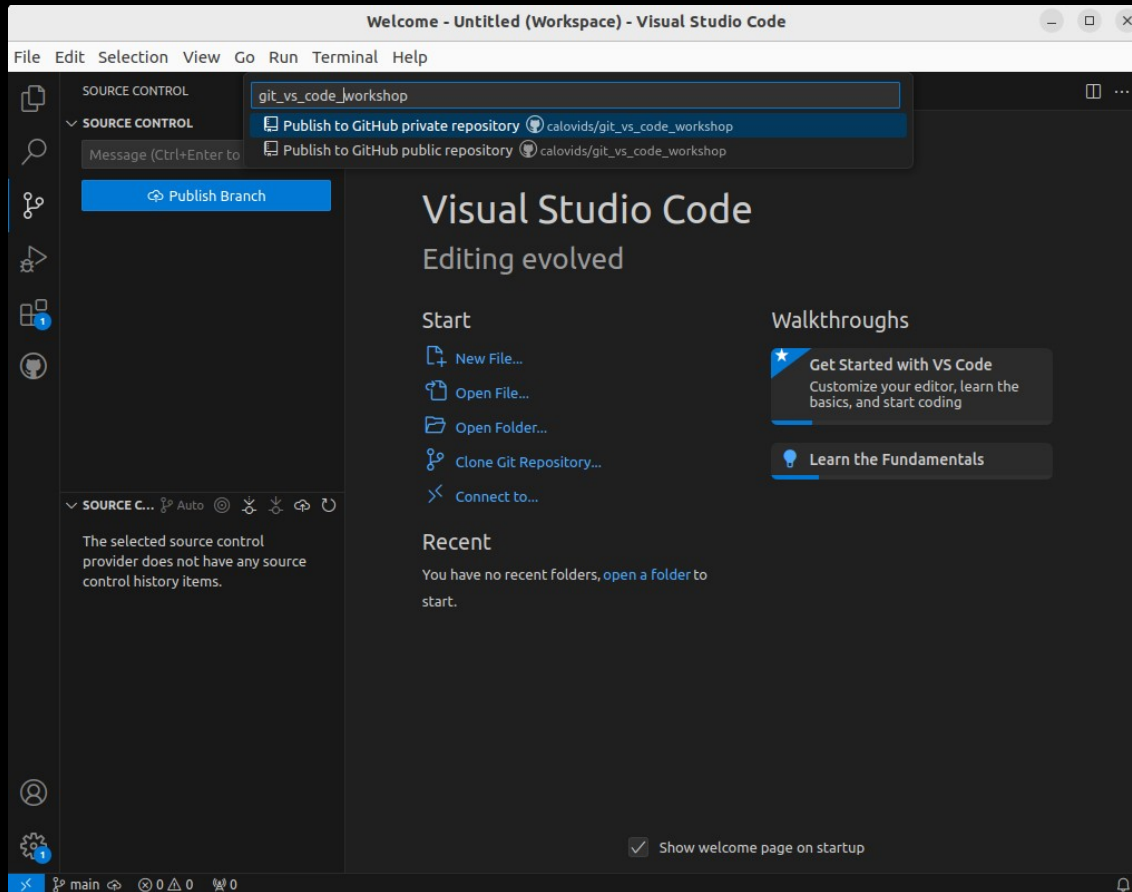
- Short subjects
- Separate a body from the subject by adding a blank line

Commit Amend is your friend – Just submitted and there were files or changes missing?

- `$ git commit -m 'Initial commit'`
- `$ git add forgotten_file`
- `$ git commit --amend`



# Publish (pushing for the 1<sup>st</sup> time)



# Branching

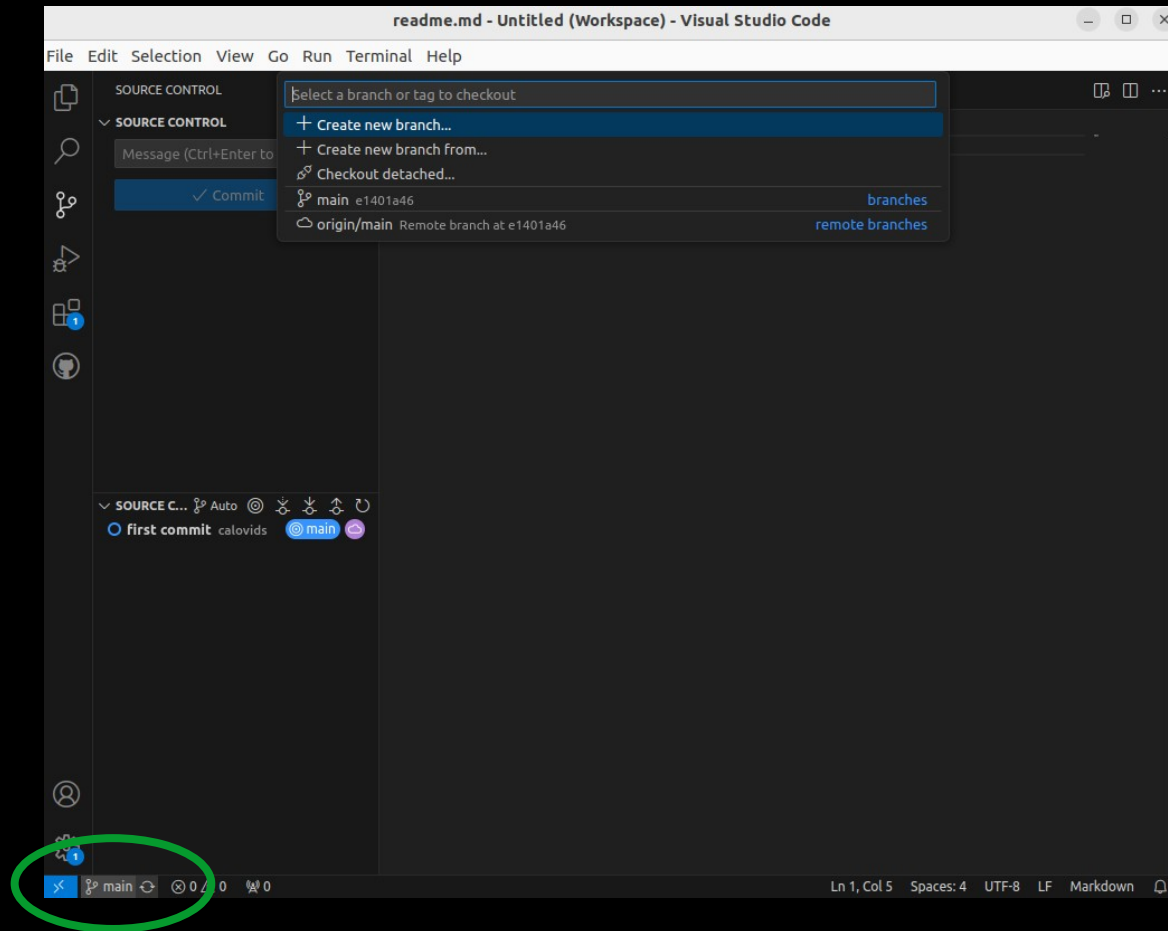
A branch is like a parallel timeline of your project

A new feature, a bug correction, or multiple fixes for releasing a new version

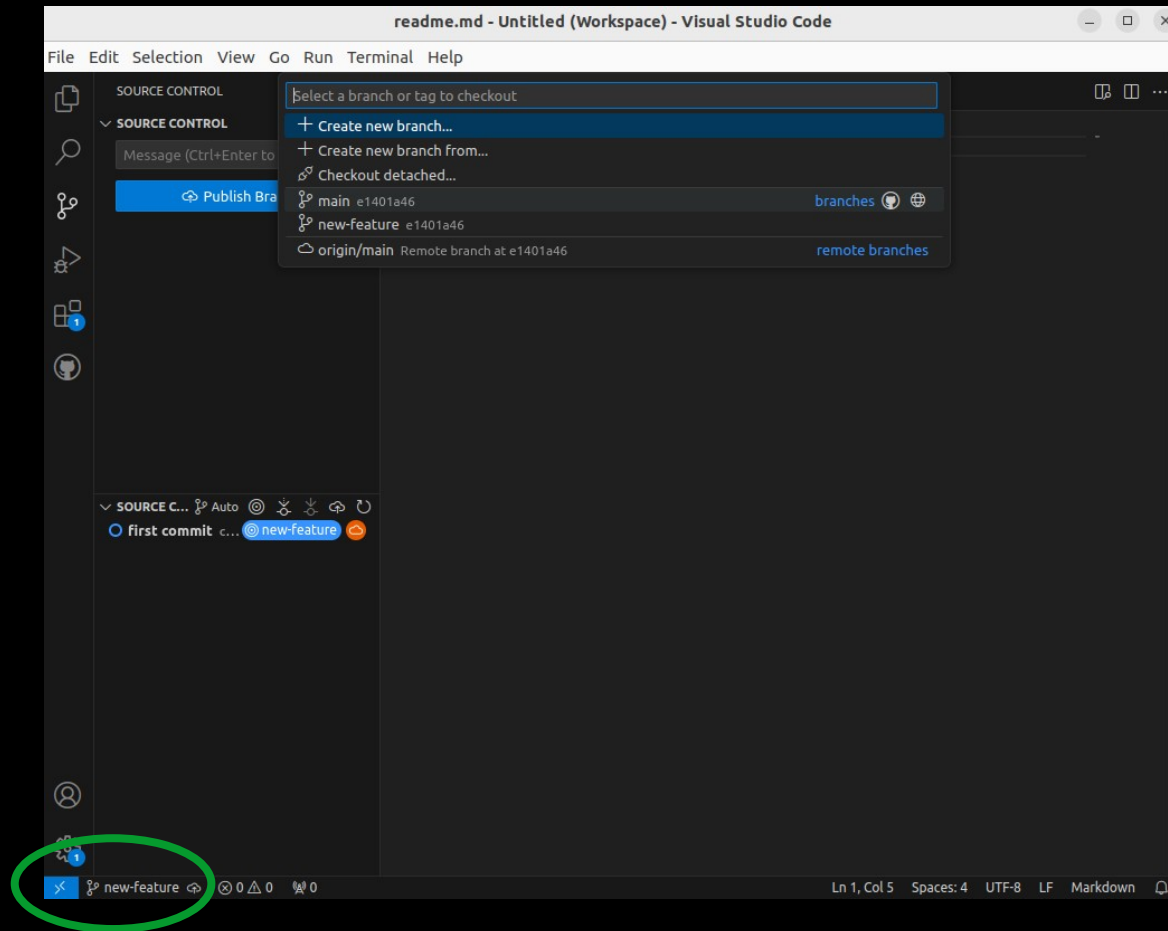
Imagine this case

- You want to add a new feature to your code
- After weeks of trying it's not working out
- You need go to back and use the old code for something else

# Branching



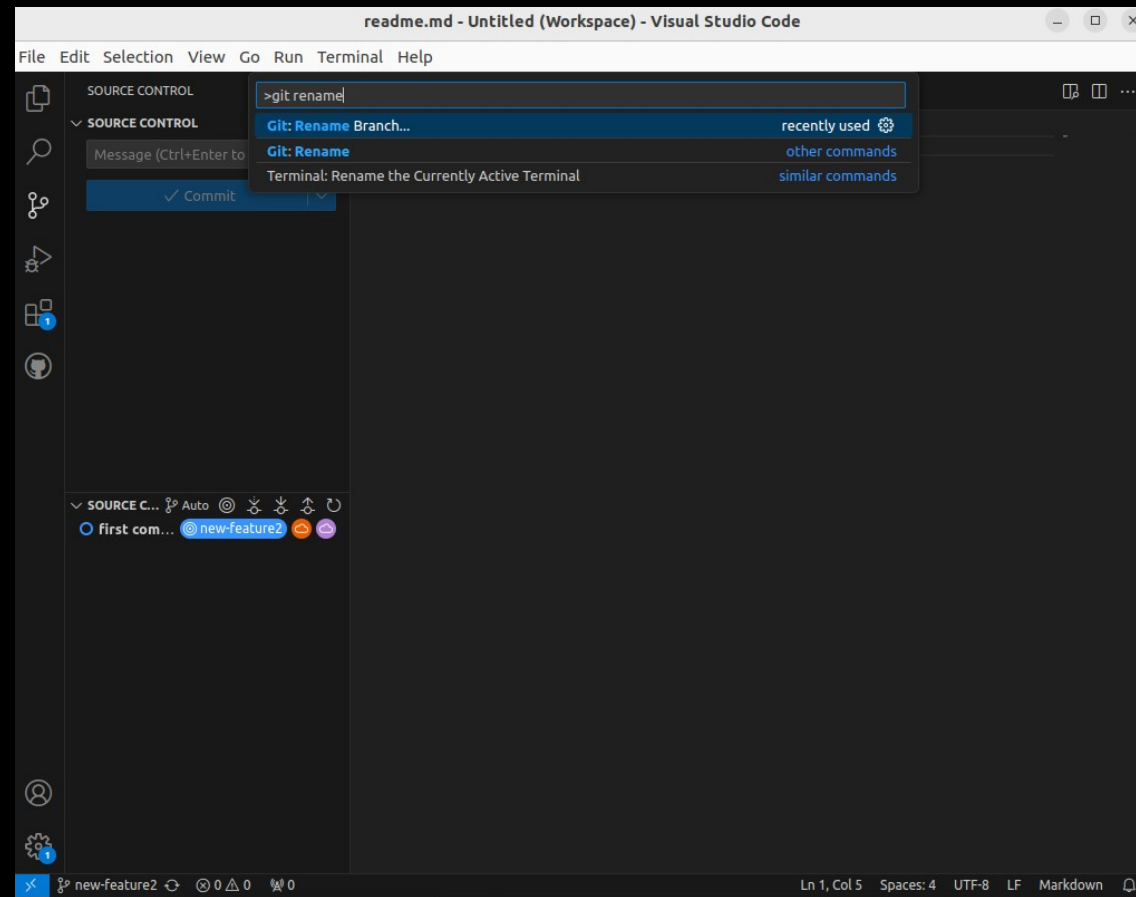
# checkout



# Renaming a branch

A simpler way to rename branches in VS Code:

- Ctrl+Shift+P (Windows/Linux)
- Cmd+Shift+P (MacOS)

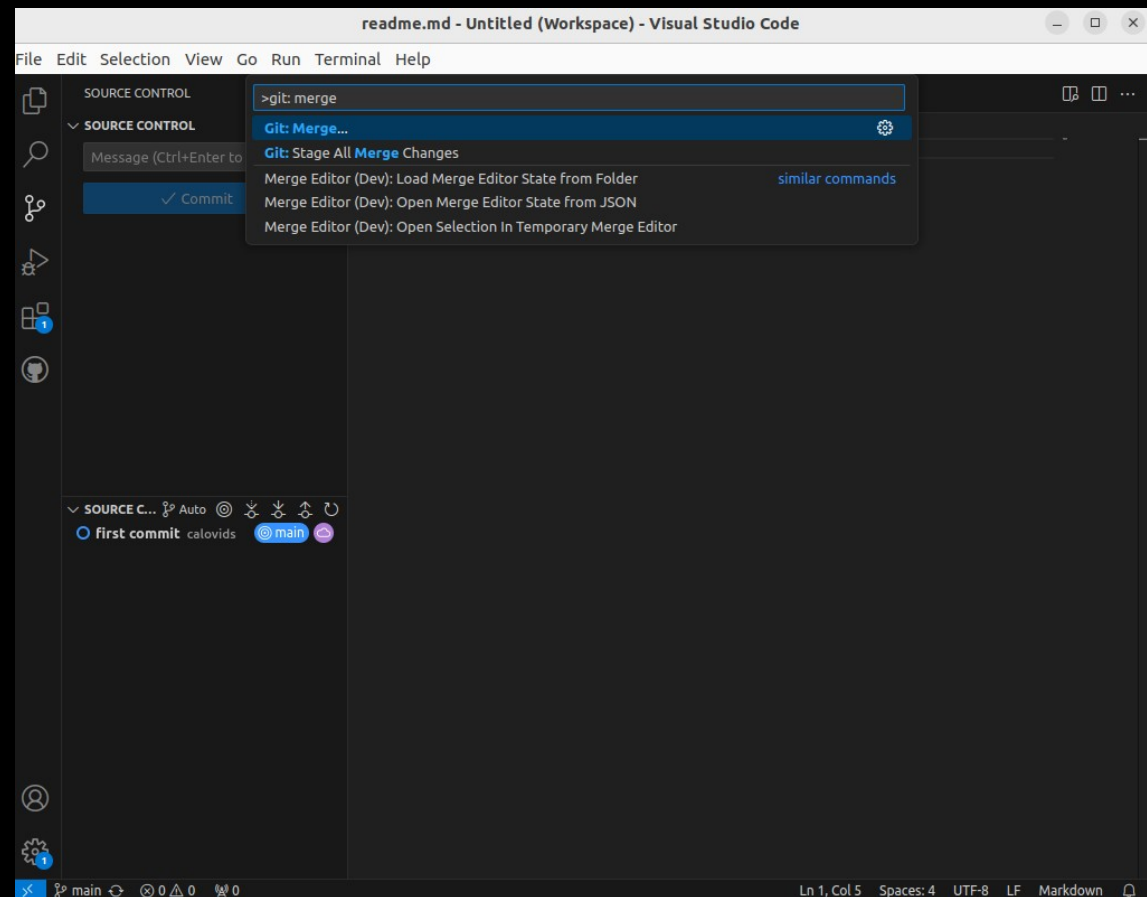




# Merging

Merges a different branch into current branch

- Ctrl+Shift+P (Windows/Linux)
- Cmd+Shift+P (MacOS)



# .gitignore File

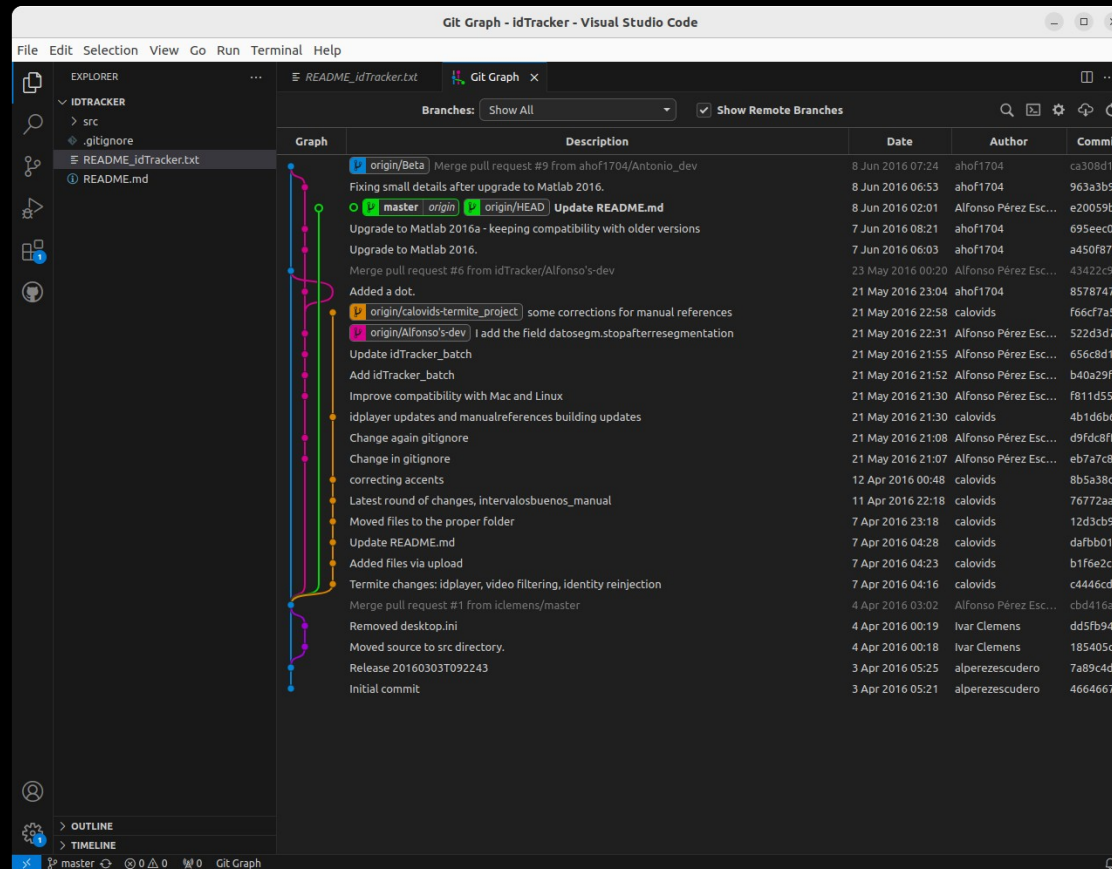
In your working folder there will be files that change often and don't need to be tracked

- Log files
- Thumbnail libraries
- Recently changed files

Also files that you simply do not want to be tracked

- Environment files (.env)
- Secrets

# VS code extensions – Git Graph



# Test time

Create a new repository

Add a readme file

- Stage and commit

Create a new branch

- Add a new file
- commit & sync

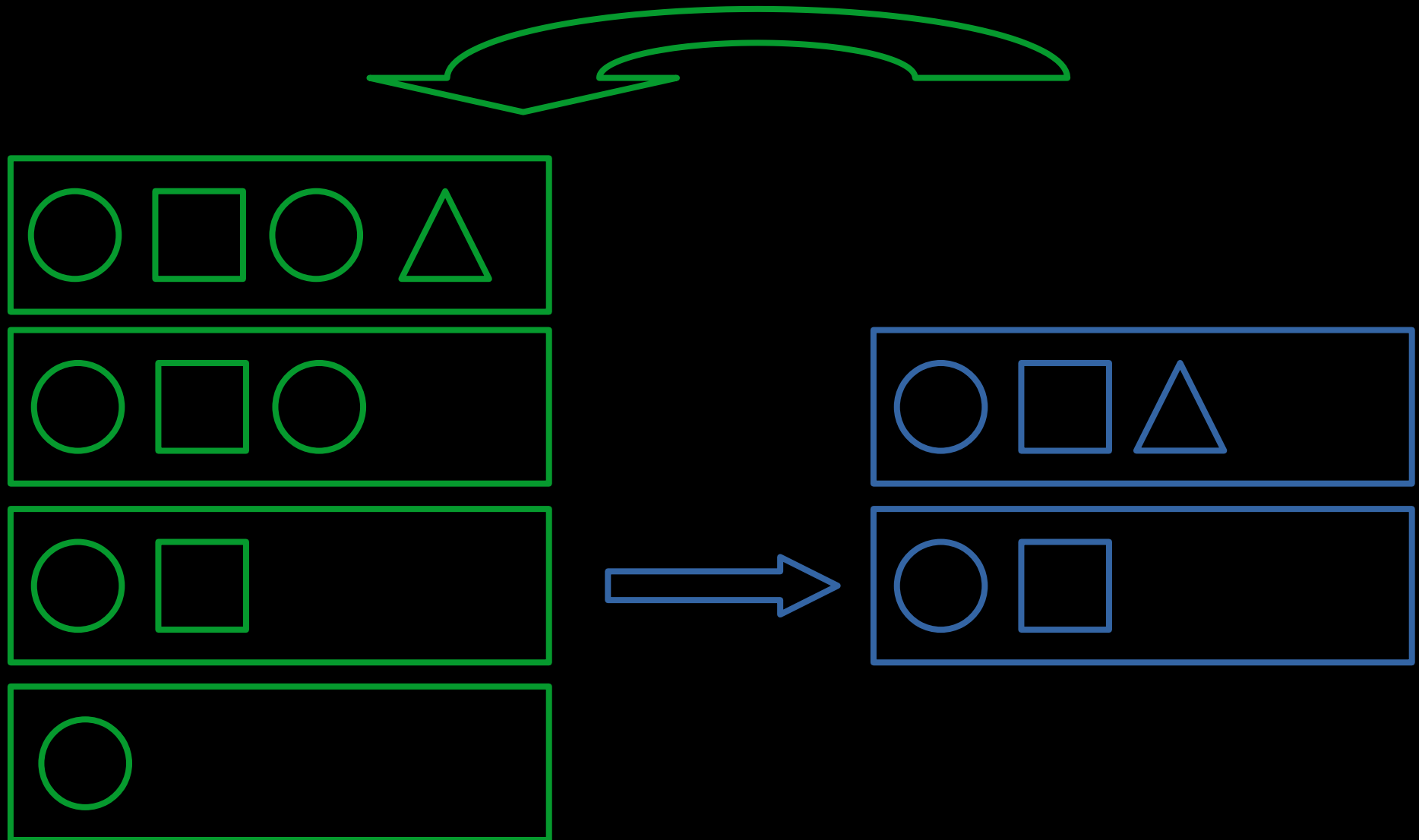
Merge back into main branch

Sync

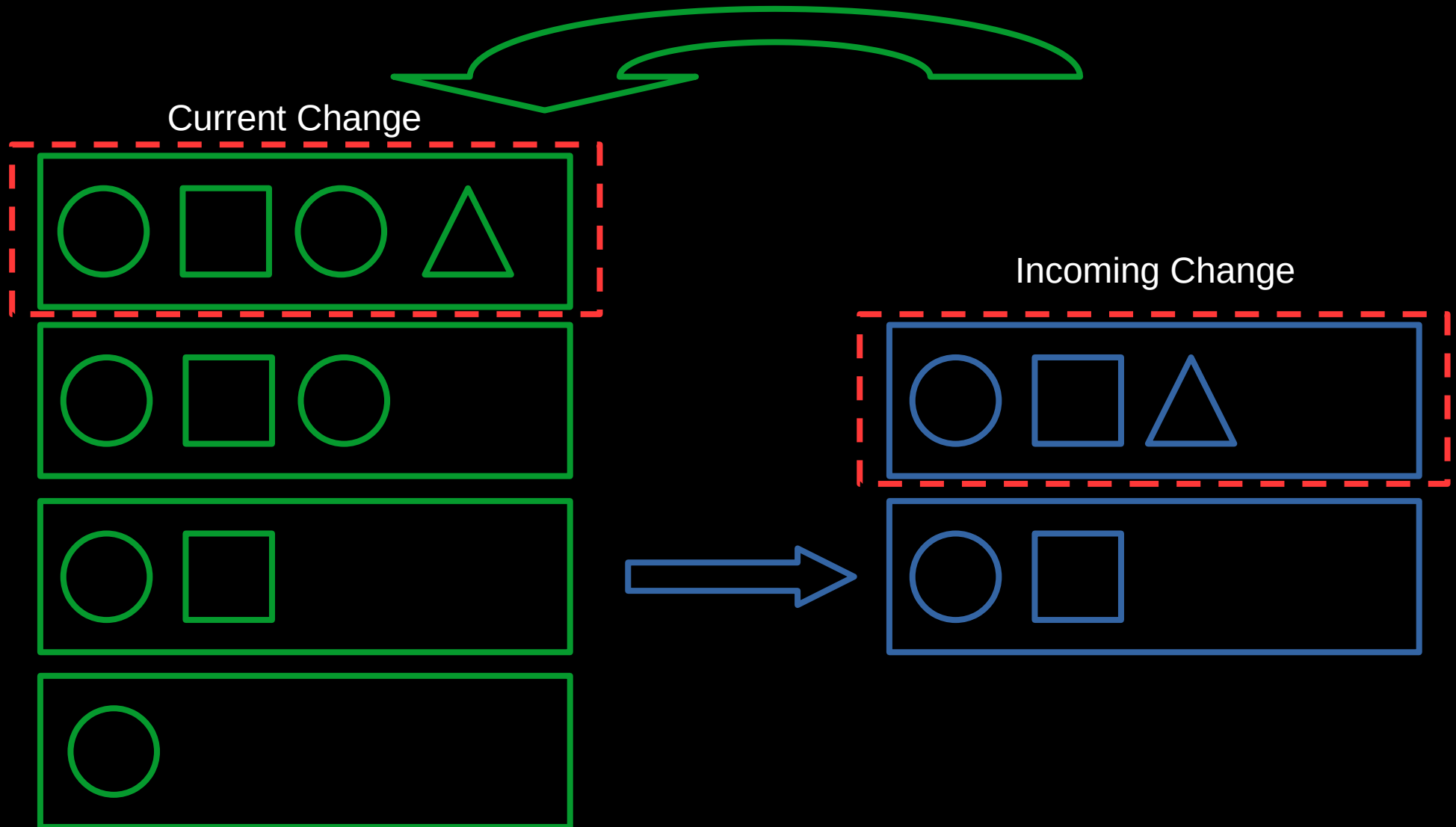
Send me a screenshot of your Git Graph by email

# What are conflicts?

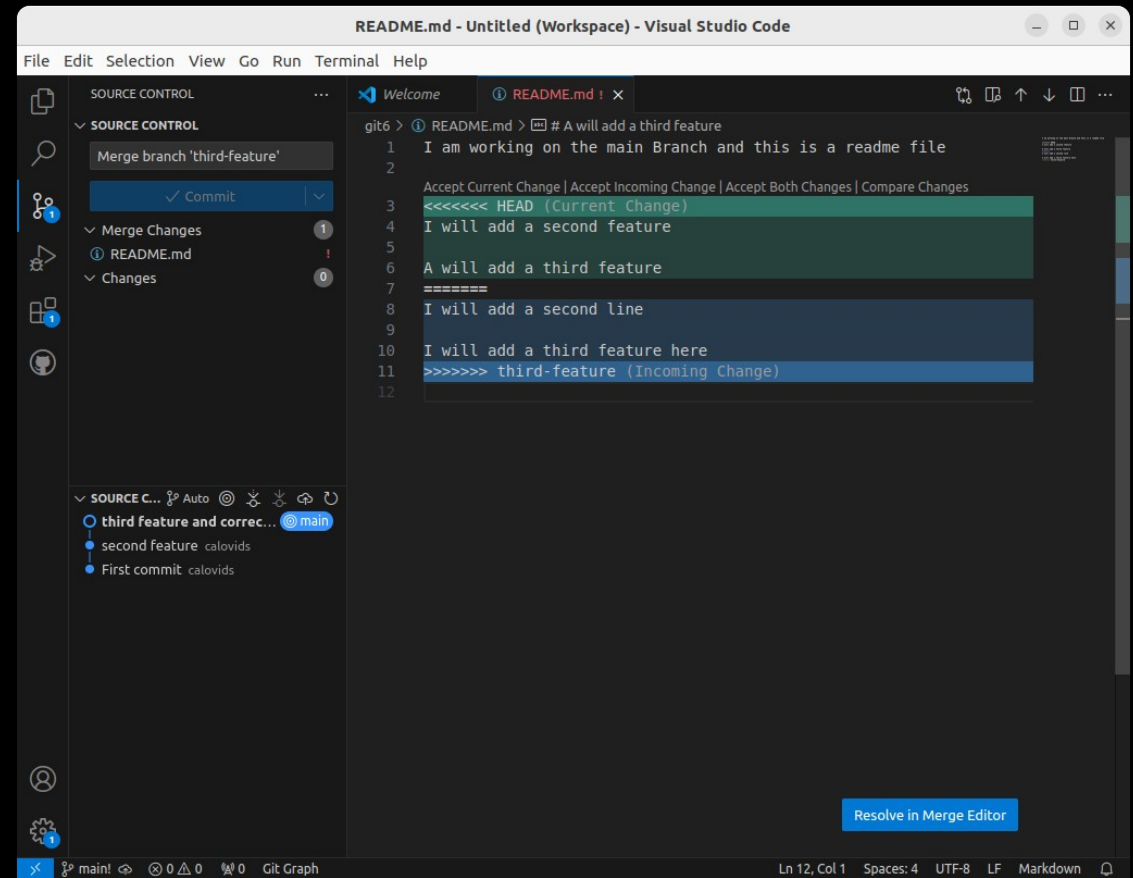
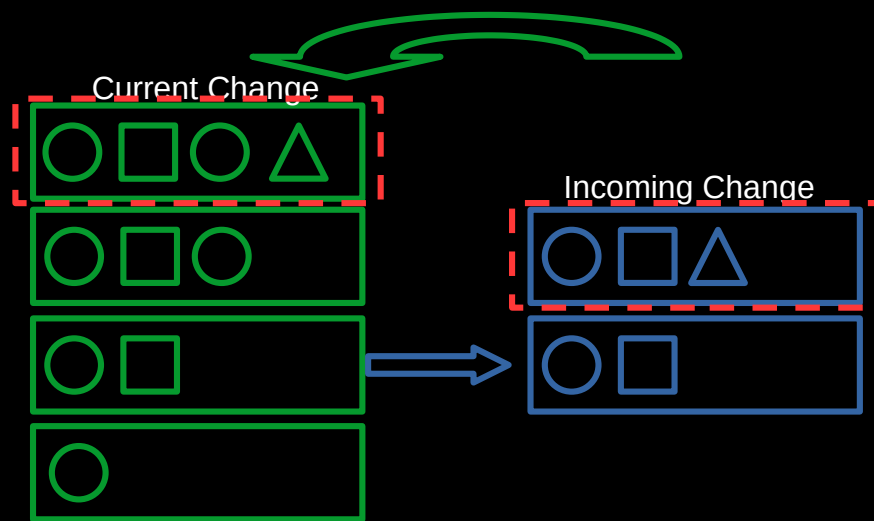
# Conflicts



# Conflicts

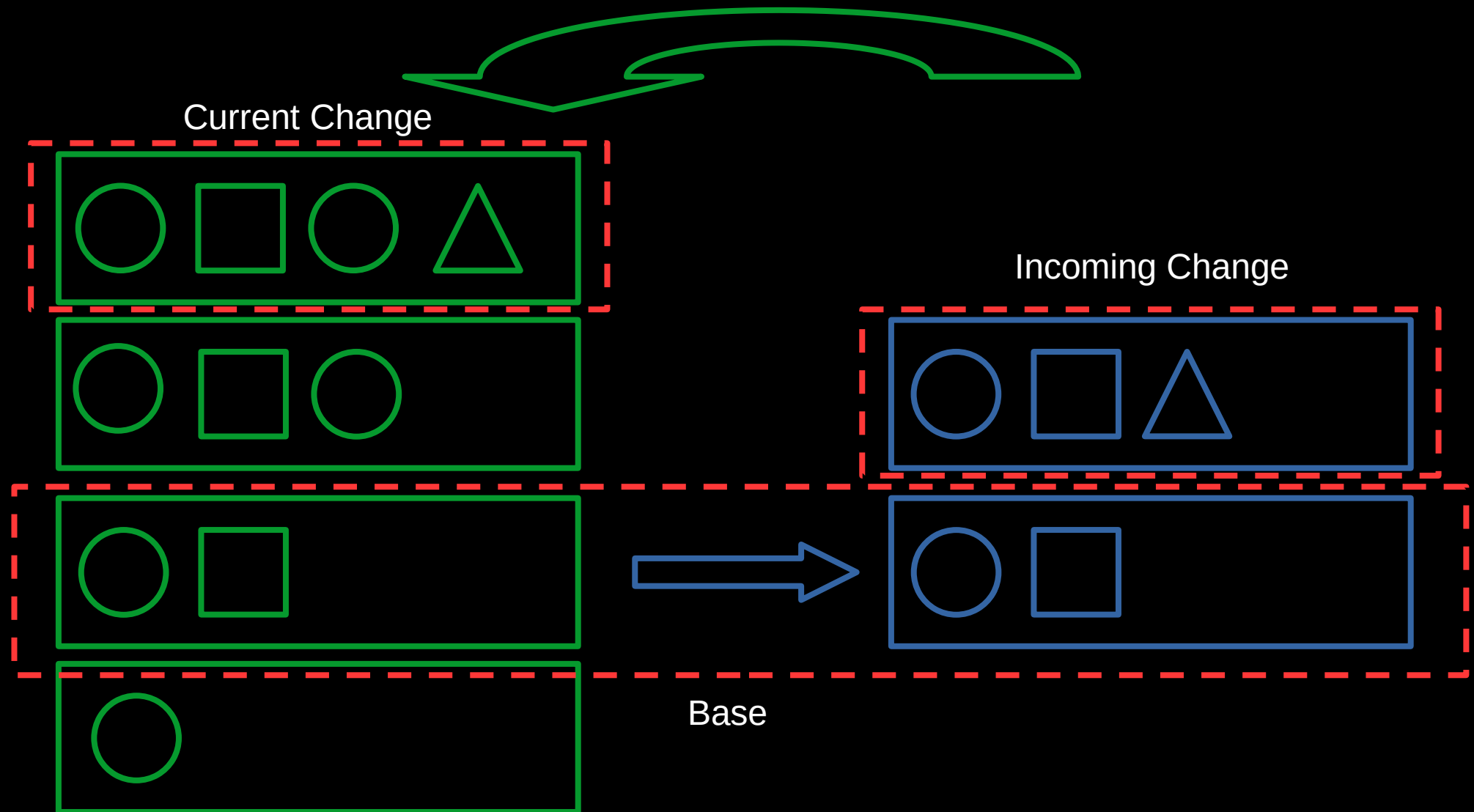


# Conflicts

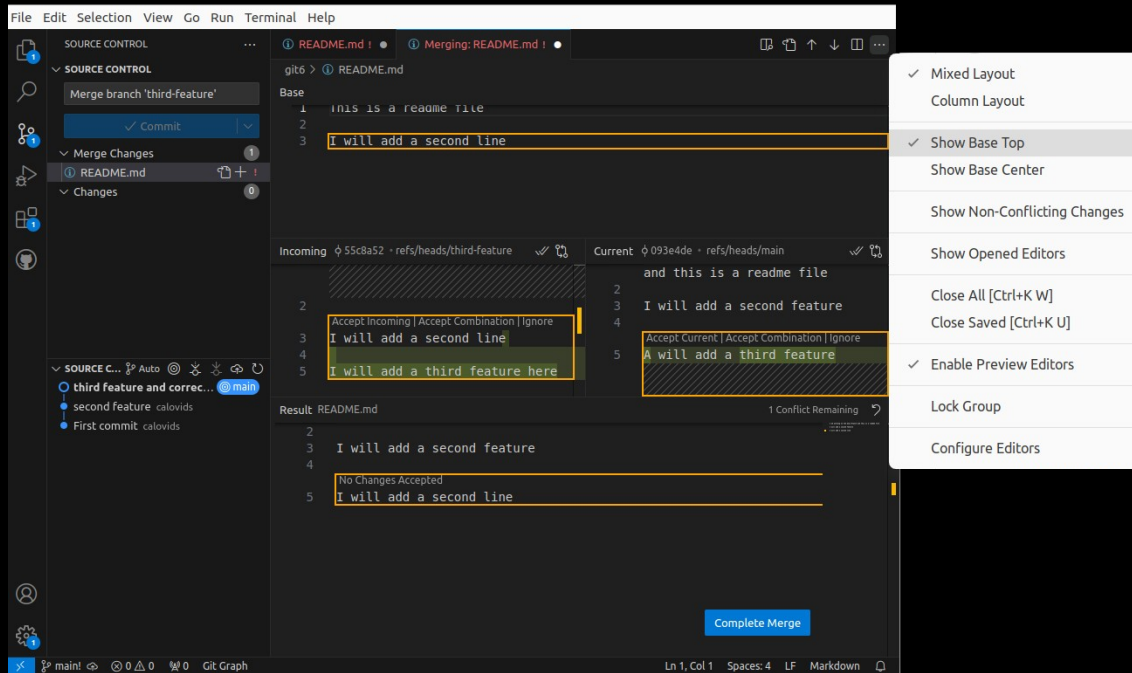




# Conflicts



# Conflicts



# Questions/Feedback

Please send any feedback to: [daniel.calovi@uni-konstanz.de](mailto:daniel.calovi@uni-konstanz.de)