



Data Science Lab

GitHub tutorial

DataBase and Data Mining Group

Andrea Pasini, Elena Baralis

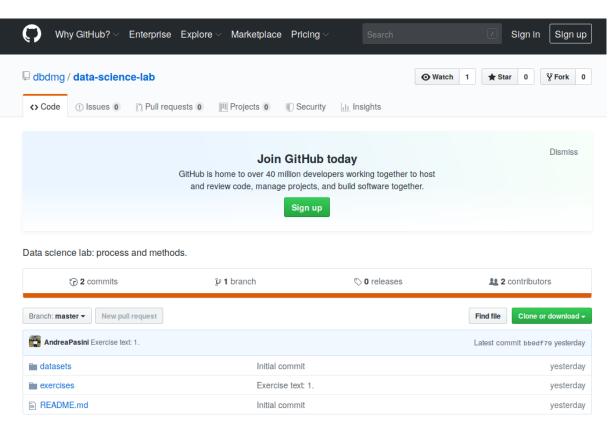






- GitHub is a website/service for code versioning
- Projects are stored in "repositories"

Give the URL of a repository, you can browse files with your browser:

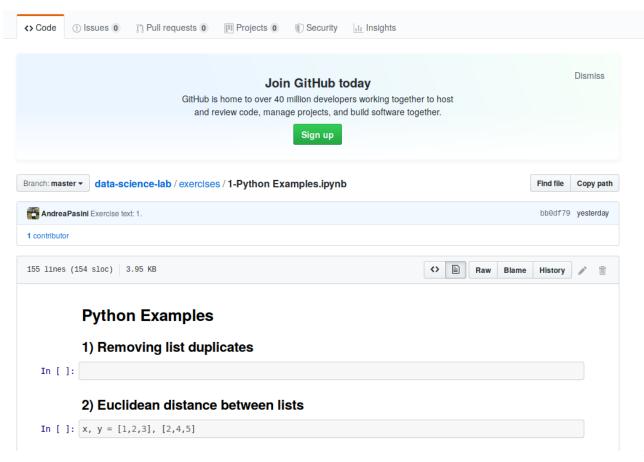








You can preview a Python notebook by clicking on its link:

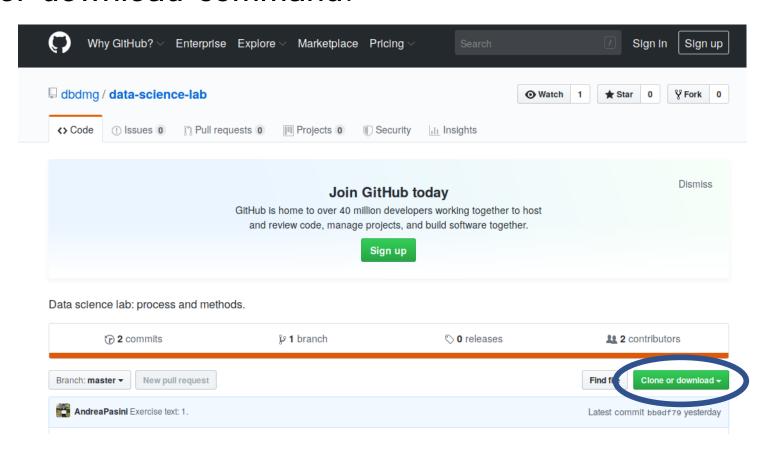








To download a repository you can use the clone or download command:









- Using the Git command on your terminal
 - This command allows keeping updated the repository on your pc without downloading it every time from your browser
- Installation:
 - For Windows:
 - Download and install Git:
 - https://git-scm.com/
 - For Ubuntu/Mac
 - Run on your terminal:
 - sudo apt-get install git







- Using the Git command on your terminal
 - Create a new folder (e.g. dsl_repository) where you want to store the repository
 - Open a terminal, change directory to the folder you created
 - Example (if you created dsl_repository in Documents):
 - cd ~/Documents/dsl repository







- Using the Git command on your terminal
 - Copy the URL of the repository from your browser
 - Example: https://github.com/dbdmg/data-science-lab
 - Type in your terminal
 - git clone https://github.com/dbdmg/data-science-lab
 - This operation will download the repository to your folder

```
File Modifica Visualizza Cerca Terminale Aiuto

andrea@andrea :~/Documenti/Dottorato/didattica/DataScienceLab/dsl_repo$ git

clone https://github.com/dbdmg/data-science-lab

Cloning into 'data-science-lab'...
remote: Enumerating objects: 12, done.
remote: Counting objects: 100% (12/12), done.
remote: Compressing objects: 100% (12/12), done.
remote: Total 12 (delta 1), reused 11 (delta 0), pack-reused 0

Jnpacking objects: 100% (12/12), done.
andrea@andrea :~/Documenti/Dottorato/didattica/DataScienceLab/dsl_repo$
```





- Using the Git command on your terminal
 - Whenever someone changes the online content of the repository, you have to **pull** the changes, to see them in your file system
 - To do that you have to move inside the repository folder and run the pull command:
 - cd ./data-science-lab
 - git pull

```
File Modifica Visualizza Cerca Terminale Aiuto

andrea@andrea :~/Documenti/Dottorato/didattica/DataScienceLab/dsl_rep
o/data-science-lab$ git pull
Already up to date.
```







- Using the Git command on your terminal
 - Already up to date means that there were no changes to update on your local version of the repository

```
File Modifica Visualizza Cerca Terminale Aiuto

andrea@andrea :~/Documenti/Dottorato/didattica/DataScienceLab/dsl_rep
o/data-science-lab$ git pull
Already up to date.
```

If it is not up to date, the git pull command will update the files according to the content that can be found on the online repository







- Changing the content of a repository
 - Note: Git also allows to update the content of a repository, but:
 - The repository of this course is read only
 - Note: if you modify some files in your local repository, then you run the git pull command, you will loose all the modifications