



POLITECNICO
DI TORINO



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:

The screenshot shows the GitHub interface for the repository `dbdmg / data-science-lab`. At the top, there is a navigation bar with links for 'Why GitHub?', 'Enterprise', 'Explore', 'Marketplace', and 'Pricing', along with a search bar and 'Sign In' and 'Sign up' buttons. Below the navigation bar, the repository name is displayed, followed by 'Watch 1', 'Star 0', and 'Fork 0' buttons. A navigation menu includes 'Code', 'Issues 0', 'Pull requests 0', 'Projects 0', 'Security', and 'Insights'. A prominent green banner encourages users to 'Join GitHub today' with a 'Sign up' button. Below the banner, the repository description reads 'Data science lab: process and methods.' and shows statistics: '2 commits', '1 branch', '0 releases', and '2 contributors'. A 'Branch: master' dropdown and a 'New pull request' button are visible. A 'Find file' button and a 'Clone or download' button are also present. The file list includes:

File Name	Commit Message	Commit Date
<code>datasets</code>	Initial commit	yesterday
<code>exercises</code>	Exercise text: 1.	yesterday
<code>README.md</code>	Initial commit	yesterday



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

The screenshot shows a GitHub repository page for a Python notebook. At the top, there are navigation tabs: Code, Issues (0), Pull requests (0), Projects (0), Security, and Insights. Below this is a green banner with the text "Join GitHub today" and a "Sign up" button. The repository path is "data-science-lab / exercises / 1-Python Examples.ipynb". The notebook content is displayed in a preview mode, showing the title "Python Examples" and two sections: "1) Removing list duplicates" and "2) Euclidean distance between lists". The first section has an input field for code, and the second section has an input field containing the code "x, y = [1,2,3], [2,4,5]".



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

The screenshot shows the GitHub interface for the repository 'dbdmg / data-science-lab'. At the top, there is a navigation bar with links for 'Why GitHub?', 'Enterprise', 'Explore', 'Marketplace', and 'Pricing', along with a search bar and 'Sign In' and 'Sign up' buttons. Below the navigation bar, the repository name 'dbdmg / data-science-lab' is displayed, followed by 'Watch 1', 'Star 0', and 'Fork 0' buttons. A secondary navigation bar includes 'Code', 'Issues 0', 'Pull requests 0', 'Projects 0', 'Security', and 'Insights'. A large light blue banner with the text 'Join GitHub today' and 'Sign up' is visible. Below the banner, the repository description 'Data science lab: process and methods.' is shown. A summary bar indicates '2 commits', '1 branch', '0 releases', and '2 contributors'. At the bottom, there is a 'Branch: master' dropdown, a 'New pull request' button, a 'Find file' button, and a 'Clone or download' button which is circled in blue. Below this, a commit by 'AndreaPasini' is listed with the message 'Exercise text: 1.' and the commit hash 'bb0df79'.



- 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
Unpacking 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_repo/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_repo/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**