

Lecture 4: numpy, Git/GitHub

LING 1340/2340: Data Science for Linguists

Jevon Heath

Objectives

- ▶ GitHub class organization
 - ◆ Housekeeping our repos and forks
- ▶ To-do2 review: study notes in JNB
- ▶ Python's `numpy` library

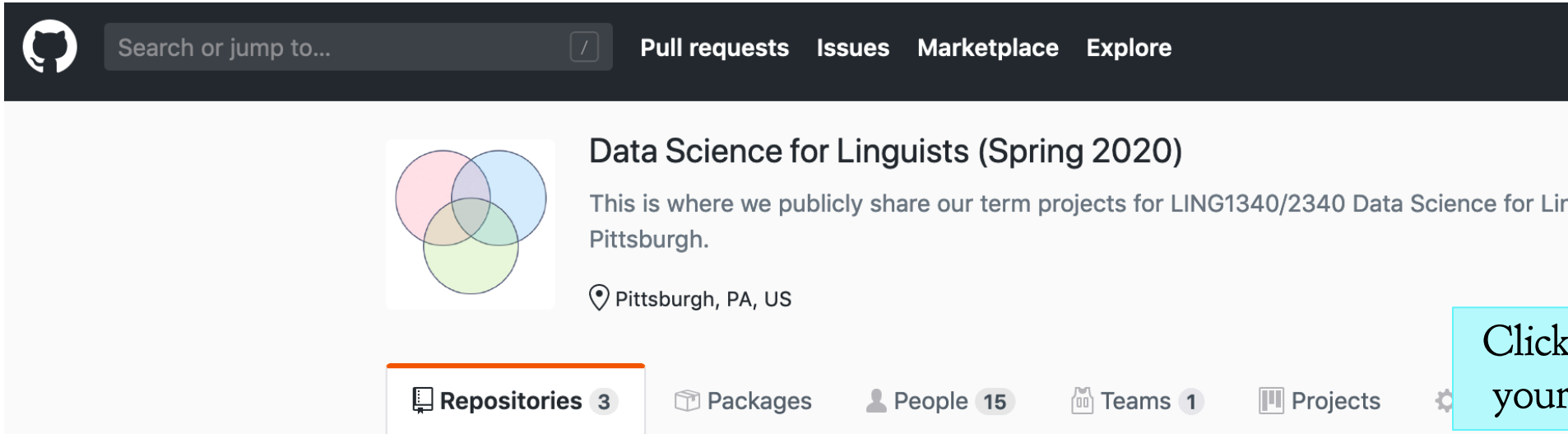
- ▶ Tools:
 - ◆ Git and GitHub
 - ◆ Jupyter Notebook

Bits and things

- ▶ Office hours working?
 - ◆ Cassie: Th → F?
 - ◆ John: T → M?
- ▶ Syllabus review:
 - ◆ Final project
 - ◆ Todos
- ▶ Jupyter Notebook "style" tips

GitHub Class Organization

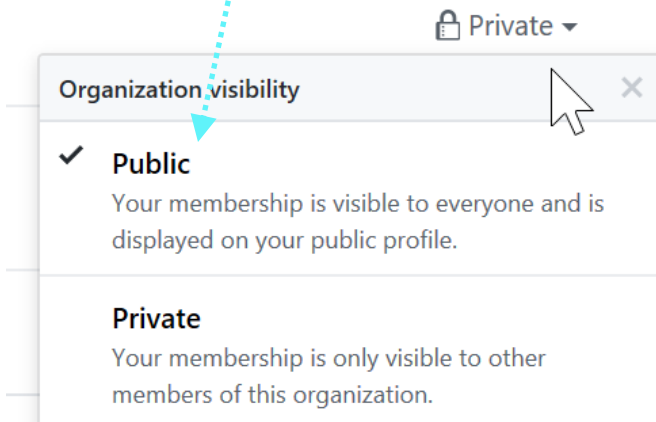
► <https://github.com/Data-Science-for-Linguists-2020>



The screenshot shows the GitHub organization page for "Data Science for Linguists (Spring 2020)". The header includes the GitHub logo, a search bar, and navigation links for Pull requests, Issues, Marketplace, and Explore. The organization's profile picture is a Venn diagram with four overlapping circles in pink, blue, green, and yellow. The bio states: "This is where we publicly share our term projects for LING1340/2340 Data Science for Ling Pittsburgh." The location is listed as "Pittsburgh, PA, US". Below the profile information, there are tabs for Repositories (3), Packages, People (15), Teams (1), and Projects. A red box highlights the "People" tab.

Click "People", and make your membership public

- So we can:
- ◆ have everyone in one spot.
 - ◆ have all class materials in one spot.
 - ◆ have everyone's term project in one spot.
 - ◆ share *private* repos as a group.



The screenshot shows the "Organization visibility" dropdown menu. At the top, there is a "Private" button with a lock icon. The dropdown menu has two options: "Public" (checked with a green checkmark) and "Private". The "Public" option description is: "Your membership is visible to everyone and is displayed on your public profile." The "Private" option description is: "Your membership is only visible to other members of this organization." A red arrow points from the text "Click 'People', and make your membership public" to the "Public" option.

Shall we try re-forking?

► On GitHub

- ◆ **Delete** your forks: HW1-Repo and Class-Exercise-Repo.
- ◆ Head to our GitHub class organization. **Fork** the two repositories.

► On your laptop

- ◆ Your local repos' remote setting (`git remote -v`):
 - ◆ The "origin" URL is still the same.
 - ◆ The "upstream" URL can be updated:

- ◆ In HW1-Repo/ folder, execute:

```
git remote set-url upstream https://github.com/Data-Science-for-Linguists-2020/HW1-Repo.git
```

- ◆ In Class-Exercise-Repo/ folder, execute:

```
git remote set-url upstream https://github.com/Data-Science-for-Linguists-2020/Class-Exercise-Repo.git
```

Back to Class-Exercise-Repo

- ▶ <https://github.com/Data-Science-for-Linguists-2020/Class-Exercise-Repo>
- ▶ To-do2
 - ◆ Who has the best study notes?

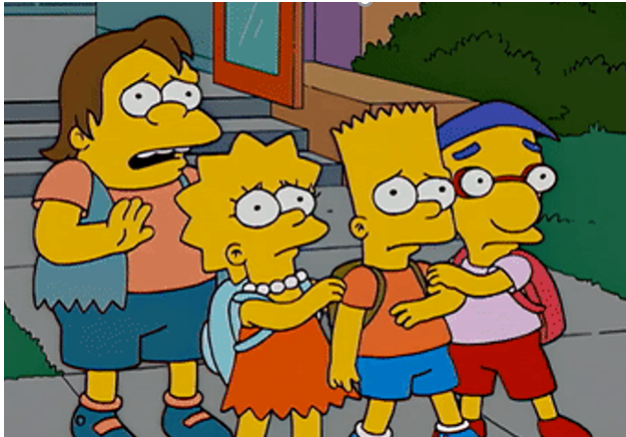
numpy practice with the Simpsons

Activity 1
40 minutes



- ▶ In Class-Exercise-Repo, activity1/ folder:
 - ◆ You will find `numpy_simpsons_BLANK.ipynb`
 - ◆ Make a copy for yourself as `numpy_simpsons_YOURNAME.ipynb`:
 - ◆ `cp xyz_BLANK.ipynb xyz_jevon.ipynb` ←
 - ◆ Let's get to work!

`cp file1 file2`
makes a copy of file1 as file2.



Wrapping up

- ▶ To-do 3: due on Tuesday.
 - ◆ Study pandas, make your own JNB.
 - ◆ Try two spreadsheet files, one your choice. (Make it small!)

- ▶ Start thinking about project ideas