



# Lecture 10: Git/GitHub

LING 1340/2340: Data Science for Linguists

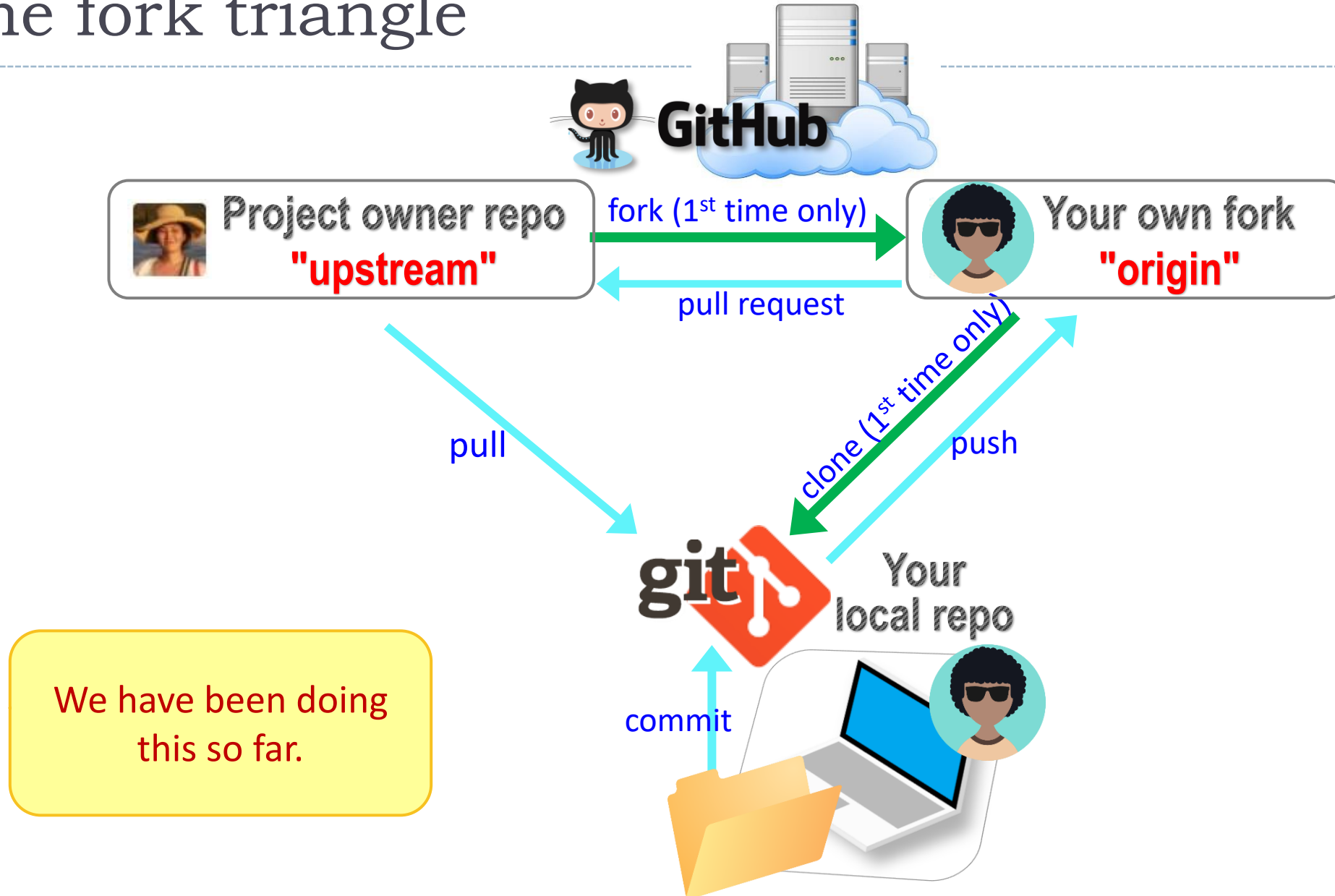
Na-Rae Han

# Objectives

---

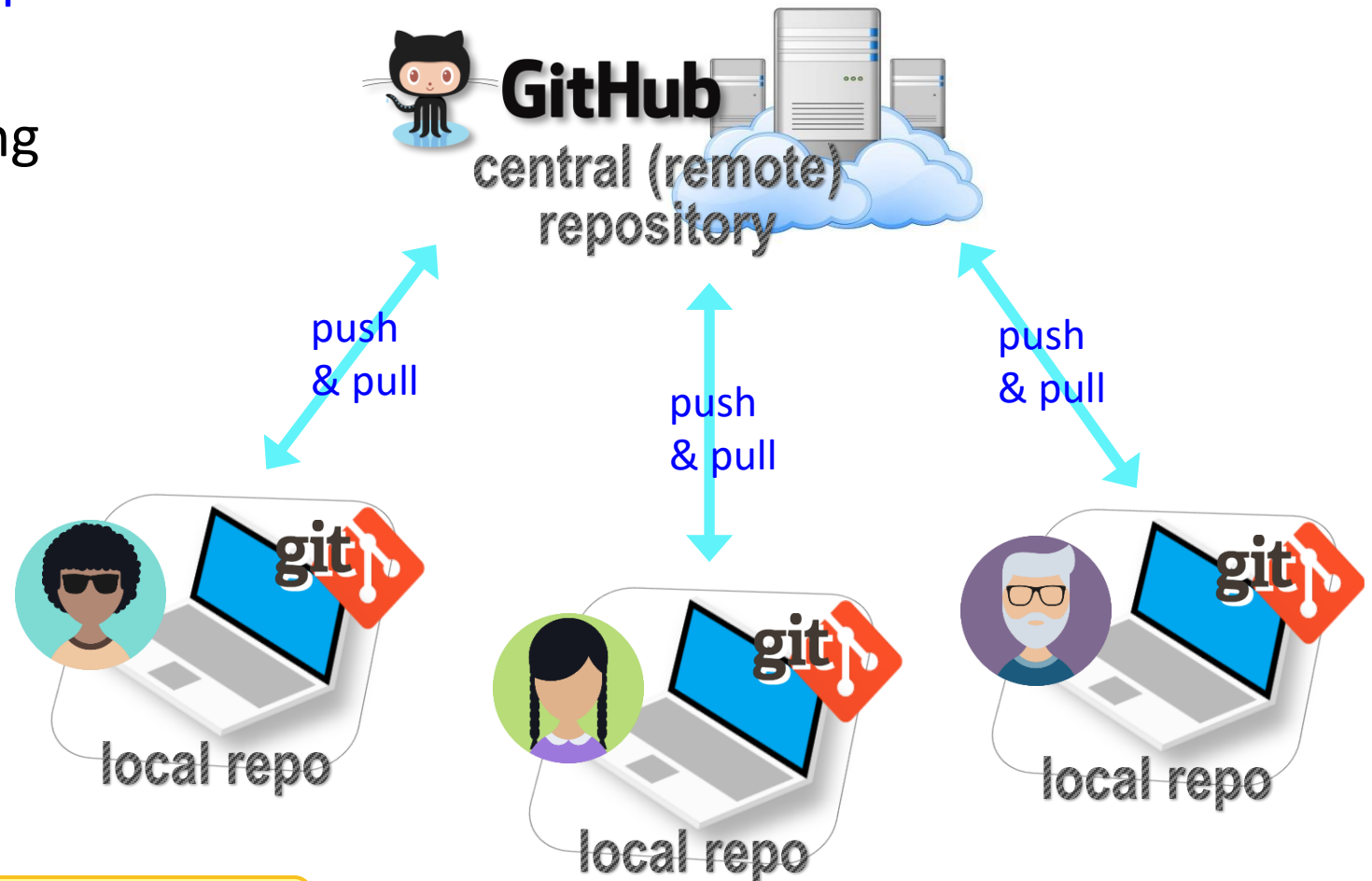
- ▶ **Git/GitHub**
  - ◆ Collaborating on the *same* repo – push access
- ▶ **HW3 review: Machine learning with ETS data**
  - ◆ Task 2: Prompt classification
  - ◆ Task 1: L1 classification
  - ◆ Task 3 → Thursday
  - ← Share your HW3! Copy your HW3 into Class-Exercise-Repo!

# The fork triangle



# GitHub: a *social*, remote repository

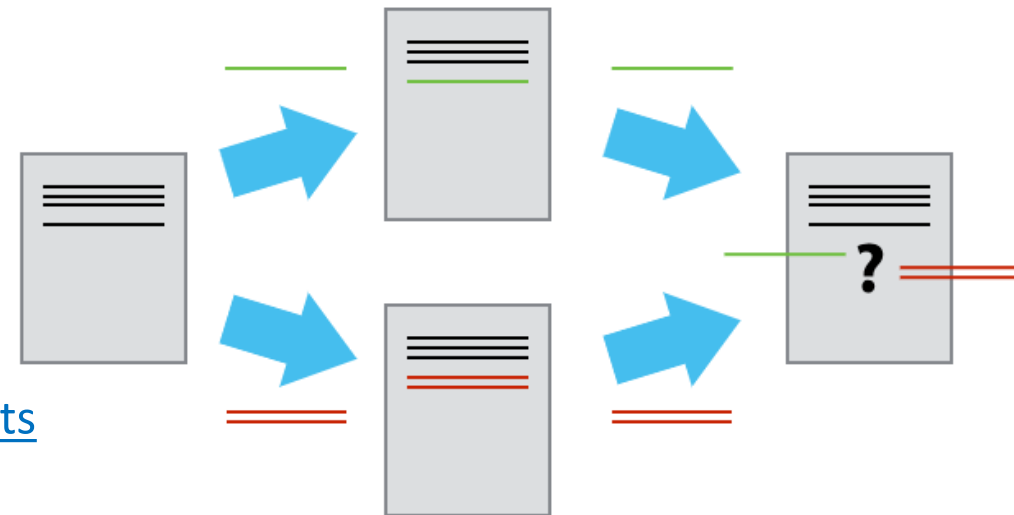
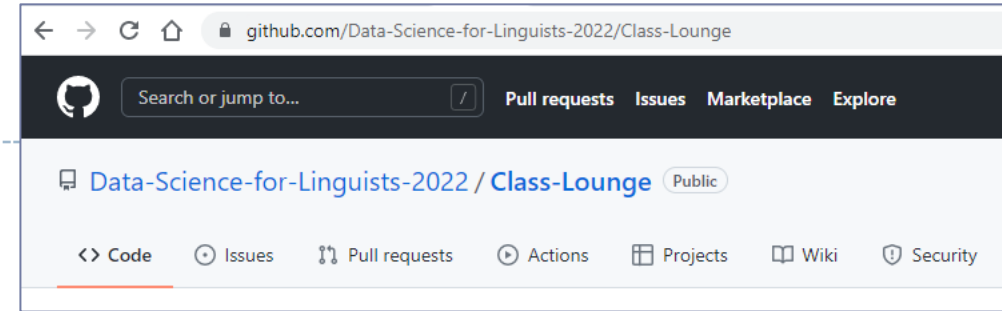
- ▶ GitHub also works as a **central remote repository** among a group of **collaborators** working on a shared project.
  - ◆ Everyone works on their own *local* copy of the repository, making changes.
  - ◆ Git is able to keep track and merge changes submitted by everyone.
  - ◆ Everyone is an **equal collaborator** with push (=write) access.



**We are now ready!**

# Introducing... "Class-Lounge"

- ▶ <https://github.com/Data-Science-for-Linguists-2022/Class-Lounge>
- ▶ Public.
- ▶ Everyone is listed as a "collaborator".
  - ◆ Meaning, everyone has push access.
  - ◆ No need to fork: pull and push directly.
  - ◆ We will also truly collaborate: **edit shared files**.
- ▶ This means: **CONFLICTS**
  - ◆ Na-Rae's tutorial on Git conflicts:
    - ◆ [https://github.com/mcdonn/LSA2019-Reproducible-Research/blob/master/linking\\_git\\_and\\_github.md#conflicts](https://github.com/mcdonn/LSA2019-Reproducible-Research/blob/master/linking_git_and_github.md#conflicts)



# When there is a conflict



- ▶ After you pull, Git changes your file, which then looks like:

```
<<<<<<< HEAD
There was copyright information here.

=====
>>>>>>> c954b23f86b629c569223e2c0c38e32a0d870d22
"RTlexdec", "RTnaming", "Familiarity", "Word", "AgeSubject", "WordC
tenFrequency", "WrittenSpokenFrequencyRatio", "FamilySize", "Deri
y", "InflectionalEntropy", "NumberSimplexSynsets", "NumberComplex
thInLetters", "Ncount", "MeanBigramFrequency", "FrequencyInitialD
elV", "ConspelN", "ConphonV", "ConphonN", "ConfriendsV", "Confriend
ConffN", "ConfbV", "ConfbN", "NounFrequency", "VerbFrequency", "CV"
Frication", "Voice", "FrequencyInitialDiphoneword", "FrequencyIni
lable", "CorrectLexdec"
```

You must **manually edit this file** and tidy it up.  
(== resolve conflict)

# A GitHub race: our favorite animals

---

Activity  
10 minutes



1. Everyone was already added to the repo as a collaborator.
2. Clone the repo to your laptop.
3. Edit "animals.md", add your line.
4. Do your usual local git routine: adding, committing.
5. Try pushing. It is likely you have a conflict (someone else pushed in the meantime) and Git tells you to pull first.
6. Pull to receive the new updates.
7. Open "animals.md". Resolve conflict.
8. Go back to step 4. Hope you were quick enough this time!

# Wrapping up

---

- ▶ To-do #10: Visit your classmates
- ▶ 2<sup>nd</sup> progress report → pushed back to next Thursday (3/24)
- ▶ "Careers in Language Data" symposium on 4/1:
  - ◆ <https://www.linguistics.pitt.edu/event/careers-language-data-symposium>

The poster features a dark green background with white and yellow text. At the top left, the title 'Careers in Language Data' is written in large white font, with the subtitle 'how to prepare our language students for the data-focused job future' in yellow. A red banner on the right side contains the text 'Featuring language specialists working at Google, Duolingo, Amazon, and more!' with logos for Google, Duolingo, and Amazon. Below the banner, the event details are listed: 'April 1 (Fri) 1-5pm' and 'University Club Ballroom A'. A quote in a speech bubble asks 'How can we language majors build data skills for such jobs?'. Another quote in a speech bubble asks 'What roles do language specialists play at tech companies?'. A curriculum panel will discuss the topic. The event is free with refreshments and will be streamed live. Pre-registration is required, with a link to <https://tinyurl.com/careers-language-data> and a QR code.

**Careers in Language Data:**  
how to prepare our language students for the data-focused job future

Year of Data and Society

Featuring language specialists working at Google, Duolingo, Amazon, and more!

An industry panel will answer:  
"What roles do language specialists play at tech companies?"

A curriculum panel will discuss:  
"How can we language majors build data skills for such jobs?"

**April 1 (Fri) 1-5pm**  
University Club Ballroom A  
Free event with refreshments!  
(Event will also be streamed live)

Pre-registration required:  
<https://tinyurl.com/careers-language-data>