«Exercise Human-Centered Data Science»
Session #2

November 10th, 2020

Agenda

1. First assignment
   - Setting up the environment, GitHub & git workflow - *me*
   - Results of first assignment (final plot) - *small groups*
2. Presentation by students - *plenum* // **BREAK**
3. Podcast discussion - *small groups*
4. Next assignment - *me*
5. (Organisation) - *me/all*
6. Outlook next week

1. First Assignment

- Access to github?
- Setting up the environment (linux/windows/mac, conda/anaconda, poetry, …)

  → `git clone git@github.com:FUB-HCC/hcds-winter-2020.git`

  README.md → poetry setup
- GitHub wiki locally

  → `git clone git@github.com:FUB-HCC/hcds-winter-2020.wiki.git`
- GitHub workflow (no branches) - this is not best practice!

```
git pull                 # get all recent changes from GitHub
...                      # work on your assignment
git add FILENAME         # add all files you want to push to github (staging)
git commit -m “MESSAGE”  # create a commit, some version of your work
git push                 # push your work to GitHub
```

https://sdqweb.ipd.kit.edu/wiki/Git_pull_--rebase_vs._--merge,
accessed: 2020-11-10 2:00 PM

1. First Assignment

- Access to github?
- Setting up the environment (linux/windows/mac, conda/anaconda, poetry, …)

  → `git clone git@github.com:FUB-HCC/hcds-winter-2020.git`

  README.md → poetry setup
- GitHub wiki locally

  → `git clone git@github.com:FUB-HCC/hcds-winter-2020.wiki.git`
- GitHub workflow (no branches) - this is not best practice!

```
...                         # work on your assignment
git add FILENAME            # add all files you want to push to github (staging)
git commit -m "MESSAGE"     # create a commit, some version of your work
git pull --rebase           # get all recent changes from GitHub and put your work on top
git push                    # push your work to GitHub
```

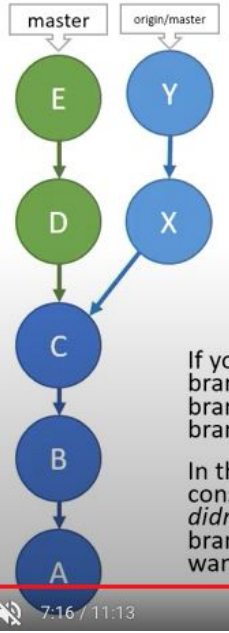
Please no force pushing!

https://sdqweb.ipd.kit.edu/wiki/Git_pull_--rebase_vs._--merge,
https://www.slideshare.net/lemiorhan/git-antipatterns-how-to-mess-up-with-git-and-love-it-again-devoxxpl-2017,
accessed: 2020-11-10 2:00 PM

Catching up a collaborative branch with `--rebase`

A similar scenario can occur when collaborating on a branch

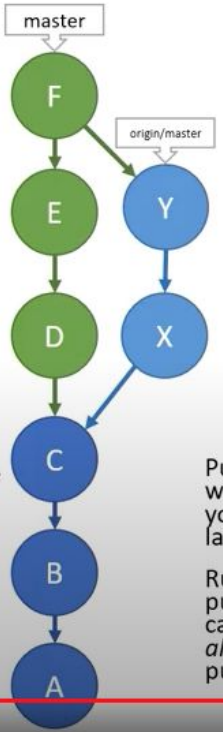
Suppose you and a collaborator both make commits to *master*, but the collaborator pushes first:



If you attempt to pull from a remote branch that is *ahead* of your local branch, `git` defaults to merging the branches with a merge commit.

In this scenario a merge commit is considered accidental noise. You *didn't* really mean to merge one branch with itself here, you just wanted to synchronize history.

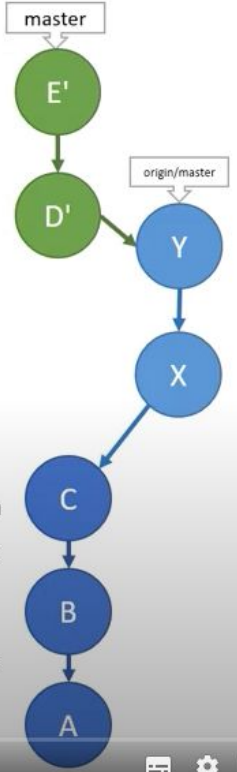
`git pull origin master`



Pulling from the same remote branch with the `--rebase` flag will cause your commits to be rebased onto the last commit of the remote branch.

Rule of Thumb: If you attempt to push and receive an error that you cannot because the remote branch is *ahead* of your branch, you should pull with the `--rebase` flag.

`git pull --rebase origin master`



7:16 / 11:13

https://www.youtube.com/watch?v=-H2U3kJ_urw&ab_channel=KrisJordan

1. First Assignment - Results

- **Small Groups** - Wikipedia Plot
  - What wikipedia language versions?
  - What article?
  - Looking at the plot - any ideas for improvement?
  - Any other experiences/challenges you had?

You can also use screen sharing to look at your notebooks!

Use video!

Be aware of the time!
Take notes!

Please put your figure (png/jpg/…) or screenshot here:

→ https://projects.invisionapp.com/freehand/document/pO9uUJhFz

1. First Assignment - Results

- One person presents their assignment please.
- Someone used *altair*?

3. Podcast

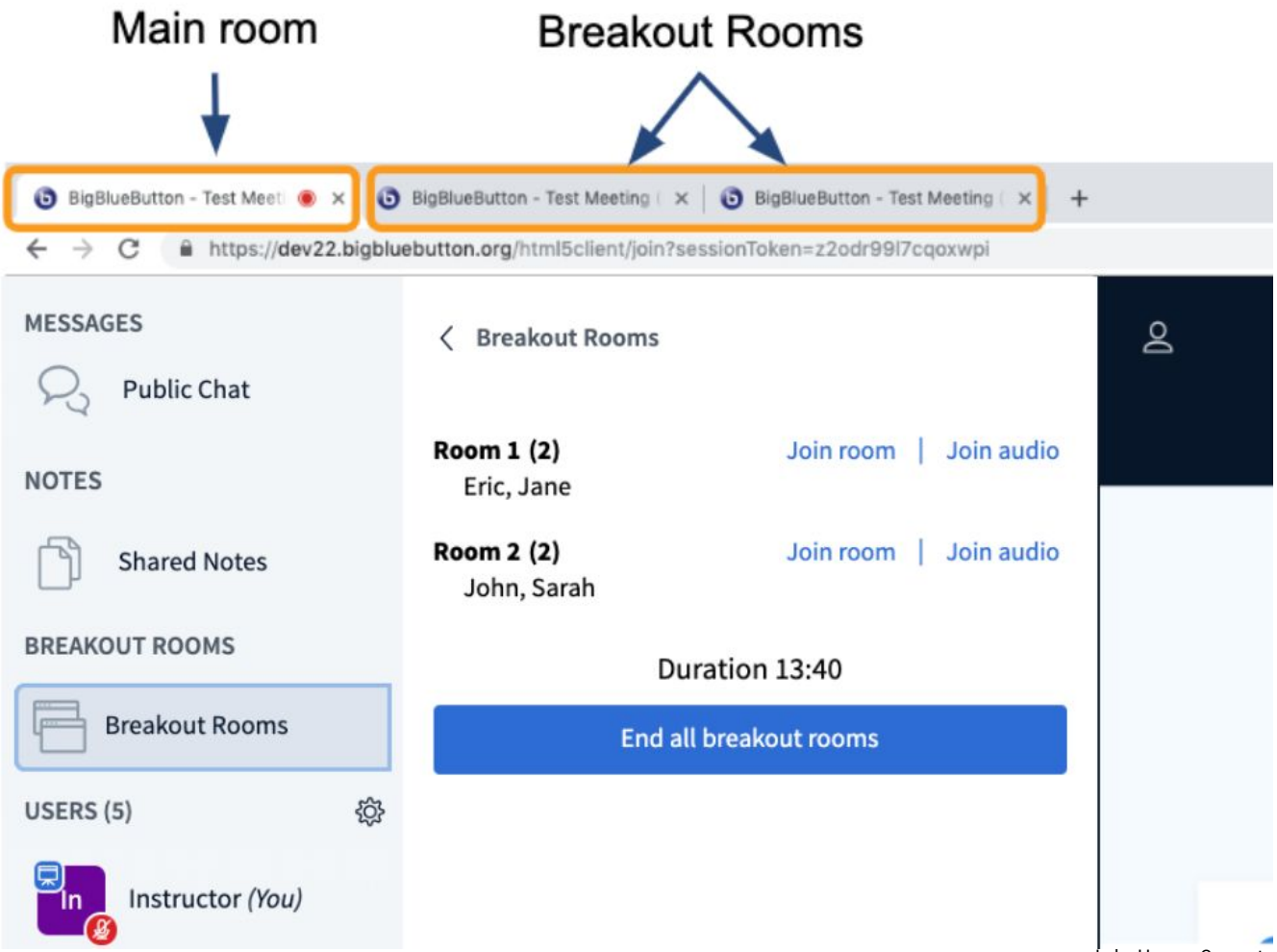
1. What are the main topics?
2. What are **your** main takeaways?
3. Choose **one** question you want to discuss in your group.

Group 1: https://projects.invisionapp.com/freehand/document/WwYCn4qR6

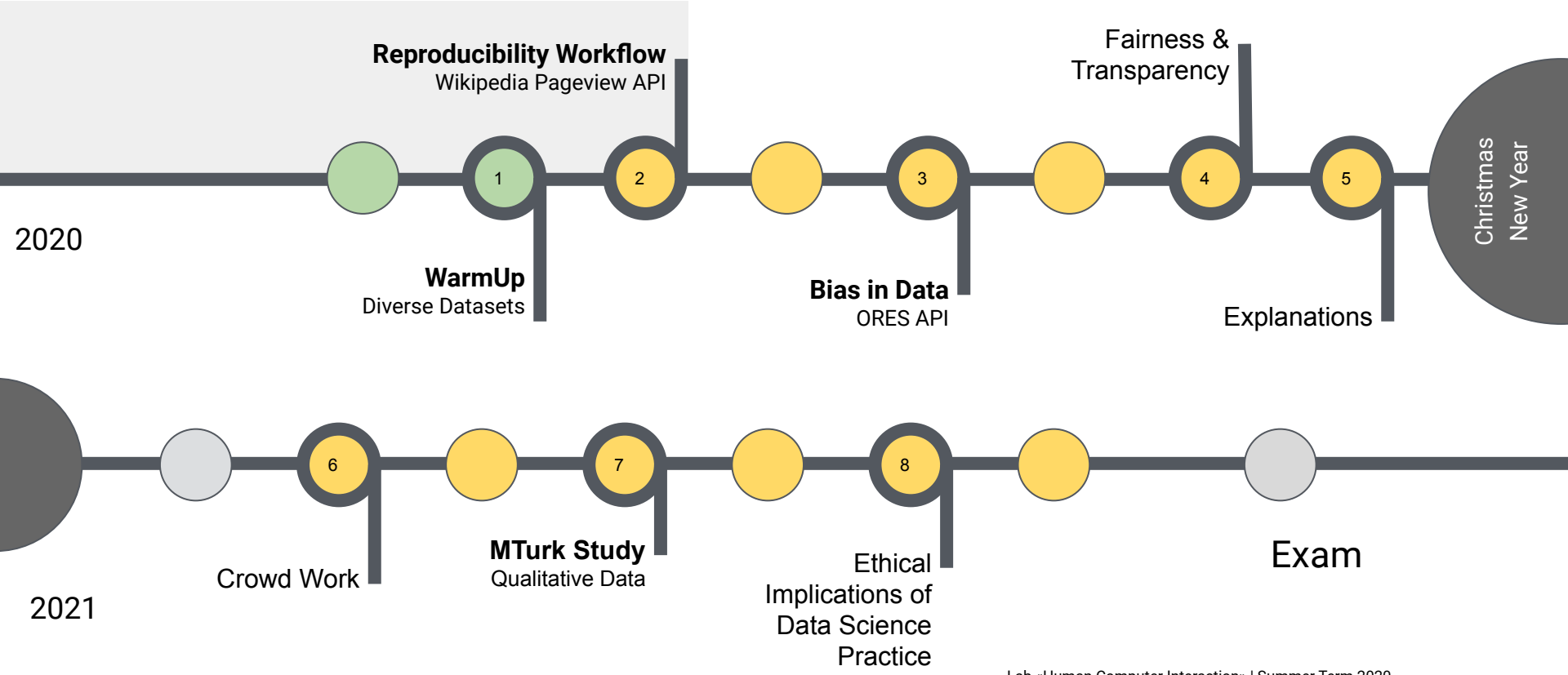
Group 2: https://projects.invisionapp.com/freehand/document/Bpbl7AA3o

Group 3: https://projects.invisionapp.com/freehand/document/5dK2r4uth

Group 4: https://projects.invisionapp.com/freehand/document/30AUJbxyB


Main room
Breakout Rooms
BigBlueButton - Test Meet
BigBlueButton - Test Meeting
BigBlueButton - Test Meeting
https://dev22.bigbluebutton.org/html5client/join?sessionToken=z2odr99l7cqoxwpi
MESSAGES
Public Chat
NOTES
Shared Notes
BREAKOUT ROOMS
Breakout Rooms
USERS (5)
Instructor (You)
Breakout Rooms
Room 1 (2)
Eric, Jane
Join room | Join audio
Room 2 (2)
John, Sarah
Join room | Join audio
Duration 13:40
End all breakout rooms

4. Next Assignment



4. Next Assignment

Wikipedia Traffic Data

The goal of this assignment is to 🅰️ construct, analyze, and publish a data set of the monthly traffic on Wikipedia and 🅱️ to learn and internalize the **reproducibility workflow** we discussed in the lecture.

1. **The Legacy Pagecounts API** provides access to desktop and mobile traffic data from December 2007 through July 2016.
2. **The Pageviews API** provides access to desktop, mobile web, and mobile app traffic data from July 2015 through last month.

→ https://wikimedia.org/api/rest_v1/

5. Outlook

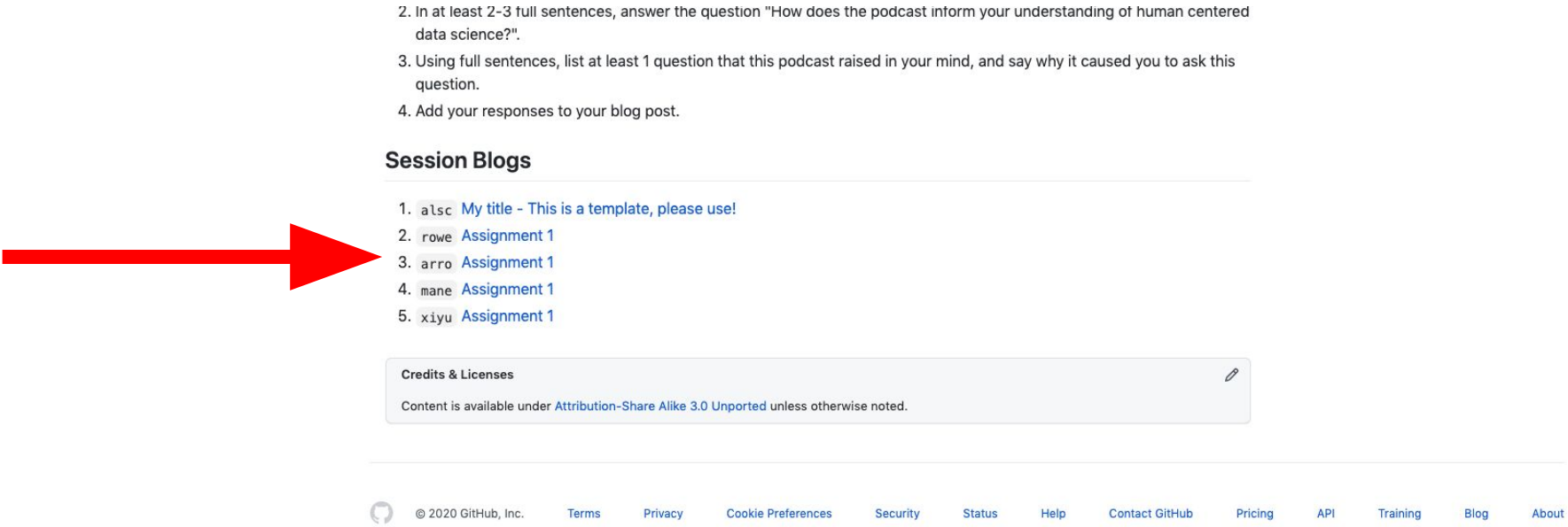
Discuss Reading!

Students present assignment

Introduction to ORES API

2. Organisation

- Please link your blog here:
  https://github.com/FUB-HCC/hcds-winter-2020/wiki/01_exercise#session-blogs



2. In at least 2-3 full sentences, answer the question "How does the podcast inform your understanding of human centered data science?".
3. Using full sentences, list at least 1 question that this podcast raised in your mind, and say why it caused you to ask this question.
4. Add your responses to your blog post.

Session Blogs

1. `alsc` My title - This is a template, please use!
2. `rowe` Assignment 1
3. `arro` Assignment 1
4. `mane` Assignment 1
5. `xiyu` Assignment 1

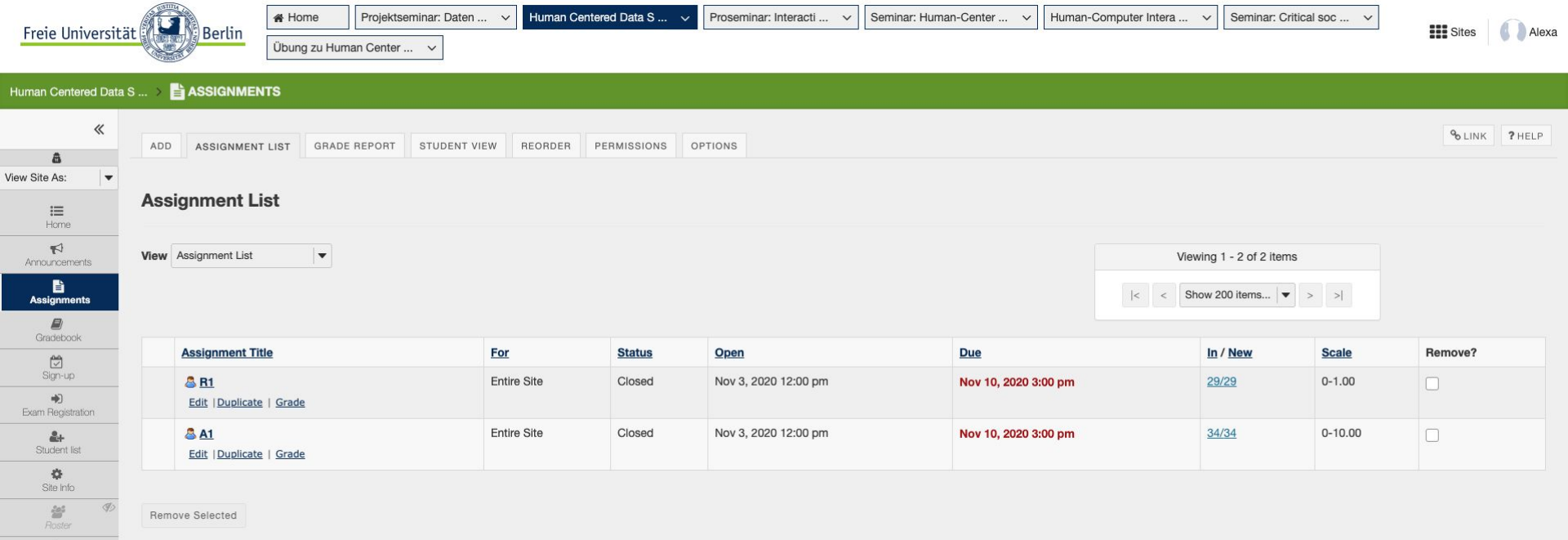
Credits & Licenses

Content is available under Attribution-Share Alike 3.0 Unported unless otherwise noted.

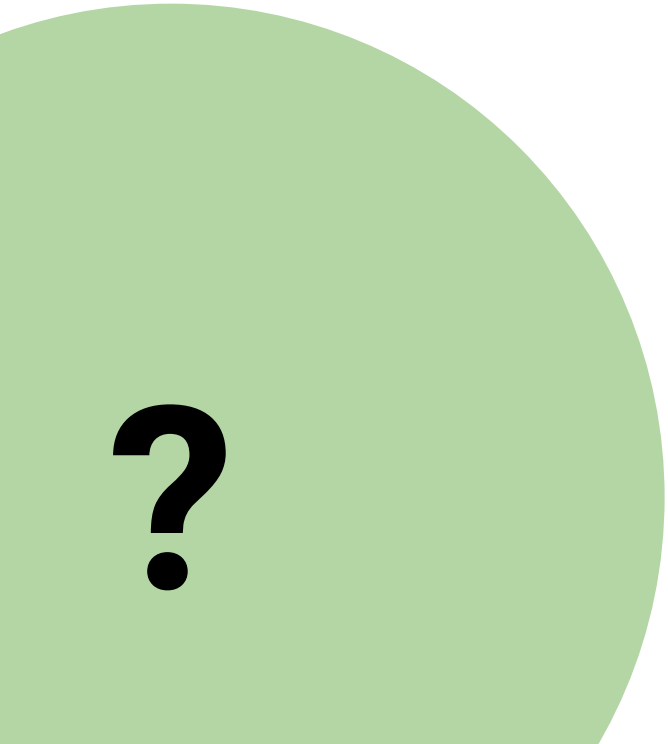
© 2020 GitHub, Inc. Terms Privacy Cookie Preferences Security Status Help Contact GitHub Pricing API Training Blog About

2. Organisation

- Overview of assignments and points (always 10 points per An, 1 for Rn)
  https://mycampus.imp.fu-berlin.de/portal/site/df87a97d-d36f-4c7b-9a03-82a1016c8045/tool/7a40f6c9-7b55-46c0-b69a-1ba1f520c064



2. Organisation

- **Discord:**
  - #hcds-lecture
  - #help
  - Anything else, we need?

Questions