- ~~Convert missing gender info to third label (US - unspecified)~~
    - ~~We want to get rid of instructors/TS but keep students who don't have a specified gender~~
- Cr~~eate label to extrapolate type of commit from the message~~
    - ~~Create a type of 'untagged'~~
    - ~~Could reuse commit message classifier if needed~~
        - ~~*Optional* Could test commit message classifier (using the student's [type])~~
    - ~~**1st1st level**: Get bracket messages (or unspecified)~~
    - ~~**1st level: Tagged brackets or not**~~
    - ~~**2nd level: Tag unspecified (and specified?) messages**~~
        - ~~**Styling/Implementation/Testing/Documentation**~~
        - ~~**or use tags from 1st level**~~
        - ~~**How many are left unspecified still?**~~
    - ***Optional* 3rd level: Do bracket tags match up with unique tags**
        - ([Implementation] Finished Testing - ***this would be incorrect match***
- ~~Filter by project type (get commits per project for both sections and analyze)~~
    - ~~I filtered by project/lab/team project~~
- ~~Analyze makeup of teams (all female/male teams, or mixed)~~
- Mixed effect regression (run analysis on all projects, run analysis on specific projects)
    - Analyze how females work within the team projects (type of work done within csc216-T)

        - especially if they are on a team within a male
- Add to Paper: Cite prior paper to highlight that the bulk of commits is around milestones or deadlines after confirming if that is a millstone around that time
- Machine learning models
    - Model1: Using tags, commit time, length of commit messages (added and deleted) as features can we predict the gender
    - Model2 : Using gender, commit time, length of commit messages (added and deleted) as features can we predict the tag
- Statistics of tags
    - ~~How many commit messages weren't tagged? 2~~
    - **X # of females tagged messages as implementation, testing, etc**