Computational Methods for Linguists

Ling 471

Olga Zamaraeva (~~Cat lady~~ Instructor)
Yuanhe Tian (TA)
Murka (style assistant)
03/30/21

Hi there!

- Let's say hi to each other!
  - This is my cat Murka
    - Murka is a common Russian cat name
    - It means "Someone who purrs"
    - *Onomatopoeia*
  - Olga
    - PhD in (computational) linguistics, UW (2021)
  - Yuanhe
    - PhD student, computational linguistics, UW
- Please raise your hand and introduce yourself!
  - Or say something in the chat

Today:
Introduction and class structure

Class goals

- Learn about data science

  - Especially as applied to language data

  - Social and technical aspects

- Learn basic programming/technical skills

- Learn basic probability and stats

- Improve presentation and communication skills

  - Communication mostly via the Discussion board...

No prerequisites!

- We will learn things from scratch

- This class doesn't count as a programming intro requirement for e.g. CS major

  - Just a Quantitative/Symbolic Reasoning credit

- This class doesn't count towards the computational linguistics Master's
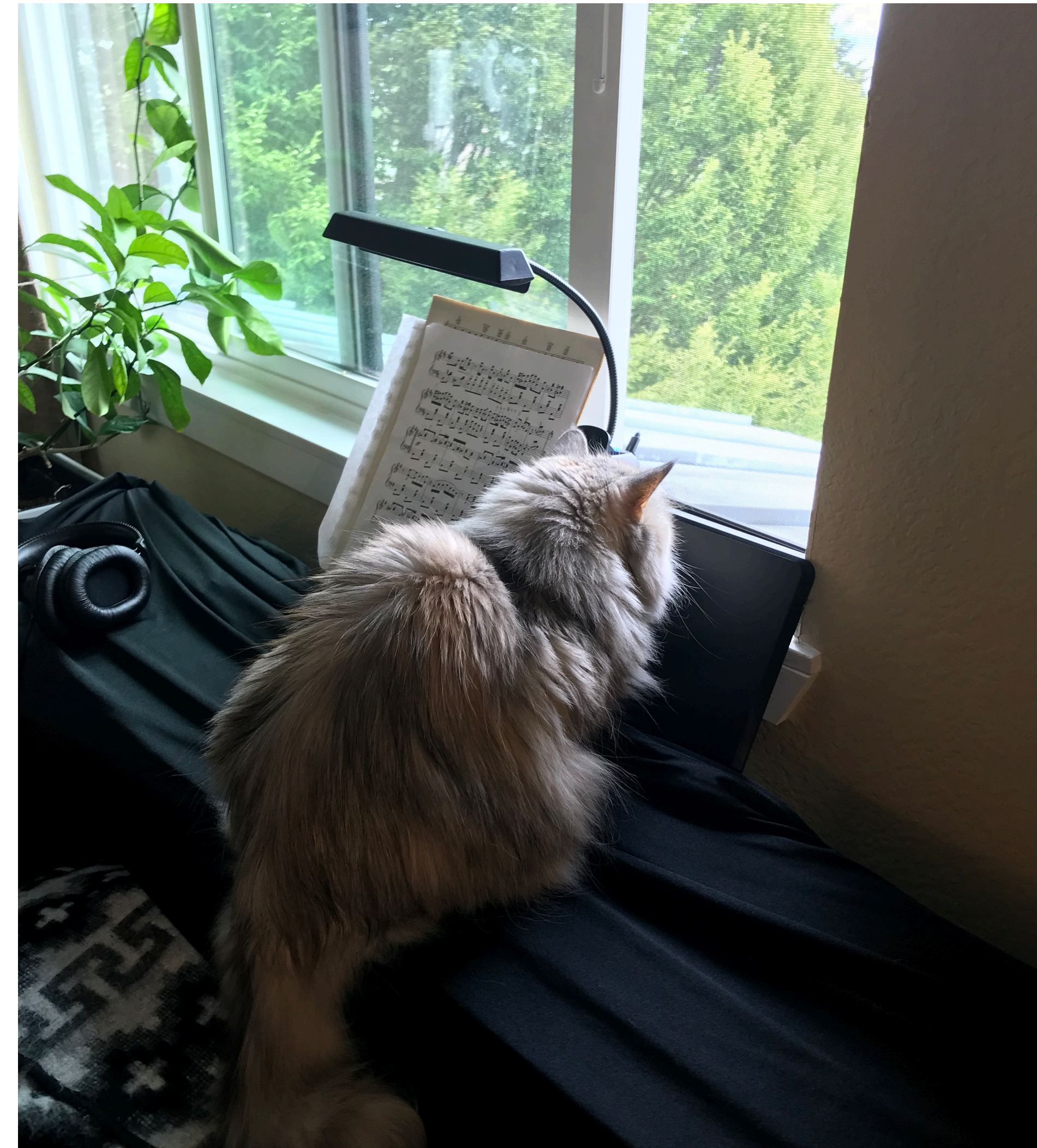
Class policies

- Online education is not ideal

  - But let's do what we can!

- Consider having your video on

  - Use Zoom backgrounds if you like (or not!)

- Stay muted when not talking but feel free to unmute and interrupt at any moment

- Use headphones

  - Otherwise, unpleasant echo happens

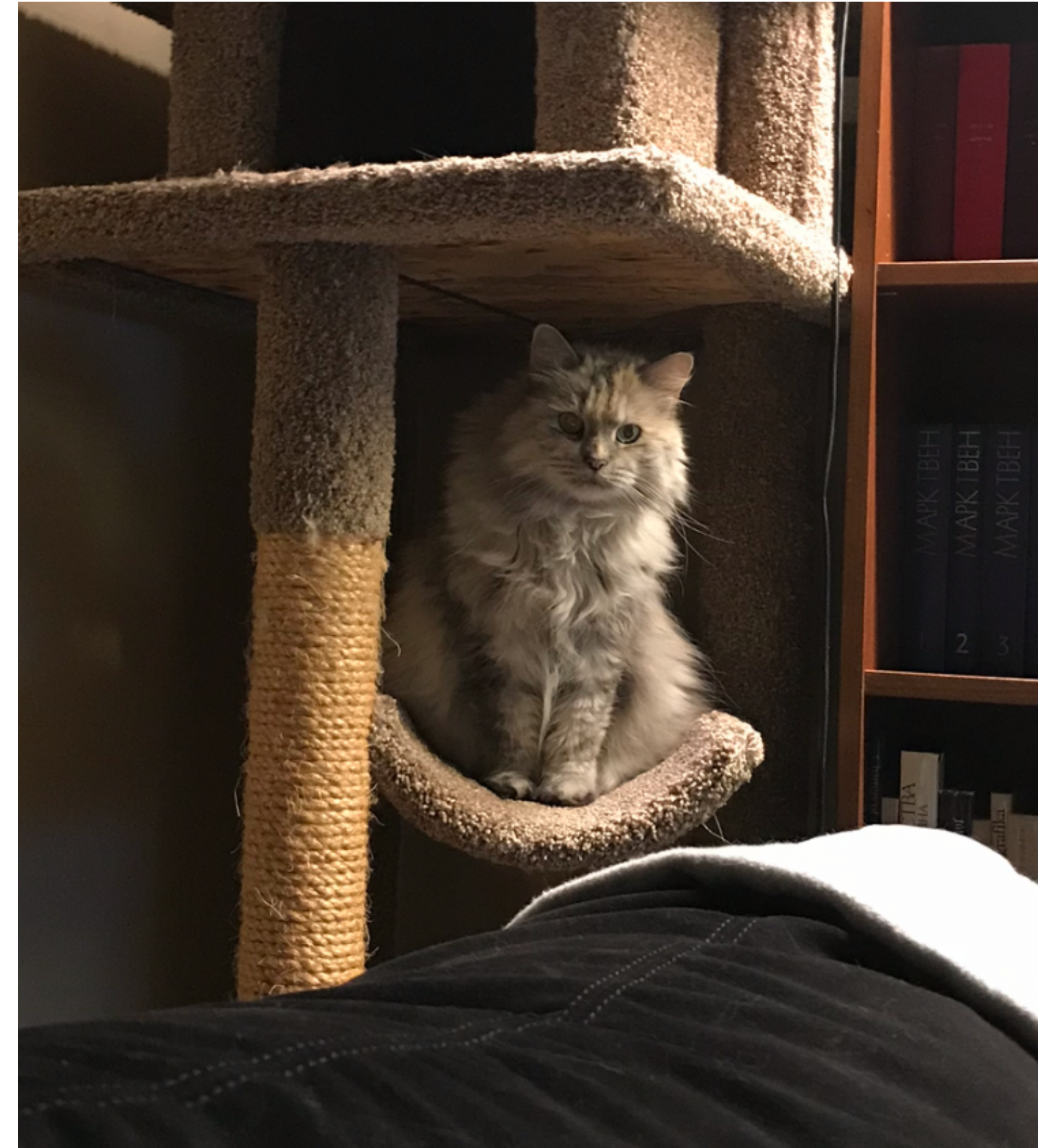- We will take a 10-min break at 11:20

Class info

- Public website

  - https://olzama.github.io/Ling471/

  - Canvas doubles the syllabus

  - Schedule on public website!

  - Assignments etc., on public website!

- Submit homework and get grades on Canvas

- Discussion Board and Blog on Canvas

Class structure

- Lecture TTh 10:30

- No section

- Extended office hours
  - Every weekday! (See Canvas)
  - No exams

- 5 assignments and 1 presentation (95% of grade)

- Blog (5%)

- Participation: up tp 2% positive adjustment

Me, Myself, and Office Hours
Activity, 10 min

- In small groups, discuss:

  - Your worst experience with office hours

    - You don't need to name people :)

  - Your best experience with office hours

  - What helps you go to office hours?

  - What is in the way of you going?

  - Formulate for yourself and share with the group:

    - How will you be using OH in this class?

  - Assign one group member to **summarize** what you talked about as a group

    - Share with the class!

Assignments
80% of the grade

- Organized around the IMDB review dataset

- Target different skills

- Due about every 2 weeks, on average

- Assignment 1 (due April 13, will publish by April 1)

  - No programming or anything

  - Somewhat lengthy setup, to prepare for the next assignments

- "Assignment 0" (due this week!)

  - Claim a Blog topic! Request an account on the "patas" cluster

  - Without it, you won't be able to complete Assignment 1!

Assignments algorithm

- Start asap

- Get blocked

- Post on the Discussion board

  - See guidelines!

- Come back next day, get unblocked

- Still blocked?

  - Attend the next Office Hour!

Late Assignments Policy

- Assignments are due strictly as indicated on Canvas
- If you submit in the next 24 hours:
  - We will grade the submission at 80%
    - E.g. a perfect submission will get 80%
    - A submission which would otherwise be 80 points, will get 64 points. Etc.
- If you submit between 24—48 hours after HW is due:
  - We will grade the submission at 70%
- If you submit after 48 hours after HW is due:
  - No credit! :(
- Exceptions granted on case-by-case basis **only if prior arrangements are made** with Olga no later than the day before HW is due

Presentation
June 1—3, 15% of grade

- Each student will do a short presentation:
  - Must present a project (such as a research paper) that involves statistical analysis of language data
    - Somebody else's existing project! We will talk about where to find them. (E.g. one of the Blog readings in the last Blog section.)
  - Must relate to/reflect on social aspects
  - Otherwise, can discuss systems, programming, ML…
  - Suggest your presentation topic by May 25 on Canvas.
- The presentation will be peer reviewed for clarity and effectiveness of communication and visualization
- Submit your presentation slides (in June) after addressing feedback
- Your original presentation can be prerecorded or not

Blog assignment
5% of the grade

- Write ONE blog post

  - Choose from suggested options

    - Look under "Pages"->"View All Pages" in Canvas

  - Or suggest your own, on one of the topics

  - Submit three choices, in order of preference, as "Assignment 0" (this week)

  - We will confirm your option and add you to the Blog schedule which will be accessible to class members

- Respond to ONE post every week

Questions?