ALC Text Mining

UCLA Extension Group 4
Bella Lee, Joy Harjanto, Kathy Fu, Kristee Song

Abstract

ALC is the American Language Center associated with UCLA that caters international students towards an academically rigorous learning experience while simultaneously practicing English in their daily lives. In the datasets provided by Professor Michael Thomas, there are 2 main datasets (Program Evaluations and Elective Preferences) that we work with to analyze class preference trends and mine text. Our objective was to optimize survey responses (mainly that of open-ended question responses), and find patterns in the students' elective preferences. This was achieved through visualization, data cleaning, and text mining, which can be defined as determining the frequency of open ended questions answered per survey. Thus this process involved three steps: cleaning, merging, and analysis, with data cleaning being the backbone of the study.

Research Questions and Data

1. Do we have enough data to do text-mining in our survey responses? Among the responses we have, what patterns can we find?
    - **Program Evaluations**
        - Contains multiple choice and open-ended questions asking about student experiences with the teachers, classes, etc.
    - **Elective Preferences**
        - Separated by intermediate students' preferences and advanced students' preferences
        - Intermediate students take one elective and advanced students take two electives
2. How can we improve the survey to increase response rate?

Variables

- The Program Evaluation data were structured with the variable names as the questions that the students answered. We focused primarily on the variables that contained open-ended questions which allowed students to write freely.
- The Elective Preferences data were structured with variables of 1st choice and 2nd choice, where the observations contained the students' choices for electives.

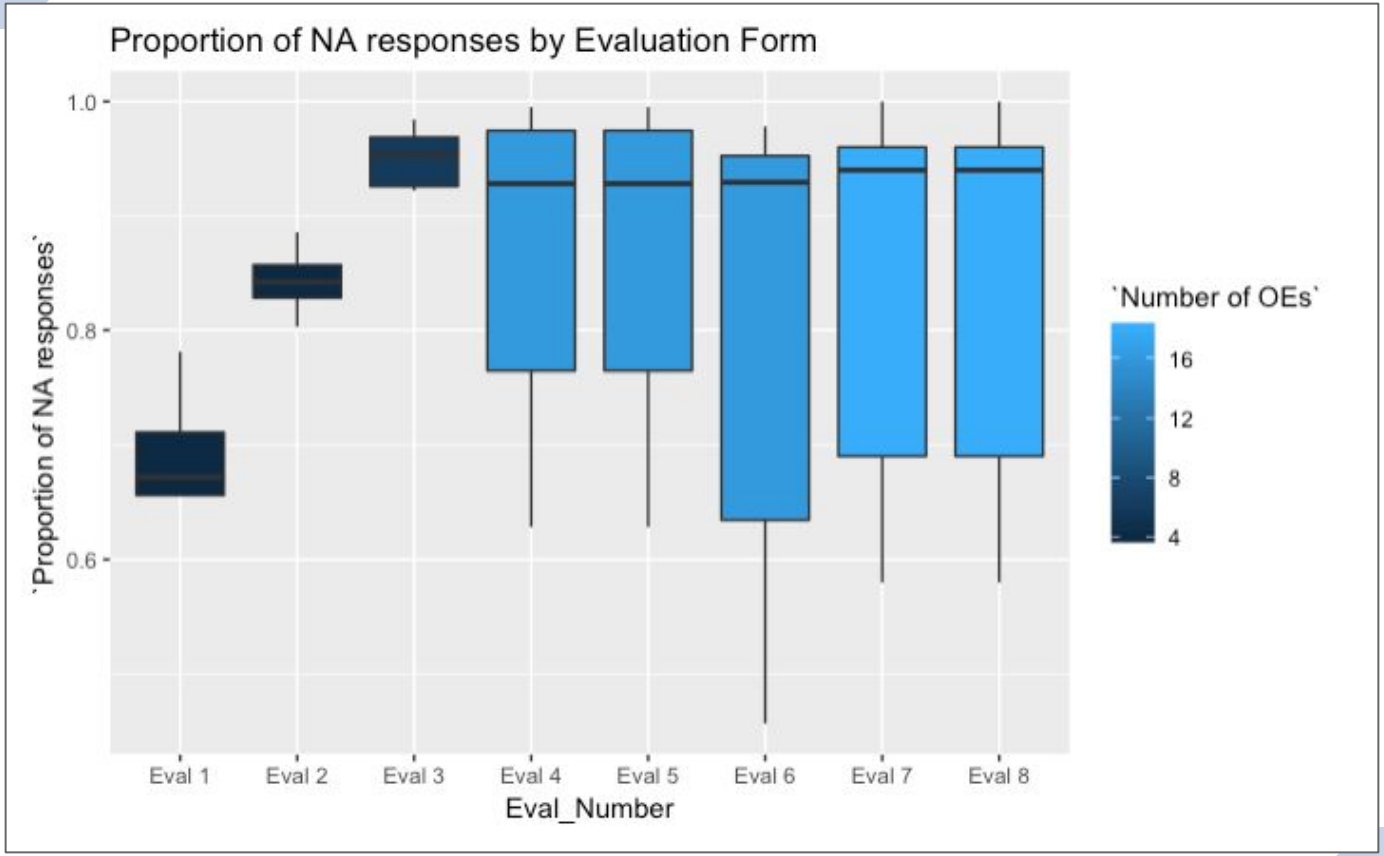
Program Evaluations

Do we have enough data to do text-mining in our Program Evaluation survey responses?

No, and this is because there are an overwhelming number of NULL responses.

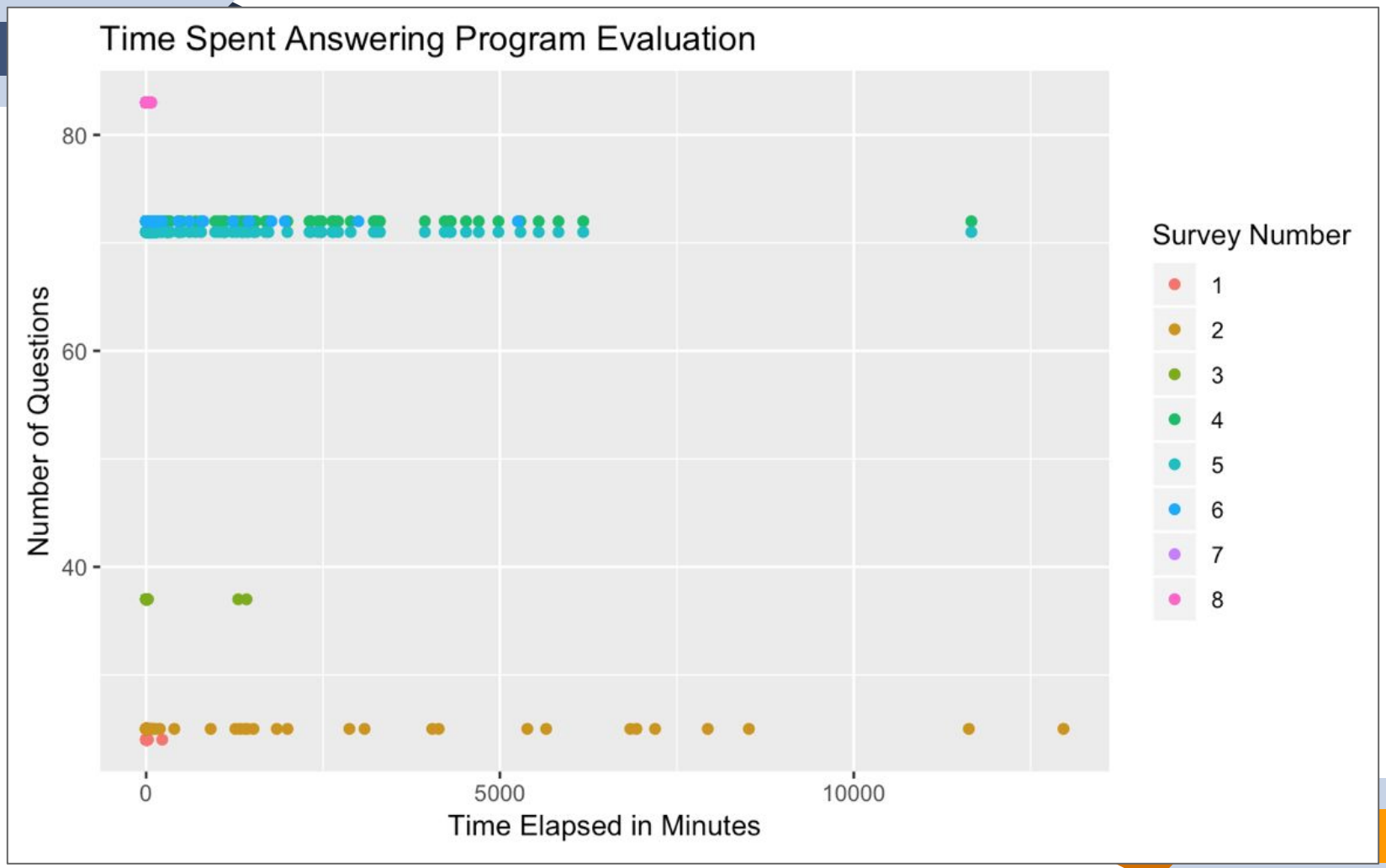
We will evaluate how these NULL responses vary across:

- Each program evaluation form
- Time
- Question number


Proportion of NA responses by Evaluation Form
`Proportion of NA responses`
1.0
0.8
0.6
Eval 1
Eval 2
Eval 3
Eval 4
Eval 5
Eval 6
Eval 7
Eval 8
Eval_Number
`Number of OEs`
16
12
8
4

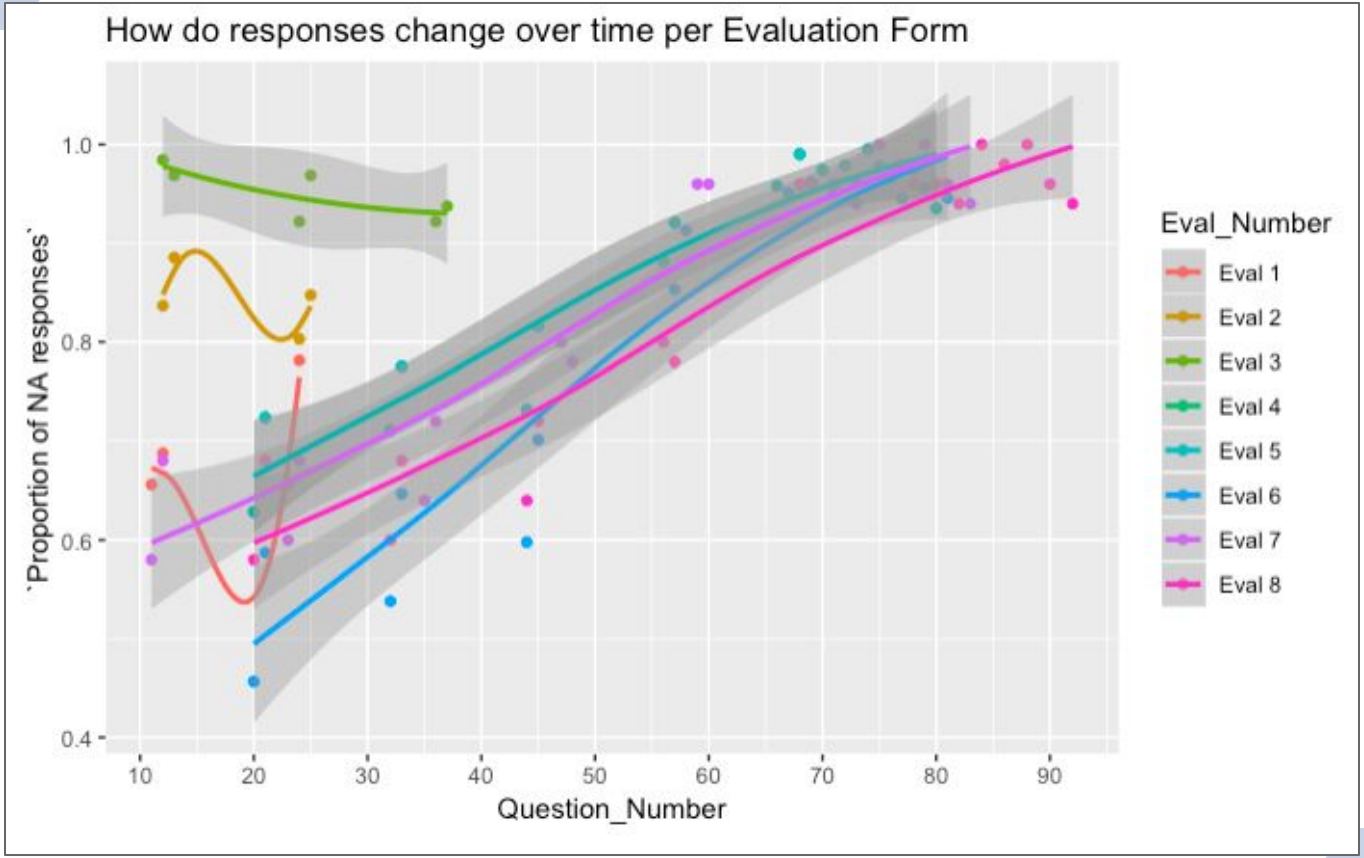
Evaluating NULL Responses per Eval Form

- We evaluated each Program Evaluation dataset to explore the trends on NULL responses depending on how many open-ended questions were on the survey.
- The fewer open ended questions that the survey had, the smaller of a proportion of NULL values we had, concluding that the more questions there are, the smaller response rate exists.
- Program Evaluation 1 had the lowest proportion of NA responses → Should model survey most closely to the format of Program Eval 1


Time Spent Answering Program Evaluation
Number of Questions
80
60
40
0
5000
10000
Time Elapsed in Minutes
Survey Number
1
2
3
4
5
6
7
8

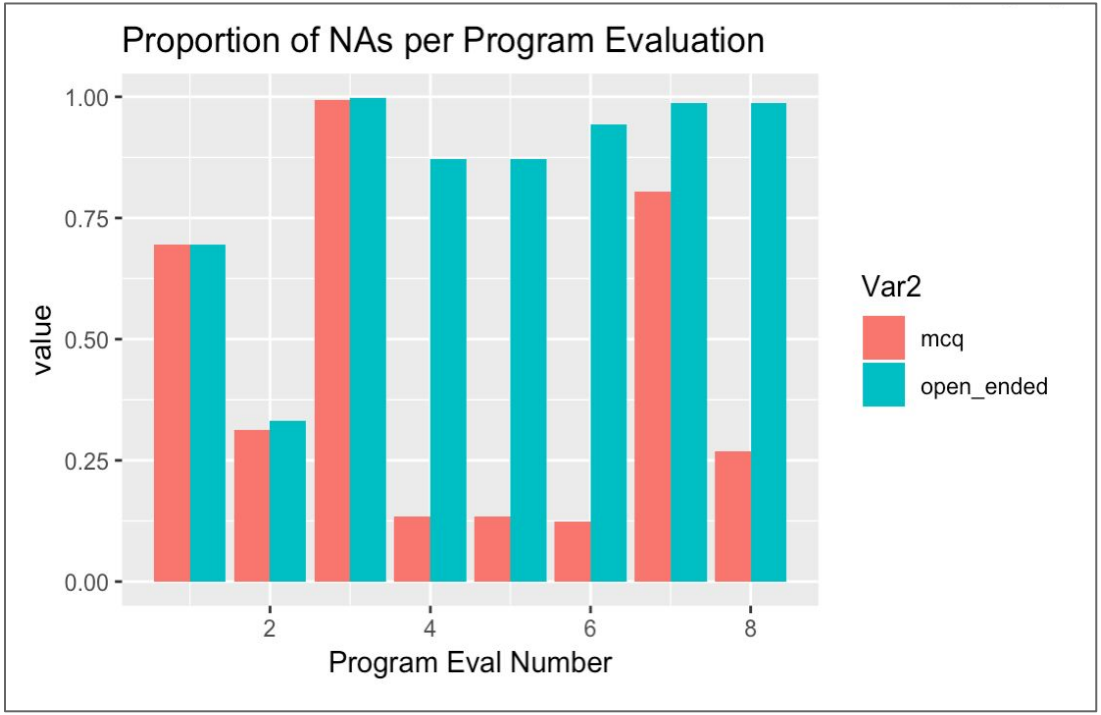
Time Spent on Program Evaluation

- For Program Evaluation 2, we found that there were a wider span of times that people spent completing the Program Evaluation.
- Many students, however, spent close to 0 or 1 minute filling out the survey. This may be the case due to the fact that some students open the link and close out of it, or students click through the survey since questions are not required to be answered.
- The newer Program Evaluations have more questions, and therefore could be a reason as to why students spend less time on them.


How do responses change over time per Evaluation Form
`Proportion of NA responses`
1.0
0.8
0.6
0.4
10
20
30
40
50
60
70
80
90
Question_Number
Eval_Number
Eval 1
Eval 2
Eval 3
Eval 4
Eval 5
Eval 6
Eval 7
Eval 8

Responses change over time during the Survey

- Based on the plot, after the Program Evaluations hit around question 60, there is a cutoff point where NULL responses are significantly high.
- With the first few Program Evaluations that contained less questions, they had better response rates compared to the more recent Program Evaluations where there were more questions, and as the survey goes on, less students respond.



MCQ vs. Open-Ended

- Based on the plot, more students skip out on open-ended questions
- There were more questions asked as time went on, which explains the dip in the number of MCQs answered

Why are there so many NULL responses?

- There is a trend of people stopping at some point into the survey and then all their responses become NULL.
- The survey collecting system is flawed. Every row in the Elective Preferences and Program Evaluations forms represents a time that any extension student merely clicked on the survey. This means that if a student clicked out, their entry is still there but there was no meaningful information recorded.

Survey suggestions

- Make the Open Ended responses REQUIRED on SurveyMonkey:
- Make survey shorter.
  - At max 60 questions
- Force students to click “Submit” at the end of the survey, which will prevent recording “responses” from students who had simply opened/closed the link.
- Each teacher has their own survey for their one class
  - From there, the teacher can send survey results to ALC directly
- And then ALC program evaluation can be in a big general survey at the end of the year
  - Like the Senior Exit Survey for UCLA

Welcome to Online Evaluation from the Evaluation of Instruction Program

Your careful and candid evaluation is important. Student course evaluations provide feedback to the instructor to help improve teaching and they provide information for faculty evaluation and promotion. Summaries of these evaluations are returned to the instructor and to the department chair. Please provide thoughtful and constructive comments regarding the instructor and course at the end of this evaluation.

1 Background Information:

1.1	Year in School:	○ Freshman	○ Sophomore	○ Junior	○ Senior	○ Graduate	○ Other					
1.2	UCLA GPA:	○ Below 2.0	○ 2.0 - 2.49	○ 2.5 - 2.99	○ 3.0 - 3.49	○ 3.5+	○ Not Established					
1.3	Expected Grade:	○ A	○ B	○ C	○ D	○ F	○ P	○ NP	○ ?			
1.4	What requirements does this course fulfill?	○ Major	○ Related Field	○ G.E.	○ None							

2 To What Extent Do You Feel That:

			1	2	3	4	5	6	7	8	9		
2.1	Instructor Concern – The instructor was concerned about student learning.	Very Low or Never	○	○	○	○	○	○	○	○	○	Very High or Always	○ Not Applicable
2.2	Organization – Class presentations were well prepared and organized.	Very Low or Never	○	○	○	○	○	○	○	○	○	Very High or Always	○ Not Applicable
2.3	Interaction – Students felt welcome in seeking help in or outside of the class.	Very Low or Never	○	○	○	○	○	○	○	○	○	Very High or Always	○ Not Applicable
2.4	Communication Skills – The instructor had good communication skills.	Very Low or Never	○	○	○	○	○	○	○	○	○	Very High or Always	○ Not Applicable
2.5	Value – You have learned something you consider valuable.	Very Low or Never	○	○	○	○	○	○	○	○	○	Very High or Always	○ Not Applicable
2.6	Overall – Your overall rating of the instructor.	Very Low or Never	○	○	○	○	○	○	○	○	○	Very High or Always	○ Not Applicable
2.7	Overall – Your overall rating of the course.	Very Low or Never	○	○	○	○	○	○	○	○	○	Very High or Always	○ Not Applicable

3 Your View of Course Characteristics:

3.1	Subject interest before course	Low	○	○	○	High	○ N/A
3.2	Subject interest after course	Low	○	○	○	High	○ N/A
3.3	Mastery of course material	Low	○	○	○	High	○ N/A
3.4	Difficulty (relative to other courses)	Low	○	○	○	High	○ N/A
3.5	Workload/pace was	Too Slow	○	○	○	Too Much	○ N/A
3.6	Texts, required readings	Poor	○	○	○	Excellent	○ N/A
3.7	Homework assignments	Poor	○	○	○	Excellent	○ N/A
3.8	Graded materials, examinations	Poor	○	○	○	Excellent	○ N/A
3.9	Lecture presentations	Poor	○	○	○	Excellent	○ N/A
3.10	Class discussions	Poor	○	○	○	Excellent	○ N/A

4 Comments:

(maximum of 5,000 characters allowed)

4.1 Please identify what you perceive to be the real strengths and weaknesses of this instructor and course. (maximum 5000 characters)

PLEASE NOTE: Each year, the Academic Senate Committee on Teaching gives awards to outstanding faculty and teaching assistants. If you wish to nominate an instructor or teaching assistant for such an award, please contact the instructor's department.

THANK YOU! Your assistance in improving teaching is appreciated.

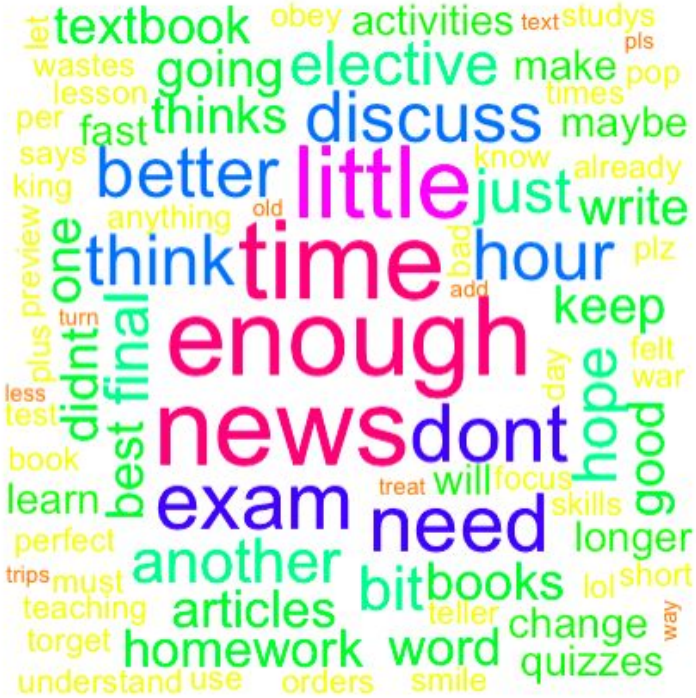
Submit

Program Eval Findings

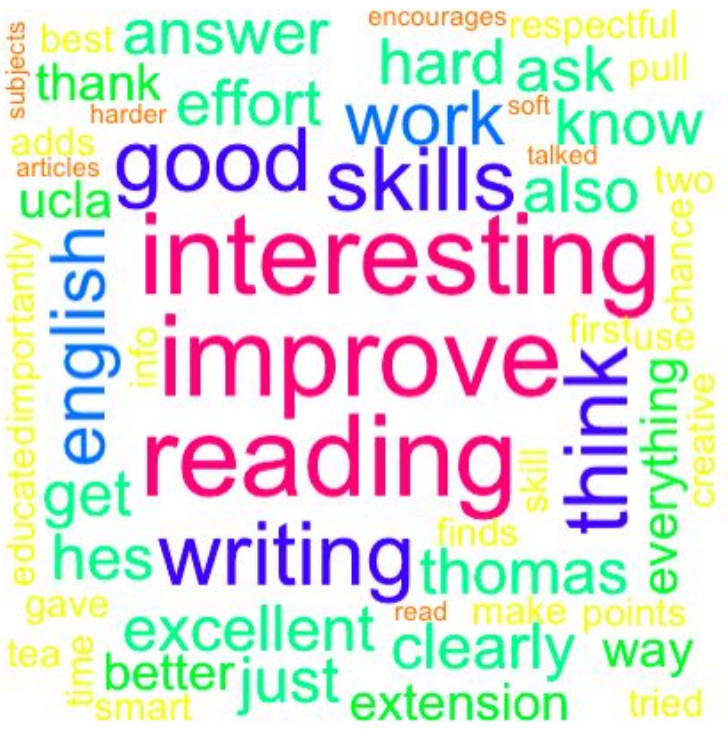
Looking at open-ended responses for teachers

Q: What do you like about Gene Kavenoki?

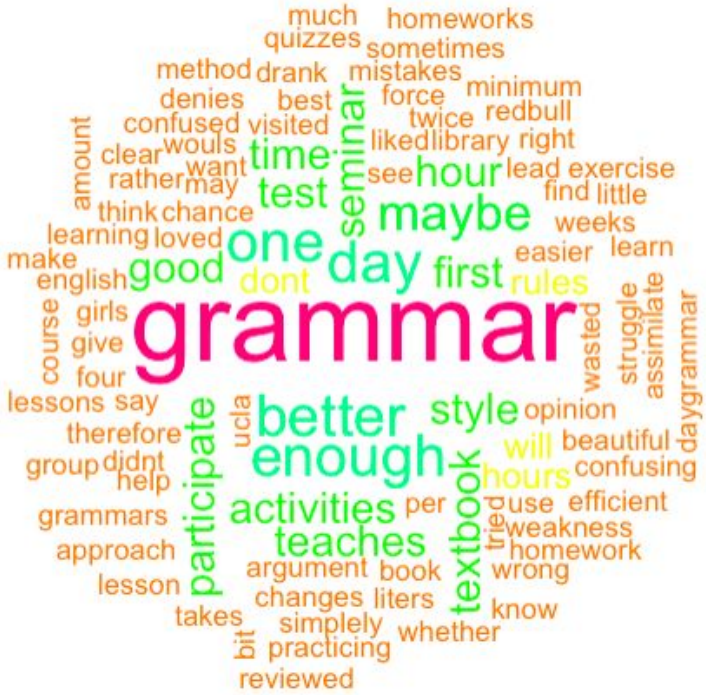
Q: What changes would make this class better?



Q: What do you like about Michael Thomas?



Q: What changes would make this class better?

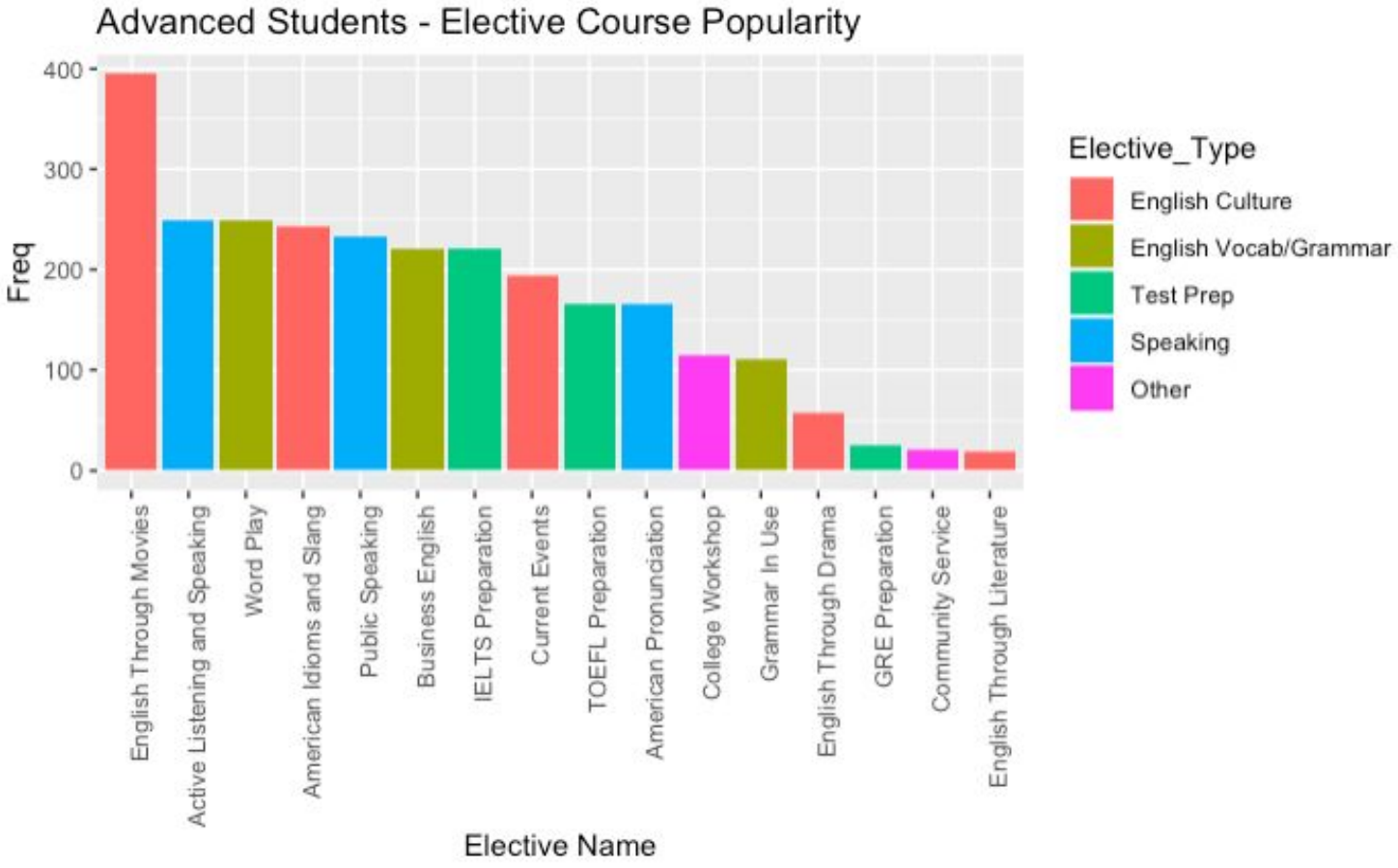


Elective Preferences

Which electives are most popular?

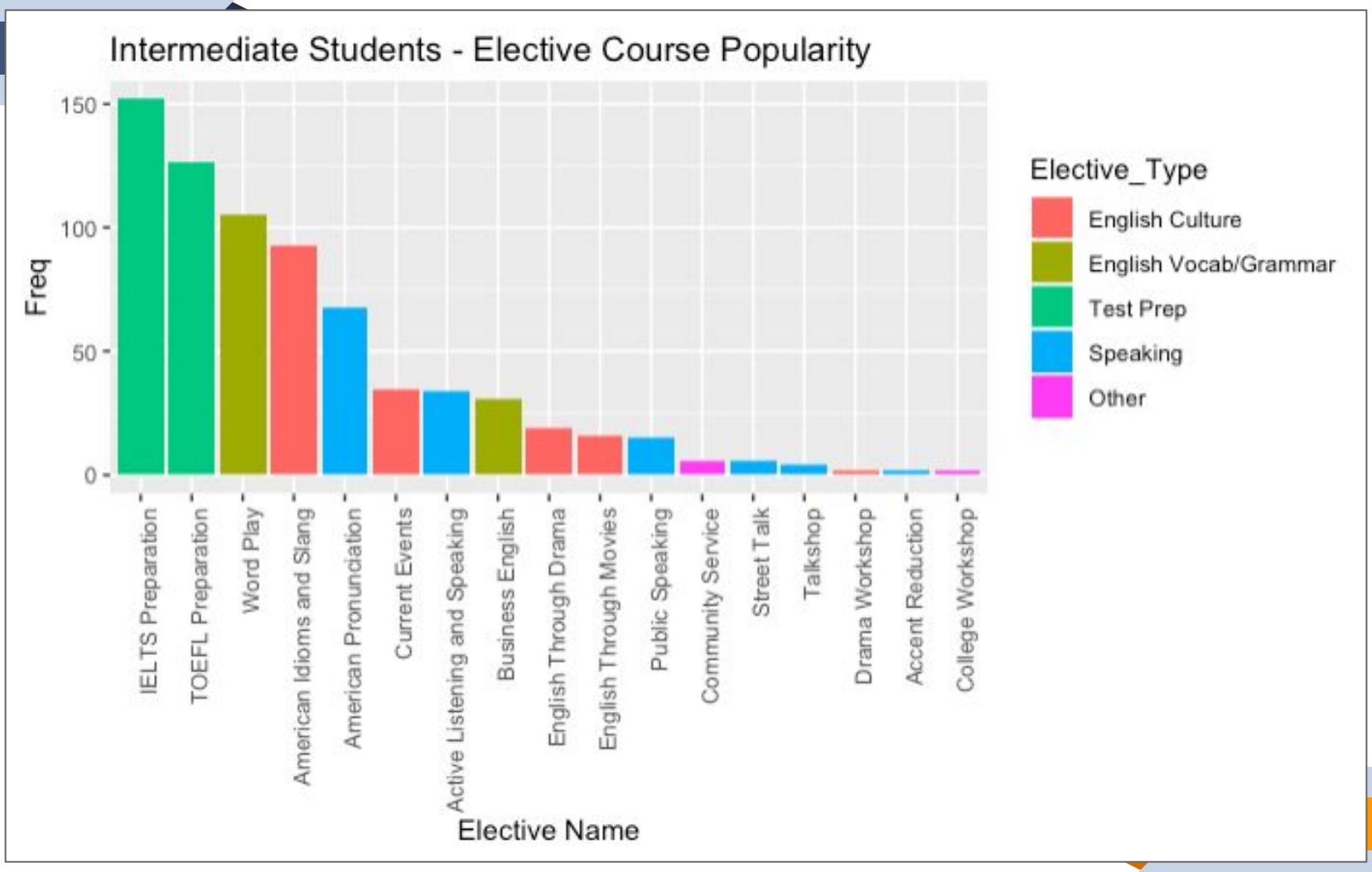
Data Reorganization

- **Categorized Elective Preferences** - Categorized the listed courses into different types of courses based on subject, which is visible through the EDA presented in this report.
  - English Culture
  - English Vocab/Grammar
  - Test Prep
  - Speaking
  - Other


Advanced Students - Elective Course Popularity
Freq
400
300
200
100
0
English Through Movies
Active Listening and Speaking
Word Play
American Idioms and Slang
Public Speaking
Business English
IELTS Preparation
Current Events
TOEFL Preparation
American Pronunciation
College Workshop
Grammar In Use
English Through Drama
GRE Preparation
Community Service
English Through Literature
Elective Name
Elective_Type
English Culture
English Vocab/Grammar
Test Prep
Speaking
Other

Popularity of Electives for Advanced Students

- English Through Movies is the most popular elective choice for advanced students.
- Overall, classes categorized under English Culture, Vocab and Grammar, and Speaking are most popular.
- Advanced students appear to want more of the “fun” electives


Intermediate Students - Elective Course Popularity
Freq
150
100
50
0
IELTS Preparation
TOEFL Preparation
Word Play
American Idioms and Slang
American Pronunciation
Current Events
Active Listening and Speaking
Business English
English Through Drama
English Through Movies
Public Speaking
Community Service
Street Talk
Talkshop
Drama Workshop
Accent Reduction
College Workshop
Elective Name
Elective_Type
English Culture
English Vocab/Grammar
Test Prep
Speaking
Other

Popularity of Electives for Intermediate Students

- Test Prep is the most popular for Intermediate students, with IELTS and TOEFL Prep being the most popular selection.
- After that comes classes classified as English Grammar and Vocab and English Culture.
- Intermediate students appear to tend towards more academic intensive courses

Conclusion

Recommendations (summarized)

- Surveys:
  - Make the surveys **shorter** and the questions **required** (at most 60 questions)
  - Follow UCLA's general model with Professor Evaluation Forms and Exit-Surveys
  - Give students **time at the end of their last class time** to fill out teacher surveys
  - **Consistency** with questions from quarter to quarter to streamline future data cleaning
- Elective Preferences:
  - Offer more IELTS and TOEFL Test Preparation courses for Intermediate students

Shortcomings

- Although we were able to do some analysis on the open-ended questions for the Program Evaluations, many of the responses were left blank, so we did not have a wide range of responses to analyze.
- As international students are the target of the ALC program, their English may not be at its best yet, so text-mining on the free-response answers provided was difficult due to frequent misspellings.
- Many of the datasets did not have the same variables, so we manually cleaned them and could only analyze them one by one instead of as a whole.