Sentiment and Mental Health Effects of the COVID-19 Pandemic on UCLA Students

Summary of Findings

Group 5

Stats 141SL – 11AM

June 5th, 2020

Our project explores the psychological effect of the COVID-19 pandemic

1 The Problem

- The extent of COVID-19's psychological impact has been difficult to measure, given the lack of structured data or controlled means of measuring its effects in the field.
- With access to UCLA student surveys, we have a unique opportunity to explore the impact the pandemic has had on student psychology and mental health.

2 Our Approach

- We leveraged sentiment analysis and textual visualization techniques to explore and characterize the emotions and ideas expressed by UCLA students.
- Our end product are a set of visualization and insights into the diverse ways UCLA students respond to COVID-19

3 Our Hypothesis

- Students' general feelings about life after this pandemic are worse than prior to the pandemic.
- STEM majors, intl. students, and those with altered living arrangements express more negativity due to the pandemic.
- Although they have access to the Zoom platform, students' interaction with their peers and instructors are often affected.

Our final data-sets are comprised by UCLA student survey responses from before and after the pandemic

Sample of Dataset

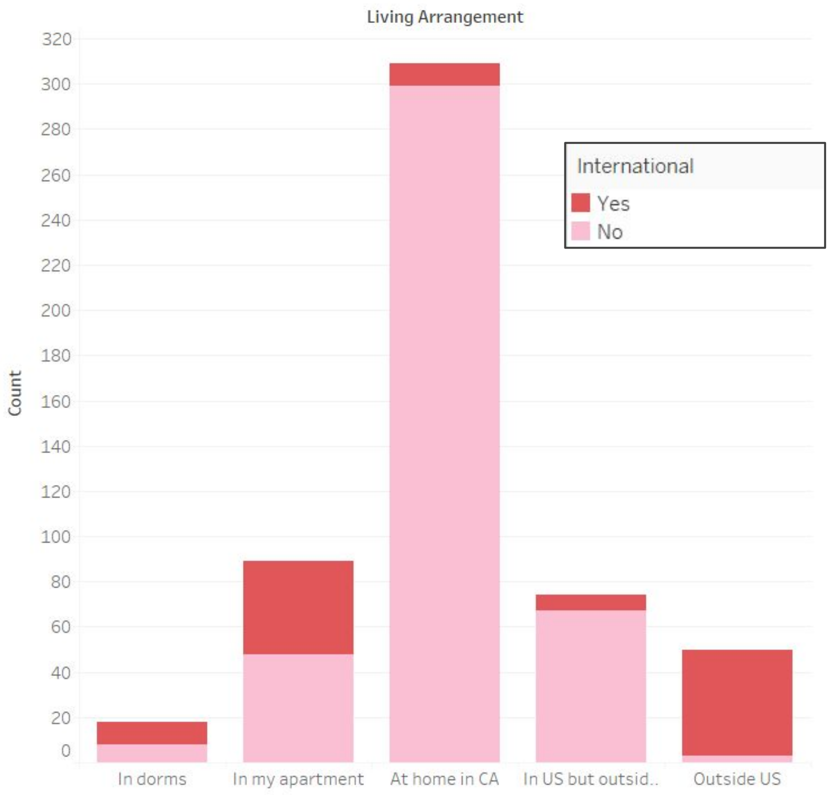
Timestamp	2020/04/19 5:19:21 PM MDT
Your major:	STEM
Are you an international student?	No
Your living arrangement during this pandemic	In my apartment
Does your living arrangement …?	No
How motivated are you…?	Less motivated than before?
How well are you able to interact with your classmates…?	2
How well are you able to interact with your instructors..?	1
Describe your sense of belonging to UCLA...	Very stressed all the time with regard to school work because of the lack of motivation. And having constant stress makes it even harder to stay motivated so it's just positive feedback loop.
What are the concerns do you have…?	How long will it last? Will I pass all my classes and graduate? How are grades going to be given at the end of the quarter?
….	…

Description

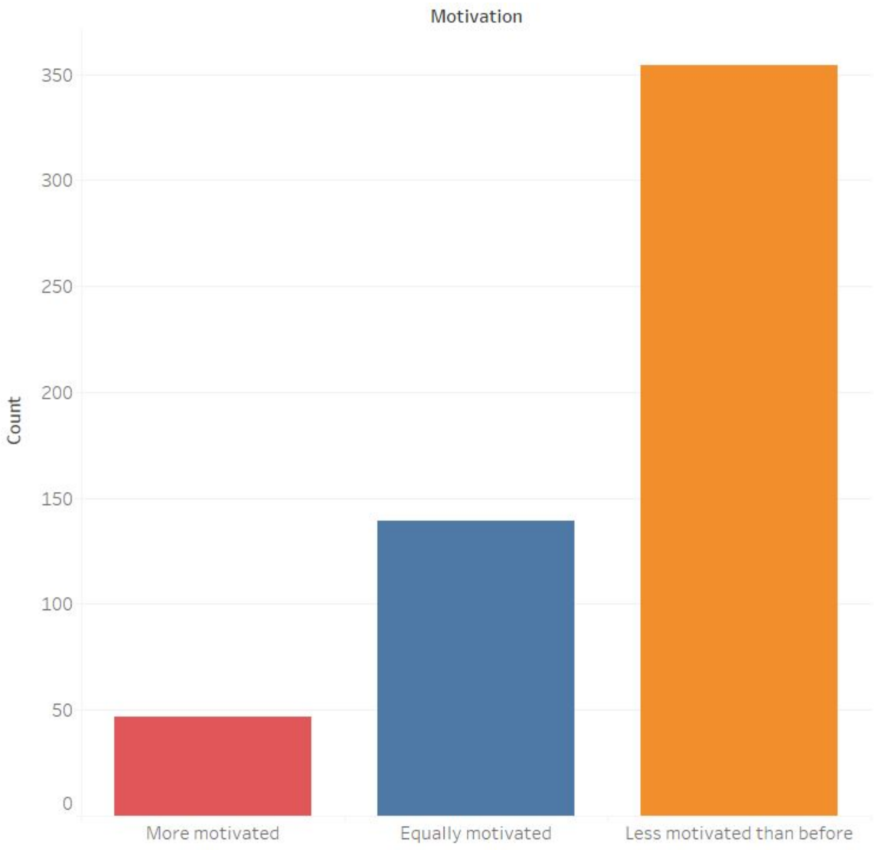
- Mainly will be looking at the data in the **post_corona_survey** file taken by various UCLA students and given to us by Professor Esfandiari (n = 545)
- Will also do some minor comparisons to two other survey data-sets given to us pre-pandemic as well (n = 277, n = 866)
- We have labeled the post_corona_survey data with sentiment values (0 – negative, 1 – positive)

Demographics & Basic Analysis of Responses

Living Arrangements & Intl. Status



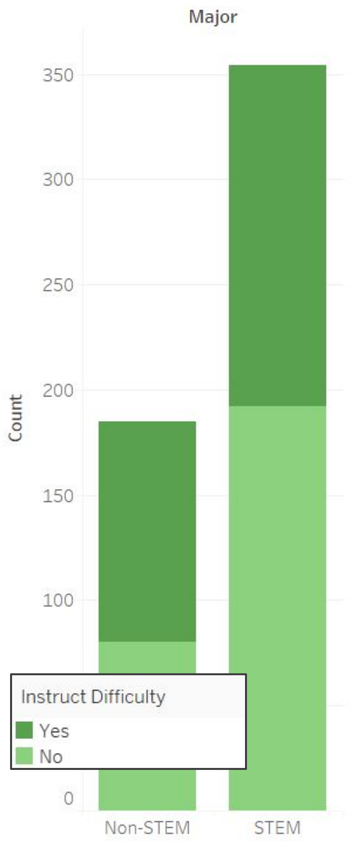
Student Motivation Compared to Pre-Pandemic



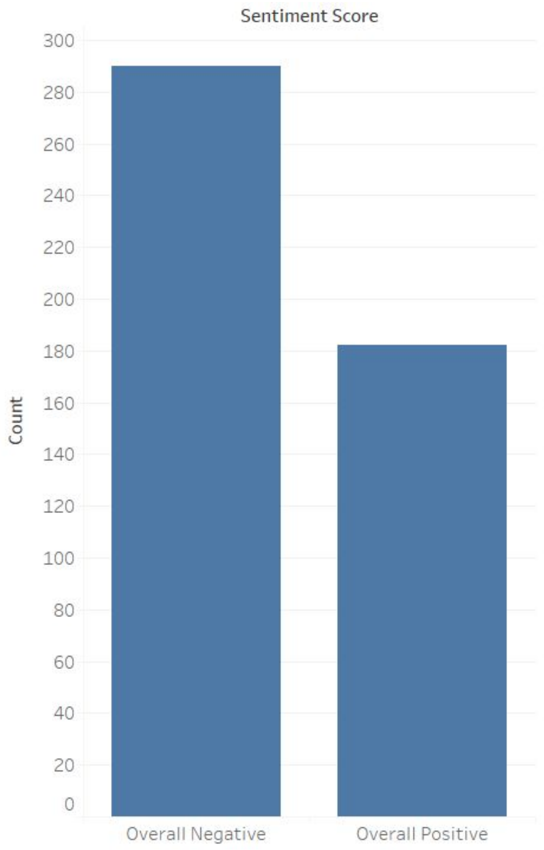
Demographics & Basic Analysis of Responses

Student & Instructor Interaction Scores

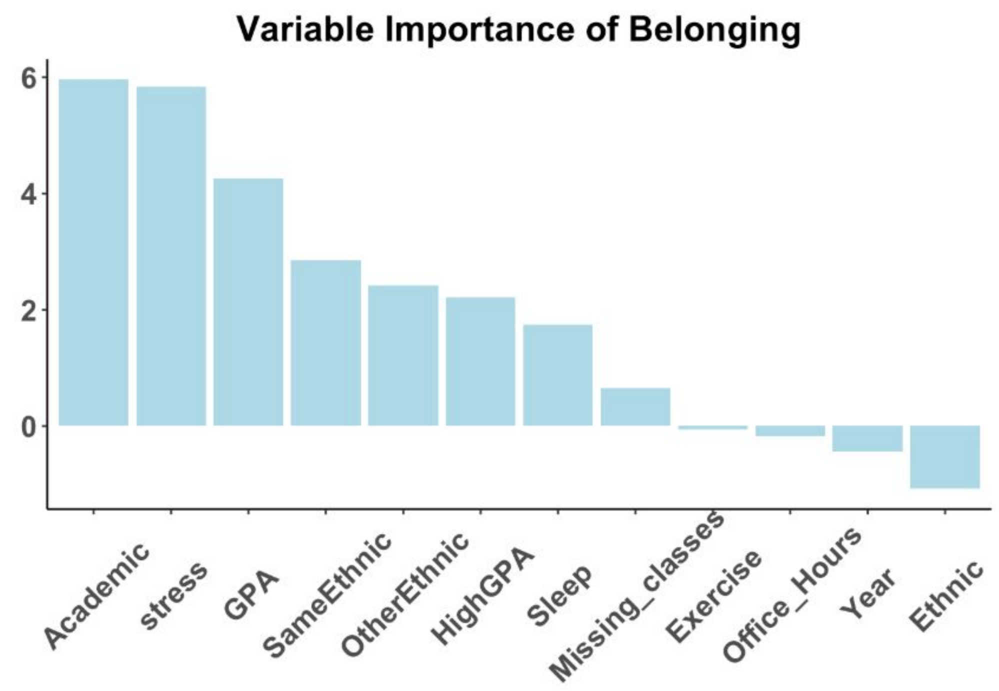
Major & Instructor Difficulty

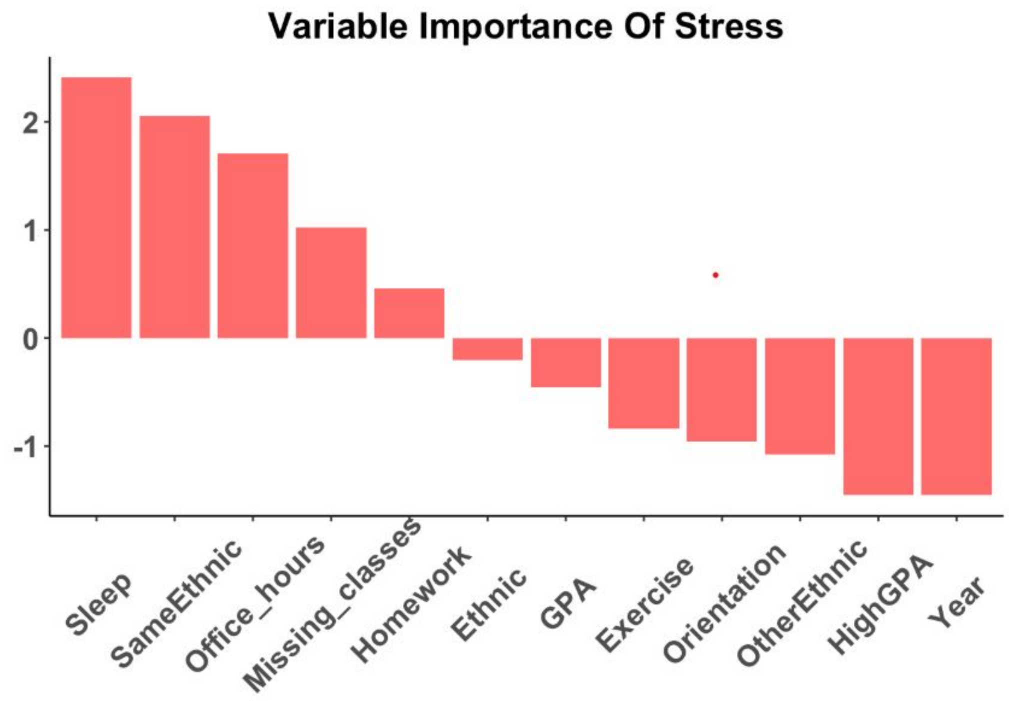


Manual Sentiment Scores



What Factors Affect Sentiment Under ‘Normal’ Circumstances?





Strong academic performance, but also high stress predict for belonging

Sleep and size of friend group predict for stress levels

To extract information from the text we compiled word counts then compared them against public ‘sentiment’ dictionaries

1 Create raw word count

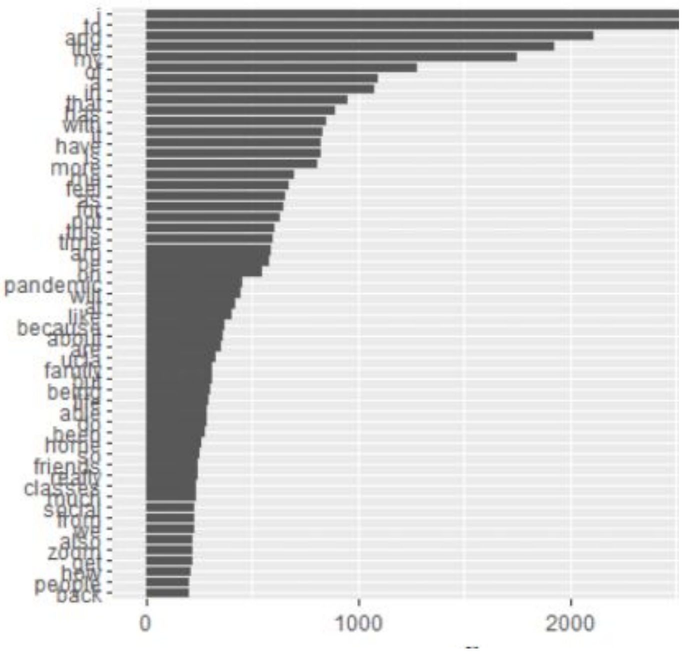
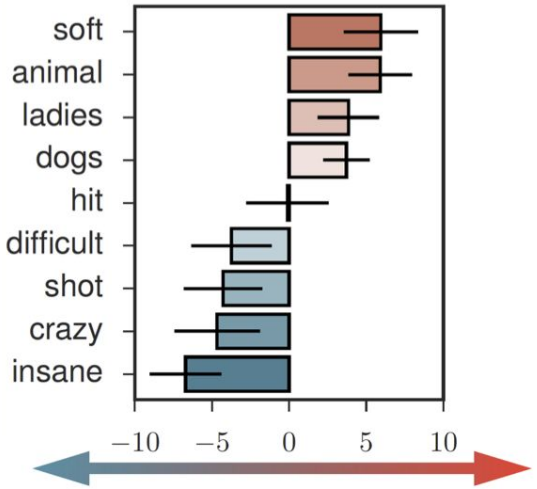


Figure: Frequency plot of raw word count for post-pandemic data

2 Group words by sentiment

- Raw word counts are filled with ambiguous or neutral words
- We used lexicons for four different emotions, joy, anger, positive, and negative to group significant words

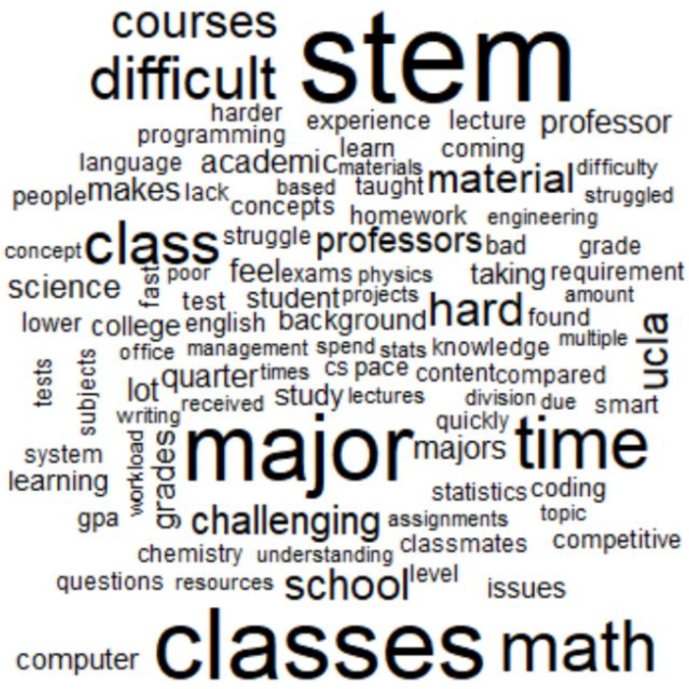


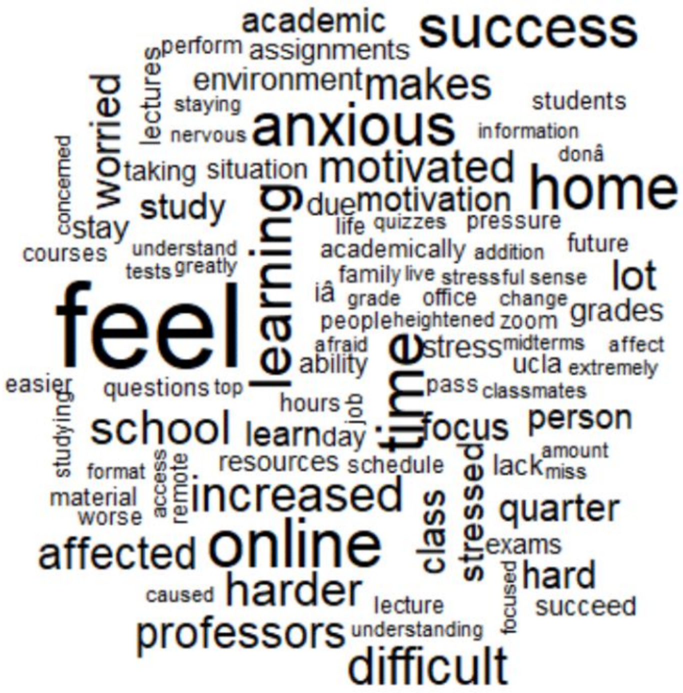
3 Segment by characteristics

- **Post/pre coronavirus** was the most pertinent factor to segment by
- **Question segmentation** was important as some questions elicit different emotions
- Factors such as **STEM/Intl. status, and living arrangement,** allowed us to follow up on hypothesis formed by our EDA
- Finally we segment by manually labeled sentiment value (0 or 1)

Word Cloud - Stress Related Questions

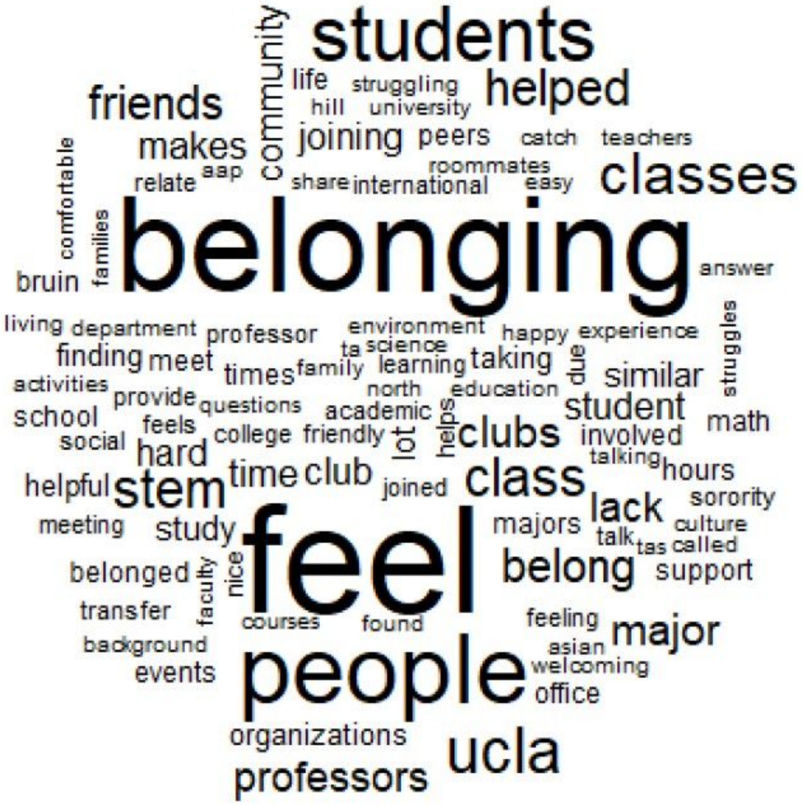
Pre-Pandemic Word Cloud	Post Pandemic Word Cloud

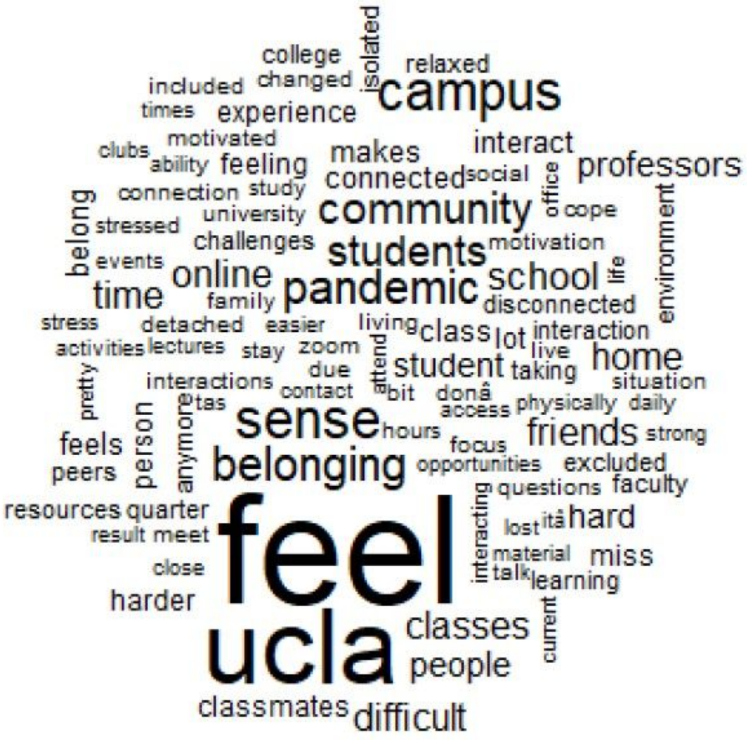




Word Cloud - Sense of Belonging Related Question

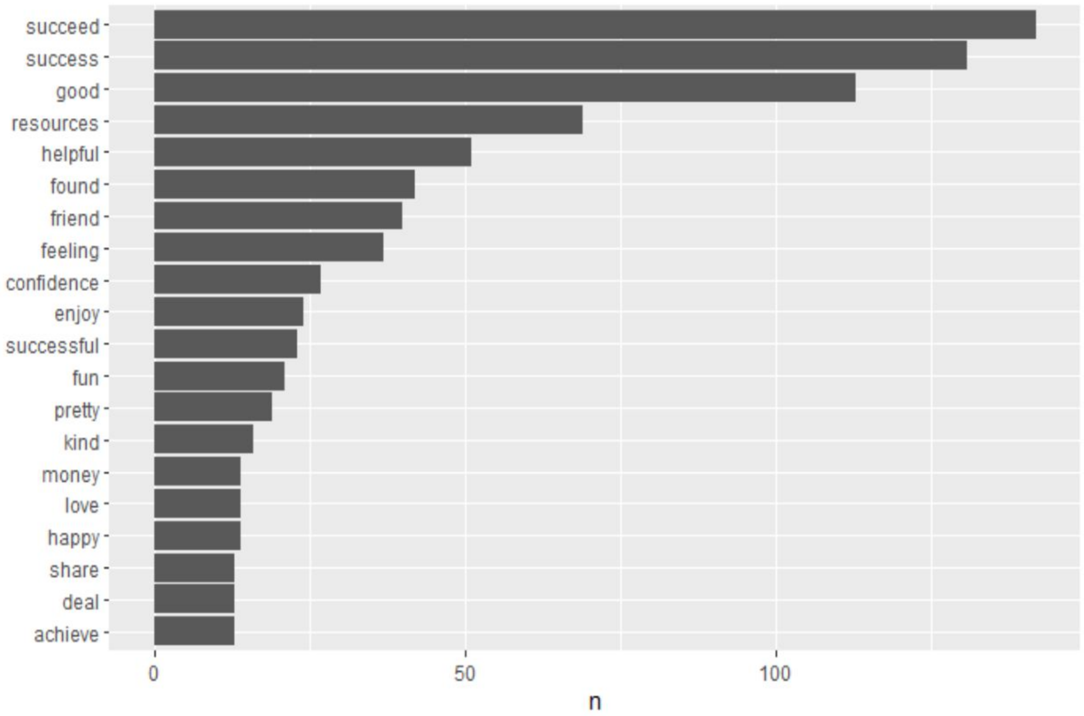
Pre-Pandemic Word Cloud	Post Pandemic Word Cloud



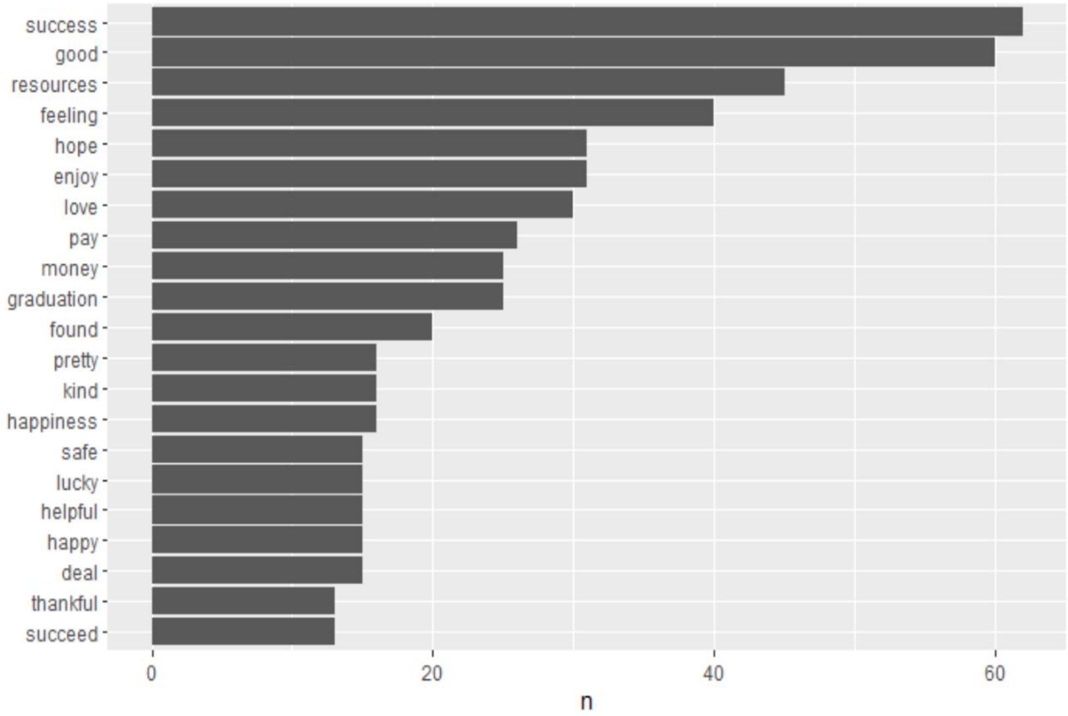


Word Frequency - Words Associated with Joy

Joy Pre Pandemic Word Frequencies

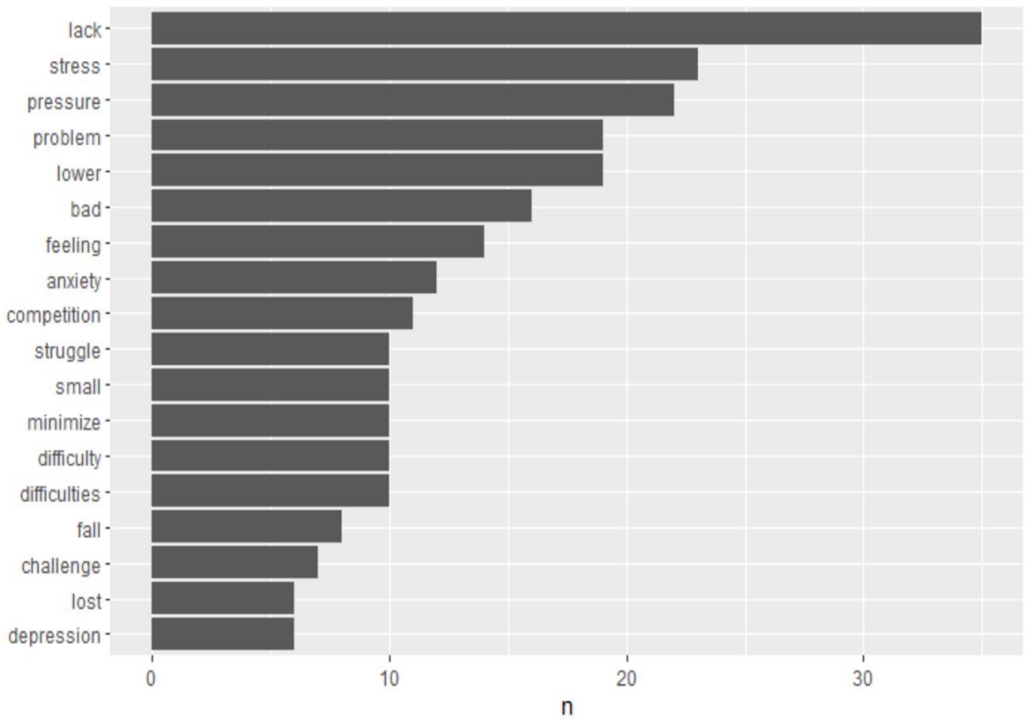


"Joy" Post Pandemic Word Frequencies

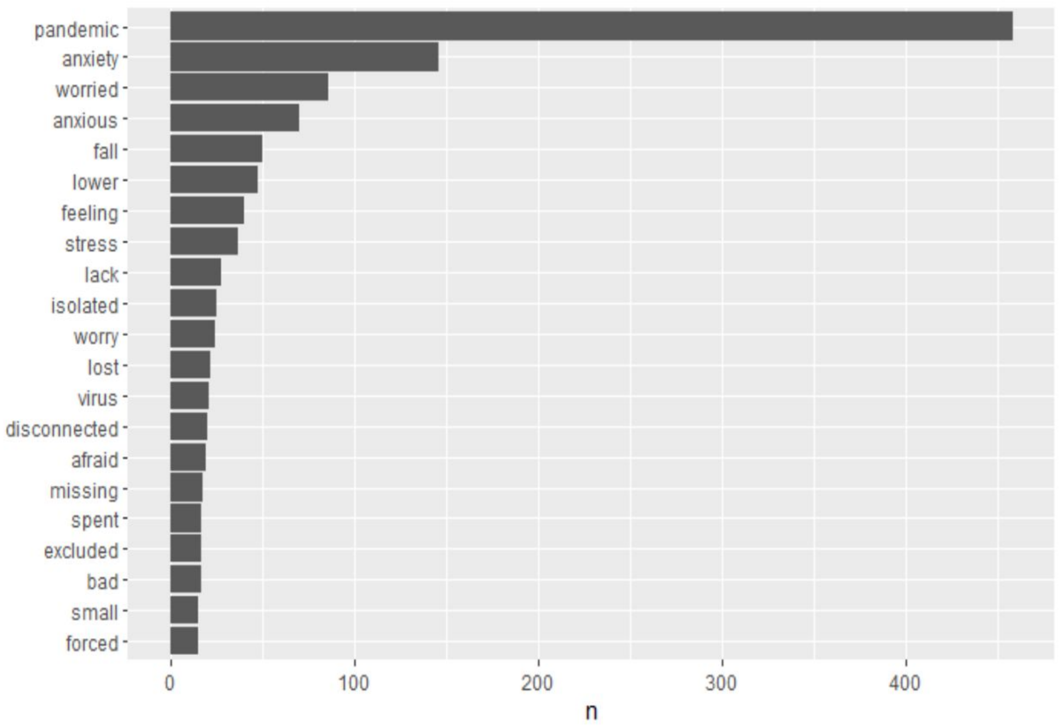


Word frequency for words associated with negative sentiment

Negative Pre Pandemic Word Frequencies

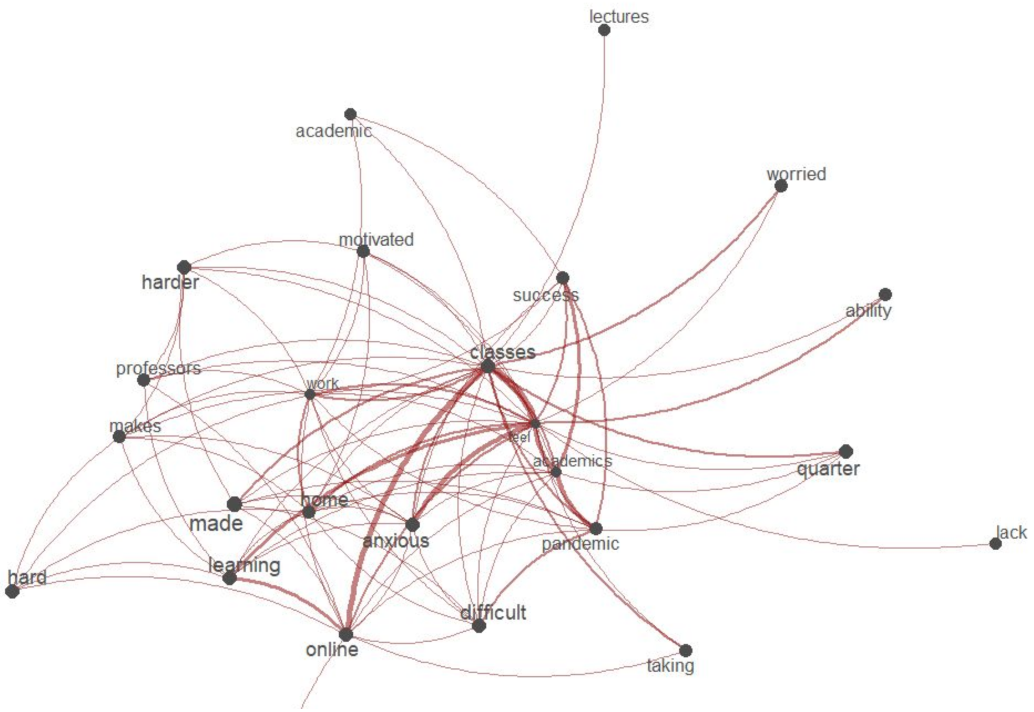


Negative Post Pandemic Word Frequencies

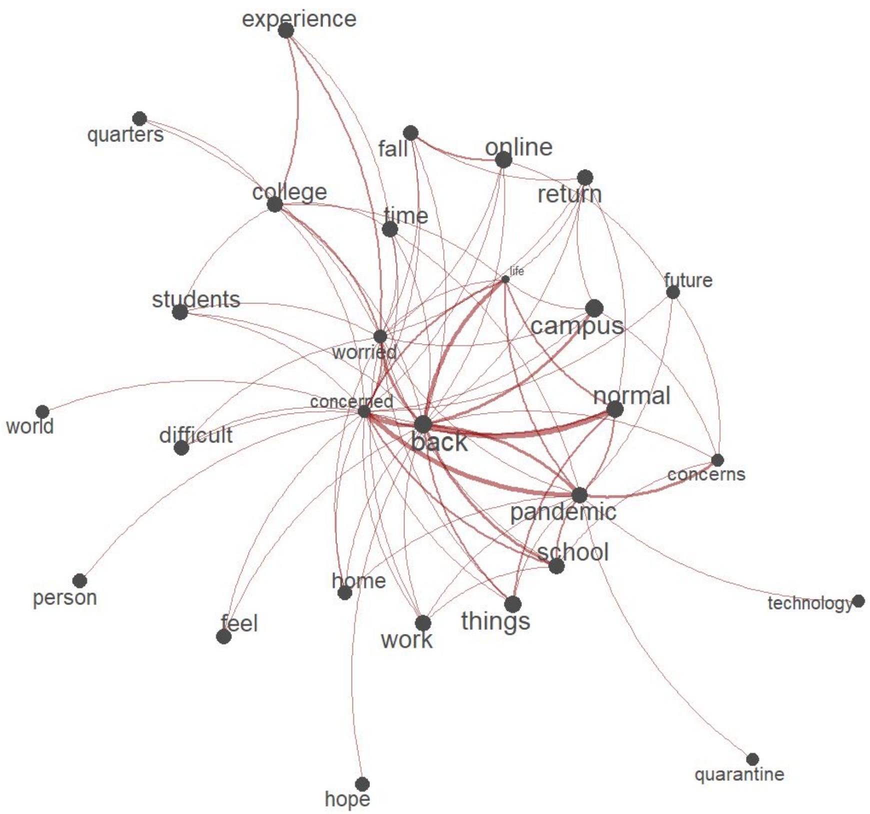


Word Networks - Stress

Academic Stress (Question 3)

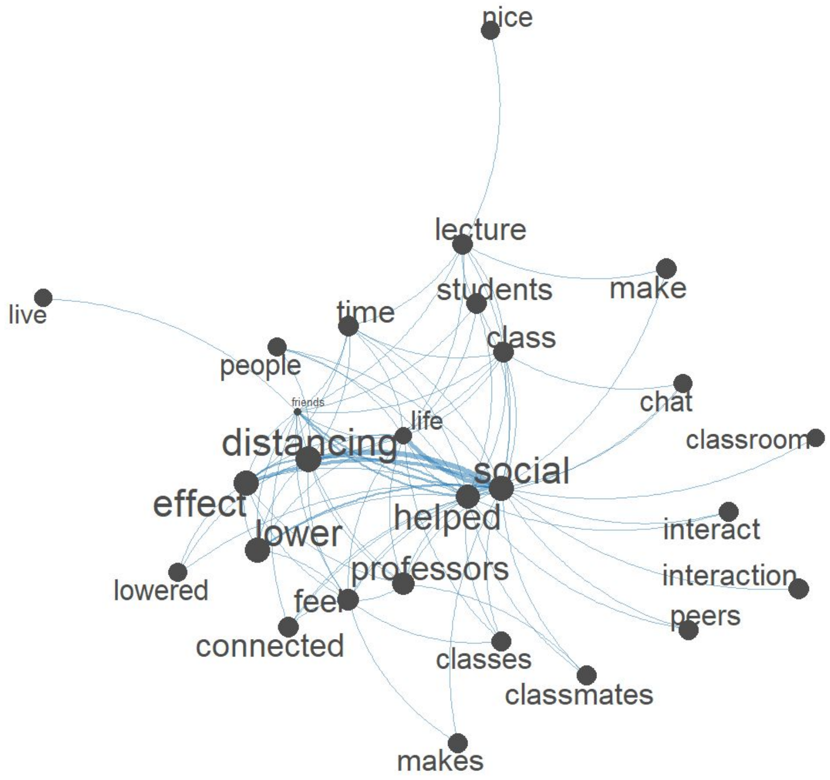


Post-Pandemic / Career Stress

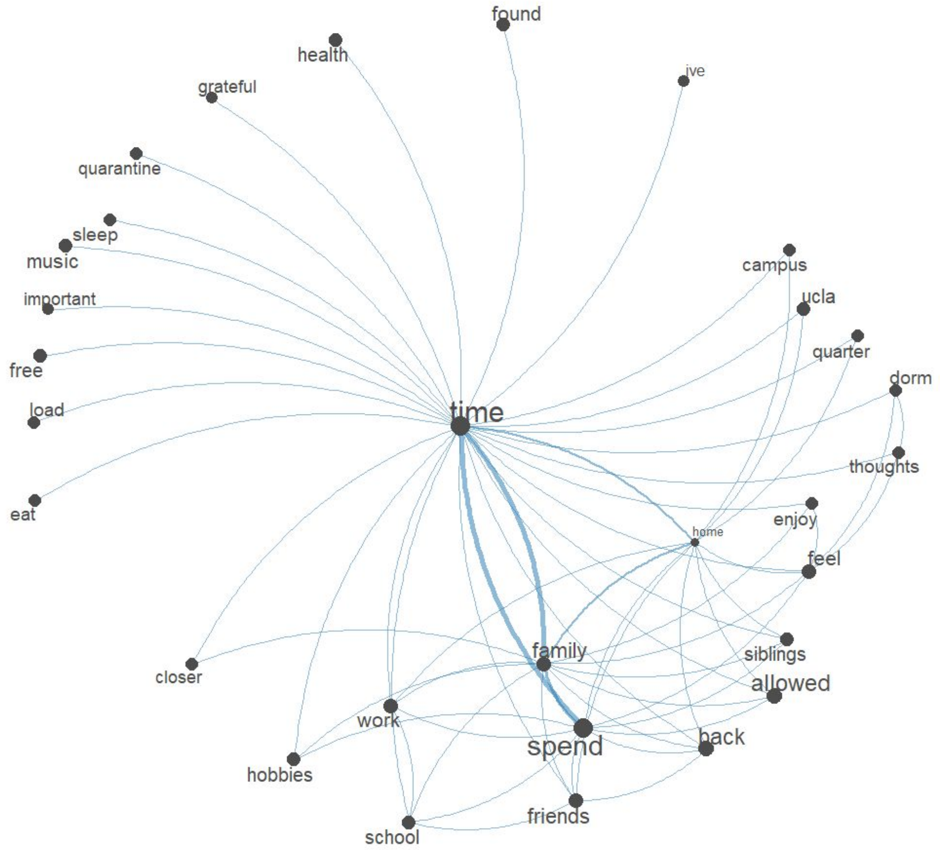


Word Networks - Positive Thoughts

Zoom Enabling Social Connection (Question 4)

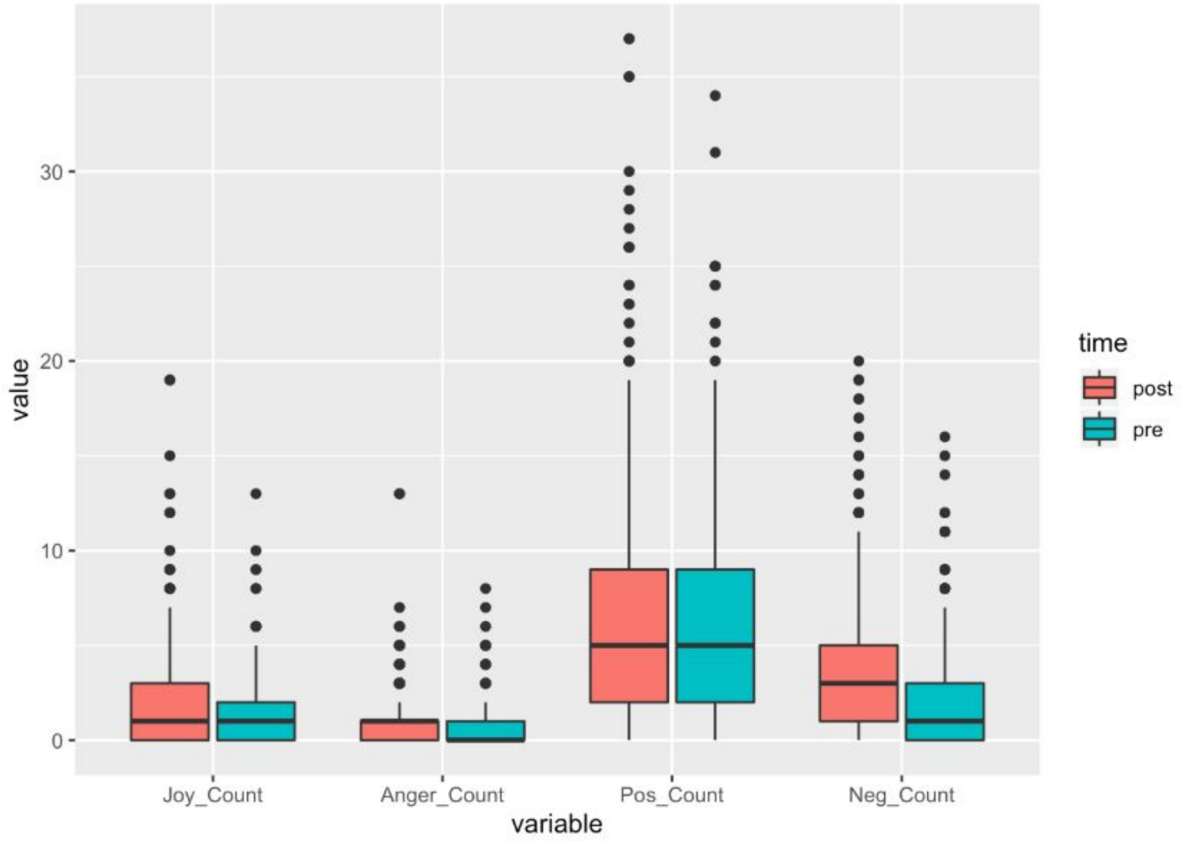


Positive Aspects of Experience (Question 6)



Sentiment Word Frequency Count

Word Count per Line by Sentiment and Pre/Post Pandemic

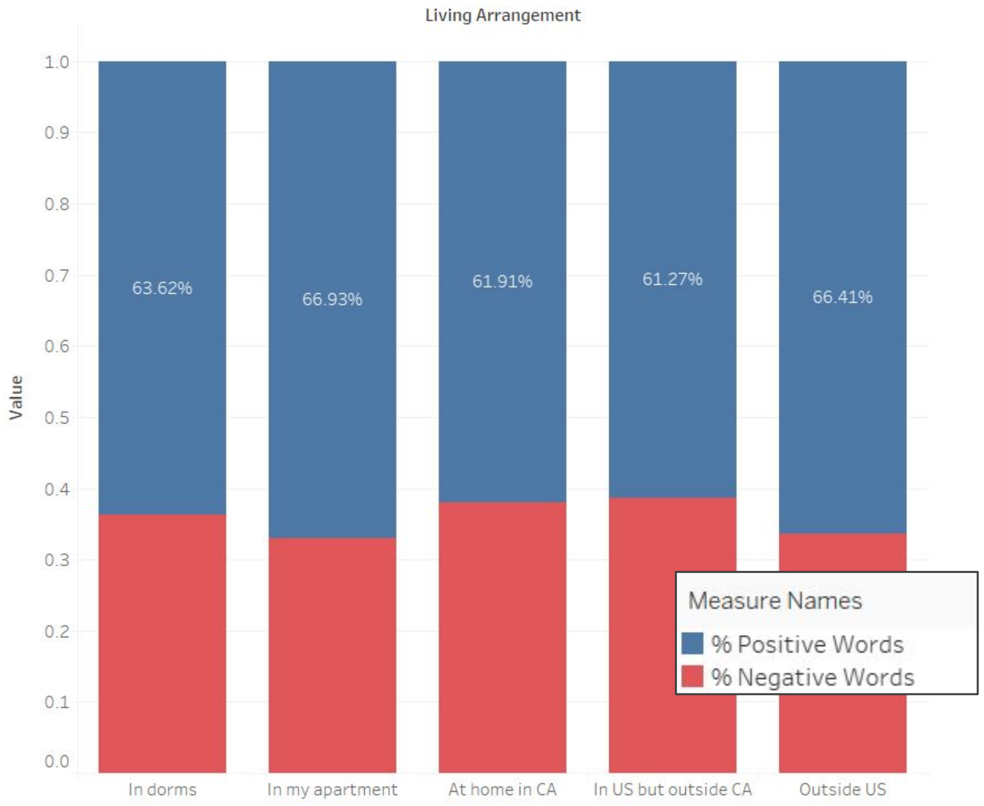


Commentary

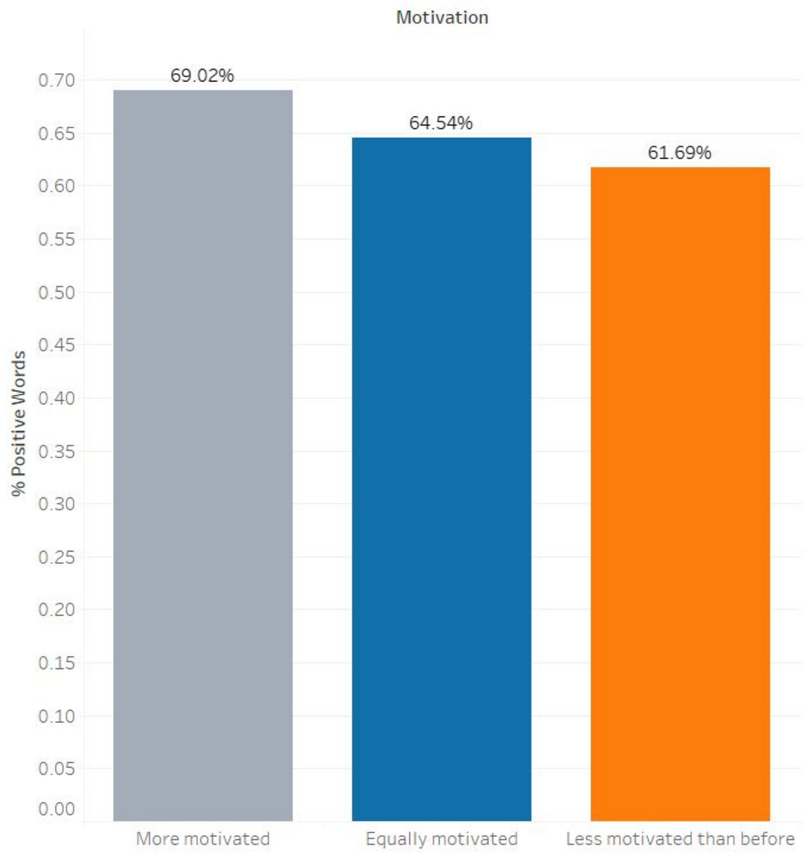
- For both data-sets positive words are most commonly used, likely due to positive words being more generally used
- Positive sentiments (joy & positive) say comparable levels of use pre and post pandemic
- However, negative sentiment count (anger & negative) saw an increase in per-line frequency post-pandemic

Sentiment Word Count by Demographic - Significant Relationships

Positive/Negative Words vs.Living Arrangement

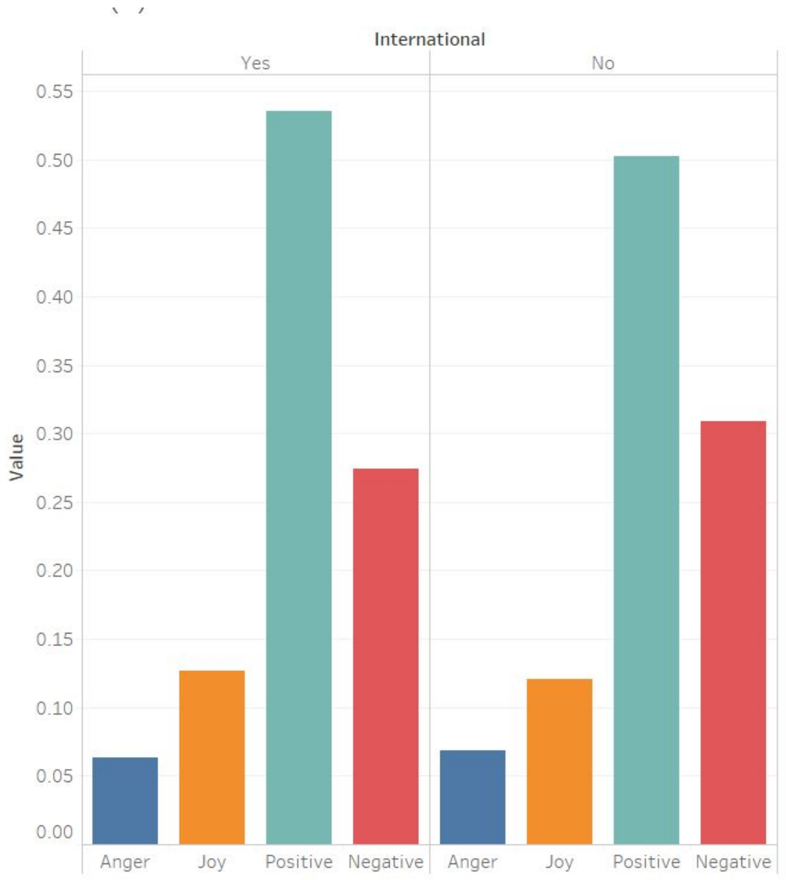


Positive Word Count vs. Motivation

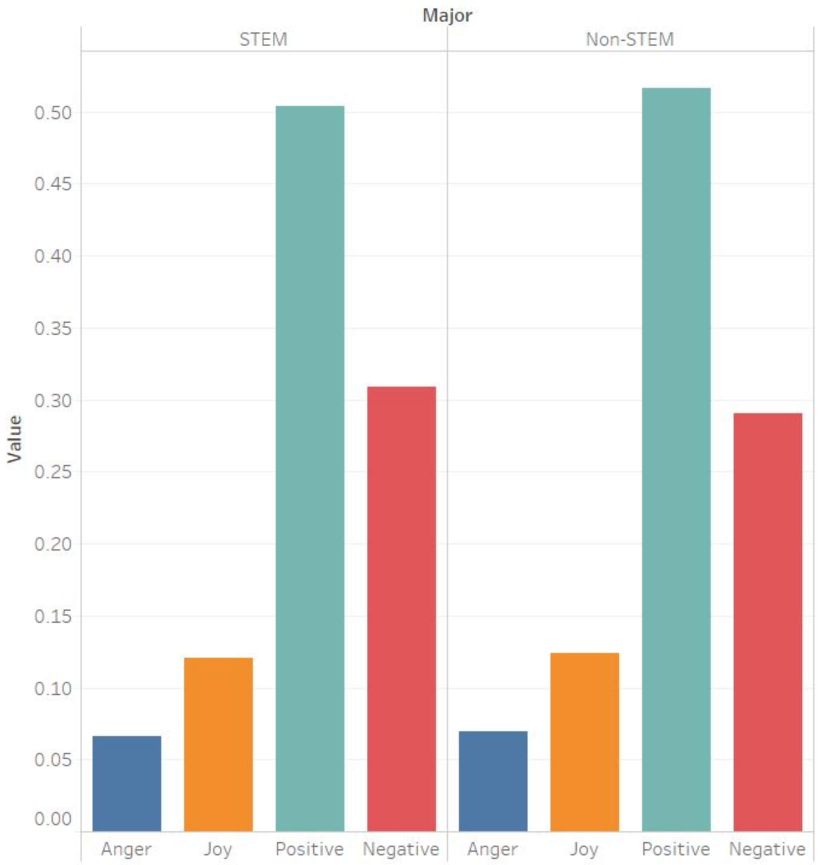


Sentiment Word Count by Demographic - Insignificant Relationships

Sentiment Word Count vs.International Status



Sentiment Word Count vs. STEM Major



Predicting Overall Sentiment with Relative Sentiment Word Counts

Pre-Pandemic Logistic Model Predicting for Sentiment

Coefficients	Estimate	Std. Error	t value	P-value
Joy_Count	0.012533	0.017913	0.700	0.484720
Anger_Count	-0.046891	0.028792	-1.629	0.104555
Pos_Count	0.049740	0.006259	7.946	5.13e-14 ***
Neg_Count	-0.051571	0.013450	-3.834	0.000157 ***

Post-Pandemic ANOVA Predicting for # of Positive Words

Factor	Df	Sum Sq	Mean Sq	F value	P-value
Motivation	3	0.2216	0.073877	1.6615	0.1746
Living_Arrange	4	0.2338	0.058460	1.3148	0.2636
Residuals	438	9.4749	0.044463		

Post-Pandemic Logistic Model Predicting for Sentiment

Coefficients	Estimate	Std. Error	t value	P-value
Joy_Count	0.0148148	0.0166335	0.891	0.374
Anger_Count	0.0005858	0.0216156	0.027	0.978
Pos_Count	-0.0006103	0.0073086	-0.084	0.933
Neg_Count	0.0008086	0.0093261	0.087	0.931

- Pre-pandemic sentiment word counts can partially predict for overall sentiment
- Post-pandemic word counts fail to predict actual overall sentiment
- ANOVA for predicting number of positive words fails to show display statistical significance, despite appearance of relationship

Conclusion

1 Main Takeaways

- Students reoriented the nature of their positive and negative sentiments after pandemic
  - Positive words like ‘money’ or ‘friends’ replaced by words like ‘time’ and ‘family’
  - Negative words like ‘stress’ and ‘pressure’ replaced by words like ‘anxiety’ and ‘isolated’
- Students increase frequency of negative sentiment, with few significant differences between demographic groups
  - Motivation and living arrangement are strongest demographic predictors, but are not statistically significant

2 How we can improve

1. Increase size of data-set with long-term aim of training sentiment analysis model to assign values to the positive/negative sentiments
2. Collect more post-pandemic demographic & survey variables to directly compare against pre-corona data
3. Build customized sentiment dictionary to better identify meaningful words
4. Benchmark against non-UCLA datasets to see how students responded relative to others

3 Final Thoughts

- COVID-19 affected UCLA student psychology both in terms of sentiment and nature of our sentiment (i.e. what we are angry/happy/sad about)
- Students across demographics appear to be facing the pandemic in a similar manner
- COVID-19 will affect our health, economy, and psychology for years to come. This is only a study of its immediate effects and follow up study is warranted to build a more complete picture

Any Questions?