Measuring Volunteering Impact on Students

Long Lin, Madeline Blasingame, Stacy Deng, Jamie Kim, Georgia Ronis von Helms, Jordan Siegel

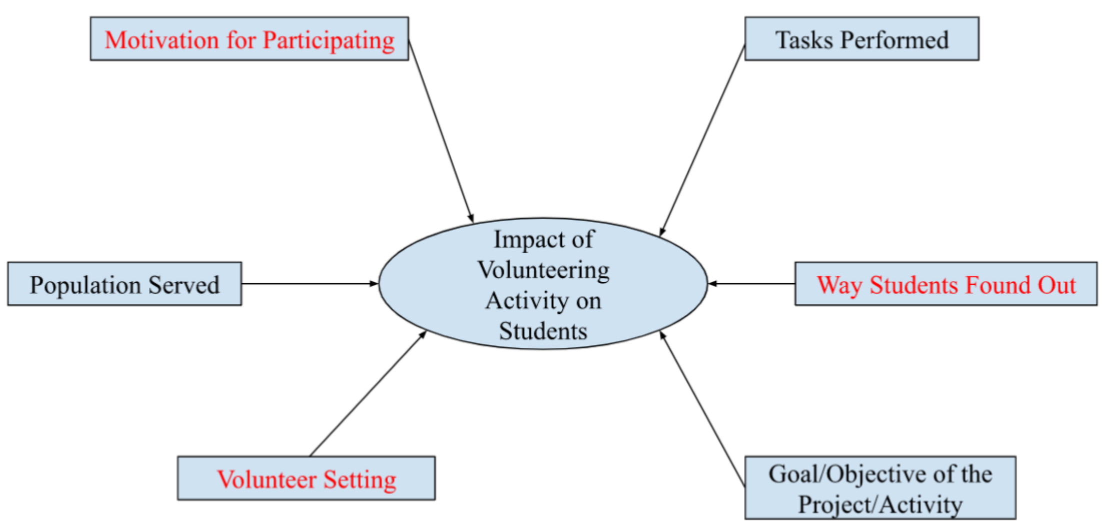
June 13, 2023

Abstract

UCLA students were surveyed and asked to choose one volunteer activity they pursued during the 2022-23 academic year and describe it. Survey respondents answered both multiple choice and short answer questions about their volunteer experiences. For this project, we aimed to use the text-based responses of the survey to answer two questions - what impact the volunteering activity had on students, and whether the impact was related to their motivation for participating, the way students found out about the volunteer opportunity, and the setting of the activity. To clean the data, we reduced the number of categories in each non-text variable. One of the most important columns in this dataset for our analysis was ImpactOnYou_Text where students wrote long-form responses to how the volunteering experience impacted them.

For our analysis, we focused on text analysis. We used bi-gram and tri-gram analysis of ImpactOnYou_Text to look at the specific word-choice of students. We also created a word cloud of words students used to describe their feelings while volunteering to determine impact on students. Next, we performed sentiment analysis using the NRC text lexicon on the ImpactOnYou_Text column, both as a whole and split into smaller data frames based on motive, setting, and how a student found out about their volunteering activity. Finally, we used ANOVA to investigate potential mean differences related to the analyzed factors.

Variables



Data Cleaning and Visualization

The Frequency, Frequency_Other_Text, and Short Term Status Variables

Pre-Cleaning Contingency Table: Frequency of Volunteering and Short Term Status

	Daily	Weekly	Monthly	1+ times per year	Other	Total
Regularly	31	240	65	44	13	393
Only Once	18	40	22	200	34	314
Total	49	280	87	244	47	707

The Frequency variable is a categorical variable that records the answers to the question asked "How frequently have you engaged in this volunteer opportunity?" The responses were Daily for a period of time, Weekly for a period of time, Monthly for a period of time, Once or more during the year, or Other. The most common response was that people engaged in their volunteer activity weekly and the second most common response was Once or more a year. It seems that monthly and daily volunteering was much less common.

The categorical variable ShortTermStatus of volunteering for survey respondents records answers to the question "Was this a short-term activity for you (i.e., did you do it only once)?" The responses were "No, I do it regularly" and "Yes, I did it only once." There was a fairly even split between responses, although slightly more people reported that they did this activity regularly, about 55.6%, and the remaining 44.4% only did the activity once.

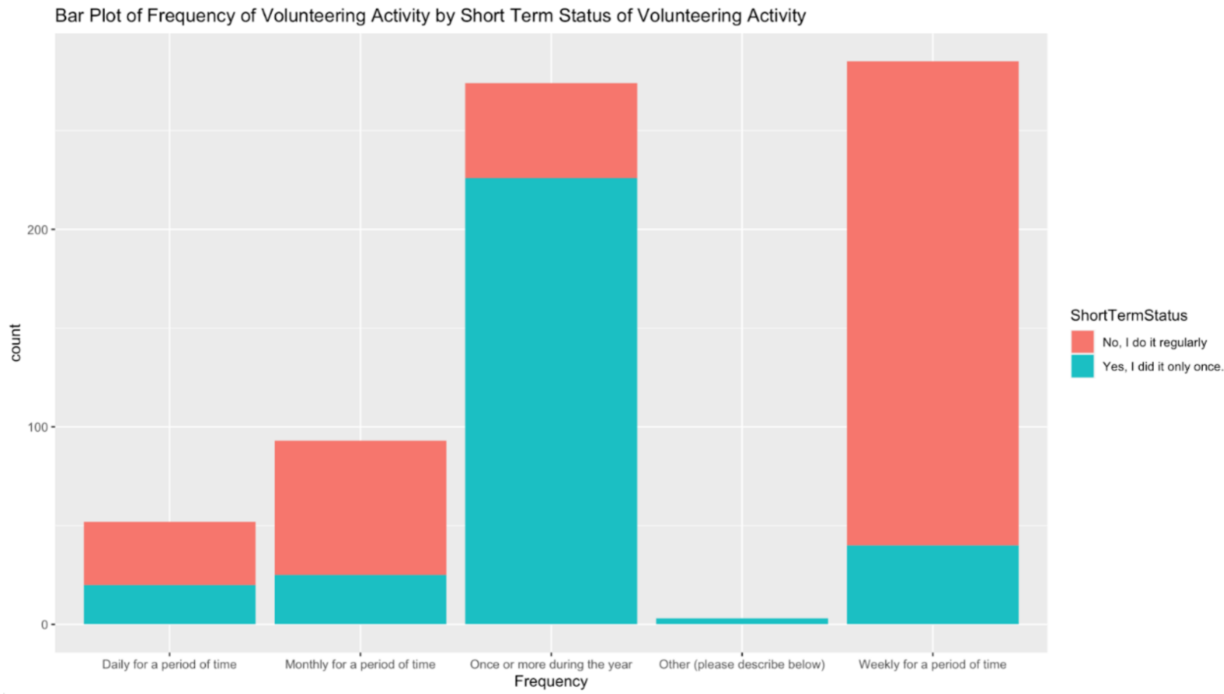
In the Frequency_Other_Text variable, those who chose "Other" for the Frequency variable provided a written explanation for their choice. Upon examining this variable, it became clear that most of the individuals who chose Other could easily be placed into one of the other categories. For example, there were a number of people who wrote that they performed their volunteer activity bi-weekly or semi-weekly. These individuals were removed from Other and placed in the Weekly category. Most of those who chose others were placed into the Once or more per Year category, because often their written explanations indicated that the volunteering activity they did was a one-off event. It is possible that the "Once or more during the year" option might have been confusing to some respondents because it does seem to imply that some amount of (infrequent) regularity is required. After re-coding the respondents who chose "Other" into alternative categories we were left with only three responses in the Other category, therefore it seems logical to disregard this category in any further analysis.

Post-Cleaning Contingency Table: Frequency of Volunteering and Short Term Status

	Daily	Weekly	Monthly	1+ times per year	Other	Total
Regularly	32	245	68	48	0	393
Only Once	20	40	25	226	3	314
Total	52	285	93	274	3	707

The Stacked Bar Plot below depicts the relationship between the variables **frequency** and **ShortTermStatus**. The plot below and table above indicate that most individuals chose either "Once or more during the year" or "Weekly for a period of time." A much higher proportion of those who only did this volunteering activity once or more a year indicated that they only did this activity once. Those who reported volunteering weekly, monthly, or daily volunteering indicated they did the activity regularly on the short term status question at a much higher rate that those who said they only did the activity one time.

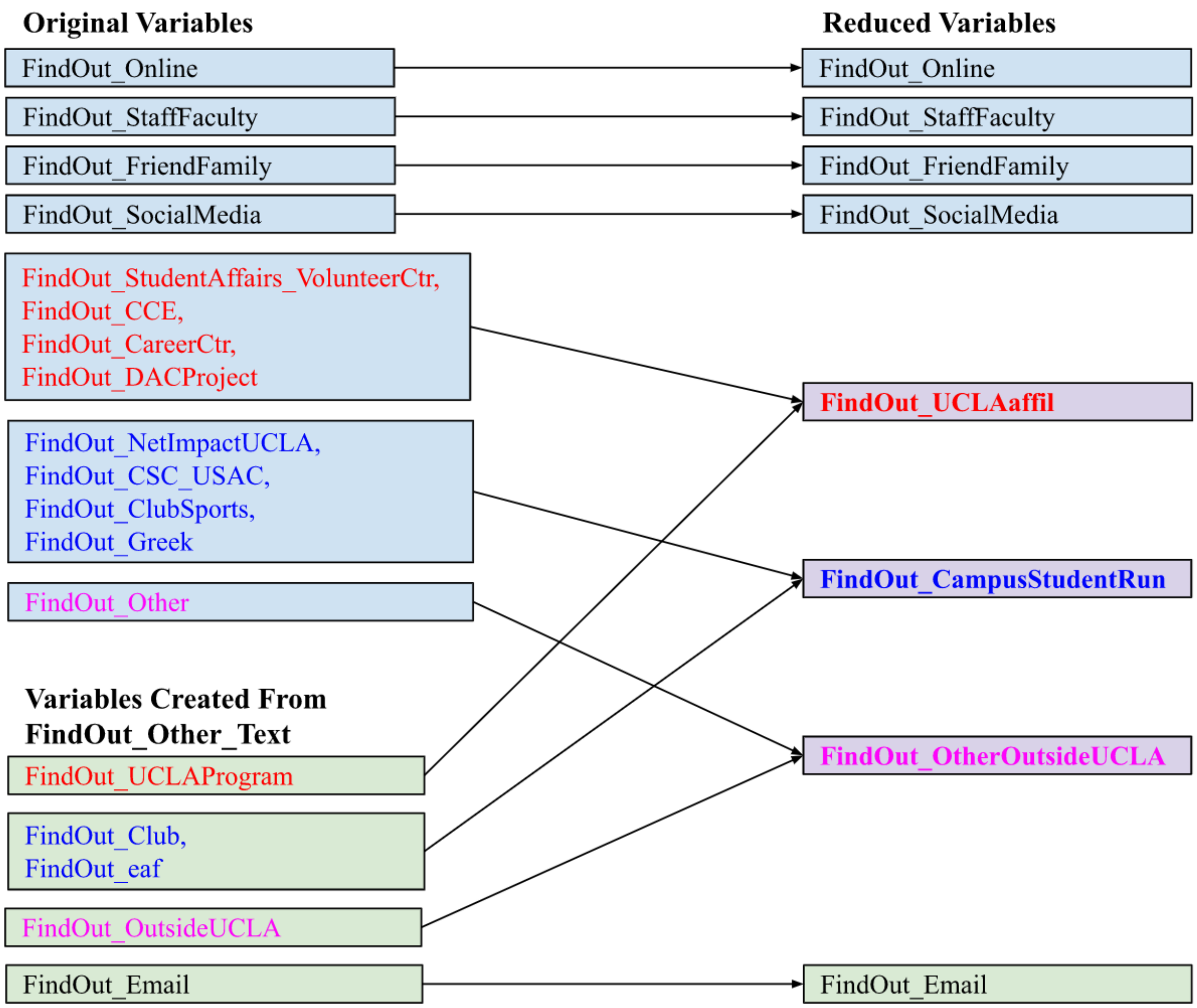
Looking at these two variables together seems to result in some contradictions in the data. For example, around 38% of those people who reported doing their volunteering activity at a daily rate, also said at the same time that they only did this activity one time. This seems to indicate that there might have been some confusion among respondents about how to answer the frequency question if the volunteer activity was a one-time-only event.



The Find Out Variable

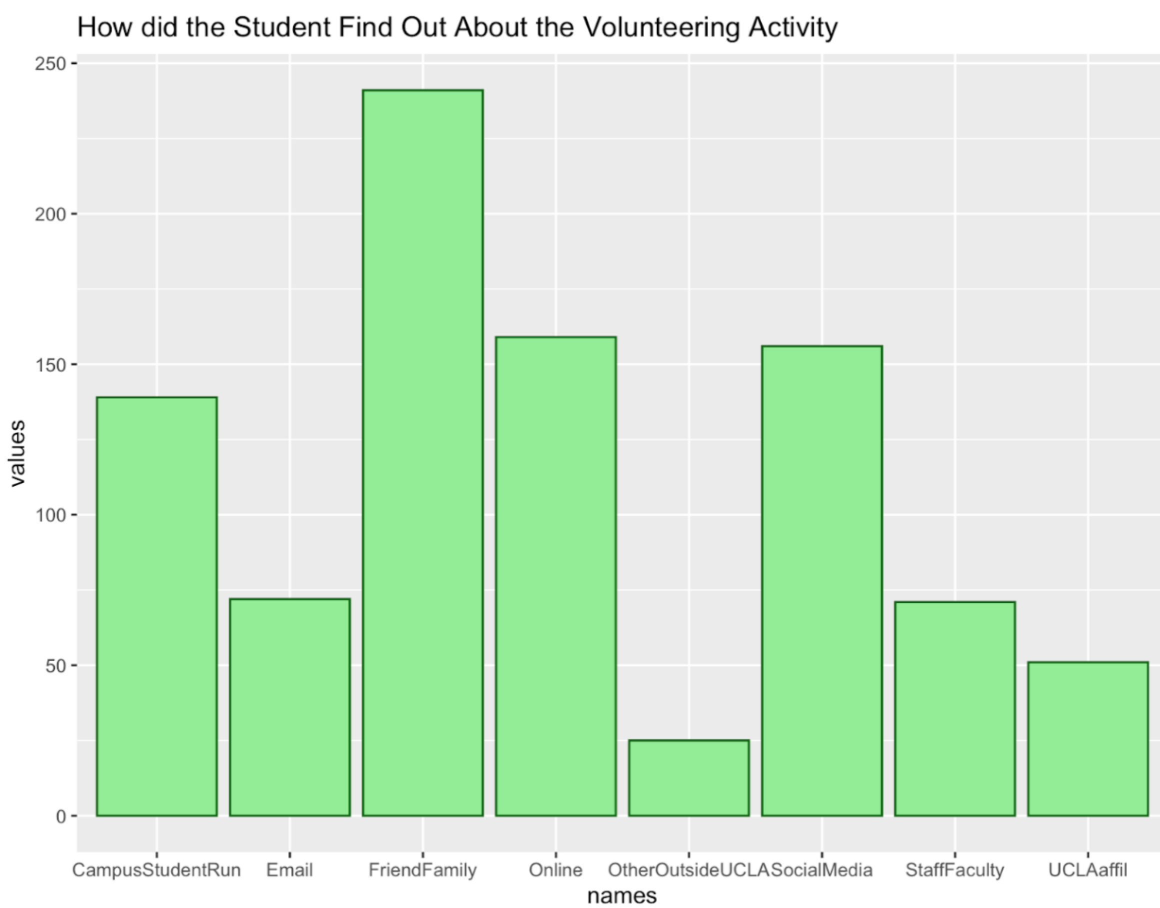
There were several variables associated with how the survey respondent found out about the volunteering activity they were involved in. The original variables were FindOut_Online, FindOut_SocialMedia, FindOut_FriendFamily, FindOut_StaffFaculty, FindOut_StudentAffairs_VolunteerCtr, FindOut_CCE, FindOut_DACProject, FindOut_Greek, FindOut_CareerCtr, FindOut_ClubSports, FindOut_CSC_USAC, FindOut_NetImpactUCLA, FindOut_Other. Respondents were allowed to select Yes for multiple options about how they found out about the activity. Upon examining the variables, we found that 177 people selected the FindOut_Other option. We examined the corresponding variable FindOut_Other_Text, wherein respondents provided a written explanation for why they chose other. There were certain recurring patterns in these explanations. Many people simply wrote that they had discovered the volunteer activity via email. Many others said they had found out from a club they are involved in on campus or from their professor or lecture. From these written responses, we created five new variables, removing the respondents from Other and into these new categories. The categories were FindOut_eaf (those who found their activity through The Enormous Activities Fair), FindOut_Email (those who found out via email), FindOut_Club (those who found about the activity through a UCLA club), FindOut_UCLAProgram (those who found out from a UCLA program or institution, FindOut_OutsideUCLA (those who found out about their activity outside of UCLA, such as through their church or job). After adding these variables, any individuals who had chosen some other category in addition to Other were removed from other. This left us with 23 "Others." Most of these were then removed from other and assigned to some of the already existing variables based on their corresponding text entry, leaving us with eight "Others."

After this work was completed it seemed prudent to reduce the number of variables by combining variables together, because many of the variables were selected by very few individuals. For example only 4 people found out about their activity through Net Impact UCLA and only two through the DAC project. We condensed any variables that related to UCLA organizations, programs, institutes, and offices into one variable named FindOut_UCLAaffil and this process of variable reduction can be seen in the figure below. The university has some official involvement in all of the categories that were placed under FindOut_UCLAaffil. We condense any variables relating to student run campus organizations such as clubs, USAC, or Greek life into the variable FindOut_CampusStudentRun. With these two new variables we have split where people found out about the activity into those who found out on campus from the student-run groups versus the school itself. We then combined the remaining others with those who found out about their activity from some off-campus source. The remaining original variables and our newly created email variable were not combined to form new variables. Each of these variables already had a fairly large number of "yes" responses.



Post-Cleaning Table: Where did Students Find Out About their Volunteering Activity

Variable	Count
Online	159
SocialMedia	156
FriendFamily	241
StaffFaculty	71
Email	72
UCLAaffil	51
CampusStudentRun	139
OtherOutsideUCLA	25

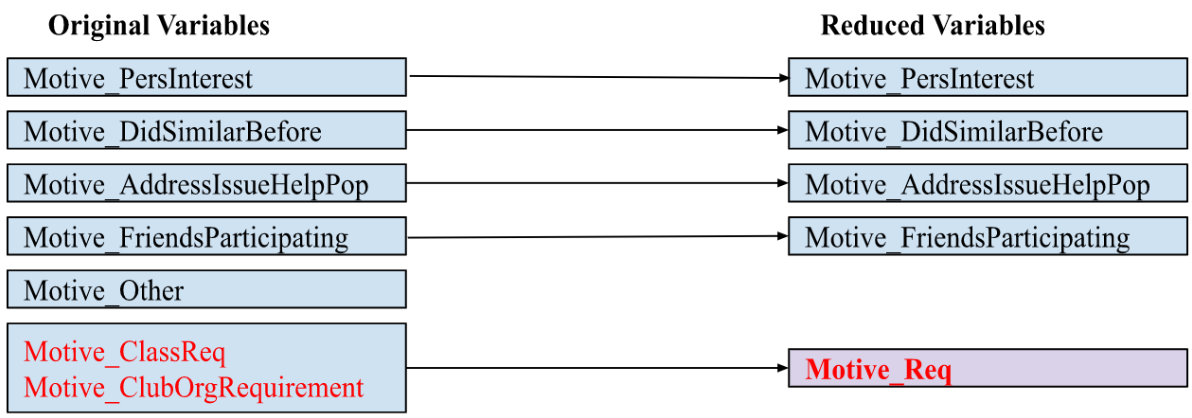


From examining the above table and bar plot we can see that the most selected option for how people found out about their volunteering activity was from family or a friend, this is followed by Online, Social Media, and student run organizations/clubs on campus which all had similar frequencies.

The Motive Variable

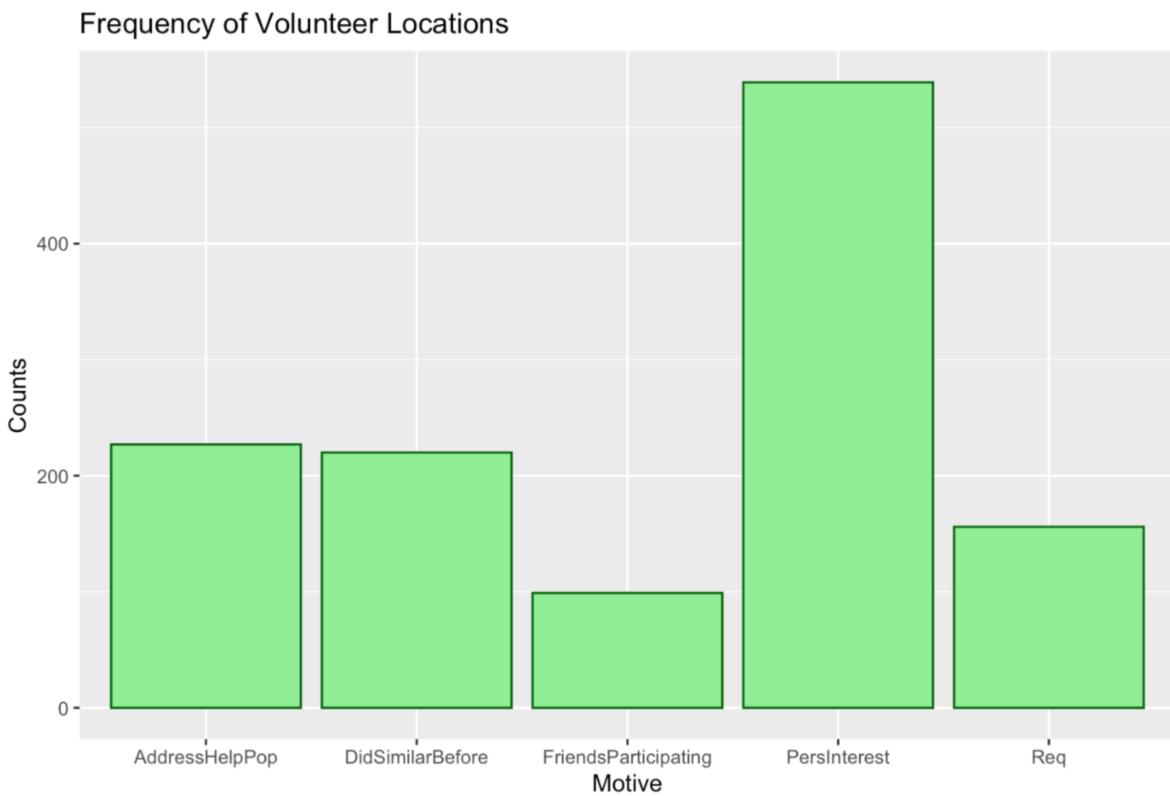
Students were able to select more than one option for what motivated them to participate in their volunteering activity. The original variables they could choose were Motive_PersInterest, Motive_DidSimilarBefore, Motive_AddressIssueHelpPop, Motive_FriendsParticipating, Motive_ClassReq, Motive_ClubOrgRequirement, and Motive_Other. We wanted to reduce the number of categories, so the first reduction that was made was with the ClassReq and ClubOrgRequirement variables. Because these variables were both related to students being required to complete the volunteering activity, we combined the two into one variable called Motive_Req.

When examining the Other category, there were only 40 students who had selected Other. Our goal was to reduce the number of students who had selected Other by looking at whether they had selected one of the already pre-set motives or had written a different motive in Motive_Other_Text. If the student had selected one of the original motive variables, we removed them from the Other variable because they had already accounted for their motive. For the few remaining individuals in Other, we were able to place them into one of the pre-set motives based on their description in Motive_Other_Text. This allowed us to reduce the number of individuals in the Other category to 0 and everyone's motive was accounted for.



Post-Cleaning Table: What Motive Made the Students Volunteer

Variable	Count
PersInterest	539
DidSimilarBefore	220
AddressIssueHelpPop	227
FriendsParticipating	99
Requirement	156

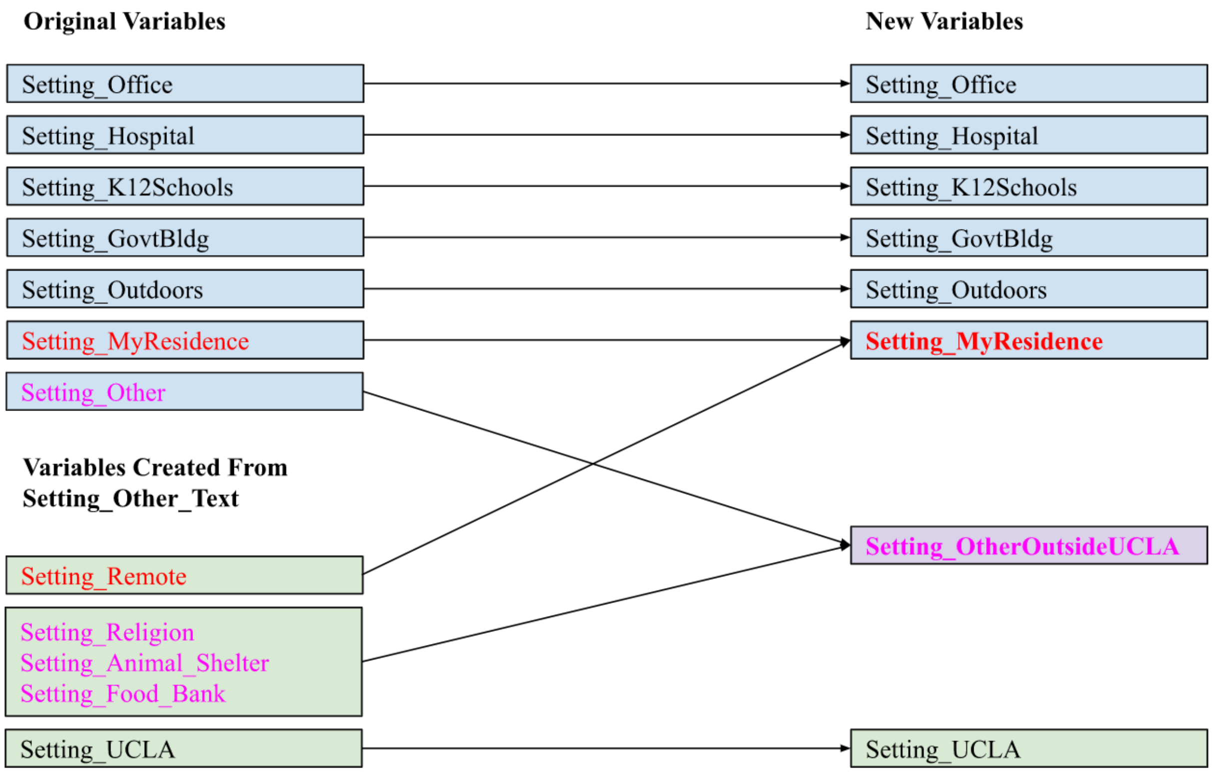


The Setting Variable

The Setting variable was another instance in which students would select multiple locations. The settings listed were Setting_Office, Setting_Hospital, Setting_K12Schools, Setting_GovtBldg, Setting_Outdoors, Setting_MyResidence, and Setting_Other. Setting_Other was the second most frequent setting option with 195 students. Because Other doesn't provide much detail on the setting, we wanted to diminish the frequency of that category as much as possible. First we examined the students who had said true to at least one of the already preset settings and selected Other. For those students, we removed them from Other because their setting was already accounted for. The remaining students who had selected Other had only chosen Other, so we wanted to make sure their settings were also examined. We looked at the variable Setting_Other_Text and found the most common locations students wrote in were zoom/remote, UCLA, places of worship, animal shelter, and food bank. We made the new variables Setting_Zoom, Setting_UCLA, Setting_Religion, Setting_Animal_Shelter, and Setting_Food_Bank. After removing students from Other and creating these new variables, we were left with 33 students in Other.

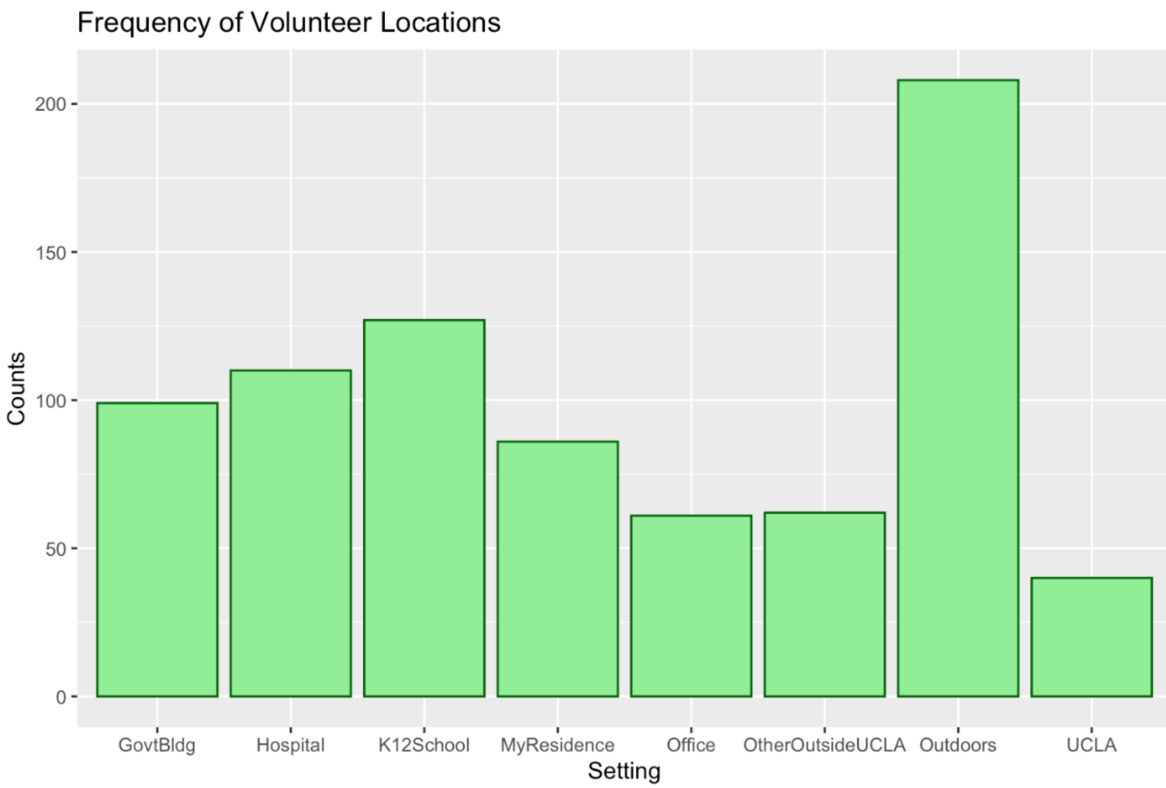
Our initial hunch was that if students had written in zoom or remote as their setting, they were completing the volunteering activities at their residence. To confirm this we looked at the students who had selected MyResidence and Other. Looking at the Setting_Other_Text for those students, they had written that it was either remote or on zoom, so they just did the activities in their home. This allowed us to move all the students from Setting_Zoom into Setting_MyResidence, increasing the MyResidence frequency from 55 to 86 and removing the Setting_Zoom variable. Once we reduced the Zoom variable, we were left with Setting_Religion, Setting_Food_Bank, Setting_Animal_Shelter, and 33 remaining in Setting_Other. The frequency for the variables Setting_Religion, Setting_Animal_Shelter, and Setting_Food_Bank were 17, 7, and 5, respectively. Because the counts for these variables, along with the 33 students in Setting_Other, is relatively low we decided to combine them into the variable Setting_OtherOutsideUCLA. It seemed impossible to create a category for each location when the frequency would be so small for each one. Setting_OtherOutsideUCLA accounts for the remaining Others but specifies that it was not completed at

UCLA because those individuals were taken out of Other into the new variable Setting_UCLA.



Post-Cleaning Table: Where Did Students Volunteer

Variable	**Count**
Office	61
Hospital	110
K12Schools	127
GovtBldg	99
Outdoors	208
MyResidence	86
UCLA	40
OtherOutsideUCLA	62



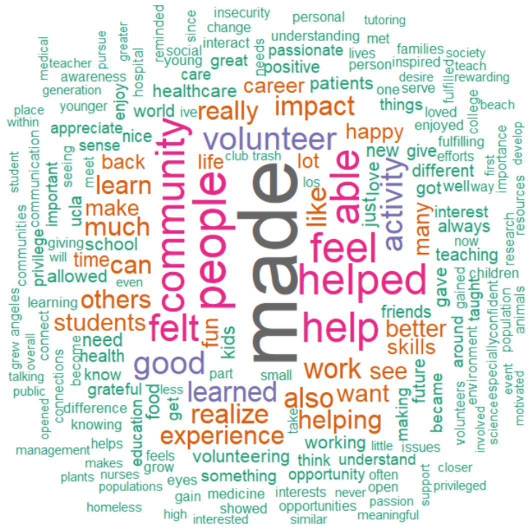
Words Clouds to Visualize text

In order to clearly visualize the data in the text responses of the impact of volunteering, the tasks undertaken, the population helped, and the goal/outcomes achieved, we discovered the most frequent words used in each response and then created word clouds for each category. We removed stop words such as "and" and "I" in order to only access words that could give greater insights.

Impact

Top 10 most frequently used words and word cloud for impact

Word	Frequency
Made	153
People	76
Helped	70
Feel	68
Help	65
Community	64
Able	63
Felt	58
Volunteer	47
Good	45



The majority of the emphasis is on different forms of “feel” and “help”, which show that the impact of volunteering is generally positive and people feel that they are able to help those in need.

Tasks Undertaken

Top 10 most frequently used words and word cloud for Tasks Undertaken

Word	Frequency
Food	102
Helped	96
Students	93
School	77
Help	56
Volunteered	49
People	46
High	44
Helping	41
Patients	39

The tasks undertaken in volunteering appear to be majority food-related, and then students or academic-oriented, and then finally patients or perhaps in hospitals.

Population Helped

Top 10 most frequently used words and word cloud for Population Helped

Word	Frequency
Helped	271
Students	128
People	94
Yes	83
School	69
Community	65
Children	48
Food	45
Homeless	41
High	40

Similarly to the frequencies in tasks undertaken, the main population helped in volunteering appears to be students and the young. After that, there is an emphasis on community, as well as helping the homeless, and perhaps tasks focused around food

Goals/ Outcomes Achieved Top 10 most frequently used words and word cloud for Tasks Undertaken

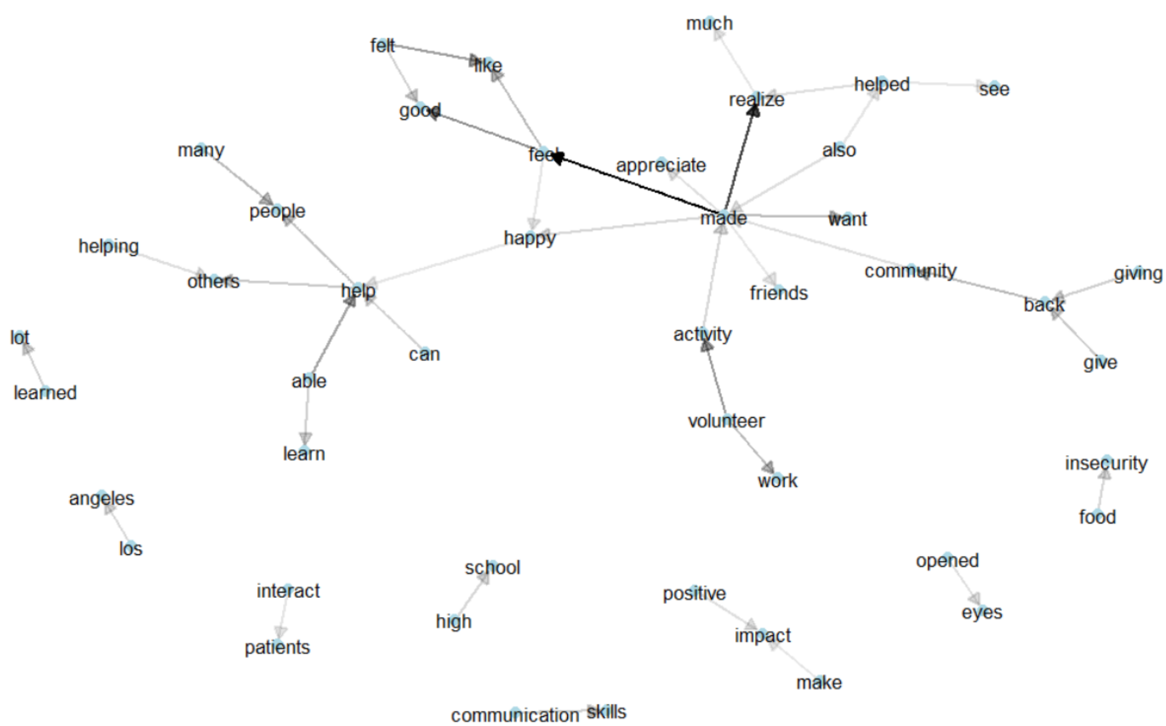
Word	Frequency
Help	144
Goal	120
Students	96
Community	81
Able	79
People	73
Helped	60
Experience	55
Get	45
Food	42

In conjunction with tasks undertaken and population helped, the outcome or goal achieved of the volunteering activity is centered around helping students and the community as a whole, and there appears to be an emphasis on food across all three response fields.

Analysis

Text Analysis

We created bigrams of the responses in the impact field of the survey in order to glean if there are mentions in the responses based on other fields such as motive for volunteering or the volunteering locations. In order to make the bigrams as effective as possible, we eliminated stop words such as “I” and “was” from the text. To effectively visualize the correlations between these words, we created a directed graph of all bigrams that appear at least five times in the impact question of the survey.



By studying the directed graph, several bigrams that relate to other fields in the survey appear. For example, “high school” appears, which suggests that volunteering at schools has a higher impact. Also, “food insecurity” suggests that goals and tasks related to food may lend itself to having a higher impact on those who performed those tasks. Lastly, “interact patients” shows that working in hospitals and interacting with patients features heavily in the impact of volunteering.

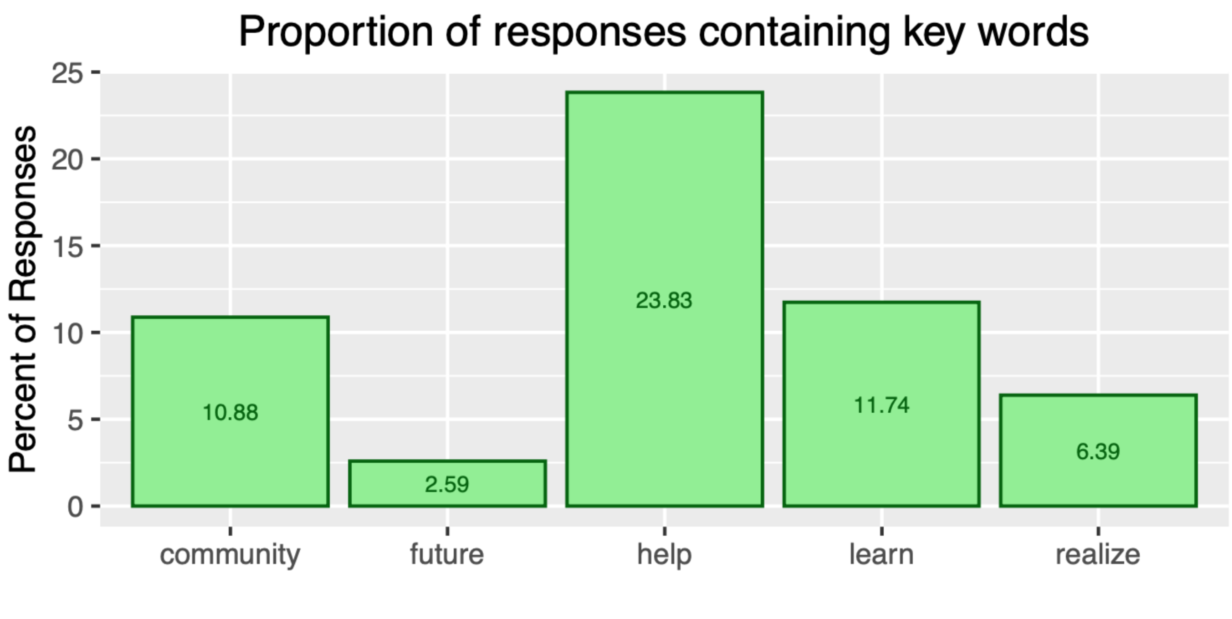
In order to understand how students were impacted by volunteering, we took a closer look at their text responses in the ImpactOnYou_Text column. In order to understand how the volunteering made the students feel, we accessed the word that came directly after “feel” or “feels” in the text. After removing filler words such as “like”, “very, “a”, and “as” we were left with a vector of words. We then created a word cloud of these remaining words.

By far, the most popular word was “good.” Many other students said volunteering made them feel “happy.” Overall, the students were impacted positively by the volunteering because it made them feel good.

By looking at the most popular tri-grams, you can also see popular keywords emerge. If you look only at the most popular words, they will simply be the most common English words such as “the” or “a” which don’t help understand what the students were writing about. However, by looking at tri-grams we are able to find keywords that have more value.

Tri-gram	**Frequency**
Made me **realize**	26
It **helped** me	21
Able to **help**	14
Learn more about	12
In the **future**	11
Me **realize** how	10
Me **realize** that	10
To my **community**	10

The keywords realize, help, learn, future, and community all stand out.



Note: percent is based on those who responded to this question (579) rather than total participants (707).

Word	Frequency	Example
realize	37	"It has made me **realize** that I love mentoring younger students from similar backgrounds as myself."
help	138	1. "It **help**ed me decide my career." 2. "It feels nice to encourage students who are also first-gen and to **help** them pave their path at UCLA."
learn	68	"I **learn**ed a lot about hospital work and heard medical terms thrown around and was able to learn through that"
future	15	"It made me much more open to volunteer work in the **future**"
community	63	"It made me realize how privileged I am and that I will always want to give back to my **community** and those who have less than me."

When looking at how students were using these specific words in their responses, a few trends emerge:

a. When students used the word “realize” or “learn” they were often discussing something new they found out either about themselves and their interests or about the work they were doing.

b. Students who used the word “future” were often discussing how it impacts their future plans either as a student volunteering or how it impacted their career choices.

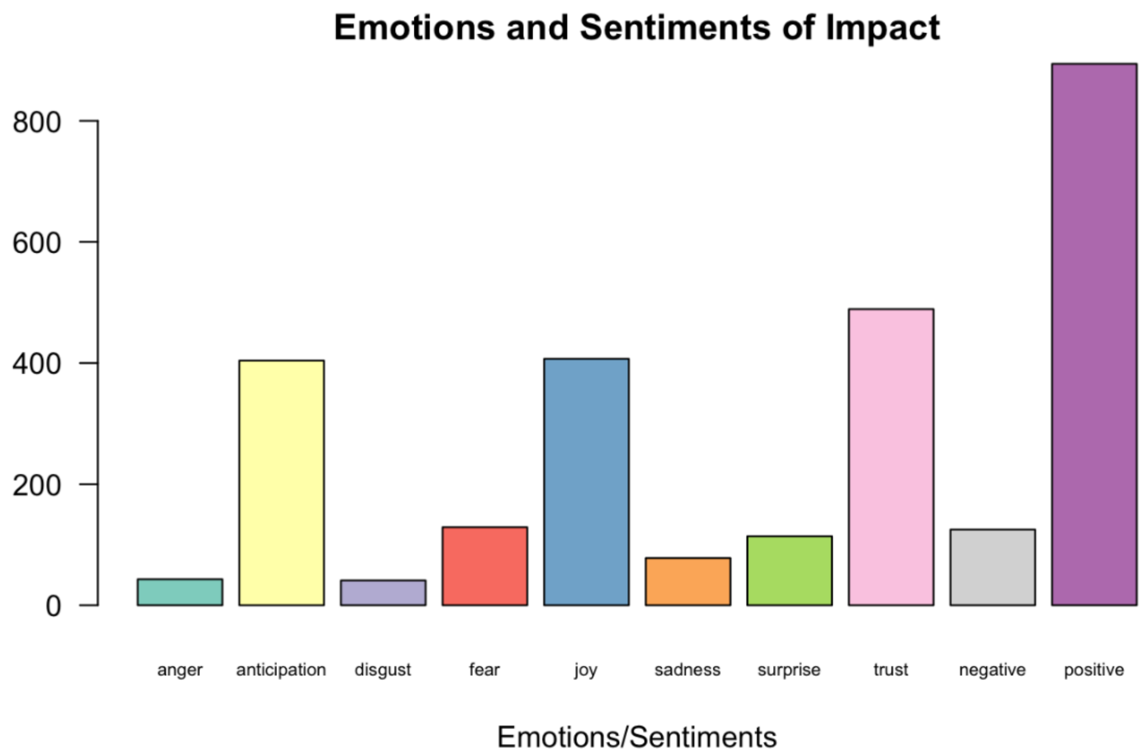
c. Students who used the word “community” emphasized how it was a positive experience to give back to their community.

d. Students used “help” in two main contexts - either how volunteering helped them or how it felt to help others.

In summation, volunteering made students feel good, helped shape their future plans and also acted as a way for students to learn something new.

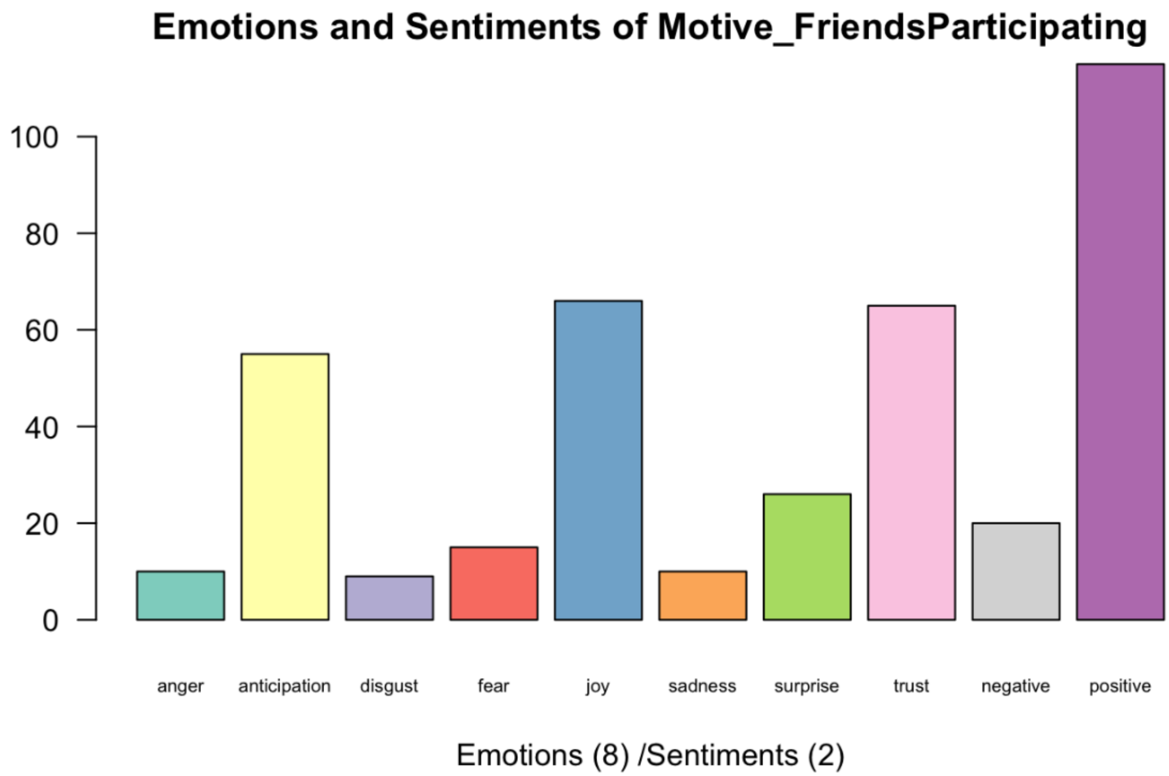
Sentiment Analysis

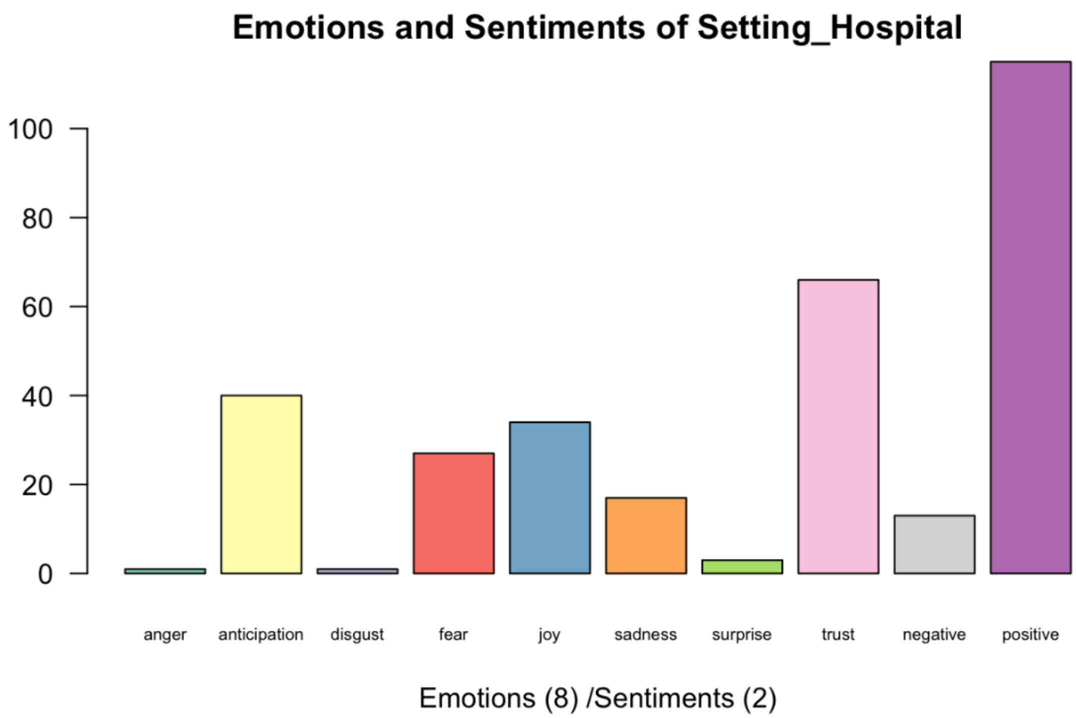
We performed sentiment analysis on the ImpactOnYou_Text column to see what the general impacts of volunteering were on students. We decided to use the NRC lexicon included in the syuzhet library to do our analysis because the NRC lexicon sorts words into eight emotion categories (anger, anticipation, disgust, fear, joy, sadness, surprise, trust) in addition to the two sentiments (positive, negative). This would allow us to analyze the emotional impact that volunteering has on students beyond just having a positive or negative impact. For instance, if the categories of emotion with the highest frequency are joy and trust, it's implied that volunteering is generally a positive experience that allows students to connect with others.

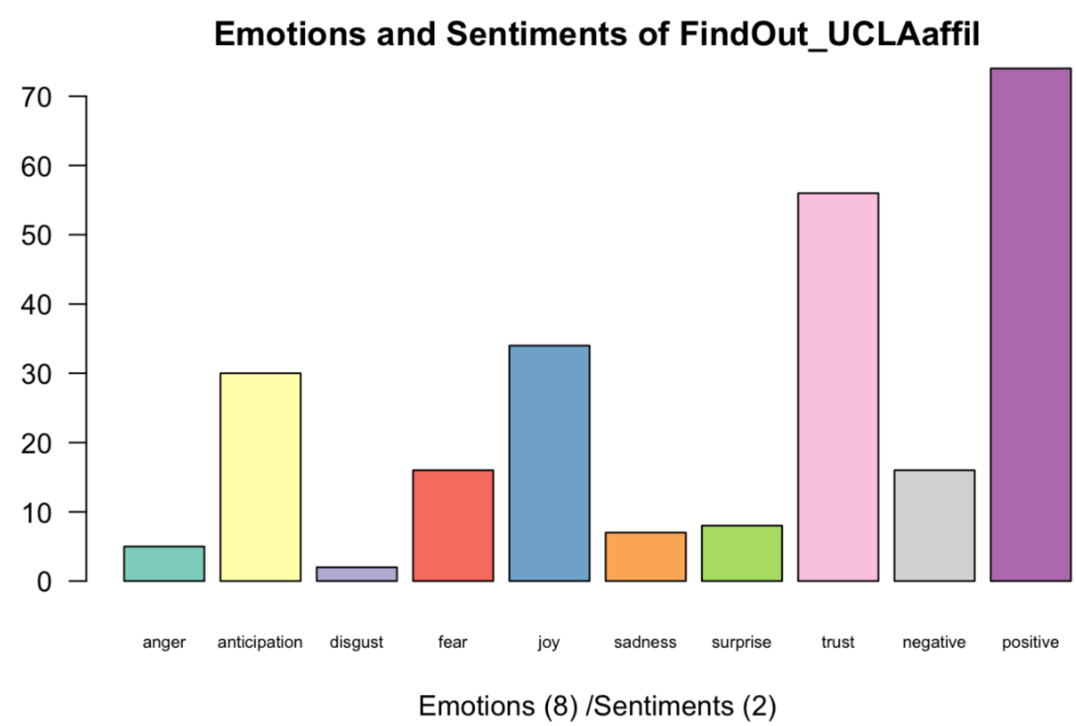


From the bar graph we can see that trust is the emotion with the highest amount of words used, followed closely by joy and anticipation. Looking at the sentiments, positive words are used far more often than negative ones, and so our graph seems to indicate that in general, volunteering is ultimately a positive experience for students.

While the general impact of volunteering seems to be positive, we also wanted to see if there were any particular motive, setting, or how a student found out about their volunteering activity that could skew the overall emotions and sentiments. To do so, we separated each individual motive, setting, and find out category into individual data frames and ran sentiment analysis on the ImpactOnYou_Text column of each data frame. The resulting bar graphs show that most of the individual data frames follow the same trends as that of the general bar graph of ImpactOnYou_Text, with little to no differences in the pattern.


Emotions and Sentiments of Motive_FriendsParticipating
100
80
60
40
20
0
anger
anticipation
disgust
fear
joy
sadness
surprise
trust
negative
positive
Emotions (8) /Sentiments (2)


Emotions and Sentiments of Setting_Hospital
100
80
60
40
20
0
anger
anticipation
disgust
fear
joy
sadness
surprise
trust
negative
positive
Emotions (8) /Sentiments (2)


Emotions and Sentiments of FindOut_UCLAaffil
70
60
50
40
30
20
10
0
anger
anticipation
disgust
fear
joy
sadness
surprise
trust
negative
positive
Emotions (8) /Sentiments (2)

As we can see from the bar graphs above, regardless of motive, setting, or how a student found out, positive sentiments outnumber negative ones. Trust seems to be the highest or one of the highest in all cases, followed by either anticipation or joy, and the emotions with the lowest count are anger and disgust.

It's worth noting that there may be a couple of issues with using sentiment analysis to measure the impact of volunteering on students. NRC assigns each individual word of the ImpactOnYou_Text column with positive or negative sentiments (or none if the word is not in the lexicon), which is why the counts in the first sentiment analysis graph (~800) outnumbers the total number of rows the data frame has (707). Since the sentiments of individual words are counted and not the overall sentiment of a single row/response, it's possible that there are more overall negative experiences, but since more positive words are used, then the overall experience is reflected as positive during sentiment analysis.

It's also important to keep in mind that the words being measured in the lexicon do not take the context of the word into account, so positive words used in a negative context will be wrongly classified as positive, and vice versa. For instance, one of the replies for ImpactOnYou_Text was "I was able to understand homeless people and got more interested in solving homeless problems", which is a sentence with a positive sentiment. However, the word homeless is considered a word with a negative sentiment by the NRC lexicon, the opposite of the actual context. Manually reading through the data does show that for the most part, the sentiment analysis we ran was accurate, but there are a few cases like the example given that may make the analysis less accurate.

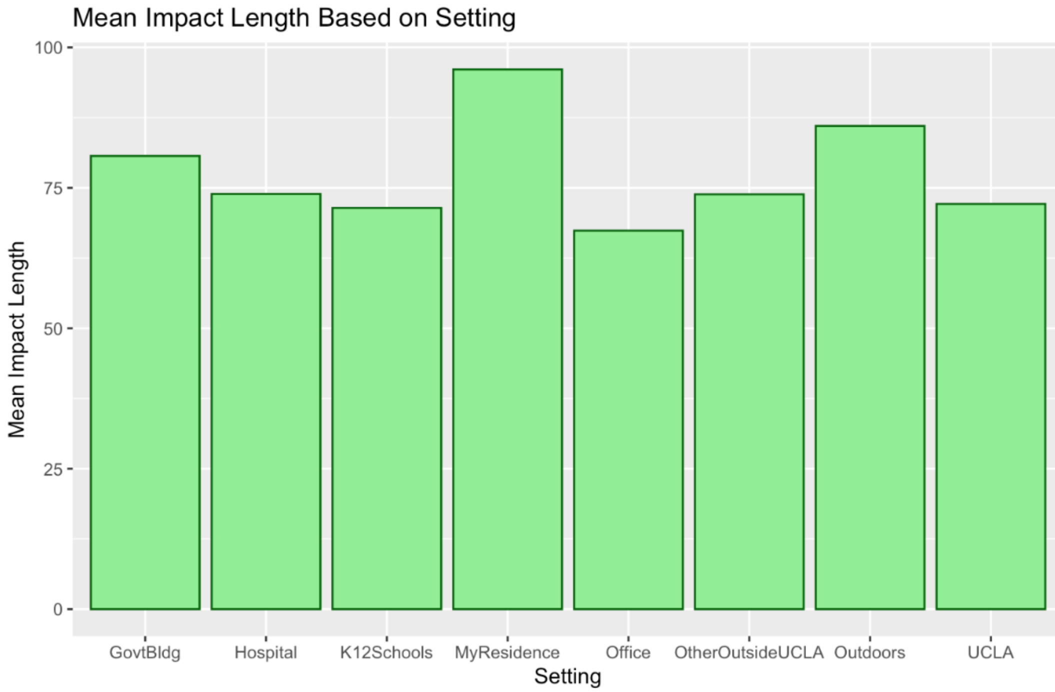
Anova

For our analysis we made the assumption that students who were more greatly impacted by their volunteer activity would write more in the ImpactOnYou_Text category. To determine if there was a difference in mean length of ImpactOnYou_text across different categories, we chose to use an ANOVA (analysis of variance test). However, one issue we ran into was that ANOVA assumes independent observations and because students could select more than one category, we could not assume independence. As the best way to work around this, we decided to remove students who selected more than one option for the setting and FindOut variables because they made up a smaller portion of total responses. However, this would not work for motive as a large portion of students selected more than one motive. Instead, for motive we decided to see if there was a relationship between length of ImpactOnYou_text and the number of motivations selected. Because we only looked at students who selected one option for setting and FindOut, our results may not be generalizable to the full population, however they still do offer insights.

Setting

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Setting	7	44794	6399	0.993	0.435
Residuals	620	3994623	6443		

The p-value for the anova of length of ImpactOnYou and setting is greater than 0.05 meaning there is no statistically significant difference in means across the setting categories.

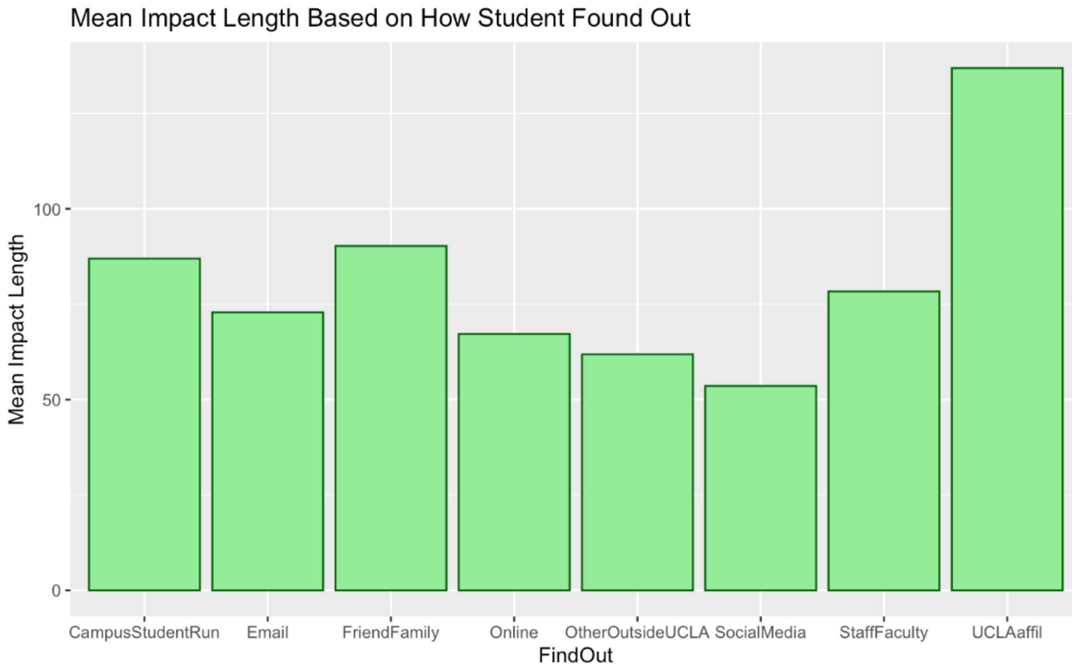


The plot of means also shows there are not significant differences in mean length across the different settings.

FindOut

```
              Df  Sum Sq Mean Sq F value   Pr(>F)    
FindOut        7  163295   23328   3.728 0.000593 ***
Residuals    544 3404506    6258                     
```

The p-value for the anova of length of ImpactOnYou and FindOut is less than 0.05 meaning there is a statistically significant difference in means across the FindOut categories.

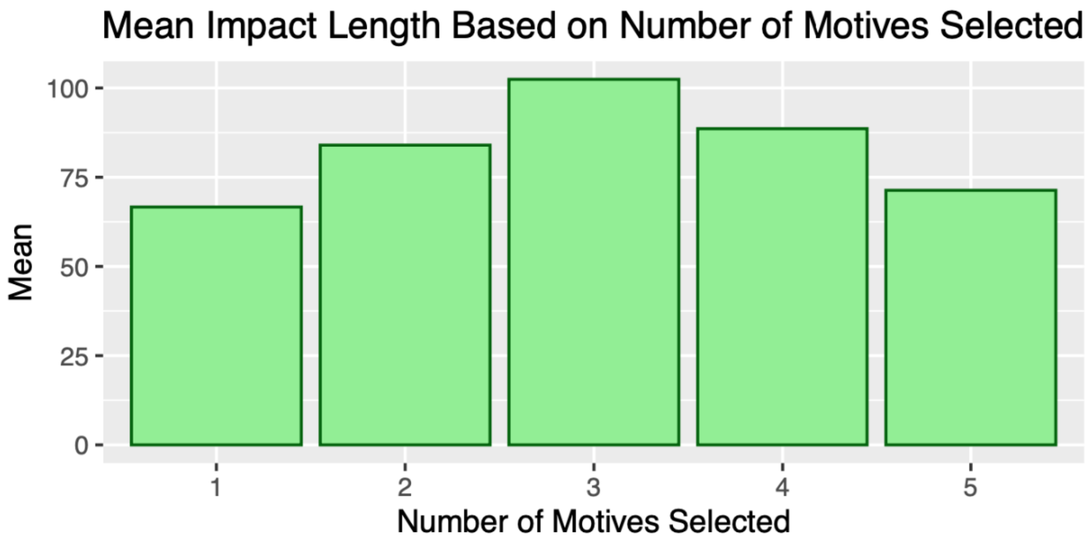


Based on out plot, we can see that on average students found out about their volunteer activity through a UCLA affiliated organization had the longest responses on average. If our assumptions are correct, this may indicate that these students were more impacted by their volunteer work than other students. Also, students who found out about their volunteer activity on social media had the shortest responses on average possibly indicating they were the least impacted by their volunteer work.

Motive

```
                             Df  Sum Sq Mean Sq   F value Pr(>F)    
as.factor(impact.length) 204 4405105   21594 2.835e+28 <2e-16 ***
Residuals                    502       0       0                     
```

The p-value for the anova of length of ImpactOnYou and Motive is less than 0.05 meaning there is a statistically significant difference in means based on the number of motives selected.



Our graph shows an increase in mean response for students who selected up to 3 motives and then a decrease in mean response for students who selected 4 or 5 motives. This was not what we expected to see. Instead, we hypothesized there would be an increase across the board. As students, we theorize that students who selected 1 option may have been trying to move through the survey as quickly as possible, but we also believe the same could be true for those who selected all 5 options. Students who selected all 5 may have also been selecting every option to move through the survey quickly. This would explain why those who selected only 3 would have the longest responses as they were focused more on selecting true options rather than all or none.

In summary, our study employed ANOVA analysis to examine the mean difference between the length of 'ImpactOnYou' and 'Setting,' 'findOut', and 'Motive'. The results indicated that there is no significant relationship between 'Setting' and 'ImpactOnYou'. However, we found a significant mean difference of the length of 'ImpactOnYou' based on the 'Motive' and 'FindOut' variables.

Conclusions

From analyzing the bi-grams and the directed graph, we can see that the population served in the volunteering activity appears multiple times in ImpactOnYou_Text, with multiple responses mentioning volunteering in high schools or interacting with patients, or even addressing food insecurity. This tells us that the impact of volunteering appears to be related to the population served, specifically with those who served high school students, patients, or those with food insecurity feeling a larger impact. By analyzing ImpactOnYou_Text, we also learned that many students described feeling good or happy while volunteering. Through tri-gram analysis we were also able to learn that many students were able to learn or realize something new and volunteering also helped some students make decisions about their future volunteering or their careers.

Our sentiment analysis graph shows that in general, volunteering is ultimately a positive experience for students, as positive sentiments and emotions of the ImpactOnYou_Text column consistently far outnumber the negative ones. Regardless of what motive or setting or how a student found out about the volunteering activity, this pattern holds, with trust, joy, and anticipation being the most common emotions.

The ANOVA analysis tells us the mean differences between the length of 'ImpactOnYou' and the variables 'Setting' or 'Motive'. The findings revealed a lack of significant association between 'Setting' and 'ImpactOnYou', whereas a notable mean difference was observed between 'Motive' and 'FindOut' and 'ImpactOnYou'.

Reccomendations

One recommendation to be made for further insights into the survey responses would be to employ natural language processing techniques such as similarity and co-occurrence matrices to operationalize impact and observe the appearances of similar phrases across the different text columns.

Another recommendation would be to do further analysis on how impact was related to tasks performed while volunteering and the goal/objective of the project or activity.

One recommendation in order to gain better insights and analysis is to change the format of the survey. It appeared that the wording of certain questions or the method in which the respondents were able to answer led to some confusion and contradictory answers. Several variables allowed students to select multiple settings, motives, etc as well as give them the option to select Other and write in their own option. This led to messy data that was difficult to decipher and clean. We would recommend that setting, motive, findout, etc. would each be one variable where the students selects one option for each rather than having the option of selecting multiple options or other.