141 Final Report

Mengxuan Sun,Yuqing Yang, Xinyi Wang, Chengcheng Xu

March 2, 2020

Contents

1   Abstract

In our study, we explored which variables affect the level of stress among STEM students and NON-STEM students. The data is obtained from a survey answered by UCLA students. Through the use of graphic visuals, text mining, and two sample t test we found that there is a significant difference between the stress level of STEM students and the stress level of NON-STEM students. Also, we studied how students' academic and daily/social life(relationship with roommates and dating life) experience affect their stress level, as well as the interaction of whether a student is international and whether or not the student is eager to find a job soon after graduation. Besides these, we also looked for some variables that we have found out that may also have some influence on our response value.

Based on what we have researched, we have concluded our findings. There are also some challenges during our work that we are concerned about, and we also think about what we can do better about this study in the future.

2   Background

Stress among college students can be overwhelming and can affect many areas of a student's life. Different majors lead to different levels of workload and cause different levels of stress. There are two main divisions of major selection nowadays: STEM and NON-STEM. STEM stands for science, technology, engineering, and math. In the survey, students report if they study STEM majors or not. Students also evaluate their own stress levels at the end of the questionnaire.

3   Clear Statement of The Problems To Be Solved

We are particularly interested in studying on the following statements in this report:
1. Overall findings from comparing STEM with NON-STEM majors.
2. What are some factors that may impact the level of stress for STEM and Non-STEM students?
3. Over findings from Q2.1-Q2.5 and Q3.1-Q3.24.
4. Overall findings from each of the 2 open-ended questions. (HelpFactor & InbibitFactor)

5. Hypothesis Testing

4   Variables of The Study

See attached code book

5   Data Cleaning

5.1   Drop Rows and Columns

[1] 846   51

- We use Excel to remove rows from 868 to 1128 because we found out this part of the data is from another survey.
- We decided to remove the columns "Academic", "SES", "Belonging", "Exclusion" because we focus on investigating the Stress as our outcome variable. We also dropped the rows containing missing values. The remaining dataset has 846 rows and 51 columns.

5.2   Data Types

- There are five numberic variables: GPA, HighGPA, SameEthic, OtherEthnic, and Stress.
- The rest of the variables are categorical.

6   Exploratory Data Analysis

6.1   Overall findings from comparing STEM with NON-STEM majors.

Ethnic of STEM or NON–STEM Major Students
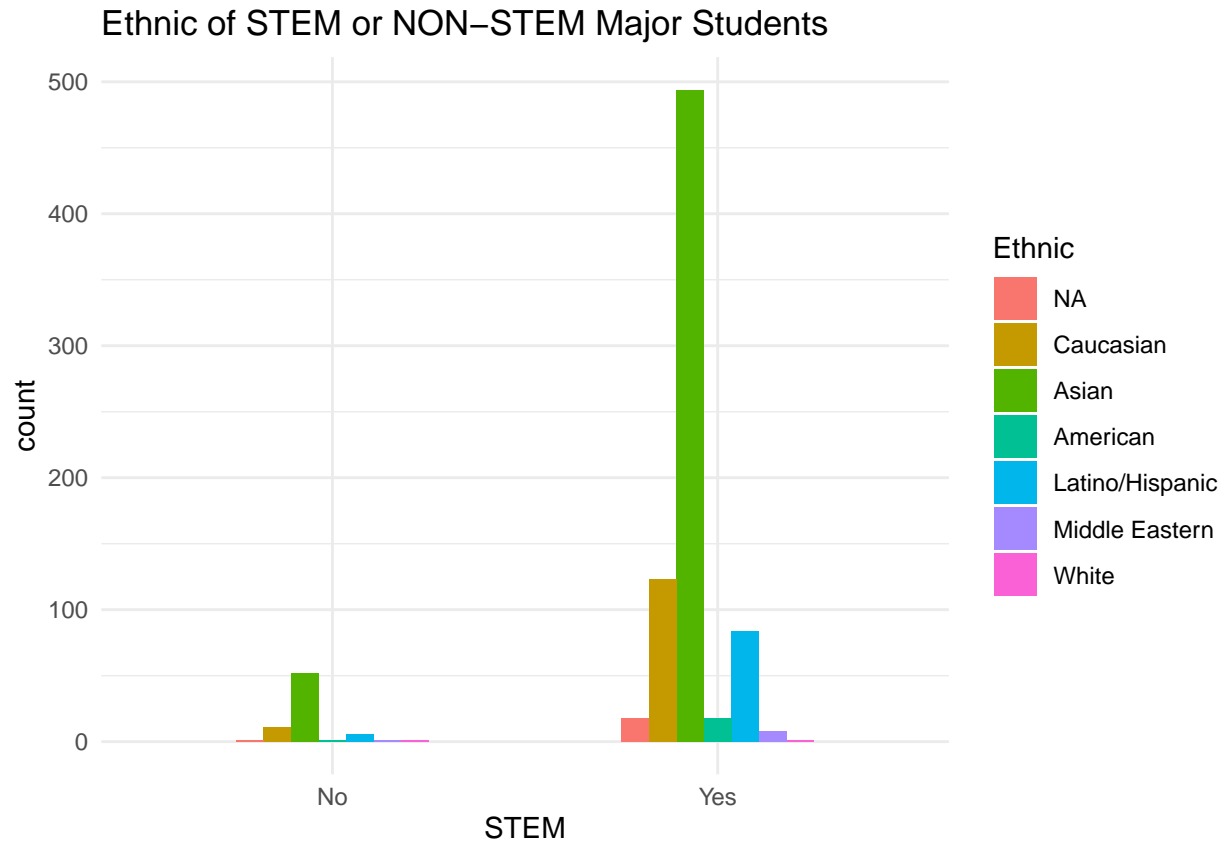
- For STEM and NON-STEM major students, Asian has the biggest population.

- There are much more STEM students than NON-STEM students in general, which shows more students tend to choose STEM majors over NON-STEM majors.

Mother's Education of STEM or NON–STEM Major Students
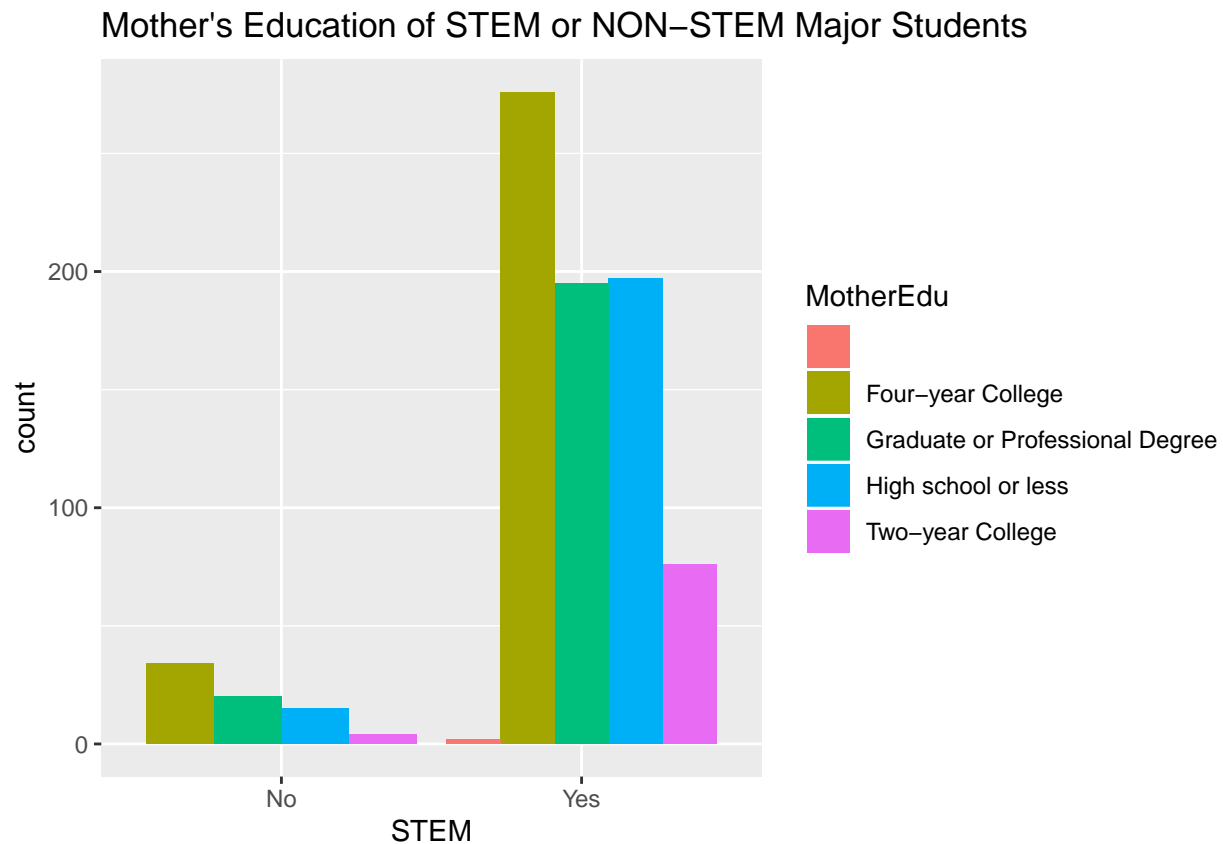
- For the vaiable of MotherEdu has 5 levels: High school or less, Two-year college, Four-year college, Graduate or professional degree, and No Response. There are some answered "yes", "no", and "Are you transfer from community college?", so we changed these answers to No Response.

- In general, the distribution of MotherEdu of NON-STEM students and the distribution of MotherEdu of STEM students are very similar. The most mothers are from Fou-year college and the fewest mothers are from two-year college.

Proportion of Father Education Level (STEM)
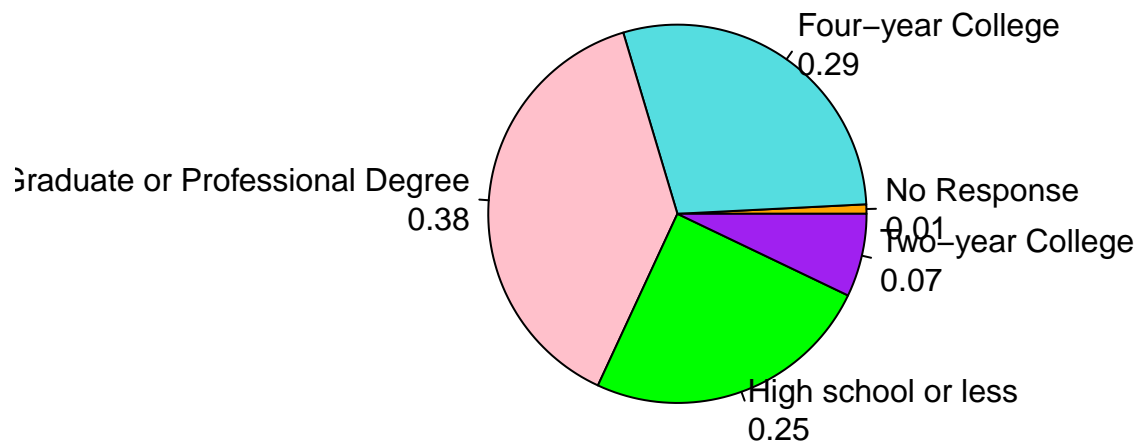
- We can see from the pie plot, most of stem majors students' FatherEdu (38%) are graduate or professional degree.

- However, most of nonstem majors students' father education (40%) are Four-year college degree.

STEM	same_ethnic	n
No	30.72857	73
Yes	30.78412	746

STEM	other_ethnic	n
No	17.24286	73
Yes	22.12117	746

- After we drop one ourlier, STEM students and NON-STEM students have about the same amount number of friends from their own ethnic. And STEM major students have more friends from other ethnic than NON-STEM students.

Barplot of Orientation of STEM students
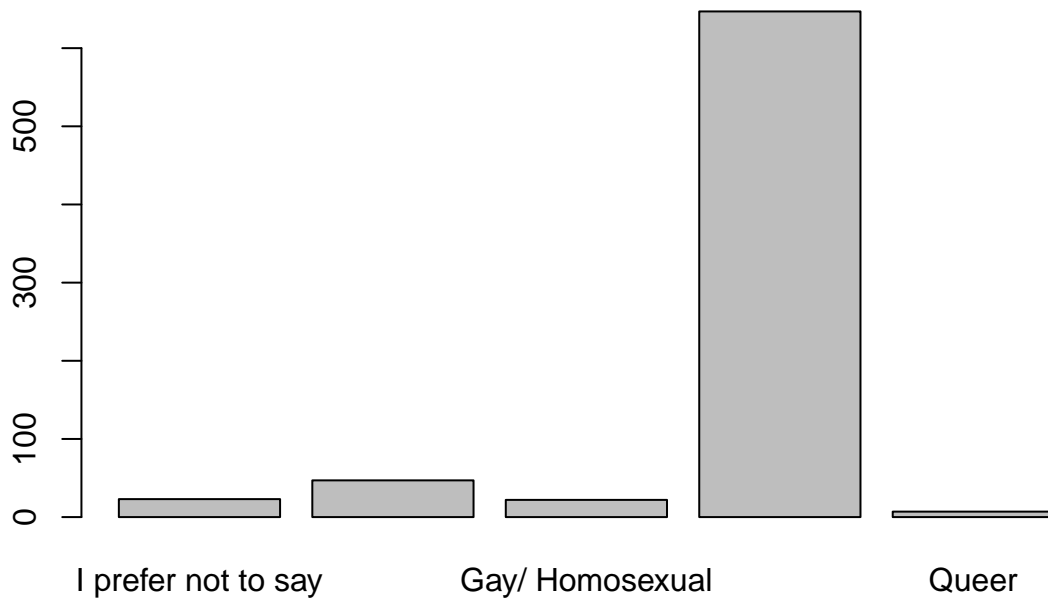
Barplot of Orientation of NON−STEM students

- The sexual orientations of the majority of people from STEM or NON-STEM majors are heterosexual.

6.2 What are some factors that may impact the level of stress for STEM and NON-STEM students?

Mean Stress Level of STEM/NON–STEM Students Among Different Standir
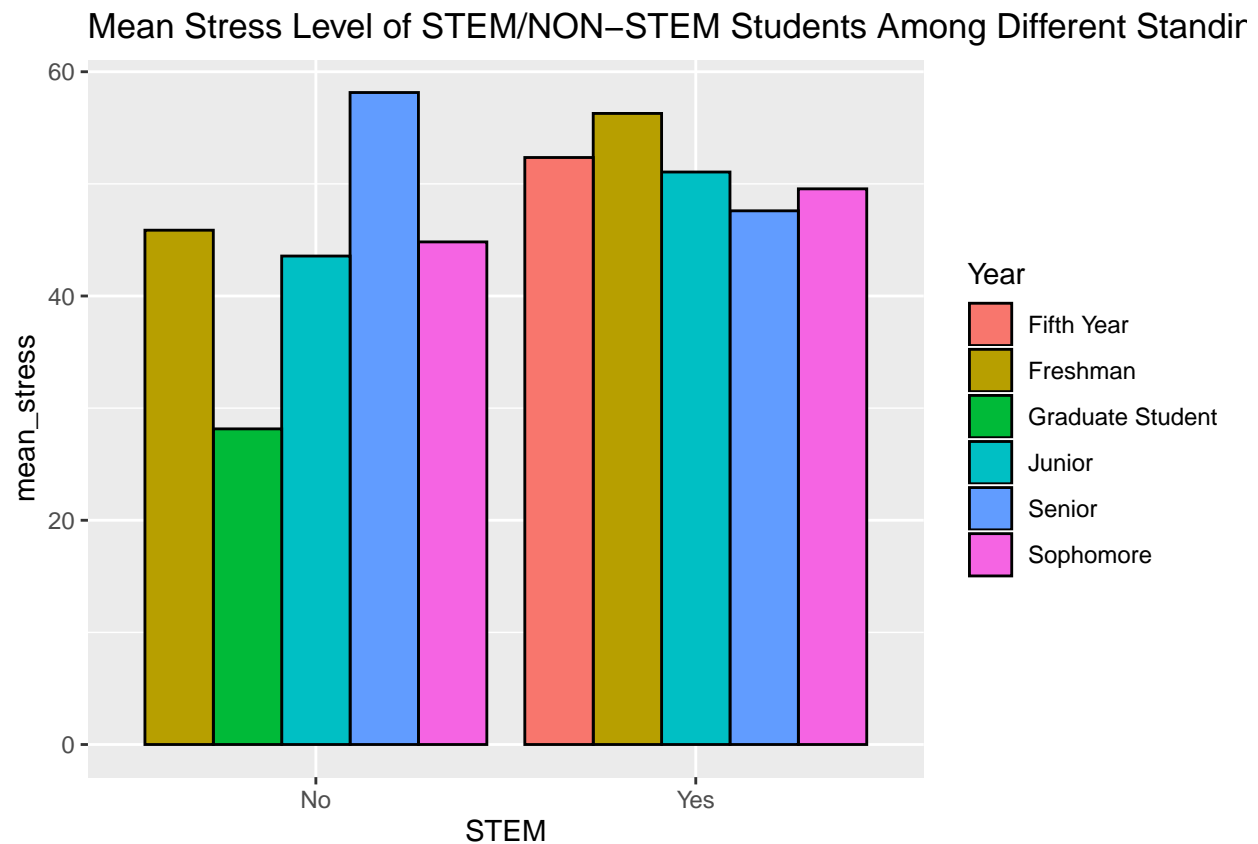
- For NON-STEM students, graduate students have the smallest stress level. Senior has the largest stress level. Freshman, junior, and sophmores have similar stress levels.

- For STEM students, freshman students have the biggest stress level and fifth-year students have the second biggest stress level. Junior, senior and sophomore have similar stress levels.

Average Stress Level of STEM/NON−STEM students with Different Expecta
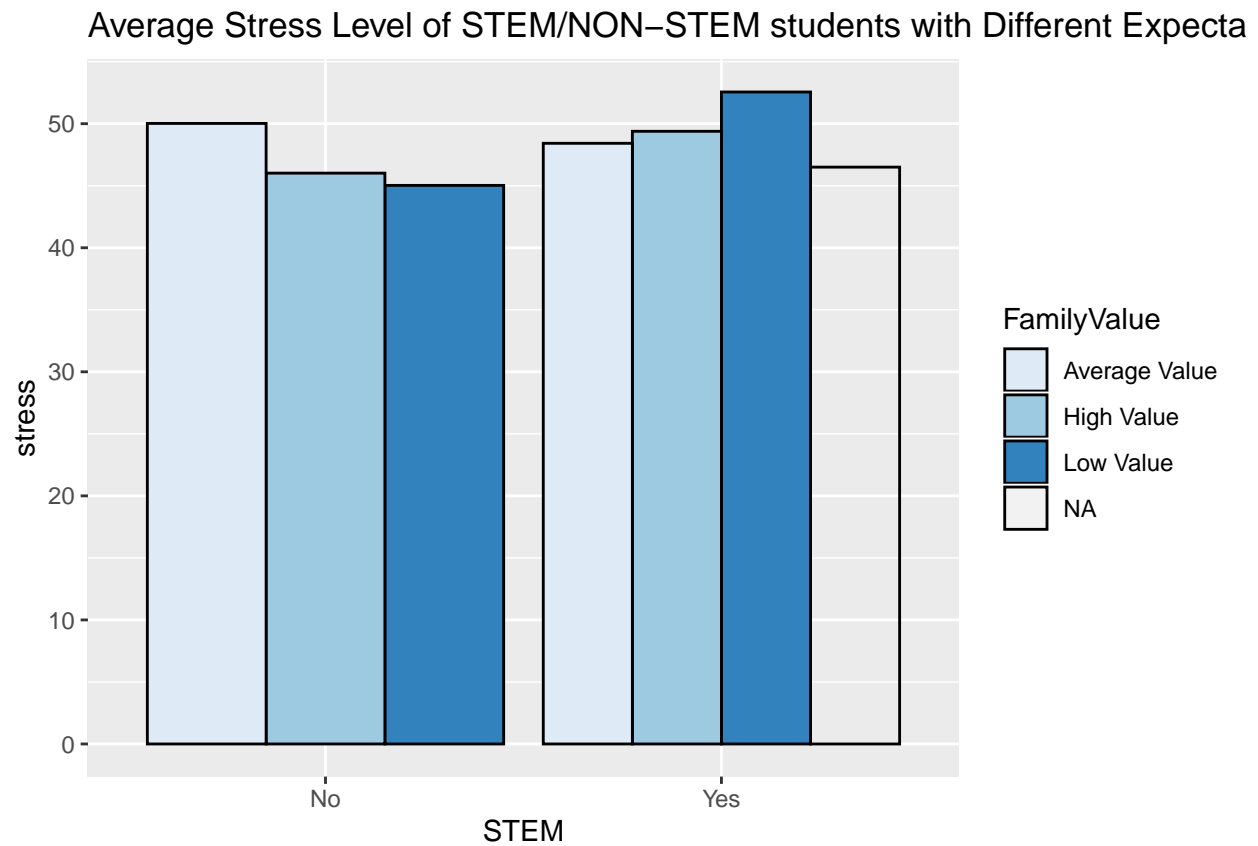
- For NON-STEM students, when their families have low expectations for them, they tend to have lower stress.

- But for STEM students, when their families have low expectations for them, they tend to have high stress.

Average Stress of STEM or NON–STEM students v.s. Transfer or Not
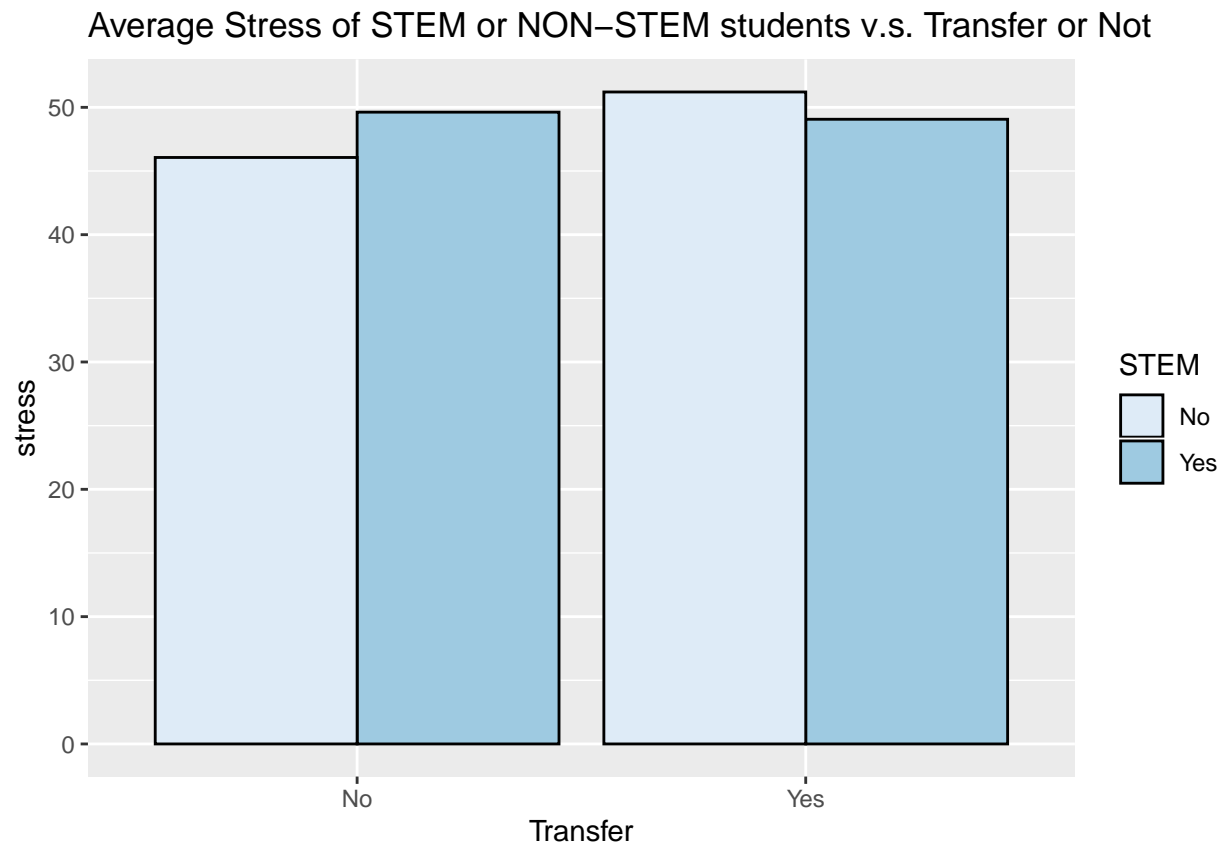
- For STEM or NON-STEM students, whether they are transfer students or not does not affect their stress level that much.

Average Stress of STEM or NON–STEM students v.s. International or Not
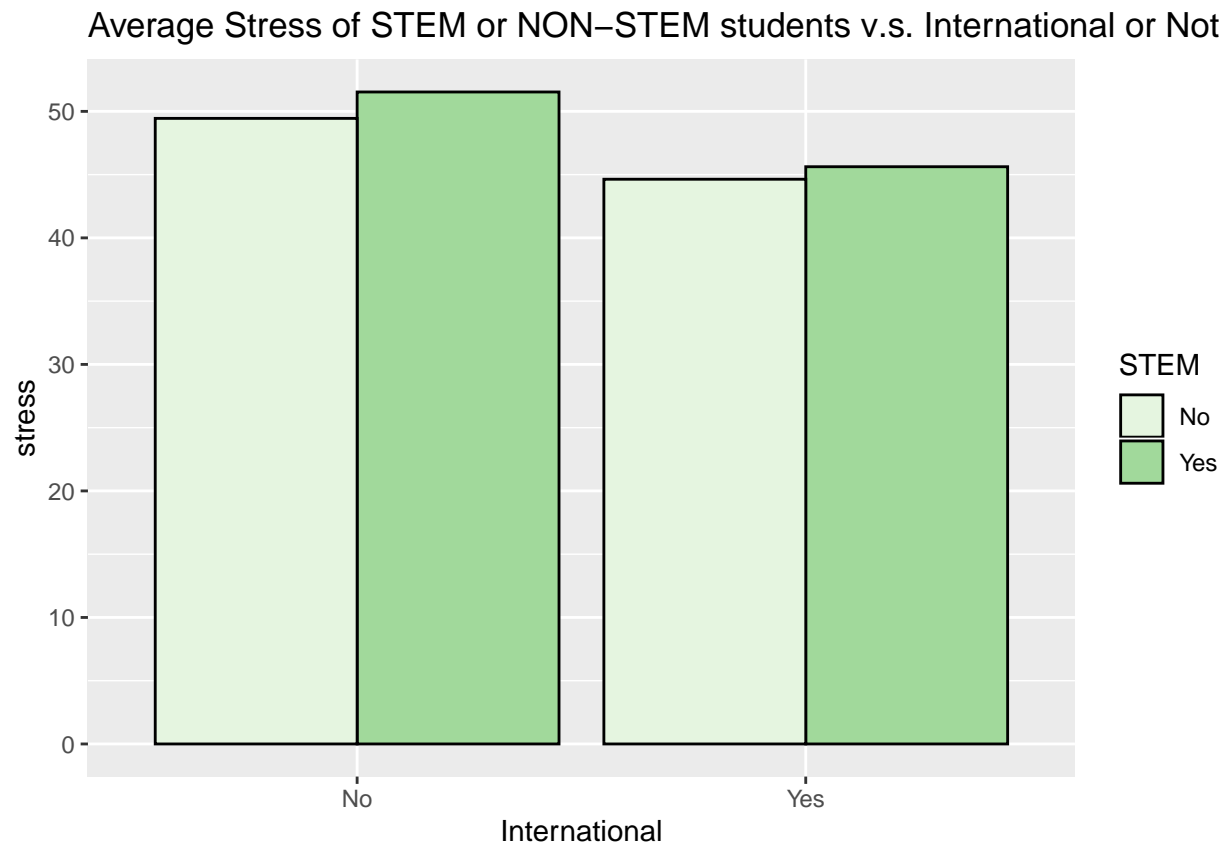
- From this graph, internation students tend to have lower stress than domestic students. And STEM students always have higher stress levels than NON-STEM students.

Stress Level in Different GPA Range of STEM Students
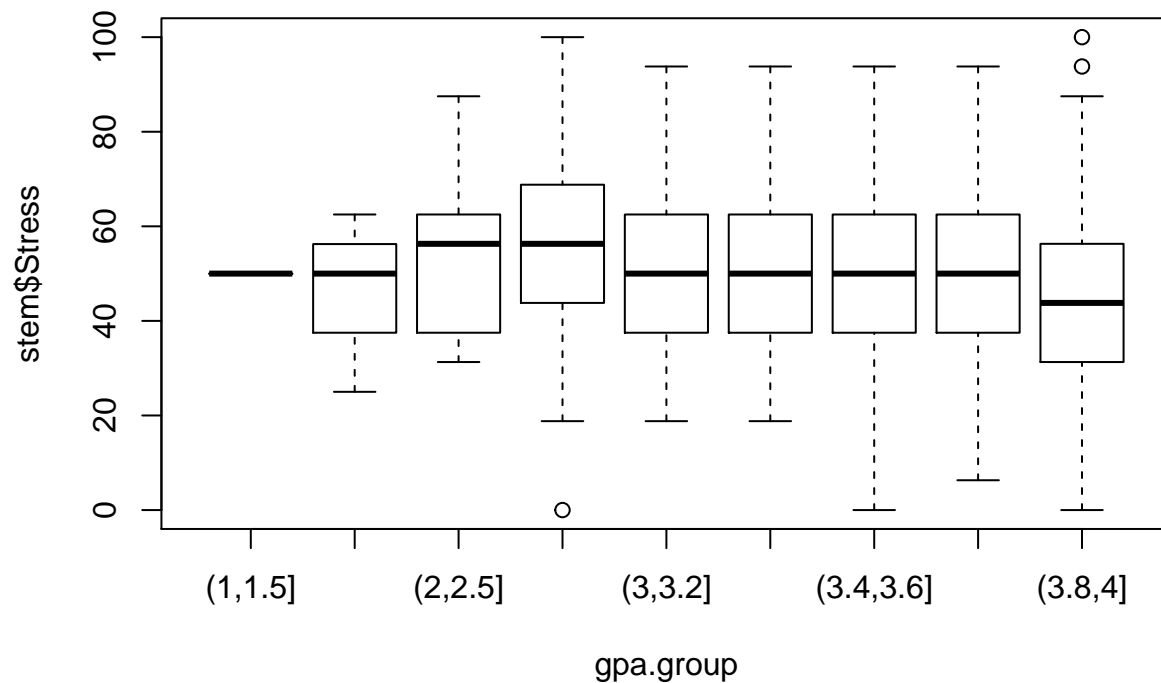
Stress Level in Different GPA Range of NON–STEM Students
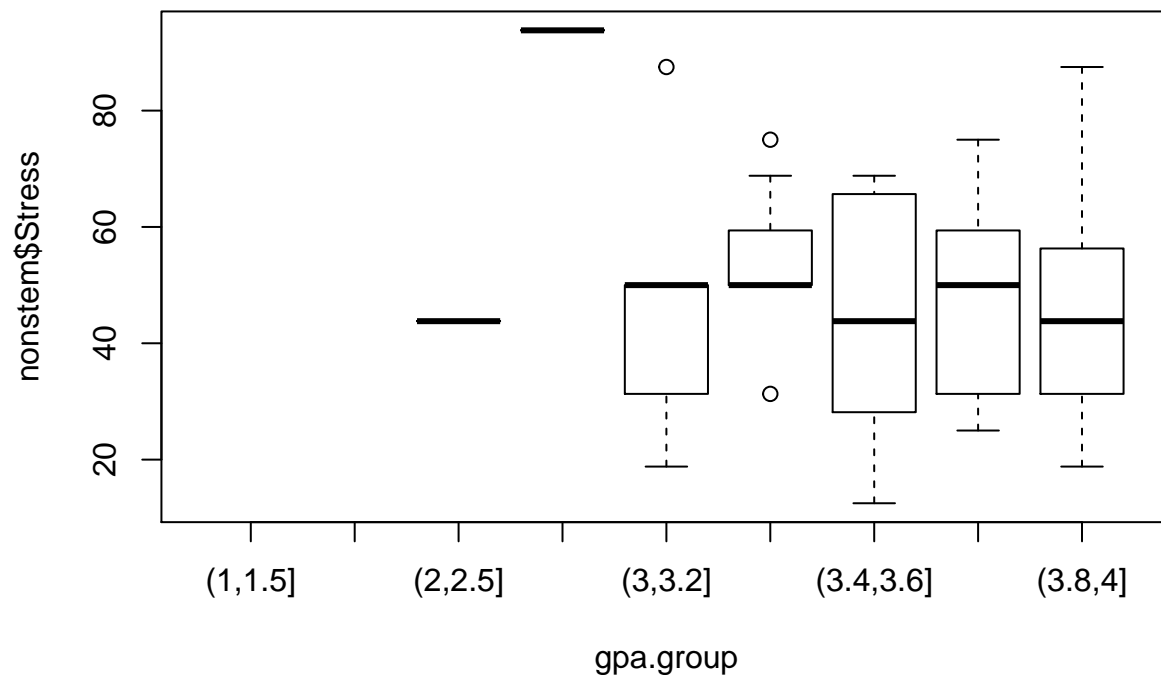
- The average value of GPA from 3.0 to 4.0 is relatively average, and the pressure average of 3.8 to 4.0 is slightly lower. Average pressure is slightly higher when students in the GPA range 2.0 to 3.0. Students in this range will be more likely to feel the pressure, probably because they worry that their GPA is not enough to graduate.

- The variation of NON-STEM students' stress levels are a lot higher than STEM students' stress levels. NON-STEM students have the highest stress level when their gpa is between 2.5 to 3.0.

6.3   Over findings from Q2.1-Q2.5 and Q3.1-Q3.24.
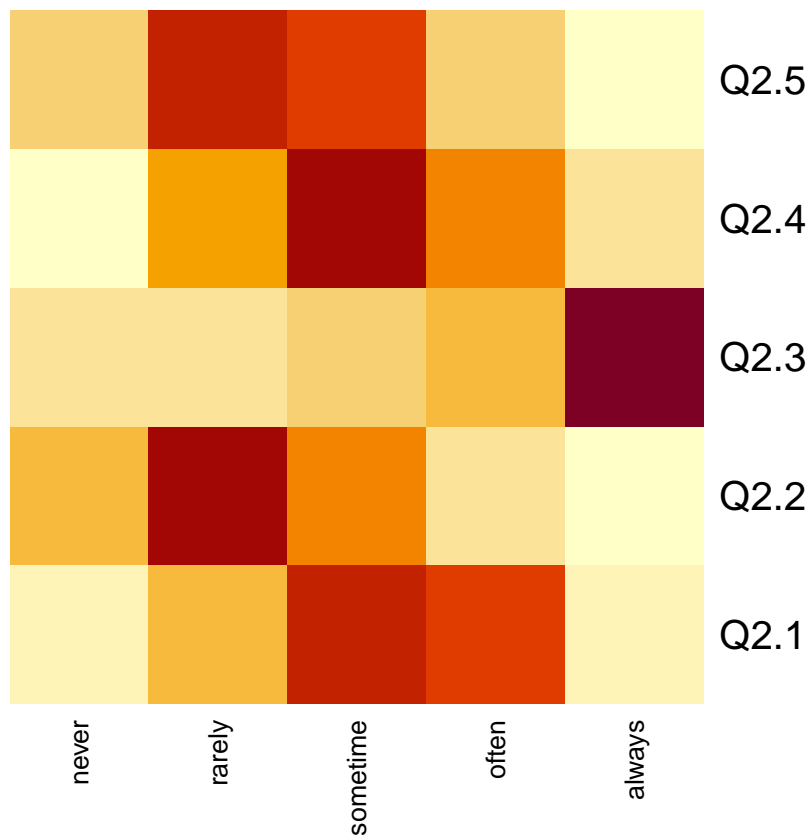
- From this heatmap, we can tell that most people sometimes get enough sleep. Most people ralely miss classes. Most people always do their homework. Most people exercise sometimes. Most people ralely go to office hours.
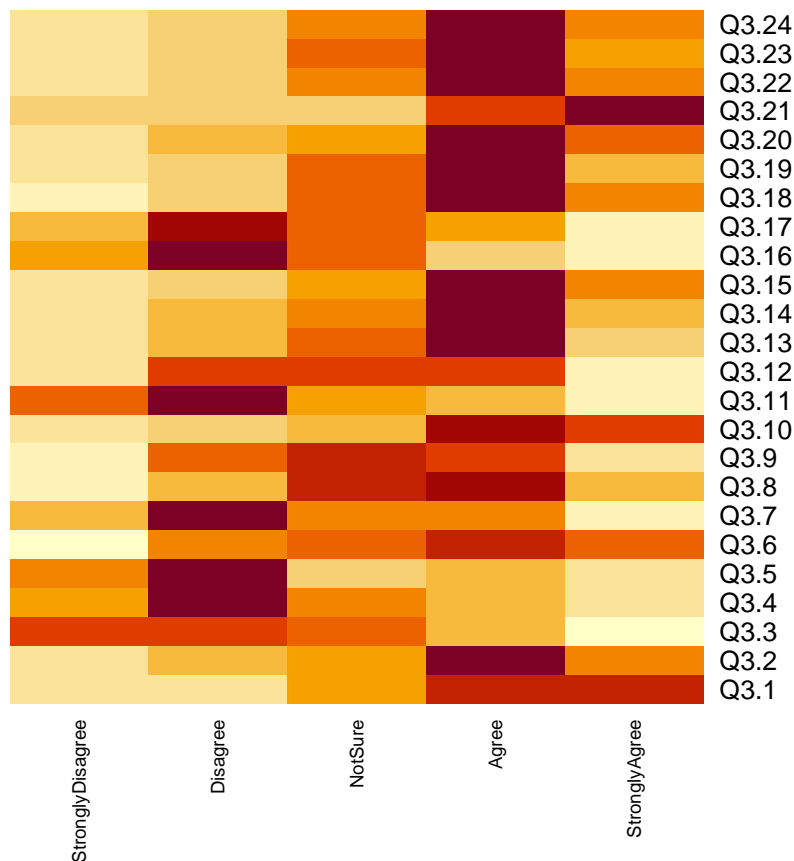
- From Q3.3 and Q3.4, most people disagree that their families have struggled when they grow up and they claim that they don't have to work to pay for their expenses. This is interesting because it shows that people who go to UCLA are less likely to come from a poor background.

- In terms of social life, (Q3.7) most people disagree that they make friends from their own majors. And (Q3.11) most people don't have trouble with roommates. From Q3.23, most people feel happy at UCLA.

6.4   Overall findings from each of the 2 open-ended questions. (HelpFactor & InbibitFactor)

- The question is asked for this variable is: Please describe the factors that have helped you succeed academically, socially, emotionally, or otherwise at UCLA? It would be great if you can provide us with examples or tell us a story about something that happened.

- The most frequent words show on student's responses are Friends, Helped, Support, Help, Family, ucla, etc.

- Friends and family usually play big roles in supporting everyone. They provide emotionally, financially, and academically support for ucla students.

- The question is asked for this variable is: Please describe the factors that have inhibited your success academically, socially, emotionally, or otherwise at UCLA? It would be great if you can provide us with examples or tell us a story about something that happened.

- Top frequent words based on responses are ucla, classes, students, people, time, etc.

- Surprisely, classes, ucla, people, students are the keywords of inhibiting students succeeding. It would be interesting to do more research or send out more surveys on this topic.

7   Hypothesis Testing

```
##
##   Welch Two Sample t-test
##
## data:  Stress by STEM
## t = -0.88745, df = 88.531, p-value = 0.3772
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
## -6.410306  2.452284
## sample estimates:
##  mean in group No mean in group Yes
##         47.54110          49.52011
```

- H0: The mean stress level among STEM and NON-STEM students are the same.

- HA: The mean stress level among STEM and NON-STEM students are different.

- Based on the results of two sample t-test, the p value is 0.3772 which is greater than 0.05. Therefore, we reject H0 and accept HA, which is there is a significant difference between the stress level of STEM students and the stress level of NON-STEM students.

8   Variables Significance Test

- In order to convinient our variables selecting and our model study, we decide to put our response variable stress in to two levels, below median and above median.

	x
Below Median	327
Above Median	236
NA's	11

- Also, we put the questions that we concern within Question 2 and 3 parts into three levels.

- After resetting the variables, we do the logistic regression model, check for pseudo R square and confidence interval and check the effect plots to check variables' significance and interaction term.

- Our main focus would be on students' academic stress, friendship(roommates) stress, and emotion(dating) stress, also we would study for the interaction of International(or not) and Q3.6(eager to work or not) since we are international students and some of us would be getting jobs very soon.

8.1 First model: add Q3.10(I experience anxiety about academics.)

Predictors	Stress.factor		
	Odds Ratios	CI	p
(Intercept)	0.37	$0.03 - 3.85$	0.402
Q2.1 [Rarely]	1.87	$1.10 - 3.19$	**0.021**
Q2.1 [Sometimes]	1.13	$0.71 - 1.78$	0.612
Year [Junior]	1.40	$0.26 - 7.02$	0.676
Year [Senior]	1.22	$0.23 - 6.08$	0.805
Year [Sophomore]	1.13	$0.20 - 5.83$	0.887
EFL [Yes]	1.31	$0.85 - 2.04$	0.226
Transfer [Yes]	1.20	$0.60 - 2.38$	0.611
Q2.4 [Rarely]	2.01	$1.21 - 3.36$	**0.007**
Q2.4 [Sometimes]	1.01	$0.63 - 1.62$	0.975
GPA	1.11	$0.67 - 1.88$	0.695
SameEthnic	1.00	$0.99 - 1.00$	0.118
Q3.6 [Disagree]	0.54	$0.33 - 0.86$	**0.011**
Q3.6 [Not Sure]	0.95	$0.55 - 1.66$	0.869
Q3.7 [Disagree]	1.09	$0.68 - 1.77$	0.715
Q3.7 [Not Sure]	1.38	$0.75 - 2.54$	0.299
Q3.10 [Disagree]	0.02	$0.00 - 0.08$	**<0.001**
Q3.10 [Not Sure]	0.08	$0.03 - 0.17$	**<0.001**
Q3.24 [Disagree]	1.98	$1.03 - 3.95$	**0.045**
Q3.24 [Not Sure]	1.41	$0.86 - 2.33$	0.177
Observations	563		
R^2 Tjur	0.261		

- Based on the logistic regression summary report, we can notice that while including Q3.10, this variable has the most significant level. Besides, Q2.4(I exercise.), Q2.1Rarely(I get enough sleep.),Q3.6 Disagree(I need to graduate and get a job as soon as possible.) and Q3.24Disagree(My experiences at UCLA have helped me to interact with diverse groups.) also have some significance.

- Now we check the pseudo R square. The value is between 0.2-0.4, which is considered fitted well for our model. So we can say that those variables we have used in our model "test1" are overall good.

- Here we check for the confidence interval(CI) for the variables. From the CI, we can interpret the followings:

- 1. The effects of Q2.1Rarely, Q2.4Rarely, Q3.6Disagree, Q3.10Disagree,

Q3.10Not Sure and Q3.24Disagree do have a significant influence on feeling stress, since these variables have CI that don't include 1.

- 2. For Q2.1(I get enough sleep.), we can conclude that, for those who answered "rarely" have 87% more chance than feeling stressed than those who have enough sleep.

- 3. For Q2.4(I exercise.) we can conclude that, for those who answered "rarely" have 2.01 times more chance of feeling stressed than those who do exercise.

- 4. For Q3.6(I need to graduate and get a job as soon as possible.), we can conclude that, for those who answered "Disagree" have 0.46 times less chance than feeling stressed than those who don't worry about getting a job as soon as possible.

- 5. For Q3.10(I experience anxiety about academics.), we can conclude that, for those who answered "Disagree" have 0.98 times less chance than feeling stressed and for those who answered "Not Sure" have 0.92 times less chance than feeling stressed than those who don't anxiety.

- 6. For Q3.24(My experiences at UCLA have helped me to interact with diverse groups.), we can conclude that, for those who answered "Disagree" have 1.98 times more chance of feeling stressed than those who think UCLA have helped them to interact with diverse groups.
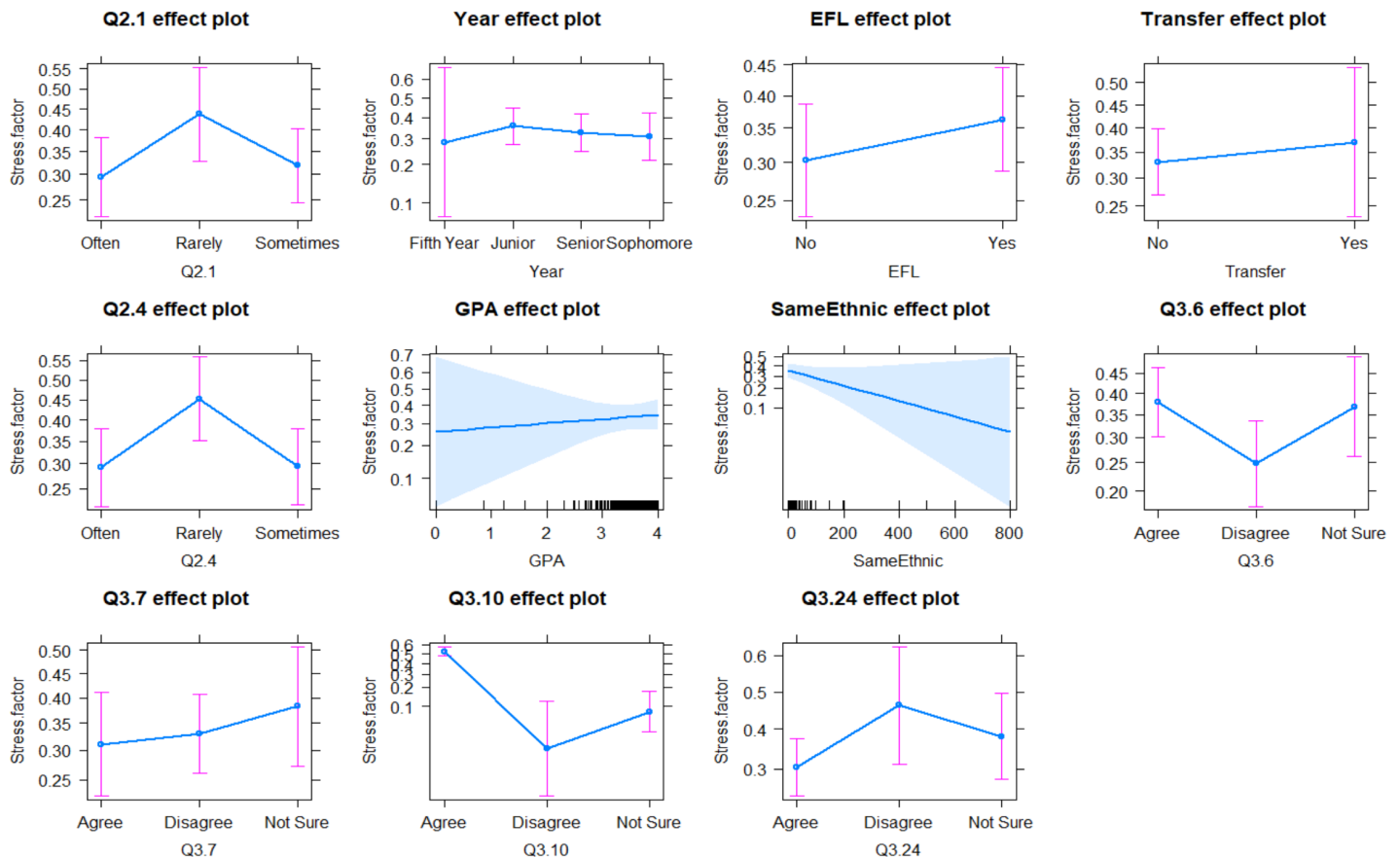
- In these plots we can interpret that:

- Those who have rarely sleep, or at their junior year, or their English are not their first language, or transfer students, or rarely excercise, or have higher GPA(in the range of 0 to 4), or have less same ethnic friends(in the range of 0 to 800), or experience anxiety about academics, or think their experiences at UCLA have not helped them to interact with diverse groups are having higher stress feeling.

- Those who think they don't need to graduate and get a job as soon as possible are having less stress feeling.

- Whether or not that most of friends are from their major does not vary their stress level overall.

- For Q3.10(I experience anxiety about academics.), those who have answered "Agree" are obviously have much higher stress feelings than who answered "Disagree".

8.2 Second model: add 3.11(I experience anxiety about getting along with my roommates.)

	Stress.factor		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	3.68	$0.28 - 46.33$	0.315
Q2.1 [Rarely]	2.11	$1.16 - 3.85$	**0.014**
Q2.1 [Sometimes]	1.09	$0.65 - 1.83$	0.744
Year [Junior]	1.58	$0.27 - 10.33$	0.613
Year [Senior]	0.87	$0.15 - 5.71$	0.882
Year [Sophomore]	1.43	$0.23 - 9.69$	0.703
EFL [Yes]	1.59	$0.97 - 2.63$	0.070
Transfer [Yes]	1.67	$0.76 - 3.63$	0.198
Q2.4 [Rarely]	2.17	$1.23 - 3.87$	**0.008**
Q2.4 [Sometimes]	0.94	$0.55 - 1.62$	0.823
GPA	0.94	$0.55 - 1.60$	0.807
SameEthnic	1.00	$0.99 - 1.00$	0.537
Q3.6 [Disagree]	0.51	$0.29 - 0.89$	**0.019**
Q3.6 [Not Sure]	0.68	$0.37 - 1.21$	0.190
Q3.7 [Disagree]	1.72	$0.98 - 3.07$	0.062
Q3.7 [Not Sure]	1.40	$0.71 - 2.80$	0.335
Q3.11 [Disagree]	0.02	$0.01 - 0.05$	**<0.001**
Q3.11 [Not Sure]	0.15	$0.07 - 0.31$	**<0.001**
Q3.24 [Disagree]	1.93	$0.94 - 3.98$	0.074
Q3.24 [Not Sure]	1.24	$0.69 - 2.19$	0.471
Observations	563		
R^2 Tjur	0.422		

- Based on the logistic regression summary report, we can notice that while including Q3.11, this variable has the most significant level. Besides, Q2.4(I exercise.), Q2.1Rarely(I get enough sleep.),Q3.6Disagree(I need to graduate and get a job as soon as possible.) also have some significance.

- Now we check the pseudo R square. The value is between 0.2-0.4, which is considered fitted well for our model. So we can say that those variables we have used in our model "test2" are overall good.

- Here we check for the confidence interval(CI) for the variables. From the CI, we can interpret the followings:

- 1. The effects of Q2.1Rarely, Q2.4Rarely, Q3.6Disagree, Q3.11Disagree, and Q3.11Not Sure do have a significant influence on feeling stress, since these variables have CI that don't include 1.

- 2. For Q2.1(I get enough sleep.), we can conclude that, for those who answered "rarely" have 2.11 times more chance of feeling stressed than those who have enough sleep.

- 3. For Q2.4(I exercise.), we can conclude that, for those who answered "rarely" have 2.17 times more chance of feeling stressed than those who do exercise.

- 4. For Q3.6(I need to graduate and get a job as soon as possible.), we can conclude that, for those who answered "Disagree" have 0.49 times less chance than feeling stressed than those who don't worry about getting a job as soon as possible.

- 5. For Q3.11(I experience anxiety about academics.), we can conclude that, for those who answered "Disagree" have 0.98 times less chance than feeling stressed and for those who answered "Not Sure" have 0.85 times less chance than feeling stressed than those who don't anxiety.
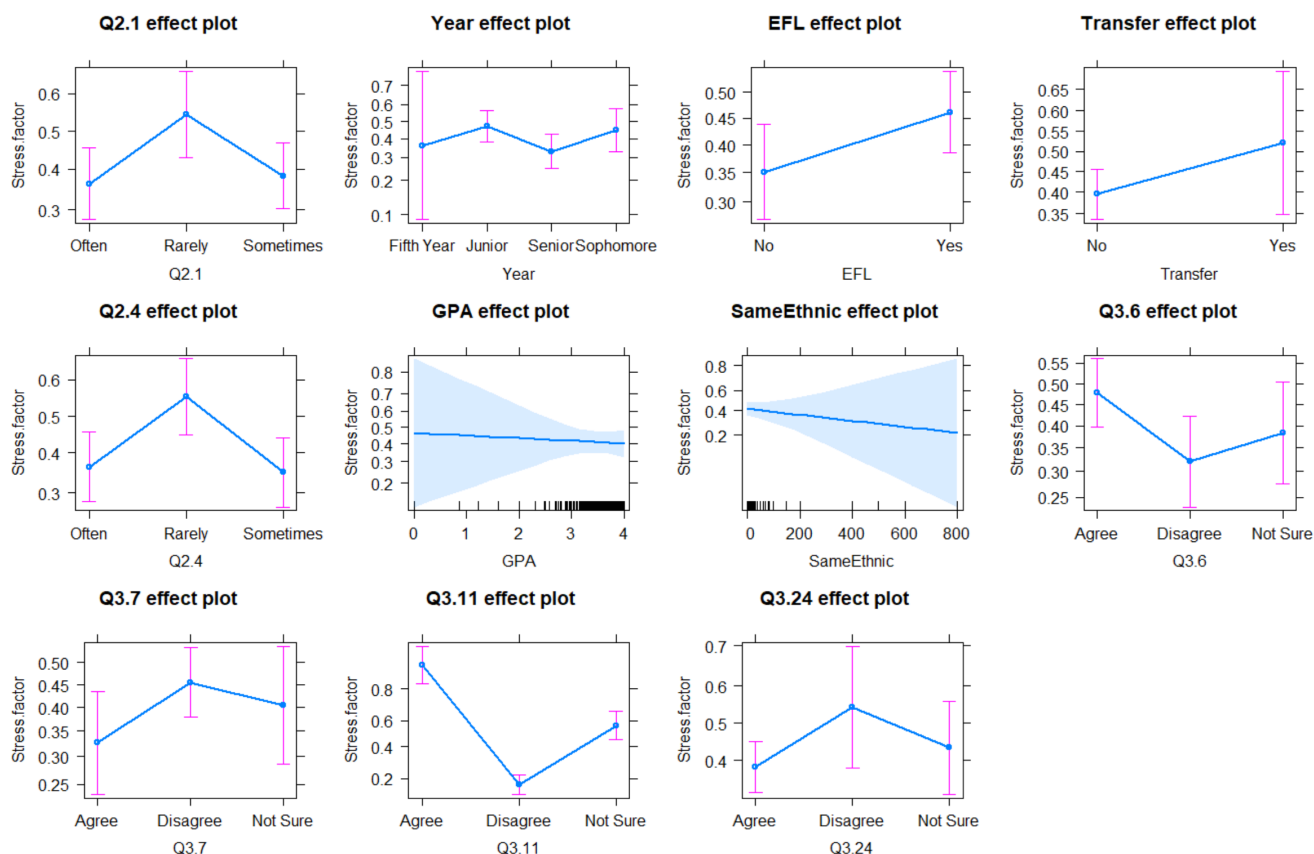
- In these plots we can interpret that:

- Most findings are similiar as what we have found above(in test1).

- For "Q3.6: I need to graduate and get a job as soon as possible.", those who have answered "Agree" have much higher stress feelings than others.

- For "Q3.7: Most of my friends are from own major.", those who have answered "Agree" are obviously have much lower stress feelings than who answered "Disagree".

- For "Q3.11: I experience anxiety about getting along with my room-mates.", those who have answered "Agree" are obviously have much higher stress feelings than who answered "Disagree".

## 8.3   Third model: add Q3.12(I experience anxiety with my dating life.)

	Stress.factor		
Predictors	*Odds Ratios*	*CI*	*p*
(Intercept)	0.77	$0.06 - 9.95$	0.842
Q2.1 [Rarely]	2.56	$1.38 - 4.80$	**0.003**
Q2.1 [Sometimes]	1.50	$0.89 - 2.53$	0.130
Year [Junior]	1.77	$0.29 - 10.28$	0.527
Year [Senior]	1.18	$0.19 - 6.77$	0.857
Year [Sophomore]	1.64	$0.26 - 9.94$	0.592
EFL [Yes]	1.17	$0.70 - 1.95$	0.560
Transfer [Yes]	1.49	$0.67 - 3.32$	0.323
Q2.4 [Rarely]	2.76	$1.52 - 5.07$	**0.001**
Q2.4 [Sometimes]	1.38	$0.80 - 2.40$	0.249
GPA	1.10	$0.64 - 1.92$	0.735
SameEthnic	1.00	$0.99 - 1.00$	0.440
Q3.6 [Disagree]	0.55	$0.32 - 0.96$	**0.037**
Q3.6 [Not Sure]	0.63	$0.34 - 1.14$	0.130
Q3.7 [Disagree]	1.10	$0.63 - 1.92$	0.738
Q3.7 [Not Sure]	1.44	$0.72 - 2.88$	0.303
Q3.12 [Disagree]	0.03	$0.01 - 0.05$	**<0.001**
Q3.12 [Not Sure]	0.12	$0.07 - 0.21$	**<0.001**
Q3.24 [Disagree]	2.49	$1.22 - 5.18$	**0.013**
Q3.24 [Not Sure]	1.56	$0.87 - 2.79$	0.134
Observations	563		
R^2 Tjur	0.438		

- Based on the logistic regression summary report, we can notice that

Q3.12, Q2.4Rarely(I exercise.), Q2.1Rarely(I get enough sleep.),Q3.6Disagree(I need to graduate and get a job as soon as possible.), and Q3.24Disagree(My experiences at UCLA have helped me to interact with diverse groups.) have significance while doing our model prediction.

- Now we check the pseudo R square. The value is between 0.2-0.4, which is considered fitted well for our model. So we can say that those variables we have used in our model "test3" are overall good.

- Here we check for the confidence interval(CI) for the variables. From the CI, we can interpret the followings:

- 1. The effects of Q2.1Rarely, Q2.4Rarely, Q3.6Disagree, Q3.12Disagree, Q3.12Not Sure, and Q3.24Disagree do have a significant influence on feeling stress, since these variables have CI that don't include 1.

- 2. For Q2.1(I get enough sleep.), we can conclude that, for those who answered "rarely" have 2.56 times more chance of feeling stressed than those who have enough sleep.

- 3. For Q2.4(I exercise.), we can conclude that, for those who answered "rarely" have 2.76 times more chance of feeling stressed than those who do exercise.

- 4. For Q3.6(I need to graduate and get a job as soon as possible.), we can conclude that, for those who answered "Disagree" have 0.45 times less chance than feeling stressed than those who don't worry about getting a job as soon as possible.

- 5. For Q3.12(I experience anxiety about academics.), we can conclude that, for those who answered "Disagree" have 0.97 times less chance than feeling stressed and for those who answered "Not Sure" have 0.88 times less chance than feeling stressed than those who don't anxiety.

- 6. For Q3.24(My experiences at UCLA have helped me to interact with diverse groups.), we can conclude that, for those who answered "Disagree" have 2.49 times more chance of feeling stressed than those who think UCLA has helped them to interact with diverse groups.
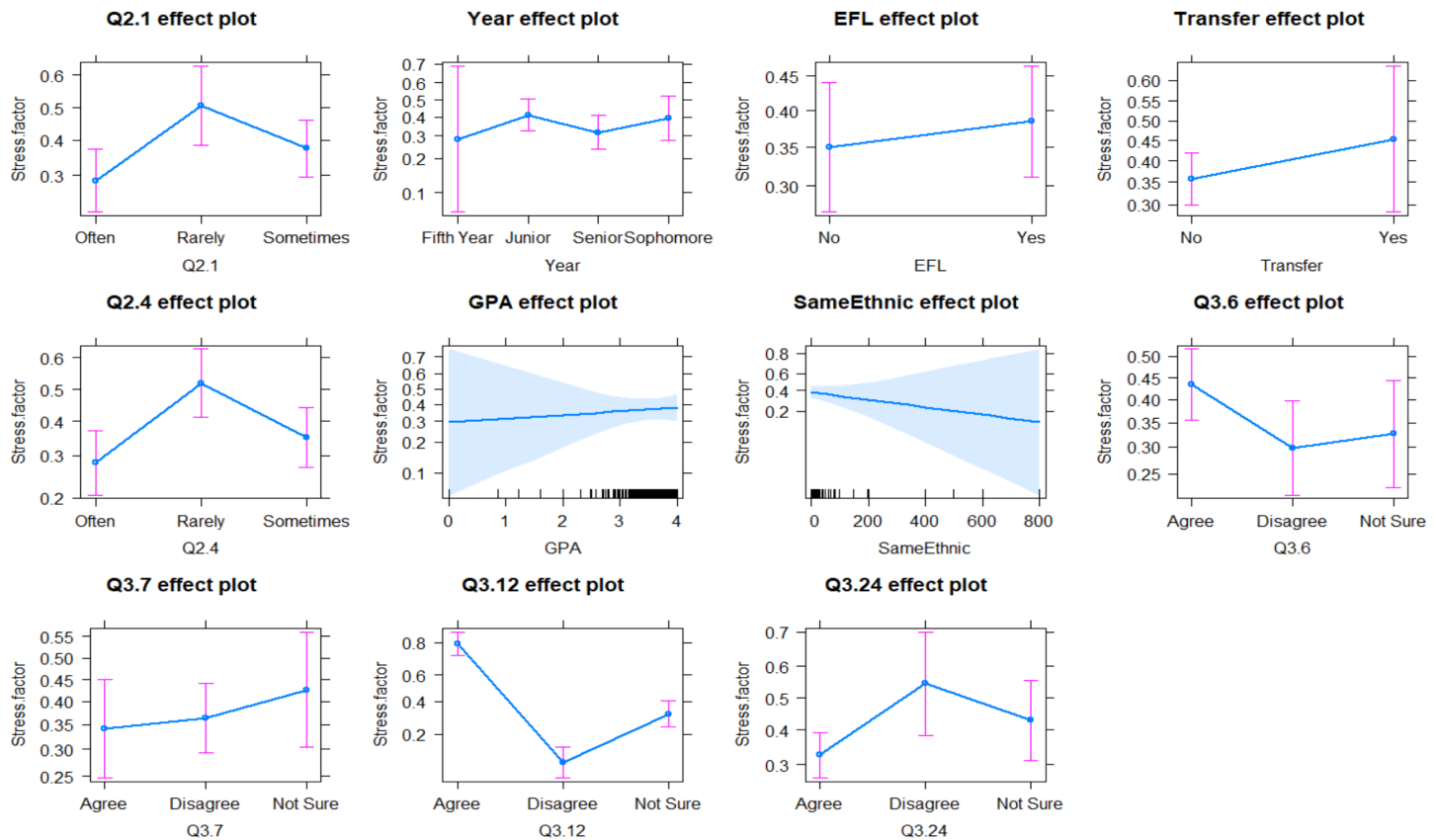
- In these plots we can interpret that:

- Most findings are similiar as what we have found above(in test2).

- Whether or not that most of friends are from their major does not vary their stress level overally.

- For "Q3.12: I experience anxiety with my dating life.", those who have answered "Agree" are obviously have much higher stress feelings than who answered "Disagree".

8.4   Interaction plot between International and Q3.6

	Stress.factor		
Predictors	Odds Ratios	CI	p
(Intercept)	1.06	$0.82 - 1.39$	0.639
International [Yes]	0.84	$0.46 - 1.52$	0.568
Q3.6 [Disagree]	0.63	$0.39 - 1.01$	0.055
Q3.6 [Not Sure]	0.63	$0.37 - 1.05$	0.077
International [Yes] * Q3.6 [Disagree]	0.36	$0.14 - 0.92$	**0.035**
International [Yes] * Q3.6 [Not Sure]	0.85	$0.30 - 2.37$	0.761
Observations	563		
R^2 Tjur	0.051		

- Based on the model analysis, we find that "InternationalYes:Q3.6Disagree" has significant related to our response value stress.

- Doing the CI test, we can support our finding above. "InternationalYes:Q3.6Disagree" can affect the "Stress" value.

- We can say that, for those who are international students and they don't feel they need to graduate and get a job soon, are having 0.64 less stress feeling.

International*Q3.6 effect plot
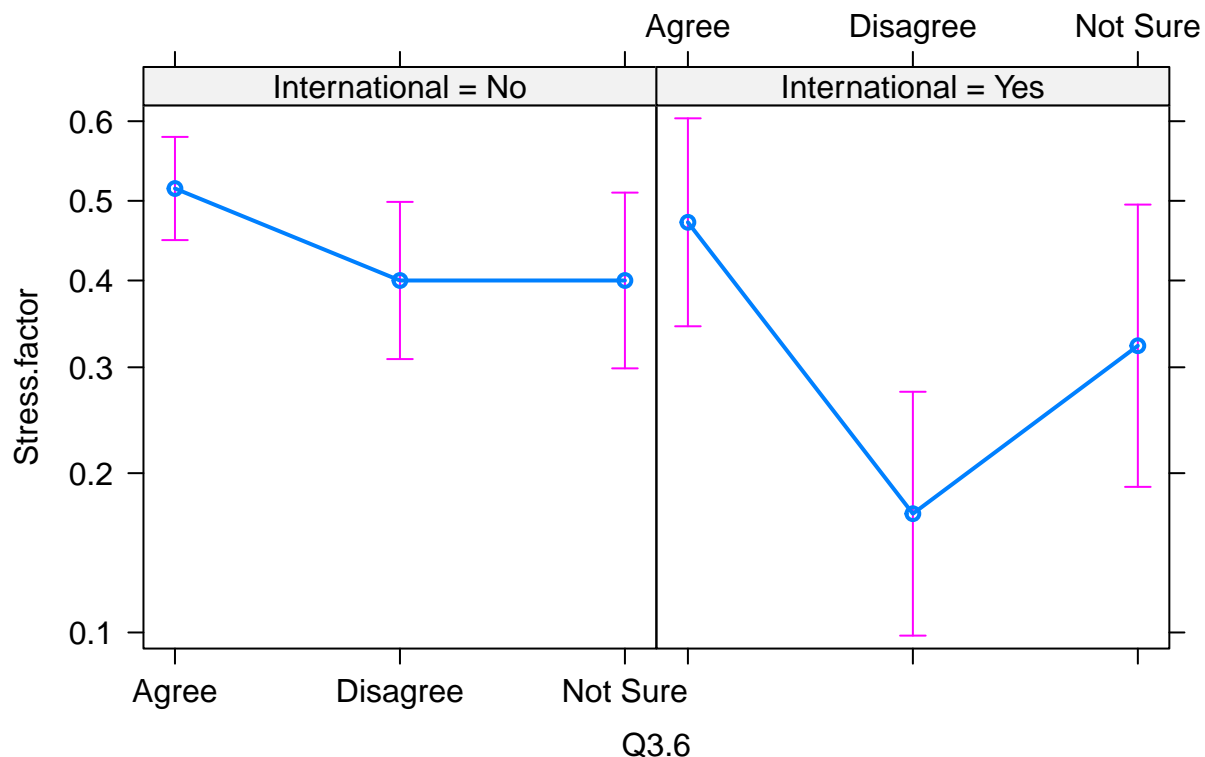
- From the interaction plot above, we conclude that:

- 1. Within those who think the need to graduate and get a job soon, International students and not international students don't vary to much on stress feelings.

- 2. Within Those who don't think the need to graduate and get a job soon, International Students are having much less stress than all other students, and not international students also feel less stress than other not international students.

9   Summary of Results

- Students have high anxiety of their academics, daily life(friendship, room-mates, and dating) would have higher stress feeling.

- People who have less sleep time and not doing exercise tend to have higher stress.

- Making more friends, especially with the same major, would help disperse stress.

- Students who are concerned about getting jobs soon after their graduation may have higher stress.

- Transferring or not doesn't show a big difference on stress feeling level.

- Junior students generally are facing more stress than other year under-graduates.

- International students who are also not considering having a job soon after graduation have much less stress than others.

10   Challenges of the Study

- Our dataset is not big enough for our variables tests. For example, we have a lot of categorical variables so that our degree of freedom for the logistic models are large considering our eight-hundred sample size.

- Our dataset is really messy, but this is also very natural, and we spend a lot of energy on cleaning variables and creating new variables.

- While doing the variables selection, we find out most of the variables don't have much significance relationship with our response variable stress. The most three significant variables are all the questions asked directly about participants' anxiety feeling level. So we decided to also look for some other variables that we are interested in as well as show some significance based on our studies.

11   Recommendation for the future

- For future study, we recommend improving the questions of the survey. After we cleaned the data, we found that many students were too free to fill in unrealistic answers. As a result, we had to abandon the data

of a small number of students. If we set the survey as a wide range of multiple-choice questions instead of open-ended questions, it will help us to get a better dataset. We can also apply this survey to students in other schools, so we can collect more data to make models.