Stats 112: Final Project Report Group 2B

Group Member Names (UID): Yaxin Tan (705121965), Charles Liu (304804942), Kevin Nguyen (305182183), Alexandra Tucker (704813360), Sam Naraghi (305188279), Wan-ling Renee So (105189595)

Research Question: Can we predict a student's Sense of Belonging at UCLA from their Major, Stress level, and Ethnicity? Additionally, what is the effect of an interaction between ethnicity and stress on the prediction model?

Interaction Effect Questions: Is the students' sense of belonging at UCLA related to their ethnicity, perception of stress, and the combined effect of ethnicity and perception of stress?

1) Is a students' ethnicity related to their sense of belonging at UCLA?
2) Is perception of stress related to students' sense of belonging at UCLA?
3) Is the effect of perception of stress on sense of belonging similar for students of different ethnicities?

Purpose: The purpose of this project is to predict a student's Sense of Belonging at UCLA from their Major, Ethnicity, and Perception of Stress. Additionally, an interaction will be used to examine the combined effect of Ethnicity and Stress on Sense of Belonging at UCLA.

Outcome: Belonging: Indicator for "sense of belonging" (continuous variable from 0 to 1 ,0= Low level of sense of belonging, 1=High level of sense of belonging) [*Item Reliability*: *Cronbach's α=0.67*]

Predictors: (1) Major- Major of the Participant (2) Stress- Indicator for "Stress" (continuous variable from 0 to 1 ,0= Low level of stress, 1=High level of stress) *[Item Reliability: Cronbach's α=0.61]* (3) Ethnicity- Ethnic background (Factor: African American, Asian, Caucasian, Latino/Hispanic, Mixed, Other)

Methods: The sample size is 866 students at the University of California, Los Angeles. The project uses statistical analysis methods like multiple linear regression, correlation, interaction effect, to illustrate the relationship between the student's sense of belonging, major, ethnicity and stress level. Programming technique through R studio is used to show the relationship between the variables and predict the student's Sense of

Belonging from Stress, Major and Ethnicity.

Libraries to Load Up in RStudio:
1. library(readr)
2. library(car)
3. library(PerformanceAnalytics)
4. library(effects)
5. library(gplots)

RStudio Codes:
- stemfinal <- read_csv("UCLA Works/UCLA Fall 2019/Stats 112/R Codes/Raw Data/stemfinal.csv")
- attach(stemfinal)
- View(stemfinal)
- table(Ethnic)
- Ethnicr <- recode(Ethnic, "'Latino/Hispanic' = 'Other';'Mixed Race' = 'Other'")
- table(Ethnicr)
- m1 <- lm(Belonging~ Majorr + Stress * Ethnicr)
- summary(m1)
- plot(m1)

- m2 <- lm(Belonging ~ Stress * Ethnicr)
- summary(m2)
- plot(allEffects(m2), ask = FALSE)

- m1vif <- lm(Belonging ~ Majorr + Stress + Ethnicr)
- vif(m1vif)
- residualPlots(m1)
- influencePlot(m1)
- NewDataSet <- stemfinal[-c(54, 69, 701, 733),]

- summary(NewDataSet)

- scatterplot(Belonging, Stress)
- cor(Belonging, Stress, use = "complete.obs")

- table(Majorr, Ethnicr)
- plot(Stress, Belonging, main="Scatterplot of Students' Stress and Sense of Belonging to UCLA")
- abline(line(Stress,Belonging), col="red")
- hist(Stress)
- hist(Belonging)

- plotmeans(Belonging~Majorr, main="Plot of the means and error bars for Sense of Belonging to UCLA + and Major of students")
- plotmeans(Belonging~Ethnicr, main="Plot of the means and error bars for Sense of Belonging to UCLA + and Ethnicity of students")
- summary(Stress)
- cutstress <- cut(Stress, br=c(0, 37.50, 62.50, 100.00),
          labels=c("Low Stress", "Medium Stress", "High Stress"), right = FALSE)
- Stress_Levels <- as.factor(cutstress)
- plotmeans(Belonging ~ Stress_Levels, main="Plot of the means and error bars for Sense of Belonging to UCLA + and Stress of students")

- summary(Belonging)
- summary(Stress)
- cutbelong <- cut(Belonging, br=c(0, 61.10, 97.20),
          labels=c("Doesn't Belong","Belongs"),right=FALSE)
- cutstress <- cut(Stress, br=c(0, 37.50, 62.50, 100.00),
          labels=c("Low Stress", "Medium Stress", "High Stress"), right = FALSE)
- mytable <- table(cutbelong, cutstress)
- prop.table(mytable, 2)

RStudio Plots/Tables (images)**:**

```
> table(Ethnic)
Ethnic
          Asian      Caucasian Latino/Hispanic          Mixed Race              Other
            558            135              88                   28                 33
```

1.

```
> table(Ethnicr)
Ethnicr
    Asian Caucasian     Other
      558       135       149
```

2.

```
Call:
lm(formula = Belonging ~ Majorr + Stress * Ethnicr)

Residuals:
    Min      1Q  Median      3Q     Max
-36.132  -6.714  -0.170   6.162  29.990

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)             85.75837    1.33882  64.055  < 2e-16 ***
MajorrMixed             -1.77803    1.56950  -1.133 0.257608
MajorrPhysical Science  -2.66150    0.79642  -3.342 0.000871 ***
MajorrSocial Science     0.10874    1.42968   0.076 0.939390
Stress                  -0.49833    0.02326 -21.421  < 2e-16 ***
EthnicrCaucasian         2.58700    2.84257   0.910 0.363045
EthnicrOther            -0.15041    2.79826  -0.054 0.957147
Stress:EthnicrCaucasian  0.05280    0.05199   1.016 0.310064
Stress:EthnicrOther     -0.04854    0.05006  -0.970 0.332559
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.1 on 809 degrees of freedom
  (48 observations deleted due to missingness)
Multiple R-squared:  0.4857,    Adjusted R-squared:  0.4807
F-statistic: 95.52 on 8 and 809 DF,  p-value: < 2.2e-16
```
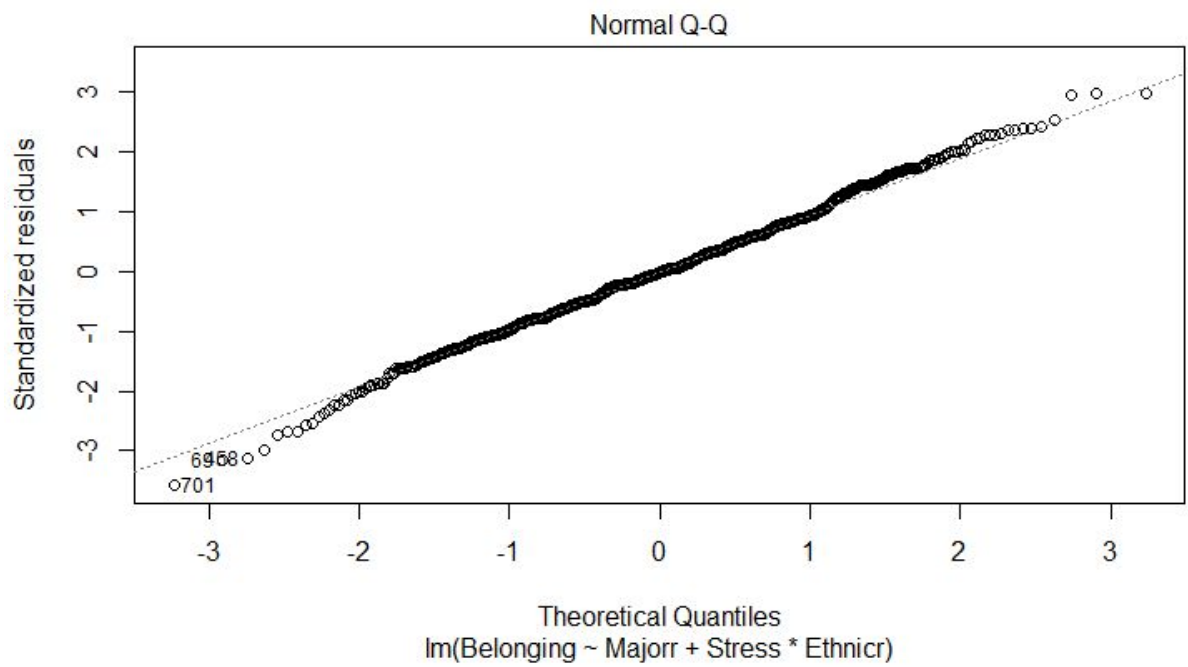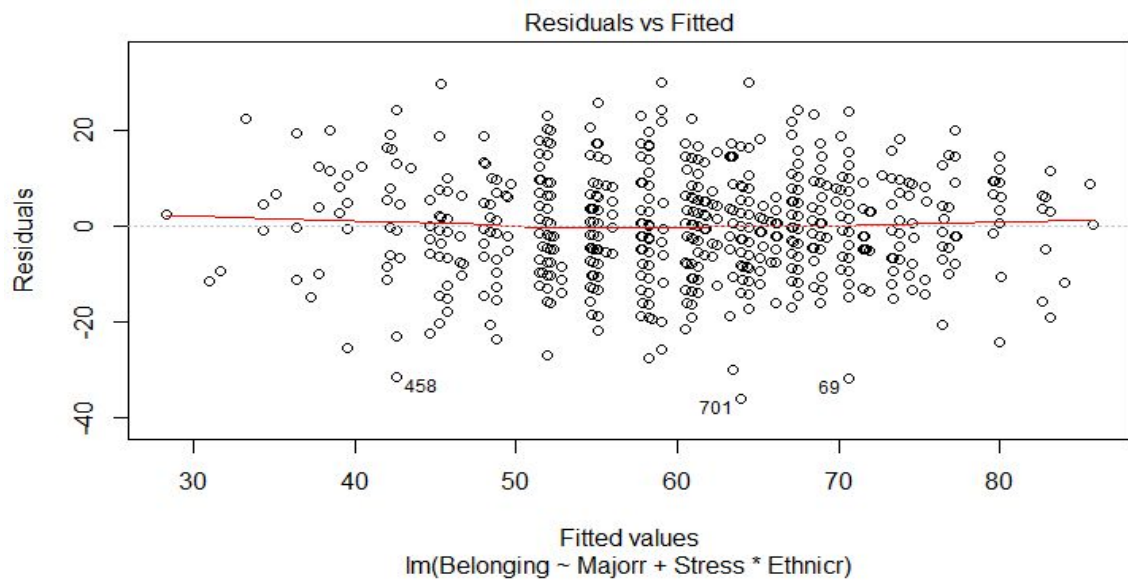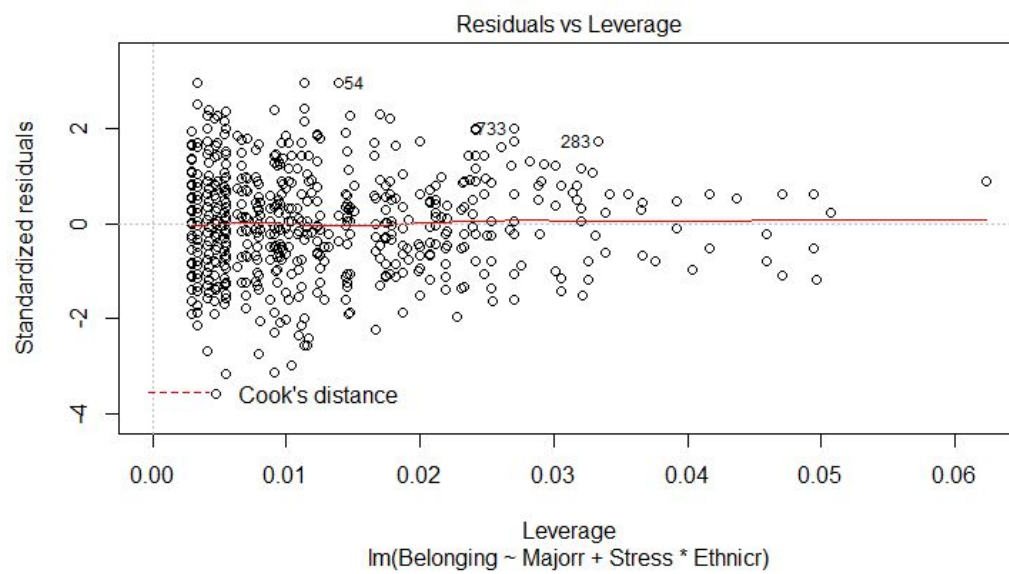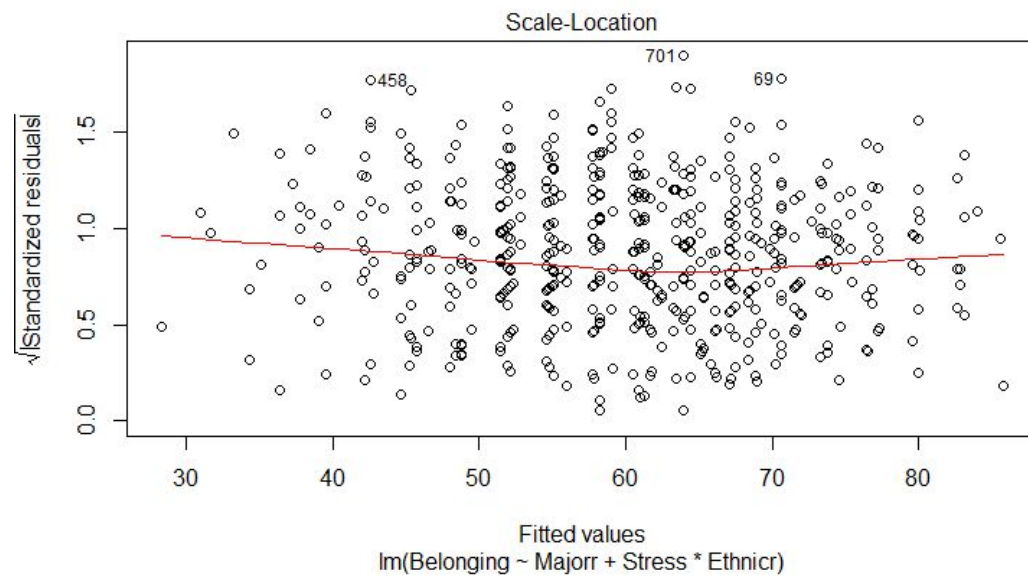
3.

Residuals vs Fitted

Residuals

Fitted values
lm(Belonging ~ Majorr + Stress * Ethnicr)

4.

Normal Q-Q

Standardized residuals

Theoretical Quantiles
lm(Belonging ~ Majorr + Stress * Ethnicr)

5.

Scale-Location

√|Standardized residuals|

Fitted values
lm(Belonging ~ Majorr + Stress * Ethnicr)

6.

Residuals vs Leverage

Standardized residuals

Cook's distance

Leverage
lm(Belonging ~ Majorr + Stress * Ethnicr)

7.

```
Call:
lm(formula = Belonging ~ Stress * Ethnicr)

Residuals:
    Min      1Q  Median      3Q     Max
-34.437  -6.728  -0.396   6.883  29.083

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)             83.65224    1.19229  70.161   <2e-16 ***
Stress                  -0.48894    0.02325 -21.031   <2e-16 ***
EthnicrCaucasian         3.60619    2.84766   1.266    0.206
EthnicrOther             1.47355    2.78665   0.529    0.597
Stress:EthnicrCaucasian  0.04729    0.05227   0.905    0.366
Stress:EthnicrOther     -0.06380    0.05023  -1.270    0.204
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10.17 on 818 degrees of freedom
  (42 observations deleted due to missingness)
Multiple R-squared:  0.474,     Adjusted R-squared:  0.4707
F-statistic: 147.4 on 5 and 818 DF,  p-value: < 2.2e-16
```
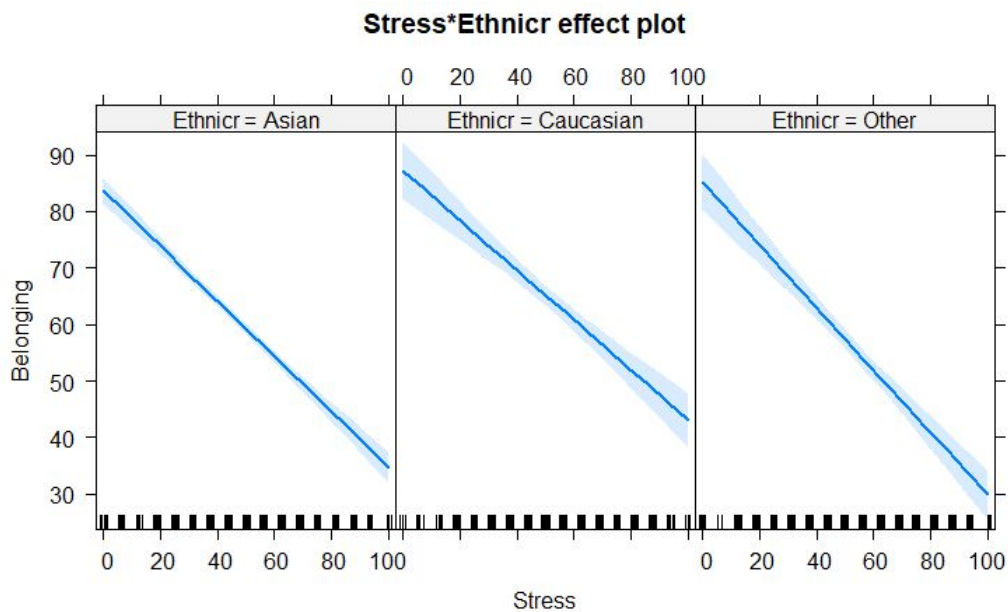
8.

Stress*Ethnicr effect plot
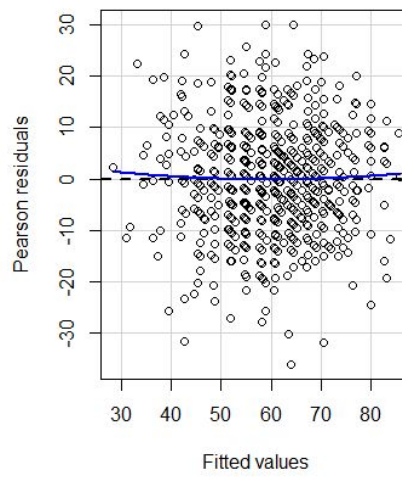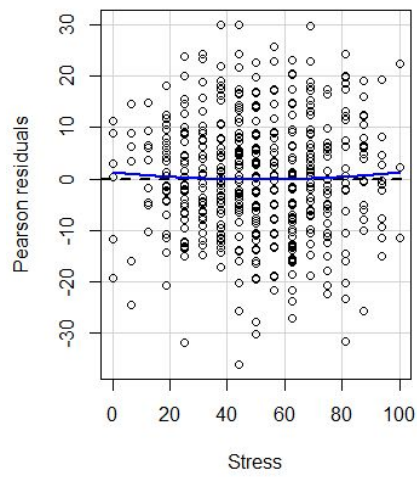
9.

```
> vif(m1vif)
            GVIF Df GVIF^(1/(2*Df))
Majorr  1.098380  3        1.015762
Stress  1.019345  1        1.009626
Ethnicr 1.106090  2        1.025528
```
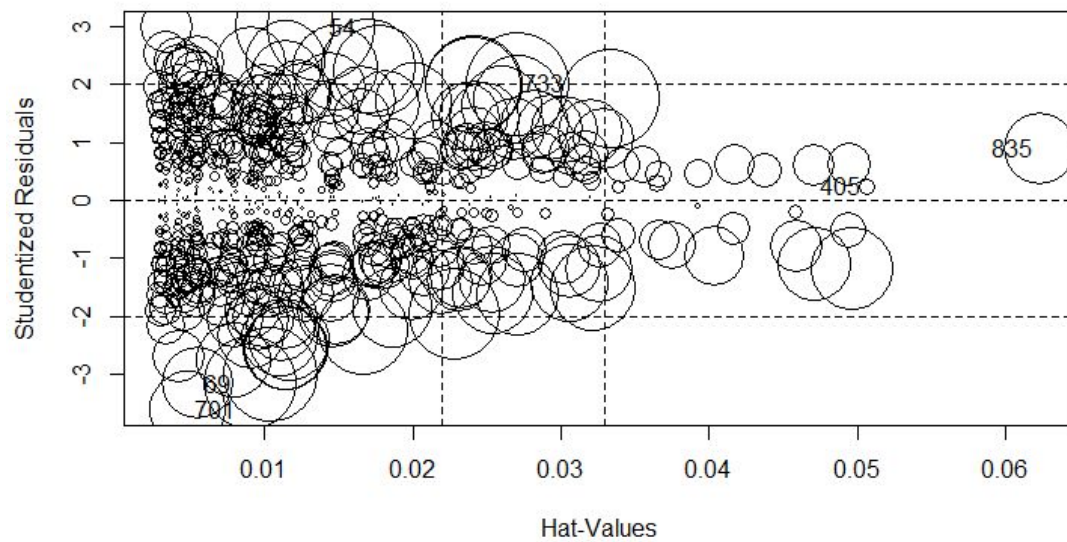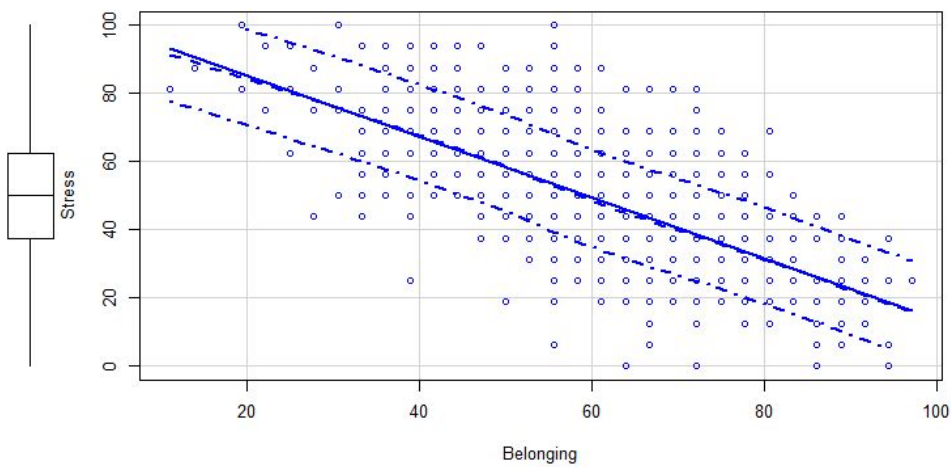
10.

11.

```
> influencePlot(m1)
        StudRes          Hat         CookD
54    2.9718645  0.013861338  0.0136615023
69   -3.1671126  0.005460594  0.0060517555
405   0.2375267  0.050697392  0.0003351737
701  -3.6108650  0.004783156  0.0068605966
733   2.0179724  0.027013104  0.0125143796
835   0.8984351  0.062323826  0.0059625994
```
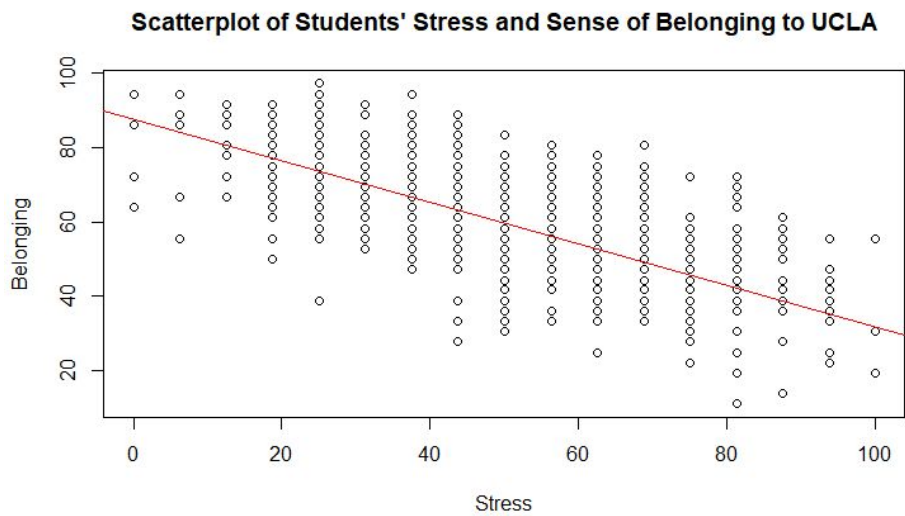
12.

13.

14.

```
> cor(Belonging, Stress, use = "complete.obs")
[1] -0.6613278
```

15.

16.

```
> table(Majorr, Ethnicr)
                 Ethnicr
Majorr            Asian Caucasian Other
  Life Science      162        77    90
  Mixed              40         7     3
  Physical Science  311        39    45
  Social Science     40        11    10
```
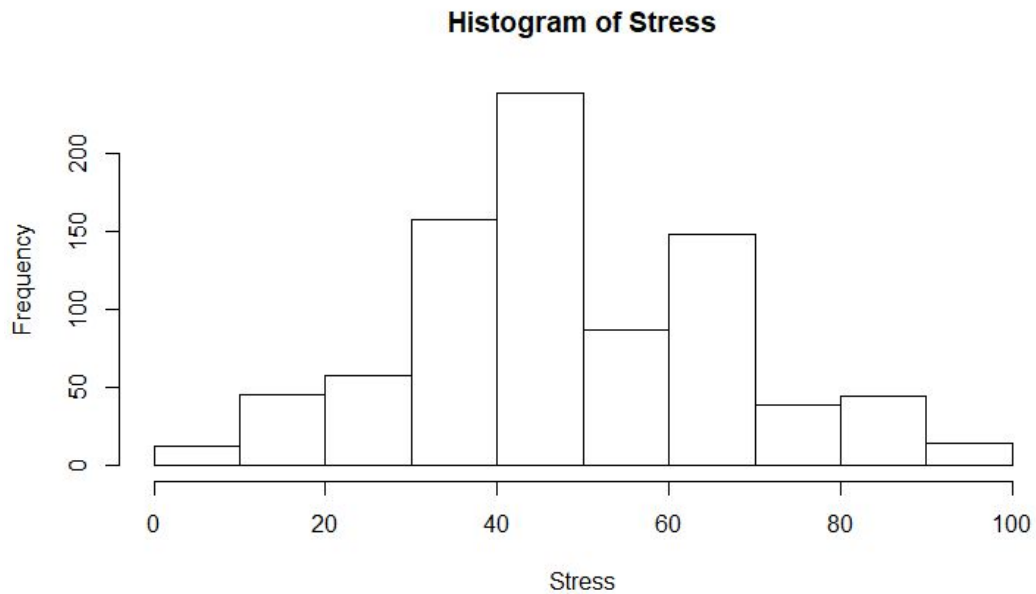
Scatterplot of Students' Stress and Sense of Belonging to UCLA

17.

Histogram of Stress

18.

Histogram of Belonging

19.

Plot of the means and error bars for
Sense of Belonging to UCLA and Major of students

20.

Plot of the means and error bars for
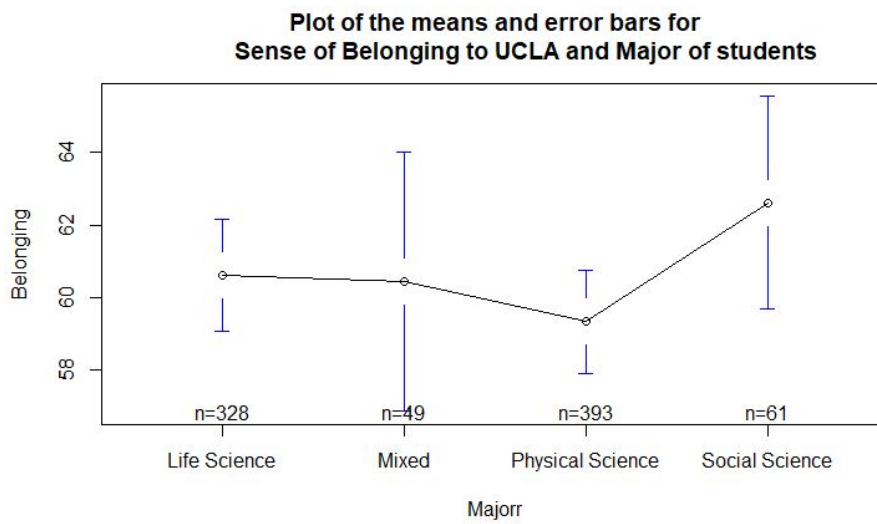**Sense of Belonging to UCLA and Ethnicity of students**
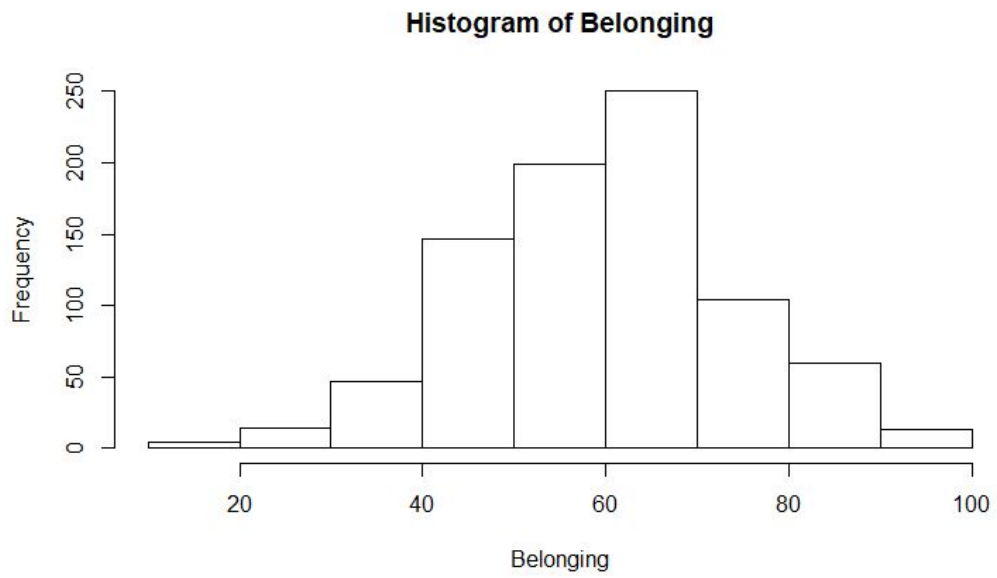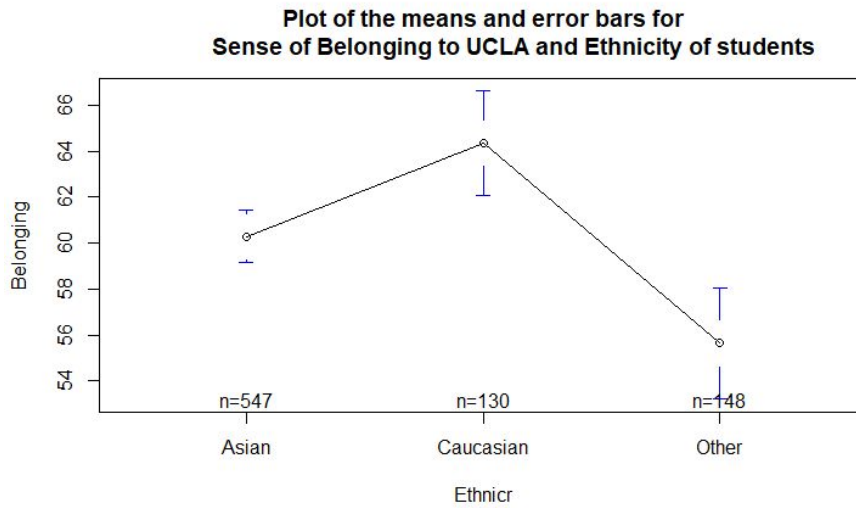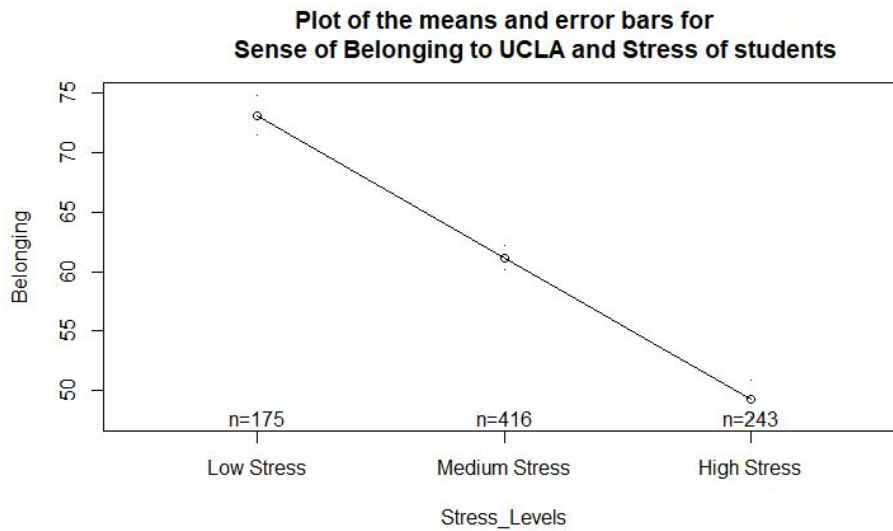

21.

```
> summary(Belonging)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
  11.10   50.00   61.10   60.11   69.40   97.20     28
> summary(Stress)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.   NA's
   0.00   37.50   50.00   49.31   62.50  100.00     20
```

22.

Plot of the means and error bars for
Sense of Belonging to UCLA and Stress of students

23.

```
> prop.table(mytable, 2)
                cutstress
cutbelong        Low Stress Medium Stress High Stress
  Doesn't Belong 0.1149425    0.4567308   0.8106996
  Belongs        0.8850575    0.5432692   0.1893004
```
24.

Interpretations of Our Model, Tables, & Plots (numbers follow respectively from above "**RStudio Plots/Tables:**"):

1. We see that for our Ethnicity variable, Latino/Hispanic, Mixed Race, and Other are all smaller compared to Caucasian and Asian.
2. We decided to combine Latino/Hispanic, Mixed Race, and Other all into one category known as "Other". We did this so the sample sizes won't cause it to skew our data. Therefore, we used the recode() function in R to recreate Ethnicity into "Ethnicr".
3. Interpretation for Our Linear Model:
    a. Due to Physical Science Majors and Stress Levels variables' significance levels, we are able to use them to predict our outcome variable.
    b. We can see that for Physical Science Majors the p-value is close to zero (0.000871), which is statistically significant. For those students who are in the Physical Science Major, we see, ON AVERAGE, a 2.66 decrease in Sense of Belonging.
    c. We can see that for Stress levels the p-value is close to zero (< 2e-16), which is statistically significant. For every 1 unit of increase for sense of Stress levels, we see, ON AVERAGE, a 0.50 decrease in Sense of Belonging.
    d. The other Variables are NOT statistically significant, and therefore, we are unable to draw any conclusions from them.
    e. We see that we have a Multiple R^2 of 48.57%. This Variance helps us explain our model, and we see that this is over 30% Variance. Therefore, we can conclude that this model is excellent for us to use.
4. The residuals seem to be equally spread around the horizontal line. However, the residuals seem to line up at approximate fitted value of 52, 55, 58 and 62, which may affect the results of the data. Suggests that there is a lack of non-linear relationship between the predictor variables and the outcome variables.
5. The QQ-line seems to be straight, so it suggests that the residuals are normally distributed.
6. The residuals seem to be spread equally and randomly around the red line. This suggests that there is equal variance, or homoscedasticity.

7. All of the residuals are within Cook's Distance, which means that there aren't any points that could influence the regression line.
8. Interpretation for the Interaction Effect:
    a. For Ethnicity, our baseline is "Asian" Ethnicity. On average, the respondents who are of Caucasian Ethnicity score 2.58700 points higher on their sense of Belonging here at UCLA. This difference is not statistically significant (p = 0.363045). Thus, we can say that being of Caucasian ethnicity is not related to their sense of Belonging here at UCLA.
    b. On average, the respondents who are of Other Ethnicity score 0.15041 points lower on their sense of Belonging here at UCLA. This difference is not statistically significant (p = 0.957147). Thus, we can say that those categorized as Other Ethnicity is not related to their sense of Belonging here at UCLA.
    c. For the Interaction Effect, we can say they are NOT related to their Belonging because Asian, Caucasian, Other are NOT statistically significant. We are unable to draw any conclusions due to the high p-value and NOT being statistically significant.
9. For our Interaction Effect Plot:
    a. As indicated by the plot, as Stress increases, Sense of Belonging decreases and this pattern is true for all three levels of ethnicity (Asian, Caucasian, Other).
    b. The interaction effect is not statistically significant, indicating that the effect of Stress on Belonging is similar for different levels of Ethnicity.
    c. Therefore, there is not enough evidence to conclude anything with Ethnicity and Stress due to it not being statistically significant.
10. Checking our VIF, we can see that our GVIF is not over 5 or under -5, so we are good! The reason why we didn't include in the interaction effect in our predictors just for checking VIF is because if we did, it would create high amounts of VIF, when in reality it should not be that high. Instead, we just added all the predictors together and checked for VIF based on our 3 predictors.
11. The plot of residuals vs. Y^ (= "Y hat") and residuals vs. the three predictors look pretty good. They are not perfect, but they look acceptable.
12. Assuming that the data set we are working with is large, none of the observations listed here have high Standardized residuals. However, with the assumption that the data set is small, there are four observations with both high leverage and high standardized residuals. We see that the possible Leverages are 54, 69, 405, 701, 733, and 835.

13. This is the plot that will help use verify whether we have Leverages (good or bad) or outliers. We can also see our possible Leverages in this plot.
14. There is a strong negative correlation correlation between Stress and Sense of Belonging. This suggests that the less stress each student feels, the more they feel like they belong to the campus.
15. We can see there is a negative correlation with a Sense of Belonging here at UCLA and Stress Levels. This is confirmed numerically and proven graphically in Interpretation #14. The higher the amount of Sense of Belonging a student feels here at UCLA, the lower amounts of Stress Levels a student feels. We can also see there is a linear relationship between Stress and Belonging.
16. This is our cross Frequency Table between Majorr and Ethnicr (both of these were recoded). We are able to see their distribution of their responses to these two variables.
17. From this plot, we can see that although sense of belonging scatter across quite wide range at each stress level, there is a downward trend as indicated by the line. which means a negative correlation between sense of belonging and stress. Students who have higher stress tend to feel lower sense of belonging. The standard deviation of Belonging and Stress are 13.952 and 18.837 respectively.
18. The histogram of Stress is bimodal. It could be explained by the disparity between the stress score of ethnic majorities and ethnic minorities, which is shown later.
19. The graph has a bell-curved shape, which indicates the frequency of Sense of Belonging is normally distributed.
20. The center of each interval is the mean score of sense of belonging to UCLA with respect to students with different majors. The major variable was recoded into four main categories (i.e. Life Science, Mixed, Physical Science and Social Science). Two of the observation was not sent to any categories. However, with a large sample size for each category, it does not affect our interpretation. Students with social science major has the highest sense of belonging to UCLA, while students with life Science has the least sense of belonging to UCLA. Students with Life Science or Mixed have similar mean score of sense of belonging to UCLA.
21. The center of each interval is the mean score of sense of belonging with respect to students with different ethnicities. The ethnicity variable was re-coded into three main categories, Asian, Caucasian and other. Ethnic minorities in the sample are combined into Other (e.g. African American and Latino/Hispanic) to make comparisons. This plot shows that caucasians have the highest sense of belonging, Asian seconds in sense of belonging, while Other has the least sense of belonging to UCLA.The average mean difference between the Caucasian and

Other is around 10. This plot illustrates that ethnic majorities have higher sense of belonging in comparison to ethnic minorities.

22. We observe from the summary that we can use the cut() function in R to turn the Stress Levels into 3 Levels (originally numeric, but we code to turn it into categorical). Using the Min, 1st Quartile, 3rd Quartile, and Max values, we cut Stress into 3 categories known as "Low Stress, Medium Stress, and High Levels".

23. The center of each interval is the mean score of stress level. This plot shows that the higher the stress level of students, the lower their sense of belonging to UCLA. The stress level was recoded, from a numerical variable to a categorical variable with three levels (i.e low, medium, and high), to make comparisons. From the previous Interpretations, we have illustrated that stress is statistically significant. We can see that Stress plays an important role in a students' Sense of Belonging. This graph helps us to reach our conclusion that Stress is a significant variable in determining Sense of Belonging.

24. To present it in a way that non-statisticians understand, we will be using the method of Odds Ratio. The code being used gives us the Column Percentages of the Odds Ratio. I've also used the cut() function for Sense of Belonging, using the Min, Median, and Max values to create 2 Levels known as "Doesn't Belong and Belongs" for their feeling of Belonging here at UCLA. Here is our results:
    a. Of the respondents with Low Stress, 11% feel like they Don't Belong here at UCLA, and 89% feel like they do Belong.
    b. Of the respondents with Medium Stress, 32% feel like they Don't Belong here at UCLA, and 68% feel like they do Belong.
    c. Of the respondents with High Stress, 70% feel like they Don't Belong here at UCLA, and 30% feel like they do Belong.
    d. We can reach the conclusion that as Stress levels elevate higher for students, their sense of Belonging here at UCLA lessens.

Conclusion:

● We can predict a students' sense of belonging from Physical science major and stress level. They are the only two statistically significant predictors in our test.
● Around 49% of the variance in Students' Sense of Belonging at UCLA is explained by Stress and Physical Science Majors
● For every one unit increase in levels of Stress, on average, there is a 0.49833 unit decrease in students' sense of belonging
● For those who are physical science majors, on average, they experience 2.6615 unit decrease in sense of belonging

- There is no statistically significant collinearity between predictor variables
- Additionally, The interaction between Stress and Ethnicity are not statistically significant

Recommendations for Stress:

1. Meditation - 5 minutes a day spent practicing meditation is proven to improve focus, self-compassion, mood, immune function and quality of sleep
2. Friends - Take a break from studying. Go out with friends or classmates
3. Art - Release built-up stress through creativity and self-expression
4. Counseling - CAPS provides workshops on self-management and stress management skills
5. Massage - Relax and Unwind

Recommendations for Physical Science Majors:

1. Group Projects - Promotes interaction, exchanging of ideas and concepts
2. Campus Organizations - Join campus clubs and organizations to meet peers and build interpersonal skills
3. Research Project - Social responsibility research projects promote campus engagement and integration
4. Discussion Sections - Require discussion sections for all physical science courses at UCLA

SIDE NOTE:
- Given more time, we would've used Model Selection to find a better fitting model that may not include interaction. It would help us find more variables that are statistically significant, and therefore, we can draw more conclusions. We would also be able to further explore more interactions that would be statistically significant.
- One of the methods we would use is the Likelihood Ratio Test to test which model fits better given we have more time. We would further explore more into how to use the LRT in RStudio and run the code with our model.