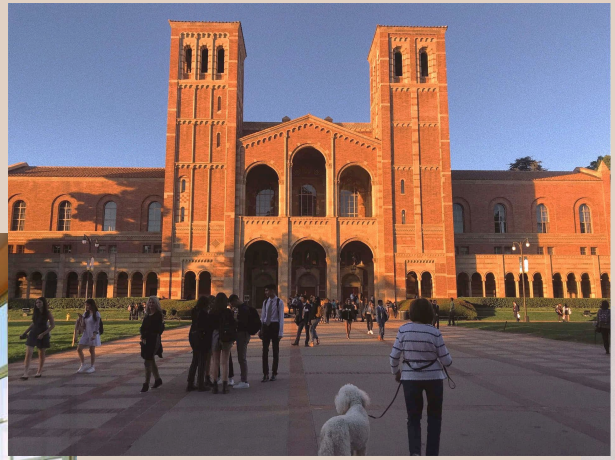
How can various factors impact a UCLA student's GPA?

Group 14: Emre, Taro, Meha, Laura, Marlene, Rita, Min



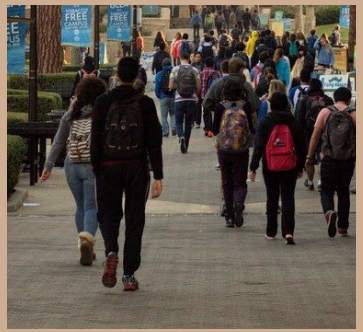
Abstract

Goals

This study aims to investigate the potential impact various factors have on the academic performance of undergraduate students. A survey was conducted among a sample of undergraduate students at UCLA, and multiple regression analysis was used to examine the relationship between the predictors and students' GPA.

Results

The results indicate that year of study, ethnicity, parental education, student engagement, language proficiency, academic respect and enrollment in a statistics course were significant predictors of GPA, while gender and relationship status did not have significant impacts.



Implications

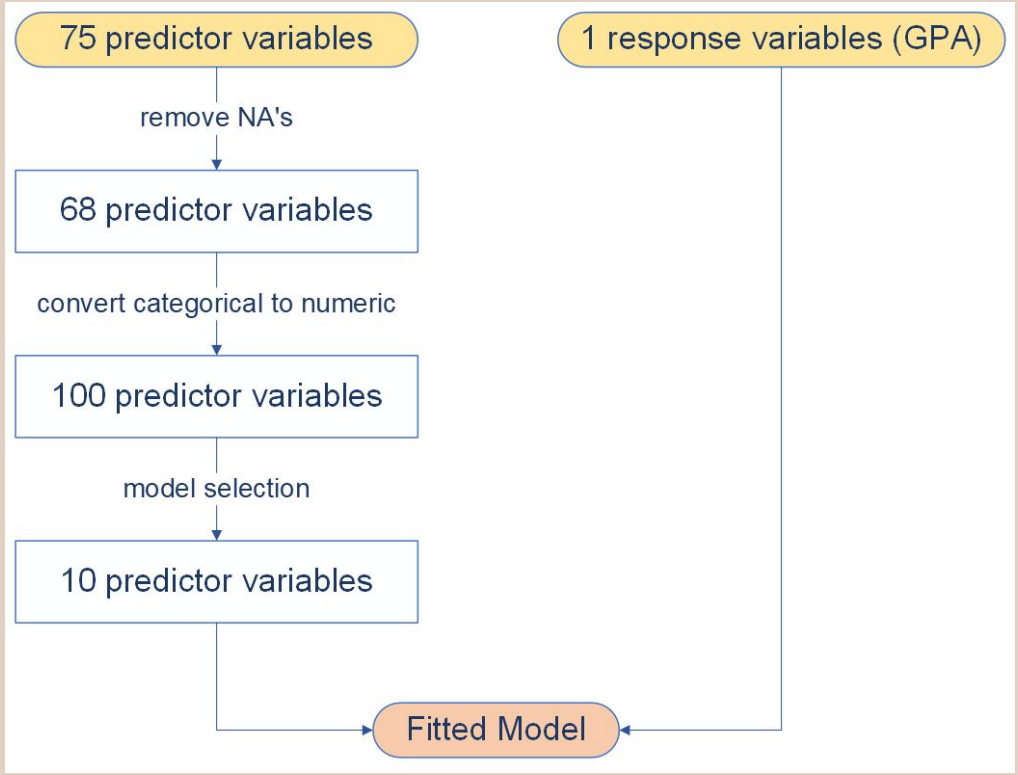
These findings have implications for universities and policymakers in developing effective interventions and policies to support students' academic success.

Data

Range: [0,1] or [1,5]

converted categorical variables into dummy variables (0,1) or ranges

Ex. year, feduc



Range: [0.0, 4.0]

Data Processing

Before selecting features, we converted categorical data to numeric data:

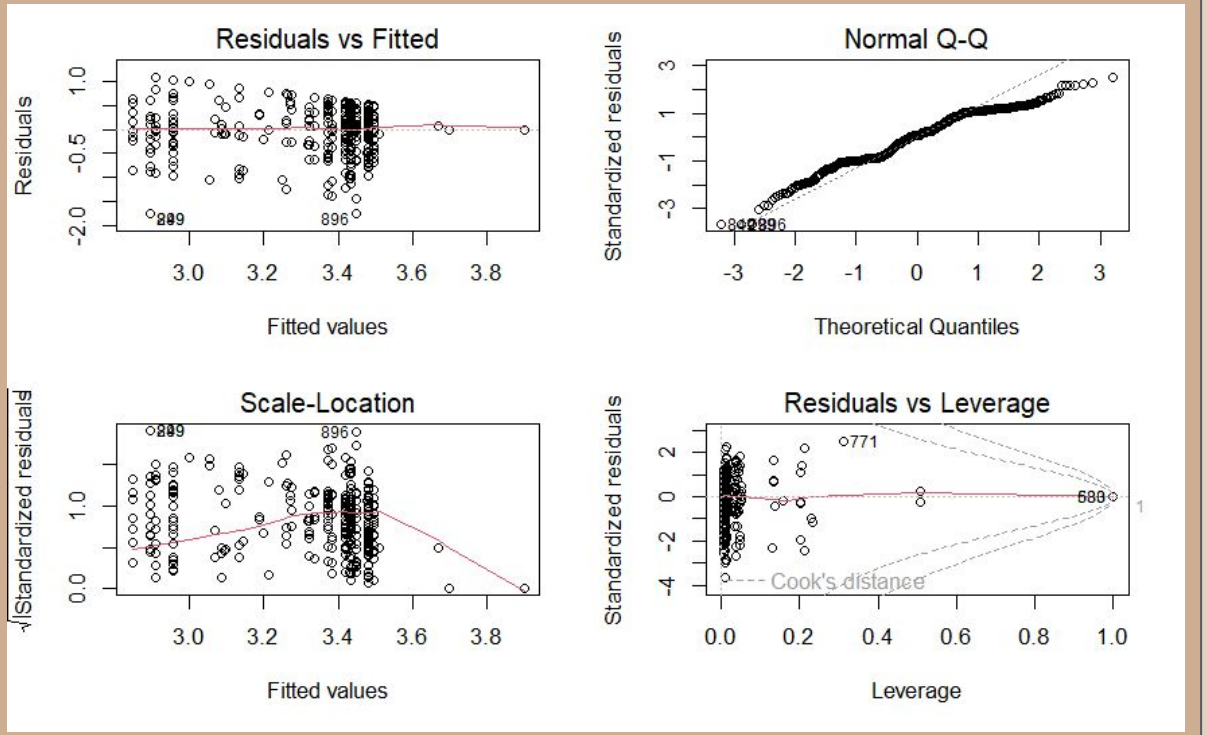
Original Variable (Categorical)	Transformed Variable (Numeric)
Course (stats10, stats13, etc.)	5 Dummy Variables
Year (freshman, sophomore, etc.)	1-5 (first year → fifth year +)
Origin (in-state, out-of-state, international)	2 Dummy Variables
Enrollment Status (full-time, part-time)	1 Dummy Variable
Language (english only, multilingual)	1 Dummy Variable
Discipline (STEM, humanities, etc)	4 Dummy Variables
Campus (North, South)	1 Dummy Variable
Relationship Status (single, relationship, married)	3 Dummy Variables
Gender (male, female, transgender, queer)	4 Dummy Variables

Data Processing Cont.

Original Variable (Categorical)	Transformed Variable (Numeric)
Ethnicity (hispanic/latino, asian, etc)	9 Dummy Variables
Parent Education Level (HS, Bachelors, etc)	1-4 (HS → PhD)
Socioeconomic Class (lower, middle, upper)	1-3 (lower → upper class)
Sexual Orientation (asexual, bisexual, etc)	6 Dummy Variables
Religion (Christian, Jewish, etc.)	5 Dummy Variables
Political View (liberal, conservative, etc)	5 Dummy Variables
Class/UCLA Climate Rating (comfortable, uncomfortable, etc)	1 → 5 (very uncomfortable → very comfortable)
Exclusion (no, yes, yes + affected learning)	1-3 (no → yes + affected learning)
Total: 17	**Total: 61**

Baseline Models

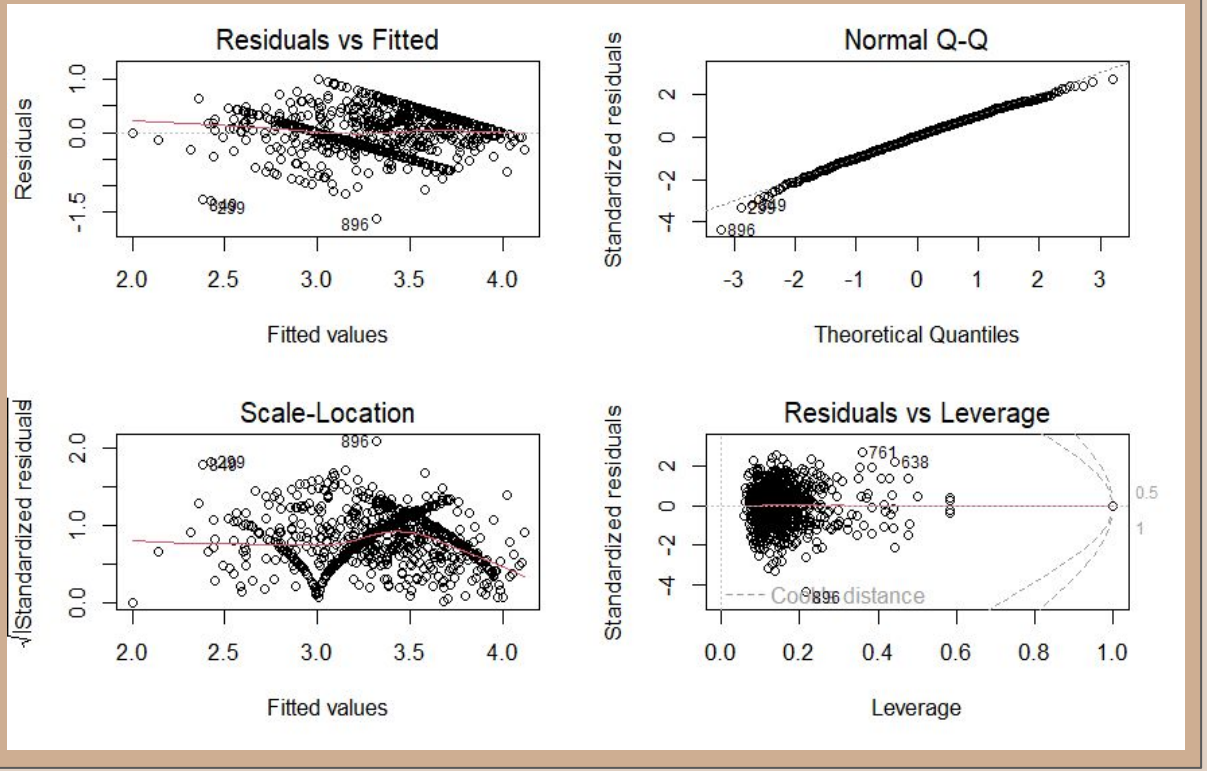
Suspected Variables: gender, relationship status, and ethnicity



Number of Predictors	**3**
F-Statistic	10.99
P-value	< 2.2e-16
Adj-R^2	0.149
MSE	0.283
Number of Significant Variables	0

Baseline Models

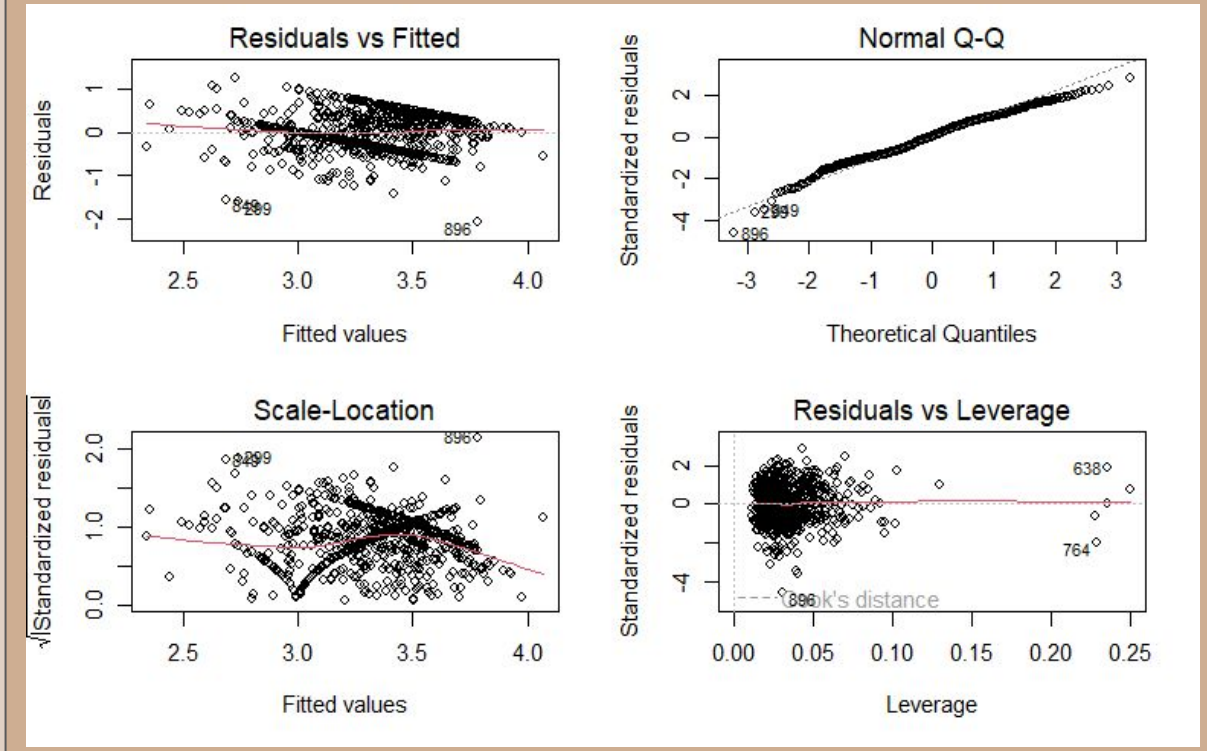
All Variables Included



Number of Predictors	**68**
F-Statistic	4.53
P-value	< 2.2e-16
Adj-R^2	0.361
MSE	0.274
Number of Significant Variables	25

Feature Selected Models

Significant Variables



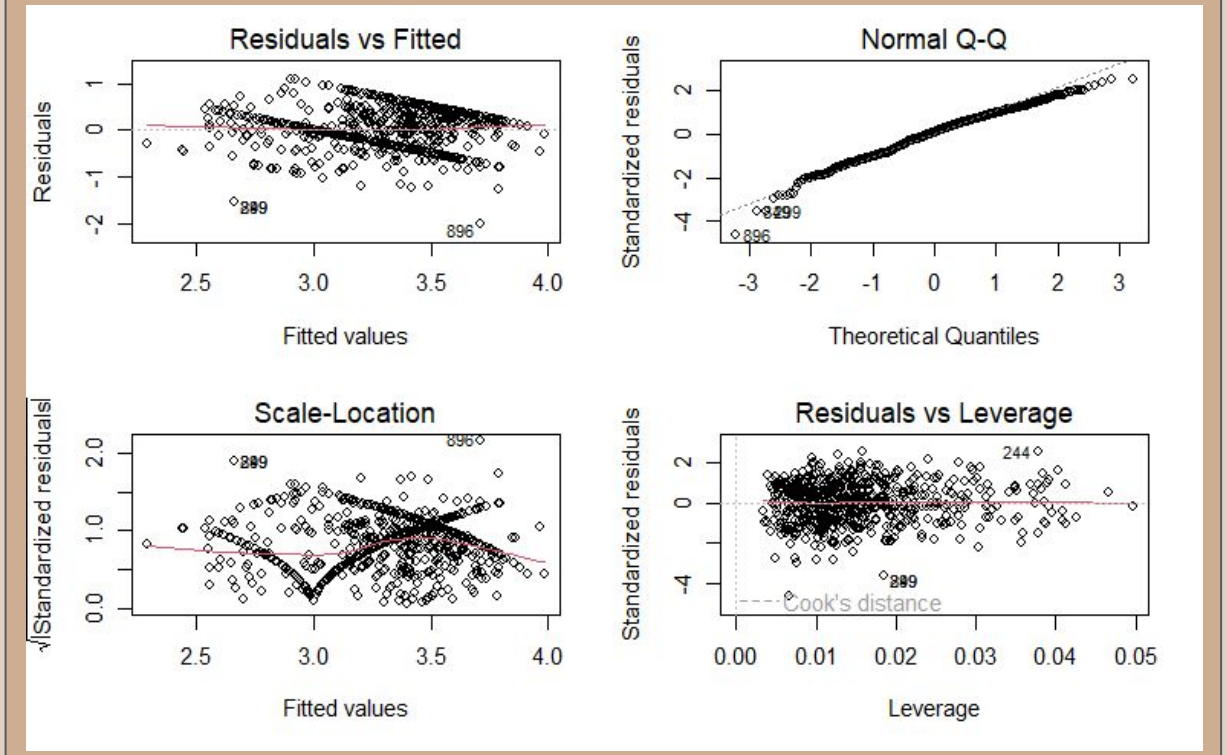
Number of Predictors	**25**
F-Statistic	10.26
P-value	< 2.2e-16
Adj-R^2	0.238
MSE	0.311
Number of Significant Variables	12

Stepwise Selected Variables

Variable Names	Variable Meaning
year	School year
feduc	Father’s education level
leavingucla	People who have considered leaving UCLA due to negative experiences
staringatyou	Feeling perceived negatively by others
excluenglish	Feeling socially accepted despite only speaking English
facunderstand	Feeling faculty relates to/understands them
appearanceresp	Feeling responsible for negative reactions to appearance
academicresp	Feeling responsible for one’s own academic success
hispanicLatino	Hispanic/Latino students
stats13M	Students who have taken class Stats 13M

Feature Selected Models

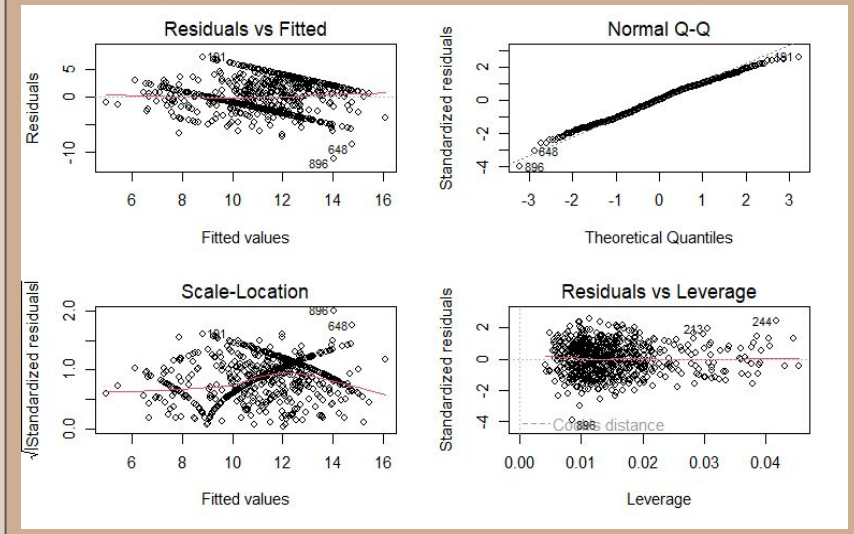
Stepwise Selected Variables



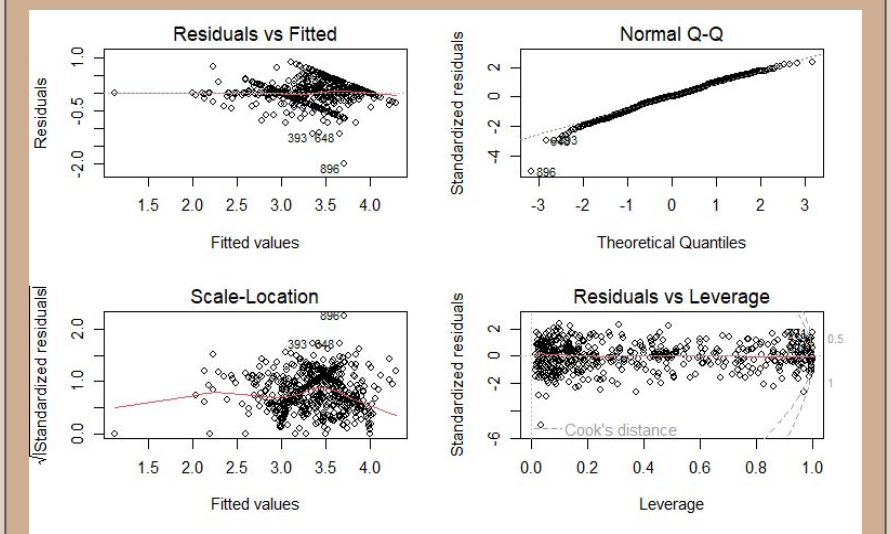
Number of Predictors	**10**
F-Statistic	35.8
P-value	< 2.2e-16
Adj-R^2	0.319
MSE	0.238
Number of Significant Variables	9

Improved Models

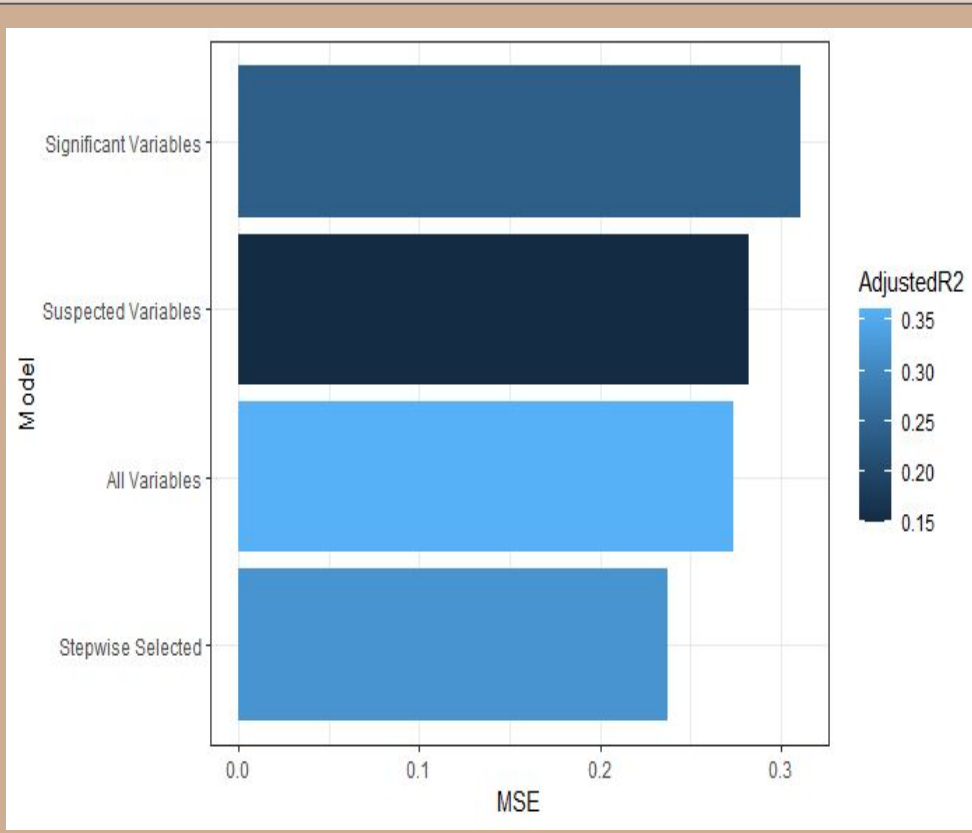
Stepwise Selection with Transformation



Stepwise Selection with Interaction



Model Summary



- Based on this plot of MSE values for each group of variables vs the Model, we can visualize the changes of the adjusted R^2 values.

- More specifically, we can see that the model including all variables have the highest adjusted R^2 whereas the model including suspected variables have the lowest adjusted R^2.

Summary of Findings

Positive Impacts (Correlation Coefficient Included)

- **feduc**: father's education level; .07493
- **academicresp**: Feeling responsible for your own academic success; 0.11837
- **excluenglish**: Only speaking English in relation to feeling socially accepted; .09706
- **facunderstand**: feeling understood by a faculty member; .02323
  → not statistically significant

Negative Impacts (Correlation Coefficient Included)

- **staringatyou:** feeling judged while being on campus; -.07622
- **appearanceresp**: feeling responsible for negative perception of personal appearance; -.08900
- **year**: student year classification; -.09788
- **leavingucla:** considered leaving UCLA due to negative experiences; -0.12799
- **stats13M:** enrolled in STATS 13M; -0.17891
- **hispanicLatino:** hispanic/latino; -.30925

Summary of Findings

Most Significant Positive Impact

Feduc: Father's Education

Paper by European Journal of Public Health:

"Parental Education is one of the best predictors of child school achievement [...] Parents with a high education are able to provide social and material resources promoting higher offspring school achievement"

Most Significant Negative Impact

Variables Associated with Mental Health:

Paper by Suicide Prevention Resource Center:

"Mental health problems can affect a student's energy level, concentration, dependability, mental ability, and optimism, hindering performance. Research suggests that depression is associated with lower grade point averages, and that co-occurring depression and anxiety can increase this association."

Limitations/Challenges

- Fixing and changing how impactful/correlated certain variables are depending on their distribution
- Having to change variables examined due to hypothesis being statistically insignificant/poor fit
- Determining what variables are best: categorical or numerical
- All variables either categorical or discrete numeric variables - difficult to interpret visualizations.
- Attempted Box Cox transformations on all variables, but model was an extremely poor fit and worse than our original model

Suggestions for Future Improvements

- Adding variables for better visuals of mental health specificities in terms of GPA
  → Examples include:
    1. Sleep Quality
    2. Stress Levels
    3. Existing Anxiety Medication

- Alternative transformation methods that are potentially more advanced
  → Visualize data more effectively/efficiently