



MINI PROJECT 2

Johan Ng - IOD



CONTENT OF PRESENTATION

- Background and Problem
- Dataset and Methodology
- Exploratory Data Analysis
- Machine Learning Modelling
- Recommendations
- Conclusion

Credit to Artist: Steve Tierney

BACKGROUND

- University Students usually have **higher depression rates** compared to general population
- High level of stress related to future, employment, studies
- Mental health if neglected could lead to serious consequences such as self-harm and depression
- A **university's mental wellbeing center** is key to early intervention for the students



PROBLEM STATEMENT

Can we predict students at risk using the data collected from the center so that the center is able to provide early intervention for these students?

Source: unsplashed.com



data

Data Descriptor

A Dataset of Students' Mental Health and Help-Seeking Behaviors in a Multicultural Environment

Minh-Hoang Nguyen ¹, Manh-Toan Ho ^{2,3,*}, Quynh-Yen T. Nguyen ^{4,*} and Quan-Hoang Vuong ^{2,3}

¹ International Cooperation Policy, Graduate School of Asia Pacific Studies, Ritsumeikan Asia Pacific University, Beppu, Oita 874-8577, Japan

² Center for Interdisciplinary Social Research, Phenikaa University, Ha Dong district, Hai

³ Faculty of Economics and Finance, Phenikaa University, Ha Dong district, Hanoi 10080

⁴ College of Asia Pacific Studies, Ritsumeikan Asia Pacific University, Beppu, Oita 874-85

* Correspondence: toan.homanh@phenikaa-uni.edu.vn (M.-T.H.); thiqng17@apu.ac.jp (Q

Received: 7 June 2019; Accepted: 17 August 2019; Published: 21 August 2019

Abstract: University students, especially international students, possess a high health problems than the general population. However, the literature regarding determinants of mental health problems as well as help-seeking behaviors of domestic students in Japan seems to be limited. This dataset contains 268 records of acculturative stress, social connectedness, and help-seeking behaviors reported by domestic students at an international university in Japan. One of the main findings from this dataset is how the level of social connectedness and acculturative stress are reported depression among international as well as domestic students. The data provide reliable materials for further study of cross-cultural public health studies in higher education.

DATASET

- Questionnaire of 268 students from international and domestic students in a International university in Japan
- Questionnaire includes:
 - Socio-demographic information
 - Mental health conditions (depression, suicide ideation, acculturative stress, social connectedness)
 - Help-seeking behaviours
 - 25 Categorical Variables and 26 Numerical Behaviours

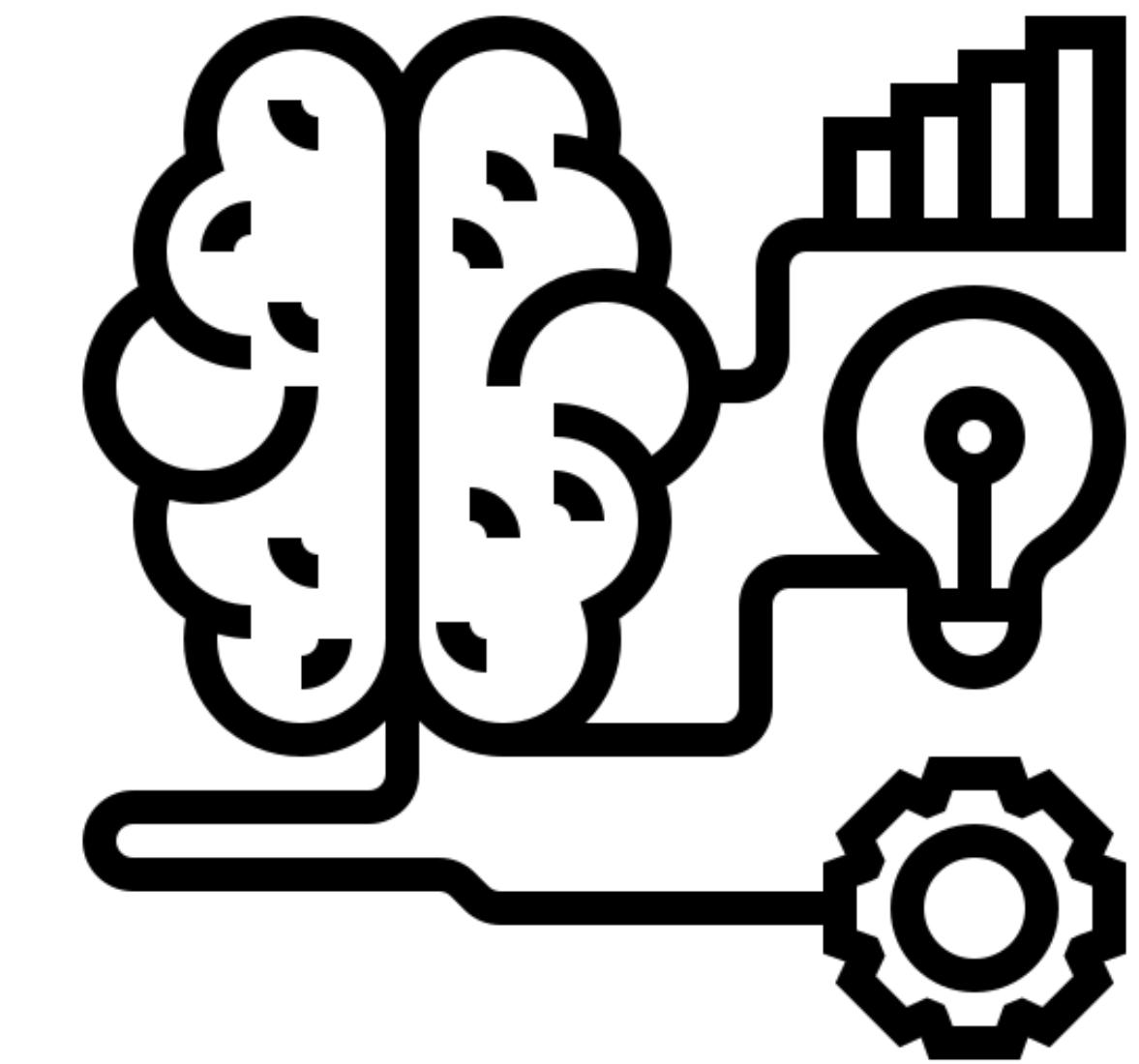
METHODOLOGY



*Cleaning and
preparing the
Data*

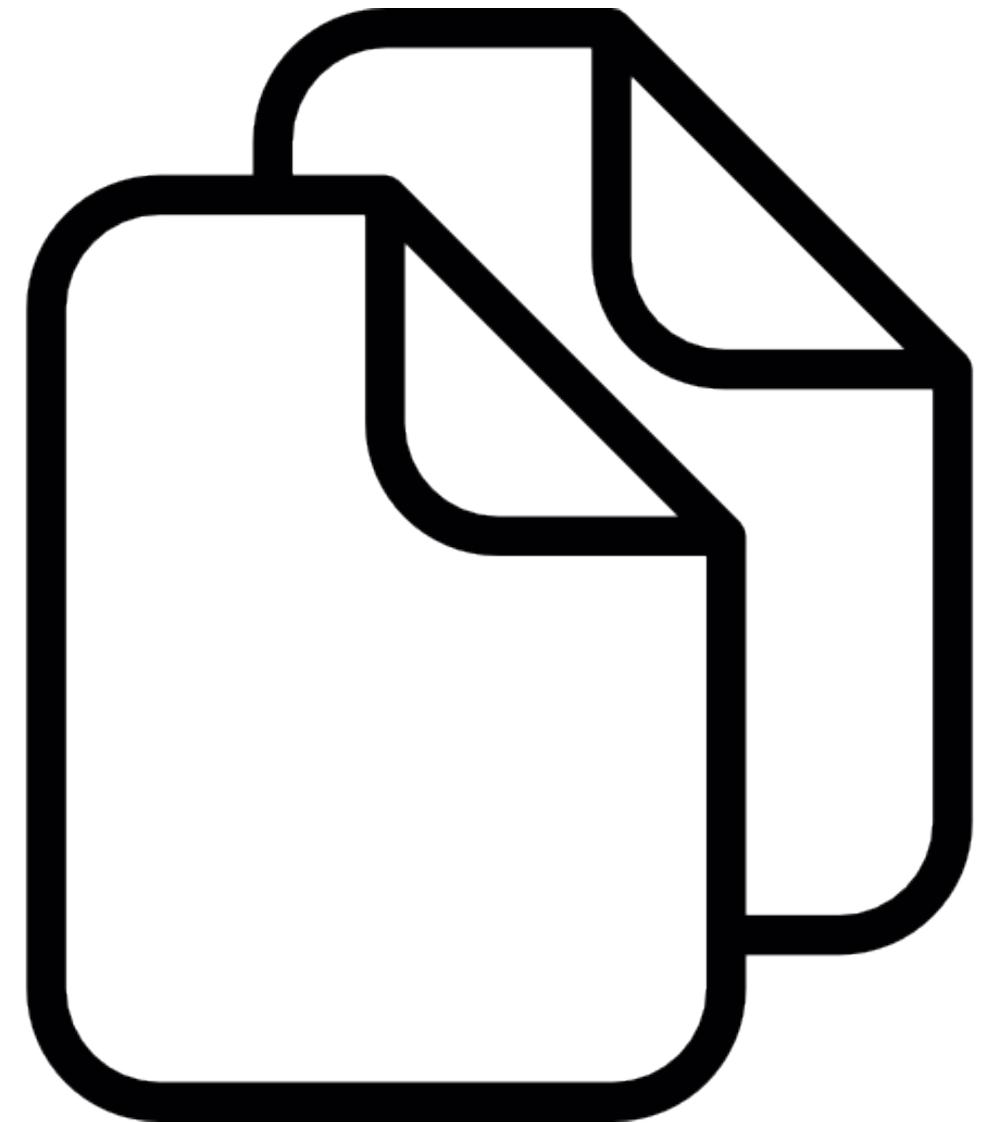


*Exploring the
Data*

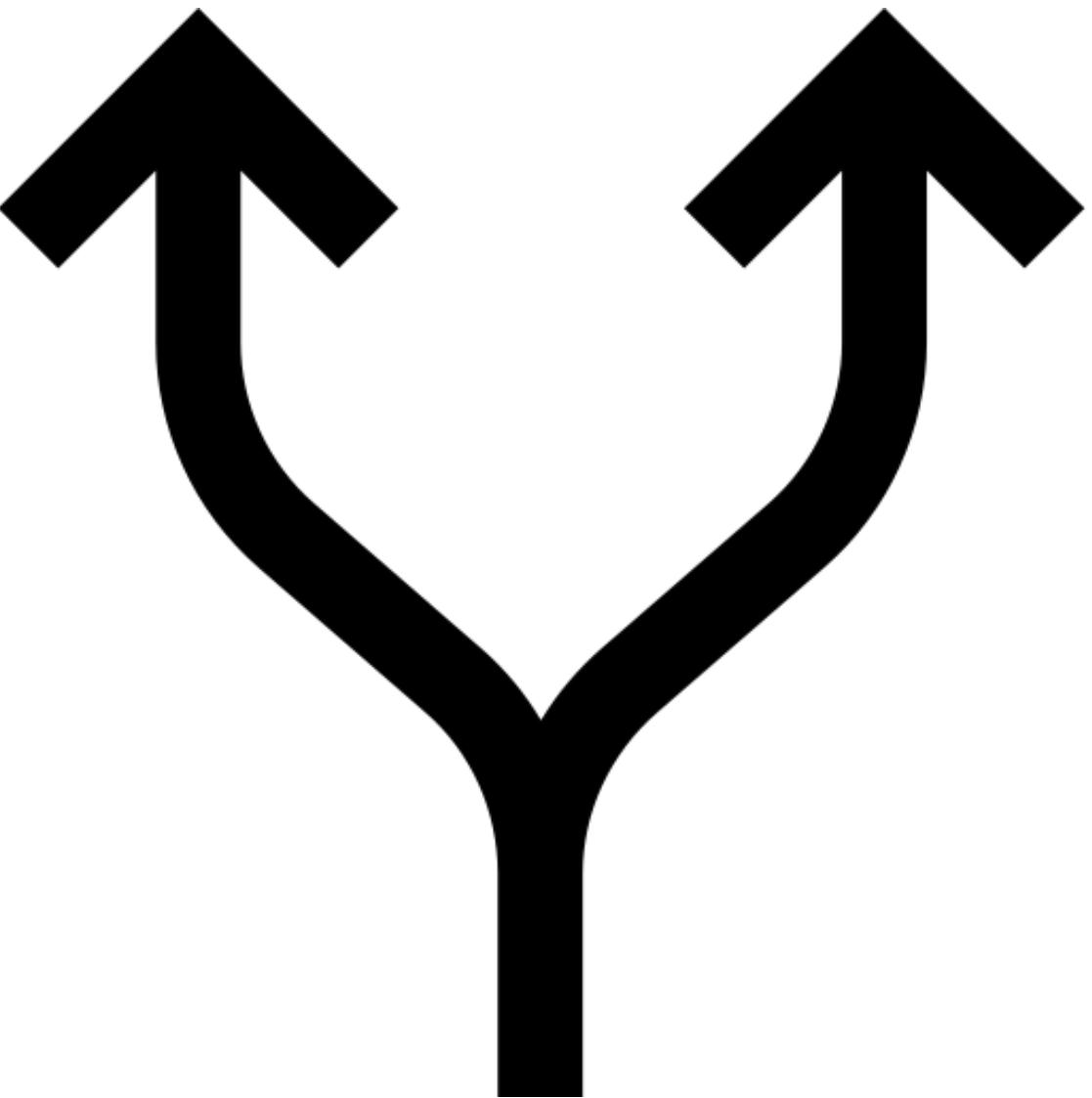


*Modelling the
data using Linear
Regression*

CLEANING AND PREPARING THE DATA



*Removing 18 invalid
questionnaire
(278 questionnaires)*



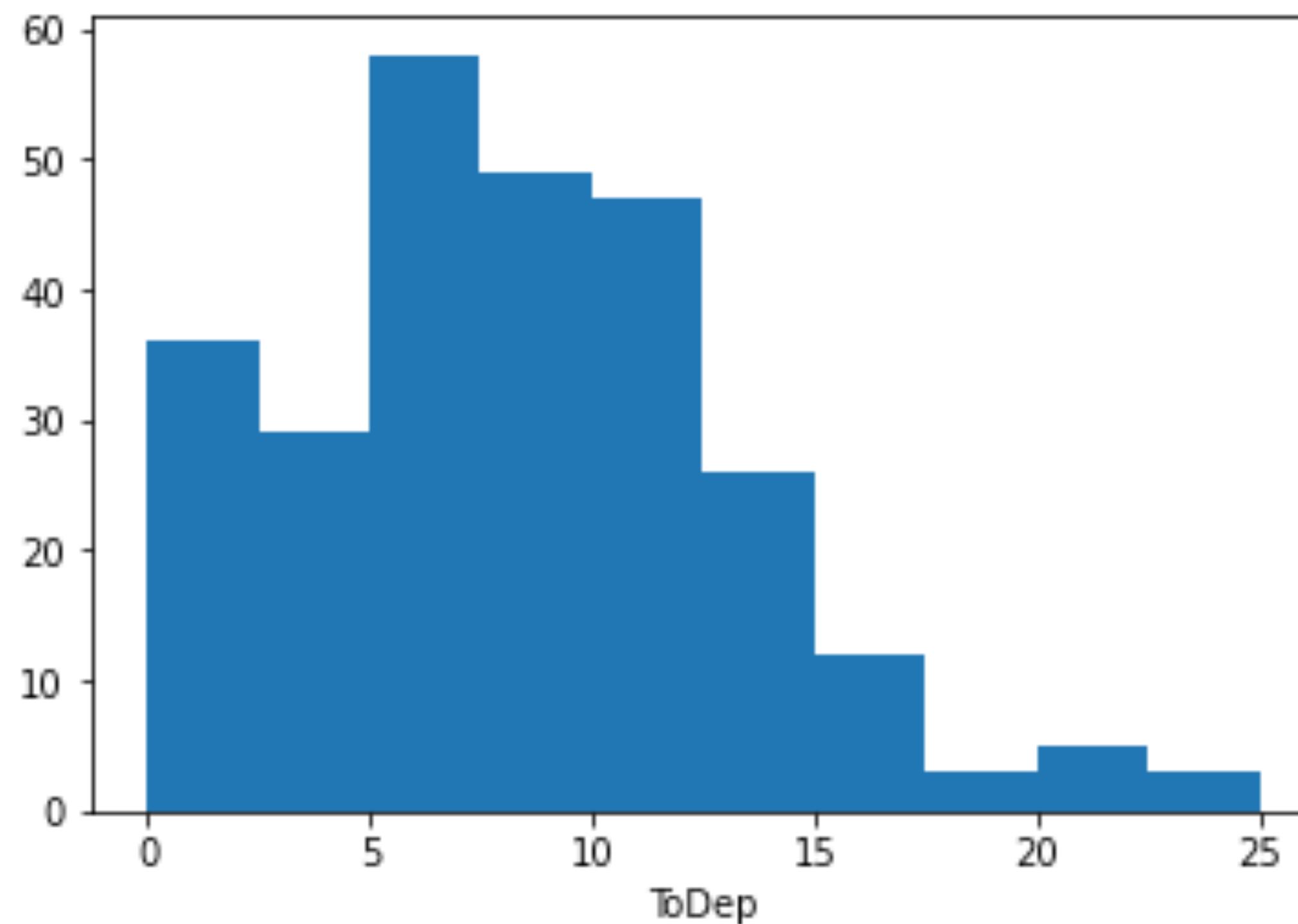
*Splitting the Data
into categorical and
numerical variables*



*Converting categorical
variables into dummy
variables*

EXPLORATORY DATA ANALYSIS - TARGET VARIABLE

Figure 1: Histogram of Depression Scores



- Patient Health Questionnaire-9 (PHQ-9)
- Self-administered questionnaire of 9 questions
 - Monitor severity of depression and response to treatment
 - Can be used to make a tentative diagnosis of depression in at-risk population
 - Scale from (1) to (4)

EXPLORATORY DATA ANALYSIS - TARGET VARIABLE

Over the last 2 weeks, how often have you been bothered by any of the following problems? 過去二週間に、次の問題のどれに悩まされましたか？ *

1. Not at all ぜんぜん
ありません 2. Several days 数日 3. More than half the days 1週間以上 4. Nearly everyday 每日

Little interest or pleasure in doing things 仕事をする興味がない

Feeling down, depressed, or hopeless 落ち込み、絶望

Trouble falling or staying asleep, or sleeping too much 寝にくい / 寝すぎ

Feeling tired or having little energy 疲労感

Poor appetite or overeating 食欲がない / 食べすぎ

Score (0) - (3)

Depression Severity

1 - 5

None

6 - 10

Mild

11 - 15

Moderate

16 - 20

Moderately Severe

21 - 28

Severe

EXPLORATORY DATA ANALYSIS - PREDICTOR VARIABLES

Socio Demographic Variables

- Gender
- Home country
- Academic Level
- Length of stay
- Religion
- Intimate Partner
- Japanese and English Proficiency

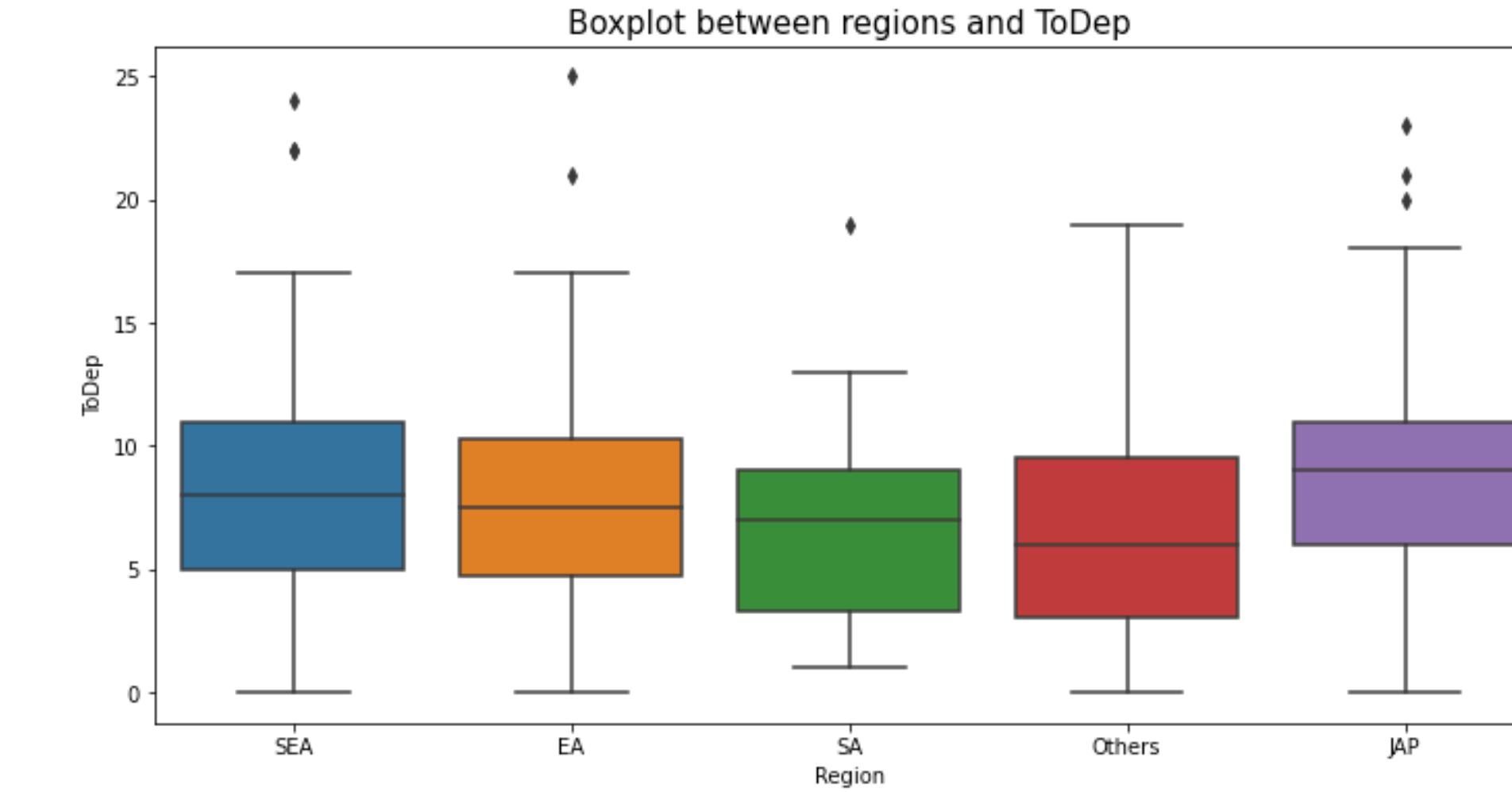
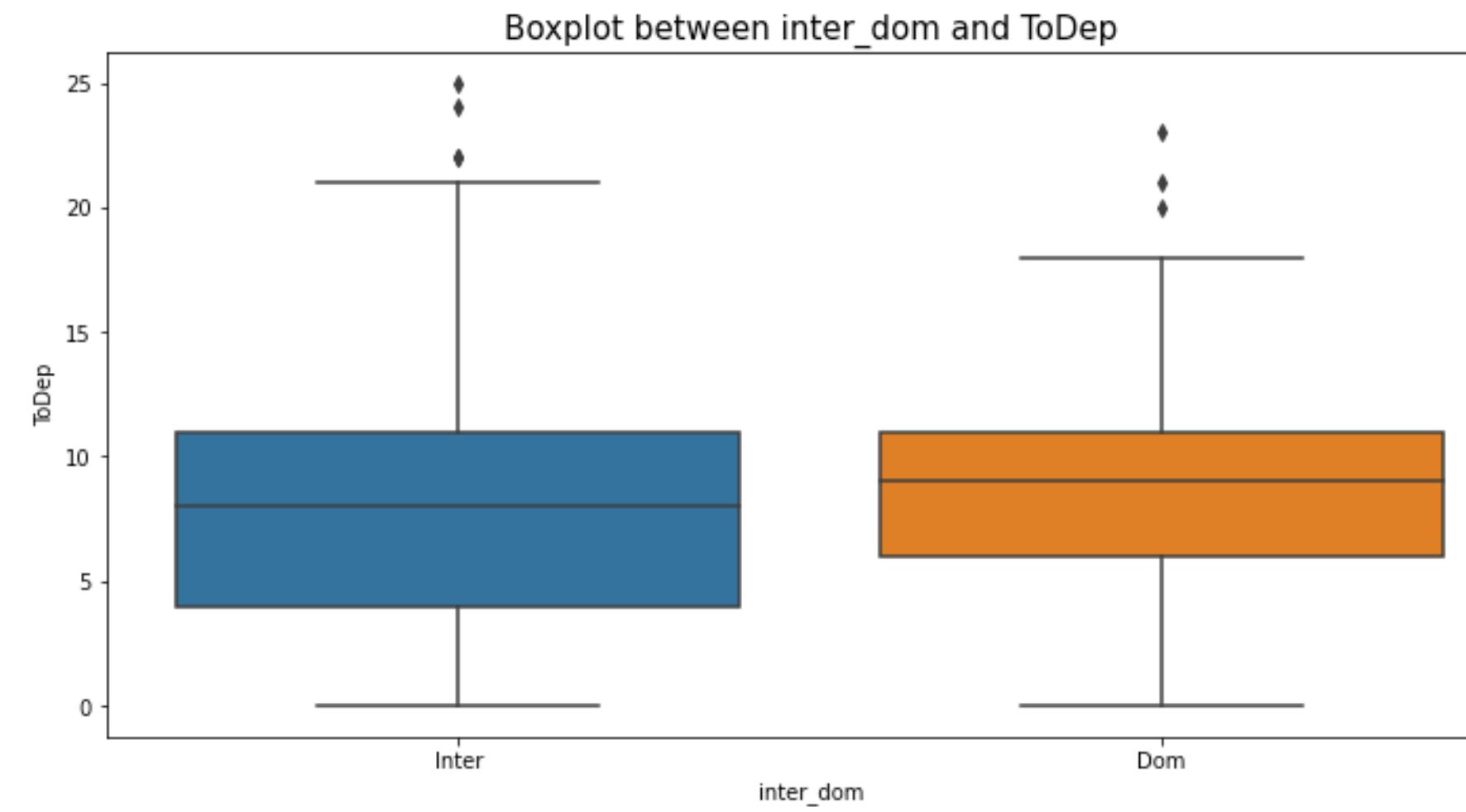
Mental Health Conditions

- Suicide Ideation
- Acculturation Stress
 - Perceived Discrimination
 - Homesickness
 - Perceived Hatred
 - Fear
 - Cultural Shock
 - Guilt

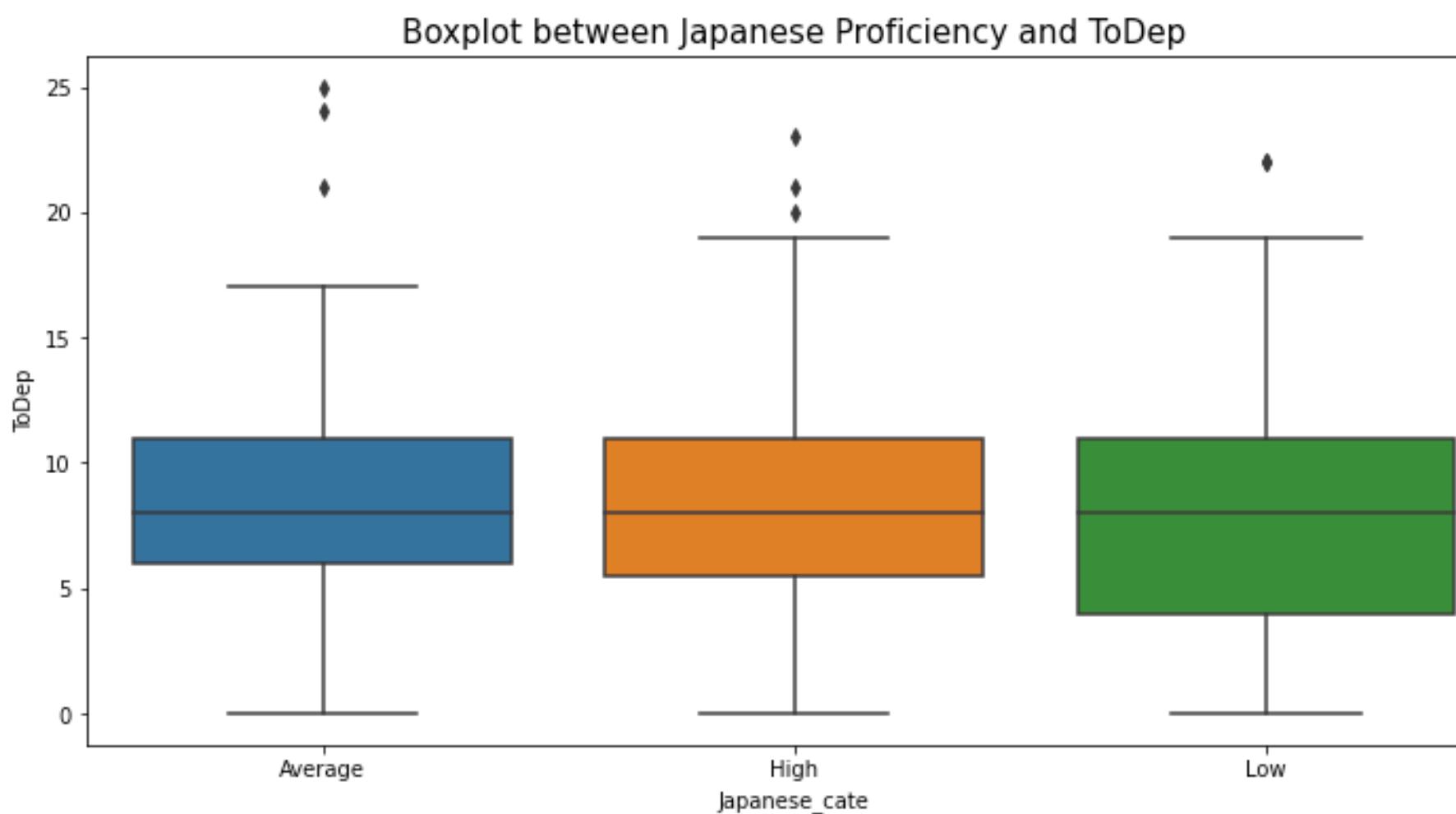
Help Seeking Behaviors

- Partner
- Friends
- Parents
- Relative
- Professional
- Phone
- Doctor
- Religion
- Alone
- Internet

EXPLORING THE SOCIO DEMOGRAPHIC VARIABLES

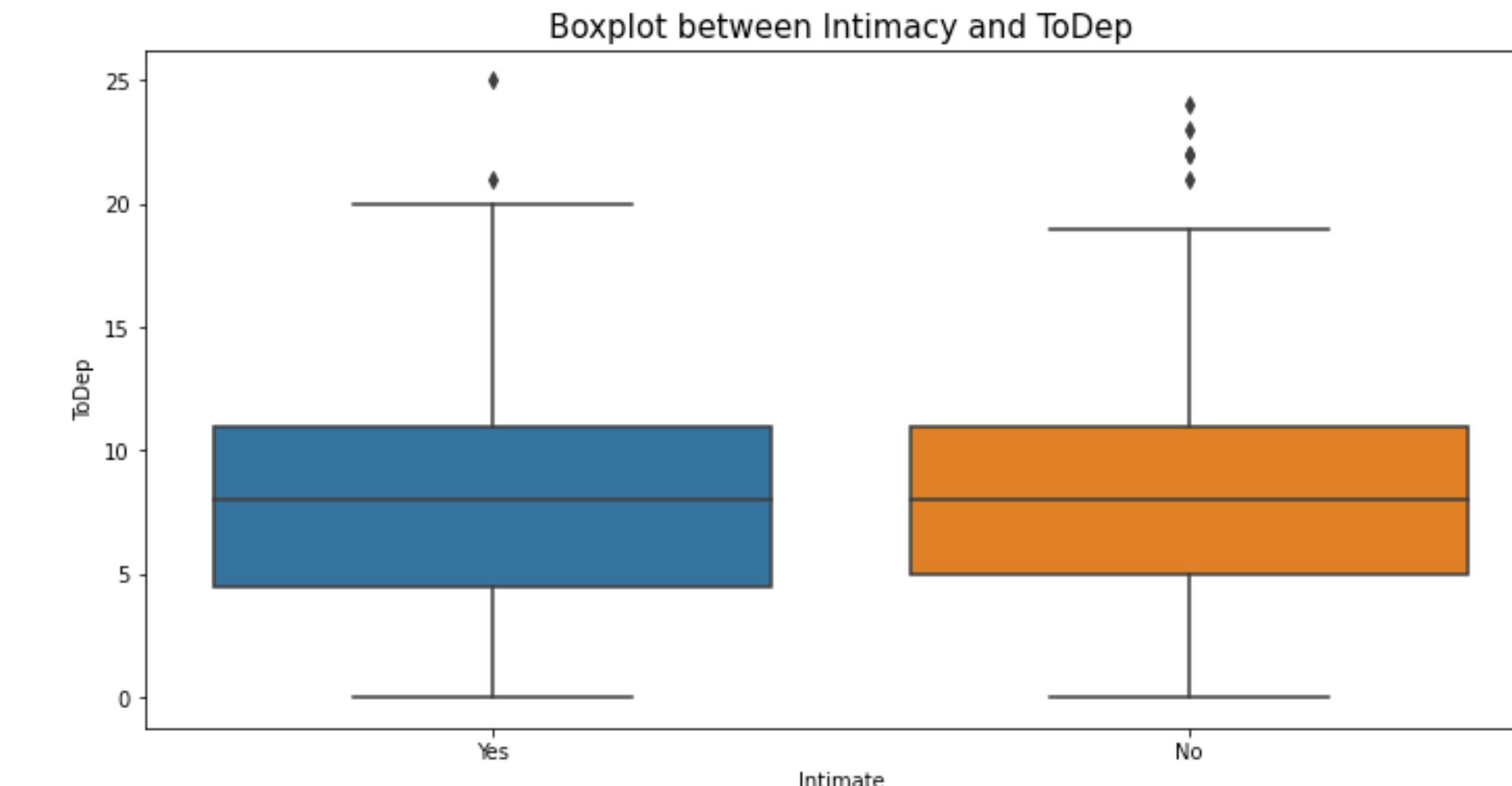


Domestic Vs International



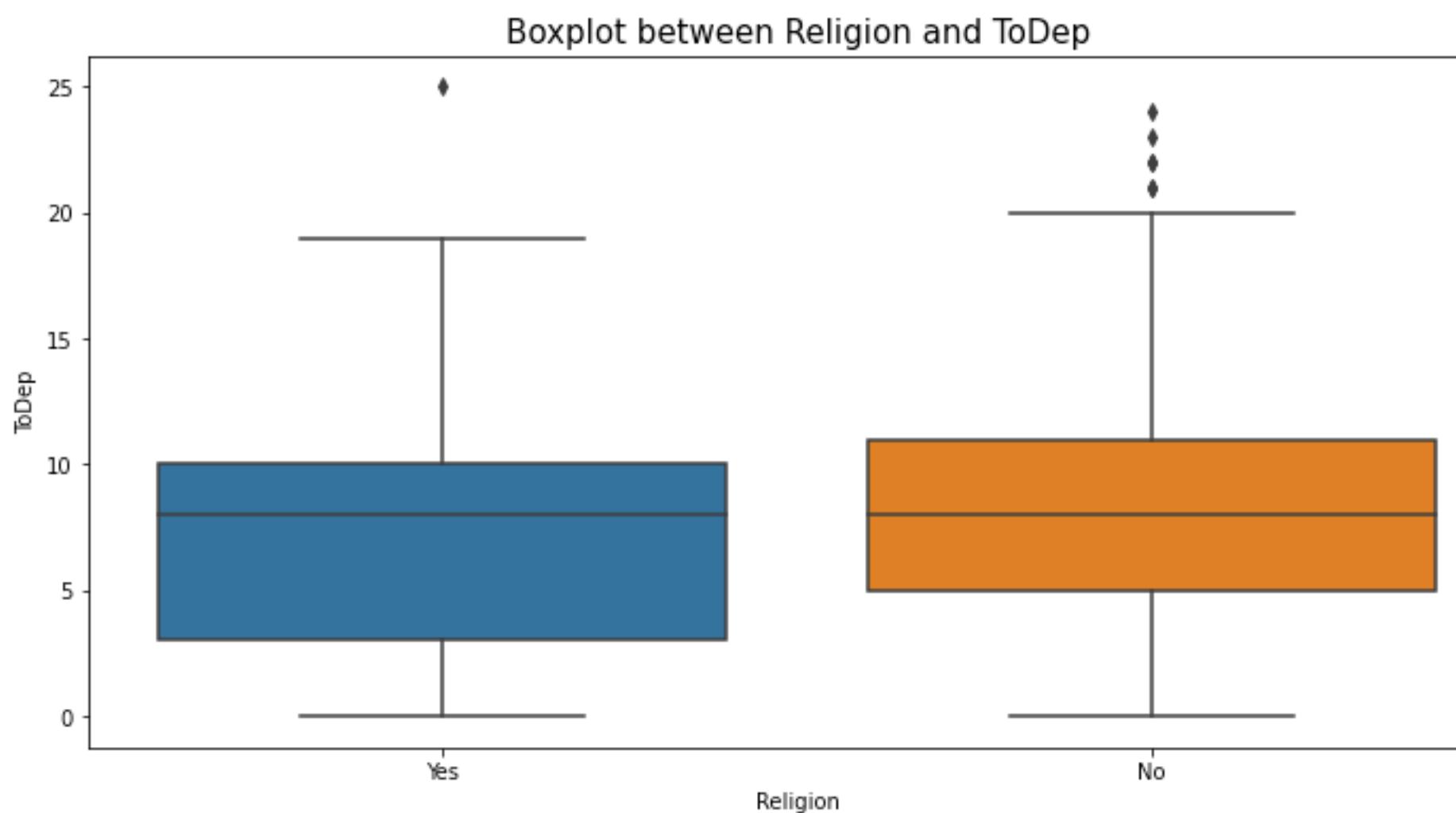
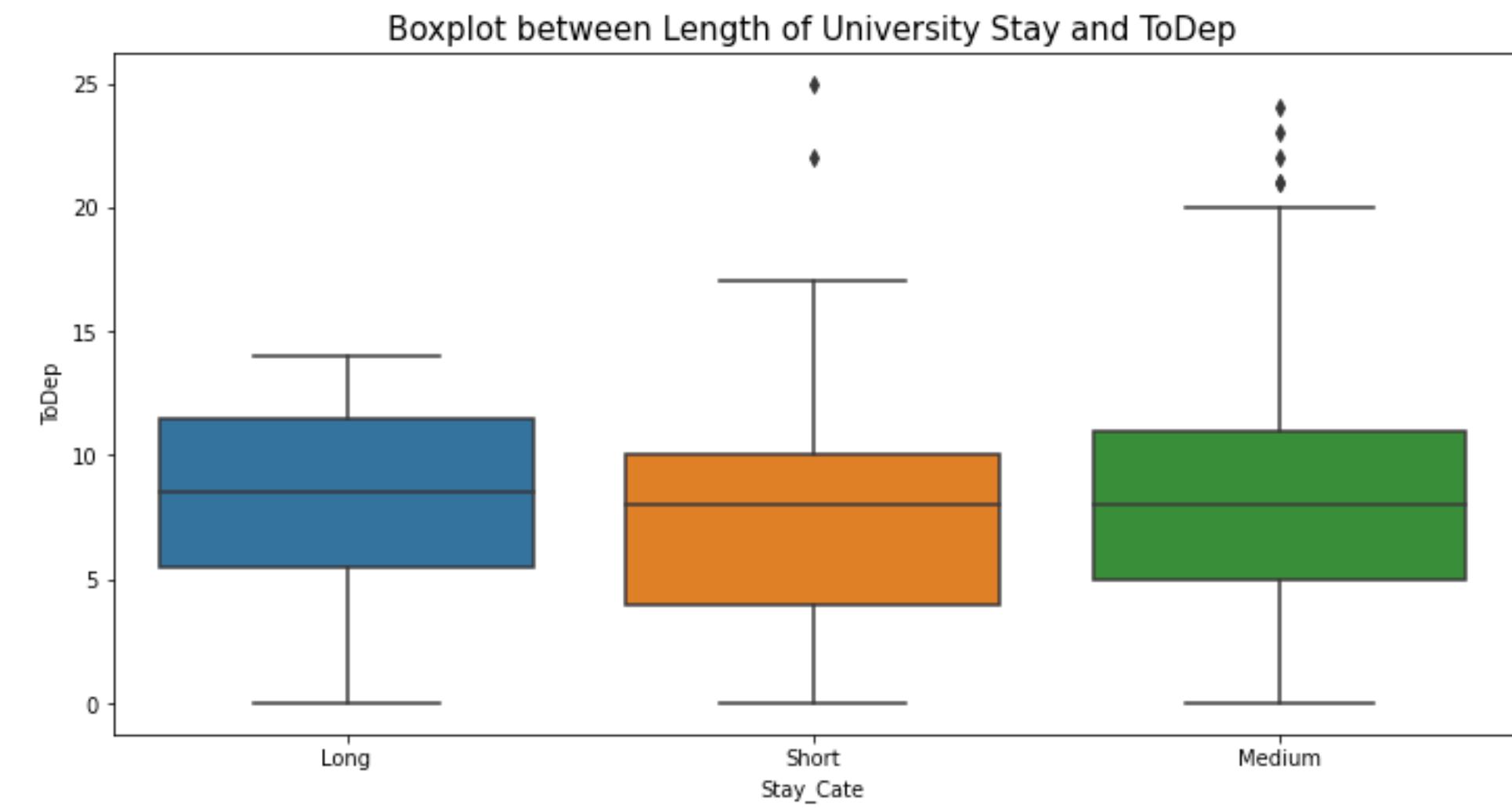
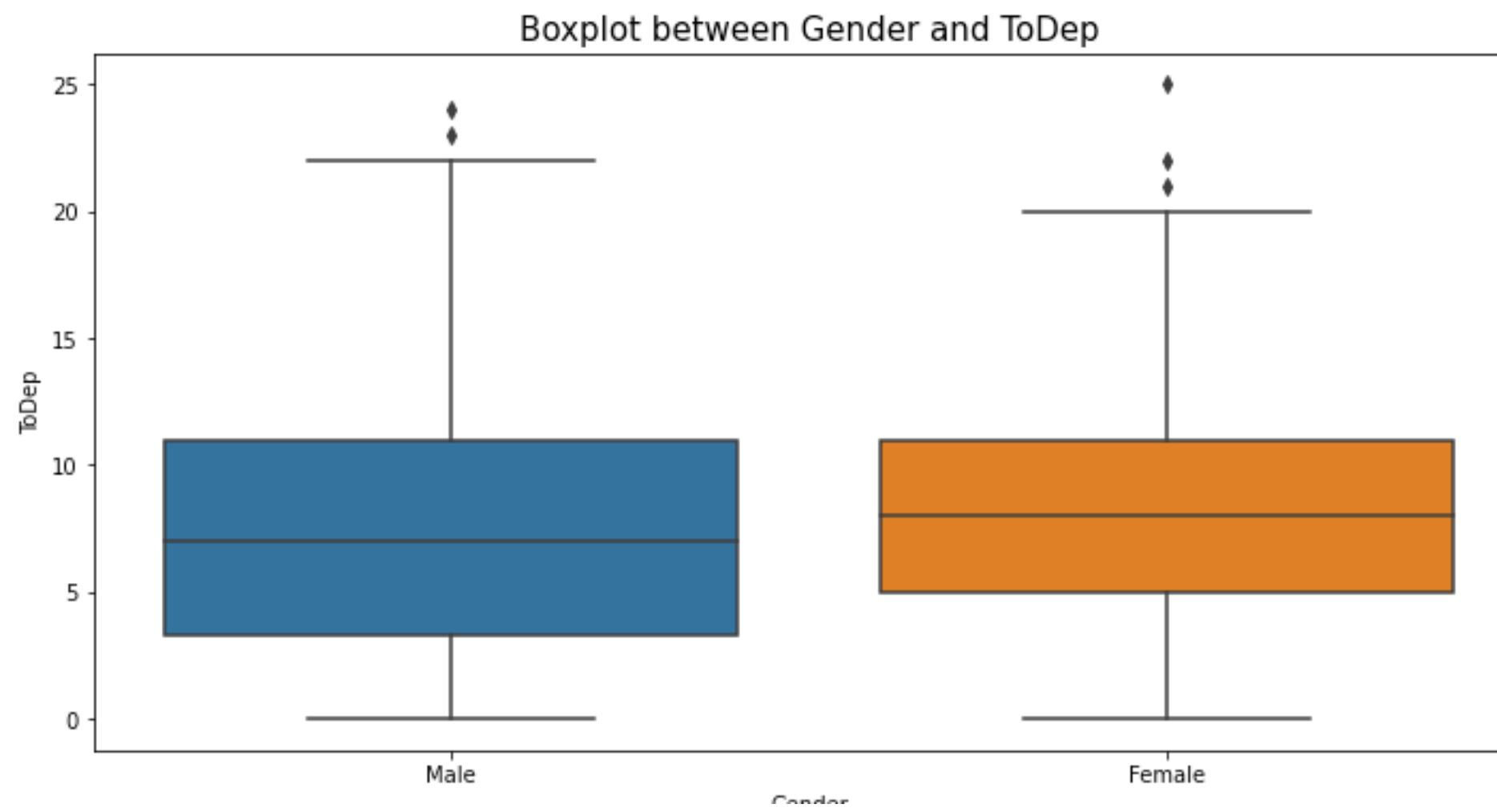
Japanese Proficiency

Regions



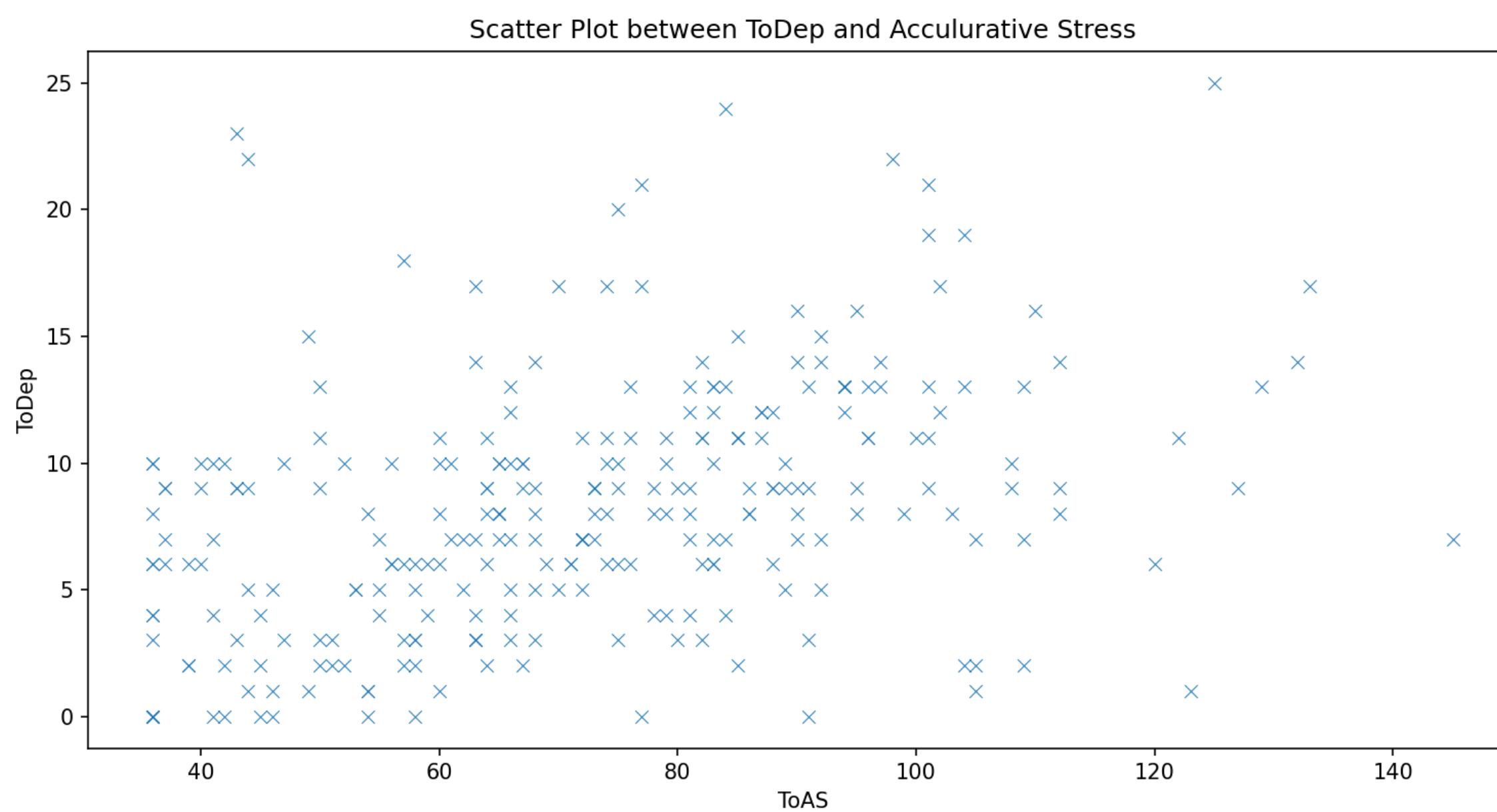
Intimate Partner

EXPLORING THE SOCIO DEMOGRAPHIC VARIABLES



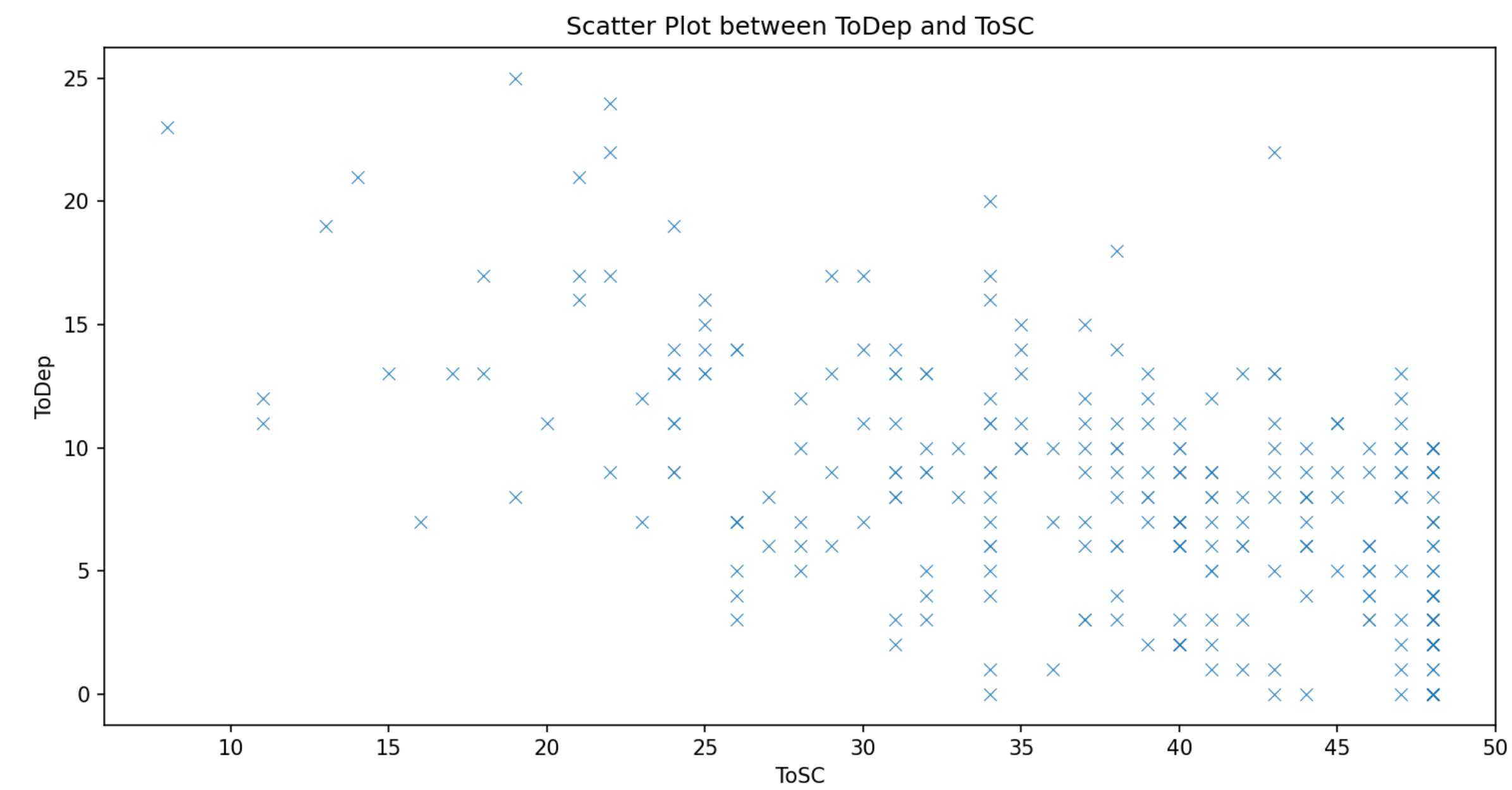
EXPLORING THE MENTAL HEALTH CONDITION VARIABLES

Depression and Acculturative Stress



*Correlation = 0.39
(Statistically significant)*

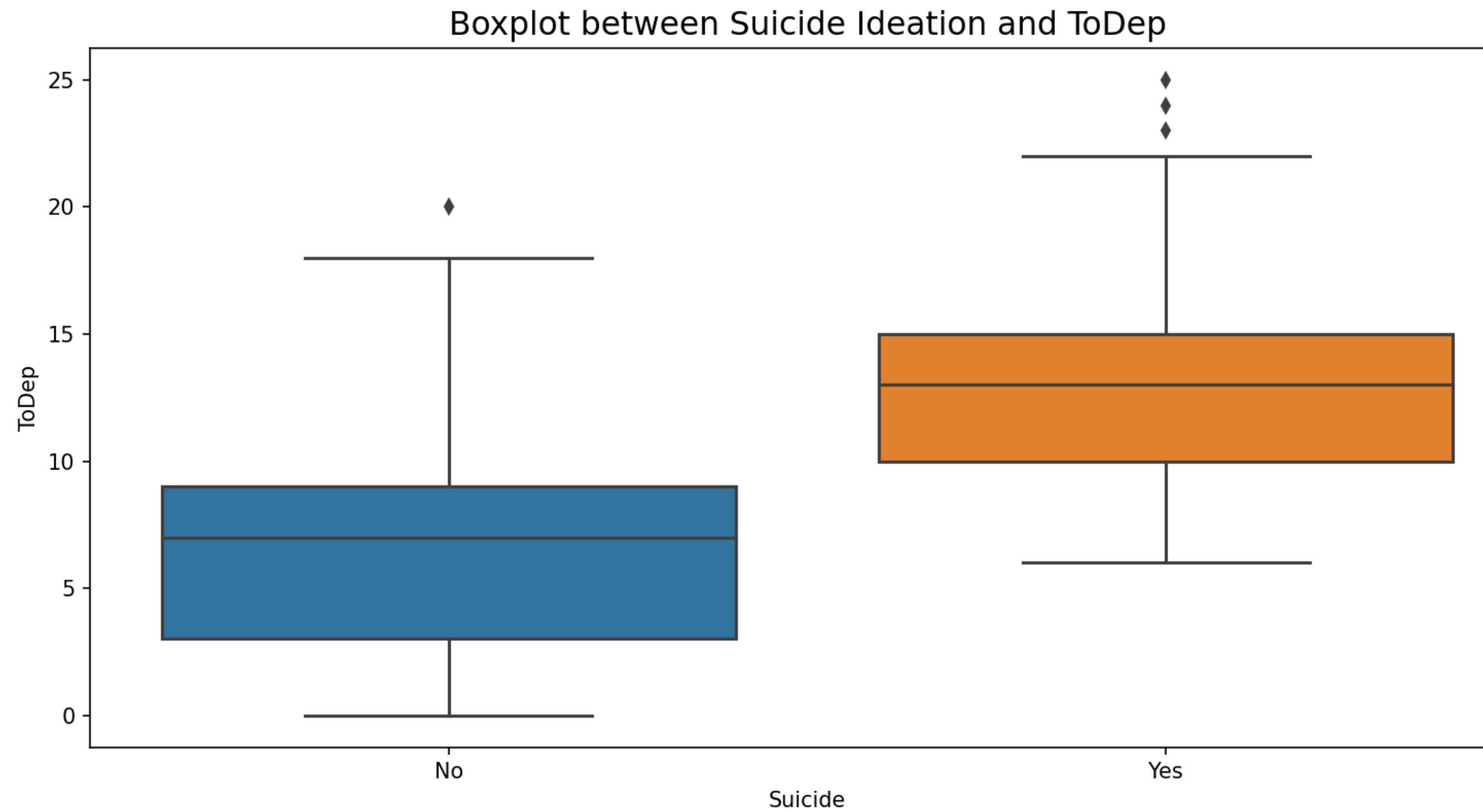
Depression and Social Connectedness



*Correlation = -0.55
(Statistically significant)*

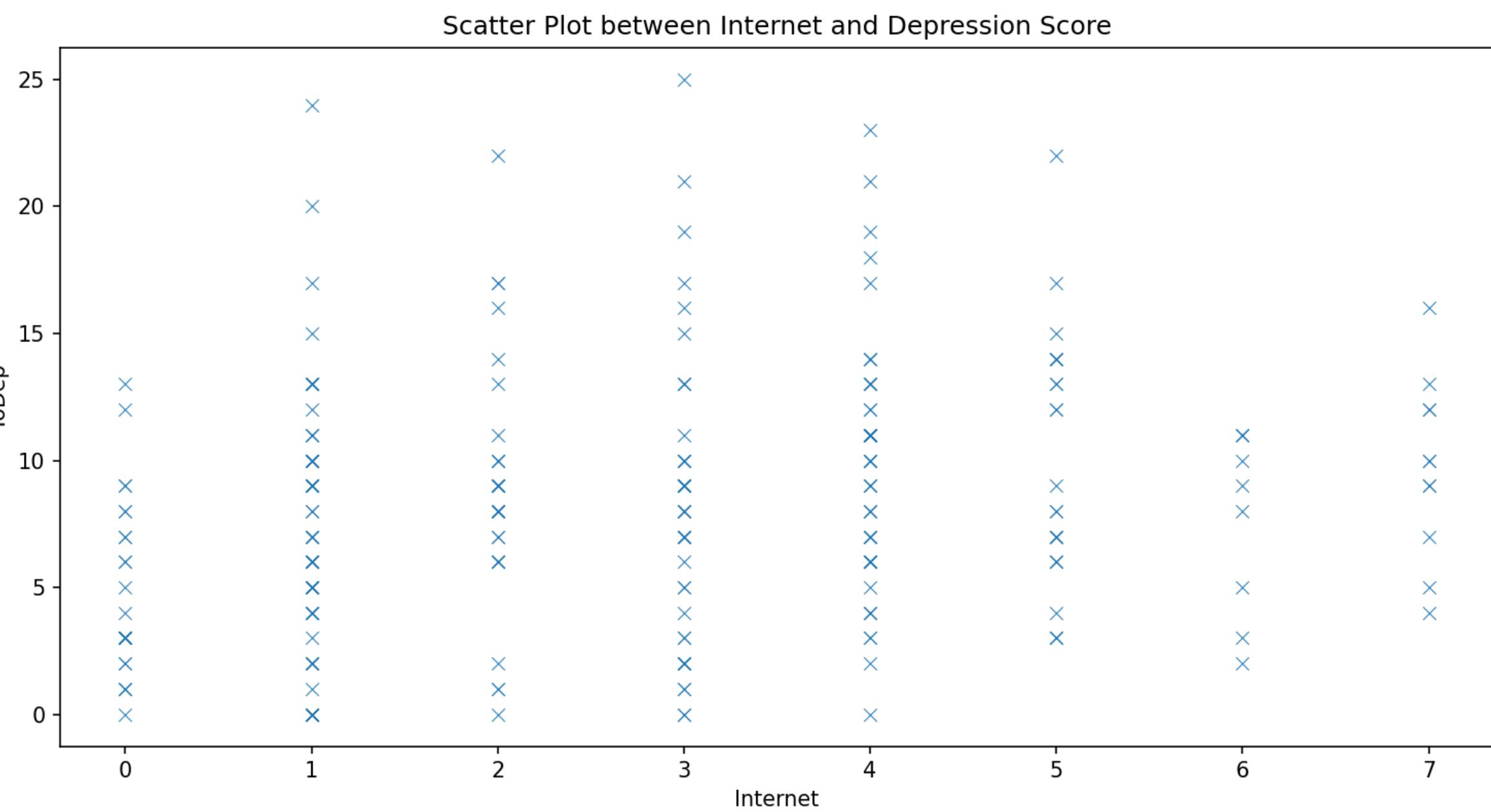
EXPLORING THE MENTAL HEALTH CONDITION VARIABLES

Suicide Ideation

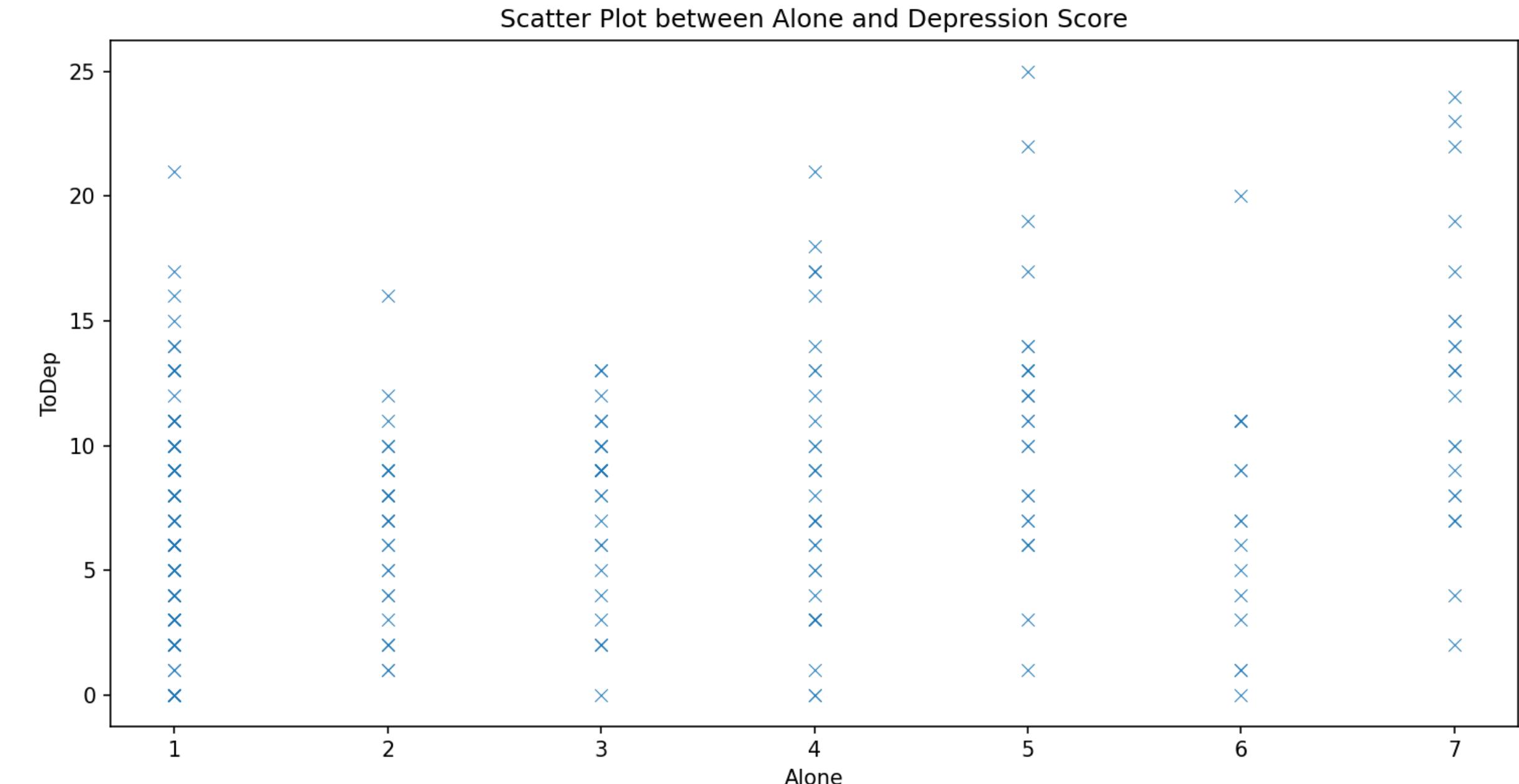


Difference in the means are statistically significant

EXPLORING HELP SEEKING BEHAVIOR VARIABLES



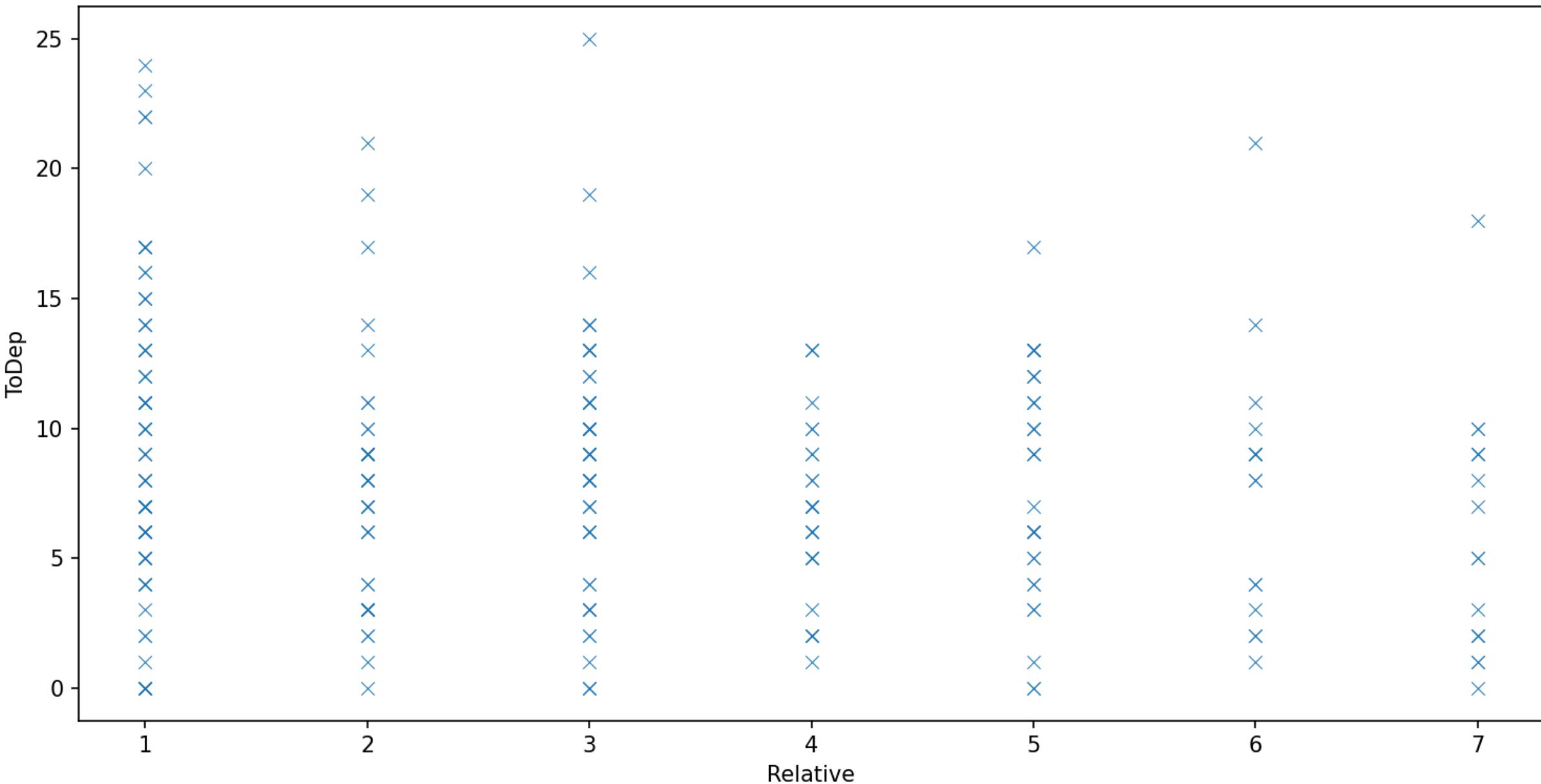
Correlation = 0.24
(Statistically significant)



Correlation = 0.31
(Statistically significant)

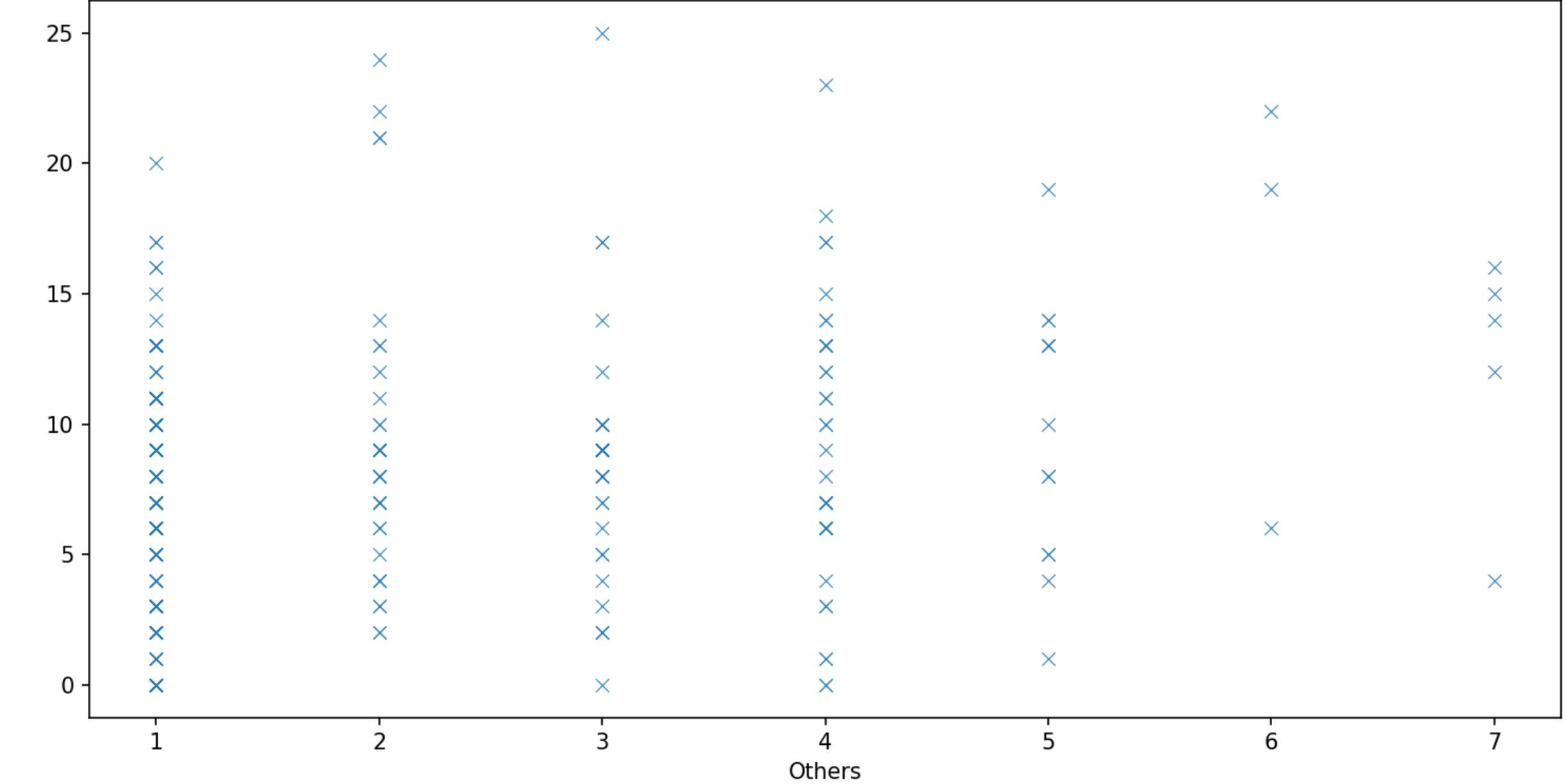
EXPLORING HELP SEEKING BEHAVIOR VARIABLES

Scatter Plot between Relative and Depression Score



*Correlation = -0.15
(Statistically significant)*

Scatter Plot between Others and Depression Score



*Correlation = 0.24
(Statistically significant)*

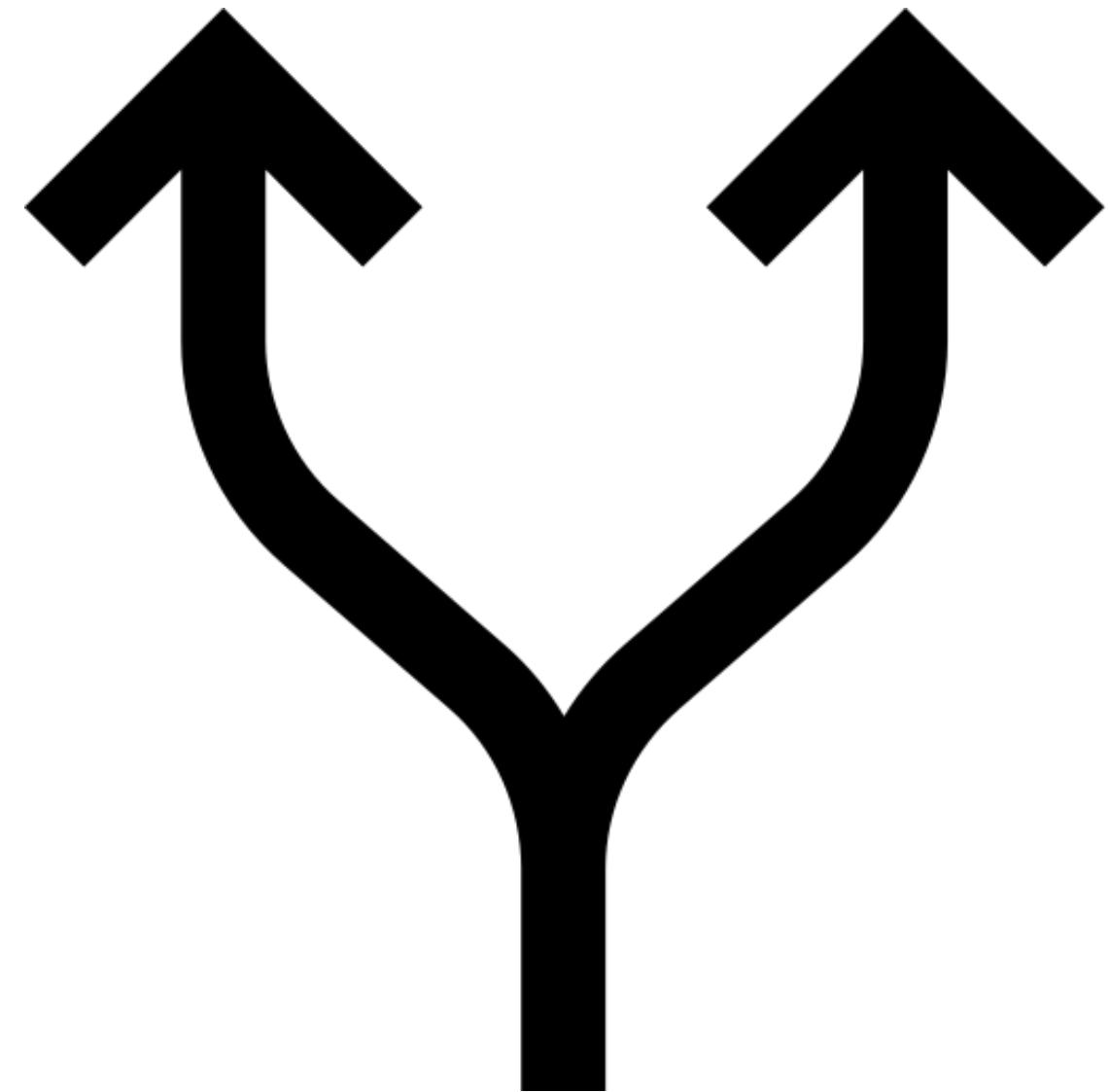


SUMMARY

- Students with suicide ideation more likely to have higher depressions scores
- Though some variables are statistically significant, their correlation is quite weak, which warrants further exploration

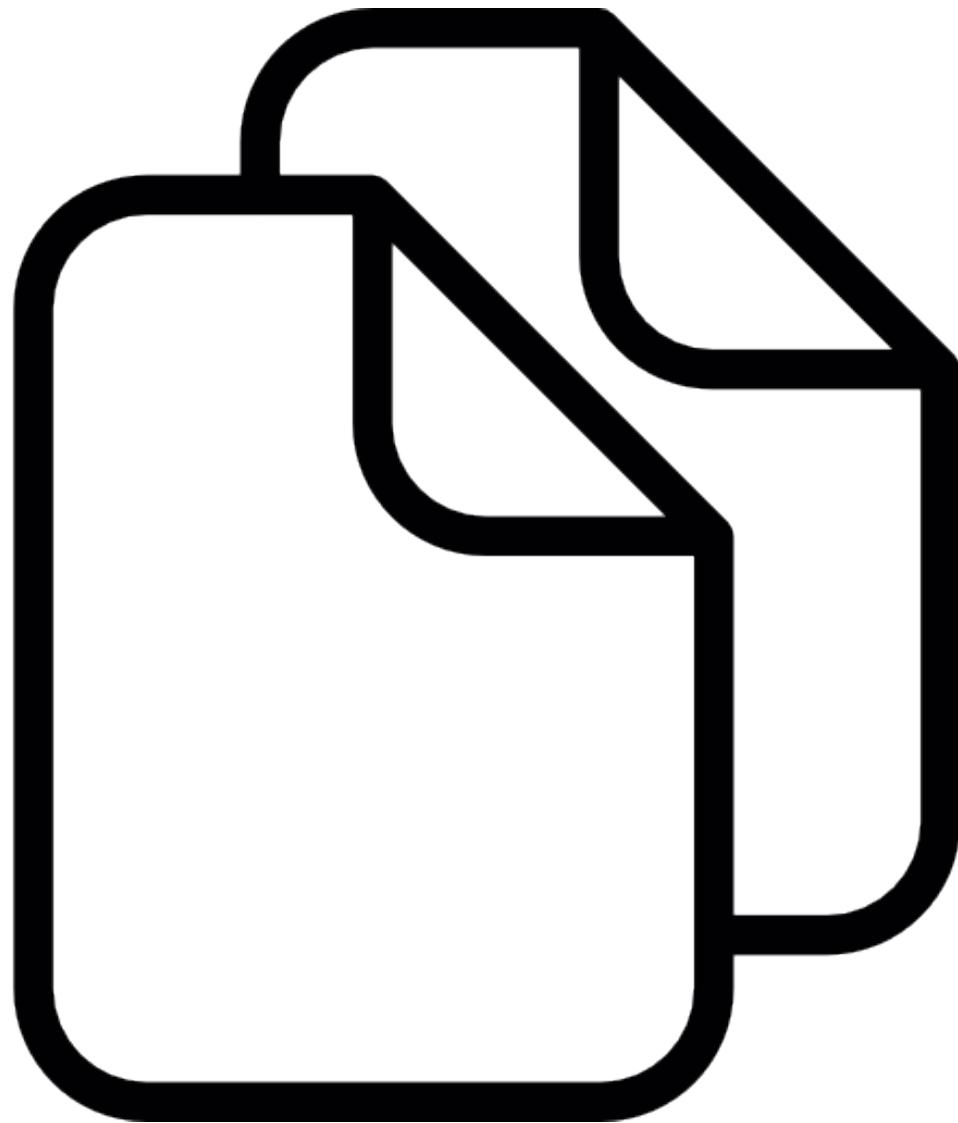
Source: *The Perspective Project*

MACHINE LEARNING METHODOLOGY



*Splitting the data into
test and training data*

Test size = 0.2



*Linear Regression
Modelling using
Forward Feature
Selection*



*10-Fold Validation
Ridge and Lasso
Regularisation*

LINEAR REGRESSION MODEL



*Target Variable:
Total Depression Score*

1. Include all features as predictor variables
 - Exclude features that are derived from the PHQ-9 questionnaire
2. Use forward feature selection to eliminate features until the top features that give good predictive power are selected

FINAL FEATURES AND EVALUATION

Top Features Selected

1. Suicide Ideation
3. Internet help-seeking behaviours
5. Total Social Connectedness score

Model Evaluation:

Adjusted R Square: 042

42% of the variance in total depression scores can be explained by these features

K-FOLD VALIDATION

- Splitting the dataset into 10 groups and fit each group

	r2_train	r2_test	rmse_train	rmse_test
0	0.410570	0.631592	3.810583	2.884136
1	0.443175	0.175742	3.797768	3.066741
2	0.419858	0.515141	3.724184	3.771276
3	0.438652	0.383322	3.635622	4.479429
4	0.450354	0.095383	3.724841	3.754219
5	0.426985	0.446382	3.606343	4.687061
6	0.447235	0.173997	3.738481	3.641305
7	0.426335	0.459967	3.699575	3.966275
8	0.428633	0.463724	3.792153	3.051272
9	0.438602	0.368700	3.691589	4.051512

- Metrics for test data seem to always perform worse than the training data
- R Square for the test data are **consistent** for some groups

LASSO AND RIDGE REGULARISATION

Purpose

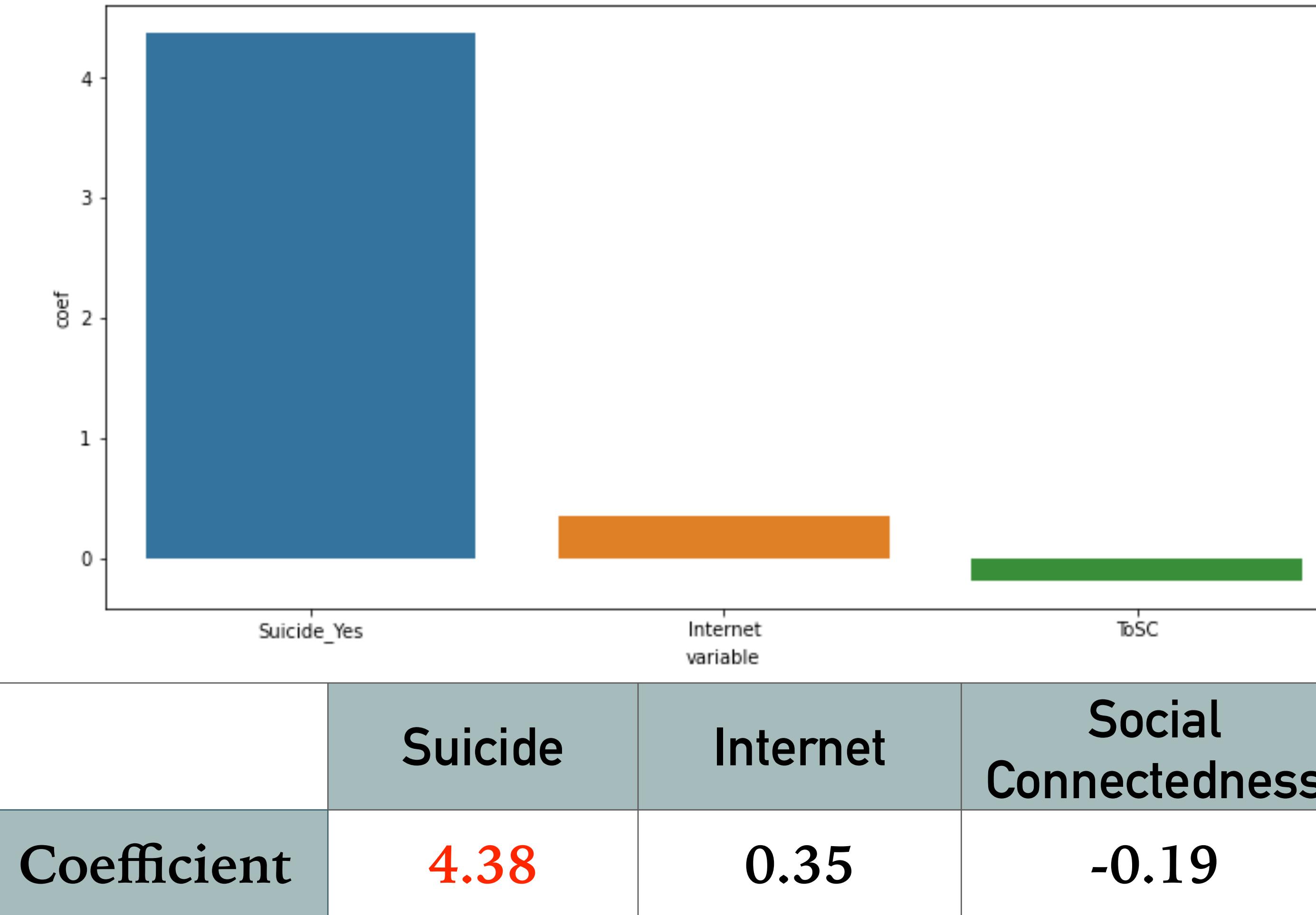
Improve the prediction accuracy and interpretability of our model

	Ridge	Lasso
R Square	0.43	0.43

- Both techniques did not really **change** much
- **43% of the variance** in total depression scores can be explained by these features

INTERPRETATION OF FINAL MODEL

Coefficients of predictor variables for final model



- Having suicide ideation will increase the total depression score by 4.38 points
- A 1-point increase in Internet Help Seeking Behavior will increase the total depression score by 0.35 points
- A 1-point increase in the total social connectedness score will decrease the total depression score by 0.19 points



RECOMMENDATIONS

- Having suicide ideation is a good indicator of students at risk of depression
- Larger sample size could lead to better predictive features
- Will be more ideal to build a classification model to classify students-at-risk

A photograph of three yellow smiley faces. One large smiley face is positioned in the upper left, another large one is in the lower right, and a smaller one is partially visible at the bottom left. All three have black eyes and a black curved line for a smile.

**End of
Presentation**