Final Report

Iciar Fernandez Boyano, Jacob Gerlofs

The Mental Health Toll of Graduate School

Introduction

For the past five years, the iconic science journal Nature has launched a survey for PhD students in STEM fields to share their experience in graduate school, hoping to illuminate the goals, challenges, and sources of satisfaction for doctoral students across seven continents. Last year's survey collected data from over 6000 graduate students, which constitutes the highest response rate in the survey's history. The full data from the survey was made publicly available following publication of an article discussing the results. It is interesting to note that the survey was offered in English, Spanish, Chinese, French, and Portuguese - open-form questions have not been translated to English if answered by the participant in another language. Available materials include anonymysed raw data, the questionnaire that was provided to PhD students, and a presentation of the survey data.

In our project, we aim to investigate the relationship between these two question areas (mental health & feelings of harrasment/bullying) and other variables that we hypothesise may be related to positive and/or negative outcomes. For example, are those pursuing a degree far from home more likely to suffer from anxiety and depression? Are instances of harrasment and/or bullying male-biased? In an effort to shed some light into the matter, we will study these questions in detail.

Research question

Rather than a single question, the many variables available as part of our dataset has allowed us to investigate several relationships.

1. **PhD Satisfaction.** We hypothesise that the level of satisfaction that a graduate student may feel with their decision to pursue a PhD program may be associated with (1) their university's long hours culture, (2) their work/life balance, and (3) their relationship with their supervisor.

2. **Suffering from anxiety or depression caused by PhD studies.** We are interested in seeing whether this variable is influenced by (1) the student's relationship with their supervisor, and (2) studying outside of their home country.

3. **Suffering from discrimination or harrassment**. We have investigated the relationship of this variable with (1) studying outside of your home country, and (2) the student's gender. ## Data and methods

Data Description

According to the script with survey information that was provided, there were a total of 65 questions. Not all questions were mandatory, and there was a mix of single choice (yes/no), multiple choice (several options) and free-form questions.

In the dataset, each row represents an individual who participated in the survey, whereas each row represents a question. We have noticed some redundancy in the dataset column that will require substantial cleanup of the data as part of our next project milestone. For instance, Q12 ("What prompted you to study outside your country of upbringing?") was presented in the survey as a multiple choice question with 11 possible answers (a-k), with the last one (k) being open-form ("If other, please specify"). In the data frame, 11 rows correspond to Q12, each one composed of 2 values: NA, and 1/11 possible answers. As such, the column named Q12_1 only contains NA values and answer "(a) To study at a specific university"; whereas Q12_2 only contains NA values and answer "(b) Lack of funding opportunities in my home country", and so on. We plan on combining columns Q12_1:Q12_11 into a single Q12 column using dplyr::coalesce(), following the same rationale for other redundant columns in the dataset. In addition, open-form questions such as (k) in this specific example will be dropped due to the difficulty in analyzing this, and the fact that they contain answers in different languages.

Due to this redundancy, the dimensions of the raw dataset when downloaded are 6812 rows (participants) by 274 columns (questions), whereas the actual survey only has 63 questions. Below is the complete list of questions, which is a simplified version of the Word document provided here, which includes all the possible answers for each question. For simplicity, we have only included the question, its type, and the category it belongs to within the survey.

```
## [1] 63  4
```

Question_No	Section	Question
1	Questionnaire	Which, if any, of the following degrees are you currently studying
2	Questionnaire	Which was the most important reason you decided to enrol in a
3	Questionnaire	Are you studying in the country you grew up in?
4	Questionnaire	Where do you currently live?
5	Questionnaire	Which country in Asia?
6	Questionnaire	Which country in Australasia?
7	Questionnaire	Which country in Africa?
8	Questionnaire	Which country in Europe?
9	Questionnaire	Which country in North or Central America?
10	Questionnaire	Which country in South America?
11	Questionnaire	What prompted you to study outside your country of upbringing
12	Questionnaire	Do you have a job alongside your studies?
13	Questionnaire	What is your main reason for having a job?
14	PhD Highs and Lows	What concerns you the most since you started your PhD?
14a	PhD Highs and Lows	Is there anything else not mentioned that has concerned you sinc
15	PhD Highs and Lows	Overall, what do you enjoy most about life as a PhD student?
16	PhD Highs and Lows	How satisfied are you with your decision to pursue a PhD?
17	Satisfaction with your PhD experience	How satisfied are you with your PhD experience?
18	Satisfaction with your PhD experience	Since the very start of your graduate school experience, would yo
19	Satisfaction with your PhD experience	How satisfied are you with each of the following attributes or asp
20	Satisfaction with your PhD experience	To what extent does your PhD programme compare to your origi
21	Your programme	On average, how many hours a week do you typically spend on y
22	Your programme	On average, how much one-on-one contact time do you spend wit
23	Your programme	Overall, how would you describe the academic system, based on y
24	Your programme	To what extent do you agree or disagree with the following stater
25	Your programme	Have you ever sought help for anxiety or depression caused by Pl
26	Your programme	Did you seek help for anxiety or depression within your institutic
27	Your programme	To what extent do you agree or disagree with the following stater
28	Mental health and discrimination	Do you feel that you have experienced bullying in your PhD prog
29	Mental health and discrimination	Who was the perpetrator(s)?
29a	Mental health and discrimination	Do you feel able to speak out about your experiences of bullying
30	Mental health and discrimination	Do you feel that you have experienceddiscrimination or harassme
31	Mental health and discrimination	Which of the following have you experienced?
32	Future career plans	How much do you expect your PhD to improve your job prospect
33	Future career plans	Which of the following sectors would you most like to work in (b
34	Future career plans	Please use the scale below to indicate how likely you are to pursu
35	Future career plans	If you're unlikely to pursue an academic research career, what ar
36	Future career plans	What position do you most expect to occupy immediately after y
37	Future career plans	What type of career you are interested in pursuing after your gra
38	Career expectations	After completing your PhD, how long do you think it will take yo
39	Career expectations	How much more likely are you now to pursue a research career th
40	Career expectations	What is the main reason why you are more likely to pursue a res
41	Career expectations	How did you arrive at your current career decision?
42	Career support	How do you learn about available career opportunities that are b
43	Career support	Which of the following 3 things would you say are the most diffic
44	Career support	Which of the following would you say are the most difficult for P
45	Career support	Which of the following resources do you think PhD students nee
46	Career support	How well is your programme preparing you to carry out each of t
47	Career support	To what extent do you agree or disagree with the following stater
48	Career support	Which, if any, of the following activities have you done to advanc
49	Career support	Which of the following social media networks have you used to b
50	Reflection	What would you do differently right now if you were starting you
51	Reflection	With the benefit of hindsight, what one thing do you know now v
52	Reflection	What is your age?
53	Reflection	Are you. . . (Gender)
54	Reflection	Which of the following best describes you? (Ethnicity)
55	Reflection	Do you have any caring responsibilities?

After careful consideration, we have only used certain variables for our analysis. Below is a dataframe with these variables:

Variables	Type	Description
Gender	factor	Female (including trans female) / Male (including trans ma
Studying in your home country	factor	Yes / No
Level of satisfaction with PhD	factor	1-5 Scale (1 - Very dissatisfied / 5 - Very satisfied)
Supervisor Relationship	factor	1-7 Scale (1 - Not at all satisfied / 7 - Extremely satisfied
Work Life Balance	factor	1-5 Scale (1 - Strongly disagree / 5 - Strongly agree
University Long Hours Culture	factor	1-5 Scale (1 - Strongly disagree / 5 - Strongly agree
Anxiety or Depression caused by PhD	factor	Yes / No / Prefer not to say
Experienced Bullying in PhD	factor	Yes / No / Prefer not to say
Experienced Discrimination or Harrassment in PhD	factor	Yes / No / Prefer not to say

Methods

Instead of linear regression, we have chosen logistic regression as the statistical method to investigate the relationships that we have outlined above in our "Research Question" section for our data. Logistic regression is used when the dependent variable is categorical, which is our case. Linear regression is not suitable for a classification problem because it is unbounded.

Results

Prior to logistic regression, we performed some exploratory data analysis in our data as part of milestone 1.

Exploratory data analysis

General Demographics of Graduate Students in the Survey   While the survey was available in several languages, which we assume was an attempt to improve outreach, it is clear that the respondents origin is highly biased to those located in Europe, closely followed by Asia and North America.

Logistic regression

We then perform logistic regression of the relationships that we attempt to investigate, described in the **research question** section.

LR1: Long Hours Culture ~ PhD Satisfaction   Interestingly, there isn't a clear trend in the relationship between long hours culture & satisfaction. It is true that those dissatisfied with their decision (1 & 2 in the x axis) seem to have a higher probability of their university having a long hours culture, but aside from that, there is no clear pattern.
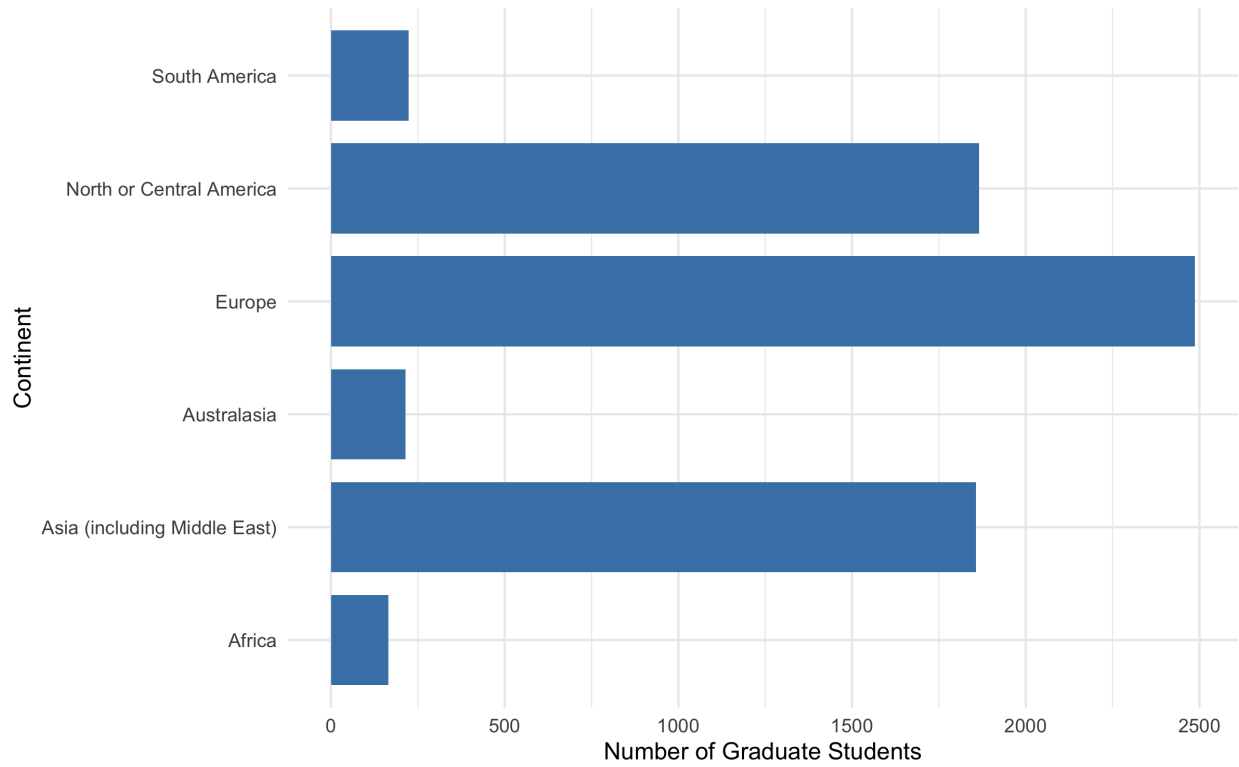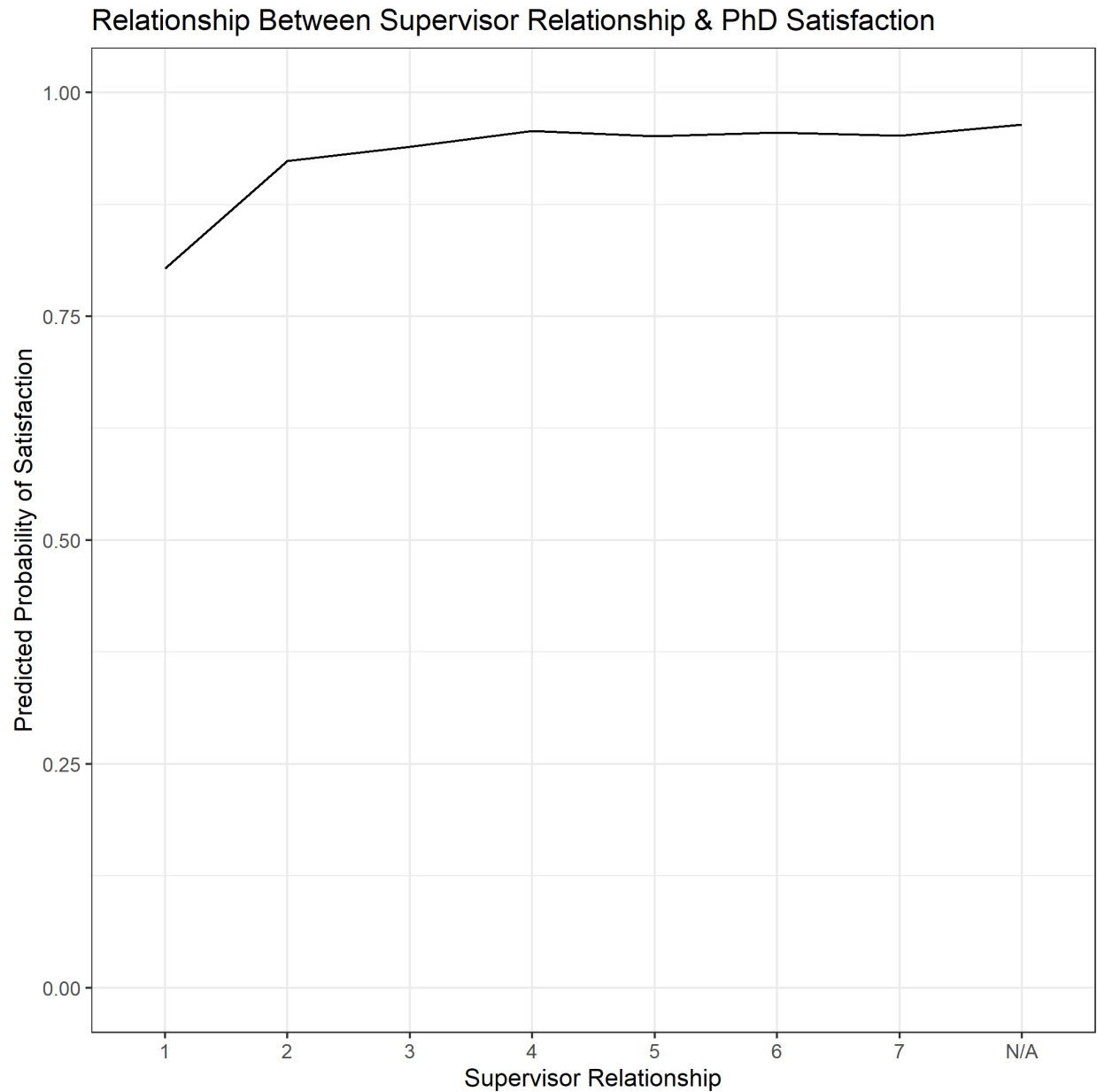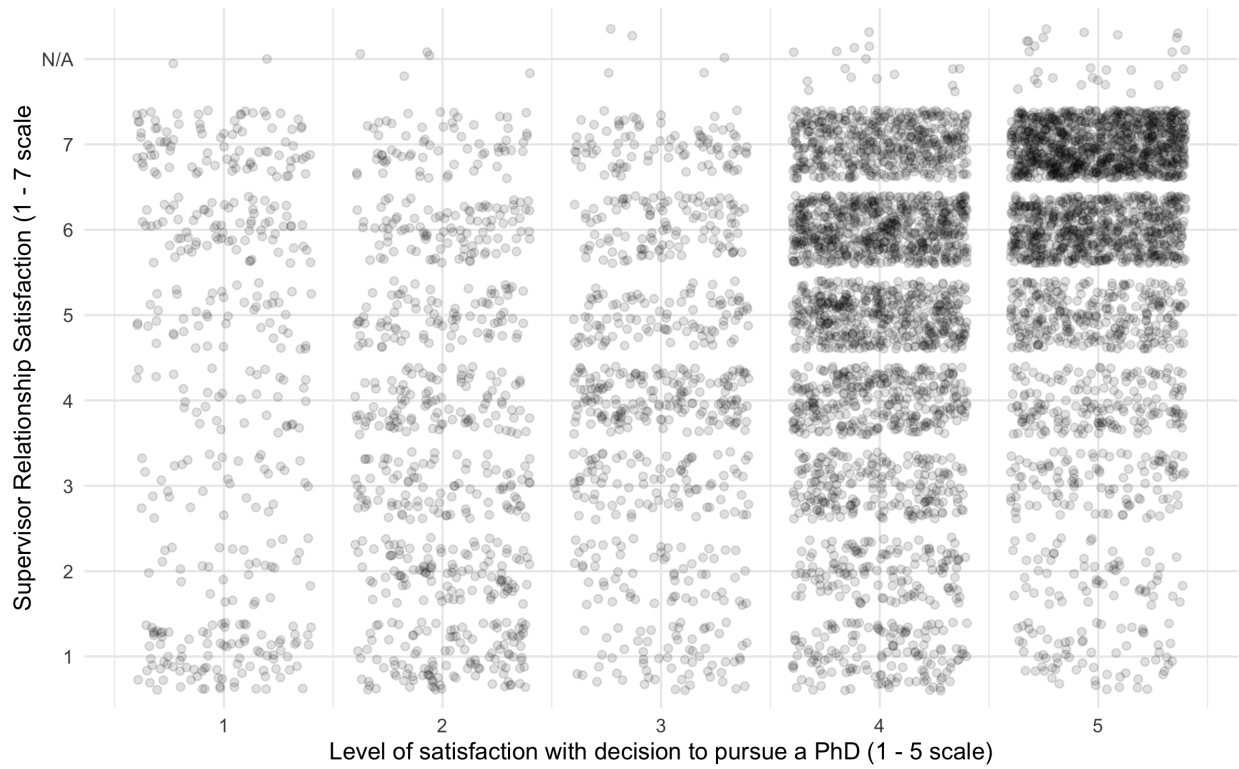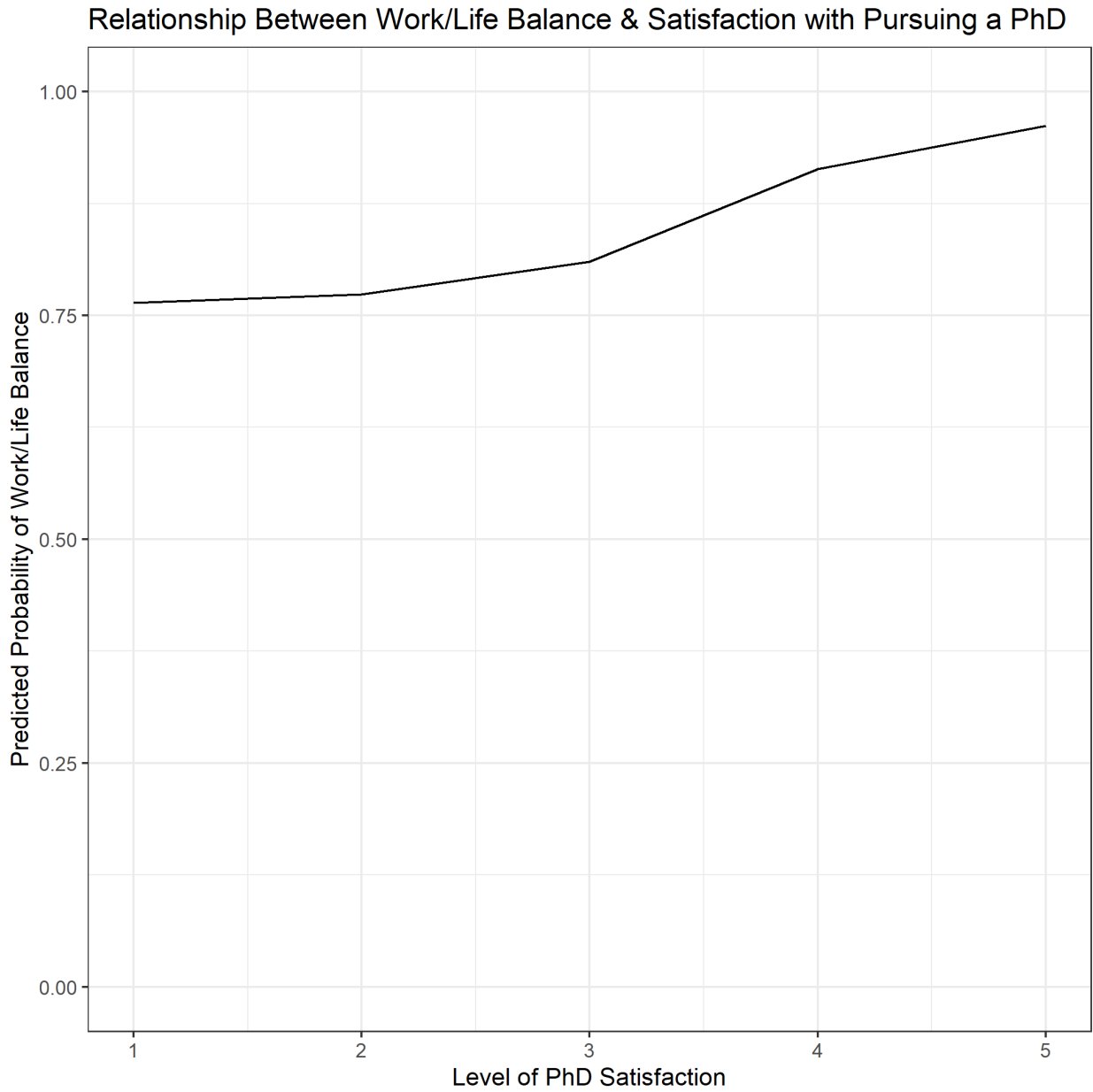
Figure 1: Basic Demographics

Relationship Between Long Hours Culture & Satisfaction with Pursuing a PhD
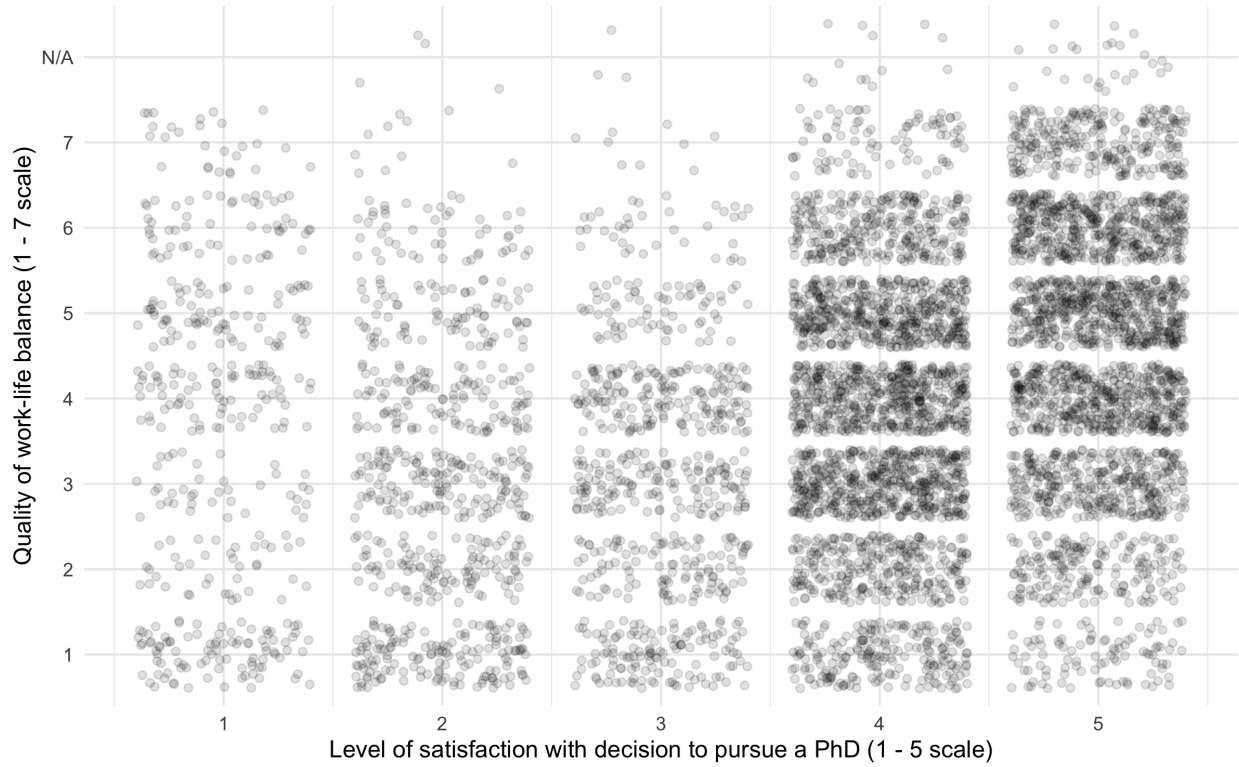
LR2: Supervisor Relationship ~ PhD Satisfaction It appears that probability of satisfaction does increase as the supervisor relationship goes from very dissatisfied (1) to very satisfied (7). However, the trend is not as pronounced as one may expect in this case either. Findings in the second graph match those that result from logistic regression, and indeed confirm that students that have a good relationship with their supervisor are more satisfied with their decision to pursue a PhD than those that don't.
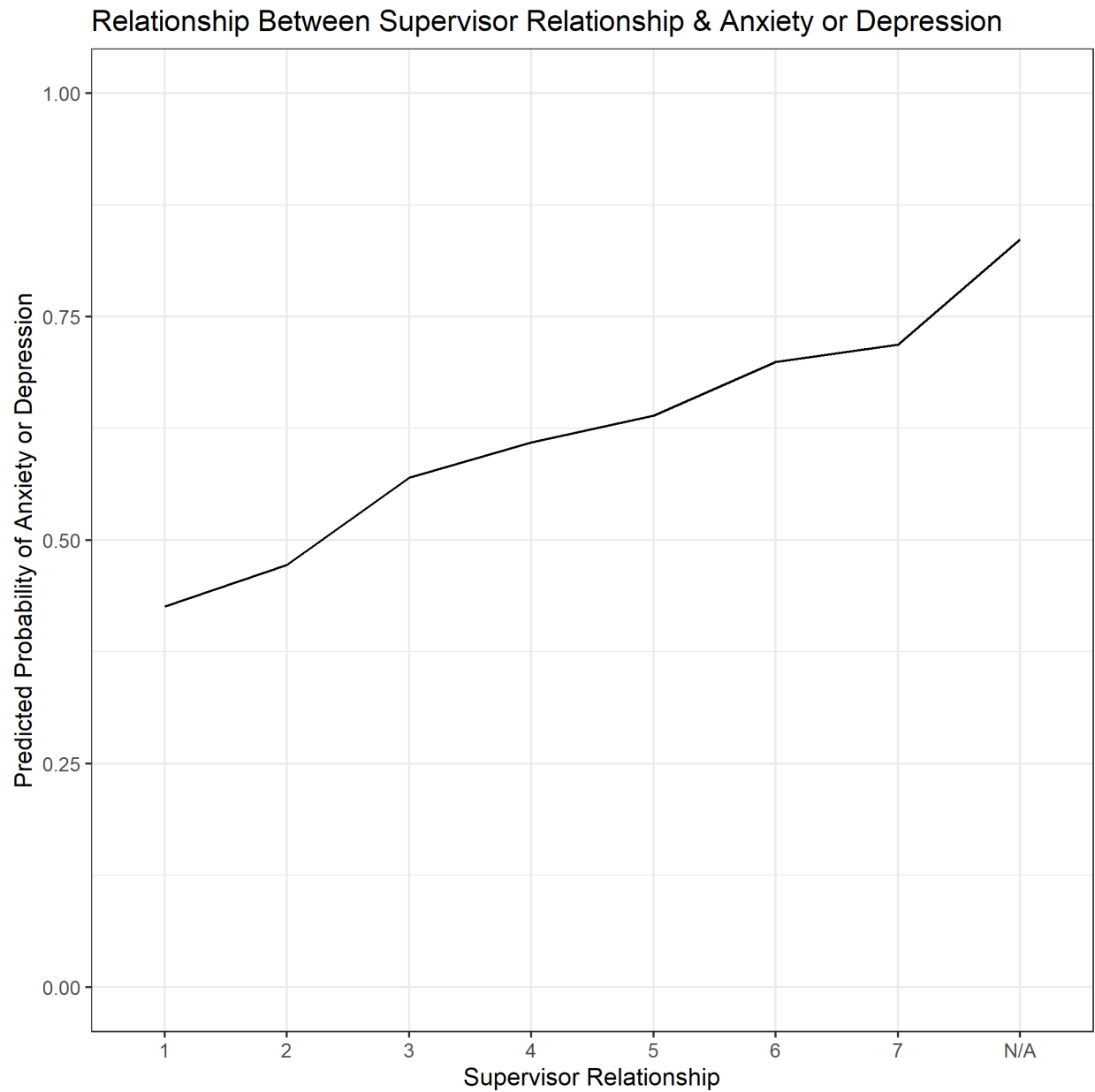

Relationship Between Supervisor Relationship & PhD Satisfaction

LR3: Work/Life Balance ~ PhD Satisfaction Out of all of our analysis, this one shows one of the most pronounced relationships - there is a strong trend between students that are more satisfied with their decision to pursue a PhD and a probability of having a good work/life balance, which stays true in our jitter graph.
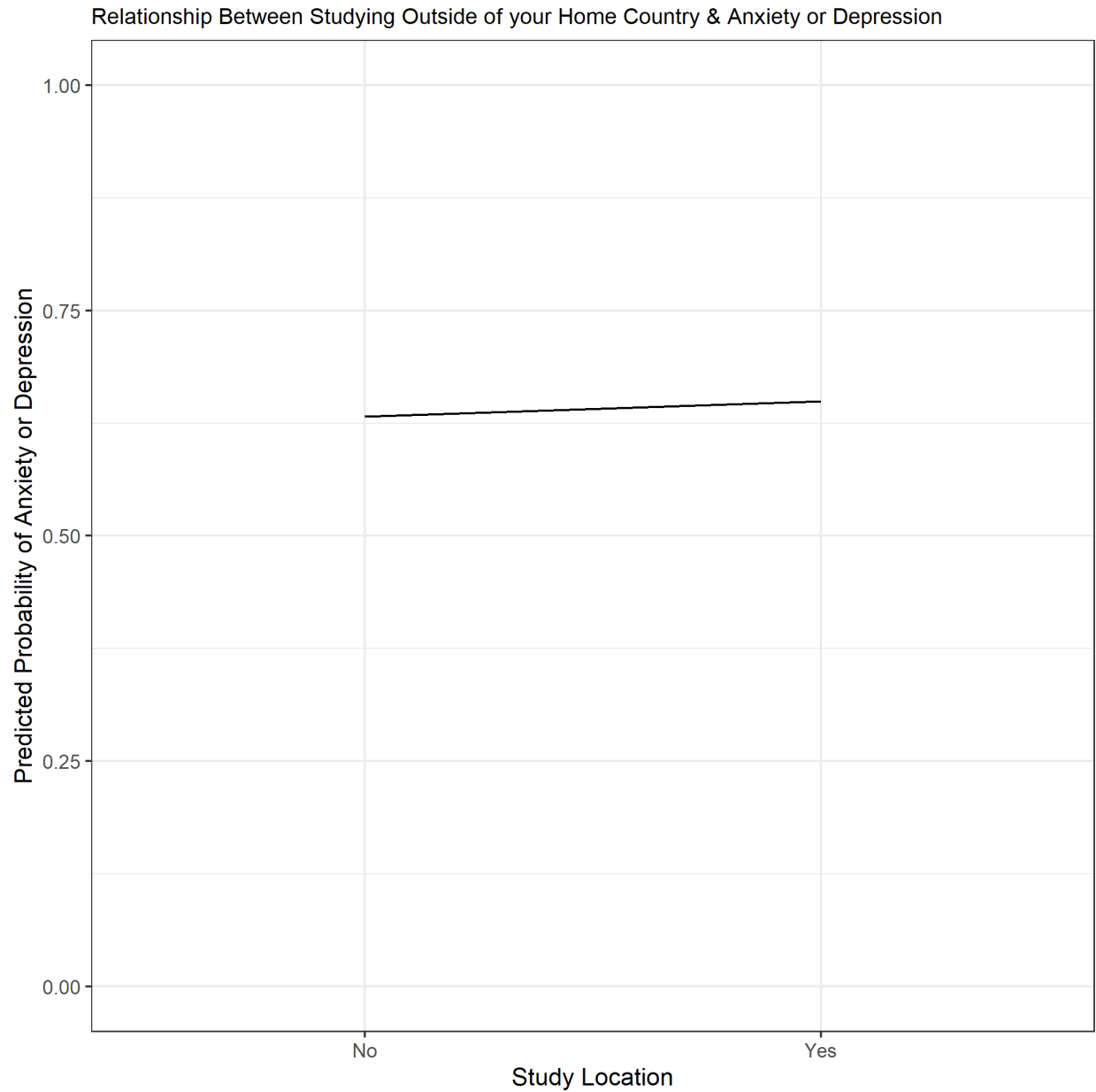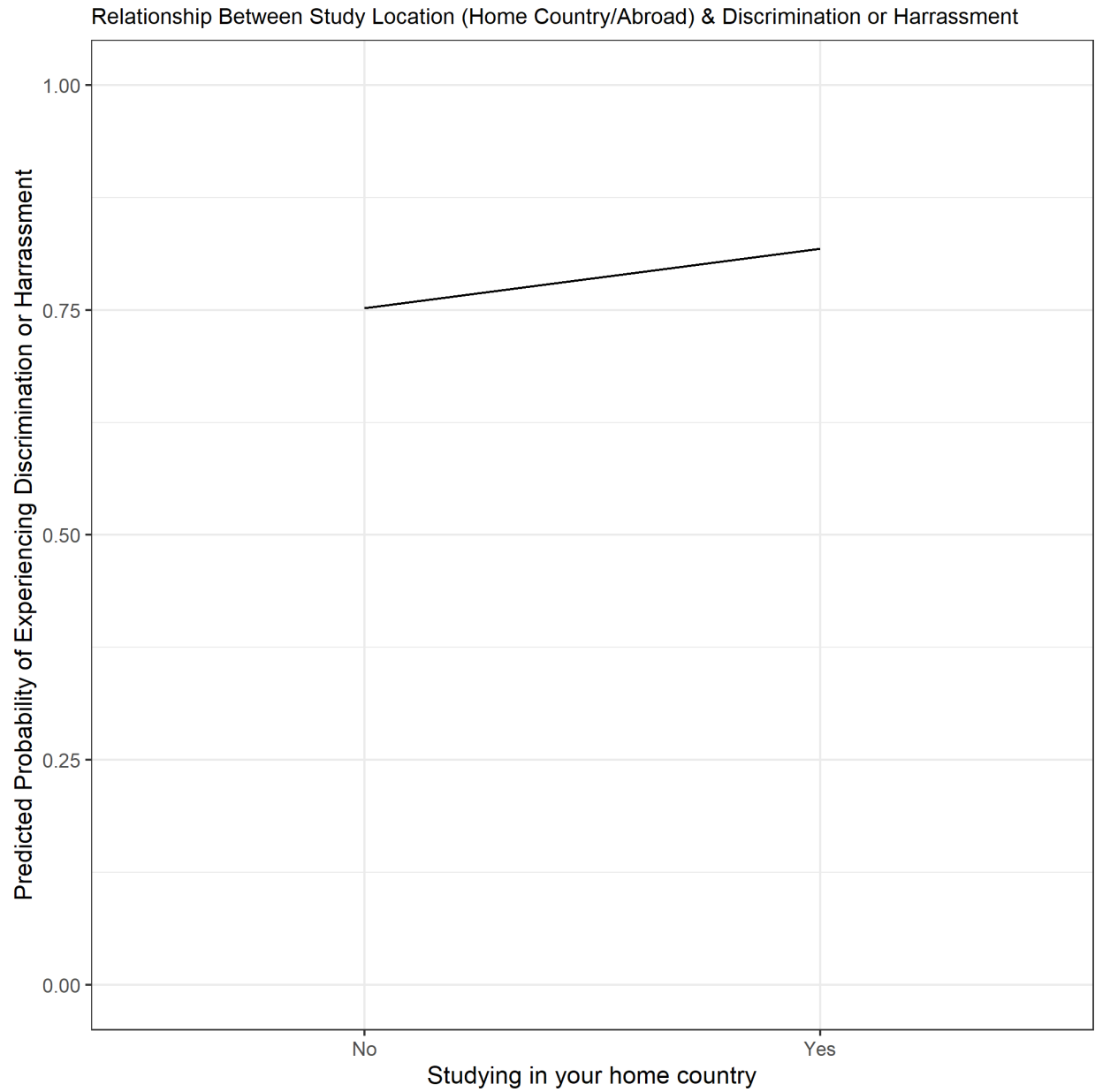
Relationship Between Work/Life Balance & Satisfaction with Pursuing a PhD

LR4: Supervisor Relationship ~ Anxiety or Depression We had a great interest in mental health markers in graduate school, with anxiety or depression being the clearest to measure out of the survey variables. There is a positive relationship between having a good relationship with your supervisor and a low probability of seeking help for anxiety or depression, which matches our hypothesis.
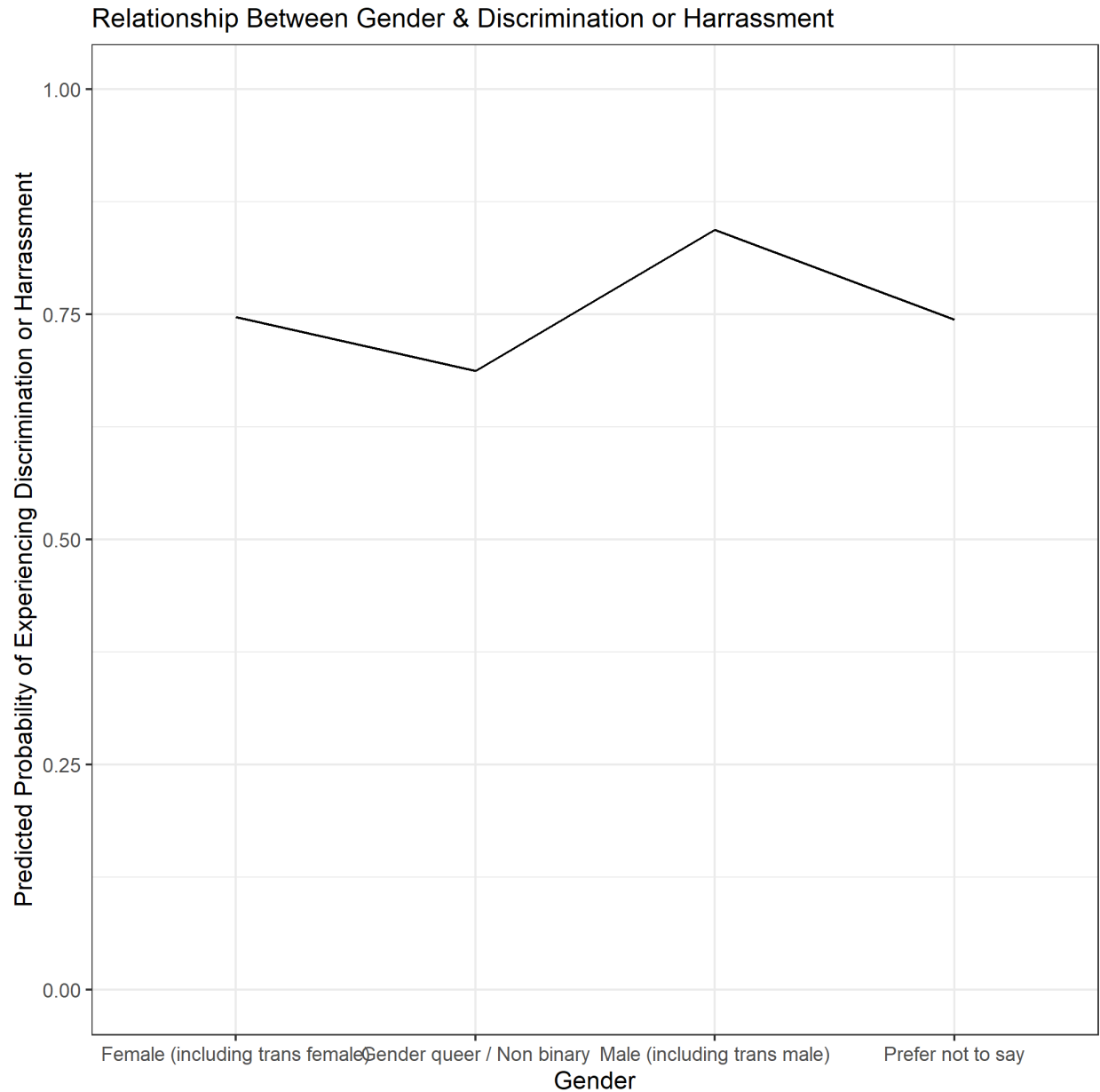
Relationship Between Supervisor Relationship & Anxiety or Depression

LR5: Studying Outside of your Home Country ~ Anxiety or Depression Studying in your home country and seeking help for anxiety or depression do not appear to be correlated when running a logistic regression.

Relationship Between Studying Outside of your Home Country & Anxiety or Depression

LR6: Studying Outside of your Home Country ~ Discrimination or Harrassment Although seemingly weak, there does appear to be a correlation between study location and having experienced discrimination or harrassment.

Relationship Between Study Location (Home Country/Abroad) & Discrimination or Harrassment

LR7: Gender ~ Discrimination or Harrassment   There are different correlations between each gender and discrimination or harrassment.

Relationship Between Gender & Discrimination or Harrassment

Discussion and Conclusion

While we have certainly found some relationships between our variables of interest, these have proven weaker than originally expected. This makes sense given the complexity of the variables - the level of satisfaction that a student may feel with their decision to pursue a PhD is probably due to the added effect of many variables rather than to a single factor, meaning that a relationship between two variables alone is an incomplete picture of everything that may be taken into account. Nonetheless, it was still interesting to investigate these relationships and work with this dataset. Unfortunately, it did prove quite challenging to use a dataset with many categorical variables that required a lot of fixing, converting answers to scales, and ultimately running logistic regression rather than linear, as the latter is unbounded and cannot be used for categorical variables.

References

Nature.com. 2020. Phds: The Tortuous Truth. [online] Available at: https://www.nature.com/articles/d41586-019-03459-7 [Accessed 17 March 2020].