DATA 550 Final Project

Tracy Huang

2024-03-26

Introduction

The Global School-Based Student Health Survey (GSHS) (https://www.kaggle.com/datasets/leomartinelli/bullying-in-schools) is a dataset containing school-based survey data. The survey was conducted in Argentina in 2018 with 56,981 student participants, using a self-administered questionnaire to obtain data on adolescent's demographics, family, and social factors related to bullying. In this R Markdown report, the objective of the analysis is to conduct exploratory data analysis on the variable relating to the severity of experiencing physical attacks in bullying and factors related to social relationships with peers, friends, and family. The results will be displayed in a table. Additionally, in order to visually represent the association between physical attack severity and loneliness, a mosaic plot will be created.

Dataset Description

The physical attack variable (`physically_attacked`) was originally recorded in different categories based on frequency (0 times, 1 times, 2 to 3 times, etc.). However, I recoded the variable to denote different severity levels (low, medium, or high) based on the frequency.

- `physical_attack_severity` - Recoded to have values "Low" if the adolescent was physically attacked 0 or 1 times, "Medium" if the adolescent was physically attacked 2 to 7 times, or "High" if the adolescent was physically attacked 8 or more times.

Additionally, categorical variables with more than two levels were recoded to have only two levels for greater simplicity in future analyses and figures.

- `were_obese` - Has values "Yes" or "No"

- `most_of_the_time_or_always_felt_lonely` - Has values "Yes" or "No"

- `sex` - Has values "Male" or "Female"

- `other_students_kind_and_helpful` - Recoded to have values "Yes" if adolescents indicate that other students are kind and helpful always or most of the time, or "No" if adolescents indicate other students are kind and helpful sometimes, rarely, or never

- `parents_understand_problems` - Recoded to have values "Yes" if adolescents indicate that parents understand their problems always or most of the time, or "No" if adolescents indicate that parents understand their problems sometimes, rarely, or never

- `close_friends` - Recoded to have values "Low" if an adolescent has 0, 1, or 2 close friends, and "High" if an adolescent has 3 or more friends
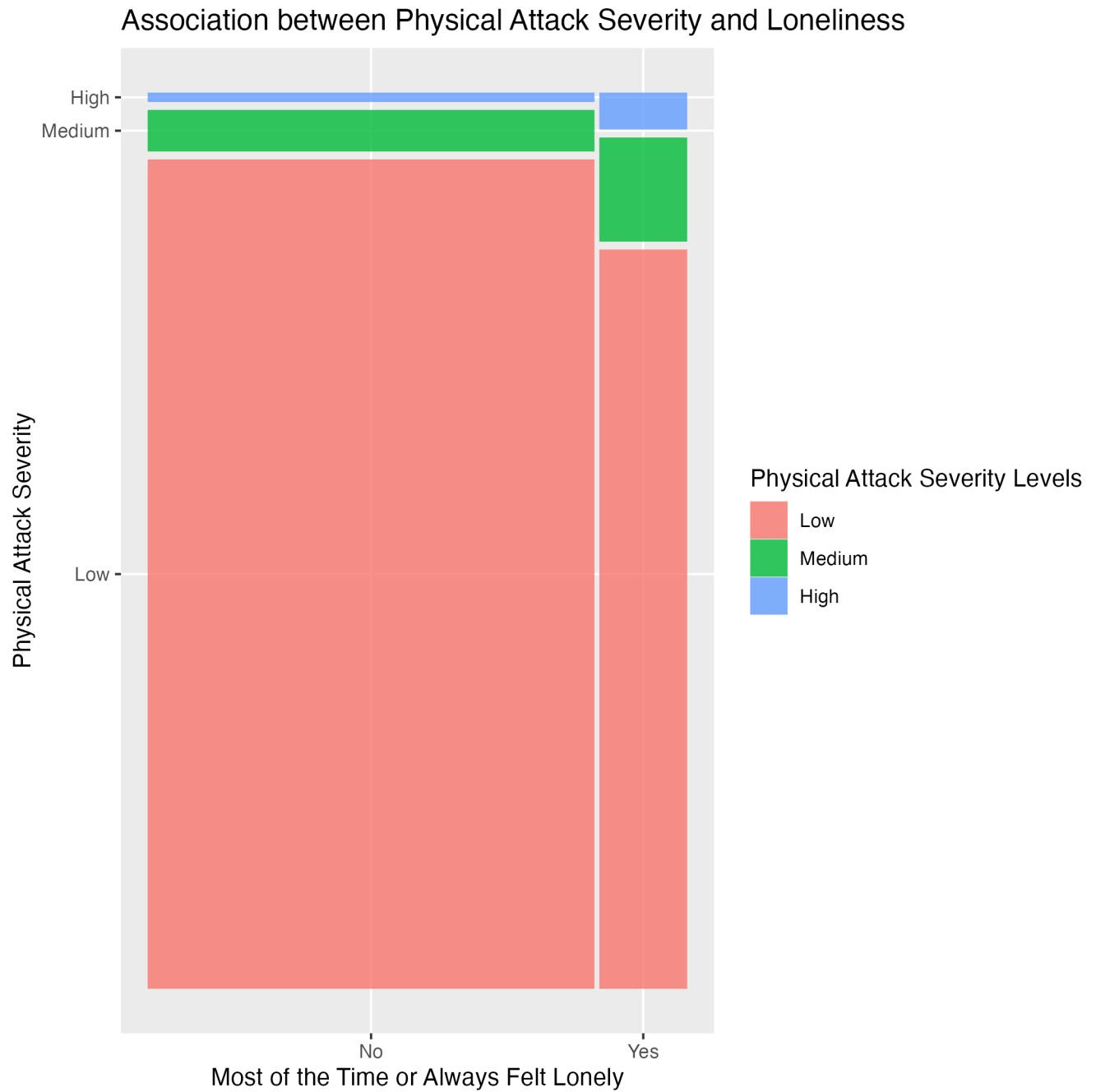
Descriptive Statistics Table

Table 1: **Distribution of Variables Across Physical Attack Severity Levels**

Characteristic	**Overall**, N = 32,938	**Low**, N = 30,521	**Medium**, N = 1,919	**High**, N = 498
Parents Understand Problems				
No	18,879 (57%)	17,145 (56%)	1,366 (71%)	368 (74%)
Yes	14,059 (43%)	13,376 (44%)	553 (29%)	130 (26%)
Close Friends				
High	22,938 (70%)	21,424 (70%)	1,204 (63%)	310 (62%)
Low	10,000 (30%)	9,097 (30%)	715 (37%)	188 (38%)
Were Obese				
No	30,552 (93%)	28,348 (93%)	1,767 (92%)	437 (88%)
Yes	2,386 (7.2%)	2,173 (7.1%)	152 (7.9%)	61 (12%)
Most of the Time or Always Felt Lonely				
No	27,548 (84%)	25,989 (85%)	1,284 (67%)	275 (55%)
Yes	5,390 (16%)	4,532 (15%)	635 (33%)	223 (45%)
Sex				
Female	17,679 (54%)	16,539 (54%)	911 (47%)	229 (46%)
Male	15,259 (46%)	13,982 (46%)	1,008 (53%)	269 (54%)
Other Students Kind and Helpful				
No	16,951 (51%)	15,440 (51%)	1,188 (62%)	323 (65%)
Yes	15,987 (49%)	15,081 (49%)	731 (38%)	175 (35%)

The table above displays descriptive summary statistics for each of the categorical variables in the dataset, in terms of the counts and percentages of each category across the different physical attack severity levels. Overall, the table shows that most students had low physical attack severity levels in bullying (defined as being physically attacked zero or one time). Additionally, for close friends, obesity, and loneliness, students overall were in the positive categories of having a high number of close friends, not being obese, and not feeling lonely most of the time or always. However, for parents understanding problems, sex, and finding other students kind and helpful, students were more evenly split between each variable's two categories.

When examining each variable's distribution across physical attack severity levels, most of them suggest that there may be an association between the variable and physical attack severity. For instance, for parents understanding problems, within low physical attack severity levels the percent of students who indicate that their parents understand their problems is comparable to the percent of students who indicate that their parents don't understand their problems. However, within the medium and high physical attack severity levels, the percent of students who indicate that their parents don't understand their problems is much higher than the percent of students who indicate that their parents understand their problems.

Mosaic Plot

Association between Physical Attack Severity and Loneliness

In the mosaic plot displayed above, it's clear that there are differences in physical attack severity levels depending on how lonely adolescents felt. Within adolescents who felt lonely most of the time or always, there was a higher proportion ot adolescents who experienced medium or high levels of physical attacks, compared to adolescents who did not feel lonely most of the time or always. This is because the green and blue sections indicating medium or high levels of physical attacks take up a larger proportion of space in the column representing adolescents who felt lonely most of the time or always.