Project 1 Analysis of Results

Group Members: Hannah Duncan, Geoffrey Johnston, Vineet Sikri, Melody Yu

**Background**

Bullying is a major social problem affecting school aged children and adolescents nationwide. Researchers on school bullying have defined it as “a systematic abuse of power”. The defining features of bullying which distinguish it from other forms of aggressive behavior are that it is (a) repeated and (b) intent to hurt, and (c) there is an imbalance of power between the bully and the victim, with the victim in a weaker and more vulnerable position. There are four main types of bullying which are most prevalent in school, including physical bullying, verbal bullying, social or relational bullying, and cyberbullying. Bullying can negatively impact youths’ overall well-being and has been linked to multiple adverse outcomes, including both externalizing and internalizing difficulties. One of the major consequences associated with being the target of bullying is lower academic performance. For this project, we as data analyst in the field of education, will be looking at national bullying statistics and determining if there is a relationship between bullying and academic performance.

**Data Sources**

To compile our data that we analyzed, we used two API calls from these sources:

<https://educationdata.urban.org/documentation/>

<https://www.census.gov/acs/www/data/data-tables-and-tools/data-profiles/2015/>

We also used data from the following sources:

<https://nces.ed.gov/programs/crime/student_data.asp>

<https://usedgov.github.io/>

Alhajji, M., Bass, S., & Dai, T. (2019). Cyberbullying, Mental Health, and Violence in Adolescents and Associations With Sex and Race: Data From the 2015 Youth Risk Behavior Survey. *Global Pediatric Health*, 6.

Our data compilations included CSVs and API calls as well as a published paper.

**Data Exploration & Cleaning**

To answer our research questions, we combined data from Urban Education Data API, NCES Crime & Student Survey, and Department of Education state assessment surveys. With such large data sets––many of which taking the form of survey responses––we had a lot of cleaning to do before our analysis. For example, when combining the six state assessment CSV files, the numeric values for assessment passing rates had such a large range of responses and data types that we had to formulate and bin valid score ranges for each response.

**Results & Analysis**

**Bullying by Location**

Because the bullying allegation data was organized by state, our group created a heat map using each state’s coordinates and Python’s gmaps module. Interestingly, this heat map showed that even though states like California and Texas have large populations––and thus conceivably may have higher allegation counts––much of the U.S’s bullying allegations in 2015 came from the Northeast.

A close up of a map

Description automatically generated

**Prevalence of Bullying in Schools**

In 2015, bullying is most prevalent in middle schools, and followed by high schools. Among the three types of bullying, verbal bullying is the most prevalent for all school levels comparing with physical bullying and social bullying.

A screenshot of a cell phone

Description automatically generated

Bullying reported in 2015 showed a bell-curved distribution for different school sizes – where medium-sized schools have more bullying reported comparing with small-sized and large-sized schools.

A screenshot of a cell phone

Description automatically generated

**Cyberbullying**

Cyberbullying is a serious issue, with growing access to technology it’s only going to increase. We found that kids over the age of 15 were more likely to be cyberbullied, gender wise female were cyber bullied more than males and overall white adolescents are at increased risk of being cyberbullied. This can relate to increased use of technology and social media.

A screenshot of a cell phone

Description automatically generated

Cyberbullying was seen as additional factor affecting mental well-being of students. Among victims, approximately 60% reported having depressive symptoms Nearly 40% of the victims reported having thoughts about suicide. Being cyberbullied was also found to be associated with violent behaviors. Of students who reported victimization, 20% reported carrying a weapon and some engaged in physical fight.

A screenshot of a cell phone

Description automatically generated

**Bullying and Academic Performance**

To better understand the relationship between bullying and academic performance, we wanted to first group assessment scores and graduation rates by state to match the bullying allegation data. In each case, the indicators with rates “Less Than 50% (LT50%)” displayed positive correlations when juxtaposed against the bullying allegation data. This correlation was especially strong (r2 = .67) with the assessment score data.

A close up of a map

Description automatically generated

A close up of a map

Description automatically generated

Another goal of ours was to see if we could find a correlation between the high school dropout rate and instances of alleged bullying. Unfortunately, we were only able to compile reliable data for two years, 2013 and 2015. Surprisingly, as bullying counts increased over this time period, instances of high school dropouts decreased.

A screenshot of a cell phone

Description automatically generated

**Demographic Factors**

Our final research question asked us to explore which demographic factors (socioeconomic status, age, ethnicity, etc) may lead to higher instances of bullying in our sample. Using the 2015 Census results, we were able to study the effects of poverty rate, per capita income, and various racial/ethnic identities on bullying allegations. For all of the aforementioned factors, the correlations were weak and even somewhat contradictory. An example of this can be seen in the following figures, where bullying allegations increased slightly with increasing poverty rate but also increased with per capita income.

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated

**Limitations**

We recognize that there were limitations on the data we compiled. Below is a listing of some of the limitations:

* Population and ethnicity data are only based on reported census data.
* Bullying data contain only instances of alleged bullying and does not include cyberbullying.
* Graduation rates, dropout counts, and test scores are not the only measures of student success.
* Assessments given to students are standardized by state; they are not nationally standardized.