# INTRODUCTION

The great city of Chicago, a place many cherish for its thriving metropolitan sector, competitive sports, and growing diversity has an unfortunate reputation and history of abnormally high crime levels. Many would assume that this type of violence would impact a variety of societal factors: i.e. education, economic mobility, citizen safety, etc. In this project, I wanted to uncover the seemingly reciprocal relationship between crime in Chicago on Chicago Public School elementary school student performance. To gain more insight on this topic I looked into previous research. For example, Dr. Julia Burdick-Will from the National Institute of Health (NIH), in her study on the effects of violent crime on student academic performance in the City of Chicago, states, "The evidence described in this study shows that violent crime is a serious issue in some of Chicago's public high schools. Many students are exposed toviolent crime on a routine basis just by entering school grounds." She found that exposure to violent crime on school grounds negatively affects test scores, but not grades. This suggests a negative relationship between the amount of crime in a student's environment and their academic performance such that it causes a direct reduction in students' ability to learn. As a mentor of Chicago Public School high school students in the Southside community, through the University of Chicago's student organization Moneythink, I wanted to look deeper into how crime can influence student performance and opportunities for success.

# THE DATASETS

For this project, I used a total of three datasets to visualize this information. First, I selected crime data from the years 2001 - 2023, produced by the Chicago Data Portal. This dataset provides fifty variables and over seven million rows that encompass crime reports around the City of Chicago from the past two decades. Included in the descriptors for each crime report, are the Wards, Primary Type (Description of Crime), and the location data for all reports (physical coordinates and information of crime location), etc. Next, I imported the GeoJSON Chicago Boundary data, necessary to create choropleth map visualizations, from the Chicago Data Portal. This dataset proved helpful in producing effective visualizations that display the impact of a single variable across the entirety of Chicago. Finally, I imported a data by Chicago Public Schools (CPS) and made public by the Chicago Data Portal as a social indicator of the people in the area of the crimes reported for a more meaningful analysis. This dataset measures variables of Chicago Elementary Public School students in the years 2013-2014, as this was the most recent and informative dataset on academic performance in CPS I could obtain. It was also helpful that the Chicago crime data spanned the years 2001-present, so I could also compare these datasets without misinterpreting the values by year.

## Cleaning the Data

The data wrangling process for these datasets was relatively simple. First, I dropped all rows with values equal to 'NaN' (approximately 731,000 rows ) from the crime data frame to ensure the accuracy and integrity of my analysis. Although a high number of rows were removed, they are only a small fraction (9.37%) of the nearly-eight million rows in the original dataset. Furthermore, for the remaining two datasets, there were no 'NaN' values in the dataset and that I could begin analyzing and comparing the data right away. Modifications to the remaining two datasets were simple, I converted the column types from'object' to 'int' to merge the separate data frames on the same column.
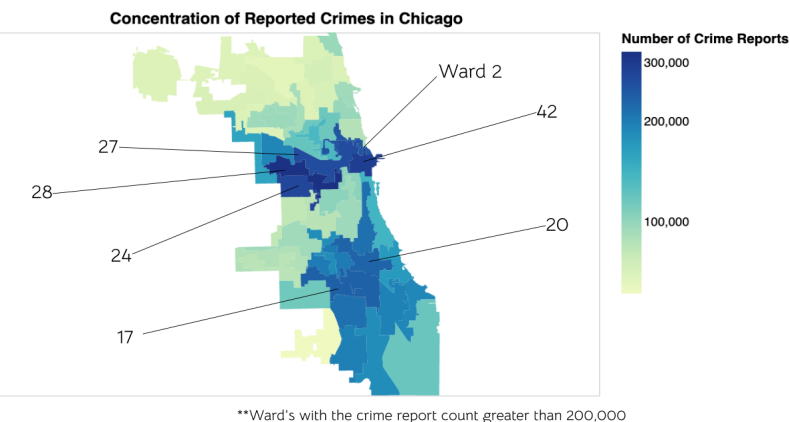
## THE ANALYSIS



**Concentration of Reported Crimes in Chicago**

Ward 2
42
27
28
24
20
17

Number of Crime Reports
300,000
200,000
100,000

**Ward's with the crime report count greater than 200,000

Figure 1 **[1]**: Concentration of Reported Crimes in the City of Chicago

### *Figure 1*

During the beginning stages of analysis for this project, I remained committed to visualizing only the information in the Chicago crime dataset. This consisted of encoding multiple choropleth maps to display which types of crime prevailed as most common in each ward. After this, I realized the most effective way to visualize crime in Chicago was to show the total count of crime reports for each ward **[1]**. To create this chart, first, I grouped the data by ward and primary type to uncover a crime report count for every ward in this dataset. I plotted the data using the Altair mark_geoshape and encoded the color of this map as the total number of crime reports in each respective ward. I then used a logarithmic scale for the color encoding so the viewer could easily recognize which wards face the highest crime. Lastly, I imported **[1]** into the Figma file editor to mark up the visual to display and label all wards with a crime report count greater than 200,000. I included a text box with the ward number and lines that are attached to each respective location so the viewer could gain a further understanding of which Chicago wards are most impacted by crime. I then included a message in a text box at the bottom of the figure such that the viewer could understand what the

numbers represent. Upon completion, I was able to draw some interesting information. For example, Wards with the highest crime counts are all seen in close proximity to one another. There are two main concentration of crime reports across Chicago in this dataset. First, many wards surrounding the North Loop
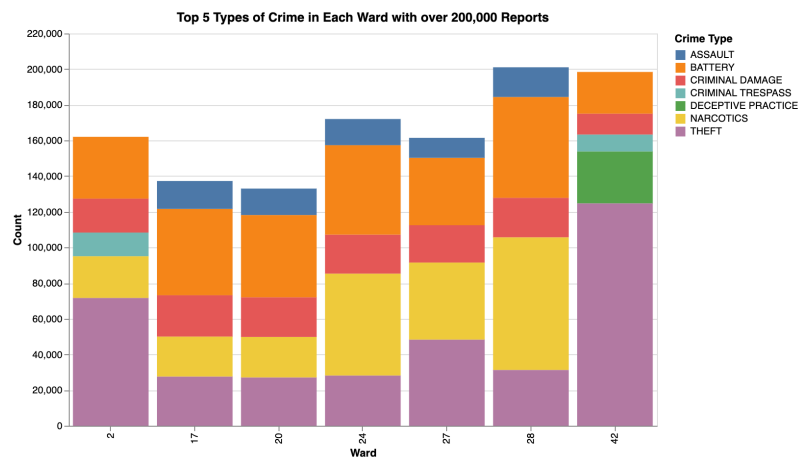


Figure 2 **[2]**: Most Common Crime Types in Wards with > 200,000 Reports

### *Figure 2*

Upon plotting **[1]**, I was able to extract some valuable information regarding the most common types of crime in all Chicago Wards with a crime report count of greater than 200,000 reports, as they prevailed in a very dark blue color. I used these seven wards as a basis to uncover more information within the Chicago crime data frame. To find the top five most common crimes within wards which have the most crime reports, I grouped the data by Chicago Ward such that it filtered the seven wards with more than 200,000 reports. I then used another groupby function by the maximum count of the Primary Type variable or each ward. This allowed me to then plot the data frame in Altair using the mark_bar encoding **[2]**. I encoded the x-axis by each Ward as ordinal values and the y-axis with the quantitative count of crime reports for each respective Ward. To visualize the top five most common crimes, Iencoded the chart color

with the nominal Primary Type variable such that it displays, for each ward, how many reports there are for each type of crime in these seven areas with the most crime. I then encoded the chart color by including the categorical color scheme 'Tableau10" because it allowed a distinct color for each crime type, which is easily navigable by the legend. From this chart, we find that Wards 27 and 42 have the highest crime in Chicago, and Ward 20 has the lowest of the group. The most common crime for all wards is theft
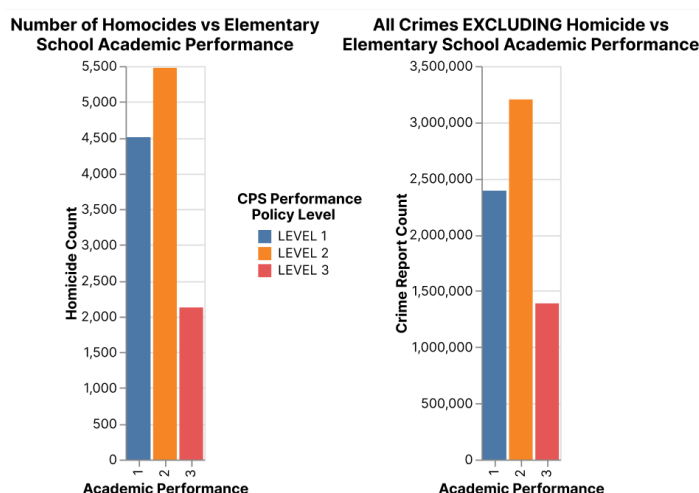


Figure 3 **[3]**: Comparing Crime Rates to Academic Performance: Homicide vs. All Other Crimes

### Figure 3

When visualizing Figure 3 **[3],** I had initially plotted the chart on the right to understand the relationship between the academic performance of Chicago Public School elementary students and the number of all crimes reported in Chicago. To visualize this, I chose the variable, 'CPS Performance Policy Level' which represents the students' performance categorically from levels one to three. The ordinal value of three represents the total CPS elementary schools with higher cumulative academic performance (Grades, Test Scores, etc.) whereas the value of one represents schools with an overall lower standing in these metrics. I then transformed the metric column's values into quantitative

dummy variables so I could find any correlation between crime and academic performance. Lastly, I plotted the data in Altair with a mark_bar encoding to create a bar chart. I encoded the x-axis as the academic performance and the y-axis as the total crime count to find any trend between crime and the performance of the groups of students within each category. I then included the CPS categorical data as the encoding for the color of this bar chart to make clear what information is being presented. Further, I wanted to discover the effect of distinctly violent crime on academic performance (as described in the study cited earlier), so I chose to analyze the homicide crime type and its effects on the same CPS overall performance metric. When plotting them side-by-side, I found surprising results. The relationship between the effects of homicide versus all other crimes on student academic performance does not differ at all. This suggests that violent crime in Chicago has zero correlational relation with students' academic performance when compared to the influence of all other crimes.This allows us to infer that violence in Chicago has no impact on CPS students' academic performance.

### Figure 4

After finding insufficient evidence for the negative relationship between distinctly violent crime and academic performance, I wanted to take a different approach and test for a relationship between the crime count and the student suspension rate. I chose this new variable because it offers information, separate from academics, that can be suggestive of the student's mental or physical state, or in this case, the student's ability or eagerness to learn. Since the education data comes from the years 2013-2014, I selected the suspension data from 2013 and filtered the crime report data frame (using the pandas

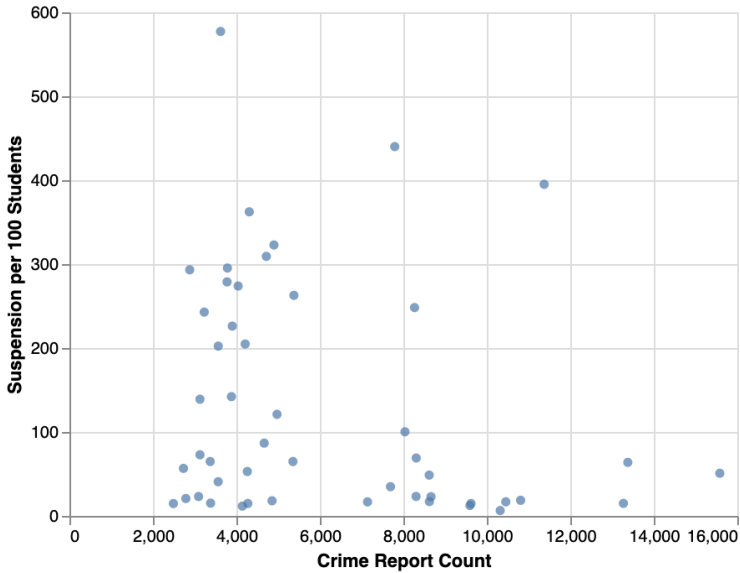**Number of Crime Reports vs Student Suspension Rate in Chicago 2013**



Figure **[4]**: Crime Report Count v. Suspension Rate in Chicago 2013

pd.date_time function) to include only data from the year 2013. I then merged the two separate data frames together on the 'Ward' column and plotted the data in Altair with mark_circle to create a scatter plot. I decided a scatter plot would be optimal to compare two categorical variables for any insight into existing trends. I encoded the x-axis as the crime report count and the y-axis as the suspension response variable to find a relationship. Most points in the distribution were centered around the crime count of 4,000 with a weak distribution and outliers surrounding this value. The fact there was immediately no obvious correlation **[4]** between the variables suggests, provided a lack of evidence for any existing trend in **[3]**, that there is no correlation between crime and student performance at all.

## DISCUSSION AND CONCLUSION

Upon rigorous analysis and interpretation of these datasets, there is enough evidence here to re-evaluate Dr. Burdick-Will's suggestion that the violent crime issue in the City of Chicago negatively impacts students' academic ability and performance. Figure **[3]** rejects any relationship between violent crime and academic performance and Figure **[4]** furthers

the argument by suggesting there is no correlation between students' performance in school and the crime in the community at all. This finding directly rejects the commonly held belief that exposure to crime can have a negative impact on the student's performance or overall success. However, these findings do not denounce the real impact of crime in the City of Chicago. They only raise a larger question at hand of the other factors that play a more substantial role in a student's well-being and overall performance in school. It also alludes to the fact that this issue is less of a problem caused by social factors today but by systemic issues implanted into the history of Chicago, which contribute greatly to the continued persistence of crime today.

**Limitations to the Dataset**

Although I was able to draw significant findings from these datasets, the data was not perfect. First, the crime data only included the crime committed for each report and no more information on the offender which could have been very useful. The crime data spans from 2001 to the present, but there might be fluctuations in crime patterns over time that are not adequately represented. Further, the education dataset is from 2013 - 2014, the most informative and up-to-date data I could find, provides little insight into the academic performance of CPS students today. Finally, another limitation of the education data was the fact that it only recorded CPS elementary schools' performance, as this was the only education information available. While this is insightful, it is not representative of the entire impact of violent crime on student academic performance in Chicago, as it excludes the performance of CPS middle schools and high schools.

## WORKS CITED

- Chicago Violence/Education Research: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3831577/
- Figma Workbook :https://www.figma.com/file/ICK5iEVoLwYu5XXrLFLism/Untitled?type=design&node-id=0-1&t=Bg2sUkYx5S0eYWUz-0
- City of Chicago Crime Data 2001 - Present — Chicago Data Portal: https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-Present/ijzp-q8t2
- City of Chicago GeoJSON Data – Chicago Data Portal: https://data.cityofchicago.org/Facilities-Geographic-Boundaries/Boundaries-Wards-2015-2023-/sp34-6z76/
- Chicago Public School Elementary School Performance Data – Chicago Data Portal (2013-2014): https://data.cityofchicago.org/Education/Chicago-Public-Schools-Elementary-School-Progress-/tj8h-mnuv