Crime and the Church: An Analysis of Crime in Chicago Based on the Location of Churches

VALPARAISO UNIVERSITY

Adali Johnson, Data Science and Economics

Abstract

Based upon publicly available data, there are approximately 43 crimes committed per 1000 residents in Chicago, IL with a percentage that is 149% higher than the average city in Illinois. For years researchers, politicians, and individuals have struggled with what can be done to deter crime. One method that is often tried is the introduction of community programs and outreach. This research aims to analyze the relationship between churches and crimes committed to determine if increasing church presence is a viable method for crime reduction. Using data from the Chicago CLEAR dataset of crimes committed in 2019 coupled with church location data from 2018, visualizations were created to illustrate the geographic relationship between crimes and churches. The churches used come from both the Department of Homeland Security's Homeland Infrastructure Foundation-Level Data and the Chicago Church Finder. Data cleaning was done with Python, exploration and visualization were done with Tableau, Python's KDTree package was used to create a KDTree that calculated the number of crimes within pre-determined radii of churches, and something was used for analysis.

Data Collection and Preprocessing

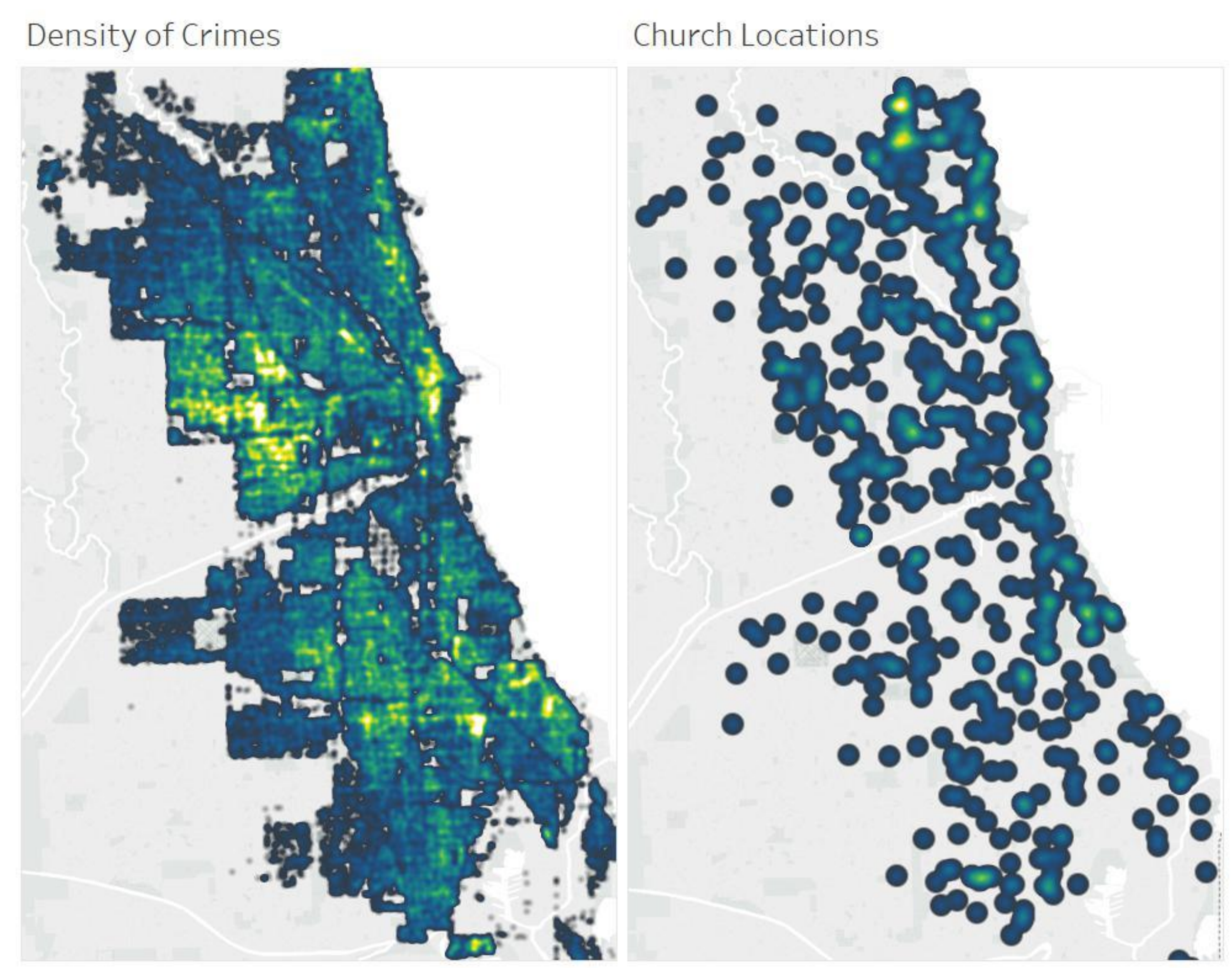
For this analysis, data was collected from two sources. The first was the Chicago CLEAR dataset. This dataset is publicly available on the city of Chicago's data portal. In its entirety, the set contains every reported incident of crime that has occurred in Chicago from 2001 to present. In the original dataset there were 22 variables present. These variables included identification, dates, crime type, and a wide range of locational data. However, given the enormity of the number of crimes that have occurred in that time span, Python was used to limit the number of columns to only those that were necessary for the software packages being used and limit the number of rows to only crimes that were committed in 2019. The resulting dataset contained the Latitude and Longitude points, with all null values removed, of 256,147 crimes committed in 2019. The second data set was collected from the Department of Homeland Security's open geospatial data. When narrowed down to only Illinois, this dataset contained 1,942 churches that have their address listed publicly. In order to make the data applicable to this analysis, Python was used to narrow down the set to 570 non-null latitude and longitude points of churches in the city of Chicago. While these 570 churches do not make up the total number of churches in Chicago, it was the best publicly available set to use for this analysis.

Methods

For data preprocessing, Python was used to remove rows of data that had null values for either latitude, longitude, or both. Following this removal, it was ensured that the latitude and longitude columns matched in both the crime data and church data. Ensuring this allowed for the church data and crime data to be vertically stacked with the church coordinates listed in rows 0-570 and the crime data in the remaining. For data exploration, the data was visualized in Tableau. This gave insight into the relationship between crime locations and church locations. It was determined that there weren't any clear hubs of churches. Although there were some areas with a higher density of crime, the equal spread of churches throughout the city should allow for the KD Tree to work appropriately. In order to ensure that the results from the KD Tree for churches and crime had a baseline for comparison, a set of random points was created to compute a KD Tree with random points versus crimes in Chicago. These random points were generated by determining furthest north and south longitudes and furthest east and west latitudes of crime points in the Tableau visualizations. From there, 570 random latitude, longitude points were generated as random number between the determined maximum and minimum coordinates.

Geographical Exploration

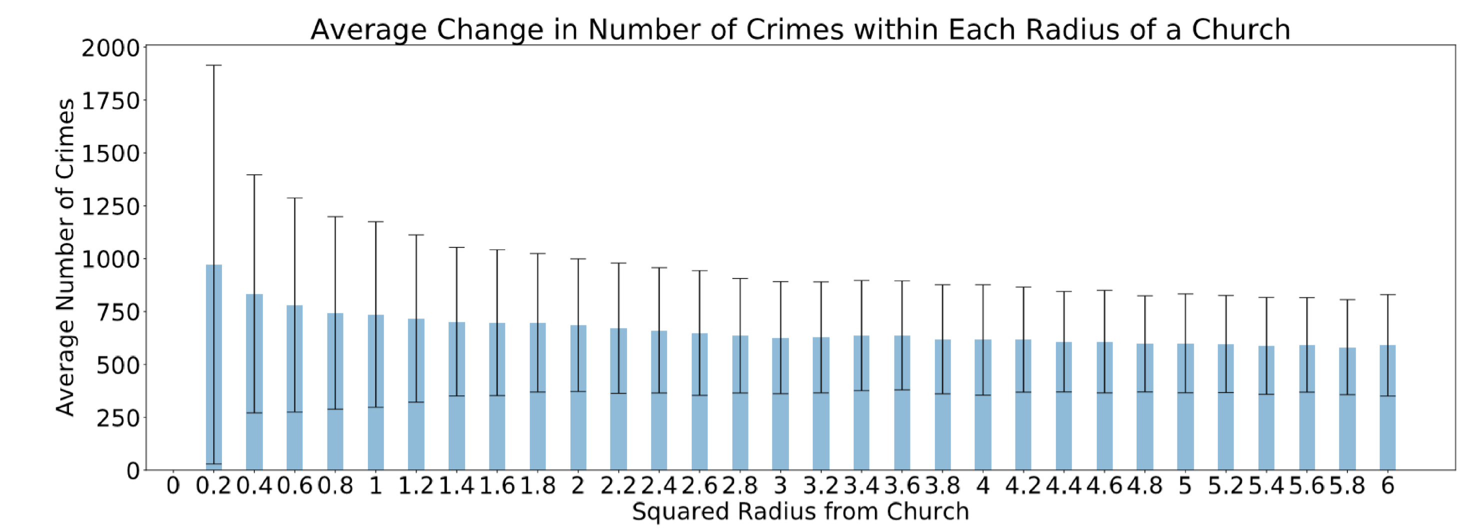
For initial exploration, both the crime data and church data was plotted in Tableau using density plots. During the preparation, it was ensured that all of the data came from the same year, 2019. Based on the visualizations below, one can notice that aside from some areas along Lake Michigan, in this set of data, there aren't many clear hubs of churches- seen in the graph on the right. This would lead one to believe that the data included is spread out evenly throughout the city.



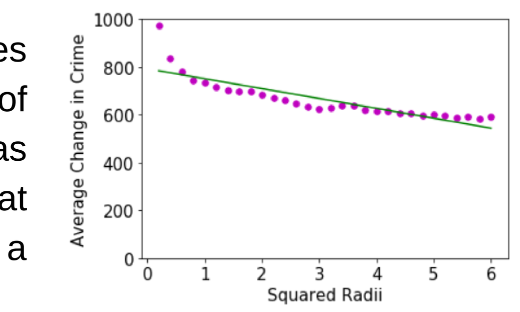
KDTree Specifications and Results

A K Dimensional Tree is a binary search tree that computes distances between n points within a k-dimensional space. Generally, k-d trees are used to compute "neighbors" of points based on certain attributes. In this specific analysis, a k-d tree was used to compute the number of crimes located within predetermined radii of church locations. The query function of Python's KDTree package allows for the return of certain specifics about nearest points. In this design, the distance of crimes from churches was determined using the package's built in Euclidean distance formula. However, given the nature of latitude and longitude points, the Haversian formula was used to convert the Euclidean distances to miles. Prior to this conversion, the square root of each radius was taken to ensure that the area expanded appropriately as the area of a circle. The radii used were r^2 in increments of 0.2 between 0 and 6, these were technically calculated as square root of 0, 0.2, 0.4, etc. Upon the completion of the tree, 286,000+ crimes were compared to each of the 570 churches. The number of crimes within each specified radii of each church was then calculated and stored in a 570 x n matrix, with n being the number of specified radii. In order to get the most meaningful results, the matrix was then recalculated to produce the change in number of crimes from radii to radii (i.e. the number of additional crimes when the radius expands from 0.2 r^2 to the 0.4 r^2). The results found that there were close to 1000 crimes within a 0.2-mile radius. There was then a decreasing trend in change in number of crimes (i.e. 1000 decreased to 900, etc.) until a radius of 1 mile. After that, the change in crimes leveled out at an increase of approximately 600 crimes as the squared radius increases by increments of 0.2 miles squared. For the random points versus crimes analysis, it was found that there was a steady increase of approximately 500 additional crimes as the squared radius increases by increments of 0.2 miles squared. In order to better portray these results, bar charts with error bars were created to display the change in average number of crimes as the squared radius increases by increments of 0.2 miles squared. Linear regression lines were also created to determine the average change in number of crimes.

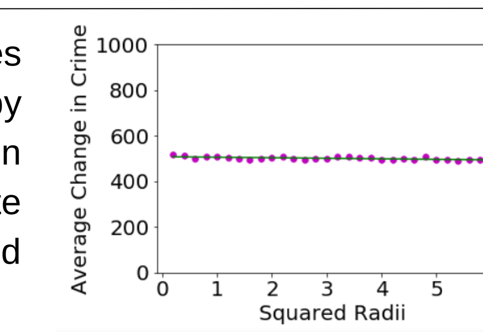
Analysis

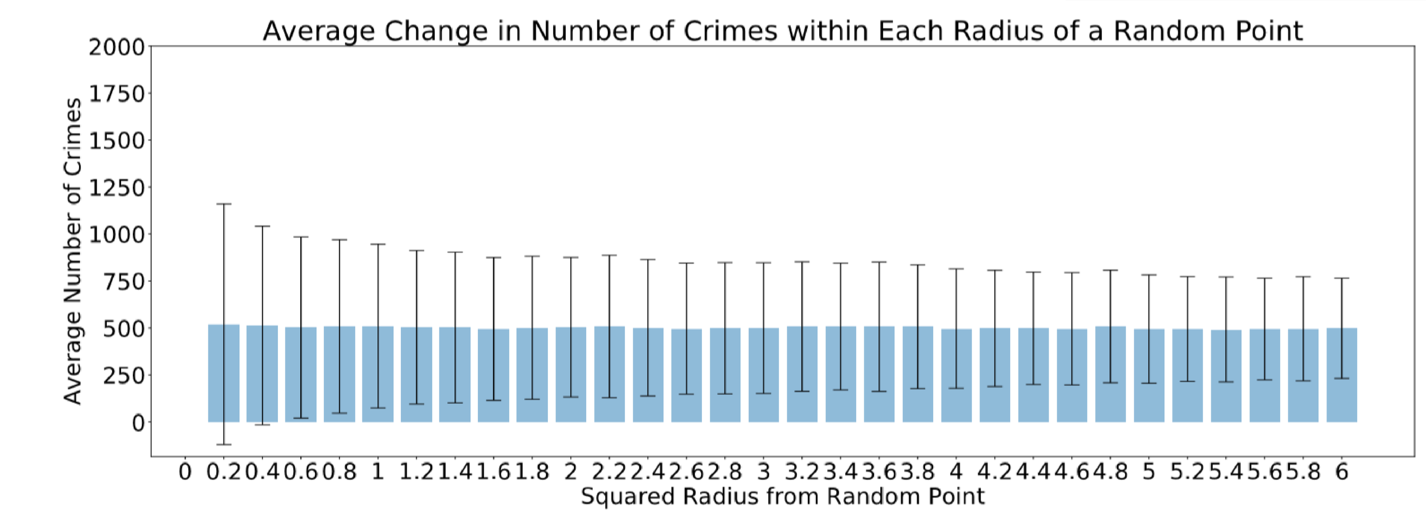


The plot above shows the change in average number of crimes from churches as the squared radius increases by increments of 0.2 miles squared. The linear regression equation was determined to be 792.498 – 41.419 r^2.This would indicate that for every increase in the radius by one mile, one would expect a 41.419 decrease in the average number of crimes.



The plot below shows the change in average number of crimes from random points as the squared radius increases by increments of 0.2 miles squared. The linear regression equation was determined to be 509.130 – 2.329 r^2. This would indicate that for every increase in the radius by one mile, one would expect a 2.329 decrease in the average number of crimes.





Summary of Results

Given the regression equations above, it seems that there are more crimes within the initial 0.2 r^2 radius of churches than there are in the initial 0.2 r^2 mile radius of random points. The bar charts also indicate that there are less additional crimes in the first 1 r^2 mile radius for random points than for churches. Additionally, at every additional 0.2 r^2 mile radius, there are 600+ more crimes per each from churches compared to the 500 more crimes per each random point.

Conclusions / Future Work

In this analysis, the primary objective was to determine whether churches could serve as natural deterrents of crime within the city of Chicago. Despite the hypothesis that churches may serve as natural deterrents of crime, the results would beg to differ. Given the bar charts and linear regression equations above, there is evidence that there may be more crime around churches than around average points in Chicago. However, this could simply be attributed to the fact that the population density may be higher around these churches. The Chicago CLEAR data set contains far more information than just the coordinates of crimes. Specifically, it includes information on type of crime, time, whether arrests were made, and more. For further work, this analysis can be expanded to see if there is are certain types of crimes that occur frequently around churches. A second area for expansion, would be to look at the data on a ward level to compare the impact of churches based on each ward.