CS5830 Project 2

Wesley Dashner
Zachary Ellis

January 2021

Introduction

The National Incident-Based Reporting System (NIBRS) contains data for any recorded incident, crime, in the United States. This data was used to look at an arbitrary city, in this case Austin, Texas. The purpose of this analysis was to look at correlation between Median Household Income for a given zip code in Austin, Texas and the amount of Aggravated Assaults committed. The main analysis techniques used were Pearson Correlations, standardization of data, and a Mann-Whitney-U-Test.

Dataset

As previously mentioned the dataset is very robust; however, we honed our approach to simply the aggravated assaults, the zip codes in which they were committed and the median household income of said zip codes. The data was given in a format that had each crime listed individual with statistics about the crime itself and the zip code in which it was committed. It is also important to note that the data was given in such a way that cleaning was simple. We only had to remove the aggravated assaults that were committed in zip codes that there was no information for the median household income.

Analysis Technique

Map: For the map graph, we removed all crimes from the data set that were not an aggravated assault. We then graphed a point for each remaining crime based on their x and y coordinate attributes. This shows areas where there are more aggravated assaults and where there are fewer aggravated assaults throughout Austin.

Aggravated Assaults compared to Income: We again removed all non-aggravated assault crimes. We wanted to compare aggravated assaults to income. We grouped the crimes by zip code and added up how many rows went into each

group to get how many aggravated assaults were in each zip code in the data set. We then standardized the data median household income and amount of aggravated assaults in each zip code together to fit them on one graph. A Pearson correlation test gave us a value of -0.46 between these two variables. Despite the graph rising and falling together for the most part, there is no statistically significant correlation between aggravated assaults and median household income.

Aggravated Assaults throughout the year: We again removed all non-aggravated assault crimes. We then separated these crimes into two groups. These groups are crimes that happened in zip codes with less than or equal to a median household income of $50,000 and zip codes with more than a median household income of $50,000. This divided the crimes into two fairly equal halves in quantity. We then graphed these distributions over time. They are pretty equal throughout the year, but there are times (for example, at the end of the year), where there are more crimes in zip codes with median household incomes of under $50,000. Comparing these two groups gave us a very high p-value of 0.87, meaning there is no statistically significant correlation. The statistic value was 0.17.

Results

Note: For ease of reading, and condensing this report to 2-3 pages, all graphs are on the presentation slides at the end of this document.

Map: This graph first confused us because we didn't understand why there was a big hole in the western center of the graph. It turns out this is just the shape of Austin. There are more aggravated assaults in the general the closer you get to the center of the city. There are fewer and fewer aggravated assaults as you get farther away from the center of the city. There are especially few aggravated assaults on the northwest and southeast sides of Austin.

Aggravated Assaults compared to the Poverty Line: In this analysis we looked at how many aggravated assaults were committed in zip codes above and below the poverty line. One major thing to note is that there are a lot more aggravated assaults being committed in zip codes below the poverty line, consistently over the year. However, when we standardized them, it shows that because there are fewer communities above the poverty line it makes sense that there would also be less aggravated assaults above the poverty line. To find a real correlation we would need to look at exact populations per zip code.

Aggravated Assaults compared to Income: We discovered that there is no strong correlation between number of aggravated assaults and income in Austin. This surprised us. There are more aggravated assaults in areas with a median household income closer to Austin's median, simply because more of Austin has a

median household income closer to Austin's median than not.

Aggravated Assaults throughout the year: Aggravated assault numbers are fairly consistent throughout the year. The amount of aggravated assaults in areas with less than $50,000 median household incomes and areas with more than $50,000 median household incomes also stay fairly close to each other. The only noticeable exception is towards the end of the year, there are more aggravated assaults in areas with less than $50,000 median household incomes.

Presentation Link

Project 2 Presentation