

Project 2

Beau Martin, Bekah Grant

Introduction

Austin, Texas, one of the largest cities in the state, is no stranger to crime. Using a crime and housing dataset provided by the City of Austin, we set out to explore three key questions: which zip codes experience the highest rates of specific types of crime, how the median salary in those areas influences the frequency and types of crime, and if the percentage of the population in these areas with disabilities is more likely to have a higher crime rate. By analyzing this data, we were able to uncover some interesting correlations between income levels and criminal activity across different neighborhoods, and demographics. Links to presentation and code are located at the bottom of the paper.

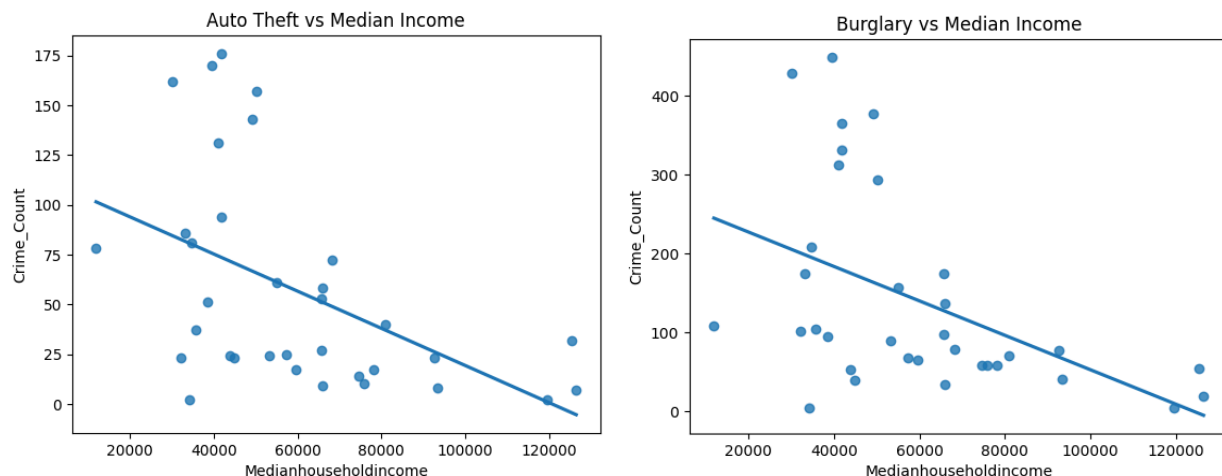
Dataset

The data we are using to explore our ideas is from “data.austintexas.gov”. This table provides details on crime reports and what housing looks like for that area. Crime reports contain information such as what type of crime was performed, what area the crime occurred in, and the date it occurred on. The housing information contains information such as median housing data for a specific zip code and the unemployment rate of that area.

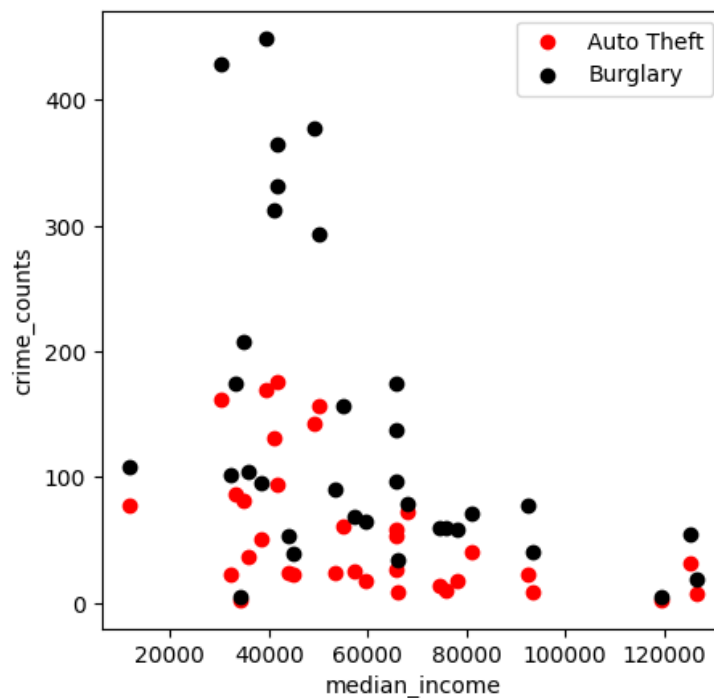
Analysis Technique

To make our data more usable for the analysis we wanted to perform, we used data manipulation techniques to group relevant features and create new ones. For visual analysis, scatter plots helped us explore relationships between two variables. We also applied Pearson correlations to see linear relationships between two variables and their strength. Additionally, we conducted t-tests to compare means between groups and determine if differences were significant.

Results



To explore whether the median household income of an area affects the number of auto thefts and burglaries, we began by creating scatter plots for each crime type. The scatter plot for auto thefts suggested a negative relationship between median income and the number of incidents. To confirm this, we ran a Pearson correlation, which resulted in a correlation of -0.48 and a p-value of 0.003, supporting the negative relationship. Similarly, for burglaries, the scatter plot indicated a negative relationship between median income and burglary incidents. A Pearson correlation confirmed this with a correlation of -0.47 and a p-value of 0.004. There were a total of 1927 auto thefts and 4791 burglaries in our database. On average there were 56 auto thefts per area and 140 for burglaries per area. Therefore, burglaries were more common than auto theft in most if not all areas. Additionally, we found that the standard deviation for auto thefts was 52, while burglaries had a standard deviation of 125, highlighting the variability in crime occurrences across different areas. These findings, combined with the p-values and correlations, help confirm that lower median income areas tend to experience higher rates of these crimes.



We then wanted to explore if there was a relationship between our two groups, auto thefts, and burglaries. To do this, we started by looking at a scatter plot with both groups listed. By looking at this scatter plot we are able to tell that usually in areas where there are a lot of burglaries there is also a lot of theft and vice versa. We also noticed that burglaries points were a lot higher than those of auto thefts so we wanted to look at the averages of both per area. To investigate further into our relationship we performed a t-test between our two groups. The t-test resulted in a p-value of 0.0007. Thus we can conclude that in areas where we see a lot of auto theft we can also expect a high amount of burglaries.

Next we wanted to answer the question, “Do criminals take advantage of disabled people? Through Pearson correlation we were able to find a positive .34 correlation between population with disabilities and the amount of crime committed in the area. Through this information we were able to conclude, as the disability population increases, crime is more likely to occur.

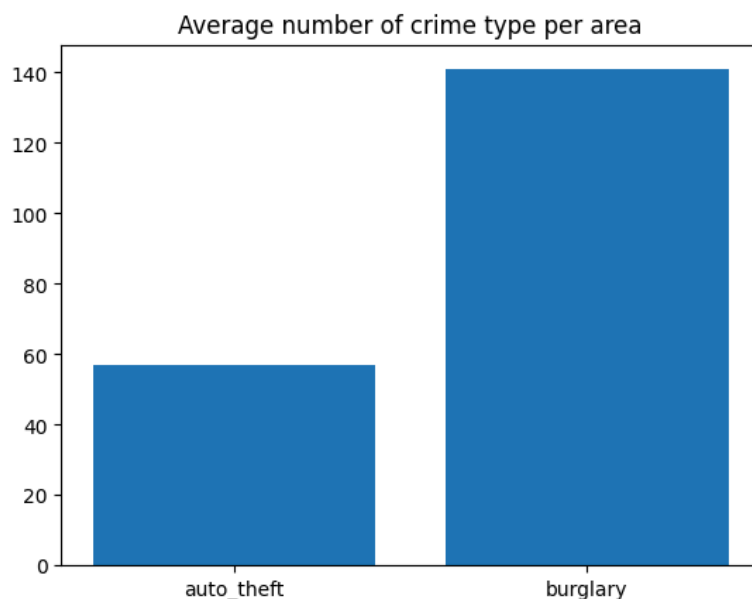
We wanted to look deeper, and compare the lowest percentage of disabled population and the highest percentage of disabled people. The city with the highest prevalence of disabilities exhibits a median income of \$32,131 and a population of 10,124. In contrast, the city with the lowest disability population, located in the 78726 ZIP code, has a median income of \$66,069 and a population of 6,480. We found that although these cities are different there was no correlation between crimes committed in the city with the highest vs lowest disability population.

The difference in median income at cities with a higher and lower income intrigued me so we did a correlation test on this data. We found that there is a negative correlation, as income increases the population with disability decreases. There was also a p-value of .00045, indicating that this was statistically significant. This conclusion was a bit surprising to me!

Tech

Initially starting this project we wanted to see if there was a correlation between median income and types of crimes committed. After analyzing the data we found that the majority of the crimes committed were theft, auto theft and burglary. To find this data we counted the number of burglaries and auto-thefts committed in each zip code. In order to find a correlation between median income and type of crimes committed, we appended the associated median income to each crime committed. We conducted a Pearson correlation on Auto Theft and median income, finding a p-value of 0.003 and a correlation of -0.48. The results confirm a statistically significant correlation, as income increases Auto-theft decreases. Conducting another Pearson correlation test on Burglaries and median income, yielded the resulting p-value of 0.004 and a correlation value of -0.47. This also confirms the negative correlation seen in the scatterplot, as income increases there are less burglaries.

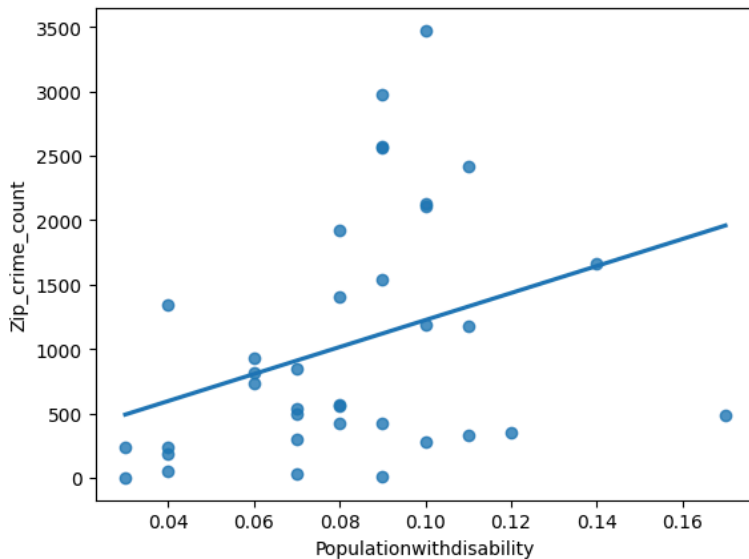
Looking at the standard deviation of auto theft and burglaries. There was a standard deviation of 125 burglaries and standard deviation of 52 auto theft crimes committed per area. On average there are more burglaries than auto theft crimes committed. The bar chart on the left shows that per area burglary is much more common than auto-theft.



We conducted a t-test between auto-theft and burglaries in each area. The t-test resulted in a p-value of 0.0007 this is statistically significant, allowing us to conclude that in areas where

auto-theft is high, burglary is also high.

In our next analysis we compiled the data to assess whether a zip code with a higher percent of individuals with disabilities were more likely to encounter a criminal offense. Thus posing the question “Do criminals take advantage of the disabled?”.



The scatter plot suggests that there is a positive correlation, indicating that as the percentage of individuals with disabilities increase, crime will also increase. The Pearson correlation coefficient was calculated to be 0.336, with a corresponding p-value of 0.0452. Given that a p-value below 0.05 is considered statistically significant, the p-value of 0.0452 confirms a statistically significant correlation between the population percentage of individuals with disabilities and the incidence of criminal offenses.

After isolating the data on the highest vs lowest cities disability population we conducted a t-test to find if there is a higher likelihood of crimes being committed in these locations. The test resulted in a T-value of 1.581 and a p-value of 0.2546. This test suggests that there isn't a strong correlation between the two groups. However, it is important to note that our analysis was based on a limited sample size of only four cities, which may contribute to the observed statistical insignificance.

Before conducting a t-test we gathered more information about the cities with the highest and lowest disability percentage. Through this analysis we found that the highest city, located in 78702 has a lower median income and higher population. The higher population and income could be a result of a higher crime rate.

Presentation:

<https://docs.google.com/presentation/d/17cfvnmU0CdK9y1u2WSzmCshheyDCxq7j24VdpgBerfU/edit?usp=sharing>

Code:

https://github.com/beaumartin25/cs5830_project2