Austin, Texas 2015 Crime and Housing Statistics

Jonah Harmon & Bishal Sarkar
Presentation:
https://docs.google.com/presentation/d/17pczCXnNDcZCzY7VMSIImXhAeEkHnmgdbJJHq6rg9WU/edit?usp=sharing
GitHub: https://github.com/JHamoni676/cs5830_project2

INTRODUCTION

In this project, we look into the intersection of crime and housing data in the city of Austin, Texas (2015). Our analysis aims to find significant correlations between crime incidents across various zip codes. These insights may help city planners, real estate developers, local law enforcement, and individuals looking to possibly live in the city of Austin, better understand how crime trends could impact housing prices and the desirability of specific areas.

DATASET

This dataset contains crime and housing statistics from the city of Austin, Texas, during the year 2015, as well as Austin zip codes. The crime dataset includes information on the types of crime, location, date, and time, while the housing sections include information like housing prices, zip codes, and property characteristics. It is relevant for our analysis because it allows us to examine potential relationships between crime rates and housing prices across different neighborhoods. By using the zip codes, we can cross-analyze crime patterns in specific areas and see how they correlate with changes in the housing market.

ANALYSIS TECHNIQUES

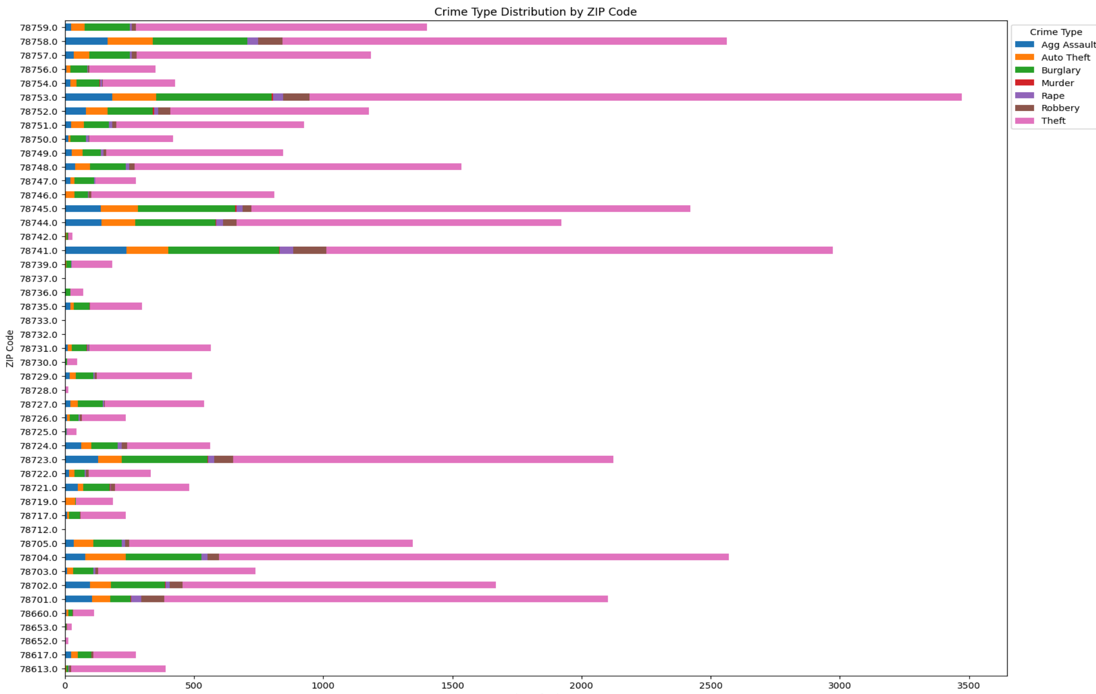
The analysis conducted in this project uses a couple of different statistical methods and charts to explore the relationships between crime counts and housing affordability factors, as well as the relationship between the average income of zip codes and crime counts across Austin, Texas. We began by thoroughly examining the two provided datasets to familiarize ourselves with the available data. The key operations included extracting relevant columns, such as crime counts by zip code, housing affordability for renters and owners, poverty levels, and median rent/home values. To understand the strength of correlations between crime rates and various housing factors, we employed the Pearson correlation. We used the Pearson correlation because it measures the strength and direction of linear relationships between variables. In this analysis, we were more interested in whether areas with more affordable housing for retail/service workers experienced higher or lower crime counts.

We used scatterplots to create the visualization of the relationships between crime counts and housing factors. Scatter plots help in identifying patterns, clusters, and outliers that might not be apparent from statistical coefficients alone. We also used a box plot to view the difference in crime counts and zip codes with an average income below the total median income and those above. This helped to see a significant difference in crime rates in those two groups of zip codes.

RESULTS

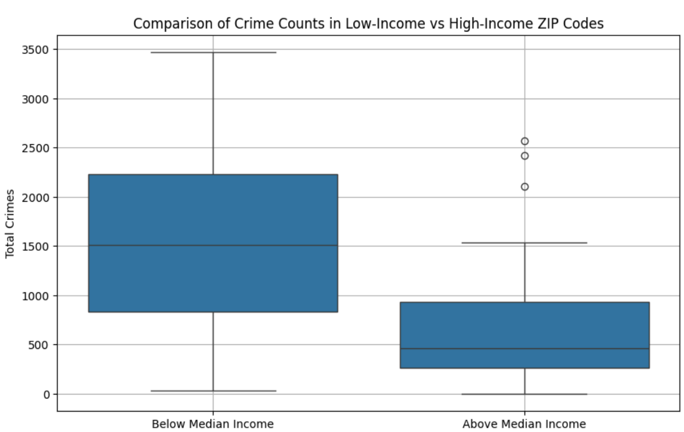
CRIME TYPE DISTRIBUTION:

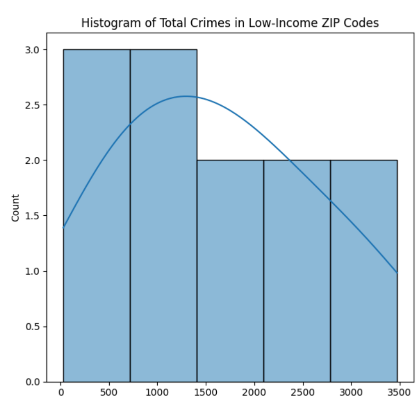
The bar chart ‘Crime Type Distribution by ZIP Code’ shows us several important trends in the distribution of different types of crime across Austin’s zip codes. It shows us the zip codes to possibly avoid when considering buying or renting a home. Also, it shows that theft is by far the most common type of crime, followed by burglary and auto theft. It may also help police departments see areas in need of focus. Luckily, the distribution of violent crimes is much less prevalent.

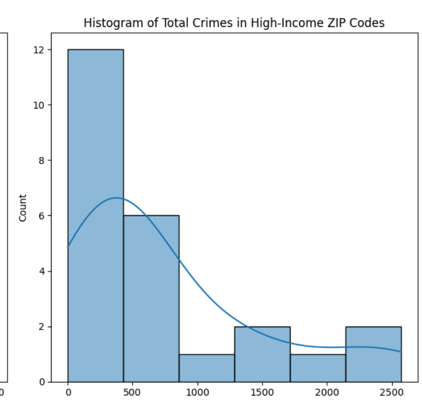


CRIME COUNT INCOME COMPARISON:

The box plot and histograms below show an important relationship between crime counts and income levels across Austin. The boxplot compares total crime counts across zip codes above and below the median average annual income per zip code ($41,869). Those zip codes below the median tend to have a significantly higher crime count than those above the

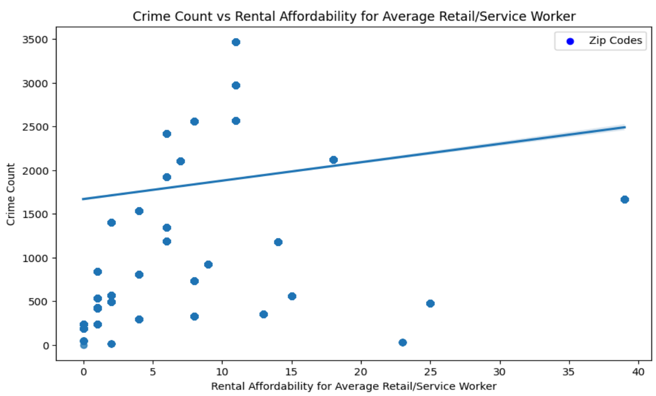


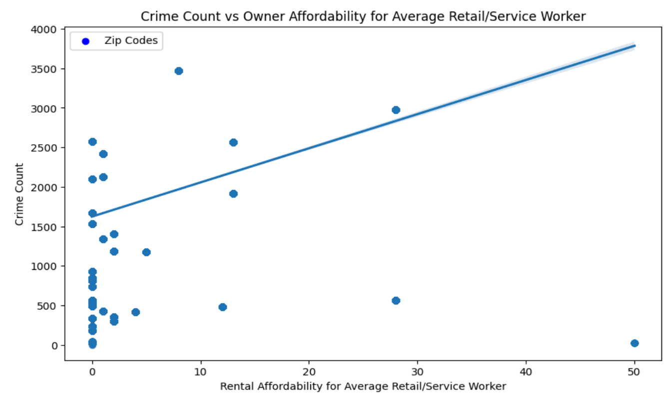




average median. And those above, have a more concentrated distribution of crime counts with fewer outliers. The histogram helps further illustrate these differences. In low-income zip codes, crime counts are more evenly distributed, with areas seeing between 1,000 and 2,000 total crimes. In high-income zip codes, the majority have crime counts clustered at the lower end of the spectrum, with most areas having fewer than 1,000 crimes. This shows a pattern where high-income areas tend to experience significantly fewer crimes overall.

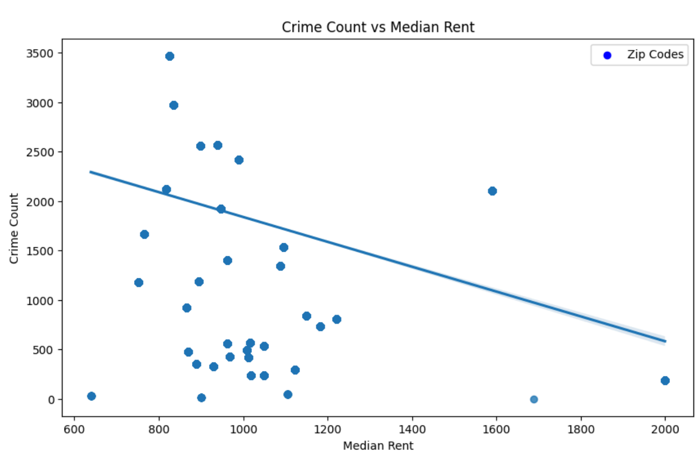
CRIME COUNT VS RENTAL/OWNER AFFORDABILITY:

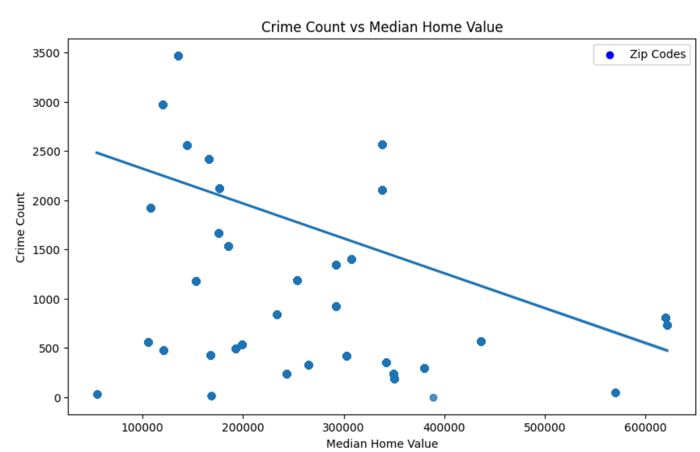




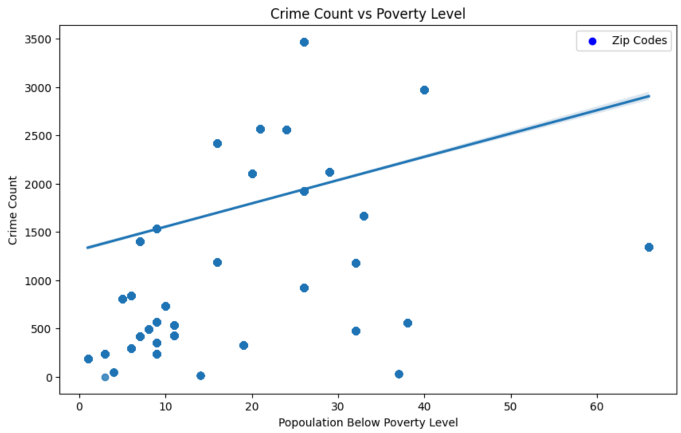
With a statistically significant p-value of 1.49783e-262 and a Pearson correlation coefficient of 0.17849, the dataset showed a relationship between rental affordability and crime rate. This shows that lower crime rates are typically seen in locations where rental accommodation is more affordable for service and retail workers. The owner affordability and crime count also had a little connection with a p-value of 0.395667, although this association was not statistically significant. Similarly, there was little to no correlation found between the median rent and median property values and the crime rate, suggesting that these variables may not have a substantial influence on crime on their own.

Additionally, the analysis of median rent and median home value concerning crime count yielded weak positive correlations. The Pearson correlation coefficients were -0.2716 for median rent and -0.448 for median home values, neither of which were statistically significant where the p-values are 0.0. As shown in the Figures below, the scatterplots for these variables indicate a lack of a clear pattern, with no strong relationships between either rent or home value and crime count. This suggests that while higher costs of living might slightly coincide with increased crime, these variables are not reliable predictors of crime rates when considered independently.





CRIME COUNT VS POPULATION BELOW POVERTY LEVEL:



When comparing crime count to the percentage of the population below the poverty level in each zip code, we find a decent correlation between the two metrics. The chart to the left indicates that as the percentage of people below the poverty line increases, crime counts tend to rise. This is an interesting correlation that could be useful to city planners to help find a way to combat poverty in places with a high poverty rate. It suggests that if poverty decreases, crime rate will also decrease.

TECHNICAL:

The zip code dataset was nice and easy to work with, while the crime-housing dataset was a bit of a mess. It contained lots of null cells, all the percentages were string values containing the '%' symbol, and the cells with USD numbers had to have the '$' filtered out. We found that the ',' were filtered easily by stripping the '$' and converting the values to a float type. Also, dropna() was used extensively. We use four different types of charts to describe the dataset, a stacked bar chart, a box plot, a histogram with KDE, and a handful of scatter plots with linear regression lines. We ran the Pearson Correlation on the scatter plots to find the statistical significance in the scatter plots, and a Mann-Whitney U test to find the significance of the split boxplot. We feel these charts are an effective way to find correlations between different areas of the data. We did have interesting Pearson correlations and p-values. For instance, we found that a few of the charts had negative correlations that suggested an inverse relationship between columns, and some p-values of 0. The p-values of 0 are a little confusing, because we believe it either means the data is bad, or there is an extremely strong significance, or more likely, we did it wrong. We started off simply exploring the dataset, and started with the idea of creating the stacked bar chart above. This helped us see interesting crime rates in the specific zip codes. Afterwards, we ended up creating the boxplot to see how income could correlate with crime count, and how crime count affected different housing metrics.