Michael Hanks
Stockton Smith
CS 5830
Project 2

Introduction:

Over the years, the city of Austin, Texas has been renowned for its high crime rates, which tend to be significantly higher than the national average. Notably, statistics have suggested that Austin is more dangerous than 95% of neighborhoods in the United States[1]. Given this issue, our analyses in this report are designed to bring awareness to the factors that are closely related to Austin's high crime rates, allowing government officials and law enforcement in the city to know how to best address criminal activity, as well as how to most effectively prepare against it.

Our findings were obtained using a collection of Austin's crime and zip code data, which are described in greater detail in the remaining sections of this report. Based on our analyses, there seems to be evidence to suggest that increased crime rates are closely related to poor living conditions, high unemployment rates, and hotter temperatures. Further information on our findings is given in our slide presentation and in our github repository.

Dataset:

We used two datasets for this report. The first given dataset is a report of the crime statistics in Austin in 2015. Each row of this dataset represents a crime that was reported and contains data for population statistics, housing statistics, and other relevant information for the zip code the crime was committed in. These attributes gave us specific details on the atmosphere and life of the area a crime was committed in, allowing us to analyze which specific factors are most closely correlated with crime rates. The second dataset we used contains information about the population per zip code. This allowed us to more effectively make comparisons by adjusting crime counts for population, which is demonstrated in further sections of this report.

Analysis Techniques:

The analysis techniques that we used included Pearson correlation tests, t-tests, scatterplots, bar charts, means, and standard deviations. The Pearson correlation was used in our analyses that compared two variables in one single population. In these analyses, we used scatterplots, as they allowed for easy interpretation of the correlation between the two factors being observed. The t-tests and bar charts were useful for categorical data, as they allowed for valid comparison between two separate datasets on the same variable. Means and standard deviations on this data were also calculated to assist in the interpretation of the significance of the results.

Results:

First, we analyzed the data with respect to crimes per zip code relative to the percentage of rental units in poor condition. To achieve this, we grouped each of the crimes based on the zip code where the crime was committed and added up the total number of occurrences of each zip code. After doing this, we merged this with a separate dataset, which contained the population per zip code. This allowed us to create a crime per population column, which was simply the number of crimes committed divided by the population of the zip code. After doing this, we ran a Pearson Correlation test and drew out a plot (Fig. 1) comparing this ratio and the percentage of units in poor condition, given the zip code. The first p-value we got was 0.2637, which showed that we cannot say that these are correlated.

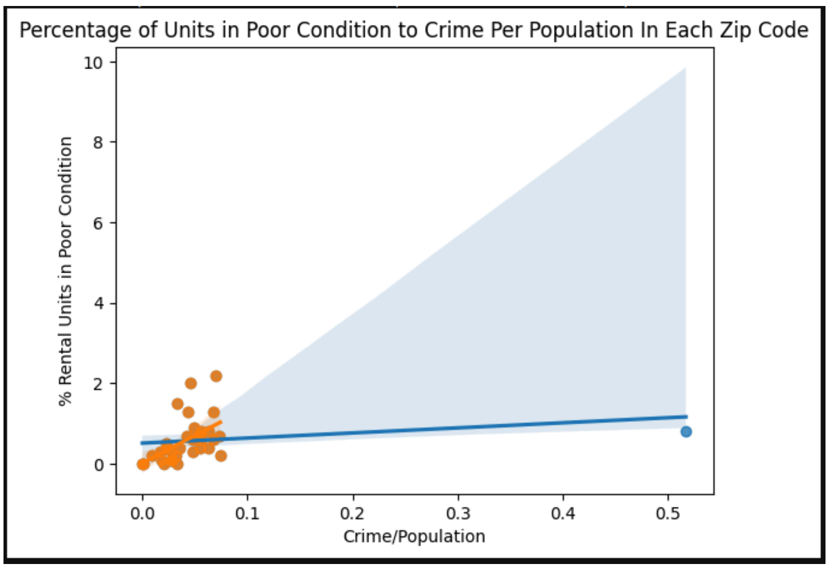

Fig. 1: Percentage of units in poor condition per population in each zip code with a major outlier

However, it is important to note that there is a clear outlier in the data; there is a zip code with over half a crime committed per person, which seems to seriously bias the data. After reviewing this, we removed that row from the dataset, generated a new plot, and ran a new Pearson Correlation test (Fig. 2). The p-value for this is about 0.0019. Such a low p-value allows us to reject the null hypothesis, implying that crime rates are higher when rental units are in poor condition.

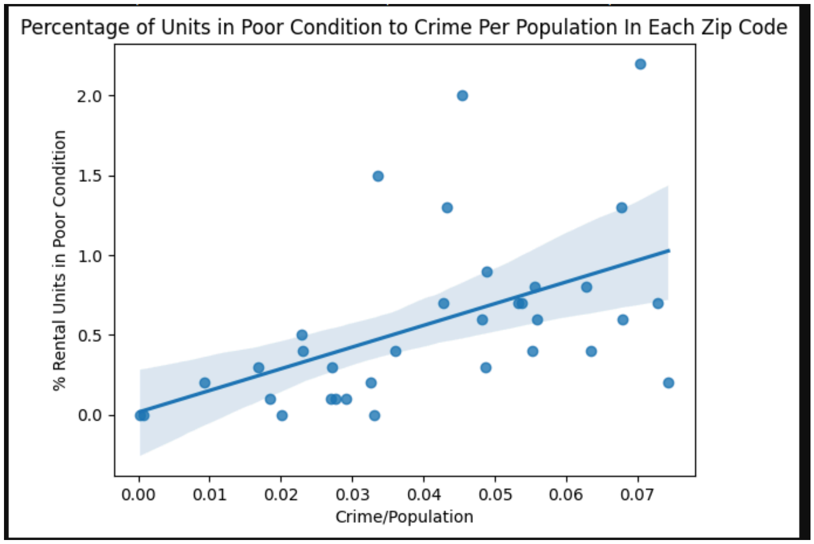

Fig. 2: Same data as shown in Fig. 1, but with the extreme outlier removed. Evidence suggests that rental units in poor condition lead to more crime in a given population.

Next, we used the crime per population data to explore the relationship between crime rates and unemployment levels. With even more significance (p-value ≈ 0.0004), it was observed that crime rates tended to be higher in zip codes with higher percentages of unemployment. The strength of this relationship was verified using a Pearson correlation test. The observed relationship is described in the scatterplot below.

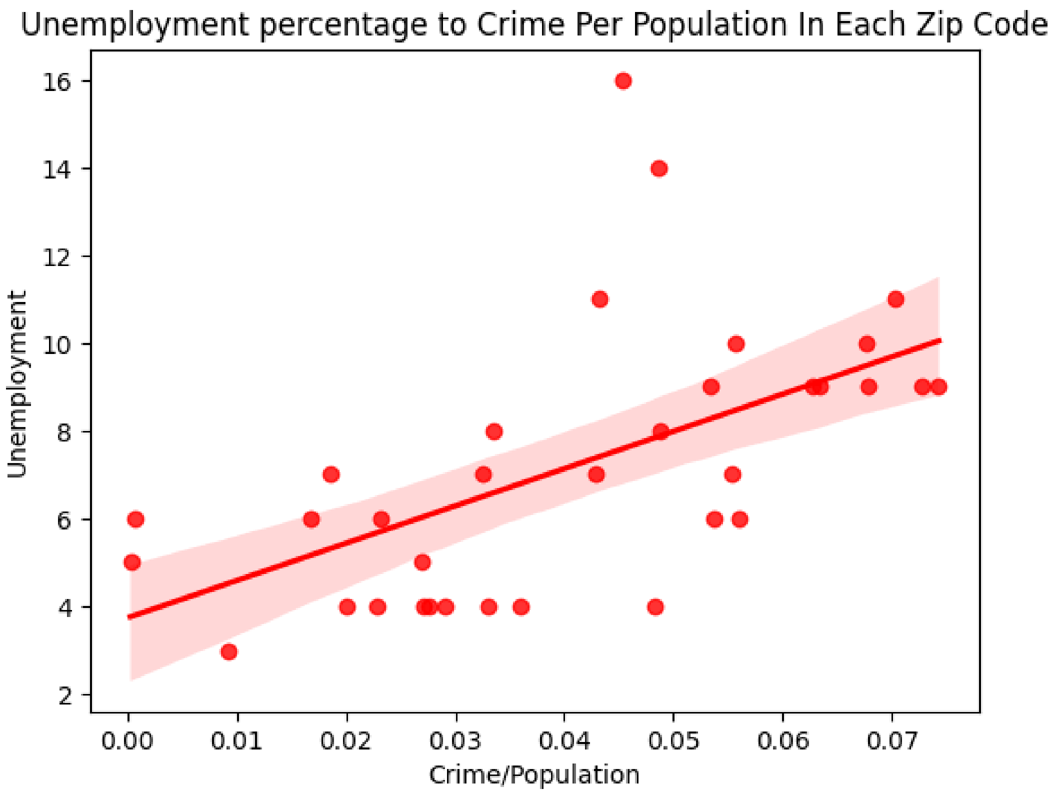

Fig. 3: High crime rates are closely tied to high unemployment rates

Finally, we elected to analyze the data in terms of which months of the year possessed the highest crime rates. To do this, we divided the months into two categories: the months from April to September (shown in red below) were designated as the hotter-temperature period, and the months from October to March (shown in blue) were designated as the regular-temperature period. Expected crime rates were calculated for each month by assigning them a proportion of the total number of crimes reported in the year, adjusted for the number of days in each month.

By comparing the rates of actual crime with those of the expected crime, it was discovered that the hotter-temperature months had statistically significantly higher crime rates than the average-temperature months, with a p-value of approximately 0.002 computed from a t-test. Out of the seven months that experienced more crimes than expected, six of them were the six months in the hotter temperature category. The hotter-temperature months averaged 102.99% of expected crime with a standard deviation of 2.53%, and the cooler-temperature months averaged 97.02% of expected crime with a standard deviation of 2.05%. The small variability for each category reinforces the idea that months with hotter temperatures are disproportionately affected by crime.

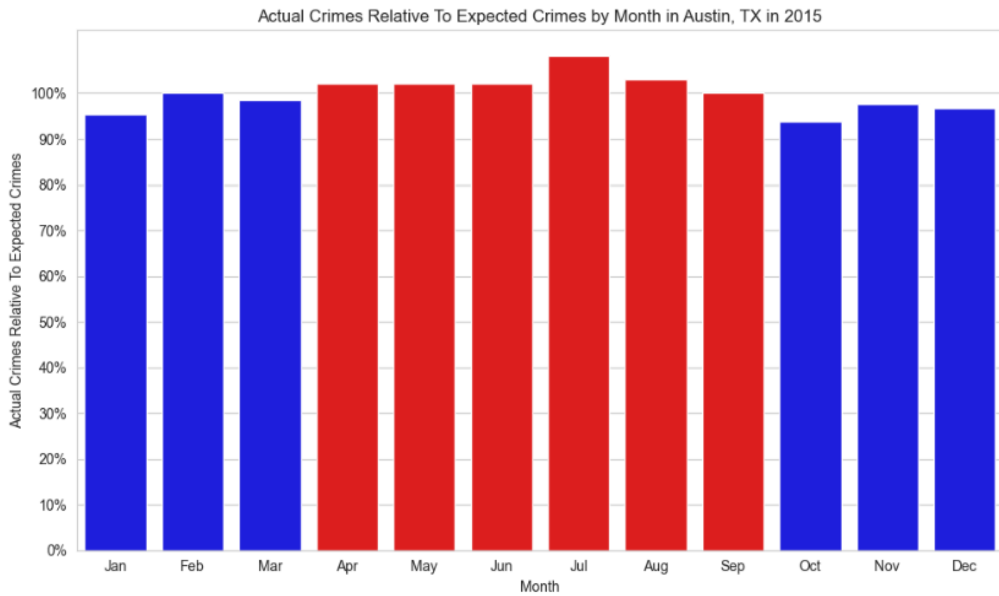

Fig. 4: Evidence suggests that hotter temperatures lead to higher crime rates

Ultimately, our results indicate potential steps that Austin's government officials and law enforcement can take to most effectively respond to criminal activity. These steps could include passing legislation so that owners of rental units are required to keep their properties in certifiable condition. Efforts by government leaders to reduce unemployment may be most beneficial in cities with the highest crime, as the two factors are strongly positively correlated. Additionally, it can be expected that crime rates will be higher during the hotter months of the year, so law enforcement and security workers should be deployed in greater numbers during these times. While such measures will never eradicate crime completely, we certainly expect that they would minimize its impact, improving the quality of life for the city's nearly one million residents[2].

Technical Section:

In order to prune our dataset for analysis, we began by taking the two provided datasets and merging them together. This enabled us to more easily make analyses based on crime per capita, as the crime counts and population tallies were originally kept in separate tables. Other preparatory steps that were taken included removing symbols (such as commas and percentage signs) from column values, condensing zip code data into a single row in order to avoid bias from duplicates, and adding new columns with custom calculations (such as crime per person) into the table.

As mentioned above, most of our claims were based on p-values generated from Pearson correlation tests and independent t-tests. The Pearson tests were appropriate when analyzing two variables from the same population, and the t-tests were appropriate when analyzing the same variable from two separate populations. Additionally, we had to correct some of our data after initially failing to normalize crime counts based on zip code population.

Had we elected to ignore this correction, our comparisons between large populations and small populations would have been meaningless, as they would have only reflected that more crimes are committed when there exist more people to commit them.

Appendix and References:

1) https://www.neighborhoodscout.com/tx/austin/crime
2) https://worldpopulationreview.com/us-cities/austin-tx-population