CS 5830 - Project 2

Anny Chen, Zion Steiner

INTRODUCTION

The purpose of this report is to communicate general statistical measures of interest found in the *Austin Crime Report 2015* dataset. In addition to the type and location of criminal reports, the dataset includes housing statistics associated with where these crimes took place. The aim of this analysis is to determine how crime is geographically distributed within Austin, TX, in 2015. Descriptive statistics such as correlations, means, standard deviations are used to supplement the findings of a hypothesis test. As a result, we found evidence that suggests crime is more common in the eastern half of Austin than the western half.

DATASET

The *Austin Crime Report 2015* dataset is mainly about crime and housing statistics for Austin, Texas in 2015. This dataset informs us of the crime position, crime data, clearance status and the wealthy status of the crime position. The dataset domain of crime and housing dataset is criminal statistics. Crime and Housing dataset is a specific part of criminal statistics, which includes specific time and position crime information. The format of the dataset is a comma-separated values file. Before using the dataset, we transform percentage or dollar amounts to the float type for further analysis.

ANALYSIS TECHNIQUE

The proportion of total reported crimes occurring in each zip code was calculated to give a measure of region-specific criminal activity. Given the coordinates for where each crime occurred, we were able to visually analyze which areas of the city experienced more reported crime. Plotting crime proportion side-by-side with median household income shows that crime is more common among lower-income areas. Fig. 1 shows that the split occurs roughly at x coordinate = 3100000, with higher income to the west and higher crime levels to the east. A two-sided hypothesis test was conducted to compare the mean crime proportion between zip codes on either side of the city. With a t-statistic of -62.0 and a p-value of ~0.000, we have evidence suggesting that less crime occurs in the western half of Austin than the eastern half.

Statistics detailing the distribution of crime proportion by zip code are shown in Table 1. These show there is a spread of ~9% between zip codes with the lowest crime occurrences and those with the highest. The minimum proportion is 0.0026%, so at least one zip code experiences almost no reported crime and another experiences 9% of all crime in the city.

RESULTS

The hypothesis test detailed above shows that crime is more prominent in the eastern, lower-income region of Austin (at least in 2015). While these results are statistically significant, the correlation between crime proportion and the x coordinate of the crime is only loosely positively correlated with a correlation of 0.294 and an associated p-value of 0.00. This can be

explained by the graphs in Fig 1. The high-crime/high-income split is not perfect; there are several regions where this split isn't apparent which leads to the loose positive correlation. However, in tandem with the visual analyses, the hypothesis test's conclusion that crime is more common in the eastern half of Austin has some practical significance.

This information may be useful for law enforcement officials or city planners to make informed decisions regarding police patrol schedules or city zoning. Such officials may also be interested in knowing there exists a positive correlation (0.541) between crime proportions and the number of owner units affordable by teachers within an area. This figure doesn't suggest causation, but that is not to say it is not in Austin PD's best interest to keep an eye on any wily teachers lurking about east Austin.

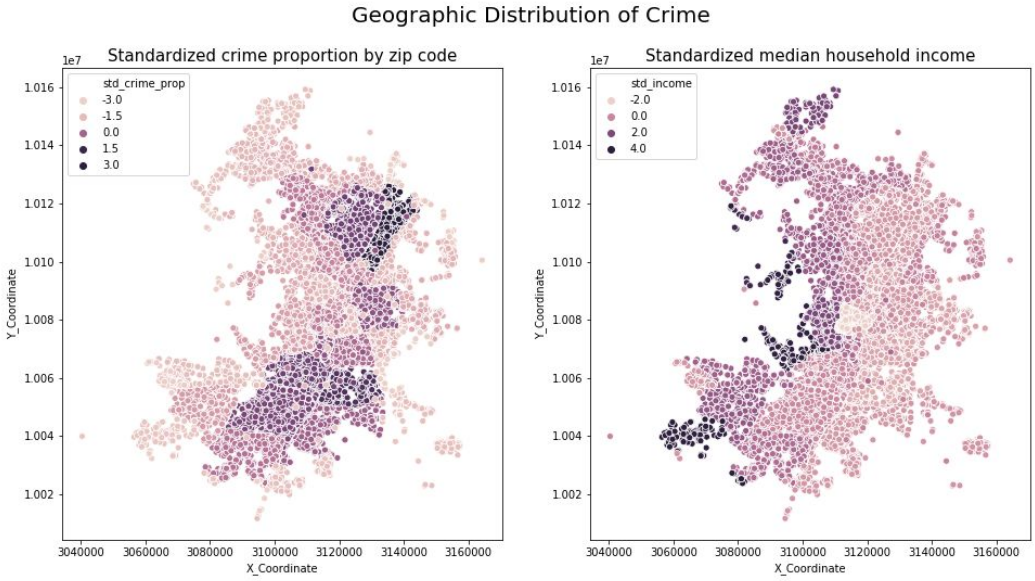

Fig 1. Criminal activity is concentrated higher in the eastern half of Austin

Table 1. The spread between the zip codes with the highest and lowest occurrences of crime is 9%

count	38573.000000
mean	0.046828
std	0.025235
min	0.000026
25%	0.024032
50%	0.049802
75%	0.066445
max	0.090011

Name: crime_prop, dtype: float64

Presentation link:
https://docs.google.com/presentation/d/199_QlGP2VxgMai--4H5K80zQmTuex1Srj5goa65AS9o/edit?usp=sharing