ANNY CHEN, ZION STEINER

CSI AUSTIN

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.

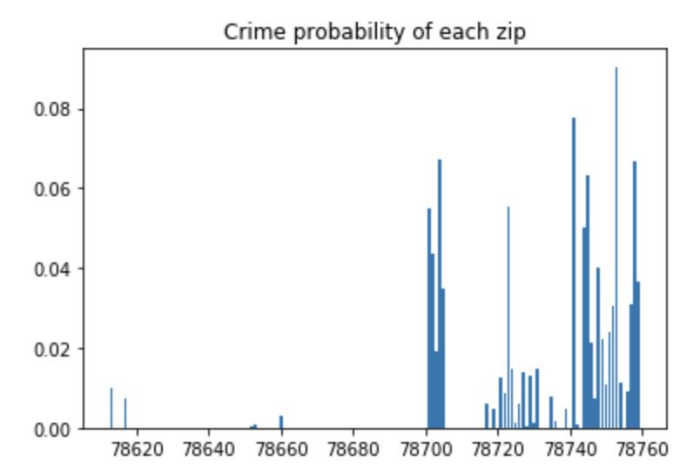
The Criminal Geography of Austin

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 and change the string type of crime date to date type.

ANALYSIS TECHNIQUE

The proportion of total reported crimes occurring in each zip code was calculated to give a measure of region-specific criminal activity. Calculate the Pearson's correlation coefficient between crime proportion and other columns and find crime proportion and median household income is loose negative correlation. Given the coordinates for where each crime occurred, we were able to visually analyze which areas of the city experienced more reported crime. Plotting standard crime proportion side-by-side with standard median household income shows that crime is more common among lower-income areas(scatterplots of standardized data). 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.



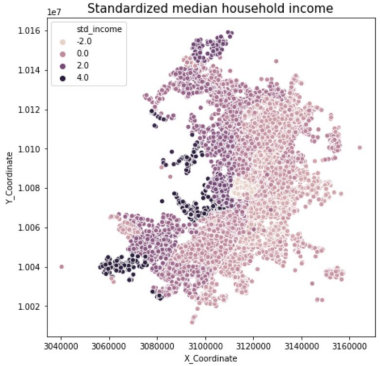
count	38573
mean	0.046828
std	0.025235
min	0.000026
25%	0.024032
50%	0.049802
75%	0.066445
max	0.090011
Name: crime_prop, dtype: float64	

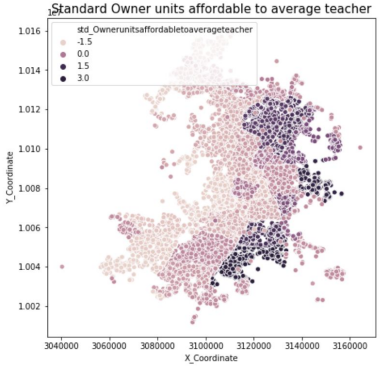
RESULTS –CRIME PROPORTION

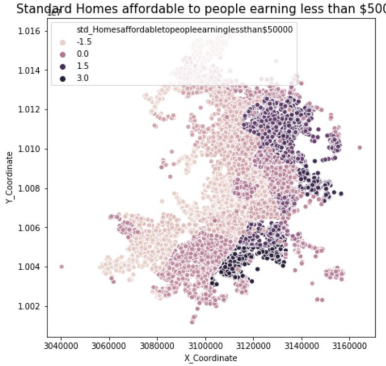
Pearson correlations with crime_prop	Pearson's correlation coefficient	p value
crime_prop	1	0
Ownerunitsaffordabletoaverageteacher	0.540611	0
Homesaffordabletopeopleearninglessthan$50000	0.520392	0
HispanicorLatinoofanyrace	0.518378	0
Ownerunitsaffordabletoaverageartist	0.500552	0
……	……	……
Medianhomevalue	-0.387767	0
Medianhouseholdincome	-0.401332	0
Name: crime_prop, dtype: float64		

Pearson correlations with Owner units affordable to average teacher	Pearson's correlation coefficient	p value
Ownerunitsaffordabletoaverageteacher	1	0
Homesaffordabletopeopleearninglessthan$50000	0.9967	0
Ownerunitsaffordabletoaverageartist	0.929891	0
HispanicorLatinoofanyrace	0.82655	0
…...	……	……
Medianhouseholdincome	-0.593866	0
Medianhomevalue	-0.724997	0
Name: Ownerunitsaffordabletoaverageteacher, dtype: float64		


Standardized crime proportion by zip code
1e7
std_crime_prop
-3.0
-1.5
0.0
1.5
3.0
Y_Coordinate
X_Coordinate
1.016
1.014
1.012
1.010
1.008
1.006
1.004
1.002
3040000
3060000
3080000
3100000
3120000
3140000
3160000


Standardized median household income
1e7
std_income
-2.0
0.0
2.0
4.0
Y_Coordinate
X_Coordinate
1.016
1.014
1.012
1.010
1.008
1.006
1.004
1.002
3040000
3060000
3080000
3100000
3120000
3140000
3160000


Standard Owner units affordable to average teacher
1e7
std_Ownerunitsaffordabletoaverageteacher
-1.5
0.0
1.5
3.0
Y_Coordinate
X_Coordinate
1.016
1.014
1.012
1.010
1.008
1.006
1.004
1.002
3040000
3060000
3080000
3100000
3120000
3140000
3160000


Standard Homes affordable to people earning less than $50000
1e7
std_Homesaffordabletopeopleearninglessthan$50000
-1.5
0.0
1.5
3.0
Y_Coordinate
X_Coordinate
1.016
1.014
1.012
1.010
1.008
1.006
1.004
1.002
3040000
3060000
3080000
3100000
3120000
3140000
3160000

T- TEST

STATISTIC=
-62.00265755032987, P-VALUE=0.0