**Project 2 – Crime and Housing in Austin, TX (2015)**

https://github.com/annmarie2/cs5830_project2
https://docs.google.com/presentation/d/1w78B53bIPPzx7FN1omCroKae0ppI9CL7tu1xc258Oac/edit?usp=sharing

1. **Introduction**

Our analysis will provide the law enforcement and local government of Austin, Texas, with insight into the correlations between crime and housing cost in the area, as well as how violent crimes are handled in their various council districts. We found no significant correlation between the percent of solved crimes in an area and a rise in the cost to buy or rent a home. We also found no significant difference in the amount of crimes reported in neighborhoods with higher affordability vs lower affordability. We did, however, find that crimes in more affluent neighborhoods are resolved quicker. We also found variations in crime clearance across council districts. To find these results, we analyzed data from the Austin Crime Report 2015 dataset using charts and statistical tests.

2. **Dataset**

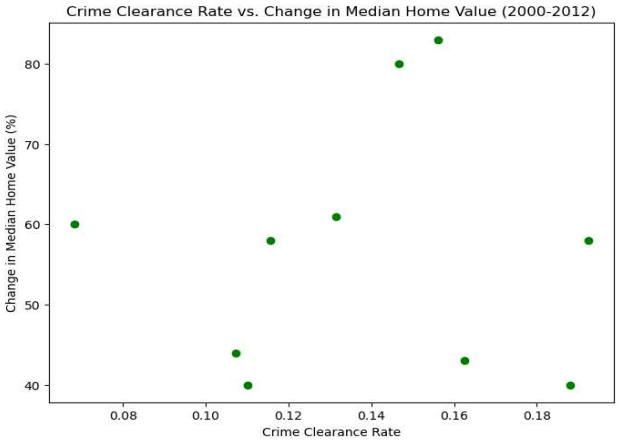
The Austin Crime Report 2015 dataset contains data on all crimes reported in Austin, Texas in 2015, as well as financial and housing information for the areas in which those crimes were reported. It is stored in a .csv file, which makes it ideal for analysis using statistical software such as pandas and seaborn.

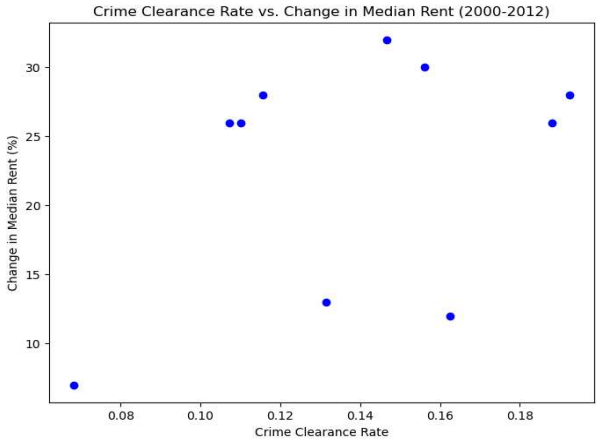
3. **Analysis technique**

Our principal aim in this analysis was to find correlations in the data between various categories. To do this, we used scatter plots, distributions, Pearson correlations, and t-tests. Scatter plots and distributions were chosen for their ability to reveal general trends and relationships between large amounts of data. After observing general trends in these charts, we used Pearson correlations and t-tests to determine the probability of any correlations.

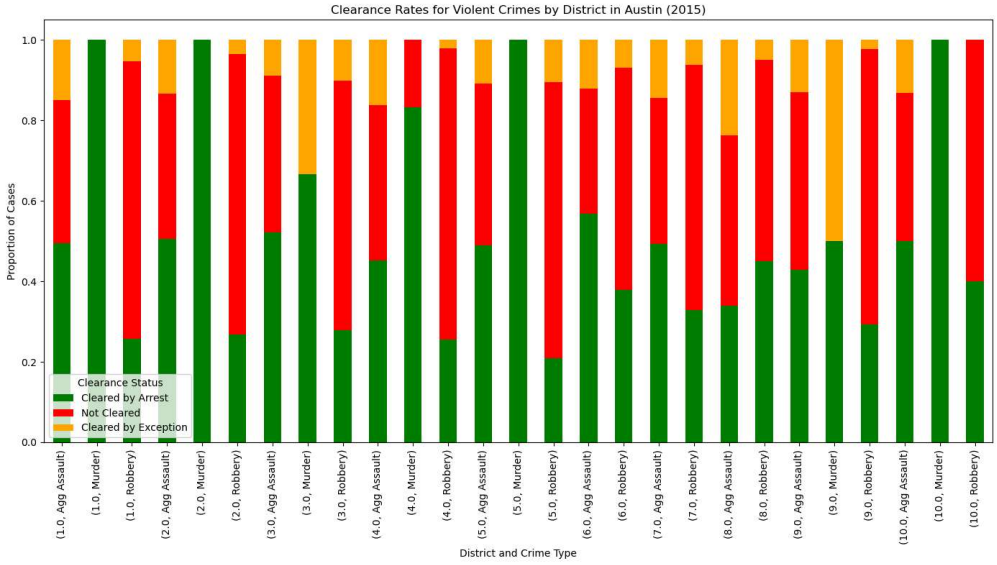
4. **Results**

In our first analysis, we wanted to see if there was any correlation between the percent of crimes that were solved in an area and a rise in home values and rental prices (from 2000-2012) in that area. We found no evidence to suggest that rising prices contribute to a change in the amount of crimes that are solved. This is helpful for local government and law enforcement because it shows that changing prices have not made a difference in crimes being solved.



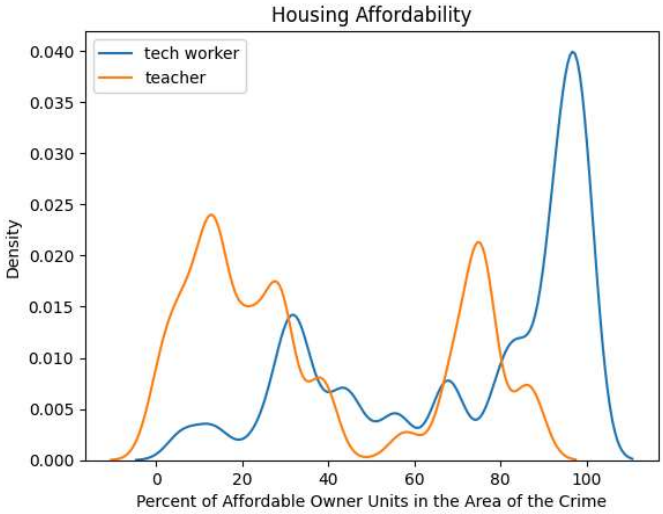




Next, we wanted to see how different council districts deal with violent crimes like robbery, aggravated assault, and murder. We found significant variability between districts. Districts 1, 2, 5, and 10 had a higher percentage of crimes cleared by arrest, while District 4 had a higher percentage of crimes uncleared. Additionally, Districts 9 and 3 had the largest proportion of crimes cleared by exception (which can happen, for example, when the offender is killed or the victim refuses to cooperate). This information suggests that District 4 could use more help clearing crimes, and Districts 9 and 3 could use further investigation into why so many of their crimes are being cleared by exception. The local government can use this as an indication of where to direct their attention.

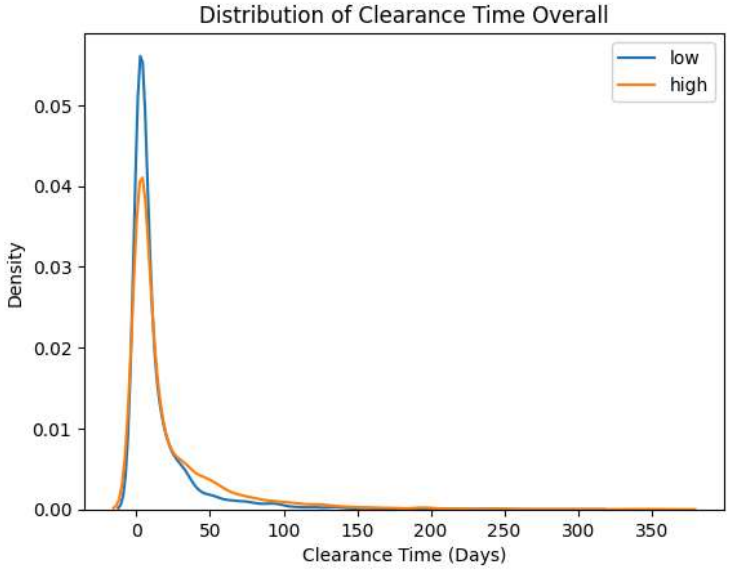



To analyze how housing affordability affects crime rates, we looked at areas where even the average tech worker cannot afford a majority of homes, and areas where a majority of homes are affordable to the average teacher. We found that the areas with more expensive housing had only 7% less crimes reported than the areas with less expensive housing. This information is useful to Austin's law enforcement, as it suggests that the crime rate is the same for areas

with expensive housing and areas with cheap housing. Law enforcement should, therefore, give equal attention to both.




From here, we looked at the time elapsed between when a crime is reported and when a crime is cleared. We found that crimes committed in areas with more expensive housing are cleared an average of seven days sooner than in areas with less expensive housing. Some possible explanations are that areas with expensive housing have more funding for law enforcement, or that wealthy neighborhoods are being prioritized when clearing crimes. The local government in Austin can use this information to root out inequality in Austin's law enforcement and give more priority to areas with cheaper housing.



5. **Technical**

This dataset contains many values stored as strings, with leading dollar or percent signs. These values had to be converted to floats and stripped of leading symbols for every analysis we conducted. We also dropped N/A values in the columns we were analyzing. For our analysis of change in home value and rent, we grouped the data by Council District and Clearance Status, as well as by Council District and Change in median home value 2000-2012. We also filtered the dataset into tables of districts with higher or lower clearance rates. For our analysis of clearance time, we made tables for all crimes in areas where >50% of housing was affordable to the average teacher, and all crimes in areas where <50% of housing was affordable to the average tech worker. We then made tables for each clearance status within these demographics. Finally, for our analysis of violent crimes, we filtered the dataset into a table of violent crimes, categorized these crimes by clearance status, and grouped by council district.

Distributions show the average across a category, so we used a distribution to analyze the relationship between crime rate and housing. This is the same method we used for analyzing clearance time. We used a t-test to verify these results, as this shows whether there is a significant difference between two populations with the same variable. Scatter plots show the relationship between two values over many data points, so we used them to analyze the correlation between change in rent and crime density. We used a bar chart to show the clearance status of violent crimes across different districts, as this easily displays data across various categories.

A setback we encountered during the analysis process was determining how to interpret the clearance status of “Not Cleared”. At first it was assumed that, since there was a clearance date for every status, crimes that were “Not Cleared” were those that had been resolved without the need of arrest. Then we changed our interpretation to be crimes that are still waiting for clearance. To that end, we have interpreted the clearance date on these crimes to be the last date that anything action was taken toward clearance.