Los Angeles Poverty and Crime Severity

Olivia Kaczmarskyj, Tyler Lynch, Liam Tolbert, and Ethan Walls

Research Question

Our research question for this study was, "How do poverty and related factors affect crime severity?" Crime is an age-old problem in cities across the world, and attempts to study its patterns and how it relates to other potential factors in a city are continuously helpful in the global battle to reduce crime. While crime rate has been studied more, crime severity is also a significant metric to study. A city with enough violent crimes has the potential to be significantly more dangerous than a city with just a high crime rate. Additionally, since there is an already established link between poverty and crime rate, if patterns can be found that uncover more about crime severity from poverty, then the causes of those patterns can be used to quash what is arguably the larger issue in the global fight against crime.

Literature Review

Past research already suggests the link between poverty and crime rate. In a study done by the American Action Forum, they came to the conclusion that, " Without reducing poverty—and more specifically, income inequality—as well as racial bias and rolling back harsh sentences for certain crimes, the United States will not meaningfully reduce its prison population." [5]. The paper itself investigates the link between incarceration and poverty by examining characteristics of the current prison population as well as exploring which crimes offenders are most often jailed for. Our goal is to add additional features, such as location and public funding, and expand on the study to see if there is a further correlation between our new features, poverty, and crime.

Some studies have already tried to analyze specific cities and find a link between certain types of crimes and poverty. According to P.B. Stretesky et al. there is not a direct correlation between poverty clusters, areas with a high density of poverty, and violent crime. However, they did find a correlation between homicides and a city's disadvantage score, a variable made up of various city statistics, including poverty. As they state it, "disadvantage has a much stronger relationship to homicide in cities with high levels of poverty clustering." [9].

While the data in the paper are useful, there are a few problems. The disadvantage score that they use in their conclusion can be seen as racially biased. The score is made up of 4 variables: percent unemployed, percent black, percent poverty, and percent of female-headed households, all of which could be skewed towards the black community. Having the crux of the paper based on the amount of black people in the city can cause harmful conclusions to be made

about black people specifically when it comes to violent crime. Another problem is that the study was conducted in 2004, so the data might not be relevant to modern times. Our goal is to conduct our own research on the Los Angeles area while trying to keep the focus solely on poverty and all types of crime, and avoid features that involve race or focus on minority groups as much as possible.

Methods

We broke the analysis of our data into two parts: location data and data over time. In terms of location data, we focused solely on 2017 data of poverty and crimes. The poverty dataset included the percentage of people receiving under the acceptable living wage [6]. The feature used was poverty rate, which is based on the MIT living wage calculator while taking geography into account for different locations within Los Angeles. The location data contains latitude and longitude data in Los Angeles county during the year 2017, which we later narrowed down to just Los Angeles city to fit with the crime dataset. The crime dataset included every crime recorded since 2010 [2]. This includes the location in latitude and longitude and the type of crime. Every crime in the dataset is marked with their respective unique crime code, designated by the LAPD in accordance with their severity. The lower a crime code is, the more severe the crime. The LAPD bases these crime codes off the FBI's official hierarchy of crime severity. We later narrowed down the dataset to only during the year 2017 to fit with the poverty dataset.

We used these datasets to create heatmaps of different features for the city of Los Angeles. We used heat maps because they are the best method to represent data about some location. First, using the crime dataset, we created a heatmap of crime frequency by location. We used crime count specifically because we wanted to be able to show the true relationship between crime count and crime severity without manipulating the data too heavily, such as adding population as a third variable to account for, and thus risk introducing bias. Since there were many crimes recorded during 2017, we decided to use 2D bins by latitude and longitude and color each bin based on crime frequency. Second, using the 2017 living wage dataset, we created a heatmap of the percentage of people living below the acceptable living wage by location. We also binned latitude and longitude to match the bins of the crime dataset and colored the heatmap based on the previously mentioned percentage. Lastly, using the crime dataset once again, we created a third heatmap to display crime severity by location, which was also binned in the same way as the previous two heatmaps. Because the heatmap showed larger numbers as brighter colors, we decided to normalize the crime codes and convert them into a severity metric that would range from zero to one; zero being the lowest severity and one being the highest. In this way, we could show potential hotspots where crimes were the most severe.

We believe that the techniques that we used to create the three heatmaps are the best possible way to represent the data in order to try to find a correlation. However, we do have some minor concerns about our data and methods that could introduce bias into our results. For

example, the way we binned our heatmaps could be an issue. The crime dataset was very large (around 200,000 rows for 2017 alone), but the living wage dataset, in comparison, was much smaller. Because of this, we had to limit the number of bins so the heatmaps were filled in enough so an accurate visual comparison could be made. Similarly, the living wage dataset included data throughout the entirety of Los Angeles County, which is larger than Los Angeles as a city, so we had to trim that dataset to match up more with the crime dataset. Secondly, the location-based living wage and crime data can only include 2017 because that is the only time when both these datasets intersect. While there ended up being enough data to make a comparison, the limited timeframe could introduce some uncertainty into the results. Lastly, although it was obtained from the FBI, the method that the LAPD used to determine crime severity could be another issue that could have affected our heatmap. The LAPD gives crime code numbers for each instance with lower crime code numbers correlating with higher severity crimes, and vice versa. While we remained consistent with the LAPD's definition of severity, the bias related to their determination of the severity of crimes carried over into our own analysis. Specifically, the code values assigned to each crime are not at regularly spaced intervals, such as 110, 115, and 200. When creating our severity metric, this proportionality of severity is retained in our heat map and could potentially influence the results. Additionally, the ucr handbook in which the FBI included its crime severity scale was last updated in 2013, so the information may be slightly out of date.

The data over time came from the LAPD crime data [2], a second poverty dataset that contained the estimated percent under the poverty line for Los Angeles over the years 2010 to 2019 [8], and datasets containing budgets for different departments in the City of Los Angeles [1]. Using this data we constructed some preliminary correlation matrices to see which types of policies we should include to help with our analysis between crime rate and poverty. Using the information we gathered from our heat maps and preliminary analysis, we made an initial linear regression model to predict the crime rate in the city of Los Angeles. We chose which departments to include in the model based on a variety of factors such as date of creation and relevance to the socioeconomic status of residents. Our initial model originally included budgets from the LAPD, Economic and Workforce Development Department, Recreation and Parks Department, Public Transportation Department, and Neighborhood Empowerment Fund along with the city's poverty level. To train our model we used an Akaike Information Criterion to optimize our linear regression model, as it is good for working with time series data and known to help protect against (not completely ignore) over-fitting [7]. The algorithm also works well with R which is where most of the preliminary statistical analysis was done.

After training the model we had to reconsider our features as one of them was causing our model to have too high of an accuracy. The Neighborhood Empowerment Fund was being weighted too highly to be reasonable, causing an R-squared of 98%. The fund could be seen as a proxy feature for poverty and crime, as the fund was raised and lowered in areas where a neighborhood could be seen as needing assistance because of a high poverty or crime rate [3].

After reevaluating the feature, we realized that it was a less reliable feature than the other departments. A large factor in this decision was the discrepancy in the amount of funding it received, which was significantly lower than the other departments. We therefore decided to limit our analysis to departments with relatively the same amount of funding in order to reduce possible bias in the model, and removed the Neighborhood Empowerment Fund. After removing the Neighborhood Empowerment Fund from the model we were left with poverty levels and budgets from the LAPD, Economic and Workforce Development Department, Recreation and Parks Department, and Public Transportation Department. When stepping our features through the Akaike Information Criterion algorithm we were left with a model that included the Economic and Workforce Development Department, Recreation and Parks Department, Public Transportation Department, and the Poverty.

Results

Looking for matching patterns of bright and dark spots, which correlate to high poverty, crime, or severity rates depending on the specific visualization, is the best method in attempting to find a correlation between data using heatmaps. In the case of our data, when the poverty and crime frequency heatmaps were compared (Fig. 1 & Fig. 2), bright spots indicating high poverty and high crime frequency were in the same locations. Our crime severity heatmap (Fig. 3), however, had no distinct bright spots or dark spots, with merely a splash of the same color with no discernible pattern across the entire city. This indicates that there is not in fact any pattern to crime severity from a geographical standpoint. Similarly, overlaying this map (Fig. 3) with that of crime and poverty (Fig. 1 & Fig. 2), there is no pattern specifically between poverty rate and crime severity nor crime rate and crime severity.

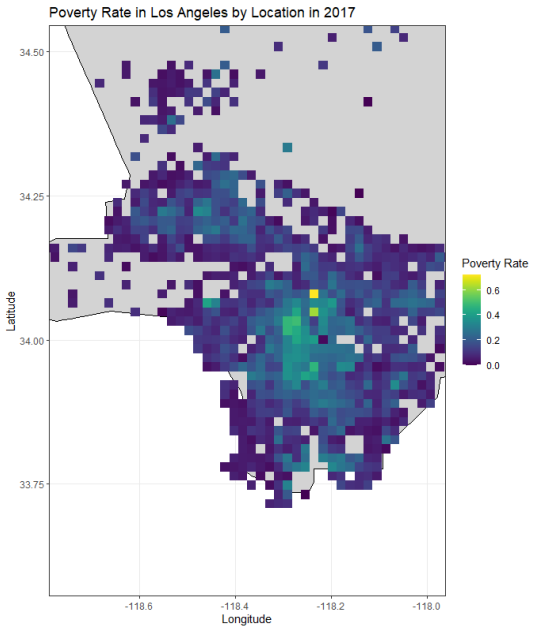

Figure 1: Poverty Rate Heatmap

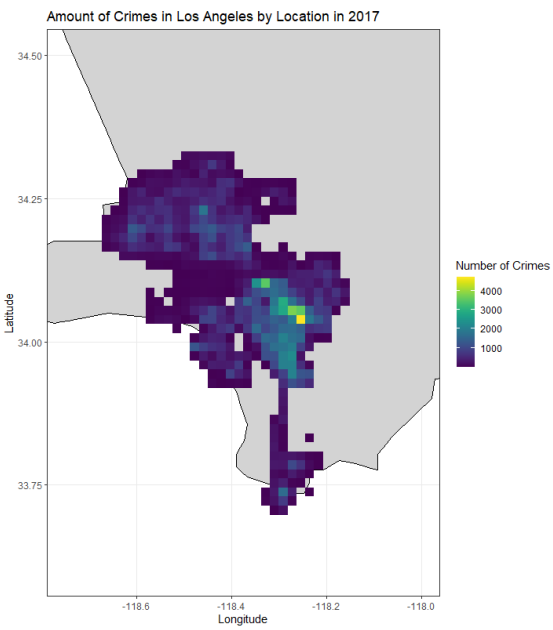

Figure 2: Crime Rate Heatmap

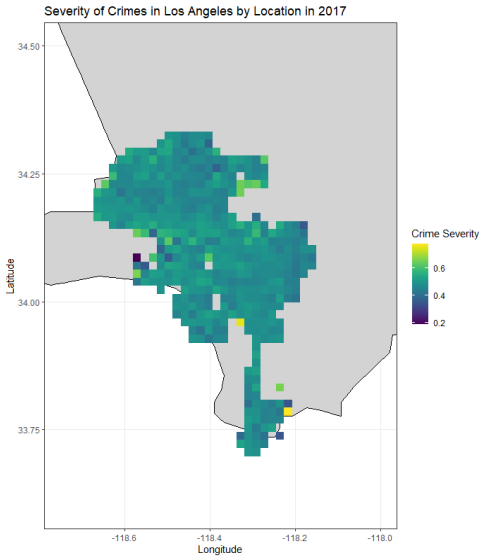

Figure 3: Crime Severity Heatmap

One interesting result of our analysis was the location of the concentration of poverty and crime in Los Angeles; this would be the bright yellow spots in Figures 1 and 2. We were curious as to what this hotspot was, and found that this particular location is Skid Row. Skid Row is a neighborhood in Los Angeles known to be riddled with homelessness and a strong police presence. This was discovered and confirmed in two steps, first by talking with a Los Angeles resident to figure out what it could be. He suggested that it may be Skid Row, since visually it looked to match where the neighborhood is in the city. We then confirmed this by overlaying a map of Los Angeles that had Skid Row marked with our poverty rate map (Fig. 4). While the outline of the city doesn't match up perfectly with our map, this is due to how we binned the data. Larger bins means larger squares on the map, so the curves aren't exact. Similarly, we overlaid the crime rate data over the map of Los Angeles with Skid Row marked (Fig. 5). Our map once again doesn't match up perfectly with the outline of the city, but this is because the data for crime is county-based rather than city-based. The bright spot in this map also is clearly on or near Skid Row. These results are particularly interesting because it connects our findings to known issues in the area. As Skid Row is known to have a high police presence, it's likely that more crimes, no matter how severe, are reported simply because there are more police around to see and report the crimes. Still, this shows that despite high poverty and high crime reporting, especially in an area as specific and well-known as this, poverty has little effect on the severity of crimes.

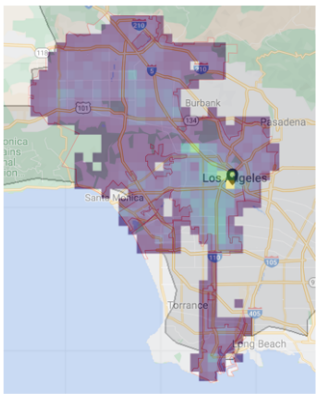

Figure 4: Poverty Rate over LA Map

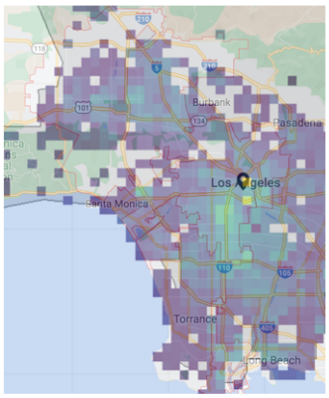

Figure 5: Crime Rate over LA Map

Analyzing our linear regression model we found that it placed the highest weight on the recreation and parks budget with the lowest weight being on the transportation budget when trying to predict the crime rate of Los Angeles City, as seen in the results of Figure 6. There was an R-squared value of 77% meaning the model accounts for 77% of the variance in Crime Rate from our features.

```
Coefficients:
                                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                                     -806.8      440.8  -1.830   0.1646
ECONOMIC.AND.WORKFORCE.DEVELOPMENT.DEPARTMENT  1078.9      274.8   3.927   0.0294
RECREATION.AND.PARKS                           3159.4      817.5   3.865   0.0306
TRANSPORTATION                                 -469.5      279.6  -1.679   0.1917
POVERTY.LEVEL                                  1935.3     1684.5   1.149   0.3339
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 30.02 on 3 degrees of freedom
Multiple R-squared:  0.9053,    Adjusted R-squared:  0.7791
F-statistic: 7.172 on 4 and 3 DF,  p-value: 0.06868
```

Figure 6: AIC Model Results

Discussion

First, with our three heatmaps of Los Angeles, we reinforced an already widely accepted notion: that poverty (or at least places where higher amounts of people live under an acceptable living wage) has a direct relationship with crime frequency. However, we also discovered something lesser-known: that the severity of crime has no relationship with either poverty *or* crime frequency. The results from our heatmaps confirm that although crime rate and poverty rate are related, there is obviously more nuance to their relationship than a simple direct correlation. They also answer our research question for the most part: “crime and related factors” do not seem to affect crime severity at all.

One ethical concern about our results is the harmful feedback loop that’s caused simply from the known relationship between crime and poverty. While this smaller portion of our results proves a known fact, this fact has significant implications. First, a strong relationship between

poverty and crime introduces the idea that impoverished areas will have more crime and thus should be policed more. With more policing comes more reported crimes, since there are more police around to see them. A criminal record makes it more difficult to obtain jobs, so then the poverty rate may further increase, and then the cycle repeats.

Our model tried to predict the crime rate of Los Angeles to see if we could find interesting results with correlations between poverty level and different department budgets for the city. After training we had a semi-accurate model with a 77% R-squared, it showed that the Transportation budget had the highest impact in reducing the crime rate prediction while the Recreations and Park Budget had the highest impact in raising the crime rate prediction (Fig. 6). What this doesn't mean is that increasing the transportation budget will reduce the crime rate in Los Angeles or that increasing the Recreation and Park Budget will increase crime. Finding correlations and trying to predict the crime rate does not suggest causation, and looking at our model in that sense could cause poor decisions in regards to crime rate and those in poverty. However, our model could still be used as a starting point for policymakers. This is why we specifically chose to look at government-related factors to compare poverty level to; these are things that policymakers can directly take action on, but only upon further inspection. Additionally, because the types of data that we used, such as department funding, poverty and crime, are not things unique to Los Angeles, this framework of analysis can be easily transferred to other cities across the country, assuming they also collect the relevant data. It is our hope that this investigation can be repeated in other cities to help local governments analyze the effects and causes of poverty and crime in their area. Finally, it should be noted that our analysis is by no means definitive; this is a very complex issue with many societal and ethical concerns. Our results should be a suggestion to begin more research into these important topics to further see what can be productively done to reduce crime and poverty.

References

[1]

"City Budget and Expenditures | Control Panel LA." https://controllerdata.lacity.org/Budget/City-Budget-and-Expenditures/uyzw-yi8n

[2]

"Crime Data from 2010 to 2019 | Los Angeles - Open Data Portal." https://data.lacity.org/Public-Safety/Crime-Data-from-2010-to-2019/63jg-8b9z

[3]

"Departments & Bureaus | City of Los Angeles." https://www.lacity.org/government/popular-information/departments-bureaus

[4]

"How exactly does AIC penalize overfitting?," *Cross Validated*. https://stats.stackexchange.com/questions/278570/how-exactly-does-aic-penalize-overfitting

[5]

"Incarceration and Poverty in the United States," *AAF*. https://www.americanactionforum.org/research/incarceration-and-poverty-in-the-united-states/

[6]

"Living Wage (2017) | LAC Open Data," *Socrata*. https://data.lacounty.gov/Sustainability/Living-Wage-2017-/f5es-twcr

[7]

S. Yang and G. Berdine, "Model selection and model over-fitting," *The Southwest Respiratory and Critical Care Chronicles*, vol. 3, no. 12, Art. no. 12, Oct. 2015.

[8]

"Percent of Population Below the Poverty Level (5-year estimate) in Los Angeles County, CA (S1701ACS006037) | FRED | St. Louis Fed." https://fred.stlouisfed.org/series/S1701ACS006037 (accessed Dec. 10, 2021).

[9]

P. B. Stretesky, A. M. Schuck, and M. J. Hogan, "Space matters: An analysis of poverty, poverty clustering, and violent crime," *Justice Quarterly*, vol. 21, no. 4, pp. 817–841, Dec. 2004, doi: 10.1080/07418820400096001.