Narcotics Crimes and Mental Health in Chicago: Analysis for the 2018-2022 period

PPHA 30536 1 - Data and Programming for Public Policy II - R Programming

Natalia Zorrilla

2024-12-07

Research Question

The primary research question for this project is: *How do access to mental health resources and socioeconomic factors influence crime rates in Chicago's neighborhoods?*

This project was inspired by an abuse of force incident perpetrated by the San Diego Police Department in 2022 and the City of Chicago Treatment Not Trauma initiative (2024), which aims to reallocate police resources to fund mental health crisis teams led by clinicians and paramedics. The goal is to address mental health emergencies through trauma-informed, holistic care to improve community safety and well-being.

Approach and Coding

The project uses three data sources: Chicago Data Portal, Chicago Health Atlas, and the City of Chicago Mental Health System Expansion Website.

From the Chicago Data Portal I retrieved the *Crime data from 2001 to present, CDPH Mental Health Resources (using API), Boundaries - Neighborhoods Shapefile.* While from the Chicago Health Atlas, I downloaded the *Socioeconomic indicators (Perceived neighborhood safety, Perceived neighborhood safety rate, Trust in law enforcement, Trust in law enforcement rate, Trust in local government, Trust in local government rate, Unemployment rate,Median household income, and Poverty rate)*

Using web scraping I retrieved the *Treatement not Trauma Report* from the City of Chicago Mental Health System Expansion Website. I used this 74 page long document for sentiment analysis.

The approach involved integrating crime and mental health resource data datasets, ensuring consistent spatial formats and identifiers across sources. Likewise, I had to incorporate the socioeconomic indicators to each dataframe that I was working with. This required significant data cleaning and transformation efforts. Most of the time spent on the project was during this part.

Data Cleaning Challenges

Several challenges were encountered during the data cleaning process. I had to handle incomplete or missing data, particularly missing geolocation information. Likewise, I encountered inconsistencies in variable names across data sets and duplicate records. During the data cleaning I had to adress missing values to maintain data accuracy. For example, to facilitate clear visualization and regression analysis, categorical bins were created for socioeconomic indicators. This included creating brackets for income levels, trust levels, and crime ratios based on population.

Initial Project Challenges

In an ideal scenario, I wanted to conduct a Difference-in-Differences (DID) research design to evaluate the impact of the mental health expansion proposal (Treatment Not Trauma report) on crime rates in Chicago. However, several challenges emerged during the data exploration and visualization process caused a shift in approach.

Firstly, the mental health expansion proposal is very recent, dating from May 2024, which means that an analysis of its effects would only be possible in the future. Secondly, when accessing socioeconomic indicators, I found that most data was representative of the period from 2018 to 2022.

Given these constraints, the focus of the study was narrowed to this specific time frame. The regression analysis was limited to narcotics-related crimes, as plotting the top 10 crimes in Chicago revealed that not all crime types might be directly influenced by mental health resources. Additionally, the available mental health resources dataset lacked information on the opening dates of clinics, further complicating a longitudinal analysis. These factors collectively led to a more focused, cross-sectional approach examining the relationship between narcotics crime rates, mental health resources, and socioeconomic factors in Chicago neighborhoods.

Results

The analysis of narcotics crime patterns in Chicago from 2018-2022 reveals several relationships across different socioeconomic indicators. The Neighborhood Safety Rate plot shows a striking peak of narcotics crimes (approximately 22,000) in areas where residents report safety rates between 40-50%, with a consistent decline in crime counts as perceived neighborhood safety increases. Areas with the highest safety rates (90-100%) demonstrate minimal narcotics crime activity.

When examining income distribution, the data indicates that middle-income neighborhoods ($40K-60K) experience the highest concentration of narcotics crimes, while both lower-income (<40K) and higher-income brackets (>80K-100K) show notably lower crime rates.

The unemployment pattern indicates that neighborhoods with 15-20% unemployment rates face the highest narcotics crime counts (around 20,000), with significantly lower rates in areas of very low (<5%) or very high (>30%) unemployment.

Perhaps most surprisingly, the trust in law enforcement data suggests a strong inverse relationship with narcotics crimes - areas with 40-50% trust levels show the highest crime counts (nearly 30,000), followed by a dramatic decrease as trust levels increase, with minimal criminal activity reported in neighborhoods where trust in law enforcement exceeds 80%.

Furthermore, the heat maps that I created, reveal contrasting spatial patterns across Chicago. The narcotics crime map shows high concentrations (4,000-6,000 crimes) in the center-west region, while most other areas have lower rates. The mental health crisis providers map indicates limited resource distribution, with most neighborhoods having 0-1 providers and only a few northern areas reaching 4 providers. This spatial mismatch between crime rates and mental health resources suggests potential gaps in service allocation across Chicago's neighborhoods.

However, the regression analysis revealed several interesting findings across the three models I performed:

Model 1: Narcotics Crime Rate and Total Providers

- There is a statistically significant positive relationship between the number of total mental health providers and narcotics crime rates (coefficient = 4.054, p < 0.1).
- However, the model explains only 5.8% of the variance in narcotics crime rates ($R^2 = 0.058$), indicating that other factors play a substantial role.

Model 2: Incorporating Crisis Providers and Population

- The relationship between total providers and narcotics crime rates remains significant and positive (coefficient = 5.830, $p < 0.05$).
- Population shows a significant negative relationship with narcotics crime rates (coefficient = -0.0005, $p < 0.05$).
- Crisis providers do not show a significant relationship.
- The model's explanatory power increases to 12.6% ($R^2 = 0.126$).
- Variance Inflation Factors (VIF) are all below 2, suggesting no serious multicollinearity issues.

Model 3: Controling for Socioeconomic Indicators

- In this more complex model, none of the individual variables show statistical significance at the conventional levels.
- The model's explanatory power increases to 21.9% ($R^2 = 0.219$), but the adjusted R^2 (0.105) suggests potential overfitting.
- VIF analysis reveals moderate multicollinearity, with Neighborhood Safety Rate and Median Household Income having VIF values above 4.

These results suggest a complex relationship between narcotics crime rates, mental health resources, and socioeconomic factors. While the simpler models indicate a positive relationship between mental health providers and narcotics crime rates, this relationship becomes less clear when additional factors are considered.

Based on the results previously mentioned, the project faces some limitations:

1. The counterintuitive positive relationship between providers and crime rates might indicate reverse causality or the presence of confounding variables.

2. The role of mental health resources in crime prevention requires further investigation, possibly with more nuanced data on the timing and nature of interventions.

3. The moderate multicollinearity in the model 3 suggests that some socioeconomic factors may be interrelated, complicating interpretation.

Finally, the sentiment analysis of the Treatment Not Trauma report reveals an interesting disconnect between the proposal's focus and its content. While the report predominantly shows positive sentiment across all three lexicons (NRC showing high positive and trust scores, Bing indicating twice as many positive as negative terms, and AFINN clustering around positive scores of +2), it notably lacks discussion of crime despite proposing reallocation of CPD resources. The high counts of "anticipation" and "trust" in the NRC analysis, combined with the overwhelmingly positive tone, suggest the report focuses on constructive healthcare solutions rather than addressing crime concerns directly. This observation underscores a potential gap in the proposal's framework - while it advocates for shifting resources from law enforcement to mental health services, it doesn't explicitly address how this might impact crime rates or public safety.

Future Research Directions

To build upon this research, future studies could incorporate longitudinal data to better understand causal relationships. Additionally, one could explore non-linear relationships between variables. Like I mentioned, ideally one could use more advanced statistical techniques, such as instrumental variables or matching methods, to address potential endogeneity issues. Finally, collecting more detailed data on the timing and nature of mental health interventions will help to better isolate their effects.