

# Relationship between housing rents and crime rates

Cheng Wu

Geraint Gui

Stephen Xiao

2024-10-19

## Project Objective

The objective of this project is to investigate the relationship between crime rates and housing rents in various regions of a city over time. Specifically, the research seeks to answer the question:

**“Do crime rates in different city areas influence housing rents?”**

This study is motivated by the idea that crime rates may impact housing demand, and consequently, housing prices. Higher crime rates could make certain areas less desirable, potentially lowering housing rents, while lower crime rates might contribute to higher demand and elevated rents.

By examining crime data alongside housing rent data across multiple years and locations, this analysis aims to uncover patterns that could help city planners, policymakers, real estate investors, and residents better understand the dynamics between urban crime and housing markets. Ultimately, insights derived from this research could contribute to more informed decision-making in urban planning and real estate investments, as well as enhance the understanding of socio-economic influences on housing.

## 2. Research Questions

The primary goal of this study is to determine whether crime rates in different city areas influence housing rents. To comprehensively address this, the analysis will explore the relationship between crime and housing rents over time, examining variations by geographic location and time period. The research is structured around the following key questions:

### 2.1 Primary Research Question

- **Do crime rates in different areas of a city impact housing rents over time?**
  - This question seeks to understand if there is a general relationship between crime rates and housing rents, assessing whether areas with higher crime experience lower housing rents, and vice versa.

### 2.2 Sub-Questions

To answer the primary question effectively, several sub-questions are considered. These sub-questions enable a more nuanced analysis by breaking down the main question into specific aspects of time, geography, and type of crime:

#### 1. Temporal Impact on Housing Rents:

- **How do housing rents vary with crime rates over time within the same area?**
  - This sub-question examines if changes in crime rates in a specific area correlate with changes in housing rents over months and years. It aims to identify whether a decrease in crime, for instance, might correspond with an increase in housing rents over the same period, or if there is a lagged effect.

#### 2. Geographic Differences in Crime and Housing Rent Relationship:

- **Are there specific areas (zip codes) where the relationship between crime rates and housing rents is stronger or weaker?**

- This question focuses on geographic variability, investigating whether certain areas are more sensitive to crime rate changes in terms of housing rent fluctuations. It will look at regional differences across zip codes to determine if some areas experience a stronger correlation between crime rates and housing rents than others.
3. **Type of Crime and Its Differential Impact on Housing Rents:**
    - **Do specific types of crime (e.g., violent crime vs. property crime) have a different effect on housing rents?**
      - Crime data often includes various types of crime, from domestic violence to theft and assault. This sub-question will analyze whether certain types of crime have a stronger influence on housing rents, offering insight into which crime categories are most impactful for housing market dynamics.
  4. **Crime Rate Thresholds and Housing Rent Sensitivity:**
    - **Is there a threshold level of crime rate that significantly impacts housing rents?**
      - This question investigates whether there is a tipping point in crime rates beyond which housing rents are notably affected. For example, do rents only decline sharply when crime rates exceed a certain level, or is the effect on housing rents more gradual?
  5. **Lagged Effects of Crime Rates on Housing Rents:**
    - **Do changes in crime rates in one period (e.g., previous month or year) predict changes in housing rents in subsequent periods?**
      - This question examines potential lagged effects, investigating whether changes in crime rates in one period influence housing rents in subsequent months or years. This analysis will help understand if there are delayed impacts of crime rate changes on housing rents, suggesting that renters and landlords may respond to crime trends with a time lag.

## 2.3 Hypotheses

Based on the above questions, the study will test the following hypotheses:

- **Hypothesis 1:** Higher crime rates are associated with lower housing rents within the same area and time period.
- **Hypothesis 2:** The relationship between crime rates and housing rents varies by area, with some areas being more sensitive to crime rate fluctuations.
- **Hypothesis 3:** Different types of crime have varying levels of impact on housing rents, with violent crime likely having a stronger negative effect than non-violent crime.
- **Hypothesis 4:** There exists a threshold in crime rate that, once crossed, significantly lowers housing rents in an area.
- **Hypothesis 5:** Changes in crime rates have a lagged effect on housing rents, with previous-period crime rates impacting housing rents in future periods.

## Summary

By breaking down the primary research question into these specific sub-questions, the analysis will provide a comprehensive examination of how crime rates may influence housing rents across different times, locations, and types of crime. These research questions and hypotheses guide the data processing, analysis, and visualization steps, ensuring a structured approach to uncover potential correlations or causations between crime and housing market trends.

## 3. Team

- **Cheng Wu:** Responsible for data collection, cleaning, and preprocessing. Also leading the report writing and final presentation preparation.
- **Geraint Gui:** Assisting with data analysis, initial exploration of trends, and involved in data collection. Additionally, creating visualizations and summarizing findings.
- **Stephen Xiao:** Supporting the statistical modeling and hypothesis testing. Also helping with proofreading and presentation edits.

## 4. Datasets

In this study, two primary datasets were utilized to explore the relationship between crime rates and housing rents across various regions of Chicago over time. These datasets were carefully sourced, processed for consistency, and integrated to enable comprehensive analysis. This section describes the sources, access methods, and preprocessing steps for each dataset, followed by the integration process.

### 4.1 Dataset Sources

#### 1. Chicago Crime Data (2001 - 2024)

- **Source:** The crime data was obtained through the City of Chicago Data Portal, which provides open-access data on reported criminal incidents in Chicago. This data is regularly updated with information from the Chicago Police Department.
- **Access Method:** The dataset was accessed via the City of Chicago’s API, allowing for automated retrieval of crime data. Using API queries, data from 2001 to 2024 was downloaded in JSON format and subsequently converted to CSV for further processing.
- **Data Fields:** Key fields in the crime dataset included:
  - **ID, Case Number:** Unique identifiers for each crime incident.
  - **Date:** Timestamp of the incident.
  - **Primary Type, Description:** Categories and descriptions of crime types (e.g., theft, assault).
  - **Arrest:** Indicator of whether an arrest was made (boolean).
  - **Domestic:** Indicator of domestic violence-related incidents (boolean).
  - **Location Description, Latitude, Longitude:** Geospatial coordinates and location details.
  - **Community Area:** Code representing different community areas in Chicago.
- **Data Coverage:** This dataset covers all reported crimes from 2001 to present, providing a comprehensive look at crime trends over time.

#### 2. Zillow Housing Rent Data (2000 - 2024)

- **Source:** Housing rent data was sourced from Zillow Research Data, which provides housing metrics across the U.S., including rental prices at various geographic levels.
- **Access Method:** The housing rent dataset was downloaded directly from Zillow’s website as a CSV file. It contains time-series data on rental prices, specifically at the zipcode level for Chicago, covering data from 2000 to 2024.
- **Data Fields:** The main fields in this dataset included:
  - **RegionID:** Unique identifier for each geographic area.
  - **SizeRank:** Rank of the region based on population size (not used in final analysis).
  - **RegionName:** Zip code for each region in Chicago.
  - **City, CountyName:** Identifiers for the city and county.
  - **Monthly rent columns (2000-01-31 to 2024-09-30):** Each column represents the average rent for a specific month, structured in a wide format.

### 4.2 Data Preprocessing and Cleaning

To ensure compatibility and analytical usability, both datasets required extensive preprocessing. The steps taken to clean and transform each dataset are outlined below.

#### Chicago Crime Data Preprocessing

1. **Handling Missing Values:** Records with missing values in crucial columns (such as **Location**, **Latitude**, or **Longitude**) were removed to maintain accuracy in geospatial analysis. Missing values in fields like **Community Area** were also handled, ensuring that each entry could be mapped accurately.
2. **Date Formatting:** The **Date** column was originally in a timestamp format (e.g., 05/07/2020 10:24:00 AM). This field was standardized to a date format (YYYY-MM-DD), and additional columns for **Year** and **Month** were extracted to support time-based aggregation.

3. **Zipcode Mapping:** Since the crime data was provided at the latitude and longitude level, zipcodes were appended to each record using spatial mapping techniques. Geocoding via a third-party API was utilized to map latitude and longitude to zipcodes, creating a `ZCTA5CE20` column to align with the housing data's zipcode-level granularity.
4. **Data Aggregation:** The crime data was aggregated by `Year`, `Month`, and `ZCTA5CE20` (zipcode). Monthly crime metrics were calculated for each zipcode, including:
  - `crime_count`: Total number of crimes reported.
  - `arrest_count`: Number of arrests made.
  - `domestic_count`: Number of domestic-related incidents.

This preprocessing resulted in a structured dataset with monthly crime statistics by zipcode, facilitating temporal and geographic analysis.

### Zillow Housing Rent Data Preprocessing

1. **Wide-to-Long Transformation:** The housing rent data was originally in a wide format, with each month represented as a separate column. To align with the crime data format, the dataset was transformed to a long format using the `pivot_longer()` function in R, resulting in columns for `RegionName` (zipcode), `Year`, `Month`, and `price` (monthly rent).
2. **Date Parsing:** The date columns (e.g., 2000-01-31) were parsed to extract `Year` and `Month`, enabling synchronization with the crime data's temporal structure.
3. **Column Selection:** Only essential columns, such as `RegionName` (zipcode), `City`, and `CountyName`, were retained. Non-essential fields, like `SizeRank`, were discarded to streamline the dataset.
4. **Handling Missing Rent Data:** Any missing values in the rent columns were either imputed or removed to ensure consistent time series data across all zipcodes. This preprocessing step ensured that each zipcode had a continuous monthly rent record, supporting time-series analysis.

### 4.3 Data Integration

After preprocessing, the crime and housing rent datasets were integrated to create a unified dataset for analysis. The steps taken in the integration process are as follows:

1. **Column Renaming:** The `RegionName` column in the Zillow housing data was renamed to `ZCTA5CE20` to match the zipcode identifier used in the crime data.
2. **Merging Datasets:** Using R's `left_join()` function, the crime and housing datasets were merged on `ZCTA5CE20` (zipcode), `Year`, and `Month`. This join operation ensured that each row represented a unique combination of zipcode, year, and month, with corresponding data on housing rents and crime metrics.
3. **Final Data Structure:** The resulting dataset includes the following columns:
  - `ZCTA5CE20`: Zipcode identifier.
  - `Year` and `Month`: Time identifiers for monthly analysis.
  - `price`: Monthly average housing rent.
  - `crime_count`: Total number of crimes recorded for that month and zipcode.
  - `arrest_count`: Number of arrests made.
  - `domestic_count`: Incidents of domestic violence.
  - Additional identifiers like `City` and `CountyName` to facilitate regional analysis.

### 4.4 Summary of Preprocessed Data

The final integrated dataset provides a comprehensive monthly record of housing rents and crime statistics across different Chicago regions. By transforming and merging the data into a consistent format, this dataset enables analysis to address the research questions of whether and how crime rates in different areas

influence housing rents over time. This structured dataset is now ready for advanced statistical analysis and visualization to reveal insights into the dynamics between urban crime and housing markets.

## 5. Timeline

### Week 1-2: Data Collection and Preparation

- **Task:** Collect crime data from the Chicago Data Portal and housing price data from Zillow's Home Value Index (ZHVI).
- **Team Member:** Cheng Wu
- **Details:** Download datasets for both crime reports and housing prices. Ensure the data covers a consistent time frame (2000-2024) and aligns by geographic location (e.g., zip codes in Chicago). Set up initial data checks and confirm data completeness.

### Week 3: Data Cleaning and Preprocessing

- **Task:** Clean and preprocess both datasets.
- **Team Member:** Cheng Wu, Geraint Gui
- **Details:** Handle missing values, format date columns to extract **Year** and **Month**, and ensure consistent geographic identifiers (e.g., zip codes). Transform housing data into long format and aggregate crime data by month and zip code. Merge the datasets based on **Year**, **Month**, and **zipcode** for a unified view.

### Week 4-5: Exploratory Data Analysis

- **Task:** Conduct an initial analysis to understand trends in crime rates and housing rents.
- **Team Member:** Cheng Wu, Geraint Gui
- **Details:** Create visualizations, such as time-series plots and heatmaps, to inspect the distribution of crime rates and housing prices across zip codes and over time. Identify any initial patterns or correlations between crime rates and rent levels.

### Week 6: Statistical Analysis and Modeling

- **Task:** Apply statistical analysis and modeling techniques to evaluate the relationship between crime rates and housing rents.
- **Team Member:** Cheng Wu, Stephen Xiao
- **Details:** Perform correlation analysis and develop regression models to investigate how changes in crime rates affect housing rents. Analyze lagged effects, seasonal variations, and potential thresholds. Test hypotheses regarding different crime types and their specific impact on housing rents.

### Week 7: Final Report and Presentation Preparation

- **Task:** Compile findings into the final report and prepare a presentation.
- **Team Member:** Cheng Wu, Stephen Xiao
- **Details:** Draft the report, summarizing the study's key insights, statistical findings, and visualizations. Prepare a presentation that outlines the research process, analysis methods, and main conclusions, with visual aids and key takeaways.

### Week 8: Project Review and Submission

- **Task:** Final review and submission of the project.
- **Team Member:** All members
- **Details:** Conduct a final review of all deliverables, ensuring completeness and accuracy. Submit the project report, code, and any required documentation or metadata for reproducibility.

## Visualizations

In this section, we explore the dataset, `Combined_Housing_Crime_Data.csv`, by creating a time-series plot and a heatmap to uncover trends and relationships in the data. These visualizations will help in understanding key patterns in housing and crime data over time and across regions in Chicago, aligning with the research questions around how housing and crime may correlate over time and by location.

### Time-Series Plot

The time-series plot displays trends in [specify variable, e.g., “housing prices” or “crime rate”] over time. This helps us identify any seasonal patterns, long-term trends, or fluctuations, which are crucial for understanding the impact of various factors on housing and crime rates. By examining these trends, we aim to answer questions such as:

- How do housing prices or crime rates fluctuate over time?
- Are there any discernible patterns that could indicate external factors influencing both housing and crime data?

**R Code for Time-Series Plot** The code below reads in the data, processes it to extract time-series data for selected variables, and then plots a line graph showing trends over time.

```
library(ggplot2)
library(dplyr)

##
## Attaching package: 'dplyr'

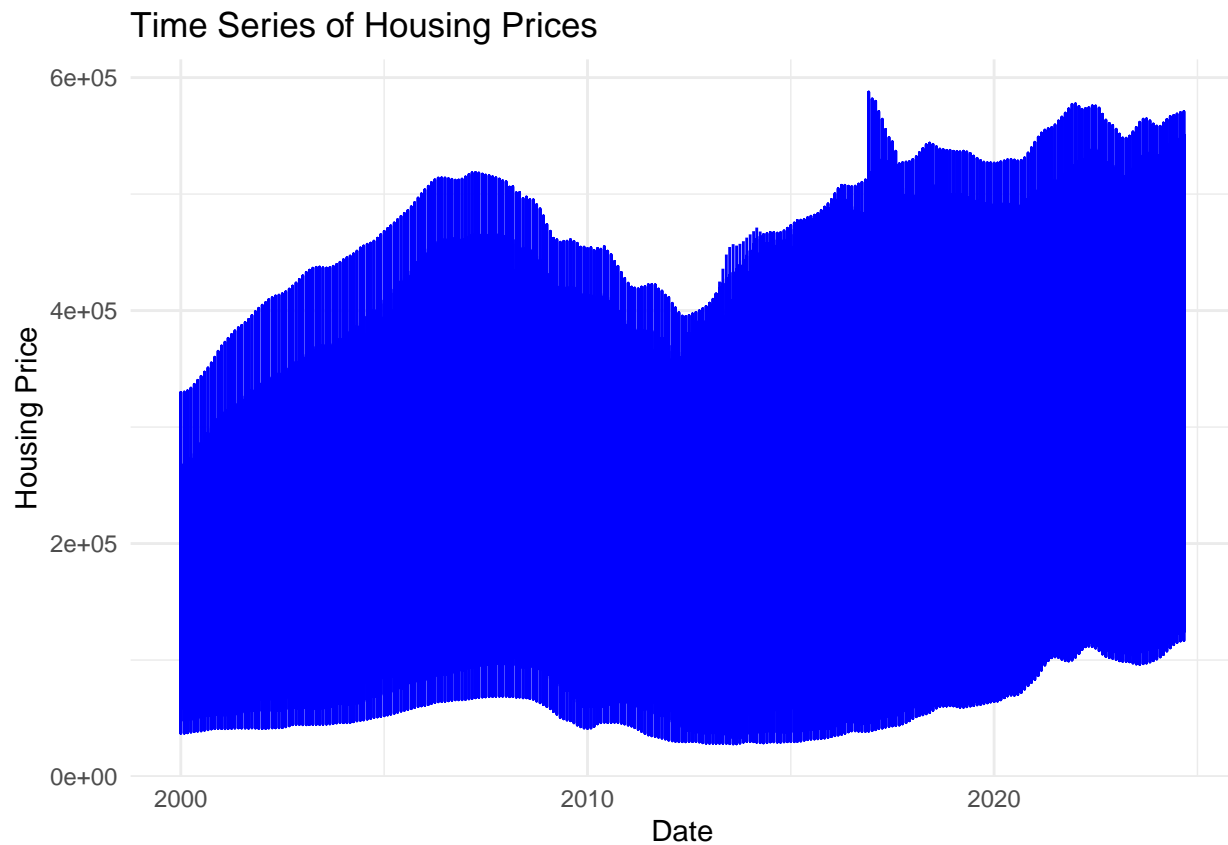
## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

data <- read.csv("Combined_Housing_Crime_Data.csv")

data$date <- as.Date(paste(data$Year, data$Month, "01", sep = "-"), format = "%Y-%m-%d")

ggplot(data, aes(x = date, y = price)) +
  geom_line(color = "blue", na.rm = TRUE) +
  labs(title = "Time Series of Housing Prices",
       x = "Date",
       y = "Housing Price") +
  theme_minimal()
```

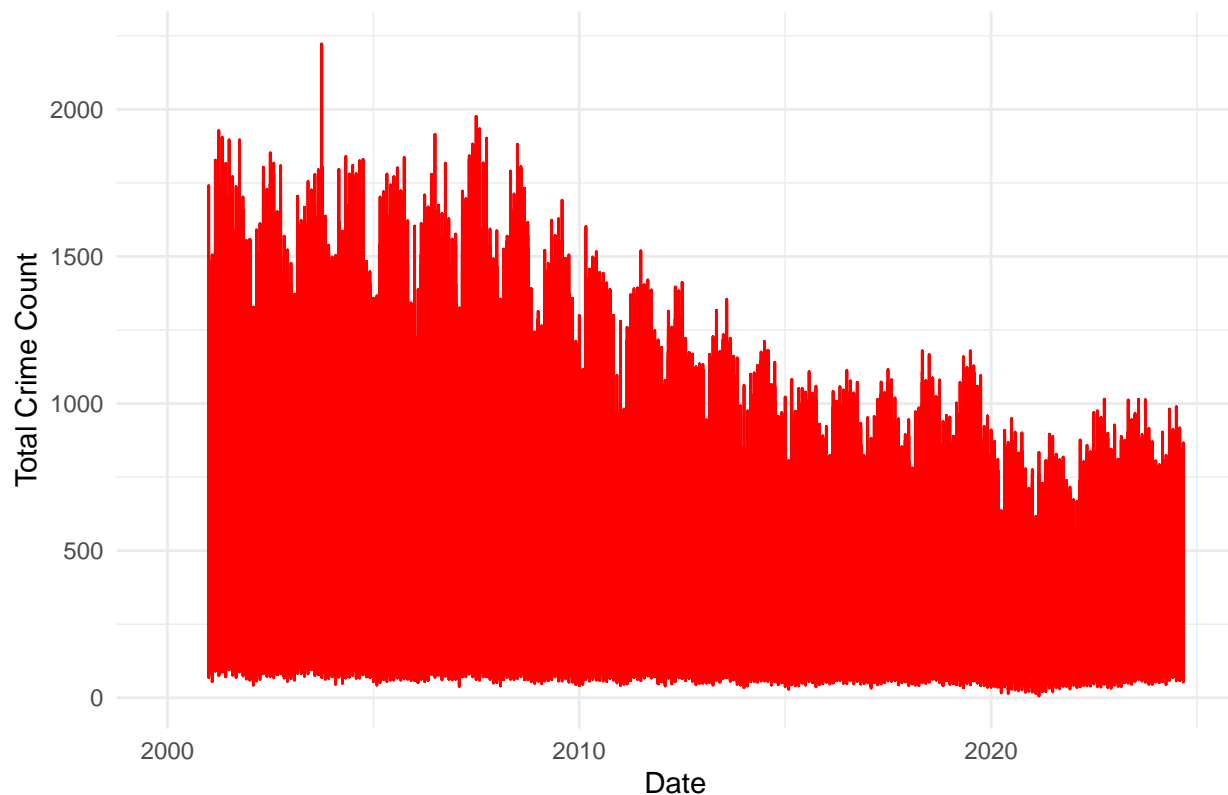


### Time Series of Total Crime Count

```
# Plot for Total Crime Count
ggplot(data, aes(x = date, y = crime_count)) +
  geom_line(color = "red", na.rm = TRUE) +
  labs(title = "Time Series of Total Crime Count",
       x = "Date",
       y = "Total Crime Count") +
  theme_minimal()
```

The time-series plot of total crime count shows the overall trend of crime incidents in Chicago over time. This plot is crucial for examining how the total volume of crimes has changed, potentially influencing or being influenced by housing market trends.

## Time Series of Total Crime Count



### Interpretation ### The plot indicates a general decline in total crime counts over the years, with some peaks that may correspond to specific events or policy changes. This decline may reflect efforts in crime reduction, which could impact public perceptions and housing demand in affected areas.

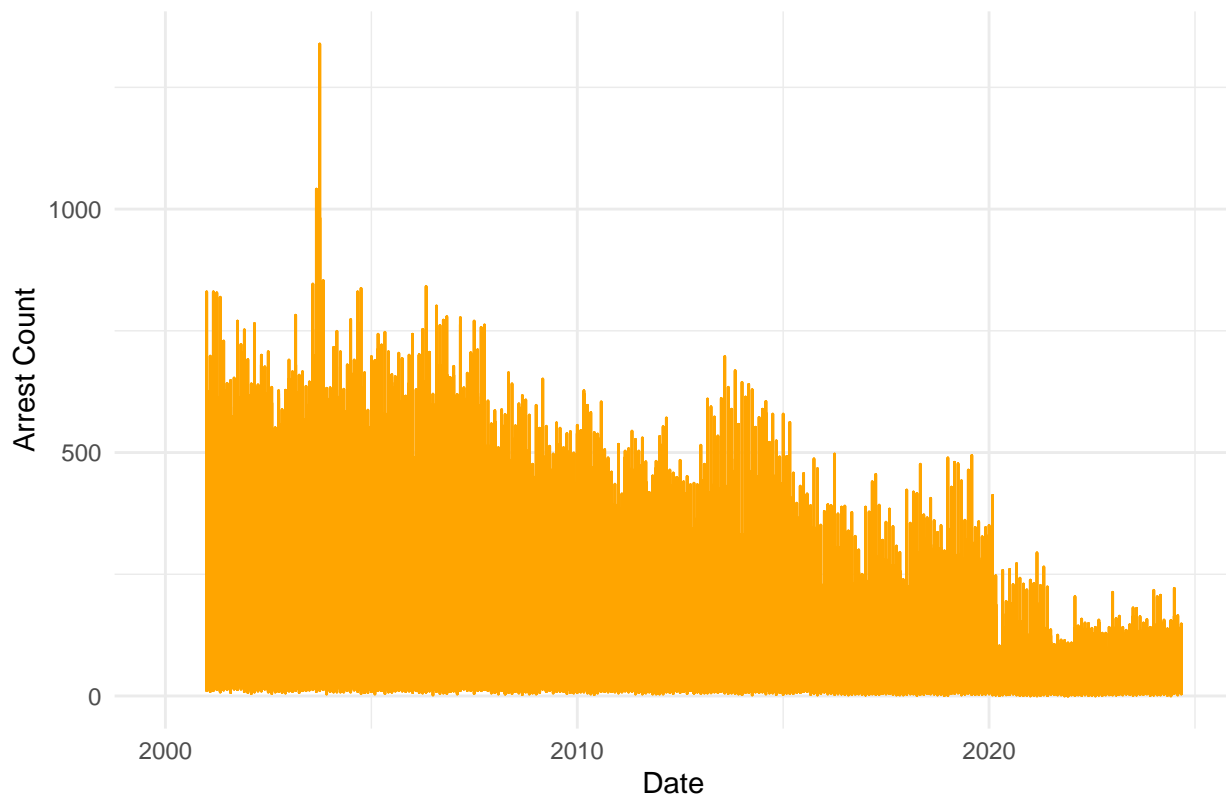
## Time Series of Arrest Counts

```
# Plot for Arrest Count
ggplot(data, aes(x = date, y = arrest_count)) +
  geom_line(color = "Orange", na.rm = TRUE) +
  labs(title = "Time Series of Arrest Count",
        x = "Date",
        y = "Arrest Count") +
  theme_minimal()
```

The arrest count time-series plot provides insights into law enforcement activity and its possible impact on housing and crime. Higher arrest rates may indicate a focus on law enforcement in certain periods, which could affect perceptions of safety and influence housing market trends.



## Time Series of Arrest Count



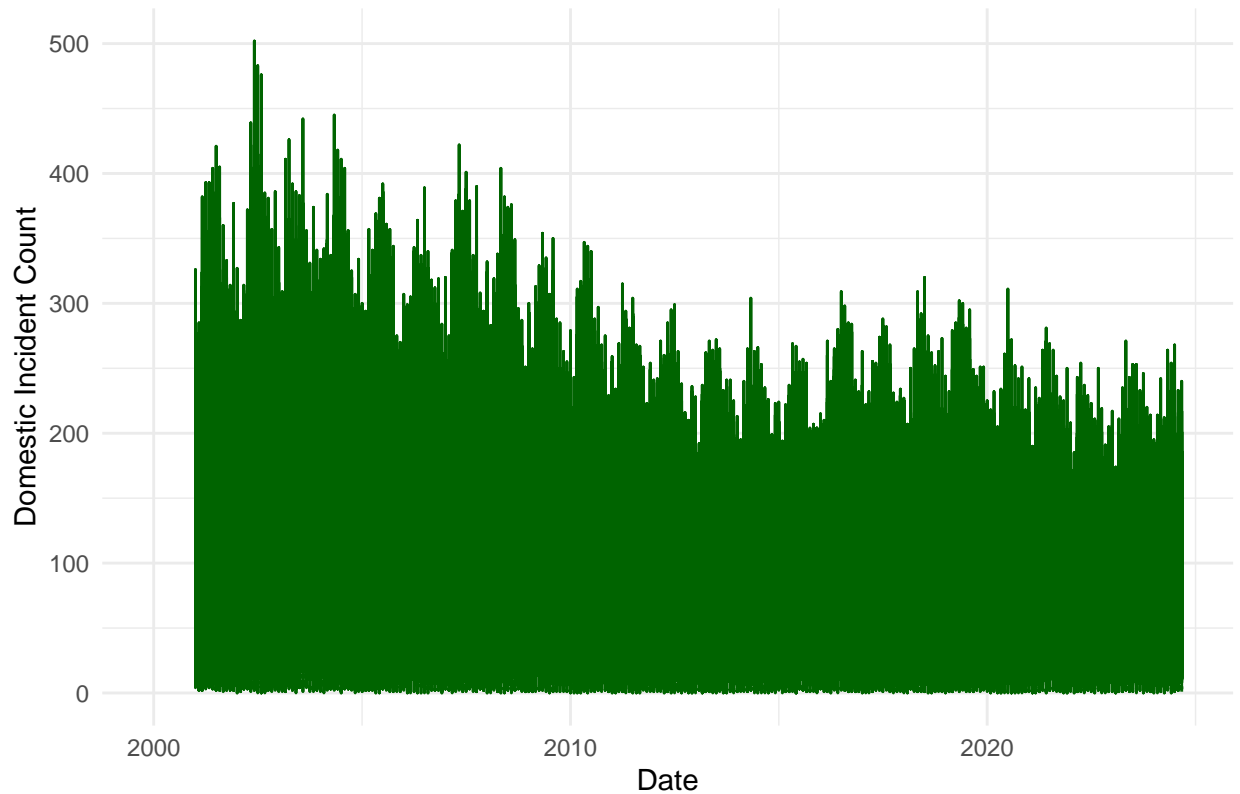
### Interpretation ### We observe fluctuations in arrest counts, with some decline over time. Changes in arrest counts may correlate with changes in crime rates or reflect shifts in policing strategies. Understanding arrest patterns helps in analyzing public safety perceptions and their effect on housing prices.

## Time Series of Domestic-Related Incidents

```
# Plot for Domestic Count
ggplot(data, aes(x = date, y = domestic_count)) +
  geom_line(color = "darkgreen", na.rm = TRUE) +
  labs(title = "Time Series of Domestic-Related Incidents",
       x = "Date",
       y = "Domestic Incident Count") +
  theme_minimal()
```

The plot of domestic-related incidents provides a focused look at specific crime types, offering insight into trends that might affect neighborhood desirability and housing values. Domestic incidents may directly impact perceptions of community safety, which in turn can influence housing

## Time Series of Domestic-Related Incidents



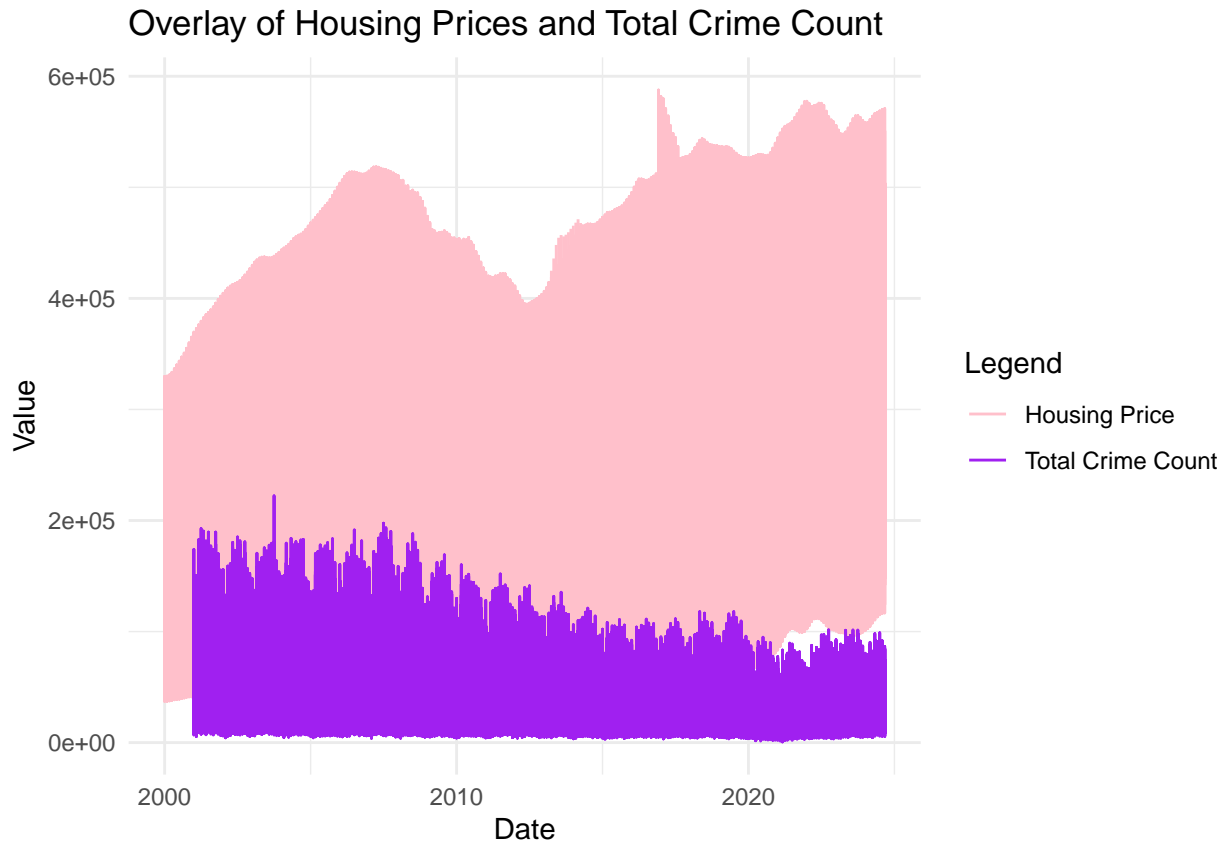
demand.

### Interpretation ### The time-series of domestic-related incidents indicates a steady trend with occasional spikes. Analyzing these patterns can help understand whether particular types of crimes have a more significant influence on housing prices than others.

```
# Define a scale factor for crime count to bring it to a similar scale as housing price
scale_factor <- 100
```

```
# Overlay Plot of Housing Price and Total Crime Count
```

```
ggplot(data, aes(x = date)) +
  geom_line(aes(y = price, color = "Housing Price"), na.rm = TRUE) +
  geom_line(aes(y = crime_count * scale_factor, color = "Total Crime Count"), na.rm = TRUE) +
  labs(title = "Overlay of Housing Prices and Total Crime Count",
       x = "Date",
       y = "Value") +
  scale_color_manual(name = "Legend", values = c("Housing Price" = "pink", "Total Crime Count" = "purple")) +
  theme_minimal()
```



### Interpretation of the Overlay Plot: Housing Prices and Total Crime Count

The overlay plot of housing prices and total crime count offers preliminary insights that align with our research objective: to understand the relationship between crime rates and housing rents in various city regions over time.

#### Observations

##### 1. Trend in Housing Prices:

- The housing prices (represented by the pink line) show a general upward trend with some fluctuation, reflecting the value growth in the housing market over the period studied. Key dips, such as those potentially associated with economic downturns (e.g., the 2008 financial crisis), suggest that housing prices are sensitive to broader economic conditions, but also demonstrate resilience and recovery.

##### 2. Trend in Total Crime Count:

- The total crime count (represented by the purple line) appears to be stable or slightly declining. While there are occasional spikes, the overall trend does not show a substantial increase, implying either stability or gradual improvement in crime levels.

##### 3. Relationship Between Housing Prices and Crime Rates:

- This overlay suggests an inverse trend, where rising housing prices correspond with stable or slightly declining crime rates. This observation aligns with the hypothesis that lower crime rates could make certain areas more attractive, thereby supporting higher housing prices. Conversely, stable or increasing crime rates may contribute to reduced demand in some regions, potentially influencing housing rents negatively.

## Relevance to Research Questions

The insights gained from this plot are relevant to our primary research question: **“Do crime rates in different areas of a city impact housing rents over time?”** The observed inverse trend between crime rates and housing prices suggests that areas with higher crime may experience suppressed demand, possibly leading to lower housing rents. Conversely, areas with stable or lower crime rates may attract more residents, driving up housing rents due to increased demand.

This finding supports the motivation behind our study, which seeks to understand how socio-economic factors like crime influence urban housing markets. By examining these temporal relationships, this analysis can help identify patterns that city planners, policymakers, real estate investors, and residents can use to make more informed decisions regarding urban development, rental pricing, and community safety investments.

## Next Steps in Analysis

To answer our research questions more thoroughly, the following steps are recommended:

### 1. Regional Analysis:

- Since our research question focuses on different city areas, it is important to disaggregate the data by geographic location. Analyzing crime rates and housing rents by neighborhood or zip code can reveal localized trends and identify regions where crime rates most strongly influence housing rents.

### 2. Correlation and Causation Analysis:

- Conduct statistical tests to determine the strength of the correlation between crime rates and housing rents. Additionally, exploring causal relationships through time-lag analysis can help assess whether changes in crime rates precede shifts in housing rents in specific areas.

### 3. Analysis of Crime Types:

- Different types of crime (e.g., violent crime, property crime) may have varying impacts on housing demand and rents. By examining these crime categories separately, we can identify if certain types of crime are more closely related to housing market dynamics.

### 4. Incorporate Housing Rent Data:

- Since the current analysis focuses on housing prices, including housing rent data specifically will provide a direct answer to our research question. This could allow for more precise insights into how crime impacts rental demand versus home purchase demand.

## Conclusion

The overlay plot of housing prices and total crime count provides an initial view of the potential relationship between crime and housing rents. By further analyzing this relationship across different regions and over time, this study aims to uncover patterns that can guide urban planning, real estate investments, and housing policy decisions, ultimately enhancing our understanding of the socio-economic dynamics between crime and housing markets in urban environments.