Title

Crime Diversity Beyond Crime Counts: A Spatial Analysis of LAPD Reporting Districts in Los Angeles

ISBN
9798190937435

Author
Yang, Wanrun

Publication Date
2026-09-09

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Crime Diversity Beyond Crime Counts: A Spatial Analysis of LAPD Reporting Districts in Los Angeles

A thesis submitted in partial satisfaction

of the requirements for the degree

Master of Applied Statistics and Data Science

by

Wanrun Yang

2026

ABSTRACT OF THE THESIS

Crime Diversity Beyond Crime Counts:

A Spatial Analysis of LAPD Reporting Districts in Los Angeles

by

Wanrun Yang

Master of Applied Statistics and Data Science

University of California, Los Angeles, 2026

Professor Xiaowu Dai, Chair

Traditional crime measures, such as crime counts and rates, effectively describe the overall volume of crime. However, they provide limited information about the composition of crime within local areas. This thesis examines whether a Crime Diversity Index (CDI) can add useful information about local crime patterns beyond conventional measures of crime volume. The analysis is based on 875,087 crime incidents reported by the Los Angeles Police Department (LAPD) between January 2020 and December 2023. These incidents are grouped into eight broad crime categories, and crime diversity is measured using a normalized Shannon Diversity Index. The main spatial unit used in the study is the LAPD Reporting District.

The study examines crime diversity from spatial, temporal, and spatio-temporal perspectives. Spatial variation is evaluated using Reporting District-level comparisons, choropleth maps, and Global Moran's I. Temporal patterns are examined across months and seasons using descriptive statistics, one-way ANOVA, and a random-intercept mixed-effects sensitivity analysis. The relationship between crime volume and diversity is assessed using correlation and ordinary least squares regression, together with fixed-count repeated subsampling and spatial error regression to evaluate the effects of unequal incident counts and residual spatial dependence. Spatio-temporal dynamics are further examined across eight consecutive half-year periods.

The results show substantial variation in crime diversity across Reporting Districts and statistically significant positive spatial autocorrelation. In the unadjusted data, crime diversity is strongly and positively associated with crime volume. However, when exactly 100 incidents are repeatedly sampled from each eligible Reporting District, the association between the count-adjusted CDI and original crime volume becomes negligible, while the relative CDI pattern across districts remains highly stable. Spatial error modeling further shows that residual spatial dependence is important in the count–diversity relationship. Temporal variation is more modest: monthly diversity remains relatively stable, and although seasonal differences are statistically detectable, their magnitude remains small after accounting for repeated observations within Reporting Districts. The half-year analysis shows that the broad spatial pattern of crime diversity persists over time while local changes remain visible.

Overall, the findings indicate that crime diversity should not be interpreted independently of crime volume, particularly in low-volume areas. Nevertheless, the CDI captures a compositional dimension of local crime that is not reducible to total incident frequency. The results support using crime diversity as a complementary rather than competing measure of local crime patterns and provide a reproducible framework for jointly examining crime volume, crime composition, and their spatial and temporal variation in Los Angeles.

The thesis of Wanrun Yang is approved.

Guang Cheng

George Michailidis

Frederic Schoenberg

Xiaowu Dai, Committee Chair

University of California, Los Angeles

2026

To my family,

whose unwavering love and support carried me through

every challenge, especially during my medical treatment.

This accomplishment would not have been possible

without your encouragement, patience, and belief in me.

With my deepest love and gratitude.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank my advisor and committee members for their guidance and support throughout this project.

CHAPTER 1

Introduction

In March 2025, the Los Angeles Police Department (LAPD) released its 2024 year-end crime statistics, reporting notable declines in several major crime categories across the city. Homicides decreased by 14%, shooting victims decreased by 19%, and property crime also declined compared with the previous year [Off24]. These statistics provide an important summary of recent changes in public safety in Los Angeles. However, they primarily describe how much crime occurred and provide less information about the composition of crime across different parts of the city. Two areas with similar crime counts may experience very different combinations of violent crime, property crime, financial crime, or public order offenses. As increasingly detailed crime data become publicly available, it is possible to examine local crime patterns from perspectives beyond traditional measures of crime volume.

Crime counts and crime rates have long been widely used measures in criminology. They provide simple and intuitive ways to compare crime levels across geographic areas and monitor changes over time. Despite their usefulness, these measures mainly describe the quantity of crime. When different offenses are combined into a single measure, differences in the types and relative composition of crime may be overlooked. For example, one area may be dominated by vehicle-related crime, while another with a similar total number of incidents may contain a more balanced mixture of property, violent, financial, and other offenses. Although their overall crime volumes may be similar, the local crime patterns and possible policing concerns can be different. Therefore, crime volume alone may not provide a complete description of crime within a local area.

Crime diversity provides a complementary way to describe this aspect of crime. Rather than focusing only on how many incidents occur, crime diversity considers how those incidents are distributed across different crime categories. Higher diversity indicates that reported crimes are distributed relatively evenly among several categories, whereas lower diversity indicates that crime is more concentrated in one or a few categories. Previous studies have introduced diversity measures into crime research and have shown that crime diversity represents a dimension of offending that is related to, but not identical to, crime frequency. However, compared with traditional crime counts and rates, the empirical application of crime diversity remains relatively limited, particularly when spatial and temporal variation are considered together.

This study uses the Shannon Diversity Index to construct a normalized Crime Diversity Index (CDI). The original LAPD data contain many detailed offense descriptions, which are consolidated into eight broader categories: Financial Crime, Other Crime, Property Crime, Public Order Crime, Sexual Crime, Vehicle-related Crime, Violent Crime, and Weapon-related Crime. The normalized Shannon index summarizes both the presence and relative distribution of these categories on a common scale from 0 to 1. The CDI is not intended to replace conventional measures of crime volume. Instead, crime count describes the amount of reported crime, while the CDI describes the heterogeneity of its composition. Examining these two measures together provides the main analytical framework of this thesis.

The analysis is based on publicly available LAPD crime incident data from January 2020 to December 2023. While the original dataset also includes records from 2024, those observations were left out so that the study would cover four complete years and allow more consistent comparisons across months and years. The main spatial unit used in the analysis is the LAPD Reporting District (RD), which is a smaller geographic unit used by the department for reporting and operational purposes. Compared with the 21 LAPD Areas, Reporting Districts provide a more detailed view of local crime patterns. This makes it possible to examine differences in crime composition at a finer spatial scale. The analysis therefore compares crime diversity across Reporting Districts, months, and seasons, while also tracking how its spatial pattern changes over successive time periods.

The central research question of this thesis is:

Main Research Question: Can a Crime Diversity Index provide additional insights into local crime patterns beyond traditional measures of crime volume?

This question is examined through two supporting research questions:

Sub-question 1: How does crime diversity vary across LAPD Reporting Districts?

Sub-question 2: How does crime diversity change over time across months and seasons?

To address these questions, the study combines descriptive, spatial, temporal, and statistical analyses. Crime count and the normalized CDI are first calculated for each Reporting District, allowing the overall volume of reported crime to be compared with its composition. Spatial patterns are examined using choropleth maps and spatial autocorrelation analysis. Monthly and seasonal analyses are then used to evaluate temporal variation in crime diversity, with one-way ANOVA and effect-size measures used to assess seasonal differences. The relationship between crime count and diversity is examined using correlation and ordinary least squares regression. Sensitivity analyses are also used to evaluate the effects of unequal incident counts, spatial dependence, and repeated observations within Reporting Districts. Finally, the study combines the spatial and temporal perspectives by comparing the geographic distributions of crime count and crime diversity across eight half-year periods from 2020 H1 through 2023 H2.

The contribution of this study is not the development of a new diversity index, since Shannon-based measures are already well established and have previously been applied to crime. Instead, this thesis applies the Crime Diversity Index within an integrated and reproducible spatial–temporal framework using LAPD Reporting Districts. By examining crime volume and crime composition together, the study evaluates whether local crime patterns contain information that is not fully represented by crime counts alone. It also connects crime diversity with spatial autocorrelation, temporal comparison, the count–diversity relationship, and descriptive spatio-temporal analysis. This framework provides a relatively straightforward way to examine the composition of reported crime while retaining conventional crime volume as an important reference measure.

The remainder of this thesis is organized as follows. Chapter 2 reviews previous research on traditional crime measures, crime diversity, and spatial and temporal crime patterns, and identifies the research gap addressed by this study. Chapter 3 describes the LAPD crime data, the variables used in the analysis, and the data preparation and aggregation procedures. Chapter 4 presents the exploratory and empirical analyses of Reporting District-level diversity, spatial autocorrelation, monthly and seasonal patterns, the relationship between crime count and diversity, and the spatio-temporal evolution of both measures. Chapter 5 presents the methodological framework, including construction of the normalized Crime Diversity Index, the spatial and temporal statistical methods, the count–diversity analysis, and the reproducible analytical workflow. Finally, Chapter 6 summarizes the main findings, discusses their practical implications and limitations, and concludes the thesis.

CHAPTER 2

Literature Review

2.1 Traditional Crime Measures and Their Limitations

Crime is commonly measured using crime counts and crime rates. These measures provide a simple and intuitive way to describe the amount of crime within a geographic area and to compare crime levels across locations or time periods. However, crime counts and rates mainly describe the quantity of crime. When different offenses are combined into a single measure, differences in crime type, composition, and severity may be overlooked.

One limitation of aggregated crime measures is that different crime types do not necessarily share the same spatial pattern. A study of multiple crime categories in Campinas, Brazil, showed that crime was highly concentrated in particular locations, but individual crime types often exhibited different spatial patterns. These differences became especially apparent at finer geographic scales, where the spatial pattern of total crime did not necessarily represent the patterns of individual offenses [PMA15]. This suggests that aggregating different offenses into a single crime measure can conceal important local variation in the types of crime occurring across an urban area.

A second limitation is that conventional crime measures generally give the same numerical contribution to offenses with very different levels of seriousness. One alternative approach is the Crime Harm Index, which weights offenses according to their severity rather than treating every recorded crime equally. A study using the Japanese Crime Harm Index compared conventional crime counts with weighted crime harm across 77 jurisdictions in central Tokyo. Although the two measures were related, the temporal and spatial variability of crime harm differed from that of crime counts [OAN23]. These results suggest that crime frequency alone does not fully describe the seriousness or distribution of crime within local areas.

Traditional measures may also overlook the relative composition of crime types. Rather than treating different crime categories as independent counts, compositional approaches consider each category as part of an overall crime profile. Research using crime statistics from European countries applied Compositional Data Analysis (CoDA) to examine the relative relationships among crime categories rather than relying only on their absolute frequencies [DK25]. The results showed that examining crime composition can reveal patterns that are not represented by conventional counts or rates alone.

Together, these studies illustrate several limitations of relying on a single aggregate measure of crime. Total crime can obscure differences in the spatial distribution of individual offenses, crime counts do not distinguish between offenses of different severity, and aggregate measures provide limited information about the relative composition of crime types. These limitations do not make traditional crime measures unnecessary; rather, they suggest that additional measures can provide a more complete description of local crime patterns. This motivates the use of crime diversity as a complementary measure, which is discussed in the following section.

2.2 Crime Diversity and Alternative Crime Measures

The limitations of traditional crime measures have encouraged researchers to consider alternative ways of describing local crime patterns. One such approach is crime diversity, which focuses on the variety and relative distribution of crime types within an area. Rather than asking only how much crime occurs, crime diversity considers whether incidents are concentrated in a small number of crime categories or distributed more evenly across multiple categories. It therefore provides a different perspective from conventional crime counts and rates.

The concept of diversity has been widely used in fields such as ecology to describe the composition of a population or community. Diversity measures generally consider two related characteristics: richness, or the number of categories present, and evenness, or how observations are distributed among those categories. A review of commonly used diversity indices shows that different measures emphasize these characteristics in different ways and that the choice of index should depend on the purpose of the analysis [KVT23]. Among these measures, the Shannon Diversity Index incorporates both richness and evenness and provides a single summary of the distribution across categories. This makes it suitable for crime data, where both the number of crime types and their relative frequencies are of interest.

The application of diversity measures to crime was developed more directly through the concept of crime diversity. Crime diversity treats the mixture of offenses occurring within an area as an additional dimension of local crime patterns. Research introducing this approach showed that crime diversity is related to, but distinct from, other characteristics such as crime concentration and poverty [Bra16]. This distinction is important because two areas with similar levels of crime may contain different combinations of offenses. Crime diversity can therefore complement measures that focus primarily on the overall amount or concentration of crime.

Later research applied the concept of crime diversity to Los Angeles and examined whether differences in the local environment were associated with differences in the diversity of crime. Using the Shannon Diversity Index together with Google Street View imagery and demographic information, the study found that environmental and population characteristics were associated with variation in crime diversity [KCM21]. This application demonstrates that crime diversity can be measured at a local geographic scale and can reveal systematic variation between areas rather than simply reflecting random differences in crime composition.

Crime diversity has also been examined from the perspective of crime opportunity. A recent theoretical framework suggests that diversity can emerge through the interaction between offender mobility and the distribution of crime opportunities across environments [BV25]. Offenders who encounter a wider range of opportunities may become involved in a broader variety of offenses, while environments containing more heterogeneous opportunities may also produce more diverse crime patterns. This perspective provides a possible explanation for why crime diversity may vary across locations and reinforces the idea that diversity represents a meaningful characteristic of local crime environments.

An important issue is whether diversity simply increases with the frequency of offending. Research examining the relationship between offending frequency and diversity found that the two are related but should not be treated as equivalent measures [FH16]. Frequency describes how often offending occurs, whereas diversity describes the range and distribution of offense types. This distinction is particularly relevant when evaluating whether a diversity measure provides information beyond a conventional measure of crime volume.

Taken together, these studies provide both a methodological and conceptual foundation for using crime diversity as a complementary measure of local crime patterns. The Shannon Diversity Index summarizes the distribution of crime across different categories, while previous crime research suggests that diversity captures a dimension of offending that is related to, but not identical to, crime frequency. Building on this literature, the present study uses a normalized Shannon Crime Diversity Index to measure the distribution of crime across eight crime categories within LAPD Reporting Districts. The relationship between crime volume and crime diversity is then examined to evaluate whether the diversity measure provides additional information beyond traditional crime counts.

2.3 Spatial and Temporal Crime Patterns

Crime is not distributed uniformly across either space or time. Some locations experience substantially more crime than others, while crime levels and patterns can also change across different time periods. These variations are important for local crime analysis because a city-wide or long-term aggregate may conceal differences that occur at smaller geographic and temporal scales. Spatial and temporal analysis therefore provides an important framework for understanding how local crime patterns vary and change.

Spatial crime research has increasingly emphasized the importance of geographic scale. Crime mapping allows researchers to examine the distribution of crime across smaller areas rather than relying only on city-wide statistics. However, the observed spatial pattern can depend on the geographic and temporal units used in the analysis, creating important challenges for the interpretation of crime maps [Rat09]. This point is especially important when crime is concentrated within relatively small areas. Using a finer geographic scale can make local differences more visible, while these differences may be overlooked when crime incidents are combined into larger administrative units.

Crime patterns are also dynamic over shorter time periods. Studies of weekly crime concentration have found that crime can remain highly concentrated even over relatively short time periods [WS22]. However, the locations where crime is concentrated, as well as the degree of concentration, may still change from one period to another. This suggests that a spatial pattern observed over a long study period should not necessarily be interpreted as a fixed representation of crime. Examining multiple time periods can provide additional information about the stability and change of local crime patterns.

Seasonality represents another important source of temporal variation. An empirical study of crime in London found that different types of crime exhibit distinct seasonal patterns and that temporal fluctuations are not uniform across offense categories [TBJ24]. Earlier research similarly demonstrated that seasonal cycles vary across crime types and that the magnitude and timing of seasonal effects are not necessarily constant [MLP12]. These findings are particularly relevant to crime diversity because changes in the relative frequency of different offense types may also change the overall composition of crime. Consequently, examining diversity across months and seasons can provide information that may not be visible from annual crime totals alone.

Local crime patterns may also change as the characteristics of an area develop over time. Research taking a dynamic view of neighborhoods has shown that crime and local structural characteristics can influence one another rather than remaining fixed throughout a study period [Hip10]. Although the present study does not examine socioeconomic or structural neighborhood characteristics, this dynamic perspective reinforces the broader point that local crime environments should not always be treated as temporally stable.

More recent crime research has increasingly considered space and time jointly rather than treating spatial and temporal variation as separate problems. A systematic review of multi-scale spatio-temporal crime prediction research identified spatial heterogeneity, temporal variation, and their interaction as important components of crime analysis [DD23]. The reviewed studies used a range of spatial and temporal scales, illustrating that crime patterns can differ not only between locations or between time periods, but also in the way individual locations change over time. Although much of this literature focuses on crime prediction, it provides a broader motivation for examining crime patterns from a joint spatio-temporal perspective.

A specific example is a spatio-temporal Bayesian analysis of urban burglary that incorporated spatial effects, temporal effects, and interactions between area and time [HZD18]. The model allowed temporal crime trends to differ across geographic areas rather than assuming that all locations followed the same pattern. This illustrates an important advantage of spatio-temporal analysis: an overall temporal trend may conceal different changes occurring in individual areas. While the present study does not employ a Bayesian prediction model, the same general motivation supports examining whether spatial patterns of crime volume and crime diversity remain stable or change across time periods.

Taken together, the literature demonstrates that crime patterns exhibit spatial, temporal, and joint spatio-temporal variation. These findings provide the basis for examining crime diversity at multiple geographic and temporal scales. In the present study, spatial variation is examined across LAPD Reporting Districts, while temporal variation is evaluated across months and seasons. In addition, the spatial distributions of crime volume and crime diversity are compared across multiple time periods to examine how local patterns change over time. This combined perspective connects the spatial and temporal components of the analysis and provides a more complete basis for evaluating whether crime diversity offers information beyond traditional crime volume.

2.4 Research Gap and Study Contribution

The literature reviewed in the previous sections demonstrates that crime can be described from several perspectives. Traditional crime counts and rates provide information about crime volume, while alternative measures capture characteristics such as crime severity, composition, and diversity. Spatial and temporal research further shows that crime patterns vary across locations and over time. Together, these findings suggest that relying on a single aggregate measure may provide only a partial description of local crime patterns.

Crime diversity extends this perspective by describing the distribution of different offense types within an area. Previous research has established crime diversity as a meaningful dimension of crime and has shown that diversity is related to, but not equivalent to, crime frequency [Bra16, FH16]. However, much of the existing literature has focused on defining crime diversity or explaining its relationship with environmental and offender characteristics. This leaves room for further examination of how crime diversity itself varies across local geographic units and over time.

At the same time, research has increasingly recognized the importance of examining spatial and temporal crime patterns jointly [DD23, HZD18]. Much of this work has focused on crime counts, individual offense types, or crime prediction. The present study connects this spatio-temporal perspective with crime diversity by examining whether the spatial distribution of crime diversity remains stable or changes across different time periods.

Building on these areas of research, this study applies a normalized Shannon Crime Diversity Index to reported crime data from the Los Angeles Police Department between January 2020 and December 2023. Crime diversity is examined across LAPD Reporting Districts and across months and seasons, while conventional crime volume is retained as a reference measure. Spatial autocorrelation and changes in spatial patterns over time are also examined. By looking at crime diversity across spatial, temporal, and spatio-temporal dimensions, the study assesses whether it adds information about local crime patterns beyond what can be captured by crime volume alone.

CHAPTER 3

Data

3.1 Data Source

The data used in this study were obtained from the Los Angeles Police Department (LAPD) *Crime Data from 2020 to 2024*, which is publicly available through the Los Angeles Open Data Portal. The dataset contains officially reported crime incidents from across the City of Los Angeles. It includes information on occurrence date and time, crime code and description, Reporting District, geographic coordinates, and victim characteristics. Because the records are collected and maintained by the LAPD, they provide a detailed basis for examining crime patterns across locations and over time.

The original dataset includes crime incidents from January 2020 through December 2024. This study uses only the records from January 2020 to December 2023. The 2024 records were excluded so that the analysis covers four complete years. This also allows more consistent comparisons across months and years. After restricting the study period and completing the data preparation procedures described in Section 3.3, the final incident-level dataset contains 875,087 reported crime incidents.

The main spatial unit used in this study is the LAPD Reporting District (RD). Reporting Districts are relatively small geographic units used by the LAPD for recording crime and supporting police operations. Because they are smaller than the 21 LAPD geographic Areas, they provide a more detailed spatial view and make it easier to identify local differences in crime patterns. The crime records contain 1,200 Reporting District identifiers. For analyses requiring geographic boundaries, 1,135 Reporting District polygons with valid spatial boundaries were available and successfully matched to the crime data. Therefore, descriptive analyses use the Reporting Districts represented in the crime records, while spatial analyses requiring polygon boundaries are based on the 1,135 matched Reporting Districts.

3.2 Variable Description

The original LAPD crime dataset contains nearly thirty variables describing different aspects of each reported crime incident. Since this study focuses on the spatial and temporal patterns of crime diversity, only variables relevant to the analysis were retained. The main variables can be grouped into three categories: temporal, spatial, and crime-related variables.

Temporal variables describe when each crime occurred. The primary variables include `DATE OCC` and `TIME OCC`, which record the occurrence date and time of each incident. Additional temporal variables, including year, month, season, and half-year period, were derived from the occurrence date during data preparation. These derived variables support the temporal and spatio-temporal analyses presented in later chapters.

Spatial variables identify where each crime occurred. The most important variable is `Rpt Dist No`, which identifies the LAPD Reporting District and serves as the primary spatial unit throughout this study. Additional geographic variables, including `AREA`, `AREA NAME`, `LAT`, and `LON`, provide location information for individual crime incidents and support the preparation and validation of the spatial data.

Crime-related variables describe the type of reported offense. The variables `Crm Cd` and `Crm Cd Desc` provide the crime code and its corresponding description. Because the original dataset contains many detailed offense types, these descriptions were grouped into broader crime categories during data preparation. The resulting categories form the basis for constructing crime composition and measuring crime diversity.

Table 3.1 summarizes the main original variables used in this study.

Category	Variable	Description
Temporal	DATE OCC	Date when the crime occurred
	TIME OCC	Time when the crime occurred
Spatial	Rpt Dist No	Reporting District used as the primary spatial unit
	AREA / AREA NAME	LAPD administrative area information
	LAT / LON	Geographic coordinates of each incident
Crime	Crm Cd	Crime code
	Crm Cd Desc	Crime description

Table 3.1: Main variables used in this study.

3.3 Data Preparation

Before conducting the analysis, the original LAPD crime data were cleaned and processed to create consistent analytical datasets. The study period was restricted to January 2020 through December 2023. Duplicate incident records were removed, and observations with invalid geographic coordinates were excluded. After these preparation steps, 875,087 reported crime incidents remained for analysis.

The original dataset contains a large number of detailed crime descriptions. To make crime composition easier to compare across Reporting Districts, individual offenses were grouped into eight broader categories: Violent Crime, Property Crime, Vehicle-related Crime, Public Order Crime, Sexual Crime, Financial Crime, Weapon-related Crime, and Other Crime. This grouping reduces the complexity of the original crime classifications while retaining differences in the types of crime reported across locations and time periods.

Additional temporal variables were created from the occurrence date, including year, month, season, and half-year period. Seasons were used to examine broad seasonal differences in crime diversity, while the half-year variable divided the four-year study period into eight consecutive periods, from 2020 H1 to 2023 H2. The half-year periods were used for the spatio-temporal analysis to examine how spatial crime patterns changed over time.

Because different analyses require different levels of aggregation, three analytical datasets were constructed. First, crime incidents were aggregated by Reporting District over the full 2020–2023 study period for the overall spatial analysis. Second, Reporting District–month observations were created for the monthly and seasonal analyses. Third, Reporting District–half-year observations were created for the spatio-temporal comparison. For each aggregation level, the number of incidents in each of the eight crime categories and the corresponding total crime count were calculated.

Crime-category proportions were then calculated by dividing the number of incidents in each category by the total number of crimes within the corresponding Reporting District and time period. These proportions provide the input for the Crime Diversity Index introduced in Chapter 5. Using the same crime categories at each aggregation level allows crime volume and crime composition to be compared consistently across the spatial, temporal, and spatio-temporal analyses.

CHAPTER 4

Exploratory Data Analysis

4.1 Dataset and Analytical Overview

This chapter presents the empirical analysis of crime diversity across Los Angeles Police Department (LAPD) Reporting Districts. The analysis is organized around two related dimensions of local crime patterns: crime volume, measured by the number of reported incidents, and crime composition, measured by the Crime Diversity Index. Examining these measures together provides a basis for assessing whether crime diversity offers information about local crime patterns beyond that provided by crime counts alone.

The analysis uses crime incident records from the LAPD Crime Data from 2020 to 2024 dataset. The study period is restricted to January 1, 2020 through December 31, 2023. Records from 2024 are excluded because they do not provide a complete year for comparison. After restricting the study period, removing duplicate incident records and excluding observations with invalid geographic coordinates, the final analytical dataset contains 875,087 reported crime incidents.

The crime records cover 21 LAPD Areas and 1,200 Reporting District identifiers. Reporting Districts are used as the primary spatial unit of analysis because they provide a finer geographic scale than LAPD Areas and allow local differences in crime patterns to be examined in greater detail. For the spatial mapping analysis, 1,135 Reporting District polygons with valid geographic boundaries are available and successfully matched to the crime data. This distinction is maintained throughout the chapter: the descriptive analyses use all Reporting Districts represented in the crime records, while analyses requiring geographic boundaries use the 1,135 matched spatial units.

The original LAPD crime descriptions are grouped into eight broader categories: Property Crime, Vehicle-related Crime, Violent Crime, Weapon-related Crime, Public Order Crime, Sexual Crime, Financial Crime, and Other Crime. These categories are used to describe the composition of crime within each spatial or temporal unit. Crime diversity is measured using the normalized Shannon index, with larger values representing a more even distribution of incidents across the eight crime categories and smaller values representing greater concentration in a limited number of categories. The construction and interpretation of this measure are discussed formally in Chapter 5.

To examine crime diversity from complementary perspectives, the data are aggregated at three levels. First, crime incidents are aggregated by Reporting District over the full 2020–2023 study period to examine overall spatial differences in crime volume and diversity. Second, Reporting District-month observations are constructed to examine monthly and seasonal variation. Third, Reporting District-half-year observations are used to examine how the spatial patterns of crime count and crime diversity evolve over time. The half-year analysis divides the four-year study period into eight consecutive periods, from 2020 H1 through 2023 H2, allowing spatial and temporal variation to be considered jointly rather than as separate dimensions.

The remainder of this chapter proceeds from descriptive to spatial and temporal comparisons. Section 4.2 examines variation in crime diversity across Reporting Districts. Section 4.3 evaluates spatial patterns and spatial autocorrelation. Section 4.4 examines monthly and seasonal variation. Section 4.5 investigates the relationship between crime volume and crime diversity using correlation and regression analysis. Section 4.6 extends the analysis to the joint spatial–temporal setting by comparing the geographic distribution of crime count and crime diversity across the eight half-year periods. Together, these analyses provide the empirical foundation for the statistical framework and methodological discussion presented in Chapter 5.

4.2 Reporting District-Level Crime Diversity

To examine variation in crime composition across Los Angeles, crime incidents were aggregated over the full 2020–2023 study period by LAPD Reporting District. For each Reporting District, the total number of reported crimes and the number of incidents in each of the eight crime categories were calculated. The normalized Shannon Crime Diversity Index was then computed from the distribution of these categories. This aggregation resulted in 1,200 observations at the Reporting District level.

Crime volume differed substantially across Reporting Districts. The average district recorded 729.24 crime incidents during the study period, while the median was 617. The number of reported incidents ranged from 1 to 4,654, with a standard deviation of 589.88. The difference between the mean and median, together with the relatively large standard deviation, indicates that crime volume was unevenly distributed across Reporting Districts, with some districts recording substantially more incidents than others.

Crime diversity also varied across Reporting Districts. The normalized Shannon index had a mean of 0.717 and a median of 0.762, with values ranging from 0 to 0.880. Figure 4.1 shows the distribution of the index across Reporting Districts. Most districts were concentrated toward the upper part of the observed distribution, particularly around values between approximately 0.70 and 0.85. This indicates that, in most Reporting Districts, reported crimes were distributed across several crime categories rather than being concentrated entirely within a single category.

The distribution is also characterized by a small number of Reporting Districts with unusually low diversity values, including some values at or near zero. These observations should be interpreted cautiously. Several of the lowest-diversity districts also contained very small numbers of reported crimes. With only a few observed incidents, fewer crime categories can be represented, which mechanically limits the possible value of the Shannon index. Therefore, an extremely low diversity value in a low-volume Reporting District does not necessarily indicate a persistent concentration of crime in a single category.

This feature of the distribution also illustrates an important distinction between crime

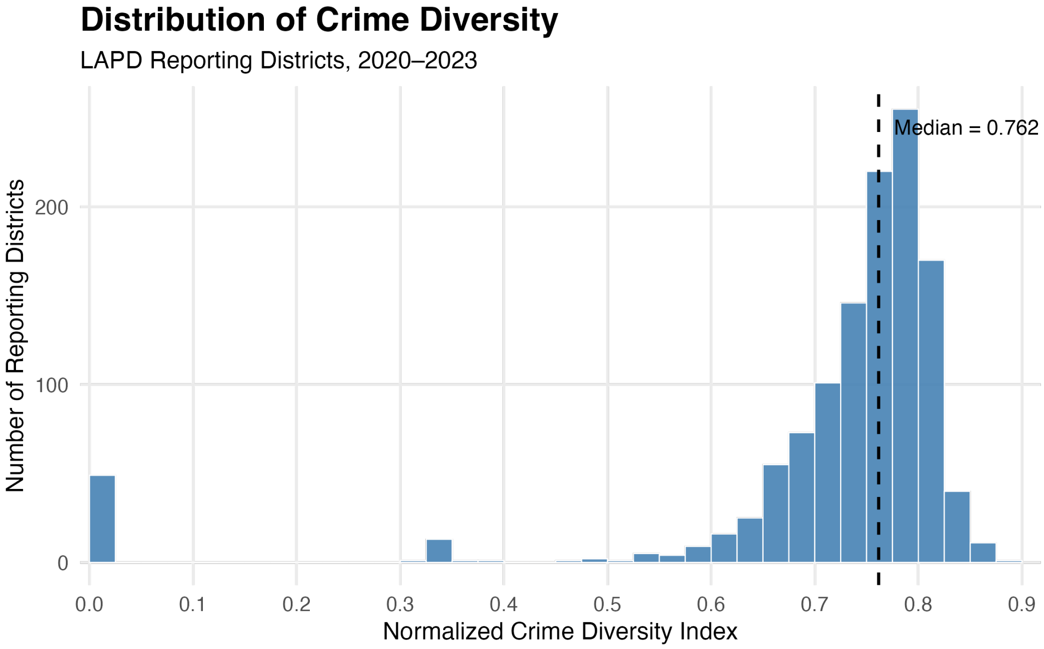

Figure 4.1: Distribution of the normalized Crime Diversity Index across LAPD Reporting Districts, 2020–2023. The dashed line represents the median normalized Shannon index (0.762).

volume and crime composition. Two Reporting Districts may record different numbers of crimes while still having similar distributions across crime categories, or they may have similar crime counts but different levels of diversity. The crime count and the Crime Diversity Index, therefore, describe different characteristics of local crime patterns. Their empirical relationship is examined more directly in Section 4.5.

In general, the analysis at the Reporting District-level shows substantial variation in both the volume and composition of reported crime in Los Angeles. The concentration of most diversity values within a relatively high range suggests that multiple crime categories are represented in most Reporting Districts, while the lower tail identifies a smaller set of districts with more concentrated observed crime compositions or very small crime counts. These descriptive differences provide the basis for the spatial analysis in the following section, which examines whether similar levels of crime volume and crime diversity tend to occur in neighboring Reporting Districts.

4.3 Spatial Patterns and Spatial Autocorrelation

The Reporting District-level analysis in the previous section established that both crime volume and crime diversity vary across Los Angeles. This section examines whether these differences also exhibit geographic structure. The spatial analysis is based on 1,135 LAPD Reporting District polygons with valid geographic boundaries, all of which were successfully matched to the district-level crime data.

Figure 4.2 compares the spatial distribution of crime count and crime diversity. Because crime counts are highly right skewed, the count map uses $\log(1 + \text{crime count})$ to reduce the influence of Reporting Districts with exceptionally large numbers of incidents. Crime diversity is represented by the normalized Shannon index.

The two maps show that neither crime volume nor crime diversity is distributed uniformly across Reporting Districts. Areas with relatively high and low values appear in different parts of the city, suggesting that local crime patterns have a geographic component. At the same time, the two maps are not identical. Some Reporting Districts that appear similar in crime

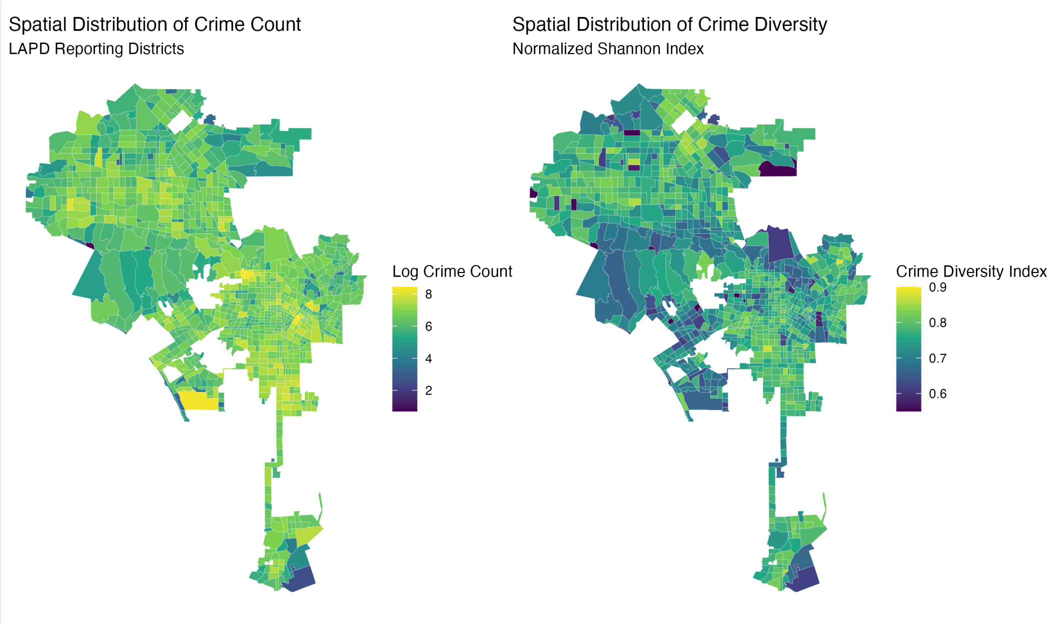

Figure 4.2: Spatial distributions of crime count and Crime Diversity Index across LAPD Reporting Districts, 2020–2023. Crime count is displayed on a log-transformed scale, while crime diversity is measured using the normalized Shannon index.

volume show different levels of crime diversity, while districts with comparable diversity may differ in the number of reported incidents. This visual comparison provides initial evidence that crime volume and crime composition capture related but not identical spatial characteristics.

To evaluate the spatial patterns more formally, Global Moran's I was calculated for both measures. Reporting Districts were defined as neighbors using Queen contiguity, under which two districts are considered neighbors when their polygons share either a boundary or a vertex. The spatial structure contained 7,008 nonzero neighbor links. On average, each Reporting District had about 6.17 neighboring districts. Row-standardized spatial weights were then used to calculate Moran's I.

For log-transformed crime count, Global Moran's I was 0.2925. For the normalized Crime Diversity Index, Moran's I was 0.2862. Both statistics were positive. This indicates that Reporting Districts with similar values tended to be located near one another rather than being randomly distributed across space. A Monte Carlo permutation procedure with

999 simulations produced $p = 0.001$ for both measures, providing evidence of statistically significant positive spatial autocorrelation.

Figure 4.3 provides a graphical representation of these results. In each Moran scatterplot, the horizontal axis shows the standardized value for a Reporting District. The vertical axis shows its spatial lag, which represents the weighted average of values in neighboring districts. The positive slopes are consistent with the positive Global Moran's I statistics.

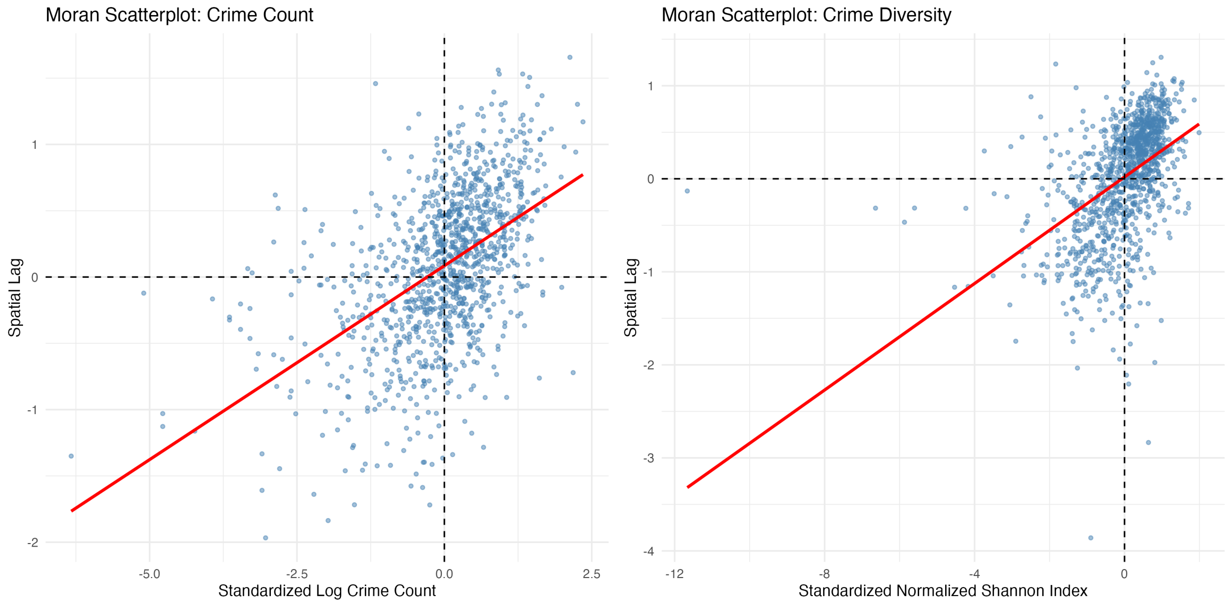

Figure 4.3: Moran scatterplots for log-transformed crime count and the normalized Crime Diversity Index across LAPD Reporting Districts.

The similarity between the two Moran's I values is notable. Both crime count and crime diversity show moderate positive spatial autocorrelation. However, this does not mean that the two measures follow the same geographic pattern. Moran's I measures the degree to which similar values are spatially clustered within each measure; it does not measure whether the two variables are spatially identical. The differences visible in Figure 4.2 therefore remain important when interpreting the two measures.

Overall, these results show that crime composition, like crime volume, has a meaningful spatial structure across LAPD Reporting Districts. Crime diversity is not simply distributed randomly across the city; neighboring districts tend to have more similar diversity levels than would be expected under spatial randomness. This finding supports the use of Reporting Districts as a meaningful spatial unit for examining crime diversity and provides evidence relevant to the first research question concerning how crime diversity varies across Reporting Districts. The next section turns to the temporal dimension of the analysis and examines whether crime diversity also varies across months and seasons.

4.4 Temporal and Seasonal Patterns of Crime Diversity

The previous sections examined variation in crime diversity across space. This section addresses the temporal dimension of the analysis by examining how crime diversity changed across months and seasons during the 2020–2023 study period. Crime incidents were aggregated by Reporting District and month, and the normalized Shannon index was calculated separately for each Reporting District-month observation. The resulting dataset contained 53,253 Reporting District-month observations across 48 months.

Figure 4.4 presents the average normalized Crime Diversity Index for each month. Monthly averages fluctuate throughout the study period, but the magnitude of these changes is relatively limited. The LOESS curve provides a smoothed representation of the temporal pattern and shows gradual changes rather than a strong sustained upward or downward trend. Overall, the figure suggests that average crime diversity remained relatively stable at the citywide level between 2020 and 2023, despite short-term month-to-month variation.

To examine whether systematic differences were present across seasons, the monthly observations were classified into Winter, Spring, Summer, and Fall. Figure 4.5 compares the distributions of the normalized Crime Diversity Index across the four seasons. The distributions overlap substantially, suggesting that seasonal differences are small relative to the overall variation among Reporting District-month observations.

The descriptive statistics show a small but consistent seasonal pattern. The mean normalized Crime Diversity Index was 0.569 in Winter, 0.576 in Spring, 0.584 in Summer, and 0.579 in Fall. Summer therefore had the highest average diversity, while Winter had the lowest. The difference between the highest and lowest seasonal means was approximately

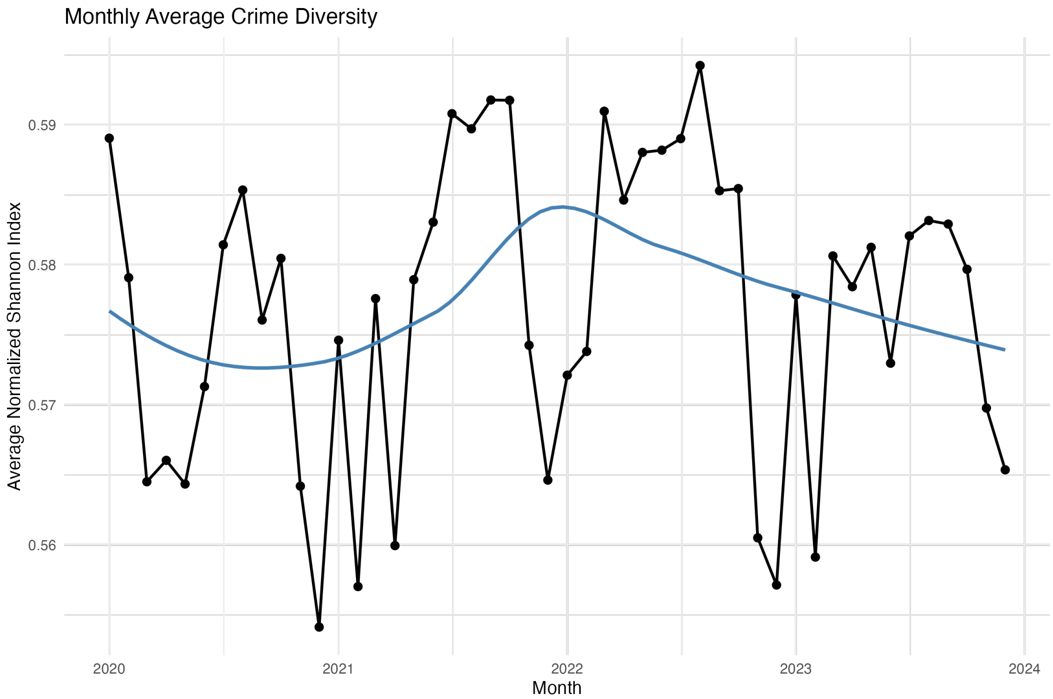

Figure 4.4: Monthly average Crime Diversity Index across LAPD Reporting Districts, 2020–2023. The smooth curve represents the LOESS trend across the study period.

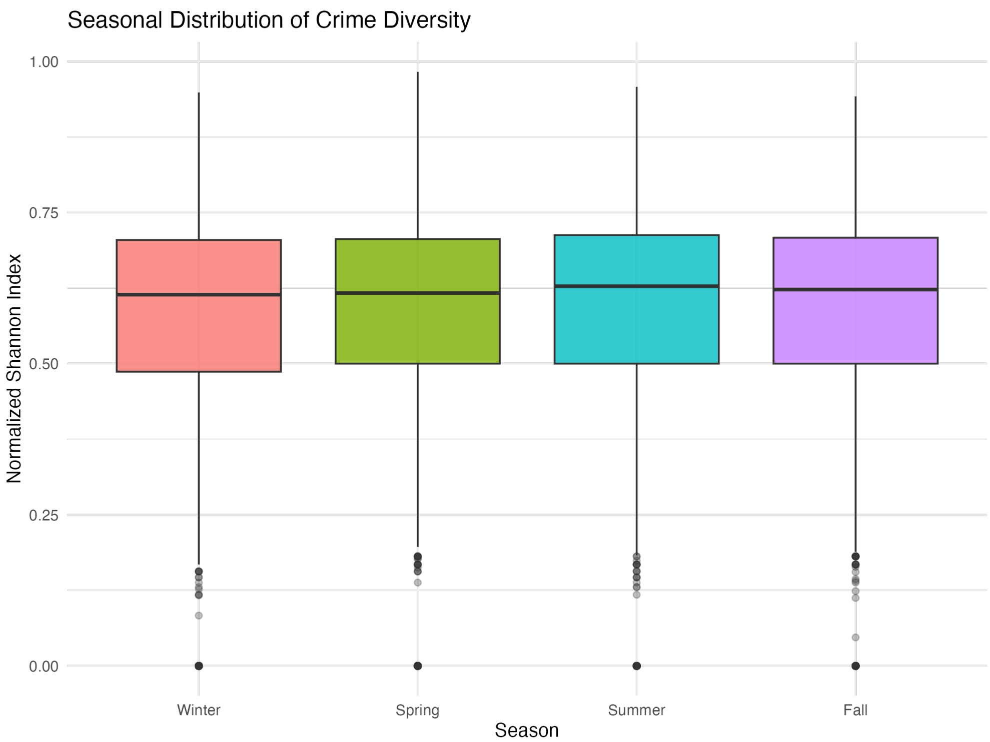

Figure 4.5: Distribution of the normalized Crime Diversity Index across seasons.

0.016 on the normalized 0–1 scale, indicating that the magnitude of the seasonal variation was small.

Figure 4.6 presents the mean Crime Diversity Index for each season together with 95% confidence intervals. The pattern is consistent with the distributional comparison: average diversity increases from Winter to Spring and Summer before declining slightly in Fall. However, the differences between seasonal averages remain modest.

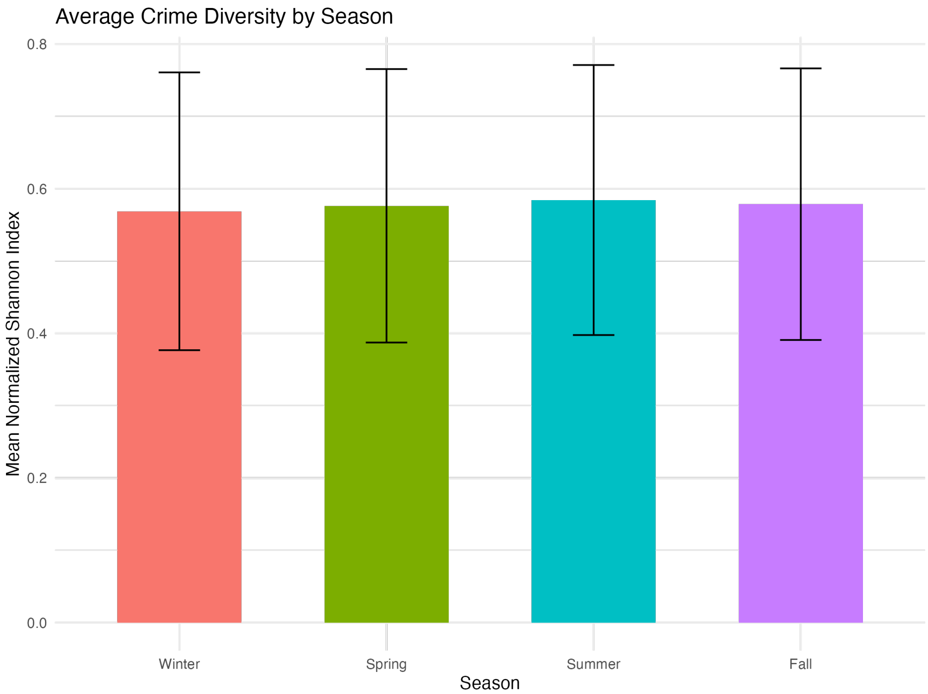

Figure 4.6: Mean normalized Crime Diversity Index by season with 95% confidence intervals.

A one-way analysis of variance (ANOVA) was used to evaluate whether mean crime diversity differed across the four seasons. Before interpreting the ANOVA, homogeneity of variance was assessed using Levene's test. The test produced $F = 3.40$ and $p = 0.017$, indicating some evidence of unequal variances across seasons. The conventional ANOVA results should therefore be interpreted with this assumption in mind.

The one-way ANOVA nevertheless indicated a statistically significant association between season and crime diversity ($F(3, 53249) = 15.48$, $p < 0.001$). Statistical significance alone, however, does not indicate that the seasonal differences were substantively large. The corresponding effect size was $\eta^2 = 0.000872$, meaning that season accounted for less than 0.1%

of the observed variation in the normalized Crime Diversity Index. Thus, although differences among seasonal means were statistically detectable in the large sample, their practical magnitude was very small.

Tukey's HSD post-hoc comparisons provide additional detail about these differences. Relative to Winter, mean crime diversity was significantly higher in Spring ($p = 0.006$), Summer ($p < 0.001$), and Fall ($p < 0.001$). Summer was also significantly higher than Spring ($p = 0.003$). In contrast, the difference between Fall and Spring was not statistically significant ($p = 0.770$), and the difference between Fall and Summer did not reach the 0.05 significance level ($p = 0.063$). The largest estimated difference was between Summer and Winter. Even then, the difference in the normalized index was only about 0.016.

Because the same Reporting Districts appear across multiple months, the observations are not fully independent. As a result, the independence assumption of a conventional one-way ANOVA is not completely appropriate for the RD-month data. To evaluate whether the seasonal result remained after accounting for these repeated observations, a linear mixed-effects model was therefore fitted as a sensitivity analysis. Season was included as a fixed effect, while Reporting District was included as a random intercept, allowing each district to have its own baseline level of crime diversity.

The mixed-effects model continued to identify a statistically significant overall seasonal effect,

$$F(3, 52050) = 33.78, \qquad p < 0.0001.$$

Using Winter as the reference season, the estimated differences were 0.0076 for Spring, 0.0163 for Summer, and 0.0103 for Fall. All three coefficients were statistically significant, but their magnitudes remained small on the normalized 0–1 CDI scale. The estimated standard deviation of the Reporting District random intercept was 0.168, compared with a within-district residual standard deviation of 0.134, indicating that persistent differences among Reporting Districts account for an important component of the observed variation.

The mixed-effects sensitivity analysis therefore leads to the same substantive conclusion as the original ANOVA. Seasonal differences in crime diversity are statistically detectable even after accounting for repeated monthly observations within Reporting Districts, but their estimated magnitude remains modest. In particular, the largest estimated difference, between Summer and Winter, remains approximately 0.016.

Overall, the temporal analysis indicates that crime diversity was relatively stable during the 2020–2023 study period. Monthly averages showed short-term fluctuations, but there was no large sustained change over time. Seasonal differences were statistically significant, although their magnitude was small. The mixed-effects sensitivity analysis reached the same general conclusion after accounting for repeated observations within Reporting Districts. This suggests that temporal variation is present, but it explains only a limited share of the overall differences in crime diversity. In relation to the second research question, crime diversity does change across months and seasons, but the seasonal differences are modest compared with the broader variation across Reporting Districts. Section 4.5 then examines another source of variation by focusing on the relationship between crime volume and crime diversity.

4.5 Relationship Between Crime Count and Crime Diversity

A central objective of this study is to determine whether the Crime Diversity Index provides information about local crime patterns beyond traditional crime volume. The descriptive results in Section 4.2 showed that both measures vary substantially across Reporting Districts, while Section 4.3 showed that both also exhibit significant spatial autocorrelation. This section examines their relationship directly using correlation and regression analysis and then evaluates two important sources of dependence through sensitivity analyses.

Because total crime counts are strongly right-skewed across Reporting Districts, the count variable was transformed using $\log(1 + \text{crime count})$. Figure 4.7 shows the relationship between log-transformed crime count and the normalized Crime Diversity Index. The fitted regression line indicates a clear positive association: Reporting Districts with larger crime volumes generally have higher levels of observed crime diversity.

The Pearson correlation between log-transformed crime count and normalized crime di-

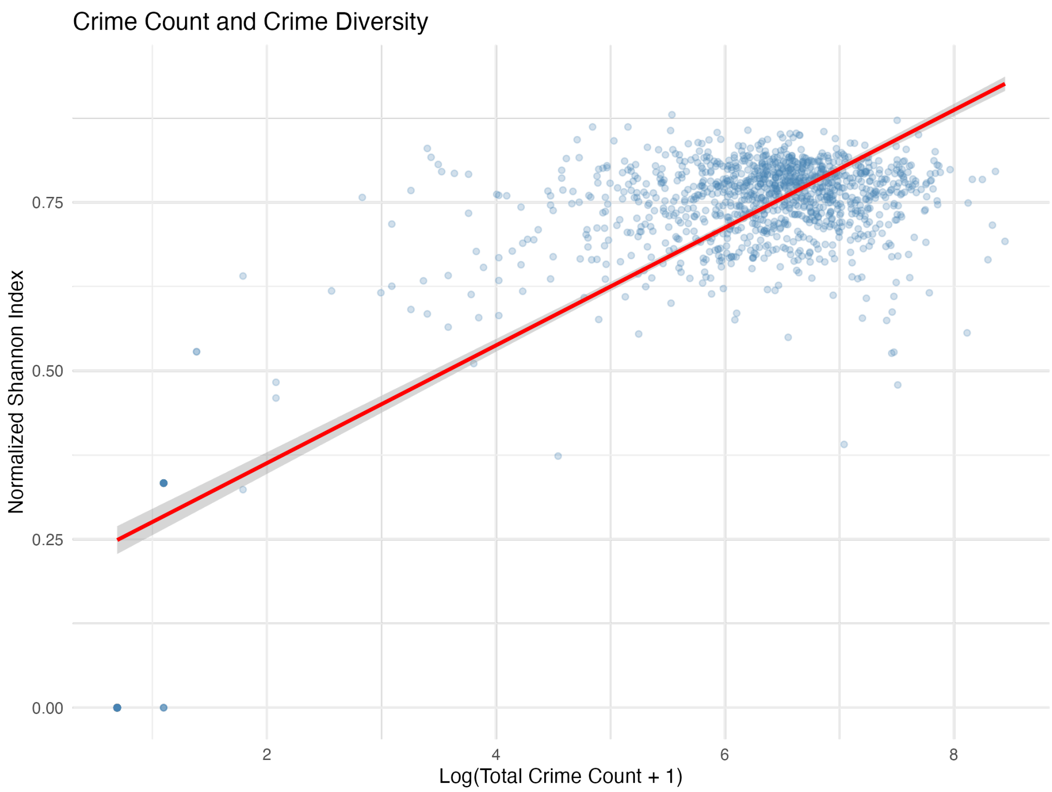

Figure 4.7: Relationship between log-transformed crime count and the normalized Crime Diversity Index across LAPD Reporting Districts, 2020–2023. The fitted line represents the ordinary least squares regression relationship.

versity was

$$r = 0.801,$$

with a 95% confidence interval from 0.780 to 0.820 and $p < 0.001$. This indicates a strong positive association between the two measures. Reporting Districts with more reported crimes therefore tend to contain a more diverse mixture of observed crime categories. This result also helps explain the low-diversity observations identified in Section 4.2, many of which occur in districts containing very small numbers of reported incidents.

To quantify this relationship further, an ordinary least squares regression model was fitted with the normalized Crime Diversity Index as the response variable and log-transformed total crime count as the explanatory variable. The fitted model can be written as

$$\widehat{D}_i = 0.1882 + 0.08735 \log(1 + C_i),$$

where $\widehat{D}_i$ represents the predicted normalized Crime Diversity Index and C_i represents the total crime count for Reporting District i. The estimated coefficient for log crime count was positive and statistically significant ($p < 0.001$). Thus, within this descriptive model, higher crime volume is associated with higher estimated crime diversity.

The model produced an R^2 of approximately 0.641, indicating that about 64.1% of the cross-district variation in the normalized Crime Diversity Index is associated with variation in log-transformed crime count in this simple linear model. This relatively strong relationship is important for interpreting the CDI. Crime diversity is clearly associated with crime volume, and part of the observed variation in diversity may result from differences in the number of incidents available to represent the eight crime categories.

At the same time, the relationship is not perfect. Approximately 35.9% of the observed cross-district variation remains unexplained by the simple count-based regression, and Reporting Districts with similar crime volumes can still exhibit different diversity values. Therefore, the CDI should not be interpreted as either completely independent of crime count or simply equivalent to it. Crime count measures the amount of reported crime, while the diversity index summarizes how observed incidents are distributed among crime categories.

Regression diagnostic plots were also examined to evaluate the adequacy of the simple linear specification. Figure 4.8 shows the residual-versus-fitted, normal Q–Q, scale-location, and leverage diagnostics. The plots indicate departures from the ideal assumptions of a simple linear model, particularly among observations with very low crime counts and diversity values. The regression is therefore used primarily as a descriptive model of the association between crime volume and crime diversity rather than as a predictive or causal model.

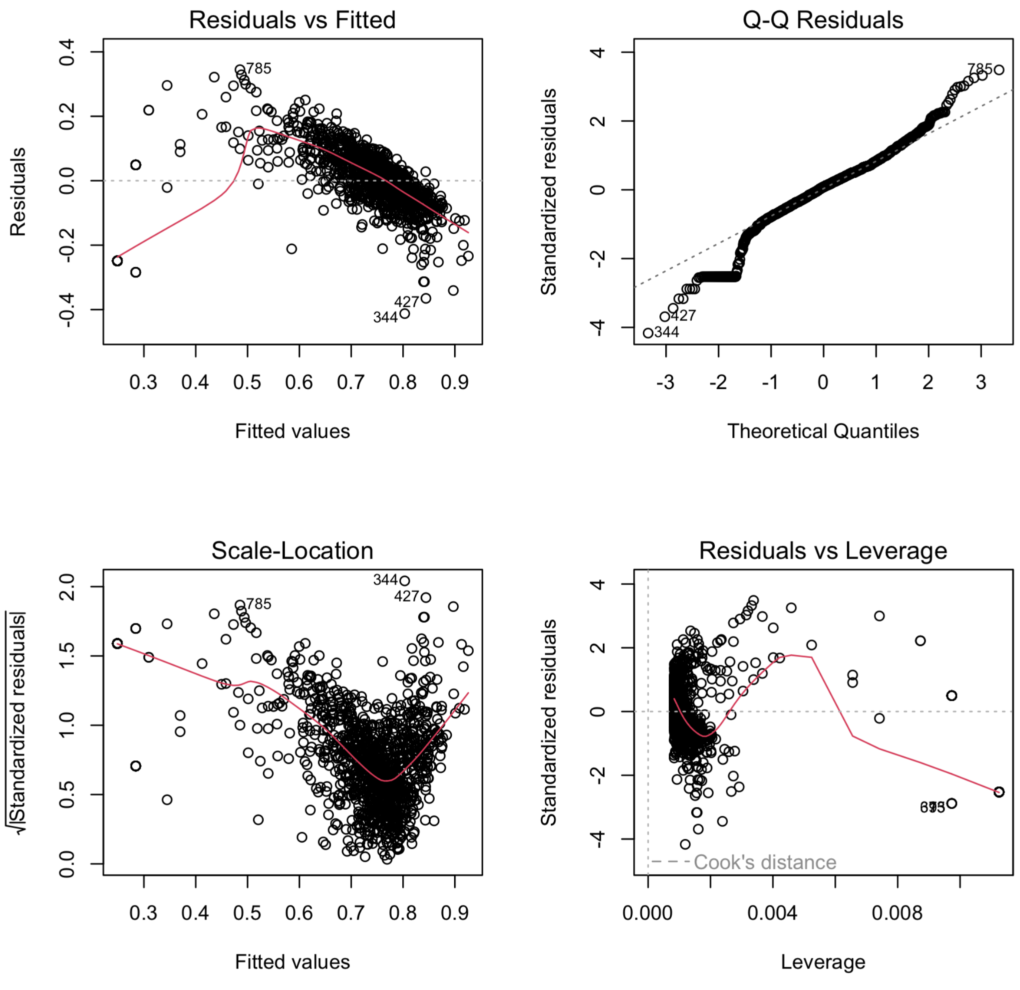

Figure 4.8: Diagnostic plots for the linear regression of normalized Crime Diversity Index on log-transformed crime count.

An additional concern is that the empirical CDI can be mechanically affected by the number of observed incidents. A Reporting District containing only a small number of crimes has fewer opportunities for multiple crime categories to appear, even if its underlying crime composition is relatively diverse. To examine how strongly the observed count–diversity relationship depends on this feature, a fixed-count subsampling sensitivity analysis was conducted.

Reporting Districts containing at least 100 incidents over the full study period were retained for this analysis, resulting in 1,085 eligible districts. For each eligible Reporting District, exactly 100 incidents were randomly sampled without replacement and the normalized Shannon CDI was recalculated. This procedure was repeated 50 times, and the mean of the 50 subsampled CDI values was used as a count-adjusted diversity estimate for each district. Using the same sample size for every district removes the direct advantage that higher-volume districts have in representing a larger number of observed crime categories.

The resulting count-adjusted CDI was highly consistent with the original CDI. The Pearson correlation between the original normalized Shannon CDI and the mean count-adjusted CDI was

$$r = 0.995,$$

with a 95% confidence interval from 0.994 to 0.995 and $p < 0.001$. The count-adjusted CDI had a mean of 0.741 and a median of 0.753. The average standard deviation of the CDI across the 50 repeated samples was approximately 0.035, indicating relatively limited variation across repeated random subsamples.

More importantly, after fixing the number of sampled incidents at 100 per district, the association between the adjusted CDI and the district's original crime volume was essentially absent:

$$r = 0.018,$$

with a 95% confidence interval from -0.042 to 0.077 and $p = 0.562$. This contrasts sharply with the original correlation of $r = 0.801$ between log-transformed crime count and the unadjusted CDI. The result indicates that a substantial portion of the original count–diversity association reflects the unequal number of observed incidents across Reporting Districts.

At the same time, the extremely strong correlation between the original and count-adjusted CDI values indicates that the relative diversity pattern across most Reporting Districts remains highly stable after controlling the number of sampled incidents. The subsampling analysis therefore provides an important qualification to the original regression result. Crime volume influences the observed level of empirical diversity, particularly through differences in sample size, but the cross-district pattern captured by the CDI is not solely a consequence of unequal crime counts.

A second issue concerns spatial dependence. Because the CDI itself exhibits significant positive spatial autocorrelation, residuals from the ordinary least squares regression may also be spatially structured. To evaluate this possibility, the OLS count–diversity model was fitted to the 1,135 Reporting Districts with matched spatial boundaries, and Global Moran's I was calculated for its residuals using the same Queen-contiguity spatial weights employed in Section 4.3.

The OLS residuals exhibited substantial positive spatial autocorrelation,

$$I = 0.266,$$

with a Monte Carlo permutation p-value of 0.001. Thus, after accounting for the linear relationship between crime count and diversity, neighboring Reporting Districts still tended to have similar unexplained CDI values. This indicates that the simple OLS model does not fully account for the geographic structure present in crime diversity.

Because of this remaining residual spatial dependence, a spatial error regression was fitted as a sensitivity analysis. The model retained normalized Shannon CDI as the response variable and log-transformed crime count as the explanatory variable while allowing the regression error to follow a spatially correlated process. The estimated spatial error parameter was

$$\lambda = 0.482,$$

and was statistically significant ($p < 0.001$), providing additional evidence that spatial dependence was present in the OLS error structure.

After accounting for this spatial dependence, the coefficient for log-transformed crime count remained positive and statistically significant:

$$\hat{\beta}_1 = 0.01795, \qquad p < 0.001.$$

The magnitude of this coefficient should not be compared directly with the original full-sample OLS coefficient because the spatial model is estimated only for the 1,135 districts with matched geographic boundaries and uses a different error specification. The relevant result is that the positive count–diversity association remains detectable after spatial dependence is incorporated into the model.

Model fit also improved under the spatial error specification. The spatial error model had an AIC of -3252.4, compared with -3093.5 for the corresponding ordinary linear model fitted to the same spatial sample. The lower AIC favors the spatial error specification for these data.

Most importantly, the residual spatial autocorrelation was substantially reduced after fitting the spatial error model. Global Moran's I for the spatial-error residuals was

$$I = -0.029,$$

and the 999-permutation test produced $p = 0.960$. The remaining residual spatial structure was therefore not statistically significant. This result suggests that the spatial error model successfully accounts for the principal global spatial dependence left unexplained by the ordinary regression.

These two sensitivity analyses address different sources of dependence: subsampling evaluates the mechanical influence of unequal incident counts, whereas the spatial error model evaluates spatial dependence in the regression residuals.

Taken together, these analyses provide a more nuanced interpretation of the relationship between crime volume and crime diversity. The original CDI is strongly associated with crime count, but fixed-count subsampling shows that this association largely disappears when the number of observed incidents is held constant. At the same time, the adjusted CDI remains extremely strongly correlated with the original CDI, indicating that the broad cross-district diversity pattern is robust to this count adjustment. The spatial analysis further shows that a simple OLS model leaves substantial geographic structure in its residuals, while a spatial error specification substantially removes that dependence and retains a positive association between crime count and diversity.

The CDI should therefore not be interpreted as independent of crime volume. Observed incident frequency affects the empirical measurement of diversity, particularly in low-volume districts. However, the sensitivity analyses show that the cross-district pattern of crime composition cannot be explained solely by unequal crime counts or by unmodeled spatial dependence. Crime count and crime diversity remain related measures of different characteristics of local crime: one summarizes the amount of reported crime, while the other summarizes its relative composition across crime categories. The following section extends this comparison by examining how the spatial distributions of crime count and crime diversity evolve jointly over time.

4.6 Spatio-Temporal Evolution of Crime Patterns

The previous analyses examined spatial and temporal variation separately. To evaluate whether the geographic patterns of crime also changed over time, a spatio-temporal analysis was conducted by dividing the 2020–2023 study period into eight consecutive half-year intervals: 2020 H1, 2020 H2, 2021 H1, 2021 H2, 2022 H1, 2022 H2, 2023 H1, and 2023 H2. Crime count and the normalized Crime Diversity Index were calculated separately for each

Reporting District within each period.

The half-year interval provides a balance between temporal resolution and the number of crime incidents available for estimating diversity. Shorter intervals would provide greater temporal detail but would also produce more Reporting District-period observations with very small crime counts, for which the diversity index may be unstable. The eight half-year periods therefore allow changes in spatial patterns to be examined while retaining a sufficient number of observations within most Reporting Districts.

Figure 4.9 shows the spatial distribution of crime count across the eight periods. Crime count is displayed using $\log(1 + \text{crime count})$ to reduce the influence of Reporting Districts with exceptionally large numbers of incidents and to make differences among districts more visible.

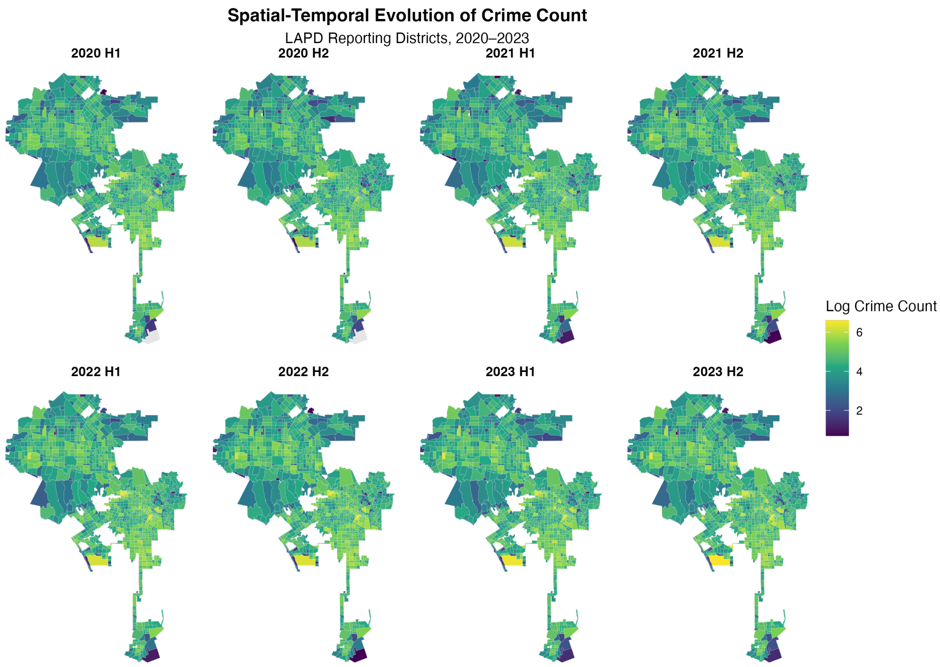

Figure 4.9: Spatio-temporal distribution of crime count across LAPD Reporting Districts in eight half-year periods from 2020 H1 through 2023 H2. Crime count is displayed on a log-transformed scale.

The maps show that the spatial distribution of crime volume is not completely constant over time. Differences in intensity can be observed across the eight periods, indicating that the number of reported incidents within individual Reporting Districts changes during the study period. At the same time, the maps retain a recognizable overall spatial structure across successive periods. This suggests that the geography of crime volume contains both persistent spatial differences and shorter term temporal variation.

Figure 4.10 presents the corresponding spatio-temporal maps for the normalized Crime Diversity Index. The same eight periods are used, allowing the evolution of crime composition to be compared directly with the evolution of crime volume.

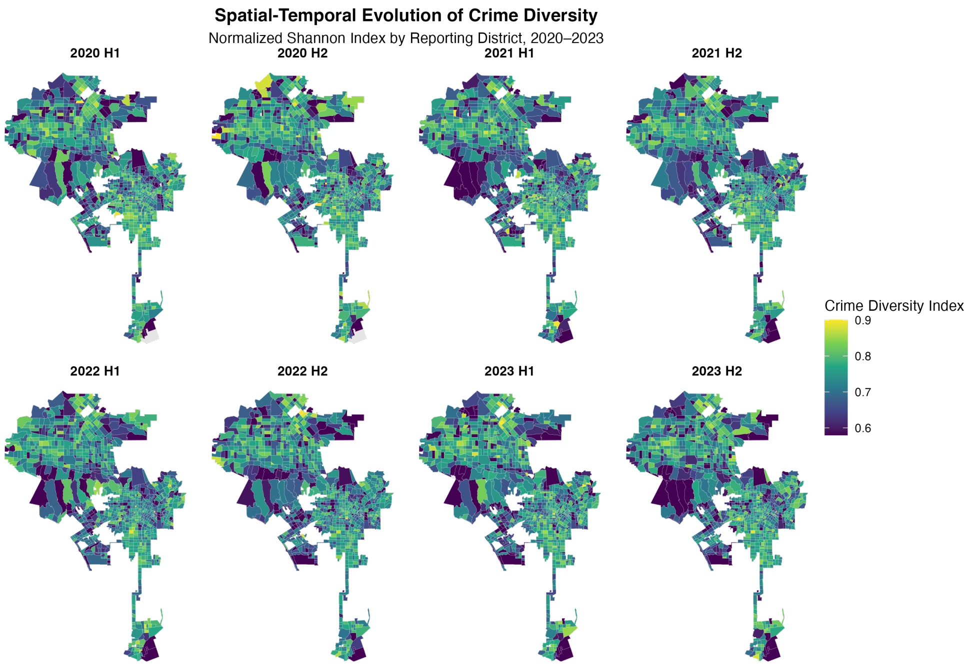

Figure 4.10: Spatio-temporal distribution of the normalized Crime Diversity Index across LAPD Reporting Districts in eight half-year periods from 2020 H1 through 2023 H2.

The crime diversity maps also show variation across both space and time. Many Reporting Districts maintain broadly similar levels of diversity across successive periods, while others exhibit visible changes in their relative diversity levels. The overall spatial pattern is therefore neither completely stable nor completely reorganized from one period to another. Instead, the maps suggest a combination of persistent geographic structure and gradual temporal variation in crime composition.

Comparing Figures 4.9 and 4.10 also illustrates that temporal changes in crime volume do not necessarily correspond to identical changes in crime diversity. A Reporting District may change in the number of reported incidents without experiencing a proportional change in the distribution of those incidents across crime categories. Conversely, changes in crime composition may occur even when the overall crime volume remains relatively similar. These comparisons reinforce the distinction between crime frequency and crime composition established in the previous sections.

The spatio-temporal results extend the findings from the aggregate spatial analysis in Section 4.3 and the temporal analysis in Section 4.4. The aggregate maps show that crime diversity exhibits significant spatial structure, while the monthly and seasonal analyses indicate relatively modest temporal variation at the overall level. The half-year maps show how these two dimensions operate jointly: the geographic pattern of crime diversity is broadly persistent, but local changes remain visible across time.

Overall, the analyses in this chapter show that crime diversity varies across LAPD Reporting Districts, exhibits significant spatial structure, and remains relatively stable over time despite modest monthly and seasonal variation. Crime diversity is strongly associated with crime volume in the unadjusted data, but the sensitivity analyses show that much of this association reflects unequal numbers of observed incidents. The two measures therefore remain related but describe different aspects of local crime patterns. The spatio-temporal analysis further shows that both measures exhibit a combination of persistent geographic structure and local change across the 2020–2023 study period. These empirical findings provide the basis for the statistical methodology and interpretation developed in the following chapter.

CHAPTER 5

Methodology

5.1 Methodological Framework and Contribution

The main methodological idea of this study is to describe local crime patterns from two perspectives: crime volume and crime composition. Conventional crime counts summarize how many reported incidents occur within a geographic unit, but they do not describe how those incidents are distributed across different types of crime. Two Reporting Districts (RDs), for example, may have similar crime counts while having substantially different crime compositions. One may be dominated by a small number of crime categories, whereas another may contain a more balanced mixture of crime types. Treating these two situations as equivalent can therefore hide an important dimension of local crime patterns.

To address this limitation, this study uses a normalized Shannon Diversity Index as the primary Crime Diversity Index (CDI). The CDI is not intended to replace conventional crime counts. Instead, the two measures describe different aspects of reported crime. Crime count represents the magnitude of crime, whereas the CDI represents the heterogeneity of its composition. Examining the two measures together therefore provides a joint volume–composition framework for describing local crime patterns.

The methodological contribution of this study is not the development of a new diversity index. Shannon-based diversity measures are well established and have previously been applied to crime. Instead, the contribution lies in applying the CDI within an integrated and reproducible spatial–temporal framework using LAPD Reporting Districts. Individual offenses are first consolidated into eight broad crime categories, after which their relative proportions are summarized using a normalized diversity measure. The use of Reporting

Districts provides a relatively fine operational spatial unit, allowing crime composition to be examined at a more localized scale than broader LAPD Areas.

The framework differs from approaches that consider crime diversity only as an isolated summary measure. In this study, the CDI is examined alongside conventional crime volume from several perspectives. The analysis considers its spatial distribution, temporal variation, relationship with crime count, and changes over time within the same framework. Looking at these dimensions together helps assess whether crime diversity captures features of local crime patterns that crime volume alone may not fully describe.

The framework contains four complementary analytical components. First, spatial analysis compares the geographic distributions of crime count and crime diversity across Reporting Districts and evaluates whether similar values are spatially clustered using Global Moran's I. Second, temporal analysis examines monthly trends and seasonal differences in crime diversity using descriptive summaries, one-way analysis of variance (ANOVA), and effect-size measures. Third, the relationship between crime volume and crime diversity is evaluated directly using correlation and ordinary least squares regression. This step is important because observed diversity can be associated with the number of recorded incidents, particularly in low-volume Reporting Districts. Finally, spatio-temporal analysis compares the geographic distributions of crime count and crime diversity across eight consecutive half-year periods from 2020 H1 through 2023 H2.

Additional sensitivity analyses are used to address three methodological concerns: unequal incident counts, residual spatial dependence, and repeated observations within Reporting Districts. These include fixed-count repeated subsampling, a spatial error model, and a random-intercept mixed-effects model, respectively.

In addition to these four components, a methodological robustness comparison is used to evaluate whether the observed diversity patterns depend strongly on the choice of diversity measure. The normalized Shannon CDI is compared with a normalized Simpson diversity index and crime richness. All three measures are constructed from the same eight crime categories and are compared using their relationships with crime volume and their spatial autocorrelation. This comparison provides numerical evidence for assessing whether the main patterns identified using the Shannon CDI remain similar under reasonable alternative measures. The construction and empirical comparison of these measures are described in Section 5.2.

These components are intended to function as an integrated analytical framework rather than as a collection of unrelated statistical procedures. Each method addresses a different aspect of the research questions. Mapping describes where crime diversity differs, spatial autocorrelation evaluates whether those differences exhibit geographic structure, monthly and seasonal analyses examine temporal variation, regression evaluates the relationship between diversity and crime volume, and the half-year maps show whether spatial patterns persist or change over time. The alternative diversity measures provide an additional robustness check on the primary Shannon-based CDI.

Figure 5.1 summarizes the primary analytical framework. The LAPD incident data are first cleaned and classified into eight broad crime categories. Crime count and crime diversity are then constructed at the spatial and temporal scales required for each analysis. Crime volume and crime composition are subsequently examined through spatial, temporal, count–diversity, and spatio-temporal analyses. The diversity component also includes comparison of the Shannon CDI with Simpson diversity and crime richness. Together, these analyses provide the evidence used to address the two supporting research questions and the main question of whether crime diversity provides additional information beyond conventional measures of crime volume.

The framework is also consistent with the broader perspective of environmental and opportunity-based criminology. Different locations may contain different combinations of activities, facilities, targets, and environmental conditions that support different types of crime. A more heterogeneous local environment may therefore be associated with a wider or more balanced mixture of reported crime categories. The present study does not directly measure these environmental mechanisms and therefore does not interpret the CDI causally. Instead, this theoretical perspective provides a conceptual basis for treating crime composition as a meaningful characteristic of place that can be examined alongside conventional

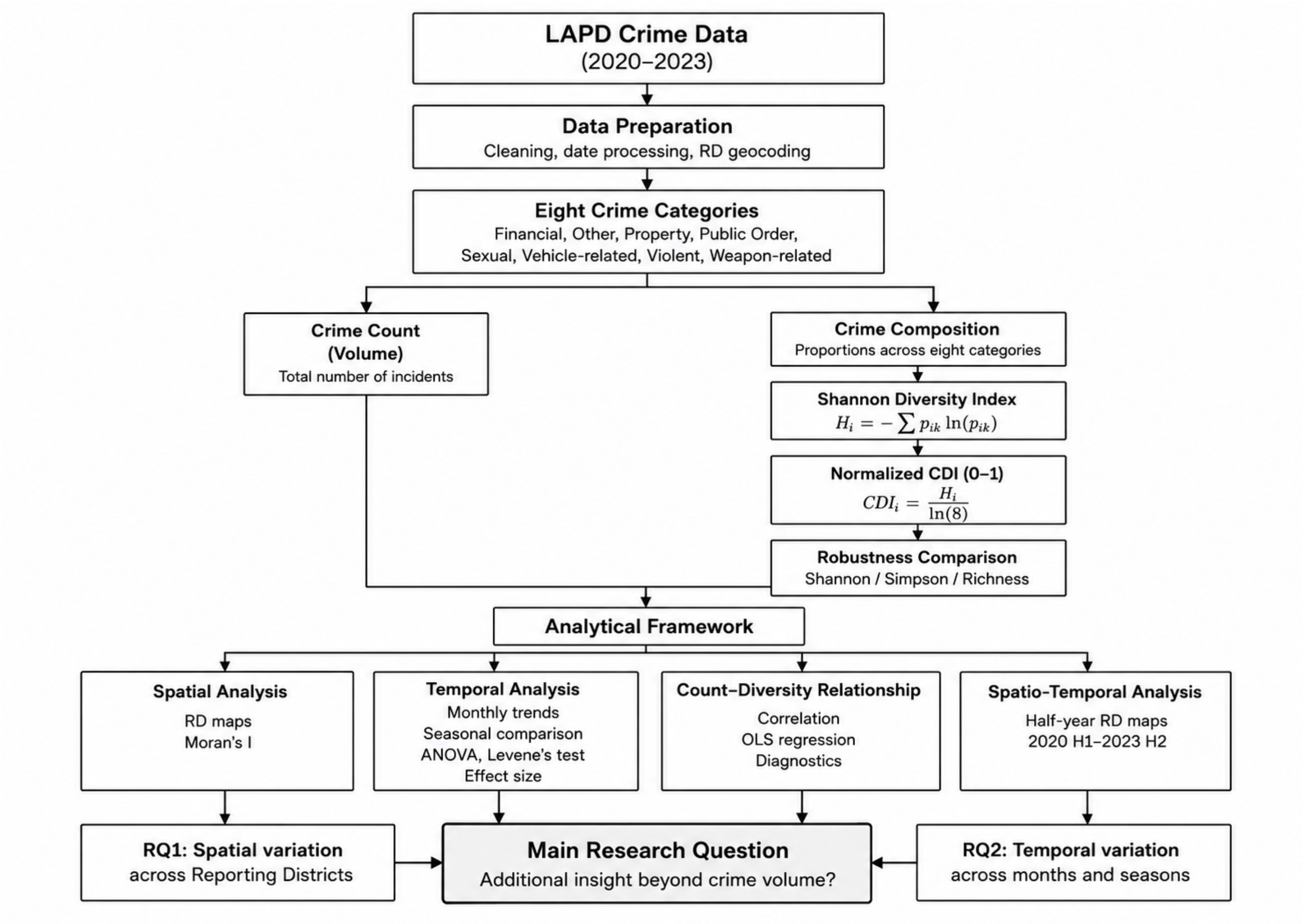

Figure 5.1: Methodological framework for the analysis of crime volume and crime diversity.

crime volume.

The objective of the framework is therefore not to demonstrate that the CDI is superior to traditional crime measures. Rather, it evaluates whether crime composition provides an additional descriptive dimension of local crime patterns. Evidence of variation across Reporting Districts, geographic clustering, temporal stability or change, an imperfect relationship with crime volume, and consistency across alternative diversity measures would support interpreting the CDI as a complementary measure rather than as a replacement for crime count.

5.2 Construction and Comparison of Crime Diversity Measures

The primary measure used in this study is the Crime Diversity Index (CDI), which summarizes the composition of reported crime within a given spatial or temporal unit. The purpose of the index is to distinguish between the volume of crime and the distribution of crime types. Whereas total crime count increases whenever additional incidents are recorded, crime diversity increases when incidents are distributed more evenly across different crime categories. The two measures therefore describe different characteristics of the same local crime pattern.

Crime categorization. The original LAPD data contain a large number of individual crime descriptions. Directly treating every description as a separate category would produce a highly fragmented measure in which rare offenses could have disproportionate influence on the number of crime types observed. To make the measure easier to interpret and more consistent across the analysis, the individual offenses were grouped into eight broader crime categories: Financial Crime, Other Crime, Property Crime, Public Order Crime, Sexual Crime, Vehicle-related Crime, Violent Crime, and Weapon-related Crime. The same classification rule was applied to all observations throughout the 2020–2023 study period.

This categorization represents a compromise between detail and interpretability. Using very broad categories would remove meaningful differences in crime composition, whereas retaining every individual LAPD offense description would make diversity highly sensitive to rare offense labels and the classification system used by the police. The eight-category structure preserves major differences among forms of reported crime while providing a common composition for comparison across Reporting Districts (RDs) and time periods.

Shannon diversity. Let n_{ik} denote the number of incidents belonging to crime category k in observational unit i, where $k = 1, \ldots, K$ and $K = 8$ in this study. The total number of crimes in unit i is

$$N_i = \sum_{k=1}^{K} n_{ik}.$$

The proportion of crime belonging to category k is then

$$p_{ik} = \frac{n_{ik}}{N_i}.$$

Crime diversity for observational unit i is measured using the Shannon Diversity Index,

$$H_i = -\sum_{k=1}^{K} p_{ik} \ln(p_{ik}),$$

where terms for which $p_{ik} = 0$ are defined to contribute zero to the summation. The index depends on both the number of categories represented and their relative frequencies. If nearly all incidents belong to one category, H_i will be relatively low. If incidents are distributed more evenly across several categories, H_i will be higher.

The maximum value of the Shannon index occurs when all K categories have equal proportions. Its theoretical maximum is therefore

$$H_{\max} = \ln(K).$$

Because eight crime categories are used in this study, the maximum possible raw Shannon value is $\ln(8)$. To provide a common and interpretable scale, the index is normalized as

$$CDI_i = \frac{H_i}{\ln(K)}.$$

The resulting CDI ranges theoretically from 0 to 1. A value approaching 0 indicates that reported crime is highly concentrated in a small number of categories, whereas a value approaching 1 indicates that incidents are distributed relatively evenly across the eight categories. Importantly, a high CDI should not be interpreted as a high level of crime or as an indication that an area is more dangerous. It indicates only greater heterogeneity in the composition of reported crime.

Alternative diversity measures. To evaluate whether the empirical findings depend strongly on the use of the Shannon index, two alternative measures were also calculated: Simpson diversity and crime richness.

The Simpson diversity index for observational unit i is defined as

$$S_i = 1 - \sum_{k=1}^{K} p_{ik}^2.$$

Like the Shannon index, Simpson diversity uses the relative proportions of all crime categories, although it places greater emphasis on more common categories. When all eight categories are equally represented, the maximum value of S_i is

$$1 - \frac{1}{K}.$$

The Simpson index was therefore normalized as

$$S_i^* = \frac{S_i}{1 - 1/K},$$

giving a theoretical range from 0 to 1 and making its scale directly comparable with the normalized Shannon CDI.

Crime richness provides a simpler alternative. It is defined as the number of crime categories represented by at least one incident:

$$R_i = \sum_{k=1}^{K} I(n_{ik} > 0),$$

where $I(\cdot)$ is an indicator function. With eight crime categories, richness ranges from 1 to 8 for observational units containing at least one incident. Unlike Shannon and Simpson diversity, richness does not consider the relative frequencies of the categories. Two RDs can therefore have the same richness value even if their crime compositions are very different.

Numerical comparison of the measures. The three measures were compared empirically at the Reporting District level in order to assess the robustness of the Shannon-based CDI. Table 5.1 summarizes their average values, their correlations with log-transformed crime count, and their spatial autocorrelation measured using Global Moran's I.

Measure	Mean	Correlation with log crime count	Moran's I
Normalized Shannon CDI	0.717	0.801	0.286
Normalized Simpson	0.808	0.777	0.325
Crime richness	7.533	0.890	0.071

Table 5.1: Comparison of crime diversity measures across LAPD Reporting Districts.

The normalized Shannon and normalized Simpson indices were extremely strongly correlated ($r = 0.986$), indicating that the two measures produced very similar assessments of crime diversity across Reporting Districts. Shannon diversity was also strongly correlated with crime richness ($r = 0.924$), although the relationship was less pronounced than that between Shannon and Simpson.

The comparison with crime volume provides an additional distinction. The correlation between the normalized Shannon CDI and log-transformed crime count was $r = 0.801$, while the corresponding correlation for normalized Simpson diversity was $r = 0.777$. Crime richness had a stronger relationship with crime volume, with $r = 0.890$. This suggests that simple richness is more strongly influenced by the number of recorded incidents than either of the two evenness-sensitive diversity indices.

The distribution of richness further illustrates this limitation. The mean crime richness was 7.533 categories, while the median, first quartile, third quartile, and maximum were all equal to 8. Thus, most Reporting Districts contained all eight broad crime categories. Once an RD reaches the maximum richness of eight, richness cannot distinguish between a district in which the eight categories are relatively balanced and a district in which one or two categories account for most incidents. The Shannon and Simpson indices retain this information because they incorporate the relative proportions of the categories.

The spatial comparison produced a similar pattern. Global Moran's I was 0.286 for the normalized Shannon CDI and 0.325 for normalized Simpson diversity. Monte Carlo permutation tests with 999 simulations produced $p = 0.001$ for both measures, indicating significant positive spatial autocorrelation. In contrast, crime richness had a much smaller Moran's I of 0.071, although the corresponding permutation test remained statistically significant ($p = 0.002$). Thus, Shannon and Simpson diversity both identify substantially stronger geographic structure than richness.

These comparisons do not imply that the Shannon index is universally superior to the alternative measures. Instead, these measures are used as robustness checks for the primary CDI. The very high correlation between Shannon and Simpson diversity, together with their similar positive spatial autocorrelation, suggests that the main pattern of crime diversity is not driven only by the choice of the Shannon formula. Richness, however, shows a stronger ceiling effect and a closer relationship with crime volume. This suggests that simply counting the number of crime categories provides less information about differences in crime composition in this dataset.

For these reasons, the normalized Shannon index is used as the main CDI in the rest of the analysis. Simpson diversity is included as an alternative measure to check whether the results remain similar, while crime richness provides a simpler benchmark. Crime count is kept separate because it measures the volume of reported crime rather than its composition.

5.3 Spatial and Temporal Analytical Methods

After constructing the Crime Diversity Index (CDI), the analysis examines how crime composition varies across space and time. These two parts correspond to the supporting research questions. The spatial analysis compares crime diversity across LAPD Reporting Districts (RDs) and tests whether districts with similar diversity levels tend to cluster geographically. The temporal analysis examines how crime diversity changes across months and seasons. Crime count is retained as a reference measure throughout the analysis so that patterns in crime composition can be interpreted alongside patterns in crime volume.

For the spatial analysis, crime incidents from 2020 through 2023 were first aggregated by RD. Total crime count and the normalized CDI were calculated for each district and joined to the LAPD Reporting District boundary data. Choropleth maps were then constructed for both measures using the same geographic units. Because crime counts are strongly right-skewed, the crime-count map uses log(1 + CrimeCount) to reduce the visual influence of districts with exceptionally large numbers of incidents. The CDI map displays the normalized Shannon index directly. These maps provide a descriptive comparison between the spatial distributions of crime volume and crime composition.

Spatial dependence was subsequently evaluated using Global Moran's I. Geographic observations may not be statistically independent because neighboring districts can share physical, social, land-use, or crime-opportunity characteristics. Global Moran's I provides a summary measure of whether RDs with similar values tend to occur near one another across the study area.

For n Reporting Districts, Global Moran's I can be expressed as

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2},$$

where x_i and x_j denote the value of the variable for RDs i and j, $\bar{x}$ is the overall mean, and w_{ij} represents the spatial relationship between the two districts. Neighboring relationships were defined using Queen contiguity, under which two polygons are treated as neighbors when they share either a boundary or a vertex. Row-standardized spatial weights were used in the calculation.

A positive Moran's I indicates that districts with similar values tend to be geographically adjacent, whereas a negative value indicates that neighboring districts tend to have dissimilar values. Values close to the expectation under spatial randomness indicate relatively little global spatial structure. Statistical significance was evaluated using a Monte Carlo permutation procedure. Moran's I was calculated separately for log-transformed crime count and the normalized Shannon CDI. For the methodological robustness comparison described in Section 5.2, the same spatial weights and permutation procedure were also applied to normalized Simpson diversity and crime richness.

Temporal variation was examined at both monthly and seasonal scales. Crime incidents were aggregated by RD and month, and the CDI was recalculated from the crime-category composition of each RD-month observation. This procedure allows the relative distribution of crime categories within an RD to change over time rather than assigning a single full-period diversity value to every month. Monthly average CDI values across RDs were then used to describe the overall temporal pattern between 2020 and 2023.

For the seasonal analysis, months were grouped into four seasons: Winter, Spring, Summer, and Fall. The distributions of RD-month CDI values were first compared graphically using boxplots and seasonal means with confidence intervals. A one-way analysis of variance (ANOVA) was then used to evaluate whether mean crime diversity differed across the four seasons. The null hypothesis was

$$H_0 : \mu_{\text{Winter}} = \mu_{\text{Spring}} = \mu_{\text{Summer}} = \mu_{\text{Fall}},$$

with the alternative hypothesis that at least one seasonal mean differed from the others. A significance level of $\alpha = 0.05$ was used as the conventional reference for statistical significance.

Because statistical significance does not necessarily indicate substantive importance, the seasonal analysis also reports an effect-size measure. This is particularly important in the present study because the large number of RD-month observations can make relatively small mean differences statistically detectable. The ANOVA result is therefore interpreted together with the magnitude of the estimated effect rather than from the p-value alone.

The homogeneity-of-variance assumption associated with conventional one-way ANOVA was assessed using Levene's test. Evidence of unequal variances was treated as a limitation when interpreting the ANOVA results. In addition, the RD-month data contain repeated observations from the same Reporting Districts over time. A conventional one-way ANOVA does not explicitly account for this within-district dependence.

To address this issue, a linear mixed-effects model was fitted as a sensitivity analysis. Season was included as a fixed effect, while Reporting District was included as a random intercept:

$$CDI_{it} = \beta_0 + \beta_1 I(\text{Spring}_{it}) + \beta_2 I(\text{Summer}_{it}) + \beta_3 I(\text{Fall}_{it}) + b_i + \epsilon_{it},$$

where CDI_{it} denotes the normalized Crime Diversity Index for Reporting District i at month t, Winter serves as the reference season, b_i is the Reporting District-specific random intercept, and ϵ_{it} is the within-district residual term. The random intercept allows each Reporting District to have its own baseline level of crime diversity. This helps account for the repeated monthly observations within the same geographic unit.

The mixed-effects model is used as a robustness check rather than as a replacement for the descriptive seasonal analysis. The conventional ANOVA provides a simple comparison of seasonal means and an easily interpretable effect-size estimate. The mixed-effects model then checks whether the seasonal pattern remains after accounting for persistent differences among Reporting Districts. Similar conclusions from both approaches strengthen the interpretation of the seasonal findings.

The spatial and temporal methods play complementary roles within the overall analytical framework. Mapping describes where crime diversity differs, Global Moran's I evaluates whether those differences exhibit geographic structure, monthly analysis describes how average diversity changes over time, and seasonal analysis evaluates whether recurring calendar periods are associated with systematic differences in diversity. The mixed-effects sensitivity analysis further accounts for repeated observations within Reporting Districts. Together, these methods address the spatial and temporal dimensions of crime composition while maintaining the distinction between crime diversity and crime volume.

5.4 Crime Count–Diversity Relationship and Spatio-Temporal Analysis

A central objective of this study is to determine whether the Crime Diversity Index (CDI) provides information about local crime patterns beyond that conveyed by crime counts. The existence of two different measures alone does not establish that they contain different information. In particular, diversity estimates may be related to the number of observed incidents because larger samples provide more opportunities for multiple crime categories to be represented. The relationship between crime volume and crime diversity must therefore be examined directly.

The association between total crime count and the normalized CDI was first evaluated graphically using a scatterplot and statistically using Pearson's correlation coefficient. Because crime count is strongly right-skewed across Reporting Districts (RDs), the count variable was represented using $\log(1 + \text{CrimeCount})$. For paired observations (X_i, Y_i), where X_i represents log-transformed crime count and Y_i represents the CDI, Pearson's correlation coefficient is defined as

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}.$$

The coefficient ranges from -1 to 1 and summarizes the direction and strength of the linear association between the two variables. A positive correlation indicates that RDs with greater crime volume tend to exhibit greater observed crime diversity. However, even a strong correlation does not imply that the two measures are interchangeable, because crime count describes incident frequency whereas the CDI is calculated from the relative distribution of incidents across crime categories.

To examine this relationship further, an ordinary least squares regression model was fitted. The normalized CDI was used as the response variable, while log-transformed crime count was used as the explanatory variable. The model is

$$CDI_i = \beta_0 + \beta_1 \log(1 + \text{CrimeCount}_i) + \epsilon_i,$$

where β_0 is the intercept, β_1 represents the estimated association between crime volume and crime diversity, and ϵ_i represents variation in CDI not accounted for by the fitted crime-count relationship.

The regression is used as an explanatory and descriptive diagnostic rather than as a predictive model. The objective is not to predict future CDI values or establish a causal effect of crime volume on diversity. Instead, the fitted relationship provides a quantitative assessment of how strongly the two measures are associated and how much cross-district variation in observed CDI is summarized by differences in crime volume. Regression diagnostics were examined to evaluate the adequacy of the linear specification, and the results are interpreted together with the scatterplot and correlation coefficient.

This interpretation is particularly important because a positive relationship between count and diversity is expected to some extent. RDs containing very few recorded incidents have fewer opportunities for all eight crime categories to be represented, whereas higher-volume districts provide a larger sample from which a broader crime composition can be observed. For this reason, the analysis includes a fixed-count subsampling sensitivity check to evaluate how strongly the observed CDI pattern depends on unequal incident counts.

For the subsampling analysis, only RDs with at least 100 reported incidents were retained. Within each eligible RD, exactly 100 incidents were sampled without replacement and the normalized Shannon CDI was recalculated from the sampled crime-category proportions. This procedure was repeated 50 times. For each RD, the mean of the 50 recalculated CDI values was used as a count-adjusted diversity estimate. Using the same number of incidents for every district removes the direct mechanical advantage that higher-volume RDs have in representing a larger number of crime categories.

The count-adjusted CDI was then compared with both the original CDI and the original crime volume. A strong association between the original and count-adjusted CDI would indicate that the broad diversity pattern remains stable after equalizing the number of observed incidents. In contrast, a weak relationship between the count-adjusted CDI and original crime volume would suggest that much of the original count–diversity association is driven by unequal sample sizes rather than by a persistent link between crime composition and total incident frequency.

A second methodological concern is spatial dependence in the regression errors. Because crime diversity exhibits significant spatial autocorrelation, the residuals from the ordinary least squares regression may not be spatially independent. To evaluate this possibility, Global Moran's I was calculated for the OLS residuals using the same Queen-contiguity spatial weights described in Section 5.3.

When significant residual spatial autocorrelation was detected, a spatial error regression was fitted as a sensitivity analysis. The spatial error model can be written as

$$CDI_i = \beta_0 + \beta_1 \log(1 + \text{CrimeCount}_i) + u_i,$$

with

$$u = \lambda W u + \varepsilon,$$

where W is the row-standardized spatial weights matrix, λ represents the strength of spatial dependence in the error process, and ε is the remaining random error. This specification does not assume that crime diversity in one RD directly causes diversity in neighboring RDs. Instead, it allows unobserved spatially structured factors to induce correlation among regression errors.

The spatial error model is used as a robustness check on the simple OLS relationship. The main questions are whether the relationship between log-transformed crime count and the CDI remains positive after accounting for spatial error dependence, whether the spatial error parameter is significantly different from zero, and whether Moran's I for the residuals is substantially reduced after fitting the spatial model. Model fit is also compared using the Akaike Information Criterion (AIC), where lower values indicate better relative fit among models estimated on the same spatial sample.

The analysis then extends from the aggregate count–diversity relationship to a joint spatio-temporal comparison. The 2020–2023 study period was divided into eight consecutive half-year intervals: 2020 H1, 2020 H2, 2021 H1, 2021 H2, 2022 H1, 2022 H2, 2023 H1, and 2023 H2. Crime count and the normalized CDI were recalculated for each RD within each half-year period.

The half-year interval was chosen to balance temporal detail with the stability of the diversity measure. Shorter periods, such as individual months, would capture more detailed changes over time, but they would also produce more RD-period observations with very small crime counts. Low incident counts can limit the number of crime categories that appear and make the CDI less stable. Grouping the data into half-year periods therefore provides a more practical way to examine how spatial crime composition changes over time.

Separate spatio-temporal maps were created for crime count and the CDI. The same Reporting District boundaries were used across all periods, and each measure kept a consistent scale over time. This makes the maps easier to compare and helps ensure that visible differences reflect changes in the data rather than changes in map scaling. Crime count was displayed using $\log(1 + \text{CrimeCount})$, while the CDI was displayed on the same normalized diversity scale across all eight periods.

This part of the analysis combines the spatial and temporal dimensions rather than treating them separately. Citywide monthly averages can hide local changes because increases in some Reporting Districts may be offset by decreases in others. Similarly, a single map covering the full four-year period may not show how the geographic pattern of crime composition changes over time. The half-year maps are therefore used to examine whether local spatial patterns remain relatively stable or shift across different periods.

The spatio-temporal analysis is descriptive rather than a formal space-time statistical model. It does not estimate specific spatial–temporal interaction effects, nor is it intended for prediction or causal inference. Instead, it is used to see whether the broad geographic patterns found in the overall spatial analysis remain visible across shorter time periods and whether changes in crime volume are accompanied by similar changes in crime diversity.

Taken together, the count–diversity and spatio-temporal analyses provide two complementary ways of evaluating the CDI. Correlation and OLS regression summarize the observed relationship between crime volume and diversity, while the fixed-count repeated subsampling analysis checks how sensitive that relationship is to differences in the number of incidents across Reporting Districts. Residual Moran's I and the spatial error model assess whether the regression relationship is affected by unmodeled spatial dependence. The half-year analysis then compares how crime composition changes geographically over time with the corresponding changes in crime count, allowing the study to see whether the two measures show similar or different spatial patterns.

The CDI is therefore useful not because it is completely separate from crime volume, but because it captures a different feature of reported crime. The CDI remains meaningful after checking its sensitivity to unequal crime counts and spatial dependence in the model residuals. These additional checks support using the CDI as a complementary measure, while crime count remains an important reference for describing the overall volume of reported crime.

5.5 Reproducibility and Software

All data preparation, statistical and spatial analyses, and visualizations were carried out in R. The workflow was divided into separate stages for data cleaning, statistical analysis, and figure generation. This structure allows the main results to be reproduced from the original LAPD crime records and the Reporting District boundary data.

The reproducible workflow consists of three primary R scripts. The first script, `01_data_cleaning.R`, imports and cleans the original crime records, restricts the analysis to the 2020–2023 study period, constructs the required temporal variables, and applies the eight-category crime classification. The second script, `02_analysis.R`, performs the principal statistical calculations, including construction of the Crime Diversity Index, Reporting District and temporal aggregation, alternative diversity-measure comparisons, spatial autocorrelation analysis, seasonal statistical tests, the random-intercept mixed-effects sensitivity analysis, the count–diversity correlation and regression analysis, fixed-count repeated subsampling, and the spatial error regression. The third script, `03_visualization.R`, generates the maps and statistical graphics used to present the results. A master script, `run_all.R`, executes these components in sequence to reproduce the analytical workflow.

Several R packages were used for specific components of the analysis. Data manipulation and aggregation were primarily performed using `tidyverse`, while `lubridate` was used to construct and manage temporal variables. Geographic boundary data and spatial joins were handled using `sf`. Global Moran's I, spatial weights, and permutation-based spatial autocorrelation analyses were implemented using `spdep`, while the spatial error regression was fitted using `spatialreg`. Seasonal variance diagnostics used `car`, and effect-size calculations used `effectsize`. The random-intercept mixed-effects model used `nlme`. Figures were constructed primarily with `ggplot2`, `patchwork`, and `viridis`.

The sensitivity analyses were also implemented programmatically within the reproducible workflow. For the fixed-count subsampling analysis, Reporting Districts with at least 100 incidents were identified and exactly 100 incidents were randomly sampled without replacement within each eligible district. This procedure was repeated 50 times using a fixed random seed, and the resulting CDI estimates were averaged within each Reporting District. The use of a fixed seed ensures that the reported subsampling results can be reproduced consistently.

For the spatial sensitivity analysis, the ordinary least squares regression was first fitted to the Reporting Districts with matched geographic boundaries. Moran's I was then calculated for the OLS residuals using the same Queen-contiguity spatial weights applied in the main spatial analysis. The spatial error model was subsequently fitted using the same spatial sample and weights structure, and residual spatial autocorrelation was evaluated again after model fitting. This sequence allows the reduction in remaining spatial dependence to be reproduced directly.

For the seasonal repeated-measures sensitivity analysis, a linear mixed-effects model was fitted using season as a fixed effect and Reporting District as a random intercept. The model used the same Reporting District-month data as the main seasonal analysis. Using both the conventional ANOVA and the mixed-effects model makes it easier to compare the results. It also helps check whether the seasonal findings remain consistent after accounting for repeated observations within Reporting Districts.

Intermediate processed datasets and final analytical tables were generated programmatically rather than edited by hand. The crime-category mapping is stored in a separate input file, making it possible to review and reproduce how the original LAPD offense descriptions were grouped. Tables and figures are saved in separate output folders, which keeps the raw data, processed data, code, and final analytical results clearly organized.

The complete analysis code is available in the GitHub repository that accompanies this thesis. The original LAPD crime data and Reporting District boundary files are not stored in the repository because they come from public external sources. Instead, the repository provides instructions for downloading these files and placing them in the correct folders. Once the source data are added, the master script can run the analysis in sequence and reproduce the processed datasets, statistical results, sensitivity analyses, and figures used in the thesis.

This reproducible structure is an important component of the methodology because it makes the construction of the CDI and the subsequent analyses transparent and verifiable. In particular, the separation of crime classification, index construction, statistical analysis, sensitivity analysis, and visualization reduces reliance on manual processing and provides a consistent implementation of the methodological framework described in this chapter.

CHAPTER 6

Conclusion

6.1 Main Findings

The primary objective of this study was to examine whether a Crime Diversity Index (CDI) can provide additional insights into local crime patterns beyond traditional measures of crime volume. Using 875,087 crime incidents reported by the Los Angeles Police Department between January 2020 and December 2023, the study measured crime composition using a normalized Shannon Diversity Index and examined its spatial and temporal variation across LAPD Reporting Districts. The analysis considered crime volume and crime diversity together through Reporting District-level comparisons, spatial autocorrelation, monthly and seasonal analysis, correlation and regression, spatio-temporal comparisons across eight half-year periods, and several sensitivity analyses designed to evaluate count dependence, spatial dependence, and repeated observations over time.

The first research question examined how crime diversity varies across LAPD Reporting Districts. The results show substantial differences in crime diversity across districts. At the full-period Reporting District level, the normalized CDI had a mean of 0.717 and a median of 0.762, with values ranging from 0 to 0.880. Most Reporting Districts exhibited relatively high diversity, indicating that reported crime was generally distributed across several crime categories. However, a smaller number of districts showed much lower diversity. Some of these districts also contained very few reported incidents, demonstrating that crime volume is an important consideration when interpreting low diversity values.

Crime diversity also exhibited meaningful geographic structure. Global Moran's I was 0.2862 for the normalized CDI, with a permutation p-value of 0.001, indicating statistically significant positive spatial autocorrelation. Neighboring Reporting Districts therefore tended to have more similar crime diversity values than would be expected under spatial randomness. Crime count exhibited a similar degree of spatial autocorrelation, with Moran's $I = 0.2925$. Nevertheless, the spatial maps showed that the geographic distributions of crime volume and crime diversity were not identical. This suggests that crime composition represents a spatial characteristic of local crime patterns that is related to, but not completely represented by, crime volume.

The relationship between crime count and diversity provides an important qualification to this conclusion. The Pearson correlation between log-transformed crime count and normalized CDI was strong and positive ($r = 0.801$, 95% CI: 0.780–0.820, $p < 0.001$). The corresponding ordinary least squares regression produced an R^2 of approximately 0.641. Thus, Reporting Districts with more recorded crimes generally exhibited greater observed crime diversity, and crime volume accounted for a substantial portion of the cross-district variation in the CDI. However, the relationship was not perfect, and Reporting Districts with similar crime volumes could still have different crime compositions.

Because empirical diversity can be mechanically affected by the number of observed incidents, this relationship was further evaluated using fixed-count repeated subsampling. Among the 1,085 Reporting Districts with at least 100 incidents, exactly 100 crimes were sampled from each district, the CDI was recalculated, and the procedure was repeated 50 times. The mean count-adjusted CDI remained extremely strongly correlated with the original CDI ($r = 0.995$, $p < 0.001$). In contrast, its correlation with the original log-transformed crime count was only $r = 0.018$ and was not statistically significant ($p = 0.562$). These results indicate that much of the strong observed relationship between crime count and the unadjusted CDI reflects unequal numbers of observed incidents. Importantly, however, the broad cross-district pattern of diversity remained highly stable after the incident count was held constant.

The ordinary least squares model also left substantial spatial structure in its residuals. For the 1,135 Reporting Districts with matched spatial boundaries, Global Moran's I for the OLS residuals was 0.266, with a permutation p-value of 0.001. This indicates that neighboring districts continued to have similar unexplained CDI values even after the linear association with crime count was taken into account. A spatial error model was therefore fitted as an additional sensitivity analysis. The estimated spatial error parameter was $\lambda = 0.482$ and was statistically significant ($p < 0.001$), while the coefficient for log-transformed crime count remained positive and statistically significant. The spatial error model also provided a lower AIC than the corresponding OLS model (-3252.4 versus -3093.5). After accounting for spatial error dependence, residual Moran's I decreased to -0.029 and was no longer statistically significant (permutation $p = 0.960$). These results show that spatial dependence is an important feature of the count–diversity relationship and that the spatial error specification substantially reduces the remaining global residual structure.

As a separate robustness check, the normalized Shannon CDI was compared with normalized Simpson diversity and crime richness. The Shannon and Simpson measures were extremely strongly correlated ($r = 0.986$), and both exhibited statistically significant positive spatial autocorrelation. Global Moran's I was 0.286 for the Shannon CDI and 0.325 for Simpson diversity, with permutation $p = 0.001$ for both measures. In contrast, crime richness showed a stronger relationship with log-transformed crime count ($r = 0.890$) and a much weaker degree of spatial autocorrelation (Moran's $I = 0.071$, $p = 0.002$). Richness also exhibited a substantial ceiling effect, with the median and both quartiles equal to the maximum value of eight categories. These comparisons indicate that the main characterization of crime diversity is robust to the use of another evenness-sensitive diversity measure, while simple category richness provides less differentiation among Reporting Districts in this application.

The second research question focuses on how crime diversity changes over time across months and seasons. Monthly average CDI values varied throughout the four-year study period, but there was no clear long-term increase or decrease. Seasonal differences were statistically detectable, with mean normalized CDI values of 0.569 in Winter, 0.576 in Spring, 0.584 in Summer, and 0.579 in Fall. The one-way ANOVA produced $F(3, 53249) = 15.48$ and $p < 0.001$. However, the effect size was extremely small ($\eta^2 = 0.000872$), indicating that season accounted for less than 0.1% of the observed variation in crime diversity.

Because each Reporting District appears in the data across multiple months, a random-intercept mixed-effects model was also fitted to account for repeated measurements. The overall seasonal effect remained statistically significant, $F(3, 52050) = 33.78$, $p < 0.0001$. Relative to Winter, the estimated CDI differences were 0.0076 for Spring, 0.0163 for Summer, and 0.0103 for Fall. Even after accounting for repeated observations within Reporting Districts, the seasonal effects remained statistically detectable, although they were still small in size. The standard deviation of the Reporting District random intercept was 0.168, compared with a residual standard deviation of 0.134, showing that persistent differences between districts were substantial. Overall, the mixed-effects analysis supports the same conclusion as the earlier seasonal analysis: seasonal variation is present, but it is relatively modest compared with the broader geographic differences across Reporting Districts.

The spatio-temporal analysis brings the spatial and temporal results together. Across the eight half-year periods from 2020 H1 to 2023 H2, the overall geographic patterns of both crime volume and crime diversity remained broadly similar, although some local shifts appeared from one period to another. Changes in crime count were not always matched by similar changes in crime diversity. The results therefore suggest a combination of persistent geographic structure and local temporal variation in crime composition.

Taken together, these findings provide a more qualified answer to the main research question. In the original data, crime diversity is strongly related to crime volume, which means the CDI should not be interpreted without considering how many incidents are observed. However, when the same number of incidents is repeatedly sampled from each eligible Reporting District, the overall diversity pattern across districts remains very stable, while its direct relationship with the original crime volume becomes much weaker. The spatial error analysis also shows that geographic dependence matters, but accounting for it does not remove the positive relationship between crime count and diversity. Similarly, the mixed-effects analysis shows that the small seasonal pattern remains after repeated observations within Reporting Districts are taken into account.

The CDI therefore captures an aspect of reported crime that is connected to crime volume but is not the same as crime volume: how incidents are distributed across different crime categories. The very similar results from the Shannon and Simpson measures also suggest that this pattern is not simply an artifact of using the Shannon index. Overall, the CDI offers a complementary view of local crime patterns, while the sensitivity analyses provide a more careful basis for interpreting how diversity relates to crime volume, spatial dependence, and temporal change.

6.2 Comparison with Previous Studies

The findings of this study are generally consistent with earlier work on crime diversity, while extending that research by examining spatial and temporal variation together across LAPD Reporting Districts. Previous studies have suggested that crime diversity captures an aspect of offending that is connected to, but not identical to, crime frequency. The results here support that idea, while also showing that the relationship between diversity and crime volume can be quite strong.

Previous work on crime diversity established the importance of considering the mixture of offenses within an area rather than relying exclusively on overall crime frequency. Related work has also shown that diversity and offending frequency are connected, but they should not be treated as the same thing. In this study, the strong positive relationship between crime count and diversity is consistent with the idea that areas with more incidents have more chances for different crime categories to appear. The fixed-count subsampling results support this interpretation: once the number of sampled incidents is kept the same across Reporting Districts, the relationship with original crime volume becomes negligible. Even so, the relative diversity pattern across districts remains highly stable after this adjustment. This reinforces the distinction between the two measures: crime count shows how much reported crime occurs, while the CDI shows how those incidents are distributed across crime categories.

The geographic variation found in this study is also consistent with earlier research showing that crime is unevenly distributed across space. Previous studies have shown that different crime types can follow different spatial patterns, and that combining them into a single total crime measure may hide important local differences. The significant positive spatial autocorrelation observed for the CDI in this study extends this perspective by showing that crime composition itself also has a geographic structure. This result is consistent with previous research examining crime diversity in relation to characteristics of the urban environment, which suggests that diversity is systematically distributed across space rather than occurring randomly. Although the present study does not attempt to identify the environmental mechanisms responsible for these patterns, its findings provide additional evidence that the geographic distribution of crime diversity contains information that is not fully captured by overall crime volume.

The temporal findings can also be considered in relation to previous research on crime seasonality and temporal variation. Earlier studies have shown that individual crime types can follow different temporal and seasonal patterns. Because the CDI depends on the relative distribution of crime categories, changes in the prevalence of individual crime types could also produce changes in overall crime diversity. In the present study, however, seasonal differences in the CDI were very small in magnitude despite being statistically significant. This suggests that temporal changes in individual crime categories do not necessarily translate into large changes in the overall diversity of crime composition.

The spatio-temporal analysis also follows earlier research that stresses the need to consider space and time together when studying crime. Citywide trends can hide very different local patterns, and changes over time may not occur in the same way across all areas. In this thesis, the eight half-year comparisons apply this idea descriptively to crime diversity. Rather than using a formal spatio-temporal prediction model, the analysis shows that the overall geographic pattern of crime diversity remained fairly recognizable across the study period, while some Reporting Districts still experienced local changes from one half-year period to another.

Overall, the findings support an important point from the existing literature: no single measure can fully describe local crime patterns. Crime counts describe how often crime occurs, harm-based measures focus on the seriousness of offenses, and diversity measures describe how crime is distributed across different categories. The contribution of this study is therefore not to replace existing crime measures, but to show how a normalized Crime Diversity Index can be examined alongside crime volume within the same spatial and temporal framework. Applying this approach to LAPD Reporting Districts provides additional evidence that crime composition can complement traditional measures when describing local crime patterns.

6.3 Practical Implications

The findings also have practical implications for how local crime statistics are presented and interpreted. Crime counts remain a simple and important measure of crime volume, and the results do not suggest replacing them. Instead, the Crime Diversity Index (CDI) can be used alongside crime counts to show more clearly how reported incidents are distributed across different crime categories within a Reporting District.

This distinction is especially important when comparing Reporting Districts. In the unadjusted data, crime count and the CDI are strongly positively associated, meaning that districts with more reported incidents also tend to show greater observed crime diversity. However, the fixed-count subsampling analysis shows that this association becomes negligible once the same number of incidents is sampled from each district. At the same time, the relative diversity pattern across Reporting Districts remains highly stable. This suggests that part of the original count–diversity relationship is due to differences in the number of recorded incidents rather than differences in crime composition alone. Even among districts with similar crime volumes, the distribution of incidents across crime categories can still differ. A district where crime is concentrated in only a few categories has a different crime profile from one where incidents are spread more evenly across several categories. Looking at crime volume and crime composition together therefore provides a more complete description of local crime conditions.

The spatial results also help explain how local crime patterns should be interpreted. The significant positive spatial autocorrelation in the CDI shows that neighboring Reporting Districts tend to have similar levels of crime diversity. This suggests that crime composition is not randomly distributed across Los Angeles. Broader geographic patterns may therefore provide useful context for understanding individual districts. The spatio-temporal analysis shows a similar pattern over time: the main geographic structure remained visible across the eight half-year periods, although some districts changed from one period to another. Considering both spatial structure and temporal change therefore provides context that a single citywide crime count cannot capture.

The temporal results also have a practical implication. Crime diversity differed significantly across seasons, but the size of the effect was very small. This shows that statistical significance does not necessarily mean that a difference is practically important, especially in a large dataset. In this study, the large number of observations made small seasonal differences detectable, but these differences explained very little of the overall variation in crime diversity. The more persistent differences across Reporting Districts therefore appear to be more informative than seasonal variation alone.

At the same time, the CDI should not be treated as a direct measure of crime severity, public safety, or policing need. A Reporting District with higher crime diversity does not necessarily face a more serious crime problem than one with lower diversity. The CDI describes how reported incidents are distributed across crime categories, while crime count measures the overall volume of crime. Other measures may be more appropriate when the goal is to capture the seriousness or harm of particular offenses. For practical use, the CDI is therefore best interpreted together with crime counts and other relevant information rather than on its own.

Finally, the framework used in this study is relatively easy to reproduce with routinely collected police incident data. Once individual offenses are grouped into consistent crime categories, the CDI can be calculated at different spatial and temporal scales and compared with conventional crime measures. This makes the framework suitable for use in other cities, as long as sufficiently detailed crime records and geographic data are available.

Overall, the practical value of the CDI is that it captures an aspect of local crime that traditional crime counts do not directly describe. Crime counts show how much reported crime occurs, while the CDI shows how those incidents are distributed across different crime categories. Considering both measures together can therefore provide a more complete, but still cautious, description of local crime patterns.

6.4 Limitations

Several limitations should be considered when interpreting the findings of this study.

First, this study relies on crimes officially recorded by the Los Angeles Police Department. These records do not include offenses that were never reported or detected, and reporting patterns may vary across locations and crime types. The Crime Diversity Index (CDI) should therefore be interpreted as a measure of recorded crime composition rather than as a complete representation of all criminal activity in Los Angeles.

Second, the construction of the CDI depends on the classification of individual LAPD offense descriptions into eight broader crime categories. This classification improves interpretability and reduces the influence of rare offense labels, but some detailed differences between individual offenses are necessarily lost during aggregation. The number and definition of categories may therefore affect the resulting diversity values. For example, a finer classification using approximately 15 categories rather than the eight categories adopted in this study could distinguish additional differences in local crime composition and may produce different CDI values, spatial patterns, or relationships with crime volume. Conversely, a smaller number of broader categories could reduce observed variation by combining offenses that have different underlying distributions. The present analysis does not systematically evaluate alternative crime classification schemes, so the findings should be interpreted as conditional on the eight-category framework used in this study. Future research could conduct sensitivity analyses using multiple levels of crime-category aggregation to examine how robust the spatial and temporal findings are to the choice of classification.

Third, empirical diversity estimates are influenced by the number of observed incidents. Reporting Districts with very small crime counts have fewer opportunities to contain multiple crime categories, which can mechanically limit the observed CDI. This issue was addressed in the present study through a fixed-count subsampling sensitivity analysis in which exactly 100 incidents were sampled from each eligible Reporting District and the CDI was recalculated over 50 repeated samples. The resulting count-adjusted CDI remained extremely strongly correlated with the original CDI, while its association with original crime volume was nearly zero. This provides evidence that the broad cross-district diversity pattern is not solely a consequence of unequal incident counts. However, the adjustment was limited to the 1,085 Reporting Districts with at least 100 incidents and depended on the selected fixed sample size. Reporting Districts with fewer than 100 incidents were excluded from this sensitivity analysis. Future work could examine alternative sample sizes or formal rarefaction and bias-correction approaches.

Fourth, spatial dependence remains an important methodological consideration. The ordinary least squares regression relating crime count to CDI produced spatially autocorrelated residuals, indicating that the independence assumption of the simple regression was not satisfied. A spatial error model was therefore fitted as a sensitivity analysis and substantially reduced the residual spatial autocorrelation. Although this provides a more appropriate treatment of global spatial dependence, the spatial error model remains a relatively simple specification. It does not identify the specific social, environmental, or land-use processes that generate the spatial structure, and it assumes that remaining spatial dependence can be represented through the chosen Queen-contiguity weights matrix. Alternative definitions of spatial relationships or more complex spatial models could yield somewhat different results.

Fifth, the seasonal analysis involves repeated observations of the same Reporting Districts across months. The conventional one-way ANOVA does not explicitly account for this within-district dependence. To address this concern, a random-intercept mixed-effects model was fitted as a sensitivity analysis, allowing each Reporting District to have its own baseline level of crime diversity. The seasonal effects remained statistically significant, but their magnitude was small, which supports the substantive conclusion from the original ANOVA. However, the mixed-effects model uses a relatively simple random-intercept structure. It does not account for more complex temporal dependence, such as autocorrelation between adjacent months or seasonal patterns that may differ across Reporting Districts. The seasonal findings should therefore continue to be interpreted primarily as modest descriptive temporal differences rather than as evidence of strong or causal seasonal effects.

Sixth, the ordinary least squares regression and the sensitivity models are explanatory and descriptive rather than causal. The regression diagnostics showed departures from ideal linear-model assumptions, particularly among low-volume Reporting Districts. Although the fixed-count subsampling and spatial error model address important sources of dependence, they do not establish that crime volume causes changes in crime diversity. Other characteristics of Reporting Districts may influence both crime volume and crime composition.

Finally, the spatio-temporal analysis remains descriptive rather than a formal space–time statistical model. The eight half-year maps allow persistent and changing geographic patterns to be compared visually, but they do not estimate explicit spatial–temporal interaction parameters or explain why individual Reporting Districts change over time. Similarly, this study does not incorporate socioeconomic, demographic, land-use, or environmental variables that could help explain the observed geographic differences. These factors are outside the scope of the present study but provide important directions for future research.

Despite these limitations, the additional sensitivity analyses strengthen the interpretation of the main findings. Fixed-count subsampling shows that the overall CDI pattern remains stable after incident counts are equalized. The spatial error model substantially reduces the remaining spatial dependence in the count–diversity regression. The mixed-effects model also confirms that seasonal differences remain small after repeated observations within Reporting Districts are taken into account. These checks do not remove all methodological limitations, but they provide a more cautious and robust basis for interpreting the CDI as a complementary measure of local crime composition rather than a replacement for conventional crime statistics.

6.5 Conclusion

This study examines whether a Crime Diversity Index (CDI) can provide additional insight into local crime patterns beyond traditional measures of crime volume. The analysis uses

875,087 crime incidents reported by the Los Angeles Police Department between January 2020 and December 2023. Crime composition is summarized using a normalized Shannon Diversity Index based on eight broad crime categories. The study then examines differences across LAPD Reporting Districts, changes across months and seasons, the relationship between crime volume and diversity, and changes in spatial patterns across eight half-year periods.

The results provide a clear answer to the first supporting research question. Crime diversity differed substantially across Reporting Districts and showed statistically significant positive spatial autocorrelation. Crime count and crime diversity had similar levels of overall spatial clustering, but their geographic patterns were not identical. The spatial error analysis also showed that the simple count–diversity regression left significant spatial structure in the residuals. After spatial dependence was included in the model, this remaining structure was greatly reduced. This indicates that geographic dependence plays an important role in the observed relationship between crime count and crime diversity.

For the second supporting research question, crime diversity remained relatively stable over time. Monthly averages changed across the four-year study period, but there was no clear sustained increase or decrease. Seasonal differences were statistically detectable, although the size of these differences was small. A random-intercept mixed-effects model reached the same general conclusion after accounting for repeated monthly observations within Reporting Districts. The eight half-year comparisons also showed that the broader geographic pattern of crime diversity remained recognizable over time. However, some individual districts still experienced local changes from one period to another.

The relationship between crime volume and crime diversity requires a more qualified interpretation. In the unadjusted data, log-transformed crime count was strongly positively associated with the normalized CDI. However, fixed-count repeated subsampling showed that this association became negligible when the number of sampled incidents was held constant. At the same time, the relative CDI pattern across Reporting Districts remained highly stable after this adjustment. These results indicate that unequal numbers of observed incidents account for an important part of the original count–diversity relationship, while the relative cross-district pattern of crime diversity remains highly stable after this adjustment.

The robustness comparison with alternative diversity measures provides additional support for this interpretation. The normalized Shannon and Simpson diversity indices produced very similar patterns across Reporting Districts and similar levels of spatial autocorrelation, whereas simple crime richness was more strongly associated with crime volume and showed a substantial ceiling effect. The main characterization of crime diversity therefore does not depend solely on the particular choice of the Shannon index.

Taken together, these results provide a qualified affirmative answer to the main research question. The Crime Diversity Index can provide additional insight into local crime patterns, but it should not be interpreted independently of crime volume or as a replacement for traditional crime measures. Crime count describes the amount of reported crime, whereas the CDI summarizes how those incidents are distributed across different crime categories. The sensitivity analyses show that this compositional pattern remains meaningful even after accounting for unequal incident counts, spatial dependence, and repeated observations over time.

The main contribution of this thesis is the use of an established diversity measure within an integrated and reproducible spatial–temporal framework at the LAPD Reporting District level. The framework brings together crime volume, crime composition, spatial structure, temporal variation, and several sensitivity checks. This provides a transparent way to describe local crime patterns from multiple perspectives. The CDI is not a measure of crime severity, public safety, or causal mechanisms. Instead, it provides a complementary description of reported crime composition and can contribute to a more complete understanding of local crime patterns in Los Angeles.

REFERENCES

[Bra16] P. Jeffrey Brantingham. "Crime Diversity." *Criminology*, **54**(4):553–586, 2016.

[BV25] P. Jeffrey Brantingham and Christophe Vandeviver. "Moving to Crime Opportunity." *Journal of Quantitative Criminology*, 2025.

[DD23] Yingjie Du and Ning Ding. "A Systematic Review of Multi-Scale Spatio-Temporal Crime Prediction Methods." *ISPRS International Journal of Geo-Information*, **12**(6):209, 2023.

[DK25] Onur Batın Doğan and Fatma Sevinç Kurnaz. "Crime in Proportions: Applying Compositional Data Analysis to European Crime Trends for 2022.", 2025.

[FH16] Brian Francis and Les Humphreys. "Investigating the Relationship Between the Diversity Index and Frequency of Offending." *Journal of Developmental and Life-Course Criminology*, **2**(4):397–416, 2016.

[Hip10] John R. Hipp. "A Dynamic View of Neighborhoods: The Reciprocal Relationship Between Crime and Neighborhood Structural Characteristics." *Social Problems*, **57**(2):205–230, 2010.

[HZD18] Tao Hu, Xinyan Zhu, Lian Duan, and Wei Guo. "Urban Crime Prediction Based on Spatio-Temporal Bayesian Model." *PLOS ONE*, **13**(10):e0206215, 2018.

[KCM21] Samira Khorshidi, Jeremy G. Carter, George Mohler, and George E. Tita. "Explaining Crime Diversity with Google Street View." *Journal of Criminal Justice*, **75**:101830, 2021.

[KVT23] O. M. Kunakh, A. M. Volkova, G. F. Tutova, and O. V. Zhukov. "Diversity of Diversity Indices: Which Diversity Measure Is Better?" *Biosystems Diversity*, **31**(2):131–146, 2023.

[MLP12] David McDowall, Colin Loftin, and Matthew Pate. "Seasonal Cycles in Crime, and Their Variability." *Journal of Quantitative Criminology*, **28**(3):389–410, 2012.

[OAN23] Takahiro Ohyama, Masayuki Amemiya, and Tomoki Nakaya. "Investigating Crime Harm Index in the Low and Downward Crime Contexts: A Spatio-Temporal Analysis of the Japanese Crime Harm Index." *Cities*, **132**:104090, 2023.

[Off24] Office of Mayor Karen Bass. "LAPD Releases 2024 End of Year Crime Statistics for the City of Los Angeles." `https://mayor.lacity.gov/news/lapd-releases-2024-end-year-crime-statistics-city-los-angeles`, 2024. Accessed July 2026.

[PMA15] Silas Pereira, Claudio M. Mota, Martin A. Andresen, and Luiz F. Matias. "Crime Concentrations and Similarities in Spatial Crime Patterns in a Brazilian Context." *Applied Geography*, **62**:314–324, 2015.

[Rat09] Jerry H. Ratcliffe. "Crime Mapping: Spatial and Temporal Challenges." In Alex R. Piquero and David Weisburd, editors, *Handbook of Quantitative Criminology*, pp. 5–24. Springer, New York, 2009.

[TBJ24] Lisa Tompson, Kate J. Bowers, and Shane D. Johnson. "Seasonal Characteristics of Crime: An Empirical Investigation of the Temporal Fluctuation of the Different Types of Crime in London." *Crime Prevention and Community Safety*, **26**(1):1–23, 2024.

[WS22] Andrew P. Wheeler and Wouter Steenbeek. "Weekly Crime Concentration." *Journal of Quantitative Criminology*, **38**(4):863–885, 2022.