THE GEOGRAPHIC CONNOTATIONS OF REINCARCERATION:
A SPATIAL ANALYSIS OF RECIDIVISM IN WASHINGTON STATE

by

Robert P. Woodmark

A Thesis Presented to the
FACULTY OF THE USC DORNSIFE COLLEGE OF LETTERS, ARTS AND SCIENCES
UNIVERSITY OF SOUTHERN CALIFORNIA
In Partial Fulfillment of the
Requirements for the Degree
MASTER OF SCIENCE
GEOGRAPHIC INFORMATION SCIENCE AND TECHNOLOGY

December 2023

Dedicated to the incarcerated individual who aspires to make right, and the formerly incarcerated individuals fulfilling the promises that we made to ourselves. Never forget to extend an arm to the ones coming after you

Acknowledgements

I am grateful to my advisor, Dr. Elisabeth Sedano, for pushing me to strive for my best. I would also like to acknowledge the Thesis Committee which included Dr. Robert Vos and Dr. An-Min Wu. Your expertise in your respective fields paid dividends to my success. Your assistance in this process is much appreciated.

Table of Contents

List of Tables

List of Figures

Abbreviations

ACS	American Community Survey
AUA	Areal unit of aggregation
BJS	Bureau of Justice Statistics
CGP	Criminal geographic profiling
df	Degrees of freedom
DOC	Department of Corrections
EDA	Exploratory data analysis
ESDA	Exploratory spatial data analysis
GIST	Geographic information science and technology
GLM	Generalized linear model
GWR	Geographically weighted regression
HBLR	Hierarchical binary logistic regression
ICE	Influence on conditional expectation
LISA	Local indicators of spatial association
MAUP	Modifiable Areal Unit Problem
NIJ	National Institute of Justice
OAA	Offender Accountability Act
RMI	Risk management identification
SPLOM	Scatterplot matrix
SWM	Spatial weights matrix
WDQ	Weighted displacement quotient
WADOC	Washington State Department of Corrections

Abstract

Recidivism rates have important implications for public safety, the well-being of both those reentering our communities, as well as the communities that the formerly incarcerated individuals are being released back into. This study leverages a comprehensive prison admission/release dataset from the Washington State Department of Corrections in a spatial analysis looking at both individual level and county contextual variables with the intent to identify whether the county of release of a formerly incarcerated individual is correlated to the recidivism rates of the county. The analysis further considered social disorganization theory aspects as potential contributors to recidivism patterns. By incorporating these variables into a comprehensive exploratory data analysis, subsequent statistical analyses, and finally exploratory spatial data analysis methodologies the study aimed to understand how the socioeconomic context of the county of release may lead to a propensity to recidivate in Washington State. The findings of this project show that there is no evidence of a correlation between the county of release and propensity to recidivate in the State from 2012-2022, the conclusion drawn here is a finding for the null hypothesis of this study, that while many counties display statistically significant deviations from the sample mean, these deviations are not attributable to the county of release itself. Furthermore, contrary to academic literature that has found significant correlations between social disorganization and crime, no statistically significant correlations with the socioeconomic contextual variables meant to reflect social disorganization were found in this analysis. This suggests that while rates of public assistance enrollment and literacy rates or educational attainment have been found to be correlated to crime and recidivism elsewhere, they are not in Washington State at the county scale of analysis. The ultimate conclusion underscores a critical concern regarding the selected scale of analysis, emphasizing that the

County, as an areal unit of aggregation, proves to be too broad to comprehensively capture the nuances of recidivism. This assertion gains robust support from the evidence revealed by the analysis' findings.

Chapter 1 Introduction

The primary intent of this study is to investigate the correlation between the county of release and recidivism outcome in Washington State. Employing an evidence-based spatial analysis approach at the county scale, the study engages in a comprehensive exploratory data analysis (EDA), and exploratory spatial data analysis (ESDA) to assess recidivism likelihood in Washington State. By leveraging and expanding upon existing research on recidivism in the fields of Criminal Justice and Criminology within Washington State (Aos 2003; Miller, Drake, and Nafziger 2013; Miller, Drake, and Aos 2006; Gagliardi et al. 2004; Beckett, Harris, and Evans 2008; Lovell, Johnson, and Cain 2007; Manchak, Skeem, and Douglas 2008; Phipps et al. 1999; Camasso and Jagannathan 1995), this project seeks to inform decision-making processes, targeted outreach efforts, and drive changes in public policy and discourse.

This research sheds light on the spatial patterns and potential contextual factors influencing recidivism through the use of variables that reflect social disorganization and give each county further context. To analyze these research goals, utilization of various statistical methods, including descriptive statistics, a Power Analysis, Shapiro-Wilks test, T-tests, Cohen's *d*, ANOVA, Moran's I Index for spatial autocorrelation, and Getis-Ord G_i^* hot spot analysis allowed this research to delve into the intricacies of recidivism trends and spatial patterns across counties. Additionally, the identification of hotspots and spatial autocorrelation were analyzed in a way that provides insights for informed decision-making regarding recidivism through targeted interventions in areas exhibiting higher susceptibility.

This study also builds on prior research assessing the relationship between the neighborhood context and the propensity to recidivate. Specifically, building on work by Hipp (2007), this study aims to investigate whether the county of release is correlated with a propensity to recidivate while recognizing the scale of analysis emerges as a pivotal consideration. Drawing inspiration from this past work, which delved into the neighborhood effects of crime at the block and tract levels, it is clear that selecting the most appropriate areal unit of analysis is not merely a methodological choice but also a theoretical one. Prior studies have primarily focused on more granular scales of analysis. However, some have examined the neighborhood effects of crime using combinations ranging from two tracts together (e.g., Logan and Stults 1999; Morenoff et al. 2001; Sampson et al. 1997) to aggregations of nine or ten tracts (e.g., Almgren et al. 1998; Bursik 1988; Heitgerd and Bursik 1987). A consistent finding across the referenced studies is the profound impact that the unit of analysis can have on our understanding of crime dynamics.

While it is reasonable to assume that the existence of heterogeneity across blocks and within tracts, as outlined in Hipp's work, may be masked at higher levels of aggregation for certain measures and theoretical inquiries, by adopting the county as the unit of analysis, the intent is to capture a broader spectrum of contextual variables that might shape an individual's journey post-release. Beyond the mere correlation between county of release and reoffending, this study seeks to understand how the diverse landscape of county-specific attributes—socio-economic indicators, educational opportunities, or even community cohesion—might serve in the future as predictors of recidivism. The county scale, thus, offers both a challenge and an opportunity. A challenge in ensuring that critical nuances are not overlooked, and an opportunity to understand recidivism in a broader socio-spatial framework (Hipp 2007).

Applying this understanding to the context of recidivism and county of release, it raises pertinent questions: How do counties, as broader units compared to tracts or neighborhoods, capture the myriad of factors influencing recidivism? Can the heterogeneities, often noted at smaller spatial scales, be suitably represented at the county level? Another focus of this research then, is to understand if the county scale of analysis accurately captures the spatial processes of recidivism, with the ultimate goal of determining whether the county serves as an appropriate areal unit of aggregation for the purposes of analyzing recidivism in Washington State.

1.1 Background

In 2019 alone, Washington State ranked in the top ten of all states for the number of total admission and releases from incarceration (Bureau of Justice Statistics 2019). Between 2012-2022 approximately 6,500 people were released from WADOC operated prisons per year. These statistics highlight the importance of creating a more evidence-based data-driven approach to reducing recidivism rates and promoting successful reentry in the state of Washington.

Reducing recidivism rates holds significant implications for public health, safety, crime rates, tax burdens, and various measurable social benefits. It contributes to the creation of more cohesive and crime-resilient neighborhoods (Drawve and McNeeley 2021; Hipp et al. 2010; Petersilia 2009; Wang et al. 2013). However, it's crucial to acknowledge that recidivism disproportionately impacts marginalized communities, often affecting them at higher rates than their more prosperous and socially cohesive counterparts (Kubrin et al. 2007; Shaw and McKay 1942).

In the context of Washington State, the Offender Accountability Act (OAA), enacted in 1999, mandates the consideration of recidivism potential in risk assessments. The Level of Service Inventory Revised (LSI-R) is a key tool used to determine sentencing guidelines for judges (Washington State Legislature, 1999; Aos, Drake, and Miller 2006; Drake 2014). Given that the majority of incarcerated individuals will eventually re-enter society, there is a compelling argument for prioritizing rehabilitation alongside confinement as part of a holistic approach to criminal justice. Reducing recidivism not only enhances equitable outcomes but also aligns with the broader goals of public safety and community well-being.

Traditional statistical methods that are widely used in the analysis of recidivism assume that the relationship between recidivism and the other variables (ecological and demographic factors) remain constant over the entire State. This is typically not the case with dynamic variables such as crime rates or recidivism that have geographic connotations and usually exhibit spatial patterns (Cahill and Mulligan 2007). By conducting the analysis at the county level, the study acknowledges that recidivism rates and their underlying correlates may vary spatially across different counties. This approach allows the project to consider the unique social, economic, and environmental factors specific to each county that may impact recidivism. Doing so, makes the study better able to capture the localized influences on reoffending and understand the potential importance of the neighborhood context to successful reentry. This approach can help identify areas of high or low recidivism rates, detect spatial patterns or clusters, and inform targeted interventions tailored to the specific needs of each county. Studies on the neighborhood context of recidivism suggest that the social and environmental context to which the formerly incarcerated reenter the community as well as the geographical accessibility to interventions and mobility are a pivotal part of successful reentry (Hipp et al. 2010; Kubrin and Stewart 2006).

1.2 Study Area

In 2007 in Washington State, recidivism rates were recorded at 65.9% for men and 53.6% for women (Caseload Forecast Council 2007). The Washington State Department of Corrections maintains and operates 12 prisons across the State. As of June 2022, the WADOC had 12,972 person(s) incarcerated under their supervision (WADOC 2022). These statistics highlight the importance of gaining a better understanding of the factors contributing to recidivism to inform effective public policy in the State.

The geographic scope of the analysis focused on the state of Washington in the United States. Washington State has shown a commitment to evidence-based practices and innovations in correctional programs, making it a compelling case study for assessing the effectiveness of interventions aimed at reducing recidivism rates.

Washington State is located in the Pacific Northwest region of the United States. It spans an area of approximately 184,661 square kilometers and is bordered by the Pacific Ocean to the west, Oregon to the south, Idaho to the east, and British Columbia, Canada to the north. The U.S. Census Bureau's 2020 Census indicates that the population in Washington State grew by .58% year over year from 2010-2020 to roughly 7.71 million residents. Of those residents, a majority reside around the Puget Sound Region on the Western Side of the Cascade Mountain Range. Population hot spots are also centered in the southernmost Clark County which is near Portland Oregon. In addition, the eastern half of the state's population is largely centered in Spokane and Benton Counties.

In Washington State, the socioeconomic landscape varies across counties, influencing recidivism rates in distinct ways. For instance, urban areas like King County, which encompasses Seattle, have numerous employment opportunities, vocational training centers, and support services that could positively impact individuals' chances of finding stable employment post-release. On the other hand, counties with essential healthcare infrastructure and support systems, such as Thurston County, where the state capital Olympia is located, may potentially witness better outcomes in terms of reducing recidivism (Figure 1). These urban areas across the state tend to have larger populations and a more diverse economy, offering a broader range of reentry opportunities. In contrast, rural areas, such as Adams County or Franklin County in Eastern Washington, often face unique challenges associated with fewer economic opportunities, higher poverty rates, and limited access to quality education and healthcare services.

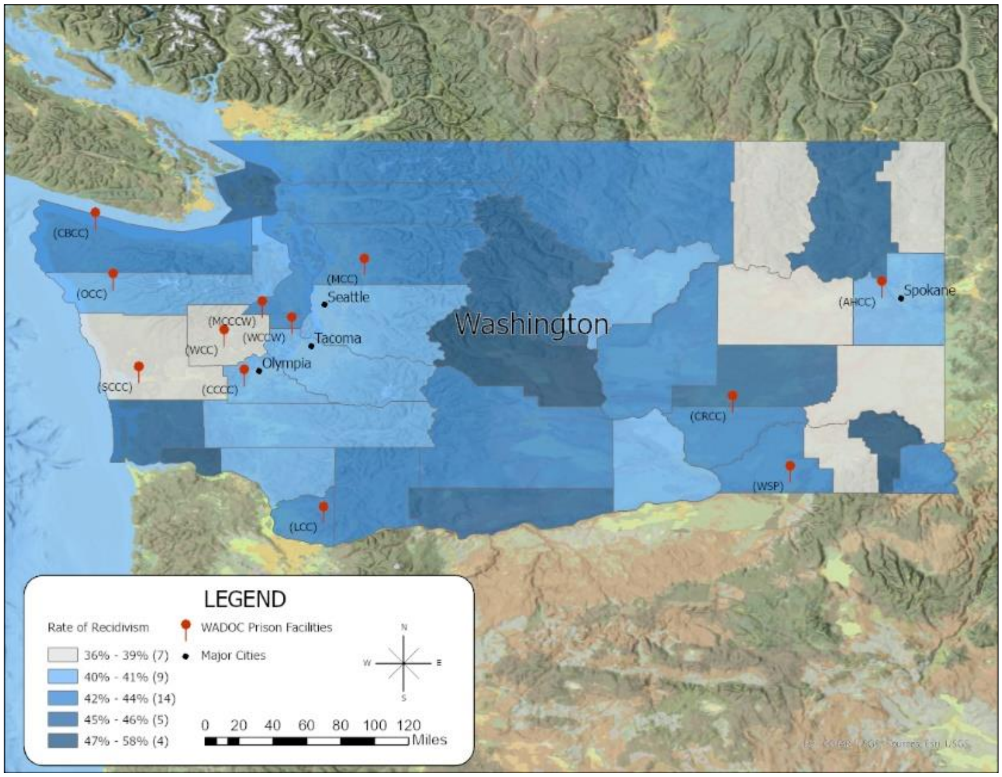


Figure 1. Rate of Recidivism (Reconfinement) 2012- 2022

The Washington State Department of Corrections operates twelve prisons across the state (Table 1). These facilities are located in 11 different counties (Figure 1). Of those twelve facilities four, Clallam Bay Corrections Center (CBCC), Coyote Ridge Corrections Center (CRCC), Monroe Correctional Complex (MCC), and Washington State Penitentiary (WSP) are designed to house incarcerated individuals with higher security classifications, including maximum security. Three of those, CRCC, MCC, and WSP are capable of housing roughly 2,400 individuals each making them the largest of the twelve facilities. The total capacity of all WADOC operated prisons as of June, 2022 is 14,698 incarcerated individuals. The average daily population across all WADOC facilities during that same time was 12,233 incarcerated individuals (WADOC, 2022). The majority of WADOC-operated prisons within the state are strategically situated around the Puget Sound Region. Upon comparing this distribution with the population density map provided above, noticeable patterns emerges between prison locations and population centers.

Table 1. WADOC facilities

Facility Name	Acronym	County	Capacity	Custody Level
Airway Heights Corrections Center	AHCC	Spokane	2258	Minimum and Medium
Cedar Creek Correctional Complex	CCCC	Thurston	480	Minimum
Clallam Bay Corrections Center	CBCC	Clallam	858	Medium, Close, and Maximum
Coyote Ridge Corrections Center	CRCC	Franklin	2468	Minimum
Larch Corrections Center	LCC	Clark	240	Minimum
Mission Creek Corrections Center for Women	MCCCW	Mason	321	Minimum
Monroe Correctional Complex	MCCCW	Snohomish	2400	Minimum, Medium, Close, and Maximum
Olympic Corrections Center	OCC	Jefferson	272	Minimum
Stafford Creek Corrections Center	SCCC	Grays Harbor	1936	Minimum, Medium, and Maximum
Washington Corrections Center	WCC	Mason	1268	Medium, Close, and Maximum
Washington Corrections Center for Women	WCCW	Pierce	738	Minimum, Medium, and Close
Washington State Penitentiary	WSP	Walla Walla	2439	Minimum, Medium, Close, and Maximum

1.3 Project Overview

The primary objective of this study was to investigate the potential correlation between the county of release and the propensity to recidivate. This exploration was based on an extensive WADOC prison admission/release dataset spanning from 2012 to 2022. As a secondary aim, the study aimed to contribute a small collection of variables linked to Social Disorganization Theory. This contextual framework aims to offer insights into county dynamics and assess if relationships found in studies from diverse jurisdictions hold true for the State of Washington.

1.3.1 Temporal Scope

The temporal scope of the study covers a period of ten years, from July 1, 2012 to June 30, 2022. The spatiotemporal granularity allows for an in-depth examination with a sufficient follow-up period to determine whether the propensity to recidivate is driven by spatial factors at the County level. The study aims to capture the geographic connotations and trends in recidivism rates over this period, being that it is the most recent data available.

1.3.2 Scale of Analysis

This analysis is conducted at the county scale to examine the variations in recidivism rates and factors across different areal units within Washington State. Washington State comprises 39 counties, each with its own unique characteristics, socioeconomic factors, and criminal justice practices. Analyzing this data at the county level enables a spatial examination of recidivism patterns and provides insights into county-specific factors that may contribute to reoffending. While the County scale may be effective at evaluating the overall effectiveness of WADOC operated prisons, it may fail to highlight the nuanced nature of recidivism. The choice of the scale of analysis was limited to data availability.

1.3.3 Primary Dataset

The study utilized a comprehensive prison admission/release dataset obtained from WADOC through a public records request. This dataset provides detailed information on the admission and release of individuals under WADOC custody or released between July 1, 2012, and June 30, 2022. The dataset comprises a total of 64,320 individuals, as received from the

WADOC prior to any data preprocessing. It also encompasses various variables, including an individual's unique identifier number (DOCNUM), admission and release dates, admission reasons, and their respective county of admission and release.

1.3.4 Limitations of the Primary WADOC Dataset

When analyzing this dataset, it became apparent that it poses certain limitations for analysis. The dataset received from WADOC, included data entry that was not standardized across the institution’s database. Each facility seemingly entered data in different ways which made preparing the data for analyses a major obstacle. The WADOC only maintains records at the county scale of analysis due to privacy concerns with finer scale data collection and dissemination. This limitation is related to the modifiable areal unit problem (MAUP). The MAUP refers to the potential distortion of results when aggregating data into different spatial units (e.g., counties) due to the arbitrary nature of the boundaries (Openshaw and Taylor, 1979; Openshaw, 1984). Aggregating data at the county level may mask finer-scale spatial patterns and heterogeneity within counties, potentially leading to unreliable conclusions.

Another major limitation was posed when attempting to extrapolate independent variables associated with individual-level characteristics. The efficiency of searching each individual DOC Number within a government database was severely limited by the volume of the data, thereby preventing the retrieval of demographic information and many other variables initially considered valuable for the analysis. It should also be noted that the dataset excludes records from county jails which typically hold incarcerated individuals until sentencing at which point WADOC takes custody of an incarcerated individual, Federal prisons, Immigration Detention facilities, and private prisons operating within the State. Consequently, the analysis is focused solely on individuals under custody or released by WADOC.

Furthermore, the primary dataset encompasses a temporal span from 2012 to 2022. While this ten-year period provides valuable insights into recidivism trends, it's important to consider the temporal limitations inherent in the data. Notably, individuals released towards the latter part of the 2012-2022 timeframe inherently have less opportunity post-incarceration to recidivate when compared to those released earlier in the time span. This temporal asymmetry could influence the observed recidivism rates and patterns, potentially skewing the analysis towards individuals who had more time to reoffend. Therefore, the findings of the study should be interpreted while keeping this temporal bias in mind.

Using the county as the scale of analysis has certain limitations that should be highlighted. The county scale might not capture the actual scale at which relevant processes are occurring. It is useful for analyzing general trends across a study area but typically does not adequately represent the nuances that are typical of social phenomenon such as crime and recidivism which are generally fluid across space. Another limitation is that it may not adequately account for variations in population density across different regions. When using counties, areas with higher populations will naturally have more incidents, which could lead to biased results if not appropriately adjusted for population differences. The areal unit of aggregation posed a major methodological limitation to the framework of this study. Furthermore, while the study focused on Washington State, the findings may not be directly generalizable to other states or jurisdictions. It is also essential to consider the unique characteristics and local context of each jurisdiction (county) when applying the study's findings to inform policy in different settings.

1.4 Document Overview

This thesis navigates through a structured exploration, beginning with a literature review in Chapter 2 that examines social disorganization theory, assesses critically relevant research conducted in Washington state, as well as studies employing spatial analysis on recidivism. Subsequently, Chapter 3 outlines the methodological approach, detailing the steps taken to achieve the research goals. Chapter 4 presents the results derived from these methods, offering insights into potential correlations between the county of release and recidivism. Finally, Chapter 5 delves into a comprehensive discussion that contextualizes the findings within real-world implications, thereby connecting research outcomes to practical applications.

Chapter 2 Literature Review

This chapter critically evaluates conventional statistical approaches to assessing recidivism, conducts an in-depth analysis of existing studies focused on recidivism within Washington State, and extends its scope to include a review of spatial analysis methodologies used in relevant literature beyond the state. Furthermore, this section explores the foundational underpinnings of social disorganization theory, which is considered a fundamental theoretical framework within the criminology and criminal justice disciplines.

2.1 Social Disorganization Theory

Two predominant theoretical dichotomies, social disorganization theory (SDT) and routine activities theory, are dominant in the academic literature on the neighborhood structural characteristics of crime and recidivism in the broader fields of criminology and criminal justice. In spatial analyses of recidivism, the predominant theoretical framework often employed is based on SDT (Drawve and McNeeley 2021). This study delves into the significance of the location—specifically the county—where an individual is released, a variable inherently tied to the social disorganization theory. Rooted in the premise that community structures influence crime tendencies, social disorganization theory aligns with our assessment.

SDT explains that communities characterized by poverty, residential instability, and racial diversity suffer from higher rates of crime (Shaw & McKay 1942). It was a theoretical shift away from the tendency to use only individual-level characteristics to describe crime and recidivism, to a recognition that place, and geographical location also contribute to these phenomena (Shaw and McKay 1972). The guiding principle of the theory is that crime and recidivism are directly correlated to ecological context and that social and economic disorganization which are typically characterized by poverty, residential instability, higher levels of gentrification or housing instability, ethnic diversity, and low literacy rates or educational attainment, lead to higher rates of crime and recidivism (Shaw and McKay 1942). Another major component of the theory is characterized by the social cohesion of a neighborhood. The neighborhoods that share common values and maintain strong social connections are more resilient to crime according to this theory (Shaw and McKay 1942). The primary criticism of this theory in criminology is that it ignores the individual-level factors that are often involved in crime and recidivism (Bursik 1988). Individual-level factors include the personal characteristics and environmental circumstances of an individual that may contribute to the likelihood of committing a crime or recidivating. Some of these factors include things such as genetics, poverty status, educational attainment, literacy, social/peer network, childhood experiences, health, and substance use or abuse.

2.2 Social Disorganization Theory in Crime and Recidivism Research

SDT is often utilized in studies on crime and recidivism by guiding the research design, influencing the selection of variables, and shaping the interpretation of data. A common approach to utilizing the SDT is building a subset of variables into an index meant to reflect social disorganization, and then testing for areas of concentrated disadvantage using the index (Kirk and Laub 2010). Studies might also start with hypotheses that connect recidivism rates to indicators of social disorganization such as residential mobility, socioeconomic status, family structure, concentrated disadvantage, and neighborhood resources. When utilized in studies on crime and recidivism a common approach is acquiring qualitative survey data from residents on the context of a specific neighborhood (Sampson et al. 1997). More recent methodologies include incorporating spatial analysis and multilevel models of census data and crime rates across neighborhoods to see how key elements of social disorganization theory predict crime rates (Kubrin and Weitzer 2003; Hipp 2007).

Many academic studies are grounded in the SDT framework. A recent study exemplifies this approach by examining how neighborhood characteristics link to gun violence using data from 2014 to 2019 (Maher et al. 2022). In another study grounded in the Social Disorganization Theory framework, researchers employed specific variables to gauge concentrated disadvantage within neighborhoods. This included assessing low income, which involved quantifying the percentage of households in the community with incomes falling below the lowest quintile of the Netherlands' income distribution in 2008. Additionally, the study considered the proportion of households receiving welfare benefits in 2006 (Gerben et al. 2013). These chosen variables are indicative of economic hardship and align with social disorganization theory's core premise that areas marked by economic disadvantage are more likely to experience heightened crime rates (Shaw and Mckay 1942).

2.3 Mathematical Methodology

Exploratory data analysis (EDA) is a set of methods that evaluate variations among different groups, identify data outliers, detect clustering, and highlight non-linear relationships between variables (David and Tukey 1977). The purpose of EDA is to visually interpret data in a way that could potentially unveil underlying patterns, leading to novel and often unexpected insights (David and Tukey 1977). Recent advances emphasize the application of multivariate EDA techniques, specifically the utility of boxplots and scatterplots for large data sets (Nicodemo and Satorra 2022).

Correlative analysis is a pivotal tool in unveiling potential relationships and associations within complex datasets. This analytical approach enables researchers to explore the degree of linear dependence between two or more variables, unveiling insights into how changes in one variable might correspond to changes in another. By calculating correlation coefficients, such as Pearson's correlation coefficient, researchers can quantify the strength and direction of relationships, shedding light on patterns that might not be immediately apparent (Ostertagova, Ostertag, and Kováč 2014). Correlation analyses then, are useful in preliminary hypothesis testing used to guide further research endeavors. It should always be cautioned that correlation does not imply causation.

Exploratory spatial data analysis (ESDA) is an extension of the traditional EDA techniques to spatial data. Spatial data has geographical or spatial information (x,y coordinates) associated with it, and ESDA techniques are designed to uncover spatial patterns, anomalies, or other interesting characteristics in such data. One common aspect of ESDA is the assessment of spatial autocorrelation, which is the tendency of nearby geographical entities to exhibit similar attributes is a principle often attributed to Waldo Tobler and cited as Tobler's First Law of Geography (Tobler 1997). At its core, the concept of spatial autocorrelation assesses how observations of the same variable are related to each other across a study area, offering insights into the spatial patterns in the data distribution. This concept becomes especially pertinent when considering crime or recidivism patterns, where geographic proximity can often translate to similar characteristics and behaviors (Cahill and Mulligan 2007). When assessing dynamic spatial patterns such as recidivism, it is uncommon that homogeneity exists across an entire, global, study area. The spatial weights matrix (SWM) represents the spatial relationship amongst all neighbors or observations in a dataset in a mathematical matrix format which makes it possible to measure autocorrelation (O'Sullivan and Unwin 2010).

Local statistics are a method for taking this heterogeneity that exists from one observation to the next across a study area and using the descriptive statistics of each to derive a more nuanced understanding of the spatial patterns that make up the global study area (O'Sullivan and Unwin 2010). Since Unwin (1996) and Fotheringham (1997) highlighted the importance of local statistics in their seminal works on the issue, the use of local statistics in spatial analysis research has been limited to the technological capabilities – especially the computing power of home computers. Within the last 15 to 20 years, access to extensive and detailed spatial as well as spatiotemporal datasets has offered the chance to acquire new insights and enhance our comprehension of intricate geographic phenomenon (Mennis and Guo 2009). Spatial statistical methods that were previously computationally burdensome have become achievable with the advances in computer science (Goodchild 2007). Common statistical measures of spatial autocorrelation include the Moran's I test (Moran 1948), Getis and Ord's G statistics (Getis and Ord 1992), Geary's C (Geary 1954), join count statistics (Besag 1974), and local indicators of spatial association (LISA) (Anselin 1995).This study has chosen to implement the Getis-Ord G_i^*Statistic to assess localized autocorrelation.

2.4 Areal Unit of Aggregation

While a wide range of interdisciplinary research approaches have centered on the significance of neighborhoods, a recurring aspect in many of these studies is the tendency to overlook the essential component of determining the appropriate level of aggregation for capturing the neighborhood effects (Hipp 2007). Choosing the areal unit of aggregation (AUA) at an appropriate scale to define the neighborhood structure and underlying spatial processes associated with such is critical to the results of any analysis. Too large of an aggregation has the potential to obscure more localized spatial patterns and effects by combining the heterogeneity of multiple neighborhoods into one aggregate (O'Sullivan and Unwin 2010). This makes the analysis susceptible to the ecological fallacy or ecological inference, which occurs when it is assumed that all observations in a dataset are homogenous across the whole of a study area, or in other words, it is a failure to recognize that the data aggregate is not representative of each observation within that aggregation (Robinson 1950).

An ideal approach to finding the most appropriate areal unit of aggregation to represent the neighborhood contextual aspects of crime and recidivism is to flexibly aggregate the data to varying geographic sizes and compare the results (Hipp 2007). The ideal aggregation would mimic the scale at which the phenomenon of recidivism occurs, for instance, one study found that Census Block Groups most adequately represented the actual geographic areal measurements of the city neighborhoods in both Los Angeles and San Francisco, California (Grannis 1998). Further research related to the proper geographical units of aggregation for neighborhood level crime and recidivism have assessed the Census tract and Census block units of aggregation (Hipp 2007). This same study also called for future research to assess the larger geographical units of aggregation such as combinations of block groups, or the County scale of analysis. That is one place where this study intends to contribute to existing literature; By assessing whether the county of release is associated with a propensity to recidivate this study seeks to contribute to research on the appropriate scale of aggregation for assessing the neighborhood contextual elements of crime and recidivism.

2.5 Overview of Previous Analytical Research on Recidivism

Understanding the underlying spatial processes of recidivism is an important component of studies in the fields of criminology and criminal justice. It not only sheds light on the effectiveness of rehabilitation programs and the challenges of reintegration but also offers insights into the broader social dynamics that contribute to criminal behavior. As the rates of recidivism continue to pose significant challenges to the criminal justice system, researchers have sought diverse statistical and modeling methodologies to unravel the complexity of factors influencing the propensity to recidivate. Prior research in Washington State has utilized various methods for assessing recidivism, a Level of Service Inventory-Revised (LSI-R) score analysis, meta-analytical approaches that calculate effects sizes, and multivariate and bivariate analysis methodologies such as correlative analyses. The LSI-R is a widely used assessment tool in the field of criminology and criminal justice. It is designed to evaluate the risk and needs of offenders, particularly those who are involved in the criminal justice system (Phipps et al. 1999). When expanding the scope to encompass studies from outside jurisdictions, it is common to observe the utilization of a wide range of statistical methods to assess recidivism. Binary logistic regression is frequently employed, offering insights into the relationship between predictor variables and the likelihood of recidivism. Geographically weighted regression and area under the receiver operator characteristic (AUC) analysis are also popular choices, allowing for the examination of spatial variations and predictive accuracy. In addition, Bayesian estimation techniques provide a probabilistic framework to assess the complex interactions contributing to recidivism. Correlative analyses further enhance understanding by investigating associations among variables. These methodologies collectively contribute to a comprehensive assessment of the multifaceted phenomenon of recidivism.

2.5.1 Washington State Specific Studies on Recidivism

The passing of the Offender Accountability Act in 1999 by the Washington State legislature mandated the use of a risk assessment for post-release sentencing guidelines and allocating resources to the highest-risk formerly incarcerated individuals, for this purpose the

DOC adopted the use of the LSI-R (Phipps et al. 1999). This assessment is used as the risk for re-offense component in DOC's Risk Management Identification (RMI) system (Phipps et al. 1999).

A common approach to assessing recidivism is to assess the effectiveness and accuracy of the LSI-R risk assessment questionnaire as a predictor of recidivism (Manchak, Skeem, and Douglas 2008; Aos 2003; Camasso and Jagannathan 1995). One study conducted by the Washington State Institute for Public Policy (Aos 2003) analyzes rates of recidivism among a sample of 22,533 offenders using the LSI-R scores as the predictor variable. The authors used three statistical measures to assess the efficacy of the LSI-R. First, they employed correlation coefficients. They then conducted an area under the receiver operator characteristic (AUC) to assess the strength of the correlation. Finally, they obtained the odds ratio from multivariate analyses and used them to assess the contribution of the variables that make up the LSI-R questionnaire for predicting recidivism.

The research reveals a gradual increase in recidivism rates with higher LSI-R scores, indicating a positive correlation between scores and re-offending likelihood. However, the absence of distinct cut-off scores suggests a lack of naturally occurring low- and high-risk categories (Aos 2003). This upward trend is supported by equations showing that each one-point increase in the LSI-R score corresponds to incremental percentage point increases in misdemeanor and felony recidivism rates, felony recidivism rates, and violent felony recidivism rates (Aos 2003). Despite low correlation coefficients, the moderate AUC values suggest a significant but not overwhelmingly strong association between LSI-R scores and recidivism, emphasizing its predictive capability (Aos 2003). The distribution of LSI-R scores, as demonstrated by a bell-shaped curve, highlights a clustering of scores around the mean, with fewer offenders having notably high scores (Aos 2003).

The study further dissects the associations between LSI-R domain scores and recidivism. It underscores that certain variables, particularly criminal history, exhibit stronger correlations with recidivism, contributing significantly to the overall predictive power of the LSI-R (Aos 2003). Criminal history emerges as a robust predictor of recidivism in this analysis (Aos 2003). The findings collectively highlight the dynamic nature of predicting recidivism and the varying strengths of association between LSI-R variables and different types of recidivism. This offers valuable insights for risk assessment and intervention strategies within the criminal justice system (Aos 2003).

Another study conducted in Washington State assesses evidence-based interventions and programs for adult incarcerated individuals to determine the efficacy of those interventions (Aos, Miller, and Drake 2006). The intent is to determine which programs are working and which are not to ultimately aid in lowering recidivism rates in the State of Washington. This study is a cost-benefit analysis of sorts. Notably, only a limited subset of the evaluated program assessments originated from within Washington State; instead, the study encompassed evaluations spanning a 40-year timeframe commencing from 1970, drawn from a wide range of English-speaking nations (Aos, Drake, and Miller 2006).

Aos, Drake, and Millers' study used a meta-analytic approach to calculate effect sizes. They employed both fixed effects and random effects modeling in their analysis. These methods involve pooling the results of multiple rigorous evaluations of different programs to calculate an overall estimate of the effect size, which indicates the magnitude of the impact of each program on recidivism rates.

The results of their study reveal that certain programs are effective in reducing recidivism rates, while others are ineffective. Among the effective programs are drug courts, in-prison therapy communities, cognitive-behavioral drug treatment while incarcerated, drug treatment in the community, and cognitive-behavioral treatment for domestic violence offenders (Aos, Drake, and Miller 2006). However, some programs, such as jail diversion for mentally ill individuals and some intensive supervision programs, did not demonstrate significant recidivism reduction (Aos, Drake, and Miller 2006). Their study emphasizes the need to focus resources on evidence-based programs to avoid ineffective approaches to successful reentry.

Other studies analyzing recidivism assess the impacts of individual-level variables on the propensity to recidivate (Miller, Drake, and Nafziger 2013; Gagliardi et al. 2004; Beckett, Harris, and Evans 2008; Lovell, Johnson, and Cain 2007). In a study examining recidivism among mentally ill formerly incarcerated individuals, researchers analyzed a sample of 333 individuals released from Washington State Prisons in 1996 and 1997 with a post-release follow-up period ranging from 27 to 55 months. The study found that approximately 77% of mentally ill offenders (MIOs) were re-arrested or charged with new crimes, with 41% convicted of another felony offense and 23% convicted of violent crimes. Comparatively, the general prison population had a 38% conviction rate for new felonies. Several independent variables were assessed in this study, including sentence length, past felonies, and past drug felonies, which showed statistically significant correlations to recidivism. Surprisingly, the study concluded that MIOs were no more likely to recidivate than the general population. The authors argued against including mental health status as a predictor in recidivism assessment tools.

2.5.2 Relevant Studies in Spatial Criminology

In the context of spatial criminology, which is often used to assess recidivism, researchers examine how crime and criminal behavior vary across specific localized environments. Typically, research investigates whether certain factors that contribute to crime remain consistent across different areas or if they are influenced by the characteristics of each location. Understanding these variations is crucial for developing effective crime prevention strategies and policies that consider the specific conditions of different places. (LeBeau and Leitner 2011).

Within the realm of spatial criminology, researchers have conducted studies that focus on diverse aspects of crime. For instance, a National Institute of Justice (NIJ) funded project explores the connection between neighborhood characteristics and juvenile delinquency/recidivism in Philadelphia. Their findings reveal that delinquency is concentrated in impoverished, violent crime-prone neighborhoods (Mennis et al. 2011). In another study, the concept of the "ambient population" was introduced to refine crime rate calculations, offering empirical evidence for its application (Malleson and Andresen 2015). In the realm of criminal geographic profiling (CGP), one study identified the center of minimum distance as a reliable model to predict recidivism. This model calculates the geographical point or area where individuals at risk of recidivism are most likely to gravitate towards, shedding light on the direction and magnitude of the relationship between spatial factors and recidivism outcomes. In that study they also explore Bayesian estimation techniques to enhance CGP accuracy (Levine and Block 2011). Another methodology that has recently surfaced is to evaluate the weighted displacement quotient (WDQ) for crime suppression operations, which are strategies aimed at reducing criminal activities in specific areas (Ratcliffe 2004). Another study examined the impact of population displacements after Hurricanes Katrina and Rita on crime rates in

Louisiana. Their study suggests that crime remains stable or declines in regions receiving emergency evacuees from the disaster zones (Leitner et al. 2011).

When assessing recidivism using spatial analysis, it is far more common to apply global statistics and regression models to the phenomenon than it is local statistics (Kim et al. 2013; Leymon et al. 2022; Wang et al. 2013; Stahler et al. 2013; Hipp, Petersilia, and Turner 2010). Only in the last decade and a half have studies begun to leverage spatial modeling to assess the spatial relationship between the community contextual environment and recidivism empirically (Gottfredson and Taylor 1988; Stahler et al. 2013; Kubrin and Stewart 2006). Those studies that do assess recidivism using local statistics tend to rely on sampling units or surveys to derive their results (Drawve and McNeeley 2021; Cahill and Mulligan 2007).

In one such study the relationship between community context and recidivism in Minnesota is assessed. The study investigated the impact of neighborhood-level factors on recidivism, extending beyond social disorganization. The analysis considered the influence of prosocial-local institutions and criminogenic establishments on recidivism outcomes. While the presence of criminogenic establishments showed no significant relationship with recidivism, an increase in the number of prosocial institutions within neighborhoods was associated with a decrease in arrest likelihood (Drawve and McNeeley 2021). This finding deviated from previous research. Moreover, the study revealed that the effect of prosocial establishments was more pronounced in less-disadvantaged neighborhoods (Drawve and McNeeley 2021). The research suggests that the structural nature of disadvantaged neighborhoods might outweigh the impact of prosocial institutions, emphasizing the need for tailored approaches in service provision. Despite certain limitations, this study contributes to the understanding of neighborhood influences on recidivism and advocates for localized research approaches. The results of this multilevel

Bernoulli model suggest that the greater number of prosocial establishments per neighborhood leads to a decrease in recidivism rates (Drawve and McNeeley 2021).

In a study published in 2007, researchers aimed to understand violence levels in Portland, Oregon's Multnomah County from 1998 to 2002. Initially using a traditional Ordinary Least Squares (OLS) regression model, they found it inadequate, explaining less than 40% of violence variance and yielding counterintuitive results (Cahill and Mulligan 2007). This underlined the limitation of global regression models for spatial data, missing local variations.

Chapter 3 Methods

This chapter provides a detailed description of the methodological approach taken in this study to assess recidivism. It begins with an introduction to the essential data required for the analysis, followed by the steps taken to prepare and clean the data for spatial analysis. Subsequently, the proposed methodological approach is outlined, highlighting how it will facilitate the project's primary objective of determining the potential correlation between the county of release and the likelihood of recidivism.

Figure 2 provides a visual representation of the project workflow. The initial phase involved detailing the data cleaning and preparation process for the primary WADOC admission/release dataset. Subsequent steps encompass dataset normalization, the presentation of baseline and summary statistics, and a power analysis to determine a minimum sample size (number of total releases for each county). Univariate and bivariate data visualization methods, such as histograms, boxplots, and a scatterplot matrix (SPLOMs), assessed data distribution and Pearson’s correlation coefficients. Both an ANOVA test and subsequent T-test and Cohen’s *d* and their corresponding p-values served to explore the correlation between the county of release and recidivism propensity in Washington State from 2012-2022.

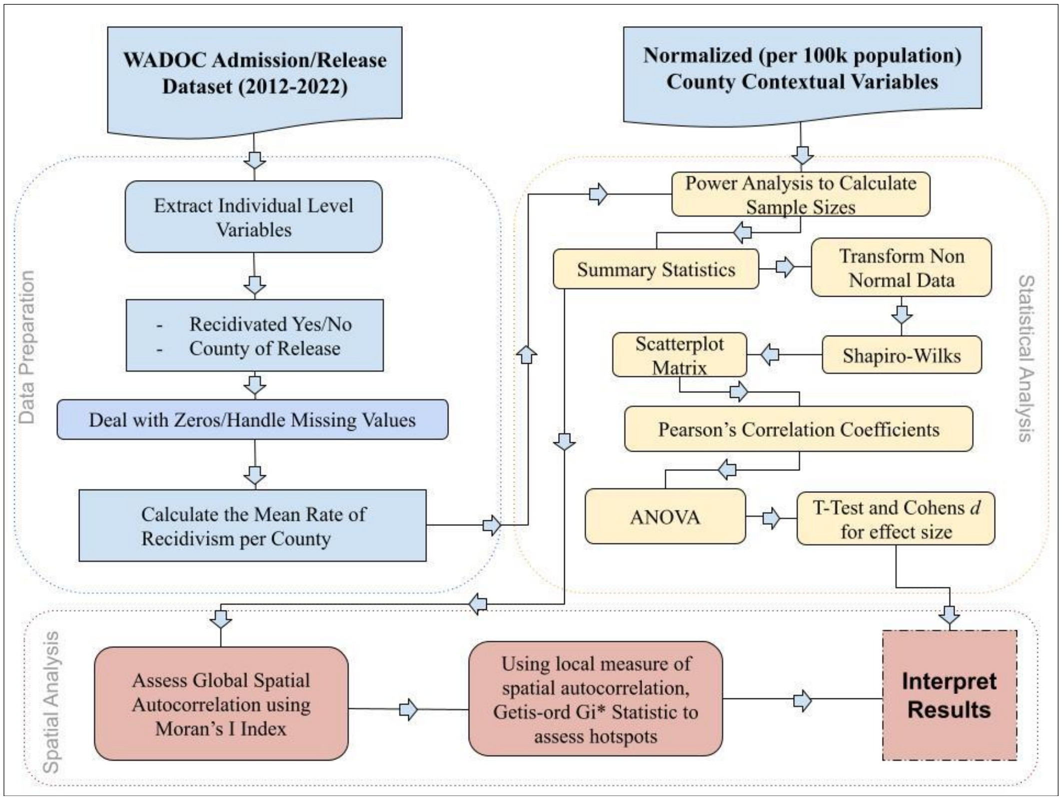


Figure 2. Workflow diagram

3.1 Variable Selection and Data

The foundation of this analysis lies in a collection of data sourced from reputable institutions, including the Washington State Department of Corrections (WADOC), the U.S. Census Bureau, the American Community Survey, and relevant datasets from the Washington State Open Data Portal. These diverse datasets were utilized to provide a comprehensive basis for the ensuing workflow.

This study's selection of contextual variables is aligned with the key tenets of social disorganization theory (SDT), with each choice grounded in the scholarly discourse on the topic. For instance, the inclusion of mass shootings and gun violence as a variable draws upon analyses that demonstrate its significant impact on the social fabric and crime rates within communities, thereby echoing SDT's focus on the influence of community context on criminal behavior, especially recidivism (Maher et al. 2022). Furthermore, the study's focus on public assistance enrollment rates draws on insights from a recent study which contextualized welfare benefits within the SDT paradigm, highlighting economic disadvantage as a pivotal factor in community social dynamics and crime occurrence. (Gerben et al. 2013). Literacy rates are also considered, given their established role in indicating community cohesion, social control, and educational attainment, factors that are fundamental to SDT's explanation of how lower literacy levels can undermine social stability and control, potentially leading to higher crime rates (Shaw and McKay 1942).

The deliberate inclusion of these variables is designed to capture the complex interplay of factors that contribute to social disorganization and its relationship with recidivism, consistent with the empirical evidence presented in SDT-related research.

Table 2. Variables

Variable	Source	Temporal Extent	Definition
Rate of Recidivism	WADOC	July 1, 2012 to June 30, 2022.	Recidivism rates by county of WADOC operated prisons from 2012-2022
County of Release	WADOC	July 1, 2012 to June 30, 2022.	The county that an incarcerated individual (DOCNUM) was released to.
Enrolled in Public Assistance (SNAP/SSI)	American Community Survey	2021 (five-year estimate)	The population enrolled in SNAP food assistance or Supplemental Income public assistance.
Lacking Basic Prose Literacy Skills	National Center for Education Statistics	2003	The percentage lacking basic prose literacy skills by county.

3.1.1 Description of the Data

In this study, a comprehensive dataset spanning from 2012 to 2022 was utilized, consisting of detailed prison release and admission records obtained from the Washington State Department of Corrections (WADOC), as discussed in Chapter 1. The WADOC dataset provided valuable information on the admission and release of individuals under WADOC custody during the specified ten-year period. While the concept of recidivism in criminal justice and criminology is often used to broadly describe a formerly incarcerated individual who was sentenced to corrections, released, and then displayed further criminal behavior (Maltz 2001), for the purpose of this analysis, it is defined instead as a formerly incarcerated individual who was released from confinement and returned to incarceration at a later date. It is operationalized in this manner to capture only the most serious relapses of criminal behavior that result in further confinement. This dataset was complemented by several county contextual variables, which were incorporated as additional social disorganization predictors to enhance the analysis.

The county contextual variables used as control variables included in this analysis were carefully selected based on academic research on Social Disorganization Theory. They are meant to capture various dimensions of the socio-economic and demographic context within each county as a measure of social disorganization. These variables comprised the percentage of the population lacking basic prose literacy skills, and the number of individuals enrolled in public assistance programs, specifically SNAP food assistance, and SSI supplemental income. These datasets were then normalized for population per 100k.

By integrating these diverse variables, the study aimed to explore potential associations and shed light on the complex interplay of factors contributing to recidivism at the county level in Washington State. The inclusion of county contextual variables enables a deeper understanding of the social, economic, and demographic factors that may influence recidivism rates across different counties. This approach allows for a more nuanced analysis, considering the unique characteristics and challenges within each county and their potential impact on reoffending rates. Furthermore, the temporal scope of the dataset (2012-2022) allows for an in-depth examination of recidivism trends over that time, providing valuable insights into what implications they might have for policy and intervention strategies. Overall, this dataset, encompassing both prison release records and county contextual variables, forms the foundation for the evidence-based spatial analysis approach used in this study to address the research objective effectively.

3.1.2 Limitations of the Data

The primary dataset used in this study presents limitations that need to be considered when interpreting the results (Figure 3). One major limitation is the use of the county scale of analysis, which may not adequately represent the nuances of the neighborhood context. At this scale, the spatial variation is constrained to the large areal measurements of the counties, masking potential spatial heterogeneity that could exist at finer, more local scales of analysis. Moreover, counties with larger populations will naturally have greater numbers of incidents for count data variables, potentially biasing the results. Additionally, the dataset obtained from the Washington State Department of Corrections (WADOC) lacks demographic-related information and specifics regarding the prison from which individuals were released. The absence of these crucial details limits a deeper exploration of individual-level characteristics and their potential influence on recidivism rates. It is essential to acknowledge these limitations to ensure a comprehensive interpretation of the findings. Despite these constraints, the dataset serves as a valuable starting point for investigating the correlation between the county of release and the likelihood of recidivism within Washington State.

DOCNUM	ADMIT_DATE	RELEASE_DATE	ADMIT_TYPE	ADMIT_REASON	ADMIT_COUNTY	RELEASE_COUNTY	RELEASE_STATE
6023	14-Jan-13	23-Dec-13	VIOLATOR	READMISSION	OKANOGAN	SNOHOMISH	
21090	8-Feb-47		FIRST ADMISSION	FIRST ADMISSION TO PRISON	KING		
26909	23-Jan-13	24-Mar-20	FIRST ADMISSION	FIRST ADMISSION TO PRISON	FRANKLIN	SPOKANE	
28062	22-May-70	4-Feb-18	REVOCATION	PAROLE REVOKE WITH A NEW SENTENCE	PIERCE	FRANKLIN	
28159	5-Aug-11	23-Oct-13	FIRST ADMISSION	FIRST ADMISSION TO PRISON	KING		OR
28255	20-Oct-87		RE-ADMIT	READMISSION	PIERCE		
28523	5-Jun-08	18-Oct-16	RE-ADMIT	READMISSION	PIERCE	FRANKLIN	
28632	11-Dec-81	20-Oct-14	RE-ADMIT	READMISSION	YAKIMA	BENTON	
29644	1-Feb-68	20-Apr-21	FIRST ADMISSION	FIRST ADMISSION TO PRISON	KING	KING	
29645	1-Feb-68	12-Nov-20	FIRST ADMISSION	FIRST ADMISSION TO PRISON	KING	KING	
40424	29-Aug-14	7-Apr-15	FIRST ADMISSION	FIRST ADMISSION TO PRISON	THURSTON	THURSTON	
40847	24-Sep-13	5-Jun-17	FIRST ADMISSION	FIRST ADMISSION TO PRISON	SNOHOMISH	SNOHOMISH	
41750	8-Jan-13	15-May-14	RE-ADMIT	READMISSION	MASON	SKAGIT	
44242	25-Sep-09	4-Mar-16	FIRST ADMISSION	FIRST ADMISSION TO PRISON		GRAYS HARBOR	
50101	8-May-14	17-Dec-14	RE-ADMIT	READMISSION	GRAYS HARBOR	KING	

Figure 3. An image of the raw dataset as received from WADOC

3.2 Data Preprocessing

The raw data received from the Washington Department of Corrections (WADOC) was initially in PDF format and needed to be converted to CSV format for further processing. The data provided by the WADOC was merged by the Public Records Unit to combine 10 years of data, resulting in a large dataset. However, the merged dataset had inconsistent cell sizing and formatting, requiring reformatting and standardization before proceeding with data exploration and preparation.

3.2.1 Data Conversion and Standardization

The dataset contained many invalid date values which required transforming them into valid dates. Many of the date values were in text format, contained leading spaces, or included special characters as a result of an error in data entry at the point it was collected and entered. Conversion to valid date values was achieved by using the TRIM and DATEVALUE functions in Microsoft Excel. The dataset was then resized to extend the range horizontally to accommodate all the new column headers associated with the admission and release dates. These headers included the first, second, third, fourth, and fifth admission dates and counties, as well as the first, second, third, fourth, and fifth release dates and counties. The objective was to consolidate multiple entries into a single entry for each unique identifier (DOCNUM variable), allowing for a more streamlined analysis.

3.2.2 Handling Missing Values

During the exploratory phase, it was observed that the dataset contained numerous missing values across various variable columns. Each missing value was carefully assessed to determine appropriate measures for handling them before proceeding with further analysis. In some cases, missing values could be explained, such as multiple admission dates and only one release date indicating a sentence extension while incarcerated. In these instances, the corresponding cells were left blank. However, there were also cases where admission or release counties were missing, indicating transfers to the custody of another jurisdiction or department, such as Immigration and Customs Enforcement or the US Marshal. These instances were removed from the analysis.

Additionally, approximately 9,000 cells had one admission date without a corresponding release date, which could be attributed to life sentences, individuals in-custody but awaiting sentencing, escaped prisoners who were never relocated, or individuals who died while incarcerated. These entries were removed from the dataset as they did not contribute to the overall analysis as there was no chance for recidivism.

3.2.3 Dependent Recidivism Outcome Variable and Date of Recidivism

A critical component of this analysis is the dependent outcome variable, which in this case, indicates whether an individual has recidivated within the given spatiotemporal granularity of the analysis. By leveraging the IF, AND, and MAX functions, individuals with admission dates that came after their first release dates were identified in a helper column and labeled as

RECIDIVATED. Additionally, all DOCNUMs that exhibited a first release date that was prior to the first admission date were identified as RECIDIVATED. These dates were stored in a helper column and used later as the date to identify the correct county of release. The individuals identified represent cases of recidivism, as they left prison and returned later between 2012 and 2022. Another column was inserted, and a formula was created to label all blank values, or individuals who had not recidivated, as "0" (“NO RECIDIVISM”), while those labeled as "RECIDIVATED" were assigned a value of "1". It's worth mentioning that out of the formerly incarcerated individuals who experienced recidivism after their release, a total of 1,117 individuals faced multiple instances of recidivism. This discovery adds an intriguing layer to the analysis, but it also brings about important considerations for the results. The decision to treat these cases of multiple recidivism as separate entries, or to consolidate them as a single entry, has significant implications on the interpretation and outcomes of the analysis. Treating multiple instances of recidivism as separate entries could potentially provide a more comprehensive understanding of the recidivism phenomenon, capturing the varied experiences of individuals. On the other hand, aggregating these cases into a single entry will simplify the analysis but could potentially mask nuances in the recidivism patterns.

Treating multiple recidivism cases as a single entry simplifies the dataset, making it easier to manage and analyze however, given that a substantial portion of the entries involve multiple recidivism instances and a significant proportion have different release counties for each instance, consolidating these cases into a single entry might not fully capture the nuanced variations present in your dataset. For this reason, the data was used with each instance treated separately.

3.2.4 County of Release

After all of the 45,731 incarcerated individuals had been screened to determine recidivism status from 2012-2022 and that status had been recorded, the next step required determining which County each individual entry was released to post incarceration. This required aligning the date of recidivism that we logged into a new helper column (admit date that most closely follows the first release date) with the county of release for that corresponding entry. The remaining admission/release entries for each individual DOCNUM (most typically sentence extensions during incarceration) were removed from the analysis. This left the instances of recidivism as a binary helper column 1,0 and the County of release next to that helper column. The next step was to calculate the counts of total releases and incidents of recidivism for each individual county. To achieve these results the IF and ANDIF functions were utilized to check both the County name matches the newly created horizontal headers for the County of release and the recidivism status was a 1 (RECIDIVATED) if that were true, then the formula added a 1 in a new helper column, otherwise it returned zeroes. This permitted the calculation of the total number of releases and incidents of recidivism per county.

3.2.5 Data Normalization

Once the recidivism data was sorted into the 39 counties with totals for the number of incidents of recidivism, and the total releases the data was normalized to the total rate of recidivism per county (Figure 4). To calculate that rate the formula employed was:

$$RATE\ OF\ RECIDIVISM = \ (\#RECIDIVATED/TOTAL\ RELEASES) \tag{1}$$

Opting against normalizing the recidivism rates per 100k as was done with the remaining variables, was a deliberate choice. The unnormalized rates inherently account for population sizes, considering that areas with larger populations naturally yield more incidents. Since densely populated regions also have more releases, the existing rates inherently incorporate population dimensions into the analysis, making normalization unnecessary.

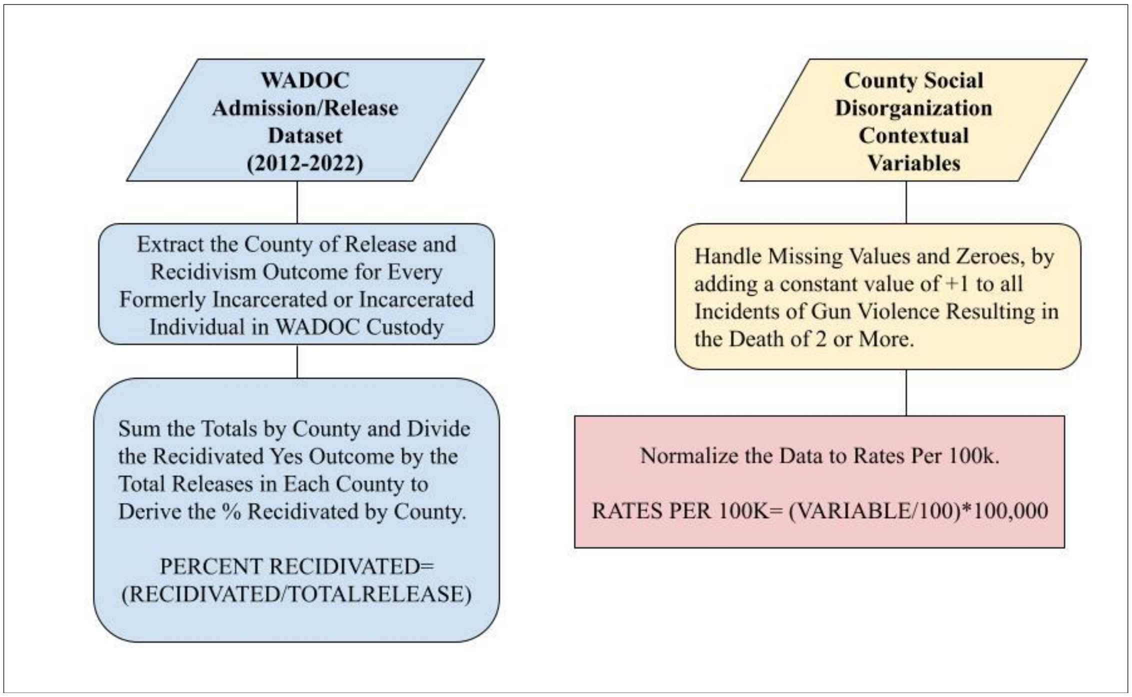


Figure 4. Workflow diagram of the data normalization process

The County social disorganization contextual variables needed to be normalized for population. The percentage of those lacking basic prose literacy skills in decimal form simply needed to be multiplied by 100,000 to derive the rate lacking basic prose literacy skills per 100k. The number enrolled in public assistance (SNAP/SSI) needed to be normalized by dividing the number enrolled by the county population and then multiplying by 100,000 to get the rate enrolled in public assistance per 100k.

3.3 Exploratory Data Analysis

In the subsequent section of this study, a transition occurred from data preparation and normalization to the statistical analysis phase. This analysis followed a step-by-step approach aimed at comprehensively examining the relationship between the county of release, social disorganization contextual variables, and recidivism rates in Washington State. Initially, baseline descriptive statistics were calculated, including the recidivism rate for each county, which was used to determine the overall recidivism rate as well as county rates for WADOC-operated facilities in the State of Washington during the period from 2012 to 2022. Subsequently, a power analysis was conducted to ascertain the appropriate number of total releases for each county (sample size) for this study. Following this, summary statistics were computed and reported to provide a clear overview of the dataset's characteristics.

To gain deeper insights into the data, univariate data visualizations were employed to assess distributional characteristics, and transformations were applied to variables that exhibited significant deviations from a normal distribution. After transformation, the Shapiro-Wilks test was utilized to reevaluate the normality of the data.

The analysis then shifted focus to exploring the linear relationships between the social disorganization contextual variables and recidivism rates through bivariate data visualizations. To quantify these relationships, Pearson's correlation coefficients were used. Finally, attention was turned to the crucial question of whether the county of release was correlated with recidivism rates in Washington State. This was achieved through the application of both ANOVA and T-tests, further supported by Cohen's d for effect size calculations.

Descriptive statistics were employed to gain initial insights into the distribution and characteristics of the data. Key summary statistics, including means, standard deviations, medians, and ranges were calculated for the variables incorporated into this study within Excel utilizing the Data Analysis Tool. This step provided a snapshot of the overall recidivism rates and a glimpse into the context of the county of release of the incarcerated individuals across Washington State, especially in terms of social disorganization.

3.3.1 Baseline Descriptive Statistics

To assess the relationship between the county of release and propensity to recidivate, a contingency table was created in Excel. A contingency table, also known as a cross-tabulation or crosstab, displays the frequency distribution of various categorical variables in tabular format (Agresti 2002). Crosstabs aid in better understanding the relationship between the variables. The creation of a crosstab table allowed for pairwise comparison between the counts of recidivism and the identification of initial patterns present in the data. Furthermore, it allowed for the calculation of the mean rate of recidivism per county which was used as the dependent variable in this analysis. After creating a contingency table of the total number of releases per county and the total number of recidivism incidents, the rate of recidivism was calculated for the 39 counties across the State of Washington, notating the counties that exhibited low numbers of total releases.

A power analysis was conducted in R.studio utilizing the 'pwr' package to derive a sample size with a .80 statistical power (Travers et al. 2021). A weighted average of Cohen's *d* ("Pooled *d*") was derived and input as the estimated effect size for the analysis. The alpha was set to .05, with 39 groups. The results of the power analysis derive adequate sample sizes per group for the selected statistical power. Once the number of total releases threshold was established, counties that did not meet that threshold were removed from the analysis.

To delve into the data's characteristics, summary statistics were calculated and analyzed for each of the four variables within Excel, utilizing the Data Analysis Tool. This process aimed to reveal central tendencies, comprehend the data distribution, and gauge variance. Furthermore, attention was paid to any significant deviations from normal distributions, which could potentially violate the assumptions of subsequent statistical tests. The results were organized in separate tables for each variable, facilitating a thorough examination and interpretation of the data's key attributes and overall distributional characteristics.

3.3.2 Univariate Data Visualization

Analyzing the data's distribution provides a foundational understanding of the dataset's characteristics. It aids in formulating hypothesis testing, assessing the normality of the data, and determining the most suitable statistical test for a given use case. This study leveraged both histograms and boxplots to further assess the data's distribution. To do so, the data was imported into R.Studio and the ggplot2 package was utilized. Charts were created for all of the variables in the analysis.

3.3.3 Data Transformation

Based on the findings of the summary statistics and the univariate data visualizations it was determined that the rate of those lacking basic prose literacy skills per 100k needed to be transformed using a natural logarithmic function, this was done in Excel using the LN() function. In Excel, the LN() function calculates the natural logarithm of a given number. The natural logarithm is the logarithm to the base "e," where "e" is the mathematical constant approximately equal to 2.71828 (Agresti and Finley 2008). To reassess for normalcy, a Wilks-Shapiro test was conducted on each of the variables, summary statistics were calculated and univariate data visualizations were created for the log transformed variables in R.studio with the ggplot2 package (Kassambara 2019). The results of the Wilks-Shapiro tests were used to reassess the data's distribution prior to further analysis.

3.3.4 Multivariate Data Visualization

Scatter plot matrices (SPLOMs) efficiently visualize pairwise relationships among multiple variables in a matrix format (David and Tukey 1977). They consolidate multivariate information, allowing for simultaneous comparisons across different variable pairs (Chambers et al. 1983). The increased interpretability of SPLOMs makes identifying linear and nonlinear relationships, clusters, and outliers more feasible when working with large datasets (Agresti 2002). This analysis constructed a SPLOM using the 'ggally' package and the ggpairs command in R.Studio. The resulting visualization showcases Pearson's correlation coefficients and scatterplots, enhancing the ability to discern and understand linear relationships within the data.

3.3.5 Correlation Analysis – Pearson's Correlation Coefficients

The Pearson's correlation coefficient assesses how closely the data points in a scatterplot align along a straight line, indicating whether one variable tends to increase as the other increases, decrease as the other increases, or shows no relationship. The formula to calculate Pearson's correlation coefficient is as follows:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \Sigma(y_i - \bar{y})^2}} \tag{2}$$

Where r is Pearson's correlation coefficient, x_i and y_i are individual observations from the two variables, and $\bar{x}$ and $\bar{y}$ are the mean averages of the corresponding group (Agresti & Finley 2008). The formula calculates the strength and direction of linear relationships between two variables (Pearson 1900). For the purpose of this analysis the Pearson's correlation coefficients are used to determine if a relationship exists between recidivism rates and the variables meant to reflect elements of SDT.

3.3.6 Analysis of Variance (ANOVA)

A single-factor Analysis of Variance (ANOVA) tests whether the means of two or more groups are significantly different from each other. The equation for a one-way ANOVA statistic can be expressed as follows:

$$F = \frac{mean\ sum\ of\ squares\ between}{mean\ sum\ of\ squares\ within} \tag{3}$$

For the purpose of this analysis, the ANOVA test is utilized to determine if there is a relationship between the county of release and a propensity to recidivate in Washington State from 2012-2022.

3.3.7 One Sample T-Test

After the initial ANOVA test was employed to assess whether there were significant differences in recidivism rates among counties, a pairwise T-test was executed within RStudio on the recidivism rates per county to both verify the findings of the ANOVA and identify potential individual counties that exhibit significant deviations from the sample mean. The t-test, when used to compare each county's recidivism rate to the sample mean rate, is a univariate test that examines whether each county's rate of recidivism significantly deviates from the overall population mean. The t-test approach treats each county's recidivism rate independently and doesn't consider potential differences between groups. It is useful in identifying the statistically significant rates of recidivism that fall above and below the sample mean. The results of this test will be leveraged to verify the results of the ANOVA, and to identify individual counties that may deviate from the population mean that could not be identified in an ANOVA.

Utilizing the Base R package and the t.test() function the T-test was used to compare each county's recidivism rate to the average rate of all counties. The T-statistic measures how many standard errors the county's rate deviates from the mean rate of all counties. This statistic indicates the extent to which the county's mean varies from the average rate of all counties. The sign (+/-) of the T-statistic denotes the direction of this variance. The P-value of the test reveals whether the difference from the collective average rate holds statistical significance. A one-sample T-test is specified as:

$$t = \frac{sample\ mean - population\ mean}{(Sample\ Std\ Dev./\sqrt{Sample\ Size}} \tag{4}$$

3.3.8 Bonferroni Correction

A Bonferroni Correction is a statistical correction made to control the familywise error rate when multiple comparisons or hypothesis tests are conducted simultaneously. It is a method used to mitigate the problem of inflated Type I error rates that can occur when conducting multiple statistical tests on the same dataset (Agresti & Finley 2008). To implement the correction, the significance level (alpha) is divided by the number of tests being conducted.

3.3.9 Cohens d – Calculating Effect Size

After determining whether there were significant differences in the mean rates of recidivism across the 30 counties in the analysis via ANOVA and pairwise T-tests, the effect size of these differences was calculated to determine the extent to which those differences are influenced by the County of release. This was achieved by utilizing Cohens *d* which is specified as:

$$d = \frac{(\bar{x}_1 - \bar{x}_2)}{s} \tag{5}$$

Where d is the effect size, $\bar{x}_1$ is the mean of the first(dependent) group, $\bar{x}_2$ is the mean of the second (independent) group, and s is the pooled standard deviation of both groups (Cohen 1988; Rosenthal and Rosnow 1991). The pooled standard deviation is calculated as:

$$S = \frac{\sqrt{(n_1-1)S_1^2+(n_2-1)S_2^2}}{n_1+n_2-2} \tag{6}$$

Where S is the pooled standard deviation, n_1 is the sample size of the first group and $n_2$is the sample size of the second group, and s_1 and s_2 are the standard deviations of their respective groups (Rosenthal & Rosnow 1991). Cohen's d provides a measure of the standardized effect size, where larger values of d indicate a larger difference between the group means. The results of the analysis are usually interpreted based on common threshold *d* values and significance is interpreted based on corresponding p-values. The results of the analysis can be read based on these threshold *d* values:

Small effect size:	$d \approx 0.2$
Medium effect size:	$d \approx 0.5$
Large effect size:	$d \approx 0.8$

3.4 Exploratory Spatial Data Analysis

To gain a deeper understanding of the underlying spatial patterns and dependencies in the dataset, an Exploratory Spatial Data Analysis was performed. This study utilized global spatial autocorrelation measures, followed by an assessment of local clusters and spatial autocorrelation. All of this is meant to provide a more meaningful understanding of the data, and to help answer the underlying research question of whether the county of release is related to recidivism rates for those incarcerated in WADOC facilities from 2012-2022.

The Moran's I test is used to assess spatial autocorrelation, which measures the degree of similarity or dissimilarity between the values of neighboring locations across a global study area. In this case, the Moran's I test was applied to the Recidivism rates to assess the underlying spatial patterns of the dataset. This study further explored the spatial patterns by employing local spatial autocorrelation measures. These local measures provided a more detailed examination of clustering or dispersion at the individual observation level, shedding light on specific areas with notable high-value hotspots or low-value coldspots.

3.4.1 Spatial Weights Matrix (SWM)

When assessing spatial autocorrelation of recidivism rates at the county scale, it is crucial to establish a Spatial Weights Matrix (SWM), which represents the neighborhood relationships in a matrix format. The selection of an appropriate SWM should align with the underlying spatial processes at play.

To identify the optimal distance threshold for this assessment, the study drew insights from previous studies in academic literature that employed similar mathematical methodologies for crime or recidivism analysis. These prior research efforts often incorporated a distance decay function to model the movement patterns of formerly incarcerated individuals (Piquero, Farrington, and Blumstein, 1999; Hipp, 2010). In addition, some studies have used the median Census Tract size to determine a plausible distance threshold (Hipp, 2010). This analysis utilizes a combination of those two elements. To further cross validate these findings a K-nearest neighbors SWM was also utilized.

3.4.2 Moran's I Index for Spatial Autocorrelation

To examine the presence of spatial autocorrelation, the Moran's I index was utilized. The Moran's I test for spatial autocorrelation measures the degree of spatial clustering or dispersion in the data and assesses whether recidivism rates in neighboring counties were more or less similar than expected under a random distribution (O'Sullivan and Unwin 2010). Positive values of Moran's I indicate spatial clustering, while negative values suggest spatial dispersion. The Global Moran's I statistic is defined as:

$$I = \frac{n}{\sum_{i=1}^{n}(x_i - \bar{X})^2} \times \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(x_i - \bar{X})(x_j - \bar{X})}{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}} \quad (7)$$

where x is the attribute value at a given location, $\bar{X}$ is the mean attribute value across the study area, w_{ij} is the weight of the distance between location *i* and its neighbor *j*, and *n* is the number of features within a specified distance from location *i* (O'Sullivan and Unwin 2010).

3.4.3 Getis-Ord Hot Spot Analysis

The Getis-Ord G_i^*statistic, a local spatial autocorrelation technique, was employed to identify spatial clusters (hot spots) of high and low recidivism rates across Washington State. This analysis considered the county-level recidivism data and explored whether certain areas exhibited significant clustering patterns. Getis and Ord have both alluded to the benefits of leveraging both the Moran's I and Getis-ord G_i^* statistics in tandem when analyzing spatial datasets (Getis and Ord 1992). Both global and local measures of spatial autocorrelation measure different things and can point to different underlying spatial processes and patterns (O'Sullivan and Unwin 2010). The global Moran's I measure of spatial autocorrelation measures across an entire study area without distinguishing between concentrations of high and low values, while the Getis OrdG_i^* statistic allows for that distinction to be analyzed. This gives a more nuanced finer resolution look into the local clusters of hotspots and coldspots that may arise across the entire study area. The Getis-Ord G_i^* is specified as:

$$G_i * (d) = \frac{\Sigma_{j-1}^{n} w_{ij}(\mathrm{d}) x_j - \bar{x}\Sigma_{j-1}^{n} W_{ij}(\mathrm{d})}{\frac{\sqrt{n\sum_{j=1}^{n} w_{ij}^2(d) - \left(\Sigma_{j}^{n}{=}1 w_{ij}(d)\right)^2}}{n-1}} \quad (8)$$

Where w_{ij} *(d)* is the spatial weight between observation *i* and observation *j* at distance *d.* In our case using a contiguity based SWM the distance would be substituted for a binary value based on whether observations *i* and *j* are neighbors. X_j is the attribute value for the J^{th} location. $\bar{x}$ is the mean of the attributes for all values. s is the standard deviation of the attribute values from all locations, n and is the total number of observations (Getis and Ord 1992). The result of a Getis-ord G_i^* hotspot analysis are G_i^*values which are essentially z-scores. Positive values indicate clustering of high values while negative values indicated clustering of low values. The magnitude of the z-score indicates the intensity of clustering (Getis and Ord 1992).

3.4.4 Interpretation of Findings

Finally, the statistical findings were interpreted to draw meaningful conclusions regarding the correlation between the county of release and recidivism likelihood in Washington State. The results from the various statistical methods were synthesized to identify significant associations, spatial patterns, and potential areas of intervention.

Chapter 4 Results

This chapter describes the results of the analytical workflow described in Chapter 3. It begins with the results of the EDA and moves on to the results of the ESDA.

4.1 Results of Exploratory Data Analysis

This study conducted a comprehensive analysis of the summary statistics for the rate of recidivism across counties. Additionally, it delved into key variables associated with social disorganization, a factor that has been linked to recidivism rates in previous research. The primary focus was on determining if the county of release is correlated with a propensity to recidivate among WADOC incarcerated individuals from 2012 to 2022. By examining essential variables such as ‘Public assistance enrollment per 100k’ and the rate ‘Lacking basic prose literacy skills per 100k’, another objective is to extract meaningful insights into the intricacies of recidivism and the prospective role of social disorganization as a predictor. Each summary statistic reveals the central tendencies, variability, and distributional properties of these variables, allowing us to draw more informed conclusions regarding the potential relationship between the county of release and recidivism rates.

4.1.1 Baseline descriptive statistics

The total recidivism rate for all WADOC Formerly Incarcerated Individuals from 2012-2022 is 41.0028%. Specifically, of 45,731 incarcerated individuals, 18,751 were reconfined in WADOC operated facilities between 2012-2022.When broken down by County, the rate of recidivism highlights unique patterns across the study area (Table 3). These rates highlight unique patterns and trends that further enhance understanding of the phenomenon of recidivism in Washington State. Notably, the range, approximately 35.71% to 58.06%, highlights significant heterogeneity in reoffending behavior across counties. These rates underscore the nuanced nature of recidivism and suggest that localized factors may be contributing to the observed variations. Upon closer examination of the data, several counties stand out with small sample sizes and outlier recidivism rates. San Juan County, for instance, records the highest rate of 58.06% with a sample size of 31 total releases, and Wahkiakum County at 50.00% has a sample size of only 18 total releases. On the other end of the spectrum, Ferry County has the lowest rate at 35.71% and a sample size of only 28 total releases. While these disparities may reflect distinct social, economic, and environmental contexts within each county, to run a meaningful statistical analysis to determine whether the county of release is related to recidivism outcome, adequate sample sizes are necessary. To determine an adequate number of total releases the weighted average of the Cohens *d* values was input into a power analysis as the estimated effect size.

Table 3 Crosstab of recidivism vs. total releases with derivative rates of recidivism by county.

COUNTY	**ADAMS**	**ASOTIN**	**BENTON**	**CHELAN**	**CLALLAM**	**CLARK**	**COLUMBIA**
RECIDIVATED YES	39	79	616	254	172	1210	15
TOTAL RELEASES	84	184	1565	565	405	2828	39
PERCENT RECIDIVATED	0.4643	0.4293	0.3936	0.4495	0.4246	0.4278	0.3846
COUNTY	GRAYS HARBOR	ISLAND	JEFFERSON	KING	KITSAP	KITTITAS	KLICKITAT
RECIDIVATED YES	353	89	32	3579	832	90	46
TOTAL RELEASES	927	206	80	8941	1973	175	102
PERCENT RECIDIVATED	0.3807	0.4320	0.4	0.4002	0.4216	0.5142	0.4509
COUNTY	PIERCE	SAN JUAN	SKAGIT	SKAMANIA	SNOHOMISH	SPOKANE	STEVENS
RECIDIVATED YES	2889	18	407	17	1504	1711	104
TOTAL RELEASES	7070	31	950	39	3625	4215	234
PERCENT RECIDIVATED	0.4086	0.5806	0.4284	0.4358	0.4148	0.4059	0.4444
COUNTY	FERRY	FRANKLIN	GARFIELD	GRANT	MASON	OKANOGAN	PACIFIC
RECIDIVATED YES	10	235	4	295	208	127	74
TOTAL RELEASES	28	559	8	681	540	300	165
PERCENT RECIDIVATED	0.3571	0.4203	0.5	0.4331	0.3851	0.4233	0.4484
COUNTY	LEWIS	LINCOLN	WHITMAN	THURSTON	WAHKIAKUM	WHATCOM	WALLA WALLA
RECIDIVATED YES	413	9	16	739	8	594	158
TOTAL RELEASES	1026	25	44	1850	16	1416	382
PERCENT RECIDIVATED	0.4025	0.36	0.3636	0.3994	0.5	0.4194	0.4136
COUNTY	**COWLITZ**	**DOUGLAS**	**YAKIMA**	PEND OREILLE			
RECIDIVATED YES	654	76	1057	18			
TOTAL RELEASES	1676	192	2538	47			

PERCENT RECIDIVATED	0.3902	0.3958	0.4164	0.3829			

4.1.2 Power analysis

To ascertain a suitable sample size of total releases for subsequent statistical analyses, a power analysis was performed. One of the decisions that needs to be made when conducting a power analysis is on the chosen statistical power of the analysis. Statistical power is the probability that a study will correctly detect a true effect or relationship if it exists in the population. To settle on a statistical power of 80% for the power analysis in this study, a literature review within the field of recidivism and crime studies was conducted. While the literature revealed limited studies that explicitly detailed power analysis parameters, a common practice emerged in human and behavioral science studies, as well as related research within this field, where a statistical power of 80% was employed (Travers et al. 2021; Nesset et al. 2020). This choice also aligns with the conventional approach adopted in most research domains, ensuring a reasonable balance between the probability of detecting true effects and the risk of making a Type I error.

The next decision when conducting a power analysis involves estimating the effect size—a crucial parameter that gauges the magnitude of the phenomenon under investigation. In this study, we opted for a robust approach by utilizing Cohen's d to derive a weighted average effect size, often referred to as the "Pooled *d*," across all 39 counties.

The calculation of the Pooled *d* involves summing the weighted Cohen's *d* values and subsequently dividing this sum by the total of the weights. The results revealed that the Pooled *d* equates to approximately -0.0291039964 (Table 4). In more concrete terms, this signifies that the estimated effect size is characterized as a small negative effect, providing a nuanced understanding of the impact being examined in our research.

Table 4 Results of the weighted average pooled d for effect size.

COUNTY	COHENS D	VARIANCE	WEIGHTS	WEIGHTED COHENS D	STD DEV
ADAMS	0.9357	0.0171	58.4230	54.6678	0.0544
ASOTIN	0.1389	0.0054	182.2412	25.3184	0.0364
BENTON	-0.6738	0.0007	1275.4633	-859.4108	0.0123
CHELAN	0.6010	0.0020	478.5530	287.6437	0.0209
CLALLAM	0.0342	0.0024	404.7631	13.8456	0.0245
CLARK	0.1070	0.0003	2811.8863	301.0310	0.0093
COLUMBIA	-0.8786	0.0355	28.137497	-24.7242	0.1153
COWLITZ	-0.7512	0.0007	1307.1741	-981.9568	0.0119
DOUGLAS	-0.6237	0.0062	160.7350	-100.2534	0.0352
FERRY	-1.5047	0.0761	13.1324	-19.7609	0.1428
FRANKLIN	-0.0636	0.0017	557.8686	-35.5280	0.0208
GARFIELD	1.7484	0.3160	3.1638	5.5318	0
GRANT	0.2277	0.0015	663.79	151.1544	0.0189
GRAYS HARBOR	-0.9652	0.0015	632.4158	-610.4092	0.0159
ISLAND	0.2003	0.0049	201.9451	40.4688	0.0345
JEFFERSON	-0.5281	0.0142	70.2095	-37.0778	0.0547
KING	-0.5212	0.0001	7871.5478	-4103.2290	0.0051
KITSAP	-0.0340	0.0005	1971.8542	-67.2205	0.0111
KITTITAS	2.0740	0.01800	55.5424	115.1951	0.0377
KLICKITAT	0.6329	0.0117	84.9782	53.7862	0.0492
LEWIS	-0.4711	0.0010	923.4843	-435.1358	0.0153
LINCOLN	-1.4387	0.0813	12.2852	-17.6750	0.14
MASON	-0.8650	0.0025	392.9725	-339.9345	0.0209
OKANOGAN	0.0023	0.0033	299.9991	0.7004	0.0285
PACIFIC	0.5760	0.0070	141.5211	81.5200	0.0387
PEND OREILLE	-0.9151	0.0301	33.1284	-30.3164	0.1170
PIERCE	-0.3323	0.0001	6700.0373	-2226.5506	0.0058
SAN JUAN	3.5833	0.2393	4.1777	14.9704	0.0806
SKAGIT	0.1184	0.0010	943.3831	111.7337	0.0160
SKAMANIA	0.2891	0.0267	37.4347	10.8254	0.0641
SNOHOMISH	-0.1888	0.0002	3561.4605	-672.7455	0.0081

SPOKANE	-0.3937	0.0002	3911.7103	-1540.3775	0.0075
STEVENS	0.4826	0.0047	209.5846	101.1641	0.0324
THURSTON	-0.5394	0.0006	1614.9835	-871.2608	0.0113
WAHKIAKUM	1.7484	0.15803	6.3277	11.0637	0
WALLA WALLA	-0.2184	0.0026	373.0945	-81.5178	0.0251
WHATCOM	-0.0841	0.0007	1411.0013	-118.7700	0.0131
WHITMAN	-1.3567	0.0436	22.9117	-31.0859	0.1363
YAKIMA	-0.1524	0.0003	2508.838	-382.5251	0.0097

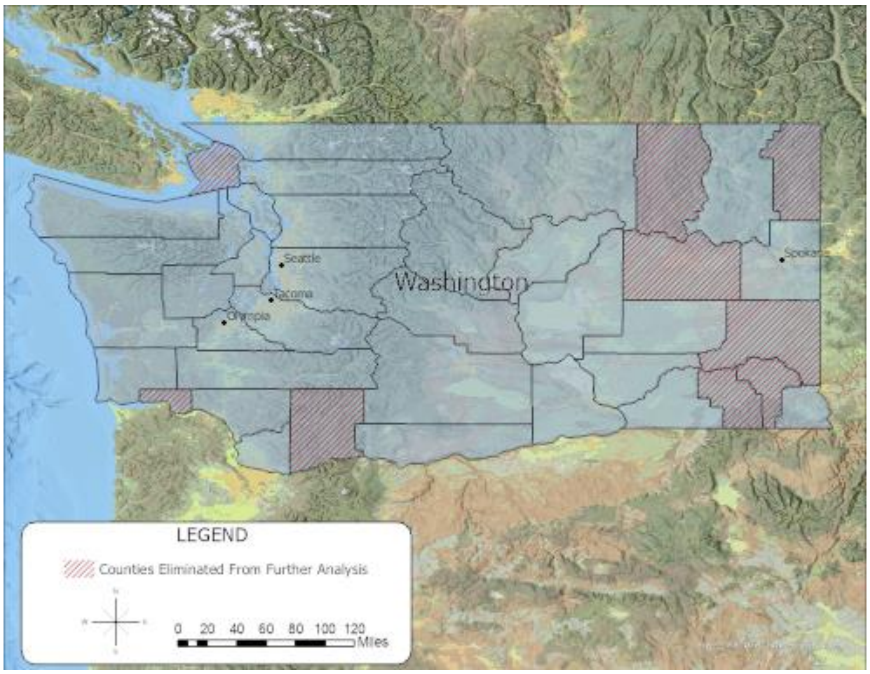
$$\text{Pooled } d = \frac{\text{SUMWEIGHTED}}{\text{SUMWEIGHTS}}$$
$$\text{Pooled } d = \frac{-12206.84499}{41942.16093}$$
$$\text{Pooled } d \approx -0.291039964$$

In crafting a rigorous power analysis for this study, essential parameters were carefully considered. The significance level was set at 0.05 and aimed for a statistical power of 0.80, ensuring a strong likelihood of detecting true effects. Our analysis included 39 distinct counties, each with unique characteristics. The estimated effect size, Pooled d ≈ -0.0291, indicates a small negative effect. These parameters establish a robust foundation aligned with our study's objectives. The analysis determined that a total number of releases sample size of n=2403 yields a power of 0.80. This implies that for a balanced one-way ANOVA, each group should have a minimum of 62 observations or total releases, providing a power of 0.8035 (Figure 5).

Figure 5 Results of a power analysis to determine sample size.

4.1.3 Eliminating the counties with insufficient sample sizes

After determining the sample size threshold of 62 total releases, the counties that exhibited less than that were eliminated from further analysis. Those counties include Columbia with 39 total releases, Ferry with 28, Garfield with 8, Lincoln with 25, Pend Oreille with 47, San Juan with 31, Skamania with 39, Wahkiakum with 16, and Whitman with 44.



COUNTIES ELIMINATED FROM ANALYSIS	
Columbia	San Juan
Ferry	Skamania
Garfield	Wahkiakum
Lincoln	Whitman
Pend Oreille	

Figure 6 Map of the counties that were removed from further analysis.

4.1.4 Summary statistics

To delve deeper into the characteristics of the data, the summary statistics of each variable were calculated and examined (Table 5). The analysis of recidivism rates reveals a considerable degree of variability. With a standard deviation of approximately 2.72%, the data's dispersion is significant. A positively skewed distribution (skewness ≈ 1.364) suggests a tendency toward higher rates of recidivism, while a peaked distribution (kurtosis ≈ 3.448) indicates the presence of extreme values. The 95% confidence interval (±0.1335) provides a reliable estimate of the mean recidivism rate, showing a diverse range of post-release experiences.

The average enrollment in public assistance programs is approximately 36,485.06 people per 100k population, indicating a moderate level of enrollment. Substantial variability is evident, with a standard deviation of approximately 3,721.79, reflecting differing levels of need and economic circumstances across regions. A mildly negatively skewed distribution (skewness ≈ -0.57) suggests regions with higher enrollment rates, indicative of reliance on public assistance. The distribution's relatively low kurtosis (≈ 0.84) indicates a moderate presence of extreme values or outliers.

The rate of individuals lacking basic prose literacy skills is approximately 12,466.67 people per 100k, indicating a moderate level of individuals with this challenge. The distribution exhibits a positively skewed pattern (skewness ≈ 2.07), implying areas with higher rates of those lacking literacy skills. A wide range of rates (from 6,000 to 34,000) underscores significant disparities in literacy levels across regions. A moderately high kurtosis value (≈ 3.91) indicates a distribution with a notable presence of extreme values or outliers.

In summary, these variables' distributions demonstrate significant variability, skewed patterns, and the presence of extreme values. These irregularities in the distributions prompt consideration of data transformation. Specifically, the highly positively skewed distribution and peaked kurtosis observed in literacy rates underscore the potential presence of extreme values. Addressing these irregularities through transformation may allow for more accurate analysis, interpretation, and identification of potential underlying factors.

Table 5 Summary statistics.

Variable	**Mean**	**Standard Error**	**Median**	**Standard Deviation**	**Sample Variance**	**Kurtosis**	**Skewness**	**Range**	**Minimum**	**Maximum**	**Sum**	**Confidence Lvl. 95.0%**

Recidivism Rates	0.4213	0.0049	0.4199	0.0272	0.0007	3.4477	1.3644	0.1335	0.3808	0.5143	12.6399	0.0102
Enrolled in Public Assistance per 100k	36485.06	679.5024	36854.47	3721.788	13851706	0.8362	-0.5725	16698.2	27081.01	43779.21	1094552	1389.72
Lacking Basic Prose Literacy Skills per 100k	12466.67	1264.669	10000	6926.876	47981609	3.9079	2.0717	28000	6000	34000	374000	2586.52

4.1.5 Univariate Data Visualization

Both histograms and box plots were leveraged to assess the data for normality and to identify outliers.

4.1.5.1 Rate of recidivism

The analysis of the histograms and boxplots revealed interesting patterns in the distributions of various variables. The rate of recidivism data reveals that the majority of counties have recidivism rates falling within the range of approximately 0.39 to 0.43 (Figure 7). This central cluster suggests consistency in these counties' rates of recidivism. There is a single county outlier with a high recidivism rate greater than 0.50. The outlier emphasizes the diversity in recidivism rates among the counties in Washington State.

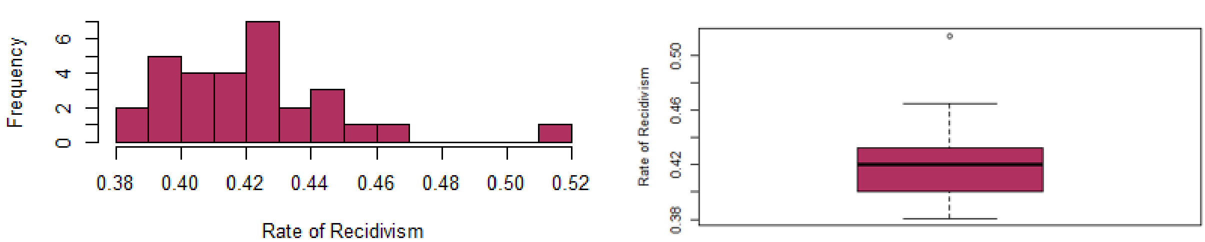


Figure 7 Histogram and boxplot of the rate of recidivism by county.

4.1.5.2 The rate enrolled in public assistance per 100k population

The distribution of the rate enrolled in public assistance per 100k population across the 30 counties in Washington State reveals a range of frequencies within distinct enrollment rate intervals. Notably, most counties exhibited enrollment rates primarily falling within two intervals of approximately 27,000 to 30,500 and 37,000 to 40,500. This suggests consistency in enrollment patterns among these counties (Figure 8). Moreover, two counties stand out with substantially lower rates, approximately 27,000 and 30,500, highlighting potential socioeconomic disparities among counties in the State.

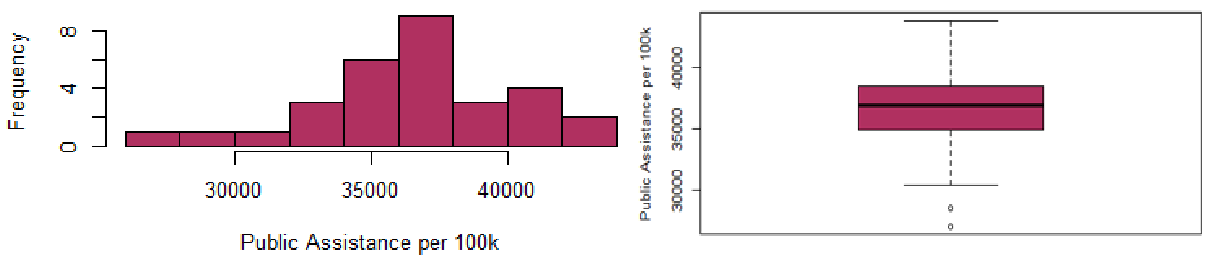


Figure 8 Histogram and boxplot of the rate enrolled in public assistance per 100k.

4.1.5.3 Lacking basic prose literacy skills per 100k population

The rate lacking basic prose literacy skills per 100k population variable exhibited positive skewness in the distribution of the histogram as well. This indicates that most counties have a rate lacking literacy between 5,000 and 15,000 people per 100k. Furthermore, there are substantial outliers present in the dataset that tail off to the right side of the distribution. This means that the tail of the distribution is elongated towards the higher values more than that of a normal distribution. This is further confirmed by visually inspecting the boxplot and histogram (Figure 9). The boxplot appears to show that there are four substantial outliers in the data above the sample mean. Based on the summary statistics and the univariate visual analysis on this variable, transformation is required.

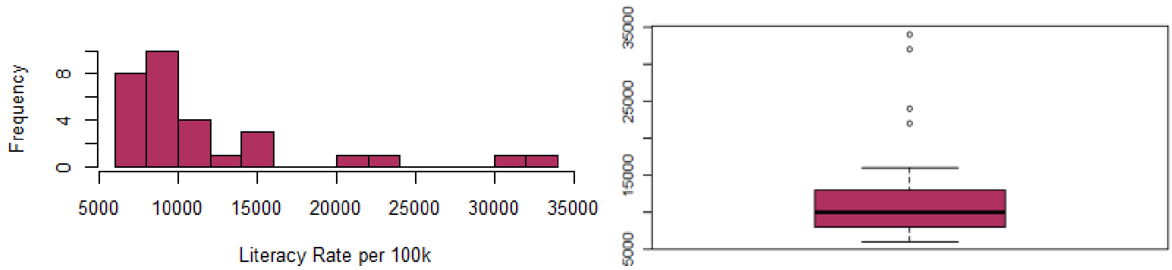


Figure 9 Histogram and boxplot of the rate lacking basic prose literacy skills per 100k.

4.1.6 Data Transformation

Based on the skewness values and the results of the univariate data visualizations, the rate lacking basic prose literacy skills per 100k exhibited a moderate to highly skewed distribution which required transformation. The symmetry of the variable's distribution improved significantly after a natural logarithmic transformation (Figure 10).

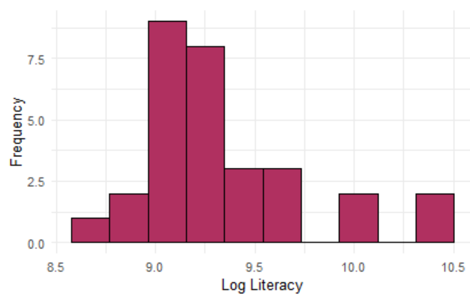


Figure 10 Histogram of the log transformed variable - highlighting the improved symmetry.

Prior to logarithmic transformation, the individuals lacking basic prose literacy skills per 100k exhibited a positively skewed distribution with a kurtosis value of 3.908 and a skewness of 2.072. However, after applying the logarithmic transformation, these distribution characteristics improved, resulting in reduced kurtosis (1.082) and skewness (1.227) values. Although it still displayed positive skewness, the distribution became more balanced. The improvement in distributional symmetry was confirmed through visual inspection of the histograms and the results of the proceeding Shapiro-Wilks tests.

4.1.7 Reevaluating Data Distribution for Normality

The Shapiro-Wilks test was employed to assess the normality of the variables' distributions, a critical consideration for the proceeding statistical analyses. While the test resulted in a p-value of 0.01623, which is below the typical alpha threshold of 0.05, suggesting statistical significance, it's crucial to look beyond this. The test's W-statistic for the rate of recidivism is 0.91149, indicating a distribution that is relatively close to normal (Table 6). There is a pronounced outlier above a 50% recidivism rate that could be impacting the p-value. This nuanced perspective suggests that while there may be some departure from perfect normality, the distribution is within a range that is compatible with the assumptions of ANOVA. Therefore, despite the statistical significance, the rate of recidivism distribution exhibits characteristics that align reasonably well with the requirements for ANOVA, supporting its applicability for subsequent analyses.

The public assistance enrollment per 100k population variable displayed a higher W-statistic of 0.96346 with a p-value of 0.3786. In this case, the p-value is not statistically significant, implying that the distribution is more consistent with a normal distribution. It is important to note that some departure from normality may still exist.

The logarithm of the rate lacking basic prose literacy skills per 100k exhibited a W-statistic of 0.8794, accompanied by a small p-value of 0.002732. While this p-value suggests statistical significance and some departure from perfect normality, it's essential to note that the W-statistic of 0.8794 is relatively high. Compared to the W-statistics of other variables in the study, which are mostly at or above 0.90, this value indicates that the deviation from normality, while statistically significant, is not as pronounced as it may seem at first glance. This means that while there is evidence of non-normality, the distribution of the logarithm of those lacking basic prose literacy skills per 100k is not dramatically far from a normal distribution, and it remains within a range that is suitable for statistical analyses.

Table 6 Results of the Shapiro-Wilks tests.

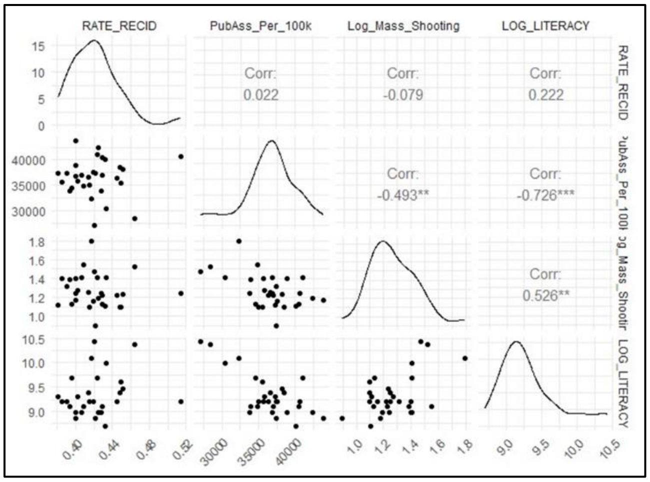
Variable	**Rate of Recidivism**	**Public Assistance**	**Log Lacks Literacy**
W-statistic	0.9115	0.9635	0.8794
P-value	0.0162	0.3786	0.0027

4.1.8 Multivariate Data Visualization and Pearson's Correlation Coefficients

This study conducted a thorough scatterplot analysis to examine the connections between recidivism and county contextual variables that align with prior research grounded in Social Disorganization Theory. The key contextual variables included the public assistance enrollment per 100k, and the rate lacking basic literacy prose per 100k (log transformed). The results yielded intriguing insights. Notably, the Pearson's Correlation Coefficients, which found no significant correlations between the rate of recidivism and the other variables, indicating an absence of linear relationships. Based on these results it appears that while these variables have contributed to a positive correlation to crime and recidivism in other studies conducted elsewhere, in Washington State, these variables do not contribute to a propensity to recidivate at the County scale of analysis.

The only noteworthy findings that surfaced in the form of robust and statistically significant correlations were seen in the variables linked to social disorganization. Notably, a strong and statistically significant negative correlation was evident between public assistance enrollment per 100k and the rate lacking basic prose literacy skills per 100k with a Pearson's correlation coefficient of -0.726 ($p < 0.001$). This correlation suggests that as levels of public assistance enrollment increase, the rate lacking literacy tends to decrease. This is a counterintuitive finding that would require further analysis to dissect, which is outside the realm of this study. While these findings are interesting, they show no relevance to our analysis.

Table 7 Scatterplot matrix with density plots and Pearson's correlation coefficients.



4.1.9 Single-Factor Analysis of Variance (ANOVA)

The one-sample ANOVA test compared the distribution of recidivism rates among different counties to determine if there are statistically significant differences in these rates based on the county of release. The test statistic is calculated to be nearly zero for the between group variation. For a one-sample ANOVA, the test statistic is used to determine whether the observed differences between the sample mean and the population mean are statistically significant across groups. The p-value associated with the F-statistic is 1. This p-value indicates the likelihood of observing these differences due to random chance alone.

Table 8 Results of the single factor ANOVA.

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	5.55E-17	29	1.91E-18	1.39E-16	1	1.8474
Within Groups	0.4142	30	0.0138			
Total	0.4142	59				

Since the p-value is very high (1), it suggests that there isn't enough evidence to conclude that the differences in recidivism rates among the counties are statistically significant. In other words, based on this test, there is insufficient evidence to reject the null hypothesis, which states that there are no differences in recidivism rates among the counties. In this case, the high p-value suggests that the differences observed are likely due to random variation rather than the effect of the county of release.

4.1.10 Two-tailed one sample T-test and Cohen's d

The t-test, when used to compare each county's recidivism rate to the sample mean rate, is a univariate test that examines whether each county's rate of recidivism significantly deviates from the overall population mean. The t-test approach treats each county's recidivism rate independently and doesn't consider potential differences between groups. It is useful in identifying the statistically significant rates of recidivism that fall above and below the sample mean. The results of this test were leveraged to verify the results of the ANOVA, and to identify individual counties that may deviate from the population mean that could not be identified in an ANOVA (Table 9).

Table 9 Results of the T-test and Cohen's *d*.

COUNTY	RATE RECID	STD DEVIATION	SQRT SAMPLESIZE	T STATISTIC	P_VALUE	BONFERRONI	COHENS_D	COHENS P_VALUE
ADAMS	0.4643	0.0544	9.1652	7.237454	0.0000	0.0017	0.7897	0.5680
ASOTIN	0.4293	0.0365	13.5647	2.96269	0.0060	0.0017	0.2184	0.1726
BENTON	0.3936	0.0124	39.5601	-88.8291	0.0000	0.0017	-2.2454	0.9751
CHELAN	0.4496	0.0209	23.7697	32.1089	0.0000	0.0017	1.3508	0.8227
CLALLAM	0.4247	0.0246	20.1246	2.7612	0.0098	0.0017	0.1372	0.1090
CLARK	0.4279	0.0093	53.1789	37.5528	0.0000	0.0017	0.7062	0.5198
COWLITZ	0.3902	0.0119	40.9390	-106.958	0.0000	0.0017	-2.6126	0.9909
DOUGLAS	0.3958	0.0353	13.8564	-10.0235	0.0000	0.0017	-0.7234	0.5296
FRANKLIN	0.4204	0.0209	23.6432	-1.0532	0.3009	0.0017	-0.0445	0.0355
GRANT	0.4332	0.0190	26.0960	16.3132	0.0000	0.0017	0.6251	0.4679

GRAYS HARBOR	0.3808	0.0159	30.4467	-77.3736	0.0000	0.0017	-2.5413	0.9887
ISLAND	0.432	0.0345	14.3527	4.43722	0.0001	0.0017	0.3092	0.2424
JEFFERSON	0.4	0.0548	8.9443	-3.48317	0.0015	0.0017	-0.3894	0.3019
KING	0.4003	0.0052	94.5569	-383.8	0.0000	0.0017	-4.0586	0.9999
KITSAP	0.4217	0.0111	44.4185	1.47826	0.1501	0.0017	0.0333	0.0265
KITTITAS	0.5143	0.0378	13.2288	32.5528	0.0000	0.0017	2.4608	0.9851
KLICKITAT	0.451	0.0493	10.0995	6.08198	0.0000	0.0017	0.6022	0.4516
LEWIS	0.4025	0.0153	32.0312	-39.3949	0.0000	0.0017	-1.2299	0.7809
MASON	0.3852	0.0209	23.2379	-40.0917	0.0000	0.0017	-1.7253	0.9149
OKANOGAN	0.4233	0.0285	17.3205	1.19614	0.2413	0.0017	0.0691	0.0550
PACIFIC	0.4485	0.0387	12.8452	9.01407	0.0000	0.0017	0.7017	0.516
PIERCE	0.4086	0.0058	84.0833	-183.085	0.0000	0.0017	-2.1774	0.9705
SKAGIT	0.4284	0.0161	30.8221	13.5728	0.0000	0.0017	0.4404	0.3402
SNOHOMISH	0.4149	0.0082	60.2080	-47.3078	0.0000	0.0017	-0.7857	0.5679
SPOKANE	0.4059	0.0076	64.9230	-132.4399	0.0000	0.0017	-2.0400	0.9585
STEVENS	0.4444	0.0325	15.2971	10.8640	0.0000	0.0017	0.7102	0.5217
THURSTON	0.3995	0.0114	43.0116	-82.4552	0.0000	0.0017	-1.9170	0.944
WALLA WALLA	0.4136	0.0252	19.5448	-5.9959	0.0000	0.0017	-0.3068	0.2408
WHATCOM	0.4195	0.0131	37.6298	-5.2511	0.0000	0.0017	-0.1395	0.1109
YAKIMA	0.4165	0.0098	50.3786	-24.8665	0.0000	0.0017	-0.4936	0.3783

Initially, the pairwise T-tests revealed deviations in 25 out of the 30 counties under examination, which raised both questions and interest. To gain a better understanding of the significance of these deviations, this study calculated effect sizes (Cohen's d) and the associated p-values. To address the challenge of multiple comparisons associated with the analysis of 30 counties, a Bonferroni correction was applied to the alpha level. Following this correction, none of the observed Cohen's d values remained statistically significant at the conventional alpha level of 0.05 (Only two did prior to correction). In essence, the fact that the deviations were not significant after the Bonferroni correction supports the null hypothesis – that the differences in rates across counties are not substantial enough to be attributed to the counties themselves. Instead, it is more likely that these observed differences are a result of omitted variables or random variation.

4.1.11 Conclusion

This study utilized both ANOVA and t-tests, along with effect size calculations using Cohen's d. The initial ANOVA test was conducted to explore whether significant differences exist in recidivism rates among various counties. Initial results were not statistically significant. The analysis suggests that, based on the available data, the county of release alone does not exert a statistically significant impact on recidivism rates. This finding indicates that other factors, such as individual characteristics, socioeconomic conditions, or more localized environmental factors, may exert more substantial influences on recidivism rates within Washington State.

Understanding the spatial distribution of these rates could still reveal spatial clusters or patterns of high and low values, highlighting counties with unique characteristics influencing recidivism rates. Incorporating further spatial visualization techniques, such as spatial autocorrelation analysis and hotspot mapping, enables a deeper understanding of the geographic aspects of the phenomenon. By exploring the spatial dimension, we can uncover spatial dependencies, spatial clusters of similar values, and potential spatial patterns related to recidivism and its contributing factors across counties. Such analyses can serve as a basis for targeted interventions and policy decisions to effectively address recidivism and improve public safety in different regions.

4.2 Results of Exploratory Spatial Data Analysis

This sections describes the results of the ESDA.

4.2.1 Spatial Weights Matrix

The selection of the Spatial Weights Matrix (SWM) for this analysis is underpinned by a rationale grounded in prior research findings. Existing literature has suggested the presence of a distance decay function for formerly incarcerated individuals (Piquero, Farrington and

Blumstein, 1999; Pyle, 1974). Other studies have leveraged the use of the median area of the scale of analysis (Hipp, 2010).

This study utilized this same methodology to derive a distance based SWM. First the median square mileage of the counties in the analysis was found to be 1,719 square miles. To estimate the approximate distance across a square or rectangular area with an area of 1,719 square miles, a straightforward method involves calculating the square root of the given square mileage. In this case, taking the square root of 1,719 yields a value of approximately 41.49 miles. To further refine this measurement into meters, a simple conversion can be applied. By multiplying the 41.49 miles by the conversion of approximately 1,609.34 meters per mile, the approximate distance is transformed into meters, resulting in a rounded value of 66,734 meters across. Thus, the choice of a SWM is a fixed distance band threshold set to 66,734 meters.

To cross validate the findings of the distance based SWM when utilized in the Getis Ord Hot Spot Analysis a K-nearest neighbors algorithm was also utilized with the neighbor parameter set to 3. It is also important to note that alternative contiguity based SWMs were initially considered. However, they were eliminated from further consideration due to a specific challenge posed by the dataset. The removal of counties with small numbers of total releases resulted in the creation of isolated island counties, rendering the contiguity based SWMs useless for this analysis (one county, Asotin, would have no neighbor).

4.2.2 Moran's I Statistic – Measuring Global Spatial Autocorrelation

The assessment of Moran's Index for recidivism rates for each county revealed insightful spatial patterns. The calculated Moran's Index value of 0.029836 suggests a slight positive spatial autocorrelation, indicating that, to a small extent, areas with higher rates of recidivism tend to be located near similar values. The Expected Index, which was found to be -0.034483, provides a baseline against which the actual Moran's Index can be compared. It represents the anticipated value of Moran's Index under the assumption of no spatial autocorrelation. The p-value of 0.754924, exceeding the significance level of 0.05, does not provide enough evidence to reject the null hypothesis of complete spatial randomness. Thus, the analysis suggests that there is no evidence of spatial autocorrelation in the distribution of recidivism cases among counties. Further investigation and consideration of the underlying spatial patterns are warranted to comprehend the complex dynamics influencing recidivism rates across Washington State.

Table 10 Results of the Moran's I index for spatial autocorrelation on the recidivism rates.

Global Moran's I Summary	
Moran's Index	0.0298
Expected Index	-0.0345
Variance	0.0425
z-score	0.3122
p-value	0.7549

4.2.3 Getis-Ord Statistics - Local Cluster Analysis

Overall, the local spatial autocorrelation analysis using the Getis-Ord G_i^* statistic provided deeper insights into the distribution and spatial patterns of recidivism rates at the county level. These results offer a more nuanced understanding of the spatial dynamics of recidivism in Washington State at the County level. The initial hot spot analysis conducted using the Getis-Ord G_i^* statistic in ArcGIS Pro with a distance band parameter of 66,734 meters, observed a statistically significant hotspot at the 99% confidence interval of recidivism rates in Kittitas County (Figure 11). The calculated Z-score of 3.48 reflects a substantial clustering of high recidivism rates compared to its neighboring areas. This strong positive Z-score indicates that the observed high rates are not just a product of random chance but represent a statistically significant hot spot. Additionally, the low p-value of 0.0005 underscores the significance of this spatial clustering, which is well below the conventional significance threshold of 0.05. While these findings highlight Kittitas County as a possible hotspot for recidivism rates in the broader context of the study area, it also invites further examination into the underlying factors contributing to this localized phenomenon. It is abnormal, but not uncommon to find no global spatial autocorrelation, yet local clusters at such a high level of significance.

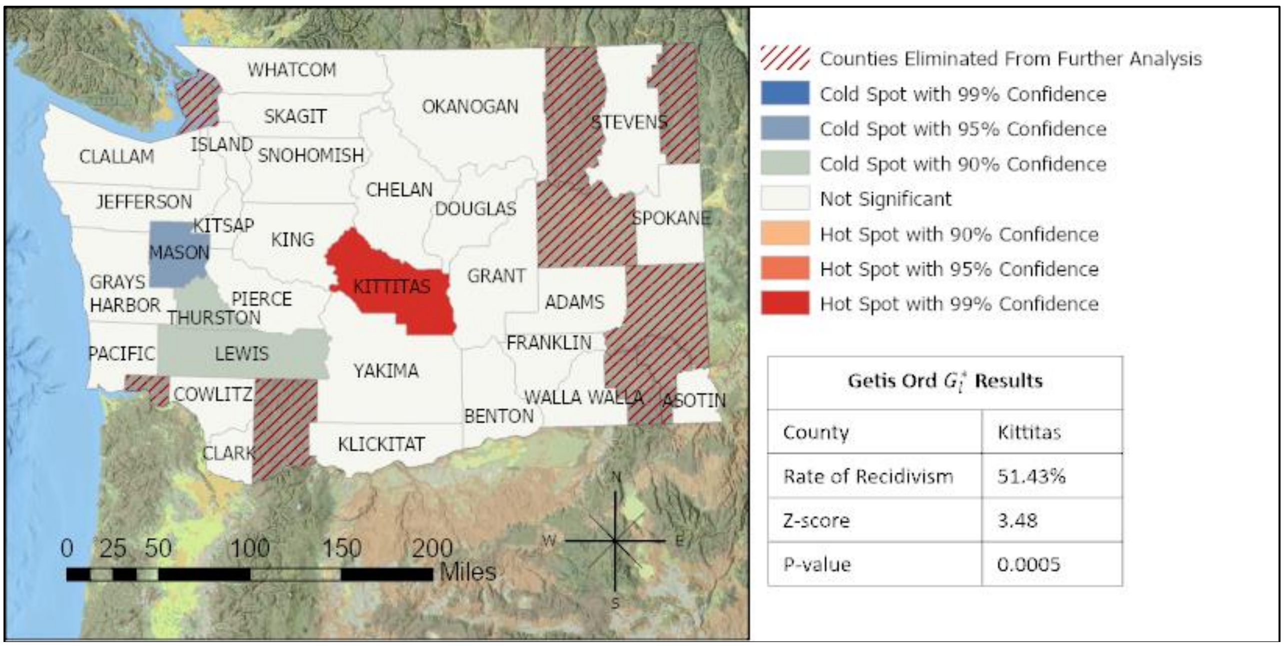


Figure 11 Results of the 66,734-meter fixed distance Getis-ord G_i^* Hot Spot Analysis

To cross-validate whether the finding in Kittitas County is an artifact, or indeed a hotspot of recidivism rates, a further Getis-ord G_i^* Hot Spot Analysis was conducted using the K-nearest neighbors algorithm utilizing 3 as the number of neighbors parameter (Figure 12). The results of this analysis resulted in a Z-score of 2.53 and a P-value of .0114, which cross-validates the findings of the distance based SWM. The only notable change detected was the hot spot in Kittitas County which was initially detected at the 99% confidence threshold, but when using the K-nearest neighbors SWM changed to the 95% confidence threshold.

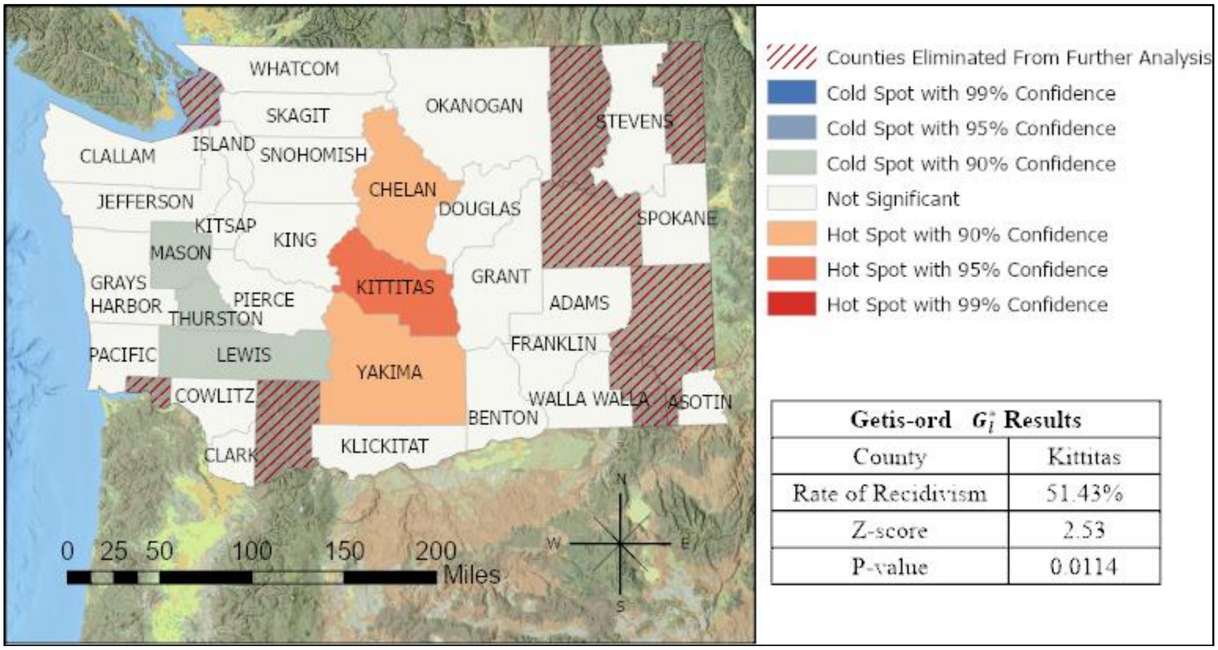


Figure 12 Outcome of Getis-ord G_i^* Hot Spot Analysis utilizing the K-nearest neighbors parameter set to 3

4.3 Conclusion

Throughout this comprehensive exploratory data analysis (EDA) and exploratory spatial data analysis (ESDA), we have gained valuable insights into the dynamics of recidivism and its potential underlying correlates with the county of release as the predictive factor. Our findings have provided a deeper understanding of the spatial distribution and patterns of various variables that have been found to be related to recidivism in other studies conducted elsewhere, in Washington State, laying the groundwork for further research and testing.

In the EDA, we analyzed multiple variables, including the rate of recidivism across counties, the rate of enrollment in public assistance programs per 100k, and the rate of those lacking basic prose literacy skills per 100k (log transformed). Notably, no statistically significant correlation was found between the rate of recidivism per county and either of the county contextual variables meant to reflect social disorganization. Additionally, the outcomes of the

ANOVA test which were then verified with a T-test and Cohen's *d* do not provide sufficient evidence to assert that the variations in recidivism rates among the different counties are statistically significant. In simpler terms, the county of release does not exhibit a significant correlation with the propensity to recidivate in Washington State from 2012-2022.

The ESDA delved into spatial autocorrelation, highlighting noteworthy spatial patterns in the data. While the Moran's I index for spatial autocorrelation revealed no global spatial autocorrelation, further measures of local spatial autocorrelation revealed a statistically significant cluster of recidivism in Kittitas County. Upon closer examination utilizing an alternative SWM it may be reasonably inferred that the clustering present is due to the low variance in the recidivism data. Thus, what the Getis-ord G_i^* Hot Spot Analysis might be detecting as hotspots are not traditional hotspots or cold spots of high or low values, but rather clusters of areas with similar rates. This could be because, with low variance, even small deviations from the mean could be statistically significant (as was found in the T-tests), leading to the identification of clusters that are not necessarily extreme but are consistently different from the overall mean.

In conclusion, this extensive EDA and ESDA has tested whether the county of release is associated with a propensity to recidivate, whether county social disorganization contextual factors influence recidivism rates in Washington State, and finally, assessed the spatial patterns of the phenomenon across Washington State. The findings of this study suggest that the County scale of analysis is too broad an areal unit of aggregation to provide the detailed nuances of the neighborhood context for crime or recidivism. Although the study offered valuable insights into the general effectiveness of WADOC-operated facilities between 2012 and 2022, it fell short in capturing the diverse nature of recidivism patterns across different microenvironmental locations.

Chapter 5 Discussion

The preceding chapters have explored the nuances of our research goals, the methodology utilized, and the outcomes achieved. As we transition into this chapter, our focus shifts from the analytical framework to the theoretical, where the findings of the study are summarized in terms of real-world implications. It should be noted that the results of this analysis are specific to Washington State and should not be applied to other jurisdictions.

5.1 Significant Findings

The primary findings of this analysis are threefold. First, the county of release of WADOC formerly incarcerated individuals from 2012-2022 showed no relationship with recidivism status. This was verified by the outcomes of the ANOVA test in combination with the subsequent T-test and Cohens *d* analyses. Second, the absence of a statistically significant correlation between recidivism and the social disorganization contextual variables implies that these factors, as measured in this study, do not appear to be strong drivers of recidivism rates in Washington State counties. This finding challenges the idea that concentrated disorganization, which has been found to correlate with crime and recidivism in some other jurisdictions, plays a similar role in explaining recidivism patterns in Washington State at this scale of analysis. Lastly, the findings of this study suggest that the County scale of analysis is too broad an areal unit of aggregation to provide the detailed nuances of the neighborhood context for crime or recidivism. Although the study offered valuable insights into the general effectiveness of

WADOC-operated facilities between 2012 and 2022, it fell short in capturing the diverse nature of recidivism patterns across different microenvironmental locations.

Translating our findings into real-world terms, the analysis suggests that recidivism is not substantially impacted by the geographical factor of the county of release. This outcome underscores the importance of considering the spatial scale of analysis in studies of this nature. It's noteworthy to acknowledge that existing academic literature has revealed significant correlations between social disorganization indicators of crime and recidivism and increased crime and recidivism rates in other jurisdictions. This contrasts with our findings and implies that the county scale, while useful for certain analyses, might be too broad of an aggregate to capture the nuanced nature of crime, which frequently occurs within subsections of cities or regions that could be more accurately expressed through finer-grained spatial units like census tracts, blocks, zip codes, or neighborhoods.

Revisiting the literature discussed in sections 1 and 2, Hipp (2007) emphasizes the importance of considering scale when investigating crime and other neighborhood effects such as recidivism. Drawing on data from a specific subset of the American Housing Survey, his study delved into how various variables impact crime and disorder at both the block and census tract levels. Notably, his research uncovers that the influence of these variables varies according to the level of aggregation. For example, racial/ethnic diversity consistently influenced perceptions of disorder, regardless of the aggregation level in this study, while economic resources exhibit only localized effects at the block level, demonstrating disparities in their impact on crime and disorder. This underscores the significance of researchers contemplating aggregation schemes when examining crime and recidivism especially from a social disorganization theory perspective.

Within the broader context, numerous studies exploring the influence of neighborhood structural factors on various outcomes, such as recidivism, often struggle to define the "neighborhood." Often, researchers poll household responses within a specific geographic unit or employ statistical estimates to create such measures, without considering whether this unit aligns with the outcome or predictor variables. Studies frequently use diverse geographic units, such as blocks, block groups, tracts, zip codes, or amalgamations of multiple tracts, as proxies for the neighborhood, without a determination of the appropriate level of aggregation. Hipp's study addressed exactly that, determining the correct geographic level of aggregation for use in studies measuring or assessing crime and recidivism.

The author warns that in cases where the unit of analysis is too large, researchers risk inadvertently including geographic units that contain multiple neighborhoods. For instance, constructs like social disorder and crime result from the aggregation of individual instances—each dilapidated building or piece of litter contributes to physical disorder, and each additional crime event contributes to the overall crime rate. Therefore, the question then arises regarding which is the "most" suitable geographic unit for aggregating these instances when constructing a neighborhood measure of crime or disorder. This dilemma is pertinent to all studies, regardless of their methodology for measuring such. Employing too high a level of aggregation carries the risk of merging crime and disorder rates from different neighborhoods into a larger unit, often masking meaningful relationships.

The study that we conducted at the county scale was a prime example of what Hipp was warning readers about when aggregating at too high of a level, the phenomenon of masking most of the heterogeneity of recidivism rates within the larger areal unit of analysis impacted the results of the analysis. However, by solely focusing on the WADOC operated facilities for purposes of this analysis, the results do speak to the efficacy of the Department of Correction's ran facilities at rehabilitating the population of incarcerated individuals that they are responsible for.

This study found that the average rate of recidivism for all WADOC incarcerated individuals released from prison between 2012-2022 was 41% (Figure 13).

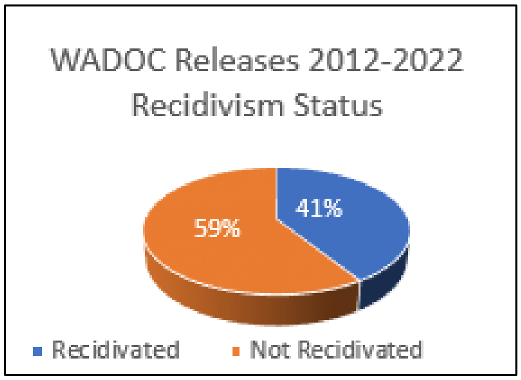


Figure 5 Pie chart of statewide WADOC recidivism rates (2012-2022).

When comparing the findings of this study to the national average, the most often-cited statistic is that within three years of release, roughly 67.8% of released prisoners were rearrested (U.S. Bureau of Justice Statistics). Considering this study specifically gauged reconfinement rather than rearrest as that of the BJS, the findings present a consistent narrative with broader national trends.

5.2 Methodological Barriers

Throughout the course of this study, methodological challenges emerged from the initial state of the raw dataset as it was received. The process of overcoming these obstacles was a time-intensive endeavor, involving months dedicated to meticulously preparing the data for analysis. Yet, the significance of these barriers extended beyond data transformation, significantly shaping the study's overall framework.

The data received from the WADOC was limited to the county scale of analysis due to privacy concerns in the public records process. This limitation caused by the way the data is reported imposed certain parameters that significantly constrained this project's ability to perform various spatial statistical analyses that could have potentially unveiled the microenvironmental nuances typically associated with crime and recidivism. These limitations are commonly known as the modifiable areal unit problem (MAUP).

In essence, the MAUP acknowledges the role of the chosen scale when aggregating data, as it can significantly impact the results of any analysis. Eliminating much of the MAUP as a constraint to our spatial analysis would require individual level point data, which, unfortunately, is not feasible due to the previously alluded to data availability limitations and restrictions imposed by the public records request process. Thus, in this project, we have effectively highlighted the impact of the MAUP; as arbitrary data aggregates become larger, many of the finer and more nuanced spatial relationships become obscured as they are aggregated within the confines of the larger unit while they were collected at a much finer scale. This limitation underscores the importance of recognizing and accounting for the MAUP in spatial analyses.

Another barrier imposed by the data was that limited sample sizes (total releases from WADOC prison facilities) prevented the inclusion of all 39 counties in Washington State, which reduced the level of detail attainable from the analysis. Additionally, the data received from the WADOC did not include any demographic information that could have been leveraged for a more comprehensive analysis of their impact on recidivism in Washington State. This limitation restricted our ability to employ many multivariate data analysis methods, especially regression analysis.

The absence of demographic information ultimately forced a shift in the project's focus from exploring contributing contextual factors to recidivism (especially the context of the geography to which the formerly incarcerated individual is released) to a more specific investigation into whether the county of release is correlated with recidivism rates. Thus, the data limitations imposed required an adaptation of research questions.

Certainly, if the dataset had contained additional information, such as the specific prison locations upon release, gender, age, race, and crime type, would have undeniably enriched the dataset, offering a more robust and comprehensive analysis of recidivism factors.

Beyond these variables, numerous other demographic and contextual factors could have significantly enhanced the depth of the analysis. These might encompass aspects such as an individual's educational background, including educational attainment, which could have facilitated an examination of the impact of education on recidivism rates rather than utilizing literacy as a measure of education. Likewise, exploring their employment history, both before and after release, would have provided valuable insights into the role of employment in reducing recidivism which is commonly found to be a substantial contributor to recidivism in other geographies.

Moreover, considering marital and family status, such as whether individuals were married, had dependents, or maintain family support systems, could have allowed for a more intricate understanding of the influence of family dynamics and peer-support networks on recidivism. Information related to substance abuse history, including previous treatment and relapse rates while incarcerated, could have been instrumental in studying the connection between substance abuse and recidivism, as well as the efficacy of programs offered while incarcerated. Additionally, data on mental health history, such as diagnoses, treatment, and access to mental health services, could have been invaluable in assessing the impact of mental health on recidivism. Integrating these additional demographic and contextual variables into the analysis would have fostered a more comprehensive understanding of the multifaceted factors contributing to recidivism rates in Washington State.

If this dataset had been available at a more nuanced scale such as the census block or census tract level, the study would have been far better positioned to gain a deeper understanding of the spatial components of recidivism. The intricacies that remained obscured when analyzed at the county scale could have potentially been unveiled had the data been accessible at a more fine-grained, microenvironmental level. It's worth noting that some states, such as California, offer access to higher precision geographical data upon public request. Assessing whether there have been any adverse consequences from the release of such data in other jurisdictions could help evaluate the risks associated with making finer-scale data more accessible in Washington State.

5.3 Future Research

In essence, the outcomes of this study emphasize the significance of context and granularity in the examination of recidivism. They underscore the necessity for future research to delve into the effects of more localized factors and spatial scales, ultimately leading to a deeper comprehension of the intricate dynamics influencing recidivism patterns within Washington State. Furthermore, the finding of a statistically significant hotspot of recidivism rates in Kittitas County suggests that further research could focus on this county specifically when assessing recidivism.

Future studies aiming to delve into more nuanced patterns of recidivism would greatly benefit from accessing higher-resolution data, which would enable a more detailed analysis at a smaller geographic scale, such as neighborhoods or streets. This finer granularity could reveal critical insights into localized environmental or social factors influencing recidivism rates, which are not discernible at broader geographic levels. However, this approach presents a challenge, as currently the most detailed data available through public records requests is at the county level. Gaining access to more granular data would require navigating privacy concerns and data sharing policies, as well as potentially collaborating with government agencies or criminal justice organizations that have the capacity to provide such detailed information. This access is crucial for future research to effectively explore and understand the complex spatial dynamics underpinning recidivism patterns, thereby informing more targeted and effective intervention strategies.

Additionally, future research could delve into the geographical histories of individuals who have been incarcerated. This exploration would encompass not only their residential locations prior to incarceration but also the specific prison locations and the sites of their release. Analyzing how these geographical factors correlate with recidivism outcomes could be particularly enlightening. For example, the study might investigate whether the proximity of the prison to the individual's home community or the characteristics of the release site have any bearing on the likelihood of reoffending. Such research would offer a deeper understanding of the complex spatial dynamics that underlie recidivism patterns, potentially leading to more effective and targeted intervention strategies.

Even when the relationships observed in EDA and ESDA do not show strong or statistically significant correlations as was the case in this study, further multivariate analyses including regression modeling could help to strengthen the findings of this study. Regression can unveil hidden and nonlinear relationships that might not be evident through basic correlation analysis. It can help uncover subtle and complex associations between variables that EDA and ESDA might not capture. Furthermore, regression allows for hypothesis testing regarding the significance of individual predictor variables. This can provide evidence for or against the importance of specific factors, regardless of the strength of correlations. Given these considerations, it is recommended that future research extends the findings of this study by employing advanced modeling techniques to evaluate the interplay of variables of social disorganization more comprehensively. In this same capacity, future research may incorporate more readily available measures of social disorganization into an analysis, or perhaps create a composite index of social disorganization to test against the recidivism rates derived from this analysis.

References

Agresti, A. 2002. *Categorical Data Analysis*. Hoboken, NJ: John Wiley & Sons.

Agresti, A., Finlay, B. 2008. *Statistical methods for the Social Sciences*. 4th ed. Pearson Prentice Hall.

Almgren, Gunnar, A. Guest, G. Immerwahr, and M. Spittel. 1998. "Joblessness, Family Disruption, and Violent Death in Chicago, 1970-90." *Social Forces* 76, no. 4: 1465.

Anselin, L. 1995. "Local Indicators of Spatial Association-LISA." Geographical Analysis 27, no. 2: 93-115.

Aos, S., Miller, Mama, and Drake, E. 2006. "Evidence Based Adult Corrections Programs: What Works and What Does Not." Olympia: Washington State Institute for Public Policy.

Aos, S. 2003. "Washington's Offender Accountability Act: Update and Progress Report on the Act's Evaluation". Washington State Institute for Public Policy Publications. 1-12. https://www.wsipp.wa.gov/ReportFile/822/Wsipp_Washingtons-Offender-Accountability-Act-Update-and-Progress-Report-on-the-Acts-Evaluation_Full-Report.pdf

Average Daily Population of Incarcerated Individuals Fiscal Year 2022, 2023. Washington State Department of Corrections Publications https://doc.wa.gov/docs/publications/reports/400-RE002-2206.pdf

Beckett, K.A., A.M. Harris, and H. Evans. 2008. *The Assessment and Consequences of Legal Financial Obligations in Washington State: Research Report*. Olympia, WA: Washington State Minority and Justice Commission.

Besag, J. "Spatial Interaction and the Statistical Analysis of Lattice Systems." Journal of the Royal Statistical Society: Series B (Methodological) 36, no. 2 (1974): 192–225.

Bursik, R.J. 1988. "Social Disorganization and Theories of Crime and Delinquency: Problems and Prospects." Criminology, 26(4), 519-551.

Cahill, M., and Mulligan, G. 2007. "Using Geographically Weighted Regression to Explore Local Crime Patterns." Social Science Computer Review 25, no. 2: 174-193.

Camasso, M.J., Jagannathan, and Radha. 1995. "Prediction accuracy of the Washington and Illinois risk assessment instruments: An application of receiver operating characteristic curve analysis." Social Work Research, 19, no. 3: 174–183, https://doi.org/10.1093/swr/19.3.174

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Earlbaum Associates.

David, F.N., and Tukey, J.W. 1977. "Exploratory Data Analysis." Biometrics 33, no. 4 (1977): 768-769. https://doi.org/10.2307/2529486.

Drake, E. 2014. "Predicting Criminal Recidivism: A Systematic Review of Offender Risk Assessments in Washington State." Olympia, WA: Washington State Institute for Public Policy.

Drawve, G., and McNeeley, S. 2021. "Recidivism and Community Context: Integrating the Environmental Backcloth." Journal of Criminal Justice 73: 101786. https://doi.org/10.1016/j.jcrimjus.2021.101786.

Fotheringham, A.S. 1997. "Local Statistics for Spatial Analysis." In Quantitative Geography: Perspectives on Spatial Data Analysis, edited by Fotheringham, A.S., Jordan, N., and Rogerson, P., 175-197. London, UK: Sage Publications Ltd.

Gagliardi, GJ., D. Lovell, P.D. Peterson, and R. Jemelka. 2004. "Forecasting Recidivism in Mentally Ill Offenders Released from Prison." *Law and Human Behavior* 28, no. 2: 133–155.

Geary, R.C. 1954. "The Contiguity Ratio and Statistical Mapping." The Incorporated Statistician 5, no. 3: 115.

Gerben J.N. Bruinsma, Lieven J.R. Pauwels, Weerman, F.M., Bernasco, W. 2013. "Social Disorganization, Social Capital, Collective Efficacy and the Spatial Distribution of Crime and Offenders: An Empirical Test of Six Neighbourhood Models for a Dutch City." The British Journal of Criminology 53:5, 942–963. https://doi.org/10.1093/bjc/azt030

Getis, A., and Ord, J.K. 1992. "The Analysis of Spatial Association by Use of Distance Statistics." Geographical Analysis 24, no. 3 (1992): 189-206.

Goodchild, M.F. 2007. "Citizens as sensors: The world of volunteered geography."

Gottfredson, S.D., and Taylor, R.B. 1988. "Community Contexts and Criminal Offenders." In Communities and Crime Reduction, edited by Hope, T., and Shaw, M., 62-82. London: Her Majesty's Stationary Office.

Heitgerd, J.L., and R.J. Bursik. 1987. "Extracommunity Dynamics and the Ecology of Delinquency." American Journal of Sociology 92, no. 4: 775–787.

Hipp, J.R. 2007. "Block, Tract, and Levels of Aggregation: Neighborhood Structure and Crime and Disorder as a Case in Point." American Sociological Review 72, no. 5: 659-680. https://doi.org/10.1177/000312240707200501.

Kassambara, A. 2019. *Practical statistics in R II - comparing groups: Numerical variables*. Datanovia.

Kim, R.H., and Clark, D. 2013. "The Effect of Prison-Based College Education Programs on Recidivism: Propensity Score Matching Approach." Journal of Criminal Justice 41, no. 3: 196-204. https://doi.org/10.1016/j.jcrimjus.2013.03.001.

Kirk, D.S., and J.H. Laub. 2010. “Neighborhood Change and Crime in the Modern Metropolis.” Crime and Justice 39, no. 1: 441–502.

Kubrin, C.E., and R. Weitzer. 2003. “Retaliatory Homicide: Concentrated Disadvantage and Neighborhood Culture.” Social Problems. 50: 157–180.

Kubrin, C.E. and Stewart, E.A. 2006. “Predicting Who Reoffends: The Neglected Role of Neighborhood Context in Recidivism Studies.” Criminology, 44: 165-197. https://doi.org/10.1111/j.1745-9125.2006.00046.x

Kubrin, C.E., Squires, G.D., and Stewart, E.A. 2007. "Neighborhoods, Race, and Recidivism: The Community-Offender Nexus and Its Implications for African-Americans." Race Relations Abstracts 32: 7-37.

LeBeau, J.L. & M. Leitner. 2011. “Introduction: Progress in Research on the Geography of Crime.” The Professional Geographer, 63:2, 161-173, DOI: 10.1080/00330124.2010.547147

Leitner, M., Barnett, M., Kent, J., and Barnett, T. 2011. “The Impact of Hurricane Katrina on Reported Crimes in Louisiana: A Spatial and Temporal Analysis.” The Professional geographer 63, no. 2: 244–261.

Levine, N. and Block, R. 2011. “Bayesian Journey-to-Crime Estimation: An Improvement in Geographic Profiling Methodology.” The Professional geographer 63, no. 2: 213–229.

Leymon, M.G., Campbell, C.M., Henning, K., and Renauer, B.C. 2022. "The Impacts of Length of Prison Stay on Recidivism of Non-Violent Offenders in Oregon." Journal of Criminal Justice 82: 102002. https://doi.org/10.1016/j.jcrimjus.2022.102002.

Logan, J.R., and Brian J. Stults. 1999. “Racial Differences in Exposure to Crime: The City and Suburbs of Cleveland in 1990*.” *Criminology* 37, no. 2: 251–276.

Lovell, D., Johnson, L.C., & Cain, K.C. 2007. “Recidivism of Supermax Prisoners in Washington State.” Crime & Delinquency. 53 no. 4, 633–656. https://doi.org/10.1177/0011128706296466

Maher, E.J., Gerlinger, J., Wood, A.D. et al. 2022. “Won’t You be My Neighbor? Neighborhood Characteristics Associated with Mass Shootings in the USA.” Race Soc Probl 14, 223–237. https://doi.org/10.1007/s12552-021-09350-3

Manchak, S. M., J. L.Skeem, and K.S. Douglas. 2008. “Utility of the Revised Level of Service Inventory (LSI-R) in Predicting Recidivism after Long-Term Incarceration.” *Law and Human Behavior* 32, no. 6: 477–488.

Malleson, N. and Andresen, M.A. 2015. “Spatio-Temporal Crime Hotspots and the Ambient Population.” Crime Science 4, no. 1.

Maltz, M. 2001. *Recidivism*. Orlando, FL: Academic Press Incorporated.

Mennis, J., and Diansheng G. 2009. “Geographic Data Mining and Knowledge Discovery - An Introduction.” Computers, Environment and Urban Systems: 403–408.

Mennis, J, Harris, P.W.,Obradovic,Z., Izenman,A.J., Grunwald, H.E. & Lockwood, B. 2011. The Effect of Neighborhood Characteristics and Spatial Spillover on Urban Juvenile Delinquency and Recidivism, The Professional Geographer, 63:2,174-192, DOI: 10.1080/00330124.2010.547149

Miller, M., Drake, E., & Nafziger, M. 2013. What Works to Reduce Recidivism by Domestic Violence Offenders? (Document No. 13-01-1201). Olympia: Washington State Institute for Public Policy.

Moran, P.A.P. 1948. “The Interpretation of Statistical Maps.” Journal of the Royal Statistical Society. Series B (Methodological) 10, no. 2: 243-251.

Nesset, M., Berg, M., Loret, L.C., Bjørngaard, J., Whittington, R. and Palmstierna, T. 2020. “Cognitive Behavioural Group Therapy versus Mindfulness-Based Stress Reduction Group Therapy for Intimate Partner Violence: A Randomized Controlled Trial.” BMC Psychiatry 20, no. 1:5.

Nicodemo, C., & Satorra, A. 2022. “Exploratory data analysis on large data sets: The example of salary variation in Spanish Social Security Data.” BRQ Business Research Quarterly, 25(3), 283–294. https://doi.org/10.1177/2340944420957335

Openshaw, S. 1984. “Ecological fallacies and the analysis of areal census data.” Environment and Planning A 16, 17–31.

Openshaw, S. and Taylor, P. 1979. *A million or so correlation coefficients: three experiments on the modifiable areal unit problem.* Statistical Applications in the Spatial Sciences. Edited by N. Wrigley. Pion, London. 127–144.

Ostertagova, E. & Ostertag, Oskar & Kováč, Jozef. 2014. “Methodology and Application of the Kruskal-Wallis Test.” Applied Mechanics and Materials. 611. 115-120.

O'Sullivan, D., and Unwin, D. 2014. *Geographic Information Analysis*. 2nd ed. Hoboken, NJ: Wiley.

Pearson, K. 1900. "On the Criterion That a Given System of Deviations from the Probable in the Case of a Correlated System of Variables is Such That it can be Reasonably Supposed to Have Arisen From Random Sampling." Philosophical Magazine Series 5, vol. 50, no. 302, pp. 157-175.

Petersilia, J. 2009. *When Prisoners Come Home: Parole and Prisoner Reentry*. Oxford, UK: Oxford University Press.

Phipps, P.A., K. Korinek, S. Aos, and R. Lieb. *Research Findings on Adult Corrections Programs: A Review*. Olympia, WA: Washington State Institute for Public Policy, 1999.

Piquero, A.R., Farrington, D.P., and Blumstein, A. 2003. "The Criminal Career Paradigm." Crime and Justice: A Review of Research 30, edited by Tonry, M., 359-506. Chicago, IL: The University of Chicago Press.

Ratcliffe, J. 2004. "The hotspot matrix: A framework For the spatio-temporal targeting of crime reduction." Police Practice and Research 5:5–23.

Recidivism of Adult Felons 2007, 2008, Washington State Caseload Forecast Council. https://cfc.wa.gov/sites/default/files/Publications/Adult_Recidivism_FY2007.pdf

Robinson, W.S. 1950. "Ecological Correlations and the Behavior of Individuals." International Journal of Epidemiology 38, no. 2: 337-341.

Rosenthal, R. & Rosnow, R.L. 1991. *Essentials of behavioral research: Methods and data analysis*. New York: McGraw Hill.

Sampson, R.J., S. W. Raudenbush, and Felton Earls. 1997. "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy." Science 277, no. 5328: 918–924.

Shaw, C.R., and H. D. McKay. *Juvenile Delinquency and Urban Areas*. Chicago, IL: Univ of Chicago Press, 1942.

Stahler, G.J., Mennis, J., Belenko, S., Welsh, W.N., Hiller, M.L., and Zajac, G. 2013. "Predicting Recidivism for Released State Prison Offenders." Criminal Justice and Behavior. 40, no. 6 pg. 690–711. https://doi.org/10.1177/0093854812469609.

Tobler, W. 1997. "The First Law of Geography." Economic Geography 47(1). Pgs. 1-7.

Travers, Áine, T. McDonagh, T. Cunningham, C. Armour, and M. Hansen. 2021. "The Effectiveness of Interventions to Prevent Recidivism in Perpetrators of Intimate Partner Violence: A Systematic Review and Meta-Analysis." Clinical Psychology Review 84: 5.

Unwin, D.J. 1996. "GIS, Spatial Analysis, and Spatial Statistics." Progress in Human Geography 20(4): 540-551. doi:10.1177/030913259602000408.

Wang, X., Hay, C., Todak, N.E., and Bales, W. 2013. "Criminal Propensity, Social Context, and Recidivism." Criminal Justice and Behavior 41(3): 300–317. https://doi.org/10.1177/0093854813507055.

Washington State Department of Corrections. 2019. Improve Lives | Decrease First-Year Rate of Return to Institutions. https://doc.wa.gov/about/agency/strategic-plan-g1o1.htm

Washington Sentencing Guidelines Commission. 2007. Recidivism of Adult Felons. By the Caseload Forecast Council. cfc.wa.gov/sites/default/files/Publications/Adult_Recidivism_FY2007.pdf

Washington State Institute for Public Policy. 1999. The Community Juvenile Accountability Act: Research-Proven Interventions for the Juvenile Courts, by Robert Barnoski. Document No. 99051201. Olympia, WA. https://www.wsipp.wa.gov/ReportFile/1311/Wsipp_The-Community-Juvenile-Accountability-Act-Research-Proven-Interventions-for-the-Juvenile-Courts_Full-Report.pdf

Washington State Legislature. 1999. Offender Accountability Act (OAA). Engrossed 2nd Substitute Senate Bill 5421, Chapter 196, Laws of 1999 56th Legislature Regular Session. July, 25.