Title

Not All Eviction Cases are Alike: How Do Contextual and Individual Characteristics Matter? A Computational Analysis of a Decade of Court Cases from Pierce County, Washington

Author

Ren, Cheng

Publication Date

2024

Peer reviewed|Thesis/dissertation

Not All Eviction Cases are Alike: How Do Contextual and Individual Characteristics Matter? A Computational Analysis of a Decade of Court Cases from Pierce County, Washington

by

Cheng Ren

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Social Welfare and the Designated Emphasis

in

Computational and Data Science and Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Julian Chun-Chung Chow, Chair

Professor Emmeline Chuang

Professor David Harding

Spring 2024

Abstract

Not All Eviction Cases are Alike: How Do Contextual and Individual Characteristics Matter? A Computational Analysis of a Decade of Court Cases from Pierce County, Washington

by

Cheng Ren

Doctor of Philosophy in Social Welfare and the Designated Emphasis in Computational and Data Science and Engineering

University of California, Berkeley

Professor Julian Chun-Chung Chow, Chair

Eviction is arguably one of today's most significant and pressing issues, affecting millions of households. On average, 3.6 million eviction filings occur annually across the United States. The eviction crisis is poised to intensify in the aftermath of the COVID-19 pandemic, displacing low-income families, restricting future housing options, and potentially leading to homelessness. A significant challenge in addressing this issue is the lack of systematically collected eviction data, with a vast amount of detailed information contained in unstructured court records, primarily in PDF format. Although filing data offer insight regarding this initial step in the eviction process, the events that follow remain largely understudied. This dissertation leveraged computational social science techniques, combining social and data science, to extract these unstructured data and analyze posteviction filing outcomes. This study explored three main issues: (a) the efficacy of computational methods in extracting information from unstructured court files; (b) the influence of individual, community, and macro-level factors on dismissal or judgments; and (c) the determinants of eviction by the sheriff following court judgment. The study utilized advanced document layout analysis based on natural language processing and computer vision to recognize information in PDF files and link this personal information to broader property, community, and county datasets. A classification model was used to identify important factors related to eviction filing outcomes. The analysis covered 56,070 unique cases derived from 772,629 PDF files spanning 2004–2022 in Pierce County, Washington, demonstrating high accuracy in data extraction (median Levenshtein similarity ratio of 1 and mean of 0.95). Key findings indicated that significant individual-level factors—such as race, property sale records, legal representation, taxable property value, and response to summons—influence eviction filing outcomes. At the community level, poverty rates and the proportion of rent-burdened households emerged as strong predictors. At the macro level, a housing price index and rent prices play a crucial role. The interaction between an individual's race and the proportion of White people in the census tract shows that people of color experience different eviction filing outcomes compared to White individuals in the same community. The discussion touches on computational social science in eviction research and how variables at different levels affect eviction filing outcomes. The study findings have implications for social welfare interventions and policy, aiming to support affected families and mitigate the eviction crisis.

Acknowledgements

I would like to express my deepest gratitude to my advisor and committee chair, Dr. Julian Chun-Chung Chow, for his unwavering support throughout my PhD journey, exemplifying what it means to be a great advisor and professor. My journey would have been impossible without the guidance and feedback from my dissertation and qualifying committee members, Dr. Emmeline Chuang, Dr. David Harding, Dr. Susan Stone, and Dr. Rediet Abebe from EECS. Special appreciation goes to Dr. Tim Thomas for providing both opportunities and funding for eviction research, making this dissertation possible.

I am immensely grateful to my professors, coauthors, and supervisors—Dr. Irene Bloemraad, Dr. Marla Stuart, Dr. Valerie Shapiro, Dr. Zachary Pardos, Dr. Sonia Katyal, Dr. Claudia von Vacano, Dr. Cassondra Marshall, and Anthony Suen at UC Berkeley; Dr. Ott Toomet, Dr. Sarah Stone, Dr. Anissa Tanweer, Dr. Bryna Hazelton, Lisa Manzer, and Annie Kucklick at the University of Washington; Thomas O'Brien, April Urban, and Dr. Claudia Coulton during my time at Case Western Reserve University; and Max Griswold from the Rand Institute—for offering me opportunities to explore data science's integration with various fields.

My family has been my cornerstone; my wife, Yaoyao Cai; our child, Ian Ren; and my parents. Yaoyao has devoted so much to our family, supporting and accompanying me through challenging moments during the past decade. Ian, as a newborn during my dissertation, slightly slowed my process, but I cherished being a new parent and almost every moment with him. Ian is truly a gift. My parents' unconditional love and support have been fundamental to my achievements.

A special thanks to my PhD cohort for creating a positive learning environment, especially Dr. Brenda Mathias, Eunkyung Chung, Dr. Joshua Gregory, Valentín Q. Sierra, Dr. Ivy Hammond, and Sergio Martinez. I'm grateful to my colleagues at UC Berkeley, including Luyi Jian, Zhi Li, Leyi Zhou, Jennifer (Xin) Chen, Caedi Seim, Nehal Eldeeb, Elizabeth Pleasants, Midori Pierce, Dr. Kelly Ziemer, Dr. Juyeon Lee, and my D-Lab Fellows.

Thanks also to my friends and professionals who offered suggestions and encouragement: Peilun Wu, Huizhi Gong, Yumeng Liu, Linyu Yu, Wei Zhang, Sicong Sun, Yang (Suzy) Su, Liancheng Shen, Ye Wang, Wenjin Situ, Zheng Tang, Jia (Lucy) Chen, Yunqi Wei, Su Wang, Linyun Fu, Chaoxin Jiang, Yuanyuan Yang, Yuncheng Man, and Quan Cao.

Having moved to UAlbany as new faculty member while finishing my dissertation, I extend my sincere thanks to all my colleagues, especially Dr. Victoria Rizzo, Dr. Eunju Lee, Dr. Wonhyung Lee, Dr. Heather Larkin Holloway, Dr. Carmen Morano, Dr. Eric Hardiman, Dr. Xin Wang, and Dr. Angie Wootton, for helping me settle in Albany, NY.

I'd like to acknowledge my reliable cloud computer, named "bit by bit," which continuously ran my models while I slept. Last, I'm grateful to my editor, Eric Lindberg, for revising several manuscripts to ensure clear communication.

If I have inadvertently overlooked anyone, please know your contributions have been invaluable. Thank you so much!

Table of Contents

Chapter 1: What is Eviction?

Introduction

Eviction, the legal process for removing a tenant from a rental property, is arguably one of today's most significant and urgent issues, affecting millions of households. The eviction crisis has the potential to worsen in the aftermath of the COVID-19 pandemic in the United States, displacing low-income families, limiting future housing options, and likely leading to homelessness (Burt, 2001; Cookson et al., 2018; Greiner et al., 2013; Humphries et al., 2019). Eviction research has only started to gain significant attention in recent years, particularly after the publication of Matthew Desmond's *Evicted: Poverty and Profit in the American City* in 2016. Garnham et al. (2022) reported that from 2000 to 2018, an average of 3.6 million eviction filings occurred annually across the United States, indicating that approximately 6 of every 100 renter households faced the risk of eviction. However, public agencies have not systematically collected eviction data in accessible databases, leaving much information unused in individual case records.

The Eviction Lab has reported collecting 99 million eviction records from 30,000 county-level annual reports on eviction filings. Yet data for 33% of renter households remain uncollected (Garnham et al., 2022). Although some states or municipalities have digitized data, most eviction studies do not extend beyond community-level research due to the lack of comprehensive individual data. It is important to note that eviction filing data only cover filing activities; cases where tenants move before filing are not reflected in these numbers. Similarly, if the case is dismissed, the number of filings does not decrease. Thus, filing data provide only a snapshot of the eviction process's beginning, with subsequent events often understudied.

The absence of relevant research raises new questions, such as how many people are evicted by the sheriff and who experiences a dismissal. This is very important because not all eviction cases are alike. If the case is dismissed, it indicates an earlier resolution, potentially because the tenants had more flexibility. Meanwhile, a trend in eviction record sealing movements allows parties of dismissed cases to request restricted public access, such as Senate Bill 873 in Oregon (Hussein et al., 2023). However, if the case moves to sheriff-managed eviction, then the tenant faces the most dangerous and potentially violent part of the eviction process (A. Eisenberg & Brantley, 2023). Many tragedies on both the tenant and sheriff sides have been reported in the media, such as a sheriff's deputy in Larimer County, Colorado, fatally shooting a man while serving an eviction writ and a sheriff's deputy being shot by tenants in Alameda County, California, in 2024. Steven Strand, executive director of the Washington Association of Sheriffs and Police Chiefs, stated that serving an eviction notice is one of the most dangerous jobs (Sumrall, 2023).

Computational social science (CSS), which combines social and data science, has been increasingly recognized in recent years in social science research as the digital age has created new opportunities with social media, massive open administrative data, and online experiments, among others. Computational methods have the potential to explicate or even address social issues from poverty to health in this digital era. CSS involves the application of novel methodologies and data analytic techniques such as machine learning to behavioral and social sciences research, allowing various approaches such as automated social information extraction and complexity modeling (Cioffi-Revilla, 2010). Several organizations have recognized its power and have built workshops or curricula to meet this change. These trends indicate that computational methods combined with social sciences have become an accepted and promising approach.

This study explored how to understand eviction filings and factors influencing outcomes after filings with the application of computational methods to analyze messy, unstructured, scanned PDF files from the court system in Pierce County, Washington. Specifically, the research questions were:

1. What kinds of computational methods can be applied to extract information, such as addresses, from unstructured court files in PDF format with high accuracy?
2. What factors at the individual, community, and macro levels influence whether a case is dismissed or judged by the court?
3. What factors at the individual, community, and macro levels influence whether a case will involve a sheriff-executed eviction after judgment?

Chapter 1 serves as the introduction to this dissertation. It also provides definitions of eviction filings and sheriff evictions. Specifically, I detail which situations during the eviction process may result in a dismissal label and outline the sheriff eviction stage in the eviction process for this study.

Chapter 2 lays out the scope of the eviction problem, including its magnitude, severity, trends, and effects, along with a review of the literature to provide an understanding of individual, community, and macro factors influencing evictions. This includes examining personal circumstances like income level and job security, alongside community and economic conditions.

Chapter 3 is the methodological chapter. It identifies the role of CSS and current gaps in research regarding both methods and findings, setting the stage for this study's contributions. The theoretical framework and research questions are also described in this chapter. The chapter outlines the use of individual, community, and macro factors as independent variables and outcomes after filings and outcomes of sheriff eviction as the dependent variables. It details the research design, data collection, data linkage, and analysis procedures to respond to the research questions.

Chapter 4 presents the findings of the study, beginning with a descriptive profile of cases that were dismissed and those leading to eviction by the sheriff, including changes at each eviction stage and spatial differences. The results from the multilevel analysis are then presented, illuminating the relationship between identified factors and sheriff eviction.

Chapter 5 features an interpretation of the findings in light of the literature and theoretical framework, paying attention to what factors significantly influence outcomes after filings and sheriff eviction. It explores the significance and implications of the findings and how they can inform future practices and policies. Finally, a limitations section outlines shortcomings to this research.

Understanding Eviction

"Without a home, everything else falls apart." –Evicted (https://evictedbook.com/)

The following sections define eviction filings and sheriff's eviction and lay out the scope of the eviction problem, including its severity, effects, and trends.

Definition

The definition of eviction varies slightly across contexts. For instance, Merriam-Webster (2021) defines eviction as "the dispossession of a tenant of leased property by force or especially by legal process." In other circumstances like court files, eviction may also be known as an "unlawful detainer." The Eviction Lab at Princeton University (2024) provides a broader definition in plain language, stating, "An eviction happens when a landlord expels people from property he or she owns. Evictions are landlord-initiated involuntary moves that happen to renters, whereas foreclosures are involuntary moves that happen to homeowners when a bank or other lending agency repossesses a home." This definition highlights three key points: First, eviction typically occurs between landlords and tenants, with the process initiated by the landlord. Second, the keyword "involuntary" signifies that tenants are forced to leave. Third, although there are exceptions like foreclosures that result in homeowners being evicted, this study focused solely on evictions involving landlords and tenants.

Thus, how can researchers measure evictions, or more specifically, determine the number of evictions in a given area? When researchers count evictions, they often refer to eviction filings, indicating that landlords have submitted their eviction request to the court and officially served a summons to tenants. Some scholars might use data from the sheriff's office to count evictions, because the sheriff or marshal in certain jurisdictions executes the eviction after receiving a writ from the court. In this study, the term "eviction" refers only to cases in which the legal process has been initiated, excluding cases before the legal process began. The term "sheriff eviction" refers to evictions carried out by the sheriff. Tenants may or may not encounter the sheriff and potential violence during this process. As previously mentioned, it is important to note that some cases do not proceed through the eviction filing process if tenants move out after receiving an eviction notice[1] from their landlord. Similarly, although some cases enter the legal process, they might be withdrawn by landlords or dismissed by judges for various reasons. Therefore, the real number of eviction (i.e. moving out) exceeds that of eviction filings, and more eviction filings occur than sheriff evictions.

Eviction Process

State laws mainly regulate the eviction process, and cities, countries, or local governments may supplement state regulations in the United States (Legal Information Institute, 2021). Although details can vary, the general steps of the eviction process are similar across jurisdictions. Typically, evictions occur due to late or nonpayment of rent, violation of lease terms, or lease expiration without renewal. The process in Washington is briefly outlined as follows:

1. The landlord gives the tenant a written notice asking them to accomplish something in a particular period, like a 3-day or 14-day notice.

2. If the tenant does not vacate the premises, the landlord proceeds to court to initiate an eviction, commonly referred to as a filing. A summons is then sent to the tenant (defendant).

3. The tenant has a specific timeframe to file a response in court. Failure to respond may result in the court making a judgment without the tenant's input.

4. Before the judge makes the decision, the case can also be dismissed for various reasons. Either the landlord or tenant can file a motion to dismiss, with the court making the dismissal decision, but motion to dismiss from landlord side is much more common in eviction cases based on the observation of court files. Typically, there are three common situations: (a) The tenants paid in full or came to an agreement with the landlord, so the landlord has no need to continue the legal process; (b) the tenants have moved out and the property has been vacated; or (c) before the hearing, the filing is improper or neither side (plaintiff or defendant) appears in the court. In the third situation, such cases are typically referred to as a "clerk's dismissal" in the

[1] This eviction notice is not a writ from the judge but a notice from the landlord. If it is due to nonpayment, it is usually called a pay or vacate notice.

state of Washington's court data system to distinguish them from a "motion to dismiss" from the landlord or tenant.

5. The judge makes a decision. If the court rules in favor of the landlord, a writ of restitution is issued, allowing local law enforcement to remove the tenant from the property. The writ of restitution, sometimes called a writ for short, is a judicial document that authorizes a landlord to reclaim a rental unit from a tenant (Washington State Legislature, 1973). In Figure 1, the judgment occurs on day 24; however, this is for default judgments. Usually, a default judgment in eviction involves the tenant automatically losing the case because they failed to respond or appear in court as required—e.g., the defendant did not file an answer to the plaintiff's argument. If there is a hearing, the date of the judgment is determined by the court's schedule.

On the other hand, if the court does not rule in favor of the landlord after the hearing, then the tenant wins the case. In court paperwork, this outcome is usually still referred to as a judgment.

6. The sheriff or other law enforcement entity sends a notice to the tenant, and the plaintiff may need to schedule the eviction.

7. If the tenant still does not leave after receiving notice from law enforcement, a law enforcement officer (e.g., sheriff, marshal) will evict the tenant and their belongings.

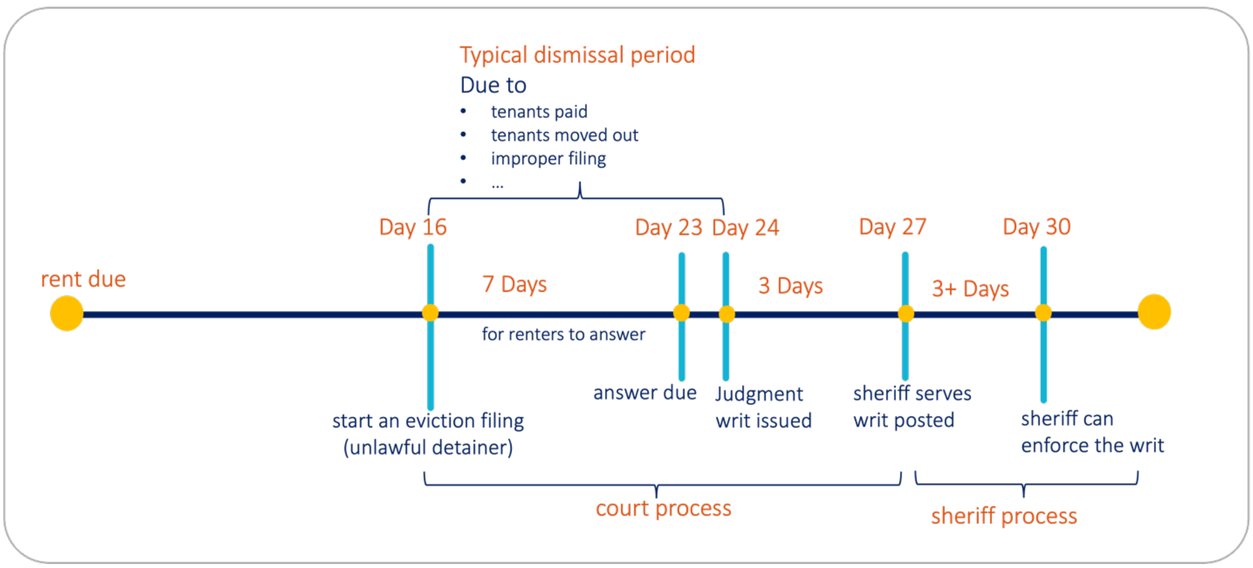


Figure 1: Timeframe of the eviction process in the state of Washington (Tenants Union of Washington State, 2024).

Ahn (2023) produced two eviction flowcharts relevant to California and New York, shown in the Appendix (Figure A1 and Figure A2). Again, the specific timeframes may vary based on state laws. For instance, some states may allow tenants 10 days to respond to the summons, whereas others offer only 7 days. These timelines give an idea of the potential speed of eviction under the current system.

Typically, the eviction process can be divided into two phases: before and after filing. Before filing for eviction, landlords issue notices to ask the tenant to pay or vacate. After filing, the legal process begins. Removal or tenant departure can occur at any stage during this process, from the first notice to the sheriff's eviction. However, what happens to tenants after the eviction filing remains understudied. Specifically, three key stages occur after eviction filings: the period after receiving a summons, judgment from the court, and eviction by the sheriff. Determining how the number of eviction cases changes at each stage remains challenging.

Summary

This chapter provided a brief introduction to this study, offering a broad background, defining eviction as it pertains to eviction research (specifically, eviction filings and sheriff evictions), and outlining the eviction process. The eviction process varies by state, but the overall structure is similar, typically resulting in one of three main outcomes: dismissal, judgment, or eviction by the sheriff. For instance, a case might be dismissed if tenants move out or pay in full or neither party attends the hearing before judgment. This study primarily investigated factors related to these three outcomes, making it essential to understand the meanings of these terms.

Chapter 2: Eviction Dynamics: Scope and Contributing Factors

This chapter first outlines the scope of the eviction problem, covering its magnitude, severity, effects, and trends. This analysis clarifies why eviction constitutes a significant social concern. Because eviction remains a critical societal challenge, scholars are dedicated to investigating the underlying mechanisms through empirical research and evidence, aiming to mitigate eviction risks, prevent eviction, and bolster housing stability. Understanding this part of the literature is beneficial for developing models that treat the eviction rate or outcome as a dependent variable. Thus, the factors associated with or potentially causing eviction in the literature are reviewed.

Scope of the Eviction Problem

Magnitude and Severity

Eviction is a significant and urgent issue affecting millions of households today. According to the American Community Survey (ACS) 2019 5-year estimation report in the United States, 36% of households are occupied by renters (U.S. Census Bureau, 2021a). In major metropolises like San Francisco County, California, renters outnumber homeowners, comprising 60% of the population (U.S. Census Bureau, 2021b). Regarding gross rent as a percentage of household income, the median range is between 25.0% to 29.9%, with 39.4% of renters spending 35.0% or more of their income on rent. These individuals fall into the category of rent-burdened families, because they spend more than 30% of their household income on housing (U.S. Census Bureau, 2021a). Rent increased approximately 15.3% in early 2022 but began to stabilize in the United States in 2023 (Joint Center for Housing Studies of Harvard University, 2024). Nationally, although rent growth is slowing, the median rent was $1,967 in November 2023, compared to $1,637 in February 2019 (Leckie, 2023). The median rent on the West Coast increased from around $2,000 in 2019 to $2,500 in 2023, representing a 25% increase (Leckie, 2023). As companies encourage or require their workers to return to the office, some urban areas may become competitive again, potentially fueling further rent increases. Desmond (2020) highlighted how rent burdens exacerbate difficulties for low-income families, leading to less food because “rent eats first.” In addition to those struggling with rent, it is important to recognize the substantial number of unhoused individuals. According to the U.S. Department of Housing and Urban Development (HUD), 582,000 people experienced homelessness in the United States in 2022, equating to approximately 18 per 10,000 people (De Sousa et al., 2022).

Due to the large number of tenants in the United States, even a low eviction rate results in a large absolute number of tenants suffering from eviction. According to national estimates from the Eviction Lab at Princeton University, in a typical year, particularly before the COVID-19 pandemic, landlords filed 3.6 million eviction cases from 2000 to 2018, resulting in an approximate 8% eviction filing rate (number of filings per renting households). Because some families may face several evictions a year, the estimated number of annual evictions represents around 2.7 million unique households. This indicates that nearly 6 of 100 renting households were threatened with eviction at least once. Leung et al. (2021) also revealed an unexpected pattern whereby close to one third of the households facing eviction had cases filed against them repeatedly at the same address across multiple months in 2014. Additionally, a large portion of the population is on the edge of eviction. A recent report noted 52% of workers indicated they do not have a 3-month savings cushion (Metlife, 2023). This situation is further supported by data from the Federal Reserve Bank and the ACS, showing the median savings account balance for Americans in 2019 was around $5,300, whereas the median gross rent (2017–2021) was $1,163, barely enough to sustain rent for 3 months after accounting for other expenses. Furthermore, approximately 68 million individuals, accounting for about one third of the U.S. population, had debt in collections in 2019, according to a report from the Urban Institute (Warren et al., 2020).

After the pandemic began, the eviction filing rate changed significantly due to various tenant protection policies, such as eviction moratoria. For example, from March to December 2020, eviction filings were 65% lower compared to the historical data for the same range. Data from 2020 and 2021 showed a similar pattern, but when tenant protection policies were eliminated, the number of eviction filings surged (Hepburn et al., 2022). A preliminary analysis found landlords filed 78.6% more cases compared to 2021 (Vallejo et al., 2023).

Eviction filings also vary geographically. First, the eviction filing rate varies among states due to different eviction policies and laws. For example, eviction filing rates are usually higher in the Southeastern United States, like South Carolina and Georgia, which are close to 20%, whereas some states with high-population metropolitan areas do not always have high rates. Because the eviction filing rate is not an absolute number, smaller cities sometimes have a higher eviction filing rate due to the smaller number of rental properties (Eviction Lab, 2018). Although New York City has the most eviction filings, the rate is relatively low due to its large number of rental properties. Within states, urban areas usually have higher rates than rural areas. Recently, there has been a trend of increasing eviction rates in suburban areas, possibly related to urban rental markets and segregated suburbs (Hepburn et al., 2022).

At the community level, census tracts with a higher concentration of people of color, especially Black residents, have higher eviction rates than non-Hispanic Whites (Nelson et al., 2021; Thomas, 2017). Medina et al. (2020) estimated that people living in minority-dominated communities are 66% more likely to experience eviction compared to non-Hispanic Whites-dominated communities. Similar evidence emerged in King County, Washington, which includes Seattle (Thomas, 2017). Partially due to the history of segregation and discrimination, people of color usually have lower household incomes and live in the same communities. Beyond race, immigration status also is linked to disparities, with immigrants being more vulnerable to eviction compared to citizens. Tesfai and Ruther (2023) argued that immigrants are more vulnerable based on a natural experiment that compared eviction rates to community characteristics related to immigration, such as the percentage of foreign-born residents. The study showed that immigration status was significantly correlated with higher eviction filing rates and mediated by settlement patterns.

Racial and ethnic minorities and women are more vulnerable compared to other groups of tenants. Thomas (2017) analyzed individual cases in King County, Washington, and found that people of color, particularly Black women, had a higher risk of eviction. Similar results emerged in national analysis: Although Whites comprised about half of adult renters (51.5%), they experienced only 42.7% of eviction filings (Hepburn et al., 2021). Hepburn et al. (2021) estimated that 341,756 women were evicted annually, approximately 16% more than men, based on their sample of 1,195 counties. In a recent study connecting eviction filing data to census data, Graetz et al. (2023) showed that 28.3% of Black women with children experienced eviction filings, compared to 16.3% without children.

The rent burden also varies among different races and ethnicities. Wedeen (2021) analyzed data from the U.S. census and concluded that more than half of Hispanic and Black renter households are rent burdened (spending more than 30% of their income on housing), compared to around 40% of White households. In addition, 26.1% of Hispanic and 29.3% of Black households are severely burdened (spending more than 50% of their income on housing), relative to 21% of White households in this category. Leung et al. (2021) found that serial eviction filing rates are higher in Black and Hispanic communities. Serial eviction means the landlord repeatedly filed against the same address, mostly to regain late fees.

Effects of Evictions

Because eviction is so prevalent, researchers have begun to explore the effects of being evicted. Knowable Magazine (2020) quoted Emily Benfer, a housing expert at Wake Forest University, as stating: "Nothing good comes out of eviction." Eviction negatively affects many aspects of life, including housing, poverty, physical well-being, and mental health. Additionally, eviction as a household event also affects child welfare and development, such as food security and academic performance. This section discusses the short-term and long-term consequences of eviction in terms of housing, poverty, health, and other aspects.

Housing. The most direct impact of eviction is housing challenges. If tenants cannot find a place to live in time, they may face homelessness. This argument, although seemingly straightforward, lacks solid evidence to establish causality, such as the percentage of eviction cases that result in experiences of homelessness. Treglia et al. (2023) argued that there is currently no causal evidence that eviction is a direct cause of homelessness. Burt (2001) used data from 1996 and concluded that 15% of people experiencing homelessness in the survey were homeless due to " could not pay for rent." Both pieces of research may support the notion that eviction leads to homelessness, but there might be several steps between these two events.

However, empirical studies have discussed the causality between eviction and shelter service use. Collinson and Reed (2019) linked eviction data with administrative data covering all shelter use in New York and estimated that eviction increases the probability of applying to a homeless shelter by 14%. Collinson et al. (2024), in their analysis of New York City and Cook County, Illinois, observed that although evictions cause no substantial increase in homelessness, the use of emergency shelters or other housing-related temporary services increased. A 2013 New York City homelessness report found that about 25% of families who arrived at emergency shelters had been evicted from their apartments (Hartman & Robinson, 2003). Data from Columbus, Ohio, showed that more than 33% of families in emergency shelters reported eviction as the primary factor (Hartman & Robinson, 2003). These pieces of evidence show that eviction begins to shake the foundation of stable housing.

Even if the situation does not become as bad as homelessness, eviction also limits tenants' ability to find a new place to rent due to a lower budget and lower credit score. Regarding budget constraints, tenants are required to pay extra fees in addition to rent, such as attorney and court fees if they experience eviction fillings. According to a report from Seattle, these extra fees can double the cost of rent (Cookson et al., 2018). The median rent owed in eviction cases is $1,500 (59.9%), and the rest includes court costs ($358.98, 14.3%), attorney fees ($416.19, 16.6%), and nonrent charges ($228.03, 9.1%; Cookson et al., 2018). Thus, if tenants lose their case, they will have less cash for the next security deposit or first month's payment. Another study compared earnings between people who were evicted and those who were not and found that eviction decreased their average quarterly earnings by $323 (7% of the nonevicted mean of $4,300; Collinson et al., 2024). Moreover, the total balance of collections peaks at around $3,000 at the time of eviction filings.

Moreover, when eviction records appear on tenants' credit reports, future landlords can view these records. Humphries et al. (2019) conducted a quasi-experimental study and analyzed credit score changes during the eviction filing process, finding that credit scores dropped by about 15 points in the prior eight quarters at the time of eviction filing. It took an additional 12 quarters (about 3 years) to recover to a normal score of around 545. Thus, their eviction record constrains their choices due to this history. This record can also exclude them from applying for welfare benefits. Greiner et al. (2013) noted that an eviction record could jeopardize an individual's or family's ability to obtain public housing. For example, public housing agencies cannot admit tenants if they were evicted from federally assisted housing (Code of Federal Regulations, 1999). However, a study by Gromis et al. (2022) suggested that public housing is a significant player in evicting tenants. Besides losing the benefits of public housing, eviction records also negatively influence tenants' ability to rent new houses. Several online tenant screening tools, like First Advantage, include eviction history. Although eviction records do not appear on credit reports, tenants' credit scores can decrease dramatically if the landlord starts debt collection, decreasing their ability to rent other housing in the future.

Health. Housing has been regarded as a social determinant of health. According to a systematic review, Vásquez-Vera et al. (2017) found that individuals affected by the threat of eviction have worse outcomes in both mental health, such as depression, anxiety, psychological distress, and suicide, and physical health, including chronic diseases, high blood pressure, domestic violence, and child maltreatment. However, the term *eviction* used in this literature review included foreclosure and rent-related eviction. When narrowed to rental-related eviction, which is the definition of eviction in this study, similar results were observed. A longitudinal study with 120 tenants appearing in eviction court in New Haven, Connecticut, assessed mental health at baseline and 1, 3, 6, and 9 months. The study reported that 39% of the tenants screened positive for generalized anxiety disorder, but less than 25% received mental health assistance (Tsai et al., 2021). Beyond the process of eviction, the perceived risk of eviction is also associated with elevated mental health problems. Acharya et al. (2022) analyzed the U.S. Census Bureau's Household Pulse Survey from July 2021 to March 2022 and found a higher prevalence of depression (59.33% vs. 37.01%) and anxiety (67.01% vs. 43.28%) among residents who believed they were at risk of eviction compared to those not at the risk of eviction. Displacement, a potential result of eviction, was associated with severe maternal morbidity risk, according to an investigation of 72,718 cases in California between 2006 and 2017 (Gao et al., 2023). Graetz et al. (2024) linked eviction and mortality records and found eviction filings were positively associated with mortality.

Eviction also influences the health outcomes of minors through their mothers and household. Children in these households also experience worse health status compared to nonevicted households. A systematic review of 11 studies related to eviction experience suggested that childhood exposure to eviction is associated with overall poorer child health (Ramphal et al., 2023). These adverse outcomes include the risk of prematurity, low birth weight, or even infant mortality (Hazekamp et al., 2020; Himmelstein & Desmond, 2021; Khadka et al., 2020). Hazekamp et al. (2020) used eviction data from Chicago in 2016 to assess the relationship between eviction and pediatric health outcomes in 653 census tracts. In terms of cognitive development, eviction experience in middle childhood was associated with lower cognitive scores at age 9 (G. L. Schwartz et al., 2022). Bullinger and Fong (2021) used block-level administrative data in Connecticut and found that reports of maltreatment were positively associated with the rate of eviction notices in the neighborhood. Another national study (N = 9,029; 1994–2008) provided further evidence that as children grow up, adverse outcomes of eviction extend into adolescence or early adult life, with statistically more depressive symptoms in the group with eviction experience (Hoke & Boen, 2021).

Poverty. Eviction is closely associated with poverty. Because this concept's definition and measurement can vary by location, in this study, poverty is defined and measured as household income. Poverty is a broad term and is affected by factors such as residential mobility, making it directly and indirectly related to eviction in many aspects. Desmond (2012) observed an association between eviction and the perpetuation of urban poverty. A study in Milwaukee County using both quantitative and qualitative analyses concluded that in poor Black neighborhoods, eviction is a typical factor in the reproduction of urban poverty. Collinson and Reed (2018) suggested that eviction might not be the main driver of poverty; rather, living in poverty makes tenants more likely to experience eviction, based on an analysis of eviction data in New York City. The legal process of eviction also increases the financial burden on tenants after adding various non-rent-related court fees. Desmond (2016) and Desmond and Kimbro (2015) highlighted that eviction often forces people to move to poorer neighborhoods, increasing their hardship.

Eviction represents a household change and is often an indicator that children, if present, are experiencing poverty. Lundberg and Donnelly (2018) used data from the Fragile Families and Child Wellbeing Study, a longitudinal panel study ($N = 4{,}898$) of children born in 20 large U.S. cities. Analysis of one wave of data (1998–2000) concluded that 1 in 7 children experiences at least one eviction before age 15, and 1 in 4 experiences eviction in deeply impoverished households. However, overall investigating the impact of eviction on children is challenging, because their names are not specified in eviction records. Given the mutual relationship between poverty and eviction, experiencing eviction in childhood could place children at high risk of reproducing poverty in their future personal and household life.

Beyond the main effects on housing, health, and poverty, which have been widely discussed, eviction has effects on other domains such as on social mobility and election participation. For instance, eviction filings often result in extra costs, like attorney fees, increasing the financial burden for residents and reducing the possibility of upward mobility (Cookson et al., 2018). Another study found that residential eviction rates negatively influenced voter turnout during the 2016 presidential election, potentially affecting democratic participation (Slee & Desmond, 2023).

This section highlighted the various adverse effects of eviction. Although categorized into three main areas, these effects are part of a complex system, interacting with one another and sharing confounding factors. For example, eviction is associated with both poverty and health, which may also influence each other, with either being a cause for the other. Furthermore, poverty and health can also lead to eviction, indicating a direct or indirect relationship or cause and effect.

Trends in Eviction

This section explores trends related to rent prices and policy interventions in recent years, as discussed in the literature.

Rent Price Change. According to analysis from the Federal Reserve Bank (2023), rents have been increasing faster than overall prices and have kept pace with general inflation since data collection began in 1984. Location also affects the rate of increase; for example, the East Coast and West Coast have seen steeper rate increases than the Midwest. When the COVID-19 pandemic started, the rate flattened for around 1.5 years in major cities like New York and San Francisco. Rent prices began to bounce back in 2021, but the rate of increase varied geographically. For instance, the rent price index in Hillsborough County, Florida, rose by more than 40% compared to prepandemic levels in 2022, whereas rents in San Francisco had yet to return to prepandemic levels (Martin, 2022). As this trend continues, tenants will face more pressure from rent expenses. Additionally, the return to the office by big companies will push rent higher in major metropolises, creating a seller's market by making it easier for landlords to find tenants (White, 2023). A report by the Joint Center for Housing Studies of Harvard University (2022) showed that the vacancy rate in prime urban markets is at historical lows.

In the renter group, rent increases will disproportionately influence people with different income levels. For households earning below $30,000, around 80% are burdened by rent, compared to around 10% of households with incomes exceeding $75,000 (Joint Center for Housing Studies of Harvard University, 2022). Higher-income households might decide to rent instead of buying a home due to decreased affordability of homeownership, which may also drive up the average rent. Data from the ACS in 2019 indicate that the number of renters making at least $75,000 jumped by 48% compared to 2009, reaching 11.3 million, adjusted for inflation (U.S. Census Bureau, 2023). Meanwhile, as interest rates and housing prices climb rapidly compared to prepandemic levels, potential buyers are finding it more difficult to afford a house or save for a down payment. This situation will likely keep more higher-income households in the rental market in the coming years.

After the homeownership rate peaked in 2005 at 69%, it declined to a historical low of 63% in 2016 due to the 2008 financial crisis. According to Federal Reserve Economic Data (2022), the homeownership rate has stabilized around 65%, except for a jump to 68% in 2020 due to the COVID-19 pandemic, but it quickly fell back to 65%. Thus, the percentage of renters could be stable in the near future. Internationally, renters also represent a significant percentage of the population. For example, the English Housing Survey reported that 36% of households are renters; in the European Union, 30% live in rented housing. Similarly, in China, 248 million residents are renters (17.6%), an increase of 20% compared to 2017.(CBNData, 2018; Department for Levelling Up Housing & Communities, 2023; Eurostat, 2021)

Policy Change. As the pandemic began, tenants' rights garnered significant public attention, leading to policy changes. The most impactful was the federal eviction moratorium, issued by the Centers for Disease Control and Prevention under the CARES Act and the Emergency Rental Assistance program from the Consolidated Appropriations and American Rescue Plan Act. These policies and programs aimed to prevent immediate displacement or assist in paying back rent. According to estimates from Eviction Lab, the eviction moratorium prevented around 1.55 million eviction filings between September 4, 2020, and July 31, 2021 (Rangel et al., 2021). However, this does not imply that these households will never face eviction; it might only have delayed evictions to the date the moratoriums ended, which was September 30, 2021, in most states. Despite the termination of some policies, attention to tenants' rights has somewhat persisted postpandemic, leading to further policy changes to enhance tenant protection (White House, 2023).

At the federal level, the Biden–Harris Administration published a memo for tenant protection in July 2023, urging federal departments like HUD and the U.S. Department of Agriculture to guide fair tenant-screening practices. Changes in eviction timelines in some states have also started to influence federal policy, with HUD encouraged to issue a notice about allowing more time for public housing tenants to avoid eviction. Leung et al. (2023) noted that tenants in public housing do not have a lower risk of eviction compared to those in the private market and face a far higher risk of serial eviction filing. The new proposal aims to give residents more time to respond, especially for late payments, because in some states, landlords can file for eviction only 5 days after rent is due.

State and local governments have begun to enact policies to protect tenant rights. For general cost burden, Connecticut Senate Bill 998 limited the amount of late fees and California AB 12 set a limit on direct deposit fees, easing the cash flow for tenants to some extent. For legal aid, Colorado House Bill 23 prohibited rental agreement clauses that limit renters from seeking legal recourse, whereas St. Louis Board Bill No. 59 provided access to legal representation for tenants facing eviction. New York City became the first jurisdiction to implement a universal access to counsel policy, ensuring that low-income tenants can receive free legal aid when facing an eviction filing. This policy has led to a modest decrease in both the proportion and absolute number of executed warrants.

Other policies indirectly affecting eviction may also enhance tenant protection. For instance, a report from the Rand Institute argued that crime-free rental policies did not improve local safety but increased eviction rates (Griswold et al., 2023; Griswold et al., 2024). Such policies were popular nationally and have started to be abandoned in cities like Riverside and San Bernardino, California, due to their adverse impact on eviction. California AB 1418 also limited the impact of local crime-free policies and extended tenant protection.

Thus, the trend for the next few years is twofold: On one hand, the rental market is heating up as fewer people can afford to purchase housing and compete for rental properties, especially in locations with job opportunities. On the other hand, policies aimed at protecting tenants' rights and establishing programs to decrease eviction rates are being implemented. The former may harm tenants, whereas the latter benefits them.

Factors and Causes

Various factors influence or cause eviction. This section reviews the literature to provide an understanding of both contextual and individual factors influencing evictions. This includes societal and economic conditions alongside personal circumstances like income level and job security.

Contextual Factors

Policy Factors. As mentioned in the previous chapter, policy and macroeconomy are two influential factors associated with the eviction rate. For example, the federal eviction moratorium issued by the Centers for Disease Control and Prevention decreased the eviction rate temporarily during the emergency situation of the COVID-19 pandemic.

Usually, eviction-related laws are issued at the state level. Overall, more tenant-friendly policy environments are associated with lower eviction rates (Merritt & Farnworth, 2021). Bradford and Bradford (2023) tested the impact of state housing policies on eviction filings across a large proportion of U.S. counties from 2001 to 2018 and found that some policies may be more effective in reducing eviction filings than others but found less evidence that they lead to lower eviction judgments. Specifically, policies prohibiting landlord retaliation and the availability of low-income housing tax credits were statistically significantly linked to a decrease in the eviction filing rate, whereas housing and urban development permissions and policies were associated with increased eviction filing rates. Additionally, each state may allow different periods for tenants to respond to a summons. Intuitively, if the state allows more time for tenants to respond, they may be more likely to prepare for the case and acquire legal resources. Some states now require notice, such as issuing a pay or vacate notice, before filing for eviction (e.g., Washington). According to Eviction Lab's estimation, the longer the required period for notice, the lower the eviction rate. However, lower eviction rates do not mean fewer people are displaced from their rental homes. This only indicates fewer formal evictions because tenants have more time to move before eviction filings or judgments.

On the housing supply side, some state laws may alter the housing inventory, such as SB 9 in California, which allowed for the conversion of single-family units to multifamily units (California State Senate, 2021). Another approach is to increase the cost of filing to decrease the eviction rate. Gomory et al. (2023) observed that each $76 increase in eviction filing fees was associated with a 1.71% decrease in the eviction filing rate. The researchers explained that some tenants might owe a small amount of money, and the higher filing cost could make landlords reconsider the purpose of filing—e.g., prevent serial filing by landlords to collect on late payments, fees, and fines without intending to remove the tenant. For tenants, if other costs in their daily life decrease, their housing will become more affordable. Gerr (2019) conducted a natural experiment by comparing counties along the borders of different states that did or did not expand Medicaid and found that Medicaid expansion was associated with lower eviction rates.

A wide range of policies can affect the landlord–tenant relationship. Most were mentioned in the section on policy trends, such as limitations on security deposits or collecting late fees. Although these policies may or may not influence the eviction rate directly or indirectly, empirical research supporting these practices is rare. Thus, it might be too early to see the difference due to the policy change, or partially due to variations in the definition of eviction rate or how eviction data are collected from state to state, which makes it difficult to compare states accurately.

Neighborhood Factors. Because this study concentrated on one county in the state of Washington, there is no significant difference in policy implementation. Thus, this section narrows the perspective to the neighborhood level, discussing findings related to gentrification and neighborhood socioeconomic status (SES) in the literature.

Gentrification, the process of transforming working-class areas for middle-class use, has been identified as one factor leading to eviction and displacement (Atkinson, 2004; Chum, 2015). As gentrification begins in a neighborhood, especially with housing demolition and the introduction of newer properties, the direct impact is often an increase in housing costs, including rent, property taxes, and maintenance, which are eventually passed onto the tenant. For example, in a neighborhood in Atlanta, Georgia, investor purchases of multifamily rental housing led sale prices to increase by an average of $5.5 million for the same parcel from 2005 to 2018 (Raymond et al., 2021). Scholars have linked gentrification and eviction in various research studies and contexts. Chum (2015) analyzed 59,415 eviction applications in Toronto from 1999 to 2001 and concluded that the eviction rate was positively associated with neighborhoods in the early stages of gentrification, indicating that the number of eviction filings increased as a neighborhood began the gentrification process. Sims (2021) found that large multifamily housing developments, a typical phenomenon in gentrification, increased the number of available units but produced increased eviction filings in a small radius (tenth of a mile) in Madison, Wisconsin. Evidence of gentrification-induced eviction also appeared in Detroit, Michigan (Mah, 2021).

Hepburn et al. (2023) examined gentrification's role in the eviction crisis and found that eviction rates fell more in places that were gentrifying but still less significant than in low-SES neighborhoods, based on around 6 million eviction filings across 72 U.S. metropolitan areas between 2000 and 2016.

However, even though the concepts of gentrification and SES have been used in prior research, the measurement of these two concepts remains vague. One is based on demographic information for residents, and the other focuses on economic activities associated with properties. Chum (2015) used several indicators, like average personal income and the percentage of employed residents, and applied principal component analysis to create a gentrification score. Mah (2021) used the number of permits to reflect gentrification status. Han (2023) used the ratio of building permits for new construction to reflect the level of investment. In terms of SES, Hepburn et al. (2023) also used principal component analysis to convert four variables into a single measure of SES. These four variables were neighborhood education level (percentage of residents with a high school degree or higher), employment status (percentage) in technical or professional occupations, median home value, and median rent.

Beyond gentrification and SES, scholars have sought to identify other concrete and measurable factors related to eviction. Goodspeed et al. (2021) reviewed eviction case filings in the literature and identified several levels of neighborhood characteristics: geographic scope, demographics, economics, housing, and the legal system. In terms of geographic scope, most studies have focused on the city or county level or metropolitan region. Recently, more research has expanded this factor to include urban versus rural or urban versus suburban areas (Goodspeed et al., 2021; Hepburn et al., 2022). At the demographic level, almost all eviction studies related to neighborhood characteristics have considered racial and ethnic composition (Goodspeed et al., 2021). Some studies also included educational attainment, single-mother households, percentage of children, and crime rate. Regarding economics, common measurements are poverty rate, median household income, and unemployment rate. Regarding housing status at the community level, some measurements are vacancy rate, percentage of burdensome housing cost, rate of foreclosure, percentage of subsidized housing units, and percentage of mobile homes and parks. To show the relationship between these variables and eviction rate or eviction filing rate, Table 1 displays their coefficients. The sign (positive or negative) of the coefficient only if its *p*-value is less than .05.

Table 1 was inspired by Goodspeed et al (2021)'s study and generated from data across 16 studies from 2012 to 2023 focusing on neighborhood characteristics and various regression models. These characteristics are divided into three broad categories: demographics, economics, and housing factors. Most variables were sourced from the ACS, an annual demographic survey by the U.S. Census Bureau. Some local variables were obtained from local government open data platforms. As research progresses, an increasing number of variables have been considered and found to be statistically significant with eviction. Generally, the observed relationships have been consistent across studies.

In the demographic category, a higher percentage of Black residents has been positively associated with eviction filing rate or eviction rate. Neighborhoods with a higher percentage of single mothers and more children in the household also have been linked to higher eviction rates. However, a higher education level—specifically, the percentage of residents with a high school diploma or college degree—was statistically significantly related to a decreased eviction rate.

Characteristics Level	Variables	Desmond, 2012, Milwaukee, WI	Desmond et al., 2013, Milwaukee, WI	Greenberg et al., 2016, Milwaukee, WI	Desmond & Gershenson, 2017, Milwaukee, WI	Johns-Wolfe, 2018, Cincinnati, OH
Demographics	% black population	+	+			+
	% Single mothers		+			+
	% High or College Education					
	% Hispanic population	+				+
	% Asian population					
	# of Kids in household		+	+	+	
Economic	Unemployment rate			+	+	
	Median income					
	%Poverty rate		+			+
	Job opportunities					
	Rent burdened households					
	Average Household size					
	Level of investment (%)					
Housing	Renting households (%)					
	Household density					
	Median rent					
	Homeownership rate (%)					
	Number of leaseholders		+			
	Vacant for sale					
	Mobile homes as % of HH					
	# of subsidized housing units					

Characteristics Level	Variables	Raymond et al., 2018, Fulton Co., GA	Clark et al., 2018, Charlotte, NC	Immergluck et al., 2020, Atlanta, GA	Tan 2020, San Francisco, CA	Robinson and Steil, 2021, Boston, MA
Demographics	% black population	+	+	+	+	+
	% Single mothers	+				
	% High or College Education			-	-	-
	% Hispanic population		+			
	% Asian population			-		
	# of Kids in household	+	+		+	
Economic	Unemployment rate					
	Median income					-
	%Poverty rate					
	Job opportunities				+	
	Rent burdened households					+
	Average Household size					
	Level of investment (%)					
Housing	Renting households (%)					
	Household density					
	Median rent					
	Homeownership rate (%)					-
	Number of leaseholders					
	Vacant for sale					
	Mobile homes as % of HH					
	# of subsidized housing units					

Characteristics Level	Variables	Merritt and Farnworth, 2020, US	Goodspeed et al., 2021, MI	Kim et al., 2021, Salt Lake Co., UT	Preston and Reina 2021, Phila., PA	Eviction Lab, 2022, US	Han, 2023 Kansas City, MO
Demographics	% black population	+			+	+	+
	% Single mothers		+				
	% High or College Education			-			-
	% Hispanic population			+	+		
	% Asian population						
	# of Kids in household				+		
Economic	Unemployment rate	+		+		+	+
	Median income	-				-	
	%Poverty rate	+		-	+		
	Job opportunities		+				
	Rent burdened households	+			+		
	Average Household size	+					
	Level of investment (%)						+
Housing	Renting households (%)					-	
	Household density					+	
	Median rent					+	
	Homeownership rate (%)		-				+
	Number of leaseholders						
	Vacant for sale						+
	Mobile homes as % of HH		+				+
	# of subsidized housing units			+			-

Table 1: Relationship of Neighborhood Factors to Eviction Filing or Judgment Rate

In the economic category, neighborhoods with more households with stable and higher incomes have been linked with lower eviction rates. The unemployment rate and median household income have had similar effects associated with eviction. Lower unemployment rates and higher median household incomes in a neighborhood are usually associated with lower eviction rates. The poverty rate is another indicator, with most studies concluding that it is positively related to eviction rates, although Kim et al. (2021) reported a negative relationship between the poverty rate and eviction rate. Kim et al. (2021) found a high-density area of eviction in the central business district of Salt Lake City, Utah, where the poverty rate was diluted by higher-income groups in other parts of the census tract. Spending on housing for rental properties is reflected by an indicator called rental burden. The percentage of households with rental burden in a neighborhood has been positively associated with the eviction rate, with more households at risk of missing or making late payments.

In the housing category, certain factors reflect the housing situation of a neighborhood, such as the percentage of rental versus owned properties, median rent, and percentage of mobile units. More number of rental properties are usually associated with a lower eviction rate, and median rent is typically positively associated with the eviction rate. Higher homeownership rates have been related to lower eviction rates, likely because fewer people are renting in the neighborhood. Findings regarding the percentage of renting households and homeownership may seem in conflict, because both have been negatively related to the eviction rate. Han (2023) observed that a higher homeownership rate increased the eviction rate, suggesting this indicator may be affected by specific geographic areas. Specifically, because the denominator is the total number of rental properties, the more rental properties, the smaller the eviction rate might be. In another extreme situation, if a neighborhood with a very high homeownership rate has only two rental properties and one eviction, this will result in a 50% eviction rate. This might explain why both relationships can be observed in the percentage of rental versus owned properties in neighborhoods regarding the eviction rate. Moreover, the percentage of mobile homes is another strong indicator positively related to the eviction rate.

When exploring neighborhood factors associated with the eviction rate, most studies have selected similar variables from the ACS, HUD, or local data to establish a connection with the eviction rate, ultimately demonstrating consistent relationships and even replicating results of significance. For example, the percentage of Black residents has had a strong positive relationship with the eviction rate. When narrowing even further from the neighborhood level, the next unit of analysis is the individual property.

Property and Individual Factors

Property Factors. When analyzing neighborhood-level variables, housing characteristics are already reflected to some extent. Yet in any given neighborhood, special properties might exist that don’t necessarily align with overall neighborhood housing characteristics. A large share of eviction filings is submitted by a relatively small share of landlords. For example, Teresa and Howell (2021) reported that the top 10 property owners were responsible for 39% of all evictions ($N = 9{,}269$) in Richmond, Virginia, from 2015 to 2018. Rutan and Desmond (2021) found that evictions in several buildings were responsible for large portions of the overall eviction rate in Cleveland, Ohio; Fayetteville, North Carolina; and Tucson, Arizona. The property or address is a strong variable to reveal information related to eviction. By comparing the text or geocoding of addresses, scholars can find which properties are often related to eviction and explore the property characteristics. Scholars have found the serial eviction filing rate is higher in Black and Hispanic communities by exploring which addresses have repeated eviction filings (Leung et al, 2020). Property characteristics create another perspective to view the eviction crisis and identify specific property uses, landlords, and other information.

Few studies have explored factors at this level, given that case-level data are not easy to access and external datasets are required to link them to geographic information. In other words, other datasets can add extra value to address information and explicate property characteristics. Some related factors include sales history, property type, and property funding sources.

As previously mentioned, gentrification influences neighborhood development, which is reflected by permit applications, construction, and renovation. Robinson and Steil (2021) linked eviction filings in Boston Housing Courts from 2014 to 2017 with property history and found that eviction filings were more likely in recently constructed or renovated cooperative-owned properties compared to other properties in the same neighborhood. Ramiller (2021) combined eviction with property-level turnover data in Seattle, Washington, and revealed that evictions were more likely to occur at properties sold in the same year or with remodeling and demolition permit applications filed.

Property type and property funding sources are directly related to property characteristics. These two factors are often considered together due to their high correlations. For example, subsidized housing units are usually part of multifamily properties and owned by large corporate entities or local housing authorities, whereas single-family housing is more commonly owned by private landlords. In certain situations, a firm may primarily invest in single-family housing. Overall, large corporate owners are more likely to evict tenants than individual landlords (Decker, 2023; Raymond et al., 2016; Seymour, 2022). Regarding subsidized housing, which is designed to increase housing affordability, Harrison et al. (2021) showed that subsidized multifamily properties have statistically significantly lower eviction rates (10.7% lower) than market-rate properties in the order adults group but not statistically significant lower (1.4% lower) in the non-senior group in Atlanta, Georgia. Preston and Reina (2021) delved into different subsidized housing programs, concluding that public housing and project-based rental assistance properties were linked with a decreased eviction rate, with no decrease observed for low-income housing tax credit properties. Although there is evidence that public housing decreases the eviction rate, a disproportionate relationship persists between eviction rate and percentage of public housing units. Gromis et al. (2022) combined eviction filing data from 2006 to 2016 with federal registers of public housing authorities and estimated that public housing accounted for 5.8 of every 100 eviction filings, despite only representing 3.5 of every 100 rentals in public housing. A similar result was observed in Richmond, Virginia, where public housing constituted 7% of the rental housing but 14% of evictions (Teresa & Howell, 2020).

Extracting and linking the addresses of properties with evictions to other datasets provides a new viewpoint to examine the characteristics of these properties, such as permit application or ownership, revealing other factors associated with the eviction rate.

Individual Factors. After analyzing factors at macro levels and neighborhood characteristics, researchers have delved into individual factors not only to explore correlations but also to identify potential causes, because direct connections to eviction may be easier to detect at the individual level compared to neighborhood or higher levels.

Typically, evictions occur due to late payment or nonpayment of rent. Payment issues are estimated to account for 93% of eviction filings in Washington, DC; 86.5% in Seattle, Washington; and 90% in Cleveland, Ohio (Cookson et al., 2018; McCabe & Rosen, 2020; Urban et al., 2019). In Boston, approximately three fourths of eviction filings are due to nonpayment (City of Boston, 2019). Other factors such as violation of lease terms, illegal activity, and lease termination lead to eviction, but their overall proportion is small. Given that nonpayment is the major cause, researchers have explored what leads to nonpayment. However, data quantifying specific causes of nonpayment are lacking. Urban et al. (2019) conducted interviews with 87 residents with eviction experiences in Cleveland, Ohio, and identified three main causes: attempted payments (e.g., payment plan), housing condition issues (e.g., water leakage), and financial instability. The first two are usually due to disagreements between tenants and landlords, and financial instability often involves factors like jobs, medical bills, or unexpected expenses. Chen et al. (2024) also tried to identify the conflicts between tenants and landlords using the GPT-4 model and concluded that the concerns vary across states. Some of the top conflicts are utility disputes, deposit disputes, and evictions by landlords.

Race and ethnicity information is not usually collected by the court or related departments, especially in large administrative datasets, which prevents research connecting race and ethnicity to eviction cases. Thomas (2017) used an R package called wru and Bayes's rule to estimate the race and ethnicity of people based on their last name and location, which can reflect the census tract's racial composition. Similar methods have been applied to estimate gender(Thomas et al., 2024). According to his analysis, Black women were the most vulnerable group, with Black adults being 5.5 times more likely to be evicted than Whites in King County, Washington, from 2012 to 2017. Another national study of filing patterns in 2023 linking U.S. Census Bureau individual data revealed that Black tenants experienced a disproportionate share of eviction filings and Black women with children were the most vulnerable group (Graetz et al., 2023). Desmond and Gershenson (2017) selected a block in Milwaukee, Wisconsin, and collected individual characteristics such as race, job status, family structure, income, educational level, and payment history. They found that the number of children, payment history, and time since the last eviction were significant variables, with the first two positively correlated with the eviction rate. However, due to the homogeneity of a block, some statistically significant variables may be nonsignificant in a larger area.

Legal representation is another factor if the landlord has filed for eviction. A significant gap in legal representation exists between landlords and tenants. According to data from Albany, New York, only 2.47% of tenants were represented in court, compared to 88.6% of landlords. Several pilot tests about legal representation for tenants have occurred in New York City and Boston. Seron et al. (2001) observed New York's Housing Court and found that legal counseling and representation resulted in fewer eviction writs. However, Greiner et al. (2013) found no statistically significant evidence of a relationship between the extent of representation and eviction court case outcomes, because most legal aid only involves advice, given the high demand for service.

This review explored various factors associated with the eviction rate. First, most factors previously tested have been at the neighborhood or census tract level, seldom narrowing to the individual case level because individual data are difficult to collect comprehensively. Second, the dependent variable in most research has been the eviction filing rate, eviction rate, or sometimes the number of evictions. Research exploring the process or outcome after an eviction filing, except for studies on the legal representation of defendants, is rare. Third, most research could only establish associations, facing challenges building causal inference because the data were collected by other entities, making it difficult to design experimental or quasi-experimental research. Thus, it often remains challenging to determine causality from observational studies due to the presence of confounding variables and the absence of randomization that would help isolate the effect of a single variable. Some conceptual frameworks, theories, innovative designs, or sophisticated statistical methods may need to be employed to infer causal relationships.

Eviction is a complex issue. Scholars have tried to connect factors at various levels with the eviction rate to better understand the eviction crisis, but many factors cannot be explored due to challenges in eviction data collection and analysis processes. The next section explores CSS, current research methods, and challenges in eviction research—most importantly, which aspects of CSS could be a promising approach to fill the gaps in eviction research.

Summary

This chapter outlined the scope of the eviction problem, including its magnitude and severity, effects, and trends. Regarding magnitude and severity, eviction affects a vast population in the United States and disproportionately affects certain groups based on race, ethnicity, and gender. Eviction may be associated with or even cause housing insecurity, physical and mental health issues, and the reproduction of poverty. Regarding trends, rental costs and progressive tenant protections are both increasing, with indications that eviction will likely persist in the near future.

This chapter also discussed factors associated with or causing eviction at the individual, community, and macro levels. In current eviction studies, most factors have been examined at the neighborhood or census tract level, with few researchers focusing on the individual case level. The dependent variable in most research has been the eviction filing rate, eviction rate, or in some cases, the number of evictions. Most studies have established associations between factors and eviction but struggled to construct causal inferences.

The next chapter explores how CSS has been used in eviction research, along with existing methodological issues and challenges. It also highlights the research methods used for this dissertation.

Chapter 3: Methodology

This methodology chapter briefly introduces CSS, covering its definition, applications, and benefits. It discusses how CSS has been using in eviction research, including for data collection, variable creation, prediction with models, and machine learning, along with associated challenges.

Following this, the chapter details the benefits of CSS in this study, from dataset creation to analysis—specifically, how raw data from PDF court files can be converted into tabular data suitable for modeling. This section provides a comprehensive overview of the methods for data collection, extraction, and linkage and the analytic models employed.

Computational Social Science

Definition and Applications of CSS

CSS is analogous to computational approaches in other fields, like computational biology and finance. All computational approaches fall under the umbrella of computational science, which is defined as a domain of study that creates new knowledge through networks, computers, software, algorithms, simulations, and so on. It is hard to track when the term *computational social science* was coined. It roughly dates to the late 20th century and the invention of modern computers (Cioffi-Revilla, 2010). For example, Schrodt (1987) introduced statistics and numerical processing with three mainframe packages—SPSS, BMDP, and SAS—for social scientists. However, CSS has no constant definition due to its continually evolving technology and theory development. Lazer et al. (2020) defined CSS as the "development and application of computational methods to complex, typically large-scale, human (sometimes simulated) behavioral data" (p. 1060). Cioffi-Revilla (2017) defined CSS as an "interdisciplinary investigation of the social universe on many scales, ranging from individual actors to the largest groupings, through the medium of computation" (p. 1). Despite slight differences between these two definitions, they share many commonalities. The term *computational* indicates the research method, whereas *social science* reflects the field for application or the purpose of research. Moreover, to distinguish CSS from traditional quantitative social science that primarily focuses on the structure of two-dimensional spreadsheet data like rows of cases and columns of variables, CSS emphasizes investigation in various formats, such as languages, images, and locations, through the medium of computing techniques (Lazer et al., 2020).

Social scientists are gradually employing CSS in their research. Metzler et al. (2016) surveyed social scientists ($N = 9{,}412$) and found that a third reported some involvement with research using computational methods, 79% participated in interdisciplinary teams, and more than half used administrative data.

Given the diversity of researchers devoted to social science fields, their computational techniques also vary. Cioffi-Revilla (2010) divided CSS into five main areas: automated social information extraction, social networks analysis, geospatial analysis, complexity modeling, and social simulation modeling. For instance, Onnela et al. (2007) examined the communication patterns of mobile phone users to understand information diffusion. Wimmer and Lewis (2010) tested racial homophily through social network analysis via Facebook. Ihara et al. (2012) used agent-based modeling to test role capabilities and motivations in caregiving. Satellite imagery and mobile phone data have become an alternative but common method to measure poverty rates (Aiken et al., 2020; Khan & Blumenstock, 2019; Smith-Clarke et al., 2014).

With the development of new techniques, other areas emerged, like data visualization, natural language processing (NLP), and machine learning. For example, Gentzkow et al. (2019)

described how to use text as an input in economic research. Evans and Aceves (2016) applied NLP and text analysis to discover hidden regularities worth theorizing. Pleasants et al. (2023) used text analysis for women's health and contraception research.

CSS has also been used in the social welfare field. Ren et al. (2020) noted the application of computational methods in social welfare should include a commitment to social justice and equity compared to general CSS. Applications in social welfare have addressed racial segregation, social services, child welfare, foster care, and other areas (Elgin, 2018; Israel & Wolf-Branigin, 2011; Jayaprakash et al., 2009; Ren & Bloemraad, 2022; I. M. Schwartz et al., 2017). These applications bring new possibilities for social scientists to advance research and understanding of society.

CSS Advances Social Research

Conte et al. (2012) stated the emergence of CSS is an appropriate response to the increasing complexity of the social world. Big data and advanced computational approaches benefit scientists by expanding their understanding of social phenomena and their mechanisms. Due to its research paradigm, CSS provides opportunities for interdisciplinary collaboration, accelerates the development of research data infrastructure, and introduces computational methods to new fields (Lazer et al., 2020).

First, CSS has encouraged collaboration among fields such as sociology and computer science and in domains beyond academia. For instance, the Solve for Good program at Carnegie Mellon University recruits nonprofits and public-sector actors with challenges related to data and links them with volunteers (mostly scholars) to find solutions. This project evolved from the Data Science for Social Good initiative at the University of Chicago.

Second, CSS and data infrastructure mutually accelerate social science research. On the government side, beyond census data, more administrative data became available on government websites after the Open Government Initiative. Currently, 294,477 datasets are available at Data.gov (as of February 21, 2024). In academia, the Harvard Dataverse provides a platform for researchers to manage and share their data with the academic community (Dataverse, 2024). Private companies like real estate marketplace Zillow have built application programming interfaces for their users to access their data easily (Zillow, 2024). Other infrastructure such as open-source programming languages like Python and R and version control websites like Github and Gitlab have also become popular in the CSS community. These tools increase the transparency and replication of social science research and help social science researchers follow reproducibility standards (Camerer et al., 2018).

Third, CSS has fused computational methods from other fields with researchers' domains, informing new research questions or designs. For example, NLP methods and machine learning in computer science have been applied to massive social media data to explore critical information, evaluate public sentiment, analyze drug-dealing characteristics, and so on (Kouloumpis et al., 2021; Li et al., 2019). Researchers' domain knowledge can also reveal issues related to computational methods and evaluate negative aspects of technology, such as bias embedded in algorithms (Eubanks, 2018).

Some scholars in social welfare have shared similar ideas about applying computational methods in welfare research. Berzin et al. (2015) encouraged social workers to understand how technology can be incorporated in their practice. Coulton et al. (2015) addressed how technology generally changes how government and nonprofits make decisions and emphasized computational methods like data science will be required in social welfare research, too.

In recent years, the development of pretrained NLP models has the potential to reshape research methods in CSS concerning natural language. For instance, OpenAI's GPT models have seen exponential improvements, from 0.17 billion parameters in the first-generation GPT-1 in 2018 to approximately 1 trillion parameters in the latest GPT-4 in 2023 (Achiam et al., 2024). This rapid evolution of large language models (LLMs) has enabled machines to understand complex natural language, consequently transforming workflows in social science research. Two observable trends include transitioning from supervised to unsupervised learning and shifting from meticulously crafted to more flexible, less-designed methods. People previously had to manually label images and instruct machines on where and how to extract pertinent content. Although these supervised models were designed well, they were typically only applicable to the task to which they were tailored. However, with the recent emergence of LLMs, researchers no longer have to train the models personally. Instead, they can retrieve information through suitable prompts, such as "return the defendant's name." This advancement has streamlined the process, broadening the applicability of AI in data retrieval tasks.

However, there are some concerns related to CSS. Generally, data privacy and ethics standards or rules are not rigorous during data sharing in CSS, especially data that may include personal information like health records (Pietri, 2013; Salganik, 2018). Many questions need to be answered before sharing, like what kind of data should be shared, what kind of documentation is necessary when sharing, and how to evaluate ethics in collecting and sharing data. Some algorithms and libraries are available for programming languages like Python or R. CSS scientists may call on these resources for various tasks. However, these resources may include biases or issues that may cause inaccurate analysis or conclusions and worsen social inequities, such as pretrained word-embedding models for NLP that show gender and racial bias (Bolukbasi et al., 2016; Manzini et al., 2019; Zhao et al., 2019). Algorithms or models may not fit social science and make meaningless predictions (Salganik et al., 2020). Finally, Heiberger and Riebling (2016) stated that social scientists are generally not well prepared to use CSS due to a lack of training in programming techniques. Traditional training in one field of expertise in academia may also decrease effective communication and increase conflicts in real-world collaboration.

Despite its challenges, CSS presents an opportunity and promising path to explore society's patterns, behaviors, and rules. Embracing computational methods but testing them critically is one lesson from the literature. The next section discusses what kinds of computational methods exist and how these methods can be used to explore and estimate eviction issues.

CSS Applications and Challenges in Eviction Research

This section covers the typical data sources in eviction research, how research methods have evolved in this domain, the kinds of applications that integrate with computational methods, and potential issues that might arise.

Data Sources. Data sources in eviction research can be categorized into two main types: eviction filing records and supplemental datasets. Eviction filing records primarily contain information about specific eviction cases, whereas supplemental datasets provide information on associated variables or causes of eviction, such as housing and neighborhood information.

Given the complexity of the eviction process, collecting data on all eviction cases is challenging. For example, illegal lockouts occurring outside the courtroom may not be documented. Thus, the data discussed in eviction research mostly involve filed (i.e., legally documented) evictions. Moreover, some states, like California and Wisconsin, restrict access to eviction records from the public in certain circumstances to protect tenants (Desmond et al., 2018). Eviction filing records are the primary data sources for current eviction research. For instance, the Eviction Lab's methodology report indicated it had compiled more than 80 million records (Desmond et al., 2018). Regarding these records, 16 states have centralized systems that hold eviction records. Twenty-seven states, New York City, and the District of Columbia report aggregate county-level data. For county-level eviction data, the Eviction Lab submits annual requests to state and county court systems, receiving data from 2,204 counties across 46 states. The lab also acquires public eviction records from two private companies, LexisNexis Risk Solutions and American Information Research Services, documenting local-level cases (Gromis, et al., 2022).

Legal Services Corporation, a nonprofit corporation established by the U.S. Congress, scrapes digital data from county court systems to create structured datasets. It has collected data from 32 states, 18 of which offer complete state coverage. Data from 13 of these 32 states include name and address information, with five providing complete state coverage in 2024.

Other regional studies have used court eviction files, especially from county court systems. For example, Cookson et al. (2018) and Thomas (2017) used data from King County, Washington. McCabe and Rosen (2020) acquired data from the Washington, DC Superior Court system. The Evicted in Oregon team at Portland State University focused on evictions in Oregon and gained insights into the local legal landscape through mixed methods, including data analysis, interviews, and courtroom observations (Bates et al., 2021). If researchers access individual or case-level data, they may acquire additional information such as plaintiff and defendant names, addresses, amounts owed, and reasons for eviction, compared to aggregated data. Descriptive analysis of this data (e.g., address) can be very informative and help build an understanding of eviction issues.

Although national-scale efforts by the Eviction Lab and Legal Services Corporation have been successful in collecting high-level data for many jurisdictions and detailed records for a smaller number of regions, these organizations cannot capture less accessible records or data only available in unstructured formats. For example, as seen in Figure 2, a screenshot from

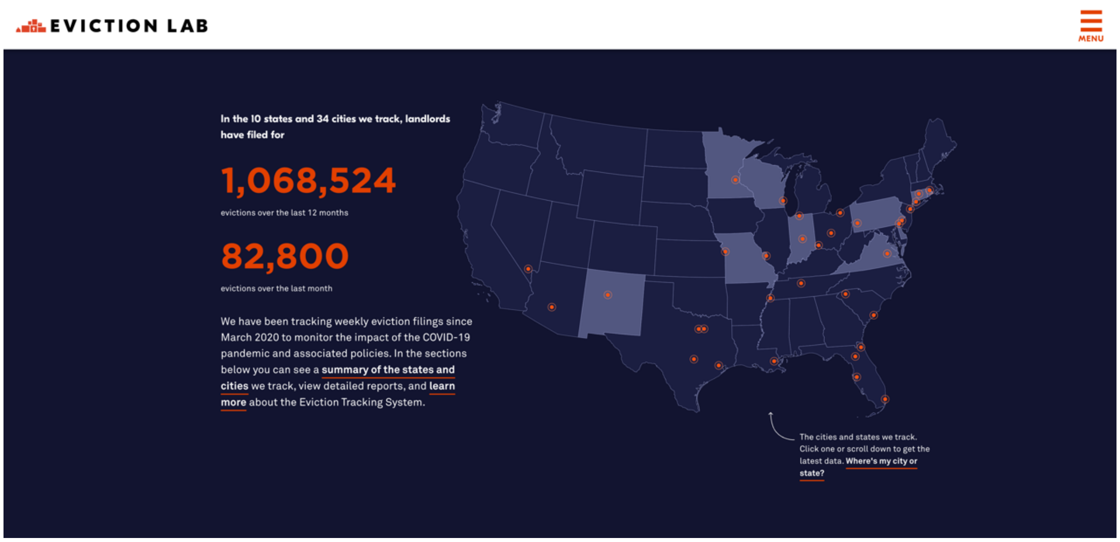


Figure 2: A screenshot from Eviction Lab tracking system website. The tracking system tracks weekly eviction filing since March 2020. However, many Western states are missing (Eviction Lab, 2024).

Case Number	Court Number	Action Type	Defendant First Name	Defendant Last Name	Plaintiff First Name	Plaintiff Last Name	Street	City	State	Zip
123	456	Filing	Jane	Doe	John	Adams	12 Main Street	Albany	NY	12345
123	456	Filing	John	Doe	John	Adams	12 Main Street	Albany	NY	12345
123	456	Judgment	Jane	Doe	John	Adams	12 Main Street	Albany	NY	12345

Figure 3. A screenshot of Eviction Lab database sample

Eviction Lab's tracking system website, many eviction-related documents on the West Coast are in unstructured formats, thus the absence of Eviction Lab's tracking system in many Western states.

This court information can be linked to supplemental datasets to explore associations between eviction data and other factors under the guidance of theories and concepts (e.g., ecological theory, neighborhood effects concept). Census data, including the ACS and Household Pulse Survey, represent one of the most popular options. Census data can be used to link neighborhood or community-level demographic and economic data to eviction information. Housing data can also be merged. For instance, McCabe and Rosen (2020) matched property records from the Office of Tax and Revenue to eviction data in Washington, DC. These housing data can reveal property owners and housing structures (e.g., number of units). Scholars also matched other datasets for specific research purposes. Richter et al. (2021) and Collinson and Reed (2019) linked administrative data, particularly on public assistance programs like Temporary Assistance for Needy Families and Medicaid with eviction filing data. Humphries et al. (2019) linked eviction data to credit history data. Ramiller (2021) connected evictions with property-level data in Seattle, Washington, to investigate the relationship with property sales.

Eviction research relies heavily on eviction records, and the following subsection discusses how these eviction records become available for linkage and study.

Data Structuring. Scholars can request eviction records from the court system. As mentioned previously, states or countries may provide these data. Usually, data at the individual level are difficult to acquire, but aggregated information is often available through annual court reports. If personal or case-level data can be shared and have been transferred into electronic databases, then scholars could use these records after simple data wrangling. Most original data that the Eviction Lab acquired come in this format (see Figure 3; Desmond et al., 2018).

However, data usability usually varies by provider. In other words, providers may collect different variables. For example, sometimes the databases may include tenants' names and addresses, but sometimes addresses are not included. In one study in Washington, DC, researchers needed to hand-code certain key information, including the amount of rent owed and presence of legal representation (McCabe & Rosen, 2020).

Another situation is that scholars can access case-level information, but this information is nested in court files instead of a structured database. When researchers retrieve court files, there are usually several PDF or image (e.g., TIFF) files for each case. Each case represents a story of an individual or family and can be used to partially reconstruct the eviction process. Acquiring this valuable information from court files often involves considerable time and effort. Cookson et al. (2018) reported that "a team of researchers reviewed each court document in these

1,218 cases and created a database" (p. 16). After structured data are generated, linkage, application, and estimation are possible.

Popular Applications. Addresses in natural language are not enough to reveal meaningful information until geocoded. Geocoding is a process of transferring text-based addresses into a coordinate system. ArcGIS, Google Map, HERE, and OpenStreetMap are popular tools for geocoding. After geocoding, latitude and longitude values can be applied to various base maps for visualization and spatial analysis (e.g., most eviction cases occurred on the south side of a certain city). Geocoding also links these cases to census data at the block, tract, or ward level. Then, that demographic information can be associated with eviction records.

Race, ethnicity, and sex estimation is another common application in eviction research. Usually, eviction records do not have demographic information in these domains, but several tools can be applied for imputation to test disparities in evictions. Desmond (2012) used first names in the Social Security system and research assistants to impute the sex variable. Cookson et al. (2018) and Thomas (2017) used an R package produced by Imai and Kabir (2016) to estimate sex via first name and race and ethnicity via last name using Bayesian prediction with location data. Hepburn et al. (2021) used the same package for race and ethnicity but used gender (Mullen, 2021), genderizeR (Wais, 2016), and Gender API (not specify) for sex prediction. If names happened to be in other datasets with sex, race, and ethnicity information, then either extract or fuzzy matching can be applied to assign demographic information to each name.

After these imputation processes, several conclusions can be derived from estimation. For instance, Hepburn et al. (2021) observed that Black tenants experienced the highest average eviction filing rate (6.2%) and women were more likely to be evicted than men. Scholars who apply these similar methods typically come up with analogous results, like Black women have higher risk of eviction than other groups.

Another popular application is eviction rate estimation. Usually, the eviction rate is based on simple calculation, description, and comparison. The Eviction Lab's methodology report described how it assigned several states a low estimation label after comparing the number of cases from court files and the number of cases from private companies such as LexisNexis (Desmond et al., 2018). Cunningham et al. (2021) from the Urban Institute determined the eviction rate during the pandemic based on their Coronavirus Tracking Survey. The Urban Displacement Project at the University of California, Berkeley also tried to build a housing precarity risk model to estimate communities at risk of postpandemic eviction across 17 metro areas via a Bayesian additive regression tree model with variables like the percentage of Black residents.

As machine learning becomes a popular method for estimation or prediction, it has been increasingly applied in eviction research. Ye et al. (2019) tested machine learning on rent-stabilization policies in New York to predict potential landlord harassment. Another project in San Francisco used machine learning (e.g., recurrent neural network) to identify populations at high risk of eviction (Tan, 2020). A student capstone project called HOME at the University of California, Berkeley examined eviction filings in 16 U.S. metropolises to predict eviction rates in California (Kambath et al., 2021).

Current Issues in Eviction Research Methods. Indeed, computational methods have provided many possibilities for computational research, but several issues cannot be ignored. This section discusses issues ranging from data collection to prediction.

Current data collection is heavily based on the court system. If the court system does not have a data system with structured data, then research possibilities are limited. The Eviction

Lab's solution is to pay for the data, which might not be feasible for other research teams. Even though state court systems build these kinds of data systems, Kleysteuber (2007) found that housing court records contain data-entry mistakes. Porton et al. (2021) systematically reviewed the inaccuracy of 3.6 million eviction records in 12 states and found that 22% of eviction records contained ambiguous information or falsely represented a tenant's eviction history. Data collection issues may cause inaccurate estimation.

Researchers also have imputed demographic information, like race and gender. The accuracy of these processes needs more validation. Xie (2022) compared several packages for race prediction by name in a dataset. The results showed an obvious difference: One package predicted that 20.83% of the dataset was Black, whereas another package predicted 8.93% was Black (Chintalapati et al., 2023). A point-by-point comparison is necessary to ensure disparity estimations are not significantly wrong. Moreover, is it ethical to use defendants' names for prediction? An example in South Africa indicated an IBM tool resulted in unethical application and worsened segregation (Eubanks, 2018).

Moreover, geocoding is a typical process for spatial analysis, and studies have explored geocoding accuracy. According to Eviction Lab's methodology report, 93.7% of addresses in its data can be geocoded into coordinates or street-level data. Although the remaining percentage is low, the absolute value is close to 4.5 million cases. So far, the cause and impact of this accuracy are unknown.

Finally, the eviction rate is represented as a number, yet there are real individuals behind the statistic. Although making predictions is always challenging, they are nonetheless essential. Eviction research can have a significant impact, reaching audiences beyond the academic realm. Therefore, presenting this information accurately is crucial. In the Urban Displacement Project's housing precarity risk model, San Francisco was omitted from its map because an advocacy group argued that the estimations were not accurate in some areas. Consequently, the research team removed the map of that area from its presentation. However, feedback loops like this are not always common. Given these issues, the question arises: Why have we not been able to resolve them? The following section discusses challenges in current eviction research.

Methodological Challenges in Eviction Research. These issues prompt questions about the challenges we still face. How do these challenges affect current research or potentially mislead researchers? This section discusses challenges encountered throughout the research pipeline, including data collection, estimation, and the potential for integration with other fields.

Data Collection. Data collection is one of the most challenging yet crucial steps in the research process. Even for cases in the structure format in the legal system, their value has yet to be fully explored. Due to infrastructure, legislation, and political issues, researchers may access these data in various formats and at different levels of granularity. For instance, Washington provides structured data on eviction filings but does not include address data, and some cases are missing when compared to cases in the county system (Cookson et al., 2018). Additionally, researchers can access scanned court files in PDF or TIFF formats. Taking California as another example, the counties surrounding the San Francisco Bay Area do not provide court documents for each case. Instead, they offer a summary table from the sheriff with basic information for eviction cases, including plaintiff and defendant names, dates, and addresses in PDFs (T. Thomas, personal communication, September 15, 2021). These factors necessitate a significant investment from research teams to extract essential information.

Then, even if addresses can be extracted, to increase geocoding accuracy, researchers need to standardize them. However, the address formatting is often messy. For example, abbreviation formats are not consistent. Sometimes the apartment number is preceded by a number sign (i.e., #) and sometimes by an abbreviation (e.g., Apt.).

Moreover, information like room number and closest street intersection may also decrease geocoding accuracy. The Eviction Lab's solution was to create many rules via regular expression (e.g., replace "st" with "street") to make addresses clean enough for the next step. However, new mistakes can always arise, and it is hard to account for all situations. Thus, getting a clean address efficiently is a challenge in eviction research (Desmond et al., 2018).

In terms of the causes of evictions, the typical method is to send out surveys for data collection. Although most cases are related to nonpayment, the causes of nonpayment have not been explored in court documents. Two issues may be behind this challenge. First, a lot of time is needed for qualitative analysis. Second, several documents (response or answer files) are handwritten, which can be hard to decipher.

The outcomes of evictions might be ambiguous or incorrect in state datasets, as Porton et al. (2021) highlighted. Regarding the eviction process, the sheriff typically handles evictions. Although court files usually include a sheriff return, which is a report from the sheriff to the court, this information may not be updated in the court system. Communication loss can occur between the sheriff's office and court office. For example, eviction data from Washington are from the court system, whereas eviction data for California are collected through sheriff's records, and scholars may not be aware of cases that were dismissed in the court system.

Estimation and Prediction. Although data collection can cause challenges related to estimation and prediction, these processes also have their challenges. For estimation, CSS researchers typically rely on Python or R software, which each offering several options. Although several packages are available for predicting race, ethnicity, and gender, it can be unclear which option is the best and whether its predictions are accurate. Most packages come with validation datasets, but there is often little systematic comparison or guidance in these packages' documentation. Therefore, CSS requires researchers to critically evaluate and judge the functionality or performance of these packages.

In terms of prediction, especially for machine learning applications, many teams lack an eviction scholar with solid domain knowledge, which can be a significant limitation. Data shortages may also hinder sound predictions. For instance, Kambath et al. (2021) used metropolitan data from other states to predict California's eviction rate, including data from rural areas in California. Moreover, although theoretical frameworks can guide the selection of predictors, supplemental resources, like property-level data, are still not rich enough. Moreover, in these model estimations or machine learning predictions, the dependent variable is the eviction filing rate or judgment rate at the census tract level, with no consideration of the outcome of each case after eviction filing.

Theoretical Framework

According to the literature reviewed previously, a significant challenge in eviction research from a methodological perspective is collecting eviction records in counties without a digitized system at the individual level and converting this information into a meaningful format for research. In eviction research, most studies have focused only on eviction filings or eviction data—a one-time snapshot—at the community level. The events occurring after eviction filings and factors associated with these events remain understudied. Thus, I aimed to answer the following three questions in this study:

1. **RQ1: What kinds of computational methods can be applied to extract information, such as addresses, from unstructured court files in PDF format**

with high accuracy? The purpose of this research question is to explore the possibilities of making long-term unused PDF files in the court system analyzable. Structuring the data is foundational for further research. Especially in eviction research, this will help fill the gap in at least three states where I have data access so far and contribute to the national eviction database. A pipeline that combines multiple computational methods will facilitate the collection and analysis of these documents for other court file-related research, such as medical bill costs.

2. **RQ2: What factors at individual, community, and macro levels will influence whether an eviction case is dismissed or judged by the court?** The purpose of this research question is to examine factors related to the first-stage outcome of an eviction filing, whether the case is withdrawn by the plaintiff (dismissed by the court) or moves to judgment by the court. There are two implications of dismissal or judgment. First, in some states, the defendant can request the records of a dismissal case be sealed from public access, which can mitigate the potential negative impacts of having an eviction filing on their record on their future ability to rent. Second, a dismissal indicates resolution before judgment. It does not necessarily mean the tenant stayed in the property, but it may indicate the tenant moved to another place or came up with a payment plan with the plaintiff. Thus, I am interested in what kinds of factors influence the outcome at this stage.
3. **RQ3: What factors at individual, community, and macro levels influence whether the case moves to eviction by the sheriff after the judgment?** The purpose of this research question is to examine factors associated with sheriff eviction. After the judge issues a judgment and orders a writ, the case moves to the sheriff, who typically posts a writ of restitution on the tenant's door and can enforce the writ after 72 hours. In Pierce County, Washington, the enforcement of the writ typically occurs 1 to 2 weeks after posting, depending on the availability of the sheriff. After judgment, some tenants leave and some stay. At this stage, their eviction records usually become publicly available. The difference is that those who choose to stay may be forcibly evicted by the sheriff, which can result in violence, temporary homelessness, or the need for emergency shelter. Thus, addressing this question could identify high-risk factors for sheriff eviction.

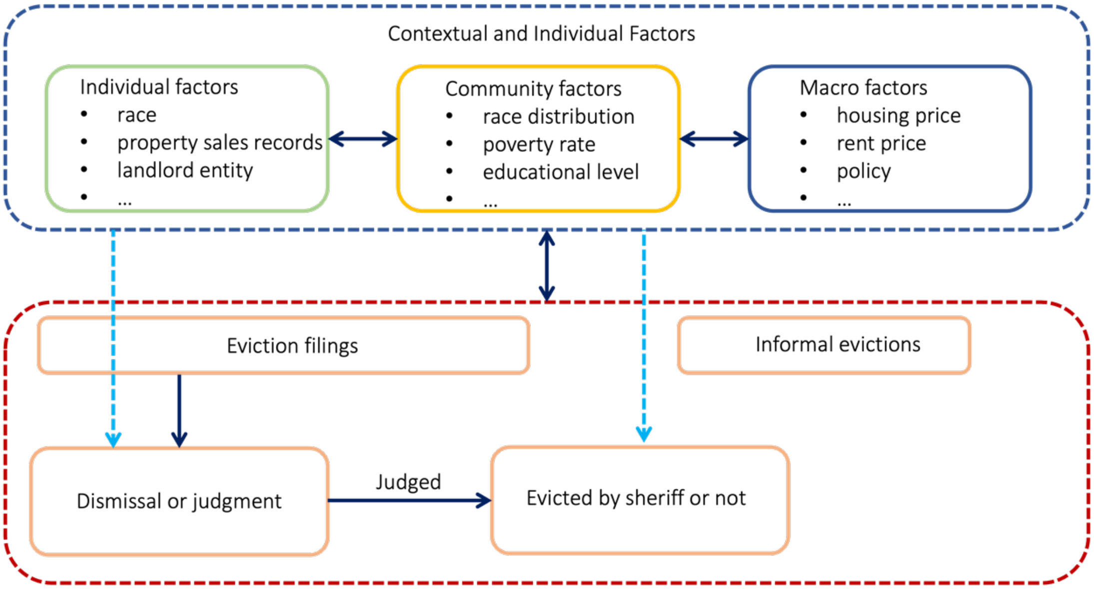


Figure 4: Theoretical Framework

The theoretical framework (see Figure 4) shows the interactions between multilevel factors and evictions. These two major components influence each other. Most studies build associations between these two big components, such as which neighborhood factors are associated with the eviction filing rate. In this framework, I also attempt to build connections between these factors and outcomes at different stages, including whether the case is dismissed or results in eviction by the sheriff.

Research Design

Data Collection

The data for this research were sourced from the state of Washington, primarily focusing on Pierce County, covering 2004–2022. Pierce County, encompassing the Seattle–Tacoma metropolitan area, has an estimated population of 928,696, according to the U.S. Census Bureau (2023). Tacoma is its largest city. Eviction policies vary among states and counties, making a county-based sample a viable starting point for studying eviction. Most counties in Washington use the same data format, PDF, to store eviction data, so the method in this study can be expanded to the entire state. Pierce County has established an account for researchers, allowing them access to court files through the Pierce County Legal Information Network Exchange.[2] Although the public usually pays for each document download, researchers with an account can download these court files for free. On the website, researchers need a case number to locate details (PDF files) of superior court cases. Thus, Pierce County provided a separate digitized dataset with case numbers of unlawful detainer in the county. This dataset was formatted as comma-separated values and featured key columns such as county name, case number, names of defendants and plaintiffs, and names of attorneys (if applicable) from 2004 to 2022. However, the information in digital form was very basic and lacked important details, such as the location of the eviction. After obtaining the spreadsheet with eviction case numbers, I wrote a web scraper that automatically logged into the website, navigated to the correct webpage, input the

[2] https://linxonline.co.pierce.wa.us/linxweb/Main.cfm

case number, select all related PDF files, downloaded the files, and renamed them (i.e., by case number).

After downloading these PDF files, each case received a folder and PDF files were placed in their corresponding folder (see Figure 5). The dataset featured 56,070 unique cases and

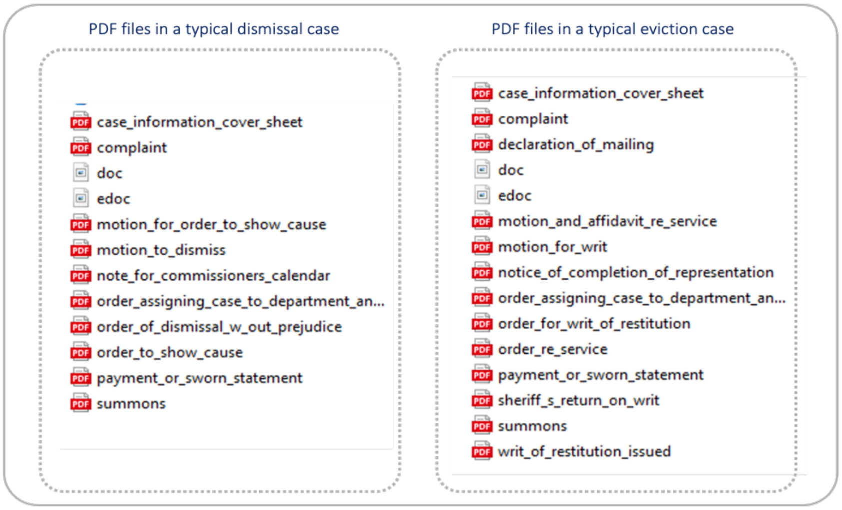


Figure 5: PDF files in different cases

approximately 3,000 eviction court cases per year in Pierce County, except for 2020, 2021, and 2022, during the pandemic. Each case featured printed and occasionally handwritten documents. The number of PDF files per case typically ranged from four to 20, depending on the complexity of the eviction process. The total number of PDF files was 772,629. The PDF files were scanned versions of court documents and usually included summons, complaints, judgments, sheriff writs, or returns of service for each case. Notably, eviction notices from landlords before filing were not included in the dataset, reducing the number of eviction cases that could be tracked.

Data Extraction from PDFs

Extracting information was central to data collection, because most court files contained unstructured data. Converting these data into a structured dataset was foundational for this eviction research. Key pieces of information, including names of plaintiffs and defendants and housing locations, are typically extracted for this kind of research.

The conventional method involves training research assistants to read through the documents carefully and manually type or copy and paste crucial data into a spreadsheet. This method, although effective, is labor intensive, time consuming, and challenging to replicate. For instance, given a sample of 10,000 cases and the need to extract information from each summons, requiring 20 seconds per file, it would take at least 55 hours to complete this task, not including time to process other documents like sheriff's writ returns.

Thomas et al. (2024) sought to streamline the process of extracting addresses from court summonses by developing rule-based methods. They used optical character recognition technology to convert summonses into text files, then crafted rules to identify address patterns using regular expressions, such as strings that begin and end with numbers. Thomas et al. (2024) also experimented with customizing named entity recognition techniques for addresses. This is an NLP technique used to extract entities from sentences. For example, given the sentence "John grew up in Berkeley, CA," this technique can recognize "John" as a name and "Berkeley, CA" as a location. They customized the technique to recognize address patterns. For instance, in the sentence “Jane lives at 123 ABC Street, Berkeley, CA, and likes to play basketball,” the customized recognition technique would extract “123 ABC Street, Berkeley, CA.” However, this approach has limitations in real-case applications for summonses. For example, if a page contains multiple addresses, such as those of the defendant, attorney, and court, it may not extract the correct one as the eviction address.

Thus, in this study, I explored two advanced techniques: one based on computer vision and the other on NLP, specifically focusing on LLMs. First, I conducted document layout analysis combining computer vision and NLP. This method learned both the location and content of interest, scanned the page, and transformed the information into a structured dataset. Given that summonses in one county generally follow a similar format, this approach tended to yield reliable results. However, if there was a significant change in the format of the summonses, it became necessary to relabel images and retrain the supervised neural network model to accommodate the new format.

Second, LLMs, like the GPT models, offered new possibilities for this task. The previous method relies on supervised models with well-defined structures, typically tailored to specific tasks. In contrast, LLMs can adapt to new formats they have not previously encountered. This flexibility means that if a law firm introduces a new summons format unfamiliar to the model, LLMs might still be able to process it without making errors. In the subsequent subsections, I introduce how to set up these two different methods, document layout analysis, and LLMs and how to extract information accurately.

The goal was to extract names of both plaintiffs and defendants and any amount of money owed if a judgment summary was present. If a case file included a PDF with a name suggesting it was a sheriff return writ, this document likely contained the outcome of the eviction, such as “moved before the sheriff showed up,” “ejected by sheriff,” or some other outcome. Optical character recognition (OCR) techniques were employed to convert these results into a structured dataset. A general preprocessing step involves converting PDFs to images, because images are more manageable for information extraction tasks, such as selecting the first page of a PDF.

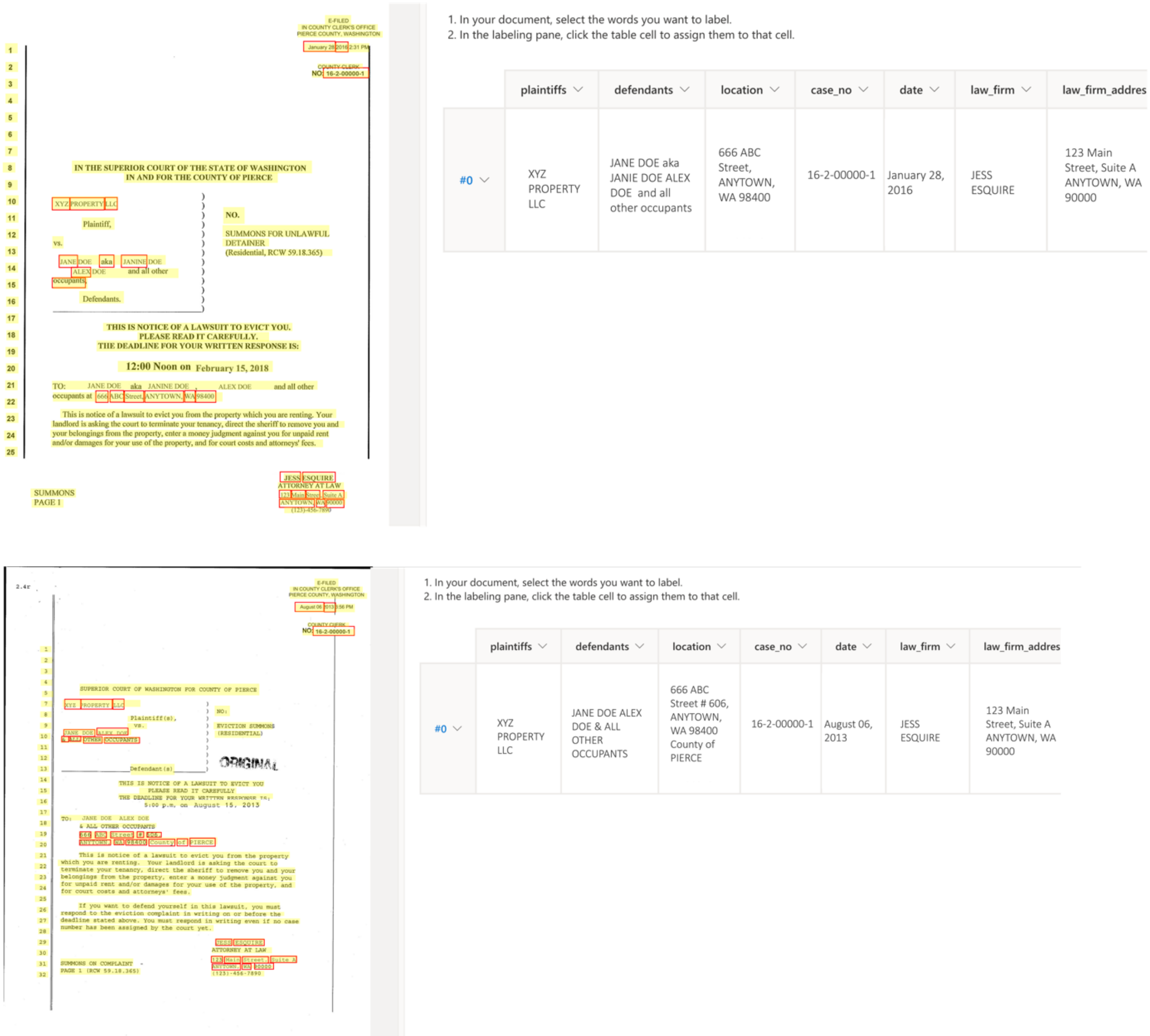


Figure 6: A labelling case in the Microsoft Azure Document Intelligence Studio

Document Layout Analysis. The document layout analysis used the Microsoft Azure Document Intelligence Studio platform. This online tool facilitates easy labeling of information and training of the neural network model using Azure Cloud Computing resources.

Summons. Because summonses are initiated by law firms and these firms use roughly similar but not identical templates, it was necessary to train the model on as diverse a format range as possible. Additionally, e-filing was not popular before 2010, and documents not e-filed often present more complicated situations. Therefore, I developed two models. One identified information from summonses issued after 2010. However, for those not e-filed—indicated by a seal on the first page—these documents were processed again by the model using a training set derived from summonses issued before 2010.

For example (see Figure 6), first, I uploaded the image to Microsoft Azure Document Intelligence Studio, reviewed as many law firm formats as possible, and selected around 20 pages with a unique format for each model. Then, on the platform, I created a table with columns like defendant's name, plaintiff's name, location, law firm name, and law firm address. I highlight the content of interest on the image, and the chosen content was transferred as text into the table I created. After labeling all the images, I trained the model in the cloud. Once the model was successfully trained, I tested it with 10 random first pages of summonses. If one image failed, I added that image to the training set and retrained the model until all new random images passed the test. After successful deployment on the cloud, I used Python to communicate with the model online and retrieve the results after processing.

Motion for Judgment. The judgment amount of money is from the motion for default judgment. This file usually appears in the folder if the attorney for the plaintiff hasn't heard anything from the defendant after serving the summons. As shown in Appendix Figure A3, the motion for default judgment typically includes a table-like summary showcasing the principal amount of money owed, attorney fees, and filing fees that the defendants are asked to cover. Similar to the method for summonses, different types of judgment summary files were selected and labeled in Microsoft Azure Document Intelligence Studio. Variables such as case number, principal judgment amount, attorney fee, filing fee, total amount, attorney for judgment creditor, and law firm address were extracted. After the model passed all 10 randomly selected pages at this stage, the model was deployed and communicated through codes built by Python, a programming language.

Sheriff Return Writ. As previously discussed, after the court issues a judgment, a writ for restitution is granted, allowing the plaintiff or their party to schedule an eviction with the sheriff. The relevant document is the page that the sheriff returns regarding the writ. Specifically, the sheriff reports back to the court, informing it of outcomes such as the status of posting. In the sheriff return writ, the format tends to be more consistent compared to the judgment summary and summons, because it originates from the sheriff's office rather than various law firms. I randomly selected about 10 pages for training and another 10 pages for testing to ensure the model accurately extracted the information I need. The variables extracted from this page included the case number, defendant's name, plaintiff's name, posted status, and return status. Then, the model was deployed on the cloud. The labeling sample for a sheriff return writ is shown in Appendix Figure A4.

For return status, different case files report different categories:

1. Ousted and Ejection: A typical description for this category is, "On February 28, 2010, I ousted and ejected, with the aid of the County, the defendant from possession of the premises described in said Writ of Restitution, and placed the plaintiff in possession of the premises." This indicates the case ended in eviction by the sheriff.
2. Peaceful Possession: "On August 22, 2007, I placed the plaintiff in peaceful possession of the above described premises." This also indicates the case resulted in eviction by the sheriff, but with a peaceful status.
3. Return Per Attorney: "At the request of the attorney for the plaintiff, on April 4, 2007, I am returning said Writ of Restitution. Per the attorney, the defendants have moved." This indicates the defendant moved before encountering the sheriff.
4. Return Per Plaintiff: "At the request of the plaintiff, on February 20, 2007, I am returning said Writ of Restitution. Per the plaintiff, the defendant has moved." This also indicates the defendant moved before encountering the sheriff.
5. Expired: "On December 9, 2014, I am returning said Writ of Restitution unsatisfied. The Sheriff's Office has not been contacted to schedule an eviction, and as of

December 8, 2014, the Writ of Restitution has expired." Because the case expired, very likely the defendant had moved and the landlord did not need to request sheriff service.

Large Language Models. LLMs offer an alternative method for extracting the necessary information from images. Given that LLMs are designed for language processing and understanding rather than image processing, I used Tesseract, an optical character recognition engine, to convert the content from images to text. However, this conversion process eliminates the structure of the text, meaning that table-like information is transformed into a continuous paragraph of text. Following the OCR process, I employed the OpenAI API through Microsoft Azure. According to its privacy terms, the prompts (inputs) and completions (outputs) are not available to other customers and not accessible by OpenAI. The model I used was text-davinci-003, which is based on GPT-3.5, similar to the initial version used in ChatGPT. Python was used to communicate with this model, which is hosted on the cloud as an application programming interface. The responses were received and saved in a structured table format. I tested the LLM only on the summons documents, because using the model is significantly costly in both financial and temporal terms.

Dependent Variables

Dismissal or Judgment. The first dependent variable was a categorical variable. In this context, 1 indicated the case was dismissed, whereas 0 indicated the case moved to judgment. The criterion for identifying a dismissal was the presence of a PDF file containing the string "dismiss." Similarly, if a file contained keywords like "judgment" or "restitution," then the case was labeled as a judged case. In some special situations, the attorney for the plaintiff may submit a motion to dismiss (see Figure 5) but ultimately decide to continue with the case. Therefore, if a case was labeled both 1 and 0, then the case was considered to have proceeded to judgment.

Sheriff Return Status. As previously described, there are several statuses for the sheriff return on the writ of restitution. I labeled "ousted and ejection" and "peaceful possession" statuses as 1, which indicated an eviction by the sheriff. The rest—i.e., "expired," "return per attorney," and "return per plaintiff"—were labeled as 0, which meant they were resolved before eviction by the sheriff. The distinction here is that in scenarios categorized as 0, the tenants moved before facing sheriff enforcement, thereby avoiding direct confrontation, whereas in scenarios categorized as 1, the tenants were forcibly or peacefully removed by the sheriff, which could result in immediate homelessness and limited options for housing. Although "ousted and ejected" and "peaceful possession" represent different severities of sheriff intervention, they were grouped together because both involve tenants moving at the last minute, possibly rendering their housing situation more vulnerable compared to those who relocated before sheriff involvement. Several circumstances could have result in a category 0 status, such as no follow-up action from the plaintiff after the writ was posted (expired) or the tenants having already moved (return per attorney or return per plaintiff).

Data Linkage

Individual Level. Meta data were contained in an HTML file in each folder (i.e., case) that showed the list of files in each case. The information included case number, defendant name, plaintiff name, case filing date, return writ date (if present), and the number of days between filing and return writ. Typically, the difference between filing and return indicates the processing time of the case.

In the extraction process, case number, defendant name, plaintiff name, address, law firm, and law firm address were extracted from summonses. Case number, defendant name, plaintiff name, address, posting and return statuses, and content were obtained from the sheriff return writ. The defendant and plaintiff names may not always be consistent at different stages. For example, the meta data may use "Jane Doe VS ABC Property," whereas the summons may use "Jane Doe and all other people in the property." Thus, I used meta data for names, and if the meta data were missing, I used the name from the summons or return writ. Beyond basic information, racial and ethnic data are not generally collected in court cases. Some eviction studies have attempted to infer race and ethnicity based on the defendant's name. Despite known biases with this method, it is still a potential approach for current practice because other independent variables may limit its influence in the model if there is strong bias. A similar method was used for estimating the gender of the defendant. The relevant R packages are "rethnicity" for race and ethnicity and "gender" for gender. The race and ethnicity assignment also considered census information on racial and ethnic distribution, with the final assignment based on the highest conditional probability from "rethnicity" and "census race/ethnicity distribution" packages.

Regarding plaintiffs in individual cases, the plaintiff's name indicated whether the property was owned by a private individual or an organization. Thus, I applied named entity recognition with BART, a pretrained NLP model produced by Facebook, to classify whether the plaintiff was an organization or individual (Lewis et al., 2019). I ran the BART model on three sources (meta data, summons, return writ) because they might have differences, such as ABC Company or ABC LLC. Then, I used a voting system. If two sources identified the plaintiff as an organization and one as an individual, the plaintiff was classified as an organization.

Judgment data were extracted from the judgment summary file with variables such as case number, principal amount, attorney fees, court fees, total fees, and law firm. The principal amount is usually the rent owed, and the court fees are usually charged by the court for filing for eviction. At the individual level, data were connected by case number because they all came from the same case at different processing stages.

Property Level. To connect individual cases to the properties they involved, spatial join techniques are necessary. Visualize each property as a polygon on a map, referred to as a parcel. If an address is converted into a point with latitude and longitude that falls within a parcel polygon, then the property can be linked with the individual case.

First, I used the address extracted from summonses as the primary address. If the address was missing for any reason, I used the address from the sheriff return writ. Second, I applied batch geocoding in ESRI software to convert text-based addresses into latitude and longitude. If the latitude and longitude did not return as "point of address" or "subaddress," which means the address fell outside the parcel polygon due to being incomplete, mismatched, or for other reasons, I then processed the remaining addresses through Azure Maps. The advantage of Azure Maps is that it attempts a best guess, although the accuracy is generally lower than that of ESRI.

After the addresses were geocoded, they were connected with parcel records through spatial matching. If the location existed and was correctly geocoded, then the point fell within the parcel polygon. The address may not be 100% correct due to potential mismatches, such as an incorrect apartment number but a correct street address, and the point may fall on the street. In these cases, the nearest parcel number was assigned to that address. The following are the data sources of the property level data.

Parcel Record. Prior research suggested that eviction rates may be higher for large rental properties, such as multifamily units, and the parcel number serves as the unique identifier to connect property sales records (Rutan & Desmond, 2021). The parcel record[3] was sourced from the Pierce County WA Open GeoSpatial Data Portal (v2.1). As of December 22, 2023, the dataset featured 336,079 records and was updated daily. In addition to the polygon location of each parcel, the other relevant columns were parcel number, taxable value, and land use description. The land use description had various categories, with "single family" being the predominant category.

Sales Record. Property sales records[4] were obtained from the Pierce County Assessor-Treasurer's Office, documenting property transactions. The dataset was from 1997 to 2023, with 597,369 records. From this table, I used variables such as parcel number and sales date. By grouping parcel number and sales year and counting the frequency, a table was generated showing the number of sales for each parcel each year. This allowed me to connect the eviction case to indicate the number of sales that occurred 1 year before eviction, 1 year after eviction, and the year the eviction occurred.

Permit Record[5]. Previous studies indicated that permit approvals—for instance, demolition permits—are significant indicators of gentrification and displacement (Ramiller, 2022). Permit records were obtained from the Planning and Public Works PALS Plus application. PALS Plus is an end-to-end system for permitting, covering building permits, development engineering permits, environmental permits, etc. Cases that were canceled were filtered out, resulting in 531,324 valid records from 1985 to 2023. Each case was geocoded into latitude and longitude, allowing for linkage to the parcel number using spatial matching. After assigning the parcel, I grouped the parcel number and date of permit to count the number of permit approvals for each parcel each year. This data was then connected to the eviction case to show how many permits were accepted 1 year before eviction, 1 year after eviction, and the year the eviction occurred.

Low-Income Housing Tax Credit[6]. According to several studies, tenants in subsidized housing are disproportionately likely to be evicted (Gromis et al., 2022; Leung et al. 2023). The low-income housing tax credit (LIHTC) is a federal program that provides tax credits to housing developers who agree to have a certain number of rent-restricted units for lower-income tenants. The Property Level Data LIHTC, administered by HUD, contains 52,006 projects and 3.55 million housing units placed in service between 1987 and 2021 (HUD, 2022). I filtered the LIHTC projects in Pierce County ($N = 97$), converted addresses into geolocations, and then linked them with parcel polygons. If a parcel was linked, it was labeled 1.

---

[3] Parcel Record: https://gisdata-piercecowa.opendata.arcgis.com/datasets/piercecowa::tax-parcels/explore
[4] Sales Record: https://www.piercecountywa.gov/736/Data-Downloads
[5] Permit Record: https://open.piercecountywa.gov/dataset/Permits-Pierce-County/nhnt-v7ka/about_data
[6] LIHTC: https://www.huduser.gov/portal/datasets/lihtc/property.html

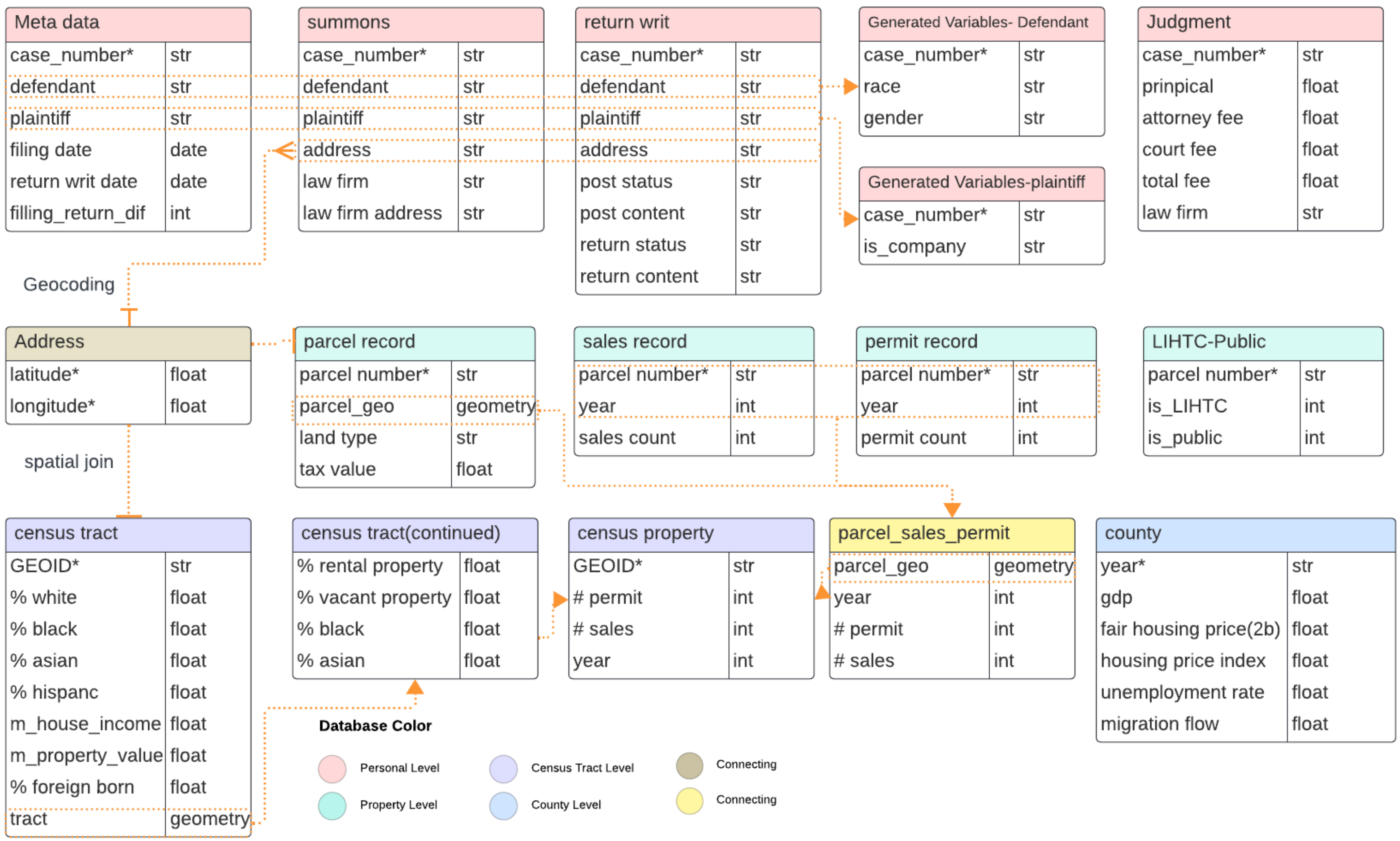


Figure 7: The database structure diagram. This figure shows how datasets at different levels connect.

Public Housing. The list of public housing[7] was sourced from the Pierce County Housing Authority and Tacoma Housing Authority. The former's website listed eight properties, and the latter listed 15 properties. The addresses of these apartments were geocoded and linked to the corresponding parcels, which were labeled 1.

Community Level. Community-level data were sourced from the ACS. In this research, I collected 5-year ACS data at the census tract level, because some required variables were not collected at the block group level and the 5-year data are more reliable (U.S. Census Bureau, 2024a). Tracts typically have a population size between 1,200 and 8,000 people, with most averaging about 4,000 people (U.S. Census Bureau, 2020a). Although ACS data collection began in 2005, it only covered areas with populations exceeding 65,000. Therefore, comprehensive data for Pierce County at the census tract level start in 2009. Consequently, the community characteristics for eviction cases from 2004–2009 used data from 2009, whereas other years used the data from the end year, such as ACS 2011–2015 for 2015. Similar to merging geopoints to parcels, I also used spatial join to merge addresses with census tracts and attach corresponding community characteristics for each case.

Race and Ethnicity. The population composition by race and ethnicity was defined by U.S. Census Bureau data. There were five categories for race: White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or other Pacific Islander, and others. The ethnicity category counted the percentage of Hispanic or Latino residents, which is often abbreviated to "Hispanic." This ethnicity is defined "as a person of Cuban, Mexican, Puerto Rican, South or Central American, or other Spanish culture or origin regardless of race" (U.S. Census Bureau, 2020b).

Poverty. Poverty was calculated as the proportion of the population living under the 100% poverty threshold in each census tract. The U.S. Census Bureau assigns each person or family one of 48 possible poverty statuses based on their age, family size, inflation, and other factors, then compares this to their income. If the income-to-poverty threshold ratio is less than 1, then the person or family is considered to be living in poverty.

Foreign-Born Status. This variable refers to the percentage of the population born outside the United States. This percentage is close to the percentage of noncitizens but not exactly the same in some situations. For example, a foreign-born person can become naturalized as a U.S. citizen.

Median Household Income. The median household income of each tract was the midpoint of the household incomes in the tract. It is an important measure of the economic status of an area.

Percentage of Rental Properties. The percentage of rental properties in a census tract was calculated by dividing the number of rental properties by the total number of occupied housing units in that tract.

Percentage of Vacant Properties. The percentage of vacant properties was calculated by dividing the number of vacant units by the total number of housing units in the census tract. Here, the denominator was total housing units, including both occupied and vacant, whereas the denominator for the percentage of rental properties was only occupied housing units.

Rent Burden. HUD defines households as "rent burdened" when they spend more than 30% of their income on rent. This percentage may influence the affordability of other necessities (HUD, 2018). Although whether 30% is a reasonable measurement remains debated, it is a

---

[7] Public housing data were from the from Pierce County Housing Authority: https://www.pchawa.org/apartments. Public housing lists were from the Tacoma Housing Authority: https://www.tacomahousing.org/housing/properties/.

common standard to assess rent burden. The percentage of rent burden reflects the proportion of households paying more than 30% of their income on rent.

Count of Sales and Permits. Because the geometry of parcels can overlap with census tracts spatially, I used sales records and permit records to create new variables by counting how many sales and permits applications occurred in each census tract.

County or Macro Level. At the county or macro level, focusing on Pierce County, economic indicators such as the unemployment rate and gross domestic product were sourced from Federal Reserve Economic Data. Regarding the housing market, the House Price Index was collected to illustrate overall sales price trends in the county.

The House Price Index, provided by the Federal Housing Finance Agency, spans from 1975 to 2022. The calculation of the index is complex, typically based on the Repeat-Sales Index, and is calibrated using appraisal values and sales prices for mortgages purchased. The data from this category were merged based on the year when the eviction occurred.

Data Validation

To answer RQ1 regarding what kinds of advanced computational methods can be applied to extract information and test their performance, a validation process was essential to demonstrate the reliability of the information for subsequent analysis. Two critical validation steps were designed: one for the extraction results and another for the geocoding results.

Extracting Validation. To validate the performance of the information extracted from summonses, judgment summaries, and sheriff return writs, I randomly selected 200 pages from summons and sheriff return writs and 100 pages from judgment summaries, then manually extracted the correct information into a spreadsheet. In the spreadsheet, one column contained machine-extracted information, and another contained human-extracted information. Then, I used the Levenshtein distance to compare each pair in Python (Bachmann, 2023; Levenshtein, 1966). For example, if the location was “123 S. ABC Street, Tacoma, WA, 95555” per human extraction and “123 5 ABC Street, Tacoma, WA, 95555” per machine extraction, the difference would be “S.” and “5,” requiring two steps to reconcile: (1) Replace “5” with “S” and (2) add the period after “S.” The first string features 36 characters and the second string has 35 characters. So, the total length is 36 + 35 = 71, and the number of matching characters is 71 – 2 = 69. Thus, the Levenshtein similarity ratio is 69 ÷ 71 = 0.97. After comparing each pair, I calculated the mean and median for each category, such as the mean score for addresses in summonses.

Geocoding Validation. The geocoding validation assessed the accuracy of the address type after geocoding by ESRI and Azure Maps. I validated and demonstrated how many addresses accurately represented an eviction location, with some located in the middle of the city or even outside of Pierce County. If the geocoding was incorrect, then the associated community characteristics would not be considered reliable.

Data Analysis

The first part of the analysis was descriptive, focusing on filing trends from 2004 to 2022, the rate of dismissal and judgment by year, and the outcome of sheriff executions. I visually represented eviction counts based on geographic density and provided a descriptive table for the variables used in the models. To make easier convergence of the models, I also standardized the independent variables before feeding them into the models.

In the second step, I conducted several models to examine the relationship between independent variables and the dependent variables: case dismissed or judged for RQ2 and the outcome of sheriff execution for RQ3. I only used data from 2010–2019 because census tracts change every 10 years and could mismatch the community-level variable. Community data before 2009 were not collected by ACS and 2020 was a special year due to the pandemic. Models for analysis were logistic regression, hierarchical logistic regression (multilevel), and random forests. A multilevel model is necessary due to its ability to account for relationships at the individual, census tract, and county levels, distinguishing the weight of different levels. The intra-class correlation coefficient was calculated for each model to assess the presence of clustering in the data and whether a multilevel model was necessary. Random forest was chosen because it may reveal nonlinear relationships or interactions and tolerate multicollinearity between variables that logistic regression may not detect. To address concerns of multicollinearity in the logistic models, the variance inflation factor (VIF), a measure of multicollinearity, was used to filter out certain variables before finalizing the model. The VIF cutoff was set at 5 for most models and 10 for the macro model due to the smaller number of variables involved (Belsley et al., 2005; Hallgren, 2012). To evaluate the performance of logistic regression, I used the Hosmer and Lemeshow test to examine the goodness of fit by calculating the chi-square between the original and predicted results (Hosmer & Lemesbow, 1980). To investigate the random forest model and assess how each variable contributed to it, I analyzed the feature importance and used partial dependence plots to explain their relationships. To highlight commonalities and differences between logistic regression and random forest, I also calculated the marginal effects of the logistic regression.

Summary

In this methodology chapter I briefly introduced CSS, including its definition, applications, and benefits. Next, I discussed how CSS has been used in eviction research for data collection, variable creation, and prediction with models and machine learning. However, despite using CSS in eviction research, some issues persist, such as data collection from unstructured data, and new issues have emerged regarding how to ensure the reliability of the created variables. Next, I proposed three research questions based on the theoretical framework, seeking to determine: (a) how data collection from unstructured data can be optimized with CSS for RQ1 and (b) what factors are related to case outcomes after an eviction filing for RQ2 and RQ3.

Then, I introduced the process of establishing the dataset for analysis, specifically detailing how I converted raw data from PDF court files into tabular formatted data suitable for modeling. This chapter provided a comprehensive overview of data collection under CSS using the document layout analysis method, creating new variables based on current variables, and connecting data from individual, property, census tract, and county levels. The method for validation was also introduced to ensure the quality of the extraction and outline the types of models that were used to address RQ2 and RQ3.

Chapter 4: Findings

This chapter presents the findings from the data analysis. First, because advanced CSS methods were used to extract data from images, it was essential to validate these results to demonstrate their accuracy and reliability. Second, descriptive analysis are presented, including changes in the number of cases by year and by each stage of the eviction process, the distribution of dismissals and judgments, and the outcomes of sheriff evictions. Additionally, the processing times for eviction cases and the geographic distribution of eviction filings are examined. Third, results from various classification models, such as logistic regression and random forest, are shown to explore which factors influenced a case being dismissed, judged, or handled by the sheriff after judgment.

Validation Results Analysis

The first finding corresponds to RQ1 regarding the performance of methods used to extract information from PDF files and construct structured datasets. The validation results are presented, guiding which method was employed to extract information from the remaining documents in this study. Second, the results from geocoding are validated. Primarily, this involved checking whether the addresses were extracted accurately.

Information Extraction Validation

First, I needed to decide whether to use the layout analysis method or the LLM-based GPT method. To make this decision, I compared the results of addresses from the validation dataset I had created. The addresses were from summons files, and both results from the layout analysis method and the GPT method were compared to the manually extracted results. I report both the mean and median Levenshtein ratios. The layout method had a median ratio of 1.00 and a mean of 0.95, whereas the GPT model had a median ratio of 0.98 and a mean of 0.90. Although the median ratios were close, the mean ratio of the layout method was better than that of the GPT model; thus, I chose the layout method for the following tasks. Potential reasons for the GPT model's poorer performance are discussed in the discussion section. The following sections outline the performance of the layout model for different files.

Summons. As mentioned in the methodology chapter, the five main categories extracted from the summons were defendant names, plaintiff names, addresses, law firm names, and the law firm addresses. All five categories achieved a median Levenshtein ratio of 1.00 and a mean Levenshtein ratio greater than 0.95. The location category achieved the lowest mean ratio, which makes sense because its format is much more diverse than the other categories. For instance, sometimes the location may be expressed as "123 45th Street," but the image may not be of very high resolution, leading to the misrecognition of "th" as a quotation mark. Alternatively, the machine might extract extra information such as "resident at." This may not affect the final results significantly, because the geocoding software may correct the errors or automatically clean up some information, such as deleting "resident at." However, based solely on the string comparison, I still accounted for these differences. The law firm addresses achieved a higher mean ratio because there were not many law firms in Pierce County and unlike law firm names, which may use logo-like letters, law firm addresses are more standard, and the training model may already recognize the pattern.

Judgment. The performance on judgment files was more diverse compared to the summons files. Although the median Levenshtein ratio remained 1.00 for all six categories, the mean ratios for attorney for creditor (i.e., plaintiff) and law firm were lower, because some law firms included these two pieces of information on their motion for default judgment whereas others not. Thus, in some instances with no such information, the model may mistakenly select surrounding, related information. However, the primary purpose of analyzing these files was to extract monetary amounts owed, such as principal money (rent owed), attorney fees, and filing fees. The performance here is impressive; because the numbers were usually concise, even a small discrepancy could significantly affect the ratio. For example, in one case, the rent owed was initially recorded as 192, then corrected to 190 with a line for deletion. The model recognized it as "192 190," resulting in a Levenshtein ratio of 0.60.

Sheriff Return Writ. The sheriff return writ had a well-organized format. Despite different statuses, the ratios from the validation data for posting status, posting content, return status, and return content were almost 1.00 for both median and mean. There were some minor issues in the dataset, such as an extra space in the return status, but these issues were easily fixed before merging.

Location Validation

Not every case had a successfully extracted address from the summons and subsequent files, either due to insufficient specificity of the location or because the address for mailing the summons was outside of Pierce County. Additionally, even when an address was extracted, the geocoding process may have resulted in errors. For instance, the address may not be found in the map system or the map system only partially matched the address, such as an incorrect street number. Therefore, it was necessary to validate the locations and check how many cases had addresses and how many were successfully assigned census tract community characteristics.

Of 56,070 cases, 56,058 contained PDF files; 1,684 cases lacked an address from either the summons or sheriff return file. Among the remaining 54,374 cases, 96.97% of the locations could be converted into a specific location using either the ESRI or Azure Maps geocoding system, whereas the rest, 3.03%, could only be matched to a higher level, such as a street or postal area. Although some addresses were geocoded at the street level, it was still possible to assign the nearest property data or census tract data. When spatially merging with the census tract map file of Pierce County, 3,863 of the cases lacked a census tract ID. In other words, although some addresses were successfully geocoded, their locations might be outside of Pierce County. Indeed, I identified some cases in King County (adjacent to Pierce County). These 3,863 cases were excluded from the analysis requiring census tract information.

Descriptive Analysis

Change by Time and Category

Number of Filings by Year. After analyzing the cases during an 18-year period (2004–2022), the total number of eviction filings was 55,918. The annual number of filings was relatively stable, ranging from 3,000 to 3,500 from 2004 to 2019 (see Figure 8). The filings decreased to 2,892 after the 2008 financial crisis, likely reflecting improvements in employment, and then increased again, reaching a peak in 2015 with 3,630 filings. The peaks in 2005 and 2015, before and after the financial crisis, could be attributed to the housing boom, which led to higher housing values and consequently, influenced rent costs. After 2018, the number of filings never surpassed 3,000. Although there may not be empirical research on this issue, in 2019, Washington Gov. Jay Inslee signed a new eviction reform bill into law, extending the notice period for tenants to pay or vacate the property from 3 days to 14 days. This policy gave tenants more time to react before an eviction filing could be made. Later, the COVID-19 pandemic and related eviction protection policies drastically changed the number of filings, causing the number to decrease to around 500 in 2020 and 2021. However, after those policies, like the eviction moratorium, expired on November 1, 2021, in Washington, the number of filings jumped back to 2,016 in 2022.

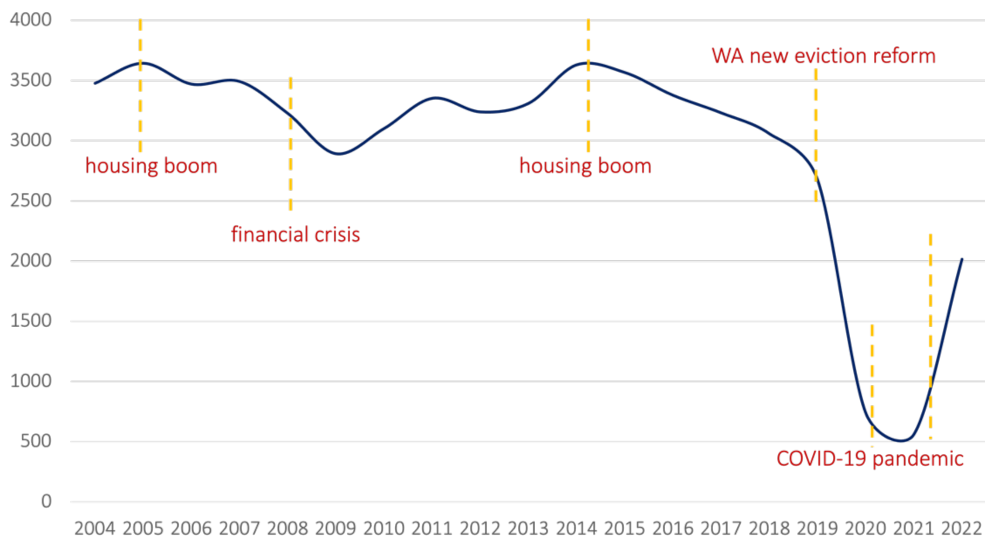


Figure 8: Number of filings by year from 2004 to 2022

Change at Each Eviction Stage. Each folder or case contained multiple files, and the presence of certain files, such as a sheriff return writ, indicated that the case had progressed through a specific process. Generally, the process follows a linear and simple sequence, including summons, judgment writ, and sheriff return. However, some cases were more complex and had multiple files with similar names like "sheriff's return writ 1" and "sheriff's return writ 2," with the first one having expired. Alternatively, some cases contained both "dismissal" and "judgment" documents, indicating that the case was initially unresolved and later took another direction.

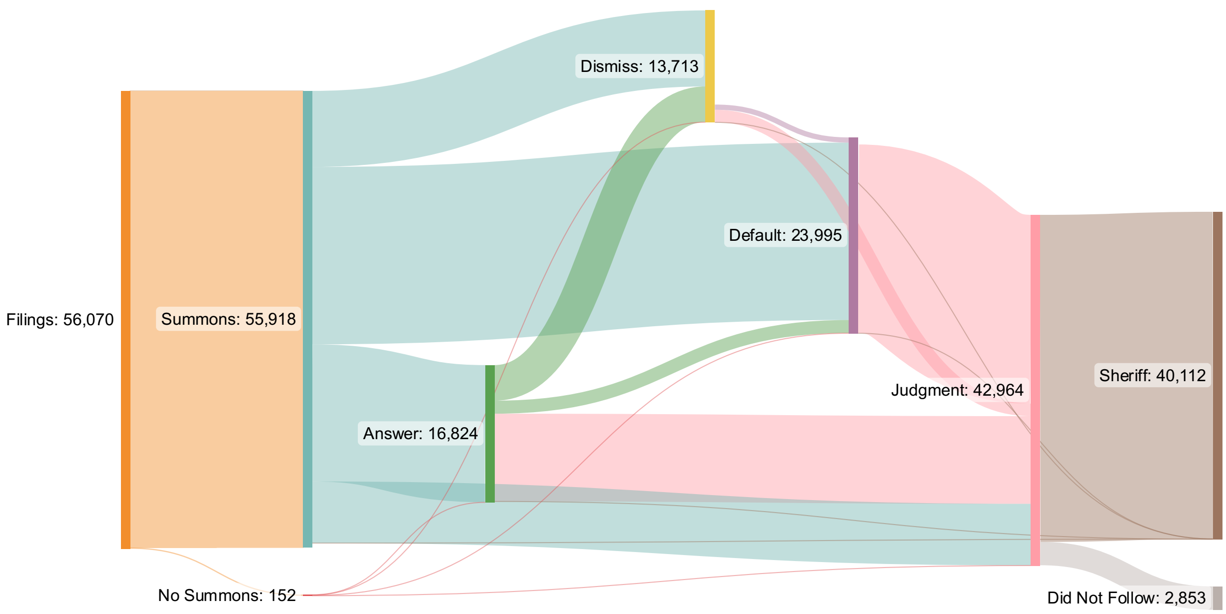


Figure 9: A Sankey diagram shows the flow of data from filling to various outcomes.

Outcome Rate by Year

Dismissal Rate by Year. According to the data, of 55,918 cases with a summons file, 30% (n = 16,824) had answers or a notice of appearance, indicating the defendant's request for a hearing (see Figure 9). However, 25.9% of cases with answers (n = 4,356) resulted in dismissal, leaving 13,713 cases. Here, dismissal didn't necessarily mean the defendant could continue living in the current place; it simply meant the case was resolved before judgment. Several situations can lead to dismissal, such as tenants moving out, paying the rent with a late fee to continue staying, or coming up with a payment plan. Approximately 43.8% of all cases were categorized as default (n = 23,995) because they contained a motion for default judgment, indicating the defendant did not respond to the summons. The default motion is almost always approved by the judge and quickly moves the case to judgment; 63.8% of the cases with an answer from the defendant also proceeded to judgment (n = 10,740). The judgment category featured 42,963 cases, comprising 76.8% of all cases with summons. Of the judged files, 93.4% were eventually handed to the sheriff (n = 40,056). For the remaining 6.6% of judged cases (n = 2,853), there was no sheriff return writ, which may indicate that the defendants won the case or the file was missing, necessitating a review of the judgment content.

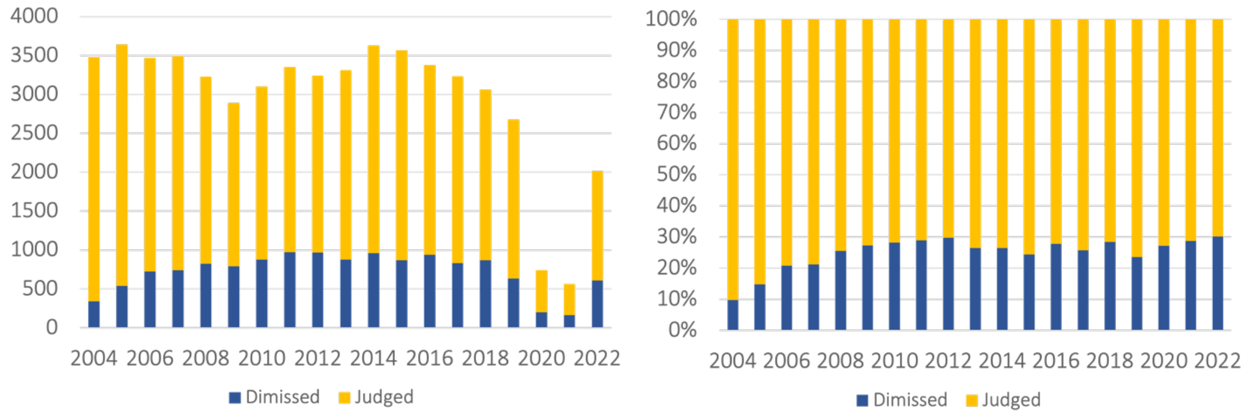


Figure 10: The dismissal rate by year; the left figure is based on the ratio of absolute value and the right figure is based on the ratio of percentage.

As the Sankey plot in Figure 9 illustrated, a case could have either been dismissed or proceeded to judgment. Thus, I calculated the ratio of dismissals to judgments by year (see Figure 10). Except for 2004 and 2005, the dismissal rate ranged from 20% to 30%, indicating that the case was resolved before reaching judgment. The dismissal rate was notably higher around 2011 and again around 2021, particularly in recent years during the pandemic, when the dismissal rate was slightly higher than the historical average. One potential explanation is that more cases were resolved outside of court because court processes slowed during the pandemic. The notably lower dismissal rates in 2004 and 2005, compared to other years, could potentially be related to the housing boom and changes in sales and landlord activities.

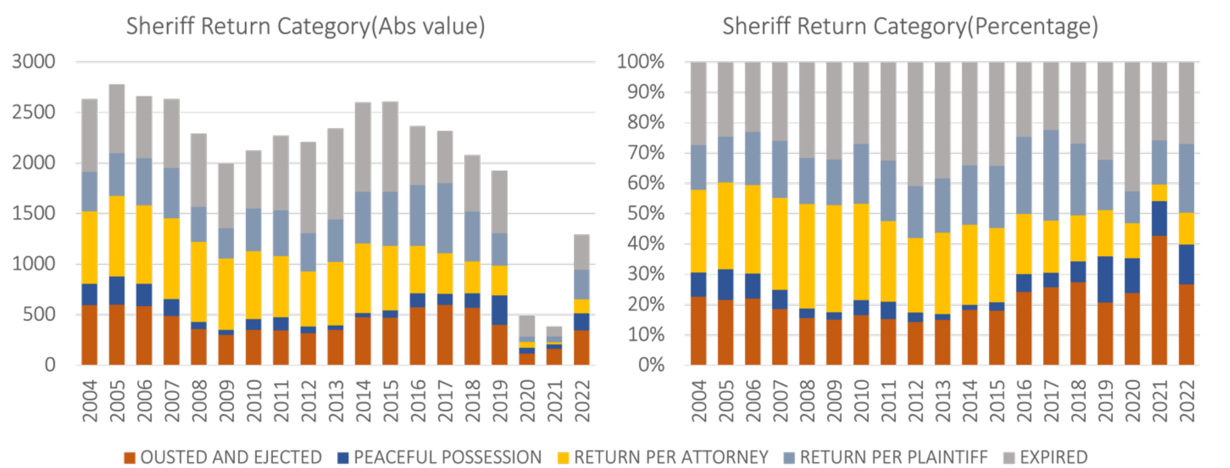


Figure 11: The different sheriff return writ rates by year; the left figure is based on the ratio of absolute value and the right figure is based on the ratio of percentage.

Sheriff Return Writ by Year. For the 71.7% of cases with summonses that eventually proceeded to the sheriff for eviction, the outcomes differed. Some tenants had already moved, whereas others faced the sheriff and were ejected at the last minute. As introduced in the method section, there were five main categories of sheriff returns: return per plaintiff, return per attorney, expired, ousted and ejected, and peaceful possession. These five categories indicate what happened when sheriff tried to process the writ. The distribution of these categories was unbalanced and fluctuated over time. The top three categories each year typically included return per plaintiff, return per attorney, and expired (sometimes referred to as return to court). These three categories usually mean that after the sheriff received the judgment writ (writ of restitution or writ of possession), they posted the notice to the tenants, asking them to leave within 3 days (The Tenants Union, 2024). If there was no contact, to schedule a date for eviction, from the plaintiff's side, the case expired. If the sheriff attempted to serve the writ but was informed by the plaintiff or plaintiff's attorney that the tenants had already moved, the return type was typically classified as return per plaintiff or return per attorney. Only if the tenants were still there and the sheriff physically executed the eviction would the return type be ousted and ejected or peaceful possession.

After analyzing the proportion by year, the proportion of last-minute actions (ousted and ejected or peaceful possession) increased since 2013, from 20% to nearly 50% in 2021, even though there were fewer filings after 2020 compared to other years. In 2020, the proportion of expired cases was higher than in other years. These two unusual findings suggest that the eviction moratorium could be a potential factor. According to the eviction moratorium initiated on February 29, 2020, in Washington, some sheriff departments may have paused serving eviction notices after the writ, and some local actions may have been taken before the state policy was implemented (Beekman & Brownstone, 2020). The data indicate that 81.8% of the expired cases occurred in February, March, and April 2020. In 2021, the proportion of last-minute moves (ousted and ejected or peaceful possession) surpassed 50% for the first time. This trend suggests that tenants may have had less capacity to find another place or move, such as not having enough money for a security deposit for a new location or they were still waiting for pandemic-related assistance, with the average wait time for assistance being 88 days, according to media reports (H. Smith, 2021).

Answers in Eviction Process

Median Processing Days of Eviction Cases. Regarding the duration from filing to sheriff execution, the median was around 25 days before the pandemic (see Figure 12). After the pandemic began, the median jumped to 40 to 50 days. As previously outlined, tenants had to respond to the summons within 7 days. If no answer was submitted, the case proceeded to default judgment; if a response was submitted, a hearing was scheduled based on availability of the court. Therefore, the median duration from filing to sheriff return with default judgment was about 20 days before the pandemic and increased to 25 to 35 days afterward. For cases with an answer filed, the median duration was 30 to 35 days before the pandemic and extended to 40 to 70 days afterward. Although the outcomes in these cases did not change, favoring the plaintiff, the duration from filing to sheriff execution with answers increased by about 10 days compared to cases without answers, providing tenants with more time to come up with other plans when facing eviction.

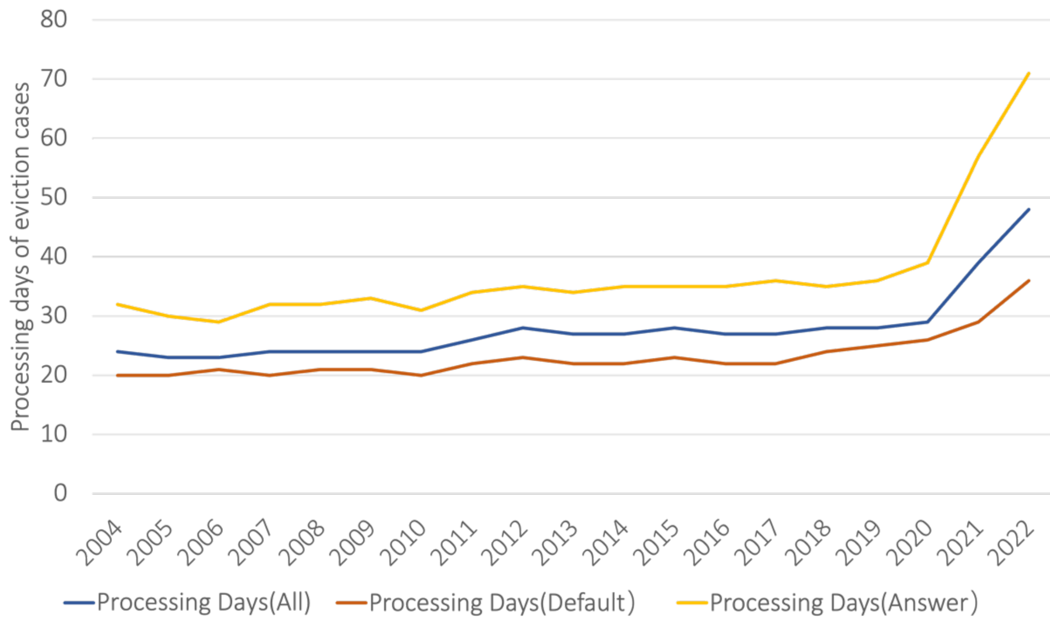


Figure 12: The median processing days of eviction cases by answering status.

Answer Rate Change over Time. Typically, the answer rate from summonses was around 30% before the pandemic, then increased after 2019 (see. Figure 12). In 2022, more than half of the cases included answer files. This partially explains the increased duration from filing to sheriff return, because more cases needed to be scheduled for hearings.

Financial Aspects

As shown in Figure 13, the median cost of cases with rent information in the judgment summary files steadily increased to 2019. The median principal, usually representing 1 month of owed rent, increased from a median of $1,000 in 2004 to $2,000 in 2019. The median attorney fee was around $400 before 2013, then increased to approximately $500 after 2013. The filing cost rose from around $200 to $400. According to the Pierce County Clerk of the Superior Court (2024) fee schedule, the civil filing fee is $240. The total cost of filing and attorney fees constitutes about 40% of the total judgment payment, with slight variations by year. The total cost and principal in 2022 were much higher than in any other year due to the accumulation from the eviction moratorium. After the moratorium expired and some tenants were sued by landlords, the principal was no longer 1 month of rent as before the pandemic but had accumulated for several months. Although the specific months of accumulation were not detailed, an estimation based on the data suggests around 3 months of owed rent. In 2022, as the number of filings and the amount of owed rent increased, the back-rent situation became not only more common but also more severe in the rental housing market. Many families may have anticipated rent relief, but it appears that some did not receive it, as indicated by the outcomes.

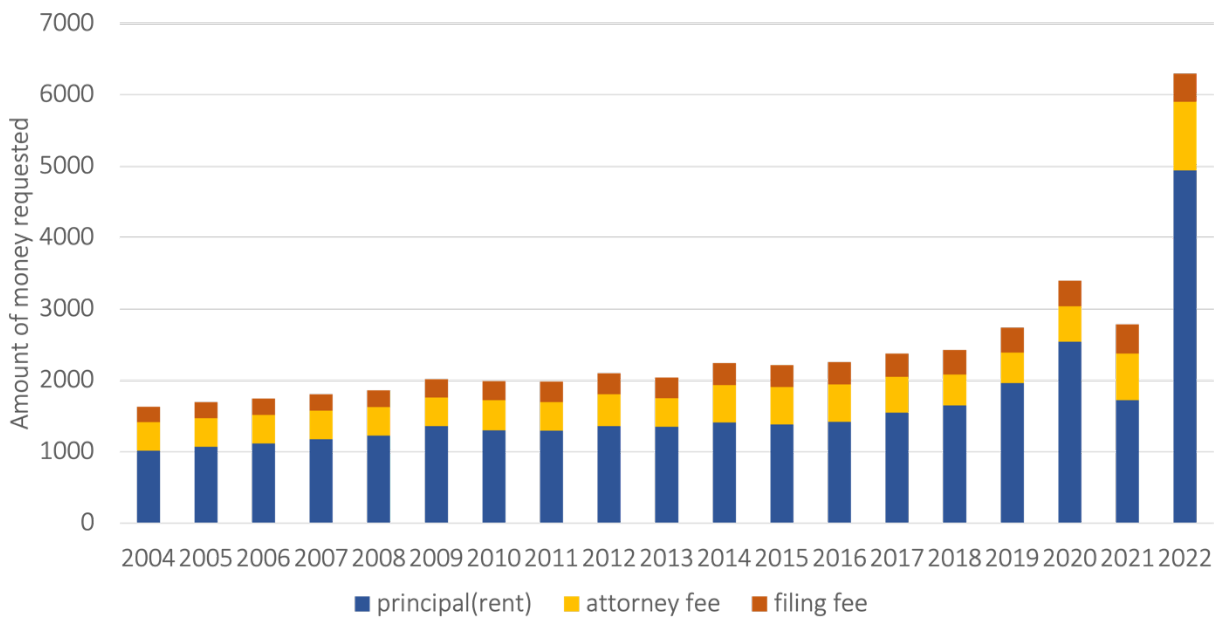


Figure 13: The amount and proportion of the money requested by the plaintiff by year.

Top Evictors

The top five evictors in Pierce County during the study period were Dobler Management Company Inc (*n* = 846), Equity Residential Properties (*n* = 665), Pierce County Housing Authority (*n* = 549), American Management Services Northwest LLC (*n* = 532), and Laurel Gardens WPIG LLC (*n* = 442). These five account for 5% of all eviction filings among 15,807 unique evictors.[8] On average, each evictor accounted for 3.5 cases, making these top five evictors obviously more active. Four of the top five evictors were companies, which owned more rental properties than small private landlords. Decker (2021) noted that larger-scale owners typically rely on highly routinized systems and are more likely to start eviction filings. The Pierce County Housing Authority, the county's public housing management agency, was among the top three evictors. This finding is not unique to Pierce County. Gromis et al. (2022) linked public housing data with estimated eviction filings and found that public housing authorities accounted for 5.8 of every 100 eviction cases, although only 3.5 of every 100 renting households lived in public housing. Leung et al. (2023) conducted in-depth interviews and analysis, revealing that rent collection is a key factor for the federal government to assess public housing authority management, but the number of evictions does not stem directly from the U.S. Housing Act of 1937 (HUD, 2021; Lead the Way, 2015).

Race and Ethnicity in Eviction Filings and Sheriff Returns

As mentioned earlier, race estimation was conducted to assign each individual a race or ethnicity. In this study, when multiple names appeared in the defendant list, only the first was used to estimate race and ethnicity based on conditional probability, using name and address for the best estimation. According to the data and Figure 14, assuming each case represented a unique person from 2004 to 2022 and using mid-year 2013 census tract race and ethnicity

[8] Some property companies may share the same parent company. This study distinguished the entity solely based on the property name in the court file.

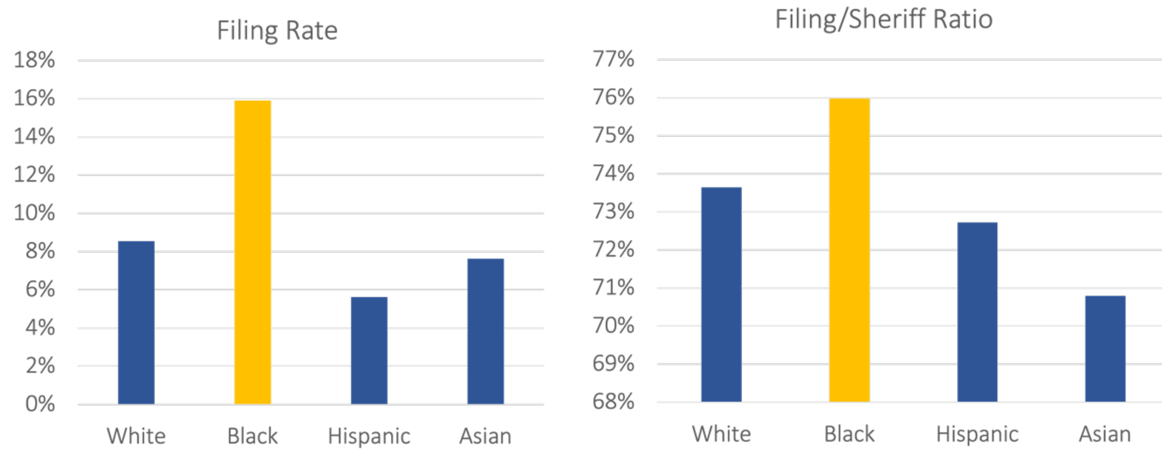


Figure 14: Filing rate and filing and sheriff eviction ratio by estimated race and ethnicity.

distribution to represent community conditions, 15.9% of Black adults experienced eviction filings compared to around 8% of White adults. The rates for Asian and Hispanic individuals were relatively lower. However, it is important to note that the race and ethnicity estimation method based on names typically performs worse in predicting people of color compared to White people. Therefore, the filing rates for Black, Hispanic, and Asian individuals were likely slightly higher than the data suggest. This aligns with other research indicating that Black individuals are more vulnerable to eviction. A recent study linking eviction records to personal census data found that Black women and men experienced the highest eviction rates, with Black renters making up around half of eviction filings nationally (Graetz et al., 2023). My estimation for Hispanics was lower than the national level. Although nationally, the rate for Hispanics is close to that of non-Hispanic Whites, Juan Pablo Garnham from the Eviction Lab suggested that data from courts represent formal evictions, but there may be many informal evictions, particularly in the undocumented Hispanic population (Linares & Telemundo, 2023). My underestimation of Hispanics could also be influenced by immigration status.

When comparing the filing to sheriff eviction ratio, which indicates how many filed cases proceeded to the sheriff's office after judgment, by race, the results show that 76% of Black individuals lost their cases, leading the court to order the sheriff to execute an eviction, whereas 73.5% of cases involving White defendants were sent to the sheriff. More Asian individuals were filed against but fewer led to judgment, whereas more Hispanic cases led to judgment despite fewer eviction filings. Overall, Black adults experienced a disproportionately high rate of eviction filings and a higher ratio of filing to sheriff eviction. This also reflects the disparity in homeownership rates, given higher homeownership reduces the need to rent and potentially face eviction. The higher proportion of Black adults living in rental properties may be associated with the disproportionate rate of eviction experienced from 2004 to 2022.

Geographic Distribution

Having discussed how eviction patterns have changed over time and the internal processes involved, I now address where these evictions occurred in Pierce County during the past 18 years. The geographic hotspots analysis provides an intuitive way to understand where eviction filings occurred and may benefit a diverse audience. For example, this information can assist public administrators in making informed decisions about resource allocation or help social work practitioners identify target areas of interventions, such as sharing knowledge about tenant rights. A heatmap (see Figure 15) based on density highlights the hotspots of eviction filings, with red indicating areas with the most filings (> 160) and blue denoting areas with fewer than 20 filings from 2004 to 2022. One hotspot is near Tacoma, with two moderate hotspots near

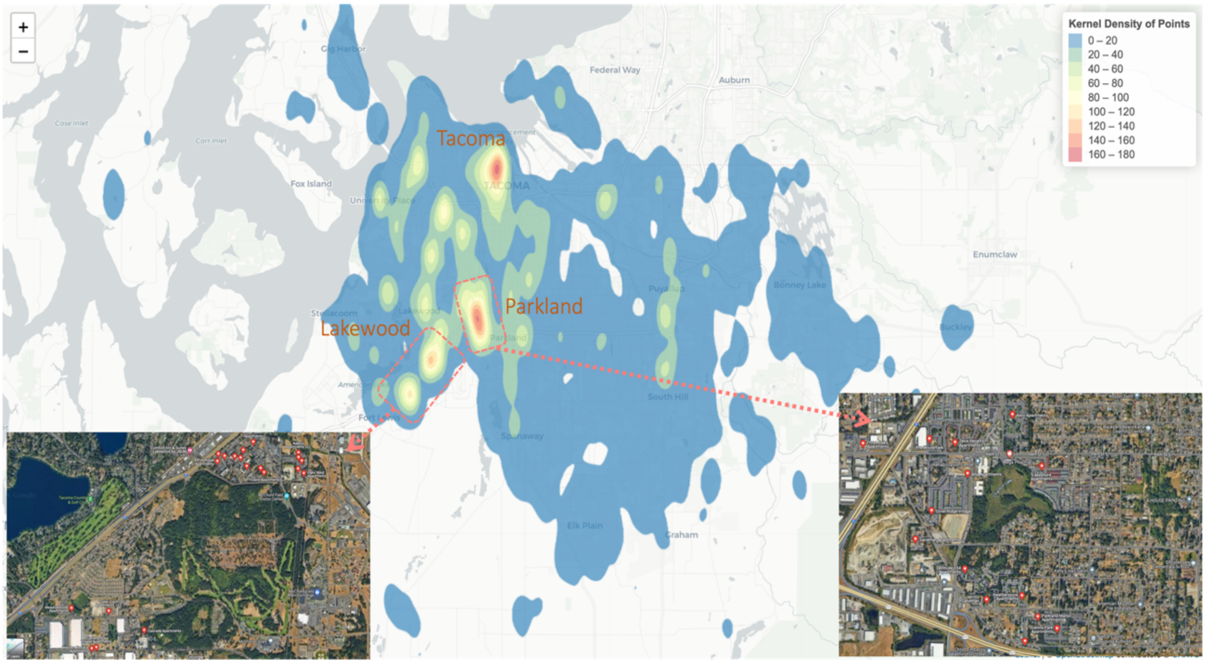


Figure 15: Heatmap based on eviction density. The red spots on the satellite map indicate apartments from a Google Map search.

Parkland and Lakewood. The primary hotspot is in downtown Tacoma, an urban setting with many single families, primarily in ZIP code 98402. According to the ACS 2022 5-year estimation, 23.6% of residents in this ZIP code lived below the poverty line, compared to 10% in Washington. The median household income was $57,486.

Another hotspot, in Parkland, is on the northeast side of the intersection of highways 5 and 512, a densely populated area with apartments. Census Tract 717.04 aligns with the orange area on the map. In this census tract, 88.5% of housing units were occupied by renters, of whom 78.6% were non-White, 19.6% were living below the poverty line, and the median household income was $49,138. The other spot, in Lakewood, exhibited characteristics similar to Census Tract 717.04, with many rental apartments. This makes sense, given evictions typically occur between renters and landlords. These apartments were usually near highways. After investigating several apartment websites, in 2023, rent was around $1,400 for a 1-bedroom, 1-bathroom apartment and $1,600 for a 2-bedroom, 2-bathroom unit. The median household income in these areas was generally below the state median of $91,306.

When comparing the eviction filing density map with the race and ethnicity map (see Appendix Figure A5), the two hotspots in the south coincide with areas where racial and ethnic minorities reside. Although the hotspot in the downtown area does not overlap with the highlighted areas where minorities live, the large green belt on the east side of Tacoma is also an area with a significant presence of racial and ethnic minorities, indicated by the light area on the race and ethnicity map.

I also examined the dismissal, judgment, and sheriff execution rates by census tract but did not identify any clear geographic patterns.

Descriptive Analysis Model Variables

Level	Variable Name	values	Sample (Total N=32540)	Count/Mean	Percentage/SD
Individual/ Property	Estimated race		31539		
		White		22998	72.92%
		Black		4565	14.47%
		Hispanic		2145	6.80%
		Asian		1831	5.81%
	Sex		30272		
		Male		15475	51.12%
		Female		14797	48.88%
	Organizational plaintiff		32540		
		Yes		23785	73.09%
		No		8755	26.91%
	Property taxable value		31631	11029422.35	20622564.73
	Law firm		32540		
		Puckett&Redford		4412	13.56%
		Donaldeallen		2273	6.99%
		Jordan		2158	6.63%
		Mattjmclain		1807	5.55%
		Everettholum		1504	4.62%
		Other		20386	62.65%
	Landuse deed		32540		
		multi-units-family		20849	64.07%
		single family		7516	23.10%
		other_org		2336	7.18%
		mobile home		1526	4.69%
		unknown		313	96.19%
	Sale record-same year		32540		
		Yes		4239	13.03%
		No		28301	86.97%
	Sale record-next year		32540		
		Yes		3699	11.37%
		No		28841	88.63%
	Sale record-last year		32540		
		Yes		2590	7.96%
		No		29950	92.04%
	Permit approved-same year		32540		
		Yes		1134	3.48%
		No		31406	96.52%
	Permit approved-next year		32540		

		Yes		1020	3.13%
		No		31520	96.87%
	LIHTC property		32540		
		Yes		1456	4.47%
		No		31084	95.53%
	Public housing		32540		
		Yes		279	0.86%
		No		32261	99.14%
	Answer file present		32540		
		Yes		9653	29.67%
		No		22887	70.33%
Community	median household income		31540	50676.78	17865.60
	median property value		31521	220112.84	77806.79
	% of non-citizen		31540	6.49	6.26
	% of foreign born		31540	12.04	8.48
	% of White		31540	59.28	24.42
	% of Black		31540	9.23	6.94
	% of Native		31540	1.31	1.59
	% of Hispanic		31540	12.71	10.44
	% of Asian		31540	6.55	5.47
	% of PI		31540	1.85	2.80
	% of Below poverty rate		31540	16.41	11.80
	% of High school and more		31540	73.89	25.95
	% of Bachelor and more		31540	17.78	11.34
	% of Unemployed		31540	8.84	5.60
	% of Rental properties		31540	44.20	24.26
	% of Vacant		31540	7.72	5.12
	% of Rental burden		31540	21.77	10.77
	# of Sales		32540	112.92	88.97
	# of Permits		32540	45.76	98.40
	Community SES		31521		
		Lower		5948	18.87%
		Middle		19404	61.56%
		Upper		6169	19.57%
County	GDP		32540	37595962.21	3969431.95
	Unemployment Rate		32540	7.01	1.57
	Fair housing price 2b		32540	1060.75	97.07
	House price index		32540	162.00	31.99

Table 2 Descriptive analysis for variables using in the model. Results with % are presented as count and percentage of distribution, otherwise are present mean and standard deviation.

Before proceeding with the models, the variables presented in Table 2 were used for analysis. This study focused on data from 2010 to 2019 due to changes in census tracts every 10 years, which could affect the consistency of community-level variables. Additionally, community data prior to 2009 were not collected by the ACS and 2020 was excluded because of its unique circumstances related to the pandemic. This study found that 73.92% of the estimated racial composition was White, compared to the county proportion of 74.3%, according to the census. Meanwhile, 14.47% of residents were estimated to be Black, whereas the total population percentage of Black individuals was only 7.7%. Additionally, 73.09% of the total cases were filed by organizational plaintiffs, such as companies and trusts. However, due to missing data on the number of rental units owned by these companies and trusts in Pierce County, estimating their relative impact is challenging. In other words, it is difficult to determine whether the fact that organizational plaintiffs filed 73.09% of the cases represents a disproportionately high number or not. Several dominant law firms for plaintiffs emerged, with the top five law firms representing more than 37% of the eviction filings in Pierce County. Furthermore, 13% of the cases matched sales records for the property in the same year. Eviction filings from LIHTC properties accounted for 4.47% of the cases, and 0.86% of cases were linked to public housing managed by either Pierce County or the City of Tacoma Housing Authority.

Exploratory Analysis

Models for RQ2

Regarding RQ2, this study sought to determine what factors at individual, community, and macro levels influenced whether a case was dismissed or judged by the court. A dismissal did not necessarily mean the defendant won the case; it signified the case was resolved before judgment, potentially with the defendant moving out. Dismissed cases imply that the defendant had options at that stage, such as relocating to a family member's home. Although not common, in some states, a dismissed case can be requested to be sealed, preventing public access to the eviction record. To explore influencing factors at various levels, the results reflect three logistic models at the individual, community, and county levels; one logistic model combining all three levels; and a multilevel logistic regression. The models at individual levels and macro level did not pass the Hosmer–Lemeshow test, indicating these models did not fit well; however, the combined model did pass, suggesting a better model fit. Moreover, the VIF was tested after each model (see Appendix Table A1). If the VIF exceeded 5 in Models 1 (i.e., individual), 2 (i.e., community), and 4 (i.e., combined), then the variable was eliminated. In Model 3 (i.e., county), the VIF tolerance was 10 due to fewer variables. The multilevel logistic regression did not feature a decent intraclass correlation coefficient ($< .05$); thus, the multilevel logistic regression (random-fixed model) results are not reported (LeBreton & Senter, 2008). The top five features of importance from the random forest model are discussed.

In Model 1 (see Table 3), which included only individual-level variables such as predicted race and ethnicity, plaintiff entity, and property sales history, property characteristics emerged as significant variables associated with the decision to dismiss or proceed to judgment. Cases involving an organization as the plaintiff, like a property management company, were more likely to be dismissed compared to those with private landlords (adjusted odds ratio [AOR] = 1.13, $p < .01$). If the property had sales records in the same year as the eviction filing, the case was more likely to be dismissed compared to those without sales records in the same year (AOR = 1.38, $p < .01$). Moreover, each unit increase in the taxable value of the property increased the estimated odds of dismissal by 18% (AOR = 1.18, $p < .01$). Compared to single-family homes, the estimated odds of dismissal decreased by 18% for mobile homes (AOR = 0.82, $p < .01$) and 30% for multiunit properties (AOR = 0.70, $p < .01$). Plaintiff representation by specific law firms also significantly affected the likelihood of dismissal; representation by Matt J. Mclain (AOR = 1.36, $p < .01$) or Puckett & Redford (AOR = 1.91, $p < .01$) increased the estimated odds of dismissal, whereas representation by Jordan reduced the estimated odds of dismissal (AOR = 0.57, $p < .01$) while the reference group is other small law firms. The Black racial group was the only group with lower odds of dismissal compared to the White group, although only approaching marginal statistical significance ($p = 0.15$). The presence of answer files in a case statistically decreased the estimated odds of dismissal (AOR = 0.94, $p < .05$). This might indicate that once they answered the summons, they sought a hearing and the case was less likely to be dismissed.

When testing the community variables against the outcomes of dismissal and judgment, each unit increase in the median property value in a census tract increased the estimated odds of a case being dismissed by 8% (AOR=1.08, $p < .01$). The percentages of Black and Hispanic residents in the census tract decreased the estimated odds of dismissal (AOR=0.94, $p < .05$; AOR=0.97, $p = .14$). Furthermore, each unit increase in the poverty rate decreased the estimated odds of dismissal (AOR = 0.81, $p < .01$). Regarding the rental situation in a community, each unit increase in the percentage of rental units in the census tract increased the estimated odds of dismissal by 25% (AOR = 1.25, $p < .01$), whereas each unit increase in the percentage of rent-burdened households decreased the estimated odds of dismissal by 5% (AOR = 0.95, $p < .05$).

For the model based solely on county-level data across years, factors such as the House Price Index and fair market rent for 2-bedroom units (40th percentile of gross rent) were considered. An increase of one unit in the fair market rent for 2-bedroom units decreased the estimated odds of a case being dismissed by 14% compared to moving to judgment, possibly due to the higher amount of money owed by tenants. Conversely, an increase in the House Price Index tended to make cases more likely to be dismissed (AOR = 1.13, $p < .01$). This could be because property owners might want to sell their properties at higher prices, using eviction as a tool to vacate the premises quickly. An interaction test between fair housing price and the House

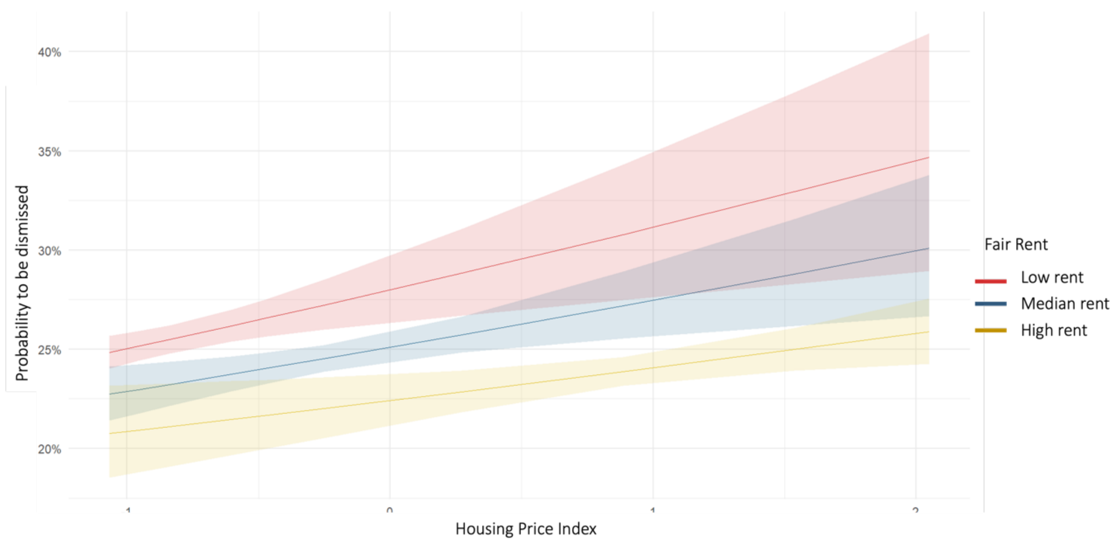


Figure 16 The interaction between Housing Price Index and fair housing price

Price Index was conducted to investigate their relationship. Overall, high housing prices coupled with low rents were associated with higher odds of dismissal. In other words, more affordable rent prices were associated with a higher likelihood of cases being dismissed, and this advantage became more pronounced when housing prices were high.

However, when employing the Hosmer–Lemeshow goodness-of-fit test, the chi-square value was relatively large and the p-value was less than .05, indicating a poor fit. Therefore, I decided to combine all three levels (individual, community, and county) and rerun the logistic regression. Most independent variables at the individual level in the combined model exhibited similar significance and relationships as in the model that included only individual-level variables. However, Hispanic ethnicity became statistically significant in the combined model. Specifically, Hispanic individuals (AOR = 1.14, $p < .05$) were more likely to have their case dismissed compared to White individuals.

When comparing the model with only community-level variables to the combined model, many variables became nonsignificant, such as presence of an answer from the defendant, percentage of Asian population, and percentage of rental burden. Independent variables often become nonsignificant upon the addition of new variables due to reasons like mediation. In this study, personal information like race and ethnicity might already encapsulate some community information, potentially related to historical segregation or other associations. Nonetheless, some variables remained significant, such as the percentage of the Black population, poverty rate, and percentage of rental properties in the census tract. Each unit increase in the percentage of the Black population in the tract decreased the estimated odds of being dismissed by 6% (AOR = 0.94, $p < .01$). The percentages of the Black population and Pacific Islander population were the only two statistically significant community characteristics regarding race and ethnicity in the combined model compared to the community-only model.

Regarding the rental situation, each unit increase in the percentage of rental properties increased the estimated odds of dismissal by 13% (AOR = 1.13, $p < .01$). More rental properties in the same tract may help renters find another place to move before judgment without changing much, such as commute distance or school district, potentially leading to a quicker resolution. County-level variables echoed the findings from the county-level only model. Interacting personal race and ethnicity against the percentage of the White population in the census tract revealed that the interaction between individual Asian race and the percentage of White residents in the tract was statistically significant. Overall, a higher percentage of White residents in the census tract decreased the estimated odds of dismissal, and these odds varied by individual race and ethnicity (see Figure 17). For instance, an Asian individual in a tract with a lower percentage of White residents had higher estimated odds of dismissal compared to an Asian individual in a tract with a higher percentage of White residents.

Variable	Value	Model 1-individual		Model 2-community		Model 3- county		Model 4- combined	
		AOR	95CI	AOR	95CI	AOR	95CI	AOR	95CI
(Intercept)		0.30*	[0.28,0.33]	0.29*	[0.29,0.3]	0.34**	[0.32,0.35]	0.28**	[0.26,0.31]
Estimated race/ethnicity (White)									
	Asian	1.13*	[1.01,1.27]					1.12	[0.99,1.26]
	Black	0.94	[0.87,1.02]					1.00	[0.91,1.09]
	Hispanic	1.08	[0.98,1.20]					1.14*	[1.02,1.27]
Estimated Sex(Male)									
	Female	0.98	[0.92,1.03]					0.98	[0.93,1.04]
	Unknown	0.95	[0.85,1.07]					0.97	[0.86,1.08]
Organizational plaintiff(No)		1.13**	[1.06,1.21]					1.14**	[1.06,1.22]
Sale record-same year(No)		1.38**	[1.27,1.49]					1.36**	[1.26,1.48]
Sale record-next year(No)		1.05	[0.97,1.14]					1.05	[0.97,1.14]
Sale record-last year(No)		1.06	[0.96,1.16]					1.07	[0.97,1.18]
Permit approved-same year(No)		0.94	[0.81,1.09]					0.97	[0.83,1.12]
Permit approved-next year(No)		1.10	[0.95,1.26]					1.10	[0.93,1.24]
Permit approved-last year(No)		1.06	[0.91,1.23]					1.09	[0.95,1.29]
Property taxable value		1.18**	[1.14,1.21]					1.15**	[1.11,1.18]
Landuse deed(Single family)									
	mobile home	0.82**	[0.71,0.94]					0.82*	[0.71,0.94]
	multi-units-family	0.70**	[0.65,0.75]					0.74**	[0.68,0.80]
	other_org	1.40**	[1.23,1.59]					1.28**	[1.1,1.45]
	unknown	1.20	[0.92,1.55]					1.16	[0.89,1.51]
Law firm (Other)									
	Donaldeallen	1.09	[0.98,1.22]					1.05	[0.94,1.17]
	Everettholum	0.94	[0.82,1.08]					0.86*	[0.75,0.99]
	Jordan	0.57**	[0.50,0.65]					0.57**	[0.5,0.65]
	Mattjmclain	1.36**	[1.21,1.53]					1.29**	[1.15,1.45]

	Puckett&Redford	1.91**	[1.77,2.07]					1.80**	[1.67,1.96]
answer (No)		0.94*	[0.88,1.00]					0.95	[0.90,1.01]
LIHTC property (No)		1.01	[0.89,1.15]					0.97	[0.85,1.11]
Public housing (No)		1.18	[0.87,1.60]					1.16	[0.86,1.58]
median household income				-	-			-	-
median property value				1.08**	[1.03,1.12]			1.05*	[1.01,1.1]
# of Sales				0.99	[0.95,1.03]			0.99	[0.95,1.04]
# of Permits				1.04*	[1.01,1.08]			1.04	[1.00,1.08]
% of non-citizen				-	-			-	-
% of foreign born				-	-			-	-
% of White				1.02	[0.98,1.07]			1.02	[0.98,1.08]
% of Black				0.94*	[0.9,0.98]			0.94*	[0.9,0.98]
% of Native				0.99	[0.96,1.03]			0.99	[0.96,1.02]
% of Hispanic				0.97	[0.93,1.01]			0.97	[0.93,1.01]
% of Asian				1.07**	[1.03,1.1]			1.02	[0.98,1.05]
% of PI				0.91**	[0.88,0.95]			0.96*	[0.93,1.00]
% of Below poverty rate				0.81**	[0.77,0.86]			-	-
% of High school and more				-	-			-	-
% of Bachelor and more				0.97	[0.92,1.02]			1.01	[0.96,1.06]
% of Unemployed				0.99	[0.95,1.04]			0.92**	[0.88,0.96]
% of Rental properties				1.25**	[1.18,1.32]			1.13**	[1.06,1.19]
% of Vacant				0.99	[0.96,1.03]			0.98	[0.95,1.02]
% of Rental burden				0.95*	[0.91,0.99]			0.96	[0.92,1.00]
Community SES(low)									
	Middle			-	-			-	-
	High			-	-			-	-
GDP						-	-	-	-
Unemployment Rate						-	-	-	-
Fair housing price 2b						0.90**	[0.80,0.92]	0.92**	[0.89,0.95]
House price index						1.13**	[1.05,1.22]	-	-

Fair housing price 2b *House price index					0.98**	[0.94,1.00]	-	-
Estimated race/ethnicity (Asian)*% of White							0.89*	[0.8,0.99]
Estimated race/ethnicity (Black)*% of White							0.95	[0.87,1.04]
Estimated race/ethnicity (Hispanic)*% of White							0.97	[0.87,1.09]
Hosmer and Lemeshow goodness of fit test	Fail		Pass		Fail		Pass	

Table 3 Logistic Regression for outcome of dismissal or judgement at different factor level(N=31,361). * $p<0.05$ ** $p<0.01$
Notes: 1) The reference group is in the parentheses () 2) All the continuous variables are standardized. 3) The "-" symptom means the variable is not included in the model due to multicollinearity.

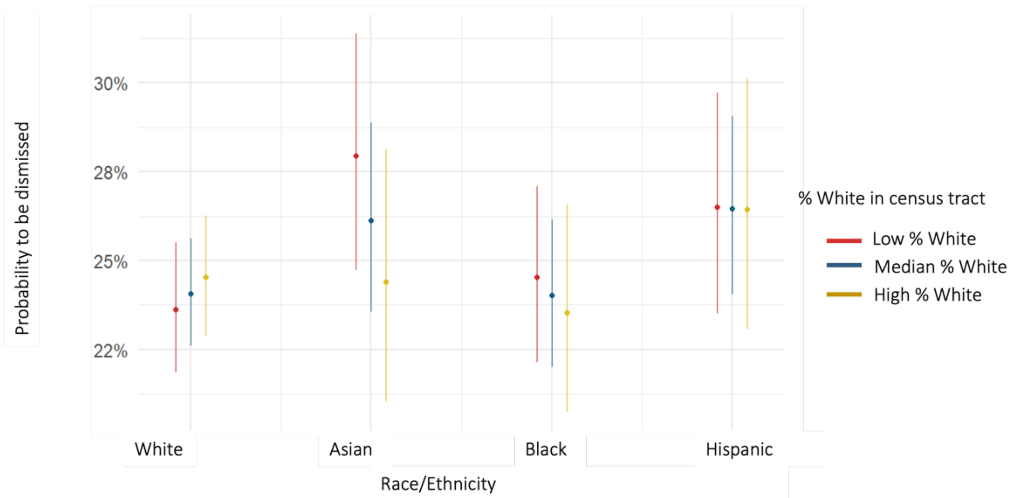


Figure 17. Estimated odds of dismissal with different proportions of White residents by race and ethnicity

The multilevel models, which treated each census tract as a random effect, showed very low ICCs, indicating that multilevel modeling was not necessary for this analysis. After considering the Hosmer–Lemeshow goodness-of-fit test, it became apparent that the distinction between dismissal and judgment was complex and data from a single level could not provide comprehensive insights into their relationship. Integrating variables from various levels was crucial to understanding their associations with dismissal and judgment. Additionally, community-level differences did not contribute significantly to the outcome variations, because the outcomes were diverse and community factors only explained a small portion of these variations.

Compared to logistic regression, the random forest model can capture nonlinear relationships between features and the target variable and naturally model interactions between features. Random forest also provides straightforward metrics for feature importance and predictor affects based on partial dependence plots (PDPs). Feature importance from random forest modeling represents the average contribution across all trees in the forest. I used random forest modeling to identify the top five features (independent variables) for dismissal and judgment classification. The top five variables ranked by the model were the taxable value of the property, presence of an answer file in the case, median property value, individual estimated race being White, and estimated percentage of rental burden. This indicates that the random forest model estimated that individual-level factors were more critical in classifying the outcome of dismissal or judgment, such as property value and how the defendant responded to the eviction filing. Taking the presence of an answer file in the case as an example, the PDP indicated the probability of being dismissed was 0.80 less when there was an answer file.

Models for RQ3

RQ3 centered on factors at individual, community, and macro levels that influence whether a case leads to eviction by the sheriff after judgment. Specifically, this involved differentiating between cases handled by the sheriff, including those categorized as ousted and ejected or peaceful possession, and those in which the tenant moved before the sheriff's intervention, including cases labeled as expired, return per attorney, and return per plaintiff. The distinction here is that in the latter scenarios, the tenant moved before facing sheriff enforcement, thereby avoiding direct confrontation, whereas in the former scenarios, the tenant was forcibly removed by the sheriff, which can result in immediate homelessness and limited options for housing. Although ousted and ejected or peaceful possession represent two different levels of severity of sheriff intervention, I combined them because they both involve moving at the last minute, likely making their housing status more vulnerable compared those who moved before sheriff involvement. Furthermore, evictions executed by the sheriff can escalate to violent situations, which are not rare nationally and even cause death. An example of such outcome occurred on January 24, 2024, when two tenants died by gunfire while an Alameda County sheriff's deputy was serving an eviction notice. The sheriff's deputy also was shot and injured.

According to the individual-level model (see Table 4), the estimated odds of sheriff eviction increased by 11% for Black individuals compared to White individuals (AOR = 1.11, $p < .01$). Property sales also influenced the likelihood of execution by the sheriff. If there was a sale history in the same year when the eviction was filed, the estimated odds of sheriff eviction increased by 25% compared to cases with no sale history in the same year (AOR = 1.25, $p < .01$). Regarding legal representation of the plaintiff, if the case was handled by the Donald E. Allen (AOR = 0.67, $p < .01$) or Everett Holum (AOR = 0.80, $p < .01$) law firms, the estimated odds of sheriff eviction decreased by 33% and 20%, respectively, compared to other law firms. Moreover, if the case was labeled as involving LIHTC property, or a property that received tax credits in exchange for offering a certain fraction of rent-restricted units to lower-income tenants, the estimated odds of enforcement by the sheriff increased by 38% compared to non-LIHTC properties (AOR = 1.38, $p < .01$). The Hosmer–Lemeshow test for Model 5 (Sheriff Eviction model at individual level) indicated marginal goodness of fit, with a p-value of .10. The VIF results are in Appendix Table A2.

Regarding the community-level model, the estimated odds of sheriff eviction increased by 6% for each unit increase in median property value (AOR = 1.06, $p < .05$). Similarly, each additional sale in the census tract increased the estimated odds of sheriff eviction (AOR = 1.10, $p < .01$). Notably, percentage of White residents in the census tract decreased the estimated odds of sheriff eviction (AOR = 0.90, $p < .01$). A higher unemployment rate decreased the estimated odds of being evicted by the sheriff (AOR = 0.78, $p < .01$). A higher percentage of rent-burdened households in the census tract increased the estimated odds of sheriff eviction (AOR = 1.07, $p < .01$). However, the Hosmer–Lemeshow goodness-of-fit test yielded a p-value of .01, which is below the .05 threshold, indicating that the community-level model did not sufficiently explain the classification between cases handled or not handled by the sheriff.

When moving the focus from census tract to county-level factors, with an increase in housing prices, the estimated odds of sheriff eviction increased (AOR = 1.32, $p < .01$). Nevertheless, the county-level model did not pass the Hosmer–Lemeshow goodness-of-fit test, because the p-value was less than .01.

Variable	Value	Model 5-individual		Model 6-community		Model 7-county		Model 8-combined	
		AOR	95%CI	AOR	95%CI	AOR	95%CI	AOR	95%CI
(Intercept)		0.31**	[0.28,0.34]	0.31**	[0.31,0.33]	0.34**	[0.32,0.35]	0.32**	[0.29,0.35]
Estimated race/ethnicity (White)									
	Asian	0.94	[0.81,1.08]					0.91	[0.79,1.06]
	Black	1.11*	[1.02,1.21]					1.04	[0.94,1.15]
	Hispanic	0.89	[0.78,1.01]					0.84*	[0.73,0.96]
Estimated Sex(Male)									
	Female	0.99	[0.93,1.06]					0.99	[0.92,1.05]
	Unknown	0.92	[0.80,1.04]					0.91	[0.80,1.04]
Organizational plaintiff (No)		0.98	[0.91,1.06]					0.96	[0.89,1.03]
Sale record-same year (No)		1.25**	[1.13,1.37]					1.26**	[1.15,1.39]
Sale record-next year (No)		1.09	[0.99,1.20]					1.10	[1.00,1.22]
Sale record-last year (No)		1.02	[0.91,1.14]					0.99	[0.88,1.1]
Permit approved-same year (No)		0.87	[0.73,1.04]					0.91	[0.76,1.09]
Permit approved-next year (No)		1.04	[0.88,1.24]					1.13	[0.95,1.35]
Permit approved-last year (No)		1.10	[0.92,1.31]					1.16	[0.97,1.39]
Property taxable value		0.96*	[0.93,1.00]					0.95	[0.92,0.99]
Landuse deed (Single family)									
	mobile home	1.14	[0.98,1.32]					1.28**	[1.10,1.50]
	multi-units-family	1.07	[0.98,1.16]					0.96	[0.88,1.06]
	other_org	1.10	[0.93,1.31]					1.07	[0.89,1.27]
	unknown	1.07	[0.78,1.48]					0.99	[0.71,1.38]
Law firm (Other)									
	Donaldeallen	0.67**	[0.59,0.77]					0.84*	[0.74,0.96]
	Everettholum	0.77**	[0.66,0.89]					0.96	[0.82,1.12]
	Jordan	1.02	[0.91,1.15]					1.04	[0.93,1.17]
	Mattjmclain	0.80**	[0.69,0.92]					1.03	[0.89,1.19]
	Puckett&Redford	1.06	[0.96,1.17]					1.04	[0.94,1.15]
answer (No)		0.98	[0.92,1.05]					0.94	[0.88,1.00]
LIHTC property (No)		1.38**	[1.20,1.59]					1.36*	[1.18,1.57]
Public housing (No)		1.14	[0.83,1.58]					1.08	[0.78,1.50]
median household income				-	-			-	-
median property value				1.06*	[1.00,1.11]			-	-

Variable	Value	Model 5-individual	Model 6-community		Model 7-county		Model 8-combined	
# of Sales			1.10**	[1.05,1.15]			1.02	[0.97,1.07]
# of Permits			0.94*	[0.90,0.98]			0.95*	[0.90,0.99]
% of non-citizen			-	-			-	-
% of foreign born			-	-			-	-
% of White			0.90**	[0.86,0.96]			0.89**	[0.84,0.95]
% of Black			1.01	[0.96,1.06]			1.01	[0.96,1.06]
% of Native			0.99	[0.96,1.03]			1.00	[0.96,1.03]
% of Hispanic			1.00	[0.96,1.05]			0.92**	[0.87,0.96]
% of Asian			0.95*	[0.92,0.99]			0.97	[0.94,1.01]
% of PI			1.03	[1.00,1.07]			0.99	[0.95,1.02]
% of Below poverty rate			1.23**	[1.15,1.30]			1.16**	[1.08,1.23]
% of High school and more			-	-			-	-
% of Bachelor and more			1.14**	[1.08,1.21]			1.08*	[1.01,1.14]
% of Unemployed			0.78**	[0.74,0.82]			0.88**	[0.84,0.93]
% of Rental properties			1.06	[1,1.13]			1.06*	[0.99,1.14]
% of Vacant			1.05*	[1.01,1.09]			1.04*	[1.00,1.09]
% of Rental burden			1.07**	[1.03,1.13]			1.08**	[1.03,1.13]
Community SES (lower)								
	Middle		-	-			-	-
	Upper		-	-			-	-
GDP					-	-	-	-
Unemployment Rate					1.08*	[1.01,1.15]	-	-
Fair housing price 2b					1.09	[1.00,1.19]	1.33**	[1.28,1.38]
House price index					1.44**	[1.32,1.59]	-	-
Fair housing price 2b *House price index					0.92**	[0.89,0.96]		
Estimated Race (Asian)*% of White							1.08	[0.92,1.26]
Estimated Race (Black)*% of White							1.00	[0.90,1.11]
(Hispanic)*% of White							0.96	[0.83,1.11]
Hosmer and Lemeshow goodness of fit test		Pass	Fail		Fail		Pass	

Table 4 Logistic Regression for outcome of sheriff eviction at different factor level (N=22,829). * $p<0.05$, ** $p<0.01$

Notes: 1) The reference group is in the parentheses () 2) All the continuous variables are standardized. 3) The "-" symptom means the variable is not included in the model due to multicollinearity.

When I combined all independent variables in one model, most statistically significant variables from the individual-only model remained statistically significant in the combined model. However, if the tenants were Hispanic, the estimated odds of sheriff eviction were less than for White tenants (AOR = 0.84, $p < .05$). Although the outcome for Black people was not statistically significant in the combined model, they were still the only race more likely to be evicted by the sheriff compared to White people. Compared to eviction properties without sales records, properties with sales records in the same year had higher estimated odds of sheriff eviction (AOR = 1.26, $p < .01$). Compared to tenants in single-family houses, tenants in mobile homes were more likely to face enforced eviction by the sheriff (AOR = 1.28, $p < .01$). In the combined model, the influence of law firms decreased. Only the Donald E. Allen law firm remained statistically significant (AOR = 0.84, $p = .01$). The presence of a LIHTC property still positively influenced the estimated odds of sheriff eviction (AOR = 1.36, $p < .01$).

At the community level, fewer independent variables were statistically significantly associated with the dependent variable. Regarding race and ethnicity in the census tract, percentage of White residents (AOR = 0.89, $p < .01$) and percentage of Hispanic residents (AOR = 0.92, $p < .01$) were negatively associated with sheriff eviction. Furthermore, the percentage of rent-burdened households in the census tract was positively associated with sheriff eviction (AOR = 1.08, $p < .01$). Interestingly, each unit increase in the unemployment rate was associated with an 12% decreased in the estimated odds of sheriff eviction (AOR = 0.88, $p < .05$). This is counterintuitive to expectations. A potential explanation is that evictions typically involve low-income families who may have jobs that are hourly or lack job security, making them more vulnerable to housing instability when employed because they do not have availability to handle an eviction, such as attending hearings (Desmond & Gershenson, 2016).

At the county level in the combined model, the relationships between the independent variables and dependent variable were consistent with the county-only model. Interestingly, each unit increase in the fair housing price was associated with an 33% increase in the estimated odds of sheriff eviction (AOR = 1.33, $p < .01$).

After including the interaction between the percentage of White residents in the tract and individual race and ethnicity in the model, the p-value for the Hosmer–Lemeshow goodness-of-fit test increased from .49 to .72. Thus, I retained this interaction even though it was not statistically significant.

Similar to the dismissal model, I also ran a random forest model to identify the top five features (independent variables) that influenced the classification outcome. The top five variables were the taxable value of the property, presence of an answer file in the case, whether the plaintiff was an organization, estimated White race, and estimated median household income (see Appendix Figure A10). Four were individual-level variables associated with whether the case ended in sheriff eviction. The taxable value of the property and whether the plaintiff was an organization indicated the type of property, whereas the presence of an answer file showed the tenant's response to the case. The PDP was used to analyze the relationship of the top two independent variables of taxable property value and answer file with the dependent variable of sheriff eviction (see Appendix Figure A11). According to the PDP, when the property value was lower, the tenant was more likely to be evicted by the sheriff. Answering the summons decreased the probability of being evicted by the sheriff. Interestingly, the presence of an answer file was not statistically significant in the logistic regression model, with 26% of all cases and 27% of cases with answer files ending in sheriff eviction. Purely from these numbers, the influence on the results appears no different. Meanwhile, by analyzing the marginal effects, the Average

Marginal Effects for answer is -0.01 without statistical significance (see Appendix Table A4). However, the random forest model assigned high importance to this variable. The PDP showed that the probability of being evicted by the sheriff decreased by approximately 0.6 if an answer file was present in the case. It is challenging to determine the reason because the random forest consists of hundreds of trees, but the model may have captured some nonlinear interactions or relationships between this variable and others that logistic regression did not.

Summary

This chapter presented the findings from the data analysis. Advanced CSS methods, including document layout analysis and LLM, were used to extract data from images, with document layout analysis achieving high accuracy for this task.

Descriptive analysis indicated that eviction filings in Pierce County have been generally stable at around 3,000 cases annually, dropping to around 500 during the pandemic but surging almost back to prepandemic levels in 2022. A Sankey plot and bar plot revealed that about 30% of cases were dismissed, with most being judged. Approximately 20% to 30% of tenants face eviction executed by the sheriff after judgment. The presence of answer files added approximately 10 days to the eviction process prepandemic and 20 days during the pandemic. Geographically, areas with high concentrations of apartments, particularly in suburban regions, often had more eviction filings. These hotspots had significant overlap with tracts with a high percentage of people of color. This is consistent with the disproportionate number of eviction filings experienced by Black adults found in another bar plot.

Various classification models such as logistic regression and random forest highlighted different factors that influenced whether a case was dismissed, judged, or led to eviction by the sheriff after judgment. Common strong factors at the individual level included being a person of color, having property sale records in the same year as the eviction filing, and representation by certain law firms. Random forest modeling also pointed to the taxable value of the property and whether there is a response to the summons as important features influencing the outcome. At the community level, the percentage of people under the poverty level and the percentage of rent-burdened households were strongly connected to the outcome. Housing type and the percentage of rental properties in a tract were also associated with the outcome of dismissal or judgment. At the county or macro level, the Housing Price Index and fair rent price influenced outcomes in complex ways. Interacting individual race with the proportion of White people in the census tract showed that people of color experienced different outcomes compared to White people in communities with more White people.

Chapter 5: Discussion and Conclusion

Discussion

The discussion section covers both the methodology and eviction research. On the methodological side for RQ1, the role of CSS in this study is examined. Regarding the research findings for RQ2 and RQ3, the discussion focuses on how variables at various levels influenced eviction outcomes, with relevant theories and arguments.

The Role of CSS

Lower Accuracy in GPT. Because this study was based on computational methods, the role of CSS was critical and formed the foundation of this research. Computational methods enable the conversion of information from various formats, such as images and text, for research purposes. Unused information in PDF files became accessible due to these methods. Specifically, in this study, I tested two methods: layout analysis for images and LLMs like the GPT model. The performance of layout analysis was better than that of the LLM in this scenario. After thorough investigation, I identified two issues. First, the GPT model sometimes failed to recognize an address, resulting in extremely low scores. This failure is a significant reason for lower mean scores compared to the layout model. Second, I realized that the lower accuracy in the GPT model might stem from preprocessing rather than the model itself. For example, many summons contain line numbers and the address may span several lines. When converting all images to text line by line, the line numbers might be included in the address. Thus, the original address “123 ABC Street, APT 456, Tacoma, WA, 98407” in a text file might appear as “19 123 ABC Street, APT 456, 20 Tacoma, WA, 98407,” where “19” and “20” are line numbers. These line numbers pollute the address pattern, making it more difficult for the GPT model to recognize the address. Comparatively, the layout method only recognized information in a specific area, resulting in more accurate outcomes compared to the GPT model.

It is common in data science and AI that although an algorithm might be developed and work very well for understanding natural language, it does not yield better performance for specific tasks. The preprocessing of information plays a significant role in the results. The fact that the GPT model did not achieve the expected results does not mean it is not as good as the layout method in this situation. Instead, it shows that computational methods form a pipeline of methods. When one method is upgraded, the related upstream or downstream processes should also be adjusted to meet expectations.

Information Extraction from PDF Files. Based on the analysis of these two methods, it is important to consider in what situation layout analysis or application of GPT for information extraction from PDF documents should be used. If the target information in the PDF files is in similar places—for example, if the defendant’s address consistently appears in the summons, as shown in Figure 6—then layout analysis can yield more accurate results. However, its drawback is that if the target information is not near the expected area, extraction might fail. On the other hand, the GPT method is more effective with unstructured PDF files. That is, it excels when information is embedded in extensive text and the pattern of its location is difficult to identify. For instance, the specified amount of rent money could appear anywhere in a paragraph and be expressed in various ways, such as “$1,000 per month” or “$50 per day.” GPT models can read and comprehend the context to extract the information accordingly. However, GPT models are more computationally intensive and the quality of the output depends heavily on the prompt provided by the user.

Pipeline Increases Geocoding Success Rate. Additionally, there is a challenge in acquiring accurate locations due to the geocoding process. The pipeline I built to integrate with

ESRI and Azure Here increased the success rate in obtaining the correct latitude and longitude. Address information may appear several times at different stages, such as in summonses and sheriff returns. Thus, I used multisource information to cross-check or impute missing information. Previously, Thomas et al. (2024) relied solely on the address information in summonses, but in some cases, the address may not appear in summonses or the shipping address for summonses is not in the designated jurisdictional area. Multisource cross-checking makes the information more accurate and comprehensive. Next, for geocoding, ESRI is usually more professional and has higher accuracy in geocoding, whereas Azure Here or Google Maps have a higher capacity to understand the address string and will provide their best guess even without a perfect address pattern. Thus, the pipeline prioritized accuracy and verified whether the location could be successfully converted into latitude and longitude coordinates at the center of a parcel. In other words, the point should be in the center of the parcel so it can accurately merge with property information. If the point was on the street or at another level, then I used Azure Here to check whether to move the point on the street to the center of the parcel. For example, given the address "123 ABC St APT 45 Tacoma, County of Pierce, WA, 94807," APT 45 might be recognized as 4S due to poor scan quality. Thus, ESRI may place the point on the street, whereas Azure may return a better guess. The best guess may not guarantee accuracy, but it provides the best estimate of which property involved the eviction. Although the quality of the addresses that Eviction Lab has received is unclear, its researchers have stated that regular expression (e.g., manually changing "st" to street") plus ESRI converted 93.4% of addresses into point or street-level addresses. In this study, if I only used summons data without multisource checking, I would have determined point-level locations for 92.1% of the cases, which is close to Eviction Lab's geocoding results. After adding Azure Maps for cross-checking, however, 96.97% of the cases matched point-level locations.

CSS for Feature Generation. Another challenge previously mentioned is data generation, often referred to as feature engineering in data science. Unlike the traditional approach in quantitative methods that links to external datasets—which although accurate, may have limited access to those datasets—computational methods enable the creation of new variables based on current information. For instance, previous eviction research used computational methods to infer race and gender variables based on plaintiff names. In this study, I also employed advanced NLP models to classify whether the plaintiff was an organization, such as an LLC or trust, or a private landlord. Traditional methods might rely on keyword searches, but such approaches can overlook some unique names. Advanced NLP models not only save time by eliminating the need to list keywords but also yield higher accuracy because the results are voted on by different language models in the case of conflicts. As computational methods develop, features embedded in the raw data can be uncovered and transformed into usable features for modeling.

Data Linkage. Data linkage is a crucial component of analysis, involving integration of datasets from multiple sources. In CSS, this enables a more comprehensive exploration of social issues. The linkage process is not solely based on string matching; it often involves spatial linkage, presenting a dilemma between excluding data that cannot be directly linked or employing fuzzy methods for best-guess approximations. For instance, when an apartment address is imprecise, leading to a geocoded point that doesn't precisely match a property, should this address be excluded or associated with the nearest property information? Similarly, when an open dataset provides only the most recent parcel numbers, making it challenging to link longitudinal sales records due to parcel splits or merges, the question arises whether to omit this information or infer it using surrounding property data for tax value estimation. These numerous minor decisions in the data linkage process may significantly affect the modeling outcomes, and it is difficult to establish a one-size-fits-all rule for exclusion or imputation. The current best practice is to maintain transparency in the linkage methodology, detailing the decisions made and their rationale. Although CSS opens new avenues for inquiry, it also introduces challenges in reporting. Ensuring the replicability of CSS-based research is a critical step toward leveraging CSS to advance scientific knowledge.

Eviction Outcomes and Tenant Response Impact

Judgment Rate. After extensive data processing, merging, and imputation for this large dataset, case information from court files was made available and linked with external data for analysis. According to a Sankey diagram, 76.6% of eviction filings ultimately resulted in judgment, and 93.4% of these judgments led to a writ of restitution and sheriff eviction. Comparatively, data from Franklin County, Ohio, which includes Columbus, showed that the average judgment rate from eviction filings was 72.4%, with 78.6% of those judgments progressing to sheriff involvement from 2011–2017 (Pierce, 2020). The judgment rates between these two counties, despite their geographical differences, are remarkably similar. Furthermore, more than 42.8% of filed cases resulted in default, meaning the defendant automatically lost the case due to failure to respond to the summons or attend the hearing. When considering all civil cases in Pierce County in 2018, of the 13,564 civil cases filed, 3,668 resulted in default judgment, equating to 27.1%. This indicates that the default judgment rate in eviction cases was higher compared to the rate across all civil cases.

The Role of Answer Files. Furthermore, findings indicated that the answer file variable held significant importance in the random forest model but was not statistically significant in logistic regression. This observation can be attributed to several reasons. First, logistic regression is a linear model assuming a linear relationship between independent variables and the dependent variable, whereas random forest is a nonlinear model capable of capturing more complex relationships. Thus, it is possible that the answer file variable interacted with the dependent variable in a nonlinear manner. Second, logistic regression might not discover interactions between variables if they are not explicitly specified by the user, but random forest can naturally capture interactions between independent variables. Therefore, the answer file variable might have interacted with other variables in complex ways, affecting the dependent variable.

Given these observations from the eviction timeline and descriptive analysis, responding to (or answering) the summons is a crucial step in eviction cases. If the defendant does not respond, a default judgment occurs and the defendant automatically loses the case. Furthermore, descriptive analysis indicated that even if a defendant is eventually evicted by the sheriff, merely filing an answer extends the eviction process, providing more time for the defendant to respond. The rate of answers has increased since 2018, from 30% to more than 50% in 2022. One argument for this increase is that more accessible information and knowledge about eviction procedures during the pandemic have helped renters understand their rights and how to handle eviction notices. For instance, Golio et al. (2022) evaluated the impact of distributing knowledge on eviction processes and found that renters who received this information were 13% less likely to receive a default judgment compared to those who did not receive the information in New Orleans, Louisiana. Although the dismissal rate in Pierce County hasn’t significantly decreased, the increased rate of answers has distinctly extended the overall processing time. Specifically, the Pierce County Eviction Prevention program provides renters with guidance on facing eviction and accessing rental and legal assistance. These programs emphasize the importance of responding to the summons or complaint, because the eviction process with default judgment is fast and offers renters only a very short window to react. Even though responding to the summons may not change the eviction outcome, it may alter the manner of the eviction. Having additional time to react after filing an answer may also increase the chance of accessing legal aid or government housing programs.

Impact of Variables at Different Levels on Eviction Outcomes

Although this section discusses the impacts of various factors at the individual, community, and county levels on case dismissal or judgment, along with outcomes following a writ, in relation to RQ2 and RQ3, it is important to acknowledge that these impacts are not solely based on factors at a single level. Single-level analyses did not pass the goodness-of-fit test, showing the complexity of eviction cases.

Individual Impact. The influence of individual characteristics predominantly reflected the SES of tenants. The findings indicated that several statistically significant variables, including sales records, tax value, and property type, were indicative of SES. Higher tax values, single-family housing, and sales records in the same year were associated with case dismissal. Tenants in mobile homes, typically a cheaper housing option, faced a higher likelihood of enforced eviction compared to those in single-family houses. Sales records in the same year also increased the estimated odds of eviction by the sheriff. Moreover, the property tax value was another significant factor in distinguishing among dismissal, judgment, and sheriff eviction.

Desmond (2012) investigated the eviction situation in Milwaukee, concluding that eviction is a key factor leading to high levels of residential mobility and the reproduction of urban poverty. This study revealed varying patterns of residential mobility following eviction filings, a subject previously underexplored. Typically, tenants in properties with higher tax values and those residing in single-family homes are more likely to move before the final sheriff eviction, resulting in records appearing as dismissed or with no action from the sheriff. The rent for single-family homes usually exceeds that of multiunit family homes and mobile homes, indicating that tenants in these properties pay higher rent before filing occurs. Upon filing, they likely have the option to reduce their housing budget and move to a more affordable location. This suggests their financial capacity is not as constrained as that of other renters facing eviction. However, renters in mobile homes, which are generally less expensive, likely have limited housing budget flexibility even before eviction filings. Thus, they often have no choice but to remain until they face the sheriff. Although eviction drives significant residential mobility and involuntary moves, the varied SES of renters mean that how, when, and where they move differ, leading to diverse life courses. Research has focused on eviction filings and impacts after eviction but seldom explored the eviction process. Specifically, understanding what occurs after eviction filings and how different decisions regarding moving influence outcomes can enhance our comprehension of the relationship between eviction and residential mobility.

Another highly influential factor affecting case dismissal, judgment, and sheriff eviction outcomes is the presence of sales records in the same year as the eviction. Ramiller (2022) analyzed eviction filings and property data in Seattle, Washington, and concluded that eviction filings are more likely at properties sold in the same year. This study found that 13% of eviction filings coincided with a sale in the same year and 19% of eviction filings occurred a year before or after the sale date.

This study examined the relationship between sales and eviction after eviction filings. I found that property sales in the same year increased the probability of both dismissal and sheriff eviction. Although these outcomes may seem contradictory, they reveal the intent behind such evictions. Whether a landlord intends to sell or a new landlord has purchased the property, their goal is to vacate the property efficiently and at a low cost. Eviction filings serve as a tool for this purpose. If the tenants decide to move, the case is quickly dismissed. If the tenants do not move, the process progresses to the final step, involving the sheriff. In Washington, the law requires a 20-day notice for month-to-month tenants to vacate. This could explain why the median number of days between the sale date and eviction filing date is 17, indicating that evictions are filed approximately 20 days after the property is sold.

Community Impact.

Rental Supply and Dismissal Rate. At the tract level, in the dismissal and judgment model, a higher percentage of rental properties at the community level was associated with an increased the probability of case dismissal. As previously discussed, a dismissal case does not necessarily mean that the defendant won. More often, it indicates that the defendant moved out before judgment. Various factors influence housing choice, including workplace proximity, rental budget, school district, and transportation access. A higher availability of rental options in the same census tract naturally offers renters more alternatives, especially those seeking lower rent to minimize changes in other aspects of their lives, such as maintaining proximity to their workplace. Therefore, the supply of rental properties provides renters with more choices and reduces the risk of last-minute eviction.

Gentrification and Outcomes. Many studies on eviction have attempted to explain eviction filings through the lens of gentrification. This study also found evidence supporting the notion that overall, more sales and permit applications correlated positively with eviction filings. Pearson correlations were .89 between sales and permit applications, .50 between sales and eviction filings, and .43 between permit applications and eviction filings. However, investigating the outcomes after filings revealed that although specific properties' sales records and permit application histories were statistically significantly associated with the outcome, the count of sales and permit applications at the census tract level did not exhibit as strong a link. Only the count of permit applications at the census tract level was associated with a decreased rate of being evicted by the sheriff. N. Smith (1979) argued that landlords typically pursue profit, adopting strategies like cutting back on maintenance to reduce building costs and seeking higher returns elsewhere. One hypothesis is that if renters live in communities with more permit applications, the community might be wealthier or have a higher homeownership rate, offering tenants more options to move before being evicted by the sheriff. This aligns with findings that higher median household income was linked with decreased estimated odds of eviction by the sheriff. Another hypothesis suggests that gentrification is a process that may start in smaller areas first, making a census tract too large to observe forced moveouts due to planned demolition or remodeling activities. Hence, the count of permit applications at the tract level would not be positively linked to evictions by the sheriff. Ramiller (2022) focused on permit applications in a 500-meter radius, which might more accurately reflect changes in smaller areas.

Furthermore, the higher the percentage of rent burden, indicating more people allocating 30% or more of their income to rent, the higher the likelihood of sheriff evictions, suggesting tenants are more likely to stay until forcibly evicted by the sheriff. Because rent increases are a common result of gentrification, rent burden also limits housing options after eviction filings. For instance, affected individuals may lack sufficient funds for a security deposit, even if there are more affordable options.

Race and Ethnicity Alignment in Census Tracts. Hispanic ethnicity reached statistical significance in both logistic regression models. Moreover, in the dismissal–judgment model, a higher percentage of the Black population in a census tract correlated with a decreased dismissal rate. Thus, race and ethnicity at both the individual and community levels influenced the outcomes of eviction filings.

For instance, in findings regarding interactions in the dismissal–judgment model, a higher percentage of White residents in a census tract decreased the estimated odds of dismissal, and these odds varied by individual race and ethnicity. Taking Asian race, a statistically significant variable, as example: Asians in tracts with a lower percentage of White residents had higher estimated odds of dismissal compared to those in tracts with a higher percentage of White residents. This reveals an intriguing phenomenon of how individual race and ethnicity—whether aligning with or differing from the majority in a census tract—affects outcomes of eviction filings. Previous studies have shown that eviction filing rates in communities of color are significantly higher than in majority-White communities. However, given the challenges in collecting data on individual race and ethnicity, few studies have explored whether eviction risk changes for racial and ethnic minorities living in majority-White communities. This phenomenon remains complex, potentially linked to factors such as social support, cultural connections, and discrimination. For instance, Rosen et al. (2021) provided insights into how landlords select tenants and how this selection process can perpetuate racial discrimination in housing. Similarly, Decker (2021) examined differing approaches between large-scale and small-scale owners regarding eviction filings and their impact on fair housing practices.

Macro or County Impact. At the macro level, factors such as housing market prices, unemployment rates, and housing policies significantly influence the rental market. However, the interactions among these macro-level factors are highly complex, requiring substantial evidence for sound decision-making rather than relying solely on the associations of one or a few variables, because relationships often can be counterintuitive. Additionally, the influence of macro-level factors on eviction filings and outcomes can differ. Consider the following example from this study.

When county unemployment rates were higher, the likelihood of being evicted by the sheriff, as indicated in Model 7, increased, likely due to decreased income. Conversely, Model 8 revealed a negative relationship between community unemployment rates and the probability of eviction by the sheriff. Although county-level unemployment may reflect a general trend, the impact of unemployment can be uneven, potentially affecting specific groups more severely unless it is a systemic downturn. For instance, recent layoffs have predominantly affected higher-income employees, which may have a lesser impact on eviction due to previous higher income. Desmond (2012) further suggested that renters at high risk of eviction often have jobs with little to no flexibility, leaving them unable to manage their case (e.g., attending hearings) or address housing instability. Thus, unemployment makes it possible for these people to fight an eviction.

Thus, when using macro-level data, such as housing trends and policies, it is crucial to carefully examine how these factors might affect the research population in specific ways, rather than making assumptions based on general trends in housing research.

Implications

CSS for Eviction Research

As previously mentioned, eviction data analysis is nearing its limit with structured data, but unstructured data remains largely inaccessible. There are two ways to deepen the research: One is by efficiently and accurately extracting new data, and the other is waiting for the court system to upgrade its infrastructure and improve its workflow, but this could be very time consuming. Thus, the cost-effective approach now is to make use of current documents for information extraction, converting unstructured data to structured data before a digitalized system is established. This study presents a potentially promising solution for the current stage. The approach used here could simplify the process for researchers exploring eviction issues if the state or county does not provide an accessible structured dataset but instead offers raw photocopied court records. This streamlined solution could be adapted for other states or regions with minor modifications if necessary. Revisiting the Eviction Lab's weekly monitoring data, the West Coast is nearly absent because its data formats are not conducive weekly updates and analysis, because most data received are unstructured. The method used here based on CSS could ensure national datasets are more representative and comprehensive.

Even if some states develop datasets, they may miss key components, such as addresses or amounts owed. This method could ensure the datasets more complete. Additionally, court documents are similar to overall civil filings, so researchers interested in other topics could also use this process to extract information and create a structured database for academic research. For instance, if researchers wish to examine the impact of medical debt using data in the court system, the pipeline I developed can be used to extract names, addresses, processing times, and most importantly, the amounts owed from court files. This would aid in understanding the costs and burdens of medical debt.

Another implication involves the open administrative data movement. This study highlights the strengths and shortcomings of local open data platforms. Eviction data alone are insufficient to support research, but linking them to other datasets opens many possibilities. Local open data platforms provide a robust foundation for research beyond census information. For instance, property data hosted by counties and available through open property data platforms offer opportunities to access property-level information relevant to eviction research. Pierce County serves as a good example. Data accessibility in the United States varies across states and counties, with varying levels of openness. For example, some counties may provide aggregate data, whereas others lack a platform to host any data. This study illustrates how local open data platforms can enhance research focused on local issues. The following observations and suggestions relate to using open administrative data.

1) Data Accessibility: The formats of data vary; some are machine-readable, such as comma-separated valued files, whereas others are not, like PDF files. I spent a great deal of time in this study making the data machine-readable. If data were prepared in a machine-readable format, they would be easier for the public to use and access, ultimately contributing to potential research and evaluation.
2) Data Understanding: It is challenging to find introductions to datasets and detailed explanations of variables, such as how the data were collected, how often they are updated, and the meanings of the variables. It is hard for the general public to understand terms used in government departments or other institutions.
3) Data Usability: Open data platforms could add variables to facilitate data linkage. This would lower the barriers for the public to use the data. For example, property-based location data usually includes longitude and latitude. However, if government experts could add an additional column for GEOID (e.g., Federal Information Processing Standard (FIPS) code), users would not need to perform spatial merges with points and census polygons but could instead use a simple merge by matching the GEOID. This minor improvement would enable those unfamiliar with geographic information systems to conduct similar research.

4) Data Privacy: Even on open data platforms, it is crucial to establish clear levels of data openness for public understanding. This concerns data privacy and security. Users may feel uncomfortable knowing their personal property permit applications can be reviewed by anyone or that the county's quality evaluation of their property on an open data platform could influence its value. In eviction cases, data are often public and identify individuals by name and location. Thus, even in public data, privacy levels vary. Providing clear levels of openness in open data will help make both users and data providers (e.g., residents) more confident about open data. Some data, like eviction data, can be open but may require a request. If data holders provide clear instructions and time estimations for requests, it will also benefit researchers and local communities.

These suggestions for open administrative data collection and sharing strategies could facilitate smoother collaboration and data sharing between governments and researchers for research on social issues.

Social Welfare Intervention

As the model analyses and discussion show, the outcomes after eviction filings are very complicated. A single level is not sufficient to explain the differences in outcomes, which also indicates that a single-level intervention is not enough to meet all the needs of people who are facing eviction filings.

Individual intervention. The most common cause of eviction is nonpayment, and although the specific reasons behind nonpayment may vary, the direct cause is often a lack of money. In eviction research, this situation is referred to by the term *rent burden*. According to an analysis from the Joint Center for Housing Studies of Harvard University (2022), a record-high 22.4 million renter households have become rent burdened as the demand for rental units bounced back in the post-COVID-19 era.

However, easing the rent burden requires a long-term multilevel intervention. Several programs before eviction filing (pre-court stage) offer opportunities for social work to play a role in the short term. The goal of the pre-court stage is to increase knowledge about tenant rights and enhance tenants' ability to address unfair treatment from landlords. Tenants, especially those from racial and ethnic minority backgrounds who face a power imbalance, often struggle to understand their legal rights in the face of improper landlord behavior or potential eviction. Education on tenant rights and knowledge sharing can help tenants manage incidents such as rent increases, disputes over utility or repair fees, deposit disputes, and even eviction. According to a study in New Orleans, this knowledge is also beneficial after eviction filings, because it can decrease the likelihood of a default judgment and give renters more time to respond (Golio et al., 2023). This study found a difference in processing days based on whether the tenant answered the complaint, thus highlighting the importance of increasing response rates after an eviction filing.

Another pre-court intervention is mediation, serving as an alternative method for resolving evictions before a filing occurs. In Pierce County, renters receive a 14-day notice to vacate for nonpayment, meaning the case can be filed in court 2 weeks after this notice. Pierce County established a no-cost Eviction Resolution Pilot Program, which ended on June 30, 2023. There is no report on outcomes for this pilot program yet. However, data from other counties and states suggest its potential benefits. For example, the Community Mediation Service of Central Ohio (n.d.) reported helping more than 3,500 residents annually avoid eviction. The most common outcome is the creation of payment plans, found to be 58.8% of the cases in

Northampton, Massachusetts, and 45% in Baltimore, Maryland (Center for Dispute Resolution at the University of Maryland Baltimore, 2017; D. T. Eisenberg & Ebner, 2020; Kurtzberg & Henikoff, 1997). This finding is also consistent with analysis indicating that 25% of eviction filings with answer files involve a payment plan (Ren & Thomas, 2023). Nonetheless, if a case is filed, the record will remain with the renter, whereas pre-court mediation does not leave renters with an eviction history. An advanced model is the Eviction Diversion Program in Kalamazoo, Michigan, which involves collaboration with human services, nonprofit organizations, and other agencies to resolve cases no more than 3 months behind on rent. Social workers might participate in these mediation or partnership organizations to support these services.

Another potential intervention at the individual level is legal aid. This study did not have access to data on the attorneys representation of defendants. However, according to a report from Albany, New York, 2.47% of tenants were represented by attorneys, compared to 88.6% of property owners, from 2016 to 2021(D. Smith, 2022). This significant gap in representation places renters in a vulnerable situation. The lack of accessibility of legal aid demonstrates how social welfare can enhance legal support for renters in need. In many cases, it is not necessary for attorneys to represent tenants at a court hearing, but their assistance can help renters understand the process, provide guidance, and offer suggestions. One implication from this study is the need to assist renters in responding to a summons, including how to answer, what information should be included, and how to attend the hearing, because this response was identified as a significant factor in the random forest model.

Community Intervention. At the community level, this study revealed that a higher percentage of Black individuals in a census tract was associated with a decreased likelihood of dismissal. Furthermore, people of color were more likely to face eviction by the sheriff if they resided in census tracts with a higher percentage of White individuals. These findings prompt a reconsideration of the community’s role following an eviction filing. As discussed previously, social support, cultural connections, discrimination, and other factors could be related to this phenomenon. Rosen et al. (2021) and Decker (2021) both observed that landlords might rely on their personal experiences, including factors such as race and country of origin, in their decision-making processes. These experiences may result in biases and influence how nonpayment issues are handled. A report drawing on two national housing surveys noted that landlords in communities of color were more likely to defer maintenance and less likely to forgive rent (Kneebone et al., 2021). These disparities highlight the uneven development and segregation in communities, making people of color more vulnerable to being targeted by landlords. Therefore, initiatives such as promoting mixed-income communities are crucial for creating more inclusive environments.

Macro-Level Intervention. During the pandemic, the Emergency Rental Assistance program became available nationally as part of the pandemic assistance package in 2021. Evidence suggests that the program helped decrease eviction filings and increased the likelihood of dismissals. For instance, a report from Detroit, Michigan, indicated that the assistance may have contributed to a significant number of nonpayment cases being resolved, with 79% ending in dismissal during the pandemic (A. Eisenberg & Brantley, 2023). However, this trend was not observed in Pierce County.

A new challenge has emerged regarding the accessibility of benefits. A report highlighted that the average application processing time for Emergency Rental Assistance program in Wayne County was 90 days (A. Eisenberg & Brantley, 2023; Rahman, 2022). This was not an isolated issue. In 2021, Pierce County Human Services reported an average processing time of 88 days for renters from application submission to receipt of assistance (H. Smith, 2021). Comparatively, the median processing time from eviction filing to sheriff eviction is 39 days. Even with an additional 14-day notice, renters may only have 53 days to address the situation after missing their first rent payment. Given the average processing time of 88 days, renters could already be evicted by the time funding is approved.

Therefore, interventions at this level should focus on how to streamline the application process to be both smooth and time efficient. Given the rapid pace of the eviction process, human services departments need to accelerate their processes to match.

Policy Implications

Eviction filings and their outcomes represent different stages of the eviction process, and it is not feasible to apply a single eviction policy universally to all eviction scenarios. Several policies at different levels could influence eviction filings and outcomes, ultimately mitigating the eviction crisis.

At the individual level, as previously noted, the rate of defendant attorney representation in eviction cases is very low. Because eviction cases are civil matters, there is generally less assistance available. A policy could be established to increase legal assistance in civil cases or specifically, eviction cases. Seron et al. (2001) observed that in New York's Housing Court, legal counseling and representation resulted in fewer eviction writs. In 2017, New York became the first city in the United States to provide legal services to all tenants facing eviction proceedings. A report from the city's Office of Civil Justice (2019) showed that in fiscal year 2019, around 30% of tenants appearing in Housing Court for eviction cases were represented by an attorney, compared to 1% in 2013. Additionally, 84% of represented tenants remained in their homes and the eviction rate declined more than 30% since implementation. However, Greiner et al. (2012) found no statistically significant evidence of a relationship between the extent of representation and eviction court case outcomes, because most legal aid services only provide advice due to the high demand for services. Thus, legal assistance needs to be balanced in terms of coverage and quality.

Another policy suggestion at the individual level is to extend the pay or vacate notice period. According to data from Pierce County, eviction filings decreased by 12.5% after the pay or vacate notice period was extended from 3 days to 14 days. This does not necessarily mean the number of people involved in involuntary moves decreased, but rather that fewer people moved as part of the court process. This small change could prevent more individuals from facing official eviction and having their eviction records made public. Additionally, this extension could provide more time for renters to respond and relocate before reaching the stage of a sheriff eviction. Furthermore, according to the model in this study, property sales were linked to an increased likelihood of being evicted by the sheriff after judgment, and the reaction time differed between long-term and month-to-month renters. Therefore, policymakers may also need to consider revising the notice period for situations beyond nonpayment like housing trade.

At the community level, examining the outcomes of eviction in relation to renter burden and gentrification-related variables revealed a heightened risk of sheriff eviction. This raises an interesting question for the community: What happens when gentrification occurs in areas with a rent control policy? Gardner (2022) found that residing in rent-controlled units in San Francisco, California, increased the likelihood of eviction by approximately 240% annually. Similarly, Geddes and Holz (2023) also analyzed data from San Francisco, comparing policy-exposed by control ZIP codes. They found that the treatment group (exposed ZIP code) experienced an additional 34 eviction notices. Although the intent behind rent control is to stabilize the rental and housing market, it paradoxically was associated with an increase in evictions in this case. Landlords may use eviction as a tool to vacate properties, allowing for new renters, bypassing rent control policies, and maximizing profit during rent hikes driven by gentrification. This dynamic suggests that although rent control aims to alleviate the cost burden on tenants, it can inadvertently provide an incentive for eviction in gentrifying areas. Therefore, it is crucial to have rent control measures with additional support such as legal aid, to prevent eviction.

The most cost-effective strategy is often prevention. At the macro level, increasing the supply of affordable housing units is an effective way to alleviate cost burden by influencing the housing supply–demand balance. Since 1986, more than 3.6 million units have been constructed under the LIHTC initiative. However, approximately 10% of these units will revert to market rates between 2024 and 2029 as they meet the minimum 30-year affordability requirement. Additionally, around 7,000 units exit the program annually due to meeting tax code qualifications after 15 years (Joint Center for Housing Studies of Harvard University, 2024). However, the sheriff eviction model in this study indicates that properties financed through the LIHTC program have a higher likelihood of being involved in sheriff eviction. Preston and Reina (2021) expanded on this research by examining different subsidized housing programs, finding that public housing and project-based rental assistance properties were linked with a decreased eviction rate, a trend not observed in LIHTC properties. LIHTC properties are often managed by professional companies, which may result in delinquency cases more frequently proceeding to eviction due to their highly routinized systems (Decker, 2021). It is unclear whether these evictions occur more often in market-rate or affordable housing, but it is likely more prevalent on the affordable side, undermining the LIHTC program's goal of housing stability for vulnerable populations.

Furthermore, the findings reveal that the Pierce County Housing Authority was among the top three entities for eviction filings. Despite its mission to "provide safe, decent, affordable housing and economic opportunity, free from discrimination," it has become one of the county's leading evictors, alongside commercial companies(Pierce County Housing Authority, n.d., para 1). As mentioned previously, Leung et al. (2023) demonstrated that rent collection is a criterion used by the federal government to assess public housing authority management, as established by the U.S. Housing Act of 1937. However, the number of evictions is not included in this evaluation system. This oversight may lead housing authorities to prioritize evictions over long-term rental or housing stability. There is a clear need to reevaluate their role and mission in light of this finding. Overall, as affordable housing policy evolves, it is crucial to revisit whether these policies are still aligned with their intended mission.

Theoretical Contributions

Last, this study enriched the theoretical framework in eviction research. Prior literature on eviction often focused on eviction filing or final eviction rates, leaving processes that occurred after filings relatively understudied. In other words, research exploring variables that influence the decision-making process for dismissal, judgment, or sheriff eviction outcomes is rare. Furthermore, it is uncommon for studies to incorporate both contextual factors (such as gentrification, housing prices, and poverty) and individual characteristics (like race and gender). However, this study sought to bridge these gaps by incorporating multiple levels of data, offering a comprehensive view of how these multilayered factors interact to affect eviction outcomes. By integrating these perspectives, the study provides a conceptual framework supported by empirical evidence, which can help scholars gain a deeper understanding of the dynamics involved in the eviction process. This contribution can foster a more complete body of knowledge in eviction research.

Limitations and Future Research

Although this study attempted to provide a comprehensive view of various data sources related to eviction based on the literature, it still lacked several key housing indicators due to data limitations. For example, some measures of housing market competitiveness or supply, such as housing inventory levels, have only been collected from 2016 onward, which did not align with the study timeframe. Consequently, the model was constructed based on data from 2010 to 2019, because they shared the same GEOID and were gathered before the COVID-19 pandemic. To generate a larger sample for model building, I had to omit some variables that might have been relevant to the outcomes based on theory and evidence, such as rent amounts owed, which were only used in descriptive analysis. Furthermore, the number of eviction filings likely underestimated the true risk of eviction because lockouts can occur before filings and are unlikely to be recorded in the court system, potentially introducing unknown bias into cases documented in the court system.

Additionally, some algorithms, like name estimation, carry inherent biases, such as high accuracy in predicting White people's name. Although I assessed the standalone performance of this algorithm and articulated the level of confidence in its use, the impact of errors in the estimation for imputation purposes on the overall model remains unclear. Consequently, the prediction mechanism can sometimes be challenging to explain or lack full transparency. As discussed in the data generation section, CSS can help create possibilities for more variables, but the accuracy of imputation or creation is unknown, because it is very hard to validate. For instance, in estimating race and ethnicity, I used only the first defendant's name to represent each case, but there are undoubtedly cases with multiple races or ethnicities involved. Therefore, I sought to maintain as much transparency as possible and provide a foundation for further discussion.

Although the LLM did not produce the best results in this study due to minor bugs in preprocessing, it remains a highly promising method for future testing and application. The format of summonses heavily depends on the law firm's template, meaning that if a law firm changes or updates its template, the layout model may need to be updated to accommodate the new format. This limitation restricts the model's ability to be expanded to other counties and states. However, the LLM can handle these more flexible formats, because it possesses a deeper understanding of the text's meaning. Relying solely on text, a simple prompt like "Where does the defendant live?" is sufficient to extract the necessary information in an LLM.

Moreover, this study primarily focused on the correlation between eviction rates and factors such as community demographics. Although the cause of eviction is a subject of ongoing research, this study did not delve deeply into this issue. Many studies on this topic relied on surveys and interviews but could not systematically analyze and classify documents that defendants submitted to the court system. Developing an effective method to analyze court files from both parties and summarizing the conflict would significantly improve eviction research. For example, approximately 10% of cases in Pierce County included tenant answer files that described the events leading to eviction, such as job loss, health issues, and even poor landlord behavior or substandard housing conditions. However, because 10% represents only a small fraction of all cases, there is a clear need for innovative approaches to collect more comprehensive data on the causes of eviction, particularly focusing on nonpayment issues, which account for around 90% of eviction filings. Legislative measures like Washington's HB 1236 passed in 2021, which requires landlords to cite just cause for terminating leases under certain conditions, offer new avenues for data collection. Therefore, accurately identifying the reasons behind evictions is crucial not only for legal compliance but also to connect affected tenants with welfare programs and helping them stay housed.

Another future direction for research is addressing the limited empirical knowledge on posteviction outcomes. A significant question is where renters reside after being evicted. Although it is possible to identify some renters who have had multiple evictions filed by the same landlord, most individuals only appear in eviction datasets once, making it difficult to track their mobility. However, external datasets, such as credit report data, might provide insights into location changes, enabling the creation of panel data to examine posteviction residential mobility. This panel dataset could be similar to the Future of Families and Child Wellbeing Study, aiding in our understanding of how eviction affects renters' lives and their likelihood of experiencing homelessness(Reichman et al., 2001).

In this study, I attempted to establish a connection between eviction cases and landlord lawyer representation. Some law firms had a significant relationship with outcomes in the models. However, explaining the reasons behind these associations remains challenging. For example, why did 16.6% of organizational entities chose Puckett & Redford, whereas only 5.1% of private landlords opted for this firm? The mechanisms underlying the interactions between lawyers and landlords require more study. Questions arise as to whether some law firms may advise landlords to pursue more aggressive or lenient resolutions or if certain firms primarily represent large property management companies as part of their business strategy. This aspect was not covered in this study, making it difficult to understand how these law firms operate and their influence on the outcomes of eviction filings. Therefore, conducting a network analysis of these law firms could provide insight into the relationships between law firms and landlords, offering a clearer picture of the eviction landscape in Pierce County and elsewhere.

Nonprofit organizations have long been recognized as critical entities in providing human and social services where markets and governments fall short (Powell & Bromley 2020). Despite the significant role of nonprofit organizations, research that bridges eviction and nonprofit organizations is rare. The interaction between residents at risk of eviction and nonprofit organizations is hard to track. Some studies have examined the impact of legal aid organizations on eviction cases (Greiner et al., 2012). However, it is challenging to identify a comprehensive list of nonprofit organizations that provide services for people facing eviction. For instance, some community organizations may offer eviction-related services but do not categorize themselves in the housing sector. Missing information on the role of nonprofit organizations limits our understanding of their influence on eviction risk.

Conclusion

As the United States transitions away from pandemic-era renter assistance and policy protections, 19 million U.S. renters burdened with housing costs (U.S. Census Bureau, 2022) are encountering new challenges, including higher inflation, increased costs of food and basic necessities, and record-breaking eviction rates in some states (Legal Services Corporation, 2024). Eviction adversely affects renters in numerous ways, influencing housing stability, health, child development, and poverty levels. In the best-case scenario, eviction cases may be dismissed due to plaintiff withdrawal or a judgment in favor of the defendant. However, the eviction mark on their records can still limit their future housing options. Although the negative effects of eviction are well documented, data for measuring current coverage and trends remain largely incomplete (Garnham et al., 2022). This makes it challenging for local governments and policymakers to make informed decisions during this swift transition. The absence of comprehensive data to answer urgent questions, such as where evictions occur and which groups are most affected, hinders effective policymaking. In Washington, as in some other states, these urgently needed answers are hidden in individual court files, often in PDF format.

Behind each eviction case lies a real person or household. However, research data often aggregate these individuals by using rates at the census scale. As a result, most studies have attempted to establish connections between various census data and other variables to eviction filing and sheriff eviction rates but neglect the outcomes for individuals after eviction filings. Specifically, they have not examined which factors contribute to a case being dismissed, how to mitigate the adverse impacts of eviction filings, and what kinds of interventions the social welfare field can provide and advocate for to support renters. Addressing what happens after eviction filings is essential.

This dissertation sought to address challenges in eviction research, including the lack of individual data accessibility, understanding of case dismissals, and evidence of who is more likely to encounter sheriff eviction. First, the findings show that document layout analysis accurately extracted key information from court files, such as defendant and plaintiff names, eviction addresses, amounts involved, case outcomes, and processing times. Manually extracting this information from about 56,000 cases in an 18-year period would have required thousands of hours. Document layout analysis, combined with other computational methods, offers a solid foundation for eviction research. Second, in addressing RQ2 and RQ3, several common factors at the individual level were identified as strongly influencing eviction filing outcomes, including being a person of color, having a property sale record in the same year as the eviction filing, representation by certain law firms, the taxable value of the property, and whether there was a response to the summons. At the community level, the percentage of people under the poverty level and the percentage of rent-burdened households were strongly connected to eviction filing outcomes. At the county level, a housing price index and fair rent price were also associated with eviction outcomes. Interactions between an individual's race or ethnicity and the proportion of White people in the census tract revealed that people of color experienced different eviction filing outcomes compared to White individuals in the same community. Both logistic regression and random forest analyses indicated that many individual variables significantly affected eviction filing outcomes, yet these variables were rarely included in previous eviction research.

This dissertation highlighted the advantages and challenges of CSS that are rarely encountered in traditional social science research. CSS offers the potential to create new variables based on existing raw data but also presents challenges in presenting data clearly and maintaining transparency throughout the process. In eviction research, computational methods open many areas for exploration. On the one hand, there are possibilities for advanced methods such as NLP to analyze text-heavy court files, or computer vision to process street views. On the other hand, basic descriptive analysis and making data visualizations accessible to the public are crucial for collectively addressing the eviction crisis, as they help the public easily understand the hidden information in the eviction data.

Last, this dissertation outlined the types of interventions that can be provided by service providers in the social welfare field at different levels, such as helping tenants understand their rights and resources, improving welfare application procedures, and accelerating the fund distribution process.

Not all eviction cases are alike, given they occur in the lives of individuals in different properties, communities, and states with different outcomes. However, at their core, all eviction cases share common roots, stemming from challenging periods in life and often leading to adverse outcomes. This dissertation is critical to finding a path toward resolving this social challenge and supporting all people with safe and secure housing.

References

Acharya, B., Bhatta, D., & Dhakal, C. (2022). The risk of eviction and the mental health outcomes among the US adults. *Preventive Medicine Reports, 29*, 101981. https://doi.org/10.1016/j.pmedr.2022.101981

Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., ... & Zoph, B. (2024). GPT-4 technical report (arXiv:2303.08774). *arXiv*. https://doi.org/10.48550/arXiv.2303.08774

Ahn, P. (2023). New York eviction process (2024): Grounds, steps & timeline. *iPropertyManagement.com*. Retrieved from https://ipropertymanagement.com/laws/new-york-eviction-process

Aiken, E. L., Bedoya, G., Coville, A., & Blumenstock, J. E. (2020). Targeting development aid with machine learning and mobile phone data: Evidence from an anti-poverty intervention in Afghanistan. In *Proceedings of the 3rd ACM SIGCAS Conference on Computing and Sustainable Societies* (pp. 310–311). Association for Computing Machinery. https://doi.org/10.1145/3378393.3402274

Atkinson, R. (2004). The evidence on the impact of gentrification: New lessons for the urban renaissance? *European Journal of Housing Policy, 4*, 107–131. https://doi.org/10.1080/1461671042000215479

Bachmann, M. (2023). python-Levenshtein: Python extension for computing string edit distances and similarities (Version 0.25.0) [Python]. Retrieved from https://github.com/rapidfuzz/python-Levenshtein

Bates, L., Zapata, M., Greene, J., & Knowlton, S. (2021). Cost of Oregon evictions report. *Homelessness Research & Action Collaborative Publications and Presentations*. Retrieved from https://archives.pdx.edu/ds/psu/34920

Belsley, D. A., Kuh, E., & Welsch, R. E. (2005). *Regression diagnostics: Identifying influential data and sources of collinearity*. John Wiley & Sons.

Berzin, S. C., Singer, J., & Chan, C. (2015). Practice innovation through technology in the digital age: A grand challenge for social work. 22.

Bolukbasi, T., Chang, K.-W., Zou, J., Saligrama, V., & Kalai, A. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *arXiv:1607.06520* [Cs, Stat]. Retrieved from http://arxiv.org/abs/1607.06520

Bradford, A. C., & Bradford, W. D. (2023). The effect of state housing policies on eviction filings and judgments in the United States, 2001–2018. *Housing Policy Debate, 0*(0), 1–21. https://doi.org/10.1080/10511482.2023.2271881

Bullinger, L. R., & Fong, K. (2021). Evictions and neighborhood child maltreatment reports. *Housing Policy Debate, 31*(3–5), 490–515. https://doi.org/10.1080/10511482.2020.1822902

Burt, M. R. (2001). Homeless families, singles, and others: Findings from the 1996 national survey of homeless assistance providers and clients. *Housing Policy Debate, 12*(4), 737–780. https://doi.org/10.1080/10511482.2001.9521428

California State Senate. (2021, June 9). SB 9: The California HOME Act. *Focus*. Retrieved from https://focus.senate.ca.gov/sb9

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., … Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour, 2*(9), 637–644. https://doi.org/10.1038/s41562-018-0399-z

CBNData. (2018, October 9). 2018 年轻人租房大数据报告. Millennial renting big data report 2018. Retrieved from https://www.cbndata.com/report/1161/detail?isReading=report&page=1

Center for Dispute Resolution at the University of Maryland Baltimore. (2017). Report on the 2016 rent court ADR pilot for the District Court of Maryland in Baltimore City (3). *C-DRUM Publications*. s. 3. Retrieved from https://digitalcommons.law.umaryland.edu/cdrum_fac_pubs/3

Chen, X., Ren, C., & Thomas, T. A. (2024). Evaluating tenant-landlord tensions using generative ai on online tenant forums. *arXiv preprint arXiv:2404.11681*.

Chintalapati, R., Laohaprapanon, S., & Sood, G. (2023). Predicting race and ethnicity from the sequence of characters in a name (arXiv:1805.02109). *arXiv*. Retrieved from http://arxiv.org/abs/1805.02109

Chum, A. (2015). The impact of gentrification on residential evictions. *Urban Geography, 36*(7), 1083–1098. https://doi.org/10.1080/02723638.2015.1049480

Cioffi-Revilla, C. (2010). Computational social science: Computational social science. *Wiley Interdisciplinary Reviews: Computational Statistics, 2*(3), 259–271. https://doi.org/10.1002/wics.95

Cioffi-Revilla, C. (2017). Introduction to computational social science. Springer International Publishing. https://doi.org/10.1007/978-3-319-50131-4

Clark, A., Zager, K., & Lane, J. (2018). Charlotte-Mecklenburg evictions part 2: Mapping evictions. *UNC Charlotte Urban Institute*. Retrieved from https://mecklenburghousingdata.org/wp-content/uploads/2018/03/Charlotte-Mecklenburg-Evictions-Part-2-Mapping-Evictions.pdf

Collinson, R., Humphries, J. E., Mader, N., Reed, D., Tannenbaum, D., & van Dijk, W. (2024). Eviction and poverty in American cities. *The Quarterly Journal of Economics, 139*(1), 57–120. https://doi.org/10.1093/qje/qjad042

Collinson, R., & Reed, D. (2019). The effects of evictions on low-income households. 83.

Conte, R., Gilbert, N., Bonelli, G., Cioffi-Revilla, C., Deffuant, G., Kertesz, J., Loreto, V., Moat, S., Nadal, J.-P., Sanchez, A., Nowak, A., Flache, A., San Miguel, M., & Helbing, D. (2012). Manifesto of computational social science. *The European Physical Journal Special Topics, 214*(1), 325–346. https://doi.org/10.1140/epjst/e2012-01697-8

Cookson, T., Diddams, M., Maykovich, X., & Witter, E. (2018). Losing home: The human cost of eviction in Seattle. 88.

Coulton, C. J., Goerge, R., & Putnam-Hornstein, E. (2015). Harnessing big data for social good: A grand challenge for social work. 21.

Cunningham, M. K., Hariharan, A., & Fiol, O. (2021). The looming eviction cliff. 11.

De Sousa, T., Andrichik, A., Cuellar, M., Marson, J., Prestera, E., & Rush. (2022). The 2022 annual homelessness assessment report (AHAR to Congress) Part 1: Point-in-time estimates of homelessness, December 2022. Retrieved from https://www.huduser.gov/portal/sites/default/files/pdf/2022-ahar-part-1.pdf

Decker, N. (2023). How landlords of small rental properties decide who gets housed and who gets evicted. *Urban Affairs Review, 59*(1), 170–199. https://doi.org/10.1177/10780874211041513

Department for Levelling Up Housing & Communities. (2023). English housing survey 2021 to 2022: Private rented sector—GOV.UK. Retrieved from https://www.gov.uk/government/statistics/english-housing-survey-2021-to-2022-private-rented-sector/english-housing-survey-2021-to-2022-private-rented-sector

Desmond, M. (2012). Eviction and the reproduction of urban poverty. *American Journal of Sociology, 118*(1), 88–133. https://doi.org/10.1086/666082

Desmond, M. (2016). *Evicted: Poverty and profit in the American city*. Crown.

Desmond, M. (2020). On the brink of homelessness: How the affordable housing crisis and the gentrification of America is leaving families vulnerable. Retrieved from https://www.congress.gov/116/meeting/house/110362/witnesses/HHRG-116-BA00-Wstate-DesmondM-20200114.pdf

Desmond, M., An, W., Winkler, R., & Ferriss, T. (2013). Evicting children. *Social Forces, 92*(1), 303–327. https://doi.org/10.1093/sf/sot047

Desmond, M., & Gershenson, C. (2017). Who gets evicted? Assessing individual, neighborhood, and network factors. *Social Science Research, 62*, 362–377. https://doi.org/10.1016/j.ssresearch.2016.08.017

Desmond, M., Gromis, A., Edmonds, L., Hendrickson, J. R., Krywokulski, K., & Leung, L. (2018). Methodology report. *Princeton University*. Retrieved from www.evictionlab.org/methods

Desmond, M., & Kimbro, R. T. (2015). Eviction's fallout: Housing, hardship, and health. *Social Forces, 94*(1), 295–324. https://doi.org/10.1093/sf/sov044

Eisenberg, A., & Brantley, K. (2023). The crisis is not temporary: Evictions after emergency rental assistance in Detroit.

Eisenberg, D. T., & Ebner, N. (2020). Disrupting the eviction crisis with conflict resolution strategies. *Mitchell Hamline Law Journal of Public Policy and Practice*, *41*(3). Retrieved from https://open.mitchellhamline.edu/policypractice/vol41/iss3/2

Elgin, D. J. (2018). Utilizing predictive modeling to enhance policy and practice through improved identification of at-risk clients: Predicting permanency for foster children. *Children and Youth Services Review, 91*, 156–167. https://doi.org/10.1016/j.childyouth.2018.05.030

Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Publishing Group.

Eurostat. (2021). Housing in Europe—House or flat – owning or renting [2021]. Housing in Europe. Retrieved from https://ec.europa.eu/eurostat/cache/digpub/housing/bloc-1a.html

Evans, J. A., & Aceves, P. (2016). Machine translation: Mining text for social theory. *Annual Review of Sociology, 42*(1), 21–50. https://doi.org/10.1146/annurev-soc-081715-074206

Eviction Lab. (2022). Supplementary information estimating eviction prevalence across the United States. Retrieved from https://evictionlab.org/docs/Eviction_Lab_Methodology_Report_2022.pdf

Eviction Lab. (2024). Why eviction matters. *Eviction Lab*. Retrieved from https://evictionlab.org/why-eviction-matters/

Gardner, M. (2022). The effect of rent control status on eviction filing rates: Causal evidence from San Francisco. *Housing Policy Debate, 0*(0), 1–24. https://doi.org/10.1080/10511482.2022.2099932

Garnham, J. P., Gershenson, C., & Desmond, M. (2022, July 11). New data release shows that 3.6 million eviction cases were filed in the United States in 2018. *Eviction Lab*. Retrieved from https://evictionlab.org/es/new-eviction-data-2022/

Geddes, E., & Holz, N. (2023). Rational eviction: How landlords use evictions in response to rent control (SSRN Scholarly Paper 4131396). Retrieved from https://doi.org/10.2139/ssrn.4131396

Gentzkow, M., Kelly, B., & Taddy, M. (2019). Text as data. *Journal of Economic Literature, 57*(3), 535–574. https://doi.org/10.1257/jel.20181020

Gerr, R. (2019). The impact of Medicaid expansion on eviction rates [M.P.P., Georgetown University]. Retrieved from https://www.proquest.com/docview/2240074970/abstract/5E7617AC91384821PQ/1

Golio, A., Daniels, G., Moran, R., Southall, Y., & Lamoza, T. (2022). Eviction court outcomes and access to procedural knowledge: Evidence from a tenant-focused intervention in New Orleans. *Housing Policy Debate, 33*, 1–20. https://doi.org/10.1080/10511482.2022.2112257

Gomory, H., Massey, D. S., Hendrickson, J. R., & Desmond, M. (2023). The racially disparate influence of filing fees on eviction rates. *Housing Policy Debate, 33*(6), 1463–1483. https://doi.org/10.1080/10511482.2023.2212662

Goodspeed, R., Benton, E., & Slugg, K. (2021). Eviction case filings and neighborhood characteristics in urban and rural places: A Michigan statewide analysis. *Housing Policy Debate, 31*(3–5), 717–735. https://doi.org/10.1080/10511482.2020.1867882

Graetz, N., Gershenson, C., Hepburn, P., Porter, S. R., Sandler, D. H., & Desmond, M. (2023). A comprehensive demographic profile of the US evicted population. *Proceedings of the National Academy of Sciences, 120*(41), e2305860120. https://doi.org/10.1073/pnas.2305860120

Graetz, N., Gershenson, C., Porter, S. R., Sandler, D. H., Lemmerman, E., & Desmond, M. (2024). The impacts of rent burden and eviction on mortality in the United States, 2000–2019. *Social Science & Medicine, 340*, 116398. https://doi.org/10.1016/j.socscimed.2023.116398

Greenberg, D., Gershenson, C., & Desmond, M. (2016). Discrimination in evictions: Empirical evidence and legal challenges. *Harvard Civil Rights*, 51.

Greiner, D. J., Pattanayak, C. W., & Hennessy, J. (2013). The limits of unbundled legal assistance: A randomized study in a Massachusetts district court and prospects for the future. *Harvard Law Review*. Retrieved from https://dash.harvard.edu/handle/1/12956304

Griswold, M., Baker, L., Hunter, S. B., Ward, J., & Ren, C. (2024). Analyzing the effect of crime-free housing policies on completed evictions using spatial first differences. *Cityscape*, 26(1), 195-240.

Griswold, M., Holliday, S. B., Sizemore, A., Ren, C., Baker, L., Howell, K., Osoba, O. A., Williams, J., Ward, J. M., & Hunter, S. B. (2023). An evaluation of crime-free housing policies: Annex.

Gromis, A., Fellows, I., Hendrickson, J. R., Edmonds, L., Leung, L., Porton, A., & Desmond, M. (2022). Estimating eviction prevalence across the United States. *Proceedings of the National Academy of Sciences, 119*(21), e2116169119. https://doi.org/10.1073/pnas.2116169119

Gromis, A., Hendrickson, J. R., & Desmond, M. (2022). Eviction from public housing in the United States. *Cities, 127*, 103749. https://doi.org/10.1016/j.cities.2022.103749

Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology, 8*(1), 23–34.

Han, H.-S. (2023). What explains variation in neighborhood evictions? Investigation of neighborhood characteristics and federal rental assistance: Case study of Kansas City, Missouri. *Journal of Urban Affairs, 0*(0), 1–25. https://doi.org/10.1080/07352166.2023.2198131

Harrison, A., Immergluck, D., Ernsthausen, J., & Earl, S. (2021). Housing stability, evictions, and subsidized rental properties: Evidence from Metro Atlanta, Georgia. *Housing Policy Debate, 31*(3–5), 411–424. https://doi.org/10.1080/10511482.2020.1798487

Hartman, C., & Robinson, D. (2003). Evictions: The hidden housing problem. *Housing Policy Debate, 14*(4), 461–501. https://doi.org/10.1080/10511482.2003.9521483

Hazekamp, C., Yousuf, S., Day, K., Daly, M. K., & Sheehan, K. (2020). Eviction and pediatric health outcomes in Chicago. *Journal of Community Health, 45*(5), 891–899. https://doi.org/10.1007/s10900-020-00806-y

Heiberger, R. H., & Riebling, J. R. (2016). Installing computational social science: Facing the challenges of new information and communication technologies in social science. *Methodological Innovations, 9*, 2059799115622763. https://doi.org/10.1177/2059799115622763

Hepburn, P., Louis, R., & Desmond, M. (2023). Beyond gentrification: Housing loss, poverty, and the geography of displacement. *Social Forces, soad123*. https://doi.org/10.1093/sf/soad123

Hepburn, P., Louis, R., Fish, J., Lemmerman, E., Alexander, A. K., Thomas, T. A., Koehler, R., Benfer, E., & Desmond, M. (2021). U.S. eviction filing patterns in 2020. *Socius, 7*, 23780231211009983. https://doi.org/10.1177/23780231211009983

Hepburn, P., Olivia, J., Fish, J., Lemmerman, E., Alexander, A. K., & Desmond, M. (2022). Preliminary analysis: Eviction filing patterns in 2021. Preliminary analysis: Eviction filing patterns in 2021. Retrieved from https://evictionlab.org/us-eviction-filing-patterns-2021/

Himmelstein, G., & Desmond, M. (2021). Association of eviction with adverse birth outcomes among women in Georgia, 2000 to 2016. *JAMA Pediatrics, 175*(5), 494. https://doi.org/10.1001/jamapediatrics.2020.6550

Hoke, M. K., & Boen, C. E. (2021). The health impacts of eviction: Evidence from the National Longitudinal Study of Adolescent to Adult Health. *Social Science & Medicine (1982), 273*, 113742. https://doi.org/10.1016/j.socscimed.2021.113742

Hosmer, D. W., & Lemesbow, S. (1980). Goodness of fit tests for the multiple logistic regression model. *Communications in Statistics - Theory and Methods, 9*(10), 1043–1069. https://doi.org/10.1080/03610928008827941

Humphries, J. E., Mader, N. S., Tannenbaum, D. I., & van Dijk, W. L. (2019). Does eviction cause poverty? Quasi-experimental evidence from Cook County, IL (Working Paper 26139; Working Paper Series). National Bureau of Economic Research. https://doi.org/10.3386/w26139

Hussein, N., Bourret, T., & Gallagher, S. (2023). Eviction record sealing and expungement toolkit. National Low Income Housing Coalition.

Ihara, E. S., Horio, B. M., & Tompkins, C. J. (2012). Grandchildren caring for grandparents: Modeling the complexity of family caregiving. *Journal of Social Service Research, 38*(5), 619–636. https://doi.org/10.1080/01488376.2012.711684

Immergluck, D., Ernsthausen, J., Earl, S., & Powell, A. (2020). Evictions, large owners, and serial filings: Findings from Atlanta. *Housing Studies, 35*(5), 903–924. https://doi.org/10.1080/02673037.2019.1639635

Israel, N., & Wolf-Branigin, M. (2011). Nonlinearity in social service evaluation: A primer on agent-based modeling. *Social Work Research, 35*(1), 20–24. https://doi.org/10.1093/swr/35.1.20

Jayaprakash, C., Warren, K., Irwin, E., & Chen, K. (2009). The interaction of segregation and suburbanization in an agent-based model of residential location. *Environment and Planning B: Planning and Design, 36*(6), 989–1007. https://doi.org/10.1068/b33029

Johns-Wolfe, E. (2018). “You are being asked to leave the premises”: A study of eviction in Cincinnati and Hamilton County, Ohio, 2014-2017.

Joint Center for Housing Studies of Harvard University. (2022). *America’s rental housing 2022*.

Joint Center for Housing Studies of Harvard University. (2024). *America’s rental housing 2024*.

Kambath, A., Zhou, E., Archer, J., & Pola, S. (2021). Project HOME. Retrieved from https://jearcher.github.io/project-home/HTML/public/index.html

Khadka, A., Fink, G., Gromis, A., & McConnell, M. (2020). In utero exposure to threat of evictions and preterm birth: Evidence from the United States. *Health Services Research, 55*(S2), 823–832. https://doi.org/10.1111/1475-6773.13551

Khan, M. R., & Blumenstock, J. E. (2019). Multi-GCN: Graph convolutional networks for multi-view networks, with applications to global poverty. *Proceedings of the AAAI Conference on Artificial Intelligence, 33*(01), Article 01. https://doi.org/10.1609/aaai.v33i01.3301606

Kim, K., Garcia, I., & Brewer, S. (2021). Spatial relationship between eviction filings, neighborhood characteristics, and proximity to the central business district: A case study of Salt Lake County, Utah. *Housing Policy Debate, 31*, 1–26. https://doi.org/10.1080/10511482.2020.1838598

Kleysteuber, R. (2007). Tenant screening thirty years later: A statutory proposal to protect public records. *The Yale Law Journal, 116*(6), 1344. https://doi.org/10.2307/20455760

Kneebone, E., Decker, N., de la Campa, E., & Herbert, C. (2021). The impact of the pandemic on landlords: Evidence from two national surveys. Retrieved from https://www.jchs.harvard.edu/sites/default/files/research/files/harvard_jchs_impact_on_landlords_two_national_surveys_kneebone_et_al_2021.pdf

Knowable Magazine. (2020, September 29). The pandemic will likely make America’s eviction crisis even worse. *PBS NewsHour*. Retrieved from https://www.pbs.org/newshour/economy/the-pandemic-will-likely-make-americas-eviction-crisis-even-worse

Kouloumpis, E., Wilson, T., & Moore, J. (2021). Twitter sentiment analysis: The good, the bad, and the OMG! *Proceedings of the International AAAI Conference on Web and Social Media, 5*(1), 538–541. https://doi.org/10.1609/icwsm.v5i1.14185

Kurtzberg, J., & Henikoff, J. (1997). Freeing the parties from the law: Designing an interest and rights focused model of landlord/tenant mediation. 1997.

Lazer, D. M. J., Pentland, A., Watts, D. J., Aral, S., Athey, S., Contractor, N., Freelon, D., Gonzalez-Bailon, S., King, G., Margetts, H., Nelson, A., Salganik, M. J., Strohmaier, M., Vespignani, A., & Wagner, C. (2020). Computational social science: Obstacles and opportunities. *Science, 369*(6507), 1060–1062. https://doi.org/10.1126/science.aaz8170

LeBreton, J. M., & Senter, J. L. (2008). Answers to 20 questions about interrater reliability and interrater agreement. *Organizational Research Methods, 11*(4), 815–852. https://doi.org/10.1177/1094428106296642

Leckie, J. (2023). December 2023 rent report. Retrieved from https://www.rent.com/research/average-rent-price-report/

Legal Information Institute. (2021). Eviction. Legal Information Institute. Retrieved from https://www.law.cornell.edu/wex/eviction

Legal Services Corporation. (2024). Legal services corporation civil court data initiative. Eviction Tracker. Retrieved from https://civilcourtdata.lsc.gov/data/eviction

Leung, L., Hepburn, P., & Desmond, M. (2021). Serial eviction filing: Civil courts, property management, and the threat of displacement. *Social Forces, 100*(1), 316–344. https://doi.org/10.1093/sf/soaa089

Leung, L., Hepburn, P., Hendrickson, J., & Desmond, M. (2023). No safe harbor: Eviction filing in public housing. *Social Service Review, 97*(3), 456–497. https://doi.org/10.1086/725777

Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady, 10*(8), 707–710.

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., & Zettlemoyer, L. (2019). BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension (arXiv:1910.13461). *arXiv*. https://doi.org/10.48550/arXiv.1910.13461

Li, J., Xu, Q., Shah, N., & Mackey, T. K. (2019). A machine learning approach for the detection and characterization of illicit drug dealers on Instagram: Model evaluation study. *Journal of Medical Internet Research, 21*(6), e13803. https://doi.org/10.2196/13803

Linares, A., & Telemundo. (2023, October 19). Who's most affected by rental evictions? Kids, for starters. *NBC News*. Retrieved from https://www.nbcnews.com/news/latino/rental-evictions-impact-kids-black-women-latinos-study-finds-rcna121067

Lundberg, I., & Donnelly, L. (2018). A research note on the prevalence of housing eviction among children born in U.S. cities. *Demography, 56*(1), 391–404. https://doi.org/10.1007/s13524-018-0735-y

Mah, J. (2021). Gentrification-induced displacement in Detroit, Michigan: An analysis of evictions. *Housing Policy Debate, 31*(3–5), 446–468. https://doi.org/10.1080/10511482.2020.1800781

Manzini, T., Lim, Y. C., Tsvetkov, Y., & Black, A. W. (2019). Black is to criminal as Caucasian is to police: Detecting and removing multiclass bias in word embeddings. *arXiv:1904.04047* [Cs, Stat]. Retrieved from http://arxiv.org/abs/1904.04047

Martin, H. (2022). Making sense of eviction trends during the pandemic. *Economic Commentary, 2022–12*. https://doi.org/10.26509/frbc-ec-202212

McCabe, B. J., & Rosen, E. (2020). Eviction in Washington, DC: Racial and geographic disparities in housing instability. 42.

Medina, R. M., Byrne, K., Brewer, S., & Nicolosi, E. A. (2020). Housing inequalities: Eviction patterns in Salt Lake County, Utah. *Cities, 104*, 102804. https://doi.org/10.1016/j.cities.2020.102804

Merriam-Webster. (2021). Legal definition of EVICTION. Retrieved from https://www.merriam-webster.com/legal/eviction

Merritt, B., & Farnworth, M. D. (2021). State landlord–tenant policy and eviction rates in majority-minority neighborhoods. *Housing Policy Debate, 31*(3–5), 562–581. https://doi.org/10.1080/10511482.2020.1828989

Metlife. (2023). MetLife's 21st annual U.S. employee benefit trends study 2023. Retrieved from https://www.metlife.com/content/dam/metlifecom/us/noindex/pdf/ebts-2023/MetLife_EBTS_2023.pdf

Metzler, K., Kim, D. A., Allum, N., & Denman, A. (2016). Who is doing computational social science? Trends in big data research A SAGE white paper. *SAGE Publishing*. https://doi.org/10.4135/wp160926

Mullen, L. (2021). Gender [R]. Retrieved from https://github.com/lmullen/gender (Original work published 2014)

Nelson, K., Garboden, P., McCabe, B. J., & Rosen, E. (2021). Evictions: The comparative analysis problem. *Housing Policy Debate, 31*(3–5), 696–716. https://doi.org/10.1080/10511482.2020.1867883

Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., & Barabási, A.-L. (2007). Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences, 104*(18), 7332–7336. https://doi.org/10.1073/pnas.0610245104

Pierce County Housing Authority. (n.d.). About PCHA. Pierce County Housing Authority. Retrieved April 18, 2024, from https://www.pchawa.org/about-us/about-pcha

Pietri, R. (2013). Privacy in computational social science. 130.

Pleasants, E., Ryan, J. H., Ren, C., Prata, N., Gomez, A. M., & Marshall, C. (2023). Exploring language used in posts on r/birthcontrol: case study using data from reddit posts and natural language processing to advance contraception research. *Journal of medical Internet research*, 25, e46342.

Porton, A., Gromis, A., & Desmond, M. (2021). Inaccuracies in eviction records: Implications for renters and researchers. *Housing Policy Debate, 31*(3–5), 377–394. https://doi.org/10.1080/10511482.2020.1748084

Preston, G., & Reina, V. J. (2021). Sheltered from eviction? A framework for understanding the relationship between subsidized housing programs and eviction. *Housing Policy Debate, 31*(3–5), 785–817. https://doi.org/10.1080/10511482.2021.1879202

Rahman, N. (2022, April 27). Detroit city council wants statewide housing agency to speed up rent aid applications. BridgeDetroit. Retrieved from http://www.bridgedetroit.com/detroit-city-council-wants-statewide-housing-agency-to-speed-up-rent-aid-applications/

Ramphal, B., Keen, R., Okuzuno, S. S., Ojogho, D., & Slopen, N. (2023). Evictions and infant and child health outcomes: A systematic review. *JAMA Network Open, 6*(4), e237612. https://doi.org/10.1001/jamanetworkopen.2023.7612

Rangel, J., Hass, J., Lemmerman, E., Fish, J., & Hepburn, P. (2021). Preliminary analysis: 11 months of the CDC moratorium. Eviction Lab. Retrieved from https://evictionlab.org/eleven-months-cdc/

Raymond, E. L., Duckworth, R., Miller, B., Lucas, M., & Pokharel, S. (2016). Corporate landlords, institutional investors, and displacement: Eviction rates in single-family rentals (SSRN Scholarly Paper 2893552). Retrieved from https://papers.ssrn.com/abstract=2893552

Raymond, E. L., Duckworth, R., Miller, B., Lucas, M., & Pokharel, S. (2018). From foreclosure to eviction: Housing insecurity in corporate-owned single-family rentals. *Cityscape, 20*(3), 159–188.

Raymond, E. L., Miller, B., McKinney, M., & Braun, J. (2021). Gentrifying Atlanta: Investor purchases of rental housing, evictions, and the displacement of Black residents. *Housing Policy Debate, 31*(3–5), 818–834. https://doi.org/10.1080/10511482.2021.1887318

Reichman, N. E., Teitler, J. O., Garfinkel, I., & McLanahan, S. S. (2001). Fragile families: Sample and design. *Children and Youth Services Review, 23*(4–5), 303–326. https://doi.org/10.1016/S0190-7409(01)00141-4

Ren, C., & Bloemraad, I. (2022). New methods and the study of vulnerable groups: using machine learning to identify immigrant-oriented nonprofit organizations. *Socius*, 8, 23780231221076992.

Ren, C., Stuart, M., & Chow, J. C.-C. (2020). Computational social welfare: Applying data science in social work. Obo. Retrieved from https://www.oxfordbibliographies.com/view/document/obo-9780195389678/obo-9780195389678-0286.xml

Ren, C., & Thomas, T. (2023, January 15). Applying machine learning to understand the cause and cost of eviction. Retrieved from https://sswr.confex.com/sswr/2023/webprogram/Paper50910.html

Richter, F. G.-C., Coulton, C., Urban, A., & Steh, S. (2021). An integrated data system lens into evictions and their effects. *Housing Policy Debate, 31*(3–5), 762–784. https://doi.org/10.1080/10511482.2021.1879201

Robinson, D., & Steil, J. (2021). Eviction dynamics in market-rate multifamily rental housing. *Housing Policy Debate, 31*(3–5), 647–669. https://doi.org/10.1080/10511482.2020.1839936

Rosen, E., Garboden, P., & Cossyleon, J. (2021). Racial discrimination in housing: How landlords use algorithms and home visits to screen tenants. *American Sociological Review, 86*, 000312242110296. https://doi.org/10.1177/00031224211029618

Rutan, D. Q., & Desmond, M. (2021). The concentrated geography of eviction. *The ANNALS of the American Academy of Political and Social Science, 693*(1), 64–81. https://doi.org/10.1177/0002716221991458

Salganik, M. J. (2018). *Bit by bit: Social research in the digital age*. Princeton University Press.

Salganik, M. J., Lundberg, I., Kindel, A. T., Ahearn, C. E., Al-Ghoneim, K., Almaatouq, A., Altschul, D. M., Brand, J. E., Carnegie, N. B., Compton, R. J., Datta, D., Davidson, T., Filippova, A., Gilroy, C., Goode, B. J., Jahani, E., Kashyap, R., Kirchner, A., McKay, S., … McLanahan, S. (2020). Measuring the predictability of life outcomes with a scientific mass collaboration. *Proceedings of the National Academy of Sciences, 117*(15), 8398–8403. https://doi.org/10.1073/pnas.1915006117

Schrodt, P. (1987). *Microcomputer methods for social scientists*. SAGE Publications, Inc. https://doi.org/10.4135/9781412985062

Schwartz, G. L., Leifheit, K. M., Chen, J. T., Arcaya, M. C., & Berkman, L. F. (2022). Childhood eviction and cognitive development: Developmental timing-specific associations in an urban birth cohort. *Social Science & Medicine, 292*, 114544. https://doi.org/10.1016/j.socscimed.2021.114544

Schwartz, I. M., York, P., Nowakowski-Sims, E., & Ramos-Hernandez, A. (2017). Predictive and prescriptive analytics, machine learning and child welfare risk assessment: The

Broward County experience. *Children and Youth Services Review, 81*, 309–320. https://doi.org/10.1016/j.childyouth.2017.08.020

Seron, C., Frankel, M., Van Ryzin, G., & Kovath, J. (2001). The impact of legal counsel on outcomes for poor tenants in New York City's housing court: Results of a randomized experiment. *Law & Society Review, 35*(2), 419–434. https://doi.org/10.2307/3185408

Seymour, E. (2022). Corporate landlords and pandemic and prepandemic evictions in Las Vegas. *Housing Policy Debate, 0*(0), 1–22. https://doi.org/10.1080/10511482.2022.2125335

Sims, J. R. (2021). Measuring the effect of gentrification on displacement: Multifamily housing and eviction in Wisconsin's Madison urban region. *Housing Policy Debate, 31*(3–5), 736–761. https://doi.org/10.1080/10511482.2021.1871931

Slee, G., & Desmond, M. (2023). Eviction and voter turnout: The political consequences of housing instability. *Politics & Society, 51*(1), 3–29. https://doi.org/10.1177/00323292211050716

Smith, D. (2022, January 4). Albany eviction data. Retrieved from https://www.albanyny.gov/2009/Albany-Eviction-Data

Smith, H. (2021, September 11). Pierce County faces highest eviction rate in Washington as bridge moratorium nears expiration date. *King5.Com*. Retrieved from https://www.king5.com/article/news/local/homeless/pierce-county-rental-assistance-eviction-rate-bridge-moratorium/281-7a6acb07-c20e-41e1-a083-452924287b43

Smith, N. (1979). Toward a theory of gentrification: A back to the city movement by capital, not people. *Journal of the American Planning Association, 45*(4), 538–548. https://doi.org/10.1080/01944367908977002

Smith-Clarke, C., Mashadi, A., & Capra, L. (2014). Poverty on the cheap: Estimating poverty maps using aggregated mobile communication networks. 10.

Sumrall, F. (2023, March 22). Eviction notices one of the most dangerous jobs a police officer faces. *MyNorthwest.Com*. Retrieved from https://mynorthwest.com/3862444/eviction-notice-dangerous-job-police-officer/

Tan, J. (2020). Using machine learning to identify populations at high risk for eviction as an indicator of homelessness. 93.

Teresa, B. F., & Howell, K. L. (2021). Eviction and segmented housing markets in Richmond, Virginia. *Housing Policy Debate, 31*(3–5), 627–646. https://doi.org/10.1080/10511482.2020.1839937

Tesfai, R., & Ruther, M. (2022). Immigrant neighborhoods and eviction: Hidden housing crisis? *Cities, 131*, 104033. https://doi.org/10.1016/j.cities.2022.104033

Thomas, T. A. (2017). Forced out: Race, market, and neighborhood dynamics of evictions. 131.

Thomas, T. A., Ramiller, A., Ren, C., & Toomet, O. (2024). Toward a national eviction data collection strategy using natural language processing. *Cityscape, 26*(1), 214–259.

Treglia, D., Byrne, T., & Tamla Rai, V. (2023). Quantifying the impact of evictions and eviction filings on homelessness rates in the United States. *Housing Policy Debate, 1*–12. https://doi.org/10.1080/10511482.2023.2186749

Tsai, J., Jones, N., Szymkowiak, D., & Rosenheck, R. A. (2021). Longitudinal study of the housing and mental health outcomes of tenants appearing in eviction court. *Social Psychiatry and Psychiatric Epidemiology, 56*(9), 1679–1686. https://doi.org/10.1007/s00127-020-01953-2

Urban, A., Tyler, A., Richter, F., Coulton, C., & Chan, T. (2019). The Cleveland eviction study: Observations in eviction court and the stories of people facing eviction. Center on Urban

Poverty and Community Development. Retrieved from https://case.edu/socialwork/povertycenter/sites/case.edu.povertycenter/files/2019-11/The%20Cleveland%20Eviction%20Study-10242019-fully%20accessible%28r%29.pdf

Vallejo, C., Haas, J., & Hepburn, P. (2023). Preliminary analysis: Eviction filing patterns in 2022. Eviction Lab. Retrieved from https://evictionlab.org/ets-report-2022/

Vásquez-Vera, H., Palència, L., Magna, I., Mena, C., Neira, J., & Borrell, C. (2017). The threat of home eviction and its effects on health through the equity lens: A systematic review. *Social Science & Medicine, 175*, 199–208. https://doi.org/10.1016/j.socscimed.2017.01.010

Wais, K. (2016). genderizeR [R]. Retrieved from https://github.com/kalimu/genderizeR (Original work published 2014)

Warren, A., McKernan, S.-M., & Braga, B. (2020, April 17). Before COVID-19, 68 million US adults had debt in collections. What policies could help? | Urban Institute. Retrieved from https://www.urban.org/urban-wire/covid-19-68-million-us-adults-had-debt-collections-what-policies-could-help

Washington State Legislature. (1973). Chapter 59.18 RCW: RESIDENTIAL LANDLORD-TENANT ACT. Retrieved from https://apps.leg.wa.gov/rcw/default.aspx?cite=59.18&full=true

Wedeen, S. (2021, January 11). Black and Hispanic renters face greatest threat of eviction in pandemic | Joint Center for Housing Studies. Retrieved from https://www.jchs.harvard.edu/blog/black-and-hispanic-renters-face-greatest-threat-eviction-pandemic

White, M. (2023, May 2). Amazon's return-to-office: Could it impact how much you pay for housing? *King5.Com*. Retrieved from https://www.king5.com/article/news/local/seattle/amazon-return-to-office-housing-rent-increase/281-f5413cfa-e57b-4ed4-8ad8-8a91c01a5911

Wimmer, A., & Lewis, K. (2010). Beyond and below racial homophily: ERG models of a friendship network documented on Facebook. *American Journal of Sociology, 116*(2), 583–642. https://doi.org/10.1086/653658

Xie, F. (2022). rethnicity: An R package for predicting ethnicity from names. *SoftwareX, 17*, 100965. https://doi.org/10.1016/j.softx.2021.100965

Zhao, J., Wang, T., Yatskar, M., Cotterell, R., Ordonez, V., & Chang, K.-W. (2019). Gender bias in contextualized word embeddings. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 629–634. https://doi.org/10.18653/v1/N19-1064

Appendix

Figure A1. California eviction flowchart (Ahn, 2023)

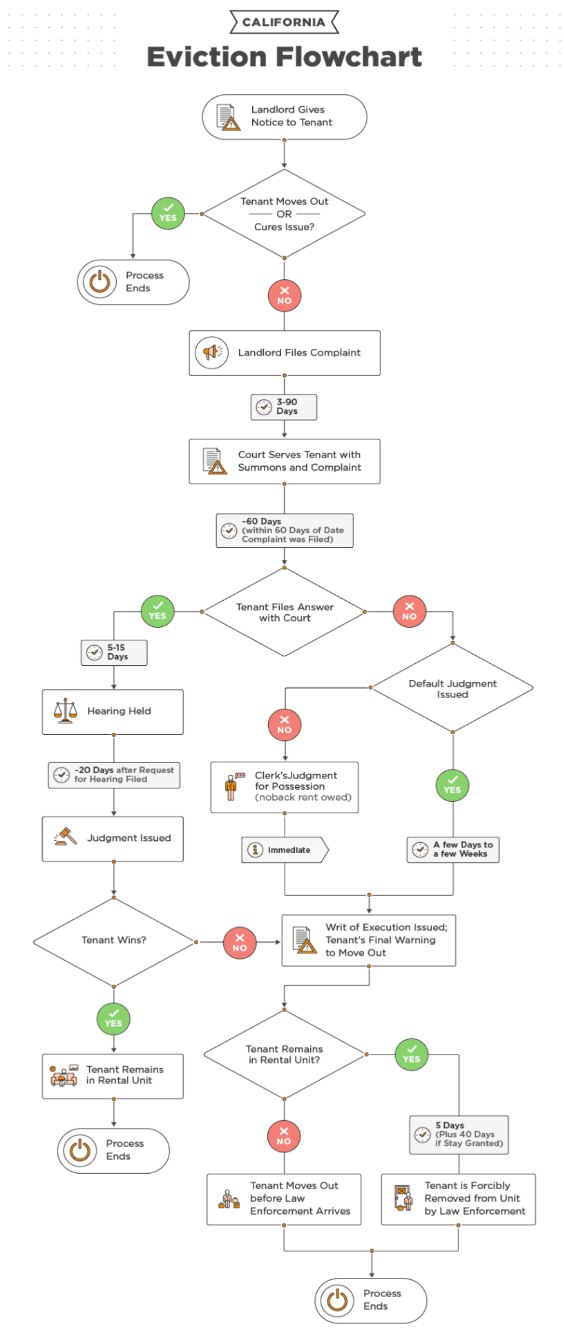


Figure A2. New York eviction flowchart (Ahn, 2023)

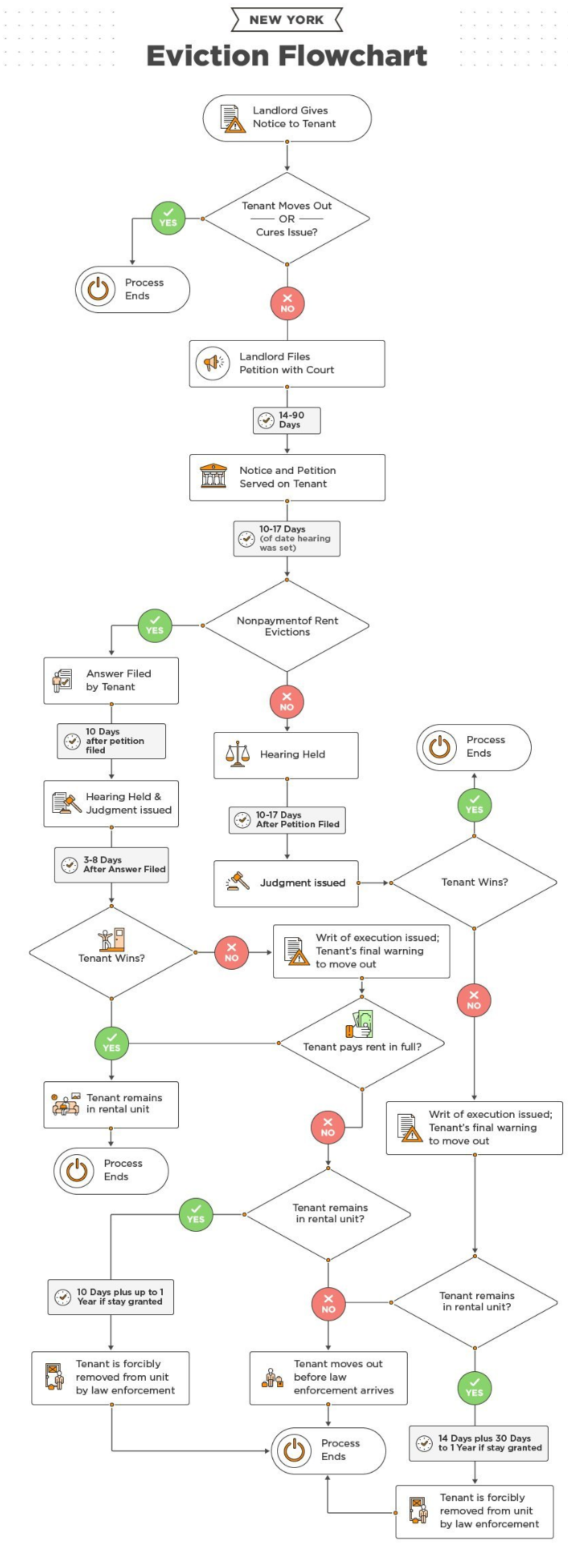


Figure A3 The labelling sample of motion for judgement file.

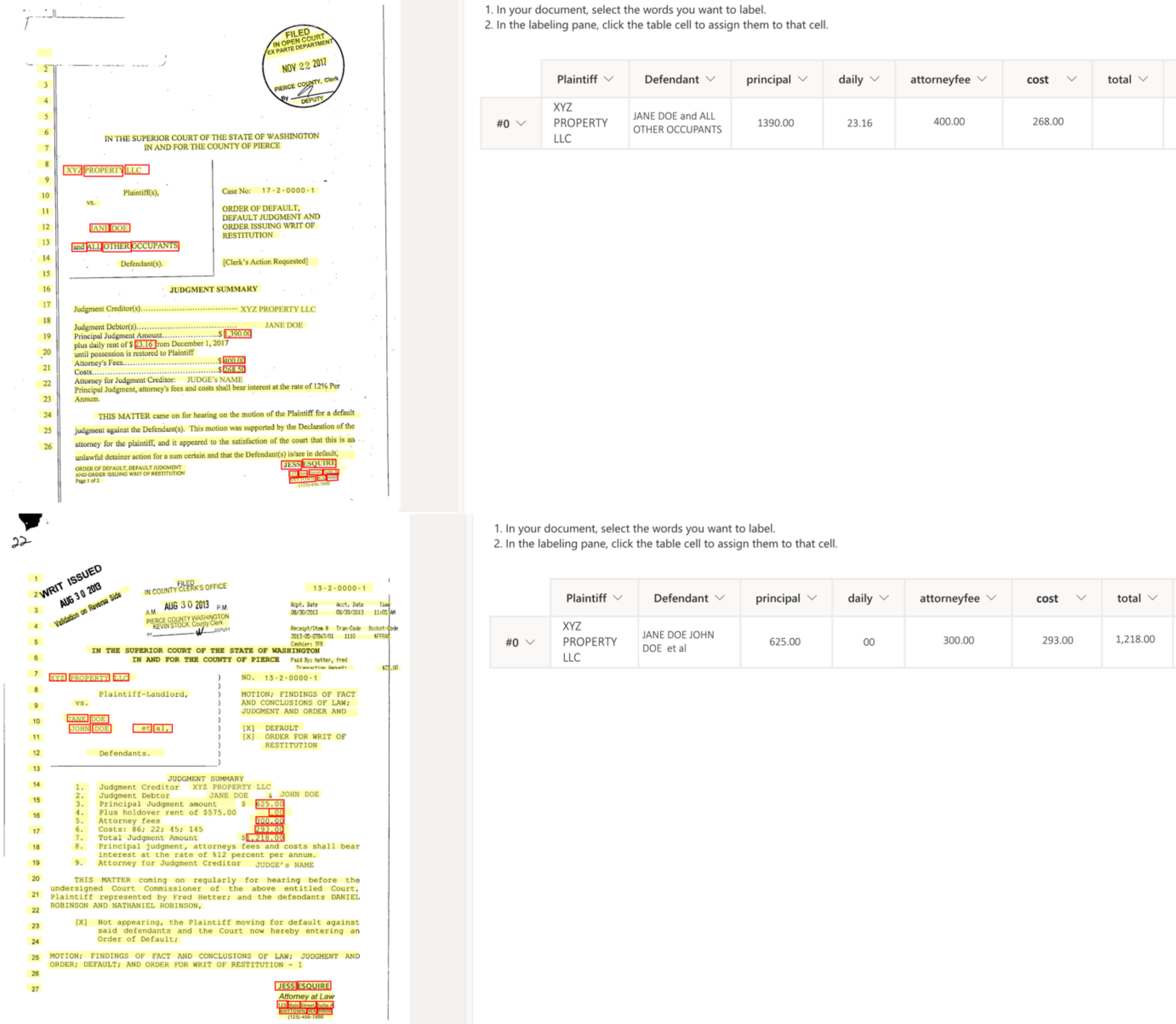


Figure A4 The labelling sample of sheriff's return writ.

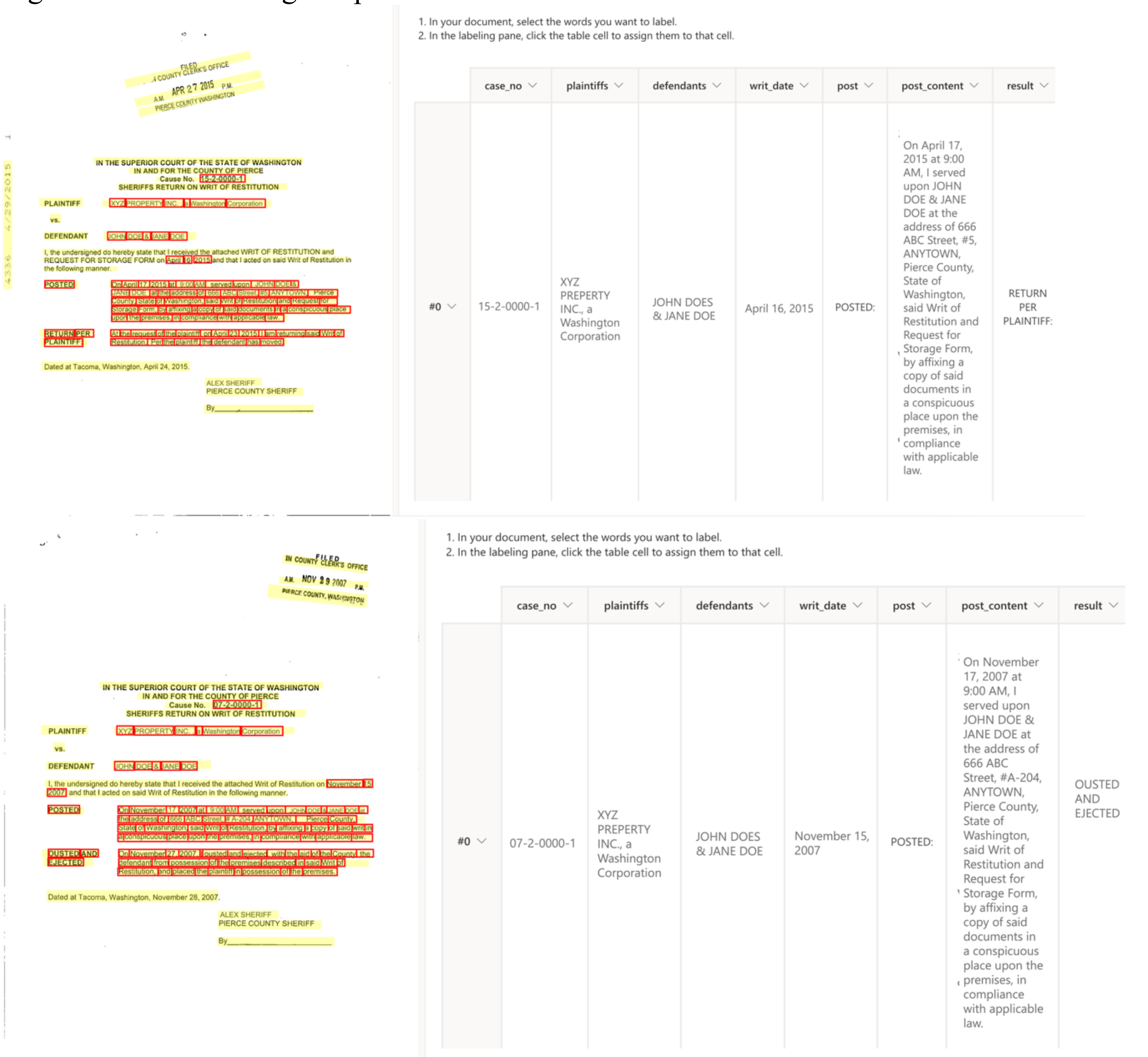


Figure A5 The distribution of Nonwhite in Pierce County, WA

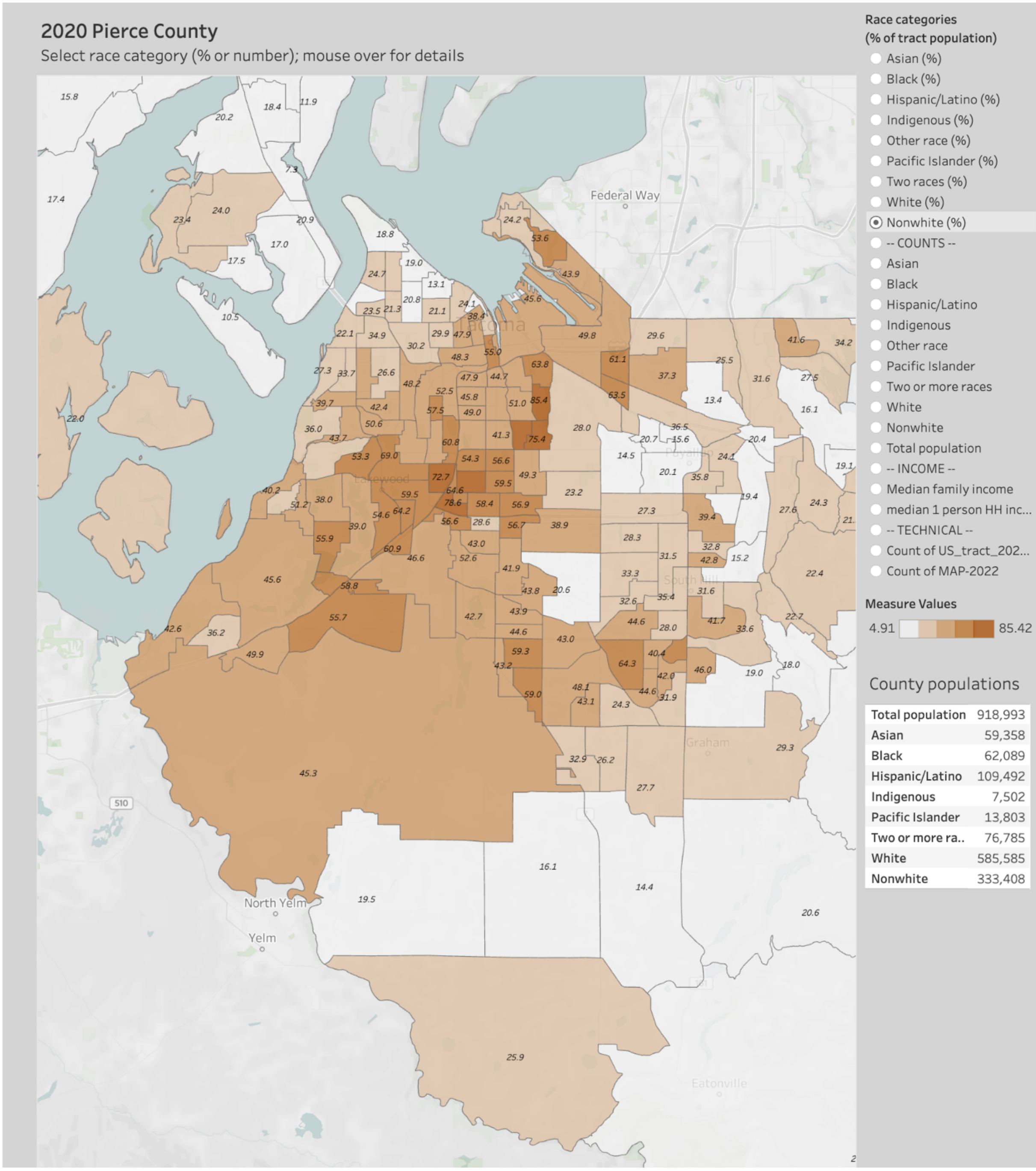


Table A1: Variance Inflation Factor (VIF) table for outcome of dismissal or judgement at different factor levels.

Variable	**Model 1- individual**	**Model 2- community**	**Model 3- macro**	**Model 4- full**
	VIF	VIF	VIF	VIF
Estimated race/ethnicity	1.13			1.52
Estimated Sex	1.12			1.13
Organizational plaintiff	1.14			1.15
Sale record-same year	1.18			1.19
Sale record-next year	1.10			1.11
Sale record-last year	1.02			1.03
Permit approved-same year	1.11			1.14
Permit approved-next year	1.10			1.12
Permit approved-last year	1.11			1.14
Property taxable value	1.33			1.44
Landuse deed	1.68			2.16
Law firm	1.42			1.65
answer	1.03			1.03
LIHTC property	1.09			1.12
Public housing	1.04			1.04
median household income		-		-
median property value		2.73		2.77
# of Sales		2.30		2.42
# of Permits		2.10		2.20
% of non-citizen		-		-
% of foreign born		-		-
% of White		3.28		3.64
% of Black		2.68		2.81
% of Native		1.29		1.28
% of Hispanic		2.35		2.50
% of Asian		1.41		1.47
% of PI		1.34		1.40
% of Below poverty rate		4.24		-
% of High school and more		-		-
% of Bachelor and more		3.81		3.87
% of Unemployed		2.37		2.38
% of Rental properties		4.20		4.40
% of Vacant		1.72		1.62
% of Rental burden		2.40		2.35
Community SES		-		-
GDP			-	-
Unemployment Rate			-	-
Fair housing price 2b			7.37	1.40

House price index	8.65	-
Fair housing price 2b *House price index	1.58	-
Estimated race/ethnicity*% of White		1.63

Note: 1) The cut-off is 5 for model 5, model 6 and model 8, but the cut-off is 10 for model 7 due to less variable. 2) The “-” symptom means the variable is not included in the model due to multicollinearity.

Table A2: Variance Inflation Factor (VIF) table for outcome of sheriff eviction at different factor level.

Variable	Model 1-individual	Model 2-community	Model 3-macro	Model 4- full
	VIF	VIF	VIF	VIF
Estimated race/ethnicity	1.13			1.56
Estimated Sex	1.12			1.13
Organizational plaintiff	1.15			1.17
Sale record-same year	1.13			1.13
Sale record-next year	1.09			1.09
Sale record-last year	1.02			1.02
Permit approved-same year	1.09			1.11
Permit approved-next year	1.08			1.09
Permit approved-last year	1.10			1.12
Property taxable value	1.26			1.39
Landuse deed	1.50			1.95
Law firm	1.34			1.56
answer	1.03			1.03
LIHTC property	1.10			1.13
Public housing	1.04			1.04
median household income		-		-
median property value		2.85		2.85
# of Sales		2.32		2.39
# of Permits		1.98		2.02
% of non-citizen		-		-
% of foreign born		-		-
% of White		3.13		3.49
% of Black		2.57		2.66
% of Native		1.31		1.31
% of Hispanic		2.36		2.52
% of Asian		1.35		1.41
% of PI		1.37		1.46
% of Below poverty rate		4.46		4.76
% of High school and more		-		-
% of Bachelor and more		3.72		3.82
% of Unemployed		2.42		2.79
% of Rental properties		4.23		5.05
% of Vacant		1.71		1.74
% of Rental burden		2.29		2.32
Community SES		-		-
GDP			-	-
Unemployment Rate			4.73	-

Fair housing price 2b	9.10	1.42
House price index	9.81	-
Fair housing price 2b *House price index	2.26	-
Estimated race/ethnicity*% of White		1.65

Note: 1) The cut-off is 5 for model 5, model 6 and model 8, but the cut-off is 10 for model 7 due to less variable. 2) Although the VIF for "% of Rental properties" is above 5, I decide to keep it as it is slightly above 5. 3) The "-" symptom means the variable is not included in the model due to multicollinearity.

Table A3: Marginal Effects table for outcome of dismissal or judgement at full model

Variable	Value	AME	SE	Z	P	CI
Estimated race/ethnicity (White)						
	Asian	0.02	0.01	1.78	0.08	[0,0.04]
	Black	0.00	0.01	-0.08	0.94	[-0.02,0.01]
	Hispanic	0.02	0.01	2.20	0.03	[0,0.04]
Estimated Sex(Male)						
			0.00	0.00	-	[-0.01,0.01]
	Female	0.98			0.54	
	Unknown	-0.01	0.01	-0.59	0.56	[-0.03,0.01]
Organizational plaintiff(No)		0.02	0.01	3.72	0.00	[0.01,0.03]
Sale record-same year(No)		0.05	0.01	7.6	0.00	[0.04,0.07]
Sale record-next year(No)		0.01	0.01	1.15	0.25	[-0.01,0.02]
Sale record-last year(No)		0.01	0.01	1.34	0.18	[-0.01,0.03]
Permit approved-same year(No)		0.00	0.01	-0.37	0.71	[-0.03,0.02]
Permit approved-next year(No)		0.02	0.01	1.26	0.21	[-0.01,0.04]
Permit approved-last year(No)		0.01	0.01	1.09	0.28	[-0.01,0.04]
Property taxable value		0.02	0.00	9.24	0.00	[0.02,0.03]
Landuse deed(Single family)						
	mobile home	-0.04	0.01	-2.88	0.00	[-0.06,-0.01]
	multi-units-family	-0.05	0.01	-6.87	0.00	[-0.07,-0.04]
	other_org	0.05	0.01	3.23	0.00	[0.02,0.07]
	unknown	0.03	0.03	1.09	0.27	[-0.02,0.08]
Law firm (Other)						
	Donaldeallen	0.01	0.01	0.85	0.39	[-0.01,0.03]
	Everettholum	-0.02	0.01	-2.15	0.03	[-0.05,0]
	Jordan	-0.08	0.01	-9.65	0.00	[-0.1,-0.06]
	Mattjmclain	0.05	0.01	4.02	0.00	[0.02,0.07]
	Puckett&Redford	0.11	0.01	13.40	0.00	[0.1,0.13]
answer (No)		-0.01	0.01	-1.59	0.11	[-0.02,0]
LIHTC property (No)		-0.01	0.01	-0.49	0.62	[-0.03,0.02]
Public housing (No)		0.03	0.03	0.97	0.33	[-0.03,0.08]
median household income		-	-	-	-	-
median property value		0.01	0.00	2.21	0.03	[0,0.02]
# of Sales		0.00	0.00	-0.28	0.78	[-0.01,0.01]
# of Permits		0.01	0.00	2.14	0.03	[0,0.01]
% of non-citizen		-	-	-	-	-
% of foreign born		-	-	-	-	-
% of White		0.00	0.00	0.34	0.73	[-0.01,0.01]
% of Black		-0.01	0.00	-2.63	0.01	[-0.02,0]

% of Native		0.00	0.00	-0.62	0.53	[-0.01,0]
% of Hispanic		-0.01	0.00	-1.33	0.18	[-0.01,0]
% of Asian		0.00	0.00	1.01	0.31	[0,0.01]
% of PI		-0.01	0.00	-2.09	0.04	[-0.01,0]
% of Below poverty rate		-	-	-	-	-
% of High school and more		-	-	-	-	-
% of Bachelor and more		0.00	0.00	0.38	0.71	[-0.01,0.01]
% of Unemployed		-0.01	0.00	-3.84	0.00	[-0.02,-0.01]
% of Rental properties		0.02	0.00	4.15	0.00	[0.01,0.03]
% of Vacant		0.00	0.00	-1.01	0.31	[-0.01,0]
% of Rental burden		-0.01	0.00	-1.94	0.05	[-0.01,0]
Community SES(low)						
	Middle	-	-	-	-	-
	High	-	-	-	-	-
GDP		-	-	-	-	-
Unemployment Rate		-	-	-	-	-
Fair housing price 2b		-0.02	0.00	-5.44	0.00	[-0.02,-0.01]
House price index		-	-	-	-	-

Note: The "-" symptom means the variable is not included in the model due to multicollinearity.

Figure A6: Marginal Effects graph for outcome of dismissal or judgement at full model

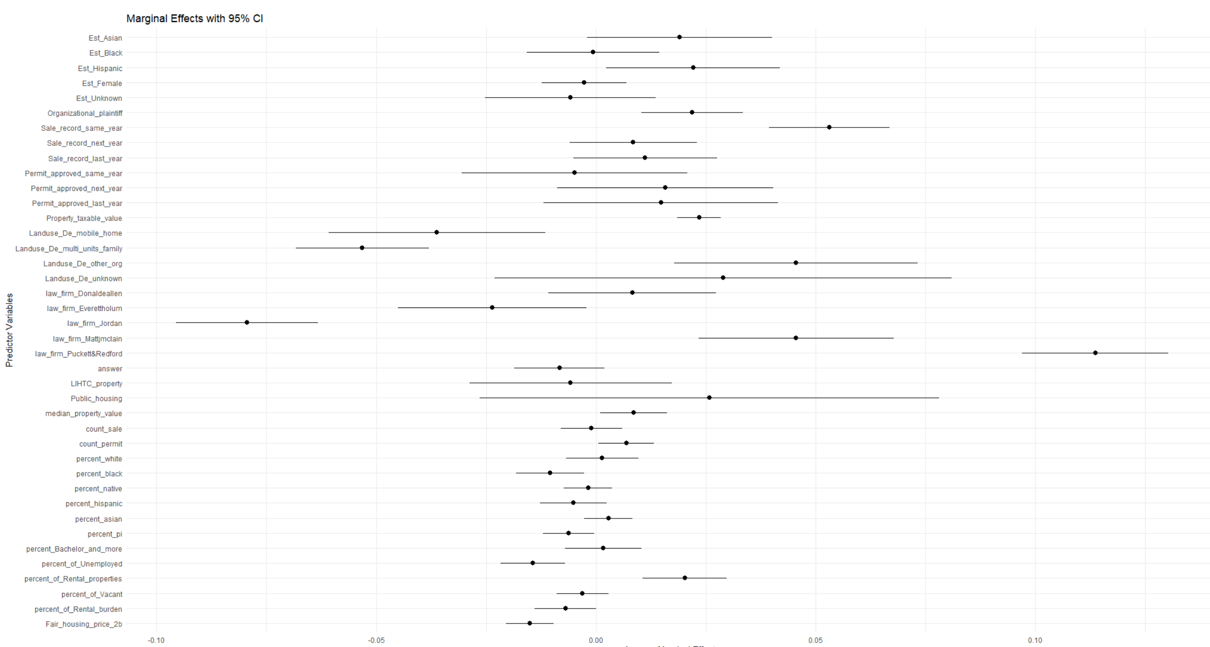


Table A4: Marginal Effects table for outcome of sheriff eviction at full model

Variable	Value	AME	SE	Z	P	CI
Estimated race/ethnicity (White)						
	Asian	-0.02	0.01	-1.25	0.21	[-0.04,0.01]
	Black	0.01	0.01	0.81	0.42	[-0.01,0.02]
	Hispanic	-0.03	0.01	-2.60	0.01	[-0.05,-0.01]
Estimated Sex(Male)						
	Female	0.00	0.01	-0.44	0.66	[-0.01,0.01]
	Unknown	-0.02	0.01	-1.43	0.15	[-0.04,0.01]
Organizational plaintiff(No)		0.02	0.01	3.72	0.00	[0.01,0.03]
Sale record-same year(No)		0.04	0.01	4.82	0.00	[0.02,0.06]
Sale record-next year(No)		0.02	0.01	1.99	0.05	[0.00,0.04]
Sale record-last year(No)		0.00	0.01	-0.26	0.80	[-0.02,0.02]
Permit approved-same year(No)		-0.02	0.02	-1.01	0.31	[-0.05,0.02]
Permit approved-next year(No)		0.02	0.02	1.41	0.16	[-0.01,0.05]
Permit approved-last year(No)		0.03	0.02	1.58	0.11	[-0.01,0.06]
Property taxable value		-0.01	0.00	-2.38	0.02	[-0.02,0.00]
Landuse deed(Single family)						
	mobile home	0.05	0.02	3.00	0.00	[0.02,0.08]
	multi-units-family	-0.01	0.01	-0.82	0.41	[-0.02,0.01]
	other_org	0.01	0.02	0.70	0.48	[-0.02,0.04]
	unknown	0.00	0.03	-0.05	0.96	[-0.06,0.06]
Law firm (Other)						
	Donaldeallen	-0.03	0.01	-2.67	0.01	[-0.05,-0.01]
	Everettholum	-0.01	0.01	-0.48	0.63	[-0.03,0.02]
	Jordan	0.01	0.01	0.67	0.50	[-0.01,0.03]
	Mattjmclain	0.01	0.01	0.39	0.70	[-0.02,0.03]
	Puckett&Redford	0.01	0.01	0.71	0.48	[-0.01,0.03]
answer (No)		-0.01	0.01	-1.87	0.06	[-0.02,0]
LIHTC property (No)		0.06	0.01	4.25	0.00	[0.03,0.08]
Public housing (No)		0.01	0.03	0.49	0.62	[-0.04,0.07]
median household income		-	-	-	-	-
median property value		0.00	0.00	0.92	0.36	[0,0.01]
# of Sales		0.00	0.00	0.75	0.45	[-0.01,0.01]
# of Permits		-0.01	0.00	-2.31	0.02	[-0.02,0]
% of non-citizen		-	-	-	-	-
% of foreign born		-	-	-	-	-

% of White	-0.02	0.01	-3.79	0.00	[-0.03,-0.01]
% of Black	0.00	0.00	0.49	0.63	[-0.01,0.01]
% of Native	0.00	0.00	-0.28	0.78	[-0.01,0.01]
% of Hispanic	-0.02	0.00	-3.55	0.00	[-0.02,-0.01]
% of Asian	0.00	0.00	-1.43	0.15	[-0.01,0]
% of PI	0.00	0.00	-0.77	0.44	[-0.01,0]
% of Below poverty rate	0.03	0.01	4.37	0.00	[0.01,0.04]
% of High school and more	-	-	-	-	-
% of Bachelor and more	0.01	0.01	2.44	0.01	[0,0.02]
% of Unemployed	-0.02	0.00	-4.72	0.00	[-0.03,-0.01]
% of Rental properties	0.01	0.01	1.64	0.10	[0,0.02]
% of Vacant	0.01	0.00	2.04	0.04	[0,0.01]
% of Rental burden	0.01	0.00	3.00	0.00	[0,0.02]
Community SES(low)					
Middle	-	-	-	-	-
High	-	-	-	-	-
GDP	-	-	-	-	-
Unemployment Rate	-	-	-	-	-
Fair housing price 2b	0.05	0.00	15.86	0.00	[0.04,0.06]
House price index	-	-	-	-	-

Figure A7: Marginal Effects graph for outcome of sheriff eviction at full model

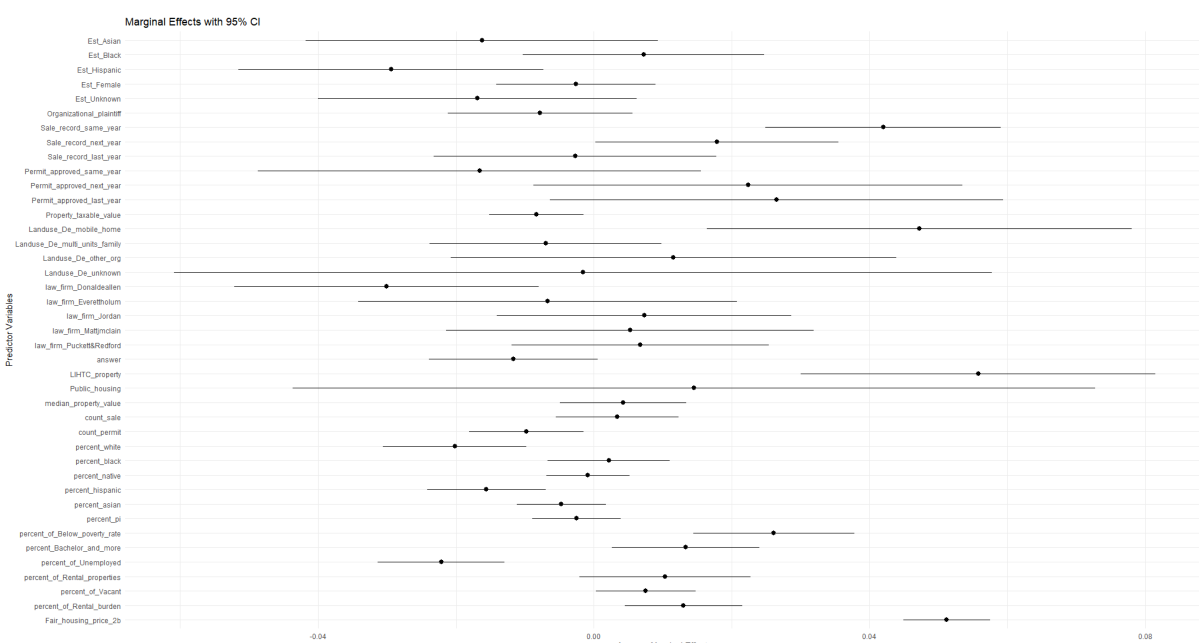


Figure A8: Feature Importance of Random Forest model for outcome of dismissal or judgement

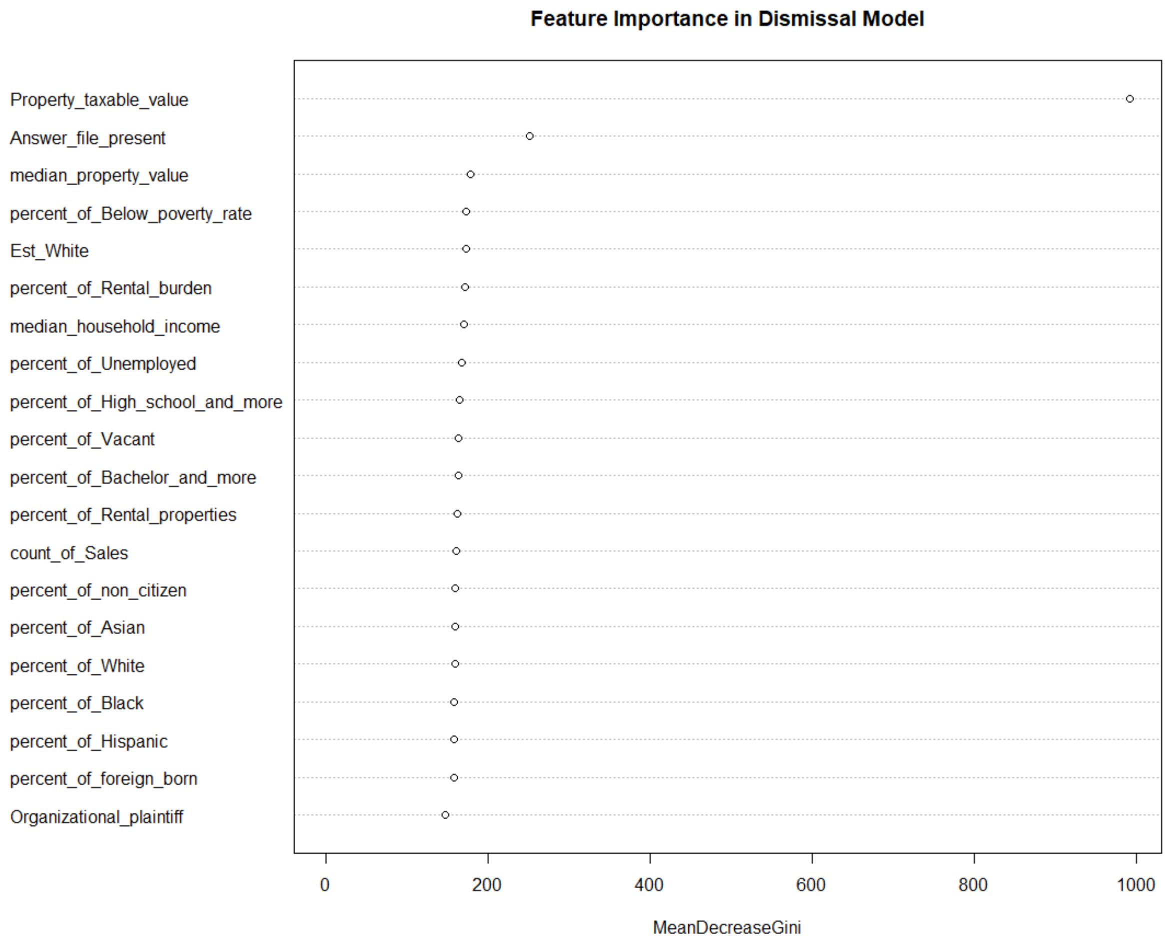
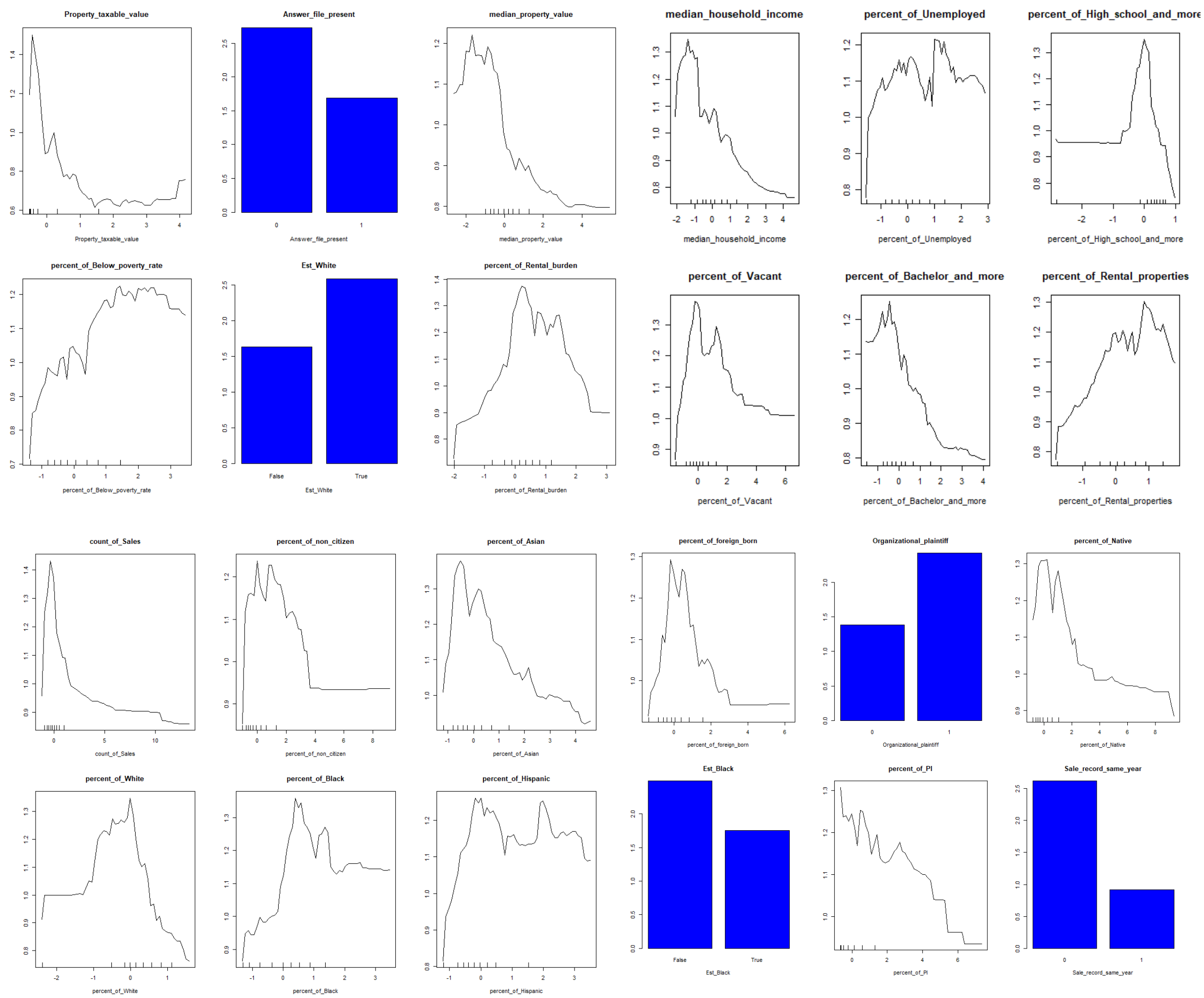
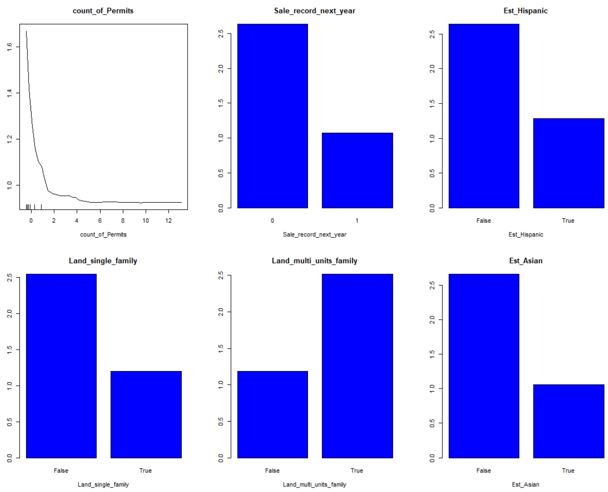
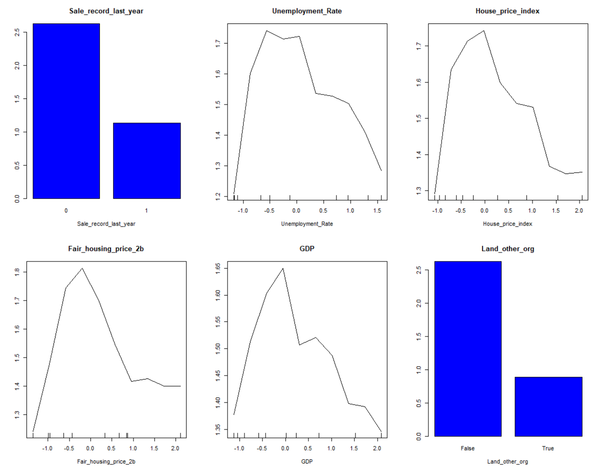
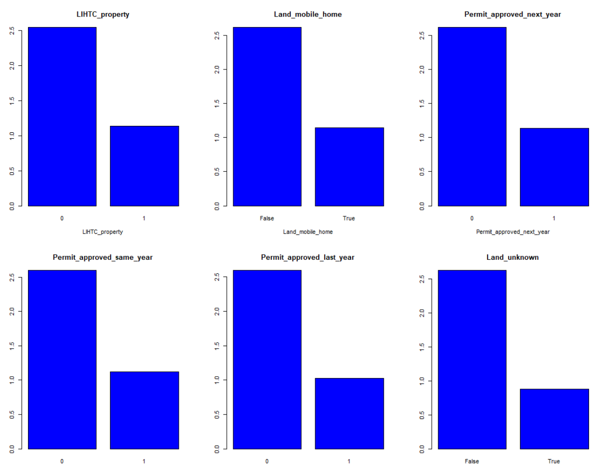


Figure A9: Partial dependence plot (PDP) for outcome of dismissal or judgment




count_of_Permits
Sale_record_next_year
Est_Hispanic
Land_single_family
Land_multi_units_family
Est_Asian


Sale_record_last_year
Unemployment_Rate
House_price_index
Fair_housing_price_2b
GDP
Land_other_org


LIHTC_property
Land_mobile_home
Permit_approved_next_year
Permit_approved_same_year
Permit_approved_last_year
Land_unknown

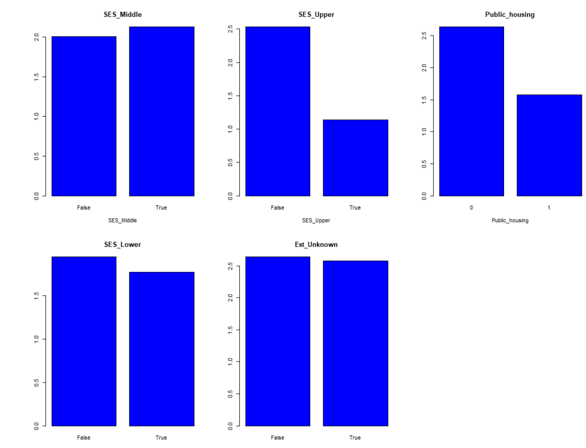

SES_Middle
SES_Upper
Public_housing
SES_Lower
Est_Unknown

Figure A10: Feature Importance of Random Forest model for outcome of sheriff eviction

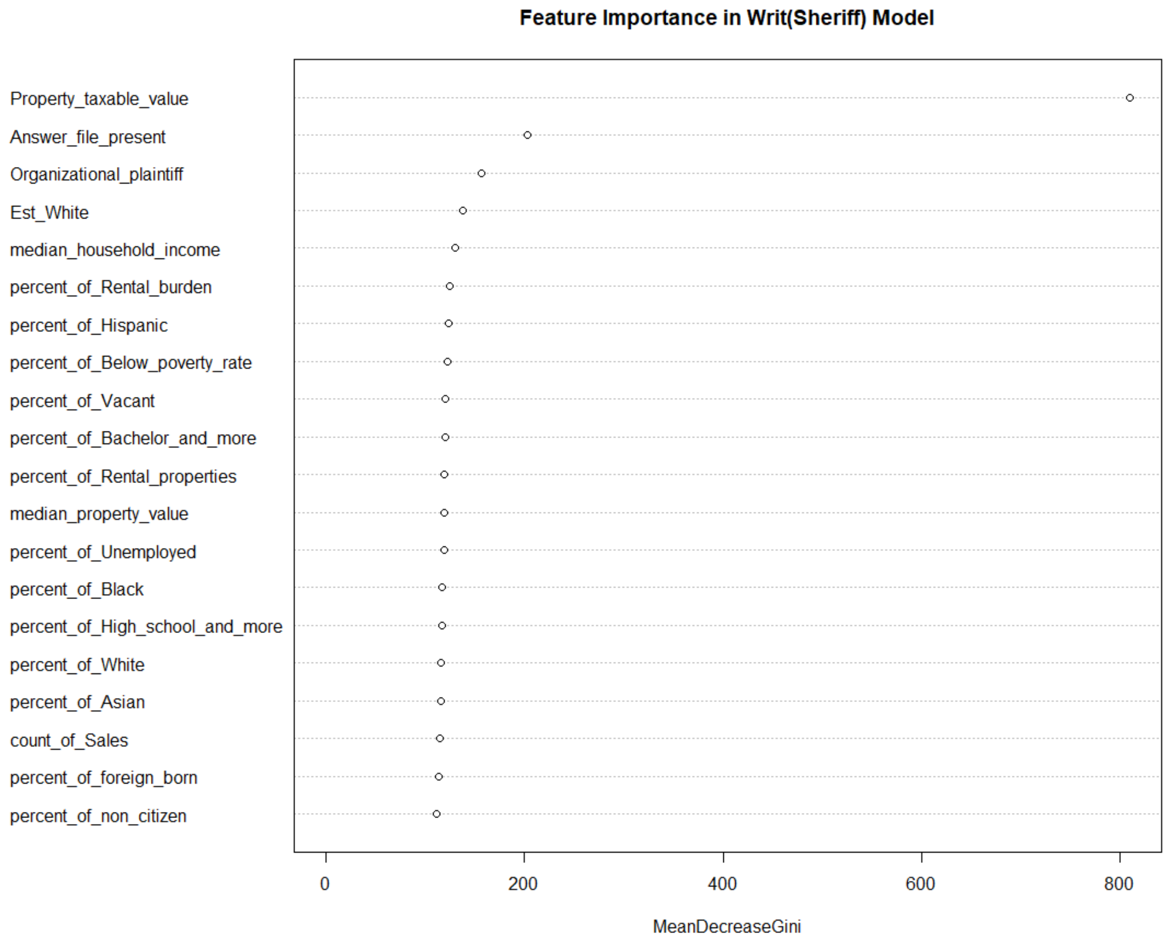
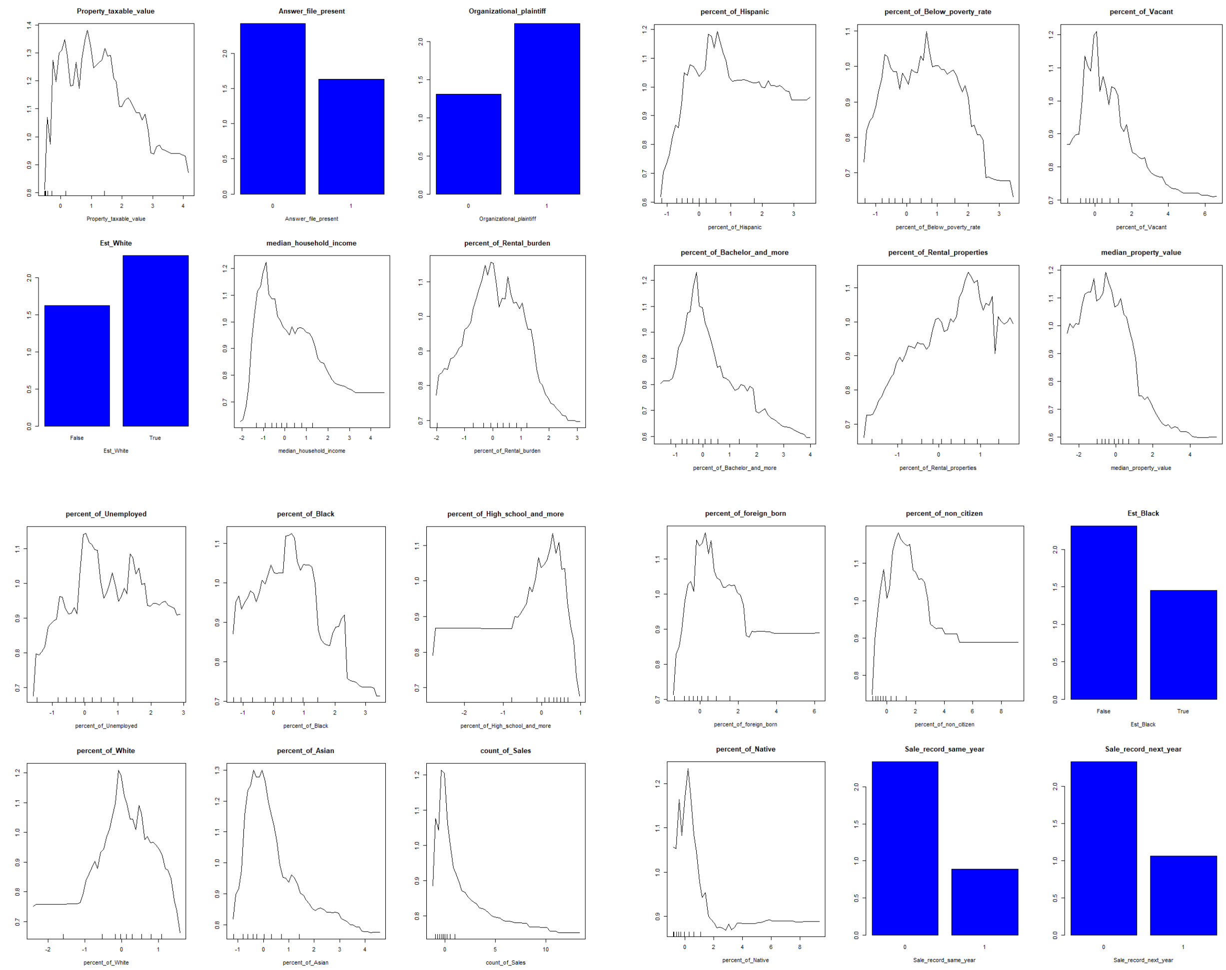
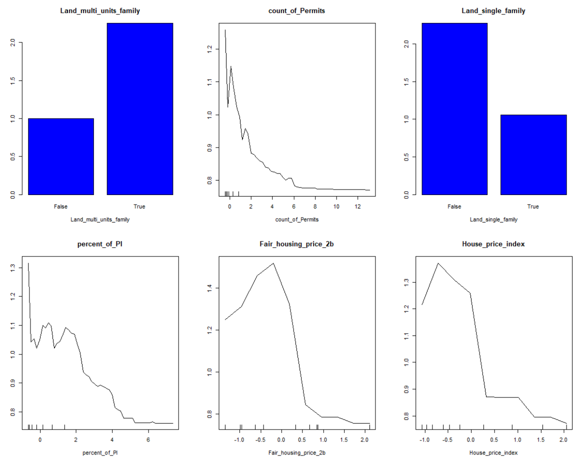
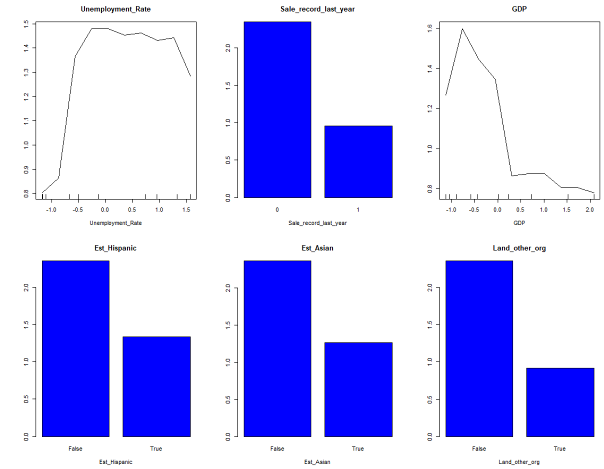
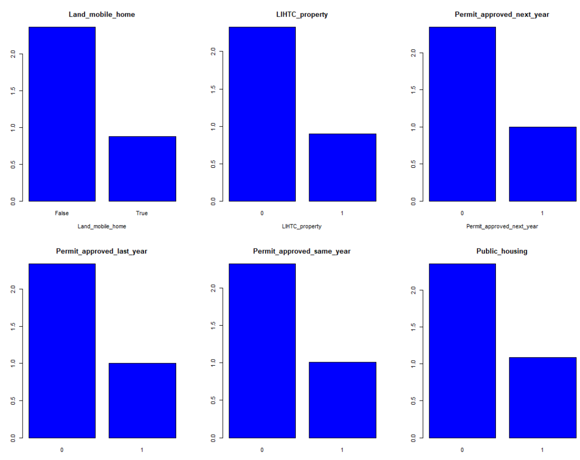


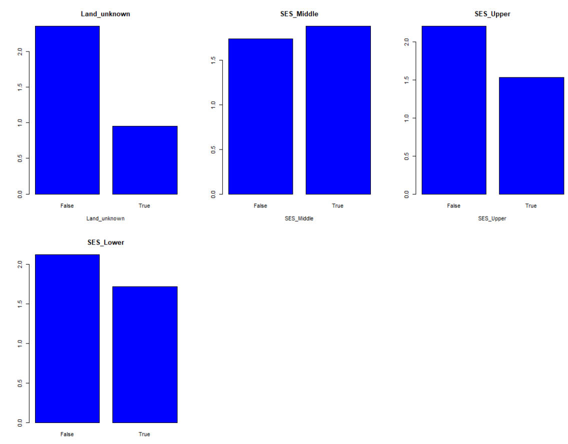
Figure A11: Partial dependence plot (PDP) for outcome of sheriff eviction




Land_multi_units_family
False
True
Land_multi_units_family
count_of_Permits
count_of_Permits
Land_single_family
False
True
Land_single_family
percent_of_PI
percent_of_PI
Fair_housing_price_2b
Fair_housing_price_2b
House_price_index
House_price_index


Unemployment_Rate
Unemployment_Rate
Sale_record_last_year
0
1
Sale_record_last_year
GDP
GDP
Est_Hispanic
False
True
Est_Hispanic
Est_Asian
False
True
Est_Asian
Land_other_org
False
True
Land_other_org


Land_mobile_home
False
True
Land_mobile_home
LIHTC_property
0
1
LIHTC_property
Permit_approved_next_year
0
1
Permit_approved_next_year
Permit_approved_last_year
0
1
Permit_approved_last_year
Permit_approved_same_year
0
1
Permit_approved_same_year
Public_housing
0
1
Public_housing


Land_unknown
False
True
Land_unknown
SES_Middle
False
True
SES_Middle
SES_Upper
False
True
SES_Upper
SES_Lower
False
True
SES_Lower