Data Shop

Data Shop, a department of Cityscape*, presents short articles or notes on the uses of data in housing and urban research. Through this department, the Office of Policy Development and Research introduces readers to new and overlooked data sources and to improved techniques in using well-known data. The emphasis is on sources and methods that analysts can use in their own work. Researchers often run into knotty data problems involving data interpretation or manipulation that must be solved before a project can proceed, but they seldom get to focus in detail on the solutions to such problems. If you have an idea for an applied, data-centric note of no more than 3,000 words, please send a one-paragraph abstract to* chalita.d.brandly@hud.gov *for consideration.*

A Statistical Machine Learning Approach to Identify Rental Properties From Public Data Sources

Daniel Kuhlmann
University of Arizona

Jane Rongerude
Iowa State University

Lily Wang
George Mason University

GuanNan Wang
College of William and Mary

Introduction

For academic researchers and practitioners alike, identifying individual rental properties can be incredibly useful but is often difficult due to insufficient and incomplete data. Although some cities have ordinances that require residential rental property owners (RRPOs) to register their properties, the availability and completeness of these registries vary dramatically from place to place. In places without rental registries, tax assessor data can provide some information but often not enough to clearly distinguish residential rentals from owner-occupied units and other commercial properties. As part of a larger project surveying RRPOs, the project team developed and tested statistical (machine) learning methods and predictive models to identify potential rental properties from existing data sources. This article describes the process for creating the models, suggests potential applications for the methods, and discusses how researchers and practitioners can use these models and methods in their work.

Background

Many policy and research contexts exist in which the ability to identify individual rental properties and the owners of those properties are useful. For example, for this project, the team needed to generate a sample of rental property owners across multiple jurisdictions for a survey-based study investigating RRPO characteristics and behaviors. Similarly, identifying who owns rental properties can help researchers track investment behavior, understand market dynamics, and study other owner-related housing questions. Identifying locations and addresses of rental properties is useful for researchers when tracking displacement and tenant mobility or contacting tenants for surveys and interviews. Likewise, differentiating rentals from owner-occupied properties can be useful for those studying property and tenant outcomes, such as property neglect and evictions.

In addition, being able to identify rental properties may be useful for practitioners and local government officials. For example, local housing officials might want to identify likely rental properties in their jurisdictions to track whether RRPOs comply with rental licensing and inspection laws. Similarly, in places without rental licensing requirements that need full disclosure of owner data, practitioners might use these methods to distribute program information and other resources to tenants or RRPOs. Local jurisdictions can benefit from knowledge about the individuals or entities investing in their rental housing markets, the location of those entities, and whether their holdings are expanding or consolidating in particular neighborhoods. The COVID-19 pandemic and the subsequent housing policy responses, such as the eviction moratoria and Emergency Rental Assistance program, demonstrated that having basic contact information for RRPOs is necessary during a catastrophe or disaster. To distribute pandemic-related rental assistance efficiently, local governments needed inventories of rental properties within their jurisdictions and contact information for RRPOs. However, in most jurisdictions in the United States, this information is not readily available.

This article describes the project team's method for identifying potential rental properties from existing data sources. Exhibit 1 summarizes different methods and data sources considered for the study. It is important to note that each project described here has unique goals, which understandably influence the methodological choice. This article does not aim to provide an exhaustive accounting of the various methods researchers have used to identify individual rental properties; instead, the project team presents a new methodology developed for its specific research needs.

Exhibit 1

Methods for Identifying Rental Property (1 of 2)

Primary Data Source	Method Description	Strengths	Weaknesses	Example
Rental license records	Use rental licensing data to identify rental properties.	In places with well-enforced rental licensing programs, rental registries are likely the best and most complete source of information on individual rental properties.	Not all jurisdictions have ordinances requiring rental licenses. Further, the types of properties and property owners who must obtain licenses vary from place to place. Even in jurisdictions that require regulations, there is variation in how well jurisdictions enforce rental licensing.	Kuhlmann et. al. (2022) used rental registry data to contact property owners in Minneapolis, MN, for an online survey.
Property assessor administrative files	Identify patterns in property assessment records, such as mismatched tax and situs addresses, use of homestead exemptions, and property use data to flag likely rental properties.	Most uniform source of data available across the country for information on property ownership. Relatively easy to identify certain types of rental properties, such as large multifamily rental buildings.	In places without rental licensing laws, assessment records rarely identify rental properties explicitly. Assessment data vary from county to county, making cross-jurisdiction analyses difficult. Likely misses some rental properties and mis-identifies certain types of ownership forms as rentals.	Travis (2019) compared identified rental properties based on mis-matched tax and situs addresses, non-individual owner names (LLCs, etc.), and properties with multiple units to identify likely rental properties in his project examine the association between LLC ownership and property upkeep.
Public program participation	Identify rental properties based on whether either the landlord or tenant participated in a public program.	Efficient method to identify a targeted subset of the rental housing stock. For example, using program participation is a direct method to identify low-cost rentals, properties with physical deficiencies, and cost-burdened tenants.	These data are not always publicly available. Only captures tenants and owners who submitted applications to public programs. As a result, they are unlikely to be representative of the rental housing stock in a particular place.	De la Campa, Reina, and Herbert (2021) used applications for the Emergency Rental Assistance program to identify landlords in Los Angeles, CA for a survey.

Exhibit 1

Methods for Identifying Rental Property (2 of 2)

Primary Data Source	Method Description	Strengths	Weaknesses	Example
Online listings	Use web-scrapping techniques to pull detailed rental listing data from websites such as Craigslist.	One of the few methods that provides real-time data on asking rents, utilities, vacancy, and other property information not generally captured in administrative data.	Requires technical expertise to set up web scrapping programs and can be computationally taxing. Not all properties are listed on online platforms and increasingly sites block web scrapping programs. Only useful for generating point-in-time data on rental listings; less useful to capture information on the full housing stock.	Boeing and Waddell (2017) and Boeing et al. (2021) created web-scrapping programs to pull rental listing data from Craigslist.
Proprietary date sources	Use data from proprietary sources to identify rental properties and property owners. Possible sources are from third-party data companies, online listing websites, and private research firms.	Efficient method to identify generally representative samples of rental properties and owners.	Access to these data is limited and expensive when available. These data sources tend to be focused on certain geographies and segments of the rental housing market, and are thus less useful for studies interested in either all the rental properties or those that attempt to draw a representative sample of rental units or owners.	Decker (2021) used contact investor contact information from the residential investment property listing platform Roofstock to draw a national sample of residential landlords. Raymond et. al. (2021) and others have relied on the real estate data company CoreLogic, which provides standardized assessment records and has internal (although unvetted) methods to identify likely investor properties.
SEC fillings	Use public filings with the SEC to identify various corporate entities associated with a particular corporation. Then match these corporate names with assessment and other public data to identify rental holdings of large, corporate owners and REITS.	Possible to identify rental portfolios nationally.	Only applicable for entities subject to SEC reporting requirements. Useful for describing the portfolios of specific corporate owners but misses the majority of rental properties owned by non-corporate entities.	Colburn, Walter, and Pfeiffer (2021) examined SEC filings study ownership and investment patterns of large, publicly traded investors in single-family rental properties.

LLC = limited liability company. REITS = real estate investment trusts. SEC = U.S. Securities and Exchange Commission.

From a review of the previous research on rental property ownership, the team identified the following questions as those important to consider when selecting a method for identifying rental properties or rental property owners, or both.

- Is the project's goal to identify all rental properties in each jurisdiction or a sample of rentals?
- How sensitive is the project to possible misidentification? Relatedly, is the project particularly sensitive to either false positives (identifying nonrentals as rentals) or false negatives (failing to identify rentals)?
- Does the project have funding to purchase third-party data or access to proprietary data sources?
- Is the project focused on a single jurisdiction, or should the method be applicable to multiple jurisdictions and data sources?

Case Study

In the spring of 2020, the team launched a new project to examine the question, "What is influencing the decisionmaking of RRPOs during the COVID-19 pandemic?" The decisions that RRPOs make during a disaster affect not only their tenants' short-term housing stability but also the composition and stability of the rental housing stock. At the time, little information existed on how the ongoing pandemic and subsequent policy responses affected the businesses of RRPOs. To study this question, the team collected data about the characteristics and behaviors of RRPOs to better understand who they were and how their decisions contributed to rental housing stability. Time, cost, and safety considerations required gathering data through an e-mail survey; however, the team quickly ran into difficulty trying to identify the RRPOs and obtain their contact information. To address this obstacle, in the first year of the study, the team limited the scope to four cities: Cleveland, Des Moines, Minneapolis, and Tampa. Each of these cities had a rental registry the team could use to identify a sample of rental properties, even when a corporatized ownership structure, such as a limited liability company (LLC), obscured the name of an individual owner.

Even after limiting the study scope to cities with available registries, the team found that the completeness and accessibility of these registries varied from city to city. For example, the Minneapolis rental registry is easily accessible online and includes inspection ratings and contact information for property owners. Compared with census records, it includes around 95 percent of the city's rental units. In contrast, the Des Moines registry was not publicly available and required a formal request to access from the city's inspections department. Further, the team found that the Des Moines registry had incomplete or missing contact information for many property owners and included less than two-thirds of the city's rental units. The team's data dashboard has more information about each city's rental registry (Rongerude et al., 2021).

In 2022, the project team expanded this study to examine how a broader range of natural disasters affect the businesses of RRPOs, adding five disaster-prone cities: Austin, Dallas, Houston, Miami, and New Orleans. In the expanded study, the team also shifted from an e-mail to a primarily physical mail survey with an online option. Because only some, not all, of these new cities had rental registries, the team was concerned about potential bias from surveying only owners who comply with registration requirements and, thus, developed a method to identify likely rental properties across a diverse set of cities. After switching to a mail-based survey, the team was particularly interested in developing a method that minimized misidentification, because unlike the team's previous study, now a marginal cost was associated with sending survey invitations. To maximize the probability that the survey reached actual RRPOs, prior to distribution, the team considered a methodological concern: how to identify rental properties from data sources available across multiple jurisdictions while minimizing the number of owner-occupied properties misidentified as rentals.

In response, the interdisciplinary team, including planners, statisticians, and data scientists, undertook the modeling exercise this article describes. The team first reviewed methods that other researchers have used to identify rental properties, then created a new modeling technique to address its specific research needs. The rest of this article describes the data the team's modeling efforts used, details the modeling process and how the accuracy of each model was assessed, and ends with a brief discussion of the limitations of the models and other potential applications.

Data

To build the predictive models, the project team first searched for variables that correlate with housing tenure and were available and uniform across the nine jurisdictions in this study. For example, some counties report whether properties in their assessment rolls claim a homestead exemption. Although claiming an exemption is likely a good indication that the property is owner-occupied, not all counties in the study report this field. Using this variable to build a model in Minneapolis, for example, likely would improve its predictive power, but doing so would prevent the model from applying to the cities lacking such indicators.

The data in this study come from two primary sources: Regrid.com, a data services company that collects and standardizes parcel, transaction, and assessment administrative data, and 5-year American Community Survey (ACS) estimates. The project team elected to purchase standardized assessment data rather than collect the data for two reasons. The first is one of expediency. Collecting and combining assessment records is a time-consuming task and one that is made particularly difficult by the scope of this analysis. Because this study includes nine large U.S. cities, several of which contain multiple counties (for example, five separate counties fall within the municipal boundary of Dallas, Texas), building this dataset would require contacting 19 different assessor offices, collecting their respective records, and formatting each county's data to create consistency across the sample. The second reason to purchase the standardized assessments is that Regrid.com, in addition to standardizing the assessment records, also combines parcel-level data with several other data sources that could be useful predictors of property tenure, such as occupancy indicators from the U.S. Postal Service (USPS).

Exhibit 2 lists all the variables included in the modeling process. In addition to standard assessment fields, the process used several variables from secondary data sources, such as a vacancy indicator and an indicator from the USPS of whether mail is deposited at the street level or in a highrise building. The project team also used several proprietary measures Regrid.com, including a count of all primary and secondary addresses at the parcel (a more consistent measure of unit count across the cities in this study) and Regrid.com's calculation of parcel sizes, building footprints, and a count of structures on the parcel. In addition to these pre-formatted fields included in the Regrid.com files, the project team also created several new variables. When reviewing the literature, the team identified two measures that serve as strong indicators of property tenure. The first is a measure of whether a corporate entity owns the property, such as a corporation, limited partnership, or LLC.[1] To create this variable, the team wrote code that searches for regular expressions matching common nonindividual ownership forms. This process was iterative, because none of the counties in the study standardized their ownership fields, resulting in small variations in ownership names.[2]

Exhibit 2

Model Variables

Variable	Type	Level	Source
Total Parcel Value	Numerical integer; dollars	Parcel	County Assessor's office; Regrid.com
Do the property situs and taxpayer billing address match?	Binary; 1 = same address	Parcel	County Assessor's office; Regrid.com; author's calculations
Does the owner name contain a corporate indicator? (e.g., LLC, LP, INC)	Binary; 1 = corporate indicator found	Parcel	County Assessor's office; Regrid.com; author's calculations
Postal delivery type (street versus highrise)	Binary; 1 = corporate indicator found	Parcel	USPS; Regrid.com
Is the property a residential address?	Binary; 1 = residential	Parcel	USPS; Regrid.com
Is the property vacant?	Binary; 1 = vacant	Parcel	USPS; Regrid.com
How many primary and secondary addresses are at the parcel?	Count	Parcel	USPS; Regrid.com
Building footprint	Numerical integer; square feet	Parcel	Regrid.com
Number of structures on parcel	Count	Parcel	Regrid.com
Share of housing units owner occupied	Percent	Census Tract	2015–2019 5-year ACS estimates
Median household income	Numerical integer; dollars	Census Tract	2015–2019 5-year ACS estimates

ACS = American Community Survey. INC = incorporated. LLC = limited liability company. LP = limited partnership. USPS = U.S. Postal Service.

The second measure is a field indicating whether the taxpayer's billing address differs from the property situs address. To create this field, the project team first parsed the billing and situs addresses, pulling each of the address components (house number, unit, street prefix direction, and so on) into separate fields using the R package "PostmastR," which parses address formats and also standardizing common spelling variants in addresses (for example, "South," "So," and "S"). The team then looked for matches in the house numbers and street names (excluding directionals and suffixes) of billing and situs addresses. The team experimented with more strict match conditions (for example, requiring that all address components match) but noticed enough instances of slight variations in ultimately matching addresses to justify a more lenient matching criterion.

[1] The project team considered including but ultimately excluded "trust" and related terms from the corporate entity search. This term often picked up living trusts, which can be used for ownership of rental investments, but are also for owner-occupiers during the estate planning process.

[2] For example, the code checks for multiple variants of "limited liability corporation," including abbreviations with and without punctuation and variations in spelling.

Finally, the team used several census tract-level variables from the 2018 ACS; using census tracts (geographies containing between 1,200 and 8,000 people) rather than the more granular census boundaries ensured consistency across the sample. The U.S. Census Bureau suppresses data for smaller geographies, particularly when the reported tabulations could potentially be used to identify individual households. The team included two estimates from the ACS in its models: the share of renter-occupied housing units and the median household income.

Method

The goal is to create a predictive model to accurately identify potential RRPOs and use the model to guide the selection of survey participants. Correctly identifying the potential rental properties required an accurate binary classifier (that is, a model that estimates a yes or no outcome), and the team considered five popular classification methods. Exhibit 3 briefly describes each method, along with their relative advantages and disadvantages.

Exhibit 3

Modeling Techniques

	Logistic	Decision Tree	Naive Bayes	kNN	(partially linear) GAM
Sensitivity	0.6701	0.7444	0.6378	0.7304	0.6983
Specificity	0.9487	0.942	0.9273	0.9481	0.9463

GAM = generalized additive model. kNN = k-nearest neighbors.
Source: Author's Calculations.

To build these classification models, the project team used the variables in the combined dataset to predict if the Minneapolis rental registry database lists a property. Essentially, the team built a binary classifier using the nearly complete list of rental properties in Minneapolis, then used this model to predict likely rental properties in the sample cities that lacked complete rental registries. The outcome variable is a binary measure of whether the Minneapolis rental registry lists the property as a licensed, long-term rental property. To predict this outcome, the team included property-level variables measuring the parcel's total assessed value, counts of the number of structures and addresses on the parcel, and a measure of the total square footage of the structures on the parcel. In addition, the team included binary measures of whether the parcel's tax and situs addresses match, whether a nonindividual owns the property, whether the property is residential and vacant, and whether the postal delivery is at street level or in a highrise building. Finally, the team included two census tract-level measures from the ACS: the share of the tract's housing units that are renter-occupied and the median household income.

Model Assessment

To contrast the classifier model's prediction accuracy, the project team used the Minneapolis data to compare how often the model correctly identified properties in the city's rental registry. The team followed a fivefold cross-validation process, randomly and evenly splitting the Minneapolis data into five portions called folds. The team then chose four folds for training and one fold for testing, repeated this procedure five times, then compared the predicted value (that is, whether the model predicted that the property was a rental) with the truth (whether the observation had an active rental license). Because using the actual registry results is necessary as a basis for comparison, the team could only perform this exercise in Minneapolis.

A confusion matrix captured comparisons that include the four mutually exclusive measurements: (1) True positives—the property is a rental, and the model correctly identified it as a rental; (2) true negatives—the property is not a rental, and the model correctly identified it; (3) false positive—the property is not a rental, but the model predicted it as a rental; and (4) false negative—the property is a rental, but the model failed to predict it. From this confusion matrix, the team calculated two additional evaluation measures: specificity and sensitivity. *Specificity* is the ratio of rental properties the model correctly predicted to the total number of rental properties in the Minneapolis registry. *Sensitivity* is the ratio of properties the model correctly identified as nonrentals to the actual number of nonrentals, plus the number of properties the model identified as rentals when they were not in the rental registry. A specificity of 100 percent indicates that the model perfectly predicted all the actual rental properties, and a sensitivity of 100 percent suggests that the model did not misidentify any nonrentals as rentals. Exhibit 4 reports these metrics. In general, the prediction accuracy is similar for all the methods, with specificities ranging from 92.7 to 94.8 percent and sensitivities ranging from 63.7 to 74.4 percent.

Exhibit 4

Confusion Matrix

Method	Main Technique	Advantages	Disadvantages	Citation
Logistic Regression	Maximum likelihood estimation	Easy to implement No assumptions on distributions of classes in feature space Easy to interpret	Assumption of linearity between the response and the explanatory variables	Hosmer, Lemeshow, and Sturdivant (2013)
kNN	Euclidean distance	Easy to implement Training is fast	Testing is slow Sensitive to noise	James et al. (2013)
Decision Tree	Splitting, stopping, and pruning	No requirements of domain knowledge Easy to interpret	Unstable	Breiman (2017)
Naive Bayes	Bayes rule	Easy to implement Does not require many data	Strong assumption on the shape of data distribution	James et al. (2013)
GAM	Smoothing	Ability to model highly complex nonlinear relationships	High computational complexity	Hastie et al. (2009)

GAM = generalized additive model. kNN = k-nearest neighbors.

To predict potential rental properties in cities without rental registries, the project team focused on the binary classifier based on the logistic regression model, which provides the highest specificity. The team was particularly interested in finding models with high specificities due to the limited survey sampling budget and the need to minimize the number of negative examples (non-RRPOs) that are incorrectly classified. The team trained the classifier using the data from Minneapolis, then applied this classifier to other cities in the study to generate a sample of likely RRPOs.

As a result of this process, the model created a probability measure for each property in the database corresponding to the predicted likelihood that it is a rental property. The team then used these probabilities to create a sample in a way that minimized the risk of sending a survey invitation to a nonrental property. To generate the survey sample, the team created a list with unique owner names, keeping only one property per owner, then created the survey sample based on two factors. First, the team oversampled owners who hold properties in corporatized entities (such as LLCs), because these owners might be less likely to respond to survey requests. The team split the sample, sending two-thirds of the 2,000 invitations to corporatized owners and one-third to those who held properties as individuals. With this split in mind, the team ordered the properties by their predicated rental likelihood for each city, then created a cutoff yielding the minimum two- or one-third split between ownership types. Second, the team randomly selected owners within this group, preserving the preferred split. This process created a sample for each city that both oversampled nonindividual owners and minimized the likelihood that survey invitations reached owners of properties that were not, in fact, residential rentals.

Conclusion

Rental housing occupies a significant portion of the housing stock in U.S. metropolitan areas, yet researchers know very little about the specific characteristics of the institutional and noninstitutional entities that hold titles to those properties and determine housing supply, rents, and the conditions of both buildings and units. This gap in knowledge related to RRPOs persists partly due to a problem of insufficient and incomplete data. No comprehensive national or statewide public database exists that contains information about who owns rental properties, how to contact them, and what types of units they own. Some cities have municipal rental registry databases connected to rental unit certification and inspection programs; however, these programs are not universal, and because the databases are resource intensive to create and maintain, they are often incomplete. As a result, RRPOs can be difficult to identify and contact, a difficulty shared by researchers, housing advocates, and local governments.

In this article, the project team describes a novel method to identify potential rental properties from existing data sources. The modeling procedure is flexible, and users can tune the parameters within each model depending on the research objectives to achieve the desired specificity or sensitivity. The case study describes how these models identify owners of likely rental properties in eight cities, with either incomplete or nonexistent rental registries. Because the objective was to create a representative sample of RRPOs, the team focused on the likelihood that any given residential property was a rental unit, then created a model that allowed the team to identify a sample of property owners based on the registered owners of those units.

Although the method described here allows researchers to identify likely rental properties in cities without complete rental registries, its effectiveness ultimately depends on the accuracy and completeness of the rental registry in Minneapolis. Although this registry is one of the most complete in the country, how representative the registry is of all property types is unknown. For example, RRPOs with lower-cost and -quality properties may be more likely to evade the registration requirements and, thus, be underrepresented in the Minneapolis data. If this scenario is the case, using the Minneapolis registry to train the predictive model may systematically misidentify these properties when applied to the other sample cities.

Relatedly, whether the associations in Minneapolis used to train the model are consistent across the sample is unknown. For example, in Minneapolis, having a mismatched tax and situs address may be a strong predictor of a rental property, but the association may be weaker and less accurate in predicting rentals in cities with warmer climates and more second homes. In future iterations of this project, the team plans to partner with other cities that have relatively complete registries to expand the training dataset outside of Minneapolis. The team also plans to use the results of initial survey responses—specifically whether respondents completed the survey or responded that their property is not a rental—to build smaller training datasets in subject cities to improve the prediction accuracy of the models.

These limitations notwithstanding, this article can be useful for researchers and practitioners interested in identifying individual rental properties. Many questions regarding RRPOs and changing trends in the nation's rental markets still need to be answered. Furthermore, as the COVID-19 pandemic put in stark relief, having the ability to identify and reach out to both RRPOs and their tenants during a disaster can be hugely useful to housing officials in developing effective policy responses and distributing aid.

Acknowledgments

The research in this article is part of a larger, longitudinal project investigating the effects of residential rental property owner decisionmaking on post-disaster housing stability. This study received funding from the Polk County Housing Trust Fund and the National Science Foundation (NSF RAPID #2050264 and NSF #2139816). The authors acknowledge the following people who have contributed to the ongoing work of the project, either as co-principal investigators, senior personnel, or graduate student researchers: Elizabeth Mueller, Jake Wegmann, Taylor Cook, Biswa Das, Wanying Zhu, and Lin Quan.

Authors

Daniel Kuhlmann, Ph.D. is an assistant professor at the University of Arizona (dok@arizona.edu). Jane Rongerude, Ph.D. is an associate professor at Iowa State University (jrong@iastate.edu). Lily Wang, Ph.D. is a professor at George Mason University (lwang41@gmu.edu). GuanNan Wang, Ph.D. is an associate professor at the College of William and Mary (gwang01@wm.edu).

References

Boeing, Geoff, Max Besbris, Ariela Schachter, and John Seungmin Kuk. 2021. "Housing Search in the Age of Big Data: Smarter Cities or the Same Old Blind Spots?" *Housing Policy Debate* 31 (1): 112–126.

Boeing, Geoff, and Paul Waddell. 2017. "New Insights into Rental Housing Markets Across the United States: Web Scraping and Analyzing Craigslist Rental Listings," *Journal of Planning Education and Research* 37 (4): 457–476.

Breiman, Leo. 2017. *Classification and Regression Trees*. New York: Routledge.

Colburn, Gregg, Rebecca J. Walter, and Deirdre Pfeiffer. 2021. "Capitalizing on Collapse: An Analysis of Institutional Single-Family Rental Investors," *Urban Affairs Review* 57 (6): 1590–1625.

De la Campa, Elijah, Vincent J. Reina, and Christopher Herbert. 2021. *How are Landlords Faring During the COVID-19 Pandemic? Evidence from a National Cross-Site Survey*. Report prepared for the Joint Center for Housing Studies, Harvard University. Cambridge, MA: Harvard University.

Decker, Nathaniel. 2021. *The Uneven Impact of the Pandemic on the Tenants and Owners of Small Rental Properties*. Report prepared for the Terner Center for Housing Innovation. Berkeley: University of California.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer.

Hosmer, Jr., David W., Stanley Lemeshow, and Rodney X. Sturdivant. 2013. *Applied Logistic Regression*. New York: Wiley.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning with Applications in R*. New York: Springer.

Kuhlmann, Daniel, Jane Rongerude, Biswa Das, and Lily Wang. 2022. *Rental Property Owner Stress During the COVID-19 Pandemic: Results from a Minneapolis, MN Survey*.

Raymond, E.L., B. Miller, M. McKinney, and J. Braun. 2021. "Gentrifying Atlanta: Investor Purchases of Rental Housing, Evictions, and the Displacement of Black Residents," *Housing Policy Debate* 31 (3-5): 818–834.

Rongerude, Jane, Lily Wang, Biswa Das, and Daniel Kuhlmann. 2021. "RAPID: Keeping Shelters in Place: Understanding Landlord Decision-Making During the COVID-19 Pandemic." https://covidrental.design.iastate.edu.

Travis, Adam. 2019. "The Organization of Neglect: Limited Liability Companies and Housing Disinvestment," *American Sociological Review* 84 (1): 142–170.

Additional Reading

Awad, Mariette, and Rahul Khana. 2015. *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*. New York: APress.

Furukawa, Norrie, and Motoharu Onuki. 2022. "The Design and Effects of Short-Term Rental Regulation," *Current Issues in Tourism* 25 (20): 3245–3260.

Hastie, Trevor, and Robert Tibshirani. 1990. *Generalized Additive Models*. New York: Chapman and Hall.