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.

Toward a Cross-Platform Framework: Assessing the Comprehensiveness of Online Rental Listings

Ana Costa
Victoria Sass
Ian Kennedy
Roshni Roy
Rebecca J. Walter
Arthur Acolin
Kyle Crowder
University of Washington, Seattle

Chris Hess
Cornell University

Alex Ramiller
Sarah Chasins
University of California, Berkeley

Abstract

Research on rental housing markets in the United States has traditionally relied on national or local housing surveys. Those sources lack temporal and spatial specificity, limiting their use for tracking short-term changes in local markets. As rental housing ads have transitioned to digital spaces, a growing body of literature has utilized web scraping to analyze listing practices and variations in rental market dynamics. Those studies have primarily relied on one platform, Craigslist, as a source of data. Despite Craigslist's popularity, the authors contend that rental listings from various websites, rather than from individual ones, provide a more comprehensive picture. Using a mixed-methods approach to study listings across various platforms in five metropolitan areas, this article demonstrates considerable variation in both the types of rental units advertised and the features provided across those platforms. The article begins with an account of the birth and consolidation of online rental platforms and emergent characteristics of several selected websites, including the criteria for posting, search parameters, search results priority, and first-page search results. Visualizations are used to compare features such as the 40th percentile of rent, rent distribution, and bedroom size based on scraped data from six online platforms (Padmapper, Forrent.com, Trulia, Zillow, Craigslist, and GoSection8), 2020 Fair Market Rents, and 2019 American Community Survey data. The analyses indicate that online listing platforms target different audiences and offer distinct information on units within those market segments, resulting in markedly different estimates of local rental costs and unit size distribution depending on the platform.

Introduction

Recent years have seen a surge in housing market research, driven by the increasing availability of data from online platforms. This increased availability of publicly available information is particularly important for rental markets in which transactions do not appear in public records—contrary to the for-sale market, in which data about sales have long made possible the development of local housing price indexes. In contrast to the traditional data sources used for tracking the rental stock—Census Bureau data products that provide time-delayed summary statistics at a limited array of geographic scales—online rental listing data offers significant spatial and temporal flexibility (Boeing and Waddell, 2017).

Although the recent shift of rental listings to online platforms has created opportunities for housing market research, several issues remain unresolved, including data quality, the comprehensiveness of online housing platforms, and the accessibility of these websites relative to traditional housing search methods. For example, the effectiveness of key programs, such as the Housing Choice Voucher program, depends heavily on accurate local Fair Market Rents (FMRs). Federal data sources, such as the American Community Survey (ACS) that is used to determine FMRs, have been found to systematically underestimate local rents relative to data from online platforms such as Craigslist and Zillow, potentially limiting the accessibility of higher rent neighborhoods for voucher holders (Boeing et al., 2020; Hess et al., 2019).

This study employs a mixed-methods approach to study listings across several predominant platforms in five metropolitan areas. First, the development of online rental platforms over the past three decades is reviewed. Emergent characteristics of several selected websites are then described. Finally, rental listings are analyzed through web scraping to assess variation in listing characteristics across platforms and to compare the findings with 2020 Fair Market Rent and selected 2019 ACS estimates. This analysis finds significant variation in both the types of units advertised and the information provided across these online platforms. Selectivity in which units are advertised across different rental platforms has considerable implications for researchers and policymakers because differences in the representation of types of units or neighborhoods imply variations in measures such as rent or housing stock composition generated from those data.

Data and Methods

Rental Listing Platform and Market Selection

To better understand some of the differences between platforms, the authors selected 17 rental platforms that are the most referenced and have the greatest number of listings from the first page of Google search results. Those platforms include American Homes 4 Rent, Apartments.com, Craigslist, For Rent, GoSection8, Homefinder, Homes.com, Hotpads, Invitation Homes, Padmapper, Realtor.com, Rent.com, Rentable, Tricon Residential, Trulia, Zillow, and Zumper. Those sites were used to illustrate how online rental platforms developed over time. Variations across platforms in terms of who can post, what is posted, how results are displayed, and what results are prioritized were also examined. The six platforms used in the web scraping analysis were selected due to their size, the market segments they represented, and differences in data availability and structure that made the analysis feasible. Five metropolitan areas at different stages of urban growth, with diverse demographics, and in different regions of the country were selected for the web scraping analysis: Cincinnati, Philadelphia, Phoenix, San Antonio, and Seattle.

Scraping and Processing

The data collection system consisted of web scraping scripts written in two programming languages, Helena and Python. The authors ran those scripts daily throughout October 2020 to collect comparable samples across six platforms (Craigslist, Forrent.com, GoSection8, Padmapper, Trulia, and Zillow). After the removal of duplicate listings, the scraping consisted of 2,732 listings for GoSection8, 10,671 for Forrent.com, 20,283 for Padmapper, 25,812 for Craigslist, 37,819 for Trulia, and 46,316 for Zillow.

A challenge to using multiple platforms is data consistency. For every source, the raw data were processed to clean fields to the proper types (e.g., converting "$1,000" to a numeric value) so the data from different platforms could be compared. For many sources, the authors also had to adjudicate differences in data structure between rows that denoted a single unit and rows that denoted a set of units within a single building. Whether the data included each possible bedroom size, inventory count, and bedroom-specific rent or simply a rent range for the building, unit-like data for each source were constructed, with varying degrees of assumptions about rent and inventory. At best, a source had bedroom-specific rents and inventory at each bedroom size,

allowing for a complete reshaping without assumptions. Less ideal cases had a rent range and a bedroom size range, with no inventory. In those cases, rents and bedroom sizes available within the building were interpolated, and an inventory of one at each bedroom size and rent combination was assumed. Those assumptions likely led to underweighting multifamily buildings in this analysis relative to the true availability of units within those buildings. However, the authors chose this path rather than omitting a given source entirely.

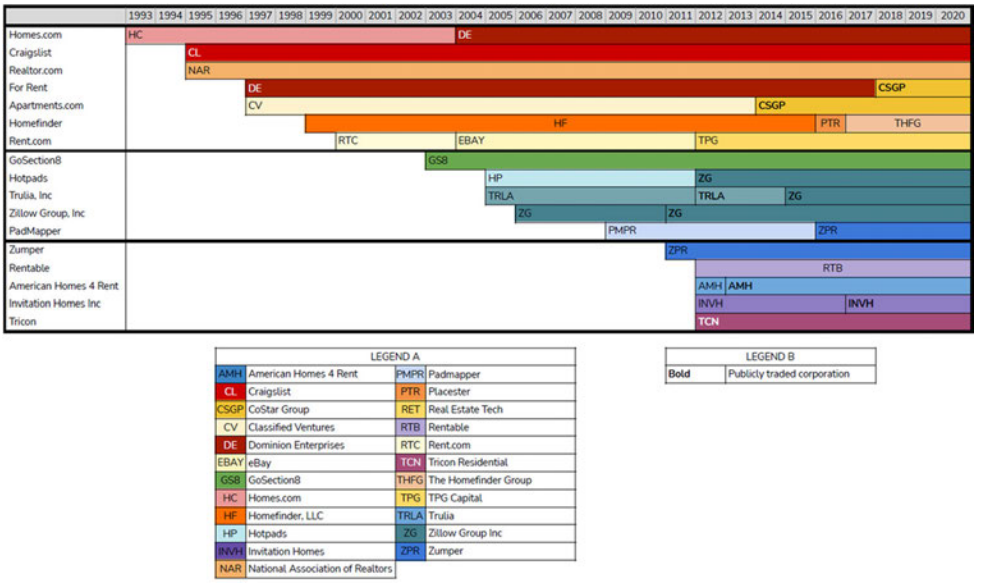
Each source has somewhat different search methods and location information, so the methods for constructing listing samples with comparable geography differed somewhat by source. For Forrent, GoSection8, Trulia, and Zillow, scrapers were constructed that collected listings from each county in the respective metros. Those sites have hard geographic constraints to searches, which allowed easy aggregation of the metropolitan areas of interest. Craigslist offers relevant but irregularly defined locations from which data were collected, so the authors had to rely on the latitude and longitude embedded within pages to assess whether listings were in one of the five metropolitan areas. Padmapper provided no means for targeting a county, only municipalities, but the authors could construct a URL with a 2-decimal degree by 2-decimal degree bounding box for each principal city. Those listings were then geocoded; listings that fell within the counties in the five selected metropolitan areas were used.

Market Consolidation

Online rental platforms emerged in the early 1990s, and three distinct time periods distinguish the founding for the websites selected for this analysis (exhibit 1). The first era represents the initial transition of rental ads from newspapers to digital spaces and includes websites founded before the turn of the century: Apartments.com, Craigslist, Homefinder, Homes.com, Forrent.com, Realtor.com, and Rent.com. The central element of this transition into digital spaces may be best represented by the launch of Craigslist, which was founded in 1995 by Craig Newmark. In 2000, Craigslist started to expand its markets to other major U.S. cities, and, by 2010, it was available in more than 700 local markets in 70 countries (Kroft and Pope, 2014; Seamans and Zhu, 2014).

Exhibit 1

Market Consolidation



LEGEND A			
AMH	American Homes 4 Rent	PMPR	Padmapper
CL	Craigslist	PTR	Placester
CSGP	CoStar Group	RET	Real Estate Tech
CV	Classified Ventures	RTB	Rentable
DE	Dominion Enterprises	RTC	Rent.com
EBAY	eBay	TCN	Tricon Residential
GS8	GoSection8	THFG	The Homefinder Group
HC	Homes.com	TPG	TPG Capital
HF	Homefinder, LLC	TRLA	Trulia
HP	Hotpads	ZG	Zillow Group Inc
INVH	Invitation Homes	ZPR	Zumper
NAR	National Association of Realtors		

LEGEND B	
Bold	Publicly traded corporation

Sources: Urban Affairs Review *(May 2020); Authors' summation based on research findings from listed sites*

The second era took place in the 2000s and included the creation of GoSection8, Hotpads, Padmapper, Trulia, and Zillow. This period was marked by the expanding presence of online housing rental platforms and an increased differentiation in terms of displayed features and target audience. Zillow.com was launched in 2006, with the idea "to aggregate and map all types of real estate data, both public and proprietary, to empower and inform consumers, and to disrupt the industry" (Green and Walker, 2017: 4). Hotpads, Padmapper, and Trulia were created to improve the home search experience through more interactive tools, a better interface, and additional resources (DeMenthon, 2008; Herel, 2010). Targeting a more specific audience, GoSection8.com is the largest online rental listing provider for housing choice voucher landlords and tenants across the United States (Bergman, Chan, and Kapor, 2020).

The third era comprises the 2010s and was primarily marked by the consolidation of corporations that own and operate these websites and an increased presence of single-family homes in the rental market. Although most websites were created by separate entities, over the past decade, mergers, acquisitions, and syndications have concentrated their ownership to a few corporations. Zillow Group is an important example of such consolidation. Although its first website, Zillow.com, was launched in 2006, rental listings were not added to the database until 2009. Since then, Zillow has significantly expanded the size of its consumer base. Not long after becoming a public company in 2011, Zillow acquired three other consumer brands: Hotpads, StreetEasy, and Trulia. After acquiring Trulia, Zillow announced the formation of Zillow Group, which it claims to be a "portfolio of the largest real estate and home-related brands" (Zillow Group, Inc., n.d.).

In addition, although single-family home rentals (SFRs) have been present in real estate markets in the United States for many decades, the past 10 years have seen a transition from a market

controlled by small investors to an increasing presence of large institutional buyers until, by 2019, real estate investment trusts (REITs) had accumulated a portfolio of more than 200,000 homes (Colburn, Walter, and Pfeiffer, 2020).

Platform Characteristics

The diversity of online rental websites presents challenges and opportunities for studying rental market dynamics. As aforementioned, online real estate marketplace companies are continually in a state of flux regarding ownership and market share, which has implications for target audiences of individual sites—for both those listing available rental properties and those in the market to rent. The authors conducted a qualitative analysis of the user interface (UI) for each website to gain greater insight into who can list rental properties, what features of each rental are prioritized, how search results are listed, and what attributes of each property are featured in the search results. Each of those analyses is described in depth on the following pages.

Most websites had relatively low barriers to entry for who could post a rental listing. On most sites, anyone can post, or a simple signup (e.g., email, rental address, or both) is required. Some sites required a verification of identity, and the three REITs included in the sample did not permit listings from external users. Figure A of exhibit 2 shows specifically which sites fell into each of the categories. The availability of listing platforms with low barriers of entry for posting and finding listings is particularly important in the rental market, given that many landlords and prospective tenants do not work with professional real estate brokers.

Exhibit 2

User Interface Features

Source: Authors' summation based on research findings from listed sites

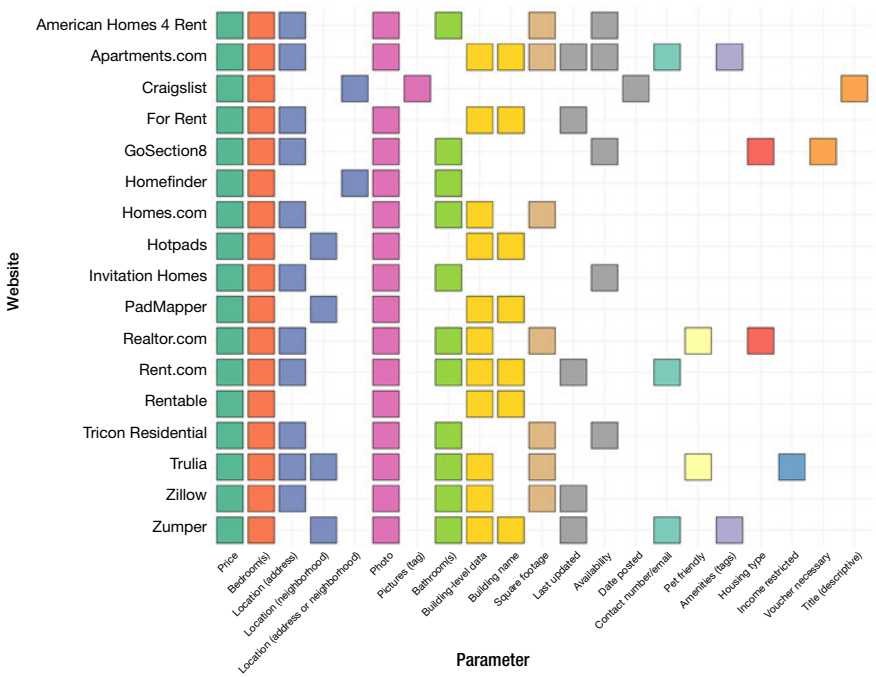
Three different UI features were evaluated from the renters' perspective. The first being the search parameters suggested to a user on the primary search page—for many websites, literally the first search page. As figure B in exhibit 2 shows, all the sites studied had options for specifying location, price, and number of bedrooms and/or bathrooms. Some sites also had options for selecting the type of housing and/or the square footage. After those basic selections, a few platforms—such as Trulia—had additional options (pets). That specificity may speak to a concerted effort to attract renters for whom those criteria are important when considering their housing options.

The second feature analyzed concerning renters was the algorithm used to list search results. Because users may not click on every property that fits their search parameters, the order of listings obviously impacts which rentals they select to inquire about further. Figure C in exhibit 2 shows that most websites employ a proprietary algorithm as the default to sort listings—usually referred to as "Just for You," "Best Match," or simply, "Default." The fact that they are using an unspecified default speaks to added variability in what prospective renters are shown above and beyond the listings available on the platform themselves.

Finally, several parameters are visible to users within the search results page before they click on an individual property listing. In exhibit 3 the price, number of bedrooms, location, and photographs are all standard across the sites we surveyed (see appendix A for differences between sites on these key features). However, websites varied in the types of additional information included in this preview page, such as building information, date-specific details, and included amenities. These differences are a crucial aspect that is not being captured by utilizing any one individual website to conduct research on online rental markets.

Exhibit 3

First-Page Search Results



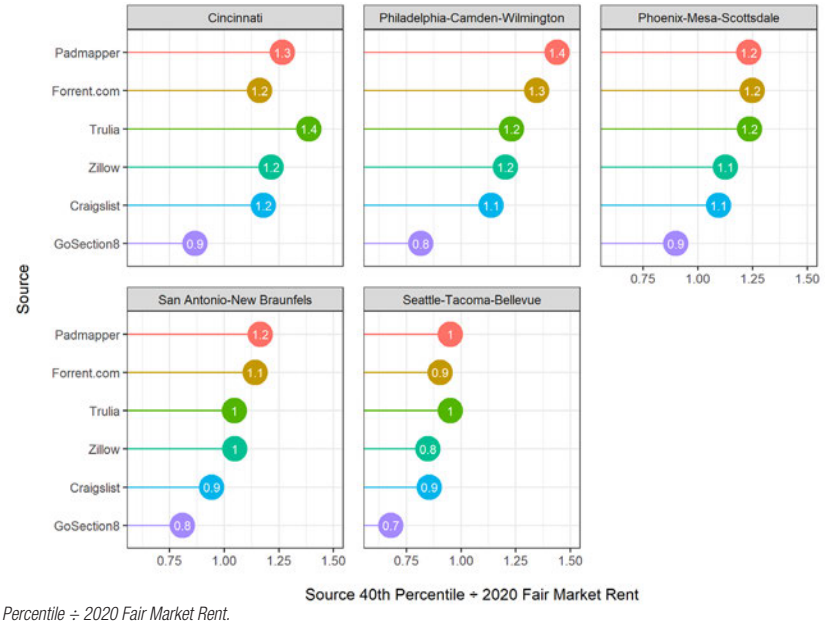
Rental Listings

For quantitative results, market-level estimates were generated for the ratio of each source's 40th percentile of rent asked to the 2020 FMR, the empirical distribution of rent asked, and the bedroom size composition for six online listings sources (Craigslist, Forrent.com, GoSection8, Padmapper, Trulia, Zillow). For the two rent analyses, the focus was on two-bedroom units to adjudicate differences in bedroom size composition by listing source. Those estimates are presented within a series of graphical displays to facilitate comparison across sources and locations.

Exhibit 4 displays ratios for the 40th percentile of rent asked among two-bedroom units relative to the 2020 two-bedroom FMR by source and metropolitan area. Across all five metropolitan areas, considerable variation is present in this ratio between the six different listing sources. Some sources—such as Forrent.com, Padmapper, and Trulia—are consistently among the most expensive listing sources, whereas GoSection8 and, to a lesser extent, Craigslist have a clear role as a source of information for more affordable rental housing opportunities. In the Cincinnati, Philadelphia, and Phoenix metropolitan areas, most sources' 40th percentile for rent asked is about 20 percent—if not more—higher than the corresponding metropolitan FMR level. By contrast, data for the San Antonio and Seattle metropolitan areas display closer alignment between the 40th percentile from online listing sources and the relevant FMR. The nearly 1:1 relationship between the platform 40th percentiles and FMR in the Seattle metropolitan area stems from this metro's FMR being based on independent surveys rather than the ACS. That difference in FMR construction from the other metropolitan areas reflects this area's exceptional rental growth in recent decades; lacking that adjustment, the ratios would be comparable to, if not greater than, those observed in the other four metropolitan areas.

Exhibit 4

Ratio of 40th Percentile for Two-Bedroom Rent Relative to the 2020 Metropolitan Fair Market Rent for Two-Bedroom Units



Source: 40th Percentile ÷ 2020 Fair Market Rent.

Exhibit 5 shows the empirical distributions for rent asked among two-bedroom units by data source and metropolitan area. Included for comparison is the rent distribution among households in the 2019 ACS who recently moved. The GoSection8 distribution confirms that this source is particularly focused on lower asking rents in each metropolitan area, given the curves' flatness across percentile ranks. The other sources display sizable differences among metropolitan areas in how spread out the various sources are. One possible explanation is metropolitan heterogeneity in how much the listings advertised overlap each other between sources, suggesting how sources might overlap relatively more in the Seattle-Tacoma-Bellevue and Cincinnati metro areas compared with Phoenix-Mesa-Scottsdale or San Antonio-New Braunfels. Variations across sources are important for researchers and policymakers to consider when working with data from particular sources to capture rental stock dynamics.

Exhibit 5

Empirical Distributions for Rent Asked, by Data Source and Metropolitan Area

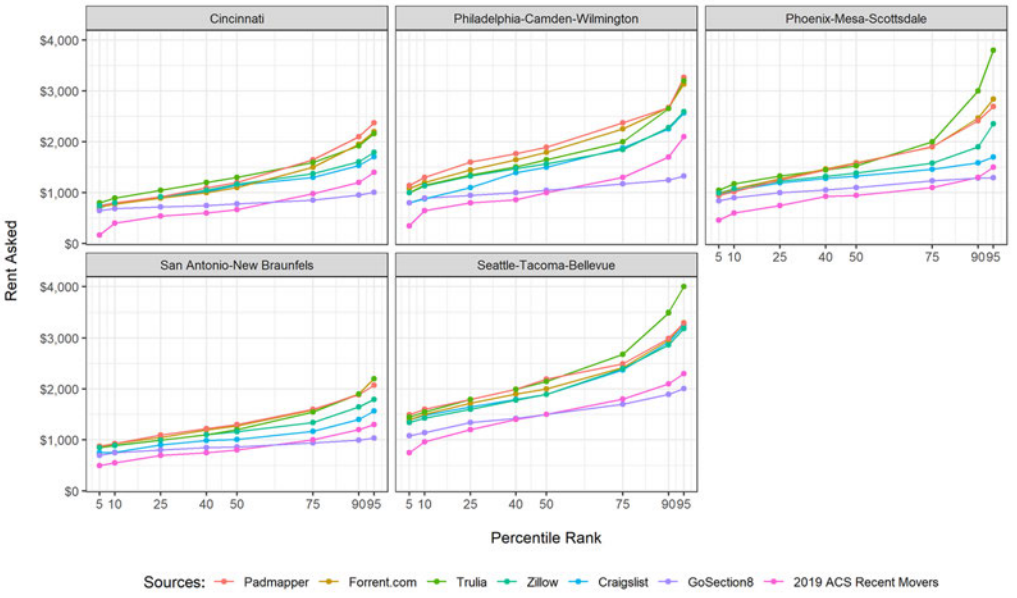
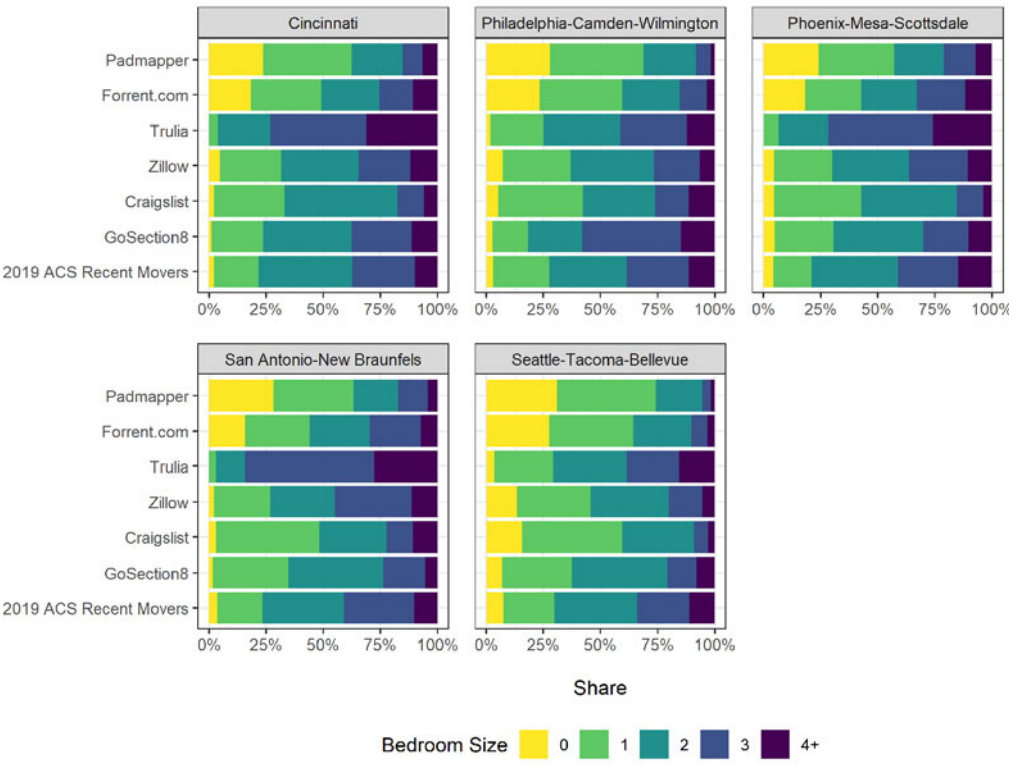


Exhibit 6 visualizes the composition of listings on each online platform in terms of bedroom size (i.e., studio or 1, 2, 3, or 4+ bedrooms) as a stacked bar graph. The composition of housing units by bedroom size for households in the 2019 ACS who recently moved is included for comparison. Whereas Trulia has a greater focus on larger-sized rental units among each of the five metropolitan areas, other sources, such as Forrent.com and Padmapper, have compositions more aligned with coverage of multifamily apartment complexes. Craigslist and Zillow—much as with rent estimates—fall somewhere between the other four sources in terms of the composition of units by bedroom size.

Exhibit 6

Bedroom Size by Data Source and Metropolitan Area



Note: ACS = American Community Survey.
Source: American Community Survey 5-Year Public Use Microdata Sample

Discussion

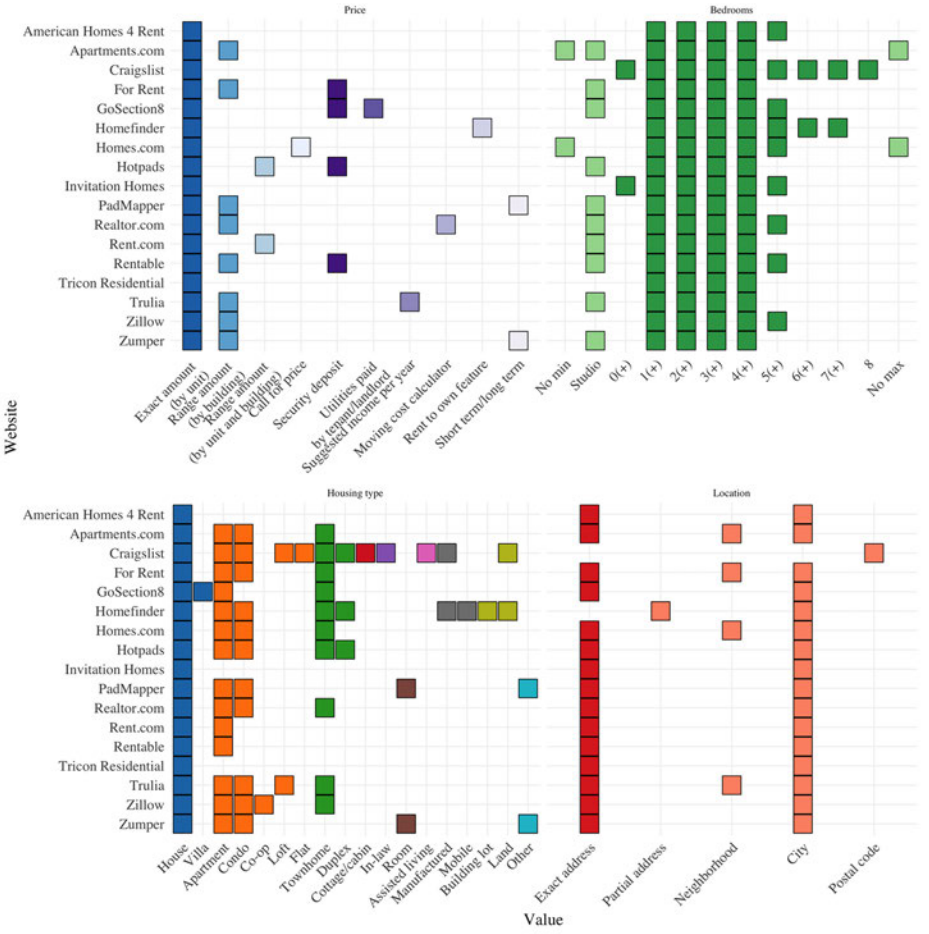
Technology brought rental ads into digital spaces, a transition that is increasing access to information for prospective renters. This shift is changing the way households search for a place to rent and, in turn, how real estate professionals reach consumers. For researchers and policymakers, these online real estate platforms represent an opportunity to capture readily available, real-time data from local rental markets across the country.

The market consolidation analysis in this study reveals the emergence of multiple rental listing websites during the 1990s and 2000s, followed by gradual amalgamation into a limited number of corporations within the past decade that host most listings. The platform characteristics reveal that although consolidation has occurred, many of the sites—even those with the same ownership—vary in terms of posting standards, search parameters, algorithms used to list search results, or search results page. The rental listing analysis shows considerable variation in typical rents, overall distributions of asking rent, and bedroom size across platforms and compared with 2020 FMRs and 2019 ACS data.

These findings have major implications for policymakers and researchers who are seeking to enhance rental housing market analyses. On one hand, the emergence of online platforms enhances the ability to track changes in local rental housing listings that, before, suffered from temporal and spatial specificity. On the other hand, this paper highlights the significant variation in both the types of units advertised and the information provided across these platforms, indicating that the use of one platform may not sufficiently represent the current state of the rental stock or may require care in developing adjustments rather than using the raw data. Given that real-time rental listing data is exceedingly valuable for understanding current rental housing dynamics and conditions, future work is needed to address the challenges in the collection, processing, storage, and dissemination of rental listings, as well as the data and methods used for tracking local rental market trends and calculating rent estimates.

Appendix A

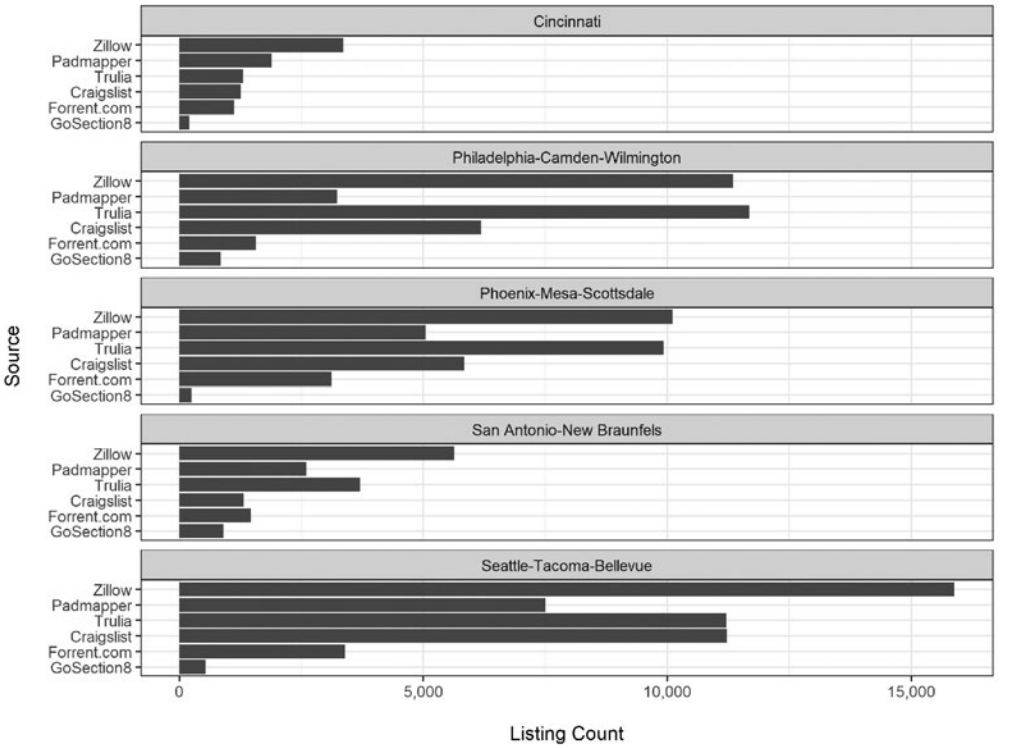
Differences in Key Features



Source: Authors' summation based on research findings from listed sites

Appendix B

Listing Counts by Source and Metropolitan Area



Acknowledgments

Partial support for this research came from a Eunice Kennedy Shriver National Institute of Child Health and Human Development research infrastructure grant, P2C HD042828, to the Center for Studies in Demography & Ecology at the University of Washington. We also thank Sanjana Gupta and Purba Roy (Information School, University of Washington, Seattle) for their work on the web scraping scripts and data infrastructure that made this project possible.

Authors

Ana Costa is a Ph.D. student in the Department of Urban Design and Planning at the University of Washington, Seattle. Victoria Sass is a Ph.D. student in the Department of Sociology at the University of Washington, Seattle. Chris Hess is a postdoctoral associate in the Department of Policy Analysis and Management at Cornell University, Ithaca, NY. Ian Kennedy is a Ph.D. candidate in the Department of Sociology at the University of Washington, Seattle. Alex Ramiller is a Ph.D. student in the Department of City and Regional Planning at the University of California, Berkeley. Roshni Roy is a master's student in the Information School at the University of Washington, Seattle.

Sarah Chasins is an assistant professor in the Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley. Rebecca J. Walter is an associate professor in the Department of Real Estate at the University of Washington, Seattle. Arthur Acolin is an assistant professor in the Department of Real Estate at the University of Washington, Seattle. Kyle Crowder is a professor in the Department of Sociology at the University of Washington, Seattle.

References

Bergman, Peter, Eric W. Chan, and Adam Kapor. 2020. Housing Search Frictions: Evidence from Detailed Search Data and a Field Experiment. Working paper No. 27209. Cambridge, MA: National Bureau of Economic Research. http://www.nber.org/papers/w27209.

Boeing, Geoff, Max Besbris, Ariela Schachter, and John Kuk. 2020. "Housing Search in the Age of Big Data: Smarter Cities or the Same Old Blind Spots?" *Housing Policy Debate 31* (1): 112–126. https://doi.org/10.1080/10511482.2019.1684336.

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. https://doi.org/10.1177/0739456X16664789.

Colburn, Gregg, Rebecca J. Walter, and Deirdre Pfeiffer. 2020. "Capitalizing on Collapse: An Analysis of Institutional Single-Family Rental Investors," *Urban Affairs Review* (May): 1–36. https://doi.org/10.1177/1078087420922910.

DeMenthon, Eric. 2008. "'Hello World!', Said PadMapper." PadBlogger blog, December 27. https://blog.padmapper.com/2008/12/27/hello-world/.

Green, Joanna, and Russell Walker. 2017. *Neighborhood Watch: The Rise of Zillow*. Evanston, IL: Kellogg School of Management, Northwestern University. https://doi.org/10.4135/9781526440822.

Herel, Suzanne. 2010. "Meet the Boss: Sami Inkinen, COO, Trulia." https://www.sfgate.com/business/meettheboss/article/Meet-the-Boss-Sami-Inkinen-COO-Trulia-3176725.php#photo-2312759.

Hess, Christian, Rebecca J. Walter, Arthur Acolin, and Sarah Chasins. 2019. "Comparing Small Area Fair Market Rents With Other Rental Measures Across Diverse Housing Markets," *Cityscape* 21 (3): 159–186. https://www.huduser.gov/portal/periodicals/cityscpe/vol21num3/ch8.pdf.

Kroft, Kory, and Devin G. Pope. 2014. "Does Online Search Crowd Out Traditional Search and Improve Matching Efficiency? Evidence from Craigslist," *Journal of Labor Economics 32* (2): 259–303.

Seamans, Robert, and Feng Zhu. 2014. "Responses to Entry in Multi-Sided Markets: The Impact of Craigslist on Local Newspapers," *Management Science 60* (2): 476–493. https://doi.org/10.1287/mnsc.2013.1785.

Zillow Group, Inc. n.d. "Our Brands & Businesses." https://www.zillowgroup.com/about-us/business/.