Forthcoming in Land Economics (provisionally accepted Mar 30, 2022)

Submission for Joint Special Issue "Property value analyses using ZTRAX: Applications under the existing sunset" (co-editors Daniel Phaneuf, Jeffrey Zabel, Kevin J. Boyle, Christoph Nolte)

Data practices for studying the impacts of environmental amenities and hazards with nationwide property data

Christoph Nolte[1,2,*], Kevin J. Boyle[3,4], Anita Chaudhry[5], Christopher Clapp[6], Dennis Guignet[7], Hannah Hennighausen[8], Ido Kushner[1], Yanjun Liao[9], Saleh Mamun[10,11], Adam Pollack[1], Jesse Richardson[12], Shelby Sundquist[13], Kristen Swedberg[3], Johannes H. Uhl[14,15]

[1] *Department of Earth & Environment, Boston University, Boston, Massachusetts, United States of America*
[2] *Faculty of Computing & Data Sciences, Boston University, Boston, Massachusetts, United States of America*
[3] *Department of Agricultural and Applied Economics, Virginia Tech, Blacksburg, Virginia, United States of America*
[4] *Program in Real Estate, Virginia Tech, Blacksburg, Virginia, United States of America*
[5] *Department of Economics, California State University, Chico, California, United States of America*
[6] *Harris School of Public Policy, University of Chicago, Chicago, Illinois, United States of America*
[7] *Department of Economics, Appalachian State University, Boone, North Carolina, United States of America*
[8] *Department of Economics, University of Alaska, Anchorage, Arkansas, United States of America*
[9] *Resources for the Future, Washington, District of Columbia, United States of America*
[10] *Department of Applied Economics, University of Minnesota, St. Paul, Minnesota, United States of America*
[11] *Natural Resources Research Institute, University of Minnesota, Duluth, Minnesota, United States of America*
[12] *College of Law, West Virginia University, Morgantown, West Virginia, United States of America*
[13] *School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, Arizona, USA*
[14] *Institute of Behavioral Science, University of Colorado, Boulder, Colorado, United States of America*
[15] *Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado, United States of America*

** Corresponding author. Email: chrnolte@bu.edu (CN)*

Abstract

We discuss data quality and modeling issues inherent in the use of nationwide property data to value environmental amenities. By example of ZTRAX, a U.S.-wide real estate database, we identify challenges and propose guidance for: (1) the identification of arm's-length sales, (2) the geo-location of parcels and buildings, (3) temporal linkages between transaction, assessor, and parcel data, (4) the identification of property types, such as single-family homes and vacant lands, and (5) dealing with missing or mismeasured data for standard housing attributes. We review current practice and show that how researchers address these issues can meaningfully influence research findings.

Introduction

Recent years have seen a rapid growth in empirical studies that use large-scale real estate data to value environmental amenities and hazards in the United States. This growth in empirical work is not novel in itself. The output of research papers using hedonic property-value methods has been trending upwards for three decades (Hanley & Czajkowski 2019), reflecting enduring interest in estimating peoples' environmental preferences, federal mandates to consider such values in regulatory cost-benefit analyses, and recent advances in computational capacity, econometric methods, and best practices (Bishop et al. 2020). However, a noteworthy recent trend is the publication of many hedonic analyses that make inferences across large and diverse geographic areas (see examples in Table S1). While there are many sources of real estate microdata (e.g., state-level databases, private data aggregators, Multiple Listing Services), an important contributor to this recent growth has been the decision of Zillow Inc., a U.S. online real estate marketplace company, to share its Transaction and Assessment Dataset ("ZTRAX") for free with U.S. academic, non-profit and government researchers between 2016 and 2023 (Zillow 2021a).

Access to large-scale real estate data has many potential benefits for economic research. It can help researchers improve and expand the set of available estimates of people's preferences for various amenities associated with property locations and characteristics (Bernstein et al. 2019; Clarke & Freedman 2019; Albouy et al. 2020; Baldauf et al. 2020; Murfin & Spiegel 2020). It can narrow gaps in the geographic coverage of evidence derived from small-scale studies (e.g., Guignet et al. 2022). And it can mitigate some risks identified as contributing to a "credibility crisis" in environmental and resource economics (Ferraro & Shukla 2020, 2022). For instance, broader access to nationwide data puts more researchers into a position to reproduce and replicate findings from prior studies and to test their generalizability across heterogeneous contexts. This greatly increases the credibility of existing results and the intellectual merit of such findings (Maniadis et al. 2014, 2017). Similarly, research efforts that would suffer from underpowered designs if conducted in a single locality can produce more defensible and insightful results when pooling real state data across a large number of sites.

However, large-scale real estate data sets also create new analytical challenges. In the U.S., such data is usually aggregated from public records provided by thousands of local data generators (county-level tax assessors, deed registries, and mapping departments). Analysts often find that large-scale public records data is provided in an only partially pre-processed state and requires substantial cleaning to be usable for empirical analyses. For ZTRAX, Zillow explicitly cautions its users that "extensive exploring on your part is required due to the detailed, rich, and nuanced

nature of the dataset" (Zillow 2021b). Researchers therefore face many data preparation choices that can affect findings but for which no published codebooks or "best practice" guidelines exist. If unreported, flexibility in data preparation adds to the range of "researcher degrees of freedom" (Simmons et al. 2011) that can leave studies vulnerable to researcher behavior that undermines the credibility of empirical findings (Christensen & Miguel 2018). Awareness of potential errors and biases, a full documentation of filtering choices, and a careful discussion of potential effects on research findings has several benefits. It can reduce the influence of such "hidden" researcher decisions (Huntington-Klein et al. 2021), enable reviewers and editors to ask the right questions, and enhance the reproducibility, replicability, and generalizability of published work.

In this article, we catalog and discuss issues related to researcher decisions when working with real estate microdata. The article is the result of a group effort by academic researchers from ten U.S. universities who have used ZTRAX for several large-scale property-level analyses and share an interest in the accuracy, reliability, and reproducibility of their empirical findings. Contributors to this article have used ZTRAX data to estimate: the cost of land acquisitions for conservation purposes (Nolte 2020); the effect of water markets on agricultural land values [Chaudhry et al., this issue (LE)]; the benefits of lake water quality (Mamun et al. 2023a) [Swedberg et al., this issue (LE)] and the cost of its impairment (Mamun et al. 2023b); the risk of flood damage to residential homes (Gourevitch et al. 2023); the effects of flood insurance policies [Hennighausen, this issue (JHE), Pollack et al., this issue (JHE)]; the cost of hazardous chemical releases and the benefits of subsequent cleanups (Guignet & Nolte 2021) [Guignet et al., this issue (JHE)]; and property value impacts of critical habitat under the U.S. Endangered Species Act [Mamun et al., this issue (LE)]. Through this work, we have identified common problems of working with large-scale property data, and experimented with potential solutions in the following areas:

1. Identifying transaction prices reflecting fair market value,
2. Geo-locating transacted properties: land and buildings,
3. Linking transactions to time-variant property characteristics,
4. Identifying specific types of properties, e.g., single-family homes or vacant lands, and
5. Dealing with missing or mismeasured data for standard housing attributes.

After a brief introduction to data types and sources, we discuss each issue, establish its relevance, and consider a range of potential solutions for each. We then conduct a literature review of recent peer-reviewed ZTRAX-based analyses to examine the extent to which researchers disclosed their decisions on each issue. Lastly, we explore the extent to which findings can be affected by analysts' choices using an illustrative hedonic analysis with different data preparation specifications to estimate the effect of flood zone location on property prices. Alongside this article, we publish a set of digital resources (deed interpretations, filtering tables, source code) to document our own choices and to help other analysts implement, scrutinize, and improve data preparation procedures and assumptions.

While ZTRAX forms the basis of our analyses, the issues and solutions we discuss generalize to other real estate microdata sources. We see this article as a starting point for the development of best data practice guidelines for large-scale property-based analyses in the United States. Our propositions should not be interpreted as an attempt to develop universal standards for all cases, as many decisions will remain specific to research questions, location, and dataset. However, by

helping researchers, reviewers, and editors develop an awareness for the potential consequences of data preparation choices on research results, we hope this article will encourage broader application of steps that can increase the credibility and transparency of empirical findings, such as a more consistent documentation of data processing choices to facilitate reproduction and replication, and a consideration of a broader range of errors and robustness checks in analyses and reviews.

Data types and sources

Real estate databases used in hedonic analyses often use at least one of three distinct types of public records data. In the U.S., these public records are produced by different branches of local government (e.g., county, town, or registry district) and serve different purposes:

1. "Assessment" data refers to tabular data compiled by local tax assessors for the purpose of assessing and collecting property taxes. Because tax assessors are tasked with maintaining a complete account of the taxable value of all properties within their jurisdiction, assessment data can usually be expected to contain a complete or near-complete list of properties within a given county or town. The set of variables collected for each property varies across geographies, but commonly includes property identifiers, such as assessor parcel numbers (APNs) and addresses; assessed or appraised values, sometimes provided separately for land and buildings; building characteristics, such as age, size, counts of stories, bathrooms, bedrooms, and other features (pool, garage); land characteristics, such as lot size, land use category, and other features (e.g., lake frontage, views); and owner identifiers, such as names, addresses, and tax account numbers.
2. "Transaction" data refers to tabular data of property transaction records, including deeds, mortgages, and foreclosures. Transaction data is generated only if and when property ownership changes and can include repeat sales (i.e., multiple transactions of the same property). Again, the set of available variables varies across geographies but often includes transaction price, date, document type (e.g., deed type), owner and seller names, property identifiers, and other information of interest (e.g., flags for intra-family, arm's-length, or partial-interest transfers, information on mortgages and loans). Property identifiers can then be used to link assessment and transaction data.
3. "Parcel boundary" data refers to geo-located polygons of parcel boundaries (i.e., vector data). Digital parcel maps now exist in all but a few U.S. counties. Parcel boundary data usually comes with an attribute table that includes variables for each property, including parcel identifiers (APNs, addresses), as well as different subsets of attributes joined from assessment data, such as owner names and assessed values. Parcels and properties are not always identical: a single property can have multiple parcels (e.g., a large ranch), and a single parcel can include multiple properties (e.g., multi-family homes).

Our exemplar data, ZTRAX, contains assessment and transaction data, but no parcel boundary data. Many of the insights we share subsequently have been obtained by comparing ZTRAX records to supplementary datasets, such as parcel boundaries, as well as satellite-derived building footprints and land cover classes. The authors affiliated with Boston University linked most ZTRAX records for the contiguous United States (CONUS) to parcel boundary data using text-based parcel identifiers and conversion algorithms developed as part of the Private-Land

Conservation Evidence System (PLACES) (Nolte 2020) and described in a subsequent section on geolocation. Because we use licensed parcel boundary data from third-party providers for approximately two thirds of U.S. counties, we are not allowed to publicly share the full parcel-level dataset underlying our claims, such as corrected parcel and building coordinates. However, with the methodological descriptions we offer below and access to similar parcel boundary data, the computationally versed reader should be able to reproduce our findings for their study region and implement the proposed corrections and filters. Unless otherwise stated, all maps and statistics in this article are derived from the ZTRAX database version made available in Oct 2019 (downloaded on Feb 3, 2020) and limited to the continental U.S. (CONUS).

According to Zillow, ZTRAX is sourced "from a major large third-party provider and through an internal initiative we call County Direct" (Zillow 2021b). Geographic coverage of transaction data (>2,750 counties, >400 million transactions) is smaller than that of assessment data (>3,100 counties, >150 million properties). The dataset also contains an archive of historical assessment that allows the tracking of changes at the property level; its temporal coverage extends back to the early 2000s in most states but varies across geographies (Figure S1).

ZTRAX exhibits substantial geographic heterogeneity in the availability of transaction price information, the dependent variable in most property-focused revealed-preference studies. The availability of price information is strongly shaped by the extent to which U.S. states require disclosure of sales prices (Figure 1). Lists of non-disclosure states commonly include: Alaska, Idaho, Kansas, Louisiana, Mississippi, Montana, New Mexico, North Dakota, Texas, Utah, and Wyoming (Wentland et al. 2020). We also find sales price data to be rare in Indiana, Maine, Missouri, and South Dakota, as well as in a large share of counties in several other states (e.g., Alabama, Nebraska, Michigan, and Minnesota). Where the density of sale price observations is scarce, some transactions might still contain price data, but these are rarely representative (e.g., they might be associated with foreclosures or public sales) and therefore warrant greater scrutiny.

States and counties also vary in the length of the time period for which transaction data is consistently available. Most counties contain transaction data for the first two decades of the 21st century. Three or more decades of data are available in Northeastern states (Connecticut, Maryland Massachusetts, New Jersey, New York, Rhode Island), much of California, Florida, Tennessee, and Ohio, as well as urban centers across the country (Figure S2).

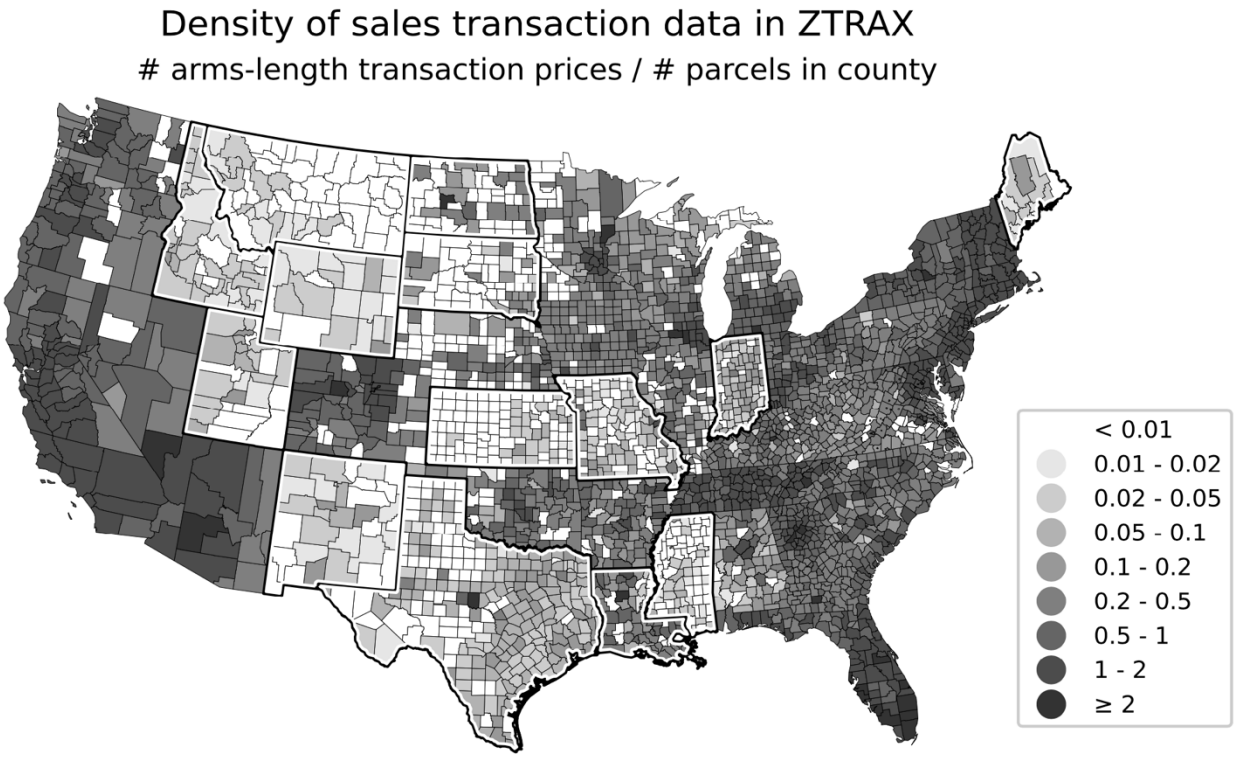

*Figure 1: Density of sales transaction data in ZTRAX. Density is computed as the county-level ratio of (i) the count of non-duplicate arms-length transaction records with prices >$1000 and (ii) the count of parcels in digital parcel maps. Black-and-white boundaries show non-disclosure states or states with unusually scarce price data. Note that the time period spanned by ZTrans transaction data varies across counties (**Error! Reference source not found.**), which contributes t o the observed differences in sales densities.*

Data preparation challenges

Identifying transaction prices reflecting fair market value

Real estate appraisals, hedonic pricing methods, risk assessments, and other property analyses often rely on the assumption that transaction prices are indicative of the fair market value (FMV) of the transacted property. FMV is the price at which a property would change hands between a willing buyer and a willing seller in a competitive market, neither being under any compulsion to buy or sell. Public transaction data often includes prices of transactions that do not fulfill these conditions, such as transactions between family members; transactions under distress (such as foreclosures); transactions below market value from public actors (e.g., by a targeted sale to veterans); or prices referring to monetary amounts other than the full property value (e.g., loans, mortgages, partial interests). To avoid biases that can affect subsequent conclusions, researchers need to be able to identify and isolate FMV transactions.

Guidelines on how to identify FMV transactions in public records data are limited. Transaction records often include fields that can be used for the development of filters, such as document

types, seller and buyer names, or arm's-length or intra-family flags. However, the interpretation of document types often requires domain expertise on the legal meaning and usage of different types of contract documents, and how that usage varies across jurisdictions. Ancillary flags developed to address these concerns are often of undocumented provenance and incomplete. For instance, the transaction data in ZTRAX contains an "intra-family transfer flag", which identifies transactions between family members using an undocumented algorithm. Based on a comparison of buyer and seller names, we estimate that this flag misses potentially 15.2 million intra-family transactions (6.1% of deed records). Zillow's FAQ confirm that their own cleaning procedure includes an internal text-matching algorithm (Zillow 2021b) that is not publicly documented.

We propose filters for nine variables available in ZTRAX, and likely in similar real estate databases, to identify transactions whose prices are more likely to reflect FMV (Data Set 1). Our approach distinguishes between transactions where the reported sales price reflects FMV with "high", "medium", and "low" confidence. After applying each filter to its respective variable, the individual filters can be combined (e.g., by only retaining transactions that obtained a "high confidence" value across all filters).

A description of filters and filtered categories is provided in Data Set 1. Six of the nine listed filters are straightforward exclusions based on discrete values (data types, document types, loan types, price types, intra-family flags) or cutoffs (for token prices). Three warrant further elaboration:

1. Name similarity: To enhance the identification of intra-family transfers, we compute a similarity index between buyer and seller names. For each transaction, our algorithm computes the percentage of identical words appearing in both fields, weighing each word by the inverse of the square root of its relative frequency within the same county. This inverse frequency weighting reduces the probability that name similarity is erroneously established from frequently occurring words (e.g., "John" or "Michael"). We omit words with ≤2 letters and remove frequent generic words (e.g., "Bank", "LCC"). Transactions for which buyer or seller names returned a similarity index of 66% or more are flagged as "low-confidence" FMV sales. We note that this threshold should be treated as a rule-of-thumb, as we do not have a way to validate predictions based on different cutoffs to determine which minimizes classification error. Using this method, we estimate that a naïve approach to identifying intra-family transactions from document types and the intra-family transfer flag underestimates the actual extent of intra-family transfers. Of 245 million deed records, ZTRAX flags 23.0% with its intra-family transfer flags. We estimate the actual number to be 29.1%, a difference of 15.2 million transactions. We provide Python code to reproduce this similarity index with this article (Digital Material).
2. Document types: Developing filters for the 161 different document types is not trivial, as they can have different meanings and usages in different states. For instance, warranty deeds are the most frequent source of FMV transactions in most states but grant deeds are more frequent in California, Nevada, and Vermont. Quitclaim deeds rarely contain sales price information in most states except in Massachusetts and Vermont, where prices are frequently provided (Figure S3). Because ZTRAX contains 3,837 unique combinations of states and document types, an in-depth assessment of each combination is not feasible. Our filters therefore combine a hierarchical exclusion filter with a data-driven follow-on: First, with the help of a land use attorney, we make deterministic choices for the most frequent unambiguous document codes, including flags for foreclosures, intra-family

transfers, loans, and cancellations. An explanation of the meanings of the most frequent deeds is provided in the Supplementary Text 1. For the remaining, non-excluded codes, we base our filters on two ancillary statistics computed for each combination of states and document codes: 1) the percentage of transactions with a price >$1000, and 2) the percentage of non-family transactions (see bullet point 1.). We flag state-code combinations where at least one of these statistics was found to be <10% or <33% as "low" and "medium" confidence, respectively.

3. Public buyers and sellers: We identify public parties in a transaction from their respective name fields using text pattern matching ("regular expressions", see Friedl 2006). There are a broad range of public organizations in the United States, and their names can be spelled in diverse ways within and across county registries. In states and regions where price data is scarce (e.g., Arkansas, Indiana, Texas), a comparatively large share of transaction data can come from public sources of sales records, which can bias estimates of property value downwards. Idiosyncrasies are common: in Arkansas, for instance, Commissioners of State Lands were identified in records by their personal names, not by their positions. Our approach is therefore best described as a hybrid of top-down (names of pre-identified agencies) and bottom-up (spelling learned from the data) identification of string patterns of public organizations. Due to the absence of an external validation dataset for testing, we do not claim that the set of expressions is comprehensive. Instead, we share a machine-readable table of text patterns alongside this article (Data Set 2) and will post feedback we receive on http://placeslab.org/hedonic-data-practices.

Geo-locating transacted properties: land and buildings

Property analyses that study localized spatial phenomena often need accurate information of the location of land, buildings, or both. For example, many hedonic property value analyses infer landowners' preferences for environmental characteristics from spatial associations between the prices of transacted parcels and environmental variables of interest. To do so, analysts need to first geo-locate transactions, and then computationally derive variables of interest from spatial data of environmental attributes. Depending on the application, demands on spatial precision can be high. For instance, Netusil et al. (2019) show that estimates of impacts of floodplain location on property values can be very sensitive to the choice of parcel boundaries vs. buildings footprints as the spatial reference. Many analyses also benefit from incorporating information on the characteristics of the land under each parcel. For example, estimates of the impacts of development restrictions (and the cost of conservation easements) need to account for a parcel's potential for future development, which is affected by its terrain, wetland presence, flood risk, existing land cover, etc. Point locations are usually insufficient for the computation of such area-based proxies; instead, a geo-located polygon of the parcel boundary is required.

Many large-scale property analyses rely on point locations (latitude and longitude) found in assessment or transaction data to geo-locate properties: for instance, 85% of peer-reviewed hedonic analyses using ZTRAX made this choice (cf. our subsequent literature review). However, the provenance and meaning of geographic coordinates is often unknown and poorly documented. Zillow describes its coordinates as "enhanced Tiger coordinates" (Zillow 2021b) and "Populated by GEOCoder", which might refer to the U.S. Census Geocoder (U.S. Census

Bureau 2021). Using these coordinates without careful attention to the coordinate system, duplicates, missing data, and building locations can lead to misleading or unrepresentative findings. We identified six issues that are particularly worthy of attention. Figure S4 illustrates these issues in two U.S. counties (Middlesex, Massachusetts, and Lane, Oregon):

1. Latitude and longitude coordinates can be derived using various geodesic datums (e.g., NAD27, NAD83, WGS84). However, the datum information is not always provided: in ZTRAX, it is entirely missing. Comparisons with geo-referenced parcel boundary data suggest that until early 2020, the predominant datum of ZTRAX varied by county in a non-predictable pattern (Figure S5). In more recent versions (Oct 2021), most counties use the more recent WGS84 datum, but exceptions remain (e.g., in New England). We also found counties using multiple datums for neighboring properties. Not correcting for these issues can lead to systematic geo-location errors that vary in magnitude across geographies: they will generally be higher on the coasts (~100m in California, ~40m in New England states), but are largely negligible in the Midwest (e.g., Indiana and Michigan).
2. Some coordinates seem to be derived from ZIP code area centroids instead of parcel locations or street addresses, without being flagged as such. Anecdotal evidence from visual inspection suggests that this issue is particularly common for recent subdivisions and properties without addresses. Using these coordinates can lead to geo-location errors of greater than 1km.
3. Many records are missing latitude and longitude data (Figure S6). Missing data is often associated with particular types of parcels (e.g., vacant parcels, rural parcels, records without addresses). Excluding them from an analysis where such parcels would otherwise be included will result in non-random selection into the sample.
4. Some counties appear to base their coordinates on parcel centroids, whereas others seem to refer to building locations. Distances between building footprints and parcel centroids vary across the country (Figure S7); mean distances of >100m are common in rural settings with large parcels. Uhl et al. (2021) compared ZTRAX locations to remote-sensing derived building footprint data (Microsoft 2018) and find positional accuracy to decrease as one moves from urban to rural settings. Analyses that assess the impact of spatially precise policies (e.g., official floodplains) are thus subject to errors of possibly large magnitudes (Netusil et al. 2019).
5. Most counties contain at least some incorrect, non-duplicate parcel locations (Figure S8). Possible observed reasons include coordinates being based on owner's mailing addresses (instead of property location addresses), as well as subdivisions of parcels.
6. Point locations can change between updates. For instance, in a comparison of Rhode Island property locations between versions of ZTRAX downloaded in 2017 and 2019, we found ZIP code area centroid placeholders to be replaced with street address geo-locations (Figure S9). For many Rhode Island properties, we found minor shifts in point locations, which were multidirectional and thus not simply attributable to changes in projection (Figure S9). While such changes appear to reflect improvements in geo-location over time, they also highlight the need to exercise particular caution in geo-location records when leveraging assessment data from multiple time periods.

There are several options to improve the geo-location of assessment and transaction data. Because geo-coordinates appear to improve in recent time, we recommend starting with the most recent available database. Analyst can then choose among the following options as a function of their resource constraints (data access and time available for data inspection and cleaning) and the anticipated sensitivity of findings to geo-location errors.

- Quick fixes: Analysts without access to digital parcel maps or without the geoprocessing skills to link assessment and transaction records to parcels and buildings can enhance the reliability of their findings with two fixes. Specifically: (i) ensure that the correct datum is used to spatially locate coordinates (Figure S5, Data Set 3), and (ii) drop entries with duplicate coordinates, especially if these coordinates are ZIP code area centroids and if dropped records have either no or unique street addresses. Figure S6 shows the prevalence of missing and non-empty duplicate coordinates in ZTRAX assessment data and, thus, of anticipated reductions in sample size and county coverage when removing these entries. Figure S8 shows how much geo-location error remains in each county after implementing these two fixes. In most counties, the median geo-location error drops to less than one meter, suggesting that most parcels are correctly located. However, mean errors can be large (often >500m), indicating that a share of parcels will remain incorrectly located, sometimes to a large degree.
- Crop to county and ZIP code boundaries: If county identifiers and ZIP codes are provided, they can be used to remove coordinates that fall outside the corresponding spatial boundary. Official boundaries of counties and ZIP codes are available through IPUMS' National Historical Geographic Information System (NHGIS) (Manson et al. 2018). We do not recommend cropping to smaller census units (e.g., census tract or block boundaries): in the case of ZTRAX, identifiers for these units appear to have been derived directly from the geo-coordinates through spatial joins.
- Geocode addresses: If records with missing geo-coordinates contain addresses (street, city, and zip code) (Figure S10), analysts can improve the completeness of geolocations by means of additional geocoding, e.g., by using the U.S. Census Geocoder or the GeoCoder API (https://geocoder.readthedocs.io). This approach remains untested and might be vulnerable to the same issues as we observe for coordinates in assessment data but will likely be able to take advantage of recent updates to geo-location databases.
- Linking assessment data to parcel boundary data: Digital parcel maps now exist in at least 3,073 (97.8%) of counties in CONUS. In most cases, parcel boundaries can be uniquely and reliably linked to assessment data using unique parcel identifiers (APNs) or unique taxpayer account numbers. This approach tends to lead to a more reliable and complete geo-location of assessment data than the previous fixes. It also allows analysts to derive important indicators of property value from the parcel boundary data (e.g., building footprint size, road access, lake frontage, wetland coverage, forest stocks). However, establishing this linkage is complicated by idiosyncratic differences in the syntaxes of APNs, which vary not only between assessor and parcel boundary datasets, but also geographically between neighboring counties and towns. We recommend a four-step approach that consists of: a) developing text pattern descriptors (regular expressions, see Friedl 2006) to identify the prevalent APN syntax in a given county or town and to extract the identifying text fragments (e.g., numbers or letters without spaceholders), b)

re-formatting and recombining the extracted text fragments to create a new parcel identifier that has the same syntax for both assessment and parcel polygon data, c) iterating over the two previous steps until the new parcel identifier produces the largest number of uniquely matched records across assessment and parcel boundary data, and d) double-checking that this linkage results in relatively small spatial distances between geographic locations provided in the assessment data and matched parcel boundaries. The last step helps identify erroneous linkages when the syntaxes of two identifier columns are apparently similar, but refer to different concepts (e.g., both APNs and tax account numbers might use numeric identifiers, but one refers to parcels and the other refers to persons). Using this approach, analysts will be able to link most parcel boundaries to assessment data in most counties (Figure S11).

- Identifying building locations: After linking assessment records to parcel boundaries, analysts can use spatial data on building footprints to identify the precise location of buildings within a parcel. A U.S.-wide open-source dataset of 130 million building footprint polygons was made available by Microsoft (2018) and has been updated multiple times since. Derived from high-resolution satellite imagery with a documented machine learning algorithm, this data is, to our knowledge, the most consistent U.S.-wide open-source indicator of building presence currently available free of charge (see also "Identifying different types of properties"). Some downsides remain: dates of observation are not always provided and can be more than a decade old in some instances. We also observe an underreporting of buildings under tree cover.

Linking transactions to time-varying property characteristics

Analyses often need to establish a reliable link between transaction prices and the characteristics of the property at the time of sale. Assessment data is an important source for these attributes, often reporting building square footage, lot size, architectural style, counts of units, rooms, bedrooms, bathrooms, as well as the presence of other features (garage, pool, etc.). Parcel boundary data can provide further information on lot size, building location, land cover, as well as access to roads, water bodies, or open space. However, these characteristics can change over time as buildings are built, remodeled, or destroyed, and as boundaries are redrawn following subdivisions or mergers. Analysts usually want to be certain that characteristics observed in assessment data and parcel boundary data are the same as those corresponding to the time period for an observation (e.g., at the time of sale).

Assessment and parcel boundary data provide only cross-sectional snapshots of property conditions at a single point in time, typically a recent one. Dataset versions for multiple years sometimes exists, and some providers of parcel boundary data also offer archives of historical data. However, synthesizing datasets from multiple time periods substantially increases data volumes and time cost for a given analysis, often with uncertain benefits. Furthermore, not all regions have historical data.

Analysts thus often consider alternative strategies to exclude transactions whose observed characteristics might not reflect those at the time of sale. Unobserved renovations between sales are particularly problematic for repeat-sales analyses, often considered a "gold standard" for evaluating property value changes (Bishop et al. 2020; Banzhaf 2021). The availability of data

on the time a building was built or remodeled varies across the United States (Figure 2, see also Leyk et al. 2020). We also observed building year values for identical properties to differ between historical and current versions of assessment data in bidirectional and idiosyncratic ways that cannot be explained by new constructions alone (Figure S12), but indicate that building year data is collected, updated, and interpreted in different ways across space and time. Filtering choices that increase the confidence in the quality of data over time (e.g., dropping counties, dropping sales, ignoring the issue) will likely affect the geographic coverage of findings (e.g., dropping observations in Vermont or Wisconsin). Satellite-based land cover change observations offer the potential of an independent detection of changes, but come with their own set of challenges, such as classification errors, insufficient resolution, or limited temporal coverage.

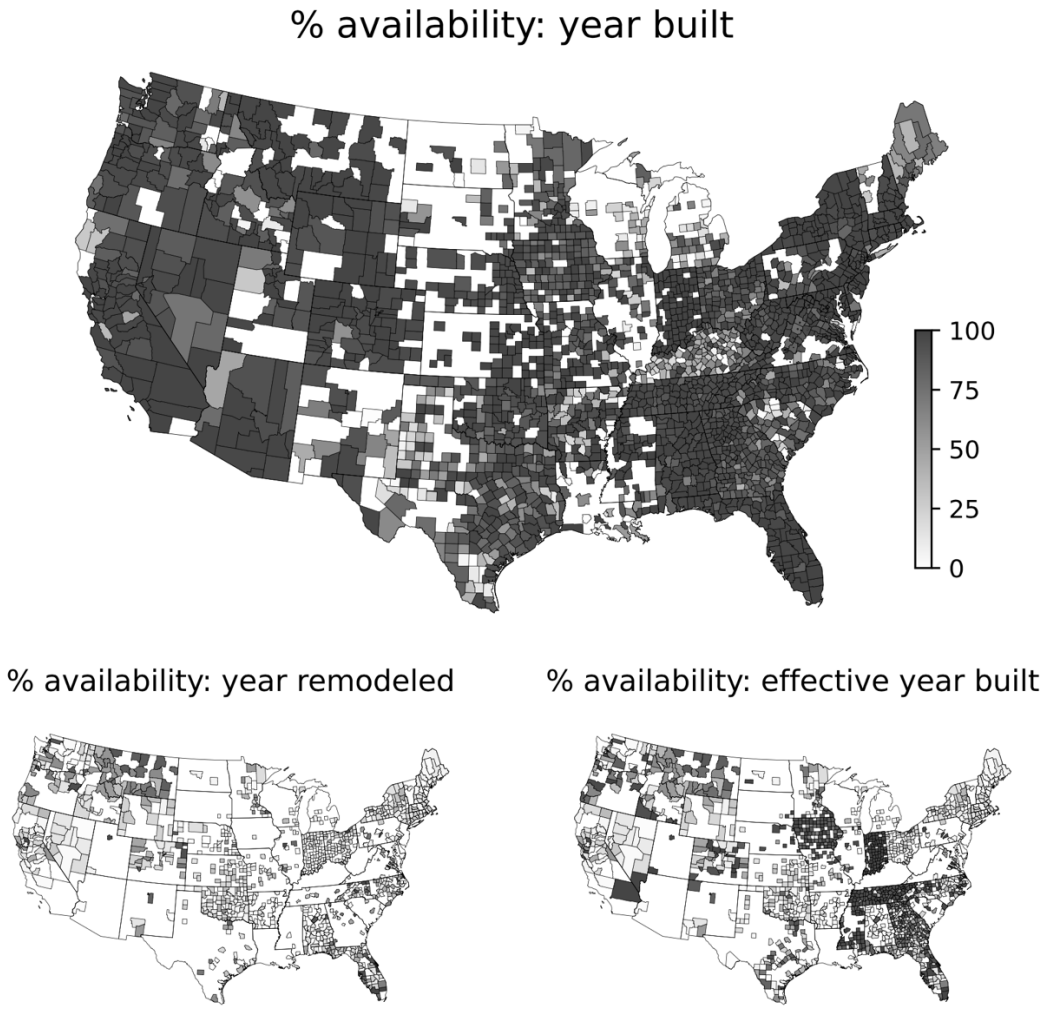

Figure 2: County-level availability of data on years of change to buildings in residential ('RR') assessment records in ZTRAX: "year built" (top), "year remodeled" (bottom left) and "effective year built" (bottom right). "Effective year built" is of unknown provenance; it could be a hybrid of "built" and "remodeled" year but might include other adjustments.

In our analyses, we consider the following solution options. Their relative utility to the analyst will depend on the application and study area. For instance, analyses of the temporal dynamics of urban growth will likely need to apply more rigorous standards than analyses of the effects of changes to nearby amenities in a stable urban core.

- Identify and account for sales with misrepresented characteristics based on years of building updates: Assessment data often contains information on (i) the year the building was first constructed and, in a minority of counties, (ii) the year of the last remodeling (Figure 2). Where both variables are available, analysts can identify transactions of properties that have been developed or remodeled since the sale. Based on available data, such sales make up 2-10% of the sample in most counties (Figure S13). There are two possible approaches to account for these transactions: (a) running the analysis after excluding such observations, and (b) controlling for an indicator of such transactions and interacting it with all hedonic variables. These approaches provide useful robustness checks that analysts can use to gauge the importance of potentially misrepresented variables in the context of their analysis. Counties that provide data on building year allow for the exclusion of new developments, but analysts will need to consider the probability of remodeling and potential biases as part of their estimation procedure. In counties where neither type of data is available, the analyst will also need to consider the extent to which new unobserved buildings might affect their results. Finally, pending a more in-depth understanding of the reasons behind idiosyncratic changes of building year data over time (Figure S12), analysts might consult with local tax assessors about the reasons for such changes, or conduct sensitivity checks that incorporate building year data from different database versions (including the database history) or external data sources such as historical maps or aerial imagery.
- Exclude sales based on remote observations: In the absence of consistent building year indicators, we considered leveraging public, satellite-based indicators of land cover change of increasing spatial-temporal resolutions and extents. Unfortunately, most nationwide historical estimates before 2013 will likely be based on products derived from medium-resolution imagery (Landsat) that are not always reliable (Brown et al. 2020) and often miss low-density development in rural, forested areas (Olofsson et al. 2016). Using the most recent public release of LCMAP, a product developed by the U.S. Geological Survey that tracks annual change to land cover between 1985 and 2017, we find low correspondence between building years in assessment data and remotely sensed transitions from undeveloped to developed land cover (Figure S14). Modern high-resolution satellites with more frequent temporal coverage (Sentinel-2, Planet Labs) will help improve observations of change. Due to the observed uncertainties associated with this approach, we currently recommend it as a robustness check only.
- Constrain the time horizon of the analysis: We expect the likelihood of unobserved changes to be higher the more time has passed since sales and the observation of property characteristics. Analysts can narrow the time horizon of the analysis by excluding sales outside a time window around the acquisition date of the property data. This likely reduces error, but at the expense of a reduced sample size, a lesser ability to observe long-term trends, and lower explanatory power of analyses estimating effects of natural events or policy changes that happened further in the past.

- Using datasets from multiple time periods: Historical assessment and parcel boundary data can sometimes be obtained from data aggregators. We have not systematically assessed the availability and quality of such historical data across the country but anticipate that both vary geographically as a function of the time at which county offices digitize their records and data aggregators expand their geographic coverage (a process that is still ongoing).

Identifying different types of properties

Analysts often need to be able to restrict the sample of transactions to specific types of properties, such as single-family homes, agricultural lands, or vacant lots. Hedonic valuation studies require the identification and delineation of a real estate market to satisfy underlying assumptions that identical properties will sell for the same price throughout that market – i.e., the "law of one price" (Bishop et al. 2020). Similarly, efforts to estimate the value of undeveloped land (e.g., Nolte 2020) need to be able to reliably identify and exclude parcels with buildings, as buildings often represent a large share of a property's value.

The availability, usage, and quality of variables used to identify properties in submarkets vary across geographies. For instance, ZTRAX contains a "property land-use standard code" with a hierarchical classification scheme, but its provenance is undocumented, and the type of properties identified with a given code varies, often across state boundaries (Figure 3). In most counties, single-family homes are identified as "RR101", but in others, "RR999" (inferred single-family), "RR000" (general residential), or "RR102" (rural residence) are also commonly used. Some counties have one code for all agricultural land ("AG000"), while others use "VL108" (Agricultural, Unimproved) to identify similar lands, or break down the agricultural land category into subcategories such as "AG101" (farm) and "AG109" (timberland / forest / trees).

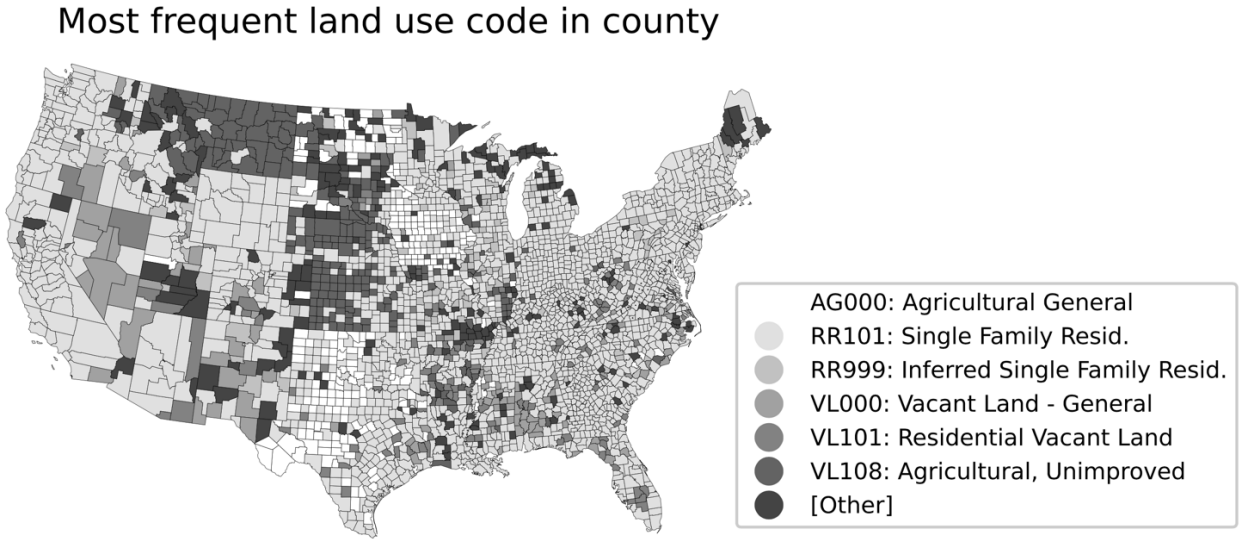

Figure 3: Most frequent land-use code across all parcels linked to ZTRAX assessment data in each county.

The identification of vacant lands is particularly challenging. For instance, ZTRAX' indicator "number of buildings" misses most buildings in hundreds of counties (Figure S15). A substantial share of parcels with "vacant" land-use codes have building footprints, and many parcels with "residential" land use codes have no building footprints (Figure S16). Alternative indicators for building presence can be derived from assessment data (e.g., property land-use codes, or the assessed value of buildings), or from remote sensing data linked to parcel boundaries (e.g., building footprints, or developed land cover). However, no single indicator is unambiguously perfect for a nationwide analysis (Figure S15).

We recommend that analysts of assessment data exercise particular caution in developing their submarket filters, test the robustness of their results to alternative plausible filtering conditions, and document their choice of filtering steps alongside published results.

- Single-family homes are likely best identified by combining several land-use codes (in ZTRAX: RR000, RR101, RR102, RR999) with indicators confirming the presence of a building, such as a positive assessed or market value for buildings, square footage, gross building area, or the presence of a remotely observed building footprint on the parcel. We note that the presence of a building does not necessarily guarantee that the building is a single-family home. We also recommend that analysts double-check whether, in their study area, land use codes are based on legal zoning or imply the presence of a building.
- Vacant parcels (parcels without any building) are most reliably identified through a combination of multiple variables. Analysts wishing to minimize the likelihood of an erroneous inclusion of buildings can exclude parcels (i) without building footprints, (ii) without a land use code indicating the presence of a building, and without a positive value for improvements in either the tax assessors' (iii) valuation or their (iv) fair market value estimates.
- The identification of agricultural parcels also benefits from combining multiple variables and can be enhanced through a judicious use of other data sources. Reducing omission and commission errors is particularly important for this category as agricultural land markets are thin, with only a small fraction of agricultural land sold annually (Bigelow et al. 2016), and small errors can lead to small sample sizes or bias. Agricultural parcels can be found under a range of land use codes; in ZTRAX, these include "AG" (agricultural), "VL" (vacant land) and "RR" (residential). Analysts filtering assessment and transaction data for agricultural sales might therefore consider additional variables – such as lot size, location, and the relative size of building footprints – as indicators of agricultural properties. Attention to regional agricultural production and institutional detail is important. For instance, in the Western United States, where irrigation is particularly important for agriculture, spatial layers for the identification of irrigated cropland from governmental sources can help distinguish between irrigated and non-irrigated agricultural areas in sample selection and analysis.

Dealing with missing or mismeasured data for standard housing attributes

Hedonic analyses need to distinguish effects of environmental attributes on property values from those of other confounding variables. Many analyses control for key characteristics of land and

buildings in a regression framework and/or through matching techniques. This requires that these characteristics are reliably and consistently observed across the study region.

The availability of standard housing attributes in assessment data varies across counties, often clustered by state. For instance, across all residential property records in ZTRAX, data gaps exist for lot size (15.1% missing), building valuation (21.4%), square footage of living area (28.6%), number of bathrooms (33.1%), number of bedrooms (53.1%) and total number of rooms (60.1%) (Figure 4). Non-sensical zero values (e.g., zero rooms, zero living area) are not uncommon. When non-zero values are observed, they can refer to different units or measurement strategies, which are not necessarily explained (e.g., frontage feet vs. lot area, square footage of building footprint vs. square footage of all floors). For instance, despite the near-complete availability of "lot size" data (Figure 4), summing the lot size of all parcels in a given county in Florida does not aggregate to the total area of the county (Clapp et al. 2018). Differences can occur both across and within jurisdictions, presumably due to variability in practices between communities or individual assessors.

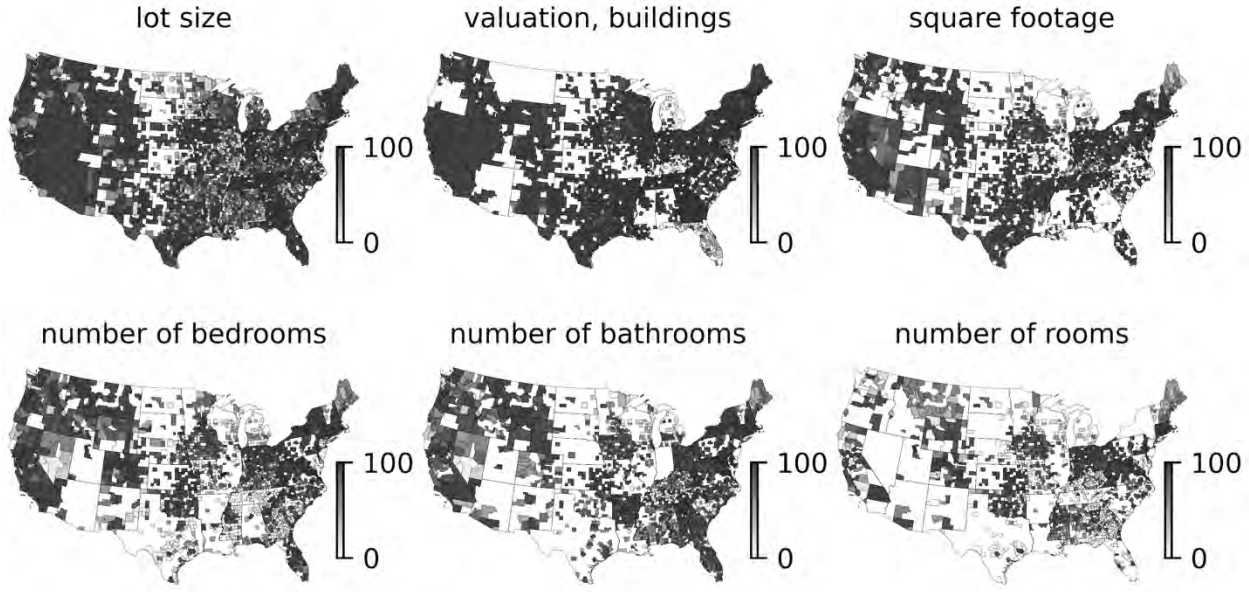

Figure 4: County-level availability (percentage of non-zero values) of standard housing indicators for residential ('RR') parcels in ZTRAX assessment records for lot size, building valuation, square footage of living area, number of bathrooms, number of bedrooms, and number of rooms.

Missing data means that researchers face sample-selection issues, while unreliable data is a measurement error issue. Both can involve trade-offs between empirical specification and geographic coverage. For instance, analysts who prefer to exclude records with missing data will likely have to work with a substantially constrained and geographically non-representative sample. A pairwise completeness analysis for a subset of attributes (Figure S17) indicates considerable heterogeneity of attribute completeness in assessment data: for example, the joint analysis of building square footage and land use type would cover a sample of 68% of its almost

150 million property records. Analysts who instead maintain those observations and somehow account for missing attribute values in their empirical models need to consider how their choice of methods might bias their estimators and affect the geographic extent of their analysis. Common examples of these methods include: using only a subset of the available measures; using spatial or temporal fixed effects, the average housing characteristics in the location, or repeat sales models to proxy for unobserved quality; using dummy variables to control for missing observations; and interpolating missing values either from available indicators, or based on out-of-sample data that contains new information (Moulton et al. 2018; Clarke & Freedman 2019; Fraenkel 2019; Gindelsky et al. 2019; Albouy et al. 2020). Analysts working with temporally varying characteristics from the historical assessment data need to be aware of data gaps resulting from sub-county level updating cycles of the underlying assessment data that lead to incompleteness patterns which vary across space and time (Figure S18a). Methods to mitigate the resulting bias may include spatial aggregation (Figure S18b) or record-level time series interpolation (Figure S18c).

A full assessment of the performance of different approaches to account for missing and mismeasured data in housing market models is beyond the scope of this study. Cameron & Trivedi (2005, chapter 26 and 27) offer a discussion of these issues and potential solutions. The key issue is understanding whether data errors are random or systematic. Determining how data errors vary spatially will help analysts account for these issues in their empirical specification. We recommend that, even in the presence of time constraints, analysts dedicate a significant amount of time to data inspection, robustness checks, and a full documentation of choices and findings. Data inspection can range from simple "sanity checks" (e.g., verifying the plausibility of values with histograms, checking for unexpected clustering with maps) to more systematic testing such as calculating correlations between data issues and either the outcomes of interest, observable characteristics (e.g., jurisdiction, income, race, etc.), or matched external validation data (e.g., lot sizes from parcel boundaries, jurisdiction, income, race, etc.). The appropriate data inspection approach should be guided by the analyst's research question and design. For instance, the pattern of missing and mismeasured data that cause bias are likely to be different between cross-sectional and panel or difference-in-difference models (e.g., see the discussion in Zhang et al. (2022)). Analysts should also include a suite of robustness checks involving different plausible combinations of data filters and models and examine the sensitivity of findings to their choices. Most importantly, we recommend that analysts fully document their sampling procedure, including choices of inclusion vs. exclusion of observations and attributes based on data availability and reliability, and the implications of that choice on the geographic coverage of findings.

Choices reported in the literature

To what extent do current hedonic analyses of environmental attributes already acknowledge and address these challenges? To answer this question, we reviewed data filtering, processing, and modeling choices reported in the 27 peer-reviewed journal articles that used ZTRAX to value an environmental attribute using hedonic property value methods and were published by August 10, 2022 (Table S1). We identified this sample by searching for the term “ZTRAX” in Google Scholar and retaining from the resulting 320 records all studies which met these criteria. We

checked whether each study reported undertaking any step from a list of 35 individual processing steps across the five previously discussed challenges (i.e., observed positives, see Figure 5).

On average, we find that reviewed studies report only a small fraction of the proposed steps (average: 5.85 of 35 potential steps; range: 1 to 11) (Figure 5). While small, this number does not by itself cast doubt on the validity of the findings of any given study, for at least three reasons: authors might have implemented a step without reporting it; not all steps are necessary in every analysis (e.g., missing data can be negligible in a given study region); and some steps are substitutes (e.g., filtering out implausible coordinates vs. linking records to parcel boundaries). However, our review also suggests that some peer-reviewed articles might have cut short the full reporting of relevant choices inherent in the analysis of large-scale property data, creating additional and likely unnecessary barriers to reproducibility and replication.

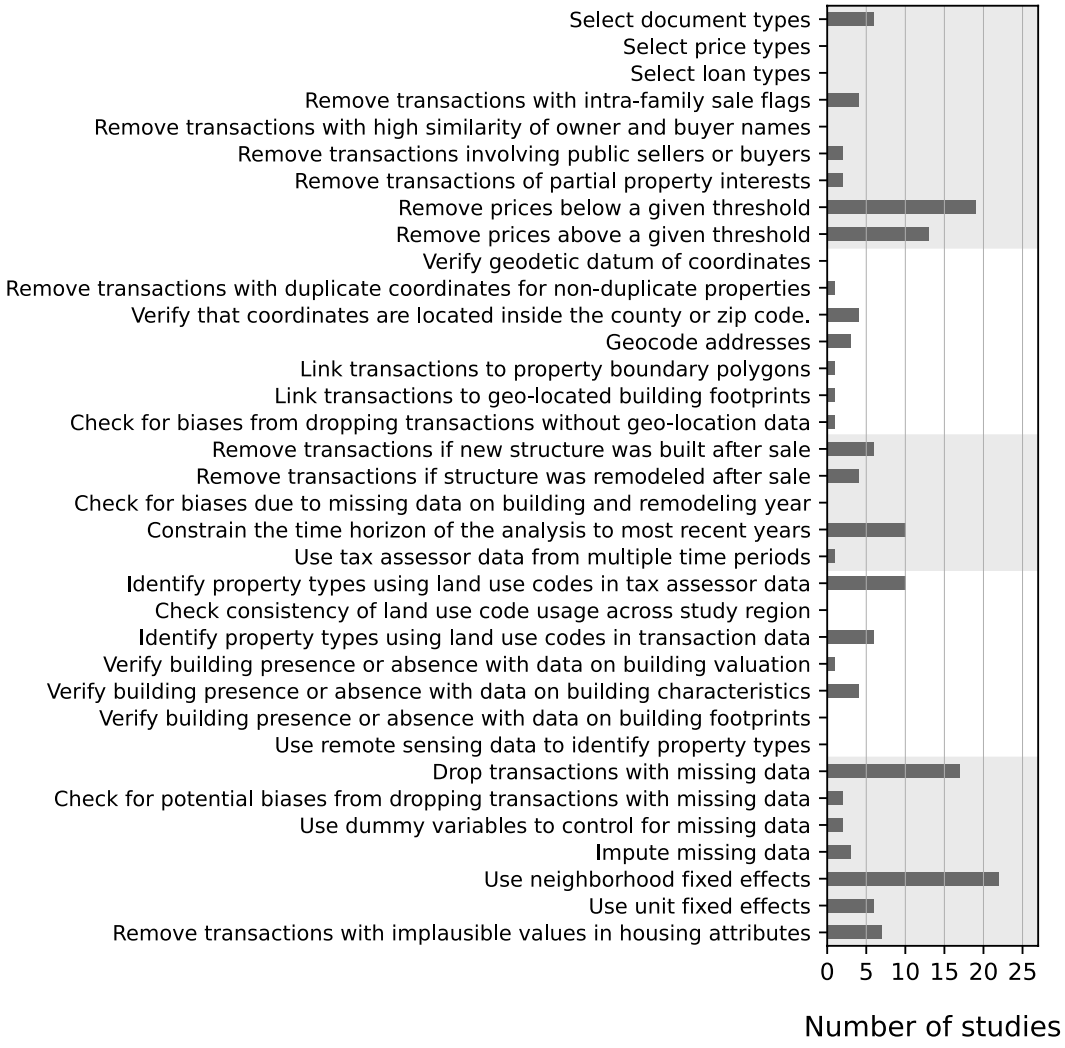

Figure 5: Number of peer-reviewed studies published by Aug 10, 2022, that report results from an environmental hedonic analysis based on ZTRAX data (total count: 27) and report to have implemented a given data processing step.

In terms of individual challenges, we find:

- Arm's-length sales filters: upper and lower thresholds on sales prices are the most frequently reported type of filter. Thresholds vary widely (lower: $1-100K; upper: $1-10M or top 0.5-5%). No study reports having excluded sales based on a high similarity between buyer and seller names or based on price types or loan types.
- Geo-location: 23 studies (85%) use property coordinates to measure spatial relationships to the environmental attribute of interest, but only eight report to have taken any step to address potential geo-location issues, and none report to have verified the geodetic datum of coordinates. Authors of three studies chose to ignore property coordinates provided in assessment or transaction data, geo-coding street addresses instead.
- Time-varying characteristics: a few studies report to have excluded sales that occurred before a house was built (n=6) or remodeled (n=4). It is also common for studies to reduce the time horizon to a more recent time period (e.g., 10 studies chose ≥ 2005).
- Property types: 26 of 27 studies focus on residential properties, predominantly single-family homes. However, only a fraction report how the sample was selected, and only four take any steps to examine omission errors (e.g., by not including the full set of building codes) or commission errors (e.g., by verifying that a building exists).
- Missing and mismeasured housing attributes: most studies include some housing attributes in their hedonic regression (n=25), including living area (n=19), age (19), bedrooms (18), bathrooms (18), and lot size (15). The most common choice to deal with missing attributes is to drop observations with missing data (n=17), whereas only two studies add missing value indicators and recode missing values as zeros. Most analyses also include neighborhood fixed effects to control for unobserved attributes (n=22). The smallest spatial scale of these fixed effects varies across studies: ZIP codes are most frequent (n=6), followed by block group (n=3), tracts (n=3), and counties (n=3).

Sensitivity of hedonic coefficients to data preparation choices: an illustration

Are the results of hedonic studies sensitive to whether and how an analyst chooses to address the five challenges we outline? The answer to this question depends on many factors specific to a given study, including its objective, geographic scope, inferential strategy, and coefficients of interest. We therefore cannot address it comprehensively. Instead, we use an illustrative case study to explore whether data processing choices matter in at least one application of interest.

Our case study focuses on the property price effects of being located inside a special flood hazard area (SFHA, 100-year flood zone), as mapped by the Federal Emergency Management Agency (FEMA) (Bin & Kruse 2006; Bin et al. 2008; Beltrán et al. 2018). Because we are interested in highlighting the importance of county-level variation in data availability and quality, we estimate this effect separately for each county within the CONUS.

In each county, we estimate the following log-linear regression:

$$ln(price_{ijt}) = \alpha + \delta\, SFHA_i + X_i\beta + \mu_j + \tau_t + \varepsilon_{ijt} \quad (1)$$

where $price_{ijt}$ is the sales price of property i in neighborhood j at time t. The indicator variable of interest is $SFHA_i$, which is 1 if the sold property was located inside the SFHA ("treated"), and 0 if FEMA considered the property to be located outside the SFHA ("control"; unmapped areas are excluded). Therefore, δ is the coefficient of interest. X_i contains property-level characteristics, μ_j are neighborhood (spatial) fixed effects, τ_t are year-quarter fixed effects, and ε_{ijt} is an error term. Definitions of $SFHA_i$, X_i, and μ_j vary across our sensitivity checks:

- $SFHA_i$: in our default model, $SFHA_i$ is based on Microsoft building footprints: it is 1 ("treated") if the centroid of the largest building footprint on a given parcel is located inside the SFHA, and 0 otherwise. We include only properties with 1 or 2 footprints. As sensitivity checks, we derive $SFHA_i$ from either parcel boundary centroids or assessment data coordinates (Oct 2021 version of ZTRAX); in both cases, we also assume that the analyst did not use any building footprint data to drop observations without observable buildings or with more than two building footprints.
- X_i includes lakefront and riverfront indicators to control for the amenity of water access. We derive both from spatial proximity of parcel boundaries and waterbody polygons from the National Hydrography Dataset (U.S. Geological Survey 2017). In our default model, X_i also includes bedroom and bathroom dummies (i.e., dummy variables for 1, 2, … 10 bedrooms and 0.5, 1, … 10 bathrooms). We vary X_i across two sensitivity checks: one that drops bedroom and bathroom dummies and one that adds total living area as a supplementary control. Our default run keeps sales of properties with missing bedroom and bathroom data, and adds separate dummies for missing values, zero values, and values above 10 to each (i.e., 2 x 3 dummy variables). A sensitivity check drops observations with missing bedroom or bathroom data or zero values in either field.
- In our default model, μ_j stands for census-tract fixed effects. As sensitivity checks, we also switch to ZIP code (coarser), block-group (finer), and no spatial fixed effects. All spatial units are derived from 2016 NHGIS data (Manson et al. 2018).

The inclusion of sales in the regression is subject to multiple filters. In all cases, we select sales of single-family homes that occurred in or after 2000 or after the most recent update to the SFHA in their county, whichever is later, and before October 1st, 2021. Due to well-documented difficulties in separating the negative price effects of coastal flood risk from the positive price effects of coastal amenities (Beltrán et al. 2018; Johnston & Moeltner 2019), we exclude sales located within 2.5km of an ocean coast. In addition:

- Our default model only includes sales that pass all of our "high-confidence" arm's-length filters. In two sensitivity checks, we also include "medium" and "low-confidence" sales. In addition, we test the effects of using only a lower price threshold (≥ $1001), the most frequently reported filter in our literature review.
- Our default model drops sales of properties whose buildings were known to be built in the same year or after the sales transaction occurred. A sensitivity check drops this filter.
- Our default model uses several document codes to identify single-family homes (RR000, RR101, RR102, and RR999). A sensitivity check uses only the simple single-family flag (RR101).

To illustrate the joint importance of multiple decisions, we also derive a "current literature" scenario with a set of changes to our default model that we consider representative of steps reported in the existing ZTRAX-based literature: 1) the only arm's-length filter used is a lower price cutoff ($1001), 2) treatment identification ($SFHA_i$) relies on property coordinates in the assessment data, 3) no building footprint data is used to select the sample, 4) sales with empty or extreme values for bathrooms and bedrooms are dropped, and 5) sales with new buildings since the transaction are kept.

After fitting each model at the county-level, we keep results from all county-level models with a minimum of statistical support, which we define here as being estimated on a sample containing at least 100 identifying treated and 100 identifying control sales. With "identifying", we mean that we count only sales belonging to categories (lakefront, waterfront, year quarter, bedroom count, bathroom count, and spatial fixed effect, if applicable) that exhibit treatment heterogeneity (i.e., that contain both treated and control sales). While our default model retains results from 297 counties, this count can range from 219 counties when dropping sales with empty or extreme bedroom or bathroom counts (variables for which data is often missing, see Figure 4) to 409 counties when adding in transactions that had not passed our arm's-length filters.

We compute all county-level differences in the estimated coefficient of interest ($\hat{\delta}$) between our default model ($\widehat{\delta_{ref}}$) and each alternative model specification with the same minimum of statistical support. To assess the nationwide magnitude of the effect of processing choices on the estimated SFHA discounts, we also report: (i) the percentage of counties in which a county-level study would have led to different conclusions regarding the statistical significance of the SFHA discount (at $\alpha = 0.05$), as well as (ii) the averages of county-level estimates of $\hat{\delta}$, weighted equally by county ($\hat{\delta}_{avg}$).

Our results demonstrate that each processing choice has discernible effects on the magnitude and significance of $\hat{\delta}$ or the geographic coverage in our application (Figure 6). Importantly, we find that the effects of choices that are rarely reported in the peer-reviewed hedonic literature – such as the choice of arm's-length filters, geo-location precision, or the removal of pre-building sales – can be of similar magnitude as the effects of choices that are more commonly reported, such as the dropping of missing-data observations or the use of different spatial fixed effects. For example, 17% of county-level estimates cross the statistical significance threshold of α = 5% (i.e., switch significance in either direction) when using only a minimum-price filter for arm's-length sales, 20% when switching to property coordinates in assessment data instead of building footprints to assign SFHA treatment, and 12% when not removing pre-building sales. These counts are of a similar order of magnitude as those observed when switching to ZIP code fixed effects (18% of county effects switch significance) or block group fixed effects (16%), and substantially larger than the consequences of dropping missing-data observations (3%). Similarly, effects on the (county-weighted) magnitude of the SFHA discount can be large: using only a minimum-price filter for arm's-length sales increases the absolute value of the discount, ($|\hat{\delta}_{avg}|$) by 8%, switching to property coordinates in the assessment data reduces it by 36%, and not removing pre-building sales increases it by 13%. Meanwhile, switching to block-group fixed effects decreases its absolute value by 11%, switching to ZIP-code fixed effects increases it by 13%, and the dropping of missing-data observations increases it by 9%.

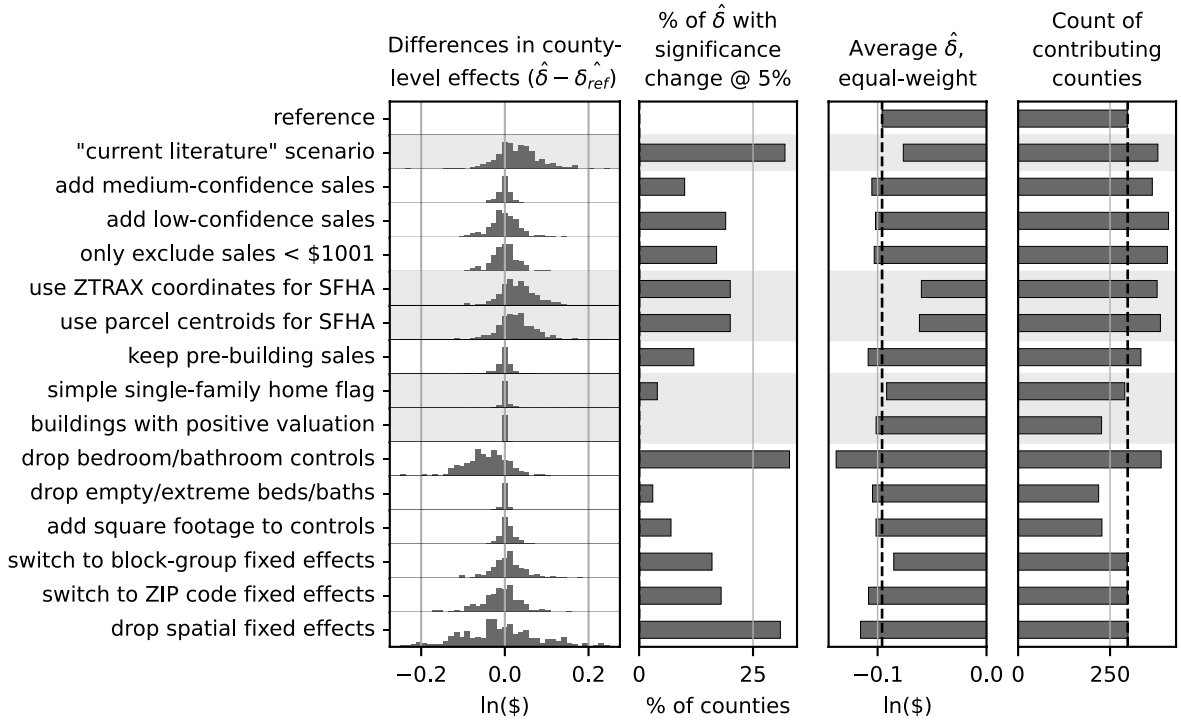

Figure 6: Effects of variations in data filtering, processing, and modeling choices on estimated flood zone discounts. Dashed lines indicate the value of the reference model.

Common strategies to strengthen the robustness of empirical estimates do not fully remove the observed sensitivity of $\hat{\delta}$ to data processing choices. For instance, if we reduce our sample to the counties with particularly strong statistical support (≥ 500 identifying treatment and control sales, n=73), our estimates remain sensitive to most specifications (Figure S19). Similarly, if we only retain counties whose estimates of $\hat{\delta}$ are robust to variations in fixed effects (defined here as $\left|\hat{\delta} - \widehat{\delta_{ref}}\right| < 0.02$ when switching to block group and ZIP code fixed effects, n=69), many estimates remain sensitive to data processing choices about arms-length filters and geo-location (Figure S20). A third potential strategy to enhance the robustness of estimates is to pool the data across larger geographic units, such as states. We therefore repeat our full analysis at the state level for all states that contain at least 500 identifying treatment and control sales (n=32). While we find that this approach greatly reduces the number of changes to the statistical significance of state-level estimates when changing spatial fixed effects (only one estimate switches, $\alpha = 0.05$), several results remain affected by choices on arms-length filters and geolocation (Figure S21).

While a full assessment of the mechanisms behind the observed sensitivities is beyond the scope of this paper, we search for potential reasons by closely examining the data for a small set of counties whose estimates of $\hat{\delta}$ are particularly sensitive to data processing choices in spite of strong statistical support. In Montgomery (Pennsylvania), Polk (Florida) and Fulton (Georgia) counties, we find that many property boundary polygons are defined such that their centroids fall inside the SFHA, although their buildings are located outside (Figure S22). This leads to an underestimation of the absolute value of $\hat{\delta}$ when using parcel centroids to allocate treatment – a

finding in line with Netusil et al. (2019). In Rock Island county (Illinois), we find that relying on coordinates in assessor data to identify SFHA location overestimated the absolute value of $\hat{\delta}$ because the county's assessor data has numerous missing and erroneous coordinates (ZIP code centroids, owner mailing addresses) whose flood zone location was associated with sales prices. Within flood zones, erroneously located sales had lower average prices than correctly located ones, while outside flood zones, erroneously located properties had higher average prices than correctly located ones. Finally, in Bucks county (Pennsylvania), we find that using lower price cutoffs as the only arm's-length filter overestimated $\hat{\delta}$ due to the inclusion of a large number of intra-family and forced sales (sheriff's deeds, executioner's deeds) whose response to flood zone location differed from those of arm's-length sales.

Taken together, these findings suggest that underreported data processing choices are not inconsequential for the results of hedonic analyses. We acknowledge an important caveat: our illustrative case study is not intended to be representative for all contexts. For instance, hedonic estimates of the location of buildings inside vs. outside of discrete flood zone boundaries are likely more sensitive to small spatial errors than estimates of the value of environmental amenities with less discrete spatial variation, such as air pollution or recreational access. However, until future empirical work begins to examine the relative importance of those data processing choices in other contexts, a more consistent and transparent reporting in the published literature can help improve shared scrutiny and advance scientific progress.

Conclusion

Large-scale property transaction and assessment data offer unprecedented opportunities for detailed empirical research into the dynamics of land ownership, land policy, property valuation and non-market valuation in the United States. After conclusion of the ZTRAX program, analysts who work with and compile data from county and state-level data sources, or who purchase similar data services from third-party providers, will be confronted by many of the issues discussed in this article. Awareness of potential errors and biases, fuller documentation of data processing and filtering choices, and discussion of the potential effects of geographic omissions will enhance the transparency, replicability, and generalizability of empirical findings. Therefore, we encourage journal editors and referees to require that authors include detailed documentation of data processing choices in their final manuscript submissions. In the absence of official best practice standards, this article can serve as a non-exhaustive checklist of potential issues.

Supplementary Material

Supplementary material includes a table of reviewed studies (Table S1), figures S1-S22 and text (descriptions of deeds). Filtering tables, search terms (regular expressions) for public buyer and seller names, and imputed geodesic datums are provided as separate Data Sets. Python code to compute owner name similarities is posted on: http://placeslab.org/hedonic-data-practices.

Acknowledgements

Data provided by Zillow through the Zillow Transaction and Assessment Dataset (ZTRAX). More information can be found at http://www.zillow.com/ztrax. The results and opinions are those of the author(s) and do not reflect the position of Zillow Group.

Parcel data for approximately two thirds of U.S. counties was provided by Regrid through its "Data with Purpose" program (https://regrid.com/purpose).

We thank participants of the ZTRAX best data practice workshop (2021 PLACES webinar, Jun 28-30, 2021) and the workshop of the ZTRAX Joint Special Issue in Land Economics and the Journal of Housing Economics (August 29-31, 2022), as well as two anonymous reviewers, for useful feedback on prior versions of this article. Christoph Nolte, Adam Pollack, Ido Kushner, and Shelby Sundquist acknowledge support from the Department of Earth & Environment at Boston University, the Junior Faculty Fellows program of Boston University's Hariri Institute for Computing and Computational Science, the Nature Conservancy, and the National Science Foundation's (NSF) Human-Environment and Geospatial Sciences (HEGS) program (grant #2149243). Johannes Uhl is funded, in part, by NSF's Humans, Disasters and the Built Environment (HDBE) program (grant #1924670).

References

Albouy, D., Christensen, P. & Sarmiento-Barbieri, I. (2020). Unlocking amenities: Estimating public good complementarity. *J. Public Econ.*, 182, 104110.

Baldauf, M., Garlappi, L. & Yannelis, C. (2020). Does Climate Change Affect Real Estate Prices? Only If You Believe In It. *Rev. Financ. Stud.*, 33, 1256–1295.

Banzhaf, H.S. (2021). Difference-in-Differences Hedonics. *J. Polit. Econ.*, 129, 000–000.

Beltrán, A., Maddison, D. & Elliott, R.J.R. (2018). Is Flood Risk Capitalised Into Property Values? *Ecol. Econ.*, 146, 668–685.

Bernstein, A., Gustafson, M.T. & Lewis, R. (2019). Disaster on the horizon: The price effect of sea level rise. *J. financ. econ.*, 134, 253–272.

Bigelow, D., Borchers, A. & Hubbs, T. (2016). *U.S. Farmland Ownership, Tenure, and Transfer*. U.S. Department of Agriculture, Economic Research Service.

Bin, O. & Kruse, J.B. (2006). Real Estate Market Response to Coastal Flood Hazards. *Nat. Hazards Rev.*, 7, 137–144.

Bin, O., Kruse, J.B. & Landry, C.E. (2008). Flood Hazards, Insurance Rates, and Amenities: Evidence From the Coastal Housing Market. *J. Risk Insur.*, 75, 63–82.

Bishop, K.C., Kuminoff, N. V., Banzhaf, H.S., Boyle, K.J., von Gravenitz, K., Pope, J.C., Smith, V.K. & Timmins, C.D. (2020). Best Practices for Using Hedonic Property Value Models to Measure Willingness to Pay for Environmental Quality. *Rev. Environ. Econ. Policy*, 14, 260–281.

Brown, J.F., Tollerud, H.J., Barber, C.P., Zhou, Q., Dwyer, J.L., Vogelmann, J.E., Loveland, T.R., Woodcock, C.E., Stehman, S. V., Zhu, Z., Pengra, B.W., Smith, K., Horton, J.A., Xian, G., Auch, R.F., Sohl, T.L., Sayler, K.L., Gallant, A.L., Zelenak, D., Reker, R.R. & Rover, J. (2020). Lessons learned implementing an operational continuous United States national land change monitoring capability: The Land Change Monitoring, Assessment, and Projection (LCMAP) approach. *Remote Sens. Environ.*, 238, 111356.

Cameron, A.C. & Trivedi, P.K. (2005). *Microeconometrics. Methods and Applications*. Cambridge University Press, Cambridge, UK.
Christensen, G. & Miguel, E. (2018). Transparency, Reproducibility, and the Credibility of Economics Research. *J. Econ. Lit.*, 56, 920–980.
Clapp, C.M., Freeland, J., Ihlanfeldt, K. & Willardsen, K. (2018). The Fiscal Impacts of Alternative Land Uses. *Public Financ. Rev.*, 46, 850–878.
Clarke, W. & Freedman, M. (2019). The rise and effects of homeowners associations. *J. Urban Econ.*, 112, 1–15.
Ferraro, P.J. & Shukla, P. (2020). Feature—Is a Replicability Crisis on the Horizon for Environmental and Resource Economics? *Rev. Environ. Econ. Policy*, 14, 339–351.
Ferraro, P.J. & Shukla, P. (2022). Credibility crisis in agricultural economics. *Appl. Econ. Perspect. Policy*, 1–17.
Fraenkel, R. (2019). Property Tax-Induced Mobility and Redistribution : Evidence from Mass Reappraisals ∗.
Friedl, J. (2006). *Mastering regular expressions*. O'Reilly Media, Inc, Sebastopol, USA.
Gindelsky, M., Moulton, J. & Wentland, S. (2019). Valuing Housing Services in the Era of Big Data: A User Cost Approach Leveraging Zillow Microdata. *Int. Conf. Real Estate Stat.*
Gourevitch, J.D., Kousky, C., Liao, Y., Nolte, C., Pollack, A.B., Porter, J.R. & Weill, J.A. (2023). Unpriced climate risk and the potential consequences of overvaluation in US housing markets. *Nat. Clim. Chang.*, Submitted.
Guignet, D., Heberling, M.T., Papenfus, M. & Griot, O. (2022). Property Values, Water Quality, and Benefit Transfer: A Nationwide Meta-analysis. *Land Econ.*, 98, 191–218.
Guignet, D. & Nolte, C. (2021). *Hazardous Waste and Home Values: An Analysis of Treatment and Disposal Sites in the U.S.* Boone, NC.
Hanley, N. & Czajkowski, M. (2019). The Role of Stated Preference Valuation Methods in Understanding Choices and Informing Policy. *Rev. Environ. Econ. Policy*, 13, 248–266.
Huntington-Klein, N., Arenas, A., Beam, E., Bertoni, M., Bloem, J.R., Burli, P., Chen, N., Grieco, P., Ekpe, G., Pugatch, T., Saavedra, M. & Stopnitzky, Y. (2021). The influence of hidden researcher decisions in applied microeconomics. *Econ. Inq.*, 1–17.
Johnston, R.J. & Moeltner, K. (2019). Special Flood Hazard Effects on Coastal and Interior Home Values: One Size Does Not Fit All. *Environ. Resour. Econ.*, 74, 181–210.
Leyk, S., Uhl, J.H., Connor, D.S., Braswell, A.E., Mietkiewicz, N., Balch, J.K. & Gutmann, M. (2020). Two centuries of settlement and urban development in the United States. *Sci. Adv.*, 6, eaba2937.
Mamun, S., Castillo-Castillo, A., Swedberg, K., Zhang, J., Boyle, K., Cardoso, D., Kling, C.L., Nolte, C., Papenfus, M., Phaneuf, D. & Polasky, S. (2023a). Valuing water quality in the US using a national data set on property values. *Proc. Natl. Acad. Sci. U. S. A.*, in press.
Mamun, S., Nolte, C. & Polasky, S. (2023b). Nationwide lake impairment and property value. *Work. Pap.*
Maniadis, Z., Tufano, F. & List, J.A. (2014). One Swallow Doesn't Make a Summer: New Evidence on Anchoring Effects. *Am. Econ. Rev.*, 104, 277–290.
Maniadis, Z., Tufano, F. & List, J.A. (2017). To Replicate or Not to Replicate? Exploring Reproducibility in Economics through the Lens of a Model and a Pilot Study. *Econ. J.*, 127, F209–F235.
Manson, S., Schroeder, J., Van Riper, D. & Ruggles, S. (2018). *IPUMS National Historical Geographical Information System: Version 13.0 [Database].* Minneapolis.

Microsoft. (2018). U.S. Building Footprints [WWW Document]. URL https://github.com/microsoft/USBuildingFootprints

Moulton, J.G., Sanders, N.J. & Wentland, S.A. (2018). Toxic Assets : How the Housing Market Responds to Environmental Information Shocks. *Work. Pap.*

Murfin, J. & Spiegel, M. (2020). Is the Risk of Sea Level Rise Capitalized in Residential Real Estate? *Rev. Financ. Stud.*, 33, 1217–1255.

Netusil, N.R., Moeltner, K. & Jarrad, M. (2019). Floodplain designation and property sale prices in an urban watershed. *Land use policy*, 88, 104112.

Nolte, C. (2020). High-resolution land value maps reveal underestimation of conservation costs in the United States. *Proc. Natl. Acad. Sci.*, 117, 29577–29583.

Olofsson, P., Holden, C.E., Bullock, E.L. & Woodcock, C.E. (2016). Time series analysis of satellite data reveals continuous deforestation of New England since the 1980s. *Environ. Res. Lett.*, 11, 064002.

Simmons, J.P., Nelson, L.D. & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci.*, 22, 1359–1366.

U.S. Census Bureau. (2021). U.S. Census Geocoder [WWW Document]. URL https://geocoding.geo.census.gov/geocoder/

U.S. Geological Survey. (2017). National Hydrography Dataset (NHD) USGS National Map [WWW Document]. URL https://nhd.usgs.gov/NHD_High_Resolution.html

Wentland, S.A., Ancona, Z.H., Bagstad, K.J., Boyd, J., Hass, J.L., Gindelsky, M. & Moulton, J.G. (2020). Accounting for land in the United States: Integrating physical land cover, land use, and monetary valuation. *Ecosyst. Serv.*, 46, 101178.

Zhang, J., Phaneuf, D.J. & Schaeffer, B.A. (2022). Property values and cyanobacterial algal blooms: Evidence from satellite monitoring of Inland Lakes. *Ecol. Econ.*, 199, 107481.

Zillow. (2021a). ZTRAX: Zillow Transaction and Assessor Dataset [WWW Document]. URL http://www.zillow.com/research/ztrax

Zillow. (2021b). ZTRAX: Frequently Asked Questions [WWW Document]. URL https://www.zillow.com/research/ztrax/ztrax-faqs/

Supplementary Material

Data Sets
1. Excel table of fair-market value filters
2. Excel table of regular expressions to identify public sellers and buyers
3. Comma-separated value (CSV) table of estimated geographic datums by county (Oct 2019)

Supplementary Table S1

Supplementary Figures S1-S22

Supplementary Text
1. Deed definitions

Digital Material (https://placeslab.org/hedonic-data-practices):
1. Python script to compute name similarity

Supplementary Table

Table S1: List of ZTRAX-based environmental hedonic analyses published in a peer-reviewed journal by Aug 10, 2022

Authors and year	Title	Attribute of interest	Location
Acolin et al. 2022	How do single-family homeowners value residential and commercial density? It depends	building density	five MSAs
Albouy et al. 2020	Unlocking amenities: Estimating public good complementarity	crime * parks	3 cities
Baldauf et al. 2020	Does Climate Change Affect Real Estate Prices? Only If You Believe In It	flood risk * belief in climate change	national
Bechard 2021	Gone with the Wind: Declines in Property Values as Harmful Algal Blooms Are Blown Towards the Shore	algal blooms coast	Florida Gulf
Bernstein et al. 2019	Disaster on the horizon: The price effect of sea level rise	sea level rise	national
Boslett & Hill 2019	Shale gas transmission and housing prices	shale gas pipeline	New York
Casola et al. 2022	Measuring the value of public hunting land using a hedonic approach	public hunting land	North Carolina
Chapple & Jeon 2021	Big Tech on the Block: Examining the Impact of Tech Campuses on Local Housing Markets in the San Francisco Bay Area	tech campuses	Silicon Valley
Chun et al. 2021	Are Foreclosure Spillover Effects Universal? Variation Over Space and Time	foreclosures	3 Ohio metros
Clarke & Freedman 2019	The rise and effects of homeowners' associations	homeowner assoc.	national
D'Lima & Schultz 2021	Residential Real Estate Investments and Investor Characteristics	investors	national
D'Lima & Schultz 2022	Buy-to-Rent Investors and the Market for Single Family Homes	buy-to-rent investors	national
D'Lima & Thibodeau 2022	Health crisis and housing market effects - evidence from the U.S. opioid epidemic	opioid dispensaries	Ohio

Dong & Lang 2022	Do views of offshore wind energy detract? A hedonic price analysis of the Block Island wind farm in Rhode Island	offshore wind energy	Rhode Island
Gindelsky et al. 2022	Valuing Housing Services in the Era of Big Data: A User Cost Approach Leveraging Zillow Microdata	housing services	national
Kuhlmann 2021	Upzoning and Single-Family Housing Prices: A (Very Early Analysis of the Minneapolis 2040 Plan	zoning changes	Minneapolis
Leonard et al. 2021	The impact of land use regulation across the conditional distribution of home prices: an application of quantile regression for group-level treatments	zoning changes	national
Lin et al. 2022	The price effects of greening vacant lots: How neighborhood attributes matter	greening vacant lots	Philadelphia
Miller & Pinter 2022	Flood risk and residential real-estate prices: Evidence from three US counties	floodplains * flood events	three counties
Moore et al. 2020	Hedonic Price Estimates of Lake Water Quality: Valued Attribute, Instrumental Variables, and Ecological-Economic Benefits	lake water clarity	113 lakes
Murfin & Spiegel 2020	Is the Risk of Sea Level Rise Capitalized in Residential Real Estate?	sea level rise	national
Richardson et al. 2022	Valuation of Wetland Restoration: Evidence from the Housing Market in Arkansas	wetland easements	Arkansas
Rivera & Loveridge 2022	Coal-to-Gas Fuel Switching and its Effects on Housing Prices	coal-to-gas switch	national
Wentland et al. 2020	Accounting for land in the United States: Integrating physical land cover, land use, and monetary valuation.	diverse	national
Zhang & Leonard 2021	External validity of hedonic price estimates: Heterogeneity in the price discount associated with having Black and Hispanic neighbors	Black and Hispanic neighborhoods	national

Zhang et al. 2022	Property values and cyanobacterial algal blooms: Evidence from satellite monitoring of Inland Lakes	algal blooms lake	national
Zheng 2022	The Valuation of Local School Quality under School Choice	school choice	multi-state

References

Acolin, A., Colburn, G. & Walter, R.J. (2022). How do single-family homeowners value residential and commercial density? It depends. *Land use policy*, 113, 105898.

Albouy, D., Christensen, P. & Sarmiento-Barbieri, I. (2020). Unlocking amenities: Estimating public good complementarity. *J. Public Econ.*, 182, 104110.

Baldauf, M., Garlappi, L. & Yannelis, C. (2020). Does Climate Change Affect Real Estate Prices? Only If You Believe In It. *Rev. Financ. Stud.*, 33, 1256–1295.

Bechard, A. (2021). Gone with the Wind: Declines in Property Values as Harmful Algal Blooms Are Blown Towards the Shore. *J. Real Estate Financ. Econ.*, 62, 242–257.

Bernstein, A., Gustafson, M.T. & Lewis, R. (2019). Disaster on the horizon: The price effect of sea level rise. *J. financ. econ.*, 134, 253–272.

Boslett, A. & Hill, E. (2019). Shale gas transmission and housing prices. *Resour. Energy Econ.*, 57, 36–50.

Casola, W.R., Peterson, M.N., Wu, Y., Sills, E.O., Pease, B.S. & Pacifici, K. (2022). Measuring the value of public hunting land using a hedonic approach. *Hum. Dimens. Wildl.*, 27, 343–359.

Chapple, K. & Jeon, J.S. (2021). Big Tech on the Block: Examining the Impact of Tech Campuses on Local Housing Markets in the San Francisco Bay Area. *Econ. Dev. Q.*, 35, 351–369.

Chun, Y., Pierce, S.C. & Van Leuven, A.J. (2021). Are Foreclosure Spillover Effects Universal? Variation Over Space and Time. *Hous. Policy Debate*, 31, 924–946.

Clarke, W. & Freedman, M. (2019). The rise and effects of homeowners associations. *J. Urban Econ.*, 112, 1–15.

D'Lima, W. & Schultz, P. (2021). Residential Real Estate Investments and Investor Characteristics. *J. Real Estate Financ. Econ.*, 63, 354–393.

D'Lima, W. & Schultz, P. (2022). Buy-to-Rent Investors and the Market for Single Family Homes. *J. Real Estate Financ. Econ.*, 64, 116–152.

D'Lima, W. & Thibodeau, M. (2022). Health Crisis and Housing Market Effects - Evidence from the U.S. Opioid Epidemic. *J. Real Estate Financ. Econ.*

Dong, L. & Lang, C. (2022). Do views of offshore wind energy detract? A hedonic price analysis of the Block Island wind farm in Rhode Island. *Energy Policy*, 167, 113060.

Gindelsky, M., Moulton, J.G. & Wentland, S.A. (2022). Valuing Housing Services in the Era of Big Data. A User Cost Approach Leveraging Zillow Microdata. In: *Big Data Twenty-First-Century Econ. Stat.* (eds. Abraham, K.G., Jarmin, R.S., Moyer, B.C. & Shapiro, M.D.). University of Chicago Press, Chicago.

Kuhlmann, D. (2021). Upzoning and Single-Family Housing Prices. *J. Am. Plan. Assoc.*, 87, 383–395.

Leonard, T., Yang, X. & Zhang, L. (2021). The impact of land use regulation across the conditional distribution of home prices: an application of quantile regression for group-level

treatments. *Ann. Reg. Sci.*, 66, 655–676.
Lin, D., Jensen, S.T. & Wachter, S.M. (2022). The price effects of greening vacant lots: How neighborhood attributes matter. *Real Estate Econ.*
Miller, R.G. & Pinter, N. (2022). Flood risk and residential real-estate prices: Evidence from three US counties. *J. Flood Risk Manag.*, 15.
Moore, M.R., Doubek, J.P., Xu, H. & Cardinale, B.J. (2020). Hedonic Price Estimates of Lake Water Quality: Valued Attribute, Instrumental Variables, and Ecological-Economic Benefits. *Ecol. Econ.*, 176, 106692.
Murfin, J. & Spiegel, M. (2020). Is the Risk of Sea Level Rise Capitalized in Residential Real Estate? *Rev. Financ. Stud.*, 33, 1217–1255.
Richardson, M., Liu, P. & Eggleton, M. (2022). Valuation of Wetland Restoration: Evidence from the Housing Market in Arkansas. *Environ. Resour. Econ.*, 81, 649–683.
Rivera, N.M. & Loveridge, S. (2022). Coal-to-gas fuel switching and its effects on housing prices. *Energy Econ.*, 106, 105733.
U.S. Geological Survey. (2019). Land Change Monitoring, Assessment, and Projection [WWW Document]. URL https://www.usgs.gov/land-resources/eros/lcmap
Wentland, S.A., Ancona, Z.H., Bagstad, K.J., Boyd, J., Hass, J.L., Gindelsky, M. & Moulton, J.G. (2020). Accounting for land in the United States: Integrating physical land cover, land use, and monetary valuation. *Ecosyst. Serv.*, 46, 101178.
Zhang, J., Phaneuf, D.J. & Schaeffer, B.A. (2022). Property values and cyanobacterial algal blooms: Evidence from satellite monitoring of Inland Lakes. *Ecol. Econ.*, 199, 107481.
Zhang, L. & Leonard, T. (2021). External validity of hedonic price estimates: Heterogeneity in the price discount associated with having Black and Hispanic neighbors. *J. Reg. Sci.*, 61, 62–85.
Zheng, A. (2022). The Valuation of Local School Quality under School Choice. *Am. Econ. J. Econ. Policy*, 14, 509–537.

Supplementary Figures

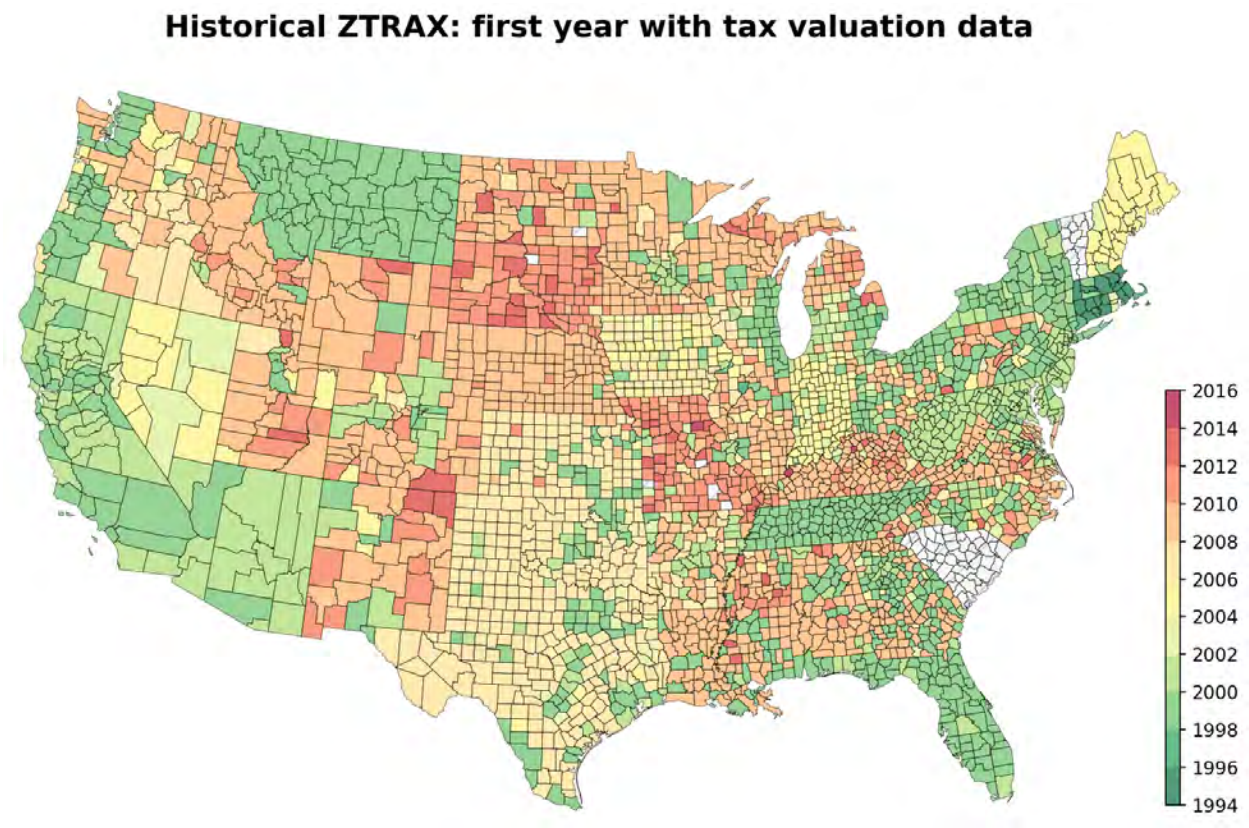

Figure S1: Earliest assessment year available in ZTRAX' historical assessment data. Grey color indicates no data on historical assessment year.

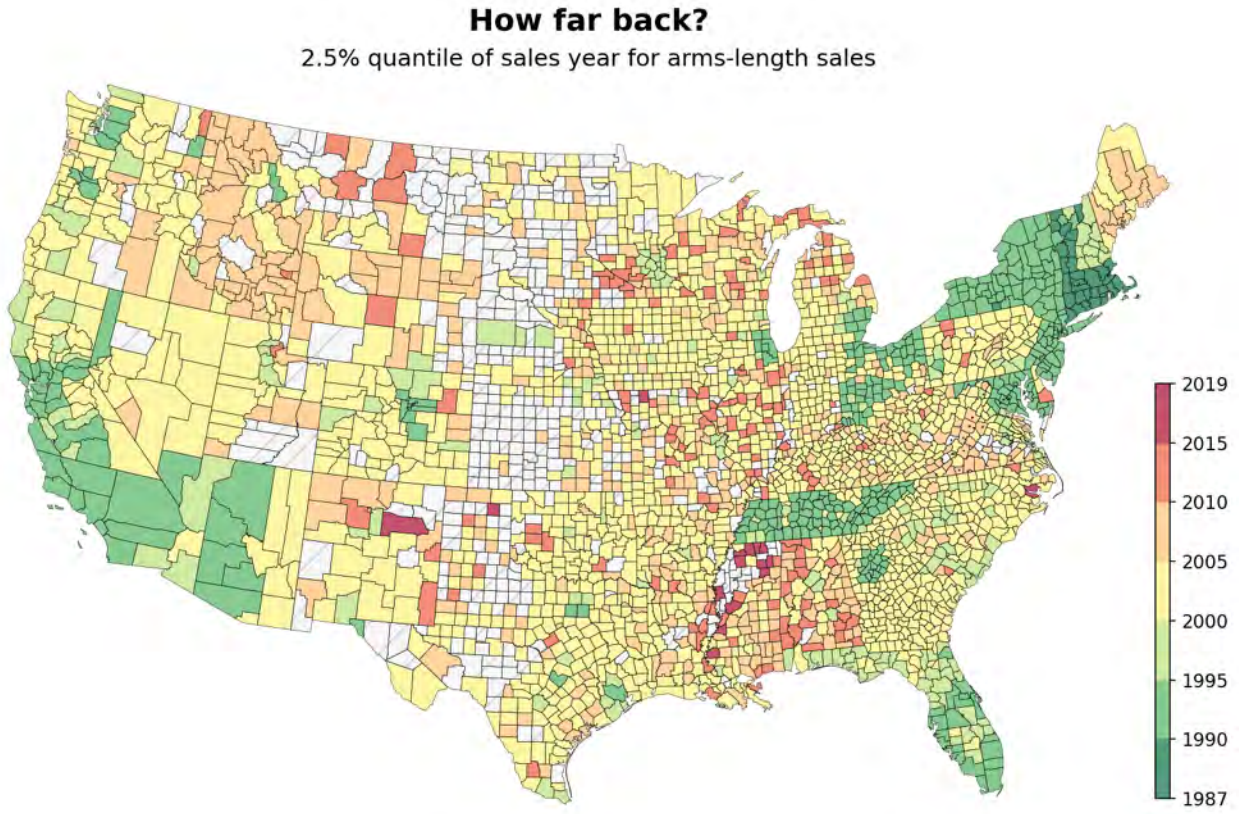

Figure S2: Temporal depth of ZTRAX transaction records. Map shows 2.5% percentile of sale dates for arms-length sales in each county. Counties often contain a long and thin tail of early transactions, many of which might be based on erroneous date records. The 2.5% percentile therefore provides a more robust visual indicator for the beginning of the time period an analyst might reasonably choose as a cutoff for their analysis.

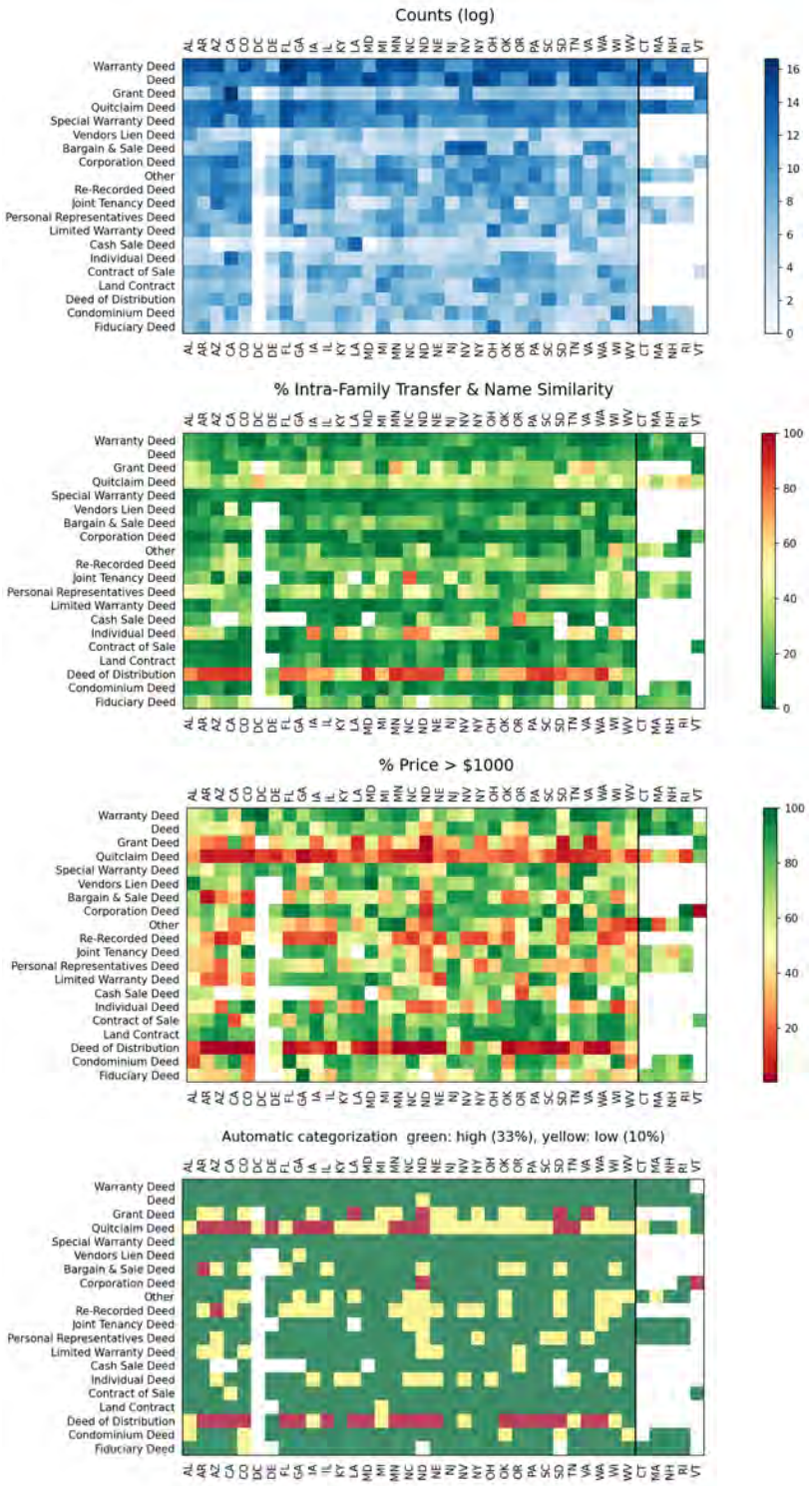

Figure S3: Illustration of our statistical approach to define confidence levels for combinations of document types and states, by example of the 20 most frequent deed types retained after the hierarchical exclusion of known document types (intra-family transfers, foreclosures). From top to bottom: frequency of document type; percentage of transactions identified as intra-family transfers (using both the intra-family transfer flag and similarity between buyer and seller names); percentage of transactions with prices larger than $1000; final categorization. States shown have price data in >20% of transaction records. New England states are shown separately to emphasize similarities in document type usage.

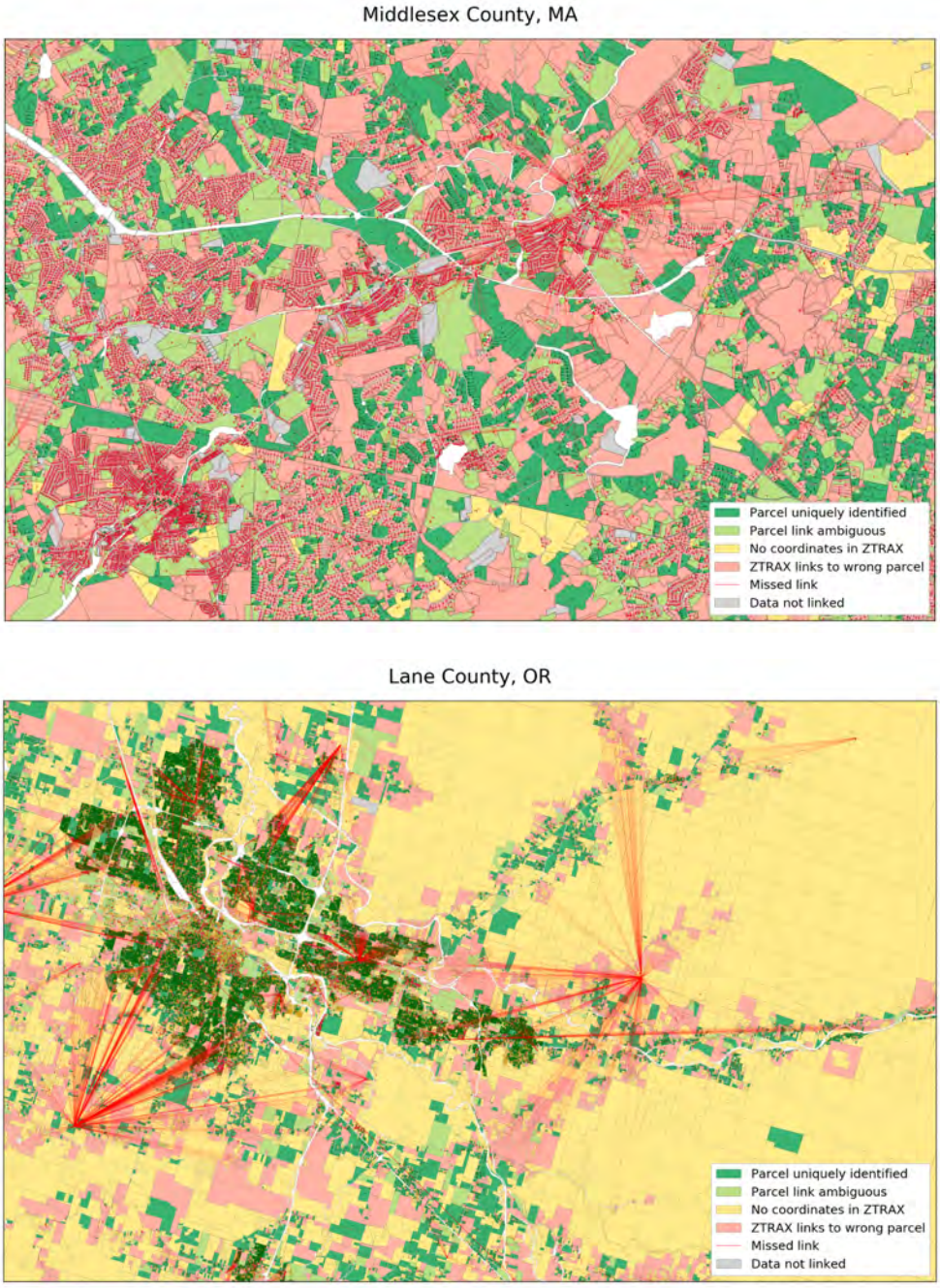

Figure S4: Illustration of potential geo-location errors of ZTRAX data by example of the town of Concord in Middlesex County, Massachusetts (top) and the city of Eugene in Lane County, Oregon (bottom). Errors shown would occur if the analyst made the (naïve) assumptions that ZTRAX coordinates are correct, based on the WGS84 datum and that they can be used to identify parcels polygons in digital maps. In Middlesex County, which uses the NAD27 datum, this approach would link up the vast majority of urban parcels incorrectly (dark red points) due to a translation error of approximately 40 meters. Most large, rural parcels would have been either incorrectly (red) or ambiguously linked (light green: multiple coordinates on one parcel, only one of which is correct). In Lane County, which predominantly uses the WGS84 datum, the analyst would correctly and uniquely identify a majority of urban parcels (dark green points). However, most non-urban parcels have no parcel coordinates (yellow). Furthermore, a non-trivial number of urban and rural parcels appear to be incorrect and derived from ZIP code area centroids (red lines show distance between correct centroid and ZTRAX coordinates and converge around a small number of points).

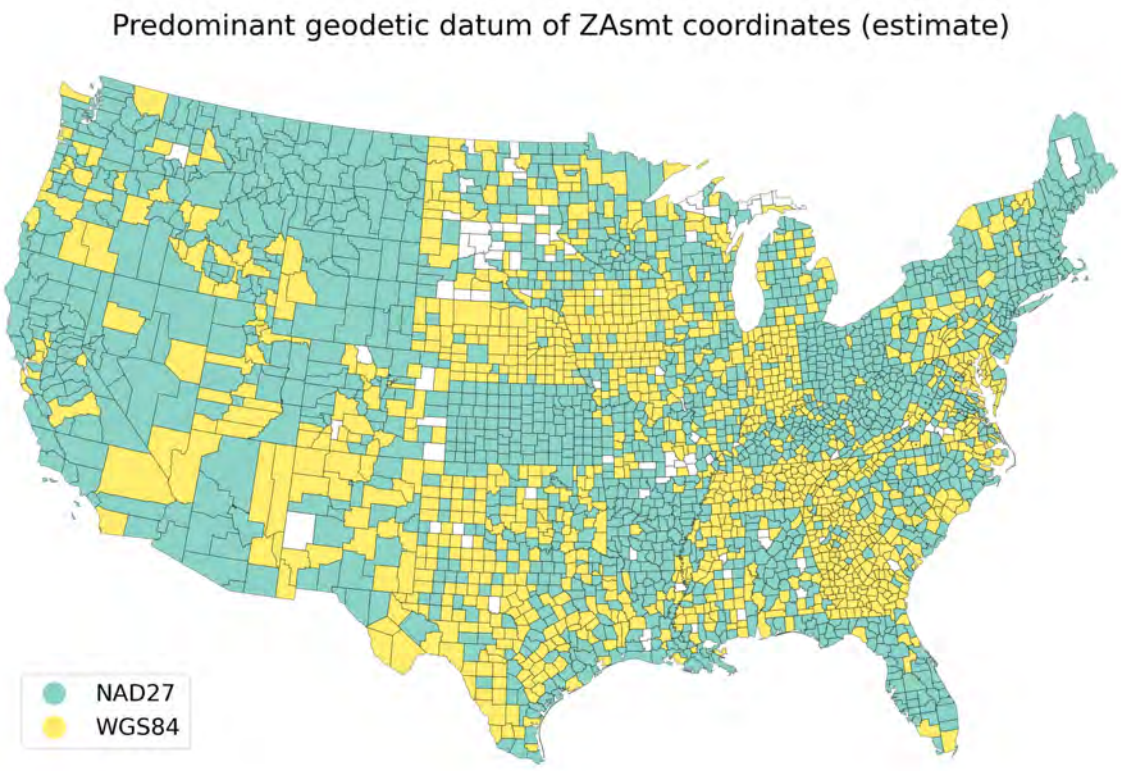

Figure S5: Estimated predominant geographic datum of ZTRAX coordinates. These estimates were obtained by projecting coordinates available in the assessment data into a geographic projection using both the NAD27 (EPSG:4267) and WGS84 (EPSG:4326) datum and selecting the datum that produces the smallest median distance between assessment data coordinates and parcel boundary centroids. Distance computation in the NAD83 Conus Albers projection (EPSG:5070).

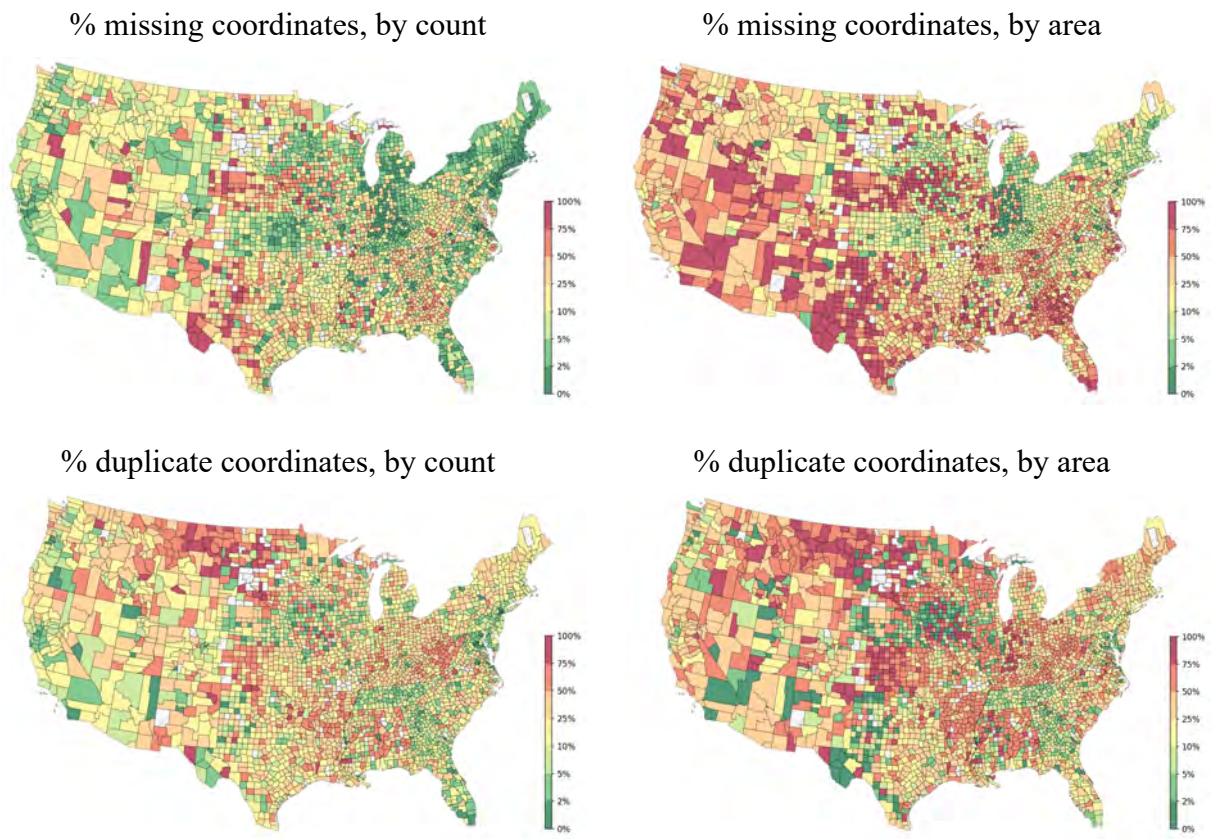

Figure S6: Prevalence of missing (top) and duplicate (bottom) coordinates in ZTRAX. Percentages are computed in reference to parcel count (left) and parcel area (right) for 131 million parcel polygons linked to assessment data.

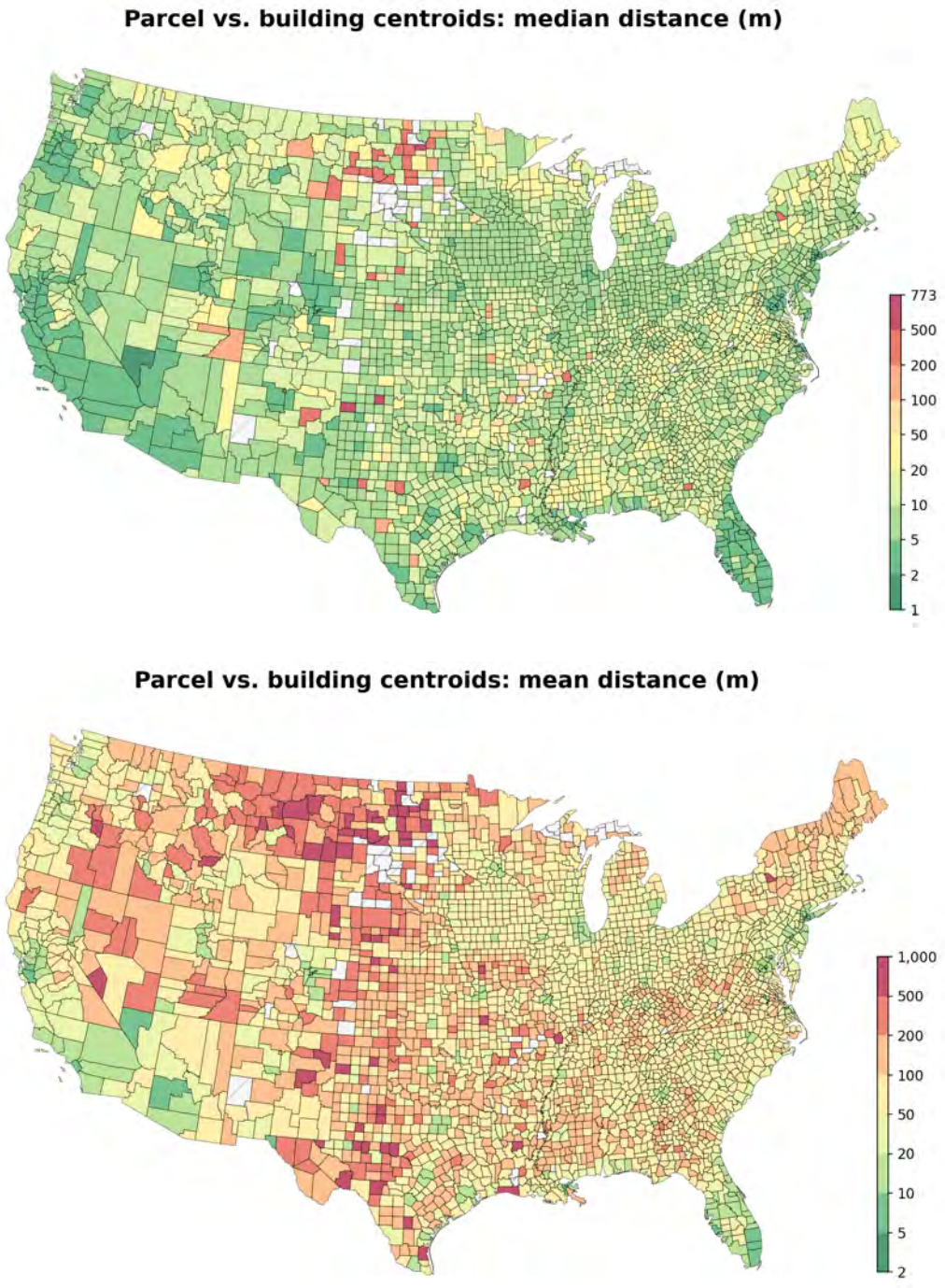

Figure S7: Median (top) and mean (bottom) distances between parcel centroids and building footprints for parcels with exactly one building footprint.

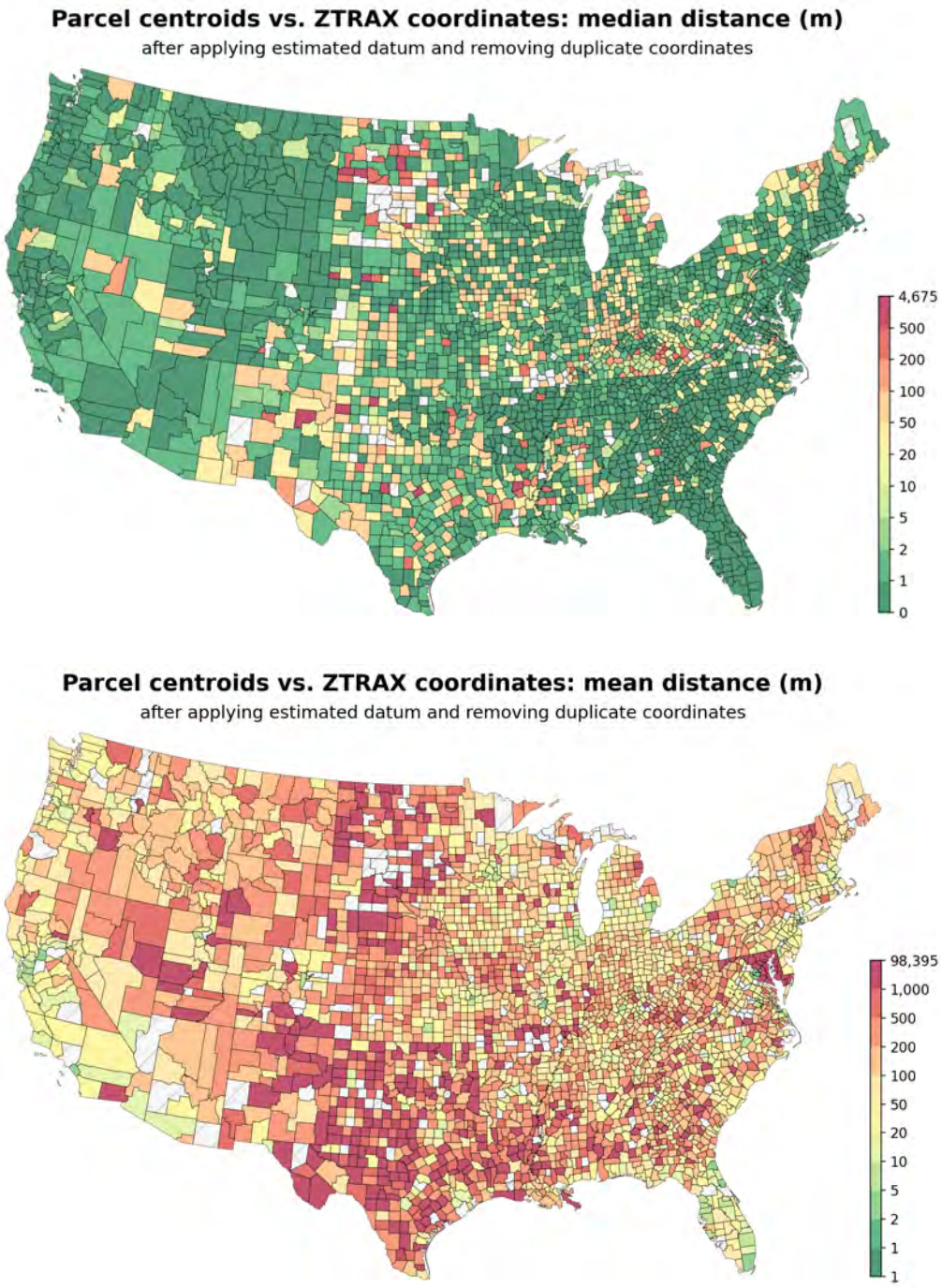

Figure S8: Median (top) and mean (bottom) distance between parcel centroids and ZTRAX assessment data coordinates for parcels uniquely linked to assessment data records. Distances are in meters and computed in EPSG:5070 projection.

Figure S9: Illustrative example of observed changes in geo-coordinates between ZTRAX versions downloaded in 2017 and 2019. Panel (a) shows a new subdivision in Rhode Island. In 2017, all properties in the subdivision had identical geo-coordinates, likely derived from the ZIP code centroid. By 2019, each property had its own unique geo-coordinate, likely derived from street addresses. Panel (b) shows minor shifts in coordinates between 2017 and 2019 database version, likely resulting from a coordinate improvement based on building footprints. Imagery source: ESRI

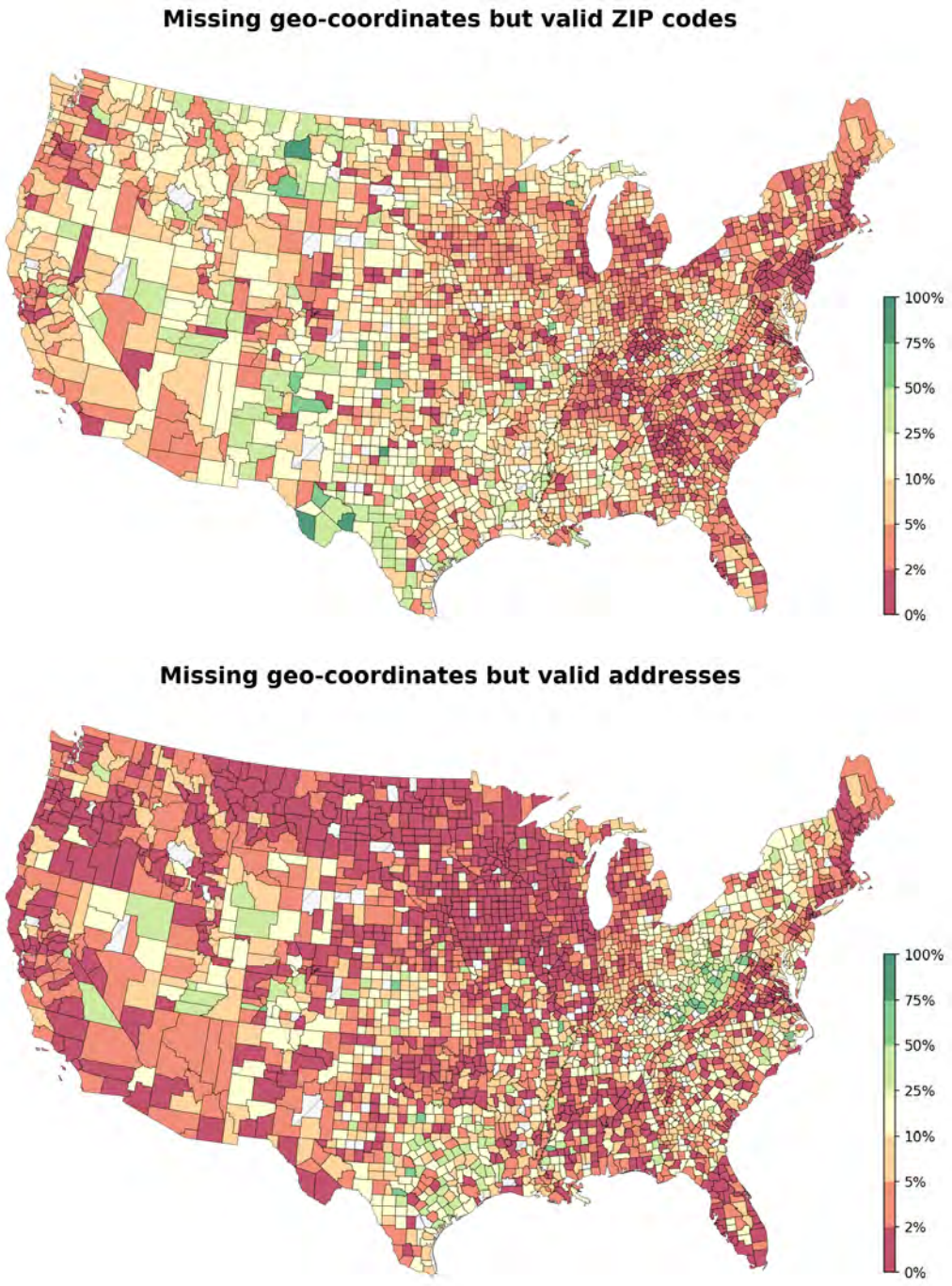

Figure S10: Percentage of records for potential improvement of geo-locations using the ZIP code and address information, shown here for a ZTRAX version from 2017.

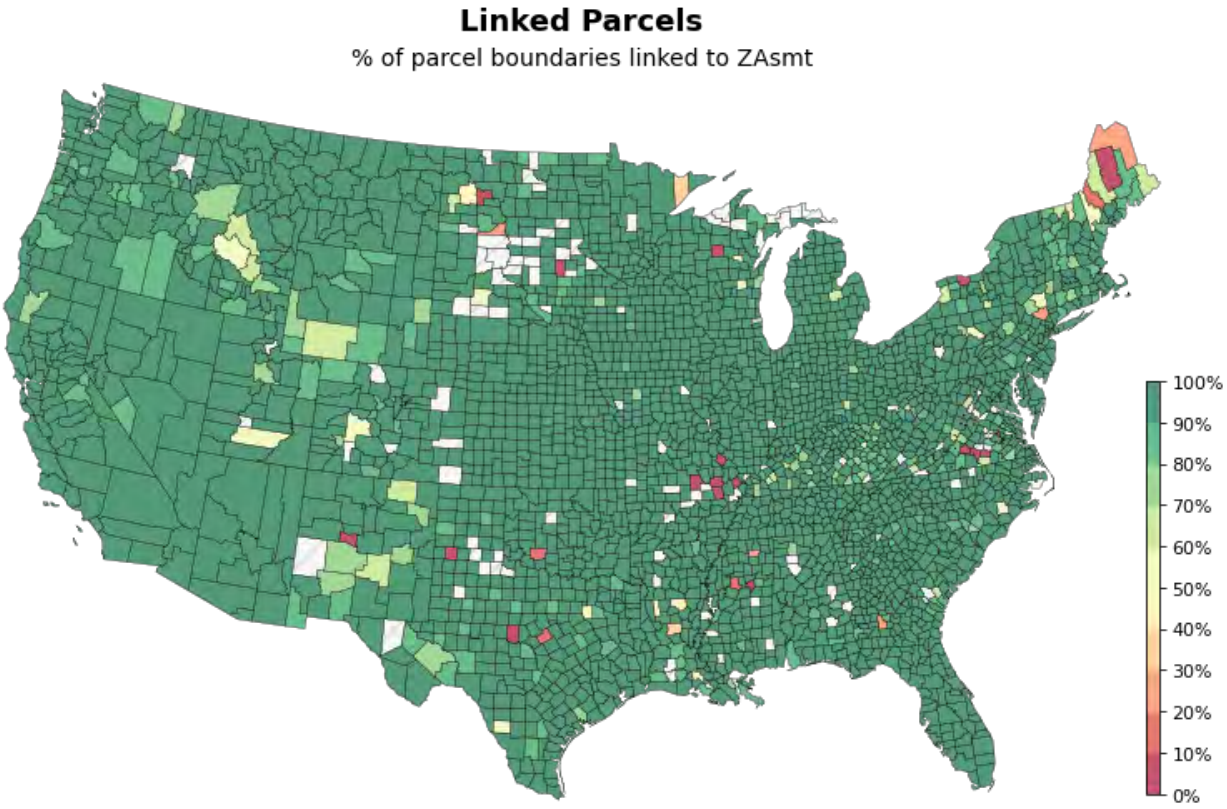

Figure S11: Percentage of parcels from digital parcel maps that PLACES lab members could successfully linked to assessment data records using string-pattern matching on assessor parcel numbers and tax account identifiers. Map shows status of PLACES data as of Jun 20, 2021.

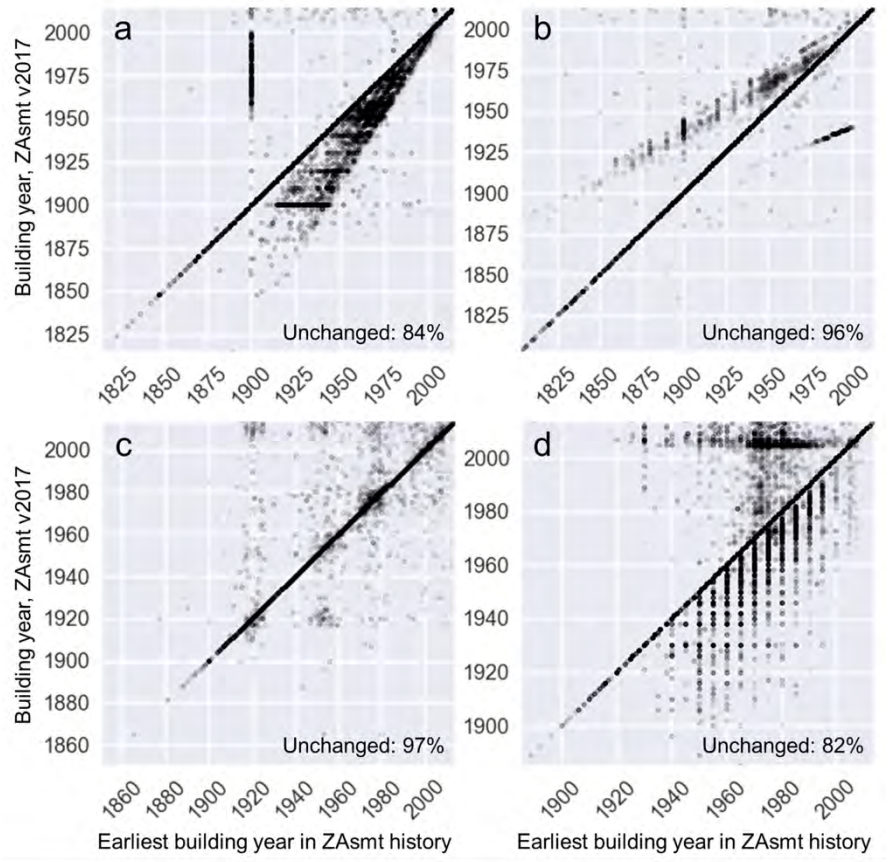

Figure S12: Differences in building year data for identical properties (matched by "ImportParcelID") between current and historical assessment data in ZTRAX (both downloaded in 2017) for (a) Mercer County (Ohio), (b) Preble County (Ohio), (c) Natrona County (Wyoming) and (d) De Soto County (Florida). Points below the diagonal show records whose building years were backdated, i.e., building years in the historical assessment data were replaced by a more recent year in the assessment data. Note that these changes across database versions are relatively rare (82% - 97% in the examples shown).

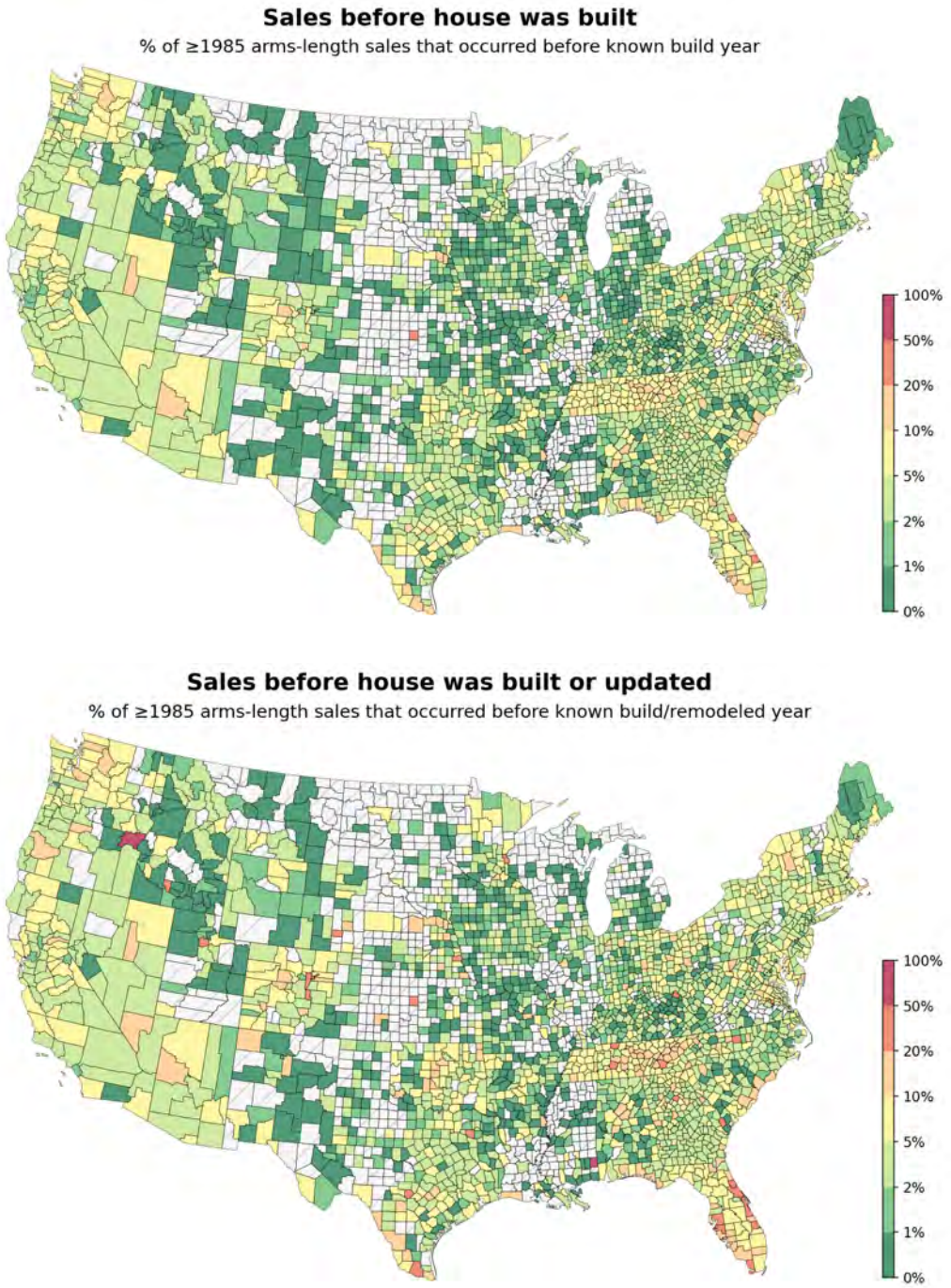

Figure S13: Percentage of sales observations after 1985 (parcel-linked, arms-length ZTRAX transaction records) that occurred before house was known to be built (top) or known to be either built or renovated (bottom).

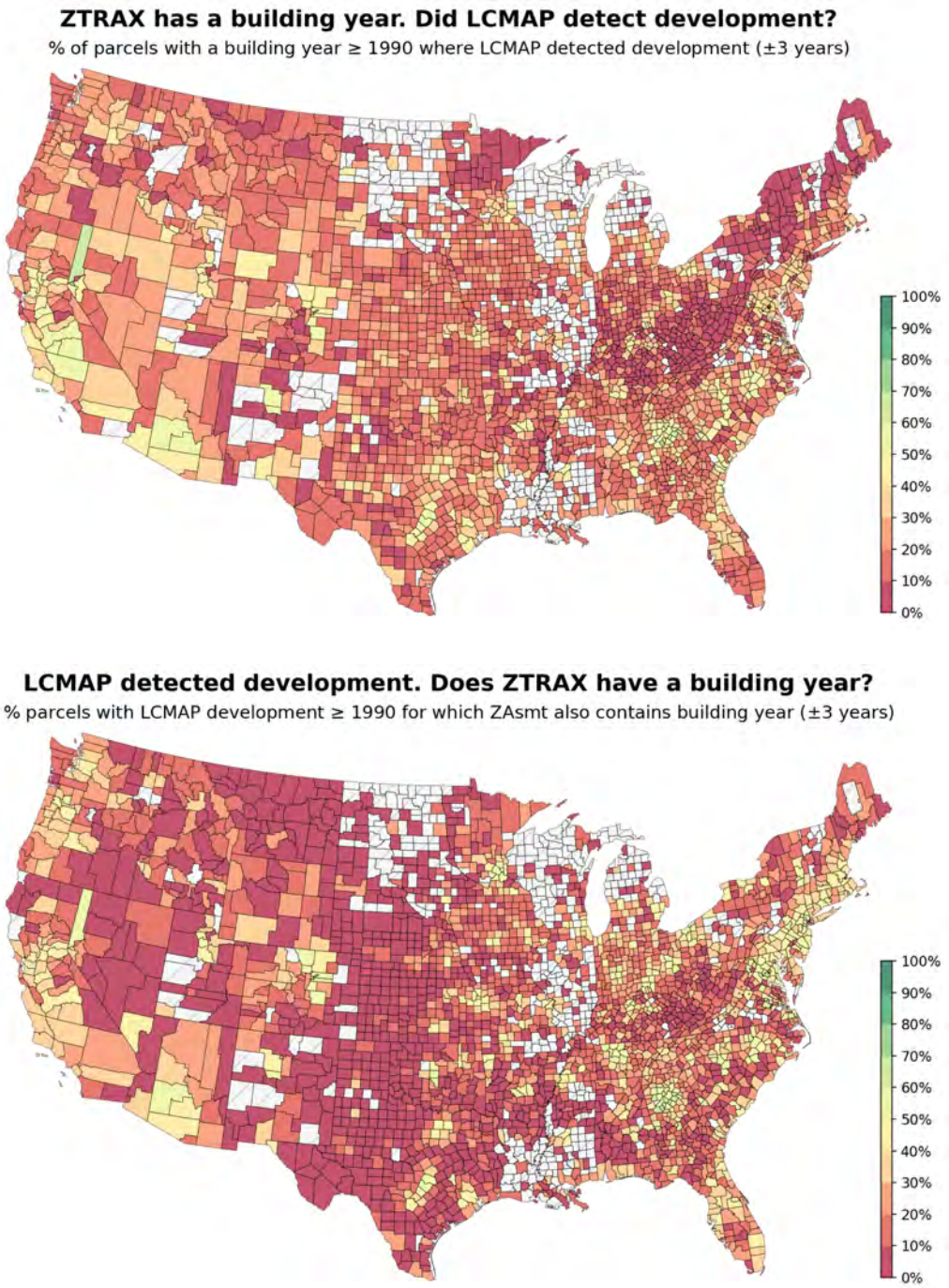

Figure S14: Discrepancies between building years reported in ZTRAX assessment data vs. most recent change of a pixel inside the corresponding digital parcel boundaries from a "non-developed" to a "developed" class in the LCMAP dataset (U.S. Geological Survey 2019).

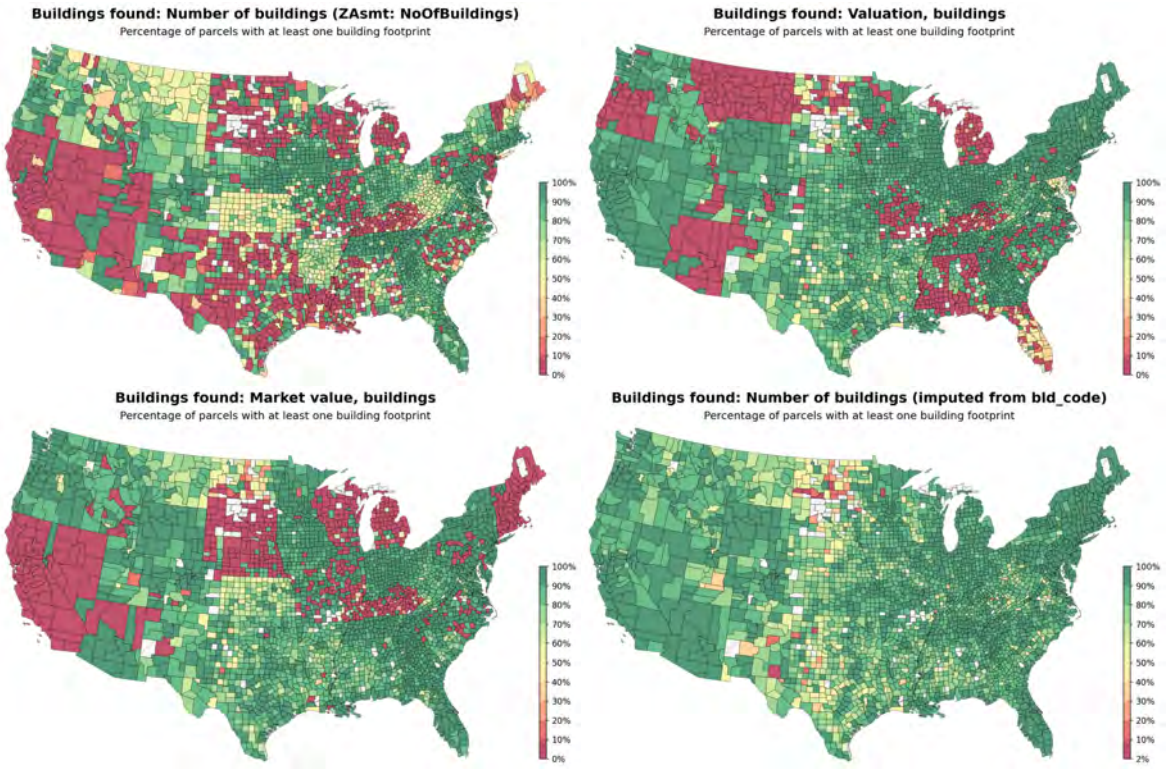

Figure S15: Percentage of parcels with at least one building footprint that would have been identified as "developed" based on the following four alternative indicators: Number of buildings (top left), positive building valuation (top right), positive building estimated market value (bottom left) and our own indicator of building presence derived heuristically from the standardized land use codes based on their text description (bottom right).

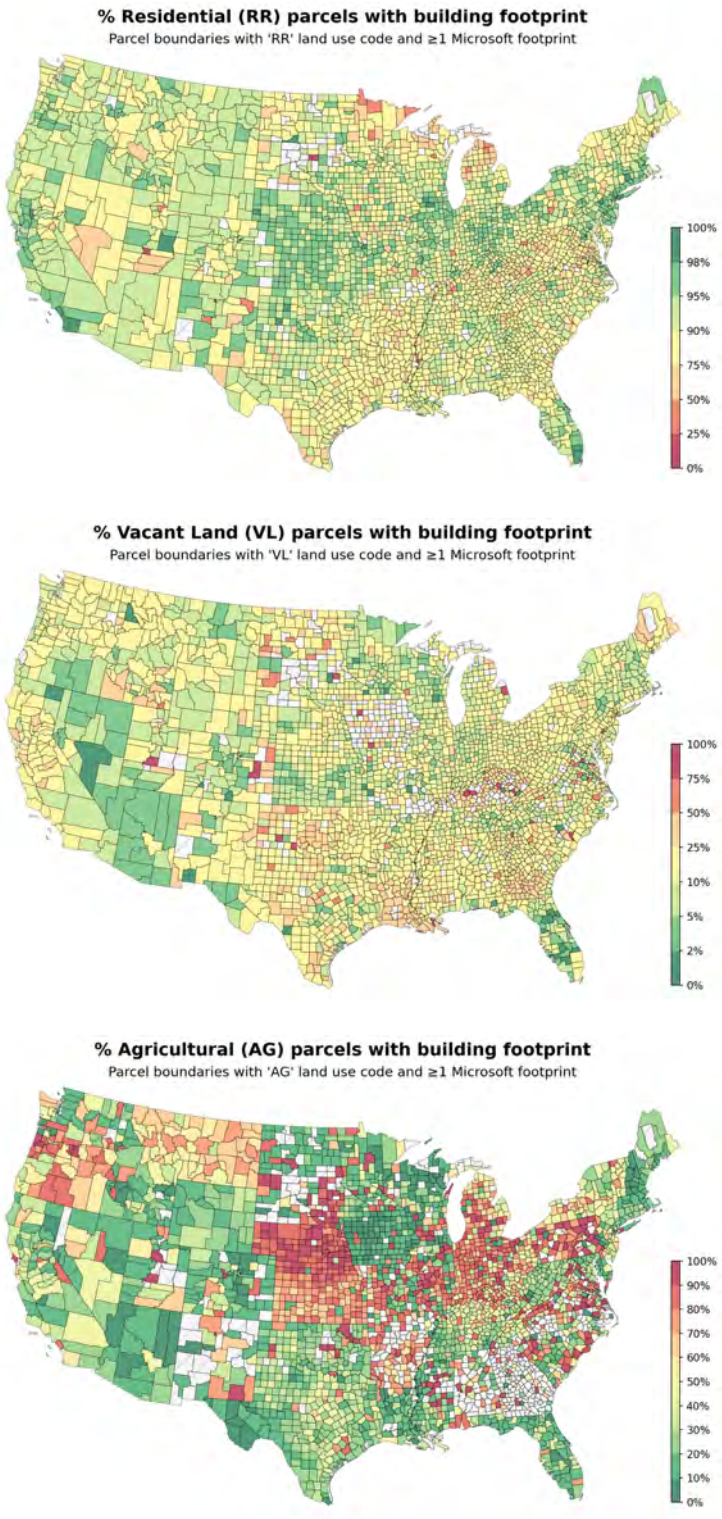

Figure S16: Percentage of building footprints found on "residential" (top), "vacant" (middle), and "agricultural" (bottom) parcels. The unique scale and direction of each map's color mapping serves to highlight concerns (absence of buildings on "residential" parcels, presence of buildings on "vacant" parcels) and heterogeneity in agricultural parcel definitions between neighboring states (e.g., Nebraska/Kansas vs. Iowa/Colorado)

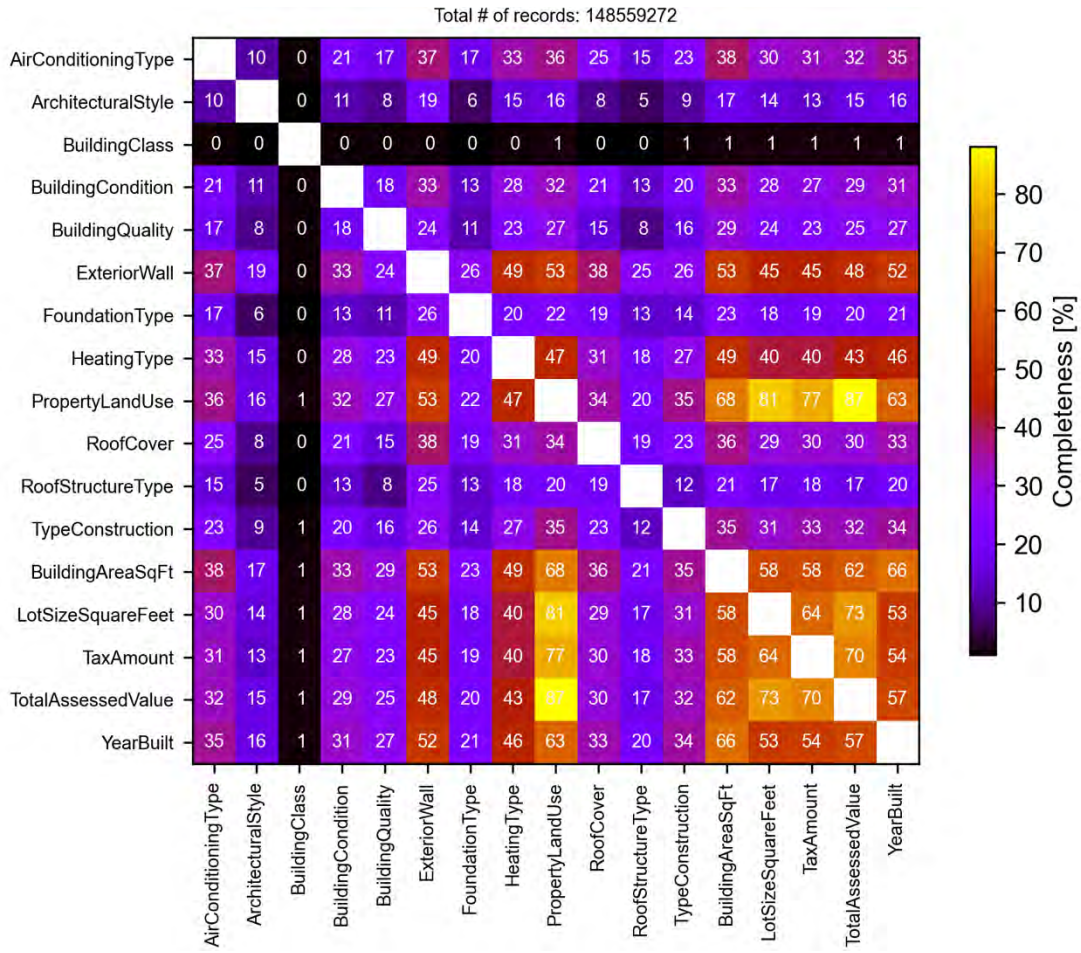

Figure S17: Mutual completeness of ZTRAX assessment records (measured in % of the total number of ZAsmt records) for selected building and property attributes, illustrating the variability of thematic data completeness across different attributes

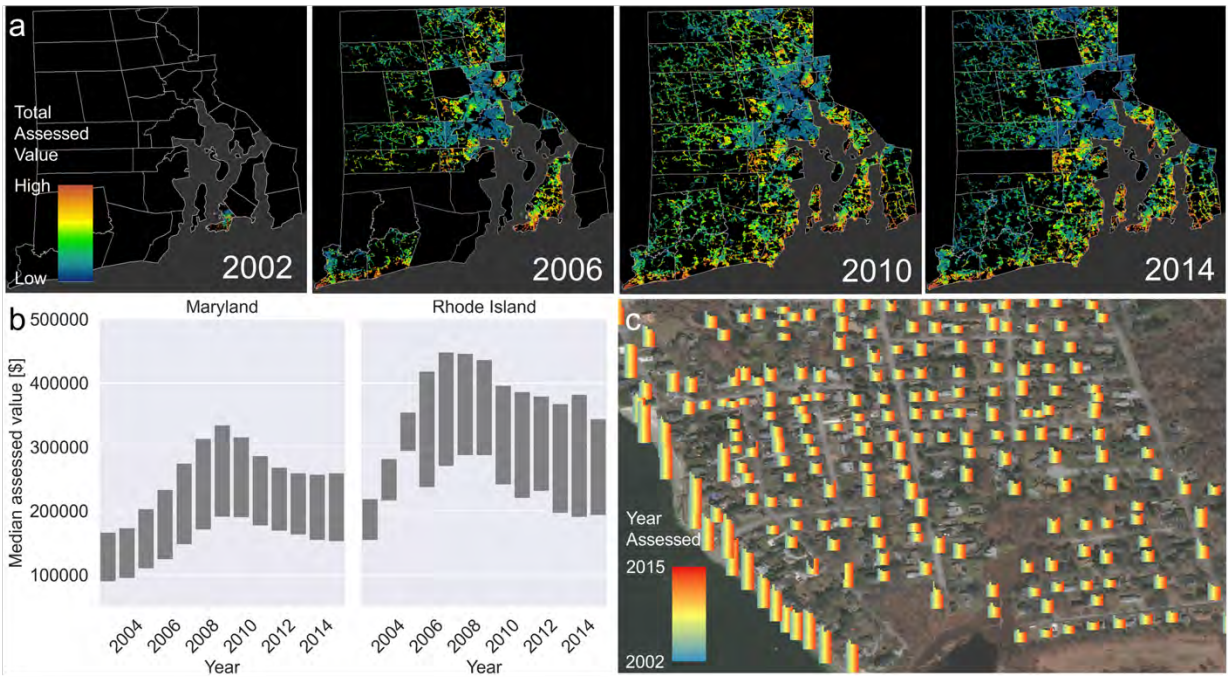

Figure S18: Illustration of data gaps in historical ZTRAX assessment data by example of the variable "total assessed value". Panel (a): Total assessed value of properties in Rhode Island, shown for different years based on the assessment year attribute, illustrating different updating cycles per town (white). Panels (b) and (c) illustrate exemplary strategies to mitigate effects of irregular updating cycles, such as (b) spatial aggregation (shown here: IQR of county-level median assessed values per state) and (c) interpolation of record-level time series for a subset of Rhode Island. The vertical dimension of the bar charts shows the total assessed value, and the horizontal component of the bar charts shows its variation over time for individual ZTRAX records at the respective locations, after applying forward-filling to the missing values in each record-level time series. Imagery source: ESRI

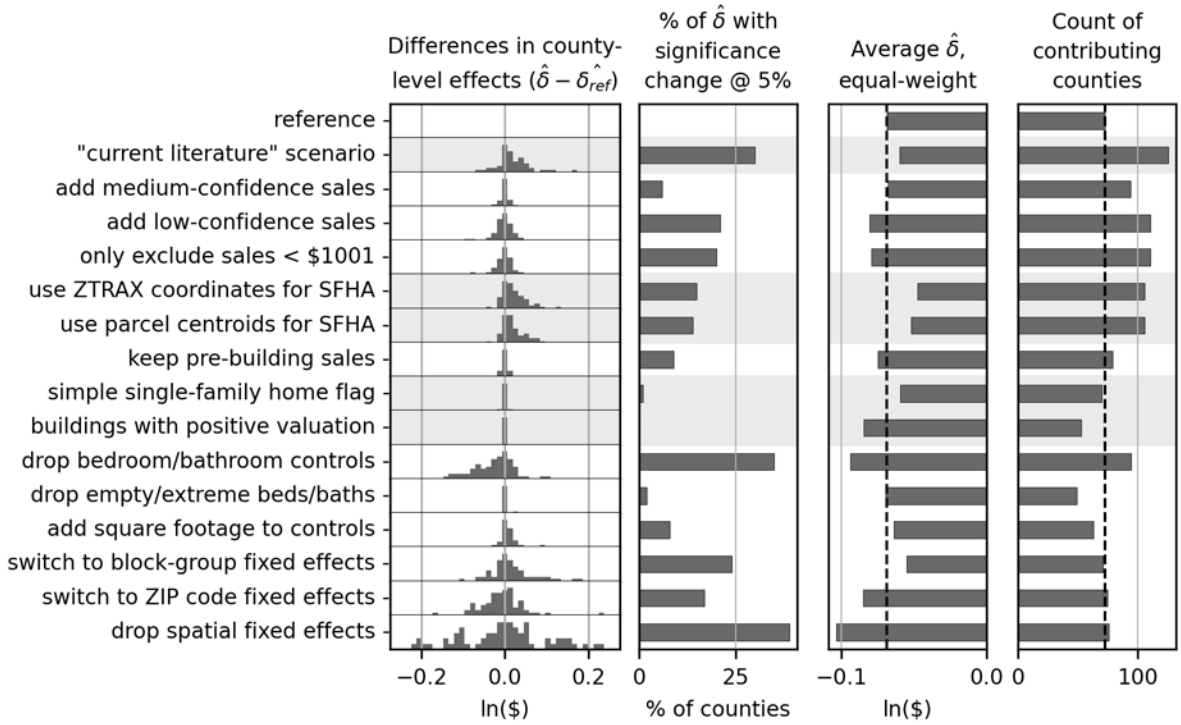

Figure S19. Effects of variations in data filtering, processing, and modeling choices on estimated flood zone discounts when including only counties with particularly favorable empirical contexts (defined here as having at least 500 identifying sales observations both inside and outside the flood zone).

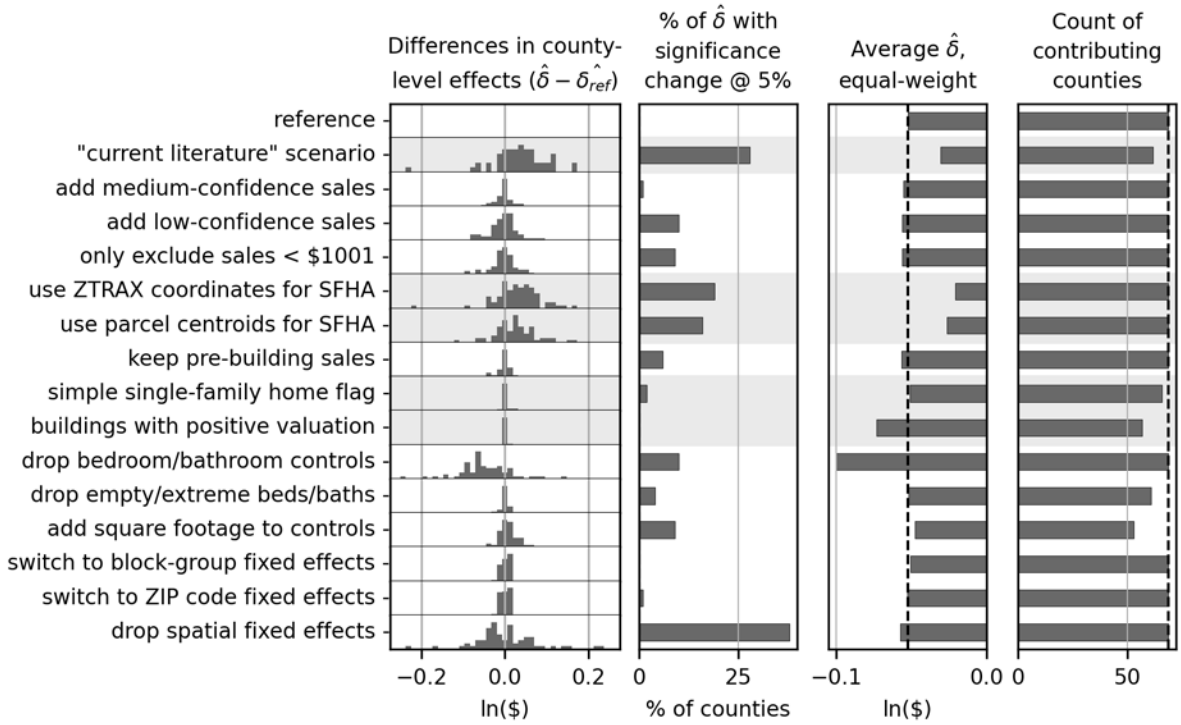

Figure S20: Effects of variations in data filtering, processing, and modeling choices on estimated flood zone discounts when including only counties whose estimates are relatively stable across variations in fixed effects (defined here as $\left|\hat{\delta} - \widehat{\delta_{ref}}\right| < 0.02$ when switching to block group or ZIP code fixed effects).

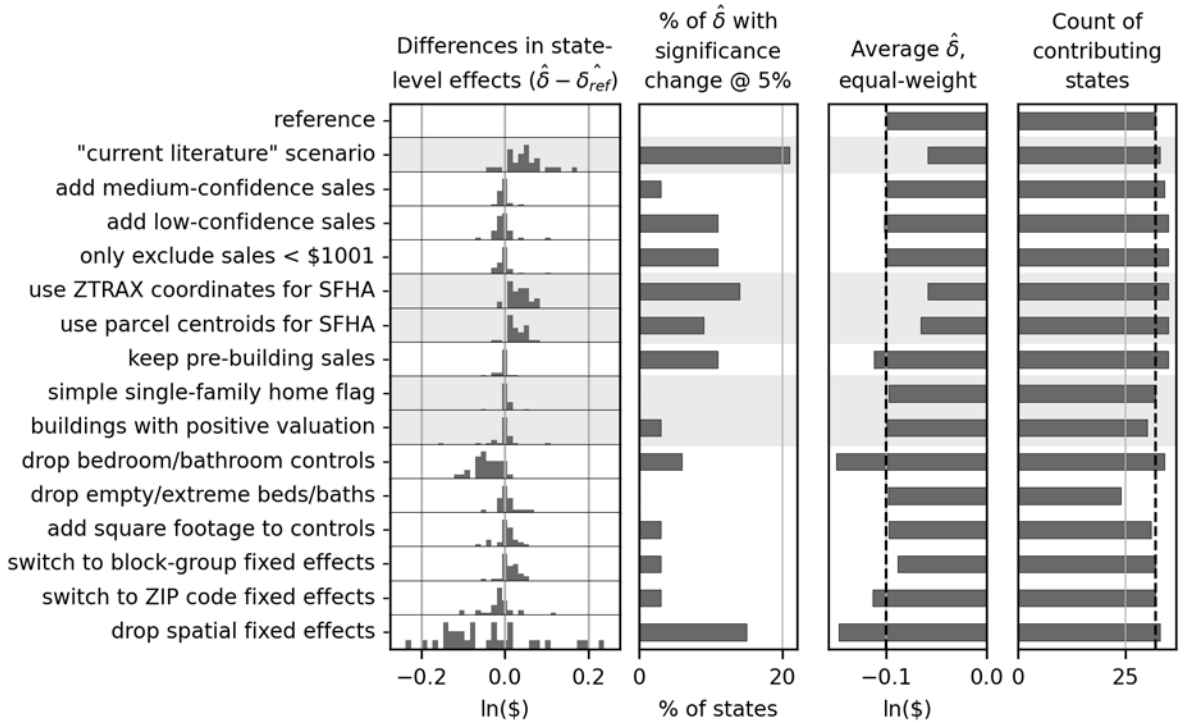

Figure S21: Effects of variations in data filtering, processing, and modeling choices on estimated flood zone discounts when pooling data at the state-level and including only states with particularly favorable empirical contexts (defined here as having at least 500 identifying sales observations both inside and outside the flood zone).

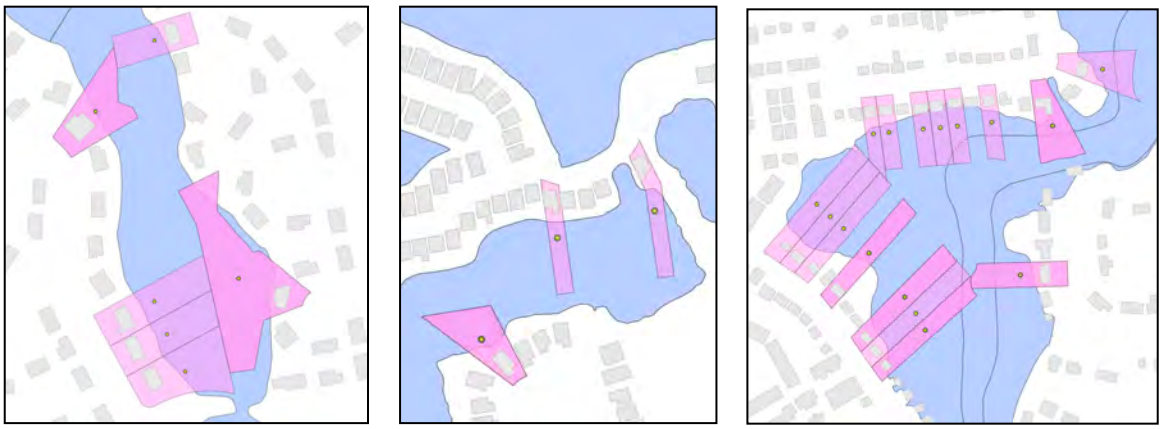

Figure S22: Illustrative examples of misallocations of property sales to special flood hazard areas (blue) when using parcel centroids (green points) instead of building footprints (grey polygons) in Montgomery county, Pennsylvania (left), Polk county, Florida (middle), and Fulton county, Georgia (right).

Supplementary Text: Deed Definitions

Contributed by: Jesse Richardson, Professor of Law, West Virginia University

DISCLAIMER: Each state has different customs and usages with respect to the types of deeds used. This document presents general legal principles. Users should consult the law of the particular state to confirm the assumptions and definitions used here.

Administrators Deed

- Also termed a *court-officer's deed*, an *administrator's deed* is an affidavit document that conveys property owned by a person who has died intestate. An affidavit of deed confirms delivery and acceptance of a deed by the grantee, and thereby its validity.[1]

This type of deed normally conveys property to an heir at the death of the owner, so does not represent an arms-length transaction.

Affidavit

- An *affidavit* is a voluntary declaration of facts written down and sworn to by a declarant, usually before an officer authorized to administer oaths.[2]

An affidavit can be necessary in a variety of situations, most not involving an arms-length transaction.

Affidavit—Death of Joint Tenant

- A *death of a joint tenant affidavit* is frequently used in lieu of a judicial proceeding to establish fact of death for the purposes of clearing title in the case of a joint tenant's death. The affidavit describes the property affected and states the recording information as to the prior document establishing affiant's and decedent's interest in the property.[3]

This affidavit is generally used when one joint tenant dies and the remaining joint tenants receive the property through survivorship. This transaction is not a market transaction.

Agreement of Sale (Agreement for Deed)

- An *agreement of sale* is an agreement that obligates someone to sell and that may include a corresponding obligation for someone else to buy.[4]

---

[1] Administrator's deed, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

[2] Affidavit, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

[3] Establishing and Reporting Fact of Death: 2 LPI Legal Professional's Handbook ¶ 1720 (West 2020).

[4] Agreement of sale, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

- An *agreement for deed* is an installment contract for the sale of land in which the seller (grantor) retains legal title until paid in full. During the contract term, the buyer (grantee) possesses only equitable title.[5] In Florida, an *agreement for deed* is considered a mortgage.[6]

Agreements of Sale can arise in a variety of situations but generally denote an arms-length transaction.

Assignment Deed

- An *assignment deed* is the instrument used to record the transfer of ownership in a property from one party to another and states that the assigned property will belong to the assignee and no longer to the assignor beginning on a specified date.[7]

Assignment deeds may arise in a variety of situations but usually involve assumption of a debt. These deeds likely do not arise from arms-length transactions.

Bargain & Sale Deed

- Typically used in foreclosure and tax sales, a *bargain & sale deed* lacks an express covenant about the validity of the title but implies that the grantor holds title to the property and conveys it to a buyer for valuable consideration.[8] Also termed a *bargain-and-sale deed without covenants* or a *deed without warranty*.

Given the lack of covenant with respect to validity of title, these types of deeds are often not used in arms-length transactions. However, in some states bargain and sale deeds are typical and represent arms-length transactions.

Beneficiary Deed

- A *beneficiary deed* automatically conveys the property to a designated person upon the property owner's death. Because it permits the property owner while living to make a grant of real property to the owner's heirs without effecting a change of ownership, a *beneficiary deed* avoids probate.[9]

---

[5] Orange County Property Appraiser, *Deed Types* (last visited Dec. 29, 2020), https://www.ocpafl.org/Searches/Lookups.aspx/Code/Deed.

[6] Fla. Stat. Ann. § 697.01 (West 1838); *see also* Lulich Attorneys & Consultants, *Florida Agreement for Deed Law: The Same as a Mortgage?* (posted Mar. 25, 2018), http://www.lulich.com/florida-agreement-for-deed-law-the-same-as-a-mortgage/.

[7] Deed of Assignment: Everything You Need to Know, UPCOUNSEL (last visited Dec. 29, 2020), https://www.upcounsel.com/deed-of-assignment.

[8] Bargain & sale deed, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

[9] Beneficiary deed, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

Beneficiary deeds convey property upon death and do not represent arms-length transactions.

Cash Sale Deed

- A *cash sale deed* is the sale of property between two parties in exchange for cash and therefore does not require the involvement of a mortgage lender or bank for a line of credit.[10]

A cash sale deed generally results from arms-length transactions.

Certificate of Transfer

- A *certificate of transfer* is the contractually agreed transference of a patent application from one court to another.[11]

This type of deed is not normally used in a land transaction and should be viewed skeptically with respect to representing an arms-length transaction.

Commissioners Deed

- A *commissioner's deed* is a deed used by a commissioner to convey property in his or her official capacity.[12]

Since this deed arises out of litigation or a forced sale, the transactions would not represent arms-length transactions.

Condominium Deed

- A condominium is created by a *master deed* that governs a condominium unit owner's rights and defines the common elements of the condominium. Created from the *master deed* are individual *condominium unit deeds*, and the acceptance of a unit deed by each unit owner constitutes an acceptance of the terms of the declaration, bylaws, and rules and regulations of the condominium. Normal rules and principles of the construction of deeds apply to *condominium unit deeds*.[13]

Condominium deeds should generally represent arms-length transactions.

---

[10] Rochford Law, *Buying Property in a Cash for Deed Transaction* (June 25, 2019), https://info.rochfordlawyers.com/resources/buying-property-with-cash-deeds; *see also* Nicole Manuel, *What is a Cash Deed?* (June 19, 2017), https://legalbeagle.com/5701773-execute-deed-power-attorney.html.

[11] Certificate of transfer, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

[12] Commissioners deed, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

[13] 31 C.J.S. Estates § 232 (West 2020).

Contract of Sale

- A *contract of sale* is a sale for the present transfer of property for a price or a contract to sell goods at a future time. Also termed *contract to sell* or a *sale contract*.[14]

Not a deed, but an agreement that results in a deed. Contracts of sale usually result from arms-length transactions.

Corporation Deed

- A *corporation deed* is a document that transfers ownership of a property from one organization to another individual or organization. The organization that makes the transfer does not need to be registered as a business corporation, and a city government that sells or grants a property may also issue a corporation grant deed.[15]

Unclear as to whether these deeds would represent arms-length transactions. Likely that some would be arms-length and some not.

Correction Deed

- A *correction deed* is a legal document used to amend errors in an existing property title. While typically used to correct minor errors like misspelled names or basic missing information like marital status, a *correction deed* may also be used to correct defects in the execution or acknowledgment of a property deed. Unlike a *re-recorded deed*, a *correction deed* is essentially a new document that is a duplicate of the original deed save for the corrected information.[16] Because its only purpose is to explain and correct minor errors, a *correction deed* does not alter the substance of the original deed but merely confirms the conveyance effected by the previously recorded deed.[17]

These deeds should be disregarded as not resulting from arms-length transactions.

Court Order/Action

- A *court order/action* is a proceeding to establish a plaintiff's title to land by compelling the adverse claimant to establish a claim or be forever estopped from asserting.[18] The

---

[14] Contract for sale, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

[15] Eric Novinson, *What is a Corporation Grant Deed?*, SAPLING (last updated Mar. 28, 2017), https://www.sapling.com/8042679/corporation-grant-deed.

[16] Goosman Rose Colvard & Cramer, P.A., *Confirmatory/Corrective Deed* (last visited Dec. 29, 2020), https://www.grcclaw.com/confirmatory-corrective-deed/.

[17] West Virginia Correction Deed (last visited Dec. 29, 2020), https://www.deeds.com/forms/west-virginia/correction-deed/. While the link is to Correction Deeds for West Virginia in particular, the definition of a *correction deed* is fairly ubiquitous across the site.

[18] Action to quiet title, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

proceeding is intended to remove, or "quiet," a claim or objection to a title by clarifying ownership of a property.[19]

These actions are compelled and do not represent arms-length transactions.

Deed

- A *deed* is the physical legal document by which the ownership rights to a property are conveyed. At common law, a *deed* is any written instrument that is signed, sealed, and delivered and that conveys some interest in property.[20]

Deed is a generic term that in and of itself does not give enough information to know whether the transaction is arms-length.

Deed in Lieu of Foreclosure

- Often referred to as a "deed in lieu," a *deed in lieu of foreclosure* is a deed by which a borrower conveys fee-simple title to a lender in satisfaction of a mortgage debt and as a substitute for foreclosure.[21]

These deeds do not represent arms-length transactions.

Deed of Distribution

- A *deed of distribution* is a fiduciary's deed conveying a decedent's real estate.[22]

These deeds transfer property at the death of an owner and do not represent market transactions.

Executors/Executrixs Deed

- An *executor's/executrix's deed* is a document that conveys property owned by a person who has died testate. Also termed a *court-officer's deed*.[23]

These deeds transfer property at the death of an owner and generally do not represent market transactions. However, in some instances the will may direct a sale of the property in which the deed would represent an arms-length transaction

---

[19] Will Kenton, *Quiet Title Action*, INVESTOPEDIA (last updated Nov. 19, 2019), https://www.investopedia.com/terms/q/quiet-title-action.asp.

[20] Deed, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

[21] Deed in lieu of foreclosure, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

[22] Deed of distribution, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

[23] Executors/Executrixs deed, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

Fiduciary Deed

- A *fiduciary deed* is a deed used by a fiduciary to convey property in a fiduciary capacity. Examples include an *administrator's deed*, *executor's deed*, *conservator's deed*, *guardian's deed*, *personal representative's deed*, and *trustee's deed*.[24]

These deeds transfer property in an official fiduciary capacity and may or may not represent arms-length transactions, depending upon the circumstances.

Foreclosure Deed/Certificate

- A *foreclosure deed/certificate* is a legal document that transfers legal ownership of a property to the purchaser at a foreclosure sale. The type of *foreclosure deed/certificate* granted depends on whether a judicial or non-judicial foreclosure is followed.
    - In a judicial foreclosure, the judge issues a foreclosure order directing the sheriff's department to hold a foreclosure sale. Following the foreclosure sale, the sheriff's department issues a *sheriff's deed* to the purchaser of the property.
    - In a non-judicial foreclosure, no judicial proceeding is required; instead, a trustee is appointed by the mortgage lender to hold a publicly noticed foreclosure sale. After the foreclosure sale, the trustee issues a *trustee's deed* naming the purchaser at the sale as the new record owner of the property.[25]

Foreclosure sales are forced sales and often result in sales at less than fair market value. These deeds should not be relied on as arms-length.

Gift Deed

- A *gift deed* is a deed transferring ownership of a property in exchange for either a nominal sum or "in consideration of love and affection."[26]

By their very nature, gift deeds do not represent arms-length transactions.

Grant Deed

- A *grant deed* contains, or has been implied by law, some but not all of the usual covenants of title, especially a deed in which the grantor warrants that he or she (1) has not previously conveyed the estate being granted, (2) has not encumbered the property except as noted in the deed, and (3) will convey to the grantee any title to the property acquired after the date of the deed.[27]

---

[24] Fiduciary deed, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

[25] Jerusha Hardman, *What is a Foreclosure Deed?*, SAPLING (last visited Dec. 29, 2020), https://www.sapling.com/8397530/foreclosure-deed.

[26] Gift deed, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

[27] Grant deed, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

Lacking some of the usual warranties, a grant deed will generally not represent arms-length transactions. However, some states may by custom use these deeds in market transactions.

Interspousal Transfer Deed

- An *interspousal transfer deed* is used in some states to transfer ownership between spouses. A gift given by one spouse to the other during the marriage is considered "separate," or separately owned, rather than "marital," or mutually owned, property.[28]

As a deed between spouses, an interspousal transfer deed does not represent an arms-length transaction.

Intrafamily Transfer & Dissolution

- An *intrafamily transfer and dissolution* is a deed that transfers ownership from one family member to another without the property being sold. Dissolution typically refers to divorce, in which ownership is transferred from the couple to just one person, but it may also refer to a change in the trust.[29]

Intrafamily Transfer and Dissolution deeds do not represent arms-length transactions.

Joint Tenancy Deed

- A *joint tenancy deed* is a deed whose language expresses that the grantees hold equal shares of the property as joint tenants, usually by providing a *right of survivorship*.[30]

Joint tenancy deeds may or may not represent arms-length transactions, depending on the circumstances.

Land Contract

- A *land contract* is a written legal contract used to purchase real estate or other real property. Also termed a *contract for deed*, a *land contract* is a form of seller financing similar to a mortgage but rather than make payments to a lender or bank, a buyer makes payments directly to the seller until the purchase price is paid in full.[31] The buyer is considered to possess an equitable title to the property while making payments and the seller retains legal title to the property until the final payment is made. The buyer's status

---

[28] Editors, *Understanding Your Deed: The Grant Deed, The Quitclaim Deed, and Interspousal Deed*, ESCROW OF THE WEST (July 31, 2018), https://escrowofthewest.com/grant-deed/.

[29] § 2:33. Constructive fraud—Common types of transfers—Intrafamily transfers, Fraudulent Transfers, Prebankruptcy Planning and Exemptions § 2:33 (West 2020).

[30] Joint tenancy deed, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

[31] Everything You Need to Know About Land Contracts, ROCKET MORTGAGE (Sep. 17, 2020), https://www.rocketmortgage.com/learn/land-contract.

as an equitable title holder grants them an interest in the property that precludes the seller from selling the property to a third party or creating a lien or encumbrance that would interfere with the buyer's interest in the property.[32]

Land contracts generally represent arms-length transactions. However, purchasers are often persons who are unable to obtain credit through usual mortgage markets due to various circumstances. Therefore, these contracts have been known to be used in abusive ways. The prices may be higher than true arms-length transactions.

Land Court (Massachusetts)

- In Massachusetts, the Land Court possesses exclusive, original jurisdiction over the registration of title to real property and over all matters and disputes concerning such title subsequent to registration.[33] After a thorough search and due process is afforded to all interested parties, the Land Court reviews, adjudicates, and decrees the state of the title to register it. Once registered, the state of the title is embodied in a certificate that not only evidences title but acts as a guarantee of title subject only to the statutory exceptions and federal law. Under the Land Court system, each registered property has a numbered current owner's certificate of title and is issued a new certificate when the property is sold or transferred. The Land Court supervises all Land Court records, and any documents affecting a certificate of title are referenced in an encumbrance list that accompanies that certificate.[34]

Land court deeds in Massachusetts may be arms-length or may not. More information would be needed to determine the nature of the transaction.

Leasehold Conversion with Agreement of Sale

- A *leasehold conversion with agreement of sale* is the conversion of a *leasehold interest* in a property to a *freehold interest*. Essentially a lease-to-own agreement, a tenant enters into an *agreement of sale* with the owner in fee of a property to pay an agreed upon portion of rent payments toward a down payment in order to purchase the property at some future point of time.[35]

Leasehold Conversion with Agreement of Sale transactions generally amount to arms-length transactions although the terms may include adjustments in the price for an option to purchase in the future.

---

[32] Kelsey Cooke, *The Basics of Land Contracts*, NOLO (last visited Jan. 2, 2021), https://www.nolo.com/legal-encyclopedia/the-basics-land-contracts.html.
[33] Commonwealth of Massachusetts, *Jurisdiction of the Land Court* (last updated Apr. 27, 2018), https://www.mass.gov/info-details/jurisdiction-of-the-land-court#overview-.
[34] Norfolk Co. Registry of Deeds, *Filing Requirements and Document Checklists* (last visited Dec. 29, 2020), https://www.norfolkdeeds.org/support/document-checklist.
[35] James Kimmons, *Using a Lease Purchase to Buy Property* (last updated July 8, 2019), https://www.thebalancesmb.com/lease-purchase-for-buyers-and-sellers-of-real-estate-2866975.

Life Estate Deed/Deed Reserving Life Estate

- A *life estate deed/deed reserving life estate* is a deed that transfers ownership of a subject property to one or more "remaindermen" while reserving in the grantor a life estate in the property. Typically employed as a way of avoiding probate, title to the property automatically vests in the remainderman/remaindermen upon the life tenant's death. While free of the life tenant's creditors, a property is still subject to any legally enforceable lien on the property upon transfer.[36]

A deed reserving a life estate often occurs between family members so probably do not amount to an arms-length transaction.

Limited Warranty Deed

- A *limited warranty deed* is a deed in which the grantor covenants to defend the title against only those claims and demands of the grantor and those claiming by and under the grantor. Also termed a *special warranty deed* and, in a few jurisdictions, a *quitclaim deed*.[37]

Special warranty deeds raise concerns as to whether the transaction is arms-length as special circumstances probably exist. For example, a trustee may use a special warranty deed. Quitclaim deeds usually indicate lack of an arms-length transaction. However, in some states in the Northeast quitclaim deeds are customary in arms-length transactions.

Partnership Deed

- A *partnership deed* is a document that outlines the rights and responsibilities of all parties to a business operation. A *partnership deed* is designed to guide partners in the conduct of business, including but not limited to how additional capital should be raised, actions to take in the event of a partner's death or withdrawal, and dissolution of the partnership.[38]

A partnership deed is used in the formation of a business organization and lacks indicia of an arms-length transaction.

---

[36] Randy Coleman, *The Problem with Life Estate Deeds*, (July 28, 2014), https://www.linkedin.com/pulse/20140728165911-7718716-the-problem-with-life-estate-deeds.

[37] Limited warranty deed, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

[38] Tom Streissguth, *What is the Meaning of a "Partnership Deed"?*, SFGATE (last updated Dec. 2, 2018), https://homeguides.sfgate.com/meaning-partnership-deed-1519.html.

Personal Representatives Deed

- A *personal representatives deed* is a deed used by a personal representative to convey property in the course of administering a decedent's estate. Compare to an *administrator's deed* or an *executor's deed*.[39]

Personal representatives deeds pass property at the death of the landowner, transactions that occur outside of the market.

Quitclaim Deed

- A *quitclaim deed* is a deed that conveys a grantor's complete interest or claim in certain real property but that neither warrants nor professes that the title is valid. Also termed *deed without covenants* or *deed without warranty*. Compare to a *warranty deed*.[40]

Quitclaim deeds usually indicate lack of an arms-length transaction. However, in some states in the Northeast quitclaim deeds are customary in arms-length transactions.

Re-recorded Deed

- A *re-recorded deed* is a replacement for a deed that has been either lost, destroyed, or removed. Unlike a *correction deed* that creates a duplicate deed save any corrected information that is filed in addition to the original deed, a *re-recorded deed* replaces the original deed entirely.[41]

The nature of the transaction underlying a re-recorded deed depends on the particular circumstances so no general rule can be set out.

Redemption Deed

- A *redemption deed* is obtained with property redeemed from a tax sale. In the case of foreclosure for unpaid taxes, the property owner has the opportunity to redeem their property during the "redemption period," the period of time that occurs after the property is placed on the foreclosure list with the courts. The redemption period varies by states, but payment of taxes, interest, penalty, and costs must be made before the property can be redeemed. In most states, only someone with a legal interest in the property can redeem it. Once payment of delinquent taxes and any additional costs have been made by a

---

[39] Personal representatives deed, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

[40] Quitclaim deed, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

[41] James Bruce Davis, *When's it OK to Re-record a Deed?*, 20 VLTA EXAMINER 14 (Spring 2014), *available at* https://www.beankinney.com/media/publication/340_March%202014%20-%20JBD%20-%20When_s%20it%20ok%20to%20re-record%20a%20deed%20_00428292xAC2B5_.pdf; *see also* § 19:84. Rerecording after correction, 2 GA. REAL ESTATE LAW & PROCEDURE § 19:84 (7th ed.) (West 2020).

person with a legal interest in the property during the redemption period, they receive a redemption deed for the property. [42]

A redemption deed occurs in the case of a tax sale, outside of market circumstances, and is not indicative of market price.

Referee's Deed

- A *referee's deed* is a document that conveys real property sold by court order, especially for a partition or a foreclosure.[43]

Referee's deeds occur in forced sale situations so lack the elements of an arms-length transaction.

Sheriff's Deed

- A *sheriff's deed* is a deed that gives ownership rights in property bought at a sheriff's sale.[44]

Sheriff's deeds occur in forced sales circumstances and do not reflect an arms-length transaction.

Special Warranty Deed

- A *special warranty deed* is a deed in which the grantor covenants to defend the title against only those claims and demands of the grantor and those claiming by and under the grantor. Also termed *limited warranty deed* and, in a few jurisdictions, a *quitclaim deed*. Compare to a *warranty deed*.[45]

Special warranty deeds raise concerns as to whether the transaction is arms-length as special circumstances probably exist. For example, a trustee may use a special warranty deed.

Special Warranty Deed with Vendors Lien

- A *special warranty deed with vendor's lien* is functionally similar to a *warranty deed with a vendor's lien* but more limited in scope. A *special warranty deed* covenants to the buyer only that the seller has not personally done anything to adversely affect the title being conveyed while the seller has owned the property whereas a *warranty deed* guarantees the property being purchased is free of encumbrances back to the property's origin.[46] A *vendor's lien* is a lien in favor of the vendor, or seller of the property. When a

[42] Julie Davoren, *What is a Redemption Deed?*, SAPLING (last updated Mar. 28, 2017), https://www.sapling.com/8654881/redemption-deed.

[43] Referees deed, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

[44] Sheriffs deed, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

[45] Special warranty deed, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

[46] Special warranty deed, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

special warranty deed contains a *vendor's lien*, it offers more protection to the vendor, or the property seller, than the buyer because the timeframe guaranteed to be free of mortgages or liens that may encumber the property is limited to the time the seller has held title. Used in situations where the buyer is taking immediate possession but paying in installments, a *special warranty deed with vendor's lien* acts like a specific kind of mortgage in permitting the seller (vendor) to retains the right to repossess the property until the buyer makes the final payment.[47] Until the buyer finishes paying, the seller possesses legal title to the property and the buyer cannot sell the property.[48]

Like a special warranty deed, a special warranty deed with vendors lien raises concerns as to whether the transaction is arms-length as special circumstances probably exist. For example, a trustee may use a special warranty deed.

Survivorship Deed

- A *survivorship deed* is a deed that creates a right of survivorship in the grantees. Specifically, a *survivorship deed* provides that ownership of a deceased tenant's share is automatically transferred to the surviving tenant.[49]

A survivorship deed may or may not indicate an arms-length transaction. The grantees are generally family members but the grantor may not be related.

Tax Deed

- A *tax deed* is a deed showing the transfer of title to real property sold for the nonpayment of taxes.[50]

Tax deeds result from forced sales of property so do not reflect an arms-length transaction.

Transfer on Death Deed

- A *transfer on death deed* is a deed that automatically conveys the property to a designated person upon the property owner's death. Also termed a *beneficiary deed*, a *transfer on death deed* avoids probate by allowing a property owner while living to make a grant of real property to the owner's heirs but does not effect a change of ownership until after the transferor's death.[51] While *transfer on death deeds* allow beneficiaries to

[47] John Willis, *What is a Vendor's Lien Deed?*, SAPLING (last updated Mar. 28, 2017), https://www.sapling.com/8466182/vendors-lien-deed.
[48] Pam McCallum, *What is a Texas Warranty Deed with Vendor's Lien?*, SAPLING (last updated Mar. 14, 2018), https://www.sapling.com/8796347/texas-warranty-deed-vendors-lien.
[49] Survivorship deed, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.
[50] Tax deed, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.
[51] Beneficiary deed, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

avoid probate, they are revocable and beneficiaries will inherit any legally enforceable lien on the property upon transfer.[52]

- As of September 2019, the District of Columbia and the following states allow some form of *transfer of death deed*: Alaska, Arizona, Arkansas, California, Colorado, Hawaii, Illinois, Indiana, Kansas, Maine, Minnesota, Missouri, Montana, Nebraska, Nevada, New Mexico, North Dakota, Oklahoma, Oregon, South Dakota, Texas, Utah, Virginia, Washington, West Virginia, Wisconsin and Wyoming. A few states, such as Michigan, have a similar but technically different document, commonly called a *Lady Bird deed*. A *Lady Bird deed* allows a property owner to transfer ownership of the property to another while retaining the right to hold and occupy the property and use it as if the grantor were still the sole owner.[53]

Transfer on death deeds convey the property to beneficiaries at the death of the landowner and do not reflect an arms-length transaction.

Trustees Deed (foreclosure sale transfer)

- A *trustee's deed (foreclosure sale transfer)* is a deed used by a trustee to convey trust property.[54] Under a trust deed, a trustee acts as a third party to a lender and lendee (buyer), holding legal title to the property for the lender as security against the loan. If the buyer fails to repay the loan (and, if in a state that requires it, the *deed of trust* contains a "power of sale" clause), the trustee may proceed with a nonjudicial foreclosure and foreclose on the property directly and sell it to pay back the loan. Upon sale, the purchaser receives a *trustee's deed*, which conveys full ownership but comes with no guarantees that the title is clean.[55]

Foreclosure sales occur as the result of a forced sale and likely do not reflect market conditions.

Trustees Deed (non-foreclosure)

- A *trustee's deed (non-foreclosure)* is a deed used by a trustee to convey trust property.[56]

A trustee's deed occurs in a variety of circumstances and may or may not reflect an arms-length transaction.

---

[52] What is a Transfer on Death Deed?, ROCKET LAWYER (last visited Jan. 2, 2020), https://www.rocketlawyer.com/article/what-is-a-transfer-on-death-deed.rl.

[53] Edward A. Haman, *Understanding the Transfer on Death Deed*, LEGALZOOM (last updated Sep. 26, 2019), https://www.legalzoom.com/articles/understanding-the-transfer-on-death-deed.

[54] Trustee's deed, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

[55] *Understanding a Trustee's Deed Upon Sale*, WESTERN RESOURCES TITLE (July 25, 2014), https://wrtca.com/understanding-a-trustees-deed-upon-sale/.

[56] Trustee's deed, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

Vendors Lien Deed

- A *vendor's lien deed* is functionally similar to a *warranty deed with vendor's lien*. Common to Texas, deeds often reserve an express *vendor's lien* to the seller in order to secure payment of a purchase money promissory note.[57] The seller and buyer agree in advance that there an interest retained by the seller in the property until the vendor is paid in full.[58] Similar to a mortgage, the transfer of legal title does not occur until the buyer completes payment of the purchase money but *vendor's lien deeds* are not executory.[59]

Vendors Lien deeds are used in a variety of circumstances. These deeds are not inherently arms-length or not.

Warranty Deed

- A *warranty deed* is a deed containing one or more covenants of title, especially a deed that expressly guarantees the grantor's good, clear title and that contains covenants concerning the quality of title, including warranties of seisin, quiet enjoyment, right to convey, freedom from encumbrances, and defense of title against all claims. Also termed *general warranty deed* or *full-covenant-and-warranty deed*.[60] Compare to *quitclaim deed* and *special warranty deed*.

Warranty deeds generally indicate an arms-length transaction but may also be used in some transactions that are not arms-length.

Warranty Deed with Vendors Lien

- A *warranty deed with vendor's lien* is an instrument used to transfer title of property with all the assurances of a *general warranty deed* but with a built-in protection for the "vendor," or seller of the property. A *warranty deed*, or *general warranty deed*, provides the most protection to buyers because it is a guarantee that title to the property is free of any mortgages or liens that could encumber the property, granting buyers recourse if that turns out not to be the case. A *vendor's lien* is a lien in favor of the vendor, or seller of the property. Used in situations where the buyer is taking immediate possession but paying in installments, a *general warranty deed with vendor's lien* permits the seller (vendor) to retains the right to repossess the property until the buyer makes the final payment. This type of deed is common where the buyer borrows the purchase price from the seller, which grants the seller with a vendor's lien the same rights as an institutional lender if the

---

[57] § 28.22. Effect on innocent purchasers of re-recording of deeds, 5 TEX. PRAC., LAND TITLES AND TITLE EXAMINATION § 28.22 (3d ed.); *see also* 44 Fla. Jur 2d Records and Recording Acts § 158.

[58] John Willis, *What is a Vendor's Lien Deed?*, SAPLING (last updated Mar. 28, 2017), https://www.sapling.com/8466182/vendors-lien-deed.

[59] § 28.22. Effect on innocent purchasers of re-recording of deeds, 5 TEX. PRAC., LAND TITLES AND TITLE EXAMINATION § 28.22 (3d ed.); *see also* 44 Fla. Jur 2d Records and Recording Acts § 158.

[60] Warranty deed, *Black's Law Dictionary* (11th ed. 2019), *available at* Westlaw.

> buyer fails to make scheduled payments. Until the buyer finishes paying, the seller possesses legal title to the property. [61]

Like warranty deeds, warranty deeds with vendors liens generally indicate an arms-length transaction but may also be used in some transactions that are not arms-length.

[61] Pam McCallum, *What is a Texas Warranty Deed with Vendor's Lien?*, SAPLING (last updated Mar. 14, 2018), https://www.sapling.com/8796347/texas-warranty-deed-vendors-lien.