The Costs of Housing Regulation: Evidence From Generative Regulatory Measurement

Alexander W. Bartik, Arpit Gupta, and Daniel Milo*

September 14, 2024

Abstract

We present a novel method called "generative regulatory measurement" that uses Large Language Models (LLMs) to interpret statutes and administrative documents. We demonstrate its effectiveness in analyzing municipal zoning codes, achieving 96% accuracy in binary classification tasks and a 0.92 correlation in predicting minimum lot sizes. Applying this method to U.S. zoning regulations, we establish five facts about American zoning: (1) Housing production disproportionately happens in unincorporated areas without municipal zoning codes. (2) Density in the form of multifamily apartments and small lot single family homes is broadly limited. (3) Zoning follows a monocentric pattern with regional variations, with suburban regulations particularly strict in the Northeast. (4) Housing regulations can be clustered into two main principal components, the first of which corresponds to housing complexity and can be interpreted as extracting value in high demand environments. (5) The second principal component associates with exclusionary zoning.

JEL-Classification: R52, R58, K11, O38, R31, C81

Keywords: housing regulation, zoning codes, large language models, natural language processing, artificial intelligence, municipal ordinances, retrieval augmented generation

*Bartik: Department of Economics, University of Illinois at Urbana-Champaign, abartik@illinois.edu. Gupta: New York University, Stern School of Business, ag5808@stern.nyu.edu. Milo: New York University, Stern School of Business, dm4766@stern.nyu.edu This paper has benefited from conversations with Elliott Ash, Leah Brooks, Sophie Calder-Wang, Tom Cui, Nestor Davidson, Tatyana Deryugina, Giles Duranton, Ingrid Gould Ellen, Chris Elmendorf, Edward Glaeser, Joseph Gyourko (discussant), John Horton, Greg Howard, Ben Hyman, Jeffrey Lin, Stijn Van Nieuwerburgh, Theresa Kuchler, Adam Osman, Franklin Qian, Julian Reif, Daniel Rock, Albert Saiz, Gregor Schubert, Allison Shertzer, Jaehee Song (discussant), Lena Song, Johannes Stroebel, Bryan Stuart, and comments from conference and seminar participants at NBER Summer Institute (Real Estate/Urban), NYU Stern Corporate Governance luncheon, NYU Law School, the Online Seminar in Economics and Data Science, Wharton real estate, UNC finance, AREUEA National Conference, the Dallas Fed, CREED Conference, and UIUC economics. Preeyonuj Boruah, David Dai, Amber Gao, Grace Getman, Drew Harrington, Karin Hobelsberger, Nagharjun Mariappan, Scott Overbey, Alok Ranjan, Ritul Soni, and Kevin Zhou provided excellent research assistance. We gratefully acknowledge funding from Emergent Ventures and the NYU Stern Center for Global Economy and Business, and thank the Pioneer Institute for providing data. Any errors or omissions are the responsibility of the authors. A replication package for the project can be found at: `https://github.com/dmilo75/ai-zoning`.

1 Introduction

Housing regulations play a crucial role in determining housing supply, significantly impacting housing affordability and urban development patterns (Glaeser and Gyourko, 2018; Gyourko et al., 2008). The influence of zoning laws and land-use policies extends well beyond real estate markets, with implications for social segregation, economic mobility, environmental sustainability, the growth of urban agglomerations, and construction sector productivity (Gyourko and Molloy, 2015; Kahn, 2000; Hsieh and Moretti, 2019; D'Amico et al., 2023). Despite their critical importance for such diverse outcomes, accurately measuring housing regulations remains challenging due to the complexity and variation of municipal ordinances. This measurement challenge hinders our ability to understand the fundamental drivers of housing regulation and their broader impacts.

Our paper argues that advances in Large language Models (LLMs) enable scalable and accurate classification of regulatory documents, a task which we refer to as *generative regulatory measurement*. We obtain municipal codes for 63% of the population governed by local zoning ordinances, and use state-of-the-art Artificial Intelligence (AI) tools to estimate housing regulation on the full text of these documents. Our core methodological innovation lies in developing a comprehensive approach to regulatory analysis using LLMs which combines Retrieval Augmented Generation (RAG), semantic embeddings, multi-step processing, and carefully engineered prompts. We apply this approach to a set of regulatory questions initially developed by the Pioneer Institute for Massachusetts (Glaeser and Ward, 2009), and benchmark our LLM-generated regulatory categorizations against the human-coded measurements from this same Pioneer Institute study.

Our results indicate that LLMs deployed on frontier models (GPT-4 Turbo by OpenAI) have achieved near-human rates of precision in classifying regulation, with an accuracy rate of 96% for binary questions. LLMs also perform strongly on numerical questions with a correlation of 0.7 between generated data and analyst responses overall, and 0.92 for residential lot size minimums, an important regulation governing density in single-family areas. To further support our methodology, we compare our generated data against existing regulatory surveys from the Wharton index (Gyourko et al., 2021). We find positive associations for specific questions common to both methodologies (affordable housing mandates and minimum lot sizes), as well as for overall indices.

We also manually verify a subset of housing regulations in California to ensure our results are not geographically biased.

We use the resulting LLM-produced dataset on national housing regulations, alongside other housing data, to establish five key facts about land use policies and their impacts across United States. First, zoning ordinances appear to be broadly binding for the new development of housing across the United States. We highlight a key distinction between incorporated areas—areas with established local governments often formed to implement stricter housing regulations—and unincorporated areas. Over the period from 2000–2020, we find that 42% of the growth in new housing units in the U.S. was concentrated in unincorporated areas, including a majority of new housing units in the South. This is despite the fact that unincorporated areas are home to only 23% of the overall population, are typically low-demand regions with low prices and rents, face high vacancy rates, and are far from municipal job centers and amenities. The heavy reliance on greenfield development in far exurban locations for a substantial portion of housing production indicates significant barriers to infill development within incorporated areas. This pattern suggests that current zoning regulations in established municipalities may be severely limiting housing supply growth where demand is highest.

Second, we turn to a closer analysis of the average level of housing regulations in incorporated regions, and argue they are characterized by a low-density bias. To do so, we generate a large set of housing regulation questions, beginning with a set of questions initially asked by the Pioneer Institute, and augment with additional questions on the process of housing regulation. Multifamily apartments are explicitly banned in the entirety of many municipalities, including 10% of municipalities in the top tercile of income. Zoning maps indicate that multifamily apartments are allowed in only 31% of the land area for a sample of the largest cities covering 18.2 million people. By contrast, 36% of the land area in these cities are reserved for single-family zoning only. Among single-family zoned areas in municipalities across the entire country, 66% have town-wide minimum lot size requirements above 5,000 square feet, 17% of requirements are above 10,000 square feet, and 7% have requirements exceeding half an acre. While exceptions and changes to these rules are possible through variance processes, these facts highlight how existing regulations mandate low-density housing uses and limit high-density housing construction. We find these regulatory restrictions on density strongly associate with observed levels of density in these areas, as well as higher prices, rents, and fewer affordable housing units. These results suggest that zoning regulations play a significant role in shaping the physical landscape of American cities.

Our third key finding is that that housing regulation varies within metropolitan areas in ways that are broadly consistent with a monocentric city model (Alonso, 1964; Muth, 1971), while also highlighting significant deviations. As standard monocentric city models predict, denser building is generally allowed in city centers, with stricter bulk regulations and lower density requirements found in inner-ring suburbs. However, we also observe substantial sorting of high-income households into expensive suburbs with higher minimum size requirements. This is in contrast to the most basic models in which all households, regardless of income, are assumed to prefer central locations to minimize commuting costs. Our findings suggest a more complicated spatial pattern, with some peripheral suburban locations able to maintain high prices and sorting by affluent households in areas with stricter zoning. We find that this pattern is particularly pronounced in the Northeast, which has substantially more onerous bulk regulation requirements than other regions of the country. This regional variation suggests that historical development patterns, local political economies, and path dependencies in urban form play significant roles in shaping contemporary zoning practices, alongside standard economic forces of agglomeration and transportation costs.

Fourth, we find that two key dimensions summarize much of the variation in housing regulations: the first of which is associated with regulatory *complexity*, and proxies for housing demand. Across our full set of regulatory questions, we find that the first principal component is high in regions that have high construction and high prices, and is low in regions with low construction and low prices. This pattern strongly suggests a factor tied to demand-side pressures, which are used by local governments to extract and redistribute some housing surplus. A typical and representative question, which receives high loading in this factor, is the presence of inclusionary zoning mandates, which require a portion of new housing developments include affordable units. Typically, these regulations are more prevalent in densely populated, centrally located cities within metropolitan areas, which also tend to support Democratic political candidates.

Our fifth and final key finding centers on the second principal component of our regulatory dataset, which we interpret as a measure of exclusionary zoning practices. This component is characterized by high loadings on bulk regulations, such as minimum lot size requirements, as well as procedural hurdles that collectively function as barriers to dense housing development. These regulatory tools are particularly prevalent in affluent, predominantly white suburban areas surrounding metropolitan centers, which tend to lean relatively more Republican politically.

We suggest a range of potential motivations for exclusionary zoning practices, including concerns about spillovers from low-income residents, such as perceived impacts on crime rates, traffic congestion, educational outcomes, and house prices more broadly. Notably, we find a robust association between these regulatory measures and indicators of local school performance and social mobility, suggesting a role for educational sorting. These regulatory practices constrain affordable housing supply in high-demand areas, but also intensify economic and racial segregation by effectively pricing out lower-income and minority households from neighborhoods with desirable amenities, particularly high-performing schools.

We interpret these facts in a framework of inter-municipal competition, in which local governments strategically and non-cooperatively select housing regulations. This view is consistent literature on Tiebout sorting and local public goods provision (Tiebout, 1956; Epple and Zelenitz, 1981), and traditional notions of zoning which emphasize sorting of households into different public goods regimes (i.e,. Fischel (1987), Hamilton (1975)), but points to distinct economic motives. High-demand areas, typically the cores of metropolitan regions and locations with elevated housing costs, are able to leverage their desirability to capture housing value. The regulatory mechanisms they employ can be viewed as implicit taxes on housing, effectively increasing the shadow price of residential development (similar to rent seeking behavior in other municipal contexts as in Diamond (2017)). A component of the housing surplus seized from new construction is redistributed back to residents in the form of public goods requirements or affordable housing mandates. Incumbents may also benefit by raising barriers on building similar units in their jurisdiction. The political economy equilibrium in these cities sustains this degree of redistribution, featuring a mix of poorer households as well as some rich households who value urban amenities sufficiently to pay the resulting implicit taxes (Brueckner, 1995, 2011). By contrast, suburban areas surrounding major metropolitan centers tend to implement exclusionary zoning practices. These regulations appear targeted at limiting perceived spillovers from the influx of lower-income residents, who are instead concentrated in urban cores, poorer suburbs, and far exurban areas.

A key economic contribution of our paper is to highlight these two distinct regimes of housing regulation, which contrasts with previous single-dimension characterizations of regulatory strictness, such as that proposed by Gyourko et al. (2008). The interaction between these two regulatory approaches creates a self-reinforcing cycle. Urban core regulations increase housing costs, especially for non-incumbents, which push some residents towards suburbs. In response, suburbs intensify exclusionary practices, thereby retaining high-income residents who are willing to trade off longer transit times to urban jobs and amenities for local public goods, while excluding low-income residents. This dual approach to regulation—value extraction in urban cores and exclusion in suburbs—results in suppressed housing supply and reduced density across U.S. municipalities. Consequently, a significant portion of housing production is displaced to far-flung exurban areas beyond the reach of restrictive municipal zoning, leading to sprawl and potentially inefficient land use patterns.

Another key contribution of our paper is the development of a general-purpose methodology to reliably and precisely measure the content of textual documents. While expert humans maintain an edge in exact classification, our LLM-based approach achieves high accuracy rates, comparable to those of skilled non-experts, and offers several advantages for researchers. First, it provides unprecedented scalability: we successfully apply our regulatory classification measure across thousands of municipalities, a task that would be expensive and time-consuming for human analysts. This scalability opens up possibilities for comprehensive regulatory analysis across multiple domains. Second, our approach ensures verifiability and auditability by prompting the LLM to provide specific supporting text from the regulatory documents, enabling independent verification of classifications. Third, our approach is highly adaptable, allowing researchers to easily incorporate changes in definitions or incorporate advancements in AI models, facilitating replication and refinement of measurements over time.

The broad applicability of our approach extends to various domains where textual analysis is crucial, including building codes, tax regulations, legal cases, financial reports, newspapers, and other uses. This versatility is particularly valuable as the volume and complexity of regulations continue to increase (Singla, 2023). Our methodology therefore represents a significant advancement in the field of regulatory analysis and text-based research more broadly.

There are several caveats to consider regarding the results generated using our approach. First, LLMs require the user to source documents about zoning, which may be stale or not encompass the broader regulatory environment (i.e., interactions between state and local law). Second, our dataset only spans a portion of the questions that make up the regulatory landscape. Third, given the ambiguity and substantial heterogeneity of regulation across municipalities, our approach requires the LLM to make judgment calls when answering questions (i.e., choosing the more common type of single-family home minimum lot size requirements within a district). Fourth, our dataset reflects what the law says, but not how it is administered in practice (i.e., whether variances are liberally granted). Still, our approach makes substantial progress in our understanding of zoning regulations and lays the groundwork for further research.

Contributions to Literature The central contribution of our project is the creation of a standardized, comprehensive dataset of zoning across the United States. Existing literature on housing regulations has primarily relied on three approaches, each with limitations. First, survey-based approaches such as the Wharton Regulatory Index (Gyourko et al., 2008, 2021; Huang and Tang, 2012) offer broad coverage on housing regulations, but are limited by response rates, fixed questionnaires, and the precision of respondents (Lewis and Marantz, 2019).

The second approach relies on other imputations of national housing regulation. Wedge-based approaches estimate housing regulations by examining the expected spatial macroeconomic distortions resulting from zoning. Examples in this literature include Hsieh and Moretti (2019), Glaeser et al. (2005), Herkenhoff et al. (2018), and Duranton and Puga (2019). Babalievsky et al. (2021) apply a similar production function based approach to impute the impact of commercial zoning impacts. Other national approaches have examined textual data; for instance Ganong and Shoag (2017) use a scaled count of judicial decisions on "land use." In a similar spirit, Stacy et al. (2023) use machine learning tools to identify newspaper articles discussing changes to zoning restrictions in eight metropolitan areas and classify them as either loosening or tightening zoning restrictions and then analyze the effects of these changes in regulation on housing supply and rents. Mleczko and Desmond (2023) use a pre-LLM natural language processing (NLP) approach to measure a set of zoning regulations for the same sample of municipalities covered by the Wharton Regulatory

Index (Gyourko et al., 2008).

The third strand in the literature uses more detailed analysis of specific regulations at more local levels. These include Shanks (2021) which focuses on Massachusetts and uses Machine Learning tools (Latent Dirichlet Allocation). California has also been the subject of detailed and specific analysis, focusing in particular on growth limitations (Quigley and Raphael, 2005; Jackson, 2016), as has Florida (Ihlanfeldt, 2007). Similarly, (Lutz, 2015) studies the effects of property tax reforms on growth limitations and impact fees in New Hampshire. Most prominent is the approach by the Pioneer Institute, which has engaged in explicit classification of zoning rules for 187 municipalities in the state of Massachusetts. Prior work by Glaeser and Ward (2009) establishes that regulatory intensity measured in this dataset does indeed associate with higher costs and lower construction. Gyourko et al. (2008) mention both the importance of this kind of detailed local analysis, as well as the challenges in scaling this approach to the national level:

"The proliferation of barriers and hurdles to development has made the local regulatory environment so complex that it is now virtually impossible to describe or map in its entirety. Glaeser et al. (2006) come closest to doing so.... However, the enormity of that effort prevents it from being replicated in other markets by a single research team."

Existing research on housing regulations therefore leaves significant gaps in our understanding of their measurement and impacts. National studies identify broad impacts of regulations on housing costs and construction but lack specificity on key drivers, while more detailed state-level analyses are geographically limited and may not be nationally representative. We argue that the practical difficulties behind the scaling up of this approach have now been addressed through the development of modern AI LLMs, which provide both the comprehensiveness and granularity of the state-based approaches along with the scale of the national regulatory studies. Reflecting this breakthrough, our approach provides detailed measured of zoning regulations for over 5800 municipalities, more than twice as many municipalities as previous national studies such as the Wharton Regulatory Index (Gyourko et al., 2021), while providing the as much granularity or more as state and local studies such as the Pioneer Institute study (Glaeser and Ward, 2009).

Additionally, we also contribute to the literature by testing the accuracy and usefulness of LLMs in creating novel regulatory and policy datasets. Existing research on AI models emphasizes both their promise in analyzing textual data (Zhao et al., 2023), as well as challenges with undesirable AI features such as "hallucination" and manufactured model output (Azamfirei et al., 2023). A broader contribution of our project is therefore a large-scale application of large language models to a complex regulatory and policy dataset generation task. A growing literature has begun to use LLMs for generative data purposes in existing textual, financial, and regulatory documents (Giesecke, 2023; Jha et al., 2023; Yang, 2023; Bybee, 2023; Hansen and Kazinnik, 2023), as well as for social science hypothesis generation (Horton, 2023). Hoffman and Arbel (2023) argues for the use of LLMs in "generative interpretation" in estimating the meaning of legal contracts. A growing literature also examines broader implications of Generative AI (Eisfeldt et al., 2023; Brynjolfsson et al., 2023), as well as the role of algorithms applied to real estate (Calder-Wang and Kim, 2023; Raymond, 2023).

2 Construction of National Housing Regulatory Database

2.1 Municipal Codes and Zoning

In the United States, local governments are "creatures of the state" subordinate to state control. Municipal corporations are authorized, subject to state law, to organize local government, and refer to cities, towns, villages, and other government units which function in that capacity. This concept largely overlaps with the Census definition of "incorporated place" which we use to organize our analysis.[1]

In most states, one of the powers granted to municipalities by the state government is control over local zoning decisions; indeed, the desire to control local zoning is a primary motive for incorporation. Zoning, broadly, consists of two key sets of regulations: land use regulations, which partition local land into distinct use classes, and bulk regulations, which restrict the density of buildings in different land use classes. Examples of bulk regulations include: coverage, setbacks, height restrictions, and floor area ratio caps. Other mandates and requirements, such as parking minimums, further constrain both commercial and residential development in different areas.[2]

[1]In several states the "township" form of government also has jurisdiction in zoning which aligns with the Census County Subdivision definition.

[2]States and municipalities also enact building codes, which govern the building and safety standards that new

Municipalities enforce laws by issuing municipal codes which outline local regulation in different domains. Some regulations apply broadly to all land within a jurisdiction; other regulations (such as minimum lot sizes) typically vary depending on the specific use class and district (i.e., single-family zoning, commonly referred to as R-1, or commercial or industrial). These ordinances are typically updated over time to reflect changes in local regulations, and are often aggregated by different companies online. Table 1 illustrates the breadth of our sample coverage. In total, we cover 25% of all municipalities in the US and 6% of all townships. This coverage is skewed to larger cities, and so of the 76% of of the population in the U.S. that live in either a municipality or a township, we have relevant municipal documents for 63% of the population. Panel B shows our underlying sources for the municipal codes in our sample. American Legal Publishing provides significant numbers of records in the Northeast and Midwest, Municode provides especially good coverage in the South as well as in the Midwest, and `Ordinance.com` provides substantial coverage of the West and Northeast.[3]

The primary dataset for our analysis consists of the full-text of municipal ordinances. At the municipality-level, we also draw on information on building permits data from the Census Building Permits Survey. We also connect to rent and price data drawn from the American Community Survey (ACS) at the municipality level.

2.2 Large Language Models

Large Language Models (LLMs) are a form of artificial intelligence that primarily handle sequential data such as sequences of words in textual data. LLMs are based on the deep learning "transformer" architecture as introduced in Vaswani et al. (2017). The key innovation is the "attention mechanism," enabling the model to focus on multiple words of the input text at once. This helps the model understand words in context, such as sentences or paragraphs. Transformers also represent a significant advancement in terms of both accuracy and runtime over previous models like Recurrent Neural Networks, which processed sequences linearly. LLMs are trained with semi-supervised learn-

construction needs to adhere to.

[3]When a municipality hosts its ordinance on multiple aggregators we prioritize Ordinance.com, and then Municode over American Legal Publishing. We show the percent of our sample that is from each source and so the share from Municode and American Legal Publishing understate the sources true coverage.

ing, first pre-training the model on a large corpus of text and subsequently fine-tuning the model with human feedback. After training, LLMs can generate human-like text, answer questions, summarize text, and generalize from their training to perform tasks they were never explicitly trained for, a concept known as zero-shot learning. This means the model does not need as an input explicit examples of additional training to perform well in an out-of-sample exercise, a key advantage we use in our analysis.

LLMs have several advantages and disadvantages relevant for our setting. The central advantage is scalability: we are able to load large quantities of municipal code data for classification and analysis, which exceeds the capacity of a typical human team to analyze at reasonable cost. Other advantages include the prospect for additional training, allowing for increased accuracy over time as LLMs improve in quality and additional training data is incorporated into the analysis. Another key benefit in creating a comprehensive and nationwide dataset is the benefit of applying a uniform and standard set of criteria for analysis, rather than relying on a group of human analysts who may employ idiosyncratic legal interpretation.

Drawbacks in using LLMs for this purpose include potentially inaccurate measurement and the need for manual sourcing of relevant documents. Inaccurate measurement stems from a failure to locate or interpret the relevant sections of legal code. LLMs can only process a limited amount of text at once, and so they require a reliable process to locate the most relevant parts of the ordinance for a given question. Additionally, legal interpretation requires many assumptions and nuances, and even though LLMs are likely exposed to legal interpretation in their training, they may need to be reprompted on them to ensure greater focus for the questions at hand. Even current state-of-the-art LLMs may inadvertently produce incorrect information, produce information with an incorrect degree of certitude, and potentially manufacture data output ("hallucination"). Possible biases in the responses are linked to the quality of training data, prompting, and multi-step processing steps, and so measurement error may or may not be classical depending on the explanatory variable of interest. Finally, relevant information to answer zoning regulation questions may be found in other legal documents. We provide the LLM with the municipal zoning ordinance for each municipality but not other potentially useful documents (i.e. state or county laws). We address these drawbacks by focusing on questions that can be answered with only a municipal zoning ordinance and measuring performance against human-defined categorizations of regulation.

2.3 Processing Municipal Codes Using LLMs

The length of typical zoning documents exceeds the context windows currently usable by current LLMs. To address this limitation, we use a standard framework in computer science known as "retrieval-augmented generation" (Lewis et al., 2020). The basic objective of this approach is to combine a large pre-trained language model with external information retrieval, in order to give the LLM the ability to "look up" information from a vast corpus of text during the generation process. We outline our general procedure in Figure 1.

The first step of our process is to download and scrape the sources of municipal codes listed in Table 1, which provides us with a large corpus of zoning documents relevant for our analysis. These municipal codes contain detailed housing and zoning regulations relevant for our study, and we filter out ordinances which do not contain zoning information by searching for key phrases, like common table headers (i.e. "Table of Uses") or zoning district names (i.e., R-1 for the first residential zoning district). We scrape each section within an ordinance separately, and partition sections so that they contain between 50 and one thousand tokens of text.[4] Any images of tables are transcribed using Amazon Textract. We then vectorize the textual content through embeddings, which are numerical representations of the text's semantic meaning.

We similarly embed the questions we want answered from the documents, which begins with the question base already used by the Pioneer Institute (i.e., "Is multifamily zoning allowed in this area as-of-right?"). We rephrase these questions from the original wording provided by the Pioneer Institute in order to produce a more simplified version which is easier for the LLM to parse. This primarily consists of breaking down compound questions.

With a question embedding and embeddings for each text section in hand, we select the most relevant information to show to the LLM. We use cosine similarity, a standard measure of distance between two vectors, to rank sections of text by how close their embedding is to the question embedding. We then refine this ranking by using a cross-encoder reranking model[5] on the top 50

[4]We use the OpenAI tokenizer where one token is roughly four characters of text.

[5]We specifically use the Cohere reranking model for this step.

sections of text, which processes the question and section text pairs simultaneously to determine the most semantically similar sections.[6] We then select text to show the LLM in order of highest relevance until a threshold of four thousand tokens is reached.

We include three key pieces of information to provide the LLMs. First, we include 4,000 tokens of relevant text to the LLMs. Second, we provide the rephrased zoning question, as described above to simplify model parsing. Third, we provide additional background information and assumptions. The background information and model assumptions were based on the Pioneer study (their "Issue Overview" and "Research Coding" sections for each question) when possible and were LLM generated otherwise. We further refined this background information to address areas of misinterpretation. Appendix 6 contains full information on the original Pioneer questions, our rephrased questions, as well as the additional background information and assumptions provided.

All three pieces of information are provided in a single call to the LLM, in order to produce model output which is our answer. Each answer consists of an open-ended argument followed by a parable answer (i.e. "Yes" or "No"). The open ended answer allows for humans to audit the reasoning path of the LLM and has been shown to increase performance by providing space for the LLM to think out loud. In many cases, to answer a specific question, we chain together multiple calls. Some pieces of information are queried prior to asking the question, which are called subtasks, to provide pre-processing or background research. For instance, when asking about the largest frontage requirement for all single family residential districts, we first ask the LLM to name all districts which allow single family housing. We do this as a separate step because the relevant text defining allowable uses in a district and the text defining frontage requirements for districts are typically in different sections of the ordinance with distant embedding vectors. Additionally, LLM performance is enhanced when it is only required to answer a direct single step question in each call. Finally, we provide a "system prompt" in which we tell the LLM that it is a municipal zoning expert, detail what the structure of the prompts for particular questions will be, and tell the LLM to think "step by step" to induce chain of thought reasoning.

We also engage in post-processing of certain questions, which functions to double-check answers.

[6] For some questions, we double-check the answer with different context creating by using keywords to rerank. Please see Appendix section 6 for more details on which questions we do this for and which keywords we check for.

For instance, an affirmative "Yes" to a question about whether townhouses/attached housing is allowed typically means the LLM has found affirmative evidence that such housing typologies are allowed, while an answer of "No" signifies either a lack of approval, or a lack of sufficient context for the LLM to answer the question. In such cases where an answer could indicate lack of information, we reprompt the LLM and directly use keywords like "townhouse" or "attached" to refine and rerank our search (instead of the reranking algorithm).

The key takeaway from our approach towards generative regulatory parsing is that, at least with models available at the time of writing, model accuracy improves substantially above simple "zero shot learning" examples given additional human input. We provide substantial human input in the areas of prompt engineering and providing background information as well as assumptions, which helps to focus the LLM on the relevant focus of the text. Additionally, we design a multi-step reasoning chain for each question to simplify the tasks required of the LLM in each sub-step. Such additional human processing is likely necessary in other contexts as well, at least until further advances in LLMs are made.

2.4 Model Validation with Pioneer Data

A critical step in assessing the performance of LLM-based approaches lies in comparing model-generated classifications against a ground truth benchmark. To do so requires a high-quality annotated reference dataset. The Pioneer dataset serves as an excellent starting point for our purposes, as previously mentioned, due to the expert classification of a large number of municipalities. The main drawback in using this dataset is the staleness of responses—with responses categorized as of 2004. Many regulations have changed in the intervening twenty years, and we have access only to the most recent zoning ordinances, not the ones that prevailed in that time period. Additionally, the Pioneer Institute relied on some outside information (i.e., directly contacting local regulatory bodies) in addition to municipal ordinance text. To address these issues, we construct a testing dataset based on 30 randomly chosen municipalities from the Pioneer Institute dataset, and 1) exclude question responses which relied on outside context, and 2) hand-correct inaccuracies in the original classification.[7]

Table 2 shows the performance results of our baseline GPT-4 Turbo model against the testing sample in Massachusetts. Among continuous questions (Panel A), our generated answers have an average correlation of 0.67 with the ground truth of expert classifications, after winsorization of our model at the 1% level and corrections of errors in the Pioneer sample. This represents a quite high benchmark and also incorporates substantial heterogeneity. When asking about the number of zoning districts in the municipality, we obtain a correlation of 0.98. When asking about the minimum of residential min lot sizes (i.e., the lot size requirement for R-1 zoned single family homes, an important zoning question determining allowable density), we find a quite high 0.92 correlation. These results suggest we are able to reach quite high model performance when matching against continuous numerical outcomes. The main challenge in our classification lies in the averages of minimum lot sizes across districts. This primarily reflects subjectivity in determining which districts to include in the calculation (i.e. whether to consider rural/estate districts as residential or whether to consider smaller fringe districts) rather than incorrectly parsing the minimum lot size within a given district. Our estimate of the minimum of all lot size requirements is more robust to these interpretative challenges that disproportionally affect inclusion of districts with large lot sizes (i.e. rural/estate districts), so we focus on this variable for the remainder of our analysis.

We find even higher model accuracy when measuring binary questions (i.e., those with a yes or no answer like whether "multi-family housing is allowed" which we measure perfectly across all municipalities). As shown in Panel B of Table 2, we observe a model accuracy of 96% across all binary questions. Because the raw accuracy measure may be biased depending on the base rate of answers, we also provide a Relative Squared Error (RSE) that compares each model result compared to a naive model which guesses the sample mode. We observe quite small RSEs as well.

[7]Due to the time-intensive nature of the expert correction step, we only check responses in which our LLM approach disagrees with the Pioneer Institute classification. This means that we potentially overstate model accuracy in cases in which the LLM agrees with the Pioneer Institute original classification; but that original classification was wrong.

2.5 Heterogeneity Across LLMs

Although our benchmark results appear quite accurate, we also contrast them with estimates drawn from other models. This analysis helps identify the strengths and weaknesses of each model, as well as any discrepancies between the model outputs and the reference data. Furthermore, performance analysis allows researchers to make informed decisions about which LLM is best suited for their specific use case and to identify areas for improvement in the models' knowledge and reasoning capabilities.

In Figure 2, we contrast model performance across GPT-4 Turbo (the benchmark model), Claude 3 Opus, and GPT-3.5 Turbo. In dark blue, we plot the percent correct for each model using the percent accuracy for binary variables (Panel A), and the correlation for continuous variables (Panel B). We also plot the frequency each model says "I don't know" in grey, which varies across each model and question type. Finally, we attribute the remainder as the incorrect percent for each model (shown in light blue). For binary questions, we find that GPT-4 Turbo is the highest performer, followed by Claude 3 and then by GPT-3.5 Turbo (which has an accuracy rate of around 80% for binary questions).[8] GPT 3.5 Turbo has a slightly better performance in terms of the correlation of continuous questions compared to Claude 3, while GPT 4-Turbo continues to have the best performance.

2.6 Understanding Model Errors

To better diagnose reasons for model error, in Figure 3 we provide a complete decomposition of all of the reasons for disagreement between GPT-4 Turbo and the original Pioneer Study on binary questions. We manually reviewed each question that GPT-4 Turbo disagreed with the Pioneer Institute, and present the reasons for discrepancies in a figure. We outline, for each of the questions, the specific reason for disagreement: whether the pioneer study was itself outdated or inaccurate and subsequently corrected, whether the LLM misinterpreted context (i.e., it was provided the correct information, and simply provided an inaccurate answer), whether the LLM missed the context,

[8] The observed performance difference may stem from fundamental features of GPT-4 Turbo and Claude 3. However, the performance gap could also be, at least in part, an artifact of our tailored optimization of prompt engineering and multi-step processing for GPT-4 Turbo. Had we invested equivalent effort in refining our approach for Claude 3, it's plausible that its performance might have been comparably enhanced or potentially superior.

and whether the answer itself was coded as incorrect but the true classification appears somewhat ambiguous. While ideally municipal regulations would identify a clear and unambiguous answer, we observe differences even among legal experts hired for the task of hand-classifying regulations. In principle, the ambiguous or unclear aspects of regulations can also be systematically classified through LLM-based approaches.

Largely, answers from the Pioneer Institute that our model did not match were due to changes in the underlying ordinance since the Pioneer Institute study roughly 20 years ago. LLMs missed the context in two cases, while in four cases the answer itself was ambiguous. The most important category for our purposes are cases in which the LLM misinterpreted the context—this happens in nine cases, most often with respect to whether townhouses are allowed and with permit caps or phasing. Six questions do not have this type of error happen at all. When considered over a large sample, these results appear promising in suggesting that errors are typically quite rare.

Importantly, the errors also appear balanced across false positives as well as false negatives. Table 3 provides a confusion matrix comparing our baseline GPT-4 Turbo model against the Pioneer classifications, separating true positives, false positives, true negatives, and false negatives. Our errors are equally represented among false positives as well as false negatives (six each), suggesting no obvious bias in our classification.

2.7 Additional Validation Checks

Comparison Against Wharton Index: To further validate our results, we compare our answers to another commonly used dataset of national housing regulation: the Wharton Index of (Gyourko et al., 2021). To do so, we scale up our generative regulatory measurement approach to the national level, asking the same set of questions in the Pioneer Institute data for a large sample of national municipalities.

In Panel A of Table 4, we first compare our questions with the Wharton approach on two questions which find overlap across the question bases: on affordable housing and minimum lot sizes. We find a sizable correlation between our measure of affordable housing and the one measured in the Wharton study of 0.38. We observe smaller, but still sizable correlations, between 0.18–0.37 when examining the minimum lot size questions. We also find positive correlations between regulatory indices that we construct (we describe our regulatory indices and their construction in more detail in Section 4.1 below) and the Wharton Index overall, between 0.11–0.33.

A positive, though not perfect, correspondence between generated data and surveys is consistent with prior literature (Lewis and Marantz, 2019) investigating the reliability of survey-based responses on land use regulation. This paper finds planners have an incomplete understanding of their own municipality's regulation, and survey responses across years are inconsistent. Additionally, our data rely on the most recent zoning codes we were able to find whereas (Gyourko et al., 2021) uses survey questions from 2018. Consequently, some of the divergence likely reflects changes in the underlying zonin codes over time.

Additional Hand Validation: To investigate the robustness of our accuracy estimates outside of the training sample state of Massachusetts, we conducted additional manual verification in the state of California. We randomly selected 30 municipalities from California and had a law student review the LLM's answers for two key questions: the minimum of residential minimum lot sizes and the longest frontage requirement. This validation process revealed that our model achieved 89% accuracy for minimum lot sizes and 89% accuracy for frontage requirements, after dropping ambiguous cases.

2.8 Scalability and Replicability of LLM Regulatory Analysis

While LLMs hold promise for regulatory analysis based on accuracy, their key advantage lies in scalability and cost-effectiveness, which enables substantially larger systematic analysis of unstructured textual datasets.

Appendix Figure A1 illustrates the cost comparison between human and LLM-based analysis across varying numbers of municipalities and total questions asked. Our analysis assumes several key parameters: law students require an average of five minutes per question-municipality pair at a rate of $50 per hour, while each LLM API calls cost $0.03 per query using GPT-4 Turbo pricing. We also account for initial setup costs, including $390 per question for preparation and $2.80 per municipality for data processing.

The key takeaway from the figure is that LLM-based approaches become substantially more cost-effective at larger scales. Human-based analysis shows a linear cost increase as the number of municipalities grows. In contrast, the LLM approach has a higher initial cost due to setup and model training, but demonstrates significantly better scalability. The cost curves intersect at approximately 300 municipalities, beyond which the LLM method becomes increasingly more cost-effective. The cost-effectiveness threshold is crossed at fewer municipalities with a higher number of questions asked. This cost structure highlights a key advantage of LLM-based analysis: its ability to handle large-scale regulatory reviews at a fraction of the cost of human analysis, which enables the novel systematic analysis of unstructured textual documents. However, it's important to note that while LLMs offer cost advantages at scale, they still require human oversight for quality control and handling complex interpretative tasks. Additionally, for smaller scale projects human-based analysis remains more cost effective given current LLMs.

Moreover, the LLM approach offers additional benefits not captured in pure cost comparisons. These include faster processing times (the entire dataset for this study can be generated in under three days), consistent application of criteria across all documents (while human research assistants might vary in their interpretation of law), and the ability to easily update analyses as regulations change.

Another important consideration in adopting LLM-based approaches for regulatory analysis is the replicability of results, which we discuss in more detail in Appendix Section 6. Unlike deterministic algorithms, LLMs can produce varying outputs for the same input due to their probabilistic nature and architectural features such as the Mixture of Experts (MoE) used in models like GPT-4. These design elements mean that two API calls are not guaranteed to deliver the same textual output, even when setting a seed.

We suggest some choices to researchers in this section to address this issue. A key choice under the control of the researcher is the "temperature," which is a hyperparameter in language models controlling the randomness of the model's output. A lower temperature (closer to 0) makes the model's responses more deterministic by picking the most probable next token. A higher temperature (closer to 1) increases randomness, resulting in more diverse and creative output. Our analysis reveals that lower temperature settings in LLM queries generally lead to more consistent responses, particularly for continuous variables. For instance, at a temperature of 0, we observe an internal consistency score of 0.9 for continuous questions, compared to 0.61 at a temperature of 1. External consistency, measured across separate API calls, also improves with lower temperatures and larger ensemble sizes (i.e., generating multiple responses to the same query and then aggregating these responses through majority rule).

Interestingly, we find that internal consistency scores are more predictive of external consistency and accuracy for higher temperature models. This suggests that within-model variation at low temperatures may not reflect the same sources of randomness as across-model runs, possibly due to factors like expert routing in MoE architectures.

To mitigate non-determinism, we find a tradeoff between a low-temperature single-shot approach with its high external consistency and cost-effectiveness, or a high-temperature ensemble approach with at least five iterations for improved reliability. These strategies balance the trade-offs between consistency, accuracy, and computational cost. Our main specification implements a low temperature single-shot approach, prioritizing external consistency and cost-effectiveness. This approach also reflects our interest in creating a single national dataset, for which consistent interpretation of zoning regulations across municipalities is important. However, for other applications requiring more creative or diverse output, the high-temperature ensemble approach may be more appropriate.

3 Spatial Patterns of Housing Regulation and Development

3.1 The Role of Unincorporated Areas in Housing Production

The first basic distinction we draw in the data is across incorporated areas, which are regions governed by a municipal corporation or a township, and unincorporated areas outside of the jurisdiction of such governments. Typically, local control over zoning decisions is a key factor behind the decision to incorporate. Unincorporated areas are still subject to housing regulations at the county and state levels, but typically face lower housing regulations. Unincorporated regions are generally lower inhabited regions located either in the exurban fringes of cities or in more distant rural areas.

To illustrate the nature of municipal boundaries, we show in Figure 4 incorporation maps around four large and representative metropolitan areas: Atlanta, Chicago, Philadelphia, and the San Francisco Bay Area. We pick these regions to illustrate patterns across a range of urban forms and geographical areas. Philadelphia represents an older, densely populated Eastern seaboard city with a long history of municipal incorporations. Chicago is an example of a Midwestern metropolis with a mix of urban density and suburban sprawl. Atlanta highlights a Southern city with extensive suburban development and growth in unincorporated areas. The San Francisco Bay Area highlights the West Coast region with complex topography and a patchwork of dense urban centers, suburbs, and unincorporated zones.

In these maps, we highlight incorporated areas in green and unincorporated areas in red, shading the areas to reflect housing unit density. Philadelphia and Chicago show a more continuous pattern of incorporation, with fewer and smaller unincorporated areas, reflecting their longer urban histories and denser development patterns. In contrast, Atlanta exhibits a more fragmented landscape with larger swaths of unincorporated land, particularly in its outer suburbs, indicative of its more recent, automobile-centric development. Perhaps surprisingly, some of this unincorporated land is relatively dense, reflecting built up activity, which is concentrated in the fringes of this unincorporated zone closest to the city itself. The San Francisco Bay Area is a more complex mosaic of incorporated cities interspersed with significant unincorporated areas, influenced by its varied topography, state parkland, and growth patterns.

We plot in Figure 5 the concentration of housing growth in unincorporated areas. Appendix Section 6 describes our construction of incorporated and unincorporated areas in more detail, which derive from shapefiles matched to the Government Units Survey of the Census of Governments. Panel A of this figure shows the percentage change in housing units from 2000-2020 as a fraction of the land incorporated in 2000 at the MSA level. Fully incorporated metros had zero net changes in housing growth over this period; the entirety of net housing growth in this period was in areas at least partially unincorporated. We show in Panel B the amount of new housing units added specifically in incorporated and unincorporated areas, generating two categories based on census block level housing data and the percent area of each block that is incorporated. We find 42% of new housing production takes place in unincorporated areas, despite these areas having only 23% of the total population. This trend is especially true in the South, where a majority of housing production takes place in unincorporated regions.

The fact that a large fraction of housing development in the United States consists of greenfield construction in unincorporated regions is potentially surprising given their remote nature and the consequent distance from local job centers and local amenities. We show in Table 5 associations between incorporated and unincorporated areas in the U.S. Unincorporated areas have substantially cheaper homes as measured prices and rents; higher vacancy rates, and residents have longer commutes.

Several factors contribute to this phenomenon. Unincorporated areas often have less restrictive zoning regulations, making it easier and less costly for developers to build new housing. Land costs tend to be lower in these areas, further incentivizing development. Unincorporated areas may also have a larger number of large plots of land for sale, reducing challenges associated with land assembly (Brooks and Lutz, 2016). Additionally, many incorporated municipalities have adopted growth control measures that limit new construction, pushing development to the fringes.

This pattern of development has significant implications for urban form and sustainability. While it allows for continued housing production in the face of restrictive zoning in incorporated areas, it also contributes to urban sprawl. This can lead to increased infrastructure costs, as new developments in unincorporated areas often require extensions of roads, utilities, and public services. Higher commute lengths contribute to limited employment opportunities and more environmental pollutants and congestion. These factors broadly suggest that development frictions in incorporated areas push housing supply to greenfield development in unincorporated areas, motivating our first fact.

Fact 1. *Housing development in the United States is concentrated in unincorporated areas.*

3.2 Prevalence of Density Restrictions

Next, we turn to the nature of regulations within incorporated regions. We focus first on density regulations, which directly limit the ability for infill development. In Panel B of Table 6, we highlight a number of density restrictions on multifamily apartments across the United States. 5% of municipalities prohibit multi-family housing entirely, which rises to 10% in the highest income tercile. 38% of municipalities ban mixed use developments (apartments above commercial units). An overwhelming majority of jurisdictions (86%) limit the conversion of single family or non-residential buildings to multi-family units. These estimates are typically lower when weighting by population, suggesting fewer constraints in more populated areas, but highlight broad hostility to apartment construction across a substantial part of the United States.

Further land use regulations restrict where dense housing can be built even in municipalities which allow it. To understand this set of regulations, we turn to land use zoning maps, which show allowable densities and housing typologies in different municipalities. We collect zoning maps for a sample of 31 large municipalities covering 18.2 million people, including Chicago, Seattle, Kansas City, Detroit, San Francisco, Austin, San Antonio, Tampa, Los Angeles, and San Diego. Our results, shown in Figure 6, highlight limitations on allowable densities even in some of the nation's largest cities. 36% of land area is zoned only for single-family zoning, while 41% of land area is zoned for single-family or duplex, and multifamily apartments are allowed in only 31% of land area.

Density restrictions also apply to single-family housing units through bulk regulations which limit lot and building size. Figure 10 shows the distribution of four different housing regulations across the US: number of zoning districts, largest frontage requirement, mean minimum lot size (across all zoning districts), and minimum minimum lot sizes (across all zoning districts). Frontage requirements specify the minimum width of a lot at the front property line, typically where it meets the street. These requirements effectively determine how wide a lot must be, influencing the overall lot size and the spacing between houses. Larger frontage requirements tend to create more spread-out neighborhoods with wider lots, potentially reducing the number of houses that can be built along a given street.

We also focus on minimum lot size requirements, which dictate the smallest allowable area for a residential plot. This regulation sets a lower bound on how small a piece of land can be for a single housing unit, impacting the potential density of an area. Larger minimum lot sizes result in fewer, more spread-out homes, while smaller minimums allow for denser development. They have been frequently estimated in prior research through bunching methods (Cui, 2024; Song, 2021) as important drivers of housing regulations; the contribution of our approach is to measure these regulations directly from municipal documents, rather than indirectly through their effects on observed housing development. We also find that the strictness of these regulations increases with municipal age of incorporation in Appendix Figure A9.

Figure 10 shows that these bulk regulations vary substantially across the U.S. In Appendix Figure A3, Panel A we show that among single-family zoned areas in municipalities, 66% have minimum lot size requirements above 5,000 square feet, 17% of requirements are above 10,000 square feet, and 7% have requirements exceeding half an acre. In contrast, the average size of a new home built in 2023 was 2,411 square feet, indicating that many jurisdictions have minimum lot size requirements considerably larger than typical of new housing construction across the United States. Panel B of this figure highlights the *maximum* residential minimum lot size requirement in the municipality: prevailing requirements for individual properties will therefore be between these two values. 89% of municipalities have a maximum requirement above 5,000 square feet; 64% have a requirement above 10,000 square feet; and 42% of municipalities have a maximum requirement above half an acre. 12% of municipalities have a maximum requirement above 100,000 square feet. These findings underscore the prevalence and magnitude of minimum lot size requirements across the United States.

3.3 Impact of Density Restrictions on Housing Markets

We then validate that these density restrictions do have implications for observed density in Table 7. In this Table, we compute the association between our measured regulations and housing values, building permits, rents, observed density, and the affordability of local housing units (rental units cost $\leq 30\%$ of state monthly median income, and owner-occupied units priced $\leq 3\times$ annual median household income). We present three sets of results on the relationship between regulations and housing outcomes. First, we use bivariate regressions to understand the unconditional associations between a given housing regulation and a particular housing outcome. Second, we estimate multivariate regression models using the full set of regulations using the Least Absolute Shrinkage and Selection Operator (LASSO) to investigate which regulations are most important and report conditional associations between these selected regulations and housing outcomes. Third, we use a random forest (XGBoost) specifications that allows for more complex interactions and potentially non-linear relationships between regulations and housing outcomes. For XGBoost there are multiple ways to measure feature importance but we use “gain" importance which is well suited to an environment that includes both binary and continuous variables.

We find that density restrictions (such as multifamily units not being allowed, minimum residential lot sizes, and frontage requirements) are strongly and robustly associated with lower observed density levels. The XGBoost model provides feature importance scores, normalized from 0 to 100, indicating the relative predictive power of each zoning regulation for various housing market outcomes, with higher scores suggesting greater influence on the predicted variable. The longest frontage requirement is identified as the most crucial factor in the random forest specification, with a maximum score of 100, closely followed by maximum (86) and minimum (84) residential lot size requirements.

While such a relationship is in principle consistent with lower demand in these regions, we also find that the same variables are strongly associated with higher house prices and rents, and a lower share of affordable housing units. In predicting median home values, the prohibition of multifamily housing is the most important predictor in the random forest specification, with a maximum score of 100. This is followed by affordable housing mandates (75) and prohibition of townhouses (56). We observe similar associations in these density regulations associating with higher median gross rents and a lower share of available affordable housing. Looking at affordability requirements has the added benefit of effectively normalizing for local income effects. These bulk regulations are also typically selected in the LASSO specifications on these outcomes, and have significant coefficients in bivariate regressions.

The consistent association between stringent zoning regulations, higher housing prices, and lower density points to an impact of these regulations on urban development patterns and housing markets. These regulations actively constrain housing supply in desirable locations. By restricting housing types, particularly in high-demand areas, these zoning practices appear to create artificial scarcity in the housing market

To further assess the predictive power of zoning regulations for housing market outcomes, we compare the performance of different modeling approaches in Table 8. This table reports the Root Mean Square Error (RMSE) for Lasso, OLS, and XGBoost models across various specifications predicting these same variables, contrasting the relative role of regulations and controls for land availability and income.

Zoning regulations demonstrate significant predictive power across all housing market outcomes. Including housing regulations improves model fit, even when other controls are already included. The XGBoost model consistently outperforms Lasso and OLS, particularly when all variables are included (column 7), achieving the lowest RMSE of 0.66 for median home value, suggesting the importance of complex, non-linear relationships between zoning regulations and housing market characteristics.

We summarize these various restrictions on density as our second fact.

Fact 2. *Housing density is limited across the United States through regulations on multifamily housing and small lot single-family housing.*

3.4 Monocentric City Model and Zoning Gradients

We next interpret municipal regulations in the context of the monocentric city model (Alonso, 1964; Mills, 1967; Muth, 1971). In these models, there is a central location in each city where production is concentrated and rents decay as one moves away from this productive center, with the rate of decay governed by transportation costs. These dynamics may also affect the benefits and costs of zoning regulations at different distances from the city center. Housing regulations, in turn, may then affect the rent gradient as one moves away from the city center.

We show various regulatory variables along the dimension of distance to city center in Figure 7. Affordable housing mandates are decreasing in distance from the center of the city, illustrating that these regulations are most commonly found at the centers of cities. Minimum lot size requirements show a different pattern, and vary markedly across regions. In much of the country outside the Northeast and Midwest, these regulations are fairly flat across the distance to city center. In some areas, these regulations are more commonly found in closer suburban regions to cities, while becoming less strict further away. The Northeast is a marked exception to this pattern, featuring an increasing relationship between these regulations and distance to the city center.

To further illustrate these patterns at the metropolitan level, Figures 11 and 12 show maps of minimum lot sizes and affordable housing incentives, respectively, for jurisdictions within the metropolitan areas surrounding four select cities in the U.S., Atlanta, Chicago, Philadelphia, and San Francisco.

These graphs document substantial variation in both minimum lot sizes and affordable housing mandates and incentives within metropolitan areas across municipalities, with the central city and inner suburbs having lower minimum lot sizes and higher rates of affordable housing policies than in jurisdictions farther from the central city. This figure illustrates a key advantage of our approach: the ability to construct measures of zoning ordinances at the level of the municipality across a wide variety of municipalities and regions in the United States.

We also show these associations in Table 9. Across all regions, we observe that the number of zoning districts decreases robustly with distance from the city center, suggesting simpler zoning structures in more peripheral areas. This pattern is consistent across all regions but is particularly pronounced in the Midwest and South. Some components of allowable density decrease with distance from the center, especially the permission of townhouses and mixed-use development. Interestingly, the allowance of multifamily housing shows a positive correlation with distance in the West and South, contrary to the general expectation of decreasing density with distance. This might reflect the presence of suburban multifamily developments in these regions.

The Northeast stands out with several distinct patterns. Unlike other regions, it shows increasing restrictiveness with distance for several measures. For instance, the longest frontage requirement, maximum residential minimum lot size, and mean residential minimum lot size all increase with distance from the city center in the Northeast, while these measures show no significant relationship in other regions. These results highlight the unique regulatory landscape of the Northeast, where bulk regulations and exclusionary zoning practices appear to intensify in suburban and exurban areas, contrary to patterns observed in other regions. These results are also especially surprising in the context of well-developed public transit and highway links in this area, which should, all else equal, facilitate greater development and density even outside of city centers in this region. The percentage of area unincorporated also increases robustly with distance from city center, as shown in Appendix Figure A2.

Several factors may help to explain this regional variation. The Northeast was the first region of the U.S. to urbanize and industrialize, leading to early adoption of zoning laws. Many of its suburbs were established earlier than in other regions, often as affluent enclaves seeking to preserve their character against urban expansion (Fischel, 2015). In addition, the region is characterized by a highly fragmented system of local governments, with many small, independent municipalities. This structure facilitates more localized and potentially more restrictive zoning policies. These suburbs also have a strong home role tradition of local control over land use decisions. The Northeast's early experience with industrial pollution and urban congestion may have fostered a culture of environmental protection that manifests itself in stricter land use controls, particularly for suburban lot size requirements, which were commonly justified on the basis of preserving natural land. Finally, the region has had particularly strong fights over access to local schools (i.e., school busing (Angrist et al., 2022)) that can increase the use of exclusionary zoning practices to maintain local school districts by limiting access to lower-income and minority households.

In contrast, California, which also has high house prices and where housing regulation is commonly thought to be tight, appears surprisingly to have more nationally typical bulk regulations. We show a map representation of these associations in Figure 8, which highlights the nature of high minimum residential lot size requirements in the northeast relative to California. In contrast, California's housing market is heavily influenced by state-level regulations such as environmental reviews and an onerous permitting process. We find some evidence of this in Appendix Figure A4, which shows that the West has the highest potential waiting time for review of a typical new multi-family building (see Mayer and Somerville (2000) for a link between regulatory uncertainty and reduced construction), and also ranks highly on public hearing requirements for multi-family building. These results therefore suggest that while the Northeast and California have high housing costs, they are the result of very different factors that may call for different policy responses. Furthermore, California's housing market faces significant constraints due to its challenging topography (Saiz, 2010), which limits developable land, and the high demand for natural amenities and local jobs. These factors can create a situation in which even relatively less restrictive local bulk regulations can result in binding constraints on housing supply, especially in combination with a challenging permitting process.

Including this spatial variation in regulation has several implications for urban economic models.

The broad patterns of zoning intensity by distance to city center is consistent with the predictions of standard monocentric city models. However, the sorting of high-income households into expensive, strictly zoned suburbs is inconsistent with the simplest models in which all households prefer central locations to minimize commuting costs, and the richest households locate centrally as they are able to pay the most for scarce urban land. The high market values of housing far from city centers and the choice of richer residents to locate there and commute long distances to urban jobs and amenities are a known challenge for this framework (Glaeser et al., 2008). This spatial pattern is more pronounced in the United States compared to other countries, where it is more typical for wealthy residents to reside in the city center. Our results suggest that this sorting is accompanied by regulatory restrictions on the minimum allowable housing size, which truncates the housing size distribution to the left, and results in sorting of higher-income residents to distant suburbs against the typical pattern expected in the monocentric city model.

The persistence of stringent zoning in affluent suburbs, often more restrictive than in central areas, suggests that regulatory requirements may lock in certain urban forms, contributing to the path dependency of metropolitan development patterns. The gradient in zoning complexity and intensity, particularly pronounced in regions like the Northeast, reflects differences in political processes, historical development, and local preferences between central cities and suburbs.

Fact 3. *Zoning regulations generally become less complex but more restrictive of density with increasing distance from city centers, with this pattern most pronounced in the Northeast where exclusionary measures intensify in suburban areas.*

4 Characterizing Housing Regulations

4.1 Principal Component Analysis of Housing Regulatory Dataset

To summarize our nationwide measure of housing regulations, we perform a PCA analysis. This technique reduces the dimensionality of our dataset by identifying key components that capture a large fraction of variation across our regulatory questions. Our analysis reveals two key principal components of housing regulation. Appendix Table A2 provides the loadings of each question on the first five principal components, and Figure 8 maps the first two principal components across the nation.[9]

In Panel B of Table 4, we find positive correlations of both PCs with the composite Wharton Index at the CBSA level. The first PC correlates at 0.33 against the Wharton Index, while the second PC correlates at 0.11. An index that sums the normalized values of each zoning regulation correlates at 0.22 with the Wharton index. These findings suggest that our regulatory measures overlap somewhat with existing measures of regulation, providing some reassurance of basic fit, but also seem to provide somewhat distinct information as reflected in the correlation being less than one.

4.2 Regulatory Complexity and Housing Markets

To better understand the economic interpretations behind these two principal components, and to disentangle the relative roles of demand and supply in housing production, we show in Figure 9 the associations between building permits, median house prices, and our two key principal components. We interpret the association of these two principal components with building permits and house prices in light of a simple model of supply and demand for housing. In places with rising demand for housing and inelastic supply, home prices will be high and building permits low (the lower, right quadrants of Panels A and B of Figure 9. In areas with rising demand for housing and elastic supply, there will be high house prices and high building permits (i.e. the upper right quadrant). In places with falling housing demand, there will be few building permits and low housing prices. In places with elastic housing supplies and constant or moderately rising demand, there will be low home prices and high permits per capita.

Panel A highlights that areas with a high value for the first principal component generally have high house prices as well as construction, while areas low in this dimension typically have low prices as well as building activity. This association suggests that the first principal component generally coincides with high housing demand environments. This interpretation is supported by the main regulatory loadings on this principal component (Appendix Table A2), which loads heavily on measures that are typically associated with more developed, high-demand housing markets. For

[9] Appendix Figure A5 also shows a heatmap of correlations between regulations at the national level.

instance, affordable housing and age restricted provisions are policy tools that are more likely to be implemented in areas with significant housing pressure and the administrative capacity to manage complex policies. Other associates of the first principal component relate to additional layers of local government: maximum review wait times and public hearing requirements. However, areas heavy in this component are much less likely to have bans or limits on multifamily and other dense housing typologies, though more likely to restrict permitting, require phasing, or restrict the inclusion of wetlands in lot size calculations.

Similarly, the first principal component correlates strongly with the number of zoning districts, which directly reflects the intricacy of the local regulatory environment. Because such counting measures are frequently used to proxy for complexity in different environments, we therefore interpret this principal component as representing "regulatory complexity." This term encapsulates not just the intricacy of the rules themselves, but also the sophistication of the regulatory approach, the diversity of policy tools employed, and the capacity of local governments to implement and manage these systems. It reflects a regulatory environment that has evolved in response to the challenges and opportunities presented by high-demand housing markets.

We show associations of the first principal component in Table 10. We focus on column 1, which controls for MSA fixed effects and compares municipalities within metropolitan areas. Consistent with largely a demand interpretation of this principal component: areas that are high on this dimension have a high college share, are higher in job density, have lower poverty rates, and have substantially higher shares of Democrats. These are high-demand areas which may be prone to extraction of value by local governments (Diamond, 2017).

To further support the notion that this PC associates with demand-side factors, Appendix Figure A12 illustrates the relationships between county-level principal components, weighted by population, and the industry-level establishments and employment per capita. Counties scoring high on the first principal component tend to host greater numbers of retail outlets (such as apparel stores and dining establishments) and professional services (including educational institutions, healthcare facilities, and cleaning services), while containing fewer gas stations, utility services, and truck transportation businesses, which are often linked to adverse externalities. This suggests that these areas are high-demand locations where certain rents can be extracted.

To be sure, such regulations may also affect housing supply. We explore this variable in more detail in columns 1–2 of Table 12, adding additional topographical and land availability controls such as the fraction of land developed in 2001, the squared fraction of land developed in 2001, and the fraction of land with a flat topography. Perhaps surprisingly, the first principal component only moderately associates with lower housing elasticities once MSA fixed effects are included. This suggests that the bundle of regulations represented by this set of housing regulations—lower bulk regulations and limits on density, and positive association with more process, complexity, and good provision requirements—nearly balance each other out in housing supply.

Fact 4. *The first principal component of housing regulation captures a dimension of housing complexity which varies with housing demand. This factor captures aspects of extraction of housing value.*

4.3 The Role of Exclusionary Zoning

Exclusionary zoning refers to land use regulations that limit housing density and types, often with the effect of excluding lower-income residents from certain areas. The role of exclusionary zoning in shaping socioeconomic patterns has been a subject of significant research and debate in urban economics and policy circles. Our analysis provides new insights into this phenomenon, leveraging our comprehensive dataset to examine its prevalence, distribution, and impacts across the United States. Regulations commonly linked to exclusionary zoning practices include measures such as large minimum lot sizes, restrictions on multi-family housing, and other bulk regulations that effectively increase the cost of housing in a given area. Our analysis reveals that the second principal component of housing regulations identified in this study correlates strongly with these measures of exclusionary zoning, particularly minimum lot sizes and other bulk regulations that limit density, the loadings for which we show in Appendix Table A2.

The spatial distribution of exclusionary zoning practices is not uniform across the United States. Our findings indicate that these practices are more commonly found in suburban areas, with a particularly strong presence in the Northeast region. This pattern suggests a spatial segregation effect, where more affluent suburbs use zoning regulations to maintain socioeconomic homogeneity and limit the influx of lower-income residents.

We show associations of the second principal component in Table 11. We focus on column 1, which controls for MSA fixed effects. The higher housing values and rents associated with areas scoring high on this component suggest that exclusionary zoning effectively creates pockets of affluence within metropolitan areas. This is further reinforced by the higher income levels and lower poverty rates in these localities. The lower share of affordable units in these areas is a direct consequence of these zoning practices, effectively pricing out lower-income households.

A plausible rationale for such practices are the spillovers of lower-income residents, including perceived impacts on crime rates, traffic congestion, educational outcomes, and house prices more broadly. Consistent with this rationale, the demographic patterns associated with exclusionary zoning correspond to a higher proportion of white residents and lower foreign-born share. This suggests racial and ethnic segregation effects of these policies. This aligns with analyses of zoning as a tool for maintaining racial homogeneity in the absence of explicit racial covenants (Rothwell, 2011; Cui, 2024). Density patterns favor single-family residents, with owner-occupied fractions substantially higher and housing unit density substantially lower.

Areas characterized by more exclusionary zoning practices also show higher average math test scores as well as rates of economic opportunity (Chetty et al., 2014). By contrast, there are lower rates of free lunch eligibility among students. This pattern suggests that exclusionary zoning effectively creates enclaves of educational privilege, where resources and positive peer effects are concentrated. Consistent with a public goods motive, we also observe higher local revenue per student, property tax rates, and total expenditures per capita. These results are consistent with classic theories of fiscal zoning which emphasize the role of zoning and property taxes in conjunction to support local public goods (Fischel, 1987; Hamilton, 1975, 1976). The presence of such public goods helps to explain the sorting of higher-income residents in these areas, which are commonly not in the city center, helping to explain the spatial patterns discussed in the previous section. Ultimately, the key factor driving these associations is the lack of school finance equalization in the United States, unlike in other countries, which allows for more local differentiation in school quality reinforced through land use regulations which impact the composition and tax base of the local population.

Fact 5. *The second principal component of housing regulation captures exclusionary zoning practices, which limit the presence of low-income residents through bulk regulations on minimum housing quantities.*

5 Discussion and Framework

Having discussed the construction of a national housing regulation dataset and five key facts which arise from the data, we discuss in this section an interpretive framework to put these facts in context. The key question that arises is: how can we characterize the tools of housing regulation in the hands of municipalities, and what accounts for the observed spatial variation in these regulations?

The first principal component of our housing regulation dataset, corresponding to higher complexity, can be seen roughly as a set of regulations on the shadow price of construction. This is because regulations such as mandatory inclusionary zoning can be seen as an additional implicit tax on new development, with the proceeds either redistributed to other residents (as in the case of affordable housing units), or else extracted by the government for the purpose of public goods or private benefits. While these are not regulations explicitly on the price of development, they may be interpreted as a shadow price or cost on housing.

By contrast, the second principal component, associated with exclusionary zoning, primarily consists of regulations on the quantity of housing built in the form of various bulk regulations, particularly on minimum lot sizes. These regulations, as discussed above, effectively truncate the housing quality distribution on the left side of the distribution, reducing the number of lower-income residents able to live locally. We connect such exclusionary motives to educational sorting at local levels.

The complexity or extraction motive for housing regulation shows up most strongly in the centers of American cities, where demand for living is highest and so the ability for local governments to extract value is correspondingly higher as well. By contrast, exclusionary regulations are commonly found in suburban areas in the peripheries of those cities across the United States, but are particularly pronounced in the Northeast.

This spatial regulatory pattern contributes to the sorting of Americans along the dimensions of age and income. Cities are home to poorer residents to access smaller housing, wealthy residents, as well as younger working households. By contrast, suburban areas are more typically home to young families with children who pay particular attention to the level of school spending and composition of the local school district.

The combination of these two sets of regulations combine to limit housing production and density across American cities, and thereby pushes out additional housing production to the far exurban fringes of cities where demand is low, but supply constraints are less binding.

Our results help to connect classic previous theories of zoning. (Fischel, 1987) and (Hamilton, 1975, 1976) argue that housing regulation and property taxes can create efficient public goods provision in the context of (Tiebout, 1956) sorting. By contrast, (Zodrow and Mieszkowski, 1986) argue instead that local taxes are distortionary and function like excise taxes, while more recent scholarship has emphasized the segregation motives of zoning, especially exclusionary zoning (Rothstein, 2017; Cui, 2024). Our two sets of regulatory controls by municipalities spans this prior literature, and helps to explain the circumstances under which housing regulation can appear extractive, and the conditions under which they sustain local public goods investment under exclusionary environments.

Figure A8 provides a visual scatterplot representation of the relationship between these two principal components across U.S. municipalities. The plot also highlights the heterogeneity in zoning practices across municipalities. Suburban areas around high-demand metropolitan areas (such as Darien, CT or Billerica, MA) rank highly on both PCs. High-demand urban areas (like Santa Ana, CA) score low in PC2, but are relatively high in PC1. Low-demand urban areas (like Cleveland, OH) rank low on both PCs. While exclusionary suburbs of low-demand areas (like Novi, MI) rank high in PC2 but low in PC1.

6 Conclusion

This study makes significant progress in using large language models (LLMs) to accurately measure and analyze complex zoning regulations across a broad sample of U.S. municipalities. Our results demonstrate that state-of-the-art LLMs can achieve near-human levels of accuracy in classifying zoning rules from textual documents, with accuracy levels of 96% for binary questions and correlations as high as 0.92 for continuous questions like minimum lots sizes relative to existing hand-classification from the Pioneer Institute. Our approach also correlates with existing measures of regulation from the Wharton Index. This generative regulatory measurement approach therefore enables the creation of a comprehensive, nationwide dataset of municipal zoning regulations. By leveraging LLMs to extract structured information from unstructured textual data, our methodology opens up new avenues for analyzing vast amounts of previously untapped regulatory documents across multiple domains.

We make all collected data and the associated replication code publicly available. Our AI-driven approach is scalable, auditable, and allows for refinement as LLMs continue to advance. With further development, this generative regulatory measurement framework can be extended to building codes, regulations in other domains, across different countries and languages, and to other regulatory contexts.

We use the resulting dataset on housing regulations across the United States to establish five key facts about housing regulation. Our results point to the importance of distinguishing between two typologies of housing regulation: one which is extractive in nature, and another which is instead primarily focused on exclusion within neighborhoods. These can be thought of as regulations on the shadow price and quantity of housing, respectively. Spatially, we find that extractive regulations are found in high-demand areas, such as the centers of cities, while exclusionary zoning practices are found in the suburban peripheries around city centers. These two sets of regulations combine to reduce allowable density within incorporated areas, pushing incremental development in the United States outside these jurisdictions into unincorporated regions.

References

Alonso, William, *Location and land use: Toward a general theory of land rent*, Harvard university press, 1964.

Angrist, Joshua, Guthrie Gray-Lobe, Clemence M Idoux, and Parag A Pathak, "Still worth the trip? School busing effects in Boston and New York," Technical Report, National Bureau of Economic Research 2022.

Azamfirei, R, SR Kudchadkar, and J Fackler, "Large language models and the perils of their hallucinations," *Critical Care*, 2023, *27* (1), 1–2.

Babalievsky, F, K Herkenhoff, LE Ohanian, and EC Prescott, "The Sky is Not the Limit: The Macroeconomic Consequences of Commercial Zoning Laws," Technical Report 2021.

Baum-Snow, Nathaniel and Lu Han, "The microgeography of housing supply," *Journal of Political Economy*, 2024, *132* (6), 1897–1946.

Brooks, Leah and Byron Lutz, "From Today's City to Tomorrow's City: An Empirical Investigation of Urban Land Assembly," *American Economic Journal: Economic Policy*, August 2016, *8* (3), 69–105.

Brueckner, Jan K, "Strategic control of growth in a system of cities," *Journal of Public Economics*, 1995, *57* (3), 393–416.

_ , *Lectures on Urban Economics*, MIT press, 2011.

Brynjolfsson, Erik, Danielle Li, and Lindsey R Raymond, "Generative AI at work," Technical Report, National Bureau of Economic Research 2023.

Bybee, J Leland, "The ghost in the machine: Generating beliefs with large language models," Technical Report, Technical report, Working Paper 2023.

Calder-Wang, Sophie and Gi Heung Kim, "Coordinated vs efficient prices: The impact of algorithmic pricing on multifamily rental markets," Technical Report, Working Paper 2023.

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez, "Where is the land of opportunity? The geography of intergenerational mobility in the United States," *The Quarterly Journal of Economics*, 2014, *129* (4), 1553–1623.

Cui, Tianfang, "Did Race Fence Off The American City? The Great Migration and the Evolution of Exclusionary Zoning," Technical Report, Working Paper 2024.

D'Amico, Leonardo, Edward L Glaeser, Joseph Gyourko, William Kerr, and Giacomo AM Ponzetto, "Why Has Construction Productivity Stagnated? The Role of Land-Use Regulation," Technical Report, Working Paper 2023.

Diamond, Rebecca, "Housing supply elasticity and rent extraction by state and local governments," *American Economic Journal: Economic Policy*, 2017, *9* (1), 74–111.

Duranton, Gilles and Diego Puga, "Urban growth and its aggregate implications," Technical Report, National Bureau of Economic Research 2019.

Eisfeldt, Andrea L, Gregor Schubert, and Miao Ben Zhang, "Generative AI and firm values," Technical Report, National Bureau of Economic Research 2023.

Epple, Dennis and Allan Zelenitz, "The implications of competition among jurisdictions: Does Tiebout need politics?," *Journal of Political Economy*, 1981, *89* (6), 1197–1217.

Fischel, William, *Zoning Rules!: The Economics of Land Use Regulation*, Lincoln Institute of Land Policy, 2015.

Fischel, William A, *The economics of zoning laws: A property rights approach to American land use controls*, JHU Press, 1987.

Ganong, Peter and Daniel Shoag, "Why has regional income convergence in the US declined?," *Journal of Urban Economics*, 2017, *102*, 76–90.

Giesecke, Oliver, "State Government Debt Structure," *Available at SSRN*, 2023.

Glaeser, Edward and Joseph Gyourko, "The economic implications of housing supply," *Journal of Economic Perspectives*, 2018, *32* (1), 3–30.

Glaeser, Edward L and Bryce A Ward, "The causes and consequences of land use regulation: Evidence from Greater Boston," *Journal of Urban Economics*, 2009, *65* (3), 265–278.

_ **, Joseph Gyourko, and Raven Saks**, "Why is Manhattan so expensive? Regulation and the rise in housing prices," *The Journal of Law and Economics*, 2005, *48* (2), 331–369.

_ **, Matthew E Kahn, and Jordan Rappaport**, "Why do the poor live in cities? The role of public transportation," *Journal of urban Economics*, 2008, *63* (1), 1–24.

Gyourko, Joseph, Albert Saiz, and Anita Summers, "A new measure of the local regulatory environment for housing markets: The Wharton Residential Land Use Regulatory Index," *Urban Studies*, 2008, *45* (3), 693–729.

_ **and Raven Molloy**, "Regulation and housing supply," in "Handbook of regional and urban economics," Vol. 5, Elsevier, 2015, pp. 1289–1337.

_ **, Jonathan S Hartley, and Jacob Krimmel**, "The local residential land use regulatory environment across US housing markets: Evidence from a new Wharton index," *Journal of Urban Economics*, 2021, *124*, 103337.

Hamilton, Bruce W, "Zoning and property taxation in a system of local governments," *Urban Studies*, 1975, *12* (2), 205–211.

_ , "Capitalization of intrajurisdictional differences in local tax prices," *The American Economic Review*, 1976, *66* (5), 743–753.

Hansen, Anne Lundgaard and Sophia Kazinnik, "Can ChatGPT Decipher Fedspeak?," *Available at SSRN*, 2023.

Herkenhoff, Kyle F, Lee E Ohanian, and Edward C Prescott, "Tarnishing the golden and empire states: Land-use restrictions and the US economic slowdown," *Journal of Monetary Economics*, 2018, *93*, 89–109.

Hoffman, David A and Yonathan A Arbel, "Generative interpretation," Technical Report, Working Paper 2023.

Horton, John J, "Large language models as simulated economic agents: What can we learn from homo silicus?," Technical Report, National Bureau of Economic Research 2023.

Hsieh, Chang-Tai and Enrico Moretti, "Housing constraints and spatial misallocation," *American Economic Journal: Macroeconomics*, 2019, *11* (2), 1–39.

Huang, Haifang and Yao Tang, "Residential land use regulation and the US housing price cycle between 2000 and 2009," *Journal of Urban Economics*, 2012, *71* (1), 93–99.

Ihlanfeldt, Keith R, "The effect of land use regulation on housing and land prices," *Journal of Urban Economics*, 2007, *61* (3), 420–435.

Jackson, Kristoffer, "Do land use regulations stifle residential development? Evidence from California cities," *Journal of Urban Economics*, 2016, *91*, 45–56.

Jha, Manish, Jialin Qian, Michael Weber, and Baozhong Yang, "ChatGPT and corporate policies," Technical Report, Chicago Booth Research Paper 2023.

Kahn, Matthew E, "The environmental impact of suburbanization," *Journal of Policy Analysis and Management*, 2000, *19* (4), 569–586.

Lewis, Patrick, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela, "Retrieval-augmented generation for knowledge-intensive nlp tasks," *Advances in Neural Information Processing Systems*, 2020, *33*, 9459–9474.

Lewis, Paul G and Nicholas J Marantz, "What planners know: Using surveys about local land use regulation to understand housing development," *Journal of the American Planning Association*, 2019, *85* (4), 445–462.

Li, Junyou, Qin Zhang, Yangbin Yu, Qiang Fu, and Deheng Ye, "More agents is all you need," Technical Report, Working Paper 2024.

Lutz, Byron, "Quasi-Experimental Evidence on the Connection between Property Taxes and Residential Capital Investment," *American Economic Journal: Economic Policy*, February 2015, *7* (1), 300–330.

Mayer, Christopher J and C Tsuriel Somerville, "Land use regulation and new construction," *Regional Science and Urban Economics*, 2000, *30* (6), 639–662.

Mills, Edwin S, "An aggregative model of resource allocation in a metropolitan area," *The American Economic Review*, 1967, *57* (2), 197–210.

Mleczko, Matthew and Matthew Desmond, "Using natural language processing to construct a National Zoning and Land Use Database," *Urban Studies*, 2023, *60* (13), 2564–2584.

Muth, Richard F, "The derived demand for urban residential land," *Urban studies*, 1971, *8* (3), 243–254.

Quigley, John M and Steven Raphael, "Regulation and the high cost of housing in California," *American Economic Review*, 2005, *95* (2), 323–328.

Raymond, Lindsey, "The Market Effects of Algorithms," Technical Report, Working Paper 2023.

Rothstein, Richard, *The Color of Law: A Forgotten History of how our Government Segregated America*, Liveright Publishing, 2017.

Rothwell, Jonathan T., "Racial Enclaves and Density Zoning: The Institutionalized Segregation of Racial Minorities in the United States," *American Law and Economics Review*, 2011, *13* (1), 290–358.

Saiz, Albert, "The geographic determinants of housing supply," *The Quarterly Journal of Economics*, 2010, *125* (3), 1253–1296.

Shanks, Brendan, "Land Use Regulations and Housing Development," 2021. manuscript.

Singla, Shikhar, "Regulatory costs and market power," Technical Report 2023.

Song, Jaehee, "The effects of residential zoning in US housing markets," Technical Report, Working Paper 2021.

Stacy, C, C Davis, YS Freemark, L Lo, G MacDonald, V Zheng, and R Pendall, "Land-use reforms and housing costs: Does allowing for increased density lead to greater affordability?," *Urban Studies*, 2023. 00420980231159500.

Tiebout, Charles M, "A pure theory of local expenditures," *Journal of Political Economy*, 1956, *64* (5), 416–424.

Vaswani, Ashish, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez,

Łukasz Kaiser, and Illia Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, 2017, *30.*

Yang, Stephen, "Predictive Patentomics: Forecasting Innovation Success and Valuation with ChatGPT," Technical Report, Working Paper 2023.

Zhao, Wayne Xin, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong et al., "A survey of large language models," *arXiv preprint arXiv:2303.18223*, 2023.

Zodrow, George R and Peter Mieszkowski, "Pigou, Tiebout, property taxation, and the underprovision of local public goods," *Journal of Urban Economics*, 1986, *19* (3), 356–370.

Tables

Table 1: Sample Coverage

Panel A: Sample and Local Government Coverage Metrics

Coverage Metrics:	National	Northeast	Midwest	South	West
Total Munis	19,488	2,101	8,481	6,587	2,319
% of Munis in Sample	25	32	19	22	48
Total Townships	16,213	4,111	12,102	0	0
% of Townships in Sample	6	23	0	-	-
Total Pop. (Millions)	331	57	69	127	77
% of Pop. Under Local Gov.	76	100	95	55	78

Panel B: % of Pop. Under Local Gov. Covered By Sample

Ordinance Aggregator:	National	Northeast	Midwest	South	West
American Legal Publishing	11	15	15	6	8
Municode	23	1	19	54	12
Ordinance.com	30	52	12	1	60
Total	63	68	46	61	80

Notes: For local governments available in multiple datasets, we prioritize using Ordinance.com and then Municode and reflect that in the population count. We also adjust for geographical overlap between townships and municipalities in tallying population by using census block level population data and corresponding shape files. We use population estimates from the 2022 Census of Governments for municipality population, and 2022 State-Level Census Population Data for census region and national population.
Links to data sources are American Legal Publishing, Municode, and Ordinance.com.

Table 2: Accuracy Metrics of GPT-4 Turbo on Validation Sample

Panel A: Continuous Questions

Question	RSE	Correlation
How many zoning districts, including overlays, are in the municipality?	0.06	0.98
What is the longest frontage requirement for single-family residential development in any district?	1.16	0.70
Minimum of Min Lot Sizes (Square Feet)	0.73	0.61
Mean of Min Lot Sizes (Square Feet)	14.77	0.39
Minimum of Residential Min Lot Sizes (Square Feet)	0.16	0.92
Mean of Residential Min Lot Sizes (Square Feet)	11.80	0.44
Cumulative Average	4.78	0.67
Cumulative Median	1.16	0.67

Notes: For Relative Squared Error (RSE) we compare the model's results to the naive model that guesses the sample mean. The correlation column is the correlation between the model answer and the Pioneer Institute answer. We calculate performance metrics and sample means (for RSE) only on the set of question municipality pairs that GPT-4 Turbo does not say "I don't know." We winsorize data from our models at the 1% level but do not winsorize data from the Pioneer Institute. The Cumulative Average and Cumulative Median are calculated across questions giving equal weight to each question.

Panel B: Binary Questions

Question	RSE	% Accuracy
Is multifamily housing allowed, either by right or special permit (including through overlays or cluster zoning)?	0.00	100%
Are apartments above commercial properties (mixed-use) allowed in any district?	0.07	96%
Is multifamily housing listed as allowed through conversion (of either single-family houses or nonresidential buildings)?	0.08	96%
Are attached single-family houses (townhouses, 3+ units) listed as an allowed use (by right or special permit)?	0.30	90%
Does the zoning include any provisions for housing that is restricted by age?	0.14	96%
Are accessory or in-law apartments allowed (by right or special permit) in any district?	0.09	96%
Is cluster development, planned unit development, open-space residential design, or another type of flexible zoning allowed by right?	0.00	100%
Is cluster development, planned unit development, open-space residential design, or another type of flexible zoning allowed by special permit?	0.00	100%
Does the zoning bylaw/ordinance include any mandates or incentives for development of affordable units?	0.00	100%
Is there a town-wide annual or biannual cap on residential permits issued, and/or is project phasing required?	0.33	90%
Are there restrictions on counting wetlands, sloped land or easements in lot-size calculations?	0.14	96%
Cumulative Average	0.11	96%
Cumulative Median	0.09	96%

Notes: For Relative Squared Error (RSE) we compare each model's results to the naive model that guesses the sample mode. The accuracy column is calculated as the percent of municipalities where the model matches the adjusted Pioneer Institute answer for each question. We drop any errors where the answer is considered ambiguous. For details on adjustments to the Pioneer data, see Figure 3.

Table 3: Confusion Matrix For Binary Performance Results

Question	True Positive	False Positive	True Negative	False Negative	True Positive Rate	False Positive Rate	Precision
Multifamily Allowed	28	0	2	0	1.00	0.00	1.00
Mixed-Use Buildings	15	0	14	1	0.94	0.00	1.00
Conversion to Multifamily	12	1	17	0	1.00	0.06	0.92
Townhouses Allowed	18	1	9	2	0.90	0.10	0.95
Age-Restricted Provisions	22	0	7	1	0.96	0.00	1.00
Accessory Apartments Allowed	18	0	11	1	0.95	0.00	1.00
Flexible Zoning by Right	1	1	27	0	1.00	0.04	0.50
Flexible Zoning by Permit	26	0	3	0	1.00	0.00	1.00
Affordable Housing	22	0	7	0	1.00	0.00	1.00
Permit Cap or Phasing	8	2	19	1	0.89	0.10	0.80
Wetlands Restricted in Lot-Size Calc	23	1	6	0	1.00	0.14	0.96
Total	193	6	122	6	0.97	0.05	0.97

Notes: True Positive refers to an outcome where the model correctly predicts the positive class. False Positive is an outcome where the model incorrectly predicts the positive class. True Negative denotes an outcome where the model correctly predicts the negative class. False Negative represents an outcome where the model incorrectly predicts the negative class. The true positive rate (also known as sensitivity or recall) is the proportion of actual positive cases correctly identified by the model. The false positive rate (also known as the false alarm rate or fall-out) is the proportion of actual negative cases incorrectly identified as positive by the model. Precision (also known as positive predictive value) is the proportion of positive identifications that are actually correct. See Panel B footnote of Table 2 for details about the sample.

Table 4: Additional Validation Checks

Panel A: Averages and Correlation For Wharton Questions

Question		Wharton Average	Our Average	Correlation
Affordable Housing		0.20	0.06	0.38
Minimum Lot Size	Less than 1/2 acre	0.50	0.49	0.37
	1/2 to 1 acre	0.17	0.13	0.18
	1 to under 2 acres	0.12	0.17	0.26
	2 acres or more	0.22	0.16	0.27

Panel B: Correlation Matrix for Wharton Index

	Wharton Index	PC 1	PC 2	Overall Index
Wharton Index	1.00	0.33	0.11	0.22
PC 1	0.33	1.00	0.07	0.36
PC 2	0.11	0.07	1.00	0.74
Overall Index	0.22	0.36	0.74	1.00

Panel C: Manual Validation on 30 California Municipalities

	Correct	Ambiguous	Incorrect
Minimum of Residential Min Lot Sizes	83%	6%	10%
Longest Frontage Requirement	80%	10%	10%

Notes: The sample overlap between our study and Gyourko et al. (2021) is 1,171 municipalities. We drop municipalities that do not have any minimum lot size requirements. The Affordable Housing questions refers only to affordable housing mandates, not incentives, and the minimum lot size questions refers only to residential districts. In Panel C we draw a random sample of 30 municipalities from California and have a law student review each answer. Answers that are ambiguous include situations where the bulk regulations depend on whether a lot is interior or corner and the model chose one of the cases that could be correct. The Overall Index in Panel B sums the z-scores of all zoning questions.

Table 5: Characteristics of Unincorporated Regions

Independent Variable:	Block Group is Unincorporated			
Median Home Value	-6.0***	-2.6***	-1.9***	-0.3
	(0.1)	(0.8)	(0.1)	(0.7)
Median Year Built	13.4***	7.2***	11.9***	5.6***
	(0.1)	(0.7)	(0.1)	(0.6)
Median Gross Rent	-98.2***	-81.1***	-8.8**	-37.4**
	(4.2)	(26.5)	(3.8)	(19.0)
Vacancy Rate	2.7***	2.2***	1.9***	1.6***
	(0.0)	(0.4)	(0.0)	(0.3)
Rental Rate	-12.6***	-15.1***	-9.9***	-12.5***
	(0.1)	(0.8)	(0.1)	(0.8)
Percent Commute Over 60	1.2***	2.9***	1.6***	2.8***
	(0.0)	(0.2)	(0.0)	(0.2)
Percent Over 65	2.0***	2.5***	1.3***	1.7***
	(0.1)	(0.3)	(0.1)	(0.3)
Specification:	Bivariate	Metro FE	Distance FE	Metro and Distance FE

Notes: This specification regresses an indicator on whether a block group is incorporated against a variety of block group characteristics with different levels of fixed effects. Block groups that are entirely unincorporated are assigned an indicator value of 1, while those that are entirely incorporated are assigned a value of 0. Partially incorporated block groups are split into both an incorporated and an unincorporated observation. We use regression weights that are the product of the block group population and the percent unincorporated for the unincorporated observations and the percent incorporated for incorporated observations. For each block group, we identify the nearest metropolitan center city hall, measure the distance from this center, and record the corresponding metro area. Distance fixed effects are defined as quintiles of the distance from the city center to account for potential non-linear effects of distance on the dependent variables. The coefficients displayed are for the indicator variable of whether a block group is unincorporated. The vacancy rate, rental rate, percent commute over 60 minutes, and percent aged 65 and over, are all expressed as percentages ranging from 0–100. For more information on how we measure whether an area is incorporated, see Appendix section 6. Asterisks denote significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are shown in parentheses. We use robust standard errors and cluster standard errors at the metro level for specifications that include metro fixed effects.

Table 6: National Sample Question Means

Panel A: Continuous Questions

Question	National			Income Tercile			Urban/Rural		
	Mean	Weight	Count	Low	Mid	High	Rural	Mix	Urban
How many zoning districts, including overlays, are in the municipality?	14	19	5,471	13	14	14	10	16	13
What is the longest frontage requirement for single-family residential development in any district?	92	69	5,213	74	86	117	93	97	79
Mean of Residential Min Lot Sizes	24639	17527	5,424	17071	21802	34965	31039	25746	15539
Min of Residential Min Lot Sizes	10104	5894	5,440	6974	8982	14305	12416	10057	7987
How many mandatory steps are involved in the approval process for a typical new multifamily building?	4.5	4.3	5,791	4.4	4.5	4.6	4.5	4.5	4.5
For a typical new multifamily building project in this jurisdiction, how many distinct governing bodies or agencies must give mandatory approval before construction can begin?	3.1	3.0	5,759	3.2	3.1	3.1	3.1	3.2	3.1
What is the maximum potential waiting time (in days) for government review of a typical new multifamily building?	218	211	5,109	195	223	236	200	222	226

Panel B: Binary Questions

Question	National			Income Tercile			Urban/Rural		
	Mean	Weight	Count	Low	Mid	High	Rural	Mix	Urban
Is multifamily housing allowed, either by right or special permit (including through overlays or cluster zoning)?	95	99	5,703	99	97	90	95	96	92
Are apartments above commercial properties (mixed-use) allowed in any district?	63	71	5,717	65	67	57	55	66	61
Is multifamily housing listed as allowed through conversion (of either single-family homes or nonresidential buildings)?	14	20	5,766	13	14	15	10	15	13
Are attached single-family houses (townhouses, 3+ units) listed as an allowed use (by right or special permit)?	80	89	5,795	80	82	79	65	84	83
Does the zoning include any provisions for housing that is restricted by age?	45	60	5,129	34	43	58	24	50	52
Are accessory or in-law apartments allowed (by right or special permit) in any district?	33	39	5,781	27	33	39	23	41	22
Is cluster development, planned unit development, open-space residential design, or another type of flexible zoning allowed by right?	9	10	5,797	8	8	10	5	10	8
Is cluster development, planned unit development, open-space residential design, or another type of flexible zoning allowed by special permit?	80	80	5,679	79	81	80	69	86	73
Does the zoning bylaw/ordinance include any mandates or incentives for development of affordable units?	24	50	5,540	10	22	41	9	28	27
Is there a town-wide annual or biannual cap on residential permits issued, and/or is project phasing required?	17	18	5,803	11	18	21	10	19	16
Are there restrictions on counting wetlands, sloped land or easements in lot size calculations?	10	7	4,617	4	9	17	7	12	7
Do developers have to comply with the requirement to include affordable housing, however defined, in their projects?	7	10	5,784	1	4	16	2	7	10
Are there town-wide requirements for public hearings on any type of multifamily residential projects?	30	32	5,709	23	31	37	27	30	32

Notes: This table reports the averages of sample questions from our generated national regulatory dataset across a range of demographic associates. We define the count (sample size) as the number of municipalities where the model does not say "I don't know" as the answer. The "Weight" column weights each municipality by its population in the 2022 census of governments. We designate Urban/Rural using the percent overlap of the 2022 shape file for the municipality with the 2020 shape file for urban areas. Specifically, we define Urban as a municipality being 100% in an urban area, Mix as a municipality being partially in an urban area, and Rural as a municipality being 0% in an urban area. We use median income from the 2021 Five-Year American Community Survey (`B19013_001E`). For continuous questions we upper winsorize at the 1% level for frontage, minimum lot sizes, and maximum potential review waiting time.

Table 7: Predicting Housing Market Outcomes With Zoning Regulation

Panel A: Individual Housing Regulations

Dependent Variable:	Median Home Value			Total Building Permits			Median Gross Rent			Housing Unit Density			Share Housing Units Affordable		
	XGBoost	LASSO	Bivariate	XGBoost	LASSO	Bivariate	XGBoost	LASSO	Bivariate	XGBoost	LASSO	Bivariate	XGBoost	LASSO	Bivariate
Minimum Res Min Lot Size	50	0.02	0.12***	32		0.01	66	0.03	0.08***	100	-0.13	-0.24***	100	-0.11	-0.17***
Multifamily Not Allowed	99	0.17	0.24***	18		0.00	100	0.14	0.16***	49	-0.04	-0.10***	73	-0.08	-0.12***
Affordable Mandate	62	0.01	0.05***	100		-0.00	79	0.06	0.09***	58	-0.01	-0.02	35	-0.01	-0.05***
Max Res Min Lot Size	51	0.01	0.09***	11		0.02	49	0.02	0.07***	96	-0.11	-0.21***	74	-0.06	-0.15***
Affordable Incentive	62		0.01	35		0.02	61	0.03	0.08***	59		-0.02	48	-0.03	-0.07***
Townhouses Not Allowed	100	0.02	0.12***	9		-0.03**	56	-0.02	-0.01	28	-0.04	-0.07***	48		0.01
Longest Frontage Requirement	47		0.06***	37		0.01	52	0.01	0.06***	50	-0.12	-0.21***	43	-0.04	-0.12***
No Conversion to Multifamily	49		0.02*	23		0.02	64		0.02	36	-0.03	-0.06***	48		-0.03***
Age-Restricted Provisions	67		-0.07***	18		0.01	58	0.00	0.02	28	-0.00	0.00	48		-0.02*
Zoning District Count	56		-0.04***	16	0.00	0.06***	58	0.06	0.06***	40	0.03	0.03***	48	-0.04	-0.06***
Max Review Waiting Time	30		-0.00	47		0.02	42		0.01	48	-0.01	-0.02	42		-0.02*
No Mixed-Use Buildings	59		0.07***	29		-0.02	45	0.03	0.04***	33	-0.06	-0.09***	41		-0.02*
Public Hearing Requirements	57		-0.04***	12		0.04**	56	-0.00	-0.01	24	0.00	0.00	45		0.01
Wetlands Restricted in Lot Size Calc	32		0.04***	17		0.01	44	0.02	0.04***	33	-0.06	-0.11***	53	-0.05	-0.10***
Mandatory Approval Steps	43		0.03**	19		0.03*	44	0.01	0.03**	29		-0.01	41	-0.00	-0.04***
Flexible Zoning By Permit	33		0.08***	12		-0.05***	37		-0.01	49	0.09	0.08***	44	0.03	0.07***
Permit Cap Or Phasing	19		-0.04***	26		0.03**	37		0.01	33	-0.03	-0.05***	49		-0.04***
Distinct Approval Bodies	37		0.02	17		0.01	29	0.01	0.03**	36	-0.01	-0.04***	43		-0.02
Accessory Apartments Banned	22		-0.03***	21		-0.05***	31	-0.02	-0.05***	27	0.06	0.08***	23	0.08	0.12***
No Flexible Zoning By RIght	15		-0.00	13		0.00	35		-0.01	19	0.01	0.04***	38	0.02	0.06***

Panel B: Housing Regulations Indices

Dependent Variable:	Median Home Value			Total Building Permits			Median Gross Rent			Housing Unit Density			Share Housing Units Affordable		
	XGBoost	LASSO	Bivariate	XGBoost	LASSO	Bivariate	XGBoost	LASSO	Bivariate	XGBoost	LASSO	Bivariate	XGBoost	LASSO	Bivariate
Overall Index	82	-0.05	-0.01	73	0.02	0.03**	91	-0.06	0.00	100		0.01	91	0.08	0.00
First PC	91		0.02	92		-0.02	96	-0.02	0.00	80		0.02	91	0.02	-0.00
Second PC	100	0.05	0.01	100		0.02	100	0.07	0.03*	76		0.01	100	-0.10	-0.03**

Notes: This specification reports the results of three sets of regressions: a random forest specification, LASSO, and bivariate regressions of our generated regulatory variables against four dependent variables. Median house value and median gross rent are drawn from the 2022 ACS. Total building permits are defined as the number of housing units permitted divided by the population of the local government, averaged over 2019–2023, from the Census Building Permits Survey. Housing unit density is the number of housing units in a local government in the 2022 ACS divided by the area from its shape file. The share of affordable housing units is defined as the percentage of housing units affordable to someone earning the state median income. Rental units are considered affordable if the monthly rent does not exceed 30% of the monthly median household income, and owner-occupied units are affordable if their value is less than three times the annual median household income. For bivariate regressions, stars indicate statistical significance: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. LASSO coefficients are shown where selected, with blank cells indicating variables not retained in the model. XGBoost scores represent "gain" importance from a random forests specification, measuring the average gain of splits using each feature, normalized so the most important feature for each dependent variable has a score of 100, with others scaled relatively. Higher scores indicate greater importance in the model's predictions. All variables are first demeaned at the MSA level, or for municipalities not within an MSA at the state level, then transformed into z-scores (mean=0, std=1). We use imputed regulations from our PCA analysis for LASSO and Bivariate regressions when the LLM reports "I don't know," see footnote of Appendix Table A2 for further details. We express variables so that a more positive value is associated with stricter zoning regulations, i.e. we transform the question of whether multi-family housing is allowed to whether it is not allowed. We allow missing data for XGBoost letting the algorithm both impute and predict. The Overall index in Panel B is a sum of normalized individual housing regulations for a municipality.

Table 8: Comparative Model Performance in Predicting Housing Outcomes

LHS Variable	Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Median Home Value	Lasso	1.07	1.11	0.73	1.05	0.71	0.73	0.71
	OLS	1.06	1.11	0.72	1.05	0.70	0.72	0.70
	XGBoost	1.00	1.12	0.74	0.96	0.66	0.77	0.67
Total Building Permits	Lasso	1.07	1.07	1.06	1.07	1.07	1.07	1.07
	OLS	1.06	1.06	1.06	1.06	1.06	1.06	1.06
	XGBoost	1.32	1.31	1.31	1.32	1.31	1.31	1.31
Median Gross Rent	Lasso	1.08	1.10	0.82	1.07	0.81	0.82	0.81
	OLS	1.08	1.10	0.81	1.07	0.81	0.81	0.81
	XGBoost	1.06	1.08	0.80	1.04	0.76	0.80	0.77
Housing Unit Density	Lasso	1.08	1.16	1.13	1.08	1.07	1.13	1.07
	OLS	1.08	1.16	1.13	1.08	1.07	1.13	1.07
	XGBoost	0.98	1.13	1.10	0.98	0.98	1.08	0.98
Share Housing Units Affordable	Lasso	0.99	1.02	0.70	0.97	0.69	0.70	0.69
	OLS	0.99	1.02	0.69	0.97	0.68	0.69	0.68
	XGBoost	0.99	1.03	0.63	0.95	0.53	0.62	0.52
Variables Included	Regulations	Yes	No	No	Yes	Yes	No	Yes
	Land	No	Yes	No	Yes	No	Yes	Yes
	Income	No	No	Yes	No	Yes	Yes	Yes

Notes: These specification report the RMSE of models predicting the housing market outcomes in Table 7. We include housing regulation measures in columns 1, 4, 5, and 7. We also include land availability controls (the share of land that is flat plains) in columns 2, 4, 6, and 7. We include median household income from the 2022 ACS as a control in specifications 3, 5, 6, and 7. All variables are first demeaned at the MSA level, or when not available at the state level, then transformed into z-scores (mean=0, std=1). We use imputed regulations when the LLM reports "I don't know" to avoid different imputation values across different controls, see footnote of Appendix Table A2 for further details.

Table 9: Housing Regulation Intensity and Distance from Metropolitan Centers

Independent Variable	US Census Region				All Regions
	West	South	Midwest	Northeast	
Accessory Apartments Allowed	0.02	0.01	0.00	0.10***	0.05**
	(0.04)	(0.02)	(0.03)	(0.03)	(0.02)
Flexible Zoning By Right	-0.03	-0.03	-0.01	0.04*	0.01
	(0.05)	(0.03)	(0.03)	(0.02)	(0.02)
Flexible Zoning By Permit	0.02	0.04	0.00	0.07*	0.05*
	(0.03)	(0.04)	(0.05)	(0.04)	(0.03)
Affordable Incentive	0.01	-0.14**	-0.15***	0.01	-0.02
	(0.03)	(0.06)	(0.04)	(0.03)	(0.02)
Affordable Mandate	0.01	0.01	-0.08	0.01	0.01
	(0.02)	(0.07)	(0.06)	(0.01)	(0.01)
Zoning District Count	-0.06	-0.12***	-0.08	-0.06	-0.08***
	(0.04)	(0.04)	(0.05)	(0.06)	(0.03)
Permit Cap Or Phasing	0.01	-0.00	0.01	0.03*	0.02
	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)
Wetlands Restricted in Lot Size Calc	-0.01	-0.01	0.03	0.07***	0.05**
	(0.05)	(0.04)	(0.05)	(0.02)	(0.02)
Longest Frontage Requirement	0.05	-0.02	0.09**	0.12***	0.09***
	(0.05)	(0.04)	(0.04)	(0.03)	(0.02)
Maximum Res Min Lot Size	0.02	0.01	0.09***	0.13***	0.08***
	(0.02)	(0.04)	(0.03)	(0.01)	(0.02)
Mean Res Min Lot Size	0.03	0.02	0.10***	0.17***	0.12***
	(0.04)	(0.06)	(0.02)	(0.03)	(0.02)
Minimum Res Min Lot Size	0.05	0.02	0.14***	0.10***	0.10***
	(0.07)	(0.04)	(0.03)	(0.02)	(0.02)
Mandatory Approval Steps	-0.04*	-0.01	0.01	0.02	0.00
	(0.02)	(0.02)	(0.03)	(0.03)	(0.01)
Distinct Approval Bodies	0.05**	0.01	0.04	0.02	0.02**
	(0.03)	(0.03)	(0.04)	(0.02)	(0.01)
Public Hearing Requirements	0.05*	-0.03	0.03*	0.06***	0.04***
	(0.02)	(0.04)	(0.01)	(0.02)	(0.01)
Max Review Waiting Time	0.01	-0.02	-0.03	0.04**	0.00
	(0.02)	(0.03)	(0.03)	(0.02)	(0.01)
Multifamily Allowed	0.09**	0.04	0.04	-0.02	0.01
	(0.05)	(0.03)	(0.03)	(0.01)	(0.01)
Mixed-Use Buildings	-0.01	-0.06	-0.04**	-0.02	-0.03*
	(0.04)	(0.04)	(0.02)	(0.03)	(0.02)
Conversion To Multifamily	-0.05	-0.06	-0.00	-0.02	-0.02*
	(0.04)	(0.05)	(0.03)	(0.01)	(0.01)
Townhouses Allowed	0.02	0.02	-0.04	-0.05	-0.03
	(0.04)	(0.07)	(0.04)	(0.04)	(0.02)
Age-Restricted Provisions	-0.02	-0.04	-0.07***	-0.01	-0.03
	(0.03)	(0.04)	(0.02)	(0.04)	(0.02)
First Principal Component (Regulatory Complexity)	0.01	-0.11*	-0.05	0.07*	0.03
	(0.03)	(0.06)	(0.07)	(0.04)	(0.03)
Second Principal Component (Strictness)	0.02	0.02	0.11***	0.14***	0.11***
	(0.07)	(0.05)	(0.03)	(0.03)	(0.02)
MSA Fixed Effects	Yes	Yes	Yes	Yes	Yes

Notes: We perform this regression on the set of municipalities within 50 miles of the center of a metropolitan area which accounts for 3,605 observations in our sample. The dependent variable is log distance to metro center. A positive coefficient indicates that the variable increases with log distance from the metro center and a negative coefficient means that the variable decreases with log distance from the metro center. See Appendix Table A1 for full definitions of zoning questions. Asterisks denote significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are shown in parentheses. We cluster standard errors at the metro level.

Table 10: Socioeconomic and Geographic Correlates of Regulatory Complexity (PC1)

Dependent Variable:	First Principal Component (Regulatory Complexity)			
Auto Commute Share	-0.03	-0.18***	-0.03	-0.01
	(0.06)	(0.02)	(0.03)	(0.03)
Foreign Born Share	0.03	0.16***	0.06**	0.01
	(0.04)	(0.02)	(0.02)	(0.03)
Median Household Income	0.04	0.23***	-0.08	-0.01
	(0.06)	(0.02)	(0.06)	(0.05)
Share Population 65 and Over	-0.09***	-0.07***	-0.00	-0.02
	(0.02)	(0.01)	(0.02)	(0.03)
Median Gross Rent	0.09	0.29***	-0.01	-0.07*
	(0.06)	(0.02)	(0.04)	(0.04)
Median Home Value	-0.06	0.18***	0.04	-0.15*
	(0.09)	(0.02)	(0.04)	(0.08)
Share Units Owner Occupied	-0.05	0.02	-0.13***	-0.08**
	(0.04)	(0.01)	(0.03)	(0.03)
Share Population Under 18	-0.00	-0.07***	-0.01	-0.01
	(0.02)	(0.01)	(0.02)	(0.02)
White Share	-0.01	-0.03***	0.14***	0.16***
	(0.03)	(0.01)	(0.03)	(0.04)
Poverty Rate	-0.08***	-0.21***	-0.08***	-0.07**
	(0.03)	(0.01)	(0.03)	(0.03)
College Share	0.14***	0.27***	-0.05	0.09*
	(0.04)	(0.02)	(0.04)	(0.06)
Share Structures Built Before 1970	-0.28***	-0.19***	-0.06***	-0.10***
	(0.04)	(0.01)	(0.02)	(0.03)
Share Structures with 2 or More Units	0.09**	0.13***	0.05	0.02
	(0.04)	(0.01)	(0.03)	(0.03)
Vacancy Rate	-0.13***	-0.15***	-0.06**	-0.02
	(0.04)	(0.01)	(0.03)	(0.04)
Share with Commute Over 30 Minutes	-0.08***	0.13***	0.04**	0.00
	(0.03)	(0.01)	(0.02)	(0.03)
Job Density	0.04	0.12***	-0.01	-0.01
	(0.05)	(0.02)	(0.02)	(0.02)
Opportunity Index	-0.01	0.08***	-0.06**	-0.05**
	(0.03)	(0.01)	(0.03)	(0.02)
Average Math Test Scores	0.15***	0.23***	-0.04	0.03
	(0.03)	(0.01)	(0.03)	(0.08)
Math Learning Rate	0.08***	0.10***	0.00	0.03*
	(0.03)	(0.01)	(0.02)	(0.02)
Percent Eligible for Free Lunch	-0.15***	-0.25***	-0.16***	-0.07
	(0.03)	(0.01)	(0.04)	(0.05)
Local Revenue Per Student	0.00	0.12***	-0.04*	0.00
	(0.09)	(0.03)	(0.02)	(0.02)
Property Tax Rate	-0.03	0.07***	0.03	0.03
	(0.02)	(0.01)	(0.02)	(0.02)
Total Expenditure Per Capita (2017)	0.04	0.12***	0.02	0.02
	(0.04)	(0.02)	(0.02)	(0.02)
Building Permits All Units 2021	0.06***	0.09***	0.02	0.03
	(0.02)	(0.02)	(0.01)	(0.02)
Year of Incorporation	-0.09***	0.01	0.00	-0.01
	(0.02)	(0.01)	(0.02)	(0.02)
Percent Democrat	0.14***	0.27***	0.22***	0.17***
	(0.04)	(0.01)	(0.03)	(0.04)
Log Land Area	0.46***	0.47***	0.36***	0.35***
	(0.03)	(0.01)	(0.02)	(0.03)
Log Neighbors within 25 Miles	0.22***	0.15***	-0.02	0.13**
	(0.07)	(0.01)	(0.02)	(0.05)
Housing Unit Density	-0.08**	0.01	0.01	0.00
	(0.03)	(0.01)	(0.02)	(0.02)
Log Miles to Metro Center	0.01	-0.14***	0.03*	0.03
	(0.05)	(0.01)	(0.02)	(0.03)
Share Units Affordable	-0.16***	-0.33***	-0.15***	0.01
	(0.04)	(0.01)	(0.04)	(0.05)
Intercept			0.05	-0.77***
			(0.08)	(0.22)
Controls:	MSA FE	None	None	MSA FE
Specification:	Bivariate	Bivariate	All Included	All Included
R-squared			0.33	0.42
N			2712	2712

Notes: This specification regresses our first principal component (regulatory complexity) against a range of socioeconomic and geographical covariates. For variable definitions, see Appendix Table A3. All right-hand side variables are measured as Z-scores to enable comparability. The first two columns regress each variable in a bivariate regression; the last two columns include all variables. Columns 1 and 4 include MSA fixed effects. Fixed effects are for MSAs with State FE for municipalities not within an MSA. Asterisks denote significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are shown in parentheses. We cluster standard errors at the MSA level when including MSA fixed effects and use robust standard errors otherwise.

Table 11: Socioeconomic and Geographic Correlates of Exclusionary Zoning (PC2)

Dependent Variable:	Second Principal Component (Exclusionary Zoning)			
Auto Commute Share	-0.01	-0.16***	-0.02	0.02
	(0.03)	(0.01)	(0.03)	(0.03)
Foreign Born Share	-0.20***	-0.04***	0.06***	0.01
	(0.04)	(0.01)	(0.02)	(0.03)
Median Household Income	0.30***	0.39***	0.20***	0.13***
	(0.03)	(0.01)	(0.06)	(0.05)
Share Population 65 and Over	0.14***	0.16***	0.03	0.03
	(0.02)	(0.01)	(0.02)	(0.02)
Median Gross Rent	0.14***	0.22***	0.04	0.05
	(0.04)	(0.01)	(0.04)	(0.04)
Median Home Value	0.24***	0.28***	0.16***	0.20**
	(0.04)	(0.02)	(0.04)	(0.09)
Share Units Owner Occupied	0.33***	0.41***	0.04	-0.02
	(0.03)	(0.01)	(0.03)	(0.04)
Share Population Under 18	-0.02*	-0.11***	-0.04*	-0.05**
	(0.01)	(0.01)	(0.02)	(0.02)
White Share	0.22***	0.22***	0.02	-0.04
	(0.04)	(0.01)	(0.02)	(0.03)
Poverty Rate	-0.14***	-0.26***	0.09***	0.07***
	(0.03)	(0.01)	(0.03)	(0.03)
College Share	0.12***	0.27***	-0.02	0.03
	(0.04)	(0.01)	(0.03)	(0.04)
Share Structures Built Before 1970	-0.09**	0.06***	0.01	-0.04
	(0.04)	(0.01)	(0.02)	(0.03)
Share Structures with 2 or More Units	-0.35***	-0.32***	-0.09***	-0.16***
	(0.03)	(0.01)	(0.03)	(0.04)
Vacancy Rate	0.06***	0.04***	-0.01	0.03
	(0.01)	(0.01)	(0.02)	(0.03)
Share with Commute Over 30 Minutes	0.12***	0.24***	0.02	0.02
	(0.03)	(0.01)	(0.02)	(0.03)
Job Density	-0.21***	-0.15***	-0.06***	-0.04**
	(0.03)	(0.02)	(0.02)	(0.02)
Opportunity Index	0.14***	0.23***	-0.01	-0.03
	(0.03)	(0.01)	(0.02)	(0.03)
Average Math Test Scores	0.14***	0.29***	-0.01	0.02
	(0.03)	(0.01)	(0.03)	(0.04)
Math Learning Rate	0.07***	0.11***	-0.03*	-0.00
	(0.02)	(0.01)	(0.01)	(0.02)
Percent Eligible for Free Lunch	-0.18***	-0.33***	0.03	0.08
	(0.03)	(0.01)	(0.03)	(0.05)
Local Revenue Per Student	0.16**	0.28***	0.10**	0.06***
	(0.08)	(0.06)	(0.04)	(0.01)
Property Tax Rate	-0.14***	0.07***	0.05***	0.01
	(0.03)	(0.01)	(0.02)	(0.02)
Total Expenditure Per Capita (2017)	-0.07***	0.11***	-0.01	-0.03**
	(0.02)	(0.02)	(0.02)	(0.02)
Building Permits All Units 2021	-0.00	-0.03*	-0.02	-0.02
	(0.01)	(0.01)	(0.02)	(0.02)
Year of Incorporation	0.13***	0.10***	-0.00	0.02
	(0.03)	(0.01)	(0.01)	(0.02)
Percent Democrat	-0.22***	0.02	-0.10***	-0.14***
	(0.04)	(0.01)	(0.02)	(0.04)
Log Land Area	-0.04	0.07***	-0.08***	-0.08***
	(0.06)	(0.01)	(0.02)	(0.02)
Log Neighbors within 25 Miles	-0.10***	0.19***	0.05***	-0.02
	(0.04)	(0.01)	(0.02)	(0.03)
Housing Unit Density	-0.24***	-0.17***	-0.05	-0.04
	(0.05)	(0.03)	(0.03)	(0.04)
Log Miles to Metro Center	0.15***	-0.03**	0.01	-0.03
	(0.04)	(0.01)	(0.02)	(0.02)
Share Units Affordable	-0.20***	-0.29***	0.08**	0.06
	(0.05)	(0.01)	(0.04)	(0.05)
Intercept			-0.41***	-0.44***
			(0.07)	(0.12)
Controls:	MSA FE	None	None	MSA FE
Specification:	Bivariate	Bivariate	All Included	All Included
R-squared			0.26	0.35
N			2712	2712

Notes: This specification regresses our second principal component (exclusionary zoning) against a range of socioeconomic and geographical covariates. For variable definitions, see Appendix Table A3. All right-hand side variables are measured as Z-scores to enable comparability. The first two columns regress each variable in a bivariate regression; the last two columns include all variables. Columns 1 and 4 include MSA fixed effects. Fixed effects are for MSAs with State FE for municipalities not within an MSA. Asterisks denote significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are shown in parentheses. We cluster standard errors at the MSA level when including MSA fixed effects and use robust standard errors otherwise.

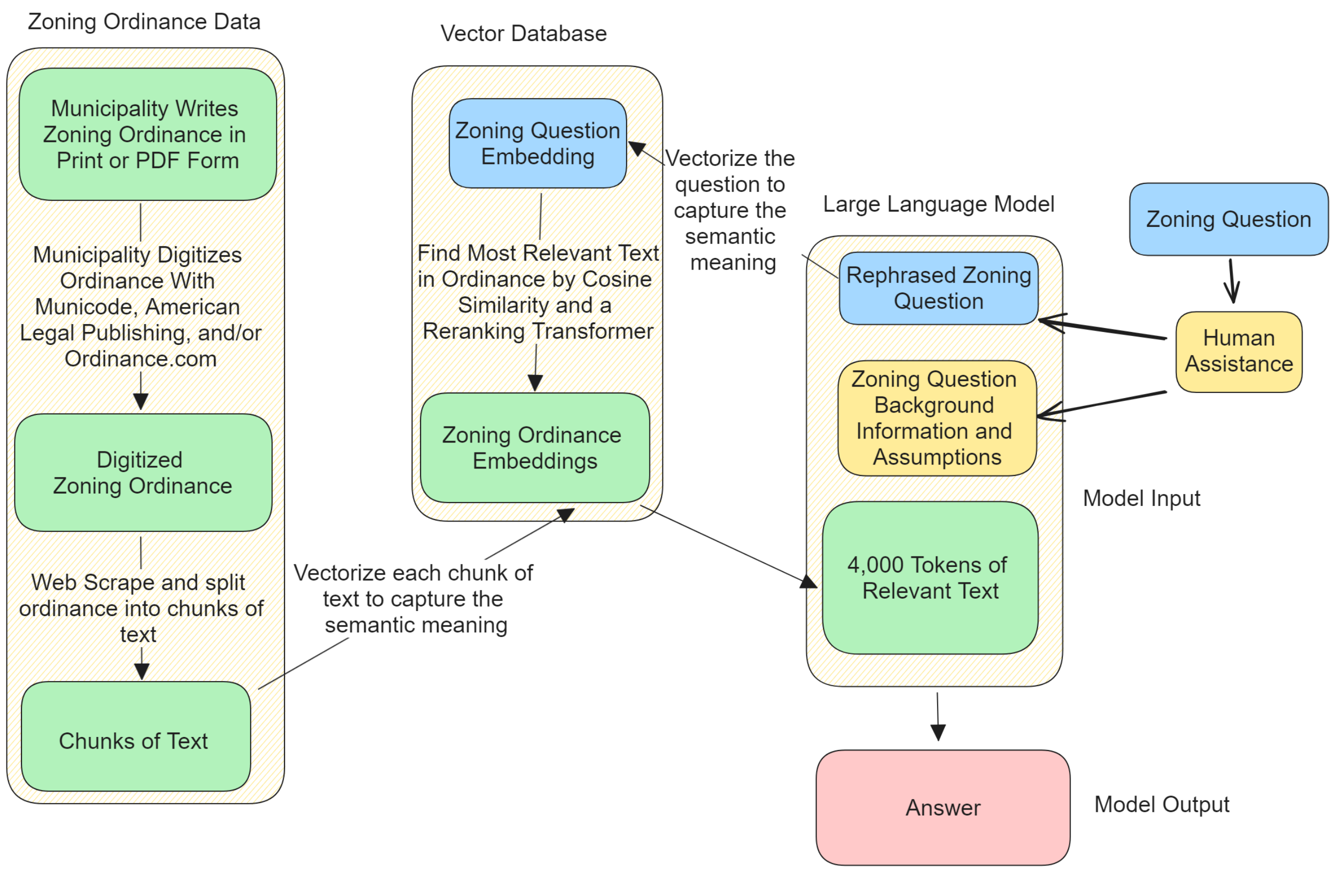
Table 12: Housing Regulation and Housing Supply Elasticity

Dependent Variable:	First Principal Component		Second Principal Component	
New Housing Unit Elasticity	-0.43***	-0.12*	-0.74***	-0.03
	(0.05)	(0.07)	(0.05)	(0.17)
Share Land Developed (2001)	-0.87***	-0.17	-2.24***	-0.80**
	(0.11)	(0.20)	(0.12)	(0.36)
Squared Share Land Developed (2001)	0.38***	-0.12	1.26***	0.36*
	(0.08)	(0.15)	(0.08)	(0.20)
Share Land Flat Plains	0.01	0.02	0.19***	0.09
	(0.03)	(0.04)	(0.03)	(0.09)
Log Miles to Metro Center	-0.10***	-0.07*	-0.15***	-0.10**
	(0.02)	(0.04)	(0.02)	(0.04)
Intercept	0.14***	-0.30***	0.06***	-0.84***
	(0.02)	(0.07)	(0.02)	(0.12)
R-squared	0.06	0.21	0.17	0.41
N	3890	3890	3890	3890
MSA Fixed Effects	No	Yes	No	Yes

Notes: This specification has as the dependent variable the first regulatory principal component (regulatory complexity, first two columns) and the second regulatory principal component (exclusionary zoning, second two columns). We regress these variables against a range of variables relating to new housing production and land availability. All variables are normalized to z-scores for the regression. Housing elasticity controls follow Baum-Snow and Han (2024) and include fraction of land developed in 2001, squared fraction of land developed in 2001, and the fraction of land with a flat topography. MSA fixed effects include state fixed effects for municipalities not in an MSA. Asterisks denote significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are shown in parentheses. Standard errors are clustered at the MSA level when using MSA fixed effects and are robust otherwise.

Figures

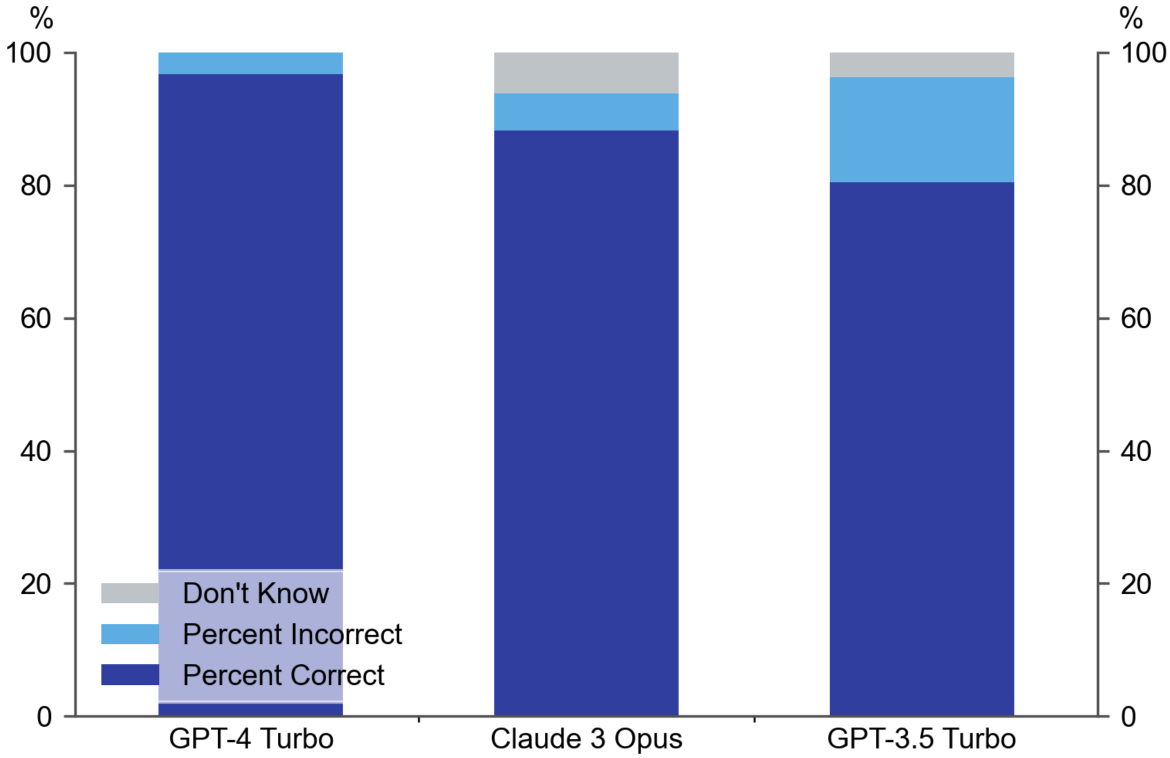
Figure 1: Model Overview



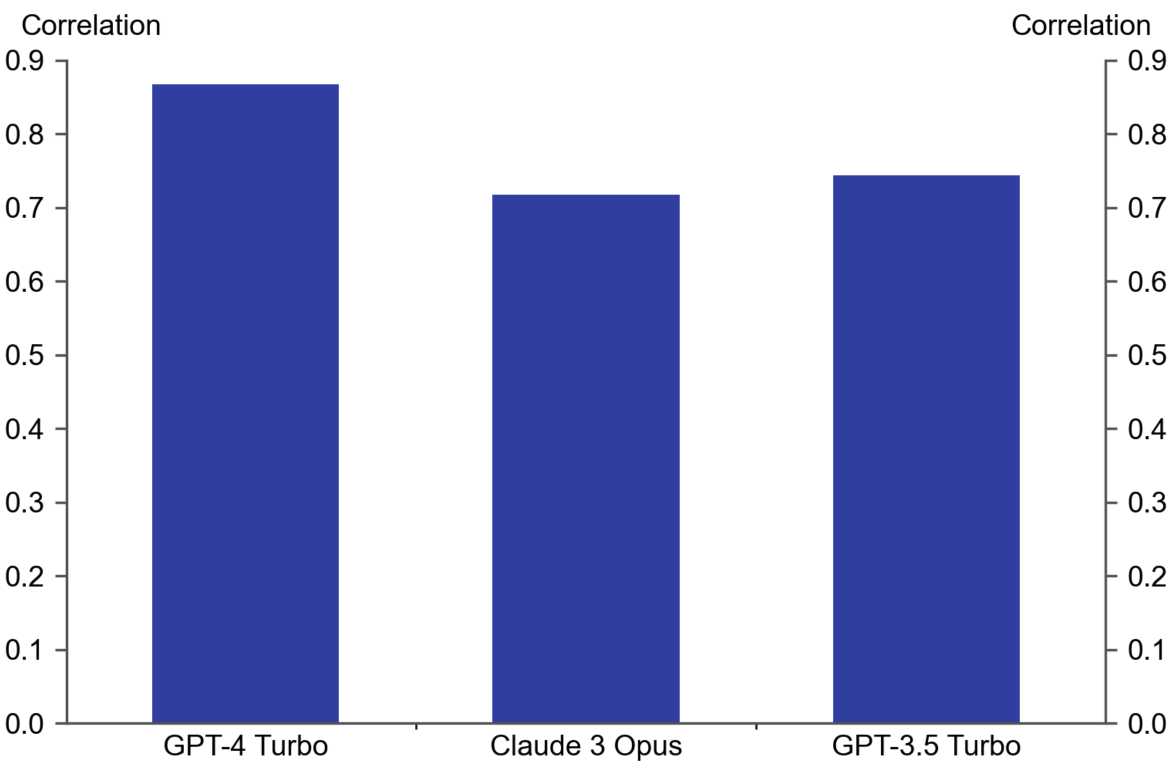
Notes: This Figure illustrates our overall process for Generative Regulatory Measurement. We scrape each section within a zoning ordinance separately. We split up sections that are longer than one thousand tokens into chunks of at most one thousand tokens. We also combine adjacent sections of less than 50 tokens. So, each section of text varies in length but is between 50 and one thousand tokens. We vectorize each chunk of text using OpenAI embeddings models (link). Specifically, we use the "text-embedding-3-large" algorithm. Sometimes digital aggregators leave tables in image form, especially the aggregator Ordinance.com. So that the model can still read the table, we transcribe images of tables using Amazon Textract. We elicit an open-ended response to each question and then use function calling to parse out a structured answer (i.e., to ascertain whether an answer is "Yes," "No," or "I don't know" to a binary question). Question background information and model assumptions are based on a combination of the "Issue Overview" and the "Research Coding" sections for each question from the Pioneer study as well as from trial and error in the training sample of municipalities. Rephrased zoning questions came entirely from trial and error on the training sample. Ordinances from digital aggregators (Municode, American Legal Publishing, and Ordinance.com) are either entirely about zoning, partially about zoning (i.e., have one or more sections about zoning), or not about zoning at all. We filter out ordinances not at all about zoning by searching for key phrases, table headers, and zoning district names (i.e., R-1 for the first residential zoning district). See Appendix Section 6 for further details on question background information and assumptions.

Figure 2: Comparison of Average Performance Across Models

Panel A: Binary Questions

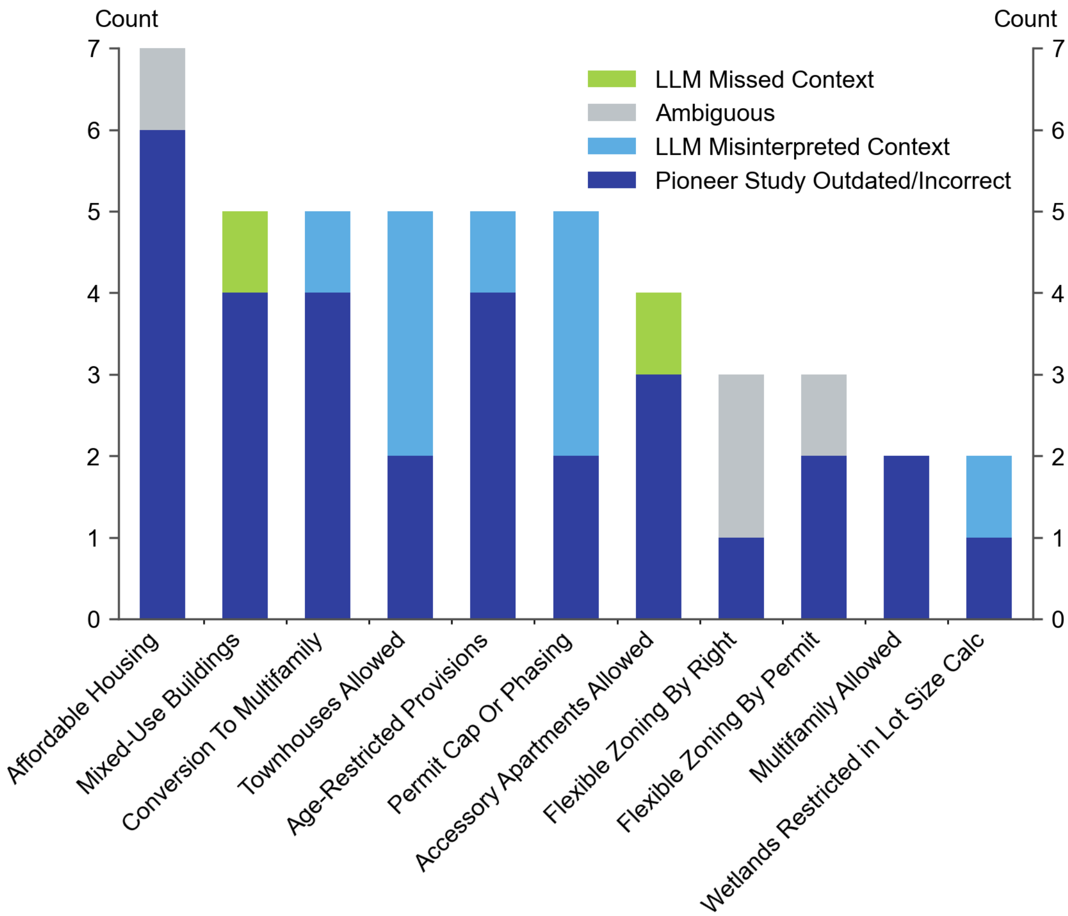


Panel B: Continuous Questions



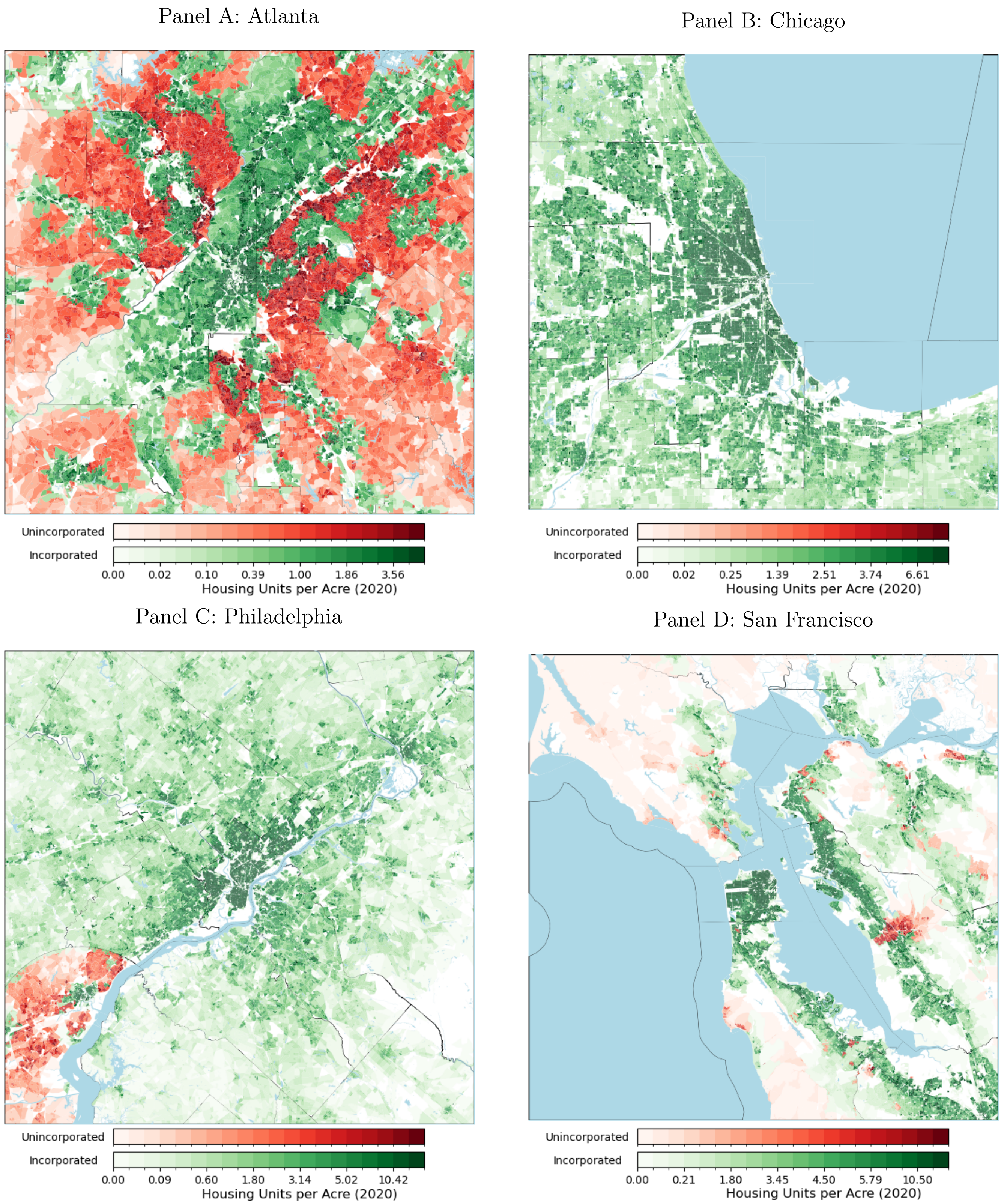
Notes: This Figure reports model accuracy measures across different LLM models. For binary questions we report the percent accurate and for continuous questions we use the correlation. We use adjusted data and drop ambiguous question-muni pairs, see Figure 3 for more details. We include the minimum residential min lot size, the maximum frontage, and the number of districts for continuous questions.

Figure 3: Sources of Discrepancy Between LLM and Human Classifications of Zoning Regulations



Notes: This Figure reports reasons for disagreement between our generated model data from GPT-4 Turbo and the original version of the Pioneer Institute data for binary questions. We first ran GPT-4 Turbo on the testing sample of 30 randomly selected municipalities that were included in the Pioneer Institute's study but were not used to train our model. We then identified the binary questions where the model responses disagreed with the Pioneer study. A law student reviewed each of these disagreements individually to determine the reason for the discrepancy, classifying them into the categories shown in the chart. When measuring the performance of the model we adjust for disagreements where the Pioneer study was outdated/incorrect and also drop ambiguous cases.

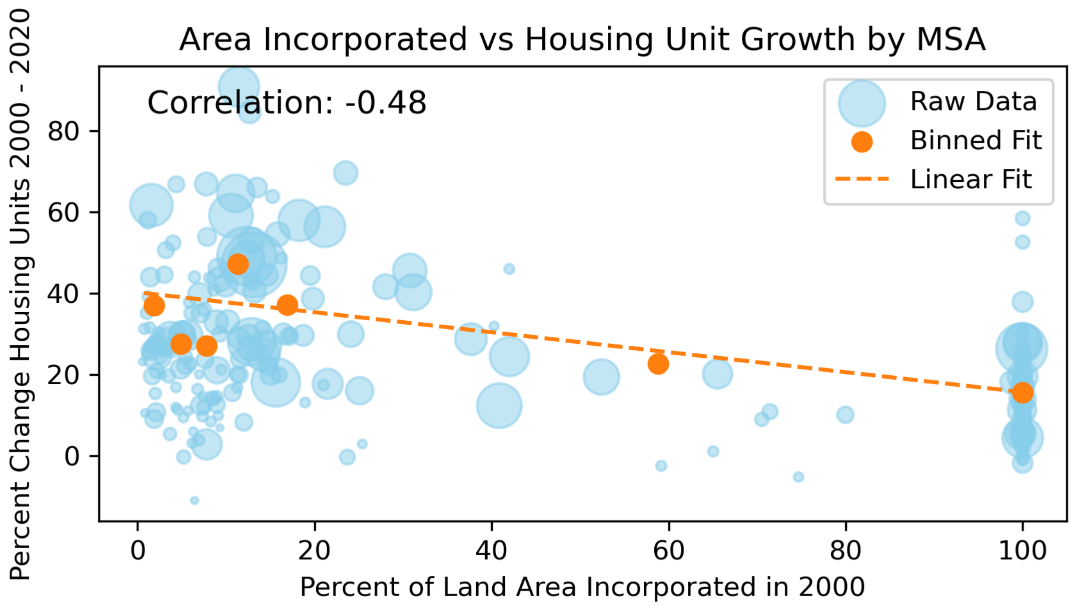
Figure 4: Unincorporated vs. Incorporated Land in Select Metropolitan Areas



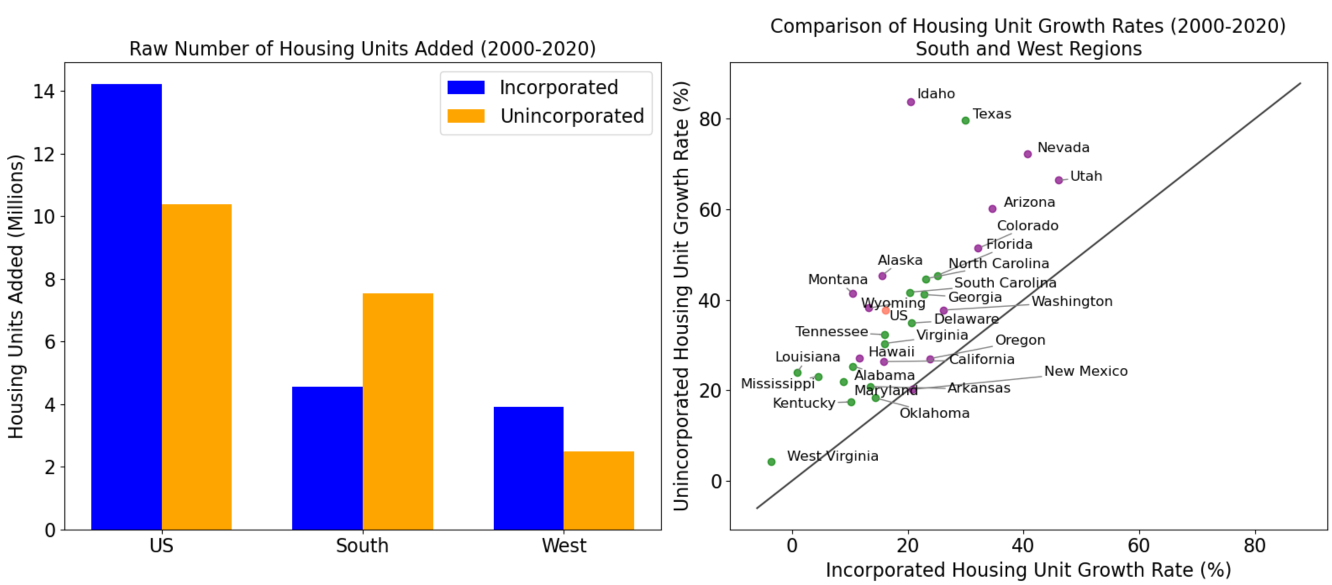
Notes: We plot housing unit density in four metropolitan regions: Atlanta (Panel A), Chicago (Panel B), Philadelphia (Panel C), and the San Francisco Bay Area (Panel D). Incorporated areas are colored in green, while unincorporated areas are colored in red. We plot housing unit density at the census block level drawn from 2022 ACS data and categorize blocks as incorporated if any of their area overlaps with an incorporated area and unincorporated otherwise. See Appendix Section 6 for further details about measuring incorporated areas.Each map shows roughly a 100km × 100km square area.

Figure 5: Housing Growth in Unincorporated Areas

Panel A: Growth in Housing Units From 2000–2020, by Fraction Incorporated



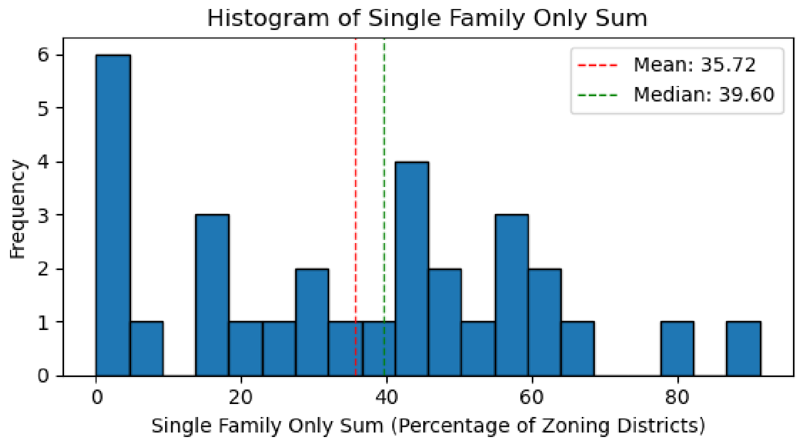
Panel B: Growth in Housing Units From 2000–2020, Incorporated vs Unincorporated



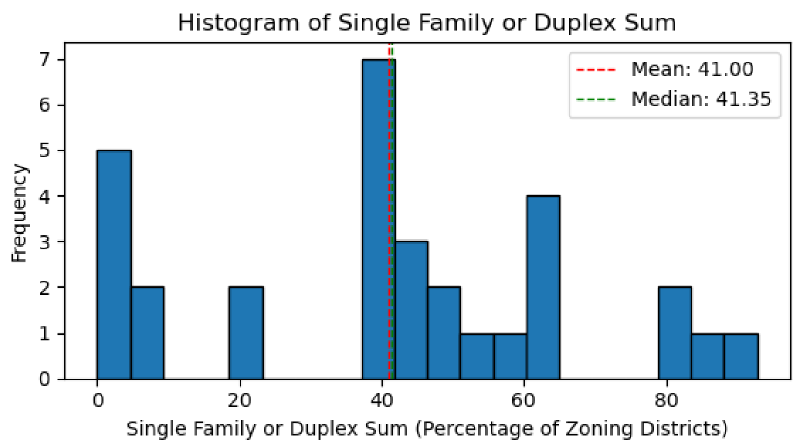
Notes: Panel A: For each MSA we calculate the percent of land area (county shape file ex water shape file) that intersects with the incorporated area in 2000 to calculate the percent of land area incorporated in 2000. Panel B: For each census block in both 2000 and 2020 we calculate the percent of its area that was incorporated in 2000. We then count the number of housing units that are incorporated by summing the product of housing units and percent incorporated for each block. We count the number of number of housing units that are unincorproated by summing the product of housing units and 1 - percent incorporated for each block. We focus on the South and West regions in this panel since practically all of the Northeast and Midwest are incorporated. See Appendix Section 6 for further details on how we measure incorporation.

Figure 6: Allowable Zoning Typologies in American Cities

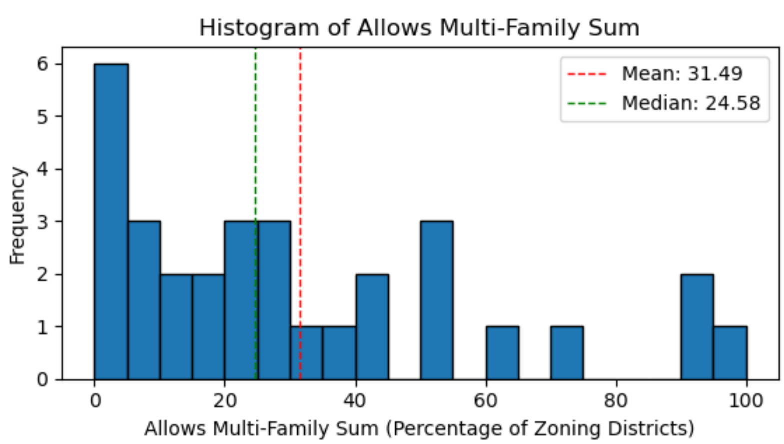
Panel A: Single-Family Only



Panel B: Single-Family or Duplex

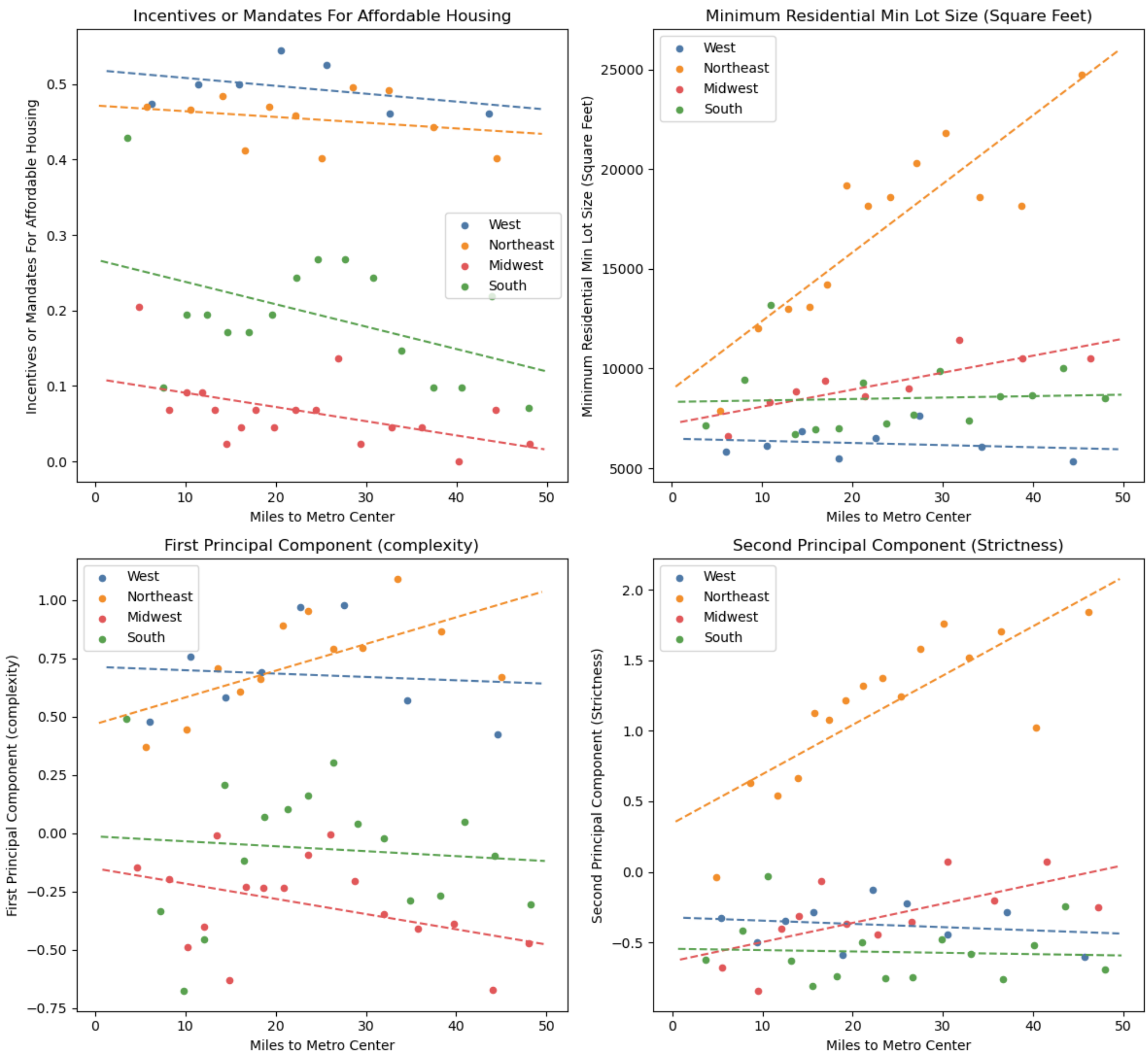


Panel C: Multi-Family



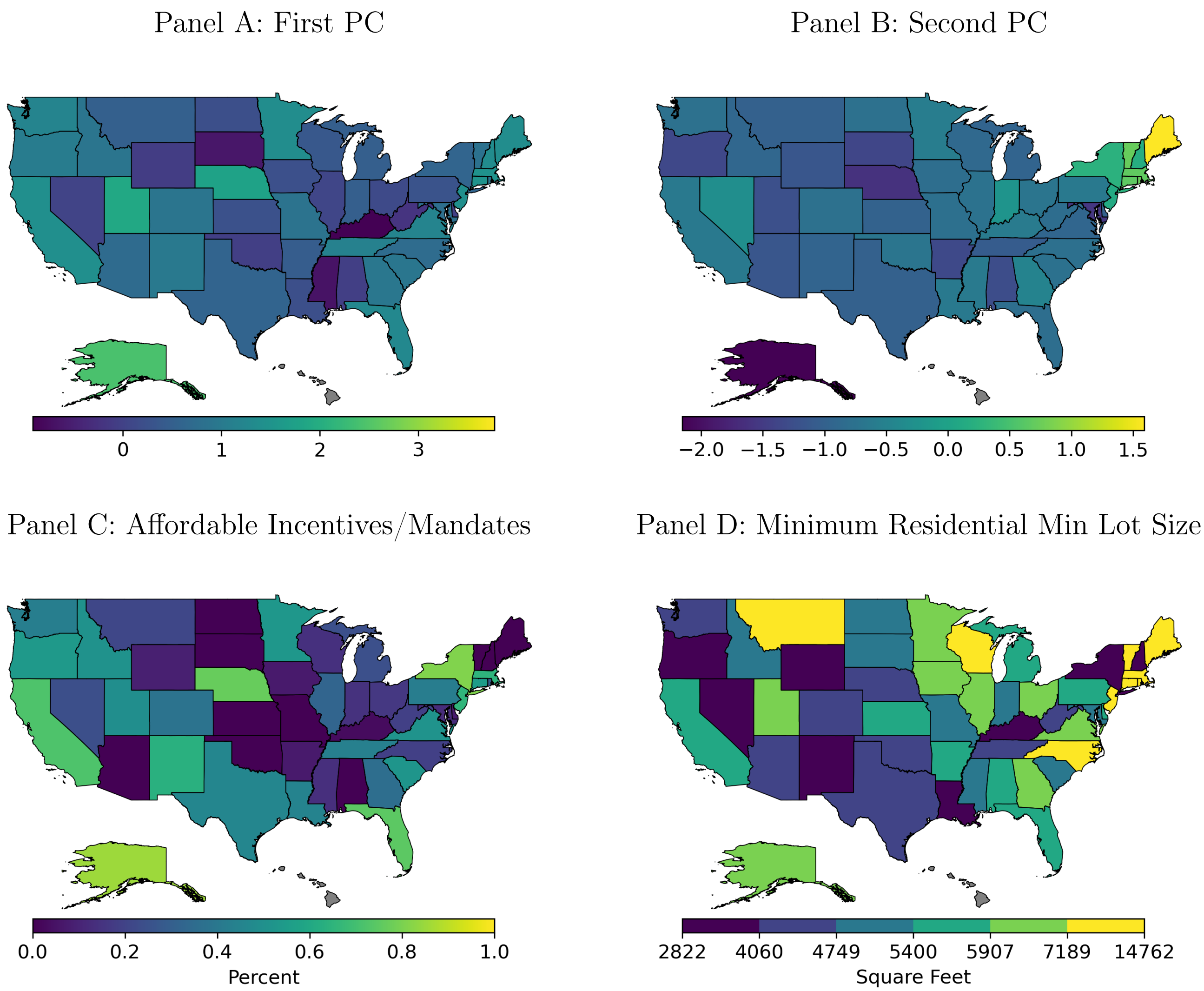
Notes: We show the distribution of overall land area zoned for three different uses: single-family only, single-family or duplex, and multi-family. We show results for 31 municipalities with 18.2 million in population, covering Chicago, Seattle, Kansas City, Detroit, San Francisco, Austin, San Antonio, Tampa, Los Angeles, and San Diego.

Figure 7: Spatial Variation of Zoning Regulations Relative to Metropolitan Centers



Notes: Note: We plot regulatory variables at the municipality level based on the distance from the center of the city for the 73 largest MSAs. We define city hall as the center of the city. We show whether a city has an affordable housing incentive or mandate; the minimum lot size; the first principal component of housing regulation (housing complexity), and the second principal component of housing regulations (exclusionary zoning).

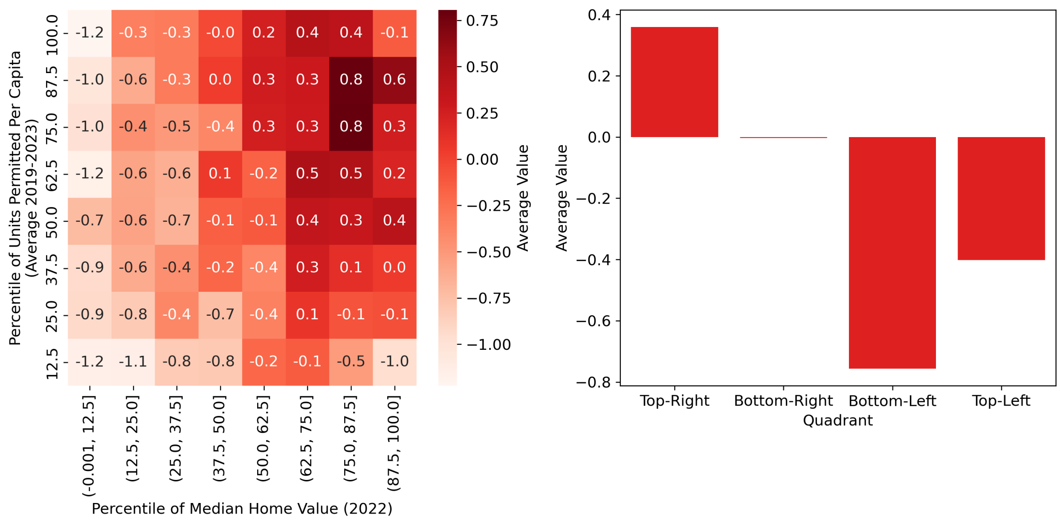
Figure 8: Nationwide Maps of Population-Weighted Averages



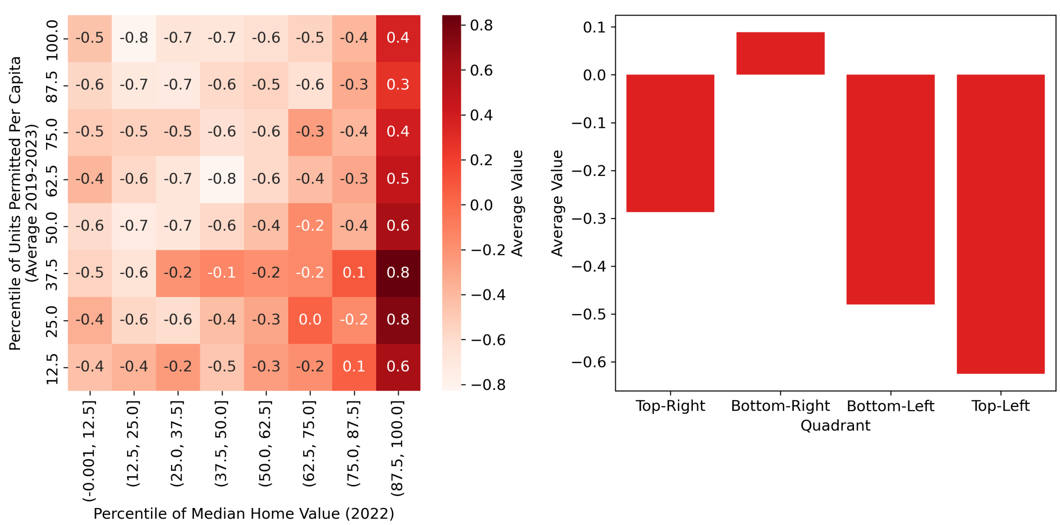
Notes: This Figure plots regulatory variables by state, weighted by local municipality population. We use the 2022 ACS Population as the population weight. Hawaii is grey because only one municipality (Honolulu) is in the dataset. For unweighted state maps please see Appendix Figure A10 and for county level maps please see Appendix Figure A11.

Figure 9: Regulatory Dimensions and Housing Market Dynamics

Panel A: First Principal Component (Regulatory Complexity)

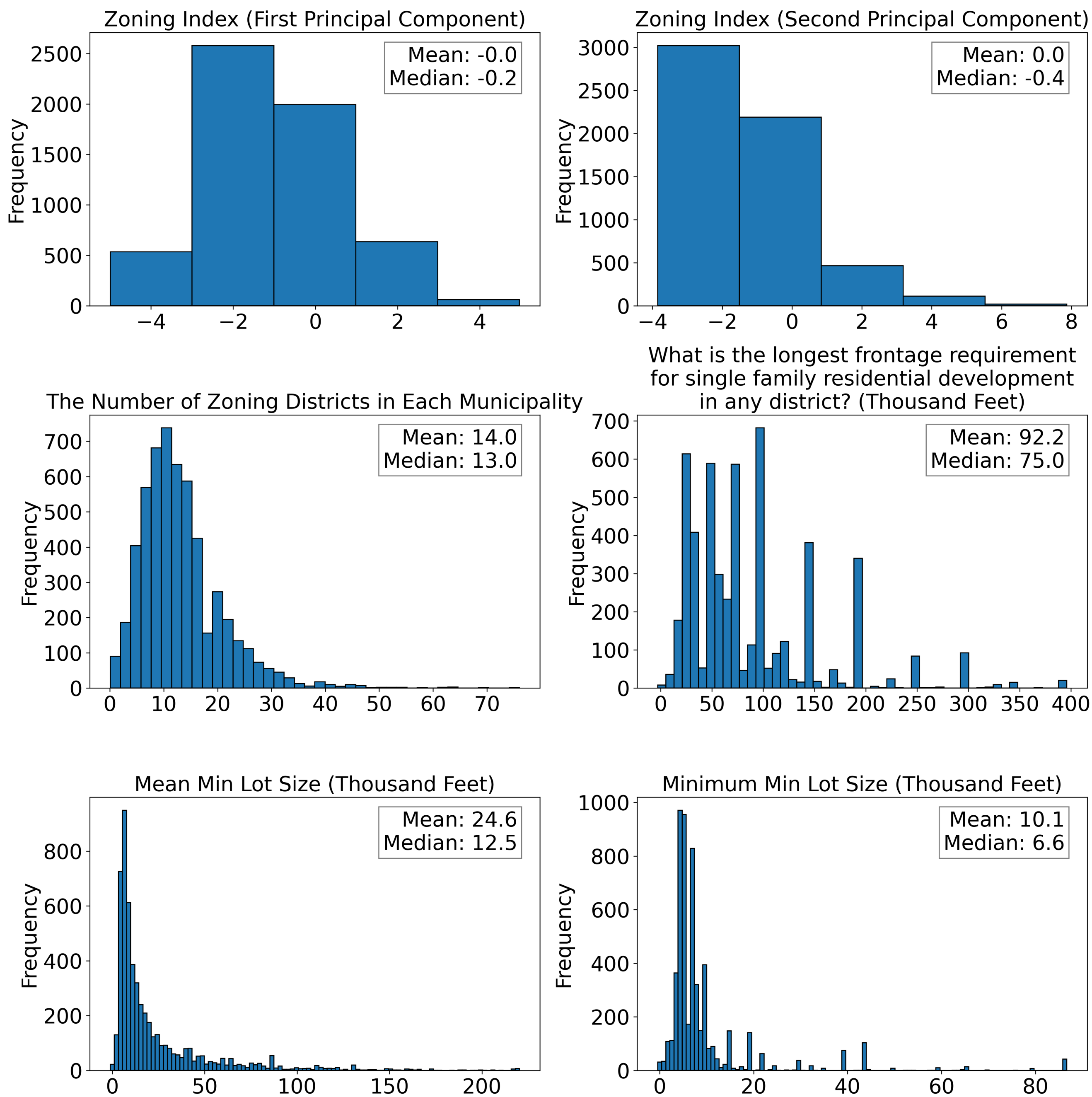


Panel B: Second Principal Component (Exclusionary Zoning)



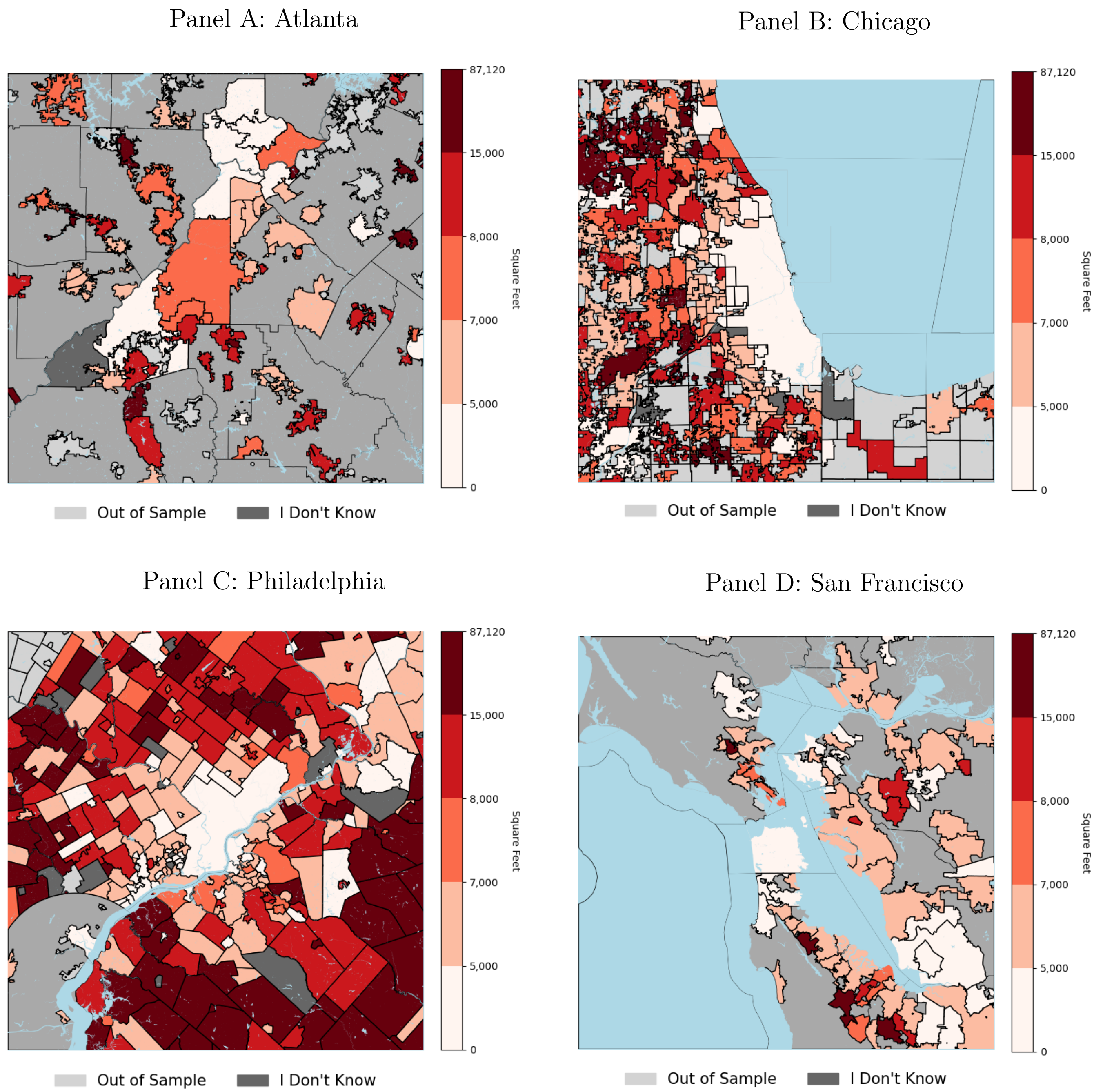
Notes: This Figure illustrates the relationship between two principal components of housing regulation and key housing market outcomes. Panel A shows the first principal component, associated with regulatory complexity. Panel B displays the second principal component, which corresponds to exclusionary zoning practices. The x-axis in both panels represents the percentile of median housing value, drawn from the 2022 ACS, while the y-axis measures the percentile of units permitted per capita (averaged from 2019–2023). Darker colors illustrate a larger correlation in the heatmap between each regulatory principal component and each coordinate of prices and building. The right figures show the overall correlation between each quadrant of the space of housing prices and quantities and the regulatory outcome. Please see Table A3 for details on variable definitions.

Figure 10: Distribution of Zoning Indices and Housing Regulations



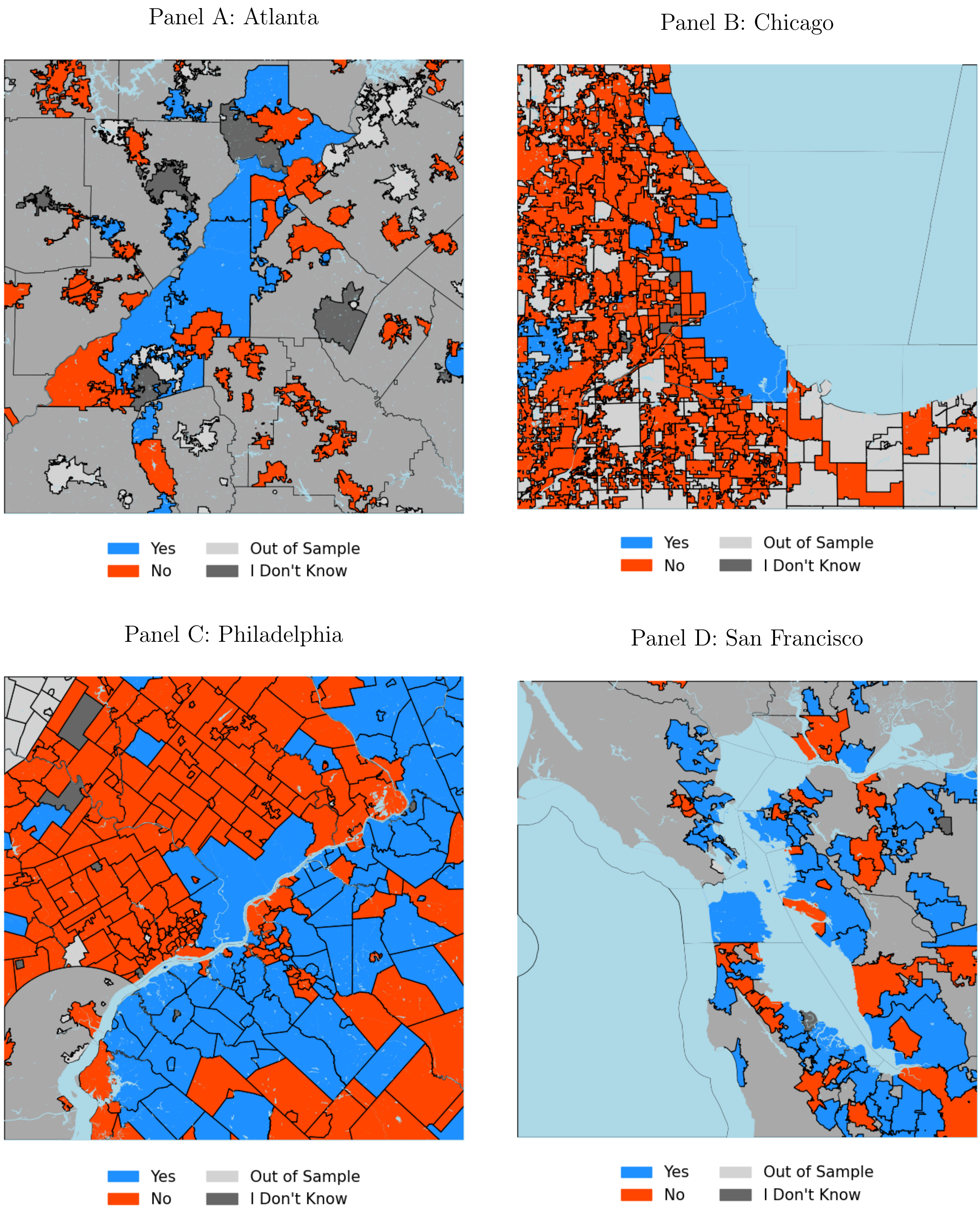
Notes: This Figure shows a histogram of the frequency of different regulatory measures at the municipality level. The first two panels show the distribution of the first two principal components. The third panel shows the distribution of the number of zoning districts. The fourth panel shows the longest frontage requirement. The fifth panel shows the average minimum lot size, and the sixth panel shows the minimum of all minimum lot size requirements. The last three variables are plotted in thousands of feet. See Table 6 footnote for details on the sample.

Figure 11: Minimum Residential Min Lot Sizes For Select Metro Areas



Notes: This Figure plots the minimum of all minimum lot size requirements for Atlanta (Panel A), Chicago (Panel B), Philadelphia (Panel C), and the San Francisco Bay Area (Panel D). The amount of affordable housing mandates are shown in red, with higher minimum requirements plotted as darker colors. Within each map we plot all Census-designated places. Both Census-designated place and Census county subdivisions data comes from the 2022 Census TIGER/Line shape files. Each map shows roughly a 100km × 100km square area. Non-Incorporated areas are shown in light grey, and areas for which the model reported "I don't know" are shown in dark gray.

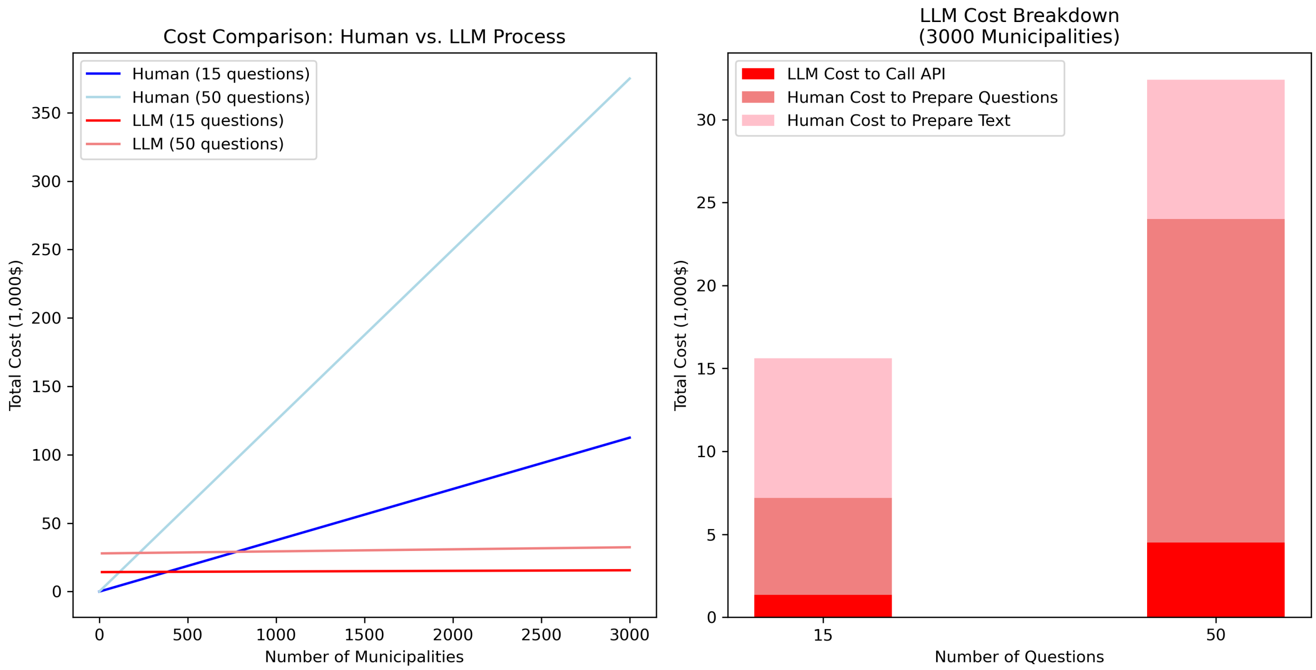
Figure 12: Affordable Housing Incentives or Mandates For Select Metro Areas



Notes: This Figure plots the presence of affordable housing incentives or mandates for Atlanta (Panel A), Chicago (Panel B), Philadelphia (Panel C), and the San Francisco Bay Area (Panel D). Areas with incentives or mandates are plotted in blue; areas without such mandates are shown in red. Within each map we plot all Census-designated places. Both Census-designated place and Census county subdivisions data comes from the 2022 Census TIGER/Line shape files. Each map shows roughly a 100km × 100km square area. Non-Incorporated areas are shown in light grey, and areas for which the model reported "I don't know" are shown in dark gray.

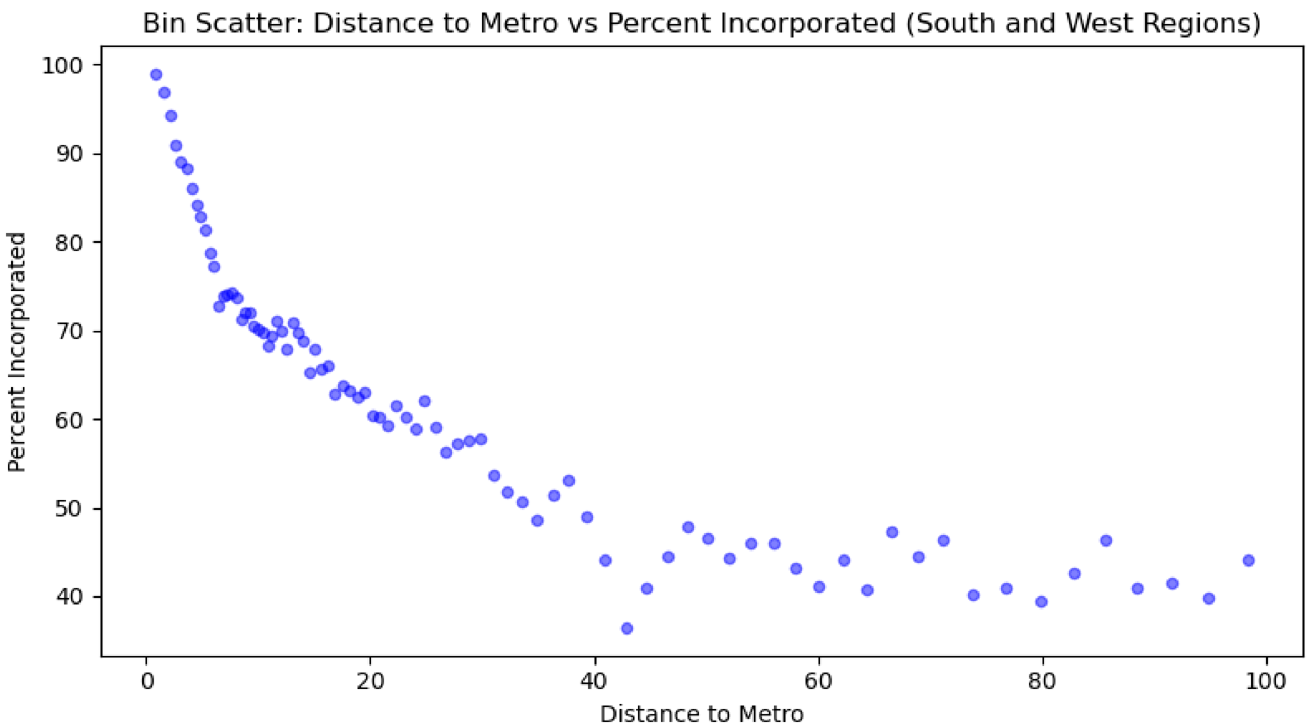
Appendix A: Additional Graphs and Tables

Figure A1: Comparison of Human vs. LLM Cost



Notes: : This figure compares the cost of human versus LLM-based analysis of zoning regulations across different numbers of municipalities. We assume: (1) Lawyers take an average of five minutes per question-municipality pair, based on observed rates for law students; (2) Lawyers are compensated at \$50/hour; (3) LLM API calls cost \$0.03 per question-municipality pair, using GPT-4 Turbo pricing for 4k token input with asynchronous batching; (4) Question preparation costs \$390/question, including lawyer time for answering 60 municipalities (30 for training, 30 for testing) and 10 hours of human labor at \$140 to train the model; (5) Scraping, cleaning, and embedding each municipality's ordinance costs \$2.80. All costs are assumed to scale linearly with the number of municipalities and questions. We contrast estimates assuming either 15 or 50 total questions

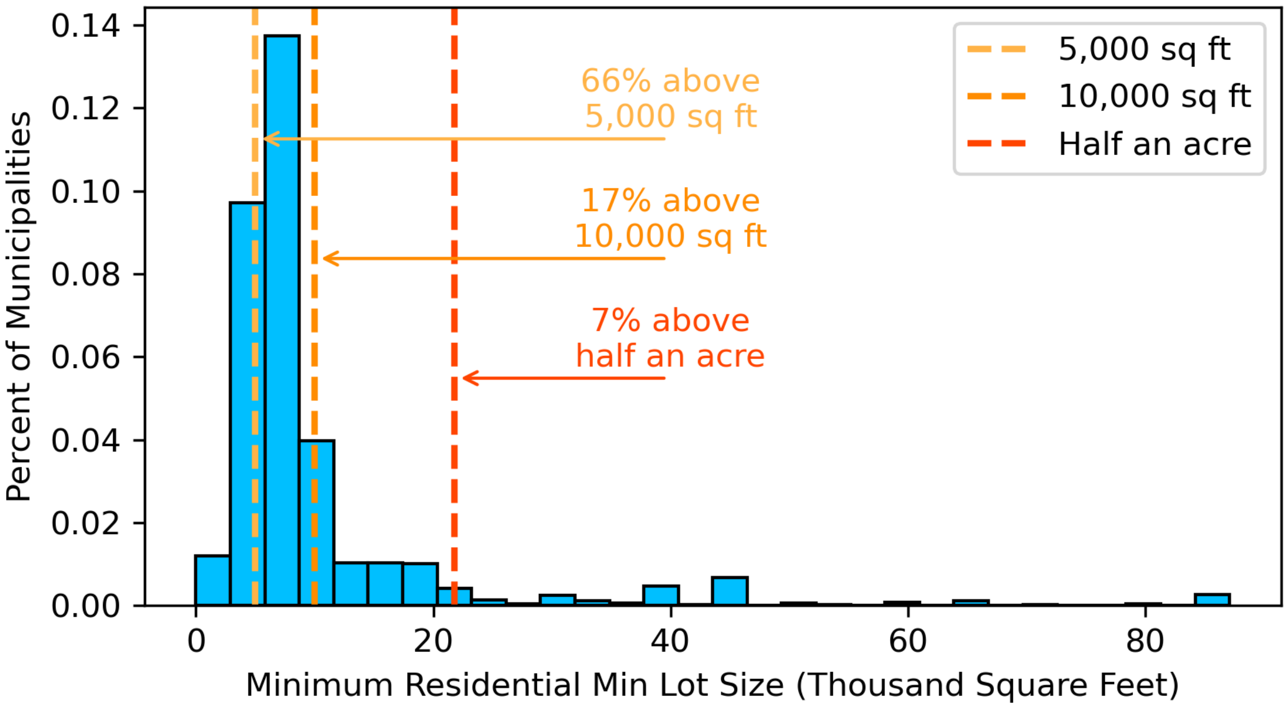
Figure A2: Municipal Incorporation Status Relative to Metropolitan Core Distance



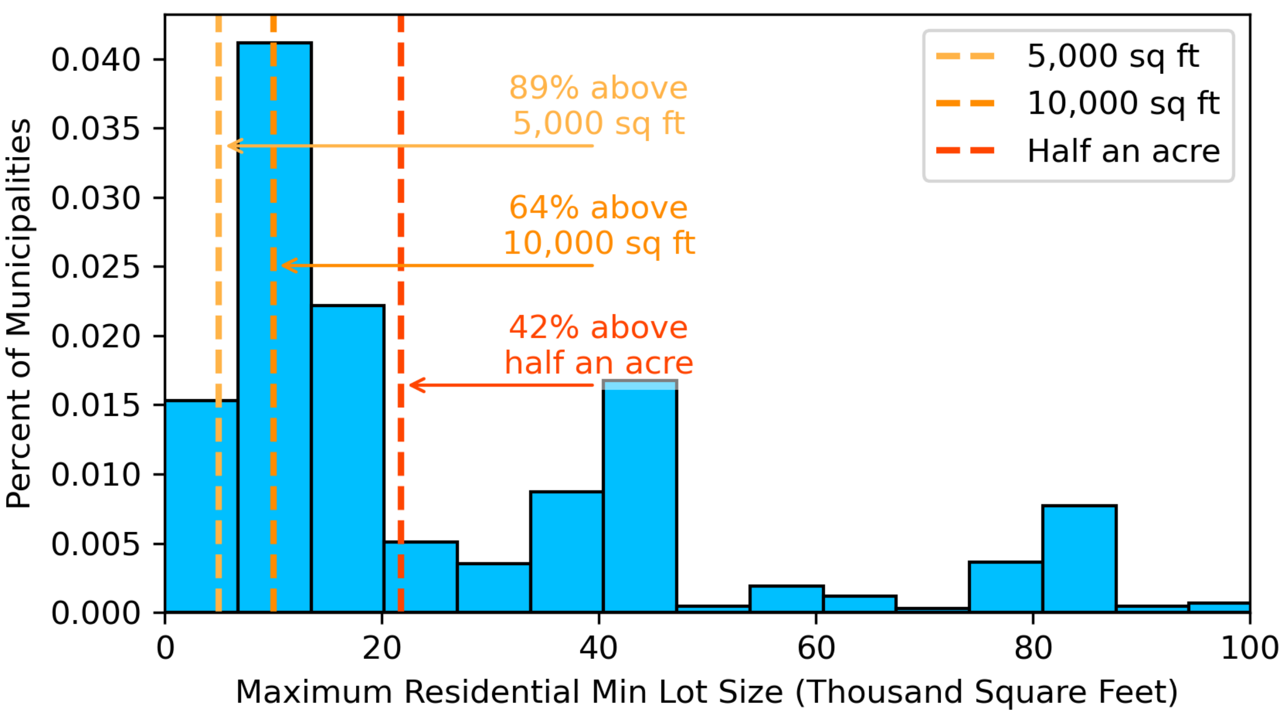
Notes: This Figure illustrates the relationship between municipal incorporation status and distance from metropolitan centers. The x-axis represents the log distance from the nearest metropolitan center, measured in miles, which is defined as City Hall. The y-axis shows the probability of a block group being incorporated, ranging from 0 to 1. Each point represents a binned average, with the blue line indicating the fitted relationship. For details on incorporation measurement methodology, see Appendix Section 6.

Figure A3: Minimum Lot Size Distribution

Panel A: Minimum Residential Min Lot Size

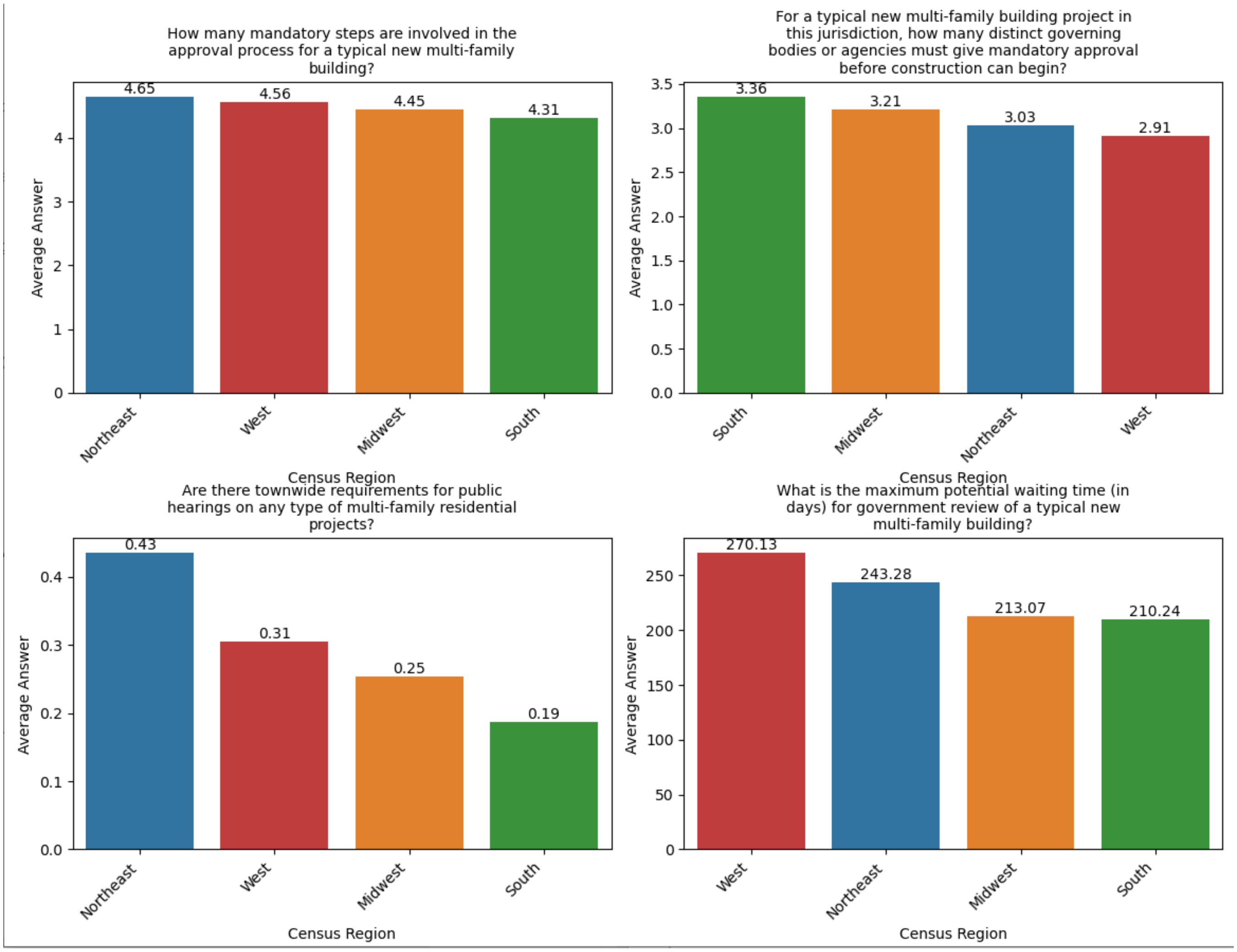


Panel B: Maximum Residential Min Lot Size



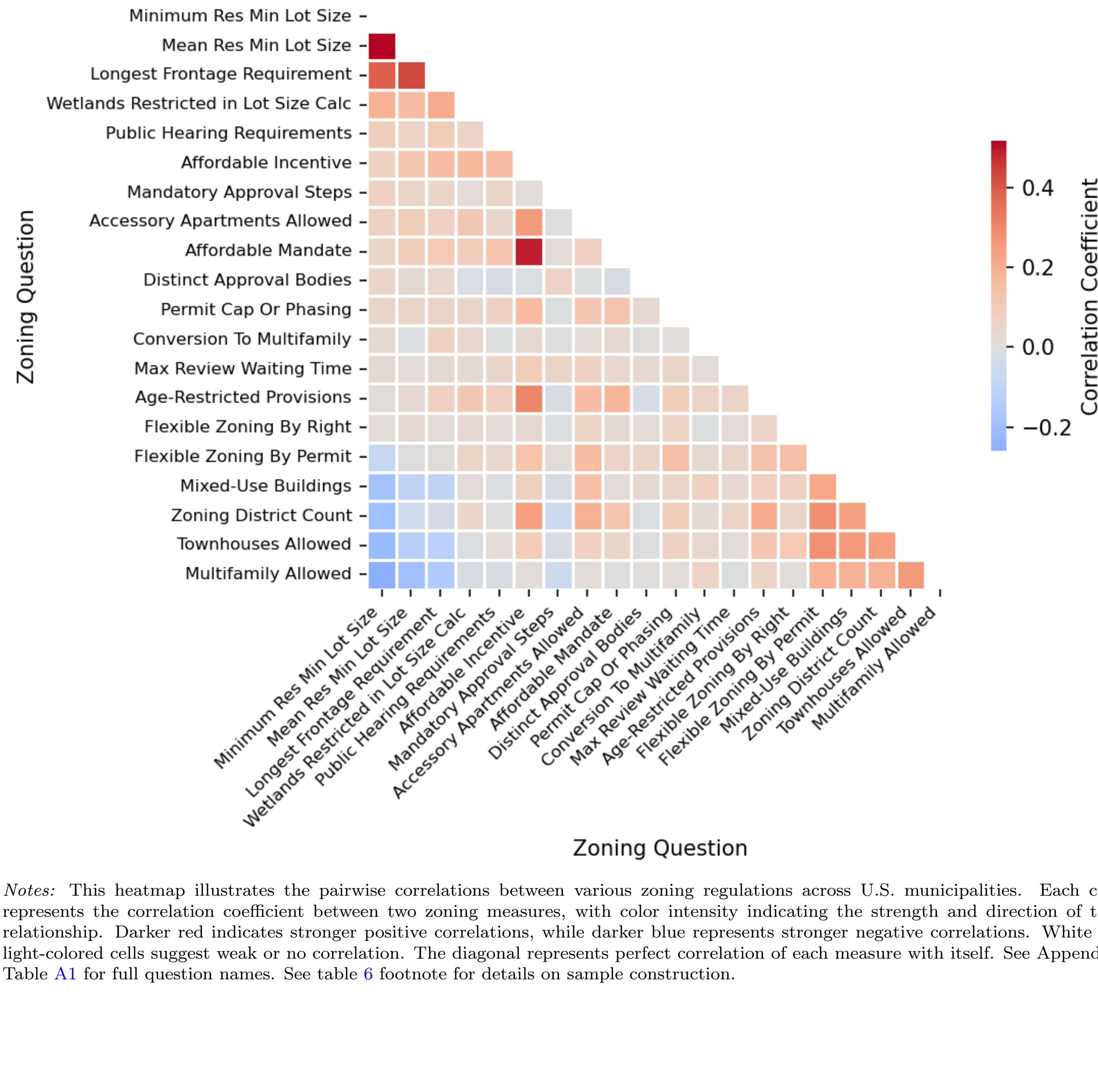
Notes: This histogram shows the distribution of minimum lot size requirements across municipalities. The x-axis represents lot size in square feet, with key thresholds labeled. The y-axis shows the percentage of municipalities falling into each lot size category. Vertical lines mark important thresholds: 5,000 sq ft (common suburban lot size), 10,000 sq ft (quarter-acre), and 21,780 sq ft (half-acre). The x-axis in Panel B stops at 100 thousand square feet, though 12% of municipalities have maximum residential min lot sizes above this level.

Figure A4: Measuring Housing Process Variation



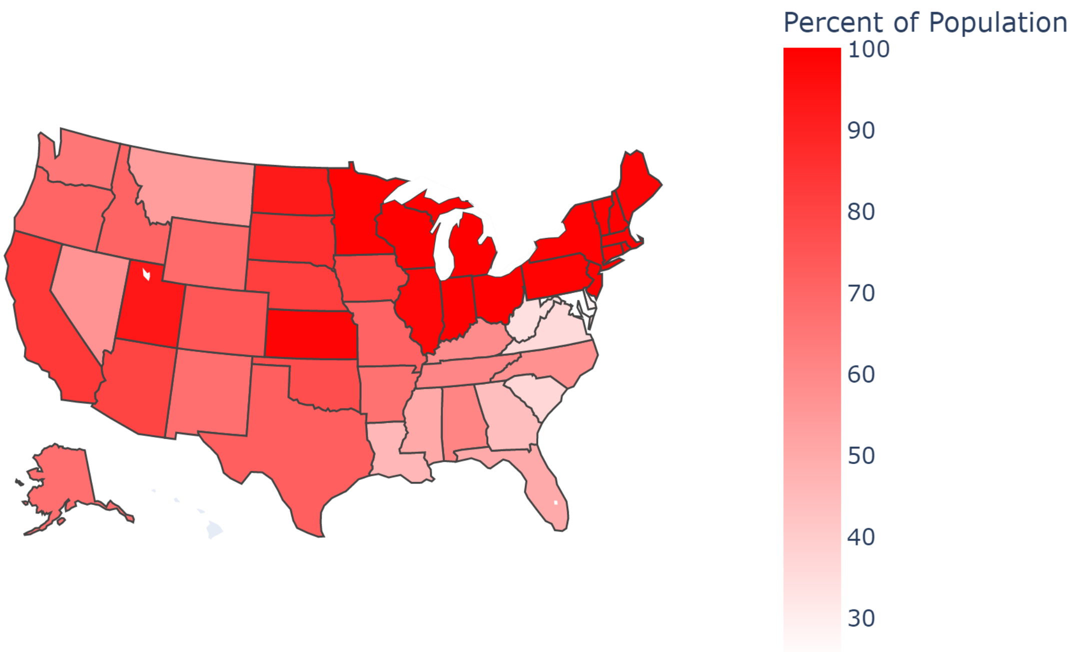
Notes:. This histogram shows the distribution of four housing process questions which augmented the initial Pioneer Institute set of questions. Survey questions are grouped into Census Region.

Figure A5: Pairwise Correlations Between Zoning Questions



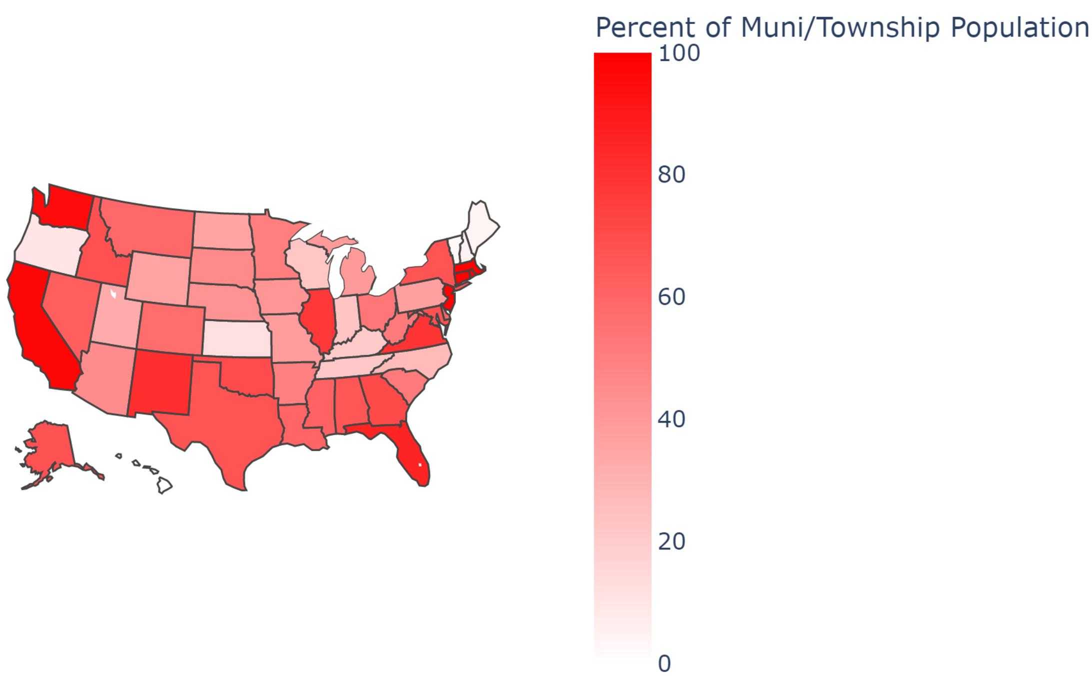
Notes: This heatmap illustrates the pairwise correlations between various zoning regulations across U.S. municipalities. Each cell represents the correlation coefficient between two zoning measures, with color intensity indicating the strength and direction of the relationship. Darker red indicates stronger positive correlations, while darker blue represents stronger negative correlations. White or light-colored cells suggest weak or no correlation. The diagonal represents perfect correlation of each measure with itself. See Appendix Table A1 for full question names. See table 6 footnote for details on sample construction.

Figure A6: Percent of the Population Living in Either a Municipality or Township Government By State



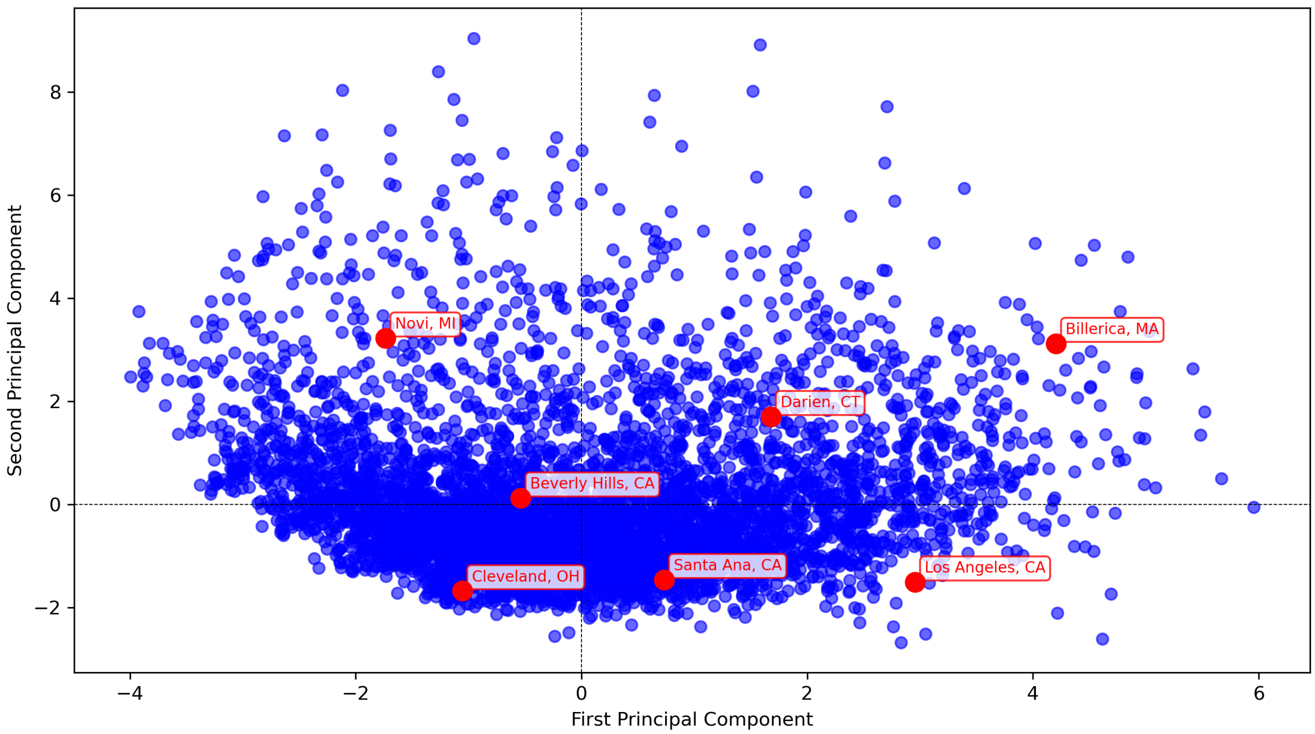
Notes: This Figure shows the fraction of the population in each state who resides in a municipality or township government by state. The remaining fraction of the population live in an unincorporated area. See Table 1 footnote for more details on sample coverage.

Figure A7: State-Level Coverage: Percentage of Municipality and Township Residents in Our Sample



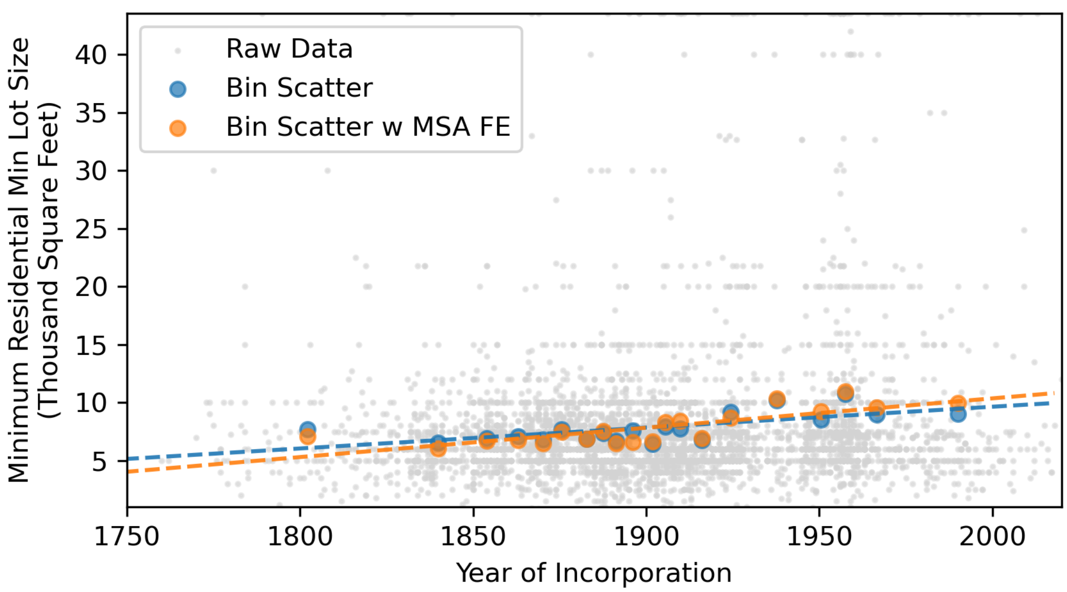
Notes: This Figure shows the fraction of the population in each state, who lives in a municipality or township, which we have municipal code coverage for. See Table 1 footnote for more details on sample coverage.

Figure A8: Relationship between First and Second Principal Components



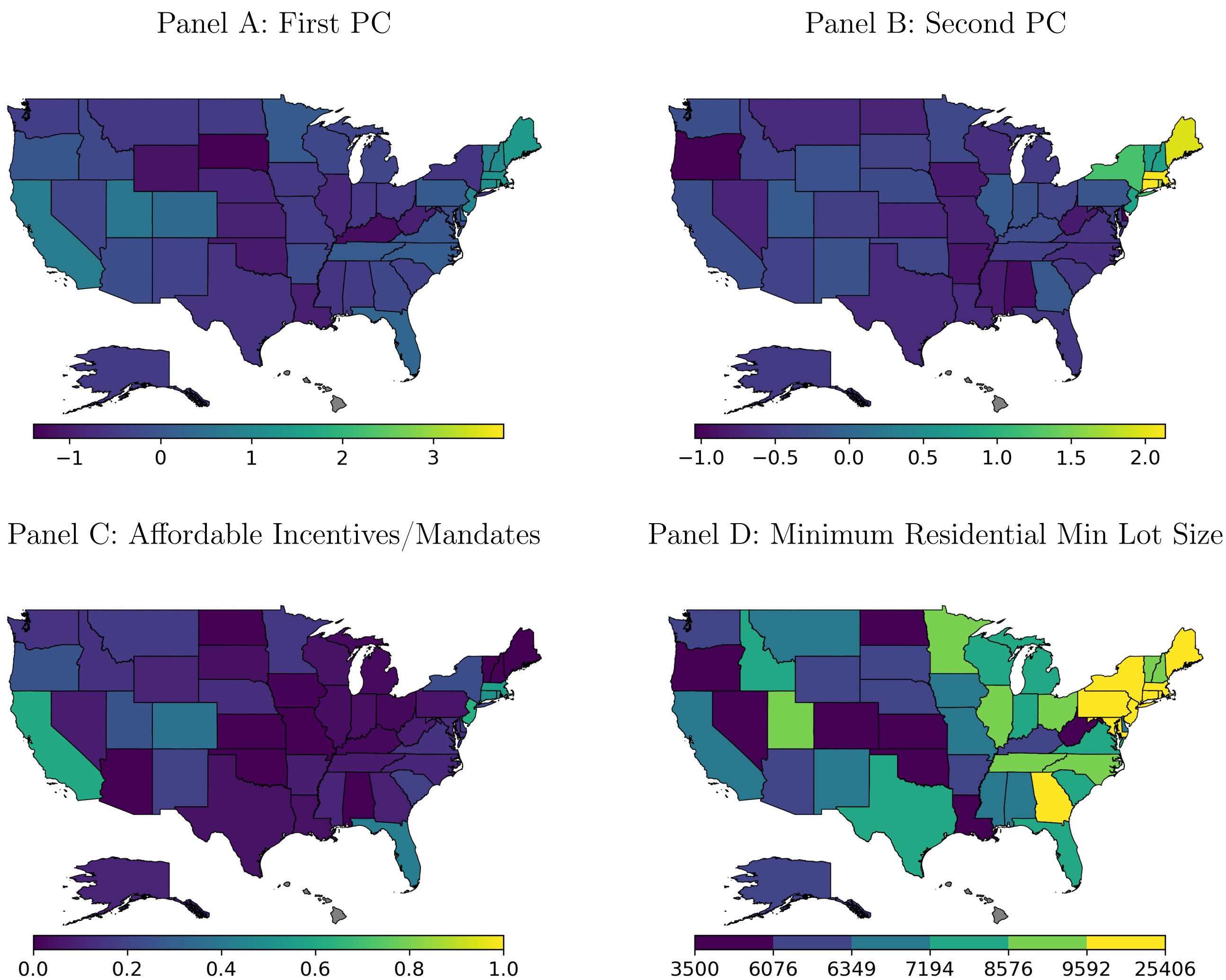
Notes: This scatter plot illustrates the relationship between the first two principal components of our zoning regulation analysis across U.S. municipalities. The x-axis represents the first principal component (PC1), which we interpret as a measure of regulatory complexity and housing demand. The y-axis shows the second principal component (PC2), which corresponds to exclusionary zoning practices. Each point represents a municipality. Municipalities in the upper right quadrant tend to have both complex regulations and more exclusionary practices, while those in the lower left have simpler, less restrictive zoning.

Figure A9: Historical Trends in Minimum Lot Size Requirements by Incorporation Date



Notes: This scatter plot illustrates the relationship between a municipality's year of incorporation and its current minimum lot size requirement. The x-axis spans from 1750 to present, capturing the majority of U.S. municipal incorporations. The y-axis shows minimum lot sizes from 1,000 to 43,560 square feet (1 acre), though some municipalities have larger requirements not shown here. Each point represents a municipality. Blue dots indicate binscatter points and blue lines indicate a linear trend in raw data, while the orange line includes a MSA fixed effect. Note that while the visible plot is truncated, the regression lines are based on the full dataset, including earlier incorporations and larger lot sizes.

Figure A10: Nationwide Maps of Regulatory Variables



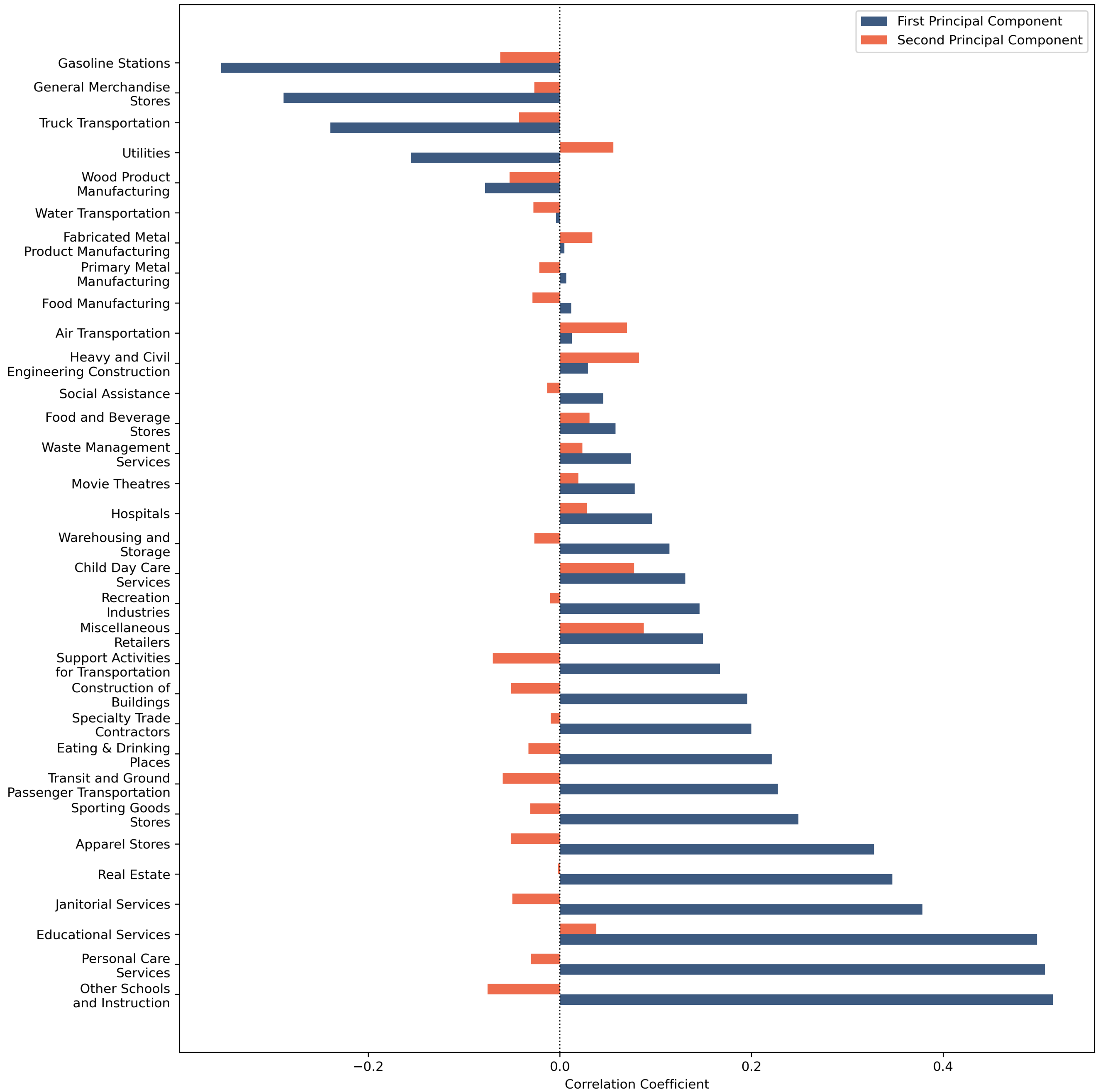
Notes: This Figure shows a comparable averages of regulatory variables across states as in Figure 8, but without population weights. This Figure therefore shows the raw average across all municipalities. Hawaii is grey because only one municipality (Honolulu) is in the dataset. For county maps please see Figure A11.

Figure A11: County Maps of Zoning Measures

Notes: This Figure shows county-level maps of four key zoning measures across the United States. Panels A–B show the first principal component (PC1) of zoning regulations, interpreted as regulatory complexity. Panels C–D display the second principal component (PC2), associated with exclusionary zoning practices. Panels E–F illustrate the prevalence of affordable housing incentives or mandates. Panels G–H depict the minimum residential lot size requirements. For each measure, we present both unweighted averages (left column) and population-weighted averages (right column). Darker colors indicate higher values or greater prevalence of each measure. We use the 2022 ACS Population as the population weight. Hawaii is grey because only one municipality (Honolulu) is in the dataset. For population weighted state maps please see Figure 8 and for unweighted state maps please see Figure A10.

Figure A12: Industry Composition Correlations with Principal Components

Panel A: Per Capita Establishments



Notes: Establishment and employment data are from the U.S. Census Bureau's County Business Patterns (CBP) 2022 dataset. Correlations are calculated at the county level, with the number of establishments and employment for each industry normalized to per capita measures using county population estimates. Principal Component indices are population-weighted averages of municipality level data, aggregated to the county level. Industries are classified using 2017 NAICS codes.

Panel B: Per Capita Employment

Table A1: Mapping of Full Pioneer Institute Study Questions to Short Names

Full Question	Short Question
Is multi-family housing allowed, either by right or special permit (including through overlays or cluster zoning)?	Multifamily Allowed
Are apartments above commercial (mixed use) allowed in any district?	Mixed-Use Buildings
Is multi-family housing listed as allowed through conversion (of either single family homes or non residential buildings)?	Conversion To Multifamily
Are attached single family houses (townhouses, 3+ units) listed as an allowed use (by right or special permit)?	Townhouses Allowed
Does zoning include any provisions for housing that is restricted by age?	Age-Restricted Provisions
Are accessory or in-law apartments allowed (by right or special permit) in any district?	Accessory Apartments Allowed
Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by right?	Flexible Zoning By Right
Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by special permit?	Flexible Zoning By Permit
Does the zoning bylaw/ordinance include any mandates or incentives for development of affordable units?	Affordable Incentive
Is there a town-wide annual or biannual cap on residential permits issued, and/or is project phasing required?	Permit Cap Or Phasing
Are there restrictions on counting wetlands, sloped land or easements in lot size calculations?	Wetlands Restricted in Lot Size Calc
What is the minimum residential minimum lot size?	Minimum Res Min Lot Size
What is the mean residential minimum lot size?	Mean Res Min Lot Size
What is the max residential minimum lot size?	Max Res Min Lot Size
How many zoning districts, including overlays, are in the municipality?	Zoning District Count
What is the longest frontage requirement for single family residential development in any district?	Longest Frontage Requirement
Do developers have to comply with the requirement to include affordable housing, however defined, in their projects?	Affordable Mandate
How many mandatory steps are involved in the approval process for a typical new multi-family building?	Mandatory Approval Steps
For a typical new multi-family building project in this jurisdiction, how many distinct governing bodies or agencies must give mandatory approval before construction can begin?	Distinct Approval Bodies
Are there townwide requirements for public hearings on any type of multi-family residential projects?	Public Hearing Requirements
What is the maximum potential waiting time (in days) for government review of a typical new multi-family building?	Max Review Waiting Time

Notes: This Table shows the mapping between the full regulatory questions and the short question name use throughout the paper. See Appendix Section 6 for further details on questions.

Table A2: Loadings on Principal Components

	First	Second	Third	Fourth	Fifth
Affordable Incentive	0.42	0.10	-0.37	-0.10	0.04
Affordable Mandate	0.32	0.11	-0.46	-0.17	0.06
Age-Restricted Provisions	0.31	0.00	-0.18	0.15	-0.02
Zoning District Count	0.30	-0.20	-0.00	0.14	-0.07
Wetlands Restricted in Lot Size Calc	0.23	0.20	0.06	0.21	0.11
Permit Cap Or Phasing	0.22	0.03	0.05	-0.22	-0.24
Maximum of Residential Min Lot Sizes	0.19	0.37	0.22	0.15	-0.02
Longest Frontage Requirement	0.17	0.40	0.21	0.15	0.03
Public Hearing Requirements	0.15	0.11	-0.13	-0.29	-0.09
Max Review Waiting Time	0.12	0.03	0.06	-0.39	0.41
Minimum of Residential Min Lot Sizes	0.05	0.47	0.19	0.05	-0.06
Distinct Approval Bodies	0.02	0.02	0.43	-0.31	-0.08
Mandatory Approval Steps	0.01	0.10	0.17	-0.52	0.39
No Conversion to Multifamily	-0.09	0.00	-0.19	-0.38	-0.48
No Flexible Zoning By Right	-0.13	0.04	-0.23	0.11	0.53
Multifamily Not Allowed	-0.14	0.35	-0.08	-0.10	-0.13
No Mixed-Use Buildings	-0.21	0.29	-0.20	-0.04	-0.21
Townhouses Not Allowed	-0.23	0.33	-0.14	0.04	0.03
Accessory Apartments Banned	-0.30	-0.02	-0.06	-0.10	0.06
No Flexible Zoning By Permit	-0.31	0.20	-0.31	0.05	0.10
Variance Explained	0.13	0.11	0.06	0.06	0.05

Notes: This Table reports loadings between the first five principal components of our regulatory dataset and specific regulatory questions. We upper winsorize at the 1% level the values for maximum residential min lot size, minimum residential min lot size, longest frontage requirement, and max review waiting time. We transform the maximum residential min lot size variable into a dummy for whether it is above one acre. Missing data, where the model output "I don't know," were imputed with k-nearest neighbors. Prior to performing principal component analysis, all variables were normalized into z-scores. Additionally, each variable was expressed in terms of its expected univariate association with stricter zoning policies, such that more positive values indicate a greater degree of restrictiveness. For example, the variable representing the allowance of multi-family housing was inverted, so that a more positive value indicates that multi-family housing is not permitted, while a more negative value suggests that it is not.

Table A3: Description of Variables Used in the Study

Variable	Source	Definition
Auto Commute Share	2022 American Community Survey	Percentage of commuters using either cars, trucks, or vans as their primary commute method.
Foreign Born Share	2022 American Community Survey	The percentage of the population that is foreign-born (B05002_013E / B05002_001E).
Median Household Income	2022 American Community Survey	The median income of all households (B19013_001E).
Share Population 65 and Over	2022 American Community Survey	The percentage of the population aged 65 and over (B01001_020E to B01001_025E and B01001_044E to B01001_049E / B01001_001E).
Median Gross Rent	2022 American Community Survey	The median gross rent for rental units (B25064_001E).
Median Home Value	2022 American Community Survey	The median value of owner-occupied housing units (B25077_001E).
Share Units Owner Occupied	2022 American Community Survey	The percentage of housing units that are owner-occupied (B25003_002E / B25003_001E).
Share Population Under 18	2022 American Community Survey	The percentage of the population under 18 years old (B01001_003E to B01001_006E and B01001_027E to B01001_030E / B01001_001E).
White Share	2022 American Community Survey	The percentage of the population identifying as White (B02001_002E / B02001_001E).
Poverty Rate	2022 American Community Survey	The percentage of the population living below the poverty line (B17001_002E / B17001_001E).
College Share	2022 American Community Survey	The percentage of the population aged 25 and over with a bachelor's degree or higher (B15003_022E, B15003_023E, B15003_024E, B15003_025E / B15003_001E).
Share Structures Built Before 1970	2022 American Community Survey	The percentage of housing structures built before 1970 (B25034_008E, B25034_009E, B25034_010E, B25034_011E / B25034_001E).

Continued on next page

Table A3 continued from previous page

Variable	Source	Definition
Share Structures with 2 or More Units	2022 American Community Survey	The percentage of housing structures with 2 or more units (B25024_004E to B25024_009E / B25024_001E).
Vacancy Rate	2022 American Community Survey	The percentage of vacant housing units (B25002_003E / B25002_001E).
Share with Commute Over 30 Minutes	2022 American Community Survey	The percentage of workers with a commute time over 30 minutes (B08303_008E to B08303_013E / B08303_001E).
Housing Unit Density	2022 American Community Survey	The number of housing units in a local government divided by the area from its shape file.
Share Units Affordable	2022 American Community Survey	The percentage of housing units affordable to households earning the state median income. This measure combines rental and owner-occupied housing affordability, determined using the state median income. Rental units are affordable if the monthly rent does not exceed 30% of the monthly median household income, and owner-occupied units are affordable if their value is less than three times the annual median household income. The total number of affordable rental and owner-occupied units is summed and divided by the total number of housing units to determine the share of units that are affordable.
Job Density	Opportunity Insights	Number of jobs per square mile in each census tract in 2013. "job_density_2013" from the Opportunity Atlas neighborhood characteristics dataset.
Opportunity Index	Opportunity Insights	The kid family rank, a measure of economic mobility.
Average Math Test Scores	The Education Opportunity Project at Stanford University	The average math test score pooled across grades (3rd-8th) and years (2008–2019) (cs_mn_avg_mth_ol).

Continued on next page

Table A3 continued from previous page

Variable	Source	Definition
Math Learning Rate	The Education Opportunity Project at Stanford University	The slope of the increase in math test scores from 3rd to 8th grade pooled across years (2008-2019) (cs_mn_grd_mth_ol).
Percent Eligible for Free Lunch	The Education Opportunity Project at Stanford University	The percentage of students eligible for free lunch (perflu).
Property Tax Rate	The Government Finance Database	The property tax rate is calculated as the total property tax revenue (Property_Tax_2017) divided by the aggregate home value from the 2017 ACS. This excludes property taxes raised from independent school districts.
Total Expenditure Per Capita (2017)	The Government Finance Database	Total expenditures per capita is calculated as the total government expenditures (Total_Expenditure_2017) divided by the population (From 2017 ACS) of the municipality or township.
Local Revenue Per Student	2022 Annual Survey of School System Finances	The local revenue for a school district divided by the enrollment of that school district. We spatially merge school districts into the Census of Governments. Local revenue for school districts includes property taxes directly raised by the school and transfers from local governments for subordinate school districts.
Building Permits All Units	Building Permits Survey	The number of housing units permitted divided by the population of the local government averaged over 2019–2023.
Year of Incorporation	Goodman, C. B. (2023). Municipal Incorporation Data, 1789-2020 (Version 1.0.0) [Computer software]. https://github.com/cbgoodman/muni-incorporation/	The year a municipality was incorporated. Not available for townships.

Continued on next page

Table A3 continued from previous page

Variable	Source	Definition
Percent Democrat	"U.S. Voting by Census Block Groups", Bryan, Michael	The share of votes that are Democrat in 2021.
New Housing Unit Elasticity	Baum-Snow and Han (2024)	The coefficient on the change in new housing units to the change in housing prices, specifically gamma01b_newunits_FMM.
Log Land Area	2022 Census Shapefiles	The area in acres of a local government with a log transform.
Log Neighbors within 25 Miles	2022 Census of Governments	The number of other local governments within 25 miles of a local government.
Miles to Metro Center	2022 Census Shapefiles	The number of miles from the centroid of a local government's shape file to the center of a metropolitan area, defined as the city hall of the center city.

Appendix B: LLM Replicability

LLMs responses are not fully deterministic and so the results of this study cannot be completely replicated for several reasons. First, LLMs sample tokens from a probability distribution leading to variation in responses to the same query across iterations. Second, the GPT-4 class of models that we use in this study follow a Mixture of Exports (MoE) architecture (see here and here for further details). This means that the specific expert that an LLM query gets routed to varies across API calls depending on supply/demand of experts. In turn, this implies that the underlying probability distribution that tokens are sampled from may change from one query call to the next, depending on the availability of experts. Moreover, OpenAI's models are closed source so there may exist other sources of randomness across API calls that we cannot explain.

In this section we quantify how deterministic LLM queries are in our use case and how ensembling many query calls may help mitigate the issue. In general, we query the LLM to respond with a detailed answer, (i.e. "Think step by step") followed by a structured output (i.e. "Yes" or "No"). Randomness in output for LLMs leads to a high frequency of variation of in open-ended responses, but many of these differences do not change the overall meaning of a response, for example just swapping synonyms. However, we do find some variation in the structured output of LLM responses in our use case.

One potential way to mitigate non-determinism is to request multiple responses from the LLM and then aggregate the answers by majority rule, an ensemble approach. Previous research has also found that ensemble methods can greatly improve LLM performance (Li et al., 2024).

With the OpenAI API there are two ways to ensemble API calls. The first way is to request multiple chat completions for a given query (by setting the API parameter n > 1). This effectively samples the distribution of tokens several times. This method is also cost effective because OpenAI only charges the user once for the input tokens regardless of how many iterations of output tokens are requested. However, this approach fails to sample from the distribution of potential experts or other potential sources of variation, for example the hardware of the server in which the LLM was run. A more costly approach is to separately query the LLM for each of the ensemble queries paying for both the input and output tokens used in each call. This second approach more broadly samples from the various sources of randomness for an LLM response.

We measure replicability with two measures in this analysis. Both measures average pairwise matching rates. We compare the final structured answer from a given LLM query across multiple API calls for all pairs (n choose 2), and take the ratio of the number of pairs that match to the total number of pairs. We call this ratio a consistency score. We measure both internal consistency, scores from requesting multiple chat completions for a given query, and external consistency, scores from comparing separate API calls.

We confirm that lower temperatures create more deterministic responses, even after ensembling. In Table A4, we compare temperatures of 0, 0.5, and 1 (lower temperatures should mean more deterministic responses) as well as whether including a random seed makes responses more deterministic. We use a random sample of 30 municipalities from our national sample and use two questions, a binary one (whether there are permits caps or project phasing) and a continuous one (how many districts there are). For each specification we run the model five times, each time requesting 10 chat completions. We measure internal consistency scores within a model run, and external consistency across aggregated majority rule answers from each model run. We find that lower temperature models are more internally consistent, especially for the continuous question, and have a lower variance of internal consistency. After aggregating responses, we find that external consistency scores are fairly similar for the binary question, but still higher for the continuous one and with lower variance. We also do not find evidence that including a random seed makes responses more deterministic.

We next show in Figure A13 that external consistency grows with ensemble size, especially for the continuous question. This suggests that answers begin to stabilize at larger ensemble sizes, though not fully. We still find that at least five percent of pairwise comparisons do not match even with ensemble sizes of 10 and a temperature of 0. We also find that the zero temperature specification already begins at a fairly high level of external consistency even with an ensemble size of 1.

Next, we ask how informative the internal consistency score is for predicting the external consistency score. In Figure A14 we find that higher temperature model internal consistency scores are highly predictive of external consistency and that this effect grows with ensemble size. However, lower temperature models internal consistency scores are not very informative for external consistency. This suggests that within model run variation for low temperature models may not reflect the same source of randomness as across model runs. For example, the variation in which expert the query is routed to may be more important than the within expert sampling distribution for predicting external consistency.

We next explore whether performance increases with ensemble size. For this analysis we return to our testing sample of 30 Massachusetts municipalities where we have a clean dataset to compare answers to. We use temperatures of 0 and 1 and request 10 chat completions from each model run. We do not find evidence of increased performance with larger ensemble sizes. In Figure A15 we find a fairly persistent outperformance of the 0 temperature model for the binary question and of the 1 temperature model for the continuous question.

We next ask whether internal consistency scores are helpful for predicting accuracy of questions. If the internal consistency score is highly predictive of the external consistency score then the measure can be used as a model confidence measure. We find in Table A5 that the internal consistency score is somewhat helpful for predicting accuracy with the temperature 1 model but not helpful for the temperature 0 model.

We suggest researchers consider one of two specifications. First, a low temperature single shot approach. This approach is cost effective by only requesting one chat completion, has a high degree of external consistency, and is straightforward to explain. If a researcher wants to use an ensemble approach then we suggest using a high temperature model with an ensemble size of at least five. High temperature ensemble models have a high degree of external consistency and have informative internal consistency scores for both predicting external consistency and accuracy. We choose to use the first approach in this paper for the higher external consistency, cost savings, and for simplicity.

Table A4: Internal and External Consistency Varying Temperature and Seed

Question	Temperature	Seed	External Consistency Mean	External Consistency Variance	Internal Consistency Mean	Internal Consistency Variance
Continuous: Number of Districts	0	No	0.91	0.05	0.9	0.05
		Yes	0.89	0.04	0.9	0.04
	0.5	No	0.83	0.09	0.68	0.1
		Yes	0.8	0.11	0.67	0.1
	1	No	0.75	0.11	0.61	0.1
		Yes	0.78	0.11	0.64	0.1
Binary: Whether Permit Caps or Phasing	0	No	0.93	0.03	0.96	0.02
		Yes	0.92	0.03	0.95	0.02
	0.5	No	0.93	0.03	0.88	0.04
		Yes	0.9	0.05	0.89	0.04
	1	No	0.91	0.04	0.88	0.04
		Yes	0.92	0.04	0.86	0.04

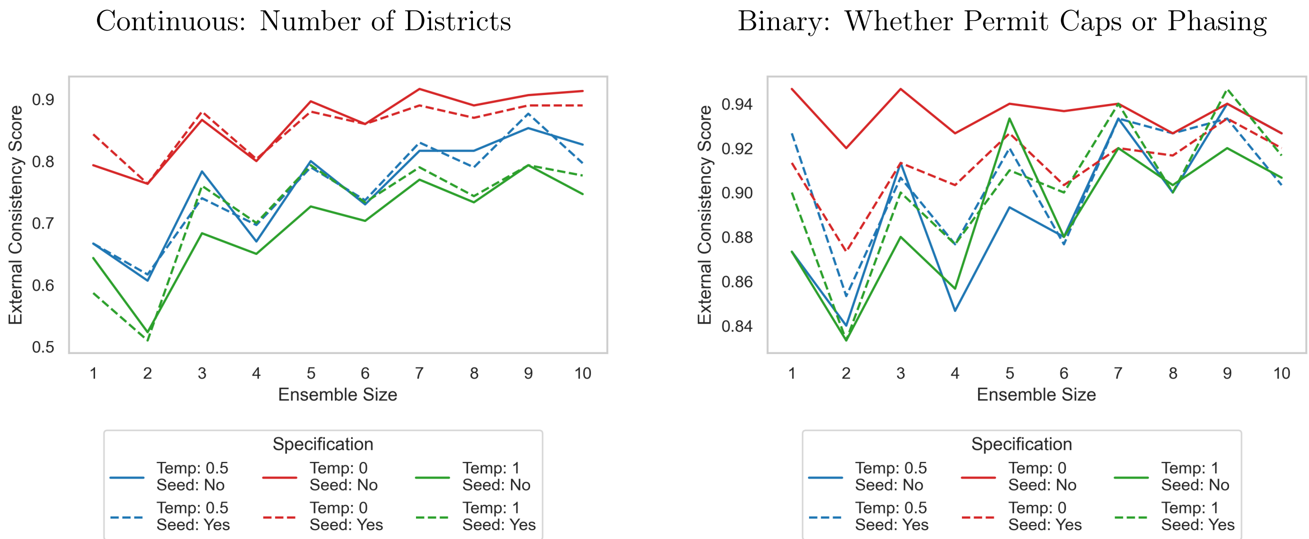
Notes: This Table compares the internal and external consistency of LLM responses across different temperature settings and with/without a random seed. "Temperature" refers to the randomness in the LLM's output (0 being most deterministic, 1 being most random). "Seed" indicates whether a random seed was used for reproducibility. "External Consistency" measures agreement across separate API calls, while "Internal Consistency" measures agreement within a single API call requesting multiple completions. Results are shown for two types of questions: a continuous question about the number of zoning districts, and a binary question about permit caps or phasing. Mean values closer to 1 indicate higher consistency. Lower variance indicates more stable results across trials.

Table A5: Regression of Internal Consistency Score on Whether Correct

	Temp = 0		Temp = 1	
	(1)	(2)	(3)	(4)
Consistency Score		0.2603		0.4175*
		(0.3614)		(0.2097)
Question FE	Yes	Yes	Yes	Yes
R-squared	0.1383	0.1461	0.2188	0.2696
N	60	60	60	60

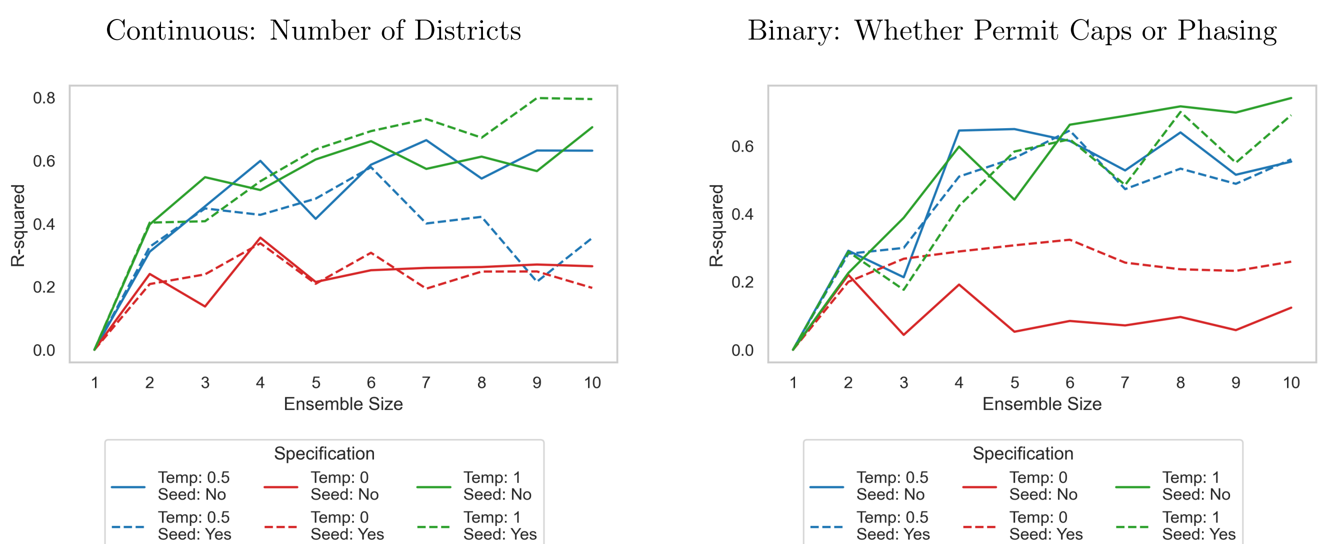
Notes: This table presents regression results examining the relationship between the internal consistency score of LLM responses and their accuracy. The dependent variable is a binary indicator for whether the LLM's response is correct. Results are shown for two temperature settings: 0 (most deterministic) and 1 (most random). Columns 1 and 3 include only question fixed effects, while columns 2 and 4 add the consistency score as an explanatory variable. Standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure A13: External Consistency vs. Ensemble Size



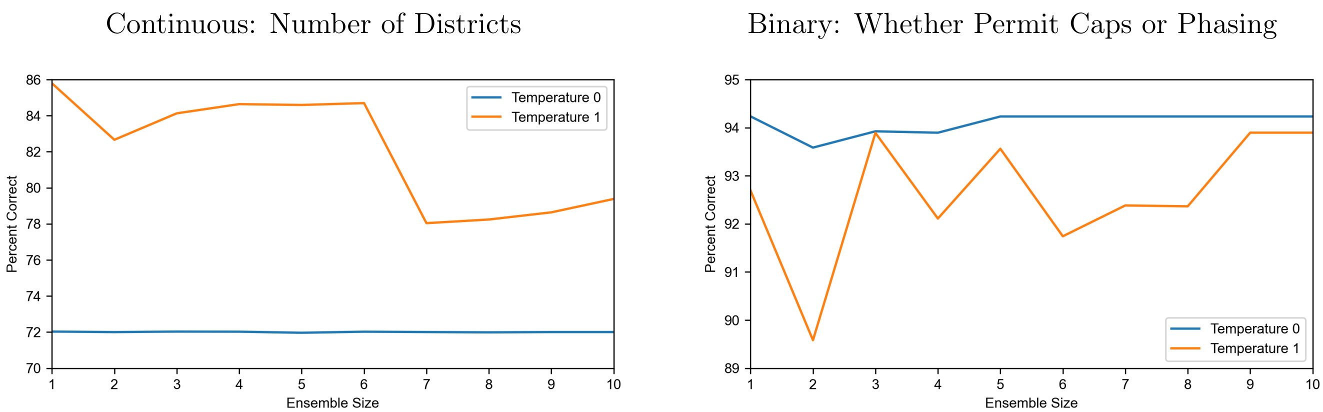
Notes: This Figure shows how external consistency of LLM responses varies with ensemble size for two types of questions: a continuous question about the number of zoning districts (left) and a binary question about permit caps or phasing (right). The x-axis represents the ensemble size (number of model runs aggregated), while the y-axis shows the external consistency score. Different lines represent various temperature settings (0, 0.5, 1) and whether a random seed was used.

Figure A14: Relationship Between Internal and External Consistency vs. Ensemble Size



Notes: This figure demonstrates the relationship between internal and external consistency of LLM responses as a function of ensemble size for two types of questions: a continuous question about the number of zoning districts (left) and a binary question about permit caps or phasing (right). The x-axis represents the ensemble size, while the y-axis shows the R-squared value, indicating how well internal consistency predicts external consistency. Different lines represent various temperature settings (0, 0.5, 1) and whether a random seed was used.

Figure A15: Ensemble Size vs Percent Correct



Notes: This Figure illustrates the relationship between ensemble size and accuracy of LLM responses for two types of questions: a continuous question about the number of zoning districts (left) and a binary question about permit caps or phasing (right). The x-axis represents the ensemble size (number of model runs aggregated), while the y-axis shows the percent of correct responses. Two temperature settings are compared: 0 (blue line, most deterministic) and 1 (orange line, most random). For the continuous question, accuracy is measured as the percentage of responses within a certain tolerance of the true value, while for the binary question, we use the percentage of correct classifications.

Appendix: Question Details

This appendix provides detailed information about each question used in the study. Each question is presented with its original phrasing by the Pioneer Institute, the text that we embed for the question, background information and assumptions, question type, and the rephrased question that the language model sees. For some questions, we also include a value that triggers double-checking if the model's answer does not match it, along with the rephrased question used for double-checking and the keywords used to build context during the double-checking process. Additionally, certain questions involve subtasks, which are described in detail.

Question 4

Question Phrased by Pioneer: Is multi-family housing allowed, either by right or special permit (including through overlays or cluster zoning)?

Question Text That We Embed: Is multi-family housing allowed, either by right or special permit (including through overlays or cluster zoning)?

Question Background and Assumptions: Multi-family housing comes in a wide variety of forms and sizes. The ways municipalities define and categorize "multi-family" housing varies widely, as do the use-regulations that govern multi-family housing development. This study includes as "multi-family" any building with three or more dwelling units. Multi-family dwelling units can be rental or condominium. They can be in a freestanding residential building or part of a mixed-use building, new construction or conversion of a preexisting building. Zoning documents usually specify what kinds of buildings qualify for conversion to multi-family housing: single family houses, two family houses, mills, schools, churches, municipal buildings or other types of facilities. Freestanding new "Multi-family" housing is defined as any building with three or more dwelling units, excluding townhouses, unless a municipality includes townhouses in its broader definition of multi-family housing and effectively permits only townhouses as such. Assisted living facilities, congregate care homes, dormitories, and lodging houses are not considered multi-family housing. If the zoning laws allow for conversion to multi-family housing, but do not comment on whether new multi-family housing is allowed, then the answer is 'YES'. Most towns allow a form of multi-family housing.

Question Type: Binary

Rephrased Question the LLM Sees: Is multi-family housing allowed at all in any district or overlay? If multi-family housing is allowed by special permission in any district or overlay then that counts allowed.

Question 5

Question Phrased by Pioneer: Are apartments above commercial (mixed use) allowed in any district?

Question Text That We Embed: Are apartments above commercial (mixed use) allowed in any district?

Question Background and Assumptions: Zoning bylaws and ordinances in various municipalities often contain provisions for combining residential dwellings with commercial uses such as retail or office spaces, creating mixed-use developments. While some zoning regulations explicitly allow multi-family housing and retail to coexist within the same district, they may not clarify whether these uses can share the same building, leaving this to be determined in practice. Certain municipalities explicitly permit "combined dwelling/retail" configurations in their use regulation tables, sometimes noting that any uses allowed within the same district can occupy the same building. Additionally, detailed provisions for mixed-use are facilitated through special zoning arrangements like overlay districts (e.g., mixed use district, downtown overlay, or planned unit development) or conversion projects, such as transforming former mills to accommodate both retail and housing. However, it's important to note that some references to "mixed use" may actually pertain to commercial and industrial combinations, excluding residential components. If you cannot find any reference to residential and commercial uses in the same building within the context then you assume that the answer is 'NO'.

Question Type: Binary

Rephrased Question the LLM Sees: Is a combination of commercial and residential uses in the same building or structure allowed in any zoning district?

Question 6

Question Phrased by Pioneer: Is multi-family housing listed as allowed through conversion (of either single family homes or non residential buildings)?

Question Text That We Embed: Is multi-family housing listed as allowed through conversion (of either single family homes or non residential buildings)?

Question Background and Assumptions: The development of multifamily housing through the conversion of existing buildings encompasses two primary approaches: transforming single-family or two-family houses into structures with at least three units, and repurposing non-residential buildings, such as mills, other industrial buildings, schools, and municipal buildings, for multi-family residential use. This is different from the ability to construct new multi-family housing. The conversion of non-residential structures often occurs through designated overlay districts, like Mill Conversion Overlay Districts, or within industrial zones, whereas the conversion of houses to accommodate more units typically takes place in residential or business districts. The question does not count the conversion of single-family homes into two-family dwellings as allowing conversion to multi-family dwellings because multi-family is defined as having at least three units. If the conversion requires a special permit then we consider that as allowing conversion. Assisted living facilities, congregate care homes, dormitories, and lodging houses are not considered multi-family housing. The allowance of multi-family housing does not imply the allowance of the conversion to multi-family housing. You must search for an explicit statement allowing the conversion to multi-family housing from another type of structure. If you do not find any mention of conversions in the context then you assume the answer is 'NO'.

Question Type: Binary

Rephrased Question the LLM Sees: In any district, is the conversion to multi-family explictly allowed under any scope?

If The Answer Is Not This Value Then We Double Check: Yes

Rephrased Question the LLM Sees When Double Checking: In any district, is the conversion to multi-family explictly allowed under any scope?

Keywords We Use to Build Context When Double Checking in Order of Importance: 'conver'

Question 8

Question Phrased by Pioneer: Are attached single family houses (townhouses, 3+ units) listed as an allowed use (by right or special permit)?

Question Text That We Embed: Are attached single family houses (townhouses, 3+ units) listed as an allowed use (by right or special permit)?

Question Background and Assumptions: The question asks whether some form of attached housing is allowed in the municipality. Common forms of attached housing are single-family attached homes, townhouses, rowhouses, and zero lot line dwelling units. Attached housing is often allowed through special zoning provisions, such as overlay districts or use provisions tailored for cluster developments, Planned Unit Developments (PUD), or communities for active adults aged 55 and over. Remember that accessory apartments to a single-family home or the ability to attach one unit to a single-family home do not count as attached housing. Duplexes also do not count as attached housing. A form of attached housing may be listed as a type of single-family or multi-family housing. However, the allowance of single-family or multi-family housing does not imply the allowance of attached housing. This context raises the question of whether any type of attached housing are allowed either as their own category of housing or explicitly as a type of single family or multi-family housing. If you do not find any mention of a type of attached housing in the context then you assume that the answer is 'NO'.

Question Type: Binary

Rephrased Question the LLM Sees: Is some form of attached housing allowed in any district of the town?

If The Answer Is Not This Value Then We Double Check: Yes

Rephrased Question the LLM Sees When Double Checking: Is some form of attached housing allowed in any district of the town?

Keywords We Use to Build Context When Double Checking in Order of Importance: 'town house', 'town houses', 'townhouse', 'townhouses', 'attached dwelling', 'attached dwellings', 'row house', 'row houses', 'rowhouse', 'rowhouses', 'attached single family', 'attached unit', 'attached units', and 'attached'

Question 9

Question Phrased by Pioneer: Does zoning include any provisions for housing that is restricted by age?
Question Text That We Embed: Does zoning include any provisions for housing that is restricted by age?
Question Background and Assumptions: Many zoning bylaws/ordinances include provisions for housing that is deed restricted to occupants 55 (or another age) and older. Some of the provisions are for developments that are entirely age-restricted, while other provisions are incentives, often density bonuses, to include age-restricted units within an unrestricted development, such as cluster or multi-family. The restricted developments are called active adult housing, adult retirement village, senior village, planned retirement community, or something similar.
The answer should be Yes if any provisions exist for age-restricted single-family, townhouse, duplex, multi-family or accessory apartments. Provisions can be in the form of an age-restricted overlay, cluster development, density bonus for age-restricted units, or other zoning requirements or incentives for age-restricted housing.
Question Type: Binary
Rephrased Question the LLM Sees: Does zoning include any provisions for housing that is restricted by age?

Question 11

Question Phrased by Pioneer: Are accessory or in-law apartments allowed (by right or special permit) in any district?
Question Text That We Embed: Are accessory or in-law apartments allowed (by right or special permit) in any district?
Question Background and Assumptions: Accessory dwellings are separate housing units typically created in surplus or specially added space in owner-occupied single-family homes. Accessory dwellings can also be attached to the primary dwelling or be situated on the same lot (for example in a carriage house or small cottage.) An accessory dwelling typically has its own kitchen and bathroom facilities, not shared with the principal residence. Many zoning bylaws/ordinances call the dwellings "in-law apartments" or "family apartments" and restrict their occupancy to relatives of the homeowner - "related by blood, marriage or adoption." Some of these also allow domestic employees, caregivers, elderly people or people with low incomes to live in the units. Some municipalities allow the apartment

by right if a family member will occupy the accessory apartment, but require a special permit otherwise. If you cannot find any reference to accessory apartments in the context then you assume that the answer is 'NO'.

Question Type: Binary

Rephrased Question the LLM Sees: Are accessory or in-law apartments allowed in any district? If they are allowed by special permit in any district then we count that as allowed.

Question 13

Question Phrased by Pioneer: Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by right?

Question Text That We Embed: Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by right?

Question Background and Assumptions: Flexible zoning, encompassing terms like open space residential design, cluster, planned unit development, or conservation subdivision, provides municipalities with a more adaptable approach to zoning beyond the traditional "as-of-right" options. This methodology allows developers to bypass the stringent requirements of standard zoning, such as specific lot sizes and setback mandates, and enables the incorporation of various residential unit types like townhouses, duplexes, and multi-family homes that might not be allowed under conventional zoning regulations. The question only considers provisions that are primarily for residential uses. Most municipalities require special permits for cluster/flexible development.

Question Type: Binary

Rephrased Question the LLM Sees: Is the answer yes to any of the following question? Question 1: Is cluster development allowed explictly by right in any district? Question 2: Is open space residential design allowed explictly by right in any district? Question 3: Is any type of flexible zoning other than cluster development and open space residential design allowed explictly by right in any district?

Question 14

Question Phrased by Pioneer: Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by special permit?

Question Text That We Embed: Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by special permit?

Question Background and Assumptions: Flexible zoning, encompassing terms like open space residential design, cluster, planned unit development, or conservation subdivision, provides municipalities with a more adaptable approach to zoning beyond the traditional “as-of-right” options. This methodology allows developers to bypass the stringent requirements of standard zoning, such as specific lot sizes and setback mandates, and enables the incorporation of various residential unit types like townhouses, duplexes, and multi-family homes that might not be allowed under conventional zoning regulations. The question only considers provisions that are primarily for residential uses. Most municipalities require special permits for cluster/flexible development so if you find suggestive evidence that the municipality allows cluster/flexible development by special permit then you assume that the answer is ’YES’.

Question Type: Binary

Rephrased Question the LLM Sees: Is the answer yes to any of the following question? Question 1: Is cluster development allowed in any district, including by special permit? Question 2: Is open space residential design allowed in any district, including by special permit? Question 3: Is any type of flexible zoning other than cluster development and open space residential design allowed in any district, including by special permit?

Question 17

Question Phrased by Pioneer: Does the zoning bylaw/ordinance include any mandates or incentives for development of affordable units?

Question Text That We Embed: Does the zoning bylaw/ordinance include any mandates or incentives for development of affordable units?

Question Background and Assumptions: Inclusionary zoning requires or encourages developers to include affordable dwelling units within new developments of market rate homes. Some municipalities call it “incentive zoning” - when provision of affordable units is voluntary. The affordable units are typically located on site, but some municipalities also allow off-site development under certain circumstances. Often, payments may be made to a trust fund in lieu of building housing. Housing designated as “affordable” must be restricted by deed or covenant, usually for a period of 30 or more years, to residents with low or moderate incomes. The deed restrictions also limit sales prices and rents as the units are vacated, sold or leased to new tenants.

Do not include provisions for entirely affordable, subsidized housing development by public or non-profit corporations. Also do not include provisions under “rate of development” headings that exempt affordable units from project phasing and growth caps.

Question Type: Binary

Rephrased Question the LLM Sees: Does the zoning bylaw/ordinance include any mandates or incentives for development of affordable units?

Question 20

Question Phrased by Pioneer: Is there a town-wide annual or biannual cap on residential permits issued, and/or is project phasing required?

Question Text That We Embed: Is there a town-wide annual or biannual cap on residential permits issued, and/or is project phasing required?

Question Background and Assumptions: Some municipalities enact town-wide caps limiting the number of units that can come on line annually or biannually. The number of permits is often set at the average in the previous years. Note that this question asks only about town-wide caps and does not consider caps exclusive to a specific district in the town. Some municipalities require phased growth for individual developments (also known as development scheduling or buildout scheduling) - a technique that allows for the gradual buildout of approved subdivisions over a number of years. Note that we only consider project phasing when it is required and not when it is optional. Project phasing is usually triggered by a minimum number of units in the project, so small subdivisions can be constructed in one year. Some phasing provisions are only triggered at the town-wide level once a threshold number of units have been permitted. Most of the "rate of development" provisions include an expiration or "sun set" date (some that have expired have been updated and re-adopted). Many include a "point system" where points are awarded for provision of community goods such as open space or affordable units, and projects with more points are given priority for permits. If you do not find any information in the context about a town-wide annual or biannual cap or about project phasing then you assume the answer is 'NO'.

Question Type: Binary

Rephrased Question the LLM Sees: Is the answer yes to any of the following question? Question 1: Is there a town-wide annual or biannual cap on residential permits issued Question 2: Is project phasing required?

Question 21

Question Phrased by Pioneer: Are there restrictions on counting wetlands, sloped land or easements in lot size calculations?

Question Text That We Embed: How is lot area defined and how is the lot size calculated?

Question Background and Assumptions: Remember to first review your research so far on how a lot size is calculated and defined. If you have already found a restriction on including wetlands, sloped land, or easements in your prior research then the answer is 'YES'.

Some municipalities require that the minimum lot size requirement be met by a percentage of land that does not include wetland resource areas, steeply sloped land or easements. A subset of those municipalities requires that the buildable area be contiguous on the lot – called "contiguous buildable area" or "contiguous upland area." Upland area is non-wetland area. It is much more common for municipalities to restrict the use of wetlands areas in meeting lot size requirements than sloped land or easements.

Note that this question only asks about whether there are restrictions on calculating the lot size. It does not ask about whether there are restrictions to buildable area or whether there are any restrictions in wetland areas.

If you do not find any restrictions for lot size calculations in the context then you assume that the answer is 'NO'.

Question Type: Binary

Rephrased Question the LLM Sees: Detail how lot area is defined and how a lot size is calculated. Then, answer the question of are there restrictions on counting wetlands, uplands, or sloped land in lot area/lot size calculation?

If The Answer Is Not This Value Then We Double Check: Yes

Rephrased Question the LLM Sees When Double Checking: Are there restrictions on counting wetlands, sloped land or easements in lot size calculations?

Keywords We Use to Build Context When Double Checking in Order of Importance: 'wetland', 'upland', 'sloped land', and 'easement'

Question 28

Question Phrased by Pioneer: What is the minimum lot size for single-family homes in each residential district?

Question Text That We Embed: What is the minimum lot size for single-family homes in each residential district?

Question Background and Assumptions: When compiling a list of minimum lot sizes for districts that permit single-family housing, prioritize clarity by selecting the specific minimum lot size for single-family homes within each district. If multiple options exist, choose the most common standard size, excluding sizes for historic properties or special cases. Report sizes in square feet over acres unless only acre measurements are available. Only include districts with a defined minimum lot size or those adhering to a town-wide minimum if no district-specific size is established. Finalize the data in a CSV format with columns for 'District Name', 'Min Lot Size', 'Unit', and 'Estate', ensuring a straightforward, single entry for each district that reflects the standard requirement for single-family homes.

Question Type: Lot Size

Rephrased Question the LLM Sees: What is the minimum lot size for single-family homes in each residential district?

Subtask:

- Subtask Question That Gets Embedded: Find the name of each district that allows single-family housing

- Rephrased Subtask Question the LLM Sees: Find the name of each district that allows single-family housing
- Additional Subtask Instructions: Please list out the name of each residential district in the town that primarily consist of detached single-family housing. If you cannot find any districts that explictly allow single-family detached housing then just assume that any residential districts allow single-family detached housing. Respond with a detailed answer followed by a CSV format with the name of the district in the first column and whether a district has the label 'Estate' in the second column as a True/False statement. Use the column headers of 'District Name' and 'Whether Estate District'.
- How The Subtask Results Are Described to the LLM Afterwards: Your previous work finding which districts to find minimum lot sizes for and whether they are estate districts

Question 2

Question Phrased by Pioneer: How many zoning districts, including overlays, are in the municipality?

Question Text That We Embed: How many zoning districts, including overlays, are in the municipality?

Question Type: Numerical

Rephrased Question the LLM Sees: How many zoning districts and overlays are in the municipality?

Question 22

Question Phrased by Pioneer: What is the longest frontage requirement for single family residential development in any district?

Question Text That We Embed: What is the longest frontage requirement for single family residential development in any district?

Question Type: Numerical

Rephrased Question the LLM Sees: What is the longest frontage requirement for single family residential development in any district?

Subtask:

- Subtask Question That Gets Embedded: Find the name of each single-family residential district
- Rephrased Subtask Question the LLM Sees: Find the name of each single-family residential district
- Additional Subtask Instructions: Please list the names of each single-family residential district. Only include districts that are primarily residential. Usually, this means districts that start with the letter R like R1. If

there is only one residential district that permits single-family zoning then just name that one district. If you are unsure whether a residential district permits single-family zoning then assume that it does, but ensure that the district is primarily residential. An agricultural (A) or industrial (I) district would not be included for example.

- How The Subtask Results Are Described to the LLM Afterwards: Only consider the frontage requirements in the following districts

Question 17w

Question Text That We Embed: Do developers have to comply with the requirement to include affordable housing, however defined, in their projects?

Question Background and Assumptions: Zoning codes may require developers to include affordable housing in market-rate residential projects, but the applicability of these requirements can vary. Some inclusionary policies apply broadly to all residential development, while others are tied to optional zoning designations, incentive programs, or specific areas.

To determine if a zoning code contains a mandatory inclusionary requirement, look for clear language stating that all or most market-rate residential projects must provide affordable units as a standard condition of approval under normal zoning rules. The requirement should not be limited to projects that opt into a special zoning designation, participate in an incentive program, or are located in a particular overlay zone.

Focus on whether the code unambiguously requires all or most market-rate residential development to include affordable housing under the generally applicable rules. Do not select "YES" if affordable housing is only mandatory in narrow, specialized situations. The mere presence of affordable housing provisions is not sufficient if they are elective or only apply in atypical circumstances. If the affordable housing requirements are not clearly universally applicable, the likely answer is "NO".

Question Type: Binary

Rephrased Question the LLM Sees: Do developers have to comply with the requirement to include affordable housing, however defined, in their projects?

Question 30

Question Text That We Embed: How many mandatory steps are involved in the approval process for a typical new multi-family building?

Question Background and Assumptions: The approval process for constructing a new multi-family building typically involves multiple mandatory steps, each representing a distinct interaction or requirement that a developer must fulfill before construction can begin. Focus on identifying only the core, pre-construction approval steps that are required for all multi-family building projects, from initial application submission to final permit issuance. Each required interaction with a distinct city department or agency should be counted as a separate step, but be careful not to artificially separate closely related actions within a single process. For example, applying for and obtaining a building permit should be considered one step, not two. Be cautious not to include optional or discretionary steps, post-approval activities such as inspections during construction or certificate of occupancy issuance, steps that are only required in specific circumstances or for certain types of properties, or internal processes within departments that don't require direct developer interaction. When analyzing the ordinances, pay close attention to language indicating whether a step is mandatory (e.g., "shall", "must", "is required") versus optional or conditional (e.g., "may", "at the discretion of", "if applicable"). The goal is to identify the minimum number of distinct, mandatory steps that every multi-family building project must go through in the approval process, avoiding redundancy and over-segmentation of closely related actions.

Question Type: Numerical

Rephrased Question the LLM Sees: How many mandatory steps are involved in the approval process for a typical new multi-family building?

Question 31

Question Text That We Embed: For a typical new multi-family building project in this jurisdiction, how many distinct governing bodies or agencies must give mandatory approval before construction can begin?

Question Background and Assumptions: When answering this question, focus on the approval process for a typical new multi-family building project as described in the provided ordinance sections. Only count distinct governing bodies or agencies whose approval is explicitly required by the ordinances for all multi-family building projects, including those allowed "by right" under existing zoning. To be counted, an entity must have clear, independent approval authority that is mandatory for the project to proceed. This approval must be specifically for the multi-family project itself. Look for unambiguous language indicating required, independent approval steps. Distinguish between actual approval authority and advisory roles; entities that only review or provide input should not be counted. Consider roles like the Planning Board, Board of Health, Building Commissioner, and special permit granting authorities, but include them only if their approval is explicitly required and independent. For coordinated review processes, determine whether they represent multiple independent approvals or a single approval incorporating multiple inputs. Provide your answer as a number, followed by a brief explanation of which entities you counted and why. Cite relevant ordinance sections, explaining why each approval is considered independent and mandatory, and how it relates specifically to the multi-family project.

Question Type: Numerical

Rephrased Question the LLM Sees: For a typical new multi-family building project in this jurisdiction, how many distinct governing bodies or agencies must give mandatory approval before construction can begin?

Question 32

Question Text That We Embed: Are there townwide requirements for public hearings on any type of multi-family residential projects?

Question Background and Assumptions: When answering this question, examine the zoning ordinances and bylaws for any townwide requirements that mandate public hearings or formal public input processes for multi-family residential developments. Focus on requirements that apply across all zones within the town. Answer YES if public hearings are required for any subset of multi-family projects, even if not all multi-family projects require hearings. For instance, if larger projects require public hearings while smaller ones don't, the answer should still be YES. Requirements specific to certain zones do not count towards a YES answer. Answer NO only if there are no townwide public hearing requirements for multi-family developments of any size or type, or if such requirements only apply in specific zones. Be sure to cite relevant ordinance sections that support your conclusion. The goal is to determine whether there is any mandated opportunity for public input on new multi-family housing developments on a townwide basis, even if this only applies to certain categories of multi-family projects.

Question Type: Binary

Rephrased Question the LLM Sees: Are there townwide requirements for public hearings on any type of multi-family residential projects?

Subtask:

- Subtask Question That Gets Embedded: Do any types of multi-family housing projects require a special permit in this jurisdiction? If so, under what conditions?
- Rephrased Subtask Question the LLM Sees: What is the typical approval process for new multi-family building projects in this jurisdiction? Please describe any required permits, reviews, or other procedures that are standard for multi-family developments.
- Additional Subtask Instructions: Do any types of multi-family housing projects require a special permit in this jurisdiction? If so, under what conditions?
- How The Subtask Results Are Described to the LLM Afterwards: Special Permit Requirements for Multi-Family Housing Developments

Question 34

Question Text That We Embed: What is the maximum potential waiting time (in days) for government review of a typical new multi-family building?

Question Background and Assumptions: The review process for constructing a new multi-family building involves several stages, each of which may have a specific waiting period. The total waiting time includes the mandatory review periods as well as any discretionary days that can be added by the governing bodies or agencies. Each agency or department that a developer must interact with, such as city government departments like fire, police, sanitation, building, and planning, has its own review timeline. Additionally, discretionary days that may be required for public hearings, environmental reviews, or other procedural requirements must also be added to the total count of government review days.

Question Type: Numerical

Rephrased Question the LLM Sees: What is the maximum potential waiting time (in days) for government review of a typical new multi-family building?

Appendix D: Unincorporated Area Construction

Making an Incorporated Area Shapefile

We use the Census of Governments (CoG), specifically the Government Units Survey (GUS) subcomponent as an index for all local governments in the U.S. Local governments are sub-county governments and are categorized into either municipalities or townships. Municipalities correspond with Federal Information Processing Standards (FIPS) place codes and exist in every state. Townships correspond with FIPS county subdivision codes and exist only in the Northeast and Midwest regions of the US. To measure the geographic area of the US covered by local governments we merge all municipalities and townships with their corresponding census shapefiles by merging municipalities with FIPS place shape files and townships with FIPS county subdivision shape files. We then take the spatial union of these shape files as the incorporated area of the US. We call this the incorporated area shapefile.

Incorporated Area in 2000

We use the 2002 GUS subcomponent of the CoG as the index for local governments. We then merge these local governments with the 2000 census TIGER/line shape files. We use the 2000 shape files because they are available in the modern Shapefile format, while the 2002 TIGER/Line files are only available in ASCII formats. In practice, this means that we drop a small number of local governments that were incorporated in 2001/2002 and thus did not yet have corresponding shape files. We also miss a small number of local governments that disbanded in the two years.

Incorporated Area in 2022

We use the 2022 GUS subcomponent of the COG for the index for local governments. We use the 2022 Census shapefiles

Measuring Share Incorporated For Counties

For each county we first isolate the area that is land by taking the spatial difference of each county's shapefile with its corresponding water area shapefiles. We do this because many counties do not incorporate water areas in their domain and this area is not relevant for housing outcomes. Then, for each county we intersect its land area with the incorporated area shapefile to measure the land area of the county that is incorporated. We then take the ratio of the land area that is incorporated to the total land area of the county as the share incorporated of the county. To measure MSA incorporation share we take the land area weighted average of all counties within an MSA.

Measuring Share Incorporated For Granular Census Geographies

For both census blocks and block groups we use the same methodology to determine the share incorporated. Specifically, we intersect the corresponding shapefile for each block/block group with the incorporated area shapefile. We take the ratio of the area incorporated within the block/block group with total area of the block/block group as the share incorporated. Note that we do not exclude water area for blocks and block groups since they do not tend to include much water area.

Using Share Incorporated For Granular Census Geographies

Most blocks and block groups are either entirely inside of an incorporated area or outside of all incorporated areas, while many are partially located inside of incorporated areas. For partially incorporated blocks/block groups, we make the assumption that variables (i.e., population) are evenly dispersed across space. In other words, we assume that there are no differences in the incorporated and unincorporated portion of a partially incorporated block/block group. As a result, we proceed by dividing up partially incorporated blocks/block groups into sub-blocks and sub-block groups that are entirely incorporated and entirely unincorporated with weights of the percent incorporated. For example, if a block has a population of 1,000 and is 25% incorporated then we say that 250 people in that block live in the incorporated area and 750 people live in the unincorporated area.