# The Costs of Housing Regulation: Evidence From Generative Regulatory Measurement

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

March 21, 2024

**Abstract**

We introduce a new approach to decode and interpret statutes and administrative documents employing Large Language Models (LLMs) for data collection and analysis that we call *generative regulatory measurement*. We use this tool to construct a detailed assessment of U.S. zoning regulations. Our approach achieves 96% accuracy for binary regulatory questions and a correlation of 0.9 for predicting residential minimum lot sizes. We estimate the association of these housing regulations with housing costs and construction. Our work highlights that LLMs can be used to achieve near-human capacity in measuring and interpreting complex regulatory datasets.

JEL-Classification: R52, R58, K11, O38, R31, C81
Keywords: housing regulation, zoning codes, large language models, natural language processing, artificial intelligence, municipal ordinances

*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 Stijn Van Nieuwerburgh, Theresa Kuchler, Johannes Stroebel, and comments from seminar participants at the NYU Stern Corporate Governance luncheon, NYU Law School, and the Online Seminar in Economics and Data Science. David Dai, Amber Gao, Grace Getman, Karin Hobelsberger, Nagharjun Mariappan, Scott Overbey, and Alok Ranjan provided excellent research assistance. We gratefully acknowledge funding from Emergent Ventures and the NYU Stern Center for Global Economy and Business. 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

From the early 2000s, housing production in the United States has declined. Shortages in housing construction are associated with increased house prices, a rising housing capital share of income, and sluggish productivity growth in the real estate construction sector (Rognlie, 2016; Goolsbee and Syverson, 2023; Glaeser and Gyourko, 2018), and many studies link these changes to stringent housing regulations (Glaeser and Ward, 2009; Gyourko et al., 2008). However, current evidence remains limited in measuring housing regulations and identifying their role in affecting housing supply. A critical challenge to answering these questions lies in understanding what current housing regulations are, given the fact that every municipality has adopted a distinct set of zoning and building code regulations, and no comprehensive dataset exists to evaluate the diversity in housing regulations across the country.

Our paper argues that advances in Large language Models (LLMs) enable the systematic analysis of regulatory documents, a task which we refer to as *generative regulatory measurement*. We use state-of-the-art Artificial Intelligence (AI) methodologies to estimate zoning regulations across a large fraction of municipalities in the United States. We document that advances in LLMs enable the scalable and accurate parsing of complicated legal and regulatory texts. Our focus is creating a national database with granular information on local zoning codes, due to their importance in shaping housing markets. However, challenges in appropriately interpreting and analyzing textual databases are common across multiple domains (in building codes, other regulations, court cases, earnings call transcripts, newspapers, etc.) and so our approach also has broader applicability in suggesting possible approaches towards the analysis of such texts more broadly. Developing such approaches has become increasingly important as the quantity and complexity of regulation has risen over time (Singla, 2023).

To do so, our study answers three main questions. First, we ask how accurate are LLMs at answering questions about zoning codes. Establishing the precision of measurement is crucial to establishing whether these AI approaches can aid in systematically organizing texts for research and policy. We compare LLM-generated answers to specific questions on zoning codes conducted on human-coded fields taken from the Pioneer Institute (see Glaeser and Ward (2009)) in the state of Massachusetts, which provides an effective training dataset for our analysis. Our results suggest that LLMs deployed on the latest models (Chat GPT-4 Turbo) have achieved near-human rates of precision in classifying regulation, with an accuracy rate around 96% for binary questions. LLMs also perform strongly on numerical questions with a correlation of 0.7 between generated data and analyst responses. This includes a considerable amount of heterogeneity, and the correlation between our data and analysts exceeds 0.90 for numerical questions on the number of zoning districts, as well as an the minimum residential lot size, an important regulation governing density in single-family areas. Importantly, these accuracy statistics vary substantially based on the size and the vintage of the model as well as the degree of question processing, indicating that generative data approaches are sensitive to model quality and domain-specific training and preparation.

Second, we assess the variation in zoning regulation across the United States. We scale up our analysis by querying the housing regulation questions from Massachusetts to the entire country, thereby creating a detailed dataset of housing regulations at the municipality level.[1] This dataset fills a knowledge gap in the literature: because of the complexity of accurately classifying and understanding zoning documents, we lack a clear understanding of how precisely these housing regulations vary at a granular level across the United States. Importantly, our analysis is conducted at the municipal level, the relevant unit of local government responsible both for the construction of zoning codes, as well as in providing public goods. While some existing research has explored proxies for housing regulations commonly used at more aggregate levels,we provide both more detailed as

[1] For reasons of cost, we currently use a national model based on an older vintage of model (GPT-3.5 Turbo). Therefore we expect the national-level accuracy to improve after we deploy the latest model.

well as granular data across a large sample of municipalities in the United States. Our data covers 63% of the population who lives under a local government subject to zoning ordinances. We validate our measures by correlating our generated data against existing surveys of regulation from the Wharton index (Gyourko et al., 2021), finding some association both for specific questions asked across both the survey-based approach as well as our generated data (for affordable housing mandates and minimum lot sizes), as well as for the overall indices.

Third, we also use our dataset to make progress on the question of whether housing regulations are associated with housing costs and construction. We find that affordable housing requirements and minimum lot sizes are strongly related with measures of housing prices and rents, while measures of frontage requirements, the legality of cluster developments, unit caps, and building conversion rules are associated with permits per capita. These subtle findings point to the promise of more granular measures of regulations enabling more fine-tuned distinctions between different regulatory regimes than just more versus less strict. Obviously, these correlations could reflect selective adoption or a causal impact of these zoning rules. In future work we will attempt to disentangle these explanations.

Our results serve as an initial proof of concept towards the use of LLMs in the systematic generation of content in regulatory and legal documents, and suggest these models are a promising tool to measure and generate data in these contexts. We estimate high accuracy rates for such an approach, close to what we might expect for a non-expert but skilled human. While expert humans still have an edge in precise classification, LLM approaches bring several advantages. First, they are far more scalable than human approaches: we are able to deploy our regulatory classification measure across a sample of thousands of municipalities, a task which would be challenging if not prohibitively expensive for humans. Our approach therefore opens up the prospect of scalable and accurate regulatory classifications across multiple domains. Second, the LLM classifications also have the benefit of verifiability and auditability. We prompt the LLM to provide the precise text in the regulatory document which supports the categorization, enabling other humans (or AI agents) to verify and check the reason for classification. Third, our approach is also flexible to changes in researcher determinations of definitions and subsequent advances in AI models. Researchers using such approaches, therefore, can easily adapt and replicate models over time, accommodating those who prefer alternate specifications or more accurate models.

Importantly, our results should be seen as illustrating a base level of performance using widely accessible tools, and have considerable scope for improvement along several dimensions. We perform some refinements of question prompting,[2] question background information, and multi-step processing.[3] Further pre-processing of documents to focus LLMs on relevant text is also likely able to improve model accuracy. Additionally, we use the highest quality LLM available at the time of writing (Chat GPT-4 and Claude 3 Opus), but these models are likely to improve over time. We also plan to expand the scope of this work to examine changes in zoning codes over time, in analyzing housing regulations across countries (including in other languages), as well as in analyzing building codes in conjunction with zoning codes. Combined, the promise of these efforts suggest that LLMs are likely to fundamentally reshape our ability to understand the content and impact of regulations broadly.[4]

[2]This entails rephrasing questions for the LLM through strategies like breaking multi-part questions into different components, and breaking compound questions into individual clauses (i.e., if the question asks about whether multi-family housing is allowed either by right or through a special overlay, we ask about those two possibilities separately).

[3]Multi-step processing entails breaking a task into multiple steps and querying the LLM separately for each step.

[4]Replication code, which can be adapted to other use cases, can be found at: `https://github.com/dmilo75/ai-zoning`.

**Contributions to Literature** The central contribution of our project is the creation of a standardized, comprehensive dataset of zoning across the United States. Much of the existing literature on housing regulations has used either indirect measures or proxies for zoning regulation. The first strand of this literature has focused on survey-based approaches to measuring housing regulations. One of the most heavily used such nationwide measures of housing regulation includes the Wharton Regulatory Index (Gyourko et al., 2008, 2021; Huang and Tang, 2012). This pioneering approach to measuring housing regulations was based on surveys sent to 2,649 distinct municipalities (there are 19,488 municipalities in the United States in total), asking for information on the regulatory process, details of local land use regulations, and outcomes of the permitting and regulatory process. The survey itself builds on earlier work which surveyed a smaller number of municipalities (Mayer and Somerville, 2000), and other research has focused on surveys given to local officials and planners (Saks, 2008). We complement this survey approach through a direct measurement of housing regulations drawn from municipal regulations. Relative to surveys, this has the advantage of being comprehensive, rather than being limited by low or biased survey response rates. Our approach is also scalable and easy to augment with new questions, while surveys are inherently limited to the set of questions which were asked and which respondents are willing to answer.[5]

The second strand of this literature includes wedge-based approaches, which instead aim to impute housing regulations by examining the expected spatial macroeconomic distortions resulting from zoning. Examples in this literature 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

[5]There is also, obviously, information captured by surveys that cannot be captured in the text of municipal zoning codes. For example, perhaps certain aspects of the zoning code are never or rarely enforced, or perhaps the zoning commission never approves particular kinds of projects even if they are legally permissible.

commercial zoning impacts.

Third, other national approaches have examined textual data, but in more limited ways. Ganong and Shoag (2017) focus a scaled count of judicial decisions on "land use." While this is surely a proxy for regulatory strictness, it leaves open the question of precisely which housing regulations are driving housing litigation. 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 and tightening zoning restrictions and then analyze the effects of these changes in regulation on housing supply and rents. Our approach, by contrast, is able to establish more cleanly the precise nature of housing regulations across a broad sample of jurisdictions in the United States.

Another literature has attempted to address the limitations in national-level approaches through more detailed analysis of specific regulations at the state level. 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. For a subset of the Boston metropolitan area, they conducted a detailed analysis of local zoning codes, permitting precise calculations of potential housing supply across communities. However, the enormity of that effort prevents it from being replicated in other markets by a single research team."

We argue that the practical difficulties behind the scaling up of this approach have now been addressed through the development of modern AI LLMs, providing both the granularity of the state-based approaches along with the scale of the national regulatory studies. Indeed, the Pioneer Institute data—the most comprehensive of these state based approaches—is a crucial test for our approach. We begin our analysis by first analyzing data in Massachusetts using the same data source for municipal documents identified by the Pioneer Institute team, which allows for a cross-validation of the accuracy of our AI-led approach against the existing housing regulation classification. This serves as an important validation check of our approach. Other detailed state-level analyses of housing regulation include Shanks (2021) which also 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).

These studies leave important gaps in our understanding of housing regulations under both the national and state-level analyses. While the national approaches establish that housing regulations appear to drive important variation across the country in housing costs and construction activity, they have less to say about which specific regulations are the key drivers. Isolating specific regulatory impacts is essential for policy seeking to remedy possible impacts of regulatory driven housing cost increases. Alternatively, more detailed state-level data offers the potential to isolate the specific aspects of housing regulation that are most binding. These approaches, however, are limited in their geographic scope outside the unique states of Massachusetts, California, and Florida. Consequently, the extent to which specific housing regulations drive costs and construction activity across the country are unclear. Both line of research are also not able to contrast costs with potential benefits or amenities, making it impossible to disentangle supply and demand side effects which are crucial to establishing the cost-benefit tradeoffs of housing regulation.

Relative to this literature, our contribution is to construct a more comprehensive and detailed measure of how zoning regulations and building codes vary across the United States. We provide the most detailed assessment to date of all relevant housing regulations (i.e., minimum lot sizes, whether multifamily apartments can be constructed, inclusionary zoning mandates, setback rules, etc.) that apply to construction in local areas.

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). Verifying whether LLMs can accurately parse large legal documents—and for which questions—is therefore a crucial step towards our understanding of the capacities of these models, with the promise of opening up the large-scale use of textual documents for quantitative research. 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. This serves as a critical test case for the efficacy and reliability of LLMs in not only understanding and processing complex legal and regulatory language but also in discovering and extracting novel, actionable insights from a vast array of documents. Prior literature has used textual data to extract information, particularly sentiment, from text (Hassan et al., 2019; Romer and Romer, 2004; Tetlock, 2007; Lopez-Lira and Tang, 2023); a few papers have 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). Hoffman and Arbel (2023) argues for the use of LLMs in "generative interpretation" in estimating the meaning of legal contracts.

The central contribution of this project is to establish a solid groundwork for the ongoing application and advancement of large language models (LLMs) in the field of legal and regulatory research. By demonstrating how these advanced models can optimize data generation, improve information accessibility, and facilitate predictive analysis, we argue for incorporating LLMs into the wider research, regulatory, and policy ecosystem.

# 2 Data and Background

## 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.[6]

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 common reason to incorporate in the first place. 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.[7]

Municipalities enforce laws by issuing municipal codes which outline local regulation in different domains. Zoning codes outline permitted uses for different classes of land as well as relevant housing regulations. Some regulations apply broadly to all land within a jurisdiction; other regulations (such as minimum lot sizes) typically vary depending

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

[7] States and municipalities also enact building codes, which govern the building and safety standards that new construction needs to adhere to. In the future, we plan on using a similar approach to analyze building codes as well.

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 aggregated by different companies. 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 US that live in either a municipality or a township, we cover 63%. 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.

The primary dataset for our analysis consists of the full-text of zoning documents. 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 represents a significant advancement in terms of both accuracy and runtime over previous models like Recursive Neural Networks, which processed sequences linearly. LLMs are trained with semi-supervised learning, 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 in applying to housing regulatory textual analysis. The central advantage is scalability: we are able to load large quantities of municipal code data for classification and analysis, which far exceeds the capacity of any human team to analyze. Other advantages include the prospect for additional training, allowing for increased accuracy over time as LLMs improve in accuracy and additional training data is incorporated into the analysis.

Potential drawbacks in using LLMs for this purpose center on the inaccuracy of measurement and classification. This can happen either through limitations in the context window used to identify relevant text from the sample corpus, or the content and lack thereof of similar questions and related texts in the underlying training sample. 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 and the 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 outside the domain (i.e., in the form of state regulation not contained within our ordinance sample). We attempt to measure these drawbacks through comparison of LLM-generated output against human defined categorizations of regulation.

## 2.3 Processing Municipal Codes Using LLMs

To conduct our analysis, we use a standard framework 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.[8] Any images in the tables are transcribed using Amazon Textract. We then use text embeddings, which are vector representations of the text's semantic meaning. This enables efficient search through zoning documents. The basic intuition behind embedding is to represent words with vectors which represent a dimension in embedding space, such that words with similar semantic meaning are closer in this space. For our zoning document, this ensures that we are able to retrieve components of the document relevant for our specific questions. Different embedding algorithms conduct this task in distinct ways; we use the `text-embedding-ada-002` algorithm from OpenAI for the national sample, and a newer algorithm `text-embedding-3-large` for the testing sample comparison with Pioneer.

We similarly embed the questions we want answered from the documents, which for ease of comparison we limit to the question base already used by the Pioneer Institute (i.e., "Is multifamily zoning allowed in this area as-of-right?"). We rephrase these ques-

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

tions 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 two separate embedded vectors in hand, the zoning documents from a particular municipality and a question we would like answered, we then isolate the parts of an ordinance most relevant to answer the question. The length of typical zoning documents exceeds the context windows currently usable by LLMs, so we need to select specific sections of text that are most likely to be relevant to the question. We use cosine similarity, a standard measure of distance between two vectors, to rank sections of text by how likely they are to be relevant to the question. We then refine this ranking by using a cross-encoder reranking model[9] on the top 50 sections of text, which processes the question and section text pairs simultaneously to determine the most semantically similar sections.[10] 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 rephrased zoning question, as described above to simplify model parsing. Third, we also provide additional background information and assumptions. The background information and model assumptions were taken directly from the Pioneer study (their "Issue Overview" and "Research Coding" sections for each question) and were based on trial and error for what information was most relevant to improve model performance. Appendix B contains full information 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

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

[10]When double checking answers on select questions we instead use keyword inclusion to re-rank section text.

produce model output which is our answer. 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 under different 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" where 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. For instance, an affirmative "Yes" to a question about whether townhouses/attached housing is allowed typically means the LLM has likely 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 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.

# 3 Model Performance

## 3.1 Comparison with Pioneer Data for Benchmark Model

Performance analysis is a crucial step in validating the effectiveness and reliability of LLMs for tasks such as zoning ordinance interpretation. By comparing the accuracy of different LLM approaches against a ground truth dataset, we can assess their ability to provide consistent and correct answers to zoning-related questions. This analysis helps identify the strengths and weaknesses of each model, as well as any discrepancies between the model outputs and the reference data.

To do so requires a high-quality reference dataset. The Pioneer dataset serves as an excellent starting dataset 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 the 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) correct inaccuracies in the original classification.[11]

[11]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. We are currently expanding our error-correction process to adjust for

Table 2 shows the performance results of our baseline Chat GPT-4 Turbo model against the testing sample in Massachusetts. Among continuous questions, our generated answers have an overall correlation of 0.67 against 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.

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). There, we observe a model accuracy of 96% across all questions. Because the raw accuracy measure may be biased depending on the base rate of answers, we also provide a Relative Squared Error (RSE) which compares each model's result compared to a naive model which guesses the sample model. We observe quite small RSEs as well.

In Figure 2, we visualize the average results across questions in Table 2. In dark blue, we plot the percent correct for each model using the percent accuracy for binary variables, the correlation for continuous variables, and adjusted percent correct for categorical questions. 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).

---

these cases as well.

## 3.2 Heterogeneity Across Models

While 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 (which we currently use in our nationwide analysis). 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 accurate rate of around 80% for binary questions).[12] Interestingly, this model order is not preserved in continuous questions, for which we actually observe the highest model performance in GPT-3.5 Turbo, followed by GPT-4 and Claude 3. However, this difference is mostly driven by differences in performance on one question, the minimum lot size question, which can tend to have extreme outliers because of districts within jurisdictions with particularly large minimum lot sizes.

## 3.3 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 Chat GPT-4 Turbo disagreed with the Pioneer Institute, and present the reasons for discrepancies in a fig-

[12]This performance may reflect fundamental features of GPT-4 Turbo versus Claude, but it could also reflect the fact that we fine-tuned our prompting and chaining strategies to optimize performance on Chat GPT-4 Turbo and it is possible that if we had instead fine-tuned to maximize performance on Claude 3 that Claude 3 would have performed better.

ure. 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, and whether the answer itself was coded as incorrect but the true classification appears somewhat ambiguous.

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.

# 4 Model Validation

## 4.1 Comparison of Specific Questions Against Wharton Regulatory 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 up 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. Unfortunately, there are small nuances which do not permit a completely clean comparison. We use the Pioneer Institute wording which classifies both mandates and incentives as constituting affordable housing (question 17), while the Wharton study only considers affordable housing mandates (question 9a). For minimum lot sizes, we currently consider minimum lot sizes across all districts, while the Wharton study (question 7b) only considers residential districts, and categorizes these into four bins (whereas we use the precise minimum lot size).

Despite these limitations, we find a sizable correlation between our measure of affordable housing and the one measured in the Wharton study of 0.36. We observe smaller, but still sizable correlations, between 0.11–0.29 when examining the minimum lot size questions. Using an updated model (i.e., GPT-4 instead of GPT-3.5) as well as more closely harmonizing the precise wordings would likely improve the concordance further between the two approaches, but the current results suggest that our approach produces regulatory estimates correlated with prior work.

## 4.2 Nationwide Index Comparison

We also attempt to construct a nationwide index of our questions to better benchmark against the Wharton study. We focus on a PCA analysis to ensure greater comparability with the Wharton Index, which engages in dimension reduction across questions to provide an omnibus index consisting of the first principal component across sub-indices that group similar questions together. We similarly examine a principal components of all of our questions at the national level, finding two key principal components which appear to drive the bulk of the cross-sectional variation in zoning answers. Appendix Table 2 provides the loadings of each question on the two principal components, and Figures 4 and 5 map the two principal components across the nation.[13]

In Panel B of Table 4, we find positive correlations of both PCAs against the composite Wharton Index at the CBSA level. The first PC correlates at 0.28 against the Wharton Index, while the second PC correlates at 0.10. 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.

# 5 Nationwide Variation in Zoning Codes

In this section, we descriptively explore the rich national data on zoning codes that our generative regulatory measurement produces, characterizing the national distribution of key zoning variables and investigating their correlation with important housing outcomes.

Figures 7 and 8 show maps of minimum lot sizes and affordable housing mandates, respectively, for jurisdictions within the metropolitan areas surrounding four select cities in the United States, San Francisco, Chicago, Atlanta, and Boston. We chose these metro-areas to span all major regions and to capture a variety of policy and legal environments. Our nationwide results were produced based on the GPT-3.5 Turbo model with a simpler methodology than our preferred one described above;[14] as previously discussed, this is not the most accurate models, but is considerably cheaper to run than the full GPT-4 model. As a result, we interpret the results with caution and seek to improve the accuracy

[13] Tables 6 and 7 highlight correlations of these principal components against housing market outcomes and socioeconomic determinants.

[14] It does not use question background information provision, dividing up the prompt into subtasks, double-checking, and a few other advanced features.

rate over time. 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 mandates 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. Appendix Figure 1 also shows a heatmap of correlations between regulations at the national level.

Figure 6 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), as well as the two different zoning indices that we created using the first two principal components of our full set of zoning measures. The figure shows that these regulations vary substantially. For example, a large mass of municipalities has no minimum lot size requirement at all, while a non-trivial share of municipalities have minimum minimum lot sizes in excess of ten thousand square feet.

Table 5 shows the association of housing regulations across income and urban categories across the United States. We observe, for instance, that affordable housing mandates are found much more often in higher income and urban areas. Lot sizes appear much higher in higher income areas, but lower in urban areas—consistent with their role in suburbs as a form of "exclusionary zoning." Other categories of regulation appear surprisingly balanced across regional attributes. For instance most municipalities do not allow multi-family housing, by right or special permit, even in the most urban areas.

## 5.1 Housing Regulation and Broader Outcomes

In Figures 9 and 10, we perform initial analysis of housing regulatory fields we measure across the country using our GPT-3.5 Turbo model, correlating these different measures of housing supply measured as levels and changes of rents, house values, and building permits. Due to the limitations in the accuracy of the model and the obvious potential for selection bias, we view these results as preliminary, and include them only to illustrate the scope of analysis possible through this procedure, which we intend to further corroborate using higher-quality LLMs, as well as through quasi-experimental methods to produce stronger evidence for causation.

Nonetheless, our analysis reveals some interesting patterns of associations. Areas with affordable housing mandates are associated with regions with substantially higher rents and prices, consistent with these regulations being clustered in more expensive housing markets, but are associated with less construction. Allowing certain housing types by right (cluster developments, planned unit developments, open space residential designs, or other types of flexible zoning) are associated with higher development, though also higher rents and prices. Median house values and gross rents are associated with higher lot sizes. Caps on residential permits are also associated with less construction. To be sure, our correlations consistent either with a causal impact on supply, or are the product of selection. Both possibilities are potentially interesting, highlighting either the impacts of housing regulation on other outcomes, or the differential adoption of housing regulation by different areas. Future work will work to better tease out these implications using more accurate models and empirical designs.

# 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. The 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. Our approach also correlates with existing measures of regulation from the Wharton Index. This generative regulatory measurement approach enables the creation of a comprehensive, nationwide dataset of municipal zoning regulations.

The 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, and across different countries and languages.

The analyses reveal substantial variation in zoning stringency both across and within metro areas. Measures such as minimum lot sizes, affordable housing mandates, and restrictions on housing types correlate with higher housing costs and lower rates of new construction, though establishing causality requires further research.

We make all collected data and the associated replication code publicly available. This open-access approach ensures that the wider research community can benefit from the project's findings, use the datasets for various analytical purposes, and even apply the LLM-based classification methodology to explore other regulatory domains.

While there are limitations to current LLM accuracy and the preliminary nature of some results, this study marks an important step forward in the application of AI to decoding complex regulation. It opens up new avenues for systematically understanding the extent and impacts of municipal zoning in shaping housing market outcomes. Continued progress in this research agenda can help inform evidence-based reforms to exclusionary and costly housing regulations.

# References


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

**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. 1

**Baum-Snow, Nathaniel and L Han**, "The microgeography of housing supply," 2019. Work in progress, University of Toronto. 7

**Bybee, J Leland**, "The ghost in the machine: Generating beliefs with large language models," Technical Report, Technical report, Working Paper 2023. 1

**Duranton, Gilles and Diego Puga**, "Urban growth and its aggregate implications," Technical Report, National Bureau of Economic Research 2019. 1

**Ganong, Peter and Daniel Shoag**, "Why has regional income convergence in the US declined?," *Journal of Urban Economics*, 2017, *102*, 76–90. 1

**Giesecke, Oliver**, "State Government Debt Structure," *Available at SSRN*, 2023. 1

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

**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. 1, 1

**_ , 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. 1

**Goolsbee, Austan and Chad Syverson**, "The Strange and Awful Path of Productivity in the US Construction Sector," Technical Report, National Bureau of Economic Research 2023. 1

**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. 1, 1

**_, 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. 1, 1, 4.1

**Hansen, Anne Lundgaard and Sophia Kazinnik**, "Can ChatGPT Decipher Fedspeak?," *Available at SSRN*, 2023. 1

**Hassan, Tarek A, Stephan Hollander, Laurence Van Lent, and Ahmed Tahoun**, "Firm-level political risk: Measurement and effects," *The Quarterly Journal of Economics*, 2019, *134* (4), 2135–2202. 1

**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. 1

**Hoffman, David A and Yonathan A Arbel**, "Generative interpretation," *Available at SSRN*, 2023. 1

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

**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. 1

**Ihlanfeldt, Keith R**, "The effect of land use regulation on housing and land prices," *Journal of Urban Economics*, 2007, *61* (3), 420–435. 1

**Jackson, Kristoffer**, "Do land use regulations stifle residential development? Evidence from California cities," *Journal of Urban Economics*, 2016, *91*, 45–56. 1

**Jha, Manish, Jialin Qian, Michael Weber, and Baozhong Yang**, "Chatgpt and corporate policies," *Chicago Booth Research Paper*, 2023, (23-15). 1

**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. 2.3

**Lopez-Lira, Alejandro and Yuehua Tang**, "Can chatgpt forecast stock price movements? return predictability and large language models," *arXiv preprint arXiv:2304.07619*, 2023. 1

**Mayer, Christopher J and C Tsuriel Somerville**, "Land use regulation and new construction," *Regional Science and Urban Economics*, 2000, *30* (6), 639–662. 1

**Quigley, John M and Steven Raphael**, "Regulation and the high cost of housing in California," *American Economic Review*, 2005, *95* (2), 323–328. 1

**Rognlie, Matthew**, "Deciphering the fall and rise in the net capital share: accumulation or scarcity?," *Brookings Papers on Economic Activity*, 2016, *2015* (1), 1–69. 1

**Romer, Christina D and David H Romer**, "A new measure of monetary shocks: Derivation and implications," *American economic review*, 2004, *94* (4), 1055–1084. 1

**Saks, Raven E**, "Job creation and housing construction: Constraints on metropolitan area employment growth," *Journal of Urban Economics*, 2008, *64* (1), 178–195. 1

**Shanks, Brendan**, "Land Use Regulations and Housing Development," 2021. manuscript. 1

**Singla, Shikhar**, "Regulatory costs and market power," Technical Report 2023. 1

**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. 1

**Tetlock, Paul C**, "Giving content to investor sentiment: The role of media in the stock market," *The Journal of finance*, 2007, *62* (3), 1139–1168. 1

**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*. 2.2

**Yang, Stephen**, "Predictive Patentomics: Forecasting Innovation Success and Valuation with ChatGPT," *Available at SSRN 4482536*, 2023. 1

**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. 1

# Tables

Table 1: Sample Coverage Metrics

**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 |

Note: 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 certain 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, 2022 State-Level Census Population Data for census region and national population, and 2022 MSA-Level Census Population for MSA population.
Links to data sources are American Legal Publishing, Municode, and Ordinance.com.

Table 2: Performance Results of Chat GPT-4 Turbo on Testing Sample of 30 Municipalities

**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 |

Note: We calculate performance metrics and sample means (for RSE) only on the set of question municipality pairs that Chat GPT-4 Turbo does not say "I don't know". For Relative Squared Error we compare the model's results to the naive model that guesses the sample mean. The correlation column is simply the correlation between the model answer and the Pioneer Institute answer. 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 multi-family housing allowed, either by right or special permit (including through overlays or cluster zoning)? | 0.00 | 100% |
| Are apartments above commercial (mixed use) allowed in any district? | 0.07 | 96% |
| Is multi-family housing listed as allowed through conversion (of either single family houses or non residential 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 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% |

Note: For Relative Squared Error 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 that the model matches the adjusted Pioneer Institute answer for each question.

Table 3: Confusion Matrix For Chat GPT-4 Turbo

| 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 |

Note: This confusion matrix is generated using the Chat GPT-4 Turbo model on the testing sample of 30 municipalities from the Pioneer study. Observations where the model responds "I don’t know" or observations we categorized as ambiguous are excluded. 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.

Table 4: Relationship between Our PCA-derived Indices and Wharton Residential Land Use Regulatory Index

**Panel A: Averages and Correlation For Individual Questions**

| | | Wharton Average | Our Average | Correlation |
|---|---|---|---|---|
| Affordable Housing | | 0.20 | 0.24 | 0.36 |
| Minimum Lot Size | Less than 1/2 acre | 0.48 | 0.39 | 0.29 |
| | 1/2 to 1 acre | 0.16 | 0.10 | 0.11 |
| | 1 to under 2 acres | 0.13 | 0.15 | 0.16 |
| | 2 acres or more | 0.23 | 0.19 | 0.22 |

**Panel B: Correlation Matrix of Index**

| | Wharton Index | Our Index PC 1 | Our Index PC 2 |
|---|---|---|---|
| Wharton Index | 1.00 | 0.28 | 0.10 |
| Our Index PC 1 | 0.28 | 1.00 | 0.05 |
| Our Index PC 2 | 0.10 | 0.05 | 1.00 |

Note: The sample overlap between our study and the Wharton study (2018 version) is 1,283 municipalities. The question on affordable housing in our study (question 17 from the Pioneer study) considers both mandates and incentives, whereas the Wharton study (question 9a) only considers affordable housing mandates. For minimum lot sizes, our study considers minimum lot sizes across all districts, while the Wharton study (question 7b) only considers residential districts. We drop municipalities that do not have any minimum lot size requirements. We follow the Wharton methodology to aggregate our index to the CBSA level by taking a simple average of all municipalities in that CBSA (only those that are in both our and the Wharton dataset).

Table 5: 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? | 12 | 16 | 4825 | 12 | 12 | 13 | 9 | 13 | 11 |
| What is the longest frontage requirement for single family residential development in any district? | 109 | 84 | 4460 | 82 | 87 | 157 | 106 | 108 | 83 |
| Mean of Min Lot Sizes (Square Feet) | 24401 | 24632 | 4774 | 15240 | 21710 | 35316 | 20044 | 26260 | 17740 |
| Min of Min Lot Sizes (Square Feet) | 6197 | 3295 | 4774 | 3951 | 5023 | 9339 | 6280 | 6348 | 4863 |

**Panel B: Binary Questions**

| Question | National | | | Income Tercile | | | Urban/Rural | | |
|---|---|---|---|---|---|---|---|---|---|
| | Mean | Weight | Count | Low | Mid | High | Rural | Mix | Urban |
| Is multi-family housing allowed, either by right or special permit (including through overlays or cluster zoning)? | 54 | 53 | 5659 | 58 | 56 | 49 | 55 | 54 | 50 |
| Are apartments above commercial (mixed use) allowed in any district? | 40 | 49 | 5553 | 41 | 42 | 37 | 34 | 42 | 39 |
| Is multi-family housing listed as allowed through conversion (of either single family houses or non residential buildings)? | 17 | 14 | 5651 | 14 | 16 | 19 | 15 | 17 | 19 |
| Are attached single family houses (townhouses, 3+ units) listed as an allowed use (by right or special permit)? | 50 | 51 | 5530 | 50 | 52 | 48 | 44 | 50 | 52 |
| Does zoning include any provisions for housing that is restricted by age? | 31 | 44 | 5677 | 21 | 29 | 44 | 16 | 35 | 39 |
| Are accessory or in-law apartments allowed (by right or special permit) in any district? | 34 | 43 | 4864 | 26 | 34 | 41 | 27 | 37 | 23 |
| Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by right? | 11 | 12 | 5493 | 10 | 11 | 11 | 7 | 12 | 8 |
| Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by special permit? | 87 | 90 | 5402 | 84 | 88 | 88 | 81 | 89 | 78 |
| Does the zoning bylaw/ordinance include any mandates or incentives for development of affordable units? | 20 | 46 | 5656 | 9 | 17 | 34 | 7 | 24 | 19 |
| Is there a town-wide annual or biannual cap on residential permits issued, and/or is project phasing required? | 31 | 35 | 5690 | 27 | 32 | 34 | 24 | 33 | 29 |
| Are there restrictions on counting wetlands, sloped land or easements in lot size calculations? | 26 | 23 | 5664 | 20 | 24 | 34 | 23 | 27 | 27 |

Note: We define the count (sample size) as the number of municipalities where the model (GPT-3.5 Turbo) 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. From the 2021 Five-Year American Community Survey we use median household income (`B19013_001E`).

Table 6: Correlation of Zoning Index (Principal Component 1) with Housing Market Outcomes and Socioeconomic Determinants

**Panel A: Socioeconomic Determinants**

| | **Bivariate** | **Income** | **All** |
|---|---|---|---|
| Median Household Income (2021) | 0.21*** | 0.20*** | 0.18*** |
| | (0.02) | (0.02) | (0.02) |
| % Change, 2021-2010 Median Household Income | 0.08*** | 0.04** | 0.05*** |
| | (0.02) | (0.02) | (0.02) |
| % Urban | 0.14*** | | 0.12*** |
| | (0.02) | | (0.02) |
| Intercept | | -0.03 | -0.03* |
| | | (0.02) | (0.02) |
| R-squared | | 0.03 | 0.03 |
| N | | 5532 | 5531 |

**Panel B: Housing Market Outcomes**

| | **Bivariate** | **Price** | **Permits** | **All** | |
|---|---|---|---|---|---|
| Median Home Value (2021) | 0.21*** | 0.15*** | | 0.05 | 0.05 |
| | (0.02) | (0.03) | | (0.03) | (0.05) |
| Median Gross Rent (2021) | 0.31*** | 0.21*** | | 0.23*** | 0.14*** |
| | (0.02) | (0.03) | | (0.03) | (0.05) |
| Median Home Value % Change, 2021-2010 | -0.04** | -0.08*** | | -0.03 | 0.06 |
| | (0.02) | (0.02) | | (0.02) | (0.04) |
| Median Gross Rent % Change, 2021-2010 | 0.08*** | -0.01 | | -0.00 | 0.04 |
| | (0.02) | (0.02) | | (0.02) | (0.03) |
| Building Permits Single Units 2021 | -0.03 | | -0.05*** | -0.04** | -0.01 |
| | (0.02) | | (0.01) | (0.02) | (0.03) |
| Building Permits Multi Units 2021 | 0.09*** | | 0.09*** | 0.07*** | 0.03 |
| | (0.02) | | (0.02) | (0.02) | (0.03) |
| Building Permits All Units 2021 | 0.02 | | 0.01 | 0.01 | 0.01 |
| | (0.02) | | (0.01) | (0.01) | (0.02) |
| Housing Elasticity | -0.23*** | | | | -0.17*** |
| | (0.03) | | | | (0.04) |
| Intercept | | -0.01 | -0.13*** | -0.10*** | 0.05 |
| | | (0.02) | (0.02) | (0.02) | (0.03) |
| R-squared | | 0.07 | 0.01 | 0.05 | 0.07 |
| N | | 5379 | 4202 | 4071 | 1577 |

Note: All right-hand side variables are in Z-scores. Asterisks denote significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are shown in parentheses. We calculate the percent of each municipality that overlaps with urban areas based on the percent overlap between the 2022 shape file for the municipality and the 2020 shape file for urban areas. Median household income data is obtained from the 2021 Five-Year American Community Survey (B19013_001E). Median gross rent data is obtained from the 2021 and 2010 Five-Year American Community Surveys, using the median gross rent variable (B25064_001E). Median home value data is also sourced from the 2021 and 2010 Five-Year American Community Surveys, using the median home value variable (B25077_001E). Building permits data is obtained from the 2022 Census Building Permits Survey, using the estimated number of units permitted in 2022. Multi-Unit includes any building with 2 or more units.

Table 7: Correlation of Zoning Index (Principal Component 2) with Housing Market Outcomes and Socioeconomic Determinants

**Panel A: Socioeconomic Determinants**

| | **Bivariate** | **Income** | **All** |
|---|---|---|---|
| Median Household Income (2021) | 0.28*** | 0.30*** | 0.32*** |
| | (0.02) | (0.02) | (0.02) |
| % Change, 2021-2010 Median Household Income | -0.02 | -0.08*** | -0.09*** |
| | (0.02) | (0.02) | (0.02) |
| % Urban | -0.07*** | | -0.11*** |
| | (0.02) | | (0.02) |
| Intercept | | -0.03* | -0.02 |
| | | (0.02) | (0.02) |
| R-squared | | 0.06 | 0.07 |
| N | | 5532 | 5531 |

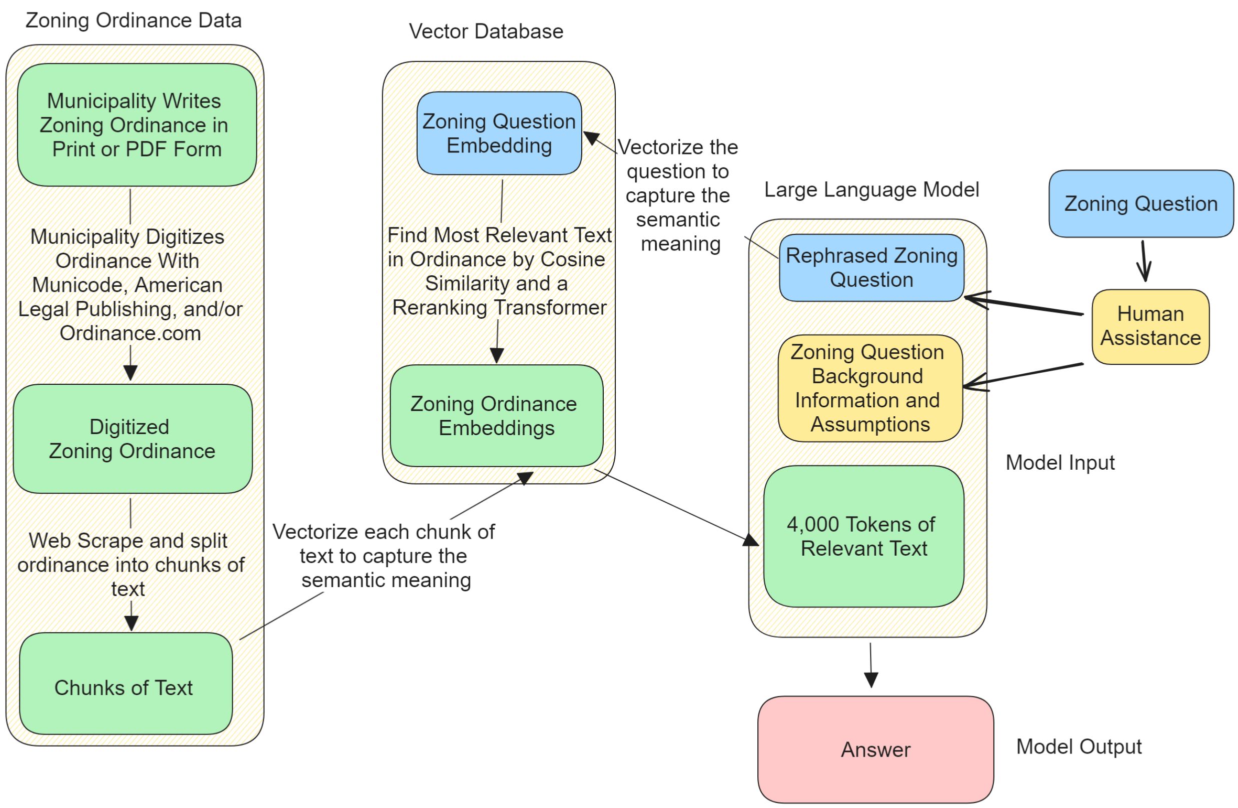
**Panel B: Housing Market Outcomes**

| | **Bivariate** | **Price** | **Permits** | **All** | |
|---|---|---|---|---|---|
| Median Home Value (2021) | 0.25*** | 0.09*** | | 0.05* | 0.05 |
| | (0.02) | (0.03) | | (0.03) | (0.04) |
| Median Gross Rent (2021) | 0.23*** | 0.19*** | | 0.17*** | 0.19*** |
| | (0.02) | (0.03) | | (0.03) | (0.05) |
| Median Home Value % Change, 2021-2010 | 0.00 | -0.04** | | 0.02 | -0.01 |
| | (0.02) | (0.02) | | (0.02) | (0.03) |
| Median Gross Rent % Change, 2021-2010 | 0.04** | -0.04** | | -0.03 | -0.01 |
| | (0.02) | (0.02) | | (0.02) | (0.03) |
| Building Permits Single Units 2021 | 0.04** | | 0.02* | 0.02 | 0.04 |
| | (0.02) | | (0.01) | (0.01) | (0.03) |
| Building Permits Multi Units 2021 | 0.02 | | 0.00 | 0.01 | 0.07** |
| | (0.02) | | (0.02) | (0.01) | (0.03) |
| Building Permits All Units 2021 | 0.04** | | 0.02** | 0.02* | 0.07*** |
| | (0.02) | | (0.01) | (0.01) | (0.02) |
| Housing Elasticity | 0.04 | | | | 0.11*** |
| | (0.03) | | | | (0.03) |
| Intercept | | -0.05*** | -0.13*** | -0.15*** | -0.16*** |
| | | (0.02) | (0.02) | (0.02) | (0.03) |
| R-squared | | 0.04 | 0.00 | 0.04 | 0.06 |
| N | | 5379 | 4202 | 4071 | 1577 |

Note: All right-hand side variables are in Z-scores. Asterisks denote significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors are shown in parentheses. We calculate the percent of each municipality that overlaps with urban areas based on the percent overlap between the 2022 shape file for the municipality and the 2020 shape file for urban areas. Median household income data is obtained from the 2021 Five-Year American Community Survey (`B19013_001E`). Median gross rent data is obtained from the 2021 and 2010 Five-Year American Community Surveys, using the median gross rent variable (B25064_001E). Median home value data is also sourced from the 2021 and 2010 Five-Year American Community Surveys, using the median home value variable (B25077_001E). Building permits data is obtained from the 2022 Census Building Permits Survey, using the estimated number of units permitted in 2022. Multi-Unit includes any building with 2 or more units. Housing supply elasticities are sourced from Baum-Snow and Han (2019), which estimates elasticities using a finite mixture model approach with a quadratic specification and 2001 tract developed fraction measures. Specifically, we use the variable (`gamma01b_units_FMM`).

# Figures

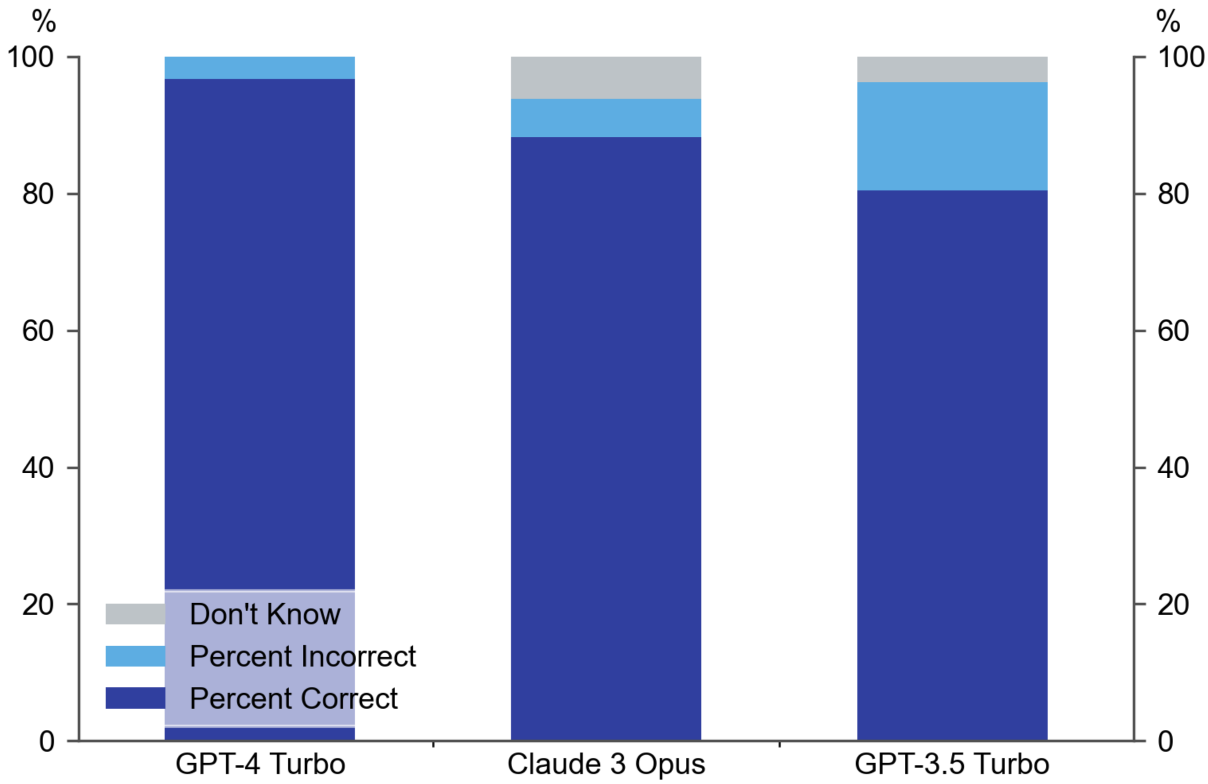
## Figure 1: Model Overview



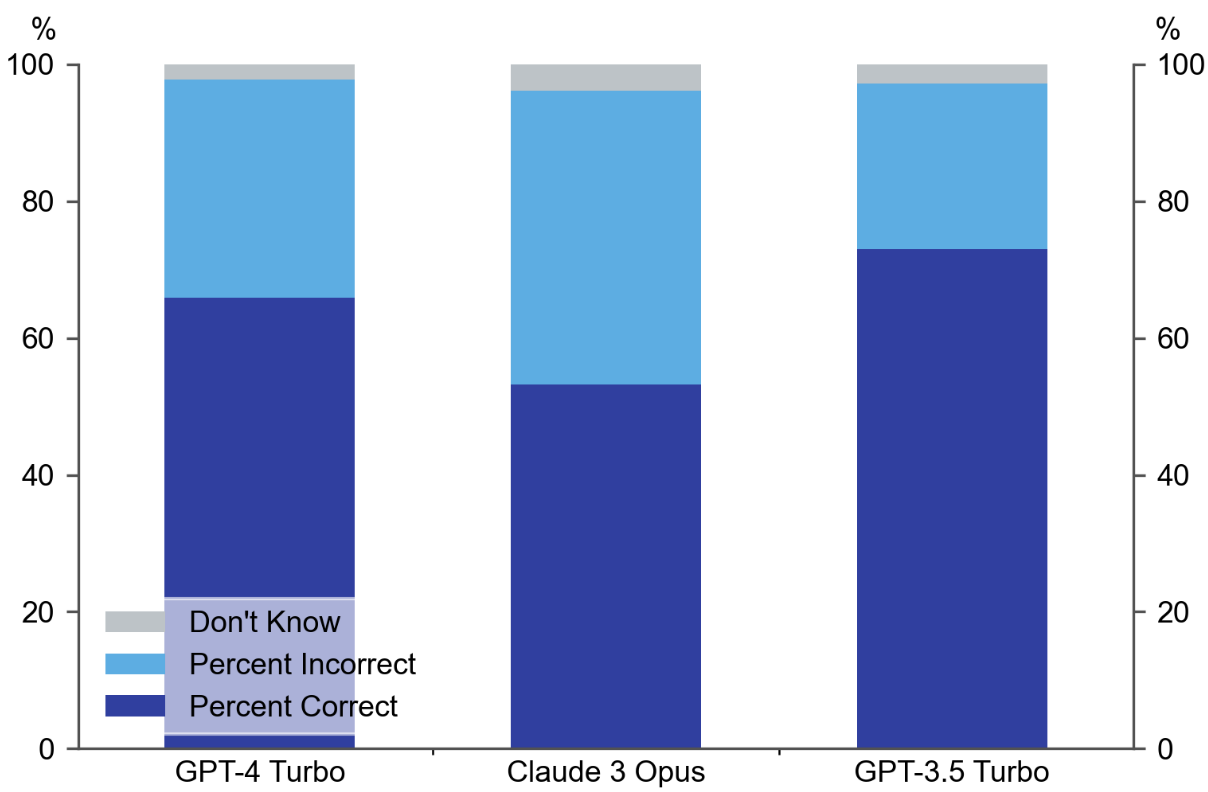

Note: We scrape each section within a zoning ordinance separately. We split up sections that are longer than one thousand tokens into chunks of at most one thousand tokens. We also combine adjacent sections of less than 50 tokens. So, each section of text varies in length but is between 50 and one thousand tokens. We vectorize each chunk of text using OpenAI embeddings models (link). For the national run, we used 'text-embedding-ada-002', and for the validation exercises in Massachusetts, we used the newer and more powerful 'text-embedding-ada-002'. 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 through key phrases, table headers, and zoning district names (i.e., R-1 for the first residential zoning district).

Figure 2: Comparison of Average Performance Across Models

Panel A: Binary Questions

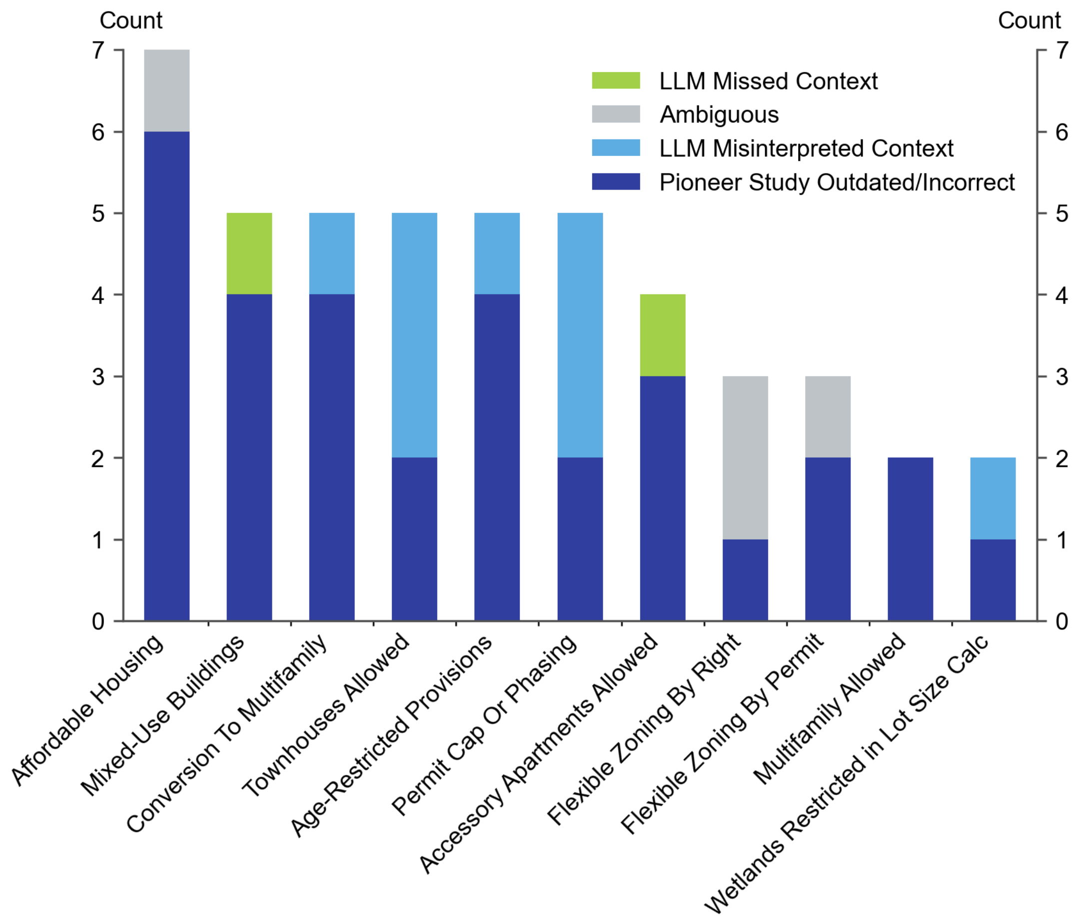



Panel B: Continuous Questions



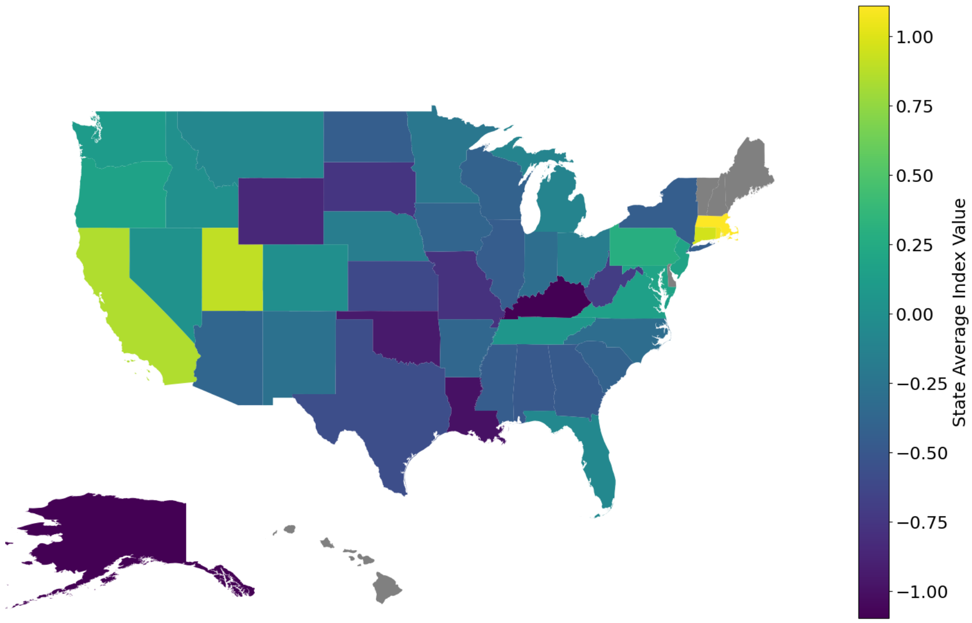

Note: For binary questions we use the percent accuracy and for continuous questions we use the correlation. We drop four question-muni pairs, which we manually categorized as ambiguous answers.

Figure 3: Reasons For Disagreement Between Chat GPT-4 Turbo and Pioneer Study on Binary Questions



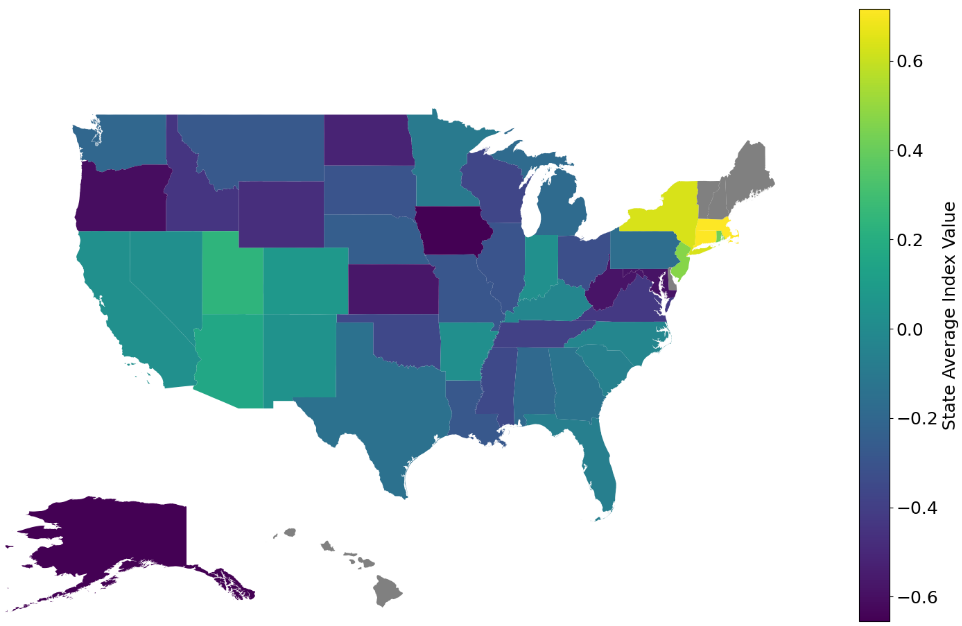

Note: We first ran ChatGPT-4 Turbo on the testing sample of 30 randomly selected municipalities that were included in the Pioneer Institute's study but were not used to train our model. We then identified the binary questions where the model responses disagreed with the Pioneer study. A law student reviewed each of these disagreements individually to determine the reason for the discrepancy, classifying them into the categories shown in the chart.

Figure 4: Nationwide Map of Zoning Index (First Principal Component)



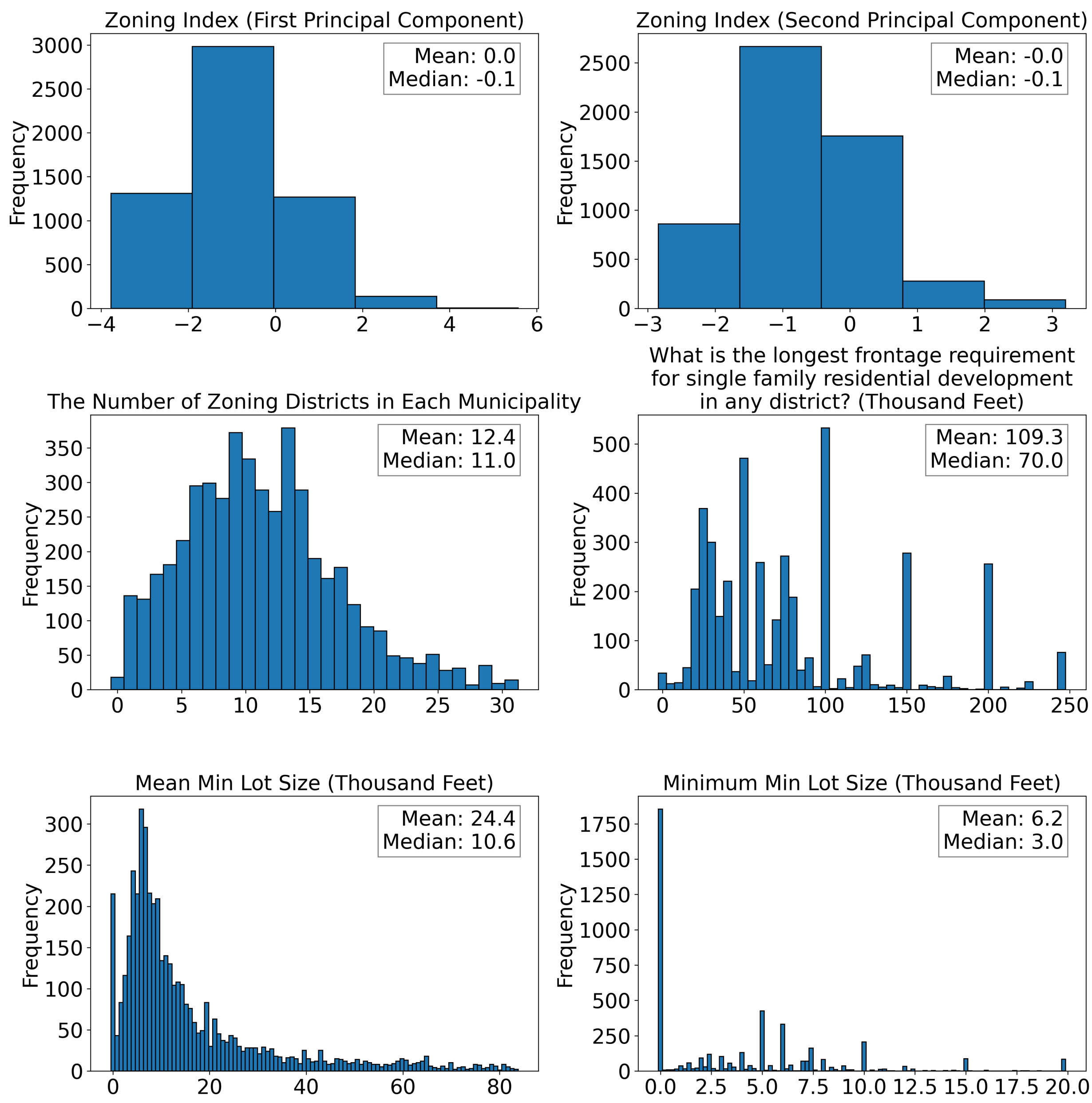

Note: The state-level zoning index value is calculated as the simple average of the index values for all municipalities and townships with available data in our dataset for each state. States shaded in grey have fewer than 10 observations and their index values are not plotted

Figure 5: Nationwide Map of Zoning Index (Second Principal Component)



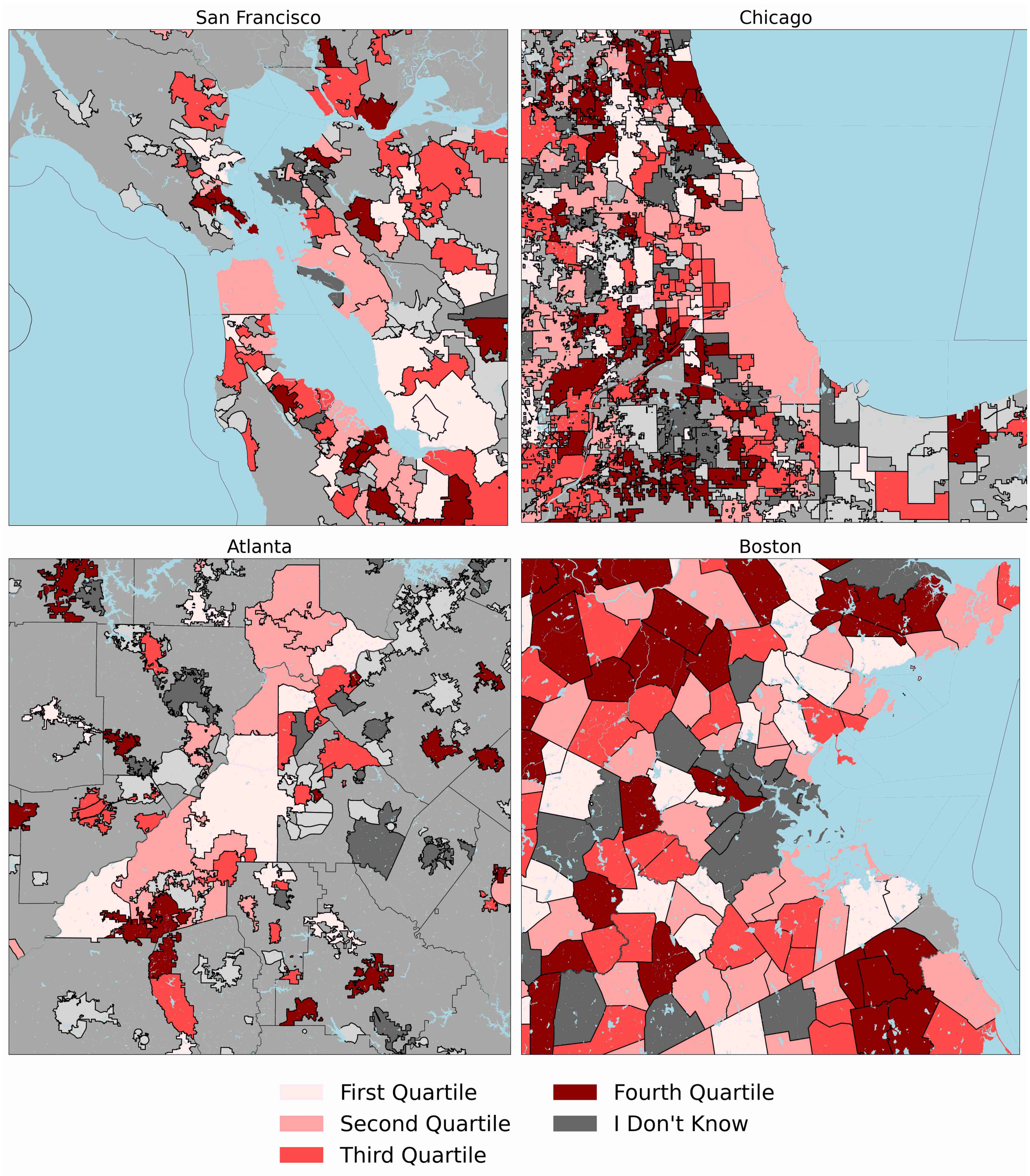

Note: The state-level zoning index value is calculated as the simple average of the index values for all municipalities and townships with available data in our dataset for each state. States shaded in grey have fewer than 10 observations and their index values are not plotted

Figure 6: Distribution of Zoning Indices and Housing Regulations



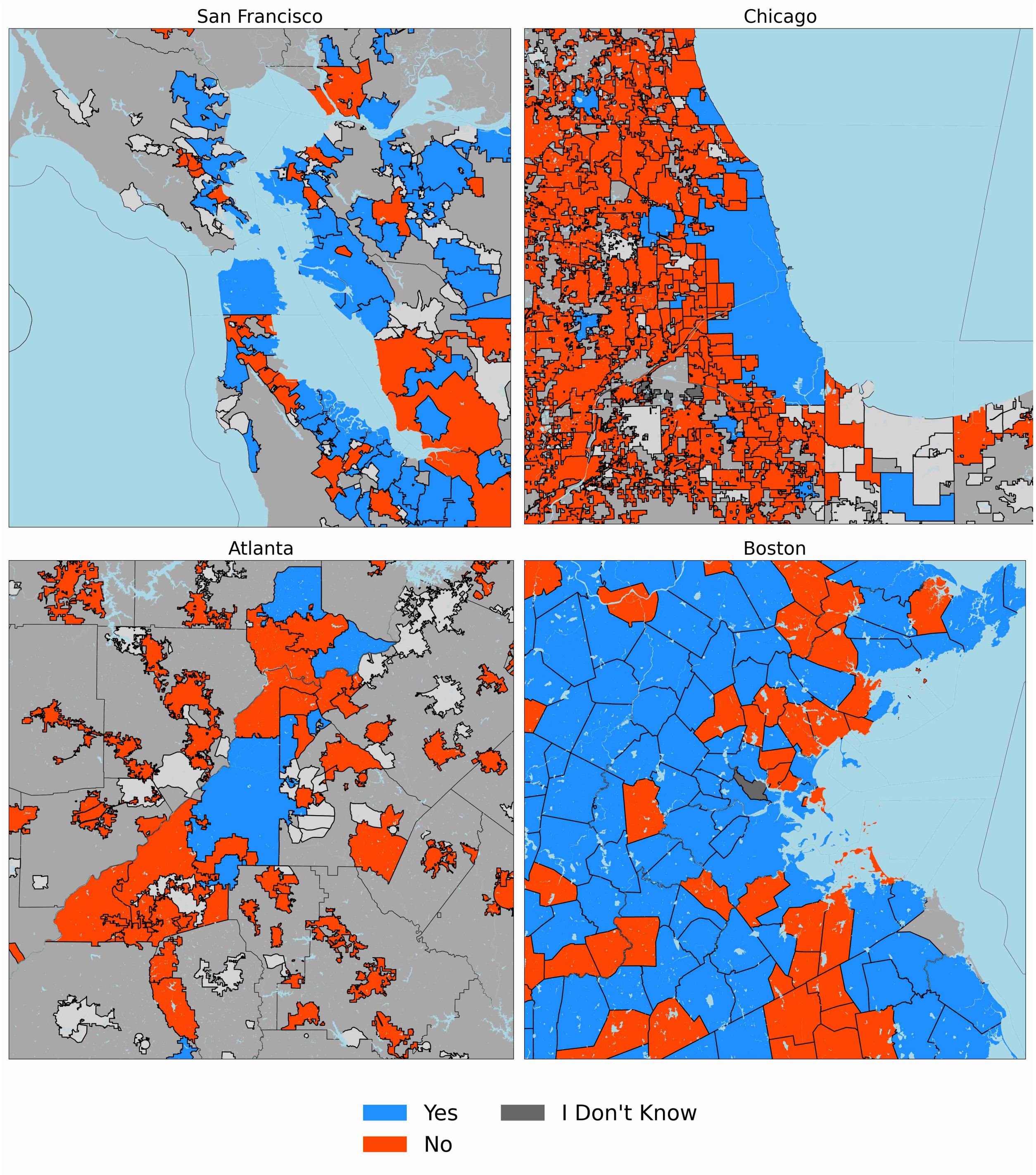

Note: See table 5 footnote for details on the sample. We cut the x-axis at the 99th percentile for the number of districts as well as the second principal component zoning index and at the 95th percentile for the minimum lot size and frontage questions. Mean and median include all data.

Figure 7: Minimum Minimum Lot Size Quartiles For Select Metropolitan Areas



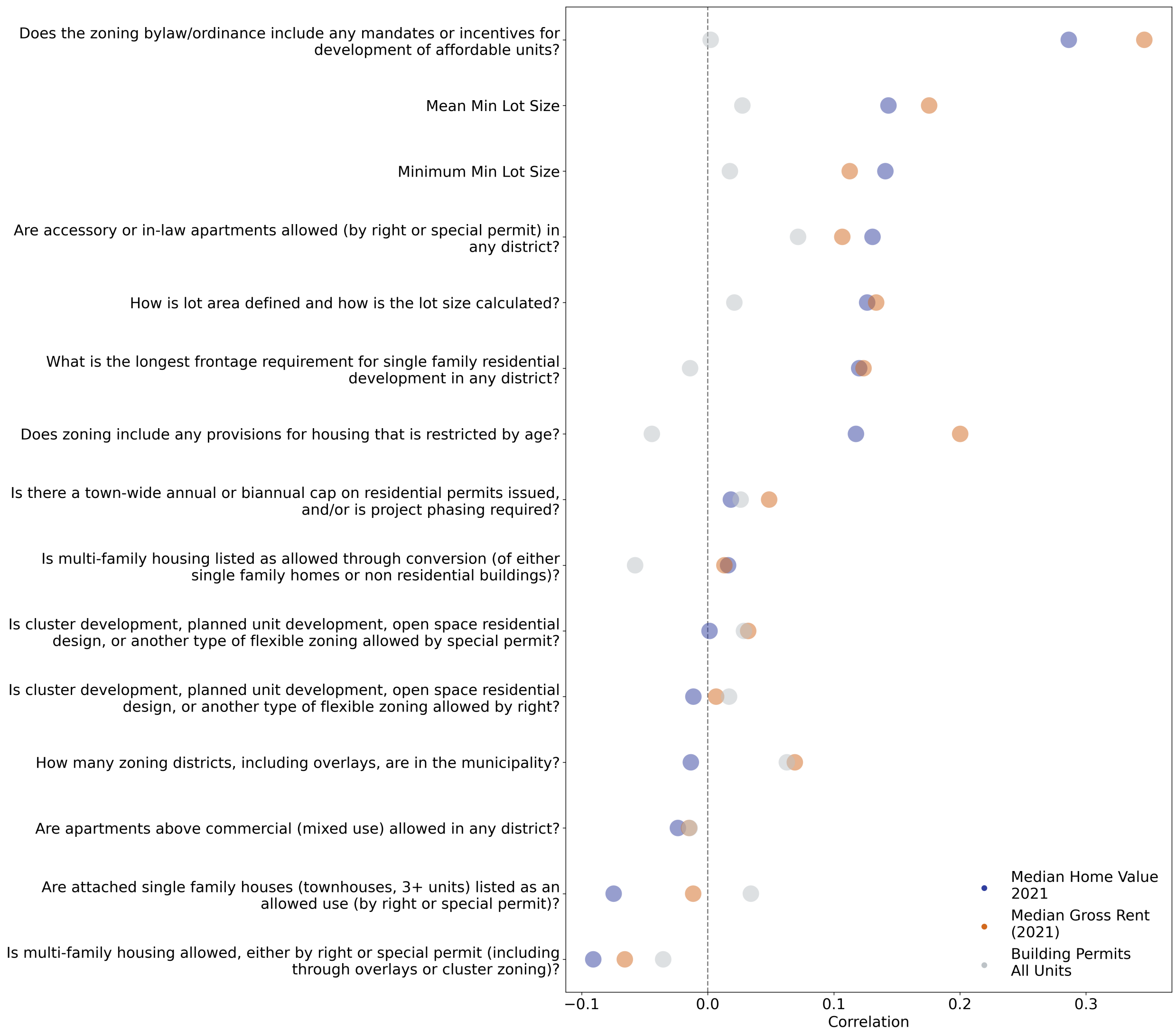

Note: Each map shows roughly a 100km × 100km square area, except for Boston where we show a 75km × 75km square area. Within each map we plot all Census-designated places, except for Boston where we also plot Census county subdivisions that correspond with townships. Both Census-designated place and Census county subdivisions data comes from the 2022 Census TIGER/Line shape files.

Figure 8: Whether There Are Mandates or Incentives For The Development of Affordable Units in Select Metropolitan Areas



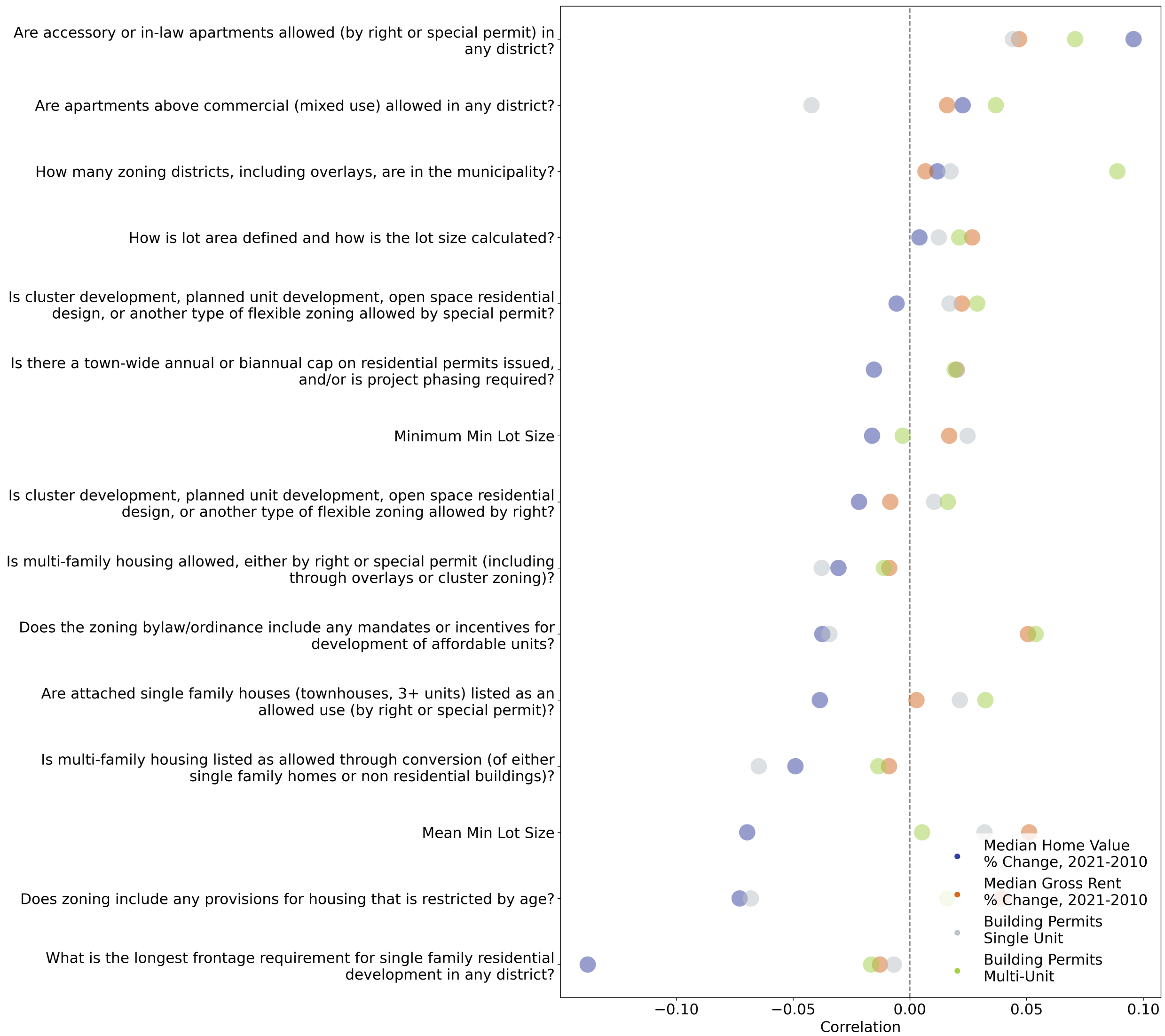

Note: Each map shows roughly a 100km x 100km square area, except for Boston where we show a 75km $\times$ 75km square area. Within each map we plot all Census-designated places, except for Boston where we also plot Census county subdivisions that correspond with townships. Both Census-designated place and Census county subdivisions data comes from the 2022 Census TIGER/Line shape files.

Figure 9: Correlation Between Median Gross Rents, Median Home Values, Building Permits Per Capita and Zoning Regulations




Note: Univariate correlations are calculated over all valid municipality question pairs (i.e. where the model does not say "I don't know") with a valid outcome variable (i.e. not missing) over our national sample with GPT-3.5 Turbo. We winsorize continuous variable answers from our model at the 5% level, but do not winsorize housing outcomes data. Median Gross Rent data comes from both the 2021 and 2010 Five-Year American Community Surveys we use median gross rent (B25064_001E). Median Home Value data comes from both the 2021 and 2010 Five-Year American Community Surveys we use median home value (B25077_001E). Building permits data comes from the 2022 Census Building Permits Survey we use the estimated number of units permitted in 2022. Multi-Unit covers any building with 2-units or more.

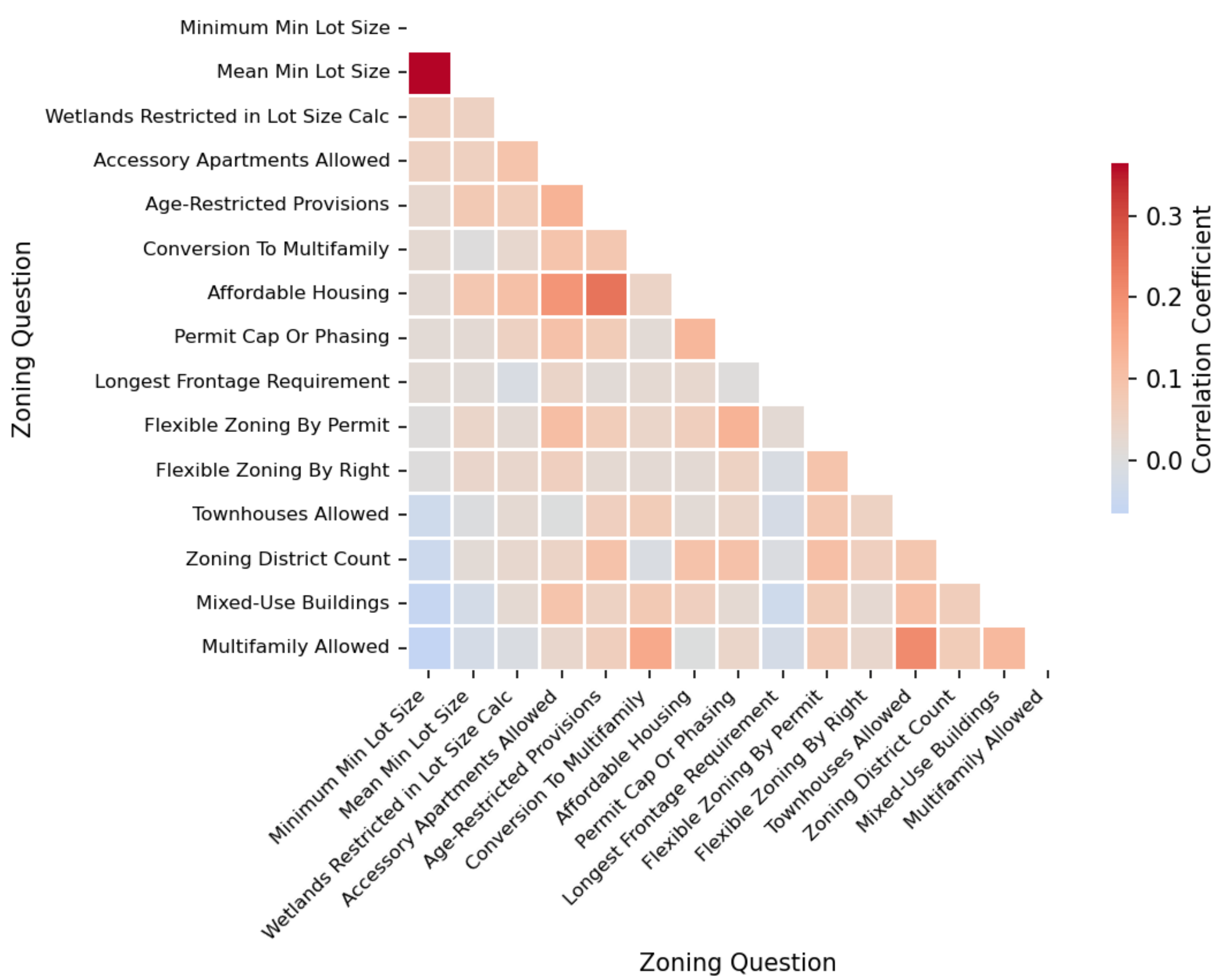
Figure 10: Correlations Between Changes in Median Gross Rents, Changes in Median Home Value, Single-Family Building Permits, Multi-Family Building Permit Units and Zoning Regulations




Note: Univariate correlations are calculated over all valid municipality question pairs (i.e. where the model does not say "I don't know") with a valid outcome variable (i.e. not missing) over our national sample with GPT-3.5 Turbo. We winsorize continuous variable answers from our model at the 5% level, but do not winsorize housing outcomes data. Median Gross Rent data comes from both the 2021 and 2010 Five-Year American Community Surveys we use median gross rent (B25064_001E). Median Home Value data comes from both the 2021 and 2010 Five-Year American Community Surveys we use median home value (B25077_001E). Building permits data comes from the 2022 Census Building Permits Survey we use the estimated number of units permitted in 2022. Multi-Unit covers any building with 2-units or more.

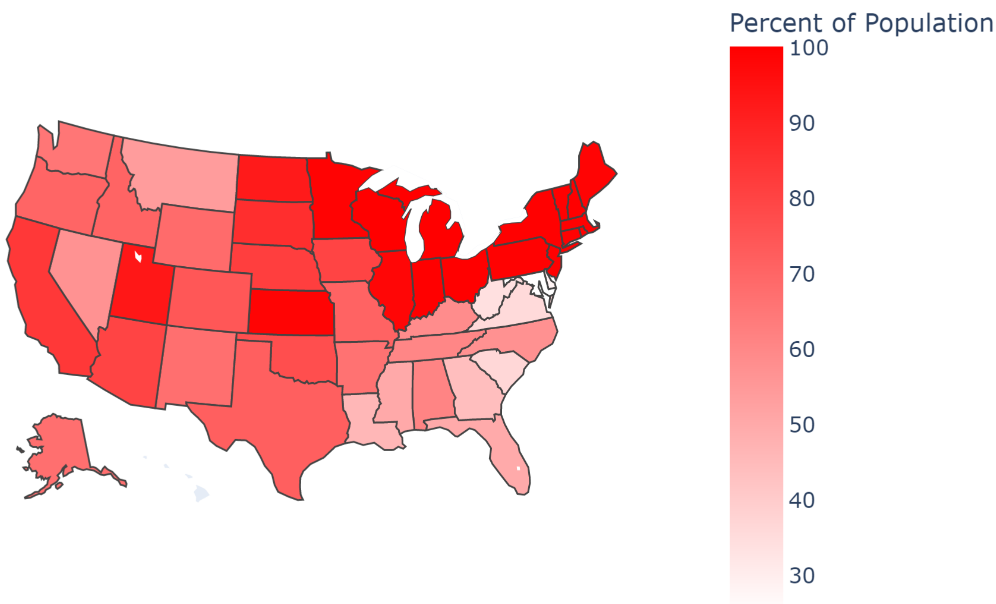
# A Data Appendix

Figure 1: Heatmap of Pairwise Correlations Between Zoning Questions



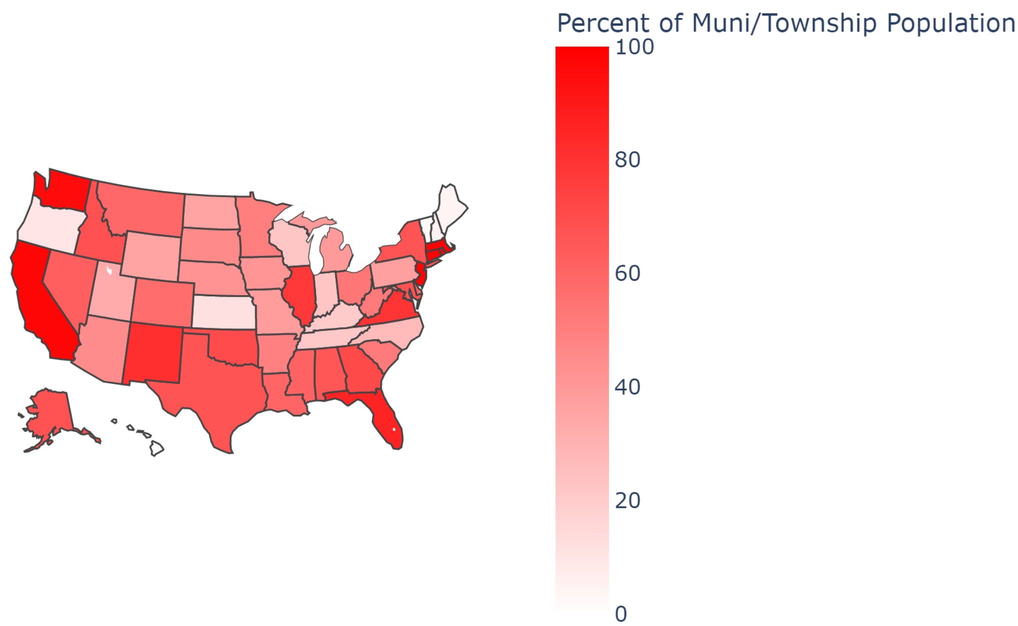

Note: Please see Appendix Table 1 for full question names. We drop observations where the model says "I don't know".

Figure 2: Percent of the Population Living in Either a Municipality or Township Government By State




Note: See Table 1 footnote for more details on sample coverage

Figure 3: Our Sample Percent of Coverage of Population That Lives Under a Municipality or Township By State




Note: See Table 1 footnote for more details on sample coverage

Table 1: Mapping of Full Pioneer Institute Study Questions to Short Names

| Full Question | Short Question |
|---|---|
| How many zoning districts, including overlays, are in the municipality? | Zoning District Count |
| 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 houses 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 Housing |
| 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 longest frontage requirement for single family residential development in any district? | Longest Frontage Requirement |
| Mean of Min Lot Sizes (Square Feet) | Mean Min Lot Size |
| Minimum of Min Lot Sizes (Square Feet) | Minimum Min Lot Size |

Note: "Full Question" refers to how each question was phrased in the Pioneer Institute study and "Short Question" refers to how we abbreviate the question in parts of the paper. Note that the Pioneer Institute study drew on external sources for information on minimum lot sizes, we create additional questions to mimic those variables.

Table 2: PCA Loadings

| | First PC | Second PC |
|---|---|---|
| Does the zoning bylaw/ordinance include any mandates or incentives for development of affordable units? | 0.38 | 0.16 |
| Does zoning include any provisions for housing that is restricted by age? | 0.38 | 0.09 |
| Is there a town-wide annual or biannual cap on residential permits issued, and/or is project phasing required? | 0.28 | 0.02 |
| Are there restrictions on counting wetlands, sloped land or easements in lot size calculations? | 0.21 | 0.15 |
| Mean of Min Lot Sizes (Square Feet) | 0.17 | 0.52 |
| Minimum of Min Lot Sizes (Square Feet) | 0.11 | 0.55 |
| What is the longest frontage requirement for single family residential development in any district? | 0.02 | 0.11 |
| Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by right? | -0.17 | 0.04 |
| Are attached single family houses (townhouses, 3+ units) listed as an allowed use (by right or special permit)? | -0.25 | 0.31 |
| Are apartments above commercial (mixed use) allowed in any district? | -0.25 | 0.28 |
| Is multi-family housing allowed, either by right or special permit (including through overlays or cluster zoning)? | -0.26 | 0.38 |
| Is multi-family housing listed as allowed through conversion (of either single family houses or non residential buildings)? | -0.26 | 0.15 |
| Is cluster development, planned unit development, open space residential design, or another type of flexible zoning allowed by special permit? | -0.32 | 0.09 |
| Are accessory or in-law apartments allowed (by right or special permit) in any district? | -0.39 | -0.08 |

Note: Prior to performing principal component analysis, all variables were normalized into z-scores. Missing data, where the model output "I don't know," were imputed k-nearest neighbors. 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 allowed.

# B 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 27

**Question Phrased by Pioneer:** What is the minimum lot size for each zoning district?

**Question Text That We Embed:** What is the minimum lot size for each zoning district?

**Question Background and Assumptions:** The question asks to provide a list of the minimum lot size in each district of the town. If a district has different minimum lot sizes depending on the type of building like for example a different minimum lot size for single-family homes than for multi-family homes, then you pick the smaller of the minimum lot sizes. If a district allows smaller minimum lot sizes for historic properties or by special permission then you pick the standard minimum lot size for current buildings. If a district only lists a minimum lot size for a specific type of housing like housing for the elderly, then you pick that minimum lot size. Your answer should be structured as a list with district name, minimum lot size, and units for the minimum lot size which are usually square feet or acres. If a minimum lot size for a district is reported in both acres and square feet then only report it in square feet. If a district does not have a minimum lot size then record the town wide minimum lot size for that district if a town wide minimum lot size exists. If a town wide minimum lot size does not exist and a district does not have a minimum lot size then exclude it from your answer.

**Question Type:** Lot Size

**Rephrased Question the LLM Sees:** What is the minimum lot size for each zoning district?

**Subtask:**

- Subtask Question That Gets Embedded: List out each district in the town

- Rephrased Subtask Question the LLM Sees: List out each district in the town
- Additional Subtask Instructions: Please list out the name of each district in the town. Do not include overlay districts.
- How The Subtask Results Are Described to the LLM Afterwards: List of all districts to find the minimum lot size for

## 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:** The question asks to provide a list of the minimum lot size in each district that permits single-family housing of the town. If a district has a specific minimum lot size for single-family homes then you choose that, otherwise you select the general minimum lot size for that district. If a district allows smaller minimum lot sizes for historic properties or by special permission then you pick the standard minimum lot size for current buildings. Your answer should be structured as a list with district name, minimum lot size, and units for the minimum lot size which are usually square feet or acres. If a minimum lot size for a district is reported in both acres and square feet then only report it in square feet. If a district does not have a minimum lot size then record the town wide minimum lot size for that district if a town wide minimum lot size exists. If a town wide minimum lot size does not exist and a district does not have a minimum lot size then exclude it from your answer.
**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 district in the town that allows single-family housing. If you cannot find any districts that explictly allow single-family housing then just assume that any primarily residential districts allow single-family housing.
- How The Subtask Results Are Described to the LLM Afterwards: List of all districts to find the minimum lot size for

## 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, just 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