

Applying Topic Modeling to Congressional Housing Legislation

Rebecca Wagner

PPOL 5205 - Fall 2025

SUMMARY

In this project, I collected data on over 1,500 congressional bills on housing policy, including their text and associated metadata. Using Latent Dirichlet Allocation, I created a 24-topic model and analyzed the relationships between these topics, party sponsors, bill outcomes, and subject assignments. I ultimately find most bills never make it past committee, a small set of topics consistently dominates the agenda, and Democrats are the primary drivers of these initiatives. Even within popular topics like public housing or community development, very few bills become law, highlighting both the procedural hurdles and the selective focus of Congress on particular aspects of housing policy.

RESEARCH OBJECTIVE

Housing represents one of the most pressing policy issues facing Americans today. For several decades, housing costs have outpaced wages, while housing supply has lagged behind demand. These factors have caused an increasing number of families to face housing insecurity, the impacts of which reach far beyond where a family sleeps at night. Even short bouts of housing insecurity have negative effects on long-term mental and physical health, employment status, economic stability, and educational outcomes. At the same time, the housing sector represents a substantial portion of the national economy, making it a central issue not only in social policy but in broader economic planning and stability (DeLuca & Rosen, 2022).

American housing policy also has a long and storied history, beginning with the National Housing Act of 1934, to the postwar G.I. Bill, to the Fair Housing Act of 1968, the Section 8 Amendments, and policy following the 2008 market crash (Leacock, 2019). Throughout this history, issues of racial and economic inequality have been deeply embedded in housing policy—both as forces that shaped policy decisions and as outcomes perpetuated by them. These policies have reflected and reinforced the political and cultural narratives of their times (Gross, 2017).

This project takes an exploratory approach, using natural language processing to analyze the textual content of housing-related congressional bills across multiple legislative sessions. By examining how language around housing has changed over time, I aim to uncover shifts in narrative focus, political framing, and policy priorities. Additionally, by incorporating metadata such as bill sponsorship, party affiliation, and legislative outcomes, I hope to identify patterns or features that may help predict a bill's progression through Congress. This study seeks to contribute to a deeper understanding of how housing is conceptualized in the legislative process and how these conceptualizations evolve in response to broader social, economic, and political trends.

DATA SOURCES & PREPROCESSING

Data for this project were collected using the Library of Congress's Congressional API (Library of Congress, 2025) and includes both the full text and associated metadata from 1,675 housing-related bills spanning the 110th through 119th congresses (January 2007–Present). Only bills tagged with the policy area "Housing and Community Development" were collected, meaning housing bills without this tag may be missing from the corpus. It is assumed that, while a bill's content may change as it makes its way through the chambers, these changes reflect congressional sentiment and the latest version is most representative of the entire Congress. For this reason only the most recent version of each bill was collected; the enrolled version if the bill passed, or the latest version considered if it did not.

Text Data

Preliminary text analysis reveals that the documents vary widely in length, ranging from 174 words to over 70,000 words, with an average of approximately 2,500 words. This distribution suggests the presence of a small number of significantly longer bills. A basic bag-of-words analysis helps identify frequently occurring terms across the corpus. Table 1 includes the top 25 words and their relative frequencies. Many of these words, such as 'housing', 'program', 'assistance', clearly point to the subject matter at hand of housing. Others, like 'agency', 'state', and 'federal' are words that refer to general government structures.

With this context as guidance, only simple preprocessing measures were taken. These include removing preambles, headers, and both standard english stop words and a custom list of common legislative terms such as 'shall', 'section', and 'paragraph'. The documents were tokenized and lematized, and a simple bigram detection model implemented to identify phrases such as 'loss mitigation', 'census tract', and 'domestic violence'. These steps resulted in a basic token count matrix, which was then used for further modeling. No other steps such as frequency transformation or normalization were applied, given the intent of topic modeling that does not necessitate such steps.

Table 1. Top 25 Most Frequent Words in Housing-Related Legislative Texts

Word	Count	Word	Count	Word	Count	Word	Count
housing	44,776	federal	9,654	including	7,921	urban	6,380
assistance	13,727	states	9,414	state	7,825	united	8,225
program	12,792	agency	8,877	42	7,524	year	6,808
public	12,328	loan	8,348	general	7,512	project	6,493
mortgage	11,118	following	8,043	grant	7,282	income	6,962
term	11,013	eligible	10,061	date	7,240	congress	6,921
development	9,899						

Metadata

The available metadata for each bill includes identifiers such as the bill id, all actions taken and their date, any involved committees, the origin chamber, the policy area, the subjects, the sponsor(s) and cosponsor(s), any related bills, among more. No preprocessing was done to this metadata, other than the steps taken to access it through the Congressional API. Some of this metadata is used for further analysis and is characterized here.

Beyond the policy area of 'Housing and Community Development', each bill may have an additional series of policy subjects assigned to it. Across all 1,675 bills, there are 646 unique subjects. These subject labels are far from perfect; most bills have multiple assigned subjects, some with greater than 15. It is difficult to know how or with what level of care these subject assignments are made. But, subjects also provide a useful validation measure for later modeling.

Table 2 displays the 10 most common subject assignments. Notably, development funding outweighs every other subject across all four congresses, which is intuitive as federal legislation often handles allocating funds to state and local agencies. In the case of housing policy, this includes local housing authorities and housing financing agencies. The proportion of these subjects in bills introduced over time are displayed in Figure 1. Though the 'other' category appears to dominate, it is important to note that this category encompasses 636 subjects, while the remaining categories are one subject each, emphasizing their relative importance.

Table 3 displays the number of bills across sessions introduced in each chamber, as well as the political party of the bills sponsor. This table shows that many more bills originate in the house (unsurprising given the relative size of the chamber), and that many more Democrats introduce bills (perhaps also unsurprising knowing each party's aptitude toward government oversight).

Table 2. Top 10 Subjects Related to Housing and Development Bills

Subject	110th	111th	112th	113th	114th	115th	116th	117th	118th	119th	Total
Housing and community development funding	0	101	54	36	54	67	162	111	91	22	698
Housing finance and home ownership	0	128	127	75	48	55	77	47	34	7	598
Low- and moderate-income housing	0	36	29	27	35	42	99	62	71	16	417
Congressional oversight	1	20	11	12	20	36	68	64	57	17	306
Public housing	12	26	20	13	25	26	67	41	39	11	280
Government information and archives	0	36	26	17	15	28	58	44	34	6	264
Government lending and loan guarantees	0	57	44	24	16	19	34	17	23	6	240
Housing supply and affordability	0	21	15	15	11	21	54	36	26	7	206
Department of Housing and Urban Development	4	29	19	14	15	18	48	30	17	6	200
Government studies and investigations	0	21	16	10	13	21	35	34	25	7	182

Note. Bills may be assigned multiple subject areas.

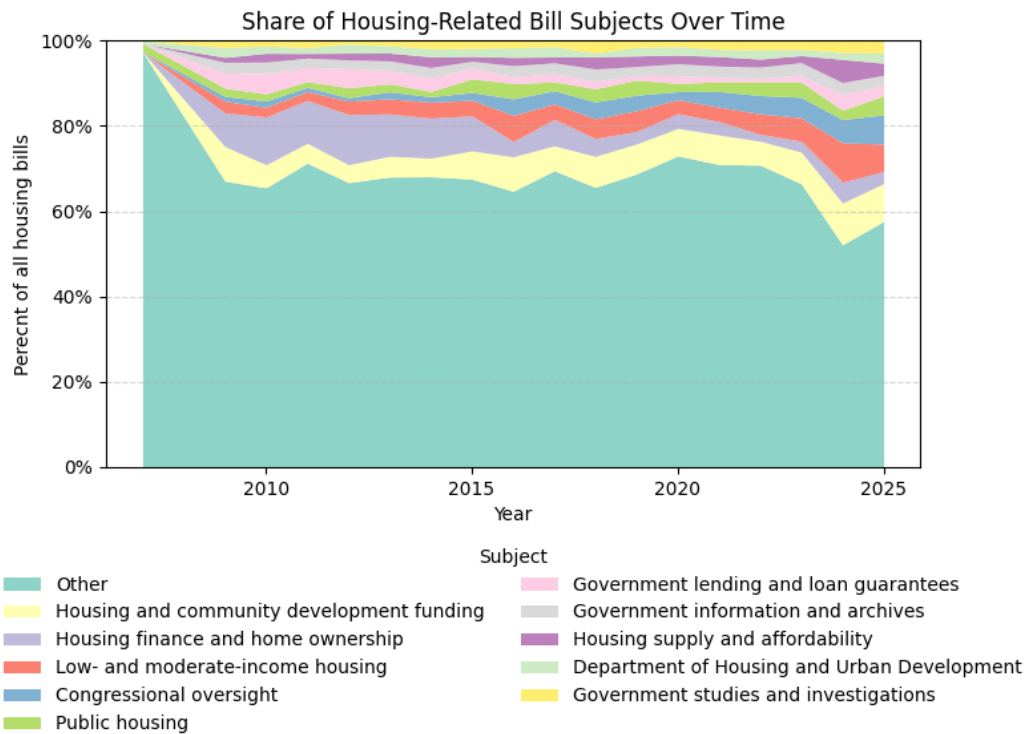


Figure 1. Share of Housing-Related Bill Subjects Over Time. The stacked area chart shows each subject’s percent of total bills per year.

Table 3. Originating Chamber and Sponsoring Party of Housing and Development Bills (110th–119th Congresses)

Category	110th	111th	112th	113th	114th	115th	116th	117th	118th	119th
Chamber										
House (H)	28	165	133	83	77	102	167	156	167	96
Senate (S)	19	42	47	26	32	39	82	73	79	43
Party										
Democrat (D)	37	165	106	68	67	85	215	192	188	94
Independent (I)	0	0	0	0	0	0	1	1	2	1
Republican (R)	10	42	74	41	42	56	33	36	56	44

MODELING TECHNIQUES

LDA & Parameter Tuning

Using the preprocessed data, I employed Latent Dirichlet Allocation (LDA) to create a topic model, in the hopes that it may uncover latent themes in the corpus. This standard topic modeling technique assigns topics to documents by modeling patterns of words across the corpus. It is unsupervised, generative, and probabilistic. In other words, it does not have predefined outcomes, but ‘generates’ the probability of identified patterns occurring. ‘Generates,’ here, is in quotation marks because the model does not *actually* create new documents. Rather, it assumes that documents are created according to an algorithm and attempts to reverse engineer this algorithm. Once this document generation reverse-engineering is complete, the LDA model can assign topic probabilities to each document and word probabilities to each topic, allowing us to analyze the underlying thematic structure of the corpus. This technique has proven effective in previous studies of legislative text. For example, Brodnax and James, 2025 use topic modeling to quantify the diffusion of political ideas across state legislatures, demonstrating the value of this approach for uncovering both explicit and implicit policy trends.

One downside to LDA is that, because it is unsupervised, the number of topics must be decided for the model as a hyperparameter. In an effort to select a value that resulted in distinct, but interpretable topics, I examined the coherence and perplexity scores of models with topic numbers ranging from 7 to 30. I selected these values with consideration for the size of my corpus. The results can be viewed in Figure 2. With some further testing, I ultimately decided on 24 topics for my model, as this value resulted in the highest coherence, important for human interpretation of the final model.

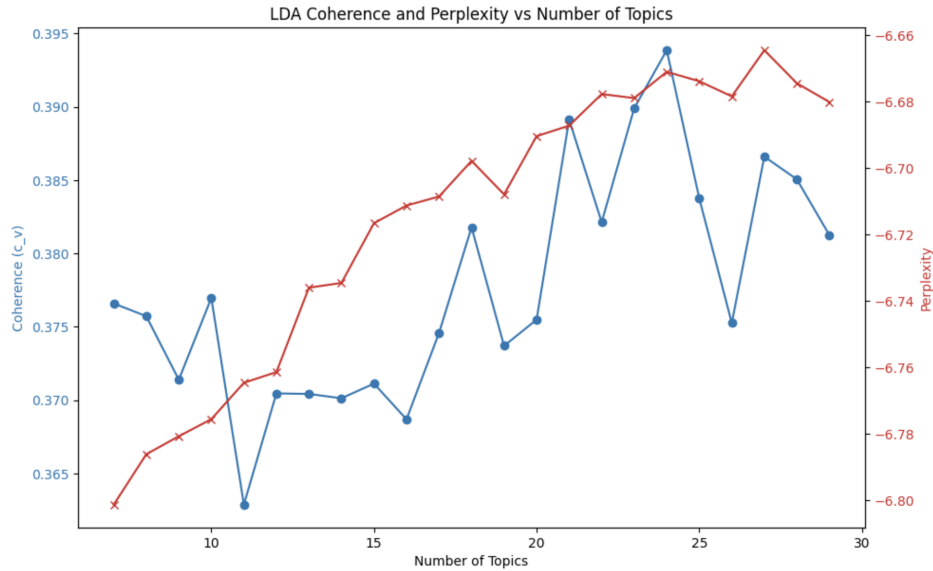


Figure 2. Model coherence and perplexity scores for different numbers of topics.

Model Validation

The final coherence of my LDA model was 0.394, and the final perplexity was -6.63. Relative to the scores of models with differing topic numbers, and with consideration for the complexity of my corpus, these values are somewhat strong. Table 4 displays both the coherence scores for individual topics and a measure of distinctiveness (calculated with a Jensen-Shannon Distance) for all 24 topics. While there is variation for both scores within topics, coherence scores tend to be similar to the overall coherence of 0.4, topic 18 having the highest coherence and topic 5 with the lowest. Jensen-Shannon scores indicate how distinctive each topic's word probability distribution is from all other topics' distributions. These scores tend to be higher than the coherence measure, indicating fairly distinct distributions. Again, topic 18 has the highest score, and topic 21 the lowest.

Further validation using subject metadata is discussed in the following section. The combination of these validation measures indicate that, while the model varies in coherence by topic, topics are generally distinct from one another. Additionally, Figure 2 indicates that more topics may lead to overfitting, leading me to believe that improvement in this topic model would necessitate further preprocessing, and is potentially limited by the complexity of the corpus.

Table 4. Topic Coherence (c_v) and Distinctiveness (Jensen-Shannon Distance)

Topic	c_v	JS	Topic	c_v	JS	Topic	c_v	JS	Topic	c_v	JS
1	0.2991	0.5020	7	0.3718	0.5254	13	0.3685	0.5049	19	0.3046	0.5267
2	0.3890	0.4744	8	0.3638	0.5065	14	0.3966	0.4781	20	0.4251	0.4885
3	0.3866	0.4808	9	0.3902	0.4809	15	0.4190	0.4861	21	0.4171	0.4610
4	0.4131	0.5143	10	0.4080	0.4782	16	0.5255	0.5550	22	0.4810	0.4864
5	0.2832	0.4973	11	0.4570	0.4991	17	0.3299	0.5044	23	0.2948	0.4913
6	0.4377	0.5242	12	0.3971	0.4837	18	0.5787	0.5966	24	0.3152	0.4786

Table 5. JS scores are calculated as an average of all pairwise JS scores between topics

FINDINGS

Topics

With an analysis of both top raw and distinctive words across topics, as well as manual investigation of associated bills, I describe the 24 topics in Table 6. Consistent with the prior coherence and distinctiveness analysis, some topics appear to overlap more than others. For instance, topic 18 (with the highest coherence and distinctiveness score) stands out from other topics, having to do with domestic violence and housing policy; no other topic is similar. Some other topics that stand out as distinct are 5 (Energy Efficiency), 21 (Rural Housing), and 24 (Low-Income Tax Credits). On the other hand, topics 4, 9, and 12 all appear to address homelessness and housing insecurity, and topics 1, 8, 13, and 16 all relate to federal oversight. These observations can be validated against the prior coherence and distinctiveness analysis.

I assign documents to the topic which has the highest probability from my lda model. Notably, these probabilities range from 23% to 99%, indicating a wide range of how "sure" the model is about these topics, and also likely topic overlap as already discussed. The average percentage is 67.5%, indicating a fairly confident model, on average. With these topic assignments, Table ?? displays the prevalence of each topic across the corpus, with raw document counts. Topics 10, 14, 13, 21, 11, and 8 combined make up over half of the corpus, containing multiple oversight and grant related topics. This is, again, intuitive for federal legislation that primarily handles financial allocations and oversight of state and local agencies.

Table 6. Topic Model Categories for Housing Legislation (24 Topics)

Topic	Category Name	Description
1	Federal Regulatory Oversight	Clarifying current law, deadline requirements, FHA oversight
2	Community Revitalization & Historic Preservation	Community development block grants, Property vacancies
3	Disaster Recovery Assistance	Housing grants for disaster-affected areas, hazard-prone zones, grantee administration
4	Homelessness Legislation (Congressional)	Legislative proposals addressing homelessness, congressional floor speeches
5	Energy Efficiency & Green Building	Renewable energy programs, efficiency standards, taxpayer incentives for green housing
6	Housing Inspections & Property Standards	Physical condition requirements, manufactured housing, contract compliance
7	Recreation & Community Facilities	Parks, recreational facilities, community infrastructure projects
8	Federal Housing Enterprise Oversight	GSE regulation (Fannie Mae/Freddie Mac), conservatorship, director authority
9	Emergency Rental Assistance	Rapid rehousing, emergency support for extremely low-income households, rental subsidies
10	Public Housing & Transportation	Public housing development with transportation connections
11	Mortgage Servicing & Foreclosure	Loan servicing processes, foreclosure procedures, Ginnie Mae pools
12	Supportive Services for Homeless	Health services, behavioral health, wraparound services for homeless individuals
13	Federal Housing Finance & Securities	FHA, GSE securities, bonds, federal housing finance structure
14	Community Development Block Grants	CDBG programs, eligible entities, housing supply initiatives

Topic	Category Name	Description
15	Public Housing Agencies & Vouchers	PHA operations, Section 8 voucher programs, small agency requirements
16	FHFA Governance & Board Structure	Federal Housing Finance Agency leadership, board compensation, governance
17	Affordable Housing Programs	Income-targeted programs, area median income metrics, demographic considerations
18	Domestic Violence Housing Protections	VAWA housing protections, survivor assistance, trafficking victim services
19	Homeowner Assistance & Foreclosure Prevention	HAMP modifications, foreclosure counseling, homeowner support programs
20	Public Housing Transformation & Lead Remediation	HOPE VI-style transformation, lead paint hazards, resident displacement
21	Rural & Multifamily Housing	USDA rural housing programs, multifamily projects, Native Hawaiian housing
22	Tenant Rights & Fair Housing	Eviction protections, landlord-tenant law, discrimination protections
23	Community Financial Institutions	CDFIs, community banks, housing finance institutions
24	Low-Income Housing Tax Credits	LIHTC program, tax incentives, qualified allocations

For a longitudinal analysis, Figure 3 displays the relative intensity of topics over each year. Standout topics include Mortgage Servicing and Foreclosures spiking in 2009 and the following years, presumably in response to the 2008 market crash. Additionally, Emergency Rental Assistance spikes in 2020, likely reflecting the surge of legislation protecting tenants during the COVID-19 Pandemic. Less intense, but notable points also include the increase in Community Development Block Grants, Tenant Rights & Fair Housing, and Public Housing in years 2020-2015. I guess that this reflects the Biden Administration's left-leaning housing initiatives.

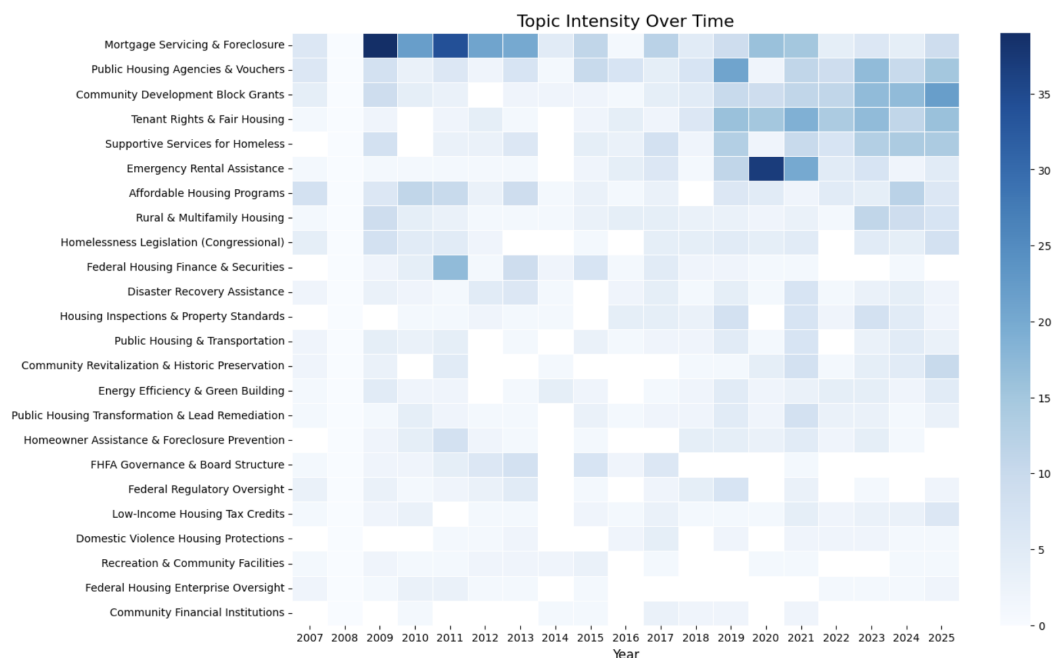


Figure 3. Heatmap of topic intensity over time. Darker values indicate more bills in the associated topic and year.

Subjects

Table 7 displays the top subject associated with each topic, providing a concrete way to interpret the latent topics identified by our model. As expected, Housing Finance and Home Ownership emerges as the dominant subject across the majority of topics, reflecting the centrality of financial and ownership issues in housing-related legislation. However, some topics exhibit notable deviations from this pattern. For instance, the topic focused on Housing Inspections and Property Standards is primarily associated with Low- and Moderate-Income Housing, suggesting that regulatory and inspection policies are especially salient in the context of housing accessible to lower-income populations. Similarly, the topic on Public Housing unsurprisingly aligns most strongly with the Public Housing subject, highlighting the consistency between the model's topic assignments and domain-specific categorizations.

These alignments provide additional validity for our topic model, as the subject labels corroborate the broader topic descriptions we assigned. By demonstrating that the top subjects reflect the semantic content implied by the topic names, we gain confidence that the LDA model is capturing meaningful patterns in legislative text. Furthermore, the presence of both expected and divergent top subjects suggests that the model is sensitive enough to distinguish nuanced subthemes within the corpus, rather than collapsing all legislation under a single general category.

Table 7. Top Subject Assigned per Topic Category

Topic Category	Top Subject (Count)
Federal Regulatory Oversight	Housing finance and home ownership (26)
Community Revitalization & Historic Preservation	Housing and community development funding (18)
Disaster Recovery Assistance	Housing and community development funding (31)
Homelessness Legislation (Congressional)	Homelessness and emergency shelter (18)
Energy Efficiency & Green Building	Housing and community development funding (22)
Housing Inspections & Property Standards	Low- and moderate-income housing (31)
Recreation & Community Facilities	Housing and community development funding (16)
Federal Housing Enterprise Oversight	Housing finance and home ownership (13)
Emergency Rental Assistance	Housing and community development funding (84)
Public Housing & Transportation	Housing and community development funding (27)
Mortgage Servicing & Foreclosure	Housing finance and home ownership (207)
Supportive Services for Homeless	Homelessness and emergency shelter (57)
Federal Housing Finance & Securities	Housing finance and home ownership (39)
Community Development Block Grants	Housing and community development funding (79)
Public Housing Agencies & Vouchers	Public housing (103)
FHFA Governance & Board Structure	Housing finance and home ownership (36)
Affordable Housing Programs	Housing finance and home ownership (41)
Domestic Violence Housing Protections	Housing finance and home ownership (14)
Homeowner Assistance & Foreclosure Prevention	Housing and community development funding (31)
Public Housing Transformation & Lead Remediation	Housing and community development funding (20)
Rural & Multifamily Housing	Housing and community development funding (39)
Tenant Rights & Fair Housing	Housing and community development funding (58)
Community Financial Institutions	Housing finance and home ownership (11)
Low-Income Housing Tax Credits	Housing and community development funding (18)

Sponsoring Party

Table 8 presents the distribution of sponsoring party across topics. Consistent with the overall trend, Democrats sponsor more housing-related bills than Republicans, so their larger counts across most topics are expected. However, certain patterns are noteworthy. Democrats are particularly dominant in topics such as Mortgage Servicing, Community Development, and Rural Housing, which may reflect their focus on consumer protections, community investment, and rural development initiatives. Conversely, Republicans show stronger sponsorship in topics like Mortgage Servicing, Federal Housing Finance, and Public Housing Agencies, suggesting a focus on financial and institutional aspects of housing policy.

These trends highlight a substantive difference in legislative priorities by party. While both parties engage heavily in mortgage-related legislation, Republican-sponsored bills tend to emphasize financial and regulatory mechanisms, whereas

Democratic-sponsored bills appear more oriented toward social and community impacts. This distinction provides insight into the partisan dimensions of housing policy, revealing how topic modeling can uncover nuanced patterns in legislative sponsorship.

Table 8. Bills by Topic and Sponsoring Party

Topic	Democrat (D)	Independent (I)	Republican (R)
Federal Oversight	19	0	18
Community Revitalization	32	0	13
Disaster Recovery	29	0	20
Homelessness	49	0	15
Energy Efficiency	32	0	13
Property Standards	18	0	1
Recreation	12	0	5
FHFA Oversight	89	2	15
Emergency Rental Assistance	39	3	4
Public Housing & Transport	39	0	4
Mortgage Servicing	193	0	46
Supportive Services	86	0	24
Federal Housing Finance	14	0	41
Community Development	117	0	16
Public Housing Agencies	80	0	66
FHFA Governance	12	0	27
Affordable Housing	75	0	20
Domestic Violence	16	0	4
Homeowner Assistance	34	0	7
Public Housing Transformation	48	0	22
Rural & Multifamily	116	0	16
Tenant Rights	7	0	5
Community Financial Inst.	20	0	15
Low-Income Housing Tax Credits	39	0	4

Bill Outcomes

Table 9 summarizes the legislative status of bills across the 24 topics, with early-stage outcomes such as advancement, calendar placement, and procedural steps combined into the “Other” category. Overall, the majority of bills die in committee, reflecting the common legislative bottleneck for housing-related legislation, with 1,252 bills falling into this category. Died-in-subcommittee counts are smaller but nontrivial (306 bills), while very few bills fail a floor vote (1 bill). Only 66 bills manage to pass one chamber, highlighting the significant attrition in the legislative process.

Some topics display notable patterns. Mortgage Servicing and Public Housing Agencies exhibit the highest counts of bills dying in committee, as well as relatively high early-stage activity, indicating heavy legislative attention but limited progression. Emergency Rental Assistance and Supportive Services also see substantial committee attrition, reflecting complex procedural hurdles. Conversely, topics such as Community Financial Institutions and Low-Income Housing Credits have fewer bills overall, suggesting more specialized or narrowly targeted legislation. The distribution emphasizes that while housing legislation covers a wide range of topics, the legislative process is highly selective, and only a small fraction of bills advance beyond early stages.

CONCLUSION

This analysis of housing-related legislation over the past decades reveals several clear patterns. First, Housing Finance and Home Ownership dominates the legislative agenda, appearing as the top subject across most topics. Other areas, such as Public Housing, Housing Inspections, and Supportive Services for Homeless, also emerge as distinct legislative clusters, confirming the interpretability and coherence of the 24-topic model. Second, the legislative process is highly selective: the vast majority of bills die in committee or subcommittee, with only a small fraction passing one chamber. Certain topics, such as Mortgage Servicing, Community Development, and Rural & Multifamily Housing, see heavy Democratic sponsorship, whereas Public Housing Agencies and Federal Housing Finance tend to be dominated by Republican sponsorship, reflecting differing party priorities. Third, early-stage procedural actions (advancements, hearings, forwarding) are common but rarely translate into

legislative success, highlighting the importance of committee-level negotiations and political alignment.

Further research on this topic could include a more in-depth temporal analysis, or using these topics to implement a predictive model for the likelihood of bills passing one chamber or another. Further party-based analysis could explore cosponsors and bipartisan bills. Overall, the combination of topic modeling and legislative outcome analysis provides a robust framework for understanding the structure, priorities, and bottlenecks in housing-related legislation, offering both descriptive insights and actionable avenues for future research.

Table 9. Legislative Status of Bills by Topic with Combined Early-Stage Outcomes. 'Other' category includes bills that were advanced, placed on calendar, forwarded by a committee, and other procedural steps

Topic	Other	Died in Committee	Died in Subcommittee	Failed Floor	Passed One Chamber
Federal Oversight	5	25	2	0	5
Community Revitalization	3	28	13	0	0
Disaster Recovery	4	35	8	0	2
Homelessness Legislation	6	39	13	0	6
Energy Efficiency	7	30	7	0	2
Housing Inspections	2	41	6	0	0
Recreation & Community	0	11	7	0	0
FHFA Oversight	3	8	5	0	1
Emergency Rental Assistance	8	86	7	0	3
Public Housing & Transit	5	30	7	0	0
Mortgage Servicing	19	160	52	0	8
Supportive Services	5	88	11	0	5
Housing Finance & Securities	1	35	9	0	1
Community Block Grants	7	108	16	0	1
Public Housing Agencies	12	117	10	0	7
FHFA Governance	5	19	8	0	4
Affordable Housing Programs	2	73	15	0	4
Domestic Violence Protections	3	12	7	0	1
Homeowner Assistance	4	25	4	1	5
Public Housing Transformation	1	34	8	0	0
Rural & Multifamily Housing	6	49	10	0	3
Tenant Rights & Fair Housing	11	99	21	0	1
Community Financial Inst.	3	5	3	0	1
Low-Income Housing Credits	3	28	4	0	0
Total	133	1252	306	1	66

REFERENCES

- Brodnax, N., & James, S. (2025). Topics as outcomes: Modeling the influence of intergovernmental grants on policy diffusion. *Policy Studies Journal*, 0(0), 1–19. <https://doi.org/10.1111/psj.70074>
- DeLuca, S., & Rosen, E. (2022). Housing insecurity among the poor today. *Annual Review of Sociology*, 48, 343–371. <https://doi.org/10.1146/annurev-soc-090921-040646>
- Gross, T. (2017, May). *A forgotten history of how the u.s. government segregated america*. National Public Radio. <https://www.npr.org/2017/05/03/526655831/a-forgotten-history-of-how-the-u-s-government-segregated-america>
- Leacock, K. (2019, October). *A brief history of housing policy in the u.s.* National Nurse-Led Care Consortium. <https://nurseledcare.phmc.org/advocacy/policy-blog/item/641:a-brief-history-of-housing-policy-in-the-u-s.html>
- Library of Congress. (2025). Congressional pai: Preliminary analysis and information. <https://www.loc.gov/congressional-pai>