Energy Policy in Flux: An NLP Approach to Media Discourse Before and After the Russia-Ukraine Conflict

Priscila Rebeca Stisman

April 29th, 2025

Abstract

This paper explores shifts in the discourse surrounding energy policy by applying natural language processing (NLP) techniques to New York Times articles published between 2021 and 2023. Using topic modeling approaches such as Latent Dirichlet Allocation (LDA) and Structural Topic Modeling (STM), alongside regression analysis, I examine whether the prominence and framing of energy-related issues changed following the beginning of the Russia-Ukraine war. My findings show that following the onset of the war, energy-related topics became significantly more prominent. However, their salience appears to decline gradually over time, suggesting a normalization or shift in media focus.

Keywords: *Energy Policy, NLP, Topic Modeling*

1. Introduction

In recent years, discussions around energy policy have become increasingly common. While topics like climate change, renewable energy, fossil fuels, and natural gas were once reserved for more technical or niche debates, they have now entered the mainstream. As climate change becomes an increasingly urgent global concern, energy security has emerged as a critical issue on the policy agenda.

However, this shift in focus is not without controversy. Some individuals remain skeptical of global warming, while others argue that pressing challenges such as poverty, inequality and economic instability should take precedence over investing in innovation for energy security. This divergence of opinion makes it particularly interesting to analyze whether public discourse around energy has shifted in recent years.

One valuable way to trace changes in public opinions and perspectives is through media reporting [3]. Media outlets play a crucial role in the policymaking process by acting as a bridge between politicians and the public. In this context, examining media coverage offers a useful proxy for understanding societal attention and concern around energy issues.

In this paper, I aim to examine whether there has been a significant shift in the media discourse surrounding energy since the onset of the Russia-Ukraine war. This conflict has had profound implications for global energy markets and security, potentially reshaping both the way energy issues are discussed in the media and the level of attention society pays to them. For this analysis, I will focus on The New York Times, one of the most influential newspapers in the world. With its extensive international readership and established role in shaping the global news agenda, The New York Times provides a valuable lens through which to assess changing narratives and shifting public attention on energy-related matters.

2. Data

For this research, I use New York Times articles published between 2021 and 2023, filtered by a set of energy-related keywords. The original dataset, sourced from the New York Times API, includes over two million articles spanning from 2000 to 2025, each containing metadata such as publication date and time, headline, abstract, lead paragraph, associated tags, department, section, authors, and word count. The API provides access to article metadata as soon as the content is published online.

Focusing on the period surrounding the onset of the Russia-Ukraine war (which began on February 24, 2022), I retain 126,109 articles from 2021 to 2023. From each article, I extract four key features: the abstract, headline, lead paragraph, and publication date.

To analyze potential shifts in the narrative around energy policy during this period, I further filter the dataset using energy-related keywords. This yields a final subset of 18,293 articles focused on energy topics.

3. Methods

The analysis begins with text pre-processing, where the raw text is converted into a corpus. Punctuation, numbers, symbols, and common stopwords are removed, and stemming is applied to normalize word forms. To improve the quality of the data, hyphenated words are retained, and additional custom filtering is performed to exclude high-frequency, semantically neutral terms such as "said," "article," "first," and "new." Following this, a document-feature matrix (DFM) is constructed, removing extremely rare and overly common terms to reduce noise and focus on more meaningful vocabulary.

Next, exploratory text analysis is conducted to uncover initial patterns and trends in the data. Sentiment analysis is then performed on articles related to energy policy to assess shifts in tone before and after the onset of the Russia–Ukraine war.

For topic modeling, Latent Dirichlet Allocation (LDA) is applied to identify latent topics within the corpus. This technique provides a broad understanding of the thematic structure in the texts. Subsequently, a regression analysis using Interrupted Time Series (ITS) is conducted to evaluate the impact of the war on the salience of energy-related topics in the media, accounting for both immediate and long-term changes in coverage.

Finally, Structural Topic Modeling (STM) is employed to capture the evolving themes within the energy discourse over time. STM incorporates document-level metadata, such as publication date, allowing for the identification of shifts in topic prevalence before and after the war, providing a nuanced view of how energy topics have changed in response to global events.

4. Results

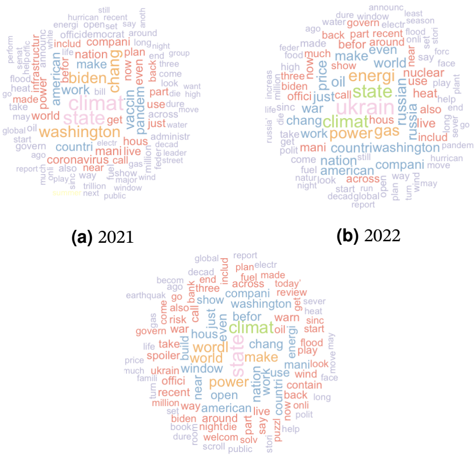
Using this pre-processed text, I generate word clouds for each year—2021, 2022, and 2023—to visually capture the most prominent terms in the energy discourse (see figure 1). In addition, I extract the top 10 most frequent words from each year to identify salient themes and possible changes in the language used around energy topics over time.

Year	Top 10 Words
2021	climat, state, chang, washington, biden, vaccin, work, pandem, american, make
2022	ukrain, state, climat, energi, power, gas, russian, oil, price, russia
2023	state, climat, wordl, power, make, world, nation, even, work, window

The most frequently used terms in energy-related articles reveal how global events have shaped the media narrative over time. In 2021, dominant words include reflect attention to U.S. politics and the COVID-19 recovery.

In 2022, there is a marked shift, with words like “Ukraine,” “energy,” “climate,” “power,” “gas,” “oil,” “price,” and “Russian” emerging at the top. This surge in energy-related vocabulary coincides with the outbreak of the Russia-Ukraine war, which triggered widespread concern over energy security, price volatility, and geopolitical tensions.

By 2023, however, only two energy-related terms—“climate” and “power”—remain among the most frequent. This suggests that the heightened media focus on energy policy seen in 2022 began to taper off, possibly as the initial shock of the crisis settled and coverage diversified into other areas.



(a) 2021 **(b)** 2022 **(c)** 2023

Figure 1. Word Clouds

4.1. Dictionary-Based Analysis

Sentiment analysis is conducted on article headlines using the VADER (Valence Aware Dictionary and sEntiment Reasoner) algorithm, which assigns a compound sentiment score ranging from -1 (most negative) to +1 (most positive). Based on these scores, each headline is classified as positive, neutral, or negative.

To examine shifts in the tone of energy policy coverage, I compare sentiment distributions before and after the onset of the Russia-Ukraine war (February 24, 2022).

As shown in Figure 2, negative sentiment increases after the war, with the share of negative headlines rising from approximately 32% pre-war to around 36% post-war. This shift likely reflects the global energy crisis triggered by the war, which led to rising oil and gas prices and concerns over supply disruptions. The increase in negative sentiment does not necessarily indicate dissatisfaction with energy policy itself but rather captures the broader challenges and anxieties surrounding energy in the context of the ongoing conflict.

4.2. Latent Dirichlet Allocation

To examine how energy-related news coverage shifted following the Russia-Ukraine war, I perform topic modeling on the full energy corpus using the Latent Dirichlet Allocation (LDA) algorithm [1]. This analysis aims to identify the dominant topics in energy discourse before and after the war, with a focus on key terms that have gained prominence. To select the optimal number of topics, I evaluate coherence and perplexity scores[1], which indicate K* = 5.

The topic modeling results reveal a clear thematic shift from pre-war to post-war coverage. Before the war, the discourse was diverse, with a focus on climate change, domestic politics, and pandemic recovery. After the war’s onset, the emphasis shifted sharply to energy security, geopolitics, and the conflict’s impact on global energy markets.

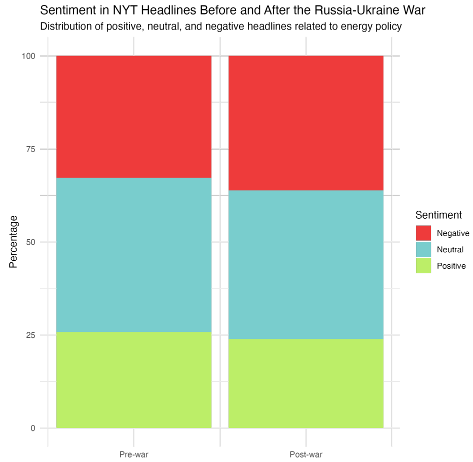


Figure 2. Sentiment Analysis of Energy Policy Discourse Pre and Post War

Notably, topics related to nuclear energy, gas, and oil became more prominent, reflecting the growing urgency around energy security and the geopolitical ramifications of the war. Terms like “nuclear,” “gas,” “oil,” and “power” emerged as central to post-war topics, highlighting a shift from pre-war concerns about climate change to a focus on energy supply, geopolitical tensions, and the role of fossil fuels in an increasingly unstable global landscape.

This shift underscores how the war has redefined energy discussions, with gas, oil, and nuclear power now central to debates on international policy and security. The post-war period reveals a broader concern for energy stability, suggesting that geopolitical tensions are now inextricably linked to global energy discourse.

In the appendix, I present the LDA topic modeling results by year, providing insights into the dominant themes of each period. In 2021, the topics remain centered on the pandemic’s impact. By 2022, the discourse shifts toward energy security, energy sources, gas prices, and the war. However, in 2023, the focus on energy diminishes, with energy only appearing prominently in Topic 3, indicating a decline in media attention to energy issues.

4.3. Interrupted Time Series Regression

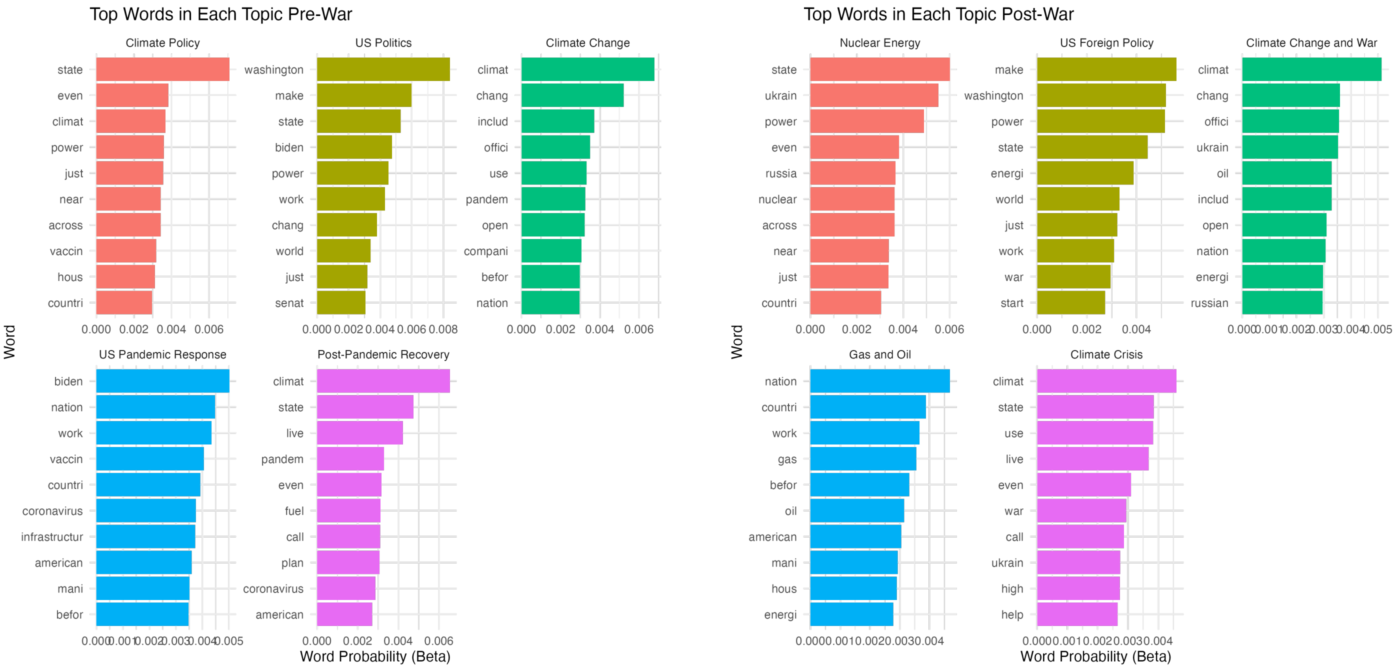
To better understand how the Russia–Ukraine conflict shaped public discourse on energy policy, I construct an energy salience score that quantifies the prominence of energy-related language in news articles. Specifically, this score is calculated as the proportion of energy-related words—such as “oil,” “gas,” “fuel,” "fossil," “solar,” and “nuclear”—that appear in the headline, abstract, and lead paragraph of each article, relative to the total number of words in those sections. The goal is to capture how central energy topics are in the framing of the news over time.

I begin the analysis by estimating a baseline regression that evaluates whether the war had a statistically significant effect on the energy salience of news articles. This initial model includes a simple binary indicator for whether an article was published after the start of the war. By comparing the average energy scores before and after the war, this approach allows me to detect whether there was a discrete shift in the importance of energy topics following the invasion.

$$EnergyScore = \beta_0 + \beta_1 War \quad (1)$$

To explore whether the impact of the war evolved over time, I extend the analysis by incorporating a variable that measures the number of days since the conflict began. This variable is negative for pre-war articles and positive for those published afterward. By including this continuous time measure, I can assess whether the

[1]Coherence measures the semantic similarity of top words in each topic, with higher scores signifying more interpretable topics. The first local maximum in coherence (Figure 6) occurs at K = 5. Perplexity, which gauges how well the model predicts unseen data, reaches its minimum around K = 22 (Figure 7). While the 22-topic model fits better statistically, K = 5 offers a balance between interpretability and complexity.



(a) Pre-War **(b)** Post-War

Figure 3. Topic Modeling using Latent-Dirichet Allocation

Variable	Energy Score
Intercept	3.2066 ***
	(0.039)
War	-0.05329
	(0.0535)
F(1,9295)= 0.9937 p-value: 0.3189	

Table 1. Regression Results for Model 1

salience of energy topics increased, decreased, or remained stable in the weeks and months that followed, providing a more nuanced view than the binary war indicator. Additionally, I introduce an interaction term between the war indicator and the time-elapsed variable. This specification allows for differential time trends before and after the war, helping to determine whether the post-war rise in energy attention was sustained or simply a temporary spike.

$$EnergyScore = \beta_0 + \beta_1 War + \beta_2 TimeTrend + \beta_3 Time * War \quad (2)$$

Variable	Energy Score
Intercept	3.334***
	(0.0832)
War	0.0009
	(0.1052)
Time Trend	0.0006
	(0.0003)
War*TimeTrend	-0.0014***
	(0.0004)
F(3,9295)= 5.162 p-value: 0.0015	

Table 2. Regression Results for Model 2

The war's effect is not statistically significant in either Model 1 or Model 2. However, Model 2 shows that the impact of the war on the energy score weakens over time. This trend is consistent with the results of previous sections, which indicate a decline in the prominence of energy-related topics in 2023.

Finally, I conducted a Chow Test to determine whether a structural break occurred at the onset of the war. The Chow statistic is 6.005, with a p-value of 0.002, allowing us to reject the null hypothesis of no structural break. This result indicates a significant shift in how energy-related content was framed in news articles following the Russia-Ukraine war. The relationship between time and the energy salience metric changed notably at this point, signaling a structural break in the data.

4.4. Structural Topic Modeling

Finally, I apply a Structural Topic Model (STM) [2] to analyze how energy-related news coverage evolved in response to geopolitical events like the Russia-Ukraine war. Unlike traditional models such as LDA, STM incorporates covariates—such as the binary war indicator and time trend—enabling a more nuanced examination of how topics change over time or across different conditions.

I fit the STM with five topics, consistent with my earlier LDA model, and estimate how their prevalence changes by including time, war, and their interaction. To capture potential non-linear trends in coverage, I model time using regression splines with 5 degrees of freedom. This flexible, nonparametric approach allows the model to adapt to complex, data-driven patterns in topic prevalence, without requiring assumptions about the underlying functional form.

This STM framework provides valuable insight into how energy-related discourse evolved from 2021 to 2023, offering a deeper understanding of media narratives and shifts in public attention.

I then analyze how the prevalence of each topic changed after the war (see Figure 4). Positive values indicate an increase in topic prevalence post-war, while negative values indicate a decline.

Topics 1 and 4, which focus on greenhouse gas emissions and the energy crisis, respectively, become more prevalent after the war, as expected given the heightened focus on energy security and geopolitical tensions. In contrast, Topic 5, which centers on gas prices, declines in prevalence post-war. This might seem counterintuitive but could reflect a shift in media framing from specific price fluctuations to broader narratives on energy security and global markets. Additionally, this trend may also be driven by the general decline in attention to gas prices over time.

Topic 2, which covers US politics, shows a slight decline after the

Table 3. LDA Topics with Highest Probability and FREX Words

Topic #	Label	Highest Probability Words	FREX Words
1	Greenhouse Emissions	energi, countri, global, mani, russian, work, war	carbon, energi, inflat, russian, conflict, greenhous, emiss
2	US Politics	climat, presid, unit, washington, nuclear, nation, biden	biden', republican, presid, democrat, tax, washington, polici
3	Lifestyle	time, home, can, wednesday, hurrican, tuesday, govern	room, grill, recip, meat, ice, ingredi, music
4	Energy Crisis	year, one, chang, oil, day, ukrain, will	oil, fuel, will, day, world', crisi, one
5	Gas Price	state, said, gas, peopl, price, thursday, month	price, peopl, electr, rain, high, area, thursday

Note: Topics extracted using LDA; FREX words help interpret topic uniqueness and frequency.

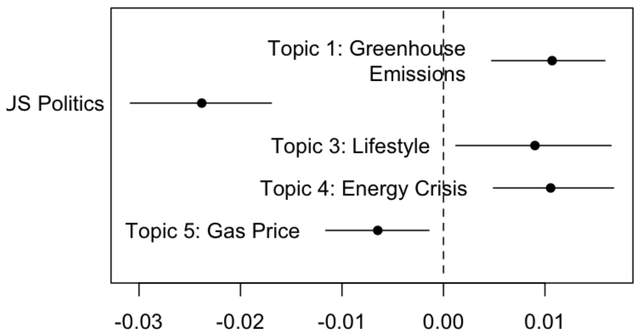


Figure 4. War Effect on Topics

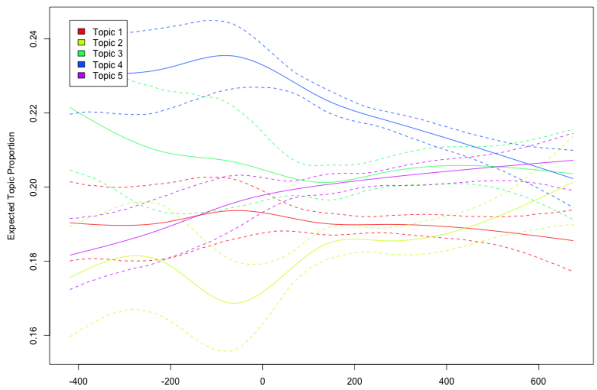


Figure 5. Topics Prevalence Over Time

war, while Topic 3, focused on lifestyle, becomes marginally more prominent. However, the magnitude of these changes is relatively small, suggesting that while geopolitical events influence media coverage, broader, longer-term trends likely play a more significant role.

To further explore these changes, I use spline regressions to model the temporal effect on topic prevalence (see Figure 5). Negative values indicate periods before the war, while positive values represent the period after.

Several topics show clear shifts around the war's onset. Topic 1 (greenhouse emissions) sees a slight increase in attention leading up to the war, followed by a gradual decline as the conflict progresses. Topic 4 (Energy Crisis) spikes just before the war, reflecting anticipatory coverage, but its prevalence diminishes over time. Interestingly, Topic 5 (Gas Prices) continues to rise both before and after the war, likely due to sustained economic impacts. Conversely, Topic 2 (U.S. Politics) experiences a sharp decline during the war period, suggesting a shift away from domestic political narratives in favor of energy and geopolitical coverage.

5. Discussion

The analysis of shifts in energy policy discourse before and after the Russia-Ukraine war reveals several key insights. First, the surge in negative sentiment following the onset of the conflict highlights how geopolitical crises can swiftly influence public perception and media narratives surrounding energy policy. LDA topic modeling shows a clear shift in focus—from broader issues like climate change, pandemic politics, and economic concerns to an increased emphasis on energy security and geopolitics. This shift signals a recalibration of priorities in response to the global crisis.

The gradual decline in the salience of energy issues, as observed in the interrupted time series analysis, suggests that while the war initially intensified media coverage, the urgency surrounding energy policy diminished over time as other global concerns came to the forefront. This pattern underscores the fleeting nature of media attention, particularly for issues like energy policy, which often recede once the immediate crisis subsides.

Structural topic modeling further highlights the evolving salience of different energy-related topics, showing how some gained prominence during the war, while others surged temporarily only to fade as the immediate crisis subsided.

In conclusion, the analysis illustrates how a major geopolitical event like the Russia-Ukraine war can significantly reshape the discourse surrounding energy policy, steering the conversation toward security and geopolitical concerns. However, this focus, while crucial during the conflict, faded once the immediate crisis waned. This is concerning, given the ongoing climate change crisis and the uncertainty surrounding energy security in the years ahead. Energy policy, being critical for the development of nations, should remain at the forefront of discussions, with a focus on long-term energy security as a priority in today's world.

A. Appendix

A.1. Metrics to choose optimal K

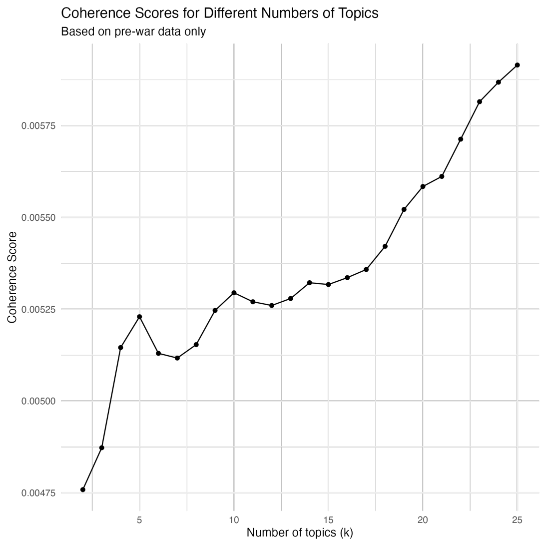


Figure 6. Coherence Score to choose optimal K

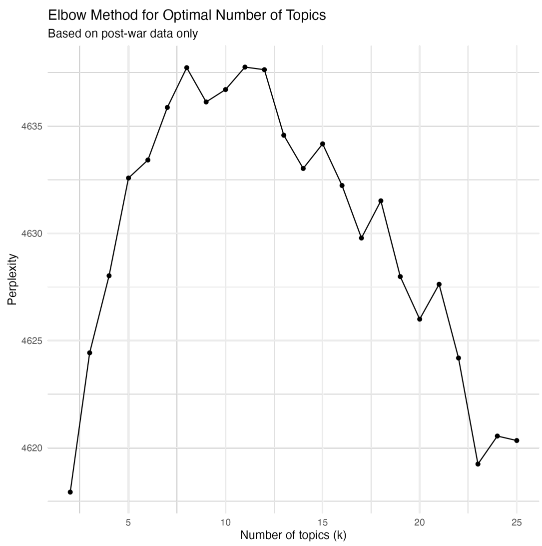


Figure 7. Perplexity Score to choose optimal K

A.2. LDA Topic Modeling by year

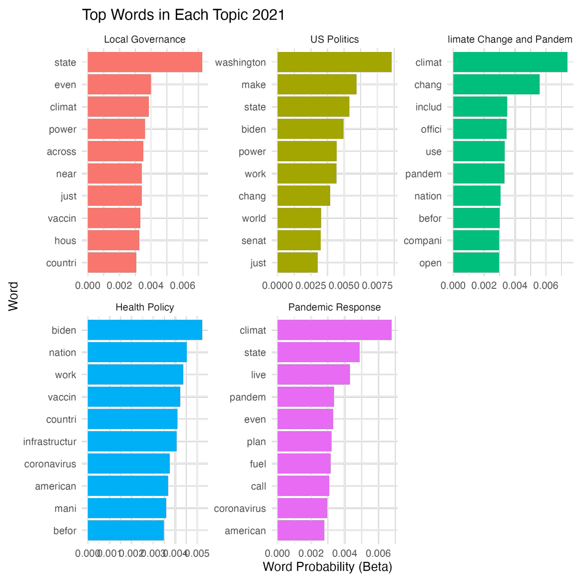


Figure 8. Topic distribution, 2021

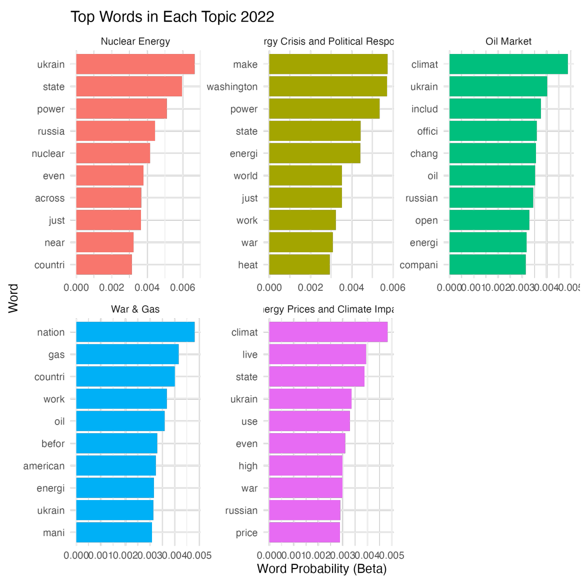


Figure 9. Topic distribution, 2022

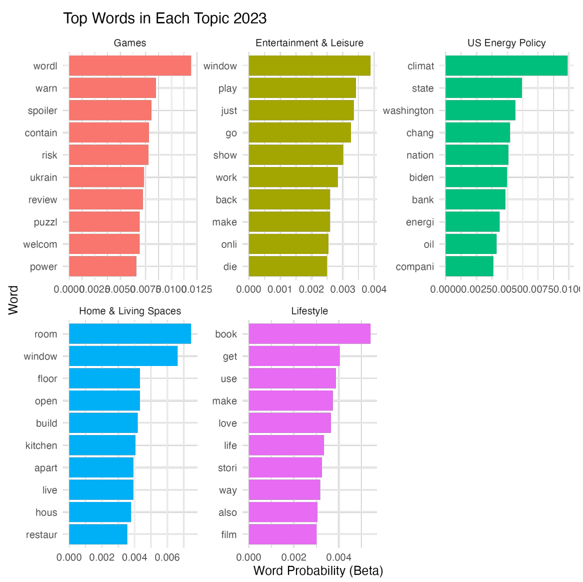


Figure 10. Topic distribution, 2023

References

[1] *Latent Dirichlet Allocation*. [Online]. Available: https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf.

[2] *R Package for Structural Topic Models*. [Online]. Available: https://cran.r-project.org/web/packages/stm/vignettes/stmVignette.pdf.

[3] *Topic Modelling Three Decades of Climate Change News in Denmark*. [Online]. Available: https://ssrn.com/abstract=4513921.