



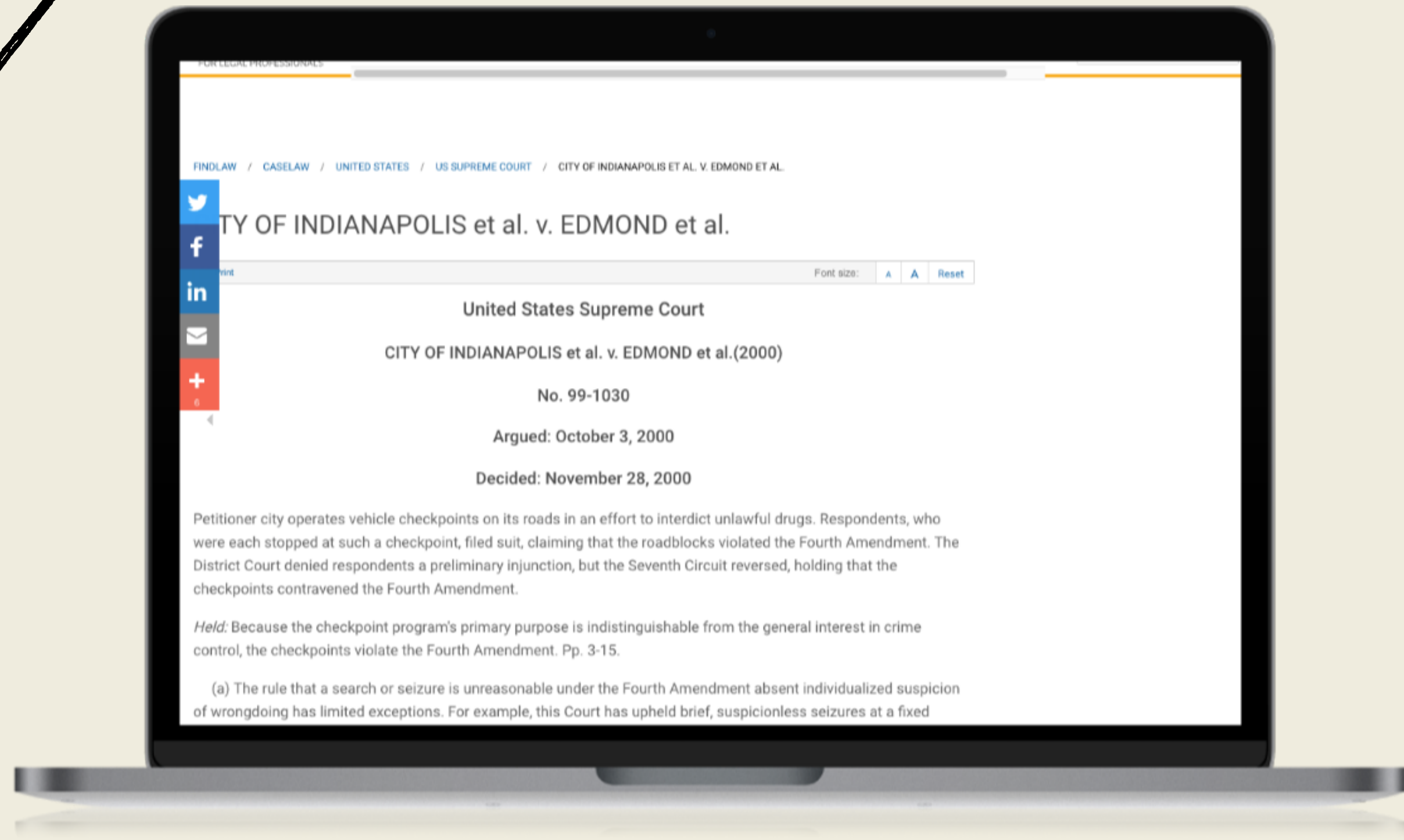
The U.S. Supreme Court

Decoding the opinions that have shaped the legal
landscape in the modern political era.

The Problem



The Opinions



Goal

Topic Modeling



Navigate the online library of cases.



Acquire opinions relevant to the modern political landscape.



Uncover hidden semantic relationships and themes.

Data

7,741 opinions
[1968 - present]

Preprocessing

1. Removed footnotes and in text citations
2. Fixed Unicode encoding errors
3. Removed '\n', punctuation, numbers and custom stop words
4. Converted all characters to lowercase
5. Applied lemmatization with NLTK

Workflow

PREPROCESS

Pandas, Numpy, Gensim, NLTK

VECTORIZE

Gensim.corpora, sklearn.CountVectorizer,
sklearn.TfidfVectorizer

TOPIC MODEL

gensim.LDAMulticore, sklearn.LDA,
CorEx

EVALUATE AND INTERPRET

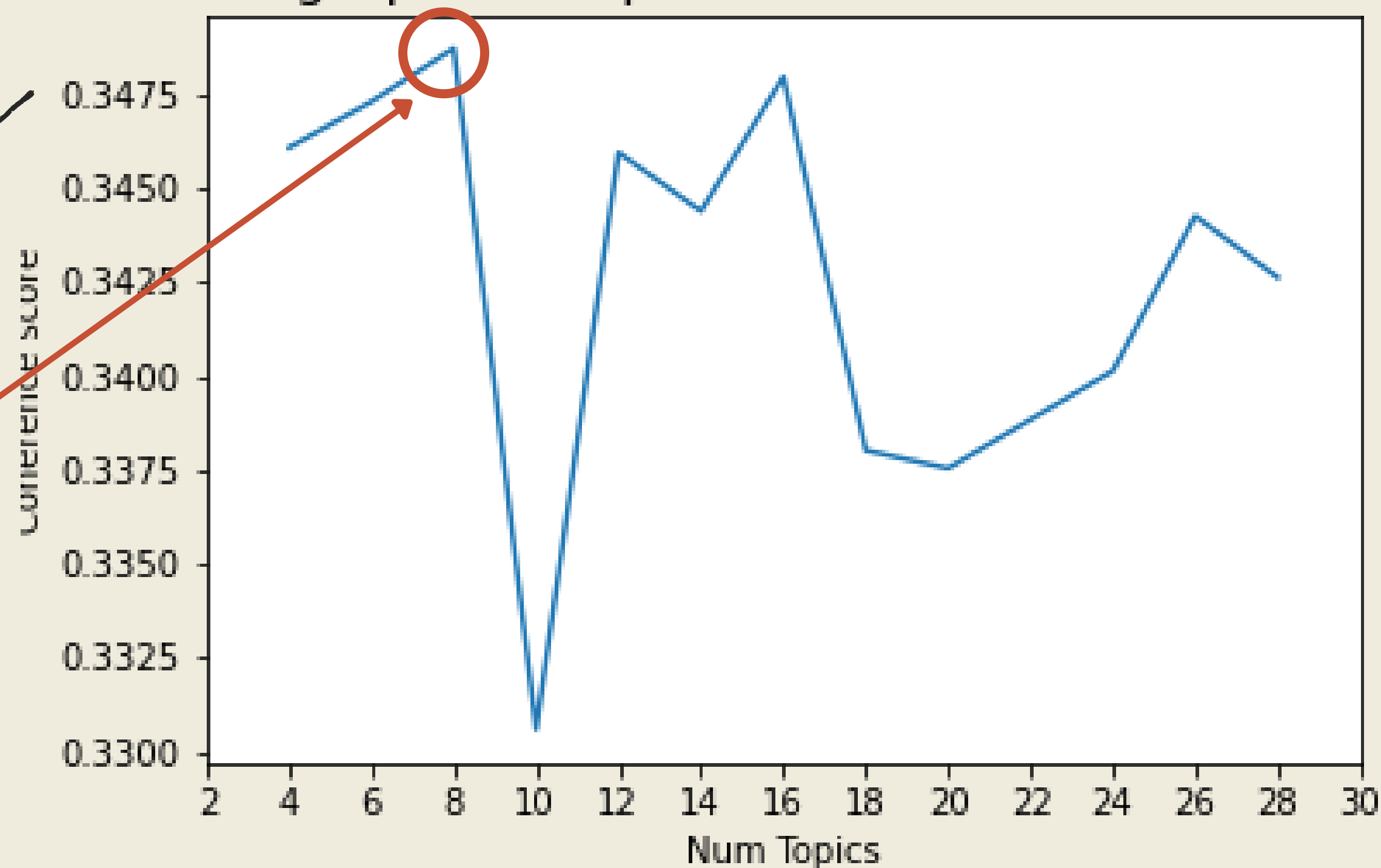
Matplotlib, pyLDavis

Tuning the

Model

Optimal
Number of
topics = 8

Choosing Optimal Topic Model with Coherence Scores



Topic 3:

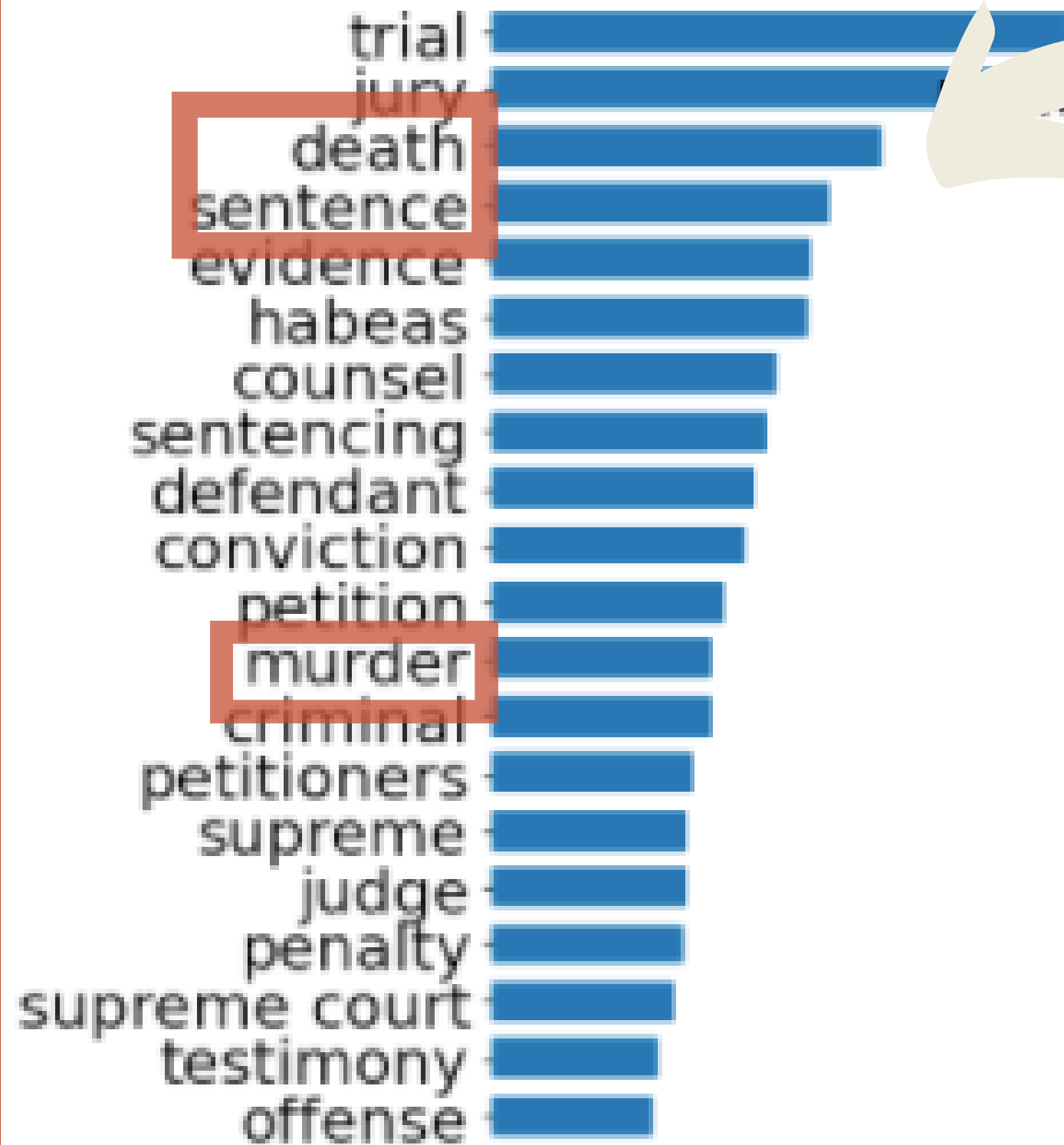
Unlawful Search and Seizure

4th Amendment: "The right of the people to be secure in their persons, houses, papers, and effects, against unreasonable searches and seizures..."

Top 20 Words for Topic 3



Top 20 Words for Topic 8



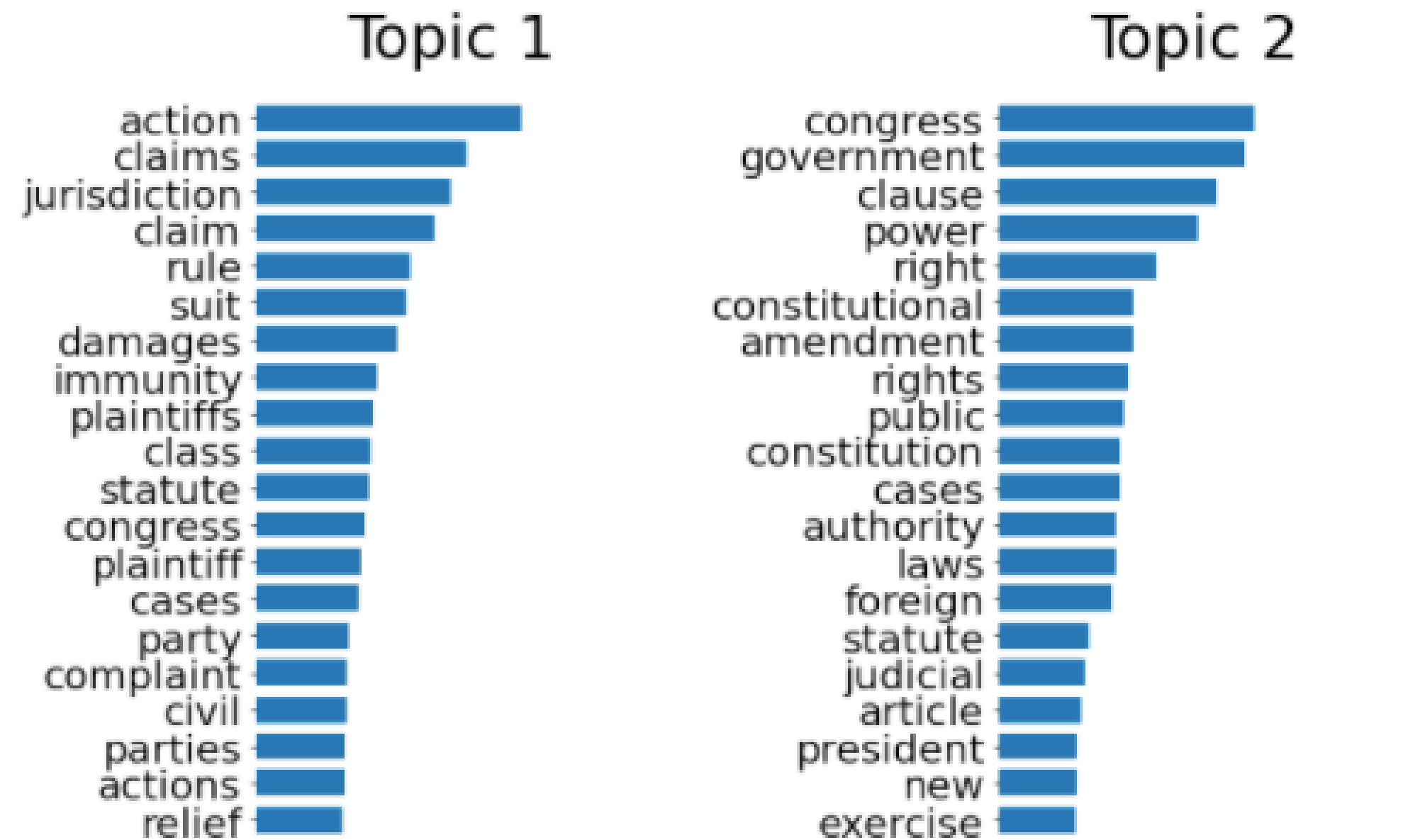
Topic 8

Eighth
Amendment:
Death Sentence

Topics 1 and 2:

Shows need
for model
improvement.

Top 20 Words for Topics 1 & 2



Future Work

Tuning and Modeling

Perform additional preprocessing to eliminate common legal terms.

Tune LDA model further and build topic models for specific time periods.

Analyze data through clustering techniques and linear regression with additional data

Thank you!