



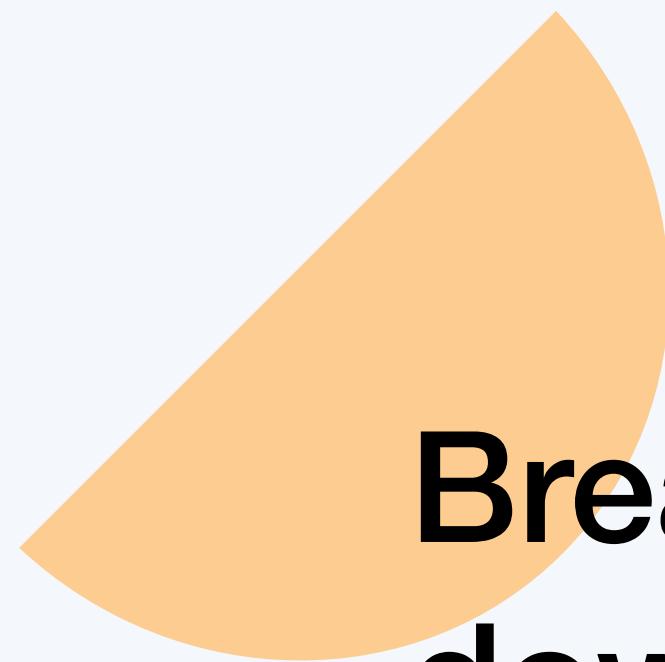
Text Analysis of NPR News Stories For ESL Learners

Leslie Cardone

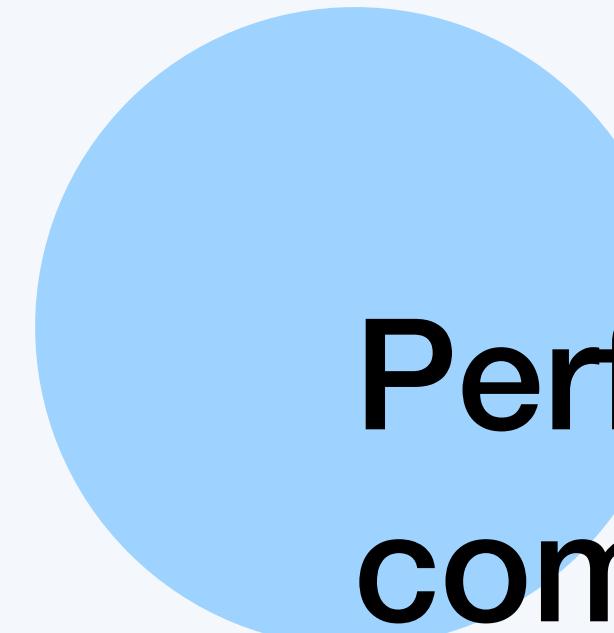
GOALS

How diverse are the topics that NPR reports on?

Are there trends in article text complexity?



Break NPR articles down into a manageable amount of topics using topic modeling



Perform a text complexity analysis by topic

DATA

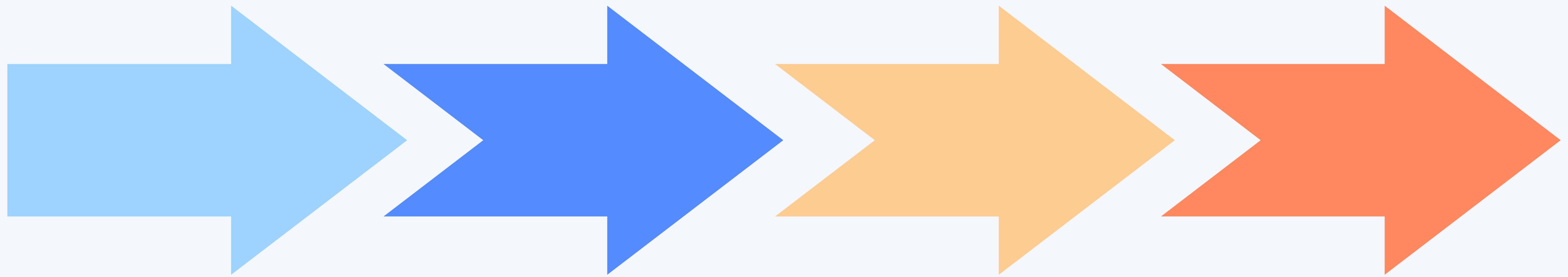
1,344 NPR news articles

6 months of stories

132 unique category tags



METHODS



Clean Up Text

**NLTK Sentence
Tokenizer**

CountVectorize

**Sci-kit Learn
Feature
Extraction**

LDA

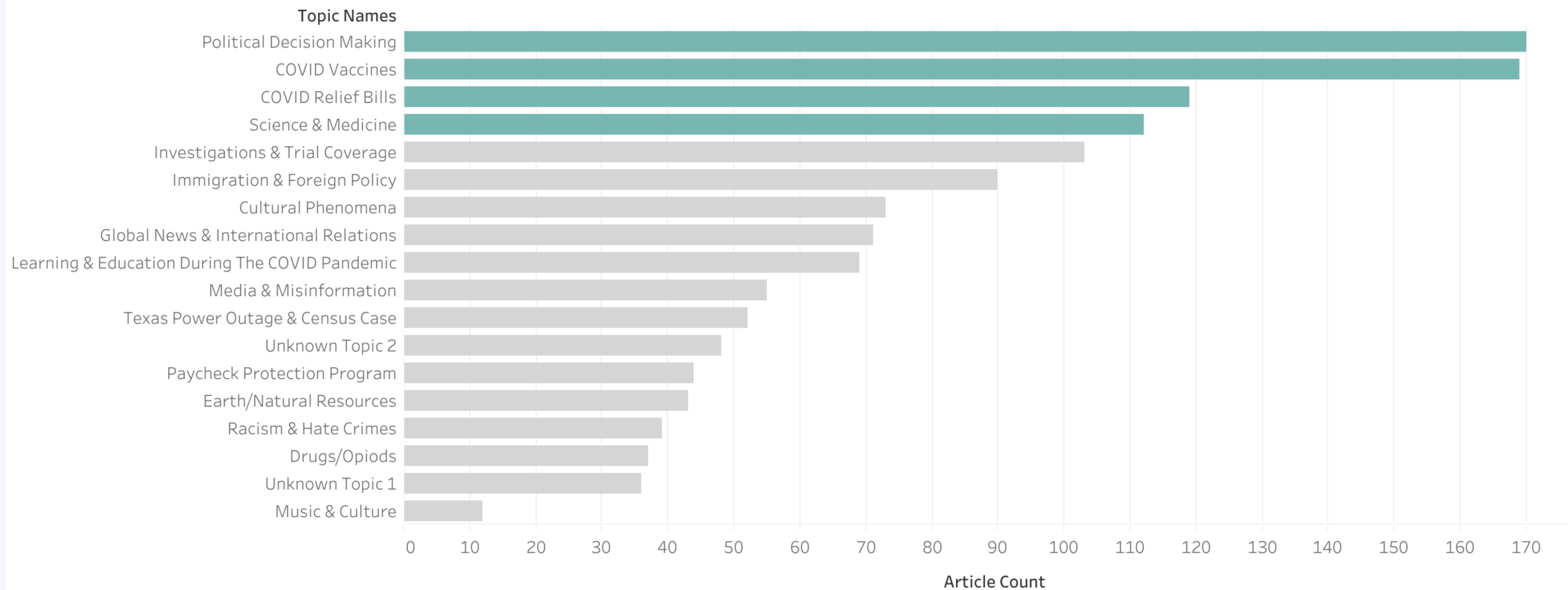
**Sci-kit Learn
Decomposition**

**Flesch Reading
Ease**

Textstat

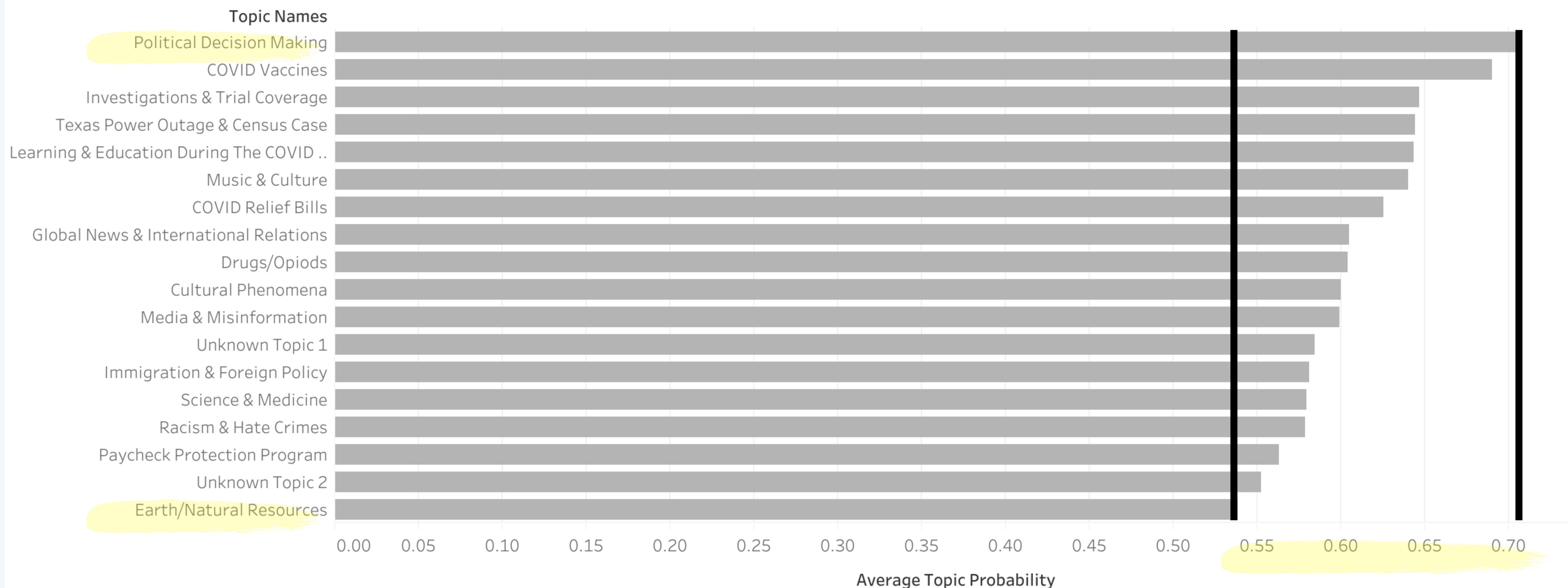
FINDINGS

Number of Articles Per Topic



FINDINGS

Average Topic Probability Across Articles Per Topic



Flesch Reading Ease

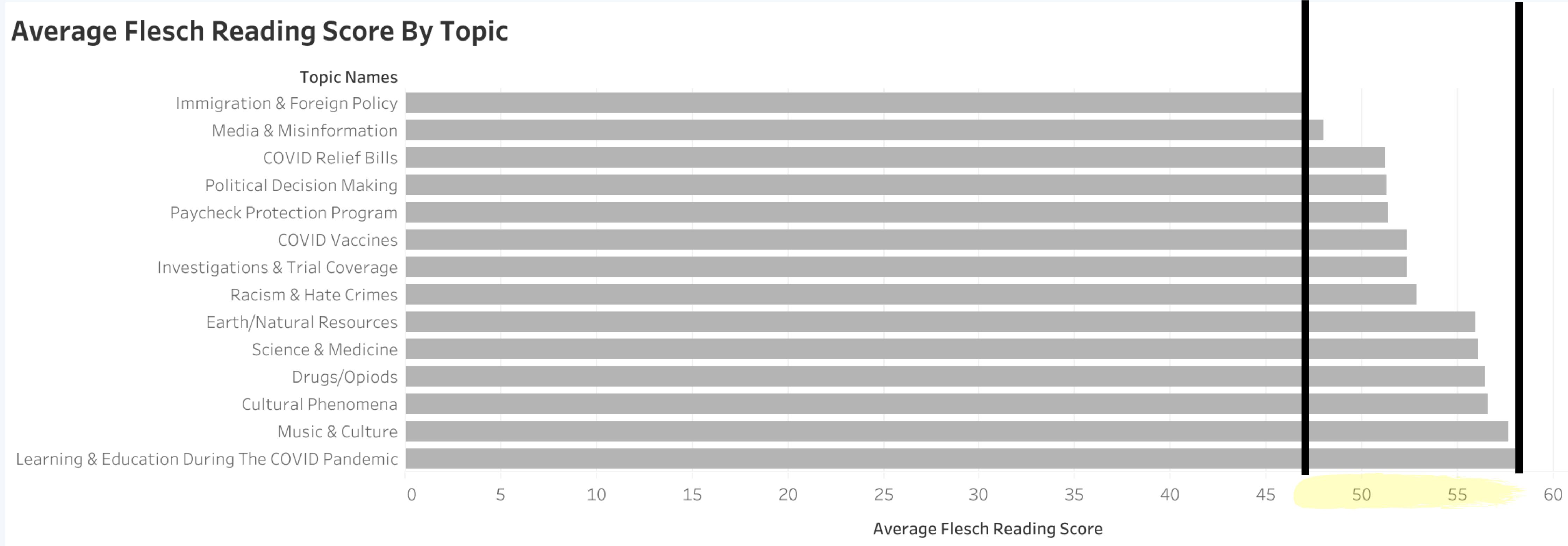
Score	Grade	Avg. Words Per Sentence	Syllables Per 100 words
90 - 100	5	8	123
80 - 90	6	11	131
70 - 80	7	14	139
60 - 70	8 - 9	17	147
50 - 60	10 - 12	21	155
30 - 50	College	25	167
0 - 30	College Grad	29	192



CONTENTWRITERS.COM

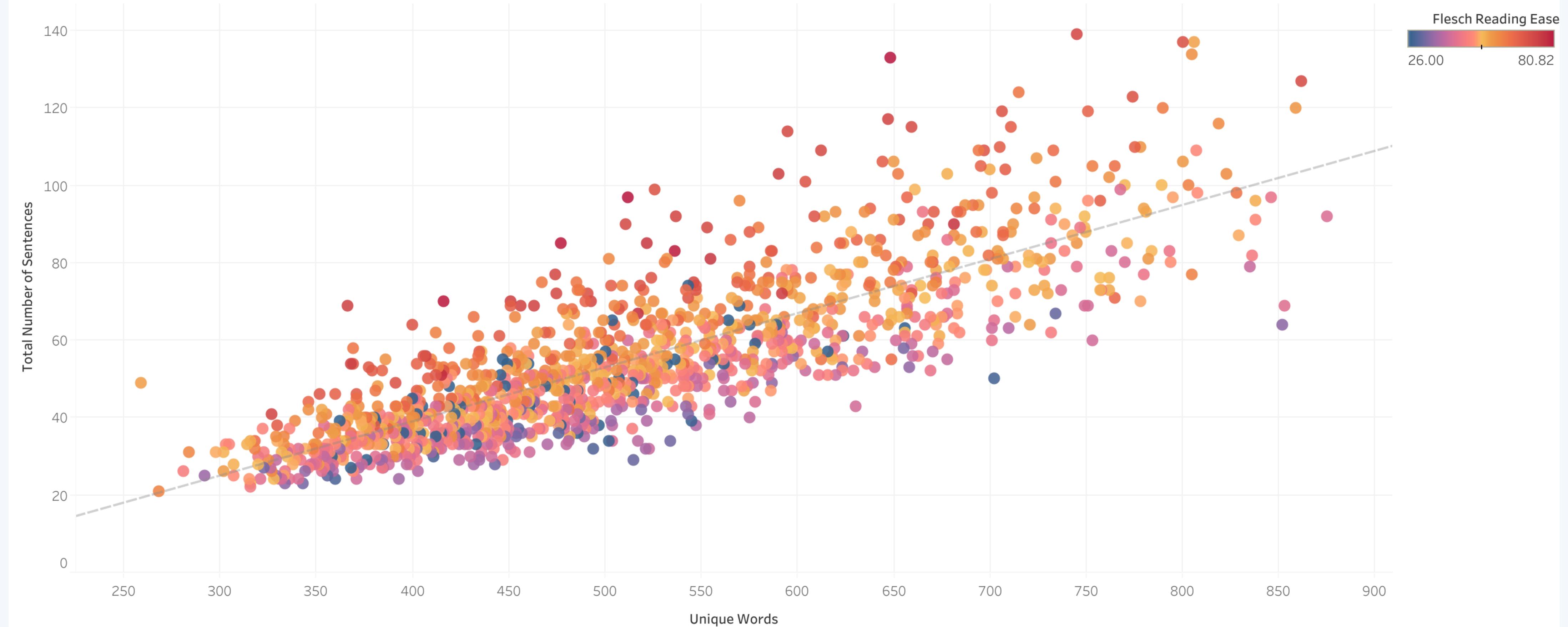


FINDINGS



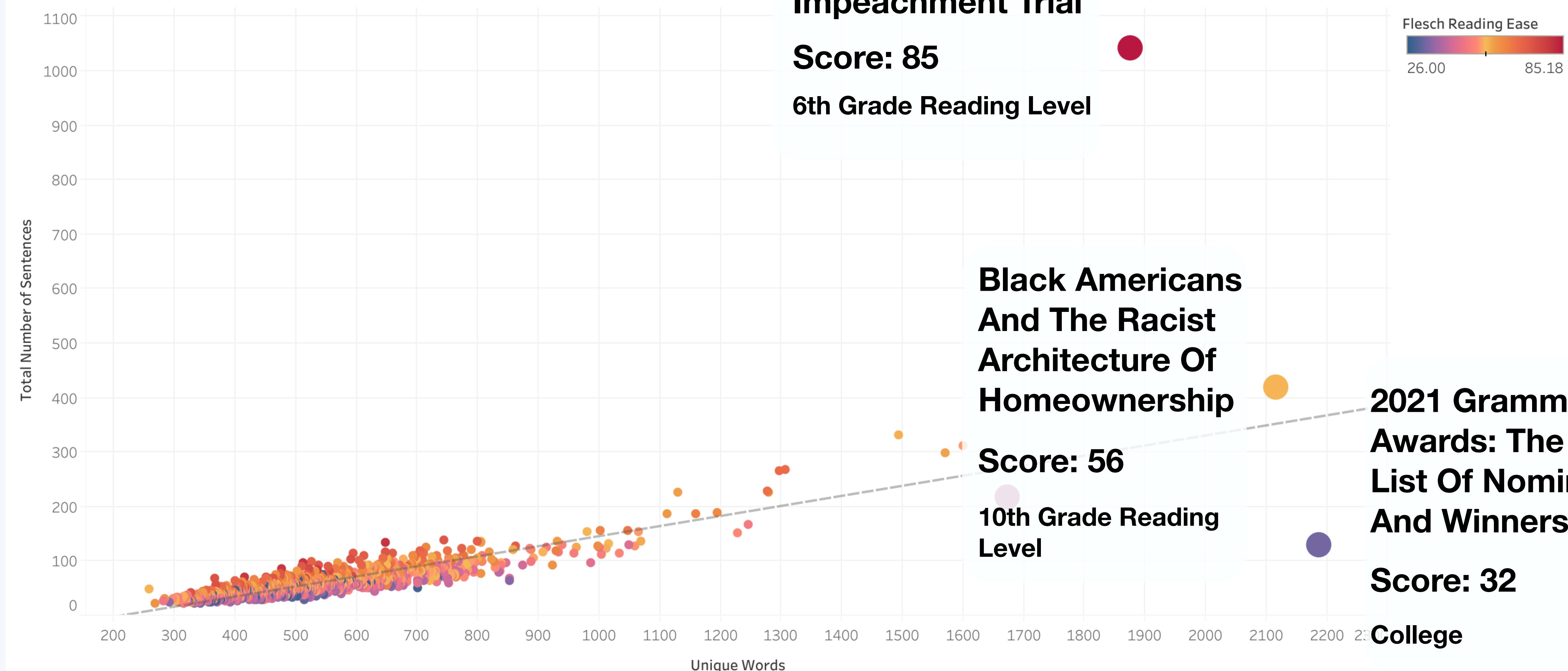
FINDINGS

Unique Words & Number of Sentences Vs Flesch Reading



FINDINGS

Unique Words & Number of Sentences Vs Flesch Reading



Read Trump's Jan.
6 Speech, A Key
Part Of
Impeachment Trial

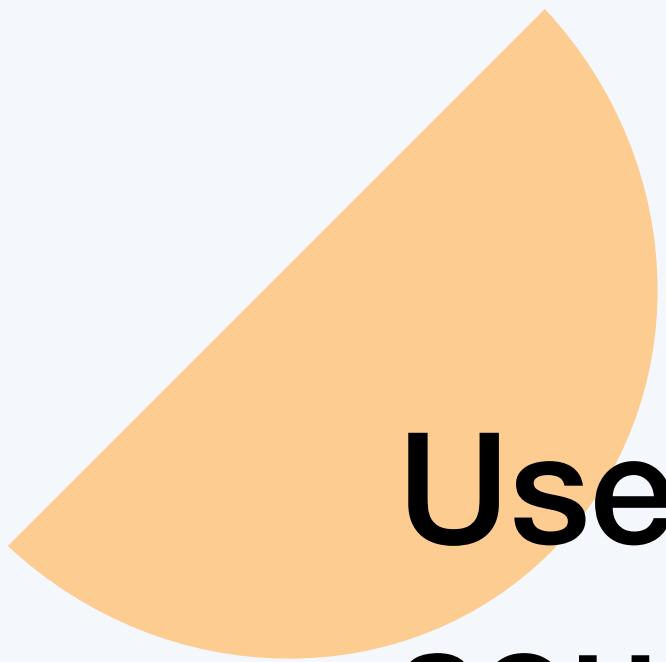
Score: 85
6th Grade Reading Level

Black Americans
And The Racist
Architecture Of
Homeownership

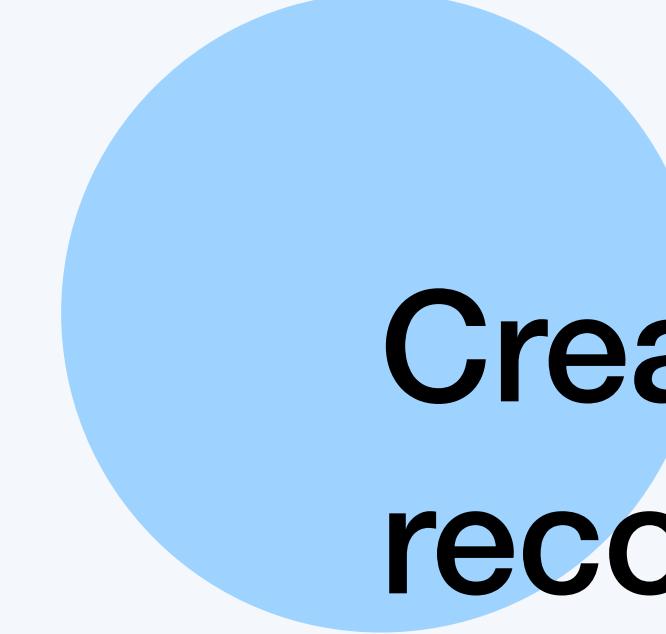
Score: 56
10th Grade Reading
Level

2021 Grammy
Awards: The Full
List Of Nominees
And Winners
Score: 32
College

FUTURE WORK



Use more diverse news sources / access a wider “reading level” and “topic” range



Create a recommendation system for ESL learners