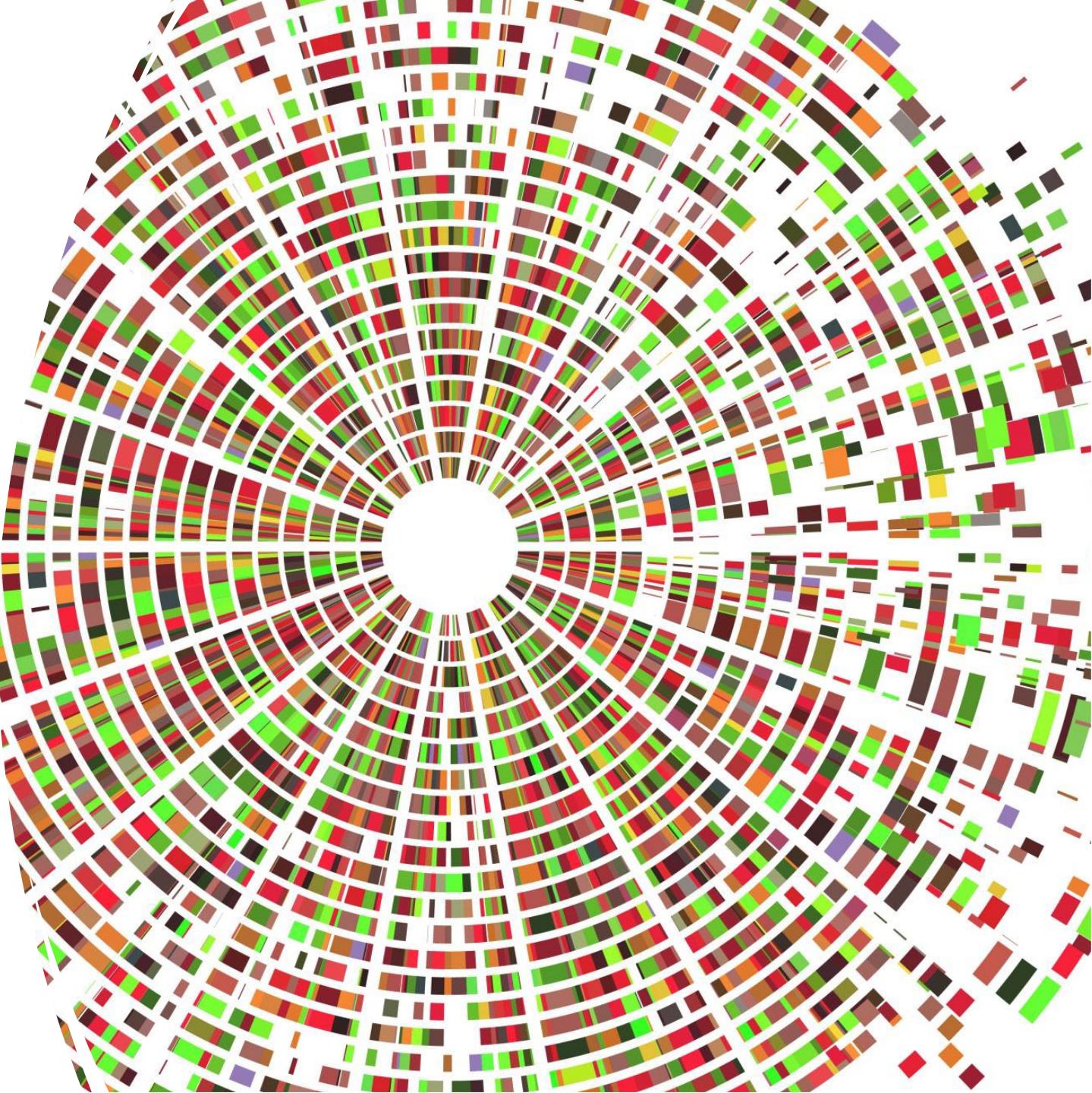


THE RISE OF FAKE NEWS IN THE POLITICAL SCENE

By Michael Mahoney





PRES

ENTATION OVERVIEW

- Quick intro to the data
- Talk about misinformation and what we found
- Explore some recommendations
- Talk about the future

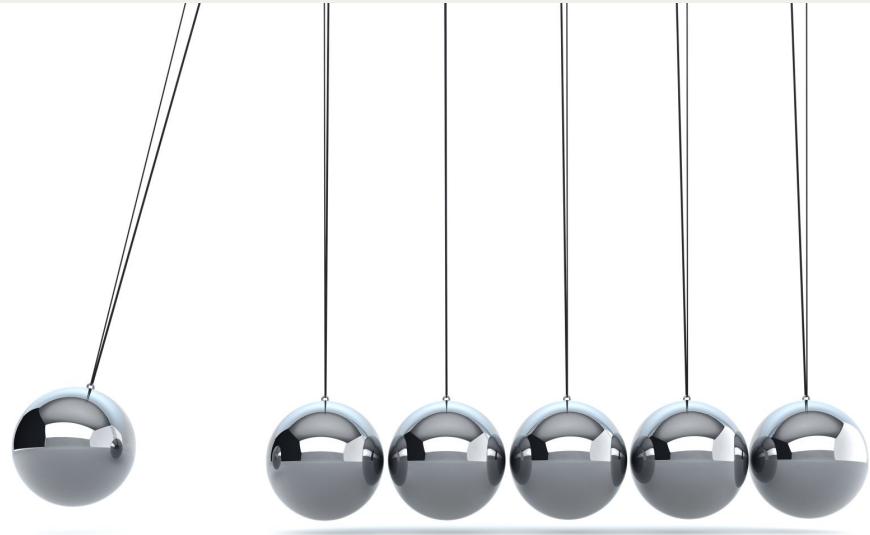


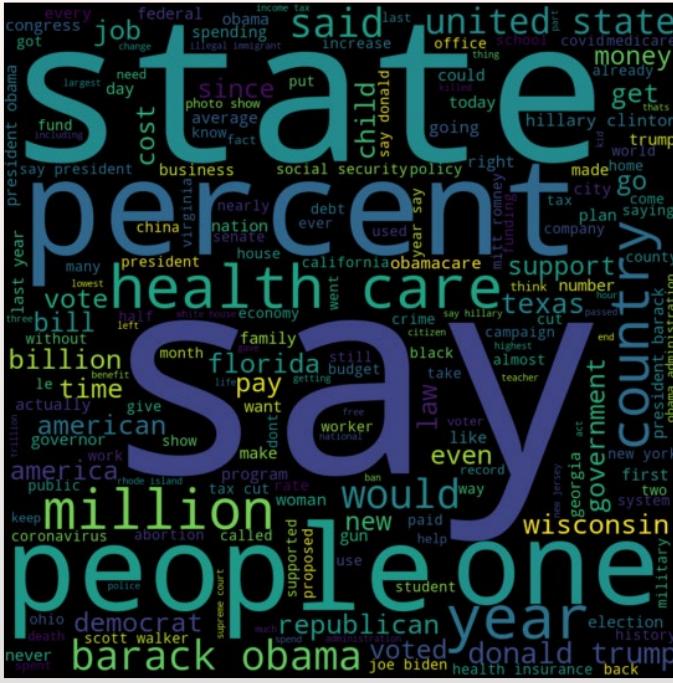
WHO IS THIS PRESENTATION FOR



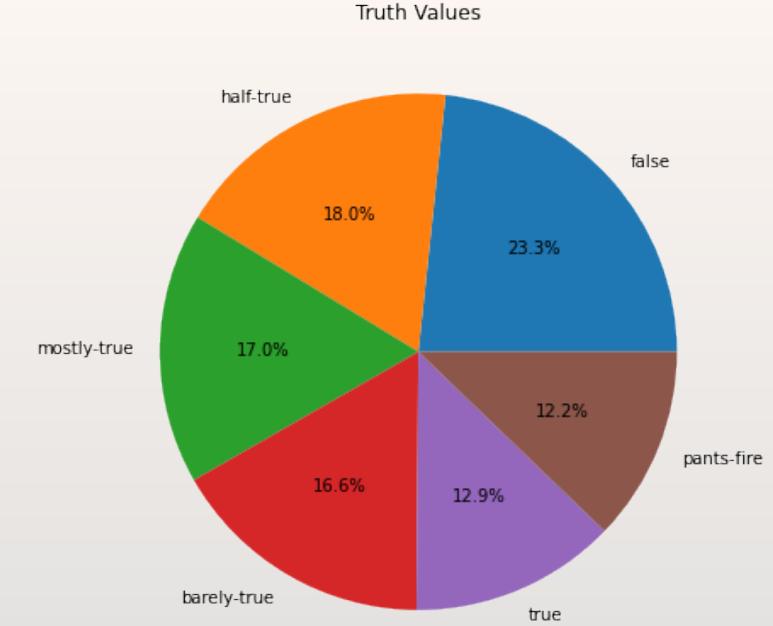
MISINFORMATION

- The Cost 79 Billion USD
 - 37 billion USD in losses on the stock market
 - 17 billion USD in losses annually from false/misleading financial information
 - 954 billion USD spent on reputation management for false information about companies.
 - 3 billion USD has been already spent to combat fake news.
 - 400 million USD spent by US public institutions defending against misinformation.
 - 250 USD million spent on brand safety measures.





OUR DATA

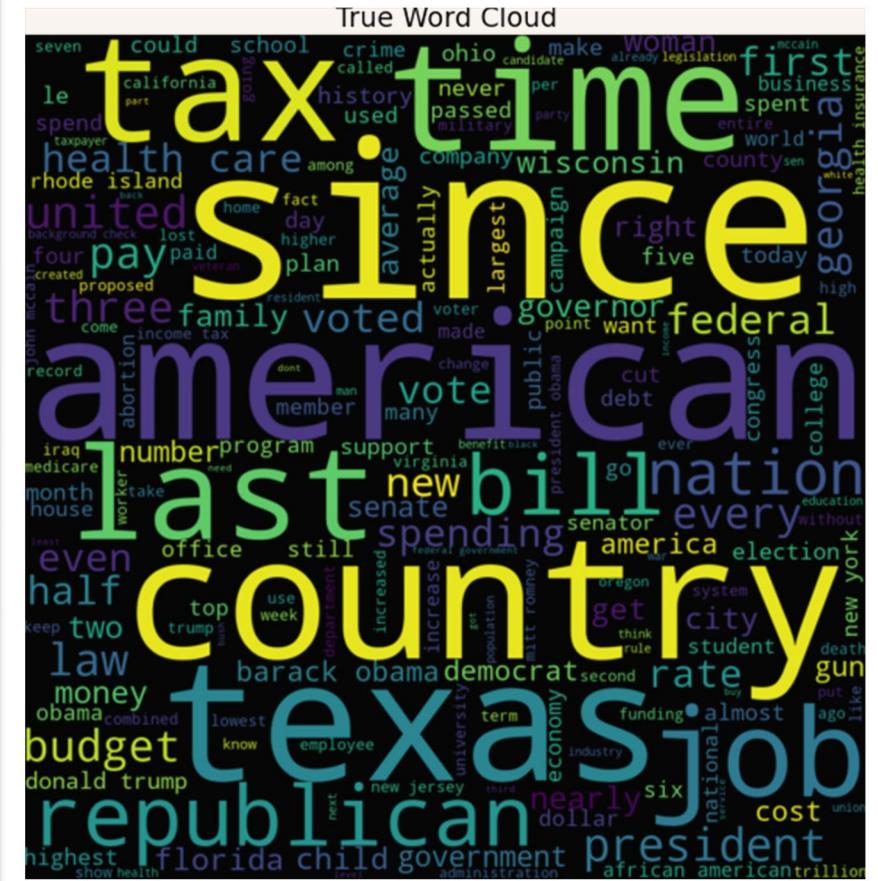
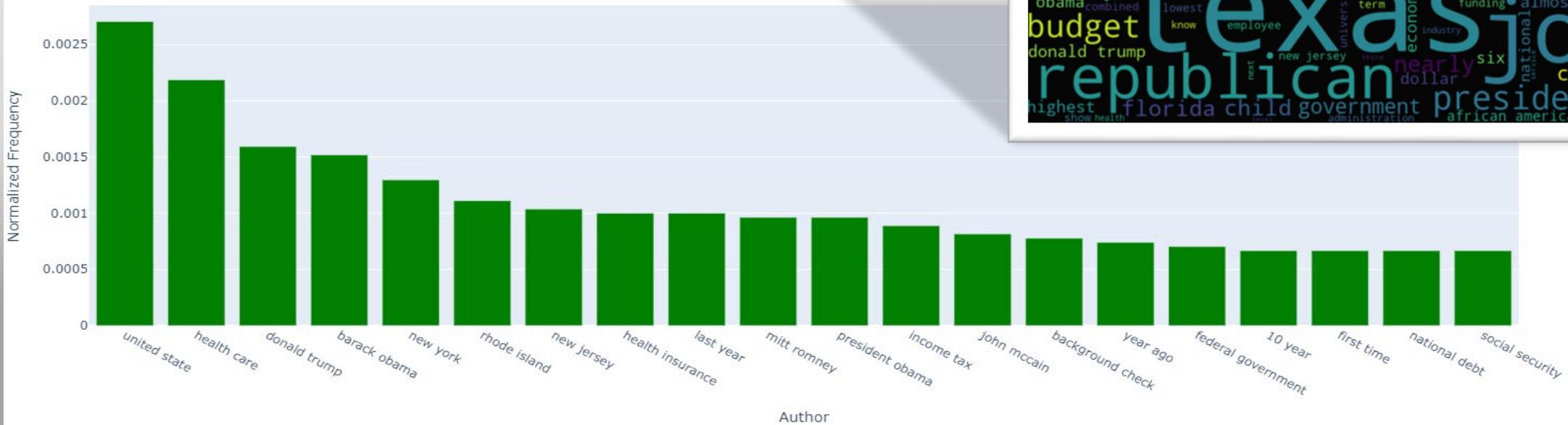


TRUE

Word pairs that are unique in the top 25:

- Per capita
- West Virginia
- National debt
- 10 year
- John McCain
- Rhode Island

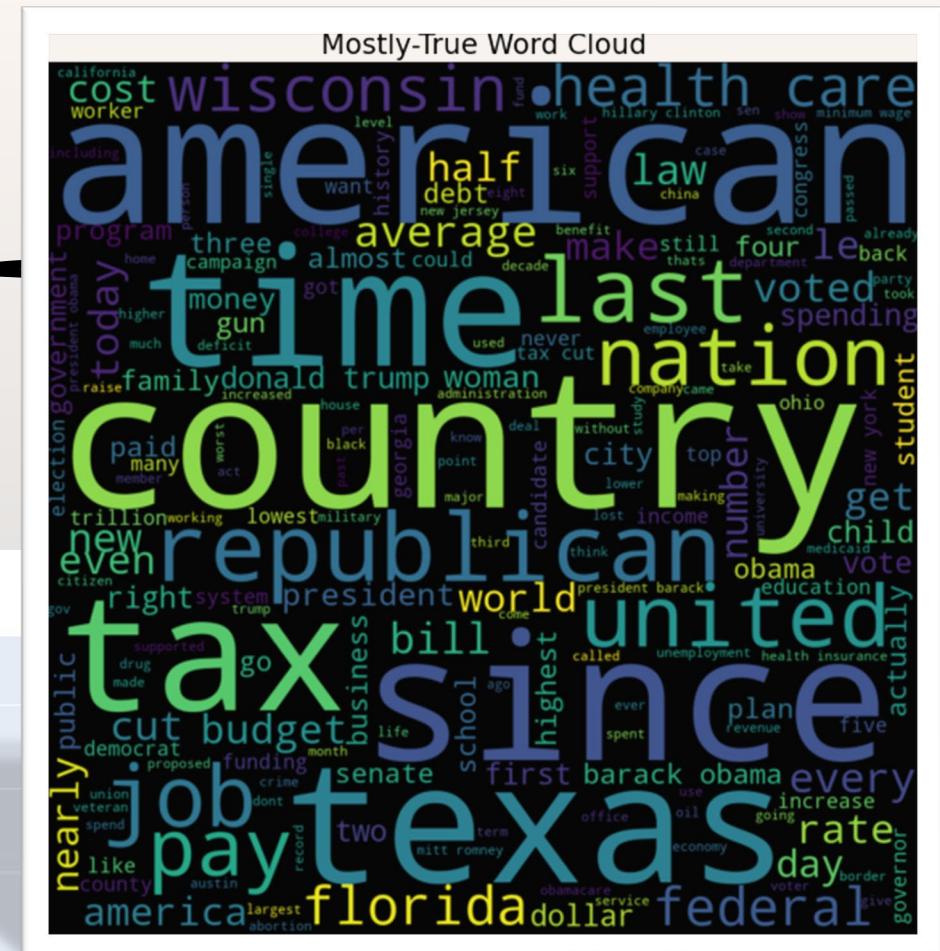
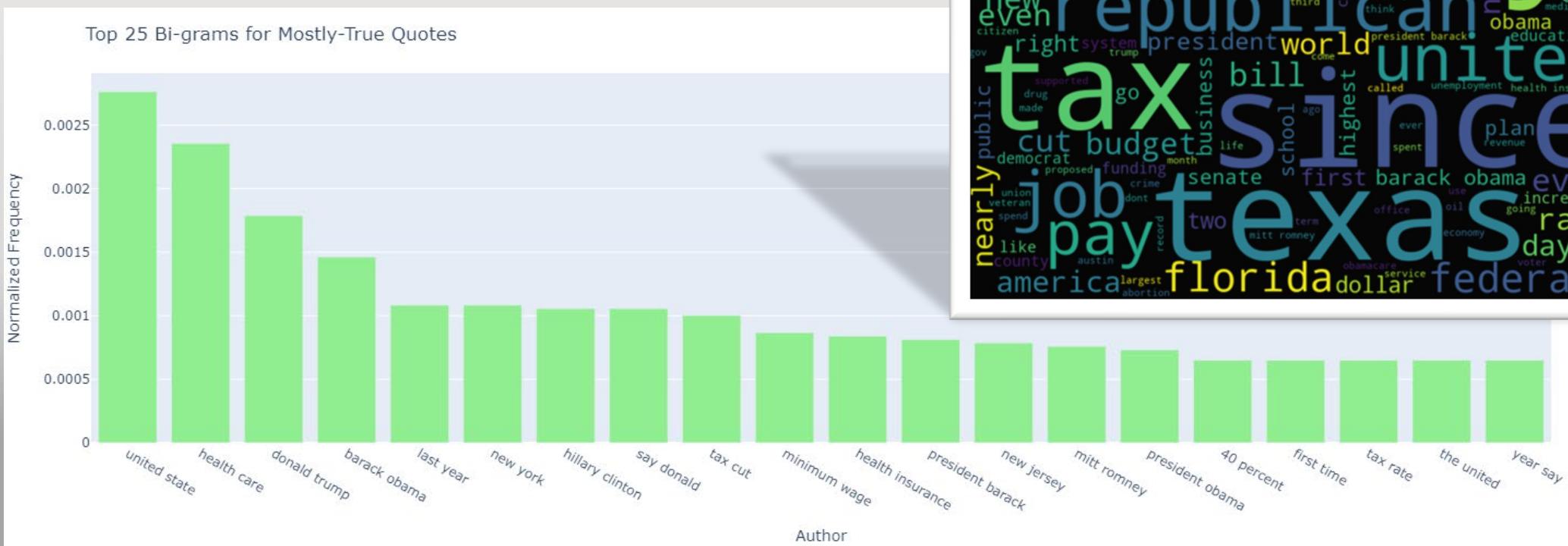
Top 25 Bi-grams for True Quotes



MOSTLY-TRUE

Word pairs that are unique in the top 25:

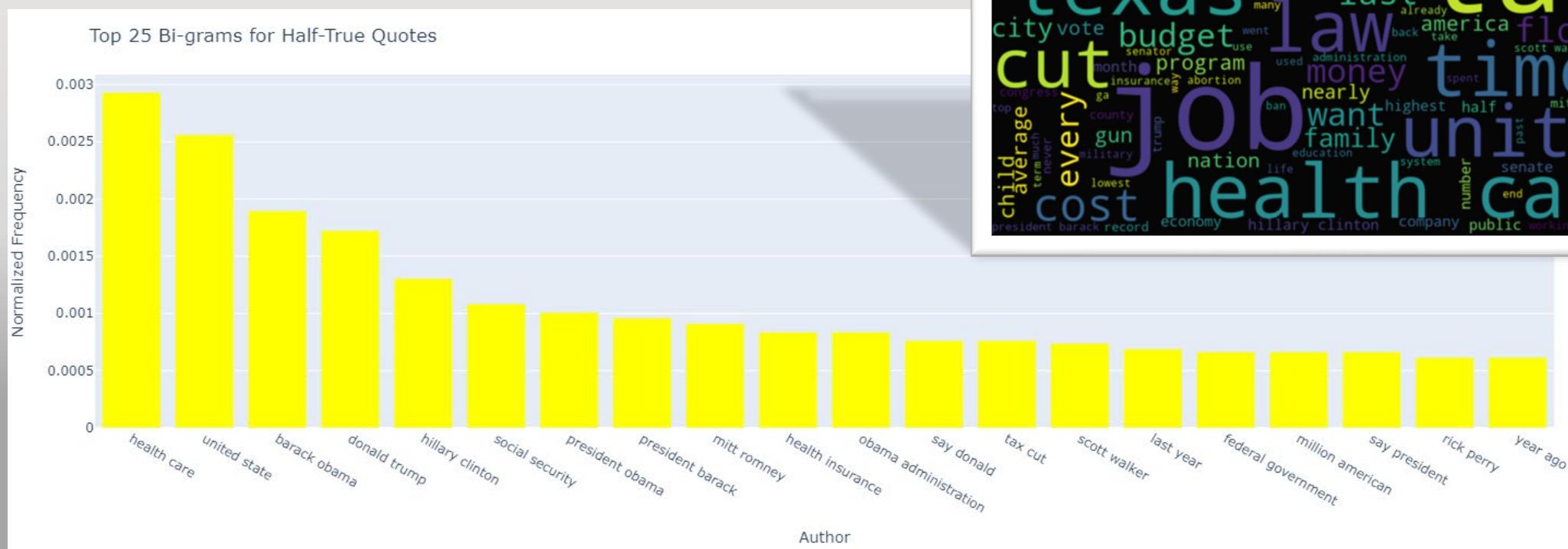
- 50 percent
 - 1 percent
 - Million people
 - Tax Rate



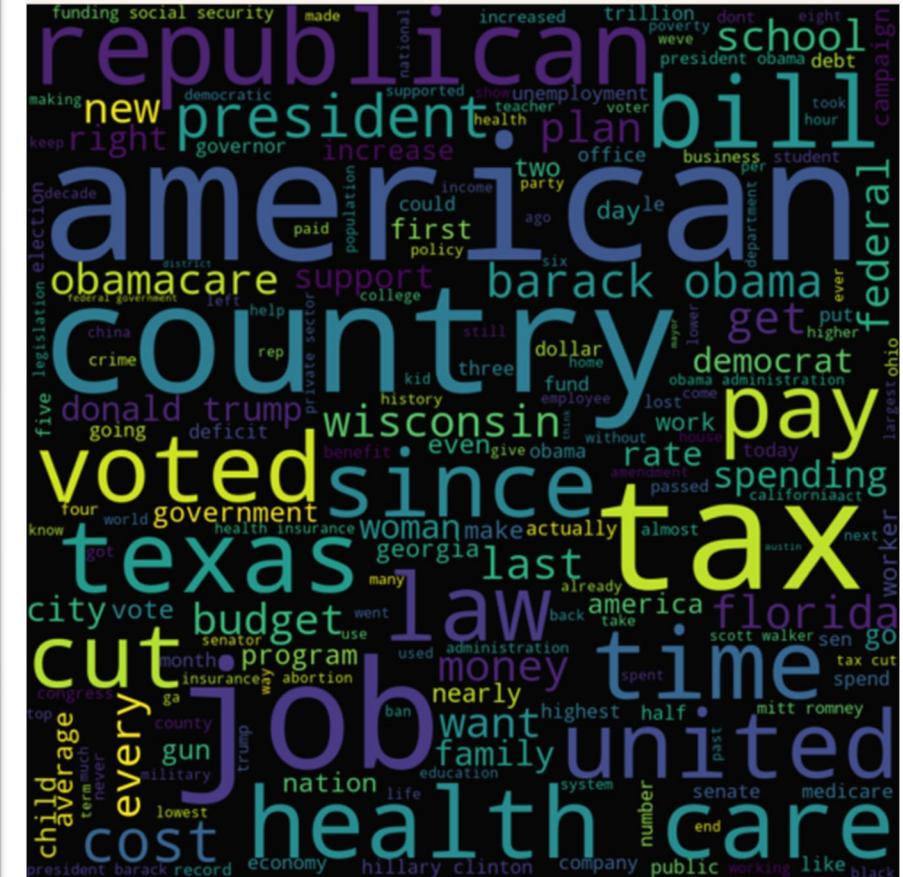
HALF-TRUE

Word pairs that are unique in the top 25:

- Cares act
 - Unemployment rate
 - Million jobs
 - Rick Perry
 - Million Americans



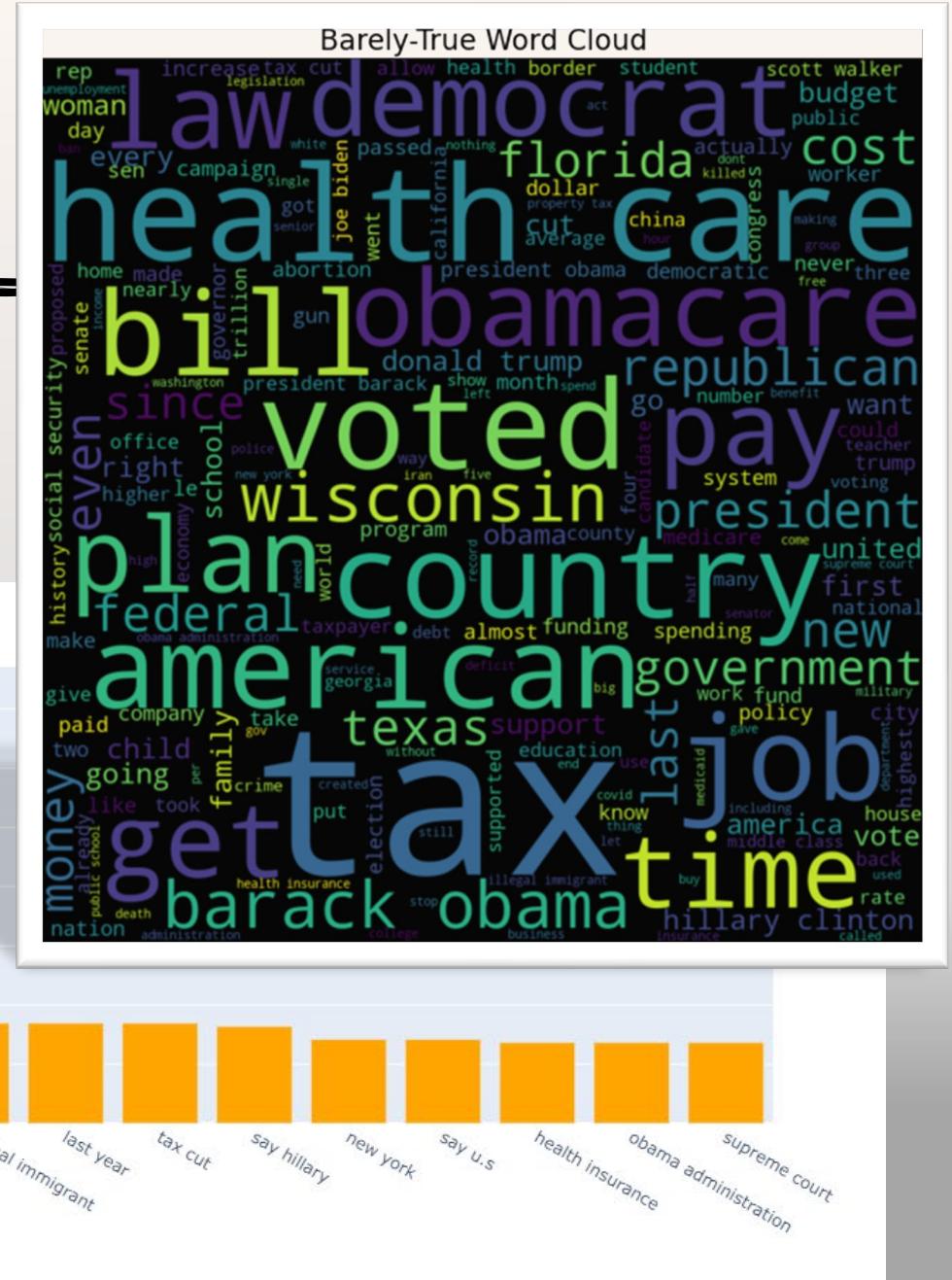
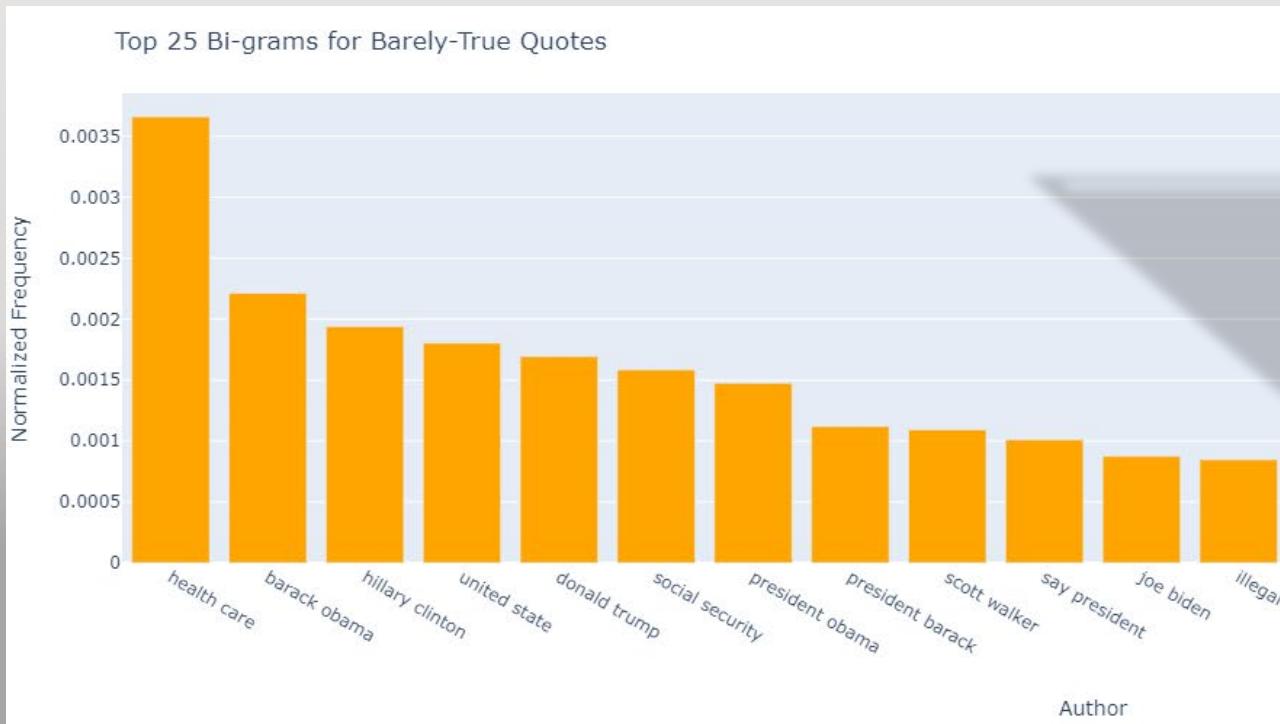
Half-True Word Cloud



BARELY-TRUE

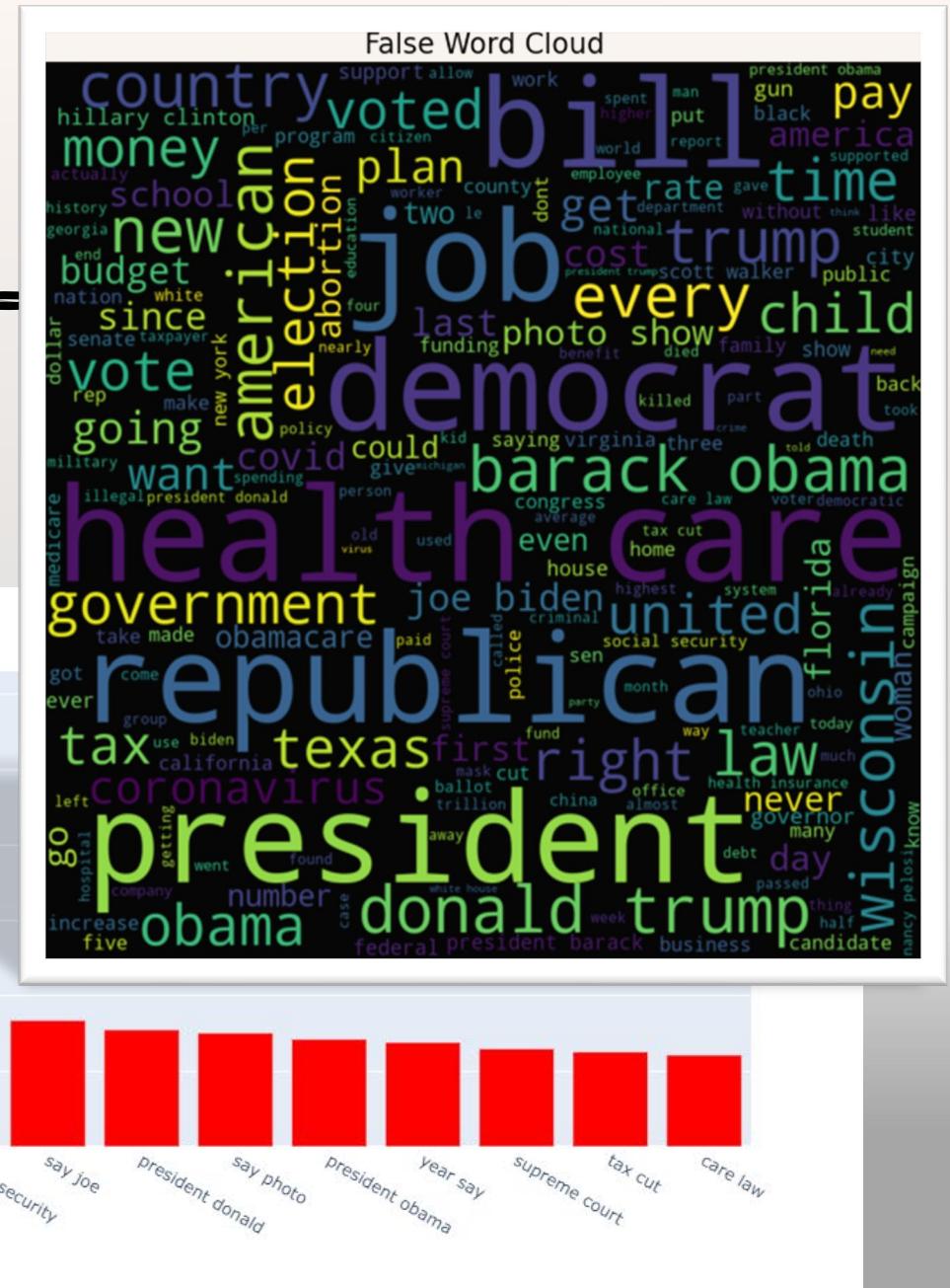
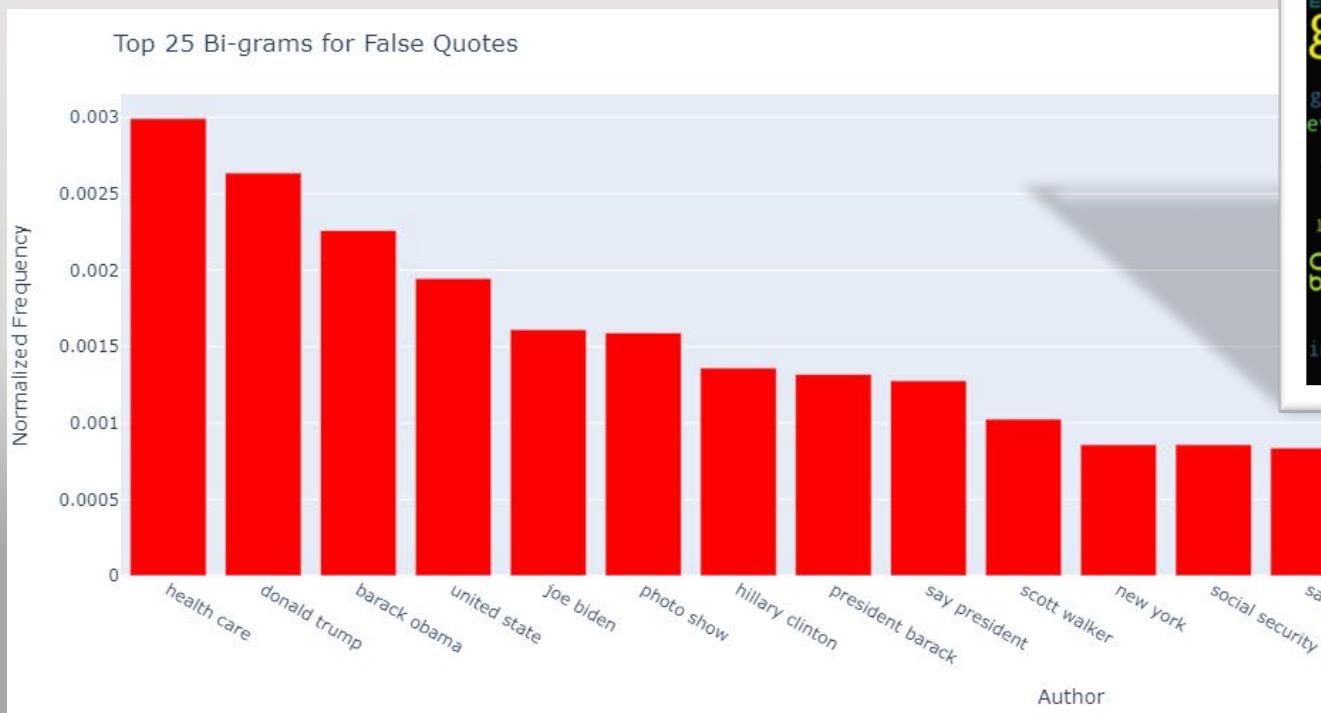
Word pairs that are unique in the top 25:

- Rick Scott
 - Property tax
 - Public school



FALSE

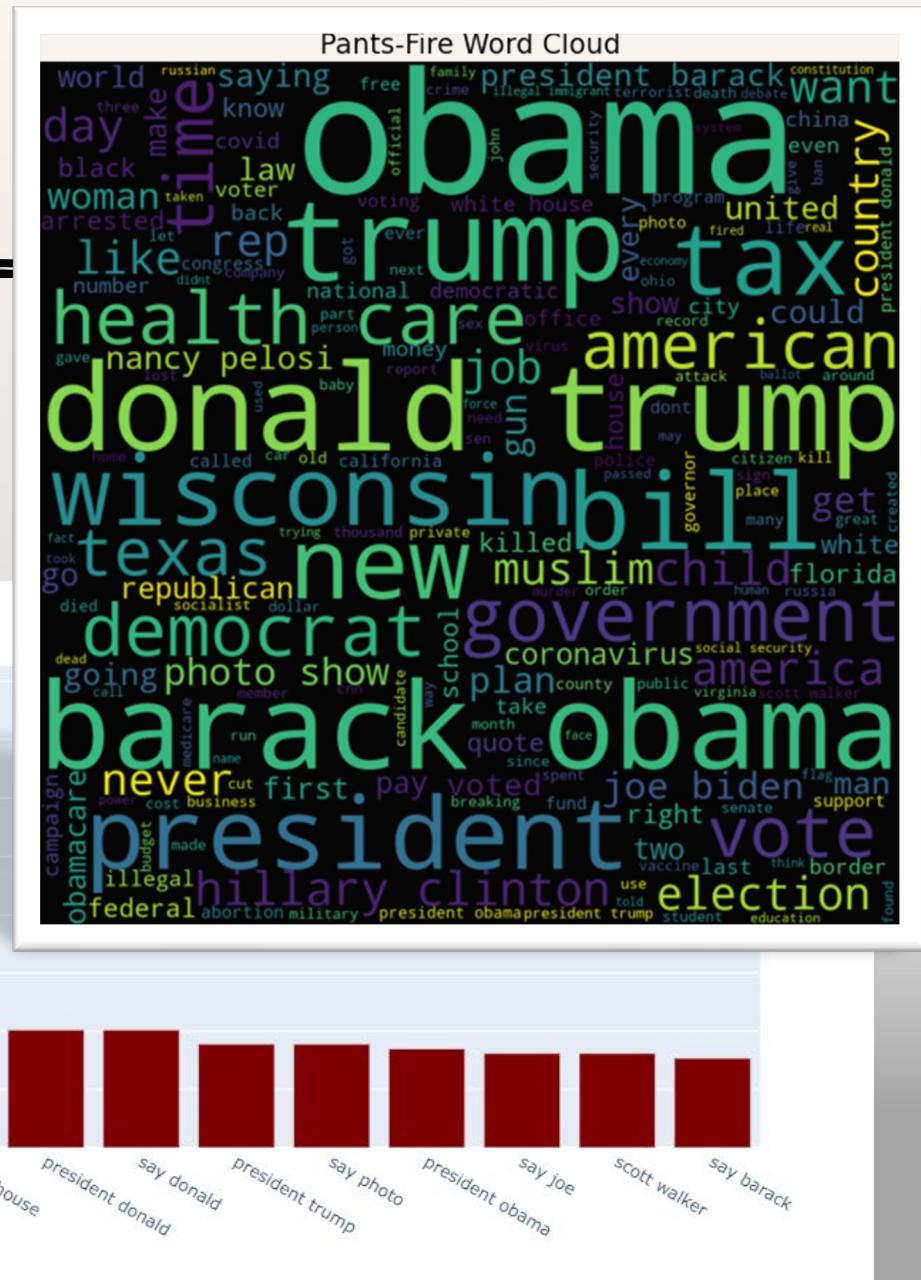
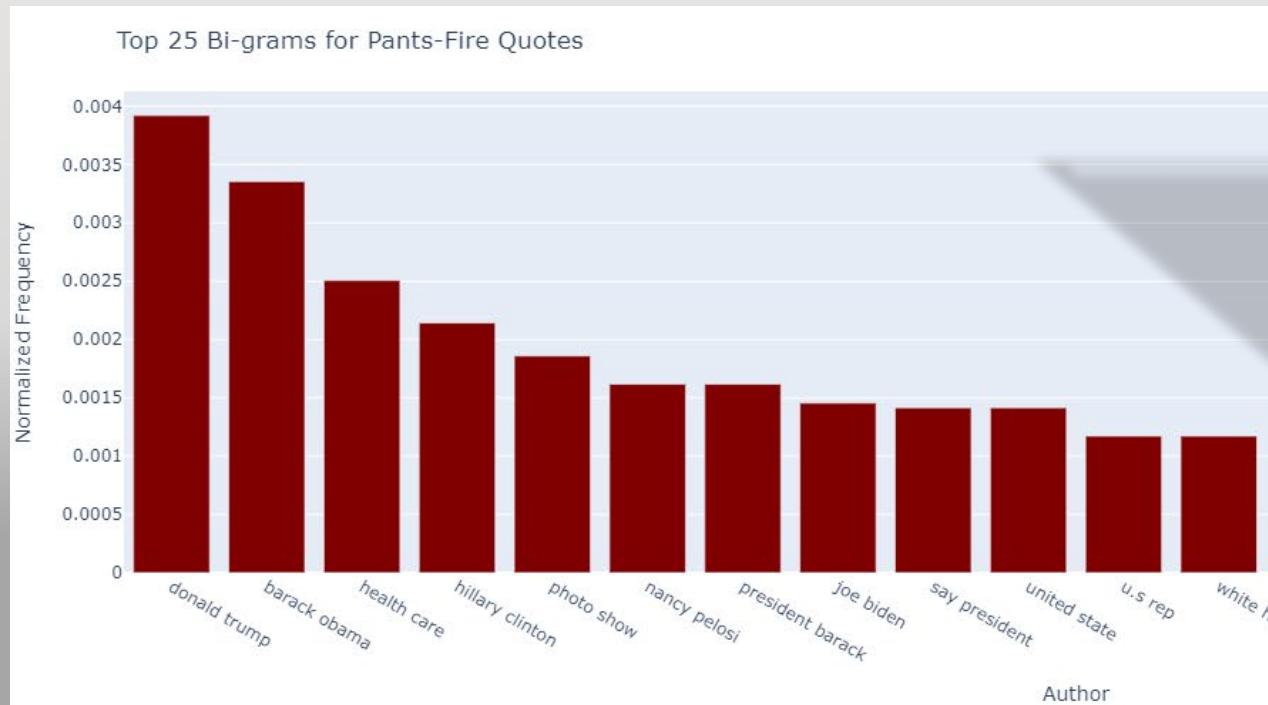
Has no unique word pairs in the top 25



PANTS-FIRE

Word pairs that are unique in the top 25:

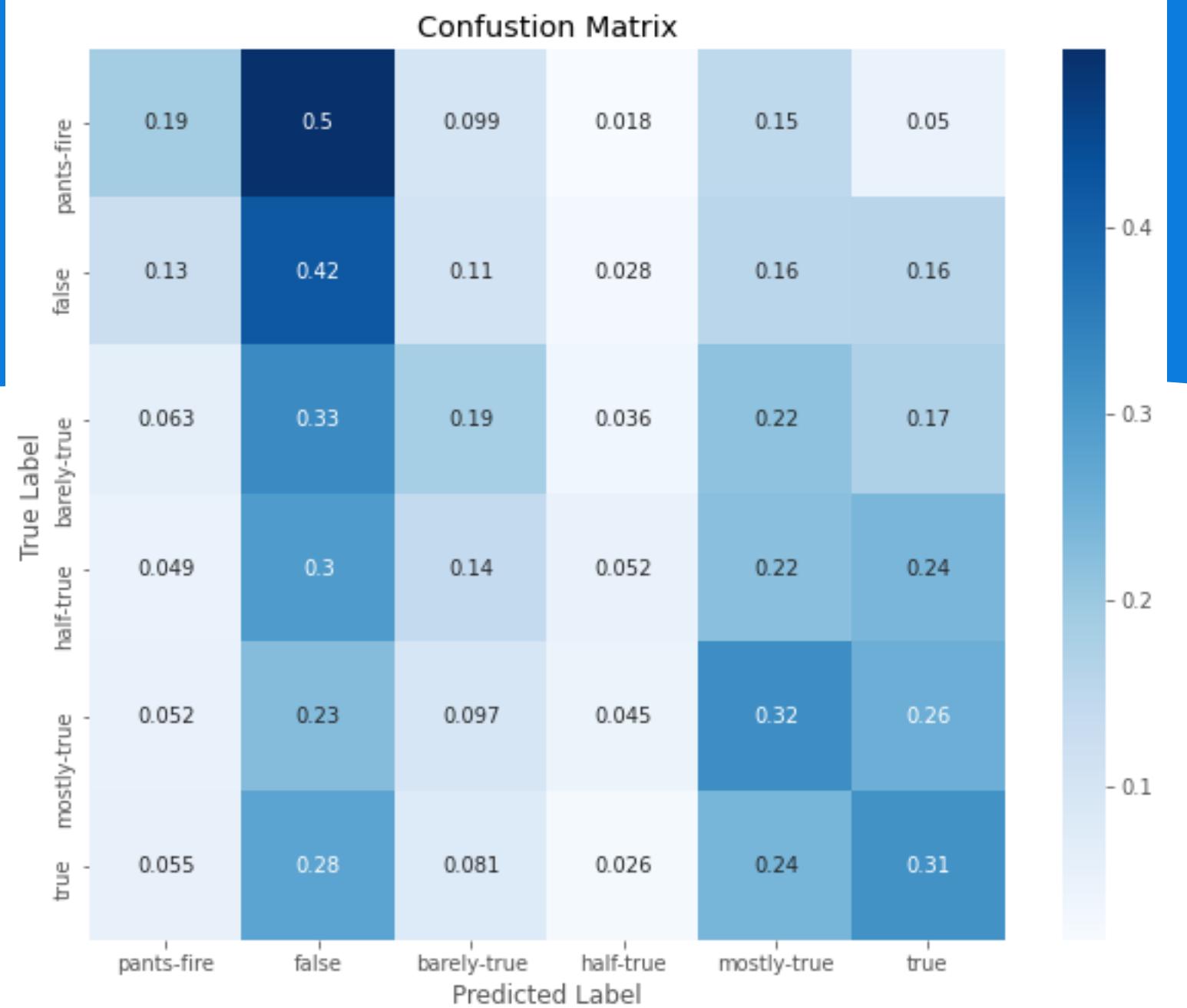
- George Soros
 - Says Barack
 - Nancy Pelosi



MODELS

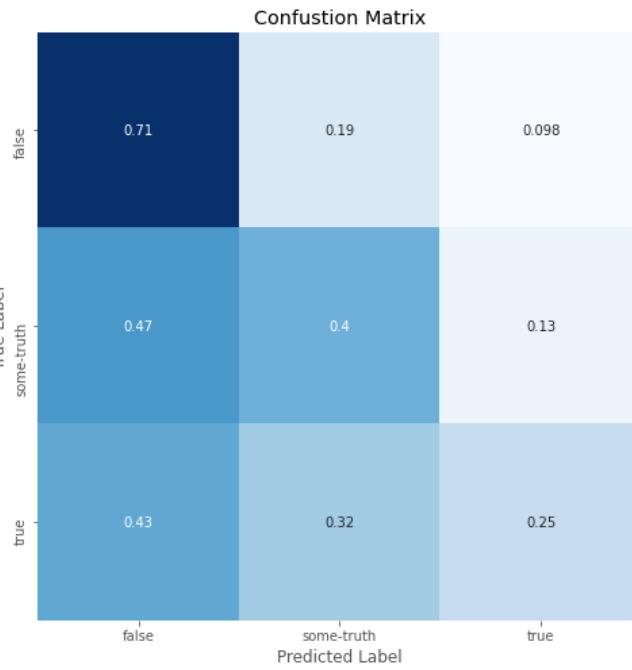
LSTM with Non-Lemmatized Tokens

Accuracy: 26%

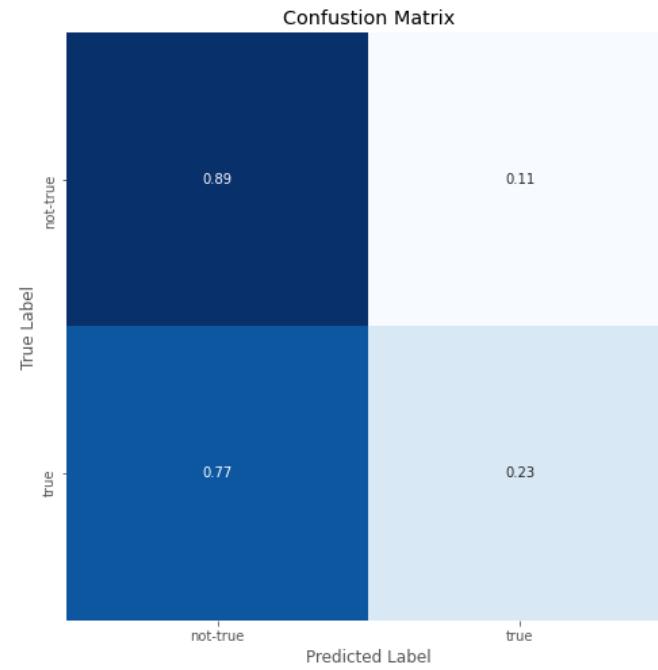


MODELS

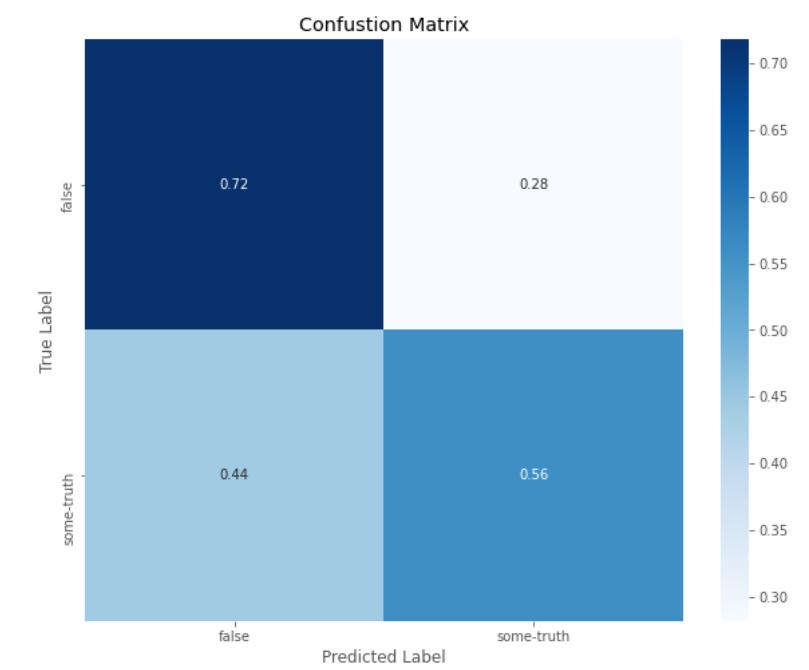
- LEFT Model:
 - Pants-fire & False mapped to false
 - True mapped to true
 - Everything else mapped to some-truth



- MIDDLE Model:
 - True mapped to true
 - Everything else mapped to not-true



- RIGHT Model:
 - Pants-fire & False mapped to false
 - Everything else mapped to not-false

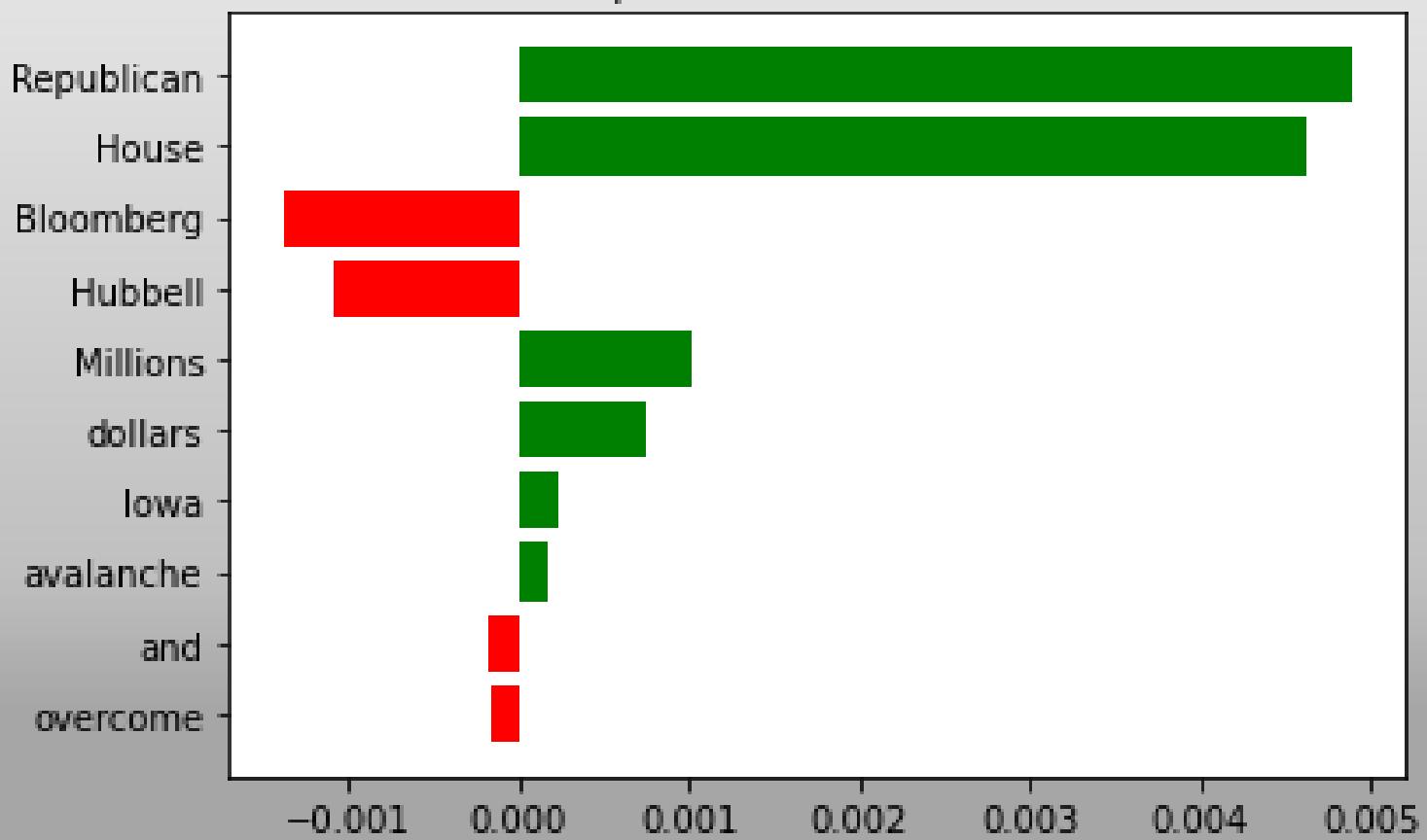


EXAMPLE CLASSIFICATION

Half-True Example

Millions of Hubbell and Bloomberg dollars wasn't enough to overcome the Iowa House Republican avalanche of success.'

Local explanation for class half-true



A photograph of a hand holding a magnifying glass. The lens of the magnifying glass is focused on a vibrant, colorful city skyline at night, showing numerous illuminated billboards and signs. The background is filled with a soft-focus, bokeh effect of various city lights in shades of yellow, orange, red, and blue. The overall composition suggests examining or zooming into a specific aspect of a larger, complex environment.

WHAT TO DO ABOUT IT?

Resource Dedication

- For those wishing to devote resources to fight misinformation; Mentions of photo content were only mentioned in the top 25 word pairs for the false and pants-fire categories. Flagging content with mentions of photos (not necessarily photos themselves) will lead to capturing more mis-information.



A photograph of several light-colored wooden blocks stacked in a stepped, staircase-like arrangement. The blocks are rectangular with visible grain patterns. One block is partially pulled out from the stack, showing its underside and the texture of the wood grain.

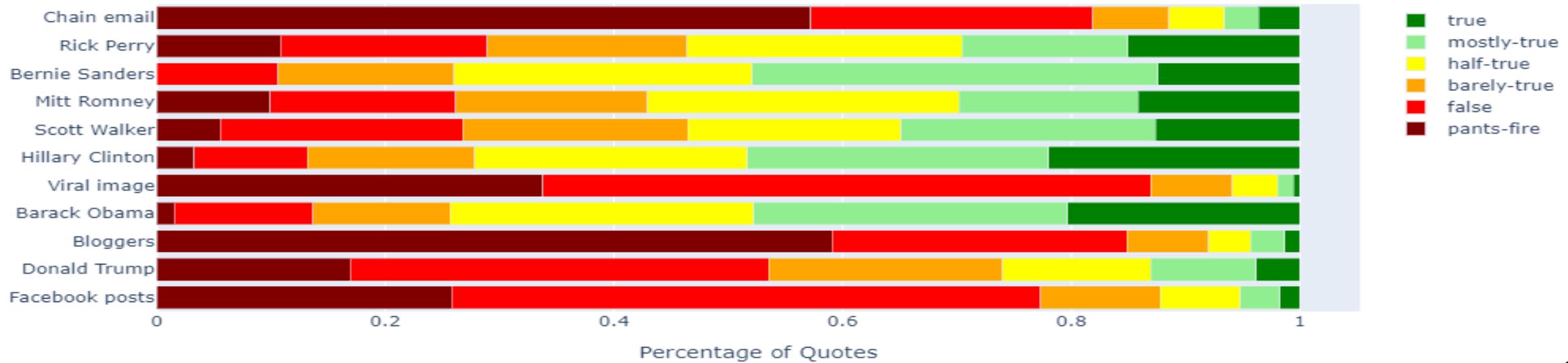
Politics

- For politicians, people and not so much issues find themselves as common associations with misinformation (with the notable exception of health care). If you wish to advance legislation, decoupling the issue from individuals will the best way to limit misinformation's impact.

Perpetuating Information

- For Journalists, bloggers and people at large: Email chains, unattributed social media posts and viral content should never be used as a source for content.

Quotes by Author: Percentages





FUTURE WORK

- I would like to bring in some more quotes.
- The second modeling goal is to implement a version of a BERT transformer.
- As for the data exploration; I would like to make more use of the various NLTK apis for text engineering.
- I would like to make the data more interactive on the app such that users can select specific quotes and interact with them through the various text models.

REFERENCES

- Politifact.com They are awesome and you should consider donating!
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin: Attention is All you Need. NIPS 2017: 5998-6008
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL-HLT (1) 2019: 4171-4186
- Keras Developers: Keras API Docs: <https://keras.io/api/>
- Ployly Developers: Plotly API Docs: <https://dash.plotly.com/>
- Scikit Learn Developers: Scikit Learn API Docs: <https://scikit-learn.org/stable/modules/classes.html>
- Pandas Developers: Pandas API Docs: <https://github.com/pandas-dev/pandas>
- Jacob Kessler: ScatterText: <https://github.com/JasonKessler/scattertext>
- Marco Tulio Correia Ribeiro: Lime: <https://github.com/marcotcr/lime>
- Radim Rehrek, Petr Sojka: Software Framework for Topic Modelling with Large Corpora: Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks: p. 45 - 50 : 2010: ELRA: Valletta, Malta
- Andreas Mueller: word-cloud: https://github.com/amueller/word_cloud

THANK YOU

