

# What's New(s)?

Creating an interest-based  
news subscription  
recommender system  
using topic modeling

● Julia Qiao



# More time at home

+

# 24/7 news cycle

—

# Information Overload



How can we  
be more  
selective  
about our  
news intake?



# Process

1

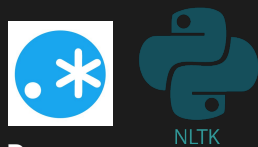
Scrape



Tweepy  
Snscraper

2

Clean



Regex  
Langdetect  
spaCy

3

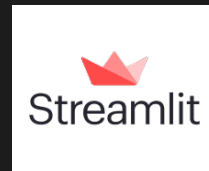
Topic  
Model



NMF  
LSA

4

Build  
App



Remove  
stopword

↑

“Analysis: The world is falling apart.  
But the stock market keeps surging.  
<https://t.co/sd9undf2xH>”

↓

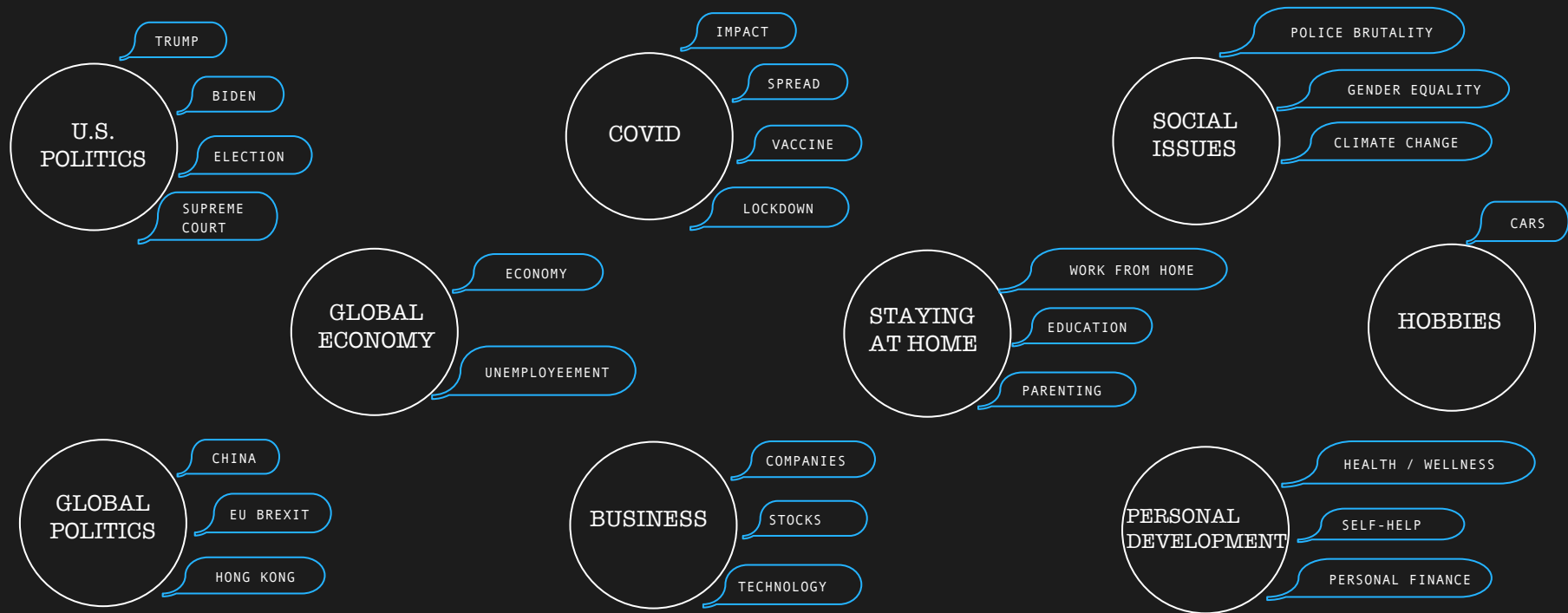
Keep  
nouns

Remove  
links,  
hashtags,  
mentions,  
emojis

-Washington Post  
June 4<sup>th</sup> 2020

200k headlines

8.8k word vectors



9 themes

26 topics

| NEW YORK TIMES    | WALL STREET JOURNAL | WASHINGTON POST  | ATLANTIC        | BUSINESS INSIDER | FINANCIAL TIMES | ECONOMIST       | NEW YORKER      | MEDIUM           | WIRED          |
|-------------------|---------------------|------------------|-----------------|------------------|-----------------|-----------------|-----------------|------------------|----------------|
| Lockdown          | Lockdown            | Trump            | Parenting       | Technology       | Brexit / EU     | China           | Trump           | Technology       | Technology     |
| Police Brutality  | Pandemic Impact     | Police Brutality | Trump           | Stock Market     | Lockdown        | Lockdown        | Parenting       | Gender Equality  | Cars           |
| Health + Wellness | Stock Market        | Supreme Court    | Gender Equality | Cars             | China           | Brexit / EU     | Gender Equality | Parenting        | Climate Change |
| Covid Spread      | Technology          | Lockdown         | Lockdown        | Personal Finance | Hong Kong       | U.S. Election   | Self-Help       | Heath + Wellness | Lockdown       |
| Biden             | China               | Biden            | Technology      | (Un)employment   | Stock Market    | Pandemic Impact | Climate Change  | Self-Help        | Parenting      |

| NEW YORK TIMES    | WALL STREET JOURNAL | WASHINGTON POST  | ATLANTIC        | BUSINESS INSIDER | FINANCIAL TIMES | ECONOMIST       | NEW YORKER      | MEDIUM           | WIRED          |
|-------------------|---------------------|------------------|-----------------|------------------|-----------------|-----------------|-----------------|------------------|----------------|
| Lockdown          | Lockdown            | Trump            | Parenting       | Technology       | Brexit / EU     | China           | Trump           | Technology       | Technology     |
| Police Brutality  | Pandemic Impact     | Police Brutality | Trump           | Stock Market     | Lockdown        | Lockdown        | Parenting       | Gender Equality  | Cars           |
| Health + Wellness | Stock Market        | Supreme Court    | Gender Equality | Cars             | China           | Brexit / EU     | Gender Equality | Parenting        | Climate Change |
| Covid Spread      | Technology          | Lockdown         | Lockdown        | Personal Finance | Hong Kong       | U.S. Election   | Self-Help       | Heath + Wellness | Lockdown       |
| Biden             | China               | Biden            | Technology      | (Un)employment   | Stock Market    | Pandemic Impact | Climate Change  | Self-Help        | Parenting      |

COVID 16%

Pandemic Impact 5%



| NEW YORK TIMES    | WALL STREET JOURNAL | WASHINGTON POST  | ATLANTIC        | BUSINESS INSIDER | FINANCIAL TIMES | ECONOMIST       | NEW YORKER      | MEDIUM           | WIRED          |
|-------------------|---------------------|------------------|-----------------|------------------|-----------------|-----------------|-----------------|------------------|----------------|
| Lockdown          | Lockdown            | Trump            | Parenting       | Technology       | Brexit / EU     | China           | Trump           | Technology       | Technology     |
| Police Brutality  | Pandemic Impact     | Police Brutality | Trump           | Stock Market     | Lockdown        | Lockdown        | Parenting       | Gender Equality  | Cars           |
| Health + Wellness | Stock Market        | Supreme Court    | Gender Equality | Cars             | China           | Brexit / EU     | Gender Equality | Parenting        | Climate Change |
| Covid Spread      | Technology          | Lockdown         | Lockdown        | Personal Finance | Hong Kong       | U.S. Election   | Self-Help       | Heath + Wellness | Lockdown       |
| Biden             | China               | Biden            | Technology      | (Un)employment   | Stock Market    | Pandemic Impact | Climate Change  | Self-Help        | Parenting      |

U.S. Politics 14%  
Trump 4%

| NEW YORK TIMES    | WALL STREET JOURNAL | WASHINGTON POST  | ATLANTIC        | BUSINESS INSIDER | FINANCIAL TIMES | ECONOMIST       | NEW YORKER      | MEDIUM           | WIRED          |
|-------------------|---------------------|------------------|-----------------|------------------|-----------------|-----------------|-----------------|------------------|----------------|
| Lockdown          | Lockdown            | Trump            | Parenting       | Technology       | Brexit / EU     | China           | Trump           | Technology       | Technology     |
| Police Brutality  | Pandemic Impact     | Police Brutality | Trump           | Stock Market     | Lockdown        | Lockdown        | Parenting       | Gender Equality  | Cars           |
| Health + Wellness | Stock Market        | Supreme Court    | Gender Equality | Cars             | China           | Brexit / EU     | Gender Equality | Parenting        | Climate Change |
| Covid Spread      | Technology          | Lockdown         | Lockdown        | Personal Finance | Hong Kong       | U.S. Election   | Self-Help       | Heath + Wellness | Lockdown       |
| Biden             | China               | Biden            | Technology      | (Un)employment   | Stock Market    | Pandemic Impact | Climate Change  | Self-Help        | Parenting      |

Social Issues 13%

Police Brutality 4%

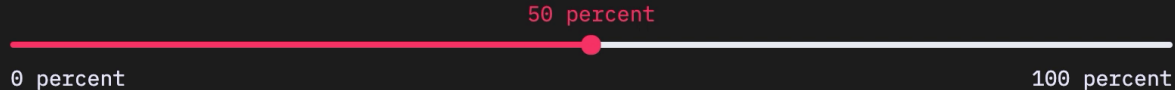
# Digital News Subscription Recommender

*Find your perfect news source, based on your interests...*

On a scale of 0% to 100%, how interested are you in reading about the subjects below?

## COVID-19

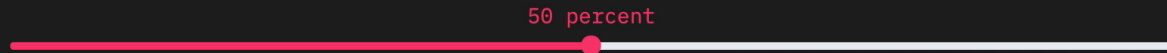
ie. pandemic impact, COVID-19 spread, vaccine research, global lockdown.



You're **50** % interested in reading about COVID-19.

## Staying At Home

ie. working from home, education, parenting.



# Thank You!

CREDITS: This presentation template was created by [Slidesgo](#), including icons by [Flaticon](#), and infographics & images by [Freepik](#).

# Appendix

| NEW YORK TIMES    | WALL STREET JOURNAL | WASHINGTON POST | ATLANTIC        | BUSINESS INSIDER | FINANCIAL TIMES | ECONOMIST       | NEW YORKER      | MEDIUM           | WIRED          |
|-------------------|---------------------|-----------------|-----------------|------------------|-----------------|-----------------|-----------------|------------------|----------------|
| Lockdown          | Lockdown            | Trump           | Parenting       | Technology       | Brexit / EU     | China           | Trump           | Technology       | Technology     |
| Police Protests   | Pandemic Impact     | Police Protests | Trump           | Stock Market     | Lockdown        | Lockdown        | Parenting       | Gender Equality  | Cars           |
| Health + Wellness | Stock Market        | Supreme Court   | Gender Equality | Cars             | China           | Brexit / EU     | Gender Equality | Parenting        | Climate Change |
| Covid Spread      | Technology          | Lockdown        | Lockdown        | Personal Finance | Hong Kong       | U.S. Election   | Self-Help       | Heath + Wellness | Lockdown       |
| Biden             | China               | Biden           | Technology      | (Un)employment   | Stock Market    | Pandemic Impact | Climate Change  | Self-Help        | Parenting      |

COVID 16%  
Pandemic Impact 5%

U.S. Politics 14%  
Trump 4%

Social Issues 13%  
Police Protest 4%

