



# Department of Justice Press Releases

Maddy O'Brien Jones



# The Data

- 13,000+ press releases from 2009 to July 2018
- Mean length: 570 words
- Goal: Figure out what the DOJ is talking about



Data  
exploration

Topic modeling

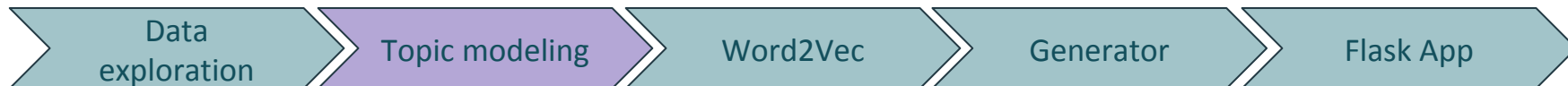
Word2Vec

Generator

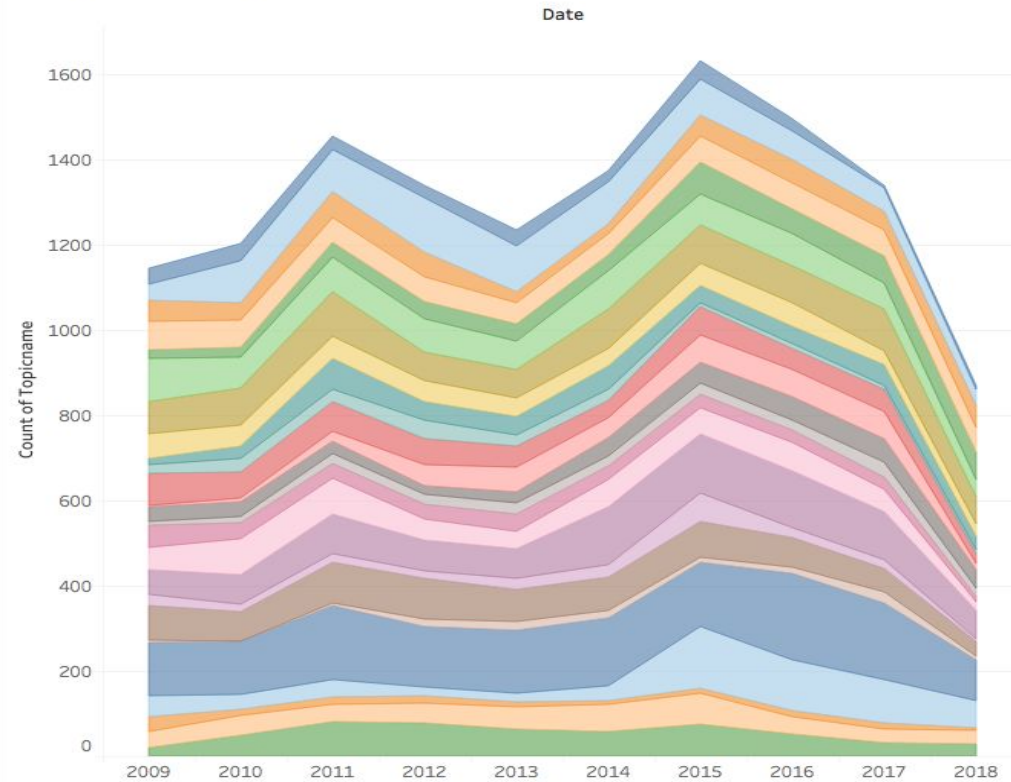
Flask App

# Topics - Non-Negative Matrix Factorization

Topic 1 - Criminal Trials	Plea, pay, special, admit, also, fbi, judg, conspiraci, prosecut, count
Topic 15 - Drugs	Cocaine, criminal, robberi, rival, sheriff, violenc, street, enterpris, drug, firearm
Topic 17 - White collar crime	Money, million, trade, victim, wire, market, credit, crime, borrow, lend
Topic 21 - Real estate	Forc, select, northern, task, mortgag, financi, counti, homeown, fraud, california
Topic 24 - Wildlife and ocean issues	Environ, illeg, marin, speci, engin, book, ocean, oili, environment, record



# Press Release Topics Over Time



- Topicname
- Accessibility
  - Benefits
  - Child Pornography
  - Civil Trials
  - Consumer Issues
  - Corruption
  - Criminal Trials
  - Discrimination
  - Drugs
  - Elections
  - Environment
  - Fraud
  - Human Trafficking
  - Immigration
  - Maritime Issues/Ocean Pollution
  - Market Regulation
  - Native Americans
  - Offshore Accounts
  - Prisoners
  - Real Estate
  - Taxes
  - Terrorism
  - Veterans
  - Whistleblowers
  - White Collar Crime

Data  
exploration

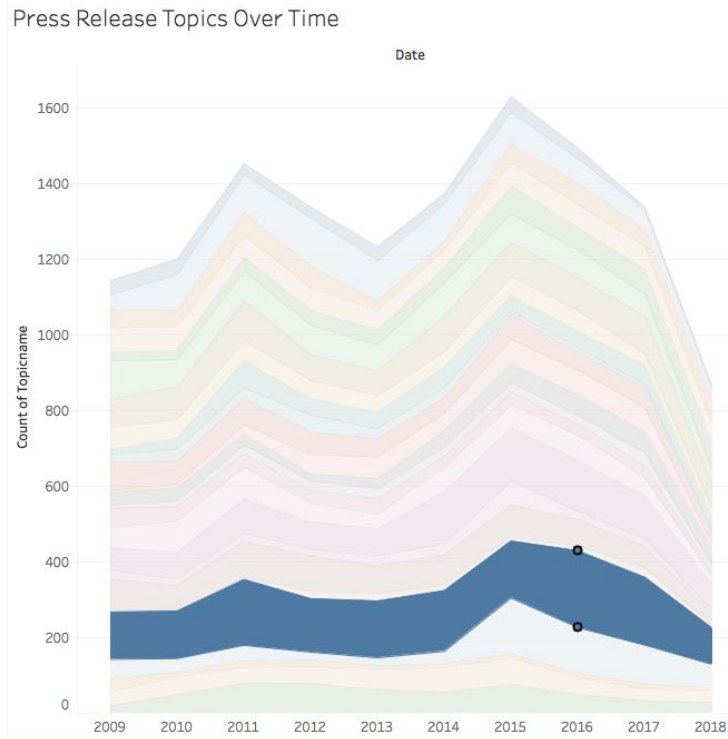
Topic modeling

Word2Vec

Generator

Flask App

# Taxes



Data  
exploration

Topic modeling

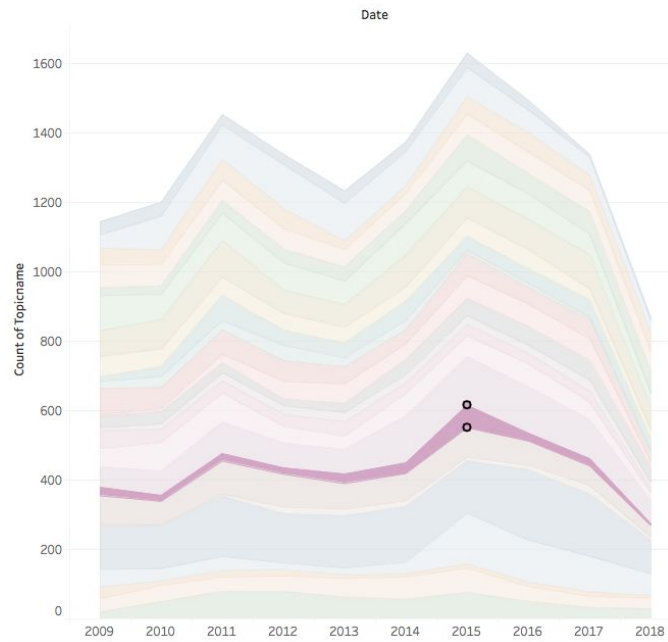
Word2Vec

Generator

Flask App

# Offshore Accounts

Press Release Topics Over Time



Data  
exploration

Topic modeling

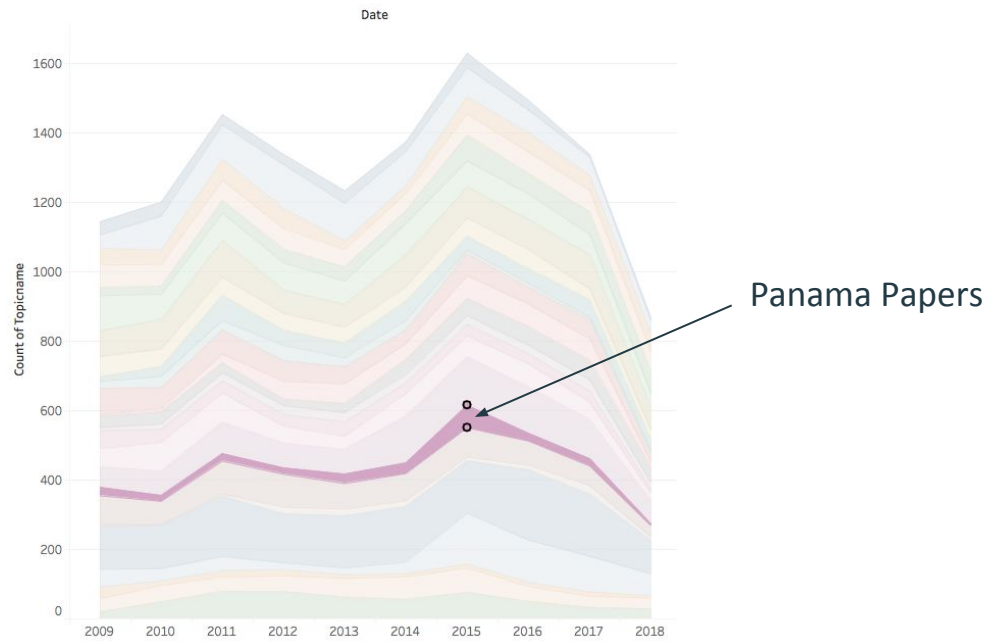
Word2Vec

Generator

Flask App

# Offshore Accounts

Press Release Topics Over Time



Data  
exploration

Topic modeling

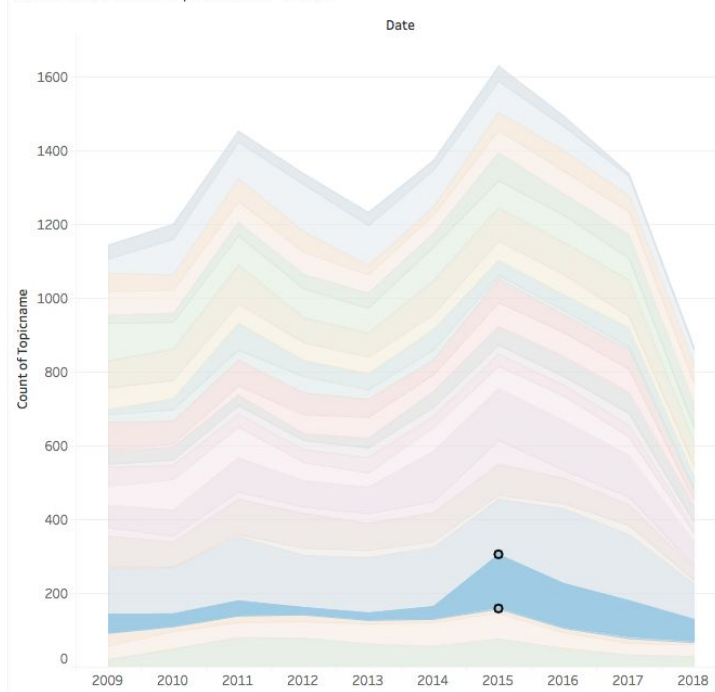
Word2Vec

Generator

Flask App

# Terrorism

Press Release Topics Over Time



Data  
exploration

Topic modeling

Word2Vec

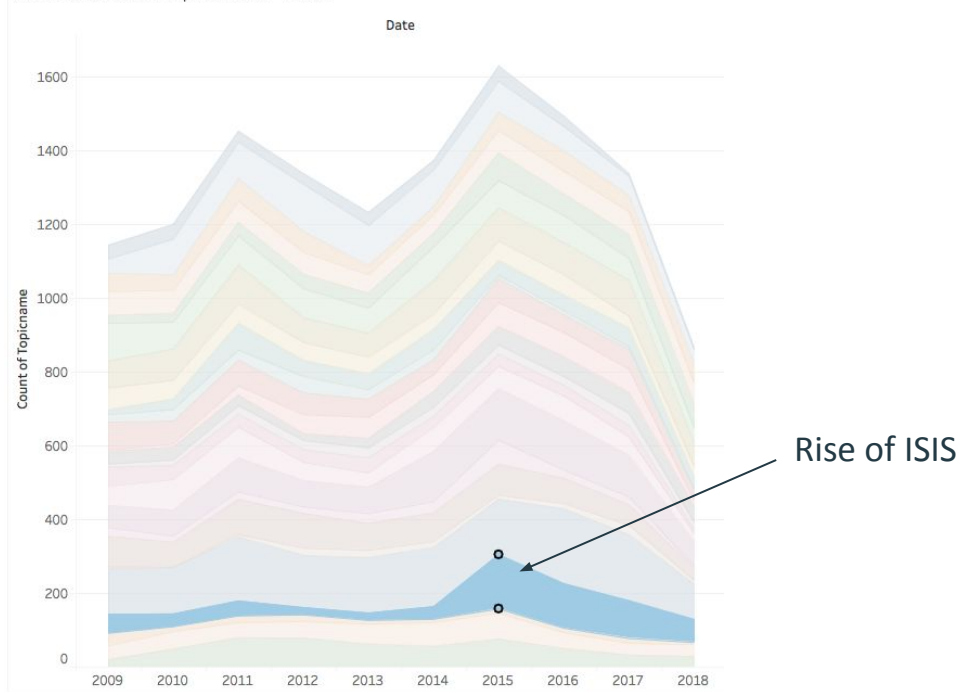
Generator

Flask App



# Terrorism

Press Release Topics Over Time



Data  
exploration

Topic modeling

Word2Vec

Generator

Flask App

# Word2Vec - Religion

## Religion

- Race
- Ethnicity
- Color
- Skin
- Beliefs
- Gender
- Regardless
- Origin
- Status
- Perceived

## Islam

- Die
- Muslims
- Allah
- Enemies
- Hatred
- Bad
- Hateful
- Crusaders
- Martyr
- Teen

Data  
exploration

Topic modeling

Word2Vec

Generator

Flask App

# Word2Vec - Places of Worship

## Church

- Congregation
- Mosque
- Christ
- Macedonia
- Neighborhood
- Motel
- Synagogue
- God
- Woman
- Restaurant

## Mosque

- Church
- Motel
- Entrance
- Restaurant
- Bomb
- Congregation
- Handgun
- Shotgun
- Synagogue
- Room

## Synagogue

- Church
- Mosque
- Congregation
- Restaurant
- Crowded
- Garage
- Motel
- Riding
- Nightclub
- Pistol

Data  
exploration

Topic modeling

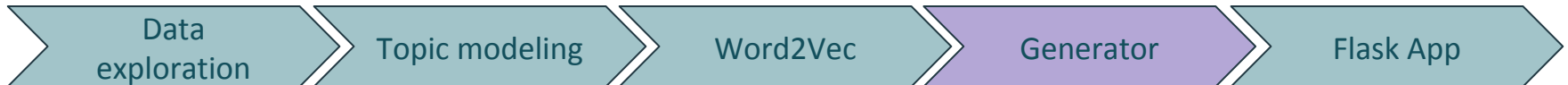
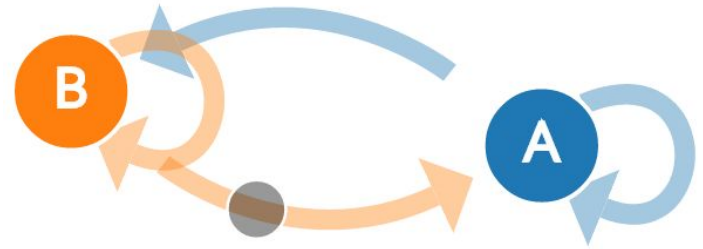
Word2Vec

Generator

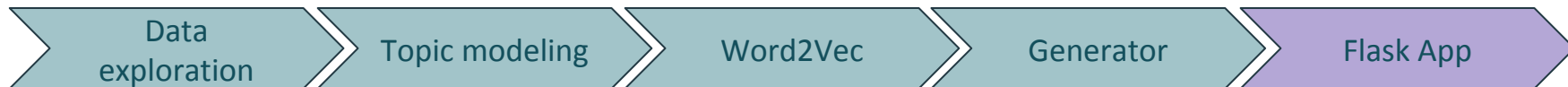
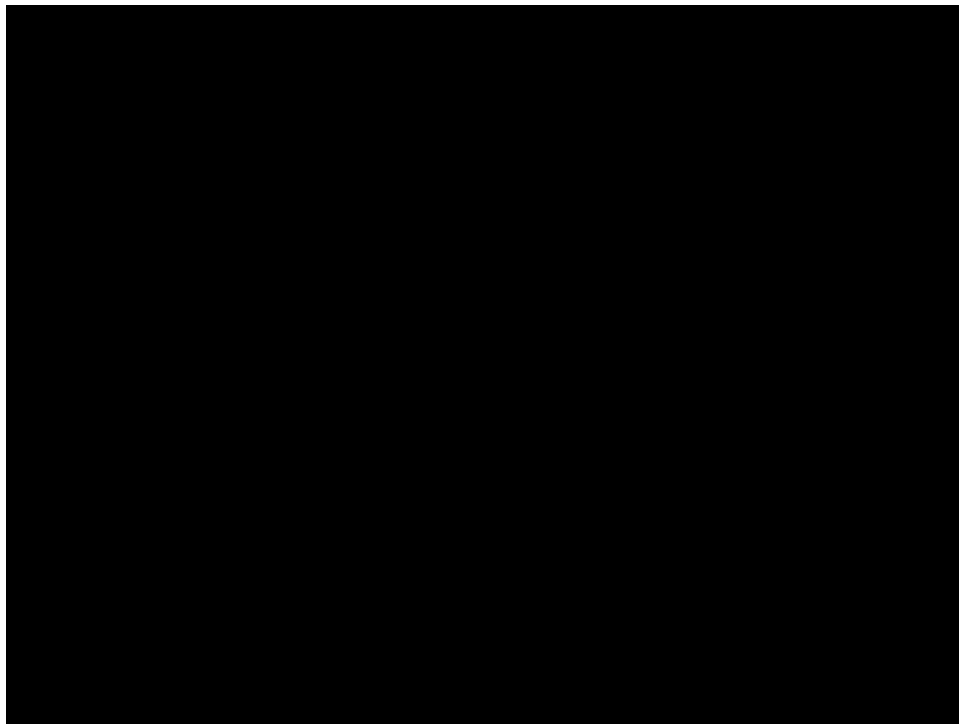
Flask App

# Generator

- Generator created using Markov process
- Press releases can be generated for some of the 25 topics



# Flask





Thank you!

Those darn dolphins

