

Final Presentation

Elizabeth Mitchell

Introduction

Getting Data

Analysis

Results

Conclusions

# Final Presentation

## What Congress Does

Elizabeth Mitchell

December 1, 2016

# Overview

Final Presentation

Elizabeth Mitchell

Introduction

Getting Data

Analysis

Results

Conclusions

1 Introduction

2 Getting Data

3 Analysis

4 Results

5 Conclusions

# Source of Interest

Final Presentation

Elizabeth Mitchell

Introduction

Getting Data

Analysis

Results

Conclusions

## H.R. 6009: Federal Agency Mail Management Act of 2016

The **House** Majority Leader indicated the bill would be considered in the week ahead on Nov 25, 2016.

## H.R. 6138: To designate the facility of the United States Postal Service located at 560 East Pleasant Valley Road, Port Hueneme, California, as the U.S. Naval Construction Battalion "Seabees" Fallen ...

The **House** Majority Leader indicated the bill would be considered in the week ahead on Nov 25, 2016.

## H.R. 6186: Follow the Rules Act

The **House** Majority Leader indicated the bill would be considered in the week ahead on Nov 25, 2016.

## H.R. 6282: To designate the facility of the United States Postal Service located at 2024 Jerome Avenue, in Bronx, New York, as the "Dr. Roscoe C. Brown, Jr. Post Office ...

The **House** Majority Leader indicated the bill would be considered in the week ahead on Nov 25, 2016.

# Tools

Final Presentation

Elizabeth Mitchell

Introduction

Getting Data

Analysis

Results

Conclusions

- Python
  - access API to download bill data
- R
  - basic data cleaning
  - tm package for text processing
  - stm package (structural topic modeling) for sorting, analysis, and plotting
- $\text{\LaTeX}$ 
  - Beamer document class for this presentation

# Source of Data

Final Presentation

Elizabeth Mitchell

Introduction

Getting Data

Analysis

Results

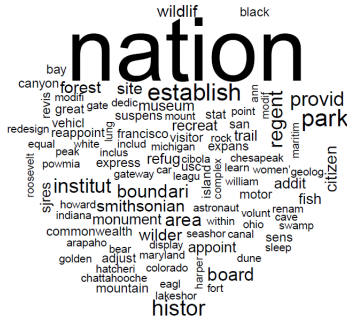
Conclusions

- GovTrack API
  - basic legislator and legislation data back to 1973
  - downloaded enacted bill titles and dates passed
- Processing for text analysis
  - lowercase, stemming, remove numbers and punctuation
  - limit to words that appear twice
  - remove stop words

```
customstop <- c("H.R.", "S.", "H.J.Res.", "S.J.Res.", "bill", "resolution", "act", "amend",  
  "extend", "joint", "authorize", "extending", "amendment", "amendments", "expand",  
  "title", "chapter", "amended", "provision", "provisions", "reauthorize", "code",  
  "section", "purpose", "purposes", "request", "authorizing", "requesting", "ensure",  
  "extending", "reauthorization", "sjres", "hjres")
```

# Analysis

- Unsupervised algorithm to sort titles into topics
  - STM package in R
  - 21 topics, 21 standing committees in House and Senate
- Explore topics for common themes



# Results

Final Presentation

Elizabeth Mitchell

Introduction

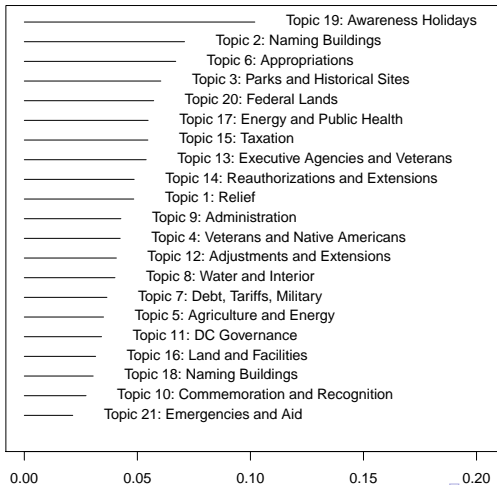
Getting Data

Analysis

Results

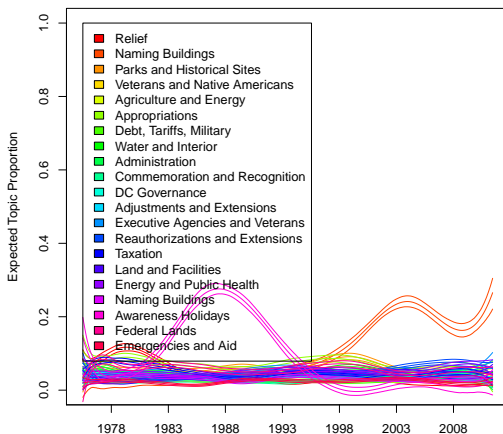
Conclusions

## Topic Frequency



# Results

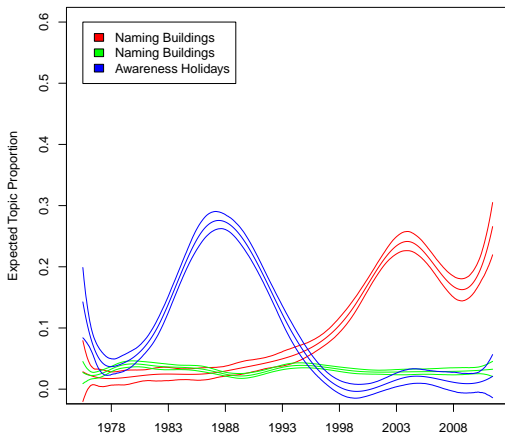
## ■ Compare topic frequency over time





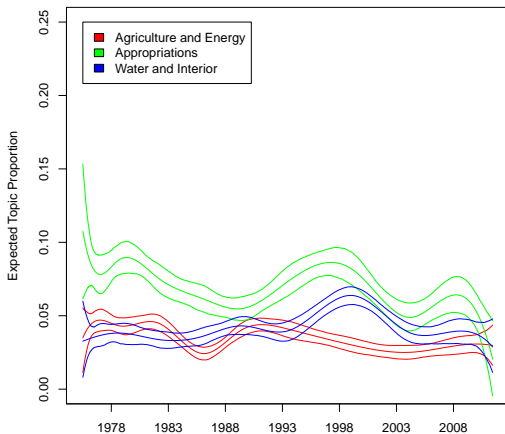
# Results

## ■ Compare topic frequency over time



# Results

## ■ Compare topic frequency over time



# Challenges and Next Steps

Final Presentation

Elizabeth Mitchell

Introduction

Getting Data

Analysis

Results

Conclusions

- stm package
  - great documentation, small community
- arbitrary decisions
  - stop words, type of processing (proper nouns? sparsity?), number of categories
- Substituting bill text for bill title
  - source, data size, computing power (topic modeling took 1.5 hours for titles alone)