README for Content Analyses

Scripts for content analysis of cable transcripts

The three R scripts in this folder run the content analysis of cable transcripts. Note that these scripts depend on raw transcripts which are licensed from Nexis and are not included in the archive. Document-term matrices and the fitted structural topic model and Gentzkow-Shapiro-Taddy slant model are included in the archive and can be used to generate the figures produced by `03_identify_tea_party_candidates_on_tv.R` without the raw transcript data.

Scripts in this folder

- `01_parse_transcripts.R`: Parse raw Lexis-Nexis transcript files into structured data.
- `02_count_literal_mentions.R`: Count literal mentions of the phrases "Tea Party" and "Occupy" in transcripts.
- `03_identify_tea_party_candidates_on_tv.R`: Apply Gentzkow-Shapiro-Taddy and Structural Topic Model methods to language from cable channel transcripts, including Congressional candidates' TV appearances.

Packages Required

The following R packages (available from CRAN) are needed to run the scripts in this folder:

- `tidyverse`
- `lubridate`
- `data.table`
- `stringr`
- `ggthemes`
- `quanteda`
- `glue`
- `stringr`
- `tokenizers`
- `stopwords`
- `distrom`
- `stm`
- `gt`

Figures and Tables Generated

From `02_count_literal_mentions.R`:

Appendix Figures: - **Figure D.1.1**: Count of mentions of the phrase "Tea Party" on cable channels. - **Figure I.1**: Count of mentions of the phrase "Occupy" on cable channels.

From `03_identify_tea_party_candidates_on_tv.R`:

Main Figures: - **Figure 1a**: Fraction of words spoken by Tea Party-affiliated candidates on each cable channel over time. - **Figure 1b**: Fraction of words spoken by mainstream Republican candidates on each cable channel over time. - **Figure 2a**: Topic weights for candidates (Tea Party topics). - **Figure 2b**: Topic weights

for channels (Tea Party topics). - **Figure 3a**: Tea Party language score using Gentzkow-Shapiro-Taddy method. - **Figure 3b**: Republican language score using Gentzkow-Shapiro-Taddy method.

Appendix Figures: - **Figure B.1.1a**: Fraction of words spoken by Tea Party-affiliated candidates on each cable channel over time, with confidence intervals. - **Figure B.1.1b**: Fraction of words spoken by mainstream Republican candidates on each cable channel over time, with confidence intervals. - **Figure B.2.1a**: Candidate topic weights for mainstream Republican topics. - **Figure B.2.1b**: Channel topic weights for mainstream Republican topics. - **Figure B.3.1a**: Tea Party language score confidence intervals. - **Figure B.3.1b**: Republican language score confidence intervals.

Tables: - **Table 1**: Structural Topic Model topics emphasized by Tea Party candidates. - **Table 2**: Structural Topic Model topics emphasized by mainstream Republican candidates.