FB Analysis

Kim, Brew, Zilinsky

- Our FB dataset consists of 293960 rows (House), and 219888 rows (Senate), but many of these ads are duplicates.
- After removing duplicate texts, we have 43694 observations for the US House, and 16828 observations for the US Senate.

Pre-processing

- We download the Facebook ads for both chambers of US Congress.
- We remove duplicates (in the ad-text sense)
- We merge in covariates, namely:
  - Party ID
  - Office
  - State
  - Vote share (a proxy for competitiveness)
- We create a corpus with the above mentioned document variables appended as covariates.
- We tokenize the text, removing English and Spanish stopwords, lower-casing the words, a removing most punctuation.
- We currently do not apply a stemmer.
- We create a document feature matrix, stacking the House and Senate data into one large martix.
- The dimensions of this matrix are: 60522 x 46417.
- That is, we there are 46417 features (types). The total number of tokens is 1853508.

Some of the analyses are at the ad level, but we also produce candidate level analyses. When each "document" is a candidate, and we prepare a document-feature matrix whose dimensions are: 716 x 46417.

The median number of words produced per candidate is 1120, the average number is 2549.2, but the maximum is 77537, so we will generally report "% of words" for a given candidate that meet some condition (e.g. belong to a particular dictionary).

Mentions of salient topics/figures

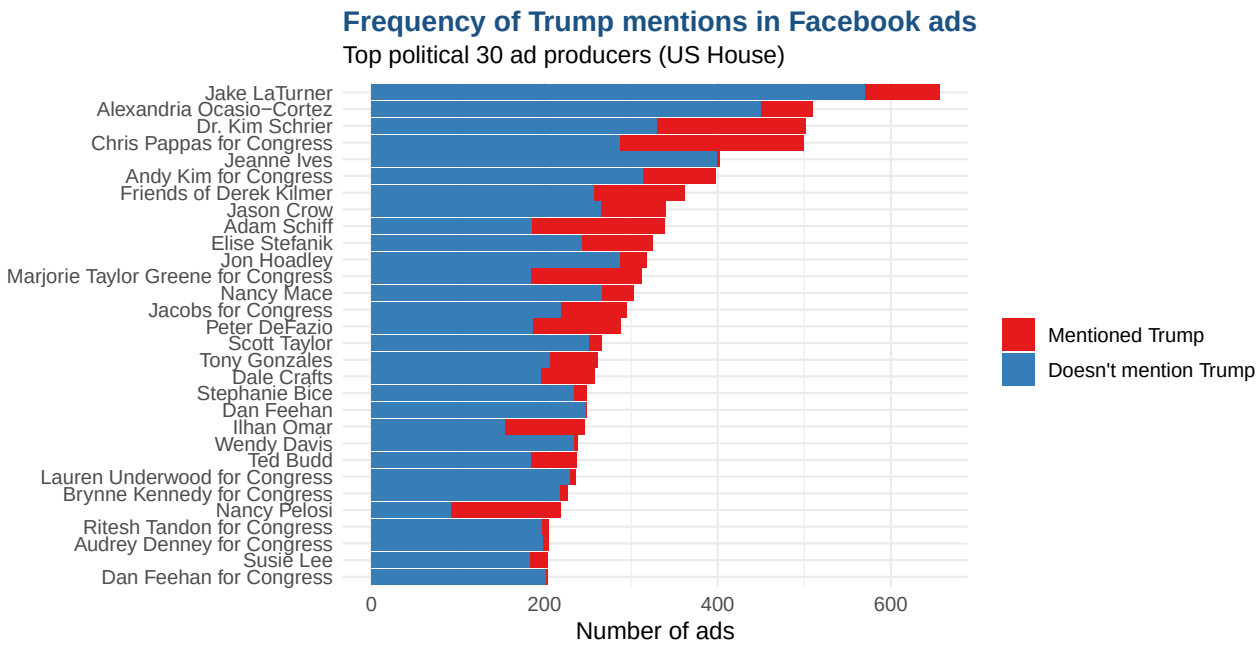


Figure 1: Mentions of Trump

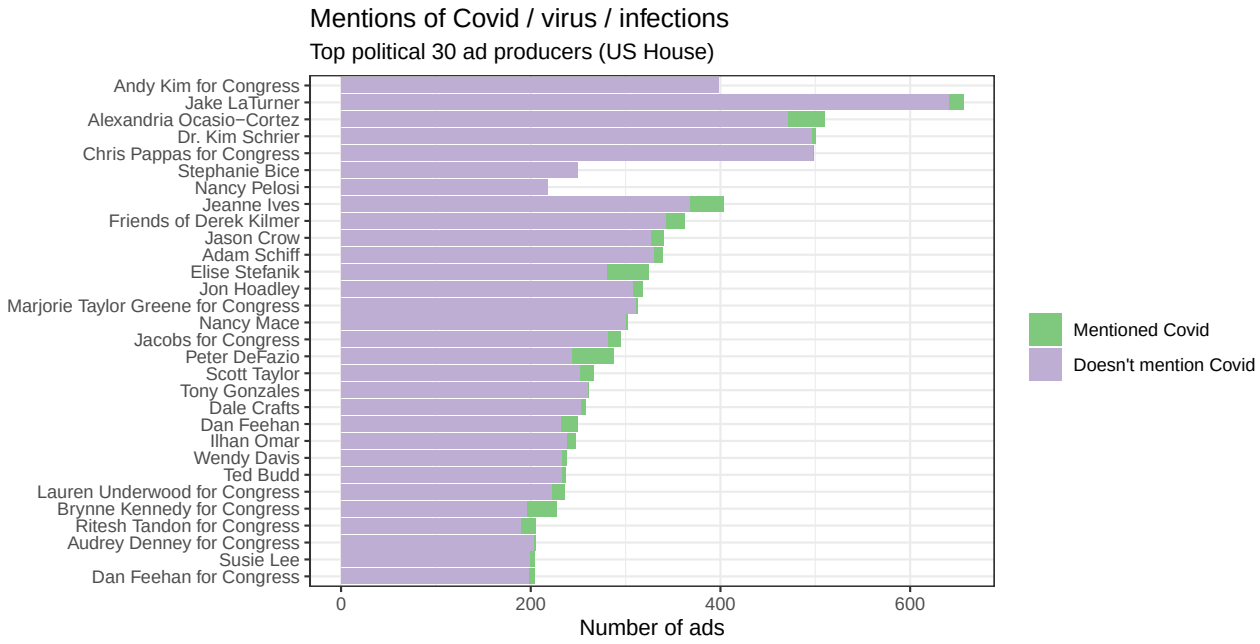


Figure 2: Mentions of Covid

Usage of salient words, broken down by PID

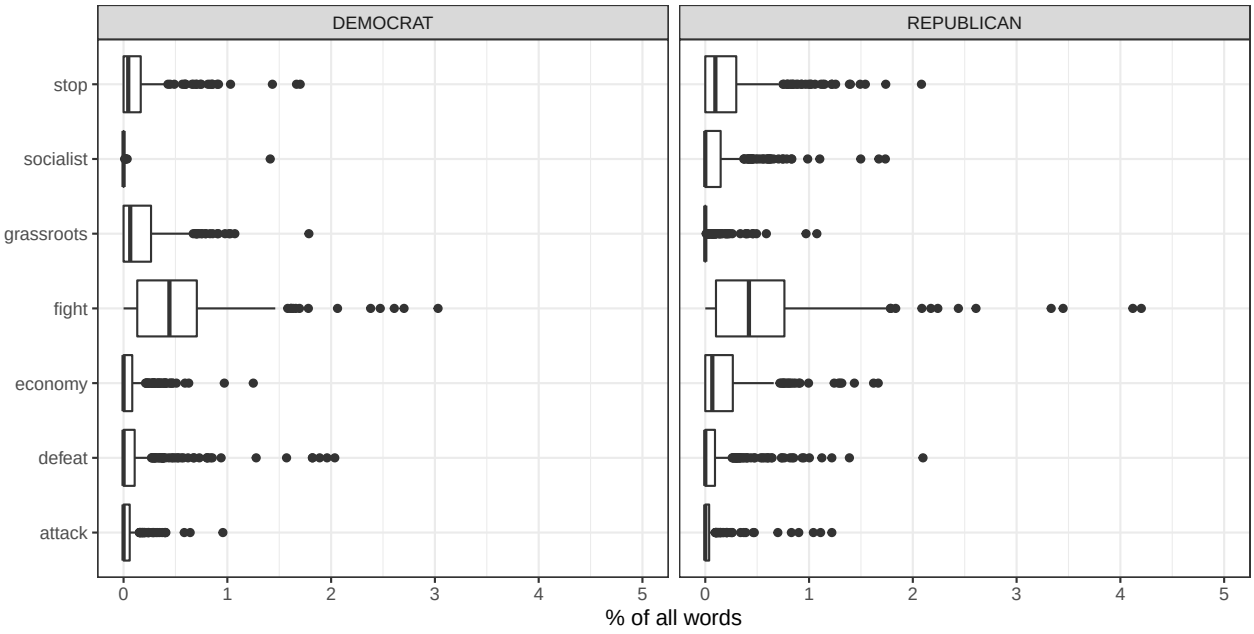


Figure 3: Candidate-level usage of selected words (% of all words used in ads)

Who produced most words for FB ads?

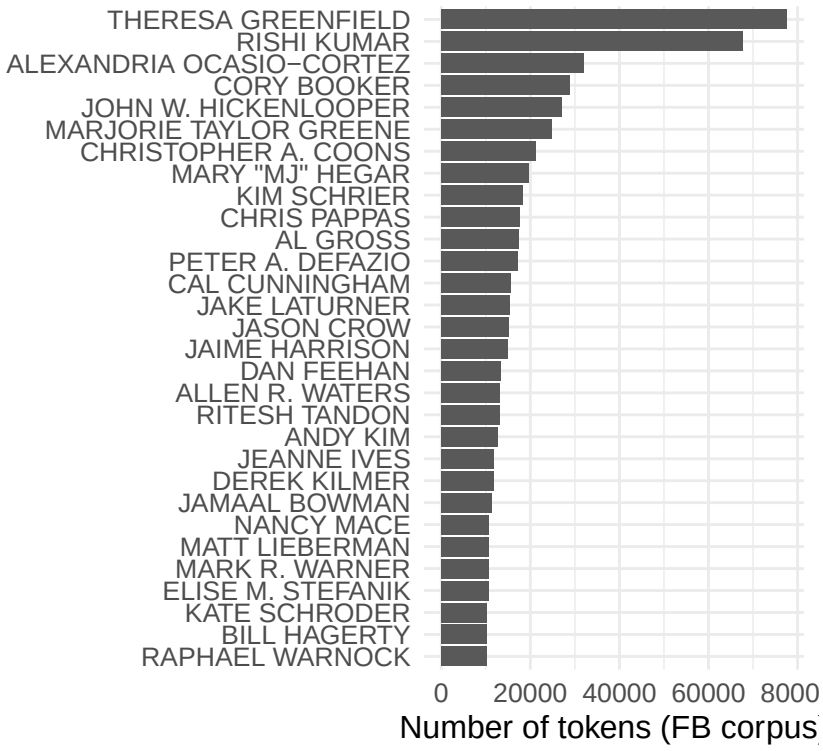


Figure 4: Number of tokens by the 30 most prolific candidates

Top features (20 most frequently occurring tokens)

All FB ads

```
##      help       can         $         >       now        us      need  congress
##     18245     15426     14861     14423     14001     13606     13065     11510
##  campaign     today     trump     fight    senate      back   support    people
##     10789     10544     10079      9592      9238      9017      8526      8467
##      chip      vote president      make
##      7962      7791      7579      7340
```

Democrats

```
##     help      can        >        $      now     need       us campaign
##    12294    11411    10752    10638    10152     9751     9696     8148
##   senate    today congress    fight     chip   people     back    trump
##     6935     6874     6791     6303     6159     6052     5638     5466
##     make     take     join     just
##     5405     4959     4731     4549
```

Republicans

```
##         help    president        trump     congress            $      support
```

```
##          5434          4837          4297          4240          3928          3729
##             >           now            us           can         today          vote
##          3598          3563          3560          3474          3270          3176
##          back         fight          need conservative         stand          join
##          3149          2997          2966          2476          2420          2401
##        senate      campaign
##          2300          2294
```

Dictionary anlysis

Trolling words by ad type

When a donation link is included, the language is on average more aggressive:

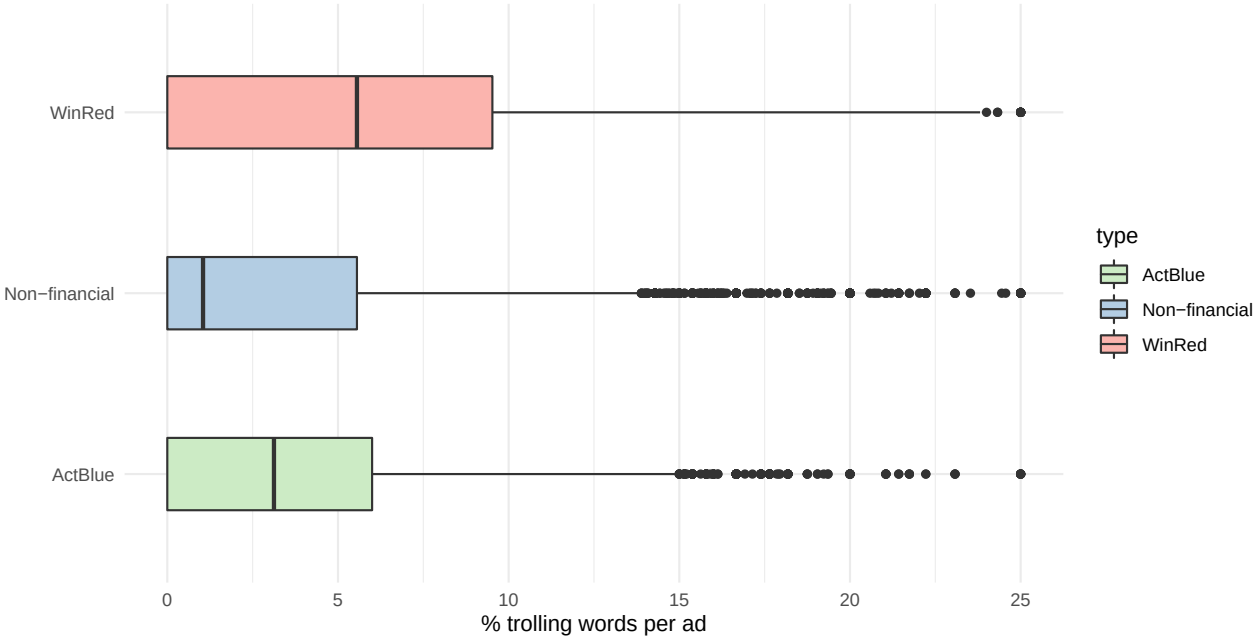


Figure 5: Percent of trolling words per ad, broken down by ad type

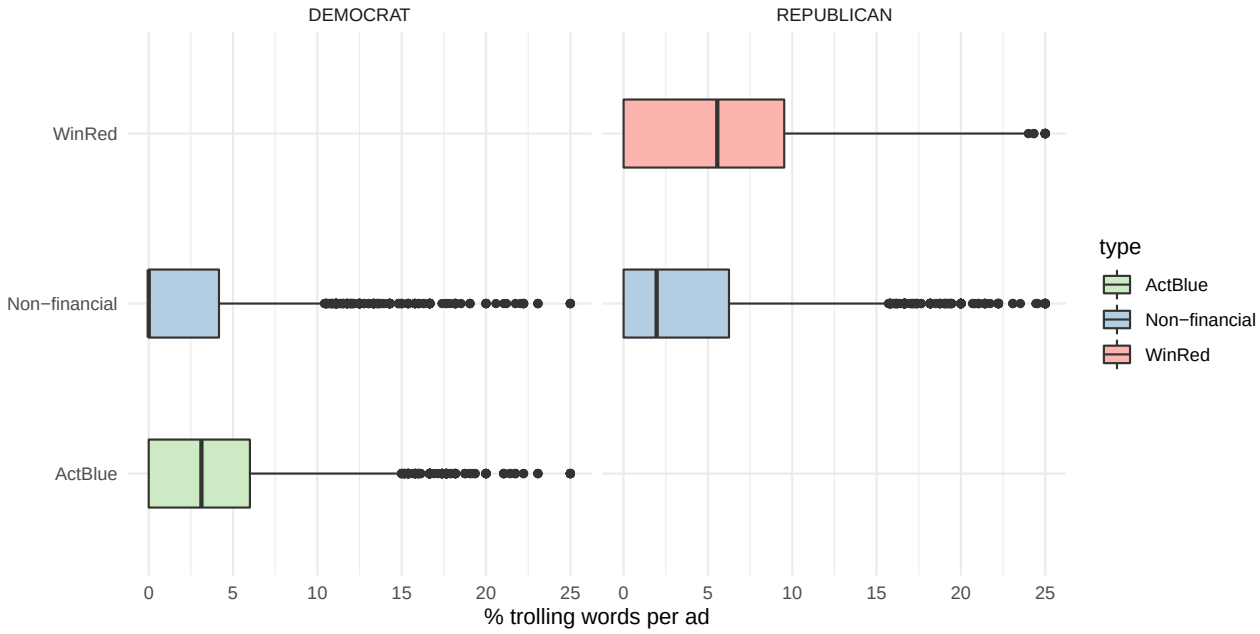


Figure 6: Percent of trolling words per ad, broken down by ad type and by Party ID

Trolling words [candidate-level analysis - PROPRTIONS]

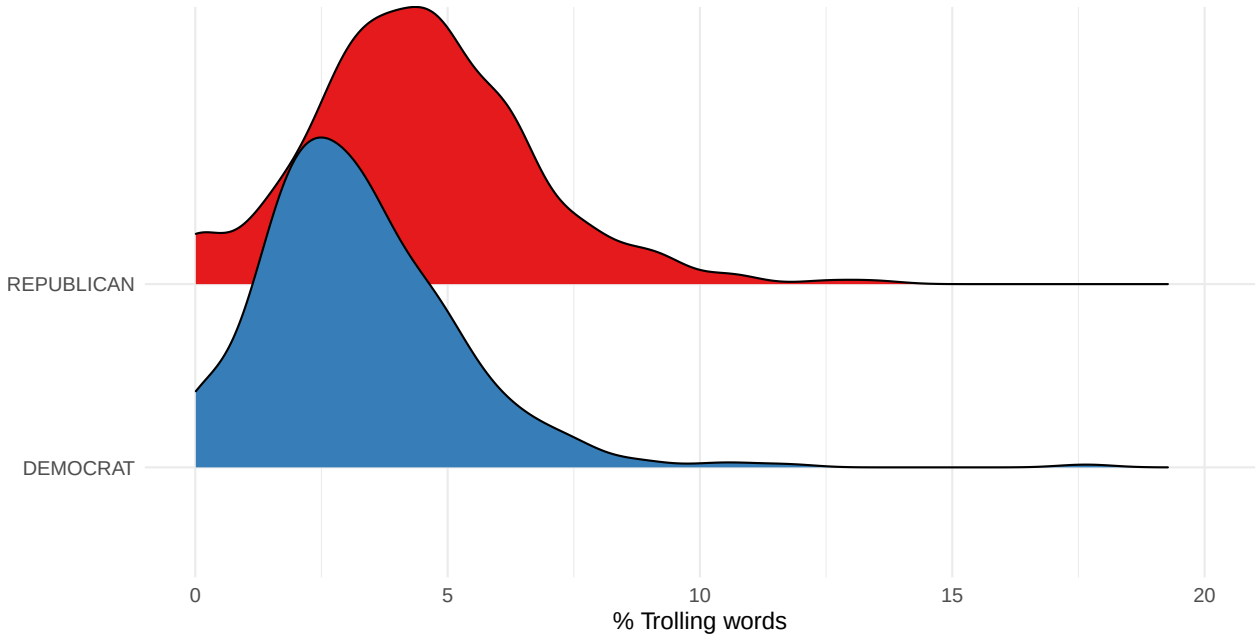


Figure 7: Distribution of candidate-level average of trolling usage (broken down by Party ID)

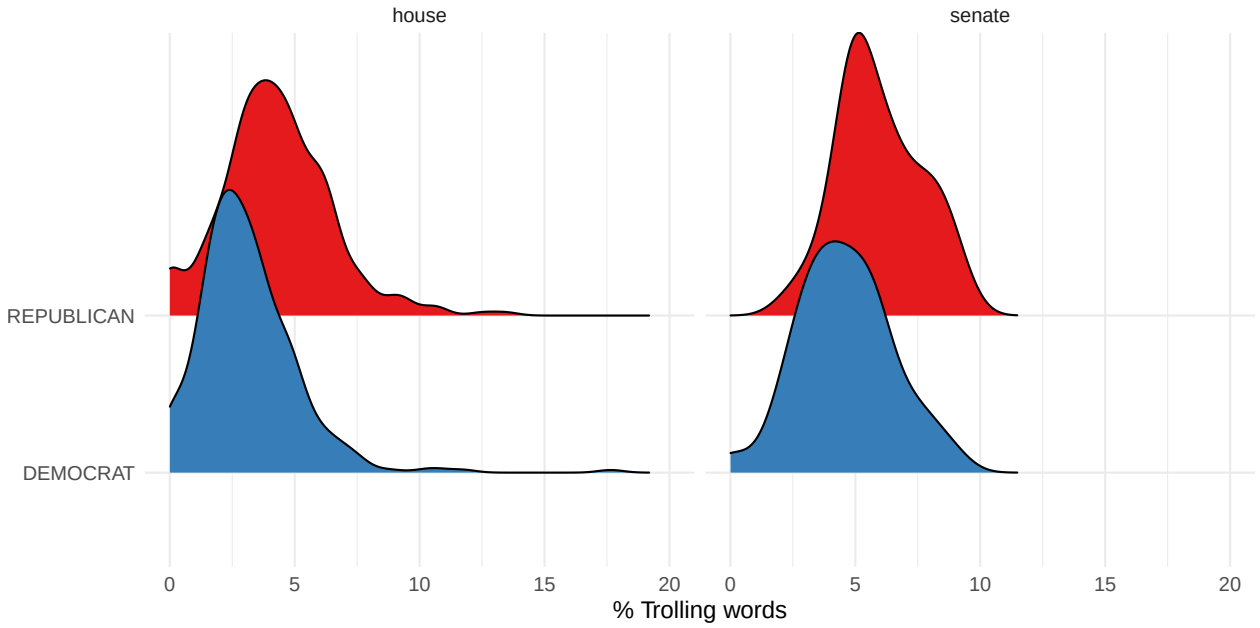


Figure 8: Distribution of candidate-level average of trolling usage (broken down by Party ID and chamber of Congress)

Figure 9: Candidate-level average usage of trolling words, broken down by chamber

Trolling words [ad-level analysis]

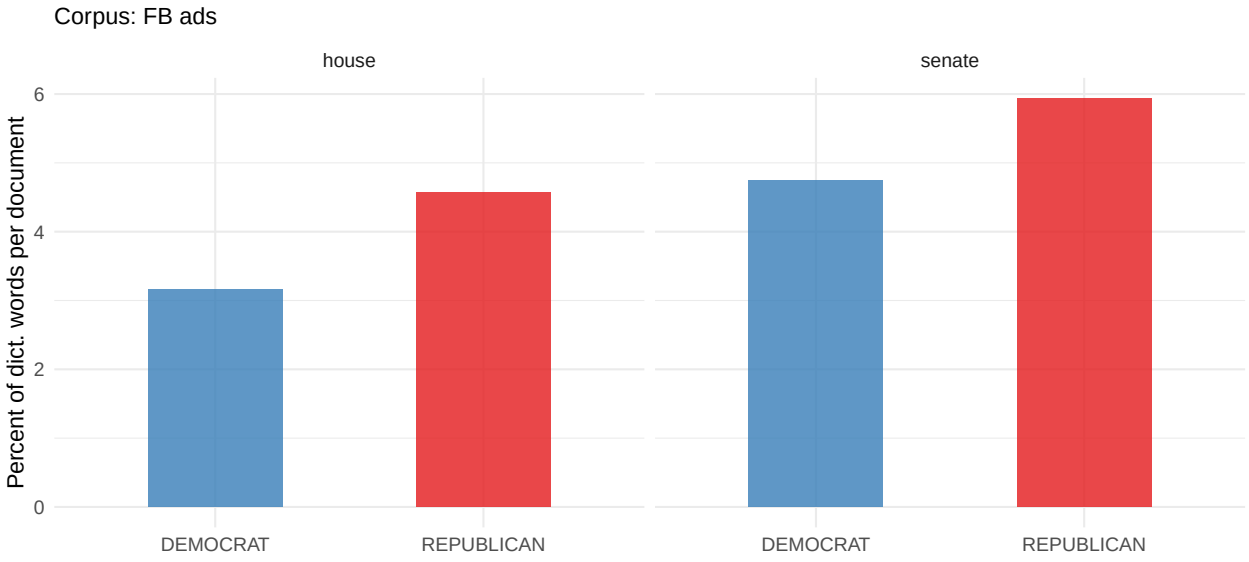


Figure 10: Average proportion of trolling words per ad, broken down by party ID and chamber

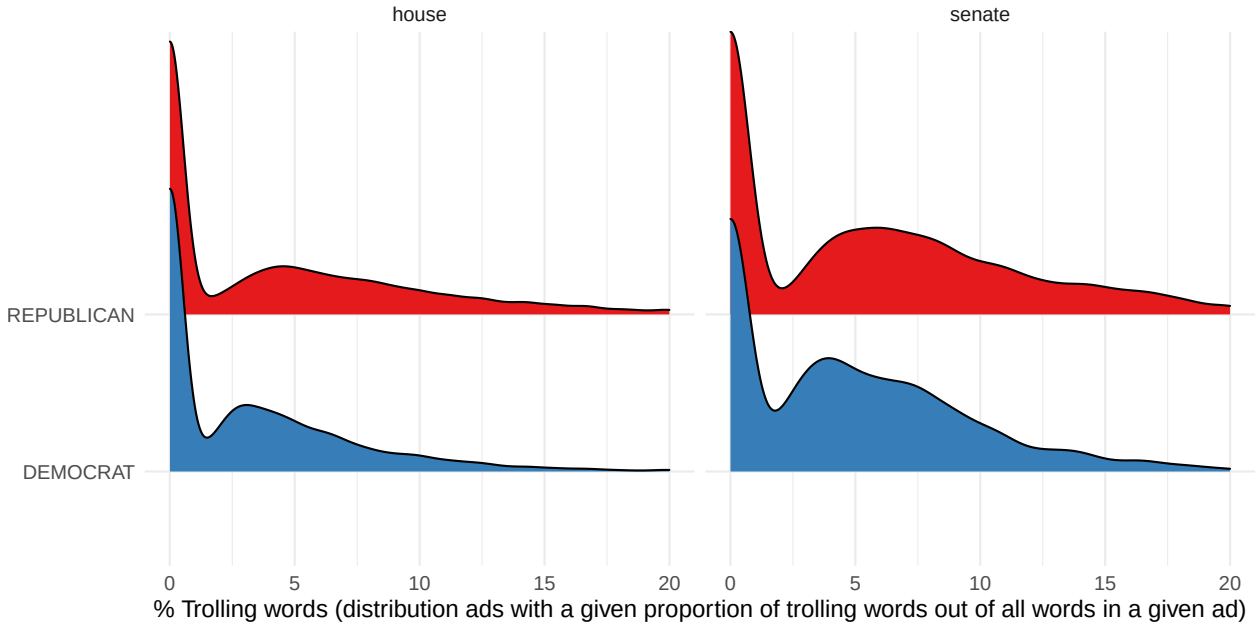


Figure 11: Distribution of trolling words per ad, broken down by party ID and chamber

Trolling words [ad-level analysis - COUNTS]

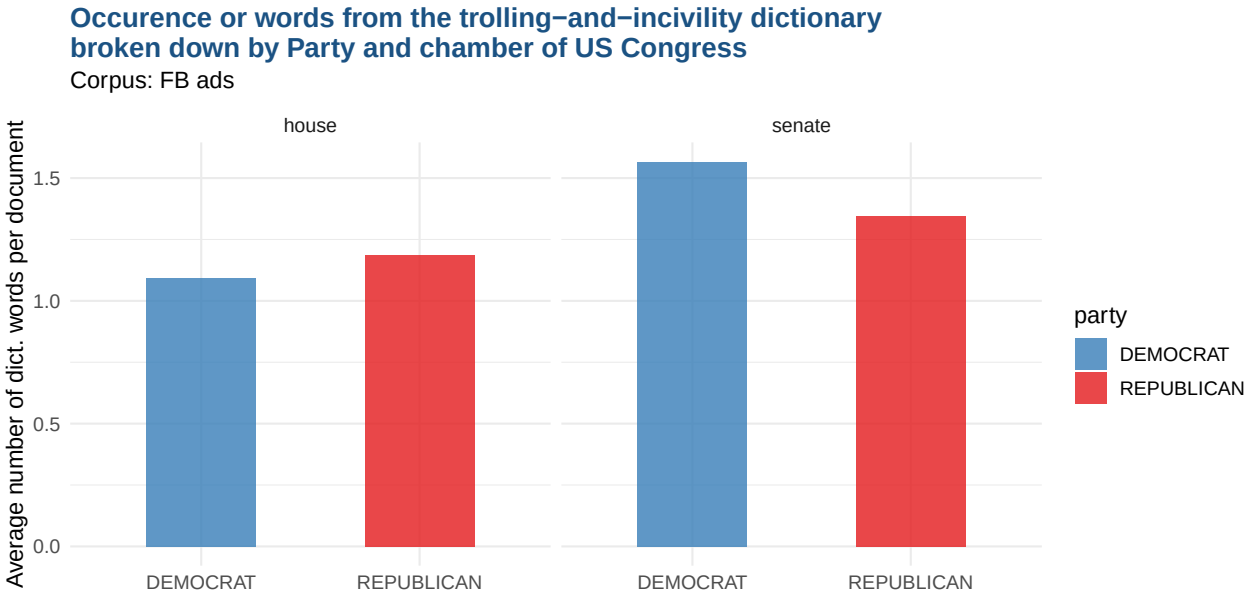


Figure 12: Average number of trolling words by party and chamber

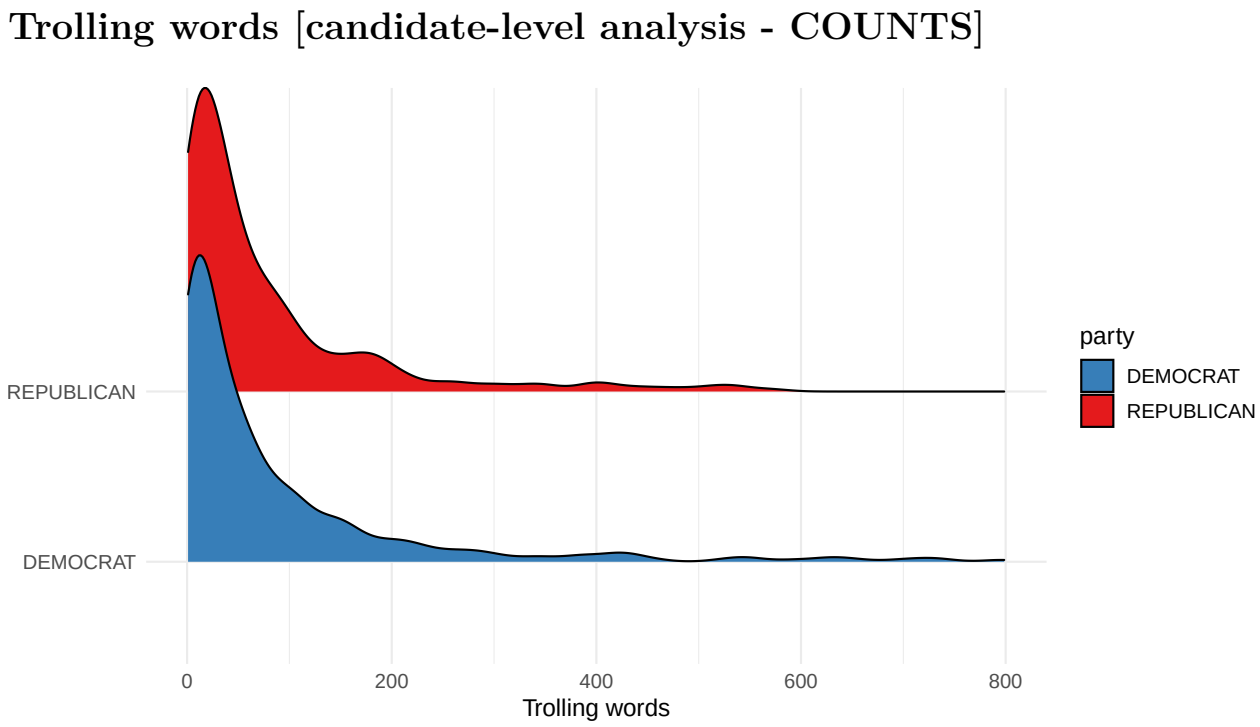


Figure 13: Distribtion of candidate-level trolling words counts (totals)

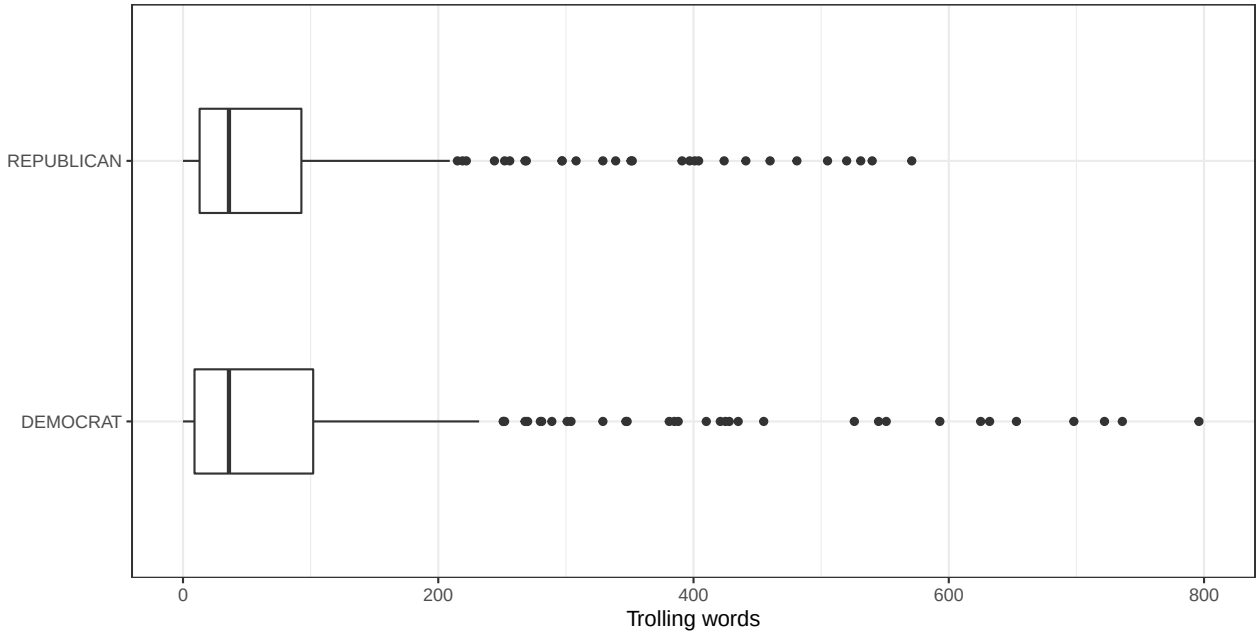


Figure 14: Distribtion of candidate-level trolling words counts (totals)

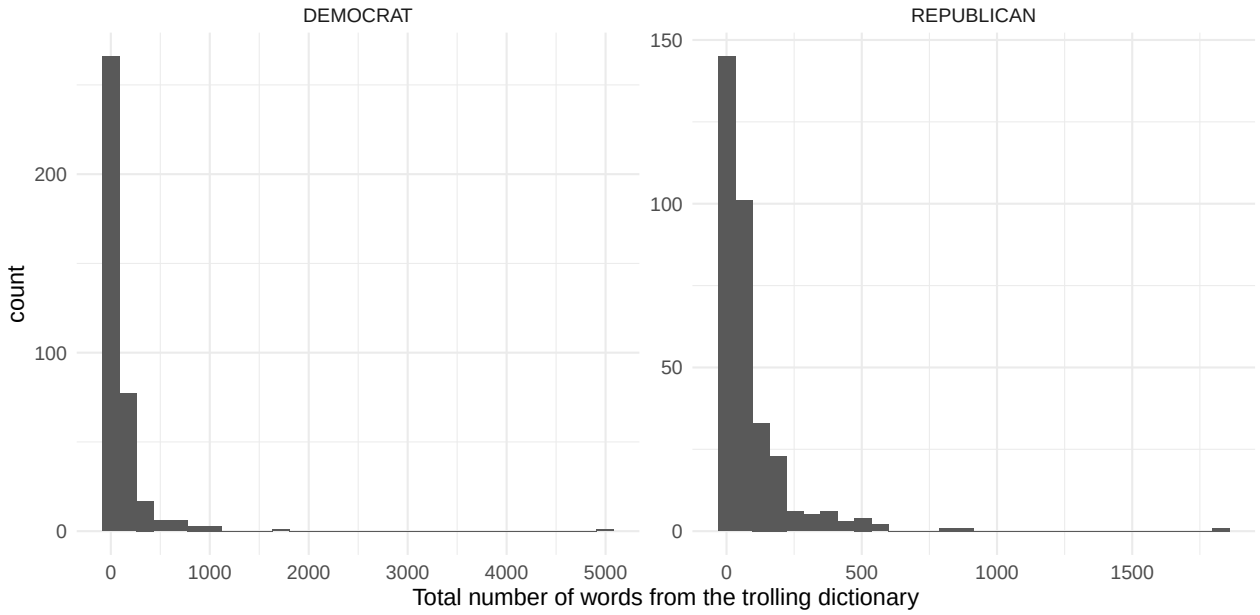


Figure 15: Total number of trolling words in the corpus produced by candidates

Moral foundations [candidate-level analysis]

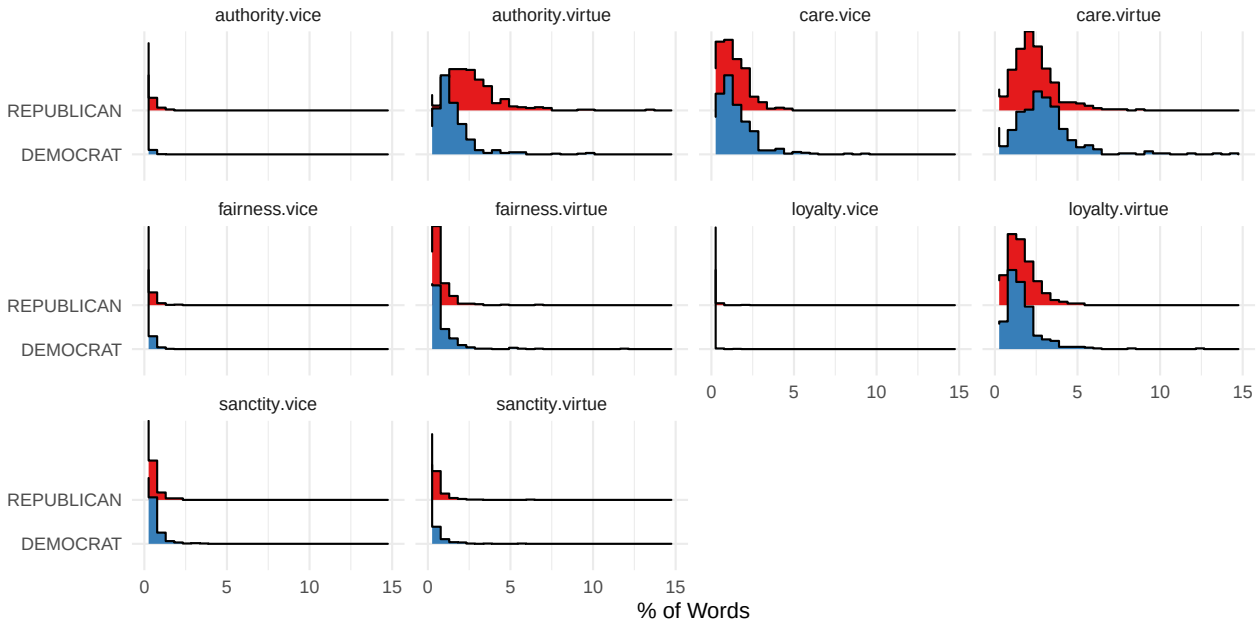
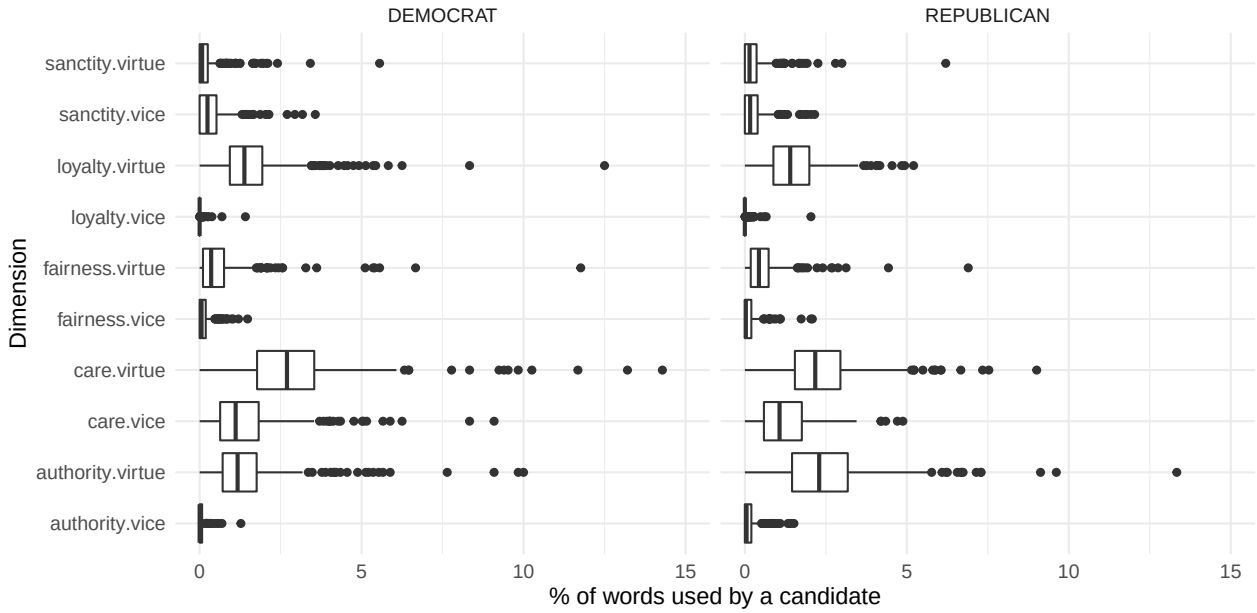


Figure 16: Distribution of Moral foundations words in FB ads, by party



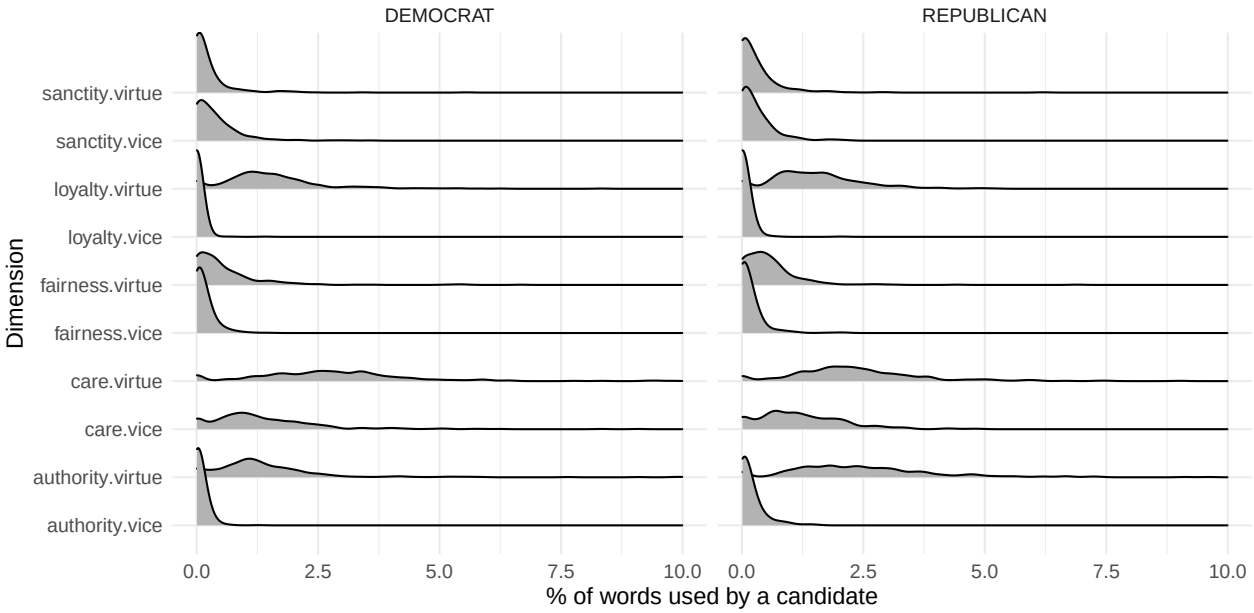


Figure 17: Candidate-level usage of moral words (MFD)

Table 1: Average Usage of words across candidates, broken down by Party ID

Average Usage of the dimension Per Candidate (in %)	DEMOCRAT	REPUBLICAN
authority.vice	0.06	0.16
authority.virtue	1.41	2.53
care.vice	1.36	1.21
care.virtue	2.94	2.33
fairness.vice	0.13	0.14
fairness.virtue	0.65	0.56
loyalty.vice	0.01	0.03
loyalty.virtue	1.56	1.52
sanctity.vice	0.38	0.28
sanctity.virtue	0.24	0.30