Replication Data Codebook for "Responsive Rhetoric: Evidence from Congressional Redistricting"

Last Modified: 7-26-24

Dataset #1: District and Member Changes (116th and 117th)

File Format: R data (.Rda)

Unit of Analysis: A member of the U.S. House of Representatives (from the 116th session)

Number of Observations: 435

Number of Variables: 36

Variable Names and Descriptions:

1. *ICPSR* – Unique ICPSR identifier for the member
2. *State* – State of origin for the member
3. *District_116* – District number represented by the member during the 116th session
4. *District_117* – District number the member ran in (or would've ran in) during the 117th session. Mappings from old to new district come from the Cook Political Report.
5. *RanForReelection* – Whether the member ran for reelection in 2022.
6. *Party* – Party of the member (D for Democratic, R for Republican)
7. *CookPVI_116* – Cook PVI rating for the member's district during the 116th session
8. *CookPVI_117* – Cook PVI rating for the member's district during the 117th session
9. *CookPVI_Change* – Difference in the Cook PVI rating between the 117th and 116th sessions
10. *CookFavorability_116* – Cook PVI rating for the member's district during the 116th session, recoded so that higher values mean more favorable to the member's party
11. *CookFavorability_117* – Cook PVI rating for the member's district during the 117th session, recoded so that higher values are more favorable to the member's party
12. *CookFavorability_Change* – The difference in *CookPVI_116* and *CookPVI_117*, recoded so that higher values indicate a more favorable district for the member's party
13. *PreviouslyCompetitive* – Binary indicator coded as 1 if the member's 116th district was competitive (Cook favorability rating of 5 or less), but the member's 117th district was not
14. *NewlyCompetitive* – Binary indicator coded as 1 if the member's 116th district was not competitive (Cook favorability rating of 5 or less), but the member's 117th district was
15. *DistIdeol_116* – Ideology of the member's district during the 116th session, estimated via MRP and produced by Tausanovitch and Warshaw
16. *DistIdeol _117* – Ideology of the member's district during the 117th session, estimated via MRP and produced by Tausanovitch and Warshaw
17. *DistIdeol_Change* -- The difference between *DistIdeol_116* and *DistIdeol _117*, coded so that higher values are more ideologically favorable to the member's party
18. *TweetSlantR_116* – A member's Tweet Partisanship Score during the 116th session, calculated as described in the article
19. *TweetSlantR_117* – A member's Tweet Partisanship Score during the 117th session, calculated as described in the article using tweets posted after the member's new district was publicly known
20. *TweetSlantR_Change* – The difference between *TweetSlant_116* and *TweetSlant_117*

21. *TweetExtremism_116* – *TweetSlant_116*, recoded so that higher values indicate more extreme tweeting (more Republican for Republicans, more Democratic for Democrats)
22. *TweetExtremism_117* – *TweetSlant_117*, recoded so that higher values indicate more extreme tweeting (more Republican for Republicans, more Democratic for Democrats)
23. *TweetExtremism_Change* – The difference between *TweetExtremism_116* and *TweetExtremism_117*
24. *Dim1_116* – A member's first dimension W-NOMINATE score during the 116th session, calculated as described in the article
25. *Dim1_117* – A member's first dimension W-NOMINATE score during the 117th session, calculated as described in the article using roll call votes after the member's new district was publicly known
26. *Dim1_Change* – The difference between *Dim1_116* and *Dim1_117*
27. *Dim2_116* – A member's second dimension W-NOMINATE score during the 116th session, calculated as described in the article
28. *Dim2_117* – A member's second dimension W-NOMINATE score during the 117th session, calculated as described in the article using roll call votes after the member's new district was publicly known
29. *Dim2_Change* – The difference between *Dim2_116* and *Dim2_117*
30. *VoteExtremism_116* – *Dim1_116*, recoded so that higher values indicate more extreme voting (more conservative for Republicans, more liberal for Democrats)
31. *VoteExtremism_117* – *Dim1_117*, recoded so that higher values indicate more extreme voting (more conservative for Republicans, more liberal for Democrats)
32. *VoteExtremism_Change* – The difference between *Dim1_116* and *Dim1_117*, calculated so higher values are more extreme (conservative for Republicans, liberal for Democrats)
33. *RetweetExtremism_116* – The average W-NOMINATE score of other members of Congress retweeted by the member during the 116th session of Congress
34. *RetweetExtremism_117* – The average W-NOMINATE score of other members of Congress retweeted by the member during the 117th session of Congress
35. *RetweetExtremism_Change* – The difference in *RetweetExtremism_116* and *RetweetExtremism_117*
36. *RedistrictingControl* – The political entity empowered to redraw Congressional district lines between the 116th and 117th sessions, as determined by the Cook Political Report

Dataset #2: Tweets by Members of Congress (116th and 117th)

File Format: R data (.Rda)
Unit of Analysis: An individual tweet
Number of Observations: 2,380,302
Number of Variables: 31
Variable Names and Descriptions:

1. *UserID* – Twitter-provided unique identifier for the Twitter account
2. *TweetID* – Twitter-provided unique identifier for the tweet

3. *ConversationID* – Twitter-provided unique identifier for the conversation the tweet is part of
4. *InReplyToID* – Twitter-provided unique identifier for the user the tweet is replying to (if the tweet is a reply to another tweet)
5. *CreatedAt* – The date and time the tweet was posted
6. *Date* – The date the tweet was posted
7. *TextOriginal* – The text of the original tweet, as downloaded (with shortened URL)
8. *TextExpanded* – The text of the original tweet, as it appeared on Twitter (with full URL)
9. *TweetSource* – Twitter-provided description of what technology was used to post the tweet (e.g., Twitter for iPhone, Twitter Media Studio, Twitter Web App, etc.)
10. *Language* – Twitter-provided description of the language the tweet was written in (e.g., en for English)
11. *AccountType* – Whether the account that posted the tweet is an official, campaign, or personal account of a member of Congress
12. *RetweetCount* – The number of times the tweet was retweeted (as of 11/11/2022)
13. *ReplyCount* – The number of times the tweet was replied to (as of 11/11/2022)
14. *LikeCount* – The number of times the tweet was liked (as of 11/11/2022)
15. *QuoteCount* – The number of times the tweet was quoted (as of 11/11/2022)
16. *MemberName* – Name of the member of Congress that posted the tweet
17. *MemberICPSR* – ICPSR unique identifier for the member of Congress that posted the tweet
18. *MemberParty* – Party of the member that posted the tweet (D for Democrat, R for Republican, or Other)
19. *MemberChamber* – Chamber the member that posted the tweet was serving in at the time (House)
20. *MemberState* – State the member represents
21. *MemberDistrict* – Congressional district the member represents
22. *Session* – Session of Congress when the tweet was posted
23. *ReferenceType* – If the tweet references another tweet, how (either replied_to or quoted)
24. *ReferencedTweetID* – Tweet ID of the tweet that was referenced
25. *ReferencedText* – Text of the referenced tweet
26. *SubreferenceType* – If the referenced tweet references another tweet, how
27. *SubreferenceTweetID* – Tweet ID of the tweet referenced by the referenced tweet
28. *SubreferenceText* – Text of the tweet referenced by the referenced tweet
29. *ProbR* – The estimated probability, based on the tweet text, that the text was posted by a Republican
30. *NOMDim1* – The first dimension W-NOMINATE score of the member
31. *NOMDim2* – The second dimension W-NOMINATE score of the member

Dataset #3: CES Perceptions

File Format: R data (.Rda)
Unit of Analysis: A CES respondent-member pair
Number of Observations: 111,277

Number of Variables: 10
Variable Names and Descriptions:

1. *CaseID* – Unique identifier for the original CES observation
2. *Year* – Year of the CES survey
3. *ICPSR* – Unique ICPSR identifier for the member of Congress
4. *PerceivedExtreme* – Binary indicator coded as 1 if the respondent rated the member as Very Conservative/Conservative (for Republicans) or Very Liberal/Liberal (for Democrats)
5. *TweetExtremism* – The member's TweetExtremism score, from the District and Member Changes (116th and 117th) dataset
6. *Copartisan* – Binary indicator coded as 1 if the respondent identifies with the party of the member
7. *Gender* – Binary indicator coded as 1 if the respondent is male, 2 if the respondent is female
8. *Age* – The age of the respondent, in years
9. *Race* – A categorical variable indicating the self-reported racial identification of the respondent
10. *FamilyIncome* – A categorical variable indicating the self-reported family income of the respondent