# Codebook for Data Described in "A Data-Driven Voter Guide for U.S. Elections: Adapting quantitative measures of the preferences and priorities of political elites to help voters learn about candidates"

March 1, 2016

**Description:**
A data repository on Congressional activity that extends the Database on Ideology, Money in Politics, and Elections (DIME) to cover detailed data on legislative voting, lawmaking, and political rhetoric. It includes data needed to replicate the analysis in "A Data-Driven Voter Guide for U.S. Elections: Adapting quantitative measures of the preferences and priorities of political elites to help voters learn about candidates".

**Principal Investigator:**
Adam Bonica

**Contact:**
Email: bonica@stanford.edu
Webpage: `http://www.stanford.edu/~bonica/`

**How to Cite:**
Bonica, Adam. "A Data-Driven Voter Guide for U.S. Elections: Adapting quantitative measures of the preferences and priorities of political elites to help voters learn about candidates" *Forthcoming.* Russell

# Contents

**1 List of Data Files** **3**

**2 Variable Codings** **3**

2.1 Congressional Text . . . . . 3

2.2 Congressional Voting . . . . . 5

2.3 Bill Database . . . . . 6

# 1 List of Data Files

**Congressional Text:** A table of text documents from the Congressional Record, which contains transcripts of all proceedings, floor debates, and extensions of remarks for Congress. The text corpus was scraped from the GPO's Federal Digital System (FDsys) before being parsed into separate documents. Each document can be linked to either a candidate in DIME or a bill or amendment contained in bill_db.csv. The table includes estimated topic weights for each document using a PLDA model trained on bill category tags assigned by the Congressional Research Service (CRS) (See paper for details).
*Source:* The GPO's Federal Digital System (FDsys) and the Congressional Research Service (CRS).

- text_db.csv - Political text from the Congressional Record (2003-2014).

**Congressional Votes:** A table of Congressional roll call vote choices. Each row records an individual legislator's vote on a specified roll call.
*Source:* Congressional voting records were obtained from Keith Poole and Howard Rosenthal and are available for download at `http://www.voteview.com`.
Cite: Poole, Keith T., and Howard Rosenthal. 2007. *Ideology & Congress*. 2nd rev. ed. New Brunswick: Transaction Publishers.

- vote_db.csv - Congressional roll call voting (2003-2014).

**Congressional Bills and Amendments:** A table of bills and amendments introduced to Congress. Each row list the name, unique identifier, and description of the bill alongside the list of sponsors and cosponsors. The estimated topic weights for bills are calculated by aggregating over all related documents (including related floor debates) and thus can differ from the topic weights found in *text_db.csv* which are based only on the text of the bill. (See paper for details).
*Source:* The GPO's Federal Digital System (FDsys) and thomas.gov.

- bill_db.csv - Bills introduced to Congress (2003-2014).

# 2 Variable Codings

## 2.1 Congressional Text

`Variable Descriptions:`

1) `doc.id:` A unique key for documents.

2) `doc.type:` Type of document ( *bill* - legislative text for bills; *amend* - legislative text for amendments; *speech* - statements made on the floor).

3) `bonica.rid:` DIME recipient ID for the author/speaker.

4) `bill.id:` Unique identifier for bills based on the congress and bill number. This can be used to search for all documents (including speeches) associated with a given bill.

5) `sponsor.rid:` DIME recipient ID for sponsor(s) of related bills. (This can differ from *bonica.rid* in cases where a legislator makes a speech that references a bill sponsored by someone else.)

6) `page.id`: Indicates the page(s) in the Congressional record where the document appears.

7) `congno`: Congress number

8) `legis.body`: Chamber of Congress (House; Senate)

9) `training.set`: Indicator variable for whether a document was included in the PLDA training set.

10) `date`: Date record in the Congressional record.

11) `text`: Raw document text.

12) `stemmed.text`:

13) `doc.labels`: Topic labels assigned to documents derived from the issue caterogies assigned by the Congressional Research Service (See paper for details).

14) `tw.latent1`: Topic weight for latent category.

15) `tw.abortion.and.social.conservatism`: Topic weight for abortion and social conservatism.

16) `tw.agriculture`: Topic weight for agriculture.

17) `tw.banking.and.finance`: Topic weight for banking and finance.

18) `tw.civil.rights`: Topic weight for civil rights.

19) `tw.congress.and.proceedural`: Topic weight for congress and proceedural.

20) `tw.crime`: Topic weight for crime.

21) `tw.defense.and.foreign.policy`: Topic weight for defense and foreign policy.

22) `tw.economy`: Topic weight for economy.

23) `tw.education`: Topic weight for education.

24) `tw.energy`: Topic weight for energy.

25) `tw.environment`: Topic weight for environment.

26) `tw.fair.elections`: Topic weight for fair elections.

27) `tw.federal.agencies.and.gov.regulation`: Topic weight for federal agencies and government relations.

28) `tw.guns`: Topic weight for guns.

29) `tw.healthcare`: Topic weight for health care.

30) `tw.higher.education`: Topic weight for higher education.

31) `tw.immigration`: Topic weight for immigration.

32) `tw.indian.affairs`: Topic weight for Indian affairs.

33) `tw.intelligence.and.surveillance`: Topic weight for intelligence and surveillance.

34) `tw.labor`: Topic weight for labor issues.

35) `tw.law.courts.and.judges`: Topic weight for law, courts, and judges.

36) `tw.transportation`: Topic weight for transportation.

37) `tw.veterans.affairs`: Topic weight for veterans affairs.

38) `tw.womens.issues`: Topic weight for womens issues.

## 2.2 Congressional Voting

1) `vote.id`: Indexes roll call votes.

2) `bill.id`: Unique identifier for bills based on the congress and bill number. This can be used to search for all documents (including speeches) associated with a given bill.

3) `bonica.rid`: DIME recipient ID for the author/speaker.

4) `name`: Name of legislator.

5) `party`: Party of legislator (100 = Dem, 200 = Rep, 328 = Ind).

6) `sponsor`: Indicator variable for whether the legislator sponsored the bill.

7) `cosponsor`: Indicator variable for whether the legislator was a co-sponsor.

8) `vote.date`: Date of vote.

9) `congno`: Congress number.

10) `vote.choice`: Legislators vote choice (1 = Yea; 6 = Nay);

11) `vs.idealpoint`: Vote-specific ideal point for the legislator from the issue-specific OC model.

12) `vs.cutpoint`: The estimated cutpoint on the roll call.

13) `vs.rcdir`: The directionality of the roll call cutpoint. Indicates that legislators with ideal points with values greater than the cutpoint are predicted to vote in favor ($Yea = 1$) or against ($Nay = 6$).

## 2.3 Bill Database

1) `bill.id`: Unique identifier for bills based on the congress and bill number.

2) `bill.str`: Name and details of bill.

3) `bill.desc`: Description of bill.

4) `date`: Date of roll call.

5) `congno`: Bill's congress number.

6) `sponsors`: List of bill sponsors.

7) `cosponsors`: List of co-sponsors.

8) `tw.latent1`: Topic weight for latent category.

9) `tw.abortion.and.social.conservatism`: Topic weight for abortion and social conservatism.

10) `tw.agriculture`: Topic weight for agriculture.

11) `tw.banking.and.finance`: Topic weight for banking and finance.

12) `tw.civil.rights`: Topic weight for civil rights.

13) `tw.congress.and.proceedural`: Topic weight for congress and proceedural.

14) `tw.crime`: Topic weight for crime.

15) `tw.defense.and.foreign.policy`: Topic weight for defense and foreign policy.

16) `tw.economy`: Topic weight for economy.

17) `tw.education`: Topic weight for education.

18) `tw.energy`: Topic weight for energy.

19) `tw.environment`: Topic weight for environment.

20) `tw.fair.elections`: Topic weight for fair elections.

21) `tw.federal.agencies.and.gov.regulation`: Topic weight for federal agencies and government relations.

22) `tw.guns`: Topic weight for guns.

23) `tw.healthcare`: Topic weight for health care.

24) `tw.higher.education`: Topic weight for higher education.

25) `tw.immigration`: Topic weight for immigration.

26) `tw.indian.affairs`: Topic weight for Indian affairs.

27) `tw.intelligence.and.surveillance:` Topic weight for intelligence and surveillance.

28) `tw.labor:` Topic weight for labor issues.

29) `tw.law.courts.and.judges:` Topic weight for law, courts, and judges.

30) `tw.transportation:` Topic weight for transportation.

31) `tw.veterans.affairs:` Topic weight for veterans affairs.

32) `tw.womens.issues:` Topic weight for womens issues.