CHORUS: A new dataset of state interest group policy positions in the United States

Galen Hall
University of Michigan
galenh@umich.edu
Corresponding author

Joshua Basseches
Tulane University
jbasseches@tulane.edu

Rebecca Bromley-Trujillo
Christopher Newport University
rebecca.bromleytrujillo@cnu.edu;

Trevor Culhane
Brown University
trevor_culhane@brown.edu

Codebook

Data

The data used in this project is available online in the SPPQ Dataverse:
https://dataverse.unc.edu/dataverse/sppq. To download it locally, run the `code/download.py` file.

Bills

Columns: `legiscan_bill_id` , `legiscan_bill` , `legiscan_url` , `status_date` , `status` , `last_action_date` , `last_action` , `title` , `description` , `year_start` , `year_end` , `prefile` , `sine_die` , `prior` , `session_title` , `state` , `bill_chamber` , `bill_number` , `bill_suffix` , `sponsor_list` , `description_ncsl` , `ncsl_summary` , `date_overlap_jaccard` , `title_overlap_jaccard` , `ncsl_topics` , `ncsl_databases` , `ncsl_metatopics` , `state_unified_bill_id`

legiscan_bill_id

Example	Description	Type	Source
225543	LegiScan bill identifier	integer	LegiScan

legiscan_bill

Example	Description	Type	Source
LR19CA	Bill chamber and number from LegiScan	string	LegiScan

legiscan_url

Example	Description	Type	Source
https://legiscan.com/NE/bill/LR19CA/2011	The URL of the bill's LegiScan page	string	LegiScan

status_date

Example	Description	Type	Source
2012-02-08	Status date for bill	datetime	LegiScan

status

Example	Description	Type	Source
3	Status value for bill	integer	LegiScan

last_action_date

Example	Description	Type	Source
2012-04-18	Date of last action	datetime	LegiScan

last_action

Example	Description	Type	Source
Indefinitely postponed	Description of last action	string	LegiScan

title

Example	Description	Type	Source
Constitutional amendment to provide that a civil officer is liable to impeachment for misdemeanors in pursuit of office	Short title of bill	string	LegiScan

description

Example	Description	Type	Source
Constitutional amendment to provide that a civil officer is liable to impeachment for misdemeanors in pursuit of office	Long title of bill	string	LegiScan

year_start

Example	Description	Type	Source
2011	Starting year of the session from LegiScan	integer	LegiScan

year_end

Example	Description	Type	Source
2012	Ending year of the session from LegiScan	integer	LegiScan

prefile

Example	Description	Type	Source
0	Flag for session being in prefile (0, 1) from LegiScan	integer	LegiScan

sine_die

Example	Description	Type	Source
1	Flag for session being adjourned sine die (0, 1) from LegiScan	integer	LegiScan

prior

Example	Description	Type	Source
1	Flag for session being archived out of production updates (0, 1) from LegiScan	integer	LegiScan

session_title

Example	Description	Type	Source

Example	Description	Type	Source
2011-2012 Regular Session	Normalized session title with year(s) and Regular/Special Session from LegiScan	string	LegiScan

state

Example	Description	Type	Source
NE	State acronym	string	CHORUS

bill_chamber

Example	Description	Type	Source
LR	The chamber/prefix of the bill	string	CHORUS

bill_number

Example	Description	Type	Source
19	The number of the bill	integer	CHORUS

bill_suffix

Example	Description	Type	Source
CA	The suffix of the bill	string	CHORUS

sponsor_list

Example	Description	Type	Source
['Bill Avery']	List of bill sponsors	string	LegiScan

description_ncsl

Example	Description	Type	Source
Election to Office Misdemeanors	Bill title from NCSL	string	Statenet/NCSL

ncsl_summary

Example	Description	Type	Source
Proposes a constitutional amendment to provide that misdemeanors related to election to office are grounds for impeachment of civil officers.	Summary of bill from NCSL	string	Statenet/NCSL

date_overlap_jaccard

Example	Description	Type	Source
0.897	A measure of how well the date range for this bill in LegiScan overlapped with the date range in StateNet	float64	CHORUS

Additional information: These fields can be used to assess the likelihood that a given bill was incorrectly aligned between LegiScan and Statenet.

title_overlap_jaccard

Example	Description	Type	Source
0.523	A measure of how well the title for this bill in LegiScan overlapped with the title in StateNet	float64	CHORUS

Additional information: These fields can be used to assess the likelihood that a given bill was incorrectly aligned between LegiScan and Statenet.

ncsl_topics

Example	Description	Type	Source
candidates_qualifications_for_office; election_crimes	Topics that the bill addresses, based on matches with NCSL's databases	string	Statenet/NCSL

ncsl_databases

Example	Description	Type	Source
state_elections_legislation_database	NCSL databases that include the bill	string	Statenet/NCSL

ncsl_metatopics

Example	Description	Type	Source
elections_and_campaigns	NCSL metatopics which contain the bill	string	Statenet/NCSL

state_unified_bill_id

Example	Description	Type	Source
NE_LR_0000000019_2011	Internal bill identifier	string	CHORUS

Clients

Columns: `client_name` , `state` , `source` , `source_eid` , `ftm_industry` , `state_client_id` , `ftm_guessed`

client_name

Example	Description	Type	Source
WISCONSIN ENVIRONMENT	The name of the client as it appeared in the original data source.	string	CHORUS, FollowTheMoney, OpenSecrets, or Google Knowledge Graph

Additional information: For CHORUS clients, this name matches the raw string parsed from the documents scraped from state legislatures, which may vary across different sources and over time for the same organization.

state

Example	Description	Type	Source
WI	The US state.	string	CHORUS

source

Example	Description	Type	Source
ftm	The dataset from which this client was sourced.	string	CHORUS, FollowTheMoney, OpenSecrets, or Google Knowledge Graph

source_eid

Example	Description	Type	Source
ftm_24761754	The client's id in its dataset of origin. Since clients in CHORUS are assigned separate state_client_ids, this field is blank for names gathered from CHORUS.	string	CHORUS, FollowTheMoney, OpenSecrets, or Google Knowledge Graph

Additional information: Since clients in CHORUS are assigned separate state_client_ids, this field is blank for names gathered from CHORUS.

ftm_industry

Example	Description	Type	Source
PRO-ENVIRONMENTAL POLICY	The client's industry, coded using Follow The Money's categories. The determination of each industry was done either by Follow the Money directly, or using text analysis (see "ftm_guessed" field).	string	FollowTheMoney

Additional information: Note the ftm_guessed field.

state_client_id

Example	Description	Type	Source
WI_773	The client's id in CHORUS.	string	CHORUS

ftm_guessed

Example	Description	Type	Source
False	If False, the industry categorization was done by Follow the Money. If True, the industry is a guess based on the name of the client, using a Naive Bayes classifier trained on the non-guessed data.	boolean	CHORUS

Positions

Columns: `client_name` , `lobbyist_rep_name` , `lobbyist_firm_name` , `position` , `start_date` , `end_date` , `year` , `record_type` , `session` , `description` , `committee` , `state` , `position_numeric` ,

`docket_number` , `docket_prefix` , `legiscan_bill_id` , `state_unified_bill_id` , `state_client_id` , `ncsl_topics` , `ncsl_metatopics` , `bill_version`

client_name

Example	Description	Type	Source
Direct Energy Services, LLC	The name recorded by the client (interest group) for this position	string	CHORUS

lobbyist_rep_name

Example	Description	Type	Source
Paul T. Donovan	The name of the lobbyist recording this position on behalf of the client	string	CHORUS

lobbyist_firm_name

Example	Description	Type	Source
Kearney, Donovan & McGee, P.C.	The name of the lobbyist's employer firm	string	CHORUS

position

Example	Description	Type	Source
NEUTRAL	The string position, i.e. what was originally written on the record.	string	CHORUS

start_date

Example	Description	Type	Source
2011-01-01 00:00:00+00:00	The first date on which the position was reported.	datetime	CHORUS

end_date

Example	Description	Type	Source
2011-06-30 00:00:00+00:00	The last date on which the position was reported, or the end of the reporting period, if applicable.	datetime	CHORUS

Additional information: Not all positions have end_dates, and in general end_dates are taken from the end of the applicable reporting period, so they should not be seen as strict bounds on the position. It may be best practice to assume a given position holds until either (a) the text of the bill changes from its value at the `start_date` or (b) the client records a new position on that bill.

year

Example	Description	Type	Source
2011	The year in which the position was reported.	integer	CHORUS

record_type

Example	Description	Type	Source	Possible Values
lobbying	The type of disclosure record from which the position was collected: "lobbying" for lobbying disclosures and "testimony" for all forms of public hearing records.	string	CHORUS	"lobbying", "testimony"

session

Example	Description	Type	Source
187	The session as reported by the legislature	string	CHORUS

description

Example	Description	Type	Source
HB37, HB223, [omitted], SD2009.	A general-purpose field for descriptive notes attached to the position.	string	CHORUS

Additional information: In this example, the positions given to us by Massachusetts included a description column containing the plaintext, comma-delimited lists of bills for each position (support/oppose/neutral) from each client, from which we extracted one row per unique bill.

committee

Example	Description	Type	Source

Example	Description	Type	Source
Telecommunications Utilities and Energy	The committee in which the position was stated, if applicable.	string	CHORUS

state

Example	Description	Type	Source	Possible Values
MA	The US state.	string	CHORUS	"MA", "MO", "NJ", "NE", "MD", "IL", "TX", "OH", "KS", "FL", "AZ", "CO", "IA", "WI", "MO", "SD"

position_numeric

Example	Description	Type	Source	Possible Values
0	A numeric representation of the position which maps "oppose" and variants thereof to -1, "support" to +1, and "neutral" and all other ambiguous positions to 0.	integer	CHORUS	-1, 0, 1

docket_number

Example	Description	Type	Source
2009	The docket number of the prefiled bill, which we can sometimes use to map to final bill numbers when, as in this case, the position was stated before bills were assigned final numbers.	integer	CHORUS

docket_prefix

Example	Description	Type	Source
SD	The chamber in which the bill was prefiled.	string	CHORUS

legiscan_bill_id

Example	Description	Type	Source
12345678	The ID in the LegiScan database, if possible	integer	LegiScan

state_unified_bill_id

Example	Description	Type	Source
MA_S_0000001958_2011	The unified identifier used to reference bills.	string	CHORUS

Additional information: Format is [state]*[chamber]*[10-digit bill number]_[session]

state_client_id

Example	Description	Type	Source
MA_548	The id used to identify clients.	string	CHORUS

ncsl_topics

Example	Description	Type	Source
candidates_qualifications_for_office; election_crimes	Topics that the bill addresses, based on matches with NCSL's databases	string	Statenet/NCSL

ncsl_metatopics

Example	Description	Type	Source
elections_and_campaigns	NCSL metatopics which contain the bill	string	Statenet/NCSL

bill_version

Example	Description	Type	Source
26242	Where available, any identifier assigned by the state legislature to distinguish versions of a bill.	string	CHORUS

Additional information: States with bill version metadata: AZ, IL, WI, FL, TX, RI. Note that not all bills will have versions in some of these states; for instance, in Texas, versions only appear where they were parsed from witness list documents (e.g. as testimony on a particular amendment).

Block Assignments

Columns: `entity_id` , `0` , `1` , `2` , `3` , `4` , `5` , `6` , `state` , `record_type`

entity_id

Example	Description	Type	Source
MA_1, or MA_H_0000000001_2020	The state_client_id or state_unified_bill_id corresponding to the node in the HBSBM graph with this block structure	string	CHORUS

0

Example	Description	Type	Source
MA_lobbying_3283	The lowest-level block assignment (i.e. most granular). With increasing levels above this one the blocks become larger and less granular. The index of the highest-level block varies in different states as the HBSBMs equilibrate to different depths depending on the amount and richness of the data.	string	CHORUS

Additional information: With increasing levels above this one the blocks become larger and less granular. The index of the highest-level block varies in different states as the HBSBMs equilibrate to different depths depending on the amount and richness of the data.

1

Example	Type	Source
MA_lobbying_150	string	CHORUS

2

Example	Type	Source
MA_lobbying_0	string	CHORUS

3

Example	Type	Source
MA_lobbying_12	string	CHORUS

4

Example	Type	Source

Example	Type	Source
MA_lobbying_3	string	CHORUS

5

Example	Type	Source
MA_lobbying_3	string	CHORUS

6

Example	Type	Source
MA_lobbying_2	string	CHORUS

state

Example	Description	Type	Source
MA	The US state.	string	CHORUS

record_type

Example	Description	Type	Source
lobbying	The type of disclosure record from which the position was collected: "lobbying" for lobbying disclosures and "testimony" for all forms of public hearing records.	string	CHORUS

References

FollowTheMoney. "Home - FollowTheMoney.Org." (https://www.followthemoney.org/). Data from FollowTheMoney is licensed under CC BY 4.0

Google Cloud. "Google Knowledge Graph Search API | Enterprise Knowledge Graph." (https://cloud.google.com/enterprise-knowledge-graph/docs/search-api).

LegiScan. "LegiScan | National Session Data Archives." LegiScan. (https://legiscan. com/datasets). Data from Legiscan (Legislative Datasets) by LegiScan LLC is licensed under CC BY 4.0

National Conference of State Legislatures. "NCSL 50-State Searchable Bill Tracking Databases." (https://www.ncsl.org/technology-and-communication/ncsl-50-state-searchable-bill-tracking-databases).

OpenSecrets. "OpenSecrets." (https://www.opensecrets.org/). Data from OpenSecrets is licensed under CC
    BY 4.0