Codebook for dataset used for
"Dynamic State Interest Group Systems: A New Look with New Data"
Interest Groups & Advocacy
8 (December): 499 – 518, 2019

Thomas T. Holyoke
Department of Political Science
California State University, Fresno
2225 East San Ramon, MS MF19
Fresno, California 93740-8029
tholyoke@csufresno.edu

Last updated – August, 2023

If you use this data, please cite the article
UPDATES: Dataset now includes data from 2006 to 2017
Dataset now includes data from 2006 to 2020 (as of August 2021)

The data in this dataset is cross-sectional, times-series, organized by both state and year. There are fifty observations, one for each state, for each year, yielding 600 observations overall.

Variable name	Full description	Variable type
year	Indicates the year of the organization observation, ranging from 2006 to 2019	Continuous
state	Indicates the state containing the observed organization. It does not include the District of Columbia	Categorical
union	The number of unions lobbying in each state for each year	Count
bus	The number of businesses engaged in lobbying in each state for each year (see the appendix below for information on how these variables were coded)	Count
nonprofit	The number of social service nonprofits engaged in lobbying in each state for each year	Count
citizen	The number of citizen groups and public interest groups engaged in lobbying in	Count
association	The number of trade or professional associations engaged in lobbying in each state for each year	Count
gov	The number of government agencies (and similar entities) engaged in lobbying in each state for each year	Count
agriculture	The number of organizations coded as lobbying for agricultural businesses / interests (see the appendix below for more information as to what kinds of businesses / organizations comprise this sector)	Count

business	The number of organizations coded as lobbying for general businesses / interests involved with general business, including the service and manufacturing sectors (see the appendix)	Count
communications	The number of organizations coded as lobbying for communication businesses / interests (see the appendix)	Count
construction	The number of organizations coded as lobbying for construction businesses / interests (see the appendix)	Count
education	The number of organizations coded as lobbying for educational interests (see the appendix)	Count
electronics	The number of organizations coded as lobbying for electronic or technological businesses / interests (see the appendix)	Count
energy	The number of organizations coded as lobbying for energy and natural resources businesses / interests (see the appendix)	Count
entertainment	The number of organizations coded as lobbying for professional entertainment businesses / interests (see the appendix)	Count
finance	The number of organizations coded as lobbying for banking, insurance, and investment businesses / interests (see the appendix)	Count
government	The number of organizations coded as lobbying for government agencies and related interests (see the appendix)	Count
health	The number of organizations coded as lobbying for medical, health insurance, and related businesses / interests (see the appendix)	Count

ideological	The number of organizations coded as lobbying for ideological or public interest causes not clearly associated with another sector (such as the ACLU) (see the appendix)	Count
lawyers	The number of organizations coded as lobbying for the legal industry and related interests (see the appendix)	Count
leisure	The number of organizations coded as lobbying for the tourism, food, drink, and related businesses / interests (see the appendix)	Count
realestate	The number of organizations coded as lobbying for real estate businesses / interests (see the appendix)	Count
socialservice	The number of organizations coded as lobbying for social service nonprofits and related interests (see the appendix)	Count
transportation	The number of organizations coded as lobbying for transportation businesses / interests (see the appendix)	Count
total	The sum of all organizations in all sectors, effectively being a count of the total number of organizations and businesses lobbying in each state for each year	Count
ig	The number of organizations that are citizen groups, public interest groups, or trade or professional associations engaged in lobbying in each state for each year	Count

Appendix

Most of the data in "Dynamic State Interest Group Systems" comes from the National Institute for Money in State Politics ("the Institute).[i] The Institute, in turn, gathered it from the agencies in each state responsible for registering entities lobbying state governments from 2006 to 2015 (processes described here were later repeated for the data from 2016 and 2017 when I received it from the Institute). While the Institute has made the data available through their website for years, it sent me the data in EXCEL spreadsheets.[ii] I used this data to create a working data set with one entry for each organization in each state for each year. The paper, however, uses a summary data set where there is on line totaling the number of organizations in each state for each year, meaning there are 500 observations in the summary data set (it is 600 now with the additional years). As I describe the data collection process below, I try to be clear as to which data set I am describing, though most of the material refers to the master data set, which, of course, was used to create the summary data set.

The Institute has tried to make the lists of interest groups from state to state as comparable as possible. Mindful of concerns regarding comparability, the Institute has twice had their data evaluated by researchers at the RAND Corporation, and RAND has validated the way it collected and audited the data (including all of the contributions data made available by the Institute). RAND did note, though, that each state has different registration and disclosure requirements, and different levels of rigor when it comes to ensuring compliance, making some cross-state comparisons difficult (see McGovern and Greenberg 2014, pp. 6-8).[iii]

This appendix outlines the procedures used to clean the raw data from the Institute forming the master data set, and aggregations of that data are used to create the summary data set, used in the paper. This appendix also explains the methods used to code interest groups and other lobbying

organizations by type and economic sector. Finally, there is additional information on how the comparison of this data with that of Gray and Lowery was done.

Addressing Errors in the Institute's Data (original data, 2006 – 2015)

After receiving the raw data, I found that for some states there were anomalies in certain years. Some states had years where the number of lobbying entities was strangely large, while for other years it was strangely small. These were:

California	-	2009
Colorado	-	2012
Iowa	-	2011
Massachusetts	-	2012, 2013, and 2014
Nevada	-	2010 and 2014
New Jersey	-	2010, 2011, and 2012
New York	-	2012
Pennsylvania	-	2013 and 2014
Tennessee	-	2012
Vermont	-	2012

To correct for these anomalies, I went to the websites for the agencies in these states that register lobbying organizations and searched for correct data. I was able to find it for Tennessee and Iowa without any problems. In New York, the new data yielded just slightly greater group numbers than what I received from the Institute, rising from 1,307 to 1,418, which still does not make much sense. I therefore averaged the 2010, 2011, 2013, and 2014 data to yield 3,709 in the summary data set used for this paper, but could not make similar changes to the master data set. For Colorado, I simply could not find a way to download a full spread sheet of data as the online system appears to only return one group at a time for each search. The average of the two years on either side of 2012 is 1,180, so this number is used in the summary data set but not the main data set. In Vermont, the website only has data beginning in 2015, so I looked at the raw 2013 data I received from the Institute and found that there were many duplicate entries unique to that

year. After these were removed, the numbers of organizations for 2014 was a logical 476, so corrections were made to both the summary and master data sets. It turns out that in Nevada during most even years the legislature does not even meet except for special session, and thus the interest groups registered for the prior year should just carry over, which I have done for both the summary and master data sets. In other words, for 2010, I used groups for 2009, and for 2014 I used groups for 2013.

It also turned out that even the data downloaded from the state agency websites for California and Massachusetts did not line up with the data from the Institute for any year. Consequently, I deleted all of the Institute data for these two states and replaced it with data from the agencies for all ten years in both data sets. The new data I obtained for the three years in New Jersey, however, was more or less the same as reported by the Institute, and thus perpetuated the odd shape of the data trend. The same is true of Pennsylvania. Since I am at a loss to explain this, New Jersey and Pennsylvania are omitted from the paper where indicated, but at this point remain in the Summary and Main data sets. Strickland (2018), however, points out that in Pennsylvania there are numerous loopholes in the registration law that *might* account for the strange variation.

A few other states also showed some odd patterns in the annual data. Corrections for these were made in both data sets unless otherwise indicated. In North Dakota it turned out that in every even year the number of groups listed was much smaller than in the odd years, though 2015 is also strangely small. The state's legislature, it turns out, only meets in odd years, so odd years are more likely to be an accurate representation of the number of groups active in the states, not just those registered that year. I therefore changed the even number states to reflect the number of groups in the prior odd year, though 2006 and 2015 were left unchanged. A comparison of the changes made for this, Nevada, and the three other states listed below are in a table on pages 5 and 6.

Maine also has the peculiar pattern where the number of registered groups in even years is always lower than in the odd. While its legislature does meet in even years, it is on a shorter, less active basis. Thus, again, the odd years are the more accurate measure of the number of lobbying organizations, so all even years are changed to be the same number as the prior odd year number. In Montana the number of groups dipped from 2011 to 2012 (to 308), and then suddenly leaped up by over 300 groups (which is big for Montana) in 2013. I averaged the two years before and the two years after 2012 and replaced the number with the average (but only in the summary data set). In South Dakota, 2014 (at 251) is 170 lower than the prior year and nearly 131 lower than the year afterwards. I therefore averaged the three prior years and 2015 to replace 2014, though this correction was only made in the summary data set (UPDATE: in later iterations, I averaged 2012, 2013, 2015, and 2016 to replace 2014 in the summary data set). Finally, Texas also has even numbered years somewhat less than odd and is a biennial legislature, so even years are replaced by the prior odd year in the summary data set.

States where the even years were oddly lower than the odd years and the corrections made in the summary data set

Year and state	**Original**	**Final**
Nevada		
2006	969	874
2007	1067	962
2008	957	867
2009	1063	985
2010	403	985
2011	1189	1084
2012	1179	1079
2013	1158	1059
2014	106	1059
2015	1276	1130

North Dakota

2006	139	135
2007	448	431
2008	196	431
2009	430	418
2010	128	418
2011	426	413
2012	190	413
2013	448	437
2014	291	437
2015	199	200

Maine

2006	324	315
2007	415	398
2008	334	327
2009	368	363
2010	265	363
2011	402	370
2012	330	370
2013	424	410
2014	303	410
2015	417	402

Texas

2006	2749	2156
2007	3518	2684
2008	2907	2684
2009	3245	2711
2010	2802	2711
2011	3444	2803
2012	3017	2803
2013	3454	3086
2014	2903	3086
2015	3511	3128

Montana

2006	450	438
2007	469	461

2008	469	461
2009	440	436
2010	589	443
2011	462	427
2012	308	533
2013	638	625
2014	649	637
2015	647	631

South Dakota

2006	404	341
2007	427	363
2008	483	399
2009	467	379
2010	428	386
2011	456	427
2012	449	437
2013	421	405
2014	251	411
2015	382	375

Coding by Economic Sector

The first batch of data received from the Institute was for the years 2006 to 2012. In 2016, the data for 2013 to 2015 was delivered. The overall number of original observations received was 1,806,218, but this raw data had three problems. First, there were many duplicate entries, often because multiple lobbyists represented the same organizations or corporations during the same year. Indeed, as Benz et al. (2011, p. 445) note in their article using state PAC contributions data from the Institute, frequent duplication is a recurring issue with the Institute's data and a great deal of "cleaning" is required before it can be used. Fortunately, it was relatively simple to use Stata statistical software to identify and delete the duplicative entries, which reduced the data set to 577,686 observations where each organization was entered once for each year in each state.

Second, each organization received a label reflecting the economic sector it operated in, which came in three columns of information ranked in order of specificity: sector, industry, and business. The majority of observations, however, were actually not coded at all. Furthermore, there is a peculiar temporal pattern in the number of uncoded observations; they seem to increase in later years, as seen in this table:

Year	Number of uncoded observations	Percentage of all observations
2006	22,156	41%
2007	23,717	41%
2008	29,398	50%
2009	28,128	52%
2010	41,516	74%
2011	46,630	77%
2012	47,670	79%
2013	38,548	66%
2014	37,616	66%
2015	41,197	68%
Total	356,576	62%

Third, many of those that were coded were done so incorrectly. When the Institute staff were asked about this in a private communication, they said that they were not aware of any real problem, and that there had not been any change in their coding rules or loss of internal funds dedicated to this project that might account for the growing number of uncoded entries.[iv] Because other scholars have made some use of this data for one or two specific years (see Witko and Newmark 2005; Gray et al. 2015), this problem with the Institute's data has become known in the community of scholars working on state interest group politics.

The decision was made to code the missing sector data. For two reasons I focused on coding the "sectors" category rather than the other, more specific categories provided by the Institute. First, since "sector" is broadest, and would therefore be the easiest to code with any

sense of validity and reliability. Second, with just a little adjustment these sector codes are made to correspond to the National Association Industry Classification Codes (NAICS) used by the U.S. Department of Commerce, which makes it possible to link other variables (such as sector employment) to these codes.

A few procedures made it possible to quickly reduce the number of uncoded organizations. For instance, it turned out that a large number of organizations not coded in one year, or in one state, were coded by sector in other years or states. Since there is no reason to believe a group or corporation would change its sector from one year to another, it was relatively simple to use statistical software to automatically find a code for an organization based on its name and then apply that same code to all other observations with the same name. Just using this method reduced the number of uncoded observations from 356,576 to 202,143, or from 62% to 35%. A second technique was to use software to pick out words in a group or corporation's name that identified it as belonging in a sector category. For example, it was possible to pick out all organization names containing words like "bank," "insurance," "hospital," "labor union," and "doctor" and categorize them, respectively, as "finance," "finance," "health," "union," and "health." Visual checking was required to make sure that a blood "bank" did not end up in the "finance" category. After spending about a month using this technique, the number of uncoded observations fell to 140,625, or 26%. Along the way, I discovered that there were a few thousand entries where there was no organization name at all, or the name was so badly misspelled that the entry was unreadable. All of these observations were removed, reducing the total number of usable observations in the main data set for the ten-year period to 545,797.

After this, six research assistants and I spent over three years working through a process of looking-up uncoded organizations on the internet and assigning a code when possible when we

could agree that it fitted one of the NAICS codes. Fortunately, we had the entire list sorted alphabetically, so after assigning a code to one organization in one year, we could assign the same code to the same organization in all other states and years. Since most of the coding was done by graduate assistants, all trained with the same instructions (see below), I randomly sampled 5% of each student's work and recoded them myself to see how well each student was doing. For five of the six students employed, my coding and the students matched in over 90% of cases. In one sad instance, however, the assistant was found to be frequently miscoding, and in a few instances had invented entirely new codes. This student was let go and I recoded all 20,000 (or so) of that student's entries. It is for reasons such as these that this coding process took nearly three years. After the work was complete, we found 6,963 cases with personal names or no name at all. These were deleted. In the end, only 26,207, or 4.8%, of entries remain uncoded by sector. In the Main Data Set now, there are 545,797 entries, with 519,590 having usable sector codes.

Here are the instructions provided to each graduate assistant along with the basic coding categories and more explicit information as to what kind of organization goes into each category:

"**General instructions**

Your focus is on the column "sectors," which will be used to identify the general economic / social sector each group primarily works in. Look at the name of the organization in the "groupname" column and try to give it a code in the "sectors" column. Using the list below of sectors, please type into the EXCEL sheet cell the sector that you think best fits the organization and its purpose. If you cannot decide, please leave it as "uncoded."

If you come across a person's name with no marks like "LLC" or "PC," and there are a lot of them in the data, please leave them as "uncoded."

Do not work too long on any single entry. If you cannot find out enough about it in about five minutes of searching, leave it "uncoded."

Sometimes you will find useful hints and clues in the columns "business" or "industry."

Agriculture

- Agricultural services and products
- Crop production and basic processing
- Dairy
- Food processing and wholesale
- Livestock
- Miscellaneous agriculture
- Poultry and eggs
- Tobacco growers (but not cigars or other tobacco products)
- Fish and fisheries
- NOT forests and forestry

Business (general)

- Business services
- Cemeteries and funeral services
- Chemical and related manufacturing
- Defense and aerospace
- General contractors
- Management consulting
- Marketing and advertising
- Miscellaneous distributing
- Miscellaneous business services
- Manufacturing (all kinds)
- Retail sales
- Large pharmacy corporations (like CVS)
- Waste management services
- Recycling services
- Otherwise unidentifiable business organizations

Communication industry

- Telecommunications services and equipment
- Cable television
- Media (print and video) such as magazines and newspapers (and therefore journalism), but not fiction or academic publishing
- Telephone utilities
- Wireless internet and fiber-optics

Construction

- Industry and business
- Home building
- Development of land for construction
- Landscaping
- Architecture
- Roads and highway building
- Relevant unions
- People who do AC / heating systems and electrical wiring for houses and buildings
- NOT raw materials used in building or real estate related services

Education

- K-12 education
- Charter schools, private schools, religious schools, and home schooling
- Higher education
- Teachers' unions
- Private school education and management companies
- Education professionals associations
- Does NOT include boards of education and education departments, which are classified as "government," but does include their associations
- Academic publishing
- School libraries and librarians
- School transportation and health

Electronics

- Computer and electronics manufacturing
- Computer equipment and services
- Other electronics services
- Software development and service

Energy and natural resources

- Utilities
- Mining of all kinds of minerals
- Distribution of all types of power
- Forest management and logging
- Environmental services

- Hunting
- Water services and water quality
- Land use issues
- Conservation
- Environmental activism

Entertainment

- Live theater and acting
- Live music (orchestras, jazz, rock, etc.)
- Musician and actors' unions
- Theater and concert venue management
- Fictional publishing
- Convention centers (but not visitors' bureaus)
- Sports and sport teams
- TV and movie filming and production

Finance and insurance

- Accounting
- Banking
- Insurance (including health insurance and vehicle insurance)
- Pay-day lending and check cashing
- Investing and finance
- Financial management
- Tax services
- Securities investing

Government

- All state and local government agencies
- Includes school districts, economic development agencies, and other special districts
- Includes associations representing government agencies
- Most "commissions" and all "authorities"
- Public sector unions

Health

- Advocates for health care and home health care services
- Medical unions (especially nurses' unions)

- Hospitals
- Medical service providers (doctors, dentists, nurses, etc.)
- Nursing homes
- Medical technology research companies
- Pharmaceuticals (production and pharmacies), but not big pharmacy chains
- Health care management systems
- Does NOT include health insurance providers or health insurance plans, which are classified as "finance"
- Health and medical certification boards
- NOT colleges and university medical teaching schools, but includes university hospitals

Ideological (not used in the paper)

- Political cause oriented organizations not clearly involved in another policy sector (usually has words like "Citizens," "Alliance," "Action," or similar words in the organization's title)
- Religious groups and churches (activist oriented)
- IMPORTANT: If the organization clearly works in another sector, then code it for that sector. For example, the American Cancer Society is essentially an ideological group, but all of its work is in the health sector, so it is coded "Health." Only ideological groups that do not clearly fit into another sector are coded "Ideological Groups"
- Organizations with words like "coalition" or "committee" in their names

Lawyers and Lobbyists (not used in the paper)

- All law firms (usually identified by a string of names followed by LLC)
- Lobbying firms may appear like law firms, but also may include words "consultants," "public affairs," or "advocates"
- Also includes all associations of lawyers involved with the judicial system, like groups representing judges, defense attorneys, prosecutors, and so forth (but not judicial administrative offices or associations)

Leisure and entertainment

- Hotels and resorts (and cruises)
- All alcohol production and distribution
- All forms of gambling (including horse and dog racing)
- Anything to do with beaches
- Anything to do with auto racing

- Food catering (but not food production or wholesale, or even retail supermarkets)
- Tourism industry (including travel agents)
- Hunting
- Art centers and museums
- Visitors' bureaus
- Libraries (except government libraries and university libraries)
- Outdoor outfitting / clothing / outdoor gear industry

Real estate

- Realtors and associations
- Realtor certification boards
- Title services (but not title insurance)
- Home inspection services
- Other home-selling and buying services

Social service (used as "nonpolitical nonprofits" in the paper)

- General nonprofit organizations
- Welfare and social policy organizations
- Organizations involved with helping marginalized populations
- Organizations dealing with social problems like drugs and alcoholism
- Counseling services
- Veterans groups
- Religious organizations (non-activist)

Transportation

- Air travel and air cargo
- Truck transportation
- Railroad transportation (people and cargo)
- Automotive industry (manufacture and sales)
- Anything in support of car travel
- Water transportation (oceans, lakes, and rivers)"
- Relevant unions

Key words used to identify group member types

Three binary variables were created capturing three kinds of lobbying organizations: businesses, trade and professional associations (including unions), and advocacy groups with both members (citizen groups) and just donors (public interest groups). The other two types used in Figure 4 of the paper, social service nonprofits and government agencies, comes from the economic sector categories. To create these variables I performed key word searches on the names of the observed organizations (using Stata's "regexm" command) to identify groups containing identifying words in their names. However, I did not solely rely on this. After the lists of organizations in each category were generated, I read through these lists and removed those organizations that clearly did not belong in that category. Those that were ambiguous were looked up online before a decision on whether to retain them on the list was made. All words were searched in both capitalized and un-capitalized versions.

<u>Advocacy groups (citizens and public interest groups):</u>

academy
Action
Advoc
Americans for
attorney
campaign
Caucus
Center
christian
Citizen
Club
collaborative
collective
committ
Cong
Demo
Forum
foundation

Fund
Grower
Guild
institute
Labor
Lawyer
League
Legal
Legion
medical
movement
Now
Nurse
Project
Society
task force
Trust
Union
United
Vet
Voice
Watch
Women

Associations of businesses, institutions, or individual professions:

academy of
alliance
associated
association
business
chamber
Church
coalition
college of
conference
congress of / for
consortium
council of / for
federation
Group

hospital
industry
institute
Insur
League
manufacturers
museum
of America
partnership
producers
professional
roundtable
society of / for
Trade
university

Businesses

These were identified using words such as "Incorporated," "Inc," "LLC," and "PC". Otherwise businesses were identified by simply looking through the complete list of organizations in the Institute's data.

Cross-Checking with the Gray and Lowery Data

As noted in the main paper, I took two approaches to comparing Gray and Lowery's data for 1975, 1980, and 1990, 1997 to 1999, and Strickland's (2018) data for 1989, with the Institute data ranging from 2006 to 2015. The first was to calculate differences each year between the state with the least lobbying entities and the state with the most. Here are the raw results:

States	Year	Lowest #	Highest #	Difference
MS and FL	1975	40	596	556
MS and FL	1980	140	872	732
MS and FL	1989	97	2,473	2,376
MS and FL	1990	107	2,969	2,862
HI and CA	1997	191	2,106	1,915
WY and CA	1998	114	2226	2,112
WY and CA	1999	72	2,272	2,200

HI and CA	2006	292	3,016	2,724
HI and FL	2007	289	3,114	2,825
RI and FL	2008	310	3,265	2,955
HI and CA	2009	263	3,045	2,782
HI and NY	2010	254	3,734	3,480
WY and FL	2011	125	3,456	3,331
WY and FL	2012	142	3,760	3,618
WY and NY	2013	169	4,000	3,831
WY and NY	2014	263	3,959	3,696
WY and CA	2015	183	3,732	3,549

I also calculated the number of states falling in quarters created by subtracting the lowest number of groups that year from the highest and dividing the resulting range into fourths. Here are the results, which arguably show that the number of organizations falling into these ranges remained more or less consistent over time.

Year	First fourth range and # of states	Second fourth range and # of states	Third fourth range and # of states	Fourth fourth range and # of states
1975	40 to 179 22	180 to 318 16	319 to 457 3	458 to 596 2
1980	140 to 323 24	324 to 506 14	507 to 689 1	690 to 872 4
1989	97 to 691 38	692 to 1,286 9	1,287 to 1,881 0	1,882 to 2,473 2
1990	107 to 823 38	824 to 1,539 10	1,540 to 2,254 0	2,225 to 2,969 1
1997	191 to 669 33	670 to 1,148 12	1,149 to 1,627 3	1,628 to 2,106 3
1998	114 to 641 28	642 to 1,169 14	1,170 to 1,697 3	1,698 to 2,226 3
1999	72 to 621 28	622 to 1,171 12	1,172 to 1,722 5	1,723 to 2,272 3
2006	292 to 973	974 to 1,654	1,655 to 2,335	2,336 to 3,016

	32	10	4	3
2007	289 to 995	996 to 1,701	1,702 to 2,407	2,408 to 3,114
	34	8	3	4
2008	310 to 1,049	1,050 to 1,788	1,789 to 2,527	2,528 to 3,266
	32	11	2	4
2009	263 to 959	960 to 1,655	1,656 to 2,350	2,351 to 2,782
	31	11	3	4
2010	254 to1,124	1,125 to 1,998	1,999 to 2,866	2,867 to 3,734
	33	12	1	3
2011	125 to 958	959 to 1,790	1,791 to 2,623	2,624 to 3,456
	31	12	2	4
2012	142 to 1,046	1,047 to 1,950	1,951 to 2,854	2,855 to 3,760
	31	13	2	3
2013	169 to 1,127	1,128 to 2,085	2,086 to 3,043	3,044 to 4,000
	33	11	1	4
2014	263 to 1,187	1,188 to 2,111	2,112 to 3,035	3,036 to 3,959
	33	11	1	4
2015	183 to 1,070	1,071 to 1,957	1,958 to 2,844	2,845 to 3,732
	31	12	2	4

UPDATED FOR 2022 DATA

Notes on problematic data in the state interest groups database and what I did about it.

For the majority of states, there were no problems. However, there were problems for these states:

- Arkansas – There is a strange dip in 2018 going almost to zero. Since this is obviously wrong, I corrected it by simply averaging 2015, 2016, 2017, and 2019 in the summary file (but not the master file).

- California – The 2009 data looked weird, so I downloaded the data for 2009 from the website of the California Secretary of State, which codes data by sector in the same way as the Institute, so it was an easy replacement. However, it turned out that other years from the California SOS did not match the Institute data, so I dumped all of the Institute data and replaced it all with SOS data. This change was made to both the master data set and the summary data set. See http://access.sos.gov/Lobbying/Employers/

  For 2016 to 2019, the California data received from the Institute seems just fine, so I use it.

- Colorado – The 2012 and 2017 data were weird looking. It was not possible to get replacement data from the state itself because it would only look up one group at a time, not yield a full list. I therefore averaged 2010, 2011, 2013, and 2014 data to replace 2012 data in the summary data set (but not the master data set). For 2017, I simply substituted the 2016 data (summary data set only).

  It turned out that 2019 was also strange, but 2018 fit the trend, so, as I substituted 2016 for 2017, I also substituted 2018 for 2019 (only in the summary file).

- Florida – The Institute data for 2016 and 2017 was strange, so replacement data was obtained from the state and substituted for those years in both the master data set and the summary data set

- Iowa – The 2011 data was strange, but I was able to get complete replacement data for both the master data set and summary data set from https://www.legis.iowa.gov/lobbyist/reports/

  For 2019 I went back to the same website and downloaded the 2019 data, coded it, and added it to both the master data file and the summary file (and removed the old, bad 2019 data).

- Louisiana – The 2017 data was strange, so new data was found and replaced in both the master and summary data set

- Maine – Here also the even numbered years were always lower than the odd numbered years. Again, even numbered years were replaced by the prior odd year, but only in the summary data set.

- Maryland – The data is weirdly low for 2019, so I averaged 2016, 2017, and 2018 in the summary data file, but not the master file.

- Massachusetts – In this case the data for 2012, 2013, and 2014 was all way off from data for the other years, and so was the data for 2017. I was able to find replacement data for the first three years from the state at http://www.sec.state.ma.us/lobbyistpublicsearch/ and was able to code that. The new data was substituted in both the master data set and the summary data set. For 2017, I simply used the 2016 data in the summary data set (no change in the master data set).

  Since the 2018 and 2019 data from the Institute was also strange, I dumped it and retrieved data from 2017 to 2020 from the website above, cleaned it, and replaced the bad data in both the Master File and the Summary File.

  Same with 2021 and 2022, got new data and replaced it in the Master File and Summary File.

- Minnesota - The 2017 data was strange, so new data was found and replaced in both the master and summary data set

- Missouri – The 2019 data was a bit strange, so I replaced it with the 2018 data.

- Montana – There is an odd pattern here, with the total number of organizations dips a bit from 2011 to 2012 (to 308), and then leaps up over 200 more groups (large for Montana) in 2013. I can't find original replacement data, so in the summary data set only I threw out 2012 and replaced it with the average of 2010, 2011, 2013, and 2014.

  Bizarre swings continued through 2019. Best not to use this state.

- Nevada – I noticed that the 2010, 2014, and 2016 data all looked strange. Nevada's legislature tends to only meet in odd years, though it meets for a special session in even years (if needed). I therefore decided to get rid of all even year data and substitute the numbers from the prior (odd) year. For 2016, the data was really weird, so I averaged the 2015 and 2017 data to replace that year. These changes were only made to the summary data set.

The Institute data for 2018 and 2019 seemed just fine, so I used it

- New Jersey – The data for this state was really strange. I went to http://www.elec.state.nj.us/publicinformation/lobby_statistics_archive.htm and downloaded the data there. It was exactly like the Institute data, with the same weird trends. I cannot explain it or account for it, so this state should probably not be used in any analysis.

  Now, though, the 2016 through 2019 data is all fine, so the problem is limited to 2010, 2011, and 2012. I will average 2008, 2009, 2013, and 2014 and use that for the three weird years (summary file only). New Jersey can be used again in analyses.

- New York – Here it was the 2012 data that was strange. I went to https://onlineapps.jcope.ny.gov/LobbyWatch/menu_reports_public2.aspx and found replacement data for 2012, but it was nearly the same as the Institute's data and still strange. I therefore averaged 2010, 2011, 2013, and 2014 to replace 2012, but only in the summary data set.

  The Institute data for 2018 and 2019 seems fine so I used it.

- North Dakota – Since the state legislature only meets in odd years, this explained why all of the even year numbers were much lower. Every even year number was dropped and replaced with the prior odd year. This change was only made to the summary data set. It still looks a bit weird, so use this data very carefully.

  Then the pattern changes again, and it is 2015 that is oddly low and 2016 appears fine, but then 2017 through 2019 are really low. Perhaps best not to use this state.

- Ohio - The 2017 data was strange, so new data was found and replaced in both the master and summary data set

- Pennsylvania – There were many slightly weird years here, but the really concerning ones were 2013 and 2014. I went to https://www.palobbyingservices.pa.gov/Public/wfSearch.aspx and found replacement data for 2013 and 2014, which was used in both the master data set and summary data set. The data for 2016 was also weird, and this was solved in just the summary data set by averaging 2015 and 2017. The same thing was done with 2018, averaging 2017 and 2019. For 2022 I went back to the above link and got replacement data used only in the summary data, but it is again surprisingly small. Something is really weird here.

NOTE: In 2022 I revamped this. The data looks completely accurate from 2006 to 2012, so no changes are made to this. Although data for the years 2013 and 2014 still look surprisingly low coming after 2012, since this is official state data collected directly from the state and replaced in the master data set, I will continue to use it. Starting in 2015, the odd years appear to have good, consistent data, but the even years are all much lower. I cannot find a reason for this, but it seems that registrations are abnormally low in all even years, so for 2016, 2018, 2020, and 2022 I replace them in the summary data set with the odd year right before each observed year (so 2022 is actually 2021 counts).

- Rhode Island - The 2016 data was strange, so new data was found and replaced in both the master and summary data set.

  2018 was also very low, so I averaged 2015, 2016, 2017, and 2019 in the summary data set.

- Tennessee – This state had weird data for 2012, so I went to https://apps.tn.gov/ilobbysearch-app/search.htm and found replacement data for both the master data set and summary data set.

- South Dakota – The 2014 total organizations was 251, but this was about 150 lower than 2013 or 2015, so I threw out the 2014 data and average data from 2011, 2012, 2013, 2015, 2016, and 2017 to replace it in just the summary data set.

- Texas – Another odd-year only legislature, so all even numbered years were thrown out and replaced by the prior odd year in just the summary data set.

- Vermont – Here the weird data was 2013. The Vermont website only has data from 2015, so I averaged data from 2011, 2012, 2014, and 2015 to replace 2013 in the summary data set, but not the master data set.

References

Benz, Jennifer, Justin H. Kirkland, Virginia Gray, David Lowery, Jennifer Sykes, and Mary Deason. 2011. "Mediated Density: The Indirect Relationship between U.S. State Public Policy and PACs." *State Politics & Policy Quarterly* 11(December): 440 – 459.

Gray, Virginia, John Cluverius, Jeffrey J. Harden, Boris Shor, and David Lowery. 2015. "Party Competition, Party Polarization, and the Changing Demand for Lobbying in the American States." *American Politics Research* 43(2): 175 – 204.

Strickland, James. 2019. "A Paradox of Political Reform: Shadow Interests in the U.S. States." *American Politics Research* 47(July): 887 – 914.

Witko, Christopher, and Adam J. Newmark. 2005. "Business Mobilization and Public Policy in the U.S. States." *Social Science Quarterly* 86(June): 356 – 36

---

[i] The organization's current website is http://www.followthemoney.org/. Overall it can be perhaps best described as a state-level counterpart to the better known Center for Responsive Politics, which tracks interests groups, lobbying, and campaign spending in Washington, DC.

[ii] After 2012 the Institute stopped making this data available, at least for a time. We received it for 203 to 2015 by special request.

[iii] The McGovern and Greenberg reference is to the second, 2014, evaluation report. The first report is not available, though the Institute published a summary paragraph from the report on its website at http://www.followthemoney.org/our-data/rand-evaluations/. In the summary paragraph RAND researchers summed by writing "Overall, NIMSP is a professional, well-managed institution that can be proud of its processes and people, and of the integrity and validity of the valuable research data and products they make available to the world in a useful form."

[iv] Even so, it is not entirely clear that this project is still ongoing at the Institute.