The new Voteview.com: preserving and continuing Keith Poole's infrastructure for scholars, students and observers of Congress

Adam Boche* Jeffrey B. Lewis† Aaron Rudkin‡ Luke Sonnet§

March 8, 2018

This project was made possible by support from The William and Flora Hewlett Foundation (Grant #2016-3870), the National Science Foundation (NSF-SBS-0611974), University of California Los Angeles Social Science Computing, and the University of Georgia.

Abstract

For the last forty years, Keith Poole has developed and curated a trove of basic data and measurements related to the United States Congress. He has made these resources freely available through his widely-used Voteview.com website since 1995. At Poole's Voteview.com, scholars, students, journalists, and the broader public could download gold-standard historical and current roll-call voting data, member rosters,

*Department of Political Science, University of California Los Angeles; aboche@ucla.edu
†Department of Political Science, University of California Los Angeles; jblewis@ucla.edu
‡Department of Political Science, University of California Los Angeles; rudkins@ucla.edu
§Department of Political Science, University of California Los Angeles; luke.sonnet@gmail.com

NOMINATE scores, and measures and visualizations of party cohesion and polarization, among many other useful things. In this article, we describe how we are preserving and continuing these vast public-goods contributions through the new Voteview.com. Developed and housed at UCLA, the new Voteview.com carries on the creation of basic roll-call data infrastructure, including the assignment of Poole's widely-used ICPSR number-like identifiers to new members, data on every roll-call vote ever taken, NOMINATE scores and other standard roll-call vote-based measures such as party-loyalty scores. In addition to serving as a platform for the continuation and dissemination of this basic data infrastructure, the new Voteview.com also provides powerful tools for exploring the history of roll-call voting, the US Congress, and American politics and political history through a simple search interface and interactive visualizations.

1 Introduction

Keith Poole is an uncommonly generous scholar who has dedicated a great deal of his career to the production of public goods. In addition to his well-known and important substantive contributions to American politics and political history and the widely-used statistical tools that he developed,[1] Poole built, curated and freely shared with the scholarly community and the public a vast trove of data on voting in the US Congress. These data and measures derived from them are staples of scholarship, pedagogy and, increasingly, journalistic reporting on American politics and government. Much of this research infrastructure has been constructed and maintained in a labor-intensive way by Poole himself.

With Poole's retirement approaching, something needed to be done to keep the infrastructure that he has been providing from falling into disrepair. Here, we describe a new

[1] For examples of Poole's substantive contributions see McCarty et al. (2016), Poole and Rosenthal (2000*a*), Poole (2007), McCarty et al. (2009) and McCarty and Poole (1995). For leading examples of Poole's methodological contributions see Poole (2005), Poole (1998) and Poole and Rosenthal (1985).

online platform that we have constructed to continue Poole's great work as the curator of the history and study of roll call voting in the United States Congress.[2] A central aim of our project is to continue the ongoing collection of new voting data, curate the existing data, and maintain updated versions of the core measures derived from these data that Poole has been providing. As described below, our new Voteview.com not only provides access to that core data infrastructure, but also offers tools for exploring and visualizing voting in Congress over its history. In this way, it supersedes not only Poole's original website, but also the companion personal computer-based software for vote exploration and visualization first developed by Poole, Rosenthal and Shor in the late 1980s and early 1990s (Voteview for DOS and Voteview for Windows; see Poole and Rosenthal (2000*b*)).

Poole's Voteview.com site also was the distribution hub for important measurements of congressional behavior derived from the roll-call voting data. In the 1980s, Keith Poole and Howard Rosenthal developed the NOMINATE model and estimator which places legislators in an abstract ideological space (for a detailed discussion see Poole, 2005, 1998; Poole and Rosenthal, 1985). They found that throughout nearly all of its history, patterns of congressional roll-call voting can accurately be summarized by locating each member along a primary liberal–conservative axis that describes preferences over fundamental issues of taxation, spending and redistribution and a secondary axis that describes preferences over social issues, such as civil rights during the 1950s and 1960s. While the algorithm that recovers the two coordinates of legislators' locations on the ideological map is computationally intensive (it was originally run on supercomputers), the interpretation is intuitive and accessible. Given the locations of legislators on this ideological Cartesian plane, each roll call can be

[2]Poole's Voteview.com also served as an informal repository for a variety of roll call voting data sets from legislative bodies other than the United States. Those data are included on our new site as part of our general archive of all Poole's Voteview.com. Our new site, as described here, is narrowly focused on voting in the United States Congress.

represented by a straight line that divides those who are expected to support the motion from those who are expected to oppose it. Across all contested roll calls, over 87% of all voting decisions are correctly classified by these predictions.[3] Any vote taken can quickly be analyzed and understood through a simple graph that shows the location of each member participating in the roll call and a line representing the division between supporters and opponents (those voting counter to the prediction also can be identified). NOMINATE scores have become a staple of the analysis of congressional voting. Over a thousand papers available on JSTOR as of this writing reference NOMINATE scores.[4] The new Voteview.com provides up-to-date DW-NOMINATE scores as well as a variety of standard measures and visualizations derived from those scores.[5]

In what follows, we describe the data and measurements that we provide on the new Voteview.com along with details about the manner in which they are produced and how they can be accessed. We then describe the tools that Voteview.com provides for exploration and visualization of roll call voting from across congressional history, highlighting the ways in which these tools can be used not only for research, but also for teaching. We conclude with a discussion of future directions that we will take and how interested scholars can contribute to the project.

[3]A contested roll call is one for which neither the yeas nor the nays received over 97.5% of the vote.

[4]The url http://www.jstor.org/action/doBasicSearch?Query=D-NOMINATE+or+DW-NOMINATE+or+W-NOMINATE, accessed on January 11, 2018, yielded 1,091 results. A similar query on Google Scholar® — https://scholar.google.com/scholar?q=%22DW-NOMINATE%22+%7C+%22D-NOMINATE%22+%7C+%22W-NOMINATE%22,accessed on January 11, 2018, yielded "about" 4,040 results.

[5]Several additional NOMINATE algorithms have followed the original in the literature, including W-NOMINATE, D-NOMINATE, Alpha-NOMINATE, DW-NOMINATE and Nokken-Poole scores. Voteview.com provides DW-NOMINATE and Nokken-Poole scores. Users who want W-NOMINATE or Alpha-NOMINATE can easily estimate those on their personal computers using R packages that implement those procedures and the `Rvoteview` package described below to access roll-call data in R.

2 Data resources provided on Voteview.com

The Voteview.com database provides combined historical and modern sources of data about roll-call votes and legislators. This section presents the core data collections, including roll-call votes, legislator biographical information, party descriptions, "key votes", presidential positions and images of original source documents.

2.1 Roll-call data

Since 1995, Poole's Voteview.com website provided a complete archive of roll-call voting data from the United States Congress. Voteview.com provided worldwide free access to such basic data as descriptions and enumerations of every roll-call vote taken in the United States House of Representatives and Senate and a complete congressional roster, including party affiliations. These data are built upon, correct and extend collections originally undertaken by the Inter-university Consortium for Political and Social Research (ICPSR). Beginning in 1991, Keith Poole, working first with Nolan McCarty and then since 2006 with Jeffrey Lewis, has continued the collection of the congressional roll-call data in a format that is compatible with the original ICPSR collections. Our Voteview.com continues that work. As of January 22, 2018, the Voteview.com database includes information on all 24,174,546 individual votes cast by 12,297 members on 105,721 rollcalls over the Congress's 229-year history.

As part of his continuation of the ICPSR roll-call voting data series, Poole has carried out such fundamental tasks as continuing the assignment of unique identifiers to each new member of Congress started by ICPSR. These ICPSR numbers are the standard identifiers for members of Congress when creating and linking datasets related to the study of the Congress. We describe below how we continue Poole's process of generating these identifiers. Because of Poole's high level of curation, the datasets provided on Voteview.com have become canonical in the discipline. The basic data are complemented on Voteview.com by other important

collections, such as issue codings for each vote taken and a complete roster of members of the Congress throughout its history.

For each roll-call vote in Congress since the first in 1789, the new Voteview.com provides a description of the issue being considered, a record of how each member voted, and further textual information. Basic roll-call data comprise the measure's title, bill number, the procedural question (e.g., "On passage"), and enumeration of the votes cast by each member (e.g., yea/nay/abstain). For current votes, the roll-call data automatically are updated from Congress throughout each day. New votes with accompanying estimates become available to the public within minutes of their availability from Congress. In addition, for bills since 1973, Voteview.com provides bill summaries produced by the Congressional Research Service. Whenever possible, each bill's sponsor is identified from the original source documents.

2.1.1 Legislation subjects

Voteview.com provides subject-matter labels for roll-call votes using several classification schemes. Poole and Rosenthal (1991) coded each roll call for the 1789-2015 period according to the six-category Clausen Classification (Clausen, 1973), the finer-grained Peltzman Classification (Peltzman, 1984), and the Poole and Rosenthal 99-category specific-issue codes (Poole and Rosenthal, 2017). The Congressional Research Service (CRS) has assigned to each bill since 1973 one of 32 "Policy Area Terms", such as "Agriculture and Food" (Congress, 2016*b*). Since 2009, CRS also has assigned at least one of more than 500 specific "Legislative Subject Terms", such as "Agricultural insurance" (Congress, 2016*a*). For each roll-call vote on a bill, Voteview.com provides the list of subjects assigned to the bill. All of these descriptors are stored in the core Voteview.com database and can be used to search for particular roll calls, as described below. In future enhancements to Voteview.com, we plan to create a crosswalk between the CRS's subject terms and the older Clausen, Peltzman, and Poole and

Rosenthal subject codes, allowing for those codes to be applied to the new roll calls without the need for additional human coding.

2.1.2 Original source documents

Voteview.com's dataset of historical roll calls for the 1789-1873 period is derived from the printed journals of Congress. Those sources were the basis of Clifford Lord's Works Progress Administration-funded effort to catalog the Congress's roll-call votes, which subsequently was digitized and archived by the ICPSR.[6] Many corrections to that original collection have been made by Poole and we continue to make efforts to uncover and correct errors.

The original ICPSR roll-call datasets include page references to the journals for each roll call. Using those page references and recently digitized versions of the journals, Voteview.com makes it simple to see the original description of each vote in the original source materials (these include the *Annals of Congress*, the *House Journal*, the *Senate Journal*, the *Congressional Globe* and the *Congressional Record*). Users can compare the summarized vote data with integrated images of the original source documents on Voteview.com for each vote during that period as well as explore the surrounding debates and legislative activity. The original documents provide context for each vote, including records of floor discussion (The Library of Congress, 2017). This also allows users to quickly identify any remaining errors in the original data collections. We hope that by making the cost of error-checking low, we can effectively crowd-source the discovery of any remaining errors in the historical roll-call data.

[6]The basis of the roll calls included in the Voteview.com database for the 1st though the 105th Congress is ICSPR Study #4; see https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/4.

2.2 Additional roll-call meta-data

Several organizations have identified certain votes or bills as politically significant. Voteview.com provides a dataset of "key votes" sourced from CQ-Roll Call, which has recognized over 1,000 specific votes since 1945. Wikipedia includes descriptions and additional information on a number of significant bills, which Voteview.com provides for each roll call on the bill. The Wikipedia descriptions are useful because they usually include the popular common names of important legislation, such as "Obamacare", that often are not part of a bill's official title, description, or summary. By including these commonly used names in the database, it is easier for users to find the votes related to the legislation they are studying.

2.2.1 Presidential position

The president, while not a legislator, often expresses a position on specific bills. These "presidential positions" as cataloged by CQ-Roll Call are provided on Voteview.com. By interpreting the president's position as a synthetic "pseudovote", Voteview.com can estimate an ideal point for each president in an ideological space shared with members of Congress. Voteview.com provides presidential positions and ideal-point estimates through 2015. We currently are evaluating replacement sources for the canonical presidential position moving forward.

2.2.2 Legislator information

Voteview.com provides summary biographical information for each member of Congress, sourced from the Biographical Directory of the United States Congress, a publication of the Congressional history offices (United States House and Senate Historical Office, 2017). These data include dates of birth and death, educational attainment, occupation, military service, other government positions held, and family relationships to other members of Congress.

For each vote cast by a legislator, the legislator is identified by office, constituency and party affiliation at the time of the vote.

Members are labeled with an identifier so that the 61 individuals assigned multiple ICPSR[7] numbers by Poole are uniquely identified, and all members' names are normalized to the official form recognized by the Biographical Directory. Voteview features photographic or portrait images for over 70% of members, including all members serving since 1950.

We continue Poole's procedure for generating and assigning ICPSR-like numbers to new members, which, for convenience and continuity, we continue to call "ICPSR numbers." Legislators who have not served in Congress previously are issued a new ICPSR number—the next available in the sequence for their Congress and chamber. For example, new members of the 115th Congress (2017) are in the sequences starting from 21700 and 41700 in the House and Senate, respectively.[8] Legislators who switch chambers retain their original number, while legislators undertaking a significant change of party are issued new numbers. Presidents are issued new numbers (in the 99000 sequence) regardless of prior service in Congress. Given the rate of incoming new members, we expect to continue the current number-assignment procedure into the foreseeable future.

2.2.3 Parties

Data related to parties include their full names and shortened names, and the commonly used "party codes" assigned by ICPSR. In total, members of Congress have identified with over 50 different parties throughout US history. Voteview.com reports the years in which parties were active in the Congress, the size of each party, the parts of the country from

[7]ICPSR numbers are the member identifiers first assigned by the ICPSR and then extended by Poole to the current Congress.

[8]Modern Senate ICPSR numbers begin with '4'; House numbers begin with '2', overflowing to '3' as needed. The second and third digits indicate the year since 2000 beginning each Congress. For members who switch parties, in all appearances after the switch, the first digit is replaced replaced with '9'.

which each party elected members, and the ideologies of members of those parties. We also track party-level summaries of ideology over time.

Each member has a party affiliation, and Voteview provides information about all of these parties. In early years, party affiliations were less distinct fixtures of politics than they are today, and judgement is required in labeling each member with a single party affiliation. The major historical work of Kenneth Martis's *Historical Atlas of Political Parties in the United States Congress, 1789-1989* (Martis, 1989) provides the primary source of information on parties and members' party affiliations. Examples of these party assignments include the decision to code members of the first two Congresses as belonging to "Pro-Administration" and "Anti-Administration" factions, and coding members who served during the fractious period between the First and Second Party Systems according to the Democratic-Republican presidential candidate they supported. For members who changed parties through their careers, Voteview.com reports the appropriate party membership for each Congress. For parties that have historically assumed multiple names, Voteview.com presents the names used by Martis. Voteview.com visualizations and maps (described in detail below) adopt the party colors used by Poole and Martis, with two exceptions: First, when multiple parties sharing colors would appear in the same Congress, smaller parties are assigned a color similar to their larger affiliates. Second, Democrats, historically red, are labeled blue, and Republicans red, to conform to the contemporary convention.

2.2.4 Districts

The Voteview.com database is spatially enabled and includes district boundary definitions for every congressional district in US history. These boundaries are provided by Lewis et al. (2013) for the pre-2000 districts and by geographic information systems (GIS) data from the US Census for post-2000 districts. These boundaries are used to render thematic maps of roll-call votes on Voteview.com and also to look up the members representing any location

in the country from throughout its history (as described in more detail below).

2.3 DW-NOMINATE scores on Voteview.com

Currently, Voteview.com generates two distinct kinds of ideal point estimates. Each measure relies on a particular version of the DW-NOMINATE algorithm. As described above, DW-NOMINATE represents legislators on a two-dimensional map by giving each an "ideal point" (secondary use of DW-NOMINATE by the academy and media have almost exclusively used only the first-dimension positions). The closer together legislators are on this map, the closer their voting records are expected to be. DW-NOMINATE also estimates a "cutting line" for each roll call. These cutting lines split the two-dimensional space such that legislators on one side of the line are more likely to vote Yea than Nay, while legislators on the other side of the line are more likely to vote Nay than Yea on the corresponding roll call. In addition to estimating the ideal point for each legislator and the cutting line for each roll call, DW-NOMINATE also includes two global parameters. The first parameter estimates the importance of the second dimension (w) relative to the primary dimension; the second parameter is proportional to the fundamental uncertainty assumed to underlie the votes that legislators cast (β). A complete description of the algorithm can be found in many other references, but was first presented in McCarty et al. (1997) and is discussed in detail in Poole (2005).

The first kind of DW-NOMINATE scores that we estimate — those used in the web visualizations and the main estimates provided in the downloadable datasets — are *Common-Space Constant* DW-NOMINATE scores. These are the scores that Poole reported at the top of his website. For these scores, House and Senate members are scaled in a single space (*Common Space*) and individual legislators have a constant ideal point throughout their time in the Congress (*Constant*).[9] Discussion of these assumptions is found elsewhere, including Poole

[9]Because we do not allow the members' locations to vary over time, our scores are based on a statistical

(2005) and Carrol et al. (2009). The sole exceptions to the constant-ideal-point constraint are the small number of members who switch from one major party to another during their time in office. These members are allowed separate locations corresponding to the periods that they served in each party to which they belonged.[10]

In the past, these Common-Space Constant DW-NOMINATE scores were updated by Poole on Voteview.com on a roughly weekly basis when Congress was in session. Poole would add the new roll calls of the week into the complete dataset which would then be processed using an implementation of the DW-NOMINATE program in the Fortran programming language, generating updated ideal point and cutting line estimates. Poole ran the entire algorithm on the full dataset, re-estimating all roll call parameters, ideal points, and the w and β hyper-parameters each week. He used the previous week's estimates for the ideal points and roll-call parameters as starting values in order to ensure a good fit and reasonable run time. Nonetheless, this process would still take several hours, was initiated manually, and required close human oversight and monitoring.

We have implemented the DW-NOMINATE algorithm in Python (Python Core Team, 2018). The new code takes advantage of opportunities for efficient parallel processing which are inherent in the DW-NOMINATE algorithm to speed estimation. The modular structure of the new code makes it very easy to estimate some of the DW-NOMINATE model's parameters while conditioning on others.[11] This allows us to quickly provide provisional

model that is identical to that developed for W-NOMINATE. However, the way in which the data are organized and the model is fit is that used in DW-NOMINATE. Because we are providing what Poole called Common Space DW-NOMINATE scores, we call our scores "DW-NOMINATE" despite the fact the each members' location is fixed over time.

[10]To be precise, the DW-NOMINATE algorithm will assign a unique ideal point to every unique ICPSR code, and members switching major parties or becoming president are the only times this happens. Examples of when a member is assigned a new ICPSR code without becoming president include Arlen Specter switching to the Democratic Party in 2009 and Strom Thurmond switching to the Republican Party in 1965.

[11]As with Poole's original Fortran algorithm, the new algorithm is freely available on the web. It can be

cutting lines and ideal-point updates as new votes are taken, and then re-estimate a fuller model each night. The added speed also makes it easier to run the algorithm through more iterations in order to ensure convergence.

More specifically, as new roll calls are added to the database throughout the day (generally within a few hours of being taken), we estimate two sets of values: cutting lines for the new roll calls and ideal points for the legislators who voted on the new roll calls. This means that when conducting the real-time analysis, we condition on, and hold as fixed, the existing roll-call cutting lines, the ideal points for legislators who did not vote on the new roll calls, and the hyper-parameters w and β. At the end of each day that sees new roll calls enter the database, we go back and jointly re-estimate all ideal points for legislators who voted on the new roll calls and the cutting-lines for all of the roll calls on which those legislators voted.

Thus, we update the ideal points of all legislators who are still actively casting votes and the cutting lines of all roll calls that those legislators ever voted on (including roll calls in past sessions of Congress). The ideal points of members who have departed from Congress and never return are not updated as new roll calls are taken. Similarly, roll calls for which all of the votes were cast by now-departed legislators also are not updated. We do this to ensure the long-run comparability of the current and future DW-NOMINATE scores with those produced by Poole in the past. Jointly estimating the complete DW-NOMINATE model over all of the votes is difficult (as described in McCarty et al. (1997, Appendix) and Poole (2005)). It has been said that Poole himself was the "outer loop" of this estimation process: his judgement and expertise were required in the estimation of the original values. By effectively locking in place the locations that Poole last estimated for past members, we guarantee that our scores maintain compatibility with the widely used DW-NOMINATE scores with which scholars are familiar.[12] In particular, this procedure

found at https://github.com/voteview/pynominate.

[12]Furthermore, the hyper-parameters w and β described above are not updated regularly in this process.

establishes the scale, location, and rotation of the ideological space to be that estimated by Poole. In principle, the estimation of previously serving members could be improved by "back propagating" information provided by recent votes all the way back to members serving in the first Congress, relying on overlapping time periods of service to identify the results. In practice, improvement (in terms of fit) of this back propagation appears to be quite small. Nonetheless, we are exploring alternative approaches and, in the future, we may fully re-fit DW-NOMINATE to the entire database of votes as new votes are ingested. Those who are interested in re-estimating the full model from scratch can use the code and data that we make available in order to do so.

Furthermore, at the end of each session of Congress, we provide estimates of uncertainty around the DW-NOMINATE ideal points using a parametric bootstrap, akin to those described in Carrol et al. (2009). Interested users can generate these bootstrapped estimates of uncertainty using our open-source Python module.[13]

The second kind of DW-NOMINATE score that we provide are known as *Nokken-Poole scores*, described in detail by Nokken and Poole (2004). These scores allow legislators' ideal points to move freely over time, and thus make less restrictive assumptions about legislator ideological fixedness.[14] In order to maintain the comparability (identification) of the scores over time and across chambers while allowing members to hold different positions in each Congress and chamber in which they serve, Nokken-Poole scores take roll-call cutting lines

They are unlikely to change in meaningful ways in the short run, and to ensure long-run comparability of estimates, they will be re-estimated periodically—most likely at the end of each Congress.

[13]See https://github.com/voteview/pynominate.

[14]Poole also regularly produced DW-NOMINATE estimates for the House and Senate separately that allowed the ideal points of long-serving members to evolve linearly over time. Voteview.com currently provides only Nokken-Poole scores for scholars interested in studying the changes in positions of members over time, and in the future we may provide other scores based on models that permit members' ideal points to change over time.

estimated by DW-NOMINATE (as described above) as fixed, and then proceed Congress-by-Congress to find the ideal point for each legislator that best describes their voting behavior on bills in that Congress. Thus, these estimates offer a useful measure of how members of Congress may have changed their ideological positions over time. Although the Poole-Nokken procedure does not fully optimize jointly over the cutlines and ideal points, they do allow members' locations to vary freely over time and Poole-Nokken scores are estimated easily from the Common-Space Constant DW-NOMINATE estimates that we produce.

While Poole regularly produced other estimates separately for the House and Senate that allowed for linear changes in ideal points over time as well as non-parametric (Optimal Classification) ideological rankings and W-NOMINATE scores for members of each chamber in each Congress, we provide only the widely used Common-Space Constant DW-NOMINATE scores. However, as described below, we provide an R API to our database that along with R packages implementing all of Poole's scaling routines make it straightforward for users to produce these other measures.

2.4 Accessing the data

A primary goal of Voteview.com is to provide access to the rich set of roll-call metadata, member ideal points, party statistics, and other information contained in the Voteview.com database. The web interface, discussed below, is a user-friendly way for anyone interested in a particular roll call or set of roll calls to access these data in familiar and easy-to-use formats. Users who want complete data for a given Congress or the entire history of Congress in a format suitable for analysis can download plain-text data files at http://voteview.com/data. This page also contains full documentation for each field included in the data. These files are rebuilt as new votes are taken and as any errors in the existing data are corrected. One major improvement present in the new Voteview.com site is that all of these files are now

produced through queries to a single underlying database. Thus, new data and corrections are reflected in every downloadable file at all times. Users can download data about roll calls, members, and political parties for the history of Congress or for specific years and chambers. We also make available an R (R Core Team, 2013) package, `Rvoteview` (Lewis et al., 2017), which allows users easily to search our database for people and roll calls and download various kinds of data directly into an R workspace for further analysis. Finally, a complete copy of Voteview.com's full database is updated weekly and available for download by power users who want unfettered access to all of the information housed on the server.

Data on individual members of Congress, downloaded either through the `Rvoteview` package or using the plain-text files, include their Common-Space Constant DW-NOMINATE ideal points and their Nokken-Poole scores. These data can easily be merged with other data using the provided ICPSR codes, Congress number and office identifiers, or member names. If users are interested in information about particular roll calls, including DW-NOMINATE cutting-line estimates or meta-data about passage and sponsorship, these data also are available as plain-text files on Voteview.com and through the R package. Users who want the full record of member votes on a specified set of roll calls can use the R package to access those data as an R `rollcall` object, a class of object used by the existing `wnominate` and `pscl` packages for ideal point analysis in R. This allows users to easily download a subset of the Voteview.com database and do bespoke ideal point analysis without having to manually build the vote matrices. More details about this and other functionality available in the `Rvoteview` package can be found in the vignette that accompanies the package.[15]

Among the many areas of the study of Congress and American Politics more generally that have been propelled by NOMINATE scores, perhaps none is more currently salient than the study of political polarization. In their 2016 book *Polarized America*, McCarty, Poole and Rosenthal draw heavily on the distribution of DW-NOMINATE scores across political

[15]See https://github.com/voteview/Rvoteview/tree/master/vignettes.

parties to document the ebb and flow of political polarization in the United States over its history and the rapid increase in polarization that has occurred over the last 30 years. Subsequently, much of the scholarly consideration of polarization can be characterized as the attempt to understand inter-party differences in DW-NOMINATE scores over time. Voteview.com provides summaries of DW-NOMINATE scores by Congress, chamber, and party that can be used to explore ideological polarization throughout congressional history. These summaries are provided in plain-text files along with downloadable plots that can be used in teaching, research or other communication. Along with the data and associated plots, we also provide the R code that is used to calculate these summaries and to make the plots in the widely used Rmarkdown (Rstudio Inc., 2018) format. Users can see how we generate each summary and plot and easily adapt the provided code to customize them.

In addition to summaries based on DW-NOMINATE scores, we also provide a growing number of other basic summaries, such as presidential- and party-support scores along with corresponding plots and the underlying Rmarkdown documents. The plots and summaries are updated as new data are ingested and corrections are made to the existing data. We are opening Voteview.com to scholars, and accept submissions of article- or note-length summaries of congressional voting using the same Rmarkdown framework through a GitHub repository.[16].

3 Exploring the US political history on Voteview.com

In addition to simple data downloads, Voteview.com provides for six major user activities: searching and exporting roll-call votes; visualizing roll-call votes; listing, exporting and visualizing congressional membership over time; visualizing individual member's voting histories; tracking party ideology over time; and visualizing representation of a geographical

[16]See https://github.com/voteview/articles

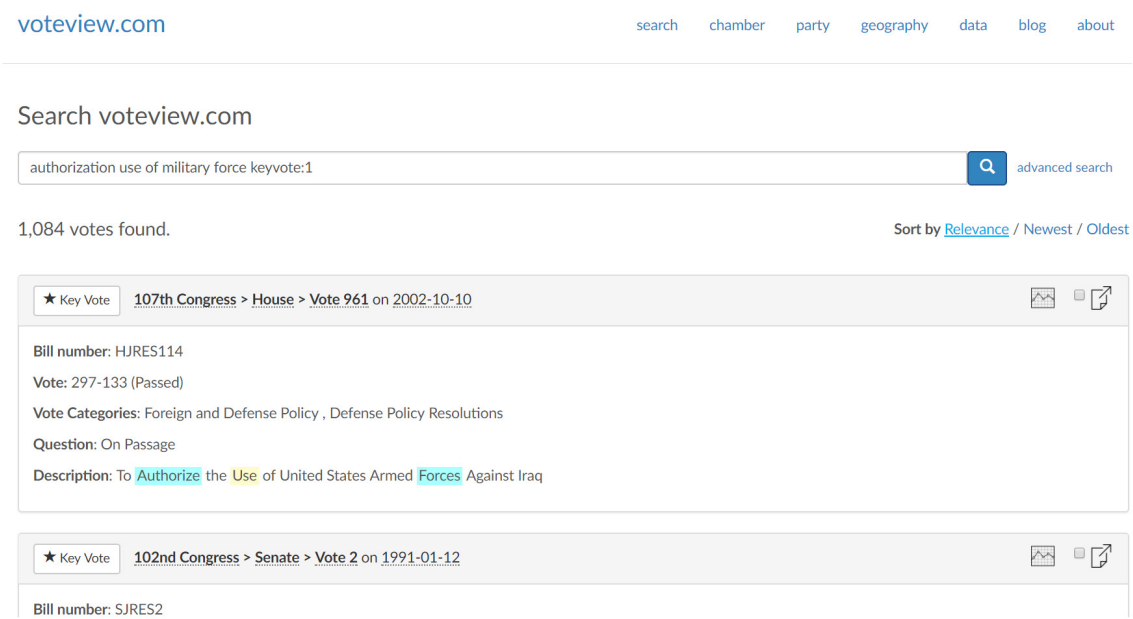

Figure 1: Searching rollcalls on Voteview.com

place over time. These functions emerged out of a desire to offer user-friendly interactive versions of some of Poole and Rosenthal's published work, Poole's work on Voteview.com and the desktop-based Voteview for DOS and Voteview for Windows software produced by Poole, Rosenthal and Shor in the late 1980s and 1990s.

3.1 Searching the roll-call database

The primary point of entry for most users is the front page search engine, which supports flexible and powerful searching of the Voteview.com database. An example of Voteview.com's search interface is shown in Figure 1. The search engine returns results related to votes, members and parties, features a full suite of boolean search algebra as well as field-based advanced searching to narrow votes by chamber of Congress, date, result, margin or category metadata. Matches to queried terms are sought in the title, question, description, other available summaries and subject codes for each vote. Users interested in exporting votes

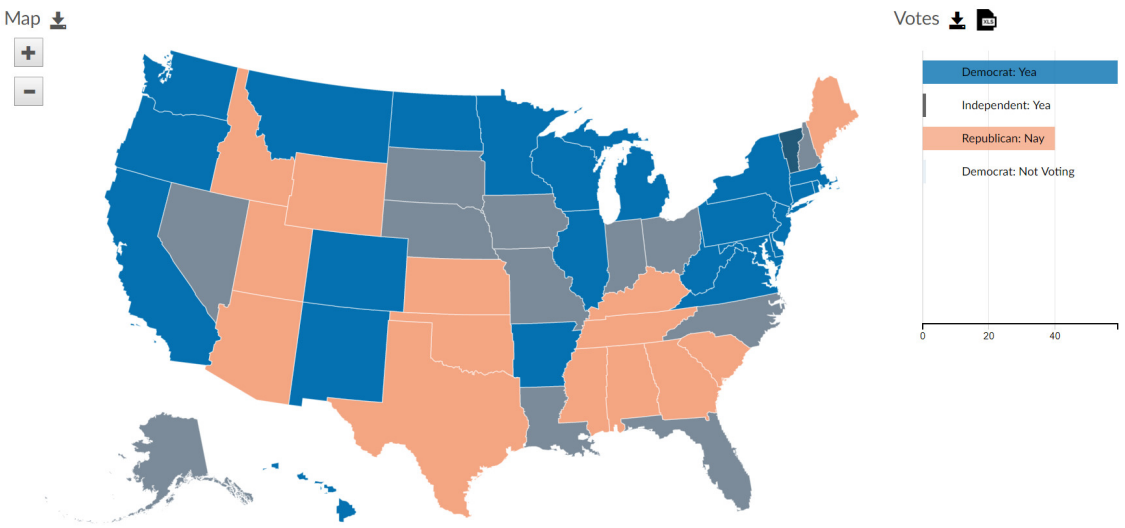

Figure 2: Thematic map of a roll call vote on Voteview.com

on a particular subject (for example, all votes concerning American entry into World War II) can select search results to "stash" for later downloading or recall. Individual votes and collections of votes can be exported to common formats, including Microsoft Excel and JSON. Specific collections of votes also can be saved for later retrieval on the site. By default, the search view displays the latest vote results, updated within minutes of their publication by Congress, including summary information about the vote result and margin. In addition to vote-searching capabilities, Voteview.com understands many plain English queries about the composition of Congress: A search for "speakers of the house" or "current senators from Virginia" will return results as expected.

3.2 Visualizing roll calls

Teachers, journalists, academics and end-users interested in visualizing the breakdown of a particular vote can click through to a page offering interactive visualization. A full map

of the United States appears with each congressional district (or state, for Senate votes), color-coded according to vote choices and party memberships of the members representing each state or district. Users can mouse over a district or state to see inline views of how that district or state's members voted. An example of one such map is shown in Figure 2. Pan and zoom functionality allows users to focus on particular areas of the map. Filtered maps can be exported to image files that can be used as pedagogical material. Users can flexibly select a subset of legislators for further examination by clicking districts or states on the map, clicking parties or vote status in a histogram chart right of the map, or based on DW-NOMINATE scores in the familiar cutting line and vote scaling plot below. An example of this ideological space plot for a roll call vote is shown in Figure 3. The triangles represent members' locations in the ideological space. Point-up triangles are members that voted Yea and point-down triangles are members that voted Nay. The cutting line appears in black and divides those predicted to vote for and against the motion. Members' DW-NOMINATE scores are constrained to fall within one unit of the origin. Thus, the feasible set of possible locations falls within a circle. We rescale the units of the second dimension by its weight parameter ($w < 1$) so that the distances between members along the first and second dimension are comparable in terms of their effect on voting. This turns the set of feasible locations into the ellipse that is shown in the plot.

The vote scaling plot contains a heat map, making it obvious to users which votes have been classified with confidence and which are more uncertain. Summary statistics about the vote, including the percentage of votes correctly classified by the NOMINATE model, are listed on the plot. A filter bar floating on the bottom of the page informs the user which filters they have selected.

A strength of the vote visualization tool is that it allows users to filter within votes interactively, and selections or filters made in one tool carry over to others on the page. By clicking and dragging on the ideological map, users can create a filter box that selects

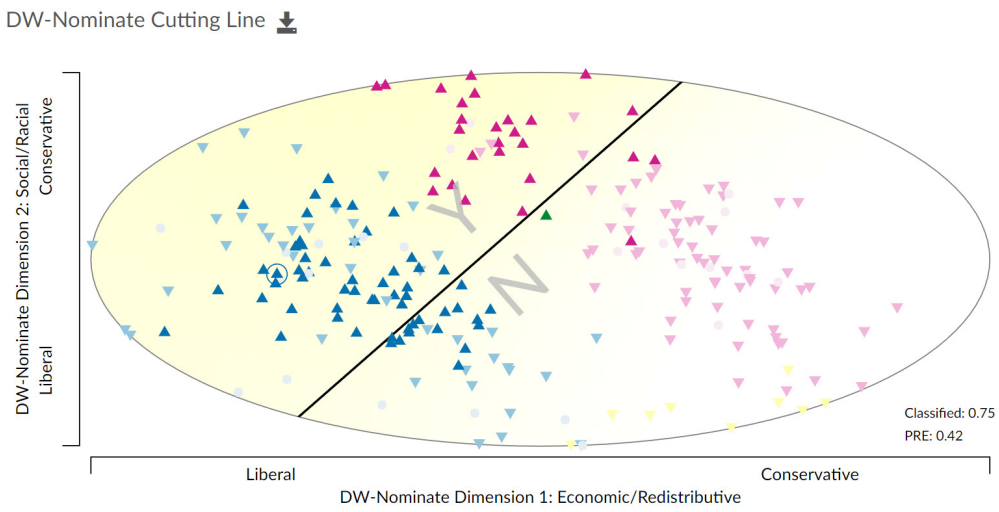

Figure 3: Example of an ideological mapping of a roll call vote on Voteview.com

and filters members from within the created rectangle. This filter not only highlights the selected members on the ideological map, but highlights those members in the geographic map display, the vote summary histogram and the list of members at the bottom of the page. By then dragging the filter box around the ideological map, users can quickly explore who the members in various regions of the plot are, where they are from, and how the selected members voted in aggregate.

A text list of members who voted can be sorted by party, state, vote choice, ideology, or vote probability. Vote probabilities allow users to identify members who voted in a way that was unlikely based on the cut-line DW-NOMINATE has fit for the vote and the legislator's ideology. The vote scaling plot quickly can visualize whether a vote was ideological along the familiar liberal-conservative first dimension of NOMINATE, or divided along the second NOMINATE dimension (as many civil rights votes were, historically), or non-ideological along both dimensions.

3.3 Visualizing congresses

Scholars and journalists who are interested in using congressional ideology data for visualization of contemporary political controversies are major consumers of these data. To that end, a list of all Senators or Representatives by Congress can easily be accessed from the main navigation bar. Users can choose how to present or report the list: they can select regions of the NOMINATE ideological plot of representatives (for example, identifying "conservative Democrats" or "the most liberal representative"), or sort a graphical list of members by party, state, ideology, or seniority. As noted above, member pages include photographic or portrait depictions for most members of Congress historically (and all representatives since 1950), including many for whom no official congressional photographs or portraits are available. In addition to the graphical list, a text-only mode displays the traditional NOMINATE score tables to which users are accustomed from Poole's Voteview.com. Each congressional view also offers a "balance of power" histogram, allowing users to see at a glance which party had control of the chamber.

3.4 Visualizing members and parties

When a user clicks on a specific member of Congress, either through the search, vote metadata, or congressional listings, they are brought to a biographical page for that member. Along with the member's congressional biography and photo, users can visualize the member's ideological score in a histogram of their contemporaries for any Congress in which the member served. Data about the member's "party loyalty" and attendance records also are provided. Below the member's summary data, a list of recent votes by the member is shown. Users can quickly identify votes that are considered unlikely by NOMINATE, understand whether the member voted with their party's majority on a given vote, and search within the member's votes to identify specific roll calls of interest. Current members also have links

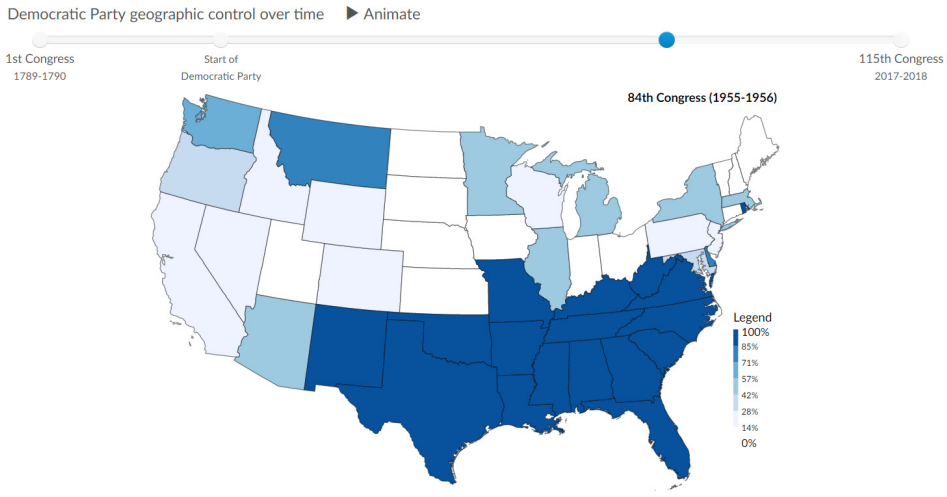

Figure 4: Thematic map of Democratic Party strength by state over time on Voteview.com

to official websites and social media accounts.

In addition to individual members, Voteview.com facilitates the exploration of party composition and ideology over time. The entrance to this view is a page that shows the ideological median and spread of the major parties across history. This page is a tribute to Poole and Rosenthal's pioneering work on congressional polarization. Users can click on a specific party to learn more about that party's historical strength and ideological positioning. Users can animate a map of the United States across time, learning how parties geographical strongholds changed over time (see Figure 4). These animations are useful pedagogical tools for illustrating political geography, including the southern realignment. Users can filter members of parties by time period and sort them, and from here return to specific member biography pages.

3.5 Congressional history of geographies

The final main functionality in Voteview.com is to view representation of a specific geographical area over time. Users who click this page automatically can detect their geographical location or enter an address of interest. Using data about congressional district boundaries over time, a user quickly can learn her current representative and senators, historical representatives, changes in which district the address was part of, or changes in the ideology of representatives over time. This page is of use to educators who want to encourage student research on their communities' political history, observe patterns of incumbency, look at partisan ideological differences within a fixed geographic area, or observe the impact of redistricting over time. States exiting Congress during the US Civil War and changes in state boundaries also are accounted for. Although non-voting representatives for US territories are not included in our data, users who search for addresses in those areas are directed to external information about their representatives.

4 Conclusion

Our core objective in developing the new Voteview.com has been to maintain and extend the infrastructure for studying Congress that Keith Poole has for so long provided. By taking advantage of advances in information technology, platforms for scientific computing and data analysis, and interactive data visualization, we are able to supply more and more reliable data, as well as more powerful means for users at all levels of technical sophistication to access and make sense of these data and what they tell us about American politics and political history. We offer the new Voteview.com in the same spirit of public-goods provision and data openness and transparency that characterize Poole's work.

All of the computer code used to build and maintain Voteview.com's database as well as

the code that powers the website and the R-based API, are publicly available.[17] The entire database of information about Congress that we have amassed can be downloaded in its entirety from Voteview.com. We encourage interested scholars and others to get involved in the project by reporting data errors, making suggestions for enhancements and participating directly in the improvement of the project. We hope that by making the entire project open, we have also provided a template for those who might wish to build similar sites for other local, state or national legislatures.[18]

The site's modular construction makes it easy to expand its functionality by adding new visualizations and data summaries in simple Rmarkdown documents into the site. These documents, data and visualizations are updated automatically and continuously to reflect the most current data, and we encourage users and scholars to contribute new summaries and visualizations in this form to the project.

In this way, we hope that the value of Keith Poole's great work can not only be maintained, but expanded upon—and that Voteview.com can continue to be an important source of insight for scholars and students of the Congress as well as journalists and broader public for another 25 years and beyond.

References

Carrol, R., Lo, J., Poole, K. T. and Rosenthal, H. (2009), 'Measuring bias and uncertatiny in dw-nominate using the parameteric bootstrap', *Political Analysis* **17**(3), 261–275.

Clausen, A. (1973), *How Congressmen Decide: A Policy Focus*, St. Martin's.

[17] See https://www.github.com/Voteview.

[18] The entire project is offered under the MIT public license, allowing for derivative works and extensions by interested users and scholars.

Congress, U. S. (2016*a*), 'Legislative subject terms'. Accessed: 2017-01-22.
URL: *http://bioguide.congress.gov/biosearch/biosearch.asp*

Congress, U. S. (2016*b*), 'Policy areas – field values'. Accessed: 2017-01-22.
URL: *https://www.congress.gov/help/field-values/policy-area*

Lewis, J. B., DeVine, B. and Pritcher, Lincoln with Martis, K. C. (2013), United states congressional district shapefiles. Data files.
URL: *http://cdmaps.polisci.ucla.edu/*

Lewis, J., Boche, A., Rudkin, A. and Sonnet, L. (2017), *rvoteview: Voteview Data in R.* R package version 0.1 — For new features, see the 'Changelog' file (in the package source).
URL: *https://github.com/voteview/Rvoteview*

Martis, K. C. (1989), *The Historical Atlas of Political Parties in the United States Congress, 1789–1989*, Macmillan Publishing Company.

McCarty, N. M. and Poole, K. T. (1995), 'Veto power and legislation: An empirical analysis of executive and legislative bargaining from 1961 to 1986', *Journal of Law, Economics, & Organization* pp. 282–312.

McCarty, N. M., Poole, K. T. and Rosenthal, H. L. (1997), *Income Redistribution and the Realignment of American Politics*, AEI Studies on Understanding Economic Inequality, AEI Press.

McCarty, N., Poole, K. T. and Rosenthal, H. L. (2009), 'Does gerrymandering cause polarization?', *American Journal of Political Science* **53**(3), 666–680.

McCarty, N., Poole, K. T. and Rosenthal, H. L. (2016), *Polarized America: The dance of ideology and unequal riches*, mit Press.

Nokken, T. P. and Poole, K. T. (2004), 'Congressional party defection in american history', *Legislative Studies Quarterly* **29**(4), 545–568.
URL: *http://www.jstor.org/stable/3598591*

Peltzman, S. (1984), 'Constituent interest and congressional voting', *Journal of Law and Economics* **27**, 181–200.

Poole, K. T. (1998), 'Recovering a basic space from a set of issue scales', *American Journal of Political Science* pp. 954–993.

Poole, K. T. (2005), *Spatial models of parliamentary voting*, Cambridge University Press.

Poole, K. T. (2007), 'Changing minds? not in congress!', *Public Choice* **131**(3), 435–451.

Poole, K. T. and Rosenthal, H. L. (1985), 'A spatial model for legislative roll call analysis', *American Journal of Political Science* pp. 357–384.

Poole, K. T. and Rosenthal, H. L. (1991), 'Patterns of congressional voting. american journal of political science', *American Journal of Political Science* **35**(1), 228–278.

Poole, K. T. and Rosenthal, H. L. (2000*a*), *Congress: A political-economic history of roll call voting*, Oxford University Press on Demand.

Poole, K. T. and Rosenthal, H. L. (2000*b*), *Voteview Manual*. Carnegie Mellon University.

Poole, K. T. and Rosenthal, H. L. (2017), *Congress: A political-economic history of roll call voting*, Oxford University.

Python Core Team (2018), *Python: A dynamic, open source programming language*.
URL: *https://www.python.org/*

R Core Team (2013), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

URL: *http://www.R-project.org/*

Rstudio Inc. (2018), *R Markdown*.

URL: *https://http://rmarkdown.rstudio.com/*

The Library of Congress (2017), 'American memory: Journals of congress'. Accessed: 2017-01-22.

URL: *https://memory.loc.gov/ammem/amlaw/lwhj.html*

United States House and Senate Historical Office (2017), 'The biographical directory of the united states congress'. Accessed: 2017-01-22.

URL: *http://bioguide.congress.gov/biosearch/biosearch.asp*