# *Legislative Explorer*: Data-Driven Discovery of Lawmaking

**Nicholas Stramp,** *University of Washington*
**John Wilkerson,** *University of Washington*

**ABSTRACT** Interactive visualizations are used widely in the natural sciences to facilitate discovery in "big-data" research. This article introduces a project that applies the same data-driven discovery approach to politics. Lawmaking is a complex process that can be difficult to convey in a constructive way, whether the audience is the general public, students, or researchers. *Legislative Explorer* (available at http://www.legex.org) visualizes the progress of more than 250,000 congressional bills and resolutions introduced since 1973. Users of all abilities can engage with actual data to discover how bills become law. Filters designed by domain experts guide them toward key insights. It is our hope that *Legislative Explorer* promotes greater understanding of a much-maligned institution.

As information about the world around us becomes increasingly accessible through technology, researchers across disciplines are challenged to develop new methodologies for making sense of large and complex data sources. Data-driven visualizations comprise an especially important area of development in the natural sciences. Effective visualizations are thought to promote the discovery of empirical laws by facilitating "detection of constancies and trends in data" (Langley 1981).

We propose that data-driven visualizations can be leveraged similarly to promote public understanding of political processes. They can enable a wide range of users to delve into large volumes of data without the usual prerequisite of advanced methodological training. Well-designed visualizations can lower the bar for discovery—not only for researchers but also for students and the general public—and promote greater interest and appreciation of the world in which we live.

We begin by briefly describing the comparative advantages of visual approaches to discovery, especially for large, complex datasets. Next, we describe the particular challenge facing instructors who teach legislative process—a process that includes hundreds of actors, scores of institutions, and thousands of bills and resolutions that vary in importance and topic. We then introduce our effort to make this process more accessible through visualization—that is, *Legislative Explorer*—and describe its potential for advancing instruction as well as research in political science.

**Nicholas Stramp** *is a graduate fellow at the Center for American Politics and Public Policy at the University of Washington. He may be reached at stramp@uw.edu.*
**John Wilkerson** *is director of the Center for American Politics and Public Policy and a professor of political science at the University of Washington. He may be reached at jwilker@uw.edu.*

## WHAT DATA-DRIVEN VISUALIZATIONS OFFER FOR THE STUDY OF POLITICS

Graphic-design theory begins with how the brain is thought to interpret signals transmitted by the optic nerve. A central finding is that the brain does not simply input what is transmitted; rather, it seeks to integrate and organize input such as dots, colors, space, and time into something more. This process of "filling in the blanks" enables people to develop an understanding of a complex world. As one of the founders of Gestalt theory stated, "There are wholes, the behavior of which is not determined by that of their individual elements, but where the part-processes are themselves determined by the intrinsic nature of the whole" (Wagemans et al. 2012). For data exploration, visualizations serve a similar purpose in helping researchers to "see" patterns and relationships that would be more difficult to detect when components are considered less holistically.

Good visualizations can be particularly effective in contexts in which the user is trying to make sense of phenomena involving many interrelated parts. Climate simulations, for example, are based on large amounts of interdependent, dynamic, and often incomplete data. A graphical visualization of the same data can help researchers to explore these interdependencies and develop hypotheses for why predictive models diverge from real-world observations (Kniss et al. 2002).

## THE HABBAL (HOW A BILL BECOMES A LAW) CHALLENGE

Data-driven visualizations are less common in the social sciences but not because there are no "big-data" challenges to be addressed. Developing an appreciation of the congressional lawmaking process is something of a big-data problem. In the 112th Congress (2011–12) alone, 553 lawmakers[1] introduced 12,299 bills and resolutions, which vary in origin, substance, and paths they take through the legislative process. In contrast, most Americans

doi:10.1017/S1049096514001644 © American Political Science Association, 2015 
https://doi.org/10.1017/S1049096514001644 Published online by Cambridge University Press

understand the legislative process in terms of a simple "one-size-fits-all" linear narrative. The conventional approach to teaching legislative process is to describe a set of procedural hurdles (i.e., HABBAL) while offering basic statistics on how unlikely it is for a bill to make it past each hurdle.[2] This procedure-based approach probably reflects the fact that actual lawmaking defies concise description. The fact that the process as described seems so straightforward also may explain why the public regards Congress with such disdain: How difficult can it be to pass a law?

## LEGISLATIVE EXPLORER

*Legislative Explorer* is an online interactive data-driven visualization of congressional lawmaking. Users can trace the progress of more than 250,000 public and private bills, simple resolutions, joint resolutions, and concurrent resolutions introduced in Congress from 1973 to the present. The data are updated nightly so that the status of current legislation also can be tracked. House and Senate members are represented by icons on the left and right borders, distinguished by gender, party, and a proxy for ideology (i.e., first-dimension DW-NOMINATE scores). Bills and resolutions are represented by color-coded particles that reflect the chamber and the party of the sponsor. The particles move across the grid based on action dates retrieved from the Library of Congress's THOMAS website.[3] For example, THOMAS might indicate that H.R. 135 was reported from the House Ways and Means Committee on June 12. *Legislative Explorer* visualizes that action by shifting the particle from the Ways and Means Committee location on the grid to the House Calendar location on that date in the timeline. As bills progress,[4] they generally move in from the left (i.e., Senate) and the right (i.e., House) borders and toward the center. Bills that pass both chambers then move upward to the president before becoming law; resolutions that do not require presidential approval move downward.

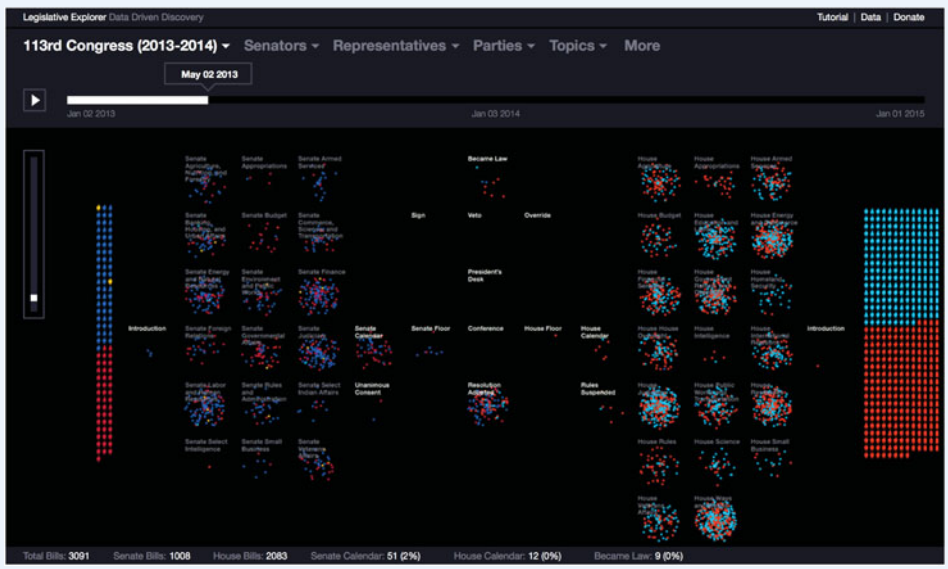
*Legislative Explorer is an online interactive data-driven visualization of congressional lawmaking. Users can trace the progress of more than 250,000 public and private bills, simple resolutions, joint resolutions, and concurrent resolutions introduced in Congress from 1973 to the present.*

*Legislative Explorer* does not incorporate all facets of the lawmaking process. However, it does make accessible substantially more information about how bills become law by capitalizing on the power of visualization to convey information simultaneously along multiple dimensions including space, time, shape, and color.

The procedural stages of the legislative process are captured on a visually appealing, two-dimensional spatial grid (figure 1).

*Figure 1*

**An example screenshot capturing activity in the 113th Congress as of May 2, 2013**




## SITE NAVIGATION

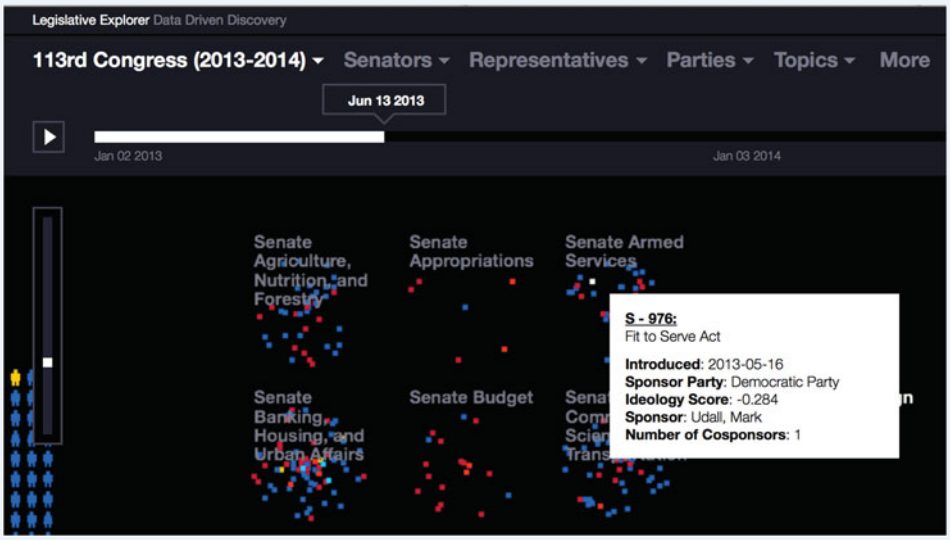
Our initial objective was to make the site engaging and easy to use. A pop-out text box directs users to a familiar "play" arrow that initiates an animation of the current Congress. Users instantly see Congress in action—a big-picture perspective on lawmaking that they cannot see anywhere else. Thousands of bill particles move across the space, corresponding to time-stamp information about their progress through various stages collected from THOMAS. For a two-year congressional session, the animation takes about 40 seconds. At the bottom of the graphic, basic bill statistics are displayed and updated based on session date (e.g., the number of bills introduced and the percentage enacted).

Users also can obtain a static view of the status of all bills on a particular day of the congressional session by pausing the animation or selecting a particular moment in time. Intuitive zoom also is included to allow users a close-up view of a particular stage of the process (e.g., only those bills that have become law). Hovering over one of the bills (i.e., a particle) displays the bill's number and title (figure 2). Clicking on the particle reveals more information—such


https://doi.org/10.1017/S1049096514001644 Published online by Cambridge University Press

*Figure 2*

**A closer view, including details about a particular bill**




as the introduction date, sponsor, sponsor's ideology, and number of cosponsors—as well as a hyperlink to full details at the Library of Congress website (i.e., http://www.congress.gov).

## FURTHER EXPLORATION THROUGH FILTERING

It is our hope that the animation of *Legislative Explorer* engages users and inspires them to delve deeper. A tutorial offers leading questions about how laws are made. Filtering options make it easy for users to investigate these and dozens of other questions. Users can select any Congress from 1973 to the present, as lawmaking patterns differ from one Congress to the next. Other top-level filters allow users to compare House and Senate bills, bills sponsored by Democrats versus Republicans, and bills addressing different major-policy topics.[5] Clicking the "More" option reveals a second level that allows for more specific filtering (i.e., a state delegation, an individual member, the type of legislation, and bills advancing beyond a specific stage in the process).

In addition, *Legislative Explorer* offers "special" filters based on domain expertise. The objective here is to direct users toward important patterns that they otherwise might overlook. For example, a substantial proportion of bills are minor in scope (i.e., naming post offices or authorizing commemorative medals). Another special filter invites users to compare the activities of committee leaders to other members; this highlights the fact that a relatively small number of committee and subcommittee chairs sponsor approximately half of all successful bills (Adler and Wilkerson 2012). Finally, a bill search tool allows users to find and follow the progress of individual bills. Compared to studying many bills at once, this option reveals the variety of paths that bills take to becoming law—or not. Approximately 15% of all bills are referred to multiple committees (Baughman 2006), whereas others bypass the committee process entirely on their way to becoming law (Sinclair 2011).

## BUILDING LEGISLATIVE EXPLORER

*Legislative Explorer* draws primarily from information made publicly available on the Library of Congress website (i.e., http://thomas.loc.gov), known colloquially as THOMAS. Details about the progress of the 250,000 bills and resolutions available on THOMAS are found in the "All Actions" section of the "All Information" page for each bill. This information is delivered as html-formatted text. *Legislative Explorer* takes advantage of substantial efforts by others to make this information available in a more accessible "machine-readable" format.[6] In particular, Joshua Tauberer—starting with his Govtrack project (https://www.govtrack.us) and later in collaboration with other open government researchers—converts and shares the "All Actions" information from THOMAS in readable formats at the /unitedstates project (http://theunitedstates.io).

We used web scrapers developed by the /unitedstates project to collect all of the actions associated with each bill from THOMAS as a single .json file and imported this information into a MySQL database. The web scrapers run nightly so that *Legislative Explorer* provides up-to-date information for the current Congress. Because the recorded actions go well beyond what was needed for *Legislative Explorer*, we then wrote custom Python scripts to extract a limited set of actions used for the visualization. Each record in this data table included the date of an "action" for a bill, the action (e.g., referred to the House Ways and Means Committee), and additional information about the bill obtained from other political science resources, such as bill topic (Adler and Wilkerson 2013; Baumgartner and Jones 2010), importance, and the attributes of thousands of members of Congress since 1973 (Carroll et al. 2011; Stewart III and Woon 2005).

The visualization then pulls information from this table of bill actions, as well as accompanying tables, on every bill and every member of Congress. When users select a specific Congress to explore, the full dataset for that Congress is downloaded to their local computer so that after an initial delay of a few seconds, they experience no additional lags. Because of *Legislative Explorer*'s unique visual and behavioral characteristics, the visualization was custom-built by the design firm Schema.[7] On the front end, the visualization was written using HTML5 and JavaScript, which in particular leverages the HTML5 Canvas element, allowing for hardware-accelerated graphics that make it possible to display tens of thousands of particles at a time. Finally, *Legislative Explorer* is hosted on an Amazon Web Services Elastic Compute Cloud (EC2) server as part of an Auto Scaling Group that ensures that the site remains responsive during heavy traffic while also keeping costs low.

## POTENTIAL RESEARCH BENEFITS: THE CASE OF MISCELLANEOUS TARIFF BILLS

As is true in the natural sciences, data-driven visualizations can provide discovery opportunities for researchers in the social sciences. We recall that a potential benefit of visualization is to

https://doi.org/10.1017/S1049096514001644 Published online by Cambridge University Press

promote the discovery of empirical laws by facilitating "detection of constancies and trends in data" (Langley 1981). *Legislative Explorer* similarly delivers existing information in a new way that suggests new research questions. For example, the visualization makes it much easier to observe variations in the timing of legislative activities; one trend that stands out is that the progress of one bill tends to be related to the progress of others. There are periods of little to no action on any bills followed by bursts of action. Some of these lulls and bursts are associated with obvious events, such as the beginning and end of a recess, but others seem to point to variations in how committees do business.

During early testing of *Legislative Explorer*, we noticed bursts in bill-sponsorship activity that seemed to indicate a problem with the data. In the second session of the 110th Congress, more than 500 bills were introduced and referred to the House Ways and Means Committee during the course of only a few days. A similar pattern occurred in the Senate of the 111th Congress. Closer examination revealed that these bills all proposed to suspend duties on imported materials and products. The streams of activity occurred in response to deadlines announced by the House Ways and Means and Senate Finance committees for possible inclusion in an omnibus Miscellaneous Tariff Bill (MTB) (Jones 2008).

This MTB process underscores why the conventional HABBAL approach of portraying bills as discrete events can be misleading in critical ways. None of the bills made it out of committee in the 110th Congress; therefore, by the usual standard of bill success, their sponsors were not effective. However, the next 111th Congress passed an omnibus MTB law that included hundreds of individual tariff-suspension provisions. Although few of the House bills of the 110th Congress were reintroduced in the 111th, many Senate bills introduced in the 111th Congress were identical to 110th House bills. Which of these hundreds of proposals made it into the omnibus law, which did not, and why or why not? We would not have thought to ask these questions without *Legislative Explorer* because we would not have noticed the same dynamics in a table or spreadsheet.[8] Additional discoveries and intriguing patterns are presented in blog format available at http://cappp.org.

## LIMITATIONS

The tutorial also informs users about important limitations of *Legislative Explorer* and discourages certain interpretations of patterns that they may observe. Despite the richness of the visualization, important stages of the lawmaking process are not captured. For example, the 200-plus subcommittees are not represented. There is no representation of the Committee of the Whole process in the House and no information about whether bills came to the floor under special rules in the House or under cloture in the Senate. Information about a small number of individual bills also is missing (due to underlying missing data on the THOMAS website). Although some of these omissions are unavoidable, they nevertheless are worth noting.

We also discourage users from equating bills with policies. The substance of a bill can change as it moves through the process, often quite radically.[9] A bill can die whereas the proposal it includes lives on, which was probably true for many of the tariff bills previously discussed.[10] Although the number of laws passed by Congress has declined over time, their length has been getting longer as Congress passes more "omnibus" laws.

There are reasons for not equating bill success with effectiveness. Sponsorship does not always indicate authorship. For example, it is fairly common for the Senate or House to replace the entire substance of a bill late in the process, as the Senate did with the Affordable Care Act. If the House and Senate pass identical bills, only one bill can become law; the other "companion" bill dies even though its substance did not. Lawmakers who fail to sponsor a single successful bill still may exercise substantial policy influence through other means, such as committee and floor amendments.

## CONCLUSION AND FUTURE PLANS

The progress of 10 thousand bills and resolutions in a typical Congress is challenging to observe all at once. The goal of *Legislative Explorer* is to take advantage of interactive data-visualization methods used widely in other fields to promote better understanding of a key representative institution. What typically is a dry and abstract topic becomes eye-catching, fun to explore, and informative. Users can quickly absorb a large volume of information about legislation that Congress is considering today and where each piece of legislation is in the process. Additionally, it is easy to observe changes in the patterns of lawmaking in the past 30 years and to investigate specific bills, topics, or venues, as well as the sponsorship activities of their own representatives and senators. Our hope is that this new tool inspires new research questions and promotes greater public appreciation of the lawmaking process.

The current version of *Legislative Explorer* is a starting point. We plan to add new features over time. For example, visualizing how the bills change in size as they move through the process provides a new opportunity to explore omnibus lawmaking and how its prevalence has changed over time. Relating language of failed bills to language found in laws may yield new insights into coalition-building and legislative effectiveness.

*We also discourage users from equating bills with policies. The substance of a bill can change as it moves through the process, often quite radically.*

More generally, data-driven visualizations can be leveraged to promote an improved understanding of government processes writ large. The enactment of a law is only one stage of the policy-making process. The practical effect of passing a law is to revise the US Code, which is the compilation of laws organized under 51 titles that serve as a primary source of authority for agencies and the regulations that they produce. The impact of laws such as the Affordable Care Act on agency regulatory activity can be visualized by highlighting sections of the US Code that the new law revises as well as the regulations subsequently proposed by agencies that cite those sections as authority sources. At the macro level, visualizing changes in the US Code and its impact on agency activities across time would provide an unprecedented dynamic view of the evolution of federal government activities. At the micro level, such a tool could be used to track more easily the status of current regulatory activities related


https://doi.org/10.1017/S1049096514001644 Published online by Cambridge University Press

to a recently enacted law. For example, Dodd-Frank includes hundreds of sections authorizing new financial-market regulations. How many have been implemented to date? Which ones have not and why? As was true for *Legislative Explorer*, the data required to build such a tool is publicly available. Integrating these different sources is likely to prove challenging, and new visualization techniques may be required. However, the end result would be a significant advance in government transparency and public appreciation of the policy-making process.

Links to all data sources, downloads, and additional information are available via the "Data" tab at http://legex.org.


## ACKNOWLEDGMENT

This project is based partly on work supported by the National Science Foundation under Grant No. SES-1243917. ■


## NOTES

1. There are 538 seats in the US House and Senate. The discrepancy is due to replacements resulting from resignations.
2. See, for example, the Library of Congress website, which offers slightly different versions of the same narrative for different education levels. Available at http://kids.clerk.house.gov/grade-school/lesson.html?intID=17.
3. Available at http://thomas.loc.gov.
4. For simplicity of presentation, by "bill," we mean a bill or resolution unless otherwise noted. *Legislative Explorer* includes public and private bills, simple House and Senate resolutions, concurrent resolutions, and joint resolutions.
5. Categorized by Policy Agendas topic codes, by the Congressional Bills Project (www.congressionalbills.org).
6. As of July 2014, the replacement for THOMAS, Congress.gov, does not provide this action information in a consistent machine-readable format. It is hoped that the Library of Congress eventually will offer application program interfaces. More information about the campaign to convince the Library of Congress to change this is available at http://congressionaldata.org.
7. Through funding provided by the Center for American Politics and Public Policy at the University of Washington, the authors collaborated extensively with Christian Marc Schmidt and Sergei Larionov of Schema Design (http://www.schemadesign.com), who developed the user interface for *Legislative Explorer*.
8. To facilitate such follow-up investigations, the underlying data are available for download.
9. As introduced, the 900-page Patient Protection and Affordable Care Act of 2009 (i.e., 111 HR 3590) was only six pages long and titled the Service Members Home Ownership Tax Act of 2009.
10. As enacted, the USA Patriot Act Reauthorization Act of 2005 contained a number of new policies that were unrelated to terrorism, such as the Combat Methamphetamine Act.

## REFERENCES


Adler, E. Scott, and John D. Wilkerson. 2012. *Congress and the Politics of Problem Solving*. New York: Cambridge University Press.

Baughman, John R. 2006. *Common Ground: Committee Politics in the U.S. House of Representatives*. Redwood City, CA: Stanford University Press.

Baumgartner, Frank R., and Bryan D. Jones. 2010. *Agendas and Instability in American Politics*. Chicago: University of Chicago Press.

Carroll, Royce, Jeff Lewis, James Lo, Nolan McCarty, Keith Poole, and Howard Rosenthal. 2011. "DW-NOMINATE Scores with Bootstrapped Standard Errors." Available at http://voteview.com/dwnomin.htm. Accessed February 1, 2014.

Jones, Vivian C. 2008. "Tariff Modifications: Miscellaneous Tariff Bills." Washington, DC: Congressional Research Service, Library of Congress.

Kniss, Joe, Charles Hansen, Michel Grenier, and Tom Robinson. 2002. "Volume Rendering Multivariate Data to Visualize Meteorological Simulations: A Case Study." In *Proceedings of the Symposium on Data Visualisation 2002*. Geneva: Eurographics Association.

Langley, Pat. 1981. "Data-Driven Discovery of Physical Laws." *Cognitive Science* 5 (1): 31–54.

Sinclair, Barbara. 2011. *Unorthodox Lawmaking: New Legislative Processes in the U.S. Congress*. Thousand Oaks, CA: SAGE Publications.

Stewart III, Charles , and Jonathan Woon. 2005. "Congressional Committee Assignments, 103rd to 109th Congresses, 1993–2005." Available at http://web.mit.edu/17.251/www/data. Accessed February 1, 2014.

Wagemans, Johan, Jacob Feldman, Sergei Gepshtein, Ruth Kimchi, James Pomerantz, and Peter van der Helm. 2012. "A Century of Gestalt Psychology in Visual Perception: II. Conceptual and Theoretical Foundations." *Psychological Bulletin* 138 (6): 1218–1252.


https://doi.org/10.1017/S1049096514001644 Published online by Cambridge University Press