RESEARCH NOTE

Congressbr: An R Package for Analyzing Data from Brazil's Chamber of Deputies and Federal Senate

Robert Myles McDonnell[1], Guilherme Jardim Duarte[2], and Danilo Freire[3]

[1] First Data Corporation, BR
[2] JOTA, BR
[3] Brown University, US
Corresponding author: Danilo Freire (danilofreire@brown.edu)

In this research note, we introduce congressbr, an R package for retrieving data from the Brazilian houses of legislature. The package contains easy-to-use functions that allow researchers to query the Application Programming Interfaces of Brazil's Chamber of Deputies and the Federal Senate, perform cleaning data operations, and store information in a format convenient for future analyses, making a previously difficult task fast and convenient. Congressbr downloads data on legislators, submitted and ratified law proposals, Senate and Chamber commissions, and other information of interest to social scientists across various fields. We outline the main features of the package and demonstrate its use with practical examples.

Nesta nota de pesquisa, apresentamos o congressbr, um pacote para R para obtenção de dados das duas casas legislativas do Brasil. O pacote contém funções intuitivas que permitem aos pesquisadores fazer buscas no API da Câmara de Deputados e do Senado Federal brasileiros, realizar tarefas de limpeza de dados, e gravar informações em um formato adequado para análises futuras, simplificando tarefas anteriormente difíceis. O congressbr baixa dados sobre legisladores, projetos de lei submetidos e aprovados, comissões na Câmara e no Senado, além de outras informações de interesse para cientistas sociais de diversas áreas. Trazemos uma breve exposição das características principais do pacote e demostramos como usá-lo por meio de exemplos práticos.

Introduction

Since the 1990s, Latin American countries have moved toward greater transparency and participation in politics (Fisher 1998; Hagopian and Mainwaring 2005; Munck 2004). As these states have become more democratic, they have devoted more attention to one of their citizens' most urgent demands: the oversight of administrative and legislative activity (Angélico 2012; Berliner and Erlich 2015; Mendez 2015; Michener 2010). However, citizens can only assess the quality of state governance effectively if they are provided with credible and comprehensive information. Hidden knowledge and hidden actions undermine the principal's ability to monitor the agent's behavior, and this problem is particularly acute in the public sphere (Downs and Rocke 1994; Miller 2005; Moe 1984; Niskanen 1971). Information asymmetries between policymakers and voters not only tilt electoral results in favor of incumbents but also lead to suboptimal outcomes in the provision of public goods. For instance, political actors can influence macroeconomic business cycles in ways unbeknownst to voters to increase their chances in future elections (Nordhaus 1975). Representatives may provide rent-seeking opportunities to special groups by introducing regulations that limit competition yet draw little attention from the general public (Tullock 1967). Similarly, civil servants can impose significant welfare losses to taxpayers by maximizing their agencies' budgets, as they have expert knowledge of the state cost functions and public finances (Migué, Belanger, and Niskanen 1974; Niskanen 1971; Tullock 1965).

Brazil has implemented a number of measures designed to reduce such practices. One notable example is the creation of *orçamento participativo*, participatory budgeting, which has fostered democratic control over public spending by encouraging citizens to engage in local fiscal administration (Baiocchi, Heller, and

Silva 2008; Koonings 2004). The Brazilian government has also improved and reformed accountability institutions such as the Comptroller General (Controladoria-Geral da União, CGU) and the Accounting Tribunal (Tribunal de Contas da União, TCU), which have increased the accountability of elected officials in all levels of government (Praça and Taylor 2014; Souza 2001). A more recent development, however, is the use of modern application programming interfaces (APIs) by Brazilian public agencies and government bodies. APIs are broadly defined as software protocols that allow machines to communicate with each other,[1] and they have been largely responsible for the recent rise in digital ecosystems and the "internet of things" (*Economist* 2014; *Wired* 2013).

Although such initiatives deserve praise, the task of collecting and managing Brazilian administrative data remains beyond most users' abilities. Groups interested in public records—such as journalists, social scientists, lawyers, or members of nongovernmental organizations—often do not have the computational skills required to interact with APIs. Even those who are familiar with that technology may find data cleaning a time-consuming task, and manual procedures for data preparation are generally burdensome and error prone (Sandve et al. 2013). Consequently, many of the benefits of providing public information to Brazilian citizens may be lost if users do not have access to the data in a timely and convenient manner. Indeed, in this article we show how researchers, in a few minutes, can replicate a study that took the authors many months of arduous data collecting.

In this research note, we present congressbr, a package for the R statistical programming language (R Core Team 2015) that enables users to download data from the APIs of the Brazilian Federal Senate and Chamber of Deputies.[2] With congressbr, we aim to fill some of this software gap in the social sciences and to lessen the workload normally necessary to collect such data. Although the same methods could be implemented in other languages, such as Python, Stata or C, we chose R because of its popularity in political science and its status as the de facto language of data analysis. There are currently over 12,500 user-contributed packages available through the R network,[3] and methods commonly used in political science have been available in R for many years (e.g., Poole et al. 2008; Stuart et al. 2011; Zeileis, Kleiber, and Jackman 2008). It is also free software, has comprehensive documentation, and easily facilitates replication (Baumer and Udwin 2015; Tippmann 2015). For newcomers to R, we have provided the code necessary to replicate the analyses herein in the online appendix.

Our package is part of a larger movement that is bringing the data of Brazilian public institutions to citizens. For example, electionsBR (Meireles, Silva, and Costa 2016) makes Tribunal Superior Eleitoral data available for users of R; GetHFData (Perlin and Ramos 2016) downloads and prepares financial data from the São Paulo stock exchange; while Julio Trecenti has a number of R packages that interact with Brazilian government APIs, such as sabesp, which downloads and plots data from the São Paulo Water Management Company (Trecenti 2015), cnpjReceita, which retrieves information from the Brazilian Internal Revenue Service (Trecenti 2016), and spgtfs/sptrans, two packages that collect data from the São Paulo City Bus Management Service (Trecenti 2017). With specific regard to political science, there are various data sets that have been introduced and offered to the research community; good examples are Alvarez et al. (1996), Linzer and Staton (2015), and Lindberg et al. (2014). The main contribution of this work is to present an easily understood framework for downloading Brazilian legislative data directly into R. This same framework can also be useful as a guide to other researchers wishing to disseminate data from other countries in a similar manner, using publicly available data from APIs such as those utilized with congressbr, which can help foster replication and reproducibility in comparative politics.

While the use of congressbr does require some basic knowledge of R, its functions are, we hope, simple and intuitive for researchers of all levels of programming experience. Moreover, the returned data are in a "tidy" format (Wickham 2014); that is, all data are organized with variables as columns and cases as rows, no encoding incompatibilities, resulting in a final data set that is as "humanly readable" as possible. This means that users can easily export the results and analyze them within R itself or with other software and spreadsheet applications.

---

[1] *Merriam-Webster*, s.v. "application programming interface," accessed May 14, 2017, https://www.merriam-webster.com/dictionary/applicationprogramminginterface; *PC Mag Encyclopedia*, s.v. "API," accessed May 14, 2017, http://www.pcmag.com/encyclopedia/term/37856/api.

[2] The current stable version is available on CRAN at https://cran.r-project.org/web/packages/congressbr. You can install the latest build of `congressbr` from its GitHub code page at https://github.com/RobertMyles/congressbr. Comments and suggestions are most welcome and can be submitted either by contacting the authors or by opening an issue at https://github.com/RobertMyles/congressbr/issues.

[3] See https://cran.r-project.org/web/packages/. R itself may also be downloaded and installed by following the instructions on its homepage: https://cran.r-project.org/.

Exploring the Brazilian Houses of Legislature

Congressbr has a series of functions that search for the details of votes, individual legislators and commissions from the websites of the Brazilian Congress.[4] Our goal is to simplify the process of obtaining online information that may be used in both qualitative and quantitative analyses. To make the functions easy to memorize, we have adopted a consistent naming pattern for the package. Every function for the Senate starts with sen_ and all functions related to the Chamber of Deputies have the prefix cham_. As of version 0.1.3, there are over forty functions in congressbr, and all of them are described in the package manual. In order to make this section more concise, here we present the functions we believe researchers will utilize more often.

The Brazilian Congress is the source of the data contained in congressbr. It is composed of two legislative houses: the Federal Senate and the Chamber of Deputies. The Senate has eighty-one members, three for each state, elected for eight-year terms by majoritarian election. The Chamber has 513 members, proportionally elected according to the size of each state, and both houses play similar roles in the legislative process. A legislator typically participates in commissions and in the plenary, proposing, discussing and voting on different issues and the national budget. Depending on the type of the issue, in order to be approved, it must be discussed successively in both houses. Collecting data on the inner workings of such legislative bodies is a complex process and justifies reliance on official data.

Both the Brazilian Chamber of Deputies and Senate maintain APIs for the dissemination of data on bills, legislators, commissions, and the budget, among other topics. For those unfamiliar with the concept, APIs are protocols to facilitate the communication between different software programs. In the present case, Brazilian legislative houses provide documentation and protocols for downloading data from a structured server. For instance, one could simply navigate to a certain URL, using the browser, to receive the data in common formats. Instead of having to download these data directly from the API, congressbr implements methods to connect directly to the APIs, collect the data and load it into the local R environment. The Chamber has two API services: one located at http://www2.camara.leg.br/transparencia/dados-abertos/dados-abertos-legislativo, and a newer API (https://dadosabertos.camara.leg.br/) released in 2017. Congressbr connects to the older API only, as the latter is still undergoing construction and has not yet reached a stable state. We do, however, intend to utilize the newer API once it is fully developed. Regarding the API of the Federal Senate, located at https://www12.senado.leg.br/dados-abertos, we have implemented all the methods available on the API. Both the APIs for the Chamber and Senate return data in XML and/or JSON format, and congressbr transforms the data requested to the data frame format, the common R data structure more suitable for handling tables. Since the package is hosted on the standard R network, CRAN (https://cran.r-project.org), to install and load it, we need only the following code:

```
install.packages('congressbr')
library('congressbr')
```

Users may familiarize themselves quickly with a certain house or legislature by using congressbr. For those new to Brazilian politics, typing statesBR into the R console will return a table of Brazilian states by name and two-letter acronym, which is useful as many requests to the API may be filtered using these same two-letter acronyms. Likewise, sen_parties() will return a data frame of the parties in the Senate, including those that have come and gone.[5] A quick look at the resulting table tells us that a total of forty-seven parties have been present at one time or another in the Federal Senate. Given the bewilderingly large number of Brazilian parties, we suggest newcomers start here. Another useful function is sen_senator_list(). By default, it returns a data frame of the senators currently serving in the Senate. This table may be requested for a certain state, for periods other than the present legislature, and by whether the senator is *titular* or a *suplente* (stand-in deputy). This table contains the gender of the senator, meaning that an analysis of the breakdown of gender in the Federal Senate is as easy as typing sen_senator_list() and then using an appropriate summary function, for example with the table() function natively available in R, which will print the following to the R console (here we also include the R commands necessary to reproduce this table):

---

[4] The package's full documentation is available at https://CRAN.R-project.org/package=congressbr.

[5] As the features of the new Chamber of Deputies API come online, we intend to add them to the package. Currently, only the API of the Federal Senate has this level of detail.

```
library(congressbr)
table(sen_senator_list()$gender)
Feminino Masculino
13          68
```

The function sen_senator_list() also returns information on senators' party membership, mandate term, and webpage and email information. For example, a journalist or an NGO member looking to quickly gather all the emails of the currently serving senators would only have to use this function to extract this data in a matter of seconds. Congressbr also contains a number of other functions that allow users to quickly familiarize themselves with the data. The function sen_bills_list(), for example, will return a table of the types of bills possible in the Senate, along with their numeric IDs and acronyms, whereas sen_bill_sponsors() produces a table of the sponsors of bills in the Federal Senate, showing that Senator Paulo Paim currently leads the way in the number of bills sponsored with a total of 310.[6] The rich detail of the API of the Federal Senate means that we have been able to make a plethora of such functions available in congressbr: information on senators' mandates may be had with sen_senator_mandates(); sen_plenary_leaderships() returns data on leadership status in the plenary; mandate data may be had with sen_senator_mandates(), and sen_budget() returns a data frame of information on the budget proposals that have passed through the House. For the full list of functions, we refer readers to the package documentation.

Another important aspect of the package is voting data. The voting behavior of legislators is an area of great interest both inside and outside of academia (e.g., Ames 1995; Attina 1990; Poole and Rosenthal 2000; Snyder and Groseclose 2000). Congressbr has two functions that describes voting patterns: cham_votes(), which returns a data frame of votes in the Chamber of Deputies; and sen_votes(), which does the same for the Senate. We should note, however, that these are not necessarily *nominal* votes, as some may be secret. In this case, the API simply records whether the legislator voted or not.

Cham_votes() returns values such as the summary of the decision (returned as the variable named decision_summary), the guidelines given by the government and the opposition to the members of their respective coalitions (GOV_orientation and Minoria_orientation) and how political parties directed their deputies to vote. A researcher could choose to analyze, for instance, how the PSDB (Partido da Social Democracia Brasileira), the PSOL (Partido Socialismo e Liberdade), and other parties oriented their members on certain votes, using variables such as PSDB_orientation and PSOL_orientation, respectively.

The function requires that users provide the type, number, and year of the bill in question. The type parameter accepts four entries: "PL" for law proposal (*projeto de lei*), "PEC" for constitutional amendments (*projeto de emenda constitucional*), "PDC" for legislative decree (*decreto legislativo*), and "PLP" for supplementary laws (*projeto de lei complementar*). Unfortunately, a particular bill can have more than one roll call, and the API does not provide a way to readily identify these repeated roll calls. We have therefore provided the variable rollcall id in the returned data frame, which is a unique identity number (ID) for each roll call. For instance, we can retrieve information about proposition 1992/2007 with the command cham_votes(type = "PL", number = 1992, year = "2007"). This will return a table of 2056 observations on 31 variables. The second column, decision_summary, contains a useful summary of the vote. We can access it with some simple R code:[7]

```
vote_table <- cham_votes(type = "PL", number = 1992, year = "2007")
vote_table$decision_summary[[1]]
```

This will print out the following: [1] "Aprovada a Subemenda Substitutiva Global oferecida pelo Relator da Comissao de Seguridade Social e Familia, ressalvados os destaques. Sim: 318; nao: 134; abstencao: 02; total: 454."; this tells us that bill 1992/2007 was approved by 318 votes to 134. The table returned also shows that the government directed deputies to vote *yea*, while several parties, from both the right and left sides of the political spectrum, ordered their members to vote *nay*.

We could also see how many legislators voted against their party on this vote. Taking the PT as an example, the following R code tells us eight PT (Partido dos Trabalhadores) deputies voted against their party (note we use the filter() function from the dplyr package [Wickham et al. 2017]). We exploit the unique roll call ID and simply filter the data:

---

[6] This number is correct as of late May 2018.

[7] For readers not familiar with R, vote_table is the object we have created in R's memory (a data frame), $ indicates we are accessing the decision_summary column, and [[1]] refers to the first line of this column.

```
install.packages('dplyr'); library(dplyr)
pt <- vote_table%>%
mutate(legislator_party = legislator_party)%>%
filter(rollcall_id == "PL-1992-2007-1",
legislator_party == "PT")
table(pt$legislator_vote)
```

The result gives us eight against the party and sixty-eight with. The function that downloads votes from the Senate, sen_votes(), works similarly. It provides variables that pertain to the time of the vote, its number, ID, year, description and the result of the roll call. Information on individual senators (their party, name, ID, gender, and the state they represent) is also returned. This function has a binary argument that if TRUE transforms the recorded (nominal) votes from *yea* to 1 and *nay* to 0, which is useful for any following quantitative analysis. Please note that dates are in yyyymmdd format, that is, to query the API for votes on September 8, 2016, we type: sen_votes("20160809"). This returns a table of 405 rows and 16 columns. Supposing we named this object sen, typing the line unique(sen$vote_round) into the R console shows us that, on this day, bills went through up to five separate rounds of voting.

The package also contains some ready data sets. The data set sen_nominal_votes(), for instance, returns a data frame of votes in the Federal Senate. Cham_nominal_votes() returns all the votes, by legislator, for the Chamber of Deputies from 1991 to 2017 with a few other columns—attributes such as party, state, and bill ID, among others. We here use this data set to calculate some simple statistics. **Figure 1** shows the number of parties in the Chamber of Deputies by year.[8] Readers may note the high level of party fragmentation in Brazil and that this fragmentation has grown worse in the last fifteen years.

Analysts can also use congressbr for collecting details on individual legislators and on commissions, both rich sources of information. Details on individual senators can be obtained with the sen_senator() function. For example, sen_senator(id = 391)[9] will return details on Senator Aécio Neves, and will show us that he was born in March 1960 in Belo Horizonte and that he joined the PMDB (Partido do Movimento Democrático Brasileiro) in 1988. Similar information for other senators is easily available, and is useful not only for qualitative analyses but also for adding description to quantitative work.
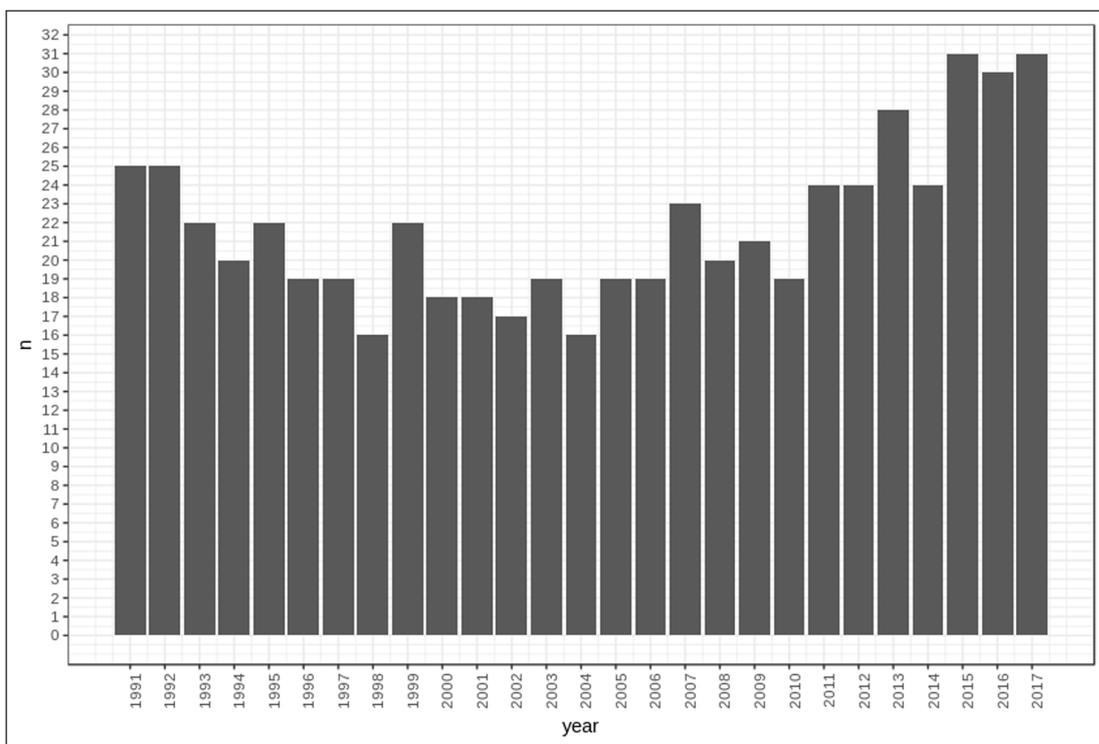

Figure 1: Number of parties by year in the Chamber of Deputies, 1991–2017.

---

[8] R code for all plots are included in the online appendix.

[9] The ID numbers of the senators, originally provided by the Senate itself, are given by the sen_senator_list() function.

The Senate API also contains information on coalitions and commissions. The function sen_commissions() will return a table of details on the commissions in the Senate. (For reasons of space, only columns 3 and 4 are shown here).

We can also see which senators serve on certain commissions. For example, the function sen_commissions_senators(code = "CCJ") will return the senators who serve on the Commission on Citizenship and Justice (Comissão de Cidadania e Justiça). Temporary coalitions may also be of interest. Sen_coalitions() will return a table of the coalitions on the Senate, including specific ID numbers for each coalition. These ID numbers can then be used to get more detailed information on that particular group. For example, 200 is the ID of the *bloco moderador*.[10] With sen_coalition_info(code = 200), we can get more complete information on the bloc, for example, on its date of formation, January 2, 2015. **Table 1** shows an example output of the function.

The APIs of the Chamber and in particular the Federal Senate are rich in details such as these, and more detailed information may be found in the package documentation.[11] We hope that congressbr can help scholars interested in qualitative studies of the Brazilian houses of legislature to more easily access this data, regardless of computer programming experience.

Producing Legislative Statistics

Congressbr also allows researchers to produce ready and direct summaries of legislative data. Much of the political science literature on the Brazilian houses of legislature has frequently utilized certain summaries of behavior (e.g., Figueiredo and Limongi 1995, 1999), and we hope that the package can help such political scientists create these summaries more easily.

For example, during the 1990s, the practice of empirically analyzing the Brazilian legislature began to grow in popularity and sophistication. Argelina Figueiredo and Fernando Limongi, pioneers in this area, studied party cohesion in the Congress and had to collect their data by hand, a time-consuming and labor-intensive process. They then used these data to analyze patterns of legislative votes, discovering that parties in the Brazilian Chamber are quite cohesive, contrary to previous findings in the literature (Figueiredo and Limongi 1995). Congressbr allows researchers to replicate these findings in a matter of minutes by collecting data from the API and employing a few R functions. For example, an important statistic, employed by the authors and often used in the political science literature to measure party cohesion, is the Rice Index (Rice 1928; Desposato 2005).[12] Calculating this measure consists of taking the absolute value of the subtraction between the number of yea votes and nay votes, and dividing this by the absolute number of votes. In mathematical terms:

$$Rice = \frac{\left| yea - nay \right|}{yea + nay}$$

Table 1: The output of function sen_commissions().

Commission name	Commission purpose
Parliamentary Inquiry Commission into "Superwages"	To investigate payments to civil servants and public officials in disagreement with the constitutional payment cap, as well as to analyze possible means for the beneficiaries to reimburse such amounts to the treasury.
Parliamentary Inquiry Commission into the BNDES	To investigate illegal procedures in BNDES-funded loans in regards to the program for the internationalization of national firms, specifically the line of credit given to firms after 1997, as well as to investigate irregularities in operations related to the public administration, mainly the line of credit denoted as BNDES/Finem Desenvolvimento Integrado dos Estados.
Parliamentary Inquiry Commission into Mistreatment	To investigate illegal procedures and crimes related to the mistreatment of children and teenagers in the country.

---

[10] The *bloco moderador*, or moderating bloc, is an informal coalition of five center-right parties: the Brazilian Labor Party (Partido Trabalhista Brasileiro, PTB), the Social Christian Party (Partido Social Cristão, PSC), the Christian Worker's Party (Partido Trabalhista Cristão, PTC), the Brazilian Republican Party (Partido Republicado Brasileiro, PRB), and the Party of the Republic (Partido da República, PR). Its main task is to coordinate voting behavior among these parties.

[11] An index of this documentation may be found by typing help(package = "congressbr") into the R console.

[12] Although there are number of methods for analyzing roll call data, such as the optimal classification (OC) method (Poole and Rosenthal 2000), Principal Component Analysis (Potthoff 2018), or variational Bayes (Imai, Lo, and Olmsted 2016), we employ the Rice index to make our results comparable with Figueiredo and Limongi (1995).

To do this in R, one can write a simple function, where votes is a numeric vector of recorded votes (traditionally coded as 1 and 0 for *yea* and *nay*, respectively):

```
rice <- function(votes){
votes <- votes[!is.na(votes)]
denominator <- length(votes)
numerator <- abs(2*sum(votes) – denominator)
numerator/denominator
}
```

This function can then be used to calculate the Rice Index for the vote data that is returned from the functions in congressbr.[13] **Figure 2** shows the historical evolution of the Rice Index for the three major parties in the Chamber of Deputies—the PMDB, the PT, and the PSDB—using votes downloaded with congressbr. The result is consistent with the view of Brazilian political history found in the literature. The PT has always been a party whose members are known for being staunch defenders of its ideology, whereas the PMDB are well-known "kingmakers" who excel at building coalitions and are not famed for ideological purity. The PSDB can be considered to be somewhere in-between.

Spatial Models of Voting Behavior

Another important application of legislative data is for use with spatial models of legislative voting (Poole 2005; Clinton, Jackman, and Rivers 2004). Analyses with spatial models usually focus on "ideal points," that is, the positions legislators take relative to one another on a scale formed by their voting records. Examples
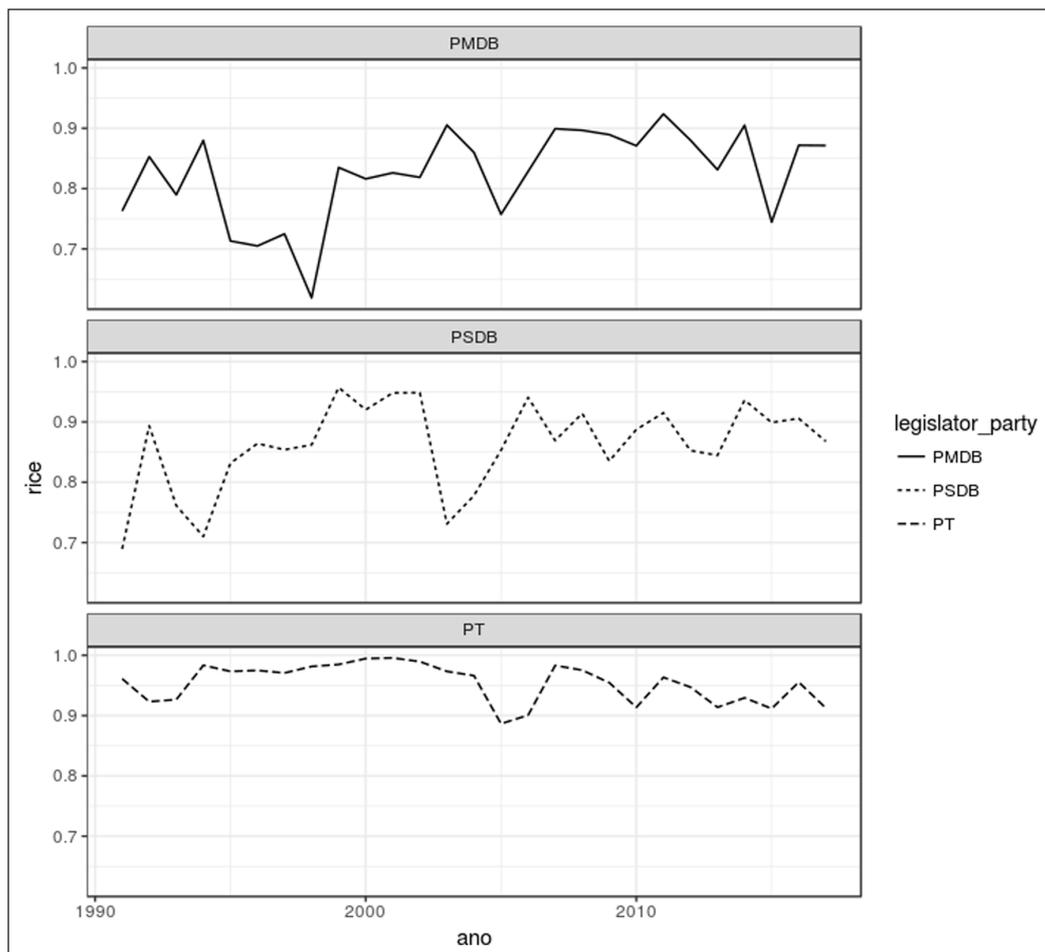

Figure 2: The Rice Index for major parties in the Chamber of Deputies, 1990–2010.

---

[13] Other simple statistics can be similarly easily constructed in the R language for use with the data provided by congressbr.

of such ideal-point analyses with Brazilian legislative data include Desposato (2006) and McDonnell (2017). These types of analyses can be easily carried out with congressbr. In order to facilitate these types of large-N nominal vote studies, we have included two data sets of nominal votes in the package, one for each legislative house, beginning in 1991 through to early 2017. The following example uses the Senate data set, which may be loaded into R with the command data("senate nominal votes"). The votes have been coded 1 for *yea* and 0 for *nay* and abstentions.

A popular way to model voting behavior utilizes Bayesian item-response theory (IRT) (Bafumi et al. 2005; Clinton, Jackman, and Rivers 2004; Martin and Quinn 2002). Bayesian IRT models estimate the probability of a yea vote ($y = 1$) as a latent regression:

$$P\left(y=1\right)=\phi\left(\beta_j \boldsymbol{x}_i - \alpha_j\right),$$

where $\boldsymbol{x}_i$ is the ideal point of senator i, and β_j and α_j are the *discrimination* and *difficulty* parameters of bill j.[14] The ideal points of the senators may be estimated in various ways in R; researchers can use probabilistic programming languages such as JAGS (Plummer 2003) and Stan (Stan Development Team 2016), or specific R functions for ideal-point analysis, such as ideal() from the pscl package (Jackman 2015) or the IRT modeling functions from MCMCpack (Martin, Quinn, and Park 2011).

For longer periods, or for when change over time is of primary interest, a dynamic ideal point model may be more suitable (Martin and Quinn 2002). Here we take a subset of the data for speed, convenience, and to avoid complications with modeling the data over time.

Ideal-point analysis requires that the data be in a particular format. We have provided a convenience function, vote_to_rollcall(), for this purpose. By default it returns data suitable for use with ideal(), but it may also be used to structure the data in a format suitable for other R packages and the programming languages mentioned above.[15] We can then use this format to run the analysis and plot the results.[16] For this example, we used the MCMCirt1d function from MCMCpack.[17] This workflow then has three simple stages: (a) load or download the data with congressbr, (b) utilize the vote_to_rollcall() function on the data, and (c) run an analysis using the IRT software the researcher has chosen.

We here present an example of this workflow. **Figure 3** shows the changes in ideal points for selected senators from an analysis of the Dilma Rousseff and Michel Temer administrations (we select only a few senators to avoid clutter in the plot). The left-hand side of the figure shows the senators' ideal points as they were when Dilma Rousseff was president. Left-wing supporters of her administration are to be found on the negative end of the scale at the bottom of the plot, while the other senators (although all except Romario were part of her coalition) can be found some distance away from their left-wing colleagues, clearly showing the disharmony in the Rousseff administration.[18]

On the right-hand side of the plot, we see evidence of the strong support some of these senators (Neves, Calheiros, Malta, Romario, and Jereissati) gave to the Temer government, whereas those who were part of the Rousseff coalition (Calheiros, for example), offered no such support to Rousseff. The two senators who opposed the impeachment process and maintained their support for Rousseff, Senators Grazziotin and Farias, display a notably different voting history. Also of note is the increased polarization seen after the impeachment, with the ideal points of each group coalescing separately, typifying closer coalitions.

Ideal points such as these can be produced quickly and easily from the voting records provided by congressbr. Data from the two houses may also be combined as in McDonnell (2017) to facilitate interesting comparisons over time and across the institutions. As may be seen from the R code in the online appendix, visualization of the results of this analysis produces the most verbose code—the actual requesting of data and its preprocessing before modeling necessitate comparatively few lines of R code.

---

[14] For more on this model, see Jackman (2001). The discrimination and difficulty parameters are analogous to the slope and intercept in regular regression models.

[15] Users may type ?vote_to_rollcall in the R console for details on how to create different data formats.

[16] The results of analyses like these can be further explored in R and users may plot the information in many ways. For instance, scholars may be interested in the senators' names, party affiliations, and state.

[17] We ran the function for 50,000 iterations, with a burn-in of 2,500 iterations. Senators Agripino (not shown) and Grazziotin were used as constraints (positive and negative, respectively). For more on constraints and identification in these models, see Rivers (2003).

[18] The use of negative numbers for left-wing legislators is for convenience to keep the ideal points on the left side of zero, for ease of interpretation. The absolute values of the ideal points do not signify anything; rather, it is the distance between legislators that is important.
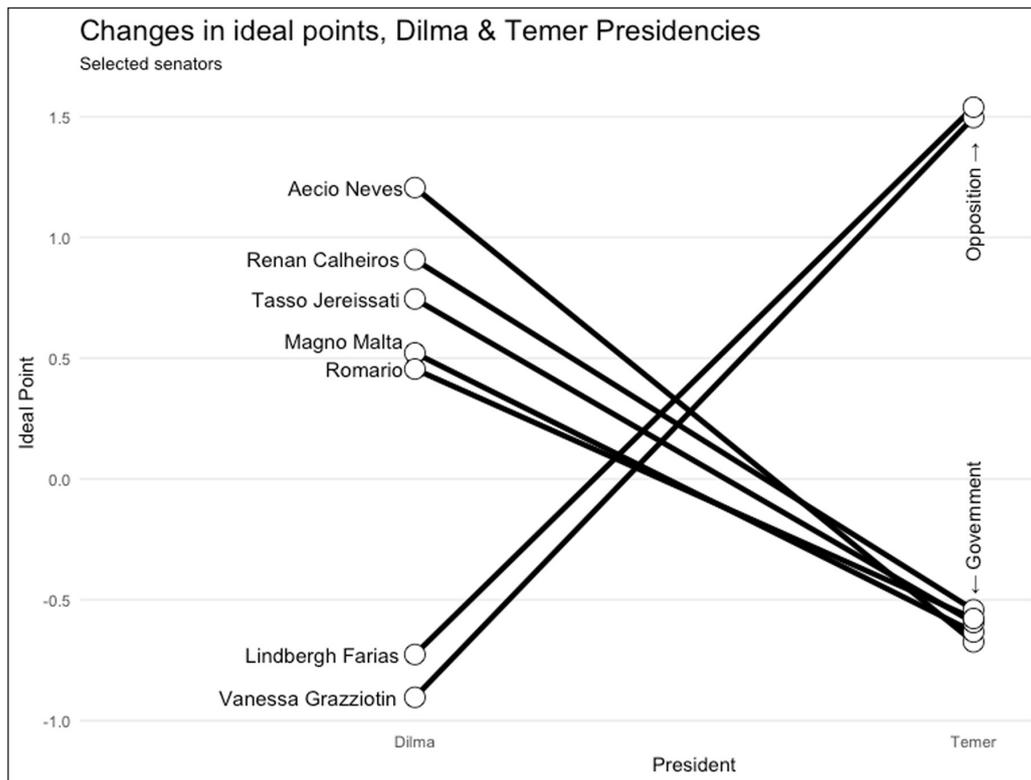
Figure 3: Differences in senators' ideal points between the Rousseff and Temer administrations.

Conclusion

We have introduced the congressbr package for R in this short research note. The purpose of making such a package is to put useful and interesting political science data in the hands of researchers. Our goal is to provide a suite of easy-to-use functions that even the novice R user can understand and use to produce analyses of Brazilian politics. This opens up the analysis of such data to more scholars than was previously possible, as studies such as those cited in the text have often been restricted to those with significant programming experience, or to those with the time and resources to collect data by hand.

In future versions of the package, we plan to include functions that download and standardize data from other levels of the Brazilian political structure, such as state and municipal legislatures. We believe that researchers will have their work greatly simplified with such an array of legislative data available with the use of only a few simple functions. We also believe congressbr can act as a useful guide for other researchers who wish to build similar packages for disseminating the data available in other countries. These types of APIs are usually quite similar in design, so that the format used by congressbr can be used for other similar APIs that make social science data available. As the source code of congressbr is freely available, researchers can copy the parts applicable to their case.

We hope users find congressbr useful for their research. Feedback and suggestions are greatly appreciated.

Additional File

The additional file for this article can be found as follows:

· **Online Appendix.** DOI: https://doi.org/10.25222/larr.447.s1

Author Information

Robert McDonnell is a data scientist at First Data Corporation. He received a PhD in International Relations from Institute of International Relations at the University of São Paulo, Brazil. His research interests include Bayesian latent variable modeling and legislative data.

Guilherme Jardim Duarte is a data editor at Jota (jota.info). He is a doctor in constitutional law at the University of São Paulo. His research interests include political methodology, judicial politics, and electoral studies.

Danilo Freire is a postdoctoral research associate in the Political Theory Project at Brown University. He holds a PhD in political economy from King's College London. His specializations are political violence and quantitative research methods.

References

Alvarez, Michael, José Antônio Cheibub, Fernando Limongi, and A. Przeworski. 1996. "Classifying Political Regimes." *Studies in Comparative International Development* 31 (2): 3–36. DOI: https://doi.org/10.1007/BF02719326

Ames, Barry. 1995. "Electoral Rules, Constituency Pressures, and Pork Barrel: Bases of Voting in the Brazilian Congress." *Journal of Politics* 57 (2): 324–343. DOI: https://doi.org/10.2307/2960309

Angélico, Fabiano. 2012. "Lei de Acesso à Informação Pública e seus possíveis desdobramentos para a accountability democrática no Brasil." Master's thesis, Fundação Getúlio Vargas. http://hdl.handle.net/10438/9905.

Attina, Fulvio. 1990. "The Voting Behaviour of the European Parliament Members and the Problem of the Europarties." *European Journal of Political Research* 18 (5): 557–579. DOI: https://doi.org/10.1111/j.1475-6765.1990.tb00248.x

Bafumi, Joseph, Andrew Gelman, David K. Park, and Noah Kaplan. 2005. "Practical Issues in Implementing and Understanding Bayesian Ideal Point Estimation." *Political Analysis* 13 (2): 171–187. DOI: https://doi.org/10.1093/pan/mpi010

Baiocchi, Gianpaolo, Patrick Heller, and Marcelo K. Silva. 2008. "Making Space for Civil Society: Institutional Reforms and Local Democracy in Brazil." *Social Forces* 86 (3): 911–936. http://www.jstor.org/stable/20430782. DOI: https://doi.org/10.1353/sof.0.0015

Baumer, Benjamin, and Dana Udwin. 2015. "R Markdown." *WIREs Comput Stat* 7: 167–177. DOI: https://doi.org/10.1002/wics.1348

Berliner, Daniel, and Aaron Erlich. 2015. "Competing for Transparency: Political Competition and Institutional Reform in Mexican States." *American Political Science Review* 109 (1): 110–128. DOI: https://doi.org/10.1017/S0003055414000616

Clinton, Joshua, Simon Jackman, and Douglas Rivers. 2004. "The Statistical Analysis of Roll Call Data." *American Political Science Review* 98 (2): 355–370. DOI: https://doi.org/10.1017/S0003055404001194

Desposato, Scott. 2005. "Correcting for Small Group Inflation of Roll-Call Cohesion Scores." *British Journal of Political Science* 35 (4): 731–744. DOI: https://doi.org/10.1017/S0007123405000372

Desposato, Scott. 2006. "The Impact of Electoral Rules on Legislative Parties: Lessons from the Brazilian Senate and Chamber of Deputies." *Journal of Politics* 68 (4): 1018–1030. DOI: https://doi.org/10.1111/j.1468-2508.2006.00484.x

Downs, George W., and David M. Rocke. 1994. "Conflict, Agency, and Gambling for Resurrection: The Principal-Agent Problem Goes to War." *American Journal of Political Science* 38 (2): 362–380. DOI: https://doi.org/10.2307/2111408

Economist. 2014. "Something to Stand On." January 16, 2014. http://www.economist.com/news/special-report/21593583-proliferating-digital-platforms-will-be-heart-tomorrows-economy-and-even.

Figueiredo, Argelina, and Fernando Limongi. 1995. "Partidos políticos na Câmara dos Deputados: 1989–1994." *Dados* 38 (3): 497–524.

Figueiredo, Argelina, and Fernando Limongi. 1999. *Executivo e legislativo na nova ordem constitucional.* Rio de Janeiro: Editora FGV.

Fisher, Julie. 1998. *Nongovernments: NGOs and the Political Development of the Third World.* West Hartford, CT: Kumarian Press.

Hagopian, Frances, and Scott Mainwaring, eds. 2005. *The Third Wave of Democratization in Latin America: Advances and Setbacks.* Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511791116

Imai, Kosuke, James Lo, and Jonathan Olmsted. 2016. "Fast Estimation of Ideal Points with Massive Data." *American Political Science Review* 110 (4): 631–656. DOI: https://doi.org/10.1017/S000305541600037X

Jackman, Simon. 2001. "Multidimensional Analysis of Roll Call Data via Bayesian Simulation: Identification, Estimation, Inference, and Model Checking." *Political Analysis* 9 (3): 227–241. http://www.jstor.org/stable/25791646. DOI: https://doi.org/10.1093/polana/9.3.227

Jackman, Simon. 2015. *pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory, Stanford University.* Department of Political Science, Stanford University, Stanford, California. R package version 1.4.9. https://github.com/atahk/pscl/.

Koonings, Kees. 2004. "Strengthening Citizenship in Brazil's Democracy: Local Participatory Governance in Porto Alegre." *Bulletin of Latin American Research* 23 (1): 79–99. http://www.jstor.org/stable/27733620. DOI: https://doi.org/10.1111/j.1470-9856.2004.00097.x

Lindberg, Staffan I., Michael Coppedge, John Gerring, and Jan Teorell. 2014. "V-Dem: A New Way to Measure Democracy." *Journal of Democracy* 25 (3): 159–169. https://www.journalofdemocracy.org/article/v-dem-new-way-measure-democracy. DOI: https://doi.org/10.1353/jod.2014.0040

Linzer, Drew A., and Jeffrey K. Staton. 2015. "A Global Measure of Judicial Independence, 1948–2012." *Journal of Law and Courts* 3 (2): 223–256. DOI: https://doi.org/10.1086/682150

Martin, Adam D., and Kevin M. Quinn. 2002. "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the US Supreme Court, 1953–1999." *Political Analysis* 10 (2): 134–153. DOI: https://doi.org/10.1093/pan/10.2.134

Martin, Adam D., Kevin M. Quinn, and Jong H. Park. 2011. "MCMCpack: Markov Chain Monte Carlo in R." *Journal of Statistical Software* 42 (9): 1–21. DOI: https://doi.org/10.18637/jss.v042.i09

McDonnell, Robert M. 2017. "Formal Comparisons of Legislative Institutions: Ideal Points from Brazilian Legislatures." *Brazilian Political Science Review* 11 (1): 1–13. DOI: https://doi.org/10.1590/1981-3821201700010007

Meireles, Fernando, Denisson Silva, and Beatriz Costa. 2016. *electionsBR: R Functions to Download and Clean Brazilian Electoral Data.* http://electionsbr.com/.

Mendez, Fabrizio S. 2015. "Right to Information Arenas: Exploring the Right to Information in Chile, New Zealand and Uruguay." PhD dissertation, London School of Economics and Political Science. http://etheses.lse.ac.uk/3361/.

Michener, Robert G. 2010. "The Surrender of Secrecy: Explaining the Emergence of Strong Access to Information Laws in Latin America." PhD dissertation, University of Texas at Austin. https://repositories.lib.utexas.edu/handle/2152/ETD-UT-2010-05-1112.

Migué, Jean-Luc, Gerard Belanger, and William A. Niskanen. 1974. "Toward a General Theory of Managerial Discretion." *Public Choice* 17 (1): 27–47. http://www.jstor.org/stable/30023151. DOI: https://doi.org/10.1007/BF01718995

Miller, Gary J. 2005. "The Political Evolution of Principal-Agent Models." *Annual Review of Political Science* 8: 203–225. DOI: https://doi.org/10.1146/annurev.polisci.8.082103.104840

Moe, Terry. 1984. "The New Economics of Organization." *American Journal of Political Science* 28 (4): 739–777. DOI: https://doi.org/10.2307/2110997

Munck, Gerardo. 2004. "Democratic Politics in Latin America: New Debates and Research Frontiers." *Annual Review of Political Science* 7: 437–462. DOI: https://doi.org/10.1146/annurev.polisci.7.012003.104725

Niskanen, William A. 1971. *Bureaucracy and Representative Government.* Chicago: Aldine Atherton.

Nordhaus, William D. 1975. "The Political Business Cycle." *Review of Economic Studies* 42 (2): 169–190. http://www.jstor.org/stable/2296528. DOI: https://doi.org/10.2307/2296528

Perlin, Marcelo, and Henrique Ramos. 2016. *GetHFData: Download and Aggregate High Frequency Trading Data from Bovespa.* https://cran.r-project.org/package=GetHFData. DOI: https://doi.org/10.2139/ssrn.2824058

Plummer, Martyn. 2003. "JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling." In *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*, March 20–22, 2003, Vienna. https://www.r-project.org/conferences/DSC-2003/Proceedings/Plummer.pdf.

Poole, Keith T. 2005. *Spatial Models of Parliamentary Voting.* New York: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511614644

Poole, Keith T., and Howard Rosenthal. 2000. *Congress: A Political-Economic History of Roll Call Voting.* New York: Oxford University Press.

Poole, Keith T., Jeffrey B. Lewis, James Lo, and Royce Carroll. 2008. *Scaling Roll Call Votes with W-NOMINATE in R.* Unpublished paper, available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstractid=1276082. DOI: https://doi.org/10.2139/ssrn.1276082

Potthoff, Richard. 2018. "Estimating Ideal Points from Roll-Call Data: Explore Principal Components Analysis, Especially for More Than One Dimension." *Social Sciences* 7 (1): 12. DOI: https://doi.org/10.3390/socsci7010012

Praça, Sergio, and Matthew Taylor. 2014. "Inching toward Accountability: The Evolution of Brazil's Anticorruption Institutions, 1985–2010." *Latin American Politics and Society* 56 (2): 27–48. DOI: https://doi.org/10.1111/j.1548-2456.2014.00230.x

R Core Team. 2015. "R: A Language and Environment for Statistical Computing." R Foundation for Statistical Computing, Vienna, Austria.

Rice, Stuart A. 1928. *Quantitative Methods in Politics*. New York: Knopf.

Rivers, Douglas. 2003. *Identification of Multidimensional Spatial Voting Models*. Unpublished paper, Stanford University.

Sandve, Geir Kjetil, Anton Nekrutenko, James Taylor, and Eivind Hovig. 2013. "Ten Simple Rules for Reproducible Computational Research." *PLoS Computational Biology* 9 (10): 1–4. DOI: https://doi.org/10.1371/journal.pcbi.1003285

Snyder, James M., Jr., and Tim Groseclose. 2000. "Estimating Party Influence in Congressional Roll-Call Voting." *American Journal of Political Science* 44 (2): 193–211. DOI: https://doi.org/10.2307/2669305

Souza, Celina. 2001. "Participatory Budgeting in Brazilian Cities: Limits and Possibilities in Building Democratic Institutions." *Environment and Urbanization* 13 (1): 159–184. DOI: https://doi.org/10.1177/095624780101300112

Stan Development Team. 2016. *Stan Modeling Language Users Guide and Reference Manual*. Version 2.14.0. http://mc-stan.org/manual.html.

Stuart, Elizabeth, Gary King, Kosuke Imai, and Daniel Ho. 2011. "MatchIt: Nonparametric Preprocessing for Parametric Causal Inference." *Journal of Statistical Software* 42 (8): 1–28. DOI: https://doi.org/10.18637/jss.v042.i08

Tippmann, Sylvia. 2015. "Programming Tools: Adventures with R." *Nature News* 517 (7532): 109. DOI: https://doi.org/10.1038/517109a

Trecenti, Julio. 2015. *Sabesp: Baixa dados da Sabesp*. https://github.com/jtrecenti/sabesp.

Trecenti, Julio. 2016. *CnpjReceita: Webscraper que realiza consulta de CNPJ na receita federal*. https://github.com/jtrecenti/cnpjReceita.

Trecenti, Julio. 2017. *Sptrans: Acesso à API olhovivo e GTFS da SPTrans*. https://github.com/jtrecenti/sptrans.

Tullock, Gordon. 1965. *The Politics of Bureaucracy*. Washington: Public Affairs Press.

Tullock, Gordon. 1967. "The Welfare Costs of Tariffs, Monopolies, and Theft." *Economic Inquiry* 5 (3): 224–232. DOI: https://doi.org/10.1111/j.1465-7295.1967.tb01923.x

Wickham, Hadley. 2014. "Tidy Data." *Journal of Statistical Software* 59 (10): 1–23. DOI: https://doi.org/10.18637/jss.v059.i10

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2017. "dplyr: A Grammar of Data Manipulation." R package version 0.7.4. https://dplyr.tidyverse.org/.

Wired. 2013. "Without API Management the Internet Is Just a Big Thing." https://www.wired.com/insights/2013/07/without-api-management-the-internet-of-things-is-just-a-big-thing/ (accessed June 16, 2017).

Zeileis, Achim, Christian Kleiber, and Simon Jackman. 2008. "Regression Models for Count Data in R." *Journal of Statistical Software* 27 (8): 1–25. DOI: https://doi.org/10.18637/jss.v027.i08