

---

# SENATEBR: COLETANDO DADOS DO SENADO FEDERAL

---

A PREPRINT

**Vinicius Santos \***

Núcleo de Dados - GLPT/SF  
Senado Federal  
Brasília, DF  
`santos.vinicius18@gmail.com`

June 14, 2024

## Abstract

Este artigo cumopre o objetivo de introduzir um novo pacote em linguagem R criado com o propósito de simplificar a interação com as APIs bem como na obtenção de dados por meio de web scraping do Senado Federal/Congresso Nacional. O objetivo central é disponibilizar à comunidade acadêmica uma ferramenta que permita o acesso eficiente a dados legislativos. A proposta é acompanhada por uma nota técnica que detalha a implementação do pacote, seu escopo e a metodologia empregada assim como estudos de caso elucidando seu potencial de uso. Assim, a iniciativa visa facilitar o processo de pesquisa e análise para estudiosos e profissionais interessados no acompanhamento das atividades legislativas do Senado Federal brasileiro.

**Keywords** programação · ciência de dados · API · senado federal

## 1 Introduction

No debate sobre governo aberto, a disponibilidade de dados é visto como um basilar para o funcionamento transparente e eficaz de qualquer democracia (Yu e Robinson, 2012; Francoli e Clarke, 2014; Sandoval-Almazan e Gil-Garcia, 2014; Abu-Shanab, 2015; De Blasio e Sorice, 2016; Kornberger et al, 2017). No caso do senado brasileiro, esses dados não apenas fornecem insights sobre o processo legislativo, mas também permitem a análise das políticas públicas a serem postas em ação, bem como do comportamento dos legisladores. Diante disso, o acesso aos dados do Senado Federal desempenha, portanto, papel fundamental na medida em que pode oferecer um conjunto de informações sobre as atividades legislativas do país (Gherghina e Katsanidou, 2013; Lupia e Elman, 2014; Gleditsch e Janz, 2016; Stockemer, Koehler e Lenz, 2018).

A motivação por trás do desenvolvimento do pacote *senatebr*, assim como iniciativas similares (Meireles, Silva e Costa, 2016; McDonnell, Duarte e Freire, 2019; Morais, 202; Saldanha, Bastos e Barcellos, 2019; Meireles e Torres, 2021), esteve ancorada na necessidade de tornar as informações da Câmara Alta acessíveis e de fácil utilização para a comunidade acadêmica, para consultores ou para qualquer pessoa interessada em compreender e analisar, no nosso caso, o cenário político brasileiro. Disso decorre que, a capacidade de acessar e analisar dados legislativos de maneira eficiente não apenas enriquece o debate público, mas também fortalece a participação cívica e a prestação de contas no sistema democrático (Gherghina e Katsanidou, 2013).

O Senado Federal brasileiro, como uma das casas do Congresso Nacional, desempenha um papel significativo na elaboração e revisão de leis e políticas públicas (Rubiatti, 2017, 2020). Portanto, seus procedimentos, debates e decisões são de central interesse para pesquisadores, acadêmicos, jornalistas, consultores e cidadãos

---

\*Doutor em Ciência Política (UFMG). Exerce o cargo de Cientista de Dados e presta Assessoria Parlamentar CTI / IA no Senado Federal - Brasília. Atuou em Pesquisas de Mercado/Notas Técnicas para Consultorias, Terceiro Setor e Governo. <http://vsantos.rbind.io/>

preocupados com questões políticas, sociais e econômicas. Ao destacar a relevância do Senado Federal como fonte de dados para análises, reconhecemos a importância de garantir que essas informações estejam disponíveis e sejam facilmente acessíveis para todos os interessados.

Diante disso, as pesquisas podem se beneficiar do acesso facilitado aos dados legislativos, eliminando, por conseguinte, a necessidade de coleta manual de informações. Ademais, o pacote fornece a possibilidade de, ao reduzir o custo de tempo do acesso à informação, os cientistas possam dedicar maior tempo e consequentemente concentrar sua atenção na análise e visualizar dados, permitindo, portanto, que os pesquisadores conduzam estudos de maior profundidade.

## 2 Escopo e propósito do pacote: acesso aos dados e reprodutibilidade

A reprodutibilidade é um princípio fundamental na pesquisa acadêmica, pois garante que os resultados obtidos possam ser verificados, validados e reproduzidos por outros pesquisadores. Isso não apenas promove transparência e confiança na pesquisa, como também permite que avanços científicos sejam construídos sobre bases sólidas e confiáveis. Nesse contexto, a reprodutibilidade desempenha um papel fundamental na validação e no avanço do conhecimento científico. Permite que outros pesquisadores verifiquem os resultados de estudos anteriores, testem hipóteses alternativas e construam sobre o trabalho existente. (Christensen e Soderberg, 2015; Da-Rt, 2012; Elman e Kapiszewski, 2014; Dunning e Rosenblatt, 2016; Freese e Peterson, 2017; Figueiredo Filho, et al, 2019).

Além disso, promove a transparência e a integridade na pesquisa, ajudando a evitar erros, vieses e má conduta científica. Portanto, garantir a reprodutibilidade dos resultados é essencial para a confiabilidade e credibilidade da pesquisa acadêmica (Gherghina e Katsanidou, 2013; Lupia e Elman, 2014; Gleditsch e Janz, 2016; Stockemer, Koehler e Lenz, 2018; Figueiredo Filho, et al, 2019).

O pacote *senatebR* foi projetado com foco na reprodutibilidade, facilitando a replicação dos resultados obtidos e a realização de análises comparativas por outros pesquisadores. Para demonstrar a reprodutibilidade, o código-fonte do pacote é estruturado de forma clara e organizada, seguindo as melhores práticas de programação e documentação<sup>2</sup>. Todas as funções e métodos são acompanhados por documentação ampla, explicando seu propósito, parâmetros e exemplos de uso. Isso permite que outros pesquisadores entendam de forma facilitada como utilizar o pacote e reproduzir os resultados (Wickham, Bryan, 2023).\*\* As principais funcionalidades do pacote incluem:

- I) Acesso à API do Senado Federal: o pacote permite a interação direta com a API do Senado Federal, facilitando a obtenção de dados atualizados sobre projetos de lei, tramitações legislativas, votações, comissões, parlamentares e outras informações.
- II) Obtenção de Dados por Web Scraping: além da API, o pacote também incorpora funcionalidades de web scraping para extrair dados diretamente do site do Senado Federal e/ou sítio eletrônico do Congresso Nacional, garantindo acesso abrangente a informações legislativas mesmo quando não estão disponíveis ou o acesso de outras informações oferecidas pela API.

Assim, foi desenvolvido para abranger uma ampla gama de funcionalidades visando atender às necessidades dos pesquisadores acadêmicos e de qualquer pessoa interessada em análises legislativas detalhadas. Este pacote oferece dados detalhados em cinco dimensões principais:

- 1) Projetos e Matérias: Este conjunto de dados permite identificar e acompanhar projetos de lei, propostas legislativas e outras matérias em tramitação no Senado Federal. Com isso, o usuário tem acesso a detalhes como título, emenda, autor, status atual e histórico de tramitação.
- 2) Informações sobre Parlamentares: nesse módulo, os utilizadores podem explorar perfis de atuais e antigos parlamentares do Senado Federal, incluindo biografias, filiações partidárias, história legislativa, entre outras informações relevantes.
- 3) Informações sobre a composição: este módulo oferece uma visão geral da composição atual do Senado Federal, incluindo a distribuição partidária dos senadores, as suas unidades federativas de origem, a duração do mandato, bem como dados demográficos e estatísticas relevantes.
- 4) Informações sobre as comissões: aqui, os usuários podem ter acesso a detalhes sobre as diferentes comissões do Senado Federal, incluindo as suas funções, membros atuais, agendas de trabalho e outras atividades relacionadas.

<sup>2</sup>Consultar site do projeto: <https://vsntos.github.io/senatebR/>

- 5) Informações sobre o Plenário: este último componente fornece informações sobre as atividades do plenário do Senado Federal, incluindo pautas de votação, transcrições de debates, vetos, medidas provisórias, decisões tomadas e outras informações relevantes.

Portanto, entre suas potencialidades de uso estão:

- 1) Análise de Dados Legislativos: uma vez que o usuário tenha familiaridade com ferramentas para limpeza, manipulação e análise de dados, o pacote permite que os usuários realizem uma ampla variedade de análises legislativas, incluindo tendências legislativas ao longo do tempo, padrões de votação, padrões de participação em comissões, entre outros.
- 2) Visualização de Dados: o pacote oferece a possibilidade de com mobilização de recursos adicionais os usuários possam visualizar os dados por meios de gráficos, mapas e outras representações visuais dos dados legislativos, tornando as análises mais acessíveis e compreensíveis.

Como dito até aqui, o *senatebR* abrange uma variedade de dados do Senado Federal, incluindo informações sobre projetos de lei, autores, tramitações, votações, comissões, parlamentares, partidos políticos, entre outros. Essa abrangência de dados permite que os usuários realizem análises multifacetadas e detalhadas do processo legislativo brasileiro facilitando pesquisas, por exemplo, de perfis parlamentares (Lemos e Ranincheski, 2008), composição e dinâmica das comissões (Nascimento, 2012; Souza e Silva, 2019; Ferreira e Rubiatti, 2022; Santos e Belém Lopes, 2022; Santos, 2024) assim como de produção legislativa (Ricci, 2008; Oliveira, 2019).

O pacote inclui uma variedade de métodos e funções para realizar diferentes tarefas relacionadas à coleta, para posterior processamento e análise de dados legislativos. Entre as principais funções temos:

#### 1. Coleta dos dados sobre os senadores por Legislatura

A função recebe como argumento o ano da legislatura de início dos dados que se busca é a legislatura de fim do intervalo desejado. O código abaixo permite, portanto, a coleta de dados das legislaturas no intervalo entre a 47 e 56.

```
library(senatebR)

df_senadores_legislatura <- obter_dados_senadores_legislatura(47, 56)

glimpse(df_senadores_legislatura)

## Rows: 929
## Columns: 13
## $ IdentificacaoParlamentar.CodigoParlamentar      <chr> "4865", "168", "5573"~
## $ IdentificacaoParlamentar.NomeParlamentar        <chr> "Abdala Karin Nabut",~
## $ IdentificacaoParlamentar.NomeCompletoParlamentar <chr> "Abdala Karin Nabut",~
## $ IdentificacaoParlamentar.SexoParlamentar         <chr> "Masculino", "Masculi~
## $ IdentificacaoParlamentar.FormaTratamento        <chr> "Senador ", "Senador ~
## $ IdentificacaoParlamentar.UrlFotoParlamentar      <chr> NA, "http://www.senad~
## $ IdentificacaoParlamentar.UrlPaginaParlamentar    <chr> NA, "http://www25.sen~
## $ IdentificacaoParlamentar.SiglaPartidoParlamentar <chr> NA, NA, "PDT", "CIDAD~
## $ IdentificacaoParlamentar.CodigoPublicoNaLegAtual <chr> NA, NA, NA, NA, "916"~
## $ IdentificacaoParlamentar.UrlPaginaParticular     <chr> NA, NA, NA, NA, "http~
## $ IdentificacaoParlamentar.EmailParlamentar        <chr> NA, NA, NA, NA, "sen.~
## $ IdentificacaoParlamentar.UfParlamentar           <chr> NA, NA, NA, NA, NA, N~
## $ Mandatos.Mandato                                <list> ["246", "DF", ["52",~
```

Com isso, uma vez seguidos esses passos, a função retorna uma base de dados de 929 observações e 12 variáveis<sup>3</sup>.

#### 2. Coleta dos dados sobre Medidas Provisórias

Para esse conjunto de dados, o pacote oferece duas opções: i) a coleta dos dados das MPs em tramitação e os dados das MPs encerradas. Sendo assim, a função a coleta feita no dia 12/06/2024 gerou um *data frame* de 24

<sup>3</sup>Disponível em:

observações e 5 (cinco) variáveis (`mpv_em_tramitacao <- coletar_medidas_provisorias_em_tramitacao()`). Já as medidas provisórias encerradas (coleta feita no mesmo dia com todas as páginas) totalizam 7268 observações e seis variáveis (`mpv_encerradas <- coletar_medidas_provisorias_encerradas(364)`).

### 3. Coleta de dados de Vetos com tramitação encerrada

Por fim, para cumprir com a tarefa inicial de apresentar exemplos que indiquem as potencialidades do pacote, é possível coletar também dados referentes aos vetos com sua tramitação encerrada (`dados_vetos <- info_vetos(pages = 20)`) somam 987 observações e 6 (seis) variáveis.

Assim, feita essa introdução, passaremos para a apresentação de um estudo de caso que demonstra o fluxo de trabalho de análise utilizando o pacote de coleta de dados do Senado Federal. Este estudo de caso oferece uma visão prática de como aplicar efetivamente o pacote em pesquisas acadêmicas e projetos de análise legislativa.

You can use directly LaTeX command or Markdown text.

LaTeX command can be used to reference other section. See Section 2. However, you can also use **bookdown** extensions mechanism for this.

#### 2.1 Headings: second level

You can use equation in blocks

$$\xi_{ij}(t) = P(x_t = i, x_{t+1} = j | y, v, w; \theta) = \frac{\alpha_i(t) a_{ij}^{w_t} \beta_j(t+1) b_j^{v_{t+1}}(y_{t+1})}{\sum_{i=1}^N \sum_{j=1}^N \alpha_i(t) a_{ij}^{w_t} \beta_j(t+1) b_j^{v_{t+1}}(y_{t+1})}$$

But also inline i.e  $z = x + y$

##### 2.1.1 Headings: third level

Another paragraph.

## 3 Examples of citations, figures, tables, references

You can insert references. Here is some text (Kour and Saabne 2014b, 2014a) and see Hadash et al. (2018).

The documentation for **natbib** may be found at

You can use custom blocks with LaTeX support from **rmarkdown** to create environment.

<http://mirrors.ctan.org/macros/latex/contrib/natbib/natnotes.pdf%7D>

Of note is the command `\citet`, which produces citations appropriate for use in inline text.

You can insert LaTeX environment directly too.

`\citet{hasselmo}` investigated\dots

produces

Hasselmo, et al. (1995) investigated...

<https://www.ctan.org/pkg/booktabs>

### 3.1 Figures

You can insert figure using LaTeX directly.

See Figure 1. Here is how you add footnotes. [<sup>^</sup>Sample of the first footnote.]

But you can also do that using R.

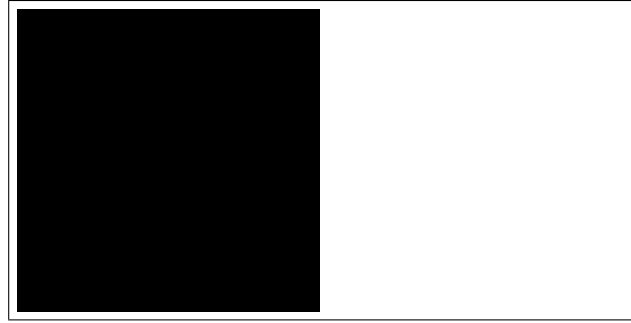


Figure 1: Sample figure caption.

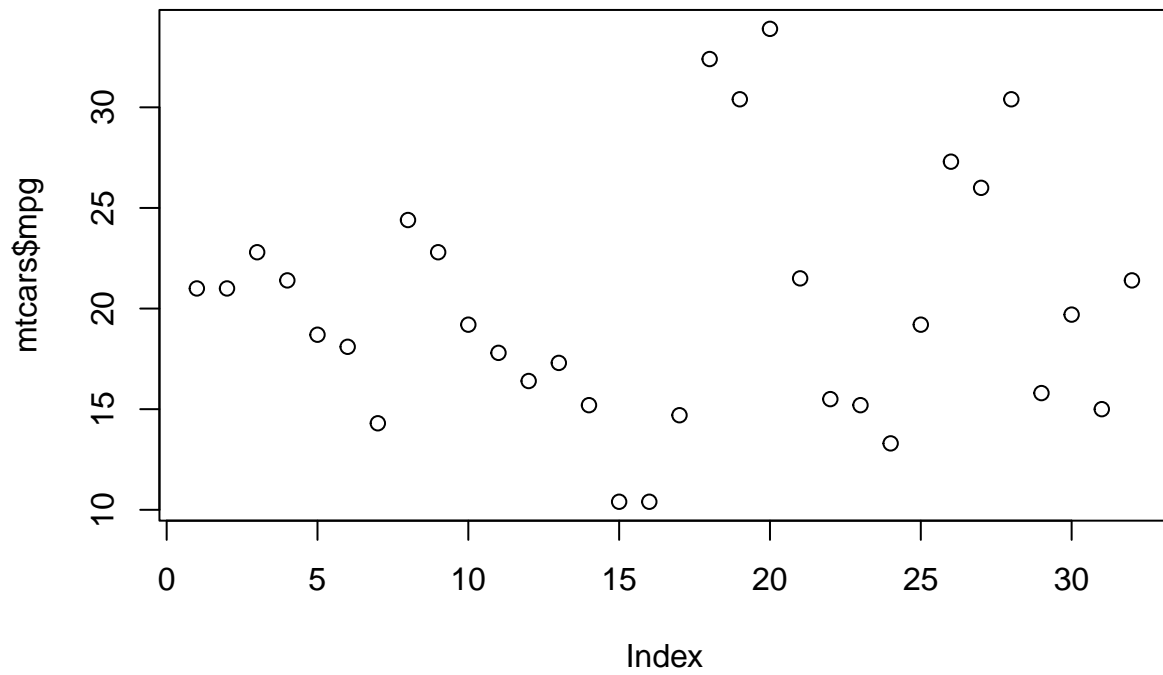


Figure 2: Another sample figure

```
plot(mtcars$mpg)
```

You can use **bookdown** to allow references for Tables and Figures.

### 3.2 Tables

Below we can see how to use tables.

See awesome Table~1 which is written directly in LaTeX in source Rmd file.

You can also use R code for that.

Table 1: Sample table title

Part		
Name	Description	Size ( $\mu\text{m}$ )
Dendrite	Input terminal	$\sim 100$
Axon	Output terminal	$\sim 10$
Soma	Cell body	up to $10^6$

```
knitr::kable(head(mtcars), caption = "Head of mtcars table")
```

Table 2: Head of mtcars table

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

### 3.3 Lists

- Item 1
- Item 2
- Item 3

Hadash, Guy, Einat Kermany, Boaz Carmeli, Ofer Lavi, George Kour, and Alon Jacovi. 2018. “Estimate and Replace: A Novel Approach to Integrating Deep Neural Networks with Existing Applications.” *arXiv Preprint arXiv:1804.09028*.

Kour, George, and Raid Saabne. 2014a. “Fast Classification of Handwritten on-Line Arabic Characters.” In *Soft Computing and Pattern Recognition (SoCPaR), 2014 6th International Conference of*, 312–18. IEEE.

———. 2014b. “Real-Time Segmentation of on-Line Handwritten Arabic Script.” In *Frontiers in Handwriting Recognition (ICFHR), 2014 14th International Conference on*, 417–22. IEEE.