

[Nota Técnica]

senatebR: Coletando Dados do Senado Federal Brasileiro*

Vinicius Santos¹

¹Pós-doutorando na Universidade do Estado do Rio de Janeiro (UERJ), Rio de Janeiro, RJ, Brasil.

✉ Email: santos.vinicius18@gmail.com  ORCID: <https://orcid.org/0000-0003-0907-7832>

DOI: <https://doi.org/10.1590/dados.2026.69.1.005>

Disponibilidade dos Dados – O autor afirma que todos os dados utilizados na pesquisa foram disponibilizados publicamente e podem ser acessados por meio da plataforma Harvard Dataverse: <https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/NIFZSU>



*O autor agradece a Thales Carvalho, Yulieth Martinez, Ranyelle Neves, Sarah Vale e aos revisores anônimos pelos seus comentários perspicazes.

Resumo

***senatebR*: Coletando Dados do Senado Federal Brasileiro**

Esta nota técnica apresenta um novo pacote em linguagem R desenvolvido com o intuito de simplificar a interação com a Interface de Programação de Aplicações (APIs) e a coleta de dados por meio de *web scraping* no contexto do Senado Federal e do Congresso Nacional. O objetivo principal é oferecer à comunidade acadêmica uma ferramenta eficiente para acesso a dados legislativos. A proposta inclui uma nota técnica que descreve a implementação do pacote, seu escopo, a metodologia utilizada, além de estudos de caso que ilustram seu potencial de uso. Dessa forma, a iniciativa visa facilitar o processo de pesquisa e análise de dados para estudiosos e profissionais interessados em monitorar as atividades legislativas do Senado Federal brasileiro.

Palavras-chave: programação, ciência de dados, API, *web scraping*, senado federal

Abstract

***senatebR*: Collecting Data from the Brazilian Federal Senate**

This technical note introduces a new R language package created with the aim of simplifying interaction with APIs and obtaining data via the web scraping site of the Federal Senate / National Congress. The main objective is to provide the academic community with a tool that allows efficient access to legislative data. The proposal is accompanied by a technical note detailing the implementation of the package, its scope and the methodology employed and case studies elucidating its potential use. Thus, the initiative aims to facilitate the research and analysis process for scholars and professionals interested in monitoring the legislative activities of the Brazilian Federal Senate.

Keywords: programming, data science, API, webscraping, federal senate

Résumé

***senatebR* : Un Outil Pour la Collecte de Données du Sénat Fédéral Brésilien**

Cette note technique présente un nouveau paquet en langage R, développé dans le but de simplifier l'interaction avec l'interface de programmation d'applications (API) ainsi que la collecte de données par *web scraping* dans le contexte du Sénat fédéral et du Congrès national. L'objectif principal est d'offrir à la communauté universitaire un outil efficace et accessible pour accéder aux données législatives. La proposition comprend une note technique qui décrit la mise en œuvre du paquet, sa portée, la méthodologie utilisée, ainsi que des études de cas qui illustrent son potentiel d'utilisation. De cette manière, ce projet vise à faciliter le processus de recherche et d'analyse des données législatives pour les chercheurs et les professionnels intéressés par le suivi des activités du Sénat fédéral brésilien.

Mots-clés : programmation, science des données, API, *web scraping*, Sénat fédéral

Resumen

***senatebR*: Recopilación de Datos del Senado Federal Brasileño**

Esta nota técnica presenta un nuevo paquete en lenguaje R desarrollado con el objetivo de simplificar la interacción con la Interfaz de Programación de Aplicaciones (API) y la recopilación de datos mediante *web scraping* en el contexto del Senado Federal y del Congreso Nacional. El objetivo principal es ofrecer a la comunidad académica una herramienta eficiente para el acceso a datos legislativos. La propuesta incluye una nota técnica que describe la implementación del paquete, su alcance, la metodología utilizada, además de estudios de caso que ilustran su potencial de uso. De esta manera, la iniciativa busca facilitar el proceso de investigación y análisis de datos para académicos y profesionales interesados en monitorear las actividades legislativas del Senado Federal brasileño.

Palabras clave: programación, ciencia de datos, API, *web scraping*, senado federal.

Introdução

No debate sobre governo aberto, a disponibilização de dados é considerada um elemento fundamental para o funcionamento transparente de qualquer democracia (Yu, Robinson, 2012; Francoli, Clarke, 2014; Sandoval-Almazan, Gil-Garcia, 2014; Abu-Shanab, 2015; De Blasio, Sorice, 2016; Kornberger *et al.*, 2017). No caso do Senado brasileiro, esses dados não apenas oferecem *insights* sobre o processo legislativo, mas também possibilitam a análise de políticas públicas assim como sobre o comportamento dos legisladores. Assim, o acesso aos dados do Senado Federal desempenha um papel essencial ao fornecer um conjunto abrangente de informações sobre as atividades legislativas do país (Gherghina, Katsanidou, 2013; Lupia, Elman, 2014; Gleditsch, Janz, 2016; Stockemer *et al.*, 2018).

A motivação para o desenvolvimento do pacote *senatebR*, assim como de iniciativas semelhantes (Meireles *et al.*, 2016; McDonnell *et al.*, 2019; Moraes, 2020; Saldanha *et al.*, 2019; Meireles, Torres, 2021), reside na necessidade de tornar as informações da Câmara Alta mais acessíveis e utilizáveis para a comunidade acadêmica, por consultores ou por qualquer pessoa interessada em compreender e analisar o cenário político brasileiro. 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 responsabilização no sistema democrático (Gherghina, Katsanidou, 2013).

Em adição, no rol de estímulos para o desenvolvimento do pacote, encontra-se o fato de que, como uma das casas do Congresso Nacional, o Senado Federal brasileiro desempenha um papel crucial na elaboração e revisão de leis e políticas públicas (Rubiatti, 2017, 2020). Assim, seus procedimentos, debates e suas decisões são de grande interesse para pesquisadores, acadêmicos, jornalistas, consultores e cidadãos interessados em questões políticas, sociais e econômicas. Por essa razão, reconhecendo a relevância do Senado, é fundamental assegurar que essas informações estejam disponíveis e sejam acessíveis aos interessados em compreender a Câmara Alta.

Disso resulta que, do ponto de vista prático, o acesso facilitado aos dados legislativos permite que as pesquisas eliminem a necessidade de coleta manual de informações, otimizando o processo de obtenção de dados. Isto posto, o pacote *senatebR* favorece, portanto, a redução do custo de tempo associado ao acesso à informação, permitindo que os cientistas se concentrem na análise e visualização de dados, o que viabiliza estudos mais aprofundados e de maior alcance.

Até o momento, mantemos o pacote no *remotes* por meio do GitHub em razão das vantagens que essa plataforma oferece na fase inicial de desenvolvimento. Essa abordagem facilita atualizações frequentes e ajustes iterativos, sem a necessidade de atender imediatamente aos requisitos do *The Comprehensive R Archive Network* - CRAN. Além disso, a hospedagem no *GitHub* permite a obtenção de *feedback* contínuo dos usuários e a colaboração da comunidade, o que é fundamental nessa etapa do projeto.

Esse processo contribui para a melhoria e estabilidade do pacote, fundamentando-se nas sugestões e experiências práticas dos usuários antes de submetê-lo ao CRAN, onde a revisão é extensa e rigorosa. Portanto, adotamos essa estratégia especialmente por acreditarmos que é benéfica para pacotes que ainda estão em fase de testes e aprimoramento. Essa abordagem possibilita a implementação ágil de novas funcionalidades, ajustes de desempenho e resolução rápida de eventuais *bugs*.

1. Escopo e Propósito do Pacote: Acesso aos Dados e Reprodutibilidade

A reprodutibilidade é um princípio fundamental na pesquisa acadêmica, pois assegura que os resultados obtidos podem ser verificados, validados e reproduzidos por outros pesquisadores. Isso não apenas promove transparência e confiança na pesquisa, mas também permite que avanços científicos se baseiem em fundamentos sólidos e confiáveis. Nesse contexto, a reprodutibilidade desempenha um papel crucial na validação e nos avanços do conhecimento científico, permitindo que outros pesquisadores verifiquem resultados anteriores, testem hipóteses alternativas e construam suas pesquisas sobre os trabalhos pré-existentes com maior clareza sobre os protocolos adotados por seus pares (Christensen, Soderberg, 2015; Elman, Kapiszewski, 2014; Dunning, Rosenblatt, 2016; Freese, Peterson, 2017; Figueiredo Filho *et al.*, 2019). Disso resulta que a reprodutibilidade, ao promover a transparência, consequentemente busca preservar a integridade na pesquisa, ajudando a prevenir erros, vieses e má conduta científica.

Assim, garantir a reprodutibilidade dos resultados é essencial para a confiabilidade e credibilidade da pesquisa acadêmica (Gherghina, Katsanidou, 2013; Lupia, Elman, 2014; Gleditsch, Janz, 2016; Stockemer *et al.*, 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 pacote¹ é estruturado seguindo as melhores práticas de programação e documentação. Dito isto, todas as funções e métodos são acompanhados por documentação, explicando seu propósito, seus parâmetros e exemplos de uso². Nosso propósito, portanto, foi permitir que outros pesquisadores entendam, de forma facilitada, como utilizar o pacote e reproduzir seus 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 não disponibilizadas por meio da API.

Assim, a ferramenta foi desenvolvida para abranger uma ampla gama de funcionalidades visando atender às necessidades dos pesquisadores acadêmicos e de qualquer pessoa interessada em análises legislativas detalhadas. O pacote oferece dados detalhados em cinco dimensões principais. A Tabela-Resumo abaixo apresenta os módulos e suas aplicações.

Tabela 1
Módulos do Pacote

Dimensão	Descrição
Projetos e Matérias	Permite identificar e acompanhar projetos de lei, propostas legislativas e outras matérias em tramitação no Senado Federal. Oferece detalhes, como título, emenda, autor, status atual e histórico de tramitação.
Informações sobre Parlamentares	Permite explorar perfis de parlamentares atuais e antigos do Senado Federal, incluindo biografias, filiações partidárias e histórico legislativo, entre outras informações relevantes.
Informações sobre a Composição	Oferece uma visão geral da composição atual do Senado Federal, incluindo a distribuição partidária dos senadores, suas unidades federativas de origem, duração do mandato, além de dados demográficos e estatísticas relevantes.

Tabela 1
Módulos do Pacote (cont.)

Dimensão	Descrição
Informações sobre as Comissões	Fornece detalhes sobre as diferentes comissões do Senado Federal, como suas funções, membros atuais, agendas de trabalho e outras atividades relacionadas.
Informações sobre o Plenário	Inclui informações sobre as atividades do plenário do Senado Federal, como pautas de votação, transcrições de debates, vetos, medidas provisórias, decisões tomadas e outras informações relevantes.

Fonte: elaborado pelo autor

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 que, 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 discutido até o momento, o *senatebR* cobre uma ampla gama de dados do Senado Federal, incluindo informações detalhadas sobre projetos de lei, autores, tramitações, votações, comissões, parlamentares, partidos políticos, entre outros elementos do processo legislativo. Essa abrangência permite que os usuários realizem análises multifacetadas e detalhadas sobre o processo legislativo brasileiro, facilitando a realização de estudos em áreas diversas. Por exemplo, o pacote possibilita análises sobre perfis parlamentares que podem incluir tanto características pessoais quanto trajetórias políticas e padrões de atuação (Lemos, Ranincheski, 2008).

Além disso, o *senatebR* oferece suporte à análise da composição e dinâmica das comissões, o que inclui investigações sobre a formação, atuação e a influência de parlamentares nas comissões permanentes e temporárias bem como o impacto de diferentes forças políticas nesses espaços (Nascimento, 2012; Souza, Silva, 2019; Ferreira, Rubiatti, 2022; Santos, Belém

Lopes, 2022; Santos, 2024). Essas informações são essenciais para estudos que buscam compreender como se articulam os interesses partidários e as alianças dentro do Senado.

Outro campo de estudo beneficiado pelo pacote é o da produção legislativa, onde se pode avaliar não apenas o volume de projetos propostos mas também a natureza e o teor das proposições, permitindo análises comparativas ao longo do tempo e entre diferentes parlamentares ou partidos (Ricci, 2008; Oliveira, 2019).

2. Implementação e Estudos de Caso

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:

I. Coleta dos dados sobre os senadores por Legislatura

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

```
> library(senatebR)

> df_senadores_legislatura <- obter_dados_senadores_legislatura(47, 56)
```

Com isso, uma vez seguidos esses passos, a função retorna uma base de dados de 929 observações e 12 variáveis³.

II. Coleta dos dados sobre Medidas Provisórias

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

III. 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)`) os quais 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.

2.1. Estudo de Caso: Pronunciamento dos Parlamentares

Antes de abordarmos o estudo de caso, é necessário revisitar o ciclo de análise de dados pelo qual um cientista atravessa. Nesse sentido, é importante considerar o modelo das etapas de um projeto típico de ciência de dados proposto por Wickham, Çetinkaya-Rundel e Grolemund (2023). Para os autores, o fluxo de trabalho inicia-se com a) importação de dados, seguido da b) limpeza e organização, passando pela c) transformação das informações, até alcançar d) visualização e modelagem, culminando na e) comunicação dos resultados obtidos (Wickham *et al.*, 2023).

Na fase de ingestão de dados, as informações são importadas para o ambiente de análise, geralmente em um formato de banco de dados, e carregadas na linguagem de programação R, que utilizamos neste estudo. A seguir, na etapa de limpeza e organização dos dados (também conhecida como “*tidying*”), organizamos as informações de forma consistente, em que cada coluna representa uma variável e cada linha uma observação (Wickham *et al.*, 2023).

Na fase de transformação dos dados (frequentemente chamada de “*wrangling*”), focamos, por exemplo, na criação de novas variáveis a partir de variáveis existentes e no cálculo de estatísticas descritivas. Assim, a etapa de visualização é reservada para revelar *insights* e levantar novas questões, enquanto a modelagem busca não apenas a extração de informações relevantes mas também, dependendo do objetivo da pesquisa, a previsão de resultados. Por fim, cabe aos cientistas a responsabilidade de comunicar, de forma clara e eficaz, as descobertas obtidas durante a análise dos dados (Wickham *et al.*, 2023).

Com base nesse quadro de referência, daremos continuidade ao nosso estudo de caso. Para a coleta de dados, retornaremos à operação de coleta dos parlamentares apresentada na seção anterior, utilizando essa base de dados como referência. Para este exercício, decidimos coletar informações dos pronunciamentos dos senadores apenas da 56ª legislatura. Veja o exemplo a seguir:

```
df_senadores_legislatura_56 <- obter_dados_senadores_  
legislatura(56, 56)
```

Com a operação anterior, obtivemos uma base de dados composta por 245 observações e 13 variáveis. Dessa base de dados, utilizaremos o vetor `df_senadores_legislatura_56$IdentificacaoParlamentar.CodigoParlamentar` para a função `extrair_pronunciamentos_multi(codigos_parlamentares, anos)`. Com essa função, conseguimos acessar todos os pronunciamentos⁴ num intervalo de tempo determinado. Veja a função abaixo:

```
dados_multi <- extrair_pronunciamentos_multi(codigos_parlamentares  
= df_senadores_legislatura_56$IdentificacaoParlamentar.  
CodigoParlamentar, anos = c(2020:2024))
```

Dito de outra forma, o que estamos buscando, portanto, por meio da função anterior, são as informações de todos os pronunciamentos de todos os senadores da 56ª legislatura entre 2020 e 2024. O código gera uma base de dados composta por 6, 528 observações e 8 variáveis⁵. Seguiremos então para a etapa de transformação dos dados. Perceba que a variável `dados_multi$Partido_UF[1]` apresenta as informações do partido e da Unidade Federativa – UF - da seguinte maneira: [1] “PDT/RO”. Para aqueles leitores já iniciados em programação/ciência de dados, podemos, portanto, separar cada uma das informações em variáveis distintas, por meio da seguinte operação:

```
dados_multi <- dados_multi %>%  
  
  separate(Partido_UF, into = c("Partido", "UF"), sep  
= "/")
```

Em outras palavras, o que estamos fazendo com essa função é separar em duas colunas (“Partido”, “UF”) a informação de Partido_e UF, tendo como referência o separador “/”. Com esse procedimento, temos, na base de dados, uma variável para Partido e outra para UF, fazendo com que a base de dados tenha agora 9 variáveis.

Os resultados da análise podem ser visualizados em gráficos, tabelas e outros tipos de representações visuais, para facilitar a interpretação e comunicação dos resultados. Isso permite que o pesquisador identifique padrões e tendências de forma mais clara e compreensível. Veja o conjunto de exemplos abaixo em que apresentamos um conjunto de estatísticas descritivas para a base de dados coletada no passo anterior (dados_multi). Abaixo apresentamos os dados sobre a frequência dos discursos por ano.

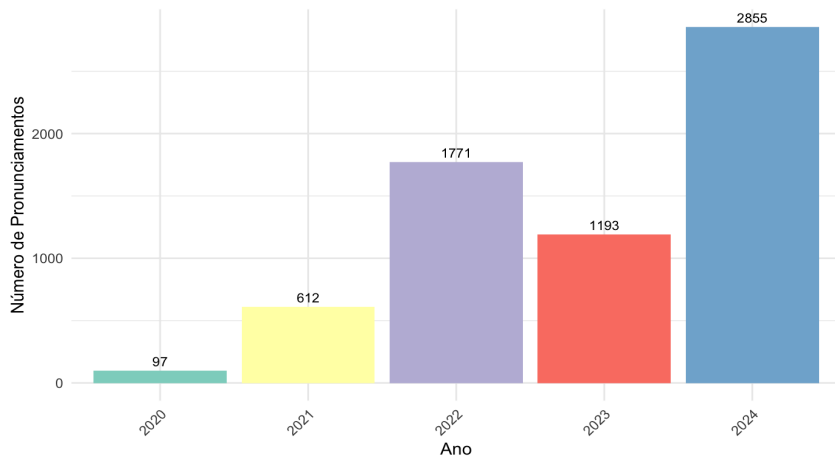
```
pronunciamentos_por_ano <- dados_multi %>%
  count(Ano)

# Plotar

ggplot(pronunciamentos_por_ano, aes(x = factor(Ano), y = n, fill
= factor(Ano))) +
  geom_bar(stat = "identity", show.legend = FALSE) +
  geom_text(aes(label = n), vjust = -0.5, size = 3) +
  scale_fill_brewer(palette = "Set3") +
  theme_minimal() +
  labs(
    title = "Distribuição de Pronunciamentos por Ano",
    x = "Ano",
    y = "Número de Pronunciamentos"
  ) +
  theme(
    plot.title = element_text(hjust = 0.5, size = 14, face =
"bold"),
    axis.text.x = element_text(angle = 45, hjust = 1)
  )
```

Gráfico 1

Distribuição de pronunciamentos por ano



Fonte: elaborado pelo autor com dados do Senado Federal

No gráfico a seguir, seguimos observando a distribuição dos discursos no tempo. Porém, nessa variação, apresentamos essa frequência em termos mensais. O que os dados sugerem é que os meses que antecedem o recesso parlamentar, quais sejam, junho e julho assim como novembro e dezembro, são os que concentram maior exposição dos parlamentares observados.

Distribuição Mensal

```
# Convertendo a coluna de Data_Pronunciamento para Date
dados_multi$Data_Pronunciamento <- as.Date(dados_multi$Data_
Pronunciamento, format = "%d/%m/%Y")
```

Extraíndo o mês

```
dados_multi$Mes <- format(dados_multi$Data_Pronunciamento, "%Y-%m")
```

Contagem de pronunciamentos por mês

```
pronunciamentos_por_mes <- dados_multi %>%
  count(Mes)
```

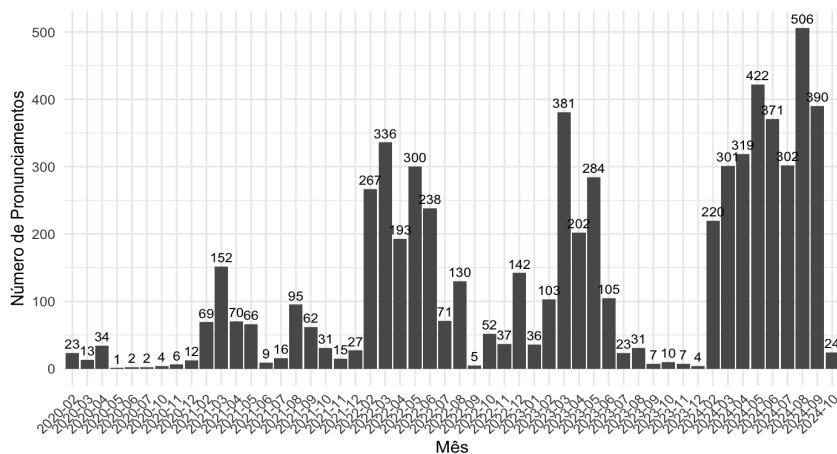
Plotar a distribuição por mês

```
ggplot(pronunciamentos_por_mes, aes(x = factor(Mes), y = n, )) +
  geom_bar(stat = "identity") +
  geom_text(aes(label = n), vjust = -0.5, size = 3) +
  labs(
    title = "Distribuição de Pronunciamentos por Mês",
    x = "Mês",
    y = "Número de Pronunciamentos"
  ) +
```

```
theme_minimal() +
theme(
  axis.text.x = element_text(angle = 45, hjust = 1), # Ajusta
a orientação do texto no eixo X
  plot.title = element_text(hjust = 0.5, size = 14, face =
"bold") # Centraliza e estiliza o título
)
```

Gráfico 2

Distribuição de pronunciamentos por mês



Fonte: elaborado pelo autor com dados do Senado Federal

Seguimos nosso exercício descritivo, observando os pronunciamentos por tipo. Os dados permitem observar, como esperado, que o uso da palavra se concentra nos discursos. No entanto, cabe observar que esse tipo de uso da palavra é seguido por pronunciamentos “Pela Ordem”, seguido dos dedicados a “Orientação à bancada”, e o uso da palavra para “Discussão” é o quarto motivo pelo qual o uso do microfone é mobilizado. Por fim, os tipos de pronunciamentos que fecham os cinco principais motivos de uso da palavra é aquele feito pelo relator, seja para proferir parecer ou não.

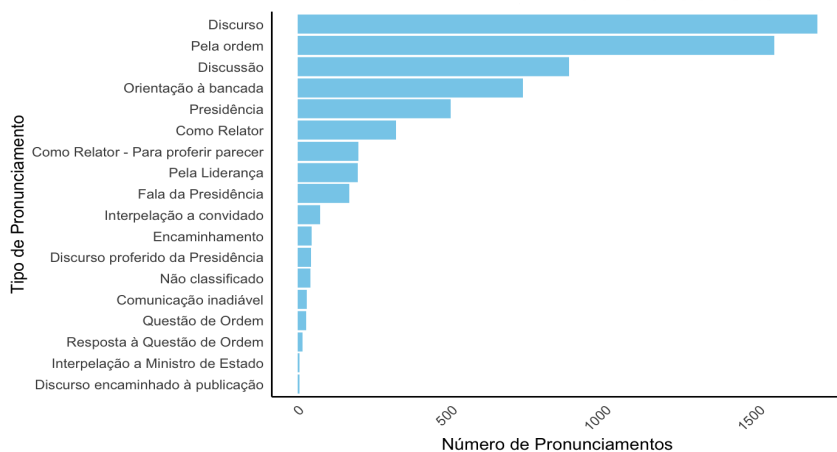
```
# Contagem de tipos de pronunciamento
tipos_pronunciamento <- dados_multi %>%
  count(Tipo_Pronunciamento)

# Ordenar os dados por ordem crescente de n
tipos_pronunciamento <- tipos_pronunciamento %>%
  arrange(n)
```

```
# Plotar a frequência com melhorias visuais
ggplot(tipos_pronunciamento, aes(x = reorder(Tipo_Pronunciamento, n),
y = n)) +
  geom_bar(stat = "identity", fill = "skyblue") +
  labs(
    title = "Frequência de Tipos de Pronunciamento",
    x = "Tipo de Pronunciamento",
    y = "Número de Pronunciamentos"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
    panel.grid.major = element_blank(), # Remove linhas de grade
    panel.grid.minor = element_blank(), # Remove linhas de grade
    axis.line = element_line(color = "black") # Adiciona linha
    de eixo x
  ) +
  coord_flip() # Inverte a coordenada para barras horizontais
```

Gráfico 3

Distribuição de frequência de tipo de pronunciamento



Fonte: elaborado pelo autor com dados do Senado Federal

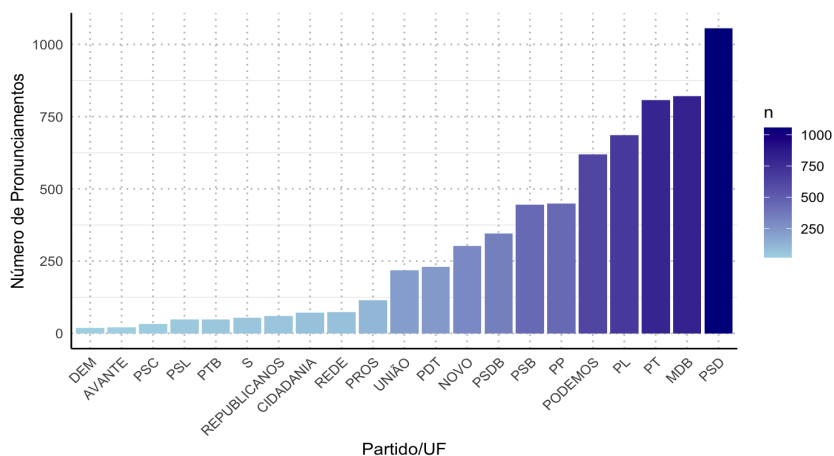
Para fechar essa apresentação de potencialidades do pacote concentrando-nos nos metadados a respeito dos pronunciamentos, podemos observar a distribuição de frequência do uso da palavra por partido. Assim, figuram no *ranking* das cinco principais agremiações partidárias: PSD, MDB, PODEMOS, PT e PL.

```
# Contagem de pronunciamentos por partido
partidos <- dados_multi %>%
  count(Partido)

# Plotar a distribuição com melhorias visuais
ggplot(partidos, aes(x = reorder(Partido, n), y = n, fill = n)) +
  geom_bar(stat = "identity") +
  scale_fill_gradient(low = "lightblue", high = "darkblue") + #
  Cores_gradientes
  labs(
    title = "Distribuição de Pronunciamentos por Partido",
    x = "Partido/UF",
    y = "Número de Pronunciamentos"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1, vjust = 1),
    # Ajuste na orientação e posicionamento do texto
    plot.title = element_text(hjust = 0.5, size = 14, face = "bold"),
    # Centraliza e estiliza o título
    panel.grid.major = element_line(color = "gray", linetype = "dotted"),
    # Linhas de grade pontilhadas cinzas
    axis.line = element_line(color = "black") # Linha de eixo preta
  )
```

Gráfico 4

Distribuição de frequência de pronunciamento por partido



Fonte: elaborado pelo autor com dados do Senado Federal

Com base nesse estudo de caso, espera-se que o pesquisador possa obter uma visão sobre o uso do pacote. Não obstante, a comunidade pode integrar as análises técnicas de Inteligência Artificial (IA) e *Machine Learning* (ML) que podem contribuir para redefinir a forma como os dados legislativos são analisados e interpretados.

Com isso, esse conjunto de técnicas pode oferecer novas angulações para pesquisa bem como novas perspectivas a respeito do desenvolvimento de uma série de potenciais aplicações que podem consumir esses dados. Os algoritmos de *machine learning* podem, portanto, ser mobilizados para analisar o comportamento dos legisladores, detectar padrões de votação e até mesmo antecipar o impacto de políticas públicas. Diante disso, usuários podem, potencialmente, explorar padrões complexos nos dados, identificar correlações e prever tendências (Batista, 2020; Rodrigues, 2022; Firebanks-Quevedo, 2022; Santos, 2024).

Conclusão

O pacote *senatebR* é uma ferramenta para o acesso a informações legislativas no contexto brasileiro. Com isso em mente, permite a simplificação de obtenção de dados, permitindo que pesquisadores, acadêmicos e profissionais realizem análises sobre diversos aspectos do processo legislativo, do perfil dos senadores, da agenda da casa, seja nas comissões seja no Plenário, por exemplo.

O acesso simplificado a esse conjunto de informações permite aos pesquisadores e consultores que mais tempo seja dedicado à análise dos dados, aumentando sua capacidade analítica e, conseqüentemente, enriquecendo as investigações legislativas, seja na identificação de tendências, padrões subjacentes aos dados seja até mesmo de *insights* significativos sobre a Câmara Alta brasileira. Isso sem mencionar o papel desempenhado na potencial democratização do acesso aos dados legislativos, com vistas a torná-los mais transparentes e compreensíveis, fazendo com que o pacote promova a transparência e a *accountability* no processo democrático.

Ainda, reconhecemos que, a despeito do pacote representar um avanço no acesso às informações do Senado Federal, ainda há espaço para melhorias. Nesta lista de aprimoramentos, figura a expansão da cobertura de dados seja do próprio Senado, seja da Câmara dos Deputados seja do Congresso Nacional. Uma vez superado esse desafio, acreditamos que

essa implementação tornaria o pacote ainda mais completo. Além disso, novos exemplos de uso são essenciais para facilitar sua adoção e utilização por uma gama ainda mais ampla de usuários.

Além das melhorias sugeridas anteriormente, cabe considerar a potencial integração de técnicas avançadas de Inteligência Artificial (IA) e *Machine Learning* (ML), na medida em que essas tecnologias possibilitem transformar a maneira como os dados legislativos são analisados e interpretados, abrindo novas perspectivas de pesquisa e desenvolvimento de aplicações. Disso decorre que, com a utilização de IA e ML, os usuários podem explorar padrões complexos nos dados legislativos, identificar correlações e prever tendências. Por exemplo, algoritmos de *machine learning* podem ser empregados para analisar o comportamento dos legisladores, identificar padrões de votação e até mesmo antecipar o impacto de determinadas políticas públicas com base em dados históricos.

Além disso, o uso de técnicas de IA e ML pode potencializar o desenvolvimento de aplicações interativas em *Shiny*, por exemplo. Sendo o *Shiny* uma ferramenta poderosa para criar *dashboards* e visualizações de dados dinâmicos, os usuários podem explorar os dados legislativos de forma mais intuitiva e interativa, permitindo uma compreensão mais profunda e uma tomada de decisão mais informada.

Sobre a pluralidade de usos, por exemplo, é possível criar *dashboards* que acompanhem o progresso de projetos de lei, monitorem a atuação de parlamentares ou até mesmo simulem cenários. Em resumo, a integração de inteligência artificial, *machine learning* e desenvolvimento de aplicações em *Shiny* representa uma oportunidade para aprimorar ainda mais o uso do pacote de coleta de dados do Senado Federal, ao aproveitar todo o potencial dessas tecnologias.

Editor responsável: Luiz Augusto Campos.

(Recebido para publicação em 08 de dezembro de 2024)

(Reapresentado em 15 de janeiro de 2025)

(Aprovado para publicação em 23 de janeiro de 2025)

Notas

1. Consultar site do projeto: <https://vsntos.github.io/senatebR/>
2. Ao adotar boas práticas de documentação em R com o uso do pacote *roxygen2*, buscamos:
i) facilitação da compreensão do código; ii) automatização da geração da documentação; iii) garantia de consistência, iv) possibilidade de geração de documentação em vários formatos; v) integração com controle de versão e vi) geração de exemplos reproduzíveis
3. Disponível em: <https://legis.senado.leg.br/dadosabertos/senador/5322?v=6>
4. O comando *extrair_pronunciamentos_multi* realiza a extração de dados de informações a respeito dos pronunciamentos parlamentares para os códigos parlamentares especificados no intervalo de anos de 2020 a 2024. Essa operação pode levar um tempo considerável em razão de vários fatores, incluindo o volume de dados a ser processado e a complexidade das chamadas de API ou consultas ao banco de dados que a função realiza para obter as informações. O tempo de resposta pode variar dependendo da disponibilidade dos serviços de dados e da carga de trabalho do sistema no momento da solicitação. Recomenda-se que os usuários atentem para esse tempo e, se necessário, considerem executar essa função em horários de menor demanda, para otimizar a eficiência.
5. Coleta de 30 de outubro de 2024.

Referências

- Abu-Shanab, Emad A. (2015), "Reengineering the open government concept: An empirical support for a proposed model." *Government Information Quarterly*, v. 32, n. 4, p. 453-463.
- Batista, Mariana. (2020), "QUAIS POLÍTICAS IMPORTAM? Usando ênfases na agenda legislativa para mensurar saliência." *Revista Brasileira de Ciências Sociais*, v. 35, p. e3510411.
- Christensen, Garret; Soderberg, Courtney. (2015), *Manual of Best Practices in Transparent Social Science Research*. Berkeley: University of California.
- Clarke, Amanda; Francoli, Mary. (2014), "What's in a Name? A Comparison of 'Open Government' Definitions across Seven Open Government Partnership Members." *Jedem-Ejournal of Democracy And Open Government*, v. 6, n. 3, p. 248-266.
- de Blasio, Emiliana; Sorice, Michele. (2016), "Open Government: a Tool for Democracy?" *Media Studies*, v. 7, n. 14.
- Dunning, Thad; Rosenblatt, Fernando. (2016), "Transparency and Reproducibility in Multi-Method Research." *Revista de Ciencia Política*, v. 36, n. 03, p. 773-783.
- Elman, Colin; Kapiszewski, Diana. (2014), "Data Access and Research Transparency in the Qualitative Tradition." *Ps: Political Science and Politics*, v. 47, n. 01, p. 43-47.
- Ferreira, Wesley Rodrigues Santos; de Castro Rubiatti, Bruno. (2022), "Caracterização da Expertise dos Senadores Membros da Comissão de Agricultura e Reforma Agrária (Cra) do Senado Federal (2005-2018)." *Novos Olhares Sociais*, v. 5, n. 2, p. 141-171.
- Figueiredo Filho, Dalson et al. (2019), "Seven Reasons Why: A User's Guide to Transparency and Reproducibility." *Brazilian Political Science Review*, v. 13, p. e0001.
- Firebanks-Quevedo, Daniel et al. (2022), "Using Machine Learning to Identify Incentives in Forestry Policy: Towards a New Paradigm in Policy Analysis." *Forest Policy and Economics*, v. 134, p. 102624.
- Freese, Jeremy; Peterson, David. (2017), "Replication in Social Science." *Annual Review of Sociology*, v. 43, n. 01, p. 147-165.
- Gherghina, Sergiu; Katsanidou, Alexia. (2013), "Data Availability in Political Science Journals." *European Political Science*, v. 12, p. 333-349.
- Gleditsch, Nils Petter; Janz, Nicole. (2016), "Replication in International Relations." *International Studies Perspectives*, v. 17, n. 04, p. 361-366.
- Kornberger, Martin et al. (2017), "When Bureaucracy Meets the Crowd: Studying 'Open Government' in the Vienna City Administration." *Organization Studies*, v. 38, n. 2, p. 179-200.
- Lemos, L. B.; Ranincheski, S. (2008), "Carreiras Políticas no Senado Brasileiro: um Estudo das Composições do Plenário e da Comissão de Justiça e Cidadania na Década de 1990," In: *Lemos, L. B. (Org.) O Senado Federal Brasileiro no Pós-Constituinte*. Brasília: Senado Federal; Unilegis.
- Lupia, Arthur; Elman, Colin. (2014), "Openness in Political Science: Data Access and Research Transparency." *Ps: Political Science & Politics*, v. 47, n. 01, p. 19-42.
- Mcdonnell, Robert Myles; Duarte, Guilherme Jardim; Freire, Danilo. (2019), "Congressbr: An R Package for Analyzing Data From Brazil's Chamber of Deputies and Federal Senate." *Latin*

- American Research Review*, v. 54, n. 4, p. 958-969. Disponível em: <https://Github.Com/DuarteGuilherme/Congressbr>
- Meireles, F.; Torres, M. V. (2021), "Siconvr: An R Package to Fetch Data from Plataforma +Brasil (Siconv)." Disponível em: <https://Github.Com/Meirelesff/Siconvr>
- Meireles, Fernando; Silva, Denisson; Costa, Beatriz. (2016), "Electionsbr: R Functions to Download and Clean Brazilian Electoral Data." Disponível em: <https://Github.Com/Silvadenisson/Electionsbr>
- Morais, João Henrique De Araújo et al. (2021), "Rtabnet: Pacote R para Extração de Indicadores de Saúde do Estado de São Paulo." *Epidemiologia e Serviços de Saúde*, v. 30, n. 1, p. E2020576. Disponível em: <https://Github.Com/Joaohmorais/Rtabnet>
- Nascimento, E. O. (2012), "O Sistema de Comissões Brasileiro: Elementos para uma Agenda de Pesquisa." *Revista de Ciência Política Brasileira*, v. 21, n. 2, p. 61-72.
- Oliveira, Danilo Amaral de. (2019), *Compreendendo e Prevendo o Processo Legislativo Via Ciência de Dados*. Tese de Doutorado, Universidade de São Paulo.
- Ricci, P. (2008), "A Produção Legislativa de Iniciativa Parlamentar no Congresso: Diferenças e Similaridades entre a Câmara dos Deputados e o Senado Federal," In: *Lemos, L. B. (Org.) O Senado Federal Brasileiro no Pós-Constituinte*. Brasília: Senado Federal; Unilegis.
- Rodrigues, Waldecy et al. (2022), "Uso de Machine Learning para a Análise de Projetos Legislativos de Desenvolvimento Regional: o Caso da Zona Franca de Manaus." *Informe Gepec*, v. 26, n. 2, p. 127-140.
- Rubiatti, Bruno de Castro. (2017), "Sistema de Resolução de Conflitos e o Papel do Senado como Câmara Revisora no Bicameralismo Brasileiro." *Revista Brasileira de Ciência Política*, n. 23, p. 35-74.
- (2020), "Para Além do Plenário: o Papel Decisório das Comissões no Senado Federal Brasileiro." *Revista de Sociologia e Política*, v. 28, n. 75, p. e005.
- Saldanha, Raphael de Freitas; Bastos, Ronaldo Rocha; Barcellos, Christovam. (2019), "Microdatasus: A Package for Downloading and Preprocessing Microdata from Brazilian Health Informatics Department (Datusus)." *Cadernos De Saúde Pública*, V. 35, P. E00032419. Disponível em: <https://Github.Com/Rfsaldanha/Microdatasus>
- Sandoval-Almazan, Rodrigo; Gil-Garcia, J. Ramon. (2014), "Towards an Evaluation Model for Open Government: A Preliminary Proposal," In: *Electronic Government: 13th Ifip Wg 8.5 International Conference, Egov 2014, Dublin, Ireland, September 1-3, 2014. Proceedings 13*. Springer Berlin Heidelberg, p. 47-58.
- Santos, Vinicius. (2025), "Inside the Fog: Assessing Foreign Policy Expertise in the Brazilian Federal Senate. *Dados*, Rio de Janeiro, vol.68 (3): e20240038.
- Santos, Vinicius; Lopes, Dawisson Belém. (2022), "Confirmação: o Lugar do Senado Federal na Política Externa Brasileira da Nova República." *Dados*, v. 66, P. E20200243.
- Souza e Silva, Jonatas Nogueira Aguiar. (2019), "A Comissão de Direitos Humanos e Legislação Participativa (Cdh) no Senado Brasileiro: um Estudo sobre sua Composição (2005-2018)." *Caos – Revista Eletrônica de Ciências Sociais*, N. 23, P. 79-112. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/caos/article/view/48175>

- Stockemer, Daniel; Koehler, Sebastian; Lenz, Tobias. (2018), “Data Access, Transparency, and Replication: New Insights from the Political Behavior Literature.” *Journal of Political Science Research*, v. 3, n. 2, p. 71-92.
- Vaqueiro, Ramon. (2023), Machine Learning Applied to Open Government Data for the Detection of Improprieties in the Application of Public Resources. In: *Proceedings of the XIX Brazilian Symposium on Information Systems*. 2023. p. 213-220.
- Weinstein, Jeremy; Goldstein, Joshua. The Benefits of a Big Tent: Opening Up Government in Developing Countries: A Response to Yu & Robinson’s The New Ambiguity of Open Government. *Ucla L. Rev. Discourse*, v. 60, p. 38, 2012.
- Wickham, Hadley. (2019), *Advanced R*. Chapman And Hall/CRC, 2019.
- Wickham, Hadley; Bryan, Jennifer. (2023), *R Packages*. O’reilly Media, Inc.
- Wickham, Hadley; Çetinkaya-Rundel, Mine; Golemund, Garrett. (2023), *R for Data Science*. O’reilly Media, Inc., 2023.