

## Réplica para a Carta às Editoras de autoria de Paiva et al.

Reply to the Letter to the Editors by Paiva et al.

Respuesta a la Carta a las Editoras de Paiva et al.

Raphael de Freitas Saldanha <sup>1</sup>

Ronaldo Rocha Bastos <sup>2</sup>

Christovam Barcellos <sup>1</sup>

doi: 10.1590/0102-311XPT117022

Agradecemos às autoras pela Carta às Editoras <sup>1</sup> referente ao artigo *Microdatasus: Pacote para Download e Pré-processamento de Microdados do Departamento de Informática do SUS (DATASUS)* <sup>2</sup>. Externamos nossa alegria em saber que o fruto desse trabalho tem contribuído para uma melhor atuação da comunidade científica.

O principal objetivo do artigo e do pacote estatístico criado é facilitar o acesso aos Sistemas de Informações em Saúde (SIS) mantidos pelo DATASUS, acrescentando a possibilidade e conveniência do acesso direto aos dados com a utilização da linguagem R. Lançado em 2019, o artigo contemplava os SIS acessíveis pelo pacote naquele momento. Desde então, com o esforço dos autores e comunidade de usuários, novos SIS foram acrescentados ao pacote, permitindo o acesso a dados de uma gama ainda maior de informações em saúde. Aproveitamos a oportunidade para agradecer a comunidade de usuários do pacote, que ativamente tem enviado sugestões e comunicado sobre erros e possibilidades de melhorias.

Uma das virtudes da linguagem R é a possibilidade de sua utilização em diversos sistemas operacionais, ainda que isso demande maior carga de trabalho no desenvolvimento e manutenção do pacote. Iremos aqui investigar os possíveis problemas relacionados ao uso e alocação de memória com o sistema operacional *Windows*.

Informamos que a seleção de variáveis no momento de importação dos dados já é possível, utilizando o argumento *vars* na função *fetch\_datasus*. Com esse argumento, após o *download* e leitura do arquivo no formato DBC fornecido pelo DATASUS, as variáveis nele listadas são selecionadas e as demais são descartadas, utilizando dessa forma uma quantidade menor de memória volátil do computador. O exemplo abaixo ilustra sua utilização. Ainda assim, iremos revisar essa função, adicionando comandos adicionais de *garbage collector*.

<sup>1</sup> Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz, Rio de Janeiro, Brasil.

<sup>2</sup> Departamento de Estatística, Universidade Federal de Juiz de Fora, Juiz de Fora, Brasil.

### Correspondência

R. F. Saldanha  
Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz, Av. Brasil 4365, Pavilhão Haity Moussatché, Rio de Janeiro, RJ 21040-900, Brasil.  
raphael.saldanha@icict.fiocruz.br



```
library(microdatasus)

dados_sim <- fetch_datasus(
  year_start = 2015, year_end = 2016,
  uf = "RJ",
  vars = c("SEXO", "CAUSABAS"),
  information_system = "SIM-DO"
)

dim(dados_sim)
```

Uma dúvida frequente dos usuários do pacote é sobre os limites de utilização de uma grande quantidade de dados. Por exemplo, os dados completos do Sistema de Internações Hospitalares do SUS (SIH/SUS), para uma dezena de anos de todas as Unidades da Federação.

A utilização de uma quantidade de dados de um recorte abrangente como esse irá demandar severamente recursos computacionais, em geral excedendo a capacidade disponível em computadores dos tipos *desktop* e *notebook*, que utilizamos usualmente.

Destacamos que isso não é uma limitação do pacote, mas da memória volátil disponível para os usuários em seus computadores. Nesses casos, sugerimos estratégias computacionais diferentes, mais apropriadas para lidar com grandes quantidades de dados.

Uma das opções é a utilização de um banco de dados. Nesse caso, os dados acessados e pré-processados pelo pacote podem ser armazenados em disco rígido, sendo posteriormente consultados dentro do R. O exemplo disponível no *link* <https://gist.github.com/rfsaldanha/a849716f4807fea7104b12e01fdf875f> demonstra essa possibilidade, utilizando o banco de dados SQLite localmente. Outras opções de banco de dados acessíveis pelo R são listadas no seguinte endereço: <https://db.rstudio.com>.

Destaca-se também a possibilidade de utilizar outras estratégias, como a Plataforma de Ciência de Dados em Saúde (PCDaS), do Instituto de Comunicação e Informação Científica e Tecnológica em Saúde, Fundação Oswaldo Cruz (ICICT/Fiocruz). Essa plataforma disponibiliza o acesso a diversos SIS do DATASUS em índices do tipo *ElasticSearch*. Dessa forma, é possível fazer consultas complexas com os dados do DATASUS e de outras fontes na infraestrutura computacional disponibilizada por essa plataforma. Mais informações podem ser acessadas no *link*: <https://pcdas.icict.fiocruz.br>.

Seguindo com os objetivos do pacote, nos comprometemos a expandir e dar manutenção às suas funções. Convidamos os usuários a contribuir ativamente para o pacote, seja com o envio de dúvidas e comentários, ou mesmo com a submissão de códigos em seu repositório: <https://github.com/rfsaldanha/microdatasus>.

O acesso e utilização de dados dos SIS é essencial para a construção e manutenção de políticas públicas em saúde. O DATASUS realiza um papel primordial com sua missão de coleta, organização e disseminação pública de dados de saúde. Nós, comunidade de usuários de dados do DATASUS, precisamos estar sempre vigilantes e prestativos para a manutenção desse órgão e consequente cumprimento de sua missão.

## Colaboradores

Todos os autores participaram igualmente na preparação do texto.

## Informações adicionais

ORCID: Raphael de Freitas Saldanha (0000-0003-0652-8466); Ronaldo Rocha Bastos (0000-0001-9597-5967); Christovam Barcellos (0000-0002-1161-2753).

1. Paiva NS, Meijinhos LS, Brochini MM. *Microdatasus*: uma ferramenta poderosa para a extração, carga e transformação de dados de saúde custodiados pelo Departamento de Informática do SUS. *Cad Saúde Pública* 2022; 38:e00096622.
2. Saldanha RF, Bastos RR, Barcellos C. *Microdatasus*: pacote para *download* e pré-processamento de microdados do Departamento de Informática do SUS (DATASUS). *Cad Saúde Pública* 2019; 35:e00032419.

---

Recebido em 22/Jun/2022  
Aprovado em 24/Jun/2022