



Implementação de novas funcionalidades ao pacote R "BrazilCrime": modelagem preditiva

Palavras-Chave: PACOTE R, PREVISÃO, CRIME

Autores(as):

GIOVANNI DOS SANTOS VARGETTE, IE - UNICAMP

Prof^(a). Dr^(a). MARCELO JUSTUS DOS SANTOS (orientador), IE - UNICAMP

INTRODUÇÃO:

O pacote de linguagem R, *BrazilCrime*, nasceu da necessidade de se facilitar o acesso às informações de criminalidade e violência ocorridas no território brasileiro. Tal demanda foi observada no projeto guarda-chuva de Iniciação Científica, intitulado Crime Violência e COVID-19, realizado na vigência de 2022 – 23 com apoio do CNPq.

No ano de 2023, demos início aos trabalhos de estruturação e implementação do pacote, tendo como finalidade principal apenas a simplificação do acesso aos dados divulgados pelo Sistema Nacional de Informações de Segurança Pública, Prisionais, de Rastreabilidade de Armas e Munições, de Material Genético, de Digitais e de Drogas (Sinesp), do Ministério da Justiça e Segurança Pública (MJ). Este objetivo foi alcançado em julho de 2024, quando foi publicado no CRAN, repositório oficial de pacotes em linguagem R, tendo atualmente 4267 *downloads*.

Na sua nova versão (2025), o pacote passou por importantes atualizações que ampliam suas funcionalidades e tornam mais acessíveis as previsões de tendências temporais em dados criminais. A atualização contempla a exclusão de funções obsoletas, melhorias de desempenho e a criação de uma função preditiva baseada em modelos estatísticos automatizados.

METODOLOGIA:

Assim como no projeto anterior, “Organização, sistematização e divulgação de dados sobre criminalidade e violência ocorrida nos estados brasileiros” (2023-24), esta atualização também foi estruturada em rotinas de programação utilizando a linguagem R¹, disponibilizadas através das funções *get_sinesp_vde_data* e *br_crime_predict*, a função *get_sinesp_data* foi descontinuada.

¹ Ver mais em: <<https://www.r-project.org>>

Inicialmente trataremos da descontinuidade de *get_sinesp_data*. Na primeira versão do pacote tínhamos a necessidade de duas funções que fizessem a busca dos dados por conta das duas metodologias distintas do Sinesp, com o sistema Sinesp Integração contendo dados de janeiro de 2015 a dezembro de 2022, enquanto o sistema Sinesp Validador de Dados Estatísticos (VDE), lançado em 2023, continha dados apenas a partir de janeiro deste ano, contando também com mais informações do que a metodologia de divulgação anterior. Com o passar do tempo, o sistema VDE foi implementado com informações de anos anteriores, tendo seu início também em janeiro de 2015. Portanto, tínhamos uma situação de redundância de informações, e já que o sistema VDE se tornou o principal meio de divulgação, contendo mais informações do que o Sinesp Integração, optamos pela exclusão da função *get_sinesp_data*, utilizando apenas *get_sinesp_vde_data* para realizar a disponibilização desses dados aos usuários.

Na função *get_sinesp_vde_data* realizamos um trabalho de atualização dos dados e de melhoria na documentação, principalmente com a inserção de exemplos elaborados de uso, gerando gráficos e análises de correlação entre variáveis de municípios específicos. Tais exemplos serão apresentados mais adiante.

A grande novidade dessa última versão do pacote BrazilCrime está na funcionalidade que nomeamos de *br_crime_predict*. Essa ferramenta se baseia na função já existente, *auto.arima*, do pacote *forecast*, desenvolvido principalmente pelo estatístico australiano Rob J. Hyndman. Ao fornecermos uma série temporal a esta ferramenta, temos algumas configurações disponíveis para selecionar o modo como será realizada a busca pelo modelo auto regressivo, integrado, de médias móveis (ARIMA) que melhor se adequa às características dos dados fornecidos, formando assim um modelo preditivo sobre a variável fornecida.

Com a funcionalidade *br_crime_predict* nosso objetivo foi facilitar a modelagem preditiva sobre séries temporais de crime. Para isso, fornecemos ao usuário uma espécie de função *auto.arima* pré-configurada, utilizando a composição mais comumente encontrada nos estudos sobre criminalidade e violência. Damos ao usuário a possibilidade de logaritmização da série, que por padrão está ativa nesta ferramenta, mas que pode ser desativada através do argumento “log = FALSE”, com isso temos uma redução da variância da série, bem como uma maior aproximação da distribuição normal, o que nos permite melhores resultados preditivos. O argumento “stepwise” é travado como TRUE, fazendo com que apenas modelos com significância estatística sejam trazidos aos usuários. Argumentos “allowdrift” e “allowmean” também são postos como “TRUE”, fazendo com que modelos com tendência e com média diferente de zero também sejam considerados. A avaliação sobre o ajuste do modelo é dada utilizando o Critério de Informação de Akaike (AIC), tal critério presa por modelos parcimoniosos, sendo assim, a escolha é feita por aquele modelo que minimize o AIC.

Além disso, *br_crime_predict* conta com um argumento de opção gráfica, caso “plot = TRUE”, além da modelagem, teremos também uma visão gráfica da predição realizada por este modelo,

facilitando, portanto, a construção de trabalhos acadêmicos, e até mesmo apontamentos oficiais sobre a tendência criminal em determinada localidade.

RESULTADOS E DISCUSSÃO:

Dada a proposta deste projeto utilizaremos exemplos para demonstrar os resultados obtidos, apresentando assim a capacidade da ferramenta criada.

Inicialmente temos a busca pelos dados de criminalidade e violência utilizando a função `get_sinesp_vde_data`, trazendo dados da categoria “ocorrências” no Estado de São Paulo, de janeiro de 2015 até dezembro de 2024.

Figura 1 – Coleta e Visualização Rápida de Ocorrências no Estado de São Paulos

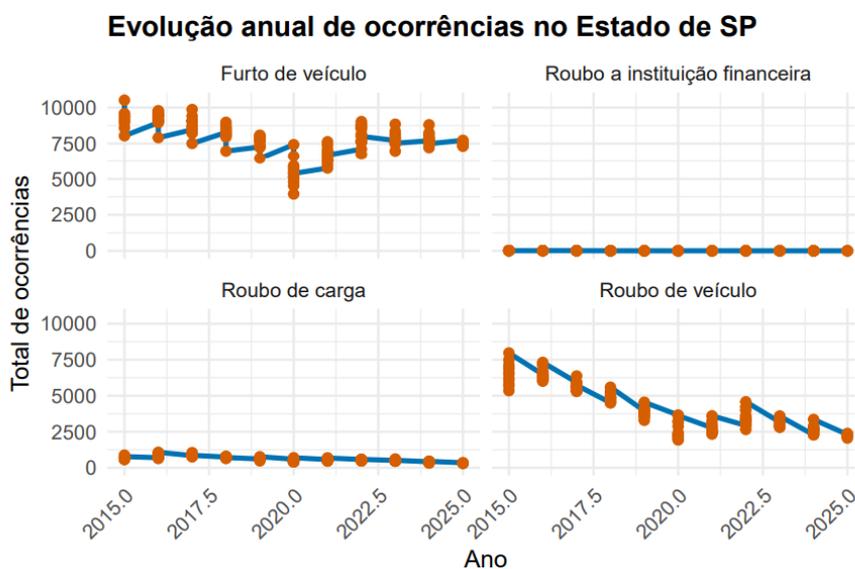
```
sp_anuais <- get_sinesp_vde_data(state = "SP", category = "ocorrencias", granularity = "year")
#> Query completed.
```

uf	municipio	ano	mes	categoria	evento	agente	arma	faixa_etaria	feminino	masculino	nao_informado	total	total_peso	total_vitima
SP	NÃO INFORMADO	2015	01	ocorrencias	Furto de veículo	NA	NA	NA	NA	NA	NA	9563	NA	NA
SP	NÃO INFORMADO	2015	01	ocorrencias	Roubo a instituição financeira	NA	NA	NA	NA	NA	NA	17	NA	NA
SP	NÃO INFORMADO	2015	01	ocorrencias	Roubo de carga	NA	NA	NA	NA	NA	NA	782	NA	NA
SP	NÃO INFORMADO	2015	01	ocorrencias	Roubo de veículo	NA	NA	NA	NA	NA	NA	7490	NA	NA
SP	NÃO INFORMADO	2015	02	ocorrencias	Furto de veículo	NA	NA	NA	NA	NA	NA	8966	NA	NA
SP	NÃO INFORMADO	2015	02	ocorrencias	Roubo a instituição financeira	NA	NA	NA	NA	NA	NA	17	NA	NA

Podemos utilizar essas informações na construção de um gráfico que apresente a evolução mensal das tipologias criminais contidas na categoria “ocorrências” no Estado de São Paulo.

Figura 2 – Evolução dos crimes da Categoria “ocorrências” em São Paulo – janeiro 2015 – dezembro 2024

```
ggplot(sp_anuais, aes(x = ano, y = total)) + geom_line(color = "#0072B2", size = 1.2) +
  geom_point(color = "#D55E00", size = 2) + facet_wrap(~evento) + labs(title = "Evolução anual de ocorrências no Estado de SP",
  y = "Total de ocorrências",
  x = "Ano") + theme_minimal(base_size = 12) + theme(plot.title = element_text(face = "bold",
  size = 14),
  axis.text.x = element_text(angle = 45, hjust = 1))
```



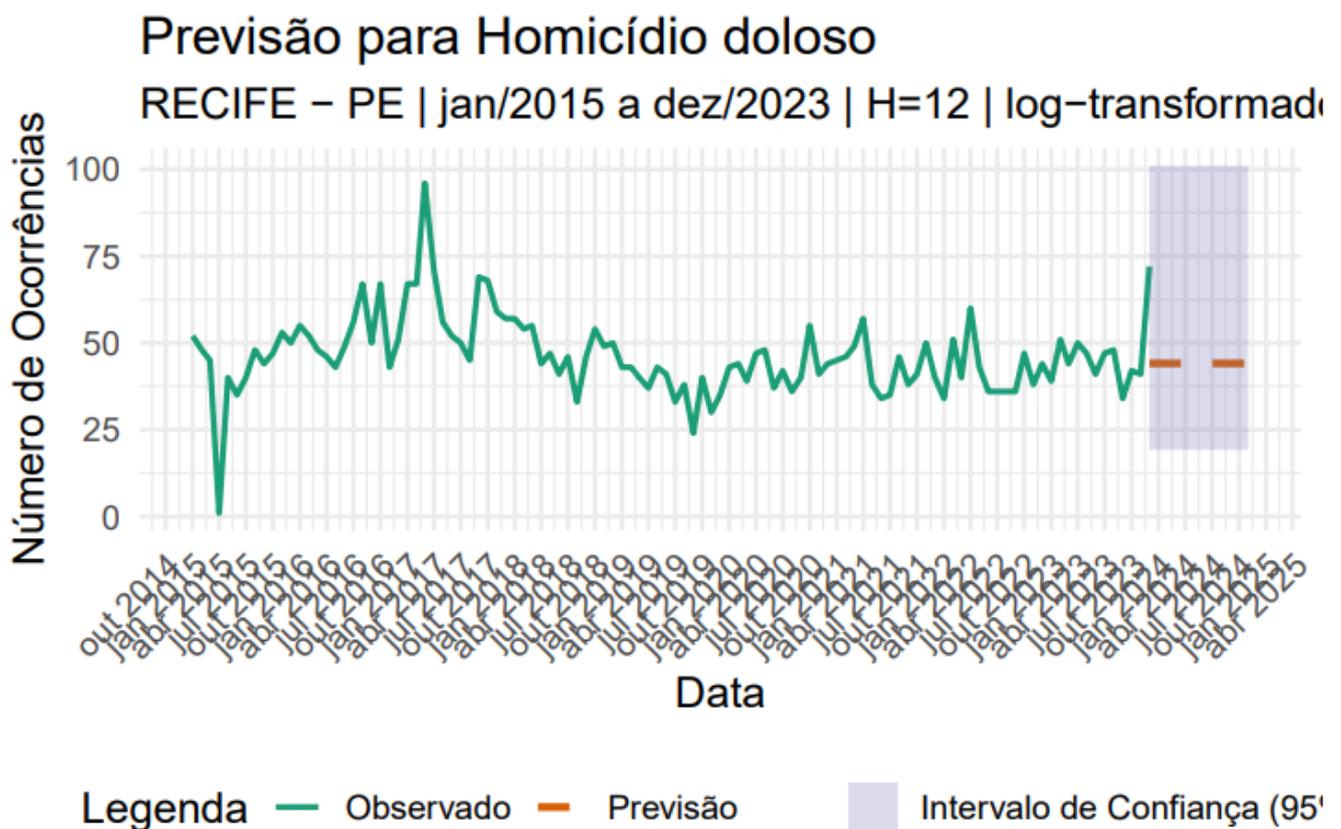
Com a utilização de *br_crime_predict*, faremos uma análise preditiva sobre os números de vítimas de homicídios dolosos em Recife – PE. Para isso, forneceremos a série temporal do total de vítimas mensal neste município, de janeiro de 2015 até dezembro de 2023, aplicaremos a transformação de Box-Cox, ou seja, aplicação de log sobre a série, e traremos a visualização gráfica gerada.

Figura 3 – Modelo preditivo de Homicídios Doloso em Recife – PE

```
dados <- get_sinesp_vde_data(state = "PE", city = "Recife", typology = "Homicídio doloso", category
↳ = "vítimas",
  granularity = "month", year = 2015:2023)
#> Query completed.

# Criar coluna de data (YYYY-MM)
dados <- dados |>
  dplyr::mutate(data = lubridate::ymd(paste0(ano, "-", mes, "-01"))) %>%
  dplyr::arrange(data)

# Rodar previsão
br_crime_predict(dados = dados, ts_col = "total_vitima", log = TRUE)
#> Convertendo dados para objeto 'ts'.
```



CONCLUSÕES:

Desta maneira conseguimos demonstrar alguns usos das funcionalidades do pacote *BrazilCrime*, e como essa implementação tem grande capacidade de auxiliar tanto na produção acadêmica, quanto de agentes públicos e privados, sobre o tema de criminalidade de violência nos territórios brasileiros.

Entendemos que existe uma grande quantidade de funções interessantes que podem ser desenvolvidas, como integrações com outras fontes de informações, busca e geração de dados populacionais para cálculos de taxas.

Portanto, esperamos que essas novas implementações agradem os mais de 4000 usuários atuais do pacote *BrazilCrime*, bem como possa cativar cada vez mais pessoas, buscando mais desenvolvimento da segurança pública brasileira, bem como a redução dos nossos índices de criminalidade.

BIBLIOGRAFIA

HYNDMAN, R., et al., (2025). **forecast: Forecasting functions for time series and linear models**. R package version 8.21, <<https://pkg.robjhyndman.com/forecast/>>.

PFAFF, B. (2008). **Analysis of Integrated and Cointegrated Time Series with R**. Second Edition. Springer, New York. ISBN 0-387-27960-1.

VARGETTE, G., LALTUF, I., JUSTUS, M. (2024). **_BrazilCrime: Accesses Brazilian Public Security Data from SINESP Since 2015_**. R package version 0.2, <<https://CRAN.R-project.org/package=BrazilCrime>>.

WICKHAM, H. **ggplot2: Elegant Graphics for Data Analysis**. Springer-Verlag New York, 2016.

WICKHAM H., et al. (2023). **dplyr: A Grammar of Data Manipulation**. R package version 1.1.1, <<https://CRAN.R-project.org/package=dplyr>>.

WICKHAM, H. **R Packages: Organize, Test, Document, and Share Your Code**. 1ª Edition. O'Reilly, 2015. KLEIBER, C., ZEILES, A. **Applied Econometrics With R**. 1ª Edition. Springer, 2008.