



TEXTO PARA DISCUSSÃO

ISSN 0103-9466

491

**BrazilCrime: um pacote R para facilitar a pesquisa em
segurança pública com dados dos estados e municípios
brasileiros**

**Bernardo Geraldini
Giovanni Vargette
Marcelo Justus
Igor Laltuf
Temidayo James Aransiola**

Novembro 2025



ie Instituto de
economia

BrazilCrime: um pacote R para facilitar a pesquisa em segurança pública com dados dos estados e municípios brasileiros

Bernardo Geraldini* Giovanni Vargette[†] Marcelo Justus[‡] Igor Laltuf[§]
Temidayo James Aransiola[¶]

Resumo

Ferramentas que tornam mais simples a extração e o tratamento de dados públicos contribuem de maneira decisiva para a produção de pesquisa de qualidade e para o desenvolvimento de políticas baseadas em evidências. Nesse contexto, apresentamos o pacote R **BrazilCrime**, desenvolvido por Vargette, Laltuf e Justus (2024). Trata-se de uma ferramenta simples que disponibiliza, de forma mais rápida e organizada, dados públicos do SINESP, uma plataforma de informações integradas desenvolvida e mantida pela Secretaria Nacional de Segurança Pública do Ministério da Justiça e Segurança Pública.

Palavras-chave: Segurança pública; Criminalidade; Séries temporais; Previsão estatística; Modelos Autorregressivos.

Códigos JEL: C32; K42; C55.

Abstract

Tools that simplify the extraction and processing of public data play a decisive role in producing high-quality research and developing evidence-based public policies. In this context, we introduce the R package **BrazilCrime**, developed by Vargette, Laltuf, and Justus (2024). It is a straightforward tool that provides faster and more organized access to public data from SINESP, an integrated information platform developed and maintained by the National Secretariat of Public Security of the Ministry of Justice and Public Security.

Keywords: Public security; Crime; Time series; Statistical forecasting; Autoregressive models.

JEL codes: C32; K42; C55.

*Instituto de Economia, Universidade Estadual de Campinas. b094671@dac.unicamp.br. ORCID 0000-0002-9132-8575.

[†]Instituto de Economia, Universidade Estadual de Campinas. g216978@dac.unicamp.br. ORCID 0009-0006-8164-5274.

[‡]Instituto de Economia, Universidade Estadual de Campinas. mjustus@unicamp.br. ORCID 0000-0001-8660-1779.

[§]Instituto de Pesquisa e Planejamento Urbano e Regional, Universidade Federal do Rio de Janeiro. igorlaltuf@gmail.com. ORCID 0000-0002-5614-4404.

[¶]Faculdade de Ciências Aplicadas, Universidade Estadual de Campinas. tjara@unicamp.br. ORCID 0000-0001-5578-0065.

1 Introdução

A pandemia de COVID-19 evidenciou a necessidade de contar com bancos de dados públicos organizados e facilmente acessíveis para uso em pesquisas. Nesse contexto, diversos estudos internacionais analisaram rapidamente os efeitos da crise sanitária sobre diferentes indicadores de segurança pública, como mostram Andresen e Hodgkinson (2020), Ashby (2020a, 2020b), Campedelli et al. (2020), Dai et al. (2021), Hodgkinson e Andresen (2020), Kemp et al. (2020), Kim e Phillips (2021), Langton et al. (2021), Mohler et al. (2020), Payne e Morgan (2020b), Shen et al. (2021) e Sutherland et al. (2021).

Ferramentas que tornam mais simples a extração e o tratamento de dados públicos contribuem de maneira decisiva para a produção de pesquisa de qualidade e para o desenvolvimento de políticas baseadas em evidências. É nesse cenário que se insere o pacote **BrazilCrime** (Vargette, Laltuf e Justus 2024), inspirado em iniciativas anteriores de pacotes específicos para acesso a dados de segurança pública, como o **crimedata** (Ashby 2018), o **ukpolice** (Odell e Tierney 2019) e o **ispdata** (Laltuf 2023).

O **BrazilCrime** tem como propósito facilitar o trabalho com informações produzidas no Brasil, permitindo que pesquisadores do país e do exterior lidem com dados de segurança pública de estados e municípios com maior rapidez. O pacote, desenvolvido inteiramente em linguagem R, organiza e disponibiliza, de forma eficiente, dados públicos do SINESP, plataforma de informações integradas mantida pela Secretaria Nacional de Segurança Pública do Ministério da Justiça e Segurança Pública. Ressalta-se que o pacote apenas extrai os dados e facilita análises utilizando a linguagem R. Portanto, tem como único objetivo apoiar a pesquisa empírica. Todos os créditos dos dados pertencem ao SINESP.

Na atualização mais recente do **BrazilCrime**, foram incluídas funcionalidades, baseadas no pacote **forecast**, que geram previsões automáticas de diversas tipologias de crimes e de outras variáveis da base, utilizando a modelagem clássica Box-Jenkins. Também foram adicionados gráficos com previsões pontuais e intervalares.

Ressalta-se que até este momento, desde que o pacote **BrazilCrime** foi publicado no CRAN (repositório oficial do R) em julho de 2024, foram registrados cerca de 5.860 downloads. A média mensal de aproximadamente 544 downloads mostra que, além de divulgar dados sobre ocorrências criminais e outras violências, é essencial garantir acesso rápido e organizado para pesquisadores que dependem dessas informações para gerar conhecimento científico que possa ser aplicado para a elaboração de políticas públicas baseadas em evidências.

2 Pacote BrazilCrime

O pacote disponibiliza duas funções centrais: uma para a coleta e organização de dados criminais a partir do Sistema Nacional de Estatísticas de Segurança Pública (SINESP), e outra que visa auxiliar na criação de modelos preditivos. A primeira delas, **get_sinesp_vde_data**, tem no seu escopo a coleta de dados em nível municipal a partir da base SINESP VDE. Cobrindo o período de janeiro de 2015 a maio de 2025 (esse período é expandido dentro do pacote trimestralmente), contempla diversas tipologias, como homicídio doloso, roubo de veículos, roubo de carga, entre outras, listadas abaixo. Essa função possibilita filtragens refinadas, incluindo parâmetros como estado, município, ano, categoria de crime e tipologia, além da escolha da granularidade temporal (anual ou mensal). Essa função facilita análises comparativas entre cidades

e investigações mais detalhadas da dinâmica espacial do crime no Brasil. Dessa forma, temos uma das possibilidades de pesquisa empírica em segurança pública e economia do crime, bem como a formulação de políticas públicas baseadas em evidências.

A principal contribuição do pacote `BrazilCrime` é disponibilizar de maneira acessível os dados sobre criminalidade e violência do Brasil por meio da linguagem R. Disponível no CRAN desde julho de 2024, a nova versão do pacote (2025) passou por importantes atualizações que ampliam suas funcionalidades e tornam mais acessível a previsão de tendências temporais em dados criminais. A atualização contempla a exclusão de funções obsoletas, melhorias de desempenho e a criação de uma função de previsão baseada na tradicional modelagem Box-Jenkins (Box, Jenkins, Reinsel e Ljung, 2016).

Nesse sentido, a principal novidade desta versão do pacote `BrazilCrime` é a funcionalidade denominada `br_crime_predict`. Essa ferramenta se baseia na função já consolidada `auto.arima`, do pacote `forecast` (Hyndman et al., 2025), desenvolvido principalmente pelo estatístico australiano Rob J. Hyndman. Ao fornecer uma série temporal à função, o usuário conta com diferentes configurações para determinar como será realizada a busca pelo modelo autorregressivo integrado de médias móveis (ARIMA), sazonal ou não sazonal, mais adequado às características dos dados, resultando em um modelo preditivo para a variável em questão. O objetivo da função é tornar o pacote um recurso facilitador tanto para a elaboração de trabalhos acadêmicos quanto para a produção de relatórios técnicos e notas oficiais sobre tendências da criminalidade.

Essa nova funcionalidade busca simplificar a modelagem preditiva de séries temporais criminais, especialmente para pesquisadores com pouco conhecimento de modelos de previsão de séries temporais univariadas. Para isso, cria uma versão pré-configurada da função `auto.arima`, estruturada segundo as composições mais comuns na literatura sobre criminalidade e violência.

Outros parâmetros também foram fixados para garantir robustez. O argumento `stepwise` permanece definido como `TRUE`, de modo que apenas modelos com todos os coeficientes estatisticamente significativos sejam retornados. Os argumentos `allowdrift` e `allowmean` foram igualmente configurados como `TRUE`, permitindo que modelos com tendência e média distinta de zero sejam considerados. A escolha final do modelo é guiada pelo Critério de Informação de Akaike (AIC), que privilegia parcimônia, de modo que a função seleciona aquele que apresenta o menor valor do AIC. Além disso, foi implementada a saída gráfica. Tal funcionalidade faz com que, além da estimação do modelo e previsões, a função gere uma visualização da predição resultante.

3 Exemplos de funcionalidades

Como forma de demonstrar o uso das funções do pacote `BrazilCrime`, apresentamos a seguir exemplos de aplicação que podem ser utilizados de maneira integrada com outros pacotes da linguagem R. Primeiro, carregamos todos os pacotes que serão necessários para rodar todo o código.

```
# Pacotes utilizados em todo o código  
library(BrazilCrime)  
library(dplyr)  
library(tidyr)  
library(ggplot2)
```

```
library(lubridate)
library(forecast)
library(patchwork)
library(purrr)
library(stringr)
library(readr)
library(vars)
library(cowplot)
library(corrplot)
library(gridExtra)
```

3.1 Variáveis disponíveis

Consta abaixo o código que gera uma lista com todas as possíveis ocorrências criminais. Deve-se notar que diversas variáveis não apresentam registro no sistema. Por exemplo, não há ocorrências relativas à apreensão de drogas nem de feminicídio. Isso se deve a limitações da própria plataforma SINESP e não do pacote BrazilCrime.

```
df <- get_sinesp_vde_data( state = "all",
                           city = "all",
                           year = "2023",
                           category = "all",
                           typology = "all",
                           granularity = "year")

unique(df$evento)
```

```
## [1] "Mandado de prisão cumprido"
## [2] "Feminicídio"
## [3] "Homicídio doloso"
## [4] "Lesão corporal seguida de morte"
## [5] "Morte no trânsito ou em decorrência dele (exceto homicídio doloso)"
## [6] "Mortes a esclarecer (sem indício de crime)"
## [7] "Roubo seguido de morte (latrocínio)"
## [8] "Suicídio"
## [9] "Tentativa de feminicídio"
## [10] "Tentativa de homicídio"
## [11] "Arma de Fogo Apreendida"
## [12] "Atendimento pré-hospitalar"
## [13] "Busca e salvamento"
## [14] "Combate a incêndios"
## [15] "Emissão de Alvarás de licença"
```

```
## [16] "Realização de vistorias"
## [17] "Pessoa Desaparecida"
## [18] "Pessoa Localizada"
## [19] "Apreensão de Cocaína"
## [20] "Apreensão de Maconha"
## [21] "Tráfico de drogas"
## [22] "Furto de veículo"
## [23] "Roubo a instituição financeira"
## [24] "Roubo de carga"
## [25] "Roubo de veículo"
## [26] "Morte de Agente do Estado"
## [27] "Suicídio de Agente do Estado"
## [28] "Estupro"
## [29] "Morte por intervenção de Agente do Estado"
## [30] "Estupro de vulnerável"
```

3.2 Análise gráfica

Iniciamos com uma análise gráfica da evolução das tipologias criminais incluídas na categoria “ocorrências” para o Estado de São Paulo. As tipologias consideradas são furto de veículo, roubo a instituição financeira, roubo de carga e roubo de veículo. Ressalta-se que, para essas tipologias, não estão disponíveis informações em nível municipal. Por meio das linhas de código abaixo, utiliza-se a função `get_sinesp_vde_data` para coletar os dados de interesse. Em seguida, essas informações são inseridas na função `ggplot`, do pacote `ggplot2` (Wickham, 2016), para construir um gráfico dessas ocorrências (Figura 1).

```
# Figura 1
sp_anuais <- get_sinesp_vde_data(state = "SP", category = "ocorrencias",
                                granularity = "year") |>
  group_by(uf, ano, evento) |>
  summarise(total_ocorrencias_ano = sum(total, na.rm = T))

ggplot(sp_anuais, aes(x = ano, y = total_ocorrencias_ano, color = "Observado")) +
  geom_line(linewidth = 0.9, na.rm = TRUE) +
  labs(
    title = "Séries de Ocorrências Criminais",
    subtitle = "São Paulo | Totais anuais por evento",
    x = "Ano",
    y = "Total de Ocorrências",
    color = "Legenda"
  ) +
  scale_color_manual(values = c("Observado" = "#1B9E77")) +
  scale_x_continuous(breaks = function(x) seq(floor(min(x)), ceiling(max(x)), by = 1)) +
  scale_y_continuous(labels = function(x) format(x, big.mark = ".", scientific = FALSE)) +
```

```
facet_wrap(~evento, scales = "free_y") +
theme_minimal(base_size = 14) +
theme(
  legend.position = "bottom",
  axis.text.x = element_text(angle = 45, hjust = 1)
)
```

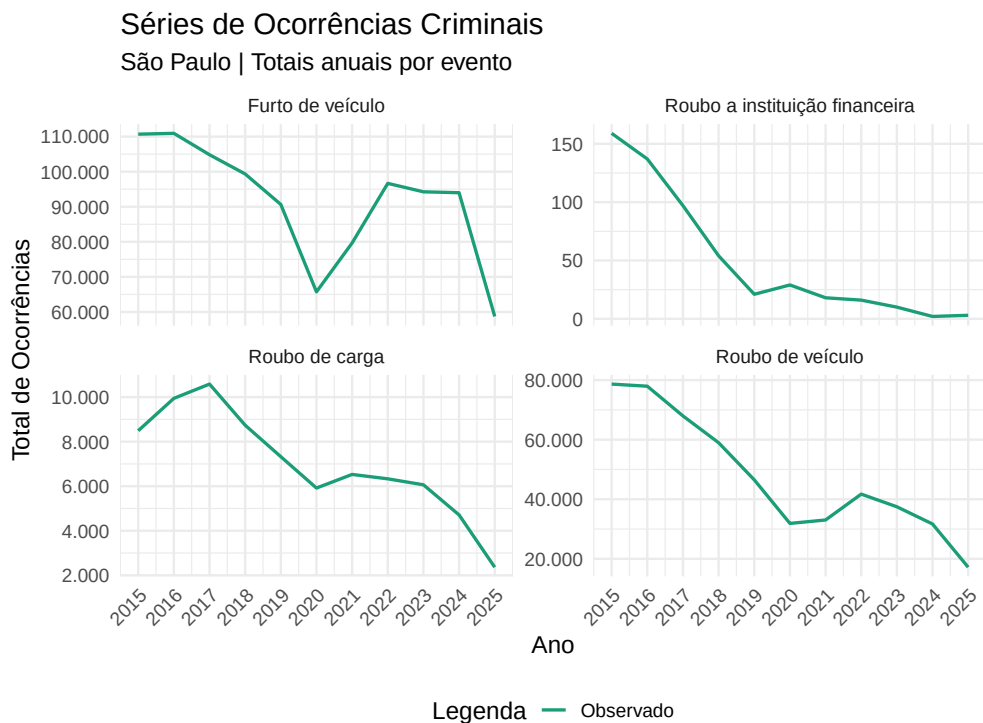


Figura 1: Número de ocorrências criminais registradas no Estado de São Paulo entre janeiro de 2015 e dezembro de 2024, por ano e tipo de crime.

3.3 Análise de correlação

Analizamos as correlações simples a partir dos dados anuais de ocorrências criminais e administrativas registradas no SINESP em 2023, agregados por Unidade da Federação.

Observam-se correlações altas e positivas entre crimes patrimoniais, como roubo de veículos e roubo de cargas (0,93), roubo de veículos e furto de veículos (0,86) e roubo de cargas e furto de veículos (0,82). A correlação entre tráfico de drogas e arma de fogo apreendida também é elevada (0,91).

Algumas correlações altas, porém, são claramente espúrias. Entre elas, tráfico de drogas e busca e salvamento, furto de veículos e combate a incêndios, furto de veículos e mandados de prisão cumpridos e realização de vistorias e furto de veículos.

```

df_estados <- get_sinesp_vde_data( state = "all",
                                   city = "all",
                                   year = "2023",
                                   category = "all",
                                   typology = "all",
                                   granularity = "year") |>

  group_by(uf, evento) |>
  summarise(total_ocorrencias_ano = sum(total, na.rm = T))

df_estados_w <- df_estados |>
  dplyr::select(uf, evento, total_ocorrencias_ano) |>
  tidyr::pivot_wider(
    names_from = evento,
    values_from = total_ocorrencias_ano,
    values_fill = 0
  )

# Removendo colunas que contêm apenas zeros

df_estados_w <- df_estados_w |> dplyr::select(-c(1:3, 9:11, 13, 14, 16:21, 26:30))

df_estados_w <- df_estados_w |>
  dplyr::ungroup() |>
  dplyr::select(-c("uf"))

# Matriz de correlação
cor_df_matriz_limpa <- cor(df_estados_w,
                           use = "complete.obs")

testRes = cor.mtest(cor_df_matriz_limpa, conf.level = 0.95)

# Plotando correlações
matriz1 <- corrplot(cor_df_matriz_limpa,
  p.mat = testRes$p,
  insig = 'blank',
  order = 'alphabet',
  diag = FALSE,
  type = "lower",
  sig.level = 0.05,
  tl.col = "grey21",
  addCoef.col = T,
  number.cex = 0.7,

```



```

number.digits = 2,
tl.srt = 45,
tl.cex = 0.7)

```

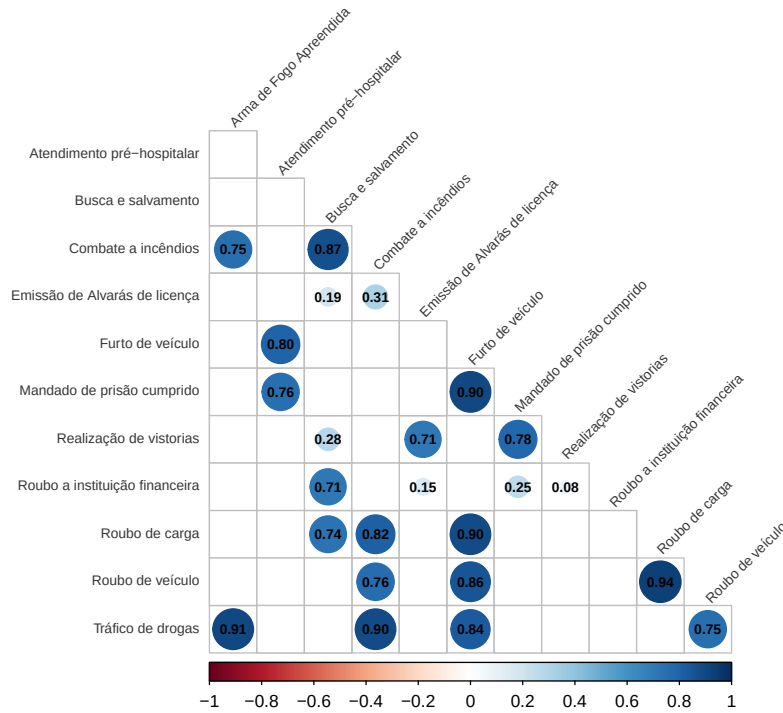


Figura 2: Matriz de correlações para ocorrências do ano de 2023, agregadas por Unidade da Federação. São exibidas somente as correlações que possuem valor $p < 0,05$.

3.4 Análise preditiva

Utilizaremos a função `br_crime_predict` para conduzir uma análise preditiva da série de homicídios dolosos na cidade de São Paulo (SP), considerando informações de janeiro de 2015 a dezembro de 2024 e projetando os 12 meses subsequentes. O principal benefício do uso da função `br_crime_predict`, em comparação a outras funções intermediárias disponíveis nesse processo, está na automatização das configurações do pacote `BrazilCrime` e na geração automática dos gráficos de previsão. Para isso, emprega-se o seguinte código:

```

hmc_d_sp <- get_sinesp_vde_data(state = "sp", city = "são paulo", year = 2015:2024,
category = "vitimas", typology = "homicídio doloso")

Acf(log(hmc_d_sp$total_vitima), lag.max = 36,
    main = "Função de Autocorrelação da série original de Homicídios em São Paulo - SP")

```

Função de Autocorrelação da série original de Homicídios em São Paulo – SP

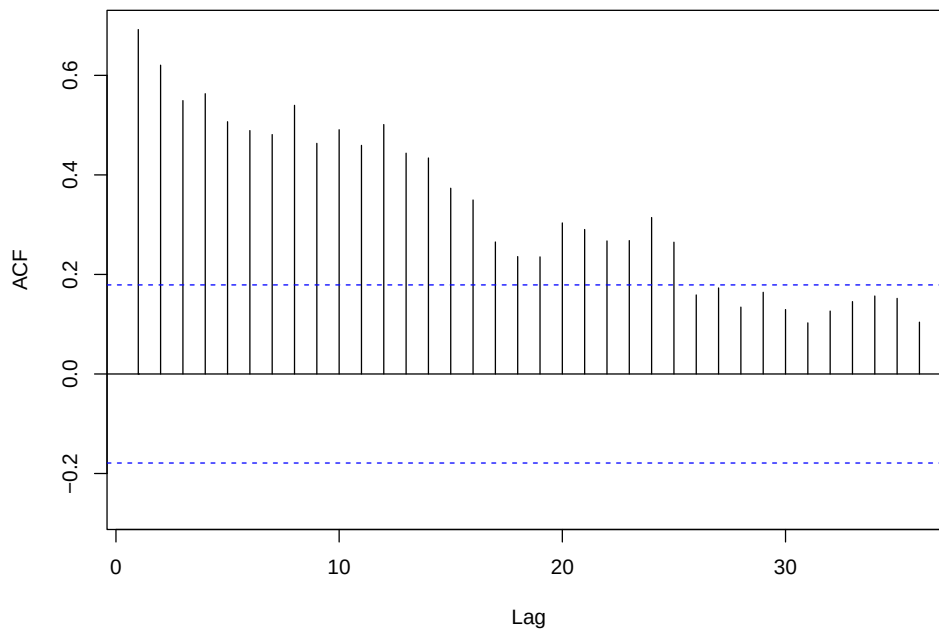


Figura 3: Função de Autocorrelação da série original de Homicídios em São Paulo - SP (jan/2015 – dez/2024).

```
br_crime_predict(dados = hmcd_sp, ts_col = "total_vitima")
```

```
Series: ts_obj
```

```
ARIMA(0,1,1)(2,0,0)[12]
```

```
Coefficients:
```

	ma1	sar1	sar2
	-0.8180	0.3056	0.3121
s.e.	0.0527	0.0893	0.0978

```
sigma^2 = 88: log likelihood = -436.81
```

```
AIC=881.61 AICc=881.96 BIC=892.73
```

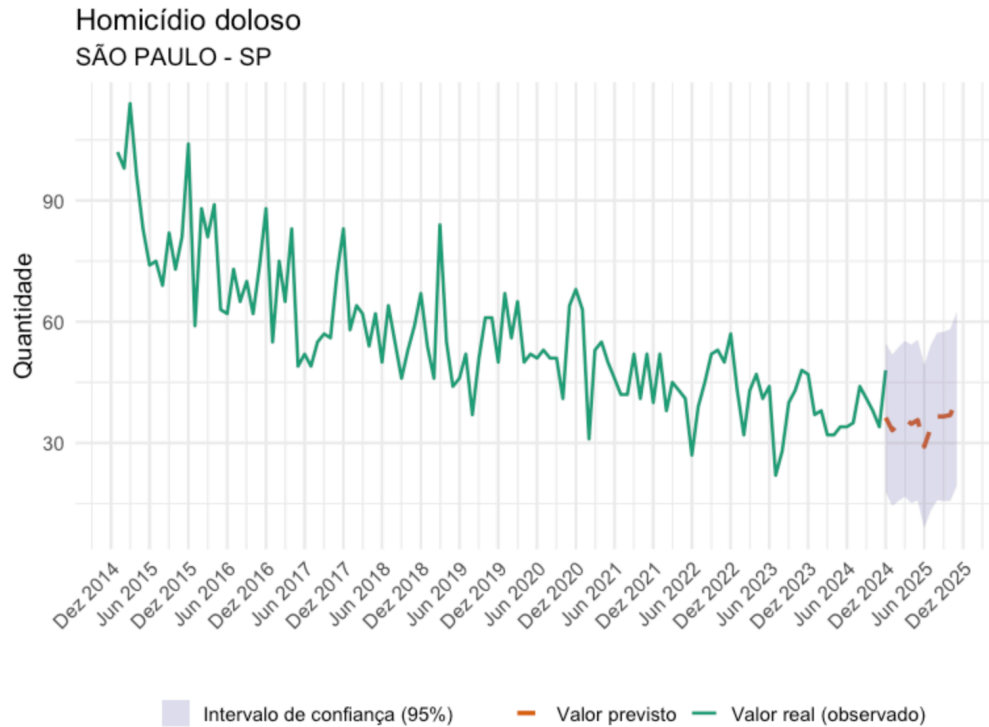


Figura 4: Valores observados até dezembro de 2024 e previsão de homicídios dolosos na cidade de São Paulo (jan/2025 – dez/2025).

Entre os outputs do código acima, destacamos o modelo ARIMA encontrado e o gráfico com a previsão de homicídios.

Vejamos, agora, a previsão para roubos e furtos de veículos nos estados de Minas Gerais e São Paulo:

Figura 5

São Paulo

```
SP_roubo_furto_veiculos <- get_sinesp_vde_data(state = c("SP"),
                                              category = "ocorrencias",
                                              typology = c("Roubo de veículo",
                                                         "Furto de veículo"),
                                              granularity = "year") |>

dplyr::filter(ano > 2017) |>
mutate(evento = "roubo e furto de veículos - SP",
       uf = "SP",
       municipio = "") |>
group_by(data, evento, municipio, uf) |>
summarise(total = sum(total))
```

```
predict_veic_SP <- br_crime_predict(SP_roubo_furto_veiculos,
                                     ts_col = "total")
```

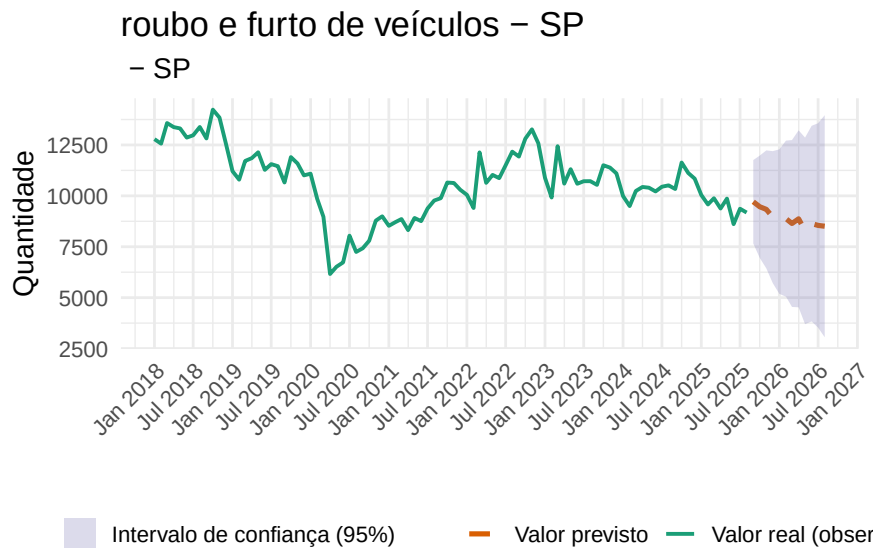


Figura 5: Valores observados até dezembro de 2024 e previsão de roubos e furtos (somados) no estado de São Paulo (jan/2025 – dez/2025).

Minas Gerais

```
MG_roubo_furto_veiculos <- get_sinesp_vde_data(state = c("MG"),
                                              category = "ocorrencias",
                                              typology = c("Roubo de veículo",
                                                            "Furto de veículo"),
                                              granularity = "year") |>

dplyr::filter(ano > 2017) |>
mutate(evento = "roubo e furto de veículos - MG",
       uf = "MG",
       municipio = "") |>
group_by(data, evento, municipio, uf) |>
summarise(total = sum(total))

predict_veic_MG <- br_crime_predict(MG_roubo_furto_veiculos,
                                     ts_col = "total")
```

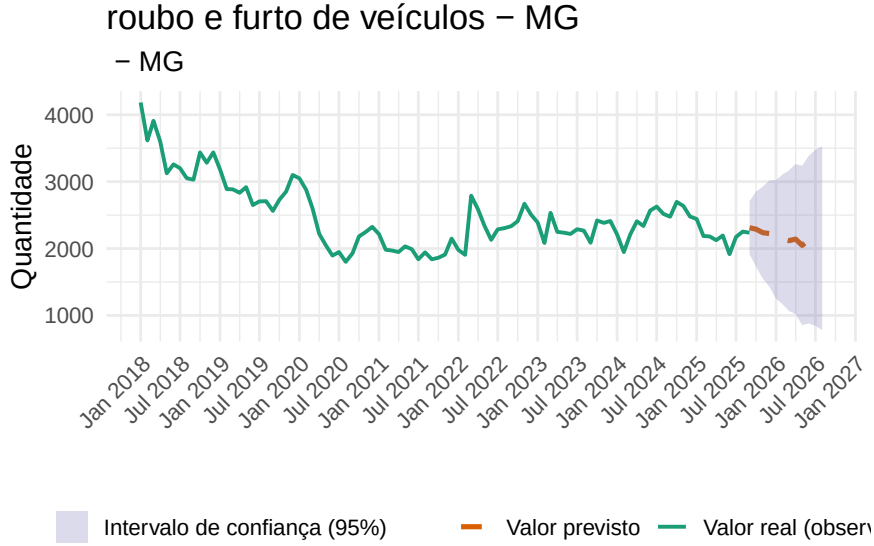


Figura 6: Valores observados até dezembro de 2024 e previsão de roubos e furtos (somados) no estado de Minas Gerais (jan/2025 – dez/2025).

3.5 Modelos vetoriais autorregressivos (VAR)

Os dados obtidos com o pacote podem ser organizados em séries temporais e, a partir disso, empregados na estimação de modelos do tipo vetor autorregressivo (VAR). No exemplo a seguir, analisamos o roubo de carga nos estados de São Paulo (SP), Rio de Janeiro (RJ) e Minas Gerais (MG). A hipótese é a de que a evolução desse tipo de crime em cada estado depende não apenas do seu próprio histórico nesse estado, mas também das ocorrências registradas nos demais estados da região. Em forma matricial, o sistema pode ser representado da seguinte maneira:

$$\begin{bmatrix} MG_t \\ SP_t \\ RJ_t \end{bmatrix} = A_1 \begin{bmatrix} MG_{t-k} \\ SP_{t-k} \\ RJ_{t-k} \end{bmatrix} + \begin{bmatrix} \varepsilon MG_t \\ \varepsilon SP_t \\ \varepsilon RJ_t \end{bmatrix}$$

em que A_1 é uma matriz que mapeia os efeitos das variáveis em $t - k$ sobre cada equação em t .

O código a seguir mostra como estimar modelos dessa natureza. Primeiro, os dados mensais são baixados do SINESP por meio do pacote, agregados por estado e filtrados apenas para a tipologia “roubo de carga”. Em seguida, as séries são organizadas em formato largo e convertidas em objeto de série temporal mensal. Por fim, aplica-se logaritmo sobre todas as séries de tempo, a fim de suavizar a variabilidade, reduzir a assimetria causada por valores extremos e permitir a obtenção de taxas de crescimento em vez de variações absolutas. Assim, é mais fácil comparar séries de diferentes magnitudes.

```
sudeste_mensais <- get_sinesp_vde_data(state = c("SP", "RJ", "MG"),
                                     category = "ocorrencias",
                                     granularity = "month") |>
group_by(uf, mes, evento, ano) |>
```

```

summarise(total_ocorrencias = sum(total, na.rm = T))

sudeste_carga <- sudeste_mensais |> filter(evento == "Roubo de carga" & ano < 2025)

sudeste_carga <- sudeste_carga %>%
  mutate(data = as.Date(paste0(ano, "-", mes, "-01"))) %>%
  dplyr::select(uf, data, total_ocorrencias)

sudeste_wide <- sudeste_carga %>%
  pivot_wider(names_from = uf, values_from = total_ocorrencias) %>%
  arrange(data)

sudeste_ts <- ts(sudeste_wide,
  start = c(min(format(sudeste_wide$data, "%Y")) %>% as.numeric(),
    min(format(sudeste_wide$data, "%m")) %>% as.numeric()),
  frequency = 12)

colnames(sudeste_ts) <- c("mes", "evento", "data", "MG", "RJ", "SP")

# aplicando log

sudeste_ts_log <- sudeste_ts
sudeste_ts_log[, c("MG", "RJ", "SP")] <- log(sudeste_ts[, c("MG", "RJ", "SP")])

```

Com o código abaixo, são executados os testes ADF para raiz unitária. Os resultados (não mostrados) sugerem que o modelo seja estimado para as primeiras diferenças das séries originais.

```

adf1 <- summary(ur.df(sudeste_ts_log[,4], type = "trend", lags = 2))
adf1

adf2 <- summary(ur.df(sudeste_ts_log[,5], type = "trend", lags = 2))
adf2

adf3 <- summary(ur.df(sudeste_ts_log[,6], type = "trend", lags = 2))
adf3

```

A função VARselect, abaixo, é usada para sugerir a defasagem ótima do modelo a partir de critérios de informação.

```

VARselect(diff(sudeste_ts_log[,4:6]), lag.max = 4, type = "both")

```

```

## $selection
## AIC(n)  HQ(n)  SC(n) FPE(n)

```

```
##      3      1      1      3
##
## $criteria
##              1              2              3              4
## AIC(n) -1.003285e+01 -1.010093e+01 -1.010613e+01 -1.009035e+01
## HQ(n)  -9.887521e+00 -9.868408e+00 -9.786415e+00 -9.683439e+00
## SC(n)  -9.674811e+00 -9.528071e+00 -9.318452e+00 -9.087851e+00
## FPE(n)  4.394018e-05  4.106922e-05  4.090049e-05  4.162928e-05
```

Optamos por estimar um VAR(1) com constante, considerando as primeiras diferenças das séries estaduais. Estimaremos, então, a equação abaixo:

$$\begin{bmatrix} \Delta \log(MG_t) \\ \Delta \log(SP_t) \\ \Delta \log(RJ_t) \end{bmatrix} = A_1 \begin{bmatrix} \Delta \log(MG_{t-1}) \\ \Delta \log(SP_{t-1}) \\ \Delta \log(RJ_{t-1}) \end{bmatrix} + \begin{bmatrix} \varepsilon MG_t \\ \varepsilon SP_t \\ \varepsilon RJ_t \end{bmatrix}$$

O comando VAR faz a estimação, e o comando diff realiza a diferenciação das séries. O comando summary(modelo_var1) apresenta os coeficientes estimados, erros-padrão e estatísticas de ajuste.

```
modelo_var1 <- VAR(diff(sudeste_ts_log[,4:6]), p = 1, type = "const")
```

```
summary(modelo_var1)
```

```
##
## VAR Estimation Results:
## =====
## Endogenous variables: MG, RJ, SP
## Deterministic variables: const
## Sample size: 118
## Log Likelihood: 105.867
## Roots of the characteristic polynomial:
## 0.481 0.1616 0.1616
## Call:
## VAR(y = diff(sudeste_ts_log[, 4:6]), p = 1, type = "const")
##
##
## Estimation results for equation MG:
## =====
## MG = MG.l1 + RJ.l1 + SP.l1 + const
##
##      Estimate Std. Error t value Pr(>|t|)
## MG.l1 -0.476605   0.083521  -5.706 9.27e-08 ***
## RJ.l1 -0.102608   0.166421  -0.617   0.539
## SP.l1  0.005179   0.247787   0.021   0.983
```

```

## const -0.007312    0.027827   -0.263    0.793
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.3018 on 114 degrees of freedom
## Multiple R-Squared: 0.2357,   Adjusted R-squared: 0.2156
## F-statistic: 11.72 on 3 and 114 DF,  p-value: 9.458e-07
##
##
## Estimation results for equation RJ:
## =====
## RJ = MG.l1 + RJ.l1 + SP.l1 + const
##
##           Estimate Std. Error t value Pr(>|t|)
## MG.l1 -0.0272383   0.0529203  -0.515    0.608
## RJ.l1 -0.0932928   0.1054479  -0.885    0.378
## SP.l1  0.0476276   0.1570028   0.303    0.762
## const  0.0002799   0.0176315   0.016    0.987
##
##
## Residual standard error: 0.1913 on 114 degrees of freedom
## Multiple R-Squared: 0.009694,   Adjusted R-squared: -0.01637
## F-statistic: 0.372 on 3 and 114 DF,  p-value: 0.7734
##
##
## Estimation results for equation SP:
## =====
## SP = MG.l1 + RJ.l1 + SP.l1 + const
##
##           Estimate Std. Error t value Pr(>|t|)
## MG.l1 -0.033586    0.034100  -0.985    0.3267
## RJ.l1 -0.152452    0.067946  -2.244    0.0268 *
## SP.l1 -0.222339    0.101166  -2.198    0.0300 *
## const -0.006225    0.011361  -0.548    0.5848
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##
## Residual standard error: 0.1232 on 114 degrees of freedom
## Multiple R-Squared: 0.1631,   Adjusted R-squared: 0.1411
## F-statistic: 7.408 on 3 and 114 DF,  p-value: 0.0001409
##

```



```
##
##
## Covariance matrix of residuals:
##      MG      RJ      SP
## MG 0.091111 0.004587 0.004574
## RJ 0.004587 0.036579 0.011959
## SP 0.004574 0.011959 0.015187
##
## Correlation matrix of residuals:
##      MG      RJ      SP
## MG 1.00000 0.07946 0.1230
## RJ 0.07946 1.00000 0.5074
## SP 0.12297 0.50738 1.0000
```

Antes de interpretar os resultados, é fundamental realizar análises de diagnóstico dos resíduos. Abaixo são apresentados gráficos com valores ajustados e residuais. Observa-se, por exemplo, que a série ajustada para o Rio de Janeiro acompanha razoavelmente bem o comportamento dos dados observados. Entretanto, os resíduos ainda apresentam alta variabilidade em alguns períodos. Além disso, o gráfico dos resíduos sugere que os erros estão centrados em torno de zero, sem tendência sistemática clara.

Embora a maioria das correlações seja estatisticamente não significativa dentro das bandas, em alguns *lags* elas são significativas, o que sugere a presença de autocorrelação residual. Portanto, o modelo estimado ainda não captura totalmente as dinâmicas de curto prazo. Contudo, como este exercício tem apenas caráter ilustrativo do uso do pacote **BrazilCrime** para facilitar o acesso aos dados, segue-se em frente com essa especificação.

```
plot(modelo_var1, names = "MG")
```

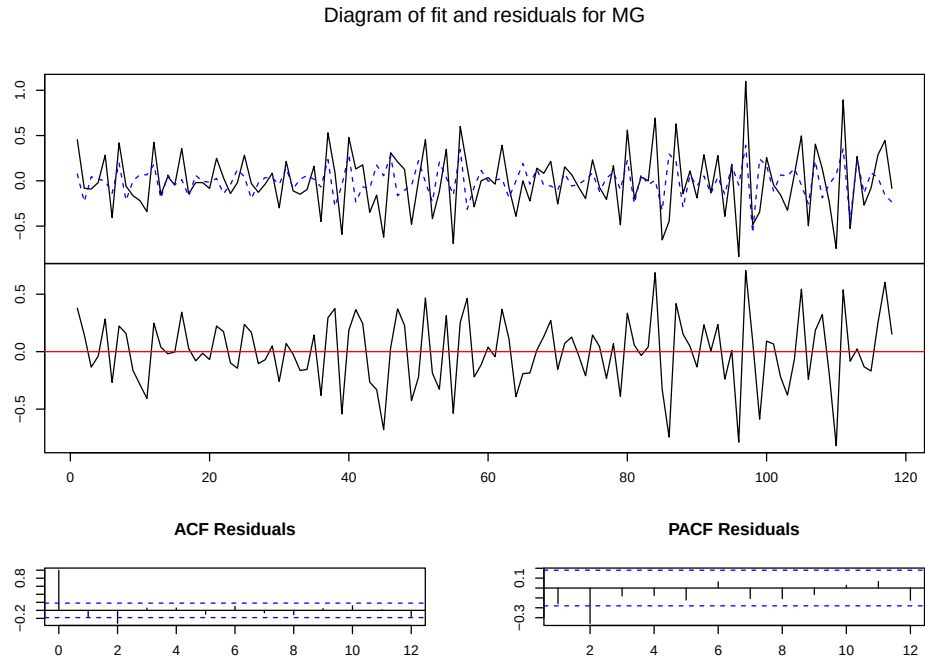


Figura 7: Valores ajustados e residuais para Minas Gerais.

```
plot(modelo_var1, names = "SP")
```

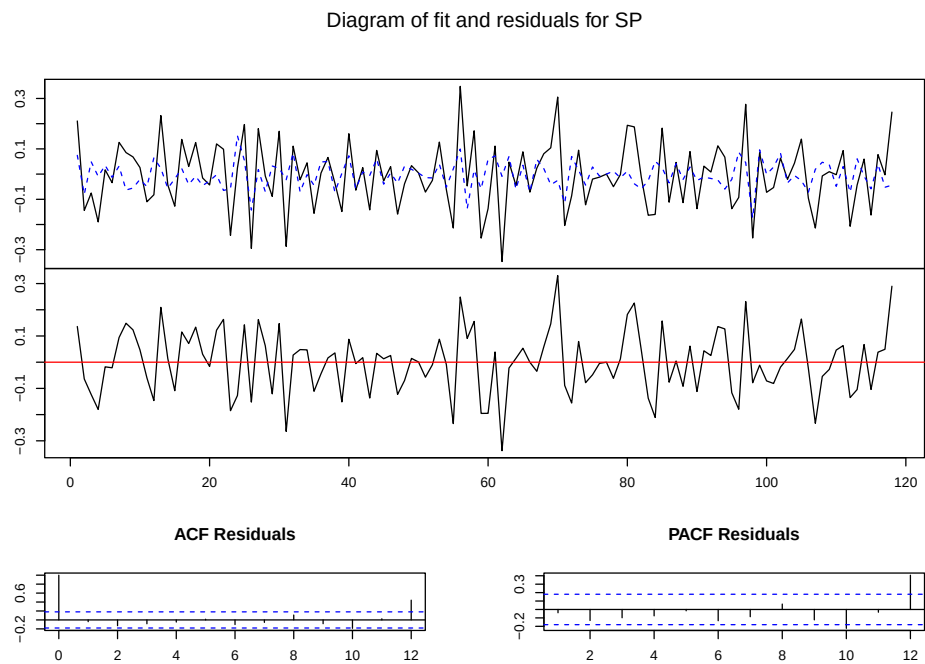


Figura 8: Valores ajustados e residuais para São Paulo.

```
plot(modelo_var1, names = "RJ")
```

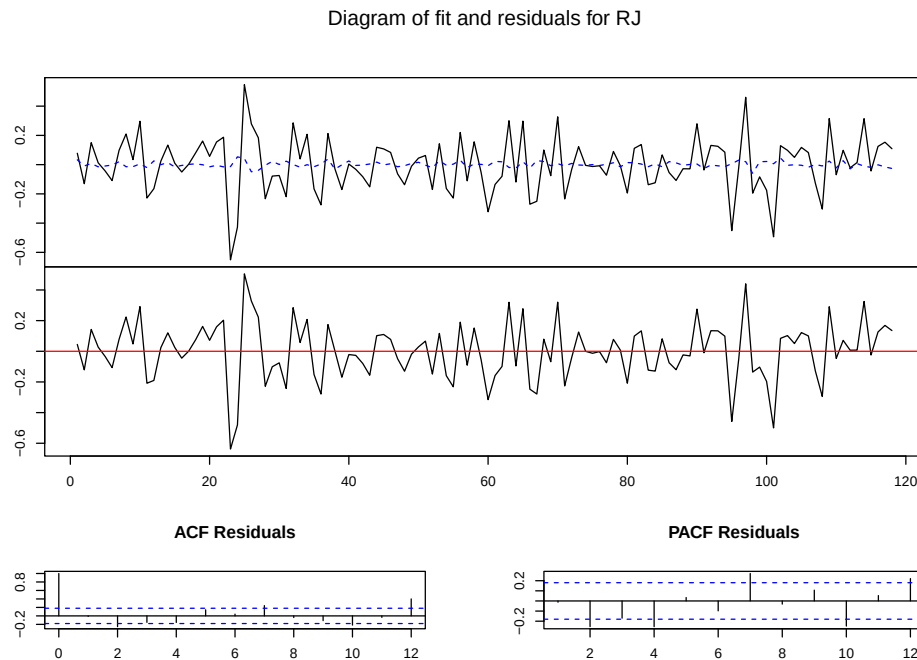


Figura 9: Valores ajustados e residuais para o Rio de Janeiro.

Os testes de diagnóstico abaixo permitem avaliar a adequação da especificação. O teste de autocorrelação rejeita a hipótese nula de ausência de autocorrelação serial dos resíduos ($p < 0,001$), indicando dependência temporal não capturada pelo modelo com apenas uma defasagem. O teste de heteroscedasticidade multivariada *ARCH*, por sua vez, não rejeita a hipótese nula ($p = 0,43$), o que indica ausência de evidências de efeitos *ARCH* relevantes nos resíduos. Os testes de normalidade multivariada (Jarque-Bera, assimetria e curtose) também não apontam violações estatisticamente significativas, com p -valores acima de 0,16. Esses resultados sugerem distribuição aproximadamente normal dos erros. No conjunto, o modelo mostra adequação quanto à normalidade e à homoscedasticidade, mas a autocorrelação residual recomenda cautela na interpretação das estimativas.

Os testes de estabilidade obtidos pelo procedimento OLS-CUSUM mostram que as equações estimadas para Minas Gerais, Rio de Janeiro e São Paulo permanecem dentro das bandas críticas. Isso indica ausência de instabilidade estrutural relevante e parâmetros consistentes ao longo do tempo. Há oscilações mais acentuadas em RJ e SP em alguns trechos, mas sempre dentro dos limites, o que reforça que não ocorreram quebras estruturais significativas.

```
serial.test(modelo_var1, lags.pt = 12, type = "PT.asymptotic")
```

```
##
## Portmanteau Test (asymptotic)
##
```

```
## data: Residuals of VAR object modelo_var1
## Chi-squared = 161.62, df = 99, p-value = 7.233e-05
```

```
arch.test(modelo_var1, lags.multi = 12)
```

```
##
## ARCH (multivariate)
##
## data: Residuals of VAR object modelo_var1
## Chi-squared = 436.35, df = 432, p-value = 0.4325
```

```
normality.test(modelo_var1)
```

```
## $JB
##
## JB-Test (multivariate)
##
## data: Residuals of VAR object modelo_var1
## Chi-squared = 9.1797, df = 6, p-value = 0.1637
##
##
## $Skewness
##
## Skewness only (multivariate)
##
## data: Residuals of VAR object modelo_var1
## Chi-squared = 4.9892, df = 3, p-value = 0.1726
##
##
## $Kurtosis
##
## Kurtosis only (multivariate)
##
## data: Residuals of VAR object modelo_var1
## Chi-squared = 4.1906, df = 3, p-value = 0.2416
```

```
plot(stability(modelo_var1), nc = 2)
```

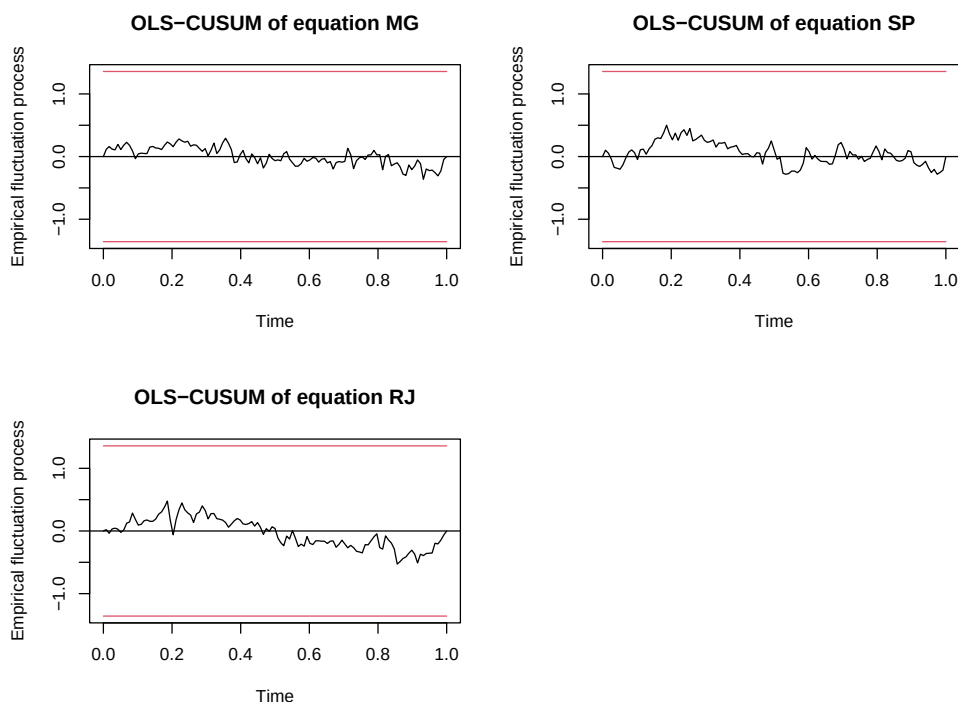


Figura 10: Testes de estabilidade.

Os resultados do VAR estimado em diferenças logarítmicas das séries estaduais mostram padrões distintos de persistência e de interação entre os estados. Na equação de Minas Gerais, o coeficiente defasado da própria variável (MG_{t-1}) é negativo, estatisticamente significativo e de magnitude elevada (-0,48), indicando forte efeito de correção. Em outras palavras, aumentos observados no mês anterior tendem a ser revertidos no mês seguinte, aproximando a série de sua média. As defasagens de RJ e SP não são significativas, o que confirma que a dinâmica mensal de Minas é essencialmente interna.

No caso do Rio de Janeiro, nenhum coeficiente apresenta significância estatística. Isso indica que as variações atuais não têm relação sistemática com valores defasados do próprio estado nem dos demais. Na prática, a série de RJ se comporta de forma mais volátil e idiossincrática, sem evidências de memória estatística no curto prazo.

Em São Paulo, dois pontos merecem destaque. O coeficiente defasado da própria série (SP_{t-1}) é negativo e significativo (-0,22), novamente sugerindo reversão parcial após picos. Além disso, a defasagem de RJ (RJ_{t-1}) tem efeito negativo e significativo (-0,15), indicando que aumentos no Rio de Janeiro em um mês estão associados a reduções em São Paulo no mês seguinte. Esse efeito interestadual é estatisticamente consistente e sugere algum grau de interdependência entre as duas maiores economias do país, ainda que com sinal negativo.

4 Considerações finais

O pacote **BrazilCrime** foi desenvolvido com o objetivo de facilitar o acesso aos dados criminais do SINESP e facilitar análises empíricas em segurança pública utilizando a linguagem R. Ao oferecer funções simples, padronizadas e automatizadas para coleta, organização, visualização e previsão de séries temporais criminais, o pacote reduz barreiras técnicas que frequentemente dificultam a entrada de novos pesquisadores na área.

Os exemplos que foram apresentados ao longo do texto mostram como o pacote pode ser integrado a ferramentas consolidadas para gerar visualizações, análises preditivas e modelos multivariados. Assim, o pacote contribui com pesquisadores, gestores públicos e instituições interessadas em compreender a dinâmica da criminalidade no Brasil.

Por fim, é importante enfatizar que o pacote **BrazilCrime** apenas extrai os dados e facilita sua manipulação na linguagem R, não realizando qualquer tipo de validação, correção ou interpretação adicional. Assim, seu único objetivo é facilitar pesquisas empíricas, oferecendo uma interface prática para organização e análise dos dados já disponibilizados publicamente pelo sistema. Todos os créditos dos dados pertencem ao SINESP, plataforma de informações desenvolvida e mantida pela Secretaria Nacional de Segurança Pública, SENASP, do Ministério da Justiça e Segurança Pública.

Agradecimentos

Marcelo Justus agradece ao CNPq pela Bolsa de Produtividade em Pesquisa vinculada ao projeto “Contribuições à Pesquisa em Segurança Pública com Dados do Brasil” (processo nº 312685/2021-1), a partir do qual o pacote **BrazilCrime** foi concebido. Giovanni Vargette agradece ao CNPq pelas bolsas de Iniciação Científica recebidas entre 2023 e 2025. Bernardo Geraldini agradece ao CNPq pela bolsa de Doutorado (processo nº 142249/2024-6). Os autores agradecem aos colegas que contribuíram com sugestões de melhorias e novas funcionalidades do pacote.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Brasil, CAPES, Código de Financiamento 001.

Referências

- Andresen, M. A., & Hodgkinson, T. (2020). Somehow I always end up alone: COVID-19, social isolation and crime in Queensland, Australia. *Crime Science*, 9(25), 1–20. <https://doi.org/10.1186/s40163-020-00135-4>
- Ashby, M. P. J. (2018). Studying crime and place with the crime open database. *Research Data Journal for the Humanities and Social Sciences*, 4(1), 65–80. <https://doi.org/10.1163/24523666-00401007>
- Ashby, M. P. J. (2020a). Changes in police calls for service during the early months of the 2020 coronavirus pandemic. *Policing: A Journal of Policy and Practice*, 14(4), 1054–1072. <https://doi.org/10.1093/police/paaa037>
- Ashby, M. P. J. (2020b). Initial evidence on the relationship between the coronavirus pandemic and crime in the United States. *Crime Science*, 9(6), 1–16. <https://doi.org/10.1186/s40163-020-00117-6>
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2016). *Time series analysis: Forecasting and control* (5th ed.). Wiley. ISBN 978-1-118-67502-1
- Campedelli, G. M., Aziani, A., & Favarin, S. (2020). Exploring the immediate effects of COVID-19 containment policies on crime: An empirical analysis of the short-term aftermath in Los Angeles. *American Journal of Criminal Justice*, 1–24. <https://doi.org/10.1007/s12103-020-09578-6>
- Dai, M., Xia, Y., & Han, R. (2021). The impact of lockdown on police service calls during the COVID-19 pandemic in China. *Policing(OxF)*, 1–15. <https://doi.org/10.1093/police/paab007>
- Hodgkinson, T., & Andresen, M. A. (2020). Show me a man or a woman alone and I'll show you a saint: Changes in the frequency of criminal incidents during the COVID-19 pandemic. *Journal of Criminal Justice*, 69, 101706. <https://doi.org/10.1016/j.jcrimjus.2020.101706>
- Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E., & Yasmeeen, F. (2025). *forecast: Forecasting functions for time series and linear models* (R package version 8.24.0) [Computer software]. <https://pkg.robjhyndman.com/forecast/>
- Kemp, S., Buil-Gil, D., Moneva, A., Miró-Llinares, F., & Díaz-Castaño, N. (2020). Empty streets, busy internet: A time series analysis of cybercrime and fraud trends during COVID-19. <https://doi.org/10.31235/osf.io/38wfy>
- Kim, D.-Y., & Phillips, S. W. (2021). When COVID-19 and guns meet: A rise in shootings. *Journal of Criminal Justice*, 73, 101783. <https://doi.org/10.1016/j.jcrimjus.2021.101783>
- Laltuf, I. (2023). *ISPDATA: The package to access public security data from the State of Rio de Janeiro* [Computer software]. GitHub. <https://github.com/igorlaltuf/ispdata>
- Langton, S., Dixon, A., & Farrell, G. (2021). Six months in: Pandemic crime trends in England and Wales. *Crime Science*, 10(6), 1–16. <https://doi.org/10.1186/s40163-021-00142-z>
- Mohler, G., Bertozzi, A. L., Carter, J., Short, M. B., Sledge, D., Tita, G. E., Uchida, C. D., & Brantingham, P. J. (2020). Impact of social distancing during COVID-19 pandemic on crime in Los Angeles and Indianapolis. *Journal of Criminal Justice*, 68, 101692. <https://doi.org/10.1016/j.jcrimjus.2020.101692>
- Odell, E., & Tierney, N. (2019). *ukpolice: Download data on UK police and crime* (R package version 0.2.2) [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.2619537>

Payne, J., & Morgan, A. (2020b). Property crime during the COVID-19 pandemic: A comparison of recorded offence rates and dynamic forecasts (ARIMA) for March 2020 in Queensland, Australia. <https://doi.org/10.31219/osf.io/de9nc>

Shen, Y., Fu, R., & Haruko, N. (2021). COVID-19's lockdown and crime victimization: The state of emergency under the Abe administration. *Asian Economic Policy Review*, 16(1), 1–22. <https://doi.org/10.1111/aepr.12339>

Sutherland, M., McKenney, M., & Elkbuli, A. (2021). Gun violence during COVID-19 pandemic: Paradoxical trends in New York City, Chicago, Los Angeles and Baltimore. *American Journal of Emergency Medicine*, 39(1), 225–226. <https://doi.org/10.1016/j.ajem.2020.05.006>

Vargette, G., Laltuf, I., & Justus, M. (2024). *BrazilCrime: Accesses Brazilian public security data from SINESP since 2015* (R package version 0.2.1) [Computer software]. CRAN. <https://CRAN.R-project.org/package=BrazilCrime>

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag. <https://ggplot2.tidyverse.org>

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2023). *dplyr: A grammar of data manipulation* (R package version 1.1.4) [Computer software]. CRAN. <https://CRAN.R-project.org/package=dplyr>