



NDSBR: UM PACOTE EM LINGUAGEM R PARA ANÁLISE DE DADOS NATURALÍSTICOS

Pedro Augusto Borges dos Santos¹; Jorge Tiago Bastos²

¹ Universidade Federal do Paraná, pedroaugusto@ufpr.br

² Universidade Federal do Paraná, jtbastos@ufpr.br

RESUMO

Este trabalho tem como objetivo apresentar o desenvolvimento do pacote *ndsbr* para a linguagem de programação R, cuja finalidade é estabelecer funções para carregar, transformar e analisar os dados do projeto de pesquisa Estudo Naturalístico de Direção Brasileiro. O projeto possibilitou a coleta de grandes volumes de dados comportamentais de um grupo de participantes, em cenários reais de condução no dia-a-dia. A amostra atual conta com mais de 9.400 km de distância percorrida e quase 382 horas de viagem. Assim, se fez necessário estabelecer um conjunto de funções para auxiliar a investigação de fatores de risco baseados nos dados naturalísticos. O R é uma linguagem de programação com foco em análise estatística e visualização de dados. Ela também proporciona uma facilidade de integração com outras tecnologias já existentes. Espera-se que a criação desse pacote auxilie na reprodutibilidade do processo de pesquisa científica, além de aprimorar a automatização dos processos e a economia de recursos humanos e computacionais.

PALAVRAS-CHAVE: Estudo Naturalístico de Direção, Segurança viária, Fatores de risco

1. INTRODUÇÃO

Estudos do comportamento humano com foco na segurança viária podem ser elaborados com base em múltiplos métodos, incluindo a elaboração de cenários de simulação e a observação do comportamento em ambientes reais. Um método observacional de investigação em condições do dia-a-dia é o Estudo Naturalístico de Direção (ou *Naturalistic Driving Study* - NDS), que permite o monitoramento de condutores dentro de seus próprios veículos, através de câmeras, sensores de GPS e medidores de aceleração (SHINAR, 2017). Em Curitiba, iniciou-se um projeto com a aplicação do NDS - o Estudo Naturalístico de Direção Brasileiro (NDS-BR) - em 2019 (CEPPUR, 2022).

As principais vantagens do NDS envolvem a imersão oferecida aos participantes e a coleta de grandes volumes de dados na composição de sua amostra (CARSTEN *et al.*, 2013). Porém, essa quantidade de dados presentes na amostra do NDS-BR exige o estabelecimento de métodos automatizados para lidar com o processamento e análise dos dados naturalísticos. Assim, foi criado um pacote para a linguagem de programação R com um conjunto de funções para auxiliar neste projeto de pesquisa: o *ndsbr*. O objetivo deste trabalho é apresentar métodos presentes no pacote. A linguagem R foi escolhida pela facilidade de integração com outras tecnologias já existentes, como bibliotecas de



sistemas de informação geográfica (SIG) (LOVELACE *et al.*, 2019), além de sua otimização para análise estatística, visualização de dados e sua natureza de código livre (R CORE TEAM, 2022).

A criação desse pacote é motivada pela importância da reprodutibilidade no processo da pesquisa científica, além da possível economia de recursos computacionais e humanos. Outros softwares de geocomputação, como o QGIS, e de edição de planilhas também podem ser utilizados para cumprir os métodos que serão apresentados a seguir. Porém, softwares com interface gráfica prejudicam a reprodutibilidade e transparência do processo de transformação dos dados naturalísticos de direção. O uso de uma linguagem de programação possibilita um processo em que os mesmos resultados de uma pesquisa possam ser alcançados por outras pessoas, a partir dos códigos e dados divulgados (PEBESMA *et al.*, 2012).

2. MATERIAIS E MÉTODOS

No NDS-BR foram coletadas imagens através de três câmeras e um sensor de GPS, com todos equipamentos sincronizados a uma frequência de 1 Hz. Uma câmera teve seu foco na parte interna do veículo, que incluiu o painel do veículo e o condutor. Duas câmeras tiveram seus focos para a parte externa frontal do veículo. Com o uso do GPS, foi possível registrar a velocidade praticada pelos condutores, além das coordenadas geográficas, data, horário, altitude e azimuth. Com o uso das câmeras foi possível verificar alguns comportamentos dos condutores, como o uso do telefone celular, a presença de carona e o uso do cinto de segurança. Todo o processo de coleta e transformação dos dados foi detalhado por Borguezani *et al.* (2020). Até o momento, o NDS-BR contém uma amostra de 1002 viagens, com quase 382 horas de viagem e mais de 9.400 km de distância percorrida. Cada viagem iniciava com a ligação do equipamento de coleta e se encerrava com o desligamento deste equipamento. O processo de ligar e desligar o equipamento foi responsabilidade dos participantes.

Os dados coletados foram transformados em arquivos de texto no formato csv. A amostra contém 34 atributos, com mais de 1,3 milhão de entradas. Cada entrada representa um ponto do GPS. A Tabela 1 apresenta uma lista reduzida com os principais atributos da base de dados do NDS-BR. Em relação ao tipo de dado, os dados primários são aqueles obtidos diretamente pelo projeto de pesquisa e os secundários foram obtidos em outras fontes e inseridos no banco de dados através de uma união espacial.



Tabela 1 – Principais atributos da amostra do NDS-BR

Tipo de Dado	Equipamento de Coleta	Atributos	Fonte
Primário	GPS	• Coordenadas (latitude, longitude, altitude); • Data; • Hora; • Velocidade; • Aceleração.	NDS-BR
Primário	Câmeras	• Tempo válido de viagem; • Uso do cinto de segurança; • Uso do telefone celular (quantidade e tipos de uso); • Presença de carona.	NDS-BR
Secundário	---	• Município; • Bairro; • Rua; • Hierarquia viária; • Limites de velocidade.	IPPUC (2021) e OpenStreetMap (2021)

Fonte: Os autores (2022)

No R, o pacote é a unidade fundamental para o compartilhamento de código. Um pacote reúne um conjunto de códigos, dados, documentação e testes. Sua principal vantagem é a facilidade de compartilhamento com outros usuários (WICKHAM, 2015). Em relação às soluções similares já desenvolvidas para analisar dados naturalísticos, Barnwal *et al.* (2021) criaram o pacote `ndsintxn` para investigar fatores de risco em interseções, utilizando dados naturalísticos como entrada. Assim como o pacote `ndsintxn`, o `ndsbr` é uma solução personalizada para as necessidades de seu projeto de pesquisa. Na Tabela 2 está presente a relação de funções incluídas no `ndsbr`. Atualmente elas estão organizadas em três grupos: carregamento, transformação e análise. Esses grupos foram definidos de forma subjetiva, levando em consideração os principais passos que ocorrem no processo de investigação dos dados do projeto.

Tabela 2 – Funções presentes no pacote `ndsbr`

Grupo	Função	Descrição
Carregamento	<code>nds_load_data()</code>	Carrega os dados naturalísticos presentes nos arquivos csv.
	<code>nds_download_sf()</code>	Faz o download de dados espaciais e carrega no ambiente do projeto.
Transformação	<code>nds_create_points()</code>	Cria um objeto espacial com geometria de pontos.
	<code>nds_create_lines()</code>	Cria um objeto espacial com geometria de linhas.
Análise	<code>nds_calc_time()</code>	Calcula o tempo de viagem, classificado por uma variável de interesse.
	<code>nds_calc_dist()</code>	Calcula a distância de viagem, classificada por uma variável de interesse.
	<code>nds_calc_speeding()</code>	Calcula a taxa de excesso de velocidade.

Fonte: Os autores (2022)



Em relação ao carregamento de dados, a função `nds_load_data` carrega os dados naturalísticos no ambiente do R e a função `nds_download_sf` auxilia no download de outros dados espaciais pertinentes para a análise (*e.g.*, eixo de vias, zonas de tráfego, limites municipais). As funções do grupo transformação tem como objetivo transformar os dados de entrada em objetos espaciais. As funções `nds_create_points` e `nds_create_lines` criam pontos e linhas, respectivamente, com base nas coordenadas geográficas. As funções de análise possibilitam o cálculo do tempo (`nds_calc_time`) e distância (`nds_calc_dist`) de viagem, classificados por outra variável de interesse (*e.g.*, condutor, viagem, hora do dia). Por fim, a função `nds_calc_speeding` calcula a taxa de excesso de velocidade, comparando a velocidade praticada e o limite de velocidade da via, já presentes na amostra de dados naturalísticos. Todas as funções foram escritas com base em funções de outros pacotes já estabelecidos para a manipulação de dados tabulares e espaciais: o `dplyr` (WICKHAM *et al.*, 2022), o `tidyr` (WICKHAM e GIRLICH, 2022) e o `sf` (PEBESMA, 2018).

O desenvolvimento dessas funções teve como inspiração as metodologias de pesquisas anteriores que utilizaram os dados naturalísticos do NDS-BR para investigar comportamentos de risco na segurança viária. Como exemplo, Bastos *et al.* (2021) e Suguinoshita *et al.* (2020) calcularam taxas de excesso de velocidade e Bastos *et al.* (2020) calcularam o tempo de uso do telefone celular.

3. EXEMPLO DE APLICAÇÃO

A fim de mostrar o funcionamento do pacote, a Figura 1 inclui um exemplo de aplicação. O objetivo deste exemplo é calcular a taxa de excesso de velocidade com base na distância percorrida, para cada classe de hierarquia viária estabelecida pelo Código de Trânsito Brasileiro (BRASIL, 1997). Os dados de hierarquia já estão incluídos na amostra do NDS-BR. O primeiro passo do processo é carregar os dados naturalísticos, utilizando a função `nds_load_data`. O pacote já inclui uma amostra reduzida de dados naturalísticos, a fim de servir de exemplo para os usuários que não possuem acesso ao conjunto completo de dados do NDS-BR. Essa amostra de exemplo contém 9 viagens de 3 condutores distintos.

O próximo passo é criar um objeto espacial de linhas a partir dos dados naturalísticos, utilizando a função `nds_create_lines`. Por fim, com a aplicação da

função `nds_calc_speeding` é possível calcular a taxa de excesso de velocidade (SP) por hierarquia de via, estabelecendo dois limiares: a velocidade acima do limite da via (parâmetro `sp`) e a velocidade abaixo do limite para ser considerado como velocidade de fluxo livre (parâmetro `exp`). As vias de trânsito rápido (hierarquia “1”) apresentaram a maior taxa de excesso, com quase 66% da distância percorrida em excesso de velocidade. As arteriais apresentaram menor taxa, com quase 20%. Locais da amostra em que não foi possível identificar a hierarquia viária foram definidos como “NPI”.

```
# devtools::install_github("pabsantos/ndsbr")
library(ndsbr) # Carrega o pacote

# Carrega os dados naturalísticos
dados_nds <- nds_load_data(
  pattern = "driver",
  folder = system.file("extdata", package = "ndsbr")
)

# Cria um objeto espacial de linhas
linhas_nds <- nds_create_lines(data = dados_nds, x = LONG, y = LAT)

# Calcula o excesso de velocidade com base nas distancias percorridas
nds_calc_speeding(
  data = linhas_nds, type = "distance", by = HIERARQUIA_CTB,
  spd = 5, exp = 10, percentage = FALSE
)

# A tibble: 5 × 2
# Groups:   HIERARQUIA_CTB [5]
  HIERARQUIA_CTB    SP
  <chr>          <dbl>
1 1              0.657
2 2              0.196
3 3              0.355
4 4              0.517
5 NPI            0.586
```

Figura 1 - Script com a aplicação de funções do ndsbr
Fonte: Os autores (2022)

4. CONCLUSÕES

O presente trabalho teve como objetivo apresentar o pacote `ndsbr` para a linguagem de programação R. A principal motivação em criar esses métodos foi auxiliar na pesquisa científica que envolve os dados naturalísticos do projeto NDS-BR, melhorando a reprodutibilidade e agilidade nos processos de análise, além de oferecer um conjunto de ferramentas gratuitas e de código livre. Uma limitação dessa ferramenta é a barreira técnica que o uso de uma linguagem de programação pode trazer para novos usuários. A



possível solução seria o desenvolvimento de uma interface gráfica ou aplicativos da web com a utilização do pacote shiny (CHANG *et al.*, 2022).

O pacote *ndsbr* se apresenta como uma ferramenta de auxílio no processo de análise de dados naturalísticos para pesquisas acadêmicas, além do desenvolvimento de diagnósticos e prognósticos relacionados à gestão da segurança viária em planos de mobilidade e demais legislações. A ferramenta facilita a análise espacial do excesso de velocidade e a análise de perfis de velocidade, com caracterização por hora, dia e hierarquia viária.

O *ndsbr* atualmente se encontra na versão 0.1.5 e pode ser instalada a partir de seu repositório (SANTOS, 2022). O desenvolvimento do pacote envolve um processo contínuo, em que novas funções vão sendo adicionadas conforme a necessidade de seus usuários. Futuras versões devem incluir funções para auxiliar na análise de outros fatores de risco na segurança viária (*e.g.*, uso do celular), na análise e recorte dos dados de vídeo e na exportação de elementos gráficos (*e.g.*, perfis de velocidade, mapas e análises estatísticas). Em adição, o pacote ainda não possui funções para importar e unir os dados dos participantes (idade, sexo, experiência de condução, etc.) e de seus automóveis (ano do modelo, potência, tipo de câmbio, etc.) com a amostra completa de dados naturalísticos, limitando a análise de fatores comportamentais e veiculares.

REFERÊNCIAS

BARNWAL, Ashirwad; MERICKEL, Jennifer; RIZZO, Matthew; *et al.* *ndsintxn: An R Package for Extracting Information from Naturalistic Driving Data to Support Driver Behavior Analyses at Intersections*. 2021. Disponível em: <<http://arxiv.org/abs/2108.04346>>. Acesso em: 27 jun. 2022.

BASTOS, Jorge Tiago; DOS SANTOS, Pedro Augusto B.; AMANCIO, Eduardo Cesar; *et al.* Is organized carpooling safer? Speeding and distracted driving behaviors from a naturalistic driving study in Brazil. **Accident Analysis & Prevention**, v. 152, p. 105992, 2021.

BASTOS, Jorge Tiago; DOS SANTOS, Pedro Augusto B.; AMANCIO, Eduardo Cesar; *et al.* Naturalistic Driving Study in Brazil: An Analysis of Mobile Phone Use Behavior While Driving. **International Journal of Environmental Research and Public Health**, v. 17, n. 17, p. 6412, 2020.

BORGUEZANI, Jackson Rossi; DOS SANTOS, Pedro Augusto B.; OSÓRIO, Fernando dos Santos; *et al.* Plataforma de coleta de dados naturalísticos de segurança viária. In: **34º congresso de pesquisa e ensino em transporte da ANPET**. [s.l.: s.n.], 2020, p. 2610–



2617. Disponível em:
<http://www.anpet.org.br/anais34/documentos/2020/Tr%C3%A1fego%20Urbano%20e%20Rodovi%C3%A1rio/Seguran%C3%A7a%20Vi%C3%A1ria%20III/5_332_CT.pdf>.

BRASIL. Lei nº 9.503, de 23 de setembro de 1997. **Código de trânsito Brasileiro**, 1997.

CARSTEN, Oliver; KIRCHER, Katja; JAMSON, Samantha. Vehicle-based studies of driving in the real world: The hard truth? **Accident Analysis & Prevention**, v. 58, p. 162–174, 2013.

CEPPUR. Estudo Naturalístico de Direção Brasileiro. Disponível em:
<<http://www.tecnologia.ufpr.br/portal/ceppur/estudo-naturalistico-de-direcao-brasileiro/>>. Acesso em: 17 jan. 2022.

CHANG, W.; CHENG, J.; ALLAIRE, J. J.; et al. **shiny: Web Application Framework for R**. 2022.

IPPUC. Dados Geográficos. Disponível em:
<<https://ippuc.org.br/geodownloads/geo.htm>>. 2021. Acesso em: 10 jul. 2021.

LOVELACE, Robin; NOWOSAD, Jakub; MÜNCHOW, Jannes. **Geocomputation with R**. Boca Raton: CRC Press, Taylor and Francis Group, CRC Press is an imprint of the Taylor and Francis Group, an informa Business, A Chapman & Hall Book, 2019.

OPENSTREETMAP. Planet dump retrieved from <https://planet.osm.org>. 2021. Disponível em: <<https://www.openstreetmap.org>>.

PEBESMA, Edzer. Simple features for r: Standardized support for spatial vector data. **The R Journal**, v. 10, n. 1, p. 439–446, 2018.

PEBESMA, Edzer; NÜST, Daniel; BIVAND, Roger. The R software environment in reproducible geoscientific research. **Eos, Transactions American Geophysical Union**, v. 93, n. 16, p. 163–163, 2012.

SANTOS, Pedro Augusto B. ndsbr: Load, Analyze and Manipulate Data of The Brazilian Naturalistic Driving Study. **Repositório do GitHub**, 2022. Disponível em:
<<https://github.com/pabsantos/ndsbr>>.

SHINAR, D. **Traffic Safety and Human Behavior**. Second edi. Bingley: Emerald Group Publishing Limited, 2017.

SUGUINOSHITA, Maria Clara; VALEIXO, Gabriela Rolim Ribeiro; SANTOS, Pedro Augusto Borges dos; et al. FATORES DETERMINANTES PARA O EXCESSO DE VELOCIDADE EM VIAS ARTERIAIS URBANAS. In: [s.l.: s.n.], 2020, p. 63–74. Disponível em: <<https://eventos.ufpr.br/stpr/2stpr/paper/view/4449/954>>. Acesso em: 5 abr. 2022.

WICKHAM, Hadley. **R packages**. First edition. Sebastopol, CA: O'Reilly Media, 2015.



WICKHAM, Hadley; FRANÇOIS, Romain; HENRY, Lionel; *et al.* **dplyr: A grammar of data manipulation.** [s.l.: s.n.], 2022. Disponível em: <<https://CRAN.R-project.org/package=dplyr>>.

WICKHAM, Hadley; GIRLICH, Maximilian. **tidyr: Tidy messy data.** [s.l.: s.n.], 2022. Disponível em: <<https://CRAN.R-project.org/package=tidyr>>

.