

# Acidentes nas Rodovias Federais Brasileiras

Gleiciano Sousa

2022-05-30

## Contents

Introdução	1
Licença	1
Objetivos	1
Código	1

---

## Introdução

Este documento consta da análise exploratória de dados inicial e visualização dos dados de acidentes nas rodovias federais brasileiras (aka BR). Os dados foram coletados no site oficial da Polícia Rodoviária Federal do Brasil (PRF). A análise foi realizada utilizando-se a linguagem R.

## Licença

“Dados Abertos são dados institucionais, disponibilizados em formato legível por máquina e sem restrição de licenças, patentes ou mecanismos de controle, que qualquer pessoa pode livremente usá-los, reutilizá-los e redistribuí-los. Os dados classificados como abertos podem ser utilizados de várias formas, seja pelo próprio governo ou pela sociedade, como, por exemplo, no desenvolvimento de aplicativos, que exibem informações de forma gráfica e interativa.” Trecho extraído do site oficial.

## Objetivos

- Analisar os dados abertos afim de encontrar possíveis padrões nos acidentes ocorridos nas BRs;
  - Desenvolver visualizações que facilitem o entedimento da análise.
- 

## Código

### Carregando os Pacotes

```
# Loading Packages
library(readr)
library(tidyverse)
library(lubridate)
```

```
library(ggplot2)
library(treemapify)
library(viridis)
```

## Importando as Tabelas

Foi preciso indicar o encoding para o idioma brasileiro, caso contrário teríamos problemas com os acentos das tabelas.

```
# Importing Datasets
## Set your working directory before importing CSVs
datatran2017 <- read_delim("datatran2017.csv", delim = ";",
                          locale = locale(date_names = "pt",
                                           decimal_mark = ",",
                                           grouping_mark = ".",
                                           encoding = "ISO-8859-1"))

datatran2018 <- read_delim("datatran2018.csv", delim = ";",
                          locale = locale(date_names = "pt",
                                           decimal_mark = ",",
                                           grouping_mark = ".",
                                           encoding = "ISO-8859-1"))

datatran2019 <- read_delim("datatran2019.csv", delim = ";",
                          locale = locale(date_names = "pt",
                                           decimal_mark = ",",
                                           grouping_mark = ".",
                                           encoding = "ISO-8859-1"))

datatran2020 <- read_delim("datatran2020.csv", delim = ";",
                          locale = locale(date_names = "pt",
                                           decimal_mark = ",",
                                           grouping_mark = ".",
                                           encoding = "ISO-8859-1"))

datatran2021 <- read_delim("datatran2021.csv", delim = ";",
                          locale = locale(date_names = "pt",
                                           decimal_mark = ",",
                                           grouping_mark = ".",
                                           encoding = "ISO-8859-1"))

datatran2022 <- read_delim("datatran2022.csv", delim = ";",
                          locale = locale(date_names = "pt",
                                           decimal_mark = ",",
                                           grouping_mark = ".",
                                           encoding = "ISO-8859-1"))
```

## Preparando e Combinando as Tabelas

A tabela de 2017 foi filtrada para que os dados não ficassem desproporcionais, ou seja, janeiro, fevereiro, março e abril teriam seis tabelas, enquanto que os restantes teriam uma tabela a menos, visto que na tabela de 2022 ainda não constam os meses que virão.

Também havia uma inconsistência no tipo de dado da coluna “latitude”, foi necessário igualar os tipos para poder combinar as tabelas.

```
# Preparing to Combine All Datasets
datatran2017_f <- filter(datatran2017, month(datatran2017$data_inversa) > 4)

datatran2021$latitude <- as.double(datatran2021$latitude)

datatran2022$latitude <- as.double(datatran2022$latitude)

all_datatran <- bind_rows(datatran2017_f, datatran2018, datatran2019,
                        datatran2020, datatran2021, datatran2022)
```

## Formatando e Organizando Colunas

Foram criadas as colunas de “mes” e “ano” extraíndo da coluna “data\_inversa” e realocadas para próximo da data

```
# Formatting and Organizing Columns
all_datatran$mes <- month(all_datatran$data_inversa,
                        label = TRUE, abbr = TRUE)

all_datatran$ano <- year(all_datatran$data_inversa)

all_datatran <- all_datatran %>%
  relocate(mes, .after = data_inversa) %>%
  relocate(ano, .after = data_inversa)
```

## Análise Exploratória dos Dados (EDA)

Foi realizada a contagem do número de acidentes por mês, ano e unidade federativa, além disso também foi obtida a média de acidentes por mês.

```
# Exploratory Data Analysis
acidentes_mes <- all_datatran %>%
  count(mes, name = "acidentes", sort = TRUE) %>%
  mutate(media = acidentes/5)
tibble(acidentes_mes)
```

```
## # A tibble: 12 x 3
##   mes      acidentes media
##   <ord>      <int> <dbl>
## 1 dez         33240 6648
## 2 out         30591 6118.
## 3 set         29573 5915.
## 4 ago         29493 5899.
```

```
## 5 jul      29483 5897.
## 6 nov      29077 5815.
## 7 jun      28713 5743.
## 8 mai      28293 5659.
## 9 jan      28052 5610.
## 10 mar     26961 5392.
## 11 fev     26407 5281.
## 12 abr     24984 4997.
```

```
acidentes_ano_mes <- all_datatran %>%
  count(ano, mes, name = "acidentes", sort = TRUE)
tibble(acidentes_ano_mes)
```

```
## # A tibble: 60 x 3
##   ano mes      acidentes
##   <dbl> <ord>      <int>
## 1 2017 dez        8623
## 2 2017 mai        7756
## 3 2017 jun        7517
## 4 2017 out        7391
## 5 2017 jul        7319
## 6 2017 set        7134
## 7 2017 ago        7085
## 8 2017 nov        6973
## 9 2018 jan        6940
## 10 2018 mar        6684
## # ... with 50 more rows
```

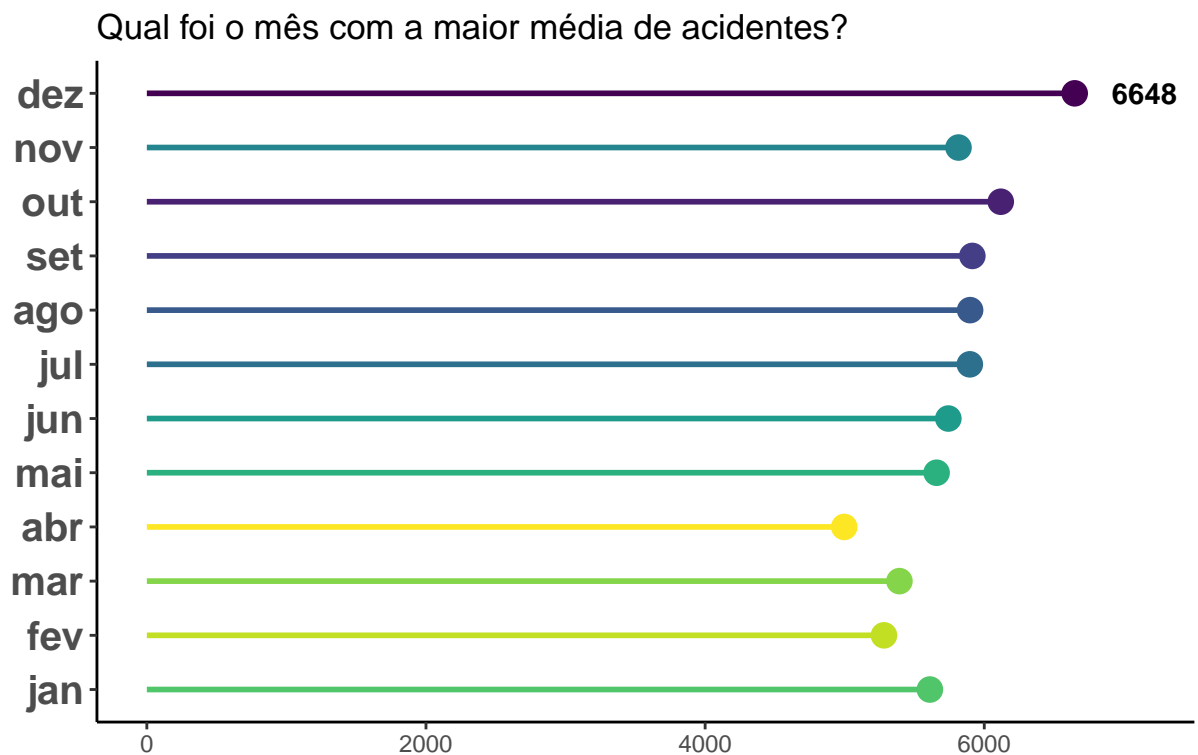
```
acidentes_uf <- all_datatran %>%
  count(uf, name = "acidentes", sort = TRUE)
tibble(acidentes_uf)
```

```
## # A tibble: 27 x 2
##   uf      acidentes
##   <chr>      <int>
## 1 MG        45428
## 2 SC        41782
## 3 PR        39605
## 4 RS        23538
## 5 RJ        23427
## 6 SP        22667
## 7 BA        17964
## 8 GO        17055
## 9 PE        13698
## 10 ES        13335
## # ... with 17 more rows
```

## Visualização 01

Na visualização abaixo é possível observar uma maior média de acidentes no mes de dezembro, isso pode ter ocorrido devido as festas de final de ano (natal e réveillon) onde geralmente as famílias viajam para se reunir com os familiares e se divertir com amigos aumentando assim o número de veículos circulando e a probabilidade de acidentes.

```
# Visualization 01
ggplot(acidentes_mes, aes(x = mes, y = media)) +
  geom_segment(aes(x = mes, xend = mes, y = 0, yend = media),
              color = viridis(12), size = 1) +
  geom_point(size = 4, color = viridis(12)) +
  coord_flip() +
  annotate("text", x = 12, y = 6648 + 500, label = "6648",
          color = "black", size = 4, fontface = "bold") +
  labs(title = "Qual foi o mês com a maior média de acidentes?",
       x = "", y = "", caption = paste(
         "Fonte: Polícia Rodoviária Federal, Brasil", "\n",
         "Período: mai/2017 a abr/2022", "\n",
         "Última atualização:", today())) +
  theme_classic() +
  theme(axis.text.y = element_text(size = 15, face = "bold"),
        plot.subtitle = element_text(size = 8),
        plot.caption = element_text(size = 5))
```



Fonte: Polícia Rodoviária Federal, Brasil  
 Período: mai/2017 a abr/2022  
 Última atualização: 2022-05-30

## Visualização 02

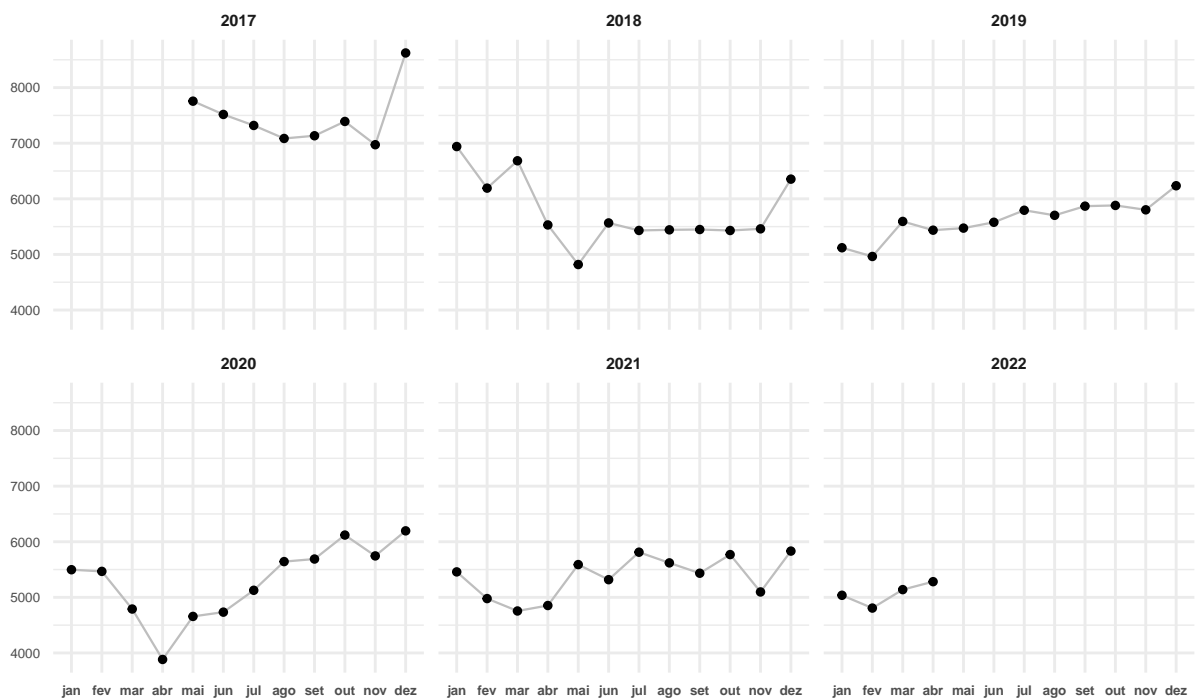
Já a visualização seguinte mostra a evolução no número de acidentes durante os meses e agrupados por ano. Aqui é possível perceber claramente o que foi visto no gráfico anterior, uma alta de acidentes ao chegar no mês de dezembro, tendo o dezembro de 2017 como o pico de acidentes.

Após o ano de 2017 observa-se uma redução no número de acidentes embora ainda é visto a alta de dezembro em cada um dos anos seguintes.

Outro ponto a sinalizar no gráfico é a baixa de acidentes no mês de abril de 2020, isso provavelmente aconteceu por ter sido o início das medidas de isolamento devido a pandemia da Covid19 no Brasil.

```
# Visualization 02
ggplot(acidentes_ano_mes, aes(x = mes, y = acidentes, group = ano)) +
  geom_line(color = "grey", size = 0.4) +
  geom_point(color = "black", size = 1) +
  facet_wrap(~ano) +
  labs(title = "Evolução no número de acidentes",
       x = "", y = "", caption = paste(
         "Fonte: Polícia Rodoviária Federal, Brasil", "\n",
         "Período: mai/2017 a abr/2022", "\n",
         "Última atualização:", today())) +
  theme_minimal() +
  theme(axis.text.x = element_text(size = 5, face = "bold"),
        axis.text.y = element_text(size = 5),
        strip.text = element_text(size = 6, face = "bold"),
        plot.subtitle = element_text(size = 8),
        plot.caption = element_text(size = 5))
```

## Evolução no número de acidentes



Fonte: Polícia Rodoviária Federal, Brasil  
 Período: mai/2017 a abr/2022  
 Última atualização: 2022-05-30

### Visualização 03

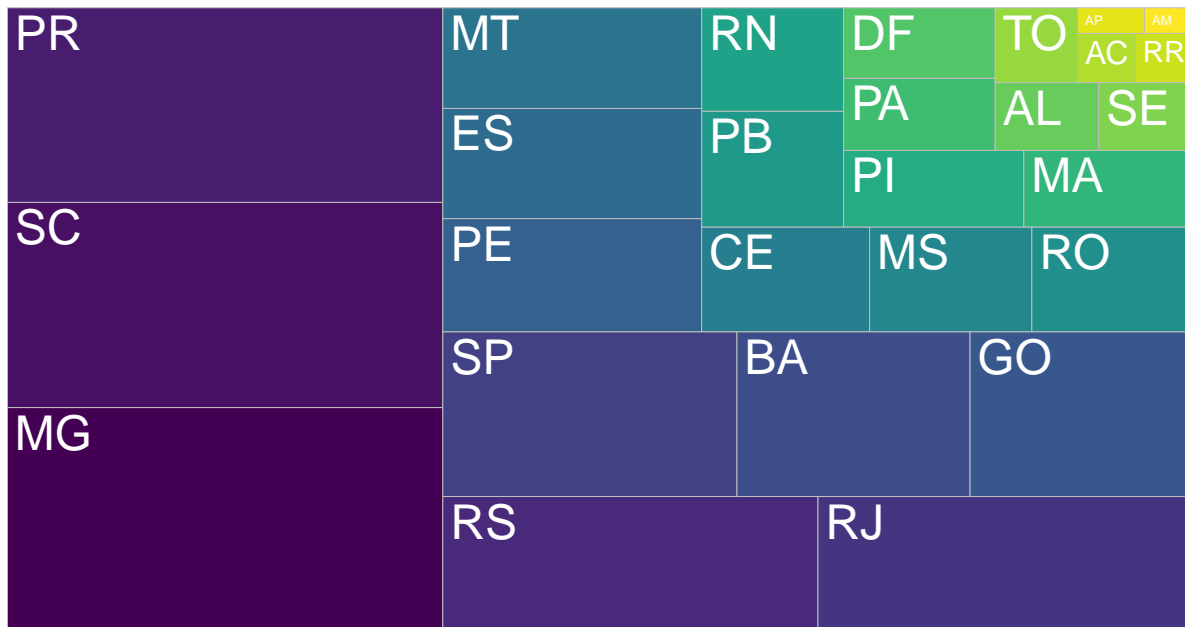
Para finalizar, o próximo gráfico mostra a distribuição no número de acidentes entre as unidades federativas do Brasil. É possível perceber dois grandes grupos por número de acidentes. O primeiro grupo liderado pelo estado de Minas Gerais, e que conta com Santa Catarina e Paraná se encontram com os maiores números de acidentes na faixa dos 40 mil. Já o segundo grupo conta com o Rio Grande do Sul, Rio de Janeiro e São Paulo com cerca de 20 mil acidentes.

Como foi observado neste gráfico, os estados da região sul e sudeste do Brasil (com exceção do Espírito Santo) brigam pelo pódio de estados com mais acidentes em rodovias federais, visto que são as regiões mais populosas do país.

```
# Visualization 03
ggplot(acidentes_uf, aes(area = acidentes, fill = uf)) +
  geom_treemap(show.legend = FALSE, fill = viridis(27)) +
  geom_treemap_text(label = acidentes_uf$uf, color = "white") +
  labs(title = "Quais os estados com mais acidentes?",
       subtitle = "Nas rodovias federais brasileiras",
       x = "", y = "", caption = paste(
         "Fonte: Polícia Rodoviária Federal, Brasil", "\n",
         "Período: mai/2017 a abr/2022", "\n",
         "Última atualização:", today()))
```

Quais os estados com mais acidentes?

Nas rodovias federais brasileiras



Fonte: Polícia Rodoviária Federal, Brasil  
Período: mai/2017 a abr/2022  
Última atualização: 2022-05-30

## **Conclusão**

Diante do que foi encontrado, é necessário que sejam desenvolvidos programas de prevenção, sobretudo nas regiões sul e sudeste do Brasil, anteriormente ao mês de dezembro, afim de educar a população e reduzir o número de acidentes.

Em conjunto com esses programas e aliado a equipe de dados da PRF, pode ser possível a criação de um sistema de recompensas para bons condutores, podendo ganhar descontos no IPVA, por exemplo.

O contínuo mapeamento dos acidentes nas rodovias é indispensável para que a PRF possa continuar garantindo um bom atendimento em casos de acidente.

---

Fonte: Polícia Rodoviária Federal, Brasil

Período: mai/2017 a abr/2022

Última atualização: 27-05-2022