

## load\_vaccine\_data

OOBr

2025-11-07

### Load vaccination microdata from the “basedosdados”

```
# Load vaccination microdata from the "basedosdados" project, specifically the  
# COVID-19 vaccination dataset from the Brazilian Ministry of Health, considering  
# general population and pregnant and postpartum women.
```

```
# Important packages  
library(basedosdados)  
library(dplyr)  
library(questionr)  
require("data.table")  
require("tidyverse")  
require("readxl")  
require("writexl")
```

```
memory.limit(999999)
```

```
# set_billing_id("token")
```

```
# Load vaccination microdata from the "basedosdados" project  
oi <- bdplyr("basedosdados.br_ms_vacinacao_covid19.microdados") %>%  
  select(sigla_uf, data_aplicacao_vacina, dose_vacina) %>%  
  group_by(sigla_uf, data_aplicacao_vacina, dose_vacina) %>%  
  summarise(n_casos = n()) %>%  
  bd_collect()
```

```
# Set vaccine doses
```

```
oi1 <- oi %>%  
  mutate(  
    num_dose = case_when(  
      dose_vacina == "1ª Dose" ~ "1a dose",  
      dose_vacina == "1ª Dose Revacinação" ~ "1o reforco",  
      dose_vacina == "1º Reforço" ~ "1o reforco",  
      dose_vacina == "2ª Dose" ~ "2a dose ou unica",  
      dose_vacina == "2ª Dose Revacinação" ~ "2o reforco",  
      dose_vacina == "2º Reforço" ~ "2o reforco",  
      dose_vacina == "3ª Dose" ~ "1o reforco",  
      dose_vacina == "3ª Dose Revacinação" ~ "1o reforco",  
      dose_vacina == "3º Reforço" ~ "1o reforco",  
      dose_vacina == "4ª Dose" ~ "2o reforco",  
      dose_vacina == "Dose Adicional" ~ "1o reforco",
```

```

    dose_vacina == "Dose Inicial" ~ "1a dose",
    dose_vacina == "Dose Única" ~ "2a dose ou unica",
    dose_vacina == "Reforço" ~ "1o reforco",
    dose_vacina == "Revacinação" ~ "1o reforco",
    dose_vacina == "Única" ~ "2a dose ou unica"
  )
)

oi2 <- oi1 %>%
  select(-dose_vacina) %>%
  group_by(sigla_uf, data_aplicacao_vacina, num_dose) %>%
  summarise(n_casos = sum(n_casos)) %>%
  select(sigla_uf, data_aplicacao_vacina, dose_vacina = num_dose, n_casos) %>%
  filter(!is.na(dose_vacina))

# Brazilian population by (federative unity) vaccinated
write_csv(oi2, "Vacinados_porUF_01-02-22.csv")
write_xlsx(oi2, "Vacinados_porUF_01-02-22.xlsx")

dados_br <- oi2 %>%
  group_by(data_aplicacao_vacina, dose_vacina) %>%
  summarise(num_casos = sum(n_casos))

# Brazilian population vaccinated
write_csv(dados_br, "Vacinados_Brasil_01-02-22.csv")
write_xlsx(dados_br, "Vacinados_Brasil_01-02-22.xlsx")

```

## Now filtering for pregnant and postpartum women

```

# Load vaccination microdata from the "basedosdados" project
ai <- bdplyr("basedosdados.br_ms_vacinacao_covid19.microdados") %>%
  # Select only the variables needed for this analysis:
  # - sigla_uf: federative unit (state) abbreviation
  # - data_aplicacao_vacina: date of vaccine administration
  # - grupo_atendimento_vacina: target group code (e.g., pregnant, puerperal women)
  # - dose_vacina: textual description of the dose type
  select(sigla_uf, data_aplicacao_vacina, grupo_atendimento_vacina, dose_vacina) %>%
  # Aggregate records by state, date, target group, and dose category
  group_by(sigla_uf, data_aplicacao_vacina, grupo_atendimento_vacina, dose_vacina) %>%
  summarise(n_casos = n()) %>%
  bd_collect()

# Filter the dataset to include only vaccination records for pregnant and puerperal women.
# The selected values of `grupo_atendimento_vacina` correspond to the codes assigned
# to these specific priority groups in the official coding system.
dados_gesta_puerp <- ai %>%
  filter(grupo_atendimento_vacina == "1901" | grupo_atendimento_vacina == "1901.0" | grupo_atendimento_vacina == "1901.1")

# Reclassify and harmonize dose descriptions into a smaller set of standardized categories.
# This step collapses multiple textual labels into interpretable dose groups:
# "1a dose", "2a dose ou unica", "1o reforco", "2o reforco".

```

```

dados_gesta_puerp1 <- dados_gesta_puerp %>%
  mutate(
    num_dose = case_when(
      dose_vacina == "1ª Dose" ~ "1a dose",
      dose_vacina == "1ª Dose Revacinação" ~ "1o reforco",
      dose_vacina == "1º Reforço" ~ "1o reforco",
      dose_vacina == "2ª Dose" ~ "2a dose ou unica",
      dose_vacina == "2ª Dose Revacinação" ~ "2o reforco",
      dose_vacina == "2º Reforço" ~ "2o reforco",
      dose_vacina == "3ª Dose" ~ "1o reforco",
      dose_vacina == "3ª Dose Revacinação" ~ "1o reforco",
      dose_vacina == "3º Reforço" ~ "1o reforco",
      dose_vacina == "4ª Dose" ~ "2o reforco",
      dose_vacina == "Dose Adicional" ~ "1o reforco",
      dose_vacina == "Dose Inicial" ~ "1a dose",
      dose_vacina == "Dose Única" ~ "2a dose ou unica",
      dose_vacina == "Reforço" ~ "1o reforco",
      dose_vacina == "Revacinação" ~ "1o reforco",
      dose_vacina == "Única" ~ "2a dose ou unica"
    )
  )

# Aggregate vaccination counts for pregnant/puerperal women at the state (UF) level.
dados_materna_UF <- dados_gesta_puerp1 %>%
  # Remove the original, more granular dose_vacina label to avoid confusion,
  # since we will now work with the standardized `num_dose`.
  select(-dose_vacina) %>%
  # Group by state, vaccination date, and standardized dose category
  group_by(sigla_uf, data_aplicacao_vacina, num_dose) %>%
  summarise(n_casos = sum(n_casos)) %>%
  select(sigla_uf, data_aplicacao_vacina, dose_vacina = num_dose, n_casos) %>%
  filter(!is.na(dose_vacina))

# Export state-level dataset for pregnant/puerperal women in both CSV and XLSX formats.
# The filenames suggest a data extraction or update up to 01-02-22.
write_csv(dados_materna_UF, "Vacinados_GestaPuerp_porUF_01-02-22.csv")
write_xlsx(dados_materna_UF, "Vacinados_GestaPuerp_porUF_01-02-22.xlsx")

# Build a national-level aggregation (Brazil as a whole) by summing across all states.
dados_materna_Br <- dados_materna_UF %>%
  group_by(data_aplicacao_vacina, dose_vacina) %>%
  summarise(n_casos = sum(n_casos))

# Export national-level dataset in CSV and XLSX formats.
write_csv(dados_materna_Br, "Vacinados_GestaPuerp_Brasil_01-02-22.csv")
write_xlsx(dados_materna_Br, "Vacinados_GestaPuerp_Brasil_01-02-22.xlsx")

#----- The End

# oil <- oi %>%
#   mutate(
#     num_dose = case_when(

```

```

#      id %in% str_which(dose_vacina, "1ª Dose") ~ "1a dose",
#      id %in% str_which(dose_vacina, "1ª Dose Revacinação") ~ "1a dose",
#      id %in% str_which(dose_vacina, "2ª") ~ "2a dose",
#      id %in% str_which(dose_vacina, "nica") ~ "dose unica",
#      id %in% str_which(dose_vacina, "Única") ~ "dose unica",
#      id %in% str_which(dose_vacina, "DoseÃ") ~ "dose unica",
#      id %in% str_which(dose_vacina, "Adicional") ~ "dose adicional",
#      id %in% str_which(dose_vacina, "DoseAdicional") ~ "dose adicional",
#      vacina_descricao_dose == "Dose" ~ "dose unica",
#      id %in% str_which(vacina_descricao_dose, "3ª") ~ "dose adicional",
#      id %in% str_which(vacina_descricao_dose, "Reforço") ~ "dose adicional"
#    )
#  )

```