Internações e vacinação

Raphael Saldanha

2023-02-24

## Pacotes

library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.0 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.1 ✔ tibble 3.1.8  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.1   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the ]8;;http://conflicted.r-lib.org/conflicted package]8;; to force all conflicts to become errors

library(arrow)

##   
## Attaching package: 'arrow'  
##   
## The following object is masked from 'package:lubridate':  
##   
## duration  
##   
## The following object is masked from 'package:utils':  
##   
## timestamp

## Leitura dos dados

### Variables schema

date\_format <- "%d/%m/%Y"  
  
var\_schema <- cols(  
 DT\_NOTIFIC = col\_date(format = date\_format),  
 SEM\_NOT = col\_character(),  
 DT\_SIN\_PRI = col\_date(format = date\_format),  
 SEM\_PRI = col\_character(),  
 SG\_UF\_NOT = col\_character(),  
 ID\_REGIONA = col\_character(),  
 CO\_REGIONA = col\_character(),  
 ID\_MUNICIP = col\_character(),  
 CO\_MUN\_NOT = col\_character(),  
 ID\_UNIDADE = col\_character(),  
 CO\_UNI\_NOT = col\_character(),  
 CS\_SEXO = col\_character(),  
 DT\_NASC = col\_date(format = date\_format),  
 NU\_IDADE\_N = col\_double(),  
 TP\_IDADE = col\_character(),  
 COD\_IDADE = col\_character(),  
 CS\_GESTANT = col\_character(),  
 CS\_RACA = col\_character(),  
 CS\_ESCOL\_N = col\_character(),  
 ID\_PAIS = col\_character(),  
 CO\_PAIS = col\_character(),  
 SG\_UF = col\_character(),  
 ID\_RG\_RESI = col\_character(),  
 CO\_RG\_RESI = col\_character(),  
 ID\_MN\_RESI = col\_character(),  
 CO\_MUN\_RES = col\_character(),  
 CS\_ZONA = col\_character(),  
 SURTO\_SG = col\_character(),  
 NOSOCOMIAL = col\_double(),  
 AVE\_SUINO = col\_double(),  
 FEBRE = col\_double(),  
 TOSSE = col\_double(),  
 GARGANTA = col\_double(),  
 DISPNEIA = col\_double(),  
 DESC\_RESP = col\_double(),  
 SATURACAO = col\_double(),  
 DIARREIA = col\_double(),  
 VOMITO = col\_double(),  
 OUTRO\_SIN = col\_double(),  
 OUTRO\_DES = col\_character(),  
 PUERPERA = col\_double(),  
 FATOR\_RISC = col\_double(),  
 CARDIOPATI = col\_double(),  
 HEMATOLOGI = col\_double(),  
 SIND\_DOWN = col\_double(),  
 HEPATICA = col\_double(),  
 ASMA = col\_double(),  
 DIABETES = col\_double(),  
 NEUROLOGIC = col\_double(),  
 PNEUMOPATI = col\_double(),  
 IMUNODEPRE = col\_double(),  
 RENAL = col\_double(),  
 OBESIDADE = col\_double(),  
 OBES\_IMC = col\_double(),  
 OUT\_MORBI = col\_double(),  
 MORB\_DESC = col\_character(),  
 VACINA = col\_double(),  
 DT\_UT\_DOSE = col\_character(),  
 MAE\_VAC = col\_double(),  
 DT\_VAC\_MAE = col\_date(format = date\_format),  
 M\_AMAMENTA = col\_double(),  
 DT\_DOSEUNI = col\_date(format = date\_format),  
 DT\_1\_DOSE = col\_date(format = date\_format),  
 DT\_2\_DOSE = col\_date(format = date\_format),  
 ANTIVIRAL = col\_double(),  
 TP\_ANTIVIR = col\_double(),  
 OUT\_ANTIV = col\_character(),  
 DT\_ANTIVIR = col\_character(),  
 HOSPITAL = col\_double(),  
 DT\_INTERNA = col\_date(format = date\_format),  
 SG\_UF\_INTE = col\_character(),  
 ID\_RG\_INTE = col\_character(),  
 CO\_RG\_INTE = col\_double(),  
 ID\_MN\_INTE = col\_character(),  
 CO\_MU\_INTE = col\_double(),  
 UTI = col\_double(),  
 DT\_ENTUTI = col\_date(format = date\_format),  
 DT\_SAIDUTI = col\_date(format = date\_format),  
 SUPORT\_VEN = col\_double(),  
 RAIOX\_RES = col\_double(),  
 RAIOX\_OUT = col\_character(),  
 DT\_RAIOX = col\_date(format = date\_format),  
 AMOSTRA = col\_double(),  
 DT\_COLETA = col\_date(format = date\_format),  
 TP\_AMOSTRA = col\_double(),  
 OUT\_AMOST = col\_character(),  
 PCR\_RESUL = col\_double(),  
 DT\_PCR = col\_date(format = date\_format),  
 POS\_PCRFLU = col\_double(),  
 TP\_FLU\_PCR = col\_double(),  
 PCR\_FLUASU = col\_double(),  
 FLUASU\_OUT = col\_character(),  
 PCR\_FLUBLI = col\_double(),  
 FLUBLI\_OUT = col\_character(),  
 POS\_PCROUT = col\_double(),  
 PCR\_VSR = col\_double(),  
 PCR\_PARA1 = col\_double(),  
 PCR\_PARA2 = col\_double(),  
 PCR\_PARA3 = col\_double(),  
 PCR\_PARA4 = col\_double(),  
 PCR\_ADENO = col\_double(),  
 PCR\_METAP = col\_double(),  
 PCR\_BOCA = col\_double(),  
 PCR\_RINO = col\_double(),  
 PCR\_OUTRO = col\_double(),  
 DS\_PCR\_OUT = col\_character(),  
 CLASSI\_FIN = col\_double(),  
 CLASSI\_OUT = col\_character(),  
 CRITERIO = col\_double(),  
 EVOLUCAO = col\_double(),  
 DT\_EVOLUCA = col\_character(),  
 DT\_ENCERRA = col\_character(),  
 DT\_DIGITA = col\_character(),  
 HISTO\_VGM = col\_double(),  
 PAIS\_VGM = col\_character(),  
 CO\_PS\_VGM = col\_character(),  
 LO\_PS\_VGM = col\_character(),  
 DT\_VGM = col\_date(format = date\_format),  
 DT\_RT\_VGM = col\_date(format = date\_format),  
 PCR\_SARS2 = col\_double(),  
 PAC\_COCBO = col\_character(),  
 PAC\_DSCBO = col\_character(),  
 OUT\_ANIM = col\_character(),  
 DOR\_ABD = col\_double(),  
 FADIGA = col\_double(),  
 PERD\_OLFT = col\_double(),  
 PERD\_PALA = col\_double(),  
 TOMO\_RES = col\_double(),  
 TOMO\_OUT = col\_character(),  
 DT\_TOMO = col\_character(),  
 TP\_TES\_AN = col\_double(),  
 DT\_RES\_AN = col\_character(),  
 RES\_AN = col\_double(),  
 POS\_AN\_FLU = col\_double(),  
 TP\_FLU\_AN = col\_double(),  
 POS\_AN\_OUT = col\_double(),  
 AN\_SARS2 = col\_double(),  
 AN\_VSR = col\_double(),  
 AN\_PARA1 = col\_character(),  
 AN\_PARA2 = col\_character(),  
 AN\_PARA3 = col\_character(),  
 AN\_ADENO = col\_character(),  
 AN\_OUTRO = col\_double(),  
 DS\_AN\_OUT = col\_character(),  
 TP\_AM\_SOR = col\_double(),  
 SOR\_OUT = col\_character(),  
 DT\_CO\_SOR = col\_character(),  
 TP\_SOR = col\_double(),  
 OUT\_SOR = col\_character(),  
 DT\_RES = col\_character(),  
 RES\_IGG = col\_double(),  
 RES\_IGM = col\_double(),  
 RES\_IGA = col\_double(),  
 ESTRANG = col\_double(),  
 VACINA\_COV = col\_double(),  
 DOSE\_1\_COV = col\_date(format = date\_format),  
 DOSE\_2\_COV = col\_date(format = date\_format),  
 DOSE\_REF = col\_date(format = date\_format),  
 FAB\_COV\_1 = col\_character(),  
 FAB\_COV\_2 = col\_character(),  
 FAB\_COVREF = col\_character(),  
 LAB\_PR\_COV = col\_character(),  
 LOTE\_1\_COV = col\_character(),  
 LOTE\_2\_COV = col\_character(),  
 LOTE\_REF = col\_character(),  
 FNT\_IN\_COV = col\_double(),  
 DOSE\_2REF = col\_character(),  
 FAB\_COVRF2 = col\_character(),  
 LOTE\_REF2 = col\_character(),  
 TRAT\_COV = col\_double(),  
 TIPO\_TRAT = col\_double(),  
 OUT\_TRAT = col\_character(),  
 DT\_TRT\_COV = col\_character()  
)

### Conversão de CSV para parquet conversio com variables schema

# read\_csv2(file = "csv\_srag/INFLUD21-20-02-2023.csv", col\_types = var\_schema) %>%  
# write\_parquet(sink = "parquets\_srag/srag\_2021.parquet")  
# srag\_2022 <- read\_csv2(file = "csv\_srag/INFLUD22-13-02-2023.csv", col\_types = var\_schema) %>%  
# write\_parquet(sink = "parquets\_srag/srag\_2022.parquet")  
# srag\_2023 <- read\_csv2(file = "csv\_srag/INFLUD23-13-02-2023.csv", col\_types = var\_schema) %>%  
# write\_parquet(sink = "parquets\_srag/srag\_2023.parquet")

### Carregar parquets como dataset

srag <- arrow::open\_dataset(sources = "parquets\_srag/", unify\_schemas = TRUE)  
dim(srag)

## [1] 2310360 173

2.310.360 casos suspeitos de SRAG reportados entre 2021, 2022 e 2023.

### Dataset de referência

Filtros: diagnóstico de Covid-19 confirmado, hospitalização confirmada, data válida de internação.

srag\_vac <- srag %>%  
 filter(DT\_INTERNA >= as.Date("2020-01-01") & DT\_INTERNA < as.Date("2024-01-01")) %>%  
 filter(HOSPITAL == 1) %>%  
 filter(CLASSI\_FIN == 5) %>%  
 select(DT\_INTERNA, VACINA\_COV, FAB\_COV\_1, DOSE\_1\_COV, FAB\_COV\_2, DOSE\_2\_COV, FAB\_COVREF, DOSE\_REF)

srag\_vac %>%  
 tally() %>%  
 collect()

## # A tibble: 1 × 1  
## n  
## <int>  
## 1 1380499

Dos 2.310.360 casos suspeitos reportados no período, 1.380.499 são casos confirmados de Covid-19 e hospitalizados (59,75%).

srag\_vac %>%  
 filter(DOSE\_1\_COV < DT\_INTERNA | DOSE\_2\_COV < DT\_INTERNA | DOSE\_REF < DT\_INTERNA) %>%  
 tally() %>%  
 collect()

## # A tibble: 1 × 1  
## n  
## <int>  
## 1 346598

Destes, 346.598 pacientes receberam alguma vacina de Covid-19 antes da data de internação (25,11%).

srag\_vac %>%  
 filter(DOSE\_1\_COV < DT\_INTERNA | DOSE\_2\_COV < DT\_INTERNA | DOSE\_REF < DT\_INTERNA) %>%  
 collect() %>%  
 mutate(  
 dias\_dose\_1 = DT\_INTERNA - DOSE\_1\_COV,  
 dias\_dose\_2 = DT\_INTERNA - DOSE\_2\_COV,  
 dias\_dose\_r = DT\_INTERNA - DOSE\_REF  
 ) %>%  
 select(dias\_dose\_1, dias\_dose\_2, dias\_dose\_r) %>%  
 summarise(  
 media\_dias\_dose\_1 = mean(dias\_dose\_1, na.rm = TRUE),  
 media\_dias\_dose\_2 = mean(dias\_dose\_2, na.rm = TRUE),  
 media\_dias\_dose\_3 = mean(dias\_dose\_r, na.rm = TRUE),  
 ) %>%  
 pivot\_longer(cols = everything())

## # A tibble: 3 × 2  
## name value   
## <chr> <drtn>   
## 1 media\_dias\_dose\_1 229.8345 days  
## 2 media\_dias\_dose\_2 218.8992 days  
## 3 media\_dias\_dose\_3 181.1703 days

srag\_vac %>%  
 filter(DOSE\_1\_COV < DT\_INTERNA | DOSE\_2\_COV < DT\_INTERNA | DOSE\_REF < DT\_INTERNA) %>%  
 select(FAB\_COV\_1, FAB\_COV\_2, FAB\_COVREF) %>%  
 collect() %>%  
 pivot\_longer(cols = everything()) %>%  
 select(name = value) %>%  
 group\_by(name) %>%  
 summarise(freq = n()) %>%  
 ungroup() %>%  
 arrange(-freq)

## # A tibble: 3,828 × 2  
## name freq  
## <chr> <int>  
## 1 <NA> 490553  
## 2 86 - COVID-19 SINOVAC/BUTANTAN - CORONAVAC 169024  
## 3 85 - COVID-19 ASTRAZENECA/FIOCRUZ - COVISHIELD 113467  
## 4 87 - COVID-19 PFIZER - COMIRNATY 89848  
## 5 CORONAVAC 30609  
## 6 BUTANTAN 21241  
## 7 ASTRAZENECA 17465  
## 8 FIOCRUZ 9697  
## 9 PFIZER 8249  
## 10 88 - COVID-19 JANSSEN - AD26.COV2.S 8087  
## # … with 3,818 more rows

res <- srag\_vac %>%  
 filter(DOSE\_1\_COV < DT\_INTERNA | DOSE\_2\_COV < DT\_INTERNA | DOSE\_REF < DT\_INTERNA) %>%  
 select(FAB\_COV\_1, FAB\_COV\_2, FAB\_COVREF) %>%  
 collect() %>%  
 mutate(  
 fiocruz\_1 = str\_detect(toupper(FAB\_COV\_1), "CRUZ"),  
 nfiocruz\_1 = !str\_detect(toupper(FAB\_COV\_1), "CRUZ"),  
 fiocruz\_2 = str\_detect(toupper(FAB\_COV\_2), "CRUZ"),  
 nfiocruz\_2 = !str\_detect(toupper(FAB\_COV\_2), "CRUZ"),  
 fiocruz\_r = str\_detect(toupper(FAB\_COVREF), "CRUZ"),  
 nfiocruz\_r = !str\_detect(toupper(FAB\_COVREF), "CRUZ"),  
 ) %>%  
 select(fiocruz\_1, nfiocruz\_1, fiocruz\_2, nfiocruz\_2, fiocruz\_r, nfiocruz\_r)

# Pessoas internadas com Covid-19 após terem recebido a primeira dose de vacina cujo fabricante tem "cruz" no nome  
sum(res$fiocruz\_1, na.rm = TRUE)

## [1] 90409

# Pessoas internadas com Covid-19 após terem recebido a primeira dose de vacina cujo fabricante não tem "cruz" no nome  
sum(res$nfiocruz\_1, na.rm = TRUE)

## [1] 239271

# Razão de pessoas internadas com Covid-19 após terem recebido a primeira dose de vacina cujo fabricante tem "cruz" no nome  
sum(res$fiocruz\_1, na.rm = TRUE)/(sum(res$fiocruz\_1, na.rm = TRUE) + sum(res$nfiocruz\_1, na.rm = TRUE))

## [1] 0.2742326

# Pessoas internadas com Covid-19 após terem recebido a segunda dose de vacina cujo fabricante tem "cruz" no nome  
sum(res$fiocruz\_2, na.rm = TRUE)

## [1] 44370

# Pessoas internadas com Covid-19 após terem recebido a segunda dose de vacina cujo fabricante não tem "cruz" no nome  
sum(res$nfiocruz\_2, na.rm = TRUE)

## [1] 88341

# Razão de pessoas internadas com Covid-19 após terem recebido a segunda dose de vacina cujo fabricante tem "cruz" no nome  
sum(res$fiocruz\_2, na.rm = TRUE)/(sum(res$fiocruz\_2, na.rm = TRUE) + sum(res$nfiocruz\_2, na.rm = TRUE))

## [1] 0.3343355

# Pessoas internadas com Covid-19 após terem recebido a segunda dose de vacina cujo fabricante tem "cruz" no nome  
sum(res$fiocruz\_r, na.rm = TRUE)

## [1] 5716

# Pessoas internadas com Covid-19 após terem recebido a segunda dose de vacina cujo fabricante não tem "cruz" no nome  
sum(res$nfiocruz\_r, na.rm = TRUE)

## [1] 81134

# Razão de pessoas internadas com Covid-19 após terem recebido a segunda dose de vacina cujo fabricante tem "cruz" no nome  
sum(res$fiocruz\_r, na.rm = TRUE)/(sum(res$fiocruz\_r, na.rm = TRUE) + sum(res$nfiocruz\_r, na.rm = TRUE))

## [1] 0.06581462