

# Processamento dos dados brutos

## Bibliotecas

```
library(tidyverse)
library(data.table)
library(yaml)
```

## Carregando os dados

```
df_raw = readxl::read_excel('./data/Dados (NOVO).xlsx')
names(df_raw) <- str_replace_all(tolower(names(df_raw)), c("-" = "_", " " = ""))

constants = yaml.load_file("constants.yaml")

categorical_columns = constants$categorical_columns

daily_columns = constants$daily_columns
```

## Processamento dos dados

```
df = df_raw %>%
  mutate_at(categorical_columns, list(~factor(.))) %>%
  mutate(ira = factor(ifelse(d_ira != 0, 1, 0)),
         kdigo = factor(kdigo),
         kdigo_agrupado = case_when(kdigo %in% c(2, 3) ~ "MODERADO/GRAVE",
                                     kdigo == 1 ~ "LEVE",
                                     kdigo == 0 ~ "SEM IRA"))

for (col in daily_columns){
  df = df %>%
    mutate(!col := if_else(ira == 1, !!sym(paste0(col, '_d_1')),
                          !!sym(paste0(col, '_media_d1_d5'))))
}

df = df %>%
  select(-matches("d_?\\d"))

dim(df)

## [1] 55 59

df[rowSums(is.na(df %>% select(-criterio_ira))) > 0,]$numero

## [1] 30 67 76 9 34
```

```
df_2 = df %>%  
  filter(!numero %in% c(30, 67, 76))
```

## Salvando dados processados

```
saveRDS(df_2, "./data/dados_processados.rds")
```