

Desafio 1 - ME315

Bruce Trevisan

2025-09-23

```
# install.packages("tidyverse")
# install.packages("dplyr")
# install.packages("tidyr")
```

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr    1.5.2
## v ggplot2    4.0.0      v tibble     3.3.0
## v lubridate  1.9.4      v tidyr      1.3.1
## v purrr      1.1.0
```

```
## -- Conflicts ----- tidyverse_conflicts() --
```

```
## x dplyr::filter() masks stats::filter()
```

```
## x dplyr::lag()     masks stats::lag()
```

```
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(dplyr)
```

```
library(tidyr)
```

```
data()
```

```
table1
```

```
## # A tibble: 6 x 4
```

```
##   country      year cases population
##   <chr>      <dbl> <dbl>      <dbl>
## 1 Afghanistan 1999    745    19987071
## 2 Afghanistan 2000   2666   20595360
## 3 Brazil      1999   37737   172006362
## 4 Brazil      2000   80488   174504898
## 5 China       1999  212258  1272915272
## 6 China       2000  213766  1280428583
```

```
table2
```

```
## # A tibble: 12 x 4
```

```
##   country      year type      count
##   <chr>      <dbl> <chr>      <dbl>
## 1 Afghanistan 1999 cases         745
## 2 Afghanistan 1999 population 19987071
## 3 Afghanistan 2000 cases         2666
## 4 Afghanistan 2000 population 20595360
## 5 Brazil      1999 cases         37737
## 6 Brazil      1999 population 172006362
```

```
## 7 Brazil      2000 cases      80488
## 8 Brazil      2000 population 174504898
## 9 China       1999 cases      212258
## 10 China      1999 population 1272915272
## 11 China      2000 cases      213766
## 12 China      2000 population 1280428583
```

```
table3
```

```
## # A tibble: 6 x 3
##   country      year rate
##   <chr>      <dbl> <chr>
## 1 Afghanistan 1999 745/19987071
## 2 Afghanistan 2000 2666/20595360
## 3 Brazil      1999 37737/172006362
## 4 Brazil      2000 80488/174504898
## 5 China       1999 212258/1272915272
## 6 China       2000 213766/1280428583
```

```
table4a
```

```
## # A tibble: 3 x 3
##   country      `1999` `2000`
##   <chr>      <dbl> <dbl>
## 1 Afghanistan    745    2666
## 2 Brazil        37737   80488
## 3 China         212258  213766
```

```
table4b
```

```
## # A tibble: 3 x 3
##   country      `1999`      `2000`
##   <chr>      <dbl>      <dbl>
## 1 Afghanistan 19987071    20595360
## 2 Brazil      172006362   174504898
## 3 China       1272915272  1280428583
```

```
# 3. Calcular a taxa de ocorrência de tuberculose por 10.000 pessoas
```

```
# Vamos criar a taxa dividindo os casos de tuberculose pela população e multiplicando por 10.000. Vamos
```

```
# Calcular a taxa de tuberculose por 10.000 pessoas
```

```
taxas <- table1 %>%
  mutate(taxa_por_10000 = (cases / population) * 10000)
```

```
# Exibir a tabela com as taxas
```

```
taxas
```

```
## # A tibble: 6 x 5
##   country      year cases population taxa_por_10000
##   <chr>      <dbl> <dbl>      <dbl>      <dbl>
## 1 Afghanistan 1999    745    19987071      0.373
## 2 Afghanistan 2000   2666    20595360      1.29
## 3 Brazil      1999  37737   172006362      2.19
## 4 Brazil      2000  80488   174504898      4.61
## 5 China       1999 212258  1272915272      1.67
## 6 China       2000 213766  1280428583      1.67
```

4. Número de casos de tuberculose por ano

Agrupar os dados por ano e somar os casos de tuberculose.

```
casos_por_ano <- table1 %>%  
  group_by(year) %>%  
  summarise(total_casos = sum(cases, na.rm = TRUE))  
# Número de casos de tuberculose por ano  
casos_por_ano
```

```
## # A tibble: 2 x 2  
##   year total_casos  
##   <dbl>      <dbl>  
## 1  1999      250740  
## 2  2000      296920
```

5. Número de casos por país

Agrupar os dados por país e somar os casos.

```
casos_por_pais <- table1 %>%  
  group_by(country) %>%  
  summarise(total_casos = sum(cases, na.rm = TRUE))  
# Número de casos de tuberculose por país  
casos_por_pais
```

```
## # A tibble: 3 x 2  
##   country      total_casos  
##   <chr>          <dbl>  
## 1 Afghanistan      3411  
## 2 Brazil           118225  
## 3 China             426024
```

6. Mudança no número de casos entre 1999 e 2000

Mudança no número de casos de tuberculose entre 1999 e 2000

```
mudanca_1999_2000 <- table1 %>%  
  filter(year %in% c(1999, 2000)) %>%  
  group_by(country, year) %>%  
  summarise(total_casos = sum(cases, na.rm = TRUE)) %>%  
  spread(key = year, value = total_casos) %>%  
  mutate(mudanca = `2000` - `1999`)
```

```
## `summarise()` has grouped output by 'country'. You can override using the  
## `.groups` argument.
```

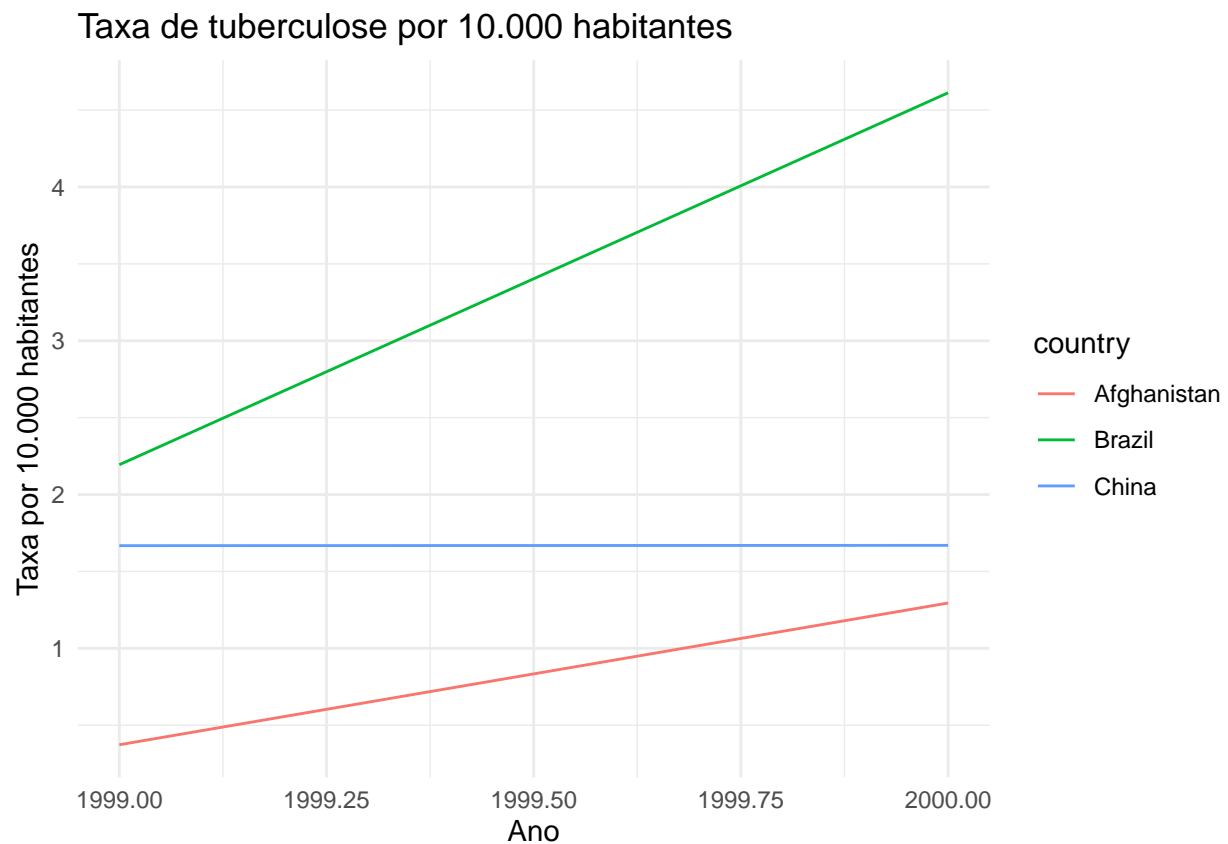
```
mudanca_1999_2000
```

```
## # A tibble: 3 x 4  
## # Groups:   country [3]  
##   country      `1999` `2000` mudanca  
##   <chr>      <dbl> <dbl>   <dbl>  
## 1 Afghanistan      745   2666    1921  
## 2 Brazil          37737  80488   42751  
## 3 China          212258 213766    1508
```

7. Gráfico de linhas da taxa de casos por 10.000 habitantes estratificado por país

```
# Gráfico de linhas para taxa de tuberculose por 10.000 habitantes

ggplot(taxas, aes(x = year, y = taxa_por_10000, color = country)) +
  geom_line() +
  labs(title = "Taxa de tuberculose por 10.000 habitantes",
       x = "Ano",
       y = "Taxa por 10.000 habitantes") +
  theme_minimal()
```



8. Calcular a taxa para table2 e table4a + table4b

```
# 1. Extrair o número de casos por país e ano

# 2. Extrair o tamanho da população

# 3. Calcular a taxa de casos (por 10.000 pessoas)

# 4. Armazenar o resultado em uma variável
```

8.a

```
table2_population <- table2 %>%
  filter(type == "population")
View(table2_population)
```

8.b Transformar table4a para formato tidy

```
tidy4a <- table4a %>%
  pivot_longer(cols = c(`1999`, `2000`),
               names_to = "year",
               values_to = "cases")
View(tidy4a)
```

8.c Transformar table4b para formato tidy

```
tidy4b <- table4b %>%
  pivot_longer(cols = c(`1999`, `2000`),
               names_to = "year",
               values_to = "cases")
View(tidy4b)
```

8.d Combinar as duas tabelas (table4a + table4b)

```
combined <- left_join(tidy4a, tidy4b, by = c("country", "year"))
View(combined)
```

Transformar table2 para formato wide

```
table2_wide <- table2 %>%
  pivot_wider(names_from = type, values_from = count)
View(table2_wide)
```

Separar a coluna rate de table3

```
table3_tidy <- table3 %>%
  separate(rate, into = c("cases", "population"), sep = "/", convert = TRUE)
table3_tidy
```

```
## # A tibble: 6 x 4
##   country    year cases population
##   <chr>      <dbl> <int>      <int>
## 1 Afghanistan 1999    745   19987071
## 2 Afghanistan 2000   2666   20595360
## 3 Brazil      1999  37737  172006362
## 4 Brazil      2000  80488  174504898
## 5 China       1999 212258 1272915272
## 6 China       2000 213766 1280428583
```

```
table3_tidy <- table3 %>%
  separate(col = rate,
           into = c("cases", "population"),
           sep = "/",
           convert = TRUE)
View(table3_tidy)
```