Grouped, stacked and percent stacked barplot

Adriano V. Werhli

Exercises

In this exercise you have to handle the R script that produce each of the figures that are presented in each question. Use as reference the tutorial here. Load the file `data_mortalidade_Regiao.csv`. This file was obtained from this link. It contains the data for the year 2021. When reading the csv file use the option `check.names = F`. For this exercises you do not need to tidy the data. The information we need is already in columns.

Before starting the exercises we investigate how to change the order in which the levels of a factor appear. You will need to do this in the data set that we will use. In R, factors are used to work with categorical variables, variables that have a fixed and known set of possible values. They are also useful when you want to display character vectors in a non-alphabetical order.

In the following example we read a csv file and one of the variables is months. If we try to sort, or plot, this data, the months will be ordered in alphabetical order. By using the function `factor` we define the levels and the order in which they should be ordered.

```
library(ggplot2)
library(tidyverse)
library(RColorBrewer)
library(showtext)
library(scales)
setwd("~/Dropbox/Aulas_FURG/PropostaDisciplinaVisualizacaoDados/Aulas_2023/7_Exercises_B
rm(list = ls())
df<-read.csv("birth_example_factor.csv")
df<-as_tibble(df)
# Months are in alfabetical order...
sort(df$Month)
```

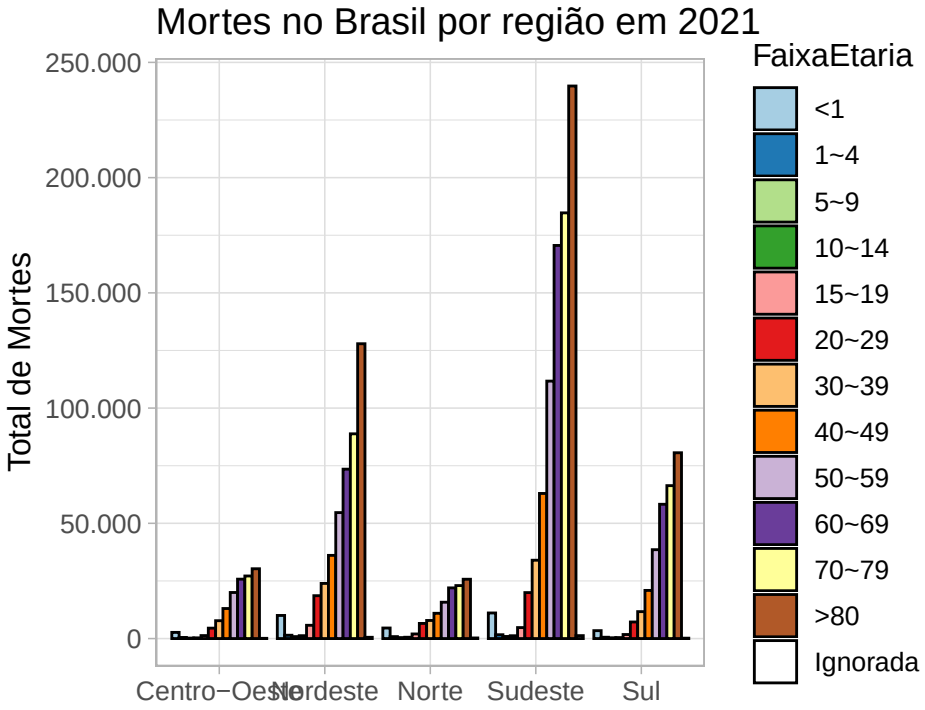
```
##  [1] "Apr" "Aug" "Dec" "Dec" "Feb" "Jan" "Jul" "Jun" "Mar" "Mar" "May" "Nov"
## [13] "Nov" "Oct" "Oct" "Sep"
```

```
# We define the order
my.levels<-c("Jan","Feb","Mar","Apr","May","Jun","Jul"
             ,"Aug","Sep","Oct","Nov","Dec")
df$Month<-factor(df$Month,levels = my.levels)
```

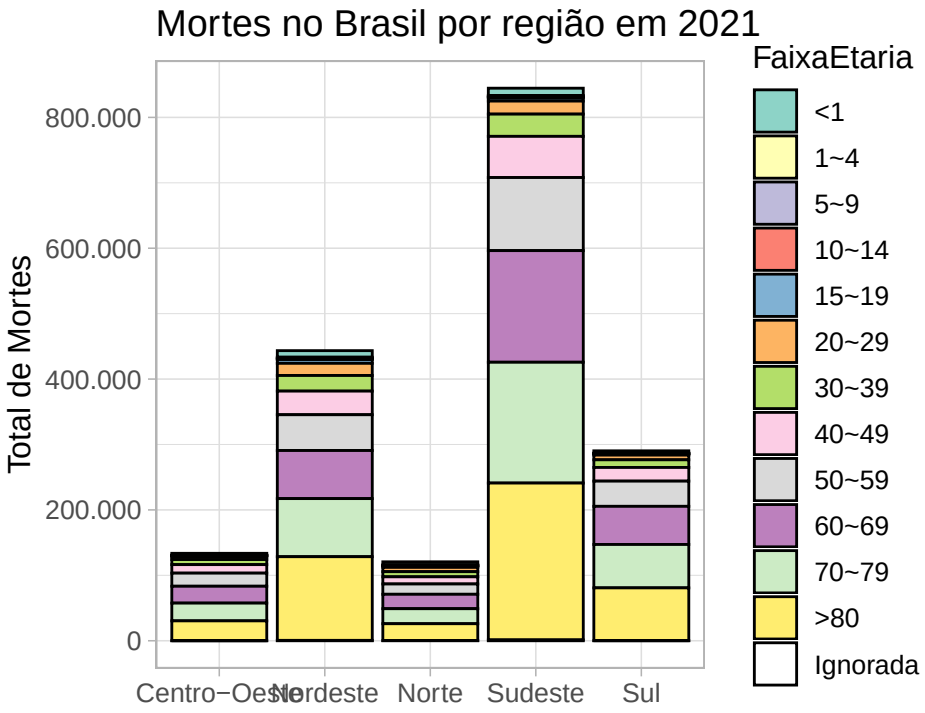
```r
# And now the months are sorted in the order that we defined
sort(df$Month)
```

```
##  [1] Jan Feb Mar Mar Apr May Jun Jul Aug Sep Oct Oct Nov Nov Dec Dec
## Levels: Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
```

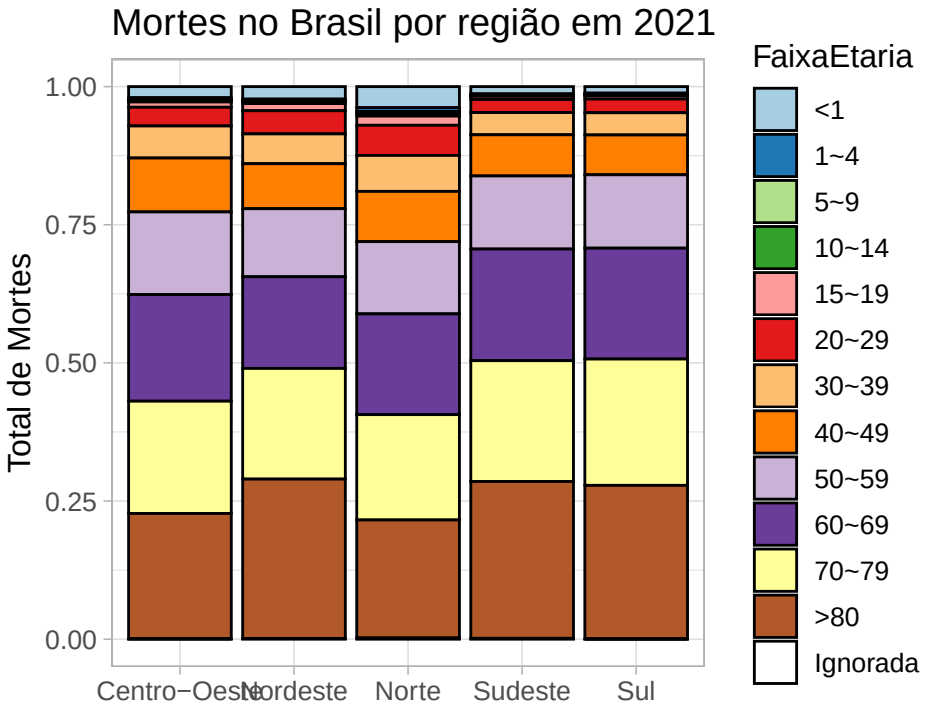
1. In this exercise you will need to change the original data. My suggestion is to change the name of the columns, so that we have the intervals with more meaningful names. After that, you will need to remove column and line with `Total` and pivot the data so that the `Região` is a group and `FaixaEtaria` is as subgroup in the barplot.



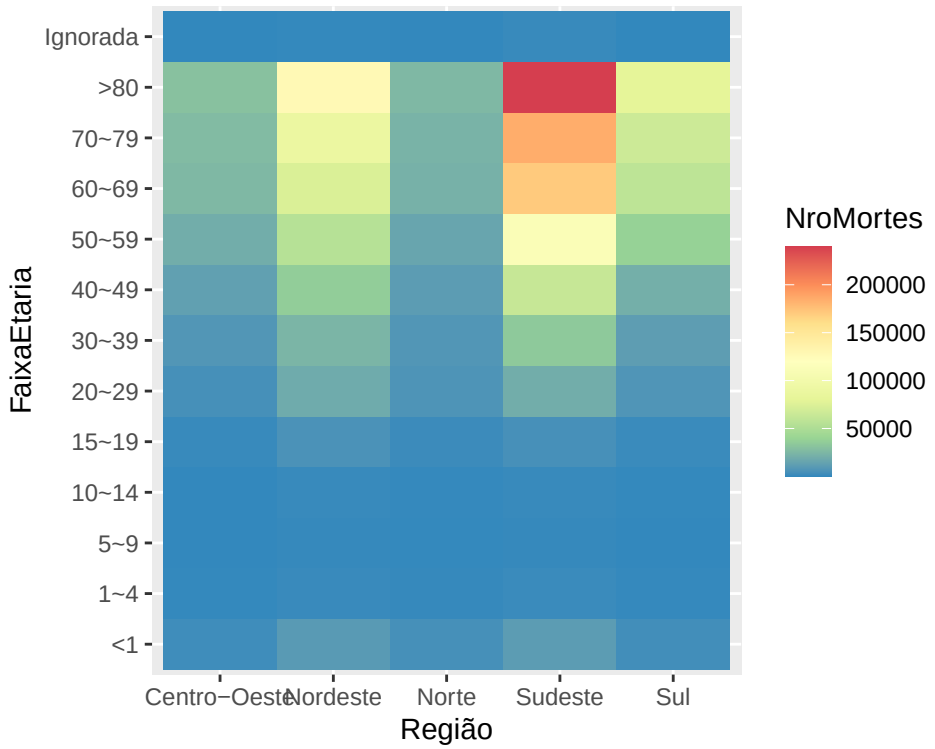
2. After producing the barplot with groups of bars, side by side as above, you can change just one parameter and obtain the figure below, where the bars are stacked.



3. Once again, changing just one parameter, you can obtain the figure below, with the bars stacked filling the whole scale, from 0% to 100%.



4. Another way of showing amounts is by using heatmap plots. In ggplot2 heatmaps are obtained with the geom `geom_tile`. Arrange the previous code and use `scale_fill_distiller(palette = "Spectral")` to obtain the graph below.



5. Modifying the previous figure by introducing a black color to `geom_tile` and using `geom_text(aes(label=NroMortes)` obtain the heatmap below.