

Introdução à Linguagem de Programação em R para tratamento de dados de poluição do ar

Mario Gavidia-Calderón, Rafaela Squizzato, Thiago Nogueira

06/02/2024

Universidade de São Paulo

Pacotes

Do Excel para R

Subset data frames

qualR

Pacotes

- Pacotes são como as **extensões** no browser.
- Conjunto de funções específicas para tarefas específicas.

- Para instalar pacotes usamos a função
`install.packages("nome_do_biblioteca")`

Instalando Rmarkdown

```
install.packages("rmarkdown")
```

- Por exemplo, para ler arquivos `.xls` podemos instalar a biblioteca `readxls`

```
# Instalando readxl  
install.packages("readxl")
```

Do Excel para R

- Um jeito de ler arquivos `.xls` é abrir os arquivos em Excel ou Google Sheet e salvar como `.csv`.
- Depois ler a tabela usando a função `read.table()`
- As vezes vale a pena abrir no Excel ou Google Sheet e mudar o nome das colunas.
- Também podemos usar a função `read_excel()` do pacote `readxl`

- Vamos usar a base de dados de qualidade do ar da WHO

```
library(readxl)
who <- read_excel("../..//data/who_aap_2021_v9_11august2022.xlsx",
                  sheet = 2) # Nome o numero do sheet no arquivo
```

Base de dados da WHO

- Este é um exemplo de nomes de coluna *complicados*.

```
str(who)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame':    32191 obs. of  15 variables:
## $ WHO Region                               : chr  "Eastern Mediterranean"
## $ ISO3                                       : chr  "AFG" "ALB" "ALB" "ALB"
## $ WHO Country Name                         : chr  "Afghanistan" "Albania"
## $ City or Locality                         : chr  "Kabul" "Durres" "Durres"
## $ Measurement Year                        : num  2019 2015 2016 2015
## $ PM2.5 (µg/m3)                           : num  119.8 NA 14.3 NA NA
## $ PM10 (µg/m3)                            : num  NA 17.6 24.6 NA NA
## $ NO2 (µg/m3)                             : num  NA 26.6 24.8 24 26.3
## $ PM25 temporal coverage (%)              : num  18 NA NA NA NA NA
```

Subset data frames

Subset

- As vezes precisamos selecionar um sub-conjunto de dados do data frame.
- Um jeito simples é usar a função `subset()`

```
novo_df <- subset(df_original,  
                  subset = condição baseado nas filas,  
                  select = vetor com as colunas)
```

Subset dados da who para Brasil

- Vamos selecionar os dados do Brasil.

```
who_br <- subset(who,  
                 subset = `WHO Country Name` == "Brazil")
```

- Olha como o nome da coluna que tem espaços é escrita usando ' '.

Subset dados da who para Brasil

- Vamos selecionar os dados do Brasil.

```
# Examinando o novo data frame
```

```
nrow(who_br)
```

```
## [1] 479
```

```
head(unique(who_br$`City or Locality`))
```

```
## [1] "Americana" "Aracatuba" "Araraquara" "Araucaria" "Barra M"
```

```
## [6] "Bauru"
```

- Olha como o nome da coluna que tem espaços é escrita usando ' '.

Subset dados da WHO para Sao Paulo

```
who_sp <- subset(who,  
  subset = `City or Locality` == "Sao Paulo" ,  
  select = c("Measurement Year",  
             "PM2.5 (µg/m3)",  
             "PM10 (µg/m3)",  
             "NO2 (µg/m3)" ))
```

Subset dados da WHO para Sao Paulo

- Vamos arrumar um pouco o data frame

```
names(who_sp)
```

```
## [1] "Measurement Year" "PM2.5 (µg/m3)"    "PM10 (µg/m3)"     "NO2 (µg/m3)"
```

```
names(who_sp) <- c("year", "pm25", "pm10", "no2")
```

```
names(who_sp)
```

```
## [1] "year" "pm25" "pm10" "no2"
```


Anos acima do standard da WHO

```
plot(who_sp$year, who_sp$no2, t = "l", lwd = 1.5,  
     xlab = "Anos", ylab = "NO2 (ug/m3)")  
abline(h = 40, lwd = 1.5, col = "red")  
legend("topright", col = "red", lwd = 1.5, legend = "NO2 WHO AQS")
```



qualR
