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

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

Bibliotecas

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

```
# Instalando Rmarkdown
install.packages("rmarkdown")
```

Bibliotecas

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

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

Do Excel para R

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

Base de dados da WHO

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

Base de dados da WHO

##

##

\$ City or Locality

\$ DM25 temporal coverage (%)

· Este é um exemplo de nomes de coluna complicados.

```
str(who)
```

```
## Classes 'tbl_df', 'tbl' and 'data.frame': 32191 obs. of 15 varia
## $ WHO Region : chr "Eastern Mediterrane
## $ TSO3
```

\$ ISO3 : chr "AFG" "ALB" "ALB" "A
\$ WHO Country Name : chr "Afghanistan" "Alba

: chr "Kabul" "Durres" "Durres

nım

1 Q NIA NIA NIA NIA NIA

\$ 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

Subset data frames

Subset

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

Subset dados da who para Brasil

· Vamos selecionar os dados do Brasil.

· 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

Subset dados da WHO para Sao Paulo

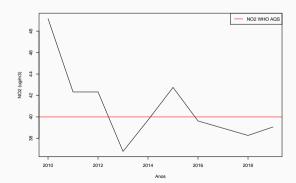
 \cdot Vamos arrumar um pouco o data frame

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

```
names(who_sp)

## [1] "Measurement Year" "PM2.5 (μg/m3)" "PM10 (μg/m3)" "NO2 (μg/m3)" "NO2 (μg/m3)" "names(who_sp) <- c("year", "pm25", "pm10", "no2")
names(who_sp)</pre>
```

Anos acima do standard da WHO



Exercicio

· Que cidades ultrapassarm o Padrão de qualidade do ar de NO2 no ano 2019?

qualR

qualR

- · Pacote desenvolvido para baixar os dados da CETESB dentro do R.
- · Gera dataset prontos e completos para análise:
 - · Horários faltantes preenchidos com NA.
 - · Coluna date tipo POSIXct para usar diretamente com openair.
- · Precisa de ter cadastro na plataforma QUALAR da CETESB.
- · Referencia qualR