

Lista 7 - Introdução a Análise de Dados

Análise de Dados

Guilherme Masuko

May 2023

Para essa lista vamos analisar o PIB per capita de alguns países. Utilizaremos os dados do banco mundial para isso. Essa base de dados está vinculada ao R através do pacote `WDI`¹.

Para acessar os dados, precisamos instalar e chamar o pacote. Os dados que queremos estão armazenados pelo `indicador = 'NY.GDP.PCAP.KD'`. O parâmetro `country` recebe as siglas dos países que estamos interessados em analisar o PIB per capita. `extra = TRUE` adiciona algumas colunas nesse dataframe. `start` e `end` referenciam o intervalo temporal dos dados. A seguir o script.

```
install.packages('WDI')
library(WDI)

help(WDI)

data = WDI(indicator='NY.GDP.PCAP.KD',
           country=c('CA', 'US', 'BR', 'AR', 'CL', 'CO', 'PY', 'PE',
                    'UY', 'FR', 'DE', 'IT', 'ES', 'GB', 'DK', '1W'),
           extra = TRUE,
           start=1960,
           end=2021)
View(data)
```

Questão 1

Para as três regiões (`region`) às quais temos dados nessa amostra, calcule as estatísticas, quantidade de países, PIB per capita médio, máximo e mínimo, referentes ao ano de 2021.

¹<<https://www.r-project.org/nosvn/pandoc/WDI.html>>

Questão 2

Compute as estatísticas média, mínimo e máximo, para cada país, durante os períodos:

- a) Todo o período da amostra.
- b) 1960-1980.
- c) 1980-2000.
- d) 2000-2021.

Questão 3

Crie um `data.frame` para cada país (cada um com o nome do país, tudo em lower case), contendo apenas as colunas `country`, `year`, `NY.GDP.PCAP.KD`.

Questão 4

Crie uma função que recebe um dataframe (no padrão da questão 3) como parâmetro. Essa função deve fazer as seguintes manipulações nesse dataframe:

- Renomear a coluna `NY.GDP.PCAP.KD` para o nome do país do respectivo dataframe.
- Manter somente as colunas `year` e a (agora) do nome do país.

Use a função para alterar todos os dataframes dos países.

Questão 5

Crie um novo dataframe chamado `canada_` a partir do drop das linhas onde o PIB per capita do Canadá é NA. Crie um novo dataframe chamado `brazil_` a partir do drop das linhas onde o PIB per capita do Brasil é maior que 8000.

Faça a união desses dois dataframes, `canada_` e `brazil_`, das seguintes maneiras:

- a) Contendo somente as linhas onde os dois dataframes contenham dados.
- b) Contendo todas as linhas onde o dataframe `canada_` contenha dados.
- c) Contendo todas as linhas onde o dataframe `brazil_` contenha dados.
- d) Contendo todas as linhas onde pelo menos um dos dois dataframes contenham dados.

Questão 6

Una todos dataframes. Renomeie as colunas dos países com nomes compostos, alterando o espaço entre os nomes por um underline "_".

Faça um gráfico apresentando a série temporal do PIB per capita para cada país, um para cada região.

- a) América Latina.
- b) América do Norte.
- c) Europa.