

Lista 8 - Introdução a Análise de Dados

Análise de Dados

Guilherme Masuko

May 2023

Para essa lista vamos analisar a população (total e urbana) 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 = `c("total_pop"="SP.POP.TOTL", "urban_pop"="SP.URB.TOTL")`. O parâmetro `country` recebe as siglas dos países que estamos interessados em analisar o PIB per capita. `start` e `end` referenciam o intervalo temporal dos dados. A seguir o script.

```
library(tidyverse)
library(WDI)

wdi = WDI(indicator = c("total_pop"="SP.POP.TOTL",
  "urban_pop"="SP.URB.TOTL"),
  country = c("CA", "US", "BR", "AR", "CL", "CO", "PY", "PE",
    "UY", "FR", "DE", "IT", "ES", "GB", "DK", "1W"),
  start=1960,
  end=2020)

View(wdi)
```

Questão 1

Crie as seguintes colunas:

- a) Taxa de população urbana.
- b) Taxa de crescimento da população (total e urbana).

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

Questão 2

Crie uma coluna contendo a região (continente) de cada país (crie utilizando código, sem utilizar o parâmetro `extra = TRUE`).

Questão 3

Calcule as estatísticas média, mínimo e máximo para as seguintes variáveis.

- a) Taxa de população urbana agrupados por região para o ano de 2020.
- b) Taxa de crescimento da população total agrupados por região para o ano de 2010.
- c) Taxa de crescimento da população urbana agrupados por região para o ano de 2016.

Questão 4

Calcule a média, mínimo e máximo da taxa de população urbana, taxa de crescimento da população total e taxa de crescimento da população urbana, para cada país durante todo o período que temos na amostra.

Questão 5

Calcule as médias de cada uma das variáveis abaixo agrupados por país (o resultado será uma média para cada país, assim como na questão anterior). A partir desse resultado, calcule o máximo e o mínimo dessas médias agrupados por região.

- a) Taxa de população urbana.
- b) Taxa de crescimento da população total.
- c) Taxa de crescimento da população urbana.

Questão 6

Crie um `data.frame` para cada país (cada um com o nome do país, tudo em lower case), contendo apenas as colunas `country`, `year` e a coluna contendo as informações sobre a taxa de população urbana.

Questão 7

Crie uma função que recebe um dataframe como parâmetro. Essa função deve fazer as seguintes manipulações nesse dataframe:

- Renomear a coluna contendo as informações sobre a taxa de população urbana 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 8

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 da taxa de população urbana para cada país, um para cada região.

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