

### Atividade 3

A desnutrição é um problema grave que torna crianças vulneráveis a doenças e à morte. Existem quatro tipos gerais de desnutrição; em inglês, são chamados de *wasting*, *stunting*, *underweight* e *overweight*.

O banco de dados a ser analisado corresponde a *porcentagens médias anuais* de crianças com menos de 5 anos que possuem determinado tipo de desnutrição para 50 países. Os dados são da UNICEF e correspondem aos anos de 1983 a 2019, mas nem todos os países foram pesquisados em todos os anos. A porcentagem média anual para um país foi calculada considerando-se os anos para os quais se tem dados para ele.

Os dados estão no arquivo `malnutrition.csv` e as variáveis disponíveis são:

- **Country:** país;
  - **Income.classification:** 0 para renda baixa, 1 para renda média baixa, 2 para renda média alta, 3 para renda alta;
  - **Severe.wasting:** porcentagem de crianças com menos de 5 anos abaixo de menos três desvios padrões a partir da razão *peso por altura* mediana;
  - **Wasting:** porcentagem de crianças com menos de 5 anos abaixo de menos dois desvios padrões a partir da relação *peso por altura* mediana;
  - **Overweight:** porcentagem de crianças com menos de 5 anos acima de mais dois desvios padrões a partir da razão *peso por altura* mediana;
  - **Stunting:** porcentagem de crianças com menos de 5 anos abaixo de menos dois desvios padrões a partir da razão *altura por idade* mediana;
  - **Underweight:** porcentagem de crianças com menos de 5 anos abaixo de menos dois desvios padrões a partir da razão *peso por idade* mediana.
1. Importe a base de dados no arquivo para o R e crie um *data frame* com as cinco variáveis correspondentes às porcentagens de desnutrição e nomeie suas linhas com os nomes dos países. Em seguida, faça uma análise de componentes principais para estas variáveis. Investigue (i) a porcentagem da variabilidade explicada por cada componente e (ii) faça um *biplot* para investigar quais variáveis são mais correlacionadas e como elas contribuem para as componentes principais. Investigue também quais países são mais parecidos de acordo com as variáveis analisadas.

2. Faça um *biplot* em que os países são identificados (cor e/ou tipo de ponto) de acordo como a renda. Você consegue observar alguma relação entre os países mais parecidos segundo a análise de componentes principais e sua categoria de renda?
3. Considerando as mesmas variáveis da análise anterior, faça um dendrograma para classificar os países em grupos (*clusters*). Experimente fazê-lo colorido selecionando diferentes números de grupos. Os resultados são condizentes com os da análise de componentes principais?
4. Ainda considerandos as mesmas variáveis das análises anteriores, utilize o algoritmo *k*-means para classificar os países em grupos. Experimente diferentes números de grupos e faça gráficos em que as observações são plotadas nas coordenadas das componentes principais e os grupos estejam identificados. Os resultados são condizentes com os das análises anteriores?

*A atividade deve ser feita no R Markdown e entregue em HTML ou PDF com os códigos explicitados e as análises comentadas. Os gráficos devem ser explicativos, com nomes corretos nos eixos, por exemplo.*