

# Introdução à Análise de Dados com Linguagem R

Este curso de introdução à análise de dados com linguagem de programação R se destina à capacitação de analistas ambientais no Núcleo de Biodiversidade do IBAMA no estado Pará, visando análise exploratória e visualização de dados ambientais.

Neste módulo o aluno dominará os conceitos básicos de codificação visando análise de dados com linguagem R. A escolha da linguagem se dá pelo fato de existirem milhões de usuários de R em todo o mundo, consolidando-a como uma linguagem de programação de destaque na área da estatística e ciência de dados.

A importância do uso de uma linguagem de programação para análise de dados no IBAMA se dá pelo seguinte fato: O IBAMA dispõe de mais de uma dezena de sistemas que geram um volume de dados que impede o uso de planilhas para processamento e análise de dados. Ao final do curso espera-se que os analistas melhorem a qualidade e tempo de análises de dados, assim como consigam obter dados de diferentes fontes.

O curso está estruturado em três módulos: 1 - Introdução à Linguagem R; 2 - R Intermediário e 3 - Importação, Exportação e Manipulação de Dados em R e 4 - Visualização Gráfica de Dados com o pacote ggplot2.

## Módulo 1 - Introdução à Linguagem R

### 1.2 Apresentação da Linguagem R e IDE RStudio

O aluno se familiarizará com a linguagem através do Ambiente de desenvolvimento integrado (IDE) *RStudio*. Aprenderá funções básicas da própria linguagem e instalação de pacotes e obtenção de ajuda na IDE.

### 1.3 Objetos, Operadores de Atribuição, Operadores Aritméticos e Lógicos em R

Será demonstrado o uso dos operadores de atribuições à objetos R, operadores aritméticos e aplicação das principais funções nativas do R para operações aritméticas e lógicas.

- Soma, Divisão, Multiplicação, Divisão Inteira, Resto da Divisão
- Raízes e logaritmos

### 1.4 Estrutura de Dados em R

Neste capítulo serão abordados os diferentes tipos de estrutura de dados em R, a saber: vetor, matriz, lista, fator e *dataframe*.

- Vetores
  - Tipos: vetor numérico, Vetor lógico e Vetor de caracteres
  - Operações com vetores
  - Funções aplicadas: `class()`, `length()`, `attributes()`, `names()`, `setNames()`, `mean()`, `max()`, `min()`, `seq()`, `rep()`, `any()`, `all()`, `head()`, `tail()`, `is.na()` etc...
  - Coerção de vetores
- Manipulação de Matrizes
- Funções aplicadas: `dimnames()`, `apply()`
- Manipulação de Listas
  - Funções aplicadas: `lapply()`, `sapply()`, `mapply()`
- Criação, importação e manipulação de Dataframes

## 1.6 Estudo de Caso

O objetivo do módulo 1 será a análise de dois vetores, sendo um contendo dados de um inventário florestal e outro com dados meteorológicos.

## Módulo 2 - R Intermediário

Neste módulo o aluno irá se familiarizar com o uso de operadores lógicos e com as estruturas de decisão `if`, `if else` e `else`; e as estruturas de repetição ou fluxo de controle (*Loop*): `for` e `while` para iterar sobre vetores e outros objetos em R visando análise de dados. Também será apresentado o uso de função para otimizar rotinas repetitivas, assim como a definição e manipulação de objetos do tipo data.

### 2.1 Operadores Lógicos em R

Operador	Descrição	Explicação	:-----: :-----: :-----:	&  AND lógico
		Comparação pareada de todos os elementos de dois ou mais vetores	&&  AND lógico	
		Compara apenas os dois primeiros elementos de vetores	OR lógico         OR lógico	
		!   NOT lógico   Negação lógica.     xor   OR exclusivo   Ou Exclusivo. Retorna valor lógico TRUE se ambos os valores de entrada forem diferentes entre si, e retorna FALSE se os valores forem iguais.		

### 2.2 Uso do if, ifelse e else

O aluno irá aprender sobre o uso das instruções `if`, `ifelse` e `else` para adicionar fluxo lógico extra à análise de dados.

### 2.3 Loops - Fluxos de Controle

Os *loops* são úteis para fazer a mesma operação em cada elemento de sua estrutura de dados. Neste capítulo, o aluno aprenderá os *loops* `while` e `for`.

## 2.4 Funções em R

Será apresentado uma introdução à syntax para criação de função em R para automatizar rotinas na manipulação de dados.

# Módulo 3 - Importação, Exportação e Manipulação de Dados em R

O aluno aprenderá como importar dados nos principais formatos de arquivos utilizados para armanezar conjuntos de dados, a exemplo de: `.xlsx`, `.csv`, `tsv`, `.txt`, `.RDS`, `.RData`, `.fst`, `.dbf`, `.sas7bdat`, `.dta` e `.sav`; além de dominar habilidades para processamento e manipulação de *dataframes* utilizando apenas comandos R e os principais verbos dos pacotes `dplyr`, `tidyverse` e `lubridate`. O aluno aprenderá a ler dados diretamente de *sites* que dispõe dados, facilitando a atualização destes quando da análise. Será dado ênfase em como acessar dados do Portal Brasileiro de Dados Abertos, o qual possui dados do IBAMA, dentre outros órgãos públicos.

## 3.1 Limpeza de Dados

Dados apesar de muitas proverem de bases de dados, os as tabelas possuem estruturação de seus dados, algumas vezes essas informações são alimentadas de forma que necessitem de algum processamento para permitir análise. Outros dados provém de fontes não estruturadas como simples planilhas ou mesmo conjunto de dados de *datalake*, o que não exige uma formatação de dados rigorosamente padronizada, necessitando de uma processamento de dados maior para análise dos dados.

## 3.2 Joining Data

O aluno aprenderá como juntar diferentes conjuntos de dados visando cruzamento de informações de interesse.

# 4 Visualização Gráfica de Dados com o pacote ggplot2

Neste módulo o aluno aprenderá que tipo de gráfico utilizar de acordo com o tipo de variáveis de seu conjunto de dados. Será utilizado o pacote `ggplot2` para geração e interpretação de gráficos do tipo:

- *Boxplot* - Utilizado para interpretar resultdo de estatística descritiva de conjunto de dados.
- *Histogram* - Utilizado para conhecer a distribuição de uma variável de interesse.
- *Scatterplot* - Utilizado para avaliar a correlação entre duas variáveis.

- *Barplot* - Utilizado para analisar analisar variáveis discretas.
- *Lineplot* - Utilizado para analisar o comportamento de uma varipavel ao longo de um determinado intervalo de tempo.

In [ ]: