

R para Ciência de Dados 1

Introdução ao curso



Março de 2022

Sobre a Curso-R

A empresa



Filosofia de código aberto!

Livros

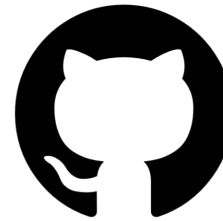


Lives



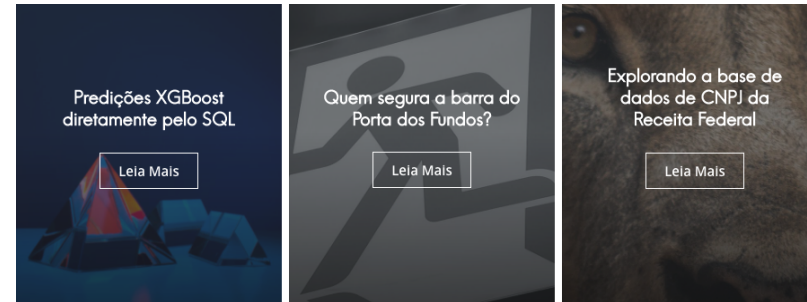
TODA
QUARTA,
ÀS 20H

Material dos cursos



Confira o
nosso GitHub

Blog



Nossos cursos

Programação para Ciência de Dados

Introdução a programação com R

R para Ciência de Dados I

R para Ciência de Dados II

Pacotes

Python para quem usa R

Web scraping

Web scraping

Dashboards e Visualização de Dados

Visualização de dados

Relatórios e apresentações

Dashboards I

Deploy

Dashboards II

Modelagem de Dados

Modelos Lineares

Introdução ao Machine Learning

Séries Temporais

Não supervisionado

Sobre o curso

Dinâmica curso

- As aulas serão em maior parte expositivas.
- As aulas terão uma seção teórica, de exposição de conceitos, e prática, de aplicação de conceitos.
- O objetivo dos exercícios é gerar dúvidas. **Com exceção do trabalho final, nenhum exercício precisa ser entregue.**
- O certificado será emitido mediante uma **entrega final**, a ser especificada nas últimas aulas do curso.
- Haverá monitoria para esclarecimento de dúvidas sempre 30 minutos antes do início das aulas.
- Usaremos os últimos minutos de cada aula para tirar dúvidas do conteúdo apresentado. Não haverá plantão de dúvidas pós aula.
- A gravação das aulas ficará disponível no Google Classroom por 1 ano após o final do curso.

Dinâmica das aulas

- Mande dúvidas e comentários no chat em qualquer momento.
- Para falar, levante a mão.
- Algumas dúvidas serão respondidas na hora. Outras serão respondidas mais tarde na própria aula ou em aulas futuras.
- Só interrompa quem estiver falando em caso de emergência.

Tire suas dúvidas

- Não fique com dúvida.
- Fora do horário de aula ou monitoria:
 - envie suas perguntas gerais **sobre o curso** no Classroom.
 - envie preferencialmente suas perguntas **sobre R** no [nosso discourse](#).
- Saber fazer a pergunta certa vai te ajudar bastante nos estudos de programação. [Veja aqui dicas de como fazer uma boa pergunta](#).

Abrace o seu sotaque

Assim como quando estamos aprendendo um novo idioma, **sempre levamos um pouco de nós para a linguagem de programação** e escrevemos códigos com o nosso *sotaque*.

Normalmente, existem várias formas de resolver um problema. Não existe o *código certo* e, principalmente no começo, não se preocupe em achar o código mais rápido, curto ou elegante.

No dia-a-dia, o melhor código é aquele que funciona. A eficiência e a elegância aparecem naturalmente com a experiência.

Dito isso, atente-se à diferença entre **regras**, **boas práticas** e **estilos**.

- **Regras:** devem ser seguidas para o código funcionar (sintaxe, vocabulário).
- **Boas práticas:** recomenda-se seguir para a criação de códigos legíveis (espaçamento, nomes, organização).
- **Estilo:** cada um pode escolher aquele com o qual se sentir mais à vontade (tipos de indentação, formatação).

Introdução

Por que analisar dados usando linguagens de programação?

Vantagens evidentes

- Linguagens de programação são gratuitas



- As melhores ferramentas de extração e manipulação de dados, modelagem e visualização são desenvolvidas primeiro dentro das principais linguagens de programação

Mas são dois os motivos pelos quais a gente realmente deve aprender a programar

Reprodutibilidade e Comunidade

Reprodutibilidade: o código é uma linguagem

Podemos documentar a nossa análise

```
image_read("https://jeroen.github.io/images/frink.png") %>%  
  image_rotate(270) %>%  
  image_background("blue", flatten = TRUE) %>%  
  image_border("red", "10x10") %>%  
  image_annotate("Linguagens de programação ainda\nsão linguagens!",  
    color = "white", size = 30)
```

Entrada



Saída



Análises reprodutíveis...

- são essenciais na Ciência.
- são transparentes, algo muito importante no mercado e no setor público para evitar fraudes e má-fé.

poupam o nosso tempo.

Comunidade: todo mundo junto!

- As pessoas se ajudam tirando dúvidas e disponibilizando exemplos e análises gratuitamente.
- A linguagem cresce com a comunidade. Ao aprender a programar, você será uma desenvolvedora ou um desenvolvedor em potencial.
- Espaços seguros e de apoio a grupos sub-representados.



Por que o R?

O ciclo da ciência de dados

Importar



Arrumar

(Armazenar os dados
consistentemente)



Transformar

(Criar novas variáveis e
agregações)

Visualizar

(Surpreende, mas não é
escalável)

Modelar

(É escalável, mas não
surpreende)

Comunicar

Automatizar

O ciclo da ciência de dados com o R

