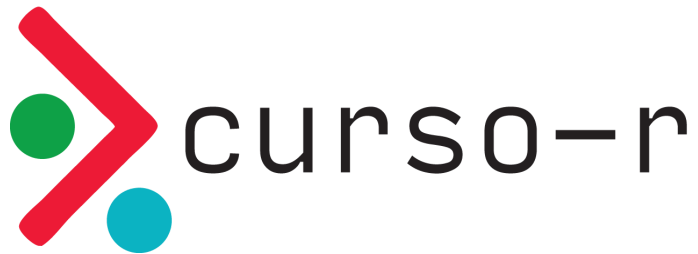


R para Ciência de Dados I



Sobre a Curso-R

A empresa



www.curso-r.com

Nossos cursos

Programação em R

Introdução à Programação em R

R para Ciência de Dados I

R para Ciência de Dados II

Pacotes

Introdução ao R com C++

Visualização de dados

Relatórios e Visualização de Dados

Dashboards

Deploy

Modelagem

Regressão linear

Machine Learning

XGBoost

Deep Learning

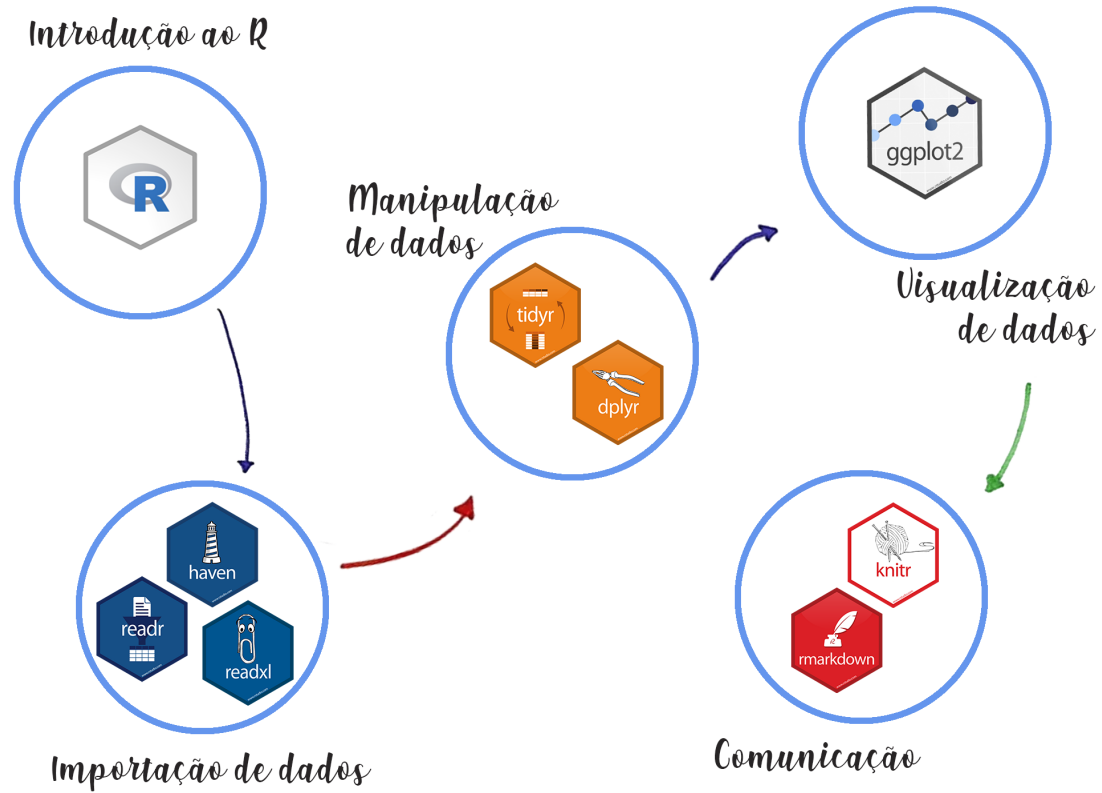
Extração de dados

Faxina de Dados

Web scraping

Sobre o curso

Conteúdo



Material

Nosso livro Ciência de Dados em R:

<https://livro.curso-r.com/>

Também temos um blog:

<http://curso-r.com/blog/>



Referência

Disponível online e gratuitamente em [Inglês](#) e [Espanhol](#).



Dinâmica do curso

- As aulas terão uma seção teórica, de exposição de conceitos, e prática, de aplicação de conceitos.
- Teremos diversos exercícios para serem feitos "em casa".
- O objetivo dos exercícios é gerar dúvidas. **Com exceção do trabalho final, nenhum exercício precisa ser entregue.**
- O certificado será emitido mediante um **trabalho final**. O prazo da entrega final será de pelo menos 2 semanas após a última aula.
- Haverá monitoria para esclarecimento de dúvidas sempre 30 minutos antes do início das aulas.

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.

Como estudar?

Pratique

- Não se preocupe com estilos de estudo. Use o que te deixar mais confortável.
- Estude um pouco todo dia. Se você tem apenas uma hora para estudar na semana, faça 3 sessões de 20 minutos em dias diferentes.
- Se teste! Coloque em prática o que você aprendeu resolvendo exercícios ou inserindo o R no seu dia-a-dia.

Referência

Nahurhodo 205 - Powerpoint é útil para a aprendizagem?

Tire suas dúvidas

- **Não existe dúvida idiota.**
- Fora do horário de aula ou monitoria:
 - envie suas perguntas gerais sobre o curso no Classroom.
 - envie preferencialmente suas perguntas sobre R, principalmente as que envolverem código, no **nosso discourse**.
- **Veja aqui dicas de como fazer uma boa pergunta.**

Por que usar o discourse?

- Muito melhor para escrever textos que possuem códigos. Com ele, podemos usar o pacote `{reprx}`!
- Mais pessoas acompanhando e respondendo as dúvidas.
- Em um ambiente aberto, as suas dúvidas vão contribuir com a comunidade.

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

Importar

readr
xml2
haven
readxl

Arrumar

tidyr
janitor
tibble

%>%
magrittr

Transformar

dplyr
stringr
forcats
lubridate
data.table

Visualizar

ggplot2

Modelar

caret
mlr
lme4
keras

Comunicar

rmarkdown shiny

Automatizar

plumber opencpu