

R para Ciência de Dados I



Sobre a Curso-R

Curso-R



Athos Damiani

*Curso-R
Bacharel em
Estatística*



William Amorim

*Curso-R
Doutor em
Estatística*



Fernando Corrêa

*Curso-R
Mestrando em
Estatística*



Julio Trecenti

*Curso-R, Terranova,
ABJ, Conre
Doutorando em
Estatística*



Daniel Falbel

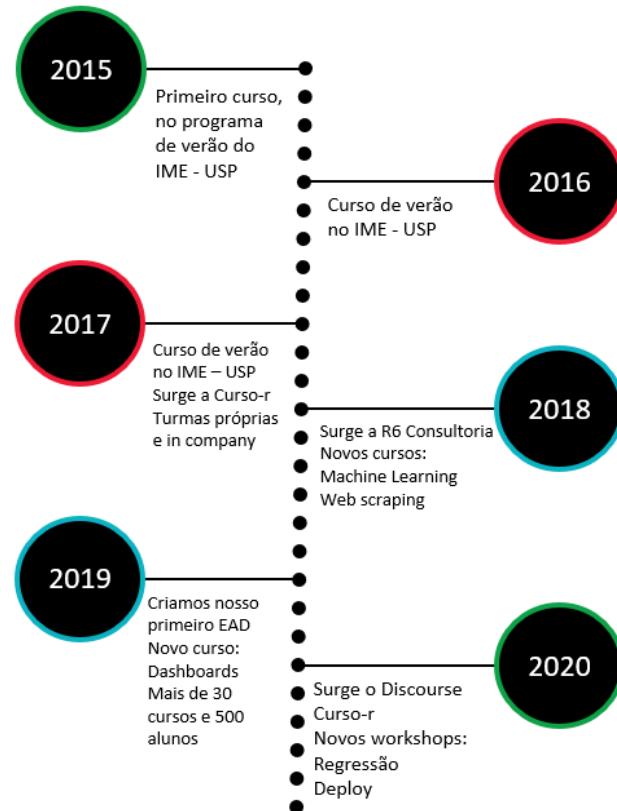
*Curso-R e RStudio
Bacharel em
Estatística*



Caio Lente

*Curso-R, Terranova,
ABJ, Mestrando em
ciências da
computação*

Linha do tempo



Nossos cursos

Programação em R

R para Ciência de dados I

R para Ciência de dados II

Introdução ao R com C++

Extração de dados

Web scraping

Modelagem

Regressão Linear

Machine Learning

XGBoost

Deep Learning

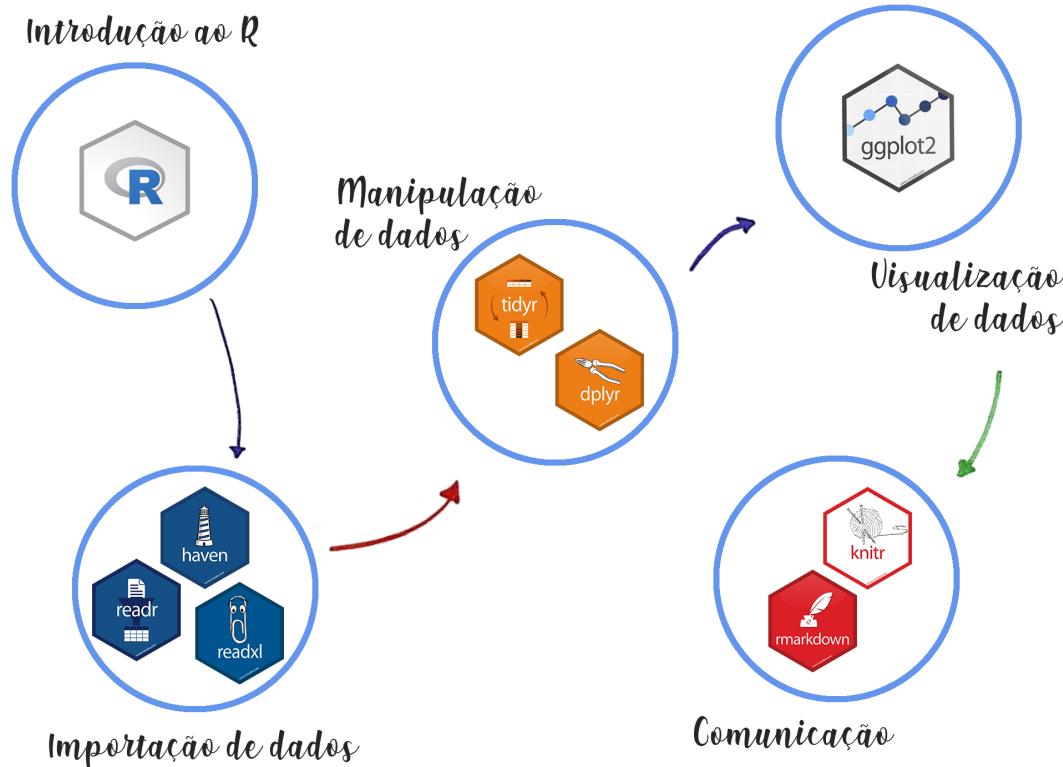
Comunicação e automação

Dashboards com R

Deploy

Sobre o curso

Conteúdo



Introdução ao R

- R e RStudio
- Trabalhando com projetos
- Objetos e funções
- Pacotes

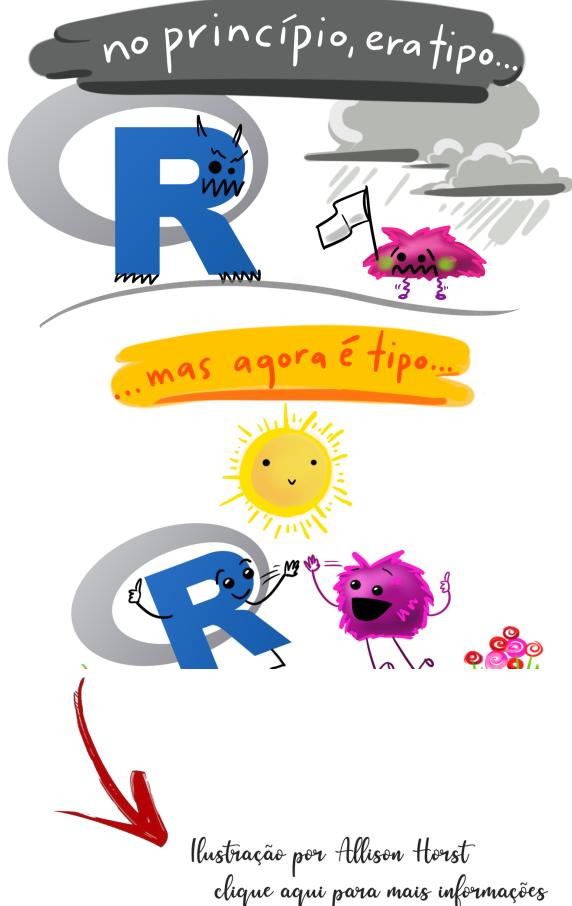


Ilustração por Allison Horst
[clique aqui para mais informações](#)

Importação

- Arquivos de texto
- Arquivos do excel
- Arquivos de outros softwares (SAS, SPSS)
- Conexão com banco de dados



Manipulação

- Selecionar colunas
- Filtrar linhas
- Criar ou modificar colunas
- Agrupar e summarizar
- Juntar tabelas

dplyr : go wrangling



Ilustração por Allison Horst
[clique aqui para mais informações](#)

Visualização



Comunicação

- Relatórios automatizados
- Dashboards estáticos



Ilustração por Allison Horst
[clique aqui para mais informações](#)

Material

Temos um material em português disponível nesta página:

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

Também temos um blog:

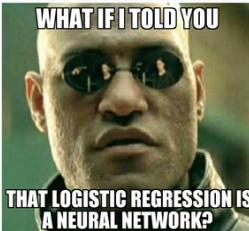
<http://curso-r.com/blog/>



PROGRAMANDO COM A COMUNIDADE

POR JULIO EM 19/08/2019

Trabalhar em grupo é difícil, mas os esforços se pagam. Aqui discuto alguns aspectos importantes para tirar proveito da melhor parte do R: a comunidade (Imagen de Allison Holst).



REGRESSÃO LOGÍSTICA EM: A MENOR DEEP LEARNING DO MUNDO

POR ATHOS EM 29/07/2017

Deep learning é uma grande novidade, mas há muitos mitos em sua volta que acaba espantando pessoas que poderiam usufruir e contribuir para a comunidade de analista de dados. Esse post tem o intuito de abrir um pouquinho a caixa preta para quebrar algumas dessas fábulas.

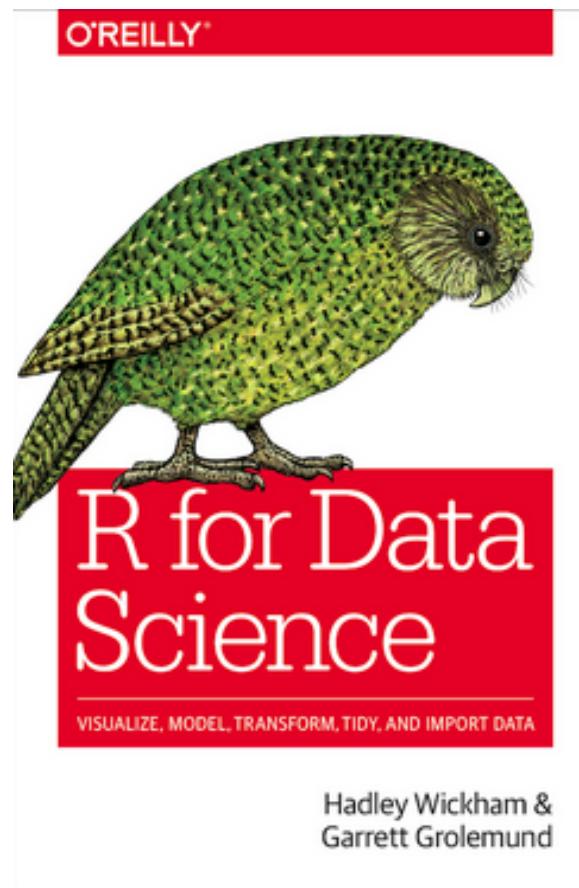


AS CORES DA MARVEL VS DC

POR WILLIAM EM 01/05/2017

A cor é uma diferença notável entre os filmes da Marvel e da DC. Enquanto a Disney/Marvel Studios costuma lançar filmes com tons mais claros e alegres, a Warner tem optado por cenários escuros, com um aspecto mais sombrios. Neste post, vamos utilizar o Kmeans para obter a paleta de cores mais utilizadas por cada estúdio.

Referência



Introdução

As vantagens de analisar dados usando linguagens de programação

As principais linguagens de programação para ciência de dados são
Código Aberto

O que significa ser código aberto?

1. Acesso gratuito.
2. Todos podem usar as melhores ferramentas independentemente do poder financeiro.
3. Estudantes podem usar as mesmas ferramentas que os profissionais.
4. Você pode corrigir problemas e aprimorar a linguagem.
5. Você pode desenvolver suas próprias ferramentas.
6. Possibilita a existência de uma comunidade ativa.

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_annotation("Linguagens de programação ainda\nsão linguagens!",  
                   color = "white", size = 30)
```

Entrada



Saída



O código é texto

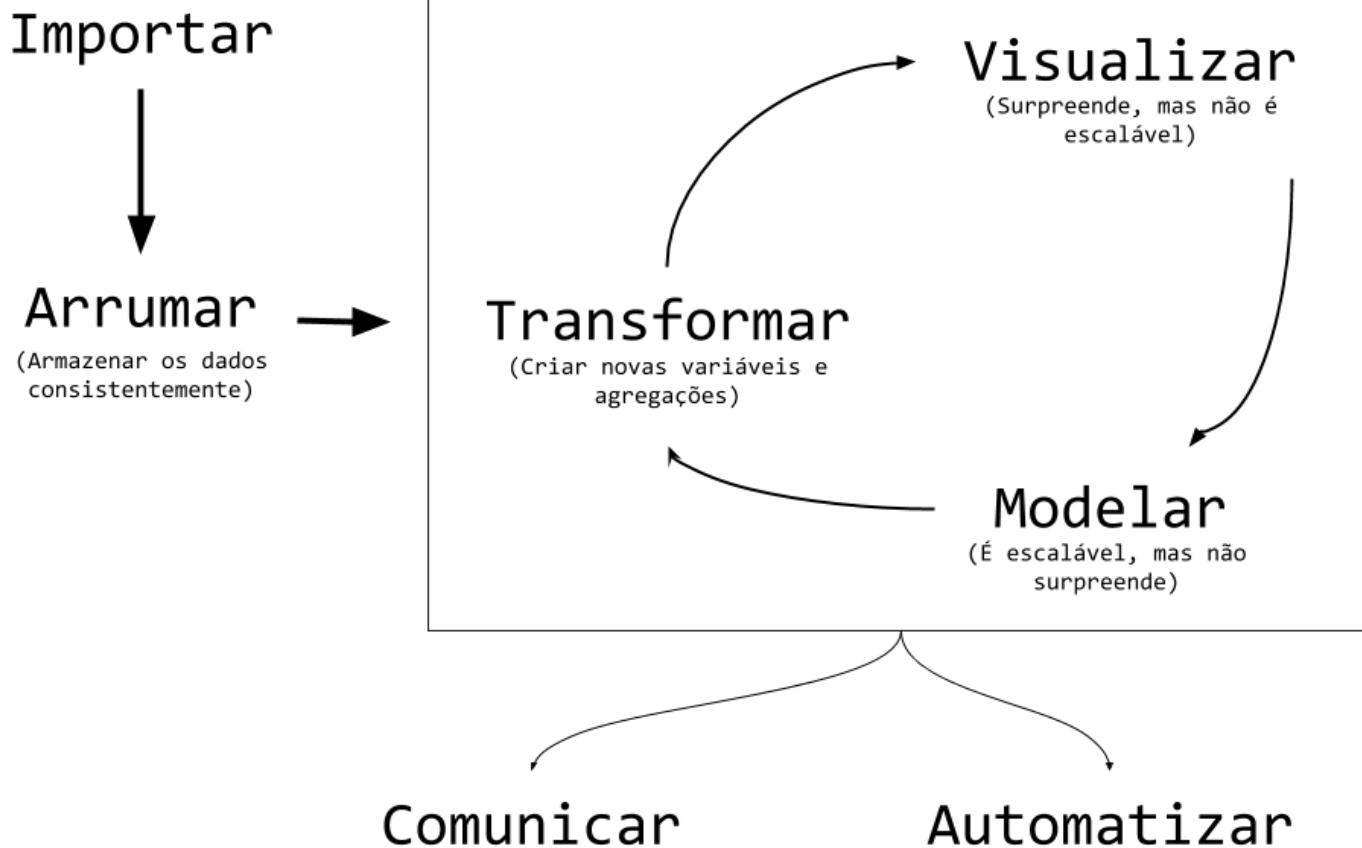
Podemos copiar e colar



Introdução

Por que o R?

O ciclo da ciência de dados



O ciclo da ciência de dados com o R

Importar

readr
xml2
haven
readxl

Arrumar →

tidyverse
janitor
tibble

%>%
magrittr

Transformar

dplyr
stringr
forcats
lubridate
data.table

Visualizar

ggplot2

Modelar

caret
mlr
lme4
keras

Comunicar

rmarkdown shiny

Automatizar

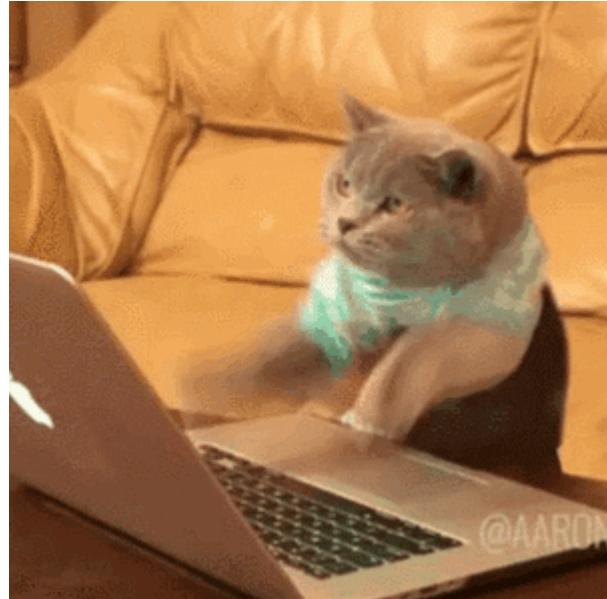
plumber opencpu

Introdução

Guia da(o) mochileira(o)

Pratique

O objetivo do curso é ajudar a quebrar inércia da etapa inicial de aprendizado. Assim como um idioma, o único jeito de aprender uma linguagem de programação é praticando.



Tirando dúvidas

- Não existe dúvida idiota.
- Nem sempre é trivial fazer a pergunta certa para que outra pessoa esclareça a sua dúvida. Neste curso, **vamos mostrar melhores práticas na hora de fazer perguntas sobre programação.**
- Fora do horário de aula ou monitoria:
 - perguntas gerais sobre o curso deverão ser feitas no Classroom.
 - perguntas sobre R, principalmente as que envolverem código, deverão ser enviadas no [nossa 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 `{reprex}`!
- Saber pesquisar sobre erros e fazer a pergunta certa é essencial para aprender e resolver problemas de programação.
- No discourse, teremos mais pessoas acompanhando e respondendo as dúvidas.
- Em um ambiente aberto, as suas dúvidas vão contribuir com a comunidade.

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

Regras, boas práticas e estilo

Na grande maioria das vezes, você poderá resolver uma tarefa de diversas maneiras. Isso porque, apesar de existirem várias **regras**, linguagens de programação também possuem **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)