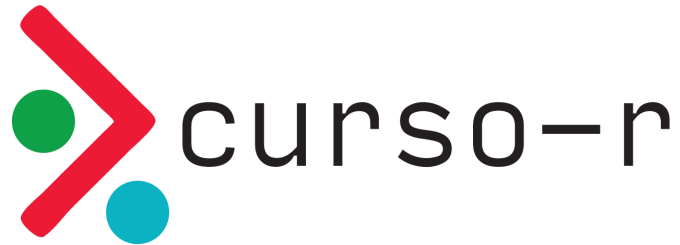


Uma introdução ao universo da Ciência de Dados


Curso-R

2020-12-03

Sobre nós



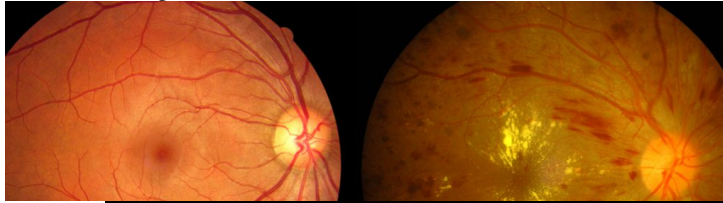
O que é ciência de dados?



Ciência de dados é a disciplina que nos permite transformar dados em intuição, compreensão e conhecimento.

Hadley Wickham

Exemplos



NÚMEROS

A Rosie, nossa Robô, em números

8.276

Reembolsos suspeitos encontrados

[Tuite sobre esse número](#)

735

Deputados diferentes encontrados com gastos suspeitos

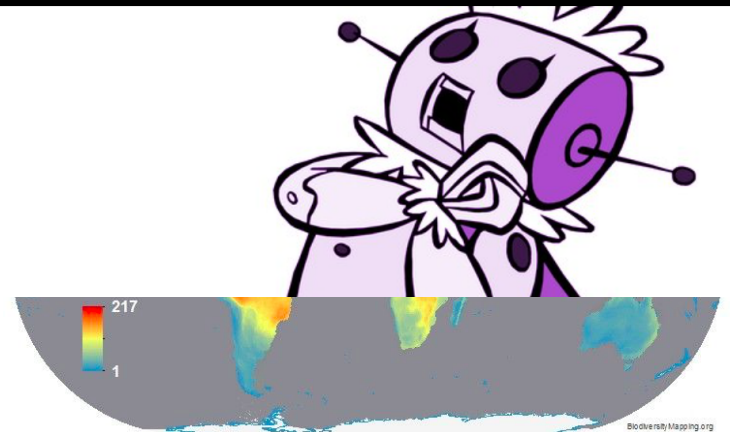
[Tuite sobre esse número](#)

R\$ 3,6 mi

Encontrados em reembolsos suspeitos

[Tuite sobre esse número](#)

Tudo isso encontrado por uma inteligência artificial.



Por que o hype?

Vivemos em uma era cada vez mais digital

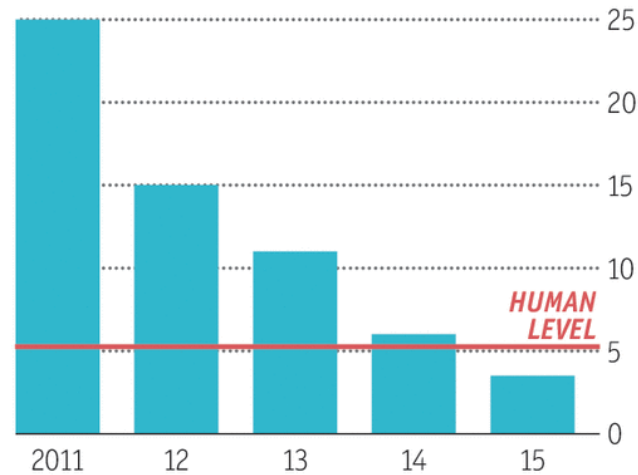
Analisar dados economiza e gera dinheiro



Problemas historicamente muito difíceis foram resolvidos nos últimos anos

Ever cleverer

Error rates on ImageNet Visual Recognition Challenge, %



Sources: ImageNet; Stanford Vision Lab

Economist.com

Perigo do hype

É preciso estudar e se aprofundar

EDUCACAO · ENEM

Filosofia e sociologia obrigatórias derrubam notas em matemática

Segundo pesquisa, exigência no ensino médio prejudica
desempenho de alunos

O marketing pode ser maior do que o produto



Data Science é só hype?

Não

- Todos os setores estão se digitalizando.
- Muito do que foi produzido é realmente útil e não vai deixar de ser usado.
- Ainda tem muito a ser aprimorado.

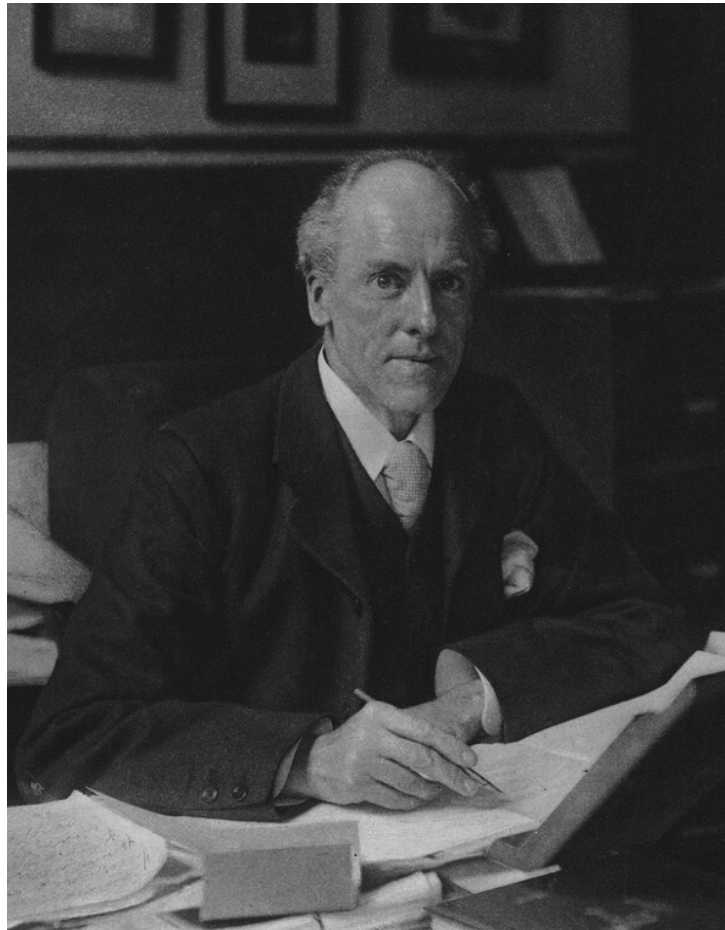
Habilidades importantes de um cientista de dados

- Curiosidade
- Análise crítica
- Conhecer o método científico
- Modelar
- Programar

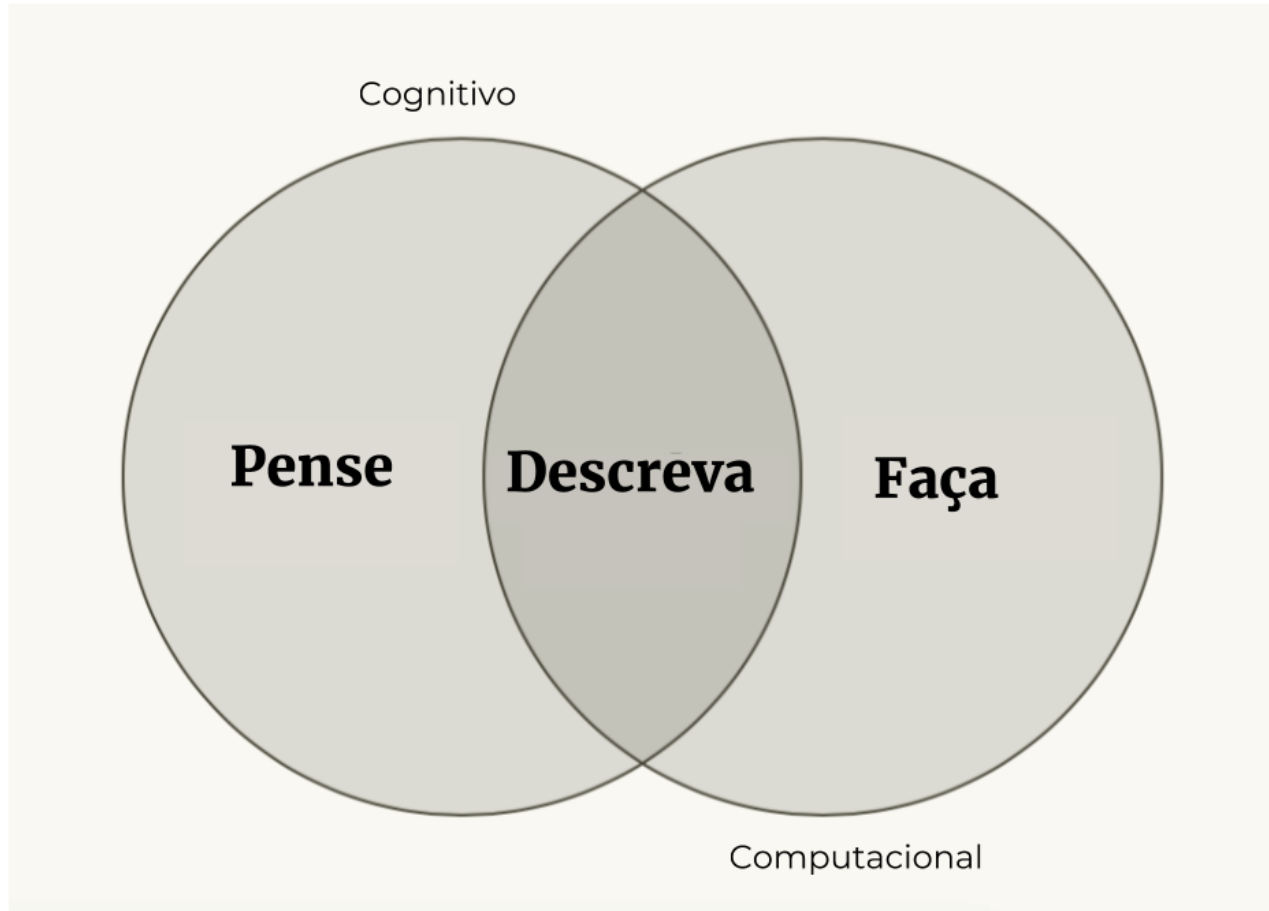


Por que programar?

Fazemos análise de dados no computador



Fazemos análise de dados no computador

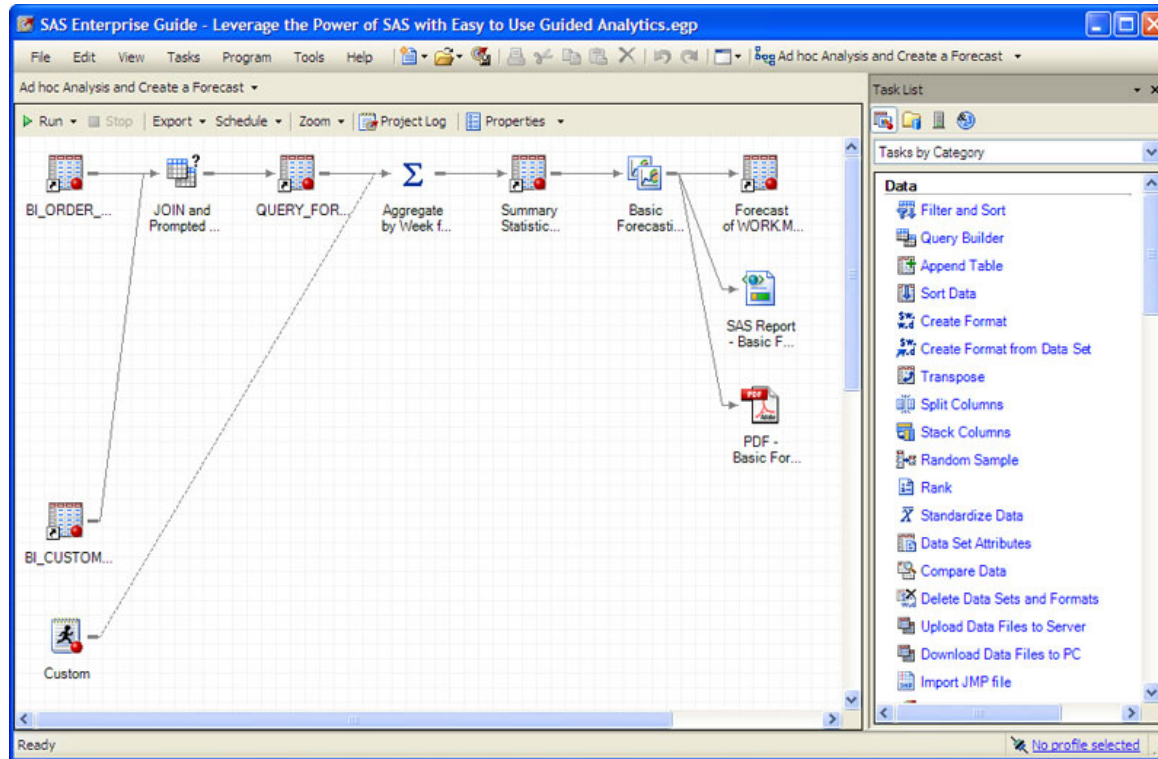


Temos duas principais interfaces: point & click

The screenshot displays the SPSS (Statistical Package for the Social Sciences) software interface. The 'Analyze' menu is open, showing various statistical analysis options. The 'Crosstabs...' option is highlighted. Below the menu, a data table is visible with columns for Name, Type, Values, Missing, Columns, Align, and Measure. The table contains 26 rows of data, including variables like Test_Score, Gender, Apt1, Apt2, Apt3, Apt4, Apt5, Exam, Grade1, Grade2, Grade3, Good1, Good2, Good3, Age, Final_exam, Ex1, Ex4, Ex5, Ex2, Ex3, Treatment, and Gift.

	Name	Type	Values	Missing	Columns	Align	Measure
1	Test_Score	Numeri	None	None	8	Right	Scale
2	Test2_Score	Numeri	None	None	5	Right	Scale
3	Test3_Score	Numeri	None	None	5	Right	Scale
4	Gender	Numeri	0, Male)...	None	8	Right	Nominal
5	Apt1	Numeri	None	None	5	Right	Scale
6	Apt2	Numeri	None	None	5	Right	Scale
7	Apt3	Numeri	None	None	5	Right	Scale
8	Apt4	Numeri	None	None	5	Right	Scale
9	Apt5	Numeri	None	None	5	Right	Scale
10	Exam	Numeri	{0, Fail)...	None	10	Right	Nominal
11	Grade1	Numeri	{1,00, A)...	None	10	Right	Ordinal
12	Grade2	Numeri	{1,00, A)...	None	10	Right	Ordinal
13	Grade3	Numeri	{1,00, A)...	None	10	Right	Ordinal
14	Good1	Numeri	{0, Not goo...	None	8	Right	Ordinal
15	Good2	Numeri	{0, Not goo...	None	8	Right	Ordinal
16	Good3	Numeri	{0, Not goo...	None	8	Right	Ordinal
17	Age	Numeri	None	None	10	Right	Scale
18	Final_exam	Numeri	{1,00, Fail)...	None	12	Right	Ordinal
19	Ex1	Numeri	{1,00, Low ...	None	10	Right	Ordinal
20	Ex4	Numeri	{1,00, Low ...	None	10	Right	Ordinal
21	Ex5	Numeri	{1,00, Low ...	None	10	Right	Ordinal
22	Ex2	Numeri	{1,00, Low ...	None	10	Right	Ordinal
23	Ex3	Numeri	{1,00, Low ...	None	10	Right	Ordinal
24	Treatment	Numeri	{1,00, Front...	None	11	Right	Ordinal
25	Gift	Numeri	{1,00, Super...	None	10	Right	Nominal
26							

Temos duas principais interfaces: **point & click**



e programar

e programar

```
23
24 # Resultado das eleições
25
26 eleicoes_2018 %>%
27   filter(turno = 2) %>%
28   group_by(candidato) %>%
29   summarise(qtd = n(),
30             prop = qtd / sum(qtd))
31
32
33 |
34
```

O que são linguagens de programação?

Uma linguagem de programação é um vocabulário e um conjunto de regras gramaticais que servem para instruir um computador a executar uma tarefa específica.

Por que programar é preferível?

1. Código é texto

2. Código é legível

3. Código é aberto

Código é texto

Copiar & Colar



Stack Overflow

pt.stackoverflow.com

stack overflow em Português Perguntas Tags Usuários

Perguntas da tag info recentes frequentes votos ativas sem resposta

R é um ambiente e linguagem de programação de código aberto para computação estatística, bioinformática e gráficos. Aconselha-se a completar a sua pergunta com um exemplo reproduzível mínimo.

Unwatch Tag Ignore Tag Saiba mais... Improve tag info Top users Sinónimos

0 votos
0 respostas
30 visitas

Erro no R com a função list2env

Tento converter minha lista (que contém 12 dataframes e aprox. 20.000 linhas cada) em dataframes separados: `list2env(mylist,envir=.GlobalEnv)` mas, a seguinte mensagem de erro aparece: `names(x) must ...`

perguntada ontem
Giovani
197 • 9

2 votos
1 resposta
29 visitas

Lapply para retirar resultados de teste ADF no R

Essa é a minha lista: `mylist=dput(mylist) list(list(c(30, 50, 35, 25, 45), c(40, 35, 35, 50, 45), c(40, 20, 40, 50, 25), c(35, 45, 45, 45, 40), c(20, 30, 50, 45, 20), c(40, 40, 50, 30, 40)), ...`

perguntada 2 dias atrás
Laura
57 • 5

1 voto
2 respostas
35 visitas

Converter todos os arquivos do Environment em '.Rda' e, após, carregá-los de uma só vez com um loop

Quero converter TODOS dataframes do meu Environment em arquivo do tipo .Rda. Meu objetivo é, após essa ação, reiniciar o R e carregar todos eles de uma vez com um loop for.

perguntada 2 dias atrás
Giovani

Código é legível

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



Código te dá confiança de que está certo!

The screenshot displays the Microsoft Excel interface with a 'Sort Warning' dialog box open. The dialog box contains the following text:

Sort Warning

Data outside your current selection won't be sorted.

What do you want to do?

- Expand the selection
- Continue with the current selection

Buttons: Cancel, Sort

The background spreadsheet shows a table with the following columns: ANO_MES, GRUPO, EMPRESA, MATRICUL, DEPENDEN, NOM, CIDADE, BAIRRO, CONTRAT/ CATEGORI PLANO, COD PROC, REG PROD SE TEM OU SE TEM OU, ACOMOD/ SEXO, USUARIO, FAIXA ETA, and IDADA. The data rows show various records with numerical values and text labels.

Código é aberto

Todas as linguagens de programação modernas são open-source.

Grátis

- Estudantes podem usar as mesmas ferramentas que os profissionais
- Todos podem usar as melhores ferramentas independente do poder financeiro
- Qualquer um pode reproduzir as suas análises

Aberto

- Você pode corrigir problemas
- Você pode desenvolver suas próprias ferramentas

A linguagem R

- R é uma linguagem de programação
- R foi criado para fazer análise de dados.
- R é Open Source
- R possui uma comunidade muito ativa de desenvolvedores

Por que usar o R?

Ciclo da ciência de dados

Importar



Arrumar

(Armazenar os dados
consistentemente)



Transformar

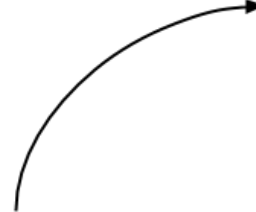
(Criar novas variáveis e
agregações)

Visualizar

(Surpreende, mas não é
escalável)

Modelar

(É escalável, mas não
surpreende)



Comunicar

Automatizar

Ciclo da ciência de dados

Importar

readr
xml2
haven
readxl

Arrumar

tidyr
janitor
tibble

%>%
magrittr

Transformar

dplyr
stringr
forcats
lubridate
data.table

Visualizar

ggplot2

Modelar

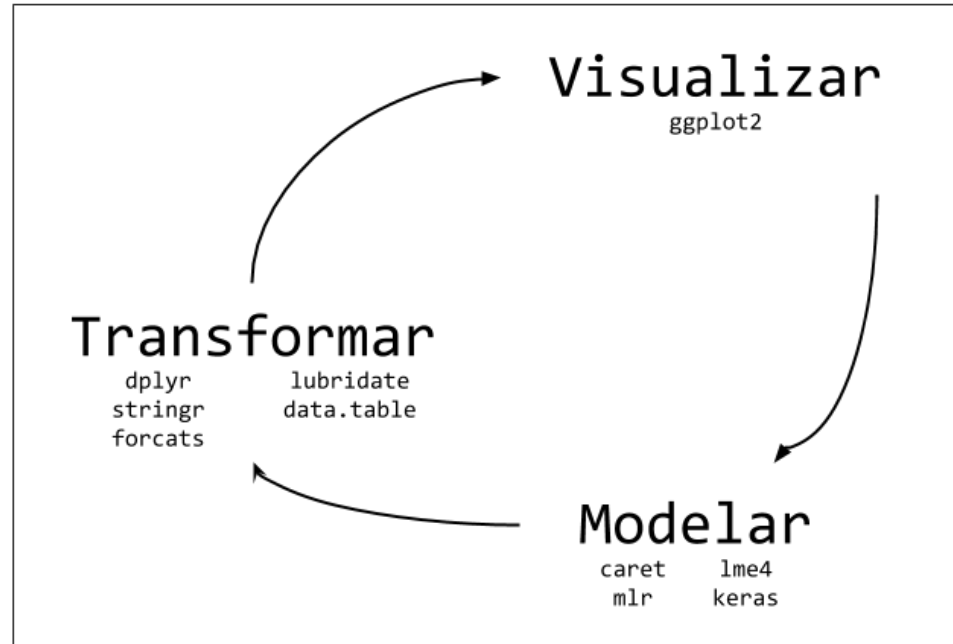
caret
mlr
lme4
keras

Comunicar

rmarkdown shiny

Automatizar

plumber opencpu



Conclusão

- Ciência de dados é muito legal!
- Programação e conhecimento técnico matemático andam lado a lado, é preciso um pouco dos dois.
- Nenhum estudo maravilhoso nasce da noite pro dia. É necessário ter as ferramentas certas e ter o objetivo claro: fazer perguntas que os dados são capazes de responder.
- A jornada é mais importante do que o destino: As aplicações precisam de rigor técnico e clareza nos objetivos, mas é muito mais importante é acertar no fluxo de produção dos resultados e no passo-a-passo.

Nossos cursos

Programação em R

Introdução a programação em R

R para Ciência de dados I

R para Ciência de dados II

Pacotes

Modelagem

Regressão Linear

Machine Learning

XGBoost

Deep Learning

Extração de dados

Web scraping

Comunicação e automação

Dashboards com R

Deploy

