

Data Science Academy

R - Fundamentos para Análise de Dados



Data Science Academy

www.datascienceacademy.com.br



Seja Bem-Vindo



Data Science Academy



Data Science Academy



Data Science Academy

- Introdução à Ciência de Dados
- Big Data Fundamentos
- Python Fundamentos para Análise de Dados
- Big Data Analytics com R e Microsoft Azure Machine Learning
- Big Data Real-Time Analytics com Python e Spark
- Engenharia de Dados com Hadoop
- Machine Learning
- Business Analytics
- Visualização de Dados e Design de Dashboards
- Gerenciamento de Dados com MongoDB
- Análise de Séries Temporais em R
- Séries Temporais
- Administração SAS
- Data Science Aplicada à Área de Saúde
- IoT Analytics



Data Science Academy



R Fundamentos para Análise de Dados



Data Science Academy

Como este curso está dividido?



Data Science Academy

Como este curso está dividido?



Como este curso está dividido?

Capítulo 9

Subsetting, Séries Temporais, Documentação e Projeto



Data Science Academy

Como este curso está dividido?

Capítulo 10

Avaliação Final e E-book



Data Science Academy



O que esperar deste curso?



Data Science Academy

O que esperar deste curso?

- Praticidade
- Dinamismo
- Muito Conteúdo
- Alguns Conceitos Avançados
- Objetividade



Data Science Academy



O que NÃO esperar deste curso?



Data Science Academy

O que NÃO esperar deste curso?

- Este não é um curso de Estatística
- Comprovar Modelos com Fórmulas Matemáticas



Data Science Academy



Quizzes e Exercícios



Data Science Academy



Scripts



Data Science Academy



Projetos



Data Science Academy

Projetos

Projeto - Análise de Redes Sociais



Data Science Academy



Avaliação Final

25 questões - 2 tentativas - 70% - 12 meses



Data Science Academy



E como deve ser sua abordagem neste curso



Data Science Academy



O SEGREDO
do seu sucesso
esta na constância
do seu ESFORÇO



Data Science Academy

E como deve ser sua abordagem neste curso



Data Science Academy



Como se conectar com a
comunidade R?



Data Science Academy



www.r-bloggers.com



R news and tutorials contributed by (573) R bloggers

[Home](#) [About](#) [RSS](#) [add your blog!](#) [Learn R](#) [R jobs](#) [Contact us](#)

WELCOME!

[Follow @rbloggers](#) 33.4K

Here you will find daily news and tutorials about R, contributed by over 573 bloggers. There are many ways to follow us -

By e-mail:

Your e-mail here

Subscribe

31,008 readers

BY FEEDBURNER

On Facebook:

 R blogg... 35 mil curtidas


 Curtiu

Você e outros 41 amigos curtiram isso

Combine ArcGIS and R – Clustering Toolbox

July 2, 2016

By Fabio Veronesi



ex...

[Read more »](#)

Last year at the ESRI User Conference in San Diego, there was an announcement of an initiative to bridge ArcGIS and R. This became reality I think early this year with R-Bridge. Basically, ESRI has created an R library that is able to communicate and

Run compiled R packages in AzureML

July 1, 2016

By David Smith

Animated world inequality map

July 1, 2016

By Peter's stats stuff - R

RECENT POPULAR POSTS

- [The useR! 2016 Tutorials](#)
- [RMarkdown and Metropolis/Mtheme](#)
- [How to write good tests in R](#)

MOST VISITED ARTICLES OF THE WEEK

- [1. How to write the first for loop in R](#)
- [2. Installing R packages](#)
- [3. R tutorials](#)
- [4. Using apply, sapply, lapply in R](#)
- [5. In-depth introduction to machine learning in 15 hours of expert videos](#)
- [6. How to perform a Logistic Regression in R](#)
- [7. Computing and visualizing PCA in R](#)
- [8. Express Intro to dplyr](#)
- [9. The useR! 2016 Tutorials](#)



Data Science Academy



https://rpubs.com

RPubs brought to you by RStudio

Sign in Register

Easy web publishing from R

Write R Markdown documents in RStudio.

Share them here on RPubs. (It's free, and couldn't be simpler!)

Get Started

Recently Published

NOAA Storm Database exploration

dfrederick

July 2, 2016

This document explores the NOAA Storm Database data. The goal of this document is to demonstrate how to use the NOAA Storm Database data to explore the relationship between the number of deaths and the number of injuries.

Loading libraries

library(ggplot2)

library(dplyr)

library(readr)

library(tibble)

library(purrr)

library(stringr)

library(cowplot)

library(patchwork)

library(scales)

library(rlang)

library(magrittr)

library(RColorBrewer)

Test

R Markdown

This is an R Markdown document. It is a text document with embedded R code. The R code is executed and the results are displayed in the document. The R code is executed in the RStudio environment.

The R code is executed in the RStudio environment.

The R code is executed in the RStudio environment.

The R code is executed in the RStudio environment.

The R code is executed in the RStudio environment.

The R code is executed in the RStudio environment.

The R code is executed in the RStudio environment.

The R code is executed in the RStudio environment.

The R code is executed in the RStudio environment.

The R code is executed in the RStudio environment.

The R code is executed in the RStudio environment.

The R code is executed in the RStudio environment.

The R code is executed in the RStudio environment.

The R code is executed in the RStudio environment.

Integer Sequence Learning

dfrederick

July 2, 2016

This document explores the Integer Sequence Learning data. The goal of this document is to demonstrate how to use the Integer Sequence Learning data to explore the relationship between the number of deaths and the number of injuries.

The Integer Sequence Learning data is a text document with embedded R code.

The R code is executed in the RStudio environment.

The R code is executed in the RStudio environment.

The R code is executed in the RStudio environment.

The R code is executed in the RStudio environment.

The R code is executed in the RStudio environment.

The R code is executed in the RStudio environment.

The R code is executed in the RStudio environment.

The R code is executed in the RStudio environment.

The R code is executed in the RStudio environment.

The R code is executed in the RStudio environment.

The R code is executed in the RStudio environment.

The R code is executed in the RStudio environment.

ence Academy





https://try.jupyter.org

jupyter

Hosted by Rackspace

Files Running Clusters

Select items to perform actions on them.

Upload New ↕

- ☐ communities
- ☐ datasets
- ☐ featured
- ☐ Welcome Julia - Intro to Gadfly.ipynb
- ☐ Welcome R - demo.ipynb
- ☐ Welcome to Haskell.ipynb
- ☐ Welcome to Python.ipynb
- ☐ Welcome to Spark with Python.ipynb
- ☐ Welcome to Spark with Scala.ipynb

- Text File
- Folder
- Terminal
- Notebooks
- Apache Toree - Scala
- Bash
- Haskell
- Julia 0.3.2
- Python 2
- Python 3
- R
- Ruby 2.1.5

Create a new notebook with R



Data Science Academy



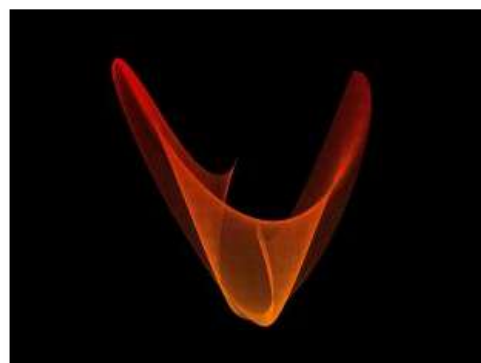
computerworld.com.br/big-data

Big Data

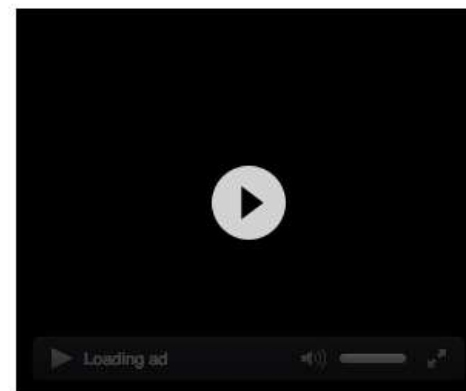
Notícias, tendências, análises, dicas e boas práticas sobre Big Data



Exemplos de aplicações de data mining no mercado brasileiro



Volume, variedade, velocidade, veracidade e valor: Os cinco Vs do Big Data



Quantidade de dados vai mudar estratégias de negócios até 2020

17/06/2016 - 12h40

Hu Yoshida, CTO da HDS, afirma que apenas 15% do volume existente de informações será realmente armazenada pelas empresas



Menos de um terço dos projetos de Big Data são lucrativos

16/06/2016 - 14h48

Pesquisa da Capgemini e Informatica aponta que 45% dos projetos envolvendo grandes volumes de dados apenas cobrem os gastos

Mais Lidas



Para ficar de olho: Dez tecnologias emergentes que revolucionarão o mundo



Como cinco minutos por semana podem mudar positivamente sua carreira



Passo a passo para evitar a

Science Academy

Grupos do facebook



- Users of R Statistical Package
- R Brasil - Programadores
- Estatística Brasil



Data Science Academy



Divirta-se!



Data Science Academy



Data Science Academy



Data Science Academy



Obrigado!



Data Science Academy

Por que Cientistas de Dados usam R?



Data Science Academy



Breve História do R



Data Science Academy

1993	Projeto de pesquisa em Auckland, na Nova Zelândia
1995	R liberado como projeto open-course
1997	Formado o grupo R-Core
2000	Liberada a versão 1.0.0 do R
2003	Criação da R Foundation
2004	Primeira conferência internacional de usuários em Vienna
2015	Formado o R Consortium (com participação da IBM e Microsoft)

Breve História do R



Data Science Academy

R possui diversas funções para:

- Extração de Dados
- Limpeza de Dados
- Carregamento e Transformação de Dados
- Análise Estatística
- Modelagem Preditiva
- Machine Learning
- Visualização de Dados



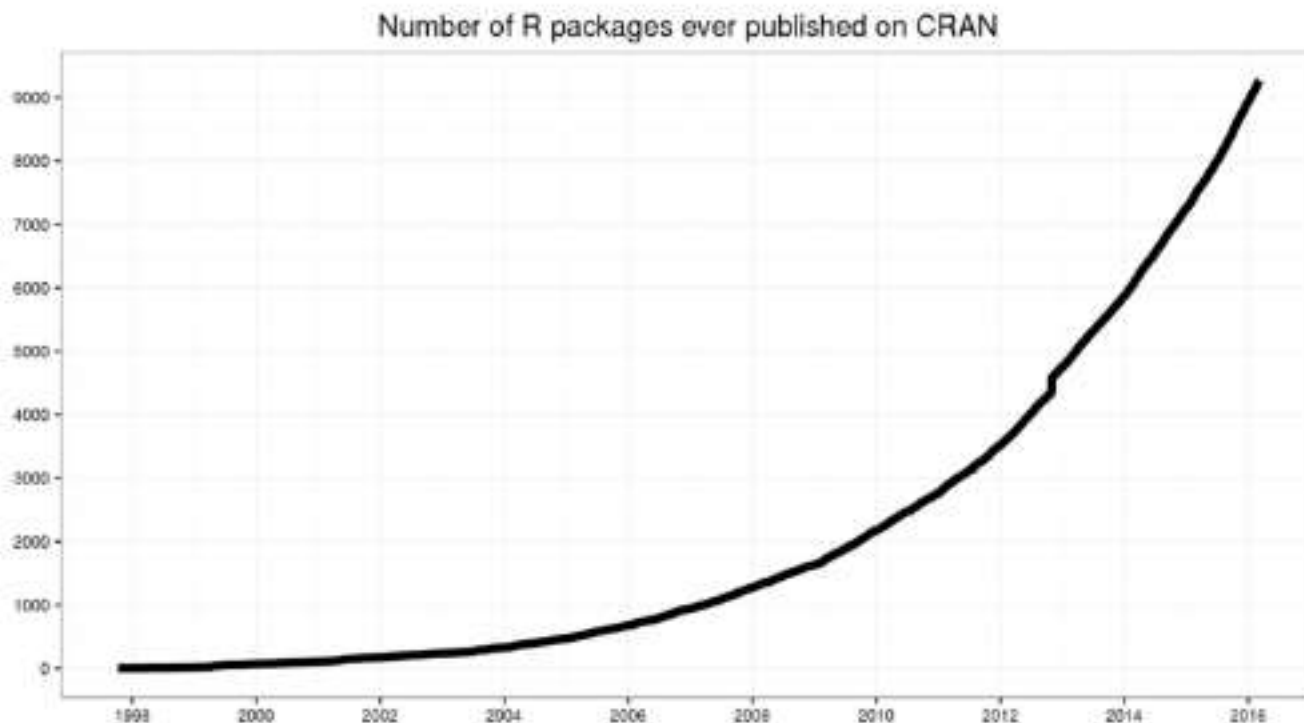
Data Science Academy

Pesquisa realizada no site KD

What programming languages you used for data mining / data analysis in the past 12 months? [570 voters]

R (257)	45%
SQL (184)	32%
Python (140)	25%
Java (139)	24%
SAS (121)	21%
MATLAB (83)	15%
C/C++ (73)	13%
Unix shell/awk/gawk/sed (59)	10%
Perl (45)	7.9%
Hadoop/Pig/Hive (35)	6.1%
Lisp (4)	0.7%
Other (70)	12.0%
None (7)	1.2%

ademy



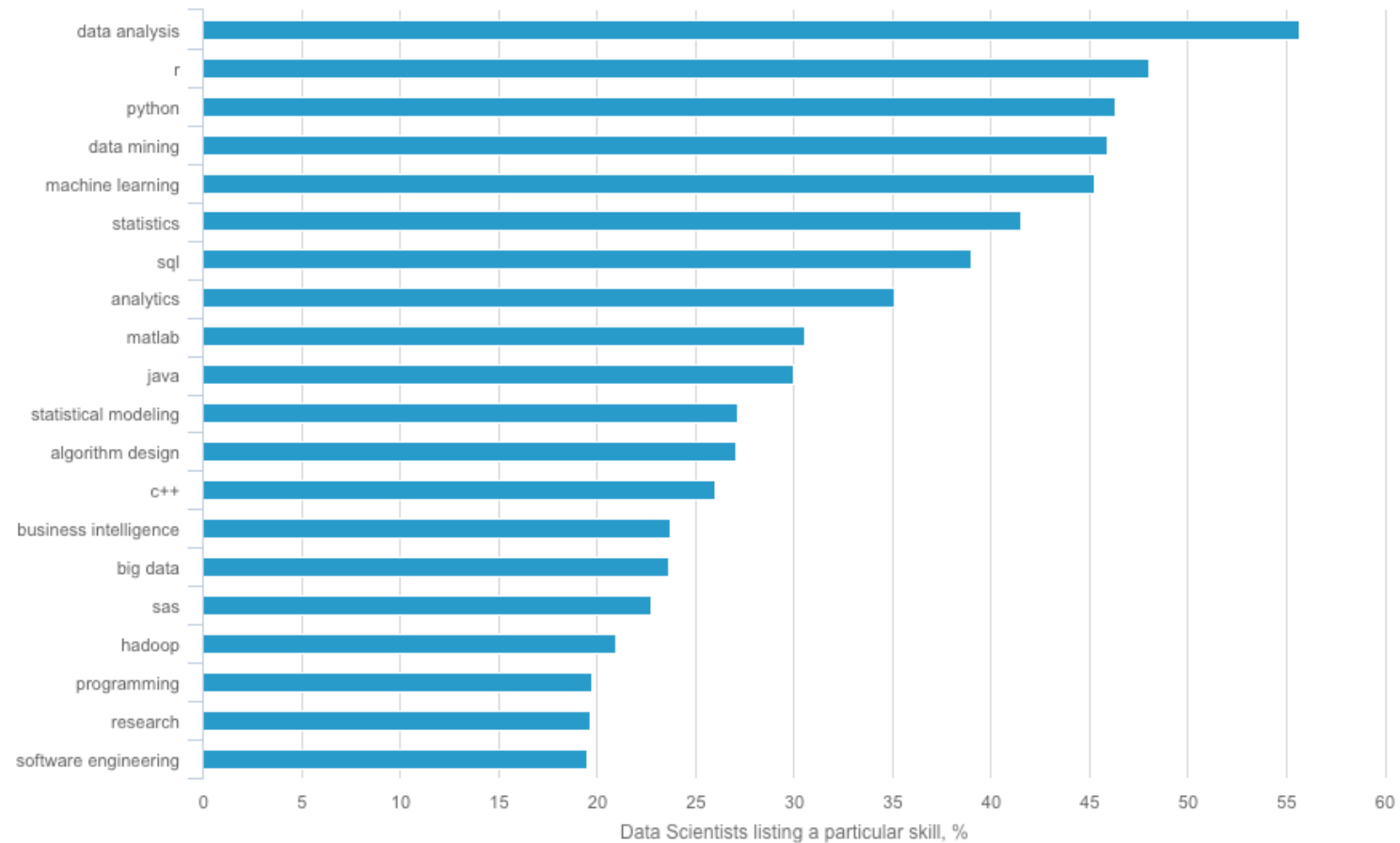
Número de
pacotes R
publicados por
ano no CRAN

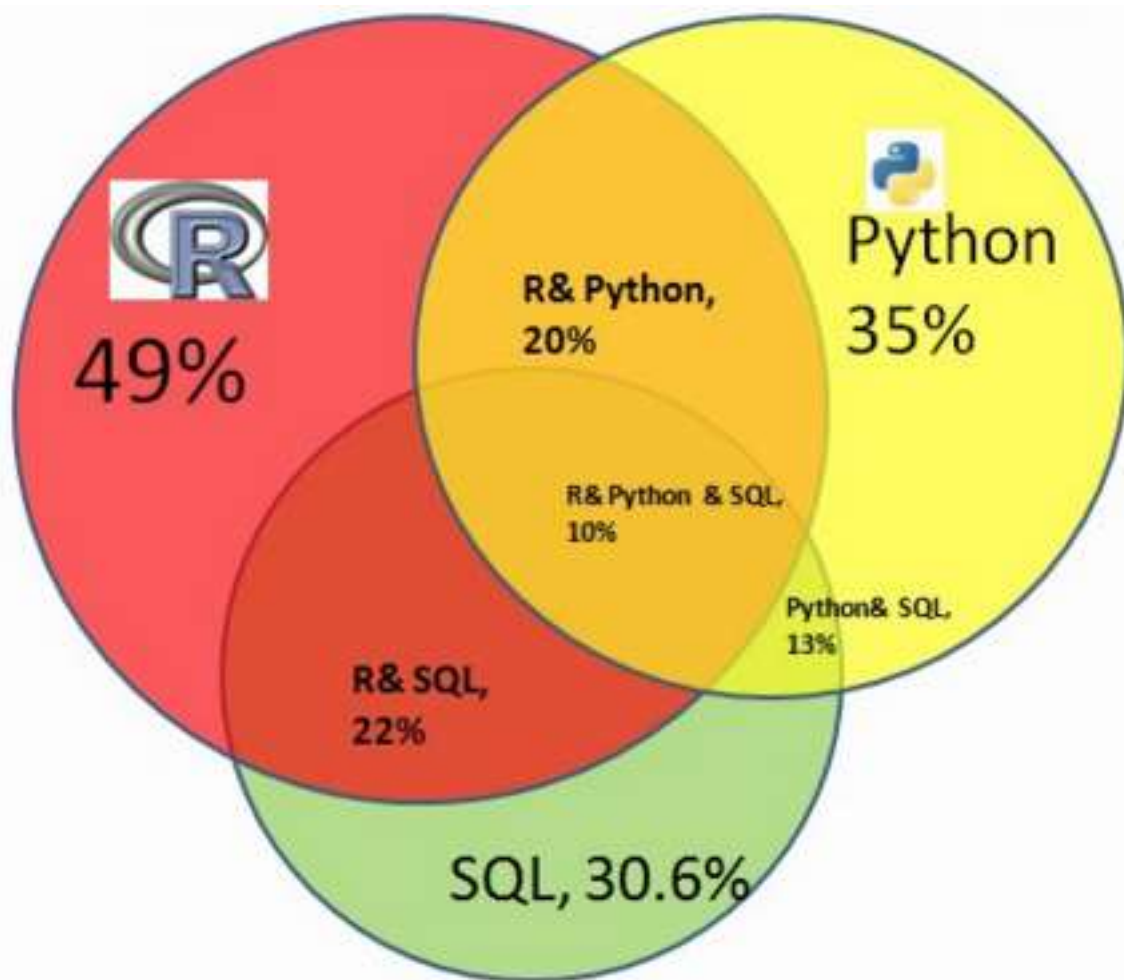
Fonte: <http://blog.revolutionanalytics.com/2016/03/16-years-of-r-history.html>



Data Science Academy

TOP 20 SKILLS OF A DATA SCIENTIST

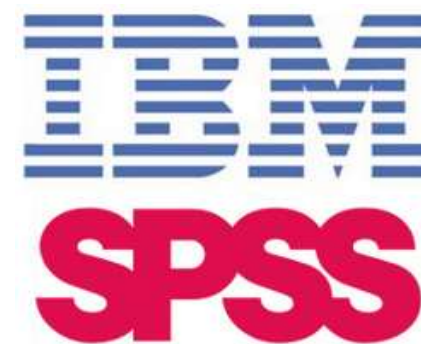




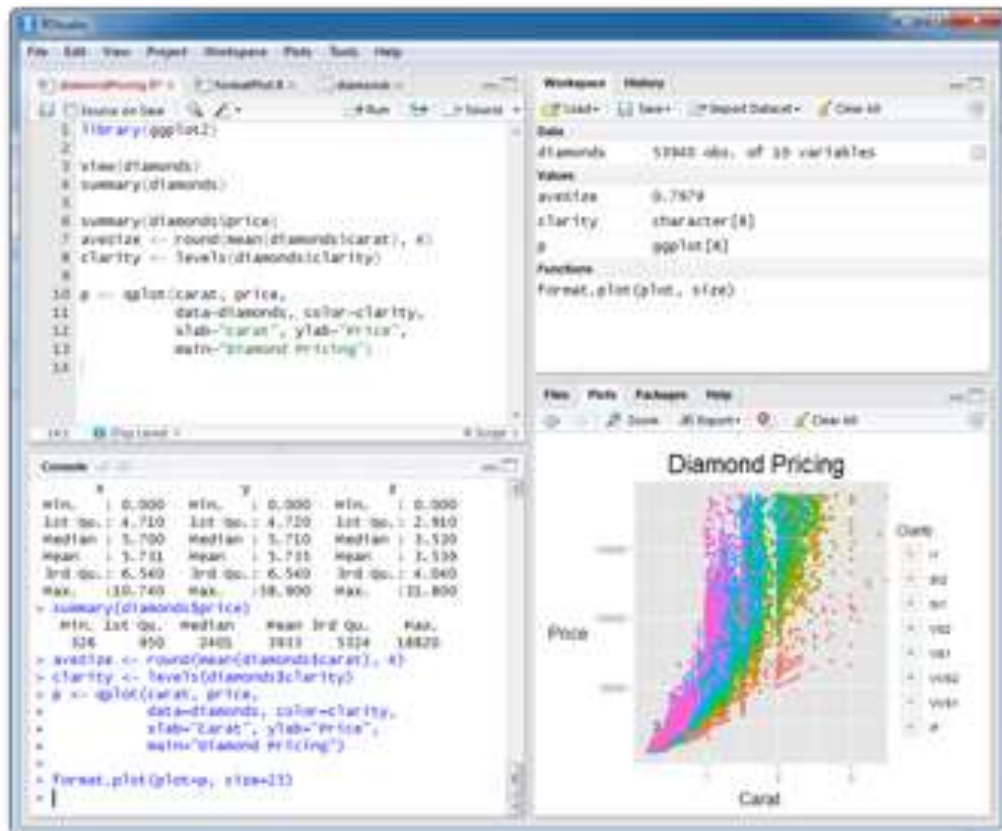
Espera-se que o Cientista de Dados domine pelo menos uma destas 3 linguagens



Data Science Academy



Data Science Academy



Sintaxe

"A" != "a"



Data Science Academy



Dificuldade



Data Science Academy



Vantagens e Desvantagens



Data Science Academy

Vantagens

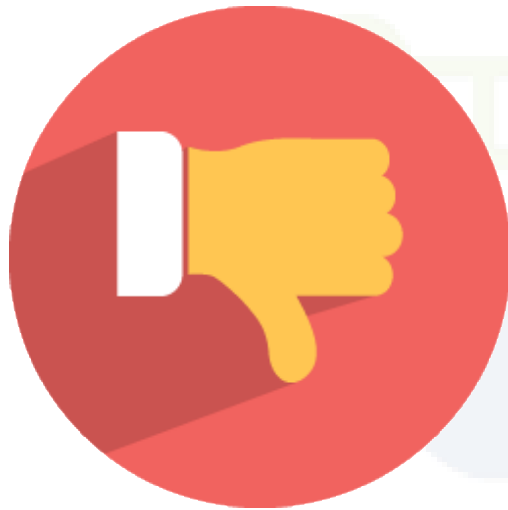


- Grande variedade de pacotes disponíveis
- Flexibilidade e Rapidez
- Machine Learning



Data Science Academy

Desvantagens



- Não há interface gráfica. Tudo é feito por linha de comando
- Limitações no uso de memória, principalmente com datasets muito grandes



Data Science Academy



www.datascienceacademy.com.br



<https://www.r-consortium.org>

MEMBERS

R Foundation Member



Platinum



Microsoft



Gold



Silver



AVANT



Data Science Academy



Oracle R Enterprise

Microsoft R Server

Oracle e Microsoft estão investindo pesado na linguagem R para suas soluções analíticas



Data Science Academy



- **sqldf** - pacote que permite realizar queries SQL em dataframes no R
- **forecast** - modelar séries temporais
- **plyr** - dividir uma estrutura de dados em grupos e aplicar funções a cada grupo
- **stringr** - manipulação de strings
- **Database drivers** - RMongo, RODB, RMySQL
- **ggplot2** - visualização de dados
- **qcc** - controle de qualidade estatístico
- **randomForest** - pacote para Machine Learning



Data Science Academy



Data Science Academy





Qual a importância do Big Data Analytics?



Data Science Academy



Dados x Informação x Conhecimento x Inteligência



Data Science Academy

Dados x Informação x Conhecimento x Inteligência

Funcionário	Idade	Salário	Cargo
Aristóteles	32	R\$ 7.500,00	Analista Financeiro
Sócrates	41	R\$ 9.300,00	Gerente Financeiro


Informação



Data Science Academy

Dados x Informação x Conhecimento x Inteligência

As pessoas trabalham com informações mas a tecnologia armazena dados



Data Science Academy

Dados x Informação x Conhecimento x Inteligência

Funcionário	Idade	Salário	Cargo
Aristóteles	32	R\$ 7.500,00	Analista Financeiro
Sócrates	41	R\$ 9.300,00	Gerente Financeiro

Conhecimento



Data Science Academy

Dados x Informação x Conhecimento x Inteligência



Data Science Academy

Dados x Informação x Conhecimento x Inteligência

Funcionário	Idade	Salário	Cargo
Aristóteles	32	R\$ 7.500,00	Analista Financeiro
Sócrates	41	R\$ 9.300,00	Gerente Financeiro

Inteligência - capacidade de resolver problemas, usando o conhecimento, através das informações disponíveis

Por exemplo: por que os funcionários do setor financeiro tem queda de performance após 1 ano trabalhando na empresa

Data Science Academy

Dados x Informação x Conhecimento x Inteligência



Big Data Analytics

Extrair conhecimento a partir dos dados



Data Science Academy

Dados x Informação x Conhecimento x Inteligência



Machine Learning

Ensinar algoritmos a usar inteligência, ou seja, usar o conhecimento para resolver problemas

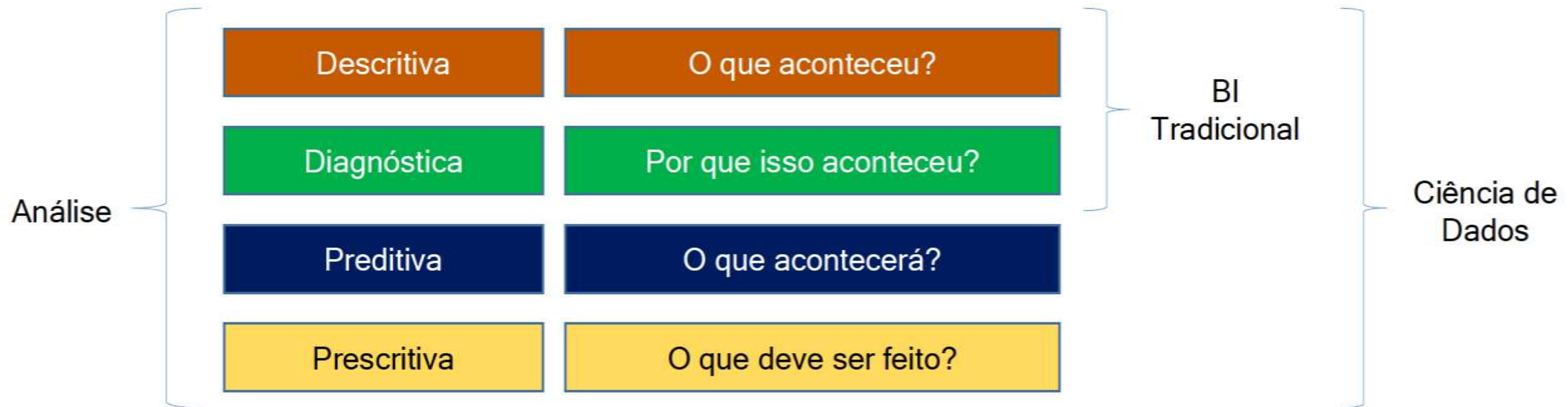


Data Science Academy

[illegible]

www.datascienceacademy.com.br

Evolução dos Sistemas Analíticos



Data Science Academy



Evolução dos Sistemas Analíticos

O conhecimento nos faz mais inteligentes;
pessoas e empresas que aprendem (e agora, máquinas)



Data Science Academy

Estamos vivendo a era da explosão dos dados

bit byte kilobyte megabyte gigabyte terabyte petabyte exabyte zettabyte yottabyte

Em 2014 a humanidade acumulou em dados
o equivalente a 1.8 Zettabyte





O que é Big Data Analytics?

O objetivo é simples: melhorar seus processos de trabalho e adquirir insights valiosos acerca das tendências de mercado, comportamento dos consumidores e suas expectativas



Data Science Academy

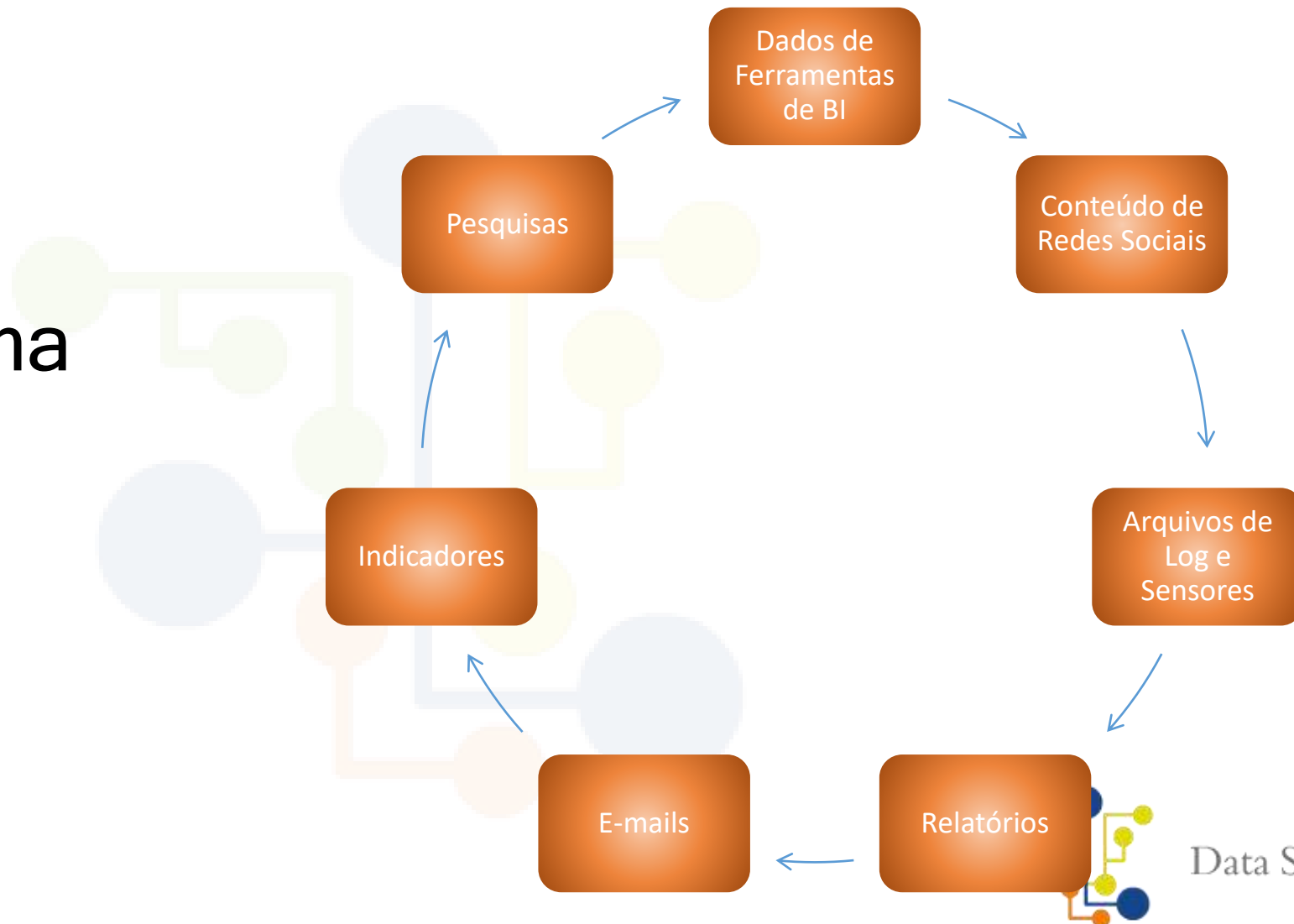
O que é Big Data Analytics?

Big Data Analytics é o trabalho analítico e inteligente em grandes volumes de dados, estruturados ou não-estruturados, que são coletados, armazenados e interpretados por softwares de altíssimo desempenho



Data Science Academy

Matéria-prima



Data Science Academy



Vantagens e Benefícios



Data Science Academy

Vantagens e Benefícios

- Direcionamento das Vendas
- Aperfeiçoamento do Processo de Logística
- Atendimento mais eficiente
- Melhoria na Gestão de Recursos Humanos
- Identificação de Padrões
- **Análise da Concorrência**




Data Science Academy



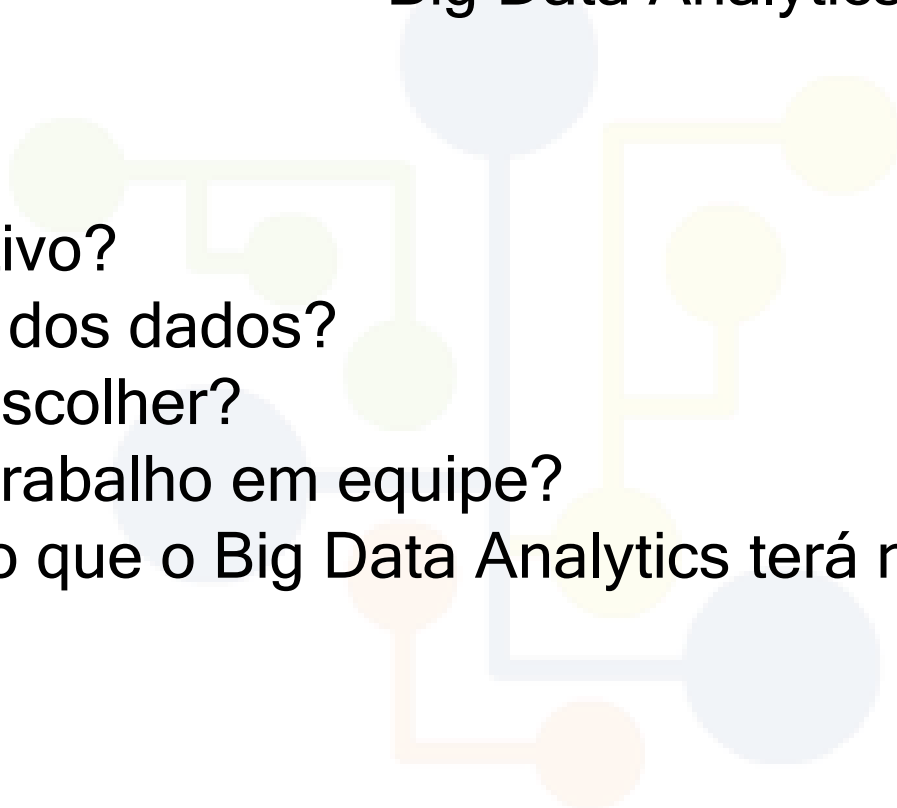
5 Perguntas que Precisam ser respondidas antes de pensar em Big Data Analytics



Data Science Academy



5 Perguntas que Precisam ser respondidas antes de pensar em Big Data Analytics

- 
- 1- Qual seu objetivo?
 - 2- Qual a origem dos dados?
 - 3- Que solução escolher?
 - 4- Este será um trabalho em equipe?
 - 5- Qual o impacto que o Big Data Analytics terá no negócio?



Data Science Academy



Big Data Analytics em Áreas de Negócio



Data Science Academy



Saúde



Data Science Academy



Saúde

- Combate a doenças e epidemias
- Dispositivos móveis de monitoramento
- Atendimento médico individualizado
- Wearables e apps



Data Science Academy

A Accenture Healthcare IT Vision 2015, uma pesquisa anual sobre tendências de tecnologia, traz dados interessantes sobre o uso de Big Data na saúde.



Data Science Academy

45% dos executivos da área consultados afirmaram que fazem análise preditiva



Data Science Academy

59% disseram que já usam algoritmos com o objetivo de fazer softwares operarem com mais inteligência



The graphic is a dark blue rectangular slide with a white line graph at the top. The Accenture logo is in the top left, and the text 'Strategy | Consulting | Digital | Technology | Operations' is in the top right. The title '2015 Healthcare IT Vision' is in orange. The main title 'Top 5 eHealth Trends' is in large white and orange font. A large orange arrow points to the right with the text 'High performance. Delivered.' inside it. A white box at the bottom contains the text: 'Five trends prove that digital is dramatically influencing the industry, today, and well into tomorrow.'

accenture

Strategy | Consulting | Digital | Technology | Operations

2015 Healthcare IT Vision

Top 5
eHealth Trends

High performance. Delivered.

Five trends prove that digital is **dramatically influencing** the industry, today, and well into tomorrow.



Data Science Academy

73% de todos os executivos de saúde relataram ter obtido um ROI positivo depois de investir em tecnologias como dispositivos móveis do tipo wearables



Data Science Academy



Construção



Data Science Academy



Logística

Estatística + Matemática + Mídias Sociais + Macroeconomia + Tecnologia + Dados

=

Big Data Analytics



Data Science Academy



A word cloud of marketing terms arranged in a grid-like fashion. The words are: CUSTOMER, ANALYSIS, ADVERTISING, MARKET, SALES, PROMOTION, PRICE, INTERNET, and TARGET. The word 'CUSTOMER' is circled in red. A red marker is positioned to the right of the word cloud.

Marketing

- Análise de mercado mais específica
- Interpretação de Dados mais eficaz
- Mais fontes de dados
- Automação do Marketing
- Experiências mais significativas no PV



Data Science Academy



Indústria

- Eficiência no Planejamento
- Gestão de Suprimentos
- Redução de Custos
- Qualidade



Data Science Academy



E-commerce



Data Science Academy



Energia



Data Science Academy



A Empresa como um todo

- Mudança de Cultura
- Visão Empresarial
- Esforço Top-Down



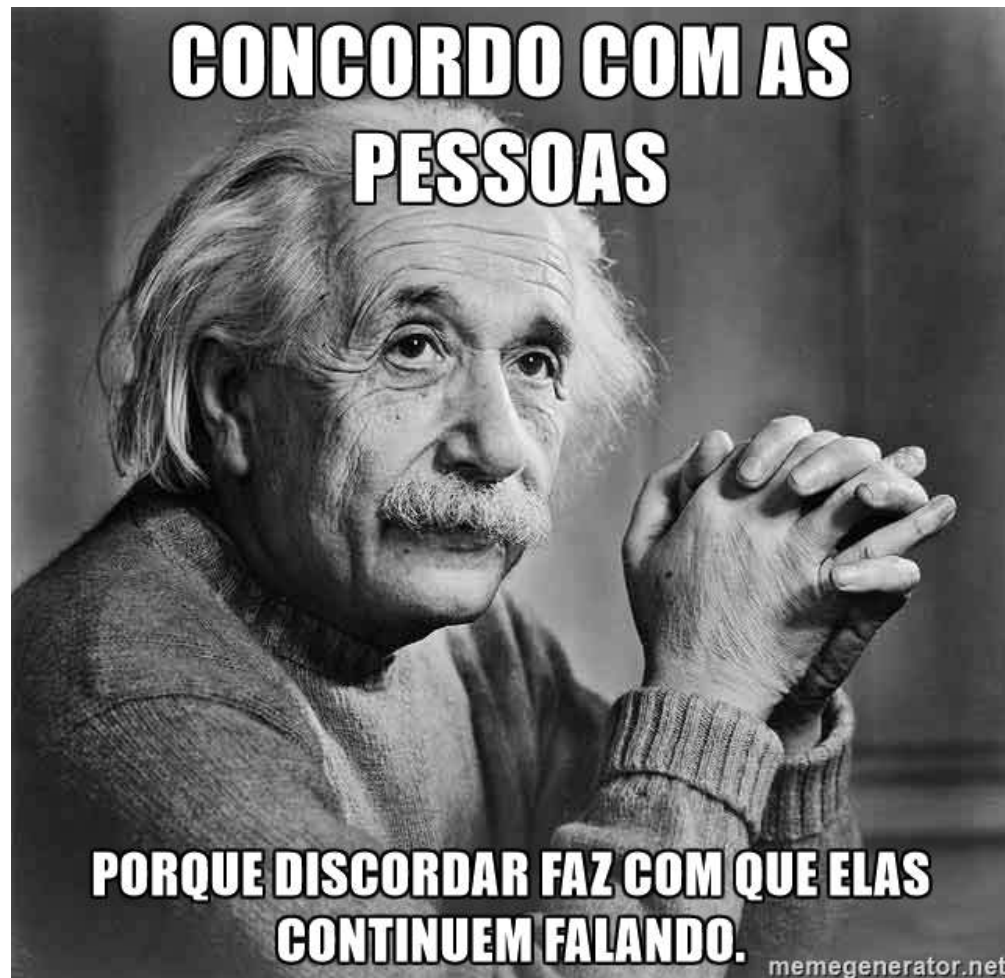
Data Science Academy



- Big Data já não é mais um tema distante
- Big Data fará toda a diferença para seu negócio
- Qualquer empresa pode se beneficiar do Big Data
- Big Data é um trabalho em equipe



Data Science Academy



Mas por onde eu começo?



Data Science Academy



Como o Big Data pode me ajudar a aumentar o Market Share da empresa?



Data Science Academy



Coletar Dados

Faturamento

Marketing

Clientes

Custos

Efetividade das Campanhas de Marketing

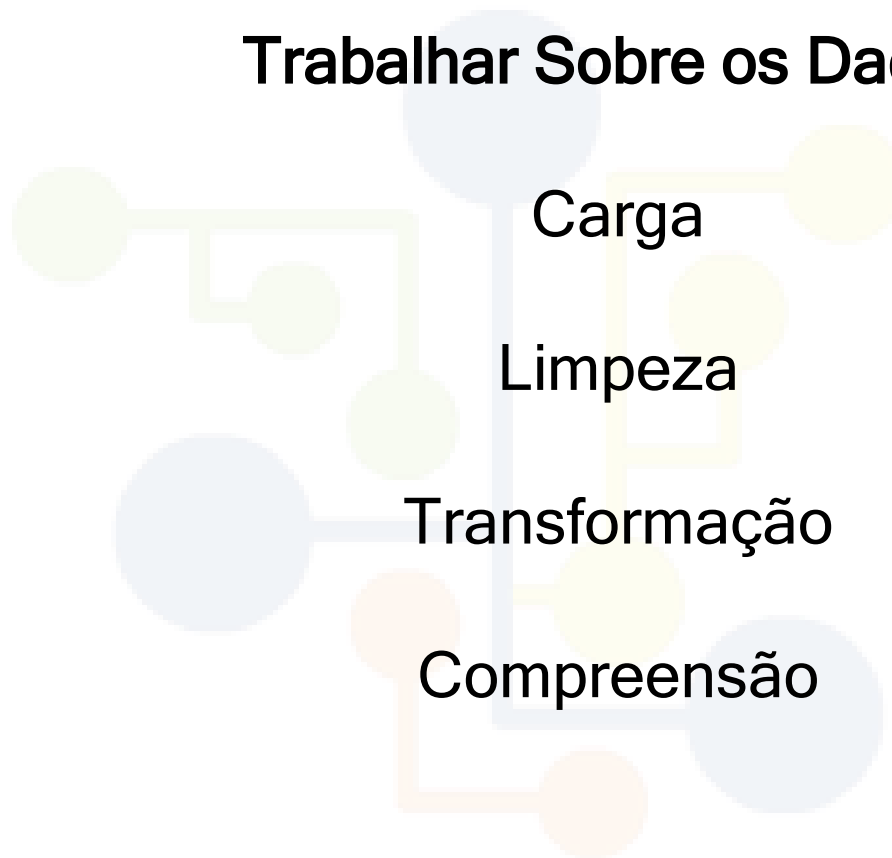
Concorrentes

Redes Sociais



Data Science Academy

Trabalhar Sobre os Dados



Data Science Academy



Técnicas de Análise

Aplicar modelos estatísticos e compreender o relacionamento entre os dados

Definir variáveis de observação e explanatórias

Buscar correlação e causalidade



Data Science Academy



Machine Learning

Juntar tudo e criar um modelo de machine learning, prevendo como estas variáveis afetam umas às outras quando alteradas

Automatizar o processo



Data Science Academy



Apresentar seus Resultados



Data Science Academy



Usando o Github



Data Science Academy

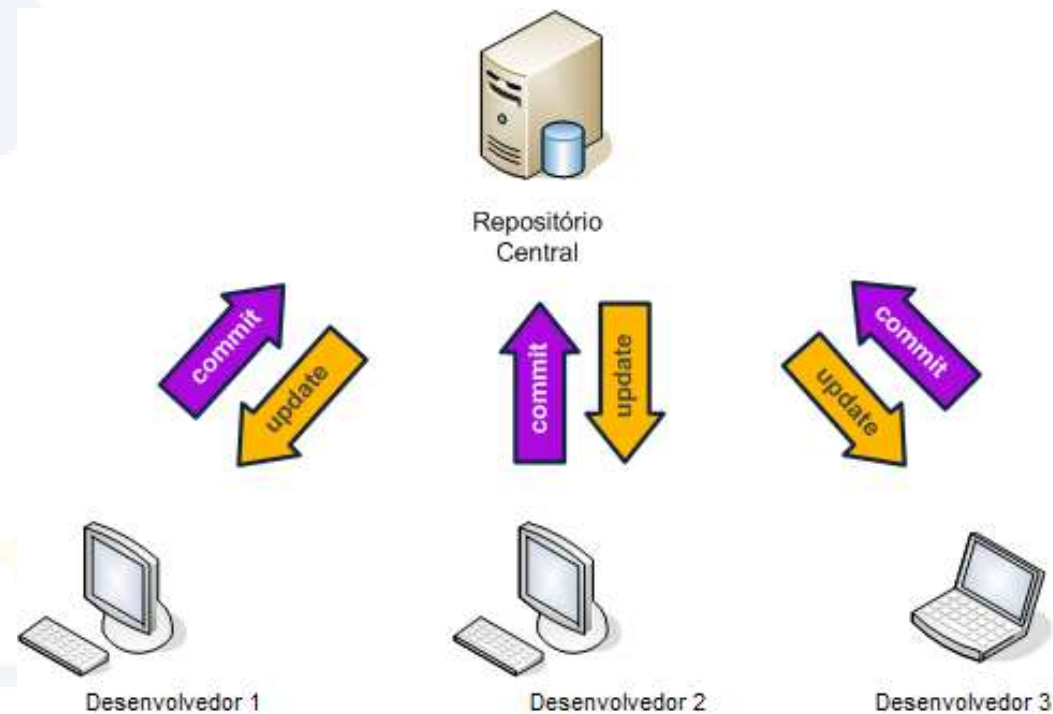


O que é um sistema de controle de versão?



Data Science Academy

Várias pessoas
trabalhando nos
mesmos arquivos



Data Science Academy



Histórico e Conceitos



Data Science Academy

Sistemas de Controle de Versão

- *Concurrent Versions System (CVS)*
- *Subversion (SVN)*
- *Visual SourceSafe (VSS)*
- *Rational ClearCase*
- *Git*



Data Science Academy



git



Data Science Academy



- Não depender de um servidor central
- Dar ênfase à velocidade
- Integridade dos dados
- Potencializar o trabalho paralelo



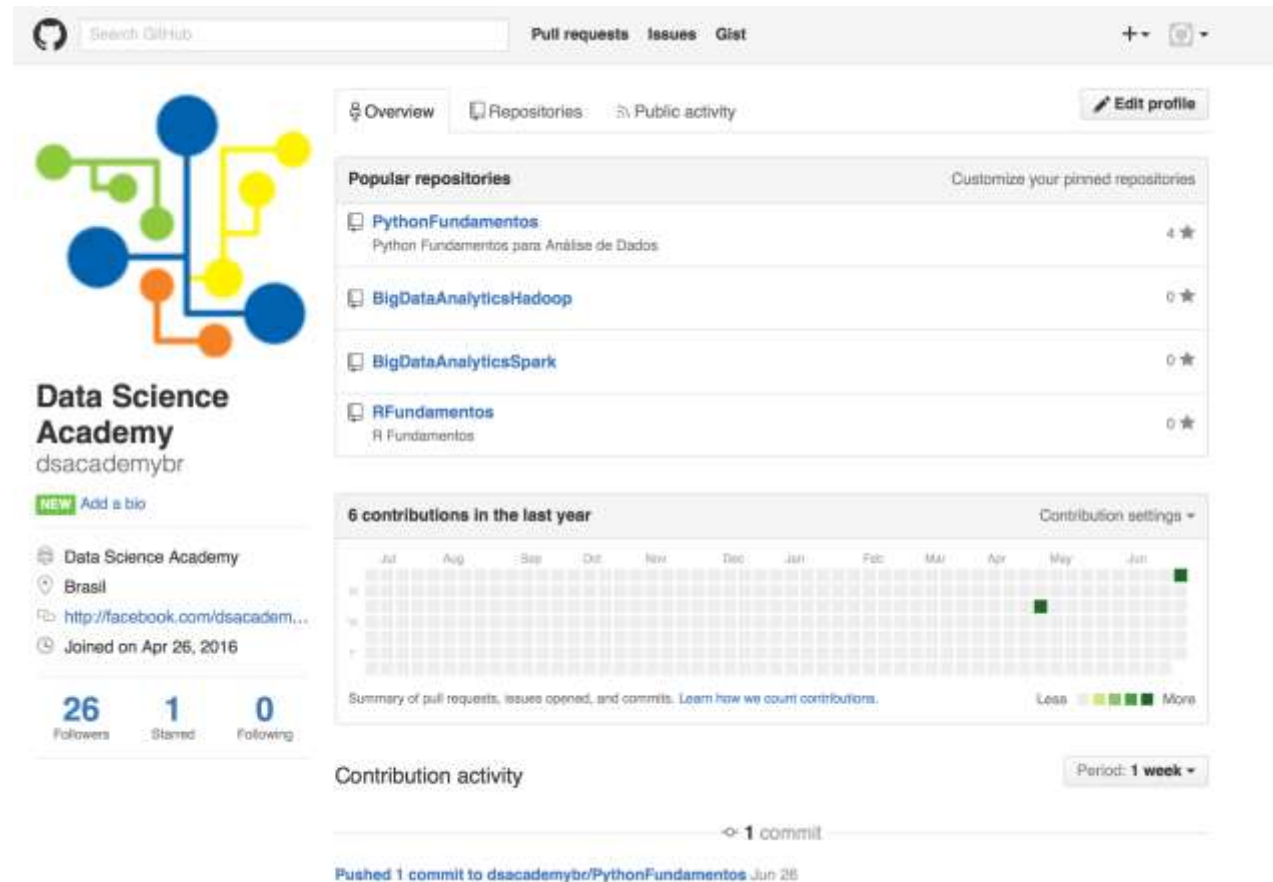

Data Science Academy



<https://github.com>



Data Science Academy



Search GitHub

Pull requests Issues Gist

Overview Repositories Public activity Edit profile

Popular repositories Customize your pinned repositories

- PythonFundamentos**
Python Fundamentos para Análise de Dados 4 ★
- BigDataAnalyticsHadoop** 0 ★
- BigDataAnalyticsSpark** 0 ★
- RFundamentos**
R Fundamentos 0 ★

6 contributions in the last year Contribution settings

Jul Aug Sep Oct Nov Dec Jan Feb Mar Apr May Jun

Summary of pull requests, issues opened, and commits. Learn how we count contributions. Less More

Contribution activity Period: 1 week

1 commit

Pushed 1 commit to dsacademybr/PythonFundamentos Jun 26

<https://github.com/dsacademybr>

www.datascienceacademy.com.br



Data Science Academy



Muitos Cientistas de Dados
estão no Github



Data Science Academy

https://github.com/rasbt



Personal

Open source

Business

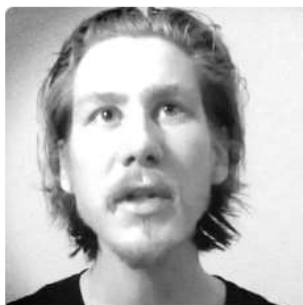
Explore

Pricing Blog Support

Search GitHub

Sign in

Sign up



Sebastian Raschka

rasbt

'Data Scientist' and Machine Learning enthusiast with a big passion for Python & open source. Author of 'Python Machine Learning'. Michigan State University.



Michigan State University



East Lansing, Michigan



mail@sebastianraschka.com



http://sebastianraschka.com



Joined on Oct 5, 2013

Overview

Repositories

Public activity

Follow

Pinned repositories

	python-machine-learning-book	3,520 ★
The "Python Machine Learning" book code repository and info resource		
	pattern_classification	1,756 ★
A collection of tutorials and examples for solving and understanding machine learning and pattern classification...		
	mlxtend	484 ★
A library of extension and helper modules for Python's data analysis and machine learning libraries.		
	python_reference	1,606 ★
Useful functions, tutorials, and other Python-related things		
	biopandas	52 ★
Working with molecular structures in pandas DataFrames		

1,510 contributions in the last year



Data Science Academy

https://github.com/rhiever



Personal

Open source

Business

Explore

Pricing

Blog

Support

Search GitHub

Sign in

Sign up



Randy Olson

rhiever

Artificial Intelligence & Machine Learning researcher. Data tinkerer. Community leader for /r/DataIsBeautiful.

University of Pennsylvania

Philadelphia, PA

rs@randalolson.com

http://www.randalolson.com

Joined on May 8, 2012

1.5k 69 14

Overview

Repositories

Public activity

Follow

Popular repositories

Data-Analysis-and-Machine-Le...	Repository of teaching materials, code, and data for my data analysis and machine learning projects.	2,457 ★
tpot	A Python tool that automatically creates and optimizes machine learning pipelines using genetic programming.	1,150 ★
TwitterFollowBot	A Python bot that automates several actions on Twitter, such as following users and favoriting tweets.	562 ★
datacleaner	A Python tool that automatically cleans data sets and readies them for analysis.	301 ★
reddit-analysis	A Python script that parses post titles, self-texts, and comments on reddit and makes word clouds out of the word freq...	182 ★


812 contributions in the last year




Data Science Academy

www.datascienceacademy.com.br


<https://github.com/hmason>


 [Personal](#) [Open source](#) [Business](#) [Explore](#) [Pricing](#) [Blog](#) [Support](#) [Sign in](#) [Sign up](#)





Hilary Mason

hmason

 [Fast Forward Labs](#)

 [New York, NY](#)

 <http://www.hilarymason.com>

 [Joined on Jul 22, 2008](#)

2k

Followers


134

Starred

50






Following

Organizations



[Overview](#) [Repositories](#) [Public activity](#) [Follow](#)


Popular repositories

 gitmarks	delicious-like bookmarks on github.	321 ★
 ml_class	code for my O'Reilly masterclass videos	254 ★
 botomatic	easily create twitter bots in python	233 ★
 tc	A command-line twitter client with smart filtering and statistical classification	167 ★
 Introbot	A python script to generate the text of Intro e-mails.	70 ★

3 contributions in the last year

	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
M	■											
W			■									
F												

Summary of pull requests, issues opened, and commits. [Learn how we count contributions.](#)

Less  More



Data Science Academy

https://github.com/mbostock



Personal

Open source

Business

Explore

Pricing

Blog

Support

Search GitHub

Sign in

Sign up



Mike Bostock

mbostock

San Francisco, CA

mike@ocks.org

http://bost.ocks.org

Joined on Mar 25, 2010

12.7k

Followers

37

Starred

13

Following

Organizations



Overview

Repositories

Public activity

Follow

Pinned repositories



d3/d3

Bring data to life with SVG, Canvas and HTML.



52,082 ★



topojson

An extension to GeoJSON that encodes topology.

2,366 ★



d3/d3-shape

Graphical primitives for visualization, such as lines and areas.

1,239 ★



d3/d3-scale

Encodings that map abstract data to visual representation.

677 ★

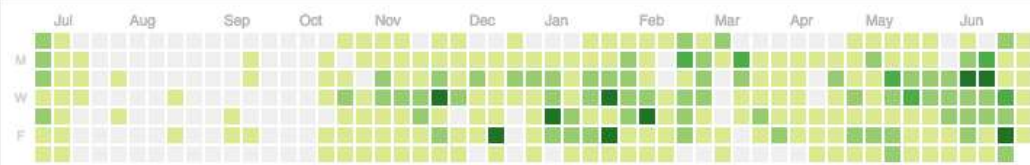


d3/d3-geo-projection

Extended geographic projections for D3.

325 ★

5,546 contributions in the last year



Summary of pull requests, issues opened, and commits. [Learn how we count contributions.](#)

Less More



Data Science Academy

https://github.com/rdpeng



Personal

Open source

Business

Explore

Pricing Blog Support

Search GitHub

Sign in

Sign up



Roger D. Peng
rdpeng

Professor of Biostatistics at Johns
Hopkins University



Johns Hopkins University



Baltimore, MD



rdpeng@gmail.com



http://www.biostat.jhsph.edu/~f...



Joined on May 7, 2008

3.2k

Followers

16

Starred

3

Following

Overview

Repositories

Public activity

Follow

Popular repositories

courses

Course materials for the Data Science Specialization: <https://www.coursera.org/specialization/jhudatascience/1>

1,088 ★

ProgrammingAssignment2

Repository for Programming Assignment 2 for R Programming on Coursera

388 ★

practice_assignment

Practice assignment for the R programming class on Coursera

225 ★

ExData_Plotting1

Plotting Assignment 1 for Exploratory Data Analysis

107 ★

RepData_PeerAssessment1

Peer Assessment 1 for Reproducible Research

44 ★

187 contributions in the last year




Summary of pull requests, issues opened, and commits. [Learn how we count contributions.](#)


Less More




Data Science Academy






www.datascienceacademy.com.br

 <https://github.com/amueller>

 Personal Open source Business Explore Pricing Blog Support [Sign in](#) [Sign up](#)







Andreas Mueller
amueller




 NYU Center for Data Science
 NYC
 t3kcit+githubspam@gmail.com
 <http://amueller.github.io>
 Joined on Oct 22, 2010

1.6k
Followers






51
Starred

40
Following

Organizations
   

 Overview  Repositories  Public activity [Follow](#)


Popular repositories

 word_cloud A little word cloud generator in Python	1,298 ★
 introduction_to_ml_with_python Notebooks and code for the book "Introduction to Machine Learning with Python"	392 ★
 scipy_2015_sklearn_tutorial Scikit-Learn tutorial material for Scipy 2015	335 ★
 kaggle_insults Kaggle Submission for "Detecting Insults in Social Commentary"	102 ★
 odscon-sf-2015 Material for ODSCON San Francisco 2015	76 ★

1,271 contributions in the last year

	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
M												
W												
F												

Summary of pull requests, issues opened, and commits. [Learn how we count contributions.](#)

Less  More

Data Science Academy



No Github você pode encontrar ainda uma infinidade de material sobre Data Science e projetos completos



Data Science Academy



Principais Conceitos do Git



Data Science Academy



Principais conceitos do Git

Branch

Ramificação do projeto,
cada *branch* representa uma versão do
seu projeto e podemos seguir uma linha de
desenvolvimento a partir de cada *branch*



Data Science Academy



Principais conceitos do Git

Clone

Cópia local de todos os arquivos de um repositório git



Data Science Academy



Principais conceitos do Git

Commit

Coleção de alterações realizadas, é uma espécie de *checkpoint*, sempre que necessário você pode retroceder até algum *commit* existente



Data Science Academy



Principais conceitos do Git

Fork

Uma bifurcação do projeto, uma cópia do projeto existente para seguir uma nova direção



Data Science Academy

Principais conceitos do Git

Master

Branch padrão de um repositório Git



Data Science Academy



Principais conceitos do Git

Merge

É a capacidade de incorporar alterações do git, quando acontece uma junção de diferentes *branches*



Data Science Academy



Principais conceitos do Git

Pull

Puxa as alterações do repositório remoto



Data Science Academy



Principais conceitos do Git

Push

Empurra as suas alterações para o
repositório remoto



Data Science Academy



Principais conceitos do Git

Repositório

Local onde ficam todos os arquivos do projeto, inclusive o histórico e versões



Data Science Academy



Como utilizar o Github?

<https://github.com>



Data Science Academy

 <https://github.com>



Personal Open source Business Explore

Pricing Blog Support

Search GitHub

Sign in

Sign up

How people build software

Millions of developers use GitHub to build personal projects, support their businesses, and work together on open source technologies.



Pick a username

Your email address

Create a password

Use at least one letter, one numeral, and seven characters.

Sign up for GitHub

By clicking "Sign up for GitHub", you agree to our [terms of service](#) and [privacy policy](#). We'll occasionally send you account related emails.

<https://github.com>



Data Science Academy

www.datascienceacademy.com.br

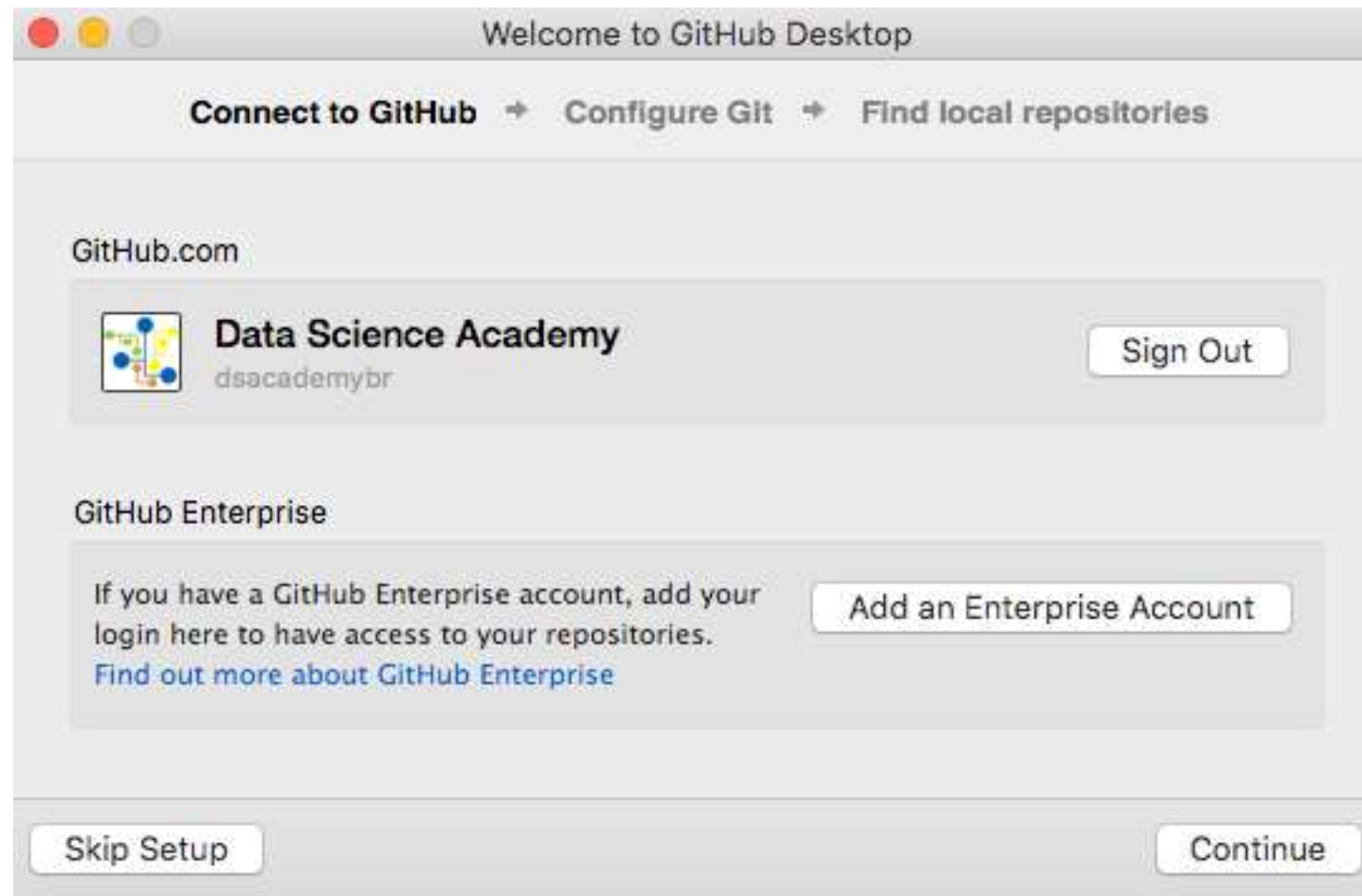


Github Desktop

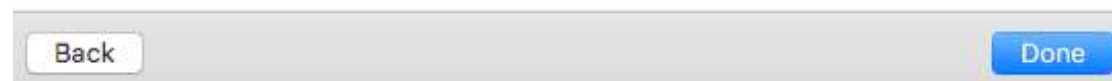
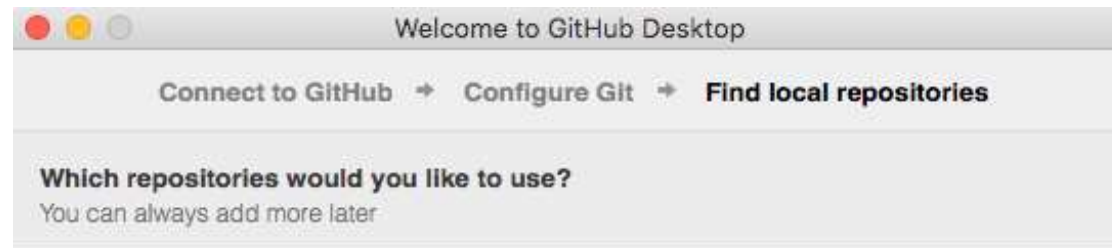
<https://desktop.github.com>



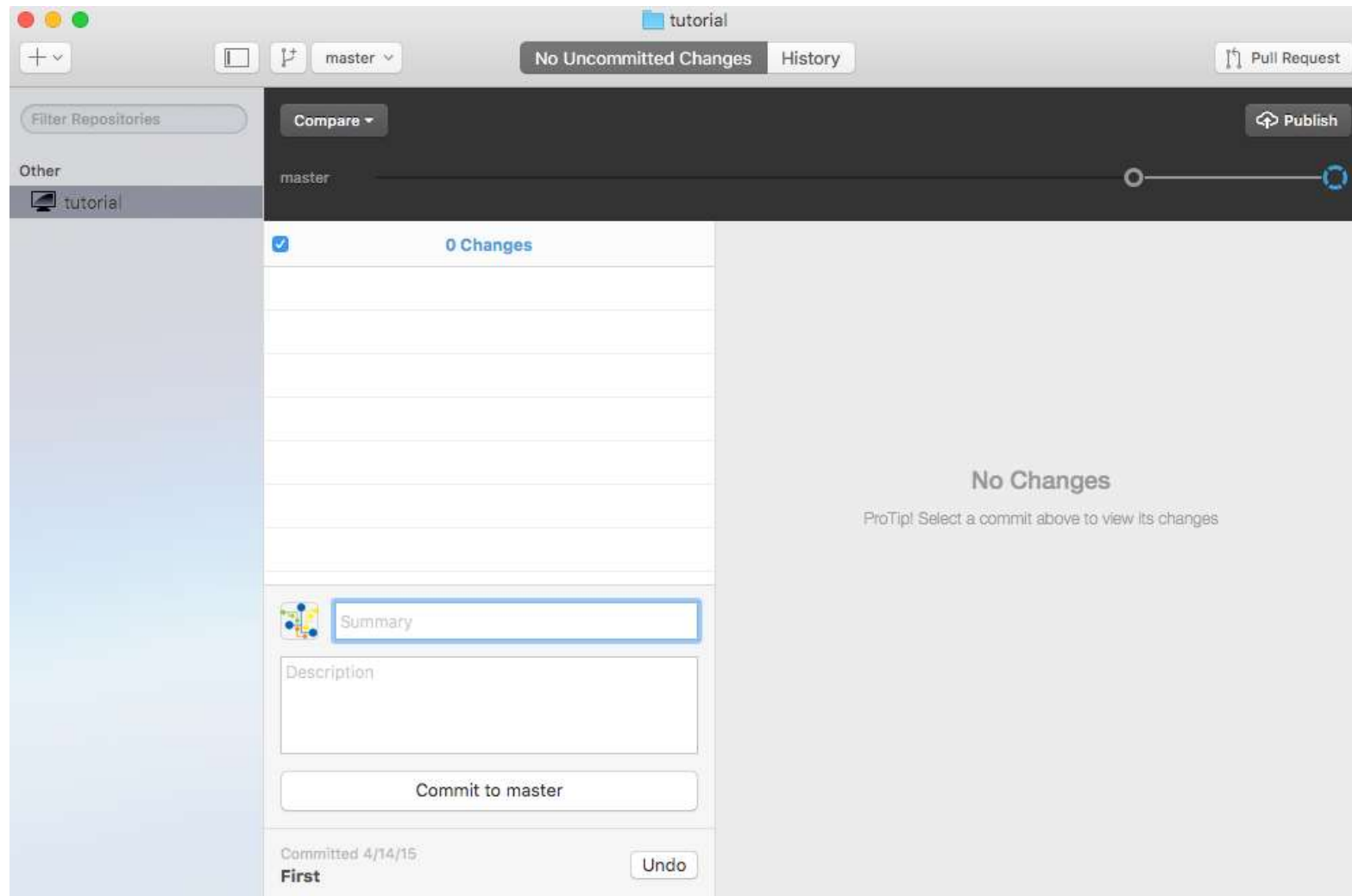
Data Science Academy



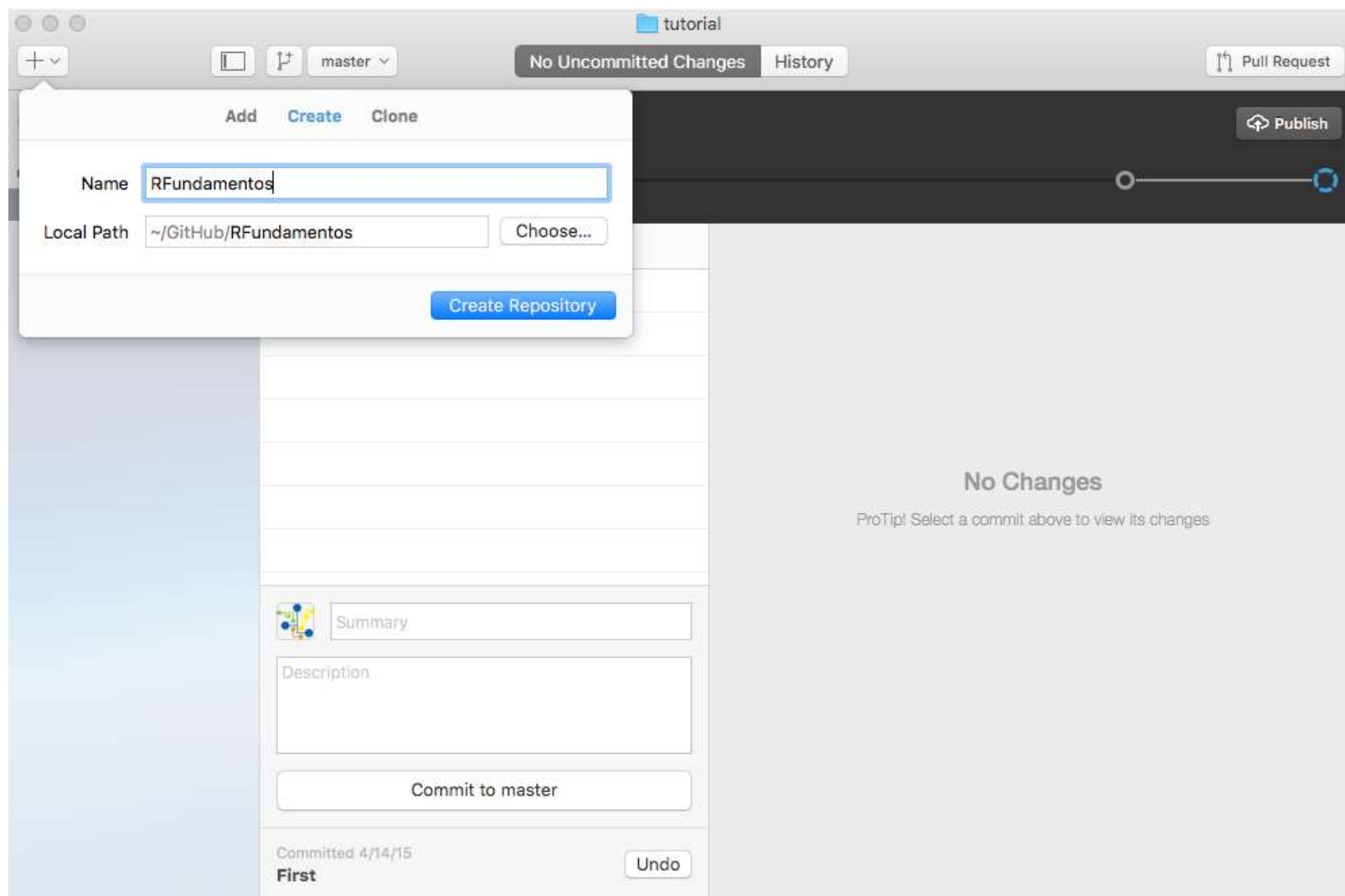
Data Science Academy



Data Science Academy

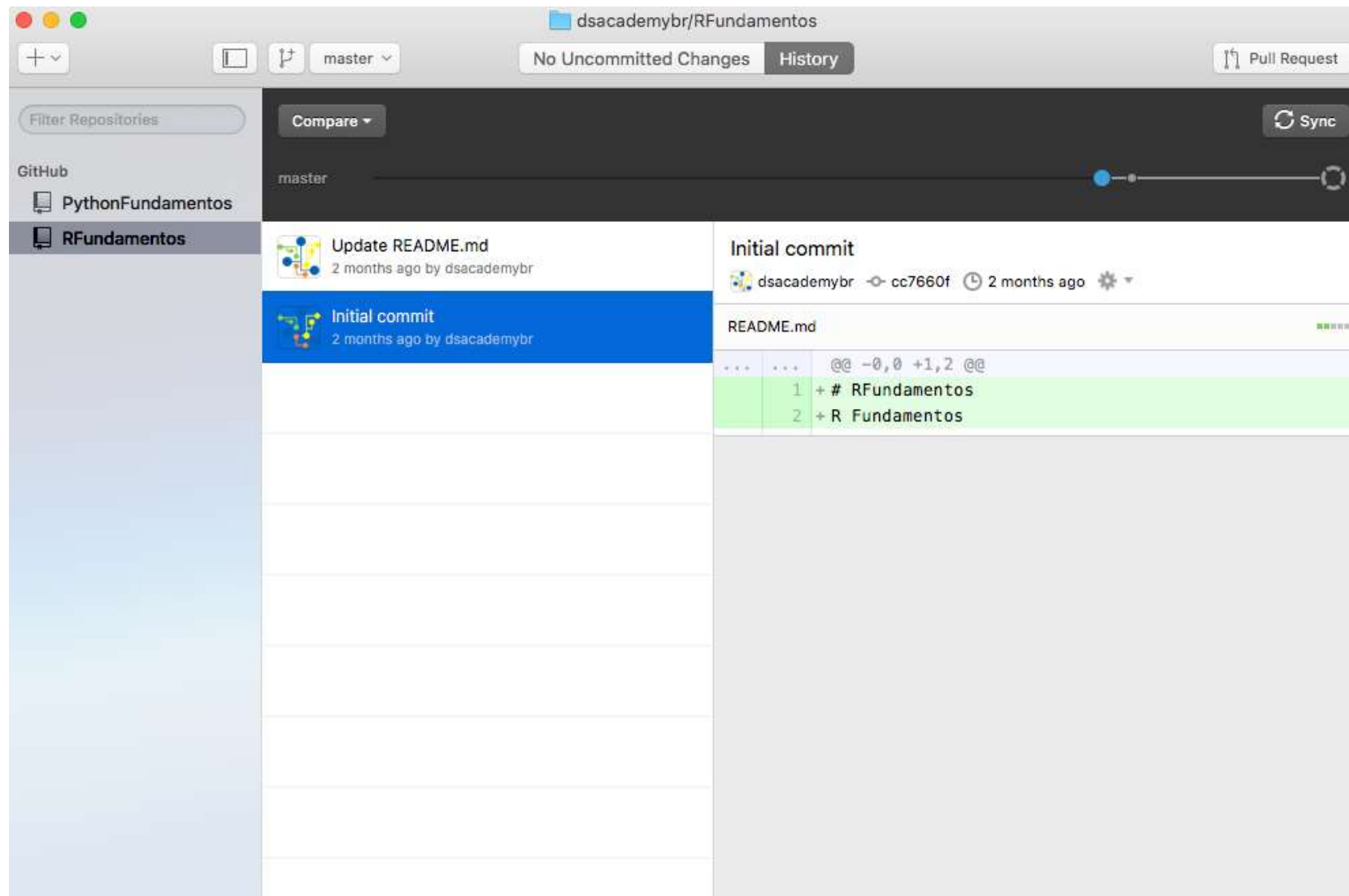


Data Science Academy



Data Science Academy







**Data Science
Academy**
dsacademybr

NEW [Add a bio](#)

 Overview

 Repositories

 Public activity

 **Edit profile**

Popular repositories

[Customize your pinned repositories](#)

 [PythonFundamentos](#)

Python Fundamentos para Análise de Dados

4 ★

 [BigDataAnalyticsHadoop](#)

0 ★

 [BigDataAnalyticsSpark](#)

0 ★

 [RFundamentos](#)

R Fundamentos

0 ★

6 contributions in the last year

[Contribution settings](#) ▼



Data Science Academy



Preparação de Documentos com R e LaTeX



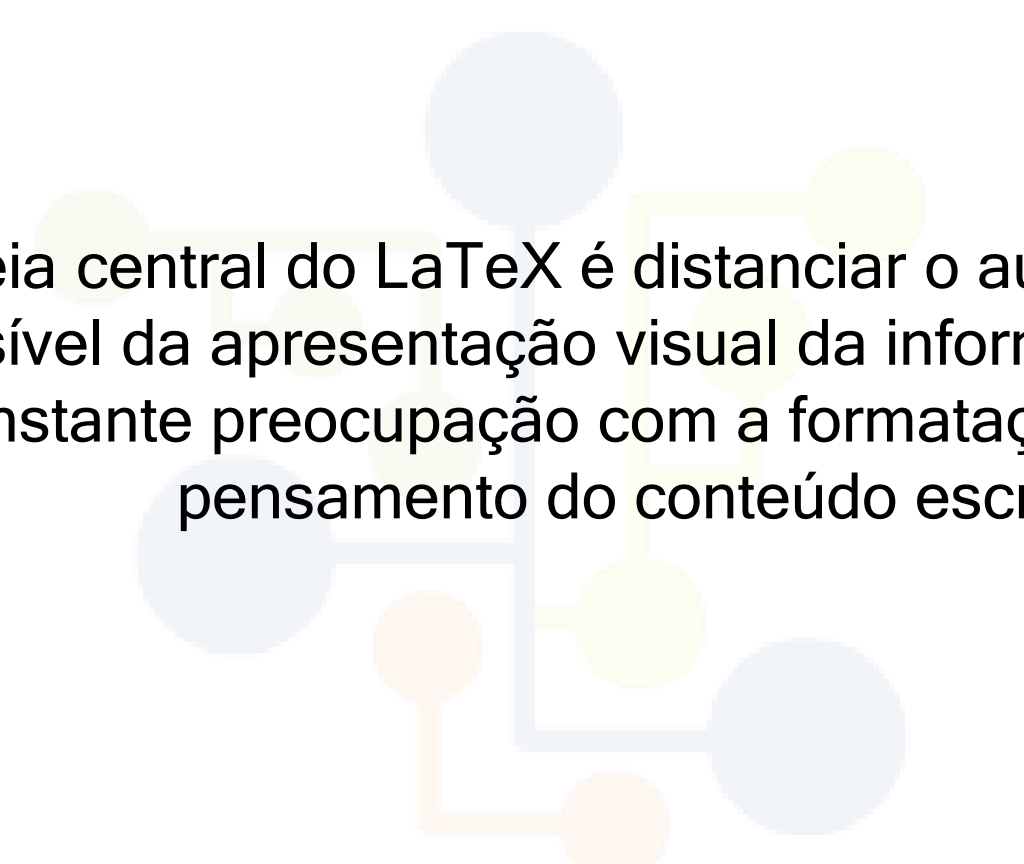

Data Science Academy



O que é LaTeX?



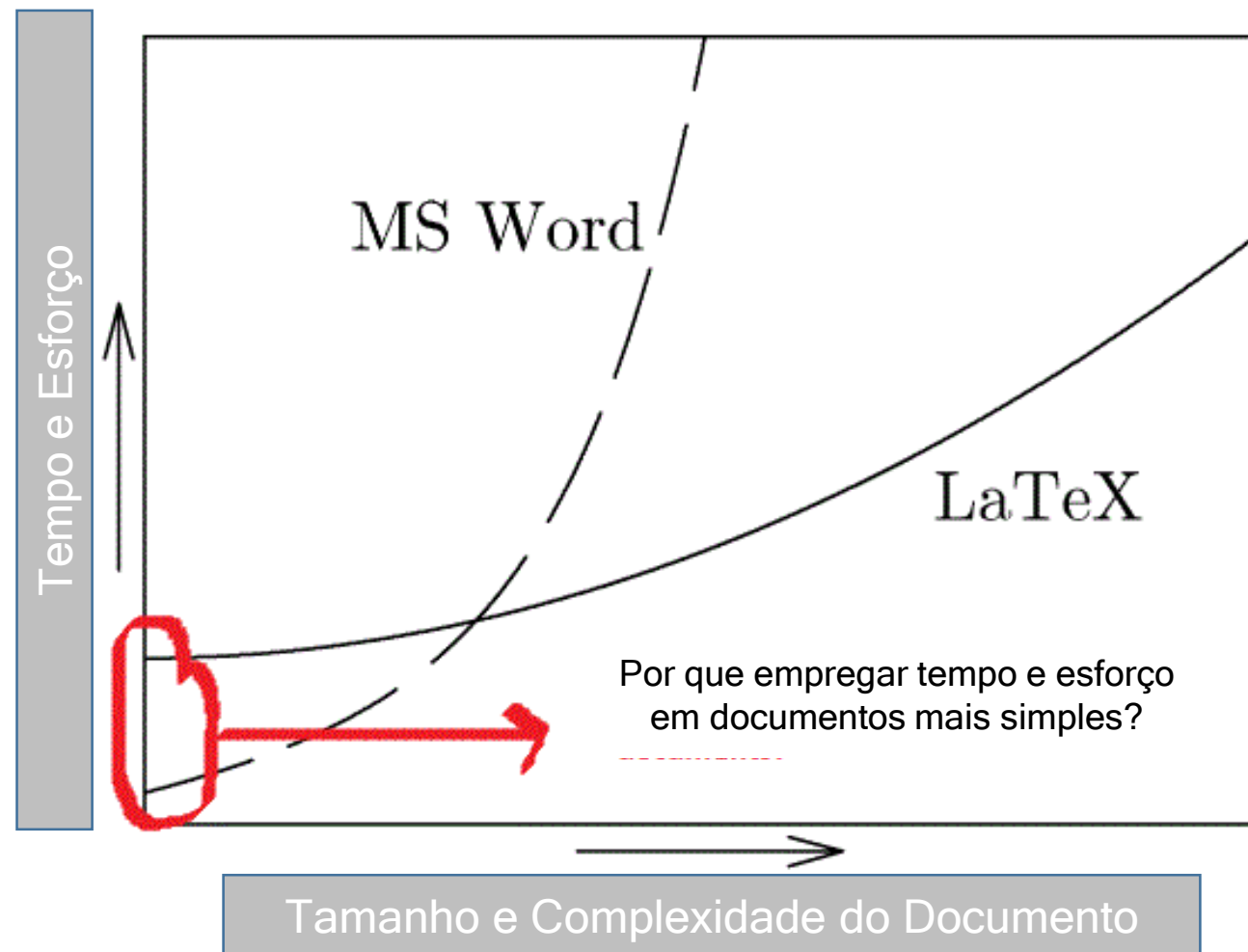
Data Science Academy



A ideia central do LaTeX é distanciar o autor o máximo possível da apresentação visual da informação, pois a constante preocupação com a formatação desvia o pensamento do conteúdo escrito



Data Science Academy



Data Science Academy



R e LaTeX

Arquivos com extensão .Rnw



Data Science Academy

R e LaTeX

```
1 \documentclass{article}
2 \usepackage[T1]{fontenc}
3
4 \begin{document}
5
6 Here is a code chunk.
7
8 <<foo, fig.height=4>>=
9 1+1
10 letters
11 chartr('xie', 'XIE', c('xie yihui', 'Yihui Xie'))
12 par(mar=c(4, 4, .2, .2)); plot(rnorm(100))
13 @
14
15 You can also write inline expressions, e.g.  $\pi = \text{\Sexpr{pi}}$ , and  $1.598673e8$  is a big number.
16
17 \end{document}
```

Arquivos com extensão .Rnw



Data Science Academy



R e LaTeX

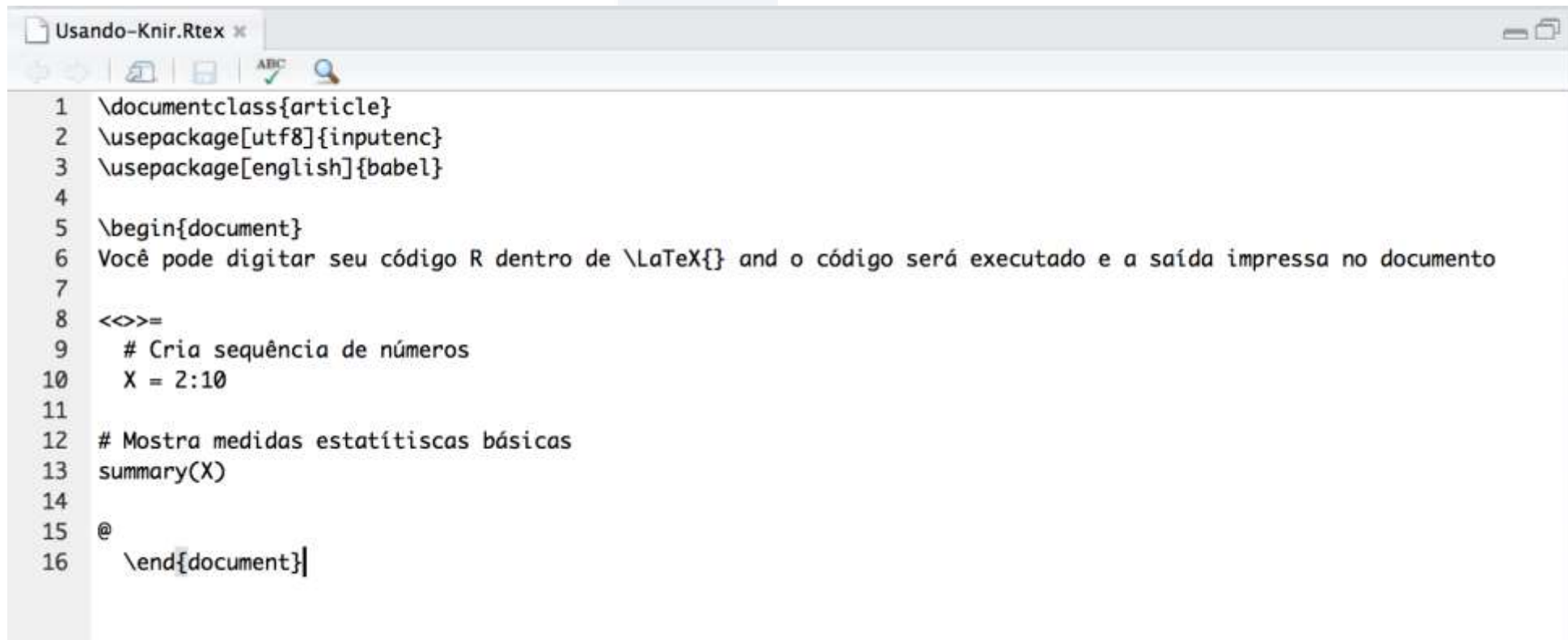
Sweave()
knit()

Reproducible Research



Data Science Academy

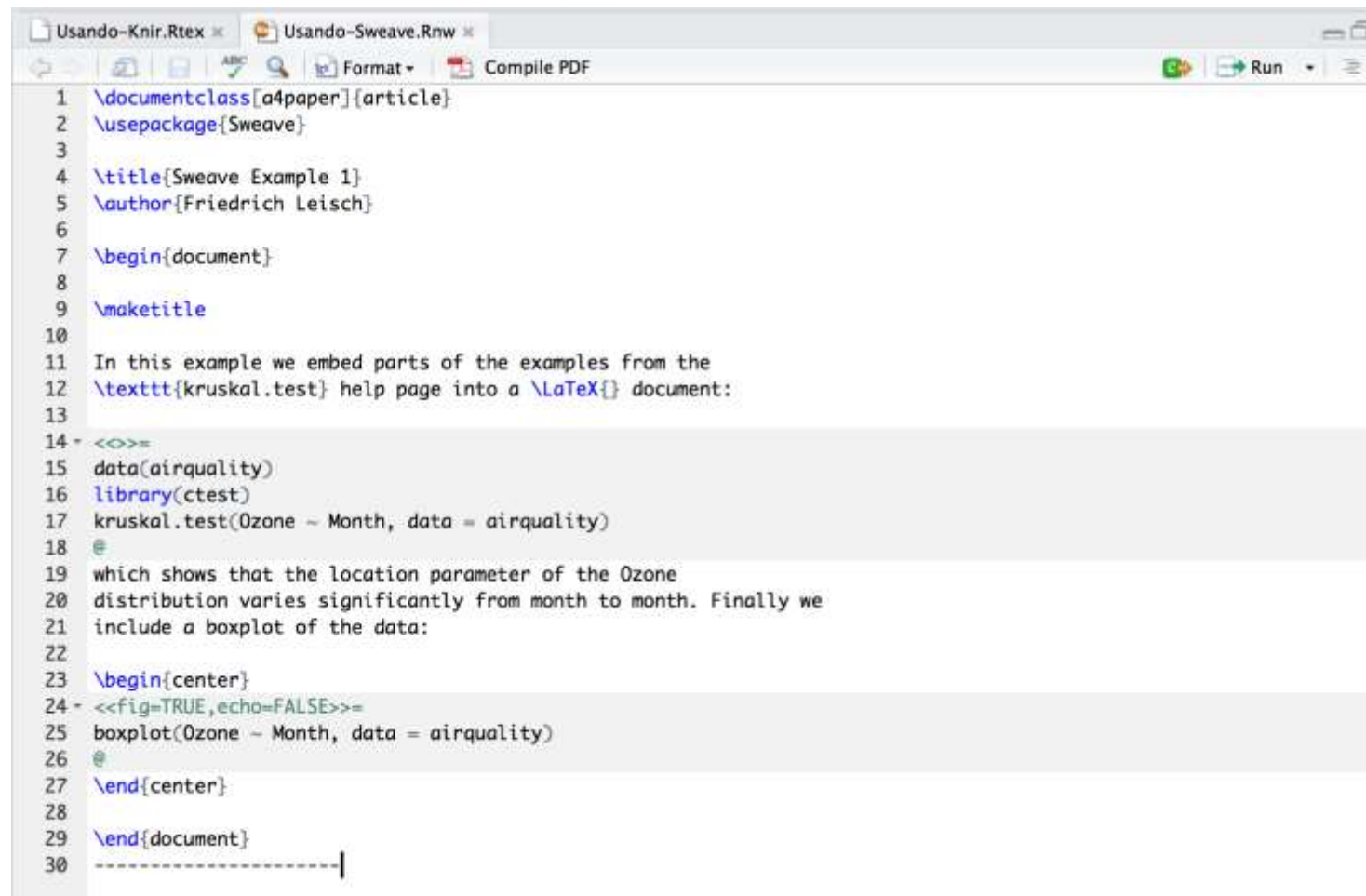
knitr()



```
1 \documentclass{article}
2 \usepackage[utf8]{inputenc}
3 \usepackage[english]{babel}
4
5 \begin{document}
6 Você pode digitar seu código R dentro de \LaTeX{} and o código será executado e a saída impressa no documento
7
8 <>=
9   # Cria sequência de números
10   X = 2:10
11
12   # Mostra medidas estatísticas básicas
13   summary(X)
14
15 @
16 \end{document}
```



Sweave()



The screenshot shows a text editor window with two tabs: 'Usando-Knitr.Rtex' and 'Usando-Sweave.Rnw'. The active tab is 'Usando-Sweave.Rnw'. The editor contains a LaTeX document structure with Sweave integration. The code is as follows:

```
1 \documentclass[a4paper]{article}
2 \usepackage{Sweave}
3
4 \title{Sweave Example 1}
5 \author{Friedrich Leisch}
6
7 \begin{document}
8
9 \maketitle
10
11 In this example we embed parts of the examples from the
12 \texttt{kruskal.test} help page into a \LaTeX{} document:
13
14 <<=>
15 data(airquality)
16 library(ctest)
17 kruskal.test(Ozone ~ Month, data = airquality)
18 @
19 which shows that the location parameter of the Ozone
20 distribution varies significantly from month to month. Finally we
21 include a boxplot of the data:
22
23 \begin{center}
24 <<fig=TRUE,echo=FALSE>>=
25 boxplot(Ozone ~ Month, data = airquality)
26 @
27 \end{center}
28
29 \end{document}
30 -----|
```



Usando o R Markdown e Criando seu Portfólio



Data Science Academy



Markdown

Linguagem de marcação de texto, que permite criar documentos a partir de texto plano



Data Science Academy

Usando o R Markdown

knitr + markdown + pandoc = rmarkdown

```
install.packages('rmarkdown')
```



Data Science Academy

Usando o R Markdown



Data Science Academy



Documentos R Markdown são totalmente reproduzíveis



Data Science Academy

Sintaxe Markdown

```
Header 1
=====
Header 2
-----
### Header 3

This is regular
text.

> This is a
blockquote.
>
> This is the
second paragraph
in the blockquote.
>
> ## This is an H2
in a blockquote
```

Saída HTML

```
<h1>Header 1</h1>
<h2>Header 2</h2>
<h3>Header 3</h3>

<p>This is regular
text.</p>

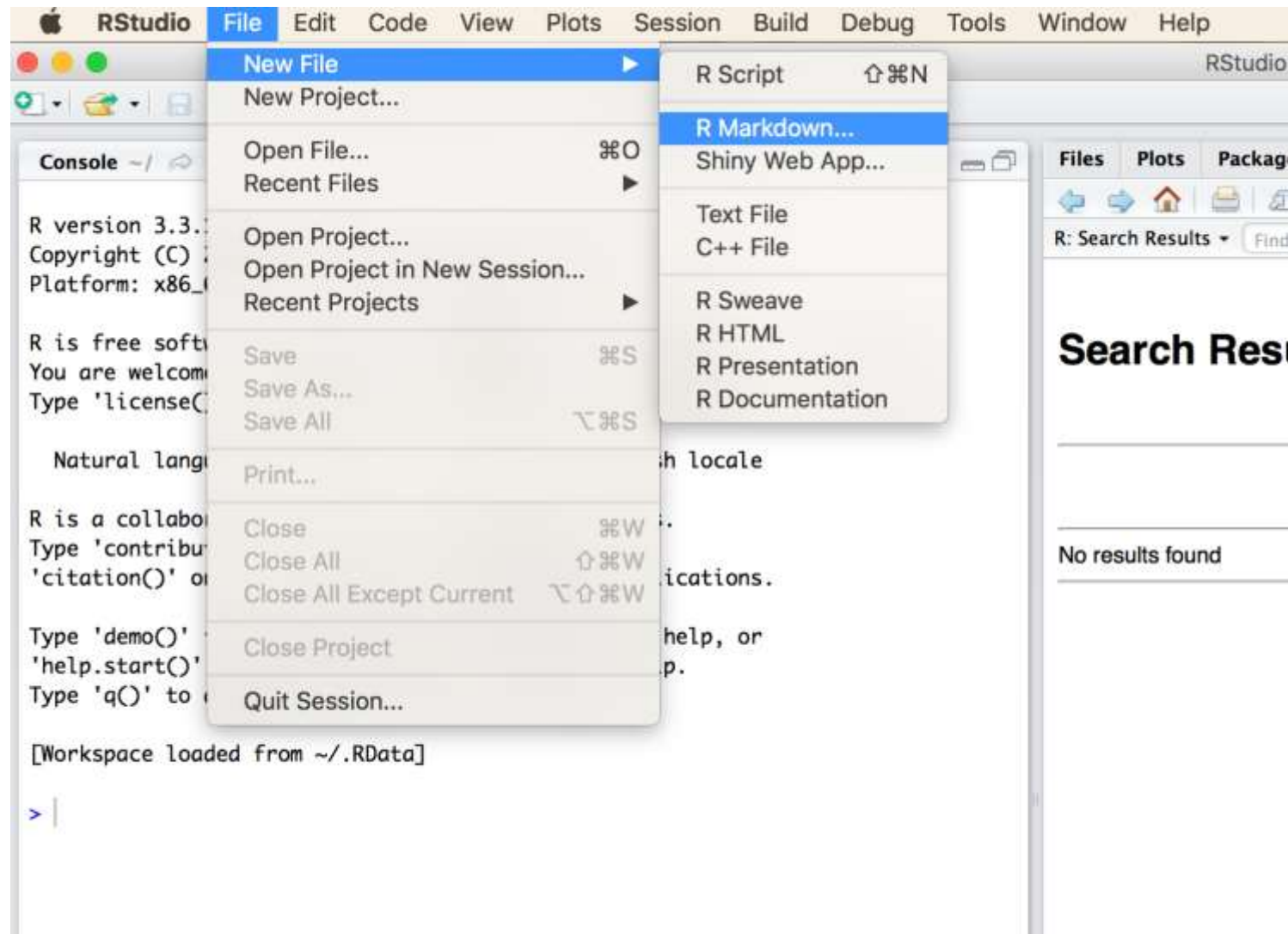
<blockquote>
<p>This is a
blockquote.</p>

<p>This is the second
paragraph in the
blockquote.</p>


<h2>This is an H2 in a
blockquote</h2>
</blockquote>
```





Data Science Academy




New R Markdown

 Document

 Presentation

 Shiny

 From Template

Title:

Author:

Default Output Format:

☒ **HTML**
Recommended format for authoring (you can switch to PDF or Word output anytime).

☐ **PDF**
PDF output requires TeX (MiKTeX on Windows, MacTeX 2013+ on OS X, TeX Live 2013+ on Linux).

☐ **Word**
Previewing Word documents requires an installation of MS Word (or Libre/Open Office on Linux).



```
Arquivo1.Rmd x
Knit Word
Run

1- ---
2 title: "Arquivo1"
3 author: "Fernanda Magalhães"
4 date: "June 15, 2016"
5 output: word_document
6- ---
7
8- ```{r setup, include=FALSE}
9 knitr::opts_chunk$set(echo = TRUE)
10- ```
11
12- ## R Markdown
13
14 Este é um documento R Markdown. Markdown é uma forma simples de formatar seu código para saída em arquivos html, pdf e MS Word.
15
16 Quando você clica no botão Knit um documento será gerado, que inclui conteúdo e o output do seu código.
17
18- ```{r cars}
19 summary(cars)
20- ```
21
22- ## Incluindo Gráficos
23
24 Você pode incluir gráficos:
25
26- ```{r pressure, echo=FALSE}
27 plot(pressure)
28- ```
29
30 |
```

```
Arquivo1.Rmd x
Knit Word
Run

1- ---
2 title: "Arquivo1"
3 author: "Fernanda Magalhães"
4 date: "June 15, 2016"
5 output: word_document
6- ---

Knit to HTML
Knit to PDF
Knit to Word
```



Data Science Academy



Arquivo1

Fernanda Magalhães

June 15, 2016

R Markdown

Este é um documento R Markdown. Markdown é uma forma simples de formatar seu código para saída em arquivos html, pdf e MS Word.

Quando você clica no botão **Knit** um documento será gerado, que inclui conteúdo e o output do seu código.

```
summary(cars)
```

```
##      speed      dist
## Min.   : 4.0    Min.   : 2.00
## 1st Qu.:12.0    1st Qu.: 26.00
## Median :15.0    Median : 36.00
## Mean   :15.4    Mean   : 42.98
## 3rd Qu.:19.0    3rd Qu.: 56.00
## Max.   :25.0    Max.   :120.00
```

Incluindo Gráficos

Você pode incluir gráficos:



Arquivo1

Fernanda Magalhães

June 15, 2016

R Markdown

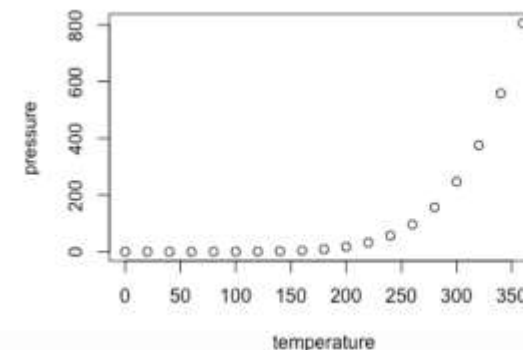
Este é um documento R Markdown. Markdown é uma forma simples de formatar seu código para saída em arquivos html, pdf e MS Word.

Quando você clica no botão **Knit** um documento será gerado, que inclui conteúdo e o output do seu código.

```
summary(cars)
##      speed      dist
## Min.   : 4.0    Min.   : 2.00
## 1st Qu.:12.0    1st Qu.: 26.00
## Median :15.0    Median : 36.00
## Mean   :15.4    Mean   : 42.98
## 3rd Qu.:19.0    3rd Qu.: 56.00
## Max.   :25.0    Max.   :120.00
```

Incluindo Gráficos

Você pode incluir gráficos:



Academy



`knit2html()`

Converte .Rmd em .Html



Data Science Academy



Obrigado!



Data Science Academy