

Inteligência Artificial, Ciência de Dados e Estatística: trajetória, dicas e desafios

Gabriela de Queiroz

@gdequeiroz | linktr.ee/gdq

slides: bit.ly/pet-ufpr

Gabriela de Queiroz

Sr. Engineering and Data Science Manager, IBM

- Fundadora do R-Ladies (rladies.org)
- Fundadora do AI Inclusive (ai-inclusive.org)



- Graduação em Estatística (UERJ)
- Mestrado em Epidemiologia (ENSP/Fiocruz)
- Mestrado em Estatística (CSUEB)

Data Scientist + Developer Advocate + Open Source Developer + Manager +
Statistician + Epidemiologist + Community Builder + Mentor + Speaker + Educator

Researcher/Statistician



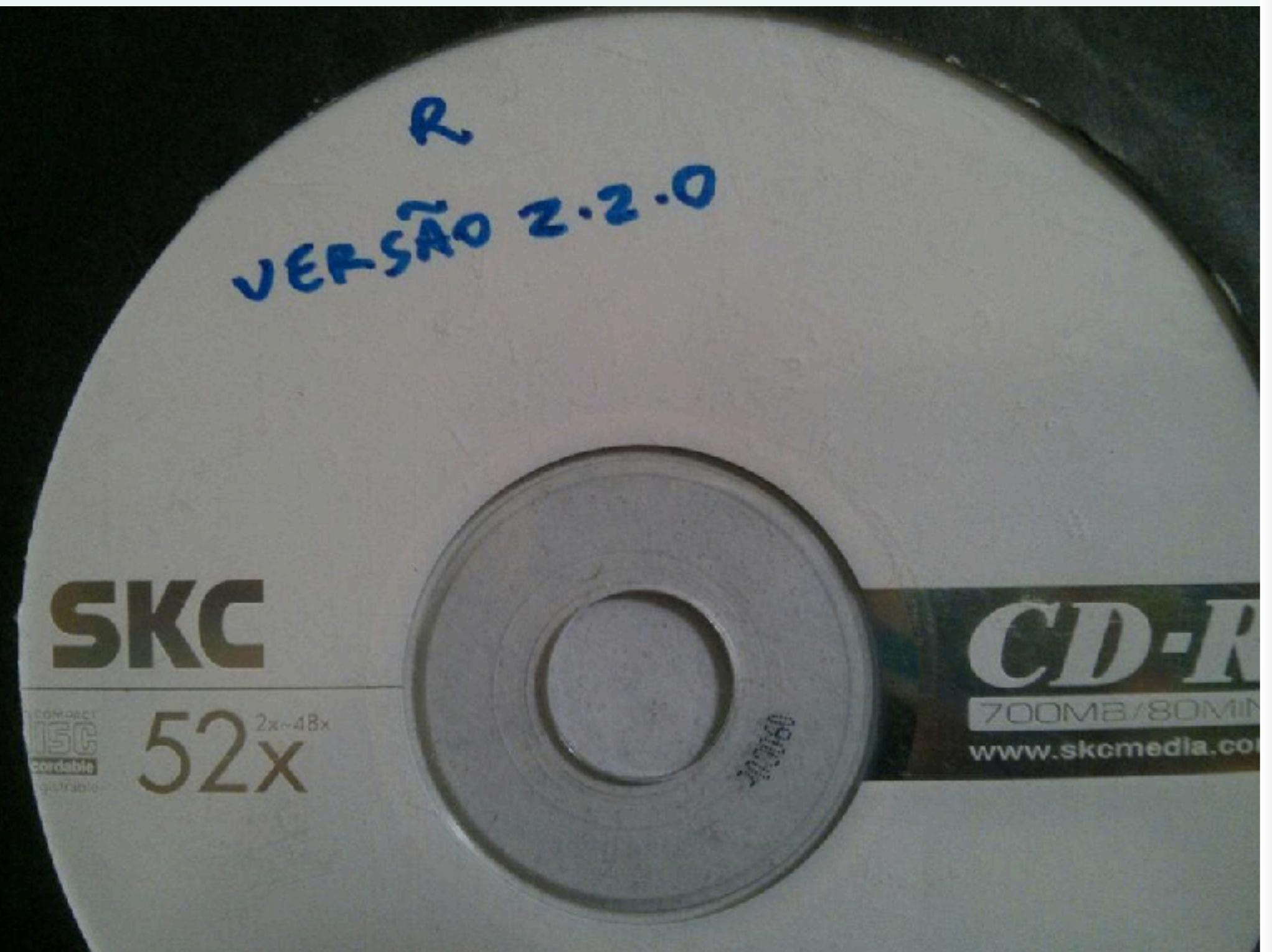
Undergraduate

- STATE UNIVERSITY
- PUBLIC AND FREE UNIVERSITY
- BACHELOR IN STATISTICS**

Grad School

- SCIENTIFIC INSTITUTION FOR RESEARCH
- PUBLIC AND FREE
- MSC. IN EPIDEMIOLOGY**

First exposure to R ❤



[R 2.2.0](#) (October, 2005)



[R 3.1.0](#) (April, 2014)
[R 3.0.3](#) (March, 2014)
[R 3.0.2](#) (September, 2013)
[R 3.0.1](#) (May, 2013)
[R 3.0.0](#) (April, 2013)
[R 2.15.3](#) (March, 2013)
[R 2.15.2](#) (October, 2012)
[R 2.15.1](#) (June, 2012)
[R 2.15.0](#) (March, 2012)
[R 2.14.2](#) (February, 2012)
[R 2.14.1](#) (December, 2011)
[R 2.14.0](#) (November, 2011)
[R 2.13.2](#) (September, 2011)
[R 2.13.1](#) (July, 2011)
[R 2.13.0](#) (April, 2011)
[R 2.12.2](#) (February, 2011)
[R 2.12.1](#) (December, 2010)
[R 2.12.0](#) (October, 2010)
[R 2.11.1](#) (May, 2010)
[R 2.11.0](#) (April, 2010)
[R 2.10.1](#) (December, 2009)
[R 2.10.0](#) (October, 2009)
[R 2.9.2](#) (August, 2009)
[R 2.9.1](#) (June, 2009)
[R 2.9.0](#) (April, 2009)
[R 2.8.1](#) (December, 2008)
[R 2.8.0](#) (October, 2008)
[R 2.7.2](#) (August, 2008)
[R 2.7.1](#) (June, 2008)
[R 2.7.0](#) (April, 2008)
[R 2.6.2](#) (February, 2008)
[R 2.6.1](#) (November, 2007)
[R 2.6.0](#) (October, 2007)
[R 2.5.1](#) (July, 2007)
[R 2.5.0](#) (April, 2007)
[R 2.4.1](#) (December, 2006)
[R 2.4.0](#) (October, 2006)
[R 2.3.1](#) (June, 2006)
[R 2.3.0](#) (April, 2006)
[R 2.2.1](#) (December, 2005)
[R 2.2.0](#) (October, 2005)

[R 4.0.3](#) (October, 2020)
[R 4.0.2](#) (June, 2020)
[R 4.0.1](#) (June, 2020)
[R 4.0.0](#) (April, 2020)
[R 3.6.3](#) (February, 2020)
[R 3.6.2](#) (December, 2019)
[R 3.6.1](#) (July, 2019)
[R 3.6.0](#) (April, 2019)
[R 3.5.3](#) (March, 2019)
[R 3.5.2](#) (December, 2018)
[R 3.5.1](#) (July, 2018)
[R 3.5.0](#) (April, 2018)
[R 3.4.4](#) (March, 2018)
[R 3.4.3](#) (November, 2017)
[R 3.4.2](#) (September, 2017)
[R 3.4.1](#) (June, 2017)
[R 3.4.0](#) (April, 2017)
[R 3.3.3](#) (March, 2017)
[R 3.3.2](#) (October, 2016)
[R 3.3.1](#) (June, 2016)
[R 3.3.0](#) (April, 2016)
[R 3.2.5](#) (April, 2016)
[R 3.2.4](#) (March, 2016)
[R 3.2.3](#) (December, 2015)
[R 3.2.2](#) (August, 2015)
[R 3.2.1](#) (June, 2015)
[R 3.2.0](#) (April, 2015)
[R 3.1.3](#) (March, 2015)
[R 3.1.2](#) (October, 2014)
[R 3.1.1](#) (July, 2014)

The R Inferno

Patrick Burns¹

30th April 2011

Contents

List of Figures

List of Tables

1 Falling into the Floating Point Trap

2 Growing Objects

3 Failing to Vectorize

- 3.1 Subscripting
- 3.2 Vectorized if
- 3.3 Vectorization impossible

4 Over-Vectorizing

- ng Functions
- ction
- city
- tency

Global Assignments

- n Object Orientation
- hods
- generic functions
- methods
- inheritance
- hods

An Introduction to R

Notes on R: A Programming Environment for Data Analysis and Graphics
Version 2.5.1 (2007-06-27)

W. N. Venables, D. M. Smith
and the R Development Core Team

1 Introduction and preliminaries

- 1.1 The R environment
- 1.2 Relational software and documentation
- 1.3 R and statistics
- 1.4 R and the window system
- 1.5 Using R interactively
- 1.6 An introductory session
- 1.7 Getting help with functions and features
- 1.8 R commands; case sensitivity, etc.
- 1.9 Recall and correction of previous commands
- 1.10 Executing commands from or diverting output to a file
- 1.11 Data permanency and removing objects

2 Simple manipulations; numbers and vectors

- 2.1 Vectors and assignment
- 2.2 Vector arithmetic
- 2.3 Generating regular sequences
- 2.4 Logical vectors
- 2.5 Missing values
- 2.6 Character vectors
- 2.7 Index vectors; selecting and modifying subsets of a data set
- 2.8 Other types of objects

3 Objects, their modes and attributes

- 3.1 Intrinsic attributes: mode and length
- 3.2 Changing the length of an object
- 3.3 Getting and setting attributes
- 3.4 The class of an object

4 Ordered and unordered factors

Multiplicity Study of Air Pollution and Mortality in Latin America (the ESCALA Study)

BACKGROUND

For nearly two decades, scientists seeking to understand the role that air pollution might play in population health effects have relied heavily on epidemiologic studies known as time-series studies. Time-series studies use information on daily changes in air pollutant concentrations and daily counts of mortality and morbidity. Although initially conducted at the individual city level, coordinated analyses across many cities have recently emerged as the tool of choice for developing more reliable and comparable estimates of the short-term effects of air pollution on health in regions around the world. HEI has a long-standing interest in these coordinated analyses; it has funded studies such as the National Morbidity, Mortality, and Air Pollution Study; Air Pollution and Health: A European and North American Approach; and Public Health and Air Pollution in Asia.

The present study, referred to hereafter by its Spanish acronym ESCALA (Estudio de Salud y Contaminación del Aire en Latinoamérica), was initiated to address underlying data and methodologic limitations in the epidemiologic literature on the health effects of air pollution in Latin America that had been identified in a 2005 review by the Pan American Health Organization. The William and Flora Hewlett Foundation, which has a strong interest in understanding air pollution and health in Latin America, provided HEI with supplemental support to address gaps in the evidence necessary to inform regulatory decisions, and in the process to build a network of health experts capable of carrying out research on air pollution in the future. The multicenter study was led by Dr. Isabelle Romieu, then at the Instituto Nacional de Salud Pública in México, in collaboration with Dr. Nelson Gouveia in Brazil and Dr. Luis Cifuentes in Chile.

With the individual city data, the investigators also explored two-pollutant models, in which PM₁₀ results were controlled for the presence of ozone and vice versa; whether the association of ozone with mortality differed by warm and cold season;

APPROACH

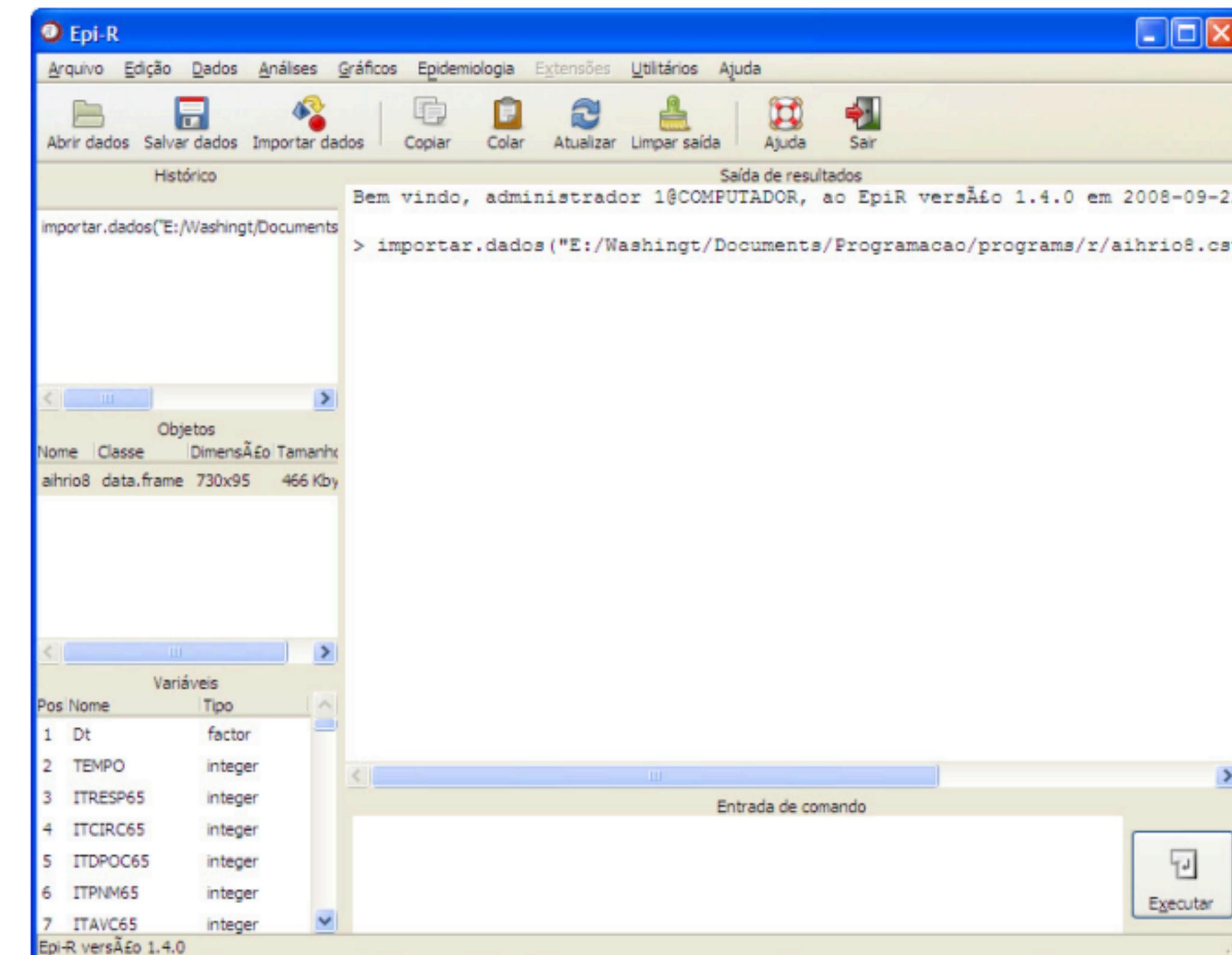
The primary objective of the ESCALA study was to estimate the effect of daily exposures to PM₁₀ (particulate matter $\leq 10 \mu\text{m}$ in aerodynamic diameter) and to ozone on daily mortality from several causes (all natural causes, cardiopulmonary disease, respiratory disease, cardiovascular disease, cerebrovascular-stroke, and chronic obstructive pulmonary disease) and for several age groups (all-age, ≥ 65 years, < 1 year, 1–4 years, 1–14 years) in nine Latin American cities, and for the region as a whole, using a common analytic framework. The nine cities were Monterrey, Toluca, and Mexico City in México; Rio de Janeiro, São Paulo, and Porto Alegre in Brazil; and Santiago, Concepción, and Temuco in Chile. Of these, three cities (Porto Alegre, Concepción, and Temuco) were excluded from the ozone analyses because of the lack of adequate ozone monitoring data.

In the first stage of the analyses, the investigators estimated the percentage change in the risk of mortality per $10\text{-}\mu\text{g}/\text{m}^3$ increase in PM₁₀ or ozone for each combination of age group and cause of death for the individual cities in each country. They followed a common protocol for fitting the widely used Poisson regression models to the air pollution and mortality time-series data in each city while controlling for other factors that might also explain the temporal patterns of mortality (e.g., temperature, humidity, season, day-of-the-week, holidays). The investigators also carried out analyses to test the sensitivity of the results to various details of the models. Ultimately, the final models used in the individual cities were chosen to fit specific patterns of mortality in those cities.

With the individual city data, the investigators also explored two-pollutant models, in which PM₁₀ results were controlled for the presence of ozone and vice versa; whether the association of ozone with mortality differed by warm and cold season;



A graphic user interface oriented to epidemiological data analysis

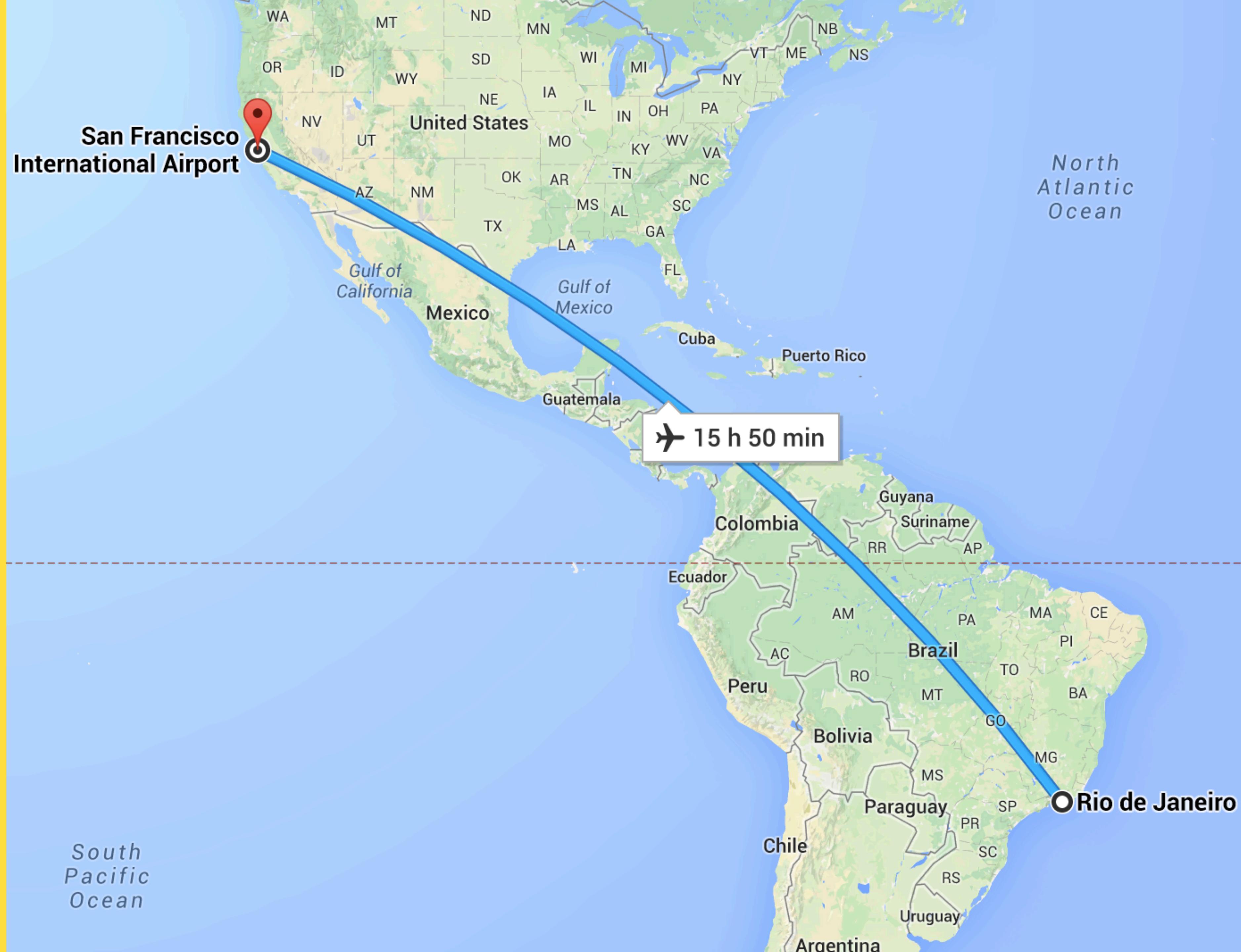


2012

Rio de Janeiro



San Francisco



Meetup

Data Visualization Group in the Bay Area

San Francisco Hadoop Users

SF Bayarea Machine Learning



#SFMysql Meetup





Fundada em outubro de 2012.
A ideia era retribuir à comunidade e criar um local onde as pessoas se sentissem confortáveis, seguras e bem-vindas.
Um lugar onde as pessoas pudessem fazer perguntas, aprender juntas e compartilhar.



31
OCT

Wednesday, October 31, 2012

Introduction to R (beginners and pre-beginners)



Hosted by
Gabriela de Queiroz

Details

Hello R-ladies!

The first meetup will take place on October 31st at the Google office in San Francisco.

For this first meetup, we'll do an introduction to R. We'll go over the following topics:

installing R setting up an R environment (RStudio) basic commands (open files, simple dataset manipulation, simple plots, etc) loading packages the help function and how to read its output

All you need is your laptop and charger.

We look forward to seeing you!

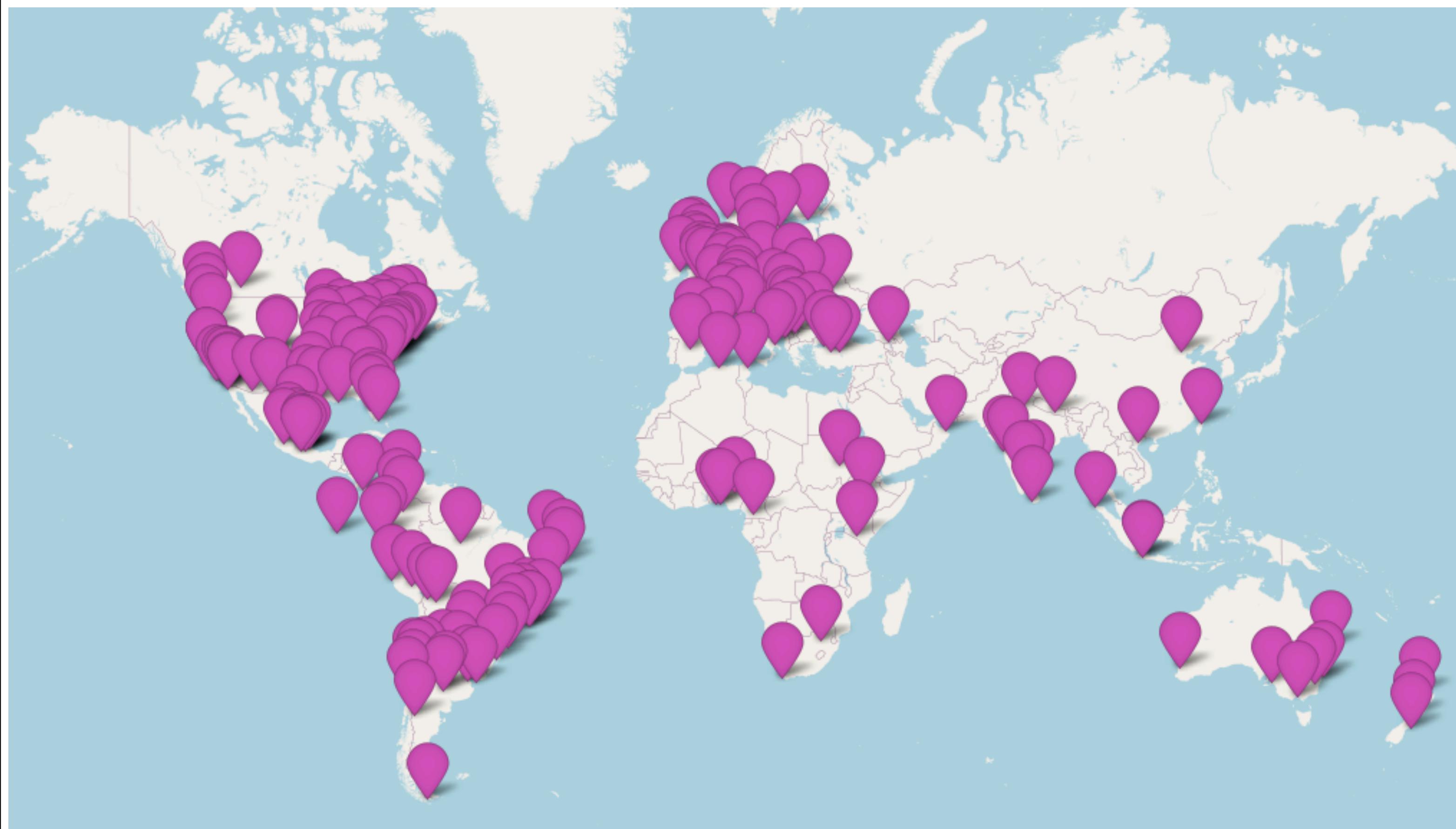
R-Ladies



54 
R-Ladies Countries

193 
R-Ladies Cities

71618 
R-Ladies members on meetup.com



	city	state	country	dt_created	members
41	São Paulo		BR	2018-02-10	1004
38	Belo Horizonte		BR	2018-04-20	800
34	Niterói		BR	2018-06-04	580
51	Rio de Janeiro		BR	2017-02-27	579
28	Florianópolis		BR	2019-04-07	441
43	Porto Alegre		BR	2017-10-30	346
7	Natal		BR	2020-06-07	241
32	Salvador		BR	2018-07-23	192
22	Recife		BR	2019-09-02	147
25	Goiania		BR	2019-05-06	146
18	Vitória		BR	2019-09-29	132
6	Fortaleza		BR	2020-06-09	69
26	Lavras		BR	2019-04-16	41
5	Curitiba		BR	2020-06-12	36
14	Ribeirão Preto		BR	2020-03-06	32
21	Manaus		BR	2019-09-11	29

52
R-Ladies groups in Latin America



The R Foundation @_R_Foundation · 26 Oct 2018

We welcome [@gdequeiroz](#), [@edzerpebesma](#) and [@henrikbengtsson](#), elected as ordinary members of the R Foundation in recognition of their services to the R community.

2018

R Foundation Membership

Ordinary members

- Douglas Bates (USA)
- Henrik Bengtsson (Sweden, USA)
- Roger Bivand (Norway)
- Jennifer Bryan (Canada)
- John Chambers (USA)
- Di Cook (Australia)
- Peter Dalgaard (Denmark)
- Dirk Eddelbuettel (USA)
- John Fox (Canada)
- Robert Gentleman (USA)
- Bettina Grün (Austria)
- Frank Harrell (USA)

- Kurt Hornik (Austria)
- Torsten Hothorn (Switzerland)
- Stefano Iacus (Italy)
- Ross Ihaka (New Zealand)
- Julie Josse (France)
- Tomas Kalibera (Czechia, USA)
- Michael Lawrence (USA)
- Friedrich Leisch (Austria)
- Uwe Ligges (Germany)
- Thomas Lumley (USA, New Zealand)
- Martin Mächler (Switzerland)
- Martin Morgan (USA)
- Paul Murrell (New Zealand)

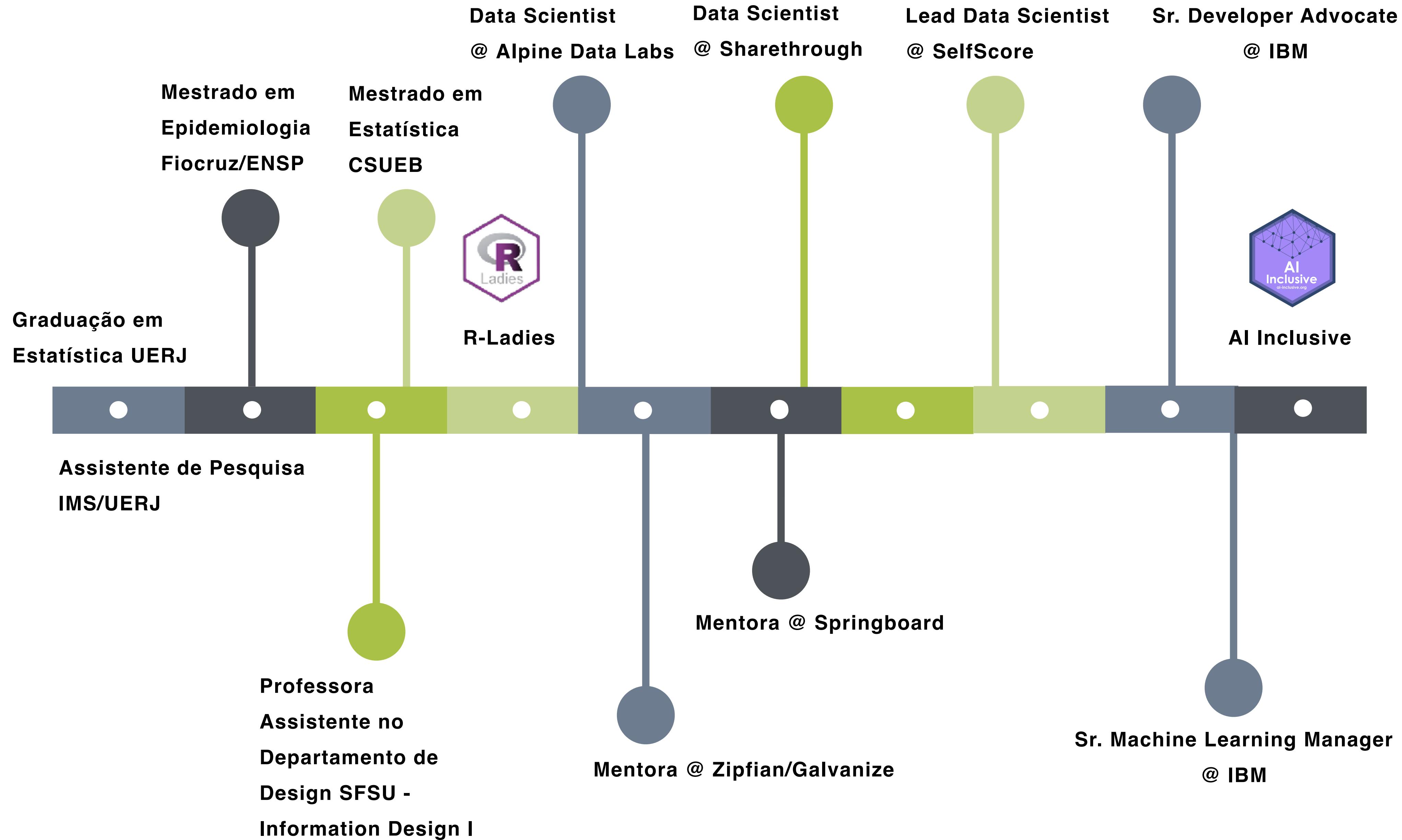
- Balasubramanian Narasimhan (USA)
- Martyn Plummer (UK)
- Edzer Pebesma (Germany)
- Gabriela de Queiroz (USA)
- Deepayan Sarkar (India)
- Marc Schwartz (USA)
- Duncan Temple Lang (USA)
- Luke Tierney (USA)
- Heather Turner (UK)
- Simon Urbanek (Germany, USA)
- Hadley Wickham (USA)
- Achim Zeileis (Austria)

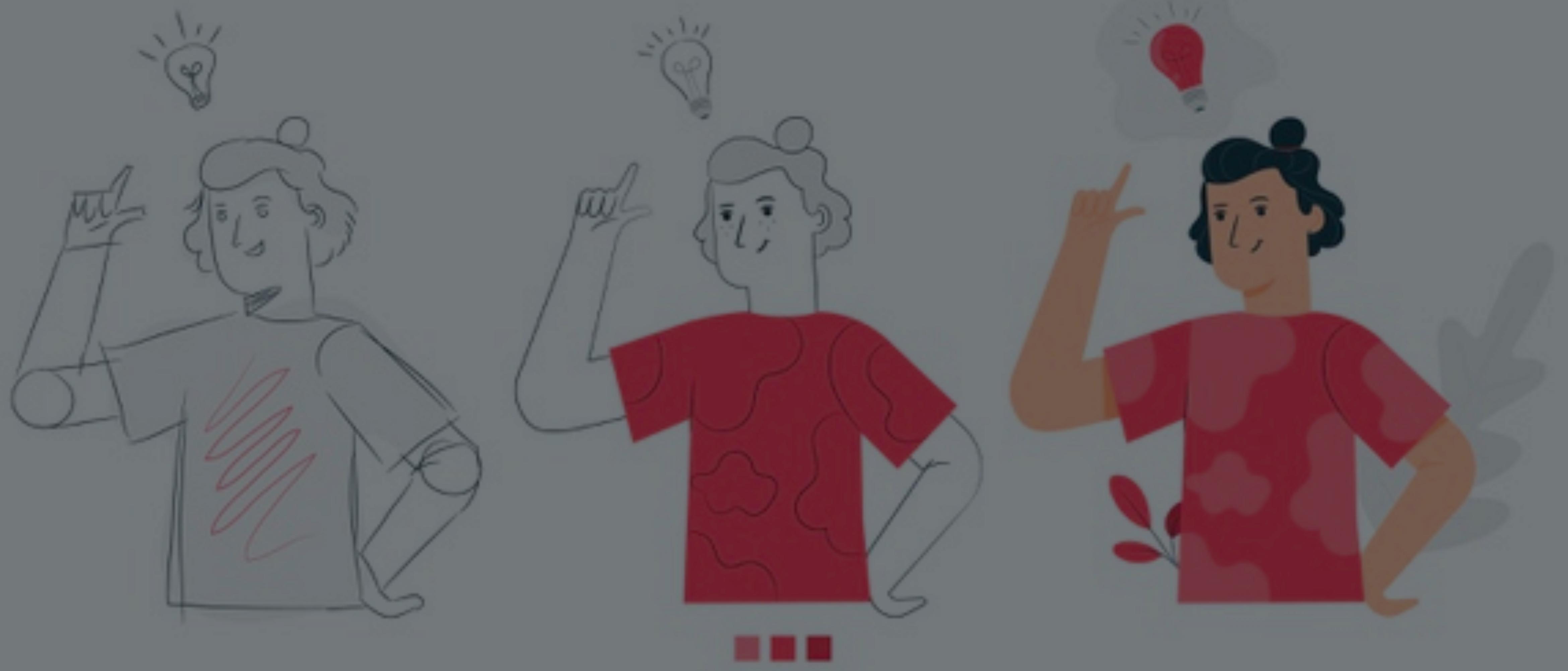
Vamos começar
com as perguntas





COMO FOI A SUA TRAJETÓRIA ATÉ
CHEGAR À IBM?





COMO É O DIA-A-DIA DE UM GERENTE?
É MUITO DIFERENTE DO QUE FAZIA ANTES?

Reuniões, muitas reuniões

Mon 3/23

Tue 3/24

Wed 3/25

Thu 3/26

9 – 10 Meeting with X Team	9 – 10 Meeting with X Team	9 – 10 Meeting with Z Team	9 – 10 Meeting with Y Team
10 – 11 Project Planning Session	10 – 11 Press Interview	10 – 11 Project Planning Session	10 – 11 Corporate Event update
11 – 12p Call with East Coast office	11 – 12:30p Presentation to executive team	11 – 12p Coaching session	11 – 12p Presentation to Mobile team
12p – 1p Management Luncheon	12:30p – 1:30p Call with Tony	12p – 1p Lunch with CEO	12p – 1p Team Lunch
1p – 2p 1 on 1 with Tony	1:30p – 3p White Board session with Antoine	1p – 2p Project Z Meeting	1p – 2p 1 on 1 with Charles
2p – 3p Review session	2p – 3p 1 on 1 with Amanda	2p – 3p 1 on 1 with Tom	2p – 3p 1 on 1 with Taylor
3p – 5p Leadership team planning session	3p – 4p Meeting with Design team	3p – 5p Company All Hands	3p – 4p Meeting with engineering
	4p – 5p HR Mandated training		4p – 5p New project Kickoff
5p – 6p Call with Australia office	5p – 6p Drinks with Thomas	5p – 6p Leadership Team Dinner	5p – 6p Speak at Event

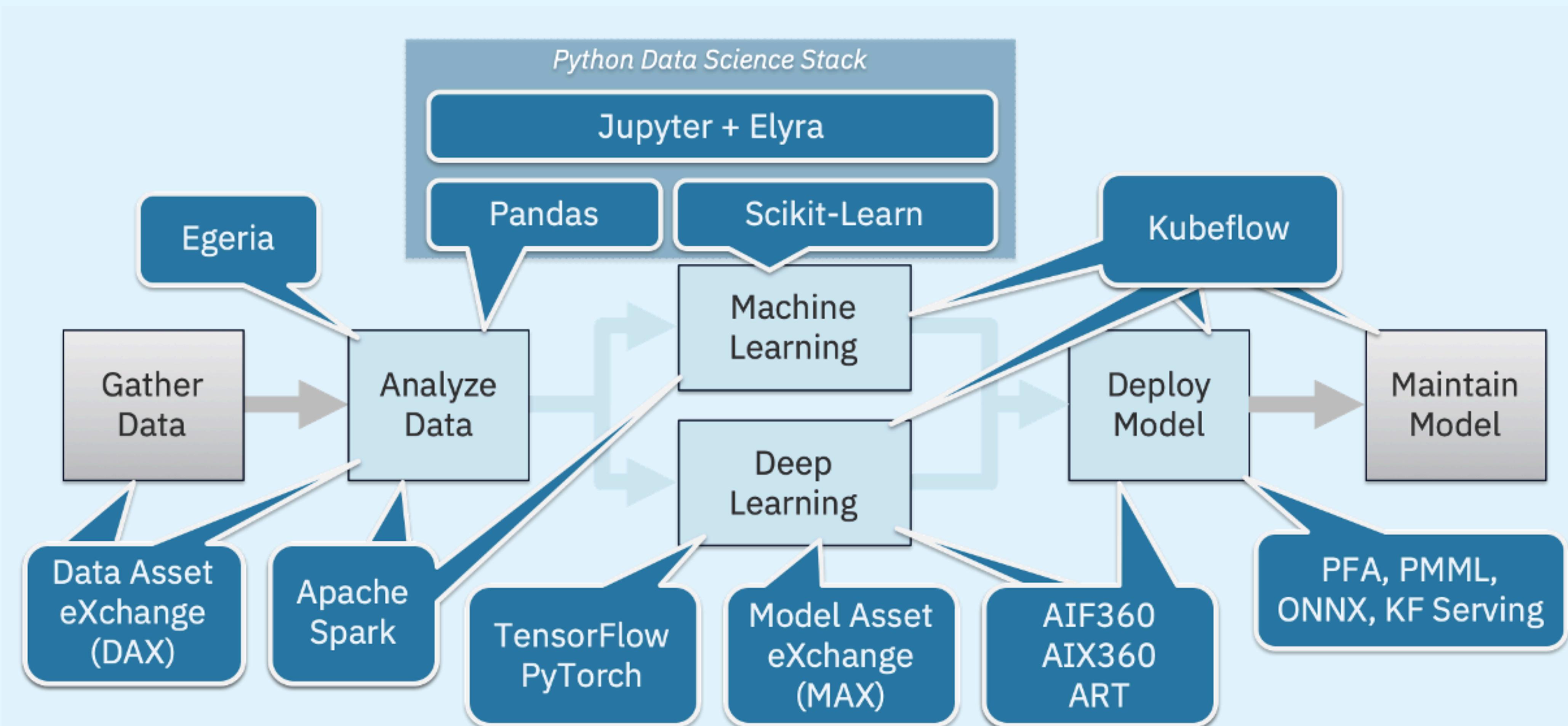
- Definir as métricas e objetivos
- Definir impacto
- Definir processo (sprints, documentação, etc)
- Criar oportunidades para o time
- Colaborar com outros times
- Criar roadmap
- Motivar a equipe
- Etc



A IBM É ENORME ... PODE CONTAR UM POUCO
DA ORGANIZAÇÃO ONDE VOCÊ TRABALHA?

We build tools to make AI accessible and available to everybody

(codait.org)

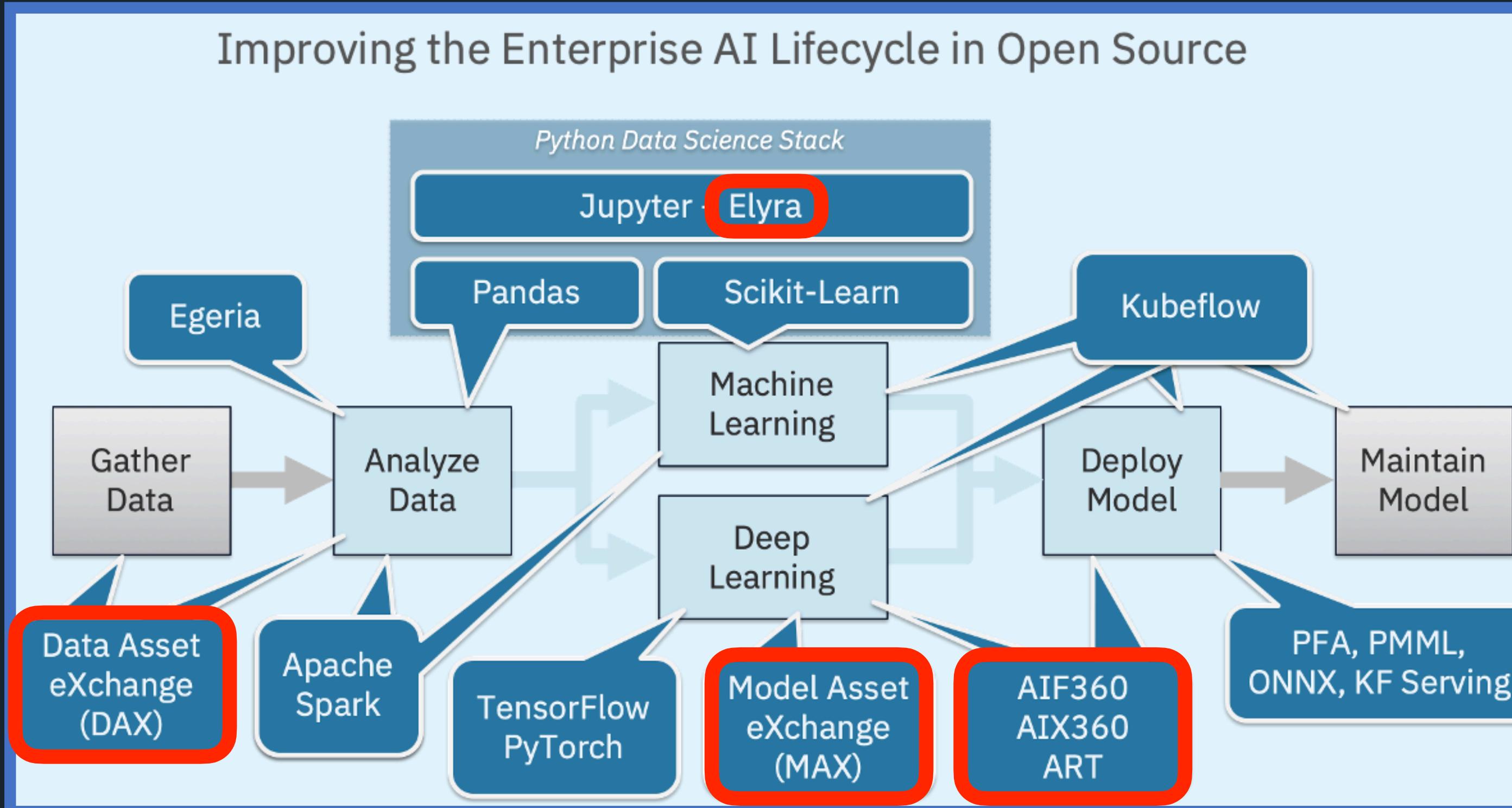


We contribute to and advocate for the open-source technologies



QUAIS SÃO OS PROJETOS QUE VOCÊ E SUA
EQUIPE ESTÃO ENVOLVIDOS NA IBM?

Center for Open Source Data and AI Technologies (CODAIT)



Alguns Projetos:

- Model Asset eXchange (ibm.biz/model-exchange)
- Data Asset eXchange (ibm.biz/model-exchange)
- Elyra (ibm.biz/elyra-ai)
- AI Fairness 360 (aif360.mybluemix.net/)
- AI Explainability 360 (aix360.mybluemix.net/)
- Adversarial Robustness Toolbox (art-demo.mybluemix.net/)
- FactSheets (aif360.mybluemix.net/)

Machine Learning Team

Model Asset eXchange

Model Asset eXchange

Free, deployable, and trainable code. A place for developers to find and use free and open source deep learning models.

Try the tutorial →
Join the community →

Featured Deployable Trainable

Model Deployable Toxic Comment Classifier Detect 6 types of toxicity in user comments Jun 04, 2019 →	Model Deployable, Trainable Text Sentiment Classifier Detect the sentiment captured in short pieces of text Mar 29, 2019 →	Model Deployable, Trainable Image Segmenter Identify objects in an image, additionally assigning each pixel of the image to a particular object. Sep 21, 2018 →
Model Deployable, Trainable Object Detector Localize and identify multiple objects in a single image. Sep 21, 2018 →	Model Deployable Audio Classifier Identify sounds in short audio clips. Sep 21, 2018 →	Model Deployable Image Caption Generator Generate captions that describe the contents of images. Sep 21, 2018 →

Data Asset eXchange

Learn More →

Data Asset eXchange

Explore useful and relevant data sets for enterprise data science

What's New →
Get Involved →

Dataset CSV NOAA Weather Data - JFK Airport June 30, 2020 →	Dataset IOB format Groningen Meaning Bank - Modified May 14, 2020 →	Dataset CSV Fashion-MNIST September 12, 2015 →
Dataset JPG, JSON PubLayNet October 25, 2019 →	Dataset WAV TensorFlow Speech Commands March 17, 2020 →	Dataset PNG, JSON PubTabNet July 20, 2020 →

ibm.biz/model-exchange

ibm.biz/data-exchange

Model Asset eXchange

Place for developers/data scientists to find and use
free and **open source** deep learning models

ibm.biz/model-exchange

Model Asset eXchange

[Try the tutorial](#)



[Join the community](#)



Free, deployable, and trainable code. A place for developers to find and use free and open source deep learning models.

[Featured](#) [Deployable](#) [Trainable](#)

Model | Deployable

Toxic Comment Classifier

Detect 6 types of toxicity in user comments

Jun 04, 2019

Model | Deployable, Trainable

Text Sentiment Classifier

Detect the sentiment captured in short pieces of text

Mar 29, 2019

Model | Deployable, Trainable

Image SegmenTer

Identify objects in an image, additionally assigning each pixel of the image to a particular object.

Sep 21, 2018

Model | Deployable, Trainable

Object Detector

Localize and identify multiple objects in a single image.

Sep 21, 2018

Model | Deployable

Audio Classifier

Identify sounds in short audio clips.

Sep 21, 2018

Model | Deployable

Image Caption Generator

Generate captions that describe the contents of images.

Sep 21, 2018

[View all models](#)

Model Asset eXchange

30+ ready to use deep learning models

- Wide variety of domains (text, audio, image, etc)
- Multiple deep learning frameworks (TensorFlow, PyTorch, Keras)
- Trainable and Deployable versions

ibm.biz/model-exchange

Model Asset eXchange

Free, deployable, and trainable code. A place for developers to find and use free and open source deep learning models.

Try the tutorial →

Join the community →

[Featured](#) [Deployable](#) [Trainable](#)

Model | Deployable

Toxic Comment Classifier

Detect 6 types of toxicity in user comments

Jun 04, 2019



Model | Deployable, Trainable

Text Sentiment Classifier

Detect the sentiment captured in short pieces of text

Mar 29, 2019



Model | Deployable, Trainable

Image SegmenTer

Identify objects in an image, additionally assigning each pixel of the image to a particular object.

Sep 21, 2018



Model | Deployable, Trainable

Object Detector

Localize and identify multiple objects in a single image.

Sep 21, 2018



Model | Deployable

Audio Classifier

Identify sounds in short audio clips.

Sep 21, 2018



Model | Deployable

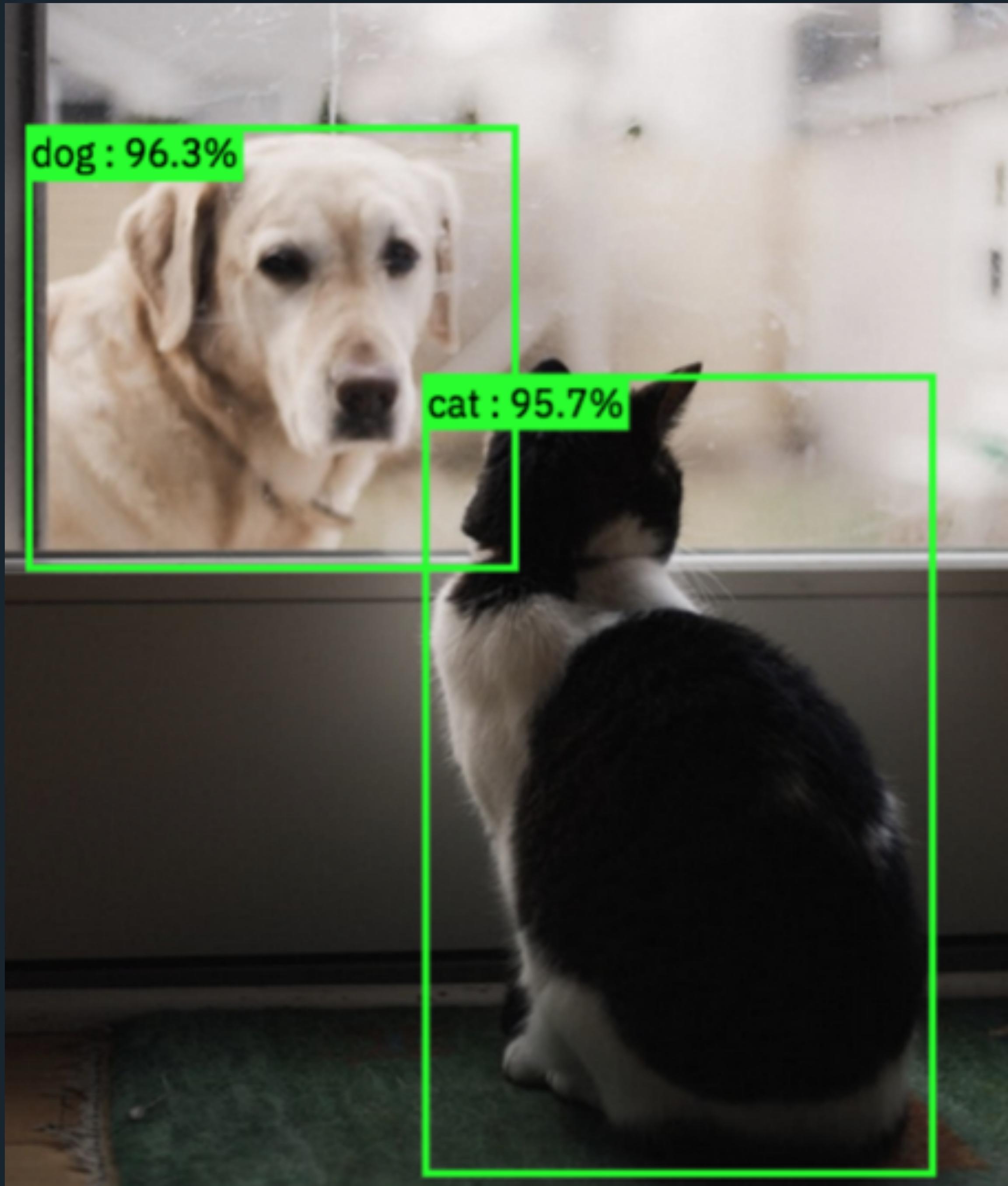
Image Caption Generator

Generate captions that describe the contents of images.

Sep 21, 2018



[View all models](#)



OBJECT DETECTOR

Localize and identify multiple objects in a single image



Signup for IBM Cloud: ibm.biz/gguniversity

Model Asset eXchange

Free, deployable, and trainable code. A place for developers to find and use free and open source deep learning models.

Try the tutorial →
Join the community →

Featured Deployable Trainable

Model | Deployable
Toxic Comment Classifier
Detect 6 types of toxicity in user comments
Jun 04, 2019 →

Model | Deployable, Trainable
Text Sentiment Classifier
Detect the sentiment captured in short pieces of text
Mar 29, 2019 →

Model | Deployable, Trainable
Image Segmente
Identify objects in an image, additionally assigning each pixel of the image to a particular object.
Mar 29, 2019 →

Model | Deployable, Trainable
Object Detector
Localize and identify multiple objects in a single image.
Sep 21, 2018 →

Model | Deployable
Audio Classifier
Identify sounds in short audio clips.
Sep 21, 2018 →

Model | Deployable
Image Caption Generator
Generate captions that describe the contents of images.
Sep 21, 2018 →

ibm.biz/model-exchange

Model Deployable, Trainable

Object Detector

Localize and identify multiple objects in a single image.

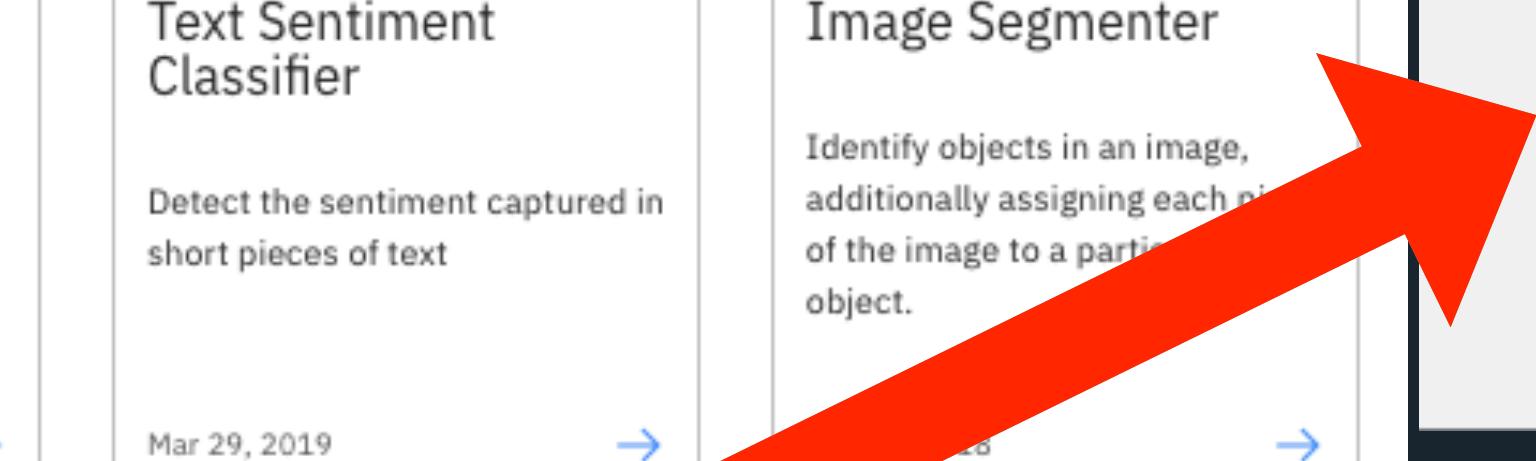
Get this model



Try the API →

Try the web app →

Try in a Node-RED flow →

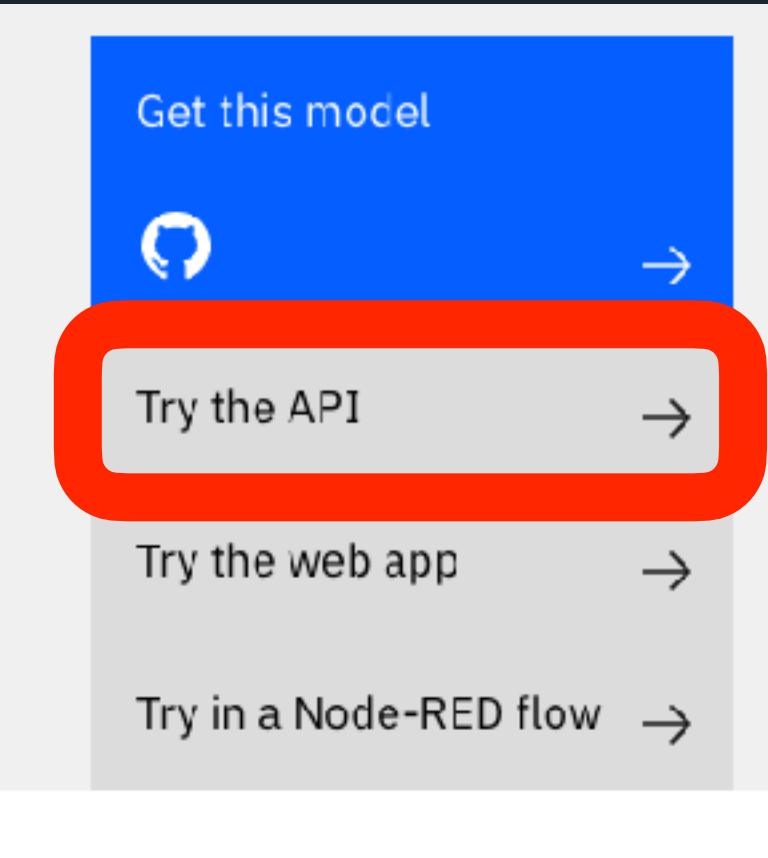


Access the API via Python

Model | Deployable, Trainable

Object Detector

Localize and identify multiple objects in a single image.



response

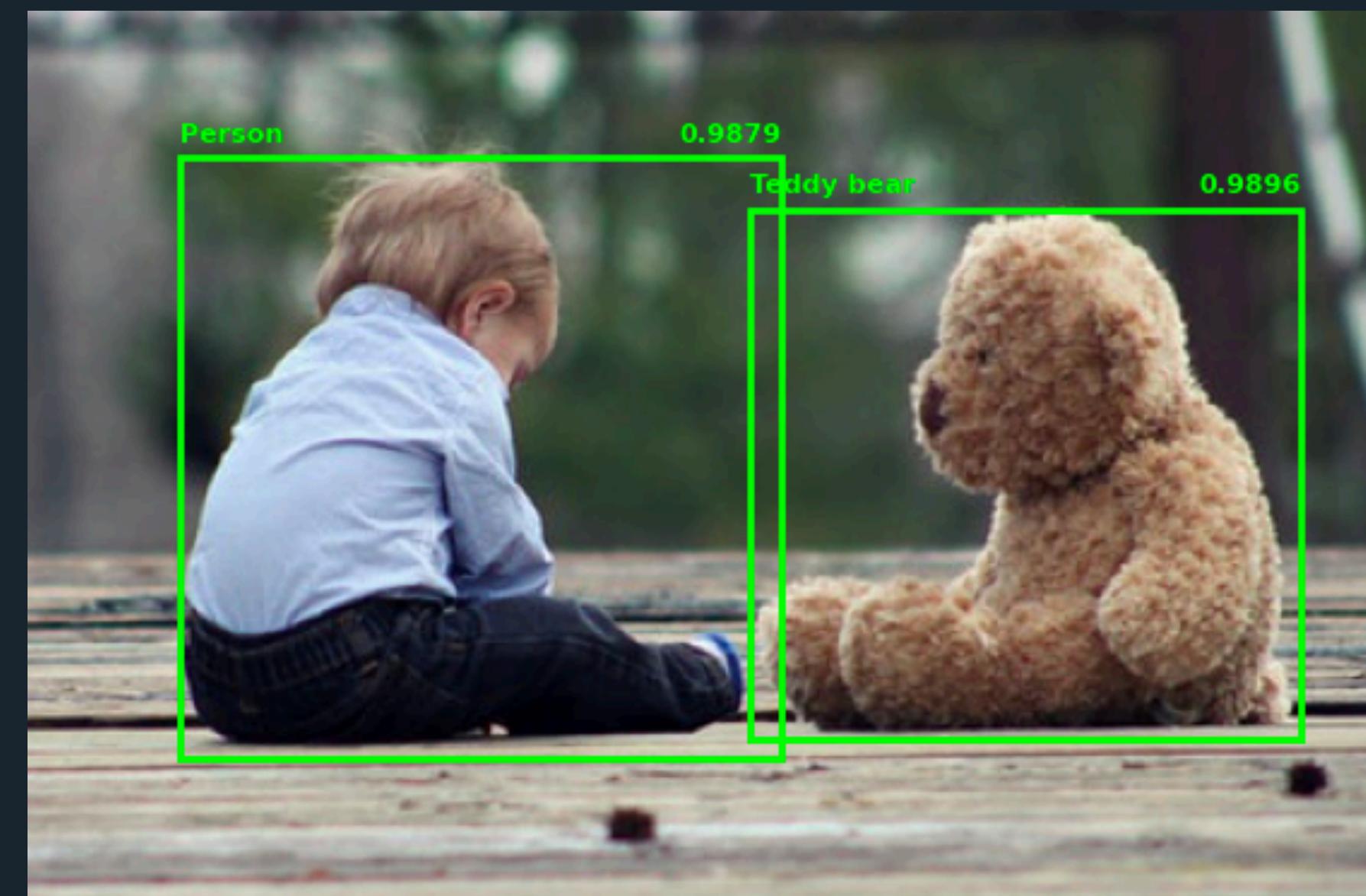
```
{'status': 'ok',
'predictions': [ {'label_id': '88',
'label': 'teddy bear',
'probability': 0.9896332025527954,
'detection_box': [0.27832502126693726,
0.5611844062805176,
0.643224835395813,
0.8432191610336304]}, {
'label_id': '1',
'label': 'person',
'probability': 0.9879012107849121,
'detection_box': [0.24251867830753326,
0.26926860213279724,
0.655893087387085,
0.5768759250640869]}]}
```

```
# Model
url = 'http://max-object-detector.codait-prod-41208c73af8fca213512856c7a09
db52-0000.us-east.containers.appdomain.cloud/'
model_endpoint = 'model/predict'
complete_url = url + model_endpoint

# Upload an image to the MAX model's rest API
path_to_input_image = 'baby-bear.jpg'

with open(path_to_input_image, 'rb') as file:
    file_form = {'image': (path_to_input_image, file, 'image/jpeg')}
    # Post the image to the rest API using the requests library
    r = requests.post(url=complete_url, files=file_form)
    # Return the JSON
    response = r.json()

IPython.display.Image(path_to_input_image, width = 450)
```



Access the API via R

```
library(dplyr)
library(httr)

# Endpoint
endpoint <- 'http://max-object-detector.codait-prod-41208c73af8fca213512856c7a09db52-0000.us-east.containers.appdomain.cloud/'
# endpoint <- 'http://localhost:5000' # if running docker locally or docker hub

object_detector <- function(path_to_img, endpoint) {
  model_endpoint <- paste0(endpoint, 'model/predict') # Model endpoint
  # POST
  response <- httr::POST(url = model_endpoint,
                           body = list(image = upload_file(path_to_img,
                                                             type = "image/jpeg")),
                           encode = c("multipart"))
  ) %>% content()
  response$predictions
}

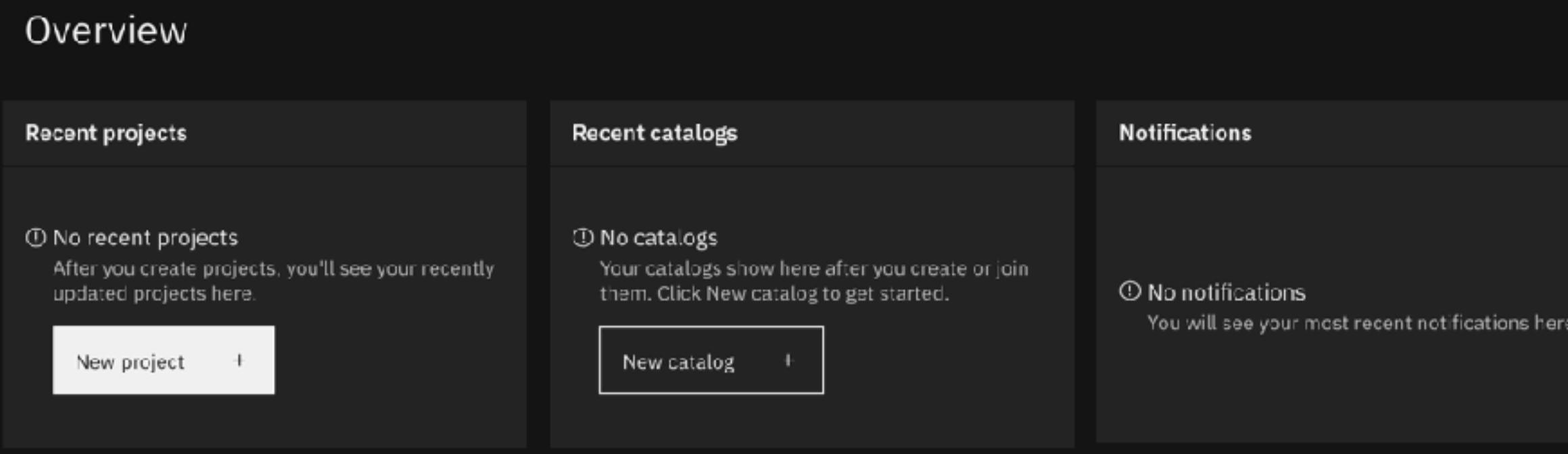
# Get the image file from GH
download.file(url = "http://github.com/IBM/MAX-Object-Detector/blob/master/samples/baby-bear.jpg?raw=true",
              'baby-bear.jpg', mode = 'wb')

object_detector("baby-bear.jpg", endpoint)
```

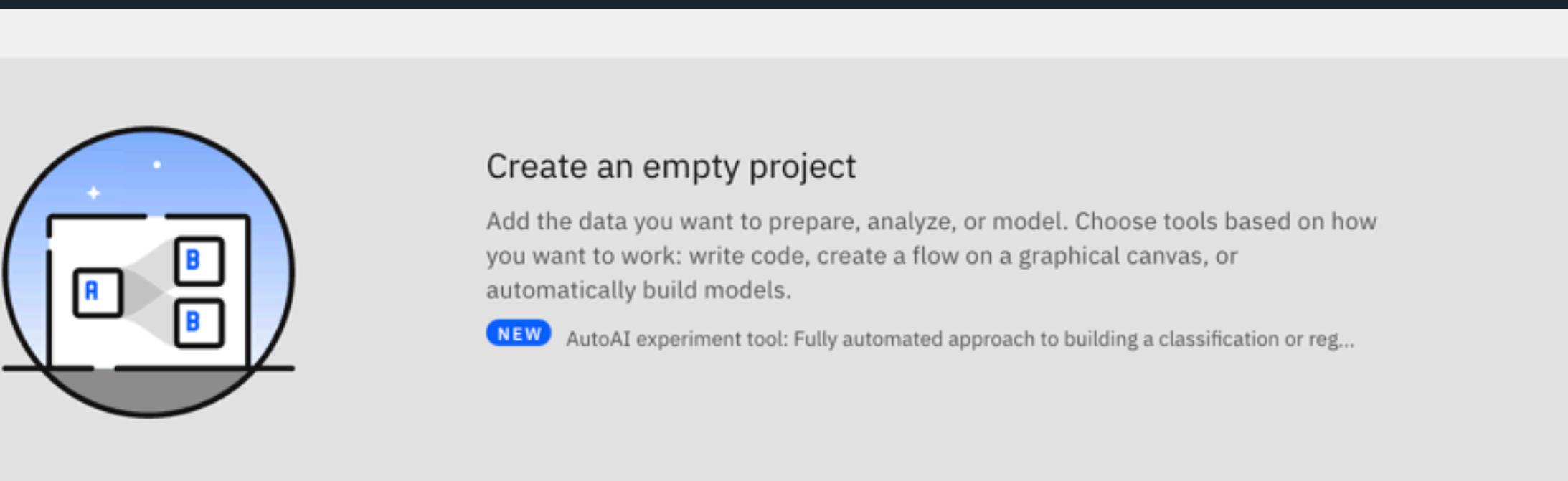
Let's try together ...

1) Go to ibm.biz/gg university and create an account

2) Login to your account and create a project



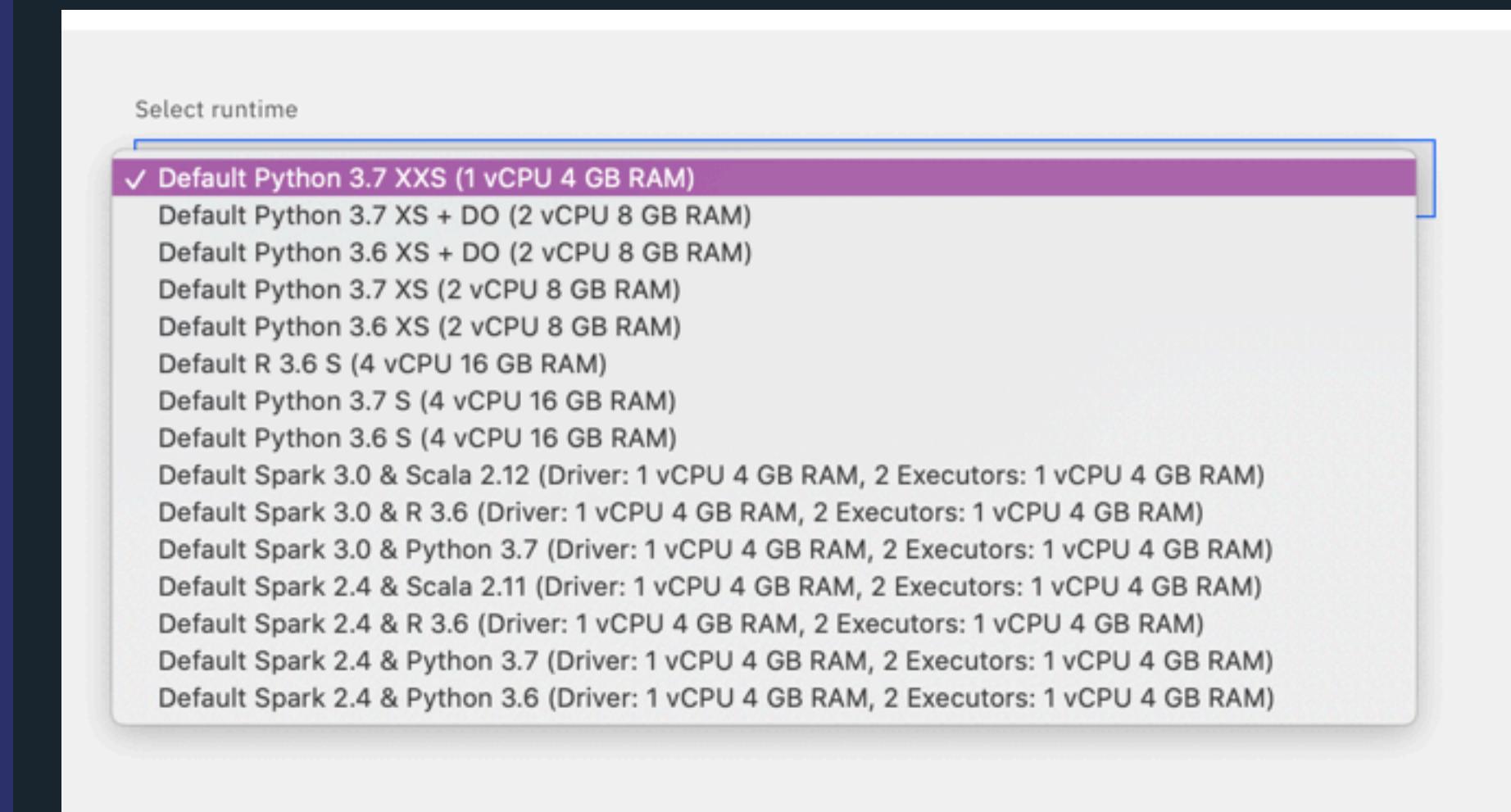
3) Create an *empty project*

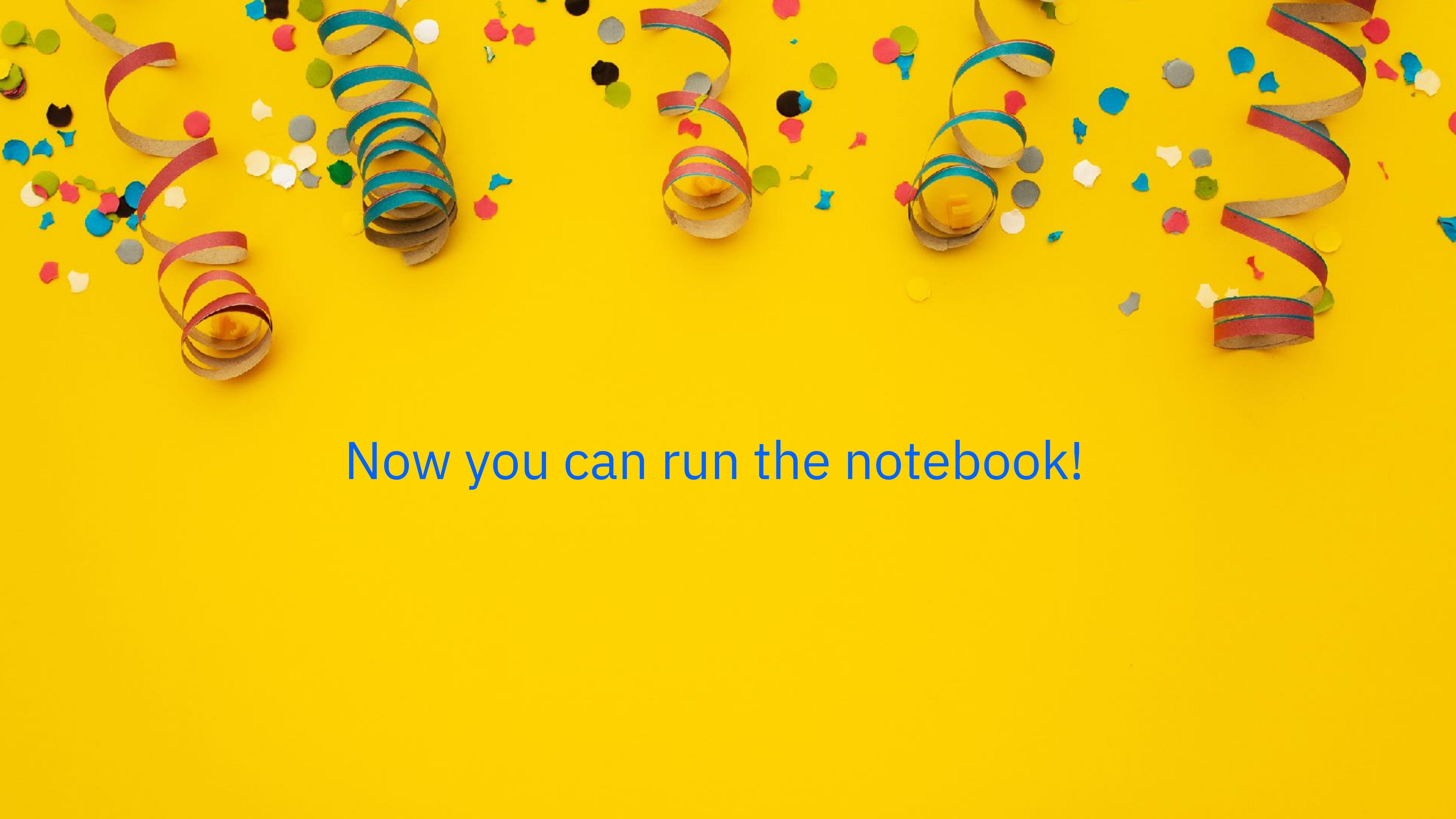


4) Go to the link ibm.biz/max-notebook and copy the project



5) Change the select runtime



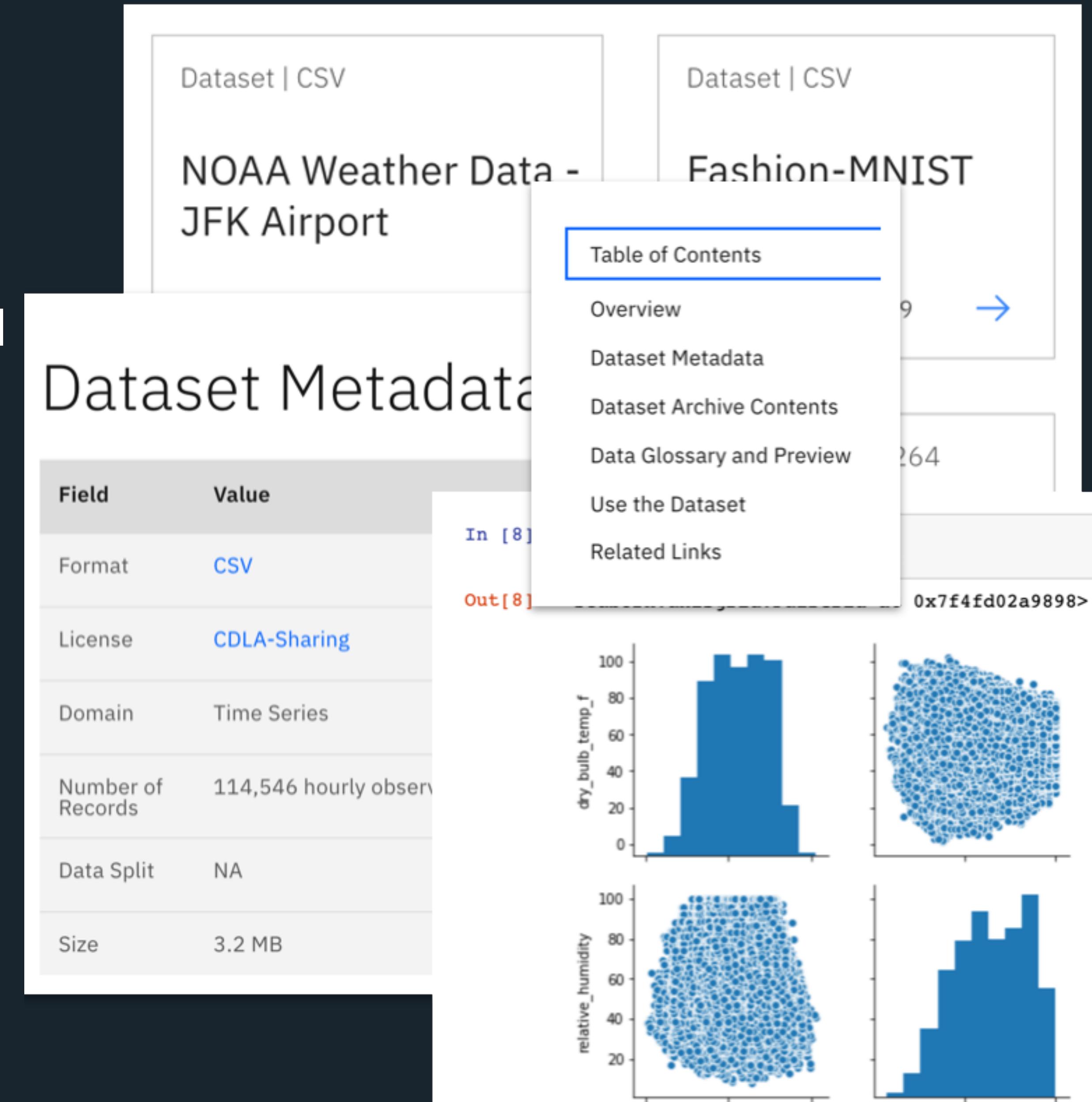


Now you can run the notebook!

Data Asset eXchange (DAX)

- Curated repository for **open** datasets from IBM Research and third-parties
- Published under data friendly licenses
- Standardized dataset formats and metadata
- Many data sets include starter notebooks (cleansing, data exploration, analysis)

ibm.biz/data-exchange



Data Asset eXchange (DAX)

MIT | CSV

Fashion-MNIST

A dataset of standardized images of fashion items from 10 classes

Get this dataset →
Try the notebook →



Technologies (1) ^

Artificial intelligence

Data Asset Technologies (2) ▾

Table of Contents ▾

Resources ▾

By Zalando

Updated September 12, 2019 | Published August 22, 2019

Overview

The Fashion-MNIST dataset contains 60,000 training images (and 10,000 test images) of fashion and clothing items, taken from 10 classes. Each image is a standardized 28×28 size in grayscale (784 total pixels). Fashion-MNIST was created by Zalando as a compatible replacement for the original MNIST dataset of handwritten digits. This version of the dataset has been converted to CSV to enable easier loading in common data science tools (see <https://www.kaggle.com/zalando-research/fashionmnist>).

Dataset Metadata

Fashion-MNIST

In this notebook we explore the Fashion-MNIST dataset from [IBM Data Asset Exchange](#). The Fashion-MNIST dataset contains 60,000 training images (and 10,000 test images) of fashion and clothing items, taken from 10 classes. Each image is a standardized 28×28 size in grayscale (784 total pixels). Fashion-MNIST was created by Zalando as a compatible replacement for the original MNIST dataset of handwritten digits. This version of the dataset has been converted to CSV to enable easier loading in common data science tools.

Reference: <https://www.kaggle.com/zalando-research/fashionmnist/home>

Table of Contents

1. Data Extraction
2. Data Exploration
3. Model development

```
In [1]: # Import required libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import tensorflow as tf
from sklearn import preprocessing
from sklearn.metrics import precision_score
from sklearn.metrics import confusion_matrix
from sklearn.metrics import recall_score
from sklearn.metrics import accuracy_score
```

ibm.biz/data-exchange

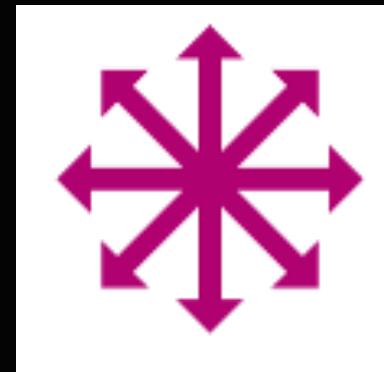
Trusted AI Lifecycle through Open Source

Pillars of trust, woven into the lifecycle of an AI application



IBM and LFAI move forward on
trustworthy and responsible AI
IBM donates Trusted AI toolkits to the Linux Foundation AI

Did anyone tamper
with it?



ROBUSTNESS

Adversarial Robustness 360
↳ (ART)

DEMO: art-demo.mybluemix.net

Is it fair?



FAIRNESS

AI Fairness 360
↳ (AIF360)

DEMO: aif360.mybluemix.net

Is it easy to understand?

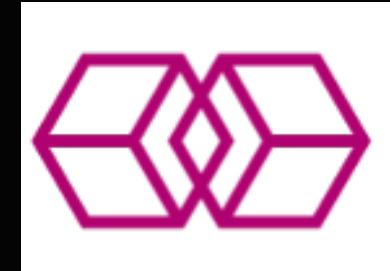


EXPLAINABILITY

AI Explainability 360
↳ (AIX360)

DEMO: aix360.mybluemix.net

Is it accountable?



LINEAGE

AI FactSheets 360

DEMO: aifs360.mybluemix.net

Trusted-AI

This GitHub org hosts LF AI Foundation projects in the category of Trusted and Responsible AI.

IBM @LFAI_Foundation info@lfaifoundation.org

Repositories 4 Packages People Projects

Pinned repositories

adversarial-robustness-toolbox
Adversarial Robustness Toolbox (ART) - Python Library for Machine Learning Security - Evasion, Poisoning, Extraction, Inference
Python 1.7k 480

AIF360
A comprehensive set of fairness metrics for datasets and machine learning models, explanations for these metrics, and algorithms to mitigate bias in datasets and models.
Python 1k 340

AIX360
Interpretability and explainability of data and machine learning models
Python 621 136

AI Fairness 360 (AIF360) R Package
CRAN 0.1.0 CRAN 0.1.0

Available in R too!

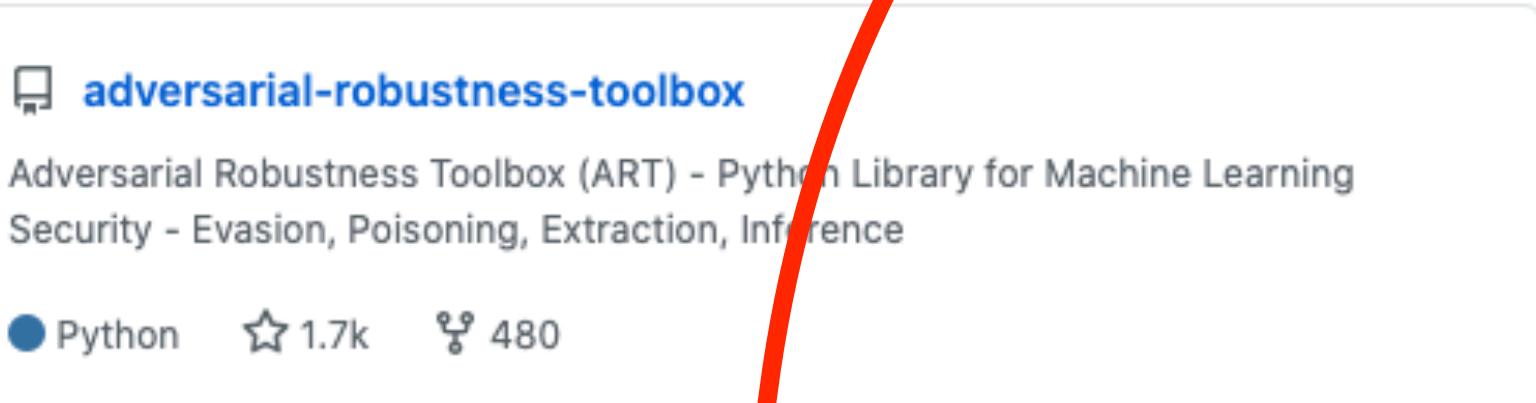
 Trusted-AI

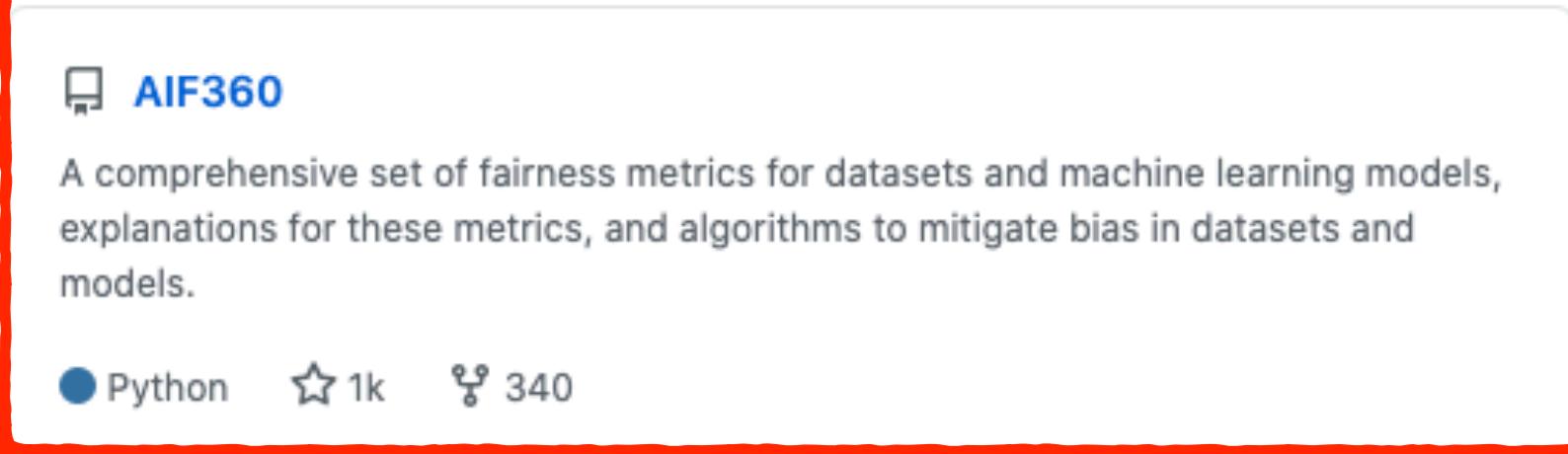
This GitHub org hosts LF AI Foundation projects in the category of Trusted and Responsible AI.

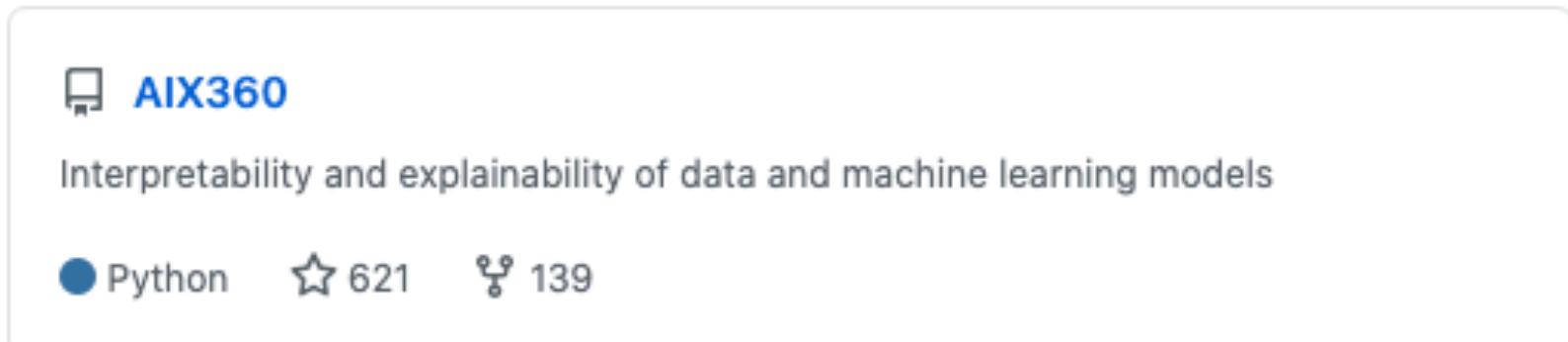
📍 IBM
🐦 @LFAI_Foundation
✉️ info@lfaifoundation.org

Repositories 4 Packages People Projects

Pinned repositories


adversarial-robustness-toolbox
Adversarial Robustness Toolbox (ART) - Python Library for Machine Learning Security - Evasion, Poisoning, Extraction, Inference
Python 1.7k 480


AIF360
A comprehensive set of fairness metrics for datasets and machine learning models, explanations for these metrics, and algorithms to mitigate bias in datasets and models.
Python 1k 340


AIX360
Interpretability and explainability of data and machine learning models
Python 621 139

Available in R too!

AI Fairness 360 (AIF360) R Package

CRAN 0.1.0
GitHub 0.1.0

Unwatch 74 Unstar 1.1k Fork 369

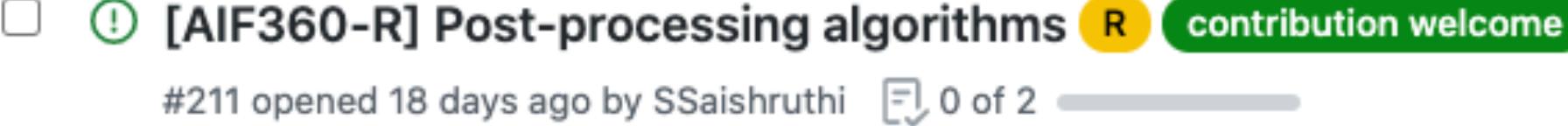
Code Issues 28 Pull requests 9 Actions Projects Wiki ...

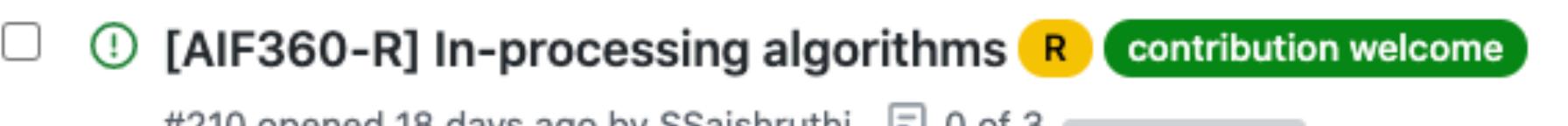
Filters is:open is:issue label:R Labels 11 Milestones 1 New issue

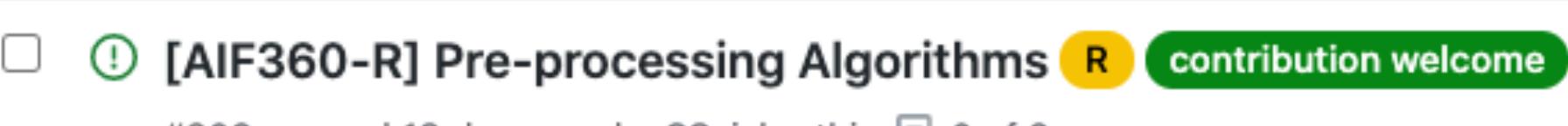
Clear current search query, filters, and sorts

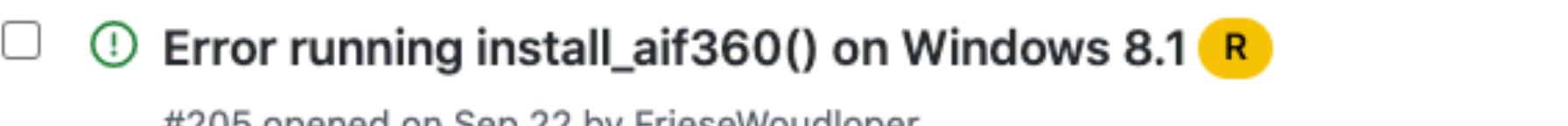
7 Open 0 Closed

Author	Label	Projects	Milestones	Assignee	Sort

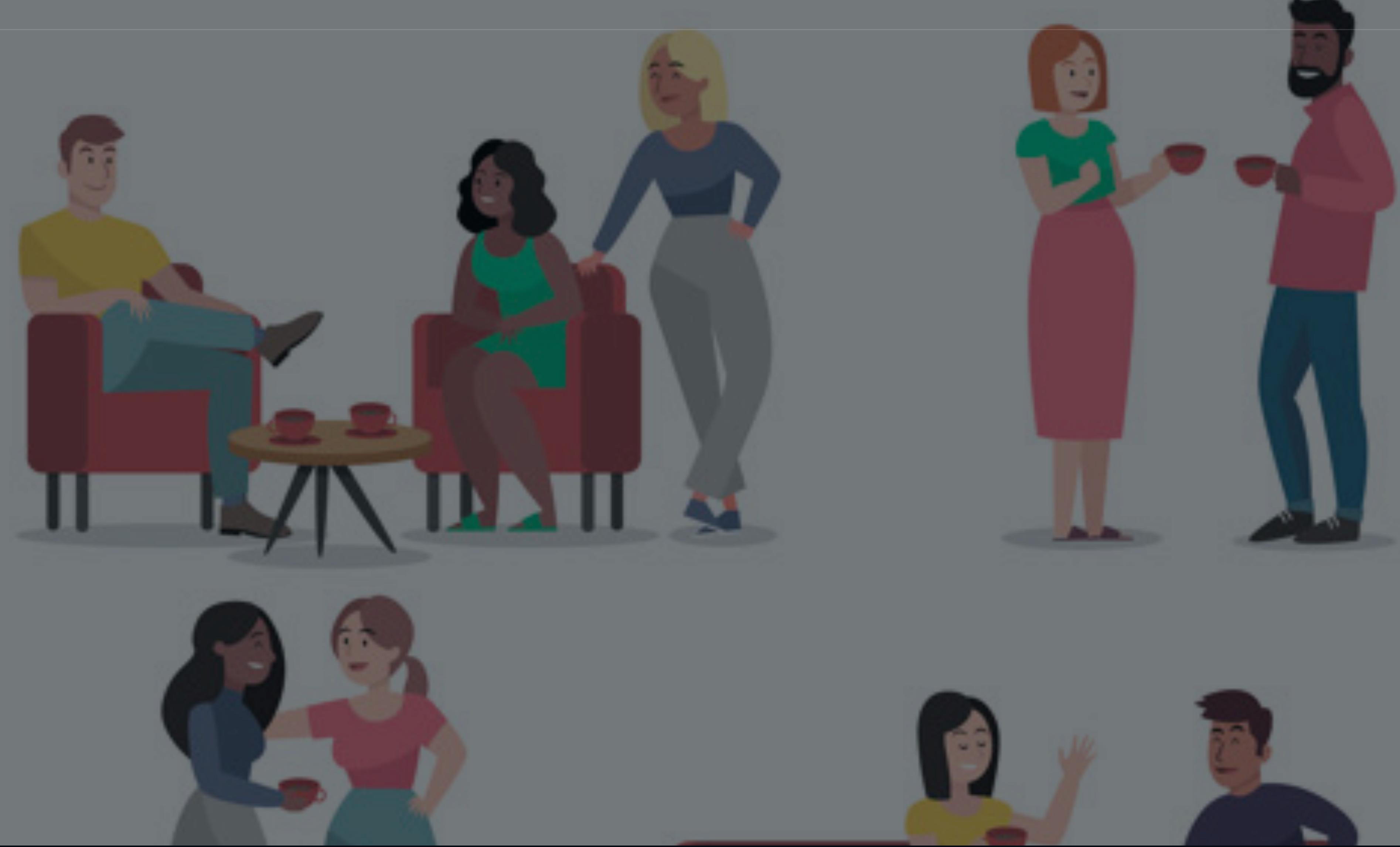

[AIF360-R] Post-processing algorithms R contribution welcome
#211 opened 18 days ago by SSaishruthi 0 of 2


[AIF360-R] In-processing algorithms R contribution welcome
#210 opened 18 days ago by SSaishruthi 0 of 3


[AIF360-R] Pre-processing Algorithms R contribution welcome
#209 opened 18 days ago by SSaishruthi 0 of 2


Error running install_aif360() on Windows 8.1 R
#205 opened on Sep 22 by Friesewoudloper


Issues in installing aif360 library on Rstudio - windows10 R
#200 opened on Sep 2 by Aloumora



QUANDO VOCÊ NÃO ESTÁ GERENCIANDO O
TIME, O QUE VOCÊ ESTÁ FAZENDO?

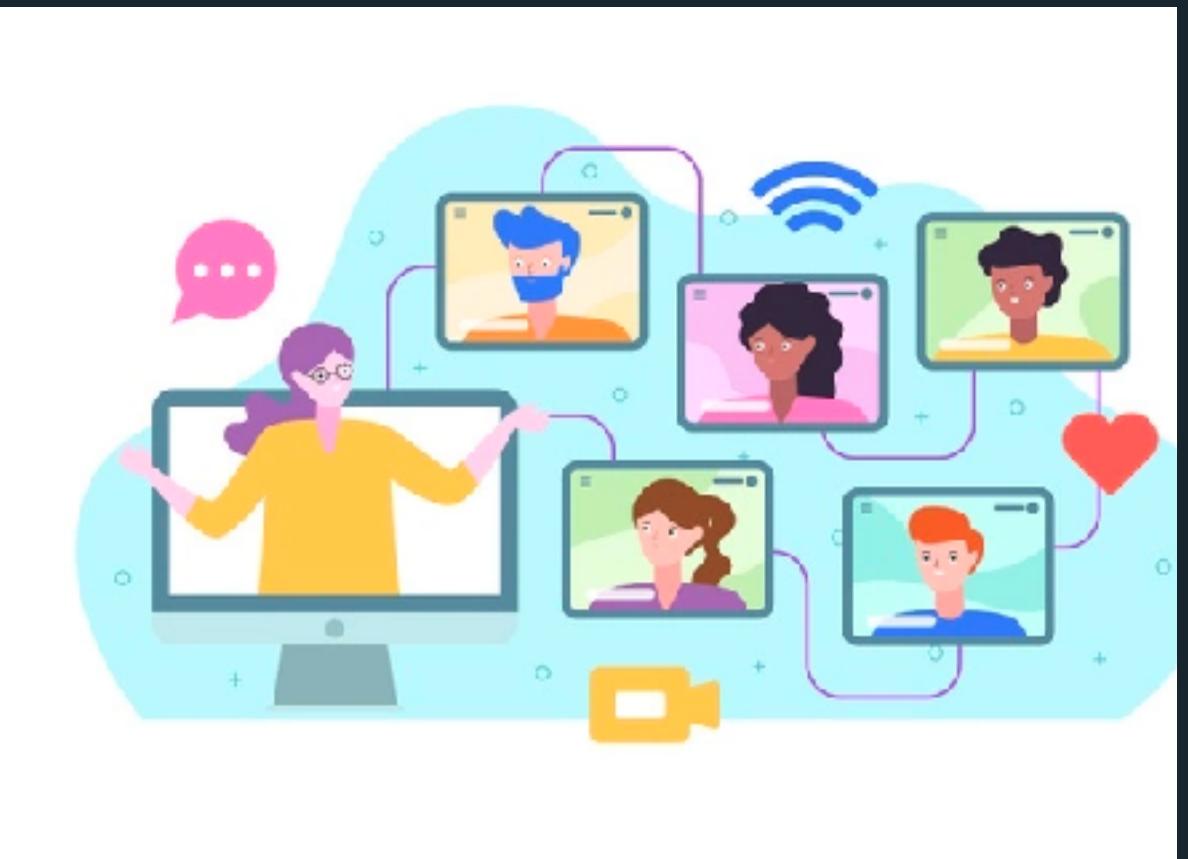
Família



Mentoria



Palestras



Cuidando de mim



AI Inclusive

Missão : Aumentar a **representatividade e participação** de minorias em Inteligência Artificial



- Website: ai-inclusive.org
- Twitter: bit.ly/ai-inclusive-twitter
- Facebook: bit.ly/ai-inclusive-facebook
- Instagram: bit.ly/ai-inclusive-instagram
- Youtube: bit.ly/ai-inclusive-youtube



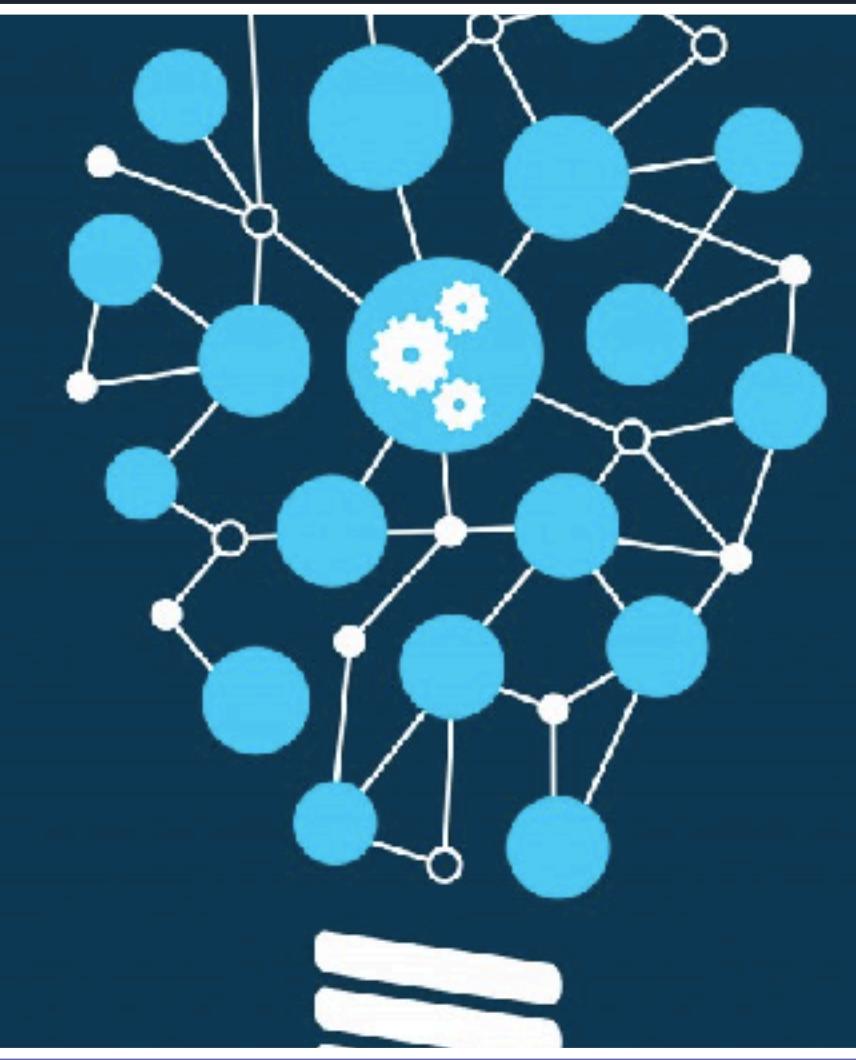
Se tiver interesse em criar um capítulo, mande-nos um email:
info@ai-inclusive.org

IDEAS



O QUE VOCÊ ACHA IMPORTANTE UM
CIENTISTA DE DADOS SABER?

MACHINE LEARNING



MACHINE LEARNING

ESTATÍSTICA

“Statistics is a **science**,
not a branch of mathematics,
but uses mathematical models
as essential tools.”

—John Tukey

A dark-themed screenshot of a computer screen showing a large amount of white and red code text, likely PHP, displayed in a terminal or code editor window.

PROGRAMAÇÃO



Communication



Critical Thinking



Curiosity
(keep asking why)



Ethics



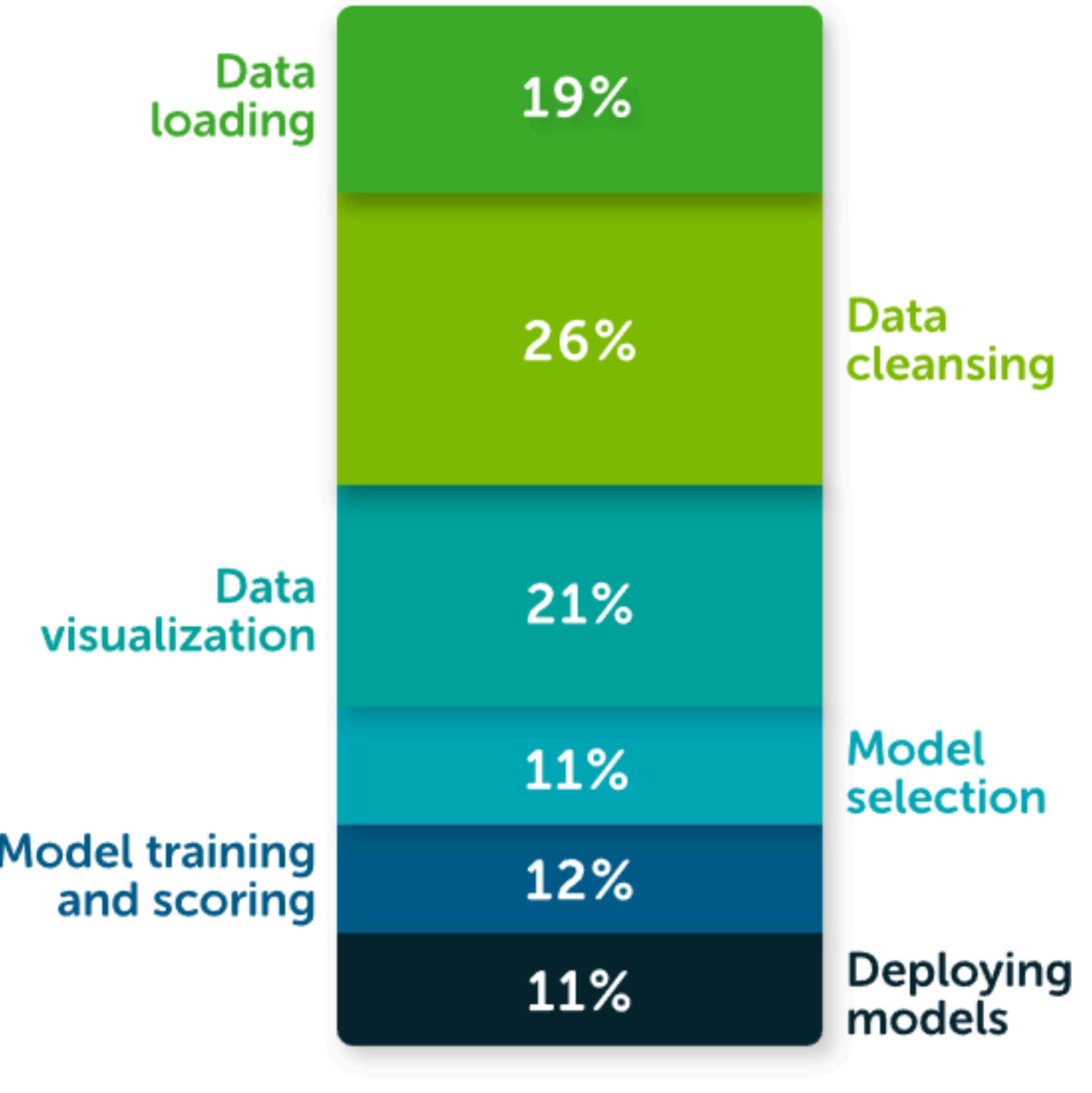
Flexibility



Be yourself

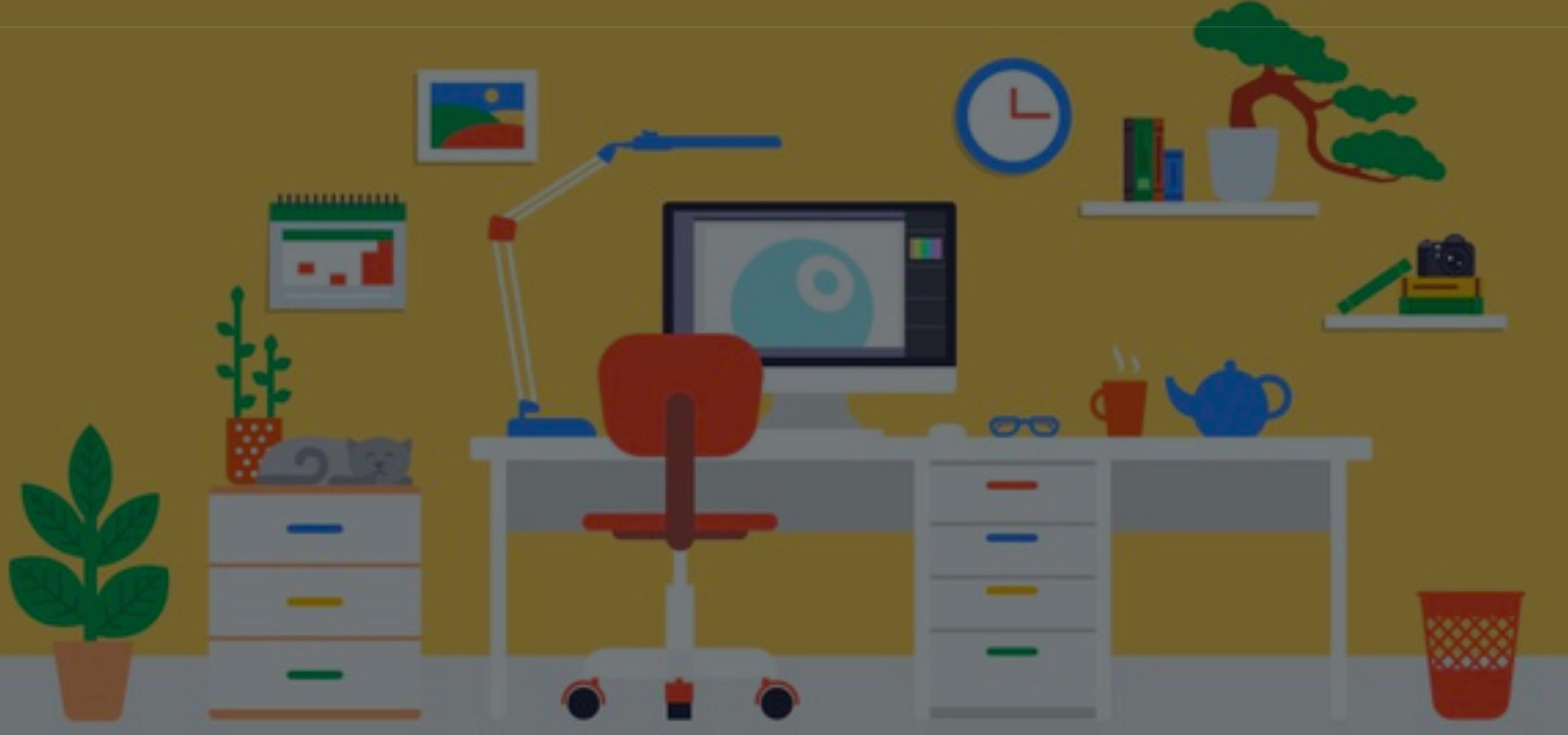


VOCÊ ACHA QUE UM CIENTISTA DE DADOS
DEVE FOCAR EM INTELIGÊNCIA ARTIFICIAL?



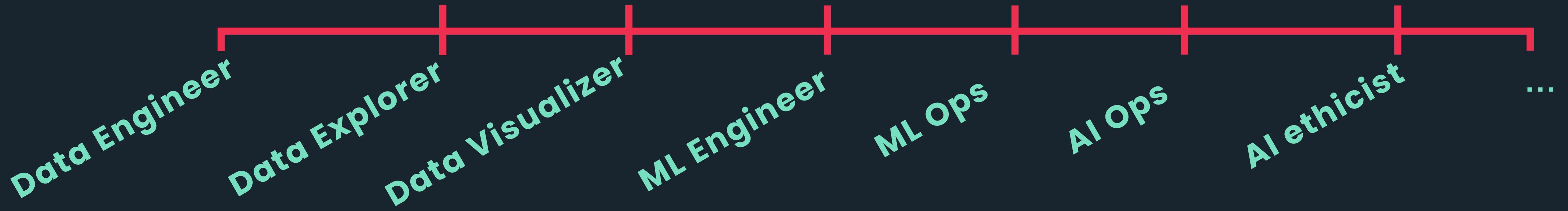
2020 State of Data Science report

For most respondents, **data management** tasks still consume a disproportionate amount of work time.



COMO VOCÊ VÊ A CARREIRA DE CIENTISTA DE
DADOS EVOLUINDO?

Specialized new roles within Data Science



Obrigada!

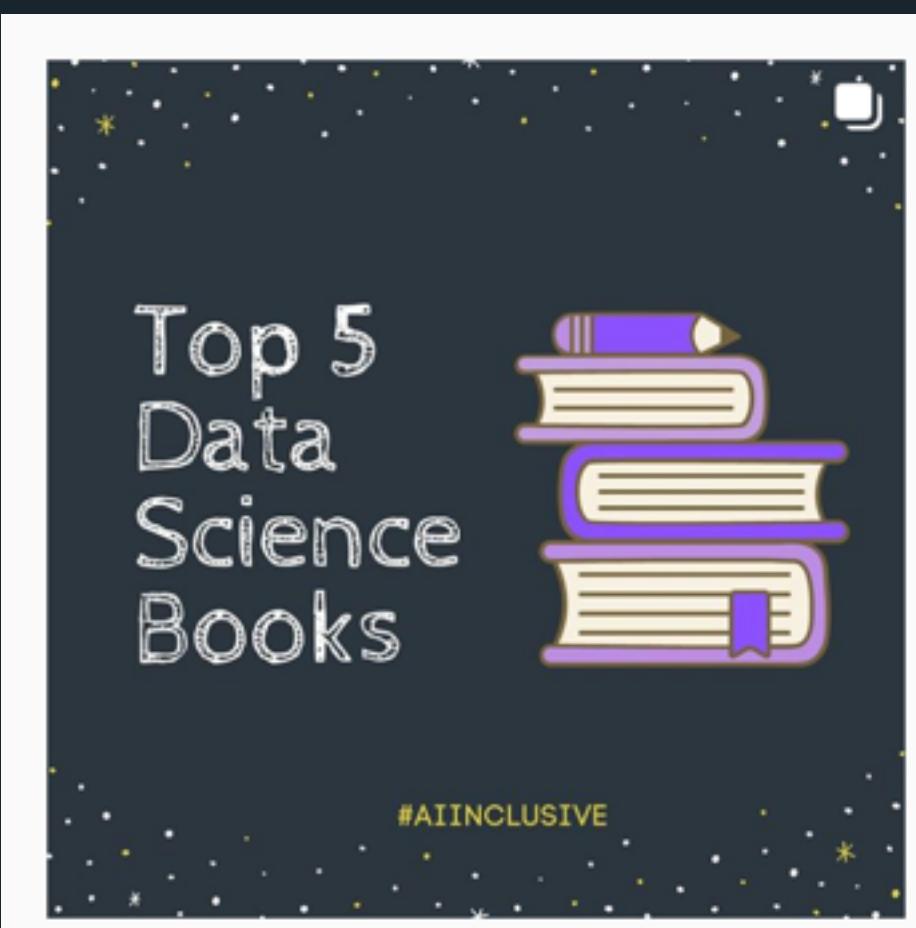
slides: bit.ly/pet-ufpr



ai-inclusive.org

Siga a gente:

bit.ly/ai-inclusive-instagram



Materiais de AI, DS, ML
Eventos, Ingressos de graça e muito mais

