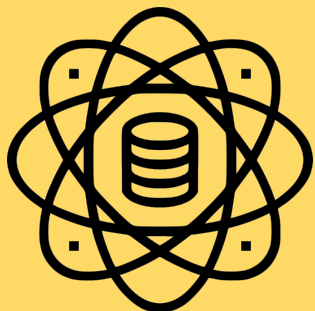


Introdução à Análise de Dados

Equipe DataAt



22 de Julho de 2022
São José dos Campos - SP

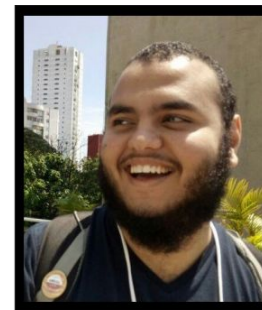
Adriano Almeida
Felipe Carvalho
Felipe Menino



ADRIANO ALMEIDA

🐙 @AdrianoPereira

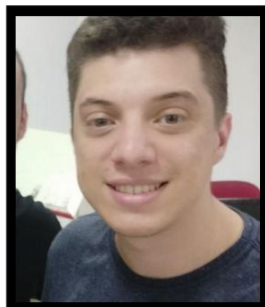
✉️ adrianoalmeida.py@gmail.com



FELIPE MENINO

🐙 @M3nin0

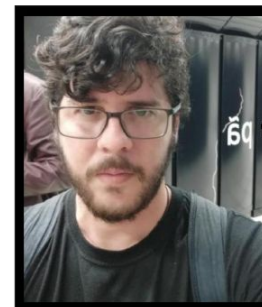
✉️ efelipecarlos@gmail.com



FELIPE CARVALHO

🐙 @oldlipe

✉️ lipecaso@gmail.com



HELVECIO NETO

🐙 @helvecioneto

✉️ helvecio.neto@live.com

Nossos minicursos - DataAt

1. Introdução à análise de dados
2. Introdução à análise de dados espaciais
3. Introdução ao docker
4. Introdução ao aprendizado de máquina

Link: github.com/dataAt

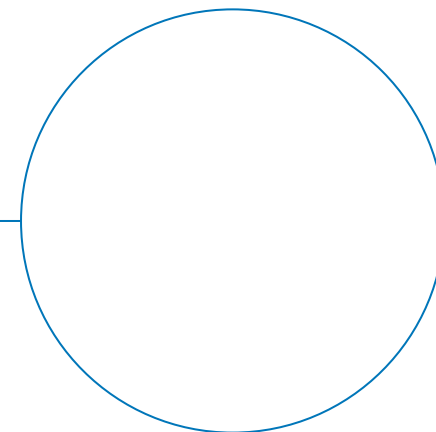


Material

Disponível no Github:

- Apresentação: [Introdução ao R](#)
- Códigos: github.com/dataAt/introducao-analise-de-dados
- Livro-Texto: dataat.github.io/introducao-analise-de-dados

INTRODUÇÃO À ANÁLISE DE DADOS



Objetivo

Objetiva-se apresentar ao aluno o processo básico de análise de dados, utilizando as linguagens de programação R e Python, de acordo com as tecnologias mais usadas no mercado e na academia.

O que são dados?



Dado

É tudo aquilo que é registrado, para posteriormente, ser analisado



Formatos

Cada formato pode apresentar etapas diferentes de organizações



Data Frame

Tabelas utilizadas para armazenar, manipular e representar dados



Peso

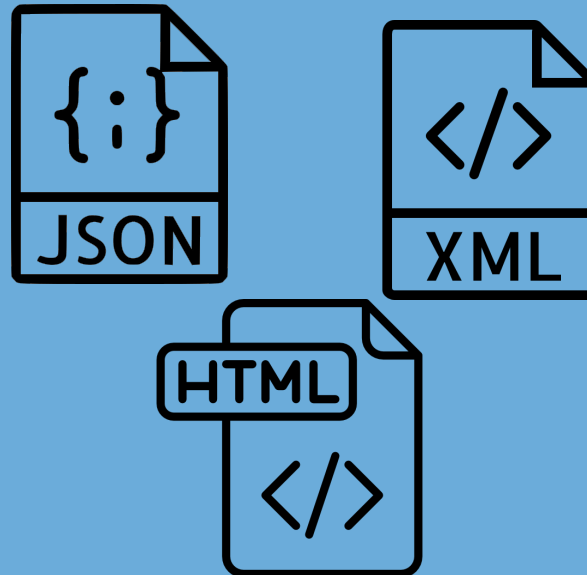
Entenda que todo dado tem um peso, contexto e impacto ao ser analisado

Formato dos dados

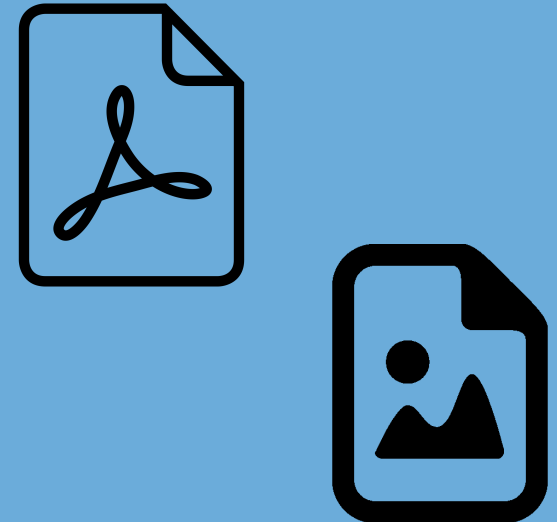
Dados Estruturados



Dados Semi-estruturados



Dados não Estruturados



Análise de dados

→ Conjunto de técnicas usadas para extrair informações de dados:

- Visualização de dados
- Análise exploratória

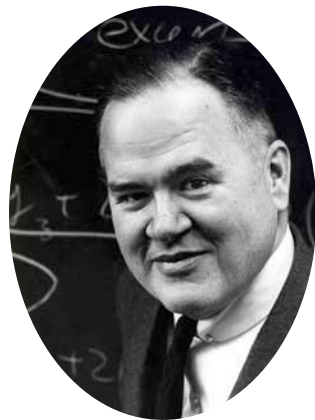


Análise de dados

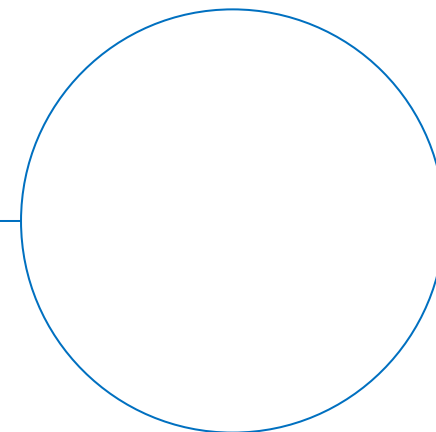
Para John W. Tukey, grande parte da análise de dados é inferencial, ou seja, o ato de extrair informações de uma amostra em relação ao conjunto todo.

→ Não é um conceito novo!

- Tukey fez essa definição em 1961! No artigo *The Future of Data Analysis*

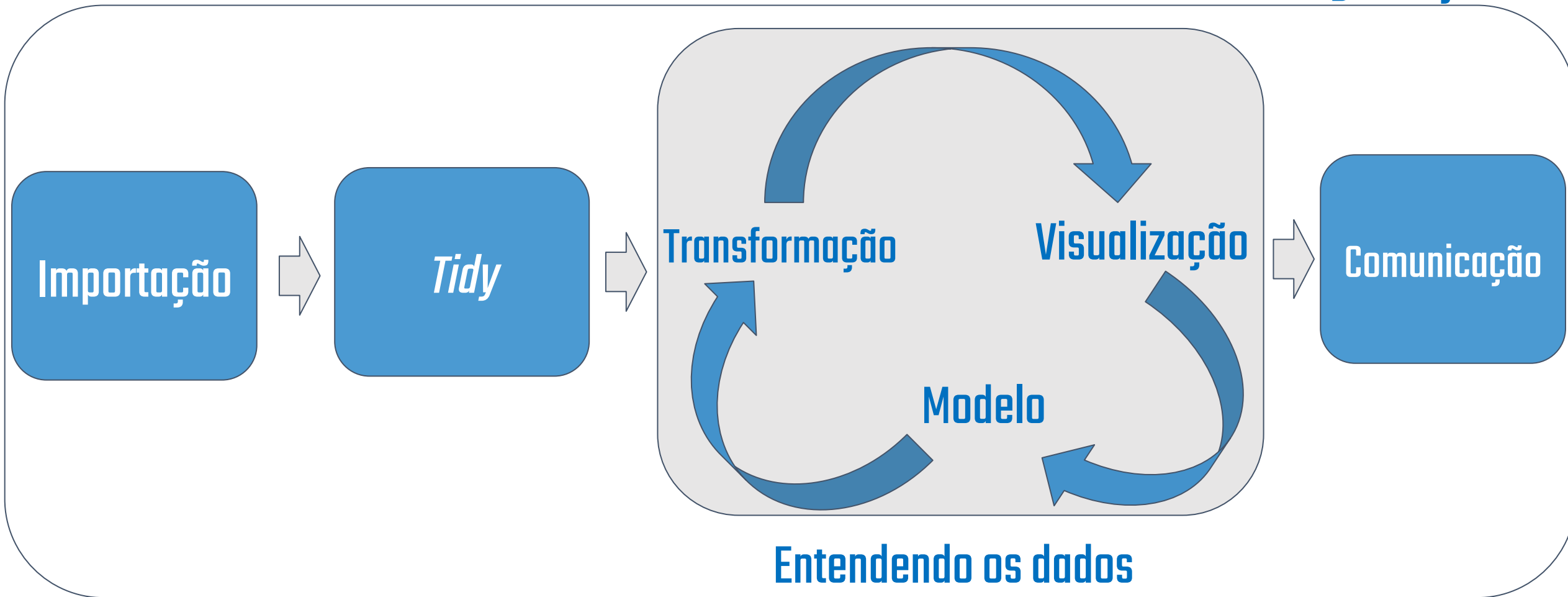


PROCESSO DE ANÁLISE DE DADOS



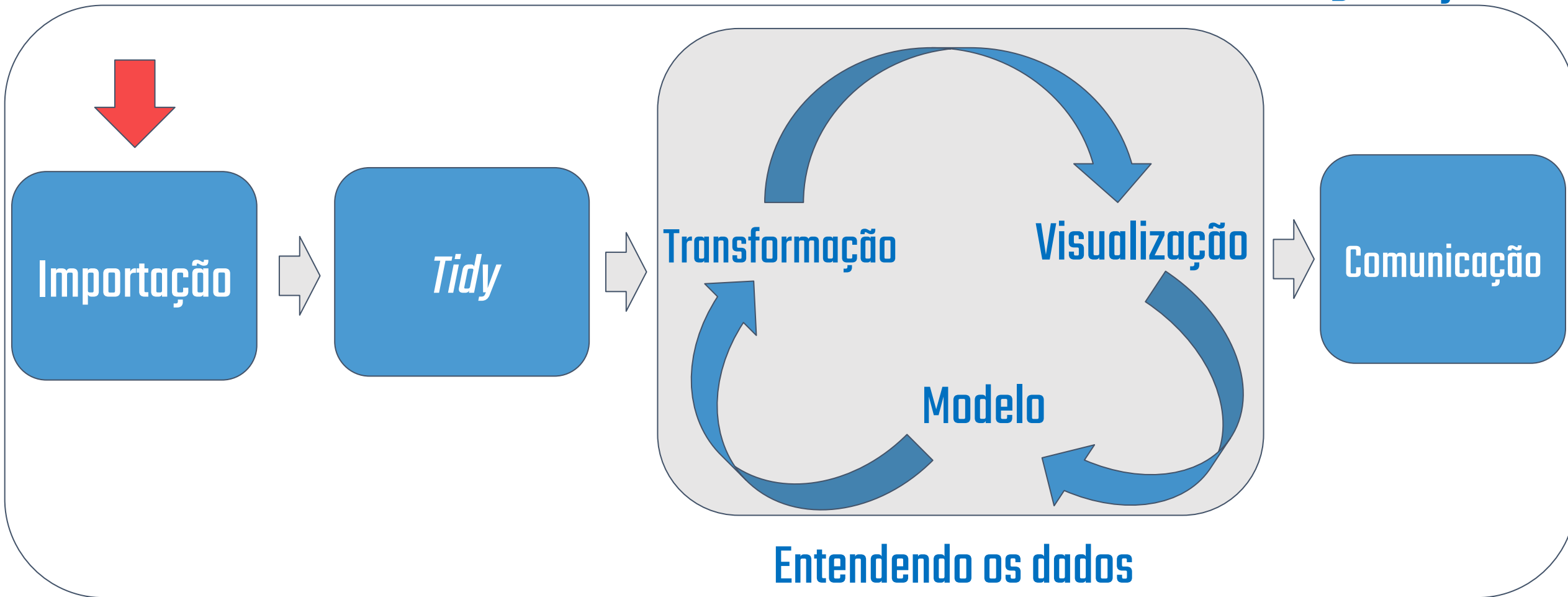
Processo de Análise de dados

Programação



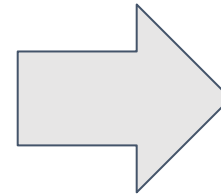
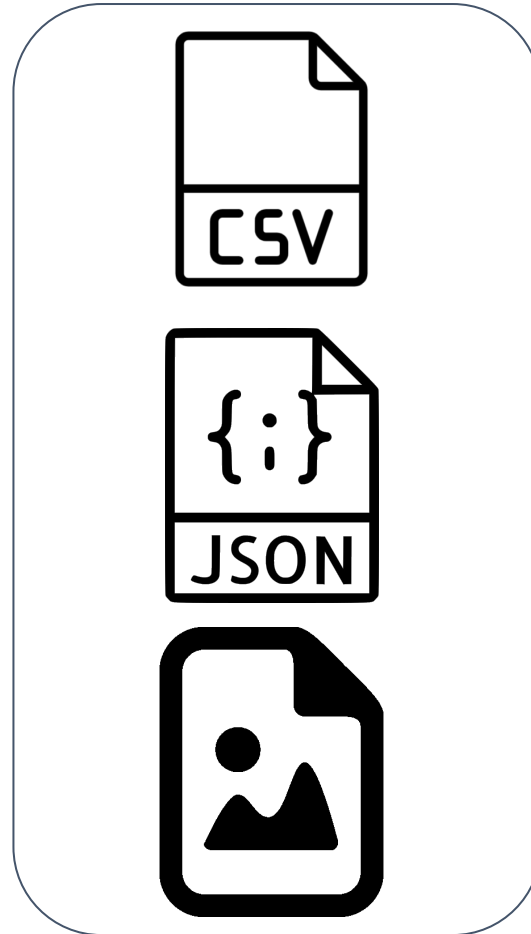
Processo de Análise de dados

Programação

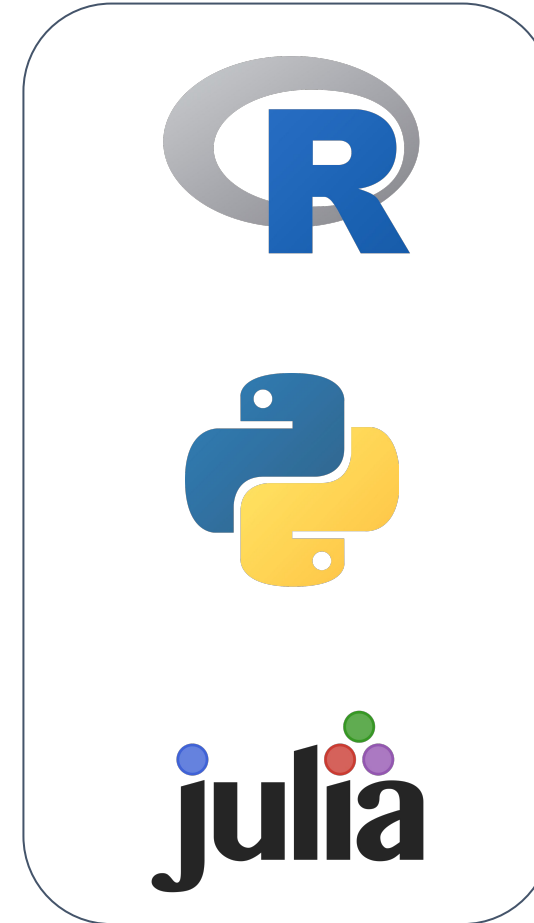


Importação

Fonte de dados



Linguagem de
programação



Processo de Análise de dados

Programação

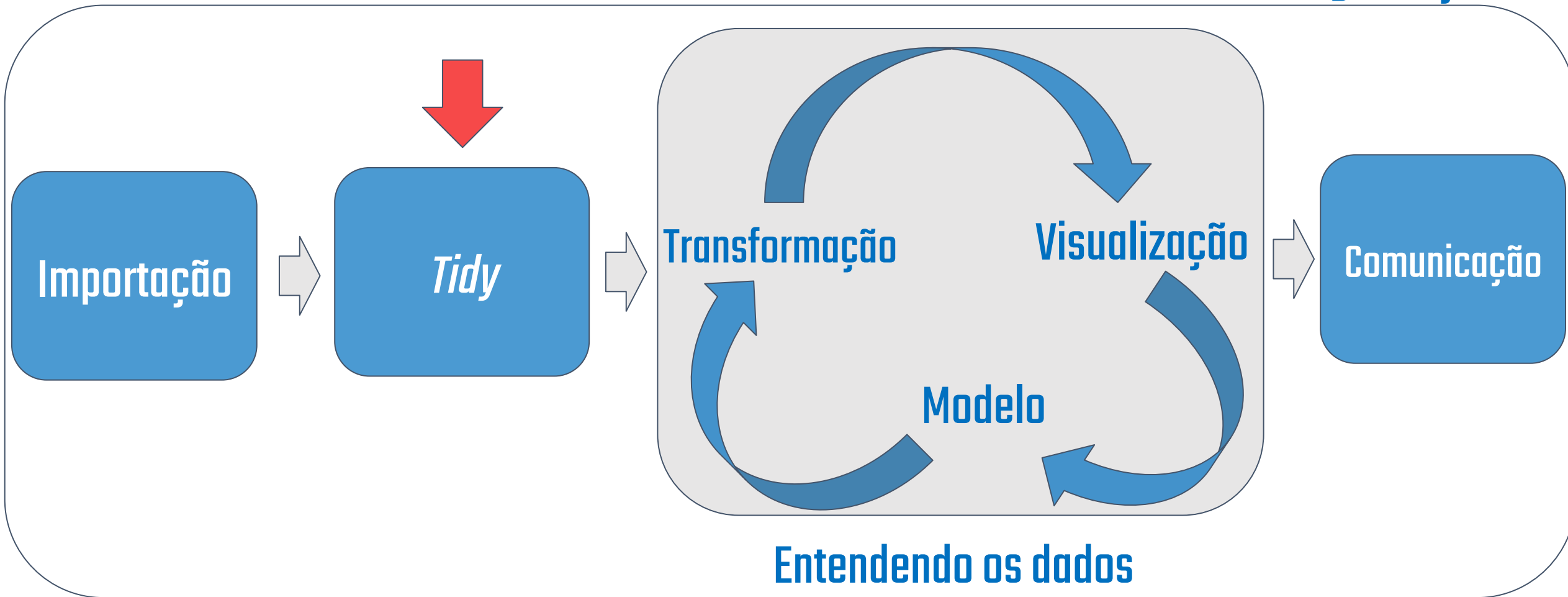


Tabela de alunos - *Wide*

Nome	Matemática	História
João	8	7
Maria	9	10
José	5	7
Helena	8	6

Tabela de alunos - *Long*

Nome	Disciplinas	Nota
João	Matemática	8
Maria	Matemática	9
José	Matemática	5
Helena	Matemática	8
João	História	7
Maria	História	10
José	História	9
Helena	História	6

Atributos/Colunas

Nome	Disciplinas	Nota
João	Matemática	
Maria	Matemática	
João	Matemática	
Helena	Matemática	
João	História	
Maria	História	1
João	História	
Helena	História	

Observação

Nome	Disciplinas	Nota
João	Matemática	8
Maria	Matemática	9
João	Matemática	5
Maria	Matemática	8
João	História	7
Maria	História	10
João	História	9
Maria	História	8

Valores

Nome	Disciplinas	Nota
João	Matemática	8
Maria	Matemática	9
José	Matemática	5
Helena	Matemática	8
João	História	7
Maria	História	10
José	História	9
Helena	História	6

Exemplo - MasterChef versão kids



baker	cinnamon_1	cardamom_2	nutmeg_3
Emma	1	0	1
Harry	1	1	1
Ruby	1	0	1
Zainab	0	NA	0

trial	Emma	Harry	Ruby	Zainab
cinnamon_1	1	1	1	0
cardamom_2	0	1	0	NA
nutmeg_3	1	1	1	0

Exemplo de Tidyficação

tame data

baker	cinnamon_1	cardamom_2	nutmeg_3
Emma	1	0	1
Harry	1	1	1
Ruby	1	0	1
Zainab	0	NA	0

trial	Emma	Harry	Ruby	Zainab
cinnamon_1	1	1	1	0
cardamom_2	0	1	0	NA
nutmeg_3	1	1	1	0

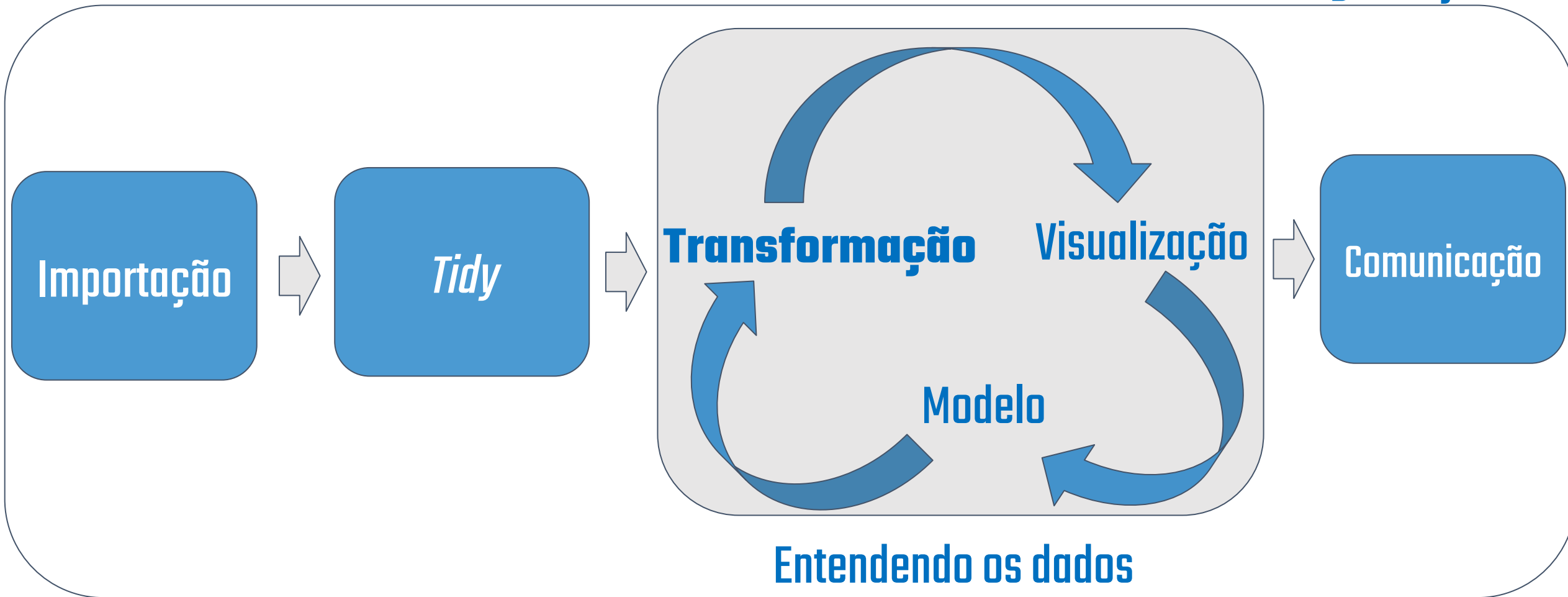
tidy data

baker	spice	order	correct
Emma	Cinnamon	1	1
Harry	Cinnamon	1	1
Ruby	Cinnamon	1	1
Zainab	Cinnamon	1	0
Emma	Cardamom	2	0
Harry	Cardamom	2	1
Ruby	Cardamom	2	0
Zainab	Cardamom	2	NA
Emma	Nutmeg	3	1
Harry	Nutmeg	3	1
Ruby	Nutmeg	3	1
Zainab	Nutmeg	3	0

Fonte: DataCamp

Processo de Análise de dados

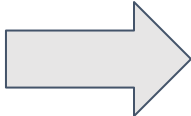
Programação



Transformação

Antes da transformação

name	idade	Alt
João	23 anos	1.7
Maria	20y.	1,68
José	19	167
Helena	18	1.72

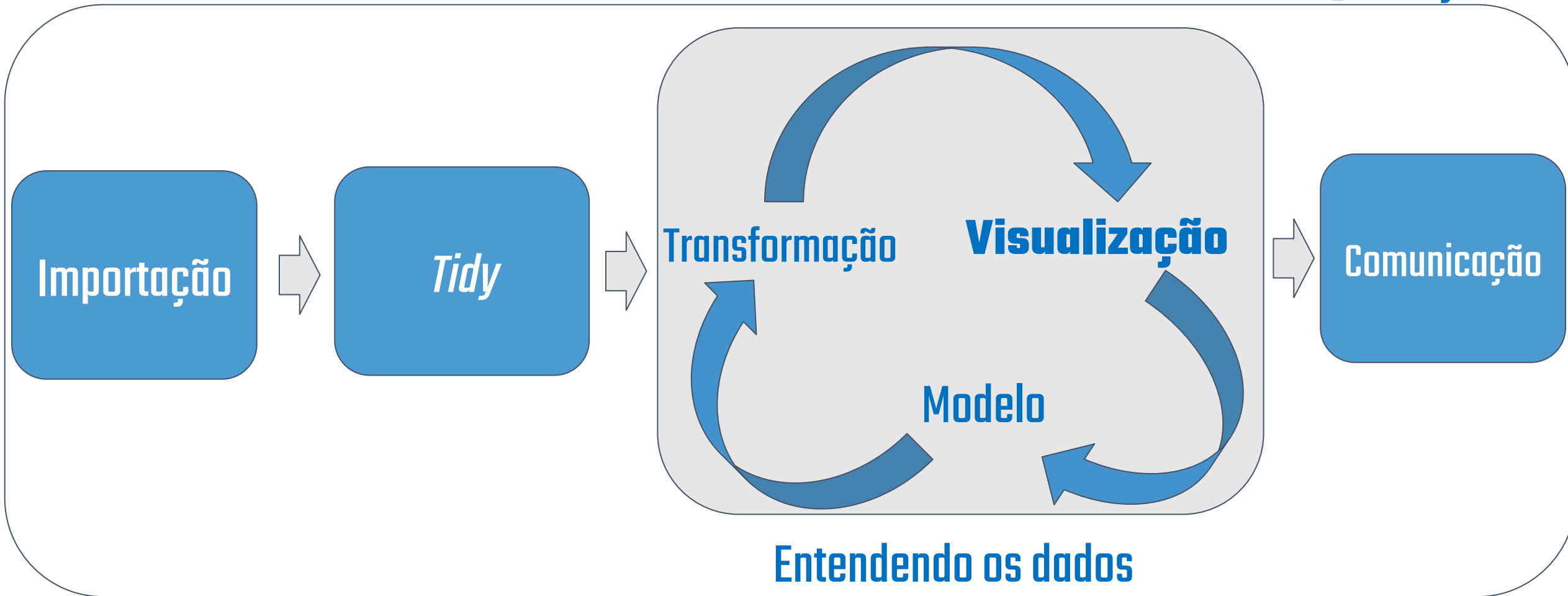


Depois da transformação

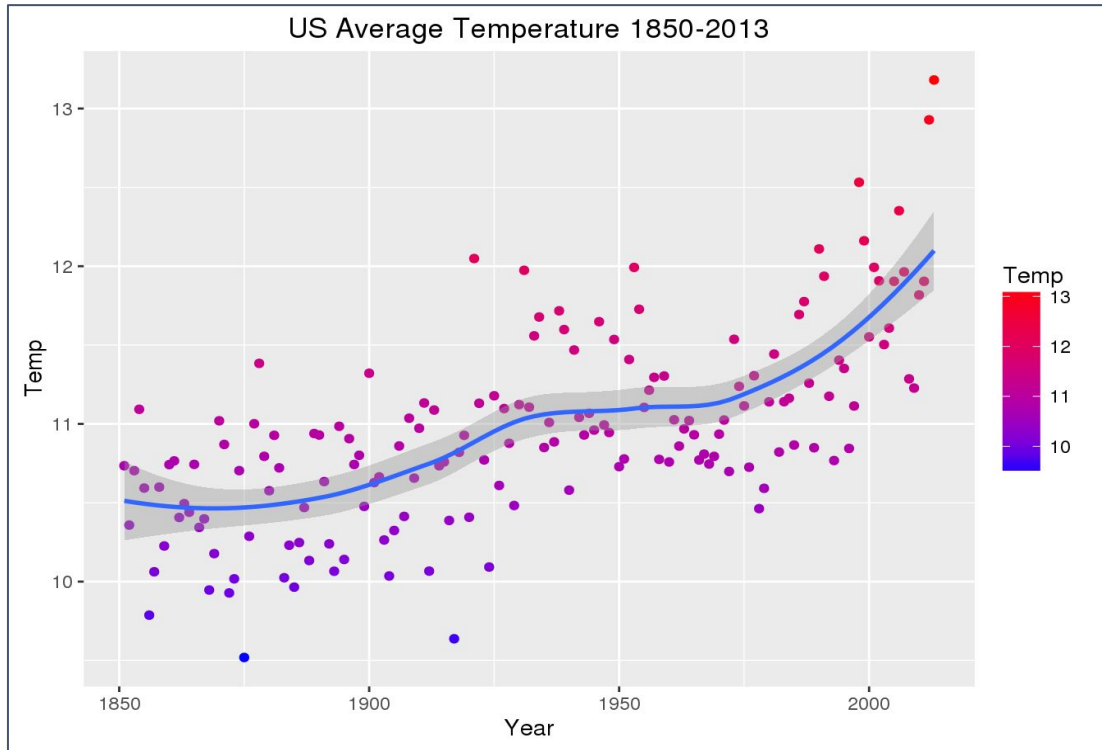
nome	idade (anos)	altura (m)
João	23	1.70
Maria	20	1.68
José	19	1.67
Helena	18	1.72

Processo de Análise de dados

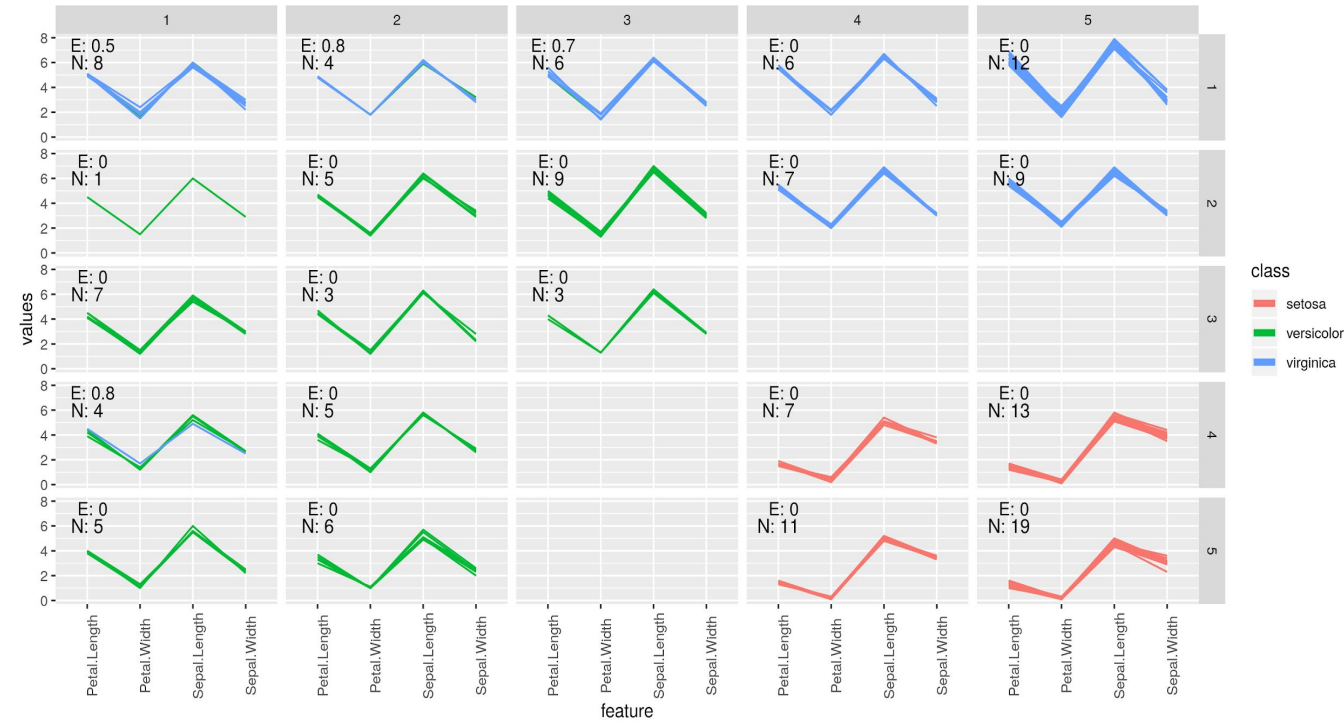
Programação



Visualização



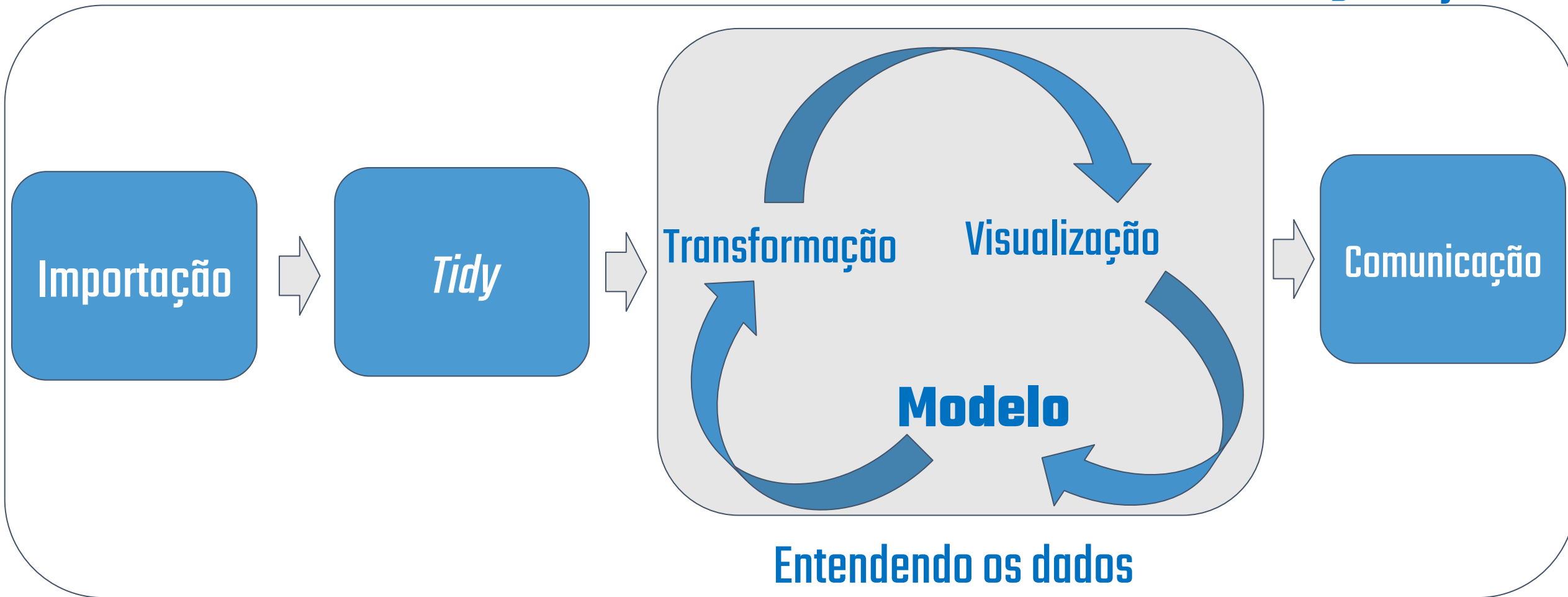
Fonte: Berkeley Earth



Fonte: Pacote ggsum

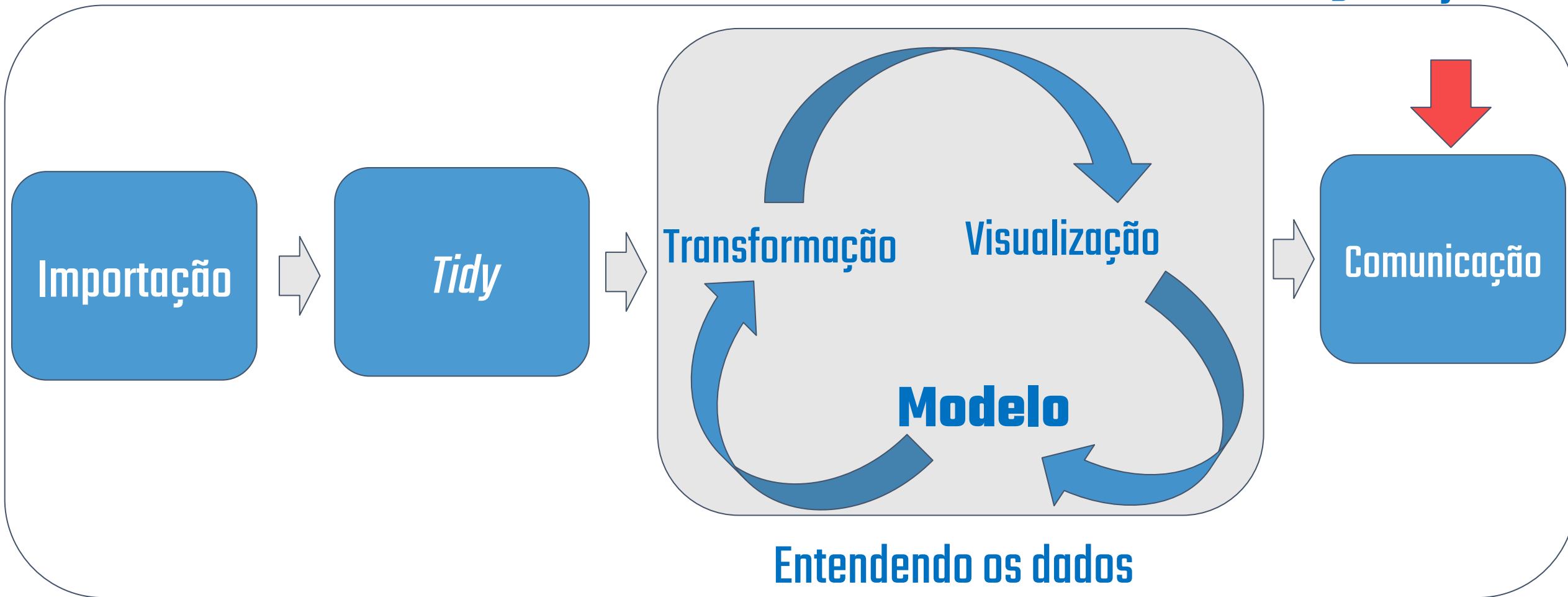
Processo de Análise de dados

Programação



Processo de Análise de dados

Programação



Plataformas:

→ Dash

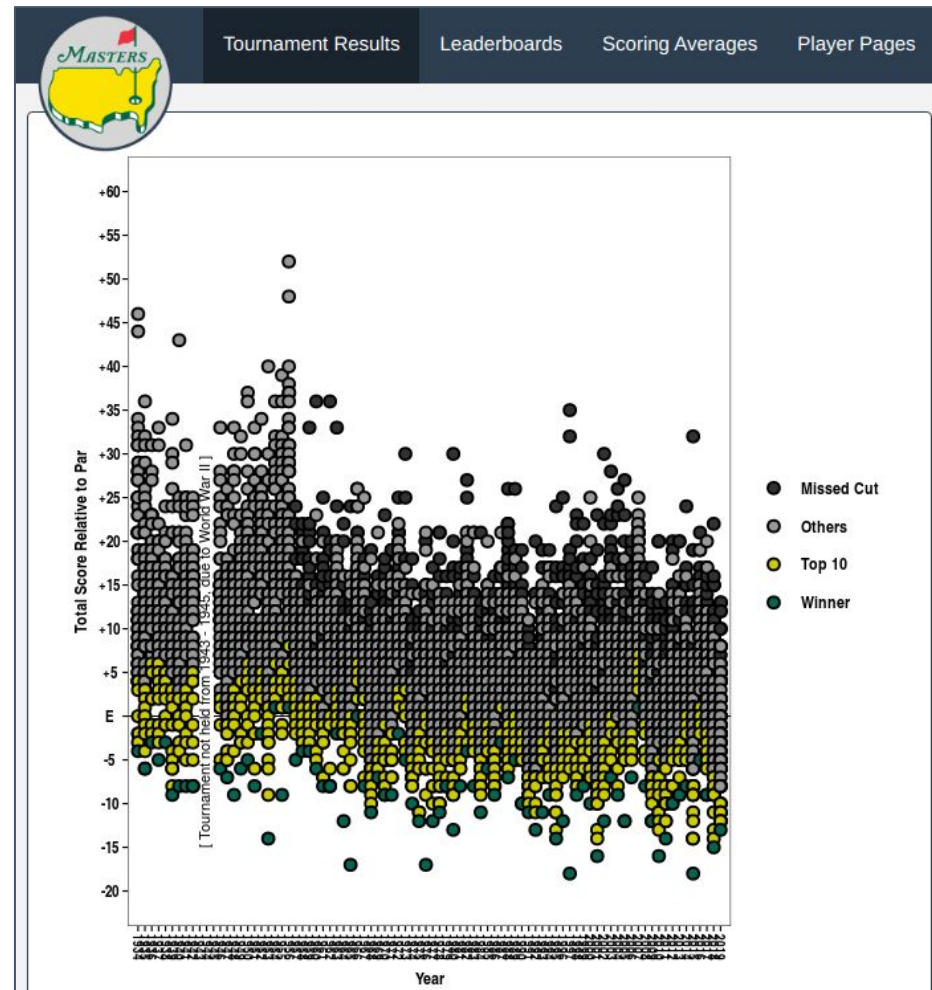
→ Shiny

Relatórios:

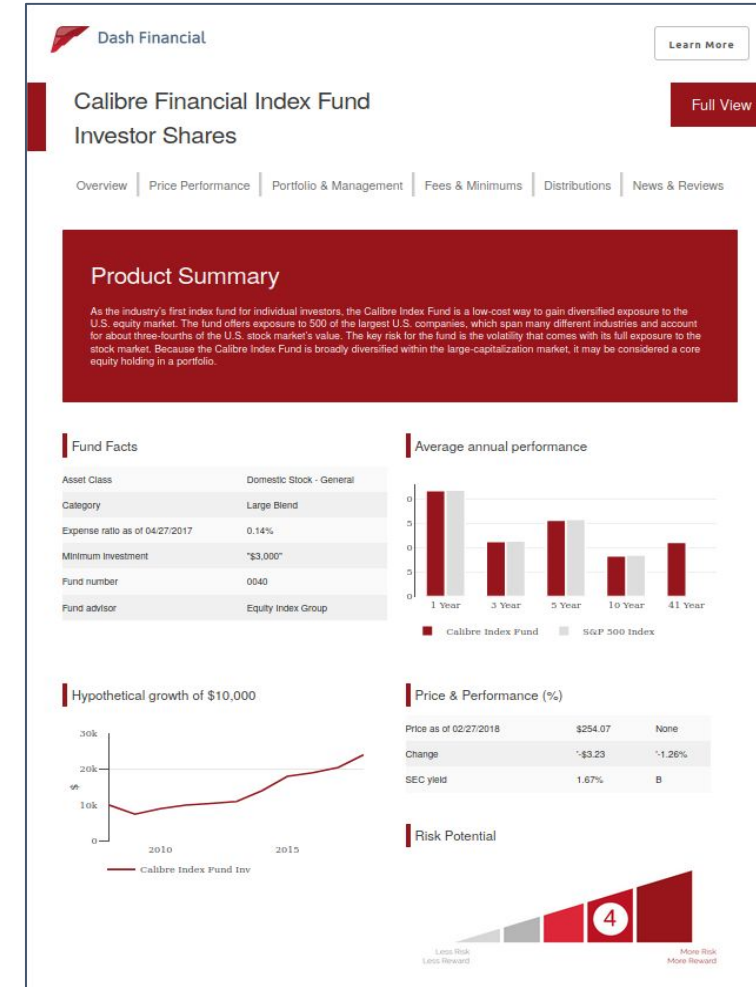
→ Rmd

→ LaTeX

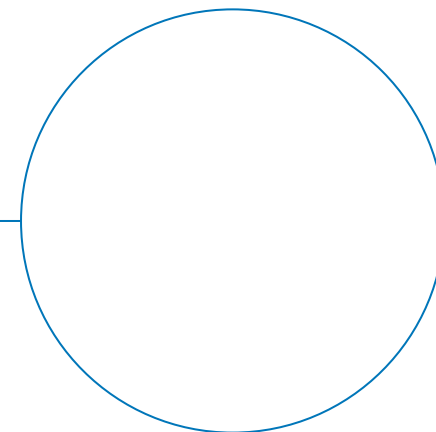
Shiny - Masters Golf Tournament



Dash - Controle financeiro



FAZENDO ANÁLISE DE DADOS

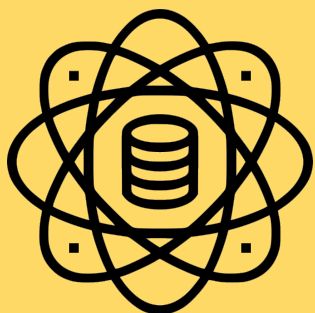


Linguagem - R ou Python ?



Equipe DataAt

Introdução à Análise de Dados



22 de Julho de 2022
São José dos Campos - SP

Adriano Almeida
Felipe Carvalho
Felipe Menino