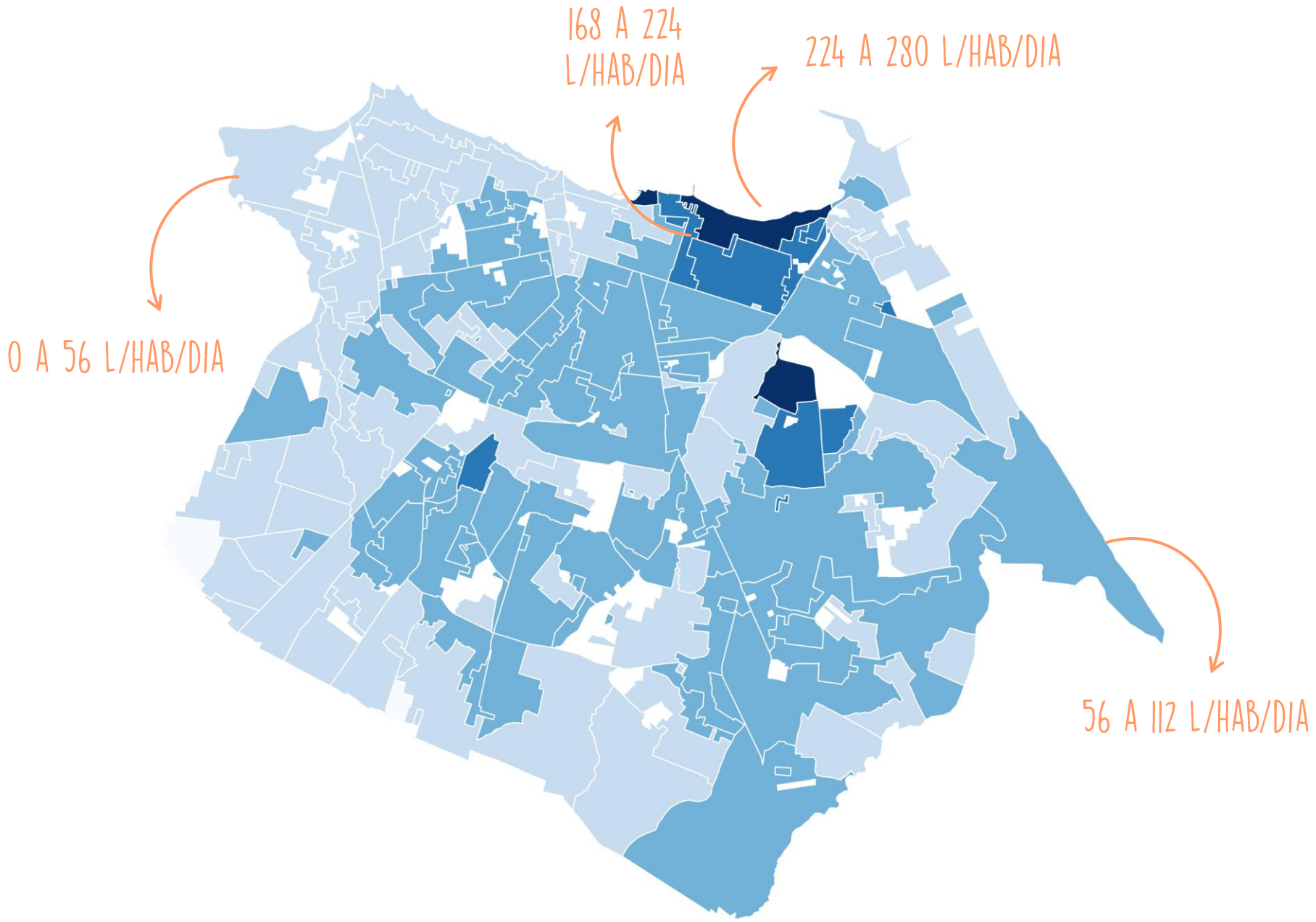
COMO A INTELIGÊNCIA ARTIFICIAL PODE NOS AJUDAR A ENTENDER A **DEMANDA DE ÁGUA** DE UMA CIDADE?

TAÍS CARVALHO

DEMANDA DE ÁGUA NOS PRÓXIMOS 30 ANOS

CONSUMIDOR PADRÃO

128 LITROS DE
ÁGUA/DIA


168 A 224
L/HAB/DIA
224 A 280 L/HAB/DIA
0 A 56 L/HAB/DIA
56 A 112 L/HAB/DIA

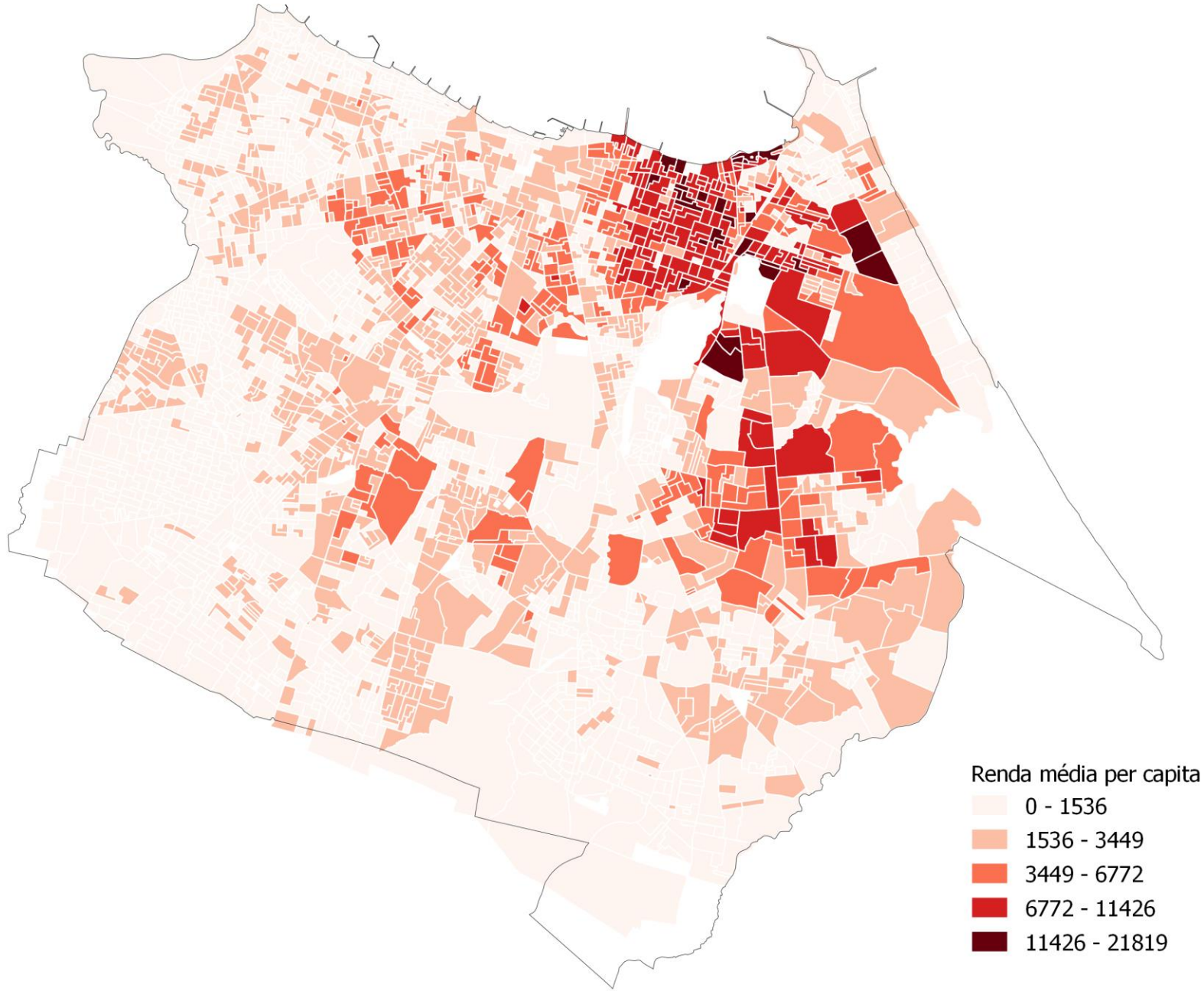
IBGE
Instituto Brasileiro de Geografia e Estatística

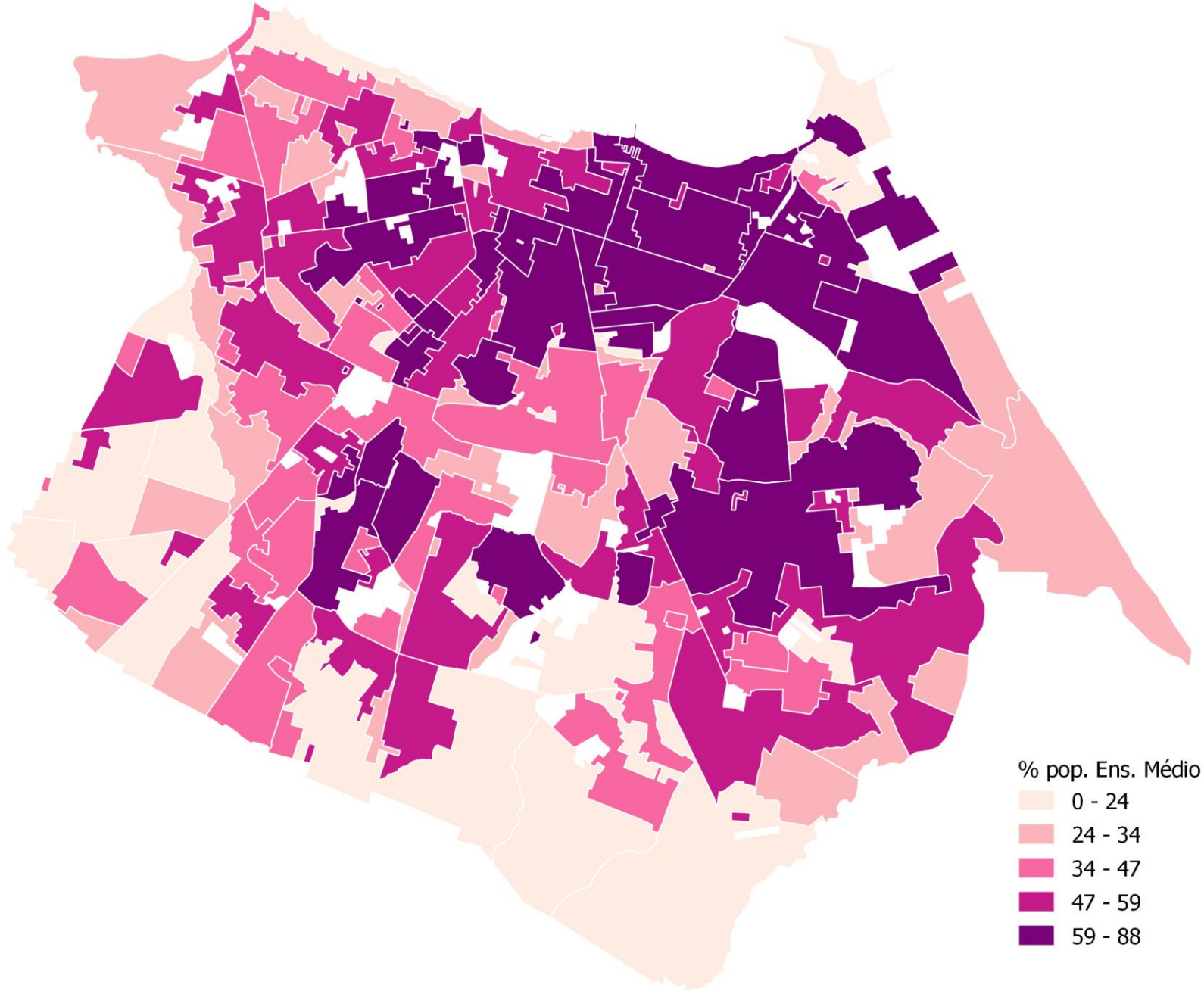
Atlas do Desenvolvimento Humano no Brasil

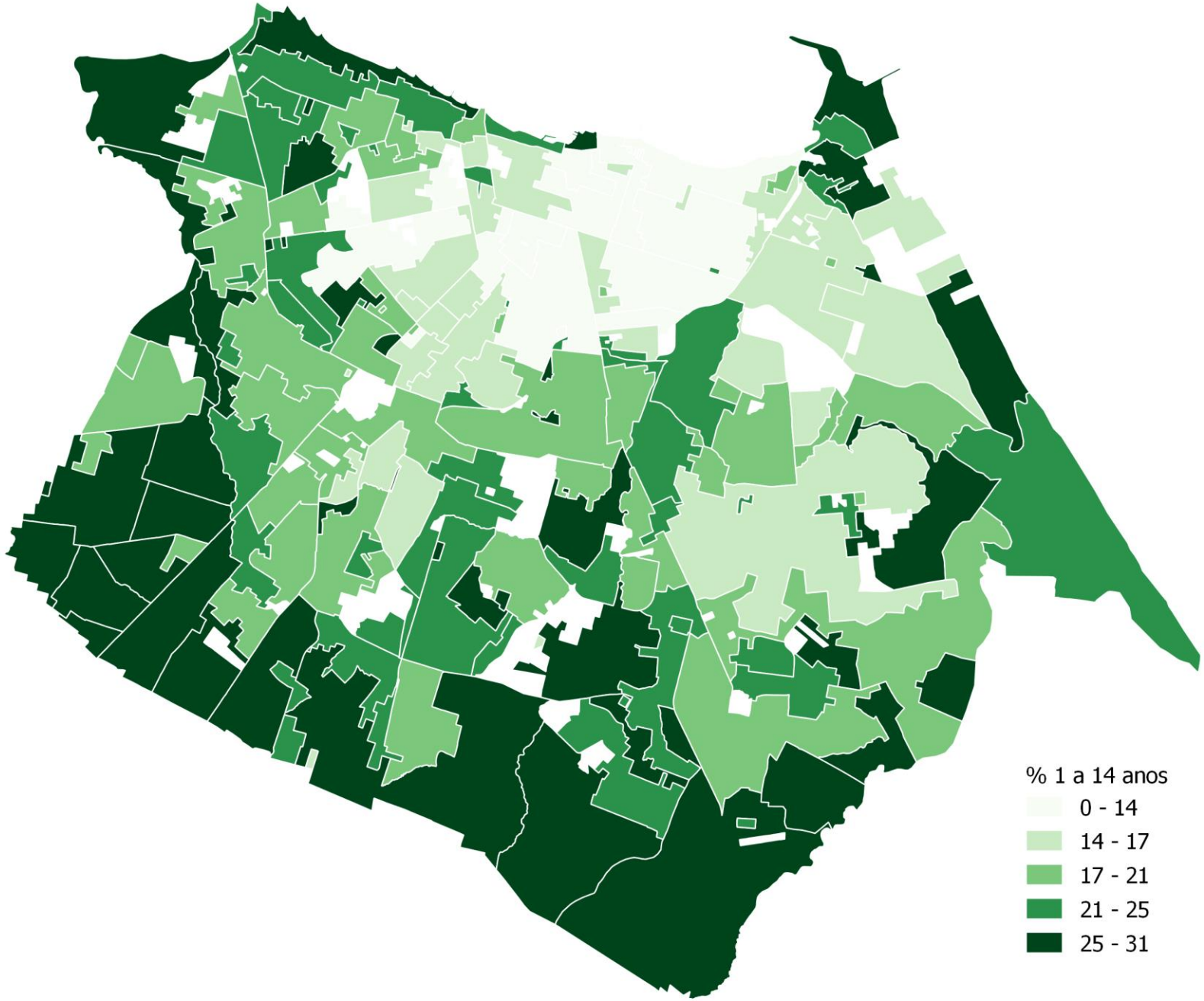
Fortaleza
PREFEITURA

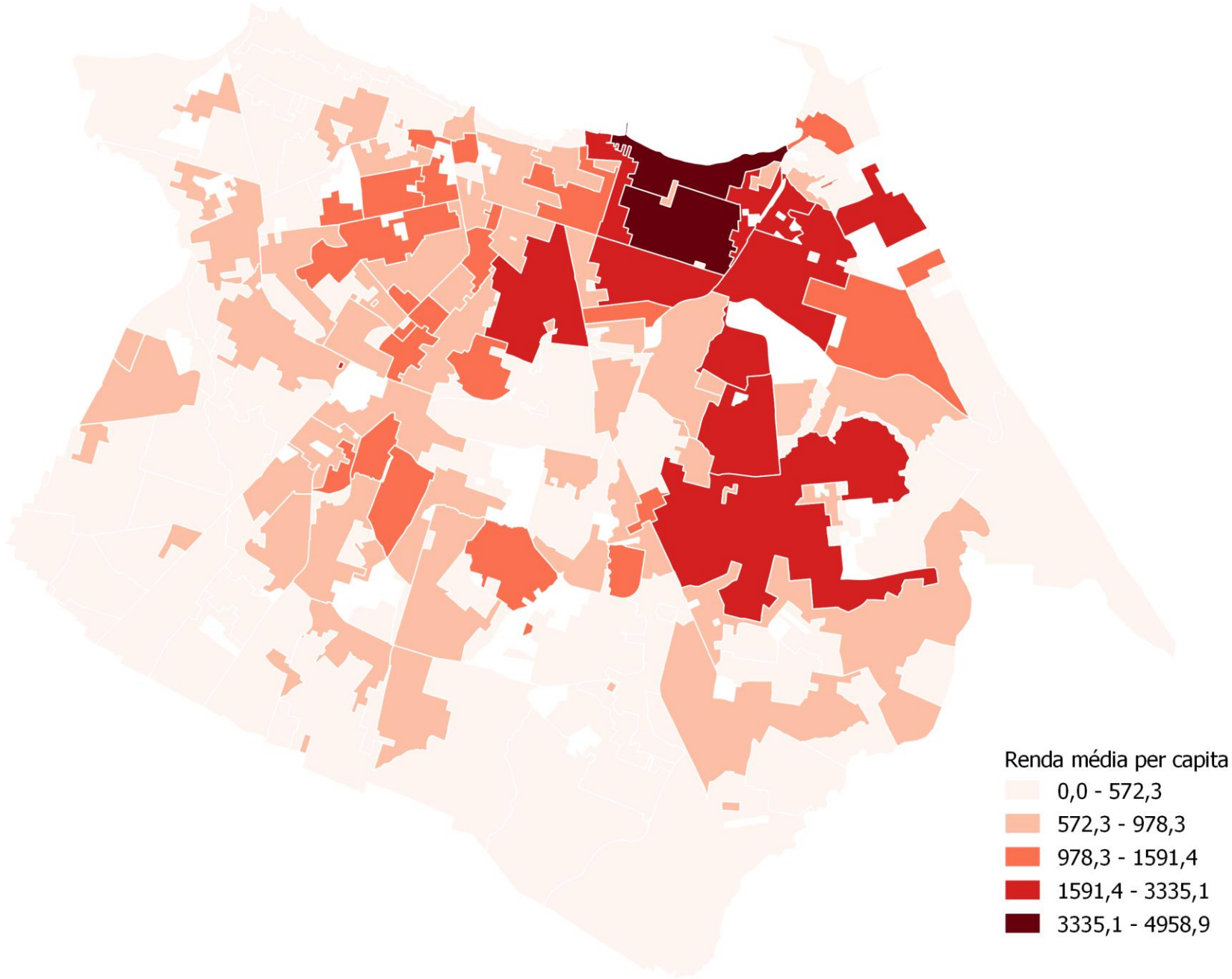
...

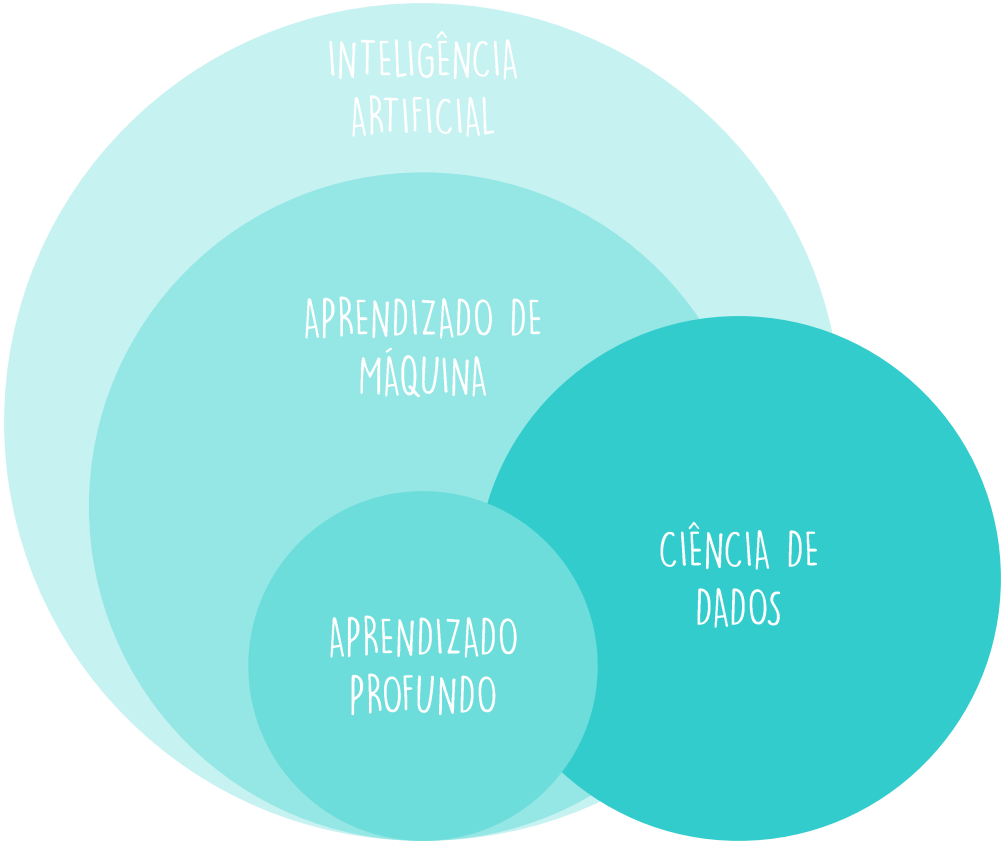
...

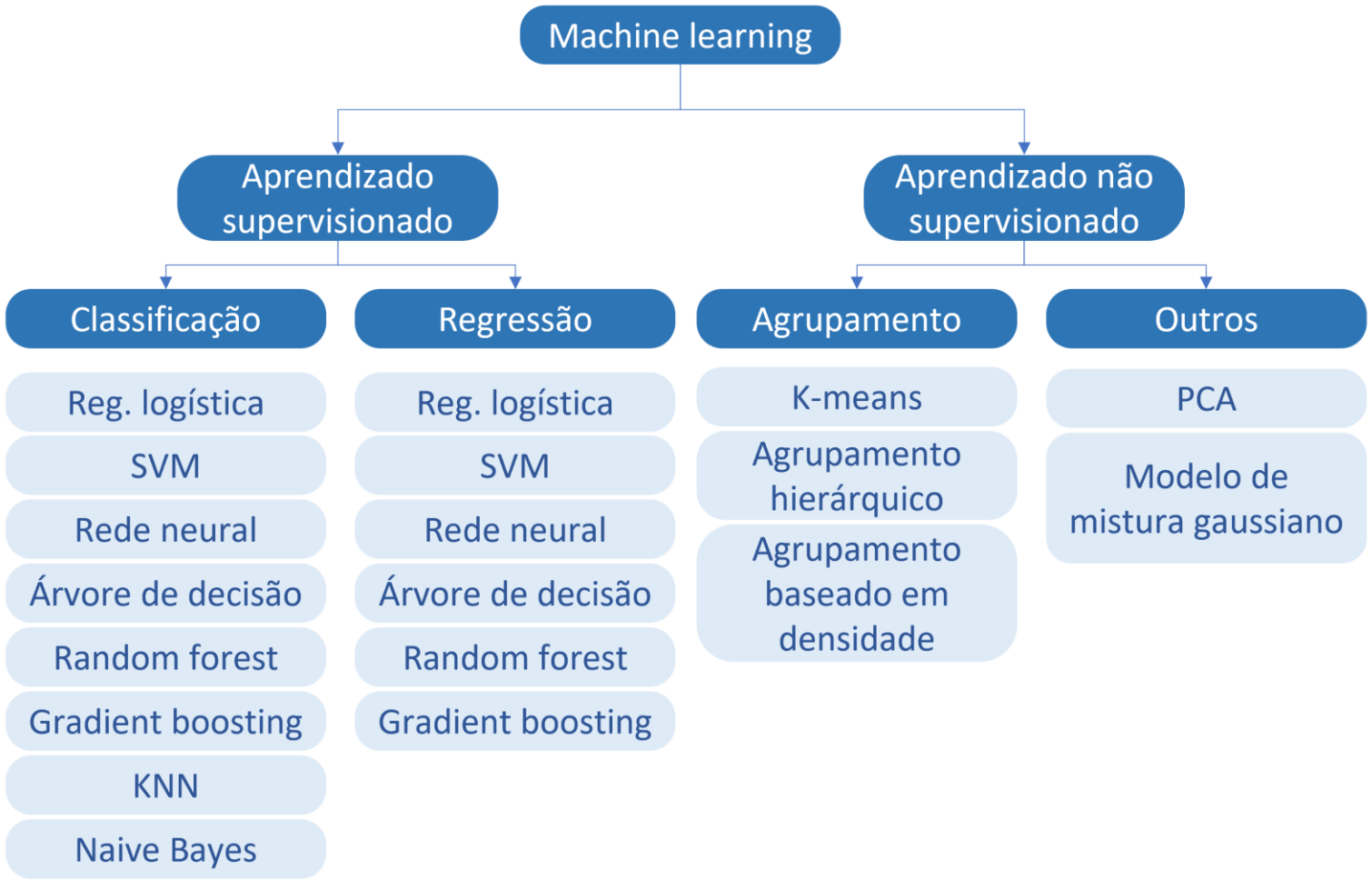

Renda média per capita
0 - 1536
1536 - 3449
3449 - 6772
6772 - 11426
11426 - 21819


% pop. Ens. Médio
0 - 24
24 - 34
34 - 47
47 - 59
59 - 88


% 1 a 14 anos
0 - 14
14 - 17
17 - 21
21 - 25
25 - 31


Renda média per capita
0,0 - 572,3
572,3 - 978,3
978,3 - 1591,4
1591,4 - 3335,1
3335,1 - 4958,9


INTELIGÊNCIA ARTIFICIAL
APRENDIZADO DE MÁQUINA
APRENDIZADO PROFUNDO
CIÊNCIA DE DADOS


Machine learning
Aprendizado supervisionado
Aprendizado não supervisionado
Classificação
Regressão
Agrupamento
Outros
Reg. logística
SVM
Rede neural
Árvore de decisão
Random forest
Gradient boosting
KNN
Naive Bayes
Reg. logística
SVM
Rede neural
Árvore de decisão
Random forest
Gradient boosting
K-means
Agrupamento hierárquico
Agrupamento baseado em densidade
PCA
Modelo de mistura gaussiano

LIMPEZA DOS DADOS

Verificação de dados faltantes

Padronização dos dados

Remoção de outliers

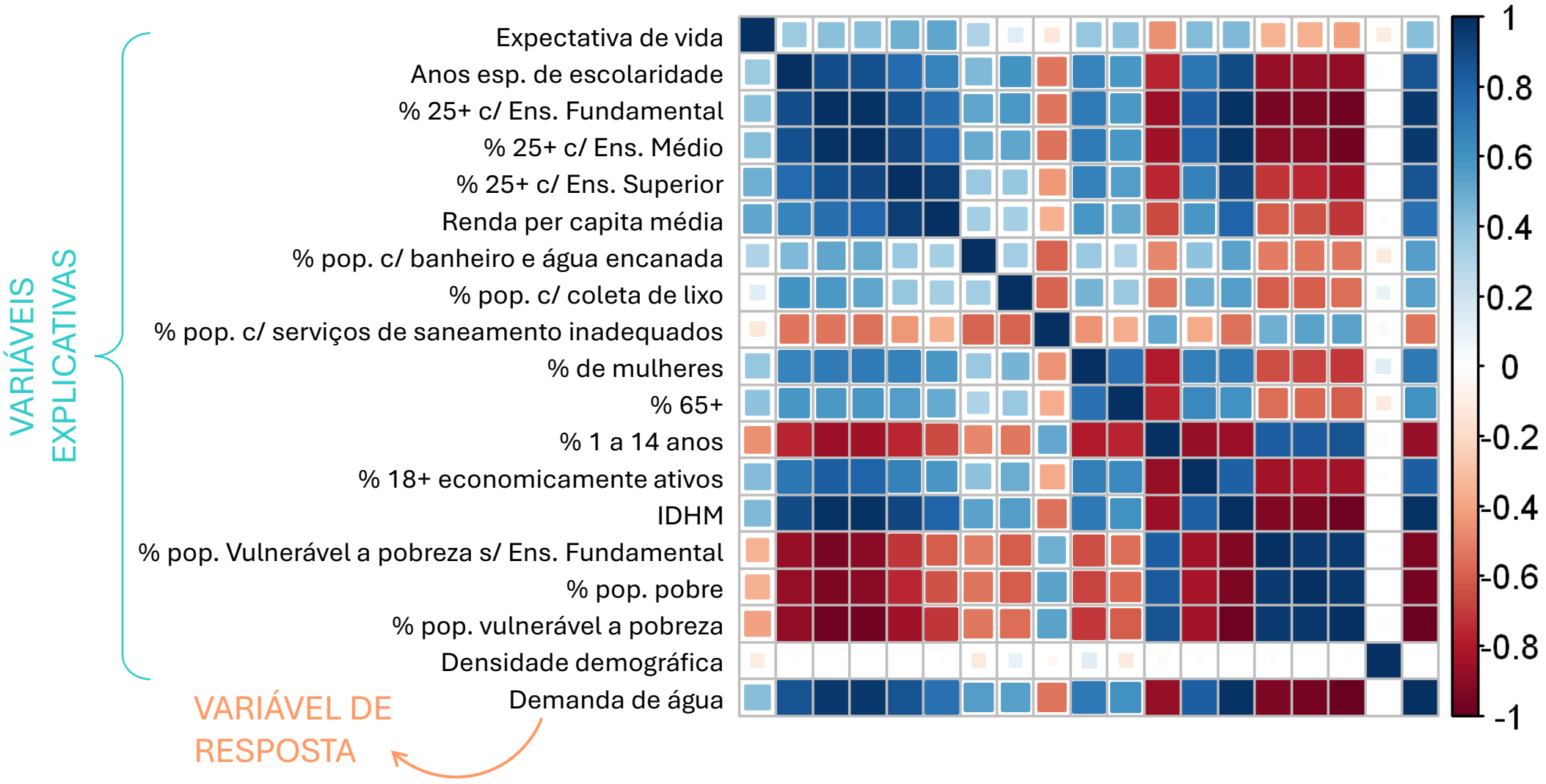
Verificação de dados redundantes

ANÁLISE DE CORRELAÇÃO

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sqrt{\sum(x - \bar{x})^2 \sum(y - \bar{y})^2}}$$

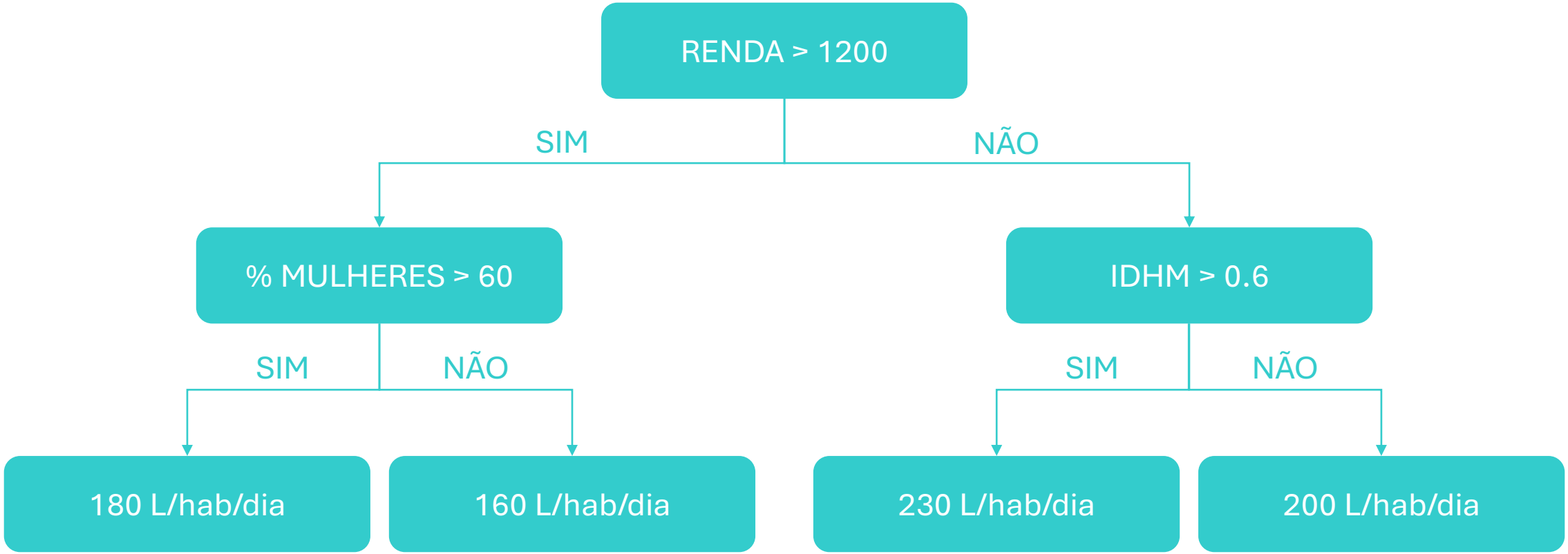
COEFICIENTE DE CORRELAÇÃO DE PEARSON

ANÁLISE DE CORRELAÇÃO

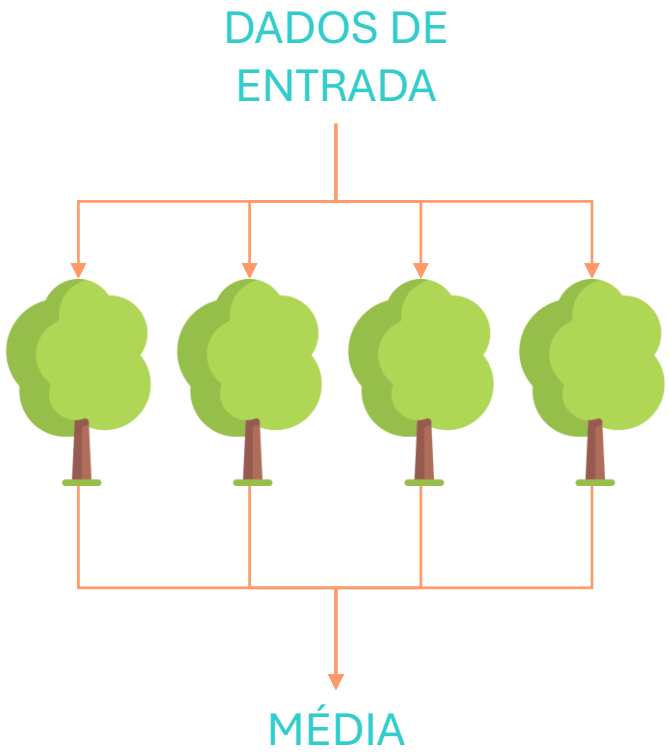


SELEÇÃO DE VARIÁVEIS

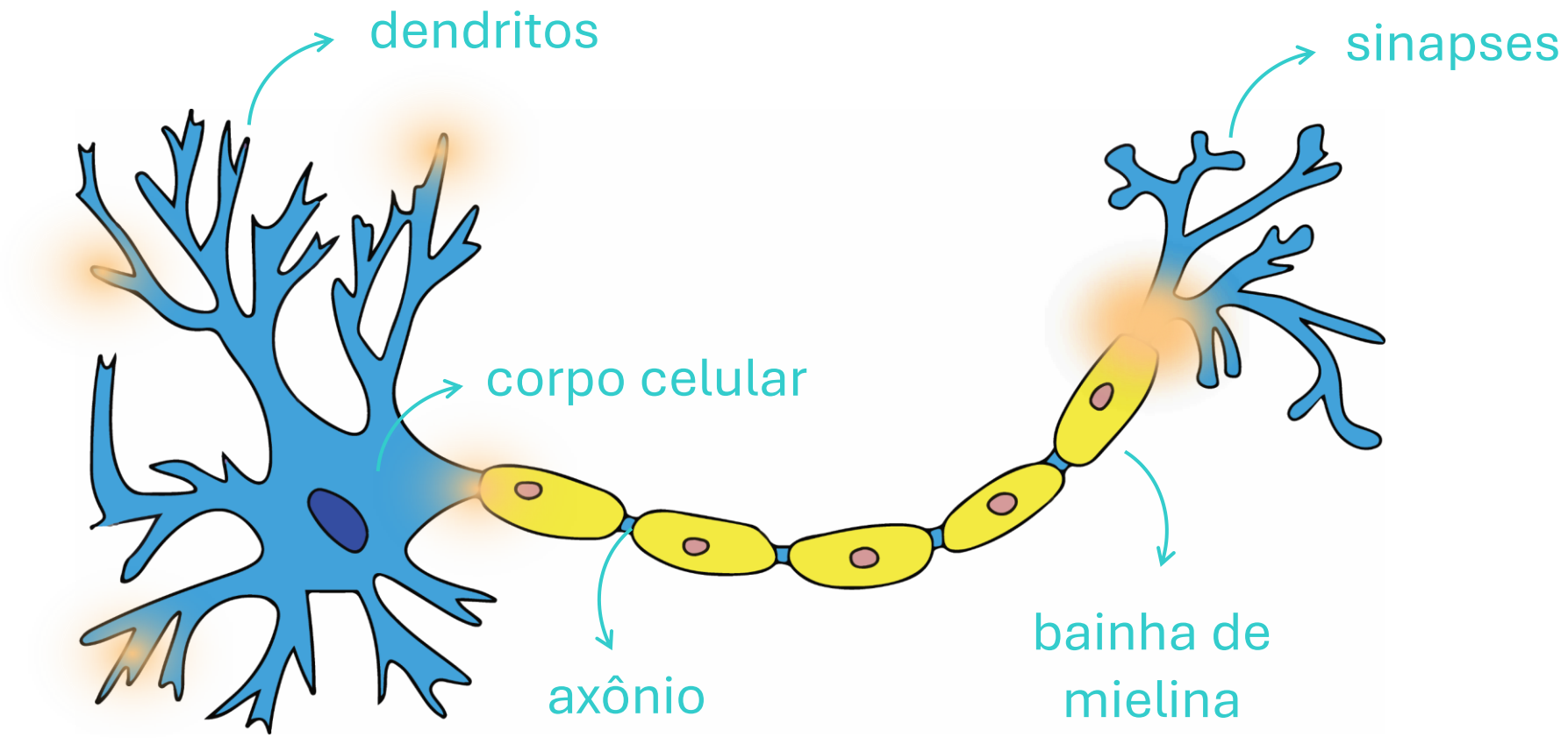
ÁRVORE DE DECISÃO



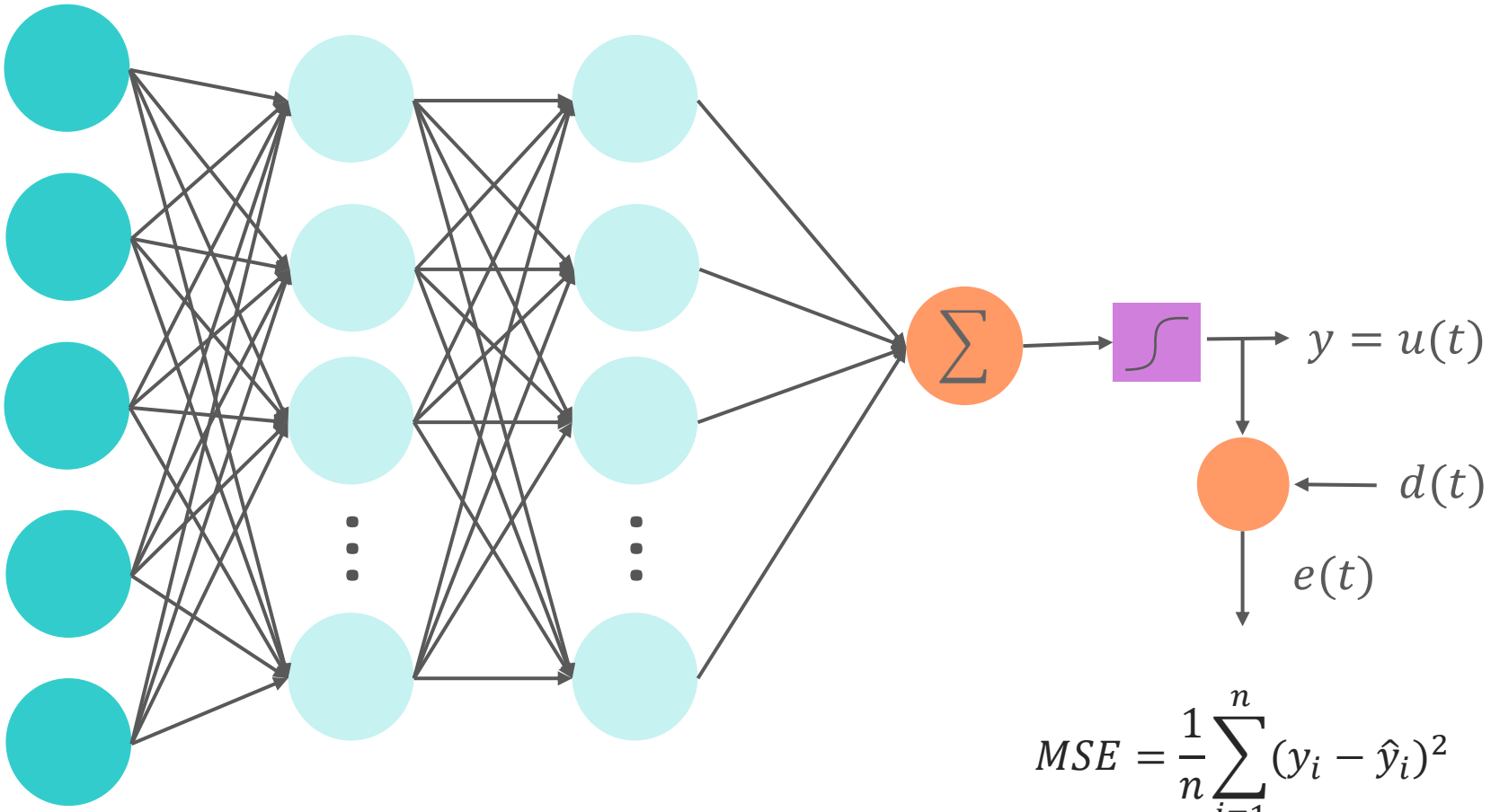
RANDOM FOREST



MODELO DE PREVISÃO:
REDES NEURAIS


dendritos
sinapses
corpo celular
axônio
bainha de mielina

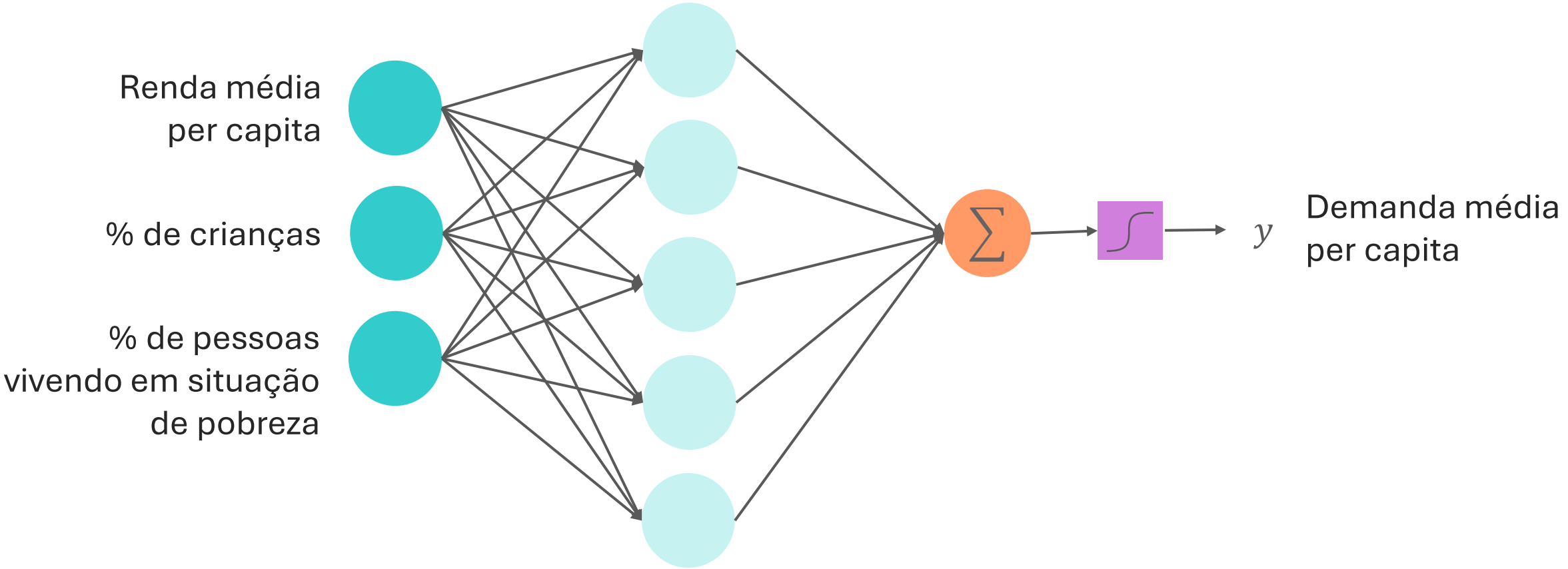
REDE NEURAL



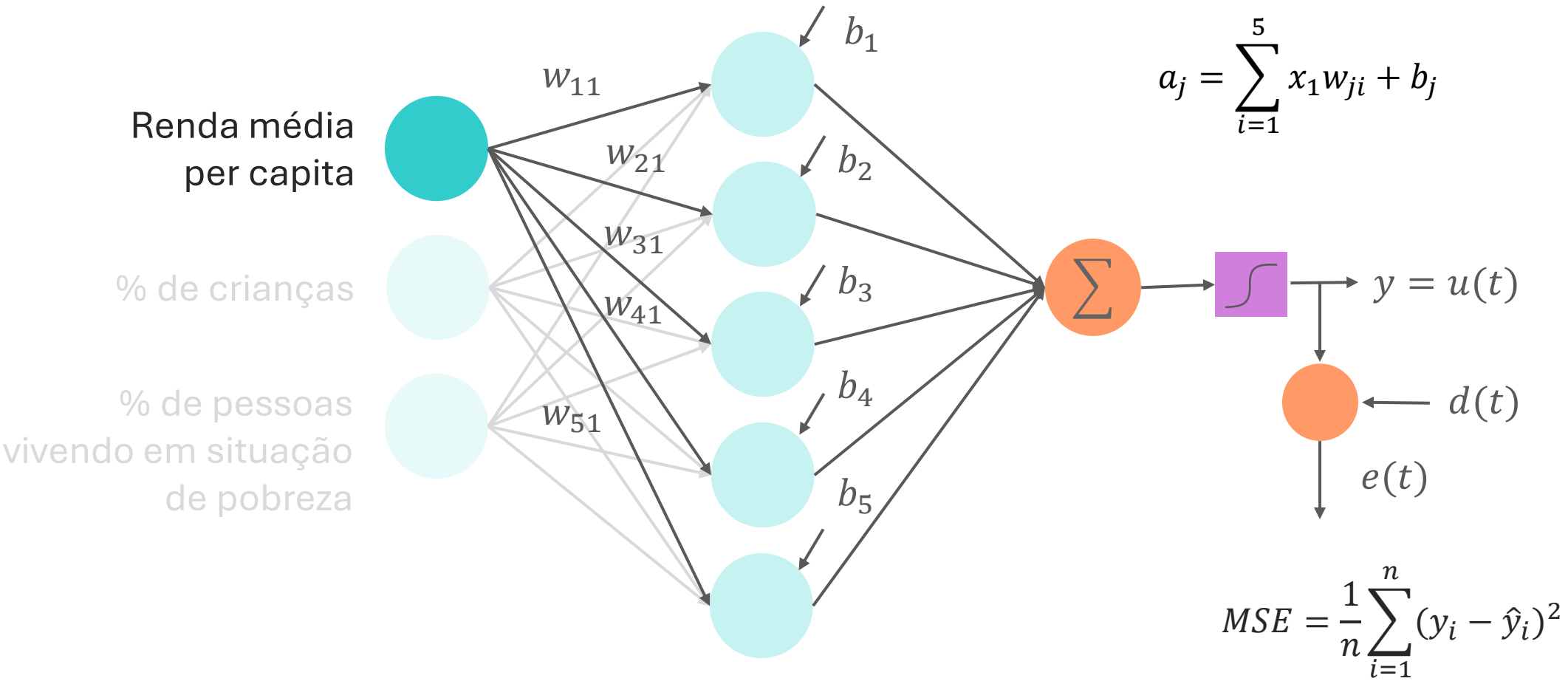
$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

ERRO QUADRÁTICO MÉDIO

REDE NEURAL


Renda média per capita
% de crianças
% de pessoas vivendo em situação de pobreza
∑
y
Demanda média per capita

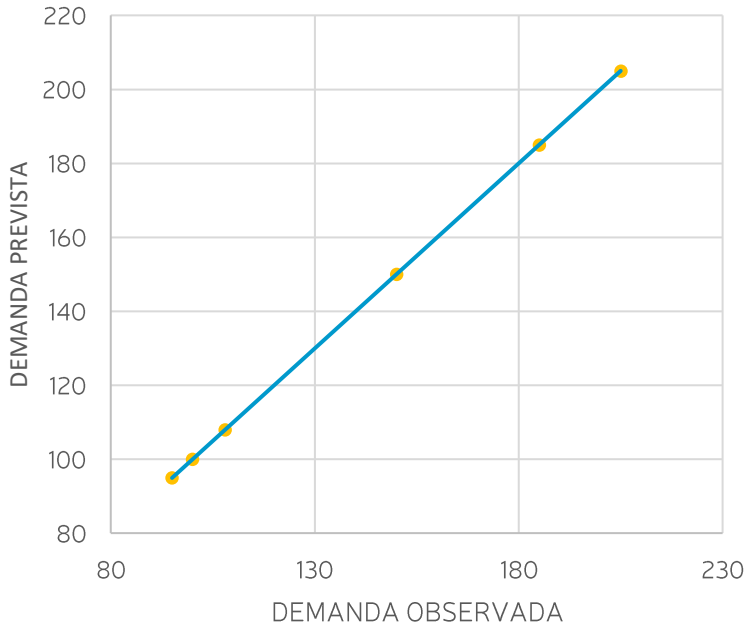
REDE NEURAL


Renda média per capita
% de crianças
% de pessoas vivendo em situação de pobreza
w_{11}
w_{21}
w_{31}
w_{41}
w_{51}
b_1
b_2
b_3
b_4
b_5
$\sum$
$\int$
$y = u(t)$
$d(t)$
$e(t)$
$a_j = \sum_{i=1}^{5} x_1 w_{ji} + b_j$
$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$

Como saber se o modelo está representando bem o fenômeno?

MEDIDA DE DESEMPENHO

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2}$$



R^2

$R^2{}_1$

$R^2{}_2$

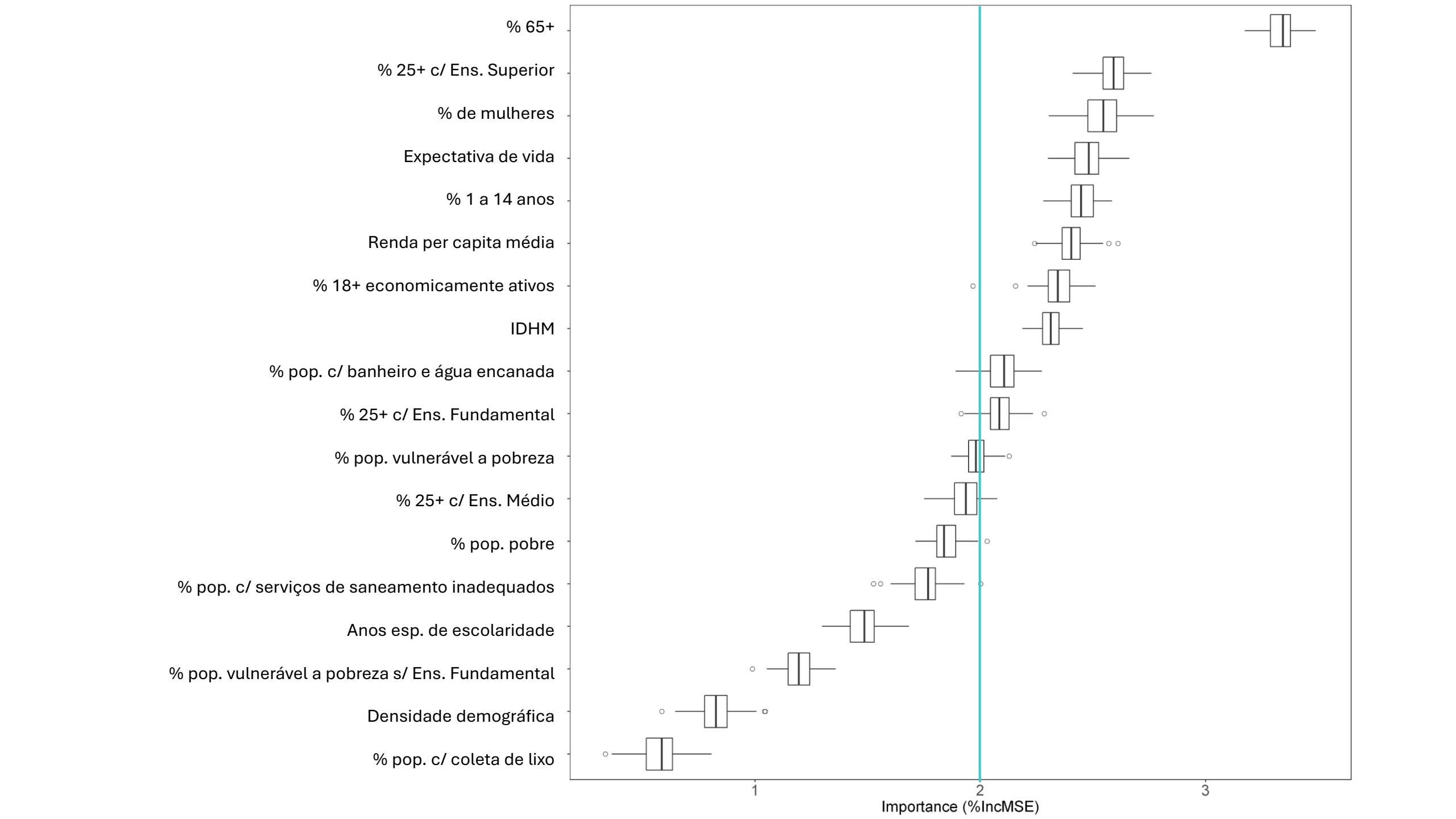
$R^2{}_3$

$R^2{}_4$

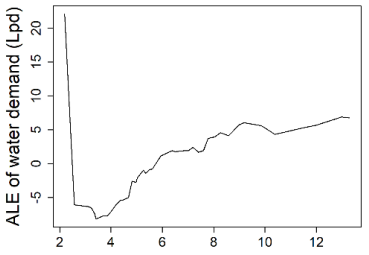
$R^2{}_5$

$$MSE = \sum_{i=1}^{5} R^2{}_i \Big/ 5$$

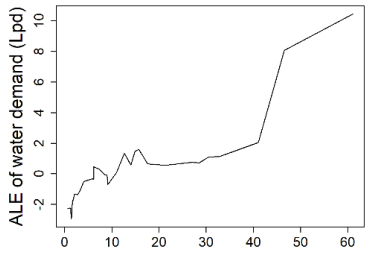
Interpretação dos Resultados


% 65+
% 25+ c/ Ens. Superior
% de mulheres
Expectativa de vida
% 1 a 14 anos
Renda per capita média
% 18+ economicamente ativos
IDHM
% pop. c/ banheiro e água encanada
% 25+ c/ Ens. Fundamental
% pop. vulnerável a pobreza
% 25+ c/ Ens. Médio
% pop. pobre
% pop. c/ serviços de saneamento inadequados
Anos esp. de escolaridade
% pop. vulnerável a pobreza s/ Ens. Fundamental
Densidade demográfica
% pop. c/ coleta de lixo
1
2
3
Importance (%IncMSE)

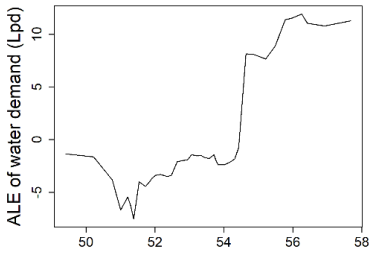
EFEITOS ACUMULADOS LOCAIS



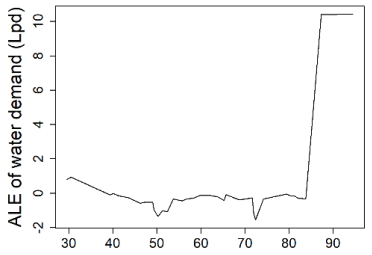
% 65+



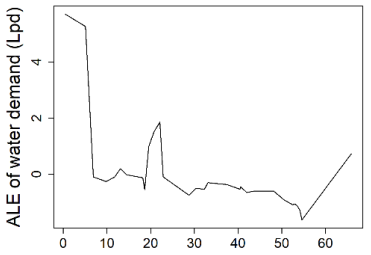
% 25+ c/ Ens. Superior



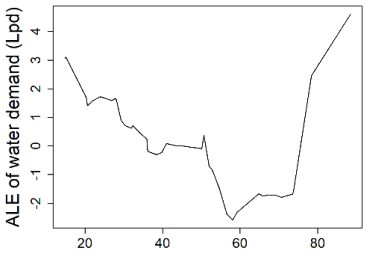
% de mulheres



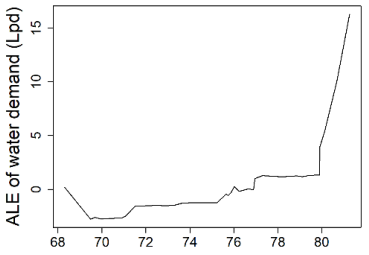
% 25+ c/ Ens. Fundamental



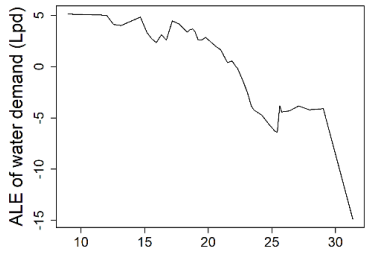
% pop. vulnerável a pobreza



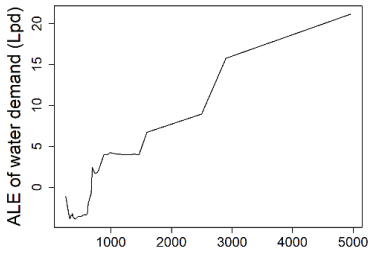
% 25+ c/ Ens. Médio



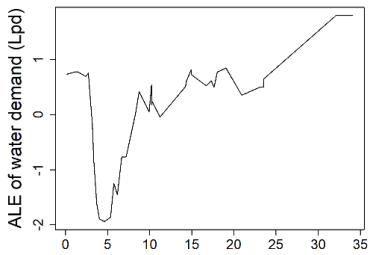
Expectativa de vida



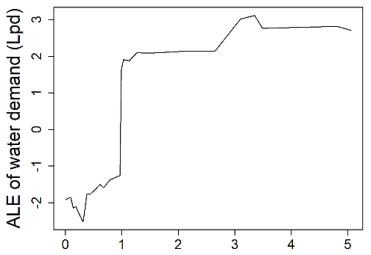
% 1 a 14 anos



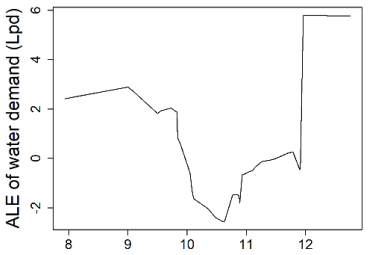
Renda per capita média



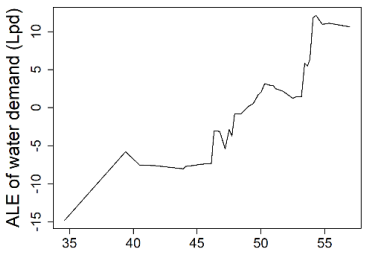
% pop. pobre



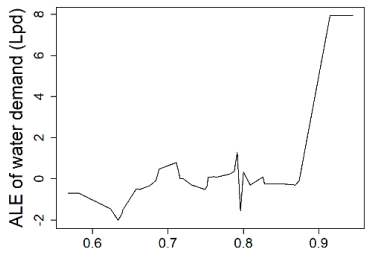
% pop. c/ serv. de saneam. inad.



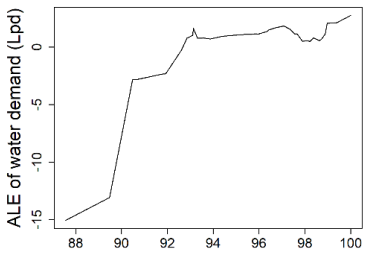
Anos esperados de esc.



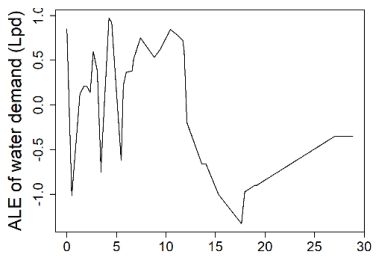
% 18+ economicamente ativos



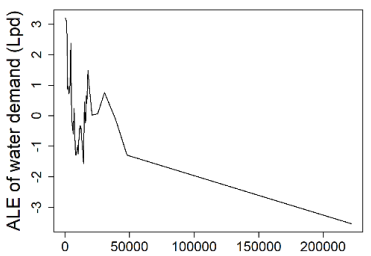
IDHM



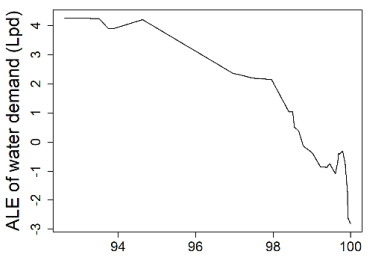
% pop. c/ banheiro e água encanada



% pop. vulnerável a pobreza + s/ Ens. Fundam.

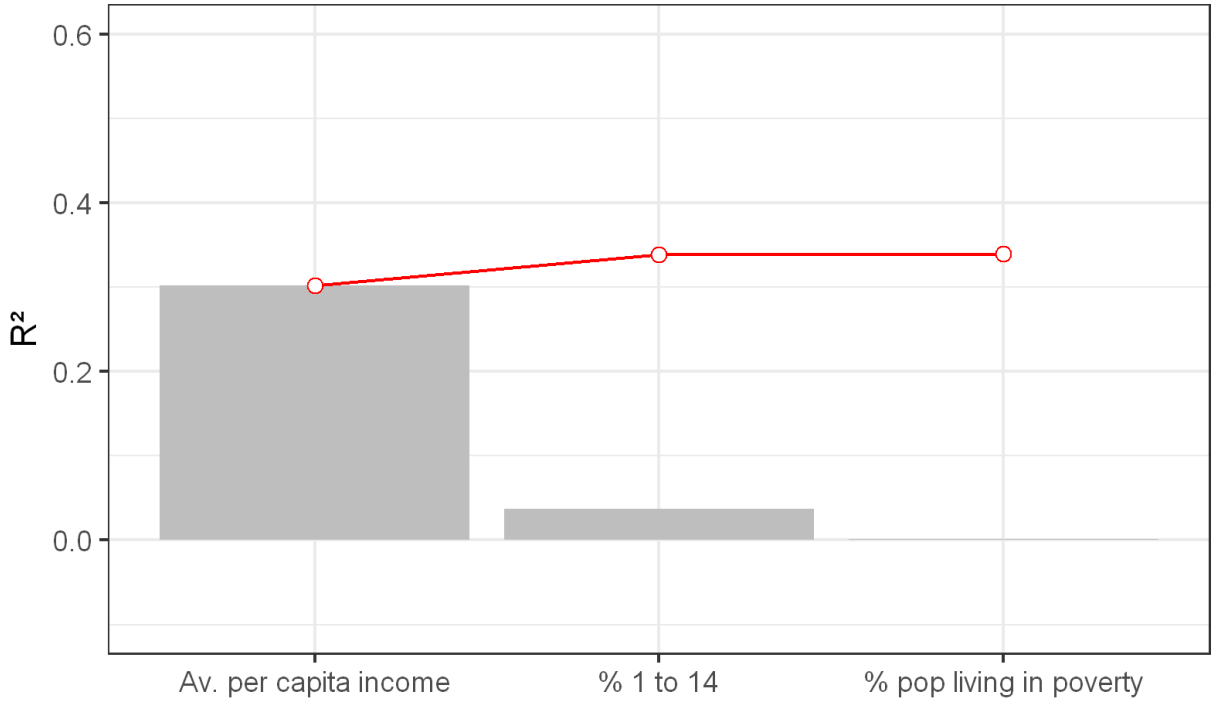


Densidade demográfica



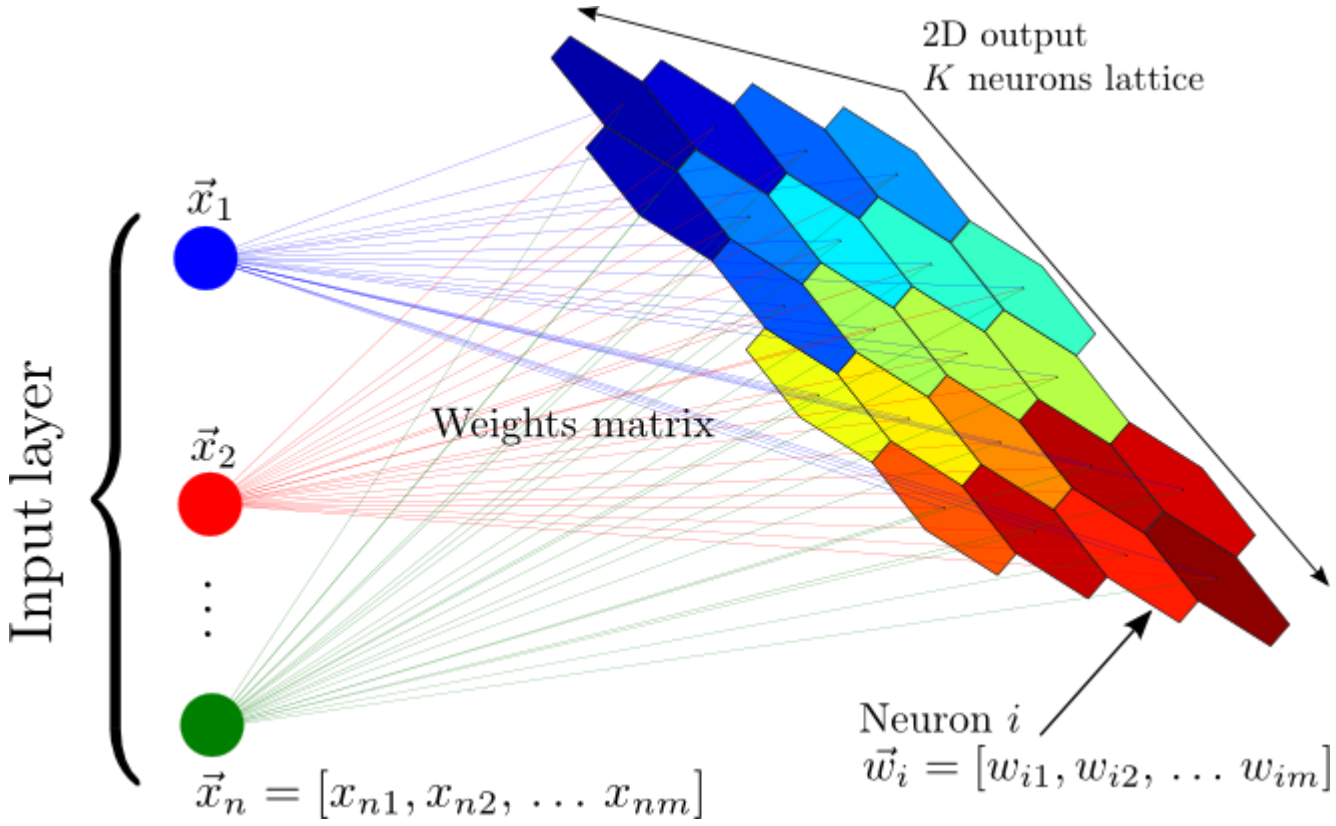
% pop. c/ coleta de lixo

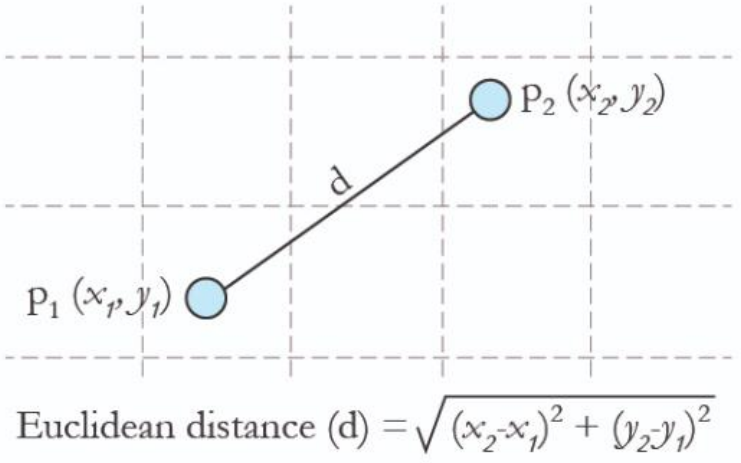
MODELO FINAL

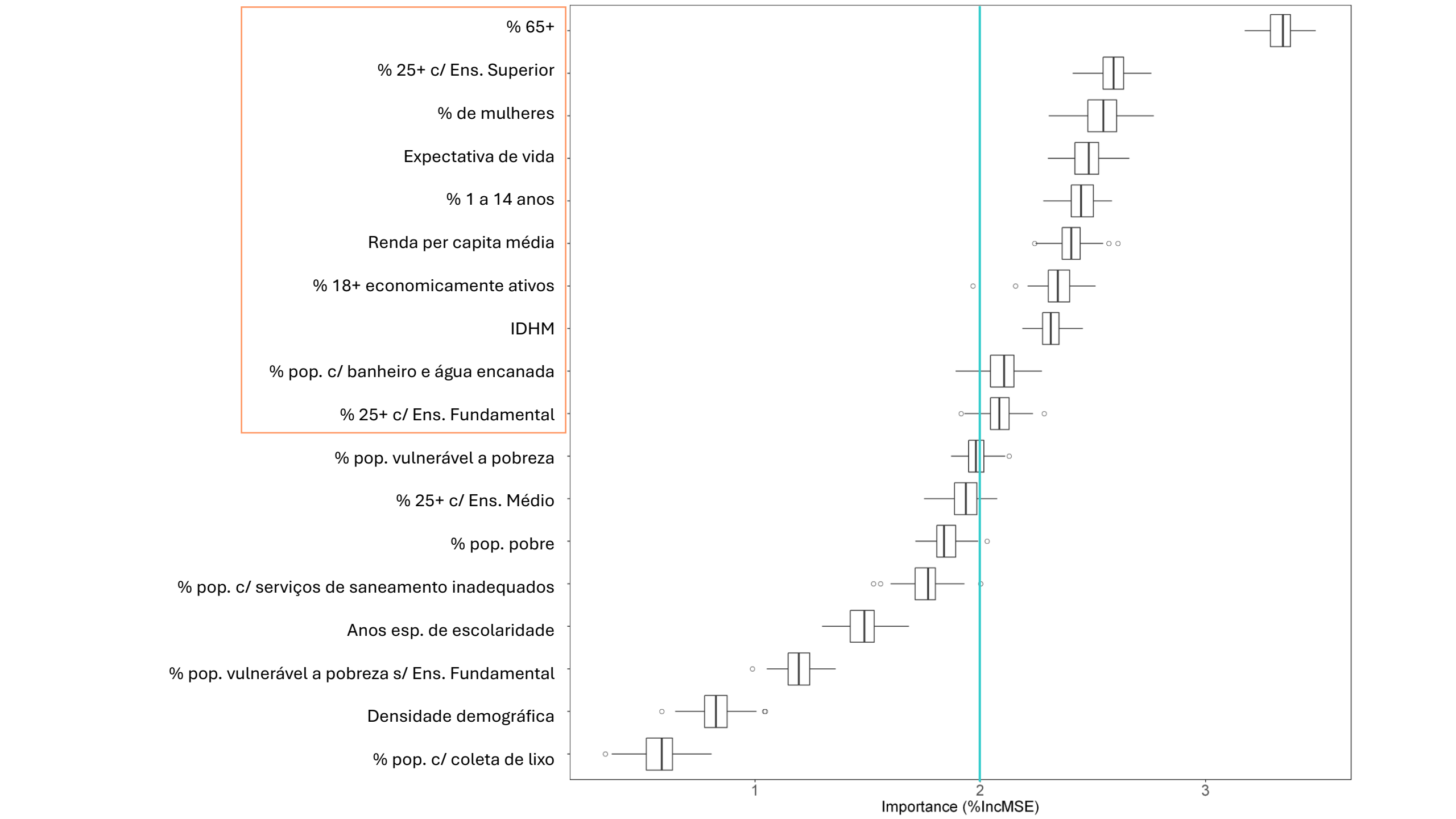


AINDA DÁ PRA APRENDER MAIS?

Análise de Agrupamento






% 65+
% 25+ c/ Ens. Superior
% de mulheres
Expectativa de vida
% 1 a 14 anos
Renda per capita média
% 18+ economicamente ativos
IDHM
% pop. c/ banheiro e água encanada
% 25+ c/ Ens. Fundamental
% pop. vulnerável a pobreza
% 25+ c/ Ens. Médio
% pop. pobre
% pop. c/ serviços de saneamento inadequados
Anos esp. de escolaridade
% pop. vulnerável a pobreza s/ Ens. Fundamental
Densidade demográfica
% pop. c/ coleta de lixo
1
2
3
Importance (%IncMSE)

U-matrix

Expectativa de vida

% 25+ c/ Ens. Fundamental

% 25+ c/ Ens. Superior

Renda per capita média

% pop. c/ banheiro e água encanada

% de mulheres

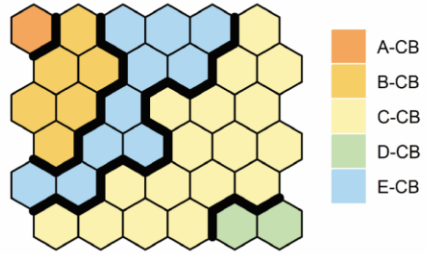
% 65+

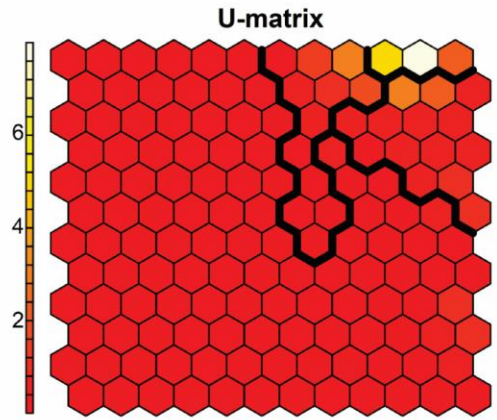
% 1 a 14 anos

% 18+ economicamente ativos

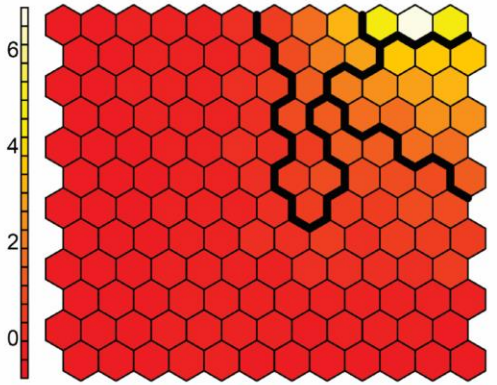
IDHM

Clusters

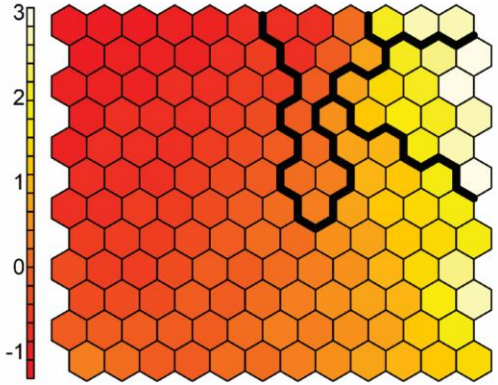


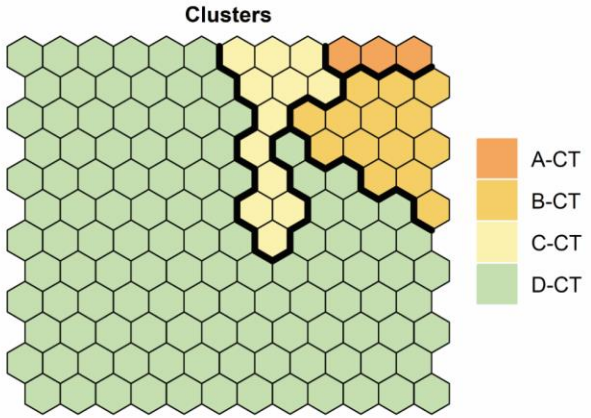


Renda per capita média

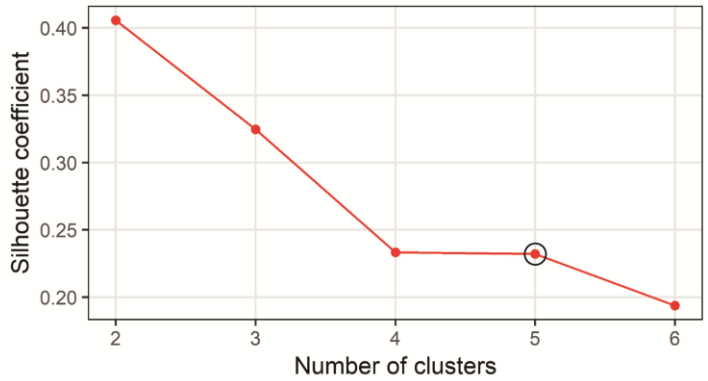


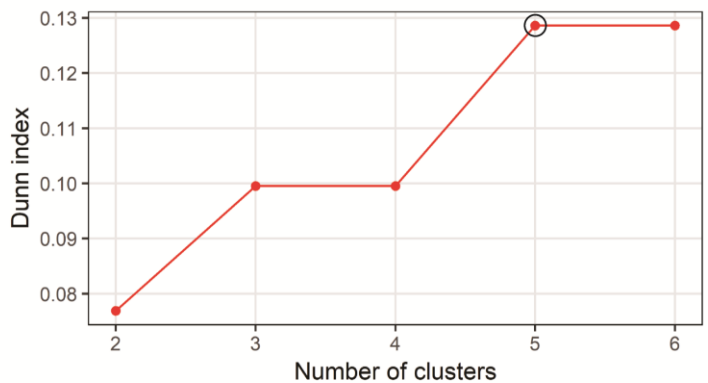
IDHM



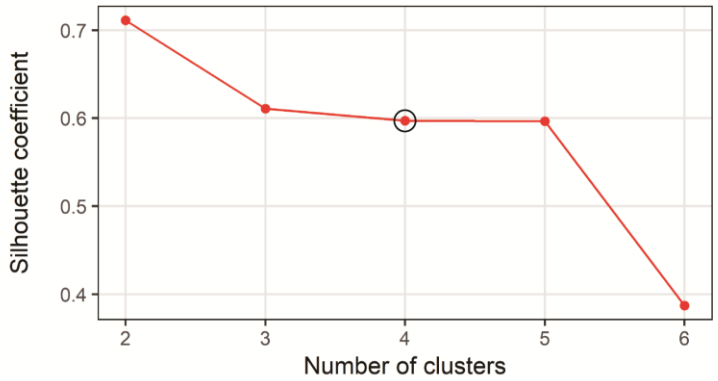


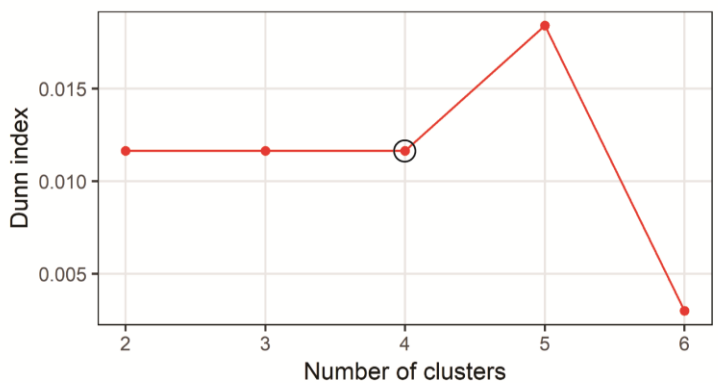
Bloco Censitário

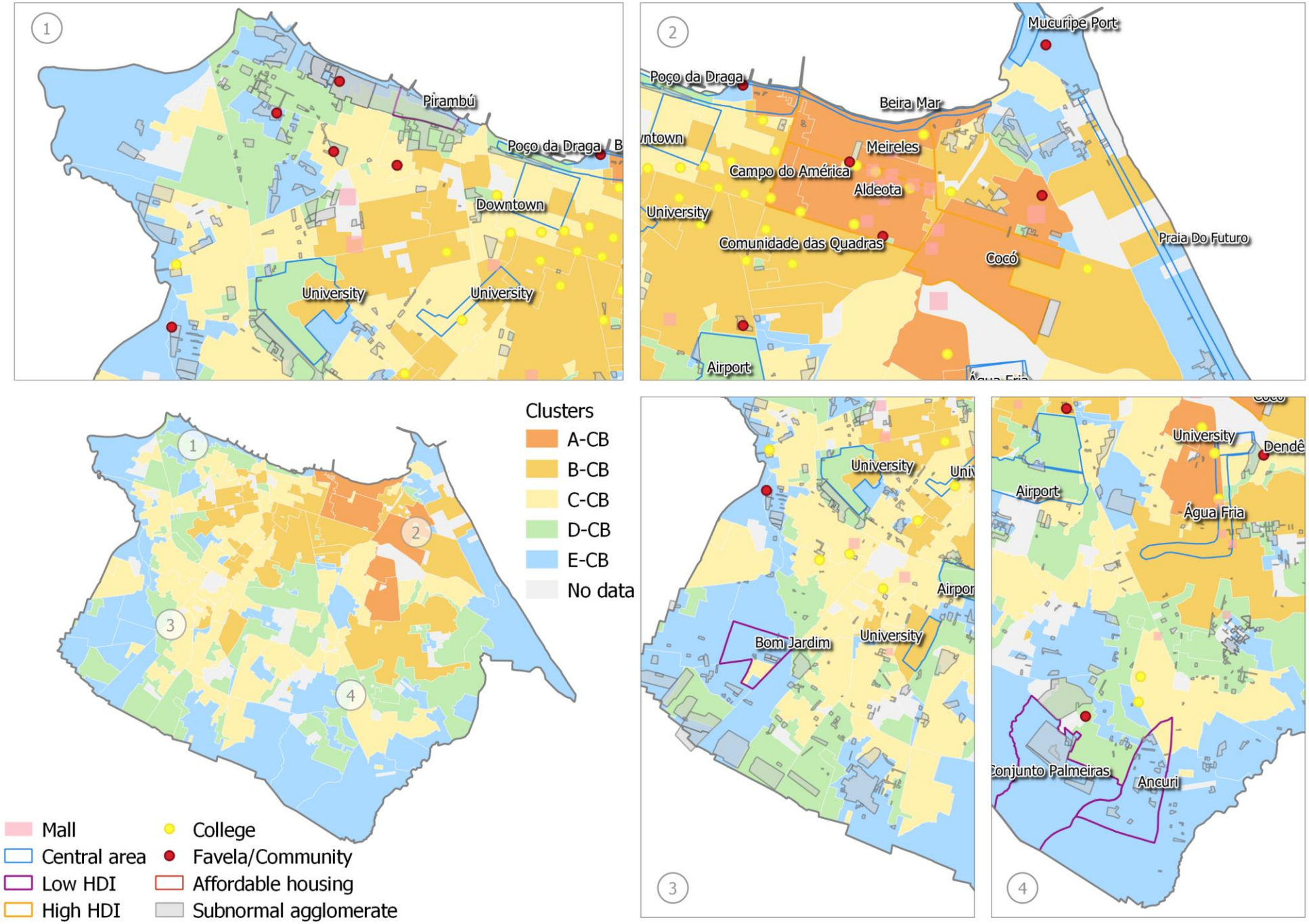


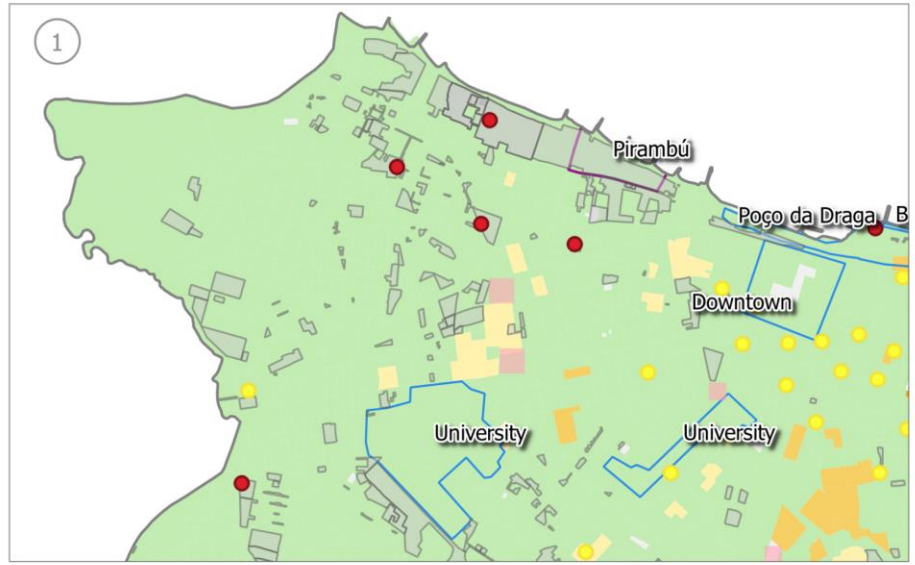


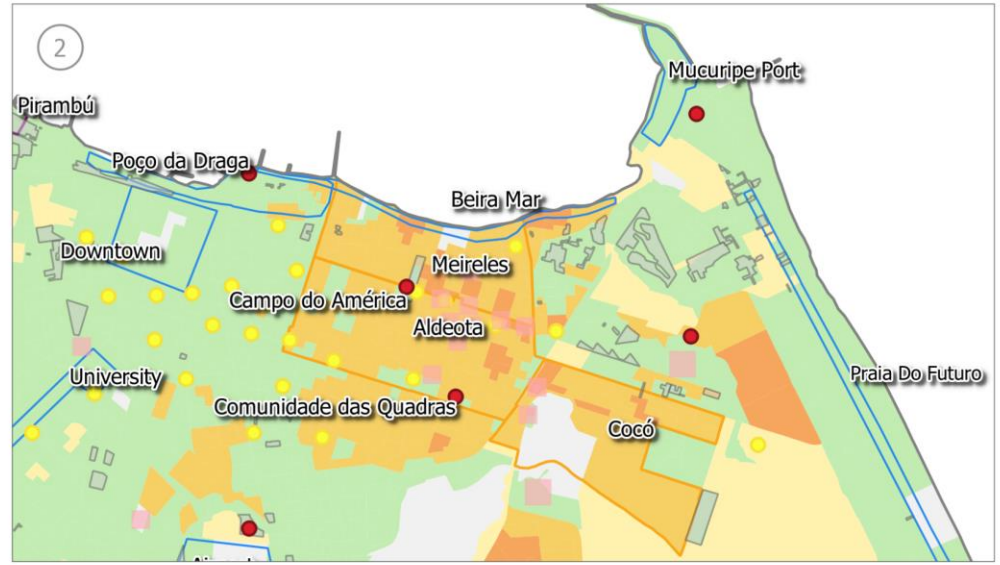
Setor Censitário

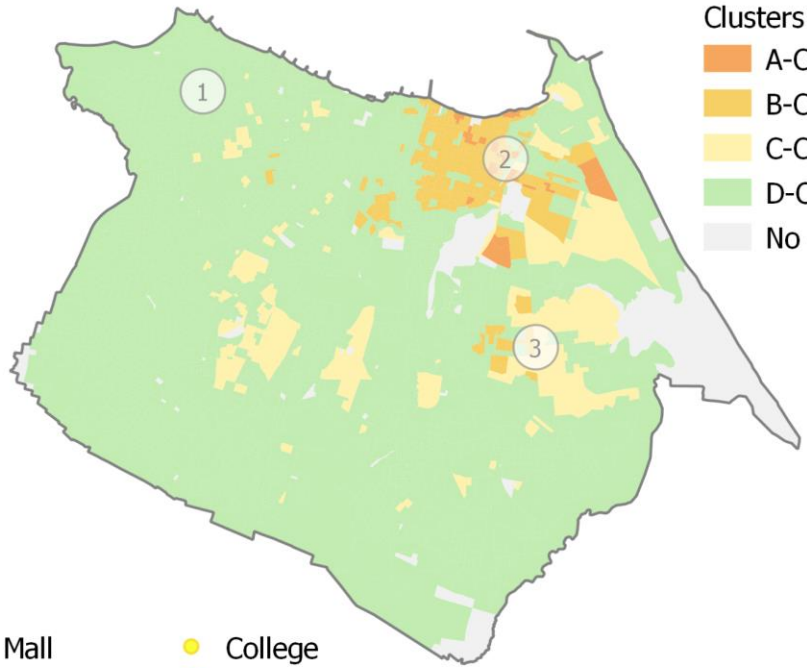




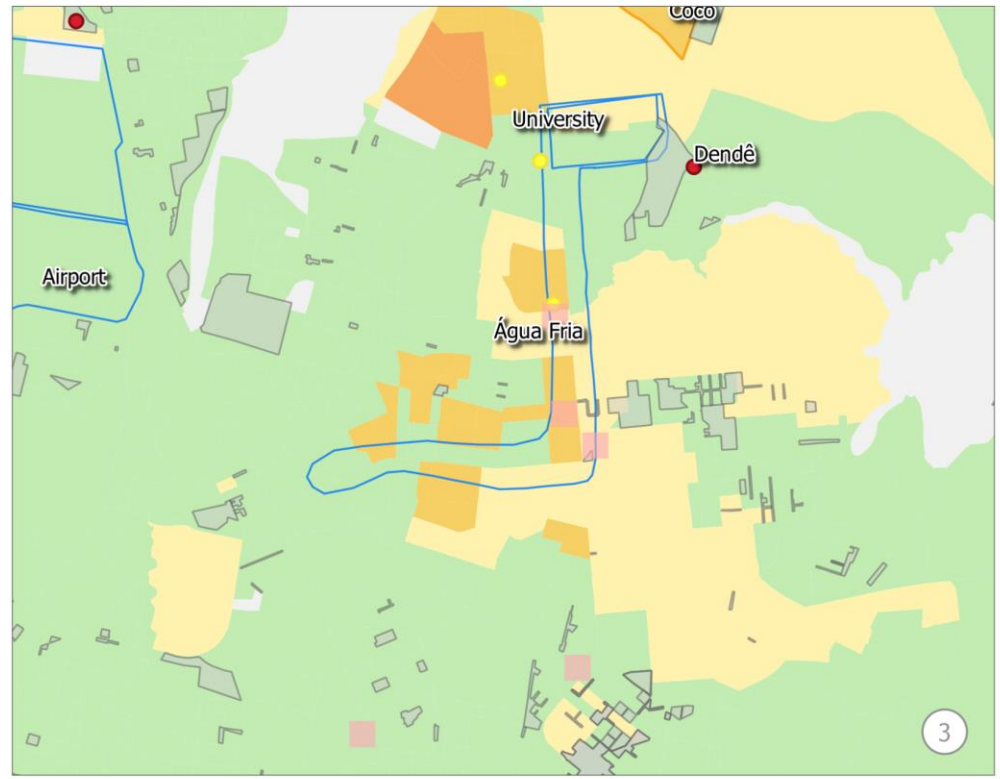

1
Pirambú
Poço da Draga
Downtown
University
University
2
Mucuripe Port
Poço da Draga
Beira Mar
Meireles
Campo do América
Aldeota
University
Comunidade das Quadras
Cocó
Praia Do Futuro
Airport
1
2
3
4
Clusters
A-CB
B-CB
C-CB
D-CB
E-CB
No data
3
University
Bom Jardim
University
4
University
Dendê
Airport
Água Fria
Conjunto Palmeiras
Ancuri
Mall
Central area
Low HDI
High HDI
College
Favela/Community
Affordable housing
Subnormal agglomerate



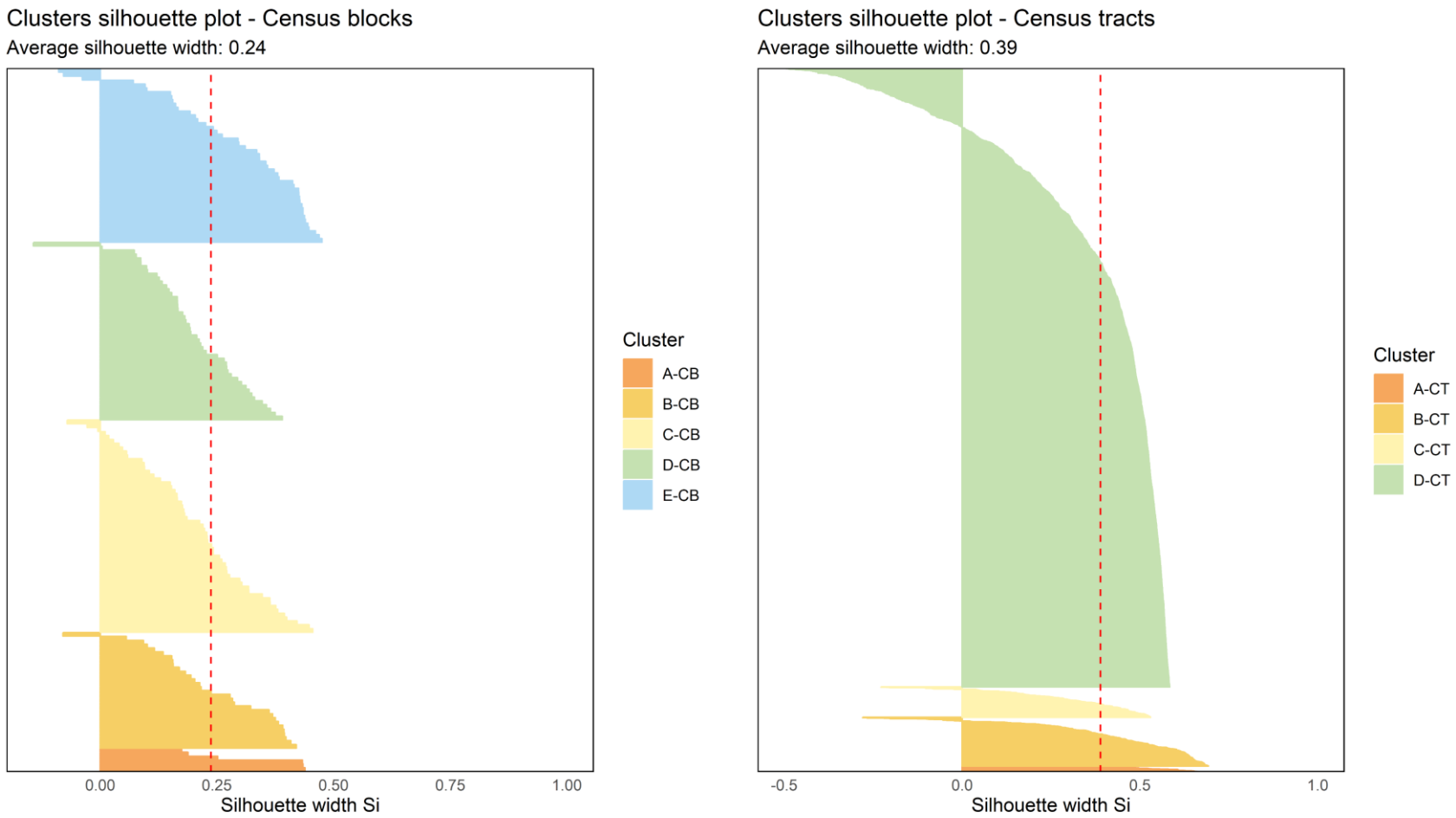




Clusters
- A-CT
- B-CT
- C-CT
- D-CT
- No data



- Mall
- Central area
- Low HDI
- High HDI
- College
- Favela/Community
- Affordable housing
- Subnormal agglomerate


Clusters silhouette plot - Census blocks
Average silhouette width: 0.24
0.00
0.25
0.50
0.75
1.00
Silhouette width Si
Cluster
A-CB
B-CB
C-CB
D-CB
E-CB
Clusters silhouette plot - Census tracts
Average silhouette width: 0.39
-0.5
0.0
0.5
1.0
Silhouette width Si
Cluster
A-CT
B-CT
C-CT
D-CT

PROJETO DE CIÊNCIA DE DADOS

FORMULAÇÃO DO PROBLEMA

COLETA DE DADOS

ANÁLISE EXPLORATÓRIA

SELEÇÃO E LIMPEZA DOS DADOS

MODELAGEM

AVALIAÇÃO

PREPARAÇÃO DE DADOS

DÚVIDAS?

MATERIAL SOBRE O ARTIGO

taismarianc@gmail.com

Links com materiais de estudo:

Materiais de estudo sobre R
https://beatrizmilz.com/project/materiais-estudo-r/

Artigo sobre o trabalho apresentado
https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29WR.1943-5452.0001310

Minicurso sobre aplicações de aprendizado de máquina em Recursos Hídricos
https://taiscarvalho.github.io/ia-rec_hidricos/

Livro sobre aprendizado de máquina
https://hastie.su.domains/Papers/ESLII.pdf

Artigo interessante com uma aplicação de aprendizado de máquina para avaliar impactos da seca
https://iopscience.iop.org/article/10.1088/1748-9326/aba4ca

Artigo sobre demanda hídrica
https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2019WR024897

Dica: acessar os artigos com o portal de periódicos da CAPES