

## A criminalidade tomou conta da cidade: o cinturão de insegurança em Fortaleza, Ceará

Marcos Renan Vasconcelos Magalhães<sup>1</sup>  | Felipe Pinto da Silva<sup>2</sup>  | Aguiamar Mendonça Neto<sup>3</sup>  | Marcelo Davi Santos<sup>4</sup>  | Christiano Modesto Penna<sup>5</sup> 

<sup>1</sup> Doutor, UFC/CAEN. E-mail: marcosrenan@caen.ufc.br

<sup>2</sup> Mestre, UNICAMP/IE. E-mail: dasilva.felipe@outlook.com

<sup>3</sup> Mestre, SEFAZ/CE. E-mail: aguimar@gmail.com

<sup>4</sup> Doutor, UFC/CAEN. E-mail: davisantos@caen.ufc.br

<sup>5</sup> Doutor, UFC/CAEN. E-mail: cmp@caen.ufc.br

### RESUMO

Este estudo investiga a criminalidade em Fortaleza, capital do Ceará, utilizando dados de boletins de ocorrências policiais para realizar uma análise de *clusterização*, que identificou a existência de uma espécie de cinturão de insegurança na cidade. O cinturão de insegurança é formado por bairros que apresentam altos índices de criminalidade, como homicídios, roubos e furtos. Os bairros que compõem essa zona estão localizados em regiões periféricas da cidade, com baixa renda e infraestrutura precária. Em seguida, a Análise de Componentes Principais foi empregada para reduzir a dimensionalidade dos dados, revelando que os principais fatores relacionados à criminalidade são pobreza e baixa educação. A complexidade do problema requer políticas públicas integradas para resolvê-lo, devendo abordar os fatores supracitados. Os resultados oferecem uma visão detalhada das dinâmicas criminais em Fortaleza, revelando que áreas mais afluentes são frequentemente circundadas por regiões com altos níveis de criminalidade, formando um cinturão de insegurança.

### PALAVRAS-CHAVE

Fortaleza, Economia do Crime, *Clusterização*, Análise dos Componentes Principais

### Crime has taken over the city: The belt of insecurity in Fortaleza, Ceará

### ABSTRACT

This study investigates criminality in Fortaleza, the capital of Ceará, using police incident reports to conduct a clustering analysis, which identified the existence of a kind of insecurity belt in the city. This insecurity belt is composed of neighborhoods that exhibit high crime rates, such as homicides, robberies, and thefts. The neighborhoods forming this zone are located in peripheral regions of the city, characterized by low income and poor infrastructure. Subsequently, Principal Component Analysis was employed to reduce the data dimensionality, revealing that the main factors related to criminality are poverty and low education. The complexity of the problem requires integrated public policies to address it, focusing on the aforementioned factors. The results provide a detailed view of the criminal dynamics in Fortaleza, revealing that more affluent areas are often surrounded by regions with high crime levels, forming an insecurity belt.

### KEYWORDS

Fortaleza, Crime Economy, Clustering, Principal Component Analysis

### CLASSIFICAÇÃO JEL

C21, C38, O11

## 1. Introdução

O 15º Anuário Brasileiro de Segurança Pública divulgou que, durante o ano de 2020, o estado do Ceará registrou a mais alta taxa de Mortes Violentas Intencionais (MVIs) no país, alcançando 45,2 ocorrências por 100 mil habitantes. Notavelmente, o Ceará apresentou também a maior incidência de homicídios infantis no Brasil, observando-se um crescimento expressivo no número de assassinatos de 2019 para 2020, conforme reportado pelo Anuário Brasileiro de Segurança Pública (Fórum Brasileiro de Segurança Pública, 2015). Tal ascensão nos índices de homicídios pode ser associada a múltiplos fatores, destacando-se o motim da Polícia Militar, que modificou a estrutura local de criminalidade e interrompeu a implementação de políticas públicas vigentes, conforme relatado no Anuário Brasileiro de Segurança Pública (Fórum Brasileiro de Segurança Pública, 2015).

Fortaleza, a capital cearense, destaca-se tanto por seu apelo turístico para visitantes nacionais e internacionais quanto por sua movimentação de passageiros, registrando 2.268.613 pessoas pelo sistema aeroportuário e 2.267.266 pelo rodoviário (Secretaria do Turismo do Estado do Ceará, 2023). Economicamente, a cidade lidera o Nordeste com o maior Produto Interno Bruto (PIB), alcançando R\$ 73,43 bilhões em 2021, o que representou 37,68% do PIB do Ceará, segundo dados do IBGE de 2023. No entanto, além de seu potencial econômico e turístico, Fortaleza enfrenta sérios desafios na esfera da segurança, posicionando-se como a quinta cidade mais violenta do Brasil em termos de homicídios dolosos por 100 mil habitantes, conforme o Anuário Brasileiro de Segurança Pública de 2022, e exibindo elevadas taxas de roubos, furtos e tráfico de drogas<sup>1</sup>.

A problemática da criminalidade em Fortaleza está intensamente marcada pela rivalidade entre facções criminosas, que disputam territórios e exercem influência por meio de confrontos violentos, incluindo tiroteios e assassinatos, comprometendo a segurança na região. O surgimento e fortalecimento dessas facções no Ceará estão atrelados às mudanças sociais e nas dinâmicas criminais observadas na capital e em sua área metropolitana, segundo Paiva (2019), Melo e Paiva (2021) e Moraes e Siqueira (2019). Antes de 2015, as facções criminosas não eram uma presença constante nas prisões e periferias cearenses, realidade que mudou com a emergência dos Guardiões do Estado (GDE) em 2016. Facções de grande porte como o Primeiro Comando da Capital (PCC) e o Comando Vermelho (CV) passaram a exercer influência significativa, com alianças e rivalidades marcantes, especialmente entre o PCC, aliado ao GDE, e o CV. A Família do Norte (FDN) também se destacou, apoiando o CV e intensificando a disputa territorial e prisional com o GDE e o PCC. Desde então, o Ceará tornou-se cenário de uma complexa disputa entre essas facções, envolvendo conflitos, alianças e rupturas, que complicam o panorama da segurança pública na área (Paiva, 2019;

<sup>1</sup>Para se ter uma ideia, em 2021, o Brasil contava com uma taxa de homicídios (Mortes Violentas Intencionais - MVI) de 23,9 por 100 mil habitantes, enquanto essa taxa em Fortaleza foi de 38,4 por 100 mil habitantes, ou seja, uma taxa 60% maior do que a brasileira, segundo o Anuário Brasileiro de Segurança Pública (Fórum Brasileiro de Segurança Pública, 2022).

Melo e Paiva, 2021; Moraes e Siqueira, 2019).

Autoridades têm adotado estratégias como desmantelamento de organizações criminosas e aumento do policiamento em áreas de risco para enfrentar o problema, Fórum Brasileiro de Segurança Pública (2022). Essa abordagem é adotada, particularmente, em metrópoles com elevados índices de criminalidade que demandam uma abordagem multidisciplinar. A análise da criminalidade e da segurança pública transcende a criminologia, abarcando contribuições de áreas como sociologia, geografia e economia. O estudo da criminalidade no campo econômico ganhou ímpeto com o trabalho pioneiro de Becker, em 1968, que examinou o comportamento criminoso sob uma ótica microeconômica. Seguidamente, as análises econômicas do crime expandiram-se para incluir aspectos espaciais, enriquecendo as investigações para além dos limites tradicionais da economia (Anselin et al., 2000; Andresen, 2006; Scott, 1986; Balaguer-Coll et al., 2019).

A incidência do crime afeta diretamente a qualidade de vida urbana, aumentando a sensação de insegurança devido à ameaça ou ocorrência de delitos como roubos (Costa, 2018). Esse impacto negativo se estende a pessoas, gerando exclusão social e perda de produtividade (Goldberg et al., 2014), assim como a empresas, que sofrem com a diminuição dos investimentos e do crescimento econômico (Loureiro e Silva, 2010).

A violência também pode estar correlacionada com o aumento da economia informal, resultado de fatores socioeconômicos complexos como falta de empregos formais, desigualdade social e falta de oportunidades. Essa realidade contribui para a percepção de atividades criminosas como alternativas factíveis, especialmente em áreas com baixo investimento público. A economia informal, por sua vez, opera com menor produtividade, impactando negativamente a produção agregada (Arbex et al., 2023).

Diante desse cenário complexo, este estudo se destaca devido à importância do aprimoramento de técnicas analíticas que transcendem a simples coleta de estatísticas oficiais, visando detectar padrões robustos e consistentes nas variáveis relacionadas a crimes (Hart, 2020). Isso contribui significativamente para a uniformidade e a acurácia dos dados (Fávero e Belfiore, 2021). Além disso, é crucial preencher lacunas de estudos que não têm enfoque em uma perspectiva mais localizada da violência urbana. Ademais, é imperativo fornecer análises detalhadas e diagnósticos que auxiliem na formulação de políticas públicas focadas em microrregiões específicas, como as de Fortaleza, conforme enfatizado por Santos (2019). Este autor aponta a necessidade de expandir as estratégias de combate ao crime e à violência, considerando não apenas as esferas federal e estadual, mas também a importância de intervenções ao nível dos bairros.

A hipótese desta pesquisa é que, teoricamente, bairros com maiores rendas *per capita* em Fortaleza, apesar de estarem próximos a pontos turísticos e, consequentemente, terem maior fluxo de pessoas, não apresentam uma predominância de delitos.

Isso se deve, em parte, às políticas públicas municipais que direcionam investimentos e esforços de segurança para áreas de interesse turístico, diminuindo a incidência de crimes nesses locais. Assim, espera-se identificar em Fortaleza um padrão de “cinturão de insegurança”, onde os bairros mais afluentes são rodeados pelos mais violentos, conforme observado por Feitoza e Costa (2019).

Assim, o objetivo deste estudo é analisar a distribuição da violência urbana em Fortaleza em 2021, com base em fatores sociais e espaciais, utilizando dados da Secretaria da Segurança Pública e Defesa Social do Estado do Ceará (2021). A metodologia adotada é exploratória, focando na identificação de grupos de bairros com padrões semelhantes de roubos de pertences, através de *clustering* não supervisionado. As contribuições incluem *insights* para políticas públicas personalizadas, uma abordagem alternativa para análise de dados criminais e a identificação dos bairros com maior incidência de violência, utilizando Análise dos Componentes Principais (PCA).

O estudo é organizado da seguinte forma: após esta introdução, a seção 2 revisa a literatura existente e as características da criminalidade em Fortaleza, analisando a conexão entre os crimes (homicídios, roubos, furtos, apreensões de drogas e de armas) e fatores socioeconômicos, como o Índice de Desenvolvimento Humano (IDH). A seção 3 descreve o uso do algoritmo k-means para *clusterização*, os métodos para determinar o número ideal de *clusters* e a aplicação da PCA. Os resultados são apresentados na seção 4, e as conclusões e recomendações de políticas são discutidas na última seção.<sup>2</sup>

## 2. Criminalidade em Fortaleza e Base de Dados

A violência nas áreas urbanas da cidade de Fortaleza tornou-se algo comum na vida diária dos seus cidadãos. Esta seção se destina a apresentar e explorar as bases de dados que serão utilizadas pelo algoritmo não supervisionado<sup>3</sup> e da análise de componentes principais (PCA), que visa reduzir a dimensionalidade dos dados, preservando ao máximo a informação original. A Tabela 1 a seguir versa sobre variáveis relacionadas ao crime (homicídios, roubos, furtos e apreensão de drogas e de armas de fogo) no ano de 2021 por bairros, além de trazer informações do Índice de Desenvolvimento Humano (IDH) segundo as dimensões/categorias, elaborado pela Secretaria Municipal de Desenvolvimento Econômico de Fortaleza (Secretaria de Desenvolvimento Econômico de Fortaleza, 2015) com base nos dados do Censo de 2010 do Instituto Brasileiro de Geografia e Estatística (Instituto Brasileiro de Geografia e Estatística, 2023).

Com relação aos dados de criminalidade, esses foram obtidos junto à SSPDS/CE que divulga mensalmente informações oficiais consolidadas sobre diversos tipos de crime, classificados por Áreas Integradas de Segurança (AIS). Neste estudo, foi utili-

<sup>2</sup>“A criminalidade tomou conta da cidade”: título em alusão à letra da música intitulada “Cachimbo da Paz” de composição de Gabriel Pensador, Lulu Santos e Marcello Mansur (DJ Memê).

<sup>3</sup>Algoritmos não supervisionado referem-se a um conjunto de técnicas de aprendizado de máquina que buscam identificar padrões e similaridades nos dados sem rótulos predefinidos (Hastie et al., 2009).

zado um painel dos 113 bairros híbridos<sup>4</sup>, contendo informações referentes aos crimes de roubos<sup>5</sup>, furtos, apreensões de drogas/entorpecentes<sup>6</sup> e apreensões de armas de fogo registrados oficialmente nos limites territoriais da capital de Fortaleza para o período de 2021. Por outro lado, as informações de óbitos por homicídios foram extraídas do Sistema de Informação sobre Mortalidade – SIM/DATASUS da Prefeitura Municipal de Fortaleza (PMF).

Os dados analisados, abrangendo os quartis de IDH-Renda dos 113 bairros de Fortaleza, revelam que áreas com IDH-Renda inferior (Q1) exibem uma incidência consideravelmente mais elevada de homicídios, roubos e apreensões de drogas e armas de fogo. Essa observação sugere uma correlação entre menor renda e elevada atividade criminosa, reforçando a teoria de que regiões menos favorecidas podem ser mais susceptíveis à criminalidade devido a condições socioeconômicas adversas. Curiosamente, os furtos constituem uma exceção, apresentando maior ocorrência em bairros com renda *per capita* superior, tanto em números absolutos quanto em taxas por 10 mil habitantes.

Esta análise é corroborada pelas pesquisas de Beato Filho (1998), Beato Filho et al. (2001) e Batella e Diniz (2010), que discutem a interação entre o desenvolvimento humano e a criminalidade, propondo que os crimes violentos contra o patrimônio tendem a estar mais associados aos níveis de desenvolvimento econômico, refletidos pelo IDH municipal. Eles argumentam que bairros mais prósperos ou com intensa atividade comercial, como observado em Fortaleza, possuem características que potencializam essa correlação, incluindo melhores condições econômicas, densidade populacional e comercial elevada e enfraquecimento dos mecanismos de controle social, criando assim mais oportunidades para a ocorrência de delitos.

Beato Filho (1998) e Batella e Diniz (2010) também abordam a relação entre riqueza e crimes contra o patrimônio, contrapondo a ideia de que a pobreza seja o principal fator determinante para o crime, sugerindo que, na realidade, é a riqueza que gera mais oportunidades para atividades criminosas. Eles concluem que ambientes mais afluentes oferecem mais alvos atraentes e lucrativos para os criminosos, além de apresentarem mecanismos de controle social e vigilância mais fracos.

Adicionalmente, observações paralelas dos quartis de IDH-Educação e IDH-Longevidade demonstram que, à medida que o IDH-Educação aumenta, os índices de homicídios, roubos e furtos tendem a diminuir, sugerindo que níveis mais altos de educação po-

---

<sup>4</sup>Vale apenas ressaltar que as análises que serão realizadas têm como foco uma abordagem que leva em consideração os transbordamentos espaciais entre os bairros, não o georreferenciamento dos delitos supracitados.

<sup>5</sup>Crimes violentos contra o patrimônio (CVP) consistem na soma de todas as ocorrências de roubo, exceto o crime de roubo seguido de morte (latrocínio), dado que já é contabilizado na categoria de CVLI. Até meados de 2019, os crimes de CVP eram divididos em dois grupos: (i) CVP-1: roubo à pessoa, roubo de documentos e outros roubos que não estão incluídos no CVP-2. (ii) CVP 2: roubo a residência, roubo com restrição de liberdade da vítima, roubo de cargas e roubo de veículos.

<sup>6</sup>A partir de 2019, houve mudanças na descrição da tabela para “Apreensão de derivados da *Cannabis* (Maconha e Haxixe)”. (3) Inclui Maconha e Haxixe.

**Tabela 1.** Quartis do IDH de Renda, Longevidade e Educação e média de variáveis de delitos por bairros de Fortaleza, em 2021.

Quartis	IDH			Nº absolutos					Taxa por 10 mil hab.				
	Renda	Long.	Edu.	H	R	F	AD	AA	H	R	F	AD	AA
Q1	0,04	0,26	0,91	16	313	136	21	22	7	140	63,1	18	10
Q2	0,09	0,36	0,95	9	265	177	47	17	5	122	90	38	8,7
Q3	0,15	0,46	0,96	8	250	219	11	13	4	126	109	4,7	5,7
Q4	0,46	0,62	0,98	2	257	323	5,7	6,5	2	185	213	8,7	4,6

Fonte: Instituto Brasileiro de Geografia e Estatística (2010); Secretaria de Desenvolvimento Econômico de Fortaleza (2015); Secretaria da Segurança Pública e Defesa Social do Estado do Ceará (2021).

Nota: H – Óbitos por Homicídios; R – Nº de ocorrências de Roubos; F – Nº de ocorrências de Furtos; AD – Massa apreendida de Drogas/Entorpecentes; AA – Nº de armas de fogo apreendidas.

dem levar a uma maior conscientização das consequências legais e sociais do crime, contribuindo assim para a redução da criminalidade.

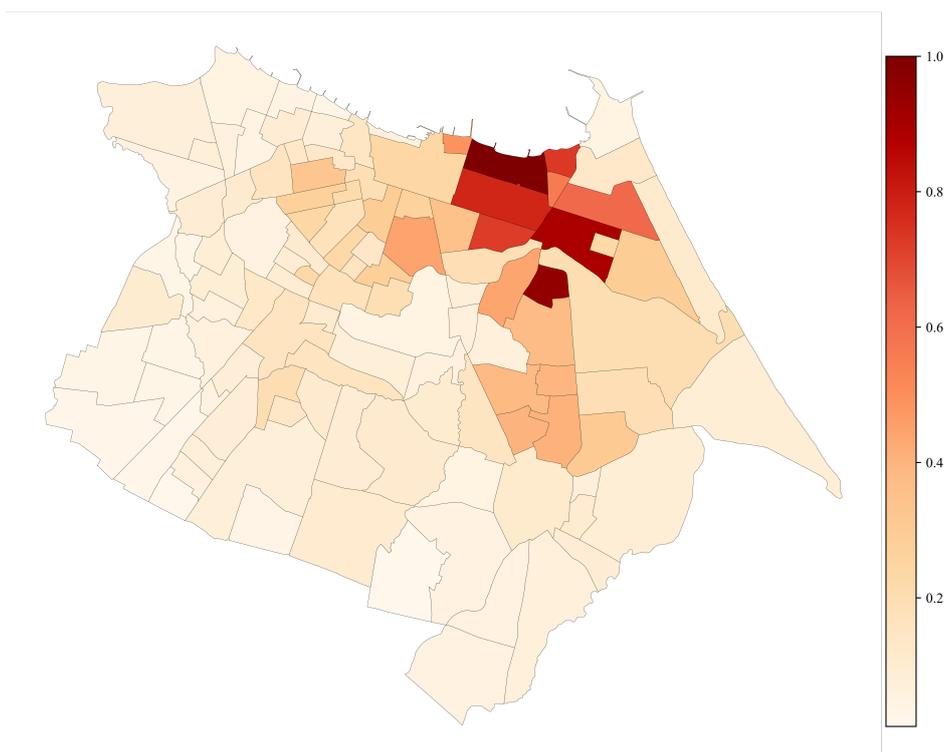
Além disso, uma possível relação inversa foi notada entre as apreensões de drogas e armas de fogo e o IDH-Renda; bairros com menor IDH-Renda mostram maiores taxas de apreensões, indicando que áreas com maior vulnerabilidade socioeconômica podem estar mais propensas ao tráfico de drogas e à posse ilegal de armas.

Em síntese, a análise dos dados dos bairros de Fortaleza, correlacionando o IDH-Renda com as variáveis criminais, destaca que fatores socioeconômicos, como renda e educação, são significativos na determinação das taxas de criminalidade.

A Figura 1 ilustra o IDH-Renda nos 113 bairros de Fortaleza, com os bairros de maior IDH-Renda, como Meireles, Guararapes, Cocó e Aldeota, predominantemente situados na região leste, indicando um *cluster* de desenvolvimento econômico e qualidade de vida superior nessa área. Essa concentração sugere não apenas o acesso a melhores oportunidades econômicas e serviços, mas também reflete as disparidades significativas em termos de emprego formal, educação de qualidade, infraestrutura e investimentos entre as diferentes regiões da cidade.

É de se esperar uma correlação negativa entre o Índice de Desenvolvimento Humano (IDH) e as taxas de homicídio, sugerindo uma possível conexão entre adversidades socioeconômicas e prevalência de violência. Apesar dos esforços contínuos para mitigar essa realidade, Fortaleza enfrenta desafios recorrentes, incluindo uma baixa taxa de responsabilização em casos de homicídio e uma tendência ao policiamento ostensivo, o que tende a engendrar em encarceramento massivo, ou seja, é possível que se direcione mais pequenos traficantes e ladrões do que homicidas à prisão.

De outro lado, a Figura 2 exibe as estatísticas de ocorrências criminais por bairro em Fortaleza para o ano de 2021, detalhando roubos, furtos, homicídios e apreensões de armas de fogo. As taxas de criminalidade variam significativamente entre os bairros, evidenciando uma distribuição desigual da criminalidade. Enquanto alguns bairros registram baixos índices em todas as categorias, sugerindo um ambiente mais

**Figura 1.** Distribuição do IDH de Renda segundo os bairros de Fortaleza.

Fonte: Secretaria de Desenvolvimento Econômico de Fortaleza (2015).

seguro, outros como Jangurussu e Bom Jardim destacam-se por altas incidências de diversos tipos de crimes.

A análise dos dados não revela um padrão uniforme de criminalidade concentrada em uma única região, mas indica uma maior incidência de homicídios e apreensões de armas no sul/sudoeste da cidade. Quanto aos roubos, esses tendem a ocorrer ao redor do centro da cidade, com o bairro do Centro registrando o maior número de casos em 2021. De forma semelhante, esse bairro lidera as estatísticas de furtos, seguido por áreas mais abastadas como Aldeota e Meireles, indicando que os crimes contra o patrimônio frequentemente ocorrem em bairros de maior renda.

De acordo com Santos (2020), duas observações são pertinentes sobre o perfil criminal dos bairros. Primeiramente, ressalta-se a qualidade de vida, medida pelo IDH, dos bairros com alta densidade criminal. Um estudo da Secretaria Municipal de Desenvolvimento Econômico (Secretaria de Desenvolvimento Econômico de Fortaleza, 2015) mostra que bairros com IDH elevado a médio, como Meireles, Aldeota e Centro, apresentaram uma média anual de 2.117 crimes contra o patrimônio entre 2010 e 2014, desafiando a noção de que bairros mais abastados são necessariamente mais seguros. Por outro lado, bairros periféricos com menor IDH, como Guajeru e Sabiaguaba, representaram apenas 0,48% das ocorrências de crimes contra o patrimônio no mesmo período, sugerindo uma distribuição complexa e multifacetada da criminalidade na cidade.

Santos (2020) destaca que nas áreas periféricas de Fortaleza, as forças policiais enfrentam desafios significativos impostos pelas facções criminosas, particularmente no contexto dos conjuntos habitacionais do programa “Minha Casa, Minha Vida”. Nestes locais, as facções frequentemente expulsam os moradores de suas residências e impõem o chamado “Tribunal do Crime”, conhecido como “Lei do vai pro saco”, penalizando severamente delitos como roubo ou furto dentro da comunidade. A dificuldade em representar visualmente as apreensões de drogas devido à dominância do bairro Demócrito Rocha nessa categoria resultou na ausência de um mapa que mostrasse a diversidade da distribuição desses crimes entre os bairros.

Para complementar os dados apresentados na Figura 2, a Tabela 2 reúne os 10 bairros com maiores índices de criminalidade por tipo de delito, revelando que as incidências criminais não estão restritas a uma região específica, mas dispersas por toda a cidade, incluindo as regiões leste e oeste. A presença de altas incidências de diversos crimes em bairros como Jangurussu e Bom Jardim sugere a existência de atividades criminosas complexas e organizadas, com Jangurussu liderando em homicídios e apreensões de armas de fogo em 2021.

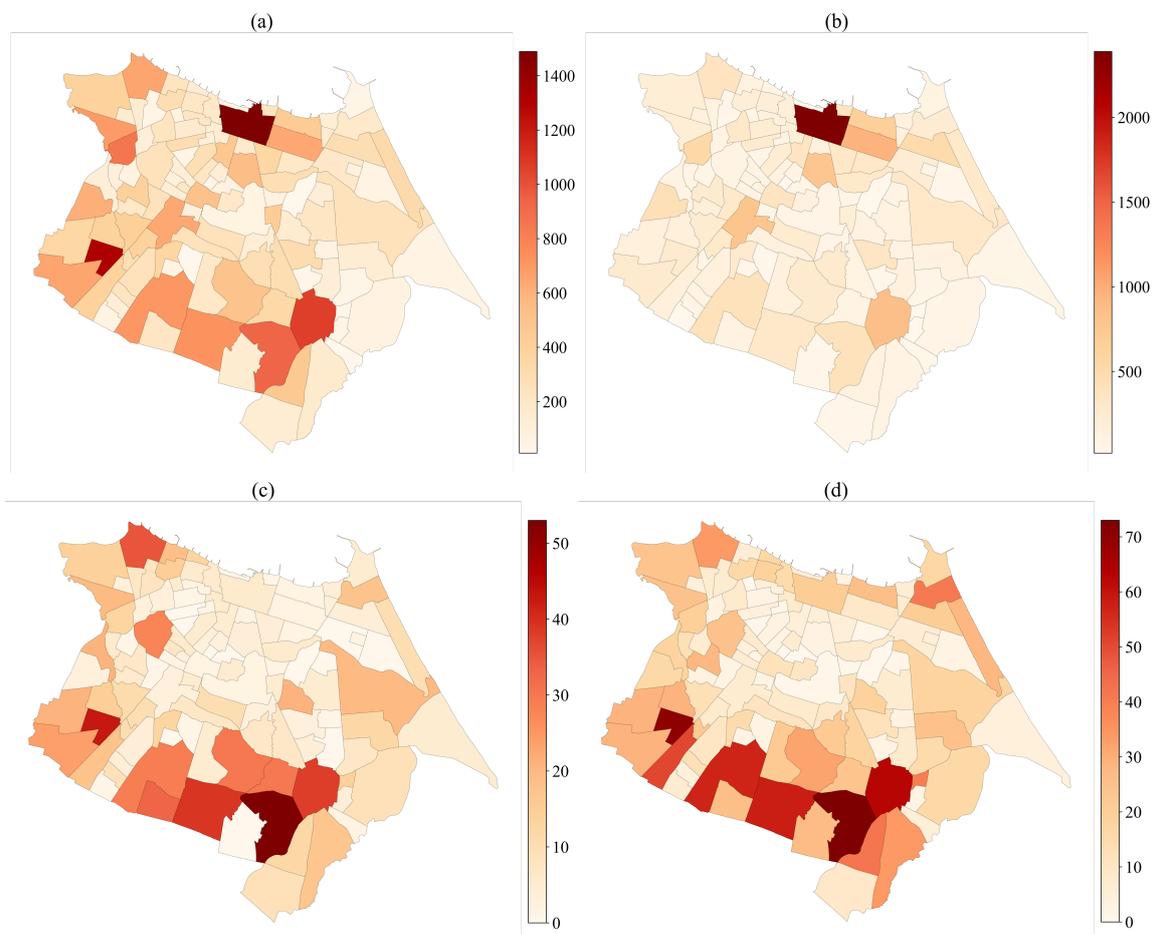
Por outro lado, bairros como Demócrito Rocha, Sapiranga/Coité e Henrique Jorge demonstram altas taxas de apreensões de drogas, apontando para uma significativa atividade de tráfico. O Centro, área de intensa movimentação econômica e comercial, registrou o maior número de roubos e furtos. Ademais, é notório que bairros de alta renda como Aldeota e Meireles também figuram na tabela devido à elevada ocorrência de roubos e furtos, possivelmente devido ao maior valor econômico dos bens na área, como discutido por Beato Filho (1998), Beato Filho et al. (2001) e Batella e Diniz (2010).

Guerra (2016) analisa a concentração geográfica de comportamentos transgressores, alinhando-se com teorias sociológicas clássicas que enfatizam a influência do ambiente urbano na prevalência criminal. Segundo essa perspectiva, altos índices de delinquência estão relacionados a comunidades heterogêneas, de baixa renda e educação, marcadas pela densa atividade industrial e comercial. Essa dinâmica é evidente em bairros da Região Grande Aldeota (Aldeota, Dionísio Torres, Meireles, Papicu, Varjota e Praia de Iracema) e no Centro, ilustrando a complexidade da distribuição espacial do crime na cidade.

A Tabela 2 revela uma concentração de roubos e furtos em bairros localizados no centro da cidade, como Centro, Aldeota, Meireles e Praia de Iracema, corroborando a hipótese da maior vulnerabilidade de áreas com maior fluxo de pessoas e atividades comerciais. Por outro lado, os homicídios e as apreensões de armas de fogo se concentram em bairros periféricos, como Jangurussu, Bom Jardim, Mondubim e Granja Lisboa, indicando uma correlação entre violência letal e áreas de menor desenvolvimento socioeconômico.

Contudo, a análise da criminalidade em Fortaleza revela um cenário complexo e

**Figura 2.** Distribuição de Roubos, Furtos, Homicídios e Apreensão de Armas de Fogo por bairros de Fortaleza, em 2021.



Fonte: Secretaria da Segurança Pública e Defesa Social do Estado do Ceará (2021).

Nota: (a) Roubos; (b) Furtos; (c) Homicídios; (d) Apreensão de Armas de Fogo.

multifacetado. A distribuição espacial dos crimes não segue um padrão linear, sendo influenciada por uma variedade de fatores socioeconômicos, demográficos e geográficos. Bairros com IDH mais elevado, como Meireles e Aldeota, não estão imunes à criminalidade, especialmente no que diz respeito a roubos e furtos, possivelmente devido à maior concentração de bens de valor e oportunidades para o crime. Por outro lado, bairros periféricos com menor IDH, como Jangurussu e Bom Jardim, apresentam maiores índices de homicídios e apreensões de armas de fogo, evidenciando a complexa relação entre pobreza, desigualdade social e violência.

### 3. Metodologia

#### 3.1 Estratégia empírica

A Análise dos Componentes Principais (PCA) e o k-means são técnicas estatísticas que atendem a objetivos diferentes por meio de abordagens metodológicas distintas. Ambas facilitam a redução da complexidade dos dados, mas enquanto o PCA identifica

**Tabela 2.** Os 10 bairros com maiores incidências em cada tipo de crime em ordem alfabética, em 2021.

Bairros	Roubo	Furto	Homicídio	Apreensão	
				Armas de Fogo	Drogas (kg)
Aldeota <sup>2</sup>	641	958 <sup>2</sup>	2	9	0,89
Ancuri	465	117	12	42	42,00
Antônio Bezerra	855	538	13	20	10,49
Barra do Ceará	654	370	35	34	9,54
Barroso	335	90	30	24	24,00
Bom Jardim <sup>2,2,2</sup>	1.320 <sup>2</sup>	269	43 <sup>2</sup>	70 <sup>2</sup>	89,54
Canindezinho	397	157	18	51	51,00
Centro <sup>1,1</sup>	1.489 <sup>1</sup>	2.388 <sup>1</sup>	6	21	5,89
Conjunto Ceará	611	427	3	16	12,96
Curió	33	26	14	40	40,00
Demócrito Rocha <sup>1</sup>	127	98	2	6	799,18 <sup>1</sup>
Fátima	551	768	4	3	1,75
Guajerú <sup>3</sup>	1.320 <sup>3</sup>	28	4	2	100,38
Henrique Jorge <sup>3</sup>	403	266	7	28	157,52 <sup>3</sup>
Jangurussu <sup>1,1</sup>	932	398	53 <sup>1</sup>	73 <sup>1</sup>	123,26
Jardim Iracema	145	122	7	7	79,29
Lagoa Sapiranga (Coité) <sup>2</sup>	2	139	128	25	176,35 <sup>2</sup>
Meireles	449	663	2	26	2,21
Messejana <sup>3,3</sup>	1.064	860 <sup>3</sup>	38	63 <sup>3</sup>	9,20
Mondubim	711	413	29	57	57,00
Papicu	290	513	4	21	10,36
Parangaba	644	789	3	7	4,69
Parque Iracema	33	41	4	2	54,14
Passaré	507	264	30	32	32,00
Paupina	175	100	17	34	125,04
Pici	250	86	28	25	25,00
Planalto Ayrton Senna	219	116	33	26	26,00
Prefeito José Walter <sup>3</sup>	725	308	39 <sup>3</sup>	58	7,58
Quintino Cunha	694	204	20	25	6,06
Siqueira	654	271	24	29	3,88
Vicente Pinzón	187	235	18	41	41,00

Fonte: Elaboração com base nos dados da Secretaria da Segurança Pública e Defesa Social do Estado do Ceará (2021).

Nota: os números em sobrescrito indicam a posição do bairro no *ranking* de cada tipo de crime (1º, 2º ou 3º).

as direções que maximizam a variância nos dados, o k-means organiza os dados em  $k$  grupos, buscando representar cada agrupamento de forma concisa.

O PCA é uma técnica de transformação linear que identifica as direções, ou componentes principais, em que a variação dos dados é mais significativa. Seu propósito é diminuir a dimensionalidade dos dados enquanto retém a maior quantidade possível de informação, principalmente a variância (Jolliffe, 2002). Os primeiros componentes principais são aqueles que capturam a maior parte da variabilidade nos dados. Por outro lado, o k-means é um método de agrupamento que classifica os dados em  $k$  grupos distintos, assegurando que os dados dentro de cada grupo estejam mais próximos entre si do que em relação aos dados de outros grupos (Bishop, 2006). Esse método

foca em identificar padrões e agrupamentos nos dados, proporcionando centróides para cada *cluster* e atribuindo rótulos que indicam a qual *cluster* cada ponto de dado pertence.

A utilização do PCA pode revelar padrões nos dados que refletem atributos socioeconômicos, demográficos, infraestruturais ou criminais, diferenciando bairros entre si. Já o k-means é eficiente na criação de grupos de bairros com características semelhantes, facilitando uma análise segmentada e subsidiando políticas públicas direcionadas. A combinação dessas técnicas demonstra uma interação eficaz. Anaraki et al. (2021) descrevem duas abordagens para essa combinação: P-K e K-P.

A abordagem P-K inicia-se com o PCA para redução da dimensionalidade e limpeza de ruídos, seguido pela aplicação do k-means para *clusterização*. Essa abordagem torna a *clusterização* mais eficiente e melhora a precisão dos *clusters* ao se basear em dados simplificados e representativos. Já a abordagem K-P inicia-se com o k-means para *clusterização* dos bairros e, em seguida, aplica-se o PCA para aprofundar a análise de cada *cluster* e identificar as variáveis mais relevantes em sua formação. Devido à sua capacidade de aprofundar a análise dos *clusters*, a abordagem K-P foi a escolhida para o presente estudo, alinhada à metodologia utilizada por Wu (2023). É importante salientar que, em ambas as abordagens, a normalização dos dados é crucial para evitar distorções analíticas.

Wu (2023) exemplifica a aplicação da metodologia K-P em um estudo sobre gestão educacional universitária, combinando dispositivos *Internet of Things* (IoT) e técnicas de *big data* para aprimorar a eficiência e a segurança dos dados. No estudo, o algoritmo k-means é empregado inicialmente para segmentar os estudantes, com o PCA sendo posteriormente utilizado para analisar os dados coletados e explorar a correlação entre a frequência de uso da biblioteca e o desempenho acadêmico dos alunos.

Seguindo essa mesma estrutura, o presente trabalho aplicou a metodologia K-P em cinco etapas: 1) Coleta de dados relevantes; 2) Tratamento e padronização dos dados para garantir uniformidade e comparabilidade; 3) Implementação do algoritmo k-means para formar grupos de bairros com características semelhantes; 4) Aplicação do PCA para sintetizar e entender as principais variáveis que influenciam cada grupo; e 5) Análise das características dos *clusters*, proporcionando *insights* sobre as dinâmicas e tendências observadas. Essa abordagem metodológica permite não apenas uma segmentação eficaz dos dados, mas também facilita uma interpretação detalhada e integrada das variáveis que caracterizam cada grupo de bairros.

### 3.2 Algoritmo de *clusterização* k-means

O *clustering* é uma técnica sofisticada de mineração de dados multivariados que tem como propósito agrupar um conjunto de  $N$  observações provenientes de um banco de dados em  $k$  grupos distintos, conhecidos como *clusters*. Dentro de cada *cluster*, as observações possuem a máxima semelhança entre si, enquanto a similaridade entre

as observações de *clusters* diferentes é minimizada. Os algoritmos de agrupamento são empregados para revelar padrões em dados espaciais, mas sua aplicação em larga escala requer uma análise minuciosa das diversas técnicas de *clusterização* existentes (Jain e Dubes, 1988). O objetivo primordial é segmentar a amostra de dados originais em grupos consideravelmente homogêneos (Hair Junior et al., 2009).

No âmbito deste estudo, o algoritmo de agrupamento selecionado foi o k-means, empregando a métrica de distância euclidiana, amplamente utilizada na análise de agrupamentos (Fávero e Belfiore, 2021). A essência do método k-means reside em estabelecer *clusters* de modo a minimizar a variação total intragrupo ( $W$ ), também conhecida como variação interna do *cluster*. Diversos algoritmos k-means estão disponíveis, sendo o algoritmo padrão o Hartigan e Wong (1979). Neste algoritmo, a variação total interna do *cluster* é definida como a soma dos quadrados ( $SS$ ) das distâncias euclidianas entre as observações e o centróide correspondente, conforme demonstrado na Equação (1):

$$\sum_{k=1}^K W(C_k) = \sum_{x \in C_k} (x_i - \mu_k)^2 \quad (1)$$

em que  $x_i$  representa uma dada variável da unidade de análise  $i$  pertencente ao *cluster*  $C_k$ ;  $\mu_k$  é a média de uma dada variável  $x$  pertencente ao *cluster*  $C_k$ ; e  $K$  é o total de *clusters*.

Cada observação ( $x_i$ ) é atribuída a um *cluster* específico de forma que a soma das distâncias ao quadrado da observação até os centros de *cluster* atribuídos  $\mu_k$  seja mínima. Define-se a variação total dentro do *cluster* da seguinte forma:

$$\sum_{k=1}^K W(C_k) = \sum_{k=1}^K \sum_{x \in C_k} (x_i - \mu_k)^2 \quad (2)$$

Para determinar o número ótimo de *clusters* no algoritmo não supervisionado k-means, é necessário definir a quantidade de *clusters* desejada, o que a literatura chama de validação externa. A literatura oferece diversos métodos estatísticos para a determinação destes *clusters*, sendo os principais: o método do Cotovelo (Elbow), o método da Silhueta e o Gap Estatístico. Entre eles, o método do Cotovelo destaca-se como um dos mais populares e amplamente utilizado. Os três métodos foram efetuados para fins de comparação: o método do Cotovelo indicou a existência de três *clusters*, assim como o método da Silhueta. Apenas o método do Gap Estatístico indicou a existência de dois *clusters*. Os resultados dos três métodos são apresentados no Apêndice. Desta feita, conforme será apresentado, optou-se por três agrupamentos de bairros.

O método do Cotovelo envolve a realização repetida do algoritmo de *clusterização* para diferentes valores de  $k$  (número de *clusters*). Em seguida, é traçada uma re-

apresentação gráfica da variância explicada ou da soma dos quadrados intra *cluster* (ou do inglês *within-clusters sum-of-squares*, comumente abreviado para WCSS) em relação ao número de *clusters*. O objetivo é identificar o ponto na curva onde ocorre uma mudança abrupta, semelhante a um “cotovelo”. Esse ponto sugere o número ótimo de *clusters*, pois a adição de mais *clusters* após esse ponto não resulta em uma diminuição substancial do WCSS. A expressão do WCSS é dada a seguir:

$$WCSS = \sum d(P_1, C_1) + \sum d(P_2, C_2) + \sum d(P_3, C_3) + \dots + \sum d(P_i, C_k) \quad (3)$$

onde  $C_k$  representa o centro de cada *cluster*  $k$ , enquanto  $P_i$  representa os elementos que estão agrupados juntos dentro de um *cluster* específico  $k$ .

O método da silhueta, introduzido por Rousseeuw (1987), avalia a adequação de cada unidade de análise ao seu respectivo *cluster*. Uma alta largura de silhueta média indica um agrupamento sólido. O cálculo da silhueta média é realizado para várias observações e diferentes valores de  $k$  (número de *clusters*). O número ótimo de *clusters*  $k$  é aquele que maximiza a silhueta média dentro de uma faixa de valores possíveis para  $i$ .

Conforme destacado por Batool e Hennig (2021), a utilização da silhueta média é recomendada devido à sua capacidade de fornecer uma medida objetiva, intuitiva e resistente a valores atípicos. O cálculo dessa métrica é realizado inicialmente ao calcular o coeficiente de silhueta,  $s(i)$ , para cada elemento  $i$ , conforme a equação a seguir:

$$s(i) = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (4)$$

Nesta equação,  $(a_i)$  representa a média das distâncias entre o elemento  $i$  e todos os outros elementos no mesmo *cluster*, enquanto  $b_i$  é a menor média das distâncias entre o elemento  $i$  e os elementos de qualquer outro *cluster*, distintos do próprio *cluster* do elemento  $i$ . O coeficiente de silhueta  $s(i)$  varia de  $-1$  a  $1$ , em que valores próximos de  $1$  indicam que a observação está bem alocada no seu *cluster* e distante de outros *clusters*, valores próximos de  $0$  indicam sobreposição ou ambiguidade e valores próximos de  $-1$  indicam que a observação foi alocada ao *cluster* errado. Assim, o método da silhueta oferece uma abordagem analítica para determinar o número ideal de *clusters*, considerando a adequação das unidades de análise dentro de seus *clusters* e o grau de separação entre os *clusters*.

Por último, o método do Gap Estatístico, proposto por Tibshirani et al. (2001), baseia-se na ideia fundamental de que, se as unidades amostrais  $i$  não formassem agrupamentos, sua distribuição se assemelharia à distribuição uniforme dentro das amplitudes das  $r$  variáveis observadas. A intuição subjacente a esse método é que, se o número escolhido de *clusters* for apropriado, a estrutura dos dados reais será

substancialmente mais proeminente do que a estrutura gerada aleatoriamente.

É importante ressaltar que a análise de agrupamentos, embora útil para diagnósticos pontuais, não garante que as similaridades intra *cluster* sejam sempre significativas. A técnica é mais eficaz quando combinada com conhecimentos específicos do problema em análise, orientando a tomada de decisões (Fawcett e Provost, 2016). Para avaliar a qualidade dos agrupamentos gerados, utilizar-se-ão métricas como o Índice de Silhueta, o Índice de Davies-Bouldin (DBI) e o Índice de Calinski-Harabasz (CHI), conforme recomenda Milligan e Cooper (1985).

O Índice de Silhueta varia de -1 a 1, sendo valores próximos a 1 indicativos de boa separação entre *clusters*, valores próximos a 0 indicam que a amostra está sobre a fronteira de decisão entre dois *clusters* e valores negativos, de baixa separação ou atribuição incorreta de amostras (Rousseeuw, 1987). O DBI mede a relação entre dispersão intra *cluster* e separação inter *cluster*, com valores menores indicando melhor qualidade de agrupamento (Davies e Bouldin, 1979). Já o CHI mede a razão entre variância inter *cluster* e intra *cluster*, com valores maiores indicando *clusters* mais densos e bem separados (Calinski e Harabasz, 1974). Essas métricas, em conjunto, fornecem uma avaliação abrangente da qualidade dos agrupamentos, auxiliando na identificação da estrutura ideal para o problema em questão.

Apesar da eficácia amplamente reconhecida do método k-means em diversos cenários, várias desvantagens foram identificadas por Gentle et al. (1991) e Fawcett e Provost (2016). O algoritmo é sensível ao número de *clusters*, exigindo uma especificação prévia, o que pode levar a agrupamentos subótimos se o número ideal não for corretamente determinado. Além disso, a sensibilidade à inicialização dos centróides pode resultar em convergência para mínimos locais inadequados, embora técnicas avançadas, como o k-means++, possam mitigar esse problema. Outras limitações incluem a incapacidade de lidar com formas complexas de *clusters*, sensibilidade a *outliers* que podem afetar negativamente a formação dos *clusters* e dificuldade em lidar com diferentes densidades de *clusters*, além da dependência da métrica de distância euclidiana, que pode ser inadequada para certos tipos de dados. Ademais, devido à natureza heurística do algoritmo, não há garantia de convergência para o ótimo global, e lidar com dados de alta dimensionalidade pode resultar em agrupamentos menos representativos.

### 3.3 Análise dos componentes principais (PCA)

A obtenção dos componentes principais envolve o processo de diagonalização de matrizes que são simétricas e semi-definidas positivas. A partir desse procedimento, é possível calcular os componentes principais de maneira direta, tornando-os aplicáveis em diversas áreas científicas variadas. Essa viabilidade deriva da disponibilidade de várias ferramentas computacionais capazes de realizar operações matriciais para efetuar a diagonalização de matrizes simétricas positivas semidefinidas (Jolliffe, 2002).

O PCA oferece vantagens notáveis, como a capacidade de eliminar a multicolinearidade entre variáveis. Isso se traduz na habilidade de converter um conjunto original de variáveis que estão correlacionadas entre si em um novo conjunto de variáveis não correlacionadas, conhecidas como componentes principais. Ademais, ele permite reduzir o número de variáveis para um conjunto de eixos ortogonais que representam uma parcela das variáveis originais. Esses eixos explicam a variabilidade dos dados em ordem decrescente e independente (Hongyu, 2015; Regazzi e Cruz, 2020).

Por outro lado, há algumas desvantagens associadas ao PCA. Ele é sensível a valores discrepantes (*outliers*) nos dados e não é recomendado quando há muitos valores ausentes na matriz ou quando se lida com dados ausentes. Também é inadequado quando o número de variáveis é maior do que o número de amostras. A redução de variáveis implica na perda de informações sobre a variabilidade original dos dados. Entretanto, essa parte retida pode ser considerada como o padrão de resposta, enquanto a parte descartada pode ser vista como ruído, ou seja, erro de medição e redundância. Vale notar que o PCA não é sempre eficaz - em algumas situações, mesmo após a redução de dimensões, a complexidade do problema pode não diminuir significativamente. Isso acontece especialmente quando as variáveis originais possuem pouca correlação, e no caso extremo quando a matriz de correlação é uma matriz identidade, nesse caso, os componentes principais se assemelham às próprias variáveis originais (Hongyu, 2015; Regazzi e Cruz, 2020).

De acordo com Hongyu (2015), sejam as variáveis  $X_1, X_2, \dots, X_k$  em cada uma das  $n$  observações. Este conjunto de  $n \times k$  medidas originais forma uma matriz de dados  $X(n \times k)$ :

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1K} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{nK} \end{bmatrix} \quad (5)$$

Para obter os componentes principais de uma forma geral, seja um conjunto de  $k$  variáveis  $X_1, X_2, \dots, X_k$  com médias  $\mu_1, \mu_2, \dots, \mu_k$  e variância  $\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2$ , respectivamente. Tais variáveis são dependentes entre si e, portanto, possuem covariância entre a  $i$ -ésima e  $j$ -ésima variável definida por  $\sigma_{ij}$ , para  $i \neq j = 1, 2, \dots, k$ . Desse modo, as  $k$  variáveis podem ser expressas em forma vetorial por:  $X = [X_1, X_2, \dots, X_k]'$ , com vetor de médias  $\mu = [\mu_1, \mu_2, \dots, \mu_k]'$  e matriz de covariância  $\Sigma$ .

$$\Sigma = \begin{bmatrix} \sigma_{11}^2 & \cdots & \sigma_{1k}^2 \\ \vdots & \ddots & \vdots \\ \sigma_{n1}^2 & \cdots & \sigma_{nk}^2 \end{bmatrix} \quad (6)$$

Então, os pares de autovalores e autovetores  $(\lambda_1, v_1), (\lambda_2, v_2), \dots, (\lambda_k, v_k)$  são encontrados, de modo que se pode definir o  $i$ -ésimo componente principal a partir dos autovalores  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$ , associados à matriz  $\Sigma$  (Hongyu, 2015).

$$Z_i = v_i'X = v_{i1}X_1 + v_{i2}X_2 + \dots + v_{ik}X_k, \forall i = 1, \dots, k. \tag{7}$$

O objetivo será determinar a variável  $Z_i$  a partir da  $k$  variáveis contidas na matriz  $X$ . A ideia envolve a transformação dos pontos de coordenadas originais em um plano, maximizando a distância entre eles, o que se traduz em aumentar ao máximo a variabilidade de  $Z_i$ . A variância de  $Z_i$  é dada por:

$$Var(Z_i) = Var(v_i'X) = v_i'Var(X)v_i = v_i'\Sigma v_i \tag{8}$$

em que  $i = 1, \dots, k$ .

Utilizando a decomposição espectral<sup>7</sup> da matriz  $\Sigma$ , tem-se que  $\Sigma = P\Lambda P'$ , em que  $P$  é a matriz composta de autovetores de  $\Sigma$  em suas colunas, enquanto  $\Lambda$  representa a matriz diagonal formada pelos autovalores de  $\Sigma$ , ou seja:

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_k \end{bmatrix} \tag{9}$$

e

$$P = [v_1'v_2'\dots v_k'] \tag{10}$$

Deste modo:

$$tr(\Sigma) = tr(P\Lambda P') = tr(\Lambda P P') = tr(\Lambda I) = tr(\Lambda) = \sum_{i=1}^k \lambda_i \tag{11}$$

Ou seja,  $tr(\Sigma)$  é dado pela soma dos elementos da diagonal principal:

$$tr(\Sigma) = \sum_{i=1}^k \sigma_{ii}^2 = \sum_{i=1}^k \lambda_i \tag{12}$$

Dessa forma, a quantidade total de variação presente nas variáveis originais equivale à quantidade total de variação presente nos componentes principais (Johnson e Wichern, 2007). A contribuição de cada componente principal ( $Z_i$ ) é expressa como uma porcentagem, e é possível calcular a explicação individual de cada componente, por exemplo, para o  $k$ -ésimo componente principal, a proporção da explicação é determinada por:

$$C_k = \frac{Var(Z_i)}{\sum_{i=1}^k Var(Z_i)} = \frac{\lambda_i}{\sum_{i=1}^k \lambda_i} = \frac{\lambda_i}{tr(\Sigma)} \tag{13}$$

<sup>7</sup>Seja  $A$  uma matriz simétrica  $n \times n$ , então  $A$  tem uma decomposição espectral  $A = CDC'$ , em que  $C$  é uma matriz  $n \times n$  nas quais suas colunas são os autovetores  $C_1, C_2, \dots, C_n$  que correspondem aos autovalores  $\lambda_1, \lambda_2, \dots, \lambda_n$  de  $A$  e  $D$  é uma matriz diagonal  $n \times n$  que possui como diagonal principal  $\lambda_1, \lambda_2, \dots, \lambda_n$  (Wansbeek e Kapteyn, 1982).

Através da avaliação da proporção pela qual o modelo de  $k$  componentes principais explica a variância total, é possível decidir quantos componentes devem ser mantidos. De acordo com Hongyu et al. (2016), em geral escolhe-se o componente principal de maior importância (o primeiro componente principal) como sendo aquele de maior variância, que explique o máximo de variabilidade dos dados, o segundo componente de maior importância, o que apresenta a segunda maior variância e assim sucessivamente, até o componente principal de menor importância (Manly, 1986). Por outro lado, os últimos componentes principais serão responsáveis por direções que não estão associadas a muita variabilidade. Em outras palavras, esses últimos componentes principais identificarão relações lineares entre as variáveis originais próximo de constante (Johnson e Wichern, 2007; Anderson, 2003).

## 4. Análise e discussão dos resultados

### 4.1 Algoritmo de clusterização de k-means

Para iniciar a análise de agrupamentos, foram definidas a métrica de dissimilaridade (ou similaridade) entre os elementos e o método estatístico de formação dos grupos (hierárquico ou não hierárquico). Para determinar o número ideal de *clusters*, foram calculados três métodos, resultando em duas ou três possibilidades (Figuras 1 e 2): dois métodos indicaram três *clusters*, enquanto um indicou dois. Para definir o número ideal de *clusters* (k-ótimo), avaliou-se uma a qualidade dos agrupamentos gerados, considerando 2 e 3 *clusters*, respectivamente.

As métricas de avaliação (Silhouette, Davies-Bouldin e Calinski-Harabasz) indicam que o agrupamento com 3 *clusters* é superior em todas as métricas (Tabela 3). O índice de Silhouette é ligeiramente superior e o índice de Davies-Bouldin é inferior para 3 *clusters*, sugerindo melhor qualidade. Além disso, o índice de Calinski-Harabasz é consideravelmente maior para 3 *clusters*, indicando uma definição superior na qualidade dos *clusters*.

Embora a literatura indique que um baixo valor no Índice de Silhueta e um alto valor no Índice de Davies-Bouldin apontam para *clusters* não bem separados, estando no limite da sobreposição entre *clusters* (Davies e Bouldin, 1979; Rousseeuw, 1987), a análise do método do cotovelo e do índice de Silhouette sugere que 3 *clusters* podem ser considerados o número ideal, mesmo com as métricas de validação interna não

**Tabela 3.** Resultados das métricas de validação interna dos *clusters*.

Métricas	2 <i>clusters</i>	3 <i>clusters</i>
Índice de Silhueta	0,023	0,063
Índice de Davies-Bouldin	4,42	2,31
Índice de Calinski-Harabasz	2,61	36,53

Fonte: Elaboração própria.

sendo ideais.

Para efetuar o processo de agrupamento, foram empregadas como variáveis de referência as dimensões relacionadas ao contexto social, econômico e criminal dos 113 bairros situados em Fortaleza. A técnica de análise de *cluster* foi empregada com o propósito de categorizar os bairros com base nas suas características intrínsecas, de maneira que dentro de cada agrupamento estejam presentes aqueles bairros que compartilham traços mais similares. Deste modo, para agrupar os bairros, foram utilizadas as cinco variáveis relacionadas à criminalidade e as três dimensões do IDH (vide Tabela 1). Em seguida, procedeu-se com o cálculo da matriz de distâncias (Equação (3)), considerando a distância euclidiana. Desta feita, foram identificados três *clusters*<sup>8</sup> cujas características são apresentadas na Tabela 4.

**Tabela 4.** Médias das variáveis por *cluster*.

Clusters	Qtd de Bairros	IDH Renda	IDH Long.	IDH Edu.	Valores absolutos				
					H	R	F	AD	AA
1	42	0,34	0,64	0,98	3	234	247	4,59	6
2	63	0,09	0,28	0,93	10	207	127	31,32	16
3	8	0,1	0,38	0,95	32	969	693	32,79	50

Fonte: Elaboração própria.

Nota 1: H - Homicídios; R - Roubos; F - Furtos; AD - Apreensão de Drogas/Entorpecentes (kg); e AA - Apreensão de Armas de Fogo.

Nota 2: a utilização de valores absolutos em vez de taxas por 10.000 habitantes visa mitigar o impacto de *outliers* na análise de agrupamento, especialmente em bairros com baixa população, como centros urbanos. Nessas áreas, as altas taxas de criminalidade resultantes do pequeno denominador podem distorcer a percepção de similaridade entre as áreas analisadas (Fávero e Belfiore, 2021). Além disso, crimes violentos, como homicídios, são considerados eventos raros (Beato Filho et al., 2001), o que justifica o uso de dados brutos (valores absolutos) para evitar flutuações e instabilidades nas taxas (Lawson, 2006).

O primeiro *cluster* compreende 42 bairros, representando 37% do total analisado, e se caracteriza por um nível mais elevado de desenvolvimento, evidenciado por um Índice de Desenvolvimento Humano (IDH) mais alto, com médias de 0,34 para IDH-Renda, 0,64 para IDH Longevidade e 0,98 para IDH-Educação. Esses bairros exibem um desenvolvimento mais balanceado nas diferentes dimensões do IDH. Embora registrem crimes, as médias são relativamente menores, refletindo o perfil mais nobre e economicamente desenvolvido dessas áreas.

O segundo *cluster* abrange 63 bairros ou 56% da amostra, mostrando um IDH geral mais baixo, com médias de 0,09 para IDH-Renda, 0,28 para IDH-Longevidade e 0,93 para IDH-Educação. Esse *cluster* revela uma mistura de indicadores sociais menos avançados, com criminalidade baixa para roubos e furtos, moderada para homicídios e apreensões de armas e alta para apreensões de drogas, comparável ao *cluster* 3 nesse último aspecto.

<sup>8</sup>Os gráficos da dissimilaridade e do método da silhueta identificaram a existência de três *clusters* e são expostos no Apêndice. O método do Gap Estatístico, por sua vez, indicou a existência de dois *clusters*. Desta feita, optou-se por trabalhar com três *clusters*.

O terceiro *cluster*, formado por apenas 8 bairros (7% da amostra), é marcado por indicadores de desenvolvimento humano mais baixos e elevadas taxas de criminalidade, incluindo as maiores médias de homicídios (32), roubos (969), furtos (693), 32,79 kg de apreensões de drogas e 50 apreensões de armas de fogo. Isso indica desafios acentuados em termos de desenvolvimento e segurança pública nesses bairros.

Em síntese, a *clusterização* dos bairros de Fortaleza reflete variações nos níveis de desenvolvimento humano e incidência criminal. O *cluster 1* tem desenvolvimento mais alto e menor criminalidade; o *cluster 2* apresenta desenvolvimento e criminalidade moderados; e o *cluster 3*, com poucos bairros, enfrenta severos desafios em ambos os aspectos, com as maiores taxas de criminalidade.

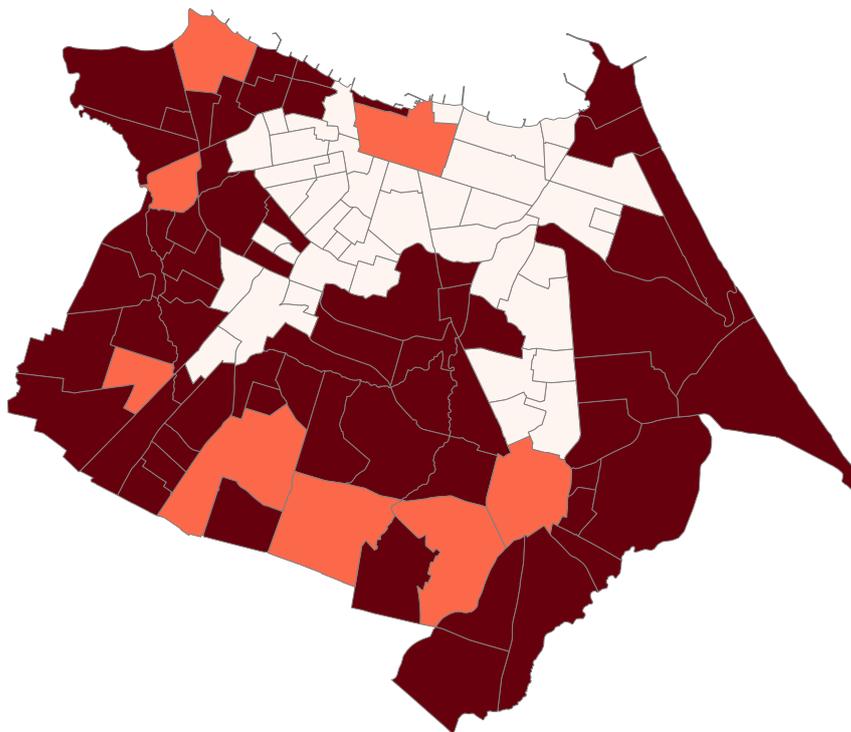
A Figura 3 ilustra a distribuição espacial dos *clusters* em Fortaleza: bairros em branco, localizados no centro, correspondem ao *cluster 1*, exibindo menor criminalidade mas com incidências moderadas de roubos e furtos; bairros em vinho representam o *cluster 2*, com níveis moderados de criminalidade; e os bairros em laranja, pertencentes ao *cluster 3*, são identificados como os mais perigosos devido à alta criminalidade. Detalhes sobre os bairros em cada *cluster* estão disponíveis no Apêndice A.

A análise do mapa revela que os bairros mais abastados de Fortaleza estão cercados por áreas com altas taxas de criminalidade, criando um “cinturão de insegurança” ao redor dessas regiões prósperas. Notavelmente, metade dos bairros marcados em laranja, indicando elevada criminalidade, situa-se na zona sul, incluindo Acarapé (AIS 09), Mondubim (AIS 09), Prefeito José Walter (AIS 09), Messejana (AIS 03) e Jangurussu (AIS 03), com Bom Jardim (AIS 02) localizando-se a oeste. Ao noroeste, encontram-se Antônio Bezerra (AIS 06) e Barra do Ceará (AIS 08). Distintamente, o bairro Centro (AIS 02), mesmo fazendo fronteira com áreas nobres como Aldeota (AIS 01), Meireles (AIS 01) e Praia de Iracema (AIS 01), também figura entre as áreas mais perigosas, principalmente devido aos elevados índices de roubos e furtos. Os bairros marcados em vinho, demonstrando uma continuidade geográfica, percorrem a extensão leste a oeste da cidade.

Essas análises fornecem uma base para identificar as regiões onde a criminalidade é mais prevalente, auxiliando as autoridades a concentrar recursos e intervenções nessas áreas para mitigar o crime. Além de apontar os bairros com maior incidência criminal, os dados oferecem *insights* sobre as condições socioeconômicas que favorecem a criminalidade, como a pobreza e os baixos níveis de educação. Esses *insights* podem orientar os formuladores de políticas na criação de programas direcionados a esses fatores, visando à redução da criminalidade.

Adicionalmente, os resultados deste estudo possibilitam a avaliação da efetividade das políticas de segurança pública existentes, permitindo que as autoridades identifiquem quais estratégias estão funcionando e quais necessitam de ajustes ou descontinuação. Conforme destacado por Felix (2002) e Melara (2008), tais achados reforçam

**Figura 3.** Cinturão de insegurança em Fortaleza – formação dos *clusters*, a partir do k-means.



Fonte: Elaboração

própria.

Nota: *Cluster 1* – criminalidade baixa (branco); *Cluster 2* – criminalidade média (vinho); *Cluster 3* – criminalidade alta (laranja).

a correlação entre pobreza e criminalidade, notadamente em regiões de baixo *status* social e econômico, exemplificado pelos bairros do *cluster 3*.

#### 4.2 Análise dos componentes principais (PCA)

Através da proporção da variância total que é explicada pelo modelo com *k* componentes principais, é possível decidir quantos componentes devem ser mantidos. Com base nos resultados obtidos pela técnica dos componentes principais, optou-se por reduzir a dimensionalidade das cinco variáveis relacionadas ao crime (Homicídios, Roubos, Furtos, Apreensão de Drogas/Entorpecentes (kg) e Apreensão de Armas de Fogo) e das três variáveis de IDH para duas dimensões, ou seja, dois componentes principais.

Foram utilizados dois componentes principais que foram responsáveis por 67,56% da variação total sobre as incidências de crime em todos os bairros da cidade de Fortaleza, em que o  $CP_1$  foi responsável por 39,26% e o segundo,  $CP_2$ , por 28,30% das variações dos dados. Como os dois primeiros componentes principais explicaram 67,58% da variação total dos dados (Tabela 5), indicando uma captura substancial da informação contida no conjunto original. de acordo com Rencher (2002), pelo menos

**Tabela 5.** Análise dos Componentes Principais: autovalores e variância acumulada.

Componentes	Autovalores	Diferença entre autovalores	Variância explicada	
			Proporção	Acumulada
1	3,1695	-	0,3932	0,3932
2	2,2850	-0,8845	0,2834	0,6766
3	0,9985	-1,2865	0,1239	0,8005
4	0,6366	-0,3619	0,0790	0,8795
5	0,4883	-0,1483	0,0606	0,9400
6	0,2318	-0,2564	0,0288	0,9688
7	0,1535	-0,0783	0,0190	0,9878
8	0,0981	-0,0554	0,0122	1,0000

Fonte: Elaboração própria. Nota: H - Homicídios; R - Roubos; F - Furtos; AD - Apreensão de Drogas/Entorpecentes (kg); e AA - Apreensão de Armas de Fogo.

**Tabela 6.** Coeficientes de ponderação dos dois componentes principais.

$X$	Variáveis	$CP_1$	$CP_2$
$X_1$	Homicídios	0,48	0,18
$X_2$	Roubos	0,26	0,56
$X_3$	Furtos	0,02	0,58
$X_4$	Apreensão de Drogas	0,06	-0,04
$X_5$	Apreensão de Armas de Fogo	0,47	0,24
$X_6$	IDH Educação	-0,42	0,33
$X_7$	IDH Longevidade	-0,41	0,29
$X_8$	IDH Renda	-0,37	0,26

Fonte: Elaboração própria.

70% da variância total deve ser explicada pelo primeiro e pelo segundo componentes principais.

Em trabalho similar para a taxa de criminalidade em algumas cidades dos Estados Unidos, Hongyu et al. (2016) reduziram a dimensionalidade de sete variáveis para dois componentes principais, que foram responsáveis por 68,13% da variância total no conjunto de dados. Segundo tais autores, os dois primeiros componentes principais resumem efetivamente a variância amostral total e podem ser utilizados para o estudo do conjunto de dados, estratégia similar à seguida neste estudo.

A escolha de dois componentes principais na presente análise é sustentada pelo Critério de Kaiser (Kaiser, 1960), refletindo uma tentativa de capturar a essência dos dados ao mesmo tempo em que se preserva a capacidade de interpretação dos resultados. Esse critério recomenda reter componentes com autovalores superiores a 1. Conforme aplicado ao estudo atual, os dois primeiros componentes satisfazem esse critério, reforçando a validade de sua seleção (Tabela 5). Essa decisão é fundamentada na análise de autovalores e na proporção de variância explicada, como demonstrado na tabela de resultados do PCA.

A avaliação das diferenças entre autovalores sucessivos revela uma redução signi-

ficativa após o segundo componente, de 2,28 para 0,99. Tal observação sugere que os primeiros dois componentes concentram as informações mais críticas, enquanto componentes subsequentes adicionam progressivamente menos ao entendimento da variância. Além disso, uma revisão prévia da matriz de correlação, realizada antes da aplicação da PCA, confirmou que todas as variáveis incluídas na análise mantêm correlações substanciais, com todas excedendo o valor de 0,3 (Figura 4). Tal constatação valida a relevância de cada variável selecionada para a análise.

O PCA aproveita as relações entre essas oito variáveis, então, se os dados fossem aleatórios e não houvesse correlação entre nenhum deles, a metodologia seria infrutífera (Greenacre et al., 2022). O PCA busca uma combinação linear de variáveis que tenha a máxima variância; em outras palavras, ele os combina de uma maneira que reflete a maior variação entre os bairros de Fortaleza.

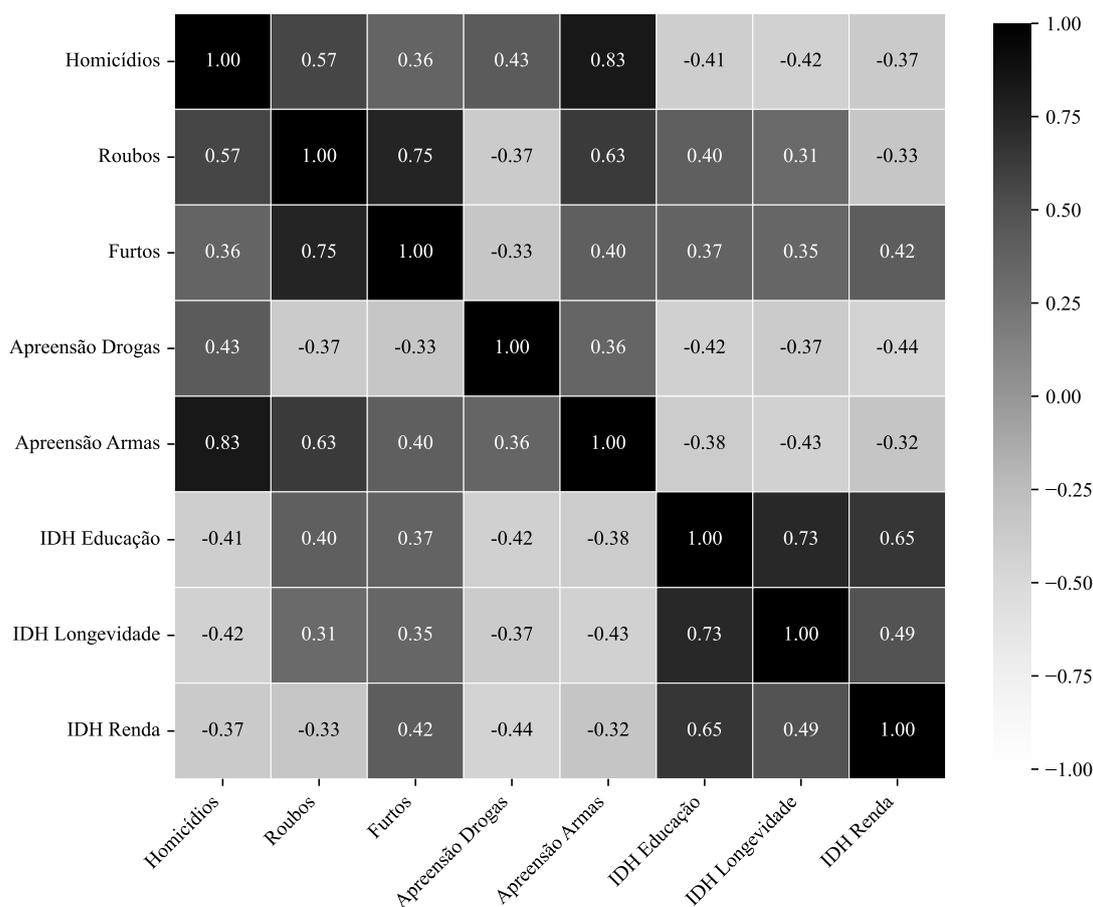
Com o intuito de se entender a importância de cada variável na construção dos dois componentes, na Tabela 6, são apresentados os coeficientes de ponderação de cada variável para a construção dos dois componentes principais. Com a seleção de dois componentes principais, a redução da dimensão de oito variáveis originais para dois componentes principais é bastante razoável.

$$CP_1 = 0,48X_1 + 0,26X_2 + 0,02X_3 + 0,06X_4 + 0,47X_5 - (0,42X_6 + 0,41X_7 + 0,37X_8) \quad (14)$$

$$CP_2 = 0,18X_1 + 0,56X_2 + 0,58X_3 - 0,04X_4 + 0,24X_5 + 0,33X_6 + 0,29X_7 + 0,26X_8 \quad (15)$$

Portanto, decidiu-se utilizar unicamente os dois primeiros componentes principais para a composição das Equações (14) e (15). De acordo com a Equação (14) e a Tabela 6, no primeiro componente principal, destacaram-se as variáveis  $X_1$  (Homicídios) e  $X_5$  (Apreensão de Armas de Fogo) e, neste caso, pode-se chamá-lo de componente de homicídios e apreensão de armas. De outro lado, no segundo componente principal, Equação (15), fica evidente a preponderância das variáveis  $X_2$  (Roubos) e  $X_3$  (Furtos), podendo ser chamado de componente de roubos e furtos. Ademais, fica evidente o contraste entre os dois componentes principais para as dimensões do IDH: bairros que apresentam maiores resultados para o  $CP_1$  detêm níveis de IDH menores, enquanto os maiores pontuadores no  $CP_2$  detêm maiores resultados de IDH.

O PCA foi aplicado para reduzir as dimensões das variáveis originais sem comprometer a integridade das informações. Por sua própria natureza, os principais componentes apresentam uma correlação nula, o que significa que a variação explicada no primeiro componente principal ( $CP_1$ ) é completamente independente da variação explicada no segundo componente principal ( $CP_2$ ) e assim por diante. Essa característica implica que a influência de um componente principal em particular não resultará em uma influência correlacionada nos outros componentes principais, evidenciando a ortogonalidade entre eles (Savegnago et al., 2011; Fraga et al., 2016).

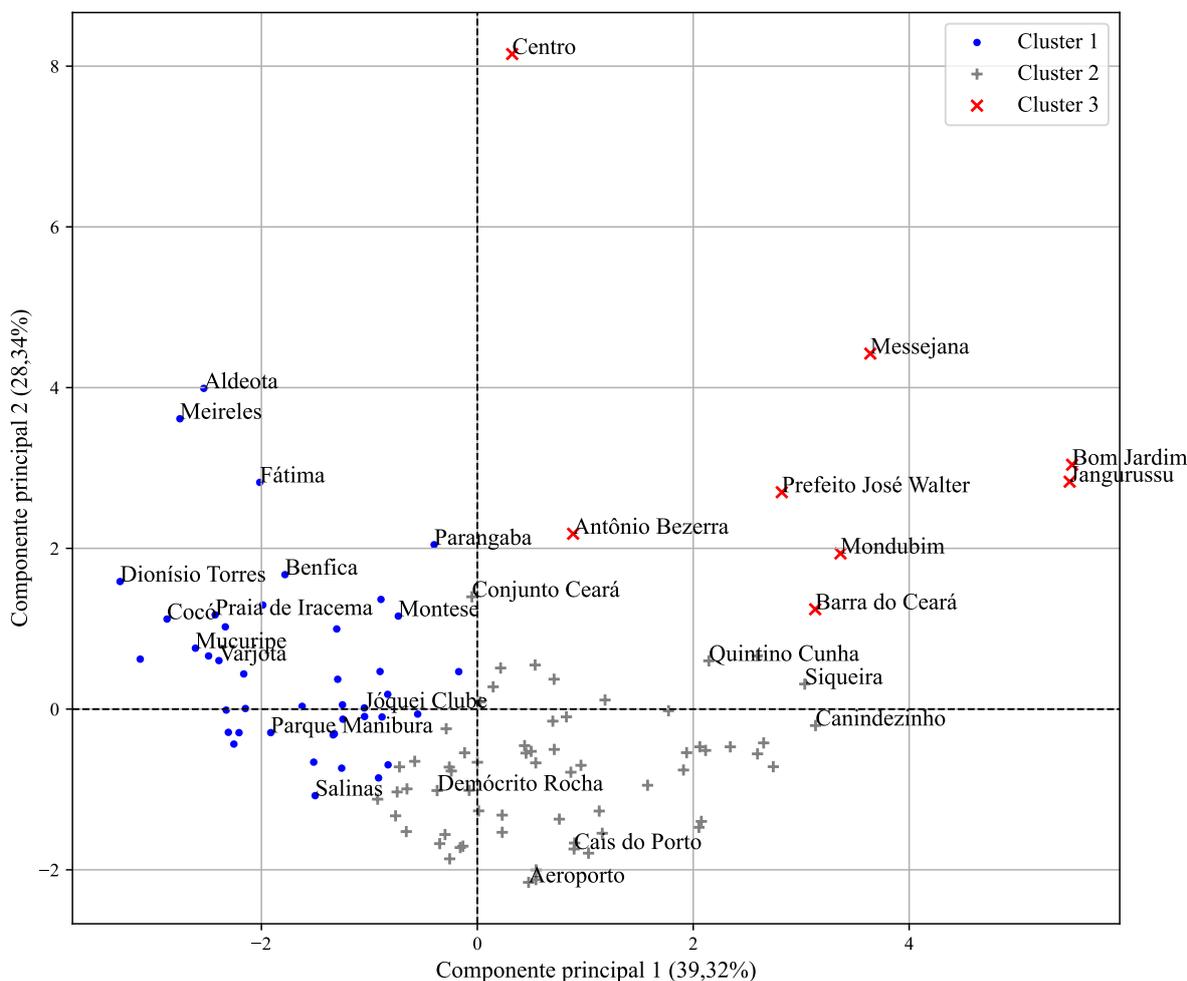
**Figura 4.** Matriz de correlação entre as variáveis utilizadas no PCA.

Fonte: Elaboração própria.

Analisando a Figura 5, que é o gráfico de dispersão  $CP_1 \times CP_2$  com os bairros sobre os índices de criminalidade e IDH, Equações (14) e (15), fica patente que não há sobreposição entre os *clusters*, pode-se concluir que, de acordo com os dados de criminalidade por bairros da cidade de Fortaleza e com o PCA, Bom Jardim, Jangu-russu, Messejana, Mondubim e Barra do Ceará, nessa ordem, são os que possuem maiores ocorrências de homicídios e apreensão de armas de fogo, a partir do  $CP_1$ . De outro lado, a partir da análise da  $CP_2$ , os bairros Centro, Messejana, Aldeota, Meireles e Bom Jardim foram os que apresentaram os maiores resultados para criminalidade relacionada a roubos e furtos, nessa ordem. Os bairros Bom Jardim e Messejana apareceram nas duas listagens, ou seja, apresentaram altos valores tanto para o  $CP_1$  quanto para o  $CP_2$ . Interessante observar que Aldeota e Meireles, bairros mais nobres da cidade, figuraram na lista dos cinco bairros com maiores resultados no  $CP_2$ .

Os oito bairros integrantes do *cluster 3*, conforme detalhado na Tabela 1, situam-se no extremo do primeiro quadrante da Figura 5, demonstrando os maiores valores para os dois componentes principais analisados. Essa constatação indica que a base de dados utilizada proporciona uma associação direta entre o crime em Fortaleza e as características locais específicas dos 113 bairros estudados, captando uma

**Figura 5.**  $CP_1 \times CP_2$  sobre índices de criminalidade por bairros de Fortaleza, em 2021.



evidente dependência espacial, um achado que ressoa com as conclusões de Guerra (2016). Uma análise preliminar, embora não exaustiva, revela uma distribuição desigual do crime na capital cearense: apenas 7% dos bairros, representados pelo *cluster* 3, concentraram cerca de 71,1% dos homicídios, 67,1% dos incidentes de roubos e furtos, 47,7% das apreensões de drogas/entorpecentes e 69,4% das apreensões de armas de fogo do total de ocorrências reportadas em 2021. Esses dados sublinham o motivo pelo qual o *cluster* 3 foi identificado como o mais inseguro entre os analisados, justificando sua distinção em relação aos demais.

Por outro lado, os bairros do *cluster* 1 estão majoritariamente localizados nos 2º e 3º quadrantes, evidenciando um perfil diferenciado em termos de incidência criminal. Enquanto isso, os bairros associados ao *cluster* 2 situam-se intermediariamente, estendendo-se desde áreas como o Conjunto Ceará até Canindezinho a leste, e alcançando o Aeroporto ao sul. Essa distribuição sugere uma gradação na intensidade e na natureza das atividades criminais entre os *clusters*, refletindo a complexa dinâmica espacial do crime em Fortaleza.

## 5. Considerações finais e propostas de políticas públicas

No Ceará, estado com o maior número absoluto de homicídios, Fortaleza destaca-se como epicentro dessa violência. O presente estudo emprega técnicas estatísticas avançadas para agrupar bairros da capital com padrões similares de criminalidade. A metodologia utilizada simplifica a complexidade dos dados, preservando informações cruciais e possibilitando a identificação das áreas mais violentas de Fortaleza. Essa abordagem inovadora analisa semelhanças e diferenças em padrões criminais e socioeconômicos entre diferentes partes da cidade, oferecendo *insights* valiosos para futuras análises e elaboração de políticas públicas direcionadas.

Os resultados do estudo destacam a concentração de bairros de alto risco nas regiões sul e oeste, com o bairro Centro sendo uma notável exceção devido à alta incidência de roubos e furtos. Foi observado também que os bairros mais afluentes são frequentemente circundados por áreas com maior criminalidade, formando um “cinturão de insegurança”.

Utilizando a Análise dos Componentes Principais (PCA), foram desenvolvidos dois indicadores que priorizam diferentes aspectos criminais e socioeconômicos. Um enfoca mais em homicídios e apreensões de armas, identificando bairros como Bom Jardim e Jangurussu como os mais afetados, enquanto o outro destaca roubos e furtos, apontando para áreas como Centro e Aldeota. Essa dualidade nos indicadores mostra que regiões como Bom Jardim e Messejana são afetadas independentemente dos critérios, e até bairros privilegiados apresentam altas taxas de determinados crimes.

A pesquisa corrobora os trabalhos anteriores de Medeiros et al. (2021) e Muggah e Pellegrino (2020) e apoia a formulação de políticas focadas na redução de Violência, Criminalidade e Marginalidade (VCM) em Fortaleza. Além disso, oferece uma base para a sinergia intergovernamental na alocação de recursos e esforços para regiões críticas, promovendo estratégias de prevenção e intervenção mais eficazes.

O estudo não só proporciona uma visão holística do panorama criminal da cidade, mas também permite orientar a alocação coordenada de recursos em segurança pública e desenvolvimento social. Ao identificar as tendências criminais, a análise possibilita o aperfeiçoamento contínuo das políticas de prevenção, além de se tornar um instrumento que passa a nortear a transparência e a responsabilidade das autoridades.

Por último, o estudo pavimenta o caminho para futuras pesquisas, particularmente com a expectativa de integrar os achados com dados do Censo Demográfico Brasileiro de 2022. Isso permitirá explorar correlações entre criminalidade, informalidade e pobreza nos bairros de Fortaleza, ampliando o entendimento dos padrões de criminalidade na cidade.

## Referências

- Anaraki, S. A. M., Haeri, A., e Moslehi, F. (2021). A hybrid reciprocal model of PCA and K-means with an innovative approach of considering sub-datasets for the improvement of K-means initialization and step-by-step labeling to create clusters with high interpretability. *Pattern Analysis and Applications*, 24(3):1387–1402.
- Anderson, T. W. (2003). *An introduction to multivariate statistical analysis*. Wiley series in probability and statistics. Wiley-Interscience, Hoboken, N. J., 3rd ed edition.
- Andresen, M. A. (2006). Crime Measures and the Spatial Analysis of Criminal Activity. *The British Journal of Criminology*, 46(2):258–285.
- Anselin, L., Cohen, J., Cook, D., Gorr, W., e Tita, G. (2000). Spatial analyses of crime. *Criminal justice*, 4(2):213–262.
- Arbex, M., Corrêa, M. V., e Magalhães, M. R. V. (2023). Tolerance of Informality and Occupational Choices in a Large Informal Sector Economy. *The B.E. Journal of Macroeconomics*, 23(1):241–278.
- Balaguer-Coll, M. T., Brun-Martos, M. I., Márquez-Ramos, L., e Prior, D. (2019). Local government efficiency: determinants and spatial interdependence. *Applied Economics*, 51(14):1478–1494.
- Batella, W. B. e Diniz, A. M. A. (2010). Análise espacial dos condicionantes da criminalidade violenta no estado de Minas Gerais. *Sociedade & Natureza*, 22(1):151–163.
- Batool, F. e Hennig, C. (2021). Clustering with the Average Silhouette Width. *Computational Statistics & Data Analysis*, 158:107190.
- Beato Filho, C. (1998). Determinantes da criminalidade em minas gerais. *Revista Brasileira de Ciências Sociais*, 13(37):74–87.
- Beato Filho, C. C., Assunção, R. M., Silva, B. F. A. D., Marinho, F. C., Reis, I. A., e Almeida, M. C. D. M. (2001). Conglomerados de homicídios e o tráfico de drogas em Belo Horizonte, Minas Gerais, Brasil, de 1995 a 1999. *Cadernos de Saúde Pública*, 17(5):1163–1171.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer, New York, NY, 6th edition.
- Calinski, T. e Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods*, 3(1):1–27.
- Costa, F. V. C. (2018). “A violência impera nessa cidade”: Reflexões sobre o medo de assaltos em Fortaleza, CE. Tese de doutorado, Universidade Federal do Ceará, Fortaleza, CE.
- Davies, D. L. e Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2):224–227.

- Fawcett, T. e Provost, F. (2016). *Data Science para Negócios*. Alta Books, Rio de Janeiro, RJ.
- Feitoza, B. M. B. e Costa, J. H. (2019). Violência urbana, insegurança e turismo na 'Cidade do sol' (Natal/RN/Brasil). *Turismo y Sociedad*, 25:93–112.
- Felix, S. A. (2002). *Geografia do crime: interdisciplinaridade e relevância*. Unesp Marília Publicações, Marília.
- Fraga, A. B., Silva, F. D. L., Hongyu, K., Santos, D. D. S., Murphy, T. W., e Lopes, F. B. (2016). Multivariate analysis to evaluate genetic groups and production traits of crossbred Holstein × Zebu cows. *Tropical Animal Health and Production*, 48(3):533–538.
- Fávero, L. P. e Belfiore, P. (2021). *Manual de análise de dados: estatística e modelagem multivariada com Excel®, SPSS® e Stata®*. Elsevier Brasil, Rio de Janeiro, RJ.
- Fórum Brasileiro de Segurança Pública (2015). Anuário do fórum brasileiro de segurança pública. São Paulo, Ano 9.
- Fórum Brasileiro de Segurança Pública (2022). Anuário do fórum brasileiro de segurança pública. São Paulo, Ano 16.
- Gentle, J. E., Kaufman, L., e Rousseuw, P. J. (1991). Finding Groups in Data: An Introduction to Cluster Analysis. *Biometrics*, 47(2):788.
- Goldberg, M., Kim, K. W., e Ariano, M. (2014). *How firms cope with crime and violence: Experiences from around the world*. The World Bank, Washington, DC.
- Greenacre, M., Groenen, P. J. F., Hastie, T., D'Enza, A. I., Markos, A., e Tuzhilina, E. (2022). Principal component analysis. *Nature Reviews Methods Primers*, 2(1):100.
- Guerra, M. B. (2016). *A Estrutura Social do Crime e o Spillover Espacial do Policiamento: O Caso de Fortaleza, Brasil*. Dissertação de mestrado, Universidade Federal do Ceará (UFC), Centro de Aperfeiçoamento dos Economistas do Nordeste (CAEN), Programa de Pós-Graduação em Economia, Fortaleza, CE.
- Hair Junior, J. F., Black, W. C., Babin, N. J., Anderson, R. E., e Tatham, R. L. (2009). *Análise multivariada de dados*. Bookman editora, São Paulo, SP.
- Hart, T. C. (2020). Hot Spots of Crime: Methods and Predictive Analytics. In: Lersch, K. M. e Chakraborty, J., editores, *Geographies of Behavioural Health, Crime, and Disorder*, volume 126, Página 87–103. Springer International Publishing, Cham. Series Title: GeoJournal Library.
- Hartigan, J. A. e Wong, M. A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. *Applied Statistics*, 28(1):100.

- Hastie, T., Tibshirani, R., e Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer series in statistics. Springer, New York, NY, 2nd edition.
- Hongyu, K. (2015). *Comparação do GGE biplot-ponderado e AMMI-ponderado com outros modelos de interação genótipo x ambiente*. Tese de Doutorado, Universidade Estadual de São Paulo, Piracicaba, SP.
- Hongyu, K., Sandanielo, V. L. M., e Junior, G. J. D. O. (2016). Análise de Componentes Principais: Resumo Teórico, Aplicação e Interpretação. *E&S Engineering and Science*, 5(1):83–90.
- Instituto Brasileiro de Geografia e Estatística (2010). Censo demográfico. Disponível em: <https://censo2010.ibge.gov.br>. Acesso em: 03 mar. 2024.
- Instituto Brasileiro de Geografia e Estatística (2023). Produto interno bruto dos municípios. Disponível em: <https://www.ibge.gov.br/estatisticas/economicas/contas-nacionais/9088-produto-interno-bruto-dos-municipios.html>. Acesso em: 02 fev. 2024.
- Jain, A. K. e Dubes, R. C. (1988). *Algorithms for Clustering Data*. Prentice Hall, Englewood Cliffs, NJ.
- Johnson, R. A. e Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis*. Prentice Hall, Upper Saddle River, NJ, 6th edition.
- Jolliffe, I. T. (2002). *Principal Component Analysis For Special Types of Data*. Springer, New York, NY, 2nd edition.
- Kaiser, H. F. (1960). The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement*, 20(1):141–151.
- Lawson, A. B. (2006). *Statistical methods in spatial epidemiology*. Wiley series in probability and statistics. John Wiley & Sons, Chichester, England, 2nd edition.
- Loureiro, P. R. e Silva, E. C. D. (2010). Does violence deter investment, hinder economic growth? *Brazilian Review of Econometrics*, 30(1):53–67.
- Manly, B. F. (1986). *Multivariate Statistical Methods: A Primer*. Chapman and Hall, Boca Raton, FL.
- Medeiros, C. N. D., Pinheiro, F. S. D. A., Souza, G. M., e Pinheiro, D. R. d. C. (2021). Correlações espaciais entre homicídios e concentração de aglomerados subnormais em Fortaleza, CE, Brasil. *Revista GeoUECE*, 2(2):92–110.
- Melara, E. (2008). A dinâmica da violência criminal no espaço urbano de Santa Maria-RS. Dissertação de Mestrado, Universidade Federal do Rio Grande do Sul, Instituto de Geociências, Porto Alegre, Rio Grande do Sul.

- Melo, J. e Paiva, L. F. S. (2021). Violências em territórios faccionados do Nordeste do Brasil: notas sobre as situações do Rio Grande do Norte e do Ceará. *Revista USP*, 129:47–62.
- Milligan, G. W. e Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50(2):159–179.
- Moraes, S. S. d. e Siqueira, I. B. L. (2019). As margens como centro no bairro Benfica: falas da violência e do matar na cidade de Fortaleza. *RBSE Revista Brasileira de Sociologia da Emoção*, 18(53):143–159.
- Muggah, R. e Pellegrino, A. P. (2020). *Prevenção da violência juvenil no Brasil: uma análise do que funciona*. Instituto Igarapé; Fundo de População das Nações Unidas (UNFPA), Rio de Janeiro, RJ.
- Paiva, L. F. S. (2019). “Aqui não tem gangue, tem facção”: As transformações sociais do crime em Fortaleza. *Caderno CRH*, 32(85):165.
- Regazzi, A. J. e Cruz, C. D. (2020). *Análise multivariada aplicada*. Editora UFV, Viçosa, MG.
- Rencher, A. C. (2002). *Methods of Multivariate Analysis*. Wiley Series in Probability and Statistics. Wiley, 1 edition.
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65.
- Santos, A. L. C. (2019). Políticas criminais, direito à cidade e capital social. Reflexões sobre a sustentabilidade de políticas públicas de tratamento da criminalidade e da violência. *Revista de Direito da Cidade*, 11(3).
- Santos, M. D. (2020). *Ensaio sobre violência & criminalidade (V&C) em uma análise teórica, empírica e espacial: Evidências para o Estado do Ceará, capital de Fortaleza e bairros*. Tese de doutorado, Universidade Federal do Ceará (UFC), Centro de Aperfeiçoamento dos Economistas do Nordeste (CAEN), Programa de Pós-Graduação em Economia, Fortaleza, CE.
- Savegnago, R., Caetano, S., Ramos, S., Nascimento, G., Schmidt, G., Ledur, M., e Murnari, D. (2011). Estimates of genetic parameters, and cluster and principal components analyses of breeding values related to egg production traits in a White Leghorn population. *Poultry Science*, 90(10):2174–2188.
- Scott, A. J. (1986). Industrial Organization and Location: Division of Labor, the Firm, and Spatial Process. *Economic Geography*, 62(3):215.
- Secretaria da Segurança Pública e Defesa Social do Estado do Ceará (2021). Indicadores de segurança pública. Disponível em: <https://www.sspds.ce.gov.br/indicadores-de-seguranca-publica>. Fortaleza, CE. Acesso em: 25 jan. 2024.

Secretaria de Desenvolvimento Econômico de Fortaleza (2015). Relatório de desenvolvimento econômico. Disponível em: [https://dados.fortaleza.ce.gov.br/dataset/desenvolvimento\\_humano\\_bairro](https://dados.fortaleza.ce.gov.br/dataset/desenvolvimento_humano_bairro). Fortaleza, CE. Acesso em: 25 jan. 2024.

Secretaria do Turismo do Estado do Ceará (2023). Indicadores turísticos do ceará: 2010/2023.

Tibshirani, R., Walther, G., e Hastie, T. (2001). Estimating the Number of Clusters in a Data Set Via the Gap Statistic. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 63(2):411–423.

Wansbeek, T. e Kapteyn, A. (1982). A simple way to obtain the spectral decomposition of variance components models for balanced data. *Communications in Statistics - Theory and Methods*, 11(18):2105–2112.

Wu, S. (2023). Research on innovation and development of university instructional administration informatization in IoT and big data environment. *Soft Computing*, 27(24):19075–19094.

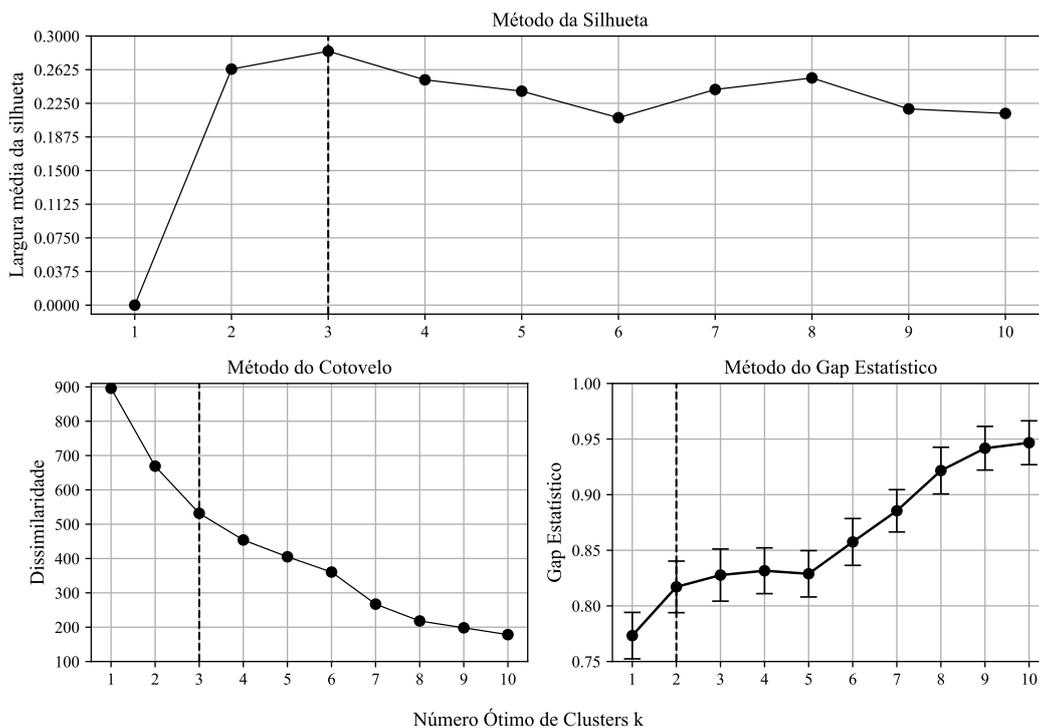
## Apêndice:

### A.1. Bairros pertencentes a cada *cluster*.

Clusters	Bairros
Cluster 1	Aldeota, Amadeu Furtado, Benfica, Bom Futuro, Cambeba, Cidade 2000, Cidade dos Funcionários, Cocó, Damas, Dionísio Torres, Engenheiro Luciano Cavalcante, Farias Brito, Fátima, Guararapes (Patriolino Ribeiro), Itaoca, Jacarecanga, Jardim América, Joaquim Távora, Jôquei Clube, José Bonifácio, Meireles, Monte Castelo, Montese, Mucuripe, Panamericano, Papicu, Parangaba, Parque Araxá, Parque Iracema, Parque Manibura, Parquelândia, Parreão, Praia de Iracema, Presidente Kennedy, Rodolfo Teófilo, Salinas, São Gerardo, São João do Tauape, Varjota, Vila Ellery, Vila Pery, Vila União.
Cluster 2	Aerolândia, Aeroporto, Alagadiço Novo (José de Alencar), Alto da Balança, Álvaro Weyne, Ancuri, Autran Nunes, Barroso, Bela Vista, Boa Vista (Castelão Mata Galinha), Bonsucesso, Cais do Porto, Cajazeiras, Canindezinho, Carlito Pamplona, Coaçu, Conjunto Ceará, Conjunto Esperança, Conjunto Palmeiras, Couto Fernandes, Cristo Redentor, Curió, Demócrito Rocha, Dendê, Dias Macêdo, Dom Lustosa, Edson Queiroz, Floresta, Genibaú, Granja Lisboa, Granja Portugal, Guajerú, Henrique Jorge, Itaperi, Jardim Cearense, Jardim das Oliveiras, Jardim Guanabara, Jardim Iracema, João XXIII, Lagoa Redonda, Lagoa Sapiranga (Coité), Manuel Dias Branco, Maraponga, Moura Brasil, Padre Andrade, Parque Dois Irmãos, Parque Presidente Vargas, Parque Santa Rosa, Parque São José, Passaré, Paupina, Pedras, Pici, Pirambu, Planalto Ayrton Senna, Praia do Futuro, Quintino Cunha, Sabiaguaba, Serriinha, Siqueira, Vicente Pinzón, Vila Manoel Sátiro, Vila Velha.
Cluster 3	Antônio Bezerra, Barra do Ceará, Bom Jardim, Centro, Jangurussu, Messejana, Mondubim, Prefeito José Walter.

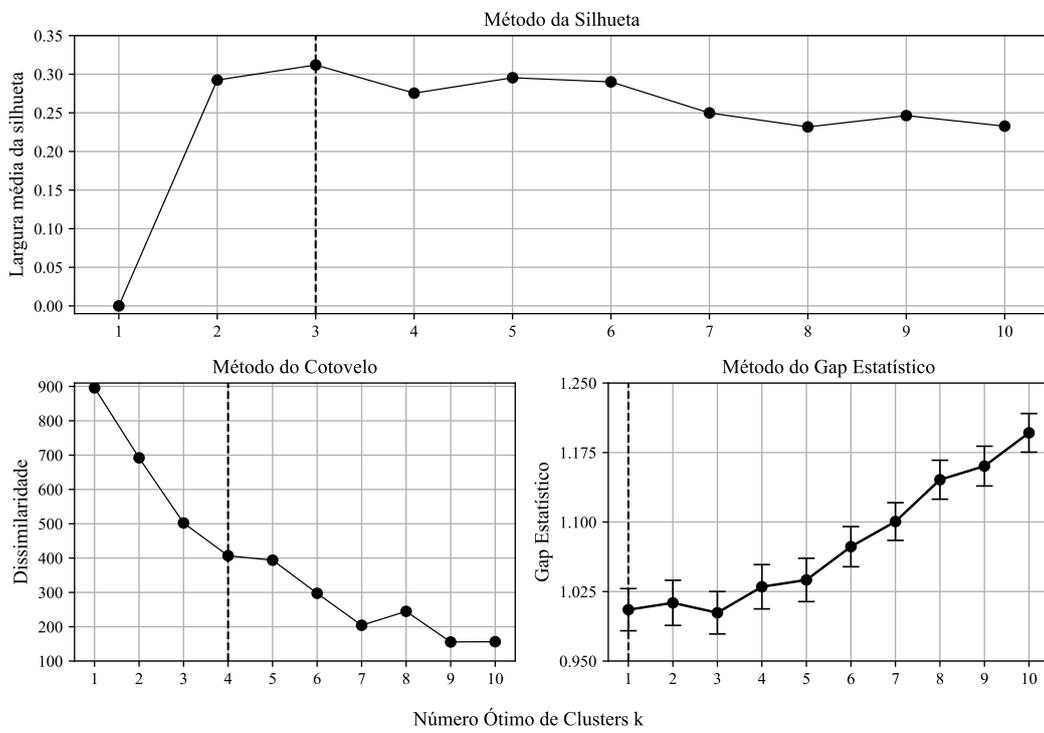
Fonte: Elaboração própria.

### A.1. Resultados para escolha do número ótimo de *clusters* (variáveis em números absolutos).



Fonte: Elaboração própria.

**A.2.** Resultados para escolha do número ótimo de *clusters* (variáveis em taxas por 10 mil hab).



Fonte: Elaboração própria.