

**Brendha Alves Gomes**

**Identificação de clusters persistentes de  
criminalidade no estado do Rio de Janeiro**

Niterói - RJ, Brasil

14 de dezembro de 2022

**Brendha Alves Gomes**

**Identificação de clusters persistentes  
de criminalidade no estado do Rio de  
Janeiro**

**Trabalho de Conclusão de Curso**

Monografia apresentada para obtenção do grau de Bacharel em  
Estatística pela Universidade Federal Fluminense.

Orientador(a): Prof. Dr. Jony Arrais Pinto Junior

Co-Orientador(a): Prof. Dr. Rafael Santos Erbisti

Niterói - RJ, Brasil

14 de dezembro de 2022

**Brendha Alves Gomes**

**Identificação de clusters persistentes de criminalidade no estado do Rio de Janeiro**

Monografia de Projeto Final de Graduação sob o título “*Identificação de clusters persistentes de criminalidade no estado do Rio de Janeiro*”, defendida por Brendha Alves Gomes e aprovada em 14 de dezembro de 2022, na cidade de Niterói, no Estado do Rio de Janeiro, pela banca examinadora constituída pelos professores:

---

**Prof. Dr. Jony Arrais Pinto Junior**  
Departamento de Estatística – UFF

---

**Prof. Dr. Rafael Santos Erbisti**  
Departamento de Estatística – UFF

---

**Prof. Dr. Victor Eduardo Leite de Almeida Duca**  
Departamento de Estatística – UFF

---

**Prof. Dr. Guilherme Augusto Veloso**  
Departamento de Estatística – UFF

Ficha catalográfica automática - SDC/BIME  
Gerada com informações fornecidas pelo autor

G633i Gomes, Brendha Alves  
Identificação de clusters persistentes de criminalidade no estado do Rio de Janeiro / Brendha Alves Gomes. - 2022.  
63 f.

Orientador: Prof. Dr. Jony Arrais Pinto Junior.  
Coorientador: Rafael Santos Erbisti.  
Trabalho de Conclusão de Curso (graduação)-Universidade Federal Fluminense, Instituto de Matemática e Estatística, Niterói, 2022.

1. Estatística Espacial. 2. SaTScan. 3. Criminalidade. 4. Rio de Janeiro. 5. Produção intelectual. I. Pinto Junior, Prof. Dr. Jony Arrais, orientador. II. Erbisti, Rafael Santos, coorientador. III. Universidade Federal Fluminense. Instituto de Matemática e Estatística. IV. Título.

CDD - XXX

# Resumo

A detecção de clusters espaciais ou espaço-temporais têm papel importante para a tomada de decisão das instituições competentes. O trabalho aqui proposto objetiva a detecção e identificação de clusters espaciais na ocorrência de crimes de roubo, furto e letalidade violenta no estado do Rio de Janeiro entre os anos de 2016 a 2020. Para tal, será utilizado a estatística Scan, proposta por Kulldorff (1997). Deseja-se identificar possíveis clusters persistentes em divisões territoriais da base de segurança feita pelo Instituto de Segurança Pública, ISP, que corresponde às menores áreas territoriais de apuração de indicadores de criminalidade, sendo as Circunscrições Integradas de Segurança Pública - CISP. As análises, para cada ano, serão realizadas de forma independente e será analisado se os possíveis clusters identificados apresentam comportamento semelhante ao longo dos trimestres. Dessa forma, foi possível observar que os municípios apresentam comportamentos diferentes em alguns trimestres. Entretanto, na grande maioria das vezes as áreas com risco mais alto de criminalidade, para os três desfechos de interesse, estão concentrados espacialmente na Região Metropolitana do estado.

Palavras-chave: Estatística Espacial. Cluster. SatScan. Criminalidade. Rio de Janeiro.

# Dedicatória

Dedico este trabalho a meus pais. Sem eles, nada disso seria possível.

# Agradecimentos

Agradeço, primeiramente, à Deus por me permitir trilhar esse caminho e chegar até aqui.

Aos meus pais, Edmir e Sandra, por todo o amor e apoio em todos os momentos da graduação, e por terem feito possível e o impossível para que eu chegasse até aqui. Obrigada por serem meu tudo, minha certeza, a melhor parte de mim. Eu não seria nada sem vocês.

À minha família, que sempre me incentivou e torceu por mim. Em especial aos meus irmãos, Wesleyne e Wesley, minha cunhada, Ingrid, meus sobrinhos, Vitor e Gabriela e a família que escolhi, Caroline, Magrão e Pedro. Quando a gente tem amor e aconchego em casa, é sempre mais fácil enfrentar as batalhas.

Ao meu amor, Gabriel, por aguentar todos os meus choros e momentos de desesperos. Obrigada por me fazer mais feliz e por estado ao meu lado.

Aos meus orientadores, Dr. Jony Arrais Pinto Junior e Dr. Rafael Santos Erbisti, por terem aceitado participar do projeto, pela dedicação e disponibilidade e principalmente, pela paciência. Obrigada por toda ajuda e orientação para que eu conseguisse desenvolver este trabalho.

Aos demais professores da Estatística da UFF e de outros departamentos, pela colaboração e pelos conhecimentos transmitidos.

Aos amigos e colegas que conheci na UFF, por tudo que vivemos juntos, obrigada por serem minha família, quando a minha estava longe para serem presentes pessoalmente. Agradeço a minha dupla durante toda a graduação, Wu Xin, e pelos amigos que carrego desde o início do curso, Hillary, Beatriz, Paola e Caio.

Aos meus queridos amigos do ensino médio, principalmente, Flávia, Duda, Julieto e Juliana pelo todo o apoio e por fazerem parte de praticamente todos os momentos da minha vida, sejam de alegria ou tristeza. Obrigada por tudo.

Ao amigos que conheci no estágio, vocês foram fundamentais para o desenvolvimento

deste trabalho e para meu crescimento pessoal. Obrigada por todos os ensinamentos de vida, obrigada por todas as palavras de força e por todo amor que vocês me dão diariamente. Hannah, Maycon, Isadora, Samanta e Thamires, eu amo vocês!

Por fim, à Universidade Federal Fluminense pela infraestrutura e oportunidades na minha formação acadêmica.

# Sumário

## Lista de Figuras

## Lista de Tabelas

<b>1</b>	<b>Introdução</b>	p. 13
1.1	Motivação . . . . .	p. 13
1.2	Objetivos . . . . .	p. 16
1.3	Organização . . . . .	p. 17
<b>2</b>	<b>Materiais e Métodos</b>	p. 18
2.1	Área de Estudo . . . . .	p. 18
2.2	Dados . . . . .	p. 20
2.3	Estatística Espacial . . . . .	p. 21
2.4	Modelo Linear Generalizado . . . . .	p. 23
2.5	Estatística Scan . . . . .	p. 24
2.5.1	Teste da Razão de Verossimilhança . . . . .	p. 26
2.5.1.1	Teste da Razão de Verossimilhança - Modelo Poisson . . . . .	p. 28
2.5.2	Teste de Monte Carlo . . . . .	p. 31
<b>3</b>	<b>Resultados</b>	p. 33
3.1	Apresentação do output do software . . . . .	p. 33
3.2	Análise e Discussão . . . . .	p. 35
3.2.1	Furto . . . . .	p. 35

3.2.2	Letalidade Violenta . . . . .	p. 44
3.2.3	Roubo . . . . .	p. 52
<b>4</b>	<b>Conclusão</b>	<b>p. 60</b>
	<b>Referências</b>	<b>p. 62</b>

# Lista de Figuras

1	Mapa do estado do Rio de Janeiro . . . . .	p. 19
2	Divisão por CISP do Estado do Rio de Janeiro . . . . .	p. 20
3	Exemplo hipotético da varredura espacial do método Scan . . . . .	p. 25
4	Exemplo de saída do Satscan para identificação do cluster primário - Primeiro trimestre de 2016 . . . . .	p. 34
5	Exemplo de saída do Satscan sobre a análise realizada - Primeiro trimes- tre de 2016 . . . . .	p. 34
6	Exemplo de saída do Satscan Teste de Monte Carlo - Primeiro trimestre de 2016 . . . . .	p. 34
7	Cluster de alto risco para casos de furto no estado do Rio de Janeiro em 2016. . . . .	p. 38
8	Cluster de alto risco para casos de furto no estado do Rio de Janeiro em 2017. . . . .	p. 39
9	Cluster de alto risco para casos de furto no estado do Rio de Janeiro em 2018. . . . .	p. 40
10	Cluster de alto risco para casos de furto no estado do Rio de Janeiro em 2019. . . . .	p. 41
11	Cluster de alto risco para casos de furto no estado do Rio de Janeiro em 2020. . . . .	p. 42
12	Persistência dos clustersde alto risco para casos de furto no estado do Rio de Janeiro. . . . .	p. 43
13	Cluster de alto risco para casos de letalidade violenta no estado do Rio de Janeiro em 2016. . . . .	p. 46
14	Cluster de alto risco para casos de letalidade violenta no estado do Rio de Janeiro em 2017. . . . .	p. 47

15	Cluster de alto risco para casos de letalidade violenta no estado do Rio de Janeiro em 2018. . . . .	p. 48
16	Cluster de alto risco para casos de letalidade violenta no estado do Rio de Janeiro em 2019. . . . .	p. 49
17	Cluster de alto risco para casos de letalidade violenta no estado do Rio de Janeiro em 2020. . . . .	p. 50
18	Persistência dos clusters de alto risco para casos de letalidade violenta no estado do Rio de Janeiro. . . . .	p. 51
19	Cluster de alto risco para casos de roubo no estado do Rio de Janeiro em 2016. . . . .	p. 54
20	Cluster de alto risco para casos de roubo no estado do Rio de Janeiro em 2017. . . . .	p. 55
21	Cluster de alto risco para casos de roubo no estado do Rio de Janeiro em 2018. . . . .	p. 56
22	Cluster de alto risco para casos de roubo no estado do Rio de Janeiro em 2019. . . . .	p. 57
23	Cluster de alto risco para casos de roubo no estado do Rio de Janeiro em 2020. . . . .	p. 58
24	Persistência dos clusters de alto risco para casos de roubo no estado do Rio de Janeiro. . . . .	p. 59

# Lista de Tabelas

1	Número de casos de furto no estado do RJ . . . . .	p. 35
2	Razão entre os valores observados e esperados e o risco relativo dos clusters, avaliando o crime de furto. . . . .	p. 36
3	Número de casos de letalidade violenta no Estado do RJ. . . . .	p. 44
4	Razão entre os valores observados e esperados e o risco relativo dos cluster, avaliando o crime de letalidade violenta. . . . .	p. 45
5	Número de casos de roubo no Estado do RJ. . . . .	p. 52
6	Razão entre os valores observados e esperados e o risco relativo dos clusters, avaliando os crimes de roubo. . . . .	p. 53

# 1 Introdução

## 1.1 Motivação

A criminalidade é uma questão que preocupa os cidadãos e os governos de qualquer país, visto que impacta, seja direta ou indiretamente, na qualidade de vida da população. Nesse sentido, os crimes impõem uma série de prejuízos à sociedade, pois carregam consigo custos econômicos e sociais, como danos patrimoniais, riscos à integridade física dos indivíduos ou até mesmo no turismo da região. Nesse sentido, a criminalidade gerou para o Brasil em 2016 uma perda de 373 bilhões de reais, de acordo com Cerqueira e Bueno (2019), o que representa 5,9% do PIB. Somado a isso, há também o aumento dos gastos com segurança pública em certas regiões, impactando ainda mais a economia.

Segundo Congress (1997), o Brasil apresenta níveis acima da média mundial no que se refere a crimes violentos, com destaque para os altos índices de violência armada e homicídios. Nesse contexto, de acordo com dados divulgados pelo Instituto de Segurança Pública (ISP) e analisados por Cerqueira e Bueno (2019), as taxas de homicídio no Brasil vêm aumentando a cada ano, sendo um dos países com maior número absoluto de homicídios do planeta. Diversos fatores podem ser apontados como causas para problemas ligados à segurança pública enfrentados pelo país. Segundo Chesnais (1999), fatores socioeconômicos, infraestrutura precária, explosão demográfica e precarização do sistema judiciário são considerados como principais motivos de aporte para o surgimento da criminalidade.

A literatura ainda aponta algumas evidências que relacionam a desorganização social ao aumento da criminalidade. Os problemas gerados pelas crises e más condições demográficas, urbanização desordenada e questões ligadas à deterioração dos espaços urbanos seriam possíveis causas da criminalidade, pois a violência seria fruto da frustração dos indivíduos, uma vez que foram privados da realização de seus objetivos sociais (PRO-COPIO; TOYOSHIMA, 2017).

No caso específico do Rio de Janeiro, a criminalidade sempre foi considerada um tema recorrente e os problemas relacionados à segurança pública assolam a população há décadas. Mesmo apontado como referência para o turismo, principalmente por suas belezas naturais, ainda existe muito receio em frequentar a região devido os rotineiros noticiários que descrevem a violência do local, atingindo pessoas de diversas classes sociais, níveis educacionais, raças e faixas etárias. O aumento das ocorrências criminosas no Rio de Janeiro tornou inegável a importância dos estudos sobre violência.

O estado possui uma das maiores taxas de homicídio por 100 mil habitantes do país, cerca de 38,4% em 2017, superando a média brasileira no mesmo período, segundo Cerqueira e Bueno (2019). Os municípios fluminenses ainda contam com o problema dos elevados níveis de tráfico de drogas. No caso da cidade do Rio de Janeiro, a grande repressão existente sobre esse tipo de crime, força a migração do mesmo para diferentes zonas, podendo afetar a incidência do crime em diferentes áreas (UCHOA; MENEZES, 2012). Desta forma, os elevados indicadores de criminalidade no Estado do Rio de Janeiro, e o seu crescimento ao longo dos anos, constituem um desafio constante para o poder público.

É possível perceber que a criminalidade se tornou um problema social ao decorrer dos anos, sendo cada vez mais alvo de grande atenção em vários países no mundo. Este é um fenômeno complexo, que pode ser estudado em diversas áreas e também por meio de diferentes abordagens. Em busca de compreender melhor essas ocorrências, a questão espacial passou a ser levada em consideração. A concepção de que os crimes não se distribuem de forma homogênea entre as diferentes localidades abriu margem para o avanço de teorias e de técnicas específicas para verificar as diferentes hipóteses de avaliação de clusters espaciais ((SANTOS et al., 2020a), (OLIVEIRA, 2008))

São diversos tipos de crimes que assolam a população fluminense nos últimos anos, porém mesmo que aparentemente o número de ocorrências aumente a cada ano, há uma queda em alguns indicadores, segundo o Instituto de Segurança Pública através do Atlas de violência publicado por Cerqueira e Bueno (2019). Em 2020 houve a menor taxa de letalidade violenta desde 1991 ao olhar o estado por completo, junto a isso, há uma redução na taxa de ocorrência de roubos e furtos. Porém, em algumas regiões o padrão não se repete, fazendo com que essas localidades sejam assoladas, cada vez mais, pela criminalidade.

Vários estudos nas mais diferentes áreas, se utilizaram da metodologia de avaliação de cluster para identificação de zonas em que o número de ocorrências de um determinado

evento era maior do que o esperado.

Filho et al. (2001) buscou identificar os conglomerados de risco de mortalidade por homicídio mais elevado nos bairros de Belo Horizonte e verificar se existe relação entre o tráfico de drogas e os aglomerados encontrados. Foram detectados conglomerados de alto risco e constatou-se que estes estavam localizados em bairros e favelas em que prevalece o tráfico de drogas e são áreas acometidas pela violência ligada ao tráfico. Esse resultado corrobora com a hipótese inicial do autor de que os homicídios são resultados da violência associada ao mercado de drogas.

Minamisava et al. (2009) analisou a distribuição espacial e potenciais conglomerados de risco para mortes de jovens com idade entre 15 a 24 anos em Goiânia, uma cidade recém-urbanizada na região central do Brasil. Embora não tenham sido detectados clusters significativos de óbitos por associação ao tráfico de drogas, foram encontrados dois aglomerados de óbitos intencionais em duas áreas emergentes e muito pobres na periferia da cidade, podendo indicar associação entre a ocorrência do crime com a qualidade de vida dos cidadãos.

No Rio de Janeiro, apesar do vasto histórico de ocorrências de crimes no estado, são poucos os trabalhos que se dedicaram a analisar e entender o fenômeno da criminalidade e sua distribuição espacial na região. Szwarcwald e Castilho (1998) fizeram uma análise espacial acerca da mortalidade por armas de fogo no estado do Rio de Janeiro no período de 1979 e 1992. A constatação de que a mortalidade era maior nos bolsões de pobreza das metrópoles brasileiras foi refutada, visto que houve expansão para todo o estado.

Santos et al. (2020b) analisou a distribuição espacial e o fenômeno do crime nos municípios fluminenses nos anos de 2006, 2010 e 2016. Foi possível observar que as cidades com maior índice de criminalidade estão concentradas, principalmente, na Região Metropolitana do estado. Os clusters com risco mais baixo localizaram-se principalmente nas regiões Serrana, Centro Fluminense e Noroeste Fluminense. No estudo, percebe-se que o comportamento ao longo dos anos foi semelhante, indicando que as mesmas zonas eram detectadas como cluster de alto risco.

Outros estudos com a mesma metodologia foram realizados na área epidemiológica, focando na detecção de cluster de doenças a fim de identificar zonas em que a doença apresenta comportamento semelhante, e assim permitindo que agentes possam atuar de maneira mais eficaz. Bacani (2013) buscou identificar padrões espaciais da ocorrência de malária no estado do Mato Grosso durante os anos de 2003 a 2009. Foi feita a identificação de aglomerados de baixo, médio e alto risco e também aglomerados espaço-temporais. Os

municípios da microrregião de Aripuanã, região Noroeste do Estado, apresentaram-se como conglomerados de alto risco em todos os anos de estudo. Já Lewkowitz (2013) aplicou técnicas de Estatística Espacial para identificar regiões espaciais e temporais em que havia maior ocorrência de casos de dengue nos municípios brasileiros de 2007 a 2011. Concluiu-se que os casos de dengue não estão distribuídos aleatoriamente pelo mapa e foi possível verificar também que a incidência da doença é maior no primeiro semestre do ano, de janeiro a julho. Dessa forma, as autoridades competentes enxergam melhor a dinâmica da doença no país, possibilitando ações de combate mais eficientes.

A detecção de clusters espaciais tem papel importante para a decisão das instituições competentes, uma vez que permite que as áreas mais atingidas sejam identificadas e possam receber maior atenção. Dentro desse contexto, tem-se o interesse em detectar zonas do mapa onde a ocorrência de crimes de roubo, letalidade violenta e furto sejam significativamente maiores do que em outras zonas, ou seja, verificar se esses crimes estão distribuídos de forma aleatória ou se seguem um padrão, gerando um aglomerado de casos em alguma região do mapa. Junto a isso, deseja-se verificar se o possível padrão de agrupamento encontrado persiste, ou não, ao longo do tempo.

Portanto, destaca-se a relevância de compreender as dinâmicas espaciais e a identificação do perfil da criminalidade no estado do Rio de Janeiro. Ao analisar o impacto que a criminalidade traz para a população e para a economia, é imprescindível que o comportamento da distribuição espacial de crimes seja estudado, a fim de auxiliar o planejamento e execução de políticas, aumentando a eficiência de combate à criminalidade e permitindo que os gestores de segurança pública consigam direcionar políticas para as diferentes localidades de forma mais eficaz.

## 1.2 **Objetivos**

O presente trabalho tem como objetivo principal identificar a existência de clusters persistentes para crimes de roubo, letalidade violenta e furto no estado do Rio de Janeiro. Para tal, o espaço será dividido em áreas, de acordo com uma divisão territorial definida no Capítulo 2, e serão considerados os anos de 2016 a 2020, sendo analisados de forma independente por trimestre. Deste modo, será verificado se os clusters encontrados persistem nos diferentes períodos analisados, ou seja, se as zonas onde são encontrados os aglomerados são as mesmas ao longo do período analisado, formando clusters espaciais persistentes.

Detectada a presença de algum aglomerado, é de interesse identificar a janela espacial que a contém, assim como verificar sua significância estatística. Dessa maneira, é possível que os gestores de segurança pública consigam criar políticas mais eficazes nas cidades do Rio de Janeiro.

## 1.3 Organização

Este trabalho está organizado como descrito a seguir. No Capítulo 2, encontram-se os ‘Materiais e Métodos’, dividido em 5 seções, onde será apresentada a região de estudo, a natureza dos dados a serem estudados e a metodologia estatística aplicada para as análises. No Capítulo 3, encontram-se as ‘Análise de Resultados’ onde são apresentados os resultados da pesquisa com suas respectivas interpretações. No Capítulo 4, está a ‘Conclusão’, em que será feito um breve resumo do estudo e dos resultados encontrados neste trabalho.

## 2 Materiais e Métodos

Este capítulo contém uma apresentação sucinta da metodologia que será usada neste trabalho. Nele será apresentada a área de estudo de interesse e os materiais utilizados para a realização das análises e também os métodos estatísticos a serem implementados para alcançar os objetivos deste estudo.

### 2.1 Área de Estudo

A área de estudo compreende o estado do Rio de Janeiro localizado na região Sudeste do Brasil. O território estadual do Rio de Janeiro é o menor em extensão entre os estados da região Sudeste, dividido em 92 cidades. Segundo dados do Censo Demográfico de 2010, o estado é o terceiro mais populoso do Brasil, atrás apenas de São Paulo e Minas Gerais. A estimativa populacional calculada pelo IBGE foi de 17.463.349 habitantes, sendo essa população distribuída ao longo de sua área de aproximadamente 43.696 km<sup>2</sup>. Na Figura 1 é possível observar a divisão territorial do estado do Rio de Janeiro.

O Instituto de Segurança Pública (ISP) divulga dados e estatísticas referentes aos crimes realizados no Estado do Rio de Janeiro em uma frequência mensal, e com o objetivo de aperfeiçoar as ações de segurança pública do estado do Rio de Janeiro, foram criadas divisões territoriais que são as Regiões Integradas de Segurança Pública (RISP) e as Circunscrições Integradas de Segurança Pública (CISP), complementando as esferas da divisão territorial junto às Áreas Integradas de Segurança Pública (AISP) (ISP, 2011).

Os dados são aferidos, considerando quatro níveis geográficos distintos. Do mais amplo para o mais reduzido, têm-se: o estado do Rio de Janeiro; as Regiões Integradas de Segurança Pública (RISPs); as Áreas Integradas de Segurança Pública (AISPs) e as Circunscrições Integradas de Segurança Pública (CISPs). A divisão foi feita para auxiliar na integração do planejamento e coordenação operacional das organizações policiais, a fim de obter maior efetividade das ações operacionais em áreas próximas (RUEDIGER, 2019).

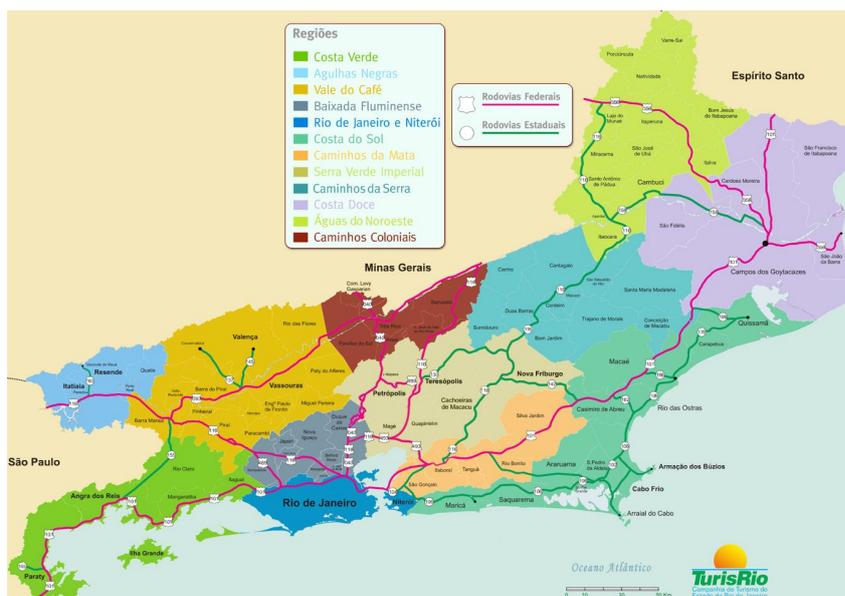


Figura 1: Mapa do estado do Rio de Janeiro (TURISRIO, 2020)

Sumariamente, as CISPs correspondem à jurisdição de uma Delegacia de Polícia (DP). As AISP são áreas sob responsabilidade de um Batalhão de Polícia Militar (BPM). Já uma RISP é uma região mais abrangente, criada para coordenar a atuação da Polícia Militar, responsável pelo policiamento ostensivo, com a da Polícia Civil, responsável pelas investigações criminais. Atualmente, o estado se encontra dividido em 7 RISPs, 41 AISPs e 137 CISPs.

As Circunscrições Integradas de Segurança Pública, correspondem às menores áreas territoriais de apuração dos indicadores de criminalidade. As CISPs têm como princípio básico o conceito de que a responsabilidade pelo policiamento de uma subárea da companhia de Polícia Militar deverá coincidir com a circunscrição de uma delegacia de Polícia Civil. Nesse sentido, as CISPs constituem a esfera de integração territorial, em nível operacional, das Companhias Integradas da Secretaria de Estado de Polícia Militar, ou SEPM, com as Delegacias de Polícia da Secretaria Estadual de Polícia Civil, ou SEPOL (ISP, 2011).

A divisão do território no nível de Circunscrições Integradas de Segurança Pública não corresponde sempre a divisão territorial de um município, pois uma mesma localidade pode conter mais de uma CISP. Por exemplo, a cidade de Duque de Caxias engloba quatro CISPs (59, 60, 61 e 62), enquanto o município de Volta Redonda abrange uma CISP (93) que atende toda a cidade. A Figura 2 apresenta a divisão territorial do estado do Rio de Janeiro pela CISP.

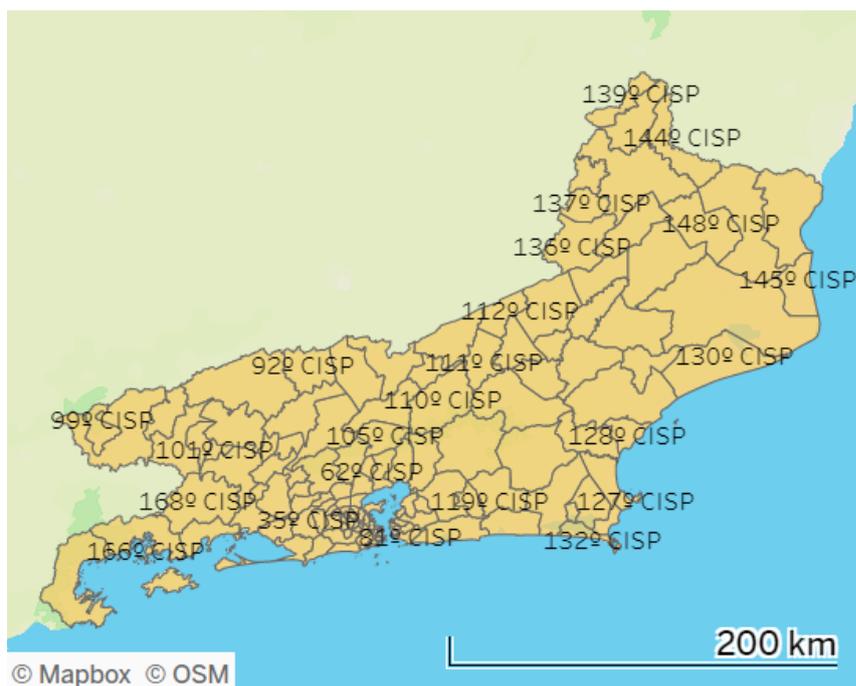


Figura 2: Divisão por CISP do Estado do Rio de Janeiro (ISP, 2011)

## 2.2 Dados

O Instituto de Segurança Pública (ISP) realiza análises e pesquisas referentes à segurança pública no estado do Rio de Janeiro, visando auxiliar na gestão de segurança pública. Nesse contexto, o ISP possui um banco de dados, desde de 2003, onde estão compiladas as diferentes ocorrências de crime no estado.

Neste trabalho, os três desfechos de interesse são: letalidade violenta, roubo e furtos. Para esses casos, a base contém a CISP em que ocorreu o crime e a data em que o crime foi registrado. Dessa forma, é possível contar o número de casos de interesse por CISP, e avaliá-los em uma determinada janela de tempo.

Os crimes de letalidade violenta englobam homicídio doloso, lesão seguida de morte, latrocínio e morte pela polícia, estão referenciados na base como “vida-ação” e “vida”. Os crimes referentes a roubos, que consiste na apropriação indevida de um bem alheio de maneira violenta, estão na base de dados como “Patrimônio-violento-móvel”, “Patrimônio-violento-rua” e “Patrimônio - Violento - Fixo”. Por fim, o crime de furto, que consiste também na apropriação de um bem alheio, porém sem o uso da violência, está na base de dados como “Patrimônio-sem violência-móvel”, “Patrimônio - Sem violência - Rua” e “Patrimônio - Sem violência - Fixo”.

## 2.3 Estatística Espacial

Em diversos estudos, um dos principais interesses consiste em investigar os dados disponíveis, verificando, por exemplo, possíveis relações entre eles. Todos os dados possuem referência temporal e espacial, e em alguns casos existe relevância em observar o padrão de distribuição dessa ocorrência ao longo da região de estudo. Métodos que possibilitem detectar um conjunto de zonas que tenham uma ocorrência maior, ou menor, de um determinado evento são de extrema importância. À título de exemplo, uma metodologia que aponte a existência de bairros que tenha uma maior quantidade de casos de crime ou de dengue de uma cidade. Nesse sentido, diversas técnicas estatísticas podem ser empregadas, técnicas, essas, que pertencem à área de Estatística Espacial.

A Estatística Espacial é um ramo que estuda métodos estatísticos para coleta, descrição, visualização, modelagem e análise de dados que possuem coordenadas geográficas. Essa análise é utilizada quando os dados são espacialmente localizados e é explicitada a importância da disposição espacial na interpretação e análise dos dados.

O interesse dessa análise é centrada nos processos que ocorrem no espaço, e as técnicas a serem aplicadas buscam compreender o comportamento das ocorrências de um evento de interesse e realizar inferências sobre o comportamento a respeito desses fenômenos. Sendo assim, a Estatística Espacial é utilizada quando se deseja investigar a existência de padrões espaciais na distribuição dos dados em uma determinada localidade, e avaliar a significância desses padrões. Nesse contexto, é de interesse avaliar se a dependência espacial existe ou se as ocorrências de um determinado fenômeno são puramente aleatórias.

Os diferentes tipos de dados espaciais são tradicionalmente classificados de acordo com sua tipologia. Essa caracterização diz respeito à natureza da observação. De acordo com Cressie (1993), a Estatística Espacial pode ser dividida em 3 grandes áreas:

- Dados de processos pontuais,
- Dados de superfícies contínuas,
- Dados de área.

Os dados de processos pontuais são fenômenos expressos por meio de ocorrências identificadas como pontos localizados no espaço geográfico. É definido como um conjunto de dados consistindo de uma série de localizações pontuais (coordenadas) que indicam a ocorrência de eventos de interesse na área de estudo. O objetivo é avaliar a distribuição

espacial dos pontos, testando hipóteses sobre o padrão observado. Por exemplo, se o padrão de pontos está distribuído de forma aleatória ou se existe um padrão espacial. Um exemplo para esse tipo de estudo são casos de câncer que ocorreram nas proximidades de instalações de uma indústria de energia nuclear.

Já os dados de superfícies contínuas referem-se a fenômenos contínuos no espaço que são mensurados em estações de monitoramento, isto é, um conjunto de localizações fixas no espaço. Neste caso, tem-se o interesse em avaliar e compreender o comportamento de uma determinada variável sobre uma determinada região de interesse, dados valores conhecidos em algumas localizações, e também modelar e estimar valores em localidades não abrangidas pelo estudo. O conceito básico é de que as observações próximas no espaço geográfico tendem a possuir valores de atributo semelhantes, e à medida que a distância vai aumentando, essa similaridade vai diminuindo. Uma análise que pode ser feita é realizar previsões de medidas pluviométricas da cidade do Rio de Janeiro, a partir do estudo de apenas alguns bairros.

Nos dados de área, o espaço é particionado em áreas, como bairros ou municípios, em que acredita-se que o evento de interesse se comporte de maneira homogênea. Estas repartições podem representar divisões de caráter político ou administrativo ou feitas com base em características geofísicas. Para cada uma dessas áreas tem-se informações agregadas de um determinado fenômeno (geralmente, médias ou somas) e o interesse é entender e explicar as estruturas espaciais nos valores observados. Uma aplicação é no estudo do número de jovens assassinados pela UF (Brasil) em 2021.

A forma usual de apresentação de dados de área é por meio de um mapa coroplético, onde a representação visual empregada deve mostrar as quantidades correspondentes às diferentes áreas, sendo feita a partir da associação do maior valor de atributo com a cor mais escura, o que estabelece uma ordem crescente que deve ser seguida na variação das cores.

Neste trabalho será utilizado o processo de dados de contagem agregados por regiões, ou dados de área, a fim de avaliar o número de ocorrências de roubos, letalidades violentas e furtos nas CISPs do estado do Rio de Janeiro, definidas anteriormente, ao longo de um período de tempo de maneira independente. Assim, serão calculados os números de ocorrências dos crimes de furto, roubo e letalidade violenta para cada uma das 137 CISPs e verificada a existência de clusters persistentes no decorrer dos trimestres por meio de uma análise de cluster espacial.

## 2.4 Modelo Linear Generalizado

A análise de regressão tem como principal objetivo verificar se o comportamento de uma variável  $Y$ , chamada de variável dependente ou resposta, é explicada pela associação dela com outras variáveis, comumente chamadas de variáveis explicativas. A maneira mais usual de abordar tal problema é por meio de um estudo de um modelo de regressão que irá relacionar as variáveis, e verificar a relação existente entre a variável resposta e uma, ou mais, variáveis explicativas.

Os modelos lineares generalizados, ou MLG's, englobam distribuições pertencentes à família exponencial e podem ser utilizados para diferentes tipos de variáveis, sejam elas dados de contagem, binárias, contínuas entre outras, mudando os pressupostos e características de acordo com  $Y$ . Para este estudo, outras variáveis poderiam ser inseridas no modelo de regressão, para posteriormente ser verificada a relação entre elas, porém, não será realizada essa análise.

De forma geral, a função de densidade de uma variável aleatória na classe dos modelos lineares generalizados pode ser descrita da seguinte forma:

$$f(\mathbf{y}; \theta) = \exp[a(y)b(\theta) + c(\theta) + d(y)], \quad (2.1)$$

em que  $a$ ,  $b$ ,  $c$  e  $d$  são funções conhecidas e  $\theta$  é o parâmetro desconhecido da variável.

Além disso, os MLG's são compostos por três componentes:

1. **Componente aleatório:** Variáveis respostas  $Y_1, \dots, Y_n$  que possuem a mesma distribuição pertencente à família exponencial

2. **Componente Sistemático:** Um conjunto de parâmetros  $\beta$  e variáveis explicativas (preditor linear)

$$\eta_i = X^T \beta, \quad (2.2)$$

em que  $X$  é uma matriz coluna de variáveis explicativas.

3. **Função de Ligação:** Uma função monótona e diferenciável que irá relacionar a média da variável com o preditor linear tal que:

$$g(\mu_i) = \eta_i = x_i^T \beta, \quad (2.3)$$

em que  $\mu_i = E(Y_i)$

Assim, para se especificar um modelo têm-se os três componentes. Para este estudo

será definido, mais a frente, a distribuição de Poisson, modelo usualmente utilizado para modelar dados de contagem ou de frequência.

## 2.5 Estatística Scan

A análise espacial pode detectar áreas, independentemente do tamanho, que apresentam diferenças significativas de risco, sendo um dos principais métodos para determinar o grau de dependência espacial entre as observações. Os métodos de detecção de clusters espaciais de algum evento de interesse teve origem no trabalho de Naus (1965), sendo posteriormente desenvolvidos os métodos de varredura espacial por Kulldorff (1997), que estabelecem uma metodologia para a identificação de clusters.

As análises da metodologia de estatística Scan, ou também chamada de estatística de varredura, podem ser feitas em diferentes dimensões: puramente espacial, puramente temporal ou espacial-temporal. No teste puramente espacial é avaliado a concentração de casos em determinada zona delimitada no mapa, onde o interesse é analisar apenas a disposição geográfica dos eventos e os clusters formados. No teste puramente temporal, o interesse é a análise de janelas de tempo dentro do qual o número de casos observados é maior do que o esperado. Por fim, o teste espaço-temporal é um método para avaliar se os casos que estão próximos no espaço estão próximos também na janela temporal e vice-versa. O foco deste trabalho se dá na análise de cluster puramente espacial.

O método de varredura espacial é normalmente aplicado para testar se a localização dos eventos de interesse na região  $D$  são puramente aleatórios ou se algum conglomerado pode ser localizado, além de verificar a significância estatística dos clusters encontrados. Esta metodologia é utilizada para confirmação estatística dos agrupamentos, os quais correspondem a um conjunto de áreas que apresentam um risco significativamente alto em relação a um determinado evento (FILHO et al., 2001). Este problema é bastante comum nas mais diversas áreas do conhecimento, como epidemiologia, demografia, psicologia, medicina veterinária, criminologia, detecção precoce de epidemias, entre outras.

A estatística Scan procura identificar uma zona da região de estudo em que a incidência de um determinado evento seja maior do que se comparada à todas as outras zonas do mapa, relacionando o número de casos observados em relação aos esperados. Para obtenção dos clusters na análise puramente espacial, a metodologia aplicada cria um número ilimitado de janelas circulares no mapa da área  $D$  de interesse. As janelas são centradas em cada um dos vários pontos de uma grade posicionada no centroide da

região, e irão variar em tamanho, do menor até um limite superior, com o raio do círculo aumentando até esse tamanho máximo definido. Assim, são identificados agrupamentos para diferentes raios de varredura, sendo que essas zonas terão uma porcentagem máxima da população sob risco definido pelo usuário, de modo a não incluir mais do que 50% da população em seu interior. Cada janela é avaliada como um possível candidato a cluster.

Para a especificação da metodologia discutida nas subseções a seguir, considere que  $Z$ , denota o conjunto das  $z$  áreas candidatas a formarem um aglomerado. Estes  $z$  candidatos são janelas/círculos de raios  $r$  arbitrários situados em um centróide.

A Figura 3 exemplifica o método de estatística de varredura espacial. O raio do círculo vai aumentando, e novas áreas vão sendo criadas até um tamanho máximo. As grades, onde estariam posicionadas as janelas, são os centroides do cluster definidos como  $\xi_i$ . Além disso, os casos observados seriam o número de ocorrências do evento de interesse que houve naquela zona.

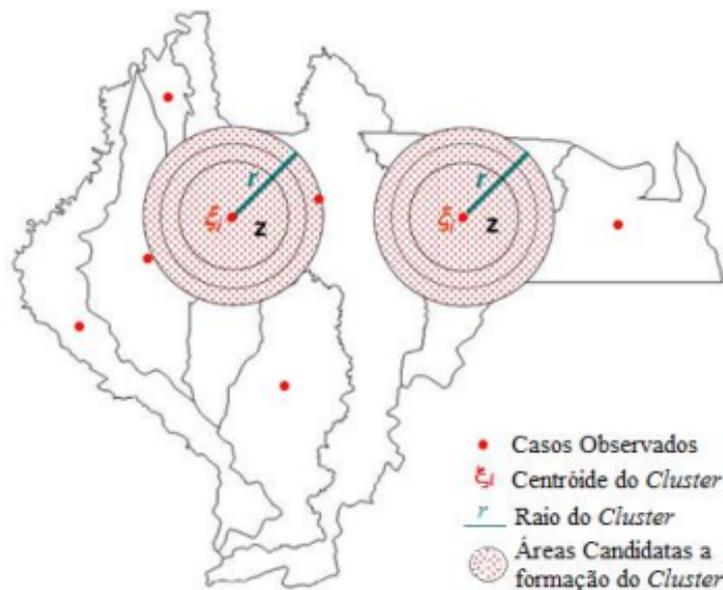


Figura 3: Exemplo hipotético da varredura espacial do método Scan (MORAES, 2010)

Para cada cluster identificado, é necessário verificar se ele é o de risco mais elevado e para isso é feito o teste de razão de verossimilhança, que irá comparar as verossimilhanças dos agrupamentos e selecionar o mais alto. A partir daí, é aplicado o método de Monte Carlo para verificar a significância do cluster selecionado. Os testes serão definidos formalmente nas próximas seções.

Além disso, os testes estatísticos podem ser segmentados em duas categorias para identificação de clusters de risco mais elevado: os focados e os genéricos (KULLDORFF;

LAWSON, 1999).

- Testes focados: verificam a existência de conglomerados em poucas regiões definidas antes das observações dos eventos, ou seja, é determinado, *a priori*, as possíveis localidades em que os conglomerados poderiam estar.
- Testes genéricos: não há uma suposição prévia de um local em específico, sendo verificado em toda a região de estudo os possíveis cluster de risco mais elevado, ou seja, eles testam: se não existe conglomerado na região de estudo ou se existe conglomerado na região do estudo. Sem especificação de onde esse cluster estaria localizado.

Neste estudo, serão usados apenas os testes genéricos e focados na aplicação destas técnicas na área criminológica com uma análise puramente espacial.

### 2.5.1 Teste da Razão de Verossimilhança

A partir do método de Estatística Scan são encontradas as zonas candidatas a clusters. Porém, apenas algumas delas podem ser consideradas significativas, sendo de interesse a identificação de trechos que apresentam uma intensidade maior do que a esperada. Para identificar essas áreas, é testada a hipótese de que o *cluster* ocorreu ao acaso contra a hipótese de que existe uma dependência espacial na ocorrência. O teste utilizado para esta finalidade é o da razão da verossimilhança.

No teste da razão de verossimilhança, para cada círculo encontrado na Estatística Scan, calcula-se a razão da verossimilhança, de acordo com o modelo proposto no estudo, e seleciona-se o que apresenta o valor mais alto. A utilização da razão de verossimilhança é para equilibrar o quão grande é o risco dentro e fora do *cluster* ao se comparar com o risco calculado na condição em que não há *cluster*. Assim, o cálculo é feito levando em conta o número de casos dentro e fora dos círculos com a quantidade de eventos esperados, e a zona que excede o limite de casos esperados, é apontado como o agrupamento mais verossímil, sendo o principal candidato a *cluster* espacial (KULLDORFF et al., 1998).

Para cada *cluster*  $d$ , calcula-se o número de casos esperados. Seja  $C$ ,  $N$  e  $n_z$  o número de casos totais, o tamanho populacional e número de indivíduos na zona  $z$ , respectivamente. Usando um modelo de Poisson, o número de casos esperados dentro de cada zona dados por:

$$\mu_z = C \frac{n_z}{N}. \quad (2.4)$$

O teste é feito, assumindo sob hipótese nula, a não existência de *clusters* espaciais na região de estudo e sob hipótese alternativa assumindo a existência de uma zona tal que o risco nela é significativamente maior do que na região de estudo, ou seja, a probabilidade de um caso ocorrer dentro de  $z$  é maior do que a probabilidade de um caso ocorrer fora de  $z$ . Portanto, definimos as seguintes hipóteses:

$$H_0 : p = q \quad \text{vs} \quad H_1 : p > q,$$

Em que  $p$  é a probabilidade de uma determinada observação ser um caso dentro da zona  $z$  e  $q$  a probabilidade de uma observação ser um caso fora da zona  $z$ . Então, pode-se concluir que a zona  $z$  será considerada um *cluster* caso o teste rejeite a hipótese nula.

A estatística de teste da razão de verossimilhança LR é calculada a partir da equação:

$$LR(z) = \frac{\text{Modelo Sob } H_1}{\text{Modelo Sob } H_0} \rightarrow \frac{\max_{z, H_1} L(Z, p_1, \dots, p_k, q_1, \dots, q_k)}{\max_{z, H_0} L(Z, p_1, \dots, p_k, q_1, \dots, q_k)} \rightarrow \max_z \frac{L(z)}{L_0}, \quad (2.5)$$

em que  $L(z)$  é a verossimilhança calculada considerando que a existência de um *cluster* espacial e  $L_0$  é a verossimilhança sob  $H_0$ , ou seja, quando não supõe a existência de um *cluster* de risco.

Normalmente, pelo fato da razão do teste de razão de verossimilhança crescer muito rápido, é utilizado o seu logaritmo, isto é,  $LLR = \log(LR)$ . Portanto, a estatística do teste é dada por:

$$T = \max_z LLR \quad (2.6)$$

A função de verossimilhança é maximizada sobre todas as zonas e o *cluster* mais provável é aquele com maior verossimilhança, ou seja, o menos provável de ter ocorrido ao acaso.

Resumindo, o procedimento do teste de razão de verossimilhança consiste em partir de uma região e ir adicionando, a cada iteração, a região com a menor distância da primeira. Assim, partindo da região  $r_1$ , fazemos com que  $z_1 = r_1$ . Calcula-se o valor de  $T = LLR(z_1)$ . Em seguida, considera-se  $z_2 = (r_1, r_2)$ , onde  $r_2$  é a região mais próxima de  $r_1$ . Calcula-se o valor de  $LLR(z_2)$  e compara-se com o valor de  $T$ , sendo este último substituído por  $LLR(z_2)$  caso  $LLR(z_2) > T$ .

As regiões vão sendo acrescentadas às zonas, de acordo com a proximidade da região inicial, até que se atinja uma zona que contenha uma porcentagem específica da população

definida pelo usuário. Em seguida, o procedimento é reiniciado, partindo de cada uma das regiões do mapa. Ao final, o candidato a *cluster* mais verossímil é aquele cujo *LLR* foi maior.

A razão de verossimilhança para esse círculo constitui a estatística de teste do método, porém, a distribuição da estatística de teste  $T$  é desconhecida, fazendo-se necessário o uso do método de Monte Carlo, processo de simulação que será descrito na seção adiante, para o cálculo da significância do *cluster*.

### 2.5.1.1 Teste da Razão de Verossimilhança - Modelo Poisson

Neste trabalho, o modelo proposto para a variável aleatória e posteriormente, definição da verossimilhança, é o de Poisson, variável muito utilizada para dados de contagem ou também de taxas (FARAWAY, 2006). Dessa forma, o número de ocorrências dos crimes de roubo, furto e letalidade violenta para cada CISP do estado do Rio de Janeiro ao longo dos anos de 2016 a 2020, avaliados de forma independente, irá seguir uma distribuição de Poisson e sob a hipótese nula, e quando não há covariáveis, o número esperado de casos em cada CISP será proporcional ao tamanho de sua população nessa área.

Considerando  $Y$  como o número de ocorrências de um determinado evento, a função de densidade e o valor esperado do modelo de Poisson podem ser descritos da seguinte forma:

- Função de densidade:

$$f(y) = \frac{(\lambda^y \exp^{-\lambda})}{y!}, y = 0, 1, \dots \quad (2.7)$$

- Valor esperado

$$E(Y) = \lambda \quad (2.8)$$

em que  $\lambda$  representa a taxa média de ocorrência do evento.

Para definição da razão de verossimilhança para o modelo de Poisson, considere as seguintes notações:

- $N$  = Número total de indivíduos na população,
- $C$  = Número total de casos observados na região,

- $Z$  = Conjunto de todas as combinações possíveis de regiões,
- $z$  = Conjunto de zonas (regiões específicas),
- $n_z$  = População na zona  $z$ ,
- $\mu_z$  = Número total de casos esperados na região  $z$ ,
- $p$  = Probabilidade de uma determinada observação ser um caso dentro da zona  $z$ ,
- $q$  = Probabilidade de uma observação ser um caso fora da zona  $z$ .

Será utilizado o modelo de Poisson para descrever a distribuição do número de casos observados nas regiões do mapa. Considere que o número de total de casos na zona  $z$ ,  $C_z$ , é uma variável aleatória com distribuição de Poisson, cujo parâmetro  $\lambda_z$  é tal que  $\lambda_z = pn_z$ . Assim:

$$C_z \sim Poi(\lambda_z)$$

Portanto, temos que a função de densidade de casos  $C_z$  é dada por:

$$f_z(c) = \frac{e^{-\lambda_z} \lambda_z^c}{c!}, c \geq 0. \quad (2.9)$$

A partir das definições feitas no início da seção, pode-se determinar o tamanho da população fora da zona  $z$ ,  $N_{\bar{z}}$ , e o número de casos fora da zona,  $C_{\bar{z}}$ , da seguinte maneira:

$$N_{\bar{z}} = N - n_z \quad (2.10)$$

$$C_{\bar{z}} = C - c_z \quad (2.11)$$

Como visto na seção anterior, considerando a não existência de conglomerados ( $p = q$ ) tem-se que:

$$\mu_z = pn_z \rightarrow \mu_{\bar{z}} = p(N - n_z) \quad (2.12)$$

Assim, expressa-se a verossimilhança por:

$$\begin{aligned} L_0(z; p) &= \frac{\mu_z^{c_z} e^{-\mu_z}}{c_z} \frac{\mu_{\bar{z}}^{c_{\bar{z}}} e^{-\mu_{\bar{z}}}}{c_{\bar{z}}} \\ &= \frac{(pn_z)^{c_z} e^{-pN_z}}{c_z} \frac{(p(N - n_z))^{C - c_z} e^{-p(N - n_z)}}{(C - c_z)!}. \end{aligned} \quad (2.13)$$

Para encontrar o estimador de máxima verossimilhança para  $p$ , primeiro define-se a

função de log-verossimilhança da Poisson:

$$\begin{aligned} l_0(z; p) &= \log[L_0(z; p)] \\ &= c_z[\log p + \log n_z] - pn_z - \log c_z! + (C - c_z)[\log p + \log(N - n_z)] \\ &\quad - p(N - n_z) - \log[(C - c_z)!]. \end{aligned} \quad (2.14)$$

Aplica-se a derivada de 2.14 uma vez que deseja-se encontrar o ponto que maximiza a função de verossimilhança. Logo:

$$\frac{\partial l_0(z; p)}{\partial p} = \frac{c_z}{p} - n_z + \frac{C - c_z}{p} - (N - n_z) = 0 \rightarrow \frac{C}{p} - N = 0 \rightarrow p = \frac{C}{N}. \quad (2.15)$$

Agora substitui-se  $p = \frac{C}{N}$  em  $L_0(z; p)$ :

$$L_0(z) = \frac{\frac{C}{N} n_z e^{-\frac{C}{N} n_z}}{c_z!} \frac{(\frac{C}{N} (N - n_z))^{C - c_z} e^{-\frac{C}{N} (N - n_z)}}{(C - c_z)!}. \quad (2.16)$$

Considerando  $\mu_z = \frac{C}{N} n_z$ , é possível reescrever 2.16 da seguinte forma:

$$L_0(z) = \frac{\mu_z^{c_z} e^{-\mu_z}}{c_z!} \frac{(C - \mu_z)^{C - c_z} e^{-C}}{c_z! (C - c_z)!} = \frac{\mu_z^{c_z} (C - \mu_z)^{C - c_z} e^{-C}}{c_z! (C - c_z)!}. \quad (2.17)$$

Sob a suposição de existência de conglomerado, tem-se que a probabilidade de que um indivíduo venha a ser um caso dentro da zona  $z$  é maior do que fora:

$$\mu_z = pn_z \rightarrow \mu_{\bar{z}} = q(N - n_z), p \geq q. \quad (2.18)$$

Dessa forma, o cálculo da função de verossimilhança sob  $H_1$  é:

$$L(z; p; q) = \frac{(pn_z)^{c_z} e^{-pn_z} [q(N - n_z)]^{C - c_z} e^{-q(N - n_z)}}{c_z! (C - c_z)!}. \quad (2.19)$$

Calcula-se, agora, a função de log-verossimilhança sob a hipótese  $H_1$ :

$$l(z; p; q) = \log L(z; p; q) \quad (2.20)$$

$$\begin{aligned} &= c_z[\log p + \log n_z] - pn_z - \log c_z! + (C - c_z)[\log q + \log(N - n_z)] - q(N - n_z) \\ &\quad - \log[(C - c_z)!]. \end{aligned} \quad (2.21)$$

Novamente, maximizando 2.20 :

$$\frac{\partial l(z; p; q)}{\partial p} = \frac{c_z}{p} - n_z = 0 \rightarrow p = \frac{c_z}{n_z}. \quad (2.22)$$

$$\frac{\partial l(z; p; q)}{\partial q} = \frac{C - c_z}{q} - (N - n_z) = 0 \rightarrow q = \frac{C - c_z}{N - n_z}. \quad (2.23)$$

Agora, basta aplicar  $p = \frac{c_z}{n_z}$  e  $q = \frac{C - c_z}{N - n_z}$  em  $L(z; p; q)$ :

$$L(z) = \frac{c_z^{c_z} e^{-c_z}}{c_z!} \frac{(C - c_z)^{(C - c_z)} e^{-(C - c_z)}}{(C - c_z)!} = \frac{c_z^{c_z} (C - c_z)^{(C - c_z)} e^{-C}}{c_z! (C - c_z)!}.$$

Portanto, encontra-se a seguinte razão de verossimilhança para o modelo de Poisson descrito anteriormente:

$$LR(z) = \frac{L}{L_0} = \begin{cases} \left(\frac{c_z}{\mu_z}\right)^{c_z} \left(\frac{C - c_z}{C - \mu_z}\right)^{C - c_z} & , \text{ se } c_z > \mu_z \\ 1 & , \text{ caso contrário.} \end{cases} \quad (2.24)$$

A partir da verossimilhança encontrada para o Modelo de Poisson, basta aplicar no teste de razão de verossimilhança para encontrar os valores da estatística de teste  $T$  para cada zona  $z$  encontrada na região de estudo. Pode-se definir também o logaritmo da função de verossimilhança, ou seja,  $LLR = \log LR(z)$ :

$$LLR(z) = \log \frac{L}{L_0} = \begin{cases} c_z(\log c_z - \log \mu_z) + (C - c_z)[\log(C - c_z) - \log(C - \mu_z)] & , \text{ se } c_z > \mu_z \\ 0 & , \text{ caso contrário.} \end{cases} \quad (2.25)$$

A partir de  $LLR(z)$  obtém-se a estatística  $T$  do teste de razão de verossimilhança e é possível notar que o cálculo da estatística de teste se concentra na comparação entre o valor observado de casos com o valor esperado de casos dentro da região analisada em contraste com o cálculo da estatística  $T$  para fora da região estudada.

## 2.5.2 Teste de Monte Carlo

A metodologia discutida a seguir, visa encontrar uma forma de avaliar a significância da estatística de teste  $T$ . Entretanto, para testar se o conglomerado é significativo a um dado nível de confiança, precisa-se conhecer sua distribuição considerando a não existência de conglomerados, porém a distribuição da estatística de teste não é conhecida. Dessa forma, faz-se necessário a aplicação do teste de Monte Carlo, sendo uma forma de conduzir

o teste quando não é possível determinar a distribuição empírica de  $T$  (METROPOLIS, 1949).

O teste de Monte Carlo consiste em repetir o mesmo procedimento analítico em um grande número de replicações aleatórias do conjunto de dados geradas sob a hipótese de não existência de cluster. Para o cálculo do p-valor para os clusters detectados, faz-se o uso de simulações, com a quantidade a ser especificada, para gerar um número de replicações aleatórias do conjunto de dados. Se a razão do teste de máxima verossimilhança calculada para o cluster mais provável no conjunto de dados reais for elevada em comparação às razões das máximas verossimilhanças calculadas para os clusters mais prováveis nos conjuntos de dados aleatórios, há evidências para que a hipótese nula seja rejeitada, isto é, evidências a favor da hipótese em que se considera que os casos observados não estão distribuídos aleatoriamente ao longo do mapa.

O teste é utilizado para estimar os possíveis resultados de um evento incerto, avaliando o impacto do risco em diferentes cenários. O teste de Monte Carlo prevê um conjunto de resultados possíveis baseado em um intervalo de valores que foram estimados por meio de amostras extraídas da população. Sendo assim, é criado um modelo de resultados possíveis para qualquer variável em que não se conhece a distribuição ou as suposições do modelo foram violadas. Ela, então, recalculará os resultados sucessivamente, cada vez usando um conjunto diferente de números aleatórios.

Desse modo, a partir do teste de Monte Carlo é possível avaliar se os clusters espaciais encontrados são, ou não, significativos e assim, detectar as áreas com risco mais elevado. Porém, para realizar o teste, a verossimilhança da variável de interesse deve ser conhecida, fazendo-se necessário a definição da distribuição que mais se adequa. Como o interesse é no número de ocorrências de criminalidade para cada CISP do estado do Rio de Janeiro, faz-se apropriado o uso da distribuição de Poisson, modelo usual para dados de contagem.

## 3 Resultados

Neste capítulo serão discutidos os resultados das análises realizadas para o estudo. Para tal, a organização dos resultados está dividida em duas seções, de modo que a primeira consiste na apresentação da saída bruta obtida com o software e a segunda parte traz uma discussão e os resultados da análise espacial de persistência de clusters dos crimes de furto, letalidade violenta e roubo no Estado do Rio de Janeiro ao longo dos anos de 2016 a 2020.

A técnica Scan descrita minuciosamente no capítulo anterior respalda o procedimento que será implementado e aplicado na base de dados com auxílio do software SatScan. O objetivo é detectar conglomerados espaciais de casos de crimes no período de 2016 a 2020, buscando verificar se os locais de alto risco se mantiveram os mesmos ao longo dos anos.

Para obter os resultados, foram necessário três bancos de dados:

- Base de casos: obtida através da parceria entre a Universidade Federal Fluminense, UFF e o ISP, onde foi disponibilizado um banco de dados com as ocorrências de crime, a CISP e data em que houve o caso.
- Base de população: tamanho populacional em cada CISP, por ano.
- Shapefile: divisão territorial do estado do Rio de Janeiro, por CISP.

Os arquivos foram inseridos no software e selecionou-se o tipo de consulta a ser realizada para este trabalho. Um exemplo de saída é apresentado na próxima seção.

### 3.1 Apresentação do output do software

Neste seção será apresentado o resultado fornecido pelo software SaTScan para os trimestres dos anos de 2016 a 2020, que serão analisados na seção subsequente.

O seguinte output refere-se ao primeiro trimestre do ano de 2016, de primeiro de janeiro a 31 de abril.

```

1.Location IDs included.: 11, 15, 19, 14, 13, 20, 12, 7, 10, 6, 18, 9, 25, 26, 16, 23, 4, 17,
                        5, 41, 24, 21, 44, 32, 29, 28, 22, 76, 30, 27, 77, 40, 38, 1, 37, 33,
                        78, 79, 39, 81, 31, 59, 42, 73, 64, 57
Overlap with clusters.: 2, 4, 5, 6, 8
Coordinates / radius.: (22.988890 S, 43.248629 W) / 26.21 km
Gini Cluster.....: No
Population.....: 6585803
Number of cases.....: 1908
Expected cases.....: 1410.60
Annual cases / 100000.: 116.3
Observed / expected...: 1.35
Relative risk.....: 1.76
Log likelihood ratio..: 141.867914
P-value.....: < 0.0000000000000001

```

Figura 4: Exemplo de saída do Satscan para identificação do cluster primário - Primeiro trimestre de 2016

Nesta primeira parte do output, representada pela Figura 4, são fornecidas informações importantes como: o cluster mais verossímil, que é a janela espacial buscada neste trabalho, assim como a estatística do teste  $LLR(z)$ , o risco relativo, uma razão entre os casos observados e esperados, que são medidas importantes no estudo da criminalidade. O p-valor, probabilidade de obtermos uma estatística de teste igual ou mais extrema que a observada quando assumimos a hipótese nula verdadeira, também é fornecido.

```

Analysis
-----
Type of Analysis      : Purely Spatial
Probability Model     : Discrete Poisson
Scan for Areas with  : High Rates

```

Figura 5: Exemplo de saída do Satscan sobre a análise realizada - Primeiro trimestre de 2016

Observando a segunda parte do output na Figura 5, é possível verificar o tipo de análise realizado nos dados, sendo ela puramente espacial para altos riscos relativos e considerando um modelo de Poisson.

```

Inference
-----
P-Value Reporting      : Default Combination
Number of Replications : 999
Adjusting for More Likely Clusters : No

```

Figura 6: Exemplo de saída do Satscan Teste de Monte Carlo - Primeiro trimestre de 2016

Por fim, na Figura 6, existem informações a respeito da simulação de Monte Carlo e sobre as áreas consideradas para o estudo.

## 3.2 Análise e Discussão

O algoritmo foi executado por meio do SaTScan versão 9.1.1. O processamento ocorreu em uma máquina com sistema operacional OS X 10.8.2 com dois núcleos de 2GHz e 4GB de memória RAM. Para cada trimestre o algoritmo precisou de cerca de 8 segundos de processamento para ser completado.

A análise será dividida por crimes, sendo considerado uma subseção para o crime de furto, outro para letalidade violenta e outra para roubo. As coordenadas dos clusters espaciais geradas pelo SatScan foram processadas no RStudio, tornando possível que fossem gerados mapas e possibilitando uma visualização mais consistente dos clusters. Todos os clusters apresentados neste trabalho foram tidos como significativos e a totalidade com um p-valor menor que  $0,001 \times 10^{-15}$

### 3.2.1 Furto

O crime de furto é descrito como subtração, ou seja, diminuição do patrimônio de outra pessoa, sem que haja violência. Com isso, sabe-se que o furto é uma crime presente no dia a dia dos moradores do estado do Rio de Janeiro, necessitando de constante atenção dos agentes públicos em relação a seu combate. A Tabela 1 nos fornece uma visão inicial mais clara em relação à magnitude do problema.

Ao avaliar a distribuição do número de casos de furto ao longo dos trimestres de 2016 a 2020 é possível notar que de 2016 a 2020 houve uma queda significativa no total de casos, de aproximadamente 20%. Essa queda no número absoluto de casos pode ser explicada pelo cenário pandêmico instaurado em 2020, onde houve diminuição de circulação de pessoas nas ruas. Quanto ao total de casos por trimestre, não existe um comportamento semelhante entre eles, havendo muita variação ao compará-los.

Tabela 1: Número de casos de furto no estado do RJ

Ano	1º Trimestre	2º Trimestre	3º Trimestre	4º Trimestre	Total
2016	3.548	3.484	3.076	2.231	12.339
2017	3.166	2.829	2.664	2.846	11.505
2018	2.688	2.923	2.958	3.025	11.594
2019	3.099	3.022	3.008	2.792	11.921
2020	2.808	2.127	2.448	2.474	9.857

A Estatística Scan identificou conglomerados significativos de alto risco em todos os trimestres de todos os anos englobados pelo estudo. Para uma visão mais completa do

cenário a Tabela 2 apresenta a razão entre os valores observados e esperados e o risco relativo para a ocorrência do crime de furto para cada cluster identificado nos anos do estudo.

Tabela 2: Razão entre os valores observados e esperados e o risco relativo dos clusters, avaliando o crime de furto.

Ano	Razão	Risco Relativo
1 <sup>o</sup> Trimestre 2016	1,35	1,76
2 <sup>o</sup> Trimestre 2016	1,29	1,76
3 <sup>o</sup> Trimestre 2016	1,31	1,77
4 <sup>o</sup> Trimestre 2016	1,33	1,84
1 <sup>o</sup> Trimestre 2017	1,36	1,75
2 <sup>o</sup> Trimestre 2017	1,39	1,90
3 <sup>o</sup> Trimestre 2017	1,38	2,19
4 <sup>o</sup> Trimestre 2017	1,58	2,14
1 <sup>o</sup> Trimestre 2018	1,66	2,01
2 <sup>o</sup> Trimestre 2018	1,45	2,34
3 <sup>o</sup> Trimestre 2018	1,35	2,06
4 <sup>o</sup> Trimestre 2018	1,84	2,18
1 <sup>o</sup> Trimestre 2019	1,49	1,90
2 <sup>o</sup> Trimestre 2019	1,50	2,03
3 <sup>o</sup> Trimestre 2019	1,44	2,27
4 <sup>o</sup> Trimestre 2019	1,40	1,87
1 <sup>o</sup> Trimestre 2020	1,29	1,71
2 <sup>o</sup> Trimestre 2020	1,39	2,22
3 <sup>o</sup> Trimestre 2020	1,43	2,37
4 <sup>o</sup> Trimestre 2020	1,44	2,28

Claramente nota-se que a incidência, nos clusters encontrados ao longo dos trimestres é aproximadamente 1,5 vezes maior que a esperada, ou seja, a população encontra-se uma vez e meia mais exposta ao furto do que deveria estar. Observando cada trimestre separadamente vê-se que há um número de casos observados maior do que o número de casos esperados e que o 4<sup>o</sup> trimestre de 2018 possui o maior valor de razão, com praticamente duas vezes mais casos do que deveriam existir.

Em seguida, é apresentada uma medida interessante: o risco relativo. O risco relativo é claramente maior que 1 para todos os períodos de análise, isto é, o risco da ocorrência de furto é maior em indivíduos expostos do que em indivíduos não expostos (indicando uma relação da exposição do indivíduo do crime em pauta). O maior risco ocorreu no 3<sup>o</sup> Trimestre 2020 com um valor de 2,37, ou seja, existia para a população exposta um risco aproximado de duas vezes e meia maior de ser furtado neste período, porém nos outros trimestres o comportamento é semelhante, com um risco relativo em torno de 2.

Seguindo a análise, em todos os testes realizados e nos clusters encontrados, os p-valoros foram significativos a um nível de significância de 5%, comprovando com grande segurança a hipótese acerca dos dados em estudo, validando que os conglomerados encontrados não ocorrem por mero acaso.

Com isso, sabe-se que os mapas de riscos relativos podem ser bastante afetados pela variabilidade, o que torna uma mera inspeção visual pouco confiável. Nesse cenário, implementou-se a saída do SaTScan a fim de encontrar os conglomerados estatisticamente significativos, caso eles existam e representá-los visualmente, através de mapas.

Dessa forma, optou-se por definir a busca por conglomerados, abrangendo círculos de 50% da população sob risco, pois dessa forma, áreas com risco muito baixo não são identificadas como *clusters*. Para isso, foram geradas imagens dos *clusters* encontrados para os diferentes trimestres e verificados quais CISPs apresentavam risco elevado. A Figura 7 representa o cluster obtido para o período de 2016.

Quanto ao cluster mais significativo para os trimestres de 2016, é observado um comportamento bem diferente entre eles. O 1º trimestre tem alto risco na região Metropolitana, mas esse risco se estende para algumas regiões do Norte Fluminense, principalmente nas regiões litorâneas. Já nos outros trimestres de 2016, existem diferentes regiões afetadas, mas elas se concentram na Região Metropolitana do Rio, diferenciando apenas no 4º Trimestre em que há uma região do Sul Fluminense com alto risco.

Para analisar os anos seguintes, a fim verificar se o comportamento dos trimestres são semelhantes no decorrer do período de análise, são apresentados os clusters para os anos de 2017- 2020 representados pelas Figuras 8 - 11.

O 1º Trimestre de 2017 difere de 2016, pois mais CISPs localizadas na região Metropolitana são afetadas, porém assim como em 2016, existe uma região no Norte Fluminense com alto risco de ocorrências. O 2º Trimestre de 2017 não se dá de maneira semelhante com 2016, sendo apenas poucas localidades afetadas e todas elas na região Metropolitana. Em 2017, o 3º Trimestre é bem semelhante ao de 2016. Por fim, o 4º Trimestre apresenta comportamento distinto ao observado em 2016, incluindo uma localidade no Noroeste Fluminense, o que não ocorreu no ano de 2016.

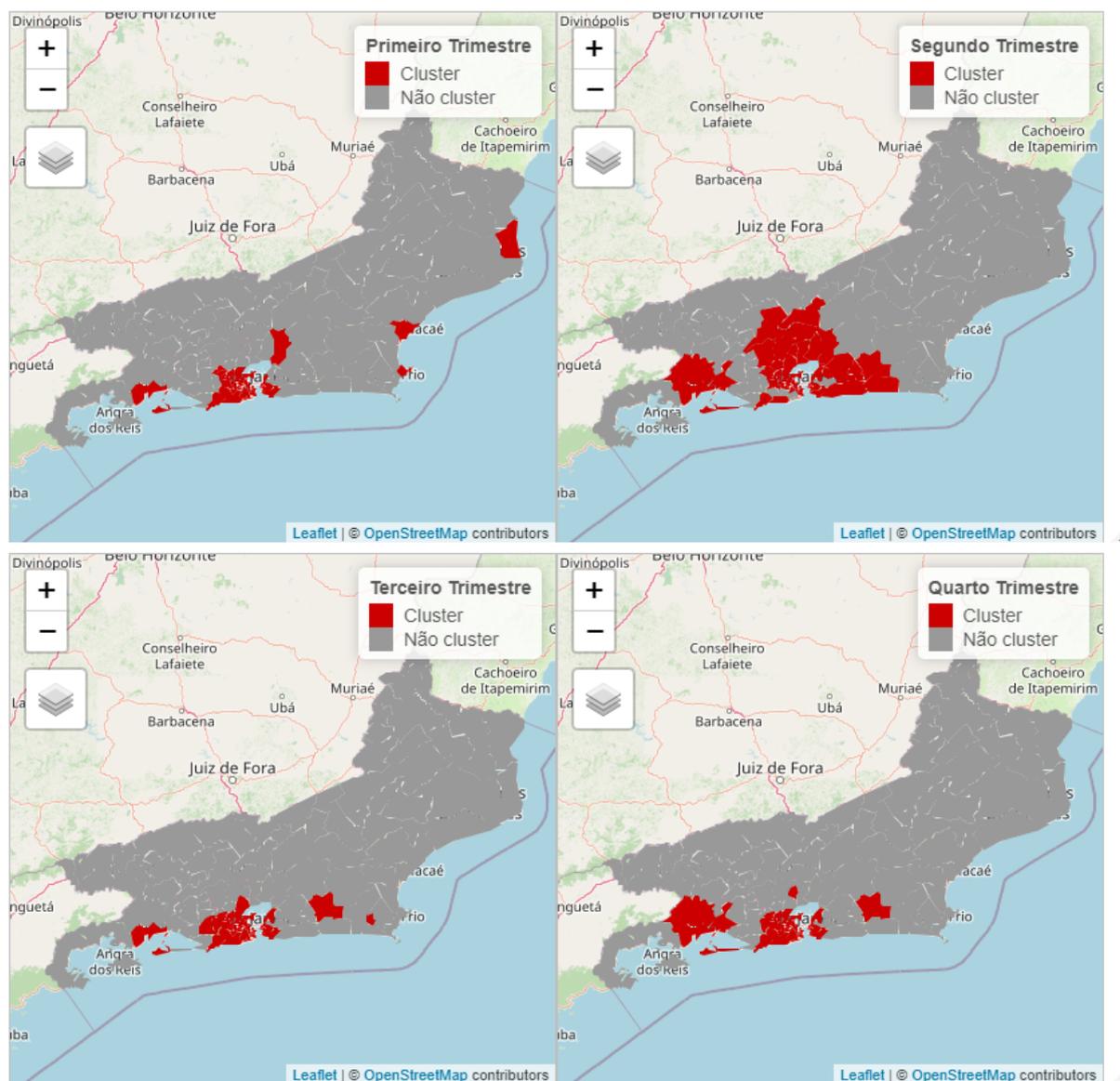


Figura 7: Cluster de alto risco para casos de furto no estado do Rio de Janeiro em 2016.

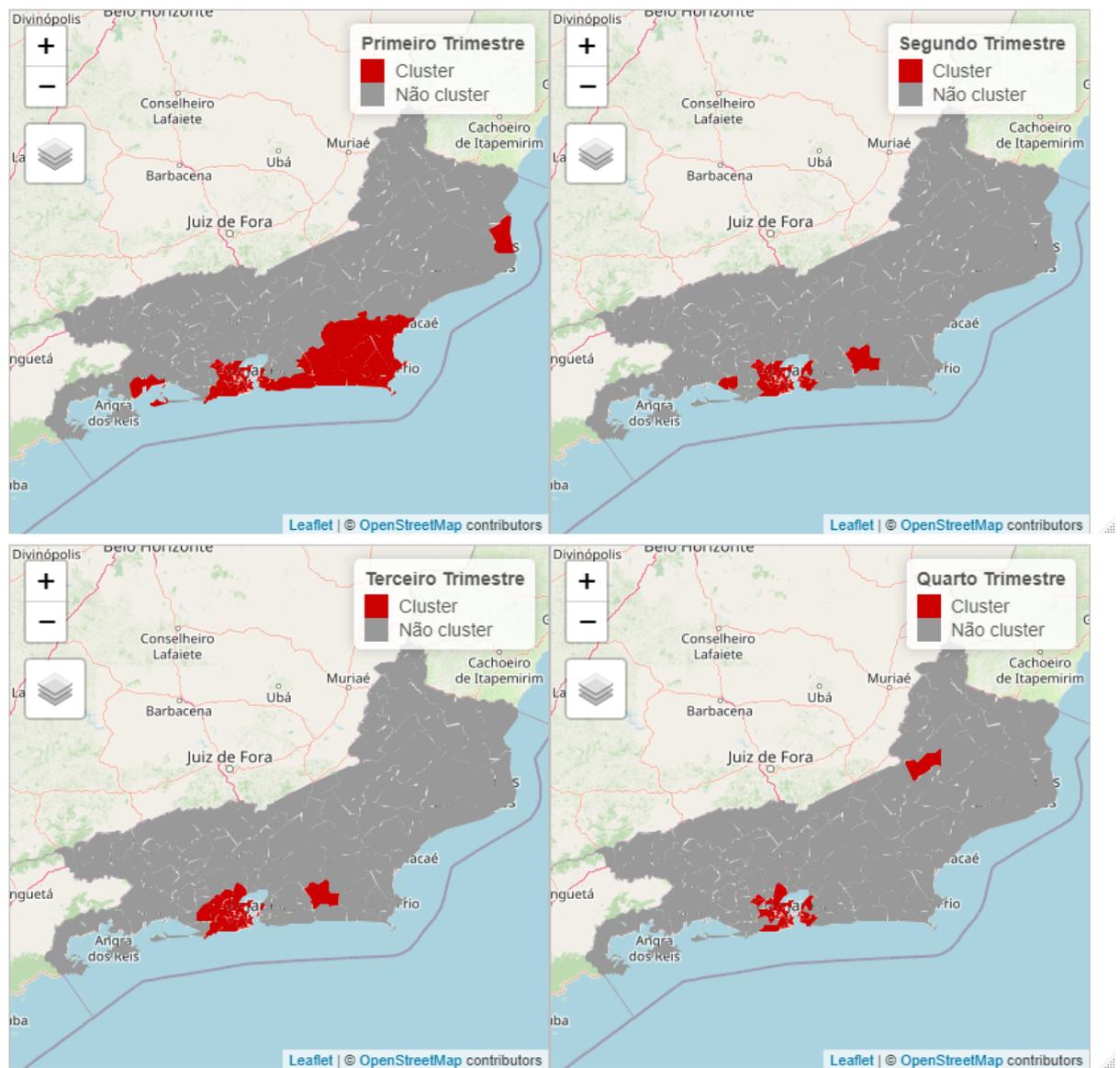


Figura 8: Cluster de alto risco para casos de furto no estado do Rio de Janeiro em 2017.

Em 2018, observa-se uma disposição de regiões com alto risco um pouco diferente do que nos anos anteriores. Mesmo que a Região Metropolitana seja a mais afetada, é possível constatar que houve uma disposição em regiões localizadas mais próximas ao Sul Fluminense.

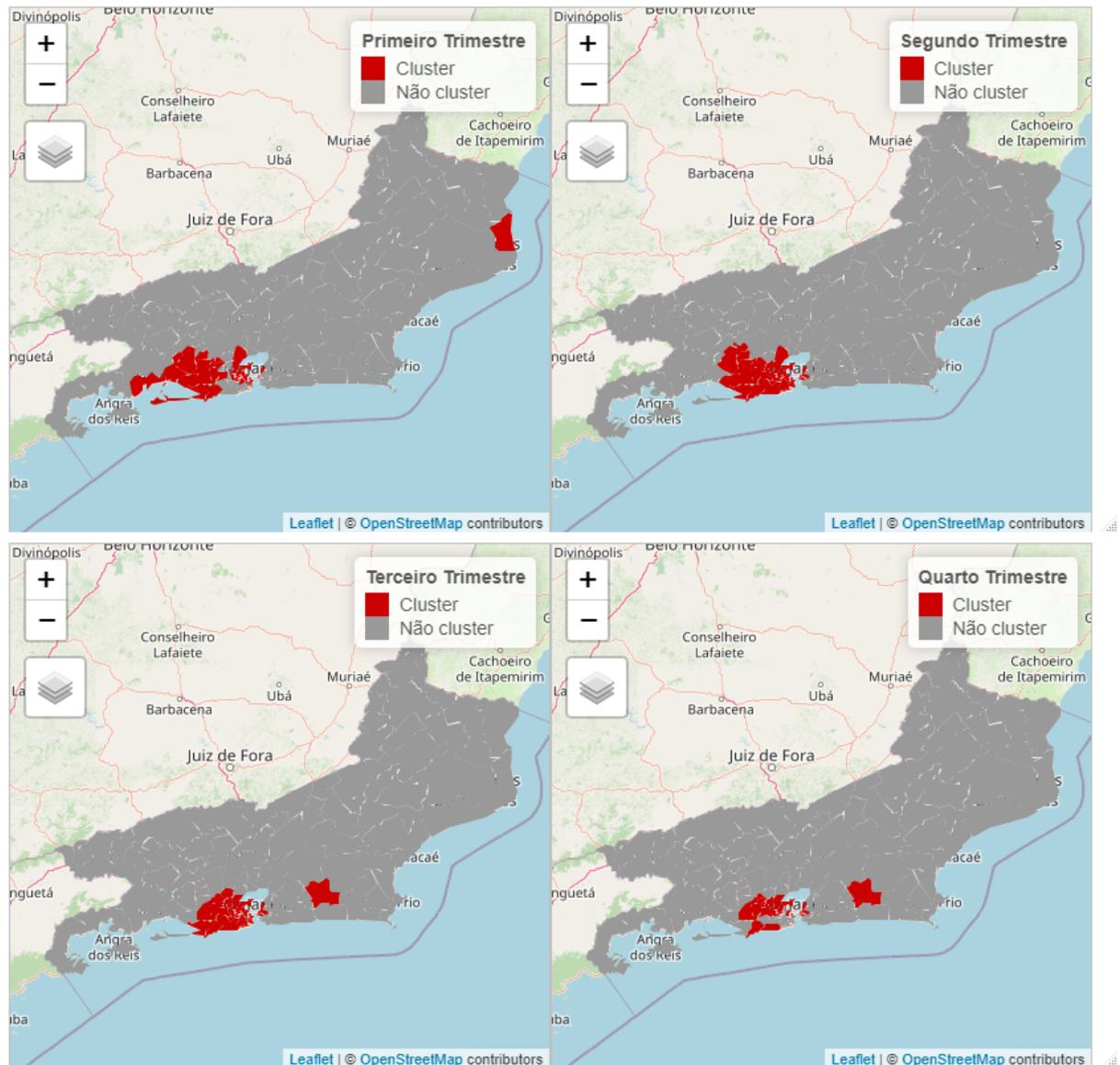


Figura 9: Cluster de alto risco para casos de furto no estado do Rio de Janeiro em 2018.

Seguindo a análise para o ano de 2019, nota-se que no 1º trimestre há uma disposição diferente de todos os outros anos analisados, com uma distribuição mais espalhada ao longo do estado. O 2º trimestre de 2019 foi bem semelhante a 2017, atingindo apenas a Região Metropolitana, assim como o 3º trimestre e o 4º trimestre. Além disso, o 4º trimestre apresentou muita semelhança com o ano de 2018.

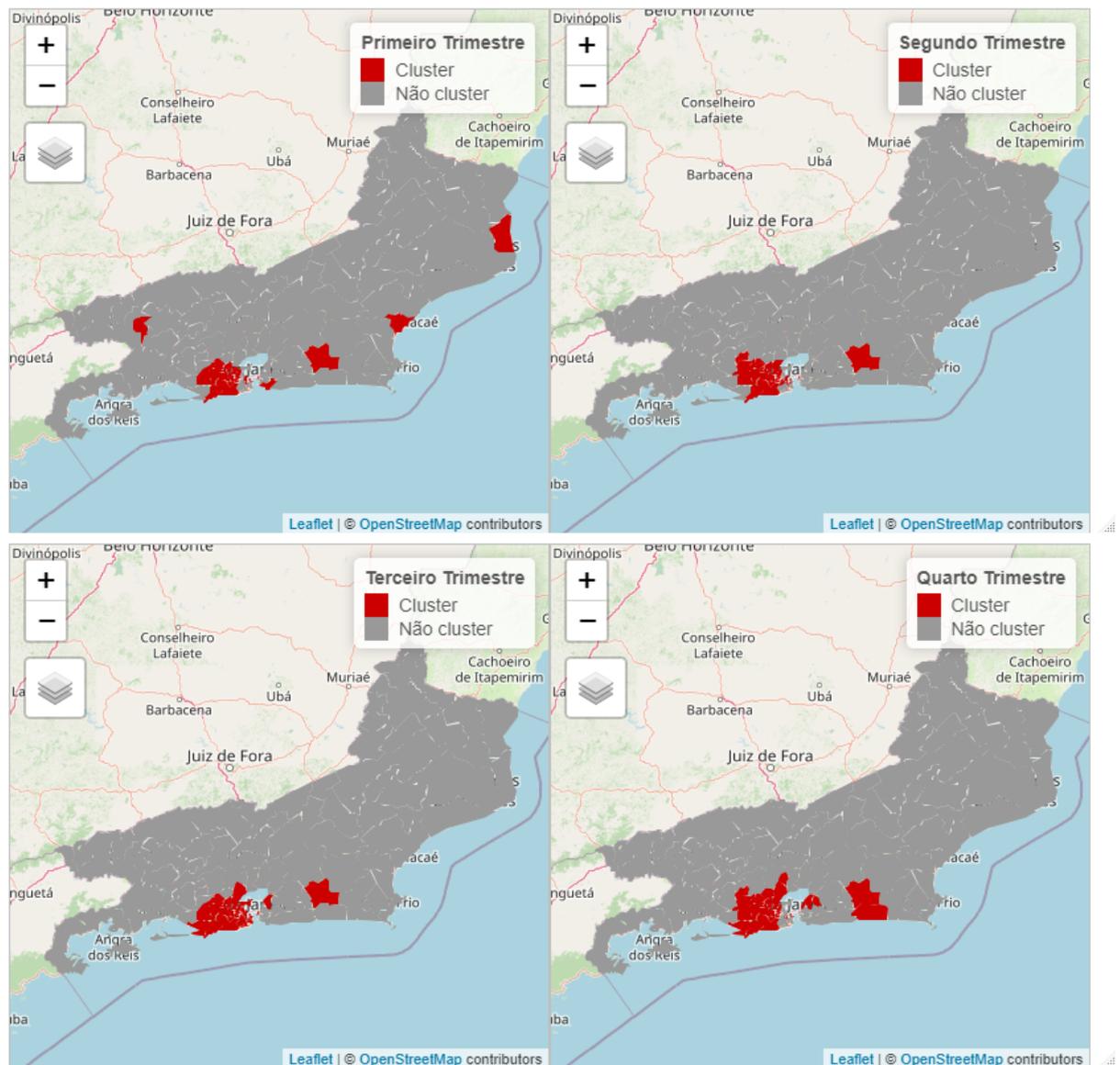


Figura 10: Cluster de alto risco para casos de furto no estado do Rio de Janeiro em 2019.

Por fim, no ano de 2020 os *clusters* de alto risco estavam dispostos de maneira bem semelhante aos anos de 2017 e 2019. Apenas no 1º trimestre há uma disposição mais espalhada ao longo do estado, comportamento que não é visto nos trimestres posteriores.

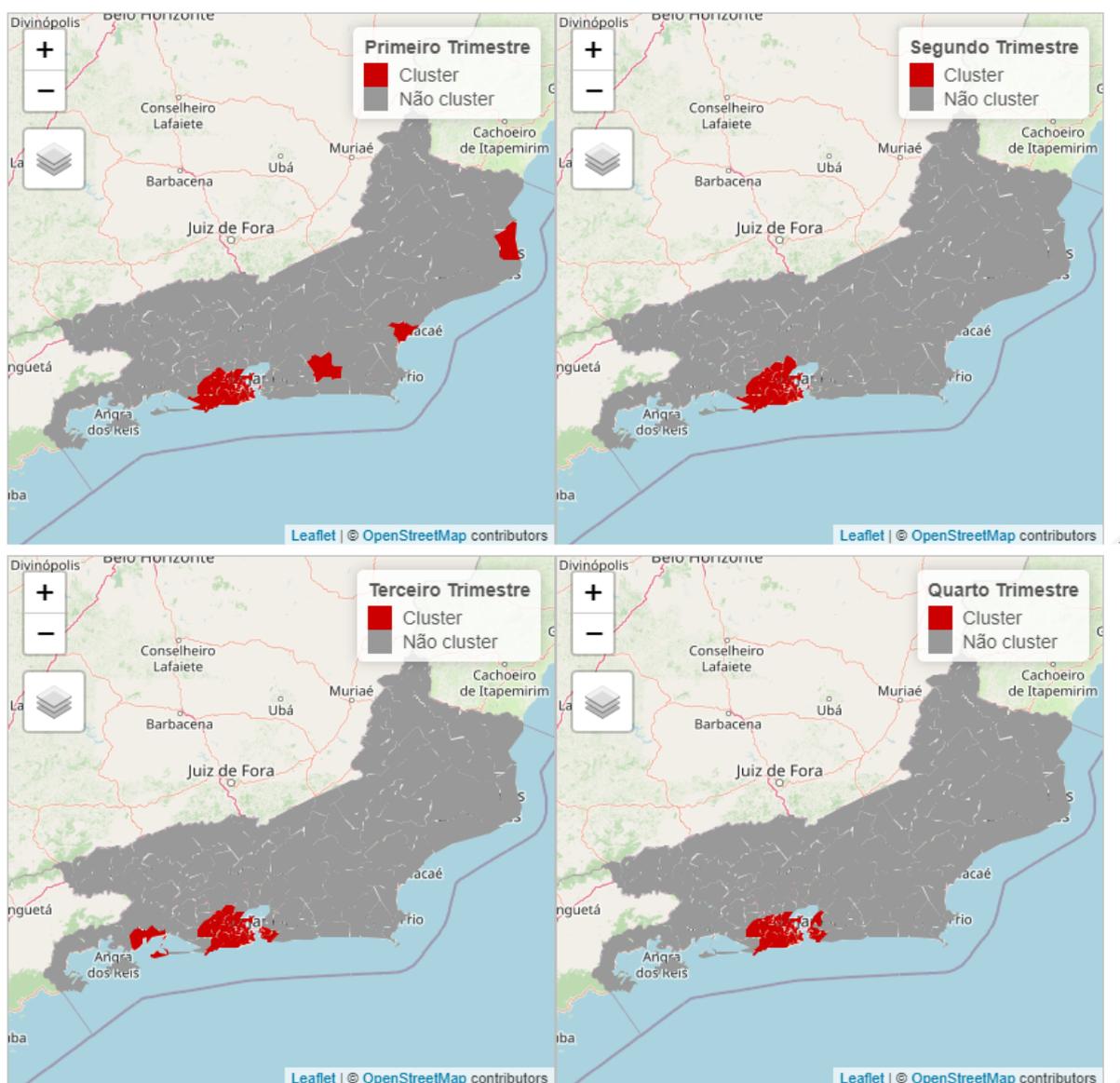


Figura 11: Cluster de alto risco para casos de furto no estado do Rio de Janeiro em 2020.



### 3.2.2 Letalidade Violenta

O índice de letalidade violenta engloba homicídio doloso, lesão seguida de morte, latrocínio e morte pela polícia. O Estado do Rio de Janeiro vem apresentando queda no número de ocorrências de letalidade violenta, porém continua sendo um crime que merece atenção especial dos agentes públicos. A Tabela 3 nos dá uma visão inicial mais clara em relação à relevância do problema para os trimestres dos anos de 2016 a 2020.

Ao avaliar a distribuição do número de casos de letalidade violenta é possível notar que de 2016 a 2020 houve redução no número de ocorrências, sendo de aproximadamente 28%. Essa queda no número absoluto de casos pode ser explicada pelo cenário pandêmico instaurado em 2020, onde houve redução de circulação de pessoas nas ruas. Em relação ao total de casos por trimestre, não existe um comportamento semelhante entre eles, pois houve muita variação ao compará-los.

Tabela 3: Número de casos de letalidade violenta no Estado do RJ.

Ano	1º Trimestre	2º Trimestre	3º Trimestre	4º Trimestre	Total
2016	1.039	1.068	999	796	3.902
2017	1.075	902	877	1.005	3.859
2018	1.087	925	905	859	3.776
2019	828	922	885	810	3.445
2020	878	688	532	723	2.821

A Estatística Scan identificou conglomerados significativos de alto risco em todos os trimestres de todos os anos englobados pelo estudo. Para uma visão mais completa do cenário a Tabela 4 apresenta a razão entre os valores observados e esperados e o risco relativo para a ocorrência do crime de letalidade violenta para cada cluster identificado nos anos do estudo.

Claramente nota-se que a incidência, nos clusters encontrados até o 2º trimestre de 2018, é aproximadamente 2 vezes maior que a esperada, ou seja, a população encontra-se duas vezes mais exposta ao crime de letalidade violenta do que deveria estar. Ao observar cada trimestre separadamente vê-se que há um número de casos observados maior do que o número de casos esperados e no 2º Trimestre de 2018 tem-se o maior valor de razão, com praticamente duas vezes e meia mais casos do que deveriam existir.

Em seguida, pode-se observar o risco relativo. O risco relativo é claramente maior que 1 para todos os conglomerados encontrados, isto é, o risco da ocorrência de letalidade violenta é maior em indivíduos expostos do que em indivíduos não expostos (indicando uma relação da exposição do indivíduo do crime em pauta). O maior risco ocorreu no 2º

Tabela 4: Razão entre os valores observados e esperados e o risco relativo dos cluster, avaliando o crime de letalidade violenta.

Ano	Razão	Risco Relativo
1 <sup>o</sup> Trimestre 2016	1,91	2,32
2 <sup>o</sup> Trimestre 2016	1,93	2,62
3 <sup>o</sup> Trimestre 2016	1,88	2,74
4 <sup>o</sup> Trimestre 2016	1,94	2,64
1 <sup>o</sup> Trimestre 2017	1,80	2,32
2 <sup>o</sup> Trimestre 2017	2,08	2,99
3 <sup>o</sup> Trimestre 2017	1,95	2,65
4 <sup>o</sup> Trimestre 2017	1,86	2,45
1 <sup>o</sup> Trimestre 2018	1,76	2,23
2 <sup>o</sup> Trimestre 2018	2,19	2,41
3 <sup>o</sup> Trimestre 2018	1,71	2,28
4 <sup>o</sup> Trimestre 2018	1,56	2,01
1 <sup>o</sup> Trimestre 2019	1,59	1,96
2 <sup>o</sup> Trimestre 2019	1,62	1,95
3 <sup>o</sup> Trimestre 2019	1,30	1,82
4 <sup>o</sup> Trimestre 2019	1,56	1,92
1 <sup>o</sup> Trimestre 2020	1,57	1,93
2 <sup>o</sup> Trimestre 2020	1,72	2,15
3 <sup>o</sup> Trimestre 2020	1,38	1,70
4 <sup>o</sup> Trimestre 2020	1,47	2,05

Trimestre 2017 com um valor de 2,99, isto é, existia para a população exposta um risco aproximado três vezes maior de ser vítima de letalidade. Contudo, nos outros trimestres o comportamento é semelhante, com um risco relativo variando entre 2 e 3.

Seguindo a análise feita na subseção anterior, em todos os testes realizados e nos clusters encontrados, os p-valores foram significativos ao nível de significância de 5%, o que comprova com grande segurança a hipótese acerca dos dados em estudo, validando que os conglomerados encontrados não ocorrem por mero acaso.

Assim, como os mapas de risco relativos podem ser bastante afetados pela variabilidade, implementou-se a saída do SaTScan nesse estudo a fim de encontrar os conglomerados estatisticamente significativos, caso eles existam, de maneira visual, por meio de um mapa.

Dessa forma, optou-se por definir a busca por conglomerados, abrangendo círculos de 50% da população sob risco. Para isso, foram geradas imagens dos clusters encontrados para os diferentes anos e verificado quais CISPs apresentavam risco elevado. As Figuras 13 - 17 representam os clusters obtidos para os períodos de análise, de 2016 a 2020.

Observando os conglomerados obtidos, é notório que não existe um padrão nos re-

sultados. A distribuição espacial dos conglomerados em 2016 permite a verificação de uma maior concentração dos crimes na região Metropolitana do estado e na região Sul Fluminense, com uma CISP localizada no Norte Fluminense, que se repete no 2º e 3º Trimestre. Nos trimestres seguintes houve um comportamento semelhante ao ocorrido no 1º Trimestre, com as mesmas mesoregiões sendo mais afetadas.

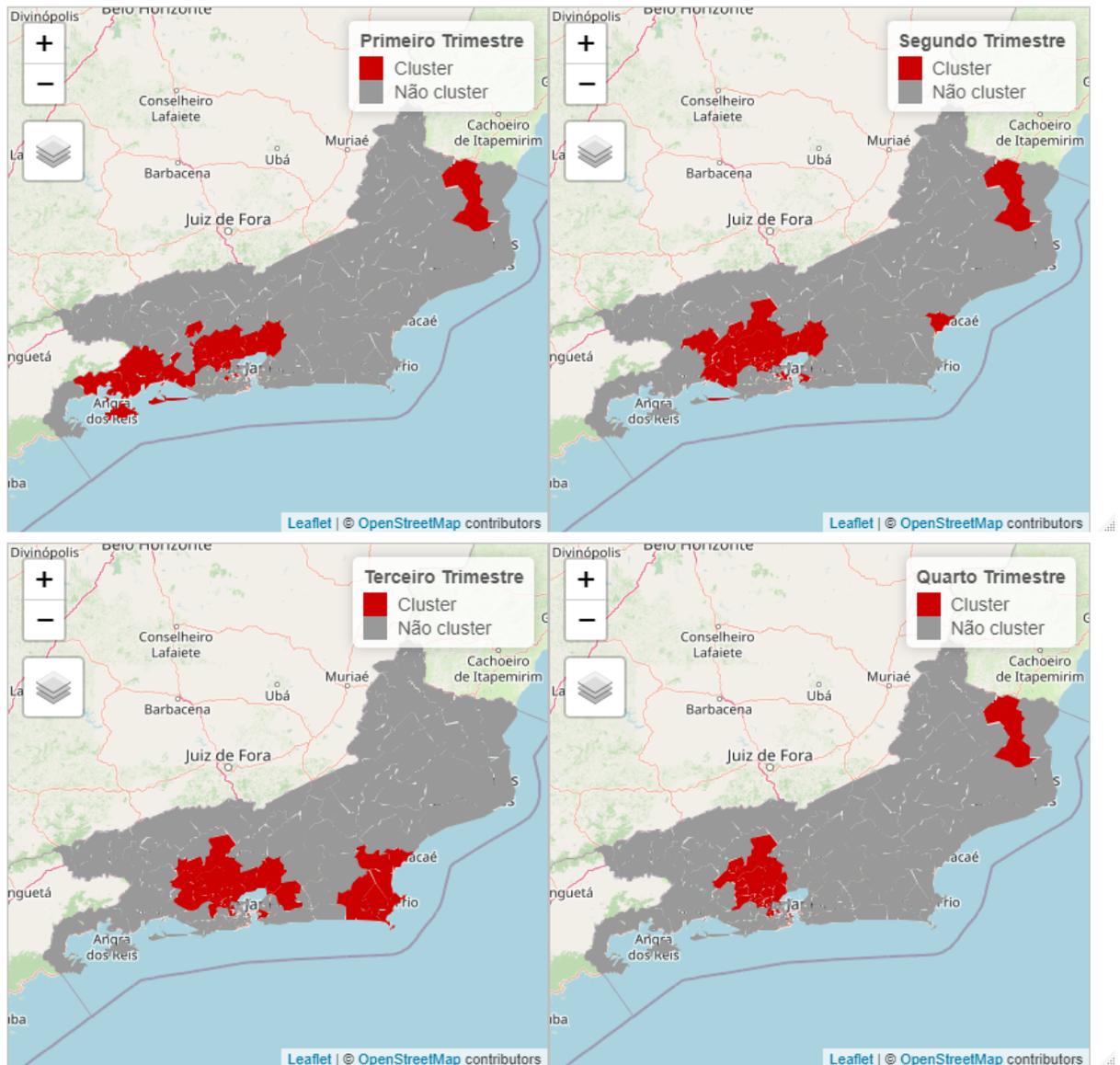


Figura 13: Cluster de alto risco para casos de letalidade violenta no estado do Rio de Janeiro em 2016.

Os conglomerados do ano de 2017 são espacialmente distribuídos de forma desigual e encontram seus níveis mais elevados nos municípios das regiões Metropolitana, sendo que no 1º trimestre se estende para as regiões do Sul Fluminense e Norte Fluminense, assim como em 2016.

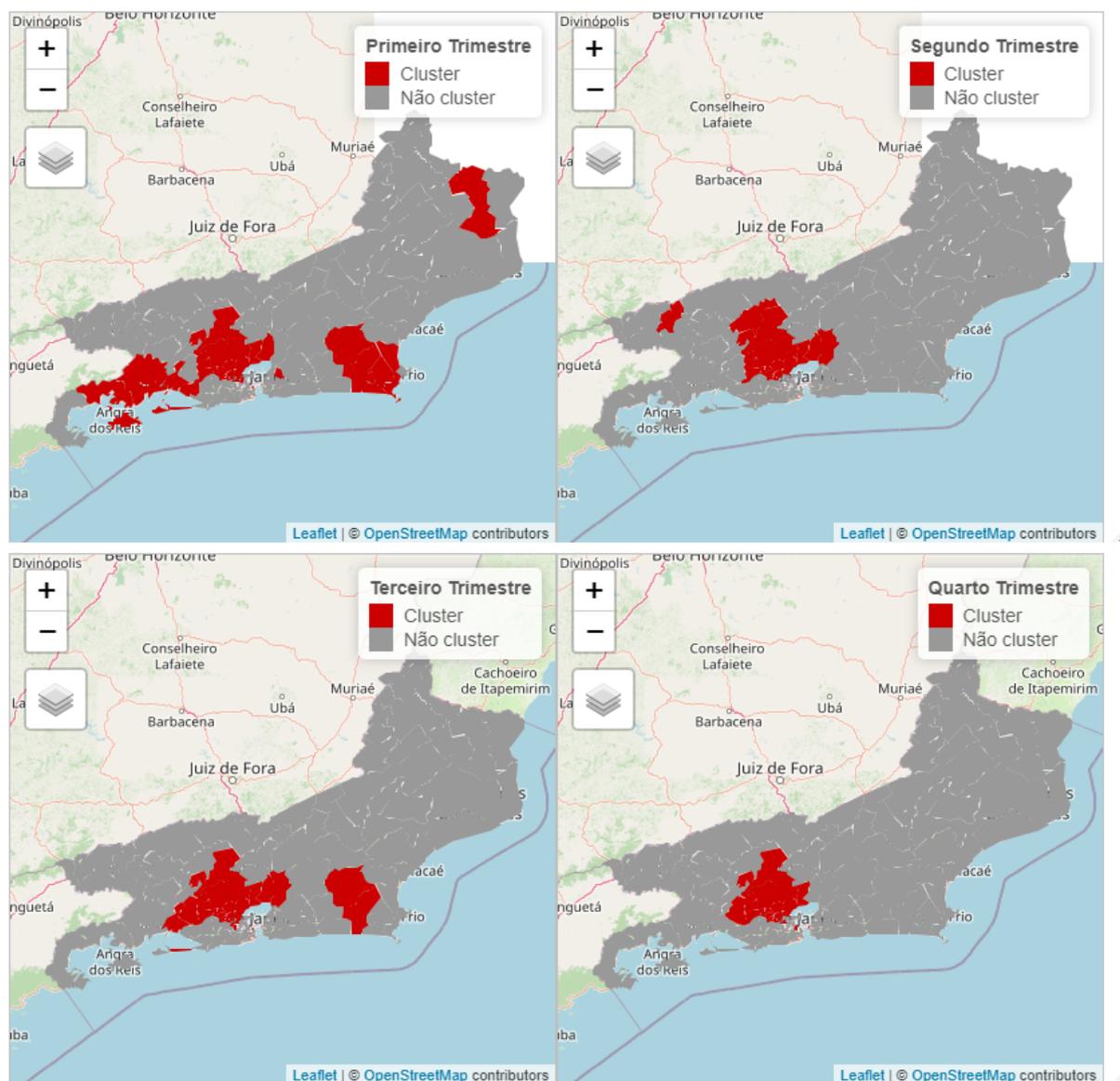


Figura 14: Cluster de alto risco para casos de letalidade violenta no estado do Rio de Janeiro em 2017.

Em 2018, os conglomerados identificados como de alto risco foram localizados nas regiões Metropolitana e Norte Fluminense. Como foi observado nos anos anteriores, há a persistência da CISP 146 na região Norte Fluminense que se repete ao decorrer dos períodos de análise.

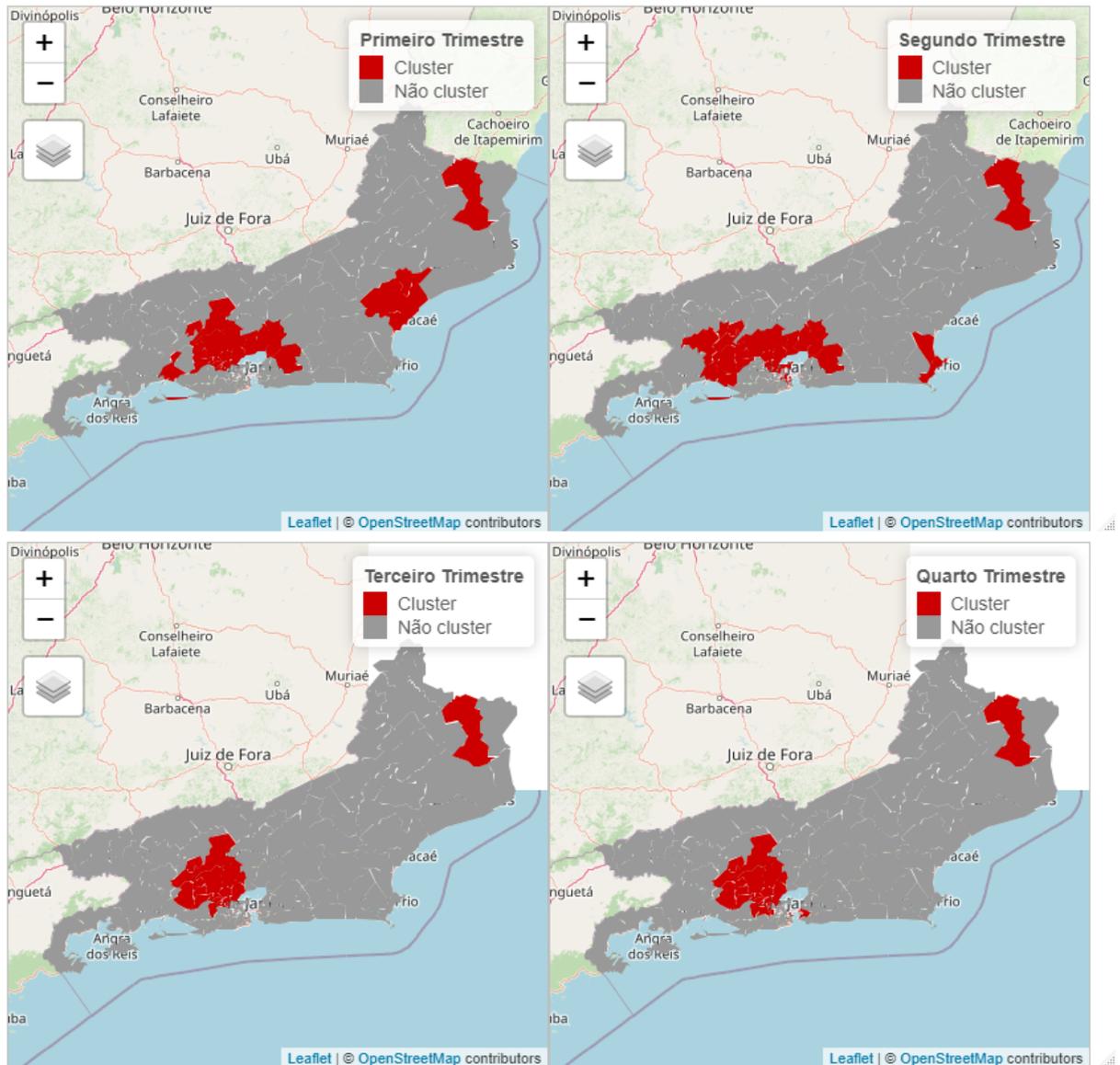


Figura 15: Cluster de alto risco para casos de letalidade violenta no estado do Rio de Janeiro em 2018.

Assim como nos anos anteriores, os conglomerados obtidos em 2019 e 2020 permitem a constatação de uma maior concentração dos crimes na região Metropolitana do estado e na região Norte Fluminense. O comportamento, ao analisar por trimestre nos dois anos, é diferente entre si, tanto dentro do mês de referência, como comparando os dois anos. Assim como nos anos anteriores, os *clusters* estão mais dispersos ao longo do território, diferente do que foi observado para o crime de furto.

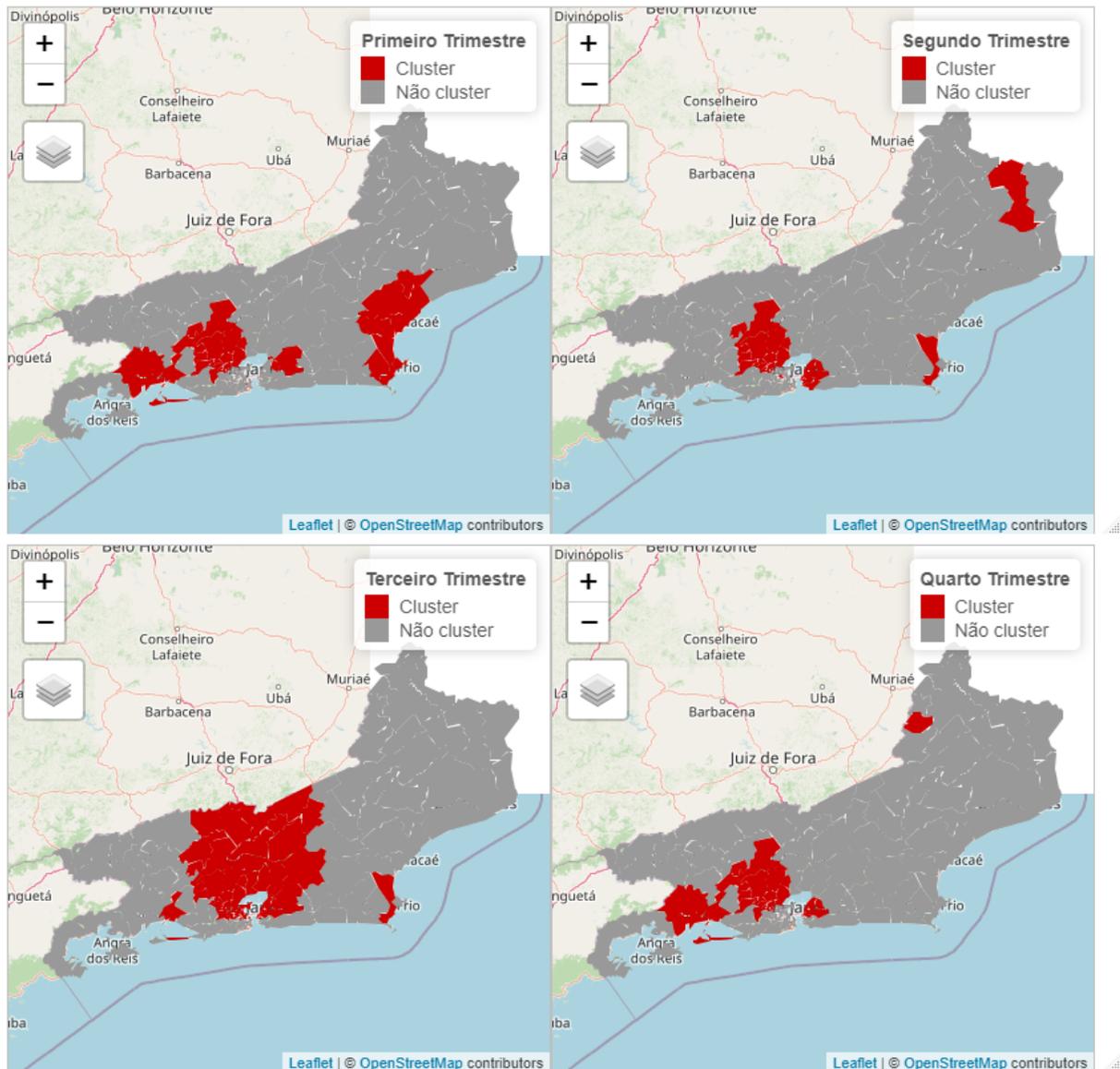


Figura 16: Cluster de alto risco para casos de letalidade violenta no estado do Rio de Janeiro em 2019.

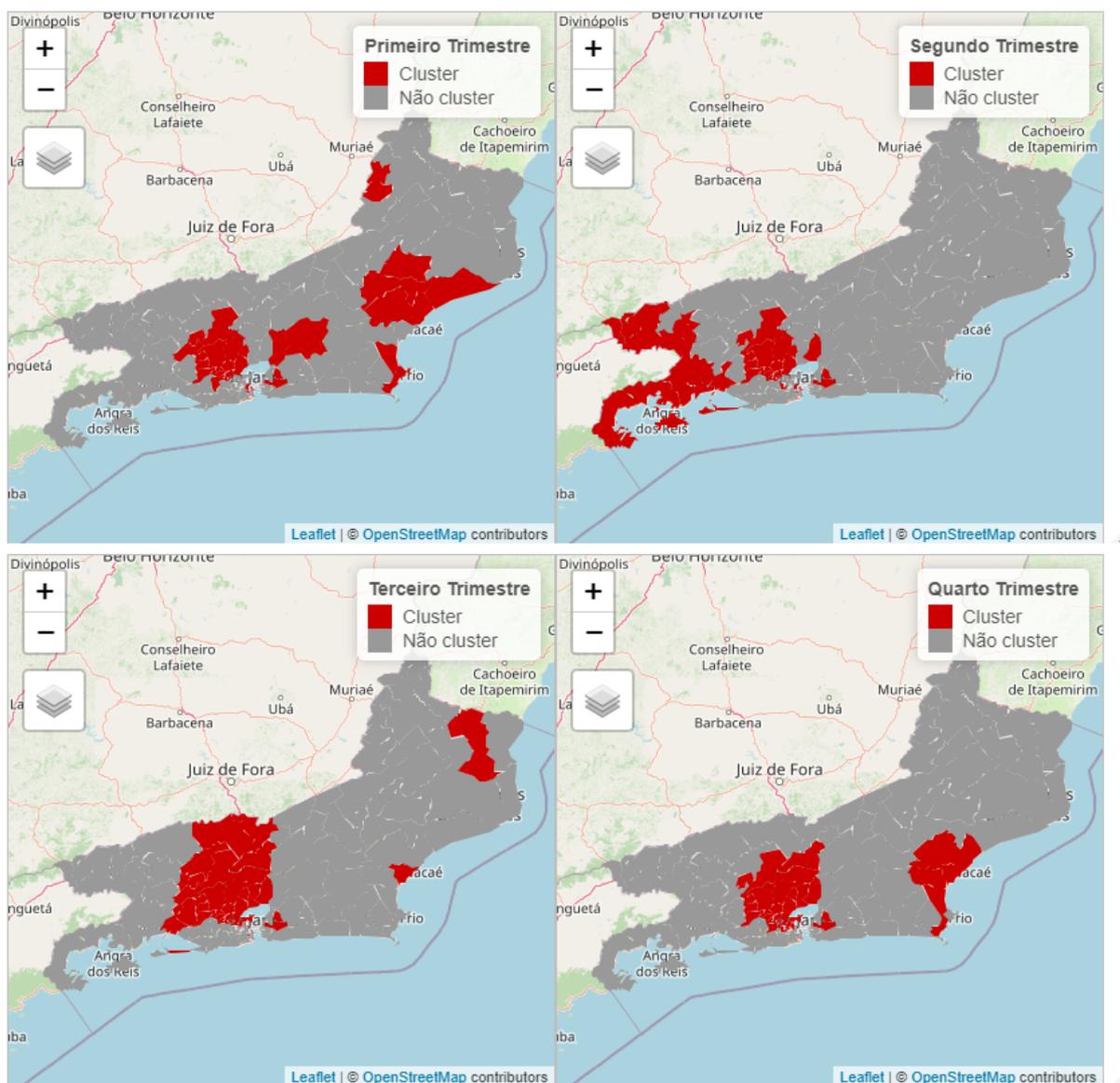


Figura 17: Cluster de alto risco para casos de letalidade violenta no estado do Rio de Janeiro em 2020.

Dessa forma, em todos os anos analisados há muita dispersão nas CISPs que foram indicadas como de alto risco, havendo variação, até mesmo, ao se analisar no decorrer dos trimestres. Entretanto, quando a análise é feita pela divisão de mesoregiões, os anos apresentam semelhanças nas áreas mais atingidas, sendo a região Metropolitanas o local mais atingidos pelo crime de letalidade violenta. Porém, ao analisar a Figura 18, observe-se que os *clusters* estão concentrados ao centro da Região Metropolitana e um persistente no Norte Fluminense, indicando as áreas que são mais atingidas.

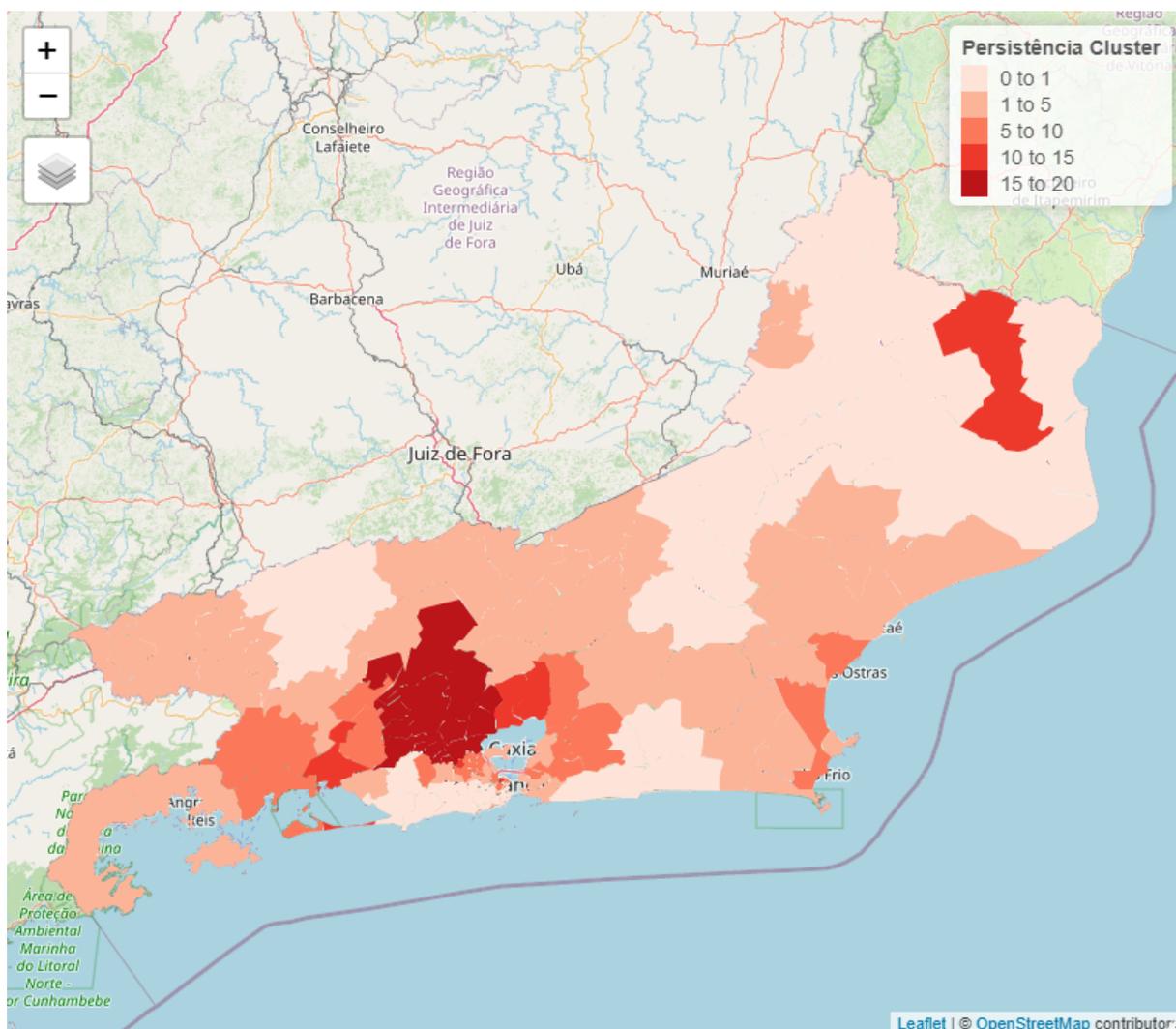


Figura 18: Persistência dos clusters de alto risco para casos de letalidade violenta no estado do Rio de Janeiro.

### 3.2.3 Roubo

O crime de roubo é descrito como diminuição do patrimônio de outra pessoa, mas diferente do furto, existe violência. Com isso, ao acompanhar os diversos noticiários, vemos diariamente notícias sobre casos de roubo no estado do Rio de Janeiro necessitando, ainda mais, atenção dos agentes públicos em relação a seu combate e conseqüentemente, diminuição no número de ocorrências. A Tabela 5 nos dá uma visão inicial mais clara em relação à magnitude do problema para os trimestres dos anos de 2016 a 2020.

Ao avaliar a frequência do número de eventos ocorridos é possível notar que de 2016 a 2020 houve uma grande queda no total de casos, de aproximadamente 49%. Essa queda no número absoluto de casos pode ser explicada pelo cenário pandêmico instaurado em 2020, onde houve diminuição de circulação de pessoas nas ruas. Quanto ao total de casos por trimestre, não houve um comportamento semelhante entre eles, havendo muita variação no decorrer dos anos.

Tabela 5: Número de casos de roubo no Estado do RJ.

Ano	1º Trimestre	2º Trimestre	3º Trimestre	4º Trimestre	Total
2016	35.173	35.287	37.295	30.488	138.243
2017	30.756	38.249	37.346	34.003	140.354
2018	38.142	36.150	33.998	34.488	142.778
2019	34.820	34.707	30.555	27.422	127.504
2020	26.109	11.838	15.751	17.539	71.237

A Estatística Scan identificou conglomerados significativos de alto risco em todos os trimestres de todos os anos englobados pelo estudo para o evento em questão. Para uma visão mais completa do cenário a Tabela 2 apresenta a razão entre os valores observados e esperados e o risco relativo para a ocorrência do crime de roubo para cada cluster identificado nos anos do estudo.

Através da Tabela 6, nota-se que a incidência, nos clusters encontrados como de alto risco, é aproximadamente 2 vezes maior do que a esperada, ou seja, a população encontra-se duas vezes mais exposta ao crime de roubo do que deveria estar. Ao observar cada trimestre separadamente vê-se que há um número de casos observados maior do que o número de casos esperados e no 2º Trimestre 2020 temos o maior valor de razão, com aproximadamente duas vezes mais casos do que deveria ocorrer.

Em seguida, têm-se os valores do risco relativo. O risco relativo é claramente maior que 1 para todos os conglomerados encontrados, isto é, o risco da ocorrência de roubo é maior em indivíduos expostos do que em indivíduos não expostos (indicando uma relação

Tabela 6: Razão entre os valores observados e esperados e o risco relativo dos clusters, avaliando os crimes de roubo.

Ano	Razão	Risco Relativo
1 <sup>o</sup> Trimestre 2016	1,55	3,34
2 <sup>o</sup> Trimestre 2016	1,70	3,08
3 <sup>o</sup> Trimestre 2016	1,53	3,24
4 <sup>o</sup> Trimestre 2016	1,68	3,34
1 <sup>o</sup> Trimestre 2017	1,68	3,89
2 <sup>o</sup> Trimestre 2017	1,60	3,72
3 <sup>o</sup> Trimestre 2017	1,69	3,81
4 <sup>o</sup> Trimestre 2017	1,60	3,81
1 <sup>o</sup> Trimestre 2018	1,60	3,91
2 <sup>o</sup> Trimestre 2018	1,58	3,54
3 <sup>o</sup> Trimestre 2018	1,58	3,49
4 <sup>o</sup> Trimestre 2018	1,60	3,72
1 <sup>o</sup> Trimestre 2019	1,58	3,63
2 <sup>o</sup> Trimestre 2019	1,58	3,74
3 <sup>o</sup> Trimestre 2019	1,56	3,48
4 <sup>o</sup> Trimestre 2019	1,68	3,75
1 <sup>o</sup> Trimestre 2020	1,64	4,22
2 <sup>o</sup> Trimestre 2020	1,73	4,20
3 <sup>o</sup> Trimestre 2020	1,70	3,98
4 <sup>o</sup> Trimestre 2020	1,64	4,18

da exposição do indivíduo do crime em pauta). O maior risco ocorreu no 1<sup>o</sup> Trimestre 2020 com um valor de 4,22, ou seja, existia, em 2020, para a população exposta, um risco aproximado, de ser vítima de roubo neste período e nos locais indicados como de alto risco, quatro vezes maior. Contudo, nos outros trimestres o comportamento é semelhante, com um risco relativo variando entre 3 e 4.

Seguindo a análise feita nas subseções anteriores, em todos os testes realizados e nos clusters encontrados, os p-valores foram significativos ao nível de significância de 5%, o que comprova novamente, com grande segurança, a hipótese acerca dos dados em estudo, validando que os conglomerados encontrados para o crime em questão não ocorrem por mero acaso.

Com isso, sabe-se que os mapas de risco relativos podem ser bastante afetados pela variabilidade, o que torna uma mera inspeção visual pouco confiável. Nesse cenário, implementou-se a saída do SaTScan a fim de encontrar os conglomerados estatisticamente significativos, caso eles existam e apresentá-los de maneira visual, através de mapas.

Dessa forma, optou-se por definir a busca por conglomerados, abrangendo círculos de 50% da população sob risco. Para isso, foram gerados mapas dos clusters encontrados para

os diferentes anos e verificado quais CISPs apresentavam risco elevado, onde as localidades de vermelho são os clusters e os de cinza são regiões que o número de casos observados é menor que o número de casos esperados. As Figuras 19 - 23 representam os clusters obtidos para o período de análise.

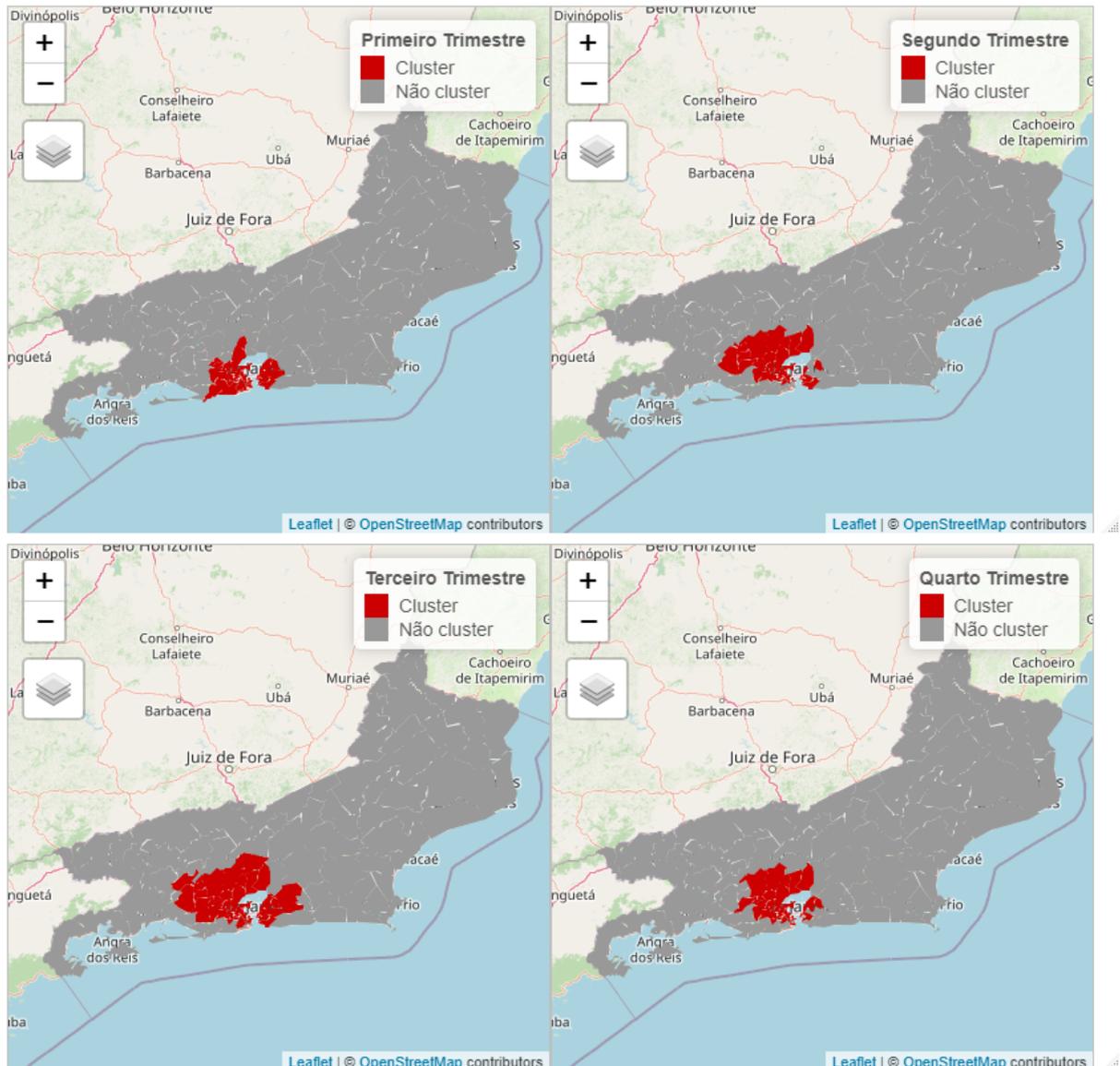


Figura 19: Cluster de alto risco para casos de roubo no estado do Rio de Janeiro em 2016.

Diferente do ocorrido pelos crimes de furto e letalidade violenta, vistos anteriormente, os casos de roubo se distribuem de maneira semelhante durante os anos de estudo. Através da análise visual dos mapas, podemos perceber que as ocorrências se concentram na região Metropolitana, e não há grande diferença no decorrer dos trimestres.

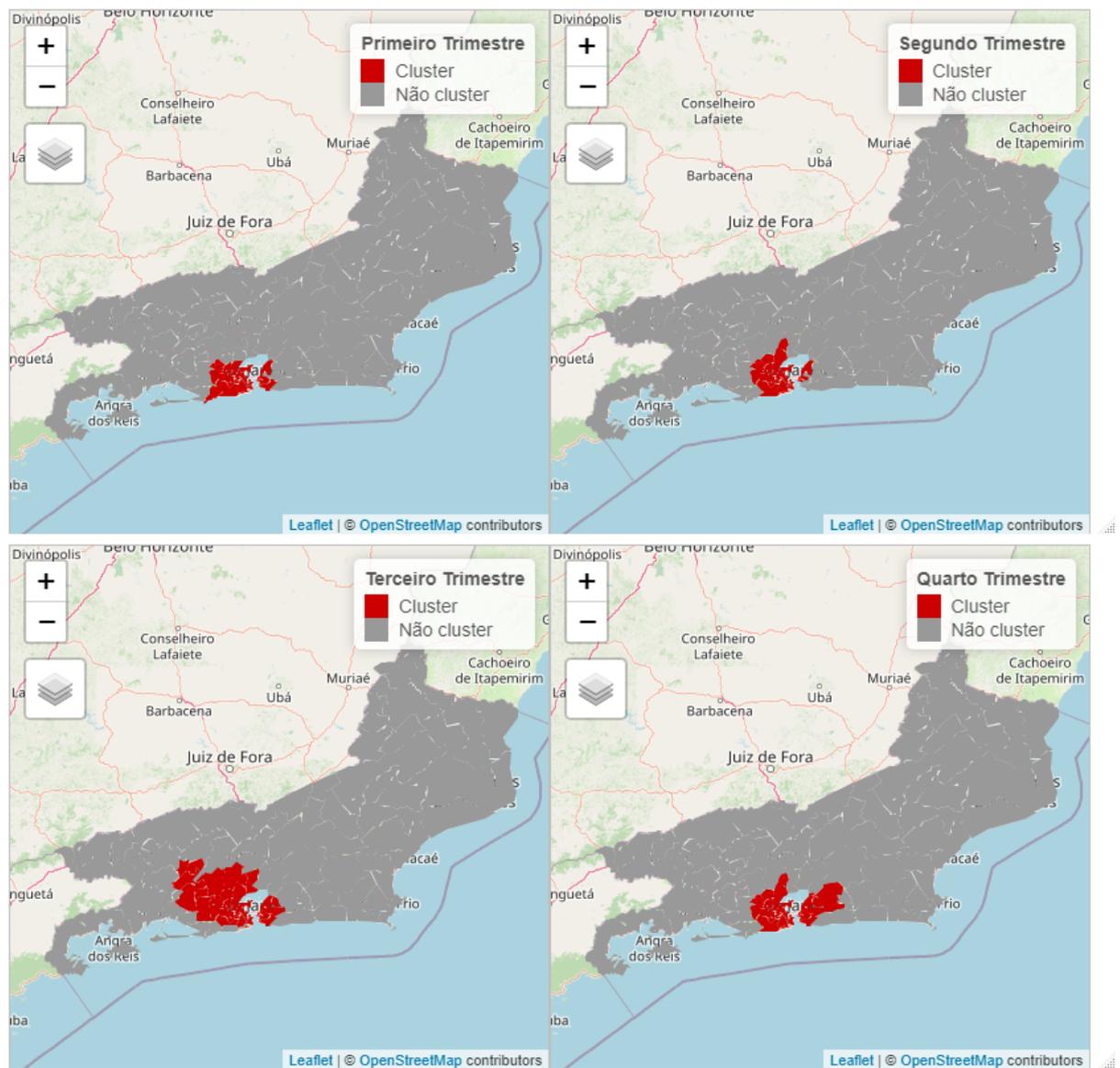


Figura 20: Cluster de alto risco para casos de roubo no estado do Rio de Janeiro em 2017.

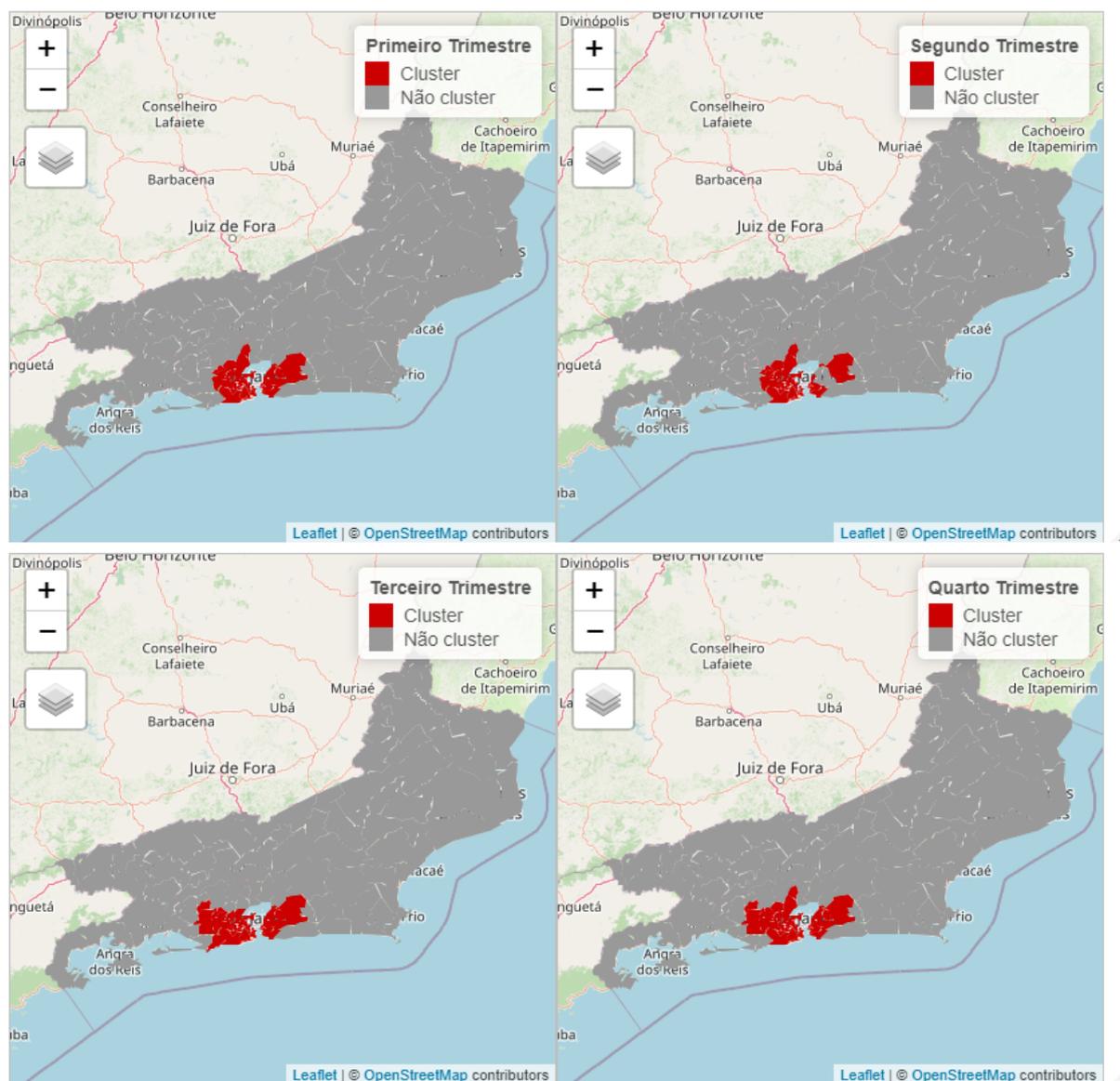


Figura 21: Cluster de alto risco para casos de roubo no estado do Rio de Janeiro em 2018.

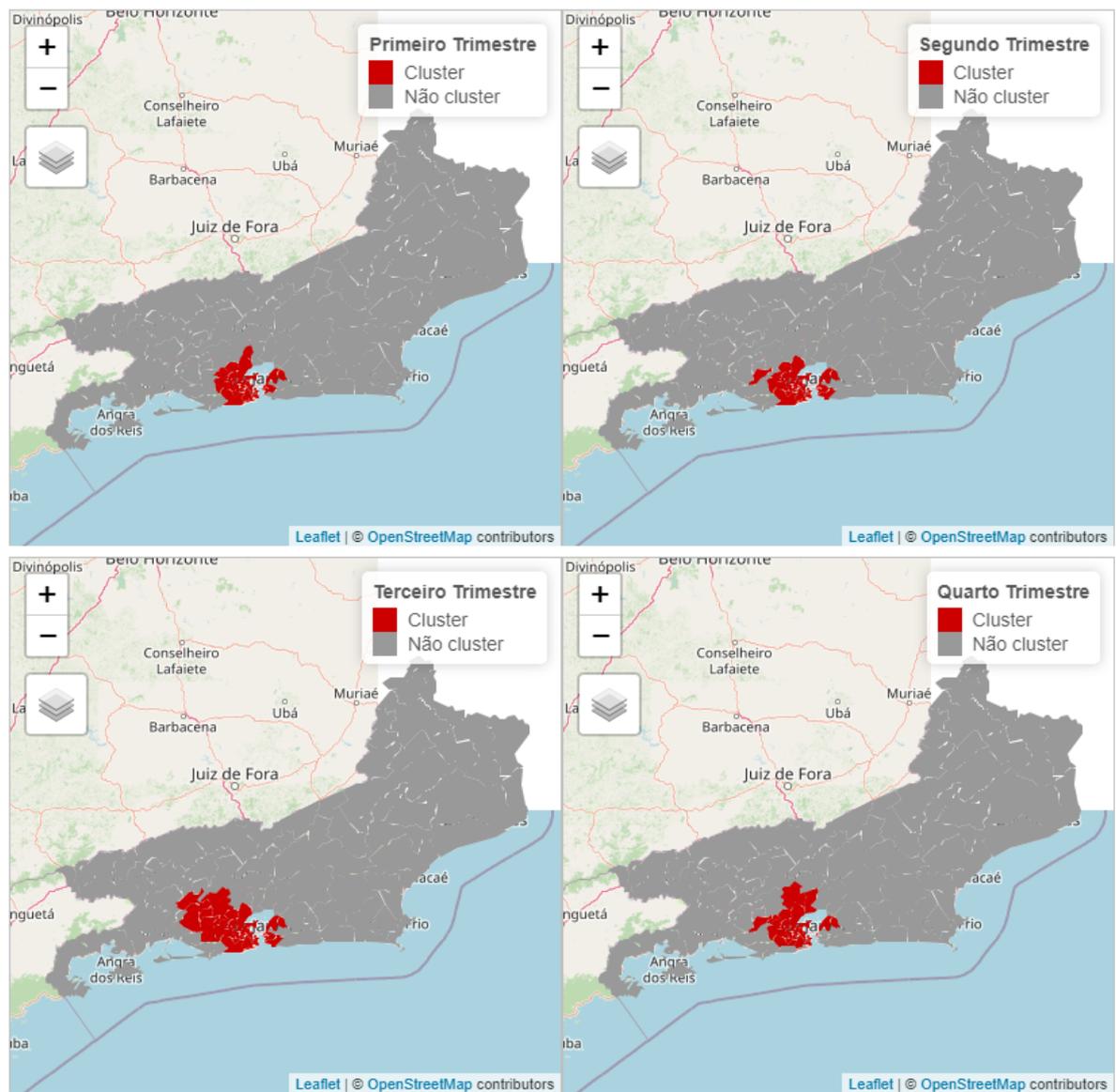


Figura 22: Cluster de alto risco para casos de roubo no estado do Rio de Janeiro em 2019.

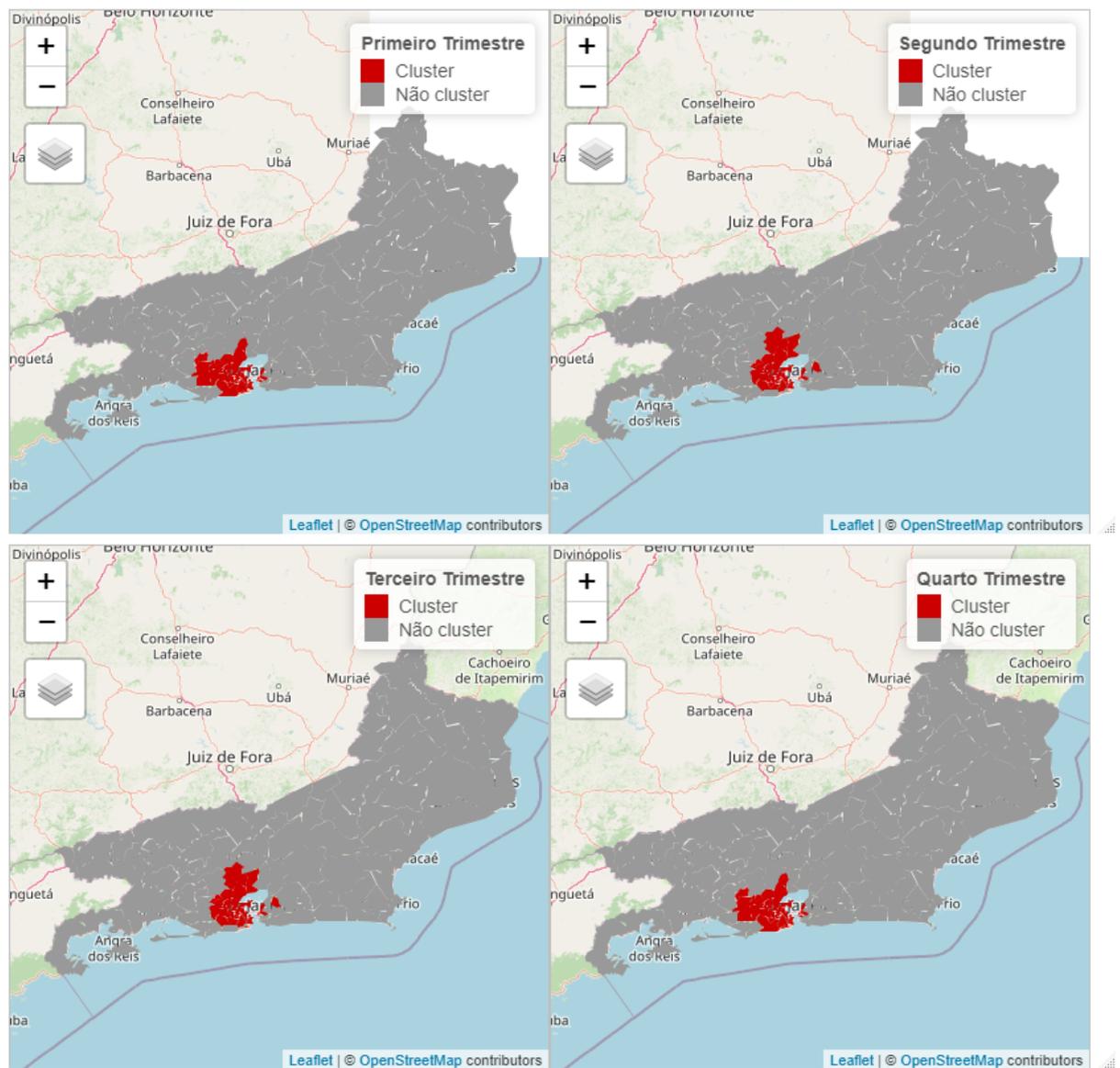


Figura 23: Cluster de alto risco para casos de roubo no estado do Rio de Janeiro em 2020.

A Figura 24 apresenta uma legenda de quantas vezes a CISP foi identificada como *cluster* de alto risco para o crime de roubo. As áreas identificadas como de alto risco apresentam algumas diferenças na comparação entre os trimestres, porém, nesse caso, são poucas as variações, e todas elas estão todas situadas na mesma mesoregião. Assim, é possível notar que há uma persistência de clusters de alto risco para os casos de roubo na Região Metropolitana do estado, sendo totalmente concentrados na parte litorânea da mesoregião.

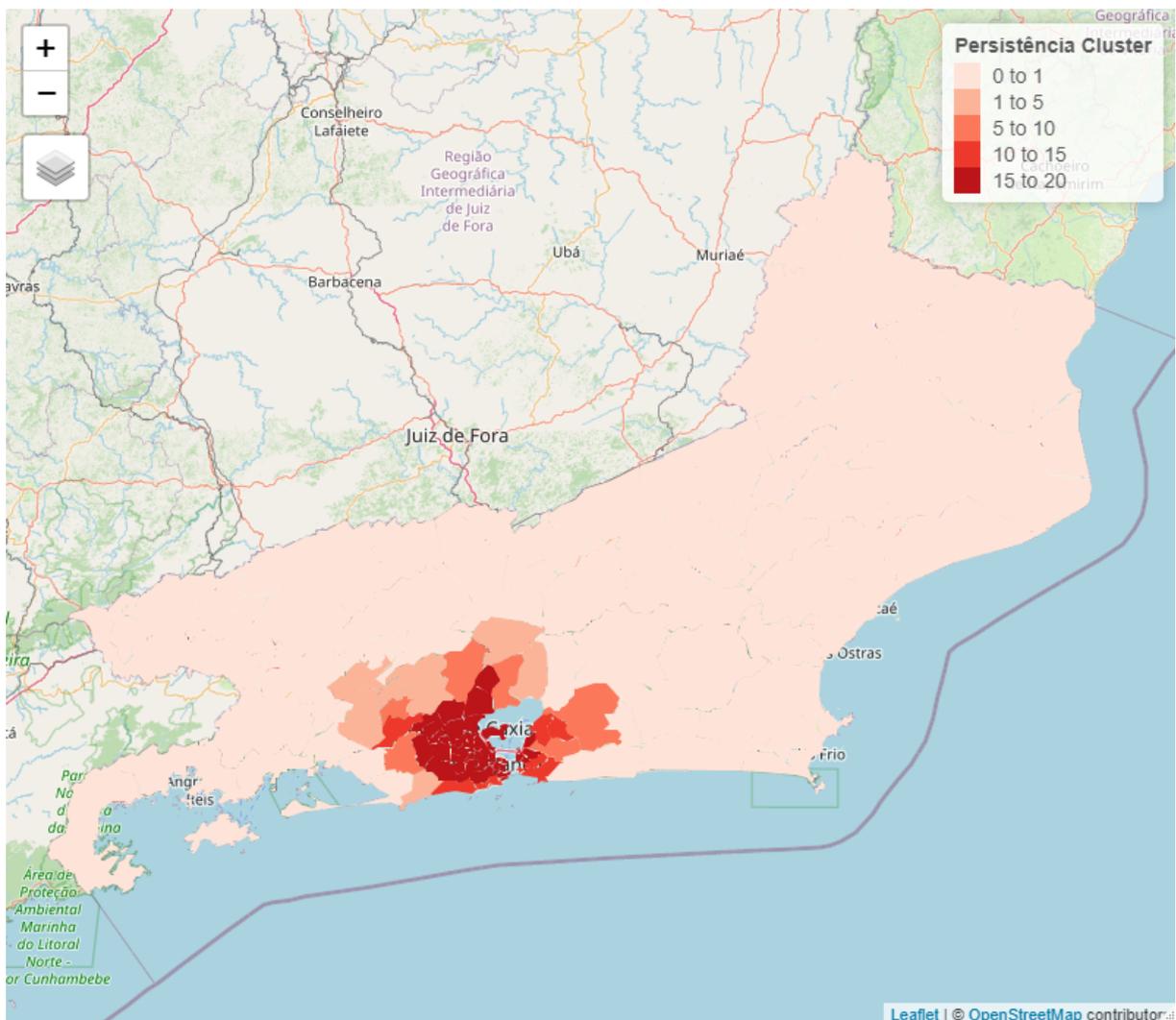


Figura 24: Persistência dos clusters de alto risco para casos de roubo no estado do Rio de Janeiro.

## 4 Conclusão

Este estudo teve por objetivo construir e analisar três diferentes tipos de crime para os municípios do Estado do Rio de Janeiro.

Dessa forma, neste trabalho foi verificada e validada a hipótese de persistência de clusters na ocorrência de casos de roubo no Brasil, utilizando a Estatística de Scan proposta por Kulldorff para identificação de clusters espaciais. Uma vez que a distribuição dessa estatística considerando a não existência de clusters é desconhecida, foram feitas diversas simulações com base nesta hipótese construindo-se uma distribuição  $LLR(z)$  para testar a significância dos conglomerados encontrados.

Para os outros dois tipos de crime, furto e letalidade violenta, foram aplicadas as mesmas técnicas estatísticas e análises, possibilitando observar uma leve variação nas regiões afetadas. O crime de letalidade violenta foi o que apresentou maior diferença nas regiões mais atingidas. Contudo, na maioria dos casos a região Metropolitana e Norte Fluminense foram as mais afetadas, indicando as áreas que devam ser o foco das políticas públicas ligadas a segurança.

Por meio desse estudo, foi possível enxergar melhor a dinâmica espacial dos crimes no Estado do Rio de Janeiro, algo que impacta a população fluminense há décadas. Foi identificado que a região Metropolitana é a localidade que deve receber maior atenção dos agentes públicos, com o objetivo de frear as ocorrências de criminalidade no estado.

Para além da alta densidade demográfica, a Região Metropolitana do Rio de Janeiro conta ainda com diversos problemas criminais que podem influenciar de forma direta ou indireta outros tipos de crime. Esses resultados podem ser explicados devido à alta densidade demográfica existente nessas regiões. Conforme Oliveira (2008) demonstrou, o tamanho e a densidade das cidades é capaz de explicar parte da criminalidade existente em uma determinada localidade. Assim sendo, a alta concentração do crime na Região Metropolitana era algo que já previsto.

Por fim, espera-se que os formuladores de políticas públicas se atentem ao fenômeno

da distribuição espacial do crime, de forma a se evitar o prolongamento da alta concentração criminosa em mais municípios. Para isso, faz-se necessário que os gestores de segurança pública do Rio de Janeiro destaquem seu olhar principalmente para a Região Metropolitana, mas também para os demais municípios. Além do mais, recomenda-se a atualização de soluções das políticas de segurança pública, através de maior foco no bem-estar da população de todo o estado.

Uma vez que realizou-se a análise espacial, torna-se possível e necessário uma análise espaço-temporal, a fim de verificar se existe dependência temporal na ocorrência de casos de um determinado crime em uma dada localidade. Com isso, há a possibilidade de investigar se o número de casos e sua localização apresentam dependência, ou não.

Além disso, um vez que os métodos já estão implementados, torna-se possível a atualização do estudo para que se possa seguir com o acompanhamento dos eventos. Ou seja, verificar, para os anos subsequentes, o comportamento dos crimes na cidade, e estabelecer possíveis dependências espaciais. Junto à isso, a análise pode ser aplicada considerando outras janelas temporais, com o objetivo de verificar se, dentro desse intervalo de tempo definido, o comportamento espacial dos clusters de alto risco são semelhantes ou não.

# Referências

BACANI, D. A.

*Análise espaço-temporal da malária no estado de Mato Grosso no período de 2003 a 2009* — Universidade Federal de Mato Grosso, 2013.

CERQUEIRA, D.; BUENO, S. Atlas da violência. 2019. Disponível em: [https://www.ipea.gov.br/portal/images/stories/PDFs/relatorio\\_institucional/190605\\_atlas\\_da\\_violencia\\_2019.pdf](https://www.ipea.gov.br/portal/images/stories/PDFs/relatorio_institucional/190605_atlas_da_violencia_2019.pdf).

CHESNAIS, J. C. A violência no brasil: causas e recomendações políticas para a sua prevenção. 1999. Disponível em: <https://www.scielo.br/j/csc/a/F8JZF8kmYgdmsS3PdFtLCPt/?lang=pt>.

CONGRESS, F. R. D. of the Library of. *Crime in Brazil*. [S.l.]: Washington, D.C., 1997.

CRESSIE, N. A. C. *Statistics for Spatial Data*. [S.l.]: John Wiley Sons, 1993.

FARAWAY, J. J. *Extending the Linear Model With R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*. [S.l.]: Chapman Hall, 2006.

FILHO, C. C. B. et al. Conglomerados de homicídios e o tráfico de drogas em belo horizonte, minas gerais, brasil, de 1995 a 1999. 2001. Disponível em: <https://www.scielo.br/j/csp/a/qmzW6Vv7FpZnt8Cm5FpLBXm/?lang=pt>.

ISP. *Divisão Territorial da Base de Segurança*. [S.l.], 2011.

KULLDORFF, M. *A spatial scan statistic, Communications in Statistics - Theory and Methods*. [S.l.]: Taylor Francis, 1997.

KULLDORFF, M. et al. Evaluating clusters alarms: A space-time scan statistic and brain cancer in los alamos. *American Journal of Public Health*, 1998.

KULLDORFF, M.; LAWSON, B. A. *A review of cluster detection methods. In: Disease Mapping and Risk Assessment for Public Health*. [S.l.]: John Wiley Sons, 1999.

LEWKOWICZ, R.

*Deteção de clusters: Uma análise de sazonalidade de surtos de dengue nos municípios do Brasil de 2007 a 2011* — Universidade de Brasília, 2013.

METROPOLIS, N. The monte carlo method. *Journal of the American Statistical Association*, 1949.

MINAMISAVA, R. et al. Spatial clusters of violent deaths in a newly urbanized region of brazil: highlighting the social disparities. 2009. Disponível em: <https://ij-healthgeographics.biomedcentral.com/articles/10.1186/1476-072X-8-66>.

MORAES, S. E. de Figueirêdo Lucena Ronei Marcos de. Método scan espaço-temporal: uma avaliação dos modelos poisson e permutação espaço-tempo. UFPB, 2010.

NAUS, J. I. The distribution of the size of the maximum cluster of points on a line. 1965. Disponível em: <[JournaloftheAmericanStatisticalAssociation](#)>.

OLIVEIRA, C. A. de. Análise espacial da criminalidade no rio grande do sul. 2008. Disponível em: <<https://revistas.ufpr.br/economia/article/view/13824/9315>>.

PROCOPIO, D. P.; TOYOSHIMA, S. H. Fatores associados à criminalidade violenta no brasil. 2017. Disponível em: <<https://seer.ufrgs.br/index.php/AnaliseEconomica/article/view/51903/42804>>.

RUEDIGER, M. A. Interiorização da violência no rio de janeiro (2003-2018). FGV DAPP, 2019.

SANTOS, P. S. dos et al. Criminalidade nos municípios do rio de janeiro: uma análise multivariada e espacial. 2020. Disponível em: <<https://revistas.ufpr.br/economia/article/view/71723>>.

SANTOS, P. S. dos et al. Criminalidade nos municípios do rio de janeiro: uma análise multivariada e espacial. 2020. Disponível em: <<https://revistas.ufpr.br/economia/article/view/71723>>.

SZWARCWALD, C. L.; CASTILHO, E. A. de. Mortalidade por armas de fogo no estado do rio de janeiro, brasil: uma análise espacial. 1998. Disponível em: <<https://iris.paho.org/bitstream/handle/10665.2/7802/4n3a3.pdf?sequence=1&isAllowed=y>>.

TURISRIO. *Divisão Territorial do Rio de Janeiro*. [S.l.], 2020.

UCHOA, C. F. A.; MENEZES, T. A. D. Spillover espacial da criminalidade: Uma aplicação de painel espacial, para os estados brasileiros. 2012. Disponível em: <[https://www.anpec.org.br/encontro/2012/inscricao/files\\_I/i9-fa2b8ebb634f17963afe1f307ae436f3.pdf](https://www.anpec.org.br/encontro/2012/inscricao/files_I/i9-fa2b8ebb634f17963afe1f307ae436f3.pdf)>.