

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE GOIÁS
DEPARTAMENTO DE ÁREAS ACADÊMICA III
CORDENAÇÃO DE GEOMÁTICA
CURSO DE ENGENHARIA CARTOGRÁFICA E DE AGRIMENSURA

REIZANE MARIA DAMASCENO DA SILVA

**RELAÇÃO DOS CASOS DE DENGUE E FATORES SOCIOECONÔMICOS E
AMBIENTAIS EM APARECIDA DE GOIÂNIA-GO BASEADO EM ANÁLISE
ESPACIAL E MINERAÇÃO DE DADOS**

GOIÂNIA, GO
2019

REIZANE MARIA DAMASCENO DA SILVA

**RELAÇÃO DOS CASOS DE DENGUE E FATORES SOCIOECONÔMICOS E
AMBIENTAIS EM APARECIDA DE GOIÂNIA-GO BASEADO EM ANÁLISE
ESPACIAL E MINERAÇÃO DE DADOS**

Trabalho de Conclusão de Curso (TCC)
apresentado à Coordenação de Geomática,
do Instituto Federal de Educação, Ciência
e Tecnologia de Goiás, Campus Goiânia,
como parte dos requisitos necessários para
a obtenção do título de Engenharia
Cartográfica e de Agrimensura.

Orientador: Prof. Dr. Giovanni de Araújo
Boggione

GOIÂNIA, GO

2019

Si381r Silva, Reizane Maria Damasceno da.
Relação dos casos de Dengue e fatores socioeconômicos e ambientais em Aparecida de Goiânia-GO baseado em análise espacial e mineração dos dados / Reizane Maria Damasceno da Silva. – Goiânia: Instituto Federal de Educação, Ciência e Tecnologia de Goiás, Câmpus Goiânia, 2019.
82 f.: il.

Orientador: Prof. Dr. Giovanni de Araújo Boggione.

TCC (Trabalho de Conclusão de Curso) – Curso de Engenharia Cartográfica e de Agrimensura, Departamento de Áreas Acadêmicas III, Instituto Federal de Educação, Ciência e Tecnologia de Goiás, Câmpus Goiânia.

1. Sistema de Informação Geográfica. 2. Dengue. 3. Aparecida de Goiânia – GO. I. Boggione, Giovanni de Araújo (orientador). II. Instituto Federal de Educação, Ciência e Tecnologia de Goiás, Câmpus Goiânia. III. Título.

CDD 910.285

Ficha catalográfica elaborada pela Bibliotecária Karol Almeida da Silva Abreu CRB1/ 2.740
Biblioteca Professor Jorge Félix de Souza,
Instituto Federal de Educação, Ciência e Tecnologia de Goiás, Câmpus Goiânia.

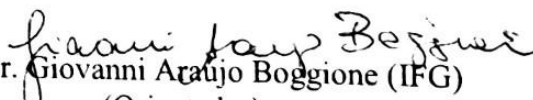
TERMO DE APROVAÇÃO

REIZANE MARIA DAMASCENO DA SILVA

RELAÇÃO DOS CASOS DE DENGUE E FATORES SOCIOECONÔMICOS E AMBIENTAIS EM APARECIDA DE GOIÂNIA BASEADA EM ANÁLISE ESPACIAL E MINERAÇÃO DE DADOS

Trabalho de Conclusão de Curso apresentado à Coordenação de Geomática, do Instituto Federal de Educação, Ciência e Tecnologia de Goiás, Câmpus Goiânia, como parte dos requisitos necessários para a obtenção do título de Bacharelado em Engenharia Cartográfica e de Agrimensura.

Aprovado em, 19 de dezembro de 2019.


Prof. Dr. Giovanni Araujo Boggione (IFG)
(Orientador)


Prof. Dr. Edipio Henrique Cremon (IFG)
(Membro da Banca Examinadora)


Prof. Dr. João Paulo Magna Junior (IFG)
(Membro da Banca Examinadora)

GOIÂNIA, GO

2019



INSTITUTO FEDERAL
Goiás

MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
SISTEMA INTEGRADO DE BIBLIOTECAS

TERMO DE AUTORIZAÇÃO PARA DISPONIBILIZAÇÃO NO REPOSITÓRIO DIGITAL DO IFG - ReDi IFG

Com base no disposto na Lei Federal nº 9.610/98, AUTORIZO o Instituto Federal de Educação, Ciência e Tecnologia de Goiás, a disponibilizar gratuitamente o documento no Repositório Digital (ReDi IFG), sem ressarcimento de direitos autorais, conforme permissão assinada abaixo, em formato digital para fins de leitura, download e impressão, a título de divulgação da produção técnico-científica no IFG.

Identificação da Produção Técnico-Científica

- | | |
|--|---|
| ☐ Tese | ☐ Artigo Científico |
| ☐ Dissertação | ☐ Capítulo de Livro |
| ☐ Monografia – Especialização | ☐ Livro |
| ☒ TCC - Graduação | ☐ Trabalho Apresentado em Evento |
| ☐ Produto Técnico e Educacional - Tipo: _____ | |

Nome Completo do Autor:

Matrícula:

Título do Trabalho:

Autorização - Marque uma das opções

1. ☒ Autorizo disponibilizar meu trabalho no Repositório Digital do IFG;
2. ☐ Autorizo disponibilizar meu trabalho no Repositório Digital do IFG somente após a data ____/____/____;
3. ☐ Não autorizo disponibilizar meu trabalho no Repositório Digital do IFG.

Ao indicar a opção **2** ou **3**, marque a justificativa:

- ☐ O documento está sujeito a registro de patente.
☐ O documento pode vir a ser publicado como livro, capítulo de livro ou artigo.
☐ Outra justificativa: _____

DECLARAÇÃO DE DISTRIBUIÇÃO NÃO-EXCLUSIVA

O/A referido/a autor/a declara que:

- i. o documento é seu trabalho original, detém os direitos autorais da produção técnico-científica e não infringe os direitos de qualquer outra pessoa ou entidade;
- ii. obteve autorização de quaisquer materiais inclusos no documento do qual não detém os direitos de autor/a, para conceder ao Instituto Federal de Educação, Ciência e Tecnologia de Goiás os direitos requeridos e que este material cujos direitos autorais são de terceiros, estão claramente identificados e reconhecidos no texto ou conteúdo do documento entregue;
- iii. cumpriu quaisquer obrigações exigidas por contrato ou acordo, caso o documento entregue seja baseado em trabalho financiado ou apoiado por outra instituição que não o Instituto Federal de Educação, Ciência e Tecnologia de Goiás.

Goiânia, 02/03/2020
Local Data

Reizane Maria Damasceno da Silva

Assinatura do Autor e/ou Detentor dos Direitos Autorais

DEDICATÓRIA

À memória de José Maria Vieira da Silva, por ter sido um pai incrível, um amigo fiel,
o melhor cozinheiro do mundo, por não me deixar desistir dos meus sonhos, por ter sido um
herói... meu herói do sorvete.

AGRADECIMENTOS

Agradeço a Deus, a Ele toda honra e toda glória. À minha mãe, que é a mulher mais forte e maravilhosa de todo o meu mundo, por ter ficado acordada do meu lado durante todas as noites em claro e por financiar todos os meus calmantes, aos meus irmãos pelas maratonas do senhor dos anéis e por serem tão legais, principalmente a minha irmã Vanessa, que me legou o amor pela leitura e pelo estudo, e por todas as vezes que, com paciência, me orientava, moldando meu caráter e me enviava *memes* para me fazer sorrir!

Agradeço ao professor Dr. Giovanni Boggione, por ter aceitado me orientar neste trabalho, pelas risadas, comentários maldosos, e valiosos conselhos em toda a trajetória. Obrigada por ter sido o Gandalf, por me deixar ser o Legolas e por me conduzir por essa jornada pela Terra média do geoprocessamento.

Agradeço a Dra. Flavia Feitosa, pelo suporte no pouco tempo que lhe coube e por ter contribuído de forma grandiosa para a conclusão desse trabalho. Seu olhar crítico e sua experiência foram determinantes na parte de análise estatística.

Agradeço aos meus parentes, amigos e professores por tornarem essa etapa da minha vida mais leve e divertida. A Leandro Gonçalves de Sousa, por todas as dicas de geoprocessamento.

Ao IFG pela oportunidade e apoio, bem como a todos os seus servidores, independentemente de cargos ou títulos.

E por último, mas de modo especial, aos meus colegas e amigos de turma que se tornaram uma extensão da minha família nesses cinco anos. Obrigada por todas as partidas de Uno, pela parceria nos trabalhos, pelos debates sobre a Marvel, pelas risadas e conversas bobas. Foi uma honra conhecer e dividir essa etapa da vida com vocês.

EPÍGRAFE

Não é sobre mim, não é sobre você também.
É sobre o legado, o legado que deixamos
para as gerações futuras. Não é sobre nós.

(Tony Stark- Homem de Ferro)

RESUMO

RELAÇÃO DOS CASOS DE DENGUE E FATORES SOCIOECONÔMICOS E AMBIENTAIS EM APARECIDA DE GOIÂNIA-GO BASEADO EM ANÁLISE ESPACIAL E MINERAÇÃO DE DADOS

AUTORA: Reizane Maria Damasceno da Silva

ORIENTADOR: Prof. Dr. Giovanni de Araújo Boggione

Considerada um grande problema de saúde pública do Brasil, a dengue apresentou uma difusão incontrolada no país ao longo dos últimos anos, revelando tendências ascendentes da transmissão da doença. É atualmente uma das maiores preocupações da gestão de saúde devido a sua transmissão ser essencialmente urbana, ambiente no qual estão presentes todos os fatores essenciais para a sua ocorrência: o homem, o vírus, o vetor e as condições econômicas, políticas e culturais que favorecem o estabelecimento e propagação da contaminação. Devido a essas peculiaridades, analisar os fatores de risco bem como sua relação para o desenvolvimento do vetor da doença, o mosquito *Aedes aegypti*, é um ponto crucial para medidas de intervenção na prevenção do mosquito. O objetivo desse estudo foi analisar quais fatores socioambientais são mais presentes nos ambientes com maior incidência de casos, caracterizando o ambiente favorável a proliferação da doença. Para atingir o objetivo foi realizado um processo de mineração de dados para descobrir correlações, padrões e tendências entre as informações dos casos de dengue e os indicadores socioambientais. Uma vez identificadas as variáveis com maior influência, foi possível analisar, a partir do cruzamento de dados em um ambiente de SIG associado a um Banco de Dados Geográfico, a relação entre a dengue e os indicadores de renda, de educação e de saneamento básico, bem como, do adensamento demográfico. A correlação espacial entre os casos de dengue e os indicadores sociais foi realizada e ao contrário do que se esperava, as variáveis escolhidas demonstraram baixa correlação espacial, o que sugere que para o Município de Aparecida de Goiânia, variáveis como renda, saneamento básico, densidade populacional e vazios urbanos não são suficientes para explicar a dengue.

Palavras-chave: Dengue, Sistema de Informação Geográfica, indicadores sociais, Correlação espacial.

ABSTRACT

SOCIOECONOMIC AND ENVIRONMENTAL FACTORS OF RATIO OF DENGUE CASES IN APARECIDA DE GOIÂNIA-GO BASED IN SPATIAL ANALYSIS AND DATA MINING

AUTHOR: Reizane Maria Damasceno da Silva

ADVISOR: Prof. Dr. Giovanni de Araújo Boggione

Considered a important public health problem in Brazil, dengue has spread uncontrollably throughout the country over the past few years, revealing upward trends in the disease transmission. It is currently one of the major concerns of health management because its transmission is essentially urban, an environment in which all the essential factors for its occurrence are present: the man, the virus, the vector and the economic, political and cultural conditions that contribute for the establishment and spread of contamination. Due to these peculiarities, analyzing risk factors as well as their relationship to the development of the disease vector, the Aedes aegypti mosquito, is a crucial point for intervention measures in mosquito prevention. The aim of this study was to analyze which socioenvironmental factors are more present in the environments with higher incidence of case, characterizing the favorable environment for the proliferation of the disease. To achieve the goal, a data mining process was conducted to find correlations, patterns and trends between dengue cases information and socioenvironmental indicators. Once the most influential variables were identified, it was possible to analyze, from the crossing of data in a GIS environment associated with a Geographic Database, the relationship between dengue and the indicators of income, education and sanitation, as well as of the demographic densification. The spatial correlation between dengue cases and social indicators was performed and contrary to what was expected, the chosen variables showed very low spatial correlation, which suggest that for the municipality of Aparecida de Goiânia, variables such as income, basic sanitation, density Population and urban voids are not sufficient to explain dengue.

Keywords: Dengue, Geographic Information System, social indicators, Spatial correlation.

LISTA DE FIGURAS

Figura 1 - ciclo de vida do <i>Aedes aegypti</i> e qual sua área de ação.....	23
Figura 2 – Convenção de contiguidade Torre e Rainha.....	33
Figura 3 – Diagrama de dispersão de Moran.....	34
Figura 4 - Etapas do processo de KDD definidas por Fayyad.....	38
Figura 5 - Localização da Área de Estudo – Aparecida de Goiânia-GO.....	41
Figura 6 - Fluxograma Metodológico.....	45
Figura 7 – Fluxograma da classificação orientada a objeto do ortomosaico do município de Aparecida de Goiânia.....	49
Figura 8 – Modelo OMT-G do banco de dados geográfico.....	52
Figura 9 – Fluxograma da mineração de padrões frequentes com o Algoritmo Apriori.....	54
Figura 10 – Área de influência das unidades de saúde.....	58
Figura 11 – Taxa de casos de Dengue nos setores Censitários em Aparecida de Goiânia.....	59
Figura 12 – A) Taxa de setores com coleta de Lixo e B) Taxa de setores sem coleta de lixo e esgoto.....	60
Figura 13 – A) Taxa de setores com abastecimento de água e B) Taxa de residências com coleta de esgoto.....	60
Figura 14 – A) Densidade populacional e B) Taxa de dependente.....	61
Figura 15 – A) Renda média e B) Renda do responsável.....	62
Figura 16 – A) Média de moradores por domicílio e B) Taxa de vazios urbanos.....	62
Figura 17 – Variáveis selecionadas pelo minerador de dados.....	64
Figura 18 – Histograma de conectividade para o critério contingência <i>queen</i>	65
Figura 19 – Gráfico de espalhamento de Moran.....	67
Figura 20 – LISA <i>map</i> , indicando clusters de distribuição de alto e baixos valores.....	69
Figura 21 –Taxa de água dos setores com correlação high-high.....	69
Figura 22 –Densidade dos setores com correlação high-high.....	70
Figura 23 – Media de moradores dos setores com correlação high-high.....	70
Figura 24 – Renda do responsável dos setores com correlação high-high.....	71
Figura 25 – Taxa de residências sem coleta de lixo e esgoto dos setores com correlação high-high.....	71

Figura 26 – Taxa de vazios urbanos dos setores com correlação high-high.....	72
Figura 27 –Residual da Ponderação geograficamente ponderada.....	73

LISTA DE TABELAS

Tabela 1 – Teste de autocorrelação espacial univariada.....	66
Tabela 2 – Teste de autocorrelação espacial bivariada.....	67

LISTA DE QUADROS

Quadro 1 – Técnica de análise espacial em SIG para aplicação de saúde e ambiente...	27
Quadro 2 – Tipo de dados e problemas em Análise Espacial.....	28
Quadro 3 – Municípios com maior coeficiente de incidência das durante as semanas 20 a 23 de 2019.....	42
Quadro 4 – Municípios com maior número de casos notificados de dengue em Goiás no período da semana 01 a 23 de 2019.....	43
Quadro 5 – Municípios com óbitos suspeitos e confirmados de dengue em Goiás no período da semana 01 a 23 de 2019.....	43
Quadro 6 – Variáveis.....	46
Quadro 7 – Classes associadas a cada uma das variáveis.....	55

LISTA DE ABREVIATURAS E SIGLAS

AEDE	Análise Exploratória de Dados Espaciais
CAR	Spatial Error Model
DEN 1	Sorotipo 1 do vírus da Dengue
DEN 2	Sorotipo 2 do vírus da Dengue
DEN 3	Sorotipo 3 do vírus da Dengue
DEN 4	Sorotipo 4 do vírus da Dengue
DF	Dependências funcionais
IBGE	Instituto Brasileiro de Geografia e Estatística
KDD	Knowledge Discovery in Databases
LISA	Local Indicators of Spatial Association
MAUP	Modifiable Areal Unit Problem
MI	Mortalidade Infantil
OMS	Organização Mundial da Saúde
OPAS	Organização Pan Americana da Saúde
RGP	Regressão Geograficamente Ponderada
SAR	Spatial LAG Model
SE	Semana Epidemiológica
SES-GO	Secretária de Estado da Saúde de Goiás
SGBD	Sistema Gerenciador de Banco de Dados
SIA	Sistema de Informação Sobre Procedimentos Ambulatoriais
SIG	Sistema de Informação Geográfica
SIM	Sistema de Informação de Mortalidade
SINAN	Sistema de Informação de Agravos e Notificação
SINASC	Sistema de Informações de Nascidos Vivos
SNIS	Sistema Nacional de Informações sobre Saneamento
SUS	Sistema Único de Saúde
UML	Unified Modeling Language

SUMÁRIO

INTRODUÇÃO	17
1.1 TEMA.....	19
1.2 JUSTIFICATIVA.....	19
2 OBJETIVOS	20
2.1 OBJETIVO GERAL.....	20
2.2 OBJETIVOS ESPECÍFICOS	20
3 REVISÃO DA LITERATURA	21
3.1 EPIDEMIOLOGIA	21
3.1.1 Dengue	21
3.1.2 Aedes aegypti	22
3.2 SANEAMENTO BÁSICO	23
3.3 BANCO DE DADOS GEOGRÁFICO APLICADO À SAÚDE	24
3.4 ESTATÍSTICA ESPACIAL	26
3.4.1 Análise espacial aplicada em estudos de saúde.....	27
3.4.2 Análise Exploratória de Dados Espaciais (AEDE).....	27
3.4.3 Dependência Espacial.....	31
3.4.4 Análise de variáveis Espaciais de Áreas	31
3.5 ANÁLISE INFERENCIAL.....	32
3.5.1 Matriz de proximidade espacial.....	32
3.5.2 Autocorrelação.....	33
3.5.3 Representação Gráfica: <i>BOX MAP</i> , <i>LISA Map</i> e <i>Moran Map</i>	35
3.5.4 Teste de Significância.....	35
3.5.5 Regressão espacial.....	35
3.5.6 Regressão geograficamente ponderada.....	36
3.6 DESCOBERTA DE CONHECIMENTO EM BANCO DE DADOS	36
3.7 NORMALIZAÇÃO	38
3.8 MINERAÇÃO DE DADOS	39
3.9 APARECIDA DE GOIÂNIA	41
4. METODOLOGIA.....	44

4.1 SELEÇÃO DE VARIÁVEIS.....	45
4.2 VAZIOS URBANOS.....	47
4.3 ESTRUTURAÇÃO DO BANCO DE DADOS GEOGRÁFICO	51
4.4 MINERAÇÃO DE DADOS	53
4.5 POLIGONOS DE THIESSEN.....	55
4.6 ANÁLISE DE DADOS DE ÁREA	56
4.7 REGRESSÃO ESPACIAL.....	57
5 RESULTADOS E DISCUSSÕES	57
6. CONSIDERAÇÕES FINAIS.....	73
6. REFERÊNCIAS	75

1 INTRODUÇÃO

A dengue é uma *arbovirose* urbana de notificação compulsória, considerada uma grande ameaça à saúde pública na maioria dos países tropicais, devido às suas condições climáticas e ambientais, favoráveis à proliferação do vetor da doença, a fêmea do mosquito *Aedes aegypti*, que deposita seus ovos próximos às superfícies de água limpa. Existem quatro tipos de vírus em circulação, que podem causar no infectado desde sintomas brandos como febre, dor de cabeça, dor nas articulações e atrás dos olhos, até sintomas mais graves como febre hemorrágica, vômitos persistentes, queda abrupta das plaquetas e delírios, que quando não tratados podem levar a óbito.

A dengue apresentou uma difusão incontrolada no país ao longo dos últimos anos, revelando tendências ascendentes da transmissão da doença. Segundo a Secretaria de Vigilância em saúde, até a Semana Epidemiológica (SE) 11 (30/12/2018 a 16/03/2019) foram registrados 229.064 casos prováveis de dengue ao longo de todo território nacional. Um dado que assusta em comparação ao mesmo período da análise anterior, que registrou 62.904 casos. O país tem dedicado recursos significativos para a erradicação do vetor da dengue, no entanto desde a década de 1990 o intervalo entre surtos da doença tem sido cada vez menor e as áreas de abrangência desses surtos tem sido cada vez maior. A tendência de aumento de casos da doença, apesar das medidas de saúde, controle e campanhas de conscientização, tem preocupado a opinião pública

A transmissão da dengue é essencialmente urbana, ambiente no qual estão presentes todos os fatores essenciais para a sua ocorrência: o homem, o vírus, o vetor e as condições econômicas, políticas e culturais que favorecem o estabelecimento e propagação da contaminação. O adensamento de moradias devido ao crescimento urbano propicia um grande volume de indivíduos suscetíveis à infecção e o estabelecimento de uma cadeia de transmissão. Fatores socioambientais como moradias inadequadas e condições precárias de saneamento básico, associados a ineficácia de políticas de saúde pública, tem influenciado a propagação do vetor e aumento dos casos de dengue no país.

Segundo dados da Organização Pan Americana da Saúde (OPAS), os cinco estados do Brasil que mais concentraram casos de dengue nas seis primeiras semanas do ano de 2010 foram Rondônia, Acre, Mato Grosso do Sul, Mato Grosso e Goiás (OPAS, 2010). Dos 108.640 casos em todo o país, 77.117 ocorreram nesses estados, dos quais 30% se concentraram nos municípios de Goiânia (GO), Aparecida de Goiânia (GO), Campo Grande (MT), Rio Branco

(AC) e Porto Velho (RO). A falta de saneamento básico foi apontada como uma explicação para a alta incidência da doença. A falta de regularidade no abastecimento de água, leva as pessoas à armazenarem água, criando pontos suscetíveis a proliferação do vetor (OPAS, 2010).

De acordo com a Secretaria de Estado da Saúde de Goiás (SES-GO), os municípios de Goiânia e Aparecida de Goiânia mantêm-se em estado de alerta. As secretarias municipais têm intensificado as ações de combate à doença e o atendimento aos casos suspeitos tem sido aprimorado nas unidades de saúde de atendimento 24 horas e as visitas domiciliares tem se intensificado. No entanto não há registro de intensificação de políticas para o aumento de taxa de saneamento básico. No ranking divulgado pelo Instituto Trata Brasil (2016), Aparecida de Goiânia está na décima posição entre os piores índices de coleta de esgoto, e dados do Sistema Nacional de Informações sobre Saneamento (SNIS), apontam que o município está na lista dos piores em fornecimento de água, ocupando a 92ª posição de um total de 100 municípios (OLIVEIRA; SCAZUFCA; PIRES, 2018).

Segundo Antunes (2017), até 2017, 65,57% da população do município de Aparecida de Goiânia tinha atendimento de água tratada e 20,8% recebia atendimento de esgoto. No mesmo período a SES-GO confirmou 8.169 casos de dengue no município, o que o tornou o segundo município do estado de Goiás com maior ocorrência da doença, perdendo apenas para a capital, classificando como médio risco.

A secretaria de vigilância em saúde, atenta à disseminação da doença, tem gerado relatórios com periodicidade mensal, com o quantitativo de casos registrados da doença em cada região, com o intuito de entender o fenômeno e conseguir detectar precocemente novas epidemias, controlar as epidemias em curso, e reduzir o risco de transmissão da doença em áreas endêmicas. A frequência de epidemias da doença em algumas localidades e o fato da doença ser causada por agente transmissor que se adapta melhor a um determinado ambiente chama atenção ao estudo da distribuição espacial dos casos registrados da doença, bem como a interação dos fatores de risco para o desenvolvimento do vetor de transmissão da doença. Analisar dados, é descobrir *insights* [grifo da autora] ocultos, é dar significado aos dados. A análise espacial se configura no princípio da descoberta de informações, mas levando em consideração a localização espacial do fenômeno em estudo.

Conhecer a estrutura e a dinâmica espacial da população é o primeiro passo para identificação de situações de saúde. Uma das maneiras de conhecer as condições de saúde da população é por meio de mapas que permitam observar a distribuição espacial de situações de risco e dos problemas de saúde (CARVALHO; PINA; SANTOS, 2000). Há pelo menos três

décadas, a área da saúde vem discutindo, estudando e implementando técnicas que abordam a localização espacial e os Sistemas de Informação Geográficas (SIG). O Objetivo é usar ferramentas de manipulação de informações espacialmente referidas e aplicá-las a questões de saúde pública, permitindo mapeamento dos fenômenos e a análise de riscos. Segundo Leite (2010), o uso de tecnologia de informações geográficas tem se tornado comum para subsidiar a tomada de decisões, pois trazem resultados precisos com certa velocidade, além da possibilidade de auxiliar análises de situações específicas com a integração de dados de diferentes fontes e de diversas áreas.

Uma abordagem espacial proporciona a integração de dados demográficos, ambientais e socioeconômicos, e promove a inter-relação de dados de diversos bancos de dados, que fornecem a matéria prima para compreensão dos fenômenos. Uma base de dados espaciais é composta por elementos de diversas naturezas e a devida integração desses dados é essencial para o diagnóstico dos fenômenos e a análise de risco. Dessa forma é necessário um estudo do tipo de dado que será utilizado e definir um arranjo de banco de dados geográfico totalmente preparado que possibilite a integração e manipulação desses dados.

Nesse contexto de distribuição espacial e relação entre variáveis, a utilização de técnicas que visam à análise da distribuição espacial, cruzamento de dados e mineração de dados, permite produzir informações que levem a compreensão da dinâmica da doença e suas relações, e que podem ser usadas para definir estratégias de ação para mitigar a atuação do vetor e prevenção da dengue.

1.1TEMA

Análise das relações espaciais existentes entre saneamento básico, indicadores socioeconômicos e ambientais, e casos registrados de dengue no município de Aparecida de Goiânia - GO.

1.2 JUSTIFICATIVA

As análises das relações entre os casos de dengue e agentes externos têm potencial para contribuir com a orientação de ações em Saúde Pública no país.

A luta pela eliminação do vetor tem sido intensa desde o início do século passado, os tempos de Oswaldo Cruz, o pioneiro no estudo de doenças tropicais e coordenador de

campanhas para eliminação dos focos de insetos transmissores de doenças tropicais. No entanto, mesmo após várias campanhas e tentativas de erradicação, o vetor continua presente em todo o território nacional (BARBOSA; LOURENÇO, 2010).

As peculiaridades da dengue, como o fato da doença ser causada por agente transmissor que se adapta melhor a um determinado ambiente traz uma impressão de que ela atinge a todos por igual. A ineficácia na eliminação do vetor tem despertado interesse na compreensão da distribuição espacial dos casos, no estudo de fatores externos ao vetor que possam estar contribuindo para a disseminação da doença e nas relações existentes entre essa distribuição e fatores externos.

Nesse cenário, é necessário lançar mão de técnicas de análise espacial utilizando análises exploratórias, empregando técnicas de mineração de dados e modelos estatísticos, com o intuito de compreender os fatores implícitos à doença que expliquem as condições epidemiológicas de algumas áreas e a desigualdade da distribuição, de forma que seja possível usar a informação gerada para identificação de áreas prioritárias para intervenção de saúde e também servir de base para definição de planos para prevenção da doença, bem como criar cenários para entender a dinâmica da doença no município.

2 OBJETIVOS

2.1 OBJETIVO GERAL

O objetivo principal deste trabalho consiste em produzir/analisar informações que permitam o entendimento das relações espaciais existentes entre saneamento básico, indicadores sociais e casos de dengue em Aparecida de Goiânia, a partir de técnicas de cruzamento de dados, e mineração de dados.

2.2 OBJETIVOS ESPECÍFICOS

- Analisar a ocorrência de caso de dengue, e o padrão de distribuição espacial da doença no município de Aparecida de Goiânia;
- Demonstrar as técnicas de Análise Estatística Espacial de Áreas para a obtenção de respostas não alcançadas em meio a métodos convencionais de análise;
- Verificar a correlação espacial de indicadores socioambientais, utilizando índice global de Moran;

-Determinar os principais indicadores e quantificar a influência dos determinantes sociais e ambientais na distribuição espacial dos casos.

3 REVISÃO DA LITERATURA

Neste capítulo, baseado em uma revisão de literatura, é apresentado os principais conceitos sobre: dengue, como é sua transmissão e sua relação com o ser humano e o ambiente no qual está inserido, dando ênfase ao saneamento básico, banco de dados geográfico e sua aplicação na área da saúde, análise espacial e destacando seus princípios básicos e sua aplicação em estudos de saúde. Um destaque também é dado à descoberta de conhecimento em banco de dados e mineração de dados abordado.

3.1 EPIDEMIOLOGIA

3.1.1 Dengue

A dengue tem ganhado destaque entre as doenças emergentes em virtude da quantidade de casos registrados no mundo. Trata-se de uma doença viral do gênero *Favivírus* transmitida de uma pessoa infectada por um dos quatro tipos do vírus **DEN 1, DEN 2, DEN 3 e DEN 4**, para outra sadia pelo vetor da doença. Esse vetor trata-se da fêmea do mosquito *Aedes aegypti*, que pica para amadurecer seus ovos, e se torna contaminada quando suga o sangue de um indivíduo já portador do vírus. O mosquito tem hábitos diurnos, vive perto do ser humano, se alimenta de sangue e o vetor se prolifera em ambientes com água limpa e parada. Segundo Bio-Manguinhos (2014) a infecção por dengue pode ter vários níveis de sintomas, inaparentes, quando há ausência dos sintomas leves ou graves. Os sintomas mais leves incluem febre alta, mal-estar, dor de cabeça, dor nos olhos, falta de apetite e manchas vermelhas. Nos casos mais graves a dengue pode levar ao óbito.

Devido a proliferação do vetor ser em ambientes com água limpa e parada, características climáticas tornam países tropicais mais suscetíveis a surtos de contaminação. Segundo Barreto e Teixeira (2008), a doença se distribui em uma faixa de 35° ao norte e 35° ao sul do Equador, em países das Américas Central e do Sul se destacam no cenário mundial, contribuindo com mais da metade dos casos registrados. No Brasil, há registro de casos de dengue na metade do século XIX, ocorridos em São Paulo e Rio de Janeiro. Mas os primeiros

casos de epidemias registrados datam do início do século XX na cidade de São Paulo. Nas décadas seguintes os surtos da doença se espalharam por todo o território nacional.

De acordo com o Ministério da Saúde (2002), a dengue é a principal *arbovirose*, doença transmitida por insetos, que afeta o ser humano e tem gerado um sério problema de saúde pública no mundo. Devido a peculiaridade da transmissão o controle da doença só é possível com a ausência do vetor, por meio de métodos que reduzam ou eliminem as áreas suscetíveis a focos de proliferação, regiões com água estagnada. Um desses métodos de combate é o manejo ambiental, que se trata de um rearranjo do ambiente para minimizar ou eliminar a propagação do vetor como participação comunitária na redução de criadouros e melhoria de saneamento básico (MINISTÉRIO DA SAÚDE, 2002).

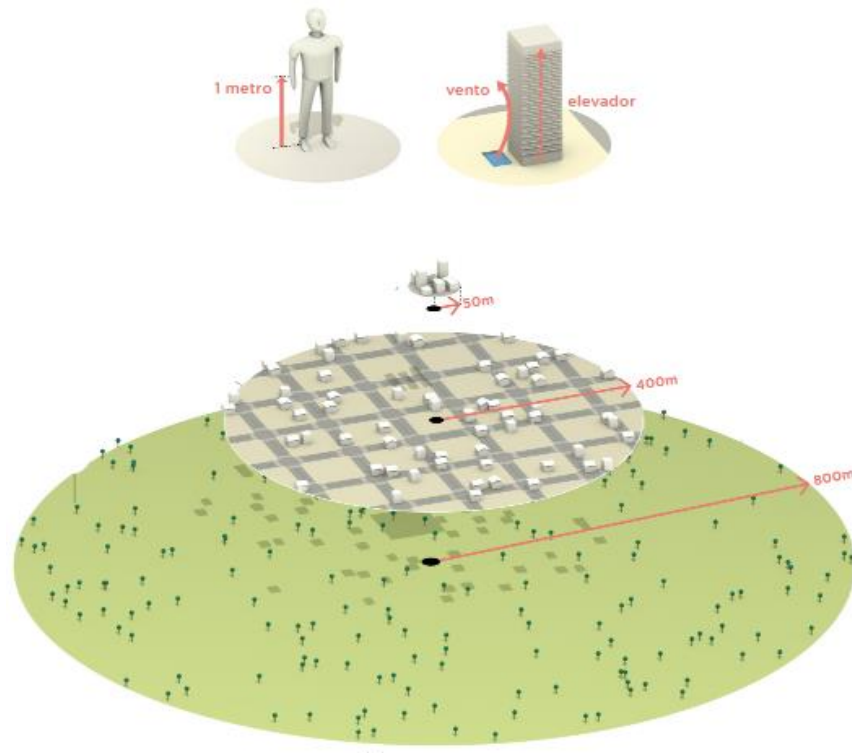
3.1.2 *Aedes aegypti*

O vetor de transmissão da dengue é o mosquito *Aedes Aegypti*, ele também é responsável por transmitir outras doenças como a *Zika* e *Chikungunya*. O mosquito tem hábitos diurnos, se alimenta principalmente de substâncias açucaradas, como a seiva das árvores e o néctar das flores. Segundo Christophers (1960), a fêmea do mosquito é a única que consome sangue, uma vez que para o desenvolvimento dos ovos são necessárias substâncias presentes no sangue. Desenvolvidos, os ovos são depositados pela fêmea próximo a superfície de água parada com baixa concentração de matéria orgânica. O tempo estimado entre o consumo de sangue pela fêmea, e a consequente contaminação do ser vivo pelo vírus da dengue, até o depósito dos ovos é de 3 dias, e até o mosquito chegar no período de vida tido como adulto é de mais 10 dias, com uma longevidade de aproximadamente 45 dias.

A média da altitude de voo do mosquito é de um metro, a baixa altitude é devido a atração do mesmo por gás carbônico, que se concentra mais próximo ao chão. O que não impede que atinja andares mais elevado de um edifício, pois pode ser levado pelo vento ou de "carona" em um elevador.

O alcance do mosquito está relacionado à densidade populacional. Em grandes aglomerados o raio de ação é menor, uma vez que a busca por sangue é mais fácil. Em ambientes descampados, abertos e com baixo aglomerado de gente, o mosquito tem um raio de ação maior, já que o acesso ao sangue é limitado e não há obstáculos, como árvores e prédios, para restringi-lo, como mostra a Figura 1.

Figura 1 - Ciclo de vida do *Aedes aegypti* e sua área de ação



Fonte: Christophers (1960).

3.2 SANEAMENTOS BÁSICO

Segundo a Organização Mundial da Saúde (2018), o saneamento é essencial para a saúde, desde a prevenção da infecção até a melhoria e manutenção do bem-estar mental e social. A falta de sistemas de saneamento leva à infecções e doenças, como o caso da cólera, que atingiu países da América do Sul como Brasil, Bolívia, Peru, entre outros locais, em meados da década de 1990.

O *Vibrio Cholerae*, por exemplo, é transmitido tanto por ingestão de água contaminada quanto por contato. O surto da doença ocorreu em lugares com bom sistema de saneamento e em regiões com sistemas ruins ou até ausentes. Em regiões com sistema de saneamento eficiente, coleta de esgoto e tratamento de água, a bactéria era excretada nas fezes, levada a estação de tratamento onde era eliminada e não contaminava outras pessoas, restringindo a transmissão eficiente nessas regiões apenas ao contanto, o que provocou uma menor quantidade de infectados, e não gerou agravos nos sintomas da doença. Nos locais com déficit de saneamento, a bactéria era excretada nas fezes que iam para a fossa séptica e contaminava a

água que depois seria ingerida pelas pessoas, o que provocou aumento no contágio e agravos nos sintomas das pessoas já contaminadas.

Devido ao impacto gerado pelo saneamento básico na qualidade de vida, na saúde, bem-estar social e na educação, esse conjunto de medidas envolve a atuação de vários agentes em uma ampla rede institucional. No Brasil o investimento nessa área ocorreu de forma pontual ao longo do século XX, com destaque para as décadas de 1980 e 1990, o que promoveu uma grande desigualdade e déficit ao acesso (LEONETI; PRADO; OLIVEIRA, 2011).

Mesmo com os surtos de epidemias de dengue, chycunguya e Zika vírus, que tem como vetor o mosquito que se prolifera em ambientes de água parada, os indicadores de saneamento para o Brasil são alarmantes. Os dados disponibilizados pelo Ministério da Saúde no Sistema Nacional de Informações sobre Saneamento (SNIS) que tem como base o ano de 2014, ainda há mais de trinta e cinco milhões de brasileiros sem saneamento básico (INSTITUTO TRATA BRASIL, 2016). O estudo feito sobre os investimentos de saneamento básico no país mostrou que o dinheiro público destinado ao saneamento básico está nas 100 maiores cidades do Brasil e que 90% do esgoto não tratado fica em regiões com moradores de baixa renda, e vivendo em moradias irregulares, o que segundo Saiani (2007), mostra que a distribuição de investimento nessa área está intimamente ligada ao perfil de renda dos consumidores. Não só o avanço em saneamento básico é muito lento no país, como está cada vez mais concentrado nas regiões que já possuem mais que o mínimo.

O déficit de saneamento básico relacionado a epidemias é um elo fraco na corrente de combate à cadeia de transmissão de doenças epidemiológicas e uma barreira na eliminação dos locais de focos de reprodução do vetor da dengue. A localização desses focos e o contexto espacial no qual estão inseridos têm servido de base para estudos de espacialização da doença. A gravidade da Dengue como problema de saúde pública em países em desenvolvimento e o fato da doença ser causada por agente transmissor que se adapta melhor a um determinado ambiente fez pesquisadores analisarem sua distribuição (ARAÚJO et al., 2008).

No caso da dengue a distribuição desigual da doença, gera lacunas na compreensão do fenômeno dificultando a busca por soluções. A explicação tem sido estudada mediante a análise da distribuição espacial da doença associada a dados especializados de determinantes socioeconômicos.

3.3 BANCO DE DADOS GEOGRÁFICO APLICADOS À SAÚDE

Todas as informações, sejam elas geográficas ou não, devem ser armazenadas de forma lógica e ordenada, de modo que expresse algum significado, e não uma coleção aleatória sem um objetivo específico. Um banco de dados é construído e preenchido com dados que têm um determinado objetivo. Trata-se de uma representação do mundo real.

Um banco de dados geográfico é o componente de armazenamento de um SIG. É um banco de informações totalmente preparado para o armazenamento de dados espaciais. Sua função é estruturar e armazenar os dados de forma a possibilitar a realização das operações de análise e consulta usando não só dados alfa numéricos como também geometrias. (LISBOA FILHO; IOCHPE, 1999).

Um modelo de dados é um conjunto de conceitos para descrever as operações e estrutura em um banco de dados. O objetivo do modelo é sistematizar o entendimento que é desenvolvido a respeito de fenômenos e objetos que serão sistematizados. No entanto, esses objetos e fenômenos reais, são complexos e não permitem uma representação completa, de forma que é necessário abstrair os objetos e fenômenos do mundo real, criando representações simplificadas, mas convenientes, que se adequem a finalidade da aplicação (CÂMARA et al., 1996).

No universo de aplicações geográficas são considerados quatro níveis de abstração, nível do mundo real, nível de representação conceitual, nível de apresentação e nível de implementação. No nível do mundo real, encontramos o fenômeno que será discriminado, tal como ele é. O nível conceitual oferece conceitos formais com os quais as entidades geográficas podem ser modeladas da forma como são percebidas pelo o usuário em um alto nível de abstração (BORGES; DAVIS JÚNIOR; LAENDER, 2005).

O nível de apresentação em um banco de dados geográfico oferece ferramentas com as quais é possível especificar os diversos aspectos visuais que as entidades assumem ao longo do seu uso em aplicações. No nível de implementação, se definem padrões e formas de armazenar estruturas de dados para implementar cada tipo de representação, bem como os relacionamentos entre elas e as funções e métodos necessários (BORGES; DAVIS JÚNIOR; LAENDER, 2005).

Os modelos de dados usados em SIG podem ser modelos semânticos e orientados a objetos, como o modelo Entidade Relacionamento, OMT (*Object Modeling Technique*) e UML (*Unified Modeling Language*), que procuram refletir melhor a necessidade de aplicações geográficas. O modelo adotado deve conter os seguintes componentes para gerar as análises, base de dados, processamento e resultados.

A Base de dados compreende a definição dos dados necessários para a análise, bem como a identificação das propriedades cartográficas do modelo. O processamento destina-se a montagem e a preparação da base de dados, construção ou adequação da tabela de atributos e especificações dos geocódigos, para as operações de análise que se destina a atingir o objetivo do estudo. Os resultados podem ser parciais, as componentes geradas ao fim das etapas, e final.

A partir do modelo gerado é possível efetuar análise espacial, mensurando propriedade e relacionamentos, considerando a localização espacial do fenômeno. O setor de saúde já há algum tempo vem se beneficiando do uso de banco de dados geográficos com investigações em epidemiologia, dando ênfase à doença da população e não do indivíduo, estudos ecológicos, detecção de aglomerados espaciais e aplicados ao planejamento de serviços de saúde.

Os bancos de dados da saúde abrangem desde informações vitais até gerenciais. Bases como Sistema de Informações de Mortalidade (SIM), Sistema de Informação Sobre Procedimentos Ambulatoriais (SIA), Sistema de Informações de Nascidos Vivos (SINASC) entre outros contam com séries históricas que servem de fonte para infinitas análises. O Sistema Único de saúde (SUS) possui bancos de dados de cobertura nacional onde as informações advêm da gestão municipal que devem captar informações pertinentes a doenças e tratamentos e também localização geográfica

Alguns autores desenvolveram sistemas de informação utilizando banco de dados geográficos em saúde pública com o intuito de oferecer um ambiente computacional gratuito, totalmente possível de ser elaborado e implementado como o trabalho de Cavicchioli Neto et al. (2008), que desenvolveram um banco de dados geográficos para o gerenciamento e vigilância em saúde com a intenção de melhorar a qualidade da informação fornecida por meio da modernização do o sistema de vigilância em saúde. Furlan et al. (2009), desenvolveram um sistema de gerenciamento da informação sobre a Mortalidade Infantil (MI) na região metropolitana da Baixada Santista. O sistema auxilia no monitoramento e mortalidade infantil e integra dados de óbitos e nascidos vivos e fornece relatórios com tabelas, gráficos e mapas.

3.4 ESTATÍSTICA ESPACIAL

Estatística espacial é o ramo da estatística que possibilita analisar a localização espacial de eventos. Permite a identificação, localização e visualização da ocorrência de fenômenos que se materializa ao longo do espaço. Os SIGs comumente utilizam a estatística espacial,

possibilitando modelar a ocorrência dos fenômenos, identificando padrões, incorporando fenômenos entre outros (ANDRADE ET AL., 2007).

3.4.1 Análise espacial aplicada em estudos de saúde

Para Barcellos et al. (2002, p. 130), “Se a doença é uma manifestação do indivíduo, a situação de saúde é uma manifestação do lugar. ” A produção de condições de doenças está ligada as condições dos lugares, como seu contexto histórico, ambiental e social que são expressas nos indicadores sociais. Identificar e refletir de forma correta as condições de risco à saúde que esses indicadores promovem configura um ponto importante para o diagnóstico de situações de saúde. Analisar as situações de saúde corresponde a analisar o lugar e a dinâmica espacial dos fatores socioambientais em relação à saúde.

Análise espacial pode ser compreendida como a análise de fenômenos posicionados no espaço. O interesse não se restringe ao dado, mas o contexto espacial no qual está inserido buscando descrever ou explicar os processos que ocorrem nesse espaço bem como sua relação com outros fenômenos. Serve para incorporar o espaço à análise que se deseja fazer, mensurando propriedades e relacionamentos que levam em conta a localidade do fenômeno em estudo. O problema faz parte da análise espacial de dados geográficos (DAVIS; CÂMARA, 2001).

Análise espacial, se configura no princípio da descoberta de informações, mas levando em consideração a localização espacial do fenômeno em estudo. Conforme Leite (2010) a necessidade de analisar e gerenciar o espaço de forma eficaz gera a necessidade de integração de dados diversos e tecnologias de informação geográfica ou geotecnologias que tem subsidiado a concatenação e análise dessas informações. As tarefas envolvendo geotecnologias são executadas por SIG que são sistemas destinados ao processamento de dados georreferenciados, e que os tornou um instrumento útil, por exemplo, na vigilância epidemiológica.

Os tipos de análise espacial podem ser de diferentes níveis de complexidade como mostrado no Quadro 1, e composta por um conjunto de procedimentos encadeados com a finalidade de escolher um modelo que considere o relacionamento espacial presente no fenômeno.

Quadro 1 - Técnicas de análise espacial em SIG para aplicações em saúde e ambiente

TÉCNICA	DESCRIÇÃO	EXEMPLO DE APLICAÇÃO
Pontos num polígono	Identificar a interseção entre pontos e a área (polígono) em que eles estão	Para identificar todos os casos dentro de uma zona de exposição especificada

Linhas num polígono	Identificar a interseção entre linhas e a área (polígono) que elas cruzam	Para identificar fontes lineares (ex. estradas) que cruzam uma área especificada
Áreas de influência (buffer)	Construção de zonas de largura especificada ao redor de pontos, linhas ou áreas	Para definir áreas de exposição em torno de fontes de risco (ex. usinas nucleares)
Interpolação	Estimação de condições em locais não amostrados	Mapeamento de superfícies de poluição
Estimação de proximidade	Análise de condições em determinado ponto, baseada em condições de uma vizinhança especificada	Estimativa de níveis de poluição baseada no uso do solo da região em torno
Alisamento (Smoothing)	Construção de uma superfície alisada (generalizada)	Mapeamento de superfícies generalizadas de exposição
Sobreposição (overlay)	Combinação de um mapa com outro por sobreposição	Combinação entre mapas de densidade de poluição e população para identificar populações expostas

Fonte: adaptado de Briggs & Elliott (1995).

Os modelos espaciais são classificados em três grupos: variação discreta, contínua e pontual e são definidos segundo o problema da análise espacial como mostra o Quadro 2. Os fenômenos expressos através de ocorrências identificados como pontos no espaço, como os dados notificados de dengue, são usados em análises com o objetivo de testar hipóteses sobre os padrões da distribuição espacial.

Quadro 2 - Tipos de Dados e Problemas em Análise Espacial

	Tipos de Dados	Exemplo	Problemas Típicos
Análise de Padrões Pontuais	Eventos Localizados	Ocorrência de Doenças	Determinação de Padrões e Agregamentos
Análise de Superfícies	Amostras de Campo e Matrizes	Depósitos Minerais	Interpolação e Medidas de Incerteza
Análise de Áreas	Polígonos e Atributos	Dados Censitários	Regressão e Distribuições Conjuntas

Fonte: Davis e Câmara, (2001)

Conhecer a estrutura e a dinâmica espacial da população é o primeiro passo para identificação de situações de saúde. Todos os eventos relacionados à saúde ocorrem em indivíduos, desde o nascimento, passando por infecções até a morte, no contexto de análise de processos pontuais. Esses pontos não têm uma distribuição regular no espaço e podem ser

descritos em termos de efeito de primeira ordem, que podem ser considerados globais ou de larga escala e variação de valor médio do processo no espaço, e segunda ordem, também considerados efeitos locais ou de pequena escala, representam a dependência espacial no processo, proveniente da estrutura de correlação espacial (JÁCOMO et al, 2010). Nesse caso, ao se trabalhar com registros de saúde, é necessário avaliar a probabilidade de um evento ocorrer, sendo ponderado pela distribuição da população (BARCELLOS; RAMALHO, 2002).

Essa distribuição pode ser considerada à partir da agregação de dados demográficos e de saúde em unidades espaciais discretas, no entanto, essa estratégia apresenta grandes limitações como a desconsideração das interações entre as unidades espaciais e as instabilidades dos indicadores criados em pequenas áreas (KING, 1979).

A distribuição da população pode ser analisada estatisticamente por meio de modelos teóricos de referência, como a aleatoriedade espacial completa, que divide a região de estudo em subáreas e avalia a distribuição de eventos pontuais como um processo aleatório (CAMARA et al, 2002).

Outra forma de analisar o comportamento dos padrões de distribuição da população é pelo método de densidade de casos (número de casos por área), um modelo estatístico de efeito global, que produz uma superfície de probabilidade, onde a proximidade entre os casos apresenta maior risco.

A estatística espacial é aplicada nos estudos de saúde para modelar a realidade. Os modelos estatísticos usados para estudar saúde e os fatores relacionados a ela são úteis na análise espacial para entendimento, estimativa e modelagem de como o contexto derivado de fatores socioeconômicos afeta a saúde das pessoas. Modelos estatísticos são especificados na modelagem de dados para tentar estimar parâmetros. Esses modelos se preocupam com fenômenos estocásticos, ou seja, sujeitos a incertezas.

Dados que possuem alguma referência espacial como, por exemplo, os dados censitários que estão relacionados ao endereço da moradia do entrevistado, podem ser analisados sob a óptica da estatística clássica, resultando em massas de dados numéricos. Mas sob a óptica da estatística espacial, que usa explicitamente referências geográficas no modelo, os resultados são diferentes, geralmente mais robustos por incorporar a dimensão espacial. Esse ramo da estatística possibilita modelar a ocorrência de fenômenos incorporando fatores determinantes à estrutura de distribuição espacial.

Com o propósito de inserir o espaço como uma categoria totalmente importante na análise, alguns trabalhos têm abordado a análise espacial aplicado a epidemiologias, doenças

que atingem uma parcela significativa da população de uma região e que acarreta grandes transtornos como a cólera, surtos de gripes e dengue.

A dengue é considerada um grande desafio à saúde pública, já que existem várias questões relacionadas à doença que precisam ser esclarecidas, que vão desde a determinação dos locais de proliferação do vetor da doença, o mosquito *Aedes Aegypti*, até a determinação das políticas de combate à cadeia de transmissão. Algumas explicações sobre a dengue, bem como de outras doenças, têm sido dadas à partir de estudos que avaliaram a relação entre a doença e determinantes sociais.

A compreensão correta dos padrões de distribuição espacial da localização da residência de casos de dengue e suas relações com fatores econômicos e socioambientais, possibilita explorar o potencial da abordagem ecológica em saúde pública, através da aplicação de ferramentas de análise apropriada.

Em Resendes et al. (2010) os autores usaram análise por conglomerados para identificação de áreas de risco da dengue e foram escolhidos 13 indicadores sociais construídos com base nas informações do censo de 2000, para análise da correlação, dos quais dois foram removidos por apresentarem alta correlação com outras variáveis. A escolha das variáveis teve como critério ser um macro determinante da dengue e os resultados mostraram que menores condições de infraestrutura de serviços de saneamento e alto adensamento populacional são os indicadores macro determinantes que mais contribuem para a proliferação do mosquito e por consequência, propagação da doença.

No trabalho de Barbosa e Silva (2016), foram escolhidos 16 indicadores socioeconômicos. A análise escolhida para testar a correlação entre os casos e os indicadores, foi a bivariada pelo índice de Moran local, que necessita de uma estimativa magnitude da autocorrelação. Foram selecionadas para análise apenas as variáveis que apresentaram correlação superior a 0,20. Como resultado, a variável com a maior correlação foi o percentual de responsável pela família sem instrução ou com apenas o nível fundamental de instrução, que mostra uma alta relação entre baixa escolaridade e a doença.

No seu estudo, Leite (2010) analisou a possibilidade de relação entre a dengue os indicadores sociais: renda, saneamento básico, densidade populacional e escolaridade. E os resultados apontaram que a maioria dos casos ocorreu em locais de residência de moradores de baixa renda, baixo nível de escolaridade, baixo índice de saneamento básico e alta densidade demográfica, demonstrando uma realidade diferente, para aquela localidade e na época do estudo, da disseminada pelo senso comum de que a dengue é uma doença demográfica.

3.4.2 Análise Exploratória de Dados Espaciais (AEDE)

Os recentes avanços em computação em relação à capacidade de armazenamento e processamento, bem como o aprimoramento dos SIGs, juntamente com o emprego de técnicas de estatística adaptada aos dados espaciais, possibilitaram utilização de técnicas mais elaboradas de análise espacial.

A análise de dados espaciais pode ser empregada quando os dados estiverem espacialmente localizados e quando o arranjo espacial necessitar ser levado em consideração. A AEDE é composta por um conjunto de técnicas que descreve a distribuição espacial, identificando padrões (*clusters*) e áreas atípicas (*outliers*) (Anselin & Bao, 1997).

A AEDE utiliza dados georreferenciados e com frequência é empregada para verificar a existência de padrões espaciais, dependência espacial ou heterogeneidade espacial que possa indicar semelhança entre regiões. Além de considerar a distribuição e os relacionamentos espaciais, a AEDE é crucial em estudos de processos de difusão espacial porque indica padrões de autocorrelação espacial (GOODCHILD *et al.*, 2000).

3.4.3 Dependência Espacial

É uma característica dependente à representação de dados por meio de subdivisões territoriais. Pode ser entendida como a tendência de que um valor de uma região específica se assemelhe mais com o valor das áreas vizinhas do que de áreas mais distantes. A definição de dependência espacial é importante para a compreensão e também análise dos fenômenos espaciais e é baseada na afirmação de que todos os objetos que estão no espaço estão relacionados, no entanto, objetos mais próximos são mais semelhantes, uma referência a primeira lei da geografia (Cressie, 1992).

3.4.4 Análise de variáveis Espaciais de Áreas

Quando se trabalha com eventos agregados por setores censitários, bairros, zonas, municípios ou que estejam delimitados por polígonos, onde o evento não tenha um lugar correto, é frequente o uso de métodos de análise de dados espaciais de área, por procurar padrões de distribuição espacial, sinais de dependência espacial e aglomerados.

No entanto, para o uso dessa análise é necessário o conhecimento de alguns problemas possíveis, como a descontinuidade de fronteiras, como o valor atribuído não é pontual, mas sim contíguo para uma área, valores próximos à fronteira tendem a ser parecidos, o que pode acabar distorcendo a análise nessas regiões de fronteira (Miller, 1999). Outro problema é o de Unidade de área Modificável (*Modifiable Areal Unit Problem - MAUP*)

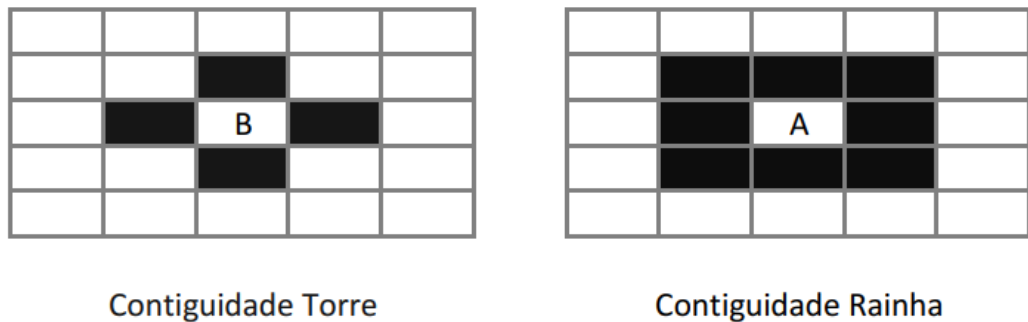
3.5 ANÁLISE INFERENCIAL

A estatística inferencial está interessada em usar as informações relativas à uma amostra para concluir sobre um grupo maior, que por alguma questão, não se tem acesso. À análise inferencial tem o objetivo de edificar modelos que possam representar da melhor forma a distribuição espacial dos dados de área com o propósito de gerar resultados aplicáveis ao estudo em questão.

3.5.1 Matriz de proximidade espacial

Para comparar a similaridade entre atributos espaciais em feições vizinhas é crucial estabelecer um critério de vizinhança. A análise da variabilidade espacial de um evento de dados de área necessita primeiro da construção de uma matriz de proximidade espacial. As relações existentes entre as áreas são explicadas por meio de pesos espaciais contido em uma matriz no qual cada elemento W_{ij} mede a proximidade ou a vizinhança entre i e j , onde i representa a linha e j a coluna.

Um critério de vizinhança que costuma ser adotado implica a contiguidade $W_{ij}=1$ quando as regiões i e j são contíguas ou vizinhas e indica o contrário quando $W_{ij}=0$ (ALMEIDA, 2012). A matriz de vizinhança é feita segundo critérios, podendo ter a contiguidade Rock (torre) e Queen (rainha) como mostra a Figura 2, a ordem de contiguidade também é relevante, podendo ser de primeira ordem, que leva em consideração apenas os vizinhos mais próximos, segunda ordem, que considera os vizinhos dos vizinhos mais próximos e assim sucessivamente.

Figura 2 - Convenção de contiguidade Torre e Rainha

Fonte: Almeida (2012)

3.5.2 Autocorrelação

A autocorrelação pode ser definida como uma correlação positiva ou negativa entre a distância entre elementos e a similaridade nos valores medidos de uma variável em uma intensidade superior à esperada pelo acaso (LEGENDRE, 1993).

O índice de Moran calcula a autocorrelação espacial que descreve o padrão de distribuição dos dados, evidenciando a magnitude da correlação. Esse índice varia de -1 a 1, onde valores negativos expressão autocorrelação espacial negativa e valores positivos expressão autocorrelação positiva. O valor zero indica que não há padrão espacial nos dados.

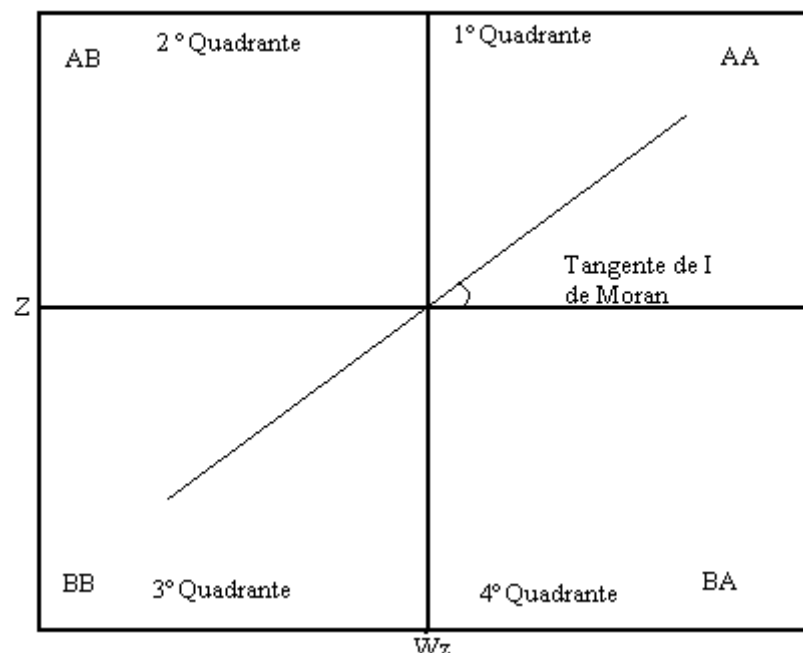
O índice de Moran global é um indicador global de autocorrelação, o que significa que ele fornece um valor único para todo o conjunto de dados, que pode ser interessante na caracterização da área como um todo, mas é possível que existam regiões que possuem associações espaciais distintas, isso ocorre quando se trabalha com várias áreas. É possível identificar diferenças espaciais através das análises locais, que desagregam a estatística do todo, de acordo com constituintes locais.

Em alguns casos é interessante descobrir se os valores de uma variável em uma determinada área possuem alguma relação com outra variável em áreas do entorno, calculando assim o grau de correlação entre duas variáveis (ALMEIDA, 2012). Nos casos em que se queira calcular a correlação de mais de duas variáveis em áreas vizinhas, uma análise multivariada, é necessário usar a matriz de diagrama de dispersão de Moran, sendo que o objetivo dessa matriz é analisar o padrão espacial de cada variável com outras variáveis ou com ela mesma.

Esse diagrama é dividido em quatro tipos de associação espacial que corresponde padrões de associação espacial local (Almeida, 2004), sendo que no primeiro quadrante (AA) trata de agrupamento de valores acima da média para o atributo em análise, valores *High-High* (Alto-Alto), que tem como vizinhos valores também acima da média, apresentando

autocorrelação positiva. O terceiro quadrante (BB) está relacionado ao agrupamento de valores que ficaram abaixo da média para o atributo em análise, valores *Low-Low* (Baixo-Baixo), que tem como vizinhos valores que também ficaram abaixo da média, apresentando autocorrelação negativa. O segundo (AB) e o quarto quadrante (BA) mostra agrupamento de valores acima da média vizinhos de valores abaixo da média, valores *High-Low* (Alto-Baixo), e agrupamento de valores abaixo da média cercado por vizinhos acima da média, valores *Low-High* (Baixo-Alto), respectivamente como mostra a Figura 3, indicando que as variáveis não dependem linearmente uma da outra.

Figura 3 - Diagrama de dispersão de Moran



Fonte: Almeida (2004)

A existência de padrões globais de autocorrelação espacial pode estar em concordância com algum padrão local. Pode ocorrer de não haver correlação espacial global, mas haver algum *cluster*. Em nível local. Estudando os padrões locais para autocorrelação é possível obter informações sobre a presença de heterogeneidade espacial, que é a tendência de lugares serem diferentes uns dos outros.

Assim como é possível calcular a autocorrelação espacial global multivariada, também é possível calcular a correlação espacial local multivariada. O índice de Moran local multivariada é um indicativo do grau de associação linear, seja ela positiva ou negativa, para o valor de uma variável em uma determinada localização e o valor de outra variável nas localizações vizinhas (ANSELIN, 2004).

3.5.3 Representação Gráfica: *BOX MAP*, *LISA Map* e *Moran Map*

Box Map, *LISA Map* e *Moran Map* são dispositivos gráficos de visualização fundamentados nos resultados obtidos para o gráfico de espalhamento de Moran e os indicadores locais. No *Box Map* os objetos são classificados conforme sua posição em relação ao gráfico de espalhamento. O *LISA Map* avalia a significância dos valores do índice de Moran local em relação a hipótese nula de não haver correlação. No *Moran Map* é destacado os objetos cujo valores *LISA* são considerados significantes e são classificados em quatro grupo conforme o diagrama de espalhamento de Moran (ANSELIN, 2004).

3.5.4 Teste de Significância

Um valor de índice de Moran positivo próximo de 0 não necessariamente significa ausência de correlação. É possível, por meio de um teste de significância, identificar a existência de autocorrelação espacial. A execução do teste é por meio de permutações, onde os valores da variável analisada são realocados aleatoriamente nas diversas áreas em estudo e a estatística é computada novamente.

3.5.5 Regressão espacial

A regressão é uma ferramenta estatística que usa a relação entre variáveis de tal forma que uma variável possa ser explicada por outra e sua aplicação ocorre quando a correlação entre as variáveis seja significativa. Existem diferentes formas para inserir efeitos espaciais em modelos de regressão, uma delas é comumente conhecida como efeitos espaciais globais, e tem o objetivo de capturar a estrutura da correlação espacial em um parâmetro podendo assim adicioná-lo no modelo de regressão clássica. Duas alternativas são possíveis, o modelo *SAR – Spatial LAG Model*, e o modelo *CAR– Spatial Error Model* (BAILEY e GATRELL, 1995).

O modelo de Regressão *SAR* é explicado por Câmara et al. (2005) como a dependência espacial no meio da adição ao modelo de regressão de um novo termo no formato de uma relação espacial para a variável dependente. O modelo de Regressão *CAR* leva em consideração que os efeitos espaciais são uma perturbação, ou ruído, o que faz com que esse erro necessite ser removido. Sendo assim, os efeitos da autocorrelação espacial são associados ao termo de erro ϵ .

3.5.6 Regressão geograficamente ponderada

Quando existe autocorrelação espacial da variável dependente em estudo é necessária a incorporação do atributo espacial na regressão, os modelos de regressão linear que costumam ser utilizados, sem a incorporação do atributo geográfico, podem conduzir a uma superestimação da significância dos modelo, e para se evitar este tipo de problema na modelagem, a inserção do caráter espacial na calibração do modelo de regressão linear pode ser realizada com a técnica de ponderação geográfica (LOUREIRO; SILVA; CARVALHO, 2006).

A regressão geograficamente pondera (RGP) é uma técnica utilizada para modelar processos que variam no espaço, ou seja, heterogêneos espacialmente. O objetivo consiste em ajustar um modelo de regressão para cada uma das áreas em análise utilizando a localização espacial das outras áreas para determinar as estimativas dos parâmetros.

3.6 DESCOBERTA DE CONHECIMENTO EM BANCO DE DADOS

A informática e as tecnologias têm mudado o modo de vida humana. A quantidade de dados gerados e disponibilizados tem crescido exponencialmente. Os avanços na área de tecnologia da informação viabilizam o armazenamento de várias bases de dados. No entanto o volume de dados criados diariamente pela web, redes sociais, notícias, sensoriamento remoto, comércio eletrônico, câmeras digitais, mercado de ações têm gerado a necessidade de manusear esses dados de forma eficiente. Dados são registros soltos e aleatórios, eles constituem a matéria prima da informação, que é a estruturação ou organização desses dados, e essa é a matéria de que é feita o conhecimento.

A quantidade de dados de doenças, cuidados e procedimentos médicos, assistência à saúde, e monitoramento epidemiológicos gerados todos os dias torna impossível a manipulação e descoberta de informações mais consistentes do que os obtidos por meio de um Sistemas Gerenciadores de Banco de Dados (SGBD). O volume de dados demanda organização com técnicas computacionais que auxiliem a análise, interpretação e no relacionamento dessas informações, porque o valor de uma decisão estratégica depende fundamentalmente das informações disponíveis ao tomador de decisão.

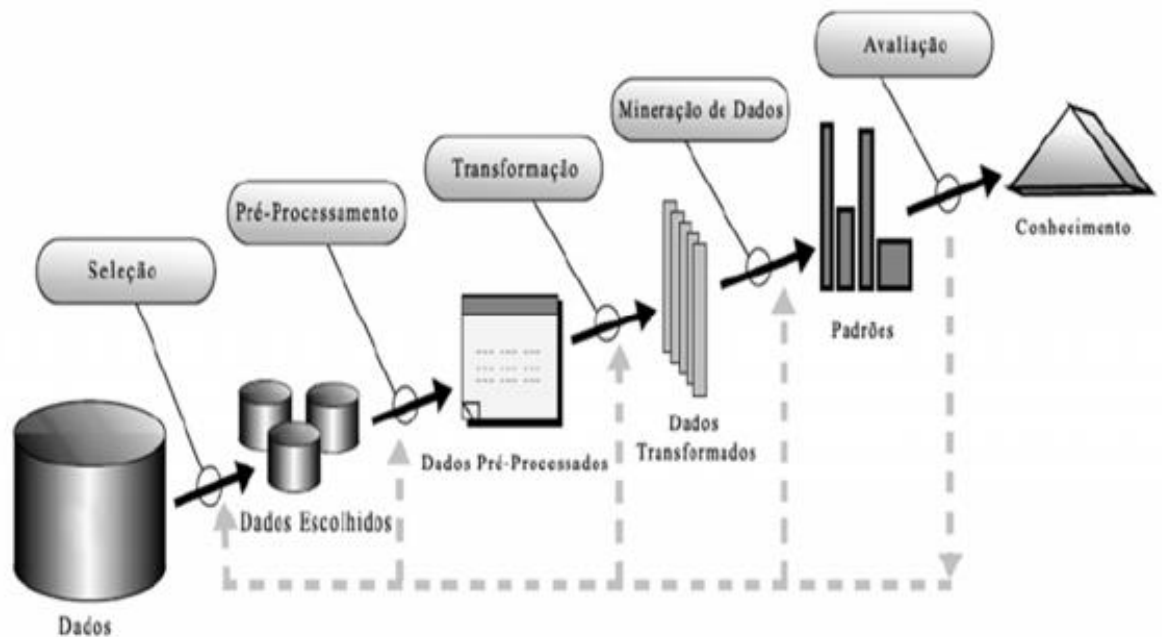
A análise de grandes volumes de dados pelo o homem é inacessível sem o auxílio de ferramentas computacionais apropriadas (GOLDSCHMIDT; PASSOS, 2005). Um processo

que aperfeiçoa a gestão dos dados é a Mineração de Dados, o qual encontra anomalias, extrai padrões e correlaciona conjuntos de dados para prever resultados, ou seja, analisar um conjunto de dados sobre diferentes perspectivas para gerar informação. Isso faz dessa ferramenta a principal etapa da metodologia para descoberta de conhecimento em banco de dados (*Knowledge Discovery in Databases* - KDD) (SANTOS; NETO, 2011).

KDD é caracterizada como um processo composto por várias etapas operacionais. A Figura 4 mostra as etapas de um processo em KDD, sendo que as etapas anteriores à mineração de dados fazem parte do pré-processamento dos dados que compreendem a captação, organização, tratamento e preparação de todos os dados que serão submetidos à mineração de dados. A etapa subsequente à mineração compreende o pós-processamento e envolve a visualização, análise e interpretação do modelo de conhecimento gerado.

A extração do conhecimento é uma área dinâmica e evolutiva, envolvendo integrações com outras áreas de conhecimento como Estatística, Inteligência Artificial e Banco de Dados. Um dos objetivos da descoberta de conhecimento em bancos de dados é tornar padrões compreensíveis para humanos, visando promover uma melhor compreensão dos próprios dados. Embora seja um tanto subjetivo medir compreensibilidade, um dos fatores frequentes é a medida de simplicidade. O fator de compreensão dos dados está relacionado à intuitividade da representação destes, bem como da granularidade alta o suficiente para que estes sejam compreendidos. É frequente os dados do mundo real não serem representados de forma completa, e é necessário que na etapa de pré-processamento dos dados, a limpeza dos dados esteja envolvida.

Figura 4 - Etapas do processo de KDD definidas por Fayyad



Fonte: Camilo e Silva (2009) apud Fayyad (1996)

3.7 NORMALIZAÇÃO

Normalização consiste no ajuste da escala de valores de cada atributo de um banco de dados, de forma que os valores fiquem em pequenos intervalos como de 0 a 1 ou -1 a 1, minimizando problemas de unidades de dados diferentes ou dispersões muito heterogêneas. Esse ajuste é necessário para evitar tendências em métodos de mineração de dados para atributos com escala de valores maior que a dos outros (GOLDSCHMIDT; PASSOS, 2005).

Um dos métodos de normalização consiste na normalização linear que considera os valores de máximo e mínimo de cada atributo e mapeá-los para um intervalo de 0 e 1. A normalização por desvio-padrão, considera a posição média dos valores de um atributo e a dispersão dos demais valores em relação à média. A normalização pela soma dos elementos pode ser entendida como a divisão do valor do atributo pela soma de todos os valores do atributo. A normalização pelo valor máximo dos elementos consiste em dividir o valor de cada atributo que esteja sendo normalizado pelo valor máximo entre esse atributo. Normalização por escala decimal consiste em deslocar o separador decimal do atributo a ser normalizado (GOLDSCHMIDT; PASSOS, 2005).

Para os dados espaciais, a escala de investigação é a variável mais importante a ser considerada. Cada critério utilizado no cruzamento de dados deve ser mensurado considerando

suas próprias características para o dimensionamento, o que significa que pode haver dados de diferentes escalas. Para serem usados num processo de mineração de dados, minimizando a influência de erros no processo de inter-relacionamento e garantindo a operacionalização entre esses dados, é necessária a padronização dos dados em uma escala única. Assim, a multiescalaridade dos dados e informações deve ser avaliada e os dados em diferentes escalas operacionais deverem ser generalizado. O nível de alteração que cada base de informação deve sofrer para a criação de uma base única, vai depender da escala do dado de pior resolução.

3.8 MINERAÇÃO DE DADOS

A execução da etapa de mineração de dados compreende a aplicação de algoritmos de busca, que procuram, segundo determinados paradigmas, explorar os dados de forma a produzir modelos (padrões) de conhecimento. A forma de representação do conhecimento está ligada diretamente ao algoritmo de mineração utilizado (GOLDSCHMIDT; PASSOS, 2005).

A busca executada na mineração de dados se manifesta por meio de regras de associação, que Segundo Brusso (2000), são padrões descritivos que representam a probabilidade de que um conjunto de itens apareça em uma transação visto que outro conjunto está presente. A quantidade de regras formadas costuma ser volumosa, o que se faz necessário identificar quais delas são realmente interessantes.

Uma forma de identificar as regras de interesse é identificando o conjunto de itens frequentes e gerar regras que contenham suporte e confiança iguais ou maiores do que os valores determinados pelo usuário. O suporte (Equação 1) representa a probabilidade de uma transação, itens que foram consultados, satisfazer ao mesmo tempo dois conjuntos de itens. A confiança (Equação 2) é a probabilidade de que uma transação satisfaça um conjunto de dados baseada no fato dela satisfazer outro conjunto de itens (GREGORY; PRETTO, 2004).

O suporte de uma regra $X \Rightarrow Y$, onde X e Y são conjuntos de itens, e T é o número de transações em que ambos os conjuntos ocorrem simultaneamente, é dado pela seguinte fórmula:

$$\text{Suporte} = \frac{\text{Frequência de X e Y}}{\text{Total de T}} \quad (1)$$

A sua confiança é dada pela seguinte fórmula:

$$\text{Confiança} = \frac{\text{Frequência de X e Y}}{\text{Frequência de X}} \quad (2)$$

As regras de associação baseadas no modelo suporte/confiança não são usuais para dados volumosos, medidas de correlação são mais interessantes para esses casos. Métodos

baseados em redes neurais, regressão, previsão de séries temporais e clusterização podem estimar erro ou distância entre o resultado produzido e o esperado.

Métodos específicos trabalham com algoritmos desenvolvidos especificamente para implementar alguma tarefa de Mineração de Dados voltada à descoberta de conhecimento. O algoritmo *Apriori*, é um clássico em regras de associação. Ele emprega busca em profundidade, encontra todos os conjuntos de itens frequentes, que satisfazem à condição de suporte mínimo, e gera as regras de associação, que satisfaçam à condição de confiança mínima.

Para um processo adequado de mineração de dados é necessário que os dados sobre os quais se deseja extrair algum conhecimento estejam completos, sem ruídos e consistentes. A qualidade dos dados influencia a qualidade dos modelos de conhecimento a serem abstraídos a partir desses dados (GOLDSCHMIDT; PASSOS, 2005).

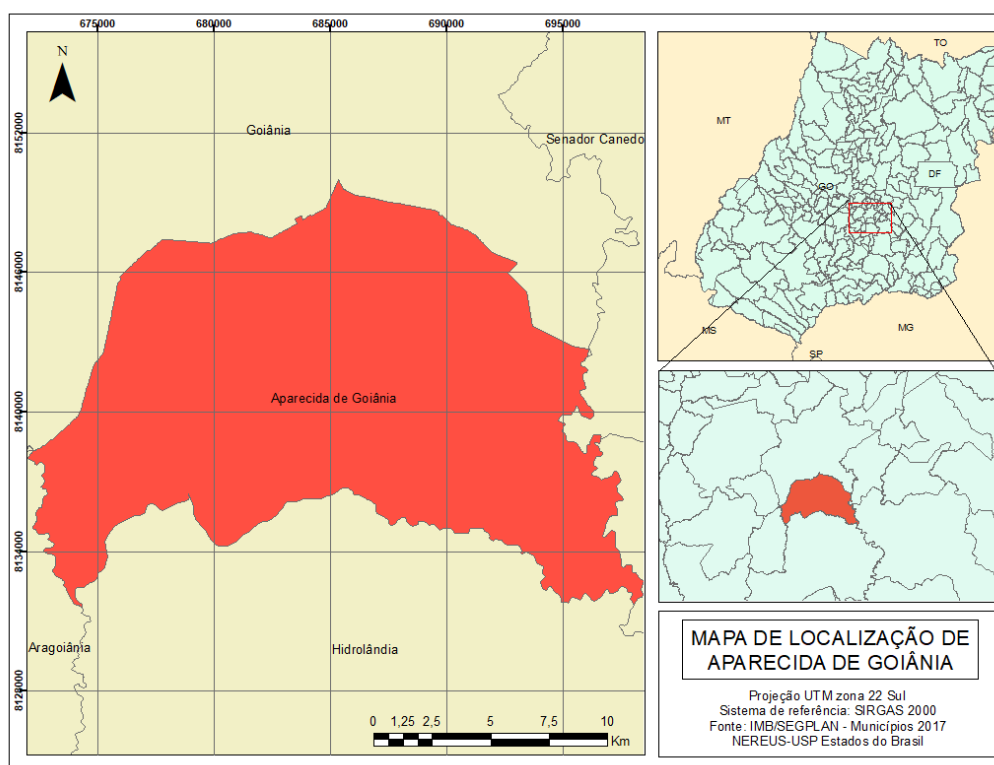
O uso de mineração de dados tem ingressado em bases de dados geográficos por ser capaz de captar a semântica agregada à componente espacial dos dados em análise, desempenhando um papel importante na identificação de relacionamentos existentes. Os algoritmos de mineração de dados extraem padrões de dados e relacionamento que representam conhecimento nas bases de dados (SANTOS et al, 1999).

O uso da mineração em bases de dados de saúde tem desempenhado um papel fundamental, uma vez que o volume de dados produzido por esse setor é considerável. Santos e Neto (2011) desenvolveram um aplicativo de mineração de dados para aplicação na base do Sistema de Informação de Agravos e Notificação – SINAN. O objetivo foi criar uma ferramenta que buscasse aplicar alguns dos algoritmos de mineração de dados gerando regras de associação ou de classificação relativas ao paciente, para auxiliar os gestores de saúde no diagnóstico de pacientes com suspeita de dengue. Os autores apresentaram os conceitos mais importantes da mineração de dados e KDD, bem como aplicação dessas ferramentas na saúde pública.

3.9 APARECIDA DE GOIÂNIA

A área escolhida para a realização do presente trabalho é a área urbana do Município de Aparecida de Goiânia, GO, situado na microrregião de Goiânia a 18 km da capital e a 270 km de Brasília. Está situada na Região Centro Oeste do Brasil, conforme mostra a Figura 5. O município possui densidade habitacional de 1962 hab/km², clima tropical. Sua altitude é de 804 metros. Suas terras são do tipo sílico argilosa com pedreiras.

Figura 5 - Localização da Área de Estudo – Aparecida de Goiânia-GO



Fonte: A autora (2019).

Sua hidrografia é formada pelo rio Meia ponte, que serve de limite natural com outros municípios, e os Ribeirões das Lages, Santo Antônio e o Córrego da Serra. Não possui uma área geográfica grande que possibilita o setor primário e tem sua base econômica no setor industrial. Com precipitação média de 200 mm e temperaturas variando de 19°C a 28° C no período chuvoso, que se estende de dezembro a março.

A renda média dos Aparecidenses é de dois salários mínimos por mês e 29,6% da população tem um rendimento mensal de até um salário mínimo. O percentual de alunos de 6 a 14 anos de idade regularmente matriculados é de 95,3%, com um relativo decréscimo dos alunos do fundamental nos últimos 10 anos (IBGE, 2018).

Com mais de 579.179 habitantes, apenas 24,93% dos Aparecidenses contam com coleta de esgoto (ARAÚJO, 2018) o que coloca o município no ranking dos piores índices de coleta de esgoto, com condições sanitárias mínimas. A ausência desse serviço indispensável a um padrão de vida saudável expõe os moradores de vários bairros a conviver à proximidade de esgoto exposto, problemas como lixões, mau cheiro, proliferação de vetores de doenças tais como ratos, baratas e micróbios e a iminência de um contágio que a negligência do tratamento do esgoto acarreta.

A expansão urbana do município, que é marcado pelo acréscimo populacional e que se desenvolveu de forma desordenada, foi um grande contribuinte para a baixa taxa de serviços de saneamento básico, bem como educação e saúde. A ocupação de áreas inadequadas, juntamente com a degradação ambiental e a ausência ou precariedade de serviços básicos à população, expõe cotidianamente os cidadãos ao contágio de doenças.

Nas últimas duas décadas o município foi alvo de várias epidemias, entre as principais doenças, a diarreia teve surtos mais frequentes, uma vez que a doença é causada por uma gama de agentes patógenos, incluindo bactérias que muitas das vezes está presente na água contaminada, por falta de tratamento adequado da água, proximidade de poços e fossas, e contaminação da água por lixo.

Outras doenças que são acometidas pela ausência ou escassez de saneamento básico é a leptospirose que tem o número de casos da doença em geral associada ao acúmulo de água com baixo teor de matéria orgânica e a precariedade dos serviços de saneamento básico. Aparecida de Goiânia possui mais de 700 casos de dengue registrados durante as semanas 20 a 23 de 2019, sendo 140 incidências, o número de casos novos, que são prováveis, dividido pela população de determinada área geográfica, número expresso por 100 mil habitantes, como mostra o Quadro 3.

Quadro 3 - Municípios com maior coeficiente de incidência das durante as semanas 20 a 23 de 2019

Municípios	Casos	População	Classificação	Incidência
Goiânia	2607	1495705	Médio Risco	174
Anápolis	2186	381970	Alto Risco	572
Águas Lindas de Goiás	816	207070	Alto Risco	394
Aparecida de Goiânia	791	565957	Médio Risco	140
Valparaíso de Goiás	577	164723	Alto Risco	350
Luziânia	552	205023	Médio Risco	269
Cristalina	443	57759	Alto Risco	767
Rio Verde	425	229651	Médio Risco	185
Cocalzinho de Goiás	321	19971	Alto Risco	1607
Cidade Ocidental	184	69829	Alto Risco	407
Formosa	276	119506	Médio Risco	231
Caldas Novas	261	89087	Médio Risco	293
Santo Antônio do Descoberto	246	73636	Alto Risco	334

Fonte: Adaptada de SES-GO (2019)

Na atualidade a Secretária de Estado da Saúde de Goiás (SES-GO), realiza o monitoramento de epidemias, especialmente dos casos de dengue nos municípios com maiores

picos epidemiológicos como Aparecida de Goiânia que concentra o terceiro maior número de casos registrados de Dengue em todo o estado de Goiás entre a 1ª e a 23ª semana de 2019 como mostra a Quadro 4.

Quadro 4 - Municípios com maior número de casos notificados de dengue em Goiás no período da semana 01 a 23 de 2019

#	Municípios	Casos Notificados
1º	Goiânia	19.658
2º	Anápolis	10.048
3º	Aparecida de Goiânia	9.794
4º	Rio Verde	3.890
5º	Jataí	3.537
6º	Águas Lindas de Goiás	3.399
7º	Catalão	2.947
8º	Posse	2.573
9º	Luziânia	2.552
10º	Itumbiara	2.379
11º	Caldas Novas	1.725
12º	Formosa	1.596
13º	Valparaíso de Goiás	1.519
14º	Inhumas	1.236
15º	Senador Canedo	1.156

Fonte: Adaptada de SES-GO (2019)

Apesar da quantidade de casos e os óbitos confirmados como mostra o Quadro 5, a SES-GO classifica o município como risco médio. No entanto existe uma grande mobilização por parte do poder municipal para eliminar o vetor da doença, por meio de campanhas de conscientização, visitação às residências para verificação quanto a limpeza dos lotes e das residências desocupadas, bem como acompanhamento por meio do programa saúde na família que tem por objetivo promover uma melhor qualidade de vida dos moradores.

Quadro 5 - Municípios com óbitos suspeitos e confirmados de dengue em Goiás no período da semana 01 a 23 de 2019

Municípios	Óbitos Suspeitos	Óbitos Confirmados
Goiânia	12	8
Inhumas	1	3
Anápolis	2	2
Posse	0	2
Valparaíso de Goiás	0	2
Águas lindas de Goiás	3	1
Aparecida de Goiânia	4	1
Caldas Novas	1	1
Ceres	1	1

Cidade Ocidental	2	1
Corumbá	0	1
Formosa	1	1
Itapaci	0	1

Fonte: Adaptada SES-GO (2019)

A SES-GO, realiza os monitoramentos da dengue a partir de campanhas para levantamentos de infestações residenciais e prediais do vetor da doença, e a partir das informações dos casos de doenças que estão disponíveis no Sistema de Informação de Agravos Notificados (SINAN).

Mesmo com o monitoramento da dengue pela SES-GO, a extinção do vetor da dengue se arrasta por anos. O sistema da SES-GO, além de não possibilitar integração com variáveis externas não permite a realização de qualquer tipo de análise espacial e denota limitações tanto para entrada de dados históricos e recentes, quanto para extração de informações.

Sendo assim, destaca-se que SIGs são adequados para a solução de problemas complexos que envolvam dados espaciais e que apresentam dinamicidade, pois possibilitam as entrada e armazenamento de dados, conversão, processamento de imagens, manipulação e integração de dados, visualização de dados, consulta, combinação e análise espacial de dados, entre outros. Sendo assim, a elaboração de um SIG com banco de dados robusto e que apresente ferramentas de análise espacial, passa a exercer importante papel, tanto para a vigilância epidemiológica da dengue em Goiás, como também para auxiliar a compreensão da dinâmica espacial desta enfermidade.

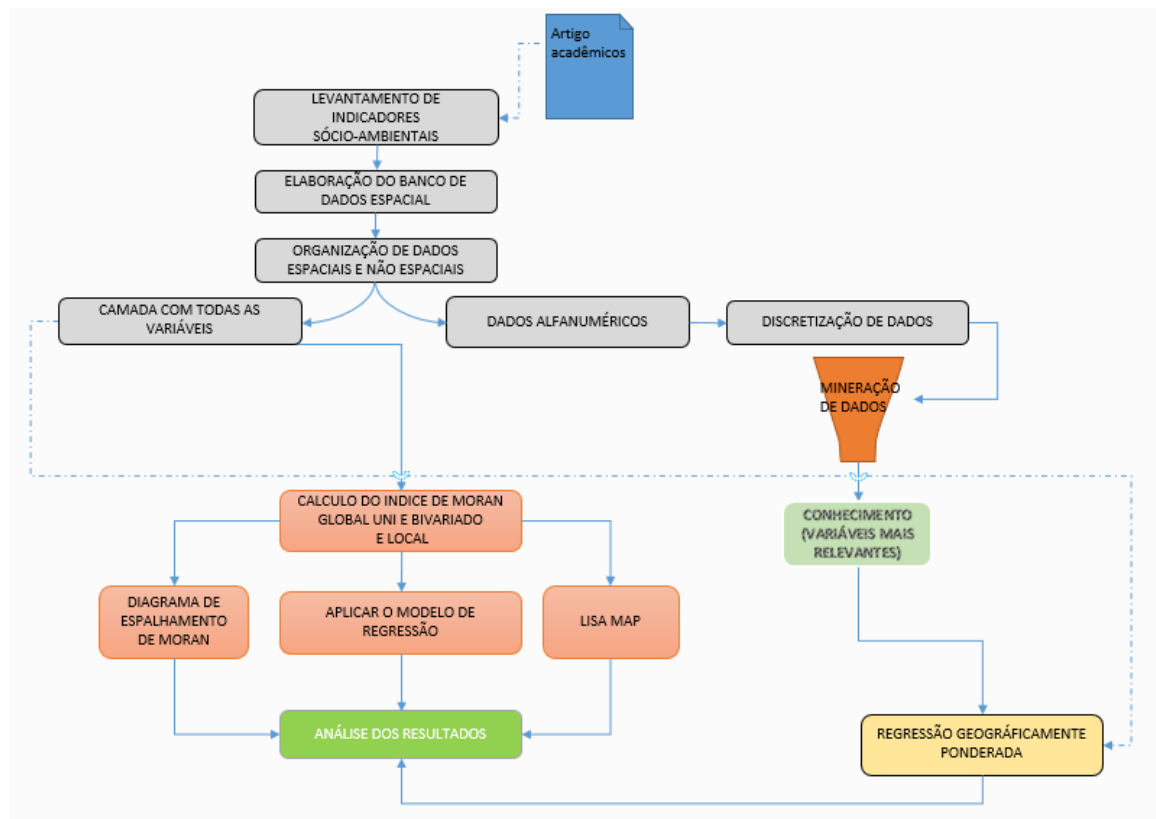
4. METODOLOGIA

O problema a ser resolvido refere-se à caracterização das áreas com maior índice de casos de dengue na cidade de Aparecida de Goiânia, de modo a identificar correlações espaciais em áreas com características semelhantes que são passíveis de uma crise epidemiológica, afim de que possam servir de base para intervenções para prevenir o desenvolvimento dessa epidemia.

Este capítulo descreve a metodologia aplicada para atingir os objetivos apresentados nessa dissertação. Este estudo baseou-se em um conjunto de procedimentos, como mostra o fluxograma na Figura 6, que demonstra as etapas de:

- i. Seleção de variáveis: Inventário das bases de dados cartográficos e alfanuméricos;
- ii. Pré-processamento de dados: Elaboração do Banco de Dados Geográfico utilizando modelagem OMT-G como suporte; preparação dos dados, normalização dos dados; Consultas espaciais;
- iii. Mineração de dados: Mineração de Dados (regras de associação baseadas em dados discretos usando algoritmo *Apriori*); Interpretação dos padrões encontrados;
- iv. Análise exploratória de dados espaciais: índice de Moran e regressão espacial.
- v. Regressão Geograficamente Ponderada

Figura 6: Fluxograma



Fonte: A autora (2019)

4.1 SELEÇÃO DE VARIÁVEIS

A dengue é considerada uma doença em extensa expansão pelo mundo (BIO-MANGUINHO, 2014). Por ser transmitida principalmente em ambiente urbano, analisar o

ambiente urbano como um todo pode não ser interessante, já que cada cidade ou município se caracteriza de forma distinta, tendo questões socioeconômicas e ambientais distintas.

Diversos fatores de risco podem estar relacionados com a presença da doença e podem ter mais ou menos influência dependendo da situação socioeconômica e ambiental do lugar. Taui (2001) destacou como principais fatores de risco para caracterização do padrão de transmissão: migrações, crescimento populacional, mau funcionamento dos postos de saúde, migrações e densidade populacional.

As variáveis usadas nesse estudo foram baseadas nos estudos de Rezendes et al. (2010), Leite (2010) e Mordini e Churavalloti Neto (2007) e demais autores, que consideraram fatores como renda, densidade populacional, saneamento básico entre outros.

As variáveis escolhidas foram agrupadas em seis categorias: renda, domicílio, população, saneamento básico, vegetação e saúde, como mostra o Quadro 6. As variáveis relacionadas a renda, domicílio, população e saneamento básico tem como base as tabelas de dados do Censo demográfico de 2010 e foram especializadas nos setores censitários por meio da ferramenta *Join* do programa *Qgis* versão 2.18.14.

Quadro 6: variáveis

Categoria	Variáveis	Abreviações / Observações	Baseado
Renda	Rendimento nominal mensal por pessoa responsável por domicílio particular permanente	TX_renda	Leite (2010); Paulino e Natal (1998); Mordini e Churavalloti (2007); Barcellos et al. (2005); Almeida (2004); Gonçalves Neto (2004); Machado (2007)
	Média do rendimento nominal mensal das pessoas responsáveis por domicílios particulares permanentes	TX_renda_media	Ferreira (2006); Paulino e Natal (1998); BARCELLOS et al. (2005); Almeida (2004); Barcellos et al. (2005); Almeida (2004); Gonçalves Neto (2004); Machado (2007)
Domicílio	Média do número de moradores em Domicílios particulares permanente	TX_med_moradores	Mordini e Churavalloti (2007); Ferreira (2006); Leite (2010); Santos (2004); Barrera et al. (2000)
População	Densidade populacional	TX_dens_pop (população residente no setor/área do setor censitário em Km ²)	Rezendes et al. (2010); Paulino e Natal (1998); Barcellos et al. (2005); Almeida (2004); Gonçalves Neto (2004); Machado (2007); Santos (2004); Barrera et al. (2000)
	Razão de dependência	TX_razao_dep ((população com menos de 15 anos+população com mais de 64 anos)/população entre 15 e 64 anos)*100	Paulino e Natal (1998); Barcellos et al. (2005); Almeida (2004); Gonçalves Neto (2004); Machado (2007)
Saneamento Básico	Esgoto	TX_esgoto (residencias que possuem coleta de esgoto no setor/residencias do setor)*100	Rezendes et al. (2010); Leite (2010); Costa e Natal (1998) Paulino e Natal (1998); Barcellos et al. (2005); Almeida (2004); Gonçalves Neto (2004); Machado (2007); Santos (2004); Barrera et al. (2000)

	Abastecimento de Água pela rede pública	TX_água (residências que possuem abastecimento de água tratada no setor/residências do setor)*100	Rezendes et al. (2010); Leite (2010); Costa e Natal (1998); Paulino e Natal (1998); Barcellos et al. (2005); Almeida (2004); Gonçalves Neto (2004); Machado (2007)
	Coleta de lixo	TX_lixo (residências que possuem coleta de lixo no setor/residências do setor)*100	Rezendes et al. (2010); Leite (2010); Costa e Natal (1998); Paulino e Natal (1998); Barcellos et al. (2005); Almeida (2004); Gonçalves Neto (2004); Machado (2007)
	Sem coleta de lixo e esgoto	TX_sem_lixo_esgoto (residências que não possuem coleta de lixo nem esgoto no setor/residências do setor)*100	Ferreira (2006);
Vegetação	Vazios urbanos	TX_vazios_urb (área dos vazios por setor censitário em Km ² /área do setor Km ²)*100	Santos (2004); Barrera et al. (2000)
Saúde	Casos de dengue	Tx_dengue quantidade de casos por setor / quantidade de moradores do setor	Leite (2010); Paulino e Natal (1998); Mordini e Churavalloti (2007); Barcellos et al. (2005); Almeida (2004); Gonçalves Neto (2004); Machado (2007)
	Unidades de saúde	Unidades de saúde	

Fonte: A autora (2019)

Os dados de casos de dengue se referem ao ano de 2018 e foram adquiridos junto à prefeitura de Aparecida de Goiânia, no formato vetorial contendo a quantidade de casos registrados por bairros. Como as demais variáveis se encontravam na escala de setores censitários (escala maior) os casos de dengue (escala menor) deveriam ser degradados para o mesmo nível de representação. O problema de incompatibilidade de escala é abordado pelo *Modifiable Areal Unit Problem - MAUP*, um viés estatístico usado para atribuir valores pontuais de fenômenos espaciais e atribuí-los a distritos ou polígonos (WONG, 2009). A quantidade de casos de dengue de cada setor foi considerada em taxa dividida (os casos pela quantidade de moradores do setor), e foi feito por meio da calculadora de campo do *software Qgis* na versão 2.18.14.

4.2 VAZIOS URBANOS

O município de Aparecida de Goiânia experimentou nas últimas décadas um crescimento intenso e desordenado, associado à implantação e consolidação de um polo industrial. Segundo Freitas (2019), Aparecida de Goiânia é uma das principais locomotivas do desenvolvimento econômico e social do estado de Goiás. Tanto a localização geográfica, no Centro-Oeste do país, como investimentos públicos de infraestrutura e modernização da saúde e educação são fatores que explica o porquê das pessoas de outras regiões do país estarem migrando para o município.

A ocupação do solo foi determinada pela expansão urbana gerando a heterogeneidade do espaço. Nota-se, em seu sítio urbano, a existência de barreiras, naturais e artificiais, que tornaram a mancha urbana descontinuada, dificultando a interligação entre os diversos setores da cidade. Entre essas barreiras estão: os rios Dourados e Meia Ponte, os ribeirões, Santo Antônio e Lages, a rodovia BR – 153, as redes de transmissão e grandes indústrias.

A retenção de terra para fins especulativos contribui para a descontinuidade do meio urbano. Alguns bairros possuem diversos lotes vagos, desprovidos de paisagens naturais e isso dificulta e encarece tanto a implantação como a manutenção de serviços públicos como rede de água e esgoto, pavimentação de ruas e área de lazer. E em muitos casos esses lotes vazios acabam sendo utilizados para descarte de lixo.

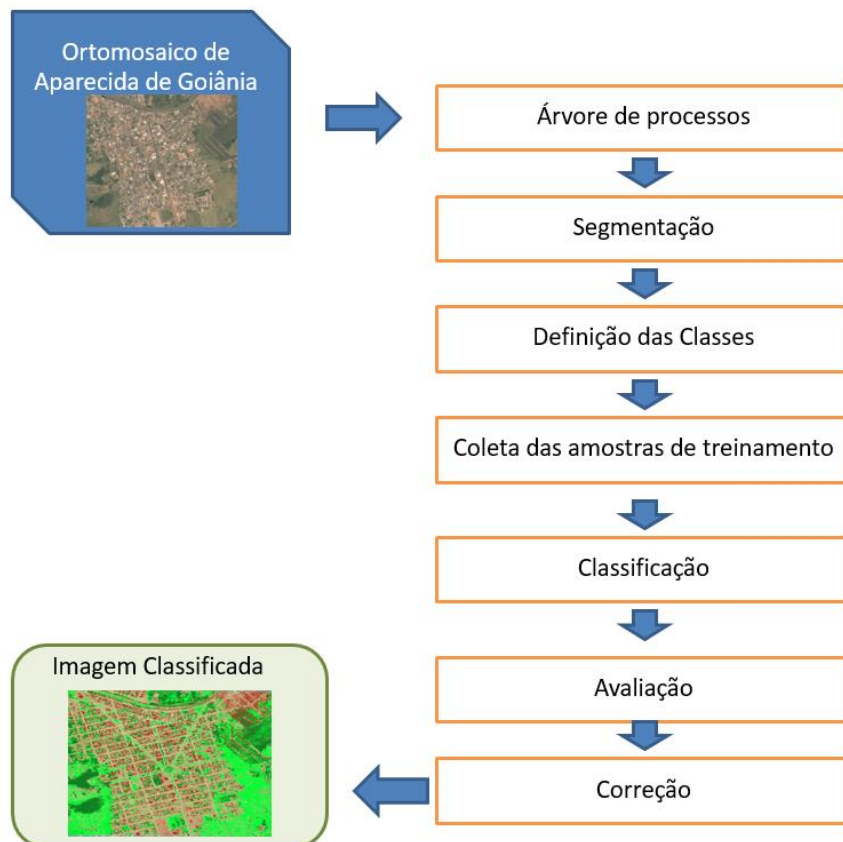
O descarte inadequado de lixo além de ser crime, acaba sendo depósito para água parada e moradia do *Aedes Aegypti*, mosquito transmissor da dengue (SOBRAL; Sobral, 2017). O depósito de lixo por períodos superiores há sete dias contribui para a criação do mosquito e, no entanto reduzir os números de dia entre coletas, bem como a limpeza e manutenção de lotes vazios e residências é capaz de reduzir o número de larvas do mosquito.

A coleta de lixo e o devido descarte do lixo é uma ferramenta contra a dengue. Vazios urbanos são normalmente definidos como aqueles terrenos localizados em áreas providas de infraestrutura que não realizam plenamente a sua função social e econômica, seja porque estão ocupados por uma estrutura sem uso ou atividade, seja porque estão de fato desocupados, vazios (BORDE, 2003).

Para identificar se existe correlação espacial entre os casos de dengue e os vazios urbanos foi gerado uma variável vegetação a partir da classificação orientada a objeto do ortomosaico do município de Aparecida de Goiânia do ano de 2011, adquirida juntamente a prefeitura do município, com tamanho de pixel de 50 cm. Para efeito da classificação, foi considerado como vazio urbano toda a área não ocupada existente dentro da Zona Urbana, excluído as áreas que possuem vegetação arbórea natural, áreas ociosas, sem construção, ou com área construída mínima, com presença de vegetação ou solo exposto, que possam servir de acúmulo de entulho e lixo, passíveis de serem criadores do mosquito que transmite a dengue e outras doenças.

Para a execução foram utilizadas tecnologias de processamento de imagens e técnicas de análise espacial e as atividades foram caracterizadas de acordo com o fluxograma da Figura 7.

Figura 7 – Fluxograma da classificação orientada a objeto do ortomosaico do município de Aparecida de Goiânia



Fonte: A autora (2019)

Tendo como base a imagem de alta resolução, foi criada uma *process Tree* implementada no programa eCognition 8.2 onde a base dos processos foi criada. O primeiro processo aplicado foi o de segmentação, o algoritmo usado foi o *chessboard segmentation*, e como parâmetro estabeleceu o tamanho do Object Size para 100.000 e inserido a camada vetorial com o formato dos lotes e das ruas. O algoritmo de segmentação utilizado permite a geração de objetos com formas geométricas mais bem definidas e delimitadas. Tendo uma forma quadrática de segmentação.

Depois de executado a segmentação o resultado foi de segmentos “grandes”. Foi necessário definir as classes do vetor, para ruas e lotes e classificá-los a partir dos atributos que constam na camada dos objetos. Foram usados os atributos da tabela de informação contida na camada de vetor, que designou quais polígonos eram ruas, lotes e os que não eram nenhum dos dois. As demais classes não possuíam padrões de forma, e, devido a isso, a camada de vetor possui apenas essas duas classes.

O processo foi executado, e a fim de melhorar o resultado dos objetos para a classificação, foi aplicada uma nova segmentação. O Algoritmo utilizado foi o *Multiresolution Segmentation*, que trata de uma segmentação baseada em objeto (multiresolução). Basicamente

esse tipo de segmentação leva em consideração várias características da imagem, como heterogeneidade espacial (forma) que também é subdividido em compacidade e suavidade, assim como espectral (cor) e diferença dos objetos de entorno, utilizando em síntese o método de crescimento de regiões para agregar pixels vizinhos. Esse processo termina quando o objeto excede o limiar (escala) definido pelo usuário.

O processo de segmentação pode ser aplicado por três parâmetros definidos pelo usuário: escala, forma e compacidade. O parâmetro de escala determina a heterogeneidade máxima permitida para os objetos da imagem, e seu valor varia de acordo com o tamanho dos objetos resultantes; um parâmetro de escala com valores maiores produz objetos maiores.

O parâmetro de forma consiste na união dos parâmetros de compacidade e suavidade, em que o parâmetro compacidade é definido pela razão da área do objeto e o raio da circunferência abrangente, e o parâmetro de suavidade expressa o limite entre objetos, em que, quanto mais fractal, maior o índice de forma (ESCH et al., 2008), (EVANS et al., 2010).

Por se tratar de uma imagem de altíssima resolução o resultado seria um dado bastante pesado. Para contornar esse problema, vários testes foram realizados em uma área de amostra da imagem, variando os parâmetros de forma, compacidade e escala, até encontrar um limiar aceitável para o nível de detalhe necessário para a determinação dos vazios e que tornasse o resultado mais leve.

Como parâmetros para a segmentação, foram usados um fator de forma de 80, a escolha desse tamanho se baseou em testes com 10, 20, 50, 80, 100 e 200, sendo o antepenúltimo o que mais se adequou. Os critérios de forma e compacidade foram definidos como 0.3 e 0.6 respectivamente, sendo que a determinação desses parâmetros é baseada em teste, uma vez que os mesmos variam de imagem para imagem devido a características espectrais e condições atmosféricas e climáticas locais. O modelo foi gerado para cada região.

Uma nova segmentação foi aplicada sobre a segmentação anterior, o algoritmo utilizado foi o *Spectral Difference Segmentation* onde ele uniu os segmentos a partir da diferença espectral definida como parâmetro, o valor estabelecido *do maximum spectral difference*, para esse caso foi, 40.

Para o processo de classificação foi usado o segundo nível de segmentação e foram inseridas as seguintes classes: água, asfalto, telha de cerâmica, telha de concreto, solo exposto, sombra, vegetação rasteira e vegetação como mostra a Figura 5.

Com o classificador *Standard Nearest Neighbor (generated)* definido para as classes, foram coletadas amostras de treinamento para a classificação. As amostras foram coletadas de

forma dispersa por toda a imagem afim de uma maior homogeneização de cada classe por toda a cena. A classificação direcionada a objeto utiliza a informação espectral de cada pixel da imagem e a informação espacial que envolve a relação entre o pixel e seus vizinhos. Estes classificadores procuram simular o comportamento de um fotointérprete, ao reconhecer áreas homogêneas, baseados nas propriedades espectrais e espaciais dessas áreas na imagem (ROLIM, 2003).

Uma vez executada a classificação, alguns aspectos da classificação foram editados, como os polígonos e a união desses. O passo seguinte foi a união dos polígonos dentro de cada classe correspondente na classificação, e o algoritmo escolhido para a execução do mesmo foi o *merge region*. A ferramenta *Swipe View* foi acionada para que se pudesse verificar o resultado da classificação.

Uma vez classificado, foram excluídas as classes que não compunham vazios urbanos como água, asfalto, cobertura de concreto, cobertura de cerâmica e sombra. Deixando apenas vegetação rasteira, vegetação e solo exposto.

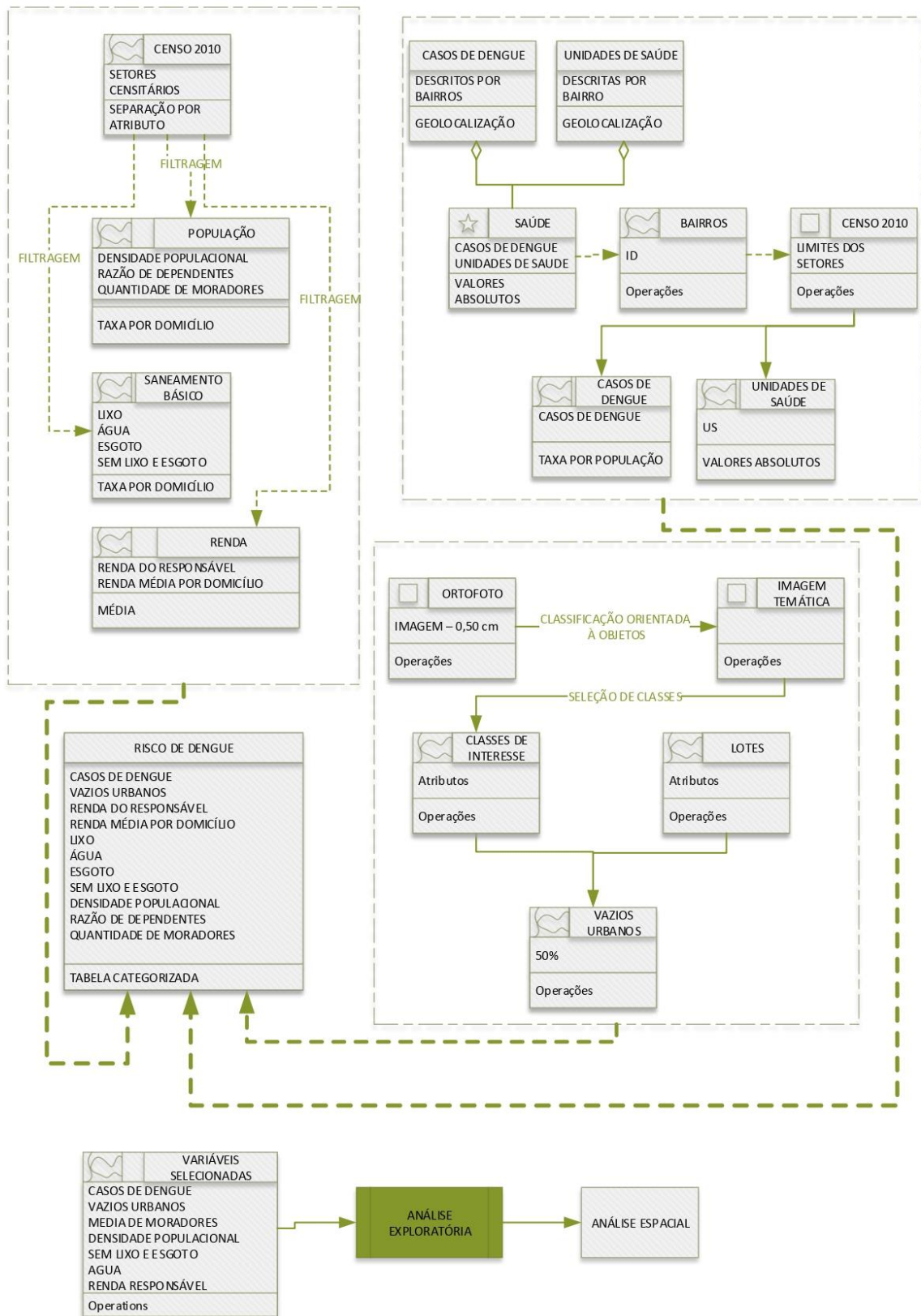
As classes resultantes foram unidas formando pré-vazios e unidas à camada de lotes de Aparecida de Goiânia. Os lotes com áreas de pré-vazio superior a 60% da área do lote foram considerados como vazios urbanos, os demais foram eliminados. A taxa de vazios urbanos para cada setor censitário considerou a área total dos vazios de cada setor dividido pela área do setor.

4.3 ESTRUTURAÇÃO DO BANCO DE DADOS GEOGRÁFICO

Para o estudo, foi criado um banco de dados geográficos que teve por objetivo armazenar as variáveis e possibilita consultas espaciais que sejam pertinentes a determinação de áreas para a intervenção de combate ao vetor da dengue. Ele possibilita visualizar as áreas com maior incidência de dengue e qual a condição socioeconômica dessas áreas. E o cenário contrário também é pertinente. As condições das áreas com baixa incidência de dengue podem ser bons indicativos de que tipo de abordagem deve e pode ser utilizadas.

Essas indicações só são possíveis de serem alcançadas porque o banco de dados geográfico criado levou em consideração um modelo do mundo real que utiliza tanto a abstração de *geo-campo* quanto a de *geo-objetos*. A Figura 7 mostra o modelo OMT-G em forma de Diagrama de Transformação. Os programas utilizados para elaboração do banco de dados foram o *pgAdmin III* versão 4.1 e o *Qgis* versão 2.18.14.

Figura 8- Modelo OMT-G em forma de Diagrama de Transformação.



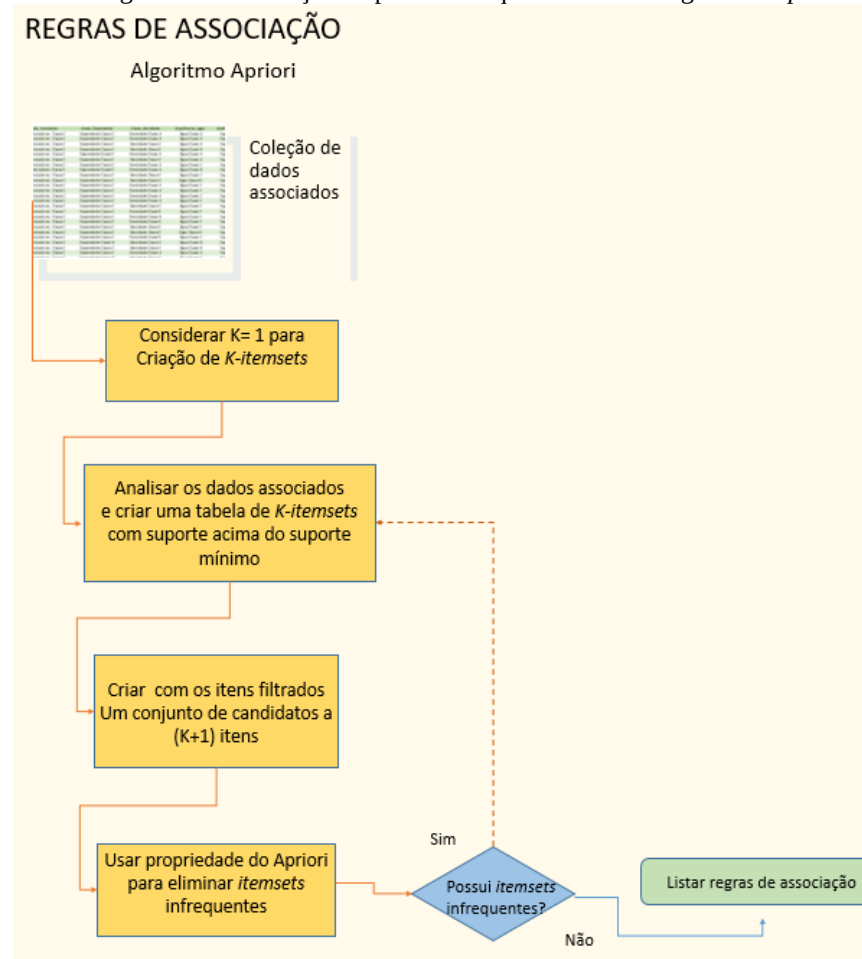
Fonte: A autora (2019)

4.4 MINERAÇÃO DE DADOS

A mineração de um conjunto de itens frequentes é um dos principais temas em KDD, seu objetivo é encontrar tendências que possam ser usadas para compreender e explorar padrões de comportamento dos dados. Seja $A = \{a_1, \dots, a_n\}$ o universo de itens, que pode ser qualquer conjunto de dados com itens frequentes, o conjunto de itens é chamado *Itemset*.

O algoritmo *Apriori* cria um conjunto F_1 de *itemsets* de tamanho 1 de forma que faz parte do conjunto de itens candidatos. Os *itemsets* $k+1$ são obtidos a partir dos *itemset* de tamanho k e são produtos da autocombinação sobre o conjunto dos itens com ele mesmo e filtragem. Somente os subconjuntos gerados que ocorrem em F_k serão os itens candidatos. O banco de dados é varrido em cada uma das transações e os *Itemsets* não frequentes são eliminados (FACELI, 2015). Uma vez eliminado todos os itens não frequentes, o algoritmo gera as regras de associação como mostra o fluxograma da Figura 8. O programa usado para a mineração foi o *Weka* (*Waikato Environment for Knowledge Analysis*; <http://www.cs.waikato.ac.nz/ml/weka>) versão 3.8.

Figura 9 - Fluxograma da mineração de padrões frequentes com o Algoritmo Apriori



Fonte: A autora (2019)

A regra de associação tem o formato, **se** antecedente **então** consequente. As regras são calculadas com base nos dados e são de natureza probabilística. A premissa básica é encontrar, na mesma transação, elementos que implicam na presença de outros elementos (BRUSSO, 2000). No caso desse estudo, a premissa é encontrar quais variáveis são frequentemente associadas a variável de casos de dengue.

A referência geográfica da informação usada nesse estudo é deixada de lado neste ponto e cada uma das variáveis usadas é representada de forma que se enquadre no formato de itens. Cada uma das variáveis possui valores reais e a etapa de mineração de dados com o uso do algoritmo *Apriori* procura por itens frequentes, e para que a possível associação entre as variáveis fosse possível cada variável teve seus valores discretizados de tal forma que o algoritmo pudesse gerar as regras.

Cada variável foi classificada arbitrariamente em intervalos que expressasse seus valores. A técnica escolhida para delimitar as classes de cada variável foi do tipo ponderável

com intervalos iguais e 5 grupos: Baixo (E), baixo a médio (D), médio (C), médio a alto (B) e alto (A). Essa técnica foi aplicada nas variáveis oriundas do Censo de 2010, já os casos de dengue, foram classificados como: Sem risco, Risco médio e Risco alto. As variáveis relacionadas a renda seguiram a classificação do IBGE, em que rendas inferiores a um salário mínimo se enquadram na classe Social E, de um a três salários mínimos se enquadram na classe social D, entre três e cinco salários mínimos, classe social C, entre cinco e quinze salários mínimos, classe social B e superior a 15 salários mínimos classe social A (DATOS MARKETING, 2012). O Quadro 7 mostra as classes associadas a cada variável. As unidades de saúde não foram utilizadas nessa etapa, devido a isso a variável não foi associada a nenhuma classe.

Quadro 7 – Classes associadas a cada uma das variáveis

ATRIBUTOS	CLASSES				
TX_RENDA	A	B	C	D	E
TX_RENDA_MEDIA	A	B	C	D	E
TX_MED_MORADORES	A	B	C	D	E
TX_DENS_POP	A	B	C	D	E
TX_RAZ_DEP	A	B	C	D	E
TX_ESGOTO	A	B	C	D	E
TX_AGUA	A	B	C	D	E
TX_LIXO	A	B	C	D	E
TX_LIXO_ESGOTO	A	B	C	D	E
TX_VAZ_URB	A	B	C	D	E
TX_DENGUE	SEM RISCO		RISCO MÉDIO	ALTO RISCO	

Fonte: A autora (2019)

Para a mineração de dados propriamente dita, foi utilizado o algoritmo J48 do *Weka* que possibilita a descoberta de padrões representada como uma árvore de decisão na forma de um classificador.

4.5 POLÍGONOS DE THIESSEN

Os casos de dengue são confirmados por um médico que faz um levantamento da história epidemiológica do paciente, observa quais os sintomas que já se manifestaram, se o paciente já teve dengue outras vezes e se o exame de sangue feito pelo paciente confirma a suspeita da doença (MINISTÉRIO DA SAÚDE, 2009).

Em caso de suspeitas de dengue o Comitê Municipal de Investigação de óbito e Controle da Dengue em Aparecida de Goiânia adverte as pessoas para que se encaminhem para a unidade de saúde mais próxima. A busca por atendimento médico não precisa ser exclusivamente à uma

Unidade Básica de Saúde (UBS), podendo ser também em Centros de Atenção Integrada à Saúde (CAIS) e Unidades de Pronto Atendimento (UPA) (MARTINS, 2019). Aparecida de Goiânia conta com dois CAIS que funcionam 24 horas por dia, quarenta UBS, com funcionamento das 7 às 17 horas e três UPAs que funcionam 24 horas por dia, além do hospital municipal e do centro de atendimento ambulatorial que contém atendimento de emergência 24 horas por dia (PREFEITURA DE APARECIDA DE GOIÂNIA, 2019).

No entanto, a busca por tratamento tem gerado uma demanda maior em alguns pontos do que em outros. Levando em consideração apenas a proximidade entre unidades de saúde e as residências e aplicando a técnica de polígonos de *Thiessen*, Também conhecido por diagramas de Voronoi, para uma simulação da área de abrangência de cada unidade de saúde é possível analisar quais unidades estariam sendo super-utilizadas e quais estariam sendo sub-utilizadas.

O objetivo dos polígonos de Thiessen é que, se considerarmos um plano, existem pontos que estão mais próximos a uma determinada fonte geradora do que outra, resultando em um polígono em que as distâncias entre a fonte e o ponto são as menores possíveis (Moura, 2003). Sendo assim, os polígonos gerados demonstram a área de cobertura de cada unidade de saúde considerando apenas a distância mínima e quando sobreposto à camada espacial contendo a quantidade de casos de dengue é possível visualizar quais polígonos das unidades de saúde que recobrem os setores com a maior quantidade de casos registrados, sendo essas unidades a primeira escolha para se procurar atendimento em casos de suspeita.

4.6 ANÁLISE DE DADOS DE ÁREA

A análise espacial de área trata de eventos aonde a localização espacial do dado está associada a área, como é o caso de estudos com fenômenos de agregados, objeto desse estudo no qual os atributos estão associados a cada polígono que representa um setor censitário, o objetivo dessa análise é buscar padrões espaciais na distribuição dos valores. Por tanto foi verificado se existe autocorrelação espacial (padrões espaciais) dos seguintes atributos: TX_dependentes, TX_agua, TX_esgoto, TX_lixo, TX_lixo_sem_esgoto, TX_vazios_urbanos, TX_renda, TX_media_moradores, TX_renda_media e TX_densidade entre os setores Censitários de Aparecida de Goiânia. O prefixo TX, neste trabalho, significa taxa.

Para a verificação da autocorrelação espacial global univariada, utilizou-se como ferramenta estatística o Índice de Moran global que analisou se os dados estavam

correlacionados ou se estavam distribuídos aleatoriamente no espaço. No caso do resultado ser uma correlação positiva, um agrupamento High-High, isso indica que os setores próximos obtiveram resultados semelhantes para a variável em análise. O mesmo ocorre quando o resultado da correlação é um agrupamento Low-Low. Caso resulte em uma correlação negativa, agrupamento High-Low e Low-High, isso indica que os setores próximos possuem valores diferentes para a variável em estudo. O índice de Moran global foi calculado para as variáveis citadas a cima.

A autocorrelação espacial global bivariada verifica a existência de correlação espacial entre duas variáveis de interesse. O intuito é analisar se os valores de uma variável em certa localidade tende a se relacionar com outra variável em locais vizinhos e se ambas as variáveis têm um comportamento semelhante.

Por Aparecida de Goiânia ser um município grande em extensão territorial foi necessário aplicar o índice de LISA. Por ser uma análise local é importante, pois o índice global pode ocultar nas estatísticas de autocorrelação padrões de associação local.

As análises exploratórias dos dados foram geradas no software Geoda.

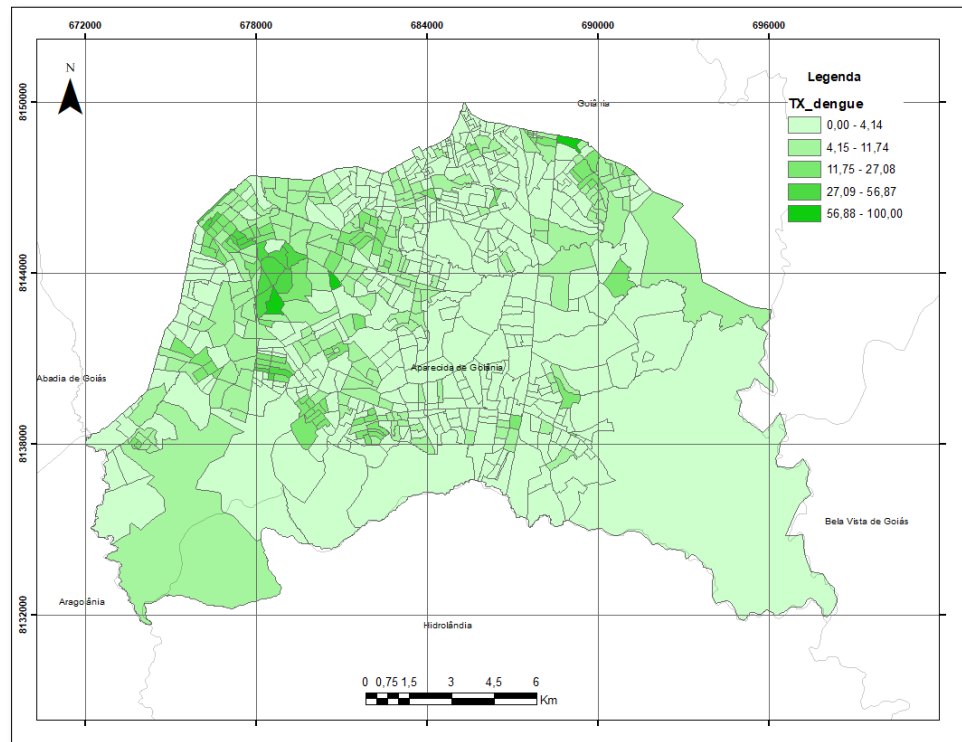
4.7 REGRESSÃO ESPACIAL

A regressão é uma ferramenta estatística que usa a relação entre variáveis de tal forma que uma variável possa ser explicada por outra. Como boa parte das variáveis geográficas está espacialmente correlacionadas, os modelos de regressão que não levam em consideração o caráter espacial deixam a desejar. Para este estudo foi aplicada uma regressão espacial pelo método clássico, onde a variável dependente é a que representa a taxa de casos de dengue e as variáveis independentes são as demais variáveis com exceção das unidades de saúde. Aplicou-se também, uma regressão geograficamente ponderada, técnica utilizada para modelar processos que variam no espaço, ou seja, heterogêneos espacialmente, Esta é uma técnica que tem por objetivo ajustar um modelo de regressão para cada uma das áreas em análise utilizando a localização espacial das outras áreas para determinar as estimativas dos parâmetros.

5 RESULTADOS E DISCUSSÕES

Com a espacialização e normalização das variáveis, foi organizadas as bases cartográficas em um ambiente SIG e realizada uma análise das áreas de influência das unidades de saúde por meio dos polígonos de Thiessen, que considera um plano, contendo pontos que

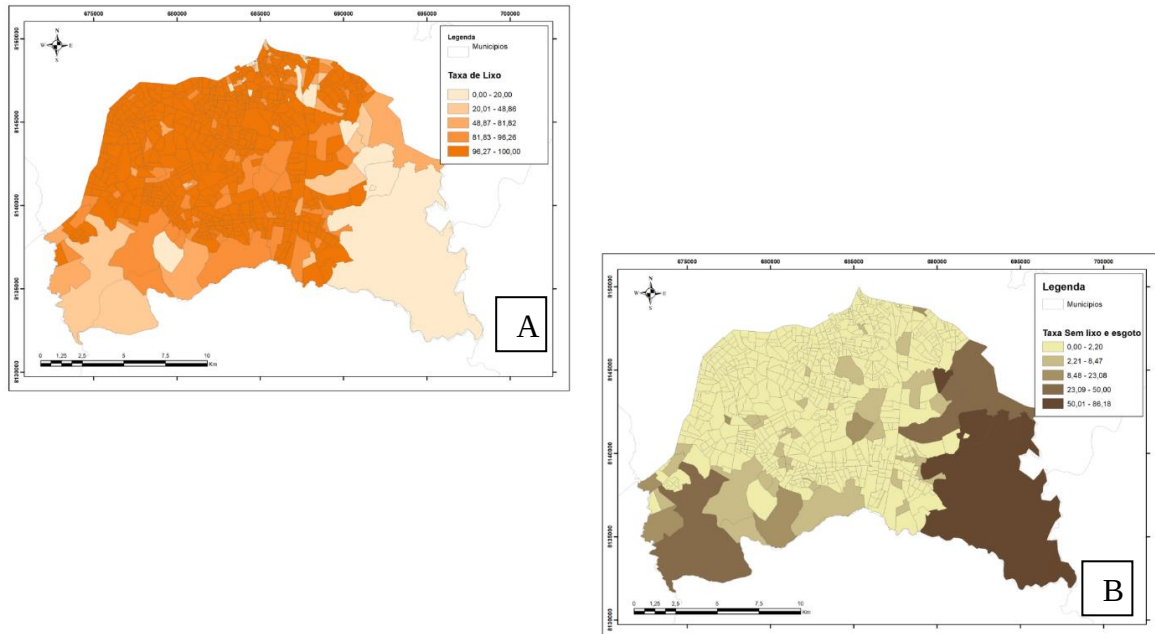
Figura 11 - Taxa de casos de Dengue nos setores Censitários em Aparecida de Goiânia



Fonte: A autora (2019)

Comparando visualmente o resultado da taxa de dengue por área de influência das unidades de saúde com os dados de taxa de esgoto, taxa de lixo, taxa de água, taxa de residências sem coleta de esgoto e lixo é possível identificar que os setores com maiores taxas de dengue são setores com baixa taxa de esgoto sem abastecimento de água ou com pouco abastecimento como mostra a Figura 12.

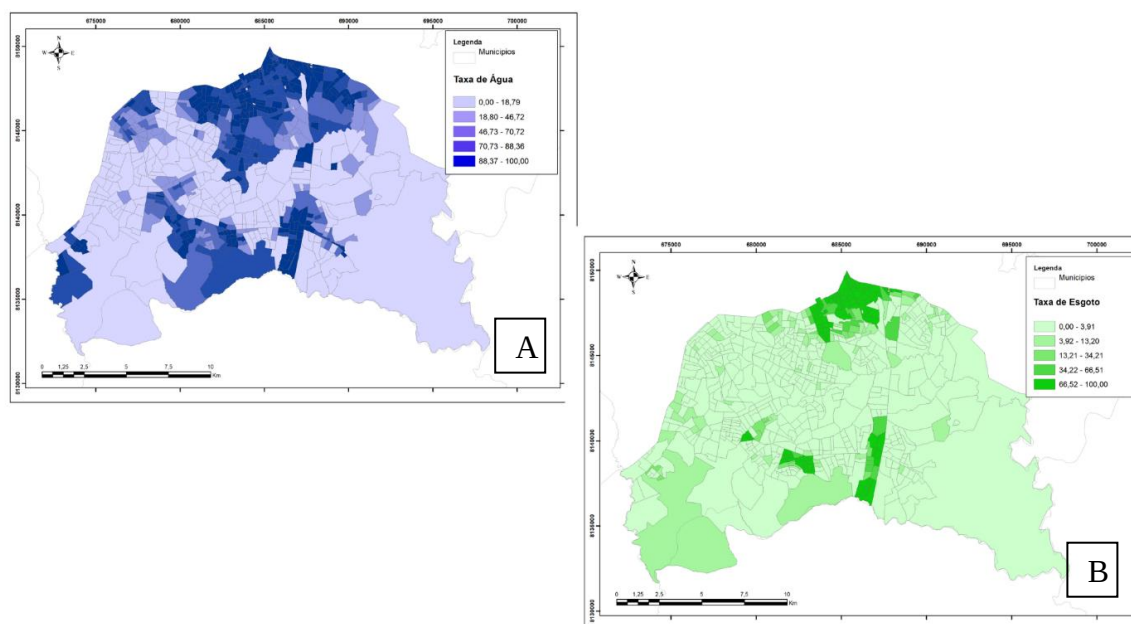
Figura 12- A) Taxa de setores com coleta de Lixo e B) Taxa de setores sem coleta de lixo e esgoto



Fonte: A autora (2019)

Em relação ao abastecimento de água e coleta de esgoto, Figura 13, a região com a maior concentração de casos de dengue fica em uma região com baixa taxa de residências com abastecimento de água e coleta de esgoto. Um cenário propício à infecções e doenças como descreve a World Health Organization (2018) quando aponta que a falta de saneamento básico ou a precariedade do mesmo resulta em contágio e afirma a importância do saneamento para a saúde.

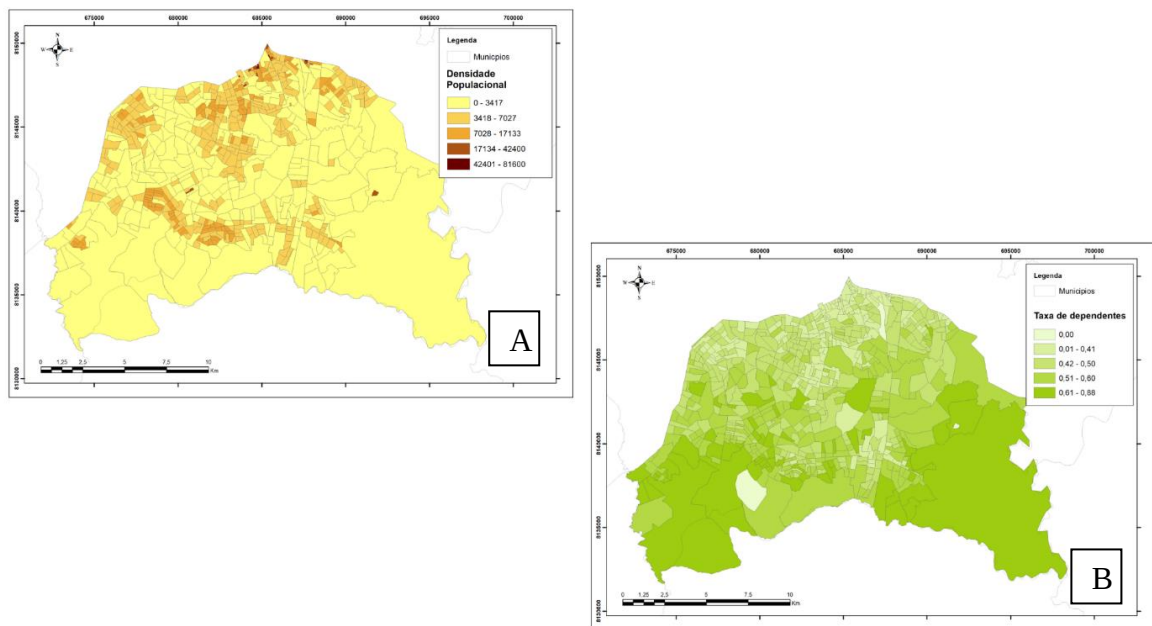
Figura 13 – A)Taxa de setores com abastecimento de água e B)Taxa de residências com coleta de esgoto



Fonte: A autora (2019)

Quanto ao adensamento populacional, nem todos os setores mais adensados ficam em regiões com a maior quantidade de casos de dengue como mostra a Figura 14, o que para o caso de Aparecida de Goiânia contraria a ideia de que a dengue afeta principalmente locais adensados. As regiões com maior número de casos de dengue têm taxas baixas de dependentes o que indica que a maior parte dos moradores dessas regiões tem idade entre 15 e 65 anos de idade.

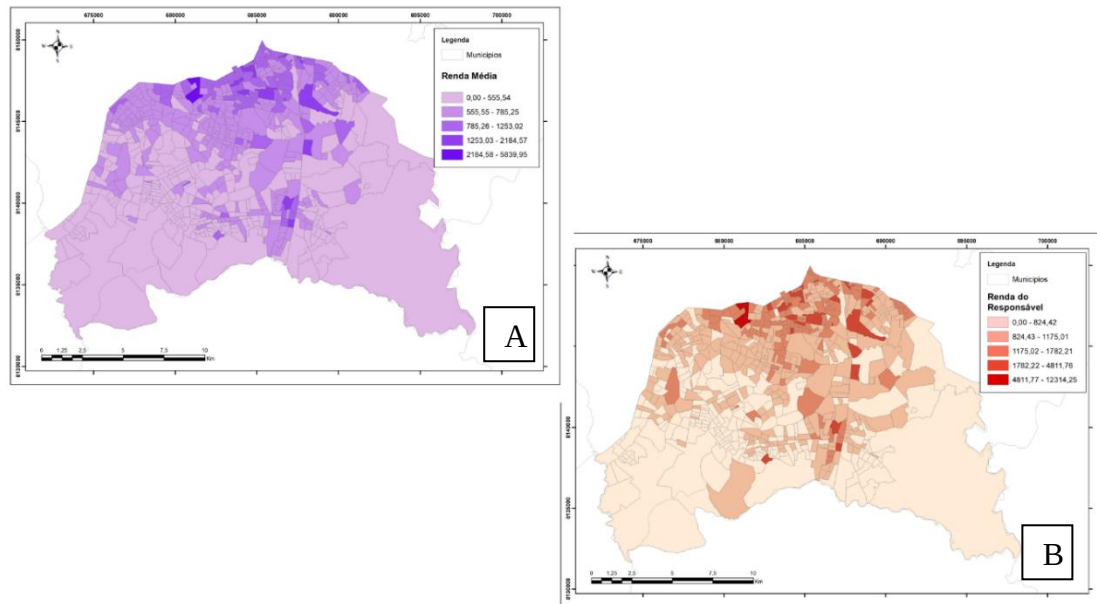
Figura 14: A) Densidade populacional e B) Taxa de dependente



Fonte: A autora (2019)

A renda média nesses setores é considerável, alguns dos setores com taxa elevada de dengue possui renda do responsável superior a 1175 aproximadamente a dois salários mínimos no ano de 2010, e renda média superior a 785 como mostra a Figura 15.

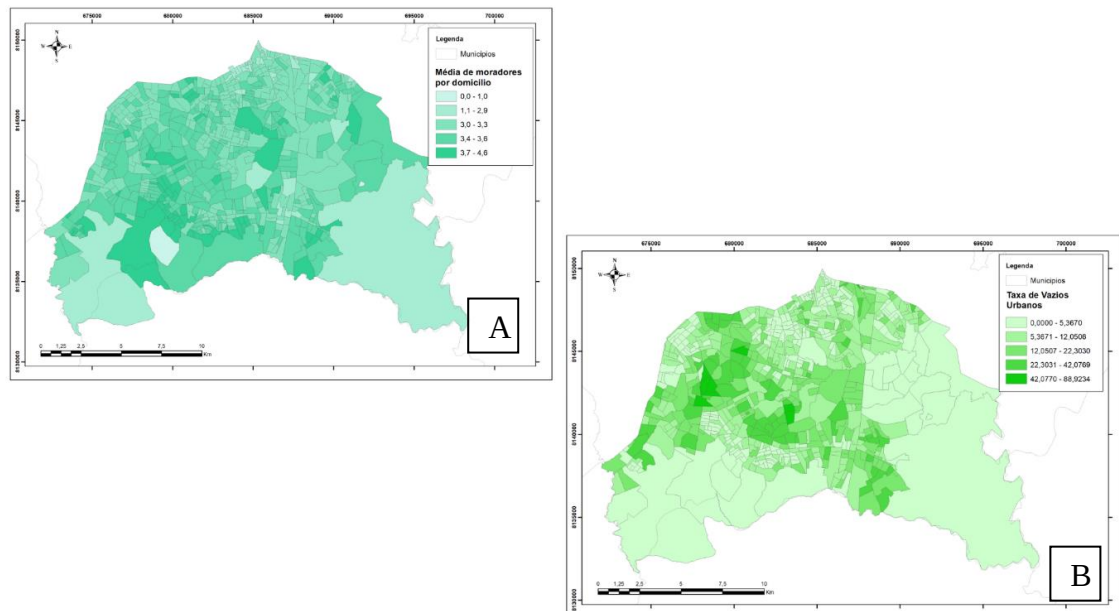
Figura 15 – A) Renda média e B) Renda do responsável



Fonte: A autora (2019)

A média de moradores por domicílio e a Taxa de vazios urbanos não expressa um padrão significativo como mostra a Figura 16.

Figura 16 – A) Média de moradores por domicílio e B) Taxa de vazios urbanos



Fonte: A autora (2019)

Quando o atributo escolhido foi a variável risco de dengue, a mineração de dados resultou em 1482 regras de associação, com 49,2492 % classificadas corretamente. O *Weka* retorna as melhores regras, algumas das regras são apresentadas a seguir:

- RISCO_DENGUE=Sem_Risco, RENDA_RESPONSAVEL = Renda_RESPONSAVEL_Classe_D, MEDIA = Media_moradores_Classe_C, DENSIDADE = Densidade_Classe_C, AGUA = Agua_Classe_A, DEPENDENTES = Dependente_Classe_D

Quando não existe risco de dengue a renda do responsável é de um a três salários mínimos, com uma média de 2 a 3 moradores por residência, uma densidade de 40 a 60%, ou seja, 36.640 a 48.960 moradores/Km² e a quantidade de residências do setor que recebem água tratada é superior a 80%, então à pelo menos um dependente em cada residência.

- RISCO_DENGUE=Alto_Risco, RENDA_RESPONSAVEL = Renda_RESPONSAVEL_Classe_D, MEDIA = Media_moradores_Classe_C, DENSIDADE = Densidade_Classe_A, AGUA = Agua_Classe_C, "ESGOTO = Esgoto_Classe_C"

Quando o risco de dengue é alto, a renda do responsável é entre um e três salários mínimos, a média de 2 a 3 moradores por residência, a densidade de moradores é superior a 80%, ou seja, 65.280 moradores/Km² superior a e a quantidade de residências do setor que recebem água tratada é de 40 a 60%, então somente 40 a 60% das residências do setor possuem coleta de esgoto.

- RISCO_DENGUE=Alto_Risco, RENDA_RESPONSAVEL = Renda_RESPONSAVEL_Classe_D, MEDIA = Media_moradores_Classe_C, DENSIDADE = Densidade_Classe_A, AGUA = Agua_Classe_C, "ESGOTO = Esgoto_Classe_B"

Quando o risco de dengue é alto, a renda do responsável é entre um e três salários mínimos, a média de 2 a 3 moradores por residência, a densidade de moradores é superior a 80%, com população/Km² superior a 65.280 e a quantidade de residências do setor que recebem água tratada é de 40 a 60%, então somente 60 a 80% das residências do setor possuem coleta de esgoto.

- RISCO_DENGUE=Risco_Medio, LIXO=Lixo_Classe_A, VAZIO_URBANO=Vazios_urbanos_Classe_E, "RENDA_RESPONSAVEL=Renda_RESPONSAVEL_Classe_D"

Quando o risco de dengue é médio, a porcentagem de coleta de lixo é de 80 a 100% e a porcentagem de vazios urbanos do setor censitário é inferior a 20% então a renda do responsável é de um a três salários mínimos.

- RISCO_DENGUE=Risco_Medio, DENSIDADE=Densidade_Classe_A, VAZIO_URBANO= Vazios_urbanos_Classe_E, "SEM_LIXO_ESGOTO= Sem_lixo_e_sem_Esgoto_Classe_E"

Quando o risco de dengue é médio, a porcentagem de densidade populacional é de 80 a 100% e a porcentagem de vazios urbanos do setor censitário é inferior a 20% então a porcentagem de residências sem coleta de lixo e esgoto é inferior a 20%.

A mineração de dados mostra uma correlação existente entre os casos de dengue e as variáveis escolhidas, os casos de dengue classificados como médio risco estão frequentemente associados a certas variáveis. Entre as variáveis propostas, as mais frequentes são Renda do responsável, média de moradores por residência, densidade populacional, Taxa de residências com abastecimento de água, Taxa de residências sem coleta de lixo e esgoto e os vazios urbanos.

Esses resultados expressam um cenário baseado nos valores categorizados e sem levar em conta a distribuição espacial.

Outro resultado observado é a seleção de variáveis pelo minerador de dados. O resultado é apresentado na Figura 17. As variáveis mais significativas foram: Renda_Responsável, Densidade, Sem_Lixo_Esgoto, água e Vazio_Urbano.

Figura 17 - Variáveis selecionadas pelo minerador de dados

```

Relation:  DENGUE
Instances: 666
Attributes: 11
    RENDA_RESPONSAVEL
    REND_MEDIA
    MEDIA
    DEPENDENTES
    DENSIDADE
    AGUA
    ESGOTO
    LIXO
    SEM_LIXO_ESGOTO
    VAZIO_URBANO
    RISCO_DENGUE
Evaluation mode: 10-fold cross-validation

=== Attribute selection 10 fold cross-validation (stratified), seed: 1 ===

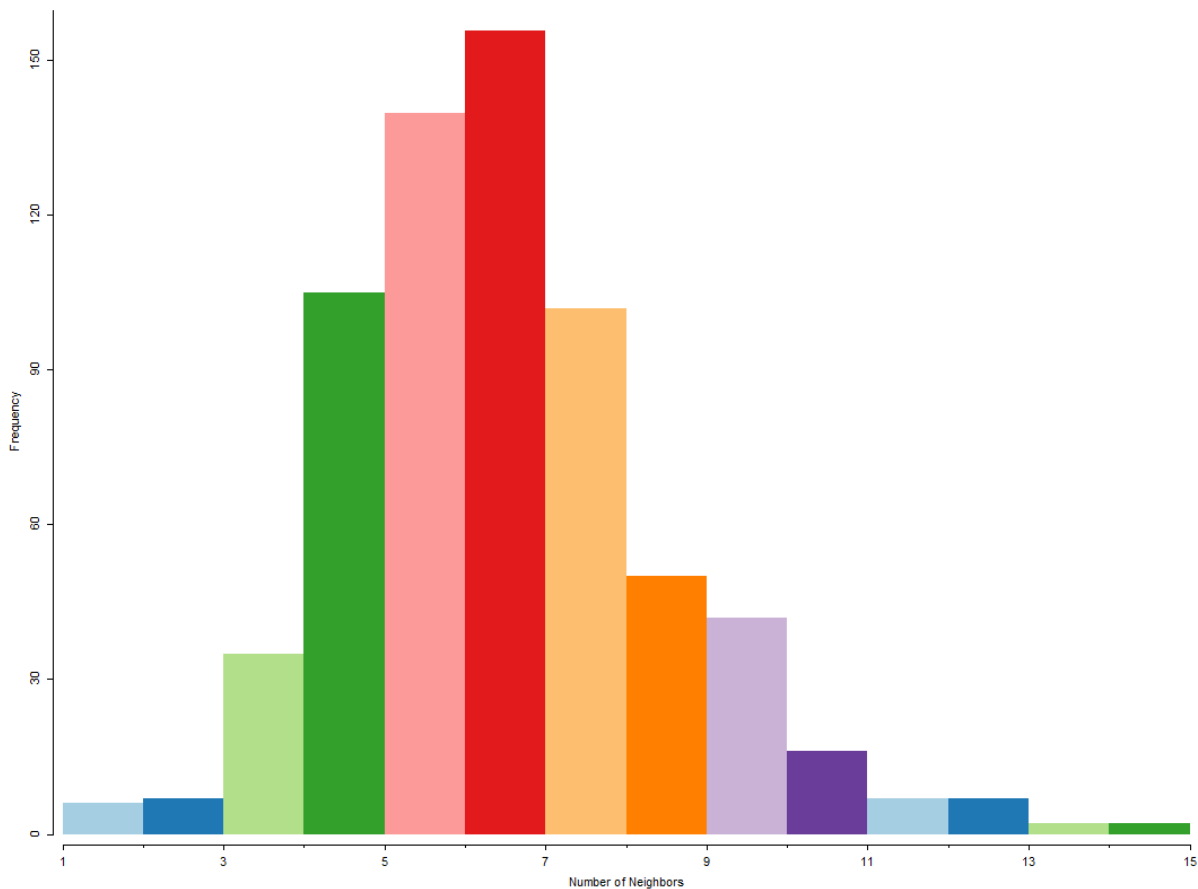
number of folds (%) attribute
10(100 %) 1 RENDA_RESPONSAVEL
1( 10 %) 2 REND_MEDIA
6( 60 %) 3 MEDIA
0( 0 %) 4 DEPENDENTES
10(100 %) 5 DENSIDADE
9( 90 %) 6 AGUA
3( 30 %) 7 ESGOTO
0( 0 %) 8 LIXO
10(100 %) 9 SEM_LIXO_ESGOTO
8( 80 %) 10 VAZIO_URBANO

```

Fonte: A autora com base no programa *Weka* (2019)

O primeiro passo para a aplicação do índice de Moran para calcular a autocorrelação foi gerar a matriz de proximidade espacial para identificar as regras de proximidade para cada polígono. Os critérios de vizinhança usados foram de contiguidade Queen com uma ordem de contiguidade, resultando em uma matriz de vizinhança que atribuiu valor 1 para polígonos que são vizinhos e 0 para aqueles que não são vizinhos. A Figura 18 ilustra o histograma de contiguidade. Nota-se que neste critério a maioria dos municípios possuem entre 5 e 7 vizinhos em comum, com poucos 1, 13 ou 15 vizinhos.

Figura 18 - Histograma de conectividade para o critério contigência Queen



Todas as matrizes abordadas neste capítulo foram criadas pelo software GeoDa.

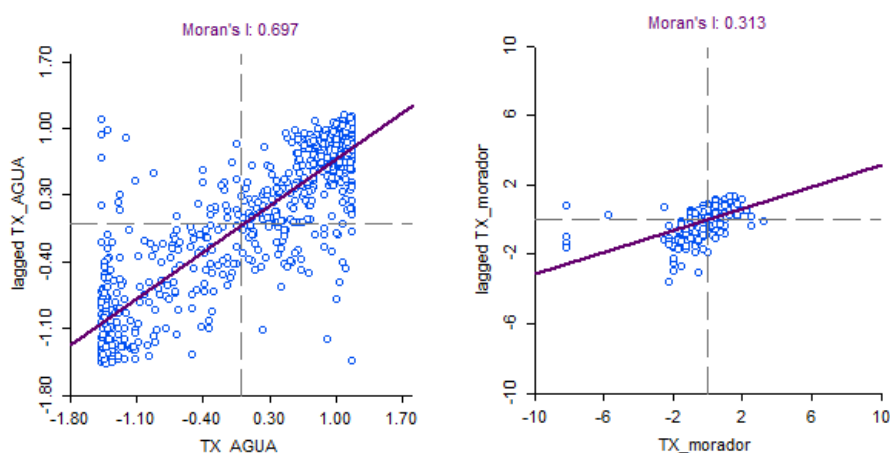
Para a análise exploratória, primeiramente, foi calculado o índice global de Moran para os casos de dengue e as demais variáveis como a Tabela 1 demonstra. Podemos observar que todos os índices são positivos, indicando existir uma autocorrelação positiva em todos os casos. O maior índice foi obtido para a variável de taxa de água que representa a percentagem de residências que tem acesso ao abastecimento de água em cada setor Censitário, enquanto o menor índice foi obtido para o atributo de taxa da média de moradores, que indica qual a média de moradores por domicílio em cada setor Censitário.

Tabela 1 - Teste de autocorrelação espacial univariada

INDICADORES	ÍNDICE GLOBAL DE MORAN
Dengue	0,352
Água	0,697
Densidade	0,317
Dependente	0,511
Esgoto	0,762
Lixo	0,369
Média de moradores	0,313
Renda média	0,417
Renda	0,380
Sem Lixo e sem esgoto	0,370
Vazios Urbanos	0,486

Podemos observar que o índice global para as variáveis foi positivo, embora em algumas não seja uma correlação muito alta é significativa, indicando existir, em todos os casos, uma autocorrelação espacial positiva.

O gráfico de espalhamento de Moran, Figuras 19, apresenta características bem diferentes para as variáveis de água e média de moradores. No caso da taxa de água há uma forte relação entre o valor de um setor em relação ao valor dos seus vizinhos. Em contra partida, na variável de taxa de média de moradores, existe concentrações de dados nas regiões Low-High e High-Low, que representam setores mais afastados dos centros comerciais e históricos. A situação discrepante pode ser explicada pelo fato de muitos desses setores serem em regiões periféricas, onde não existe um planejamento familiar e as famílias acabam sendo constituídas por um número maior de indivíduos, ao passo que há setores com uma melhor infraestrutura, mais bem consolidados e histórico que são compostos por famílias com um número menor de indivíduos ou com uma população mais idosa morando sozinhos.

Figura 19 - Gráfico de espalhamento de Moran

O índice de Moran bivariado faz a comparação de duas variáveis, nesse caso, as demais variáveis foram comparadas com a variável de taxa de casos de dengue e seus resultados são apresentados na Tabela 2. Em relação às variáveis de taxa de água, esgoto, renda média, renda e sem lixo e sem esgoto, os índices de Moran Bivariados indicam que essas variáveis estão, predominantemente no segundo e no quarto quadrante do diagrama de dispersão de Moran, nas regiões High-Low e Low-High respectivamente, de modo que os setores com as maiores taxas de residências com abastecimento de água são vizinhos dos setores com as menores taxas de residências com abastecimento de água, e o contrário também é verídico e se aplica as variáveis de esgoto, renda média, renda e sem lixo e sem esgoto. O índice de autocorrelação espacial bivariado, ainda que positivo para algumas variáveis, apresenta impacto espacial na taxa de casos de dengue nos setores Censitários de Aparecida de Goiânia, mas não expressam valores significativos, acima de 0,20.

Tabela 2 - Teste de autocorrelação espacial bivariada

INDICADORES	ÍNDICE GLOBAL DE MORAN
Água	-0,057
Densidade	0,013
Dependente	0,019
Esgoto	-0,097
Lixo	0,037
Média de moradores	0,019
Renda média	-0,059
Renda	-0,058

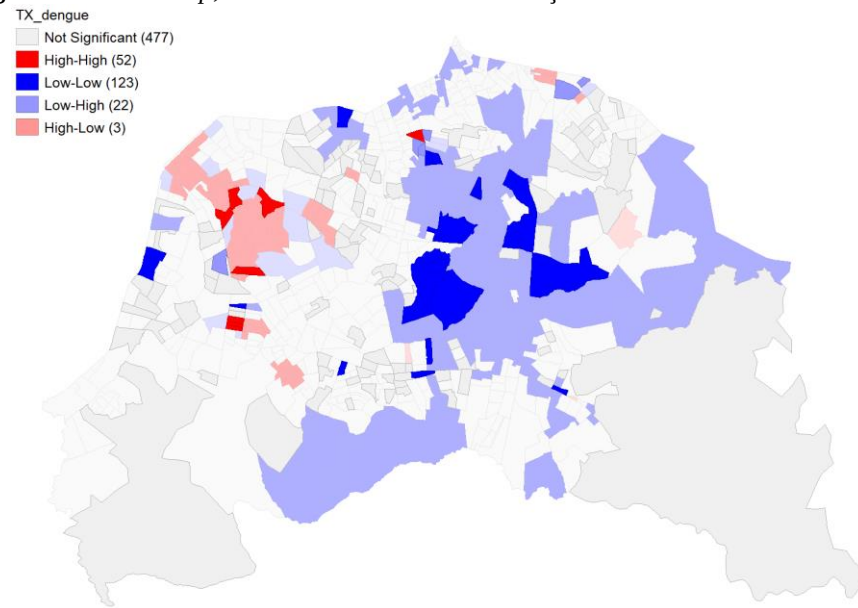
Sem Lixo e sem esgoto	-0,063
Vazios Urbanos	0,081

Os padrões de distribuição dos casos de dengue também foram examinados em uma escala menor por meio do Moran local, *Local Indicators of Spatial Association* (LISA), resultando em um valor específico para cada setor censitário, possibilitando a visualização de agrupamentos de setores com valores bem parecidos para os casos de dengue.

A Figura 20 representando o índice de Moran Local da variável de taxa de casos de dengue mostra as correlações do tipo “High-High”, “Low-Low”, “High-Low” e “Low-High” e auxilia na localização dos clusters. É possível verificar duas regiões com dependência espacial que são significantes estatisticamente. Uma representa o comportamento do tipo Low-Low, situa-se basicamente na região central do município, são setores que possuem elevadas taxas de coleta de lixo, abastecimento de água e coleta de esgoto, com baixa densidade populacional, com taxa considerável de dependente, com média de 4 a 5 moradores por residência e com taxa de vazios urbanos superior s 40 %. A segunda região é composta por poucos setores e são representadas por um comportamento do tipo High-High. Essas regiões são bem localizadas no município de Aparecida de Goiânia, próximo a grandes avenidas, comércios, feiras, grandes hospitais, onde a prefeitura tem atuado em campanhas relacionadas a saúde. São setores relativamente valorizados justamente pelo acesso a infraestrutura básica. No entanto, esses setores possuem áreas não ocupadas, proximidade a corpos hídricos e vegetação, locais que são passíveis do descarte inadequado de lixo que acabam servindo como depósito para água parada e moradia do mosquito transmissor da dengue.

O LISA *map* comparou os valores de casos de dengue de localização específica com os valores dos seus vizinhos e foi usado para testar a hipótese nula de distribuição aleatória dos dados.

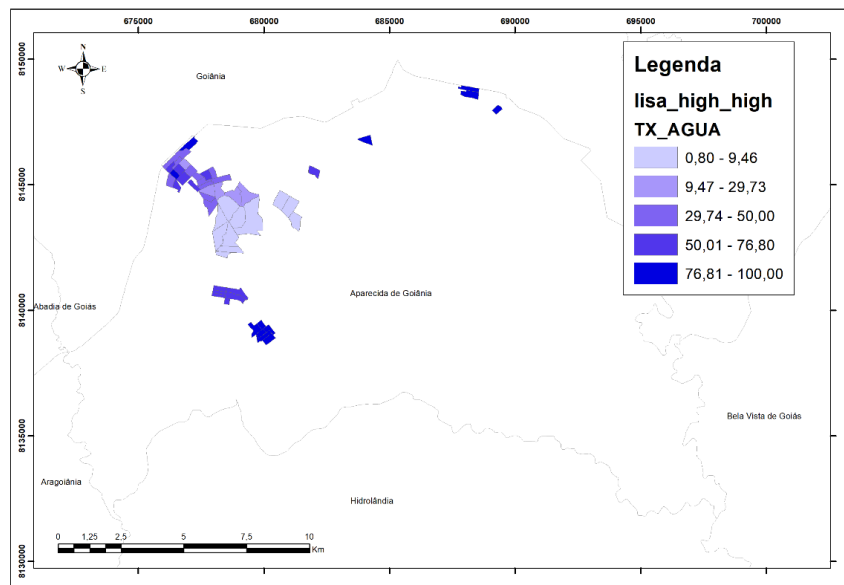
Figura 20 - LISA map, indicando clusters de distribuição de alto e baixos valores



Fonte: A autora com base no programa *GeoDa* (2019)

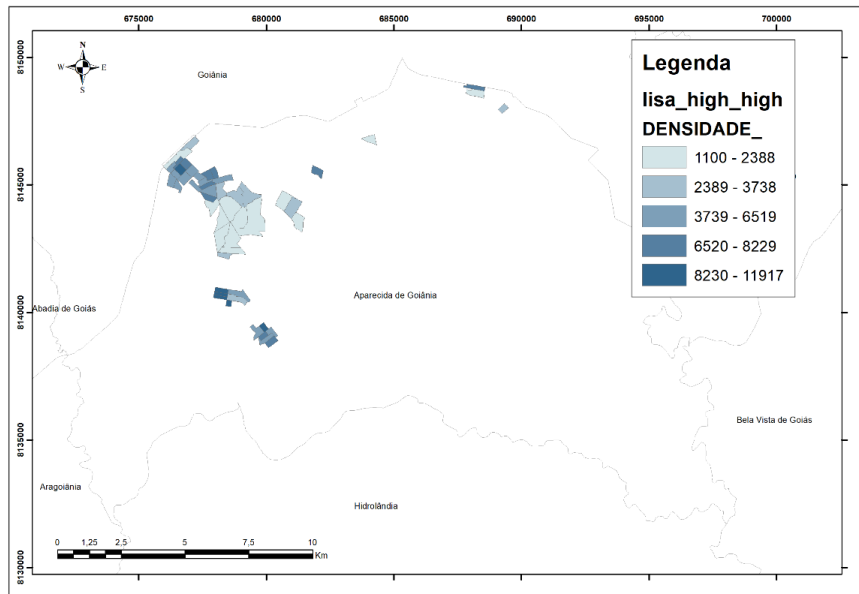
As Figuras 21, 22, 23, 24, 25 e 26 mostram as variáveis resultantes da etapa de mineração de dados para os setores com correlação high-high.

Figura 21 - Taxa de água dos setores com correlação high-high



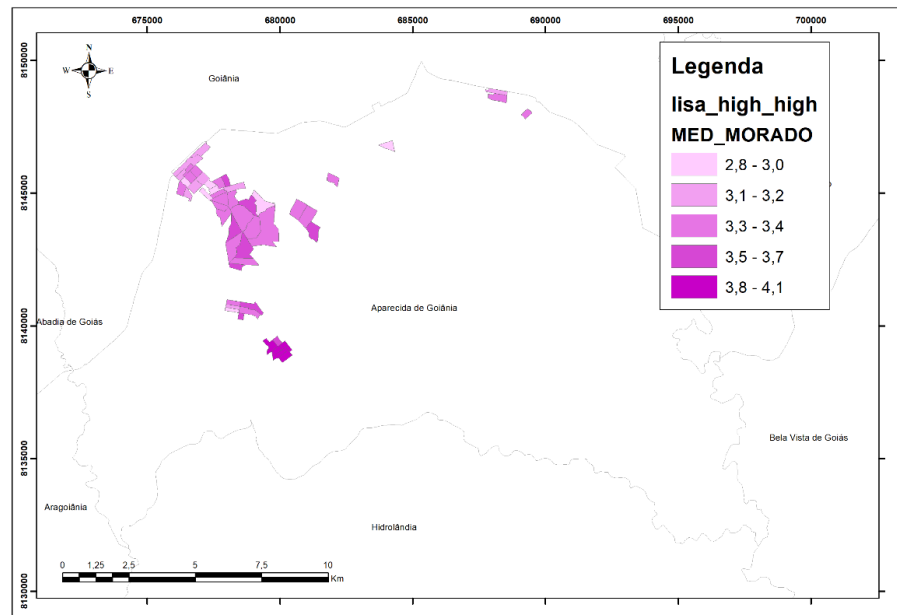
Fonte: A autora (2019)

Figura 22 Densidade dos setores com correlação high-high



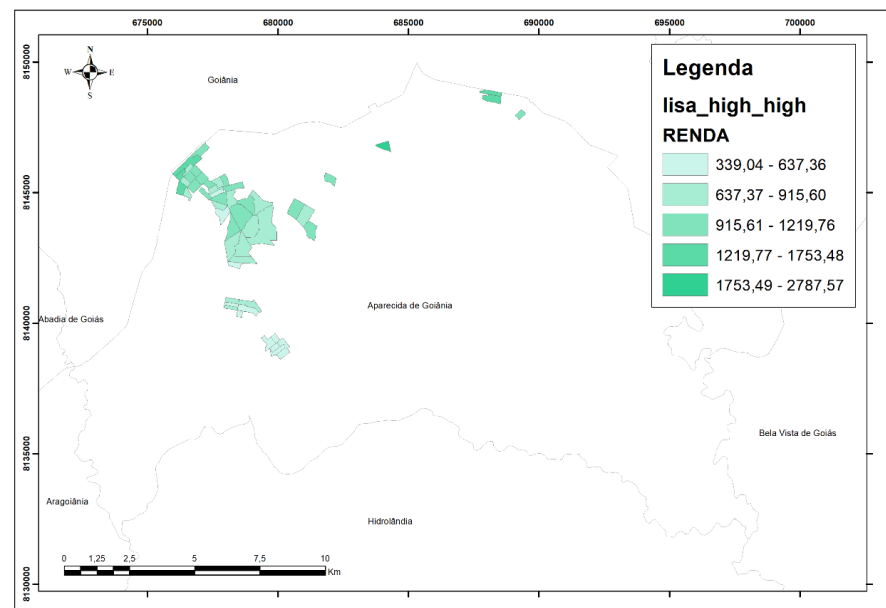
Fonte: A autora

Figura 23 - Media de moradores dos setores com correlação high-high



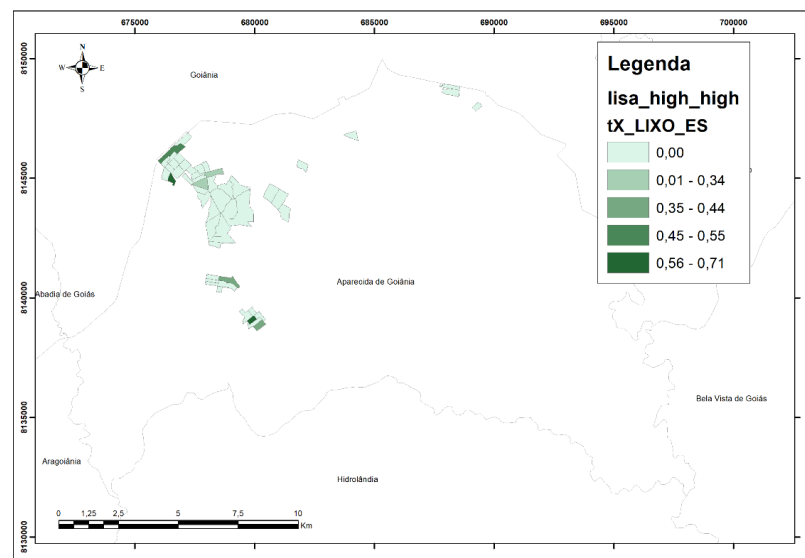
Fonte: A autora (2019)

Figura 24 - Renda do responsável dos setores com correlação high-high



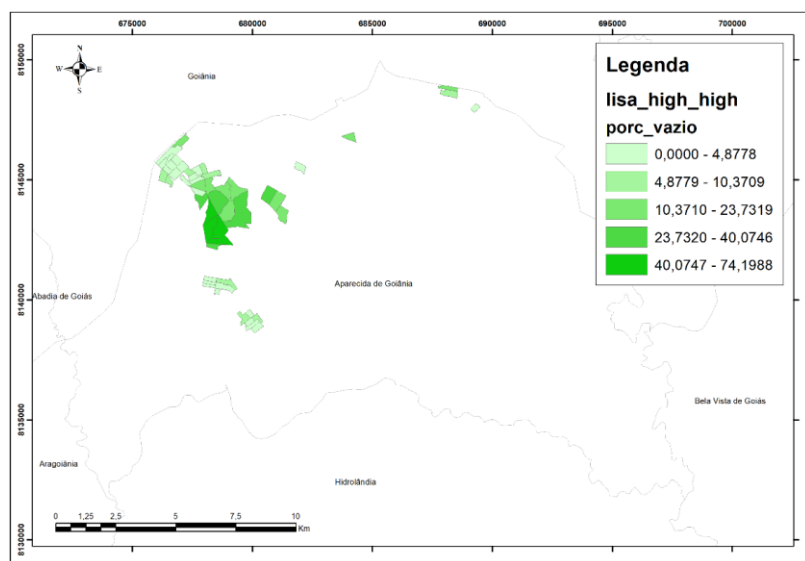
Fonte: A autora (2019)

Figura 25 - Taxa de residências sem coleta de lixo e esgoto dos setores com correlação high-high



Fonte: A autora (2019)

Figura 26 - Taxa de vazios Urbanos dos setores com correlação high-high



Fonte: A autora (2019)

Analizando os atributos determinados como mais frequentes para esses setores não fica muito claro um padrão exato que descreva os ambientes em que a taxa de dengue é mais frequente. Para cada uma das variáveis há uma dispersão considerável entre os valores de cada setor, não demonstrando uma contiguidade.

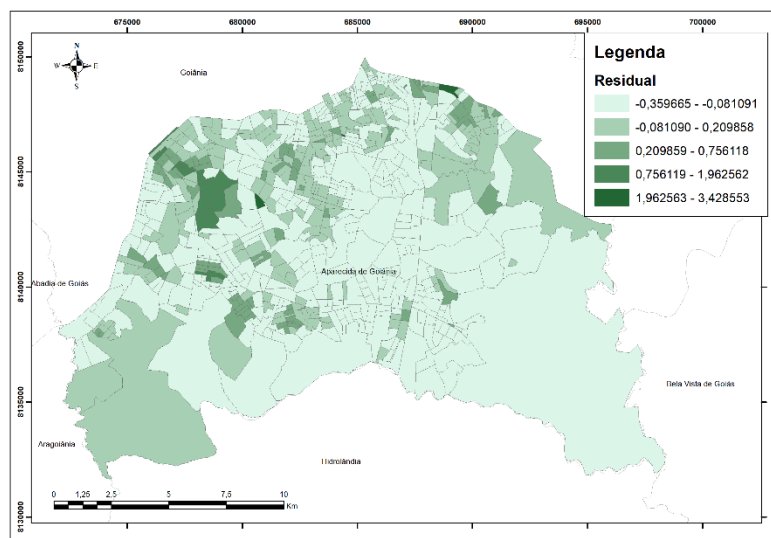
A correlação da taxa de dengue com as demais variáveis mostrou uma correlação baixa, quase nula, próxima de zero, com valores de 0,08 que indicam que quando a taxa de casos de dengue é comparada com as demais variáveis, ambas as variáveis movem-se em direções opostas. Isso pode ser explicado com a ausência de correlação entre as variáveis com os casos de dengue ou pelo equívoco na escolha do método utilizado, a correlação linear.

Em relação a regressão, foi executado tanto a regressão clássica quanto à SAR e realizando o diagnóstico sugerido por Anselin (2005) de autocorrelação espacial optou-se pela escolha do modelo regressivo SAR, que resultou em um pseudo-coeficiente de determinação, R^2 igual a 0,3196.

Para a regressão geograficamente ponderada, a variável dependente foi definida como os casos de dengue, e as variáveis explicativas foram as variáveis de caráter socioeconômico e ambiental usadas na regressão espacial SAR. A RGP foi calculada, aplicando individualmente a cada uma das variáveis. Os cálculos foram executados com a ferramenta *Geographically Weighted Regression* encontrada nas ferramentas de estatística espacial do ArcGis.

Quando aplicada a RGP, o R-Squared (r^2) apresentou um índice baixo, com valores inferiores a 0,04. Apesar dos valores da sua residual apresentar característica espacial aleatória, uma variável com valores tão inferiores não consegue explicar o fenômeno da dengue, nesse caso serão necessárias outras variáveis para dar robustez ao modelo. Os valores residuais são apresentados na Figura 28.

Figura 27 - Residual da Ponderação geograficamente ponderada



Fonte: A autora (2019)

6. CONSIDERAÇÕES FINAIS

As ferramentas de SIG, juntamente com as técnicas de análise exploratória espacial aplicada, proporcionaram a confecção de tabelas e mapas, que oferecem de forma clara e precisa informações coerentes com a realidade, demonstrando a eficácia na utilização dessas ferramentas e técnicas nas análises das ocorrências dos casos de dengue e na compreensão do cenário do padrão de distribuição espacial.

Os mapas favorecem uma rápida e simples visualização do cenário espacial demonstrando algumas explicações sobre a dengue a partir do estudo que avalia a relação entre a doença e determinantes sociais. No entanto, visualmente não é possível ter certeza se existe autocorrelação espacial nas variáveis utilizadas. A estatística espacial é aplicada nos estudos de saúde para modelar a realidade. Os modelos estatísticos usados para estudar saúde e os fatores relacionados a ela são úteis na análise espacial para entendimento, estimativa e modelagem de como o contexto derivado de fatores socioeconômicos afeta a saúde das pessoas. Com a análise

estatística espacial foi possível verificar que existe autocorrelação positiva para todas as variáveis utilizadas. Em contra partida, a correlação bivariada indicou que as variáveis escolhidas não expressam uma correlação positiva considerável com a variável de casos de dengue. Os resultados não apontaram nenhuma variável macro determinante que mais contribui para a propagação da doença, ao contrário do estudo de Resendes et al. (2010).

As técnicas utilizadas mostraram-se úteis na identificação de agrupamentos (clusters), de altos e baixos valores e também áreas de transição entre os clusters. O trabalho permitiu concluir que o conjunto de ferramentas usadas fornece, de fato, a possibilidade de explorar os dados espaciais, extraindo informações adicionais não diretamente perceptíveis, como análise visual ou as técnicas de classificação.

Este estudo difere de estudos anteriores realizados em outras localidades, tanto por aspectos de sua metodologia, com abordagens aprofundadas, quanto pelos resultados. Contradiz os achados de boa parte das pesquisas sobre a correlação de indicadores socioeconômicos e ambientais com os casos de dengue, por apontar uma associação inversa.

Apesar dos inúmeros estudos sobre a correlação entre a dengue e indicadores sociais que confirmam a existência de correlação, os resultados dos modelos de regressão linear e espacial mostram como as variáveis podem apresentar comportamentos bastante diferente para a modelagem estatística de acordo com o uso de métodos lineares convencionais.

A análise das correlações entre os indicadores e os casos de dengue, mostrou que os casos de dengue apresentam baixas associações estatísticas com as variáveis escolhidas, o que indica que as variáveis escolhidas não explicam a dengue em Aparecida de Goiânia. Seria o caso de tentar outras variáveis ou com dados mais atualizados, uma vez que oito anos separam os dados das variáveis socioeconômicas e ambientais e os casos de dengue, muita coisa pode ter mudado nesse meio tempo e isso pode ter contribuído para os resultados do trabalho.

Por fim, não esgotando este assunto, recomenda-se que este estudo seja refeito com dados mais atualizados e com outras variáveis que poderão delinear ainda mais as disparidades socioeconômicas verificada nos setores censitários e talvez seu contexto histórico, ambiental e social possam descrever ou explicar os processos que ocorrem nesse espaço em relação à dengue.

6. REFERÊNCIAS

ALMEIDA, E. **Econometria espacial aplicada**. Campinas–SP. Alínea, 2012.

ALMEIDA, E. S. **Curso de econometria espacial aplicada**. Piracicaba: Esalq-USP,. 127 p. 2004

ALVES, William Pereira. **Fundamentos de Banco de Dados**. São Paulo: Érica, 2004.

ALMEIDA MCM. **Distribuição espacial de casos notificados de dengue em Belo Horizonte, Minas Gerais**, 1996–2002: Identificação de conglomerados e fatores de risco associados [dissertação]. Minas Gerais: Universidade Federal de Minas Gerais; 2004

ANDRADE, Ana Lúcia et al. **Introdução à estatística espacial para a saúde pública**. Brasília: ministério da Saúde, 2007.

ANSELIN, L; BAO, S. **Exploratory Spatial Data Analysis Linking SpaceStat and ArcView**. In: Fischer, M. M. & Getis, A. Recent developments in spatial analysis. New York: Springer, 1997, p. 35-59.

ANSELIN L.; SYABRI, I.; SMIRNOV, O. **Visualizing multivariate spatial correlation with dynamically linked windows**. Urbana-Champaign: Spatial Analysis Laboratory, Department of Agricultural and Consumer Economics, University of Illinois, 2004. 13 p. Disponível em: <<http://www.real.illinois.edu/d-paper/01/01-t-10.pdf>>. Acesso em: 14 dez. 2019.

ANTUNES, Valdir. **Prefeitura de Aparecida trabalha para ampliar acesso ao saneamento básico**. Aparecida de Goiânia, 9 maio 2017. Disponível em: <http://www.aparecida.go.gov.br/prefeitura-de-aparecida-trabalha-para-ampliar-acesso-ao-saneamento-basico/>. Acesso em: 21 de maio de 2019.

ARAÚJO, Gabriel. **Em Aparecida de Goiânia 25% da população têm esgoto canalizado**. 2018. O hoje.com. Disponível em: <<http://www.ohoje.com/noticia/cidades/n/155209/t/em-aparecida-de-goiania-25-da-populacao-tem-esgoto-canalizado>>. Acesso em: 15 set. 2019.

ARAÚJO, José Raimundo de; FERREIRA, Efigênia Ferreira; ABREU, Mauro Henrique Nogueira Guimarães de. **Revisão sistemática sobre estudos de espacialização da dengue no Brasil**. Revista Brasileira de Epidemiologia, v. 11, p. 696-708, 2008. Disponível em: <http://www.scielo.br/pdf/rbepid/v11n4/15>. Acesso em: 6 de maio de 2019.

BAILEY, T. C., GATRELL, A. C. **Interactive spatial data analysis**. Essex: Longman Scientific and Technical, 1995

BARBOSA, Gerson Laurindo; LOURENÇO, Roberto Wagner. **Análise da distribuição espaço-temporal de dengue e da infestação larvária no município de Tupã, Estado de São Paulo**. Revista da Sociedade Brasileira de Medicina Tropical. Sociedade Brasileira de Medicina Tropical - SBMT, v. 43, n. 2, p. 145-151, 2010. Disponível em: <http://www.scielo.br/pdf/rsbmt/v43n2/08.pdf>. Acesso em: 16 maio 2019.

BARBOSA, I. R.; SILVA, L. P. da. **Influência dos determinantes sociais e ambientais na distribuição espacial da dengue no município de Natal-RN**. Revista Ciência Plural, v. 1, n. 3, p. 62-75, 2 fev. 2016.

BARCELLOS, C.; PUSTAI, A.K.; WEBER, M.A.; Brito, M.R.V. **Identificação de locais com potencial de transmissão de dengue em Porto Alegre através de técnicas de geoprocessamento**. Rev Soc Bras Med Trop. 2005;38(3):246–50.

BARCELLOS, C.; SABROZA, P.C.; PEITER, P.; ROJAS, L.I. (2002) **Organização espacial, saúde e qualidade de vida: A análise espacial e o uso de indicadores na avaliação de situações de saúde**. Informe Epidemiológico do SUS, 11(3): 129-138

BARCELLOS, C. RAMALHO, W.M. **Situação atual do geoprocessamento e da análise de dados espaciais em saúde no Brasil**. Revista de Informática Pública 2002;4(2):221-230.

BARRERA, R. DELGADO, N. JIMÉNEZ, M. VILLALOBOS, I. ROMERO, Y. **Estratificación de una ciudad hiperendémica en dengue hemorrágico**. Ver Panam Salud Publica. 2000;8(4):225–33.

BARRETO, Maurício L.; TEIXEIRA, Maria Glória. **Dengue in Brazil: epidemiological situation and contribution to a research agenda**. Estudos Avançados 22, nº. 64 (2008): 53-72. Bases Geográficas. Disponível em: http://www.scielo.br/pdf/ea/v22n64/en_a05v2264.pdf. Acesso em 21 de maio de 2019.

BIO-MANGUINHOS. **Dengue: sintomas, transmissão e prevenção**. Rio de Janeiro-RJ, 2014. Disponível em: <https://www.bio.fiocruz.br/index.php/dengue-sintomas-transmissao-e-prevencao>. Acesso em: 16 maio 2019.

BORDE, Andréa de Lacerda Pessoa. **Percorrendo os vazios urbanos**. In: Simpósio Perspectivas da forma urbanística no século XXI. Cadernos de resumos. Florianópolis, 2003.

BORGES, K. A. V.; DAVIS JÚNIOR, C. A.; LAENDER, A. H. F. **Modelagem conceitual de dados geográficos**. In: CASANOVA, M. A. et al. (org.) Bancos de dados geográficos. Curitiba: MundoGEO, 2005. p. 93-146.

BRASIL. **Lei nº 11.445**, de 5 de janeiro de 2007. Diário Oficial da União 2007. 8 de janeiro:3. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2007/Lei/L11445.htm. Acesso em: 16 de maio de 2019

BRIGGS, D.J.; ELLIOT, P., 1995. **The use of geographical information systems in studies on environment and health.** *Wld hlth quart*, 48: 85-94.

BRUSSO, M. J. **Access Miner: Uma proposta para a Extração de Regras de Associação Aplicada à Mineração do Uso da Web.** Master's thesis, PPGC da UFRGS, Porto Alegre - RS, 2000.

CÂMARA, G. et al. **Anatomia de Sistemas de Informação Geográfica.** Campinas: Instituto de Computação, UNICAMP, 1996. (X Escola de Computação).

CÂMARA G, CARVALHO MS, CRUZ OG, CORREIA V. **Análise espacial de áreas.** In: Druck S, Carvalho MS, Câmara G, Monteiro AMV, editores. *Análise espacial de dados geográficos [monografia on-line].* São Paulo: Instituto Nacional de Pesquisas Espaciais; 2002. p. 1-32. Disponível em URL: <http://www.dpi.inpe.br/gilberto/livro/analise/cap5-areas.pdf> Acesso em: 16 de maio de 2019.

CAMILO, Cássio Oliveira; SILVA, João Carlos da. **Mineração de Dados: Conceitos, Tarefas, Métodos e Ferramentas.** Instituto de Informática. Universidade Federal de Goiás, 2009. Disponível em: <http://www.inf.ufg.br/sites/default/files/uploads/relatoriostecnicos/RT-INF_001-09.pdf>. Acesso em: 11 jul. 2019.

CARVALHO, M. S.; PINA, M. F. & SANTOS, S. M., 2000. **Conceitos Básicos de Sistemas de Informações Geográficas e Cartografia Aplicados à Saúde.** Brasília: Organização Panamericana de Saúde/Ministério da Saúde.

CAVICCHIOLI NETO, Virgílio; CHIARI, Natália Santana; CARVALHO, Isabelle; PISA, Ivan Torres; ALVES, Domingos. **Desenvolvimento e integração de mapas dinâmicos georreferenciados para o gerenciamento e vigilância em saúde.** In: Anais do 11^a Congresso Brasileiro de Informática em Saúde; São Paulo, Brasil. São Paulo: Sociedade Brasileira de Informática em Saúde; 2008 Disponível em <http://www.sbis.org.br/cbis11/arquivos/1036.pdf> Acesso em: 6 de outubro de 2019.

CHRISTOPHERS, S. R. *Aedes aegypti* (L.) **The yellow fever mosquito.** 1. ed. Cambridge: The Syndics of the Cambridge University Press, 1960.

COSTA, A. I. P. da; NATAL, D. **Distribuição espacial da dengue e determinantes socioeconômicos em localidade urbana no Sudeste do Brasil.** *Revista Saúde Pública.* vol. 32, n. 3. São Paulo, 1998. p.232-236.

CRESSIE, N. **Statistics for Spatial Data** Chichester, John Wiley. 1992.

DATOS MARKETING. **Lista de classes sociais IBGE**. São Paulo, [2012]. Disponível em: <http://www.datosmarketing.com.br/listas-detahes-classes-sociais.asp>. Acesso em: 2 nov. 2019.

DAVIS, Clodoveu; CÂMARA, Gilberto. **Arquitetura de Sistema de Informação Geografica**. In: CÂMARA, Gilberto; DAVIS, Clodoveu; MONTEIRO, Antônio Miguel Vieira. Introdução à ciência da geoinformação. São José dos Campos: INPE, 2001. E-book.

DRUCK, S.; CARVALHO, M.S.; CÂMARA, G.; MONTEIRO, A.V.M. (eds) "**Análise Espacial de Dados Geográficos**". Brasília, EMBRAPA, 2004 (ISBN: 85-7383-260-6).

ELMASRI, Ramez; NAVATHE, Shamkant B. **Sistemas de Banco de Dados**. São Paulo: Pearson Addison Wesley, 2011.

FACELI, Katti et al. **Inteligência Artificial: Uma abordagem de aprendizado de máquina**. Rio de Janeiro: LTC, 2015.

FAYYAD, Usama; Piatetsky-Shapiro, G; Smyth, P. **Data mining to Knowledge Discovery in Databases**. American Association for Artificial Intelligence, 1996.

FURLAN, Luciana Benzoni; CAVICCHIOLI NETO, Virgilio; BOUSQUAT, Aylene; SILVA, Paulo Sérgio de Andrade, PISA, Ivan Torres, ALVES, Domingos. **Desenvolvimento de um sistema para o monitoramento e análise da mortalidade infantil na região metropolitana da baixada santista**. Epidemial Serv. Saúde, Brasília, V. 20, n3, p. 287-296. 2009. Disponível em: Disponível em: http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_004-04.pdf. Acesso em: 6 de outubro de 2019.

FERREIRA, Aline Chimello. **Identificação de fatores de risco para infestação por Aedes aegypti e estratificação de áreas e domicílios, São José do Rio Preto, SP**. Orientador: Dr. Francisco Chiaravalloti Neto. 2006. 71 p. Dissertação (Mestre em Ciências) - Coordenadoria de Controle de Doenças da Secretaria de Estado da Saúde de São Paulo, São Paulo, 2006.

FREITAS, Rafael. **Aparecida é a 17ª maior cidade do país, exceto as capitais, segundo IBGE**. In: Prefeitura de Aparecida de Goiânia. Aparecida de Goiânia, 28 ago. 2019. Disponível em: <http://www.aparecida.go.gov.br/aparecida-e-a-17a-maior-cidade-do-pais-exceto-as-capitais-segundo-ibge/>. Acesso em: 20 nov. 2019.

GOODCHILD, M., ANSELIN, L., APPELBAUM, R. AND HARTHORN, B. **Towards spatially integrated social science**. International Regional Science Review 23, 139-159. 2000.

GOLDSCHMIDT, Ronaldo; PASSOS, Emmanuel. **Data Mining Um guia prático**. Rio de Janeiro: Campus, 2005.

Gonçalves Neto VS, Rebelo JM. **Aspectos epidemiológicos do dengue no Município de São Luís, Maranhão, Brasil**, 1997–2002. Cad Saude Publica. 2004;20(5):1427–31.

GREGORY, L. M. R. de ; PRETTO, C. L. de. **Aplicação de Regras de Associação para Mineração de Dados na Web**. Technical Report, Goiás, 2004. Disponível em: http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_004-04.pdf.

Acesso em: 6 de maio de 2019.

Instituto Brasileiro de Geografia e Estatística. **Censo 2010**. Disponível em: <https://censo2010.ibge.gov.br/>. Acesso em: 6 de maio de 2019.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATISTICA (Aparecida de Goiânia). **Aparecida de Goiânia Panorama**. 2018. IBGE. Disponível em: <<https://cidades.ibge.gov.br/brasil/go/aparecida-de-goiania/panorama>>. Acesso em: 25 set. 2019.

INSTITUTO TRATA BRASIL. **Ranking do Saneamento Instituto Trata Brasil 2016**. São Paulo. GO Associados/Instituto Trata Brasil. 2016. Disponível em: <http://www.tratabrasil.org.br/datafiles/estudos/ranking/2016/relatorio-completo.pdf> Acesso em: 6 de maio de 2019.

JÁCOMO, C.A.; TACHIBANA, V.M.; IMAI, N.N.; FLORES, E.F. 2010. **Aplicação de técnicas de estatística espacial na caracterização dos casos de dengue no município de Presidente Prudente-SP no período de 1999 a 2007**. In: III SIMPÓSIO DE CIÊNCIAS GEODÉSICAS E TECNOLOGIAS DA GEOINFORMAÇÃO, Recife, Anais, p.1-9.

JESUS, Silvia Cristina de; MIURA, Adalberto Koiti. **Análise espacial de fragmentos de Cerrado como subsídio para a seleção de áreas prioritárias para conservação**. Anais XIV Simpósio Brasileiro de Sensoriamento Remoto, Natal - RN, 25-30 de abril 2009. Anais XIV Simpósio Brasileiro de Sensoriamento Remoto, Natal, Brasil, 25-30 abril 2009, INPE, p. 3943-3950.

KING, P.E. **Problems os spatial analysis in geographical edpidemiology**. Social Science & Medicine 13: 249-252, 1979.

LEITE, Marcos Esdras. **Análise da correlação entre dengue e indicadores sociais a partir do SIG**. HYGEIA, Revista Brasileira de Geografia Médica e da Saúde, [S. l.], 2010

LEGENDRE, P. **Spatial autocorrelation: trouble or new paradigm?** Ecology, v.74, n.6, p.1659-1673, 1993.

LEONETI, Alexandre Bevilacqua; PRADO, Eliana Leão do; OLIVEIRA, Sonia Valle Walter Borges de. **Saneamento básico no Brasil: considerações sobre investimentos e sustentabilidade para o século XXI**. Revista de Administração Pública, Rio de Janeiro, p. 332-348, 2011. rap — rio de Janeiro 45(2):331-48, mar./abr. 2011.

LISBOA FILHO, J.; IOCHPE, C. **Um estudo sobre modelos conceituais de dados para projeto de banco de daos geograficos**. Revista IP-Informatica Pública, Belo Horizonte, v.1, n.2, 1999.

LOUREIRO, Carlos Felipe Grangeiro; SILVA, Henrique Nogueira; CARVALHO, Luís Eduardo Ximenes. **Metodologia de análise de regressão geograficamente ponderada aplicada ao fenômeno das viagens intermunicipais**. In: *Anais do XX Congresso de Pesquisa e Ensino em Transportes*. 2006. p. 479-491.

Machado JP. **Dengue e condições de vida no município de Nova Iguaçu: uma abordagem espacial** [dissertação]. Rio de Janeiro: Escola Nacional de Saúde Pública, Fundação Oswaldo Cruz; 2007.

MARTINS, Polliana. **Unidades Básicas de Saúde de Aparecida atendem aos casos suspeitos de dengue**. Aparecida de Goiânia, 29 maio 2019. Disponível em: <http://saude.aparecida.go.gov.br/unidades-basicas-de-saude-de-aparecida-atendem-aos-casos-suspeitos-de-dengue/>. Acesso em: 28 dez. 2019.

MILLER, H. J. **Potential contribution of spatial analysis to Geographic Information System for Transpotation (GIS-T)**. *Geographical Analysis*, 31, 373-399. 1999

MINISTÉRIO DA SAÚDE. **Dengue Aspectos Epidemiológicos, Ministério da Saúde Diagnóstico e Tratamento**. Brasília: Fundação Nacional de Saúde, 2002.

MINISTÉRIO DA SAÚDE (Brasil). Secretária de Vigilância em Saúde. Secretária de Atenção à saúde. **O agente Comunitário de Saúde no controle da dengue**. 1. ed. Brasília/DF: Ministério da Saúde, 2009. 36 p. Disponível em: http://bvsmis.saude.gov.br/bvs/publicacoes/agente_comunitario_saude_controle_dengue.pdf. Acesso em: 28 dez. 2019.

MORDINI, Adriano; CHUARAVALLOTTI NETO, Francisco. **Variáveis socioeconômicas e a transmissão da dengue**. *Rev. Saúde Pública* vol.41 no.6 São Paulo Dec. 2007. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-89102007000600006 Acesso em: 25 de outubro de 2019.

MOURA, A. C. M.. **Geoprocessamento na gestão e planejamento urbano**. Belo Horizonte : Ed. Da autora, 2003, 294p

OLIVEIRA, G.; SCAZUFCA, P.; PIRES, R. C. **Ranking do saneamento**. São Paulo: INSTITUTO TRATA BRASIL, 2018.

OPAS - Organização Pan-Americana de Saúde. **Iniciativa Regional de Datos Básicos em Saúde**. 2010 Disponível em: https://www.paho.org/bra.../index.php?option=com_content&view=article&id=1050:cinco-estados-concentram-70percent-dos-casos-de-dengue-neste-ano&Itemid=463. Acesso em 22 de maio de 2019.

Paulino AIC, Natal D. **Distribuição espacial da dengue e determinantes sócio-econômicos em uma localidade urbana no sudeste do Brasil**. *Rev Saude Publica*. 1998;32(3):232–36.

PREFEITURA DE APARECIDA DE GOIÂNIA. **Unidades de Atendimento**. Aparecida de Goiânia, 2019. Disponível em: <http://saude.aparecida.go.gov.br/unidades-de-atendimento/>. Acesso em: 28 dez. 2019.

RESENDES, A. P. C; SILVEIRA, N. A. P. R.; SABROZA, P. C.; SOUZA-SANTOS R. **Determinação de áreas prioritárias para ações de controle da dengue**. Rev Saude Publica. 2010 abr;44(2):274-82.

REVELLE, Charles; HOGAN, Kathleen, 1989. **The Maximum Availability Location Problem**, Transportation Science, INFORMS, vol. 23(3), pages 192-200, August. Disponível em: <https://ideas.repec.org/a/inm/ortrsc/v23y1989i3p192-200.html> Acesso em 29 de dezembro de 2019

SAIANI, C. C. S. **Restrições à expansão dos investimentos em saneamento básico no Brasil: déficit de acesso e desempenho dos prestadores**. 2007. 315 f. Dissertação (Mestrado em Economia Aplicada) - Faculdade de Economia, Administração e Contabilidade de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2007.

Santos A. **Geografia da dengue em Uberlândia (MG) na epidemia de 1999**. Caminhos de Geografia. 2004;3(11)35–52

SANTOS, S. M.; SOUZA, W. V. **Introdução à estatística espacial para a saúde pública**. Brasília: Ministério da Saúde/Fundação Oswaldo Cruz; 2007. (Série Capacitação e Atualização em Geoprocessamento em Saúde, 3).

SANTOS, M., L. Amaral, e P. Pimenta, **A Descoberta de Conhecimento em Bases de Dados Geográficas através da Explicitação Semântica**, GISBrasil'99 - V Congress and Exhibition of Latin America Geo-processing Users, Salvador, Brasil, 19-23 July, 1999. Disponível em: <http://repositorium.sdum.uminho.pt/handle/1822/5598>. Acesso em: 6 de dezembro de 2019.

SANTOS, Marcelo Silva ; NETO, José Craveiro da Costa. **AMAGODIS: Algoritmos de Mineração para Apoio à Gerência de Ocorrências de Dengue a partir de Informações presentes na base dados do SINAN, Ilhéus -BA**, 2011. Disponível em: https://www.enacomp.com.br/2011/anais/trabalhos-aprovados/pdf/enacomp2011_submission_55.pdf. Acesso em: 6 de maio de 2019.

SES–GO. **Secretaria de estado da saúde de goiás. Superintendência de vigilância em saúde gerência de vigilância epidemiológica boletim semanal de dengue - Goiás 2019 Semana Epidemiológica 1 a 23 (30/12/2018 a 08/06/2019)**. Goiânia: SES-GO, 2019.

Sobral, Marcos Felipe Falcão, Sobral, Ana Iza Gomes da Penha. **Casos de dengue e coleta de lixo urbano: Um estudo na Cidade do Recife..** Cien Saude Colet [periódico na internet] (2017/Mai). Está disponível em: <http://www.cienciaesaudecoletiva.com.br/artigos/casos-de-dengue-e-coleta-de-lixo->

[urbano-um-estudo-na-cidade-do-recife/16258?id=16258&id=16258](http://www.scielo.br/urbanismo-um-estudo-na-cidade-do-recife/16258?id=16258&id=16258) Acesso em: 23 de dezembro de 2019.

TAUIL, Pedro Luiz. **Urbanização e ecologia do dengue**. Cad. Saúde Pública [online]. 2001, vol.17, suppl., pp.S99-S102. ISSN 0102-311X. Disponível em: <http://www.scielo.br/pdf/csp/v17s0/3885.pdf> Acesso em: 6 de maio de 2019.

WANDERLEY, Maria Fernanda Barbosa. **Estudos em Estimação de Densidade por Kernel: Métodos de Seleção de Características e Estimação do Parâmetro Suavizador**. 2013. 15 p. Dissertação (PROGRAMA DE PÓS GRADUAÇÃO EM ENGENHARIA ELÉTRICA)-UNIVERSIDADE FEDERAL DE MINAS GERAIS ESCOLA DE ENGENHARIA, Belo Horizonte, 2013. Disponível em: <https://www.ppgee.ufmg.br/defesas/277D.PDF>. Acesso em: 06 de maio de 2019.

WONG, David. **The modifiable areal unit problem (MAUP)**. The SAGE handbook of spatial analysis, v. 105, n. 23, p. 2, 2009.