

Análise espacial da leptospirose nos municípios do Rio de Janeiro

Raíssa Caroline dos Santos Teixeira¹

¹Instituto Nacional de Pesquisas Espaciais – INPE
Caixa Postal 515 - 12227-010 - São José dos Campos - SP, Brasil
raissa.teixeira@inpe.br

Abstract: Leptospirosis is a disease that occurs almost worldwide and displays many routes of exposure and reservoirs. Climate change may affect the number of cases, but socioeconomic and environmental context are important factors to understand its behavior. In this study, an exploratory spatial analysis was done with leptospirosis number of cases and incidence, and sanitation factors, typically explored to characterize the risk of transmission. A Geographic Information System was used for mapping the leptospirosis incidence in the geographic space, and the disease data with sanitation variables were spatially and statistically analyzed in GeoDA. The multiple linear regression did not present good explanation of the variables and areas with higher incidence showed a heterogenous spatialization along the state.

Palavras-chave: Análise espacial, leptospirose.

1. Introdução

A água pode ser veículo de doenças de forma direta e indireta, conhecidas como doenças de veiculação hídrica. De acordo com o Ministério da Saúde em sua Diretriz para atuação em situações de surtos de doenças e agravos de veiculação hídrica (Brasil, 2018), as mudanças no perfil epidemiológico das doenças de transmissão hídrica podem ser influenciadas por diversos fatores, como a distribuição de água não tratada ou em desconformidade com o padrão de potabilidade para a população.

Entre diversas doenças existentes no país, existe um grupo classificado como doenças de veiculação hídrica, que de acordo com o Ministério da Saúde, ocorrem porque a água pode veicular substâncias químicas e agentes biológicos nocivos à saúde da população, podendo adentrar o organismo humano de diferentes formas, seja por meio da ingestão, inalação, pelo contato da água contaminada com a pele ou mucosas ou por alimentos lavados com água contaminada que pode causar danos à integridade física ou mental, ou até mesmo o adoecimento do indivíduo (Brasil, 2018). Incluída nesse grupo de doenças, está a Leptospirose.

A leptospirose é uma doença endêmica e de notificação compulsória no Brasil, e isso significa que sua comunicação é obrigatória à autoridade de saúde, realizada pelos médicos, profissionais de saúde ou responsáveis pelos estabelecimentos de saúde, públicos ou privados, sobre a ocorrência de suspeita ou confirmação de doença, agravo ou evento de saúde pública. Segundo Hino et al. (2006), a análise da distribuição de doenças e seus determinantes nas populações, no espaço e no tempo é um aspecto fundamental da Epidemiologia. De acordo com o Guia de Vigilância Epidemiológica do Ministério da Saúde, a notificação é a comunicação da ocorrência de determinada doença ou agravo à saúde, feita à autoridade sanitária por profissionais de saúde ou qualquer cidadão, para fins de adoção de medidas de intervenção pertinentes, sendo que essas informações abastecem o Sistema de Informação de Agravos de Notificação - Sinan, mediante formulários padronizados, e podem ser consultados no Departamento de Informática do Sistema Único de Saúde (DATASUS).

A leptospirose é uma doença bacteriana que afeta humanos e animais, que podem ser infectados pelo contato direto com a urina de animais contaminados ou com um ambiente contaminado pela urina (World Health Organization, 2019). As bactérias entram no corpo através de cortes ou abrasões na pele ou através das membranas mucosas da boca, nariz e olhos, sendo que a exposição através da água contaminada pela urina de animais infectados é a via mais comum de infecção e a transmissão de pessoa para pessoa é rara (World Health Organization, 2019). A leptospirose é uma doença infecciosa febril de início abrupto, cujo espectro pode variar desde um processo assintomático até formas sintomáticas graves (Brasil, 2009). É uma doença transmitida ao homem pela urina de roedores, principalmente por ocasião das enchentes, e é causada por uma bactéria chamada *Leptospira*, presente na urina de ratos e outros animais (Brasil, 2018).

Trata-se de uma doença de veiculação hídrica, considerada complexa porque pode ter sua ocorrência afetada por diversos fatores como variáveis climáticas, por exemplo o regime de chuvas considerando que inundações propiciam sua disseminação, assim como fatores socioeconômicos que envolvem a população como um todo, qualidade de vida, saneamento básico, acesso a água tratada, entre outros. A leptospirose trata-se de uma zoonose de grande importância social e econômica, porque apresenta elevada incidência em determinadas áreas, causa altos custos hospitalares e perdas de dias de trabalho, e, além disso, sua letalidade pode chegar a 40%, nos casos mais graves (Brasil, 2009). Nos estágios iniciais da doença, os sintomas incluem febre alta, dor de cabeça intensa, dor muscular, calafrios, vermelhidão dos olhos, dor abdominal, icterícia, hemorragias na pele e membranas mucosas, vômitos, diarreia e erupção cutânea (World Health Organization, 2019). A figura 1 apresenta o ciclo de transmissão da leptospirose e os agentes e vetores envolvidos no contágio da doença.

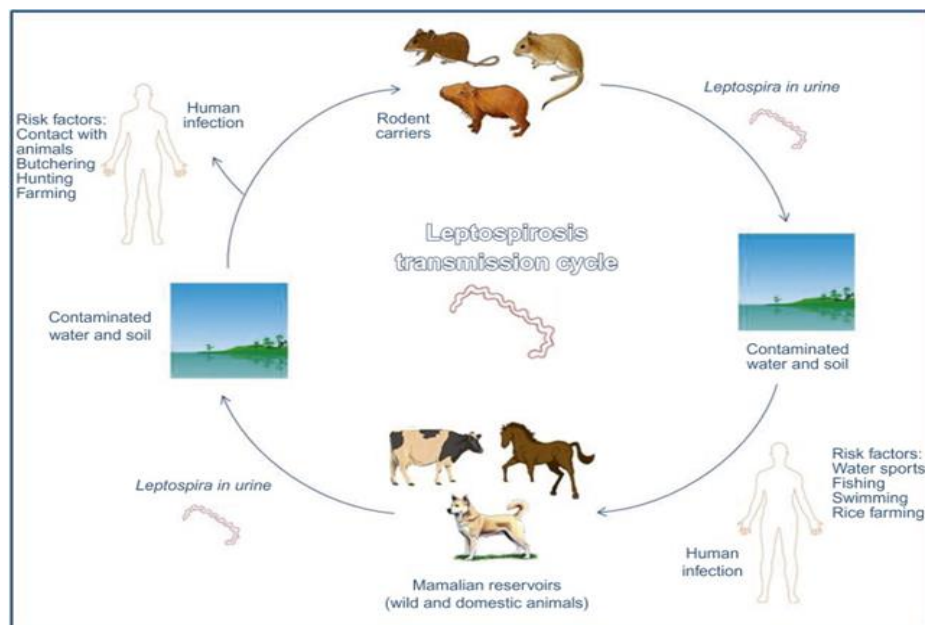


Figura 1. Ciclo da leptospirose. Fonte: Molecular Medical Microbiology (2015).

A leptospirose apresenta distribuição universal e, no Brasil, é uma doença endêmica que torna-se epidêmica em períodos chuvosos, principalmente nas capitais e áreas metropolitanas, devido às enchentes associadas a aglomerações da população de baixa renda, às condições inadequadas de saneamento e à alta infestação de roedores infectados (Brasil, 2009). As formas de transmissão das doenças de veiculação hídrica em escalas endêmicas e epidêmicas devem-se

principalmente ao acesso à água, infraestruturas inadequadas de saneamento e drenagem, e falta de acesso a cuidados de saúde agravados por desastres naturais ou convulsões sociais (Hasan et al., 2018), ou seja, é uma doença que pode ser associada a vários fatores, que podem incluir o clima, condições de saneamento básico, coleta de resíduos domiciliares, relevo da região, condição socioeconômica da população, condições de drenagem urbana, entre outros. A quantidade de casos de leptospirose notificados a cada ano varia de região para região, sendo que o ano de 2013 foi o ano em que houve maior número de notificações na região sudeste do Brasil comparada a outras regiões, como pode ser observado na figura 2.

Mesmo que o número de casos tenha diminuído ao longo dos anos, a leptospirose ainda é uma doença preocupante, e o estudo de seu comportamento e áreas de incidência é importante para a tomada de ações pertinentes de saúde pública. Entender como a ocorrência da doença pode ser influenciada, considerando anos anteriores e condições passadas dos municípios, pode ser uma forma de orientação para estudos futuros. Considerando todos esses fatores que podem estar envolvidos com a doença, a região de estudo selecionada foi o estado do Rio de Janeiro, com dados agregados a nível municipal no ano de 2013.

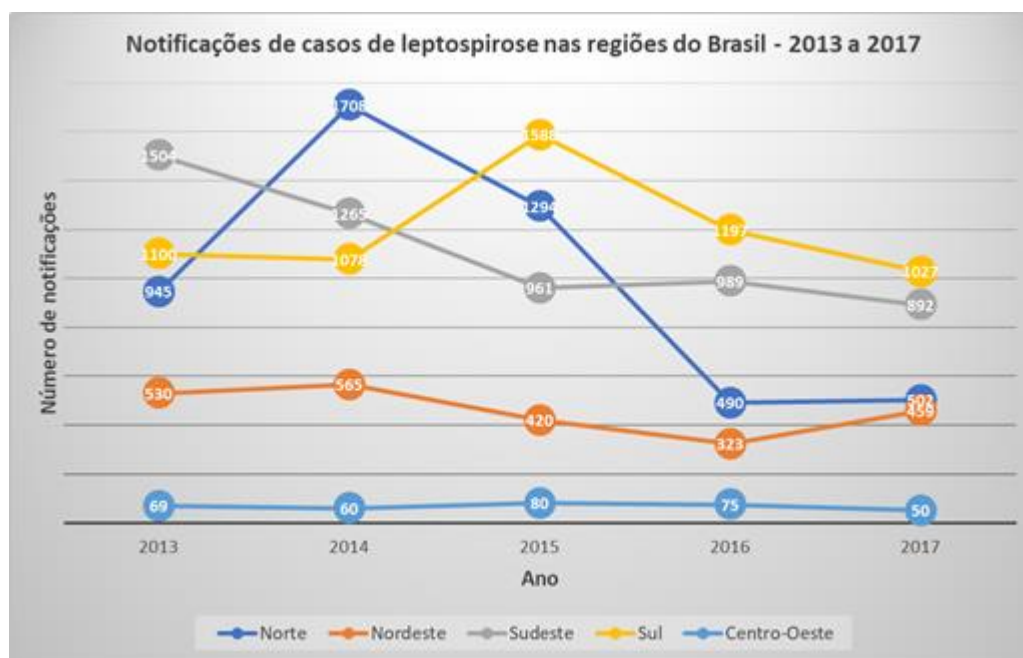


Figura 2. Número de casos confirmados de leptospirose no Brasil por região e ano.
Fonte de dados: Ministério da Saúde/SVS - Sistema de Informação de Agravos de Notificação – Sinan.

O mapeamento de doenças consiste na descrição do processo de distribuição espacial, visando avaliar a variação geográfica na sua ocorrência para identificar diferenciais de risco, orientar a alocação de recursos e levantar hipóteses etiológicas (Andrade et al., 2007). O presente trabalho teve como objetivos principais descrever a distribuição espacial dos casos de leptospirose no estado do RJ e a incidência da doença e avaliar se existe autocorrelação espacial da incidência nos diferentes municípios. Além disso, buscou-se relacionar dados de saneamento básico como tratamento de esgoto e coleta de resíduos nos municípios com a incidência da doença através de regressão linear múltipla e avaliar se diferentes matrizes de vizinhança promovem resultados diferentes para análise de autocorrelação espacial.

2. Metodologia

2.1. Área de estudo

A área de estudo ilustrada na figura 3, é o Estado do Rio de Janeiro, composto por 92 municípios. De acordo com o IBGE, o estado do Rio de Janeiro é composto por 6 mesorregiões: Noroeste Fluminense, Norte Fluminense, Centro Fluminense, Baixadas Litorâneas, Sul Fluminense e Metropolitana do Rio de Janeiro, sendo elas compostas por 18 microrregiões no total.

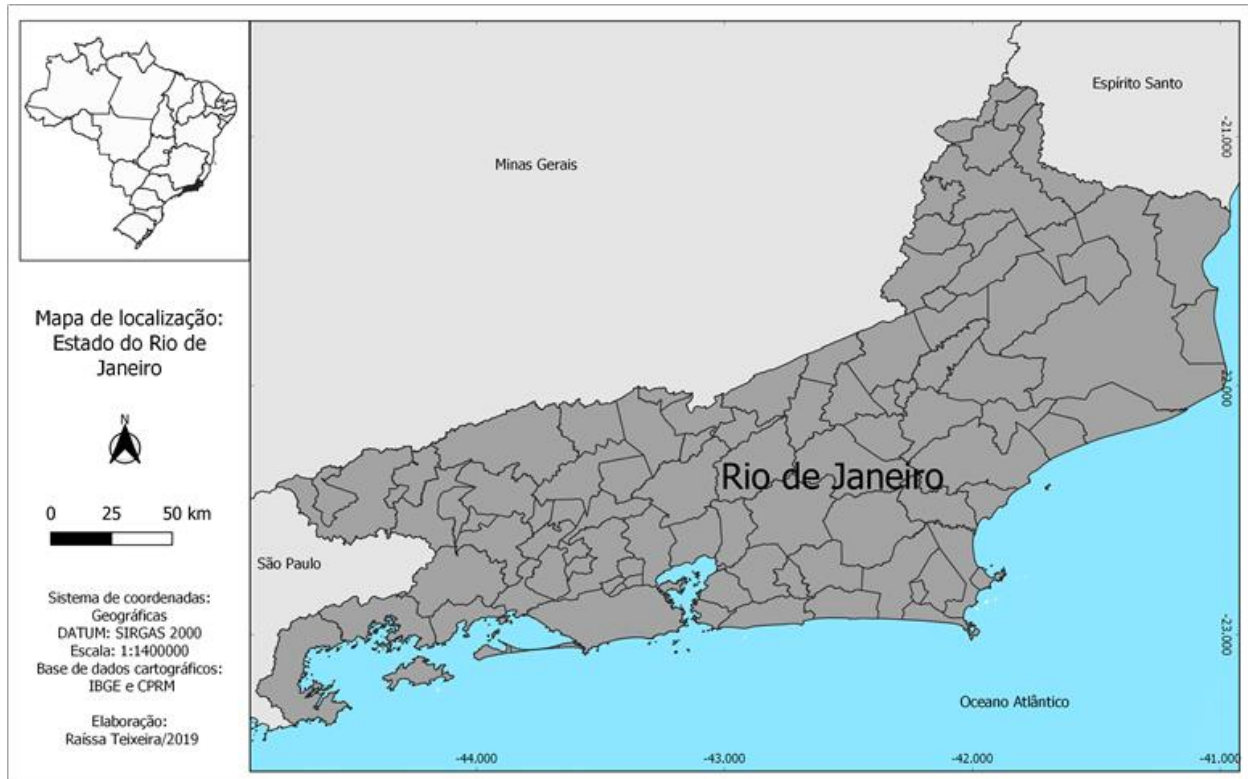


Figura 3. Área de estudo.

Grande parte da população do estado se concentra na Região metropolitana do Rio de Janeiro, também conhecida como Grande Rio, que é a segunda maior área metropolitana do Brasil atrás apenas da Grande São Paulo, e a terceira da América do Sul. Como pode ser observado pelo mapa, é um estado banhado pelo Oceano Atlântico e que faz divisa com três estados, sendo eles São Paulo, Minas Gerais e Espírito Santo, todos estados que constituem a região sudeste do Brasil.

2.2. Materiais e métodos

Os materiais e métodos foram determinados visando aplicar uma análise de autocorrelação espacial que visa indicar o grau de associação espacial presente no conjunto de dados, e posteriormente uma regressão linear múltipla, a qual é um método utilizado para avaliar a relação entre a variável de interesse (dependente) e variáveis exploratórias (independentes). O fluxograma das etapas pode ser visualizado na figura 4.

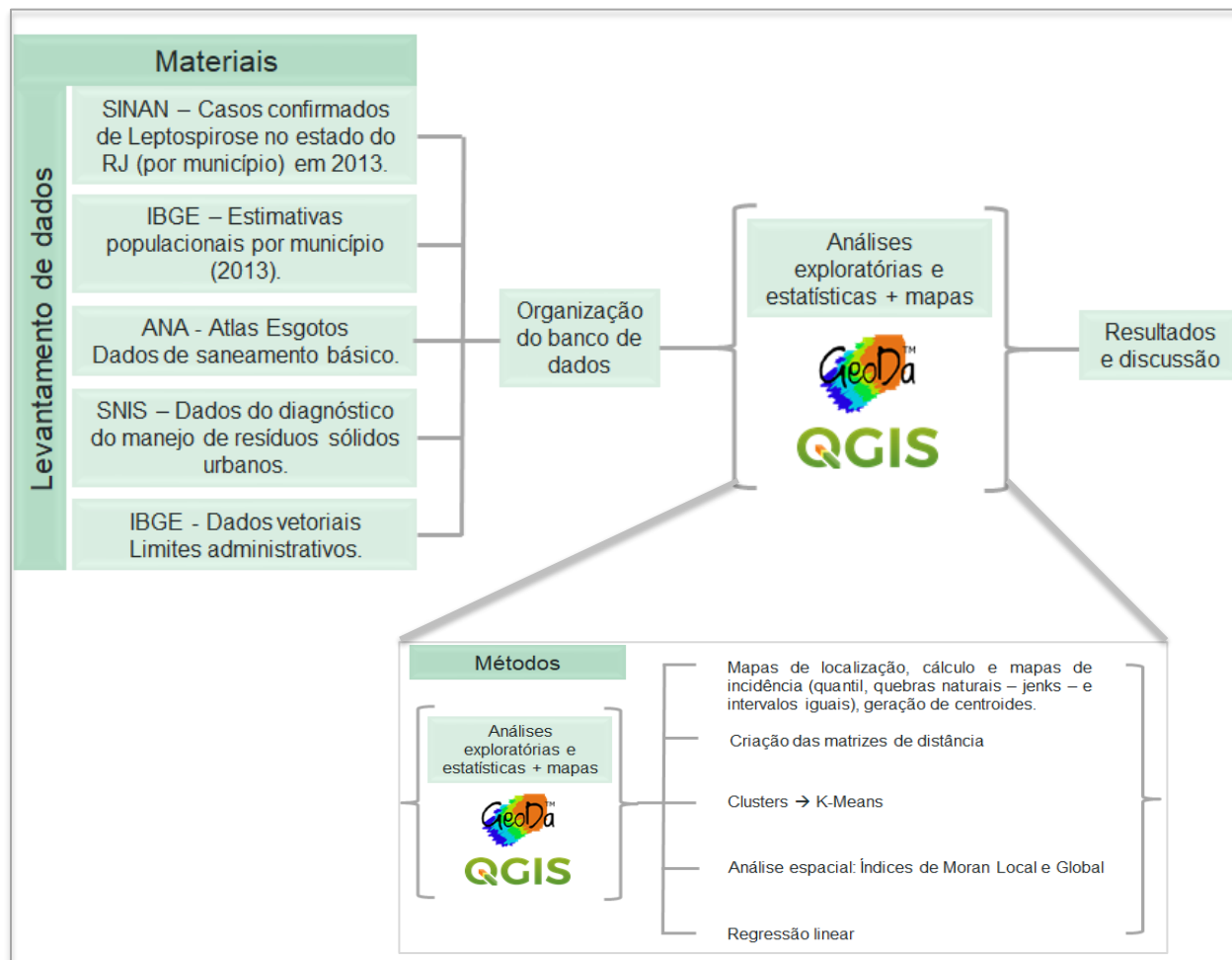


Figura 4. Fluxograma de trabalho.

Os dados de saúde utilizados foram os dados de notificação de leptospirose extraídos do DATASUS, o Departamento de Informática do Sistema Único de Saúde do Brasil. O DATASUS disponibiliza informações que podem servir para subsidiar análises objetivas da situação sanitária, tomadas de decisão baseadas em evidências e elaboração de programas de ações de saúde. São provenientes do Ministério da Saúde/SVS por meio do Sistema de Informação de Agravos de Notificação - Sinan, que tem como objetivo coletar, transmitir e disseminar dados gerados rotineiramente pelo Sistema de Vigilância Epidemiológica das três esferas de Governo, por meio de uma rede informatizada, para apoiar o processo de investigação e dar subsídios à análise das informações de vigilância epidemiológica das doenças de notificação compulsória. Os dados epidemiológicos obtidos foram os casos registrados como casos confirmados considerando a data do primeiro sintoma, assim como a localização utilizada por município de residência do paciente, e, após explorar os dados em anos diferentes para a mesma área de interesse, foi selecionado o ano de 2013 e os casos ocorridos nesse período, principalmente pelo número de casos ser relativamente alto, comparado a outros anos.

Para posterior cálculo da incidência foram extraídas as estimativas populacionais do Instituto Brasileiro de Geografia e Estatística - IBGE para o ano selecionado, além de obter da mesma fonte,

os dados vetoriais do Brasil e dos municípios, onde foi criado o shapefile de interesse através de operações de geoprocessamento utilizando o software QGIS.

Os dados de saneamento básico foram retirados do Atlas Esgotos, que foi publicado pela Agência Nacional de Águas (ANA) em parceria com a Secretaria Nacional de Saneamento Ambiental. É uma publicação referente a um estudo que apresenta o cenário atual, analisa dados e propõe ações e uma estratégia para investimentos em esgotamento sanitário com o horizonte de 2035. O Atlas Esgotos traz a análise da situação do esgotamento sanitário nas 5.570 cidades brasileiras e dos impactos do lançamento dos esgotos nos rios, lagos e reservatórios do País. Inclusive, o estado do Rio de Janeiro foi escolhida como área de estudo porque, além de ser um local com ocorrência relativamente alta de leptospirose, é um estado que possui uma média de índices relativos ao saneamento básico insatisfatórios, ou abaixo do esperado, como mostra a figura 5, onde menos de 50% do esgoto tem coleta e tratamento adequados. Para cada município que pode ser consultado no portal, o gráfico que é representado na figura se apresenta de forma diferente de acordo com suas porcentagens específicas dos índices.

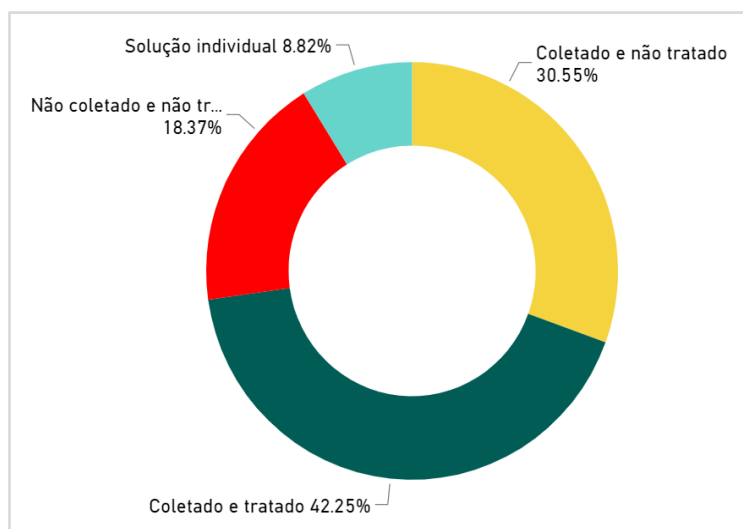


Figura 5. Índice de atendimento de esgoto - Estado do Rio de Janeiro (2013).

Fonte: Atlas Esgotos.

As variáveis escolhidas para serem utilizadas no trabalho foram o índice de coleta sem tratamento, índice de sem coleta e sem tratamento e a eficiência média de tratamento do município, esta última que é calculada a partir das cargas geradas e remanescentes de esgoto em cada local.

Já os dados sobre coleta de resíduos foram retirados do Sistema Nacional de Informações Sobre Saneamento - SNIS, que é o maior e mais importante sistema de informações do setor de saneamento brasileiro, pois possui uma base de dados que contém informações e indicadores sobre a prestação de serviços de Água e Esgotos, de Manejo de Resíduos Sólidos Urbanos e Drenagem e Manejo das Águas Pluviais Urbanas. Criado em 1996, o SNIS é uma unidade vinculada à Secretaria Nacional de Saneamento (SNS) do Ministério do Desenvolvimento Regional (MDR). Com abrangência nacional, reúne informações de caráter institucional, administrativo, operacional, gerencial, econômico-financeiro, contábil e de qualidade da prestação de serviços de saneamento básico em áreas urbanas das quatro componentes do saneamento básico. A tabela completa com os dados de um total de 64 municípios do Rio de Janeiro, declarantes no SNIS, foi retirada do Diagnóstico do manejo de Resíduos Sólidos Urbanos - 2013, disponibilizado no Painel

Resíduos Sólidos Urbanos. Entre uma variedade de dados, a variável escolhida foi a taxa de cobertura da coleta de resíduos sólidos domiciliares em relação à população total.

A escolha de utilizar essas variáveis de tratamento de esgoto e coleta de resíduos se deve principalmente ao fato de que a água e esgoto não tratados podem ser veículos de doenças de veiculação hídrica, como a leptospirose, e o acúmulo de resíduos em locais inapropriados pode causar a proliferação de vetores transmissores da doença, como os roedores. Na literatura são comumente encontrados trabalhos que se referenciam à coleta de resíduos, descarte de lixo incorreto ou adequada disposição dos resíduos como questões relacionadas à leptospirose (Gracie et al., 2014; Barcellos & Sabroza, 2000; Dias et al., 2007; Bacallao et al., 2014). Assim como também são utilizados diferentes dados referentes à questão sanitária dos municípios e existência do tratamento de esgoto como questões ligadas à leptospirose em outras muitas publicações pelo mundo (Gracie et al., 2014; Bacallao et al., 2014; Hagan et al., 2016; Barcellos e Sabroza, 2000; Ashford et al., 2000; Mwachui et al., 2015; Dias et al., 2007). A tabela 1 apresenta os dados que foram utilizados na regressão, suas fontes e descrições.

Fonte dos dados	Descrição	Unidade	Abreviação
ANA – Atlas Esgotos	Índice de Coleta sem Tratamento	Porcentagem (%)	MUN_ICS
	Índice de sem Coleta e sem Tratamento	Porcentagem (%)	MUN_JSCST
	Eficiência Média de Tratamento do município	Porcentagem (%)	MUN_EFIC
SNIS – Diagnóstico do manejo de resíduos sólidos urbanos	Taxa de cobertura da coleta de RDO em relação à pop. total	Porcentagem (%)	TX_COL_RDO

Tabela 1. Variáveis utilizadas na regressão linear múltipla.

Todos os dados coletados neste estudo foram consolidados em uma tabela que teve sua união ao shapefile da área de estudo, onde o código do município foi o dado utilizado como referência para essa união. No QGIS, utilizando a calculadora de atributos, foi realizado o cálculo da incidência da leptospirose, onde o cálculo é realizado através da equação 1.

$$\frac{\text{número de casos notificados no ano}}{\text{tamanho populacional no ano}} \times 10.000 \quad [1]$$

Ainda no QGIS, também foram produzidos os mapas de localização, mapas coropléticos de incidência com suas diferentes formas de visualização para explorar as possibilidades de representação. Já no software GeoDA, foram criadas as matrizes de distâncias, em que três foram selecionadas para serem utilizadas e verificar seus impactos nos resultados, sendo elas: Queen (queen_2_contiguity), Vizinho mais próximo de 50km (nn_50km) e Contiguidade Rook (rook_contiguity). Depois, foram realizadas análises exploratórias dos dados de incidência, e então análise de clusters com 5 e 3 classes por K-Means. O agrupamento K-means é um dos algoritmos não supervisionados mais simples e populares, um método de Clustering que objetiva particionar n observações dentre k grupos onde cada observação pertence ao grupo mais próximo da média. Assim, é possível obter um mapa que a partir dos dados, nesse caso a incidência da leptospirose, agrupa as áreas em “áreas parecidas”, dependendo do número de classes, formando *clusters*.

Em seguida buscou-se explorar a dependência espacial, mostrando como os valores estão relacionados no espaço. O conceito mais utilizado é o de autocorrelação espacial, que mede o quanto o valor observado de um atributo numa região é independente dos valores desta mesma variável nas localizações vizinhas. Uma das formas de detecção de similaridade entre áreas é através do índice global de Moran (I). O Índice Global de Moran (I) é definido como:

$$I = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad [2]$$

Em que n corresponde ao número de áreas; y_i é o valor do atributo considerado na área i ; \bar{y} representa o valor médio do atributo na região de estudo; w_{ij} são os pesos atribuídos conforme a conexão entre as áreas i e j . Este índice varia de -1 a +1 e quantifica o grau de autocorrelação espacial existente, sendo que -1 significa autocorrelação espacial negativa ou inversa; 0 significa aleatoriedade e +1 significa autocorrelação espacial positiva ou direta. A forma escolhida para visualização do I foi pelo Diagrama de Espalhamento de Moran, que é dividido em 4 quadrantes que podem ser especializados na forma do Mapa de Espalhamento de Moran. Q1 (valores positivos, médias positivas) e Q2 (valores negativos, médias negativas) indicam pontos em que uma área possui vizinhos com valores semelhantes, enquanto Q3 (valores positivos, médias negativas) e Q4 (valores negativos, médias positivas) indicam áreas que possuem vizinhos com valores distintos, como pode ser visto na figura 6.

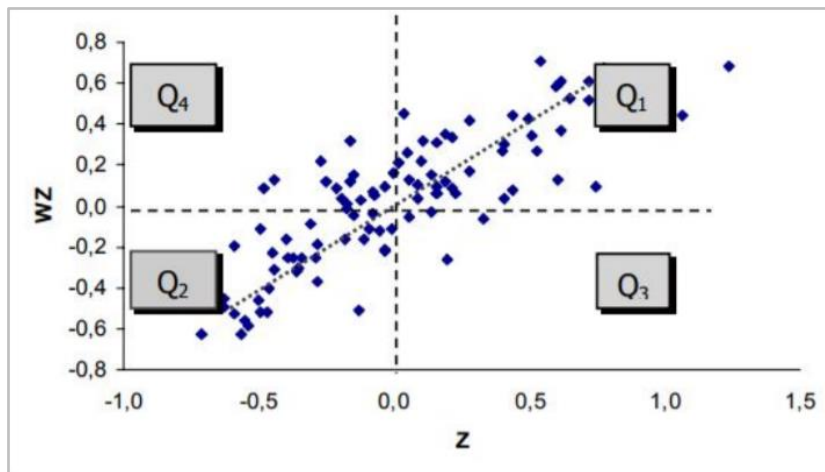


Figura 6. Diagrama de espalhamento de Moran.
Fonte: Druck et al. (2004).

Como este índice fornece apenas um único valor para todo o conjunto de dados, é interessante e indicado realizar uma análise em uma escala local para examinar padrões. Neste caso, é preciso utilizar indicadores locais de associação espacial que possam ser associados a diferentes localizações de uma variável distribuída espacialmente, pois eles produzem um valor específico para cada objeto (área). Isto acarreta na identificação de clusters (agrupamentos de objetos com valores semelhantes), outliers (objetos anômalos) e a presença de mais de um regime espacial. Um indicador local de associação espacial (LISA) tem que atender a dois objetivos (Anselin, 1995): a) permitir a identificação de padrões de associação espacial significativos; b) ser uma decomposição do índice global de associação espacial. O indicador local de associação espacial escolhido foi o Índice Local de Moran, para examinar padrões de áreas, dado por:

$$I_i = \frac{(y_i - \bar{y}) \sum_{j=1}^n w_{ij} (y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad [3]$$

Se $I_i > 0 \rightarrow$ clusters de valores similares (altos ou baixos); Se $I_i < 0 \rightarrow$ clusters de valores distintos. Assim, nesta análise foram gerados os mapas de Moran e mapas de significância.

E por fim, após a seleção, análise exploratória e transformações necessárias das variáveis, foi utilizada a regressão linear múltipla – Equação 4 (Neter et al., 1996) para verificar a relação entre a variável dependente (incidência de leptospirose) e as variáveis independentes escolhidas. Os estudos ecológicos visam estudar a relação entre a incidência de doenças e potenciais fatores etiológicos, e, para analisar o lado estatístico, esses estudos consistem, essencialmente, em modelos de regressão, em que o objetivo é buscar explicar a variação na incidência da doença por meio de outras variáveis. A equação 4 explica a equação da regressão linear múltipla.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \epsilon_i \quad [4]$$

Onde: Y_i = valor da variável resposta na i -ésima observação; β_0, \dots, β_p = parâmetros; X_{i1}, \dots, X_{ip} = valores das variáveis preditoras na i -ésima observação; ϵ_i = termo de erro aleatório.

As análises de autocorrelação espacial global e local, assim como a regressão linear múltipla e análise exploratória da incidência foram realizadas no software GeoDA, que é um pacote de software livre que realiza análise de dados espaciais, geovisualização, autocorrelação espacial e modelagem espacial pensado inicialmente pelo professor Anselin.

3. Resultados e discussão

A análise exploratória inicial da incidência de leptospirose mostrou que grande parte dos municípios possui incidência zerada. A incidência máxima encontrada foi de 1,64, sendo que a média foi de 0,1895, como pode ser observado no boxplot e histograma apresentados na figura 7.

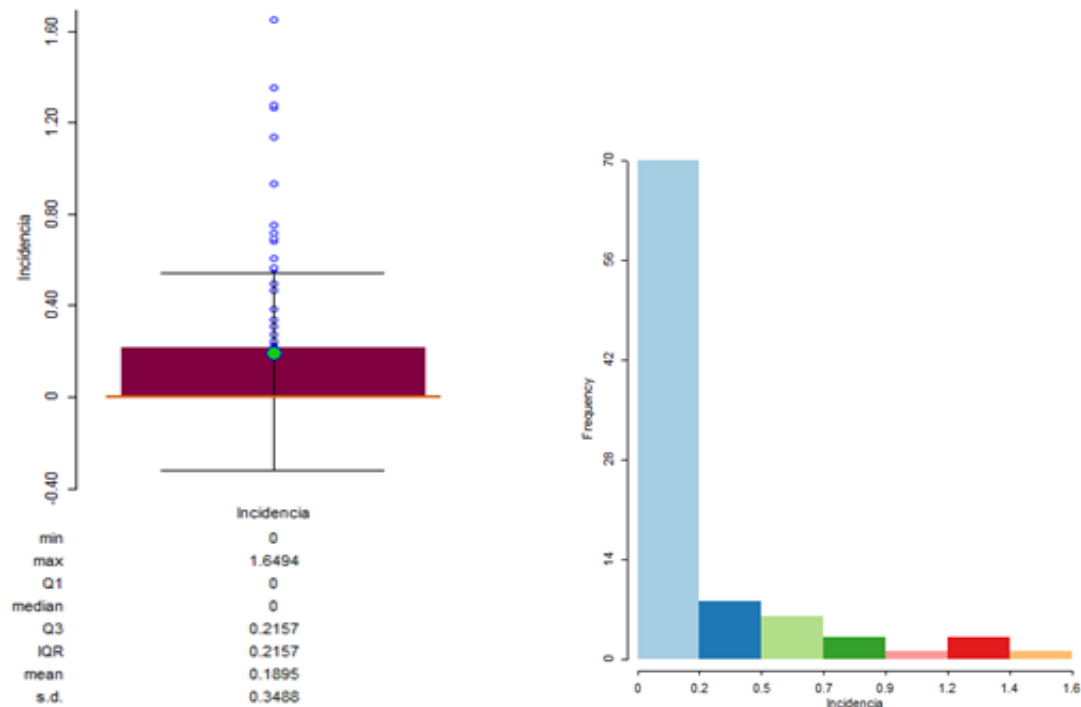


Figura 7. Boxplot e histograma dos valores da incidência.

E já no scatterplot da incidência com o número de casos (Figura 8), foi possível observar que não necessariamente a incidência será alta quando se tem um maior número de casos. Por exemplo, o município do Rio de Janeiro registrou 55 casos de leptospirose, mas devido à grande população residente, a incidência comparada a outros municípios é bem baixa. Assim como, municípios que possuem uma menor população, quando registram casos da doença elevam rapidamente valor da incidência.

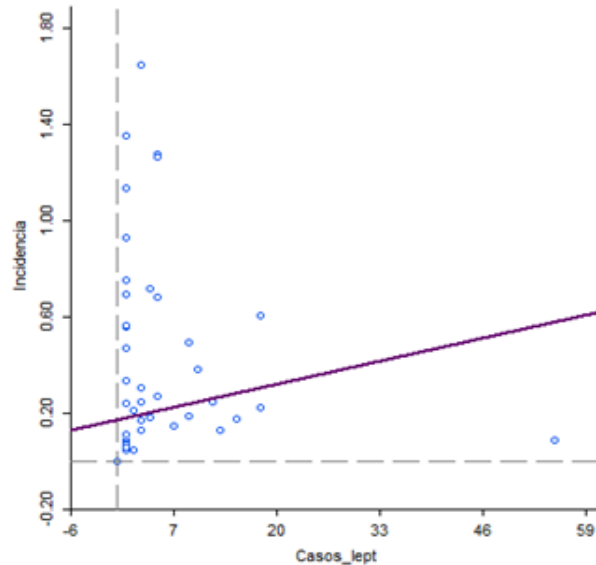


Figura 8. Gráfico de dispersão (scatterplot) de incidência e casos de leptospirose.

Inicialmente, para buscar observar a distribuição espacial dos casos notificados de leptospirose no estado do Rio de Janeiro, foi construído um mapa coroplético considerando como dado o número de casos por município. O observado como visto na figura 9, foi de acordo com o esperado, considerando que na Região Metropolitana do Rio de Janeiro haveria um destaque para um número maior de casos, e isso se deve principalmente às grandes populações residentes nesses municípios. Além dessa região, outras áreas mostraram um certo destaque comparada a outras, como por exemplo áreas mais a oeste e nordeste do estado.

Já no mapa de incidência da leptospirose, os municípios que se destacam se mostraram mais dispersos espacialmente ao longo do estado, como pode ser observado nas figuras 10, 11 e 12, onde cada uma apresenta um tipo de visualização de Padrão de Áreas. A figura 10 apresenta a visualização por intervalos iguais, e apresentou destaque em alguns municípios a oeste como Paraty e Mangaratiba, assim como no nordeste do estado, nos municípios de Porciúncula e Laje do Muriaé. Já na figura 11, a visualização por quantis, mantendo o mesmo número de classes (4 no total), acarretou na visualização de muitos municípios destacados, onde fica mais complexo o entendimento do mapa e sua interpretação. E, na figura 12, a visualização por quebras naturais mostrou, além dos municípios destacados na figura 10, outros que puderam mostrar regiões maiores em que deve-se ter uma atenção maior para a incidência da doença. Mesmo que ainda seja uma padrão mais disperso, é possível observar agrupamentos de municípios com incidência semelhante em áreas maiores.

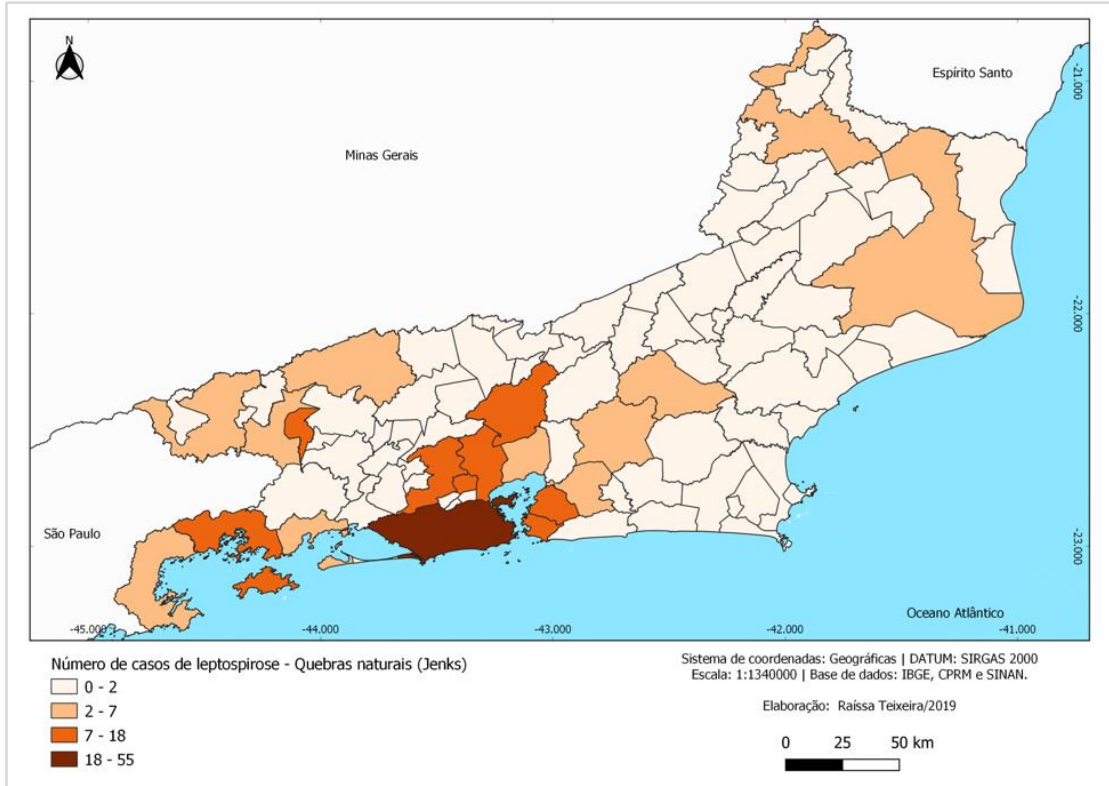


Figura 9. Mapa do número de casos notificados de leptospirose.

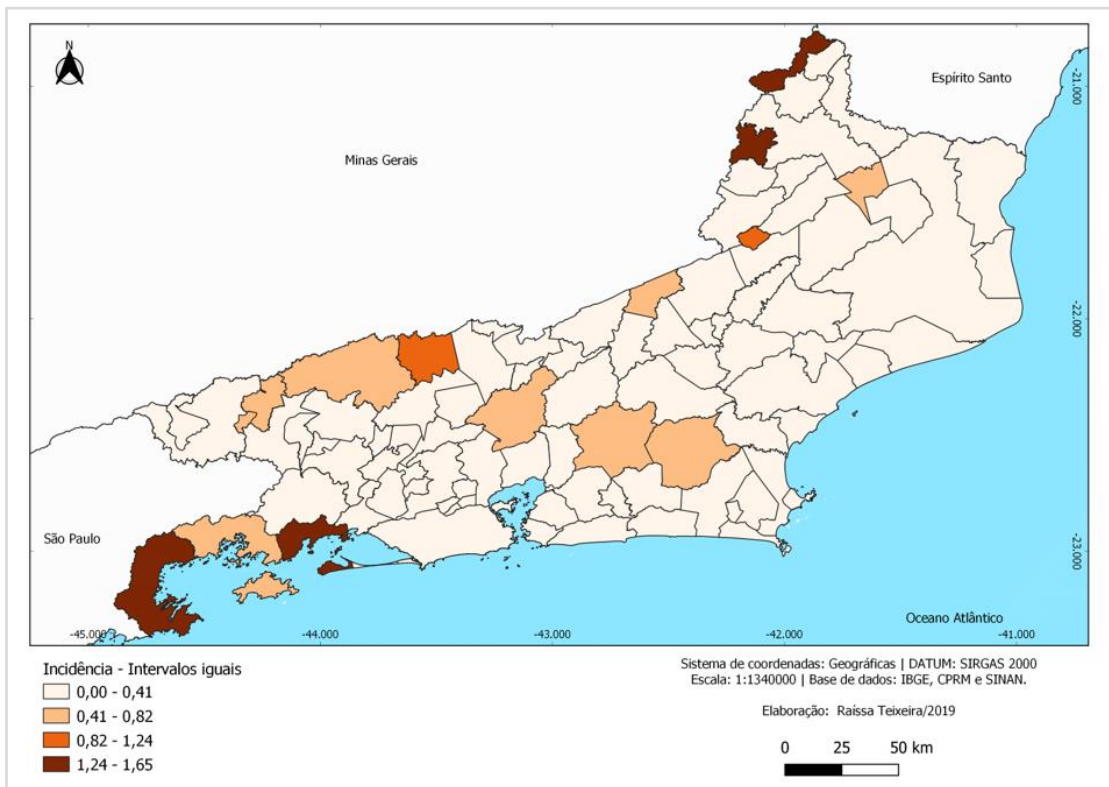


Figura 10. Mapa de incidência da leptospirose - Visualização por intervalos iguais.

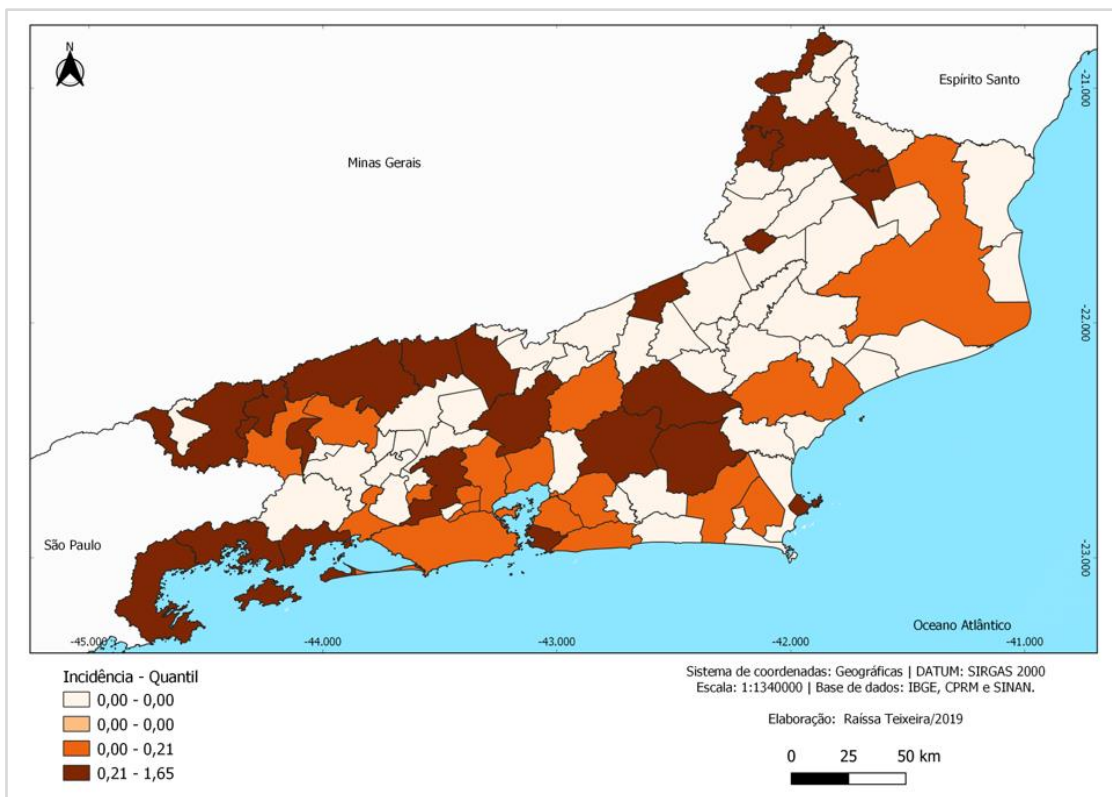


Figura 11. Mapa de incidência da leptospirose - Visualização por quantis.

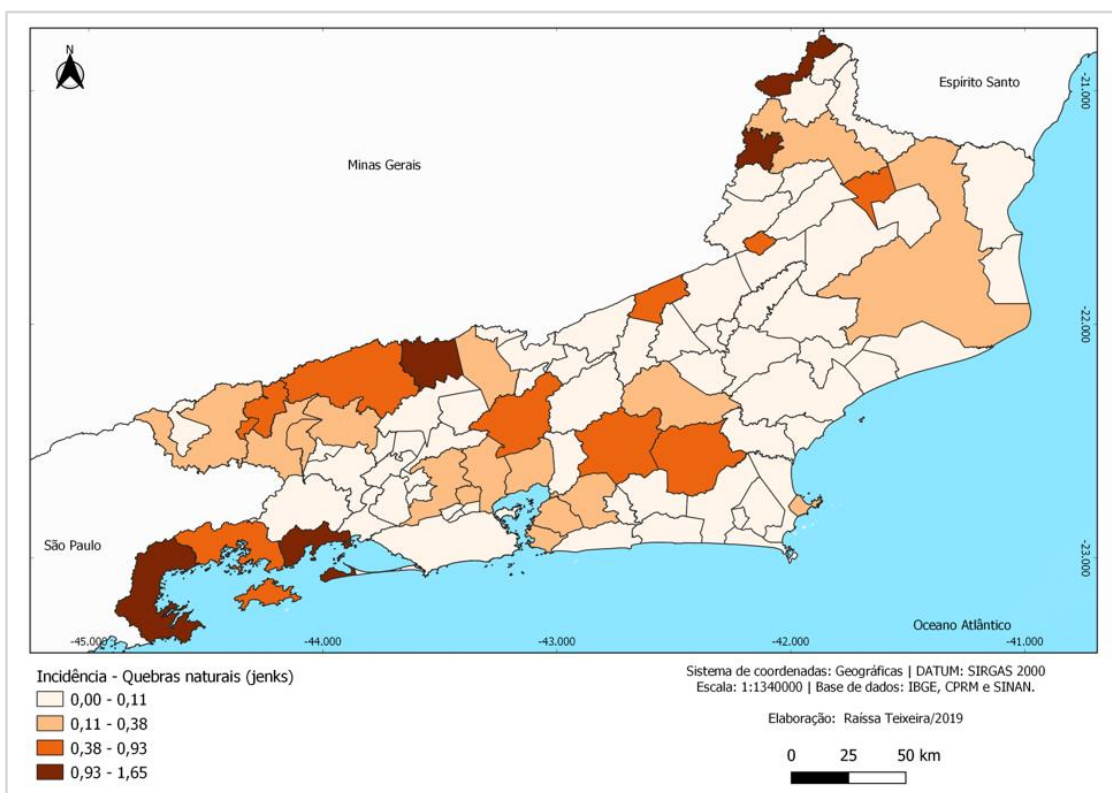


Figura 12. Mapa de incidência da leptospirose - Visualização por quebras naturais (jenks).

Um procedimento necessário para a análise de dados de área é a definição de uma matriz de vizinhança, também conhecida como matriz de distância, matriz de conectividade ou matriz de proximidade. Essa matriz pode ser composta, por exemplo, apenas pela lista de vizinhos de cada município; ou a distância entre municípios ligados por estradas; ou ainda a conectividade ponderada pelo comprimento da fronteira comum (Andrade et al., 2007). Foram definidas três matrizes a serem utilizadas, apresentadas na figura 13 juntamente com suas tabelas de propriedades e valores. Foi possível observar que a matriz do vizinho mais próximo e queen possuem valores mais parecidos, e se assemelham mais nas ligações no mapa. Ambas se diferem da matriz rook, principalmente em relação ao número médio e máximo de vizinhos.

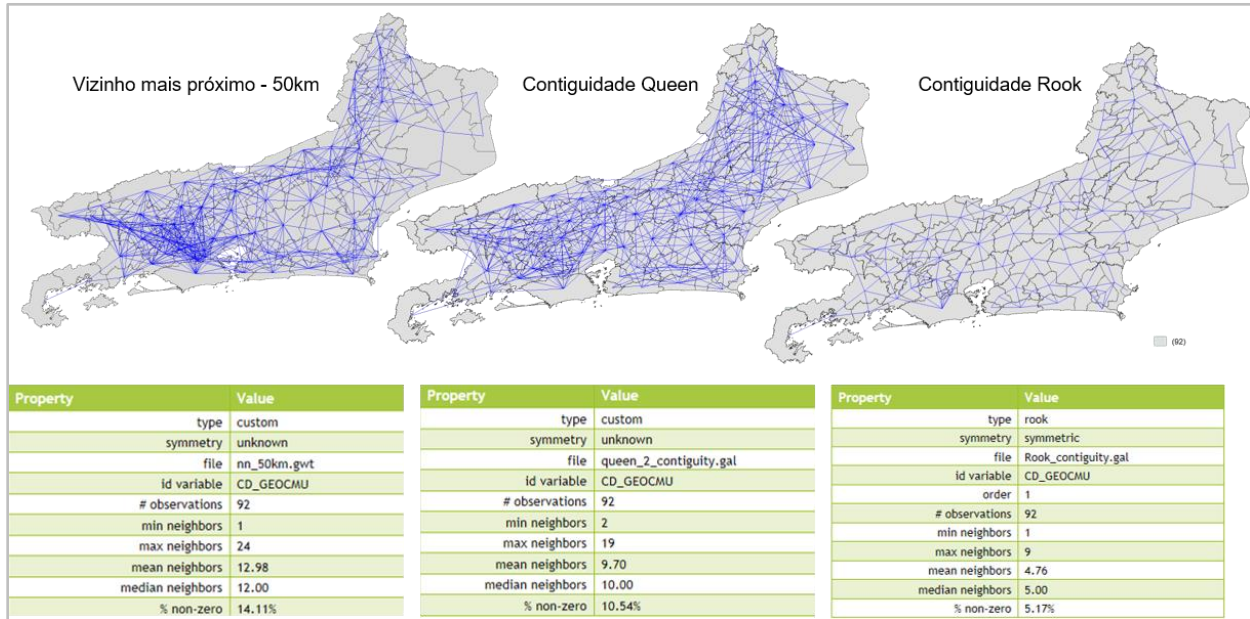


Figura 13. Matrizes de vizinhança.

Na análise de clusters, foram evidenciadas áreas com incidência de leptospirose semelhantes. O primeiro mapa (figura 14) foi construído com 5 classes. A classe 5 foi a que destacou unicamente o município de Porciúncula, que é o que possui a maior incidência no estado. Além disso, nas figuras 14 e 15 foram formados alguns clusters que podem ser observados na região nordeste, central, e oeste, sendo essa última tanto em municípios da região litorânea, quanto municípios que fazem divisa com o estado de Minas Gerais. Nota-se que alguns clusters foram possíveis de serem observados nos mapas de incidência e que se assemelham. Nestes abaixo, observa-se que os municípios destacados em cores verdes são os que apresentam maiores valores de incidência.

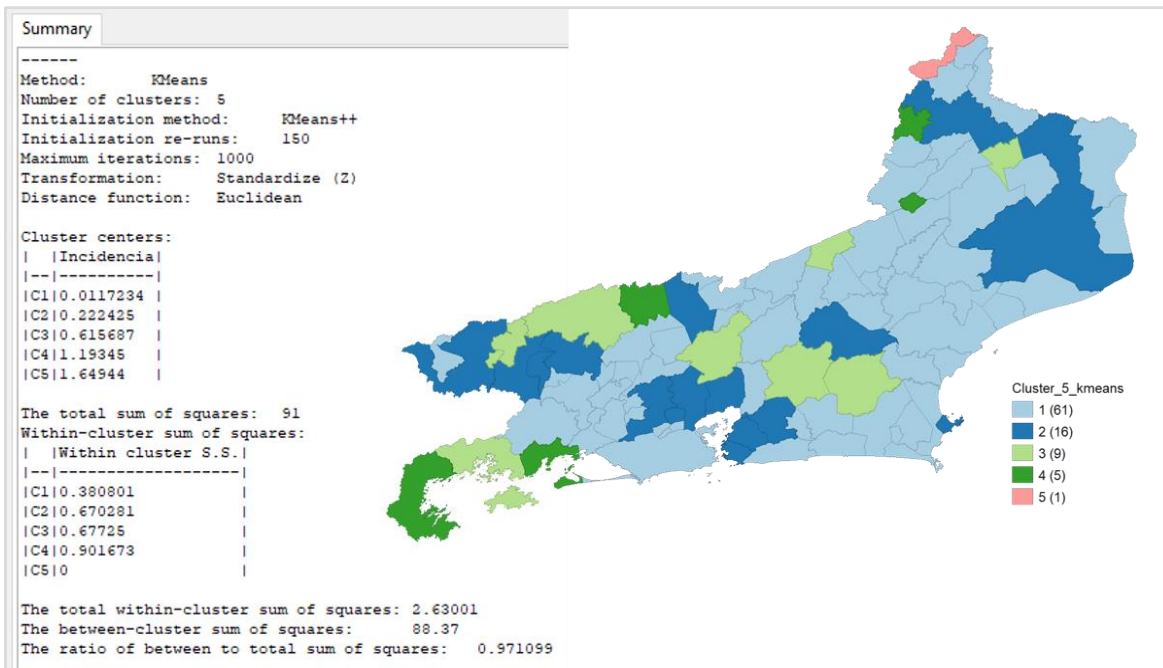


Figura 14. Mapa de clusters - K-means - 5 classes.

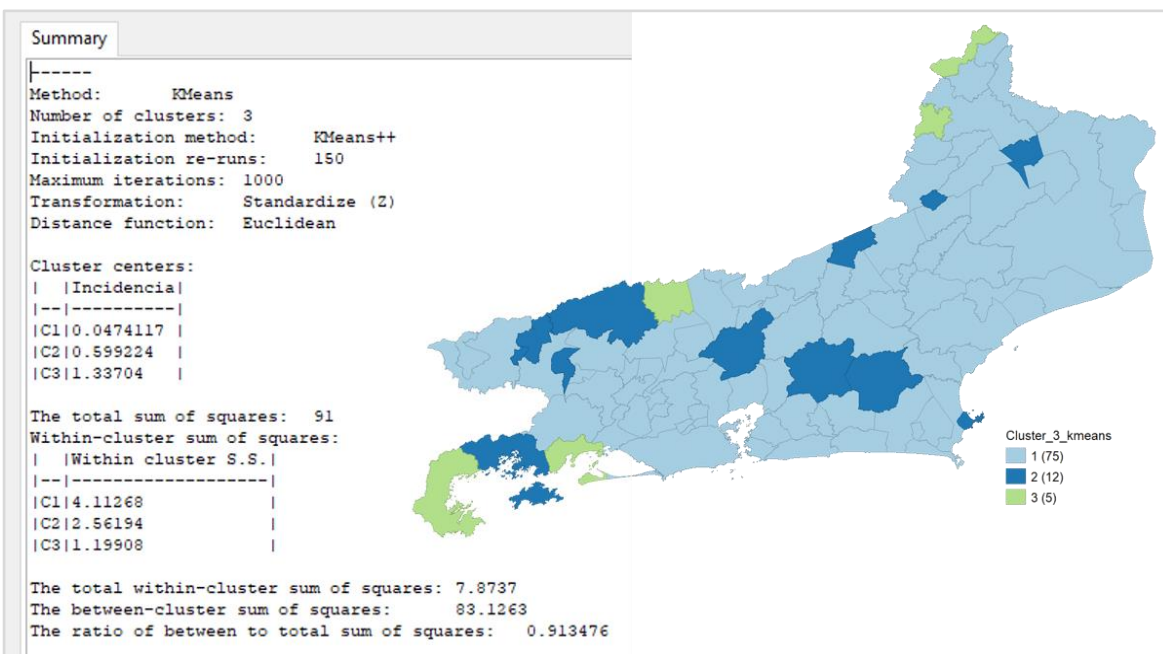


Figura 15. Mapa de clusters - K-means - 3 classes.

Na análise de autocorrelação espacial foram gerados os índices de Moran global representados pelo diagrama de espalhamento, e os mapas LISA (Índice Local De Associação Espacial), para visualização da correlação espacial local com sua significância, onde é possível observar quais municípios possuem alta ou baixa incidência em relação aos seus vizinhos, e vice

versa. Esse procedimento foi realizado para as três diferentes matrizes de vizinhança para verificar a influência das mesmas. As figuras 16, 17 e 18 apresentam os resultados.

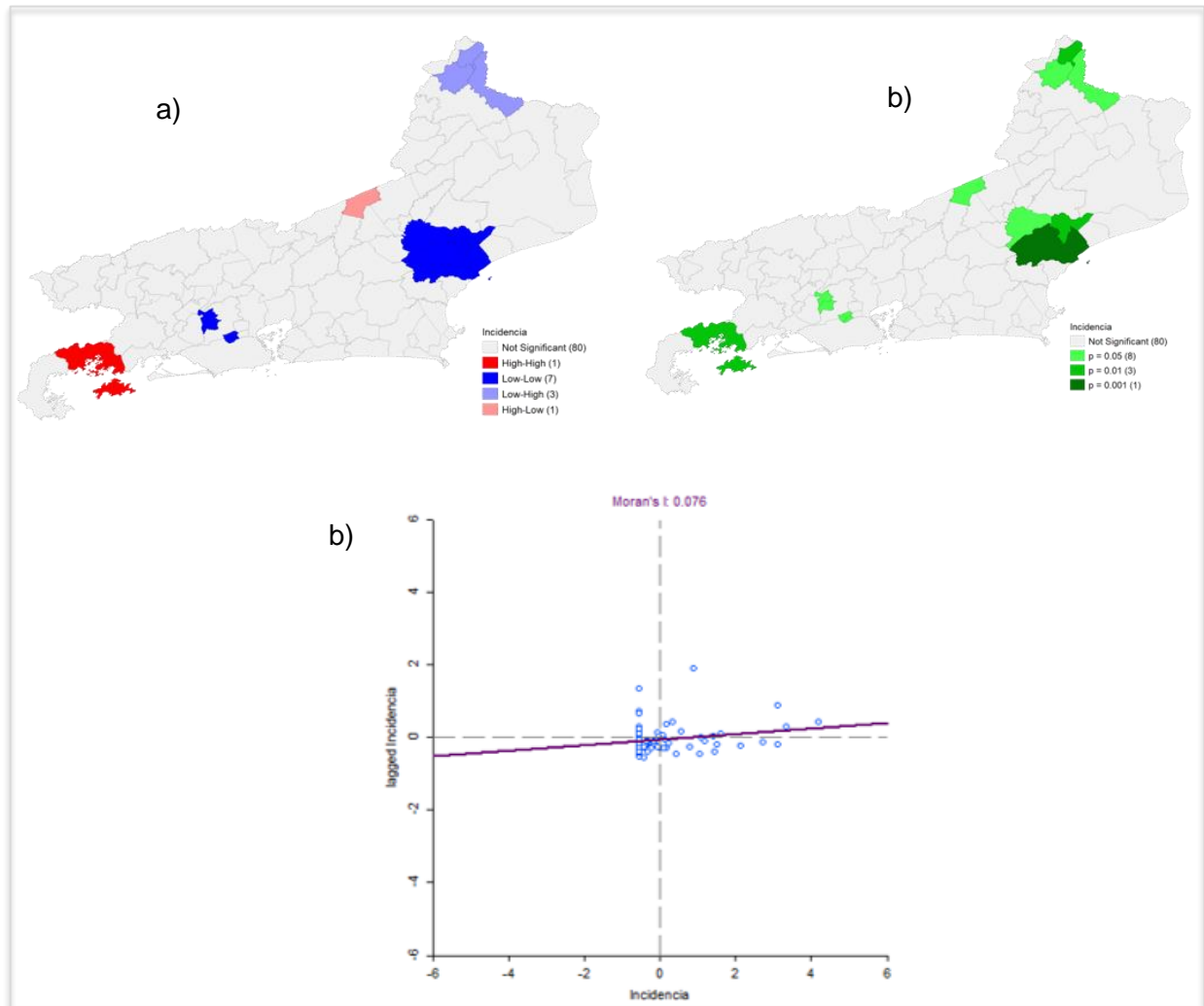


Figura 16. Matriz vizinho mais próximo - 50km. a) Mapa índice de Moran Local; b) Mapa de significância; c) Diagrama de espalhamento de Moran.

É possível observar que no resultado utilizando a matriz vizinho mais próximo - 50km, o índice de Moran apresenta um valor muito próximo a zero (0,076) ainda que positivo, o que pode indicar aleatoriedade na incidência entre os municípios, ou no máximo uma fraca correlação positiva. Já o Moran local apresentou alguns municípios em destaque, como Angra dos Reis, que possui alta incidência com vizinhos também com alta incidência (high-high), assim como alguns municípios em azul escuro (Trajano de Moraes, Macaé e Conceição de Macabu) que possuem baixa incidência, com vizinhos também de baixa incidência (low-low).

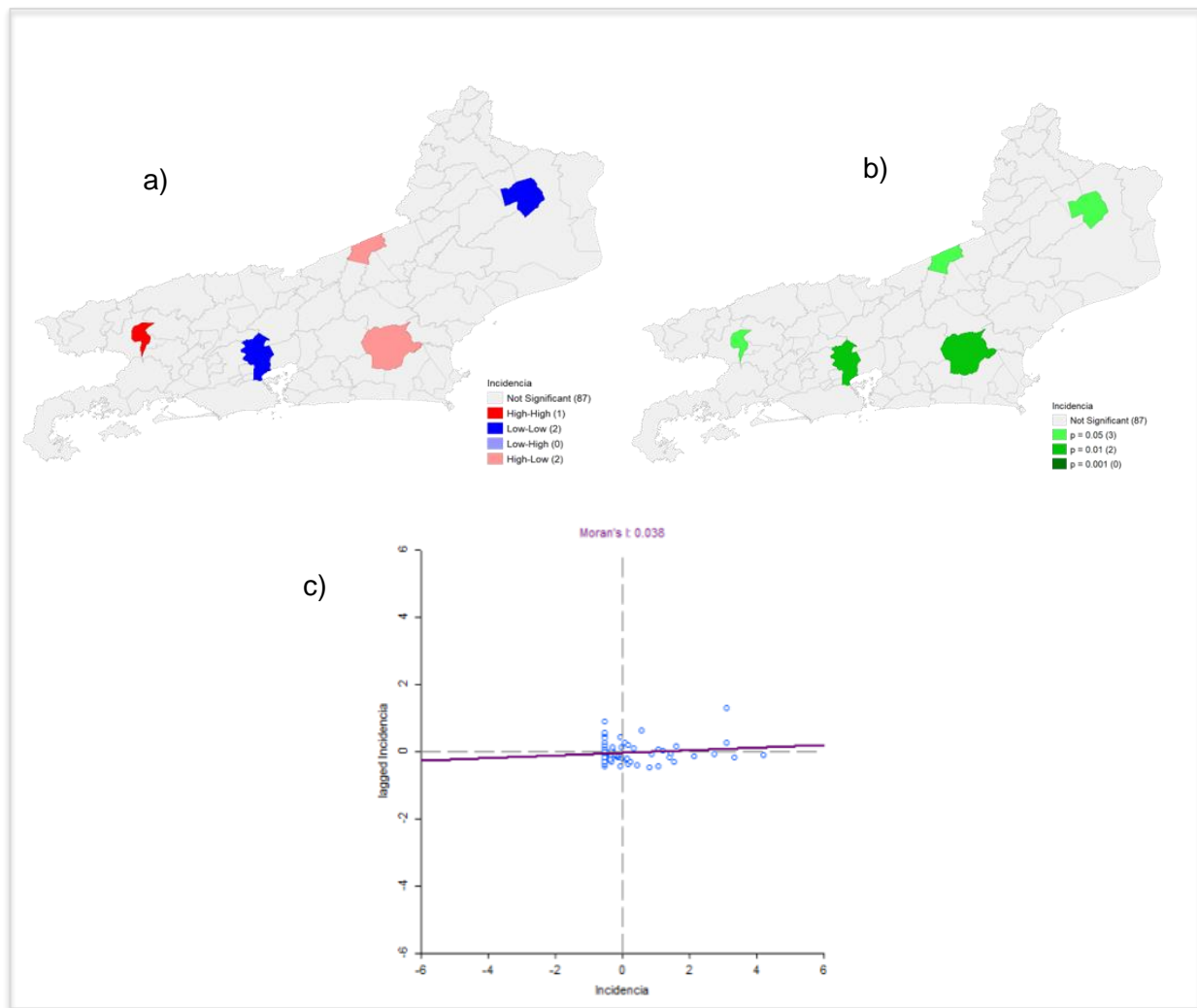


Figura 17. Matriz de contiguidade Queen. a) Mapa índice de Moran Local; b) Mapa de significância; c) Diagrama de espalhamento de Moran.

Na figura 17 é possível observar que utilizar uma matriz diferente (queen), afetou o resultado. Municípios diferentes foram destacados no mapa LISA, com níveis de significância também diferentes. O valor do índice de Moran foi de 0,038, menor do que o índice utilizando a matriz do vizinho mais próximo, o que indica mais proximidade à aleatoriedade.

A figura 18 apresenta os resultados da análise de autocorrelação utilizando a matriz de contiguidade rook. Novamente, a matriz revelou resultados diferentes. O índice de Moran foi de 0,081, o maior entre os resultados. Ele se aproxima mais de 0,1, o que indicaria uma correlação positiva (entre 0 e 1), porém, ainda é muito próximo a zero, o que leva a entender que a incidência de leptospirose no estado do Rio de Janeiro não possui forte correlação espacial, ou seja, está mais próxima da aleatoriedade no espaço. O mapa LISA apresentou também diferentes municípios significantes na análise, e mais uma vez Angra dos Reis foi identificado como high-high, e nessa matriz tiveram municípios com maior significância no resultado. Por fim, através do LISA em todos os casos, é possível ver os locais, em cinza, onde a autocorrelação espacial não é significativa, o que ocorre em grande parte do estado. Assim, não foram identificados muitos clusters ou aglomerados que poderiam indicar áreas de maior risco, pois os municípios em destaque se apresentaram mais dispersos.

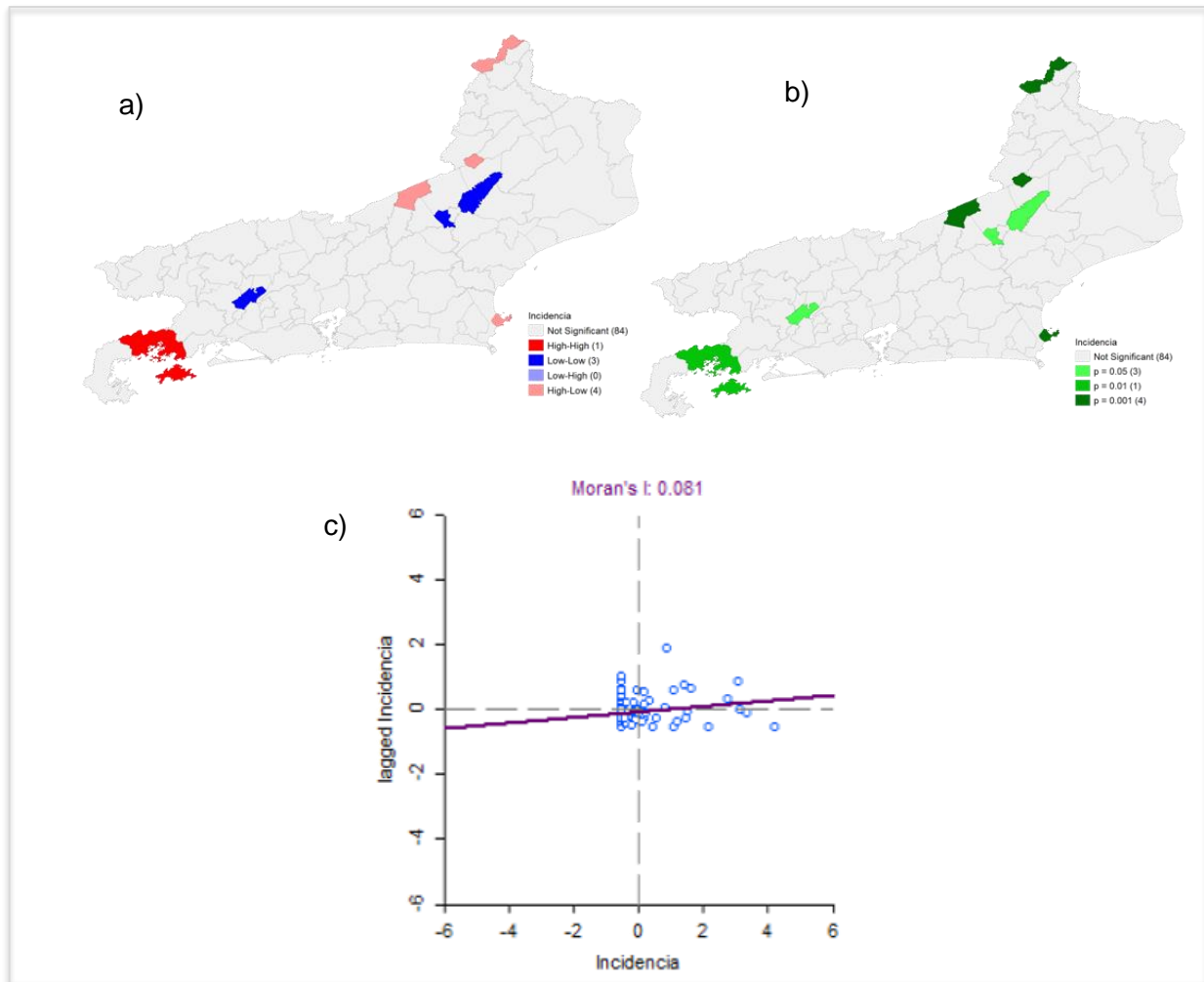


Figura 18. Matriz de contiguidade Rook. a) Mapa índice de Moran Local; b) Mapa de significância; c) Diagrama de espalhamento de Moran.

Por fim, foi realizada a aplicação de uma regressão linear múltipla para buscar examinar a relação entre a variável de interesse (dependente) que nesse caso foi a incidência da leptospirose, e variáveis exploratórias (preditores, independentes) que foram as variáveis de tratamento de esgoto e coleta de resíduos, já citadas anteriormente. Após a seleção das variáveis e testes de diferentes modelos, onde nenhum mostrou grande êxito na explicação da variável dependente, o resultado do modelo escolhido é apresentado nas figuras 19 e 20 a seguir.

```

REGRESSION
-----
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set      : municipios_RJ_shape-final
Dependent Variable : Incidencia Number of Observations: 64
Mean dependent var : 0.202245 Number of Variables : 5
S.D. dependent var : 0.362906 Degrees of Freedom : 59

R-squared      : 0.021257 F-statistic      : 0.320353
Adjusted R-squared : -0.045098 Prob(F-statistic) : 0.863245
Sum squared residual: 8.24969 Log likelihood    : -25.2534
Sigma-square    : 0.139825 Akaike info criterion : 60.5069
S.E. of regression : 0.373932 Schwarz criterion   : 71.3013
Sigma-square ML  : 0.128901
S.E of regression ML: 0.359028

```

Variable	Coefficient	Std.Error	t-Statistic	Probability
CONSTANT	0.266738	0.565456	0.471723	0.63886
MUN_ISCST	0.000454796	0.00444435	0.102331	0.91884
MUN_EFIC	0.00278687	0.00504316	0.552604	0.58262
MUN_ICS	0.00284812	0.00368092	0.773752	0.44217
Tx_col_RDO	-0.00290074	0.00596105	-0.486616	0.62833

```

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER 30.923897
TEST ON NORMALITY OF ERRORS
TEST      DF      VALUE      PROB
Jarque-Bera 2      109.3631  0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST      DF      VALUE      PROB
Breusch-Pagan test 4      9.4959   0.04983
Koenker-Bassett test 4      2.8711   0.57963

```

Figura 19. Resultado da regressão linear múltipla.

```

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : queen_2_contiguity
(row-standardized weights)
TEST      MI/DF      VALUE      PROB
Moran's I (error) -0.0097    0.0807    0.93570
Lagrange Multiplier (lag) 1      0.1401    0.70816
Robust LM (lag) 1      2.8730    0.09007
Lagrange Multiplier (error) 1      0.0177    0.89425
Robust LM (error) 1      2.7506    0.09722
Lagrange Multiplier (SARMA) 2      2.8907    0.23566
===== END OF REPORT =====

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : nn_50km
(row-standardized weights)
TEST      MI/DF      VALUE      PROB
Moran's I (error) 0.0674    1.5445    0.12246
Lagrange Multiplier (lag) 1      1.0702    0.30090
Robust LM (lag) 1      0.3850    0.53493
Lagrange Multiplier (error) 1      0.9485    0.33011
Robust LM (error) 1      0.2633    0.60787
Lagrange Multiplier (SARMA) 2      1.3335    0.51338
===== END OF REPORT =====

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : Rook_contiguity
(row-standardized weights)
TEST      MI/DF      VALUE      PROB
Moran's I (error) -0.0010    0.2766    0.78210
Lagrange Multiplier (lag) 1      0.0381    0.84527
Robust LM (lag) 1      3.6491    0.05610
Lagrange Multiplier (error) 1      0.0001    0.99250
Robust LM (error) 1      3.6111    0.05739
Lagrange Multiplier (SARMA) 2      3.6492    0.16128
===== END OF REPORT =====

```

Figura 20. Resultado da regressão linear múltipla para as diferentes matrizes.

Observando os resultados, foi possível perceber que o modelo não obteve sucesso. O R^2 (R-squared) mostrou-se muito baixo, e os valores do log likelihood, Akaike (AIC) e Schwarz que indicam sobre o poder explicativo do modelo, mostraram que esse modelo de regressão não consegue explicar muito a variável dependente (incidência). Sobre os coeficientes que funcionam como um coeficiente de força, a partir do teste T, é possível observar que eles não possuem muita significância, por estarem em um intervalo de confiança menor do que 90%, ou seja, com pouca precisão. Também foi possível observar a existência de multicolinearidade, por mostrar um valor maior do que 30, o que é suspeito e deve ser investigado mais a fundo, pois pode significar um problema comum em regressões, no qual as variáveis independentes possuem relações lineares exatas ou aproximadamente exatas. Os resíduos do modelo não estão apresentando variância constante (homocedasticidade) e normalidade, o que era esperado para resíduos aleatórios. E, os resíduos mostram que ainda há algum viés no modelo, mas que pode ser originário de vários fatores. Na regressão, a utilização de matrizes diferentes não mostrou uma grande diferença nos resultados, como observado na figura 20. Nem sempre se consegue confirmar relações entre os problemas de saúde e os fatores ambientais e sociais, mas a ausência de relação entre os indicadores, ao contrário de ser um resultado negativo de uma investigação é, antes de tudo, uma pista para identificar padrões de proteção ou de agravamento de riscos (Resendes et al., 2017).

Com os resultados estatísticos, deve-se considerar que para tentar entender essas relações, outras variáveis poderiam ser utilizadas, assim como a realização de uma análise temporal para observar como a doença se comporta ao longo do tempo. Outros tipos de análise também poderiam ser realizados, como o cálculo do índice de Moran bivariado para verificar se é possível encontrar alguma outra estrutura entre as variáveis.

Em análises como essas, deve-se levar em consideração as incertezas das notificações lançadas no sistema DATASUS, principalmente em municípios menores, pois podem ocorrer subnotificações de casos. Ainda, quando se observa a distribuição espacial da incidência, é mais provável que se visualize outliers, enquanto quando se busca visualizar apenas a distribuição do número de casos, são facilmente observados agrupamentos maiores devido ao adensamento populacional, como ocorreu na Região Metropolitana do Rio de Janeiro. Além disso, é importante considerar que, um indicador isolado não pode explicar a incidência de uma doença infecciosa complexa como a leptospirose, como citado por Gracie et al (2014).

4. Conclusão

Com esse trabalho foi possível concluir através dos resultados, que diferentes matrizes de vizinhança promovem resultados diferentes em análises de autocorrelação espacial, sendo que as matrizes de vizinho mais próximo e queen demonstraram resultados mais parecidos entre si quando comparados à matriz rook. Ainda assim, em todos os resultados do Índice Global de Moran, a incidência mostrou ausência de correlação espacial significativa entre os municípios. Os mapas LISA mostraram algumas poucas áreas que poderiam receber mais atenção na saúde para essa doença, onde municípios com relativa alta incidência fazem vizinhança com municípios também de alta incidência (high-high).

Na regressão, foram encontrados resultados insuficientes para ter conclusões concretas, o que mostra que no modelo de regressão linear múltipla escolhido, as variáveis independentes não explicam a variável dependente. Assim, outros modelos e dados diferentes poderiam ser testados.

Referências bibliográficas

Andrade, Ana Lúcia et al. **Introdução à estatística espacial para a saúde pública**. 2007.

Ashford, David A. et al. Asymptomatic infection and risk factors for leptospirosis in Nicaragua. **The American journal of tropical medicine and hygiene**, v. 63, n. 5, p. 249-254, 2000.

Bacallao, Jorge et al. Socioeconomic factors and vulnerability to outbreaks of leptospirosis in Nicaragua. **International journal of environmental research and public health**, v. 11, n. 8, p. 8301-8318, 2014.

Barcellos, Christovam; Sabroza, Paulo Chagastelles. Socio-environmental determinants of the leptospirosis outbreak of 1996 in western Rio de Janeiro: a geographical approach. **International Journal of Environmental Health Research**, v. 10, n. 4, p. 301-313, 2000.

Brasil. Sistema Nacional de Informações Sobre Recursos Hídricos. Agência Nacional de Águas. **Atlas Esgotos - Despoluição das Bacias Hidrográficas**. Disponível em: <<http://www.snirh.gov.br/portal/snirh/snirh-1/atlas-esgotos>>.

Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância em Saúde Ambiental e Saúde do Trabalhador. **Diretriz para atuação em situações de surtos de doenças e agravos de veiculação hídrica**. 2018. Modo de acesso: World Wide Web: ISBN 978-85-334-2668-9.

Brasil. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância Epidemiológica. **Guia de vigilância epidemiológica** / Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Vigilância Epidemiológica. 7. Ed. – Brasília, Ministério da Saúde, 2009. 816 p. – (Série A. Normas e Manuais Técnicos).

Brasil. Ministério do Desenvolvimento Regional. Sistema Nacional de Informações Sobre Saneamento – SNIS. **Diagnóstico do manejo de Resíduos Sólidos Urbanos – 2013**. Disponível em: <<http://www.snis.gov.br/diagnostico-residuos-solidos/diagnostico-rs-2013>>.

Dias, Juarez Pereira et al. Factors associated with *Leptospira* sp infection in a large urban center in northeastern Brazil. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 40, n. 5, p. 499-504, 2007.

Druck, S.; Carvalho, M.; Câmara, G.; Monteiro, A. M. V. Análise espacial de dados geográficos. **Brasília: Embrapa**, 2004.

Gracie, Renata et al. Geographical scale effects on the analysis of leptospirosis determinants. **International journal of environmental research and public health**, v. 11, n. 10, p. 10366-10383, 2014.

Hagan, José E. et al. Spatiotemporal determinants of urban leptospirosis transmission: four-year prospective cohort study of slum residents in Brazil. **PLoS neglected tropical diseases**, v. 10, n. 1, p. e0004275, 2016.

Hino, P., Villa, T., Sasaki, C., Nogueira, J., & Santos, C. (2006). Geoprocessamento aplicado à área da saúde. **Revista Latino-Americana De Enfermagem**, 14(6), 939-943.
<https://doi.org/10.1590/S0104-11692006000600016>

Mwachui, Mwanajaa Abdalla et al. Environmental and behavioural determinants of leptospirosis transmission: a systematic review. **PLoS neglected tropical diseases**, v. 9, n. 9, p. e0003843, 2015.

Neter, J., Kutner, M.H., Nachtsheim, C.J., & Wasserman, W. (1996). **Applied linear statistical models**. Chicago: Irwin.

Resendes, Ana Paula da Costa et al. **Abordagens espaciais na saúde pública**. 2017.

World Health Organization, 2019. Zoonoses, **Leptospirosis**. Disponível em:
<<https://www.who.int/zoonoses/diseases/leptospirosis/en/>>.