



UNIVERSIDADE DE SÃO PAULO
Faculdade de Filosofia, Letras e Ciências Humanas
Departamento de Geografia

Rodrigo Aroni Siquette
NºUSP: 9372346

**ANÁLISE ESPACIAL DE DOENÇAS RELACIONADAS AO SANEAMENTO
AMBIENTAL INADEQUADO (DRSAI) NA SUBPREFEITURA CAPELA DO SOCORRO**

Trabalho de Graduação Individual entregue como requisito para conclusão do Bacharelado em Geografia oferecido pelo Departamento de Geografia da Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo.

Orientadora: Profa. A Dra. Ligia Vizeu Barrozo

São Paulo, 2023

*Este trabalho é todo dedicado aos meus pais,
pois é graças ao seu esforço que hoje posso
concluir o meu curso.*

*Foi pensando nas pessoas que executei este
projeto, por isso dedico este trabalho a todos
aqueles a quem esta pesquisa possa ajudar de
alguma forma.*

*Dedico este trabalho aos meus colegas de
curso, e todos meus/minhas amigos(as) do
atletismo, e da vida, principalmente ao Alan,
Felipe, Vinicius, Brunos, Emile, Mariana e
meu primo Cleber*

We invade tropical forests and other wild landscapes, which harbor so many species of animals and plants — and within those creatures, so many unknown viruses, We cut the trees; we kill the animals or cage them and send them to markets. We disrupt ecosystems, and we shake viruses loose from their natural hosts. When that happens, they need a new host. Often, we are it.”

David Quammen

Resumo

Com metade da população mundial vivendo em áreas urbanas , as cidades oferecem oportunidades de emprego, educação, acesso a serviços como saúde e cultura e contribuição econômica para os países. No entanto, o crescimento urbano não planejado resultou em pobreza, degradação ambiental e esgotamento dos recursos naturais que ameaçam a saúde humana e a qualidade de vida. Os avanços tecnológicos permitiram investigações mais profundas sobre saúde urbana e estatísticas em todo o mundo, que apontaram para um agravamento considerável de problemas associados à expansão urbana não planejada em países de baixa e média renda, como moradias precárias, poluição do ar e da água e condições inadequadas serviços de saneamento. Gerenciar uma crise endêmica, portanto, abrange muitos aspectos diferentes, incluindo saúde, meio ambiente, habitação e planejamento urbano.

A expansão urbana não planejada resulta em pobreza urbana, especialmente em países com economias de renda baixa e média-baixa. No entanto, cidades bem geridas oferecem melhores economias de escala em termos de uso do solo, provisão e eficiência dos serviços e gestão dos recursos ambientais. A invasão do tecido urbano em áreas de vida silvestre tem levado ao aumento da disseminação de doenças, enquanto os serviços de saúde sofreram um declínio na distribuição. Os problemas de saúde no âmbito urbano estão intimamente associados às condições ambientais do entorno, por isso é fundamental o monitoramento conjunto tanto dos riscos ambientais quanto dos riscos e impactos à saúde humana ao se aprofundar o tema.

Patógenos restritos a ambientes florestais estão se tornando comuns devido à rápida urbanização, levando à invasão de áreas antes selvagens e intensificando o contato com o homem. O saneamento inadequado tem sido o principal agente causador de doenças como diarreia e helmintíases, além da psoríase - transmitidas por meio de fezes humanas. O acúmulo de lixo também atua como um terreno fértil para vários vetores de doenças infecciosas. Aproximadamente 75% das doenças infecciosas emergentes se originam de animais para humanos e surgem diretamente de interações homem-animal.

As mudanças ambientais estão forçando humanos e animais a interagirem de novas maneiras, levando ao surgimento de novas doenças. O Programa das Nações Unidas para o

Meio Ambiente (PNUMA) identifica o desmatamento, a agricultura intensiva, o comércio ilegal e a má regulamentação da vida selvagem, além das mudanças climáticas, como fontes centrais de surtos zoonóticos. A perda de florestas é uma das formas mais significativas de degradação ambiental, facilitando novas interações entre humanos e animais.

Em conclusão, a gestão da expansão urbana requer uma abordagem holística que considere o meio ambiente, a habitação, a saúde e o planejamento urbano. Os impactos do crescimento urbano desordenado são numerosos, levando à pobreza, à degradação ambiental e ao esgotamento dos recursos naturais. Como resultado, as cidades devem ser geridas de forma sustentável para proporcionar melhores economias de escala e mitigar o risco de surtos zoonóticos. Ao monitorar conjuntamente os riscos ambientais e seus impactos na saúde humana, podemos criar um ambiente urbano mais saudável para todos.

Abstract

With half of the world's population living in urban areas since 2007, cities offer opportunities for jobs, education, access to services such as health and culture, and economic contribution to countries. However, unplanned urban growth has resulted in poverty, environmental degradation, and depletion of natural resources that threaten human health and quality of life. Technological advances have enabled deeper investigation into urban health and statistics around the world, which has pointed to a considerable worsening of problems associated with unplanned urban sprawl in low- and middle-income countries, such as precarious housing, air and water pollution, and inadequate sanitation services. Managing a pandemic crisis, therefore, encompasses many different aspects, including health, environment, housing, and urban planning.

Unplanned urban expansion results in urban poverty, especially in countries with low-income and lower-middle-income economies. However, well-managed cities offer better economies of scale in terms of land use, provision and efficiency of services, and management of environmental resources. The invasion of the urban fabric in wildlife areas has led to increased spread of diseases, while health services experienced a decline in distribution. Health problems in the urban sphere are closely associated with the surrounding

environmental conditions, which is why it is essential to jointly monitor both environmental risks and risks and impacts on human health when analyzing this topic in depth.

Pathogens restricted to forest environments are becoming common due to rapid urbanization, leading to the invasion of previously wild areas and intensifying contact with humans. Inadequate sanitation has been the main causative agent of diseases such as diarrhea and helminthiasis, in addition to psoriasis - which are spread through human feces. Waste accumulation also acts as a breeding ground for several vectors of infectious diseases. Approximately 75% of emerging infectious diseases originate from animals to humans and arise directly from human-animal interactions.

Environmental changes are forcing humans and animals to interact in new ways, leading to the emergence of new diseases. The United Nations Environment Program (UNEP) identifies deforestation, intensive agriculture, illegal trade and poor regulation of wildlife, in addition to climate change, as central sources of zoonotic outbreaks. The loss of forests is one of the most significant forms of environmental degradation, facilitating new interactions between humans and animals.

In conclusion, managing urban expansion requires a holistic approach that considers the environment, housing, health, and urban planning. The impacts of unplanned urban growth are numerous, leading to poverty, environmental degradation, and the depletion of natural resources. As a result, cities must be managed in a sustainable way to provide better economies of scale and mitigate the risk of zoonotic outbreaks. By jointly monitoring environmental risks and their impacts on human health, we can create a healthier urban environment for everyone.

SUMÁRIO

Introdução	3
Processo de contágio de vírus por causa da degradação ambiental - spillover	5
Vulnerabilidade socioambiental e a relação entre doenças por água contaminada	7
Objetivo da pesquisa	9
Local de estudo, sua história e formação	9
Dados	10
Metodologia	13
1) Análise de agrupamento espacial	16
A) Função K de Ripley	17
B) Função F - Espaços vazios	19
C) Função G - Vizinho mais próximo	21
2) Autocorrelação espacial	22
A) Índice global de Moran	23
B) Correlograma de Distância	25
C) Índice Locais de associação Espacial (LISA)	26
3) Regressão espacial:	28
A) Regressão Geograficamente Ponderada - GWR	29
Resultados	31
Conclusão	44
Referências bibliográficas:	47
Script R	49

Introdução

Atualmente, metade da população mundial vive em áreas urbanas¹. As cidades são atrativas por conta dos empregos, da educação, do acesso a serviços como saúde e cultura e pela enorme contribuição econômica que geram ao país como um todo. No entanto, o crescimento urbano veloz e muitas vezes desordenado está diretamente relacionado à pobreza, à degradação ambiental e ao esgotamento dos recursos naturais disponíveis, o que tem cada vez mais ameaçado a saúde e a qualidade de vida humanas.

O avanço tecnológico possibilitou uma investigação mais profunda sobre saúde urbana por meio de dados estatísticos criteriosos e facilmente disponíveis em todo o mundo. Nos países de renda baixa e média, eles apontam para um agravamento considerável dos problemas associados à expansão não planejada da mancha urbana, como habitação precária, adensamento populacional, poluição do ar e dos recursos hídricos, serviços inadequados de saneamento básico, doenças trazidas por vetores, entre outros.

As cidades desempenham um papel importante na estrutura social e nas economias dos países. Como já mencionado anteriormente, elas atuam como pólos de educação, emprego e serviços, e tendem a possibilitar uma expectativa de vida maior do que nas áreas rurais. Porém, na realidade, essas oportunidades não demonstram ser tão promissoras quanto antes, uma vez que as demandas populacionais estão excedendo a capacidade dos serviços ambientais e prejudicando a distribuição de água potável, saneamento e disposição de resíduos em vários lugares.

A pobreza urbana é uma das consequências mais significativas decorrentes da expansão urbana não planejada, especialmente nos países com Economias de baixa renda,

¹ <https://news.un.org/pt/story/2019/02/1660701>

Economias de renda média-baixa²¹. Mas considerando a importância das cidades para a dinâmica econômica global, fica clara a necessidade urgente de se pensar espaços mais sustentáveis a médio e longo prazos, até porque cidades bem geridas oferecem melhores economias de escala em termos de uso do solo, disposição e eficiência dos serviços, gerenciamento dos recursos ambientais etc.

Quando o desenvolvimento dos centros urbanos se dá de forma improvisada, a habitação marginal precária, a superlotação e os níveis crescentes de poluição são apenas uma face desse contratempo enfrentado pelos países com Economias de baixa renda, Economias de renda média-baixa pois à medida em que os desafios ambientais se intensificam, mais as atividades comerciais e privadas dentro dos limites das cidades são prejudicados também.

A invasão da malha urbana nas áreas de vida silvestre ilustra bem o exposto acima. Tal cenário aumentou substancialmente a propagação de doenças como leishmaniose, leptospirose e febre amarela, ao mesmo tempo em que os serviços de saúde tiveram uma decaída de distribuição. Inclusive, isso também evidencia como os problemas de saúde na esfera urbana estão intimamente associados às condições ambientais do entorno. Por isso é essencial monitorar conjuntamente tanto os riscos ambientais, quanto os riscos e impactos à saúde humana ao analisar essa temática a fundo.

A Organização Mundial da Saúde (OMS), por exemplo, possui modelos que vinculam estudos ambientais a doenças infectocontagiosas. Hoje sabe-se que patógenos restritos aos ambientes florestais estão se tornando comuns por causa da rápida urbanização. A invasão de áreas antes *selvagens*, especialmente de florestas tropicais, está aproximando em ritmo acelerado os núcleos urbanos de vetores de doenças. Como consequência, hospedeiros como

²¹ O Banco Mundial classifica os países com base em sua Renda Nacional Bruta (RNB) per capita, que é a renda média auferida por uma pessoa em um país em um determinado ano. Com base nesse critério, o Banco Mundial classifica os países em quatro grupos de renda: Economias de baixa renda: RNB per capita de US\$ 1.036 ou menos; Economias de renda média-baixa: RNB per capita entre US\$ 1.036 e US\$ 4.045; Economias de renda média alta: RNB per capita entre US\$ 4.046 e US\$ 12.535; Economias de alta renda: RNB per capita de US\$ 12.536 ou mais. Os países classificados como economias de renda baixa ou média-baixa são normalmente considerados países em desenvolvimento, enquanto aqueles classificados como economias de renda média-alta e alta são considerados mais desenvolvidos. O Banco Mundial atualiza sua classificação de renda a cada ano com base nos últimos dados disponíveis. <https://blogs.worldbank.org/opendata/should-we-continue-use-term-developing-world>

roedores e hemípteros estão se reproduzindo em novos habitats e intensificando o contato com humanos.³

A esse cenário soma-se outros: o saneamento inadequado tem sido o principal agente causador de doenças como diarréia e helmintíase, além da psoríase - que se propagam através de fezes humanas. O acúmulo de resíduos sólidos (pneus, latas, móveis inutilizados) também atua como terreno fértil para vários vetores de doenças infecciosas, tendo em vista que o não recolhimento desses itens prejudica a drenagem adequada da água e gera criadouros de mosquitos como o *Aedes aegypti*.

Segundo pesquisas do Centers for Disease Control (2017), aproximadamente 75% das doenças infecciosas emergentes se originam de animais para humanos e surgem diretamente das interações homem-animal — são as chamadas doenças zoonóticas. Várias dessas doenças ganharam manchetes nos últimos anos: Zika, Ebola, SARS, gripe aviária, MERS e, mais recentemente, COVID-19. São fungos, vírus e bactérias trazidos por animais frequentemente culpados pelas epidemias decorrentes do contágio, quando na verdade são resultado das mudanças ambientais que estão forçando humanos e animais a interagirem de novas maneiras.

O Programa das Nações Unidas para Meio Ambiente (PNUMA) identifica o desmatamento, a agricultura intensiva, o comércio ilegal e o regulamento precário da vida selvagem, além das mudanças climáticas, como fontes centrais de surtos zoonóticos (PNUMA, 2016).

Processo de contágio de vírus por causa da degradação ambiental - *spillover*

Novas doenças geralmente emergem em *hotspots* de biodiversidade, como florestas tropicais. Por esse motivo, a perda de florestas é uma das formas mais significativas de degradação ambiental, facilitando novas interações entre humanos e animais. Muitas vezes são morcegos, camundongos e mosquitos que permanecem inócuos até o equilíbrio do

³ <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7119955/>

ecossistema onde vivem ser perturbado e esses animais começarem a causar doenças zoonóticas a humanos. (Centers for Disease Control, 2017, Bloomfield et al., 2020).

Embora há muito tempo as pesquisas científicas tenham demonstrado a ligação entre as interações humano-animal e a transmissão de doenças entre espécies (muitas vezes negligenciadas), só recentemente as condições mais amplas que facilitam, aceleram e potencializam essas relações em determinadas áreas estão sendo vistas com mais atenção.

A atual urbanização mundial apresenta importantes desafios e oportunidades a serem desbravados. Pobreza urbana, megacidades, degradação ambiental e deterioração social podem até parecer assustadores num primeiro momento, no entanto, são imperativos humanitários que precisam ser encarados para se criar cidades mais habitáveis e sustentáveis a longo prazo. Para isso, um bom caminho possível inclui: planejamento urbano, compartilhamento de experiências bem-sucedidas, estudos de modelos de sucesso e trocas de conhecimentos entre profissionais de todas as disciplinas, setores, países e comunidades de diferentes partes do mundo. Isso é importante, pois o desmatamento está diretamente ligados à fragmentação do habitat e ao salto de novos patógenos entre diferentes espécies hospedeiras (incluindo humanos e animais silvestres), um evento chamado *spillover*.

O *spillover* de patógenos é o processo pelo qual um patógeno é transmitido de uma espécie hospedeira reservatório para uma espécie hospedeira receptora e pode causar novos contágios a seres humanos. Nesse processo, o patógeno deve superar muitas barreiras físicas, moleculares e ecológicas, pulando entre diferentes hospedeiros, até encontrar uma espécie a partir da qual migrará para outras áreas e se desenvolverá em outros receptores. Essa dispersão é um evento complexo que depende da distância filogenética entre os hospedeiros, da frequência e intensidade do contato entre as espécies e de fatores genéticos do patógeno e do hospedeiro.

Alguns patógenos podem contaminar uma grande variedade de hospedeiros e se adaptar facilmente a novos hospedeiros, incluindo humanos. Características especiais, como o genoma de RNA (altos ritmos de mutação) ou transmissão vetorial, concedem aos patógenos maior facilidade para contaminar novos hospedeiros e achar novos nichos ecológicos, crescendo a transmissibilidade entre humanos. São essas transmissões mediadas por vetores que sustentam epidemias virais em humanos. A disseminação endêmica da

dengue no Brasil, da malária na Amazônia e do vírus Zika na América Latina ilustram esse ponto

As regiões de fronteira do ecossistema têm sido propostas como *hotspots* espaciais do *spillover* (DESPOMMIER et al. 2006). Essa hipótese é baseada principalmente sobre a ocorrência de doenças zoonóticas e taxas mais altas de disseminação em áreas fragmentadas e onde os humanos vivem próximos à vida *selvagem*. Juntamente com fatores como concentração populacional, acredita-se que a proximidade com os limites do ecossistema intercede a velocidade e o risco de propagação de doenças infecciosas. A teoria ecológica dos efeitos de borda (FAGAN et al., 1999) também prevê um crescimento da biodiversidade nas fronteiras do ecossistema, incluindo a existência de espécies em territórios específicos. A taxa de disseminação através dos limites do ecossistema depende da probabilidade de que os hospedeiros fonte e receptor, bem como o vírus, estejam presentes em ou perto de uma região de fronteira.

A probabilidade de transmissão pelo *spillover* é determinada pela interação entre vários fatores, incluindo a ação do vírus no hospedeiro reservatório e sua interação com o hospedeiro final. Na primeira etapa, a quantidade de vírus que está disponível para o hospedeiro humano em um determinado ponto no espaço e no tempo é determinada pela interação entre a distribuição do hospedeiro reservatório e a sobrevivência, mutação e disseminação do vírus fora do hospedeiro do reservatório. Em segundo lugar, a exposição ao patógeno é determinada pelo comportamento humano e pelos vetores, especialmente a oportunidade, caminho e quantidade de exposição. Terceiro, as características genéticas, fisiológicas e imunológicas do receptor humano bem como a dose de exposição intensifica a probabilidade e a gravidade do contágio.

Vulnerabilidade socioambiental e a relação entre doenças por água contaminada

Para Jacobi (1995), havia uma ligação direta entre exposição a riscos ambientais e acesso inseguro aos serviços públicos. Falta de infraestrutura urbana expõem as pessoas que vive nessas áreas a riscos ambientais, como doenças de veiculação hídrica.

A vulnerabilidade socioambiental é definida como a coexistência ou sobreposição espacial entre grupos populacionais muito pobres e muito desfavorecidos (vulnerabilidade social) e áreas de risco ou degradação ambiental (vulnerabilidade ambiental). Nesse sentido, é a mistura desses dois aspectos que é considerada uma condição de vulnerabilidade no meio social. Não é por acaso que as áreas de risco e degradação ambiental são também, na maioria das ocorrências, áreas de pobreza e privação social. Assim, nossa hipótese é que a vulnerabilidade ambiental é um componente relevante na configuração da distribuição espacial das condições de pobreza e privação social na cidade paulista. Apesar da melhora dos indicadores sociais médios das periferias nas últimas duas décadas, verifica-se a existência de grandes diferenças nos status de vida e acesso aos serviços públicos, com o comportamento de bairros de extrema pobreza carentes de serviços distribuídos por toda a periferia metropolitana. (TORRES e MARQUES, 2001).

Segundo dados da PNAD-IBGE de 1998, existe um grande contingente de pessoas próximo a 1,7 milhão de pessoas (10 % de pessoas da RMSP), com renda familiar inferior a dois salários mínimos, que não têm acesso à habitação em áreas tradicionalmente periféricas e mesmo em favelas consolidadas, forçadas a habitar nas franjas e interstícios urbanos mais instáveis, habitualmente em áreas de risco ambiental, com indicadores sociais e de saúde baixos (TORRES e MARQUES, 2001).

O nível de questões sociais e ambientais em algumas áreas é impressionante, sobrepondo-se, em termos espaciais (e sociais), aos piores indicadores socioeconômicos com riscos de inundações e deslizamentos de terra, ambiente intensamente contaminado e serviços sociais enormemente precários. (TORRES et al., 2003, p. 5).

Assim, em algumas áreas da periferia há intensa densidade de indicadores negativos, que indicam a presença de “pontos críticos” de vulnerabilidade social (e ambiental),

mostrando a existência de uma espécie de periferia da periferia (TORRES e MARQUES, 2001).

As regiões de risco ambiental (definidas como setores censitários localizados a menos de 100 metros dos rios, concentram maior população pobre e domicílios com condições sanitárias mais instáveis “principalmente acesso à rede de esgoto” Viver próximo a cursos d’água implica exposição real a diversos riscos ambientais, pois além das inundações (sazonais e de extensão geográfica limitada), há contato direto com água contaminada e exposição a vetores de doenças de veiculação hídrica (TORRES, 1997).

A categoria vulnerabilidade

O conceito de vulnerabilidade é normalmente definido como uma condição em que existem dois elementos: falta de resposta e a dificuldade de adaptação à percepção do risco (MOSER, 1998)

Vulnerabilidade às fontes de recursos hídricos: falta de água potável, condições anti-higiênicas e exposição às doenças trazidas pela água. A população pobre geralmente não tem acesso a saneamento adequado (água e esgoto) e, muitas vezes, é forçada a residir em áreas expostas a altos níveis de poluição hídrica. Estima-se que 20% da população mundial não tem acesso a água potável e 50% não dispõe de saneamento adequado. Doenças de veiculação hídrica representam séria ameaça à saúde humana, principalmente para as crianças, que são as mais vulneráveis a estas doenças (IHDP, 2001).

Tendo em vista a precariedade e vulnerabilidade social de populações que habitam as margens das represas no município de São Paulo, faz sentido investigar se há relação entre habitar tais áreas e a ocorrência de internações por causas relacionadas ao saneamento ambiental inadequado.

Objetivo da pesquisa

O presente trabalho tem como objetivo geral verificar a relação entre proximidade das represas e a ocorrência de internações hospitalares por doenças causadas por saneamento

básico inadequado. Outrossim, a pesquisa também pretende rever as técnicas estatísticas para processos pontuais, almejando identificar as análises mais propícias para identificarmos os padrões espaciais de doenças infecciosas emergentes na Subprefeitura Capela do Socorro

Local de estudo, sua história e formação

A área de estudo escolhida foi o distrito do Grajaú, localizado dentro da região administrativa da Zona Sul de São Paulo. Seus limites são os distritos de Pedreira, Cidade Dutra, Parelheiros e os municípios de São Bernardo do Campo e Diadema

A região foi escolhida devido à sua localização geográfica representativa: além de atuar como fronteira entre a área urbana e a área verde, que atualmente encontra-se inserida na área de proteção aos mananciais, tem grande relação com a represa Billings, localizada na borda leste da área de estudo.

Dados

Os dados utilizados são referentes a Base de dados geocodificados pelo CEP do endereço dos casos referente a Autorizações de Internações Hospitalares (AIH) AIH 2006 ao 2010 da Região Metropolitana de São Paulo, e aos dados do Censo Demográfico de 2010, (https://centrodametropole.fflch.usp.br/pt-br/download-de-dados?f%5B0%5D=facets_temas%3Asaude) selecionando crianças de 0 a 5 anos,

As doenças consideradas foram baseadas no livro “Variação geográfica do saneamento básico no Brasil em 2010: domicílios urbanos e rurais”, publicado pela Embrapa.

No capítulo 8 são abordadas as Doenças Relacionadas ao Saneamento Ambiental Inadequado no Brasil, segundo tabela 1:

Categoria	Grupos de doenças
Doenças de transmissão feco-oral	1. Diarreias 1.1 Cólera 1.2 Salmonelose 1.3 Shigelose 1.4 Outras infecções intestinais bacterianas (<i>Escherichia coli</i> , <i>Campylobacter</i> ssp., <i>Yersinia enterocolitica</i> , <i>Clostridium difficile</i> , outras e as não especificadas - NE) 1.5 Amebíase 1.6 Outras doenças intestinais por protozoários (Balantidíase, Giardíase, Criptosporídíase) 1.7 Isosporíase, outras e as NE 1.8 Doenças intestinais por vírus (enterite por rotavírus, gastroenteropatia aguda p/agente de Norwalk, enterite por adenovírus, outras enterites virais e as NE) 2. Febres entéricas 2.1 Febre tifóide 2.2 Febre paratifóide 3. Hepatite A
Doenças transmitidas por inseto vetor	4. Dengue 5. Febre Amarela 6. Leishmanioses 6.1 Leishmaniose tegumentar 6.2 Leishmaniose visceral 7. Filariose linfática 8. Malária 9. Doença de Chagas
Doenças transmitidas através do contato com a água	10. Esquistossomose 11. Leptospirose
Doenças relacionadas com a higiene	12. Doenças dos olhos 12.1 Tracoma 12.2 Conjuntivites 13. Doenças da pele

	<p>13.1 Dermatofitoses (Tinha da barba e do couro cabeludo, Tinha das unhas, Tinha da mão, Tinha dos pés, Tinha do corpo, Tinha imbricada, Tinea cruris, outras dermatofitoses e as NE)</p> <p>13.2 Outras micoses superficiais (Pitirias versicolor, Tinha negra, Piedra branca, Piedra negra, outras e as NE)</p>
Geohelmintos e teníases	<p>14. Helmintíases</p> <p>14.1 Equinococose</p> <p>14.2 Ancilostomíase</p> <p>14.3 Ascaridíase</p> <p>14.4 Estrongiloidíase</p> <p>14.5 Tricuríase</p> <p>14.6 Enterobiase</p> <p>15. Teníases</p> <p>15.1 Teníase</p> <p>15.2 Cisticercose</p>

Fonte: Adaptado de COSTA *et al.* (2002, p. 3).

Tabela 1: Categoria e grupos de Doenças Relacionadas ao Saneamento Ambiental Inadequado no Brasil

Analisando os dados Base de dados geocodificados das Autorizações de Internações Hospitalares “AIH” a única doença com número expressivo de casos foi a CID A09 – Diarreia e Gastroenterite de Origem Infecciosa Presumível como demonstrado na tabela 2.

	DIAG_PRINC	DESCRICAO	n
1	J180	Broncopneumonia não especificada	3407
2	J459	Asma não especificada	1274
3	J128	Outras pneumonias virais	912
4	P073	Outros recém-nascidos de pré-termo	877
5	A09	Diarréia e gastroenterite de origem infecciosa presu...	629
6	P221	Taquipnêia transitória do recém-nascido	529
7	P220	Síndrome da angústia respiratória do recém-nascido	478
8	N47	Hipertrofia do prepúcio, fimose e parafimose	436
9	P599	Icterícia neonatal não especificada	404
10	K429	Hérnia umbilical sem obstrução ou gangrena	402
11	J219	Bronquite aguda não especificada	392
12	J158	Outras pneumonias bacterianas	303
13	J353	Hipertrofia das amígdalas com hipertrofia das aden...	297
14	P399	Infecção própria do período perinatal não especifica...	253
15	G409	Epilepsia, não especificada	243
16	J960	Insuficiência respiratória aguda	228
17	K409	Hérnia inguinal unilateral ou não especificada, sem ...	215
18	P285	Insuficiência respiratória do recém-nascido	210
19	N390	Infecção do trato urinário de localização não especi...	203
20	J189	Pneumonia não especificada	171

Tabela 2: Casos de Autorizações de Internações Hospitalares (AIH) de crianças de 0 a 5 anos da Subprefeitura Capela do Socorro

Metodologia

John Snow foi um médico inglês considerado o pai da epidemiologia moderna. Com o surto de cólera na cidade de Londres em 1854, Snow criou um estudo que foi um dos exemplos mais famosos de análise espacial. Ele usou um mapa para mostrar que as mortes por cólera se concentravam em áreas específicas — principalmente perto das bombas de água pública, onde a maioria da população frequentava para beber água e lavar roupa.

Este caso é um bom exemplo de como o mapeamento pode representar de forma clara e objetiva situações diversas, incluindo aspectos da natureza, delimitação de áreas de risco e localização de doenças, contribuindo para a busca de possíveis soluções. Por essa razão, o presente estudo busca compreender a relação entre patógenos, seus hospedeiros e o ambiente a fim de compreender o que pode inibir ou estimular sua propagação.

Em geral, espera-se que, identificando os elementos dessa relação, seja possível reconhecer e controlar a epidemia causada pela doença, assim como compreender mais a fundo cada uma dessas variáveis, que têm sido um dos principais focos da Cartografia da Saúde. Tais medidas são necessárias para entender os padrões de transmissão da doença e os caminhos para sua inibição.

As doenças contemporâneas, especialmente as catalogadas no século XXI, surgem em um momento em que novas tecnologias de mapeamento proporcionam estudos mais específicos e detalhados sobre saúde, doenças e afecções. Os sistemas de informação geográfica (SIG), por exemplo, em conjunto com a “cartografia tradicional”, permitem não só a definição de análises gráficas e cartográficas, como também a produção de dados estatísticos importantes. Seus principais pontos fortes são: clareza na apresentação dos dados e relações mais precisas. No caso de pesquisas médicas, é possível estabelecer relações entre uma dada doença e sua população hospedeira com a finalidade de gerar indicadores que podem embasar decisões rápidas e seguras.

Fazer mapas é como escrever uma história que pode ser lida, interpretada, compartilhada e modificada. Não existe cartografia sem um pensamento lógico, uma narrativa por trás. E mais: o pensamento cartográfico inevitavelmente convida a um desfecho concreto. Ou seja, mapear significa localizar, descrever e associar elementos para responder a uma questão real. É por isso que usar essa abordagem para compreender as interações entre patógeno e hospedeiro no ambiente é geograficamente relevante.

A epidemiologia espacial é a descrição e análise da variação geográfica de determinada doença e envolve fatores como riscos demográficos, ambientais, comportamentais, socioeconômicos, genéticos e infecciosos. A disseminação de doenças infecciosas está intimamente relacionada à ideia de proximidade espaço-temporal; nesse sentido, abranger o fator de risco ambiental é igualmente necessário. Desse modo, a

compreensão das variações espaciais e temporais da doença e a caracterização de sua estrutura espacial contribuem para que os epidemiologistas tomem medidas cabíveis.

Existe uma série de métodos usados pela epidemiologia espacial com o objetivo de descrever padrões espaciais de doenças, identificar grupos de enfermidades e gerar explicações ou previsões.

A estatística espacial é utilizada para fazer inferência para tentar tirar conclusões sobre processos espaciais e seus processos de geração e assim conseguir comportamentos futuros. Seguindo essa funcionalidade, a análise foi dividida da seguinte forma:

- 1) Análise de agrupamento espacial:
 - a) Função K de Ripley
 - b) Função F - Espaços vazios
 - c) Função G - Vizinho mais próximo
 - d) Mapa de Kernel
- 2) Autocorrelação espacial:
 - a) Índice Global de Moran
 - b) Correlograma de Distância
 - c) Índice Locais de associação Espacial (LISA)
- 3) Regressão espacial:
 - a) Regressão Geograficamente Ponderada - GWR

Para todas as análises propostas, foi utilizado o pacote estatístico R (R Core Team, 2012). O R é um software livre para análises estatísticas baseada em linha de comando que, devido à sua natureza de fácil entendimento e disponibilidade do código fonte, possibilitou que fossem criados diferentes pacotes complementares visando incrementar ainda mais as possíveis análises. Para as análises de processos pontuais, foi, portanto, necessário utilizar o

pacote estatístico 'spatstat'. Para dependência espacial foi utilizado o pacote estatístico 'spdep'. Para Correlograma de distância foi utilizado o pacote estatístico 'pgrmess' e para regressão geograficamente ponderada foi utilizado o pacote 'spgwr'.

1) Análise de agrupamento espacial

A análise de agrupamento espacial é um conjunto de técnicas usadas para identificar padrões de agrupamento ou agrupamento em dados espaciais. É usado para identificar áreas onde a concentração de um determinado fenômeno, como crime, doença ou espécie, é maior do que o esperado ao acaso. Este método agrupa pontos que estão próximos uns dos outros no espaço e na densidade. Esses métodos podem ser usados para identificar grupos de pontos que são mais densamente agrupados do que seria esperado ao acaso e para mapear a distribuição espacial do fenômeno. Os resultados podem ser usados para identificar áreas onde é necessária intervenção ou estudo mais aprofundado e para planejar intervenções ou alocação de recursos. (Spatial Point Patterns book. Methodology and Applications with R. By Adrian Baddeley, Ege Rubak, Rolf Turner.)

Buscando entender o processo, primeiro foi utilizado uma análise de agrupamento buscando entender como se encontra o padrão de agregação dos pontos, onde refere-se à distribuição de um conjunto de pontos em uma determinada área espacial. Descreve a maneira como os pontos são agrupados ou dispersos dentro de uma área de estudo. Existem vários tipos de padrões de agregação de pontos, incluindo:

- Agrupado: Os pontos estão próximos uns dos outros e formam grupos ou *clusters* distintos
- Regular: os pontos são espaçados uniformemente e formam um padrão regular, como uma grade.
- Aleatório: os pontos são distribuídos aleatoriamente e não exibem nenhum padrão perceptível

Os padrões de agregação de pontos podem ser analisados usando vários métodos estatísticos, como a função K de Ripley, Função F, Função G para determinar o grau de agrupamento ou dispersão. Esses padrões podem fornecer informações sobre os processos subjacentes que os geraram

Padrões de Agregação

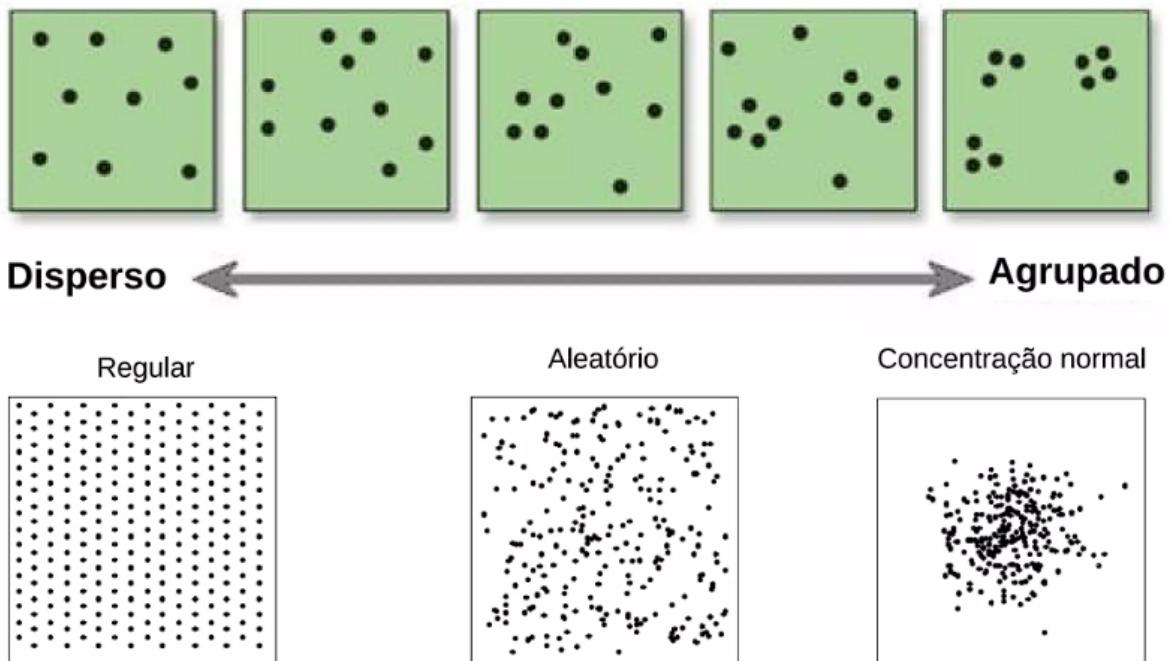


Figura 1: padrões de agregação fonte:(Spatial Point Patterns book. Methodology and Applications with R. By Adrian Baddeley, Ege Rubak, Rolf Turner.)

A) Função K de Ripley

A função K de Ripley é uma ferramenta estatística usada para medir o agrupamento ou a tendência de agrupamento de um conjunto de pontos em uma determinada distribuição espacial. É frequentemente usado em estatísticas espaciais e análise de padrão de pontos para determinar se um conjunto de pontos é distribuído aleatoriamente ou se há grupos de pontos que são mais densamente agrupados do que seria esperado por acaso. A função K é calculada

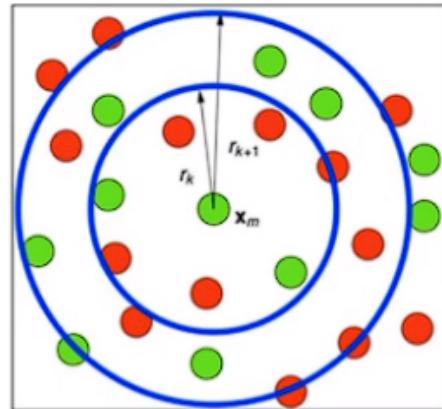
comparando o número observado de pares de pontos dentro de uma determinada distância com o número de pares de pontos que seriam esperados em uma distribuição aleatória. Se o número observado de pares de pontos for maior que o número esperado, os pontos serão considerados agrupados.

A função K de Ripley, também conhecida como análise de padrão de ponto de segunda ordem, é um método estatístico usado para medir o agrupamento ou a tendência de agrupamento de um conjunto de pontos em uma determinada distribuição espacial. A função compara o número observado de pares de pontos dentro de uma certa distância com o número de pares de pontos que seriam esperados em uma distribuição aleatória. A função K é calculada contando o número de pares de pontos que estão dentro de uma certa distância um do outro, chamada de "largura de banda", e comparando isso com o número de pares de pontos que seriam esperados em uma distribuição aleatória. Se o número observado de pares de pontos for maior que o número esperado, os pontos serão considerados agrupados.

A função K pode ser usada para identificar padrões em dados pontuais, como *clusters*, regularidade e aleatoriedade. Também pode ser usado para comparar o agrupamento de diferentes conjuntos de dados de pontos. A função K , proposta por Ripley (1976), é definida como $K(t) = \lambda^{-1} E [número\ de\ eventos\ contidos\ dentro\ de\ uma\ distância\ t\ a\ partir\ de\ um\ evento\ arbitrário]$, em que E é o operador esperança e λ é a intensidade, ou número médio de eventos por unidade de área, em uma região assumida como constante. O mais simples e natural estimador de $K(t)$ que corrige o efeito de bordas é dado por:

Função K de Ripley

$$L(d) = \sqrt{\frac{A \sum_{i=1}^n \sum_{j=1, j \neq i}^n k(i, j)}{\pi n(n - 1)}}$$



d = distância

A = área de estudo

K(i,j) = peso -> se a distância < "d", então peso é um, senão o peso é zero
n = número total de pontos na área de estudo

Ripley, B.D. Modelling spatial patterns. *J. R. Stat. Soc. Series B Stat. Methodol.* 1977; 39: 172–192

Figura 2: Função K de Ripley

A saída da função K pode ser plotada como um gráfico, K(d) contra d, que é chamado de gráfico K de Ripley. A inclinação do gráfico estará relacionada ao agrupamento dos pontos. Se os pontos forem distribuídos aleatoriamente, a parcela será linear com inclinação de 0,5, se os pontos forem agrupados, a parcela terá uma inclinação maior que 0,5 e se os pontos forem dispersos, a parcela terá uma inclinação menor que 0,5 .

B) Função F - Espaços vazios

A função F, também conhecida como função de espaço vazio ou função *void*, é um método estatístico usado para medir a distribuição de espaço vazio ou não utilizado dentro de um padrão de pontos. É frequentemente usado em estatística espacial e análise de padrão de pontos para determinar se os pontos são distribuídos aleatoriamente ou se há grandes áreas de espaço vazio entre os pontos.(Spatial Point Patterns book. Methodology and Applications with R. By Adrian Baddeley, Ege Rubak, Rolf Turner.)

A função é calculada medindo a distância de cada ponto até o ponto mais próximo e contando o número de pontos que estão a pelo menos uma certa distância do ponto mais próximo, chamada de "largura de banda". A função pode ser definida matematicamente como:

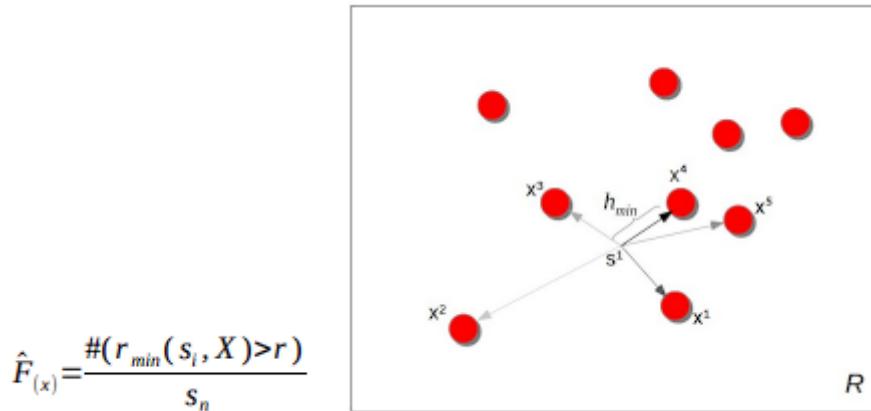


Figura 3: Esquema ilustrativo do cálculo de vizinhança da Função F.

Onde: # é “o número de eventos”, r_{min} é a menor distância da localidade si ao evento mais próximo do processo pontual X estudado, dividido pela quantidade de localidades aleatoriamente amostradas (s_n)(BAILEY e GATREL, 1995; LLOYD, 2007). Assim como na função G, o valor esperado de $\hat{F}()$ considerando a completa aleatoriedade espacial é igual a 1 (BAILEY e GATREL, 1995; LLOYD, 2007; PROTÁZIO, 2007)

Se os pontos forem distribuídos aleatoriamente, a função terá um valor próximo de 0 para todas as distâncias, indicando que não existem grandes áreas de espaço vazio entre os pontos. Se os pontos estiverem agrupados, a função terá um valor alto para distâncias pequenas, indicando que existem grandes áreas de espaço vazio dentro dos agrupamentos. Se os pontos estiverem dispersos, a função terá um valor alto para grandes distâncias, indicando que existem grandes áreas de vazio entre os pontos dispersos.

A função F pode ser usada para identificar padrões em dados pontuais, como *clusters*, regularidade e aleatoriedade, e também pode ser usada para comparar a distribuição do

espaço vazio em diferentes conjuntos de dados pontuais. Também pode ser usado para detectar mudanças na distribuição de pontos ao longo do tempo.

A função F é usada em combinação com a função K de Ripley e juntas fornecem uma imagem completa do padrão de pontos. Enquanto a função K de Ripley mede o agrupamento ou tendência de agrupamento dos pontos, a função F mede a distribuição do espaço vazio dentro do padrão de pontos.

C) Função G - Vizinho mais próximo

A função G, também conhecida como função do vizinho mais próximo, é um método estatístico usado para medir a distribuição de distâncias entre pontos em um padrão de pontos. É frequentemente usado em estatísticas espaciais e análise de padrão de pontos para determinar se os pontos são distribuídos aleatoriamente ou se há grupos de pontos que são mais densamente agrupados do que seria esperado por acaso.

A função é calculada medindo a distância de cada ponto ao vizinho mais próximo e contando o número de pontos que estão a pelo menos uma certa distância do vizinho mais próximo, chamada de "largura de banda". A função pode ser definida matematicamente como:

$$\hat{G}_{(r)} = \frac{\#(r_{\min}(x_i) > r)}{n}$$

Onde: # é “o número de ventos”, r_{\min} é a menor distância do evento x_i ao evento vizinho mais próximo, dividido pela quantidade de eventos observados (n) (BAILEY e GATREL, 1995; LLOYD, 2007). Assim, o valor esperado de $\hat{G}()$ assumindo a completa aleatoriedade espacial é igual a 1 (BAILEY e GATREL, 1995; LLOYD, 2007; PROTÁZIO, 2007). Como se pode perceber, tal definição em pouco se distingue da equação da função F , uma vez que ambas as funções estimam as distâncias médias entre vizinhos mais próximos. A única diferença, portanto, é a definição conceitual, onde a função $\hat{F}()$ estima distância média do vizinho mais próximo de uma localidade da área de estudo, e não entre eventos observados

Se os pontos forem distribuídos aleatoriamente, a função terá um valor próximo de 0 para todas as distâncias, indicando que os pontos estão igualmente espaçados. Se os pontos estiverem agrupados, a função terá um valor alto para pequenas distâncias, indicando que os pontos dentro dos clusters estão mais densamente agrupados do que o esperado ao acaso. Se os pontos estiverem dispersos, a função terá um valor alto para grandes distâncias, indicando que os pontos estão mais espaçados do que o esperado pelo acaso.

A função G pode ser usada para identificar padrões em dados pontuais, como *clusters*, regularidade e aleatoriedade, e também pode ser usada para comparar a distribuição de distâncias entre pontos em diferentes conjuntos de dados pontuais. Também pode ser usado para detectar mudanças na distribuição de pontos ao longo do tempo

A função G é usada em combinação com a função K de Ripley e juntas fornecem uma imagem completa do padrão de pontos. Enquanto a função K de Ripley mede o agrupamento ou tendência de agrupamento dos pontos, a função G mede a distribuição das distâncias entre os pontos.

2) Autocorrelação espacial

A autocorrelação espacial refere-se à correlação entre os valores de uma variável em diferentes localizações no espaço. Descreve o grau em que os valores de uma variável são semelhantes entre locais próximos. A autocorrelação espacial pode ser positiva ou negativa, indicando se valores semelhantes tendem a ser encontrados próximos ou distantes, respectivamente.

Existem várias medidas de autocorrelação espacial, incluindo:

- A) Moran's I: Esta medida varia de -1 a 1 e indica o grau de autocorrelação espacial positiva ou negativa. Um valor de 1 indica autocorrelação espacial positiva perfeita, um valor de -1 indica autocorrelação espacial negativa perfeita e um valor de 0 indica nenhuma autocorrelação espacial.
- B) Indicador Local de Associação Espacial (LISA): Esta medida identifica agrupamentos locais de valores altos ou baixos e é usada para identificar áreas de autocorrelação espacial alta ou baixa.

A autocorrelação espacial pode ser causada por vários fatores, como dependência espacial, interação espacial ou heterogeneidade espacial. A dependência espacial refere-se à relação entre os valores de uma variável em locais diferentes devido à proximidade dos locais. A interação espacial refere-se à relação entre os valores de uma variável em locais diferentes devido à influência de um local sobre outro. A heterogeneidade espacial refere-se à relação entre os valores de uma variável em diferentes locais devido à variação nos processos subjacentes que geraram os valores.

Autocorrelação espacial é um conceito importante em estatística espacial, pode ser usado para identificar padrões em dados espaciais, como clusters, regularidade e aleatoriedade, e para comparar a similaridade de valores entre diferentes regiões. Também pode ser usado para detectar mudanças na distribuição de valores ao longo do tempo. Por exemplo, em epidemiologia, a autocorrelação espacial pode ser usada para identificar grupos de doenças e monitorar mudanças na distribuição de doenças ao longo do tempo.

A) Índice global de Moran

O Índice de Moran Global é uma medida de autocorrelação espacial usada para determinar o grau de similaridade ou dissimilaridade entre os valores de uma variável em diferentes localizações no espaço. O Índice de Moran Global é calculado como a razão da soma do produto dos valores das variáveis e seus pesos espaciais para a soma dos valores das variáveis, multiplicado pela soma dos pesos espaciais. Os pesos espaciais são tipicamente binários, indicando se duas unidades são vizinhas ou não, ou baseados na distância, indicando o grau de similaridade entre duas unidades com base em sua distância. É usado para determinar se valores semelhantes tendem a ser encontrados próximos ou distantes uns dos outros. O Índice de Moran Global varia de -1 a 1, onde um valor de 1 indica autocorrelação espacial positiva perfeita, um valor de -1 indica autocorrelação espacial negativa perfeita e um valor de 0 indica nenhuma autocorrelação espacial.

O Índice Global de Moran é calculado usando as seguintes etapas:

- A variável de interesse é medida em cada local da área de estudo.
- Uma matriz de pesos espaciais é criada para definir as relações entre os locais. A matriz pode ser baseada em distância, conectividade ou qualquer outro critério.

- A média da variável de interesse é calculada
- O desvio da variável de interesse em cada local da média é calculado.
- O produto do desvio da variável de interesse em cada local da média e da matriz de pesos espaciais é calculado.
- A soma do produto da etapa anterior é calculada.
- A variância da variável de interesse é calculada.
- O Índice de Moran Global é calculado como a razão da soma da etapa 6 para a variação da etapa 7.

Assim, o índice de Moran é comumente utilizado como medida da autocorrelação espacial para dados contínuos, de acordo com a seguinte expressão (Burt et al. 2009, p. 554):

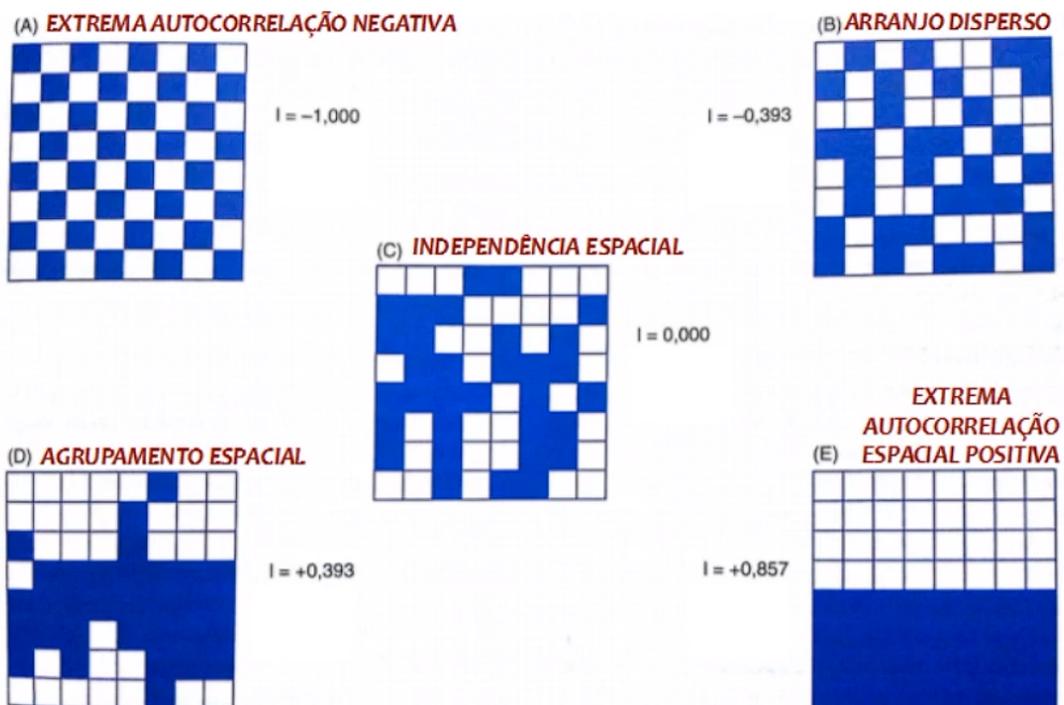
$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij}(X_i - \bar{X})(X_j - \bar{X})}{\left(\sum_{i=1}^n \sum_{j=1}^n w_{ij} \right) \sum_{i=1}^n (X_i - \bar{X})^2}$$

Onde n é o número de pontos de dados, X é a variável aleatória contínua e wij são os elementos da matriz W dos pesos espaciais que representa a relação geográfica entre todos os pares de pontos, segundo os autores mencionados.

O Índice Global Moran pode ser utilizado para identificar padrões de autocorrelação espacial em uma variável de interesse. A autocorrelação espacial positiva indica que valores semelhantes tendem a ser encontrados próximos uns dos outros, enquanto a autocorrelação espacial negativa indica que valores semelhantes tendem a ser encontrados distantes uns dos

outros. Um valor próximo de zero indica que não há autocorrelação espacial na variável de interesse.

Índice Global de Moran



Moran, P. A. P. (1950). "Notes on Continuous Stochastic Phenomena". Biometrika. 37 (1): 17–23.

Figura 4: Padrões de autocorrelação espacial

B) Correlograma de Distância

Um correlograma de distância de índice Moran é uma representação gráfica da autocorrelação espacial de uma variável em diferentes distâncias. Ele é criado plotando o Índice de Moran em diferentes distâncias (ou distâncias de defasagem) no eixo x e o valor I de Moran correspondente no eixo y.

O procedimento para criar um correlograma de distância de índice de Moran é:

- A variável de interesse é medida em cada local da área de estudo.
- Uma matriz de pesos espaciais é criada para definir as relações entre os locais. A matriz pode ser baseada em distância, conectividade ou qualquer outro critério.
- Para cada distância, é criada uma nova matriz de pesos espaciais que inclui apenas as relações entre os locais que são menores ou iguais à distância.
- O Índice Global de Moran é calculado usando a nova matriz de pesos espaciais.
- O valor calculado do Índice de Moran Global é plotado no eixo y e a distância no eixo x.

O gráfico resultante mostra como a autocorrelação espacial da variável muda à medida que a distância entre os locais aumenta. Se a variável for espacialmente autocorrelacionada, o gráfico mostrará um pico ou vale nas distâncias onde a autocorrelação espacial é mais forte. A forma do gráfico fornece uma indicação da escala na qual a autocorrelação espacial é mais forte. Um pico em pequenas distâncias sugere que a autocorrelação espacial é mais forte em pequenas distâncias, indicando que a variável está agrupada em uma escala fina, enquanto um pico em grandes distâncias sugere que a autocorrelação espacial é mais forte em grandes distâncias, indicando que a variável está agrupada em escala grosseira.

C) Índice Locais de associação Espacial (LISA)

O Índice Local de Moran (LISA) é uma medida de autocorrelação espacial usada para identificar agrupamentos locais de valores altos ou baixos de uma variável em uma área específica. É uma extensão do Índice Global de Moran (Moran's I) que fornece uma medida global de autocorrelação espacial, mas o LISA decompõe essa medida global em medidas locais, fornecendo uma imagem mais detalhada do padrão espacial da variável.

O LISA é calculado da seguinte forma:

- A variável de interesse é medida em cada local da área de estudo.

- Uma matriz de pesos espaciais é criada para definir as relações entre os locais.
A matriz pode ser baseada em distância,
- A média da variável de interesse é calculada
- O desvio da variável de interesse em cada local da média é calculado.
- O produto do desvio da variável de interesse em cada local da média e da matriz de pesos espaciais é calculado.
- A soma do produto da etapa anterior é calculada para cada local.

O mapa LISA resultante pode ser usado para identificar áreas de alta ou baixa autocorrelação espacial, onde os valores da variável tendem a ser semelhantes entre locais próximos. Esse mapa pode ser dividido em diferentes categorias, como Alto-Alto(HH), Alto-Baixo(HL), Baixo-Alto(LH) e Baixo-Baixo(LL) com base nos valores da variável.

O mapa LISA resultante pode ser usado para identificar áreas de alta ou baixa autocorrelação espacial, onde os valores da variável tendem a ser semelhantes entre locais próximos. Esse mapa pode ser dividido em diferentes categorias, como Alto-Alto(HH), Alto-Baixo(HL), Baixo-Alto(LH) e Baixo-Baixo(LL) com base nos valores da variável.

O Índice Global de Moran (Moran's I) fornece uma única estatística de resumo que descreve o nível geral de autocorrelação espacial em uma variável em toda uma área de estudo. Ele varia de -1 a 1, onde um valor de 1 indica autocorrelação espacial positiva perfeita, um valor de -1 indica autocorrelação espacial negativa perfeita e um valor de 0 indica nenhuma autocorrelação espacial. É calculado como a razão da soma dos produtos dos desvios da variável de sua média e da matriz de pesos espaciais para a variância da variável.

Por outro lado, o Índice de Moran Local (LISA) decompõe a medida global de autocorrelação espacial em medidas locais, fornecendo uma imagem mais detalhada do padrão espacial da variável. É calculado como a razão da soma dos produtos dos desvios da variável de sua média e da matriz de pesos espaciais para cada local para a variância da variável. É um mapa que pode ser utilizado para identificar áreas de alta ou baixa autocorrelação espacial, onde os valores da variável tendem a ser semelhantes entre localidades próximas.

Em resumo, enquanto o Índice Global de Moran (Moran's I) fornece uma única estatística resumida que descreve o nível geral de autocorrelação espacial em uma variável em toda uma área de estudo, o Índice Local de Moran (LISA) fornece uma imagem detalhada do padrão espacial de a variável dividindo a medida global de autocorrelação espacial em medidas locais.

3) Regressão espacial:

A regressão espacial é uma técnica estatística que permite a modelagem de relações espaciais entre variáveis. É usado para analisar a relação entre uma variável dependente e uma ou mais variáveis independentes, levando em consideração a dependência espacial entre as observações. Modelos de regressão espacial são usados para identificar e quantificar a influência de variáveis espaciais na variável dependente. Existem vários tipos de modelos de regressão espacial, incluindo:

Spatial lag model- Um modelo de defasagem espacial é um tipo de modelo estatístico usado para analisar a relação entre uma variável dependente e uma ou mais variáveis independentes, levando em consideração as relações espaciais entre as observações. O modelo de *spatial lag* inclui um termo adicional, chamado de *lag* espacial, que é uma soma ponderada dos valores da variável dependente para localizações vizinhas. O termo de defasagem espacial captura a ideia de que o valor da variável dependente para um determinado local é influenciado pelos valores da variável dependente para locais vizinhos. Os pesos usados no termo de defasagem espacial são normalmente baseados nas relações espaciais entre as observações, como a distância ou adjacência entre os locais.

Regressão Geograficamente Ponderada (GWR): Este modelo permite a estimativa de relações locais entre as variáveis, ao invés de uma única relação global.

A diferença entre o GWR e o *Spatial lag* é que o *spatial lag* considera a dependência espacial, incluindo um termo de atraso espacial na equação. O termo de *Spatial lag* representa a média ponderada da variável dependente para as observações vizinhas. Este modelo assume que a relação entre as variáveis independentes e dependentes é a mesma em

toda a área de estudo. Em outras palavras, assume que os coeficientes do modelo são os mesmos para todas as observações.

A Regressão Geograficamente Ponderada (GWR), por outro lado, permite a estimativa de relações locais entre as variáveis, em vez de uma única relação global. Ele começa ajustando um modelo OLS separado para cada local ou bairro na área de estudo, usando as observações próximas a esse local. O modelo estima os coeficientes para cada localidade separadamente, permitindo detectar a não estacionariedade, ou seja, a variação dos coeficientes ao longo da área de estudo. O GWR fornece uma imagem mais detalhada do padrão espacial da variável, considerando a não estacionariedade espacial da variável.

Em resumo, enquanto o modelo de defasagem espacial considera a dependência espacial incluindo um termo de defasagem espacial na equação e assume que os coeficientes do modelo são os mesmos para todas as observações, o GWR permite a estimativa de relações locais entre as variáveis, em vez de um relacionamento global único ajustando um modelo separado para cada local ou bairro na área de estudo. Por esse motivo foi escolhido o modelo GWR como o mais adequado para as análises

A) Regressão Geograficamente Ponderada - GWR

A regressão geograficamente ponderada (GWR) é uma das várias técnicas de regressão espacial usadas em geografia e outras disciplinas para abordar a heterogeneidade espacial. O GWR avalia um modelo local da variável ou processo que estamos tentando entender ou prever ajustando uma equação de regressão para cada recurso no conjunto de dados. O GWR constrói essas equações separadas incorporando as variáveis dependentes e explanatórias dos recursos que se enquadram na vizinhança de cada recurso de destino ou dos k vizinhos mais próximos. O GWR deve ser aplicado a conjuntos de dados com mais de centenas de recursos. Não é um método apropriado para pequenos conjuntos de dados e não funciona com dados multiponto.

O modelo GWR começa ajustando um modelo de regressão de mínimos quadrados ordinários (OLS) separado para cada local ou vizinhança na área de estudo, usando as observações próximas a esse local. O modelo estima os coeficientes para cada localidade separadamente, permitindo detectar a não estacionariedade, ou seja, a variação dos coeficientes ao longo da área de estudo. O modelo GWR é definido como:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

Onde a Variável Dependente (y): é o fenômeno, o objeto, e o que você está querendo relacionar. Variáveis explicativas (x): São as variáveis que influenciam diretamente, ou que hipoteticamente ajudam a explicar a variável dependente. Coeficientes (β): São os valores, calculados pela ferramenta de regressão, que indicam a relação e o peso de cada variável explicativa para a variável dependente; Resíduos: (ε): a parte da variável dependente que não é explicada pelo modelo.

O GWR supera a limitação do modelo de regressão OLS de gerar um conjunto global de estimativas. A ideia básica por trás do GWR é examinar a maneira pela qual as relações entre uma variável dependente e um conjunto de preditores podem variar no espaço. O GWR opera movendo uma janela de pesquisa de um ponto de regressão para o próximo, trabalhando sequencialmente por todos os pontos de regressão existentes no conjunto de dados. Um conjunto de regiões é então definido em torno de cada ponto de regressão e dentro da janela de pesquisa. Um modelo de regressão é então ajustado a todos os dados contidos em cada uma das regiões identificadas em torno de um ponto de regressão, com pontos de dados mais próximos do ponto de amostragem sendo mais ponderados do que aqueles mais distantes. Este processo é repetido para todos os pontos de amostra no conjunto de dados. Para um conjunto de dados de 150 observações, o GWR ajustará 150 modelos de regressão ponderados.

Graficamente, o GWR envolve o ajuste de um kernel espacial aos dados conforme descrito na imagem abaixo. Para um determinado ponto de regressão x , o peso (C) de um ponto de dados está no máximo no local do ponto de regresso. O peso diminui gradualmente à medida que a distância entre dois pontos aumenta. Um modelo de regressão é assim calibrado localmente movendo o ponto de regressão ao longo da área em estudo. Para cada local, os dados são ponderados de forma diferente para que as estimativas resultantes sejam exclusivas de um determinado local.

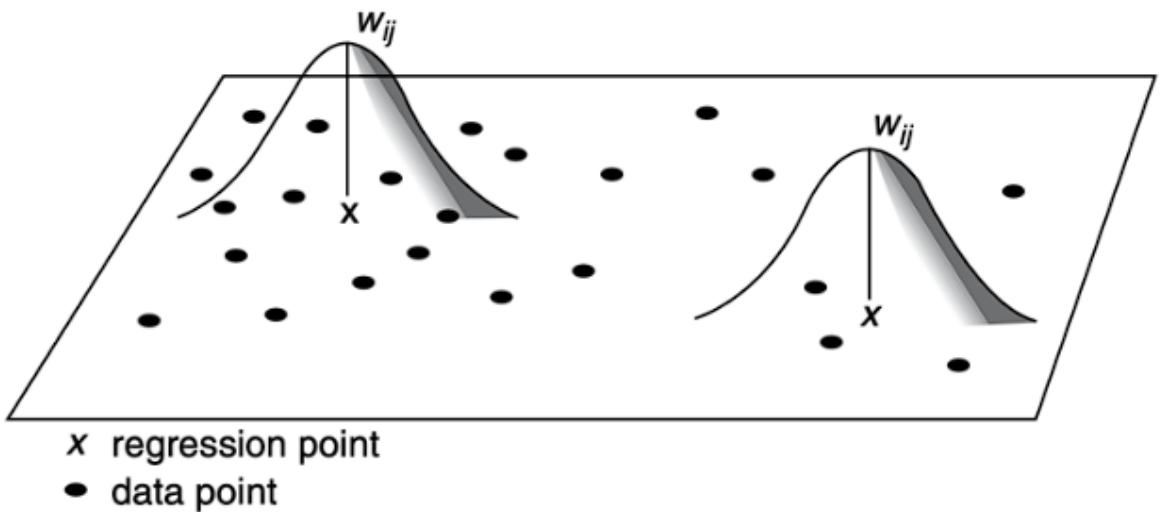


Figura 5 : Fonte: Fotheringham et al. (2002, 45)

Resultados

Para compreender os resultados é preciso entender o funcionamento dos gráficos, a função K e G tem um comportamento semelhante, onde o índice positivo mostra o agrupamento e o índice negativo mostra segregação, a faixa próximo a linha zero indica um comportamento aleatório

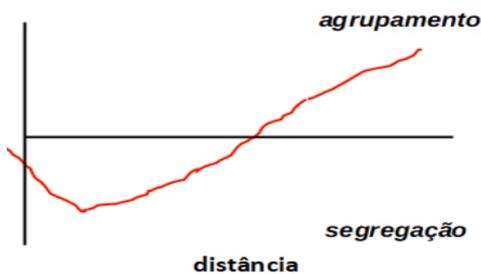


Figura 6

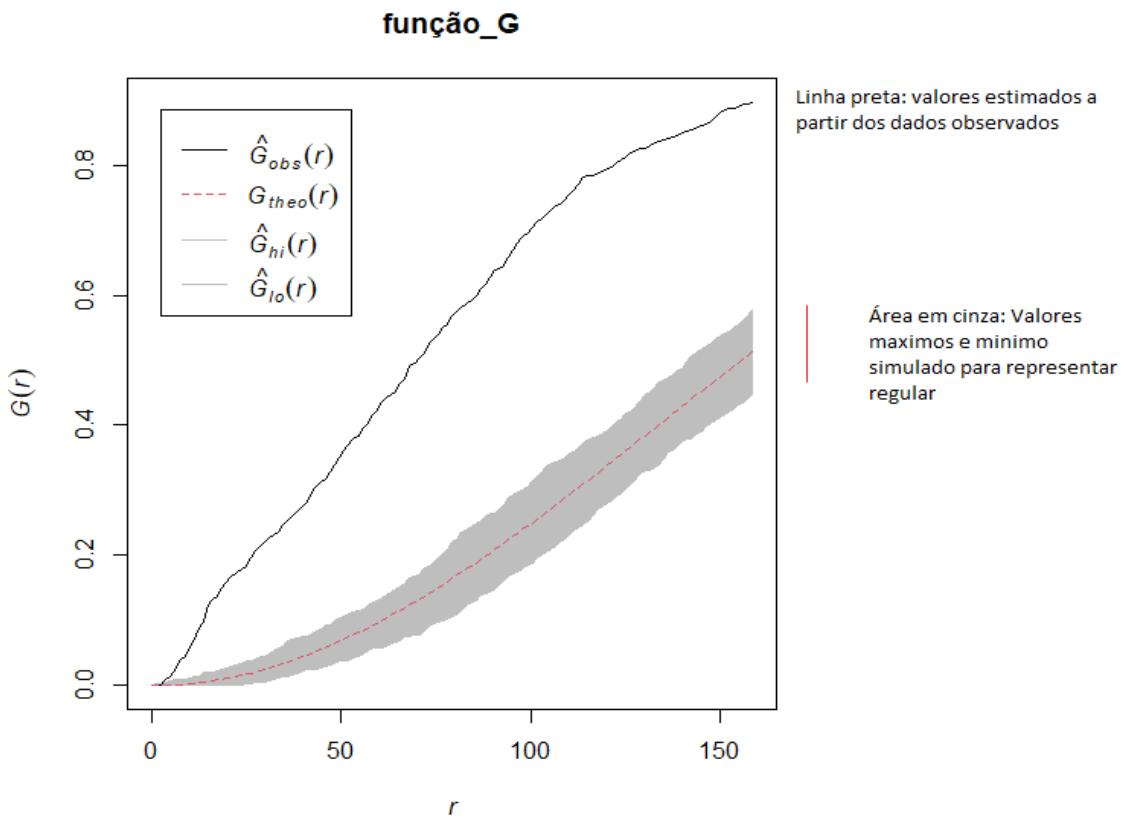


Figura 7: Função G - no eixo das ordenadas representa o $G(r)$, taxa referente a função G, no eixo das abscissas representa a distância em metros

Olhando para o gráfico figura 7, podemos ver que os valores observados de $G(r)$ são maiores que os valores esperados em pequenas distâncias, indicando agrupamento de pontos no padrão espacial. Isso se reflete no desvio positivo entre os valores observados e esperados de $G(r)$.

Podemos ver que o valor de $G(r)$ inicialmente diminui rapidamente, indicando um forte agrupamento de pontos no padrão. Porém, por volta dos 200 metros, a curva começa a se nivelar e se torna quase constante, indicando uma distribuição aleatória ou uniforme de pontos em distâncias maiores.

Isso sugere que o padrão de pontos exibe uma forte tendência de agrupamento em pequenas distâncias, mas torna-se mais aleatório em distâncias maiores. O ponto de inflexão onde a curva se nivela pode ser indicativo de um processo espacial ou uma característica geográfica que afeta a distribuição dos pontos na área.

Vale notar que o valor de $G(r)$ nunca chega a zero, indicando que sempre existem alguns pares de pontos que estão mais próximos que a distância r . Isso contrasta com a função K de Ripley, que pode chegar a zero se houver aleatoriedade espacial completa no padrão de pontos.

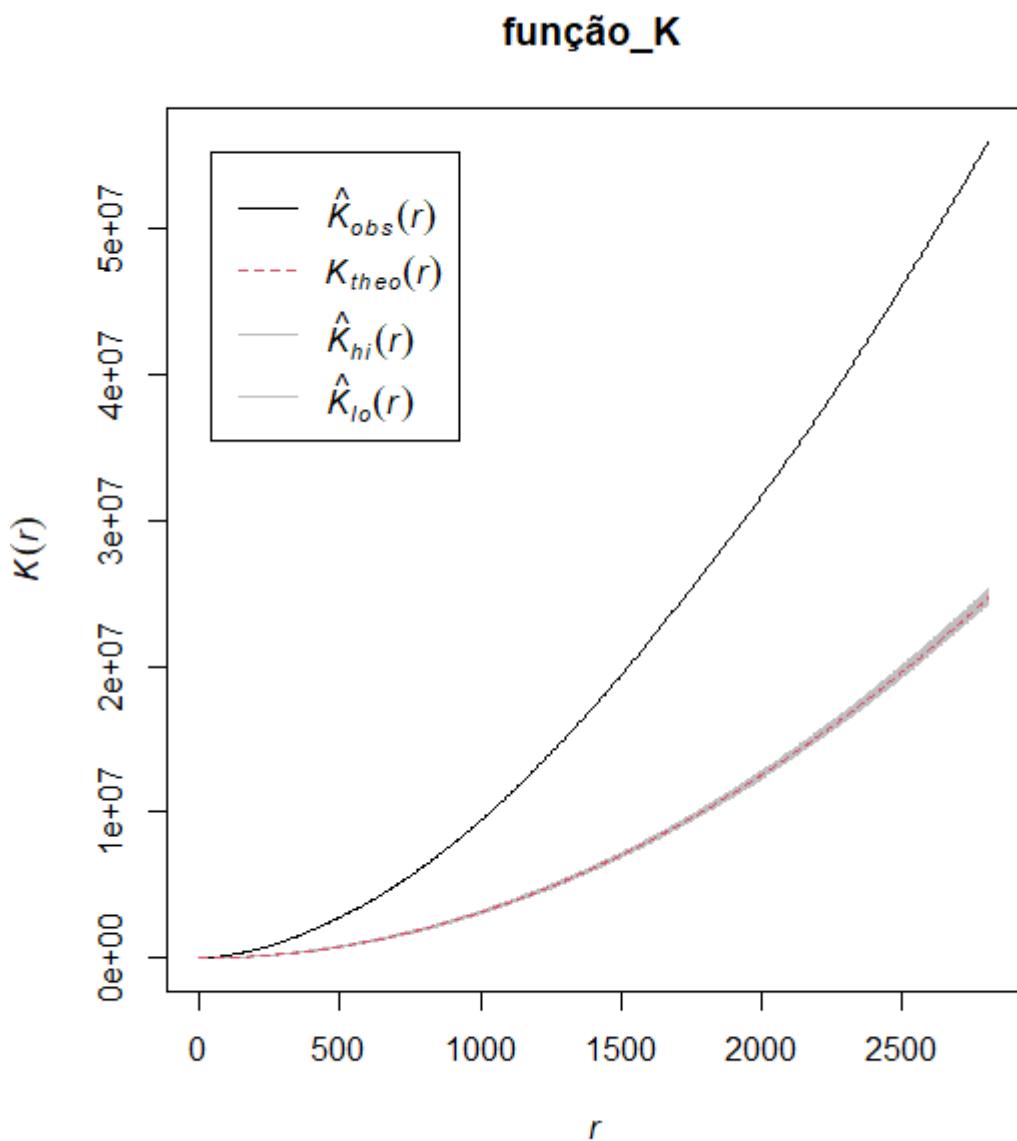


Figura 9: Função K - no eixo das ordenadas representa o $K(r)$, taxa referente a função K , no eixo das abscissas representa a distância em metros

Neste gráfico figura 9 em particular, podemos observar que o valor de $K(r)$ começa abaixo da linha diagonal, o que indica que os pontos estão agrupados a uma distância inferior a 100 metros. À medida que a distância aumenta, o valor de $K(r)$ se move acima da linha diagonal, o que significa que os pontos estão mais dispersos nessa faixa de distância.

Após uma distância de aproximadamente 300 metros, o valor de $K(r)$ cruza a linha do valor esperado, indicando que os pontos estão distribuídos aleatoriamente além desta distância. Isso sugere que não há padrões espaciais fortes na distribuição de pontos em distâncias superiores a 300 metros.

No entanto, é importante observar que o valor de $K(r)$ flutua em torno da linha de valor esperado além de 300 metros, o que pode indicar a presença de alguns padrões espaciais fracos ou fatores subjacentes que influenciam a distribuição de pontos.

No geral, esta análise sugere que o padrão de pontos nesta área particular está agrupado em distâncias inferiores a 100 metros e mais disperso além de 100 metros, com alguns padrões espaciais fracos além de 300 metros.

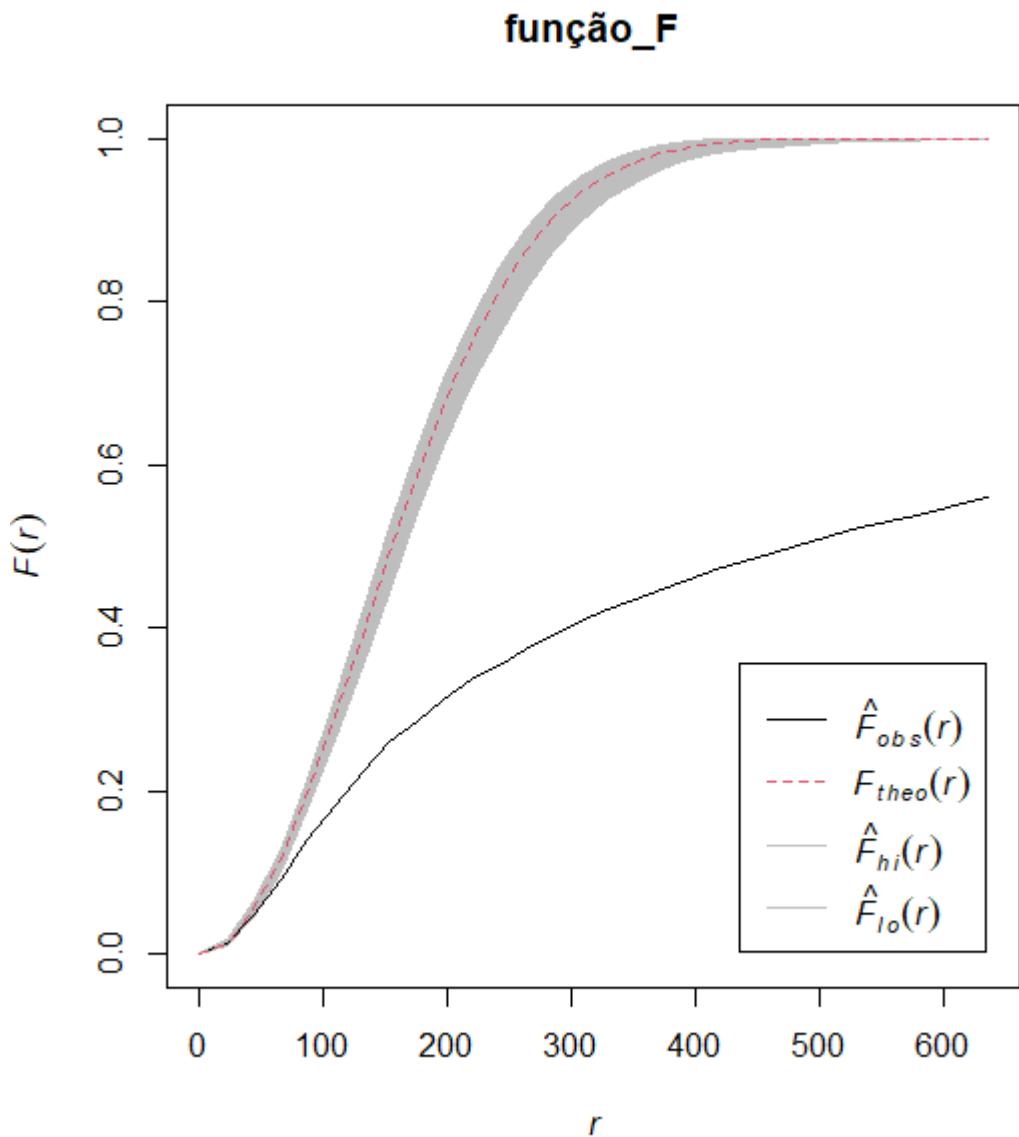


Figura 10: Função F - no eixo das ordenadas representa o $F(r)$, taxa referente a função F, no eixo das abscissas representa a distância em metros

No gráfico apresentado figura 10 , podemos ver um gráfico da Função F para um conjunto de dados pontuais. O eixo x representa a distância entre pares de pontos em metros, enquanto o eixo y representa o valor da Função F, que mede o grau de agrupamento espacial ou dispersão dos pontos.

O gráfico mostra que a Função F começa com o valor 0 quando a distância entre os pares de pontos é 0, indicando que não há agrupamento ou dispersão nessa distância. À

medida que a distância aumenta, o valor da Função F também aumenta, indicando que os pontos ficam mais dispersos ou agrupados.

A uma distância de cerca de 60 metros, a Função F atinge seu valor máximo em torno de 0,2, indicando um forte agrupamento espacial de pontos nessa distância. Isso sugere que existe um grupo de pontos fortemente agrupados, possivelmente indicando a presença de um padrão espacial ou fenômeno de agrupamento.

Após o pico a 60 metros, o valor da Função F diminui gradativamente à medida que aumenta a distância entre os pares de pontos. Isso indica que o agrupamento espacial ou dispersão dos pontos torna-se menos pronunciado à medida que a distância aumenta. No entanto, o valor da Função F permanece positivo, indicando que ainda existe algum grau de agrupamento ou dispersão espacial nos dados, mesmo a grandes distâncias.

No geral, a análise do gráfico usando a Função F - Espaços Vazios sugere a presença de um forte padrão de agrupamento espacial a uma distância de cerca de 60 metros. No entanto, ainda há algum grau de agrupamento ou dispersão nos dados, mesmo em grandes distâncias

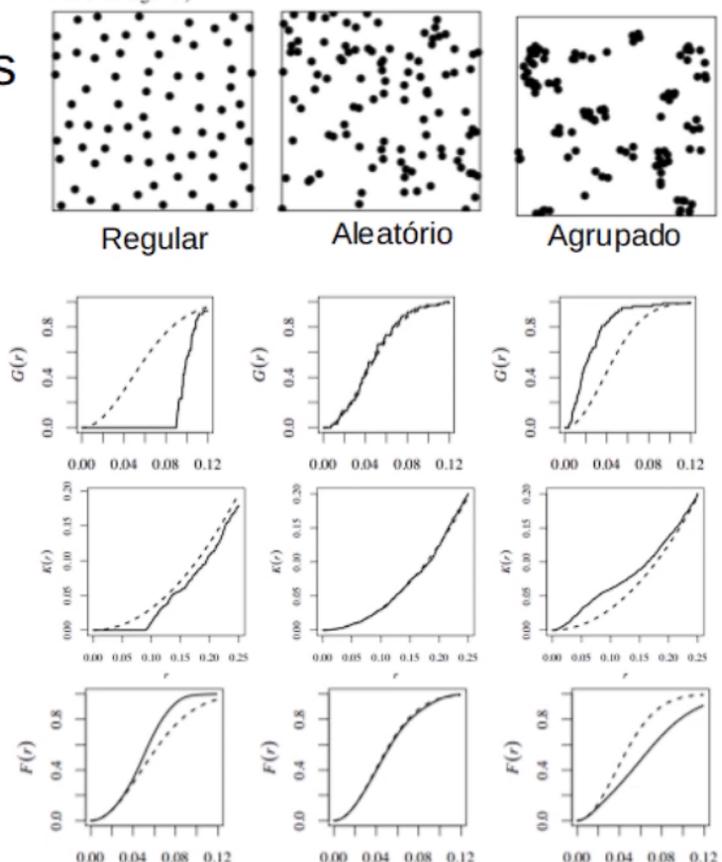
Na imagem abaixo demonstra a comparação entre as três funções

Comparando as três funções

G

K

F



— Baddeley, Adrian, Ege Rubak, and Rolf Turner. Spatial point patterns: methodology and applications with R. CRC Press, 2015.

Figura 11

Como observado as internações hospitalares têm uma tendência de agrupamento, utilizei um mapa de densidade de kernel para identificar espacialmente onde se encontra esse agrupamento

Densidade de Kernel

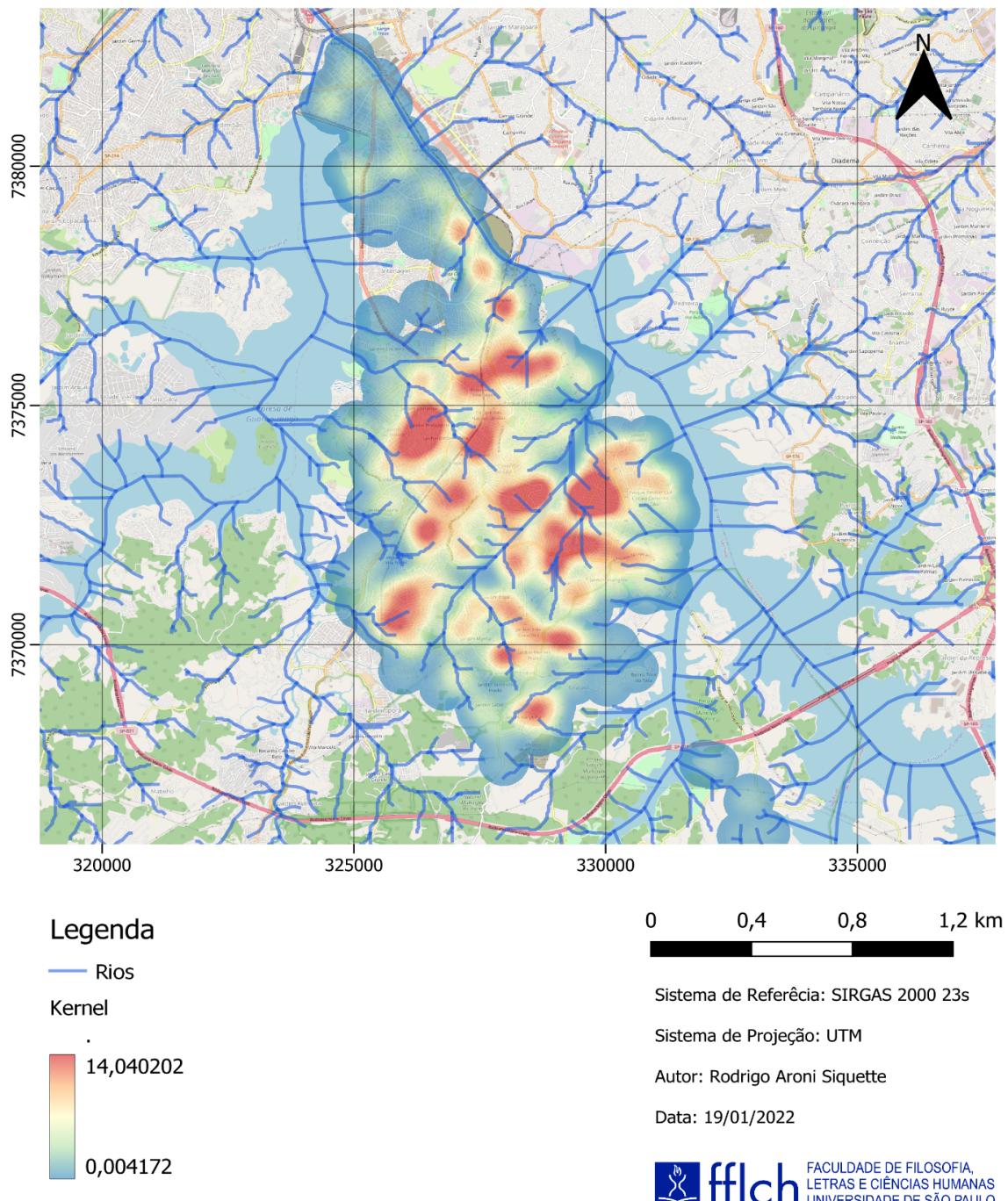


Figura 12: Densidade de Kernel para internações por diarreia presumida e 2006 a 2010

Após a configuração da existência de lugares de aglomeração, houve o interesse de descobrir se esses pontos tinham uma autocorrelação e buscar entender o motivo desse comportamento. Utilizando o índice de Moran mostra que a uma autocorrelação de 0,1343 outro dado importante é o p-value. O valor p em um Índice de Moran é uma medida da significância da autocorrelação espacial observada. Em estatística espacial, a autocorrelação espacial refere-se à presença de um padrão na distribuição espacial dos valores de uma variável, de modo que valores próximos uns dos outros tendem a ser mais semelhantes do que valores distantes. O valor p em um Índice de Moran é uma medida da significância da autocorrelação espacial observada. Na análise foi obtido o valor de “1.544e-05” ou seja “0.00001546”, mostrando que há uma forte significância de autocorrelação

```
Moran I test under randomisation

data: hexa_g_sp$NUMPOINTS
weights: hexa_pesos

Moran I statistic standard deviate = 4.1666, p-value = 1.546e-05
alternative hypothesis: greater
sample estimates:
Moran I statistic      Expectation      Variance
0.134342551     -0.001937984   0.001069820
```

Figura 13: Índice Moran

O correograma de distância de índice Moran é uma representação gráfica da autocorrelação espacial de uma variável em diferentes distâncias. Ele é criado plotando o Índice de Moran em diferentes distâncias (ou distâncias de defasagem) no eixo x e o valor I de Moran correspondente no eixo “y”

Moran I statistic = f(distance classes)

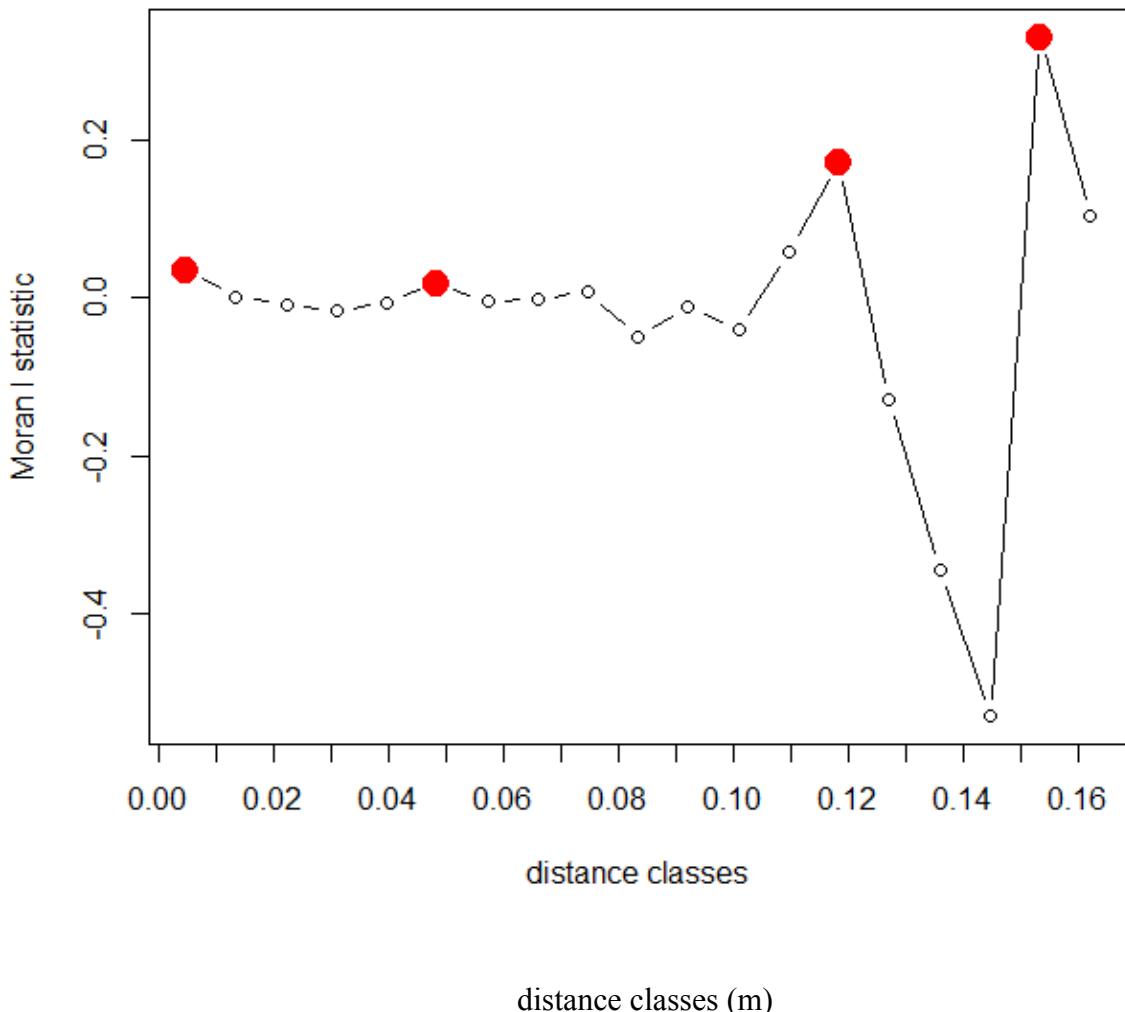


Figura 14: Correlograma de distância de índice Moran para os casos de CID A09

O Índice de Moran Local (LISA) decompõe a medida global de autocorrelação espacial em medidas locais, fornecendo uma imagem mais detalhada do padrão espacial da variável. É calculado como a razão da soma dos produtos dos desvios da variável de sua média e da matriz de pesos espaciais para cada local para a variância da variável. É um mapa que pode ser utilizado para identificar áreas de alta ou baixa autocorrelação espacial, onde os valores da variável tendem a ser semelhantes entre localidades próximas.

Olhando para o mapa figura 15, podemos ver que existem vários grupos de valores altos e baixos, representados pelos diferentes símbolos do índice LISA. No centro do mapa,

há um grande aglomerado de regiões de HH, indicando que essas regiões possuem valores altos e estão cercadas por outras regiões com valores também altos. Isso sugere uma forte autocorrelação espacial positiva, onde regiões com valores semelhantes tendem a se agrupar.

Por outro lado, também existem vários pequenos aglomerados de regiões LL e HL espalhados pelo mapa, indicando áreas de valores baixos e valores altos circundadas por regiões de valores baixos, respectivamente. Esses agrupamentos sugerem áreas de autocorrelação espacial negativa, onde regiões vizinhas possuem valores diferentes.

Além de identificar esses clusters, o índice LISA também fornece informações sobre a significância estatística dos clusters. Regiões com um índice LISA significativo indicam que é improvável que o agrupamento de valores altos ou baixos tenha ocorrido apenas por acaso. Essas regiões são marcadas com cores mais escuras no mapa, indicando um maior nível de significância

Lisa

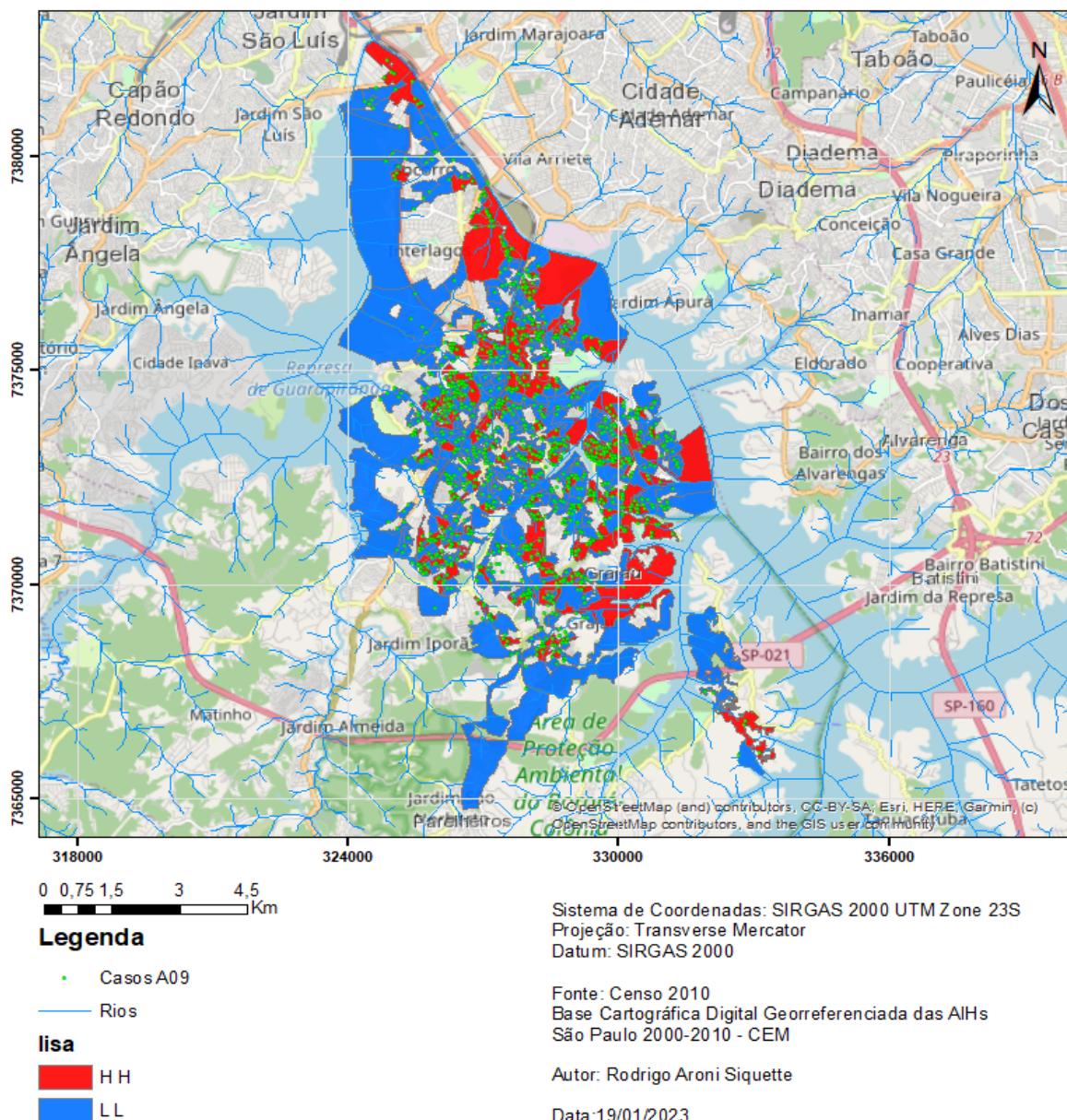
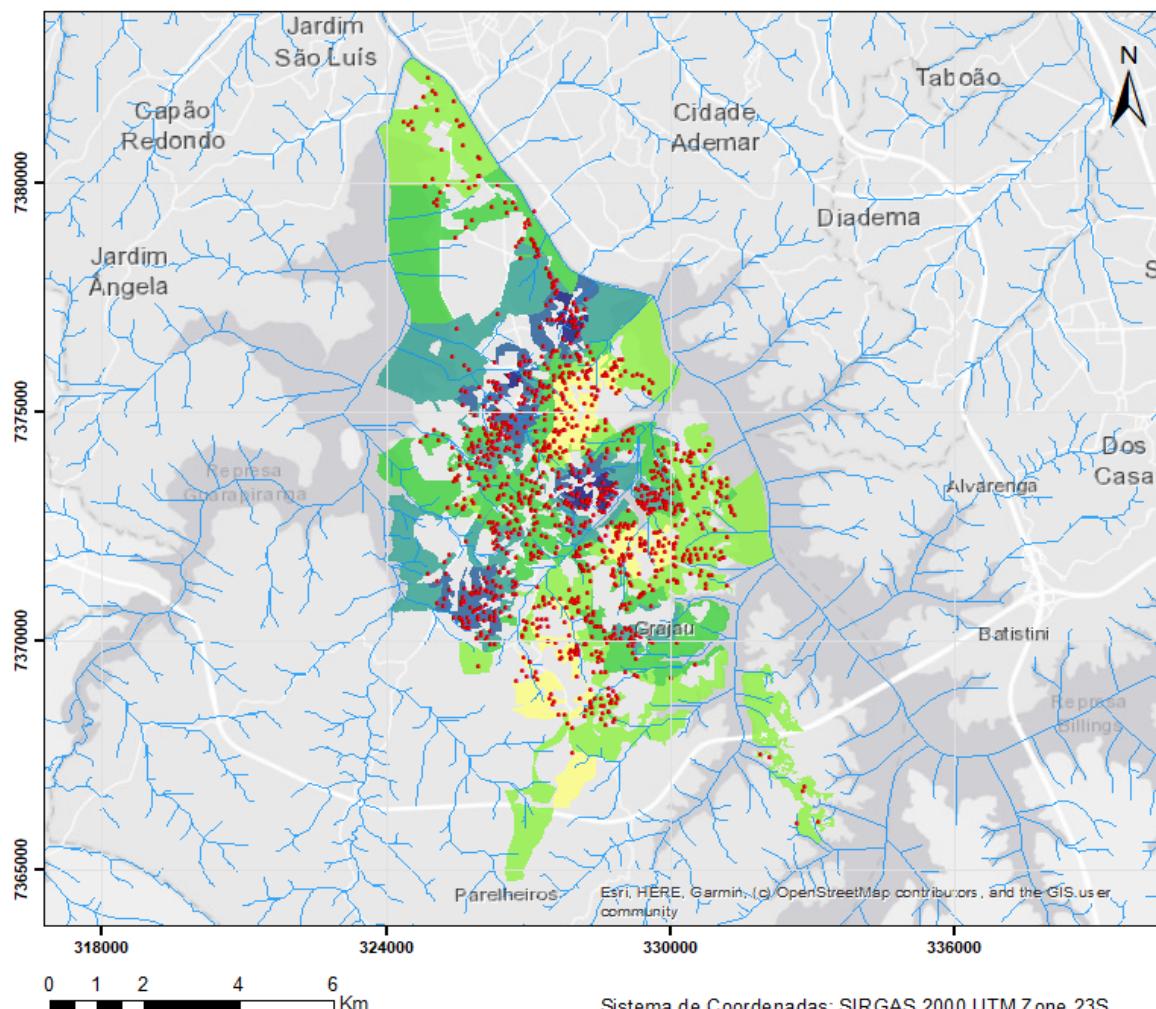


Figura 15: Mapa Lisa

Ao contrário da regressão tradicional, onde a relação entre as variáveis dependentes e independentes é estimada com base em todo o conjunto de dados, a regressão local estima a relação com base em um subconjunto dos dados, ou uma vizinhança local, em torno de cada ponto de dados. A ideia básica por trás da regressão local é ajustar um modelo de regressão aos pontos de dados mais próximos do ponto de destino, em vez de ajustar um único modelo a todos os pontos de dados

Para tentar entender o comportamento das internações foi feito uma regressão espacial local, utilizando os dados de internações hospitalares entre crianças de 0 a 5 anos por setor censitário e como variável independente, a população de até 5 anos em cada setor censitário. O resultado é mostrado no mapa a seguir.

Regressão Geograficamente Ponderada



Legenda

- Casos A09

- Rios

GWR

- | |
|-----------------------|
| -0,009963 - -0,000915 |
| 0,000916 - 0,006588 |
| 0,006589 - 0,012175 |
| 0,012176 - 0,019559 |
| 0,019560 - 0,028640 |
| 0,028641 - 0,043772 |

 **fflch** FACULDADE DE FILOSOFIA,
LETTRAS E CIÉNCIAS HUMANAS
UNIVERSIDADE DE SÃO PAULO

Figura 16 : Mapa Regressão Geograficamente Ponderada (GWR)

O R-quadrado, também conhecido como coeficiente de determinação, é uma medida estatística que representa a proporção da variância na variável dependente que é explicada pelas variáveis independentes em um modelo de regressão. No contexto da Regressão Geograficamente Ponderada (GWR), o R-quadrado é usado para avaliar a qualidade do ajuste dos modelos locais ajustados em diferentes locais dentro da área de estudo.

No GWR, um modelo de regressão de mínimos quadrados ordinários (OLS) separado é ajustado para cada local na área de estudo, usando as observações próximas a esse local. O modelo estima os coeficientes para cada localidade separadamente, permitindo detectar a não estacionariedade, ou seja, a variação dos coeficientes ao longo da área de estudo.

O valor de R quadrado para cada um desses modelos locais é calculado e pode ser usado para avaliar a qualidade do ajuste do modelo em cada local. Um alto valor de R-quadrado indica que as variáveis independentes no modelo são bons preditores da variável dependente naquele local, enquanto um baixo valor de R-quadrado sugere que as variáveis independentes não explicam muito da variação na variável dependente naquele local. Isso permite detectar se o modelo se ajusta bem em algumas áreas e não em outras.

Os resultados do R² mostram que em alguns locais a população de crianças de 0 a 5 anos chega a explicar até 53% da distribuição das internações (em tons de azul mais escuro). Em outros locais, no entanto, o ajuste é bem mais fraco. Isso indica que outros fatores são espacialmente explicativos da distribuição das internações, além da distribuição da população

de 0 a 5 anos.

Coeficiente de Determinação R²

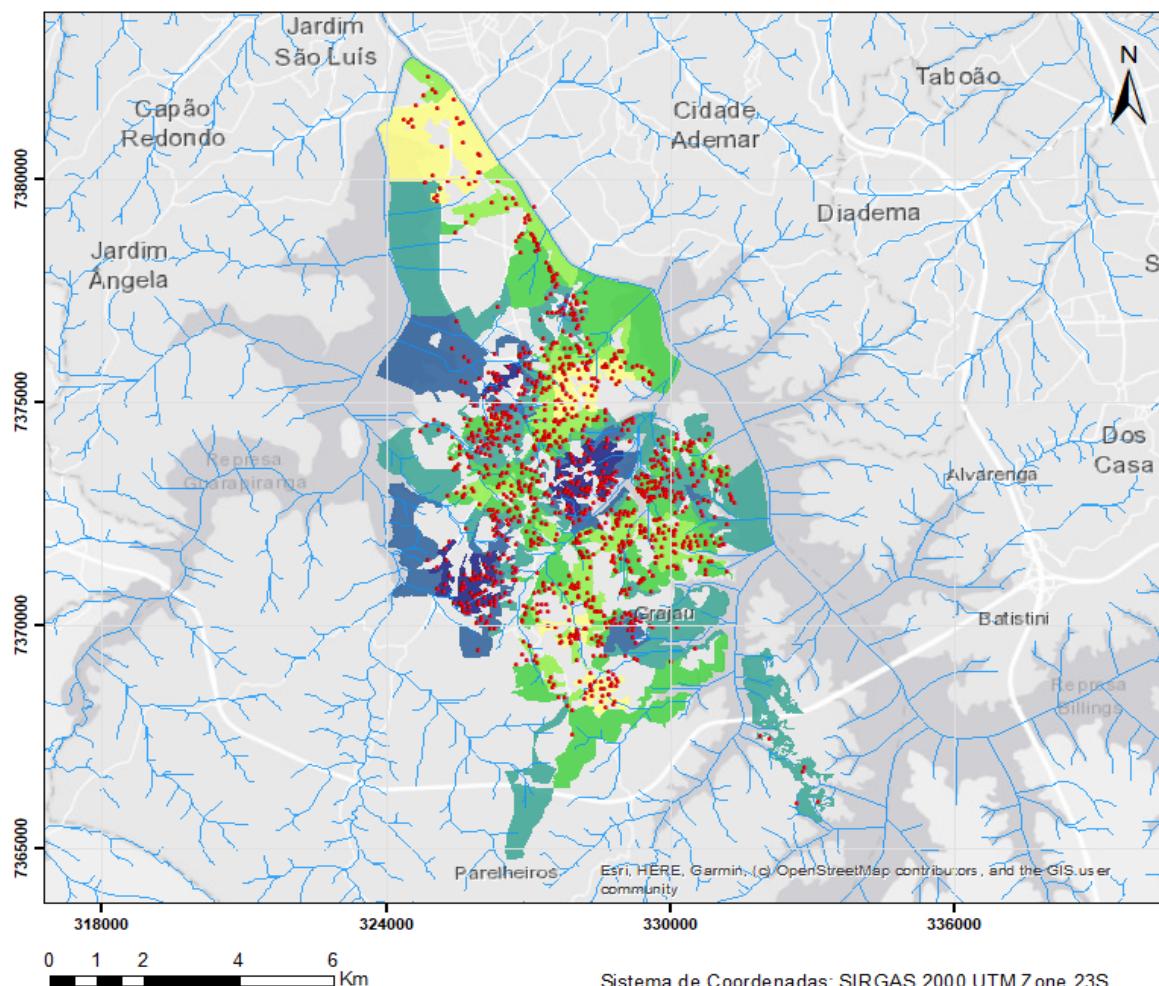


Figura 17: Mapa valor de R quadrado

Conclusão

A pesquisa buscou analisar os casos de Doenças Relacionadas ao Saneamento Ambiental Inadequado, usando como princípios a degradação ambiental nas áreas verde ao redor da área urbana e o contato com água contaminada, segundo a literatura escolhida, os ponto de maior risco de contágio é na fronteira entre o urbano e a vegetação e em torno dos cursos de água, com as análises finalizada foi possível observar hotspot de casos, contudo a distribuição dos casos não se mostra em áreas que prevista pela literatura, foi observado agrupamento na área central da subprefeitura, como mostrado na figura 19, para entender o motivo desse comportamento foi buscado relação com outro fatores, como o dados de acesso à água potável e esgoto coleta e tratamento de esgoto por setor censitário - figura 20 - portanto não foi possível afirmar que há uma relação entre casos de contaminação com falta de saneamento, com a análise de GWR foi possível observar que há uma relação com a quantidade de casos com o número de habitantes.

A Região Metropolitana de São Paulo tem 99,3% da população é atendida com abastecimento de água, frente a média de 96,51% do estado e 83,96% do país; e 96,3% da população é atendida com esgotamento sanitário, frente a média de 90,61% do estado e 66,04% do país (

<https://www.aguaesaneamento.org.br/municipios-e-saneamento/sp/sao-paulo#:~:text=S%C3%A3o%20PAULO%20possui%2067%2C28.tratamento%20ou%20de%20solu%C3%A7%C3%A3o%20individuais>). Mostrando que mesmo com os problemas encontrados na Região Metropolitana de São Paulo o saneamento básico se encontra em um nível superior em comparação com o restante do país. Utilizando matriz de Correlação Pearson - Figura 18 - é possível analisar os casos de CID A09 não tem uma correlação direta com déficits em saneamento básicos, a matriz mostra com “x” onde o *P-valor* é inferior a “0.001”

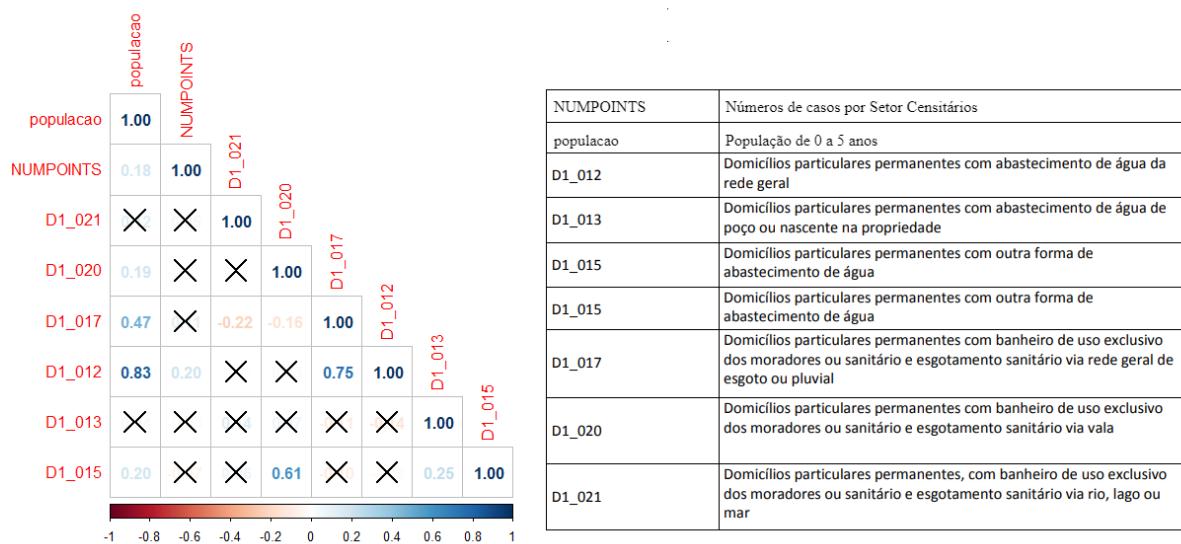
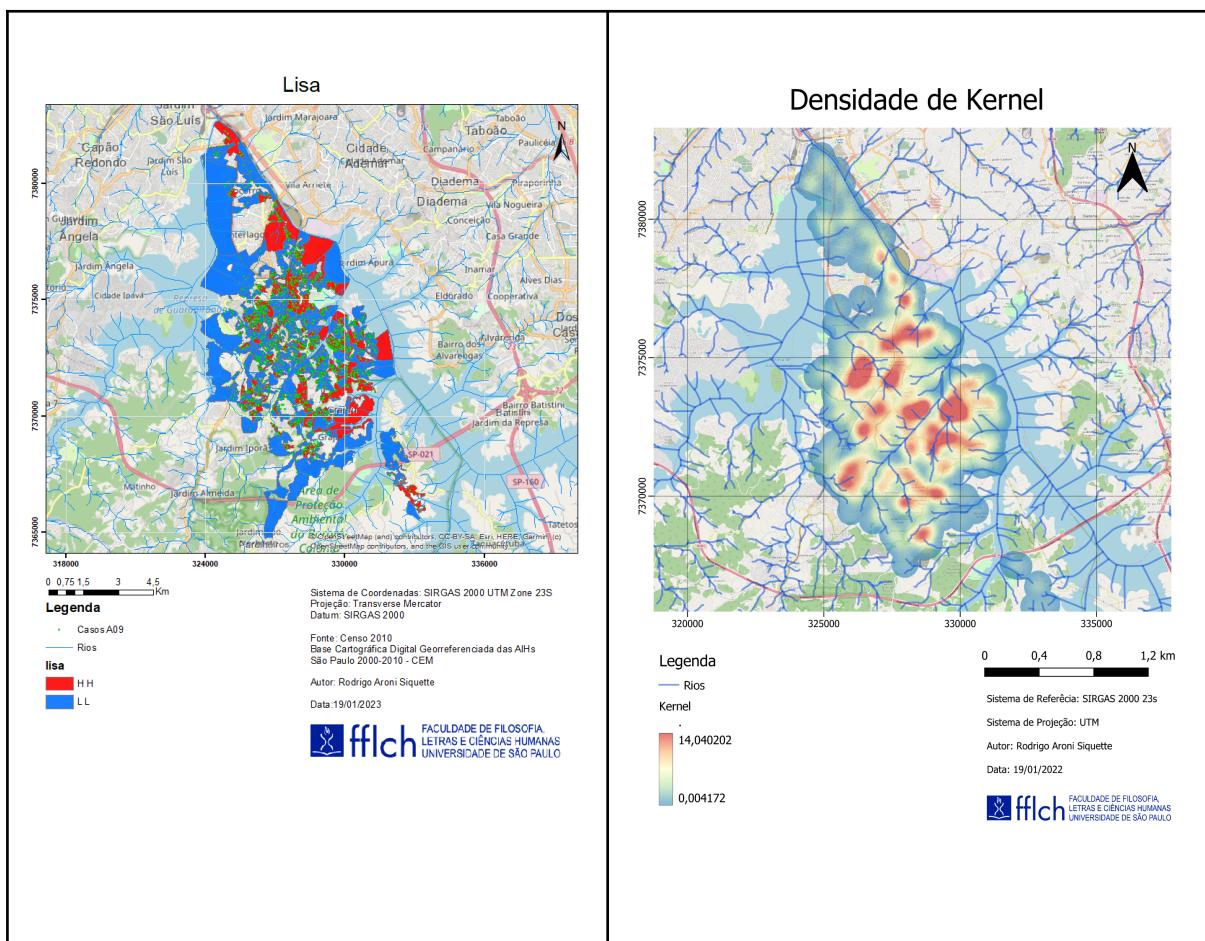


Figura 18: Matriz de Correlação Pearson



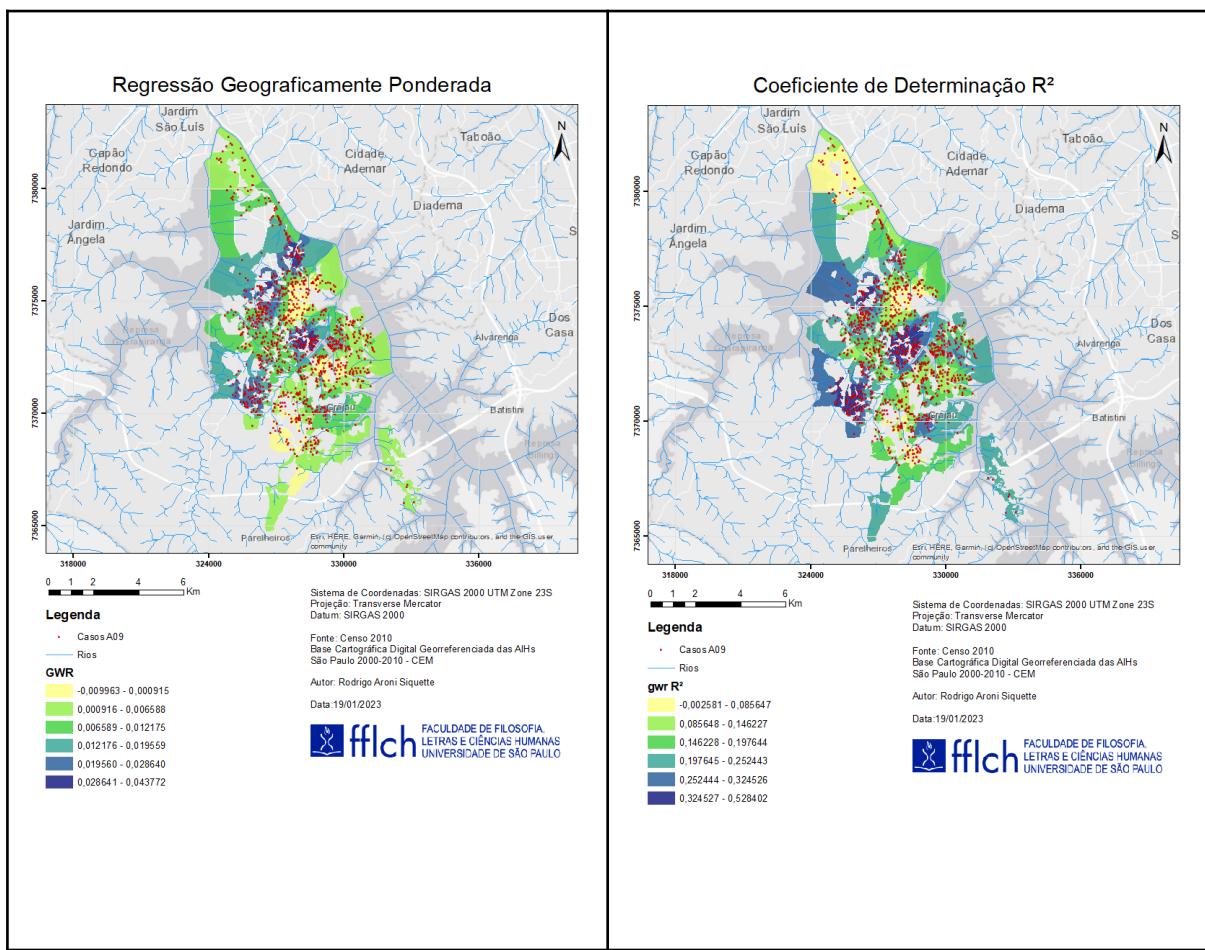


Figura 19: Mosaico de comparação entre os mapas

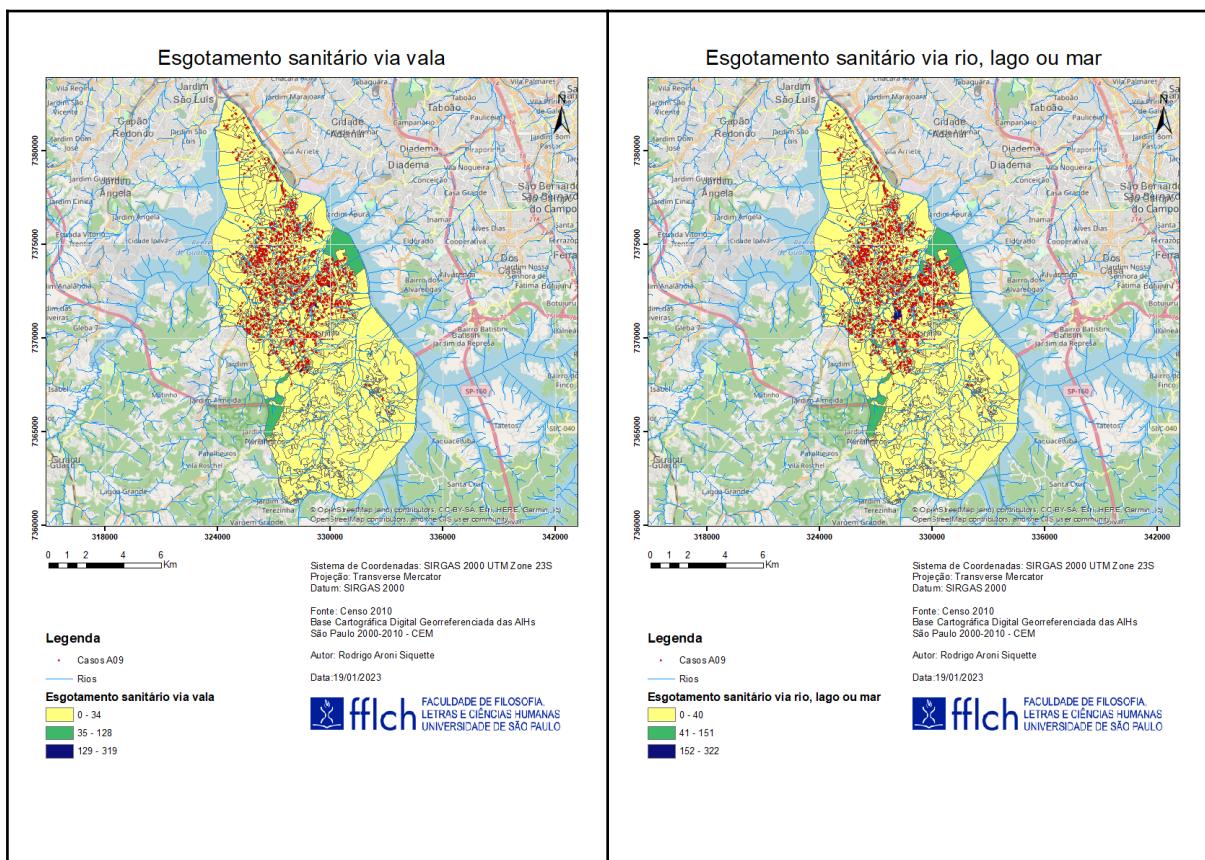


Figura 20; Mapas esgotamento sanitário

Referências bibliográficas:

AUSTIN, K. F. **Degradation and disease: Ecologically unequal exchanges cultivate emerging pandemics.** World Development Volume 137, Janeiro, 2021. DOI: <https://doi.org/10.1016/j.worlddev.2020.105163>. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0305750X20302904#bib176>>. Acesso em: 10/11/2022.

CENTERS FOR DISEASE CONTROL - CDC. Zoonotic Diseases. Disponível em: <<https://www.cdc.gov/onehealth/basics/zoonotic-diseases.html>>. Acesso em: 10/11/2022.

UNITED NATIONS ENVIRONMENT PROGRAMME (UNEP). Frontiers 2016: Emerging issues of environmental concern. In: UNEP Frontiers 2016 Report: Emerging Issues of Environmental Concern. United Nations Environment Programme, Nairobi. Disponível em: <<https://www.unep.org/resources/frontiers-2016-emerging-issues-environmental-concern>>. Acesso em: 10/11/2022.

BLOOMFIELD, L.S.P., MCINTOSH, T.L.; LAMBIN, E.F. Habitat fragmentation, livelihood behaviors, and contact between people and nonhuman primates in Africa. *Landscape Ecol* 35, 985–1000 (2020). DOI: <https://doi.org/10.1007/s10980-020-00995-w>. Disponível em: <https://link.springer.com/article/10.1007/s10980-020-00995-w?utm_source=getftr&utm_medium=getftr&utm_campaign=getftr_pilot>. Acesso em: 10/11/2022.

TORRES, H.; MARQUES, E.; FERREIRA, M.P.; BITAR, S. Pobreza e espaço: padrões de segregação em São Paulo. **Revista do Instituto de Estudos Avançados, IEAUSP**, v.17, n. 47, p. 97-128, jan./abr., 2003.

TASCHNER, S. P. Degradação ambiental em favelas de São Paulo. In: **TORRES, H.; COSTA, H. (Orgs.).** População e meio ambiente: debates e desafios. São Paulo: Editora Senac, 2000, p.271-297.

TORRES, H. A demografia do risco ambiental. In: **TORRES, H.; COSTA, H. (Orgs.).** População e meio ambiente: debates e desafios. São Paulo: Editora Senac, 2000, p.53-73.

Variação geográfica do saneamento básico no Brasil em 2010: domicílios urbanos e rurais, **LANDAU, E. C.; MOURA, L. 2016**

Spatial Point Patterns book. Methodology and Applications with R. **By Adrian Baddeley, Ege Rubak, Rolf Turner. 2015**

BAILEY e GATREL, 1995; LLOYD, 2007; PROTÁZIO, 2007

Ripley, B. D. (1977). Modeling Spatial Patterns. Journal of the Royal Statistical Society. Series B (Methodological), 39, 172-212.

Burt et al. 2009, p. 554)

Notes on Continuous Stochastic Phenomena. Author(s): **P. A. P. Moran.** Source: **Biometrika , Jun., 1950,**

Fotheringham et al. 2002

Geographically Weighted Regression The Analysis of Spatially Varying Relationships Wiley 2002

Script R

```
1  """{r}
2
3
4  library("rgdal")
5  library("spatstat")
6  library("maptools")
7  library("raster")
8  library("sf")
9  library("readr")
10 library("dplyr")
11 library("geobr")
12 library("maptools")
13 library("tmap")
14 library("cleangeo")
15 library("spdep")
16 library("pgirmess")
17 library("spatialreg")
18 library("spgwr")
19 """
20 """{r}
21 # Load data ----
22
23 grajau_sf <- read_sf("merge_ponto.shp",) %>%
24   st_transform(crs = 32723)
25
26 grajau_sp <- as(grajau_sf, "Spatial")
27
28
29 ppp_g<- as(grajau_sp,"ppp")
30
31 plot(ppp_g)
32
33 ppp_unmark_17 <- unmark(ppp_g)
34
35 plot(ppp_unmark_17)
36
37
38 #polgno
39
40 sp_g <- read_sf("agoravai_grajau.shp",) %>%
41   st_transform(crs = 32723)
42
43 sp_gwin <- as.owin(sp_g)
44
45
46 #####
47 #####
48
49
50 g_ppp <- ppp(
51   st_coordinates(grajau_sf)[,1],
52   st_coordinates(grajau_sf)[,2],
53   window = sp_gwin)
54
55
56 L_g <- envelope(g_ppp, Lest, nsim = 100, verbose = T)
57
```

```

58 fun o_G <- envelope(g_ppp, Gest, nsim = 1000, verbose = T)
59
60 fun o_K <- envelope(g_ppp, Kest, nsim = 1000, verbose = T)
61
62 fun o_F <- envelope(g_ppp, Fest, nsim = 1000, verbose = T)
63
64
65 plot(L_g)
66 plot(fun o_K)
67 plot(fun o_G)
68 plot(fun o_F)
69
70
71
72
73
74 ***{r}
75 hexa_g <- st_read("merge.shp")
76
77 hexa_g <- hexa_g %>% filter(populacao >= 1)
78
79 nao_nulos <- subset(hexa_g, is.na(hexa_g$NUMPOINTS)==FALSE)
80
81 hexa_g_validos <- subset(nao_nulos, (nao_nulos$NUMPOINTS) !=0 )
82
83 hexa_g_sps <- as(hexa_g_validos,"Spatial") #setores clan
84
85 clgeo_IsValid(hexa_g_sps) #corre ao topologica
86
87
88 #hexa_g_validos <- subset(hexa_g_valido, (hexa_g_valido$NUMPOINTS)
     !=1)
89
90
91 #hexa_g_sp <- as(hexa_g_validos,"Spatial")    #setores_juntos
92
93 #plot(hexa_g_sps)
94
95 hexa_g_vizinhancaa <- poly2nb(hexa_g_sps)
96
97 #plot(hexa_g_vizinhanca, coord= hexa_g_sps)
98
99 View(card(hexa_g_vizinhancaa))
100
101 hexa_g_sps$vizinhos <- card(hexa_g_vizinhancaa)
102
103
104
105 hexa_g_sp <- hexa_g_sps
106 hexa_g_sp <- subset(hexa_g_sps, hexa_g_sps$vizinhos !=0 ) #setores
_juntos
107
108
109
110 hexa_g_vizinhanca <- poly2nb(hexa_g_sp) #vizinhan a2
111
112 setores_xy <- coordinates(hexa_g_sp)

```

```

113
114 #plot(hexa_g_sps)
115 #plot(x=hexa_g_vizinhanca, coord = setores_xy, cex=0.6, add=TRUE)
116
117 vizinhanca_sp <- nb2lines(hexa_g_vizinhanca, coords = setores_xy,
118                             proj4string= crs(hexa_g_sps))
119
120 plot(vizinhanca_sp)
121
122 hexa_pesos <- nb2listw(hexa_g_vizinhanca) #vizinhan a normalizada
123
124
125
126 moran.test(x = hexa_g_sp$NUMPOINTS, listw = hexa_pesos ) #test
127     moran
128
129 #correlograma_contiguidade <- sp.correlogram(neighbours = hexa_g_
130     vizinhanca, var = hexa_g_sp$NUMPOINTS, order = 5, method = "I")
131
132 #plot(correlograma_contiguidade)
133
134 correlograma <- correlog(setores_xy,hexa_g_sp$NUMPOINTS )
135
136 plot(correlograma) #correlograma de distancia
137
138 #####      LISA #####
139
140 localmoran_g <- localmoran(x = hexa_g_sp$NUMPOINTS, listw = hexa_
141                               pesos)
142
143 View(localmoran_df_g)
144
145 localmoran_df_g <- as.data.frame(localmoran_g)
146
147 hexa_g_sp$moran_local <- localmoran_df_g$II
148
149 hexa_g_sp$moran_p <- localmoran_df_g$'Pr(z != E(II))'
150
151 View(hexa_g_sp@data)
152
153
154 tm_shape(hexa_g_sp) + tm_fill("moran_local", style = "fixed",
155     breaks=c(-3,0,0.2,0.5,30),palette=c("red","lightblue", "blue",
156     "blue4" ))
157
158 tm_shape(hexa_g_sp) + tm_fill(col="moran_p", style = "fixed",
159     breaks=c(-0,0.01,0.05,1),palette=c("darkblue", "blue", "gray" )
160     ) # p moran onde o moran mais confiavel

```

```

161 tm_shape(hexa_g_sp) + tm_fill(col="spatial_lag", style = "quantile")
162
163
164 moran.plot(x=hexa_g_sp$NUMPOINTS, listw = hexa_pesos, cex=0.6,
165             labels = FALSE)
166 L1 <- factor(hexa_g_sp$NUMPOINTS < mean(hexa_g_sp$NUMPOINTS),
167               labels = c("H", "L"))
168 L2 <- factor(hexa_g_sp$NUMPOINTS < mean(hexa_g_sp$spatial_lag),
169               labels = c("H", "L"))
170 hexa_g_sp$lisa <- paste(L1, L2)
171
172 tm_shape(hexa_g_sp)+tm_fill("lisa", palette = c("blue", "red"))
173
174 #hexa_g_sp$popula ao = 684757
175
176 bayes_global <- EBest(n=hexa_g_sp$NUMPOINTS, x=hexa_g_sp$populacao)
177
178 hexa_g_sp$pes_bayes <- bayes_global$raw
179
180 hexa_g_sp$bayes_gl <- bayes_global$estmm
181
182 tm_shape(hexa_g_sp)+ tm_fill("bayes_gl", style = "fisher")
183
184 bayes_local <- EBlocal(ri=hexa_g_sp$NUMPOINTS, ni=hexa_g_sp$populacao,
185                         nb=hexa_g_vizinhanca)
186
187 hexa_g_sp$bayes_ic <- bayes_local$est
188
189 tm_shape(hexa_g_sp)+ tm_fill("bayes_ic", style="fisher")
190
191 saveRDS(hexa_g_sp, file = "hexa_g_sp.Rds")
192
193 View(hexa_g_sp@data)
194
195 # GWR
196
197 raio <- gwr.sel(data = hexa_g_sp, NUMPOINTS ~ populacao )
198
199 setores_gwr <- gwr(data = hexa_g_sp, NUMPOINTS ~ populacao,
200                      bandwidth = raio)
201
202 raio_adaptativo <- gwr.sel(data = hexa_g_sp, NUMPOINTS ~ populacao,
203                               adapt = TRUE )
204
205 setores_gwr_adaptativo <- gwr(data = hexa_g_sp, NUMPOINTS ~
206                                   populacao, adapt = raio_adaptativo)
207
208 View(setores_gwr_adaptativo$SDF@data)
209
210 sqrt(mean(setores_gwr$SDF$gwr.e^2))
211 sqrt(mean(setores_gwr_adaptativo$SDF$gwr.e^2))

```

```

210 tm_shape(setores_gwr_adaptativo$SDF) + tm_fill("populacao", style =
211   "fisher")
212 tm_shape(setores_gwr_adaptativo$SDF) + tm_fill("localR2", style =
213   "fisher")
214 setores_gwr_adaptativo
215
216 hexa_g_sp$gwrlocalR2 <- setores_gwr_adaptativo$SDF$localR2
217
218 hexa_g_sp$gwrpopulacao <- setores_gwr_adaptativo$SDF$populacao
219
220 write.csv (hexa_g_sp, file = "C:/Users/rodri/OneDrive/Desktop/tgi/
221   r3/moran.csv")
222
223 View(hexa_g_sp@data)
224
225 ***
226
227 ***{r}
228
229 #Figura 18: Matriz de Correlação Pearson
230 taxa <- moran %>% select(populacao, NUMPOINTS, D1_021, D1_020, D1_
231   017, D1_012, D1_013, D1_015 )
232 mm <-cor(taxa)
233 .
234
235 m <-rcorr(as.matrix(taxa))
236
237 m
238
239 corrplot(mm, type = "lower", method = "number")
240
241 corrplot(m$r , p.mat=m$P, sig.level = 0.001, method = "number",
242   type = "lower")
243 ***

```