



**UNIVERSIDADE FEDERAL DO PARÁ
CAMPUS UNIVERSITÁRIO DE CASTANHAL
FACULDADE DE COMPUTAÇÃO**

PEDRO IGOR OLIVEIRA CARVALHO

**DESENVOLVIMENTO DE UM ALGORITMO PARA GEOCODIFICAR CASOS
DE HANSENÍASE POR SETORES CENSITÁRIOS**

**CASTANHAL-PARÁ
2018**

PEDRO IGOR OLIVEIRA CARVALHO

**DESENVOLVIMENTO DE UM ALGORITMO PARA GEOCODIFICAR CASOS DE
HANSENÍASE POR SETORES CENSITÁRIOS**

Trabalho de conclusão de curso apresentado à
Universidade Federal do Pará como requisito parcial
para obtenção do grau de Bacharel em Engenharia
de Computação.

Orientador: Prof. Dr. Josafá Gonçalves Barreto

**Castanhal, Pará
2018**

PEDRO IGOR OLIVEIRA CARVALHO

**DESENVOLVIMENTO DE UM ALGORITMO PARA GEOCODIFICAR CASOS
DE HANSENÍASE POR SETORES CENSITÁRIOS**

Trabalho de conclusão de curso apresentado à
Universidade Federal do Pará como requisito parcial
para obtenção do grau de Bacharel em Engenharia
de Computação.

Data de aprovação: ___/___/_____

Banca Examinadora:

Prof. Josafá Gonçalves Barreto
(Orientador – Universidade Federal do Pará)

Prof José Jailton Henrique Ferreira Júnior.
Universidade Federal do Pará

Prof(a) Bruno Souza Lyra Castro.
Universidade Federal do Pará

À minha mãe. Mérito todo seu.

AGRADECIMENTOS

À Universidade Federal do Pará, Faculdades de Computação do Campus Universitário de Castanhal. Por todo o conhecimento adquirido e pela oportunidade de evolução.

Ao professor Marcos Seruffo, pela primeira oportunidade de participar em um projeto de pesquisa, o que me incentivou a continuar nesta empreitada.

Ao professor João Claudio Chamma Carvalho, pelos conselhos e incentivos para progredir.

Aos demais professores da Faculdade de Computação, por todo o conhecimento compartilhado.

Aos amigos Joahannes Bruno, Daniel Souza, Neto Landy, Edylle Landy, Vinicius Pedrosa. Vocês estão sempre à frente, mas estou logo atrás!

Ao professor e amigo Josafá Gonçalves Barreto, que tive o privilégio de ser seu orientando em projeto de pesquisa e depois neste Trabalho de Conclusão de Curso, e aos amigos do Laboratório de Dermato-Imunologia (LDI), de Marituba. Aprendi que podemos fazer a diferença na vida das pessoas com ações simples, porém consistentes.

À Nevoli Telecom, em especial aos meus chefes e amigos Orivaldo e Dailane, por acreditar em mim e me ajudar a crescer.

Aos amigos Gustavo Gonçalves, Nathalia Souza, Matheus Teixeira, Emanuel Pantoja e Auriane Castro, por todos os momentos.

À minha família, em especial minha mãe Glória de Fátima e minha mãe/tia/avó Maria do Carmo. As mulheres que me inspiram.

Ao meu primo/tio/irmão Rogério Costa, por todos os conselhos e aprendizados.

À minha querida prima Sophia Laura, hoje com 3 anos. Daqui a pouco é você quem vai escrever linhas como essas.

Ao meu amigo Lucas Alves Costa, por nossa amizade.

À minha namorada Evelyn Moraes, por todos os momentos!

Por último, mas não menos importante, agradeço a todos que participaram de alguma forma nesta jornada, sejam conhecidos ou desconhecidos.

“Impasse

*Se ficar onde estou não faço nada,
Se sair por aí corro perigo,
Se me calo Minh'alma é sufocada,
Se disser o que sei faço inimigo...*

*Se pensar vou trair a madrugada
E se sonho demais vem o castigo,
Se quiser subo até o fim de escada,
Mas precisa brigar, e eu não brigo!*

*Se cantar atropelo o contracanto,
Se não canto maltrato o coração,
Se me faço sofrer me desencanto,*

*Se reprimo o ideal perco a razão,
Se perder a razão, resta-me o pranto
E meu pranto não faz uma canção.”*

(ORLANDO TEJO,

Revista Poesia Sempre, Número 29, Ano 15, 2008)

RESUMO

A hanseníase é uma doença infectocontagiosa crônica, capaz de provocar importantes incapacidades físicas quando não diagnosticada e tratada precocemente. Dentre as ferramentas disponíveis para ajudar este diagnóstico precoce, o uso de Sistemas de Informações Geográficas (softwares que possibilitam trabalhar com camadas de dados georreferenciados agregadas com bases de dados) é estimulado pela Organização Mundial da Saúde para auxiliar o estudo da epidemiologia espacial, que busca entender como um agravo se desenvolve em relação aos fatores socioambientais. A análise espacial demanda recursos e tempo para ser concluída, adiando a análise e o planejamento de ações de combate e controle deste agravo, pois as bases de dados disponíveis não possuem a informação de longitude e latitude dos casos notificados, requisitando a coleta individual destas coordenadas geográficas para então se utilizar o SIG para análises. O objetivo deste trabalho foi desenvolver um algoritmo para mapeamento de casos por setores censitários, que são uma subdivisão do território nacional utilizada pelo Instituto Brasileiro de Geografia e Estatística para realização das atividades de recenseamento populacional. Apesar de não apresentar a precisão do mapeamento no local (utilizando um aparelho receptor de sinal GPS), ou do mapeamento remoto com imagens de satélite, esta proposta pode mapear os casos dentro do seu contexto socioeconômico, cultural, e reduzir o tempo entre elaboração de mapas a partir de dados tabulados e a análise estatística, auxiliando no planejamento das estratégias de saúde. Em um primeiro momento foi elaborado o algoritmo em linguagem python para ser executado no terminal integrado do SIG QGIS. Em seguida escolheu-se a cidade de Castanhal, localizada na região nordeste do estado do Pará, para comparar os métodos de mapeamento atuais com o algoritmo proposto. De um total de 1322 casos registrados na zona urbana castanhalense, do período de 2001 a 2016, foram mapeados 1042 casos, o algoritmo final deste trabalho levou por volta de 2 minutos para concluir o geoprocessamento e não apresentou erros ou mapeamentos incorretos.

Palavras-chave: Hanseníase. Epidemiologia Espacial. Georreferenciamento. Python. QGIS.

ABSTRACT

Leprosy is a chronic infectious disease, capable of causing important disabilities when undiagnosed and treated early. Among the tools available for the early diagnosis, the use of geographic information systems (softwares which aim to work with georeferenced data layers aggregated with databases) is stimulated by the World Health Organization to assist the study of spatial epidemiology, who wants to understand how a disease develop beside of socio-environmental factors. The spacial analysis need resources and time to be concluded, postponing actions analysis and planning to combat and control of the disease, since the availabes databases do not contain the longitude and latitude information of the notified cases, requesting an individual sampling of these coordinates to use GIS for analysis. The objective of this work was to develop an algorithm for the mapping of leprosy cases by census tracts, which is a subdivision of the national territory used by the Brazilian Institute of Geography and Statistics for the accomplishment of population census activities. Although it does not provide the precision of on-site mapping (using a GPS signal receiver), or by mapping with satellite images, this proposal can map cases within their socioeconomic, cultural, and reduce the time used to elaborate maps from tabulated data and statistical analysis, assisting in the planning of health strategies. First, the python language algorithm was written to run on the integrated QGIS SIG terminal. Next, the city of Castanhal, located in the northeast region of Pará, was chosen to compare the mapping methods with the proposed algorithm. From a total of 1322 cases registered in the urban city, from 2001 to 2016, 1042 cases were mapped, the final algorithm of this work took about 2 minutes to complete the geoprocessing and did not present errors or incorrect mapping.

Keywords: Leprosy. Spatial Epidemiology. GIS. Python. QGIS.

LISTA DE ILUSTRAÇÕES

Figura 1 – Reprodução do mapa elaborado por John Snow em 1854 – As barras representam os casos de morte.	11
Figura 2 - Caderno e mapa utilizados no mapeamento participativo.....	16
Figura 3 – A: Malha digital da Zona Urbana de Castanhal sobreposta à imagens de satélite. B: Localização das UBS no município	17
Figura 4- Cobertura da ESF na Zona Urbana de Castanhal – PA	17
Figura 5 – Setores Censitários de Castanhal. Cada setor está representado com seu código, este sendo composto por 15 dígitos no formato UFMMMMMDDSDSSSS. Os dois primeiros números indicam o Estado, seguido por 5 dígitos identificando o município, os próximos 2 dígitos identificam o distrito e em seguida dois dígitos identificam o subdistrito. Os 4 últimos dígitos identificam o setor censitário	18
Figura 6 – Casos mapeados sobrepostos à camada de Setores Censitários.....	20
Figura 7 – Casos mapeados sobreposto à cobertura da ESF.....	20
Figura 8 - Diagrama apresentando a lógica implementada no algoritmo.....	23
Figura 9 – Erro ao utilizar a função centroid()	25
Figura 10 – Erro corrigido ao utilizar a função pointOnSurface() - em verde os pontos marcados corretamente	25
Figura 11 – Mapa final com os casos mapeados.....	27
Figura 12 - Mapa gerado com o mapeamento manual	28
Figura 13 – Mapa gerado com o algoritmo	29
Figura 14 – Tabela de Atributos	30

LISTA DE GRÁFICOS

Gráfico 1 - Representação da taxa de detecção em menores de 15 anos. Apesar da redução, o valor ainda é alto.	13
--	----

LISTA DE ABREVIATURAS E SIGLAS

ACS - Agentes Comunitários de Saúde

API - *Application Programming Interface*

ESF - Estratégia Saúde da Família

MA – Microáreas

OMS - Organização Mundial de Saúde

SC - Setores Censitários

SIG - Sistemas de Informações Geográficas

SUS - Sistema Único de Saúde

UBS - Unidades Básicas de Saúde

SINAN - Sistema de Informação de Agravos de Notificação

IBGE – Instituto Brasileiro de Geografia e Estatística

SUMÁRIO

1. INTRODUÇÃO	11
<i>1.1 OBJETIVO</i>	<i>14</i>
2 - METODOLOGIA	15
<i>2.1 LEVANTAMENTO DE INFORMAÇÕES</i>	<i>15</i>
3. ALGORITMO DESENVOLVIDO	22
4. RESULTADOS	27
<i>4.1 ALGORITMO X MAPEAMENTO REMOTO</i>	<i>27</i>
5 - DISCUSSÃO	31
6 - CONCLUSÕES	32
<i>6.1 TRABALHOS FUTUROS</i>	<i>33</i>
REFERÊNCIAS	34
APÊNDICE A - ALGORITMO PARA REPRODUÇÃO	36

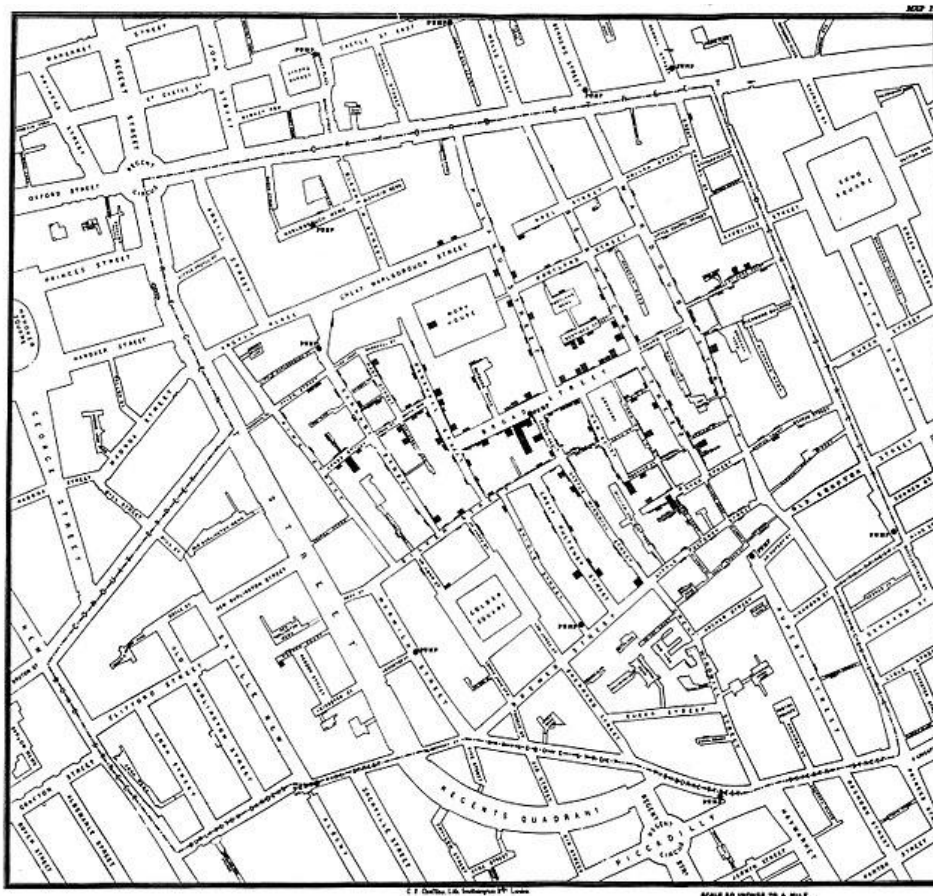
1. INTRODUÇÃO

Sistemas de Informações Geográficas (SIG) são conjuntos de ferramentas computacionais usadas para capturar, armazenar, gerenciar e apresentar informações georreferenciadas. A aplicação de SIG pode resultar em uma melhor visualização e tomada de decisão sobre questões relacionadas a diversas áreas do conhecimento, com menor custo de tempo e recursos (BARRETO et al., 2015).

Dentre as áreas que podem se beneficiar com o uso de SIG, a epidemiologia espacial descreve e analisa as variações geográficas do estado de saúde e doença das populações correlacionando a fatores de risco demográficos, ambientais, comportamentais, socioeconômicos, genéticos e infecciosos. Os avanços tecnológicos relacionados à análise espacial aplicados à área da saúde têm sido fortemente empregados para o melhor entendimento e manejo dos agravos à saúde da população (BARRETO et al., 2014), inclusive com apoio da Organização Mundial de Saúde (OMS) para análises geoespaciais do comportamento da endemia de hanseníase com objetivo de contribuir no programa de controle da doença em áreas endêmicas (WORLD OF HEALTH ORGANIZATION, 2016).

O uso de mapas e aplicação de conceitos primitivos de epidemiologia espacial foi fundamental para o médico John Snow (York, 15 de março de 1813 – Londres, 16 de junho de 1858) montar sua teoria sobre a transmissão da cólera e combater um surto em Londres, no ano de 1854. Em seu tempo ainda não se conhecia o agente causador da doença, e a "Teoria dos Miasmas" era a corrente aceita e difundida pela comunidade médica, onde se acreditava que o ar seria o responsável pela transmissão da enfermidade (HALLIDAY, 2001). Na época, a água para consumo humano era disponibilizada através de poços espalhados pela cidade. Ao mapear os casos de morte ao redor dos poços e visualizar em mapa, dr. Snow constatou que havia uma grande concentração de casos nas proximidades do poço localizado na Broad Street (SNOW, 1855) e elaborou um mapa para visualizar essa concentração (Figura 1). Snow, juntamente com as autoridades locais, interditaram o poço que julgaram ser o “causador” das mortes, e de fato o número de óbitos diminuiu drasticamente naquela área.

Figura 1 – Reprodução do mapa elaborado por John Snow em 1854 – As barras representam os casos de morte.



Fonte: Snow, John. On the Mode of Communication of Cholera, 2nd Ed, John Chruchill, London, England, 1855.

Dentre as opções de software SIG disponíveis atualmente, o QGIS (<http://www.QGIS.org>), é uma alternativa de código aberto (*software* livre), com uma comunidade ativa e que permite desenvolver algoritmos e complementos (*plugins*) para estender suas funcionalidades.

Diversos formatos podem ser lidos em um SIG. Os arquivos que compõem o formato *Shapefile* são utilizados para exibir dados espaciais vetoriais (pontos, linhas, polígonos) com seus respectivos dados não espaciais (banco de dados, tabela de atributos). Foi desenvolvido pela ESRI® com licença aberta possibilitando abrir estes arquivos em SIG da preferência do operador (ENVIRONMENTAL SYSTEMS RESEARCH INSTITUTE, 1998), sendo que três dos arquivos individuais (.shp, .shx, .dbf) são obrigatórios para armazenar os dados do núcleo que compreende um *shapefile* (“Shape – Definições e Conversão | SEMACE”, 2011).

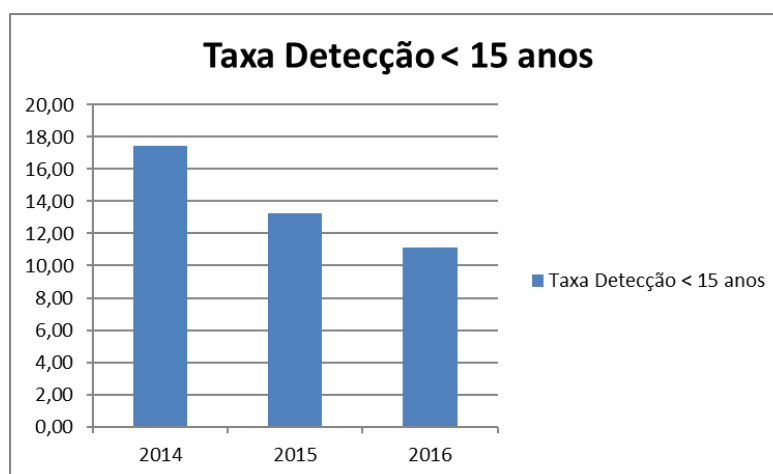
A hanseníase é uma doença infectocontagiosa crônica causada pelo *Mycobacterium leprae*, capaz de provocar importantes incapacidades físicas e deformidades quando não diagnosticada e tratada precocemente (BARRETO et al., 2012). A doença existe em todos os continentes, mas está sob controle nos países desenvolvidos. O Brasil possui uma população

estimada em 2016 de 207,7 milhões de habitantes (IBGE, 2017), e é o 2º país com maior número de casos reportados de Hanseníase, sendo 25.218 somente em 2016, possuindo uma taxa de detecção anual de 12,23 casos para cada grupo de 100.000 habitantes em 2016 (WHO, 2017).

Diante disso, o Brasil ainda não conseguiu atingir a meta de controle estabelecida pela OMS de 1 caso/10.000 habitantes e é categorizado como um país de endemia alta. A distribuição da hanseníase no Brasil não apresenta uniformidade, existindo a formação de clusters de casos em algumas regiões específicas do país (PENNA; OLIVEIRA; PENNA, 2009).

Somente no estado do Pará foram diagnosticados mais de 80.000 casos nos últimos 20 anos e, ainda hoje, com um coeficiente de detecção anual de 30,17/100.000 habitantes em 2016 (Sinan/SVS-MS), a doença permanece como um grave problema de saúde pública neste Estado, onde também é hiperendêmica entre crianças e adolescentes (BARRETO et al., 2014).

Gráfico 1 - Representação da taxa de detecção em menores de 15 anos. Apesar da redução, o valor ainda é alto.



Fonte: Sinan/SVS-MS (adaptado)

O processo de mapeamento, tanto *in loco* quanto remoto, pode ser demorado. As atividades de localizar o endereço, fazer a visita - quando possível-, e então identificar sua longitude e latitude para marcar o ponto correspondente tomam bastante tempo e recursos (CARVALHO; JUNIOR; BARRETO, 2016).

Existem importantes questões éticas envolvidas quanto ao mapeamento exato do endereço do caso, pois mesmo após exportados como imagens, os endereços são passíveis de

reengenharia e podem ser convertidos novamente em coordenadas geográficas (CURTIS; MILLS; LEITNER, 2006). Quando se trabalha com análise por setores, o que realmente importa é se o caso está ou não dentro daquele polígono correspondente, pois serão aplicados tão somente análises estatísticas.

Foram feitas pesquisas no portal de periódicos da CAPES e no Google Scholar, utilizando como palavras-chave “Geocoding”, “GIS”, “Leprosy”, “Python”, visando encontrar soluções similares já aplicadas para o mesmo tema. Entretanto, por ser uma proposta nova, mesmo as soluções aplicadas para a atividade são empíricas, baseadas na heurística dos pesquisadores.

1.1 OBJETIVO

O objetivo deste trabalho foi desenvolver um algoritmo para mapeamento dos casos de hanseníase em seus setores censitários correspondentes. Esta correspondência será identificada utilizando uma estrutura de repetição que deve comparar o *shapefile* de polígonos do IBGE e uma tabela CSV disponibilizada pelo Sistema de Informação de Agravos de Notificação (SINAN) contendo os casos de hanseníase. Esta solução será aplicada em um caso de análise espacial e então serão comparadas esta proposta e as soluções atualmente em uso, com o objetivo de identificar suas vantagens e desvantagens.

Este método pode representar uma economia significativa de tempo e recursos para projetos de análise espacial de agravos. Ao georreferenciar os casos com base nos SC, o caso estará o mais próximo possível do seu endereço correto, possibilitando que sejam feitas análises com dados disponibilizados pelo IBGE, agregados por áreas, tais como: escolaridade, saneamento básico, renda *per capita*, entre outros, sem grande esforço humano, computacional, e não comprometendo a privacidade do paciente.

2 - METODOLOGIA

O embrião deste trabalho surgiu durante as atividades do projeto de pesquisa “Malha digital dos territórios da Estratégia de Saúde da Família no município de Castanhal – PA” (CARVALHO; BARRETO, 2016) (CARVALHO et al., 2015) (CARVALHO; JUNIOR; BARRETO, 2016), onde foram elaborados a malha de cobertura da Estratégia Saúde da Família (ESF) da cidade de Castanhal utilizando a técnica de mapeamento participativo, método utilizado para obter e registrar dados espaciais em parceria com os atores sociais, neste caso os membros de equipes da ESF (GOLDSTEIN et al., 2013), e então analisada a distribuição espacial dos casos de hanseníase em relação à cobertura da ESF.

Todas as atividades foram desenvolvidas utilizando o mesmo computador, para fins de controle e comparação: Notebook Dell Inspiron 3437 – Processador Intel Core® i5-4400U, 1.6GHz, 8gb RAM ddr3 1600MHz, Placa de vídeo NVIDIA 1gb gddr3, 120gb SSD + 1tb HDD e Internet fibra óptica de 20mbps na residência do discente. Software QuantumGIS – QGIS 2.14.20-Essen.

Para aquisição do conhecimento e como referencial técnico para o desenvolvimento das atividades, o Curso Python CodeCademy (codecademy.com) foi fundamental para aprendizado da linguagem de programação. O Curso Python QGIS, da editora Pakt Pub (pakt.com) hospedado no site educacional Udemy (udemy.com), foi elementar para entender as aplicações disponíveis na documentação do SIG QGIS utilizando o Terminal Python integrado ao *software*.

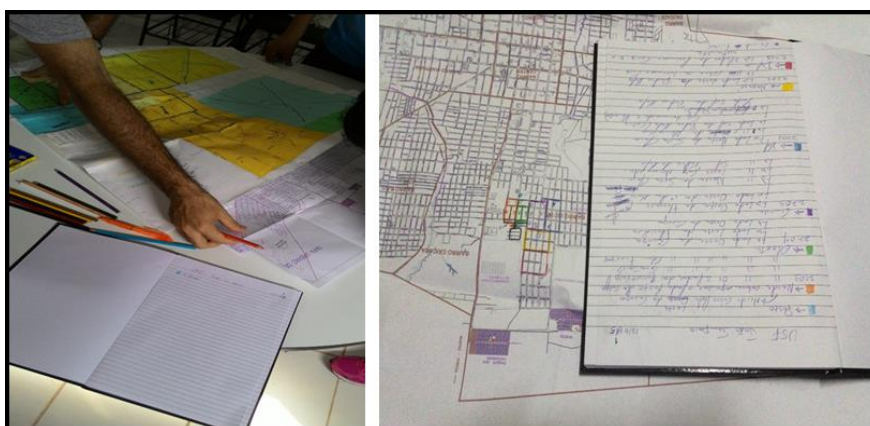
Foram realizadas revisão bibliográfica e da plataforma do SIG QGIS para identificar complementos que realizem a tarefa proposta, semelhanças e diferenças. Por fim, foi realizado um estudo comparativo com os resultados desta solução e os métodos atualmente em uso.

2.1 LEVANTAMENTO DE INFORMAÇÕES

A Estratégia Saúde da Família (antigo Programa de Saúde da Família) foi criada em 1994 como iniciativa do Ministério da Saúde que visa à reorganização da atenção básica no Brasil objetivando a prevenção e o combate de doenças, tais como a hanseníase, de acordo com os preceitos do Sistema Único de Saúde (SUS). Operacionalizada mediante a implantação de equipes multiprofissionais, como os Agentes Comunitários de Saúde (ACS), em Unidades Básicas de Saúde (UBS), responsáveis pelo acompanhamento de um número definido de famílias, localizadas em uma área geográfica delimitada – microáreas (MA) (BRASIL, 2012).

A Malha Digital da Cobertura da ESF foi elaborada em 3 etapas. Em um primeiro momento, foram levantadas as localizações das UBS georreferenciando-as utilizando um receptor portátil de sinal de GPS (Garmin eTrex H, Olathe, KS, USA), e os contatos dos enfermeiros responsáveis pelas equipes implantadas nos postos de saúde mapeados. Em seguida foi aplicado o Mapeamento Participativo, onde a equipe levou um mapa da cidade e uma caixa de lápis coloridos para que cada ACS pintasse as ruas e quadras correspondentes à sua MA de atuação, como pode-se observar na figura 2, e foram coletadas informações pessoais (nome completo, data de nascimento, endereço, telefone) e sobre a microárea (código, número de famílias atendidas, número de pessoas atendidas) para assim, armazená-las em banco de dados objetivando o levantamento de estatísticas a partir desses dados.

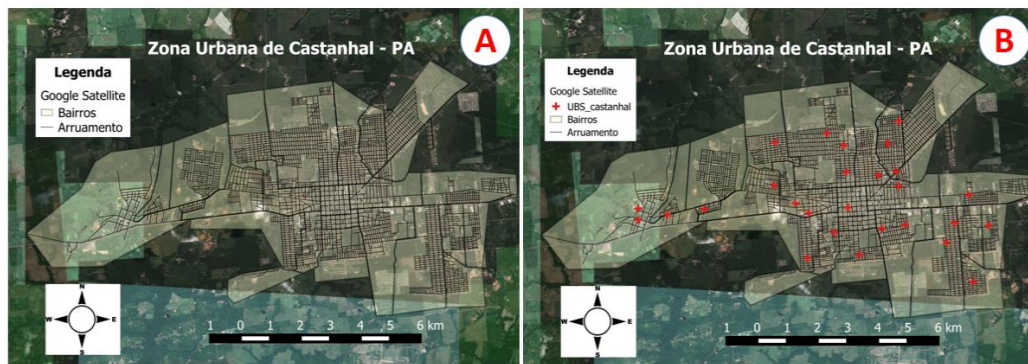
Figura 2 - Caderno e mapa utilizados no mapeamento participativo



Fonte: LabEE

Na terceira etapa, as informações coletadas foram levadas ao laboratório onde foram criados mapas georreferenciados relacionados às MA de atuação de cada ACS. Posteriormente, foram agregadas as informações armazenadas no Banco de Dados referente à MA e aos ACS.

Figura 3 – A: Malha digital da Zona Urbana de Castanhal sobreposta à imagens de satélite. B: Localização das UBS no município



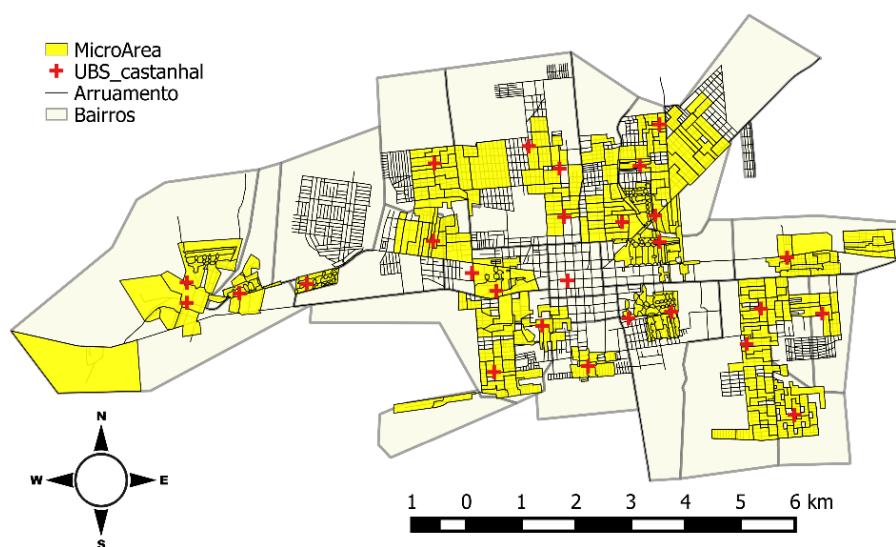
Fonte: LabEE

A camada de MA foi desenvolvida utilizando uma função de atração de camadas para que a delimitação das MA acompanhasse o arruamento. Esta função também inibe a sobreposição de MA, mantendo assim a integridade da informação. A figura 4 apresenta todas as camadas desenvolvidas pelo projeto.

Ao todo, foram mapeadas as MA de 301 ACS implantados na ESF em Castanhal, distribuídos entre as 40 equipes de 25 UBS na zona urbana da cidade, área de interesse desse estudo.

Figura 4- Cobertura da ESF na Zona Urbana de Castanhal – PA

Zona Urbana de Castanhal - Área de atuação da ESF



Fonte: LabEE

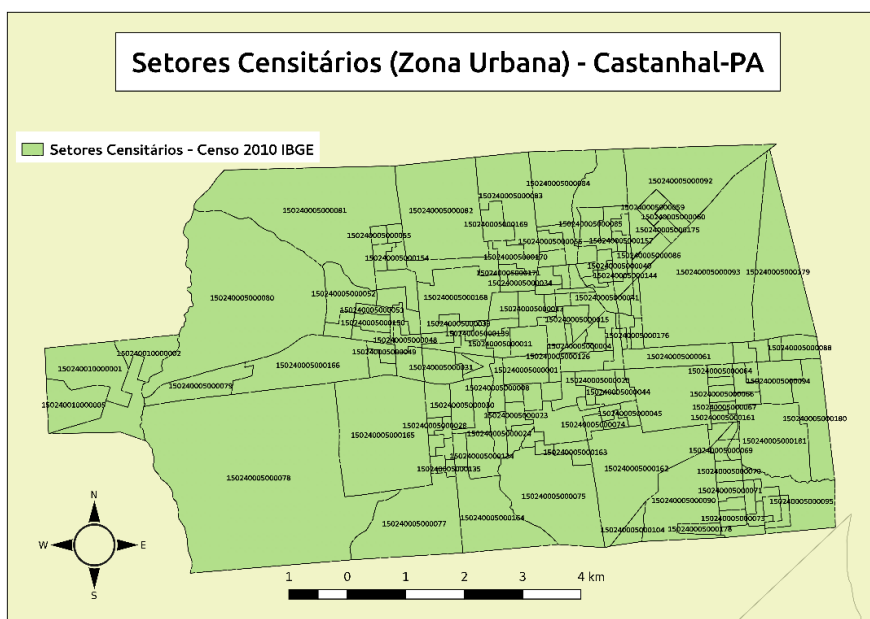
Os casos de hanseníase foram obtidos através da base de dados de casos notificados no Sistema de Informação de Agravos de Notificação do Ministério da Saúde – SINAN, onde seus endereços foram mapeados utilizando aparelhos receptores de sinal GPS.

No processo de análise espacial podem existir desafios que dificultam ou impossibilitam o mapeamento *in loco*. Quando isto ocorre, uma alternativa é mapear manualmente por meio remoto, Em (BARRETO et al., 2014) se utilizou a base nacional de endereços, disponibilizada pelo IBGE, com apoio de imagens de satélite para de alta precisão para visualizar e identificar a longitude e a latitude aproximada. Este método alternativo de mapeamento pode resultar em uma perda de precisão posicional acima de 100 metros, mas permite mapear um endereço dentro seu respectivo setor censitário (BARRETO et al., 2015).

Os Setores Censitários são subdivisões do território nacional estabelecidas pelo IBGE com dimensões e número de domicílios que permitam o levantamento por um recenseador. Para o censo 2010, a base territorial foi dividida em cerca de 314 mil SC. (IBGE, [s.d.]

Os dados levantados pelo IBGE são agregados por setores e constituem uma base de informações extremamente rica, disponibilizada a partir de seu repositório ftp (ftp://geoftp.ibge.gov.br/organizacao_do_territorio/malhas_territoriais/malhas_de_setores_censitarios__divisoes_intramunicipais/censo_2010/setores_censitarios_shp/pa/) no formato *shapefile* para apresentação em ambiente SIG. Cada SC é representado como um polígono, ou seja, um desenho sem forma definida com no mínimo 3 pontos conectados, sendo o mesmo ponto para o começo e o final da área (LAPIG UFG, [s.d.]).

Figura 5 – Setores Censitários de Castanhal. Cada setor está representado com seu código, este sendo composto por 15 dígitos no formato UFMMMMDDSDSSSS. Os dois primeiros números indicam o Estado, seguido por 5 dígitos identificando o município, os próximos 2 dígitos identificam o distrito e em seguida dois dígitos identificam o subdistrito. Os 4 últimos dígitos identificam o setor censitário



Fonte: LabEE/IBGE

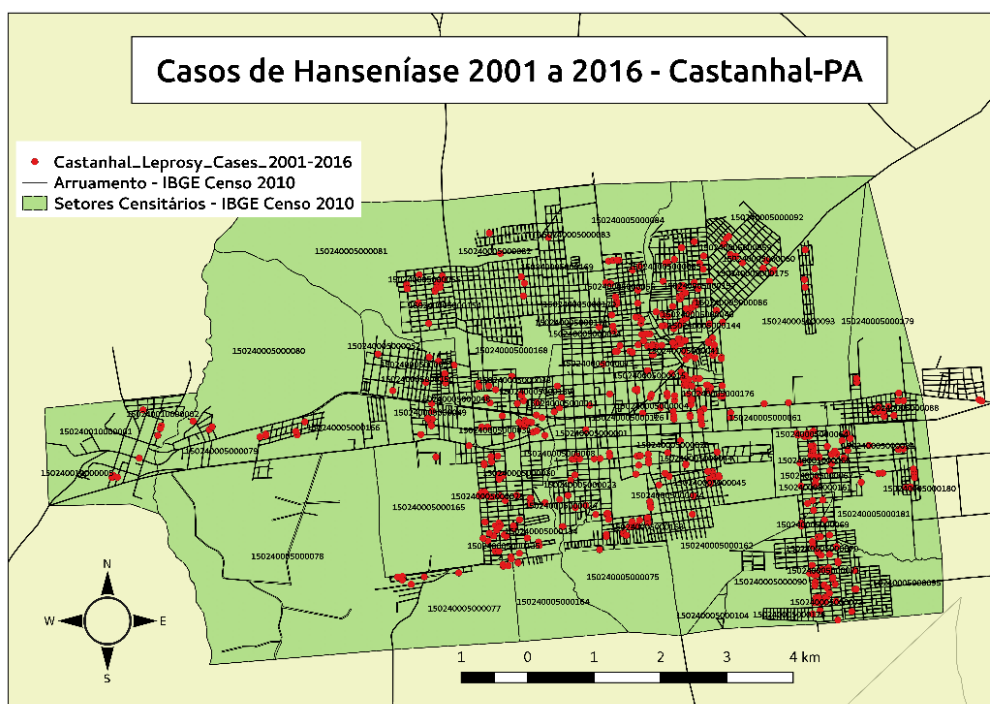
Os casos de hanseníase foram mapeados utilizando o seguinte procedimento:

- I. Através de uma planilha do SINAN contendo os casos notificados no período, foram selecionados aqueles residentes na Zona Urbana de Castanhal. Esta planilha contém uma ficha de cadastro com informações de endereço e um identificador único do caso junto ao Ministério da Saúde;
- II. Utilizando a base de endereços e as camadas *shapefile* disponibilizadas pelo IBGE, foram identificados os endereços e os setores censitários correspondentes;
- III. Um aparelho receptor de sinal GPS e um veículo para visitar os endereços dos casos notificados
- IV. Quando identificado o endereço, o operador marcava um ponto e adicionava então o código do SINAN correspondente.
- V. No laboratório o operador exportou os dados armazenados na memória do dispositivo para um computador, e então importou para visualização no SIG QGIS.

Com base nos dados obtidos, analisou-se primeiramente que das 189.784 pessoas que vivem em Castanhal, 88,6% ou 168.149 vivem na zona urbana de Castanhal, segundo estimado pelo IBGE para o ano de 2015 (IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA, 2016). Do total de 1589 casos notificados no período

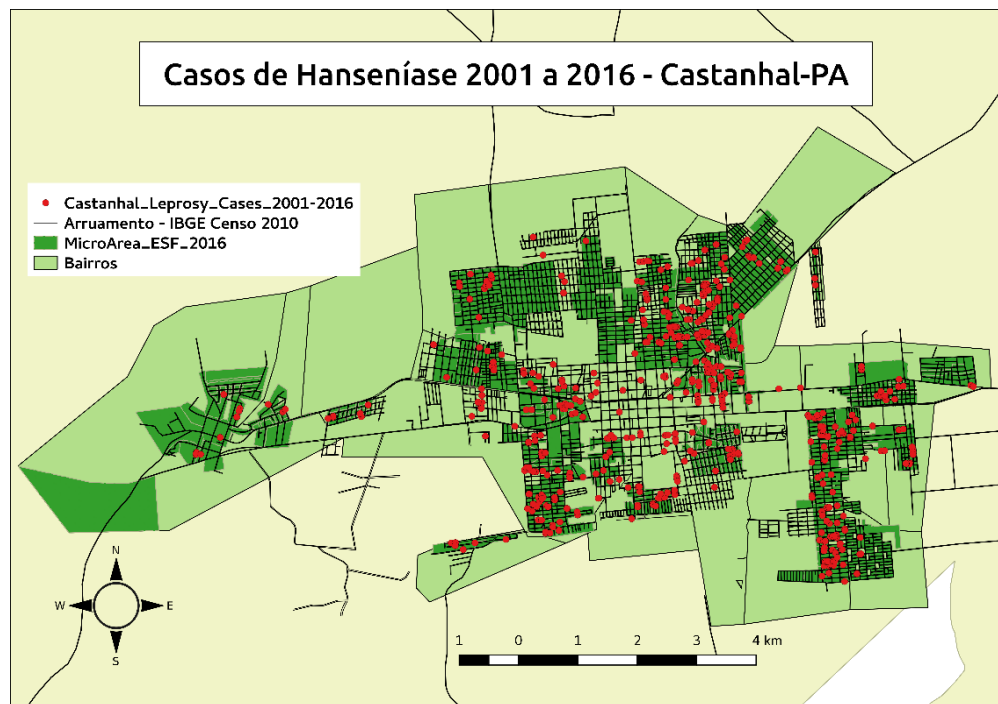
estudado (1322 na Zona Urbana e 322 na Zona Rural), 78,8% dos casos residentes na zona urbana foram mapeados (1042 casos) (Figura 5). Foram criados os mapas digitais georreferenciados da cobertura da ESF na zona urbana e observou-se que 76,38% da população está coberta pela ESF. Cerca de 80,13% (835) dos casos de hanseníase notificados no período estão dentro da área de cobertura da ESF em Castanhal), como apresentado na Figura 6.

Figura 6 – Casos mapeados sobrepostos à camada de Setores Censitários



Fonte: LabEE

Figura 7 – Casos mapeados sobreposto à cobertura da ESF



Fonte: LabEE

3 - ALGORITMO DESENVOLVIDO

As atividades de mapeamento executadas no projeto supracitado demandaram tempo e recursos para serem concluídas. Ao fim do projeto, observou-se um ponto interessante: São atividades repetitivas, que podem – e devem – ser automatizadas para acelerar a coleta de resultados e reduzir a possibilidade de erros.

O desenvolvimento do algoritmo iniciou com importações de módulos com classes que serão utilizados para o processamento dos dados e apresentação dos resultados ao usuário (QtCore e QtGui, respectivamente). Em seguida definiu-se uma variável para receber o caminho do *shapefile* de SC. Utilizou-se a classe *iface* para apresentar o polígono na tela do usuário.

Através da classe *QgsVectorLayer* foi criada uma camada de pontos na memória cache do SIG, definiu-se também o tipo da camada e de sua base de dados. Deste modo, o usuário tem liberdade para salvar a camada ou não. Esta classe também foi útil para acessar o arquivo *csv* contendo a base de informações dos casos de hanseníase para mapear.

Os Sistema de Referência de Coordenadas (SRC) utilizado foi o mesmo configurado no *shapefile* de SC disponibilizado pelo IBGE: EPSG 4019 (<http://spatialreference.org/ref/epsg/unknown-datum-based-upon-the-grs-1980-ellipsoid/>).

Através do comando `source_layer.crs().toWkt()` o SRC adotado foi copiado do *shapefile* de SC.

A base do SINAN não contém originalmente o código do setor censitário. Para o mapeamento dos casos de hanseníase via algoritmo. Foram dispensados tempo para inserção manual deste valor, seguindo o processo descrito a seguir:

- I. Foi adicionado uma coluna à planilha do SINAN, com o título “CD_GEOCODI” (o mesmo que consta na base do IBGE para identificação do setor censitário)
- II. Através do endereço “ftp://ftp.ibge.gov.br/Censos/Censo_Demografico_2010/Cadastro_Nacional_de_Enderecos_Fins_Estatisticos/PA/” obteve-se a base de endereços para o município de Castanhal, Pará (códigos 15024000500 e 15024001000), elaborada a partir do Censo 2010 (IBGE, 2010).
- III. Estas bases foram organizadas em um arquivo *.csv* para facilitar a leitura

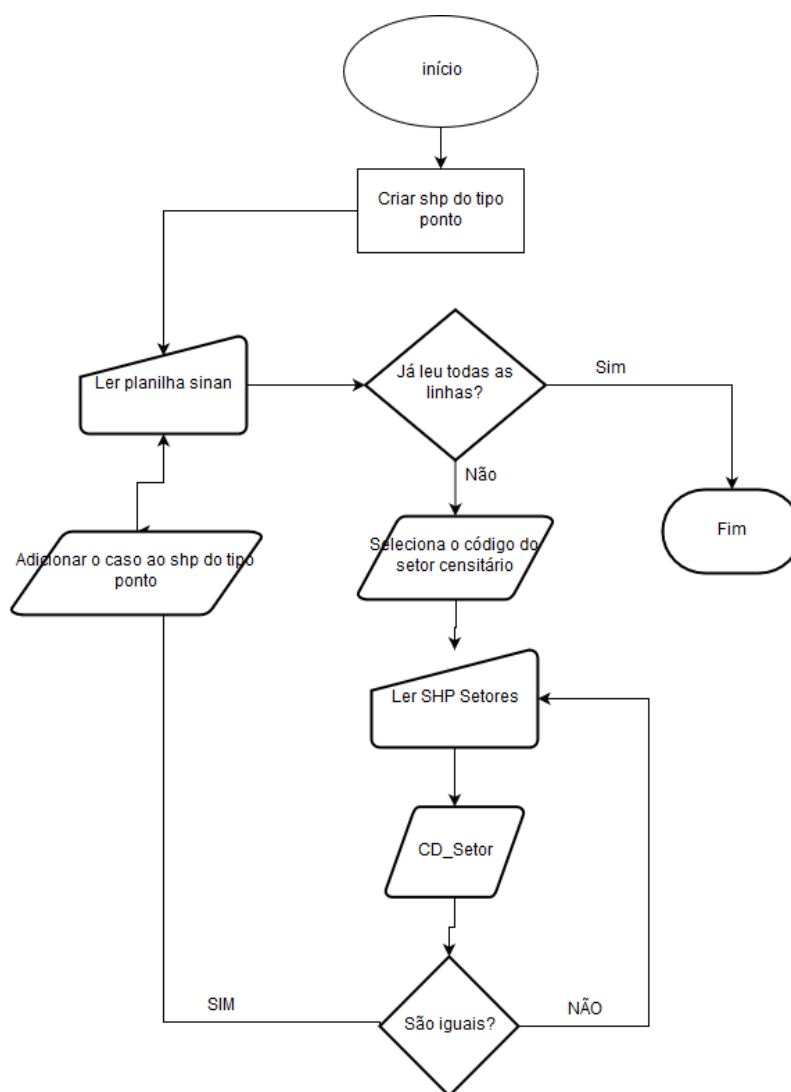
IV. Com a duas planilhas abertas, utilizou-se o campo de filtro para identificar se o endereço descrito na planilha do SINAN estava registrado na planilha do IBGE.

1. Quando positivo, o código do setor correspondente foi copiado para a coluna criada no processo I, e a linha em questão foi realçada com azul.
2. Quando negativo, a linha em questão foi realçada com vermelho.

Processo IV foi repetido até o fim da planilha de casos de hanseníase do SINAN.

Uma estrutura de repetição foi criada para acessar a base de dados do Sistema de Informação de Agravos de Notificação. Dentro desta, um novo "for" foi utilizado para então acessar a camada de SC. Primeiramente, o algoritmo leu um caso da planilha de registros do SINAN, identificando o código do setor censitário, e em seguida percorreu a base de dados do SC para encontrar correspondência. Quando os valores eram iguais, um ponto foi criado com as coordenadas do polígono correspondente e as informações do caso foram registradas na tabela de atributos para identificação. O processo foi repetido até o fim do arquivo de casos e por fim apresentado ao usuário. O diagrama abaixo representa os passos executados pelo algoritmo desenvolvido (apêndice A).

Figura 8 - Diagrama apresentando a lógica implementada no algoritmo



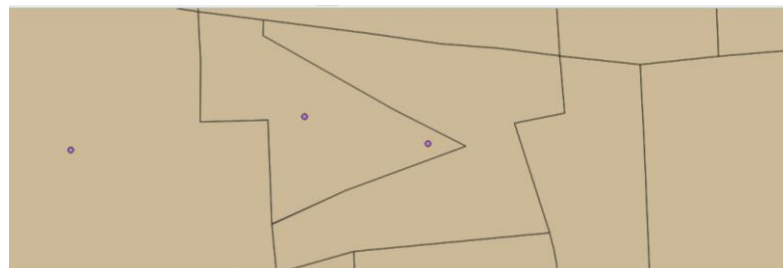
Fonte: Autor

Ao iniciar, o algoritmo deve criar uma camada *shapefile* do tipo ponto para armazenar os futuros casos mapeados. Em seguida o algoritmo vai ler a planilha do SINAN no formato CSV, o caminho para a planilha foi previamente disponibilizado. Após a leitura da planilha do SINAN o algoritmo entrará em uma estrutura de repetição, onde o objetivo é verificar se todas as linhas da planilha já foram lidas. Caso positivo, o algoritmo encerra sua execução. Caso negativo, inicia-se uma outra estrutura de repetição onde será lida a primeira linha com registro de um caso de hanseníase, e selecionada a coluna contendo o código do setor censitário, o algoritmo vai comparar com a base de dados do shp de setores censitários para verificar se o código registrado na planilha do SINAN corresponde a algum SC existente. Caso positivo, um ponto exatamente no centro deste setor será criado e suas coordenadas serão registradas no “shape” criado no início deste algoritmo e em seguida o algoritmo retorna

para a primeira estrutura de repetição a fim de verificar se todas as linhas da planilha de casos já foram lidas e caso negativo segue para a próxima linha executando a segunda estrutura de repetição. Se o código do SC não combinar com nenhum registro no shp do IBGE, o algoritmo vai ler a planilha de casos e testar o registro seguinte. Este processo se repete até o fim da planilha de casos e então o shp é exibido na tela para o usuário, e as duas estruturas de repetição são encerradas.

Logo no primeiro teste foi identificado um problema grave: A documentação da API do QGIS define o centroide como uma função que retorna o centro de massa de uma geometria (<https://qgis.org/api/classQgsGeometry.html#a9af6067caef47d380d45c32569abc10b>). Para áreas côncavas, este centro de massa pode estar externo à área do polígono.

Figura 9 – Erro ao utilizar a função centroid()

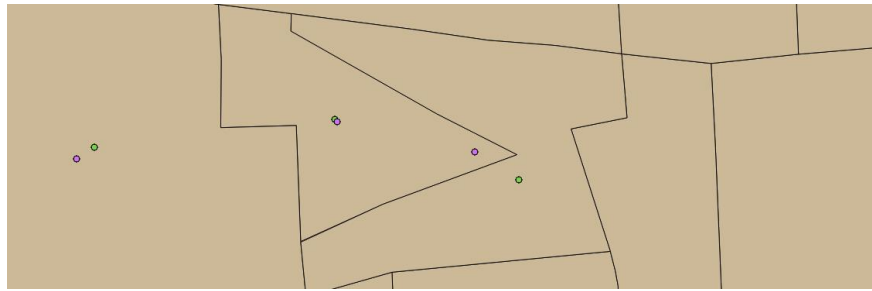


Fonte: Autor

Como apresentado na Figura 9, dois pontos foram marcados em um mesmo polígono e outro polígono não tem nenhum ponto dentro de sua respectiva área. O erro ocorre quando o polígono possui área disforme.

Para corrigir este problema se utilizou a função `pointOnSurface()`, que retornou um ponto inserido dentro do polígono, independentemente de sua forma (<https://qgis.org/api/classQgsGeometry.html#a3975e1ac16832371854703dfe699d5f2>).

Figura 10 – Erro corrigido ao utilizar a função `pointOnSurface()` - em verde os pontos marcados corretamente



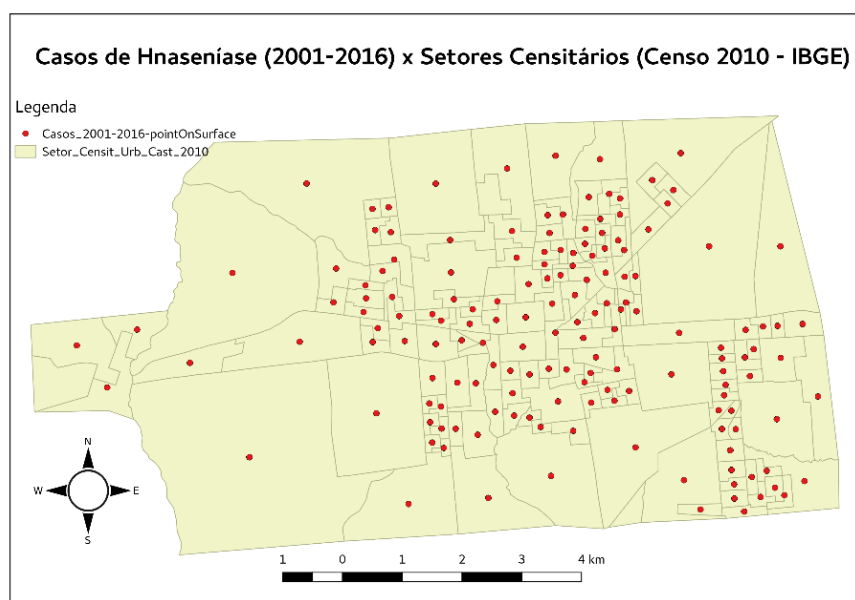
Fonte: Autor

A figura 10 apresenta os pontos marcados utilizando a função `centroid()` (em lilás) e a função `pointOnSurface()` (em verde). É possível verificar que o erro foi corrigido e mesmo que o polígono seja disforme, o ponto correspondente ao seu centroide será marcado exatamente dentro de sua área.

4. RESULTADOS

Ao final do processo, foram mapeados os mesmos 1042 casos presentes na Zona Urbana de Castanhal, de um total de 1589 (1322 ZU / 267 ZR) casos descritos na planilha do SINAN. Também foram excluídos do mapeamento os casos localizados fora da Zona Urbana da cidade de Castanhal, pois não estavam nos objetivos do projeto de pesquisa que deu base para este Trabalho. Também não foram mapeados os casos que o endereço estava incompleto ou não constavam na base do IBGE.

Figura 11 – Mapa final com os casos mapeados

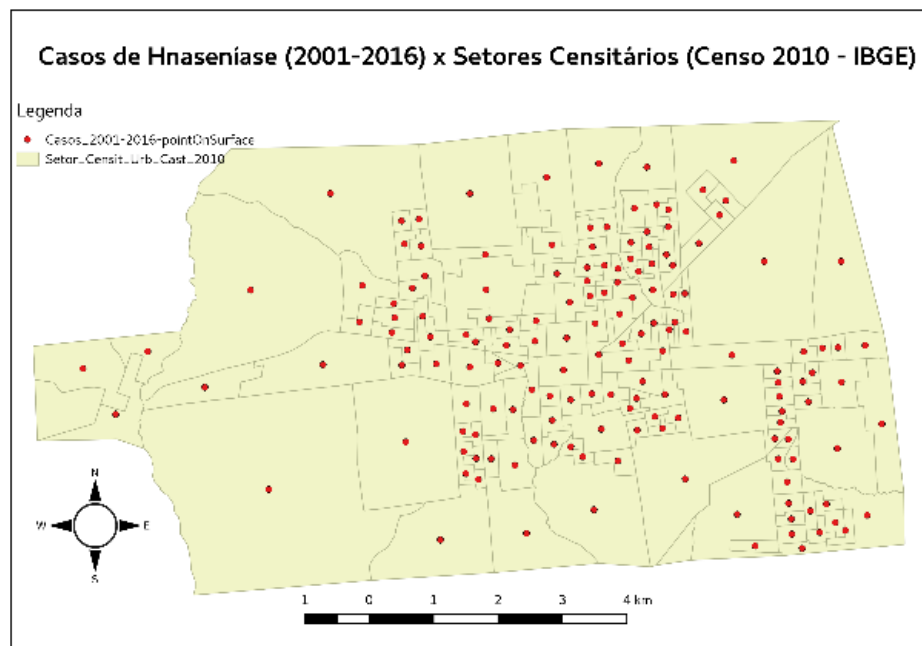


Fonte: Autor

4.1 ALGORITMO X MAPEAMENTO REMOTO

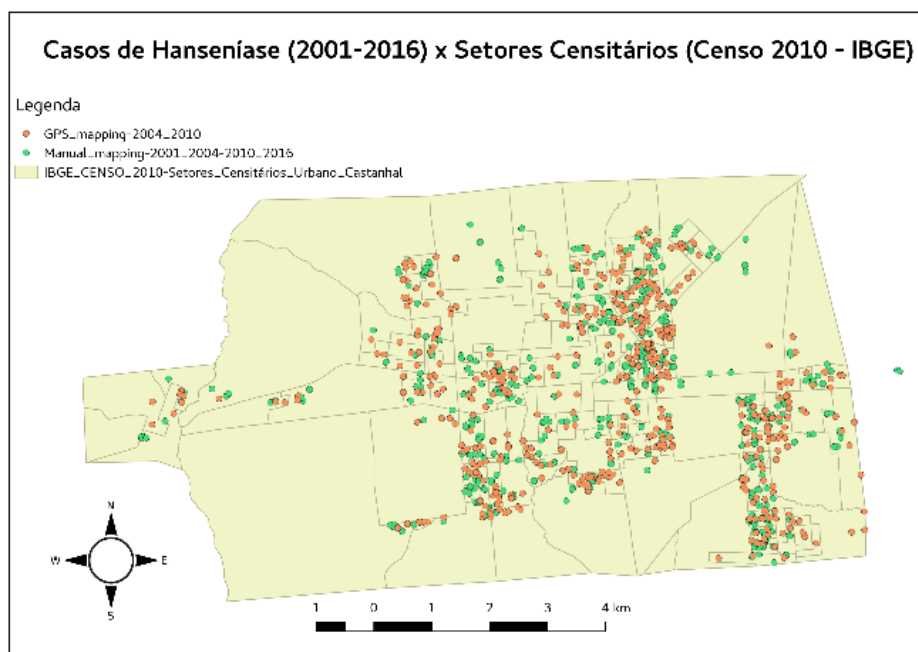
Quando se comparam os métodos atuais com o algoritmo proposto, a impressão é que existe apenas um ponto marcado em cada área no segundo processo. A imagem abaixo evidencia essa suposição.

Figura 12 - Mapa gerado com o mapeamento manual



Fonte: Autor

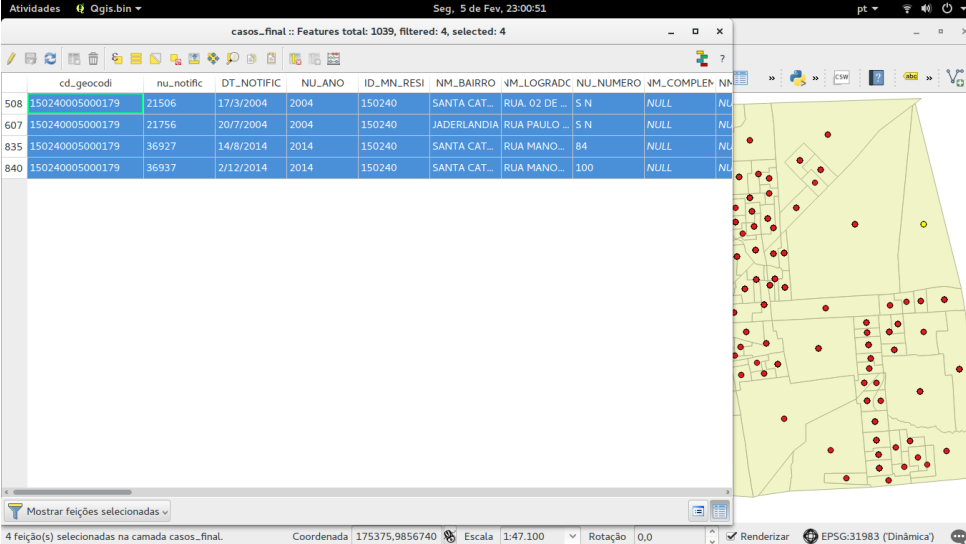
Cada camada apresenta aparentemente um ponto, ou seja, 1 caso registrado, devido todos os pontos marcados apresentarem o mesmo atributo de longitude e latitude. Isto não ocorre no mapeamento utilizando aparelho receptor de sinal GPS ou imagens de satélite, onde objetiva-se mapear o endereço do caso o mais fiel possível, como apresentado na figura abaixo.

Figura 13 – Mapa gerado com o algoritmo

Fonte: Autor

Em laranja, são apresentados os casos mapeados utilizando aparelho receptor de sinal GPS, e em verde são apresentados os casos mapeados remotamente, no laboratório, utilizando imagens de satélite de alta precisão.

Entretanto, todos os casos mapeados estão registrados com a mesma informação espacial no setor correspondente. A quantidade de casos e seus respectivos atributos podem ser acessados facilmente via função “Identificar Feições”. Ao clicar em um – conjunto de – pontos no mapa, será aberto sua respectiva tabela de atributos, conforme figura 14.

Figura 14 – Tabela de Atributos

	cd_geocodi	nu_notific	DT_NOTIFIC	NU_ANO	ID_MN_RESI	NM_BAIRRO	NM_LOGRADC	NU_NUMERO	NM_COMPLEN	NM_COMPLEN
508	150240005000179	21506	17/3/2004	2004	150240	SANTA CAT...	RUA. 02 DE ...	S N	NULL	NU...
607	150240005000179	21756	20/7/2004	2004	150240	JADERLANDIA	RUA PAULO ...	S N	NULL	NU...
835	150240005000179	36927	14/8/2014	2014	150240	SANTA CAT...	RUA MANO...	84	NULL	NU...
840	150240005000179	36937	2/12/2014	2014	150240	SANTA CAT...	RUA MANO...	100	NULL	NU...

Fonte: Autor

Todas as informações registradas são exatamente as mesmas contidas na planilha disponibilizada pelo SINAN, uma vez que o algoritmo registra inicialmente o código do caso, o operador pode utilizar a ferramenta de união de tabelas, onde o QGIS importa a base de dados registrada em uma planilha CSV.

5 - DISCUSSÃO

Este trabalho gerou um método inovador para mapeamento e análise da distribuição espacial dos casos de hanseníase, podendo ser replicado para outras realidades sem grandes adaptações.

O uso de SIG tem se difundido nos últimos anos, apoiado pela popularização dos computadores – que a cada dia ficam mais potentes, - e da Internet. Esta combinação permite a comunicação entre pesquisadores do mundo todo e a benéfica troca de experiências mais rapidamente.

Dentre suas aplicações, destaca-se o uso de SIG no planejamento urbano, de melhores rotas para transporte de cargas (BAROUDY et al., 2014) monitoramento do transporte público (DIN et al., 2016), acompanhamento da qualidade e disponibilidade de recursos hídricos (ZENDE; PATIL; PATIL, 2018), a aplicações na saúde com apoio da epidemiologia espacial (LEITE; FONSECA; BRAZ, 2008).

Aplicar ferramentas de geoprocessamento para entender a distribuição espacial de enfermidades pode ser uma tarefa bastante difícil, considerando os recursos necessários para realizar georreferenciamento *in loco* em determinadas áreas, por diversos motivos; e remotamente devido as ferramentas disponíveis (*google maps, bing maps, open street map*) apresentarem grande margem de erro, principalmente quando se trata de endereços afastados dos grandes centros urbanos.

Utilizar a malha de setores censitários e a base de endereços disponibilizada pelo IBGE reduz esta margem de erro a um nível aceitável, uma vez que possibilita agregar os casos em cada polígono que representa um setor específico e então aferir estatísticas de cada área. Entretanto, o processo de mapear manualmente caso a caso pode demandar tanto tempo quanto o processo físico com um aparelho de GPS.

6 - CONCLUSÕES

Este trabalho apresentou um algoritmo desenvolvido em python para o SIG Qgis, aplicado no mapeamento de casos de hanseníase dentro de setores censitários. Este método não possui a mesma precisão geográfica ao se utilizar aparelho receptor de sinal GPS para mapeamento, porém apresenta o caso dentro do seu contexto sociocultural e econômico, de acordo com os dados coletados através dos censos realizados pelo IBGE. Após a aquisição e organização das informações, o algoritmo geoprocessou corretamente toda a base de dados em poucos minutos, adicionando à cada caso, como informação espacial, coordenadas geográficas que representam o centro do polígono correspondente.

Todo o processo para mapeamento remoto envolveu a identificação do setor censitário correspondente ao caso a ser mapeado e em seguida, utilizando um SIG, adicionar o ponto ao setor, adicionando o código de identificação do SINAN ao ponto marcado. Somente esta atividade pode levar vários minutos por caso, além de ser passível de erro por parte do operador ao digitar o código incorretamente. O mapeamento *in loco* pode demorar até mesmo horas para encontrar o endereço corretamente, além de envolver questões de segurança que podem inviabilizar o trabalho em alguns casos.

O algoritmo realizou o geoprocessamento dos casos em aproximadamente 2 minutos, sem apresentar erros ou pontos inseridos em polígonos incorretos, graças à função `pointOnSurface()`, que corrigiu o erro relacionado à representação de centroides de polígonos disformes.

Automatizar o mapeamento dos casos de hanseníase utilizando um algoritmo que gera coordenadas dentro dos setores censitários correspondentes pode representar uma economia significativa de tempo e recursos, podendo contribuir para dar maior celeridade para análises de distribuição espacial de casos residentes em um mesmo setor censitário. Utilizar esta solução pode ser relevante principalmente para áreas onde as API's de geocodificação (como a do google) não são precisas em indicar a localização do endereço, além de ser possível utilizar *offline* (não tem consultas à internet) o que é mais rápido e torna possível utilizar geoprocessamento em áreas com pouco acesso à internet e endereços que ainda não estão listados nestas API's.

6.1 TRABALHOS FUTUROS

Este trabalho ainda está em desenvolvimento e tem diversos estudos e melhorias a serem feitas. Os seguintes itens estão em fase de estudo ou desenvolvimento e serão divulgados em trabalhos futuros:

- Desenvolver um algoritmo para normalização dos dados da tabela do SINAN e inserção automática dos códigos de Setores Censitários, com base na planilha de endereços do IBGE.
- Acrescentar ao algoritmo apresentado uma função para exportar em planilha ou documento de texto os casos que não foram mapeados e o respectivo motivo.
- Acrescentar ao algoritmo apresentado maior precisão, mapeando os casos na rua e/ou quadra correspondente.
- Aplicar o algoritmo em casos de outras realidades e/ou doenças.

REFERÊNCIAS

- BARRETO, J. G. et al. High rates of undiagnosed leprosy and subclinical infection amongst school children in the Amazon Region High anti-phenolic glycolipid-I IgM titers and hidden leprosy cases, Amazon region. **Mem Inst Oswaldo Cruz**, v. 107 Suppl, n. 1999, p. 60–67, 2012.
- BARRETO, J. G. et al. Spatial Analysis Spotlighting Early Childhood Leprosy Transmission in a Hyperendemic Municipality of the Brazilian Amazon Region. **PLoS Neglected Tropical Diseases**, v. 8, n. 2, 2014.
- BARRETO, J. G. et al. Spatial epidemiology and serologic cohorts increase the early detection of leprosy. **BMC infectious diseases**, v. 15, n. 1, p. 527, 2015.
- BAROUDY, A. A. EL et al. Geographic Information System (GIS) modeling approach to determine the fastest delivery routes. *Computers and Geosciences*, v. 62, n. 3, p. 128–135, 2014.
- BRASIL. **Política Nacional de Atenção Básica**. [s.l: s.n.]. v. I
- CARVALHO, P. I. O. et al. Georreferenciamento dos Territórios da Estratégia Saúde da Família Auxiliando na Vigilância Epidemiológica da Hanseníase. v. 40, n. Suppl 1, p. 19825161, 2015.
- CARVALHO, P. I. O.; BARRETO, J. G. **Malha digital dos territórios da Estratégia de Saúde da Família no município de Castanhal – PA**. [s.l: s.n.].
- CARVALHO, P. I. O.; JUNIOR, F. E. DE A. F.; BARRETO, J. G. Estratégia Saúde da Família e Hanseníase: Análise Territorial. **Journal of Health Informatics - ISSN 2175-4411**, v. 8, n. I, p. 353–360, 2016.
- CURTIS, A. J.; MILLS, J. W.; LEITNER, M. Spatial confidentiality and GIS: Re-engineering mortality locations from published maps about Hurricane Katrina. **International Journal of Health Geographics**, v. 5, p. 1–12, 2006.
- DIN, M. A. M. et al. The Use of Geographical Information System in the Assessment of Level of Service of Transit Systems in Kuala Lumpur. *Procedia - Social and Behavioral Sciences*, v. 222, p. 816–826, 2016.
- ENVIRONMENTAL SYSTEMS RESEARCH INSTITUTE, I. ESRI Shapefile Technical Description. **Computational Statistics**, v. 16, n. July, p. 370–371, 1998.
- GOLDSTEIN, R. A. et al. A experiência de mapeamento participativo para a construção de uma alternativa cartográfica para a ESF. **Ciência & saúde coletiva**, v. 18, n. 1, p. 45–56, jan. 2013.
- HALLIDAY, S. Death and miasma in Victorian London: an obstinate belief. **BMJ: British Medical Journal**, v. 323, n. 7327, p. 1469–1471, 2001.
- IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Cidades**. Disponível em: <<http://cod.ibge.gov.br/32T>>. Acesso em: 27 maio. 2016.
- IBGE, I. B. DE G. E E. **Censo 2010 - Operação Censitária**. Disponível em: <<https://censo2010.ibge.gov.br/materiais/guia-do-censo/operacao-censitaria.html>>. Acesso em: 30 jan. 2018.
- IBGE, I. B. DE G. E E. **IBGE divulga as estimativas populacionais dos municípios para**

2017. Disponível em: <<https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/16131-ibge-divulga-as-estimativas-populacionais-dos-municipios-para-2017.html>>. Acesso em: 28 set. 2017.

KHAN, S. et al. Positional accuracy of geocoding from residential postal codes versus full street addresses. **Health Reports**, v. 29, n. 2, p. 3–9, 2018.

LAPIG UFG. **Dados Vetoriais**. Disponível em: <https://www.lapig.iesa.ufg.br/lapig/cursos_online/qgis/dados_vetoriais.html>. Acesso em: 30 jan. 2018.

LEITE, M. E.; FONSECA, D. DE S. R.; BRAZ, C. K. R. Uso do SIG na Análise da Dengue: Aplicação na Microrregião de Montes Claros/Bocaiúva (MG). **HYGEIA - Revista Brasileira de Geografia Médica e da Saúde**, v. 3, n. 6, p. 126–141, 2008.

PENNA, M. L. F.; OLIVEIRA, M. L. V. D. R.; PENNA, G. O. The epidemiological behaviour of leprosy in Brazil. **Leprosy review**, v. 80, n. 3, p. 332–344, 2009.

Shape – Definições e Conversão | SEMACE. Disponível em: <<http://www.semace.ce.gov.br/2011/06/shape-definicoes-e-conversao/>>. Acesso em: 29 jan. 2018.

SNOW, J. **On the Mode of Communication of Cholera** (J. Churchill, Ed.) **John Chruchill**. London, England: [s.n.].

WHO. **Number of reported cases**. Disponível em: <<http://apps.who.int/gho/data/view.main.95300>>. Acesso em: 28 set. 2017.

WORLD OF HEALTH ORGANIZATION. **Global Leprosy Strategy 2016–2020, Accelerating towards a leprosy-free world** New Delhi, 2016. Disponível em: <<http://apps.who.int/iris/bitstream/10665/205149/1/B5233.pdf?ua=1>>

ZENDE, A. M.; PATIL, R. A.; PATIL, V. M. Water Security Assessment in Semi-arid Region using Geospatial Techniques. **Materials Today: Proceedings**, v. 5, n. 1, p. 620–627, 2018.

APÊNDICE A - ALGORITMO PARA REPRODUÇÃO

```

from QGIS.PyQt.QtGui import *
from QGIS.PyQt.QtCore import *
import processing

#Use # to write comments
#define the shp input like those examples above.
#src_shp = "/media/user/path/polygon.shp" #for linux
src_shp = "D:\path\polygon.shp" #for windows
source_layer = iface.addVectorLayer(src_shp, "Setor", "ogr")
crs = source_layer.crs().toWkt()

# Define the output. It will be saved on memory for now. The user can store in the
end.
pt_layer =
  QgsVectorLayer('Point?field=cd_geocodi:string&field=nu_notific:string&crs=' +
    crs, 'MyPoint' , 'memory')

# read a .csv
uri = "file:///media/user/path/database.csv?delimiter=%s" % (",") #for linux
#uri = "file:///C:/Users/User/database.csv?delimiter=%s" % (",") #for windows
db = QgsVectorLayer(uri, 'DataBase','delimitedtext')
db.fields()[0].name()
QgsMapLayerRegistry.instance().addMapLayer(db)

#start edition of point layer
pt_layer.startEditing()

for feat in db.getFeatures():
  nmSet = feat.attribute('CD_SETOR') #or just set the position like [0] or [1]
  attrs = feat.attribute('NU_NOTIFIC') #or just set the position like [0] or [1]
  iface.mainWindow().statusBar().showMessage("Please, wait!")
  for source_feature in source_layer.getFeatures():

```

```
# select the attribute
iface.mainWindow().statusBar().showMessage("Reading a new census trackt")
name = source_feature.attribute("CD_GEOCODI")

if nmSet == name:
    geometry = source_feature.geometry()
    pt_surf = geometry.pointOnSurface()
    #https://QGIS.org/api/classQgsGeometry.html#a3975e1ac16832371854703dfe699d5f
    2
    # Create the new feature with the fields of the memory layer
    # And set geometry and attribute before adding it to the target layer

    pt_feature = QgsFeature(pt_layer.fields())
    pt_feature.setGeometry(pt_surf)
    pt_feature['cd_geocodi'] = name
    pt_feature['nu_notific'] = attrs
    pt_layer.addFeature(pt_feature)

    pt_layer.commitChanges()
    # Add the layer to the registry
    QgsMapLayerRegistry.instance().addMapLayer(pt_layer)

    #https://docs.QGIS.org/2.8/en/docs/user_manual/processing_algs/QGIS/vector_analy
    sis_tools/countpointsinpolygonweighted.html
    #counting points inside polygons

    Result = "D:\path\output.shp"
```