



TEXTO PARA DISCUSSÃO Nº567

**GEOLOCALIZAÇÃO DE INTERNAÇÕES: UMA SOLUÇÃO BASEADA NO PROGRAMA
ESTATÍSTICO R PARA A AMPLIAÇÃO DE POSSIBILIDADES DE ANÁLISE BASEADAS
NO SISTEMA DE INFORMAÇÃO HOSPITALAR**

**Thiago Augusto Hernandez Rocha
Núbia Cristina da Silva
Pedro Vasconcelos Amaral
João Ricardo Nickenig Vissoci
Erika Bárbara Abreu Fonseca Thomaz
Rejane Christine de Sousa Queiroz
Matthew Harris
Christopher Millet
Thomas Hone
Antonio Augusto Moura da Silva
Luiz Augusto Facchini**

Setembro de 2017

Universidade Federal de Minas Gerais

Jaime Arturo Ramírez (Reitor)

Sandra Regina Goulart Almeida (Vice-reitora)

Faculdade de Ciências Econômicas

Paula Miranda-Ribeiro (Diretora)

Lizia de Figueirêdo (Vice-diretora)

Centro de Desenvolvimento e Planejamento Regional (Cedeplar)

Mônica Viegas Andrade (Diretora)

Eduardo da Motta e Albuquerque (Vice-Diretor)

José Irineu Rangel Rigotti (Coordenador do
Programa de Pós-graduação em Demografia)

Marco Flávio da Cunha Resende (Coordenador do
Programa de Pós-graduação em Economia)

Laura Lúcia Rodríguez Wong (Chefe do
Departamento de Demografia)

Edson Paulo Domingues (Chefe do Departamento
de Ciências Econômicas)

Editores da série de Textos para Discussão

Aline Souza Magalhães (Economia)

Adriana de Miranda-Ribeiro (Demografia)

Secretaria Geral do Cedeplar

Maristela Dória (Secretária-Geral)

Simone Basques Sette dos Reis (Editoração)

<http://www.cedeplar.ufmg.br>

Textos para Discussão

A série de Textos para Discussão divulga resultados preliminares de estudos desenvolvidos no âmbito do Cedeplar, com o objetivo de compartilhar ideias e obter comentários e críticas da comunidade científica antes de seu envio para publicação final. Os Textos para Discussão do Cedeplar começaram a ser publicados em 1974 e têm se destacado pela diversidade de temas e áreas de pesquisa.

Ficha catalográfica

G345 2017	Geolocalização de internações: uma solução baseada no programa estatístico R para a implantação de possibilidades de análise baseadas no sistema de informação hospitalar / Thiago Augusto Hernandes Rocha ... [et al.]. - Belo Horizonte: UFMG/CEDEPLAR, 2017. 17 p. : il. - (Texto para discussão, 567) Inclui bibliografia (p. 27-31) ISSN 2318-2377 1. Sistemas de recuperação da informação - Saúde pública - Brasil. 2. Indicadores de saúde - Brasil I. Rocha, Thiago Augusto Hernandes. II. Universidade Federal de Minas Gerais. Centro de Desenvolvimento e Planejamento Regional. III. Título. IV. Série. CDD: 614.981
--------------	--

Elaborada pela Biblioteca da FACE/UFMG - JN /2017

As opiniões contidas nesta publicação são de exclusiva responsabilidade do(s) autor(es), não exprimindo necessariamente o ponto de vista do Centro de Desenvolvimento e Planejamento Regional (Cedeplar), da Faculdade de Ciências Econômicas ou da Universidade Federal de Minas Gerais. É permitida a reprodução parcial deste texto e dos dados nele contidos, desde que citada a fonte. Reproduções do texto completo ou para fins comerciais são expressamente proibidas.

Opinions expressed in this paper are those of the author(s) and do not necessarily reflect views of the publishers. The reproduction of parts of this paper or of data therein is allowed if properly cited. Commercial and full text reproductions are strictly forbidden.

**UNIVERSIDADE FEDERAL DE MINAS GERAIS
FACULDADE DE CIÊNCIAS ECONÔMICAS
CENTRO DE DESENVOLVIMENTO E PLANEJAMENTO REGIONAL**

**GEOLOCALIZAÇÃO DE INTERNAÇÕES: UMA SOLUÇÃO BASEADA NO PROGRAMA
ESTATÍSTICO R PARA A AMPLIAÇÃO DE POSSIBILIDADES DE ANÁLISE BASEADAS
NO SISTEMA DE INFORMAÇÃO HOSPITALAR**

Thiago Augusto Hernandes Rocha

Universidade Federal de Minas Gerais, Faculdade de Ciências Econômicas, CEPEAD,
Belo Horizonte, Minas Gerais, Brasil.

Núbia Cristina da Silva

Universidade Federal de Minas Gerais, Faculdade de Ciências Econômicas, CEPEAD,
Belo Horizonte, Minas Gerais, Brasil.

Pedro Vasconcelos Amaral

Centre for Development and Regional Planning, Federal University of Minas Gerais

João Ricardo Nickenig Vissoci

Division of Emergency Medicine, Duke University Health System,
Duke Global Health Institute, Duke University – Durham, North Carolina, United States.

Erika Bárbara Abreu Fonseca Thomaz

Department of Public Health, Federal University of Maranhão

Rejane Christine de Sousa Queiroz

Department of Public Health, Federal University of Maranhão

Matthew Harris

Faculty of Medicine, School of Public Health, Imperial College London

Christopher Millet

Faculty of Medicine, School of Public Health, Imperial College London

Thomas Hone

Faculty of Medicine, School of Public Health, Imperial College London

Antonio Augusto Moura da Silva

Department of Public Health, Federal University of Maranhão

Luiz Augusto Facchini

Universidade Federal de Pelotas, Faculdade de Medicina,
Departamento de Medicina Social, Pelotas, Rio Grande do Sul, Brasil.

CEDEPLAR/FACE/UFGM

BELO HORIZONTE

2017

SUMÁRIO

INTRODUÇÃO	6
MÉTODOS (DESENVOLVIMENTO DA SOLUÇÃO)	7
RESULTADOS	12
DISCUSSÃO	14
AGRADECIMENTOS	15
CONTRIBUIÇÃO DOS AUTORES	15
REFERÊNCIAS	16

RESUMO

Objetivo: Este trabalho detalha um método de geolocalização dos dados das autorizações de internação hospitalar, divulgadas através do sistema de informação hospitalar.

Métodos: o presente trabalho pode ser classificado como de desenvolvimento metodológico. Para espacializar os dados das AIHs foi estruturado um script em linguagem R, baseando-se nos pacotes download DataSUS e CepR. O script foi aplicado para levantamento das AIH sem Goiás, para o ano de 2015. Após o download e pré-processamento dos dados, foi realizada a espacialização das AIHs que puderam ser geolocalizadas.

Resultados: Foram geolocalizadas 97,58% das AIHs processadas em 2015 em Goiás. Dentre as 295.670 AIHs processadas foi possível extrair 24.220 CEPs diferentes com uma taxa de sucesso de geolocalização de 98,72%.

Conclusão: O método detalhado nesse trabalho abre um novo leque de possibilidades para o desenho de estudos avaliativos, formulação de políticas e planejamento de ações de cuidado em saúde.

Palavras chave: Hospitalização, Sistemas de Informação, Processamento Automatizado de Dados, Análise de Dados, Pesquisa em Sistemas de Saúde Pública.

Classificação JEL: I180

ABSTRACT

Objective: This work details a method of geolocation of data on hospital admission authorizations, disclosed through the hospital information system.

Methods: the present work can be classified as methodological development. In order to spatialize data of the AIHs we created a script using the software R, based on the packages download Data SUS and CepR. The script was applied to survey the AIHs in Goiás, for the year 2015. After downloading and pre-processing of the data, the IAHs were spatialized and could be geolocated.

Results: 97.58% of the AIHs processed in 2015 in Goiás were geolocated. Of the 295,670 AIHs processed, it was possible to extract 24,220 different Zip code with a success rate of geolocation of 98.72%.

Conclusion: The method detailed in this paper opens a new range of possibilities for the design of evaluative studies, formulation of policies and planning of health care actions.

Key words: Hospitalization, Information Systems, Automatic Data Processing, Data Analysis, Public Health Systems Research.

JEL Classification: I180

INTRODUÇÃO

O Brasil se configura como um espaço de discussão privilegiado, no que tange à avaliação em saúde, em função da publicização de diversas informações sobre aparato prestador de serviços de saúde no país. Nesse sentido, destaca-se o papel do Departamento de Informática do SUS (DATASUS) (1) que disponibiliza diversos conjuntos de informações em saúde, sobre mais de um dúzia de sistemas de informação diferentes. Apesar da relevância dessa iniciativa, pesa ainda alguma dificuldade quanto à documentação das informações divulgadas, limitações quanto à possibilidade de desagregação dos dados e dificuldades operacionais para o manejo de um grande número de arquivos.

Dentro do universo de sistemas de informação que compõem o rol de dados públicos divulgados pelo DATASUS um merece destaque: o Sistema de Informações Hospitalares (SIH). O SIH é o sistema responsável pelo processamento das informações originárias junto às autorizações para internação hospitalar (AIHs) processando cerca de 1 milhão de autorizações por mês (2). Este mesmo sistema dispõe de recursos pagos a cada hospital do SUS, códigos de classificação internacional de doenças (CID) associados às internações, tempo de permanência média, código de endereçamento postal (CEP) de cada AIH, dentre outras informações essenciais para atividades de monitoramento, vigilância e avaliação em saúde. Graças às suas especificidades, esse sistema de informação tem sido amplamente utilizado para subsidiar estudos avaliativos em saúde no Brasil(3–7).

A despeito da potencialidades associadas ao SIH, a maior parte dos dados divulgados possui nível de desagregação máximo até a esfera municipal (2). Realizar análises mais granulares, observando-se características intramunicipais, é geralmente mais difícil e quase sempre implica a necessidade de coleta de dados que extrapolam aqueles usualmente divulgados pelo DATASUS. Além disso, a necessidade de manipular um grande volume de dados, com pouca documentação, acaba por exigir do pesquisador o manejo de ferramentas com capacidade de flexibilização restrita, permeadas por dificuldades de acesso, barreiras para tratamento e limitadas do ponto de vista de divulgação das informações, como é o caso do TABWIN e TABNET(8). Alguns esforços tem sido realizados para melhoria da documentação relacionada ao SIH, como o trabalho de Santos (9). Apesar disso, os desafios ainda persistem. Superar tais limitações, para propiciar o desenvolvimento de estudos mais aprofundados, é um entrave que se coloca para os acadêmicos e profissionais do SUS.

A premência de se poder avaliar os eventos que ocorrem dentro da esfera municipal é essencial para a realização de estudos avaliativos em saúde. Incorre em equívoco aquele que considera o território geográfico como apresentando características uniformes, especialmente, no que concerte a eventos de saúde. Cidades como Belo Horizonte, por exemplo, possuem bairros com Índice de desenvolvimento humano (IDH) equivalente ao da Suíça, bem como regiões com IDH equivalente àqueles observados na África Subsaariana(10). Ao se analisar indicadores municipais agregados, ambos os extremos são considerados de forma agregada, produzindo um efeito de homogeneização de diferenças catalisador de erros associados à falácia ecológica(11).

A capacidade de espacializar eventos de saúde abre possibilidade para a obtenção de explicações adicionais sobre a forma como a saúde da população é afetada por outros elementos que não apenas o cuidado médico. A importância de determinantes sociais e geográficos é crucial para

melhorar a compreensão das formas de articulação da díade saúde-doença (12). Levando-se em conta a importância de estudos baseados na localização geográfica de eventos em saúde(13–15)e a relativa incipiência de trabalhos amparados nesse tipo de metodologia no contexto brasileiro,(12) pesa a necessidade de desenvolvimento estratégias mais propositivas para a espacialização de eventos de saúde. Guimarães (12) destaca que há uma importante agenda a ser abordada,voltada para a disseminação de métodos de inferência espacial, focados no suporte à decisão em ações de monitoramento e vigilância em saúde.

A importância de se desenvolver soluções capazes de permitir o exame de características intramunicipais, amparadas na espacialização de eventos em saúde, associada às limitações e dificuldades inerentes ao manejo do TABWIN e TABNET contribuíram para o desenho do presente trabalho, que almeja apresentar uma solução para permitir a geolocalização de internações processadas junto ao SIH. Ao fazê-lo, almeja-se abrir a possibilidade, para que aspectos mais granulares associados às internações possam ser melhor compreendidos.

MÉTODOS (DESENVOLVIMENTO DA SOLUÇÃO)

O presente estudo se configura como de natureza metodológica, uma vez que procura definir etapas, regras para validação e a avaliação de métodos que possam ser utilizados para fins de pesquisa(16).

Para permitir a espacialização das internações processadas junto ao SIH alguns desafios precisam ser superados, a destacar: necessidade manejo de múltiplos arquivos ‘RD*.dbc’ através dos quais são divulgados os dados; extração dos dados do CEP referente a cada AIH, assim como proceder a sua geolocalização e, por último, realizar o pareamento, limpeza e categorização dos dados até então analisados para permitir a sua utilização em estudos avaliativos. Uma vez cumpridas essas etapas é possível obter as coordenadas de latitude e longitude de cada AIH, com precisão restrita à rua de origem do paciente. Dessa forma se torna factível a realização de análises mais granulares, que aquelas realizadas apenas com indicadores municipais agregados.

Para estruturação da solução optou-se pela utilização da linguagem do software R. O R funciona com objetos em um espaço de trabalho, nos quais dentre os objetos mais utilizados encontram-se os *data frames*(17). Estes objetos são como os bancos de dados utilizados em outros softwares estatísticos como STATA, MINITAB e SPSS. Além disso, o R permite a inclusão de pacotes. Cada pacote reúne um conjunto de funções que adicionam capacidades ao R, de forma a tornar possível a realização de novas tarefas e operações. O Comprehensive R Archive Network (CRAN) é o repositório oficial de pacotes para o R. Não obstante, pacotes de outras fontes podem ser instaladas, dentre as quais se destaca o Github (18). Graças ao desenvolvimento de pacotes em linguagem R direcionados para o DATASUS é que a superação dos desafios relacionados à geolocalização das AIHs se tornou possível.

O primeiro desafio perpassa o download dos dados divulgados junto ao SIH. As informações sobre as internações são divulgadas no endereço ftp:

<ftp://ftp.datasus.gov.br/dissemin/publicos/SIHSUS/200801/>. Os arquivos são nomeados segundo o padrão: RDUFAAMM, no qual RD é a abreviação para reduzida, UF para unidade da federação, AA para ano e MM para mês. Assim, para cada estado, há um arquivo reduzido mensal contendo os dados referentes às internações realizadas pelos estabelecimentos de cada unidade da federação. Santos fez uma descrição de grande qualidade sobre o detalhamento das informações contidas em cada variável divulgada nos arquivos ‘RD*.dbc’ do SIH (9). Para se analisar um único ano há a necessidade de manejo de 12 arquivos (um por mês) para cada um dos 26 estados mais o Distrito Federal. Pesa ainda, a necessidade descompactar cada um desses arquivos, pois eles são divulgados em formato ‘*.dbc’. Para que as análises sejam feitas, realiza-se a conversão em ‘*.dbf’. Para abordar os desafios vinculados a esse processo oneroso, foi desenvolvido por Rafael Saldanha o pacote: download DataSUS(19).

O pacote para o R download DataSUS apresenta funções para download dos arquivos de microdados do DataSUS (formato DBC), leitura dos arquivos através do pacote read.dbc e pré-processamento para utilização(19). A descompactação do formato *.dbc para *.dbf é realizada de forma automatizada com base em outro pacote intitulado Read.dbc(20). Esses pacotes podem ser instalados a partir do Github através do uso do pacote devtools, diretamente na front-end/GUI utilizada (própria do R, RStudio, RKward.)(17) com os comandos:

```
install.packages("devtools") # instala o pacote devtools
devtools::install_github("rfsaldanha/downloadDataSUS")
devtools::install_github("danicatt/read.dbc")
```

O pacote download DataSUS possui apenas duas funções: `datasusFetch` e `datasusProcess`. A primeira se destina a automatizar os downloads de arquivos a partir do DATASUS, já realizando o processo de descompactação e agregação dos arquivos individuais em um data.frame único. Ao realizar tal tarefa, essa função reduz de modo substancial o número de operações necessárias para criar um banco de dados processável dos dados disponíveis no DATASUS. A segunda função se destina ao pré-processamento dos dados baixados do DataSUS, atribuindo rótulos às variáveis brutas obtidas por meio da utilização da função: `datasusFetch`.

A função `datasusFetch` possui sete argumentos:

- `anoIni` : Ano inicial.
- `mesIni` : Mês inicial.
- `anoFim` : Ano final.
- `mesFim` : Mês final.
- `uf` : Unidade Federativa.
- `sistema` : Sistema de informações.
- `vars` : Variáveis de interesse.

Os argumentos iniciados com “ano” ou “mês” são destinados à definição do ano e mês de início e fim do conjunto de dados que devem ser baixados, respectivamente. O argumento UF define quais estados devem ser baixados. O argumento sistema define de quais sistemas os dados devem ser baixados. Apesar de, neste trabalho, ter sido utilizado os dados referentes ao SIH, o pacote download DataSUS é capaz de realizar downloads automatizados de outros sistemas, tais como: Sistema de informação de mortalidade (SIM), Sistema de Informações de Nascidos Vivos (SINASC), Cadastro Nacional de Estabelecimentos de Saúde (CNES). Futuramente, será incluído o Sistema de Informações Ambulatoriais (SIA). Por último, o argumento vars define quais variáveis serão baixadas.

A função datasusProcess é responsável pelo pré-processamento dos dados baixados do DataSUS através da função datasusFetch(19). Essa função somente possui três argumentos:

- data : data.frame criado pelo função datasusFetch.
- sistema : Sistema de informações
- dadosMunRes: Dados do município de residência

O argumento data define o data.frame que recebeu o resultado da função datasusFetch sem nenhuma modificação. A função datasusProcess deve ser usada logo após a função datasusFetch. O argumento sistema detalha de qual fonte são os dados armazenados no data frame a ser processado. Por fim o argumento dadosMunRes acrescenta ao data.frame informações sobre o município de residência, como nome completo, latitude e longitude do município (19). Para que as funções operem de modo satisfatório é preciso estar conectado à internet e que o FTP do DATASUS esteja funcionando. Informações mais detalhadas podem ser obtidas no Wiki do pacote no endereço: <https://github.com/rfsaldanha/downloadDataSUS/wiki>. A seguir são detalhados os comandos utilizados para carregamento do pacote download DataSUS no ambiente de trabalho do R e para download dos dados de 2015 do SIH para o estado de Goiás.

```
library(downloadDataSUS) # carrega o pacote downloadDataSUS no
ambiente de trabalho do R.

internacoes<-datasusFetch(anoIni=2015,      mesIni      =      1,
anoFim=2015,mesFim = 12,uf="GO", sistema="SIH-RD")

str(internacoes)
```

Os comandos acima geraram um *data frame* com 361.213 interações para o ano de 2015 no estado de Goiás. Ao todo foram reunidas 113 variáveis associadas a cada interação com a estrutura de dados exemplificada junto às seis primeiras variáveis geradas. O descritivo completo das variáveis pode ser obtida no trabalho de Santos (9).

```
>str(internacoes)

'data.frame': 361213 obs. of 113 variables:

 $ UF_ZI:Factor w/ 172 levels "520000","520013",...: 1 1 1 1 1 1 1 1 1
1 1

 $ ANO_CMPT: Factor w/ 1 level "2015": 1 1 1 1 1 1 1 1 1 1 ...

 $ MES_CMPT: Factor w/ 12 levels "01","02","03",...: 1 1 1 1 1 1 1 1 1
1 1

 $ ESPEC :Factor w/ 12 levels "01","02","03",...: 1 1 1 1 1 3 3 3 3 2

 $ CGC_HOSP :Factor w/ 106 levels "00029827000128",...: NA
NANANANANANANANANA

 $ CEP: Factor w/ 24220 levels "03145010","03728210",...: 6464 6447
6458 6388 6447 138 138 138 138 5911
```

Cabe destacar que as internações não permitem a identificação unívoca dos pacientes, sendo que um mesmo paciente pode ter sido responsável por múltiplas internações. Há soluções para o *linkage* probabilístico de internações como aquela realizada pelo OpenRecLink, entretanto a aplicação desse tipo de técnica foge ao escopo do presente trabalho. (21). A função *datasusProcess* não foi aplicada ao *data frame* gerado, pois a mesma só contém, por enquanto, rotinas para pré-processamento de dados dos SIM. Dentre as variáveis registradas na AIH está o CEP referente à residência do paciente admitido no estabelecimento de saúde. Essa informação é de cunho obrigatório para registro da AIH. Com a utilização de outro pacote R é possível geolocalizar cada AIH com base no CEP fornecido pelo paciente.

A base de dados do código de endereçamento postal brasileiro é monopolizada pela Empresa Brasileira de Correios e Telégrafos. Assim a utilização da mesma somente é feita mediante de pagamento de licença. Ainda assim, a versão paga, não contém os parâmetros de latitude e longitude associados a todos os CEP. Para contornar essas dificuldades foi desenvolvido de modo colaborativo o projeto: CEP Aberto(22). Esse projeto visa prover o acesso gratuito, bem como construir de maneira colaborativa uma base de dados com os Códigos de Endereçamento Postal (CEP) geolocalizados de todo o Brasil. Atualmente, encontram-se registrados 980.955 CEPs no Brasil. Além disso, esse projeto elaborou uma Application Programming Interface (API), de uso gratuito, para permitir pesquisar, a partir de um CEP, dados como: estado de localização, município, código telefônico de área, bairro, logradouro, latitude, longitude e altitude. Para utilizar a API é necessário fazer um cadastro no endereço: <http://cepaberto.com/>. Uma vez realizado o cadastro é fornecido um token de acesso. Como o projeto é mantido de forma colaborativa a API possui limitações de volume de consulta, somente sendo possível realizar uma consulta a cada 3 segundos e no máximo 10 mil consultas por token/dia. Considerando as funcionalidades do projeto CEP Aberto, Robert Myles desenvolveu um pacote R, denominado *CepR*, que consulta os dados de CEP junto ao projeto CEP Aberto, diretamente do R

(23). A integração da solução do CepR e do download DataSUS feita por nós possibilitou criar uma rotina de download, pré-processamento e geolocalização de AIHs.

A rotina de integração entre os dois segue a um conjunto de passos para otimizar o processo de consulta. Primeiramente, como análises voltadas para as AIHs geralmente apresentam um número elevado de consultas é preciso sintetizar os CEPs obtidos junto ao *data frame* obtido junto à função `datasusFetch` para minimizar o número de consultas na API do CEP Aberto, diminuindo o tempo do processo de geolocalização. Os exemplos de comandos a seguir, exemplificam como fazê-lo a partir de um *data frame* não modificado logo após a aplicação da função: `datasusFetch`. Utilizamos o mesmo *data frame* de interações definido anteriormente.

```
install.packages("cepR") # instala o pacote CepR
```

Para a realização das consultas é necessário iterar sobre o vetor de CEPs (`cep_unicos`). Para a realização dessa operação é necessário a configuração de um loop que irá consultar, para cada entrada de CEP no vetor `cep_unicos`, quais são as respectivas informações de localização geográfica. Os passos a seguir definem como criar um *data frame* que irá receber os dados de cada consulta de CEP realizada, bem como parametrizam o loop de consulta iterativa.

```
#cria um data frame temporário que será utilizado para ser populado
durante as iterações relativas a cada CEP.
geo_coded<- data.frame(estado=character(),cidade=character
bairro=character(),cep=character(),logradouro=character(),latitude=c
haracter(),longitude=character(),altitude=character(),ddd=character(
),cod_IBGE=character(),quality      =      logical(),cep_buscado      =
character(),stringsAsFactors=FALSE)

cep_unico<- as.character(unique(internacoes$CEP)) # cria um vetor
com os CEPs sem repetição para minimizar a realização das consultas
na API do CEP Aberto
#Loop de consulta de cadaceps listado no veotr: cep_unico
for (i in 1:10000) {
  sys1 <- Sys.time()
  consulta<-busca_cep(cep=(cep_unico[[i]]), token= 'seu token')
  consulta$quality<- anyNA(c(consulta$latitude,consulta$longitude))
  consulta$cep_buscado<- cep_unicos[[i]]
  geo_coded<- rbind(geo_coded, consulta, make.row.names=FALSE)
  if(Sys.time()-sys1 <=4.0) Sys.sleep(4.0-(Sys.time()-sys1))}
```

O loop em questão já considera o intervalo mínimo de tempo a ser respeitado para a realização de consultas na API do CEP Aberto. Assim apenas é preciso modificar os parâmetros iniciais em negrito da seção: (**`i in 1:10000`**) para que sejam fornecidos os índices de no máximo 10 mil CEPs por dia. Além disso, é preciso substituir esse trecho - **`seu token`** - pelo token fornecido ao realizar o cadastro no site do projeto CEP Aberto. O loop para consulta guarda um registro do resultado da

consulta realizada, para fins de rastreamento da qualidade da geolocalização – coluna quality. Uma vez terminado o processo de consulta é gerado o *data frame* *geo_coded* com a estrutura a seguir.

```
str(geo_coded)
Classes 'tbl_df', 'tbl' and 'data.frame': 24220 obs. of 12
variables:
 $ estado : chr "GO" "GO" "GO" "GO"
 $ cidade :chr "São Simão" "Mineiros" "Quirinópolis" "Jataí"
 $ bairro :chr NA NANA "Centro"
 $ cep : chr "75890000" "75830000" "75860000" "75800970"
 $ logradouro:chr "São Simão" "Mineiros" "Quirinópolis" "AC Jataí,
Avenida Dorival de Carvalho, 1007"
 $ latitude :chr "-18.9964906" "-17.5624415" "-18.4476442" "-
17.8851316"
 $ longitude :chr "-50.547432" "-52.5489206" "-50.4551598" "-
51.7153316"
 $ altitude :chr "478.400000" "787.900000" "512.900000"
"704.000000"
 $ ddd : chr NA NANA "64" ...
 $ cod_IBGE : chr "5220405" "5213103" "5218508" "5211909"
 $ quality : logi TRUE TRUETRUETRUETRUE ...
 $ cep_buscado: chr"75890000" "75830000" "75860000" "75800970"
```

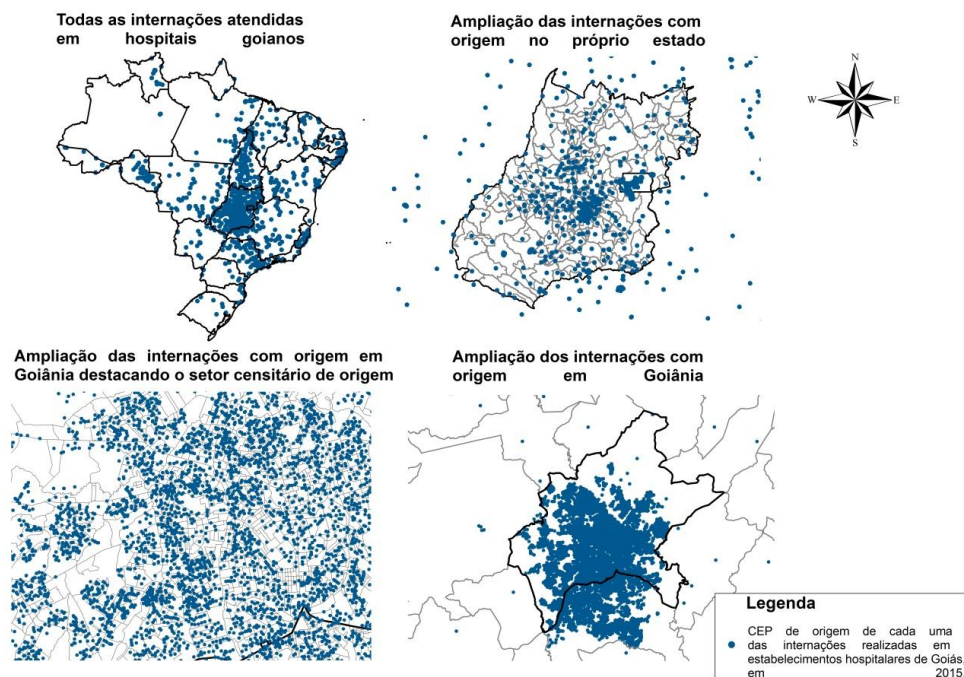
Os passos detalhados foram aplicados junto aos dados de AIH referentes a 2015 para o estado de Goiás. Os resultados obtidos no processo de espacialização e AIHs foram plotados com o auxílio do software ARCMAP(24) e uma análise de densidade de kernel foi realizada para exemplificar a potencialidade do método.

RESULTADOS

Dos 24.220 CEPs diferentes, 23.910 (98,72%) foram geolocalizados, correspondendo a 295.670 de 302.992 internações possíveis (97,58%). Não houve nenhuma ocorrência de CEP extraído do SIH com número de dígitos inferior a 8. Os CEPs que não foram geolocalizados correspondem, em sua maioria, a CEPs genéricos de cidades e não propriamente de ruas. Quanto à espacialização dos pontos referentes às AIHs pode-se notar que pacientes originários de todos os estados brasileiros foram admitidos em algum hospital goiano ao longo do ano de 2015 - FIGURA 1. Os pontos indicam o local de residência do paciente admitido em um estabelecimento de saúde goiano.

FIGURA 1

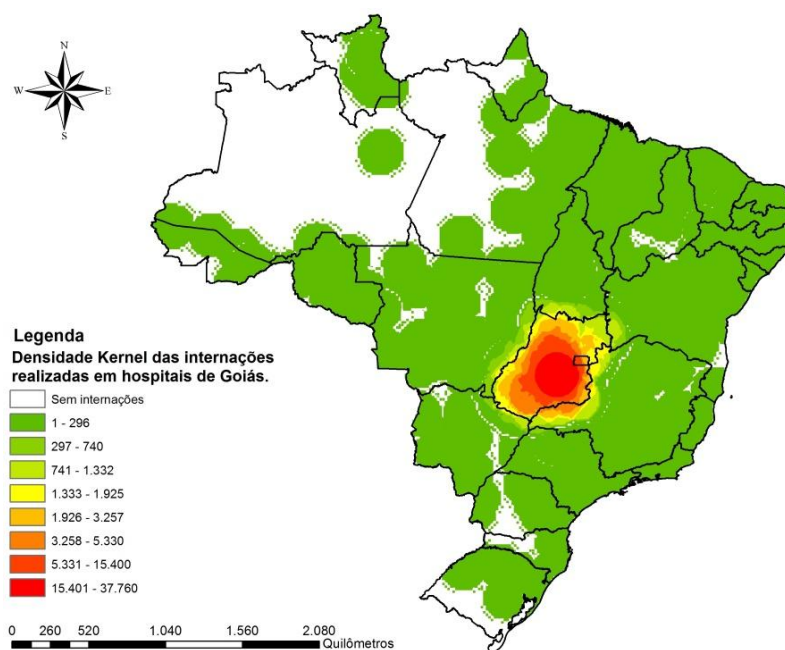
Distribuição de AIHs autorizadas em estabelecimentos goianos, por residência do paciente, 2015



A análise de densidade de incidência de pontos de Kernel (FIGURA 2) demonstrou que, apesar da amplitude de origem, a maior parte das AIH autorizadas em Goiás são de moradores do próprio estado, com importantes vazamentos para o Triângulo Mineiro.

FIGURA 2

Densidade de incidência de pontos das AIHs realizadas no estado de Goiás, 2015



DISCUSSÃO

Os desafios associados à melhor aplicação de métodos da geografia da saúde, para fins de solução de problemas metodológicos associados a estudos avaliativos se configura como uma agenda ainda inconclusa. Nesse sentido, o presente trabalho almejou fornecer uma pequena contribuição para ampliação das possibilidades analíticas vinculadas ao SIH.

Graças ao delineamento aqui proposto abre-se a possibilidade para melhor investigar uma miríade de questões. Análises sobre determinantes sociais de saúde poderão ser melhor embasadas, uma vez que a geolocalização das AIHs permite a análise concomitante de perfis de incidência de agravos e outros elementos sociais como: perfil de população residente, nível socioeconômico, grau de urbanização, avaliação de serviços de saúde, implementação de reformas sanitárias, distribuição de recursos humanos em saúde, efeitos do Programa de Melhoria do Acesso e da Qualidade (PMAQ), intervenções verticais como programas de vacinação, Programa Mais Médicos, investimento em ações preventivas, estabilidade política do município. Quanto a variáveis ambientais, é possível confrontar a proximidade com poluentes, arrumamento, cursos fluviais e a ocorrência de doenças associadas.

A análise das redes de cuidado também pode ser potencializada com o método em tela. O registro do código do Cadastro Nacional de Estabelecimento de Saúde que atendeu a AIH, comparado com a localidade de origem do paciente, pode fornecer um insight poderoso sobre os fluxos reais de pacientes em busca de cuidado. Fluxos regulares e constantes de pacientes com determinadas condições pode fornecer a profissionais de saúde, gestores e formuladores de políticas insights sobre regiões que vem enfrentando carência de serviços e por sua vez, insights sobre regiões que podem desenvolver ações em rede para diminuir as distancias viajadas por pacientes.

Análises de acesso a serviços e equidade podem ser mais bem detalhadas e áreas de abrangência a serviços podem ser definidas com base em técnicas de geoprocessamento. Assim, a qualidade do cuidado prestado por hospitais, unidades básicas de saúde e unidades de pronto atendimento podem ser avaliadas com base em parâmetros de proximidade geográfica e áreas de influência potencial. Técnicas de linkage probabilístico de dados podem ser otimizados com a inclusão de parâmetros de localização geográfica aumentando seu grau de confiabilidade. Cargas de internações podem ser vinculadas a serviços de saúde específicos, como no caso da atenção primária. O planejamento da oferta de cuidado pode ser feito de forma mais detalhada, uma vez que é possível analisar o perfil epidemiológico de uma determinada população adscrita a uma região geográfica.

A disponibilização de dados mais granulares abre um leque de opções para gestores, profissionais de saúde, formuladores de políticas públicas e acadêmicos. A articulação de métodos interdisciplinares contribui para o desenho metodológico de uma solução capaz de subsidiar decisões de gestão com base em evidências.

Apesar dos avanços e da potencialidade do método desenvolvido pesam algumas limitações a serem abordadas. A primeira delas remete à presença de algum grau de imprecisão na localização da origem do paciente, em função do uso do CEP. Como a geolocalização é feita com base no código do arruamento, a mesma é marcada por algum grau de imprecisão que pode afetar análises mais sensíveis à imprecisão, como, por exemplo, ações de vigilância epidemiológica. Além disso, em função das

limitações do volume de consultas da API o processo desenvolvido aqui é de realização lenta. Assim, grandes volumes de dados podem consumir um tempo substancial para serem geolocalizados. Por último, cabe destacar a dependência de uma solução colaborativa sem financiamento regular. O projeto CEP Aberto é fundamental para permitir a operacionalização da solução aqui proposta. Sem o seu funcionamento esse método perde sua capacidade propositiva.

A despeito das limitações apresentadas a solução aqui detalhada é marcada por mais potencialidades que limitações. O objetivo ao discutir os detalhes técnicos foi o de permitir a instrumentalização de interessados para a ampliar o leque de opções analíticas associadas aoSIH.

AGRADECIMENTOS

Agradecemos a Daniel Petruzalek, Daniel Saldanha e a Robert Myles pela doação de tempo, esforço e dedicação para o desenvolvimento de soluções voltadas para a promoção de melhorias no SUS.

CONTRIBUIÇÃO DOS AUTORES

Rocha TAH e participou da concepção, programação, elaboração do manuscrito, revisão bibliográfica, análise de dados, discussão dos elementos textuais e revisão crítica. Silva NC, Amaral, PVM, Vissoci, JRN, participaram da elaboração do manuscrito, revisão bibliográfica, análise de dados, discussão dos elementos textuais e revisão crítica. Thomaz EBA, Queiroz RCS, Harris M, Millet C, Silva AAM, Hone T e Facchini LA, participaram da discussão dos elementos textuais e revisão crítica. Todos os autores aprovaram a versão final do manuscrito e declaram serem responsáveis por todos os aspectos do trabalho, garantindo sua precisão e integridade.

REFERÊNCIAS

- Brasil. Departamento de Informática do SUS - DATASUS. Informações de Saúde (TABNET) [Internet]. 2015. Available from: <http://www2.datasus.gov.br/DATASUS/index.php?area=02>
- Brasil. Ministério da Saúde. Departamento Nacional de Auditoria do SUS. Coordenação-Geral de Desenvolvimento N e CTC de N e CT. Auditoria no SUS: noções básicas sobre sistemas de informação TT - Auditorship in SUS (BR): basic notions about systems of information [Internet]. A. Normas e Manuais Técnicos. 2004. p. 93. Available from: http://bvsms.saude.gov.br/bvs/publicacoes/auditoria_sus.pdf
- Gerhardt TE, Pinto JM, Riquinho DL, Roese A, Santos DL dos, Lima MCR de. Utilização de serviços de saúde de atenção básica em municípios da metade sul do Rio Grande do Sul: análise baseada em sistemas de informação. *Cien Saude Colet* [Internet]. 2011;16(suppl 1):1221–32. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1413-81232011000700054&lng=pt&tlng=pt
- Bittencourt SA, Camacho LAB, Leal M do C. O Sistema de Informação Hospitalar e sua aplicação na saúde coletiva Hospital. *Cad Saude Publica* [Internet]. 2006;22(1):19–30. Available from: <http://www.scielo.br/pdf/csp/v22n1/03.pdf>
- Loyola Filho AI De, Leite Matos D, Giatti L, Afradique ME, Viana Peixoto S, Lima-Costa MF. Causas de internações hospitalares entre idosos brasileiros no âmbito do Sistema Único de Saúde. *Epidemiol e Serviços Saúde* [Internet]. 2004;13(4):229–38. Available from: scielo.iec.pa.gov.br/scielo.php?script=sci_arttext&pid=S1679-49742004000400005&lng=en&nrm=iso&tlng=en
- Escosteguy CC, Portela MC, Medronho Rde A, de Vasconcellos MT. The Brazilian Hospital Information System and the acute myocardial infarction hospital care. *Rev Saude Publica*. 2002;36(4):491–9.
- Schramm JM, Szwarcwald CL. Sistema hospitalar como fonte de informações para estimar a mortalidade neonatal e a natimortalidade. *Rev Saude Publica*. 2000;34(3):272–9.
- Silva NP. A utilização dos programas TABWIN e TABNET como ferramentas de apoio a disseminação das informações em saúde [Internet]. FIOCRUZ; 2009. Available from: <http://www.arca.fiocruz.br/handle/icict/2300>
- Santos AC dos. Sistema de informações hospitalares do Sistema Único de Saúde: documentação do sistema para auxiliar o uso das suas informações [Internet]. FIOCRUZ; 2009. Available from: https://www.arca.fiocruz.br/bitstream/icict/2372/1/ENSP_Dissertação_Santos_Andréia_Cristina.pdf
- Fonseca B, Silva K. Atribuição de IDH aos bairros de Belo Horizonte [Internet]. *Revista Transite*. 2017 [cited 2017 Sep 5]. Available from: <http://transite.fafich.ufmg.br/idh-bairros-de-belo-horizonte/>
- Roux AVD. A glossary for multilevel analysis. *J Epidemiol Community Health*. 2002;56:588–94.

- Guimarães RB. Geografia e saúde coletiva no Brasil. *Saúde e Soc* [Internet]. 2016;25(4):869–79. Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0104-12902016000400869&lng=pt&tlng=pt
- Kearns R, Moon G. From medical to health geography: novelty, place and theory after a decade of change. *Prog Hum Geogr*. 2002;26(5):605–25.
- Macintyre S, Ellaway A, Cummins S. Place effects on health: How can we conceptualise, operationalise and measure them? In: *Social Science and Medicine*. 2002. p. 125–39.
- Dummer TJB. Health geography: supporting public health policy and planning. *CMAJ* [Internet]. 2008 Apr 22 [cited 2017 Sep 5];178(9):1177–80. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18427094>
- Lima DVM. Research design: a contribution to the author. *Online Brazilian J Nurs*. 2011;10(2):1–18.
- Nedel FB. csapAIH: uma função para a classificação das condições sensíveis à atenção primária no programa estatístico R*. *Epidemiol e Serviços Saúde* [Internet]. 2017 Jan [cited 2017 Sep 5];26(1):199–209. Available from: http://revista.iec.gov.br/template_doi_ess.php?doi=10.5123/S1679-49742017000100199&scielo=S2237-96222017000100199
- Pipinellis A. *GitHub Essentials*. Packt Publishing Ltd. 2015. 190 p.
- Saldanha R. downloadDataSUS [Internet]. Github; 2017. Available from: <https://github.com/rfsaldanha/downloadDataSUS>
- Petruzalek D. READ.dbc - Um pacote para importação de dados do DATASUS na linguagem R. Goiânia; 2016.
- Camargo Jr KR de, Coeli CM. Going open source: some lessons learned from the development of OpenRecLink. *Cad Saude Publica*. 2015;31(2):257–63.
- CEP Aberto. CEP aberto [Internet]. 2017 [cited 2017 Sep 5]. Available from: <http://cepaberto.com/>
- McDonnell RM. CepR [Internet]. Github; 2017. Available from: <https://github.com/RobertMyles/cepR>
- ESRI. ArcGIS Desktop: Release 10.3. Redlands CA. 2014. p. Environmental Systems Research Institute.