

Utilização do Google Maps para o georreferenciamento de dados do Sistema de Informações sobre Mortalidade no município do Rio de Janeiro, 2010-2012*

doi: 10.5123/S1679-49742017000400018

Use of Google Maps for geocoding data from the Mortality Information System in Rio de Janeiro municipality, Brazil, 2010-2012

Uso de Google Maps para georreferenciación de datos del Sistema de Informaciones sobre Mortalidad en el municipio de Río de Janeiro, Brasil, 2010-2012

Ismael Henrique da Silveira¹
Beatriz Fátima Alves de Oliveira²
Washington Leite Junger¹

¹Universidade do Estado do Rio de Janeiro, Instituto de Medicina Social, Rio de Janeiro-RJ, Brasil

²Fundação Instituto Oswaldo Cruz, Escola Nacional de Saúde Pública Sergio Arouca, Rio de Janeiro-RJ, Brasil

Resumo

Objetivo: descrever os resultados da aplicação de um procedimento de baixo custo, baseado em *software* livre, para o georreferenciamento de dados do Sistema de Informações sobre Mortalidade (SIM) no município do Rio de Janeiro, Brasil. **Métodos:** estudo descritivo, com uso da base de endereços do Google Maps para georreferenciamento de dados de óbitos registrados no SIM, ocorridos no município entre 2010 e 2012; o trabalho foi realizado em três etapas, (i) padronização dos endereços, (ii) georreferenciamento pelo Google e (iii) intervenção manual. **Resultados:** do total de 26.081 endereços submetidos ao procedimento, 18.646 (71,5%) foram automaticamente georreferenciados; os 7.435 remanescentes foram submetidos a intervenção manual, encontrando-se 5.250; 70,6% dos endereços não foram localizados inicialmente; afinal, 91,6% dos endereços foram georreferenciados. **Conclusão:** o procedimento apresentou alta proporção de acertos automáticos e, apesar de demandar maior tempo, a intervenção manual reduziu consideravelmente as perdas.

Palavras-chave: Mapeamento Geográfico; Sistemas de Informação Geográfica; Análise Espacial; Epidemiologia Descritiva.

*Manuscrito originado da dissertação de Mestrado de Ismael Henrique da Silveira, intitulada 'Espaços verdes e mortalidade por doenças cardiovasculares no município do Rio de Janeiro', defendida junto ao programa de Pós-Graduação do Departamento de Epidemiologia do Instituto de Medicina Social da Universidade do Estado do Rio de Janeiro em 2015.

Endereço para correspondência:

Ismael Henrique da Silveira – Rua São Francisco Xavier, nº 524, sala 7013D, Maracanã, Rio de Janeiro-RJ. CEP: 20550-013
E-mail: ismaelhsilveira@gmail.com

Introdução

No Brasil, existem Sistemas de Informações em Saúde (SIS) consolidados, a exemplo do Sistema de Informações sobre Mortalidade (SIM) e do Sistema de Informação de Agravos de Notificação (Sinan),¹ que disponibilizam dados para análises de situação de saúde. Contudo, o georreferenciamento desses dados em cidades brasileiras ainda é um desafio. Entre os problemas mais comuns destacam-se a qualidade dos dados, a falta de sistemas de informações geográficas (SIG) municipais estruturados, a falta de bases cartográficas digitais e cadastros oficiais de endereços, além da complexidade da infraestrutura urbana, principalmente em áreas carentes como ocupações irregulares e favelas, onde é comum a não padronização dos endereços.²

Os SIG são usados para descrever, analisar e prever padrões espaciais, dispondo de diversas aplicações em epidemiologia, como o mapeamento de doenças, a investigação de surtos e análises espaciais.³ Seu uso depende, sobretudo, da disponibilidade de dados georreferenciados. O georreferenciamento, por sua vez, é o processo pelo qual informações textuais descritivas de uma localidade, como o endereço relacionado a um evento de saúde, são convertidas em representações geográficas válidas. Ele pode ser realizado por meio da associação a um par de coordenadas geográficas ou a unidades espaciais, como setores censitários, bairros, municípios, entre outras.^{4,5}

Os SIG são usados para descrever, analisar e prever padrões espaciais, dispondo de diversas aplicações em epidemiologia, como o mapeamento de doenças, a investigação de surtos e análises espaciais. Seu uso depende, sobretudo, da disponibilidade de dados georreferenciados.

As ferramentas de georreferenciamento, concomitantemente aos avanços da informática, têm apresentado melhorias em termos de qualidade, acessibilidade e redução de custos. A escolha do método depende da disponibilidade de informações e bases cartográficas, da precisão desejada, do volume de dados e dos recursos disponíveis.^{5,6}

O objetivo deste trabalho foi descrever os resultados da aplicação de um procedimento de baixo custo, ba-

seado em *software* livre, para o georreferenciamento de dados do Sistema de Informações sobre Mortalidade – SIM – no município do Rio de Janeiro, Brasil

Métodos

Estudo descritivo de um método de georreferenciamento de dados sobre óbitos registrados no SIM, ocorridos no município do Rio de Janeiro entre 2010 e 2012.

Em 2010, a população do município era de 6.320.446 habitantes, dos quais 2,62% vivendo em extrema pobreza. No mesmo ano, o índice de desenvolvimento humano municipal era de 0,799. O produto interno bruto *per capita* era de 46.461,82 reais em 2014.⁷ A cobertura da Atenção Básica no município era de 52% em fevereiro de 2017.⁸

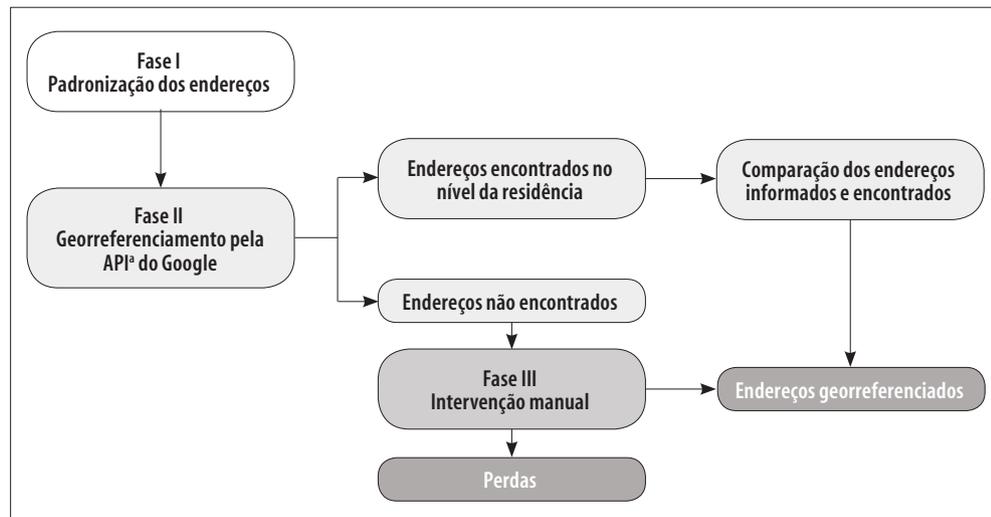
Os dados de óbitos ocorridos no período, por doenças isquêmicas do coração (DIC) e cerebrovasculares (DCBV), correspondentes aos códigos I20 a I25 e I60 a I69 da Décima Revisão da Classificação Internacional de Doenças (CID-10), foram obtidos do SIM da Secretaria Municipal de Saúde do Rio de Janeiro.

O trabalho foi realizado em três etapas: (I) padronização dos endereços, (II) georreferenciamento por meio do Google Maps e (III) intervenção manual (Figura 1).

A padronização dos endereços, programada em R,⁹ buscou corrigir erros frequentes, remover caracteres estranhos e padronizar componentes relacionados aos tipos (avenida, praça, rua) e aos títulos (Presidente, Professora, Princesa) dos logradouros. As substituições realizadas constam em material suplementar. Embora essa etapa não tenha eliminado todos os erros, ela foi útil para aumentar a quantidade de endereços georreferenciados automaticamente.

O georreferenciamento foi realizado por meio da Interface de Programação de Aplicativos (API) do Google, acessada com um *script* programado em R. A API de georreferenciamento compara os endereços informados com a base do Google Maps, para a captura das coordenadas geográficas. A utilização gratuita permite a requisição diária de até 2.500 pares de coordenadas.¹⁰ A API também retorna o endereço localizado e sua precisão, como residência, edificação específica (condomínio, parque, aeroporto), logradouro, bairro, cidade etc.

Para verificar a qualidade do georreferenciamento, os endereços encontrados pelo Google receberam o mesmo tratamento dos informados, sendo comparados entre si. Quando não correspondentes, foram transferidos



a) API: Application Programming Interface (Interface de Programação de Aplicativos).

Figura 1 – Diagrama representando as três fases do procedimento de georreferenciamento de dados do Sistema de Informações sobre Mortalidade no município do Rio de Janeiro, 2010-2012

para intervenção manual, juntamente com os demais endereços não localizados. Ademais, uma amostra de 100 endereços georreferenciados automaticamente foi analisada manualmente, para verificar se a coordenada correspondia ao endereço informado.

Na etapa manual, erros ortográficos remanescentes foram corrigidos e novamente submetidos ao georreferenciamento automático. Casos mais complexos exigiram pesquisas manuais, por meio do Google Street View, que permite a visualização do logradouro, e da base de logradouros do município, obtida do Instituto Pereira Passos. Endereços sem número foram georreferenciados nos pontos médios dos logradouros, quando contidos em um setor censitário, devido ao uso desses dados em um estudo epidemiológico. Quando não contido, o endereço foi considerado perda. A verificação baseou-se na sobreposição das bases de logradouros e de setores censitários, obtida da Fundação Instituto Brasileiro de Geografia e Estatística (IBGE).

As perdas foram analisadas pelo mapa de distribuição do percentual por bairro, dado pelo quociente entre o número de perdas e de óbitos (Figura 2). Foram utilizados os *softwares* R 3.3.2 e QGIS 2.14.

O estudo foi aprovado pelos Comitês de Ética em Pesquisa do Instituto de Medicina Social da Universidade do Estado do Rio de Janeiro (UERJ) e da Secretaria Municipal de Saúde do Rio de Janeiro (Parecer nº 531.635, em maio de 2014).

Resultados

Dos 26.081 endereços submetidos ao procedimento, 18.646 (71,5%) foram georreferenciados automaticamente. A correspondência entre os endereços informados e os encontrados, na amostra analisada manualmente, foi de 100%. Os restantes, 7.435, foram encaminhados para intervenção manual, pela qual foram recuperados 5.250, 70,6% dos endereços não localizados inicialmente. Ao final, 91,6% foram georreferenciados. A etapa manual dispendeu o maior tempo, devido à análise individual de cada endereço. A Tabela 1 apresenta os principais resultados.

A Figura 2 mostra a distribuição espacial do percentual de perdas por bairro. A maior proporção correspondeu à Rocinha, maior favela do país. Os bairros que tiveram perdas acima de 15% foram: Barra de Guaratiba; Paquetá; Jacarezinho; Vidigal; Alto da Boa Vista; Pitangueiras; Itanhangá; Complexo do Alemão; Galeão; Parada de Lucas; Gamboa; Manguinhos; Mangueira; Maré; Jacaré; Tauá; Caju; Jacarepaguá; Acari; Inhaúma; Vargem Pequena; Saúde; Santa Cruz; Barros Filho; Curicica; Sepetiba; e Costa Barros.

Discussão

O georreferenciamento de dados do SIM no Rio de Janeiro, utilizando a API do Google, teve alta proporção de acertos, apesar das dificuldades relacionadas à

Tabela 1 – Número de endereços georreferenciados e tempo gasto em cada etapa do georreferenciamento de dados do Sistema de Informações sobre Mortalidade no município do Rio de Janeiro, 2010-2012

Etapas do georreferenciamento	Endereços submetidos	Endereços georreferenciados	Tempo gasto
Etapa I – Padronização dos endereços	26.081	–	<1 dia
Etapa II – Georreferenciamento pela API ^a do Google	26.081	18.646	1-2 semanas
Etapa III – Intervenção manual	7.435	5.250	Aproximadamente 4 meses

a) API: Application Programming Interface (Interface de Programação de Aplicativos).

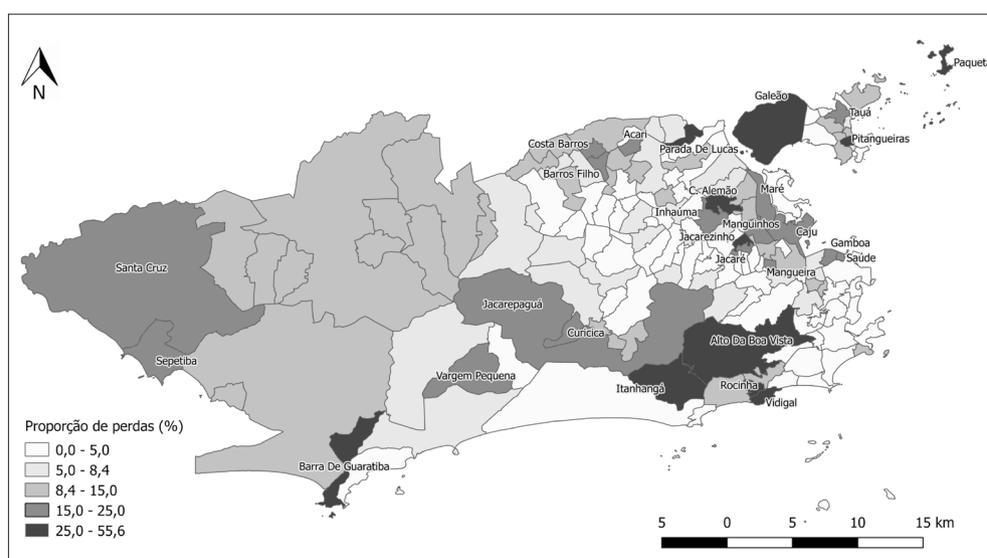


Figura 2 – Distribuição do percentual de perdas segundo bairros, no georreferenciamento de dados do Sistema de Informações sobre Mortalidade no município do Rio de Janeiro, 2010-2012

qualidade dos dados e às condições da infraestrutura urbana, marcada por inúmeras ocupações irregulares. A etapa manual foi importante por aproveitar 70% dos dados não georreferenciados automaticamente, reduzindo as perdas a 8,4%. As perdas foram maiores em bairros menos urbanizados, com assentamentos recentes e/ou ocupações irregulares, e de menor nível socioeconômico.

Foram identificados erros de inúmeras naturezas, possivelmente devidos ao informante, ao responsável pelo registro ou digitação do banco, e à base do Google. Houve erros ortográficos, no tipo e/ou título do logradouro, grafias diferentes, nomes diferentes com alguma semelhança fonética, ou mesmo o fato de o logradouro ser reconhecido por um nome diferente do cadastro oficial. A confusão entre bairros foi frequente, especialmente quando vizinhos. Outro problema recorrente foi a falta de preenchimento dos campos de endereço, principalmente do número. Endereços pouco informativos (Rua Um,

Rua Projetada etc.), frequentemente, eram perdidos por haver mais de um logradouro com o mesmo nome no bairro. Em alguns desses casos, o CEP, quando presente, facilitava a localização.

Programas de SIG e serviços comerciais, geralmente, são capazes de georreferenciar de 70 a 80% de endereços.¹¹ Em 2003, experiências em municípios brasileiros como Belo Horizonte e Goiânia, onde há SIG estruturados para georreferenciar endereços dos SIS, apresentaram desempenho acima de 90%; em Porto Alegre e no Rio de Janeiro, utilizando-se de técnicas de interpolação em logradouros, esse desempenho variou entre 60 e 90%.² Em um trabalho de georreferenciamento por interpolação em logradouros, sobre agravos notificados pelo Sinan em Campinas, Macapá e Rio de Janeiro, datado de 2004, os acertos automáticos foram, respectivamente, de 49, 72 e 48%, e a busca manual acrescentou 34, 20 e 17% dos totais de endereços.¹² Outro trabalho, realizado no Rio de Janeiro em 2014, envolvendo

dados de tuberculose do Sinan, comparou técnicas baseadas na API do Google e na interpolação com programa comercial: a primeira apresentou melhor desempenho, atingindo 69% – contra 64% de acertos do programa comercial.¹³

Conforme ressaltado por Magalhães *et al.*,¹³ diante da realidade cartográfica do país, em algumas localidades a utilização da API do Google pode ser a única opção para o georreferenciamento. Além disso, a base do Google Maps pode estar mais atualizada que muitas bases cartográficas oficiais de municípios, embora não atenda a todos com igual precisão.

O georreferenciamento pode conter erros, em função do mapeamento ou da acurácia do Google. Erros no mapeamento tendem a ser pouco frequentes entre os dados georreferenciados automaticamente, visto que são locais com melhor endereçamento. A acurácia, por sua vez, diz respeito à proximidade da coordenada obtida – por meio do Google – do valor real. Esses erros têm sido cada vez menores, devido aos esforços do Google no mapeamento de cidades e à crescente utilização das API para inúmeras finalidades de geolocalização.

Referências

1. Ministério da Saúde (BR). Biblioteca Virtual em Saúde. Vigilância em Saúde. Sistemas de Informação: informações estratégicas [Internet]. 2008 [citado 2016 jun 15]. Disponível em: http://bvsmis.saude.gov.br/bvsmis/svs/inf_sist_informacao.php.
2. Barcellos C, Ramalho WM, Gracie R, Magalhães MAFM, Fontes MP, Skaba DA. Georreferenciamento de dados de saúde na escala submunicipal: algumas experiências no Brasil. *Epidemiol Serv Saúde*. 2008 mar;17(1):59-70.
3. Kirby RS, Delmelle E, Eberth JM. Advances in spatial epidemiology and geographic information systems. *Ann Epidemiol* [Internet]. 2017 Jan [cited 2017 Jan 10];27(1):1-9. Available from: [http://www.annalsofepidemiology.org/article/S1047-2797\(16\)30495-1/fulltext](http://www.annalsofepidemiology.org/article/S1047-2797(16)30495-1/fulltext)
4. Goldberg DW. A geocoding best practices guide [Internet]. Springfield: North American Association of Central Cancer Registries; 2008 [cited 2017 Ago 28]. 287 p. Available from: https://20tqt36s1la18rvn82wcmppn-wpengine.netdna-ssl.com/wp-content/uploads/2016/11/Geocoding_Best_Practices.pdf
5. Goldberg DW, Wilson JP, Knoblock CA. From text to geographic coordinates: the current state of geocoding. *J Urban Reg Inf Syst Assoc*. 2007 Jan;19(1):33-46.
6. Rushton G, Armstrong MP, Gittler J, Greene BR, Pavlik CE, West MM, et al. Geocoding in cancer research: a review. *Am J Prev Med*. 2006 Feb;30(2 Suppl):16-24.
7. Instituto Brasileiro de Geografia e Estatística. Cidades@ [Internet]. 2014 [citado 2017 mar 20]. Disponível em: <http://cidades.ibge.gov.br/xtras/home.php?lang=>
8. Ministério da Saúde (BR). Departamento de Atenção Básica. Nota técnica: sistema de nota técnica do DAB [Internet]. 2017 [citado 2017 mar 20]. Disponível em: <http://dab2.saude.gov.br/sistemas/notatecnica/frmlistaMunic.php>.
9. R Core Team. R: a language and environment for statistical computing [Internet]. Viena: R Foundation for Statistical Computing; 2017 [cited 2017 Mar 20]. Available from: <https://www.r-project.org/>
10. Google. Google maps geocoding API [Internet]. 2016 [citado 2017 mar 24]. Disponível em: <https://developers.google.com/maps/documentation/geocoding/>

Contribuição dos autores

Silveira IH foi responsável pela concepção do estudo, análise e interpretação dos resultados e redação do manuscrito. Oliveira BFA colaborou na análise dos resultados e revisão do manuscrito. Junger WL colaborou nas etapas de programação e revisão do manuscrito. Todos os autores aprovaram a versão final do manuscrito, sendo responsáveis por todos os aspectos do trabalho e garantindo sua precisão e integridade.

11. Sonderman JS, Mumma MT, Cohen SS, Cope EL, Blot WJ, Signorello LB. A multi-stage approach to maximizing geocoding success in a large population-based cohort study through automated and interactive processes. *Geospat Health*. 2012 May;6(2):273-84.
12. Skaba DA, Carvalho MS, Barcellos C, Martins PC, Terron SL. Geoprocessamento dos dados da saúde: o tratamento dos endereços. *Cad Saúde Pública*. 2004 nov-dez; 20(6):1753-6.
13. Magalhães MAFM, Matos VP, Medronho RA. Avaliação do dado sobre endereço no Sistema de Informação de Agravos de Notificação utilizando georreferenciamento em nível local de casos de tuberculose por dois métodos no município do Rio de Janeiro. *Cad Saúde Coletiva*. 2014 abr-jun;22(2):192-9.

Abstract

Objective: to describe the results of the application of a low cost procedure, using free software, for geocoding data from the Mortality Information System (SIM), in the municipality of Rio de Janeiro. **Methods:** descriptive study using Google Maps database for geocoding deaths data recorded at SIM, occurred from 2010 to 2012, in Rio de Janeiro; the study was carried out in three stages, (i) standardization of addresses, (ii) geocoding by Google Maps, and (iii) manual intervention. **Results:** from the total of 26,081 addresses submitted to the procedure, 18,646 (71.5%) had exact matches; the remaining 7,435 were submitted to manual intervention, which found 5,250; 70.6% of the addresses were not initially found; a total of 91.6% of the addresses were geocoded. **Conclusion:** the procedure presented high proportion of automatic matches and, although it demanded much time, manual intervention allowed a considerable reduction of losses.

Keywords: Geographic Mapping; Geographic Information Systems; Spatial Analysis; Epidemiology, Descriptive.

Resumen

Objetivo: describir los resultados de la aplicación de un procedimiento de bajo costo basado en software libre para la georreferenciación de datos del Sistema de Informaciones sobre Mortalidad (SIM), en el municipio de Río de Janeiro. **Métodos:** estudio descriptivo utilizando la base de direcciones de Google Maps para la georreferenciación de datos de óbitos registrados en el SIM, ocurridos del 2010 al 2012, en el municipio de Río de Janeiro; el trabajo se llevó a cabo en tres etapas, (i) estandarización de las direcciones, (ii) georreferenciación por Google e (iii) intervención manual. **Resultados:** de 26.081 direcciones sometidas al procedimiento, 18.646 (71,5%) fueron automáticamente georreferenciadas; el restante, 7.435, se enviaron a intervención manual, donde fueron encontrados 5.250; al final, se georreferencio 91,6%. **Conclusión:** el procedimiento presentó una alta proporción de aciertos automáticos y, a pesar de demandar más tiempo, la intervención manual permitió una considerable reducción de las pérdidas.

Palabras-clave: Mapeo Geográfico; Sistemas de Información Geográfica; Análisis Espacial; Epidemiología Descriptiva.

Recebido em 14/03/2017
Aprovado em 25/07/2017