

AVALIAÇÃO DA QUALIDADE DE UM BANCO DE DADOS PARA REALIZAÇÃO DE GEORREFERENCIAMENTO

LETYCIA BARROS GONÇALVES¹; ERIKA COLLISCHONN²; EDUARDO DICKIE DE CASTILHOS³

¹Universidade Federal de Pelotas – letyciabgoncalves@gmail.com

²Universidade Federal de Pelotas – ecollischonn@gmail.com

³Universidade Federal de Pelotas – eduardo.dickie@gmail.com

1. INTRODUÇÃO

O Código de Endereçamento Postal (CEP) inicialmente foi projetado para acelerar e reduzir o custo da entrega de correspondência, mas acabaram desempenhando uma função mais ampla, sendo um mecanismo de organização espacial necessário para projetos de georreferenciamento (ARANHA, 1997).

É considerado um sistema nominal, pois o código é atribuído a uma localização e também um código hierárquico, pois divide o território em regiões progressivamente menores. Com isso, a correspondência entre um número de CEP e uma localização não é de todo aleatória: segue uma sequência com lógica interna e hierarquia específica (ARANHA, 1997).

Todos os CEPs são compostos por oito números, sendo que os cinco primeiros formam um radical e são separados por um hífen do sufixo que contém os outros três números. Cada dígito tem uma correspondência, sendo que o primeiro dígito do CEP, tomado isoladamente, define Regiões Postais; os dois primeiros, em conjunto, definem Sub-Regiões; os três primeiros delimitam setores. Assim, o nível de detalhamento da descrição geográfica que se pode obter a partir do CEP aumenta pela sucessiva subdivisão das regiões maiores em regiões menores (BRASIL, 2017).

O georreferenciamento dos eventos de saúde tem sido utilizado em diversos estudos, sendo uma importante ferramenta para a análise e avaliação de riscos à saúde coletiva (BARCELLOS et al., 2008; MOREIRA, NICO, TOMITA, 2007; COLLISCHONN et al., 2018), e é o processo pelo qual informações textuais descritivas de uma localidade, como o endereço referente a um evento de saúde, são convertidas em representações geográficas válidas (SILVEIRA et al., 2017). Uma das tecnologias utilizadas para o georreferenciamento são os chamados Sistemas de Informações Geográficas (SIG), que são um conjunto de ferramentas utilizadas para o entendimento dos fatos e manipulação de informações que ocorrem no espaço geográfico (PINA, 2000; SKABA, et al., 2004).

Barcellos e colaboradores (2008), relataram experiências de georreferenciamento de dados de saúde em alguns municípios brasileiros, e analisaram a disponibilidade e atualização de bases cartográficas e o tratamento de dados de endereço nos sistemas de informações em saúde. Em função dos problemas encontrados, uma das ferramentas discutidas consiste na padronização dos endereços, que pode ser realizada durante a digitação ou após, em uma fase de análise crítica e consolidação dos dados.

Os dados em relação a endereço e CEP analisados para o presente estudo pertencem a um banco de dados fornecido para uma dissertação de mestrado. Sendo assim, o objetivo é avaliar a qualidade das informações do banco fornecido.

2. METODOLOGIA

Trata-se de um estudo metodológico descritivo desenvolvido a partir da análise de um banco de dados de uma pesquisa.

O banco de dados fornecido apresentava as informações com um número de identificação, endereço, bairro, CEP, cidade, Estado e referência sobre o endereço, porém sem as informações de identificação dos participantes: nome e nome da mãe. O banco estava consolidado sem duplicidades. O dado sobre bairro que se encontra no banco de dados refere-se ao bairro declarado pelo paciente no momento da entrevista; da mesma forma o CEP. Os dados foram fornecidos em planilha do Programa Microsoft Office Excel®. Para o estudo serão avaliados apenas os endereços pertencentes ao município de Pelotas.

Para obtenção da localização geográfica dos endereços adquiriu-se uma outra planilha do Programa Microsoft Office Excel® contendo, uma coluna com os códigos de endereçamento postal, uma segunda coluna com a latitude e uma terceira coluna com a longitude em datum WGS84.

Para o georreferenciamento a planilha dos endereços foi aberta utilizou-se uma fórmula PROCV para buscar da planilha com a localização geográfica, para cada registro, primeiramente, a latitude e depois a longitude dos CEP em correspondência.

Após a primeira rodada de georreferenciamento verificou-se uma série de inconsistências nos dados da planilha do banco como: sem CEP, CEP não se localizava no município de estudo ou o CEP por alguma razão não foi localizado. Desse processo resultaram duas planilhas, uma com as colunas com as coordenadas definidas, e outra que, por alguma inconsistência resultou erros nestas colunas.

A partir da planilha que constavam os erros, foi realizada a verificação de todos os CEP em relação ao endereço e localização dos participantes do estudo com o uso do aplicativo “Busca CEP” versão DNE 2007 do Sistema de Correios do Brasil, sendo que para os que não foram localizados foram utilizados outros aplicativos de localização como o *Google maps* e *Here*.

3. RESULTADOS E DISCUSSÃO

A análise em relação a planilha de dados fornecida mostrou que dos 4230 registros 530 foram excluídos devido a pertencerem a outros municípios e 3 por estarem sem nenhuma informação, restando assim 3697 para análise de correspondência do CEP com o endereço. Destes, 2592 (70,11%) estavam corretos e não necessitavam de correção. E 1105 apresentavam problemas que necessitavam de correção. Analisando os endereços e CEP, os motivos de alteração mais frequentes foram: 457 (41,36%) CEP eram de outro local; 265 (23,99%) apresentavam o CEP geral do município; 256 (23,16%) estavam sem CEP; 88 (7,96%) CEP não foram encontrados; 28 (2,54%) CEP haviam sido desmembrados; e 11 (0,99%) CEP estavam com falta de número, com espaço ou número a mais.

Além disso, analisando o banco outros problemas foram verificados como: bairro errado, ruas que mudaram de nome, endereços que não foi possível localizar, números de identificação com dois endereços, alguns com falta de identificação do município, outros endereços do município de Pelotas com a identificação de outro município no banco.

Sendo assim, 1017 números de identificação tiveram os CEP alterados e 88 não foi possível localizar. E a amostra que vai ser utilizada no estudo contará



3609 endereços e CEP, e uma perda de 12,60% da amostra inicial devido aos endereços pertencerem a outros municípios e 2,08% devido à impossibilidade de localização dos endereços contidos no banco.

O georreferenciamento possibilita o cruzamento de indicadores epidemiológicos entre si, auxilia na inserção de dados intersetoriais e apóia recortes analíticos que expressem as relações entre as condições de vida e os níveis de saúde e doença (PAIM, 1997). Com isso estes endereços e CEP, convertidos um par de coordenadas geográficas são ferramentas importantes para os estudos com esta finalidade, e por isso é importante uma análise criteriosa destes dados antes da realização do georreferenciamento.

Os dados possuem uma importância crescente, e são considerados um recurso essencial, porém constata-se que uma parcela dos dados são “dados sujos”. A presença de erros, redundâncias ou inconsistências faz com que esses dados apresentem problemas de qualidade e podem levar a perdas de 1-5%, podendo chegar até 30% (BARCELLOS et al., 2008; OLIVEIRA, 2009). No presente estudo a perda devido à impossibilidade de localização dos endereços foi de 2,08%, semelhante ao estudo de Oliveira (2009).

Dentre os principais fatores que contribuem para a baixa qualidade de dados, estão a introdução incorreta de dados ocasionada por erro humano. Alguns exemplos de erros humanos incluem: erros na digitação de dados; omissões de dados supostamente obrigatórios; introdução involuntária de valores errados; entre outros (OLIVEIRA, 2009). Os problemas relatados no estudo em relação ao banco fornecido, podem ter ocorrido em função destas informações terem sido fornecidas pelos participantes do estudo, além de possíveis erros de digitação. Uma conferência dos dados fornecidos evitou a perda de 1017 endereços da amostra.

4. CONCLUSÕES

Considerando que tratam-se de dados secundários, esta avaliação da qualidade dos dados torna-se imprescindível, pois as informações sobre o endereço e/ou CEP são utilizadas para permitir a construção do elemento geométrico do SIG. A demanda do setor Saúde por ferramentas de geoprocessamento é crescente. A plena utilização dessas técnicas depende, todavia, da disponibilidade e qualidade de dados georreferenciados. No presente estudo a correção do banco com o uso de ferramentas disponíveis evitou que a perda de informação para o estudo de georreferenciamento fosse maior.

5. REFERÊNCIAS BIBLIOGRÁFICAS

ARANHA, F. Atlas dos setores postais: uma nova geografia a serviço da empresa. **RAE - Revista de Administração de Empresas**, São Paulo, v. 37, n. 3, p. 20-27, 1997.

BARCELLOS, C.; RAMALHO, W. M.; GRACIE, R.; MAGALHÃES, M. A. F. M.; FONTES, M. P.; SKABA, D. Georreferenciamento de dados de saúde na escala submunicipal: algumas experiências no Brasil. **Epidemiol. Serv. Saúde**, Brasília, v. 17, n. 1, p. 59-70, 2008.

BRASIL, C. **Estrutura do CEP**. 2017. Disponível em: <
<https://www.correios.com.br/enviar-e-receber/ferramentas/cep/estrutura-do-cep> >. Acesso em: 19 de julho de 2020.

COLLISCHONN, E.; SIQUEIRA, K. C. D.; CASTILHOS, E. D.; BIGHETTI, T. I. Rede de atenção à saúde bucal nos setores público e privado: desigual distribuição geográfica no município de Pelotas/RS. **Hygeia**, v. 14, n. 30, p. 29-40, 2018.

MOREIRA, R. S.; NICO, L. S.; TOMITA, N. E. A relação entre o espaço e a saúde bucal coletiva: por uma epidemiologia georreferenciada. **Ciencia & Saude Coletiva**, v. 12, n. 1, p. 275-284, 2007.

OLIVEIRA, P. **Deteção e Correção de Problemas de Qualidade dos Dados: Modelo, Sintaxe e Semântica**. 2009. Tese de Doutorado. Universidade do Minho, Portugal.

PAIM, J. S. **Abordagens teórico-conceituais em estudos de condições de vida e saúde: notas para reflexão e ação**. In: Barata RB, organizador. Condições de vida e situação de saúde. Rio de Janeiro: Abrasco; 1997. p. 7-30.

PINA, M. F. **Conceitos básicos de Sistemas de Informação Geográfica e Cartografia aplicados à saúde**. Maria de Fátima de Pina e Simone M. Santos. Brasília: OPAS, 2000.

SILVEIRA, I. H.; OLIVEIRA, B. F. A.; JUNGER, W. L. Utilização do Google Maps para o georreferenciamento de dados do Sistema de Informações sobre Mortalidade no município do Rio de Janeiro, 2010-2012. **Epidemiol. Serv. Saúde**, Brasília. v. 26, n. 4, p. 881-886, out-dez 2017.

SKABA, D.A.; CARVALHO, M.S.; BARCELLOS, C.; MARTINS, P.C.; TERRON, S.L.. Geoprocessamento dos dados da saúde: o tratamento dos endereços. **Cad Saúde Pública** novembro-dezembro; v. 20, n. 6, p. 1753-1756, 2004.