Ciência & Saúde Coletiva
cienciaesaudecoletiva.com.br
ISSN 1413-8123. v.29, n.11
DOI: 10.1590/1413-812320242911.02672024

GeoCNES: healthcare mapping in Brazilian cities – a computational tool for improved decision-making

GeoCNES: mapeamento da saúde em cidades do Brasil – uma aplicação automatizada para auxiliar na tomada de decisão

GeoCNES: mapeo de la salud en ciudades de Brasil – una aplicación automatizada para ayudar en la toma de decisiones

Lucas Brandão Monteiro de Assis (https://orcid.org/0000-0002-8338-8535) [1,2]
Francisco Roza de Moraes (https://orcid.org/0000-0002-7413-6467) [1]
Paulo Cesar Lima Segantine (https://orcid.org/0000-0003-1012-0666) [1]
Miguel José das Neves Pires Amado (https://orcid.org/0000-0002-9152-4226) [2]
Irineu da Silva (https://orcid.org/0000-0001-5775-6683) [1]

Abstract *Ensuring equitable access to healthcare facilities is crucial for urban well-being, but geographical barriers often impede this access. This paper introduces GeoCNES, an open-source tool developed in Python to address this challenge. GeoCNES establishes a connection to the Brazilian national healthcare establishments register and the census data, to process and geocoding them to automatically generate an interactive map that display the distribution of healthcare facilities and a heat map of the same facilities in Brazilian municipalities. To do so the user must enter the municipality code and facility type, then GeoCNES retrieves, geolocates, and exhibit the information in interactive maps. This paper details the development process, functionalities, and limitations of GeoCNES, demonstrating its application in the Brazilian cities of São Carlos-SP, Rondonópolis-MT, Chapecó-SC, Parnamirim-RN and Parauapebas-PA. While challenges related to data inconsistency were encountered, GeoCNES successfully maps healthcare facilities, offering valuable insights for urban planning and promoting equitable access to healthcare.*
Key words *Healthcare mapping, Geographic information systems, Spatial analysis, Healthcare planning, Python*

Resumo *Garantir acesso equitativo a unidades de saúde é crucial para o bem-estar urbano, mas barreiras geográficas muitas vezes impedem esse acesso. Este artigo apresenta o GeoCNES, uma ferramenta de código aberto desenvolvida em Python para enfrentar esse desafio. O GeoCNES se conecta ao CNES e aos dados censitários brasileiros e aplica técnicas de geocodificação para gerar automaticamente mapas interativos que mostram a distribuição de unidades de saúde e sua concentração por meio de mapas de calor, em municípios brasileiros. Os usuários utilizam código do município e o tipo de unidade a ser analisado como parâmetros, e o GeoCNES recupera, geolocaliza e exibe os dados em mapas. Este artigo detalha o processo de desenvolvimento, funcionalidades e limitações do GeoCNES, demonstrando sua aplicação nas cidades de São Carlos-SP, Rondonópolis-MT, Chapecó-SC, Parnamirim-RN e Parauapebas-PA. Embora tenham sido encontrados desafios relacionados à inconsistência de dados, o GeoCNES é capaz de mapear com sucesso as unidades de saúde de diferentes regiões do país e gerar mapas com potencial para auxiliar no planejamento urbano voltado para a equidade na saúde.*
Palavras-chave *Mapeamento dos cuidados de saúde, Sistema de informação geográfica, Análise espacial, Planejamento de saúde, Python*

Resumen *Garantizar un acceso equitativo a las unidades de salud es crucial para el bienestar urbano, pero las barreras geográficas a menudo obstaculizan este acceso. Este artículo presenta GeoCNES, una herramienta de código abierta desarrollada en Python para abordar este desafío. GeoCNES se conecta al CNES y a los datos censales brasileños y aplica técnicas de geocodificación para generar automáticamente mapas interactivos que muestran la distribución de las unidades de salud y su concentración a través de mapas de calor en municipios brasileños. Los usuarios utilizan el código municipal y el tipo de unidad a analizar como parámetros, y GeoCNES recupera, geolocaliza y muestra los datos en mapas. Este artículo detalla el proceso de desarrollo, las funcionalidades y las limitaciones de GeoCNES, demostrando su aplicación en las ciudades de São Carlos-SP, Rondonópolis-MT, Chapecó-SC, Parnamirim-RN y Parauapebas-PA. Aunque se encontraron desafíos relacionados con la inconsistencia de datos, GeoCNES es capaz de mapear con éxito las unidades de salud de diferentes regiones del país y generar mapas con potencial para ayudar en la planificación urbana orientada a la equidad en la salud.*
Palabras clave *Mapeo de la atención de salud, Sistema de información geográfica, Análisis espacial, Planificación de salud, Python*

[1] Departamento de Engenharia de Transportes, Escola de Engenharia de São Carlos, Universidade de São Paulo. Av. Trab. São Carlense 400, Parque Arnold Schimidt. 13566-590 São Paulo SP Brasil. lucasbrandao@usp.br
[2] CITUA, Instituto Superior Técnico, Universidade de Lisboa. Lisboa Portugal.

Introduction

A population's well-being is intrinsically related to the ease to access healthcare facilities, which can be influenced by both non-spatial and spatial factors. Non-spatial factors encompass aspects such as the health status of an individual, socio-economic circumstances, and social support network. While, spatial factors involve the location and distance of healthcare facilities from users' origins, playing a crucial role in guaranteeing equitable access to healthcare services[1,2].

Despite the availability of healthcare services, many individuals face geographical obstacles, which impede their ability to access these essential services[3,4]. For instance, an estimated 18.1% of the population faces a critical level of accessibility, often resulting in incomplete care and compromising their health outcomes[5]. This is a concerning experience, considering that providing equitable access to all services and resources and overcoming the inequities in the urban space is one of the biggest challenges to the cities nowadays[6].

In this context, the third Sustainable Development Goal (SDG), proposed by the United Nations (UN), becomes even more crucial. This goal aims to ensure healthier lives and promote well-being for all, specifically focusing on achieving universal health coverage (UHC) through access to essential health services[7]. Therefore, considering geographical barriers to healthcare access is not only essential for fulfilling the objectives of the SDGs but also critical for ensuring equitable health outcomes and addressing the larger issue of urban inequities.

One noticeable issue in Brazil involves the connection between the allocation of new public facilities and the electoral political interests of the public stakeholders involved. These stakeholders, often lacking knowledge of technical aspects and methodologies that could enhance the effectiveness of healthcare network, are primarily focused on immediate social promotion that could yield future political advantages[8]. Moreover, the lack of access to adequate software and the difficulty of training professionals become another challenge in utilizing appropriate techniques[9].

The current approach to planning health facilities, of distinct kinds, is inadequate and involves different interests, leading to several detrimental consequences. From a financial standpoint, this lack of proper planning results in an inefficient use of public resources. Locating healthcare facilities with an improperly method often require more funding, diverting resources away from other essential healthcare services. Moreover, inadequate planning can result in an insufficient supply of healthcare facilities, which affects the underserved areas and decrease the accessibility to the users. This shortage of services disproportionately impacts those who are most in need, exacerbating health inequities and compromising the overall quality of healthcare delivery[10].

The literature holds several metrics and methods that can provide effective planning for the establishment of new health facilities[11-14]. These tools consider factors such as population density, healthcare needs, accessibility, and existing infrastructure to identify optimal locations for new facilities. The Geographic Information Systems (GIS) is an valuable tool to spatialize different aspects related to both demand and supplies, and in this way, it presents to be a great tool to assist the planning of public equipment, however, GIS' software used to be often expensive, when there was no open-source alternatives, and still today, can be difficult to use, which can turns its use unfeasible in the daily life of public administration[15].

In response to these challenges, this paper introduces GeoCNES, an automated, open-source computational tool meticulously designed to enhance the planning process for healthcare facilities. GeoCNES is based on official, open-source and up-to-date data of the existing healthcare infrastructure, provided by the National Register of Health Establishments – CNES and the Brazilian Institute of Geography and Statistics – IBGE, and focus on generate interactive maps illustrating the distribution of facilities of distinct levels of care across municipalities.

Bibliographic review

The rapid urban growth, especially in the low- and middle-income countries (LMICs), already threatens to worsen the existing inequalities[16]. Sustainable urban planning is crucial for healthy and equitable development in these regions. However, current practices in LMICs, like Brazil, often prioritize short-term political gains over long term planning[17,18], which leads to an inefficient use of resources.

Effective urban planning necessitates a comprehensive approach that utilizes data, metrics, and continuous monitoring to understand community needs and adapt public policies accordingly[19]. The Department of Information Technol-

Ciência & Saúde Coletiva, 29(11):1-12, 2024

ogy of the SUS (DATASUS) serves as a central repository of extensive information, including data on healthcare utilization, health facility infrastructure, and staff demographics. This data not only informs decision-making but also plays a crucial role in developing effective health action programs[10].

DATASUS constitutes an extensive array of information systems that encompasses a diverse range of data and sectors within the SUS. The CNES plays a pivotal role in providing public managers with a comprehensive overview of the social and healthcare landscape across Brazilian regions and municipalities. This invaluable data enables the formulation and implementation of effective public health policies based on accurate and up-to-date information on health facilities nationwide[20].

However, despite its extensive data repository, the CNES web implementation presents challenges in effectively disseminating information on health facilities[21]. The reliance on highly specific terms without visual support often hinders the accurate representation of healthcare facilities' locations, this spatial limitation impedes user access to the system and its valuable data[22].

There are several ways to address the spatial visualization of healthcare systems, one of them is geocoding, which as defined by the Earth Science Research Institute – ESRI[23], involves transforming location descriptions, such as addresses or geographic coordinates, into precise spatial information referenced to the earth's surface. By incorporating geocoding techniques, the CNES can effectively translate descriptive data into meaningful spatial representations, thereby facilitating the identification of the health facilities' location and distribution.

The quality of geocoding can be evaluated through various means, not solely restricted to the positional accuracy of the information. The ISO-19157 – Geographic Information – data quality standard considers important to evaluate five key groups of elements: the completeness, the logical consistency, the positional accuracy, the thematic accuracy, and the temporal quality of the features[24].

However, due to the informative nature of this standard, some authors[25] have adapted the sets of elements for quality assessment, employing only three analysis elements:

. Completeness assesses the actual correspondence of geocoding and analyzes the temporal quality of the dataset to determine the temporal validity of the data analyzed.

. Positional accuracy evaluates the positional placement of elements present in the database.

. Data repeatability indicates the consistency of the results obtained by querying the database regarding positional changes across the mapped regions.

In the current landscape, geographic databases maintained by private companies like Google and Microsoft, as well as the collaborative effort of OpenStreetMap, are the most prevalent geocoding tools, and they have made noteworthy progress in creating user-friendly and freely accessible Application Programming Interfaces (API) for users[26]. These APIs are designed to simplify the development and integration of computer applications through different computational languages, with varying degrees of geocoding functionality and quality. These variations arise from the extensive amount of data and information related to various public and private sources that constitute the databases, rendering them highly valuable commodities for businesses[27].

The integration of APIs from renowned companies, such as Google and OpenStreetMap, has revolutionized the development of comprehensive geographic applications, empowering decision-makers across various management domains. These APIs offer sophisticated geolocation and mapping capabilities, enabling easy integration with their extensive and up-to-date geographic databases. This integration promotes more reliable and high-quality query responses, enhancing the effectiveness of geospatial analysis[28].

Consequently, this integration facilitates the spatialization and conversion of information, such as the distribution of health facilities, into informative maps. These visual representations empower urban planners to identify and analyze health service infrastructure distribution patterns, revealing inequities in access to healthcare services. This capability empowers urban planners to make well informed decisions about the placement of new health facilities, ensuring equitable access to healthcare for all residents[15].

In the context of public health, these applications emerge as invaluable tools for administrative authorities engaged in the planning and implementation of new health facilities. By providing real-time data on population distribution, healthcare needs, and infrastructure availability, these tools facilitate the identification of strategic areas for the expansion of health services. This data-driven approach streamlines decision-making in defining the geolocation of new health facilities, optimizing resource allocation, and maximizing public health outcomes.

Methods

Effective health facility location planning necessitates a robust and up to date database. While accessible databases offer detailed information, managing data analysis can be challenging. To address this, GeoCNES, a Python-based application, was developed to facilitate health planning diagnosis.

GeoCNES is an automated tool based on two key parameters: the municipality IBGE code and facility CNES code. The application generates an interactive map of the health facilities distribution with a density distribution of them, and a facility-per-inhabitants' ratio of that municipality.

Because of its popularity, variety, and flexibility, Python (in its 3.12.2 version) was chosen as the programming language to develop this tool, to facilitate future expansions and functionality development by both creators and interested users. To ensure this the code was indented, allowing clarity for those who read the code, enabling replication and utilization for different applications by third users.

GeoCNES retrieves data from CNES and IBGE databases, storing essential information in algorithmic variables, processing them until the determination of the facility's coordinates that allows the creation of the interactive map. Figure 1 shows an overview of the GeoCNES's functionality, the processes presented in it are discussed during the present section.

Package install and first steps

To start using the GeoCNES is necessary to install the computer package available at GitHub (github.com/lucasbrnd/GeoCNES). Once the user accesses the page, the code can be retrieved and loaded into the Python environment and started.

The GeoCNES application operates with only two input parameters provided by the user:

1) the municipality code, as designated by IBGE, which is comprised with seven digits, in which the first two corresponding to the Federation Unit[29], that can be consulted at IBGE's Cities;

2) the specific code of the health establishment type, comprised by two digits, according to the CNES' documentation[30].

After entering the parameters, the application goes through validation process, and if the inputs are valid a new directory is created in the working directory, to store necessary files to the application functioning.

Queries on CNES' and IBGE database

The data process occurred after the input stage consists in querying the CNES' and IBGE's database. To do this the GeoCNES relies on the Numpy, Pandas, OS, Requests, URLlib, ZipFile and Xarray libraries, which enables the management of alphanumeric data stored in data frames and online repositories. Both data sources have a standardized URLs that varies around the two codes requested from the user.

The first connection of the application is to the IBGE database referent to the 2010's census, to download the census tract geographical data, which is done by replacing the state code in the following link:

*geoftp.ibge.gov.br/organizacao_do_territorio/malhas_territoriais/malhas_de_setores_censitarios__divisoes_intramunicipais/2021/Malha_de_setores_(shp)_por_UFs/**[state_code]**/ **[state_code]**_Setores_2021.zip*

Then, the algorithm unzips and scans the acquired files, maintaining only the geographic files relevant to the city of study. The second query is to the CNES system, and follows a similar process, in which the codes from city, state, and health facility type, serves as replacement in specific parts of the following URL, which provides a list of all the establishments of a specific type in the same municipality according to the month previous to the query.

*cnes2.datasus.gov.br/Mod_Ind_Unidade_Listar.asp?VTipo=**[HF_type]**&VListar=1&Vestado=**[state_code]**&Vmun=**[city_code]**&VsubUni=&Vcomp=00*

The obtained list comprises of all the facilities' register number and name, which are the two-information needed as parameter to the next query, that replace those codes in the following link, in order to obtain the address of each establishment.

*https://cnes2.datasus.gov.br/cabecalho_reduzido.asp?VCod_Unidade=**[HF_code]***

Each query conducted yields an extensive array of data on the healthcare facilities, such as their name, address, responsible organization, and many other information. To attend the GeoCNES' objectives, the data was subject to an extraction process in which unnecessary information to identify the facility location were ignored, while the necessary part was stored in a new data frame, in which each line corresponds

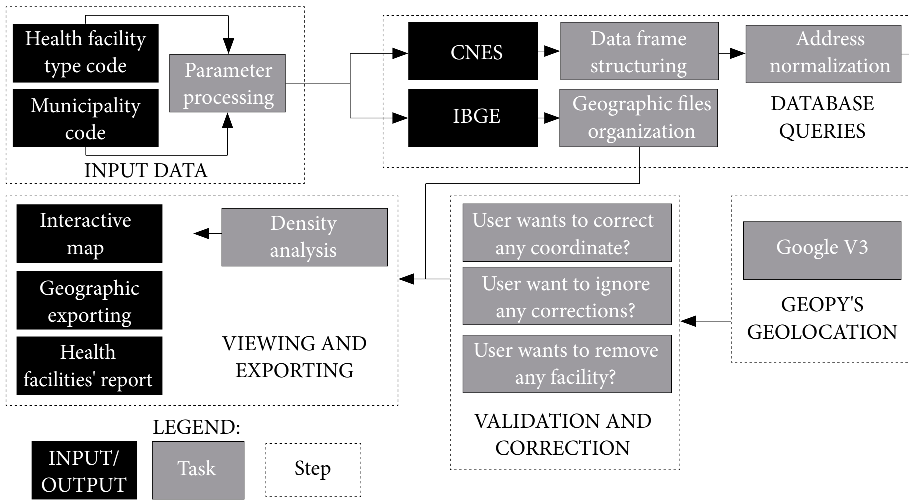

Figure 1. GeoCNES' algorithm overview.

Source: Authors.

to a health facility, and each column is a field of the address information.

The extraction process is recursive and goes through all the entries of the list. When the extraction is completed, a normalization of the address is conducted to concatenate the information into a single field, structured in the following sequence: facility's name, street, number, complement, neighborhood, postal code, city, and state, separated by commas. By this, all facilities are well identified and able to be geolocated.

Geolocation of healthcare establishments

The geolocation process consists of querying specific databases to obtain geographic coordinates correspondent to alphanumeric data, such as the address or a spatial reference[31]. The geocoding solution in GeoCNES is provided by Geopy library, a Python client that locates coordinates for addresses, cities, countries, and landmarks worldwide[32]. This library is attached to different geocoders, such as GoogleV3, Nominatim, Bing and ArcGis[32-37].

The geocoding process of this tool is related to the queries sent to the geocoding that convert the addresses in coordinates. The details about how the geocoder works are not presented here, because some of those processes are related to proprietary algorithms that are not very public for third users.

During the conception of the work two geocoders were considered: Google V3 and Nominatim, due to their worldwide recognition. However, results obtained by the Nominatim exhibited less precision and accuracy, with more failures in identification different facilities than Google V3. This problem was also identified by Clemens[38], Das and Purves[39] and Serere[40], who pointed that Nominatim's limitation to handle sentences with different spelling and error could be the cause to this imprecision, while Google V3 have a greater capacity of to locate and clean and fuzzy matches on wrong addresses, resulting in a more successful geocoding process.

Due to this, the Google V3 was selected to be the geocoder applied in the application, which requires that the user creates an account in Google Maps Platform to obtain the API key. Google V3's geolocating process is currently free but is subject to limitations on the number of addresses queried in a single search.

Once the user register is concluded and the API key is provided to GeoCNES, all the addresses in the list are geolocated and a new geo data frame, storing points features, is created. Although Google V3 is more successful in geolocate the facilities some errors could be found, and to deal with them a validation and correction process was also implemented.

Validation and manual correction

Throughout the development of the GeoCNES algorithm, particularly in the geocoding step, recurring errors were identified. Those errors primarily involved the wrong assignment of the health facility's point, either outside of the municipality's boundaries or at the city's central identification point. To mitigate such errors, a four-step validation process, reliant on user intervention, was defined:

The first step involves the comparison of the resulting coordinate to the city's central coordinate. If those coordinates match, the user is prompted to provide the correct coordinate to that facility or to exclude that point of the analysis. This error often stems from CNES's incomplete or outdated information.

The second step verifies whether the coordinates of the health facilities are within the city's boundaries. If discrepancies are noted, users are given the option to rectify the coordinates of the respective facility, or to exclude those establishments from their analysis.

The third step offers the user the option to manually update the coordinates of specific establishments.

The fourth step empowers the user to select and exclude any health facility of their choosing.

By offering this validation process, GeoCNES endeavors to ensure the autonomy of the user to observe and correct any situation derived from an imprecise data source or geolocating failure. The end of this phase marks the consolidation of the geographic database used to elaborate the interactive maps.

Viewing and exporting data

The GeoCNES' last phase is the exhibition of the results in the interactive maps. This was possible due to Geoviews and Folium libraries, which provide the necessary resources to produce a heat map analysis of the facilities distribution and an interactive map of the city.

The presented map shows three different geographic features: a point layer with the health facilities, a polygon layer with the census tract features, and a geographic layer representing the concentration of facilities. Those geographic layers are presented over a basemap.

The point layer contains the health facility information, such as their name, address and coordinate. The census tract layer is used to identify the city limits and the urban and rural areas. While the heat map helps to identify the concentration of facilities in the city in a clearer way.

The interactive display option allows the user to zoom in or out on the map window and identify facility's information with greater precision, by selecting any facility on the map that pops up a window presenting the facility information (name, address and coordinates).

Furthermore, users can export the processed data for use in different software. The export function allows the user to save the geographic files presented in the interactive map in geopackage file format.

Results and discussions

This section presents the outcomes of GeoCNES application, and the discussion derived from it. To verify the application's functionality five case studies were conducted in five different cities. The first selected city was São Carlos due to the availability of health facilities' data of Assis and Segantine[11] previous works, with a intention to compare the coordinates presented by the authors and the locations resulted by the GeoCNES method.

São Carlos is in the state of São Paulo, in the Southeast region of Brazil. The other four cities are in a different region of the country, and were selected based on their similarity to São Carlos, considering the number of inhabitants, to enrich the discussion through a fair comparison between the distribution of health facilities among them.

Since the primary healthcare facilities have a significant importance to the healthcare systems, due to its character of universal access and first contact with the health system, the presented case studies concentrated the analysis on the health facilities of this level of care.

São Carlos-SP's case study

São Carlos, the primary case study for GeoCNES, has a population estimated is 254,857 inhabitants in 2022's census[41]. Figure 2 presents the maps of distribution of the primary care facilities in São Carlos.

From the GeoCNES' map and GeoCNES's outcome it is possible to see that health facilities in São Carlos are well distributed. The main gaps are in the center region, which concentrates more commercial activities, and in the corner of the Northwest region (Parque dos Flamboyants), which are recent neighborhoods. This observation is also pointed out by Assis and Segantine[11] findings in their analysis of accessibility of the PHC facilities in São Carlos. When the health facilities location of GeoCNES and Assis and Segantine[11] work were compared only one facility presented a failed geolocating result, which resulted in a difference between the authors work, apart from this, every location was a match and a facility that was not included in their work was identified by GeoCNES.

Regarding the ratio of facilities per population, the Health Ministry defines that each familiar health team is responsible to 4.000 people and in São Carlos the ratio of facilities per population is 1 per 7079 inhabitants (as can be seen in Table 1), which is greater than the official objective, but it's not a conclusive data since more analysis is required to define the amount of a city population is in fact a regular demand of those health facilities.

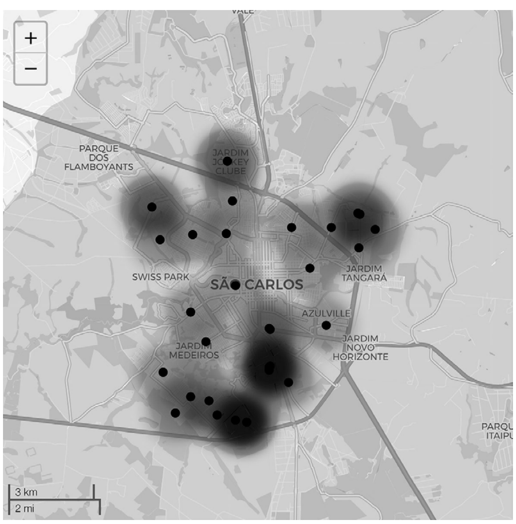

Figure 2. Primary healthcare facilities distribution in São Carlos – SP.

Source: Authors.

Other case studies

The GeoCNES was also tested with cities of different regions of the country, but with a similar population to São Carlos. The four selected cities were in Chapecó-SC (South region), Parnamirim-RN (Northeast region), Rondonópolis-MT (West Center region), and Paraupebas-PA (North region). In Table 1 the results of the facility location is shown as the ratio of inhabitants per facility.

In Figure 3 the results obtained for the four cities are presented. The city with greater number of health facilities of the PHC level is Rondonópolis-MT, which directly affects the ratio of facilities per inhabitant, making this city the only one (between the five cities mentioned) with a ratio lesser than the Ministry of Health proposition.

Considering the cities in which the GeoCNES was tested, a total of 20 facilities presented some kind of error. Those errors were most frequent among the PHC facilities, although one facility of the tertiary level also presented this error. When the address of those facilities was investigated, it was possible to note that seventeen of them had their number omitted from the address. The reason for the missing information is unknown, since the fill of those registers is made by health centers employees, but it could be a signal of some kind of error in registers' filling process, or it could represent the situations in which the address indeed does not contain any number. In any case, this reinforces the importance of de-

Table 1. GeoCNES' facility location outcomes and inhabitants per facility ratio.

City (IBGE'S code)	Population	Existing facilities			Facilities located without errors			Inhabitants by facility		
		PHC	SHC	TCH	PHC	SHC	TCH	PHC	SHC	THC
São Carlos-SP (3548906)	254,857	36	4	5	35	4	5	7,079	63,714	50,971
Chapecó–SC (4204202)	254,785	28	2	2	26	2	2	9,099	127,393	127,393
Parnamirim–RN (2403251)	252,716	29	1	4	27	1	4	8,714	252,716	63,179
Rondonópolis-MT (5107602)	244,911	66	1	5	58	1	4	3,711	244,911	48,982
Parauapebas-PA (1505536)	267,836	26	1	7	19	1	7	10,301	267,836	38,262

Source: Authors.

tailed information to a successful process of geolocation based on address.

Table 2 presents the required time to acquire the data from CNES. Regarding the time to retrieve the data from the CNES, the results were expected, since cities with more health facilities represent more data to be retrieved. Regarding this result, it is interesting to observe that the time needed to retrieve the address from hospitals of Parnamirim-RN was an outlier in those observations, which is related to the CNES system connection stability. The geocoding time follows the same rule, and queries with more entries took longer than the minor ones.

Challenges, limitations and future steps

The main challenge encountered during GeoCNES's development is related to inconsistencies of the CNES system, which initially seems to be a geocoder issue, but during the investigation and solving process it was possible to note that some establishments had incomplete registrations, with missing street numbers or names, or even spelling errors. These deficiencies sometimes made the geocoding process harder, and in some cases impossible.

Rocha *et al.*[21] have investigated the quality of information in the CNES database, and that

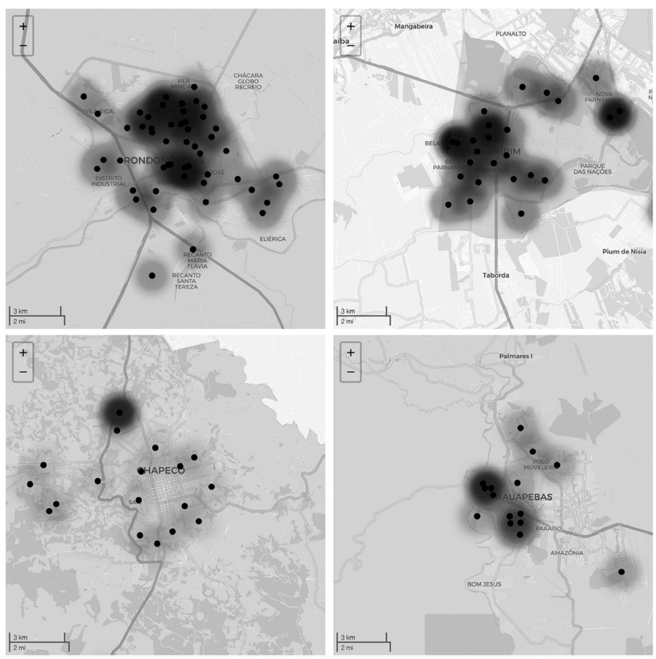

Figure 3. Primary healthcare facilities distribution in Rondonópolis-MT (top-left), Chapecó-SC (bottom-left), Parnamirim-RN (top-right) and Parauapebas-PA (bottom-right).

Source: Authors.

Table 2. Time spent in the CNES query and geocoding for each city and level of care.

City	Existing Facilities			CNES' query time (seconds)			Geolocating time (seconds)		
	PHC	SHC	TCH	PHC	SHC	THC	PHC	SHC	THC
São Carlos-SP	36	4	5	105.56	11.09	14.97	17.45	2.87	3.43
Chapecó-SC	28	2	2	76.21	5.72	6.41	9.91	1.92	1.90
Parnamirim-RN	29	1	4	74.46	2.45	43.38	9.80	1.40	2.55
Rondonópolis-MT	66	1	5	173.98	2.83	14.80	20.97	1.38	3.17
Parauapebas-PA	26	1	7	70.35	2.63	19.78	8.38	1.57	3.91

Source: Authors.

Ciência & Saúde Coletiva, 29(11):1-12, 2024

CNES' data have numerous inconsistencies, particularly regarding the number of beds and the operational status of equipment, but also in their address identification, as a consequence 63% of hospitals nationwide were located within a 1-kilometer radius of the addresses listed in the CNES database, in which for 10% of those, this distance could extend up to 5 kilometers.

It is worth noting that despite overcoming these challenges, there was a persistent random error occurring when GeoCNES attempted to communicate with CNES. This error occasionally caused application crashes, halting functionality temporarily. It is possible that the CNES system could become overwhelmed by a high volume of queries in a brief period, leading to these intermittent errors. The straightforward solution to this problem involves re-running the code.

However, this turned out to be the main limitation of the application, which is related to query information of cities with great number of facilities, such as São Paulo-SP, a city that has 518 primary healthcare facilities, and recurrently resulted in crashes when used as input. Future works could investigate the reasons to this problem and propose a way to avoid this failure.

Another limitation identified is related to the reliability of pre-geolocation data. The information utilized as enter parameter to the geolocation process are provided by CNES database, which as discussed, can present some inconsistencies. Then to enhance the process a user validation that points to the user the facilities in which some kind of problem was encountered, being the user the responsible one to correct those facilities.

Although GeoCNES presents itself as an innovative tool to retrieve alphanumeric data and convert them into a geographic information, it is important to mention that computational tools to assist multidisciplinary interests by establishing a connection to governmental data sources and retrieving different data to the user was already developed. A notable example that is Geobr[42], a R's and Python's package that allows the retrieval of official Brazilian spatial data from various sources, including healthcare infrastructure available for CNES, using a more generalized approach. Also, with a different approach, there is also the Microdatasus[43], an application that focuses primarily on analytical and statistical analysis of microdata, with less emphasis on spatial aspects.

GeoCNES emerged within the academic context of postgraduate research. Initially conceived to address the need for a streamlined and automated approach to obtain health facility locations, it has since evolved into a more sophisticated tool with multidisciplinary applicability.

The application's development remains ongoing and future enhancements will focus on refining accessibility analyses based on city street networks. An enhancement to identify the health facilities' capacity (number of physicians and nurses) and services offered is under development.

Presently, GeoCNES resides in a dedicated GitHub repository – a centralized platform that facilitates access to the software's source code and encourages contributions from users. This transparency ensures traceability throughout the development process. Once new functionalities are introduced, the code will be diligently updated.

Final considerations

The growth of urban populations has increased the need for comprehensive urban planning to optimize infrastructure and services, ensuring efficient access for all residents. This process involves analyzing data on population distribution, transportation networks, and existing facilities to identify areas requiring new or relocated facilities. By carefully placing healthcare centers, schools, and transportation hubs, urban planners can minimize travel times and enhance quality of life for all residents.

In an intention to contribute to the solution of those problems, GeoCNES is proposed as a user-friendly and accessible computer tool enabling the visualization of healthcare infrastructure distribution within the urban area of any Brazilian municipality, regardless of the user's educational background or training.

While GeoCNES does not delve into the operational aspects and types of services offered at each healthcare establishment, its significance lies in its ability to offer a product that shows the spatial distribution of health facilities.

For future endeavors, consideration should be given to adapting the method to other programming languages and expanding its capabilities to enable a micro-analysis of the healthcare system within the municipality under investigation. Additionally, utilizing the territorial division into census sectors to correlate the spatial distribution of healthcare establishments with socioeconomic data would be an intriguing avenue for future research.

Collaborations

LBM Assis conceived the work, developed and tested the program, wrote and revised the text and approved the final version; FR Moraes developed and tested the program, wrote and revised the text and approved the final version; PCL Segantine supervised the work, wrote and revised the text and approved the final version; MJNP Amado, supervised the work and approved the final version; and I Silva supervised the work and approved the final version.

Acknowledgments

This study was financed in part by the CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) - Finance Code 163811/2021-0 and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil) – Finance Code 88887.886769/2023-00, as well as the Departamento de Engenharia de Transporte (STT) da Escola de Engenharia de São Carlos (EESC) da Universidade de São Paulo (USP), and the Foundation for Science and Technology's (FCT) support through funding UIDB/05703/2020 from the research unit CiTUA at the Instituto Superior Técnico (IST) of the Universidade de Lisboa (UL).

References

1. Kanuganti S, Sarkar AK, Singh AP. Evaluation of access to health care in rural areas using enhanced two-step floating catchment area (E2SFCA) method. *J Transp Geogr* 2016; 56:45-52.
2. Guagliardo MF. Spatial accessibility of primary care: concepts, methods and challenges. *Int J Health Geogr* 2004; 3(1):3.
3. Reinhardt U, Cheng T. The world health report 2000 – health systems: improving performance. Bull World Health Organ. 2000;78(8):1064.
4. Pan American Health Organization (PAHO). Redes integradas de servicios de salud: conceptos, opciones de política y hoja de ruta para su implementación en las Américas [Internet]. 2010. [accedido 2023 jun 11]. Disponible en: https://iris.paho.org/bitstream/handle/10665.2/31323/9789275331163-spa.PDF
5. Dantas MNP, Souza DLB, Souza AMG, Aiquoc KM, Souza TA, Barbosa IR. Factors associated with poor access to health services in Brazil. *Rev Bras Epidemiol* 2021; 24:e210004.
6. United Nations Human Settlements Programme. World Cities Report 2022 [Internet]. 2022. [cited 2023 jun 11]. Available from: https://www.un-ilibrary.org/content/books/9789210028592c011
7. United Nations (UN). Ensure healthy lives and promote well-being for all at all ages [Internet]. 2021. [cited 2024 fev 6]. Available from: https://sdgs.un.org/goals/goal3#progress_and_info
8. Santos FDA, Gurgel Júnior GD, Gurgel IGD, Pacheco HF, Bezerra AFB. A definição de prioridade de investimento em saúde: uma análise a partir da participação dos atores na tomada de decisão. *Physis* 2015; 25(4):1079-1094.
9. Boeing G. The right tools for the job: the case for spatial science tool-building. *Trans GIS* 2020; 24(5):1299-1314.
10. Piccolo DM. Qualidade de dados dos sistemas de informação do Datasus: análise crítica da literatura. *Cienc Info Rev* 2018; 5(3):13-19.
11. Assis LBM, Segantine PCL. Proposal of a multicriteria method to implement new primary health care units – a case study in São Carlos-SP. *Rev Bras Cartogr* 2021; 73(4):1071-1085.
12. Goudard B, Oliveira FH, Gerente J. Avaliação de modelos de localização para análise da distribuição espacial de Unidades Básicas de Saúde. *Rev Bras Cartogr* 2015; 67(1):15-34.
13. Colaço PMPLM. *Critérios para o planeamento de equipamentos de saúde: análise de caso de estudo no contexto urbano da AML* [dissertação]. Lisboa: Universidade Nova de Lisboa; 2011.
14. Guida C, Carpentieri G, Masoumi H. Measuring spatial accessibility to urban services for older adults: an application to healthcare facilities in Milan. *Eur Transp Res Rev* 2022; 14(1):23.

15. Boeing G, Higgs C, Liu S, Giles-Corti B, Sallis JF, Cerin E, Lowe M, Adlakha D, Hinckson E, Moudon AV, Salvo D, Adams MA, Barrozo LV, Bozovic T, Delclòs-Alió X, Dygrýn J, Ferguson S, Gebel K, Ho TP, Lai PC, Martori JC, Nitvimol K, Queralt A, Roberts JD, Sambo GH, Schipperijn J, Vale D, Van de Weghe N, Vich G, Arundel J. Using open data and open-source software to develop spatial indicators of urban design and transport features for achieving healthy and sustainable cities. *Lancet Glob Heal* 2022; 10(6):e907-e918.
16. United Nations Department of Economic and Social Affairs. 2018 revision of the world urbanization prospects [Internet]. 2018. [cited 2023 dez 2]. Available from: https://population.un.org/wup/Publications/Files/WUP2018-Methodology.pdf
17. Dreux VP. *Uma avaliação da legislação urbanística na provisão de equipamentos urbanos, serviços e áreas de lazer em conjuntos habitacionais* [dissertação]. Porto Alegre: Universidade Federal do Rio Grande do Sul; 2004.
18. Moraes AF. *Análise dos processos de definição utilizados pelas prefeituras, para o local de implantação de Equipamentos Urbanos Comunitários (EUCs), em municípios do estado de Santa Catarina* [tese]. Florianópolis: Universidade Federal de Santa Catarina; 2013.
19. Lowe M, Adlakha D, Sallis JF, Salvo D, Cerin E, Moudon AV, Higgs C, Hinckson E, Arundel J, Boeing G, Liu S, Mansour P, Gebel K, Puig-Ribera A, Mishra PB, Bozovic T, Carson J, Dygrýn J, Florindo AA, Ho TP, Hook H, Hunter RF, Lai PC, Molina-García J, Nitvimol K, Oyeyemi AL, Ramos CDG, Resendiz E, Troelsen J, Witlox F, Giles-Corti B. City planning policies to support health and sustainability: an international comparison of policy indicators for 25 cities. *Lancet Glob Heal* 2022; 10(6):e-882-e894.
20. Brasil. Ministério da Saúde (MS). *DataSUS – Trajetória 1991-2002*. Brasília: MS: 2002.
21. Rocha TAH, Da Silva NC, Barbosa ACQ, Amaral PV, Thumé E, Rocha JV, Alvares V, Facchini LA. Cadastro nacional de estabelecimentos de saúde: Evidências sobre a confiabilidade dos dados. *Cien Saude Colet* 2018; 23(1):229-240.
22. Brasil. Ministério da Saúde (MS). Cadastro Nacional de Estabelecimentos de Saúde [Internet]. 2024. [acessado 2024 abr 15]. Disponível em: https://cnes.datasus.gov.br/
23. ESRI. What is geocoding? [Internet]. 2016. [cited 2023 dez 20]. Available from: https://desktop.arcgis.com/en/arcmap/10.3/guide-books/geocoding/what-is-geocoding.htm
24. ISO. ISO 19157:2023 - Geographic information - Data quality. 2023. p. 102.
25. Whitsel EA, Rose KM, Wood JL, Henley AC, Liao D, Heiss G. Accuracy and repeatability of commercial geocoding. *Am J Epidemiol* 2004; 160(10):1023-1029.
26. Bandil A, Girdhar V, Dincer K, Govind H, Cao P, Song A, Ali M. An interactive system to compare, explore and identify discrepancies across map providers [Internet]. 2020. [cited 2023 out 13]. Available from: https://dl.acm.org/doi/10.1145/3397536.3422348
27. Moraes FR. *Proposta de um modelo genérico de um SBDE que permita a interoperabilidade entre sistemas* [tese]. São Carlos: Universidade de São Paulo; 2017.
28. Präger M, Kurz C, Böhm J, Laxy M, Maier W. Using data from online geocoding services for the assessment of environmental obesogenic factors: a feasibility study. *Int J Health Geogr* 2019; 18(1):13.
29. Instituto Brasileiro de Geografia e Estatística (IBGE). Códigos dos municípios IBGE [Internet]. 2024. [acessado 2024 jan 10]. Disponível em: https://www.ibge.gov.br/explica/codigos-dos-municipios.php
30. Brasil. Ministério da Saúde (MS). Categoria: Nova Classificação de Tipos de Estabelecimento [Internet]. 2020. [acessado 2024 jan 10]. Disponível: https://wiki.saude.gov.br/cnes/index.php/Categoria:Nova_Classificação_de_Tipos_de_Estabelecimento
31. Bakshi R, Knoblock CA, Thakkar S. Exploiting online sources to accurately geocode addresses [Internet]. 2004. [cited 2024 jan 10]. Disponível em: https://dl.acm.org/doi/10.1145/1032222.1032251
32. Geopy. Geocoders [Internet]. 2023. [cited 2023 mar 6]. Available from: https://geopy.readthedocs.io/en/stable/#
33. Teske D. Geocoder accuracy ranking. In: Lamprecht AL, Margaria T, editors. *Process design for natural scientists. communications in computer and information science*. Berlin: Springer; 2014. p. 161-174.
34. Google. Geocoding API [Internet]. 2023. [cited 2023 mar 6]. Available from: https://developers.google.com/maps/documentation/geocoding/
35. OpenStreetMaps. Nominatim API [Internet]. 2023. [cited 2023 mar 6]. Available from: https://nominatim.org/release-docs/develop/api/Overview/
36. Microsoft. Bing Maps Locations API [Internet]. 2023. [cited 2023 mar 6]. Available from: https://learn.microsoft.com/en-us/bingmaps/rest-services/locations/?redirectedfrom=MSDN
37. ESRI. Geocoding service [Internet]. 2023. [cited 2023 mar 6]. Available from: https://developers.arcgis.com/rest/geocode/api-reference/overview-world-geocoding-service.htm
38. Clemens K. Enhanced address search with spelling variants [Internet]. 2018. [cited 2024 jan 10]. Available from: http://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0006646100280035
39. Das RD, Purves RS. Exploring the potential of Twitter to understand traffic events and their locations in Greater Mumbai, India. *IEEE Trans Intell Transp Syst* 2020; 21(12):5213-5222.
40. Serere HN, Resch B, Havas CR. Enhanced geocoding precision for location inference of tweet text using spaCy, Nominatim and Google Maps. A comparative analysis of the influence of data selection. *PLoS One* 2023; 18(3):e0282942.
41. Instituto Brasileiro de Geografia e Estatística (IBGE). Cidades e Estados [Internet]. 2022. [acessado 2022 dez 11]. Disponível em: https://www.ibge.gov.br/cidades-e-estados/

42. Pereira R, Gonçalves C. geobr: Download official spatial data sets of Brazil [Internet]. 2023. [cited 2023 dez 11]. Available from: https://github.com/ipeaGIT/geobr
43. Saldanha RDF, Bastos RR, Barcellos C. Microdatasus: A package for downloading and preprocessing microdata from Brazilian Health Informatics Department (DATASUS). *Cad Saude Publica* 2019; 35(9):e00032419.

Article submitted 13/03/2024
Approved 17/04/2024
Final version submitted 19/04/2024

Chief-editors: Maria Cecília de Souza Minayo, Romeu Gomes, Antônio Augusto Moura da Silva