Use of geocoding techniques for epidemiological surveillance in the Federal District, Brazil: a case study using dengue

Lucas Carvalho Sanglard,[1] Klauss K. S. Garcia,[2] Walter Massa Ramalho[3]

[1]Ministry of Health, Secretariat of Health Surveillance and Environment, Brasília; [2]The London School of Hygiene and Tropical Medicine, Department of Infectious Disease Epidemiology and International Health, Faculty of Epidemiology and Population Health; [3]Centre for Tropical Medicine Brasília, University of Brasília, Brazil

Correspondence: Lucas Carvalho Sanglard, Ministry of Health, Secretariat of Health Surveillance and Environment, SRTVN Quadra 701 Edifício PO 700, Lote D - Asa Norte, Brasília, DF, 70719-040, Brazil.
Tel.: +55 61 99844-2698
E-mail: sanglardbsb@gmail.com

Key words: georeferencing; geocoding; dengue; SINAN; epidemiological surveillance; public health; Brazil.

Ethics Committee: the project was submitted to and approved by the Research Ethics Committee of the Faculty of Medicine at the University of Brasília (CAAE 07443418.4.0000.5558), ensuring compliance with ethical guidelines for the use of sensitive data.

Availability of data and material: the datasets used and analysed during the current study are available from the corresponding author upon reeasonable request.

Conflict of interest: the authors declare that they have no competing interest.

Contributions: LS conceptualized and designed the study, and organized the data; KKSG reviewed the text and the methodology; WMS secured funding and provided supervision. All authors read, revised, and approved the final manuscript.

Funding: this project was funded by Call CNPq/MCTI/FNDCT No. 22/2024 – Knowledge Brazil Program – Support for Network Projects with Brazilian Researchers Abroad – process 445334/2024-0.

Acknowledgments: we would like to thank the National Council for Scientific and Technological Development (CNPq) for providing a scholarship during the development of this project, and the Brazilian Ministry of Health for supporting the research.

Received: 29 April 2025.
Accepted: 14 June 2025.

Geospatial Health 2025; 20:1403
doi:10.4081/gh.2025.1403

Abstract

This study aimed to compare different address geocoding services and their applicability to epidemiological surveillance using dengue as an example. We applied a cross-sectional, descriptive study based on case notifications in the Notifiable Diseases Information System (SINAN) for the Brazilian capital in 2014 that includes complete postal code (CEP) information identified in the National Address Database for Statistical Purposes (CNEFE), which is considered the 'gold standard' for accuracy analysis. For records without CEP, georeferencing was performed through linkage of the original database with four geocoding tools: Google Maps, CNEFE, OpenStreetMap (OSM) and ArcGIS. Variables used for georeferencing were 'street name', 'code for municipality/city of residency' and 'State' using accuracy rate estimate and mean spatial error (MSE) of case locations. The two most accurate models were used for kernel density (KD) analysis which is valuable for identifying priority areas for intervention. There were 18,206 dengue cases, 109 (0.6%) of which had correct CEP information and geocoded using CNEFE bases. The linkage results showed that Google Maps application programming interface (API) had an accuracy of 17.6% (MSE: 178.89km), CNEFE 9.0% (MSE: 17.24km), OSM 7.1% (MSE: 564.19km), and ArcGIS 3.7% (MSE: 2001.33km). Although overall accuracy values were modest, the best two models proven to be effective for KD analysis revealed similar patterns between Google Maps and CNEFE results but choosing the preferable geocoding technique should also financial resources. This study recommends the use of Google Maps API for georeferencing, followed by CNEFE.

Introduction

Dengue, an arboviral disease primarily transmitted by the *Aedes aegypti* mosquito that is highly adaptable to urban environments, represents a global public health challenge estimated by the World Health Organization (WHO) to have reached 390 million cases in 2023, primarily in tropical and subtropical countries (WHO, 2025). In Brazil, the Ministry of Health (MoH) reports that dengue exhibits an endemic-epidemic pattern, with recurrent outbreaks that strain healthcare systems (MoH, 2025a, 2025b). In recent years, the increased number of cases has been attributed to factors such as climate change—expanding the suitable areas for vector reproduction—and sociodemographic aspects, such as unplanned population growth, rapid urbanization and unequal access to basic sanitation services (Seixas *et al.*, 2024). In the Brazilian Capital – the Federal District, these factors create a scenario of vulnerability that underscores the need for integrated strategies for disease control and monitoring. The region recently faced a dengue epidemic in 2024, with the Secretariat of Health of the Federal District (SES-DF) recording over 284,000 probable cases and a cumulative incidence rate of 8685.9 cases per 100,000 inhabitants, the biggest in the district's history (SES-DF, 2024).

The Federal District presents stark social contrasts. As reported by the Planning Company of the Federal District (CODEPLAN), some areas present high Human Development Index (HDI) scores, while others suffer from limited access to sanitation, healthcare and infrastructure (CODEPLAN, 2020). Affected areas usually are distant one from another – which makes vulnerability an isolated problem. The lack of adequate sanitation, combined with substandard housing and social inequality, creates a fertile ground for the spread of *Ae. aegypti* that has, consequently, led to increased dengue incidence (WHO, 2025). Therefore, it is essential to understand the spatial distribution of cases—especially in micro-territories marked by high vulnerability—as this is needed for guiding and improving epidemiological surveillance efforts while enhancing health planning and contributing to dengue control. Epidemiological surveillance is a the key tool for dengue control and prevention in Brazil and it plays a critical role in ensuring the principles of equity within the Brazilian Unified Health System (SUS) (WHO, 2025; MoH, 2025b). Given that health surveillance is a continuous and systematic process of data collection, consolidation and analysis of disease dissemination, the use of georeferencing techniques stands out as an ideal approach for this purpose (MoH, 2024).

Geoprocessing refers to a set of techniques focused on managing spatial information, enabling the characterization of health-related events (Fantin *et al.* 2021). It associates data with specific geographic locations using latitude and longitude data, facilitating the use of spatial analyses for the identification of spatial patterns (Barcellos *el al.,* 2008). Geocoding is a georeferencing technique that converts address data into geographic coordinates (latitude and longitude) and enables the representation of these data on maps or spatial analysis systems. Geocoding can be performed through different computer methods, such as Application Programming Interface (API), local software or pre-mapped databases—each presenting advantages and limitations, particularly in terms of precision and cost (Skaba, 2009; McDonald *et al.*, 2017).

APIs are online services, such as Google Maps and OpenStreetMap (OSM), that retrieve and translate address information into geographic coordinates. This approach is widely used due to its ease of integration and its ability to handle large data volumes. However, it may involve associated costs and it depends on stable internet connections.

Local software and pre-mapped databases employ locally installed programs that use algorithms to process addresses based on stored datasets. While they reduce reliance on external services and may avoid request limits and usage cost, this approach typically requires more infrastructure and software capability and may be limited by outdated databases (McDonald *et al.*, 2017).

Considering the limitations of each method and the need to integrate them within public health surveillance, this study aimed to describe and compare four different geocoding methods. The analysis focused on spatial distribution of georeferenced dengue cases, accuracy of API-based services, processing time, mean location error and quality and completeness of the input data.

Materials and Methods

We applied a cross-sectional descriptive study on dengue case notifications recorded in 2014 by SINAN for Brasília, the capital of Brazil, located in the Federal District (Figure 3). The study compared the performance of four geocoding tools: Google Maps (2025), ArcGIS, ESRI, Redlands, CA, USA), OSM and the Brazilian National Address Database for Statistical Purposes (CNEFE), a georeferenced address dataset produced by the Brazilian Institute of Geography and Statistics (IBGE, 2025). CNEFE was also used as the gold standard for comparison, as it provides precise latitude and longitude coordinates for addresses with correctly filled postal codes (CEP) in Brazil.

Data processing and treatment

Data processing and analysis were performed using the R programming language within the R Studio environment (version 4.2.1), employing the various application packages, such as 'tidyverse' (Wickham *et al.,* 2019), 'tidygeocoder' (Cambon *el al.,* 2021) and shiny packages (Wickham, 2021). The methodology was structured into the following steps:

Data cleaning and standardisation

The fields for street name (NM_LOGRADO), state of residence (SG_UF), and country (PAIS) were linked together into a unified address field.

Geocoding using API services

Geocoding was performed using Google Maps, ArcGIS and OSM, converting textual addresses into geographic coordinates (latitude and longitude) for each record in the dengue dataset.

Probabilistic geocoding

Textual similarity between dengue dataset addresses and CNEFE records were compared by accounting for matching characters and necessary transpositions making it effective for handling minor typographical errors in street names, neighbourhoods and city names was computed using the Jaro-Winkler (JW) distance method (Jaro, 1989; Winkler, 1990) that calculates a similarity score, with values ranging from 0 (completely dissimilar) to 1 (identical) by the formula:

$$JW = J + (l \cdot p \cdot (1 - J)) \quad \text{Eq. 1}$$

where JW is the Jaro-Winkler similarity score, ranging from 0 (completely dissimilar) to 1 (identical); J is the Jaro Similarity Score; l is the length of the common prefix at the beginning of the string; and p is a constant scaling factor indicating the weight given to the common prefix. Commonly, p=0.1. The 0.5 threshold was selected to balance the sensitivity of the geocoding process, allowing the identification of a broader range of potential address matches as well as considering the frequent inconsistencies. In this context, more sensitive string-matching strategies were considered more appropriate to the study's objective, which is to apply geocoding techniques to support health surveillance efforts. By using a lower threshold, we aimed to maximise the detection of meaningful spatial patterns of disease distribution, even at the cost of increasing the number of matches requiring manual or probabilistic validation.

Selection of the gold standard for comparison

Gold standard coordinates were established by matching the CEP field in the SINAN database to the corresponding postal code in the CNEFE. Records with matching CEPs were considered to be reliably georeferenced. Thematic maps were created to illustrate the spatial distribution of the generated coordinates. Charts were used to display the proportions of geocoded notifications for each API, allowing identification of spatial patterns related to dengue cases.

Presentation of results

Data were presented in an interactive dashboard developed using the 'shiny package' enabling the display of the number of matched addresses, the mean location error (in metres), the accuracy percentage, the distances calculated between the obtained coordinates vis-à-vis the gold standard. This allowed a side-by-side comparison of the geocoding tools used.

Accuracy

To assess the precision of geographic coordinates generated by the different geocoding APIs, an accuracy test was conducted using CNEFE as the gold standard reference. The distances between the geocoded points and the gold standard were calculated using the 'sf' package in R. This package expresses the geodesic distances between two sets of points as latitude and longitude in the WGS84 coordinate reference system (EPSG:4326). The equation was implemented in the 'st_distance' function that accounts for the Earth's curvature, ensuring greater accuracy in distance calculations. The results were analysed based on the following metrics: i) number of matches: the count of point pairs where the calculated distance falls within a defined threshold, based on matching notifications from the APIs and the CNEFE dataset using a unique identifier (*ID_UNICO*); ii) mean distance: the average of the absolute distances measured in metres; iii) mean distance: the average distance converted from metres to kilometres; iv) accuracy (%): The proportion of observations where the distance between the API-geocoded point and the CNEFE reference falls within the predefined threshold (in metres), relative to the total number of observations, *i.e.* the number of points with distance ≤threshold/total number of points x 100.

$$\text{Mean Distance} = \frac{\sum_{i=1}^{n} \text{distance}_i}{n} \qquad \text{Eq. 2}$$

where n is a total number of matched point pairs included in the calculation, i is index representing each individual matched pair of points (a geocoded point and its corresponding reference point in the gold standard - CNEFE), and distance_i is the geodesic distance in metres between the geocoded point and its corresponding gold standard point (CNEFE).

Data analysis

The two geocoding methods with the best performance results were incorporated into a spatial analysis model using Kernel Density (KD) estimation, commonly known as a heatmap. It was conducted to identify areas with a higher concentration of georeferenced notifications and to enable a comparative evaluation of the different geocoding services analysed. This analysis was performed using R Studio (version 4.2.1) along with the 'leaflet','leaflet.extras' and 'sf' packages.

The coordinates were structured in tabular format and converted into a spatial object called Simple Features (sf) with projection in the WGS 84 spatial reference system (EPSG:4326) using the 'st_as_sf ()' function from the 'sf' package.

The KD was calculated dynamically using the 'leaflet.extras' package, which enables the creation of interactive heatmaps, which is generated by the 'addHeatmap()' function, where each point contributes to the intensity of density in its surrounding area. The parameters used were: i) intensity = 1, used for assigning equal weight to all points; ii) blur = 20, used to control blur and smooth the areas of higher concentration; iii) radius = 15, used to define the radius influence of each point in the density calculation; iv) max = 0.05, that represents the maximum value of relative intensity. As a supporting element for interpretation, a shapefile containing the boundaries of census tracts in the Federal District was incorporated (Figure 3). This shapefile was obtained from the IBGE Census Tract Grid and imported using the 'st_read()' function. The boundary lines were added to the map using the

Table 1. Characteristics of the geocoding tools and the computational infrastructure used

Geocoding software and database			
API	**Free request**	**Additional cost**	**Request/second**
Google Maps	10,000/month	US$ 20/5000 requests	50
OSM	Unlimited	Free	1
ArcGIS	Unlimited	Free	1
CNEFE	Unlimited	Free	Set by hardware capacity

Hardware Specifications	
Hardware	**Details**
Motherboard	TUF GAMING B550M-Plus
Processor	AMD Ryzen 7 5700X 8-Core Processor (16 CPUs), ~ 3.4GHz
RAM	2x16GB Kingston Fury (32768MB RAM)
Graphics card	NVIDIA GeForce RTX 3060 12 GB – INNO3D
Internal storage	SSD 1TB Kingston
Operating system	Windows 10 Pro 64 bits (10.0, Compilation 19045)

API, application programming interface; RAM, random access memory; OSM, OpenStreetMap; CNEFE, Cadastro Nacional de Endereços para Fins Estatísticos (National address database for statistical purposes).

'addPolylines()' function, configured to display the borders in black, with a line weight of 1 and an opacity of 0.8. Table 1 presents the main characteristics of the API services used in this study, including the number of free requests allowed the cost per thousand additional requests, the request rate per second, and information about the hardware setup.

Results

The dataset used in this study comprised 18,206 dengue notifications recorded at location. Among these, only 233 records (1.3%) contained a complete CEP. Of these, 109 records (46.8%) were geocoded directly via CEP using the CNEFE database and were used to define the gold standard. The results of the different geocoding approaches are detailed in Table 2, which details the proportion of geocoded cases, the time required for each method and the challenges associated with each tool. In terms of geocoding coverage, ArcGIS achieved 100% of geocoded notifications, closely followed by Google Maps, which geocoded 99.3% of the records correctly. CNEFE returned 10,430 records with a string similarity score of ≤0.5. OSM showed the lowest coverage, geocoding only 26.3% of the notifications correctly (Table 1).

When applying a threshold of 100 metres, only a small proportion of records were correctly geocoded. Google Maps demonstrated the highest accuracy (17.6%) followed by OSM (7.1%). CNEFE achieved 6.4% accuracy at this level using the probabilistic geocoding technique. The least accurate method was ArcGIS, with only 4 notifications (0.02%) falling within the threshold (Table 2). Figure 1 illustrates the number of matches across the different geocoding services highlighting cases that were incorrectly geocoded giving locations in other countries, states or municipalities. In terms of processing time, Google Maps was the fastest

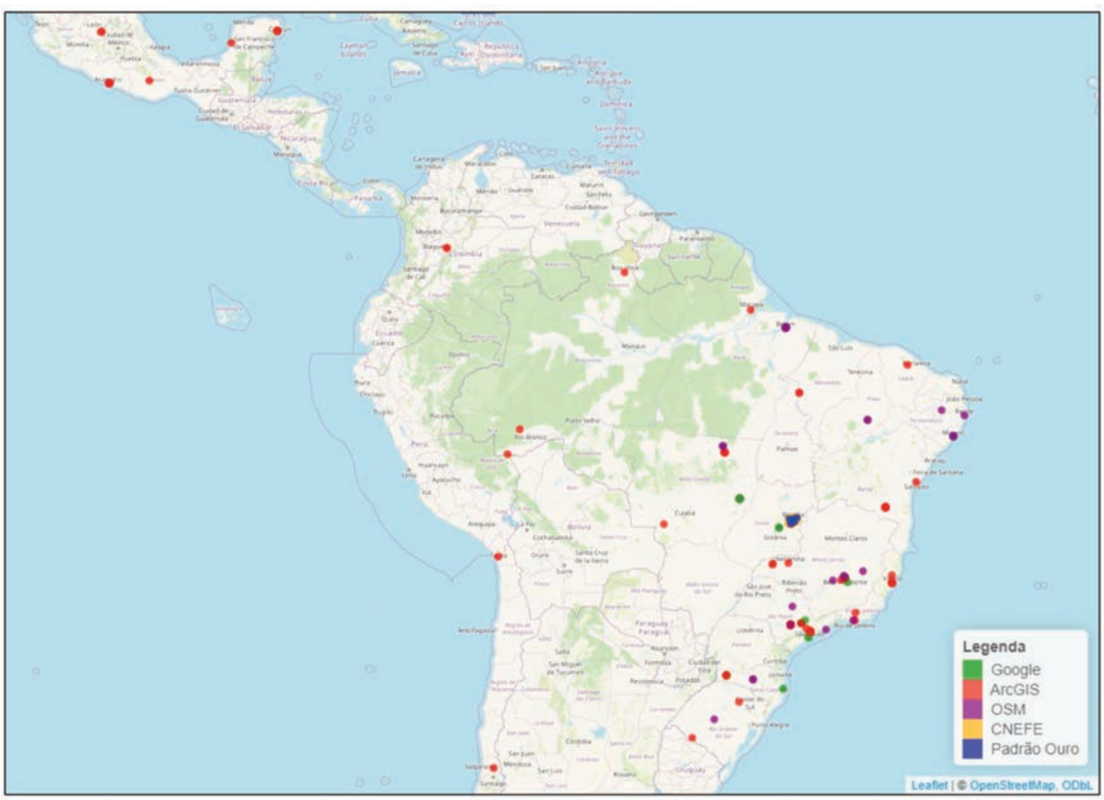


Figure 1. Geocoding of dengue cases. Caption: OSM, OpenStreetMap; CNEFE, National Address Database for Statistical Purposes.

Table 2. Geocoding results by method.

API	Notification no (%)	Accuracy no (%)	Processing time (min)	MDE (km)
Google Maps	18,073 (99.3)	19 (17.59)	47	178.89
OSM	4,785 (26.3)	4 (7.14)	307	564.00
ArcGIS	18,206 (100)	4 (3.64)	119	2001.33
CNEFE	10,430 (57.3)	6 (8.96)	83	17.24

API, application programming interface; MDE, mean distance error; CNEFE, Cadastro Nacional de Endereços para Fins Estatísticos (National address database for statistical purposes): OSM, Open Street Map.

method, completing all records in 47 min. In contrast, OSM required the longest time, taking 307 min (>5hours). ArcGIS and CNEFE had intermediate processing durations; however, the extended runtime of OSM reflects its lower request-per-second capacity (Table 1). At 17.24 km, CNEFE demonstrated the lowest mean distance error (MDE), followed by Google Maps (178.89 km) and OSM (564.19 km). Table 1 summarises the performance results of the geocoding APIs and the CNEFE database in comparison with the gold standard. The indicators used include the number of matches, accuracy (percentage of matches within the 100-metre threshold), and the mean distance in metres and kilometres. Figure 2 presents the spatial distribution of dengue notifications geocoded by each service (Google Maps, ArcGIS, OSM and CNEFE). The maps were used to visually evaluate the spatial coverage and precision of each method in relation to the gold standard.

The spatial distribution of cases geocoded using CNEFE data showed points overlapping with the gold standard in the regions of Ceilândia, Recanto das Emas and Asa Sul (Plano Piloto). Locations obtained by Google Maps also exhibited overlaps in Ceilândia, Guará and the Plano Piloto (Asa Sul, Asa Norte, and Lago Norte) (Figure 3). ArcGIS showed overlaps in Santa Maria and Asa Sul; however, a substantial number of records were widely dispersed, with some geocoded as located in other Brazilian states, even Mexico. Similarly, geocoding using OSM geocoding resulted in coordinates outside the Federal District, with locations distributed across other Brazilian states.

To assess the applicability of geocoding techniques for public health services, Figure 4 presents a KD analysis of dengue cases reported in the Federal District, using points geocoded by Google Maps API and by probabilistic geocoding based on CNEFE data. Finally, an online dashboard was developed, enabling visualization of the previously described information. The platform enables users to compare the spatial distribution of cases across the different geocoding methods and is available at: https://sanglard.shinyapps.io/Geocoding/ (Figure 5).

Discussion

The study underscored shortcomings in the address entry process in notification forms (Skaba *et al.*, 2004). The use of probabilistic methods to identify similar texts responds directly to the vulnerability of data entry (Klaus *et al.*, 2023). At present, SINAN does not support the capture of geographic coordinates (e.g., via smart devices), and the CEP field is frequently left blank, which highlights the need for improved professionals training and the incorporation of technological innovations into reporting tools to facilitate more accurate data collection and thereby improve geocoding processes (Miranda *et al.*, 2013). This undertaking emphasizes the pressing need to enhance the process of address registration process in notification forms and their integration into health information system. The study therefore investigated short-

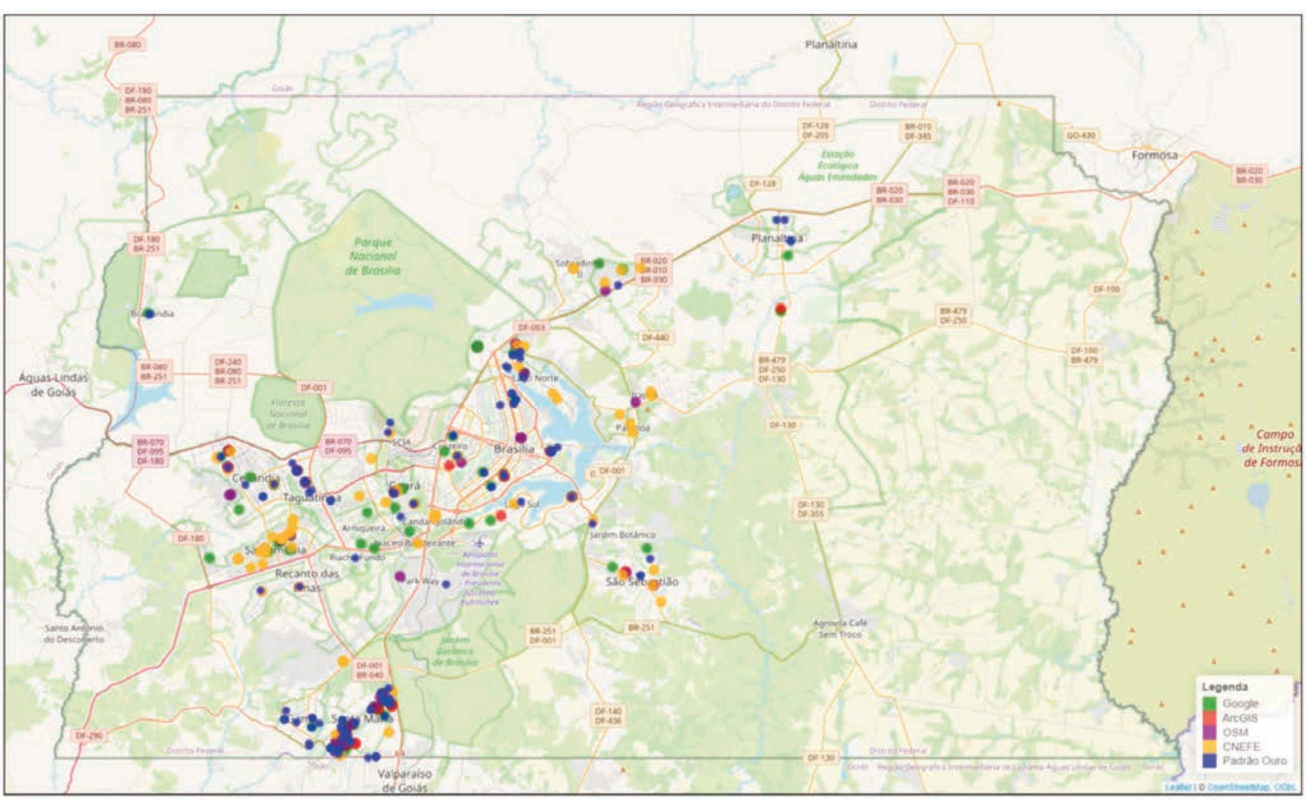


Figure 2. Spatial distribution of dengue cases geocoded in 2014 in Brasília, Federal District (DF). OSM, OpenStreetMap; CNEFE, National Address Database for Statistical Purposes.

comings in the address entry process in notification forms

Previous research has described procedures for cleaning and standardising textual in Brazilian health information systems - not only for georeferencing purposes but also to enable better performance from similarity algorithms (Skaba *et al.*, 2004; Magalhães *et al.*, 2014; Garcia *et al.*, 2022; Garcia *et al.*, 2023), and if data cleaning and standardization had not been performed, geocoding results would have been poorer. The extremely low proportion of cases with valid and complete postal code information (only 0.6% of all notifications) underscores a major limitation in Brazil's routine epidemiological surveillance systems and reflects a broader structural challenge concerning the quality of data recorded in notification forms, particularly within SINAN. The absence of detailed address information hinders the application of spatial analysis techniques—such as geocoding—and restricts the responsiveness and precision of public health interventions. This situation points to systemic underinvestment in the training of healthcare professionals responsible for case reporting, as well as a lack of robust quality control mechanisms during data entry. Bridging this gap is crucial for accuracy, usability and overall impact of surveillance data in informing localised health interventions.

With regard to accuracy, Google Maps geocoding presented the highest value among the methods evaluated, although it also showed a relatively high MDE, indicating moderate precision compared to CNEFE's gold standard. CNEFE, ranking second showed greater spatial proximity compared to the other methods,

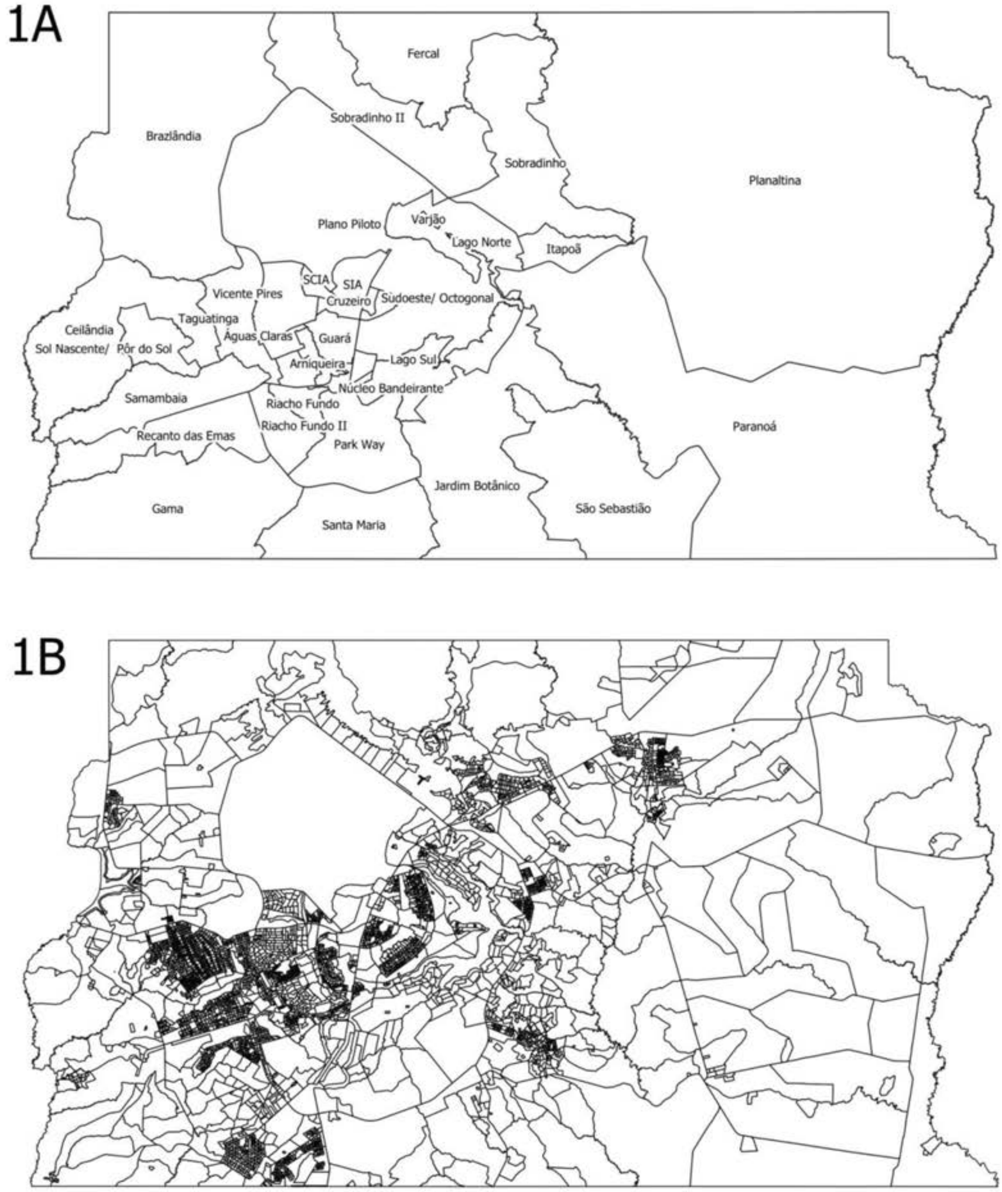
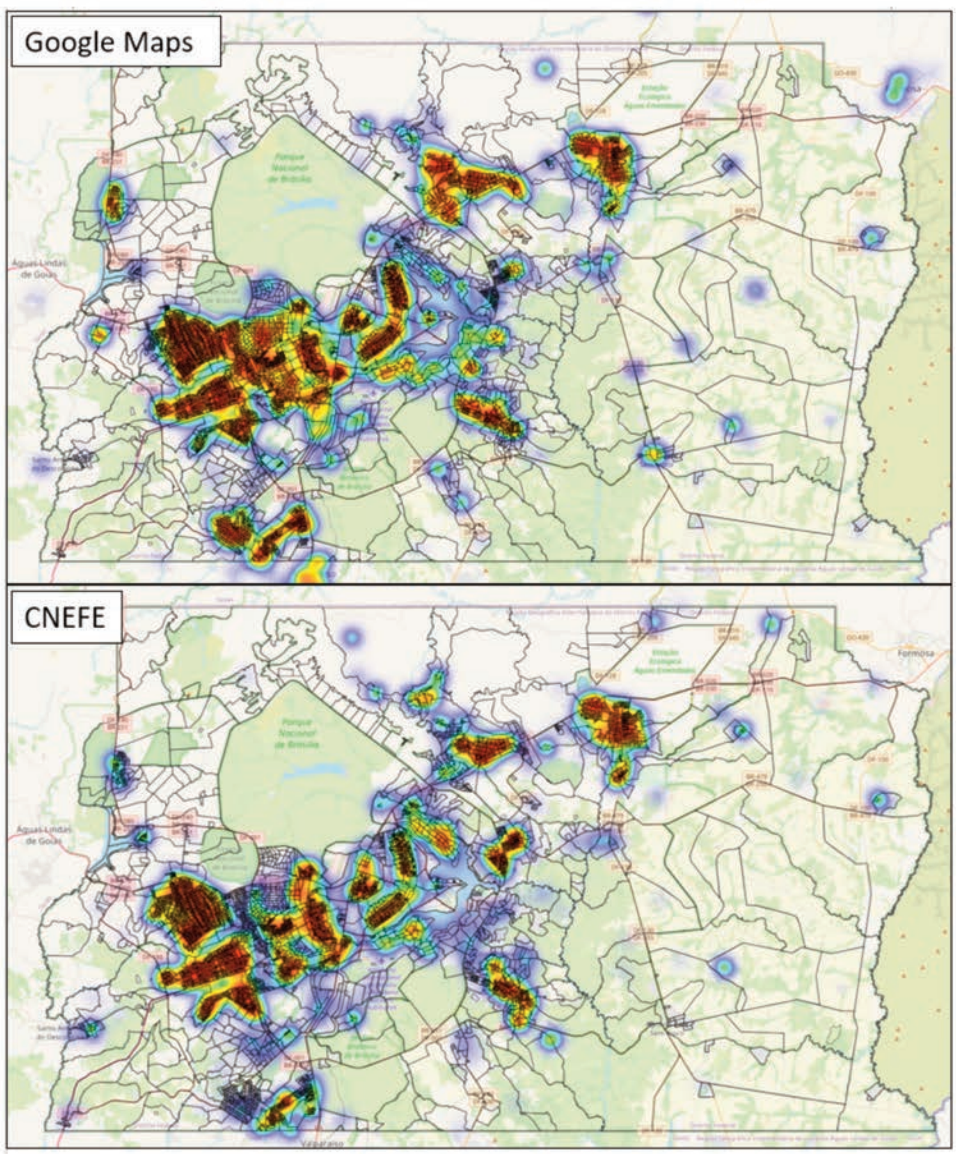


Figure 3. A) Administrative Regions of Brasília/Federal District; **B**) Census Tract Grid of Brasília, Federal District.

while OSM reflected a notable limitation in geocoding precision. ArcGIS, on the other hand, recorded the lowest performance, with both low accuracy and an extremely high mean error, a fact that either revealed substantial discrepancies in geocoded capacity or low address quality in the SINAN database. Given these results, the probabilistic geocoding technique using CNEFE addresses emerges as a viable alternative for federal and state-level applications, particularly in contexts where technological infrastructure and financial resources are limited, rendering commercial APIs less feasible. On the other hand, the limited performance of OSM highlights the need for improvements in its search engine, which restricts its large-scale applicability. Nevertheless, being a free API makes it a viable alternative for municipalities and institutions with limited resources, particularly when combined with local or probabilistic methods to improve data quality. This reinforces the importance of complementing traditional approaches with probabilistic techniques that are better suited to handling incomplete or inconsistent records.

The comparison between the KD maps generated using Google Maps and CNEFE-derived coordinates highlighted an important finding for public health practice. Although the accuracy of individual coordinates may vary between sources, the resulting spatial patterns of dengue case concentration were broadly similar. Both maps (Figure 4) reveal consistent hotspots in the south-western and northern parts of the Federal District, particularly in highly populated peripheral regions, which suggests that even geocoding



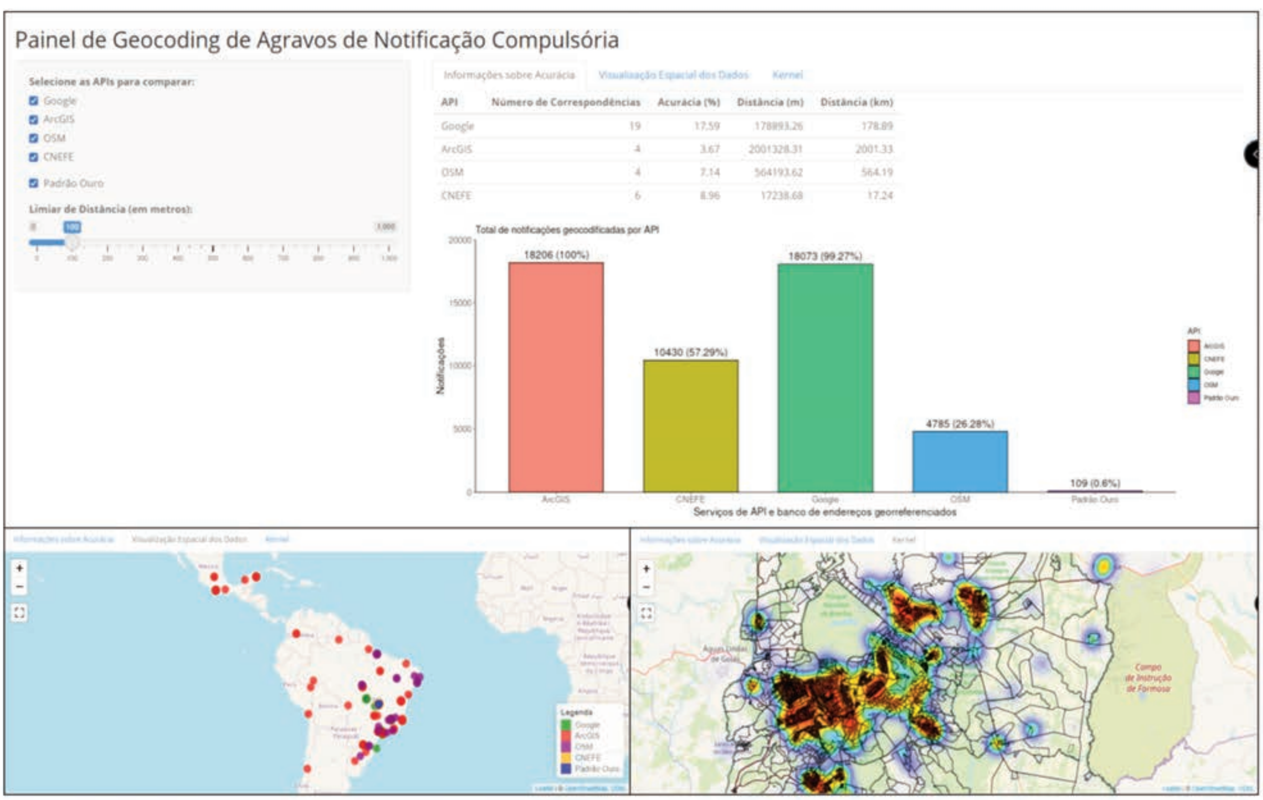
OSM, OpenStreetMap; CNEFE, National Address Database for Statistical Purposes.

Figure 4. Kernel density analysis of dengue cases in Brasília, Federal District. 2014. OSM, OpenStreetMap; CNEFE, National Address Database for Statistical Purposes.

methods with lower precision can offer meaningful information. In contexts marked by incomplete or low-quality address data—as is the case with many Brazilian surveillance systems, employing multiple geocoding strategies may help mitigate data limitations whilst still guiding the spatial targeting of control actions and resource allocation.

The choice of geocoding tool to be used must be guided by the financial constraints, and the operational reality of health services, particularly in municipalities with greater vulnerability (Battesini *et al.,* 2017; Vieira, 2020). Google Maps demonstrated superior performance in terms of both accuracy and the number of geocoded notifications. However, free usage of this tool is limited to 2,500 requests, and the high costs beyond this threshold may be unfeasible for large-scale use in public health services and research institutions in Brazil, which often operate under tight budgetary constraints (Battesini *et al.,* 2017; Silveira *et al.,* 2017; Oliveira *et al.,* 2020; Quintans-Júnior *et al.,* 2024). Conversely, the CNEFE method proved to be a more accessible and sustainable alternative, owing to the volume of geocoded matches, acceptable levels of accuracy, and satisfactory performance in spatial precision among the evaluated free services.

The use of commercial tools such as Google Maps for geocoding in public health surveillance raises critical concerns about equity, sustainability, and long-term integration into Brazil's decentralised health system. Although these platforms offer high geocoding precision, they depend on continuous internet connectivity, licensed API usage, and technical expertise that may not be uniformly available across all municipalities and states. Given the structural disparities in financial and human resource capacity across Brazil's states, the adoption of such tools may inadvertently widen the gap between better- and less-resourced regions. However, relying exclusively on commercial solutions is neither sustainable nor equitable. An alternative lies in the strategic combination of different geocoding methods—drawing on open-access government databases (such as CNEFE), probabilistic techniques and selective use of commercial APIs—to tailor the geocoding process to local capacities. This blended approach considers flexibility, cost containment, and broader applicability, ensuring that even areas with limited infrastructure can still conduct meaningful spatial analyses to inform surveillance. Moreover, combining methods can enhance both sensitivity and geographic coverage, compensating for the limitations inherent in each individual source. While technically more complex, such a strategy can be adapted incrementally and scaled according to local context, ultimately contributing to a more resilient and inclusive surveillance system. A previous study also examined geocoding for dengue in the Federal District between 2010 and 2015, utilising a significantly larger number of cases. In that research, the geocoding rate



Avaliable in: https://sanglard.shinyapps.io/Geocoding/

Figure 5. Online dashboard displaying geocoding results of dengue cases in Brasília, Federal District. Avaliable in: https://sanglard.shinyapps.io/Geocoding/

ranged from 77.2% and 89.4%, and only the Google Maps API was employed, with PHP as the programming language. That study did not compare different geocoding methods, nor did it consider the financial implications of relying on Google's services (Lustosa, 2017), but a separate study (Barcellos *et al.*, 2008), reported the need for georeferencing across several Brazilian capitals, revealing efficiency variations between cities ranging from 40% to 90%. This wide disparity reflects the sensitivity of address data entry in disease or health event notifications (SINAN) and other national health databases. The collection of high-quality address information is essential to enhance the effectiveness of these processes (Magalhães *el al.,* 2014).

Although the geocoding results from previous studies are high and broadly compatible with those observed in the present study , it is important to note that neither did these studies a gold standard for assessing accuracy, nor did they report distance error metrics in metres or kilometres. The absence of a reference benchmark in earlier research enhances the methodological innovation of the current study. The implementation of the interactive 'Shiny dashboard' enhances decision-making for health managers by enabling real-time visualization of spatial case distribution and facilitating the comparison of different geocoding methods (Katapally *et al.,* 2023). This dashboard allows users to assess the quality of geocoded notifications and identify priority areas, thereby improving the efficiency of resource allocation (Katapally *et al.,* 2023).

Potential limitations

Despite the study's contributions, certain limitations must be acknowledged. The reliance on complete and standardized data posed a challenge, as did the costs associated with commercial APIs, which may hinder their use in resource-constrained settings. Moreover, the requirement for adequate technological infrastructure and stable internet connectivity represents a critical success factors for remote geocoding.

The choice of a Jaro-Winkler similarity threshold of 0.5 may have influenced the sensitivity and specificity of the geocoding process. While this lower threshold increases the likelihood of capturing a larger number of true matches, it also raises the possibility of introducing false-positive pairs, which may have contributed to the error rates reported in the results. Nevertheless, this trade-off was considered appropriate given the study's objective of supporting health surveillance practices. By adopting a more inclusive matching criterion, the analysis aimed to better detect spatial patterns of case distribution—even in the context of incomplete or poorly standardised address data—thereby enhancing the targeting of public health interventions and resource allocation at the community and neighbourhood levels. This can be observed in the heatmap visualisations produced.

Conclusions

Georeferencing of dengue case notifications proved useful for identifying critical areas, thereby enhancing the efficiency of public health planning and interventions. The integration of different territorial databases strengthens the robustness of spatial analyses and underpins health surveillance strategies. Although technically complex, combining geocoding methods is better suited to the specific demands and available resources in each context. The approach used in this study enabled an objective comparison between geocoded coordinates against a gold standard, providing robust evidence to support the integration of spatial analysis into public health practices. Although accuracy limitations were observed across all tested geocoding methods, their application remains practical and valuable for identifying priority areas and guiding targeted surveillance and control interventions. Geocoding effectiveness is heavily dependent on the quality of the address data recorded in notification systems. The low accuracy noted with certain geocoding services highlights an urgent need to enhance data-entry practices, particularly by ensuring the consistent recording of postal codes, thereby facilitating more reliable spatial analyses. To overcome these challenges, we recommend implementing standardised address data-entry protocols within SINAN and selecting geocoding tools that align with the financial and operational capacities of local health services. Moreover, we advocate integrating mobile-enabled georeferencing technologies into national surveillance systems. Equipping health professionals with the ability to capture geographic coordinates directly via smartphones or tablets would significantly improve data accuracy, reduce reliance on retrospective geocoding and enhance the timeliness and efficacy of public health responses. Finally, sustained investment in health education initiatives—including training and sensitisation programmes focused on correctly completing notification forms—is essential. Such measures not only enhance data quality but also cultivate a culture of surveillance awareness among healthcare workers, ultimately strengthening the strategic application of spatial data in preventing and controlling notifiable diseases.

References

Barcellos C, Ramalho WM, Gracie R, Magalhães MAFM, Fontes MP, Skaba D, 2008. Georreferenciamento de dados de saúde na escala submunicipal: algumas experiências no Brasil. Epidemiol Serv Saúde. Avaliable from: http://dx.doi.org/10.5123/S1679-49742008000100006

Battesini M, Andrade CLT de, Seta MHD, 2017. Financiamento federal da Vigilância Sanitária no Brasil de 2005 a 2012: análise da distribuição dos recursos. Ciência coletiva 15:e009.

Cambon J, Hernangómez D, Belanger C, Possenriede D. 2021. Tidygeocoder: an R package for geocoding. J Open Source Softw. Avaliable from: https://jessecambon.github.io/tidygeocoder/.

CODEPLAN. 2020. Atlas do Distrito Federal – Capítulo 5: Meio Socioeconômico. Companhia de Planejamento do Distrito Federal. Marina Artes Gráficas e Editora: 74:93.

Fantin M, Vizioli SHT, 2021. Caminhos da região central paulista: articulação entre universidade e sociedade para a construção da cidadania e do desenvolvimento sustentável. Volume 1 – Tecnologia, cultura e empreendedorismo. São Carlos: Universidade de São Paulo, Instituto de Arquitetura e Urbanismo. Available from: https://doi.org/10.11606/9786586810233.

Garcia KKS, Miranda CB, Sousa FNEF, 2022. Procedures for health data linkage: applications in health surveillance. Procedimentos para vinculação de dados da saúde: aplicações na vigilância em saúde. Epidemiol Serv Saude 31:e20211272.

Garcia KKS, Xavier DB, Soremekun S, Abrahão AA, Drakeley C, Ramalho WM, Siqueira AM, 2023. Record linkage for malaria deaths data recovery and surveillance in Brazil. Trop Med Infect Dis 8:519.

Google, 2025. Geocoding API: Overview. Available from: https://developers.google.com/maps/documentation/geocoding/overview?hl=pt-br.
IBGE, 2025. Available from: https://www.ibge.gov.br/estatisticas/sociais/populacao/38734-cadastro-nacional-de-enderecos-para-fins-estatisticos.html.
Jaro MA, 1989. Advances in record-linkage methodology as applied to matching the 1985 Census of Tampa, Florida. J Am Stat Assoc, 7 pp.
Katapally TR, Ibrahim ST, 2023. Digital health dashboards for decision-making to enable rapid responses during public health crises: replicable and scalable methodology. JMIR Res Protoc 12:e46810.
Klaus CA, Henry KA, Il'yasova D, 2023. Capturing emergency dispatch address points as geocoding candidates to quantify delimited confidence in residential geolocation. Int J Health Geogr 22:25.
Lustosa JLS, 2017. Geocoding dengue cases for spatial analisys. Conference Paper. Available from: https://www.researchgate.net/publication/320372257_Geocoding_Dengue_Cases_for_Spatial_Analysis
Magalhães M de AFM, Matos VP de, Medronho R de A, 2014. 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 colet. Available from: https://doi.org/10.1590/1414-462X201400020013.
McDonald YJ, Schwind M, Goldberg DW, Lampley A, Wheeler CM, 2017. An analysis of the process and results of manual geocode correction. Geospat Health 12:526.
Miranda ML, Ferranti J, Strauss B, Neelon B, Califf RM, 2013. Geographic health information systems: a platform to support the 'triple aim'. Health Aff (Millwood) 32:1608-15.
MoH, 2024. Ministério da Saúde Brasil. Secretaria de Vigilância em Saúde e Ambiente. Departamento de Ações Estratégicas de Epidemiologia e Vigilância em Saúde e Ambiente. Guia de vigilância em saúde: volume 1. 6th ed. rev. Brasília: Ministério da Saúde. 3 v. 465 pp.
MoH, 2025a. Ministério da Saúde Brasil. Painel de Monitoramento de Arboviroses. Brasília: Ministério da Saúde. Available from: https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/a/aedes-aegypti/monitoramento-das-arboviroses.
MoH, 2025b. Ministério da Saúde Brasil,. Secretaria de Vigilância em Saúde e Ambiente. Departamento de Doenças Transmissíveis. Plano de contingência nacional para dengue, chikungunya e Zika. Brasília: Ministério da Saúde. 102 pp.
Oliveira EA, Martelli Júnior H, Silva ACSE, Martelli DRB, Oliveira MCL, 2020. Science funding crisis in Brazil and COVID-19: deleterious impact on scientific output. An Acad Bras Ciênc. Available from: https://doi.org/10.1590/0001-3765202020200700.
Quintans-Júnior LJ, Gomes FG, 2024. The abyss of research funding in Brazil. EXCLI J 23:1491-1492.
Seixas JBA, Luz KG, Pinto Junior VL, 2024. Atualização clínica sobre diagnóstico, tratamento e prevenção da dengue. Acta Med Port. Available from: https://doi.org/10.20344/amp.20569.
SES-DF. 2024. Secretaria de Saúde do Distrito Federal. Monitoramento dos casos de dengue, febre de chikungunya, doença aguda pelo vírus zika e febre amarela até a Semana Epidemiológica 52 de 2024, no Distrito Federal. Brasília, 1-16.
Silveira IH, Oliveira BFA, Junger WL, 2017. 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. Available from: http://dx.doi.org/10.5123/s1679-49742017000400018.
Skaba DA, 2009. Metodologias de geocodificação dos dados da saúde. Rio de Janeiro: Escola Nacional de Saúde Pública Sergio Arouca, Fundação Oswaldo Cruz. [Thesis] 155 pp.
Skaba DA, Carvalho MS, Barcellos C, Martins PC, Terron SL, 2004. Geoprocessamento dos dados da saúde: o tratamento dos endereços. Cad Saude Publica. Available from: https://doi.org/10.1590/s0102-311x2004000600037.
Vieira FS, 2020. Health financing in Brazil and the goals of the 2030 Agenda: high risk of failure. Rev Saude Publica 54:127.
Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, 2019. Welcome to the Tidyverse. J Open Source Software Avaliable from: https://doi.org/10.21105/joss.01686.
Wickham H. 2021. Mastering Shiny. 1st ed. Sebastopol. O'Reilly Media, Incorporated; 400 pp.
Winkler WE, 1990. String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage. Proceedings of the Section on Survey Research Methods. American Statistical Association; 8.
World Health Organization, 2025. Dengue and severe dengue. Geneva: WHO. Available from: https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue.