Journal of Internet Services and Applications, 2026, 17:1, doi: 10.5753/jisa.2026.6600

MVAR-Geo: A Proposal for Visualizing Georeferenced Association Rules

Anderson C. K. de Menezes [**Universidade Federal Fluminense** | ***ackuhl2@gmail.com***]
Antonio A. de A. Rocha [**Universidade Federal Fluminense** | ***arocha@ic.uff.br***]
Alexandre Plastino [**Universidade Federal Fluminense** | ***plastino@ic.uff.br***]
Verônica Feder Mayer [**Universidade Federal Fluminense** | ***veronicamayer@id.uff.br***]

Institute of Computing, Universidade Federal Fluminense, Av. Gal. Milton Tavares de Souza, s/n, São Domingos, Niterói, RJ, 24210-590, Brazil.

Received: 23 September 2025 • **Accepted:** 11 June 2026 • **Published:** 06 August 2026

Abstract The rapid development of geographic localization techniques has provided the means for generating and increasingly collecting massive georeferenced datasets. Geographic Information Systems (GIS) have been used to process and represent this kind of data visually on maps in the search for a better understanding of the possible relationships between geographic areas and the variables of the represented application domain. The objective of its use is to reduce the time it takes to analyze information, in addition to enabling access to spatial patterns that would be difficult to identify with simplified data listings or graphs (e.g., bars, lines, etc.) that are not very informative in a spatial context. Some modern versions of GIS have sought the support of data mining, among which Association Rule Mining (ARM) is a good option due to its relative simplicity and transparency. The combination of GIS and ARM can support the development of more sophisticated analyses, improve spatial analysis, and enable the identification of patterns of interest and the extraction of valuable knowledge. However, there are still a few works in the literature that propose this type of solution, not exploring the full potential that this approach could offer to visually communicate patterns and trends that are difficult to identify through conventional data analysis methods, such as graphs and tables. To address this gap, the objective of this work is to introduce MVAR-Geo, a method for visualizing association rules in georeferenced data. To demonstrate how this proposal can benefit the analysis of association rules with georeferenced attributes, we apply the method in three case studies to conduct spatial analyses of data from different application domains using levels of geographic granularity in accordance with each context analyzed. The case studies demonstrated that MVAR-Geo significantly enhances the ability to interpret spatial data by enabling the integration of ARM with GIS.The use of interest measures from association rules in geographic visualizations, particularly the lift measure, has proven essential for distinguishing genuinely strong associations from those expected by chance, improving the understanding of how specific geographic areas can influence the chances of occurrence of certain events. The proposed method also favors identifying spatial patterns involving agglomerations, anomalies, and geographic subspaces, which corroborates the potential contribution of this work.

Keywords: Georeferenced Data, Association Rule Mining, Data Visualization, Geovisualization, Geographic Information Systems, Spatial Patterns

1 Introduction

The recent technological advancements and the widespread use of mobile communication have provided the necessary environment for the creation and expansion of large georeferenced datasets. People are increasingly sharing their personal experiences on social networks (such as the places they visit, where they eat, or what they see) and having their mobility records collected by communication and monitoring systems (such as cell towers, presence sensors, and tolls). These facts have led to an unprecedented amount of data being stored, enabling the examination of questions that may help us understand our lives and the evolution of organizations and societies [Ciuccarelli *et al.*, 2014].

Georeferenced databases contain rich geographic information and allow us to address many challenges of modern society, such as traffic engineering, air pollution problems, tourism planning, migration studies, epidemiology, ecology, and disaster response. However, generating knowledge from massive or complex data requires more than traditional analysis methods. Due to this challenge, an effort is underway to expand the use of data mining models to process and identify patterns in massive georeferenced databases [Deng *et al.*, 2019]. Following this tendency, the present study proposes an innovative way of combining Association Rule Mining (ARM) with Geographic Information Systems (GIS) in a visualization platform to address the need for transparent and useful information for managers and policy-makers.

Currently, visualization techniques are the central focus in analyzing large amounts of data, especially when evaluating patterns, since vision is the primary sense of human perception [Madala *et al.*, 2018]. Well-designed visualization and visual analysis systems have successfully allowed users to gain insights, replacing cognition with perception and freeing up limited cognitive or memory resources for higher-level problems [Miranda *et al.*, 2020]. When dealing with georeferenced data, GIS has been used to process and visually represent it, allowing individuals to better understand the possible relationships between geographic areas and the variables of the represented application domain. This kind of solution tends to significantly reduce the time analysts take to process information, enabling access to patterns that would be difficult to identify only with data listings or even simple graphs [Andrienko *et al*., 2017]. GIS often uses frequency measures (sums and counts) in geographic views, seeking to highlight areas with values that differ from measures of central tendency. However, this approach can be misleading with biased or biased data. It also does not explain probabilistic phenomena, nor does it reflect a pattern or trend in the data.

Consequently, some more modern versions of GIS have sought the support of data mining techniques to improve spatial analysis, identifying patterns of interest and extracting helpful knowledge [Barcellos *et al*., 2017]. Association Rule Mining (ARM) is a popular data mining technique due to its relative simplicity and transparency compared to more complex and difficult-to-understand models. This technique aims to find relationships between data items in the same application domain, given a database [Han *et al*., 2022]. In this sense, it would be reasonable to presume a potential interest in discovering associations in georeferenced data.

In this way, the combination of Association Rule Mining with Geographic Information Systems could support the development of more sophisticated analyses. The geographic view of interest measures typical of ARM could be used for an eventual analyst to better understand how certain probabilistic phenomena behave from a geographical perspective. It would be possible to identify differentiated spatial patterns, such as clusters of areas with similar values or places with values that differ from their geographical context. In addition, this approach could be applied to various areas of knowledge as long as georeferenced attributes benefit society. Unfortunately, few works in the literature propose this type of solution, and those that do often fail to explore all the possibilities this approach could offer for visually communicating patterns and trends that are difficult to identify through conventional data visualization methods, such as graphs and tables.

Driven by the need to expand the means of analyzing massive georeferenced databases using ARM techniques combined with visualization solutions, this work has as its main purpose to propose a novel method for visualizing association rules that have georeferenced attributes, which, for reference purposes, will be called MVAR-Geo. The method is based on a specific type of relational association rule called a georeferenced association rule, which is a fresh approach in the field.

To evaluate the MVAR-Geo's applicability and its ability to add value in different scenarios, we apply the method in three case studies to demonstrate how it effectively assists in the perception of the spatial distribution of events, regardless of the application domain and in the revelation of correlations between variables and geographic areas, which would not be easily identified with traditional association rule analysis methods.

The remaining of this paper is structured as follows: Section 2 presents a theoretical framework, briefly summarizing the existing ARM models and the main data geographic visualization techniques currently used. Section 3 presents an analysis of the main works related to the theme of this paper. In Section 4, the methodology of this work, in which a model of association rules applies to georeferenced databases and proposes a new method of visualizing association rules in georeferenced data (MVAR-Geo), is demonstrated. Section 5 presents the results of applying MVAR-Geo to three domains by analyzing geographic visualization examples. Section 6 finally discusses conclusions and future work.

2 Background

This section provides the background necessary for a better understanding of the topic of this work. First, it provides a brief overview of the origin of Association Rule Mining and present some of the main existing models. Next, it demonstrates the advantages of using maps to analyze georeferenced data. Finally, it presents current geographic visualization techniques, citing their proposals and possible limitations.

2.1 Association Rules

Association rule is a widely used pattern type introduced by Agrawal et al. [Agrawal *et al*., 1993]. Initially, it was primarily considered for market basket analysis. Since then, it has been a widely used technique in data analysis due to its simple and easy-to-understand representation.

In the 1990s, barcode technology enabled retail organizations to collect and store large amounts of data related to customer purchases, known as "basket data." An association rule of the type *milk* ⇒ *bread* ("if milk, then bread") indicates that if customers buy milk, they are likely also to buy bread. As these rules are extracted from a set of purchase transactions, they became known as transactional association rules.

Over time, new approaches to association rules have emerged, and one such adaptation is the relational association rule model. Datasets organized into tables in a relational database apply this model. Unlike transactional association rules, which operate on sets of items, relational association rules deal with associations involving attribute-value pairs that occur in the tables of the relational database. The support factor and the confidence factor capture some aspects of the importance of an association rule (transactional or relational).

Relational association rules can be defined as follows. Let $T(A_1, A_2, ..., A_m)$ be a relation (table) formed by m attributes (columns) and n tuples (rows). Each A_j, $1 \leq j \leq m$, represents an attribute of the relation T, and each $t_i(v_{i1}, v_{i2}, ..., v_{im})$, $1 \leq i \leq n$, represents a tuple of the relation T, where v_{ij} is the value of the jth attribute of the tuple t_i. A relational association rules R defined on T is an implication of the form (Eq.1):

$$C_1 \wedge C_2 \wedge ... \wedge C_w \Rightarrow D_1 \wedge D_2 \wedge ... \wedge D_z, \tag{1}$$

where $w \geq 1$, $z \geq 1$, and each C_k, $1 \leq k \leq w$, and each D_l, $1 \leq l \leq z$, are conditions defined on distinct attributes of T of the form $A = v$, where v is a value in the domain of attribute A. We define $C = C_1 \wedge C_2 \wedge ... \wedge C_w$ as the antecedent of R and $D = D_1 \wedge D_2 \wedge ... \wedge D_z$ as its consequent.

The support of R in relation T, denoted by $sup(C \Rightarrow D)$, is given by the proportion of records in T that satisfy all the conditions of C and D. In other words, the percentage of records in T that satisfy the conditions of $C \wedge D$.

The confidence of R in relation T, denoted by $conf(C \Rightarrow D)$, is defined as the percentage of records in T that satisfy the conditions of D among the records that satisfy the conditions of C. It is also possible to establish the concept of confidence as the probability of the occurrence of D given the occurrence of C, that is, the conditional probability $P(D|C)$

In summary, let N be the number of records in T, let N_C be the number of records that satisfy the conditions of C, and let $N_{C \wedge D}$ be the number of records that satisfy the conditions of $C \wedge D$. The support of rule R is given by Eq.2:

$$sup(C \Rightarrow D) = \frac{N_{C \wedge D}}{N} \quad (2)$$

The confidence of the rule is given by Eq.3:

$$conf(C \Rightarrow D) = \frac{N_{C \wedge D}}{N_C} \quad (3)$$

For example, consider a relation (table) containing information about the historical results of students in a public elementary school network, formed by the following attributes: $IDStudent$, $SchoolYear$, $Gender$, and $Result$. Consider the following hypothetical relational association rule with the support of 10% and confidence of 80%:

$$(SchoolYear = \text{“}6^{th}\ Grade\text{”}) \wedge (Gender = \text{“}Male\text{”}) \Rightarrow (Result = \text{“}Approved\text{”})\ sup = 10\%,\ conf = 80\%$$

This indicates that 80% of the 6th-grade male students passed and that this group represents 10% of the total students in the school system. Another possible interpretation is that the chance of male 6th-year students being approved in this education system is 80%.

Due to its flexibility, this model can discover complex patterns in various contexts if the data has a relational structure. Several application domains successfully employ association rules, such as medicine, bioinformatics, recommendation systems, and software engineering, among others [Vougas *et al.*, 2019].

In 1995, Koperski and Han proposed the process of spatial association rule mining, a variant of the relational model, which focuses its approach on discovering patterns of relationships between spatial objects in a geographical space to explain another event that may or may not be spatial [Koperski and Han, 1995]. These relationships are defined based on spatial predicates, which describe spatial relationships or properties between objects or entities in a given space. This methodology can lead to rules of the type:

$$(IsA = \text{“}House\text{”}) \wedge (NextTo = \text{“}Beach\text{”}) \Rightarrow (PriceRange = \text{“}Expensive\text{”})\ sup = 2\%,\ conf = 90\%$$

This example indicates that if the spatial object is a house and it is located near a beach, then there is a 90% chance that it is an expensive house. Additionally, 2% of the total spatial objects in the database are expensive houses located near a beach. The support and confidence definitions in a spatial association rule are similar to the relational model.

This study will explore a specific type of relational association rule, called georeferenced association rules, with the aim of enabling pattern extraction from georeferenced databases. These databases contain one or more attributes whose domain values represent geographic information, either a point or a surface on a map. It will present the definition of this type of rule in Section 4.

2.2 Georeferenced Data Visualization

Data visualization is not just about managing and displaying large amounts of data quickly. It is about revealing details, patterns, and connections between the data, which in turn generates new understandings and knowledge in a short time. As Shakir Khan puts it, it's an easy way for humans to understand data, with the main advantage being the ability to interpret massive and overlapping data instantly and efficiently, represented by various graphic formats. This process facilitates the identification of patterns and trends, making it a crucial tool in the data analysis toolkit.

More specifically, geographic visualization allows for visually exploring georeferenced data, facilitating decision-making processes. Users can change the visual appearance of maps, explore different data layers, and highlight problems in areas of interest. The presentation of information is simplified, making it more accessible and understandable, as it provides a geographical context that helps to understand the spatial distribution of the data [Khan, 2021; Edsall, 2003].

Several proposals exist for the geographic visualization of data. The choice of the most appropriate geographic visualization model depends on the intended objectives, the available data, and the target audience for the visualization. The simplest ones involve heat maps, choropleth maps, and proportional symbol maps. Heat maps use a color scale to represent the value of a metric in different geographic regions. Areas with higher values use more intense colors in their representation, while areas with lower values use soft colors instead. In choropleth maps, geographic areas are predefined and colored or shaded according to the progression of the value of a statistical variable, using a visual order between colors, from softer shades to stronger shades. Proportional symbol maps allow the representation of areas or geographic points by markers whose size and color vary according to the value of a quantitative variable.

Other, more complex forms, such as cartograms, propose to distort the geographic areas of the represented regions, regardless of their original physical size, based on a specific variable, seeking to emphasize their statistical distribution. Multivariate maps propose to display two or more variables in a single visualization by combining different types of thematic maps, allowing the analysis of complex relationships between these variables in a geographic context. Finally, we can mention dasymetric maps, which propose a technical evolution of choropleth maps by using ancillary data to redistribute information only where the phenomenon actually occurs. This approach mitigates the Modifiable Areal Unit Problem (MAUP), a statistical bias defined as the sensitivity of analytical results to the definition of the territorial units for which data are reported [Openshaw, 1984]. MAUP manifests through the scale effect, where data aggregation alters results, and the zoning effect, in which the redistribution of geographic boundaries modifies the perception of the phenomenon, often causing a visual dominance of extensive areas that distorts the interpretation of the variable's true density [Wong, 2004].

3 Related Work

During the development of MVAR-Geo, a literature search was carried out to understand the current research status and establish the original contributions of our proposal. Our examination shows that previous works have not effectively combined ARM with GIS in a choropleth map platform. For instance, in the study by Torrisi *et al*. [2015], publicly available georeferenced images from the photo-sharing platform Flickr are employed as a data source to infer tourist behavior. The aim is to comprehend the concentration of tourists in terms of gender and nationality. This study proposed the application of transactional association rule mining to assist in uncovering relationships between the photographed tourist attractions. The visualization of the association rules was restricted to a graphical representation in the form of a graph to demonstrate the connection between the rules found for each tourist location and how they relate to each other, positioning the rules with the highest confidence in the higher positions of the graph [Torrisi *et al*., 2015].

A similar approach, utilizing the local/individual matrix, applied association rules on a dataset of cellular location data tracked by cellular towers in Shanghai and Shenzhen, China. The objective was to analyze how three distance-based mobility indicators, the number of connection points and frequency of movements, can be combined to obtain insights into mobility patterns in these two cities. The visualization of the rules is limited to 3D bar charts to portray the correlation between mobility indicators based on the support and confidence measures of the rules [Xu *et al*., 2018].

In the work of Schumann and Tominski [2011], ARM is employed to analyze a georeferenced dataset of human health data. They propose a method to represent association rules on maps using arrows, aiming to demonstrate the potential implications of the rules between geographical regions. The width and color saturation of each arrow highlight the support and confidence of each rule, respectively [Schumann and Tominski, 2011]. This visual representation facilitates the identification of epidemiological patterns and potential health risks in specific geographic areas. Crawford *et al*. [2021] propose a methodology to measure intrapersonal travel patterns using georeferenced data from point-to-point sensors such as Bluetooth, road cameras, and electronic toll collections. Their objective is to identify the primary origin and destination points for travelers. They employ transactional association rules to explore the primary routes used by travelers between these points. The strictest rules are visualized on an arrow map, clearly representing prevalent travel patterns [Crawford *et al*., 2021].

Sun and Wang [2017] focuses on utilizing transactional association rules based on geological data and the location of oil and gas wells in western Alberta, Canada. They aim to discover new promising areas for oil and gas extraction. The rules are extracted based on the location data of oil and gas wells and other non-spatial attributes, such as reservoir property parameters and production performance. They propose two geographic methods for visualizing the rules. Based on the points, the first method highlights wells that partially or fully meet the support and confidence criteria of the rule in question on a map. The second method applies spatial interpolation to points that meet the rule to represent the most promising areas on the map [Sun and Wang, 2017].

Based on census data from China, the work of Cui *et al*. [2007] uses association rules to correlate socioeconomic attributes of the Chinese population and highlight, on a map of the Chinese territory, the counties that meet the minimum support and confidence conditions. The generated rules show spatial distribution patterns between census data and other socioeconomic data, making it possible to visualize them in pairs of choropleth maps, where the first represents the distribution of counties that satisfy the minimum support and confidence conditions of the rule antecedent. In contrast, the second represents the counties' distribution that satisfies the rule's conditions, consequent [Cui *et al*., 2007].

Wang *et al*. [2024] introduce the concept of "Eco-location" through a Geo-Eco Knowledge Graph (GEKG), employing adaptive Delaunay networks and the Neo4j database to mine associations that link physical proximity with complex environmental attributes. Despite the innovation in its logical structure, the visualization has a high visual clutter, as the density of nodes and edges overlaid on the map creates a "web-like" structure that hinders the interpretation of large-scale patterns. The main limitation lies in its excessive abstraction, which prioritizes network connectivity over cartographic clarity: the method fails to represent rule strength (such as the lift measure) across continuous surfaces or administrative divisions, making the maps less effective for the immediate identification of spatial anomalies compared to polygon-based or choropleth mapping approaches [Wang *et al*., 2024].

Zhang *et al*. [2024] propose mapping travel patterns in border regions by employing the Apriori algorithm to mine association rules from travelogues, representing the results as a graph network projected onto a geographic map. However, the visualization is criticized for high visual clutter, as the dense network of edges creates a tangle of lines that complicates the distinction between different associations and the perception of route hierarchy. Furthermore, the representation is restricted to point-based nodes, lacking integration with administrative polygons that would enable the visualization of rule strength (such as the lift measure) across various regional scales. This limitation makes the interpretation of statistical metrics less intuitive and forces a reliance on auxiliary tables for a comprehensive understanding of the spatial phenomena [Zhang *et al*., 2024].

Canlon *et al*. [2024] recently introduced star-shaped glyphs for visualizing multidimensional association rules. In their method, each point of the star represents an attribute, with its size and color indicating its relevance to a rule within a specific geographic area. Inside the glyph, multicolored arrows depict key attribute relationships, flowing from antecedent to consequent. These glyphs are then plotted centrally on maps, allowing analysts to compare different regions effectively. To validate their approach, the authors used a dataset of Dutch traffic accidents, specifically focusing on bicycle-involved incidents given the country's prominent cycling culture. During testing with policymakers, the main feedback revealed difficulty interpreting the depicted relationships. This led to a suggestion for an interactive feature that would display textual explanations when a relationship is clicked [Canlon *et al.*, 2024].

The visualization techniques used in these studies have important limitations. In cases where studies propose only basic graphics, the need for a spatial view ends up restricting the analysis possibilities. In other publications, the proposed maps were limited to highlighting the locations that meet the minimum support and confidence patterns of one or more rules without any concern for highlighting where they were more relevant and interesting. In those with this concern, the proposed visual approach ended up restricting the number of rules represented. They also needed to demonstrate the existence of interactive resources, limiting the user experience and restricting the possibility of analysis. Table 1 summarizes the approaches of the cited works, their proposed ARM visualization methods, and their main limitations.

These limitations ultimately hinder the interpretation of data and the ability to conduct comprehensive analyses. However, we firmly believe that there are more effective ways to visually communicate the results of association rule mining in georeferenced data. This belief underscores the need for a new and more efficient methodology, the MVAR-Geo, which is introduced in the following section.

4 Methodology

This section presents the methodology of this work. First, it proposes a new Association Rule Mining model applicable to georeferenced databases. Then, it presents the MVAR-Geo, a new method for visualizing association rules with georeferenced attributes. It delimits its scope of use and addresses the key elements essential for it to fulfill its purpose. Finally, an example of its application is presented.

4.1 Georeferenced Association Rules

In this subsection, a specific type of relational association rule mining, called a georeferenced association rule, is defined to enable pattern extraction from georeferenced databases. These databases contain one or more attributes whose domain values represent geographic information. An attribute A_g is said to be georeferenced when the values of its domain represent geographical information, either a point or a surface on a map. Although A_g can be treated as a categorical reference during rule calculation, it must, obligatorily, be associated with an attribute containing the equivalent real geographic data. This implies that databases containing mere categorical references to spatial surfaces must be integrated with real geographic data (such as a shapefile) so that the rules are considered truly georeferenced, allowing their visualization on a map.

A relation T_g is considered georeferenced when at least one of its attributes is georeferenced. A relational association rule R_g defined on T_g is said to be georeferenced when at least one of the conditions that occur in the antecedent or consequent is defined over a georeferenced attribute of T_g. In summary, given the antecedent C and the consequent D, we say that a relational association rule R_g is georeferenced when at least one of the conditions that make up C and D is defined on a georeferenced attribute. Since an R_g is a relational association rule, its support and confidence measures follow the same definitions presented for relational association rules, as presented below.

The support of R_g in the relation T_g, represented by $sup(R_g)$, is given by the proportion of records in T_g that satisfy all the conditions of C and D, that is, the percentage of records in T_g that satisfy the conditions of $C \wedge D$. The confidence of R_g in the relation T_g, represented by $conf(R_g)$, is defined by the percentage of records in T_g that satisfy the conditions of D among the records that satisfy the conditions of C.

As with relational association rules, we can establish confidence as the probability of occurrence of D given the occurrence of C, that is, the conditional probability $P(D|C)$. Therefore, let N be the number of records in the relation T_g, let N_C be the number of records that satisfy the conditions of C, and let $N_{C \wedge D}$ be the number of records that satisfy the conditions of $C \wedge D$, the support of the rule R_g is given by Eq.4:

$$sup(R_g) = \frac{N_{C \wedge D}}{N}; \tag{4}$$

the confidence of the rule is given by Eq.5:

$$conf(R_g) = \frac{N_{C \wedge D}}{N_C}. \tag{5}$$

As an example, consider the various public elementary school systems in Brazil, which are identified and organized by the Ministry of Education and Culture of Brazil by State. In this case, the state would be a georeferenced attribute since a surface on the map could represent it. To identify each school system, the relational database in question has the georeferenced attribute state, which refers to the state to which the student belongs. Then consider the following hypothetical georeferenced association rule with support of 4% and confidence of 95%.

$$(SchoolYear = \text{“}6^{th}\ Grade\text{”}) \wedge (Gender = \text{“}Male\text{”}) \wedge$$
$$(State = \text{“}SC\text{”}) \Rightarrow (Result = \text{“}Approved\text{”})$$
$$sup = 4\%,\ conf = 95\%$$

This rule indicates that 4% of elementary school students in Brazil are in the 6th grade, are male, are from Santa Catarina (SC), and were approved. Additionally, it indicates that 95% of 6th-year male students from Santa Catarina were approved. Alternatively, the chances of approval for a 6th-year male student from Santa Catarina are 95%. In order to clarify the understanding of the georeferenced association rules proposed in this work, in Table 2, we provide a comparison of the main types of association rules, focusing on their characteristics and the type of information they model.

In addition to support and confidence, other proposals for measures of interest for association rules emerged over time. Finding a comprehensive compendium at Hahsler [2015] is

Table 1. Comparative Table of Association Rule Visualization Approaches in Literature.

Authors (Year)	Visual Technique	Spatial	Clear	Scalable	Low Cog.	Primary Limitation
Torrisi et al. (2015)	Graphs	✗	✓	✗	✓	Abstract nodes; no map projection.
Xu et al. (2018)	3D Bar Charts	✗	✓	✗	✓	Restricted to discrete correlations.
Sch. & Tom. (2011)	Arrow Maps	✓	✗	✗	✓	Limited rule density/overlap.
Sun & Wang(2017)	Point Maps	✓	✓	✗	✓	Lacks regional/admin context.
Cui et al. (2007)	Dual Choropleths Maps	✓	✓	✓	✗	Disjointed antecedent/consequent.
Wang et al. (2024)	Geo-Eco Knowl. Graph	✓	✗	✓	✗	High network abstraction.
Zhang et al. (2024)	Graph NET on Maps	✓	✗	✗	✓	Visual clutter (edge crossing).
Canlon et al.(2024)	Star Glyphs on Maps	✓	✓	✗	✗	High cognitive load; Multidimensional relationships.

Columns Legend: Spatial = Geospatial Visualization; Clear = Visual Clarity; Scalable = Data Scalability; Low Cog. = Low Cognitive Load.
Indicators Legend: ✓= Strength/Feature Present; ✗= Weakness/Limitation.

Table 2. Comparative Table of ARM Types (Transactional, Relational, Spatial, and Georeferenced).

Rule Type	Variables Involved	Distinctive Characteristic	Example of a Rule
Transactional (Classical)	Items in Transactions	It models the co-occurrence of discrete items in a subset of transactions or events. It ignores time and space.	Domain Context: Supermarket If (buy Milk and Bread) Then (There is a 80% increased chance of buy Butter)
Relational	Attributes and Tuples of the Entire Database	It models the co-occurrence of tuples and their descriptive attributes (e.g., age, income). It involves relationships across multiple attributes.	Domain Context: Medicine/Health If (The patient has a history of disease Y) AND (Medication Z was prescribed) Then (There is a 50% increased chance of an allergic reaction)
Spatial	Non-Geographic Objects with Spatial Relationships	It involves the co-occurrence of events or objects based on their geometric or topological relationships (e.g., proximity, intersection), that have a location in space (not necessarily geographic).	Domain Context: Public Security If (Crime Occurrence A is Near Bar) Then (There is a 20% increased chance of Crime Occurrence B)
Georeferenced (MVAR-Geo)	Geographic (Area/Point) and Non-Geographic Attributes	It models the influence of geographic attributes (locations, areas) on the occurrence of events or patterns, focusing on visualizing this influence.	Domain Context: Consumption Patterns If (customers reside in City Z) Then (They reduce the chances of beer consumption by 15%)

possible. Among these measures, one of extreme importance and applicability is the "lift". Originally called "interest" by Brin at al., this indicator reveals the dependence between the antecedent and the consequent of the rule. It aims to measure how often the presence of the antecedent increases or decreases the probability of the consequent occurring. Values above 1 indicate that the occurrence of the antecedent increases the chances of the consequent occurring, values below 1 (negative correlation) indicate that the occurrence of the antecedent decreases the chances of the consequent occurring, and values close to 1 indicate that the antecedent and consequent are independent of each other [Brin *et al.*, 1997].

For a relational association rule $C \Rightarrow D$ defined in a relationship T, the lift is defined by the ratio of the rule's confidence to the support of the consequent (or the probability of the consequent occurring). In this way, the lift reflects how much the antecedent increases or decreases the probability of the consequent occurring. Considering that the support of D in relation T, represented by $sup(D)$, is given by the percentage of records in T that satisfy all the conditions of D, the lift measure is defined by Eq.6:

$$lift(C \Rightarrow D) = \frac{conf(C \Rightarrow D)}{sup(D)} \quad (6)$$

Similarly, for a georeferenced association rule R_g, the support of D in relation T_g, represented by $sup(D)$, is given by the percentage of records in T_g that satisfy all the conditions of D. Therefore, the lift measure is defined by Eq.7:

$$lift(R_g) = \frac{conf(R_g)}{sup(D)} \quad (7)$$

Let us use an example to make the concept of lift clearer. Let us go back to the example that provides the information that 95% of the 6th grade students, male and from Santa Catarina, are approved. Assuming that the overall approval rate for an elementary school in Brazil is 73% ($sup(D)$), the lift for this example is approximately 1.3 (95% / 73%). This indicates that a 6th student, male and from Santa Catarina, is 1.3 times more likely to be approved than any randomly chosen elementary school student in Brazil.

In the case of georeferenced data, the lift is a crucial measure for evaluating the dependence between geographical areas or points and the other conditions that make up a georeferenced association rules, providing additional information beyond support and confidence. The use of this measure of interest facilitates the identification of more relevant rules with greater descriptive power, which can be crucial for making strategic decisions in various areas. For these reasons, the lift was the primary measure of interest used in this paper's case studies.

4.2 Method for Visualizing Georeferenced Association Rules (MVAR-Geo)

In this subsection, the MVAR-Geo is proposed to contribute to the advancement of knowledge and expand the means of analyzing georeferenced association rules. It aims to establish a methodology to overcome the visual limitations identified in related works. It offers a broad scope and enables the identification of spatial patterns, thus benefiting the data analysis process.

4.2.1 Scope Delimitation

In the literature, we can find two distinct approaches for visualizing association rules in georeferenced data. The first seeks mobility patterns based on relationships between different areas or geographic points [Torrisi *et al.*, 2015; Xu *et al.*, 2018; Schumann and Tominski, 2011; Crawford *et al.*, 2021]. The second approach aims to explain the occurrence of specific events based on relationships between areas or geographic points and other types of variables [Cui *et al.*, 2007; Sun and Wang, 2017; Canlon *et al.*, 2024].

The first has a more limited applicability due to the tendency of mobility pattern studies to focus on specific and contextualized objectives. Additionally, the studies found that employing this approach required extensive data manipulation to allow for the extraction of association rules, making its application quite complex and, in some cases, computationally expensive. In contrast, the second approach, on which MVAR-Geo is based, presents a more direct and comprehensive structure, encompassing virtually any application domain using georeferenced data.

4.2.2 MVAR-Geo Illustration

To validate the MVAR-Geo approach, a functional prototype was developed in Python, supported by the libraries *folium* for map rendering, *geopandas* for calculating rules with georeferenced data, and *streamlit* for creating dashboards, among others. The prototype construction went through several cycles of development, evaluation, and improvement until reaching the version proposed in this work. The evaluations were carried out by computing and management experts through applicability tests on a case study basis. At each iteration, they suggested improvements, leading to successive refinements until the prototype reached the final version. It offers the necessary interaction tools for a clear and precise understanding of the visualized association rules, facilitating the interpretation of the results and decision-making based on data analysis.

Figure 1 shows a visualization of the functional prototype, illustrating how MVAR-Geo can be applied to identify spatial patterns and significant relationships in geographic datasets. The prototype has a dashboard format, where the parameterizations and maps share the same interface, with the geographic visualization configuration parameters positioned in a sidebar and the maps updated in real time as the parameters are adjusted.

In the sidebar of Figure 1, it is possible to define to generate one or two side-by-side maps as needed. The "Map(a)" and "Map(b)" buttons are available, Figure 1 (items 1a, 1b respectively), initially unchecked. When clicked, the configuration parameters of the corresponding map are enabled. If the user marks only one map after configuration, it will occupy the entire central space. To generate the geographic visualizations, the following parameters highlighted in the sidebar of Figure 1 must also be defined:

- **Geographic Division (items 2a, 2b)** → Defines the granularity level of the map (Region, State, City, etc.);
- **Measure of Interest (items 3a, 3b)** → Which measure of interest will represent the color scale of the geographic divisions, being support, confidence, or lift;
- **Rule Antecedent (items 4a, 4b)** → The rule antecedent, which, by default, is already predefined by each geographic location represented on the map but can be complemented with values of other variables;
- **Rule Consequent (items 5a, 5b)** → This is the ultimate goal of your analysis. The value of an attribute of the studied data domain defines the rule consequent.

A geographic filter, Figure 1 (items 6a, 6b), can be defined in the sidebar to focus the analysis on a specific area. If the Southeast Region of Brazil is selected, the prototype automatically adjusts the map granularity to the state level. The user can then increase the geographic granularity as needed.

Figure 1 was constructed based on a case study of the 2018 and 2022 presidential elections, which will be detailed in Section 5.2, and where we demonstrate the visualization resources proposed by MVAR-Geo. The measure of interest represented was the lift, which is possible to observe in the map titles (items 7a, 7b) and legends (items 8a, 8b). A geographic filter was applied in the Map **(a)**, selecting the Southeast Region, detailed by state. In Map **(b)**, which serves as a detail of a state in the Map **(a)**, the State of Rio de Janeiro was filtered, being divided into Microregions. In both maps, the rule antecedent was simply the locality itself, and the consequent received the value referring to the winning political current (right or left) in the locality's ballot boxes. In this case, "right" was chosen for both Map **(a)** and Map **(b)**.

When hovering over an area on one of the maps in Figure 1, a text box appears (item 9a). In the Map **(a)**, the calculations of the rule's measures of interest (support, confidence, and lift) for the State of Rio de Janeiro are displayed. The color scale ranges for the geographic areas in Map **(a)** and Map **(b)** are synchronized (items 10a, 10b). An adjustment button can change the terrain visualization layer (items 11a, 11b). Zoom buttons (items 12a, 12b) allow zooming in and out of geographic areas to explore details or get an overview.

Finally, the auxiliary bar charts, Figure 1 (items 13a, 13b), below each map vary in hue and size according to the value of the measure of interest represented, in this case, the lift. The bars appear from left to right in descending order of value. A dividing line represents a neutral or median value, serving as a reference for analysis. Their purpose is to complement the spatial view provided by the maps with quantitative data, expanding the possibilities for data analysis. In this example, the use of lift sought to highlight which areas increase or decrease the chances of the rule's consequent occurrence.

To emphasize the differentiator of our proposal, we constructed Figure 2 based on the data from the Map **(b)** of Figure 1. Figure 2 compares three possible ways of presenting data to analyze georeferenced association rules. In Figure 2 **(a)**, the data is presented in tabular format, in descending order of the lift value. In this case, in a quick analysis, it is only possible to highlight the localities with the highest or lowest numerical values, restricting the analysis possibilities. The tabular format also makes it difficult to visualize the magnitude of the data and the differences. Finally, it would require much effort to reveal the distribution profile of the lift values in this example.

In Figure 2 **(b)**, the proposed bar chart visualization offers a visual evolution in data presentation. The proportionality of the bar heights to the numerical values favors a quick comparison and perception of the relative magnitude of the lift values without the need to analyze the numbers individually. The red line in the chart provides information about the locations above or below a lift value of 1, considered neutral for this measure of interest. The chart also reveals a distribution profile of the lift values through the color bands of the bars. However, the practical use of charts may become unfeasible in cases where there is a large number of locations. Additionally, when working with geographically referenced data, analysts are interested in identifying spatial patterns not provided by graphs.

In this sense, in Figure 2 **(c)**, the geographic visualization provides a unique and immediate visual representation of the distribution patterns and spatial correlation of the lift that cannot be observed in graphs. It is also possible to observe, through the color pattern, the locations that stand out the most for increasing or decreasing the chances of the target occurrence, although without the same quantitative precision as the graph. In particular, in Figure 2 **(c)**, it is possible to verify the existence of three clusters of locations with the lift in the last value range, standing out from the other areas of the map. In the case of opting for a more in-depth analysis, the visualization of the map could help in identifying a correlation between the rule and some socioeconomic characteristic of the location, such as the HDI, for example.

Despite the advantages of maps, in the context of spatial pattern analysis, graphs are still useful for representing quantitative data. For this reason, we opted for the combination of both techniques in the development of MVAR-Geo. In summary, the proposed method seeks to provide a comprehensive and effective way of visually communicating the results of the analysis of association rules in georeferenced data, allowing for rapid conclusions without the need for a more in-depth study of the subject in question. The functionalities proposed in the method assist in data exploration, enabling qualitative insights into patterns and trends, leading to a better understanding of the behavior of the consequence in the studied areas, supporting strategies, and future decision-making.

4.2.3 Critical Elements for Method Construction

As previously mentioned in this paper, the implementation of a prototype for visualizing georeferenced association rules that was simultaneously fast, functional, interactive, and provided information relevant to the analysis context underwent a meticulous process of development and refinement. Several cycles of development, testing, and refinement were necessary before reaching its final version. As a result of this effort, the level of maturity achieved allowed us to conclude that some key elements used have general applicability when the target is to search for association relationships between geographic areas or points with other types of variables. These key elements are techniques and concepts currently available in the literature, whose adoption was essential to ensure

The rules for Figure 1 were as follows:

Rules for Figure 1 **(a):** $(state = \text{“}x\text{”}) \Rightarrow (winner = \text{“}right\text{”})$

Rules for Figure 1 **(b)**: $(microregion = \text{“}x\text{”}) \Rightarrow (winner = \text{“}right\text{”})$

Where, in **(a)**, “x” refers to each state represented on the map and, in **(b)**, “x” refers to each Microregion.

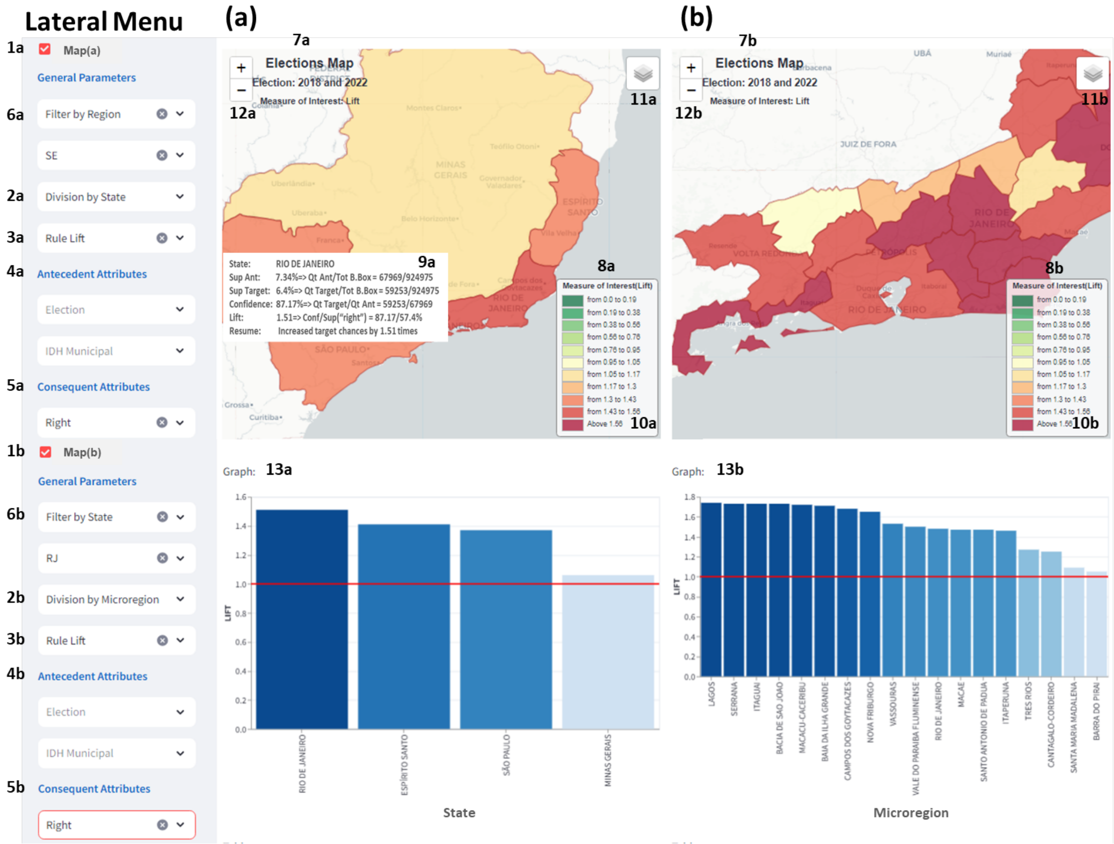

Figure 1. Example of Using MVAR-Geo: **(a)** Southeast Region, divided by State; **(b)** State of Rio de Janeiro, divided by Microregions.

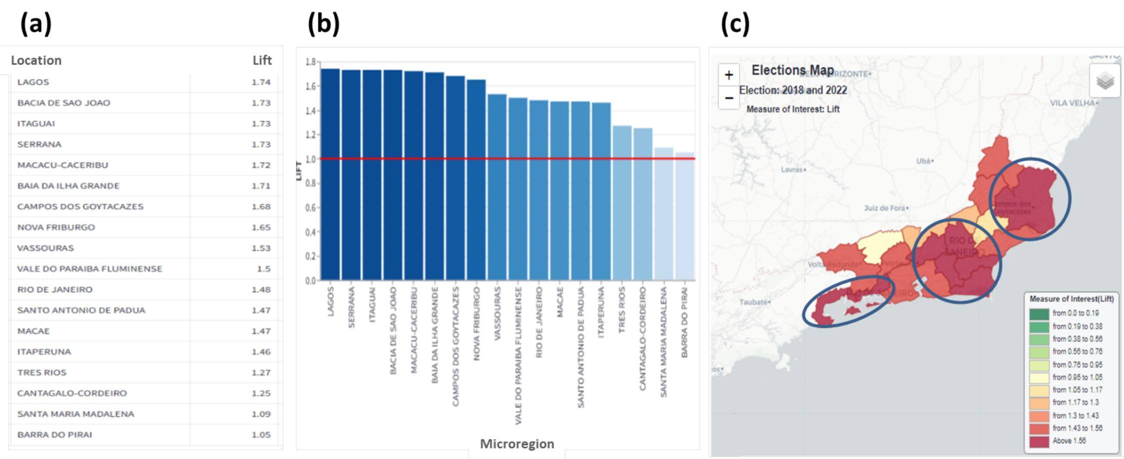

Figure 2. Comparison of Three Methods for Georeferenced Association Rules Analysis: **(a)** Table; **(b)** Bar Chart; **(c)** Geographic Visualization.

that the method could efficiently transmit relevant information from association rules, enriching the process of analyzing geo-referenced data. Therefore, for the development of MVAR-Geo, we considered the following key elements:

- **Visualization Technique:** Due to the georeferenced nature of the data, geographic visualization was the primary visualization technique employed by MVAR-Geo. It proved to be the most suitable as it provides a spatial perspective that aids in understanding the geographical distribution of data, which can be crucial for analysis and the correct interpretation of association patterns in a spatial context.

- **Type of Geovisualization:** The choice of map type depends primarily on the analysis context and the objectives set. MVAR-Geo's geographic visualizations are based on choropleth maps due to their comprehensiveness and simplicity. Choropleth maps can represent virtually all databases with georeferenced attributes.
- **Visualized Information:** We designed MVAR-Geo to allow the visualization of information related to the measures of interest of georeferenced association rules, such as support, confidence, and lift. Specifically, the lift measure received greater focus in our proposal due to its ability to reveal the degree of dependence between areas or geographic points and the other variables that make up the rule. For instance, it can quantify how much a location influences the probability of the occurrence of a specific variable or event appearing in the consequent.
- **Visualization of Rule Relevance:** Visualizing rule relevance allows for a better understanding of associations' spatial context and identifying spatial patterns that would be very difficult to detect in other visualization methods. This information is crucial for conducting in-depth data investigations and supporting strategic decision-making. MVAR-Geo proposes that the visualization of rule relevance in each geographic area should be demonstrated based on its color shade. A diverging color palette is suggested for visualizing rule strength, with its intensity scaled by the magnitude of the represented measure of interest (i.e., softer colors for weaker rules and more intense colors for stronger ones). It is recommended to follow the color palette standards proposed in Harrower and Brewer [2003]. With this visual representation strategy, it is possible to verify how the rules vary spatially and where they are most relevant.
- **Multiple Levels of Spatial Aggregation:** Aggregating or detailing geographic areas allows the choice of the most suitable granularity level for analysis needs. It can also compensate for the lack of internal variations of choropleth maps, helping to minimize the MAUP problem, mentioned in section 2.2. Together with the use of filters, it allows focusing on specific areas and visualizing important details in the data, favoring the discovery of hidden patterns.
- **Multiple Map Option:** Visualizing multiple maps enables the identification of more spatial patterns, correlations, and changes over time, among other insights. This allows for more complex visual analyses, enriching this process. In this case, when the measure of interest to be represented is the same, the color value ranges visualized on the maps should be synchronized to allow a faithful comparison.
- **Combination of Maps with Auxiliary Graphs:** Integrating maps with auxiliary graphs is a powerful technique that enriches spatial analysis. Graphs can detail quantitative information, facilitate the comparison of numerical values, corroborate the patterns and trends observed in the geographic visualization, and increase the quality and reliability of the analyses.
- **Interactivity:** This is an essential resource in map visualizations, making them more engaging, informative, and valuable. It allows the exploration of association rules in a dynamic way, adjusting visualizations, overlaying layers, and obtaining detailed information about specific areas through the use of filters, zoom buttons, and popups, among other things.

It is important to emphasize that, although some of the critical elements mentioned above — such as choropleth maps, the use of interest measures (*e.g.*, lift), or interactive interfaces — have been addressed or used in isolation in previous works, MVAR-Geo brings a synergistic combination of all these components into a single workflow, creating a unique differential for georeferenced data analysis:

- **Relevance Visualization:** The direct translation of interest measures into color symbology (choropleth gradient) eliminates the abstraction inherent in traditional tables.
- **Scalar Flexibility:** The ability to handle multiple levels of aggregation and multiple simultaneous maps allows for a holistic view of the territory that isolated tools cannot achieve.
- **Perceptual Reasoning:** By uniting interactivity and geographic context, the method does not merely present data, but enhances the user's cognitive capacity to intuitively identify spatial clusters and anomalies.

Ultimately, MVAR-Geo surpasses previous approaches by offering a complete analytical ecosystem. By integrating the rigor of association rule mining algorithms, the effectiveness of thematic mapping, and the agility of web interactivity, this proposal establishes a new standard for geographic pattern analysis, proving that the totality of its solution is significantly more powerful than the sum of its individual parts. Regardless of the data source, programming language, or GIS tool used, the diagram presented in Figure 3 summarizes the steps required to reproduce MVAR-Geo with results similar to those obtained in this work.

5 Case Studies

To demonstrate MVAR-Geo's applicability, we conducted three distinct case studies to evaluate its analytical capabilities in complex scenarios. First, we examine the Coronavirus (COVID-19) pandemic in Brazil. The pandemic presented immense challenges to healthcare systems and society globally. This study aims to show how MVAR-Geo can assist health authorities in managing epidemiological events by uncovering patterns of dissemination and lethality, in addition to illustrating its versatility in epidemiological analysis through three approaches:

- Comparison of distinct events within the same time frame, correlating infection and mortality rates to identify disparities in clinical response;
- Use of different levels of spatial granularity (States, Mesoregions, and Microregions), demonstrating how geographic detail allows for the observation of specific phenomena that would be obscured by larger aggregations;

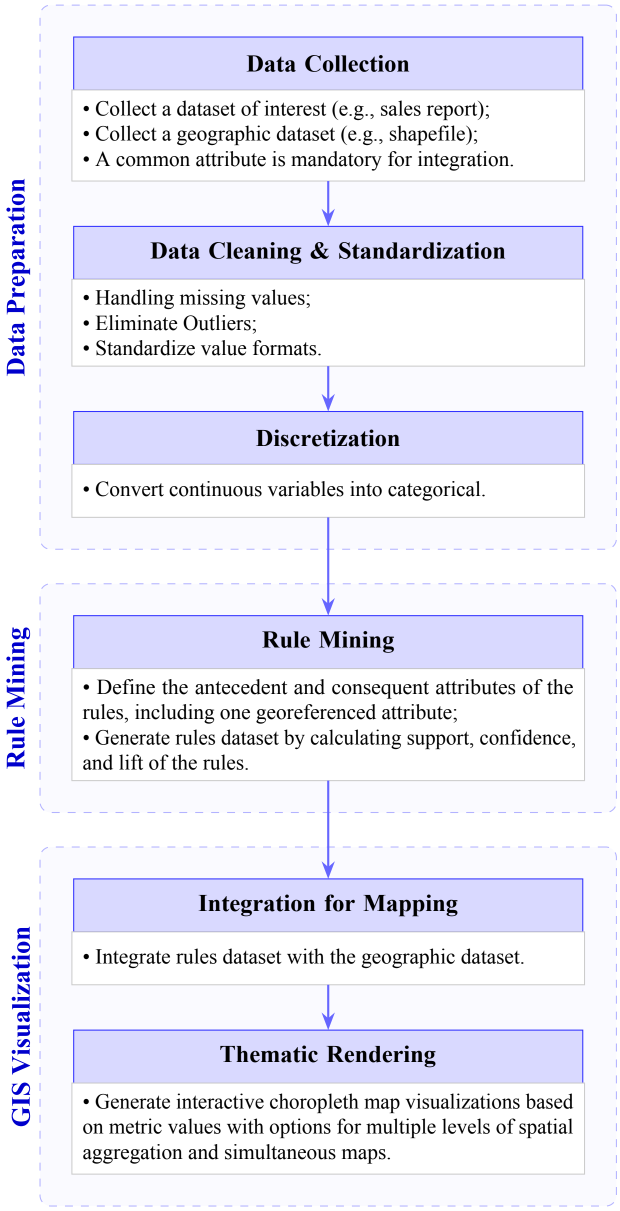

Figure 3. Workflow for the MVAR-Geo Method Reproduction, detailing the phases from Data Preparation to GIS Visualization.

- Comparative temporal analysis between the years 2021 and 2022, highlighting changes in the dynamics of disease dissemination and lethality.

Next, we analyze the 2018 and 2022 Brazilian Presidential Elections. The 2022 elections were characterized by intense political polarization, yet a remarkably narrow margin separated the two primary candidates – one from the right and the other from the left – culminating in the latter's victory in the second round [Interian and Rodrigues, 2023]. This study seeks to illustrate how MVAR-Geo can contribute to the analysis of changes in political behavior and identify atypical patterns, enabling an understanding of Brazil's recent electoral landscape through three approaches:

- A comparison between distinct electoral events, mapping the evolution of voting and the migration of party preference between 2018 and 2022;
- The detection of anomalies and local electoral phenomena through the geographic detailing of specific regions, allowing the visualization of voting niches that deviate from the state trend;
- The impact of external variables (socioeconomic or demographic) on the 2022 result, showing how MVAR-Geo integrates non-geographic attributes to enrich the extracted association rules.

Finally, our third case study explores "Smart Tourism", leveraging cell phone connection records from Rio de Janeiro. The development of Information and Communication Technology has enabled the rise of Smart Cities, which apply technological solutions to urbanization challenges. Within this context, smart tourism has recently emerged as a key component, focusing on innovative tools and technologies for tourism planning [Tripathy *et al.*, 2018]. This proposed study aims to demonstrate how MVAR-Geo can be utilized within smart tourism by using mobile connection data in Rio de Janeiro to identify patterns and trends in relevant sector data, through the following topics:

- The study of co-occurrence, analyzing the simultaneous presence of two distinct groups of foreigners in the same geographic region to understand destination affinities;
- The analysis of temporal persistence, observing the behavior of these same groups in different periods to identify seasonal trends and patterns of stay.

These studies aim to validate MVAR-Geo's utility with tangible results, facilitating quick insights. Each visualization addresses simulated stakeholder questions concerning spatial events, demonstrating the method's effectiveness in converting georeferenced databases into strategic knowledge. For consistency and leveraging its inherent properties, the lift measure was uniformly chosen as the measure of interest across all examples. Table 3 summarizes some characteristics of the three case studies that demonstrate the flexibility and potential of MVAR-Geo in different domains, spatial scales, and volumes of records analyzed.

5.1 COVID-19 in Brazil

Political polarization, misinformation, and a lack of national coordination hampered Brazil's management of the pandemic.

Table 3. Comparative Summary of MVAR-Geo Case Studies: Datasets, Granularity, and Records.

Case Study	Dataset (Domain)	Spatial Granularity	Analyzed Records (Pre-discretization)
COVID-19 Pandemic	Official Epidemiological Data (2020-2022)	**Multiscalar:** States (UF), Mesoregions, and Microregions	Universe of cases and deaths reported per municipality and date (2020-2022): Approximately 5 million records.
Presidential Elections	Brazilian Official Electoral Data Repository (2018 and 2022)	**Multiscalar:** States (UF), Mesoregions, and Microregions	For a total of 924,975 voting machines: 453,829 in 2018 and 471,146 in 2022: Approximately 12 million records.
Smart Tourism	Mobile Phone Connection Records (CDR) - Rio de Janeiro	**Specific:** Administrative Regions (AR) of the city of Rio de Janeiro	Approximately 2.67 million connections from foreigners occurred between May 2020 and October 2021.

The vaccination campaign itself was affected by the spread of fake news on digital platforms, discouraging a significant portion of the Brazilian population from adhering to social distancing and vaccination campaigns, thus exacerbating the transmission and mortality rates of the disease [Galhardi *et al.*, 2022]. These factors contributed significantly to the country having one of the highest COVID-19 mortality rates in the world [Brandão *et al.*, 2023]. Given the importance of the subject, we propose this case study to demonstrate how MVAR-Geo can assist health authorities in managing epidemiological events by discovering patterns of dissemination and lethality. This information is crucial for developing intervention strategies and resource allocation.

5.1.1 Data Preparation

The primary data sources for this study were databases provided by the Brazilian Ministry of Health (MS) for COVID-19 infection and mortality data and the Brazilian Institute of Geography and Statistics (IBGE) for population data from the 2022 Brazilian Census. Both sources are publicly available [Brasil, 2023; Cidades, 2023]. These are *.csv* files with semi-annual records of cases and deaths, organized by date and city. This study covered each of the 5570 Brazilian municipalities during the period from March 2020 to December 2022.

The Ministry of Health's public database, which provides open-access information on COVID-19 cases and deaths, is updated daily through official reports from the 27 Brazilian State Health Secretariats. This data integration process is inherently dynamic and complex, subject to fluctuations due to municipal infrastructure constraints and administrative rectifications, such as the distinction between the location of notification and the patient's actual place of residence. For this study, the dataset was processed in CSV format, comprising six consolidating files for the 2020-2022 triennium, totaling approximately 485 MB of data and 5 million records, with 0.2% inconsistencies in the correct identification of the municipality, which were disregarded from the case study.

The data was aggregated by summing the number of cases and deaths per city and year, covering the disease cycle [World Health Organization, 2021]. This methodology allowed us to establish a statistical relationship between the records of cases and deaths. Despite the annual data aggregation, 58 residual inconsistencies were identified and corrected—representing 0.024% of deaths and 0.00011% of cases—consisting of 22 negative death records adjusted in the preceding year and 36 instances where mortality exceeded infection due to the temporal lag between disease contraction and death. These statistical adjustments, performed by transferring records to the correct period to ensure a more reliable case fatality rate, had a negligible quantitative impact, thereby preserving the validity and precision of the final analytical results. Thus, it was possible to generate the following metrics for each year and city:

- **deathcase** ⇒ Ratio between the number of deaths and the number of cases;
- **deathpop** ⇒ Ratio between the number of deaths and the local population;
- **casepop** ⇒ Ratio between the number of cases and the local population.

To generate association rules, the quantitative metrics 'deathcase,' 'deathpop,' and 'casepop' were discretized into the categorical attributes 'rtdeathcase,' 'rtdeathpop,' and 'rtcasepop,' respectively. The latter aggregates the statistical distribution of the metrics' values into five value ranges: 'Very High,' 'High,' 'Medium,' 'Low,' and 'Very Low.' Each value range represents approximately 20% of the records for each metric. These categorized attributes served as a reference for generating the measures of interest (support, confidence, and lift) of georeferenced association rules, whose calculation followed the methodology defined in Subsection 4.1. Table 4 presents an example of how we organized the data for Brazil's largest city (São Paulo).

For calculating and representing the measures of interest on the maps, we used spatial aggregations representing groups of cities, which we divided into the following four granularity levels:

- **Microrregion** ⇒ A grouping of contiguous cities with similar spatial organization and economic characteristics;
- **Mesoregion** ⇒ A grouping of Microregions, considering similarities in social and economic areas;
- **State** ⇒ Brazilian Federative Unit with a certain degree of autonomy, grouping Mesoregions;
- **Region** ⇒ Macro political-administrative region of Brazil, grouping Federative Units.

Additionally, we used geographic coordinate files containing the contours of each geographic area at each spatial aggregation level as support. These files were correlated with

Table 4. Organization of COVID-19 Data

year	city	cases	deaths	deathcase	deathpop	casepop	rtdeathcase	rtdeathpop	rtcasepop
2020	São Paulo	401718	15679	3.9	0.13	3.5	Very High	High	Low
2021	São Paulo	576200	23882	4.14	0.2	5.03	Very High	Very High	Medium
2022	São Paulo	161204	4785	2.96	0.04	1.4	Very High	Medium	Very Low

the georeferenced attributes of the study database, enabling the rendering of geovisualizations.

5.1.2 Results and Discussion

"What was the situation of each Brazilian State regarding the proportion of cities with high rates of infection and mortality from COVID-19 in 2021, and what was their spatial pattern of occurrence?"

Figure 4 was generated to answer this question, featuring two geographic visualizations (Map **(a)** and Map **(b)**). In both, we defined the rule antecedent by combining the georeferenced attribute state with the attribute year equal to "2021". The consequent, in Map **(a)**, is determined by the attribute rtcasepop, which brings the annual infection rate value band, due to COVID-19, of the city's population, with the value equal to "very high." In Map **(b)**, it is determined by the attribute rtdeathpop, which brings the annual mortality rate value band, due to COVID-19, of the city's population, with the value equal to "very high."

In Figure 4, it is visible that the colors of the geographic areas have a greater tendency towards red in Map **(b)** than in Map **(a)**. The visualization reveals a geographic expansion of the high-risk areas across the national territory, suggesting a systemic spatial worsening of the pandemic. The auxiliary graphs confirmed this pattern, where it is possible to verify that the lift values are, on average, higher in Map **(b)** than in Map **(a)**. Thus, considering the behavior of these indicators throughout the study period, it is possible to conclude that, in 2021, the chances of mortality were proportionally higher than the chances of infection by the disease, pointing to a high lethality rate of the disease in 2021.

As raised in the work of Freitas *et al.* [2023], the health infrastructure was unable to adequately treat all infected individuals, showing signs of collapse due to the high contamination rate despite the beginning of the population vaccination process in the same year [Freitas *et al.*, 2023]. The map highlights that this crisis was not spatially isolated; the State of Rondônia, for instance, appears as a critical epicenter in the north. While in Map **(a)** this State presented a lift of 3.6, in Map **(b)**, the lift reached 4.8. The visual contrast between these maps allows us to identify how the mortality risk "overflowed" geographically beyond the infection hotspots, revealing the localized collapse of the health system in the Western Amazon region.

It is interesting to highlight the advantage of using lift in maps and graphs compared to a traditional numerical measure. Lift provides a much more accurate picture of the pandemic than other measures because it uses proportionality. A simple counting measure, for example, would have a bias toward numerically favoring states with a more significant number of cities.

Another point to highlight, after a closer observation of the maps, is the difference in the overall color pattern between the states located further north in Brazil (Northern and Northeastern regions, except Rondônia) compared to the states in the central-south region (including Rondônia), as if there were an imaginary line dividing the country. This macro-regional contrast and the spatial boundary of this "imaginary line" are instantly perceptible through visualization, whereas they would remain hidden in a standard sorted list of metrics for the 27 States. The states in the central-south region appear with the colors of the geographic areas tending much more towards red, evidencing higher lift values and, consequently, higher chances of occurrence of cities with a high rate of infection and mortality due to COVID-19. This visual evidence of a north-south division provides a geographic insight into how different regional public health policies or socioeconomic factors may have shaped the spatial distribution of the pandemic.

One possible cause could be the underreporting of cases and deaths in the Northern and Northeastern regions due to isolation and the precariousness of the health infrastructure in many communities in these regions. Another possibility is that the Northern and Northeastern regions adopted more effective sanitary measures to combat the pandemic than the central-south. In any case, a more accurate diagnosis requires a more in-depth analysis of the causes of the indicators that generated the spatial patterns observed in this example.

"Was the spatial pattern of cities with high COVID-19 mortality rates homogeneous across the Southeastern Region of Brazil in 2021, or were there areas that differed from the others?"

To answer this question, we created Figure 5, which presents three geographic visualizations (Maps **(a)**, **(b)** and **(c)**). In Maps **(a)** and **(b)**, a geographic filter was applied to select the Southeastern Region, and in Map **(c)**, we applied another filter to select the State of Minas Gerais. The rule antecedent, in all three maps, is defined by the combination of the attribute **year** equal to "2021" and by each visualized locality, being the **State** in Map **(a)**, the **Mesoregion** in Map **(b)**, and the **Microregion** in Map **(c)**. We defined the consequent by occurrences where the city presented the attribute **rtdeathpop** with the value "very high."

Figure 5 **(a)** reveals that in 2021, cities in Brazil's Southeastern states generally showed a strong tendency towards high lift values (indicated by red colors). However, the map clearly identifies Minas Gerais as a spatial exception within this high-risk regional block, presenting a lift of 2.9, significantly lower than the average value for the other states.

For a more precise analysis of the Southeastern Region, Figure 5 **(b)** was generated, where we increased the granular-

The rules for Figure 4 were as follows:

Rules for Figure 4 **(a):** $(state = \text{“}x\text{”}) \wedge (year = \text{“2021”}) \Rightarrow (rtcasepop = \text{“}very\ high\text{”})$

Rules for Figure 4 **(b)**: $(state = \text{“}x\text{”}) \wedge (year = \text{“2021”}) \Rightarrow (rtdeathpop = \text{“}very\ high\text{”})$

Where in both **(a)** and **(b)**, “x” refers to each state represented on the map.

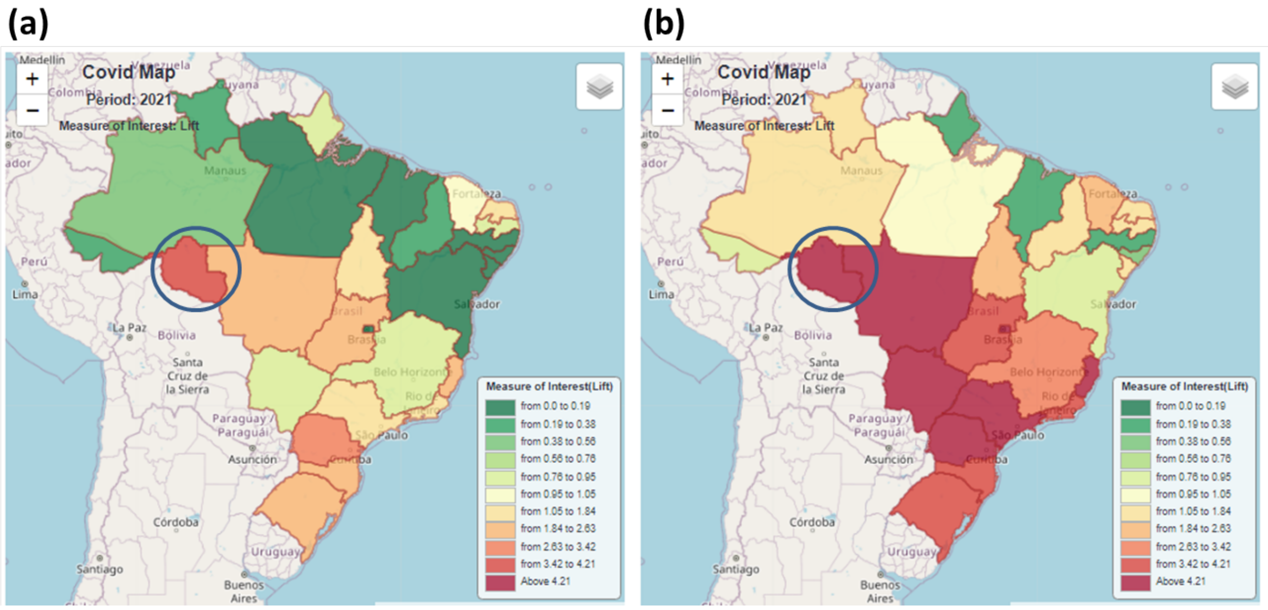

Figure 4. **(a)** Occurrences of cities with high annual infection rates of their population, by COVID-19, in 2021; **(b)** Occurrences of cities with high annual mortality rates of their population, by COVID-19, in 2021.

The rules for Figure 5 were as follows:

Rules for Figure 5 **(a):** $(state = \text{“}x\text{”}) \wedge (yr = \text{“2021”}) \Rightarrow (rtdeathpop = \text{“}v.\ high\text{”})$

Rules for Figure 5 **(b):** $(mesoregion = \text{“}x\text{”}) \wedge (yr = \text{“2021”}) \Rightarrow (rtdeathpop = \text{“}v.\ high\text{”})$

Rules for Figure 5 **(c):** $(microregion = \text{“}x\text{”}) \wedge (yr = \text{“2021”}) \Rightarrow (rtdeathpop = \text{“}v.\ high\text{”})$

Where in **a, b** and **c**, “x” refer to each State, Mesoregion and Microregion represented on the maps, respectively.

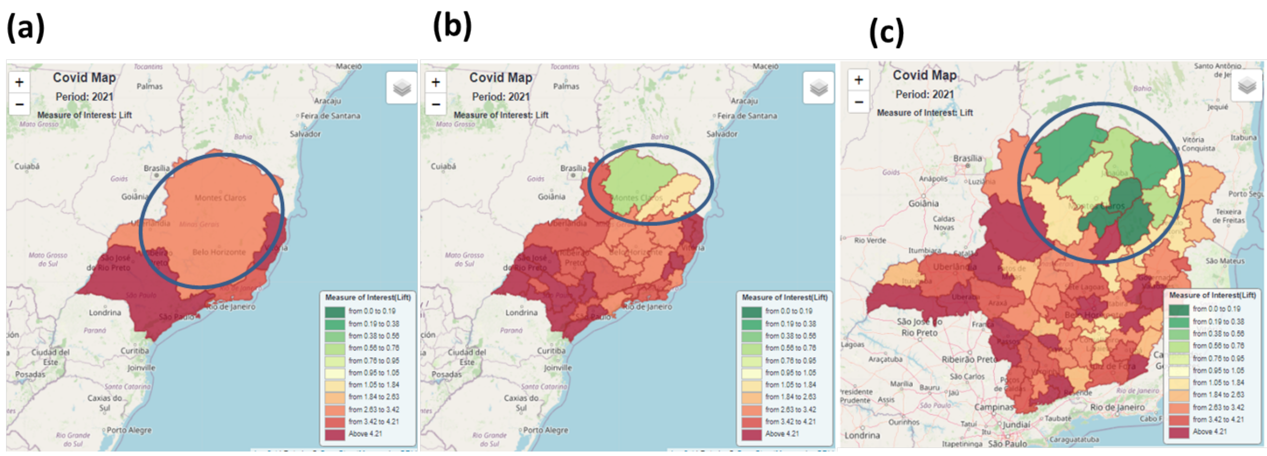

Figure 5. Occurrences of cities with high annual mortality rates of their population, by COVID-19, in 2021: **(a)** Southeast, by State; **(b)** Southeast, by Mesoregion; **(c)** Minas Gerais, by Microregion.

ity of the geographic divisions to the Mesoregion level. In this map, a relatively homogeneous pattern of high lift values was observed, except for a region in the north of Minas Gerais. The visualization highlights a significant spatial discontinuity, where three Mesoregions form a lower-risk corridor that contrasts sharply with the southern portion of the state. In order to further increase the accuracy of the spatial analysis, Figure 5 **(c)** was generated, where the state of Minas Gerais was filtered and used the Microrregion level of detail. The multiscale capability of MVAR-Geo reveals that this divergence is not random; it constitutes a spatially continuous cluster of 8 Microregions. In this visualization, it was possible to highlight this area encompassing a little over 100 cities, whose green color indicates low lift values. This visual grouping allows for an immediate assessment of regional behavior that would be impossible to synthesize through tabular data alone, as the map preserves the geographic relationship between these group of cities.

One possible explanation for this phenomenon could be the proximity of this area to the Northeastern Region of Brazil. By visualizing these patterns on the map, it becomes evident that this "atypical" zone is geographically adjacent to the Northeast, which may lead to similar political-economic characteristics, potential underreporting of deaths or an alignment of strategies to combat the Pandemic. The map effectively demonstrates that the northern border of Minas Gerais acts as a transition zone between two distinct national patterns. It is crucial to note that the use of functionalities, such as filters and navigation at different levels of spatial detail, played a key role in identifying the observed spatial patterns. Without the visual context of these nested scales, the identification of this specific geographic pocket of low mortality within a high-risk macro-region would be significantly hindered. A more comprehensive study would be beneficial to accurately diagnose the reasons for these patterns.

"What was the impact of vaccination on the occurrence of cities with high COVID-19 infection and mortality rates in 2021 and 2022?"

In response to this query, we created a visualization on Figure 6 to compare the patterns of occurrence for high COVID-19 infection rates (Maps **(a)** and **(b)**) and mortality rates (Maps **(c)** and **(d)**) across the years 2021 and 2022. We defined the rule antecedent in all maps by the combination of the georeferenced attribute **state** with the attribute **year** equal to "2021" in Maps **(a)** and **(c)**, and equal to "2022" in Maps **(b)** and **(d)**. The consequent of Maps **(a)** and **(b)** is defined by occurrences where the city presented the attribute **rtcasepop** as "very high" and, in the case of Maps **(c)** and **(d)**, by occurrences where the city presented the attribute **rtdeathpop** as "very high."

Maps **(a)** and **(b)** demonstrate that the spatial distribution pattern of contamination, dividing Brazil into north and south, observed in Figure 6, remained from 2021 to 2022. The visual persistence of this north-south polarization suggests that the factors driving infection maintained a stable geographic structure despite the passage of time. The states of the North and Northeast regions registered slight variations in the chances of occurrence of cities with high contamination rates. However, the states of the central-south region presented a slight worsening. The state of Espírito Santo was one of the most affected, presenting a variation of the lift from 2.4 in 2021 to 5.0 in 2022. The map effectively illustrates this worsening through the color transition from 2021 to 2022.

A comparison of Maps **(c)** and **(d)** shows an entirely different scenario. This pair of maps reveals a dramatic chromatic transition across the entire national territory. While in Map **(c)**, red tends to dominate, in Map **(d)**, the colors of all geographic areas are in shades of green. This global visual change provides an immediate summary of the collapse of mortality clusters from one year to the next. In 2021, several states presented lift values above 4, quadrupling the chances of occurrence of cities with high COVID-19 mortality rates. However, in 2022, most states had a zero value for this measure of interest, with no cities in this condition.

The transition of the pandemic's spatial behavior is most effectively captured through the "visual erasure" of previously high-mortality areas within our maps. Unlike a simple list of null numerical values—which often treats the absence of data as a mere statistical void—this cartographic omission provides a striking spatial signal of behavioral shifts. For instance, in the later stages of the pandemic, among the few remaining rules, only the state of São Paulo maintained a straight higher lift value of 0.25, which remains significantly lower than the 4.3 reported in 2021. By visualizing these "geographic silences" rather than relying on numerical tables, the MVAR-Geo approach transforms the absence of rules into a powerful indicator of the pandemic's evolving footprint. This interpretation is supported by recent literature on spatial dynamics; Nello-Deakin *et al.* [2024] highlight that analyzing such shifts is crucial to determining whether changes in urban mobility and spatial regimes represent a fundamental break or a continuity of prior trends [Nello-Deakin *et al.*, 2024].

The keeping of high contamination rates in 2022 can be mainly related to the appearance of the Omicron variant, which is much more infectious but less lethal than other virus variants [Freitas *et al.*, 2023]. In addition, the initial shortage of vaccines hindered the vaccination process. Another important fact was the high absenteeism rate of a portion of the population who, based on false information, distrusted the vaccines' efficacy and feared adverse reactions [Galhardi *et al.*, 2022].

Therefore, the effects of vaccination only became more relevant in 2022, partially explaining the significant decrease in mortality. The visual contrast between the persistent red in infection maps and the homogeneous green in mortality maps validates the primary characteristic of vaccines: the attenuation of symptoms rather than total immunization. The map allows the researcher to see exactly where the "decoupling" between infection and death occurred geographically. Thus, upon reaching significant vaccination coverage in 2022, mortality rates naturally decreased despite high infection rates.

5.2 Brazilian Presidential Election 2018-2022

In recent years, the far-right ideology has experienced significant political and electoral growth in Brazil, particularly from the 2018 presidential campaign onwards. That campaign can be considered a landmark due to its candidate's victory, with 55.1% of the valid votes in the second round. The 2022 presidential elections were a testament to the intense polarization that has gripped Brazil in recent decades. The first round was dominated by two candidates, one from the right and the other from the left. The second round, which saw the victory of the latter, was a nail-biting affair, with a very narrow margin of 50.9% to 49.1% of valid votes [Interian and Rodrigues, 2023]. This close contest underscores the dynamic and ever-changing nature of Brazilian politics.

Therefore, this case study aims to demonstrate how MVAR-Geo can help understand the recent electoral scenario in a geographic context in Brazil. At the same time, we brought geographic visualizations that present the most critical electoral

The rules for Figure 6 were as follows:

Rules for Figure 6 **(a):** $(state = \text{“}x\text{”}) \wedge (year = \text{“2021”}) \Rightarrow (rtcasepop = \text{“}very\ high\text{”})$

Rules for Figure 6 **(b)**: $(state = \text{“}x\text{”}) \wedge (year = \text{“2022”}) \Rightarrow (rtcasepop = \text{“}very\ high\text{”})$

Rules for Figure 6 **(c)**: $(state = \text{“}x\text{”}) \wedge (year = \text{“2021”}) \Rightarrow (rtdeathpop = \text{“}very\ high\text{”})$

Rules for Figure 6 **(d)**: $(state = \text{“}x\text{”}) \wedge (year = \text{“2022”}) \Rightarrow (rtdeathpop = \text{“}very\ high\text{”})$

Where in **(a)**, **(b)**, **(c)** and **(d)**, “x” refers to each State represented on the map.

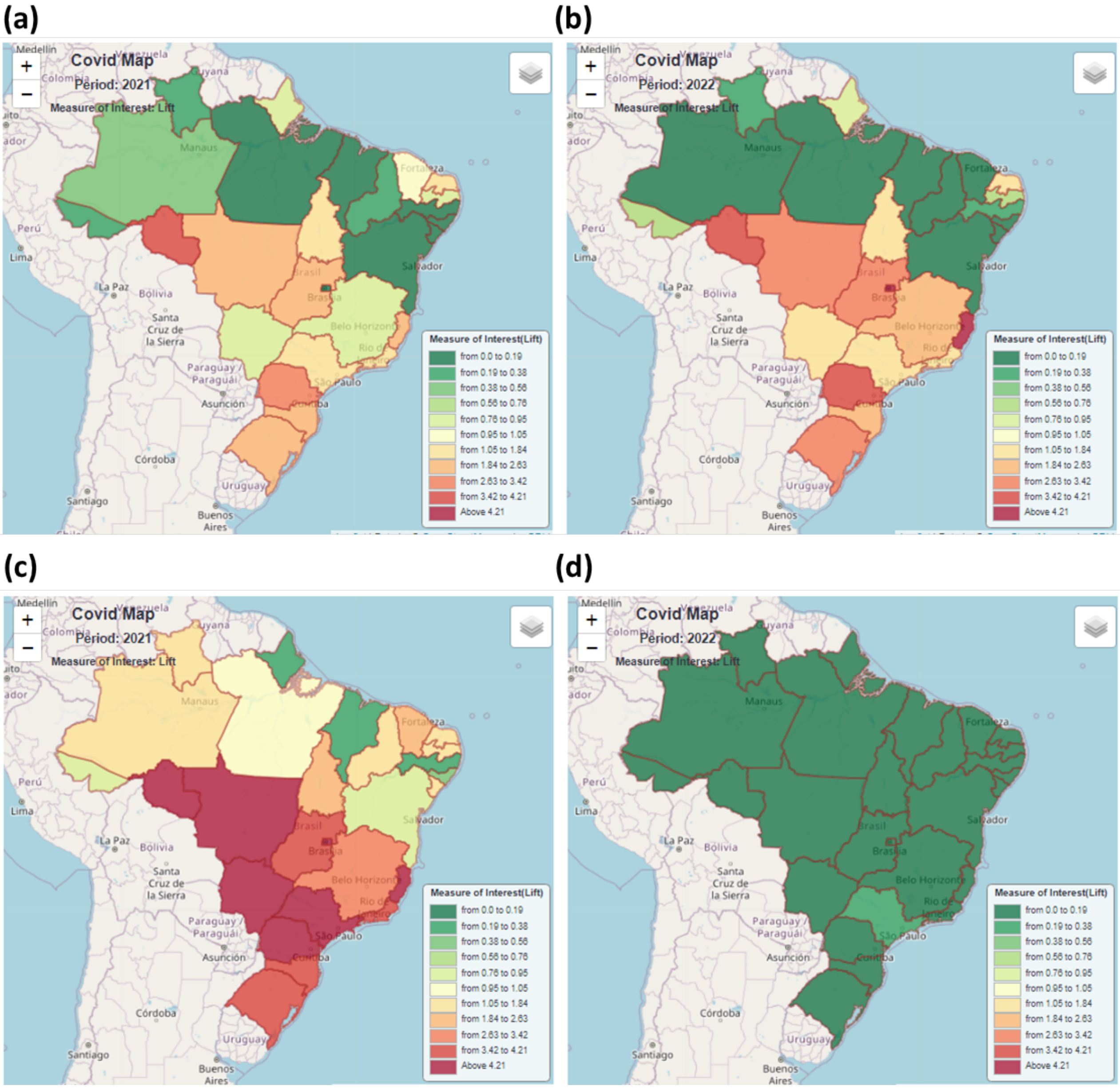

Figure 6. Occurrences of cities with high annual COVID-19 infection rates: **(a)** 2021; **(b)** 2022 - Occurrences of cities with high annual COVID-19 mortality rates: **(c)** 2021; **(d)** 2022.

strongholds of each political current. These observed changes between the two elections can explain the outcome of the latter. We also investigated whether municipal socioeconomic factors significantly influenced voter preferences, in addition to identifying specific regions that diverged electorally from their geographic context.

5.2.1 Data Preparation

The primary source used for this case study was the reports provided by the Superior Electoral Court of Brazil (TSE), which are available on this official body’s website [Tribunal Superior Eleitoral, 2024]. The TSE datasets, provided in CSV format with 22 attributes for 2018 and 26 for 2022, totaled 2.8 GB and approximately 12 million records covering information on municipalities, polling stations, and voting results.

To enable the case study, the data underwent a filtering and transformation process that selected only valid votes from the second round and excluded ballots cast abroad. This restriction had a negligible statistical impact—representing less than 0.25% of total votes—ensuring an optimized database focused on the national territory.

To generate the georeferenced association rules, we discretized the vote data into a categorical attribute that represents the result of each ballot box for the position of president in the election. We constructed the result considering the number of votes that each candidate had, generating three possible results:

- **Left** ⇒ The left-wing candidate was the winner of the ballot box;
- **Right** ⇒ The right-wing candidate was the winner of the ballot box;
- **Tie** ⇒ The candidates had the same number of votes in the ballot box.

Thus, each ballot box has only one result, which can be "Right", "Left" or "Draw". Nine hundred twenty-four thousand nine hundred seventy-five ballot boxes were used in the two elections, 453,829 in 2018 and 471,146 in 2022. Table 5 provides a sample of how we organize data from the results of each vote box in an important city (Mun.) in Brazil (Rio de Janeiro).

Table 5. Organization of 2018 and 2022 Election Data

City	B.B.	Elec.	Right	Left	Winner
Rio de Janeiro	4414	2018	87	44	Right
Rio de Janeiro	4414	2022	176	162	Right
Rio de Janeiro	4415	2018	115	119	Left
Rio de Janeiro	4415	2022	147	240	Left
Rio de Janeiro	5107	2018	68	31	Right
Rio de Janeiro	5107	2022	65	96	Left

An essential limitation of this methodology is the divergence between the proportion of ballot boxes won and the number of votes received by each candidate. Table 6 compares the proportions of total votes and ballot boxes won by each political current in the 2nd round of the two elections.

Table 6. Proportion of votes and ballot boxes won in the 2nd round of the 2018 and 2022 elections in Brazil

	2018		**2022**	
Result	% Votes	% Ballots	% Votes	% Ballots
Left	44.9%	36.5%	50.9%	47.8%
Right	55.1%	63.2%	49.1%	51.8%
Tie	-	0.3%	-	0.4%

Table 6 shows that, in both elections, the proportion of ballot boxes won by the right-wing candidate was more significant than the number of votes received, indicating that there was a better balance of voting in the ballot boxes where he won than in those where he lost. We conducted a thorough previous data analysis to ensure that this mismatch would not hinder our investigation. This analysis confirmed that the spatial distribution of candidates' votes was not significantly affected, allowing us to proceed with the study.

For the purpose of comprehensive analysis, spatial aggregations were employed to calculate and represent the measures of interest on the maps. These aggregations, which represent groupings of urns, were divided into five levels of granularity, including the municipal level, the smallest autonomous political-administrative division in Brazil. This approach facilitated more detailed and thorough analyses, similar to the COVID-19 study.

We created the scale_idh attribute to enrich the study based on the 2010 Brazilian municipal HDI. The HDI value ranges were defined based on the United Nations Development Program (UNDP) Report [Programme, 2024]. Thus, four representation values were defined, "Very High", "High", "Medium", and "Low" and their respective values are shown in Table 7. With these data, it is possible to assess whether the municipal HDI factor influenced the candidates' voting patterns.

Table 7. Range of values for the Municipal HDI

HDI Range	HDI Value
Low	up to 0.549
Medium	from 0.550 to 0.699
High	from 0.700 to 0.799
Very High	from 0.800 to 1.000

The duplicate files with geographic coordinates used in the COVID-19 study were used as support, with the addition of a file with the outlines of each Brazilian city to enable the rendering of geographic visualizations.

5.2.2 Results and Discussion

"Regarding the proportion of ballot boxes won, what were the main spatial and substantive changes in the 2022 general voting pattern, compared to 2018, that explain the return of the left to the presidency?"

To address this question, we developed two geographic visualizations (Figure 7, Maps **(a)** and **(b)**) for comparing the performance of right-wing candidates in the 2018 and 2022 elections. The antecedent of the rule, in both maps, is defined by combining the georeferenced attribute **state** with the **election** attribute equal to "2018", in Map **(a)**, and "2022", in Map **(b)**. We defined the consequence of the rule by the occurrences where the right-wing candidate won the ballot box.

The visualization provides an immediate diagnostic of electoral stability in the Northeast, where the maintenance of the green tone in Figure 7 signals that the right-wing candidate was unable to break through this solid geographic block. In these states, little has changed in 2022, indicating low values of lift and little chance of winning at the polls.

The scenario differs in the center-south, where the right-wing candidate is politically stronger. The visual transition between Maps **(a)** and **(b)** highlights a territorial contraction

The rules for Figure 7 were as follows:

Rules for Figure 7 **(a):** $(state = \text{“}x\text{”}) \wedge (election = \text{“}2018\text{”}) \Rightarrow (winner = \text{“}right\text{”})$

Rules for Figure 7 **(b)**: $(state = \text{“}x\text{”}) \wedge (election = \text{“}2022\text{”}) \Rightarrow (winner = \text{“}right\text{”})$

Where in both **(a)** and **(b)**, “x” refers to each State represented on the map.

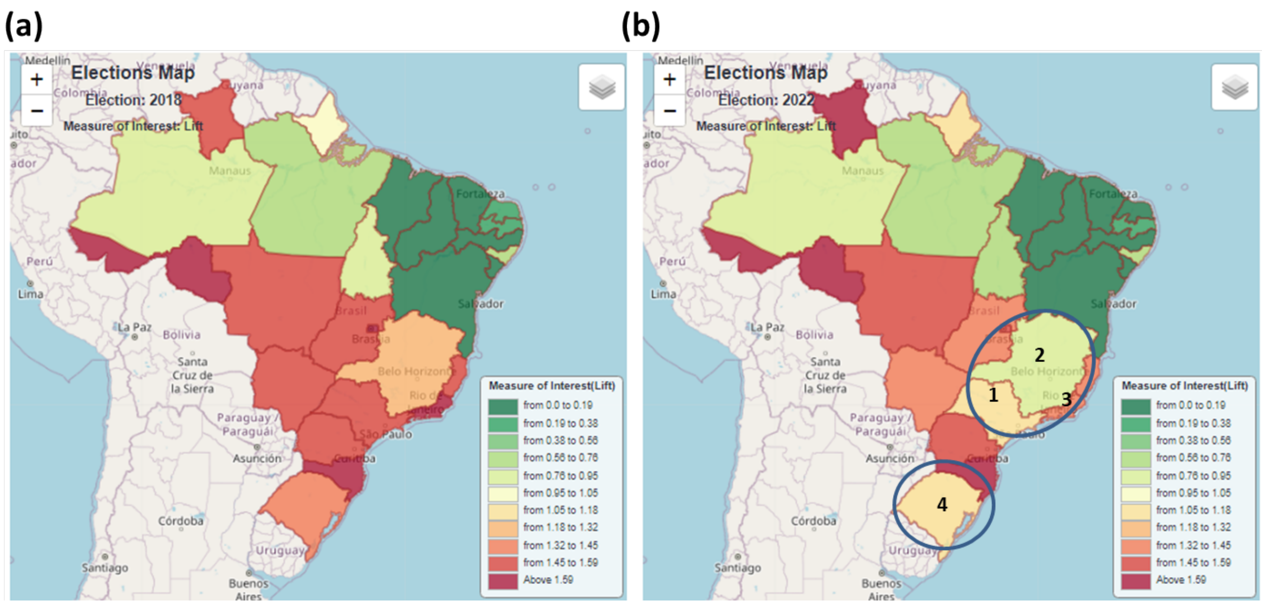

Figure 7. Occurrences where the winner of the ballot box was the right-wing candidate: **(a)** 2018 Election; **(b)** 2022 Election

The rules for Figure 8 were as follows:

Rules for Figure 8 **(a) and (b):** $(microregion = \text{“}x\text{”}) \wedge (election = \text{“}2022\text{”}) \Rightarrow (winner = \text{“}left\text{”})$

Rules for Figure 8 **(c)**: $(microregion = \text{“}x\text{”}) \wedge (election = \text{“}2022\text{”}) \Rightarrow (winner = \text{“}right\text{”})$

Where in **(a)**, **(b)** and **(c)**, “x” refers to each Microregion represented on the map.

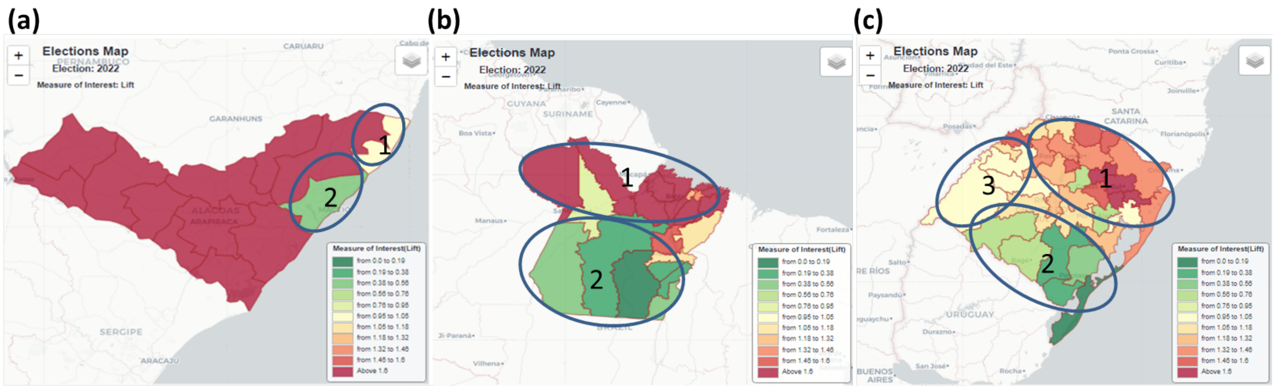

Figure 8. Proportion of occurrences, in the 2022 election - **(a)** in the state of Alagoas, where the left-wing candidate won at the ballot box; **(b)** in the state of Pará, where the left-wing candidate won at the ballot box; **(c)** in the state of Rio Grande do Sul, where the right-wing candidate won at the ballot box.

of dominance in the most populous regions of Brazil. The change in the colors of some states is straightforward, signaling a decrease in the chances of victory at the polls for this candidate, with emphasis on the states of São Paulo (1), Minas Gerais (2), Rio de Janeiro (3), and Rio Grande do Sul (4). By anchoring these lift variations on the map, MVAR-Geo allows for an instantaneous perception of how the candidate’s political erosion was spatially concentrated in a cluster that accounts for approximately 45% of the country’s votes. The right-wing candidate’s retreat in these critical states was decisive for his rival to reverse a disadvantage of almost 11 million votes in 2018 to an advantage of just over 2 million in 2022.

In this example, a traditional measure such as the count of expired ballots could also indicate the retreat of the right-wing candidate, but without the same diagnostic precision as the lift. As the number of ballot boxes increased from one election to another, the count does not reflect the actual size of its decline, unlike lift, which is based on statistical proportionality. The map provides a spatial validation of this proportionality; in the state of Acre, for instance, the subtle shift in the shade of red visually confirms a lower proportion of wins (lift of 1.63 in 2022 vs. 1.66 in 2018), even though the candidate won more absolute votes. This type of insight demonstrates how visualization enhances the understanding of electoral dynamics, revealing patterns of regional “cooling” or “heating” that raw totals or isolated metrics would fail to contextualize geographically.

“In the context of the 2022 election, within each UF, are there anomalies or spatial agglomerations that differ from the predominant voting pattern of the geographic space in which they belong?”.

To address the second question, Figure 8 was developed, showcasing three distinct geographic visualizations (Maps **(a)**, **(b)** and **(c)**), where the states of Alagoas were filtered, in Map **(a)**, from Pará, in Map **(b)**, and from Rio Grande do Sul, in Map **(c)**. The antecedent of the rule, in all maps, is defined by the georeferenced attribute **Microregion**, combined with the attribute **election** equal to “2022”. The consequent of Maps **(a)** and **(b)** is defined by the occurrences where the left-wing candidate was the winner of the ballot box and, in Map **(c)**, by the occurrences where the right-wing candidate was the winner of the ballot box.

The example in Figure 8 seeks to demonstrate that only with geographic visualization is it possible to perceive spatial patterns of agglomerations or anomalies, a type of analysis that is unfeasible to perform with graphs and tables, which only provide quantitative information. In Map **(a)**, the broad predominance of the left-wing candidate in Alagoas is demonstrated, winning 100% of the polls in some locations. However, the map clearly identifies a localized spatial divergence in the Microregions of the north coast of Alagoas (1) and Maceió (2) that breaks the state’s internal regional homogeneity. Mainly in Maceió, with approximately a third of the state’s votes, the color green, whose lift was 0.4, indicates a substantial reduction in the chances of the left-wing candidate winning at the polls, constituting a true anomaly in the face of electoral reality from the rest of the state.

On the Map **(b)**, the visualization reveals a sharp north-south territorial bisection in Pará. The first subspace (1), to the north of the state, the map shows a solid red cluster indicating a predominance of left-wing candidates. In the south of the state (2), the voting pattern was the opposite. The spatial context provided by the map suggests that this southern pattern is influenced by geographic adjacency to the state of Mato Grosso, a state that voted predominantly for right-wing candidates in the last two elections. This visual correlation between the border region and neighboring voting patterns provides a diagnostic insight that numerical tables would fail to capture, as they do not represent the physical proximity between states.

We see an interesting pattern in Map **(c)**. In Rio Grande do Sul, the map allows for the immediate identification of three distinct spatial clusters. In the first (1), located in the north of the state, the predominance of red forms a continuous geographic block favoring the right-wing candidate. On the other hand, in the southeast of the state (2), visualization reveals an agglomeration of six micro-regions where the voting pattern was the opposite, favoring the left-wing candidate. Finally, a third geographic subspace, located west of the state (3), is visually characterized as a transition zone due to a balance in candidate voting. The perception of these contiguous zones and their spatial boundaries provides a systemic understanding of the state’s electoral geography that cannot be replicated by viewing individual numeric values.

In summary, the spatial patterns observed in the three states represented by Figure 8 constitute interesting examples of anomalies and geographic subspaces that indicate divergent trends among themselves or about the larger context in which they belong. The use of MVAR-Geo transforms abstract ARM metrics into observable spatial phenomena, such as clusters, enclaves, and transition zones, which are essential for a deep understanding of geographic relationships.

“Did the municipal HDI significantly influence the 2022 election’s voting patterns? If so, where was this influence most pronounced?”.

Given the 2022 election’s intense ideological polarization, which led to right-wing and left-wing candidates being labeled “candidate of the rich” and “candidate of the poor”, respectively, this question is highly relevant. To provide an answer, Figure 9 was generated, presenting four geographic visualizations (Maps **(a)**, **(b)**, **(c)**, and **(d)**), where the antecedent of the rule, in all maps, is defined by the georeferenced attribute **state** combined with the attribute **election** equal to “2022” and with the attribute **range_hdi** equal to “high”, in Map **(b)** , and “low”, in Map **(d)**. The consequent of Maps **(a)** and **(b)** is defined by the occurrences where the right-wing candidate was the winner of the ballot box, and for Maps **(c)** and **(d)**, where the left-wing candidate was the winner of the ballot box.

Choosing municipal HDI in the “high” value range for the right-wing candidate and “low” for the left-wing candidate took into account their ideological profiles and national scope, as almost all Federation Units have cities in these HDI ranges, making it possible to carry out a more detailed analysis. Only the Federal District and the states of São Paulo, Rio de Janeiro, Espírito Santo, and Santa Catarina do not have cities with HDI in these value ranges.

The comparison between Maps **(a)** and **(b)** reveals significant changes in the colors of the geographic areas, with a tendency towards red, in Map **(b)**, indicating an increase in lift. This transition provides a visual diagnosis of a spatial association between high HDI and a preference for right-wing candidates, particularly in states where such socio-economic conditions are less frequent. For instance, in Amazonas (1) and Alagoas (2), the map identifies these high-HDI municipalities as distinct thematic clusters that contrast with their surroundings.

Similarly, the comparison between Maps **(c)** and **(d)** also reveals the same tendency towards red in Map **(d)**, showing that lower HDI levels are frequently associated with a higher lift for the leftist candidate. This is especially evident in Amazonas (1) and an area that encompasses the states of Minas Gerais, Tocantins and Goiás (2), which exhibited the most significant variations in lift values. The map reveals a continuous geographic corridor of lower socio-economic indicators, where low HDI serves as a common spatial denominator in the identified association rules, transcending state political

The rules for Figure 9 were as follows:

Rules for Figure 9 **(a):** $(state = \text{“}x\text{”}) \wedge (election = \text{“2022”}) \Rightarrow (winner = \text{“}right\text{”})$

Rules for Figure 9 **(b)**: $(state = \text{“}x\text{”}) \wedge (election = \text{“2022”}) \wedge (range\_hdi = \text{“}high\text{”}) \Rightarrow (winner = \text{“}right\text{”})$

Rules for Figure 9 **(c)**: $(state = \text{“}x\text{”}) \wedge (election = \text{“2022”}) \Rightarrow (winner = \text{“}left\text{”})$

Rules for Figure 9 **(d)**: $(state = \text{“}x\text{”}) \wedge (election = \text{“2022”}) \wedge (range\_hdi = \text{“}low\text{”}) \Rightarrow (winner = \text{“}left\text{”})$

Where in **(a)**, **(b)**, **(c)** and **(d)**, “x” refers to each State represented on the map.

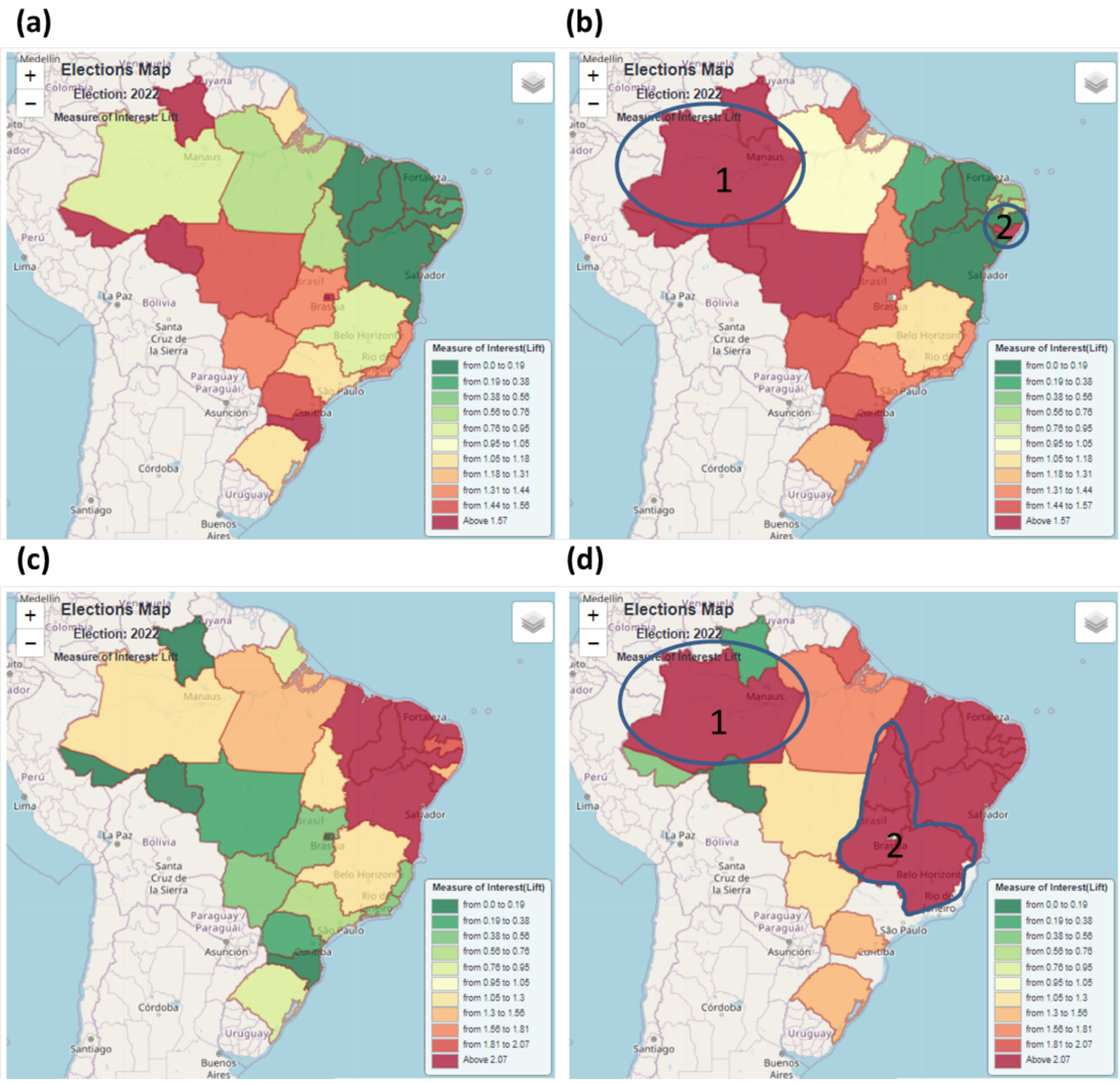

Figure 9. Occurrences in the 2022 election: **(a)** right-wing winner; **(b)** right-wing winner and high municipal HDI **(c)** left-wing winner; **(d)** left-wing winner and low municipal HDI.

boundaries.

It is important to clarify that while the inclusion of HDI in these rules identifies a strong spatial correlation between the socio-economic conditions of a municipality and its electoral profile, this does not imply a direct or definitive causal relationship. The patterns observed through Association Rule Mining (ARM) capture co-occurrences within the data rather than underlying causality.

In summary, while HDI appears to be a relevant factor in the electoral landscape of these cities, a more rigorous and comprehensive conclusion would require an in-depth study that incorporates additional regional, demographic, and cultural variables to avoid oversimplification and provide a more balanced interpretation of political behavior.

5.3 Smart Tourism

The large masses of data currently generated contain rich geographic knowledge and can help to face many urban challenges. Among these, tourism stands out, whose activity provides several benefits to cities, as it contributes to the economy, job creation, and the creation of business opportunities, especially in the service, retail, and restaurant sectors.

In this sense, it is possible to link tourist mobility data to data science approaches to understand tourist destinations and their interactions better, a practical application of the concept of "smart tourism" as a component of a "smart city" [Tripathy *et al.*, 2018]. Therefore, this case study aims to demonstrate how MVAR-Geo can be applied to data from cell phone connections in the context of smart tourism, helping to identify patterns of concentration of specific groups in certain locations, favoring the identification of preferences and possible opportunities related to tourism.

5.3.1 Data Preparation

This study used a data repository of cell phone connections from the operator Telecom Itália Mobile (TIM)[1], collected from the registry, in antennas distributed in the city of Rio de Janeiro, with a daily volume of around 130 million connections. Only the set of foreigner records was the object of this study, as it corresponds to only 0.006% of the total connections and has a greater probability of being related to tourism activity.

To survey foreign connections, approximately 21,000 raw files were processed using SSH commands, with the resulting data persisted in an intermediate repository. Since the original records contained only geographical coordinates, to identify geographical areas where connections occur, a computational routine was developed to link each antenna to its respective Administrative Region (AR) and neighborhood, utilizing geospatial data from the Data Rio portal [Data.Rio, 2024]. Following transformation and adjustment stages for the extraction of georeferenced association rules, a consolidated dataset of approximately 2.67 million connections was obtained, covering the period from May 2020 to October 2021.

The study's objective is to evaluate the pattern of occurrence of foreigners based on the daily record of their devices in geographic areas by extracting georeferenced association rules. In this way, the count of devices took into account their anonymized identifier, the date of connection, and the geographic area where the connection occurred, the unique combination of which is considered a device and counted only once, regardless of the number of connections made.

Devices are represented on the map through spatial aggregations, referring to the areas where their connections were registered, with the following levels of granularity available [Prefeitura da Cidade do Rio de Janeiro, 2024]:

- **Administrative Region (AR)** ⇒ Equivalent to a district, groups one or more neighboring neighborhoods;
- **Neighborhood** ⇒ Smallest administrative division of Rio de Janeiro.

[1] Fruit of a partnership between the University and the operator Telecom Italia Mobile (TIM)

Geospatial data files were also used as support to enable the visualization of Rio de Janeiro's administrative regions and neighborhoods.

5.3.2 Results and Discussion

"For the two foreign groups exhibiting the highest number of connections, is their spatial distribution characterized by areas of concentration, or is it proportionally homogeneous?"

To answer this question, we analyzed American and French connections, accounting for nearly 37% of the total volume. Figure 10 (Maps **(a)** and **(b)**) illustrates this analysis through two geographic visualizations, comparing the pattern of occurrence of devices of French origin with those of American origin, where the antecedent of the rule, in both maps, is defined by the georeferenced attribute **AR**. The consequent, in Map **(a)**, is defined by the occurrences of devices of French origin and, in Map **(b)**, by the occurrences of devices of American origin.

The most significant quantitative foreign connections occurred along the Rio de Janeiro waterfront. However, the geographic visualization allows us to look beyond these high-traffic areas to discover where these groups occur with a frequency above the expected norm. The use of lift on the map is essential here, as it identifies whether a specific location increases or decreases the chances of French or American occurrence.

The Maps **(a)** and **(b)** demonstrate very different patterns of occurrence of French and American devices. In Map **(a)**, the visualization clearly isolates two specific urban enclaves: the Administrative Regions of Botafogo and Santa Teresa. These areas form a localized cluster of interest, highlighted by the intense red on the map, with lift values of 1.6 and 2.0, respectively. This visual evidence reveals an increase of up to two times in the chances of French occurring in these specific neighborhoods. In contrast, by visualizing the American lift values in Map **(b)**, a pattern of relative spatial dispersion becomes evident. From the colors, a slight trend of concentration in the Barra da Tijuca Region is visible, highlighted on the map, with a lift of 1.2, indicating an increase of approximately 20% in the chances of devices from this group occurring.

In conclusion, while the spatial pattern of American device occurrences was inconclusive, the French devices revealed a strong attraction in the regions of Santa Teresa and Botafogo. The map effectively suggests the presence of potential tourist and cultural hotspots that are particularly appealing to the French. From a smart tourism perspective, the ability to visually distinguish these specific cultural centers combined with further investigation could uncover the reasons behind this pattern, paving the way for its replication or expansion in other city regions.

"Considering the two foreign groups with the most connections, how did the spatial distribution of their occurrence evolve from 2020 to 2021?"

The rules for Figure 10 were as follows:

Rules for Figure 10 **(a):** $(ar = \text{“}x\text{”}) \Rightarrow (origin = \text{“}france\text{”})$

Rules for Figure 10 **(b)**: $(ar = \text{“}x\text{”}) \Rightarrow (origin = \text{“}usa\text{”})$

Where in both **(a)** and **(b)**, “x” refers to each Administrative Region represented on the map.

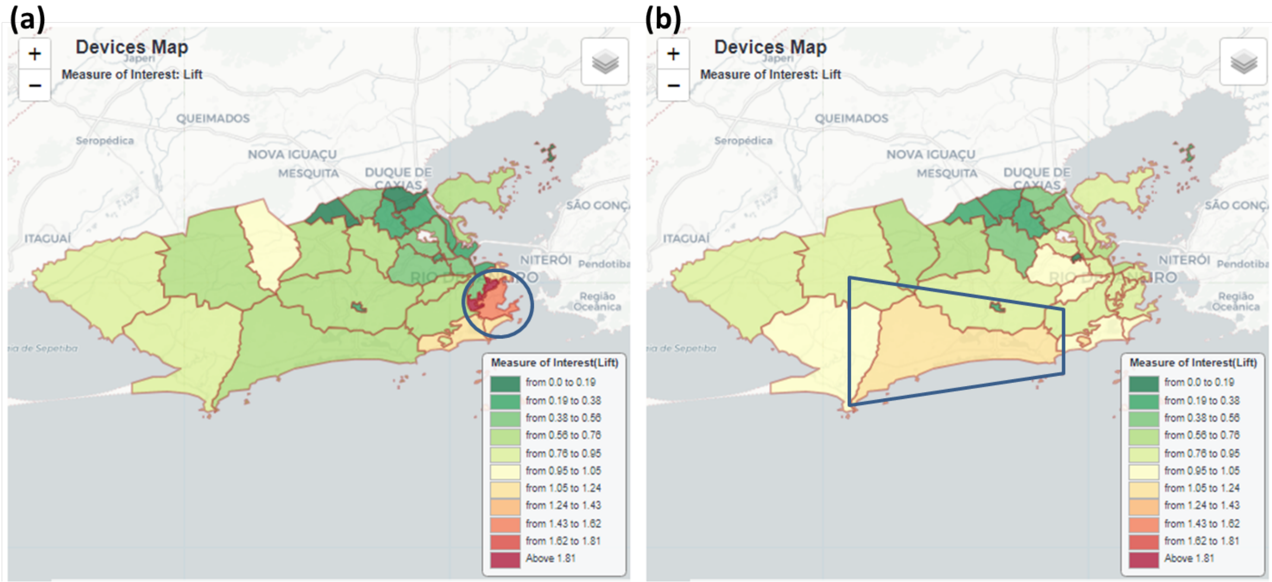

Figure 10. Device occurrences: **(a)** French origin; **(b)** American origin.

To answer this question, we created four geographic visualizations (Figure 11, Maps **(a)**, **(b)**, **(c)** and **(d)**), where the antecedent of the rule, in all maps, is defined by the georeferenced attribute **AR**, combined with the attribute **year** equal to “2020”, in Maps **(a)** and **(c)**, and “ 2021”, in Maps **(b)** and **(d)**. The consequent, for Maps **(a)** and **(b)**, is defined by the occurrences of devices of French origin, and for Maps **(c)** and **(d)**, by the occurrences of devices of American origin.

In Figure 11, data from 2020 and 2021 were separated, revealing distinct spatial trajectories for each group. The visual comparison between the evolution of occurrences of devices of French origin (Maps **(a)** and **(b)**) with those of American origin (Maps **(c)** and **(d)**), reveals completely different evolutionary patterns that are immediately identifiable through their chromatic shifts.

In the case of the French, in Maps **(a)** and **(b)**, the distribution of the occurrence of this group is relatively similar in 2020 and 2021, but the map highlights an intensification of their presence in specific enclaves. It is visible that there was an increase in concentration, indicating greater chances of occurrence of French people than other groups of foreigners in 2021. This perception is confirmed by the lift values of the Regions of Botafogo (1), which went from 1.34 to 1.75, and Santa Teresa (2), which went from 1.96 to 2.04. The persistence of these patterns across both years visually validates these neighborhoods as stable cultural anchors for this group of foreigners.

In the case of the Americans, the Maps **(c)** and **(d)** show very different patterns of occurrence. In 2020, Map **(c)**, the color pattern of the geographic areas is heterogeneous, visually denoting a solid concentration of this group on the edge of Rio de Janeiro. A different scenario is observed in 2021, Map **(d)**, in which the colors of the geographic areas are very similar, indicating a dramatic spatial transformation from a concentrated coastal footprint to a diffuse, low-intensity pattern throughout the city. Regarding the concentration intensity, the transition to green in Map **(d)** indicates a significant reduction in the chances of occurrence of this group in 2021. This perception is confirmed by the lift values of the regions highlighted on the maps, in the case of Guaratiba (1), which went from 1.67 to 0.53, and Barra da Tijuca (2), which went from 1.74 to 0.72. This visualization provides a contextualized diagnosis of territorial retreat that a simple list of decreasing values could not provide.

The difference in the evolution of the occurrence patterns of the two groups between 2020 and 2021 may be an indication of different visitor profiles. Trips for commercial purposes have a different chance of cancellation or postponement than those for tourist purposes. By using MVAR-Geo, it is possible to spatially visualize these different profiles: the resilience of the French group in specific cultural centers versus the retreat of the American group in relation to coastal business and leisure centers. In any case, the patterns observed in Figure 11 raise further studies of these issues, demonstrating that the visualization of ARM rules is a powerful tool for monitoring temporal shifts in urban dynamics.

6 Conclusions and Future Work

This work proposes an innovative way of integrating traditional ARM with GIS in an advanced geospatial visualization method called MVAR-Geo, which is a novel and efficient method for visualizing association rules with georeferenced

The rules for Figure 11 were as follows:

Rules for Figure 11 **(a):** $(ar = \text{“}x\text{”}) \wedge (year = \text{“2020”}) \Rightarrow (origin = \text{“}france\text{”})$

Rules for Figure 11 **(b)**: $(ar = \text{“}x\text{”}) \wedge (year = \text{“2021”}) \Rightarrow (origin = \text{“}france\text{”})$

Rules for Figure 11 **(c)**: $(ar = \text{“}x\text{”}) \wedge (year = \text{“2020”}) \Rightarrow (origin = \text{“}usa\text{”})$

Rules for Figure 11 **(d)**: $(ar = \text{“}x\text{”}) \wedge (year = \text{“2021”}) \Rightarrow (origin = \text{“}usa\text{”})$

Where, across all geographic visualizations, “x” refers to each Administrative Region represented on the map.

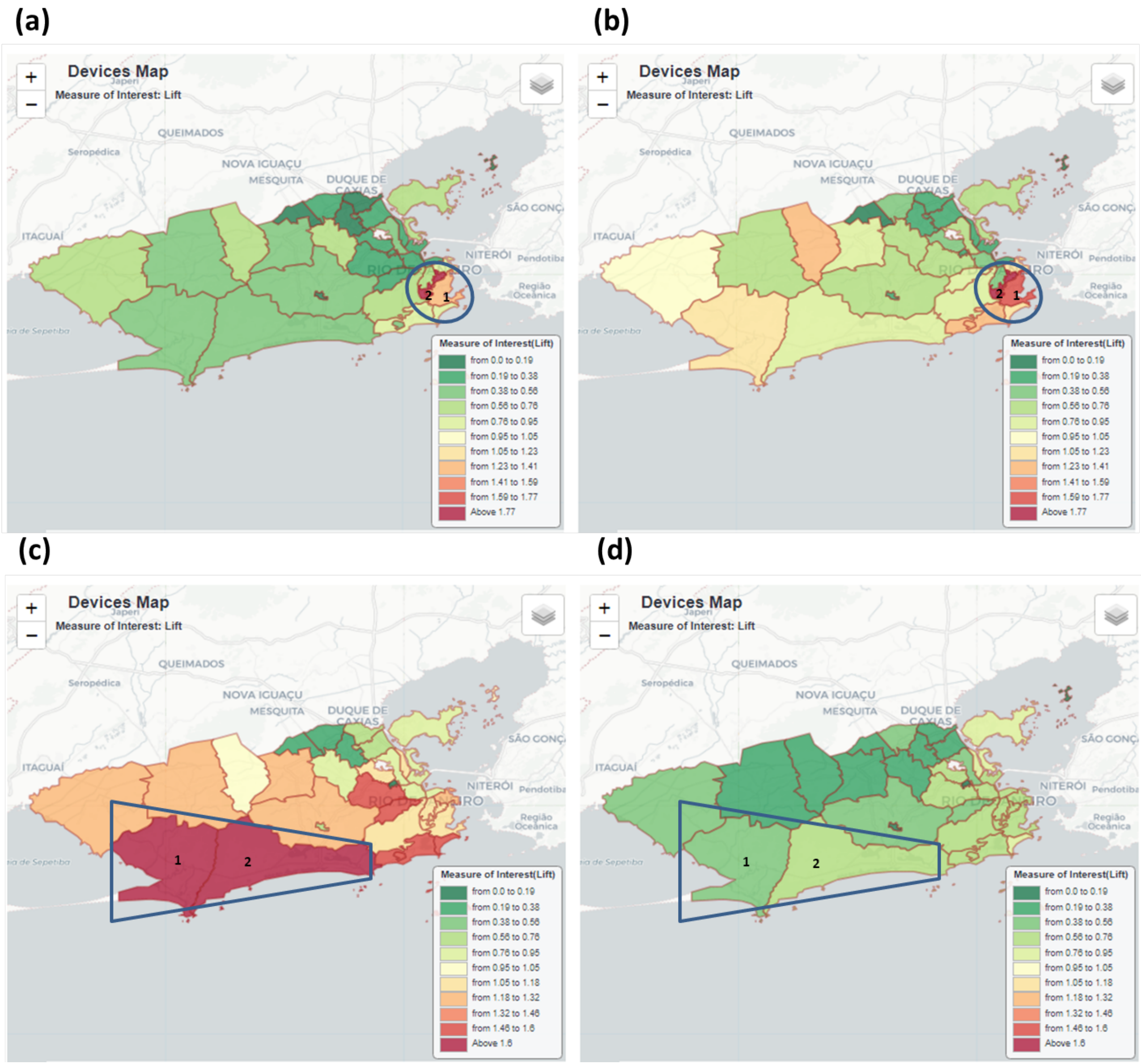

Figure 11. Device occurrences: **(a)** French origin in 2020; **(b)** French Origin in 2021; **(c)** American Origin in 2020; **(d)** American origin in 2021.

attributes, providing a clear visual representation of where these associations are most significant.

This methodology not only assists in detecting spatial trends but also supports decision-making processes in various applications, inspiring new possibilities and solutions. The proposed features facilitate the identification of interesting spatial association patterns. Because it allows for simultaneous visualization of different information, it is possible to compare different regions, detail specific locations, and perform temporal analyses, revealing spatial information that can broaden knowledge about the application domain under study, generating useful knowledge quickly.

Three case studies covering different domains were carried out to evaluate and show MVAR-Geo’s wide applicability and distinctive advantages. They explored several examples of how the analysis of association rules with georeferenced attributes can benefit from visual data analysis techniques, specifically highlighting important information, such as occurrence patterns in space, that would be hidden with traditional analysis methods, such as lists and graphs. The use of lift, as the measure of interest represented in geographic visualizations on these case studies, made it possible to improve the understanding of how certain geographic areas can influence the chances of occurrence of certain events.

In the case of the COVID-19 pandemic, the intention was to highlight how MVAR-Geo, used in epidemiological contexts, can quickly bring differentiated and timely information to health specialists, helping them to understand better, in an agile way, how and why certain diseases behave in a geographic context and, thus, better direct their efforts and decisions to combat them (e.g., resource allocation, imposition of public health measures). In the study of the 2018 and 2022 presidential elections, it was highlighted how the proposed method, used in electoral scenarios, could help political analysts better understand trends and voting patterns in the political environment studied by identifying certain interesting electoral phenomena in a geographic context, assisting in the formulation of campaign strategies or result analysis. Regarding smart tourism, using data from foreigners' cell phone connections, it sought to demonstrate how we could apply MVAR-Geo to better understand the occurrence patterns of specific groups. This information is crucial for supporting policy planning and the allocation of investments in tourism infrastructure and services.

When applying MVAR-Geo, we recommend taking a series of precautions to minimize bias and interpretation errors on the part of users. Firstly, database preparation is essential, including data cleaning and normalization, ensuring the rules extracted are based on accurate and relevant information. Adjustments to scale, size, and colors of maps are also necessary to ensure that the representation of spatial patterns occurs proportionally and understandably. Mainly, the appropriate use of colors in visualization is essential to avoid confusion or misinterpretations, such as using color scales that may mislead the user into overestimating or underestimating the strength of associations. It is recommended to follow the color palette standards proposed in Harrower and Brewer [2003].

However, it is essential to recognize the method's limitations, such as its dependence on the availability of quality georeferenced data and the possibility of not capturing contextual nuances that could influence associations. Furthermore, for scope delimitation, MVAR-Geo was restricted to the static and two-dimensional visualization of association rules. These rules aim to explain the occurrence of specific events based on relationships between geographic areas or points and other variable types. Consequently, the work did not contemplate more sophisticated models already in existence, such as geographic animations or three-dimensional maps.

Considering these precautions and limitations is the best way to ensure that MVAR-Geo offers a clear and objective visualization of the association rules without compromising the interpretation of the results. In summary, the MVAR-Geo method has significant potential to advance the field of geospatial analysis by combining it with association rule analysis to provide a comprehensive framework that integrates multiple visualization tools to highlight significant patterns and relationships in spatial data. The result is a more comprehensive, interactive, and insightful visualization of Georeferenced Association Rules, favoring their analysis and the consequent identification of differentiated spatial patterns.

Future research can focus on several promising directions to expand and deepen the findings presented in this work. In what follows, there is a list of potential directions.

Enhanced Interactivity and Visualization Techniques Firstly, exploring more sophisticated forms of interactivity, such as geographic animations, could facilitate dynamic visualization of changes in rules and patterns over time, offering clearer temporal insights. Additionally, there is potential to apply the proposed method to advanced visualization techniques, including thematic map overlays or three-dimensional geovisualizations, to reveal hidden patterns in complex data from new perspectives.

Application to Other Domains The MVAR-Geo method can also be extended to analyze data from other application domains. This includes scenarios like climate change monitoring, natural disaster prediction, urban planning based on human mobility patterns, global epidemiological surveillance, and sociodemographic data. This would enable the cross-referencing of geospatial information with new types of variables, opening up new avenues for spatial analyses of inequality and social impact studies, ultimately aiding in the formulation of effective public policies.

Integration with Other Techniques The method can be adapted to encompass other georeferenced data mining techniques. One alternative involves using predictive analytics through regression techniques to visually compare geovisualizations based on real data with those generated from model variations. This would enable a direct assessment of prediction accuracy by observing the alignment between predicted results and observed data. Another promising direction involves combining the method with Explainable Artificial Intelligence (XAI) techniques. Given that association rules are widely used in XAI to provide interpretable explanations of AI model behavior, this method could be crucial in a geospatial analysis and data visualization context for users to comprehend how decisions or patterns were generated. Another possibility would be a comparison of MVAR-Geo with alternative methods (such as observed/expected ratio maps), which could be conceptual or using empirical data.

Incorporation of Spatial Variability Techniques We propose incorporating spatial variability into the MVAR-Geo model to deepen the analysis of dependency between areas. This could be achieved by including a pre-processing stage to integrate neighborhood attributes—such as the average interest metric (e.g., lift) of adjacent zones—prior to rule mining. Furthermore, we envision adapting support, confidence, and lift measures based on spatial autocorrelation indices, such as Moran's I. This would allow for weighting the relevance of rules according to the identified degree of spatial clustering, thereby providing a more robust understanding of neighborhood effects and geographic patterns.

Acknowledgements

This work was partially supported by CNPq (grant 315750/2021-9) and FAPESP/MCTIC/CGI.br (grant 2020/09835-1) within the scope of the CPA-IA IARA - Artificial Intelligence Recreating Environments. We would like to thank the operator TIM for providing the data used in the smart tourism case study.

Funding

Funding: Alexandre Plastino and Antonio A. de A. Rocha acknowledge funding by IARA - Center of Artificial Intelligence in the Remarking of Urban Environments (FAPESP project grant 2020/09835-1). Antonio A. de A. Rocha also acknowledges funding by grants from CNPq (grants 405852/2025-8 and 312687/2025-7) and FAPERJ (LINE grant 260.279/2026).

Authors' Contributions

All authors contributed to the writing of this article, read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The data that support the findings of this study are openly available at `https://github.com/ackuhl2/MVAR_Geo` Please be advised that the data does not violate the protection of human subjects or other valid ethical, privacy, or security concerns.

References

Agrawal, R., Imieliński, T., and Swami, A. (1993). Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD international conference on Management of data*, pages 207–216. DOI: 10.1145/170036.170072.

Andrienko, G., Andrienko, N., and Weibel, R. (2017). Geographic data science. *IEEE Computer Graphics and Applications*, 37(5):15–17. DOI: 10.1109/mcg.2017.3621219.

Barcellos, R., Viterbo, J., Miranda, L., Bernardini, F., Maciel, C., and Trevisan, D. (2017). Transparency in practice: using visualization to enhance the interpretability of open data. In *Proceedings of the 18th Annual International Conference on Digital Government Research*, pages 139–148. DOI: 10.1145/3085228.3085294.

Brandão, C. C., Mendonça, A. V. M., and Sousa, M. F. d. (2023). O ministério da saúde e a gestão do enfrentamento à pandemia de covid-19 no brasil. *Saúde em Debate*, 47:58–75. DOI: 10.1590/0103-1104202313704.

Brasil, C. (2023). Painel coronavírus. Available in: `https://covid.saude.gov.br` (in Portuguese) accessed on July 12, 2023.

Brin, S., Motwani, R., Ullman, J. D., and Tsur, S. (1997). Dynamic itemset counting and implication rules for market basket data. In *Proceedings of the 1997 ACM SIGMOD international conference on Management of data*, pages 255–264. DOI: 10.1145/253262.253325.

Canlon, D. C., Paulovich, F., and Tennekes, M. (2024). Fr-glyphs for multidimensional categorical data. DOI: 10.2312/evs.20241061.

Cidades, I. (2023). Cidades e estados do brasil. Available in: `https://cidades.ibge.gov.br/` (in Portuguese) accessed on November 21, 2023.

Ciuccarelli, P., Lupi, G., and Simeone, L. (2014). *Visualizing the data city: social media as a source of knowledge for urban planning and management*. Springer Science & Business Media. DOI: 10.1007/978-3-319-02195-9.

Crawford, F., Watling, D., and Connors, R. (2021). Analysing spatial intrapersonal variability of road users using point-to-point sensor data. *Networks and Spatial Economics*, pages 1–34. DOI: 10.1007/s11067-021-09539-4.

Cui, J., Wang, Y., and Yu, Z. (2007). Association rules mining and expressing in census data based on geo-visualization. In *Geoinformatics 2007: Geospatial Information Science*, volume 6753, pages 196–206. SPIE. DOI: 10.1117/12.761827.

Data.Rio (2024). Data.Rio - Portal de Dados Abertos da Cidade do Rio de Janeiro. Available in: `https://www.data.rio/` (in Portuguese) accessed on April 11, 2024.

Deng, X., Liu, P., Liu, X., Wang, R., Zhang, Y., He, J., and Yao, Y. (2019). Geospatial big data: New paradigm of remote sensing applications. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 12(10):3841–3851. DOI: 10.1109/jstars.2019.2944952.

Edsall, R. M. (2003). Design and usability of an enhanced geographic information system for exploration of multivariate health statistics. *The Professional Geographer*, 55(2):146–160. DOI: 10.1111/0033-0124.5502003.

Freitas, C. M. d., Barcellos, C., Villela, D. A. M., Portela, M. C., Reis, L. C., Guimarães, R. M., Xavier, D. R., Saldanha, R. d. F., and Mefano, I. V. (2023). Covid-19 fiocruz observatory-an analysis of the evolution of the pandemic from february 2020 to april 2022. *Ciência & Saúde Coletiva*, 28:2845–2855. DOI: 10.1590/1413-812320232810.10412023en.

Galhardi, C. P., Freire, N. P., Fagundes, M. C. M., Minayo, M. C. d. S., and Cunha, I. C. K. O. (2022). Fake news and vaccine hesitancy in the covid-19 pandemic in brazil. *Ciência & Saúde Coletiva*, 27:1849–1858. DOI: 10.1590/1413-81232022275.24092021en.

Hahsler, M. (2015). A probabilistic comparison of commonly used interest measures for association rules. *United States. Southern Methodist University*, 27. Available at:`https://mhahsler.github.io/arules/docs/measures.pdf`.

Han, J., Pei, J., and Tong, H. (2022). *Data mining: concepts and techniques*. Morgan kaufmann. DOI: 10.5860/choice.49-3305.

Harrower, M. and Brewer, C. A. (2003). Colorbrewer. org: an online tool for selecting colour schemes for maps. *The Cartographic Journal*, 40(1):27–37. DOI: 10.1179/000870403235002042.

Interian, R. and Rodrigues, F. A. (2023). Group polarization, influence, and domination in online interaction networks: A case study of the 2022 brazilian elections. *Journal of Physics: Complexity*, 4(3):035008. DOI: 10.1088/2632-072x/acf6a4.

Khan, S. (2021). Data visualization to explore the countries dataset for pattern creation. *International Journal of Online & Biomedical Engineering*, 17(13). DOI: 10.3991/ijoe.v17i13.20167.

Koperski, K. and Han, J. (1995). Discovery of spatial association rules in geographic information databases. In *Inter-*

national Symposium on Spatial Databases, pages 47–66. Springer. DOI: 10.1007/3-540-60159-7_4.

Madala, S. R., Rajavarman, V., and Venkata Satya Vivek, T. (2018). Analysis of different pattern evaluation procedures for big data visualization in data analysis. In *Data Engineering and Intelligent Computing: Proceedings of IC3T 2016*, pages 453–461. Springer. DOI: 10.1007/978-981-10-3223-$3_4$4.

Miranda, F., Hosseini, M., Lage, M., Doraiswamy, H., Dove, G., and Silva, C. T. (2020). Urban mosaic: Visual exploration of streetscapes using large-scale image data. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–15. DOI: 10.1145/3313831.3376399.

Nello-Deakin, S., Diaz, A. B., Roig-Costa, O., Miralles-Guasch, C., and Marquet, O. (2024). Moving beyond covid-19: Break or continuity in the urban mobility regime? *Transportation research interdisciplinary perspectives*, 24:101060. DOI: 10.1016/j.trip.2024.101060.

Openshaw, S. (1984). *The Modifiable Areal Unit Problem*. Number 38 in Concepts and Techniques in Modern Geography. Geo Books, Norwich. Book.

Prefeitura da Cidade do Rio de Janeiro (2024). Rio prefeitura. Available at:`https://www.rio.rj.gov.br/` (in Portuguese) accessed on March 5, 2024.

Programme, U. N. D. (2024). Human development index (hdi). DOI: 10.1007/978-3-319-95873-$6_3$00083.

Schumann, H. and Tominski, C. (2011). Analytical, visual and interactive concepts for geo-visual analytics. *Journal of Visual Languages & Computing*, 22(4):257–267. DOI: 10.1016/j.jvlc.2011.03.002.

Sun, X. and Wang, X. (2017). Geovisualization for association rule mining in oil and gas well data. *ISPRS International Journal of Geo-Information*, 6(2):48. DOI: 10.3390/ijgi6020048.

Torrisi, A., Signorello, G., Gallo, G., De Salvo, M., and Farinella, G. M. (2015). Mining social images to analyze routing preferences in tourist areas. In *EnvirVis@ EuroVis*, pages 61–65. DOI: 10.2312/envirvis.20151093.

Tribunal Superior Eleitoral (2024). Tse - dados abertos. Available at:`https://dadosabertos.tse.jus.br/` (in Portuguese) accessed on February 15, 2024.

Tripathy, A. K., Tripathy, P. K., Ray, N. K., and Mohanty, S. P. (2018). itour: The future of smart tourism: An iot framework for the inde-pendent mobility of tourists in smart cities. *IEEE consumer electronics magazine*, 7(3):32–37. DOI: 10.1109/MCE.2018.2797758.

Vougas, K., Sakellaropoulos, T., Kotsinas, A., Foukas, G.-R. P., Ntargaras, A., Koinis, F., Polyzos, A., Myrianthopoulos, V., Zhou, H., Narang, S., *et al*. (2019). Machine learning and data mining frameworks for predicting drug response in cancer: An overview and a novel in silico screening process based on association rule mining. *Pharmacology & therapeutics*, 203:107395. DOI: 10.1016/j.pharmthera.2019.107395.

Wang, J., Li, G., and Ai, T. (2024). Revealing association rules within intricate ecosystems: A spatial co-location mining method based on geo-eco knowledge graph. *International Journal of Applied Earth Observation and Geoinformation*, 133:104116. DOI: 10.1016/j.compenvurbsys.2024.104116.

Wong, D. W. S. (2004). The modifiable areal unit problem (maup). In *WorldMinds: Geographical Perspectives on 100 Problems*, pages 571–575. Springer, Dordrecht. DOI: 10.4135/9780857020130.n7.

World Health Organization (2021). Clinical case definition of post COVID-19 condition. Available at:`https://www.who.int/publications/i/item/WHO-2019-nCoV-Post_COVID-19_condition-Clinical_case_definition-2021.1`. accessed on April 12, 2024.

Xu, Y., Shaw, S.-L., Zhao, Z., Yin, L., Lu, F., Chen, J., Fang, Z., and Li, Q. (2018). Another tale of two cities: Understanding human activity space using actively tracked cellphone location data. In *Geographies of Mobility*, pages 246–258. Routledge. DOI: 10.1080/00045608.2015.1120147.

Zhang, S., Yang, Z., and Wang, C. (2024). The border tourism hotspots network based on travelogues. *Heliyon*, 10(13). DOI: 10.1016/j.heliyon.2024.e33260.