Multidimensional Analysis of the Influence of Socioeconomic and Political Indicators on the Spread of COVID-19: A Case Study of Brazilian Cities (2020-2024)

Rôney Reis[1], Angelo Brayner[2], Miguel Ângelo[1], Ronaldo Menezes[3,1]

[1]Department of Medical Sciences – Federal University of Ceará (UFC)
Fortaleza, CE – Brazil

[2]Department of Computer Science – Federal University of Ceara (UFC)
Fortaleza, CE – Brazil

[3]BioComplex Laboratory, Department of Computer Science – University of Exeter
Exeter – England

roney@alu.ufc.br, brayner@dc.ufc.br, mans@ufc.br, r.menezes@exeter.ac.uk

Abstract. *This paper explores the influence of socioeconomic indicators and political decisions on the spread of COVID-19 across Brazilian cities from 2020 to 2024. Leveraging data on COVID-19 cases, deaths, electoral outcomes from 2020 and 2022, and the Human Development Index (HDI) from 2010, we employ a multidimensional analytical framework encompassing temporal, spatial, and statistical dimensions to uncover the correlations among these variables. Time series models, such as ARIMA, were employed to detect trends over time, while spatial correlation analyses and machine learning techniques were applied to reveal geographical variations in virus spread. Our findings highlight significant regional disparities in COVID-19 proliferation, carrying crucial implications for the formulation of targeted public policies.*

1. Introduction

The COVID-19 pandemic has precipitated an unprecedented global health crisis, profoundly impacting various aspects of societies worldwide. In Brazil, the spread of the virus and its health consequences were particularly severe, resulting in over 38 million cases and more than 700,000 deaths[1]. In response, numerous studies have sought to identify the factors contributing to the spread of COVID-19, focusing on epidemiological, socioeconomic, and political variables.

One critical aspect examined is how public policies, communication strategies, and adherence to public health measures vary significantly with the political profile of different regions. Our investigation specifically targets the influence of political and socioeconomic factors on the spread of COVID-19 across Brazilian cities between 2020 and 2024. We adopted a multidimensional approach, integrating temporal, spatial, and statistical analyses, to uncover patterns and correlations that provide deeper insights into the pandemic's dynamics within diverse political and socioeconomic contexts. In particular, we sought to understand how electoral choices, as reflected in votes for right-wing

[1]Coronavírus Brasil. `https://covid.saude.gov.br/`. Accessed: 13 August 2024

and left-wing candidates, and socioeconomic indicators, such as the Human Development Index (HDI), correlate with COVID-19 cases and deaths.

This study builds upon previous research by incorporating a comprehensive spatial correlation analysis, enabling the identification of geographic patterns linked to variations in pandemic outcomes. By synthesizing multiple data sources and employing advanced analysis techniques, including Autoregressive Integrated Moving Average (ARIMA) [Box et al. 2015] for time series forecasting, spatial correlation analysis, and machine learning models, this paper offers a detailed and nuanced perspective on the factors influencing the progression of COVID-19 in Brazil. The insights derived from this analysis are intended to inform policymakers and public health officials in crafting strategies that are attuned to local realities.

The remainder of this paper is organized as follows: Section 2 reviews related work on the impact of political and socioeconomic factors on COVID-19. Section 3 outlines the methodology employed in this study, covering data collection, preparation, and analysis techniques. The results of our analysis are presented in Section 4, with a focus on seasonality and trend analysis using ARIMA models, temporal correlations, and spatial analysis conducted with K-Means clustering. Finally, Section 5 discusses the findings, their implications for public policy, and provides suggestions for future research.

2. Related Work

Numerous studies have aimed to elucidate the impacts and spread of COVID-19 using a variety of indicators. Our research specifically investigates the relationship between political bias and socioeconomic indicators in the context of the pandemic, employing statistical methods. This section describes works that address dimension similar to those in our study.

Given the rapid emergence of COVID-19, most researchers have focused on detection, identification, and prediction of the virus using sophisticated Machine Learning (ML) algorithms [Tiwari et al. 2022]. For example, [Ak et al. 2022] and many others utilized ML techniques, such as Random Forest, to analyze the correlation between studied variables and the pandemic's outcomes.

Several studies have explored the influence of political factors on COVID-19-related deaths and cases. [Cipullo and Le Moglie 2022] demonstrated that electoral campaigns had a significant impact on COVID-19 deaths in Brazil. Similarly, [Menuzzo et al. 2021] showed that the discourse of municipalities and their respective mayors influenced the pandemic's outcomes. In the U.S., [Rönn et al. 2023] found a strong correlation between COVID-19 vaccination coverage and the proportion of votes for the Democratic candidate in the 2020 presidential election. Conversely, other studies have investigated how the pandemic itself affected electoral outcomes, as observed in Brazil [Constantino et al. 2021], Ghana [Ayifah and Ayifah 2023], and the USA [Wu 2023].

Similar to our work, the study by [Aron and Muellbauer 2022] provides strong evidence that states with higher shares of Democratic votes experienced lower excess mortality. This study employed a spatial analysis using demographic, socioeconomic, and voting data, although it was limited to the first year of the pandemic and conducted at the state level in the U.S. [Desmet and Wacziarg 2022] also examined various socioeconomic indicators alongside political indicators in the U.S., using estimation and simulations to understand the spatial variation of COVID-19, albeit with a different methodological focus compared to our study.

In the context of Brazil, [Xavier et al. 2022] investigated how political partisanship and socioeconomic factors influenced COVID-19 outcomes at the local level. Their work, similar to ours, utilized HDI and presidential election data, albeit from 2018, along with data from the Mortality Information System. They concluded that income inequalities and health infrastructure vulnerabilities shaped the dynamics of the first wave of COVID-19, while the second wave was more explicitly shaped by the political choices of municipalities. Our study builds on these findings by analyzing a longer period of data (2020-2024) and incorporating the 2020 and 2022 elections. We found that COVID-19 outcomes initially correlated more strongly with the HDI index than with party choices, but in the latter half of the period, the correlation with party choices in the 2022 elections became stronger. The exceptional nature of the 2018 elections is further discussed in [Barberia et al. 2022].

The work of [Lima et al. 2024] also examined the association between COVID-19 infections and deaths in Brazil with political choices, analyzing excess deaths during 2020 and 2021 alongside other indicators such as vaccination rates and sociodemographic information. Unlike our study, they compared the 2018 and 2022 presidential elections. As highlighted by [Barberia et al. 2022], the 2018 elections have significant limitations due to their exceptional nature [Amaral 2020, Rennó 2020]. The reliance on excess mortality as a metric has its acknowledged limitations, such as disregarding trends in population size and mortality, and delays in death reporting that render the data provisional and incomplete.

Other significant works that employ different methodologies, variables, and study periods include the study by [Fernandes and de Almeida Lopes Fernandes 2022], which showed that the number of COVID-19 cases and deaths in 2020 was higher in areas where the president received greater electoral support during the 2018 presidential elections, using negative binomial models. Additionally, [Sott et al. 2022] analyzed the role of government, the Brazilian health system, and the broader economic and social impacts triggered by the pandemic.

3. Methodology

In this section, we detail the methodology adopted to investigate the influence of political and socioeconomic factors on the spread of COVID-19 in Brazilian cities from 2020 to 2024. Our approach is multidimensional, integrating temporal, spatial, and statistical analyses to identify patterns and correlations that provide meaningful insights. The main steps of the study, from data collection and preparation to the analyses performed, are described in the following sections.

3.1. Data Collection

For this study, electoral data from the 2020 municipal elections and the 2022 presidential elections were collected from the Superior Electoral Court (TSE). This data includes the number of votes for each candidate in each city. Figure 1 presents six maps depicting

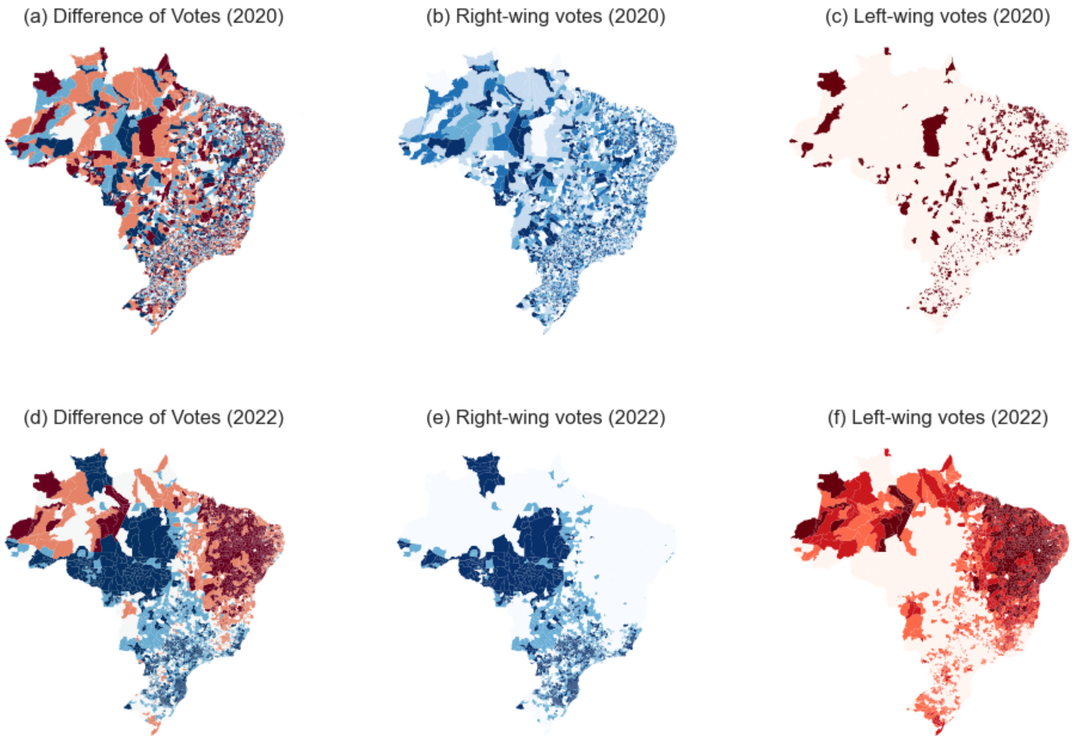

Figure 1. Spatial Distribution of Voting Patterns in Brazil (2020 and 2022)

the spatial distribution of voting patterns across Brazilian cities in the 2020 and 2022 elections. The maps illustrate the differences in votes, as well as the distribution of right-wing and left-wing votes for both election years. A comparison of these maps reveals a noticeable consistency in voting patterns across the years. Areas that showed strong support for right-wing or left-wing candidates in 2020 generally maintained similar voting behaviors in 2022. The northern and northeastern regions tend to have higher support for left-wing candidates, as indicated by the darker red areas, while the southern and southeastern regions show stronger support for right-wing candidates, evident from the darker blue areas.

Information on COVID-19 cases and deaths in all Brazilian cities was obtained from the Ministry of Health and state health departments. The analysis period spans from March 2020 to January 2024. Figure 2 presents eight maps depicting the spatial distribution of COVID-19 cases and deaths across Brazilian municipalities. These maps utilize a grayscale to represent the intensity of cases and deaths, with darker shades indicating higher values.

Additionally, data on the 2010 Human Development Index (HDI) for each city were obtained from the Brazilian Institute of Geography and Statistics (IBGE). All data were organized and integrated into a single dataset, ensuring consistency and facilitating subsequent analysis.

3.2. Data Preparation

After data collection, all information was processed and organized for further analysis. We cleaned the data by removing duplicates, handling missing values, and standardizing

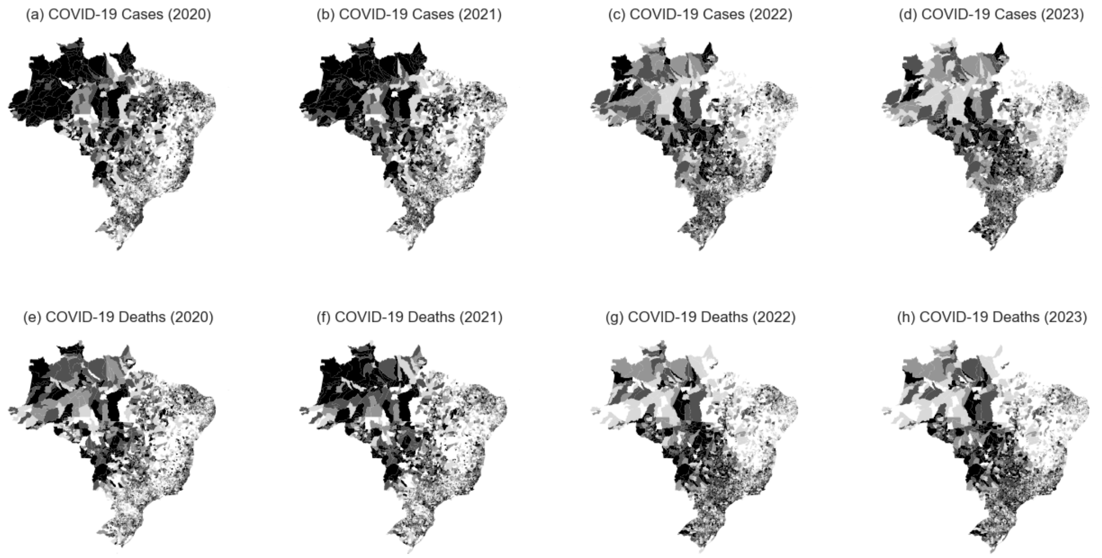

Figure 2. Spatial Distribution of COVID-19 Cases and Deaths in Brazil (2020-2023)

formats into a single data frame, using the municipality code as a key. This process was applied consistently across the three databases: COVID-19, electoral data, and socioeconomic indicators. Following this step, we consolidated the data on a monthly basis to facilitate temporal analysis.

To analyze the influence of political variables on the spread of COVID-19, we organized the electoral data as follows:

- **Candidate Classification:** We selected only the two main candidates from each election (2020 municipal and 2022 presidential) based on classifications developed by other studies in the community.
- **Political Categorization:** Candidates were classified into three categories: right, left, and center, according to their declared political ideology and advocated policies.
- **First Round Votes:** Only votes from the first round of each election were used to ensure a cleaner analysis, minimizing the influence of political alliances formed in the second round.
- **Control Variable:** Candidates classified as center were used solely as a control variable to understand the influence of political polarization (right versus left) on the impact of COVID-19.

Unlike the 2022 presidential election, which featured a clear division between left-wing and right-wing candidates, the dynamics of municipal elections are more complex, with many candidates not necessarily aligning with traditional left or right political orientations. To address this challenge, we employed the classification approach cited by the 2020 Election Observatory [Bolognesi et al. 2020], which draws on research from various Brazilian universities. This approach categorizes each party along a spectrum ranging from more left-wing to more right-wing, with a central position as well. Additionally, we conducted manual searches on the TSE website to account for name changes of some parties during the period between 2020 and 2024.

3.3. Temporal, Spatial, and Statistical Analysis

To understand the evolution of COVID-19 indicators and their correlations with political and socioeconomic variables over time, we employed a combination of time series analysis techniques, spatial analysis, and statistical and machine learning methods.

- **Time Series Decomposition:** We performed time series decomposition to visualize the seasonal and trend components of the COVID-19 case and death data. ARIMA models were applied to identify patterns, detect anomalies, and assess the predictability of the time series.
- **Spatial Analysis:** Spatial analysis was conducted to identify geographic patterns and spatial correlations among the variables of interest. We generated maps depicting the spatial distribution of COVID-19 cases and deaths, along with electoral and socioeconomic variables.
- **Statistical and Machine Learning Methods:** We utilized statistical and machine learning methods, including Pearson correlation, Random Forest Regression [Breiman 2001], and the K-Means clustering algorithm [Hartigan et al. 1979]. Pearson correlation was calculated to identify linear relationships between variables, while Random Forest Regression was used to uncover non-linear spatial patterns. K-Means clustering was employed to group the data into clusters, revealing patterns and relationships between COVID-19 cases and the variables of interest, effectively capturing complex and non-linear interactions.

Thus, our methodology integrates a range of analytical and visualization techniques to deliver a comprehensive, multidimensional analysis of the data. In the following sections, we present the results of these analyses, discussing the findings and their implications for understanding the spread of COVID-19 within the Brazilian political and socioeconomic context.

4. Results

4.1. Seasonality and Trend Analysis with ARIMA

Figure 3 was generated using the ARIMA model [Box et al. 2015] and illustrates the seasonality and trends in COVID-19 data from 2020 to 2024. The blue line represents the average monthly deaths due to COVID-19 (in thousands), while the orange line depicts the predictions made by the ARIMA model for the average monthly deaths. The decomposition of the time series with ARIMA reveals general trends of increase and decrease in COVID-19 cases and deaths over the analyzed period. Seasonality is evident, with recurring peaks potentially linked to factors such as seasonal changes, public events, or variations in health policies. These trends may also reflect the impact of government interventions and the population's adherence to preventive measures.

The ARIMA model appears to capture the overall trend of decreasing deaths following the peak in 2021. However, the model's predictions for 2024 and 2025 are very close to zero, suggesting that it does not anticipate any significant changes or new waves of COVID-19 deaths in the future.

4.2. Temporal Correlation

Figure 4 illustrates the correlation over time (2020 to 2024) between COVID-19 cases and deaths, electoral variables (votes for right- and left-wing candidates), and the HDI

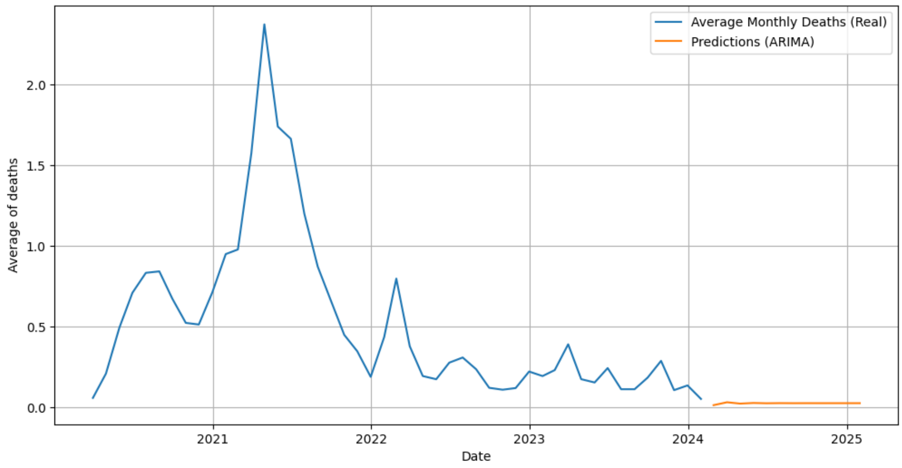

Figure 3. Average Monthly Deaths from COVID-19 and Predictions with ARIMA

as a control. The HDI is a composite index that measures average achievements in three fundamental aspects of human development: life expectancy, education, and per capita income. We conducted separate analyses for each component of the HDI, but for simplicity, and because the results were consistent across these components, we chose to use only the global HDI. The graph shows the month-to-month Pearson correlation between the variables of interest, with the x-axis representing the time period (2020 to 2024) divided by months, and the y-axis representing the correlation values, which range from -1 to 1.

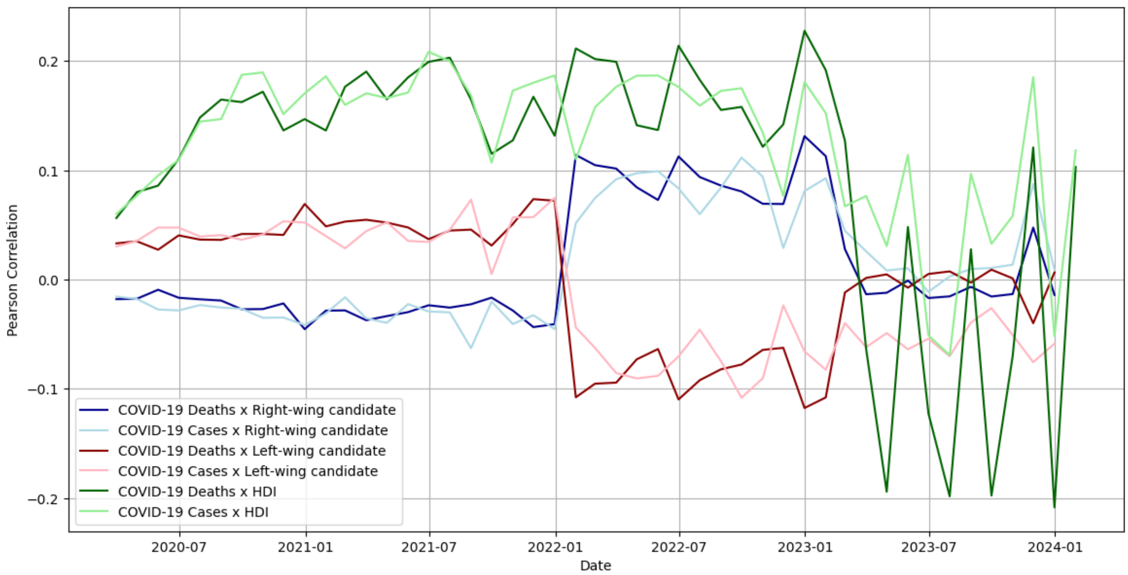

Figure 4. Monthly Correlation between COVID-19 Variables and Political Leaning

The correlation between votes for right-wing candidates and COVID-19 cases fluctuates over time, with peaks during specific periods, potentially linked to political events or changes in public health policies. The correlation with votes for left-wing candidates also varies but generally tends to be negative, indicating that regions with greater support for left-wing candidates may have responded differently to the pandemic. The correlation with the HDI, used as a control, highlights the socioeconomic influence on the spread of COVID-19. While political alignment shows weak and variable correlations, the HDI exhibits a consistently stronger positive correlation with COVID-19 cases and deaths.

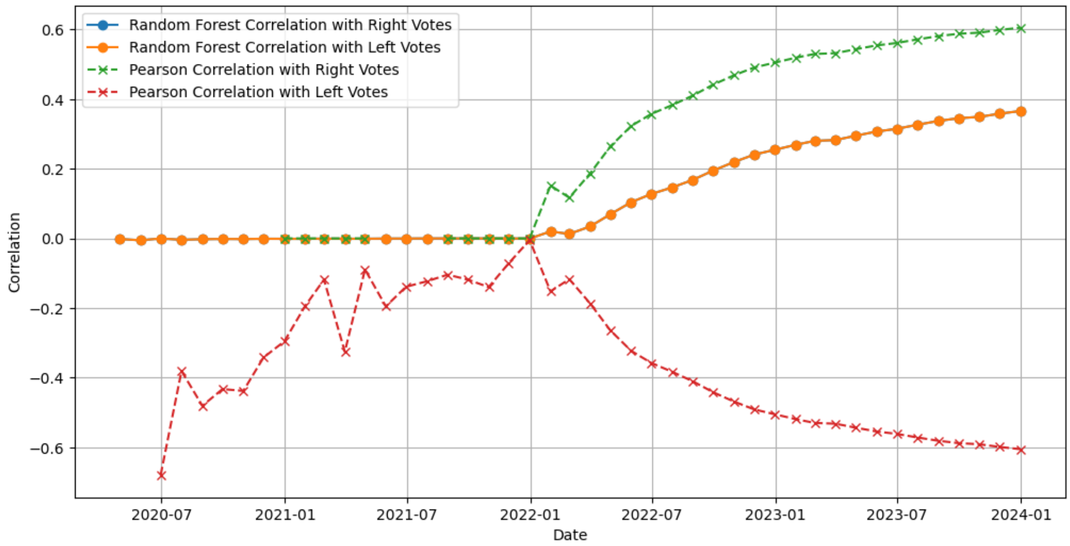

Figure 5. Correlation between Average Monthly Deaths and Votes Over Time (Random Forest and Pearson)

Figure 5 presents the correlation between monthly average COVID-19 death rates and votes for right-wing and left-wing parties, using two different methods: Random Forest Regression and Pearson Correlation. Random Forest is a machine learning method that constructs multiple decision trees during training and averages their predictions to enhance accuracy and control over-fitting. It is particularly effective for capturing complex, non-linear interactions between variables, offering a more nuanced understanding of the data. In contrast, Pearson Correlation quantifies the linear relationship between variables, providing a clear measure of direct associations in a classical statistical context.

Both Random Forest and Pearson Correlation indicate a growing positive correlation between right-wing votes and COVID-19 death rates over time. The correlation with left-wing votes remains near zero in the Random Forest model but shows a decreasing negative trend in the Pearson correlation, suggesting differing perspectives on the strength and direction of these relationships. This suggests that the Random Forest model did not identify a significant relationship between left-wing votes and average monthly death rates over time. The analysis highlights that the correlation between electoral variables and COVID-19 outcomes can be influenced by non-linear factors and complex interactions, providing a deeper and more detailed understanding of the relationships within the data.

4.3. Spatial Analysis with K-Means

K-Means [Hartigan et al. 1979] is an unsupervised machine learning algorithm that groups data points into a specified number of clusters based on their similarity. This algorithm assigns each data point to one of K clusters by minimizing the within-cluster sum of squares, effectively grouping data points that are similar to each other. The first scatter plot, labeled as **a** in Figure 6, represents the votes for left-wing candidates. Cities were grouped based on the percentage of votes and average monthly COVID-19 deaths, utilizing K-Means to identify patterns and similarities in how political preferences correlate with pandemic impact.

The clusters reveal that municipalities with moderate to high left-wing votes generally exhibit low to moderate COVID-19 death rates, while outliers with extremely high death rates are observed across all levels of left-wing votes, indicating that other factors might influence the death rates. Similarly, the second scatter **b** plot shows the clustering for right-wing votes, where municipalities with low to moderate right-wing voters tend to have lower COVID-19 death rates. However, high death rates are observed in municipalities with both low and high right-wing votes, suggesting other underlying factors impacting mortality rates.

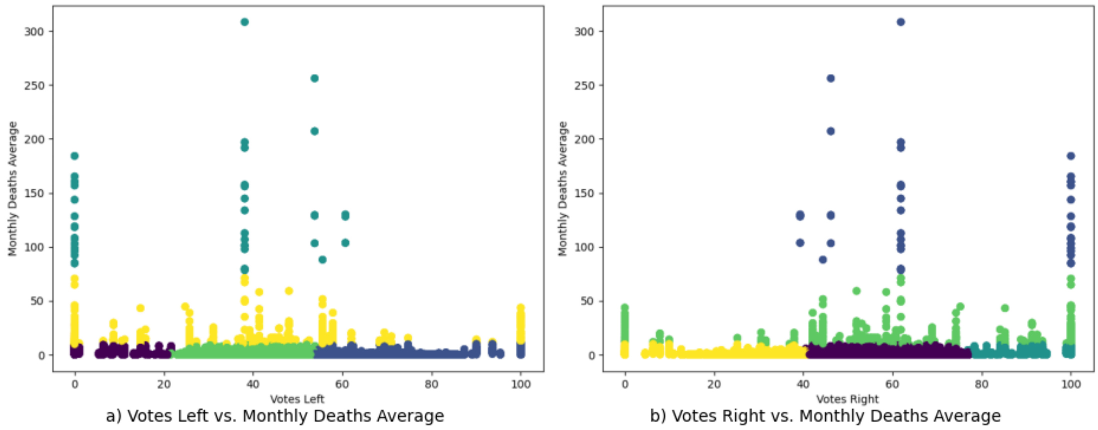

Figure 6. K-Means Clustering of COVID-19 and Political Indicators

A comparison of both plots reveals that municipalities with extreme COVID-19 death rates are distributed across various levels of political alignment, underscoring the complex relationship between political voting patterns and COVID-19 outcomes. The presence of outliers with high death rates in both plots suggests that extreme outcomes are not solely determined by political alignment but are likely influenced by a combination of factors, including healthcare infrastructure, socioeconomic conditions, and public health policies. However, a slight difference is observed, with a greater concentration of cities tending towards right-wing votes, indicating that municipalities with higher COVID-19 death rates are more often associated with right-wing voting patterns. This analysis highlights the need for a multifaceted approach to understanding the impacts of COVID-19, taking into account broader contexts such as healthcare resources and socioeconomic status, to develop effective interventions and policies. Further research is essential to unravel these complex relationships and to provide a comprehensive understanding of the factors driving COVID-19 mortality rates in different regions.

5. Discussion and Future Works

The COVID-19 pandemic has profoundly impacted global health, economies, and societies, with political and socioeconomic factors playing a crucial role in shaping responses and outcomes. In this study, we conducted a multidimensional analysis to investigate the relationships between political leanings, socioeconomic indicators, and the spread of COVID-19 in Brazilian cities.

By employing a range of statistical and machine learning methods, including Pearson correlation, the Random Forest algorithm, ARIMA forecasting, and K-Means clustering, we identified significant patterns and relationships between COVID-19 cases, deaths, and various socioeconomic and political factors.

Our findings reveal that the relationship between COVID-19 outcomes and political voting patterns is complex and multifaceted. While the correlation between COVID-19 variables and votes for right-wing candidates is weak and fluctuating, the correlation with left-wing votes is slightly more stable and negative. This suggests that regions with more right-wing votes might be experiencing higher COVID-19 deaths and cases, indicating the potential influence of political alignment on public health outcomes.

Additionally, the positive correlation between the Human Development Index (HDI) and COVID-19 cases and deaths highlights that regions with higher human development levels are experiencing higher infection and mortality rates. This may be attributed to better reporting and healthcare infrastructure in high-HDI areas, leading to more accurate data. The temporal fluctuations in these correlations reflect the dynamic nature of the pandemic, influenced by factors such as public health policies, vaccination rates, and the emergence of new variants.

The spatial distribution analysis underscores the need for targeted public health interventions and resource allocation to address persistent hotspots and mitigate the impact of the pandemic. Our analysis shows that certain municipalities consistently experience higher COVID-19 death rates, suggesting ongoing challenges in these areas that require focused attention.

Overall, this study highlights the importance of a multifaceted approach to understanding the impacts of COVID-19, considering the broader context of socioeconomic conditions, political alignment, and public health infrastructure. The insights gained from this analysis can inform policymakers and public health officials in developing effective strategies to combat the pandemic and address socio-political disparities. Further research is necessary to explore these complex relationships in greater detail and to provide a more comprehensive understanding of the factors driving COVID-19 mortality rates in different regions.

As part of our future work, we plan to investigate the influence of additional variables, such as access to healthcare services and urban mobility, and to apply our methodology to other countries or regions to validate our findings. This expansion will allow us to better understand the broader applicability of our results and to refine our strategies for managing health crises in diverse settings.

Acknowledgment

The authors would like to thank CAPES for the research support provided to Rôney Reis.

References

Ak, Ç., Chitsazan, A. D., Gönen, M., Etzioni, R., and Grossberg, A. J. (2022). Spatial prediction of COVID-19 pandemic dynamics in the United States. *ISPRS International*

Journal of Geo-Information, 11(9):470.

Amaral, O. E. d. (2020). The victory of Jair Bolsonaro according to the Brazilian electoral study of 2018. *Brazilian Political Science Review*, 14:e0004.

Aron, J. and Muellbauer, J. (2022). Excess mortality versus COVID-19 death rates: A spatial analysis of socioeconomic disparities and political allegiance across U.S. States. *Review of Income and Wealth*, 68(2):348–392.

Ayifah, R. N. Y. and Ayifah, E. (2023). COVID-19 lockdown policy and national elections: A quasi-experimental analysis of Ghana's 2020 election. *International Social Science Journal*, 73(248):685–704.

Barberia, L., Moreira, N. d. P., Carvalho, R. d. J., Oliveira, M. L. C., Rosa, I. S. C., and Zamudio, M. (2022). The relationship between ideology and COVID-19 deaths: what we know and what we still need to know. *Brazilian Political Science Review*, 16:e0002.

Bolognesi, B., Ribeiro, E., and Codato, A. (2020). Esquerda, centro ou direita? Como classificar os partidos no Brasil. *Observatório das Eleições*.

Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). Time series analysis: forecasting and control. *John Wiley & Sons*.

Breiman, L. (2001). Random forests. *Machine learning*, 45:5–32.

Cipullo, D. and Le Moglie, M. (2022). To vote, or not to vote? Electoral campaigns and the spread of COVID-19. *European Journal of Political Economy*, 72:102118.

Constantino, S. M., Cooperman, A. D., and Moreira, T. M. (2021). Voting in a global pandemic: assessing dueling influences of COVID-19 on turnout. *Social Science Quarterly*, 102(5):2210–2235.

Desmet, K. and Wacziarg, R. (2022). JUE insight: Understanding spatial variation in COVID-19 across the United States. *Journal of urban economics*, 127:103332.

Fernandes, G. A. d. A. L. and de Almeida Lopes Fernandes, I. F. (2022). Populism and health. An evaluation of the effects of right-wing populism on the COVID-19 pandemic in Brazil. *PLoS One*, 17(12):e0269349.

Hartigan, J. A., Wong, M. A., et al. (1979). A k-means clustering algorithm. *Applied statistics*, 28(1):100–108.

Lima, E. E. C. d., Costa, L. C. C. d., Souza, R. F., Rocha, C. O. d. E., and Ichihara, M. Y. T. (2024). Presidential election results in 2018-2022 and its association with excess mortality during the 2020-2021 COVID-19 pandemic in brazilian municipalities. *Cadernos de Saúde Pública*, 40:e00194723.

Menuzzo, V. A., Santanchè, A., and Gomes-Jr, L. (2021). Evaluating the cohesion of municipalities' discourse during the COVID-19 pandemic. In *Anais do XXXVI Simpósio Brasileiro de Bancos de Dados*, pages 295–300. SBC.

Rennó, L. R. (2020). The Bolsonaro voter: issue positions and vote choice in the 2018 brazilian presidential elections. *Latin American Politics and Society*, 62(4):1–23.

Rönn, M. M., Menzies, N. A., and Salomon, J. A. (2023). Vaccination and voting patterns in the U.S.: analysis of COVID-19 and flu surveys from 2010 to 2022. *American Journal of Preventive Medicine*, 65(3):458–466.

Sott, M. K., Bender, M. S., and da Silva Baum, K. (2022). COVID-19 outbreak in Brazil: health, social, political, and economic implications. *International Journal of Health Services*, 52(4):442–454.

Tiwari, S., Chanak, P., and Singh, S. K. (2022). A review of the machine learning algorithms for COVID-19 case analysis. *IEEE Transactions on Artificial Intelligence*, 4(1):44–59.

Wu, S. (2023). The spatial data analysis of determinants of U.S. presidential voting results in the rustbelt states during the COVID-19 pandemic. *ISPRS International Journal of Geo-Information*, 12(6):212.

Xavier, D. R., e Silva, E. L., Lara, F. A., e Silva, G. R., Oliveira, M. F., Gurgel, H., and Barcellos, C. (2022). Involvement of political and socio-economic factors in the spatial and temporal dynamics of COVID-19 outcomes in Brazil: a population-based study. *The Lancet Regional Health–Americas*, 10.