[Spora: A Journal of Biomathematics](https://ir.library.illinoisstate.edu/spora)

Volume 10 | Article 10

2024

Myth or Fact? An Analysis of COVID-19 Deaths in Red or Blue States of America

Kwabena S. Duku
Illinois State University, ksduku@ilstu.edu

Nishat Ara Mozid
Illinois State University, nishatmozid@gmail.com

Olcay Akman
Illinois State University, oakman@ilstu.edu

Recommended Citation

Duku, Kwabena S.; Mozid, Nishat Ara; and Akman, Olcay (2024) "Myth or Fact? An Analysis of COVID-19 Deaths in Red or Blue States of America," *Spora: A Journal of Biomathematics*: Vol. 10, 91–94.
DOI: https://doi.org/10.61403/2473-5493.1092
Available at: https://ir.library.illinoisstate.edu/spora/vol10/iss1/10

Myth or Fact? An Analysis of COVID-19 Deaths in Red or Blue States of America

Cover Page Footnote

We (the first two authors) would like to thank our professor Dr. Akman for guiding us throughout the project. We would also like to thank the Intercollegiate Biomathematics Alliance (IBA) for giving us the opportunity to present our work at the BEER symposium.

Myth or Fact? An Analysis of COVID-19 Deaths in Red or Blue States of America[†]

Kwabena S. Duku[1], Nishat Ara Mozid[1], Olcay Akman[1,*]

[1]Department of Mathematics, Illinois State University, Normal, IL

[*]Correspondence: oakman@ilstu.edu

[†]This article was published as part of the Proceedings of the 16th Symposium on BEER and reviewed for editorial clarity only.

Keywords: ESOM clustering, correlation matrix, K-means clustering, prediction, logistic classification

1 Introduction

The COVID-19 pandemic has underscored the intricate relationship between political dynamics and public health outcomes. This study delves into the correlation between political leanings and COVID-19 mortality, with a specific emphasis on states affiliated with Republican and Democratic governance. The motivation for this research stems from the observed variations in pandemic outcomes across regions with differing political affiliations, particularly noted in earlier studies indicating higher death rates in right-leaning "red" counties compared to left-leaning "blue" counties.

To elucidate these disparities, our investigation utilizes sophisticated analytical techniques, including emergent self-organizing map (ESOM), cluster analysis [2], and the logistic classification. These methods allow for a comprehensive examination of COVID-19 properties, risk patterns, and associated levels of risk. The inclusion of variables such as poverty rates, education levels, vaccination rates, and demographics enriches our analysis, considering the multifaceted nature of the pandemic's impact.

The logistic classification acts as a unifying element, summarizing findings from ESOM, and cluster analysis. Through this multi-method approach, we aim to offer a concise and nuanced understanding of the COVID-19 risk landscape in politically diverse states. This exploration holds implications for public health strategies and policy decisions, emphasizing the need for tailored approaches that account for socio-political nuances in mitigating the impact of health crises, highlighting the importance of planning ahead for future challenges.

2 Methods

2.1 Study System

In our study, we collected data from three different types of American states in terms of the variables: full vaccination rate, health score, IQ score, education score, average wage, income tax, poverty rate, and death rate. Data were collected from online newspaper source, National Institute of Health (NIH) website [3], and scientific literature [1]. Three different types of American states were red, blue, and purple states with Republican, Democratic, and fluctuating dominance, respectively.

2.2 Exploratory Analysis

At first we constructed the correlation matrix to summarize the overall dataset and proposed a logistic classification using variables for full vaccination rate, education rate, IQ score, and income tax. The logistic classification can predict the chance for an American state being red or blue.
Then we focused on red and blue states to examine the risk of COVID-19-related mortality in states with differing political affiliations using a list of unsupervised machine learning techniques: emergent self-organizing map (ESOM), K-means clustering, principal component analysis. We also used one supervised machine learning technique: logistic classification.

3 Results

3.1 Correlation Matrix

Figure 1 represents the pairwise correlations between the listed variables, where orange and purple correspond to positive and negative correlation, respectively. The intensity of the color is proportional the strength of correlation. The vaccination rate demonstrates a negative correlation (-0.634) with mortality, indicating that lower vaccination rates coincide with higher mortality rates. Likewise, the education score shows a positive correlation with wage, IQ, and health scores. These findings underscore the complex relationship between socioeconomic factors,

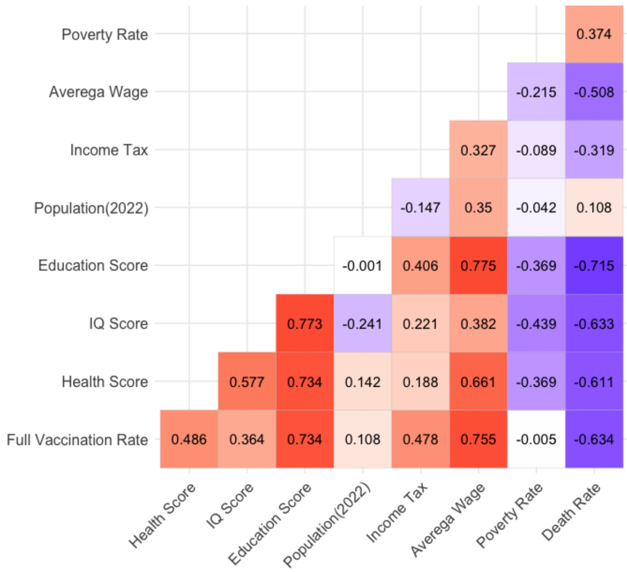

Figure 1: Correlation plot of variables.

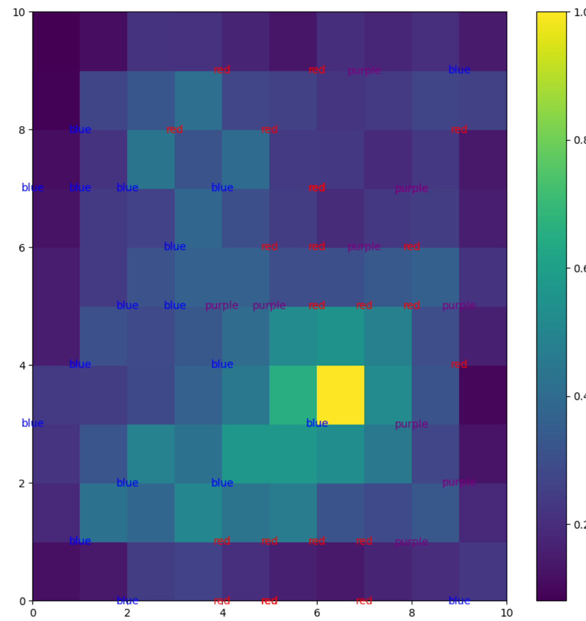

Figure 2: Emergent self-organizing map (ESOM).

healthcare access, and regional demographics, warranting further investigation.

3.2 Logistic Classification

Here we are employing log-logistic regression to classify the political tendency of a state Democratic or, Republican based on the COVID-19 related factors that are selected based on our model building algorithm. The model below provides an estimated "color" of a state indicating blue for Democratic and red for Republican when variables Vaccination Rate, Education Score, IQ Score, and Income Tax are used as inputs.

$$\begin{aligned}\log(\text{State Color}) = &-231.21156 \\ &- 0.060427 * (\text{Full Vaccination Rate}) \\ &- 0.43745 * (\text{Education Score}) \\ &+ 2.920911 * (\text{IQ Score}) \\ &- 0.06874 * (\text{Income Tax})\end{aligned}$$

In this comprehensive examination, our focus is on the intricate process of predicting the political color of a state, employing a sophisticated logistic regression model. With an impressive accuracy rate of 78%, our analysis delves into a diverse array of state-specific attributes. These include crucial metrics such as vaccination rate, education score, IQ score, and income tax, all meticulously considered alongside the state's color.

The correlation matrix showed significant insights within our data. For instance, we discovered a negative correlation of (-0.369) between Education Score and Poverty Rate. Consequently, Poverty Rate is not included in the model, due to its correlation with Education Score. Similarly, we found a correlation of (0.734) between Education Score and Health Score, leading us to exclude Health Score from our model. Finally, we observed a strong correlation of (0.755) between Average Wage and Full Vaccination Rate, resulting in the exclusion of Average Wage from our model. Recognizing the potential pitfalls of multi co-linearity, we took deliberate steps to exclude certain variables, ensuring the robustness and reliability of our analysis. This strategic decision not only enhances the accuracy of our predictions but also underscores our unwavering commitment to precision in political color forecasting.

Furthermore, by incorporating a wide spectrum of state-specific metrics, we aim to capture the nuanced interplay between socioeconomic factors, public health dynamics, and educational attainment, all of which contribute to the intricate tapestry of a state's political landscape. Through this holistic approach, we strive to offer insights that transcend mere statistical analysis, providing a deeper understanding of the complex forces shaping political affiliations at the state level.

3.3 ESOM Clustering

Emergent self-organizing map (ESOM) is a tool for clustering, data analysis and visualization. In ESOM clustering the maps (neurons in our case) visualize distance structure of multi-dimensional data set.

Within Figure 2, each square symbolizes a neuron, serv-

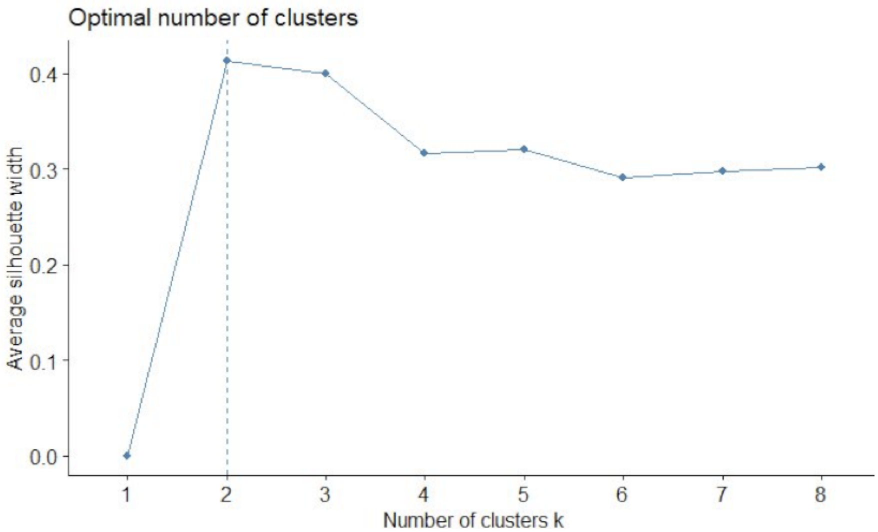

Figure 3: Elbow plot.

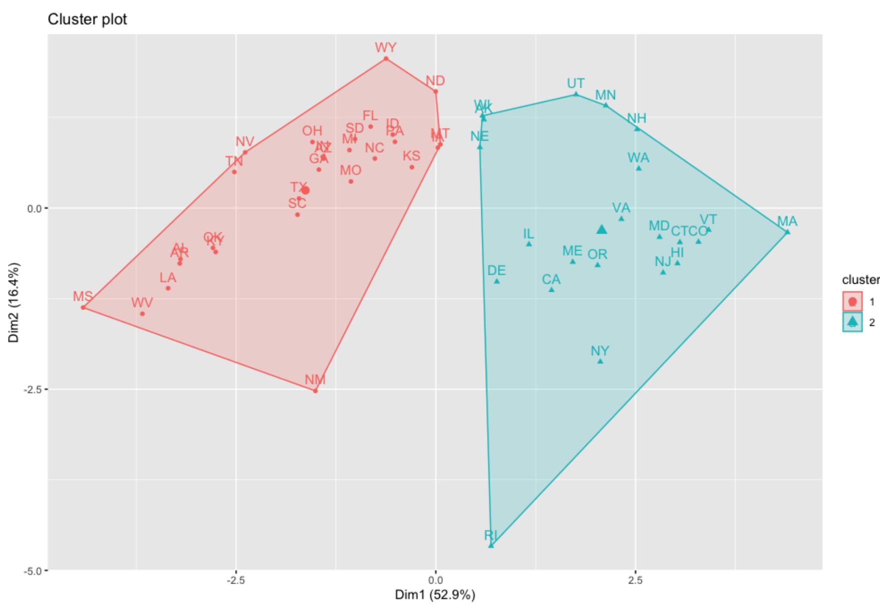

Figure 4: K-means clustering plot.

ing as a computational unit in the clustering process. The assignment of "red" and "blue" labels to these neurons corresponds to the grouping of data points based on their similarity, thus facilitating the clustering algorithm. While this method enables the identification of cohesive clusters, it's essential to acknowledge the possible existence of outliers, which may deviate from the main clusters.

Upon closer examination of the plot, a discernible pattern emerges: the majority of blue states tend to cluster tightly together, indicating a high level of similarity in their characteristics. This cohesive clustering suggests a degree of uniformity or shared traits among these states. Conversely, the scattering of red states across the plot suggests greater variability in their attributes, as represented by a vector encompassing the considered variables. This variance among red states hints at a broader spectrum of characteristics or a less cohesive grouping compared to their blue counterparts.

3.4 K-means Clustering

K-means clustering assigns similar data points to a specified number of clusters. The clustering algorithm finds out similar data points, puts them in the same set, and visualizes them accordingly.

The initial step in our analysis involved applying the K-means clustering technique to identify the optimal number of clusters for the dataset. Despite the dataset comprising three distinct state colors (red, blue, and purple), the elbow method, a commonly used heuristic, suggested that two clusters shown in Figure 3 would best capture the underlying structure of the data. This decision was based on minimizing the within-cluster sum of squares while balancing the complexity of the model.

Following the determination of the appropriate number of clusters, the K-means algorithm in Figure 4 was applied to partition the data accordingly. Within the resulting clusters, a predominant grouping of blue states was observed in one cluster, reflecting a higher degree of similarity in their characteristics. Similarly, red states tended to aggregate in another cluster, though with some instances of misclassification, where states of one color were erroneously assigned to the cluster dominated by the other color. This observation underscores the inherent complexity and variability within the dataset, warranting further exploration to refine the clustering process and enhance its accuracy.

4 Conclusion

These analyses uncover a nuanced correlation between political orientations and COVID-19 mortality rates. From our analysis, on average, red states, known for their alignment with politically right-learning ideologies, demonstrate a closer association with neurons linked to higher death rates [4], contrasting with the spatial arrangement observed in blue states, which tend to lean left politically. This suggests that red states faced a more pronounced impact from COVID-19, experiencing elevated mortality rates compared to their blue counterparts.

These findings underscore the potential impact of political affiliations on the severity of the pandemic's consequences. Factors such as public health strategies, healthcare infrastructure disparities, and socioeconomic policy variations may differ based on the policy decisions made under different political ideologies. These differences may contribute to the observed disparities. Although regional disparities in population density, demographics, and adherence to preventive measures could further shape the outcomes, our focus was on the impact of political affiliations on COVID-19.

These analyses illuminate the intricate interplay between political ideologies and public health outcomes, emphasizing the necessity for a comprehensive understanding of socio-political factors in effectively addressing health crises. It underscores the significance of the efforts across political spectra to mitigate the adverse impacts of pandemics and ensure equitable health outcomes for all communities.

Acknowledgment

We would like to thank the Intercollegiate Biomathematics Alliance for its support during this project.

References

[1] Wallace, J., Goldsmith-Pinkham, P., Schwartz, J.L. (2023). Excess death rates for republican and democratic registered voters in Florida and Ohio during the COVID-19 pandemic. *JAMA Intern Med*, 183(9):916–923.

[2] Choi, Yoon-Jung., Park, Mi-jeong., and Park, Soo Jin. (2021). Types of COVID-19 clusters and their relationship with distancing in the Seoul metropolitan area, South Korea. *International Journal of Infectious Disease*, 106:363–369.

[3] Pesta, BJ. (2022) Udated IQ and well-being scores for the 50 U.S. states. *www.ncbi.nih.gov*, 10(1): 15.

[4] C. Dominik Güss, Lauren Boyd, Kelly Perniciaro, Danielle C. Free, J.R. Free, Ma. Teresa Tuason. (2023). The politics of COVID-19: Differences between U.S. red and blue states in COVID-19 regulations and deaths. *Health Policy OPEN*, 5:2590–2296.