2026 HAWAII UNIVERSITY INTERNATIONAL CONFERENCES
SCIENCE, TECHNOLOGY & ENGINEERING, ARTS, MATHEMATICS & EDUCATION JUNE 09 - 11, 2026
PRINCE WAIKIKI RESORT, HONOLULU, HAWAII

SPATIO-TEMPORAL STATISTICAL MODELING OF SUICIDE MORTILITY IN THE UNITED STATES: STUCTURAL BREAKS, SPATIAL DEPENDENCE, AND EXPLORATORY MACHINE LEARNING

RAJAPAKSHA, CHARU
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WISCONSIN-WHITEWATER
WHITEWATER, WISCONSIN

KARUNASENA, THILINI
DEPARTMENT OF MATHEMATICS AND STATISTICS
TEXAS TECH UNIVERSITY
LUBBOCK, TEXAS

Dr. Charu Rajapaksha
Department of Mathematics
University of Wisconsin-Whitewater
Whitewater, Wisconsin
Ms. Thilini Karunasena
Department of Mathematics & Statistics
Texas Tech University
Lubbock, Texas

Spatio-Temporal Statistical Modeling of Suicide Mortality in the United States: Structural Breaks, Spatial Dependence, and Exploratory Machine Learning

Abstract

Suicide remains a major public-health concern in the United States, with substantial variation across time, demographic groups, socio-economic conditions, and geographic regions. This study develops an integrated spatio-temporal analytical framework to investigate long-term suicide mortality patterns in the United States using national and state-level data spanning 1980 to 2021. The framework combines temporal trend analysis, structural-break detection, time-series forecasting, econometric modeling, spatial statistics, clustering analysis, and exploratory interpretable machine learning techniques.

The results identify a major structural transition in national suicide trends near the late 1990s, shifting from a declining trajectory toward a persistent upward pattern throughout the 21st century. ARIMA forecasting further supports the persistence of elevated suicide mortality during recent years. Demographic analyses reveal substantial heterogeneity across age groups and gender, with particularly elevated suicide mortality among middle-aged populations and males.

Socio-economic analyses, including regression and ARDL modeling, identify unemployment and demographic aging as important factors associated with suicide mortality, while also demonstrating substantial temporal persistence in suicide-rate dynamics. Spatial analyses reveal strong and persistent geographic clustering, particularly across western and mountain-region states. Moran's I statistics confirm statistically significant positive spatial autocorrelation, and

clustering analyses further identify distinct geographic suicide-risk regimes across the United States.

Exploratory interpretable machine learning analyses additionally reinforce the importance of unemployment and demographic aging in understanding long-term suicide mortality dynamics. Overall, the findings highlight the importance of integrating temporal, demographic, socio-economic, spatial, and interpretable machine learning perspectives for understanding complex suicide-mortality patterns and developing future public-health monitoring frameworks.

Keywords: suicide mortality; spatio-temporal analysis; structural break; ARIMA; ARDL; Moran's I; spatial autocorrelation; clustering analysis; interpretable machine learning; public health

1 Introduction

Suicide is a complex and persistent public-health problem in the United States and worldwide. Its occurrence is influenced by a combination of demographic, socio-economic, geographic, psychological, behavioral, and temporal factors [1–3]. Although national suicide rates provide an overall measure of public-health burden, aggregate trends may conceal substantial heterogeneity across age groups, gender, race, and geographic regions [4–6].

Previous statistical studies of suicide mortality have frequently focused on descriptive analyses, correlation studies, or conventional linear regression models [7–9]. While such approaches are valuable for identifying broad associations, they may be insufficient for capturing nonlinear temporal patterns, structural regime shifts, and geographic dependence structures that often characterize complex public-health data. Similarly, geographic visualization can reveal regional patterns in suicide mortality, but formal spatial statistical methods are required to determine whether such patterns represent statistically meaningful clustering rather than random geographic variation [10, 11].

This study develops an integrated spatio-temporal analytical framework for investigating suicide mortality in the United States. The proposed framework combines statistical trend analysis, structural-break detection, segmented regression, time-series forecasting, econometric modeling, spatial analysis, clustering methods, and exploratory interpretable machine learning techniques to examine long-term suicide mortality dynamics and associated socio-economic patterns. The analysis addresses four primary research questions:

1. How have suicide rates in the United States changed over time?
2. Are there statistically significant structural breaks in national suicide trends?
3. Which demographic and socio-economic variables are associated with suicide mortality?

4. Do state-level suicide rates exhibit statistically significant spatial clustering and regional grouping patterns?

The primary contribution of this study is the development of a unified spatio-temporal statistical framework for suicide-mortality analysis. Rather than treating descriptive, econometric, temporal, spatial, and exploratory predictive analyses as isolated tasks, the proposed framework integrates them into a coherent public-health modeling pipeline. In addition, the study incorporates clustering-based regional classification and interpretable machine learning methods to provide supplementary insight into geographic and socio-economic heterogeneity associated with suicide mortality.

Although exploratory machine learning methods are included for interpretable purposes, the primary emphasis of the present study remains on statistical, econometric, temporal, and spatial inference. More advanced predictive machine learning analyses involving lagged features, rolling temporalficial intelligence methods are reserved for a subsequent companion study.

2 Literature Review

2.1 Suicide as a Public-Health Problem

Suicide remains a major global public-health concern and is influenced by complex demographic, psychological, socio-economic, behavioral, and environmental factors [1–3]. In the United States, suicide mortality has increased substantially during recent decades, particularly among middle-aged populations and rural communities [4, 5]. Previous studies have also shown substantial heterogeneity across gender, age groups, and geographic regions [6].

2.2 Socio-Economic Determinants of Suicide

Socio-economic conditions have frequently been associated with suicide mortality. Unemployment, economic instability, income inequality, and financial stress have consistently been identified as important contributors to suicide risk across multiple countries and populations [7–9]. Earlier statistical analyses, including conventional regression and correlation-based approaches, have also demonstrated meaningful associations between socio-economic indicators and suicide mortality.

2.3 Time-Series and Structural Break Analysis

Time-series methods are widely used in public-health surveillance and suicide research because of their ability to capture long-term temporal dynamics [12]. However, suicide mortality often evolves

through changing temporal regimes influenced by economic crises, demographic shifts, healthcare changes, and social conditions.

Structural-break analysis provides an important framework for identifying statistically significant changes in long-term trend behavior [13]. Ignoring structural changes may lead to misleading inferences and poor forecasting performance. ARIMA models are also commonly used for epidemiological forecasting because of their flexibility in modeling temporal dependence and nonstationary time-series behavior [12].

2.4 Spatial Analysis and Geographic Dependence

Spatial analysis plays an important role in suicide research because suicide mortality is often geographically clustered rather than randomly distributed across regions. Geographic disparities may reflect differences in rurality, healthcare access, demographic composition, firearm availability, and socio-economic conditions.

Moran's I is one of the most widely used measures of spatial autocorrelation in epidemiology and public-health research [10, 11]. Positive Moran's I values indicate that neighboring geographic regions tend to exhibit similar outcome behavior, thereby providing statistical evidence of spatial clustering.

2.5 Machine Learning and Interpretability

Machine learning methods are increasingly applied in public health research because of their ability to capture nonlinear relationships and complex predictor interactions [14]. Random forest methods have become particularly popular for exploratory health-risk modeling and variable-importance assessment.

Because many machine learning algorithms lack transparency, explainable artificial intelligence approaches such as SHAP (SHapley Additive exPlanations) have been developed to improve interpretability and quantify predictor contributions [15, 16]. These approaches can complement traditional statistical methods by providing additional insight into nonlinear predictor-response relationships while maintaining interpretability in public-health applications.

Overall, existing literature demonstrates the importance of integrating temporal analysis, socio-economic modeling, spatial inference, and interpretable predictive approaches for understanding complex public-health outcomes such as suicide mortality. The present study extends this literature by developing an integrated spatio-temporal statistical framework combining forecasting, econometric modeling, spatial statistics, clustering analysis, and exploratory interpretable machine learning methods.

3 Data

3.1 Data Sources

To comprehensively investigate the temporal, demographic, socio-economic, and spatial determinants of suicide mortality in the United States, this study integrated multiple publicly available datasets spanning the period from 1980 through 2023. The primary national-level statistical analyses focused on annual observations from 1980 through 2018, while later CDC surveillance data from 2019 through 2023 were used for comparison with forecasting results. State-level spatial analyses and clustering procedures primarily used annual state suicide-rate data from 2014 through 2021 because of geographic data availability constraints.

The primary suicide mortality data were obtained from the Centers for Disease Control and Prevention (CDC) [17]. These datasets included annual national suicide rates, age-specific suicide rates, gender-specific suicide rates, and state-level suicide mortality estimates. The analyses focused primarily on the "All persons" and "All ages" national suicide-rate series to maintain a consistent longitudinal framework for temporal modeling.

Socio-economic indicators were obtained from the World Bank Open Data platform [18] and the Federal Reserve Economic Data (FRED) database [19]. The socio-economic variables included GDP growth, inflation, unemployment, labor-force participation, old-age dependency ratio, consumer sentiment, urban population percentage, life expectancy at birth, and income inequality measures. These variables were incorporated to investigate the potential influence of macroeconomic conditions, demographic aging, and long-term social-development patterns on suicide mortality.

All datasets were merged using the calendar year as the common temporal index. Minor missing values within socio-economic series were addressed using linear interpolation together with forward-fill and backward-fill procedures to maintain temporal continuity for regression and time-series analyses.

3.2 Data Preprocessing

Several preprocessing procedures were performed to ensure consistency and compatibility across datasets obtained from the CDC, World Bank, and FRED databases. Variable names, formatting conventions, and temporal indices were standardized across all sources. Year variables and suicide-rate measures were converted into numeric format to facilitate statistical, econometric, and time-series analyses.

Minor missing observations within socio-economic variables were addressed using linear interpolation, together with forward-fill and backward-fill procedures where appropriate. Duplicate

yearly observations within the national suicide series were removed to ensure a unique annual time series for temporal modeling and forecasting analyses.

Datasets from multiple sources were merged using the calendar year as the common temporal key. State abbreviations were additionally converted into full state names to ensure compatibility with geographic shapefiles used in the spatial analyses.

Prior to regression modeling, multicollinearity among socio-economic variables was assessed using variance inflation factors (VIFs). Because several variables exhibited substantial multicollinearity, reduced socio-economic specifications were adopted to improve coefficient stability and interpretability in the regression and ARDL analyses.

For clustering and exploratory machine learning analyses, selected predictor variables were standardized using z-score normalization to improve comparability across variables measured on different numerical scales. In the machine learning analyses, observations were partitioned using a time-ordered training and testing framework to preserve temporal ordering within the annual suicide-rate series.

3.3 Variables

Table 1: Variables used in the analysis.

Variable	**Description**	**Role**
Suicide Rate	Deaths per 100,000 population	Response variable
Year	Calendar year	Time variable
Age Group	Age-specific suicide-rate category	Demographic variable
Gender	Male/female suicide-rate category	Demographic variable
State	U.S. state	Spatial variable
GDP Growth	Annual economic growth measure	Predictor
Inflation	Consumer price inflation rate	Predictor
Unemployment	National unemployment rate	Predictor
Labor-force Participation	Labor-force participation rate	Predictor
Old-age Dependency Ratio	Elderly dependency measure	Predictor
Consumer Sentiment	Consumer sentiment index	Predictor
Urban Population	Urban population percentage	Predictor
Life Expectancy	Life expectancy at birth	Predictor
GINI	Income inequality measure	Predictor

4 Methodology

4.1 Exploratory Trend and Demographic Analysis

The national suicide-rate series was examined using time-series plots and annual change plots. Age specific and gender specific suicide rate trajectories were also visualized to assess demographic heterogeneity across the study period [20].

4.2 Structural Break and Segmented Regression Analysis

To identify major changes in the national suicide rate trend, a grid-search structural-break procedure was applied [21]. For each candidate break year b, separate pre-break and post-break regression models were fitted, and the optimal break year was selected by minimizing

$$RSS(b) = RSS_{\text{before}}(b) + RSS_{\text{after}}(b). \tag{1}$$

A Chow-type statistic was then used to compare the single-trend model with the two-segment model:

$$F = \frac{(RSS_{\text{full}} - (RSS_{\text{before}} + RSS_{\text{after}}))/k}{(RSS_{\text{before}} + RSS_{\text{after}})/(n - 2k)}. \tag{2}$$

After identifying the main break year, segmented regression models were fitted before and after the break point:

$$Y_t = \beta_0 + \beta_1 t + \varepsilon_t, \tag{3}$$

where Y_t denotes the national suicide rate in year t. A penalized multiple-break point procedure was also used to identify additional changes in long-term trend behavior.

4.3 ARIMA Forecasting

Stationarity was evaluated using the Augmented Dickey–Fuller test [22]. While the original series was nonstationary, differencing was applied before fitting an autoregressive integrated moving average model [23]. The fitted ARIMA model was used to generate short-term forecasts and corresponding confidence intervals.

4.4 Socio-Economic Regression Analysis

Pairwise correlations and variance inflation factors (VIFs) were used to examine relationships and multicollinearity among socio-economic predictors. Because several variables exhibited substantial multicollinearity, reduced regression specifications were used for the main inferential analyses.

Multiple linear regression was used to examine associations between suicide mortality and socio-economic predictors:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \cdots + \beta_p X_{pt} + \varepsilon_t. \tag{4}$$

A log-linear model was also fitted to evaluate proportional relationships:

$$\log(Y_t) = \beta_0 + \beta_1 \log(X_{1t}) + \cdots + \beta_p \log(X_{pt}) + \varepsilon_t. \tag{5}$$

4.5 ARDL Modeling

An autoregressive distributed lag model was used to examine dynamic relationships between suicide rates and socio-economic predictors while accounting for temporal dependence [24]:

$$Y_t = \alpha + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=0}^{q} \boldsymbol{\beta}_j^\top X_{t-j} + \varepsilon_t. \tag{6}$$

Model orders were selected using the Akaike Information Criterion.

4.6 Spatial Autocorrelation Analysis

Spatial dependence in state-level suicide rates was evaluated using Moran's I statistic: [10, 11]. Queen-contiguity spatial weights were used to define neighboring states. Moran's I was computed as

$$I = \frac{n}{W} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}. \tag{7}$$

Positive values indicate that neighboring states tend to exhibit similar suicide-rate patterns.

4.7 State Clustering Analysis

Unsupervised clustering methods were applied to state-level suicide rate trajectories to identify regional risk patterns. A state-by-year matrix was standardized using z-score normalization. K-means clustering was then used to group states with similar suicide-rate trajectories, with the number of clusters evaluated using the elbow method and silhouette scores. PCA was used for visualization, and hierarchical clustering with Ward's linkage was used as a supplementary clustering assessment.

4.8 Exploratory Interpretable Machine Learning

A random forest model was used as an exploratory tool to examine nonlinear socio-economic relationships associated with suicide mortality. The analysis used a time-ordered training and testing split to preserve temporal structure. Feature importance and SHAP values were used to interpret the relative contribution and directional influence of socio-economic predictors [15]. The machine learning component was exploratory and supplementary to the main statistical, econometric, and spatial analyses.

5 Results

5.1 Overall Temporal Trends in U.S. Suicide Rates

Figure 1 shows the national suicide-rate trajectory in the United States from 1980 to 2018. Suicide rates fluctuated moderately during the 1980s, declined throughout much of the 1990s, and reached a minimum near 2000. After 2000, suicide mortality increased steadily, with the highest rates observed near the end of the study period.

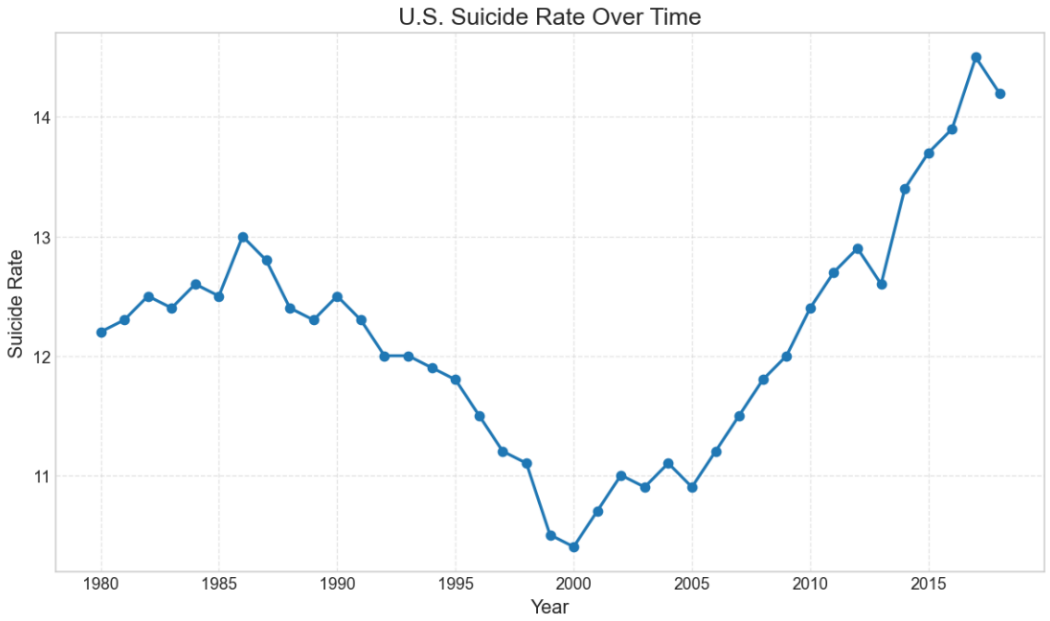


Figure 1: Overall U.S. suicide rate over time.

5.2 Annual Changes in Suicide Rates

Figure 2 presents year-to-year changes in national suicide rates. Negative annual changes were more common during the 1990s, corresponding to the prolonged decline in suicide mortality. In contrast,

positive annual changes became increasingly dominant after approximately 2000, reflecting the sustained upward trend observed during the 21st century.

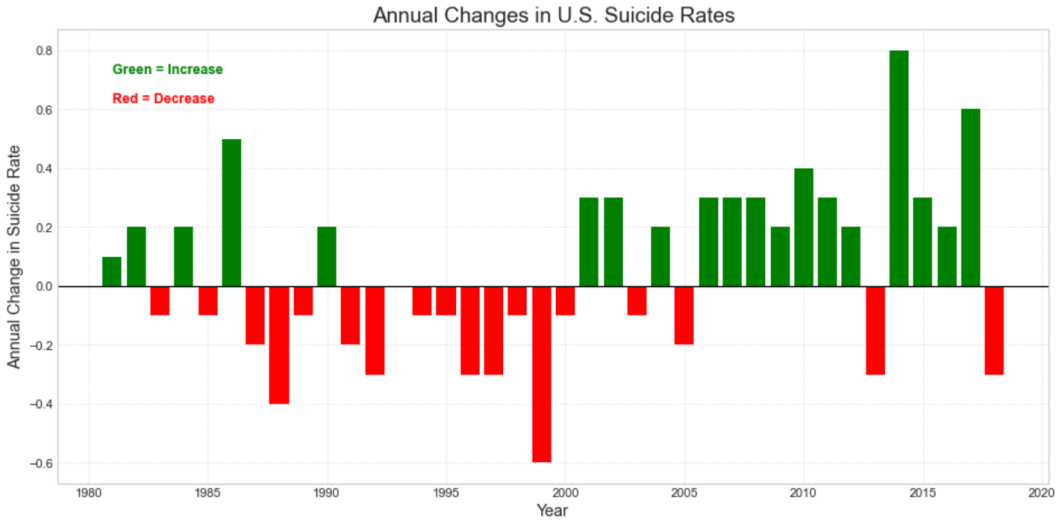


Figure 2: Annual changes in U.S. suicide rates.

5.3 Age-Specific Suicide Rate Trends

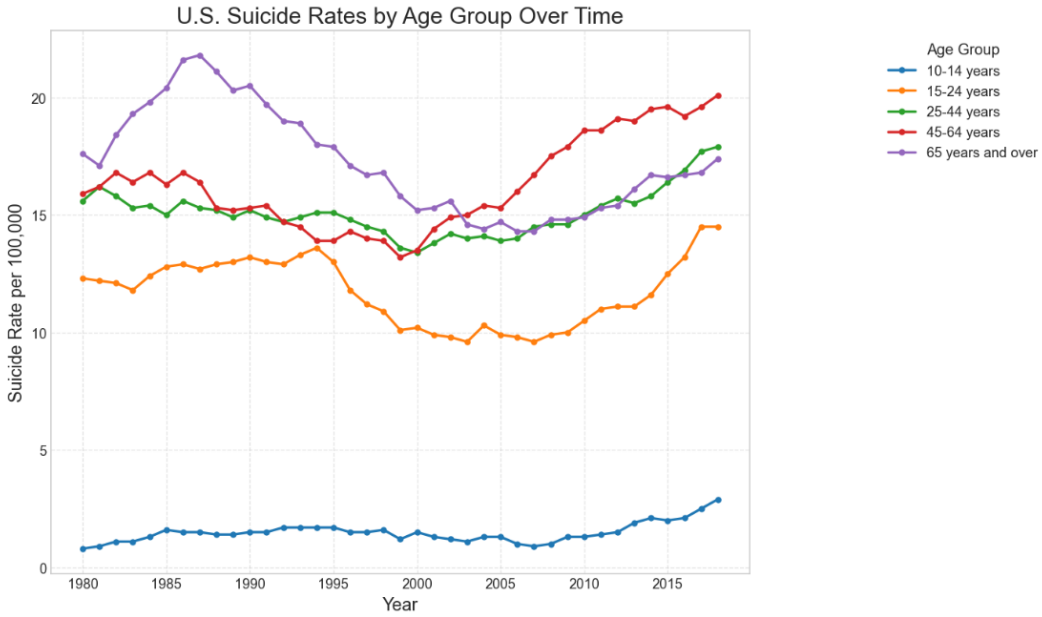


Figure 3: U.S. suicide rates by age group over time.

Figure 3 illustrates substantial heterogeneity in suicide-rate trajectories across age groups. Individuals aged 65 years and older exhibited the highest suicide rates during the early decades of the

study period, with rates exceeding 20 deaths per 100,000 population during portions of the 1980s. However, suicide mortality within the U.S. population declined through the 1990s and early 2000s before increasing moderately again.

In contrast, the 45–64 age group experienced the largest long-term increase after approximately 2000, eventually surpassing older adults in suicide mortality. These findings suggest an important demographic transition in suicide vulnerability over time.

5.4 Gender Differences in Suicide Rates

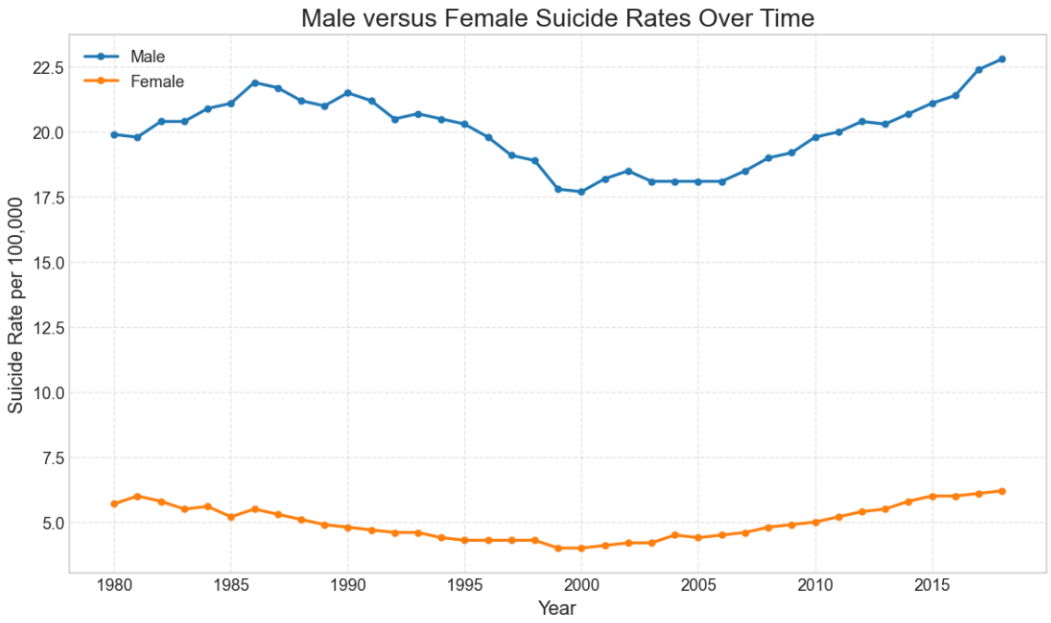


Figure 4: Male versus female suicide rates over time.

Figure 4 presents suicide rate trajectories for males and females. Males consistently exhibited substantially higher suicide rates throughout the study period, with male suicide mortality remaining approximately three to four times greater than female suicide mortality.

Although suicide rates increased for both genders starting in the year 2000, the increase was considerably more pronounced among males, indicating persistent gender disparities in suicide mortality.

5.5 Structural Break and Temporal Regime Analysis

Structural-break analysis identified a major transition in the national suicide-rate trajectory near 1999. The Chow test was highly significant ($F = 165.25$, $p < 0.001$), indicating strong evidence against a single stable long-term trend.

Figure 5 shows segmented regression models before and after the estimated break point. The pre-1999 trend was negative (−0.068), whereas the post-1999 trend became strongly positive (0.217), indicating a substantial reversal toward increasing suicide mortality.

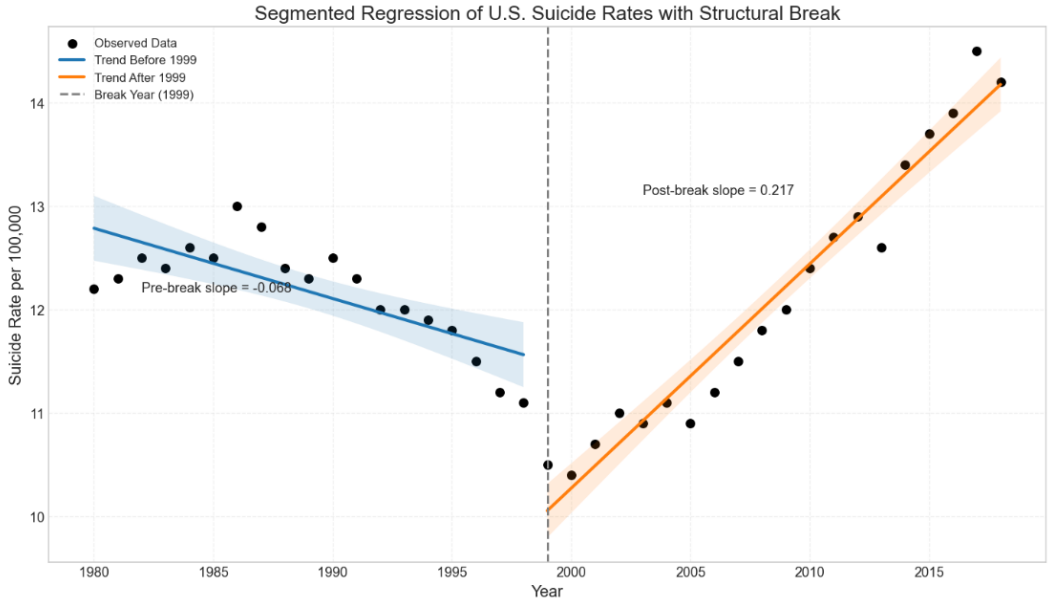


Figure 5: Segmented regression analysis illustrating the major structural break near 1999 with 95% confidence bands.

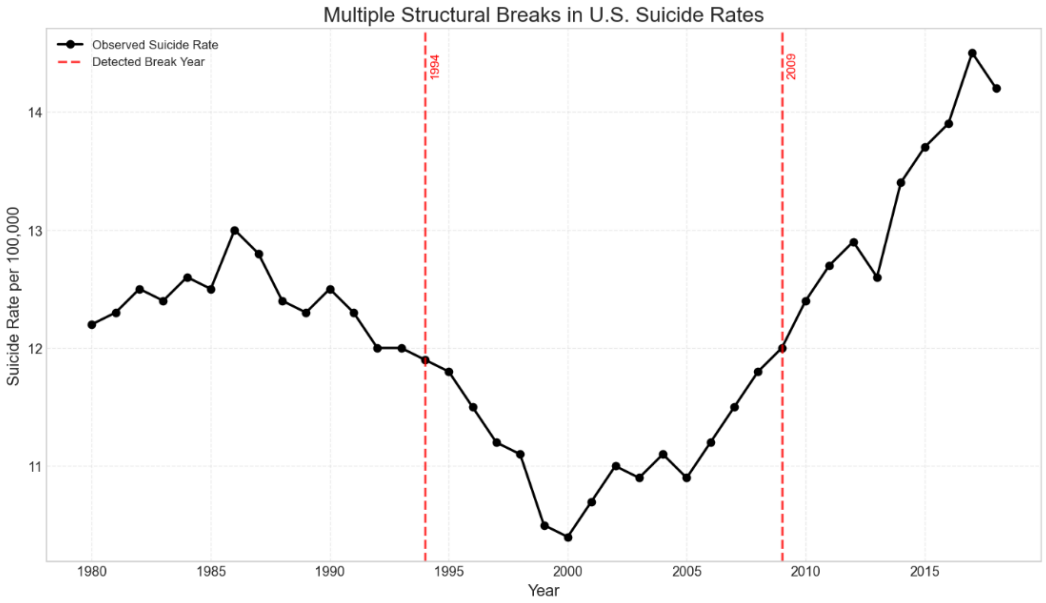


Figure 6: Multiple structural breaks detected in U.S. suicide-rate trends.

To further investigate temporal instability, a multiple-break point analysis was conducted. Figure 6 shows two additional break years that were identified near 1994 and 2009, suggesting

that the national suicide-rate series evolved through multiple temporal phases rather than a single continuous trend.

The break point near 1994 corresponded to the transition into the prolonged decline observed during the late 1990s, whereas the breakpoint near 2009 aligned with the accelerated upward trend observed during the later years of the study period.

5.6 ARIMA Forecasting Results

An ARIMA(1,2,1) model was fitted to the national suicide-rate series after second-order differencing, and the estimation results are presented in Table 2.

Table 2: ARIMA(1,2,1) estimation results.

Parameter	Estimate	Std. Error	z-value	p-value
AR(1)	-0.1920	0.173	-1.108	0.268
MA(1)	-0.7381	0.184	-4.008	< 0.001
σ^2	0.0855	0.026	3.315	0.001

Diagnostic tests did not indicate substantial residual autocorrelation (Ljung–Box $p = 0.78$) or major departures from normality (Jarque–Bera $p = 0.87$).

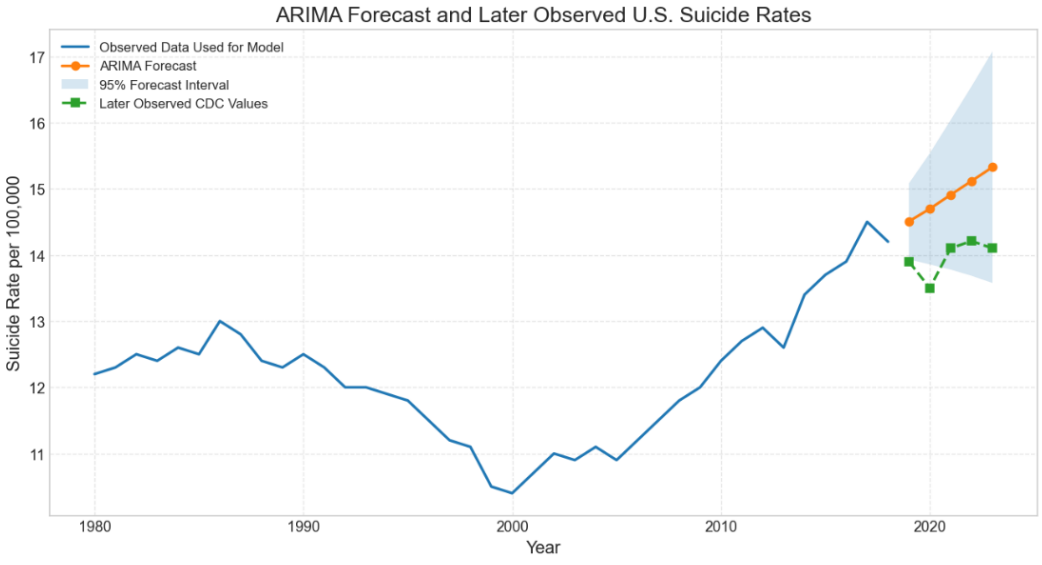


Figure 7: ARIMA forecasts of U.S. suicide rates together with later observed CDC values and 95% forecast intervals.

Figure 7 compares short-term ARIMA forecasts with later observed CDC values from 2019 through 2023. Although the observed CDC values remained somewhat lower than the projected

point forecasts, they stayed historically elevated and generally within the 95% forecast interval range.

Overall, the ARIMA analysis supports the conclusion that the post-2000 increase reflects a sustained elevated suicide-mortality regime rather than a short-term fluctuation.

5.7 Socio-Economic Correlation and Multicollinearity Analysis

Figure 8 presents the correlation matrix for the reduced socio-economic variables used in the regression and ARDL analyses. Suicide rates showed positive associations with unemployment, inflation, and the old-age dependency ratio.

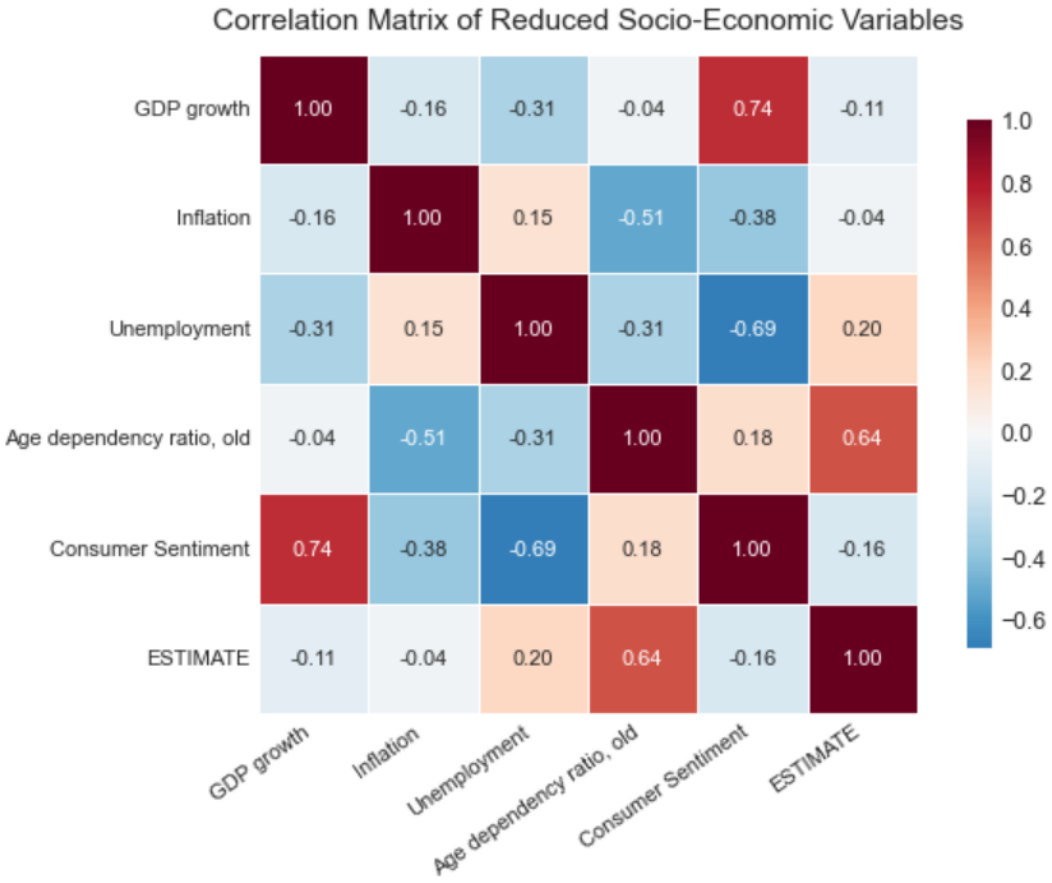


Figure 8: Correlation matrix for the reduced socio-economic variables.

Variance inflation factors presented in Table 3 indicate substantially reduced multicollinearity relative to the full socio-economic specification.

Table 3: Variance inflation factors for the reduced regression model.

Variable	VIF
GDP growth	2.82
Inflation	1.83
Unemployment	2.78
Old-age dependency ratio	1.61
Consumer sentiment	5.77

5.8 Reduced Multiple Regression Analysis

A reduced multiple linear regression model was estimated using GDP growth, inflation, unemployment, old-age dependency ratio, and consumer sentiment as predictors of national suicide rates. Table 4 summarizes the regression results. The model explained approximately 73.6% of the variation in suicide rates ($R^2 = 0.736$, adjusted $R^2 = 0.696$).

Inflation, unemployment, and the old-age dependency ratio were statistically significant positive predictors of suicide mortality.

Table 4: Reduced multiple regression results.

Variable	Estimate	Std. Error	t-value	p-value
Inflation	0.2059	0.049	4.185	< 0.001
Unemployment	0.3950	0.090	4.377	< 0.001
Old-age dependency ratio	0.6934	0.075	9.253	< 0.001
Consumer sentiment	0.0235	0.018	1.294	0.205
GDP growth	0.0063	0.081	0.077	0.939

Overall, the results suggest that labor-market conditions, inflation, and demographic aging are associated with national suicide mortality.

5.9 Log-Linear Regression Analysis

A log-linear regression model was estimated to examine proportional relationships between socioeconomic variables and suicide mortality. The log transformation additionally helped stabilize variability and evaluate relative percentage-type effects among predictors. Table 5 summarizes the estimation results. The model explained approximately 66.2% of the variation in log-transformed suicide rates ($R^2 = 0.662$, adjusted $R^2 = 0.603$).

Table 5: Log-linear regression results.

Variable	Estimate	Std. Error	t-value	p-value
ln(Unemployment)	0.2321	0.074	3.124	0.004
ln(Old-age dependency ratio)	1.0572	0.160	6.601	< 0.001
ln(Inflation)	0.0295	0.018	1.675	0.105
ln(Consumer sentiment)	0.1872	0.185	1.012	0.320
ln(GDP growth)	-0.0094	0.025	-0.371	0.714

The old-age dependency ratio remained strongly significant across both linear and log-linear specifications, suggesting that demographic aging is a stable socio-economic correlate of suicide mortality in the national time series.

5.10 Autoregressive Distributed Lag (ARDL) Analysis

Because the reduced regression models exhibited residual autocorrelation, an autoregressive distributed lag model was estimated to account for temporal dependence and lagged socio-economic effects. Model-order selection based on the Akaike Information Criterion identified an ARDL(1,3) specification as the optimal model. Table 6 presents the estimation results.

Table 6: ARDL model estimation results.

Variable	Estimate	z-value	p-value
Lagged suicide rate	0.4807	4.783	< 0.001
Old-age dependency ratio (L0)	2.6241	2.757	0.010
Old-age dependency ratio (L1)	-6.5660	-2.492	0.018
Old-age dependency ratio (L2)	8.0801	2.863	0.007
Old-age dependency ratio (L3)	-4.1792	-3.449	0.002

The ARDL analysis revealed substantial temporal persistence in suicide mortality, as indicated by the significant lagged suicide-rate term. After accounting for temporal dependence, the old-age dependency ratio remained the dominant socio-economic predictor and exhibited statistically significant effects across multiple lag periods.

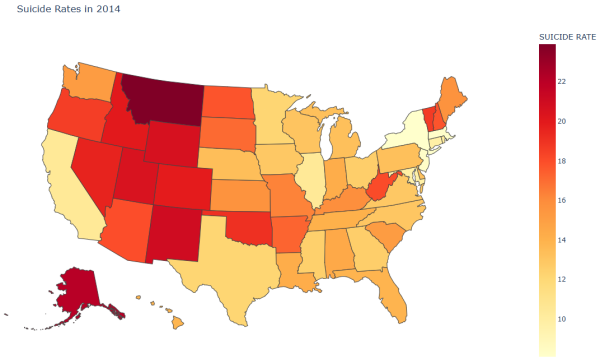
The final ARDL model produced information-criterion values of AIC $= -2.363$, BIC $= 9.100$, and HQIC $= 1.715$, supporting the adequacy of the selected dynamic specification.

Overall, the results suggest that demographic aging and temporal persistence are important components of long-term suicide-rate dynamics in the United States.

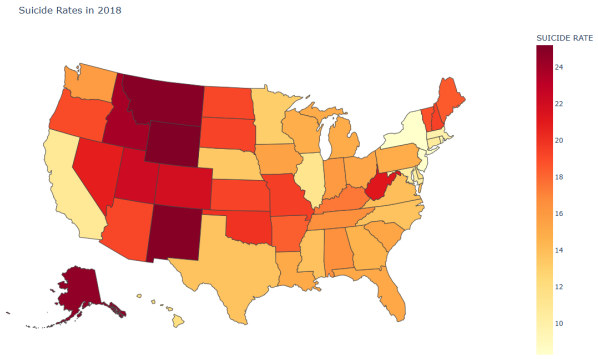
5.11 State-Level Spatial Distribution of Suicide Rates

Figure 9 presents representative choropleth maps of state-level suicide rates from 2014 through 2021. The maps reveal substantial geographic heterogeneity and persistent regional clustering, with western and mountain-region states consistently exhibiting elevated suicide rates relative to many northeastern states.

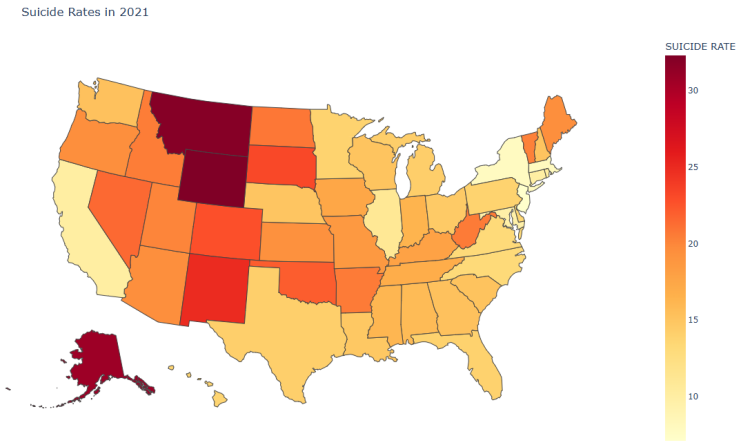
States such as Wyoming, Montana, and Alaska consistently exhibited among the highest suicide rates throughout the study period, whereas many northeastern states maintained comparatively lower suicide mortality.



(a) 2014



(b) 2018



(c) 2021

Figure 9: Representative state-level suicide-rate maps illustrating persistent geographic clustering across western and mountain-region states.

5.12 Spatial Autocorrelation Analysis Using Moran's I

Moran's I statistics were computed for each year of the state-level dataset to evaluate geographic clustering in suicide mortality. Table 7 summarizes the spatial autocorrelation results. The Moran's I statistics remained consistently positive and statistically significant across all years, indicating

persistent spatial autocorrelation in state-level suicide rates.

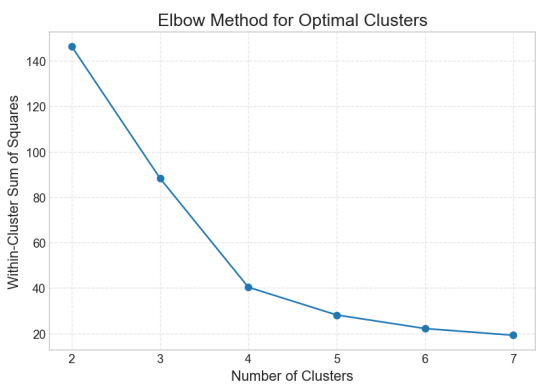
All Moran's I statistics were statistically significant at both significance levels $\alpha = 0.05$ and $\alpha = 0.01$. These findings provide strong evidence that neighboring states tend to exhibit similar suicide-rate behavior rather than random geographic variation.

Table 7: Yearly Moran's I spatial autocorrelation results.

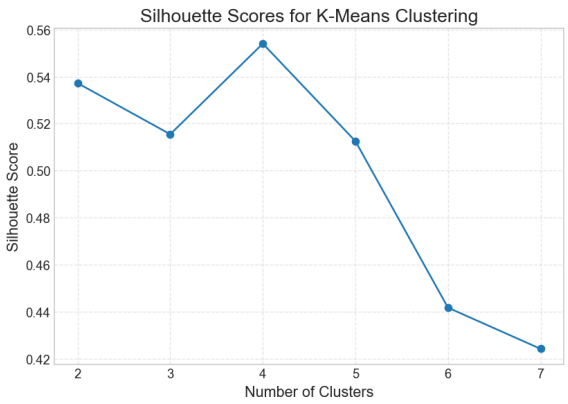
Year	Moran's I	p-value
2014	0.4434	0.001
2015	0.4547	0.001
2016	0.4376	0.001
2017	0.4584	0.001
2018	0.4583	0.001
2019	0.4815	0.001
2020	0.4691	0.001
2021	0.4119	0.001

5.13 State Clustering Analysis

Unsupervised clustering methods were applied to state-level suicide-rate trajectories from 2014 through 2021. Figures 10(a) and 10(b) present the elbow method and silhouette analysis used to determine the optimal number of clusters. Although the elbow method suggested a transition near three to four clusters, the silhouette coefficient achieved its maximum value at four clusters ($s \approx 0.554$); therefore, a four-cluster solution was selected.



(a) Elbow method



(b) Silhouette analysis

Figure 10: Methods used to determine the optimal number of K-means clusters.

Cluster 3 represented a very high-risk group consisting primarily of Wyoming, Alaska, Montana, and New Mexico, whereas Cluster 0 contained lower-risk northeastern states. Table 8 summarizes

the average suicide-rate characteristics of each cluster.

Table 8: Summary of state-level suicide-rate clusters (2014–2021).

Cluster	Risk Level	Mean Rate	Geographic Pattern
Cluster 0	Low Risk	9–10	Northeastern states
Cluster 1	High Risk	18–20	Western and mountain states
Cluster 2	Moderate Risk	13–15	Southern and midwestern states
Cluster 3	Very High Risk	22–30	Wyoming, Alaska, Montana, New Mexico

Figure 11 presents the PCA visualization of the K-means clustering structure.

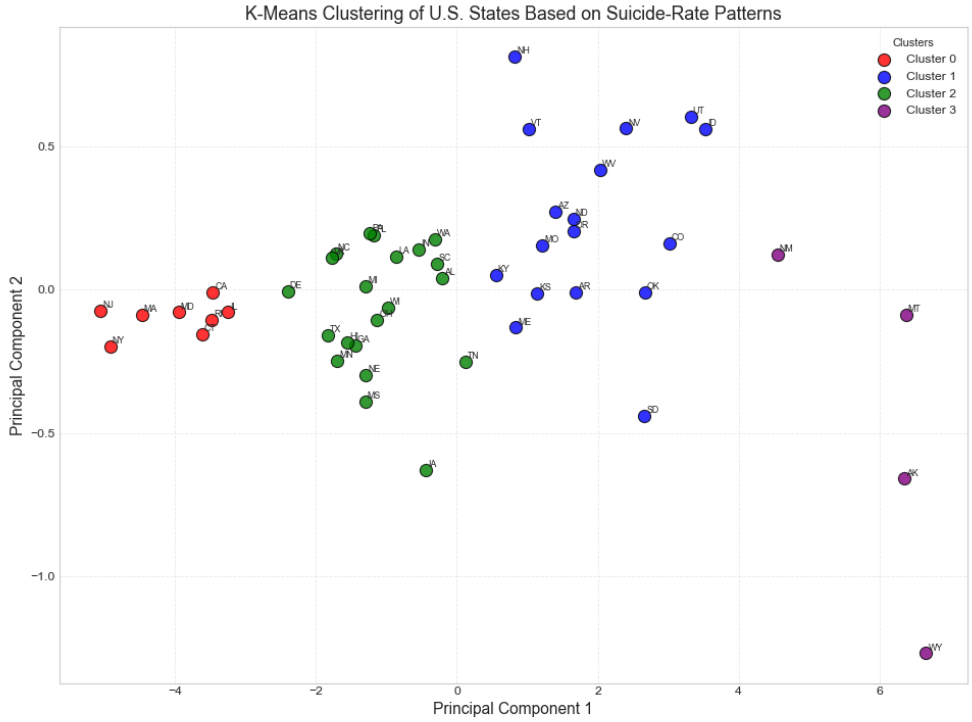


Figure 11: PCA visualization of K-means state clusters.

The clustering analysis further supports the presence of distinct geographic suicide-risk regimes across the United States.

5.14 Exploratory Interpretable Machine Learning Results

An exploratory random forest analysis was conducted to examine nonlinear socio-economic relationships associated with suicide mortality. The model used the reduced predictor set consisting of GDP growth, inflation, unemployment, old-age dependency ratio, and consumer sentiment.

Because the dataset consisted of annual observations from 1980–2018, the machine learning analysis was intended primarily for interpretation rather than high-accuracy prediction. Figure 12 presents the random forest feature-importance rankings, which identified unemployment as the dominant predictor, followed by inflation and the old-age dependency ratio.

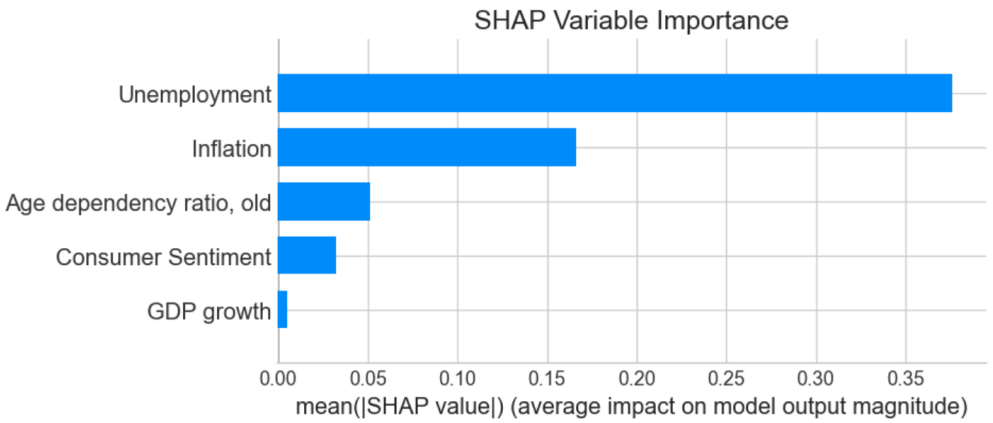


Figure 12: Random forest feature-importance rankings for reduced socio-economic predictors.

Figure 13 presents the SHAP summary plot for the exploratory machine learning analysis. The SHAP interpretation further supported the feature-importance results by showing that unemployment had the largest contribution to model predictions. Higher unemployment values generally produced positive SHAP contributions, indicating upward effects on predicted suicide mortality.

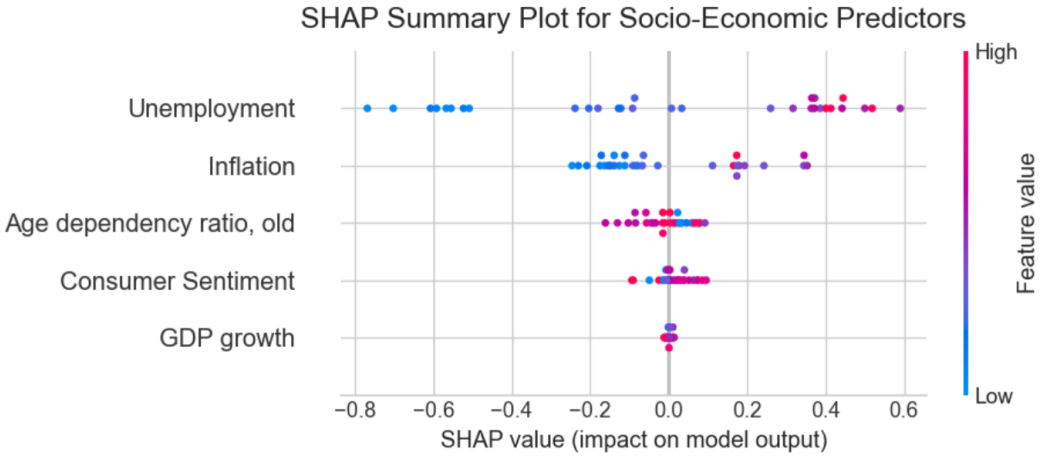


Figure 13: SHAP summary plot showing the contribution and directional influence of socio-economic predictors on suicide-rate predictions.

Overall, the exploratory machine learning results complemented the regression and ARDL analyses by identifying unemployment and inflation as important nonlinear predictors. More

advanced predictive modeling using lagged predictors, rolling-window features, hyperparameter tuning, and additional algorithms is reserved for the second paper.

6 Discussion

This study developed an integrated spatio-temporal statistical framework for investigating suicide mortality patterns in the United States by combining temporal analysis, econometric modeling, spatial statistics, clustering methods, and exploratory interpretable machine learning techniques. Collectively, the findings demonstrate that suicide mortality is influenced by interacting temporal, demographic, socio-economic, and geographic factors.

The temporal analyses revealed substantial changes in national suicide dynamics over time. Structural-break and segmented-regression analyses identified a major transition near the late 1990s, after which suicide mortality shifted from a declining pattern toward a sustained upward trajectory. The ARIMA forecasting analysis further supported the persistence of elevated suicide mortality during the 21st century.

Demographic analyses demonstrated considerable heterogeneity across age groups and gender. In particular, suicide mortality increased substantially among middle-aged populations after approximately 2000, while males consistently exhibited substantially higher suicide rates than females throughout the study period.

The spatial analyses revealed strong evidence of geographic clustering in suicide mortality. Western and mountain-region states consistently exhibited elevated suicide rates, while many northeastern states maintained comparatively lower suicide mortality. Moran's I statistics and clustering analyses further confirmed persistent regional structure and positive spatial autocorrelation in state-level suicide patterns.

The socio-economic analyses suggested that unemployment, inflation, and demographic aging are associated with national suicide mortality. The ARDL framework additionally demonstrated substantial temporal persistence and lagged demographic effects, indicating that suicide mortality evolves dynamically over time rather than through purely contemporaneous relationships.

The exploratory interpretable machine learning analyses complemented the statistical findings by identifying unemployment as the dominant nonlinear predictor associated with suicide mortality. SHAP interpretation additionally demonstrated the usefulness of explainable artificial intelligence methods for investigating complex socio-economic relationships in public-health data.

Overall, the findings highlight the importance of integrating temporal, demographic, socio-economic, spatial, and interpretable predictive perspectives when modeling suicide mortality. The proposed framework provides a flexible foundation for future public-health monitoring, spatial surveillance, and data-driven suicide-prevention research.

7 Conclusion

This study developed an integrated spatio-temporal analytical framework for investigating suicide mortality patterns in the United States. By combining temporal trend analysis, structural-break detection, econometric modeling, spatial statistics, clustering analysis, and exploratory interpretable machine learning methods, the study provides a comprehensive framework for examining complex suicide mortality dynamics.

The findings revealed substantial temporal instability in national suicide trends together with persistent demographic, socio-economic, and geographic heterogeneity. In particular, the analyses identified sustained increases in suicide mortality after the late 1990s, elevated suicide risk among middle-aged populations and males, and strong spatial clustering across western and mountain-region states. Socio-economic analyses consistently highlighted unemployment and demographic aging as important factors associated with suicide mortality.

Overall, the results demonstrate that suicide mortality in the United States is shaped by interacting temporal, demographic, socio-economic, and geographic processes. The proposed framework provides a foundation for future research involving advanced forecasting systems, spatial surveillance, and interpretable machine learning approaches for public-health monitoring and suicide-risk assessment.

8 Limitations

Several limitations of the present study should be acknowledged. First, the analyses were conducted primarily using aggregate national-level and state-level data and therefore may not capture important individual-level psychological, behavioral, and clinical risk factors associated with suicide mortality.

Second, long-term state-level socio-economic data were limited for several potentially important predictors because of missing observations, temporal inconsistencies, and differences in reporting practices across public databases.

Third, although the econometric and time-series methods provided important associative and forecasting insights, they do not establish causal relationships between socio-economic variables and suicide mortality.

Fourth, the spatial analyses were conducted at the state level and therefore may overlook important county-level or local geographic heterogeneity.

Fifth, the exploratory machine learning analyses were based on a relatively limited annual dataset and were intended primarily for interpretability and exploratory nonlinear assessment rather than large-scale predictive optimization.

Finally, suicide mortality data may be affected by reporting inconsistencies, underreporting, and changes in death-classification practices across states and time periods.

9 Future Work

Several important directions for future research emerge from the present study.

First, future analyses could incorporate county-level or sub-state geographic data, together with local spatial methods such as LISA and hotspot-detection procedures, to provide a more detailed assessment of regional suicide-risk patterns.

Second, Bayesian spatio-temporal frameworks could be developed to jointly model temporal evolution, spatial dependence, and geographic uncertainty within a unified hierarchical framework.

Third, future studies could incorporate causal-inference methodologies to evaluate the effects of economic conditions, healthcare access, public-health interventions, firearm legislation, and social policies on suicide mortality dynamics.

Fourth, the exploratory machine learning framework can be extended through lagged predictors, rolling-window temporal features, hyperparameter optimization, and advanced algorithms such as XGBoost, LSTM neural networks, and hybrid statistical–machine learning models for suicide-risk forecasting and early-warning systems.

Finally, comparative international analyses between the United States and developing countries such as Sri Lanka may provide additional insight into how cultural, economic, demographic, and healthcare factors influence suicide mortality under different public-health environments.

Acknowledgements

The authors would like to express their sincere gratitude to Professor D. D. S. Kulatunga, Emeritus Professor at the University of Kelaniya, Sri Lanka, for his valuable encouragement and academic guidance.

The authors also gratefully acknowledge the support provided by the Department of Mathematics at the University of Wisconsin–Whitewater and the Department of Mathematics and Statistics at Texas Tech University.

References

[1] World Health Organization, “Suicide worldwide in 2019: Global health estimates,” 2021. [Online]. Available: https://www.who.int/publications/i/item/9789240026643

[2] Centers for Disease Control and Prevention, "Suicide data and statistics," 2023. [Online]. Available: https://www.cdc.gov/suicide/facts/data.html

[3] M. K. Nock, G. Borges, E. J. Bromet, J. Alonso, M. Angermeyer, A. Beautrais, R. Bruffaerts, W. T. Chiu, G. De Girolamo, S. Gluzman *et al.*, "Cross-national prevalence and risk factors for suicidal ideation, plans and attempts," *The British Journal of psychiatry*, vol. 192, no. 2, pp. 98–105, 2008.

[4] H. Hedegaard and M. Warner, "Suicide mortality in the united states, 1999-2019," *NCHS Data Brief*, 2021.

[5] S. C. Curtin and M. P. Heron, "Death rates due to suicide and homicide among persons aged 10–24: United states, 2000–2017," *NCHS Data Brief*, 2019.

[6] A. L. Cammack, M. R. Stevens, R. B. Naumann, J. Wang, W. Kaczkowski, J. Valderrama, D. M. Stone, and R. Lee, "Vital signs: Suicide rates and selected county-level factors — united states, 2022," *Morbidity and Mortality Weekly Report*, vol. 73, no. 37, pp. 810–818, 2024.

[7] S. Stack, "Suicide: a 15-year review of the sociological literature part i: cultural and economic factors," *Suicide and Life-Threatening Behavior*, vol. 30, no. 2, pp. 145–162, 2000.

[8] A. Milner, A. Page, and A. D. LaMontagne, "Long-term unemployment and suicide: a systematic review and meta-analysis," *PloS one*, vol. 8, no. 1, p. e51333, 2013.

[9] S.-S. Chang, D. Stuckler, P. Yip, and D. Gunnell, "Impact of 2008 global economic crisis on suicide: time trend study in 54 countries," *Bmj*, vol. 347, 2013.

[10] L. Anselin, "Local indicators of spatial association - lisa," *Geographical analysis*, vol. 27, no. 2, pp. 93–115, 1995.

[11] P. A. Moran, "Notes on continuous stochastic phenomena," *Biometrika*, vol. 37, no. 1/2, pp. 17–23, 1950.

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

[13] J. Bai and P. Perron, "Computation and analysis of multiple structural change models," *Journal of applied econometrics*, vol. 18, no. 1, pp. 1–22, 2003.

[14] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean, "A guide to deep learning in healthcare," *Nature Medicine*, vol. 25, no. 1, pp. 24–29, 2019. [Online]. Available: https://doi.org/10.1038/s41591-018-0316-z

[15] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, vol. 30, 2017.

[16] D. A. Michael, C. Rajapaksha, and B. H. Wellalage, "A machine learning approach to identifying factors affecting students' mathematics performance in government schools in the colombo education zone, sri lanka," in *IEEE Conference Proceedings*, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/11407767

[17] Centers for Disease Control and Prevention, "Centers for disease control and prevention data portal," https://www.cdc.gov/index.html, 2024.

[18] World Bank, "World bank open data," https://data.worldbank.org/, 2024.

[19] Federal Reserve Bank of St. Louis, "Federal reserve economic data (fred)," https://fred.stlouisfed.org/, 2024.

[20] H. J. Seltman, "Experimental design and analysis: Regression," https://www.stat.cmu.edu/~hseltman/309/Book/chapter4.pdf, 2018.

[21] H. Choi, T. Kim, H. Lee, and M. Kim, "Improved identification of breakpoints in continuous piecewise regression," *arXiv preprint arXiv:2408.13751*, 2024. [Online]. Available: https://arxiv.org/abs/2408.13751

[22] J. H. Lopez, "The power of the adf test," *Economics Letters*, vol. 57, no. 1, pp. 5–10, 1997.

[23] B. E. L. Borja and M. E. M. Ventura, "Chapter 1: Autoregressive integrated moving average (arima)," https://phdinds-aim.github.io/time_series_handbook/01_AutoRegressiveIntegratedMovingAverage/01_AutoRegressiveIntegratedMovingAverage.html, 2023, time Series Analysis Handbook.

[24] M. H. Pesaran, Y. Shin, and R. J. Smith, "Bounds testing approaches to the analysis of level relationships," *Journal of applied econometrics*, vol. 16, no. 3, pp. 289–326, 2001.