Socioeconomic Structure and Marital Stability Across U.S. States:
Evidence from Birth-Origin Divorce Rates and Multivariate Clustering

Abstract: This study analyzes state-level divorce patterns in the United States using 2024 American Community Survey (ACS) 1-year estimates. We construct divorce rates for multiple birth-origin groups—including native-born, migrants from other states, U.S. citizens born abroad, and foreign-born residents—and link these rates with key sociodemographic indicators such as income, poverty, unemployment, education, and migration composition. An interactive Shiny application visualizes geographic variation through choropleth maps, histograms, and summary statistics, revealing substantial heterogeneity across states, with notable hotspots in the Mountain West and the South (e.g., Maine, New Mexico, Nevada, West Virginia)   To identify broader structural patterns, we apply K-means clustering on divorce rates and socioeconomic factors. The elbow method suggests a three-cluster solution that separates states into (1) low-divorce, high socioeconomic advantage; (2) high-divorce, economically disadvantaged; and (3) moderate-divorce, mixed-profile states. A PCA projection confirms that clusters occupy distinct regions of the multivariate space, indicating that socioeconomic conditions strongly shape divorce patterns. Cluster 1 consists of affluent, highly educated states with uniformly low divorce rates; Cluster 2 captures structurally disadvantaged southern and Appalachian states with consistently high divorce levels; and Cluster 3 represents a heterogeneous middle group with moderate outcomes.

Overall, the analysis provides clear evidence that divorce patterns across U.S. states are closely intertwined with economic and demographic conditions. The interactive platform developed in this study offers a flexible tool for visualizing and comparing state-level marital stability across diverse population subgroups.

Keywords: *American Community Survey (ACS); divorce rate; sociodemographic factors; K-means clustering; principal component analysis (PCA); regional disparities; migration and nativity; state-level analysis*

1. Introduction

Marital stability is one of the important indicators for measuring social structure and population well-being. Katherine et al.[1] pointed out that socioeconomic development, female employment rate, gender ratio, and the dominant religion in a region are four main factors affecting the divorce rate in a society. Different regions in the United States have long-standing differences in economic structure, education levels, employment opportunities, and resident diversity. Therefore, analyzing the divorce rate in American society can reveal the potential roles of demographic and socioeconomic factors. This study aims to further explore the spatial differences of these factors through a systematic analysis of state-level divorce rates and related variables.

The data source for this study is the 2024 ACS one-year data. This data is characterized by its broad time coverage and numerous variables, facilitating merging for exploratory data analysis and further clustering to observe regional characteristics. Unlike previous studies that only focused on the overall divorce rate, this study uses data that subdivides the divorce rate by place of birth, including: residents born in the state, residents born in other states, US citizens born abroad, and foreign-born residents. Choosing this categorized data helps us understand the potential relationship between resident diversity and the divorce rate in American society and its different regions.

To facilitate a more intuitive observation of the characteristics of the divorce rate in American society, this study constructed an R Shiny interactive visualization tool, allowing users to dynamically view the geographical distribution and statistical characteristics of the divorce rate, and reveal the similarities between states in socioeconomic characteristics through the K-means clustering method.

Ultimately, the research objectives of this study are:
(1) To observe whether there are systematic differences in divorce rates among different groups based on place of birth.
(2) To reveal whether socioeconomic factors, such as income level, poverty rate, and education level, are significantly correlated with state-level divorce rates.
(3) To observe whether the divorce rate in the United States exhibits a significant spatial distribution pattern through cluster analysis, thereby further demonstrating the relationship between marital status and geographical distribution in the United States.
These questions will be systematically discussed through visualization, statistical analysis, and cluster exploration.

2. Data Construction

The data used in this study comes from the 2024 American Community Survey (ACS) one-year data provided by the U.S. Census Bureau. To ensure the reproducibility of the analysis and the efficiency of the Shiny application, the necessary raw ACS variables were pre-processed and stored in an rds file.

2.1 Data Source and Variable Definition
The main data we used is Table B06008: Place of Birth by Marital Status. This table contains the number of divorced individuals and the total population for different groups based on their place of

birth. The divorce rate is denoted as the number of divorces divided by the total population of the area. We constructed the following five categories: overall divorce rate, divorce rate of residents born in this state, divorce rate of residents born in other states, divorce rate of US citizens born abroad, and divorce rate of foreign-born residents.

In addition, we also merged the following ACS data to supplement socioeconomic indicators: poverty status, unemployment status, educational attainment, median household income, and the percentage of foreign-born population.

Table 1. Merging Variables and Definitions

Variable	Definition
Poverty Rate	Population below poverty level / Total population
Median Household Income	Median of Population Income
Unemployment Rate	Unemployed Population / Labor Force
College Plus Rate(Education Level)	Population with a Bachelor's Degree or Higher / Population Aged 25 and Over
Foreign-Born Share	Foreign-Born Population / Total Population
Median Age	Median of Population Age
Owner-Occupied Housing Rate	Housing owner / Total Housing
Median Gross Rent	Median of Gross Rent
Median Home Value	Median of value of the house on the market.
Share White alone	Population of White / Total Population
Share Hispanic/Latino	Population of Hispanic or Latino / Total Population

2.2 Geographic Data Processing

All state-level data is stored in simple features (sf) format. The advantage of this is that we can use it directly to draw polygon maps, integrate it with Leaflet interactive components, and it supports interactive features such as geographical highlighting, legends, and mouseover labels.

2.3 Clustering Data Preprocessing

To perform K-means clustering, this study processed the data as follows:

(1) Removed NA values, Inf values, and unstable ratios caused by very small denominators.

(2) Standardized the data. We chose z-score standardization:

$$z = \frac{x - \mu}{\sigma}.$$

It is worth noting that users can choose whether or not to perform standardization here to eliminate differences in units of measurement.

(3) Removed variables with a standard deviation of zero to avoid unstable clustering.

(4) Constructed the final numerical matrix used for clustering.

3. Exploratory Data Analysis

3.1 Distribution and Visualization of Divorce Rates by Place of Birth

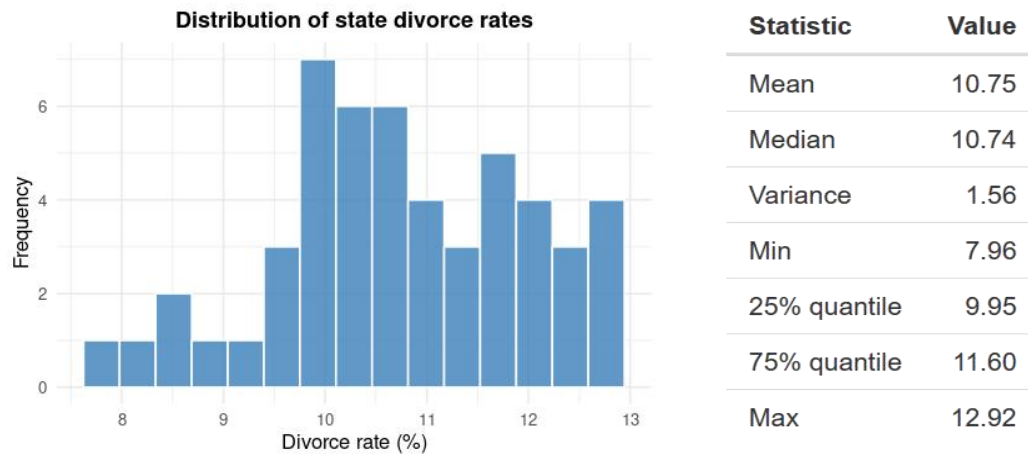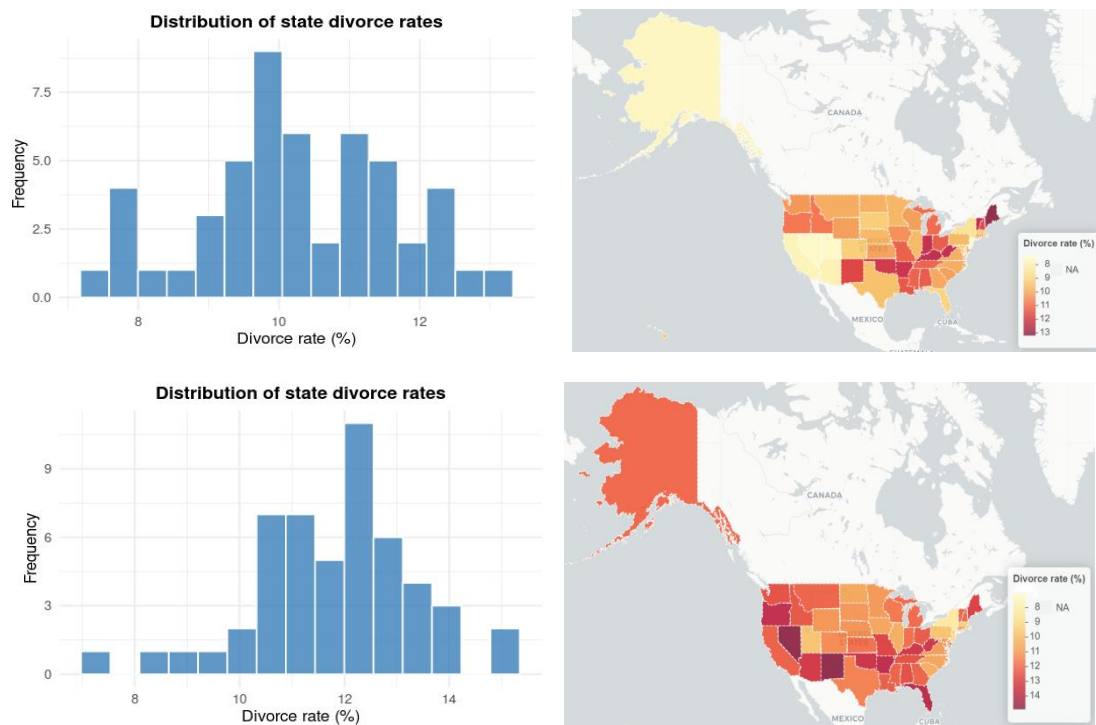| Statistic	Value
Mean	10.75
Median	10.74
Variance	1.56
Min	7.96
25% quantile	9.95
75% quantile	11.60
Max	12.92

Figure 1. Distribution and Statistics of Overall States Divorce Rates

First, we conducted an Exploratory Data Analysis (EDA) of the overall divorce rate in the United States. As shown in the graph and table, the divorce rates in most states fall within a core range of around 10%, with several states clustered between 9.5% and 11%, and a few states at lower (approximately 8%) or higher (close to 13%) rates. Overall, the distribution is relatively concentrated but slightly skewed to the right.
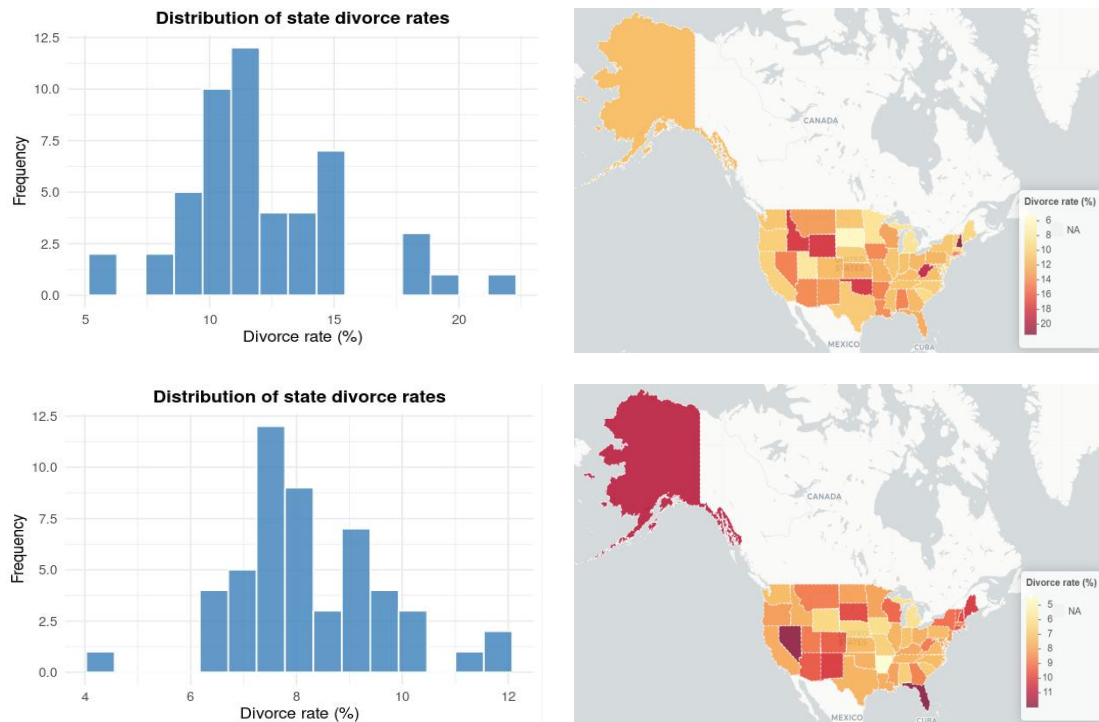
Figure 2. Distribution of States Divorce Rates and Visualization

The order of the histograms and visualizations is as follows: residents born in this state, residents born in other states, US citizens born abroad, and foreign-born population.

For residents born in the same state, the histogram shows that divorce rates in most states are concentrated in the 9%-11% range, with a relatively concentrated distribution and a slight rightward skew. Looking at the map, states with higher divorce rates among residents born in the same state are mainly concentrated in the Northeast (such as Maine and Vermont) and some Southern states (such as Kentucky and West Virginia). Many states in the Midwest and Mountain West have lower divorce rates, indicated by lighter colors. Therefore, we believe that the divorce rate for residents born in the same state is generally the lowest, with regional differences not being extreme, representing a relatively stable and concentrated group.

Residents born in other states show higher and more dispersed divorce rates. The histogram shows that divorce rates are in the 10%-14% range, with a more pronounced rightward skew. The map shows that areas with high divorce rates are clearly concentrated in the South and Southwest (such as New Mexico, Nevada, and Florida), where these states show higher divorce levels among the population who migrated from other states. Many states in the Midwest and Northeast are more stable, with lower divorce rates.

The divorce rate for U.S. citizens born abroad shows significant volatility. The histogram shows that the distribution of state-level divorce rates is extremely wide, ranging from approximately 6% to over 20%, with several states exceeding 15%, representing extremely high values. The map shows that extremely high divorce rates are mainly concentrated in states such as Wyoming, Oklahoma, West Virginia, and New Hampshire, indicated by dark red colors. Therefore, we believe that this category has the largest inter-state variation in divorce rates and the lowest concentration, showing

significantly high values in some states.

The state-level divorce rate for foreign-born residents is generally lower. The histogram shows that it is mainly concentrated in the 7%-9% range, with only a few states exceeding 10%. It shows a more pronounced concentration trend, with the smallest range of fluctuation among the four groups. The map shows that states with higher divorce rates among foreign-born residents include Florida, Nevada, and Alaska, but the overall national range is mostly light yellow, indicating lower divorce rates.

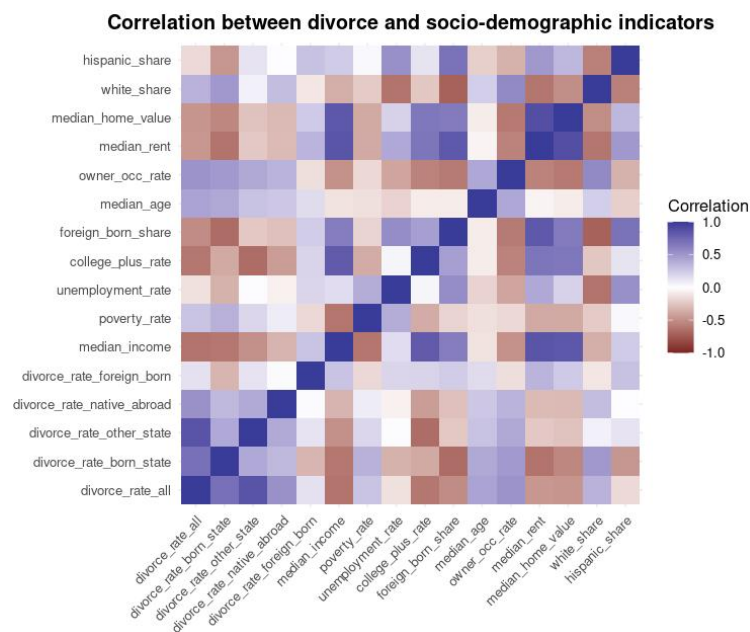3.2 Correlation between Divorce and Merged Socio-Demographic Indicators

Figure 3. The Correlation Heatmap of Divorce and Merged Socio-Demographic Indicators

The correlation heatmap shows a generally positive correlation between divorce rates across different birthplaces, indicating consistency in marital stability across states. Among socioeconomic variables, median income is negatively correlated with divorce rates, while poverty rate and unemployment rate are positively correlated, reflecting that areas with greater economic pressure tend to have higher divorce rates. Educational attainment is also negatively correlated with divorce rates, while housing stability indicators show a weaker correlation with divorce rates.

Overall, states with poorer economic conditions and a higher proportion of socially disadvantaged groups tend to have higher divorce rates; conversely, areas with greater economic prosperity, higher levels of education, and better housing stability have significantly lower divorce rates.

4. K-means Clustering
This paper employs the K-means clustering algorithm to identify U.S. state groups with similar sociodemographic characteristics and divorce rate patterns. K-means clustering groups states by minimizing within-cluster variability. Conceptually, the algorithm:
Assigns each state to the nearest cluster center based on mutivariate similarity

Updates the cluster centers until assignments stabilize

Because clustering depends on the number of groups k, model selection is guided by the elbow method, which evaluates the total within-cluster sum of squares (WSS). Lower WSS indicates tighter, more coherent clusters. Let $X=\{x_1,x_2,...,x_n\}$, $x_i \in \mathcal{R}$ be a dataset of standarlized observations and μ_j is the centroid of cluster j.

$$\text{WSS}=\sum_{j=1}^{k} \sum_{x_i \epsilon C_j} \|x_i-\mu_j\|^2, \ \mu_j=\frac{1}{|C_j|} \sum_{x_i \epsilon C_j} x_i$$

Because the objective function is not convex, the algorithm can converge to local minima. To mitigate this, our implementation uses 25 random initializations (nstart = 25) and selects the best solution.
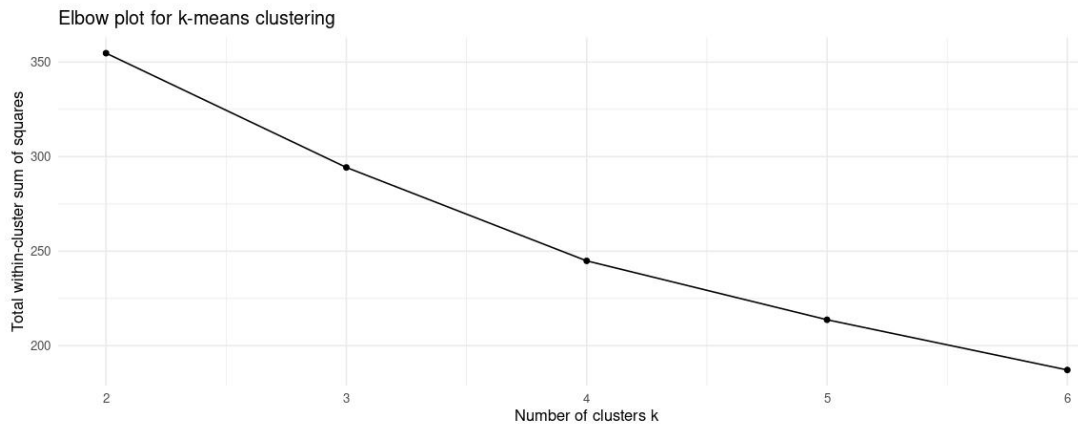
5. Choosing the Number of Clusters

Figure 4. Elbow plot

The WSS shows a steep drop from K=2 to K=3, but the marginal improvement decreases markedly for K>3. This "elbow" indicates that three clusters is an appropriate balance between compactness and interpretability. A larger K would over-partition the data without revealing substantially new structure. Though users can choose K values through Shiny App, in this paper we chose K=3 to show results.

6. K-means scatterplot

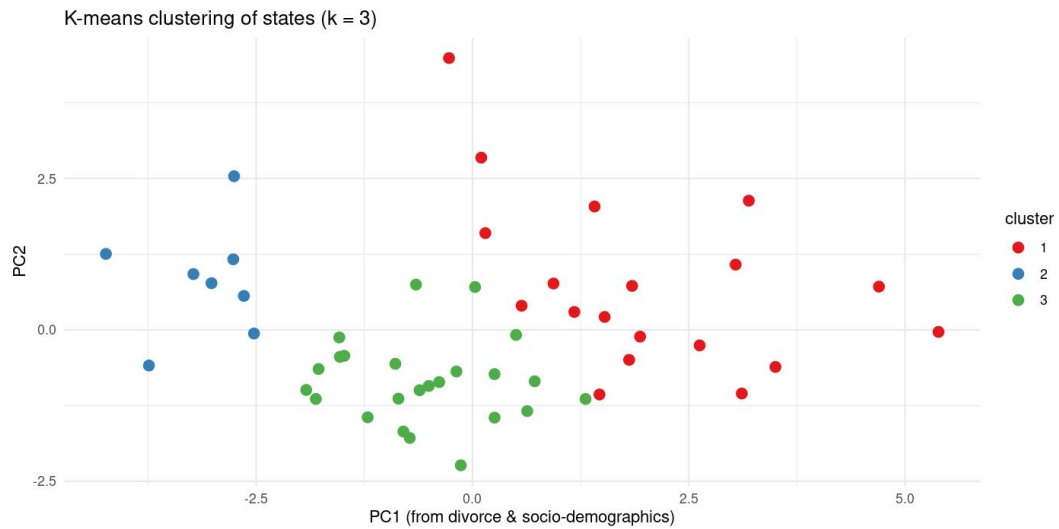**K-means clustering of states (PCA 2D projection)**

Figure 5. K-means scatterplot

Each point in this figure is a state, positioned by its scores on the first two principal components computed from all divorce and sociodemographic variables, and colored by its K-means cluster (k = 3).

The horizontal axis (PC1) clearly separates the three clusters: states in Cluster 2 (blue) lie on the far left and correspond to low divorce and strong socioeconomic conditions, states in Cluster 1 (red) lie on the far right and are characterized by high divorce and economic disadvantage, while Cluster 3 (green) occupies the middle of the axis with moderate divorce and mixed socioeconomic profiles.

The vertical axis (PC2) mainly reflects within-cluster variation rather than forming additional clearly separated groups. Overall, the plot shows that the three clusters identified by K-means correspond to distinct regions of the multivariate space, indicating that the clustering captures meaningful structure in the joint distribution of divorce rates and sociodemographic indicators across states.

7. Geographic Distribution of Clusters
Cluster Profiles: Sociodemographic and Divorce-Rate Characteristics Across U.S. States
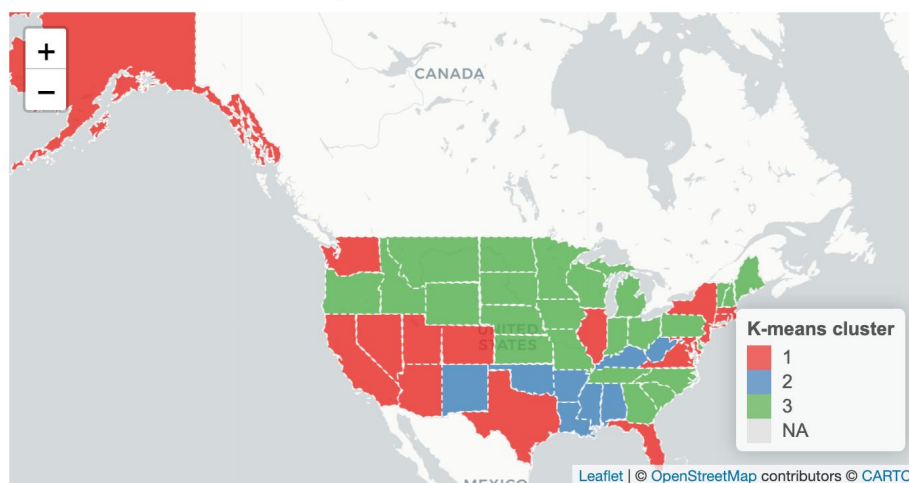

Figure 6. K-means clusters on U.S. map

Cluster 1 — Low Divorce, High Socioeconomic Advantage

Cluster 1 consists primarily of economically prosperous, highly educated states such as California, New York, Massachusetts, New Jersey, Connecticut, Hawaii, Washington, and Colorado. These states exhibit the lowest divorce rates across all birth-origin groups, with divorce_rate_all well below the national mean. Both native-born and foreign-born populations show uniformly low divorce levels, indicating broad marital stability. Socioeconomic indicators place this cluster at the top of the national distribution: median income is high, college attainment is widespread, and both poverty and unemployment rates are comparatively low. Many of these states also attract substantial domestic and international migration, yet divorce rates remain low, suggesting that strong economic conditions and educational advantages act as protective factors against marital dissolution. Overall, Cluster 1 represents affluent, mobile, and highly educated states with persistently low divorce rates.

Cluster 2 — High Divorce, High Economic Stress

Cluster 2 comprises states such as New Mexico, West Virginia, Kentucky, Oklahoma, Arkansas, Alabama, Mississippi, and Louisiana—regions historically characterized by economic disadvantage. These states have the highest divorce rates in the country across nearly all birth groups. Native-born divorce rates are especially elevated, often exceeding 1.5, and individuals born in other states also experience above-average divorce levels. Some states in this cluster exhibit exceptionally high divorce rates among U.S. citizens born abroad (in the 2.0–3.0 range). Socioeconomic conditions likewise reflect substantial vulnerability: poverty and unemployment rates are high, median household income is low, and educational attainment lags behind national averages. The cluster is concentrated in the Deep South and Appalachia, regions where structural disadvantage, limited economic mobility, and lower education levels contribute to higher marital instability. Overall, Cluster 2 represents economically strained states where systematic socioeconomic pressure corresponds to consistently high divorce rates.

Cluster 3 — Moderate Divorce, Mixed Socioeconomic Profile

Cluster 3 includes states such as Maine, Vermont, Oregon, Indiana, Idaho, Missouri, Ohio, Georgia, Wisconsin, and Montana. This group represents a middle tier in both divorce patterns and socioeconomic conditions. Divorce rates generally fall between –0.5 and +1.0, indicating moderate levels of marital dissolution. Some Mountain West states show higher divorce rates among U.S. citizens born abroad, whereas many Midwest states display relatively stable outcomes. Socioeconomic indicators in this cluster occupy an intermediate position: median incomes are moderate, poverty rates are higher than in Cluster 1 but significantly lower than in Cluster 2, and education levels vary but generally align with national midpoints. The cluster is regionally diverse, spanning the Midwest, New England, and Mountain West, reflecting a broad spectrum of cultural and structural contexts. This heterogeneity explains the cluster's mixed socioeconomic profile and moderate divorce rates. Overall, Cluster 3 represents a balanced, middle-performing group without the extremes observed in the other two clusters.

8. Conclusion

This study shows that state-level divorce patterns in the United States are strongly shaped by underlying socioeconomic conditions. By combining birth-origin–specific divorce rates with income,

poverty, education, and employment indicators, we reveal clear structural differences across states. The K-means clustering results identify three distinct groups—affluent low-divorce states, economically strained high-divorce states, and a moderately positioned mixed group—patterns that are also confirmed through PCA visualization. These findings highlight the close link between economic resources, social disadvantage, and marital stability. While descriptive in nature, the analysis provides a clear framework for understanding regional variation in divorce and offers an interactive tool for further exploration.

9. Implications

The clustering results suggest that divorce behavior is closely aligned with broader structural factors such as income, poverty, and educational attainment. The spatial distribution of clusters—including the concentration of disadvantaged, high-divorce states in the South and Appalachia—underscores the importance of regional economic conditions for marital stability. These findings may inform policymakers by highlighting the potential benefits of improving educational access, reducing economic insecurity, and understanding how demographic composition shapes divorce rates, particularly among mobile or foreign-born populations.

10. Validations

Three elements support the internal validity of the clustering solution. First, the elbow method indicates that k=3 offers a stable and parsimonious representation of the data.
Second, the PCA projection—used only for visualization—shows clear separation among the clusters, confirming that the 10-variable structure naturally forms distinct groups. Third, the geographic patterns and cluster-level standardized means are consistent with known socioeconomic differences across U.S. regions. Although silhouette scores were not computed, these combined checks suggest that the clustering result is robust.

11. Limitations

ACS one-year estimates introduce sampling variability, especially for smaller subgroups such as U.S.-born abroad, where divorce rates exhibit extreme fluctuations.
Standardization, while necessary for clustering, obscures the absolute magnitude of differences.
K-means imposes assumptions of spherical clusters and sensitivity to initialization, although multiple random starts help mitigate this.
Because the analysis is cross-sectional and unsupervised, causal interpretation is not possible, and no external ground truth exists to definitively validate the clusters.
Finally, immigration patterns, age composition, and population turnover—factors not included in the dataset—may also influence subgroup divorce rates.

12. Future Direrctions

Future research could incorporate temporal ACS data to evaluate whether these cluster patterns persist over time. Additional demographic variables—such as religious affiliation, occupational structure, housing costs, or detailed age profiles—may further refine the cluster interpretation. External validation could be performed by comparing cluster membership with existing regional classifications. Finally, supervised models could extend this work by predicting divorce rates using socioeconomic factors or using cluster assignments as higher-level predictors in regression

frameworks.

References:

[1] Trent, K., & South, S. J. (1989). Structural determinants of the divorce rate: A cross-societal analysis. Journal of Marriage and the Family, 391-404.

Appendix;
1. url: https://kenziewcm.shinyapps.io/divorce-app/
2. Cluster membership and key averages

Cluster membership and key averages

State	cluster	divorce_rate_all	median_income	poverty_rate	unemployment_rate	college_plus_rate	foreign_born_share
Nevada	1	1.63	-0.03	-0.15	2.14	-1.15	1.53
Florida	1	1.09	-0.29	0.04	0.32	-0.08	1.90
Arizona	1	0.31	0.00	-0.11	0.33	-0.24	0.48
Washington	1	0.01	1.39	-0.78	0.65	0.68	0.88
Alaska	1	-0.20	1.10	-0.69	1.80	-0.53	-0.40
Colorado	1	-0.24	1.21	-0.91	0.03	1.67	-0.01
Rhode Island	1	-0.31	0.16	0.08	0.83	0.38	0.73
Connecticut	1	-0.47	1.13	-0.68	0.73	0.90	0.83
Virginia	1	-0.56	0.82	-0.86	-0.66	1.02	0.48
Hawaii	1	-0.72	1.49	-0.76	-0.92	0.20	1.33
Maryland	1	-0.80	1.66	-1.12	0.10	1.22	1.05
Texas	1	-0.87	-0.14	0.57	0.61	-0.18	1.31
Illinois	1	-1.01	0.14	-0.15	0.92	0.42	0.79
Massachusetts	1	-1.20	1.81	-0.87	-0.19	1.74	1.23
New York	1	-1.57	0.34	0.78	1.13	0.71	2.00
California	1	-1.68	1.45	-0.06	1.77	0.26	2.76
Utah	1	-1.78	1.18	-1.41	-0.50	0.40	-0.04
New Jersey	1	-2.04	1.77	-1.07	1.22	1.19	2.29
District of Columbia	1	-2.24	2.19	2.08	1.45	4.25	0.77
New Mexico	2	1.69	-1.06	1.73	1.07	-0.67	-0.06

Showing 1 to 20 of 51 entries

Previous 1 2 3 Next

3. Divorce rates by place of birth within clusters

State	cluster	divorce_rate_all	divorce_rate_born_state	divorce_rate_other_state	divorce_rate_native_abroad	divorce_rate_foreign_born
Nevada	1	1.626	-1.850	1.982	0.846	2.507
Florida	1	1.086	-0.833	1.476	0.178	2.578
Arizona	1	0.310	-1.575	1.020	0.745	0.941
Washington	1	0.014	0.129	0.554	-0.354	-0.622
Alaska	1	-0.202	-1.771	0.186	-0.203	1.969
Colorado	1	-0.244	-0.743	-0.259	-0.145	0.851
Rhode Island	1	-0.307	0.404	-0.820	0.781	-0.213
Connecticut	1	-0.471	-0.388	-0.793	0.763	1.003
Virginia	1	-0.556	-0.195	-0.483	-0.189	-0.562
Hawaii	1	-0.725	-0.603	0.553	-1.195	-0.542
Maryland	1	-0.798	-0.729	-0.410	-0.497	-0.190
Texas	1	-0.873	-0.526	-0.145	-0.432	-0.448
Illinois	1	-1.010	-0.409	-0.870	-0.333	-0.450
Massachusetts	1	-1.197	-0.681	-1.484	-0.788	0.354
New York	1	-1.571	-1.163	-2.254	-0.351	0.596