# SURVMETH 727 Final Project
# Economic Disparities Among the U.S. States in 2023: An Analysis of Regional Data from the Bureau of Economic Analysis

Ruis Ma and Zhenjing Zhou

**Github Link:** https://github.com/Ruisi-Ma/Final-Project727-Ma-and-Zhou.git

## Introduction

Economic disparities across the U.S. states have long been a focal point for policymakers, economists, and researchers. These disparities highlight the uneven distribution of wealth and economic opportunities within a nation. This project aims to investigate regional economic disparities among the U.S. states in 2023, providing a comprehensive analysis of key economic indicators to better understand the factors driving these differences. By leveraging data from the Bureau of Economic Analysis (BEA) Regional Dataset, this study examines disparities through five key economic indicators: `PerCap personal income`, `PerCap Real GDP`, `PerCap gross operating surplus`, `PerCap Personal Consumption Expenditures`(PCE), and `Employment PerThousand`. These indicators offer a multi-dimensional view of economic performance, consumption behavior, and employment trends across the fifty states.

The paper explores how these indicators vary across the states, identifies patterns of economic clustering, and examines the role of geographic proximity in shaping the economic disparities. To achieve this, we employ various analytical approaches, including simple linear regression to assess the relationships between `Employment PerThousand` and `PerCap personal income`, as well as between `PerCap Real GDP` and `PerCap PCE`. Additionally, we conduct multiple linear regression (MLR) to model `PerCap personal income` using the other key indicators as predictors. Finally, hierarchical clustering is applied to classify states into six clusters based on significant predictors, enabling a visual and conceptual understanding of economic disparities across regions. Through these analyses, we aim to uncover meaningful relationships among these indicators and provide insights into the broader question of how economic disparity manifests geographically within the United States.

## Data Set Description

This dataset is sourced from the Bureau of Economic Analysis (BEA) via the BEA API and is processed using the bea.R package. The dataset is complied from three key BEA tables: SAINC30 (Economic Profile), SAGDP1 (GDP by State Summary), and SAPCE2 (PCE by Major Types of Product). The extracted data were cleaned, calculated, and merged into a comprehensive dataset containing 50 rows and 22 columns. Each row represents a U.S. state, while each column represents a distinct economic indicator.

The key variables of interest are the numerical variables `PerCap personal income`, `PerCap Real GDP`, `PerCap gross operating surplus`, `PerCap Personal Consumption Expenditures`(PCE), and `Employment PerThousand`. `PerCap personal income` is the per capita income measured in current U.S. dollars. `PerCap Real GDP` represents the real gross domestic product per capita, adjusted for inflation. `PerCap gross operating surplus` refers to the per capita measure of the returns to capital within each state. PCE quantifies personal consumption expenditures per capita, while `Employment PerThousand` denotes the number of employed individuals per 1,000 residents. In addition, categorical variables such as `State`, `Name`, and `Region` are used to distinguish states and their geographic regions.

This dataset is observational, meaning causal inferences cannot be made from the analysis. However, associations, patterns, and groupings among states can be explored. Of particular interest is how economic indicators vary across regions and how they relate to one another. The inclusion of hierarchical clustering provides insight into the grouping of states with similar economic characteristics, further illuminating the role of geographic proximity in regional economic disparities.

### Summary of Several Numerical Variables

Based on the numerical variables collected, we organized them in the table displayed as below to explore the min, max, sd, etc.

Table 1: Summary of Numerical Variables

| | mean | median | min | max | sd |
|---|---|---|---|---|---|
| GeoFips* | 25.5000 | 25.5000 | 1.0000 | 5.000000e+01 | 1.457738e+01 |
| GeoName* | 25.5000 | 25.5000 | 1.0000 | 5.000000e+01 | 1.457738e+01 |
| Population | 6684718.4600 | 4549951.5000 | 584057.0000 | 3.896519e+07 | 7.488403e+06 |
| PerCap personal income | 67549.3600 | 66206.5000 | 49652.0000 | 9.059600e+04 | 9.377142e+03 |
| PerCap personal current transfer receipts | 12797.1200 | 12805.0000 | 8490.0000 | 1.574100e+04 | 1.362092e+03 |
| PerCap income maintenance benefits | 973.9400 | 945.5000 | 553.0000 | 1.515000e+03 | 2.425942e+02 |
| PerCap unemployment insurance compensation | 84.2600 | 65.0000 | 16.0000 | 2.700000e+02 | 6.127016e+01 |
| PerCap retirement and other | 11738.8600 | 11723.5000 | 7712.0000 | 1.440400e+04 | 1.236208e+03 |
| PerCap dividends interest and rent | 14072.7800 | 13438.5000 | 8329.0000 | 2.989900e+04 | 3.594299e+03 |
| Employer contributions to social insurance | 16260.7380 | 9262.2500 | 1508.4000 | 1.069683e+05 | 1.945923e+04 |
| Total employment | 4308016.4800 | 2771701.5000 | 442378.0000 | 2.541730e+07 | 4.909469e+06 |
| WageSalary employment | 3203964.8400 | 2074832.0000 | 295701.0000 | 1.890064e+07 | 3.578877e+06 |
| Average wages and salaries | 66864.3800 | 64789.0000 | 50481.0000 | 9.127000e+04 | 9.756022e+03 |
| Real GDP | 447665.6600 | 255283.5000 | 35236.6000 | 3.248657e+06 | 5.813655e+05 |
| Gross operating surplus | 231055.4500 | 131096.4500 | 15375.7000 | 1.609914e+06 | 2.984519e+05 |
| PerCap PCE | 54373.5800 | 53987.5000 | 42131.0000 | 6.910100e+04 | 6.467912e+03 |

| | | | | | |
|---|---|---|---|---|---|
| PerCap healthcare expenditures | 9194.7000 | 9131.0000 | 6833.0000 | 1.301500e+04 | 1.310363e+03 |
| PerCap recreation expenditures | 2044.3600 | 2054.0000 | 1171.0000 | 3.274000e+03 | 5.456397e+02 |
| Employment PerThousand | 644.9817 | 645.3118 | 510.6095 | 7.630797e+02 | 4.873877e+01 |
| WageSalary Employment PerThousand | 482.5381 | 484.7964 | 407.0645 | 5.834109e+02 | 3.519782e+01 |
| PerCap Real GDP | 62789.4632 | 61576.0769 | 40667.0091 | 9.152271e+04 | 1.150888e+04 |
| PerCap gross operating surplus | 33147.1120 | 31755.8443 | 21047.5254 | 4.974561e+04 | 7.163792e+03 |

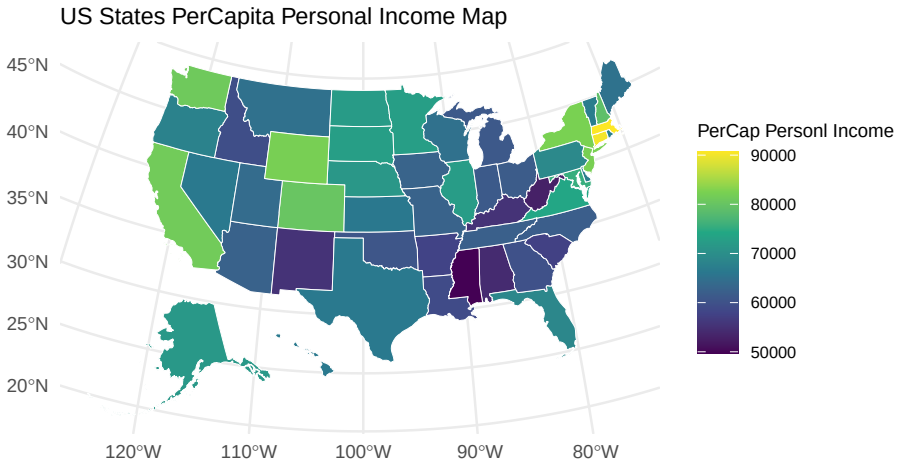
## Exploratory Data Analysis: Description and Visualization

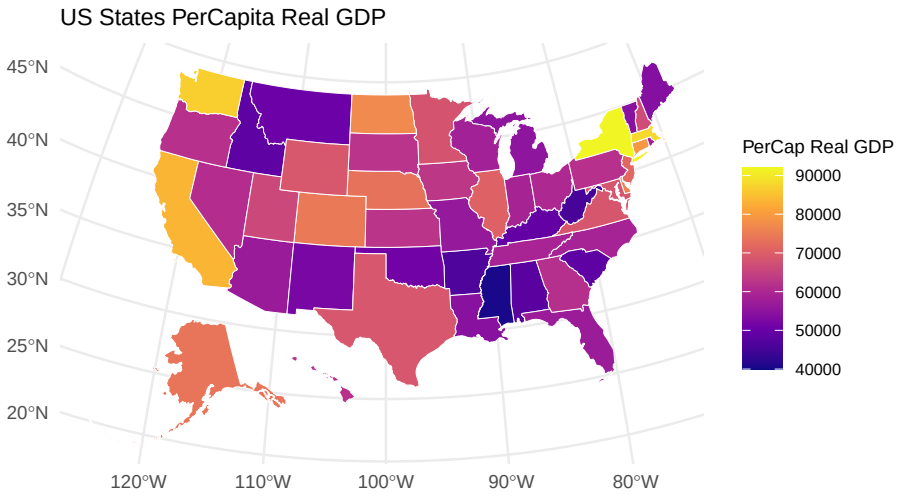
Through data cleaning, transformation, and statistical analysis, we explored the relationships among these indicators and examined the role of geographic proximity in shaping economic disparities.

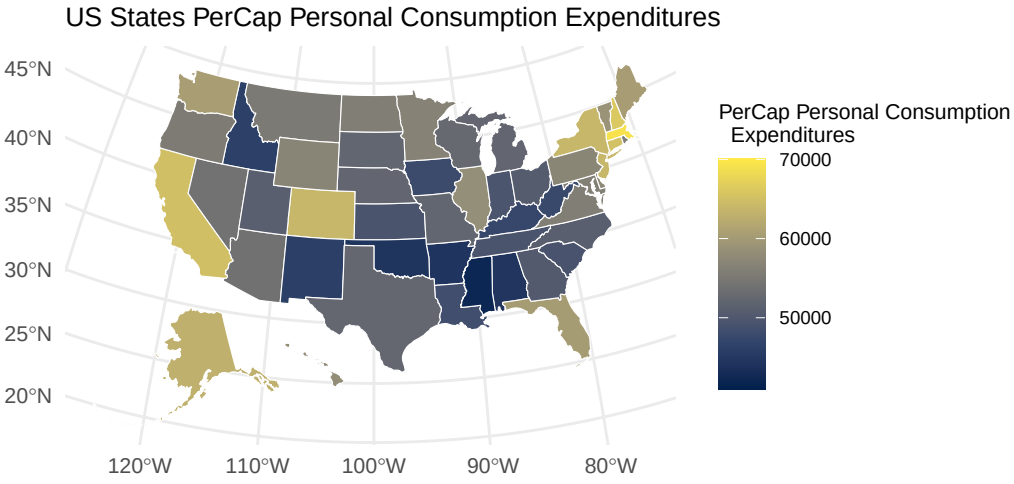
After merging the relevant BEA tables and creating a clean dataset, we first conducted descriptive analysis to explore the key variables. Using the usmap package, we visualized the geographic distribution of these five indicators.

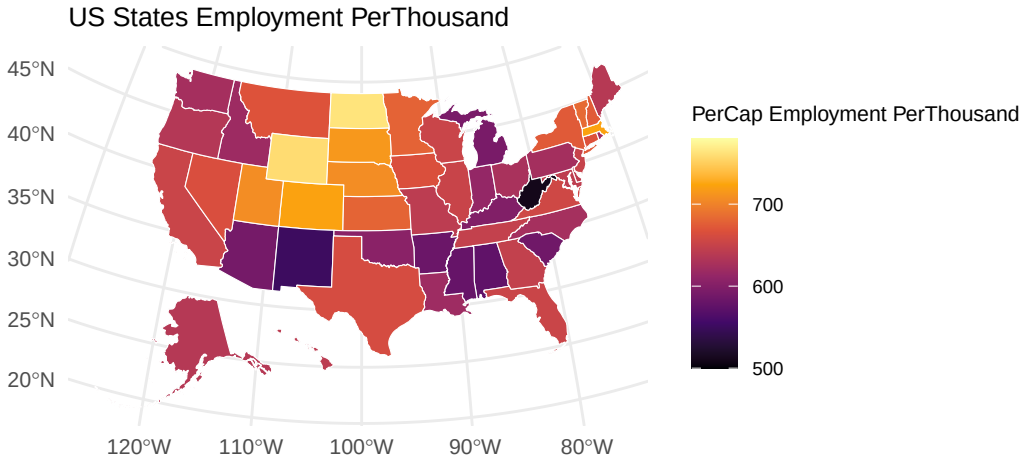
### Distribution

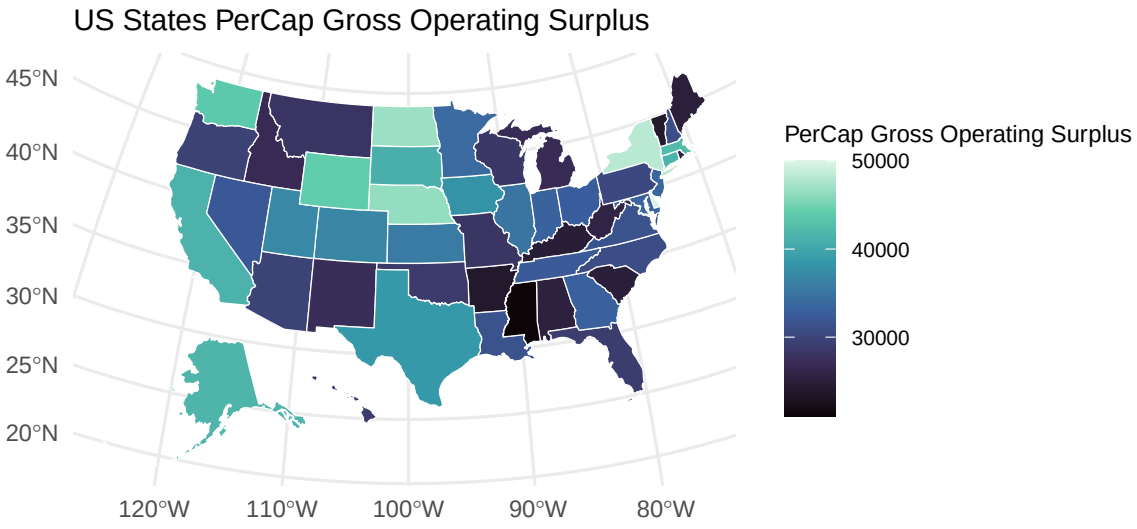
We used the geographical boundary data of 50 states from the usmap. package and our merged dataset to illustrate the geographic distribution of five key economic variables. The colors range from dark to light, representing the values of each economic indicator from low to high across the states.




US States PerCapita Real GDP
45°N
40°N
35°N
30°N
25°N
20°N
120°W
110°W
100°W
90°W
80°W
PerCap Real GDP
90000
80000
70000
60000
50000
40000


US States PerCap Personal Consumption Expenditures
45°N
40°N
35°N
30°N
25°N
20°N
120°W
110°W
100°W
90°W
80°W
PerCap Personal Consumption Expenditures
70000
60000
50000


US States Employment PerThousand
45°N
40°N
35°N
30°N
25°N
20°N
120°W
110°W
100°W
90°W
80°W
PerCap Employment PerThousand
700
600
500


US States PerCap Gross Operating Surplus
45°N
40°N
35°N
30°N
25°N
20°N
120°W
110°W
100°W
90°W
80°W
PerCap Gross Operating Surplus
50000
40000
30000

### Multiple Linear Model

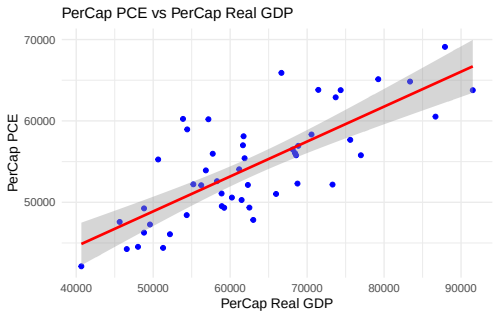
In addition, we employed a multiple linear model to examine the potential effects of various economic indicators - `PerCap Real GDP`, `PerCap PCE`, `Employment PerThousand`, and `PerCap grows operating surplus` - on `PerCap personal income`. As shown in the table, `PerCap Real GDP`, `PerCap PCE`, and `Employment PerThousand` have statistically significant effects on `PerCap personal income`. Specifically, for every one-unit increase in `PerCap real GDP`, the `PerCap personal income` is expected to increase by 0.433 units, while holding other variables constant ($p$<.05). Similarly, for every one-unit increase in `PerCap PCE`, `PerCap personal income` is expected to increase by 0.632 units, while holding other variables constant ($p$<.001), highlighting a strong positive relationship. For every one-unit increase in `Employment PerThousand`, `PerCap personal income` is expected to increase by 40.681 units ($p$<.05), indicating a statistically significant relationship. Unexpectedly, `PerCap gross operating surplus` does not show a significant effect on `PerCap personal income` ($p$=ns).

Table 2: MLR: Economic Indicators Affecting PerCap Personal Income

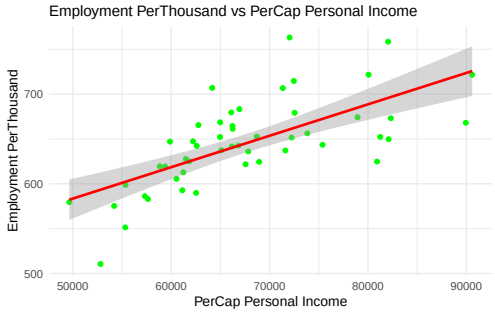
| Term | Estimate | Std. Error | t-value | p-value | Significance |
|---|---|---|---|---|---|
| (Intercept) | -13902.265 | 6689.141 | -2.078 | 0.043 | * |
| PerCap Real GDP | 0.443 | 0.134 | 3.304 | 0.002 | ** |
| PerCap PCE | 0.623 | 0.138 | 4.507 | 0.000 | *** |
| Employment PerThousand | 40.681 | 12.933 | 3.146 | 0.003 | ** |
| PerCap gross operating surplus | -0.195 | 0.172 | -1.136 | 0.262 | |

### Simple Linear Model

To investigate possible association between two specific variables, we created a couple of simple linear models. In the first model, we used `PerCap Real GDP` as the predictor and `PerCap PCE` as the response variable. The corresponding scatter plot features a red trend line, which suggests a linear relationship between these two variables. In effect, it shows a positive correlation between `PerCap Real GDP` and `PerCap PCE`. As `PerCap Real GDP` increases, `PerCap PCE` tends to increase. The blue dots are scattered around the trend line, indicating some variability in the data. As such, this visualization can be used to understand how economic growth (measured by Real GDP) impacts consumer expanding behaviors (measured by PCE) on a per capita basis.

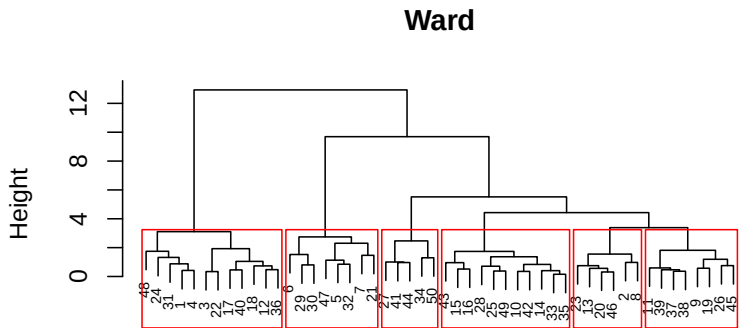


The second simple linear regression model was conducted to examine the relationship between `Employment PerThousand` and `PerCap personal income`. In the scatter plot, green dots represent individual data points, with each point corresponding to an observation where `Employment PerThousand` is plotted against `PerCap personal income`. The red line illustrates the fitted regression line, inferring a positive trend between these two variables. This indicates that as `PerCap personal Income` increases, `Employment Per Thousand` tends to increase as well. The gray shaded area around the regression (red) line represents the confidence interval, providing the range within which the true relationship between the variables is likely to lie with a certain level of confidence (95%). Despite the positive trend, the dispersion of data points around the line suggests that additional factors may also influence employment levels, indicating variability that could not be fully explained by `PerCap Personal Income` alone.



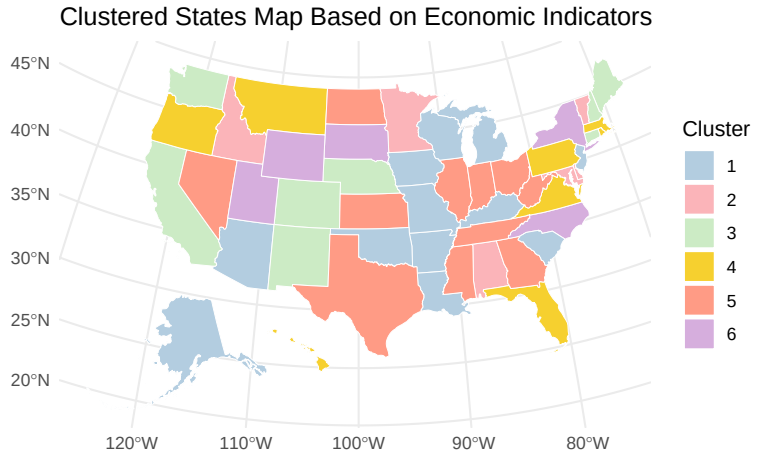

### Clustering - Dendrogram

Moreover, we present a visualization of clusterings via Dendrogram and a Clustered States Map. The dendrogram employs Ward's method for clustering, illustrating the hierarchical clustering of the U.S. states based on various economic indicators. Our hierarchical clustering is based on the following four variables: `PerCap personal income`, `PerCap Real GDP`, `PerCap PCE`, and `Employment PerThousand`. States are grouped into clusters, and highlighted by red boxes. Then, six primary clusters are identified.



### Clustering - Clustered States Map

Based on the six primary clusters, we created a 6-color choropleth map. This map visualizes how states with similar conditions are geographically distributed, colored by their respective clusters that are numbered from 1 to 6. States within the same cluster are often geographically proximate, suggesting regional economic similarities and disparities.




## Discussion and Conclusion

In this project, we aim to investigate economic disparities among the U.S. states in 2023 using data from the Bureau of Economic Analysis (BEA) Regional Dataset. Our analysis focuses on five key economic indicators: `PerCap Personal Income`, `PerCap Real GDP`, `PerCap Gross Operating Surplus`, `PerCap PCE`, and `Employment PerThousand`.

Our analysis revealed several important insights. First, economic disparities among the U.S. states are not random but follow clear geographical patterns. Economic indicators such as `PerCap Real GDP`, `PerCap PCE`, and `PerCap Personal Income` are consistently higher in certain regions, such as the Northeast and the West, while being lower in parts of the Southeast. Second, `PerCap Real GDP`, `PerCap PCE`, and `Emplyment PerThousand` significantly affect `PerCap Personal Income`, providing further insight into the drivers of economic disparities. Third, production (as measured by GDP) has a strong positive relationship with PCE, highlighting the role of economic growth in driving consumer activity. Last, geographical proximity appears to be a significant factor in economic disparity, as neighboring states tend to exhibit similar levels of economic development, possibly due to shared infrastructure, industry linkages, and trade relationships.

Unexpectedly, `PerCap gross operating surplus` does not show a significant effect on `PerCap personal income`. This lack of significance reminds us that while gross operating surplus may contribute to overall economic production, it does not directly translate into higher personal incomes on a per capita basis.The distribution of surplus within businesses, reinvestment into capital, or variations in business structures across different states could be the possible factors. It highlights the

complexity of economic dynamics and the multifaceted nature of economic disparities, suggesting that other factors that were not included in the analysis might be more or less directly relevant to personal income levels.

Noticeably, the resulting choropleth maps revealed clear regional patterns. States in the Northeast and the West generally exhibited higher `PerCap PersonalIncome` and `PerCap Real GDP`, while certain southern and central states showed relatively lower values. This visualization highlighted the clustering of states with similar economic characteristics, which provided the foundation for further statistical analysis.

However, some limitations should be noted. The observational nature of the data limits our ability to draw causal inferences. Additionally, the clustering approach relies on the choice of distance metrics and clustering methods, which could influence the final grouping of states. While Ward's method provided useful insights, alternative methods could produce slightly different results. Furthermore, unobserved variables not included in the dataset, such as industry composition or policy differences, may also contribute to the observed disparities.

In conclusion, this study provides a comprehensive examination of economic disparities among U.S. states in 2023. Our findings reveal that economic indicators such as `PerCap Real GDP`, `PerCap PCE`, and `PerCap Personal Income` exhibit clear geographic patterns, with clustering effects that align with regional proximity. The positive relationship between GDP and PCE highlights the role of production in driving personal consumption, while clustering analysis underscores the significance of shared regional economic conditions. These results provide a foundation for further exploration of the factors driving regional economic disparities and suggest that policymakers should consider the role of geographic proximity and regional interdependence when designing economic policy.