README

Christian McIntire

2025-03-15

# Introduction

Housing affordability remains a pressing issue across the United States, particularly in regions experiencing rapid urbanization and population growth. The Southeastern United States has seen significant population growth in recent years, whereas the Midwestern United States has remained much more stable or even seen population decline (Biernacka-Lievestro & Fall, 2023).

This study seeks to examine the relationship between population density, growth, and housing prices in these two regions, to determine how demographic shifts impact affordability. The Southeastern and Midwestern regions are defined as follows:

* **Southeastern U.S.**: Arkansas, Louisiana, Kentucky, Tennessee, Mississippi, Alabama, Georgia, Florida, North Carolina, South Carolina, Virginia, and West Virginia.
* **Midwestern U.S.**: Minnesota, Wisconsin, Michigan, Ohio, Indiana, Illinois, Iowa, and Missouri.

### Research Questions

1. How has population density historically affected housing prices in the Southeastern and Midwestern United States?
2. How can these patterns be measured to predict future changes in the U.S. housing market?

# Methodology

This study employs the following techniques:

* **Choropleth Maps** – Visualizing county-level home prices using spatial analysis.
* **Hedonic Regression Models** – Decomposing housing prices into key factors, including regional fixed effects and population density.
* **Panel Data Estimation** – Handling time-series data with multiple geographic units.
* **ARIMA Forecasting** – Time-series forecasting to predict future housing prices.

### Data Sources

This project uses publicly available datasets from:

1. **National Association of Realtors** – County-level median home prices (2024).
2. **Federal Reserve Bank of St. Louis (FRED)** – Housing Price Index data (2024).
3. **U.S. Census Bureau** – County-level population estimates (2024).
4. **Zillow Research Data** – Historical home values and unique price metrics (2024).
5. **TIGER/Line Shapefiles** from the **U.S. Census Bureau** – Geographic county/state boundaries.
   * **Shapefiles were too large for GitHub but can be downloaded here:**  
     🔗 [U.S. Census Bureau Shapefiles](https://www2.census.gov/geo/tiger/TIGER2024/)

### Key Findings

* Population trends in the **Southeastern U.S.** have shown significant growth, whereas the **Midwestern U.S.** has remained more stable or declined in some areas.
* Housing prices in the **Southeast** have seen steeper increases, especially in urban and suburban areas.
* **Choropleth maps** reveal strong regional patterns in home prices, with coastal and metropolitan counties exhibiting the highest median prices.
* **ARIMA Forecasting** suggests continued home price growth in both regions, albeit at different rates.

### Next Steps

#### 1. Hedonic Regression Model

* A more detailed **regression model** will be used to control for multiple factors affecting home prices, such as **income levels, employment rates, and local economic indicators**.

#### 2. Panel Data Estimation

* This method will allow for **cross-sectional and time-series analysis**, helping isolate the impact of different regional factors on housing prices.

## References

* **TIGER/Line Shapefiles:** [U.S. Census Bureau](https://www2.census.gov/geo/tiger/TIGER2024/)
* **Choropleth Mapping in R:** [R Graph Gallery](https://r-graph-gallery.com/327-chloropleth-map-from-geojson-with-ggplot2.html)
* **Customizing Legends in ggplot2:** [GeeksforGeeks](https://www.geeksforgeeks.org/control-size-of-ggplot2-legend-items-in-r/)
* **National Association of Realtors (2024)**
* **Federal Reserve Bank of St. Louis (FRED) (2024)**
* **U.S. Census Bureau Population Estimates (2024)**
* **Zillow Research Data (2024)**