A Comparative Analysis of US House Prices Using Violin Plots

This visualization illustrates the distribution of U.S. house prices across four major regions, aiming to identify regional disparities in housing market values. The data were obtained from the USA Real Estate Dataset on Kaggle, which provides detailed state-level information on property prices across the United States. In the visualization:

- Each violin represents the distribution of home listing prices within one U.S. region (Midwest, Northeast, South, and West).

- Width indicates the density of listings at a given price level.

- Inner box/line represents the median and interquartile range.
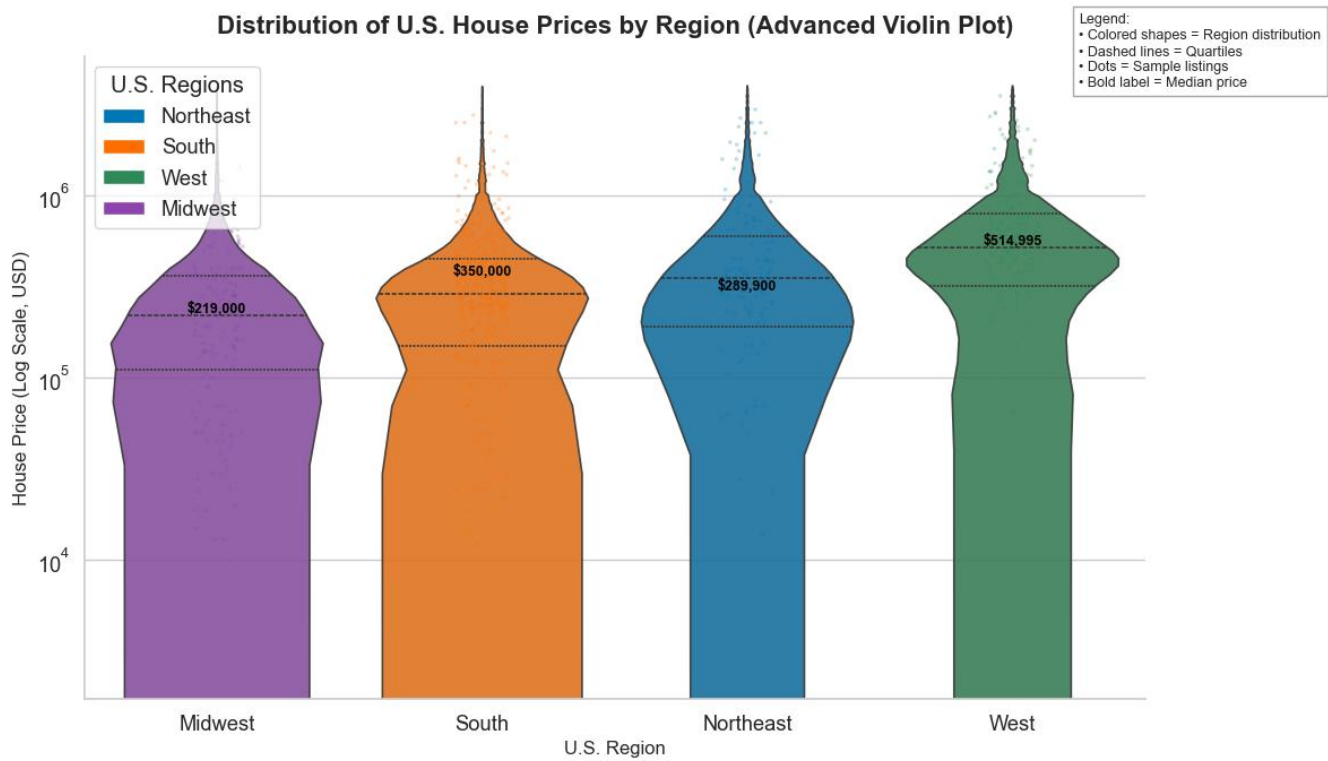
- Colors distinguish geographic regions.

Figure 1. Distribution of US House Prices by Region

Key Findings

- Median home prices vary across the regions. The South and West regions generally show higher median values than the Midwest and Northeast regions.

- The Midwest region has the lowest median home price ($219,000), indicating generally more affordable housing compared to other regions.

- South region shows a moderate median price ($350,000), with a broader spread suggesting higher variation in property values.

- The West region displays the broadest price distribution, indicating a greater range of high-end and low-end listings.

- Northeast has a median price of $289,900, slightly below the South, but also exhibits greater clustering around the median, indicating that prices are less dispersed.

- West stands out with the highest median price ($514,995) and a visibly tighter upper distribution, indicating consistently high housing costs.

- Midwest homes are generally the most affordable, as the region has the lowest overall house prices.

Data Source and Processing

The analysis uses the USA Real Estate Dataset obtained from Kaggle:

https://www.kaggle.com/datasets/ahmedshahriarsakib/usa-real-estate-dataset

The dataset contains property-level listing information, including prices, state, and regional classification across the United States. Data processing involved the following steps:

- Listings with missing or invalid price values were removed.

- States were grouped into four Census regions: Midwest, Northeast, South, and West.

- Housing prices were analyzed as continuous numerical variables.

Visualization Method

- Python was used for all analysis and visualization.

- Violin plots were chosen to represent full price distributions rather than relying solely on summary statistics.

- This approach allows comparison of central tendency, spread, skewness, and outliers across regions.

Significance Statement

Housing affordability is a critical social and economic issue in the United States. By visualizing housing price distributions across regions, this figure highlights the scale and structure of regional inequality in the housing market. Rather than focusing only on average prices, the visualization reveals how price variability and extremes differ geographically, offering insights for policymakers, urban planners, and prospective homeowners. This visualization demonstrates how information visualization techniques can transform complex market data into interpretable, decision-relevant knowledge.

GitHub Documentation

A GitHub repository accompanies this report and contains:

- Python code used for data cleaning and visualization

- The original dataset (a link to Kaggle)

- The final project report (PDF)

GitHub link: [MAlonzi25/INFSCI2415-Final-Report: Information Visualization Final Report: U.S. Housing Market](#)