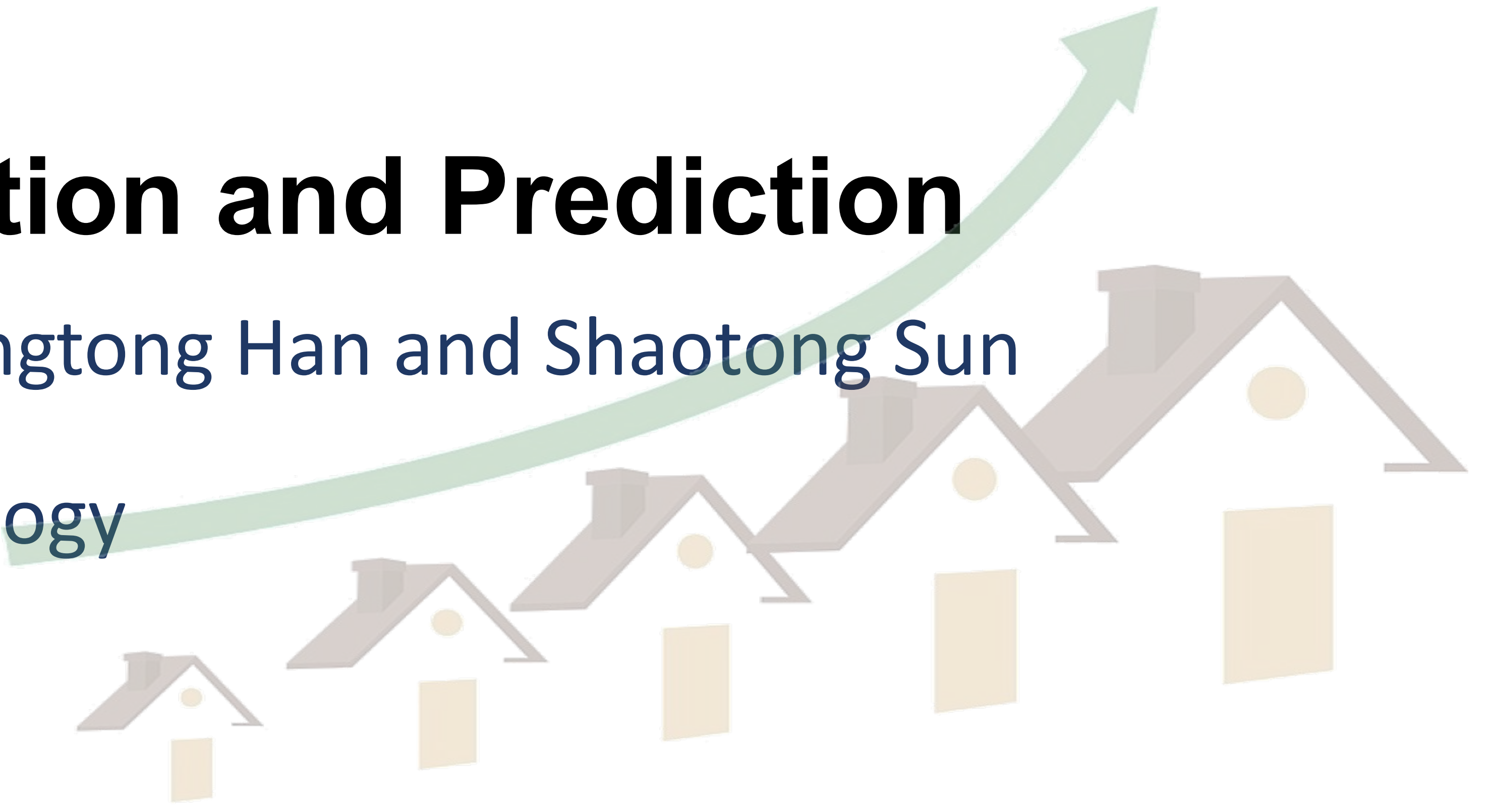
U.S. Real Estate Market Visualization and Prediction

Qianyun Chen, Chenyu Dai, Fei Ding, Shuge Fan, Lingtong Han and Shaotong Sun

Georgia Institute of Technology



Summary

We visualize the **U.S.real estate price** distributions across fine-grained geographical regions and observe their trends. Our tool reveals market insights like the price **hot/cold spots** and **future trends** to users, so they efficiently analyze the data and make informed **investment decisions**. We use the historical data to make price forecasts in the future with **regression** and **time-sequence models**. We **score and rank** the most promising investment regions.

Approach

Task	Method	Interaction
Visualizations		
Price/Change Distribution	Choropleths Sorted Bar Chart	Time Slider & Tooltips Highlighted Bars
Estate Price History	Line Chart	Region Select
State Listing Count	Piechart	Time Slider & Tooltips
Proximity Price Finder	Concentric Circle Diagram	Cursor & Sliders & Tooltips
Machine Learning		
Estate Price Prediction (1 Year)	Linear Regression Prophet Model	*Displayed on Extended Line Chart*
Region Ranking	Sharpe Ratio	*Displayed as Table*

Prediction Models

We used two models for prediction purposes, the **linear regression model** and the **additive regression model (Prophet)**. We sampled 10-15 price intervals for every county and divided them into features and targets for training purposes.

The linear regression model utilizes linear predictor functions where the parameters are learned from these training data. The additive regression model (Prophet) relies on nonparametric regression models with one-dimensional smoother to discover the relationship between features and targets we obtained for each county.

Dataset

The price dataset is downloaded from Kaggle. We combined the price dataset with the county's geographic information dataset for our project. We processed the combined data set to SQLite database that contains two tables, including a state-level table and a county-level table. The period we analyzed is from **July 2016** to **July 2021**.

Dataset Size	241.75MB **Data** + 31MB **Map**
Number of records	995,172
Number of Locations	**50 States** / **1578 Counties**
Number of Columns	25
Time Span	5 Years

Experiments & Analysis

To compare these two models, we obtained the following evaluation metrics.

Model	Avg. MSE	Mean Price Error	R2
Linear Regression	928.47 M	30.47 k	0.9445
Additive Regression	191.51M	13.84 k	0.9677

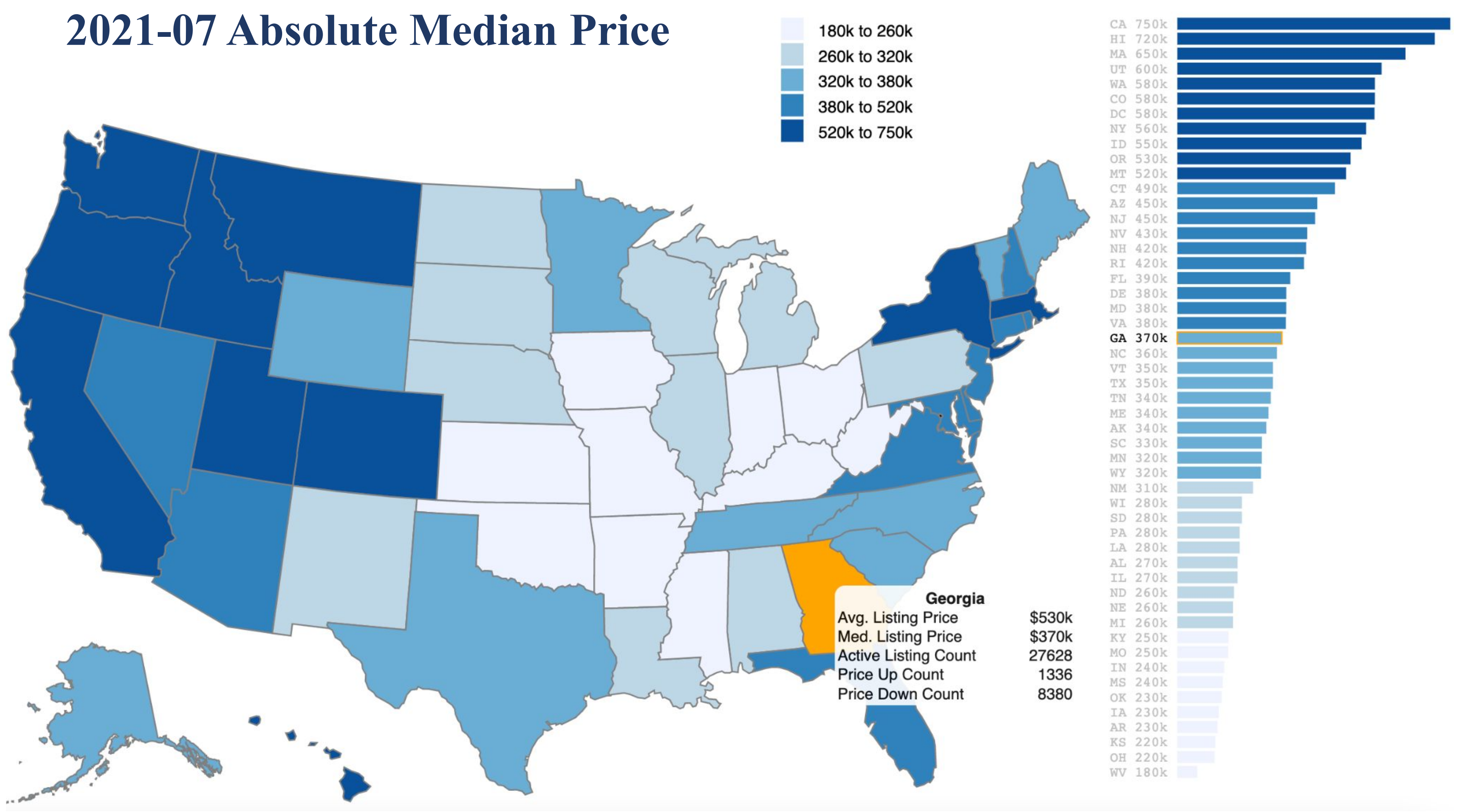
We found that the **additive regression model** is a better model for forecasting the future real estate prices.

$$S_a = \frac{E\left[R_a - R_b\right]}{\sigma_a}$$

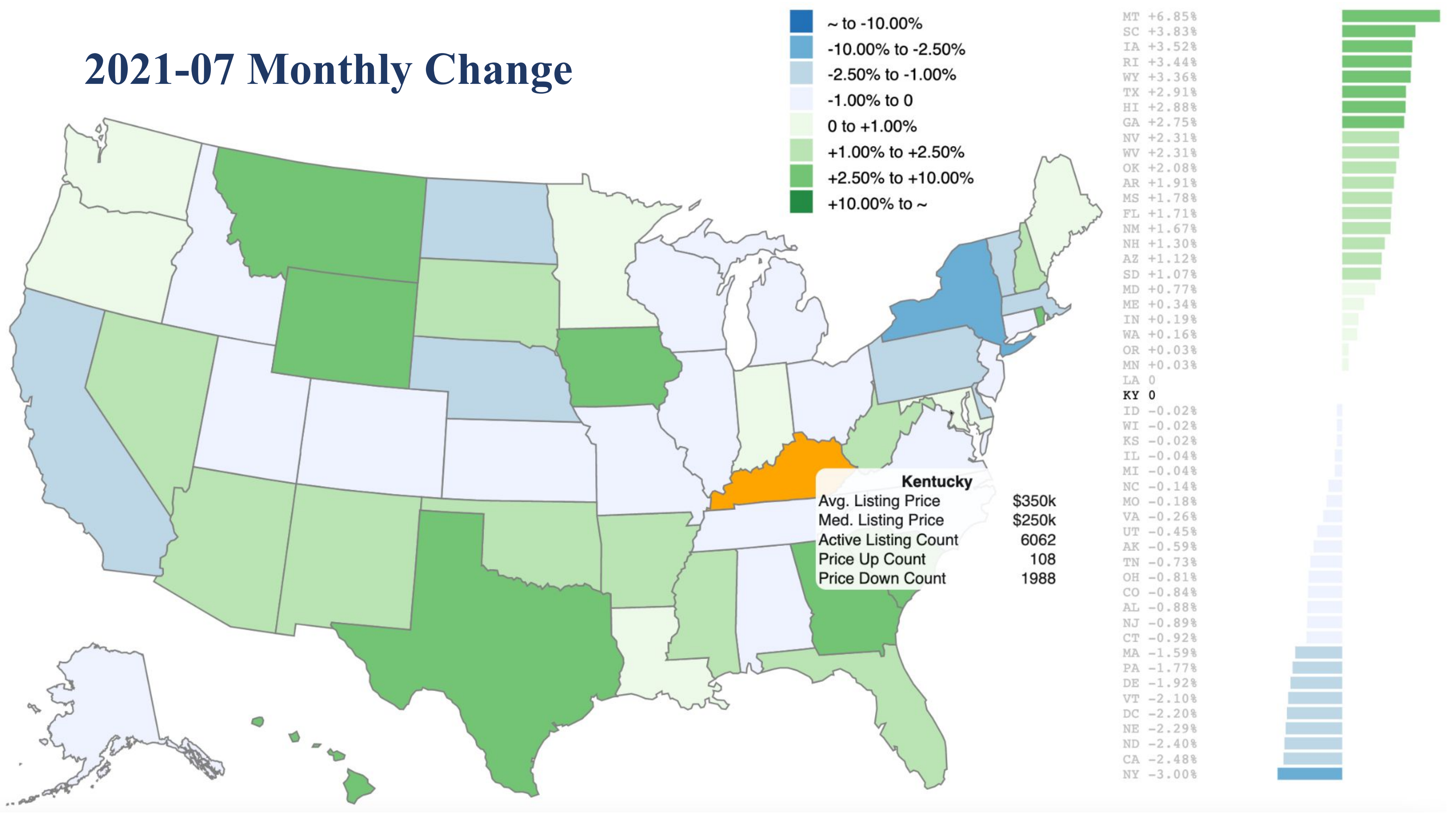
Our tool also calculates **the sharpe ratio** for each county utilizing each county's MSE data. The **higher** the sharpe ratio is, the **better** the investment would return related to its risk. This is super useful for investors to identify which area has a high real estate **investment potential.** We empirically find this approach of metric ranking matches typical human choices up to 60%.

Country Level Estate Price Visualization

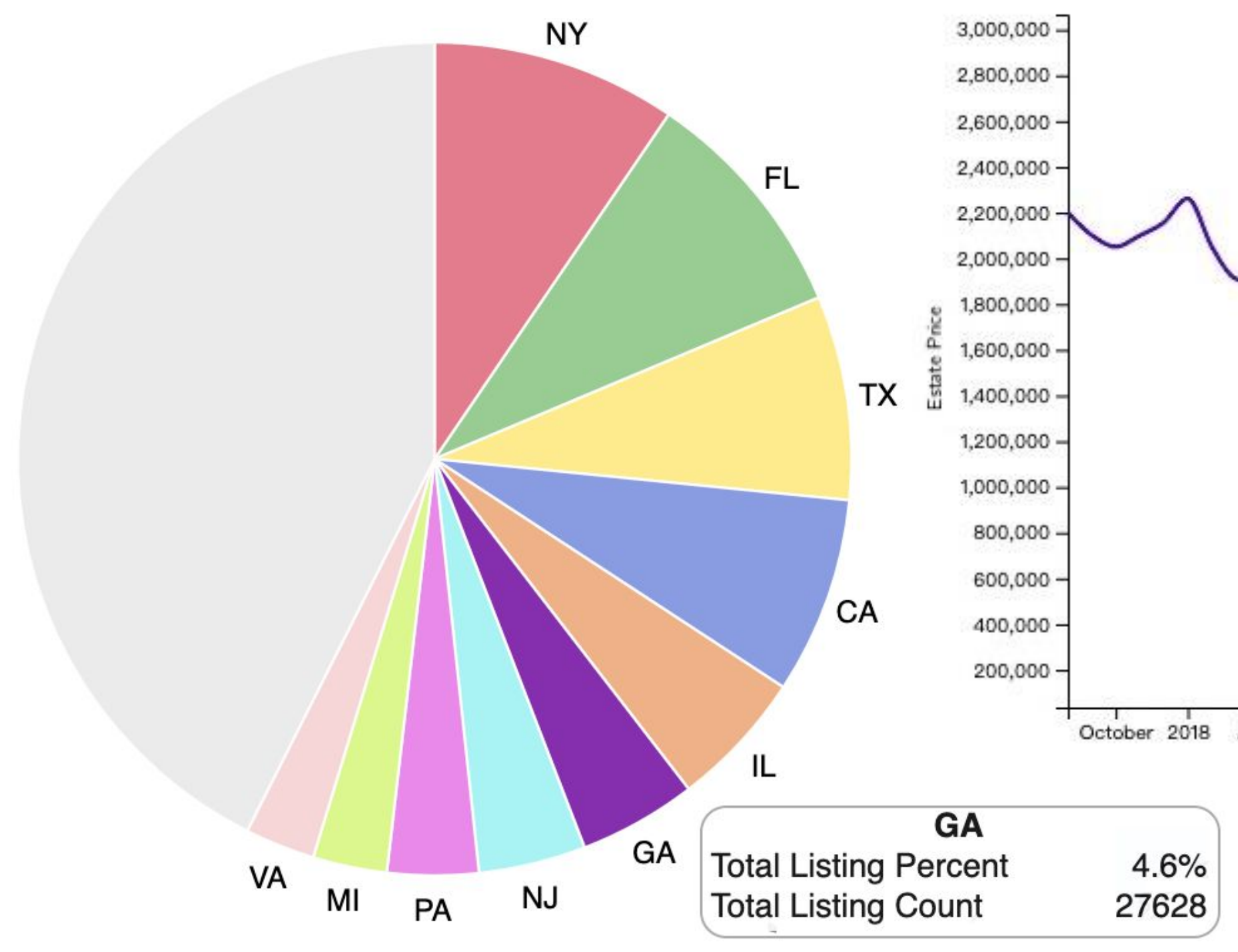
2021-07 Absolute Median Price



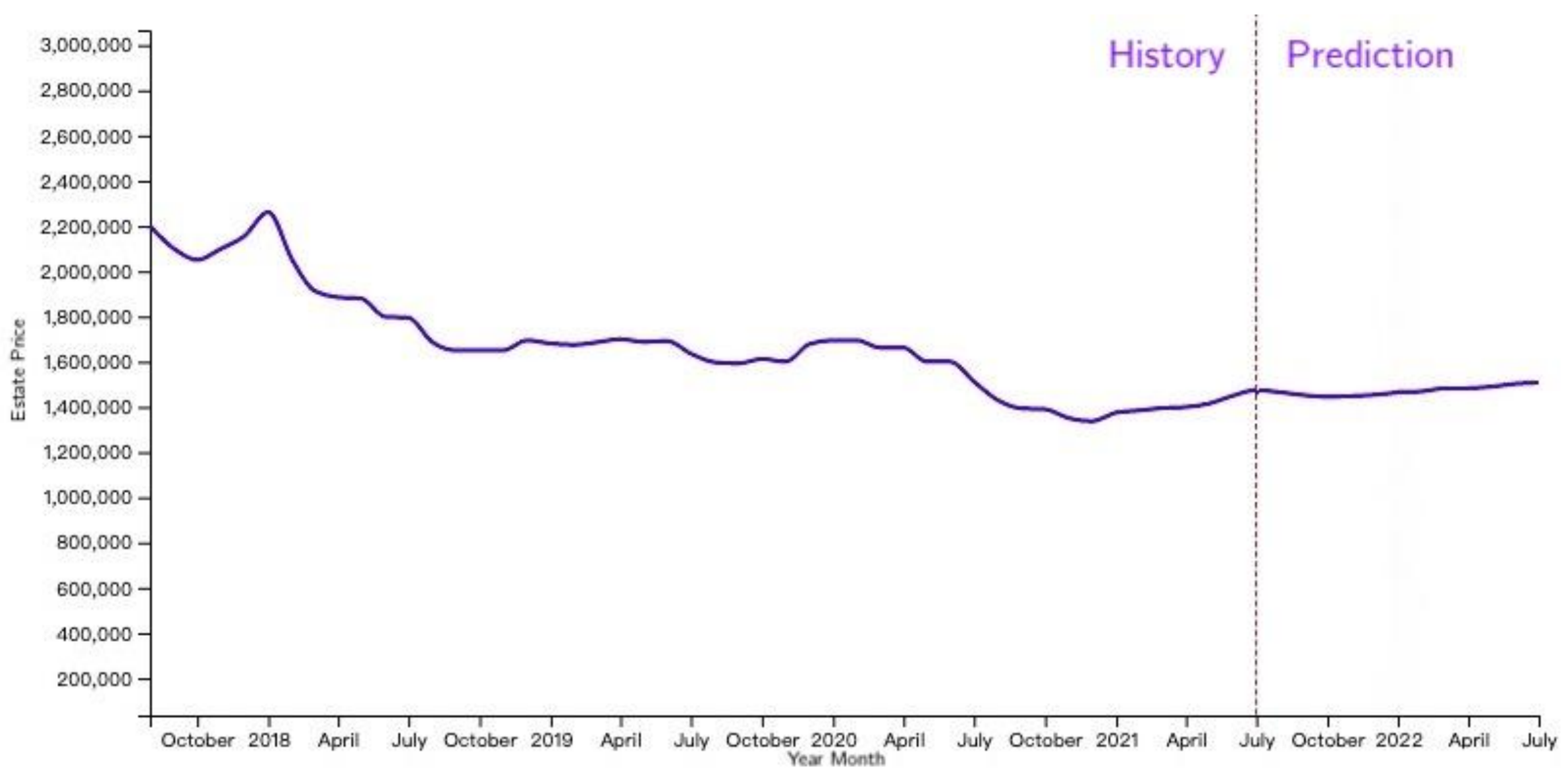
2021-07 Monthly Change



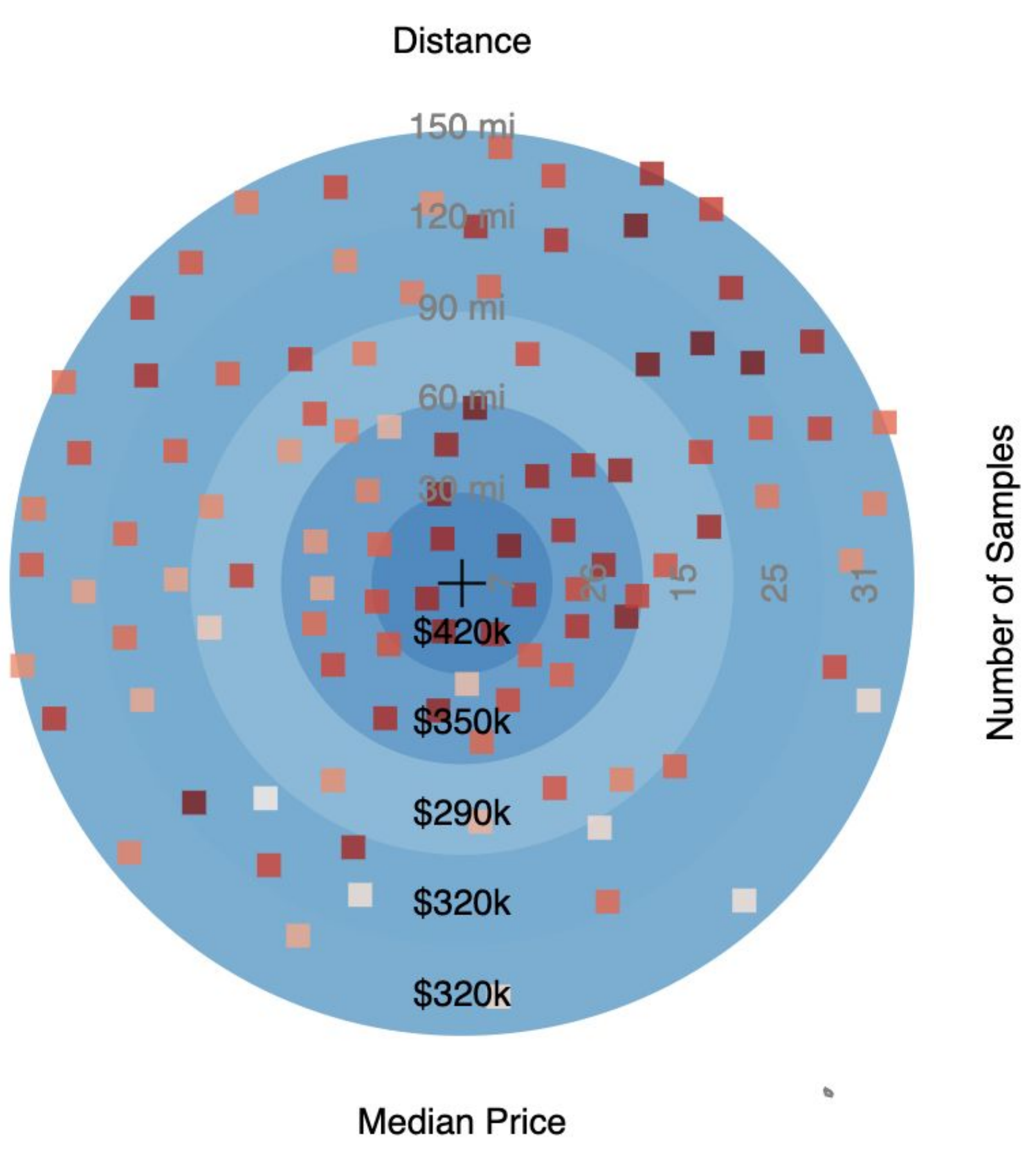
2021-07 Listing Count Distribution (Top 10)



New York Median Price Trend 2017 to 2022



2021-07 Concentric circles showing the geographic distribution of housing prices around Atlanta (150 miles radius + 5 circles)


Distance
150 mi
120 mi
90 mi
60 mi
$420k
$350k
$290k
$320k
$320k
Number of Samples
Median Price