



# A HISTORICAL ANALYSIS OF SINGLE-HOME VALUES IN THE U.S. FOR POTENTIAL INVESTMENT

USC VITERBI, DATA ANALYTICS BOOT CAMP (FALL 2019) | PROJECT 1

by Vito Perez, Gerard Tieng, & Peirang Xu

## Motivation & Summary

# I want to invest in a home...

### CORE MESSAGE

At the current median value at \$231,000 and rising\*, purchasing a home is a serious decision that should not be taken lightly.

Our group believes that the use of data analysis and visualization to evaluate historical home values will be able to better inform anyone looking to purchase a home about their potential investment opportunity.

### MAIN INQUIRY

*"If I had the money to purchase a single-family home in the U.S., where would be the best place to invest and receive the best return on investment?"*

Our primary objective with this analysis is to maximize monetary gain.

### BRIEF SUMMARY

While it is understood that home prices can be affected by numerous factors (economy, taxes, crisis, etc.), our analysis was able to yield satisfactory results in:

Determining percentage increases/decreases in home values and the ranked comparisons of such values across the nation.

\*Zillow

# Where do we start...?

## QUESTION BREAKDOWN

1. How much money should one expect to make by investing in a home?
2. Which are the best locations in which to invest for maximum return on investment?
3. Can I find suitable locations to invest based on budget/income?

## WHAT DATA IS NEEDED?

Recordings of home values over time would provide the necessary data to answer the questions posed by this project to observe property appreciation.

## WHERE WAS IT FOUND?

Data for this project was provided by American online real estate database Zillow, which listed median home values in major markets across the nation from 1996 to 2019 in a CSV file.

### Summary::

*"Zillow Home Value Index (ZHVI): A smoothed, seasonally adjusted measure of the median estimated home value across a given region and housing type. It is a dollar-denominated alternative to repeat-sales indices. Zillow also publishes home value and other housing data for local markets, as well as a more detailed methodology and a comparison of ZHVI to the S&P CoreLogic Case-Shiller Home Price Indices."*

# Data Cleanup & Exploration



# Importing Libraries in Jupyter

## NumPy (Python library)

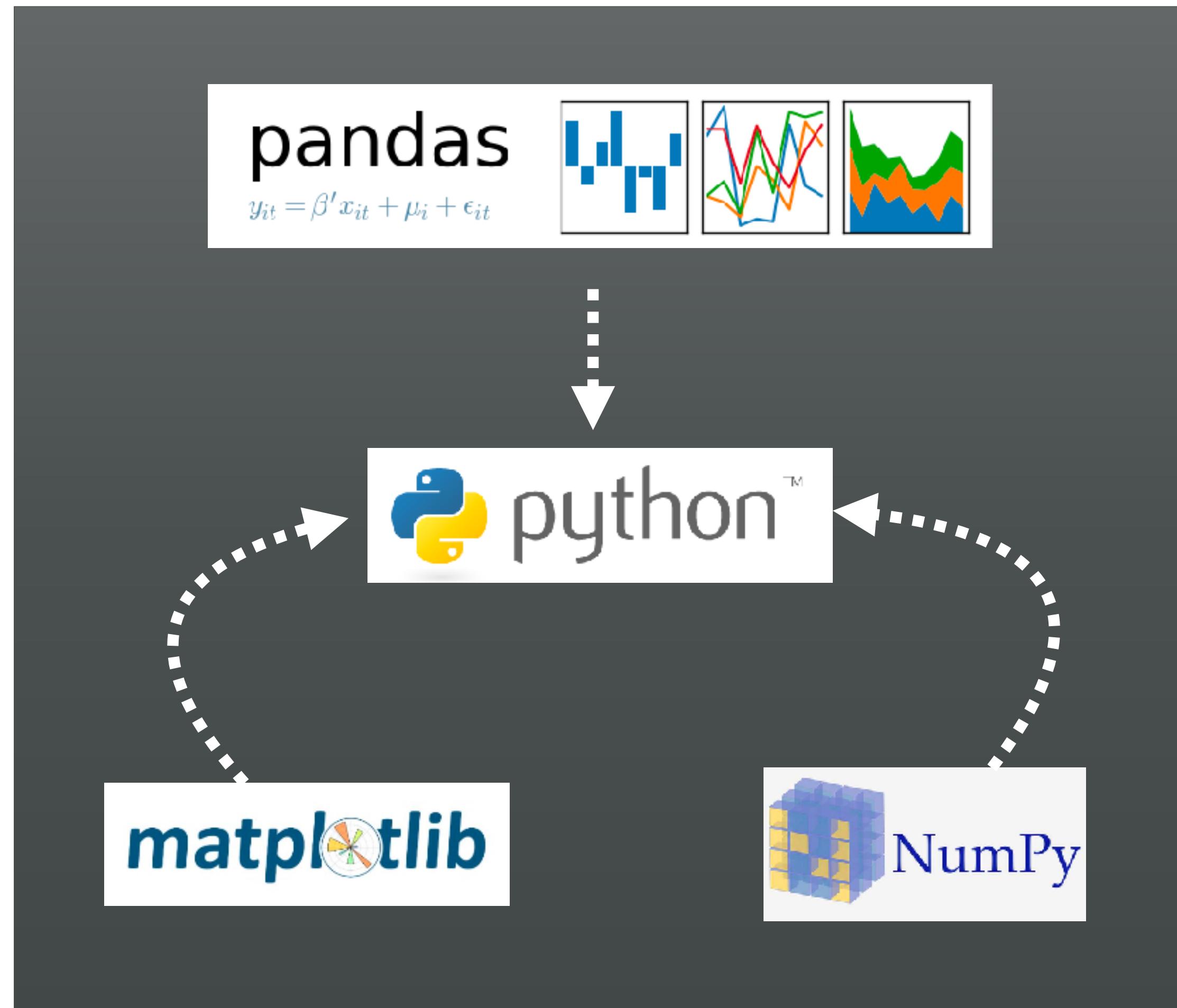
For the management and manipulation of arrays, mathematical operations, and support for Pandas.

## Pandas (Python library)

For the importation of data from CSV files, as well as its conversion to DataFrames.

## Matplotlib (Python library)

For functions relating to data visualization.



01

02

03

# DataFrame Management

## **pd.read\_csv()**

CSV data from Zillow is imported and read to the kernel in Jupyter Notebook with the Pandas library.

## **df.drop()**

Non-essential and proprietary data from columns including RegionID and SizeRank are excluded from the original data set.

## **df.fillna()**

All not-a-number (NaN) values in the data set are backfill-substituted to project 0% change in home value data.

```
filepath = 'https://raw.githubusercontent.com/vitoperez117/Housing-Data/master/house_data.csv'
house_data_original = pd.read_csv(filepath, delimiter=',')
house_data = house_data_original.fillna(method = "backfill", limit = 200, axis=0)
house_data.head()
```

	RegionID	RegionName	City	State	Metro	CountyName	SizeRank
0	274772	Northeast Dallas	Dallas	TX	Dallas-Fort Worth-Arlington	Dallas County	1
1	192689	Paradise	Las Vegas	NV	Las Vegas-Henderson-Paradise	Clark County	2
2	118208	South Los Angeles	Los Angeles	CA	Los Angeles-Long Beach-Anaheim	Los Angeles County	3
3	270957	Upper East Side	New York	NY	New York-Newark-Jersey City	New York County	4
4	192820	Sunrise Manor	Las Vegas	NV	Las Vegas-Henderson-Paradise	Clark County	5

5 rows × 289 columns

01

02

03

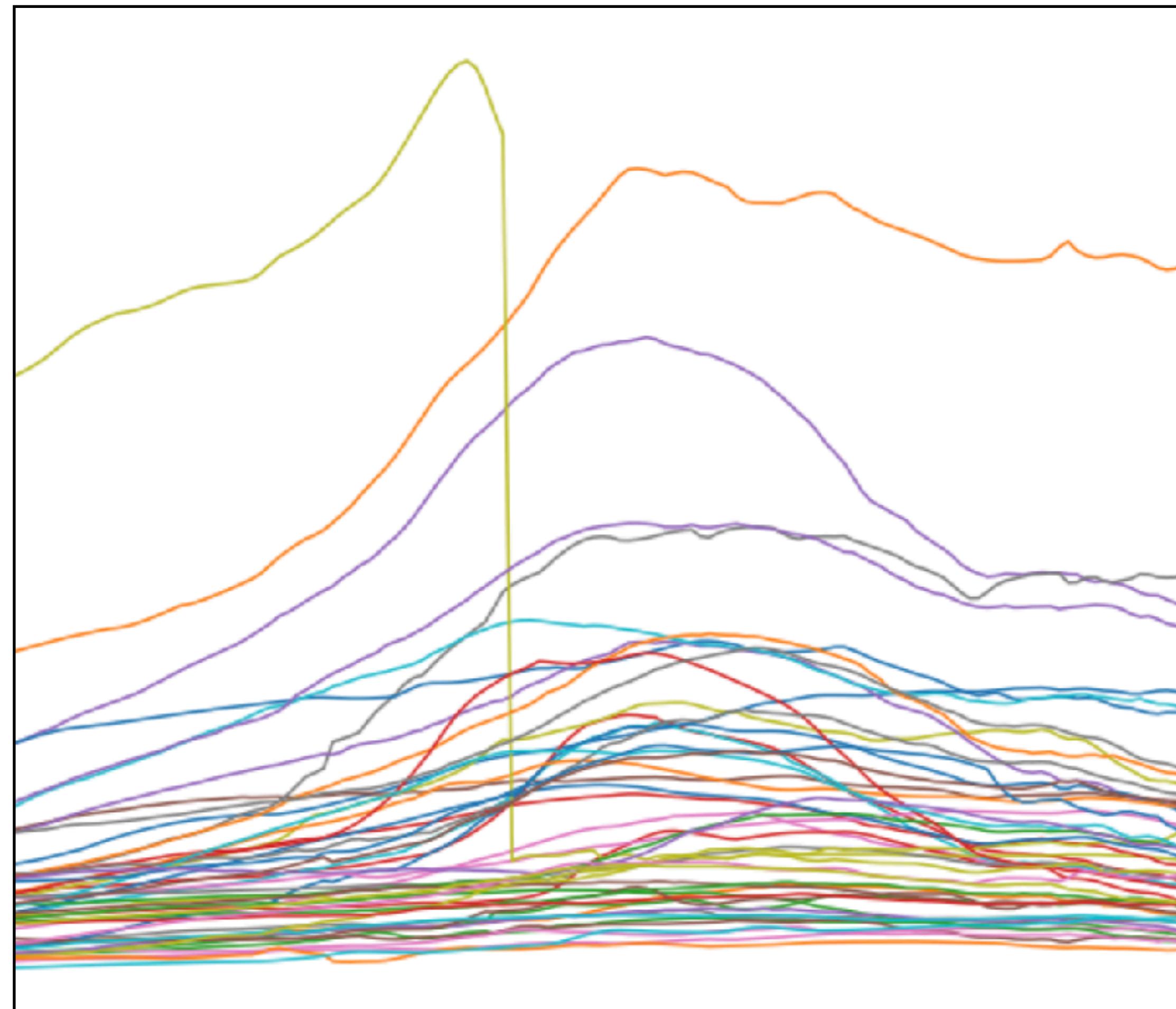
# Understanding the Data

## **df.describe()**

Returns a summary report of the now clean data frame.

Summary:

- 7275 U.S. neighborhoods
- 284 monthly periods (23 years)
- Property dollar ranges from \$10K - \$13M



01

02

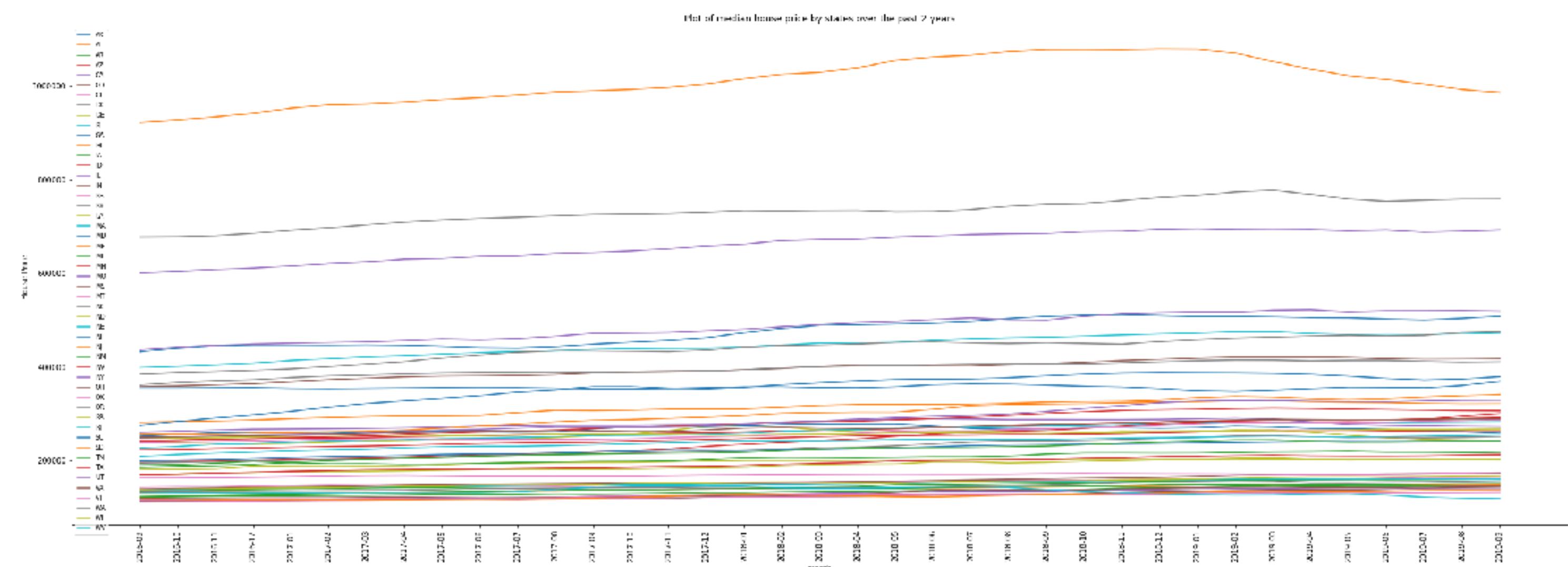
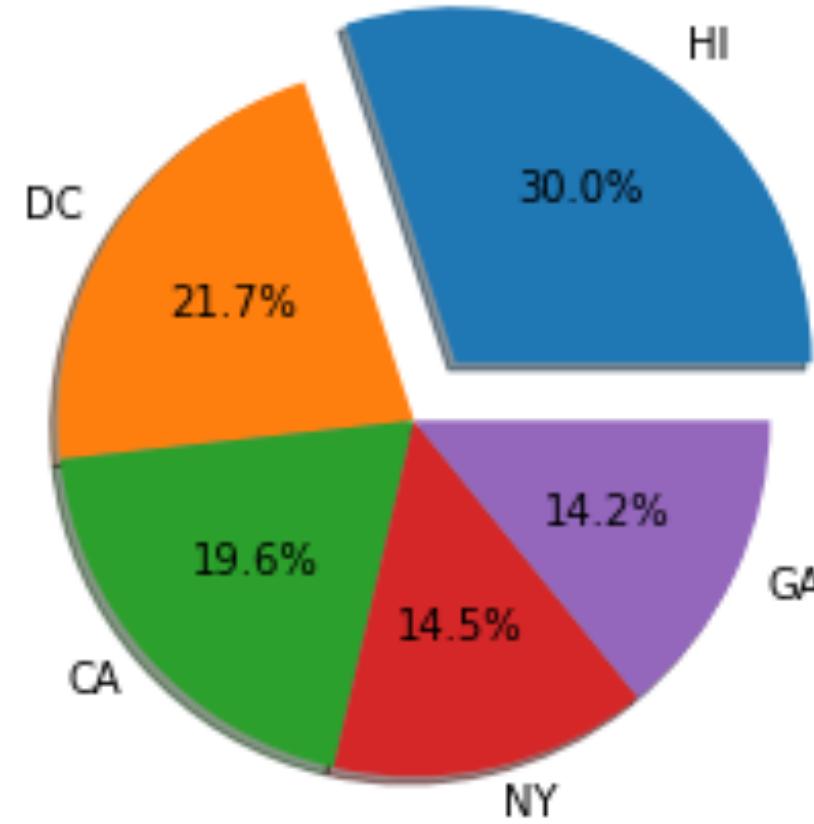
**03**

## Data Cleanup & Exploration

# Understanding the Data

The slope of price variations in the recent 3 years is not as steep comparing to that of recent 23 years.

Average of Median Price Pie Chart in last 3 years



**Does a higher price equal to a better investment?**

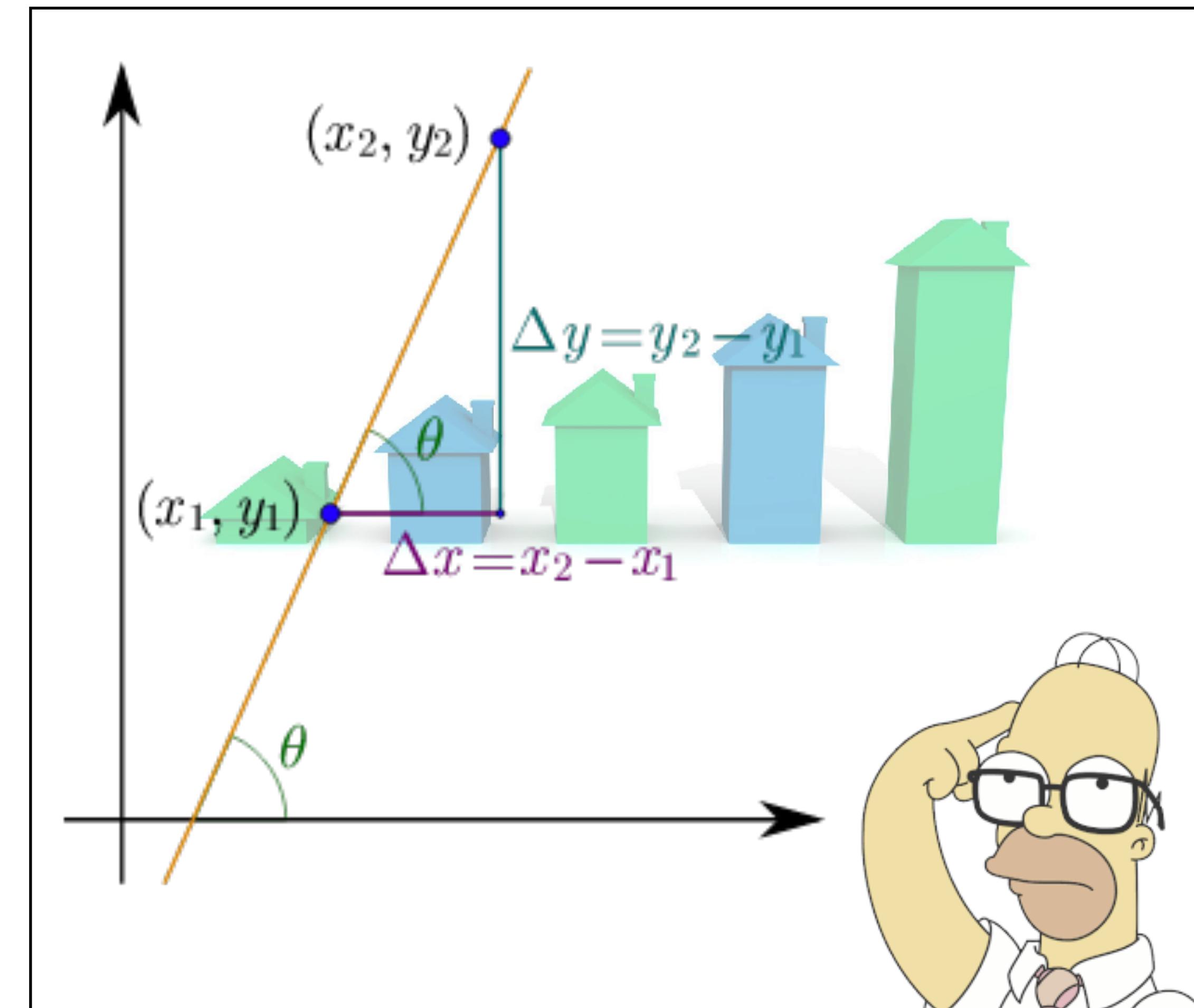


# Data Analysis

# Calculating Rate of Change

The appreciation (or depreciation) of median home values in our data will be determined and compared by calculating the rate of change between our chosen standard for long-term (10 years to recent) and short-term (3 years to recent) periods of investment.

**Long-term = 10 years**  
**Short-term = 3 years**



# Slicing & Grouping by Location

## DataFrame Subset

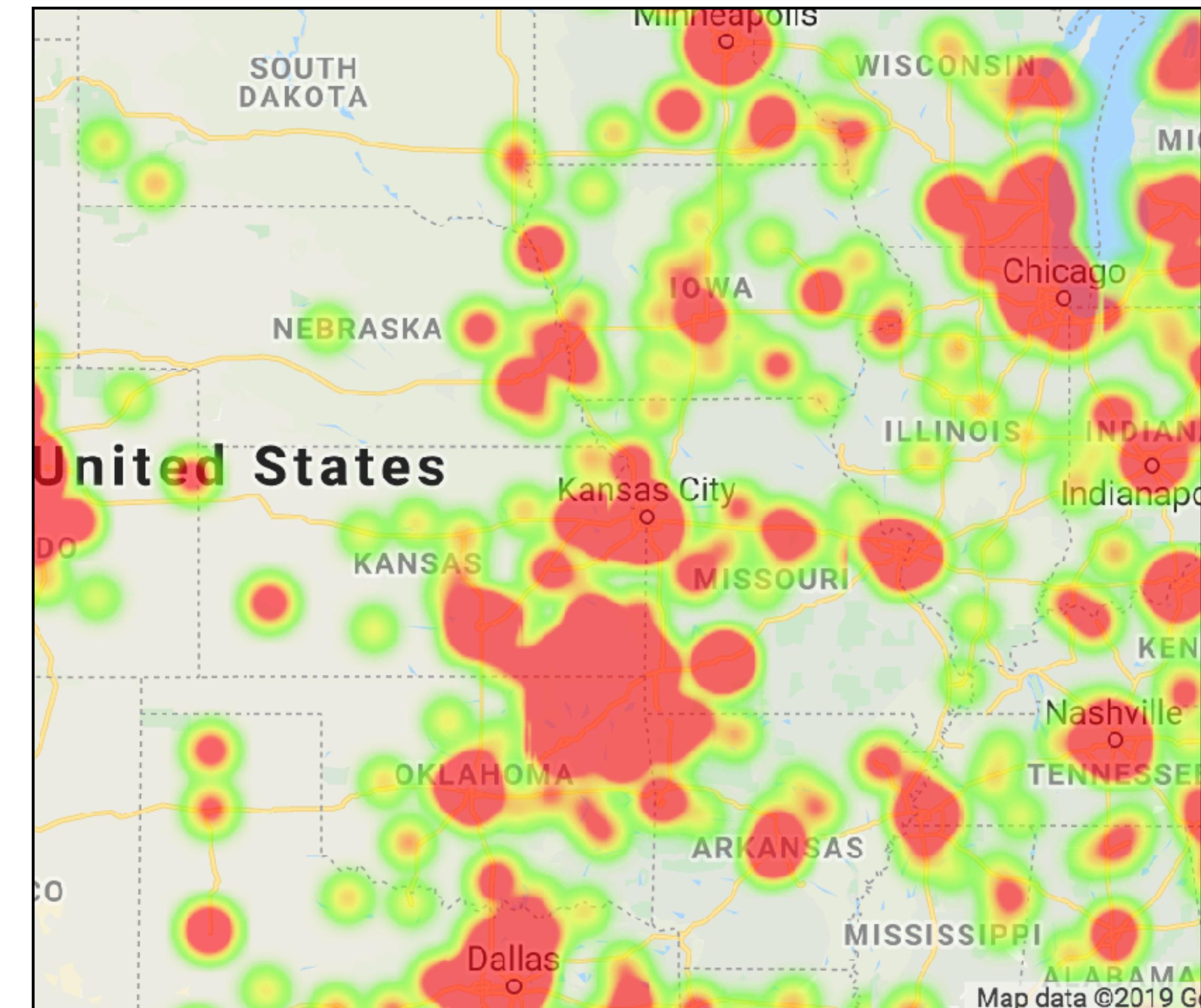
Subsetting the DataFrame streamlines the usable data by allowing for the selection of relevant columns and rows to analyze (10 years & 3 years).

## pd.groupby

Use of the groupby operation allows for median home values of neighborhoods to be combined for analysis at State and City levels.

## loc & iloc

Use of loc/iloc assists to select data index values for this project's use in data visualization.



01

02

03

# Binning into Brackets

## **pd.cut()**

This Python method allows for the sorting of median home values into custom categories for analysis and recommendations at several levels of income.

### **Bin Categories (MASH):**

“Shack” = Under \$500,000

“Apartment” = \$500,000 - \$1 Million

“Home” = \$1 Million - \$2 Million

“Mansion” = Over \$2 Million



01

02

**03**



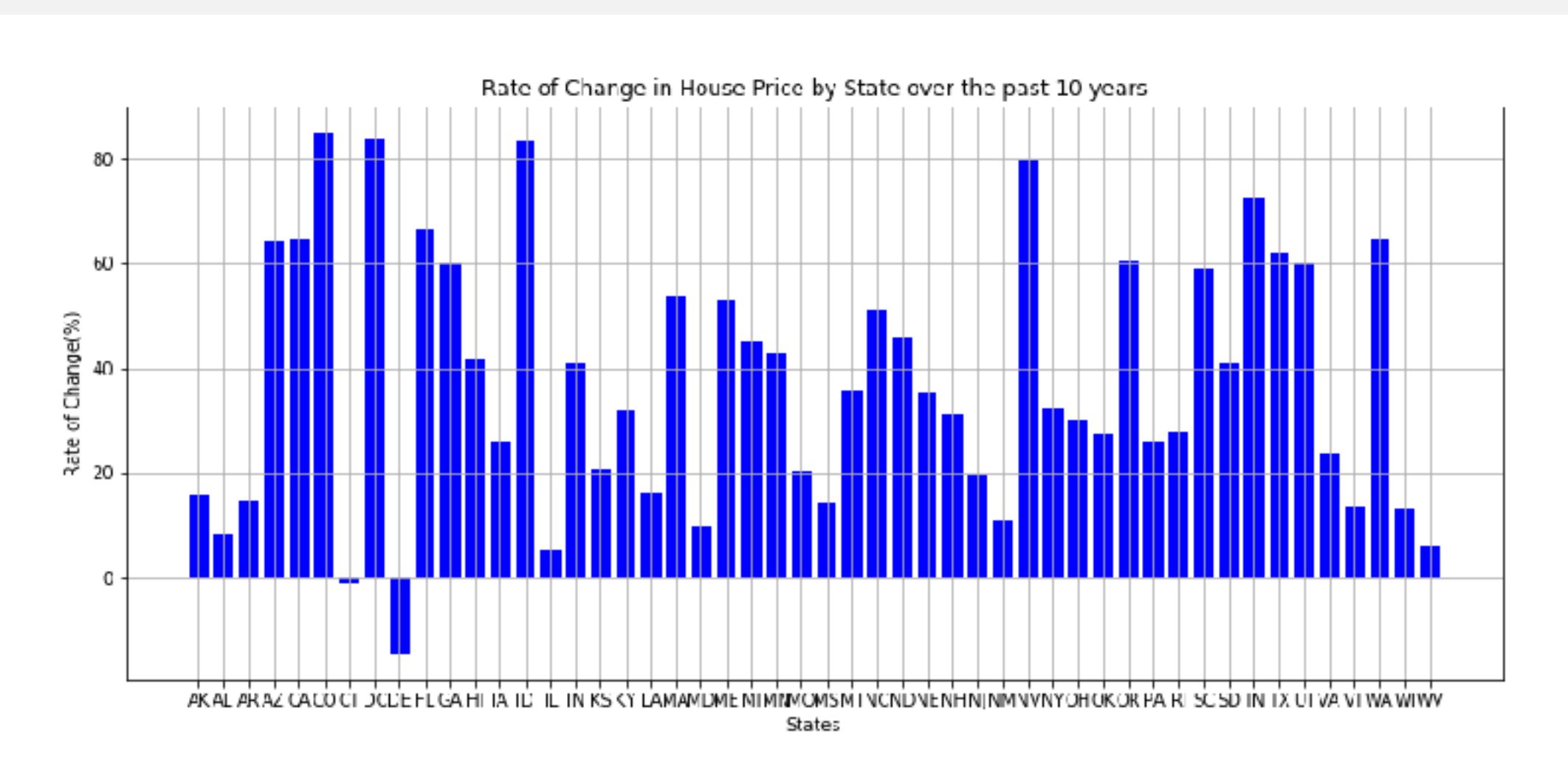
# **Findings & Discussion**

## Findings & Discussion

# Overview: 50 States in the U.S.

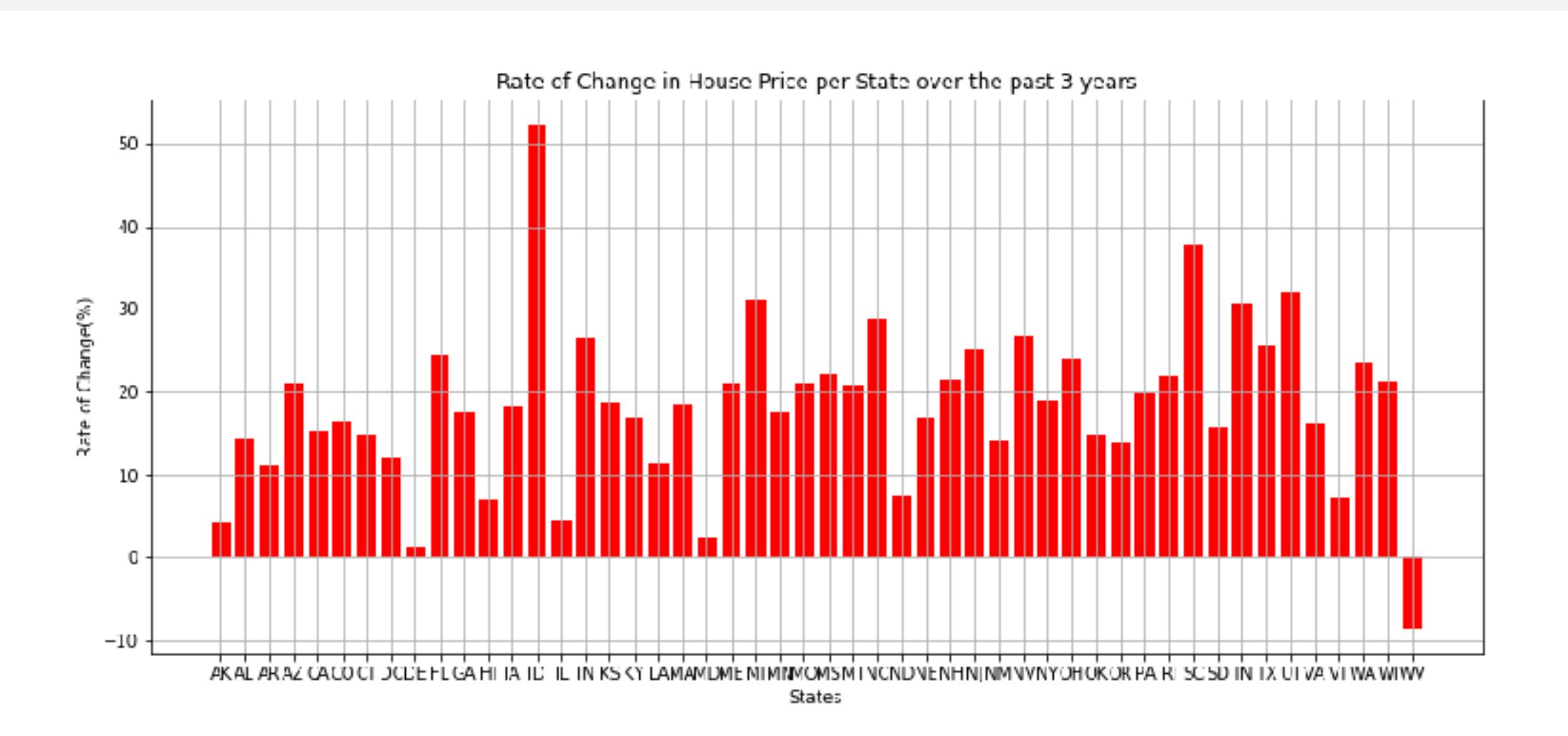
### LONG-TERM (10 YEARS)

The mean of rate of change across the 50 states shows a **54% return** over 10 years, though some states clearly fall well above and below average.



### SHORT-TERM (3 YEARS)

Investing in U.S. property in the short-term has shown a **return of 20%** with more normalized results.

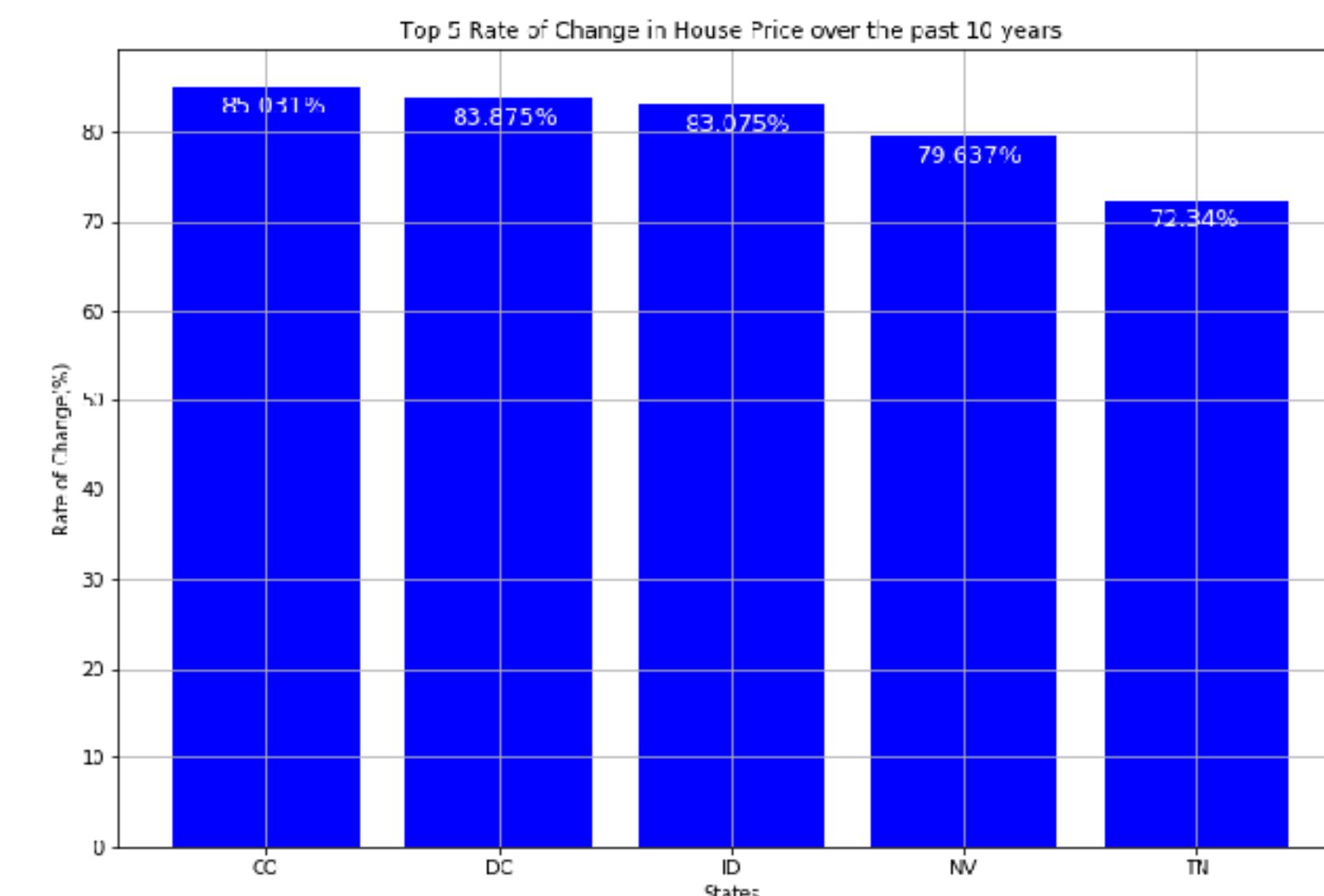


## Findings & Discussion

# Top 5 States in the U.S.

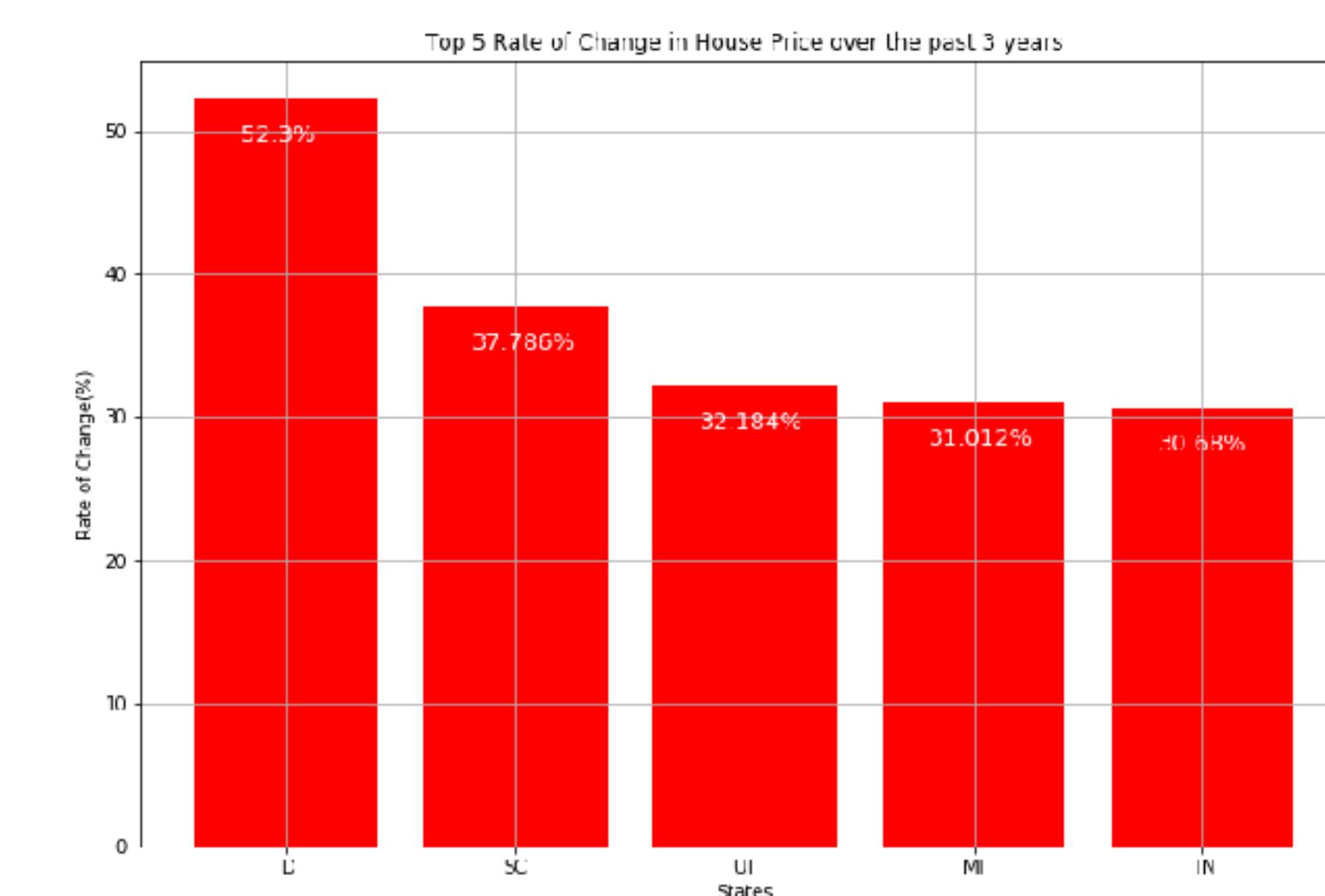
### LONG-TERM (10 YEARS)

Purchasing a home for long-term investment in any of the top 5 states averaged at least a **70% return**, based on median home value.



### SHORT-TERM (3 YEARS)

Buying short-term in Idaho averaged a runaway **+50% increase** based on median home value.



## Findings & Discussion

# Top 5 Cities (in Top States)

Top 5 cities across Idaho, South Carolina, Utah, Michigan, and Tennessee.

Rate of change for top cities range from 10% to 60%

Current property values across top cities 80k and 700k

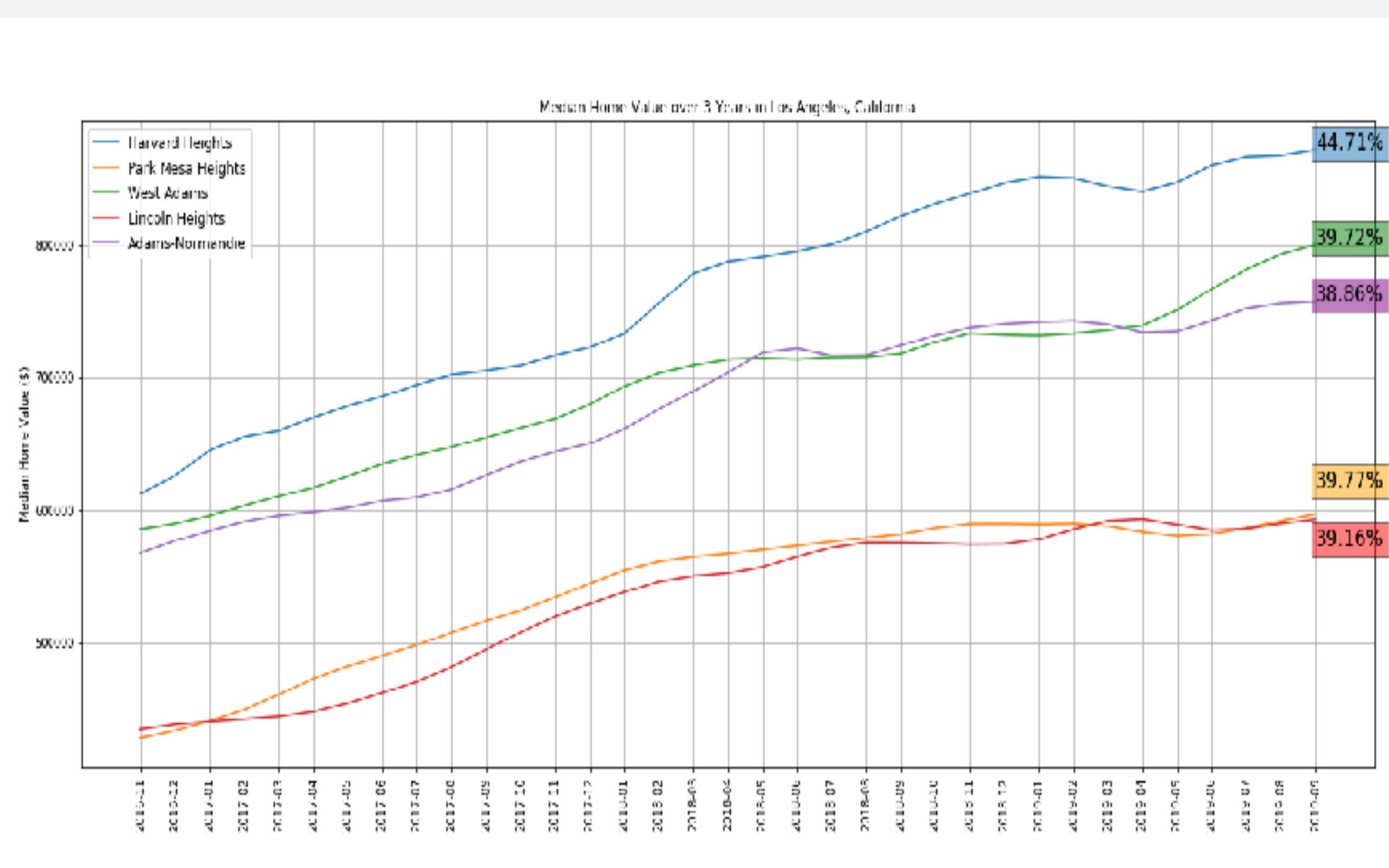
<b>Nampa</b>	55.897010
<b>Boise</b>	45.732689
<b>Spartanburg</b>	34.277559
<b>Beaufort</b>	22.798354
<b>Charleston</b>	20.836354
<b>Columbia</b>	17.806268
<b>Hilton Head Island</b>	10.428469
<b>Ogden</b>	53.187500
<b>Salt Lake City</b>	41.735426
<b>Logan</b>	37.171725
<b>Provo</b>	33.066132
<b>Orem</b>	32.317236
<b>Detroit</b>	55.601613
<b>Flint</b>	54.615385
<b>Redford</b>	52.485960
<b>Kalamazoo</b>	36.438923
<b>Taylor</b>	34.169550
<b>Knoxville</b>	62.091463
<b>Memphis</b>	37.192982
<b>Chattanooga</b>	33.640868
<b>Johnson City</b>	26.474369
<b>Bartlett</b>	22.509506

## Findings & Discussion

# Neighborhoods: Los Angeles vs. Boise

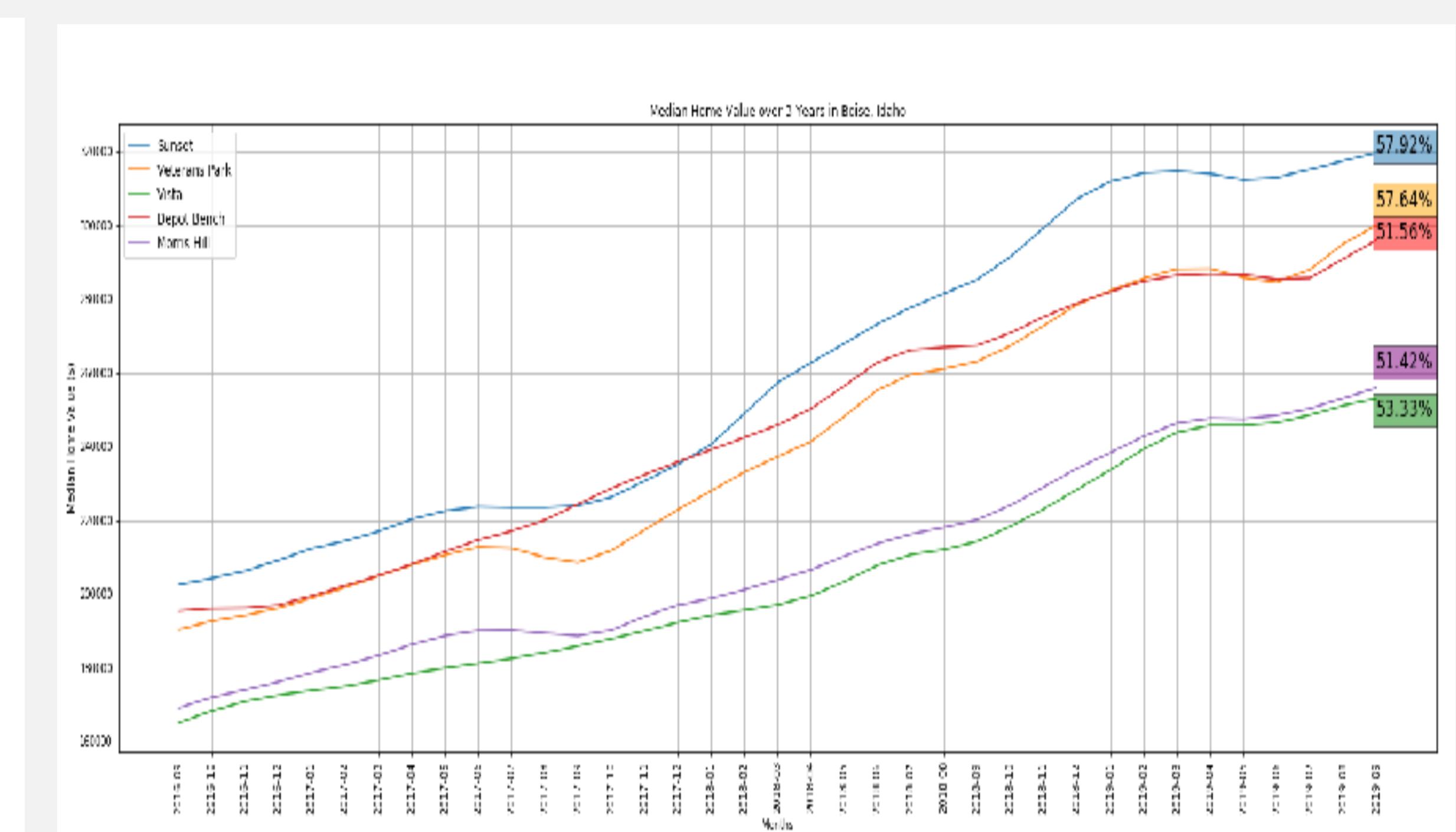
### LOS ANGELES, CA (3 YEARS)

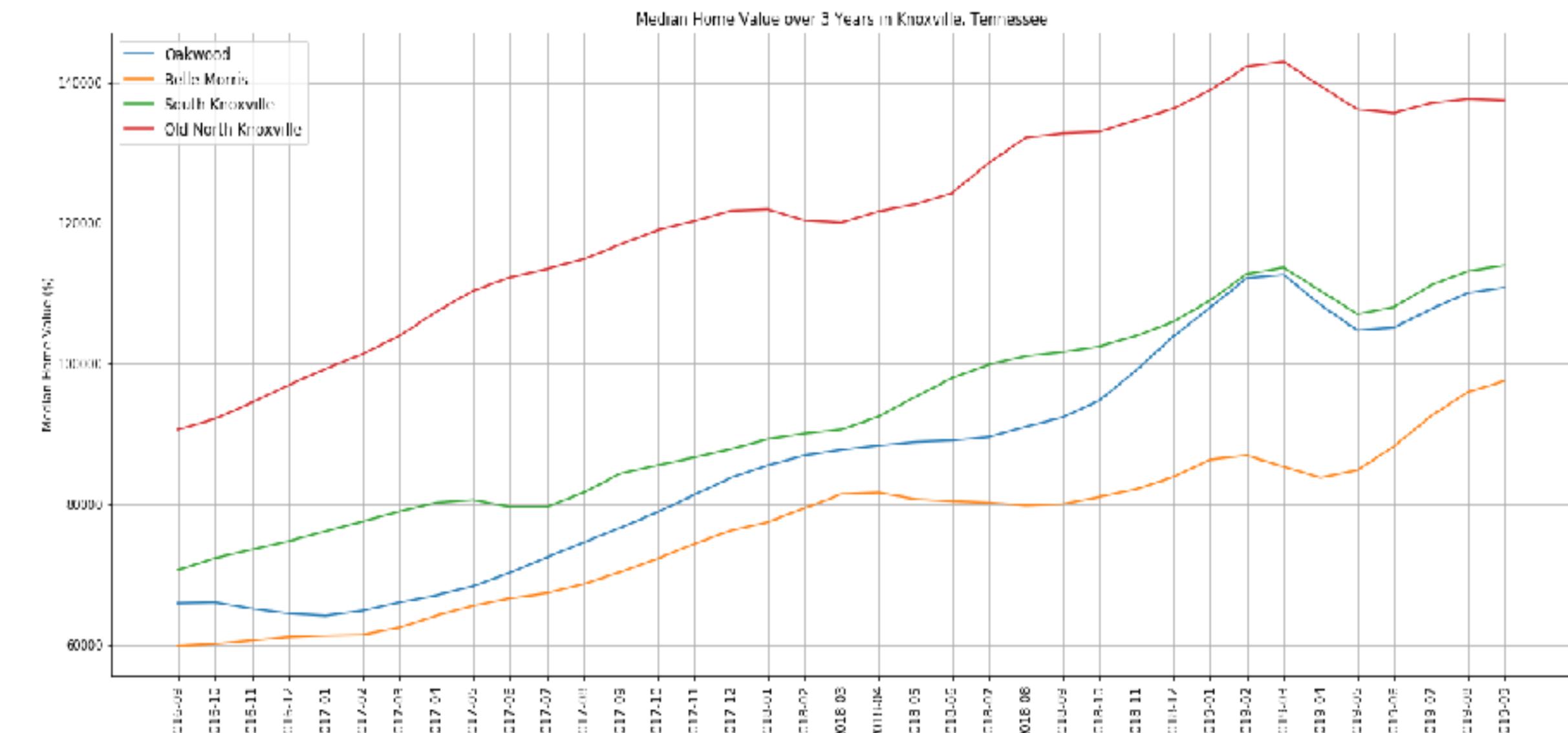
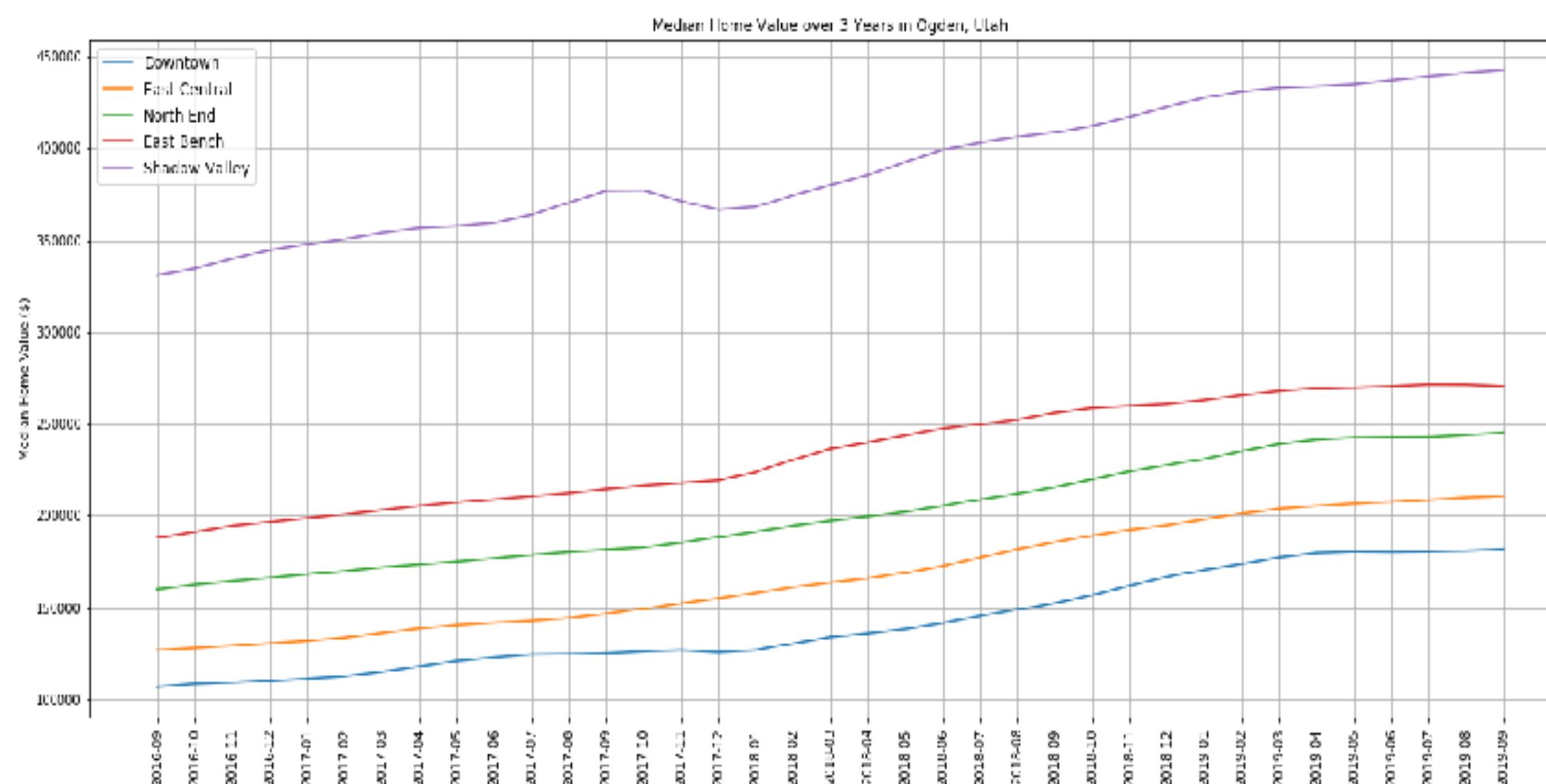
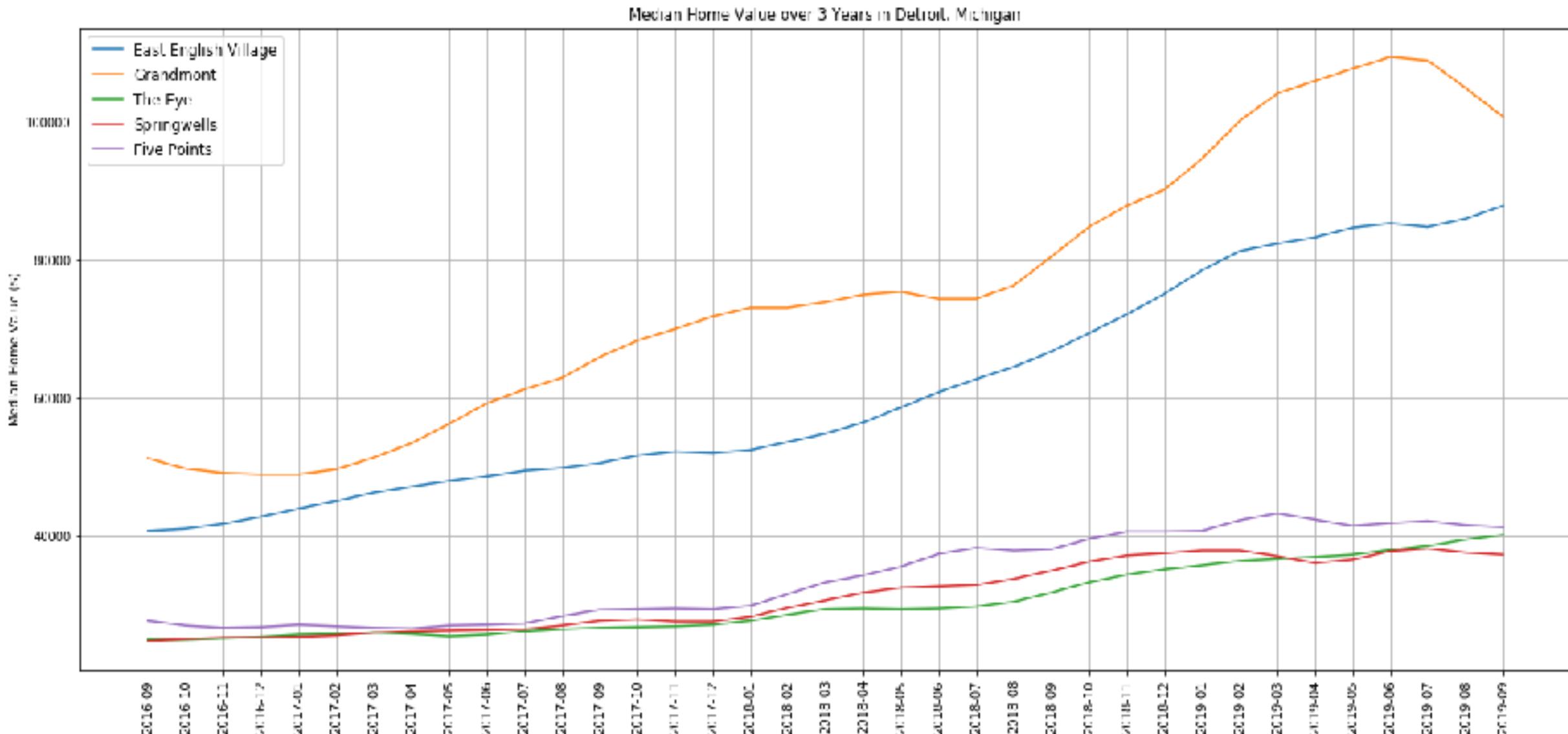
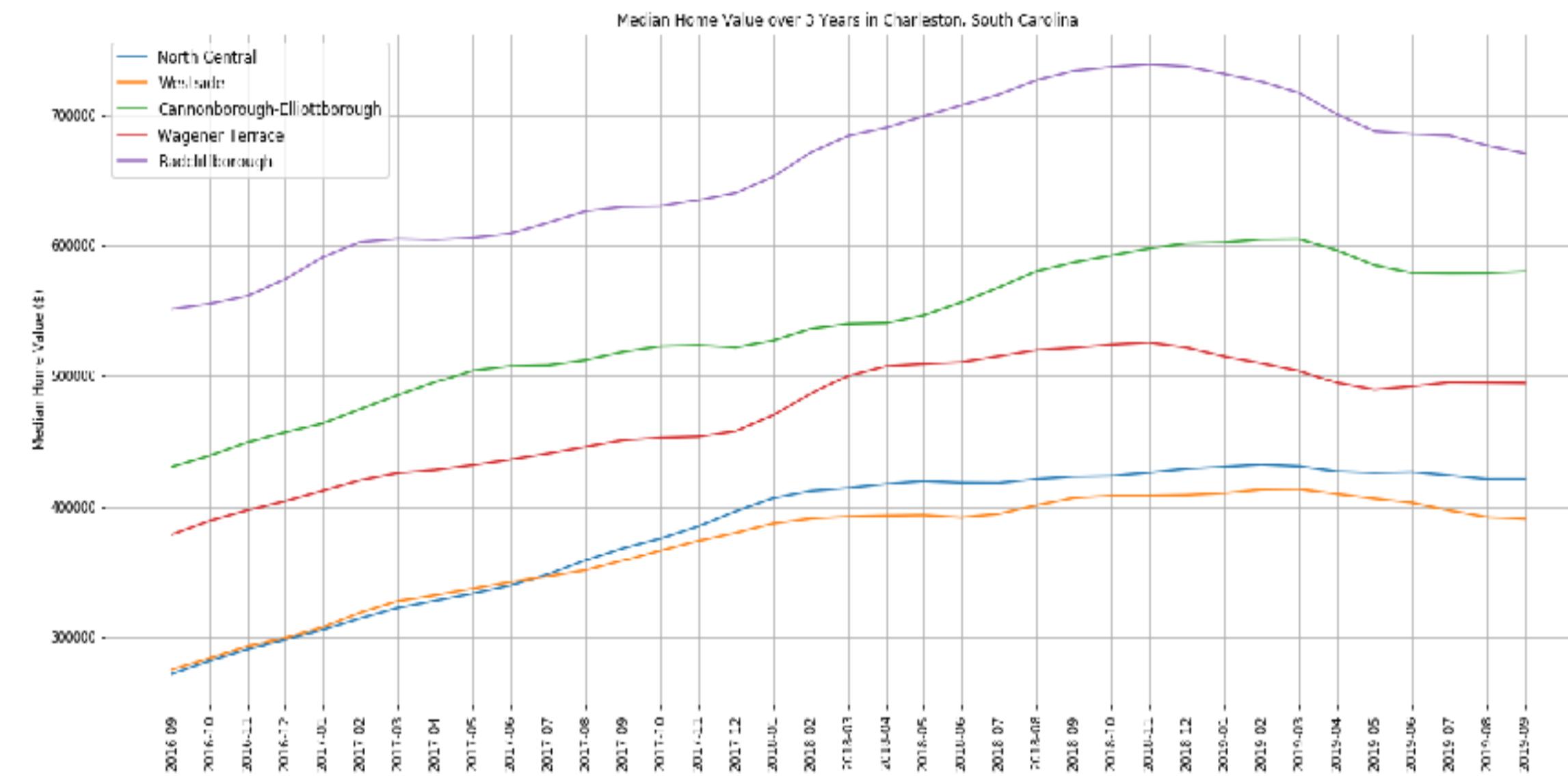
Los Angeles neighborhoods have rates of change that are wider spread out around 40%. Principal required to invest in Los Angeles: \$600k - \$900k



### BOISE, ID (3 YEARS)

Neighborhoods in Boise, ID have greater rate of change than those in L.A. Change in Boise neighborhoods cluster closer near 50%. Principal required to invest in Boise: \$240k - \$320k





## Findings & Discussion

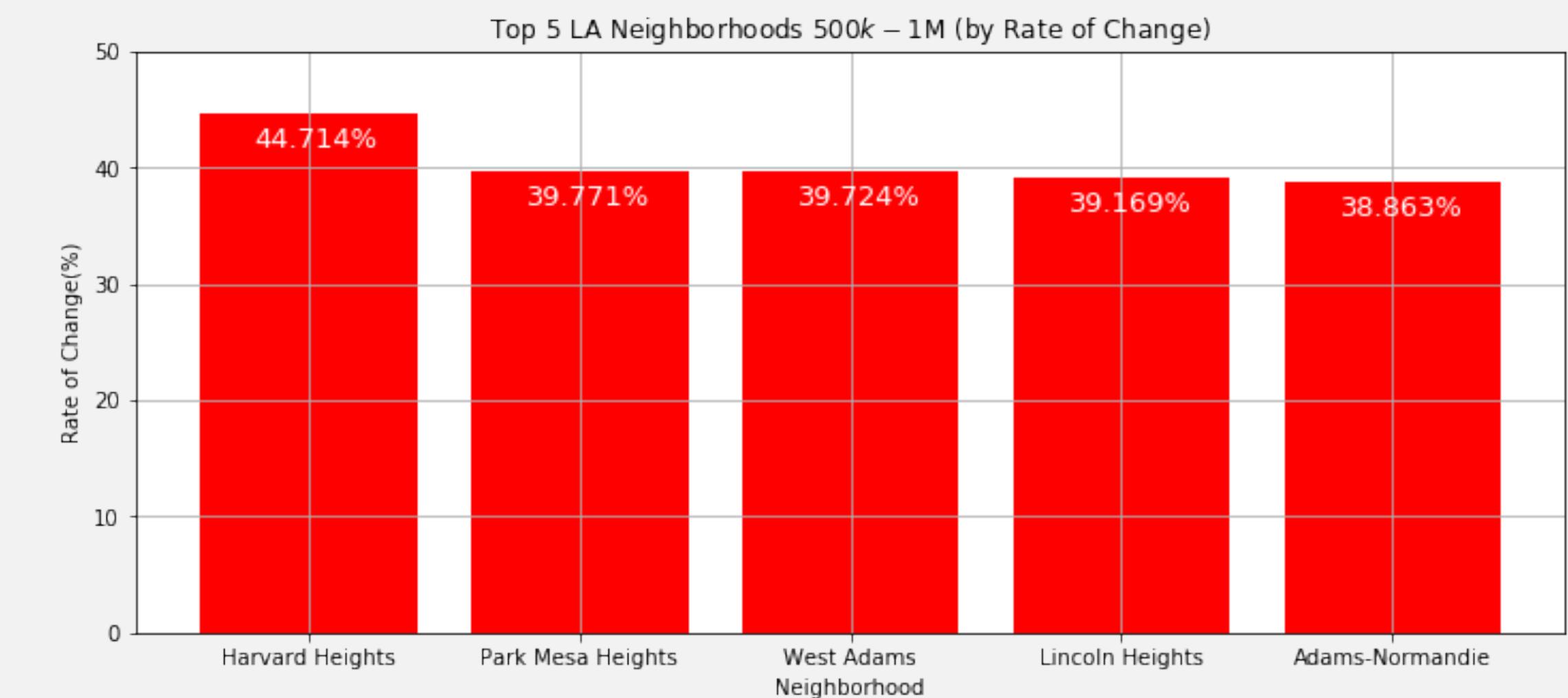
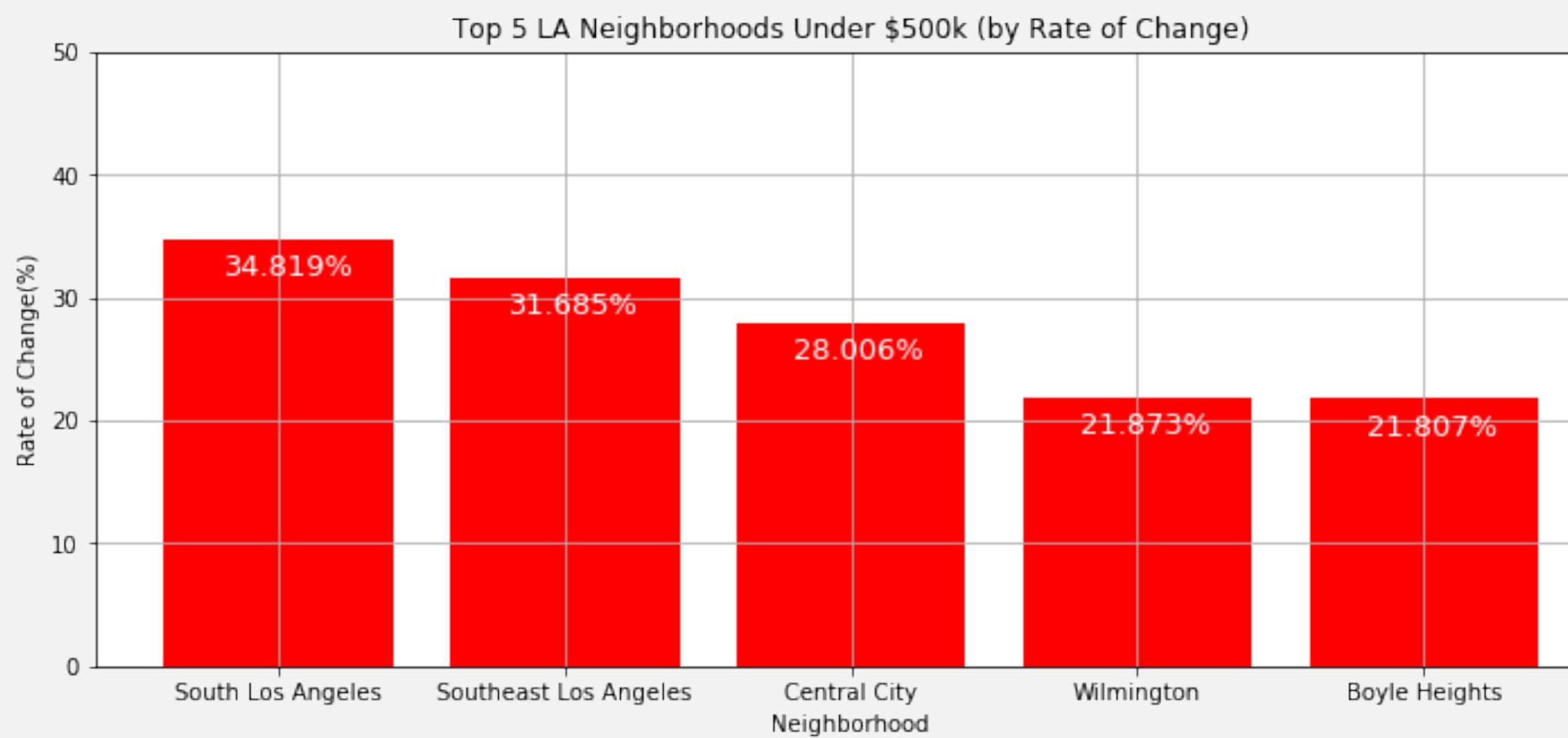
# Top 5 Places in L.A. (by Price Point)

### \$500,000 AND UNDER

The fastest growers in the sub-\$500k range has returned at least 20% of its home value in three years.

### \$500K - \$1 MILLION

Buying a Los Angeles home in the sub-\$1 million range appears likely to make the largest return on investment.



## Findings & Discussion

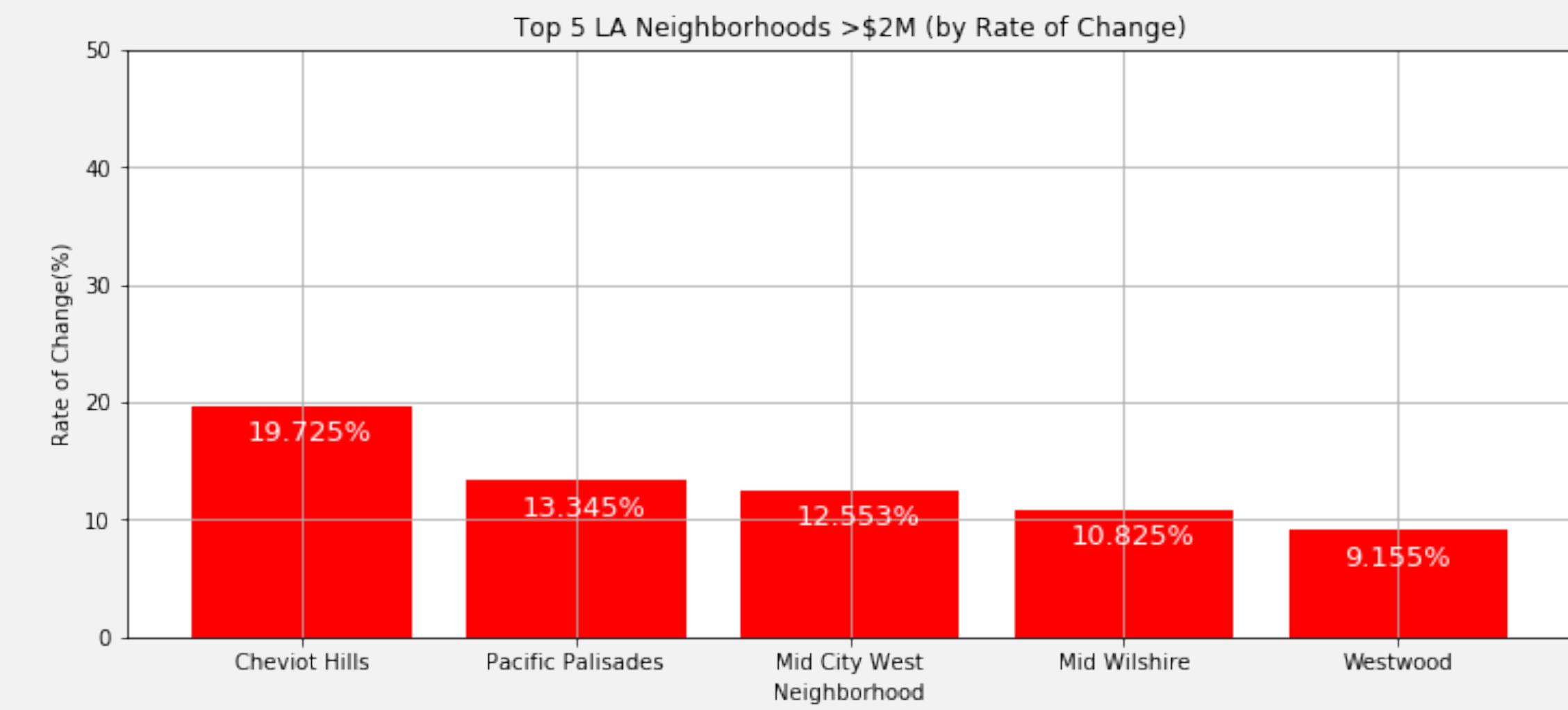
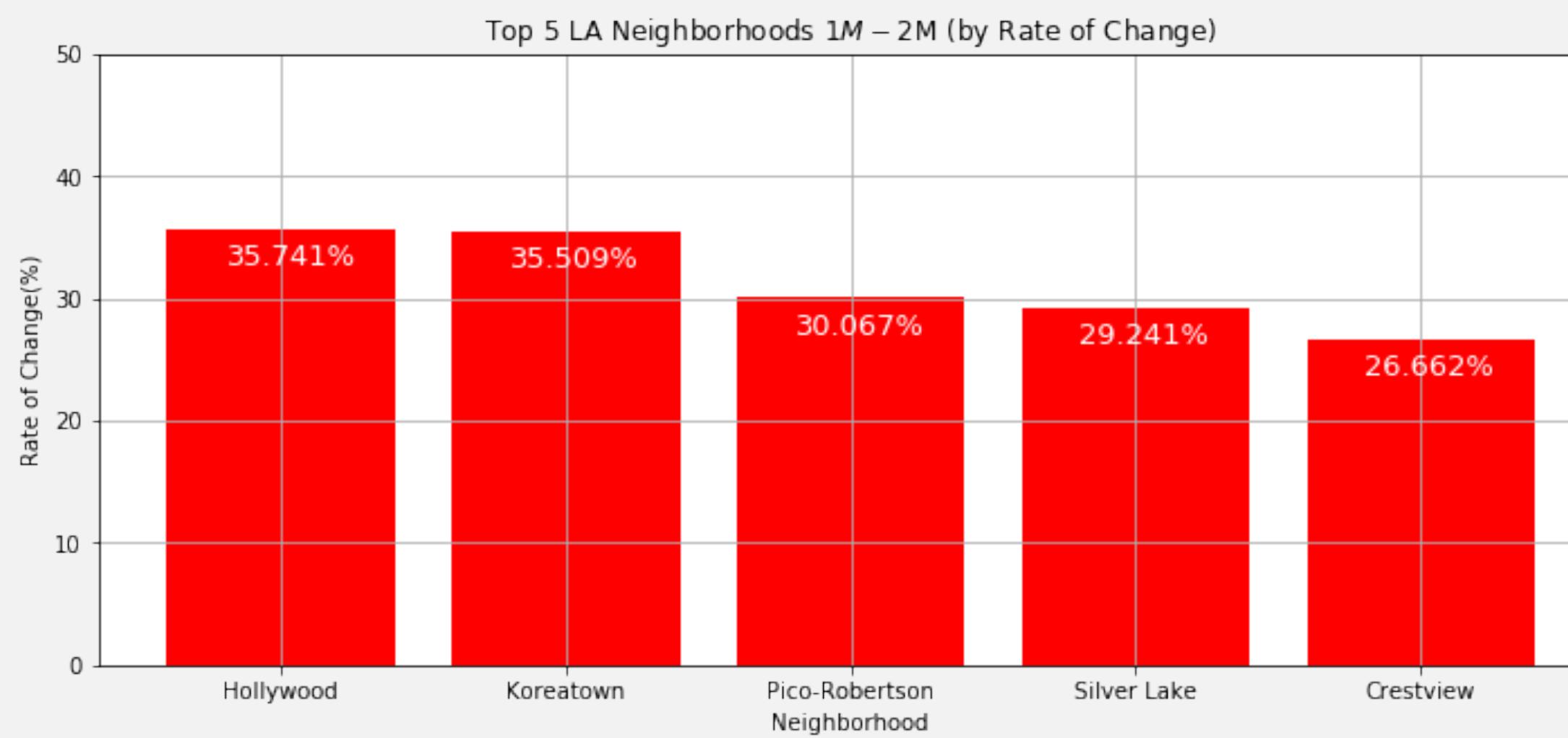
# Top 5 Places in L.A. (by Price Point)

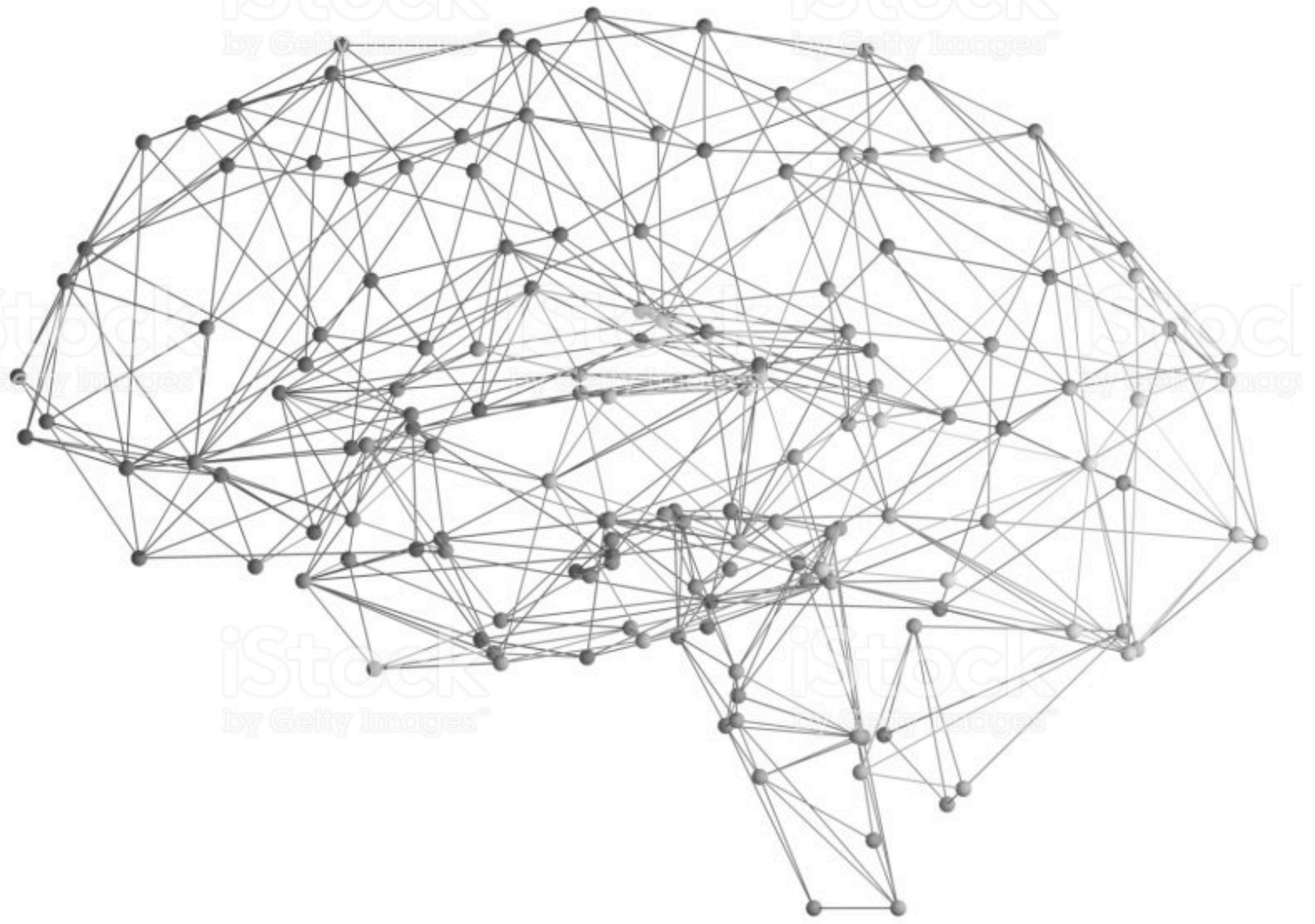
### \$1 MILLION - \$2 MILLION

Houses between \$1M and \$2M have not seen growth higher than 36% in the past three years.

### OVER \$2 MILLION

Houses over \$2 million in value has shown the lowest return on investment of the four bins created for this project.





**Post-Mortem**

# A few final points...

## CHALLENGES

1. *The data is too simple for usable action.* Expectations needed to be set that any conclusions to our findings would be the tip of the iceberg for evaluating home prices.

2. **Ensuring findings are relevant to the audience.** Research is key to determining market demand and the types of data presented are more than just interesting but also useful.

## TAKEAWAYS

1. Invest in a single-family home with the cheapest current value and the highest short term rate of change, you should pay attention to ID, SC, UT, MI, and TN..
2. Starting your search wider and drilling in ensures a higher likelihood of finding high-return properties.
3. There are neighborhoods that have similarly high short-term rates of change outside these top 5 states. However, they're likely to be infrequent or anomalous within their city or state.

## IF WE HAD MORE TIME...

1. Conduct both linear and multi variable regression to investigate more on the factors that could affect home values.
2. Modeling a prediction formula to apply to data visualization.
3. Including input functions to perform custom searches by year, price point, and location.
4. Taking in to consideration of the market value of the houses.



# QUESTIONS?

# THANK YOU!

from Vito Perez, Gerard Tieng, & Peirang Xu

*View the Code on Github:*  
<https://bit.ly/34dRxyS>