## Note to the reader:

I, Emilio Vasquez, orchestrated and directed this project during my intensive time series class over the summer. With my commitment to excellence, I was at the helm of every facet of this analysis and presentation, intricately weaving together the threads of responsibility delegation, the art of exploratory data analysis (EDA), and GARCH research and modeling. In addition, I spearheaded this presentation deck, ensuring that every slide upheld the high quality data analysis that we aim to deliver.

# GARCH and Zillow Housing Data

Presented by Emilio Vasquez, Paul (Nik) Lopez, Horace Tsai, Lucas Everett

6/28/2023

Cal State Fullerton

# Agenda

| Introduction/ EDA | GARCH Background | GARCH Housing Data Analysis/Results | Conclusion | Q & A |
|---|---|---|---|---|
| Emilio Vasquez | Paul (Nik) Lopez | Horace Tsai | Lucas Everett | All |

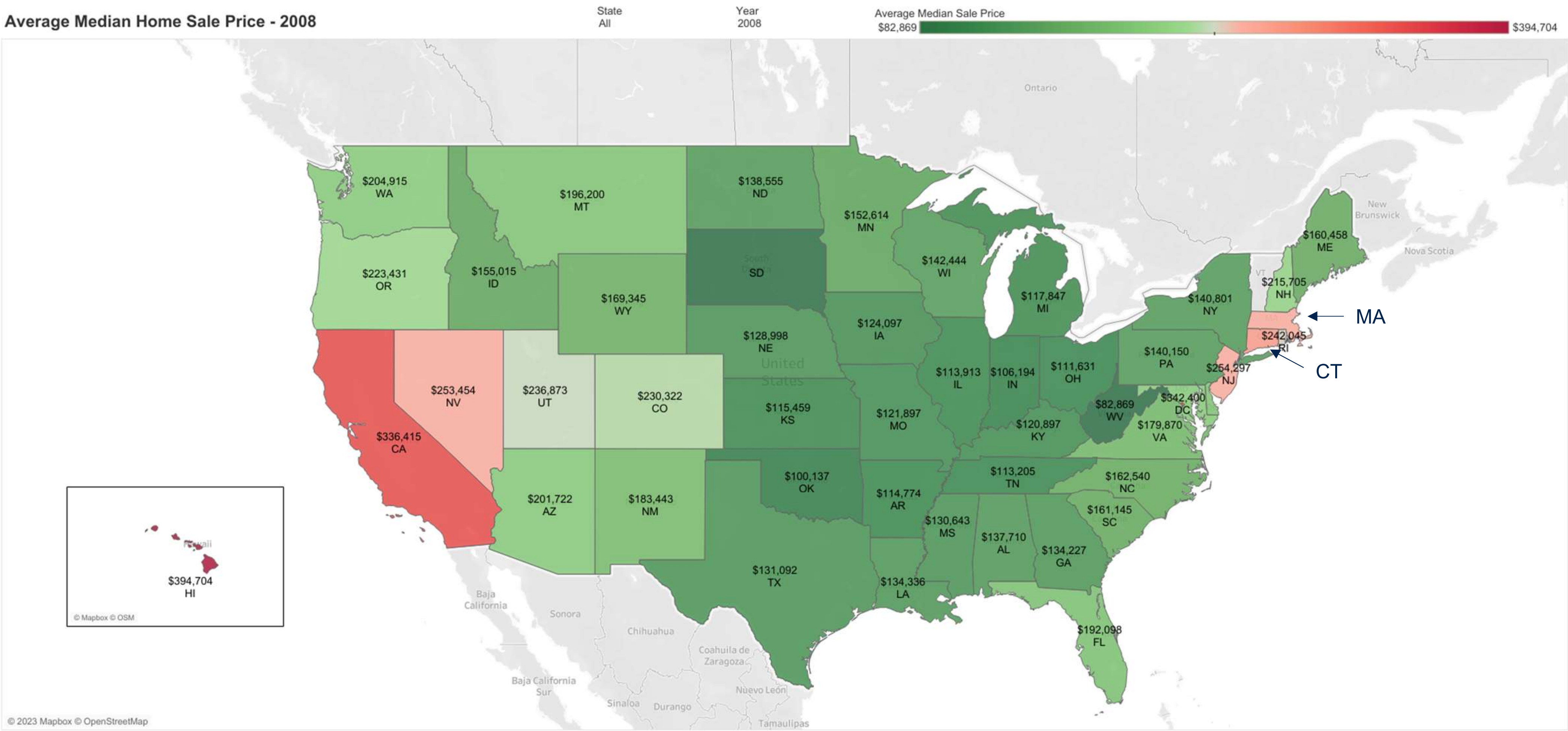
1

# Introduction/EDA

- Emilio Vasquez

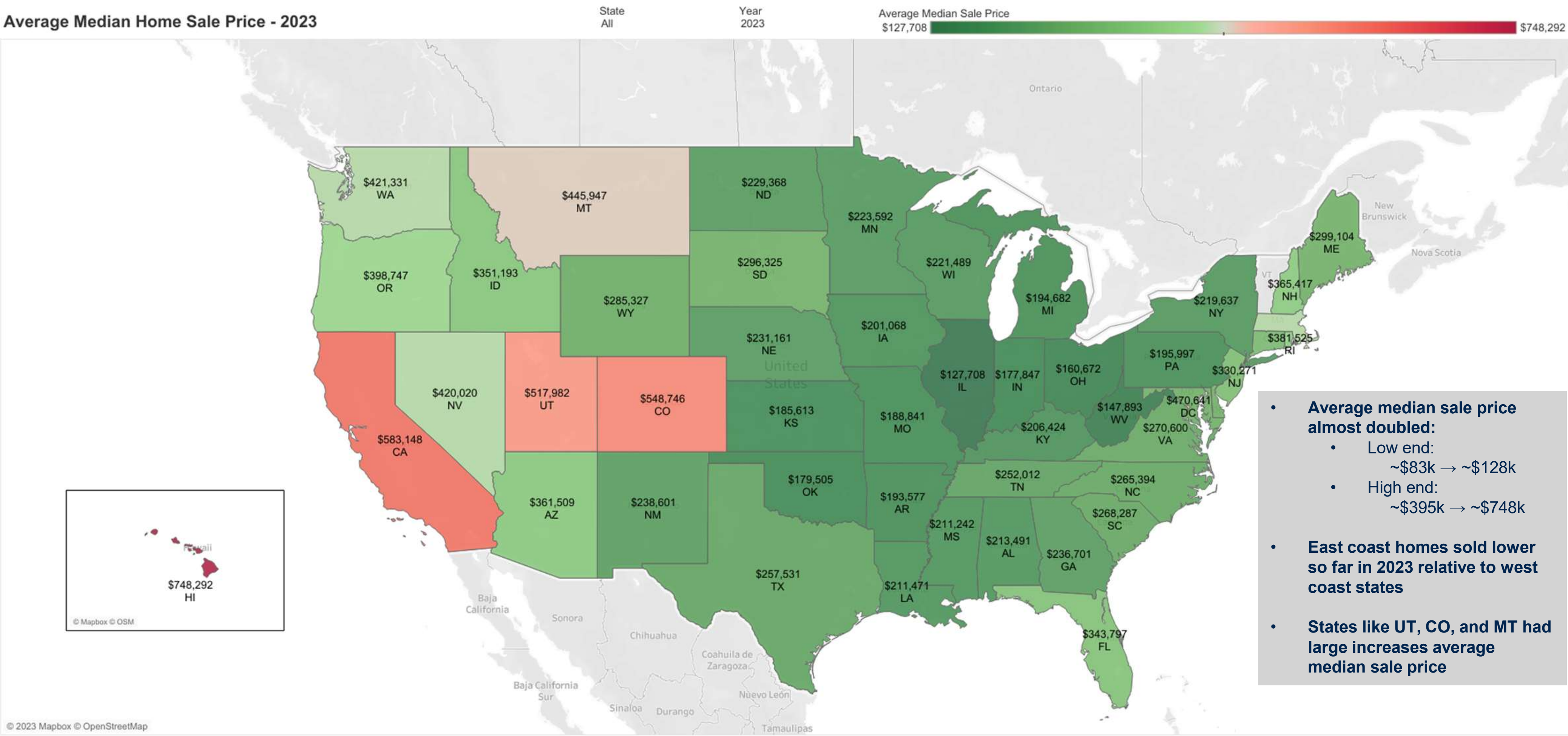
# Dataset: Median Housing Sale Price from Zillow

- Downloaded raw monthly median house sale prices from https://www.zillow.com/research/data/
- Spans from 2/29/2008 to 3/31/2023, giving us about 180 datapoints for each city
- Covers 608 different cities across the United States
- Potential bias in the data that could mispresent the true median for any given state




Dashboard link

Cal State Fullerton




- **Average median sale price almost doubled:**
  - Low end:
    ~$83k → ~$128k
  - High end:
    ~$395k → ~$748k
- **East coast homes sold lower so far in 2023 relative to west coast states**
- **States like UT, CO, and MT had large increases average median sale price**

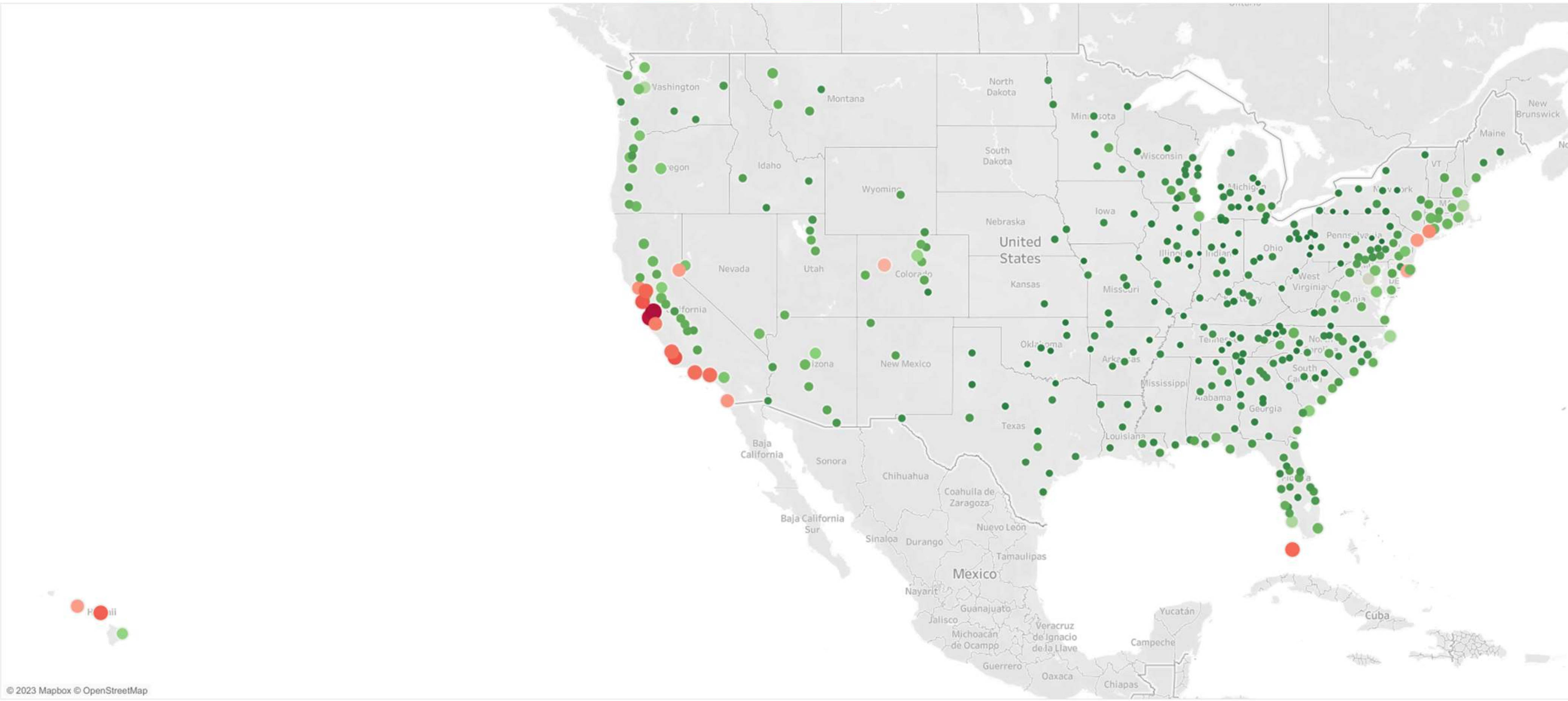
Dashboard link

**Average Median Sale Price - All - 2008**

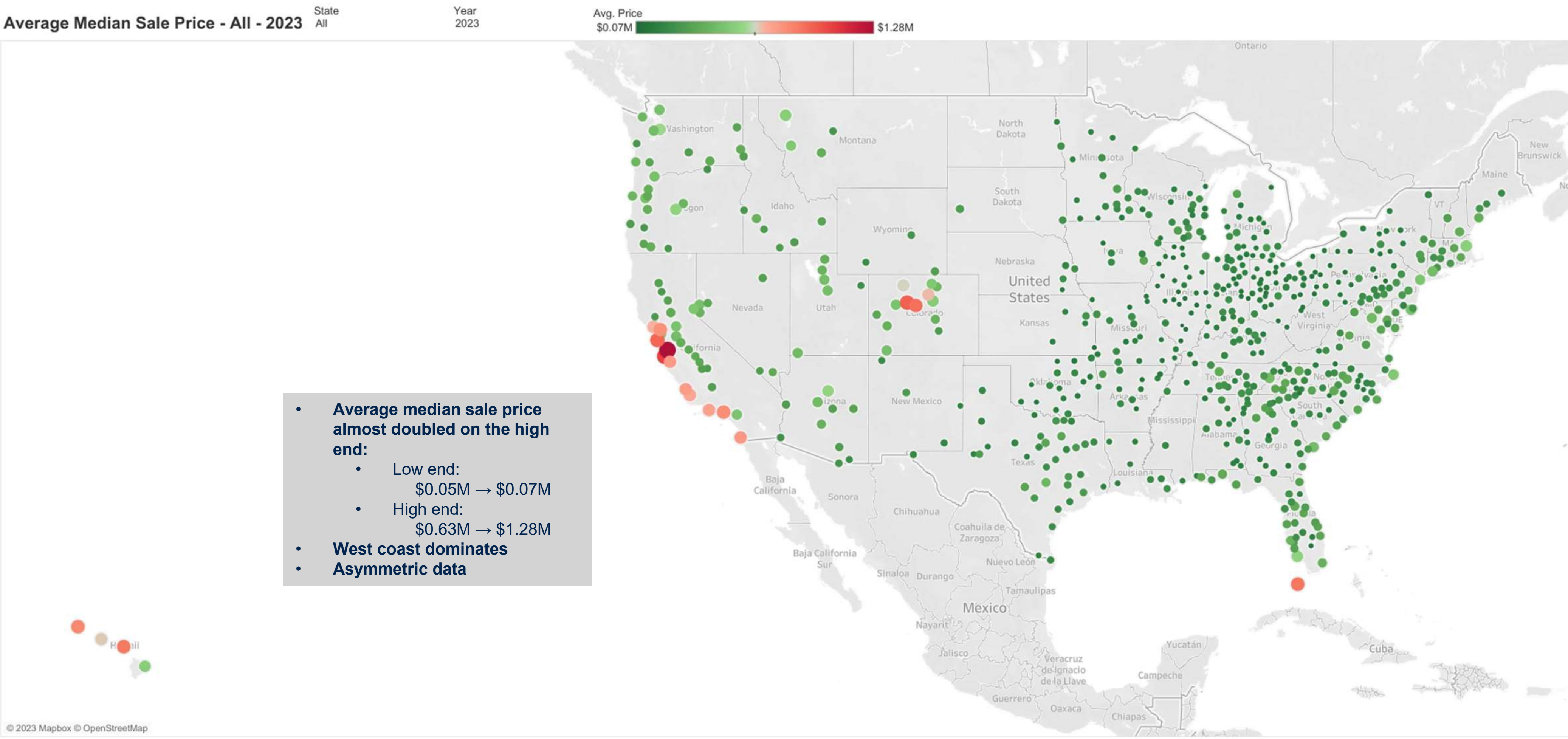
State
All

Year
2008

Avg. Price
$0.05M $0.63M




Dashboard link

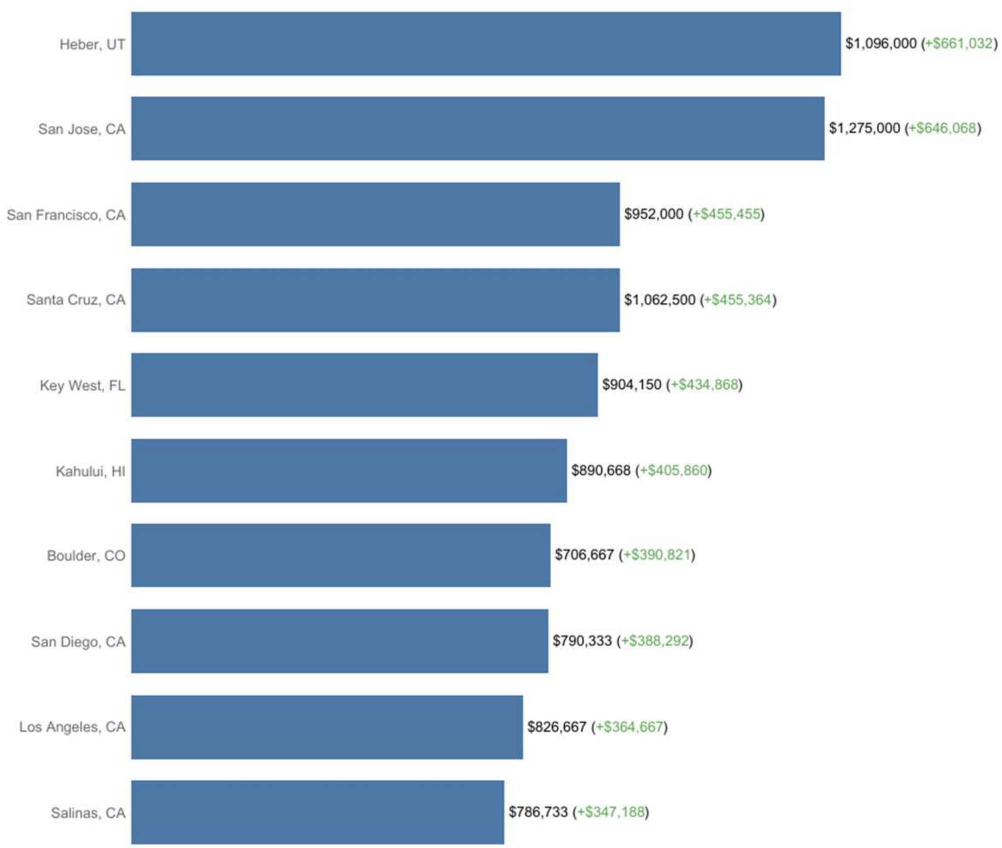



- **Average median sale price almost doubled on the high end:**
  - Low end:
    $0.05M → $0.07M
  - High end:
    $0.63M → $1.28M
- **West coast dominates**
- **Asymmetric data**

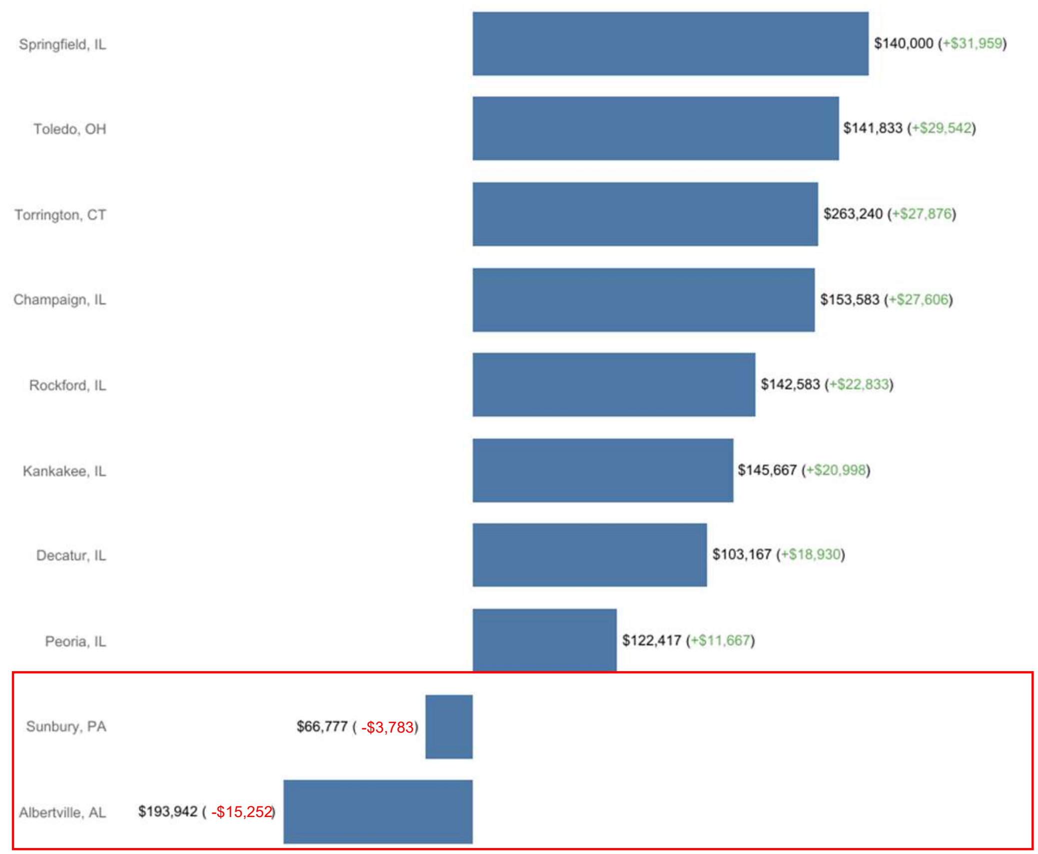
Dashboard link

## Winners and Losers Since 2008 (Sorted on Increase, by Region):

### Top 10 Largest Increases

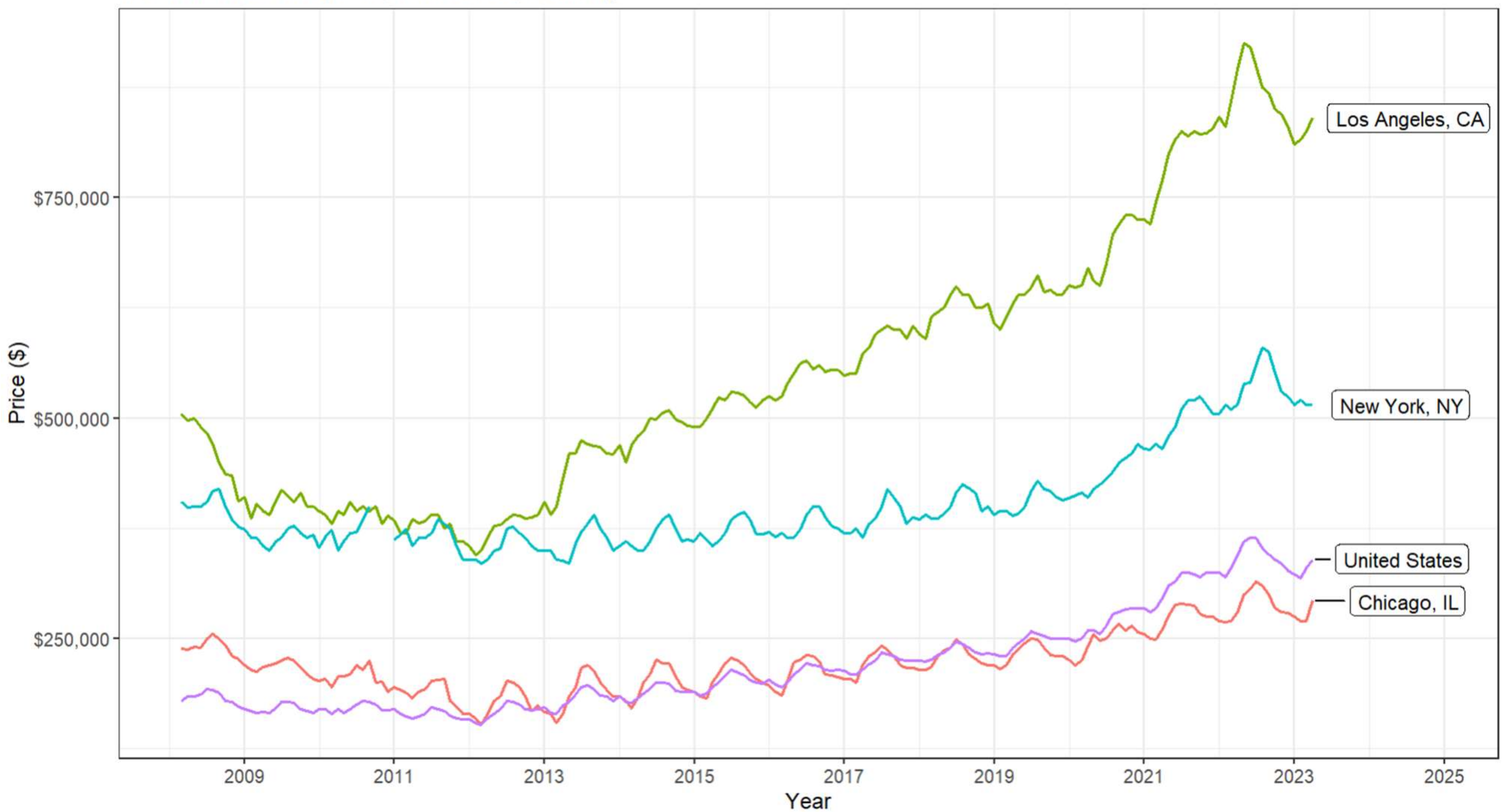



### Bottom 10 Decreases

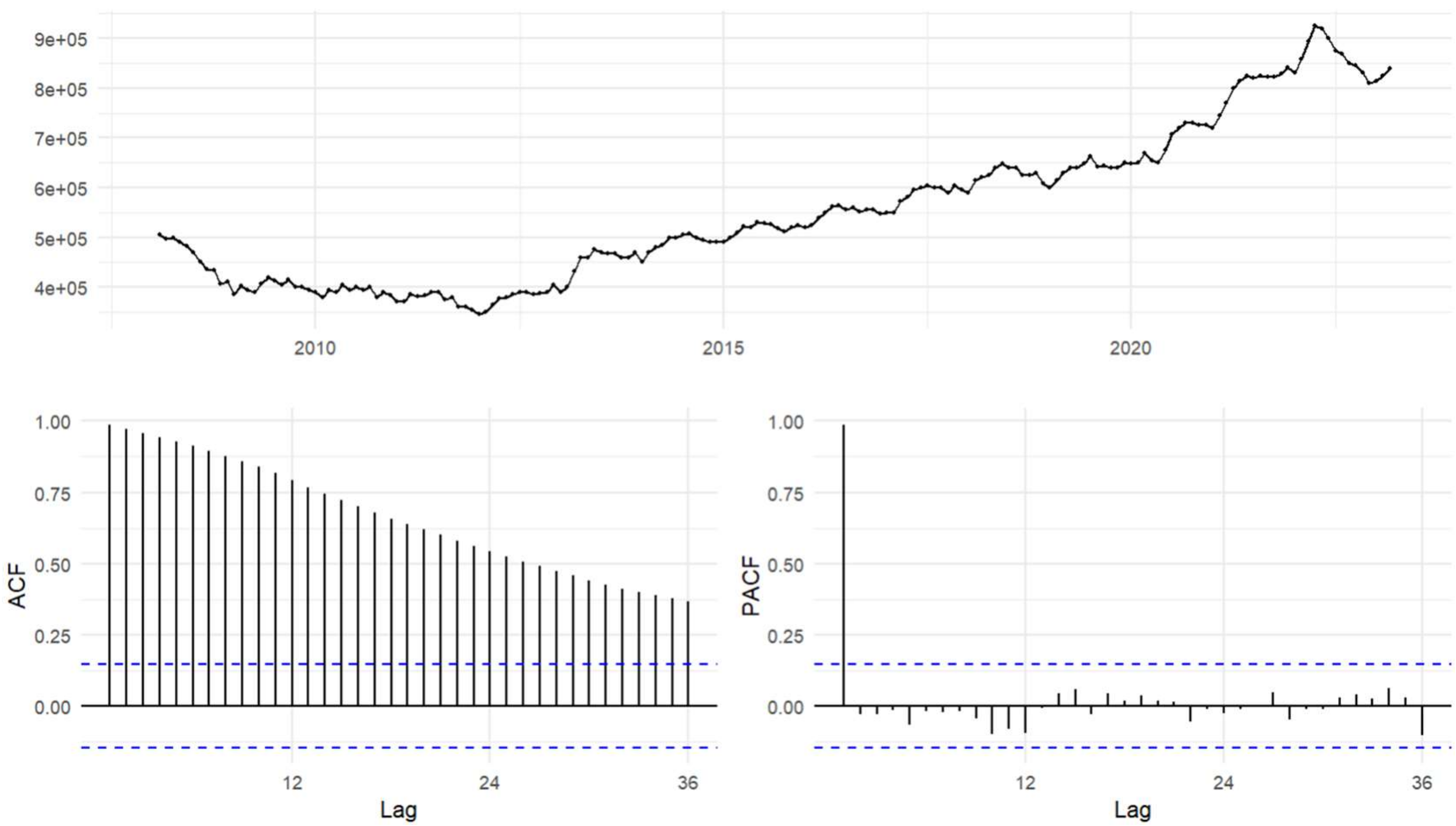


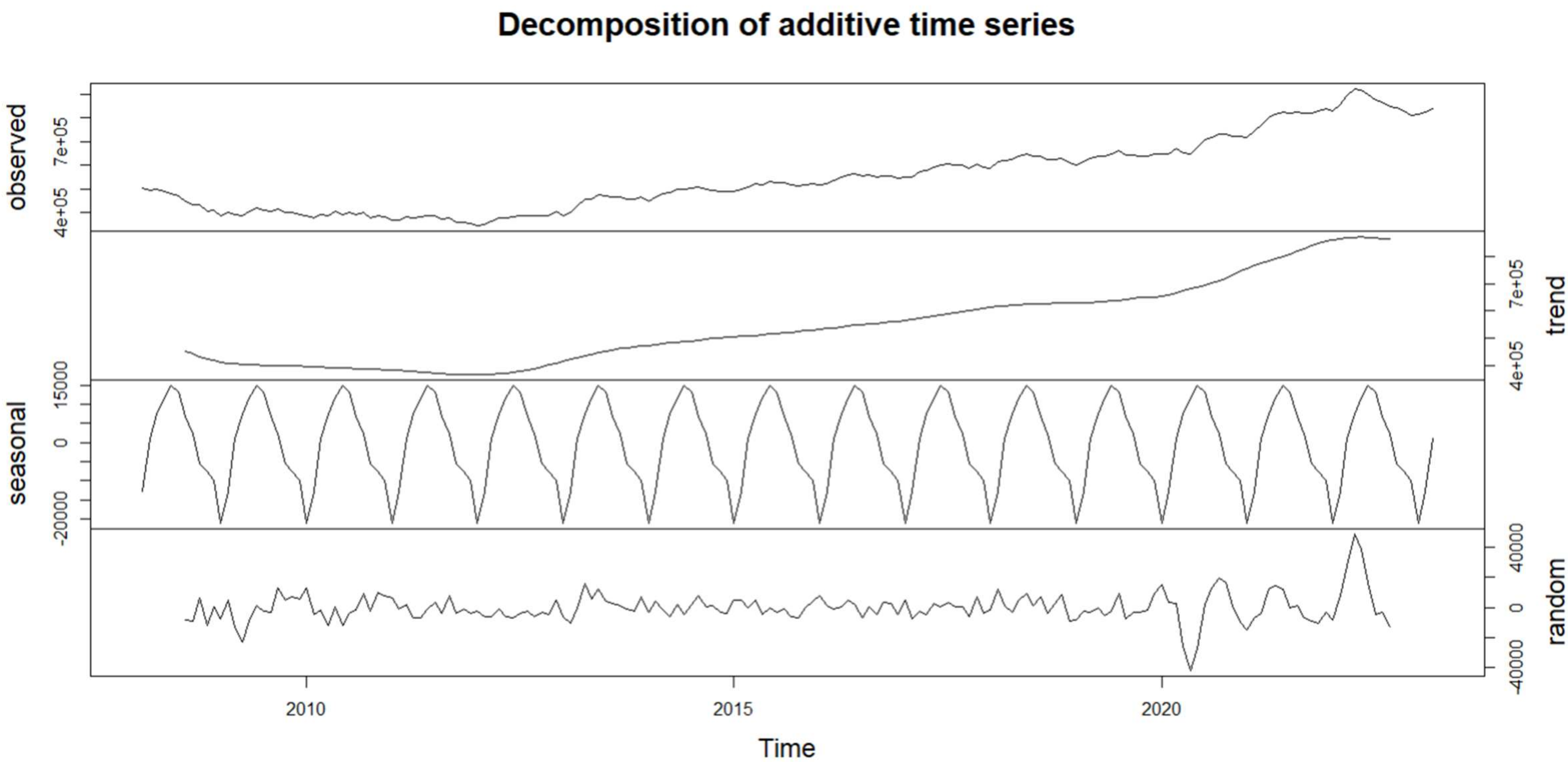

Dashboard

## Median Sale Price, Raw, All Homes, Monthly

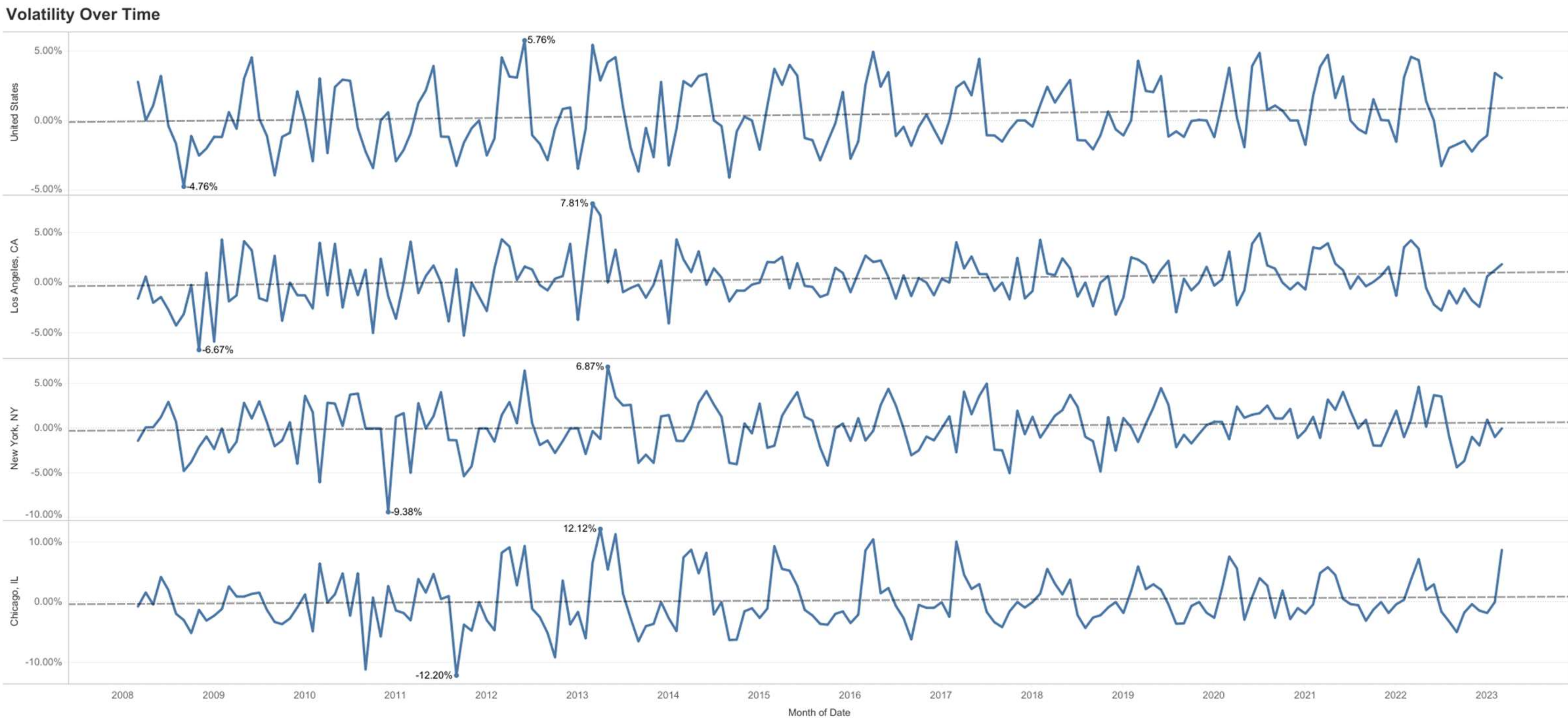


# Time Series EDA for Los Angeles, CA





Decomposition of additive time series
observed
4e+05
7e+05
trend
seasonal
15000
0
-20000
random
40000
-40000
2010
2015
2020
Time


Volatility Over Time
United States
5.00%
0.00%
-5.00%
5.76%
-4.76%
Los Angeles, CA
7.81%
-6.67%
New York, NY
-10.00%
6.87%
-9.38%
Chicago, IL
10.00%
-10.00%
12.12%
-12.20%
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
Month of Date

**G**eneralized

**A**uto

**R**egressive

**C**onditional

**H**eteroskedasticity

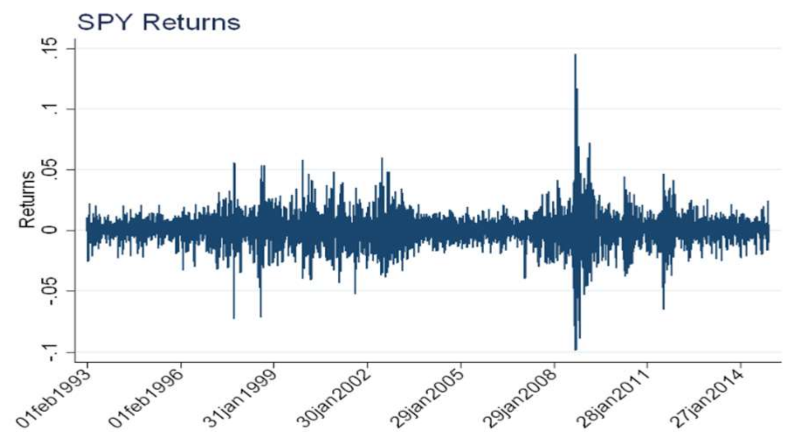
2

## GARCH Background

- Paul (Nik) Lopez

# Why are we using GARCH

1. One of the most popular models to forecast volatility in the industry
2. The model can capture volatility clustering that is often observed in time series, particularly in prices and other financial data.
3. The model can capture the fact that volatility can change over time.
4. Model flexibility.
   - The GARCH model has many extensions, such as the EGARCH (Exponential GARCH), that can accommodate asymmetric effects in volatility changes (not pursued here).



# Like AR and ARMA, GARCH is an extension of ARCH

| ARMA(1) an extension of AR(1) | GARCH(1,1) an extension of ARCH(1) |
|---|---|
| $\boldsymbol{AR}(\mathbf{1})\text{: } \boldsymbol{X_t} = \boldsymbol{\phi X_{t-1}} + \boldsymbol{Z_t}$<br>$where\ t = 0, \pm 1, \ldots, and\ Z_t \sim WN(0, \sigma^2)$<br>$X_{t-1} \rightarrow X_t$, $Z_{t-1} \rightarrow X_t$, $Z_t \rightarrow X_t$ | $\boldsymbol{ARCH}(\mathbf{1})\text{: } \boldsymbol{X_t} = \boldsymbol{Z_t\ \sigma_t} \ = \ \boldsymbol{Z_t} \sqrt{\boldsymbol{\alpha_0} + \boldsymbol{\alpha_1}(\boldsymbol{X_{t-1}})^2}$<br>$where\ t = 0, \pm 1, \ldots, and\ Z_t \sim WN(0, \sigma^2)\ and$<br>$\sigma_t\ is\ std\ dev\ of\ return\ of\ time\ t$<br>$Z_t \rightarrow X_t$, $\sigma_t \rightarrow X_t$, $X_{t-1} \rightarrow \sigma_t$ |
| $\boldsymbol{ARMA}\ (\mathbf{1}, \mathbf{1})\text{: } \boldsymbol{X_t} = \boldsymbol{\phi X_{t-1}} + \boldsymbol{\theta\ Z_{t-1}} + \boldsymbol{Z_t}$<br>$where\ t = 0, \pm 1, \ldots, and\ Z_t \sim WN(0, \sigma^2)$<br>$X_{t-1} \rightarrow X_t$, $Z_{t-1} \rightarrow X_t$, $Z_t \rightarrow X_t$ | $\boldsymbol{GARCH}(\mathbf{1}, \mathbf{1})\text{: } \boldsymbol{X_t} = \boldsymbol{Z_t\ \sigma_t} \ = \ \boldsymbol{Z_t} \sqrt{\boldsymbol{\alpha_0} + \boldsymbol{\alpha_1}(\boldsymbol{X_{t-1}})^2 + \boldsymbol{\beta\sigma^2_{t-1}}}$<br>$Z_t \rightarrow X_t$, $\sigma_t \rightarrow X_t$, $X_{t-1} \rightarrow \sigma_t$, $\sigma_{t-1} \rightarrow \sigma_t$ |

# How we forecast with the GARCH Model

Unlike other models, GARCH models returns versus the prices.

1. Take log returns or percentage changes and fit GARCH to the transformed time series
2. Use ACF and other measures to determine the correct parameters
3. Forecast volatility after the model is fit.
4. Take the last price of our data and reverse the log return or percentage change calculation to get the next predicted price

$$New\ Price = Old\ Price\ * (\ 1 + vol\ forecast)$$

3

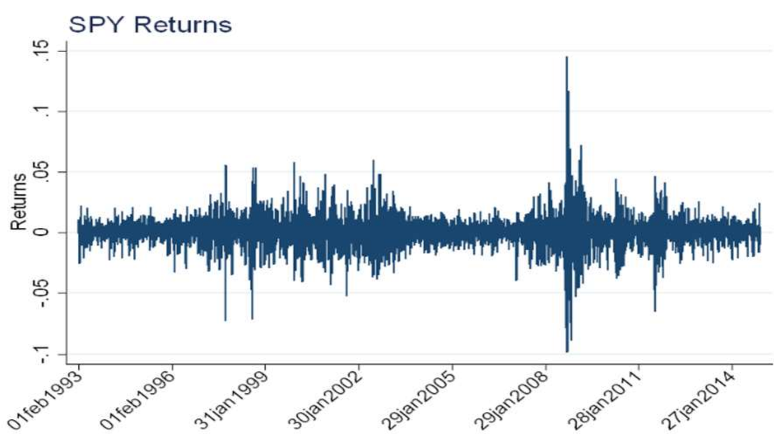
## GARCH Housing Data Analysis/Results

- Horace Tsai

# When/How to Use GARCH

## When to Use GARCH

- Widely used when trying to predict/forecast volatility such as finance or economic data
- Time series data has high points of volatility



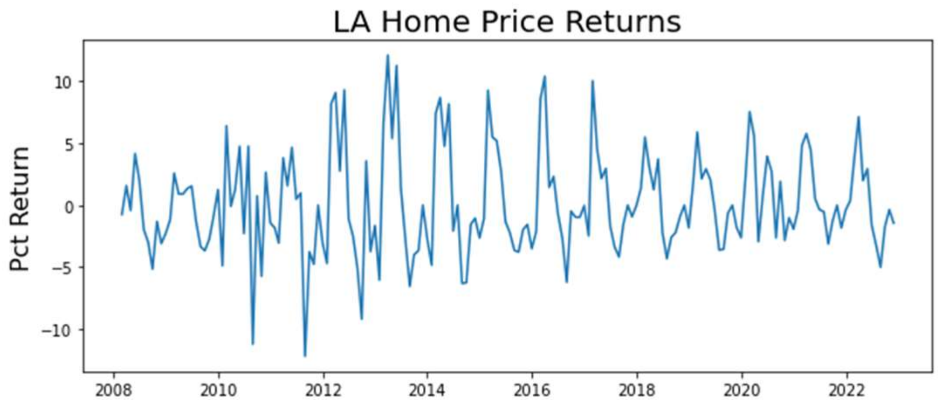

## How to Use GARCH

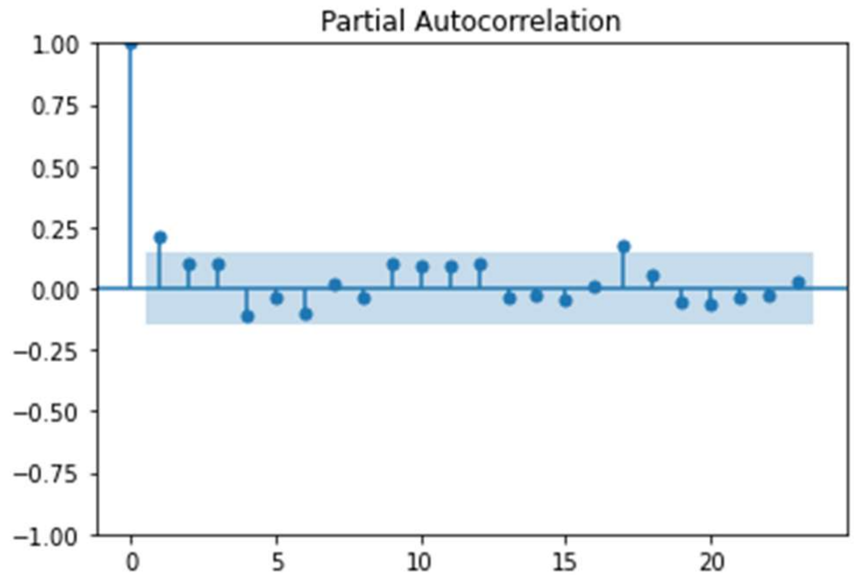
- Plot time series data to make sure data has high points of volatility
- Plot ACF and PACF
- Create GARCH Model based on analysis of ACF and PACF

# Analysis/ Results of GARCH

```
plt.figure(figsize=(10,4))
plt.plot(returns)
plt.ylabel('Pct Return', fontsize=16)
plt.title('LA Home Price Returns', fontsize=20)
```

```
Text(0.5, 1.0, 'LA Home Price Returns')
```



## GARCH(2, 2) Fit

$$X_t = Z_t\ \sigma_t\ = Z_t \sqrt{\alpha_0 + \alpha_1 (X_{t-1})^2 + \alpha_2 (X_{t-2})^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2}$$

| | | | |
|---|---|---|---|
| **Vol Model:** | GARCH | **Log-Likelihood:** | -395.646 |
| **Distribution:** | Normal | **AIC:** | 803.291 |
| **Method:** | Maximum Likelihood | **BIC:** | 822.382 |

Volatility Model

| | coef | std err | t | P>\|t\| | 95.0% Conf. Int. |
|---|---|---|---|---|---|
| **omega** | 0.1722 | 4.885 | 3.526e-02 | 0.972 | [ -9.403, 9.747] |
| **alpha[1]** | 0.0616 | 1.837 | 3.355e-02 | 0.973 | [ -3.540, 3.663] |
| **alpha[2]** | 6.9151e-10 | 2.791 | 2.478e-10 | 1.000 | [ -5.470, 5.470] |
| **beta[1]** | 2.2291e-08 | 2.735 | 8.150e-09 | 1.000 | [ -5.360, 5.360] |
| **beta[2]** | 0.8925 | 2.511 | 0.355 | 0.722 | [ -4.028, 5.813] |

## GARCH(1, 1) Fit

$$X_t = Z_t\ \sigma_t\ = Z_t \sqrt{\alpha_0 + \alpha_1 (X_{t-1})^2 + \beta \sigma_{t-1}^2}$$

| | | | |
|---|---|---|---|
| **Vol Model:** | GARCH | **Log-Likelihood:** | -396.589 |
| **Distribution:** | Normal | **AIC:** | 801.178 |
| **Method:** | Maximum Likelihood | **BIC:** | 813.906 |

Volatility Model

| | coef | std err | t | P>\|t\| | 95.0% Conf. Int. |
|---|---|---|---|---|---|
| **omega** | 0.1081 | 0.156 | 0.694 | 0.487 | [ -0.197, 0.413] |
| **alpha[1]** | 0.0325 | 0.106 | 0.306 | 0.760 | [ -0.176, 0.241] |
| **beta[1]** | 0.9405 | 0.127 | 7.396 | 1.407e-13 | [ 0.691, 1.190] |

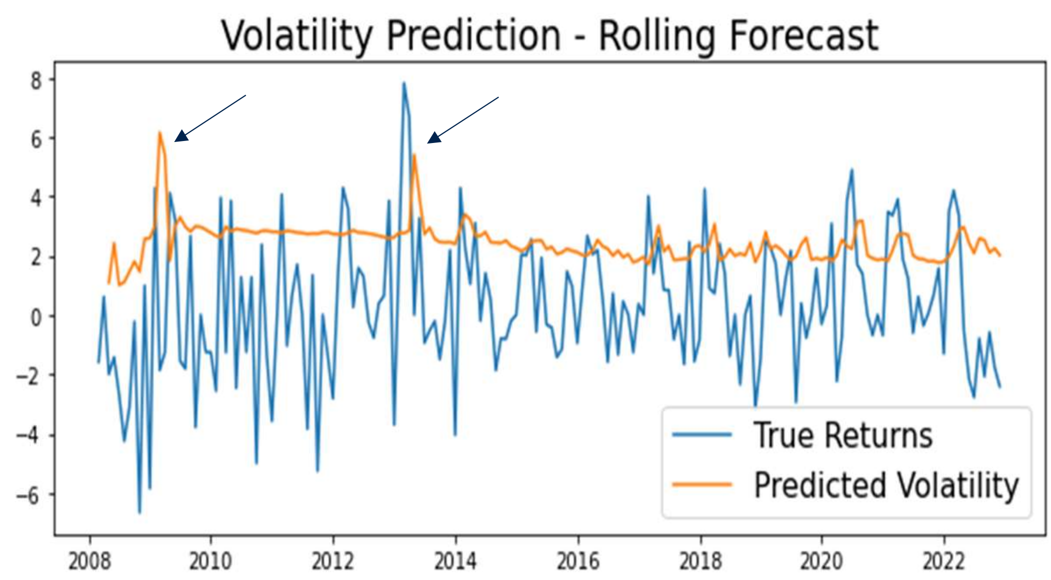
# GARCH(2, 0) = ARCH(2) Fit

$$X_t = Z_t\ \sigma_t\ = Z_t\sqrt{\alpha_0 + \alpha_1(X_{t-1})^2 + \alpha_2(X_{t-2})^2}$$

| | | | |
|---|---|---|---|
| **Vol Model:** | ARCH | **Log-Likelihood:** | -396.393 |
| **Distribution:** | Normal | **AIC:** | 800.786 |
| **Method:** | Maximum Likelihood | **BIC:** | 813.513 |

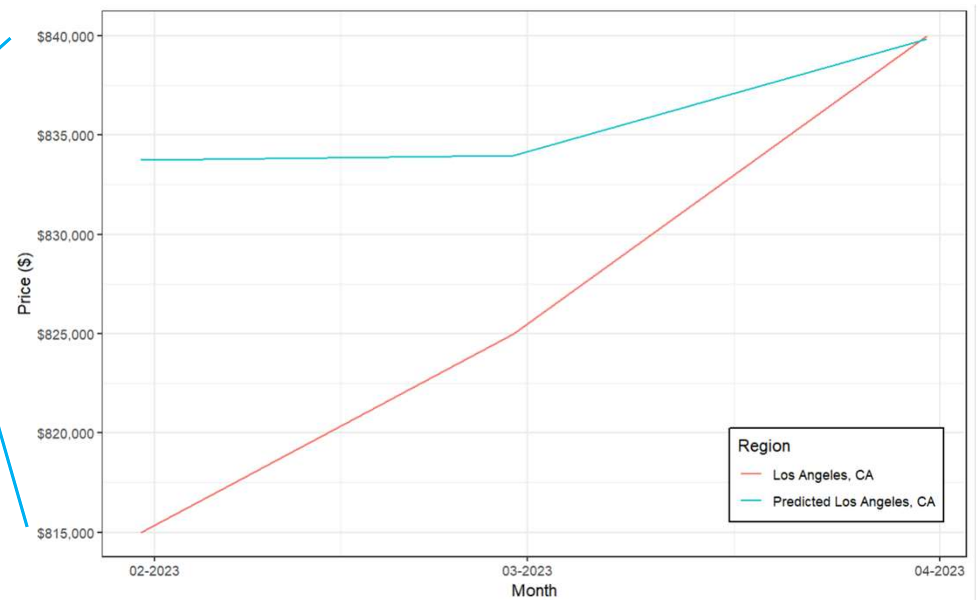
Volatility Model

| | coef | std err | t | P>\|t\| | 95.0% Conf. Int. |
|---|---|---|---|---|---|
| **omega** | 3.1326 | 0.640 | 4.894 | 9.904e-07 | [ 1.878, 4.387] |
| **alpha[1]** | 0.1488 | 8.041e-02 | 1.850 | 6.430e-02 | [-8.836e-03, 0.306] |
| **alpha[2]** | 0.2902 | 0.149 | 1.953 | 5.082e-02 | [-1.040e-03, 0.581] |




Forecast with Garch(2,0)

# GARCH(2,0) Price Predictions for First Quarter 2023


| Date | Actual | Predicted | Difference |
|---|---|---|---|
| 2023-01-31 | $815,000 | $833,745 | $18,745 |
| 2023-02-28 | $825,000 | $833,989 | $8,989 |
| 2023-03-31 | $840,000 | $839,850 | -$150 |

# Analysis/ Results of Forecast

## Model

- Our ARCH(2) fit the best with the lowest AIC and significant values

| Model | AIC | Ω (p-value) | α1 (p-value) | α2 (p-value) | β1 (p-value) | β2 (p-value) |
|---|---|---|---|---|---|---|
| GARCH (2,2) | 803.29 | 0.972 | 0.973 | 1.00 | 1.00 | 0.722 |
| GARCH (1,1) | 801.18 | 0.487 | 0.760 | N/A | 1.407e-13 | N/A |
| GARCH (2,0) | 800.79 | 9.904e-07 | 6.430e-02 | 5.082e-02 | N/A | N/A 🏆 |

## Predictions

- Ran our model to forecast volatility for the Months of January, February, and March
- Used the forecasted volatility to compute our new prices by using the following formula:

$$PercentChangeReturns = \frac{newprice - oldprice}{oldprice} * 100$$

| Date | Actual | Predicted | Difference |
|---|---|---|---|
| 2023-01-31 | $815,000 | $833,745 | $18,745 |
| 2023-02-28 | $825,000 | $833,989 | $8,989 |
| 2023-03-31 | $840,000 | $839,850 | -$150 |

4

# Conclusion

- Lucas Everett

# General Shortcomings of the GARCH Model

- The biggest shortcoming is volatility itself

# General Shortcomings of the GARCH Model

- Assumes volatility based on prior records and conditional variances
  - Models can be unstable for highly volatile data → incorrect forecasts
- Selecting the correct GARCH model can be cumbersome
- High computational cost due to intensive fitting of complex models to large data sets

# Extensions on GARCH

Could have tested other kinds of GARCH models on the data

- IGARCH (Integrated) - restricts the parameters
- NGARCH (Nonlinear) - addresses correlation
- EGARCH (Exponential) - accommodates asymmetric effects

Each distinct GARCH model tries to help access and interpret how the positive and negative changes in the series affect volatility and the return in assets/profits/losses.

# Where Can We Go From Here ...

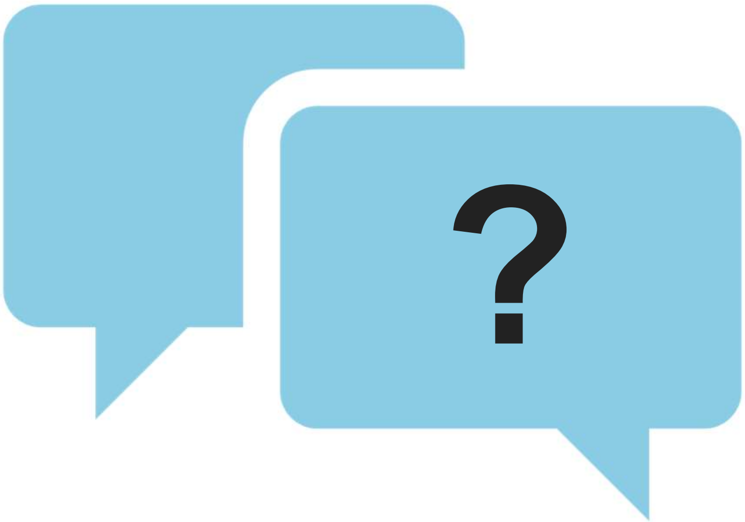
With the data available, there are a lot of different analyses one could zone in on:

- COVID-19 housing market affect
  - Was there a fundamental change in the volatility structure of the time series?
- Median housing prices of states before/after election season
- Try modeling other regions in the dataset
- Implement EGARCH model and other extensions

# Conclusion

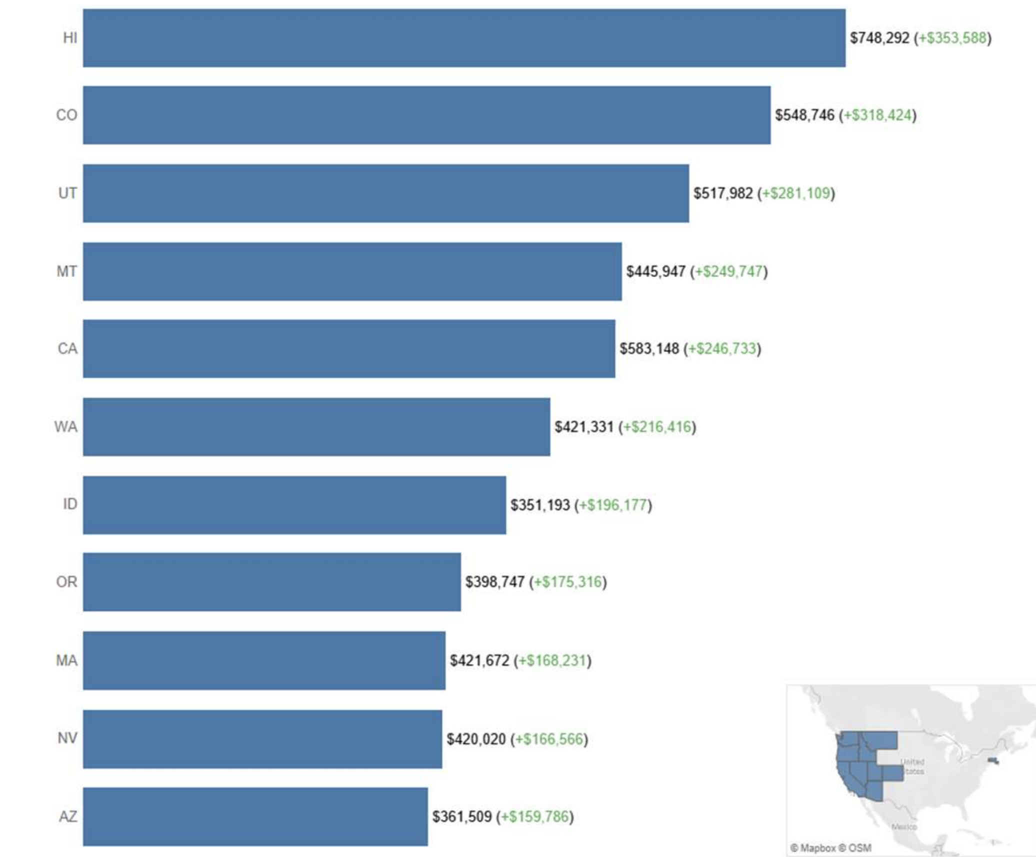
- Certain housing sale prices were shocking/counterintuitive to our initial thoughts after performing EDA
- ARCH(2) worked well for predicting Median Sale Price for Los Angeles, CA
- Small step towards learning more about ARCH/GARCH and becoming the best time series experts on the block!

# Questions & Answers?

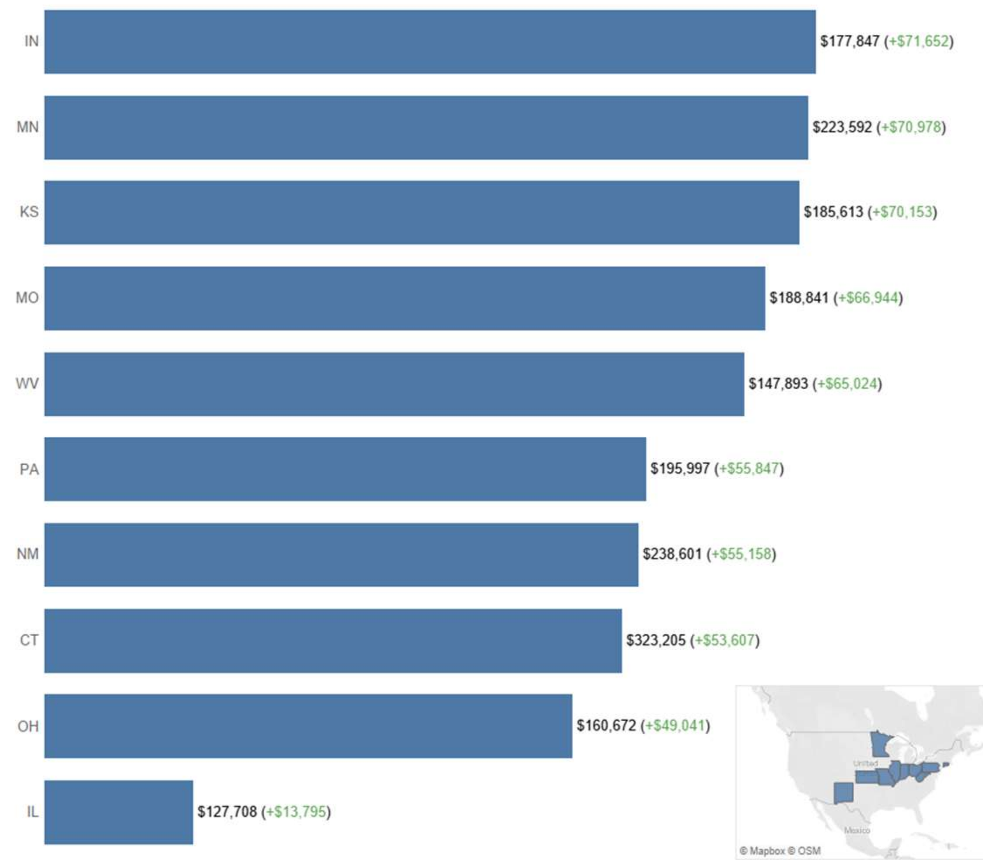


## Winners and Losers Since 2008 (sorted on increase):

### Top 10 Increases




### Bottom 10 Increases




Dashboard link

Cal State Fullerton