TBD*

TBD

Ziheng Zhong

April 8, 2024

TBD

Table of contents

*Code and data are available at: https://github.com/iJustinn/House_Price.git

Table 1: Summary statistics of the California housing dataset

Table 2: Count of missing values for each variable

1 Introduction

2 Data

Data used in this paper was cleaned, processed and tested with the programming language R (R Core Team 2022). Also with support of additional packages in R: `tidyverse` (Wickham et al. 2019), `ggplot2` (Wickham 2016), `janitor` (Firke 2023), `readr` (Wickham, Hester, and Bryan 2023), `knitr` (Xie 2014), `modelsummary` (Arel-Bundock 2023), `testthat` (Wickham Year of publication), `KableExtra` (Zhu 2023), `viridis` (Garnier et al. 2018), `lubridate` (Grolemund and Wickham 2021), `maps` (Deckmyn et al. 2021), `mgcv` (Wood 2021).

Table 3: Count of missing values for each variable after cleaning

Table 4: Modeling Results for Linear Models

	Multiple Regression	Polynomial Regression
(Intercept)	−13 969 628.019	212 976.031
	(395 357.616)	(1359.110)
Year	6958.769	
	(196.466)	
NumBedroom	59 897.327	
	(993.637)	
poly(Year, 2)1		3 866 485.706
		(105 994.679)
poly(Year, 2)2		1 601 816.793
		(105 993.208)
poly(NumBedroom, 2)1		6 591 537.312
		(105 993.313)
poly(NumBedroom, 2)2		1 365 656.076
		(105 994.574)
Num.Obs.	6082	6082
R2	0.445	0.479
R2 Adj.	0.445	0.479
AIC	158 396.4	158 018.2
BIC	158 423.3	158 058.5
Log.Lik.	−79 194.199	−79 003.096
F	2440.891	1397.712
RMSE	109 331.51	105 949.60

Table 5

	GAM Regression
(Intercept)	212 976.031
	(1322.680)
Num.Obs.	6082
R2	0.506
AIC	157 697.2
BIC	157 801.7
RMSE	103 028.50

Modeling Results for Non-linear Models

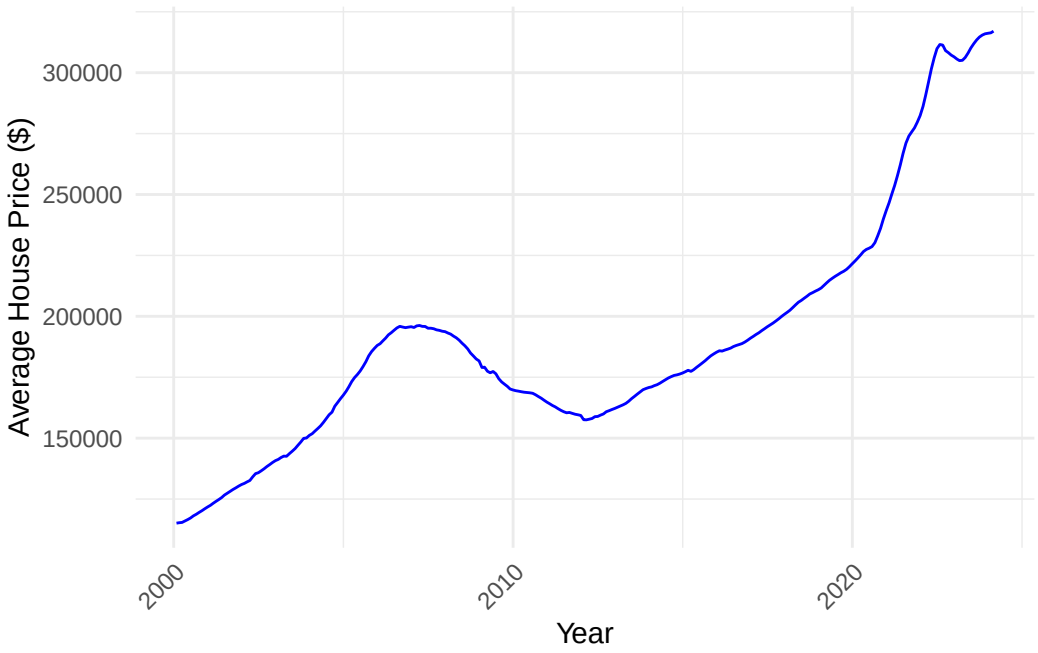


Figure 1: Trend of Average House Price from 2000 to 2024

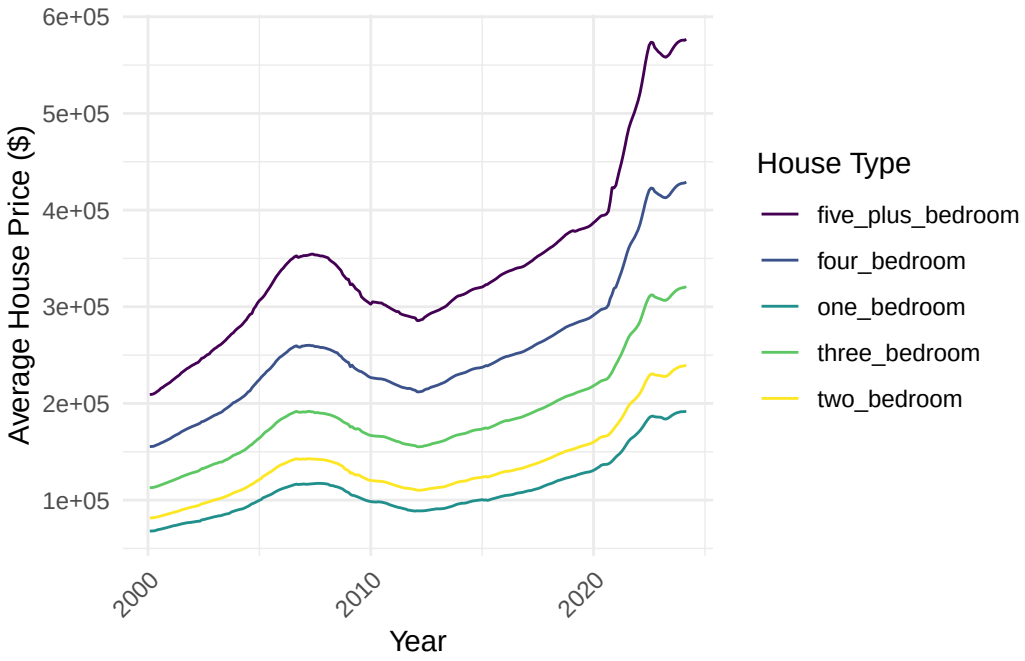


Figure 2: Trend of Average House Price from 2000 to 2024 by House Type

2.1 Source

2.2 Method

3 Results

3.1 Data Trend

3.2 Heat Maps

3.3 Modeling

4 Discussion

4.1 Demographic Shifts

4.2 Health-related Behaviors

4.3 Government Policies

4.4 Environmental Changes

4.5 Possible Improvements

5 Conclusion

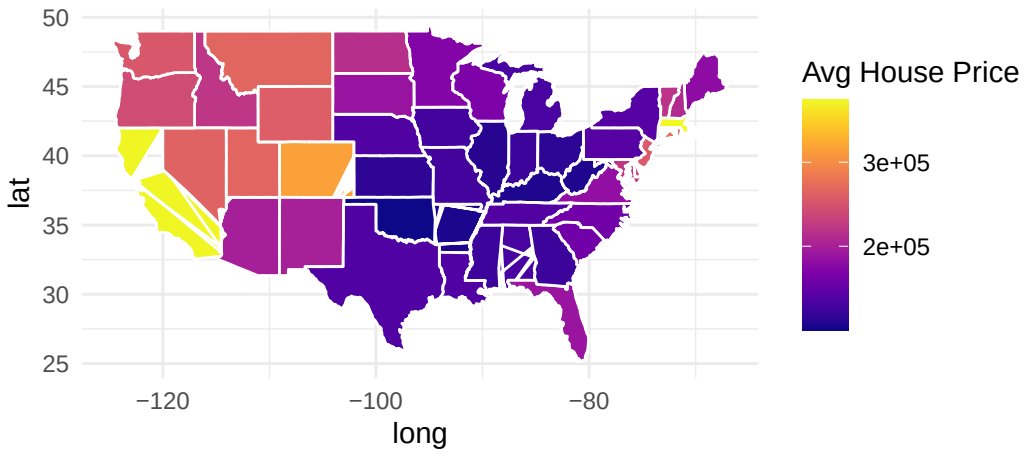


Figure 3: Average Price by State in the US for All House Types

A Appendix

A.1 Datasheet

Motivation

Composition

Collection process

Preprocessing/cleaning/labeling

Uses

Distribution

Maintenance

References

Arel-Bundock, Vincent. 2023. *Modelsummary: Summary Tables and Plots for Statistical Models and Data: Beautiful, Customizable, and Publication-Ready*. https://vincentarelbundock.github.io/modelsummary/.

Deckmyn, Alex, Original S code by Richard A. Becker, Allan R. Wilks. R version by Ray Brownrigg. Enhancements by Thomas P Minka, and Alex Deckmyn. 2021. *Maps: Draw Geographical Maps*. https://CRAN.R-project.org/package=maps.

Firke, Sam. 2023. *Janitor: Simple Tools for Examining and Cleaning Dirty Data*. https://CRAN.R-project.org/package=janitor.

Garnier, Simon, Noam Ross, Bob Rudis, and Marco Sciaini. 2018. *Viridis: Default Color Maps from 'Matplotlib'*. https://CRAN.R-project.org/package=viridis.

Grolemund, Garrett, and Hadley Wickham. 2021. *Lubridate: Make Dealing with Dates a Little Easier*. https://CRAN.R-project.org/package=lubridate.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Wickham, Hadley. Year of publication. *Testthat: Get Started with Testing*. https://CRAN.R-project.org/package=testthat.

———. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

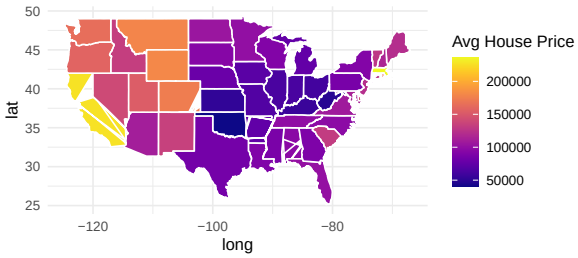
Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2023. *Readr: Read Rectangular Text Data*. https://CRAN.R-project.org/package=readr.

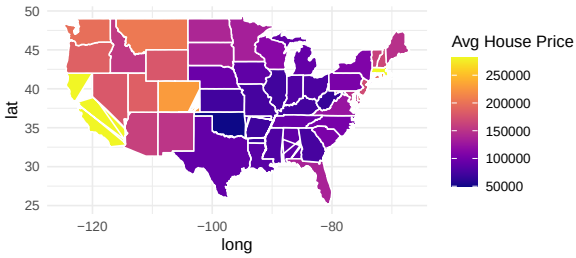
Wood, Simon N. 2021. *Mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation*. https://CRAN.R-project.org/package=mgcv.

Xie, Yihui. 2014. *Knitr: A Comprehensive Tool for Reproducible Research in R*. Edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. http://www.crcpress.com/product/isbn/9781466561595.

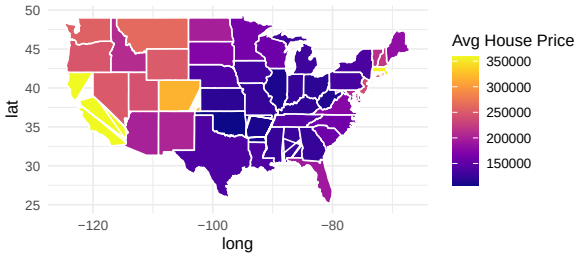
Zhu, Hao. 2023. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.



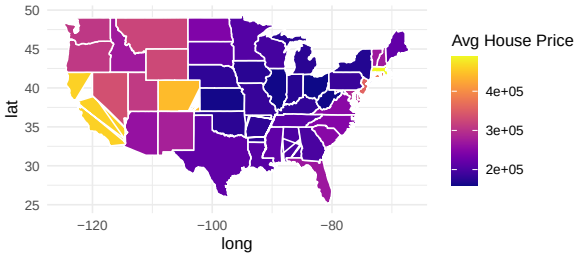
(a) one bedroom



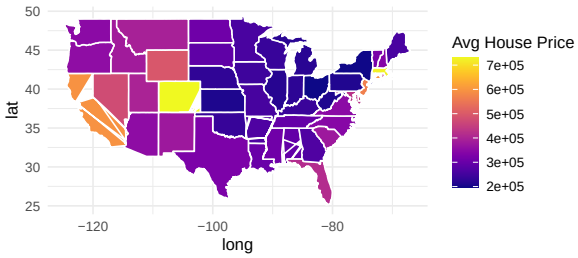
(b) two bedrooms



(c) three bedrooms



(d) four bedrooms



(e) five plus bedrooms

Figure 4: Average Price by State in the US for Different House Types

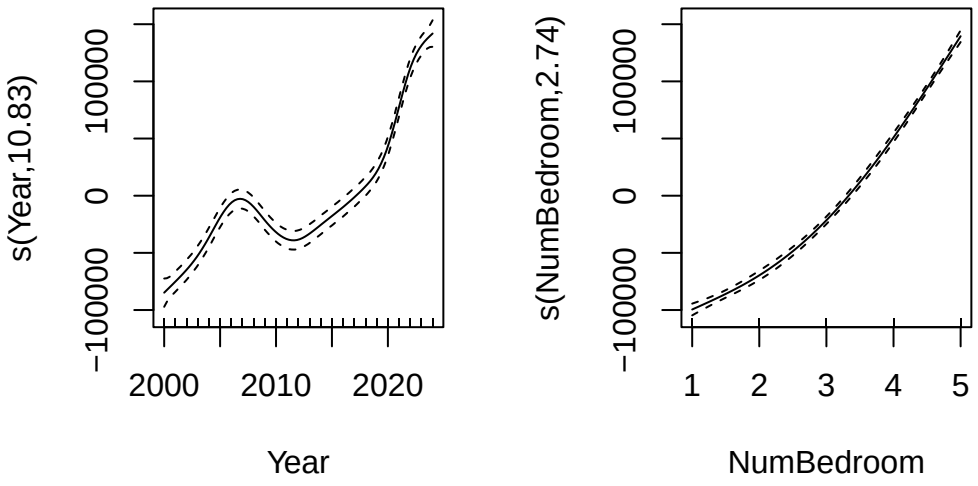

s(Year,10.83)
−100000
0
100000
2000
2010
2020
Year
s(NumBedroom,2.74)
−100000
0
100000
1
2
3
4
5
NumBedroom

Figure 5