Impact of homelessness

UCB Data Bootcamp Project 1
Alicia, Ryan & Navi
November 30th, 2020

Motivation

Homelessness is an continued issue which burden’s our community across the globe, specially elevated with COVID.

Goal of the project is to dive into factors that are related to homeleness and housing prices using various data available to us nationwide.

Variables that influence homeleness

- *Unemployment Rate*
- *Population size/infrastructure*
- *Cost of living*
- *Social services / shelters / non-profits*
- *Per capita personal income*
- *Education*
- *Climate*

Data Sources

https://www.zillow.com/research/data/
2015-2020; Zillow economic data

https://www.hudexchange.info/programs/coc/coc-homeless-populations-and-subpopulations-reports/

CoC (*Continuums of Care*) Homeless Populations and Subpopulations Reports; 2015-2019

https://www.hudexchange.info/resource/5948/2019-ahar-part-1-pit-estimates-of-homelessness-in-the-us/

PIT and HIC Data Since 2007 - HUD Exchange

2015-2019; Point in time estimates by state
2015-2019; Housing inventory count by state
(*Each of these resources is reported by Continuums of Care to the US Housing and Urban Development (HUD) department*)

https://www.census.gov/quickfacts/fact/table/

https://worldpopulationreview.com/state-rankings/homeless-population-by-state#dataTable

https://www.bls.gov/charts/state-employment-and-unemployment/state-unemployment-rates-animated.html

Data Cleanup & Exploration

➢**Unemployment Rates - U.S. Bureau of Labor Statistics**

- *Data set contains national estimates of unemployment percent or rate by year from 1960 through 2020, captured monthly*
- *We cleaned the data using pandas, matplot to use the last five years data 2015 through 2019*
- *Since the raw data was on monthly cadence, we created new dataframe to pull yearly*

	State	1/1/2015	2/1/2015	3/1/2015	4/1/2015	5/1/2015	6/1/2015	7/1/2015	8/1/2015	9/1/2015	...	12/1/2019	1/1/2020	2/1/2020	3/1/2020	4/1/202
0	Alabama	6.0	6.0	6.1	6.1	6.2	6.2	6.2	6.1	6.1	...	2.7	2.7	2.7	3	13.
1	Alaska	6.5	6.5	6.5	6.5	6.5	6.5	6.5	6.5	6.5	...	6.1	6.0	5.8	5.2	13.
2	Arizona	6.4	6.3	6.3	6.2	6.1	6.1	6.0	6.0	5.9	...	4.5	4.5	4.5	6.1	13.
3	Arkansas	5.6	5.5	5.5	5.4	5.3	5.1	5.0	4.8	4.7	...	3.5	3.5	3.5	5	10.
4	California	6.8	6.7	6.6	6.5	6.4	6.2	6.1	6.0	5.9	...	3.9	3.9	3.9	5.5	16.
5	Colorado	4.3	4.2	4.2	4.1	4.1	4.0	3.8	3.7	3.6	...	2.5	2.5	2.5	5.2	12.
6	Connecticut	6.1	6.0	5.9	5.8	5.7	5.7	5.6	5.6	5.6	...	3.8	3.7	3.8	3.4	8.
7	Delaware	5.1	5.0	5.0	4.9	4.9	4.9	4.8	4.8	4.8	...	4.0	4.0	3.9	5	14.
8	District of Columbia	7.4	7.3	7.2	7.1	7.0	6.9	6.8	6.7	6.6	...	5.3	5.2	5.1	6	11.

Graphs/Data for UE Rates

	State	Date	Year	Month	Unemployment Rate
0	Alabama	01-01-2015	2015	1	6
1	Alabama	02-01-2015	2015	2	6
2	Alabama	03-01-2015	2015	3	6.1
3	Alabama	04-01-2015	2015	4	6.1
4	Alabama	05-01-2015	2015	5	6.2
5	Alabama	06-01-2015	2015	6	6.2
6	Alabama	07-01-2015	2015	7	6.2
7	Alabama	08-01-2015	2015	8	6.1
8	Alabama	09-01-2015	2015	9	6.1

New df for UE rate, years cleaned up, Year column added, UE rate added

	Month	Unemployment Rate	Month_Name
0	1	1375.5	Jan
1	10	1112.0	Oct
2	11	1106.0	Nov
3	12	1100.9	Dec
4	2	1364.2	Feb

Month name column added

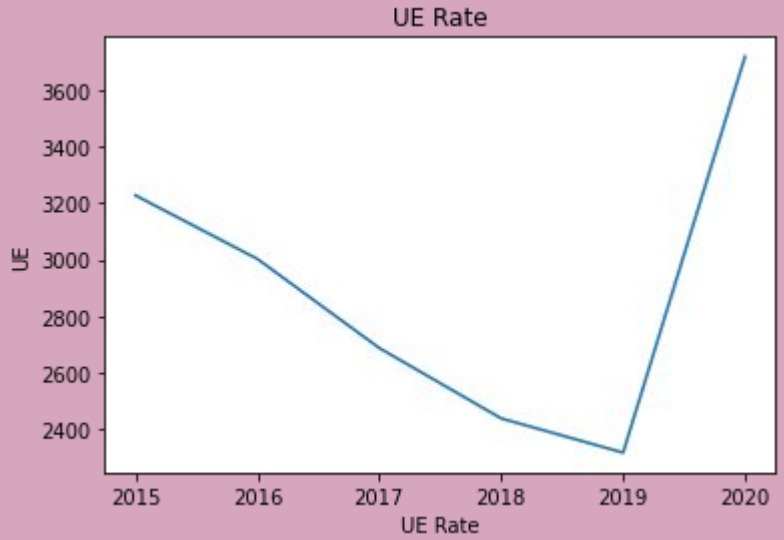
Year	
2015	3227.1
2016	3002.0
2017	2686.2
2018	2436.7
2019	2315.4
2020	3719.1

UE rate broken down by Years

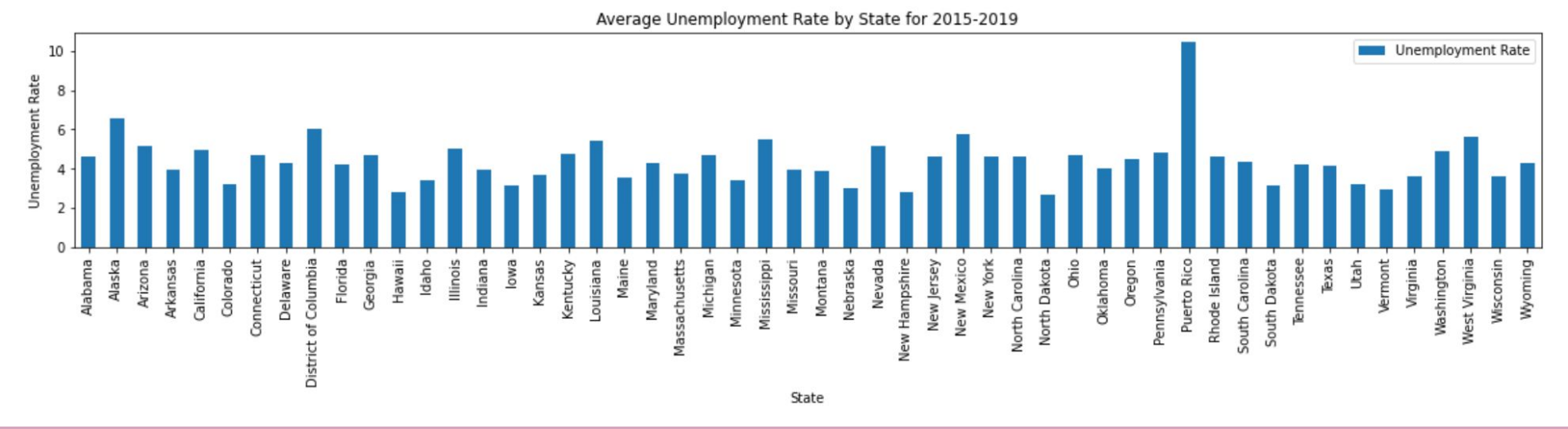
Top 10 States with highest homeless people

California (151,278)
New York (92,091)
Florida (28,328)
Texas (25,848)
Washington (21,577)
Massachusetts (18,471)
Oregon (15,876)
Pennsylvania (13,199)
Georgia (10,443)
Ohio (10,345)

Graphs/Data for UE Rates



Average UE rates by state 2015-2019



Data Cleanup & Exploration

➢ **HUD - The U.S. Department of Housing & Urban Development Dataset**

- *Data set contained national estimates of homelessness by CoC (Continuums of Care) Number and CoC Name and estimates of different categories of homelessness from 1996 - 2019*
- *Estimates of homeless veterans were included since 2019*
- *Hence the column headers count were not same across all worksheets*
- *Year information from column headers was removed and a new column 'Year' was inserted across all worksheets using VBA*
- *Data from all worksheets were consolidated using pandas concat method and removed columns that had NaN*

Graphs/Data for HUD

	Year	State	Number of CoCs	Overall Homeless	Overall Homeless - Under 18	Overall Homeless - Age 18 to 24	Overall Homeless - Over 24	Overall Homeless - Female	Overall Homeless - Male	Overall Homeless - Transgender	...	Overall Homeless Parenting Youth Age 18-24	Sheltered ES Homeless Parenting Youth Age 18-24	Sheltered TH Homeless Parenting Youth Age 18-24	Shelter To Homele Parenti You Age 18-
0	2019.0	AK	2.0	1,907	335	204	1,368	761	1,137	6	...	20	6	14	
1	2019.0	AL	8.0	3,261	629	199	2,433	1,383	1,873	3	...	22	10	11	
2	2019.0	AR	5.0	2,717	303	317	2,097	1,022	1,688	4	...	11	7	4	
3	2019.0	AZ	3.0	10,007	1,602	708	7,697	3,683	6,297	16	...	58	17	40	
4	2019.0	CA	44.0	151,278	14,305	12,673	124,300	50,467	98,404	1,764	...	1,019	457	291	7
5	2019.0	CO	3.0	9,619	1,458	685	7,476	3,238	6,289	54	...	80	19	59	

Raw data prior to clean up on the notebook

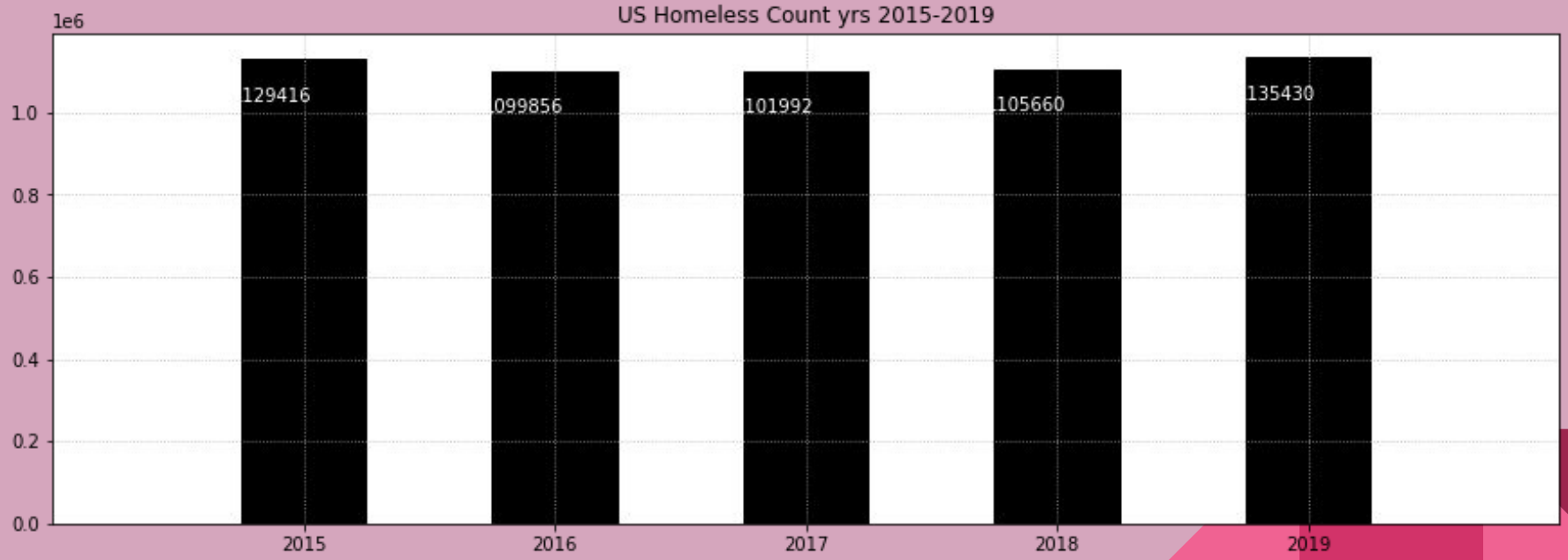
Clean data

	Year	State	Overall Homeless	Sheltered Total Homeless	Unsheltered Homeless	Overall Homeless Individuals	Sheltered Total Homeless Individuals	Unsheltered Homeless Individuals	Overall Homeless People in Families	Sheltered Total Homeless People in Families	...	Unsheltered Chronically Homeless	Overall Chronically Homeless Individuals	Sheltered Total Chronically Homeless Individuals
0	2019	AK	1907.0	1634	273	1384	1118	266	523	516	...	99	231	134
1	2019	AL	3261.0	2070	1191	2519	1429	1090	742	641	...	139	350	221
2	2019	AR	2717.0	1302	1415	2303	972	1331	414	330	...	453	676	223
3	2019	AZ	10007.0	5475	4532	7538	3099	4439	2469	2376	...	1210	1766	599
4	2019	CA	151278.0	42846	108432	128777	25323	103454	22501	17523	...	33670	39275	6483
5	2019	CO	9619.0	7431	2188	7263	5228	2035	2356	2203	...	918	2106	1211

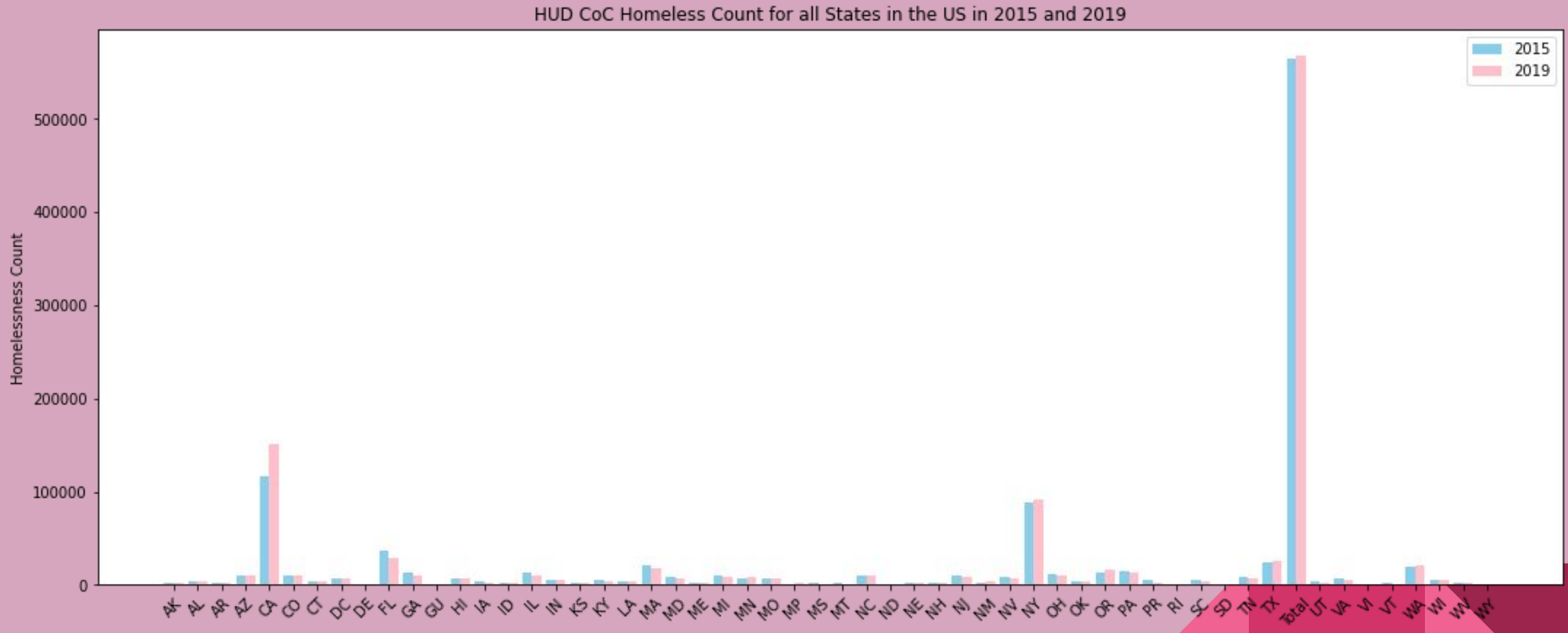
Overall Homeless Count 2015-2019

Year	
2015	1129416.0
2016	1099856.0
2017	1101992.0
2018	1105660.0
2019	1135430.0

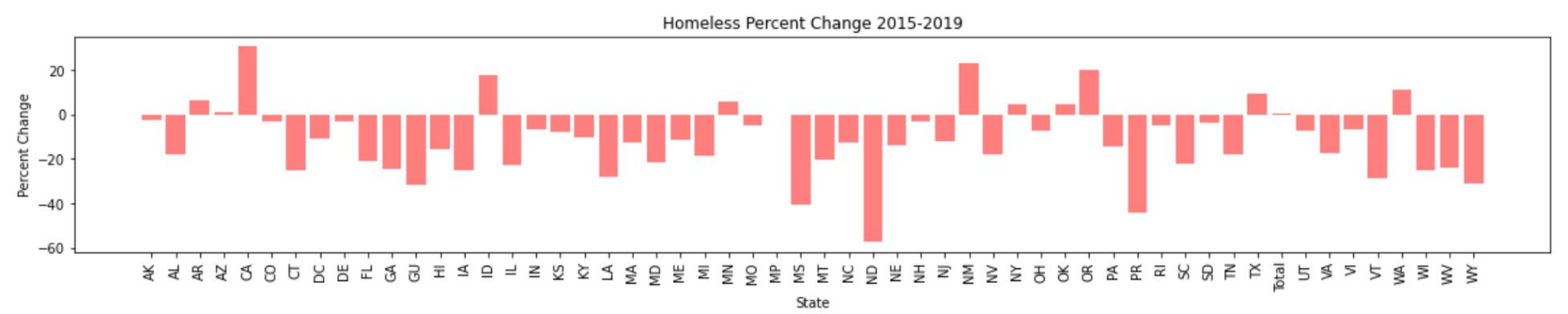
Name: Overall Homeless



Homeless Count in US 2015-2019



Homeless Change yrs 2015-2019 (nation wide)



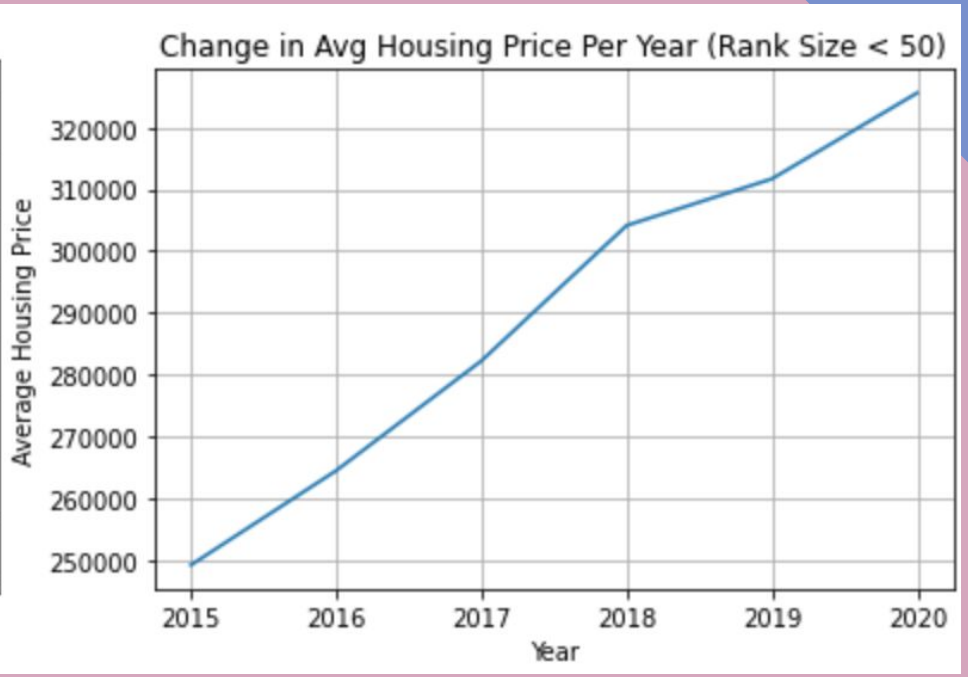
Zillow Housing Price Data 2015 to 2020

	RegionName	StateName	SizeRank	RegionID	1/31/2015	2/28/2015	3/31/2015	4/30/2015	5/31/2015	6/30/2015	...	1/31/
0	United States	Country	0	102001	190722.0	191467.0	192335.0	193235.0	194147.0	194958.0	...	24
1	New York, NY	NY	1	394913	408905.0	410065.0	410802.0	411779.0	412581.0	413688.0	...	48
2	Los Angeles-Long Beach-Anaheim, CA	CA	2	753899	526530.0	530547.0	534737.0	537106.0	539643.0	542210.0	...	68
3	Chicago, IL	IL	3	394463	206017.0	206142.0	206800.0	207924.0	208932.0	209143.0	...	24
4	Dallas-Fort Worth, TX	TX	4	394514	177328.0	178799.0	180484.0	182120.0	183830.0	185516.0	...	25
...	...	...	...	...	...	...	...	...	...	...	...	
909	Lamesa, TX	TX	929	394767	53736.0	53906.0	54068.0	54405.0	54750.0	55107.0	...	7
910	Craig, CO	CO	930	753874	164710.0	166051.0	167511.0	168959.0	170713.0	172727.0	...	18
911	Pecos, TX	TX	931	394968	NaN	NaN	NaN	NaN	NaN	NaN	...	12
912	Vernon, TX	TX	932	395188	56866.0	57187.0	57451.0	57682.0	58006.0	58341.0	...	7
913	Ketchikan, AK	AK	933	394743	275225.0	276036.0	276588.0	277246.0	277891.0	278764.0	...	31

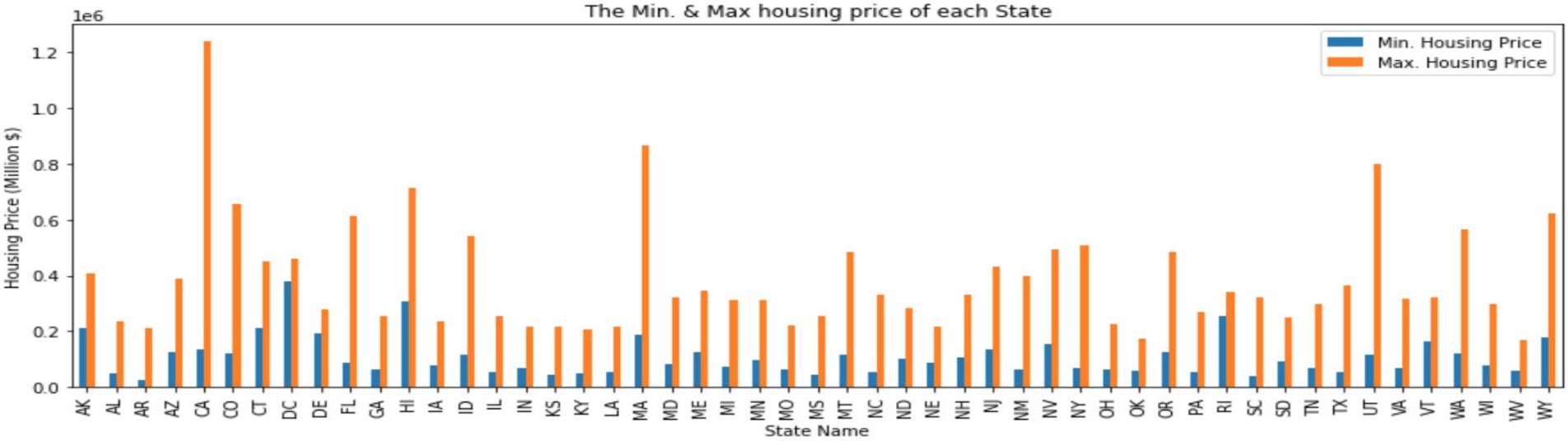
914 rows × 74 columns

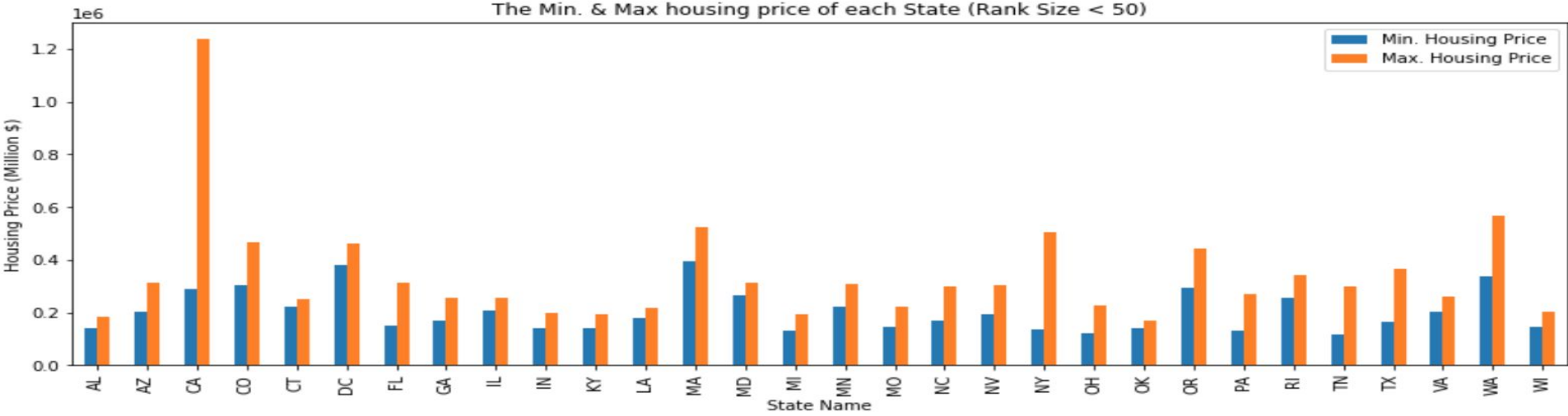
Average Housing Price 2015 - 2020

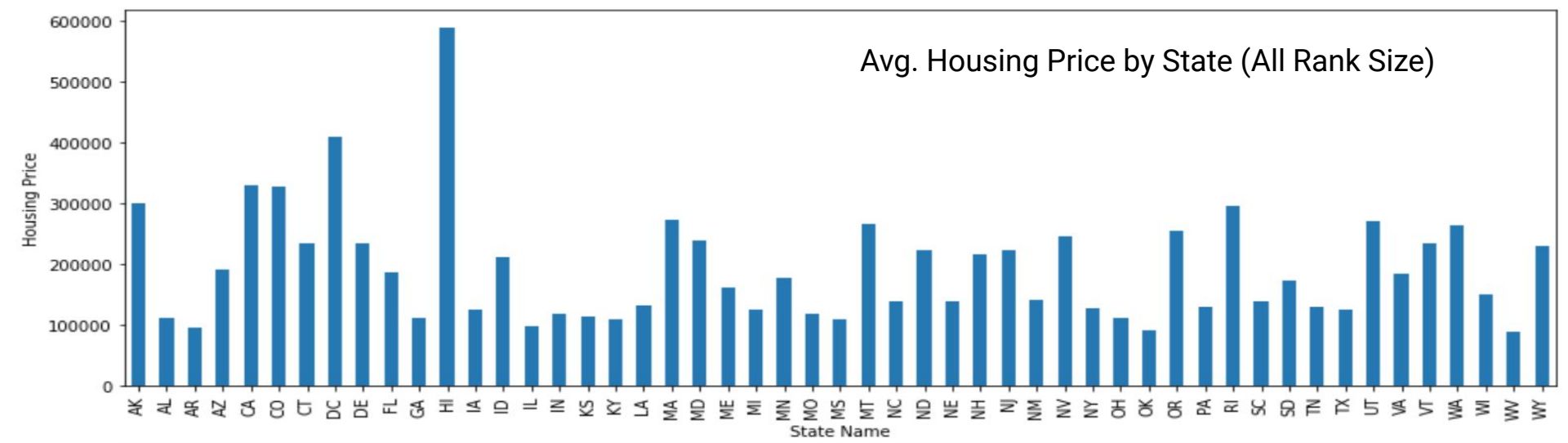
The Whole Data (All Rank Size)

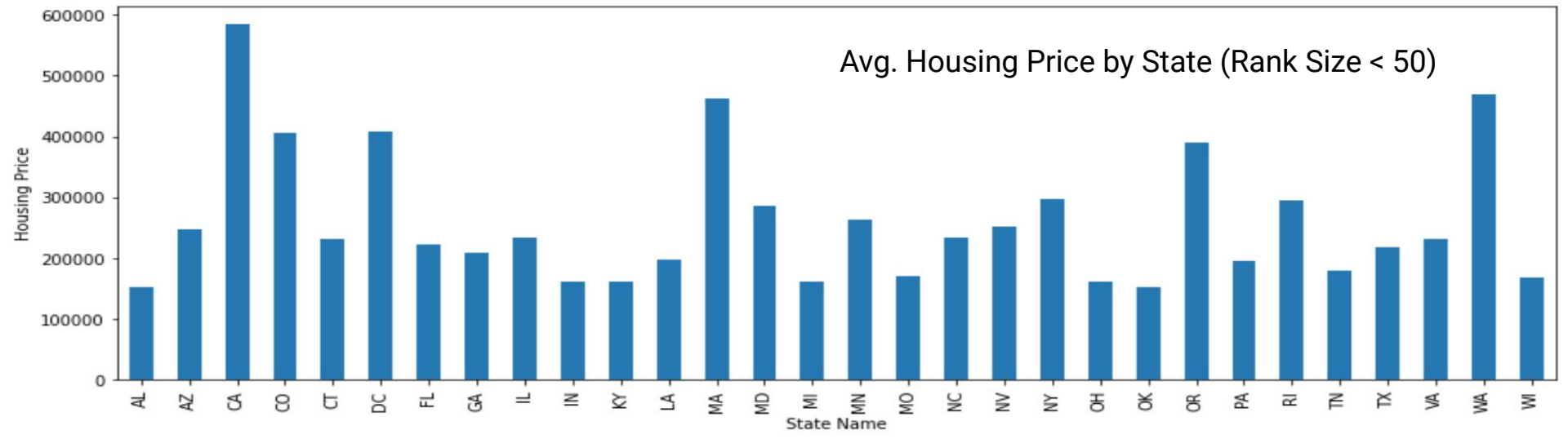


Only Data from Rank Size < 50


The Min. & Max housing price of each State
1e6
Min. Housing Price
Max. Housing Price
Housing Price (Million $)
State Name
AK AL AR AZ CA CO CT DC DE FL GA HI IA ID IL IN KS KY LA MA MD ME MI MN MO MS MT NC ND NE NH NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY

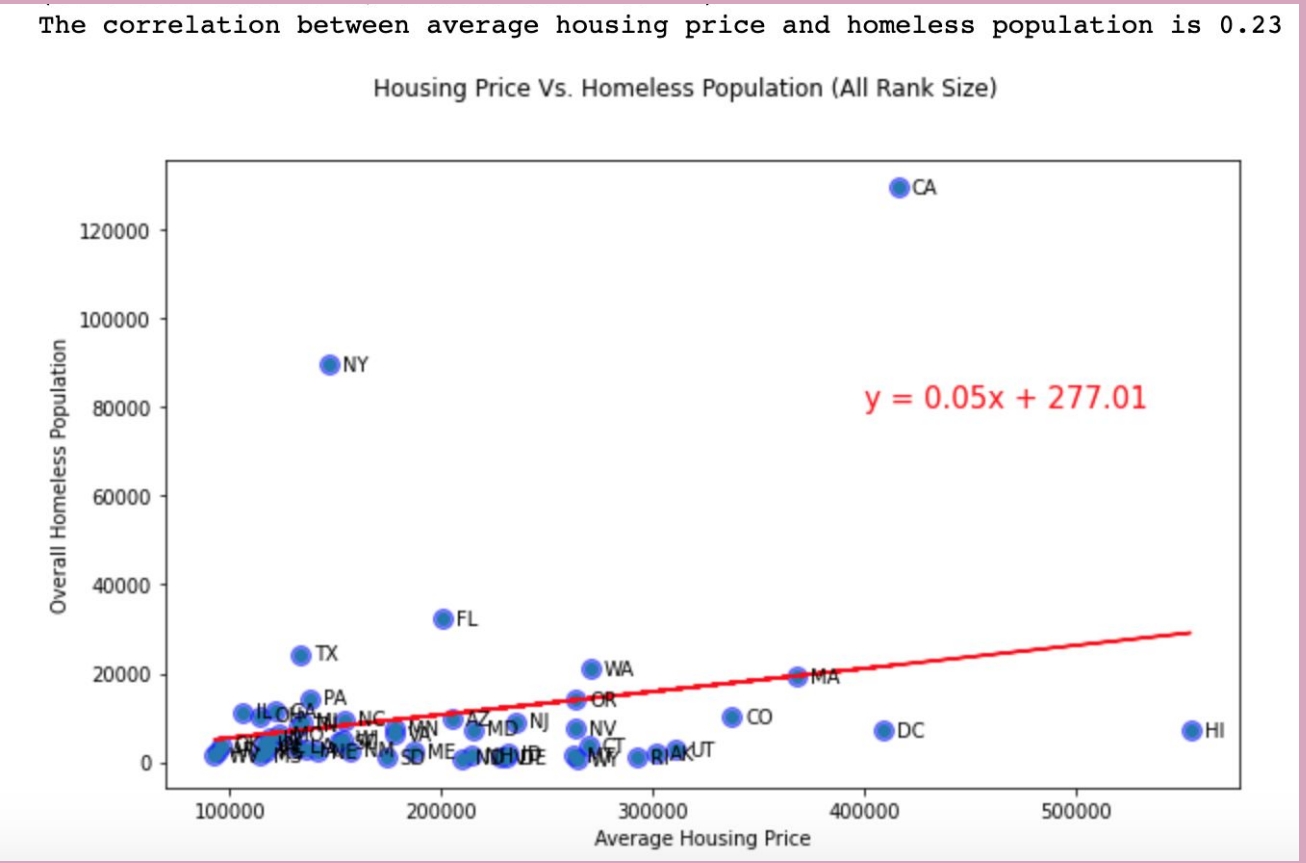

The Min. & Max housing price of each State (Rank Size < 50)
1e6
Min. Housing Price
Max. Housing Price
Housing Price (Million $)
State Name
AL AZ CA CO CT DC FL GA IL IN KY LA MA MD MI MN MO NC NV NY OH OK OR PA RI TN TX VA WA WI


Avg. Housing Price by State (All Rank Size)
Housing Price
State Name
0
100000
200000
300000
400000
500000
600000
AK AL AR AZ CA CO CT DC DE FL GA HI IA ID IL IN KS KY LA MA MD ME MI MN MO MS MT NC ND NE NH NJ NM NV NY OH OK OR PA RI SC SD TN TX UT VA VT WA WI WV WY


Avg. Housing Price by State (Rank Size < 50)
Housing Price
State Name
0
100000
200000
300000
400000
500000
600000
AL AZ CA CO CT DC FL GA IL IN KY LA MA MD MI MN MO NC NV NY OH OK OR PA RI TN TX VA WA WI

Correlation between Avg. Housing $ & Homeless Population (all rank size)

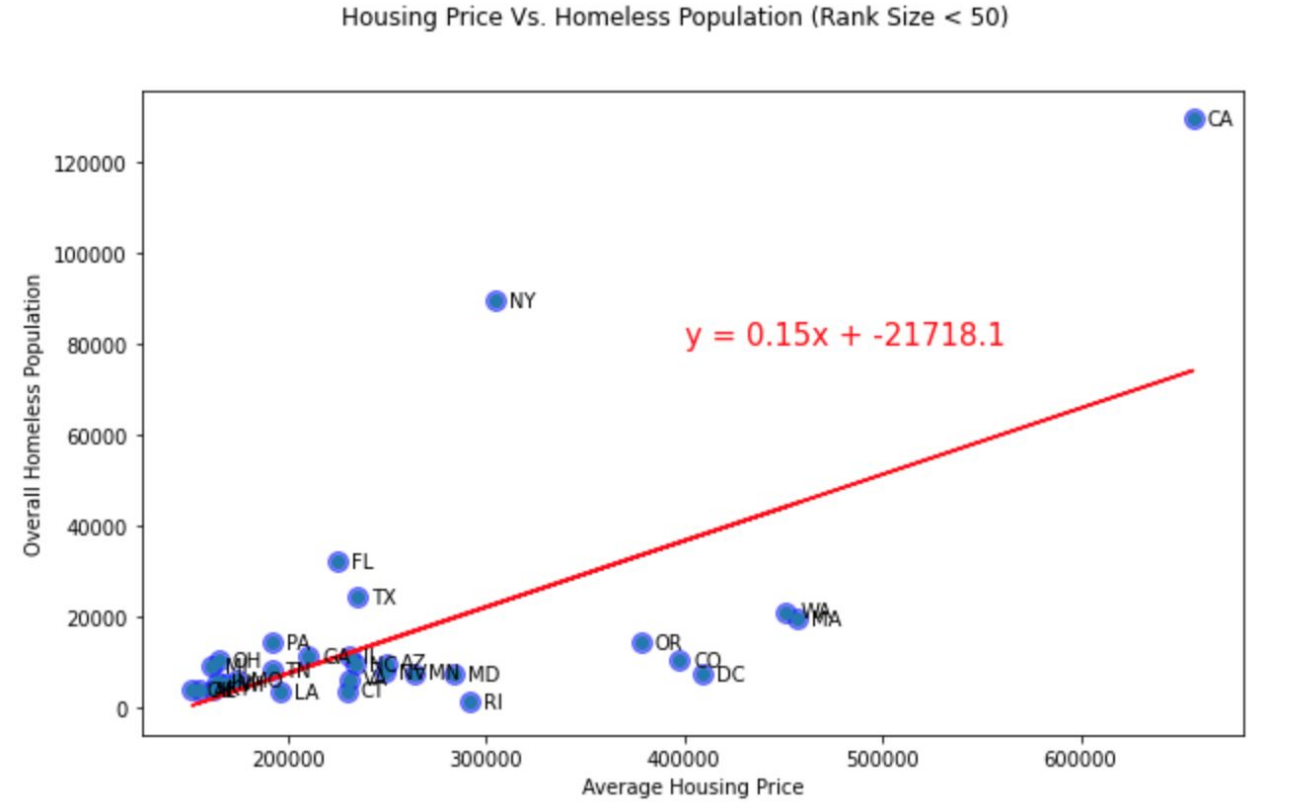
```
The correlation between average housing price and homeless population is 0.23
```



- Weak correlation
- CA and NY are really high in terms of Avg. housing price and and overall homeless population
- CA - weather is good. Silicon Valley. Subsidies program. good school, health care.
- NY - Mostly in City. Homelessness usually highest in the city / Metro area.

Correlation between Avg. Housing $ & Homeless Population (rank size < 50)

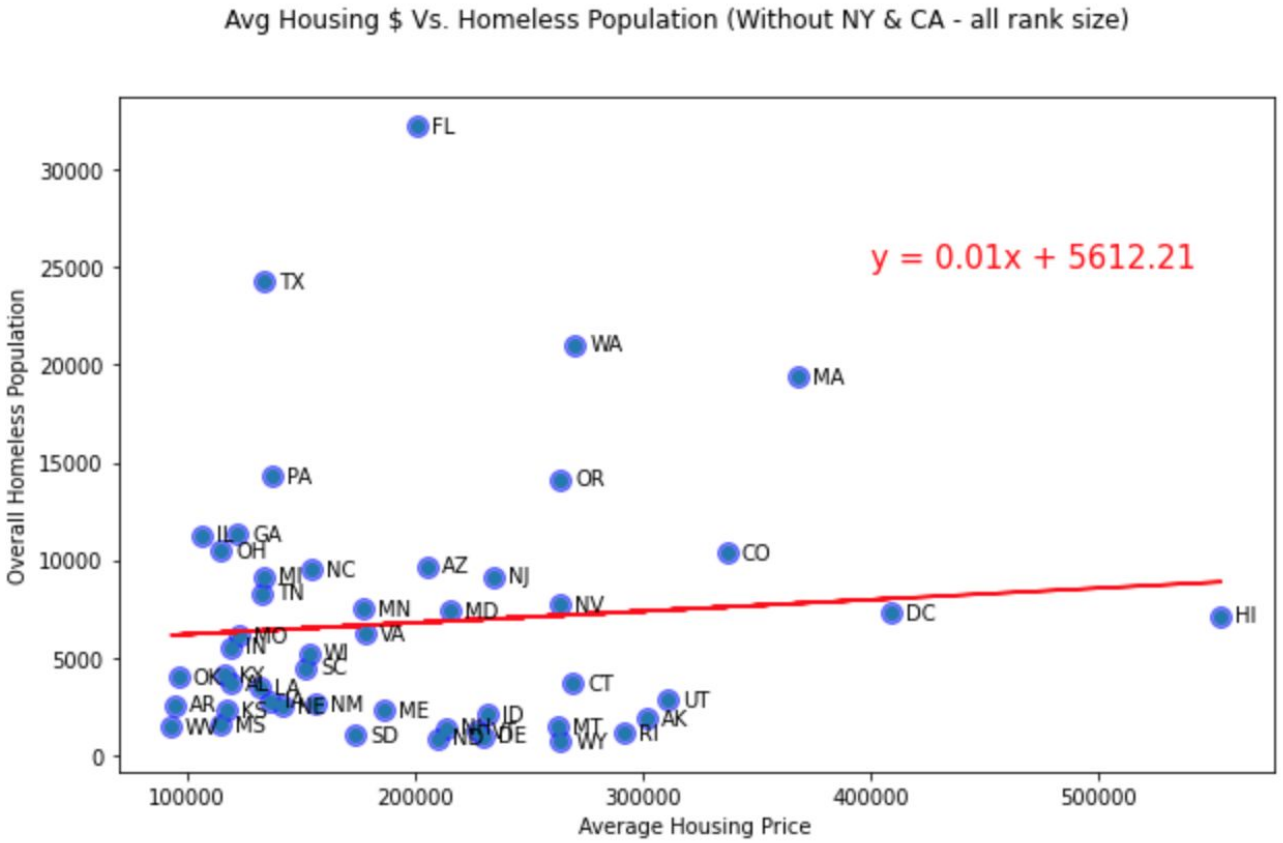
```
The correlation between average housing price and homeless population is 0.63
```



- NY cities (some area are not super expensive, mostly downtown)
- States under the has extreme weather/climate. Thus, when the homeless population is low, the housing price is still at high - still have a cluster of homeless population the avg housing of $400k.

[Without NY & CA] Correlation between Avg. Housing $ & Homeless Population (all rank size)

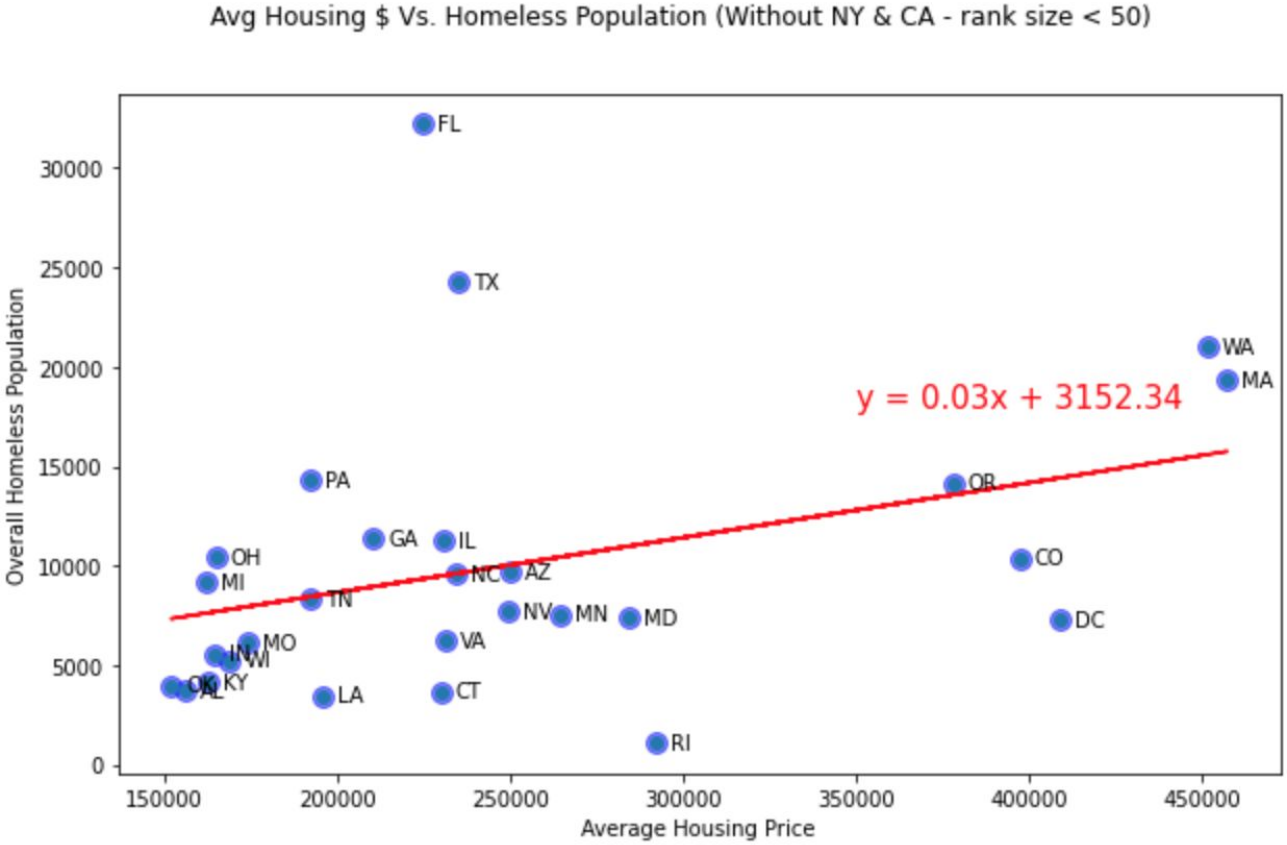
```
The correlation between average housing price and homeless population is 0.08
```



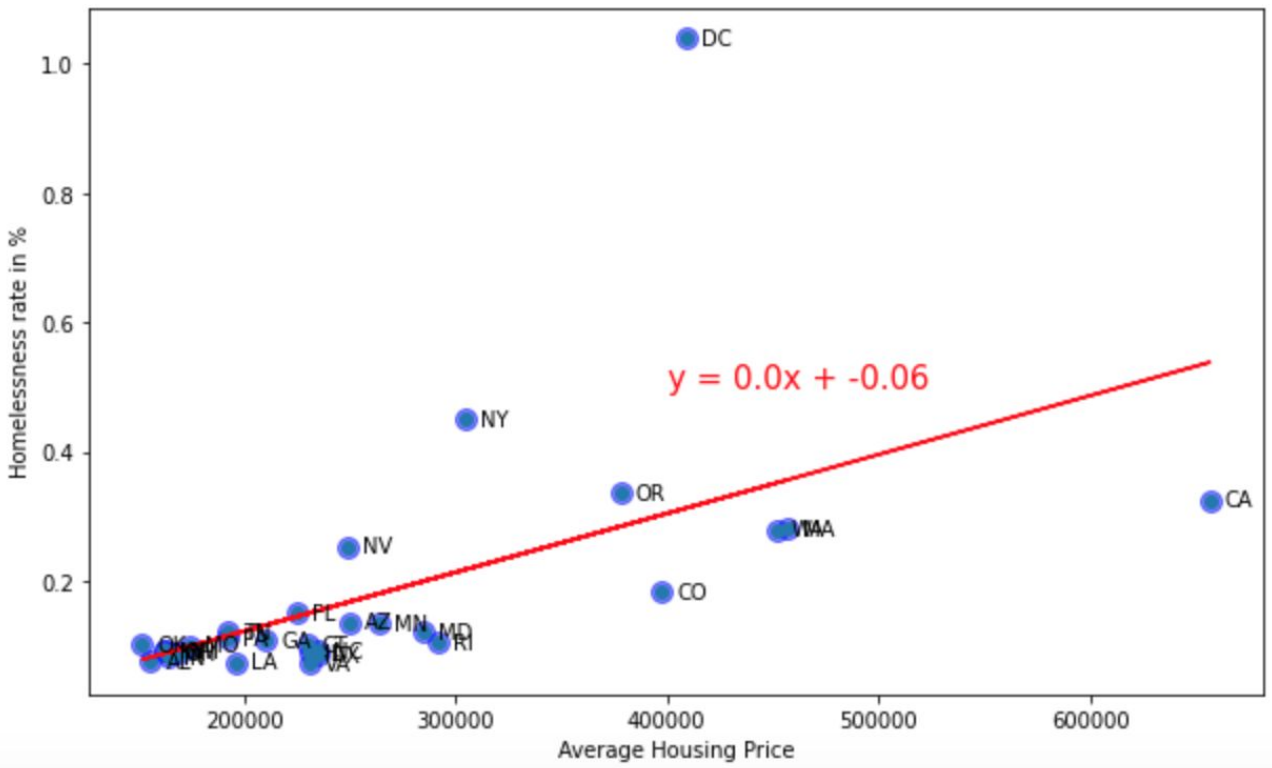
- This supported my previous statement. Homeless population has very low correlation with Avg. Housing price.

[Without NY & CA] Correlation between Avg. Housing $ & Homeless Population (rank size < 50)

```
The correlation between average housing price and homeless population is 0.36
```

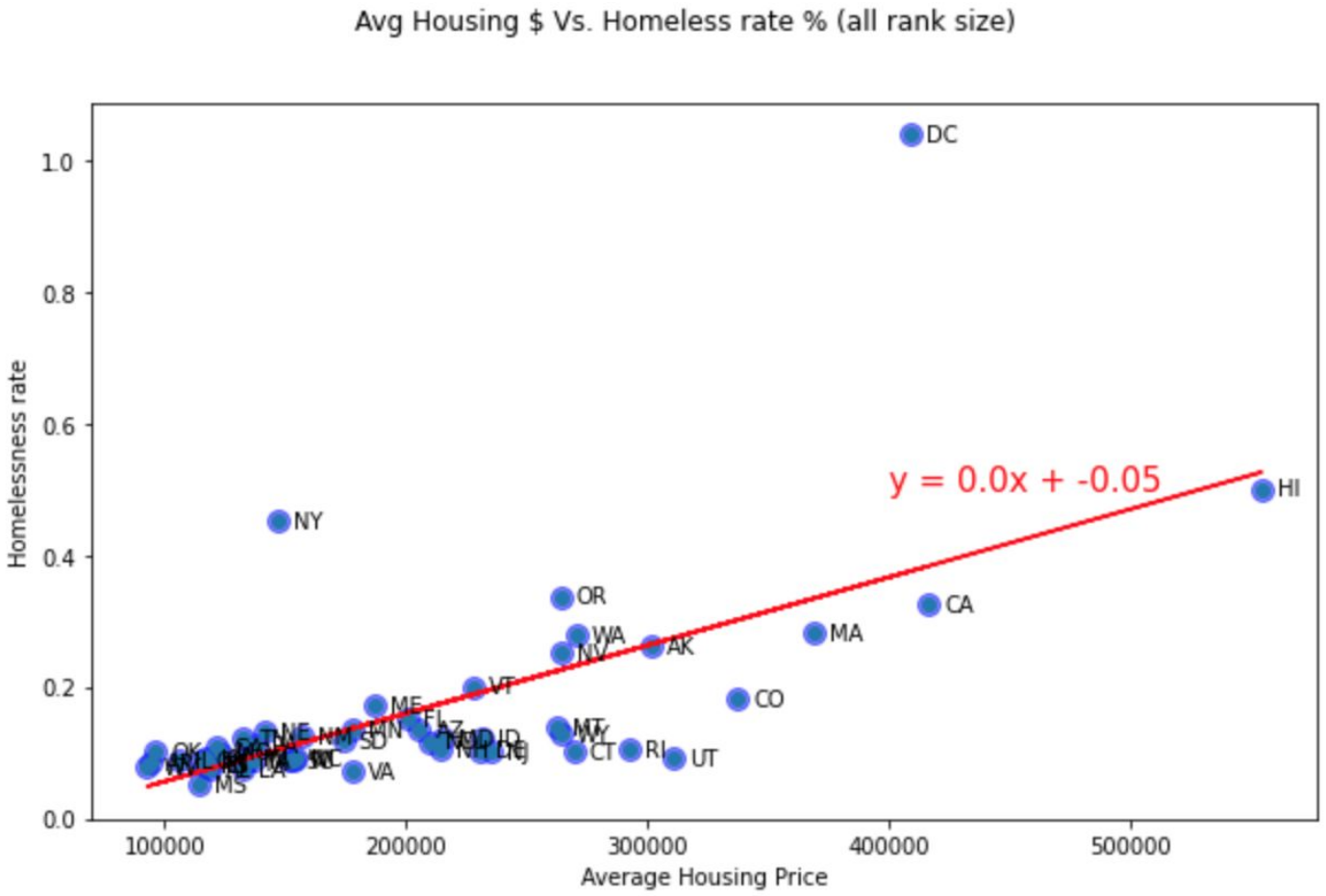


The correlation between average housing price and homeless rate is 0.56



- Homeless rate = Overall Homeless Population/Sta Population * 100
- DC is a city, homeless usu are in the city
-

The correlation between average housing price and homeless rate is 0.63



Limitation

- Data is the representation of 35th to 65th percentile
- E.g. we know NY and CA housing price are high, however, the data only showing CA as having the most expensive housing. This could be that in the NY data, it only covers upstate NY areas.
- Rank Size - didn’t specify the criteria on what this is based on. Except the state city.
- Most of the homeless population are in the cities, but the data is only describing the data in state-wide.

Conclusion:

- Homeless population Vs. Housing price. Small correlation - not causation. Housing price higher, weather is nice. Expense of living.
- Some states has decreases in homelessness, however, there is an increasing number of homeless population. There could be other factors that contributed to higher homeless rate
- Homelessness highly correlation to the population size, cost of living. E.g. CA and NY has the highest homeless rate.
- National homeless population has increased since 2016 and spiked in 2019.