MathWorks Math Modeling Challenge 2024

York Community High School

Team #17932, Elmhurst, Illinois

Coach: Ella Sak

Students: Nick Cifelli, Timothy Flisk, Jeffrey Lu, Taylor Melind, Steven Piotrowski

M3 Challenge FINALIST—$5,000 Team Prize

JUDGE COMMENTS

Specifically for Team #17932—Submitted at the close of triage judging

COMMENT 1: You provide well-balanced responses to all three questions.

COMMENT 2: Awesome! It was a great report, and I enjoyed reading it.

COMMENT 3: Your executive summary is well written and includes a nice overview of the thought processes that were used. It would be helpful to the reader to see your graph of housing units extended to see your predicted values on the curve.

COMMENT 4: Nice discussion of the strengths and weaknesses of each model.

COMMENT 5: Incorporating transitional housing in your models was creative.

M3 Challenge 2024:

A Tale of Two Crises: *The Housing Shortage and Homelessness*

Team #17932

March 4th, 2024

Executive Summary

Dear Secretary of Housing Dr. Fudge,

As homelessness continues to be one of the most potent issues in contemporary America, policymakers need to be well-informed about the housing supply and the future of homelessness in our American cities. Starting our analysis of this complex issue, we began by creating a model to predict the number of housing units in a city in a given year using a sinusoidal regression with a linear midline. For Albuquerque, we predicted that there would be 279,500 houses in 2034, 310,000 houses in 2044, and 345,400 houses in 2074. For Seattle, we predicted that there would be 424,400 houses in 2034, 496,900 houses in 2044, and 684,300 houses in 2074. Our model proved highly accurate to the data provided, which gives us high confidence in these projections of housing supply.

Our next step in the analysis was to create a model that could predict changes in the homeless population of a city over a time interval of 10 years, 20 years, and 50 years. We used parameters of Seattle and Albuquerque which we had in our data, median income, housing price, and population, to create a normal distribution of yearly income for our selected cities. By deciding on an income cutoff for homelessness based on the price of housing units, we could use the NormalCDF function to find the percentage of the population which are homeless based on the area of our normal curve below the cutoff. We then created linear regressions for all of our inputs to model how our cutoff and normal distribution evolve to model the percentage of the population that is homeless, we then used another regression to model the population change which we multiplied by our derived percentage to get a final number of estimated homeless people in each city per year. Ultimately, we estimated that both Seattle and Albuquerque will see a progressive increase in their respective homeless population.

Finally, we synthesized our findings to create a comprehensive plan for what Seattle should do to address homelessness in the long term. We first found that one of the biggest indicators of increased homelessness in Seattle was a decrease in transitional housing. So, we planned to construct enough transitional housing so that unsheltered homeless people can get back on their feet. We then calculated the price of this project over time with the changing number of unsheltered homeless and we came up with price figures that the city would need to invest 273,900,000 dollars now and 889,600,000 dollars over the next 50 years. We also calculated the maintenance costs over these periods so that you can make reasonable estimates of how much this is going to cost. Overall, we concluded that this solution is completely viable and much cheaper than any other alternative.

We believe that the data and solutions that we have compiled in this paper will be exceedingly useful in your policy decisions and that we have provided you with a secure, effective, plan to help amend Seattle's homeless crisis. We hope that this data helps inform your policy to an effective solution. Best of luck,

Team 17932.

Table of Contents

1. Global Assumptions

1. Homelessness is directly related to the economy, which is known to be periodic with an effectively constant frequency.

Obtaining a residence is guaranteed to be among the top priorities of the vast majority of individuals, so homelessness is primarily a function of an individual's capacity to obtain a home. This is directly related to the state of the economy: during a "bear" market, a greater portion of the population will not be in sufficient financial standing to purchase or begin making payments on a home; during a "bull" market, a lesser proportion of the population will be in the previously described position. Economists typically use a 10-year period to generalize the market cycles of "bull" and "bear" markets, so we assume that homelessness will follow a similar pattern.[4]

2. Historical data is predictive of future trends

Despite rapid innovation (or lack thereof), homelessness and other metrics primarily influenced by the economy tend to follow the same patterns over long periods of time. Furthermore, homelessness is, unfortunately, not among the top priorities of the most influential world leaders due to preoccupation with more pressing issues. Therefore, we assume that the trends we identify in past data will be predictive of the future.

3. The housing market will not reach its theoretical limit during our time period

Housing is a fundamental human necessity, so it's rational to assume the overall demand trends for housing with respect to population will be sustained. Furthermore, due to the prices of housing rising,[1] we can assume that the market cap for housing is not being approached due to demand increasing

Question 1: Analyzing the housing market

1.1 Defining the Problem

We aim to model the housing supply 10, 20, and 50 years in the future in Seattle, Washington, and Albuquerque, New Mexico. We define housing supply as the total amount of housing units in each city, and we used the provided data on housing unit totals.

1.2 Assumptions

1.2.1 *Available space is not a significant constraint for housing availability*

Many cities, including the likes of Albuquerque and Seattle, have resorted to using vertical expansion[5] (ex. skyscrapers) to obtain sufficient usable space for their constituents, space restrictions do not significantly influence housing availability. Furthermore, the continued innovation in housing development makes it possible to continue expanding housing supply as needed.

1.2.2 *The average longevity of housing will not drastically change*

Because there are no incentives to shift the longevity of housing, we assume that there will be no major shifts in the average longevity of houses. This is a fair assumption because the demolition of houses is largely driven by demand, not actual physical deterioration of the property.[3]

1.3 Variables

Name	Description	Units
t	Time, in years, after 2010	Years
m	Slope of the linear midline of our data	Houses per Year
β_0	The initial value of the linear midline	Houses
α	The amplitude of the sinusoidal term	Houses Per Year
ω	The frequency of the sinusoidal term	Years^{-1}
ϕ	Phase shift of the sinusoidal term	Dimensionless

1.4 The Model

Examining the available units of housing, we can immediately see a roughly linear increase. However, this doesn't consider the periodic fluctuations present in data dependent on economic factors. We decided to use a sinusoidal regression with a linear midline as our model. Sinusoidal regressions like this are often used to predict data that has periodic increases and decreases in growth rate, especially over mid-length periods such as those we aim to predict. We chose this model structure because, as described in our global assumptions, the data appears to be periodic over 10 years. Note that the use of a sine term over a cosine term does not affect the quality of our model's fit because the phase shift term gives each the same modeling capacity.

To find the optimal parameters, we minimized the mean squared error of the predictions on the housing unit availability data[2] for both Albuquerque and Seattle.

Our general function is $f(t) = mt + \beta_0 + \alpha\, sin(\omega t + \phi)$

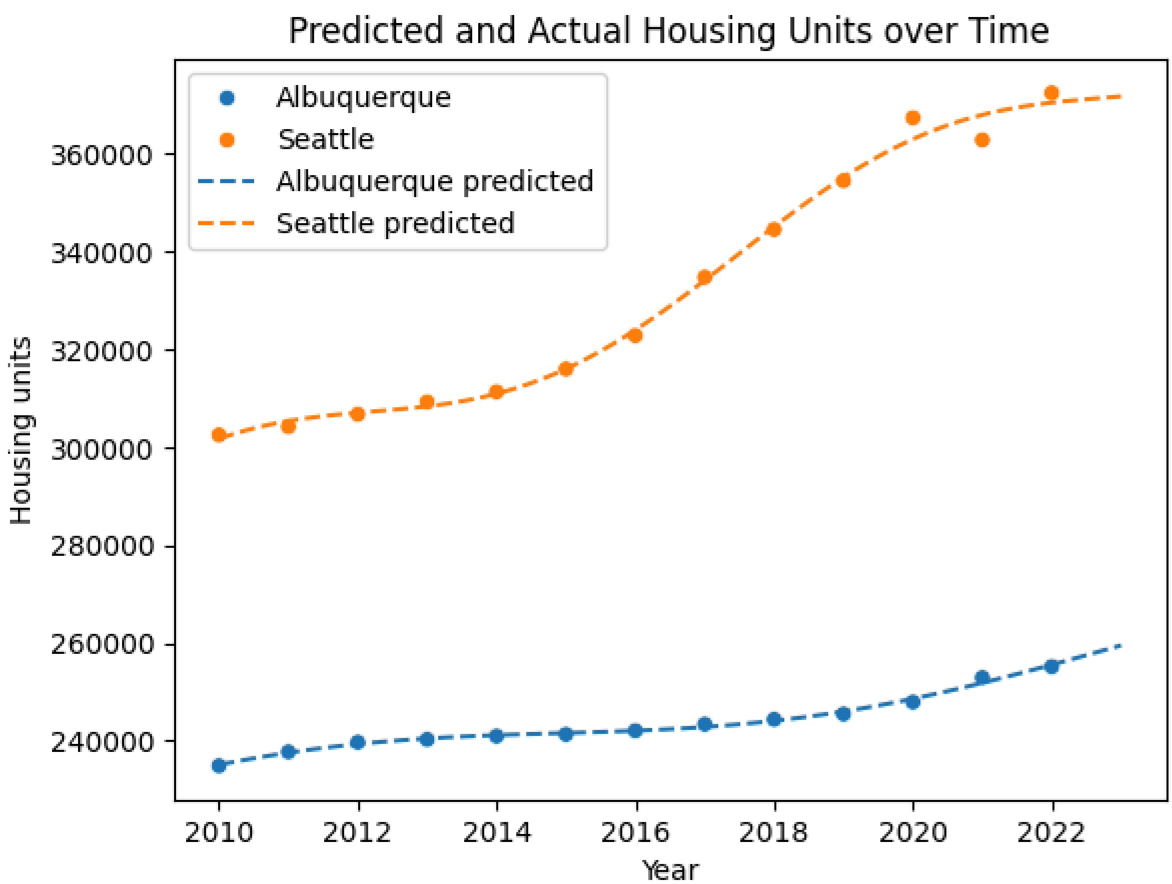
Albuquerque	$f_A(t) = 2205.8t + 230758 + 4604.53 \sin(0.3901t + 1.232)$
Seattle	$f_S(t) = 6108.7t + 293665 + 8251.45 \sin(0.5986t + 1.807)$

These functions returned the following values for 10 years from now (t = 24), 20 years from now (t = 34), and 50 years from now (t = 64)

Years since 2024	Albuquerque	Seattle
10 years	279,500 houses	424,400 houses
20 years	310,000 houses	496,900 houses
50 years	345,400 houses	684,300 houses

1.5 Discussion

Our model predicts that housing units will continue to grow, and almost double in 50 years. This matches the growing population rate, combined with periodic changes in growth with the bull and bear markets over 10 years.



Outside of the usefulness of our function in modeling the periodic nature of the housing supply, we also used the R^2 values calculated by desmos as .9936 for Seattle and .9928 for Albuquerque to evaluate the effectiveness of our function for approximating the current data points. The R^2 value being extremely close to 1 shows that our function closely approximates the data points given, and due to it fitting both the increasing nature of housing supply and the period nature of growth and decline, we are confident that it can approximate 10 and 20 years into the future, and somewhat confident about its prediction 50 years into the future. We also analyzed the residual graphs of our function: these show that our data is fairly scattered with regards to the residual and there is no leftover curve pattern. Additionally, our model also has extremely low values for the residual throughout.

1.6 Sensitivity Analysis

We checked both of our model predictions for 2022 to compare with the actual data which was the most recent data we were given.

For Albuquerque: f(12) = 2205.81(12) + 230758 + 4604.53sin(0.3901(12)+1.2315)

f(12) = 254,793 homes as compared to the actual value of 255178 homes

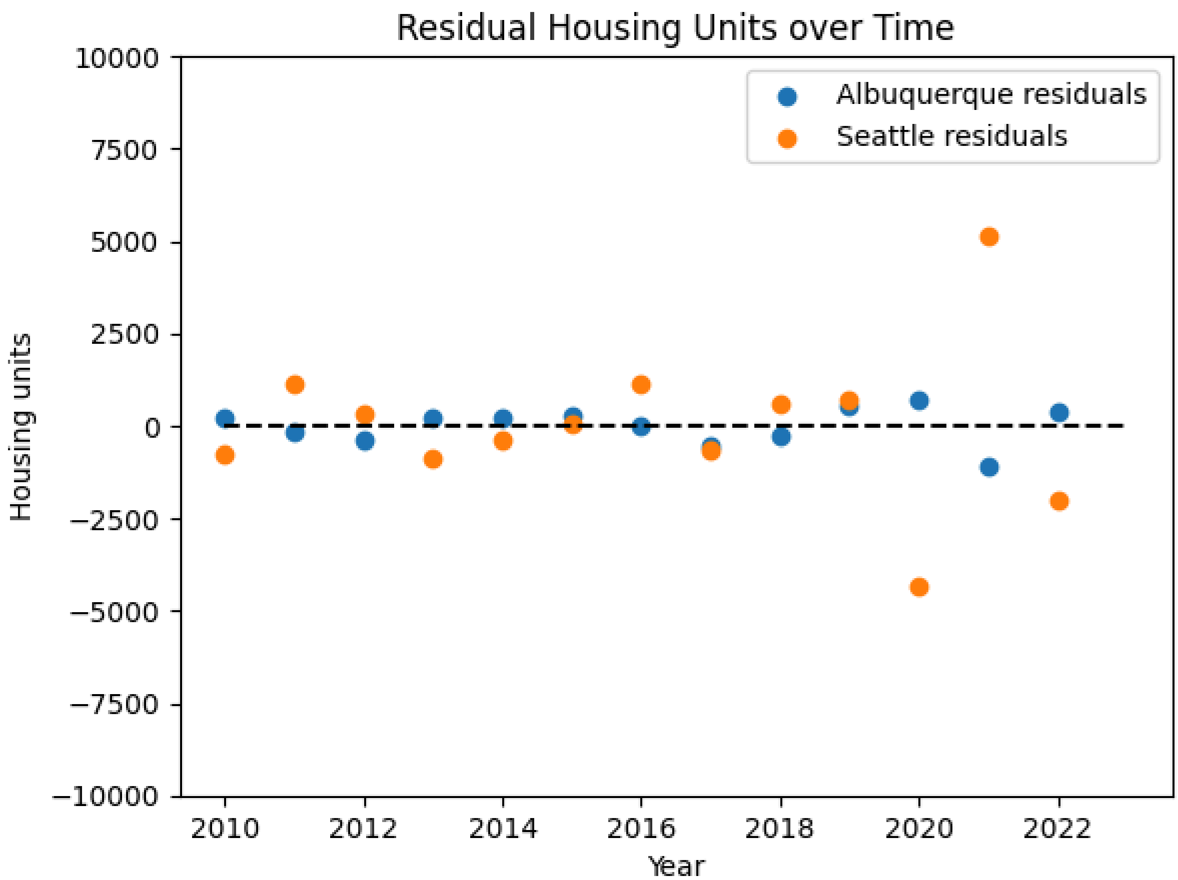
Using the percent error formula $P.E. \;=\; \frac{|Predicted - Actual\,|}{Actual} \;*\; 100$

We derived a percent error of **0.15%**

For Seattle: f(12) = 6108.7(12) + 293665 + 8251.45sin(.59856(12)+1.807)

f(12) = 374430 homes as compared to the actual value of 372436 homes

Using the percent error formula we derived a percent error of **0.54%**



The residual graph shows that the data points make no discernable pattern over the residual, showing that the sinusoidal regression effectively models the relationship between time and housing supply.

1.7 Strengths and Weaknesses

Strengths:

- *Our model is periodic, and thus can predict how the current time will affect the housing market*

This is very important for the 10 and 20 year predictions, where failing to take this periodicity into account can cause an error of up to 50,000 units.

- *Our model has a consistent rate of growth for total houses*

  This aligns with how many more units of housing are created than destroyed every year, regardless of economic period.

- *Our model has a random residual plot with no clear trends*

  If a model has a non-random residual plot or a clear trend, that implies that there is a better model to explain the data. Since our model did not have a leftover curve pattern inside of our residual plot, the randomness suggests that it was a strong fit.

Weaknesses:

- *Our model does not take into account land restrictions or if the market cap for the housing market is being approached*

  Although the current data suggests that the market cap has not been reached and will not be reached soon, this makes our confidence for our prediction for 50 years in the future less than that of the next 10 and 20 years.

- *The timeframe of our data is shorter than our predicted period for our periodic regression*

  This means that if some event we do not have context for occurred during this cycle that would alter it from a typical cycle, we would be unable to predict a "normal" cycle.

Question 2: A look at the homeless population

2.1 Defining the Problem

Problem Statement:

For the regions you chose in Q1, predict changes in the homeless population in the next 10, 20, and 50 years.

We interpreted changes in the homeless population as an estimated amount of people declared homeless, regardless of status as sheltered or unsheltered.

2.2 Assumptions:

2.2.1 *Regressions for income, normal distribution, housing price, and standard deviation are approximately linear*

For the data given, a linear approximation offers the most accuracy while still being efficient to compute over our time period in comparison to other common models, such as exponential and polynomial growth (which fails to account for growth limiting factors along longer timescales), or sinusoidal and logarithmic growth (which fail to account for a continued increasing rate of growth).

2.2.2 *Yearly income follows a normal distribution with its mean being the yearly median income.*

Income is typically modeled to be a skewed log-normal distribution,[16] however, in the case of both Seattle and Albuquerque, the difference between income being modeled as a log-normal distribution and a normal distribution is insignificant, thus using a log-normal distribution would only obfuscate our data and make computations significantly more complex.

2.2.3 *Maximum income cutoff for homelessness is defined as the median housing unit price divided by 30.*

For home mortgages, a 30-year length is the most common plan length[8]. For this reason, we are going to assume that if someone can't afford a median-priced housing unit using their entire income over a 30-year mortgage plan, they are considered homeless.

2.2.4 *The 2021 data was invalid due to a lack of data for unsheltered homeless*

Many cities did not collect data about unsheltered homelessness in 2021 and if they did collect some data, it was still incomplete[9] (like Seattle). Following this, we tried to exclude the skewed 2021 data from the Seattle model when possible.

2.3 Variables:

Name	Description	Units
N	Output of our modeling function	Number of homeless people
p	Population	Number of people
I	Median Income	Dollars
h	Median Housing price (Adjusted for inflation by converting to 2022 buying power)	Dollars

s	Standard Deviation of income	Dollars
t	Time since 2010	Years

2.4 The Model:

We first started with trying to find an applicable regression model for homelessness, but due to the data not fitting any common model, we instead decided to try and model homelessness as a function of housing price and median income, as access to affordable housing is considered the most influential predictor of homelessness.[6] We estimated the most affordable housing option for a given person or family to be the Median Household Income divided by 30 (see assumptions). Thus, we needed to find the portion of individuals with income below this and multiply that portion of individuals by the population at the given time to obtain the number of homeless people.

We decided to use a normal distribution to approximate incomes, as income in this scenario follows an approximately normal distribution (see assumptions). Our mean is already given to us for this distribution, as the median income. However, we had to find a way to model our standard deviation.

We decided to compute the standard deviation for each year using our current model and data, setting our N function to the given homeless population for a year, and using the inverse normal function to calculate our standard deviation for each year. We then used our table of standard deviations for each year to create a linear regression model for future standard deviations given a year.

Thus we created a function N(t), that found the homeless population in a given city t years after 2010.

This function is given as

$$N(p(t),\ I(t),\ h(t),\ s(t))\ =\ p(t) * NormalCDF(-\ \infty,\ \frac{h(t)}{30},\ I(t),\ s(t))$$

$$N(p(t),\ I(t),\ h(t),\ s(t))\ =\ p(t)\int_{-\infty}^{\frac{h(t)}{30}} \frac{1}{\sqrt{2\pi(s(t))^2}} e^{\frac{-(u-I(t))^2}{2(s(t))^2}}\, du$$

Where p(t) is the function for population change over time, I(t) is the function for median income change over time, h(t) is the function for median housing price change over time, and s(t) is the function for how the standard deviation of the income changes over time. For p, I, h, and s, we inputted the (cleaned) data given to us in the data table and found the linear regression for each. We computed the following function regressions for both Seattle and Albuquerque:

Function	**Regression**	**At t=24 (10 years)**	**At t=34 (20 years)**	**At t=64 (50 years)**	**R^2**
p(t), Albuquerque	2228.51(t)+539895	593,379.24	615,664.34	682,519.64	0.8324
p(t), Seattle	13406.4(t)+590255	912,008.6	1,046,072	1,448,263	0.9659
I(t), Albuquerque	1018.7(t) + 44478.2	68927.00	79114.00	109675.00	0.7816
I(t), Seattle	4457.56(t)+53233	160214.44	204790.04	338516.84	0.9286
h(t), Albuquerque	9279.92(t)+175549	398267.08	491066.28	769463.88	0.8179
h(t), Seattle	31881.6(t)+ 261495	1026653.4	1345469.4	2301917.4	0.9651
s(t), Albuquerque	275.621(t)+13671	20285.904	23042.114	31310.744	0.6855
s(t), Seattle	1711(t)+19756.7	60820.7	77930.7	129261.7	0.9088

Now, we will compute our N with these values:

City, years after 2024	**Estimated homeless population**
Albuquerque, 10 years	1,800 Homeless people
Albuquerque, 20 years	2,000 Homeless People
Albuquerque, 50 years	2,500 Homeless People
Seattle, 10 years	17,500 Homeless People
Seattle, 20 years	21,000 Homeless People
Seattle, 50 years	31,000 Homeless People

2.5 Discussion

Our model predicts that homelessness will increase in both cities, but Seattle will have a drastically larger homeless problem over the following years. Albuquerque's homeless population appears to be consistently a low overall portion of the population, suggesting that Albuquerque's homeless population does not pose a huge problem to the city. However, Seattle's homeless population is predicted to soon spiral out of control, suggesting the city must soon find a way to remedy its homeless epidemic. Our model also found that the rate of homelessness will grow in the short and mid-term, but eventually starts to decline, which matches up with what would be expected of the growth of a homeless population.

2.6 Sensitivity Analysis

We checked both of our model predictions for 2021 to compare with the actual data which was the most recent data we were given.

For Albuquerque: N(p(11), I(11), h(11), s(11)) = 1535

N(t=11) = 1535 homeless people as compared to the actual value of 1567 homeless people

Using the percent error formula $P.E. = \frac{|Predicted - Actual|}{Actual} * 100$

We derived a percent error of **2.04%**

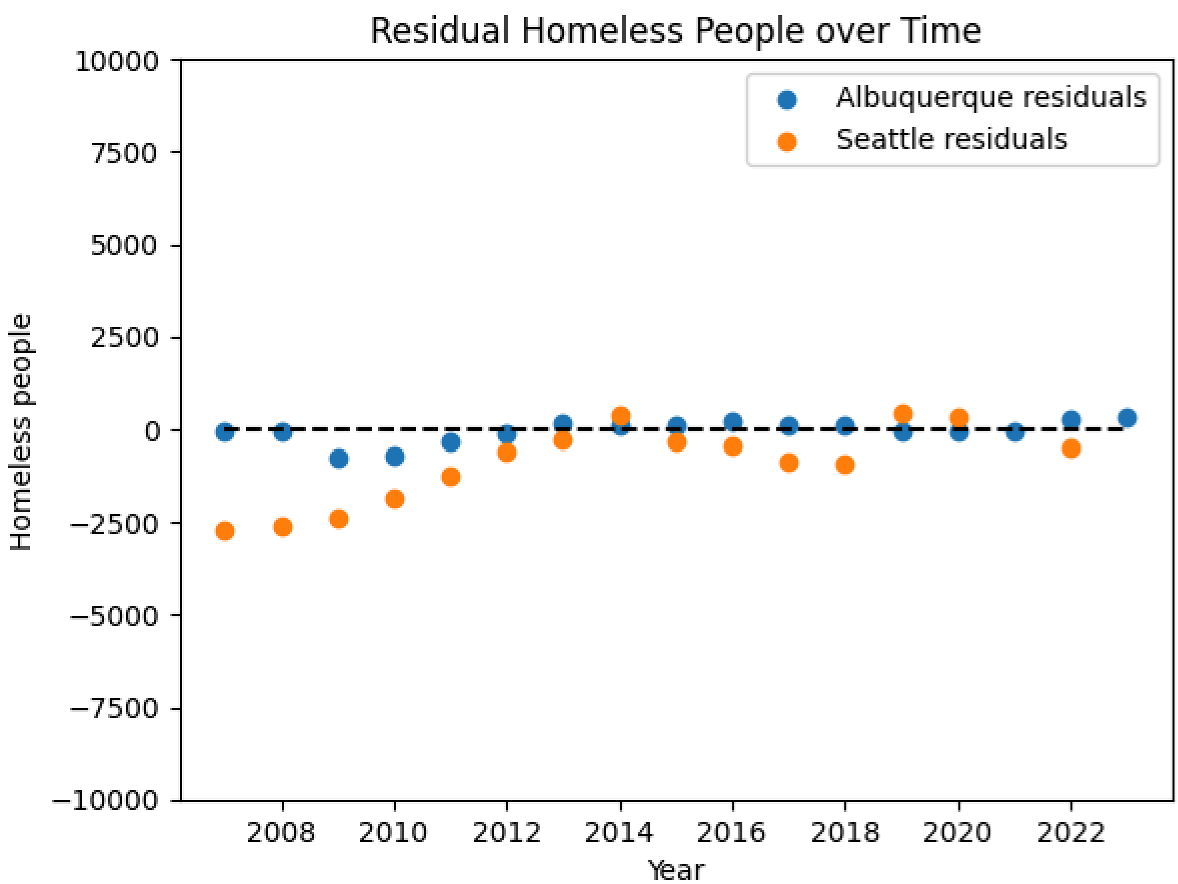
- We used the 2021 data point for Albuquerque instead of 2022, which we used for every other sensitivity analysis because the 2022 data for Albuquerque provided an outlier compared to the rest of the data trend.

For Seattle: N(p(12), I(12), h(12), s(12)) = 12895

N(t=12) = 12895 homeless people as compared to the actual value of 13368 homeless people

Using the percent error formula once again, $P.E. = \frac{|Predicted - Actual|}{Actual} * 100$

we arrived at a percent error of **3.54%**



When plotted the residuals appear to be mostly aperiodically oscillating with minimal magnitude. Other than an initial spike of error in Seattle predictions caused by an incorrect 2021 value, our data is mostly homoskedastic, validating our modeling assumptions. Overall, there is almost no discernible relationship between residuals and the time input, indicating that our model is mostly accurate.

2.7 Strengths and Weaknesses:

Strengths:

- *Incorporates multiple sources of information about population and income*

  By using multiple different variables to model our homeless population, we make our model more resistant to the effects of alterations in one variable.

- *High R^2 on the majority of regressions used for variables*

  Since a majority of our Linear Regression models for the functions had high R^2 values (Such as Seattle's population, median income, median household price, and standard deviation functions, with respective R^2 values of 0.9659, 0.9286, 0.9651, and 0.9088), our output is more accurate in the long run as it provides evidence of a true linear relationship.

Weaknesses:

- *Income is distributed log-normally with a right-skew, which our model fails to consider.*

  It is typically accepted in economics that the income distribution for a population follows a log-normal distribution[16} however for the sake of this paper and given the relatively small data set, it is well approximated by a normal distribution. Still, for complete accuracy, a log-normal distribution would have been superior to our method.

- *Our model had low R^2 values on some of the regressions.*

  The R^2 values for Albuquerque's Standard Deviation and Median Income functions were below .8 (0.6855 and 0.7816 respectively). This indicated that a linear regression may not provide a good representation of these values.

- *All factors must have the same type of regression*

  All of our regressions must have the same asymptotic growth rate, if we did not have this, our model would be skewed towards the functions with higher asymptotic growth rate greatly in the long run and would provide inaccurate results.

Question 3: Seattle rights its wrongs

3.1 Defining the problem

Using your results from the first two questions, create a model that expresses a long-term plan that would help a city combat homelessness. How adaptable is your model to unforeseen circumstances like natural disasters, economic recessions, or increased migrant populations?

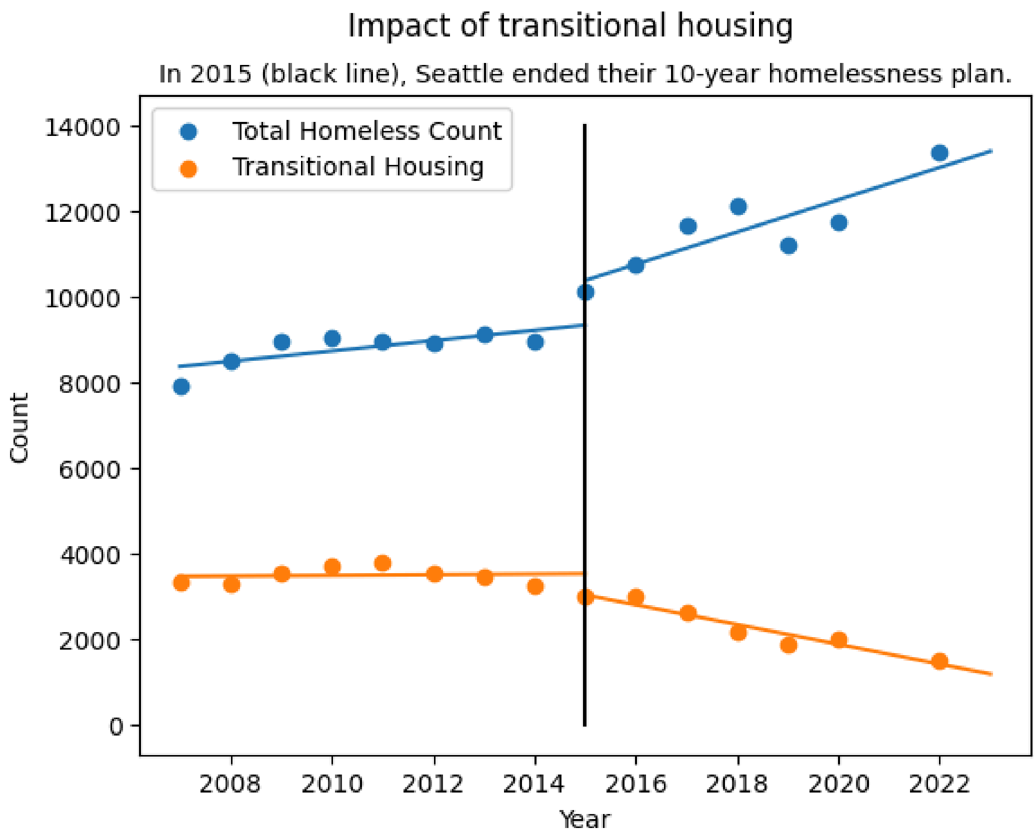
We interpreted this as finding a way to minimize homelessness rates. We also noted that the problem required only one city to focus on, so we decided to analyze Seattle, as from problem 2, Seattle has a larger homeless problem both now and in the future.

3.2 Assumptions

3.2.1 A decrease in transitional housing is correlated with an increase in homelessness

Starting in 2005, the city of Seattle implemented a "10 Year Plan" to end homelessness.[17] This is evident in an initial visual analysis of Seattle data, which shows an increase in homelessness around 2014-2016 was tied to a decrease in transitional housing.

To validate this assumption, we implemented a regression discontinuity design splitting along 2015, which is the year that the program ended.



Prior to the discontinuation of the program, the city maintained relatively constant levels of transitional housing and a slow rate of increase in homelessness. When the program ended and the transitional housing was allowed to decrease, the rate of homelessness immediately increased at a significantly greater rate. This indicates that transitional housing, a core component of the program, does indeed suppress the growth of homelessness.

3.2.2 The number of sheltered homeless people remains constant

We assume that the shelter capacity for homeless people in a given city remains constant throughout time, and that these facilities will always be at maximum occupancy due to their high demand. This assumption is supported by the data, which shows a largely constant number of sheltered homeless people that randomly varies without showing any trend up or down.

3.2.3 90% of homeless people will take advantage of transitional housing

In our research, we determined that 90% of a sample homeless population in Denver accepted housing when given the opportunity.[13] Since this trend has been observed in other cities, we thought it reasonable to infer that Seattle would follow this trend closely.

3.3 Variables

Name	Description	Units
C	Cost of the construction	Dollars
C_0	Cost of new housing construction per square foot	Dollars per ft^2
U	Amount of Unsheltered Homeless people	Number of people
S_T	Amount of sheltered homeless people	Number of people
T	Amount of necessary square feet of transitional housing	ft^2 per person
t	Years since 2010	Years
M	Maintenance Cost per year	Dollars per year

3.4 The Model

$$U(t) = N(p(t), I(t), h(t), s(t)) - S_T$$

We assume that the amount of sheltered homeless people per year remains constant (not counting the homeless population that becomes sheltered as a result of our transitional housing program), and Use our model to predict the homeless population in Seattle from Question 2 to model our unsheltered homeless population by subtracting our constant S_T from it. We calculated S_T by finding the mean value of the sheltered homeless population for every year since 2007 in Seattle. We justified this by the fact that there seemed to be no trend in the sheltered population and that it hovered around our value of $S_T = 6040$

Then, we can calculate the amount of transitional housing needed now, in 10 years, 20 years, and 50 years for Seattle's homeless population. We estimate that 90 percent of Seattle's homeless population would willingly enter transitional housing, and thus estimate our units needed at .9U. We decided that 150 square feet for each unit of Transitional housing is an adequate amount of space for each person, and found our value of cost per square foot to be 264$.[14]

More generally, the cost to construct transitional housing for U homeless people at a given point in time would be given as

$$C(t) = \frac{9}{10}U(t)C_0T$$

Thus, We find we need to employ **273,900,000** dollars to build our transitional housing now, **408,400,000** dollars in total over the next ten years, **533,200,000** dollars in total over the next twenty, and finally **889,600,000** dollars in total over the next fifty years, without adjusting for inflation. While this seems like a lot, if we construct the first set now, then we can subtract that cost from the future one. When phrased like this, it can only really take three 300-million-dollar projects to completely destroy the homeless problem in Seattle for 50 years. Compared to other government housing projects which sometimes run into the billions and fail to adequately address homelessness this seems like a good solution.

Next, we need to calculate our costs of maintenance: The cost of maintenance on housing is approximately 1.6$ per square foot,[15] so our maintenance cost per year is this value multiplied by the total square footage of our complex. So our total cost per year at a given time if we have transitional housing for all homeless people is

$$M(t) = \frac{9}{10}(1.6)U(t)(150)$$

With our two functions, Seattle would need to place \$2,475,360.00 in yearly upkeep after 10 years, \$3,231,360.00 per year after 20 years, and finally \$5,391,360.00 per year after 50 years.

3.5 Discussion

This strategy to provide enough transitional housing for every homeless person willing to use it would clearly be cost-effective for the government. Seattle has done many projects that are much more expensive such as Sound Transit 3 which is projected to cost upwards of 54 billion USD upon completion.[7] This plan to counter homelessness is barely a drop of water compared to this and it will have a large impact. Given the trend that we assumed from the data, and that it makes sense that homelessness would decrease if homeless people were given more options, we think that this plan would be incredibly cost-effective and purely effective for the city of Seattle.

Our strategy is adaptable to some, but not all unforeseen circumstances. Economic recession would not greatly affect our plan, as it is cost-effective, especially because the greatest costs are used up-front. Our plan is also well adapted for severe weather conditions such as blizzards or hot summers, as the worsened outside conditions would cause more homeless people to enter transitional housing. However, our strategy flounders with migration and natural disasters, as our transitional housing may be damaged by natural disasters, and the rapid influx of homeless population that follows a migration would not be well-suited to our model.

3.6 Strengths and Weaknesses

Strengths:

- *Our model recognizes Seattle's previous shortcomings in reducing transitional housing*

  Our rationale for increasing transitional housing is directly based on our data, which showed a direct correlation between decreases in transitional housing and increases in the homeless population for Seattle. This provided us with strong reasons to believe that increasing transitional housing availability would lead to a decrease in homelessness.

- *Our model is effective in creating employment opportunities for Seattle residents*

  Any major construction program will generate numerous jobs for local artisans in construction and design. It also offers employment for people to help the residents of transitory housing, and finally finds employment for those in transitory housing.

Weaknesses:

- *Fails to address issues with drug addiction and mental illness*

  Homeless people struggling with substance use and mental health issues may be unable to reside in transitional housing for extended periods of time.

- *The model requires voluntary compliance to transitional housing*

  From our data we made the assumption that 90% of all homeless people offered housing would accept it, however the other 10%, for various reasons (by choice, addiction, mental illness, distrust of government programs, or other obligations) will not be able to move into housing, and may remain unhoused. Also if for some reason less than 90% of the unsheltered homeless population was willing to move to a transitional shelter, then we would have to reconfigure our plan to adjust.

Conclusion:

For the first question, we modeled the changes in housing supply in Albuquerque and Seattle over the next 10, 20, and 50 years using a sinusoidal regression with a linear midline. For Albuquerque, we estimate that there will be 279,500 houses in 2034, 310,000 houses in 2044, and 345,400 houses in 2074. For Seattle, we estimate that there will be 424,400 houses in 2034, 496,900 houses in 2044, and 684,300 houses in 2074. We evaluated the effectiveness of our functions by analyzing their R^2 values and their residual plots and are highly confident in our models. Next, we modeled the change in the number of homeless people in the next 10, 20, and 50 years in the same cities using a normal approximation of the income distribution with our upper bound set as the median home price divided by 30. For

Albuquerque, we estimate that there will be 1800 homeless people in 2034, 2000 homeless people in 2044, and 2500 homeless people in 2074. For Seattle we estimate that there will be 17500 homeless people in 2034, 21000 homeless people in 2044, and 31000 homeless people in 2074. Finally, we synthesized the data that we had collected and created so far to create a plan for Seattle to combat homelessness in the long run. We decided that Seattle should target increasing transition housing because we had noticed a correlation between homelessness increasing when transition housing decreased. We created a cost-effective solution where Seattle could provide transitional housing to 9 out of 10 unsheltered homeless people for as little as 274 million dollars today, and over the next 50 years it would cost the city less than 900 million in total.

The life of a homeless person anywhere is not easy, but the USA, as the land of the free, promises to strive for everyone to have the right to life, liberty, and the pursuit of happiness. However, when one is homeless, to simply continue living is a struggle every day. This is why as policymakers, but also as Americans, it is imperative that we work to improve the conditions of the homeless. We hope the data and conclusions found here can help improve the lives of thousands of unhoused individuals across America.

Bibliography

1:

Housing is now unaffordable for a record half of all U.S. renters, study finds

https://www.npr.org/2024/01/25/1225957874/housing-unaffordable-for-record-half-all-u-s-renters-study-finds

2:

There's a massive housing shortage across the U.S. Here's how bad it is where you live

https://www.npr.org/2022/07/14/1109345201/theres-a-massive-housing-shortage-across-the-u-s-heres-how-bad-it-is-where-you-l

3:

When a house is demolished, more than the home is lost

https://theconversation.com/when-a-house-is-demolished-more-than-the-home-is-lost-42579

4:

Bull and Bear Market Cycles

https://srbadvisors.com/bull-and-bear-market-cycles/

5:

Downtown Albuquerque could see tallest building constructed in decade

https://www.krqe.com/news/albuquerque-metro/downtown-albuquerque-could-see-tallest-building-constructed-in-decade/

6:

Homelessness Is A Housing Problem

https://books.google.com/books?id=guxcEAAAQBAJ&printsec=frontcover#v=onepage&q&f=false

7:

The economics of Sound Transit 3

https://www.seattletimes.com/business/economy/the-economics-of-sound-transit-3/

8:

The Average Mortgage Length In The U.S.

https://www.rocketmortgage.com/learn/average-mortgage-length#:~:text=The%20most%20common%20amount%20of,length%20is%20under%2010%20years

9:

Why 2021 homeless counts may be incorrect

https://www.npr.org/2021/01/18/957379320/for-many-areas-count-of-homeless-population-is-canceled-or-delayed

10:

The hottest trend in U.S. cities? Changing zoning rules to allow more housing

https://www.npr.org/2024/02/17/1229867031/housing-shortage-zoning-reform-cities

11:
https://nap.nationalacademies.org/catalog/25133/permanent-supportive-housing-evaluating-the-evidence-for-improving-health-outcomes **(National Academies Press)**

12:

A Tale of Two Crises, MathWorks Math Modeling Challenge 2024, curated data,

https://m3challenge.siam.org/kdfrldh/

13:

Dismantling the Harmful, False Narrative That Homelessness Is a Choice

https://www.urban.org/urban-wire/dismantling-harmful-false-narrative-homelessness-choice

14:

Annual Affordable Housing Cost Data: Report to the Washington State Legislature

https://app.leg.wa.gov/ReportsToTheLegislature/Home/GetPDF?fileName=HousingCostDataReport2020_1e1a4d5b-0014-48d1-93c9-89bb15a6fafc.pdf

15:

How much does building maintenance cost?

https://www.consultengsurvivor.com/the-cost-of-building-maintenance

16:

Log-Normality of income distribution

https://www.aier.org/article/income-distribution-and-inequality/#:~:text=Italian%20economist%20and%20engineer%20Vilfredo,infinity%20with%20a%20central%20mean.

17:

Seattle's 10-Year-Plan to End Homelessness

http://www.cehkc.org/index.html

Code Appendix

Question 1

```
# Imports
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.optimize import curve_fit


# Load data
adf = pd.read_csv("data/alb.csv", thousands=',')
sdf = pd.read_csv("data/sea.csv", thousands=',')

adf_temp = adf.copy()
sdf_temp = sdf.copy()
adf_temp["City"] = "Alb"
sdf_temp["City"] = "Sea"

df = pd.concat([adf_temp, sdf_temp], ignore_index=True)


# Prepare data for model
adf_q1 = adf[["Year", "Housing units"]].dropna()
sdf_q1 = sdf[["Year", "Housing units"]].dropna()


# Define model function
def q1(t, m, beta_0, alpha, omega, psi):
   return m*(t-2010) + beta_0 + alpha*np.sin(omega*(t-2010) + psi)


# Optimize all parameters under given constraints (improves fit efficiency and
quality)
bounds = (0, [10000, 500000, 10000, 1, 3])
a_params, a_cov = curve_fit(q1, adf_q1["Year"], adf_q1["Housing units"],
bounds=bounds)
s_params, s_cov = curve_fit(q1, sdf_q1["Year"], sdf_q1["Housing units"],
bounds=bounds)


# Plot values
sns.scatterplot(df, x="Year", y="Housing units", hue="City")

x = np.linspace(2010, 2023, 100)
ay = np.array([q1(i, *a_params) for i in x])
sy = np.array([q1(i, *s_params) for i in x])

plt.plot(x, ay, label="Albuquerque predicted", linestyle="--")
plt.plot(x, sy, label="Seattle predicted", linestyle="--")

plt.title("Predicted and Actual Housing Units over Time")
plt.legend()
plt.show()


# Plot residual values
ares = adf.apply(lambda row: q1(row["Year"], *a_params) - row["Housing units"],
```

```
axis=1)
sres = sdf.apply(lambda row: q1(row["Year"], *s_params) - row["Housing units"],
axis=1)

plt.plot([2010, 2023], [0, 0], c="black", linestyle="--")

plt.scatter(adf["Year"], ares, label="Albuquerque residuals", c="C0")
plt.scatter(sdf["Year"], sres, label="Seattle residuals", c="C1")

plt.title("Residual Housing Units over Time")
plt.legend()
plt.xlabel("Year")
plt.ylabel("Housing units")
plt.ylim(-10000, 10000)

plt.show()
```

Question 2

```
# Imports
import numpy as np
import pandas as pd
from scipy.stats import norm
from sklearn.linear_model import LinearRegression
import matplotlib.pyplot as plt


# Load data
adf = pd.read_csv("data/alb.csv", thousands=',')
sdf = pd.read_csv("data/sea.csv", thousands=',')


# City class
class City:
   def __init__(self, df_):
       self.df_ = df_[["Year", "Population", "Median income", "Median price
inflation", "Housing std"]]
       self.df_ = self.df_[self.df_["Year"] >= 2010]
       self.df_["t"] = self.df_["Year"] - 2010

       self.p = self.lin_reg("Population")
       self.i = self.lin_reg("Median income")
       self.h = self.lin_reg("Median price inflation")
       self.s = self.lin_reg("Housing std")

   def lin_reg(self, col):
       df_ = self.df_[["t", col]].dropna()
       x = df_["t"]
       y = df_[col]

       model = LinearRegression()
       model.fit(np.array(x).reshape(-1, 1), np.array(y))

       print(model.coef_[0], model.intercept_)

       return lambda x: model.predict([[x]])[0]
```

```
    def predict(self, t):
        z = (self.h(t)/30 - self.i(t))/self.s(t)
        return self.p(t)*norm.cdf(z)


# Albuquerque
print("Albuquerque")
print("Regression values:")
alb_q2 = City(adf)

print("")

print("Population:\t\t", alb_q2.p(24), alb_q2.p(34), alb_q2.p(64))
print("Income:\t\t\t", alb_q2.i(24), alb_q2.i(34), alb_q2.i(64))
print("Housing price:\t\t", alb_q2.h(24), alb_q2.h(34), alb_q2.h(64))
print("Housing price std:\t", alb_q2.s(24), alb_q2.s(34), alb_q2.s(64))

print("Predicted:\t\t", alb_q2.predict(24), alb_q2.predict(34), alb_q2.predict(64))


# Seattle
print("Seattle")
print("Regression values:")
sea_q2 = City(sdf)

print("")

print("Population:\t\t", sea_q2.p(24), sea_q2.p(34), sea_q2.p(64))
print("Income:\t\t\t", sea_q2.i(24), sea_q2.i(34), sea_q2.i(64))
print("Housing price:\t\t", sea_q2.h(24), sea_q2.h(34), sea_q2.h(64))
print("Housing price std:\t", sea_q2.s(24), sea_q2.s(34), sea_q2.s(64))

print("Predicted:\t\t", sea_q2.predict(24), sea_q2.predict(34), sea_q2.predict(64))


# Plot residual values

ares = adf.apply(lambda row: alb_q2.predict(row["Year"]-2010) - row["Homeless total"],
axis=1)
sres = sdf[sdf["Year"] != 2021].apply(lambda row: sea_q2.predict(row["Year"]-2010) -
row["Homeless total"], axis=1)

plt.plot([2007, 2023], [0, 0], c="black", linestyle="--")

plt.scatter(adf["Year"], ares, label="Albuquerque residuals", c="C0")
plt.scatter(sdf["Year"][sdf["Year"] != 2021], sres, label="Seattle residuals", c="C1")

plt.title("Residual Homeless People over Time")
plt.legend()
plt.xlabel("Year")
plt.ylabel("Homeless people")
plt.ylim(-10000, 10000)

plt.show()
```

Question 3

```
# Imports
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.optimize import curve_fit
from scipy.stats import norm
from sklearn.linear_model import LinearRegression


# Load data
sdf = pd.read_csv("data/sea.csv", thousands=',')


# Drop 2021
sdf_ = sdf[sdf["Year"] != 2021]


# Rdd
def rdd(x, y, thres):
   # Returns the slope and intercept for the regression on each side of the
discontinuity
   x, y = np.array(x), np.array(y)

   # Linear fits
   left_model = LinearRegression()
   right_model = LinearRegression()

   mask = x <= thres
   mask2 = x > thres


   left_model.fit(x[mask].reshape(-1, 1), y[mask])


   right_model.fit(x[mask2].reshape(-1, 1), y[mask2])

   return [left_model.coef_, left_model.intercept_, right_model.coef_,
right_model.intercept_]


def rdd_data(x, y, thres):
   lc, li, rc, ri = rdd(x[:-1], y[:-1], 2014)
   lc, rc = lc[0], rc[0]

   # Left points
   x_left = [x.min(), thres]
   y_left = [lc*x.min()+li, lc*thres+li]

   # Right points
   x_right= [x.max(), thres]
   y_right = [rc*x.max()+ri, rc*thres+ri]

   return x_left, y_left, x_right, y_right


# Plot
plt.scatter(sdf_["Year"], sdf_["Homeless total"], label="Total Homeless Count")
plt.scatter(sdf_["Year"], sdf_["Transitional"], label="Transitional Housing")

xl1, yl1, xr1, yr1 = rdd_data(sdf_["Year"], sdf_["Homeless total"], 2015)
```

```
xl2, yl2, xr2, yr2 = rdd_data(sdf_["Year"], sdf_["Transitional"], 2015)

plt.plot(xl1, yl1, c="C0")
plt.plot(xr1, yr1, c="C0")

plt.plot(xl2, yl2, c="C1")
plt.plot(xr2, yr2, c="C1")

plt.plot([2015, 2015], [0, 14000], c="black")

plt.xlabel("Year")
plt.ylabel("Count")
plt.suptitle("Impact of transitional housing")
plt.title("In 2015 (black line), Seattle ended their 10-year homelessness plan",
fontsize=10)
plt.legend()

plt.show()
```