

# Assessing the Impact of Redlining on Seattle Housing Values: An Economic Analysis for Policy Guidance

## Background and Motivation:

Redlining is a discriminatory practice from the 1930s where government maps categorized neighborhoods based on race, often denying financial services to minority communities. This has left behind economic challenges and environmental issues that still affect these areas today. It is important to understand how redlining relates to current housing values to see how these past practices continue to impact communities and contribute to the racial wealth gap. This report looks at how redlining has shaped housing trends in Seattle, highlighting inequalities and offering insights for fairer urban development and policies (Source: [New York Times, 2021](#)). By examining these patterns, the report aims to shed light on the effects of past policies and explore pathways toward more equitable urban development.

## Data & Descriptive Statistics:

### *Data Source:*

**All Seattle homes sold in 2014** from the “**Mapping Inequality Project**” ([Link](#)). It explains the details about the property characteristics, demographics, and historical redlining. Moreover, this data set captures various characteristics of the house including the redlining designations.

### *Population Size:*

The population size of the dataset is **10,508** observations, which supposedly represents all Seattle homes sold in 2014.

### *Sample Size:*

There are in total **4,347** observations after removing outliers.

### *Sample Selection Criteria and Resulting Sample Size:*

The original dataset contained **10,508** observations representing all Seattle homes sold in 2014. Missing values were removed for all variables included in the model, as these are critical for modeling housing prices. Additionally, outliers were excluded based on logical inconsistencies or unrealistic values (variables such as **Sale Price, Home Age, Lot Size, Total Bedrooms, and Total Bathrooms**). For example, properties with sales prices above \$10 million were excluded as they likely represent unique or non-residential properties. Similarly, homes with 0 bedrooms or bathrooms and Home Age values of 0 or less were omitted as they indicate errors or non-standard properties. For lot size, we excluded properties below 500 sqft (too small for residential use) or above 50,000 sqft (likely estates or special-purpose properties). After these exclusions, the final sample consisted of **4,347** observations.

### *Descriptive Statistics:*

#### Descriptive Statistics for Property and Redlining Variables

Variable Name	Unit	Mean	Standard Deviation
Redlined Area	Binary	0.134	0.341
Lot Size	Square Feet	5,326.305	4,730.409
Building Area	Square Feet	2,021.643	883.044
Total Bedrooms	Count	3.236	0.940
Sale Price	Dollars	647,293.200	475,928.400
Total Bathrooms	Count	2.108	0.675
Renovation Status	Binary	0.087	0.282
Single-Family Home	Binary	0.706	0.456
Townhome	Binary	0.264	0.441
Building Condition: Excellent	Binary	0.058	0.233
Building Condition: Good	Binary	0.179	0.384
Building Condition: Average	Binary	0.758	0.428
Building Condition: Fair	Binary	0.005	0.073
Home Age	Count	43.956	36.885

### *Discussion of Descriptive Statistics:*

The average sales price for homes in the dataset was approximately **\$647,293**, with a standard deviation of **\$475,928**, reflecting the broad range of housing prices in Seattle. The mean lot size was **5,326** sqft, with significant variability of **4,730** sqft, highlighting differences between dense urban properties and suburban or rural homes. Approximately **13.4%** of homes were in redlined areas, showcasing the historical impact of discriminatory practices. Homes in redlined areas often have lower values, reflecting the long-term effects of disinvestment. Additionally, the mean number of bedrooms was **3**, and bathrooms were **2**, indicating a typical residential property profile. These variables provide a foundation for exploring the relationship between redlining and housing values, helping to uncover systemic inequalities and inform policies for urban development.

## Models:

### *Model Justification:*

To understand the relationship between redlining and housing values, two models will be run here, linear model, with Sale Price as the dependent variable and log-linear model, with natural logarithm of Sale Price as the dependent variable.

- **Linear Model** allows us to analyze the absolute dollar changes in sales price associated with the independent variables.
- **Semi-Log Model** will be useful as housing price often has skewed distribution which can be offset by taking natural logarithm. This model is insightful for understanding relative changes in sales price as the coefficient in the model represents the percentage change in sales price for a one-unit change in the independent variable.

### *Independent Variables:*

1. Redlined Area
  - Properties in redlined areas are expected to have lower values due to historical disinvestment and reduced financial opportunities.
2. Lot Size
  - Larger lot sizes generally increase property value as they offer more space for construction, landscaping, or recreation.
3. Building Area
  - Larger buildings typically translate into higher property values, reflecting increased living or usable space.
4. Total Bedrooms
  - More bedrooms increase the size and utility of a property, particularly for families.
5. Total Bathrooms
  - Like bedrooms, bathrooms are a key factor in determining sale value because of their utility.
6. Renovation Status
  - Renovated properties tend to have higher market values due to improved aesthetics and functionality.
7. Property Type
  - Single-family homes and town homes are generally more valuable than condos due to privacy, space, and ownership rights.
8. Building Condition
  - Building conditions significantly affect property value. Better conditions increase desirability and reduce maintenance costs.
9. Home Age
  - Older properties may have lower values due to wear and outdated designs. However, historical properties in good condition might hold higher value in certain neighborhoods.

## Results:

### Regression Results: Linear and Log-Linear Models of Housing Prices

	Dependent variable:	
	Sales Price (1)	log(Sales Price) (2)
Redlined Area	-84,712.200*** (11,800.800)	-0.196*** (0.021)
Lot Size	7.955 (4.392)	-0.00000 (0.00000)
Building Area	382.272*** (20.501)	0.0004*** (0.00001)
Total Bedrooms	-72,600.730*** (9,036.681)	-0.072*** (0.010)
Total Bathrooms	52,400.340*** (12,035.730)	0.063*** (0.015)
Renovation Status	-20,989.940 (29,545.960)	-0.002 (0.033)
Single-Family Home	-71,545.370 (66,809.860)	-0.018 (0.062)
Townhome	14,210.600 (71,672.840)	-0.084 (0.062)
Building Condition: Excellent	227,959.200*** (35,670.450)	0.149 (0.463)
Building Condition: Good	179,565.600*** (27,819.810)	0.132 (0.462)
Building Condition: Average	133,015.800*** (21,068.480)	0.041 (0.462)
Building Condition: Fair	155,710.200*** (55,215.510)	0.074 (0.472)
Home Age	-2,956.320*** (867.913)	-0.006*** (0.001)
Home Age Squared	44.664*** (6.718)	0.0001*** (0.00001)
Constant	-147,418.300 (82,090.570)	12.584*** (0.467)
Note	Robust SE used for Linear Model SE used for Log-Linear Model	
Observations	4,347	4,347
R2	0.538	0.393
Adjusted R2	0.536	0.391

Note: \*p<0.05; \*\*p<0.01; \*\*\*p<0.001

#### Discussion of Robust Standard Errors:

In our model, the Breusch-Pagan test revealed heteroskedasticity in the residuals of the linear model (**p-value < 2.2e-16**), suggesting that the variability of errors increases with the values of certain predictors. To address this, **robust standard errors** were applied to correct for heteroskedasticity, ensuring the validity of the t-statistics and p-values. In contrast, the log-linear model did not require robust standard errors as the residual diagnostics did not indicate significant heteroskedasticity (**p-value = 0.09464**).

### *Chosen Variable Interpretation:*

#### **Linear Model**

- **Redlined Area:** Properties in redlined areas are associated with an average decrease of **\$84,712** in sales price compared to non-redlined areas, holding all else constant. This negative relationship is **statistically significant at the 0.1% level** and is expected, as redlined areas historically suffered from disinvestment and discriminatory practices.
- **Single-Family Home:** Single-family homes are associated with a decrease of **\$71,545** in sales price compared to condos, holding all else constant. This result is **not statistically significant** and is unexpected, as single-family homes are generally more valuable than condos.
- **Home Age:** For every additional year of home age, the sales price decreases by **\$2,956**, and this relationship is **statistically significant at the 0.1% level**. This is expected, as older homes often have fewer modern amenities and higher maintenance costs. However, the quadratic term (Home Age Squared) is positive, indicating that extremely old homes might experience an increase in price due to their historical value.

#### **Log-Linear Model**

- **Redlined Area:** Being in a redlined area is associated with a **17.8% decrease in sales price** compared to non-redlined areas, holding all else constant. This negative relationship is **statistically significant at the 0.1% level** and is expected.
- **Townhome:** Being a town home is associated with an **8.0% decrease in sales price** compared to condos, holding all else constant. This result is **not statistically significant** and is unexpected, as townhomes often offer more space than condos and are typically priced higher.
- **Total Bathrooms:** Each additional bathroom is associated with a **6.5% increase in sales price**, holding all else constant. This relationship is **statistically significant at the 0.1% level** and is expected, as additional bathrooms increase utility and desirability.

### *R<sup>2</sup> Interpretation:*

#### **Linear Model ( $R^2 = 0.538$ )**

- The linear model explains **53.8%** of the variation in sales prices. This indicates that the included independent variables account for a significant portion of the variability in housing prices.

#### **Log-Linear Model ( $R^2 = 0.393$ )**

- The log-linear model explains **39.3%** of the variation in the log-transformed sales prices. While slightly lower than the linear model, it reflects the complexity of housing prices when modeled in relative terms.

*Brief Discussion on Other Coefficient Sign:*

- **Building Condition: Excellent:** Positive – Expected, as homes in excellent condition are more valuable.
- **Building Condition: Good:** Positive – Expected, since good condition adds appeal to buyers.
- **Building Condition: Fair:** Positive – Slightly unexpected, as fair condition homes may require repairs, but buyers may still value the potential.
- **Renovation Status:** Negative – Unexpected, as renovated homes are usually more desirable.
- **Building Area:** Positive – Expected, as larger homes generally have higher prices.
- **Lot Size:** Positive – Expected, as larger lots provide more space and value.
- **Total Bedrooms:** Negative – Unexpected, as more bedrooms typically increase utility and demand.

### **Implications:**

Homes in redlined areas were worth much less in 2014, averaging **\$84,712 lower** or **17.8% less** than homes in non-redlined areas. This is because “redlining” denied these neighborhoods loans and investments, which hurt property values and left them with fewer resources and opportunities. Fewer services, negative views about these areas, and being near industrial zones made the problem worse. Fixing this will need policies to improve these neighborhoods and support fair development.

## Limitation:

One limitation of this analysis is that the data does not include all factors that affect housing prices, such as school quality, local amenities, or crime rates. While the model covers important factors like lot size and building condition, missing these other details could change how much redlining impacts home values. Without this extra information, the results might not fully capture the whole picture and could lead to omitted variable bias. Addressing this limitation would require additional data to capture these factors more comprehensively.

## Sources:

**Bibliography:** New York Times. 2021. "What Is Redlining?" *The New York Times*, August 17, 2021. <https://www.nytimes.com/2021/08/17/realestate/what-is-redlining.html>

## Technical Appendix:

### Breusch-Pagan Test:

Model	BP Statistics	df	p-value	Conclusion
Linear Model	150.95	14	<2.2e-16	Reject null at 5% significance level (heteroskedasticity)
Log-Linear Model	21.281	14	0.09464	Fail to reject null at 5% significance level (homoskedasticity)

### Matrix of Robust Standard Error:

Note that we only use robust standard error for the Linear Model. Therefore, we only show the matrix for the linear model.

### Linear Model:

	(Intercept)	redlined	lotsizesquarefeet	buildingareasqft	totalbedrooms	TotalCalculatedBathCount	Renovated	Single_family
(Intercept)	6.738861e+09	-2.564539e+08	-277458.88421	-143764.50856	-94854599.821	-4.885579e+07	-6.372440e+08	-3559942829.6
redlined	-2.564539e+08	1.392588e+08	-1696.87213	68747.77180	-5033870.693	8.557865e+06	-1.584835e+07	-130204078.6
lotsizesquarefeet	-2.774589e+05	-1.696872e+03	19.29323	-19.16338	3624.247	-3.734944e+03	5.584572e+03	225249.6
buildingareasqft	-1.437645e+05	6.874777e+04	-19.16338	420.30080	-114653.981	1.095420e+03	-1.071105e+05	-556114.9
totalbedrooms	-9.485460e+07	-5.033871e+06	3624.24651	-114653.98079	81661595.972	-2.400505e+07	2.056595e+07	59500181.9
TotalCalculatedBathCount	-4.885579e+07	8.557865e+06	-3734.94371	1095.42030	-24005045.815	1.448589e+08	-6.103169e+07	-153267100.9
Renovated	-6.372440e+08	-1.584835e+07	5584.57159	-107110.47767	20565946.581	-6.103169e+07	8.729640e+08	1045123266.3
Single_family	-3.559943e+09	-1.302041e+08	225249.59082	-556114.88451	59500181.898	-1.532671e+08	1.045123e+09	4463557215.7
Townhome	-4.916348e+09	-4.421356e+07	270900.53871	-283059.01539	24996232.897	-1.296419e+08	9.338784e+08	4459148723.1
BC_Excellent	-8.368362e+08	1.103886e+08	29559.53579	-134069.81819	156470078.787	-1.609625e+08	1.340509e+08	460883096.3
BC_Good	-1.485983e+09	1.383373e+08	60940.14857	20246.42409	92261876.681	-1.212194e+08	9.806666e+07	705253004.9
BC_Average	-8.001049e+08	1.511735e+08	15090.85740	95690.76885	75403499.301	-9.953046e+07	-4.798788e+07	-33570010.2
BC_Fair	-9.607970e+08	1.535539e+08	16066.32511	106692.58786	70777340.400	-6.419901e+07	3.360623e+07	35475451.9
Home_Age	1.906461e+06	2.780107e+06	-1715.44844	7588.62080	-1823045.727	3.055952e+06	-3.032451e+06	-30616995.4
I(Home_Age^2)	4.817283e+03	-1.693509e+04	11.68144	-42.64710	10055.038	-1.233586e+04	-8.016868e+03	158836.6
		Townhome	BC_Excellent	BC_Good	BC_Average	BC_Fair	Home_Age	I(Home_Age^2)
(Intercept)	-4916347871.6	-836836160.28	-1.485983e+09	-800104886.15	-960796956.01	1906461.051	4817.28267	
redlined	-44213563.6	110388625.72	1.383373e+08	151173502.87	153553861.03	2780107.217	-16935.09352	
lotsizesquarefeet	270900.5	29559.54	6.094015e+04	15090.86	16066.33	-1715.448	11.68144	
buildingareasqft	-283059.0	-134069.82	2.024642e+04	95690.77	106692.59	7588.621	-42.64710	
totalbedrooms	24996232.9	156470078.79	9.226188e+07	75403499.30	70777340.40	-1823045.727	10055.03849	
TotalCalculatedBathCount	-129641912.4	-160962510.12	-1.212194e+08	-99530463.87	-64199010.37	3055951.694	-12335.86267	
Renovated	933878449.1	134050948.89	9.806666e+07	-47987875.65	33606225.03	-3032450.638	-8016.86786	
Single_family	4459148723.1	460883096.28	7.052530e+08	-33570010.23	35475451.87	-30616995.449	158836.57162	
Townhome	5136996530.9	492356178.28	9.544793e+08	164593518.46	260094585.91	-20782771.034	108160.77696	
BC_Excellent	492356178.3	1272381074.73	5.562842e+08	420978071.08	455039233.42	-1985548.029	-23235.88665	
BC_Good	954479288.5	556284193.40	7.739418e+08	422825894.54	478661946.83	-2757343.090	-6166.29775	
BC_Average	164593518.5	420978070.87	4.228259e+08	443880863.66	409246584.37	3582391.984	-33720.28300	
BC_Fair	260094585.9	455039233.33	4.786619e+08	409246584.48	3048753131.63	5193137.639	-56414.33880	
Home_Age	-20782771.0	-1985548.03	-2.757343e+06	3582391.98	5193137.64	753272.758	-5422.82472	
I(Home_Age^2)	108160.8	-23235.89	-6.166298e+03	-33720.28	-56414.34	-5422.825	45.12604	

Standard Error, t-stats, and p-value:

Linear Model:

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-1.4742e+05	8.2091e+04	-1.7958	0.0725958	.
redlined	-8.4712e+04	1.1801e+04	-7.1785	8.253e-13	***
lotsizesquarefeet	7.9553e+00	4.3924e+00	1.8111	0.0701872	.
buildingareasqft	3.8227e+02	2.0501e+01	18.6463	< 2.2e-16	***
totalbedrooms	-7.2601e+04	9.0367e+03	-8.0340	1.206e-15	***
TotalCalculatedBathCount	5.2400e+04	1.2036e+04	4.3537	1.369e-05	***
Renovated	-2.0990e+04	2.9546e+04	-0.7104	0.4774840	
Single_family	-7.1545e+04	6.6810e+04	-1.0709	0.2842829	
Townhome	1.4211e+04	7.1673e+04	0.1983	0.8428429	
BC_Excellent	2.2796e+05	3.5670e+04	6.3907	1.825e-10	***
BC_Good	1.7957e+05	2.7820e+04	6.4546	1.204e-10	***
BC_Average	1.3302e+05	2.1068e+04	6.3135	3.002e-10	***
BC_Fair	1.5571e+05	5.5216e+04	2.8200	0.0048235	**
Home_Age	-2.9563e+03	8.6791e+02	-3.4062	0.0006646	***
I(Home_Age^2)	4.4664e+01	6.7176e+00	6.6488	3.323e-11	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

\*Corrected using Robust Standard Error

Log-Linear Model:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	1.258e+01	4.670e-01	26.948	< 2e-16	***
redlined	-1.957e-01	2.137e-02	-9.158	< 2e-16	***
lotsizesquarefeet	-3.170e-06	2.130e-06	-1.488	0.137	
buildingareasqft	3.796e-04	1.332e-05	28.505	< 2e-16	***
totalbedrooms	-7.197e-02	9.582e-03	-7.511	7.06e-14	***
TotalCalculatedBathCount	6.288e-02	1.522e-02	4.132	3.66e-05	***
Renovated	-2.283e-03	3.334e-02	-0.068	0.945	
Single_family	-1.845e-02	6.180e-02	-0.299	0.765	
Townhome	-8.355e-02	6.243e-02	-1.338	0.181	
BC_Excellent	1.492e-01	4.632e-01	0.322	0.747	
BC_Good	1.318e-01	4.624e-01	0.285	0.776	
BC_Average	4.139e-02	4.622e-01	0.090	0.929	
BC_Fair	7.417e-02	4.720e-01	0.157	0.875	
Home_Age	-5.829e-03	9.543e-04	-6.109	1.09e-09	***
I(Home_Age^2)	7.025e-05	8.201e-06	8.566	< 2e-16	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1