Seattle Housing Values | Report

Background and Motivation

Redlining was a discriminatory practice that systematically denied financial services, particularly mortgages, to residents of specific areas based on their race or ethnicity. This practice designated neighborhoods deemed high risk for mortgage lending, predominantly immigrant communities or people of color. Redlining followed the "one-house rule", a single black household in a middle-class area could make the whole neighborhood "risky" for mortgage in the eyes of the federal government *(Camila 2016)*[(1)] .The purpose of this report is to examine the long-lasting effects of historical redlining on housing prices, which included data sets from all Seattle homes sold in 2014, and to develop multiple regression models to estimate the relationship between location in a previously redlined area and 2014 housing values while controlling for relevant housing characteristics. Understanding this relationship is critical for several reasons, such as the persistent wealth gap since Redlining has contributed significantly to the racial wealth gap in the United States. By denying access to mortgages and other financial services, it prevented many Black families and other minorities from building wealth through homeownership. Also, despite being outlawed in 1968, the legacy of redlining continues to influence today's housing market. Homes in formerly redlined areas are often worth significantly less than comparable properties in those not redlined regions *(Lathan 2023)*[(2)].

Data and Descriptive Statistics

The data used for this project is sourced from the project *"Mapping Inequality: Redlining in New Deal America" (Nelson, R. K., Winling, L, et al. – 2023)* [(3)]. Our data set consists of 10,508 Seattle homes sold in 2014, with qualitative and qualitative variables and dummy variables measure on nominal and ordinal scale. Our original data set consisted of properties with missing information for variables, which were removed listwise to preserve the model's performance by only analyzing data sets with complete values for all variables used in our model, reducing our sample size to 4,741 observations. The data also consisted of possibly miscategorized or erroneous values specifically for properties sold at unusually high prices. To minimize the effect of the outliers we limited our data to properties with sales price of at least $6.5 million after looking at sources like *The Curbed Seattle* [(4)] for reference, further reducing our sample size to 4733 observations.

The average sales price was $642,378.800, with a standard deviation of $467,842.200 reduced from $1,504,492.000 after adjusting for outliers. Of the total homes sold only 13.3% were from previously redlined neighborhoods. 65% of the homes sold were single family homes, followed by 26.5% townhomes and condominiums at the bottom with 8.5%.76.5% of the homes sold were in average building condition, followed by 17.1% in good condition and only 5.3% in excellent condition, which is low but can be explained by a possibly higher sales price.

Descriptive Statistics Table for Seattle homes sold in 2014

Time frame: 2014 (1 year)

Sample size: 4741

Time frame: 2014 (1 year)

Sample size: 4733 (After removing outliers)

Variables	***Mean***	***Standard Deviation***	***Mean***	***Standard Deviation***
Sales Price ($)	686,374.800	1,504,492.000	642,378.800	467,842.200
Building Square Footage (Sq Ft)	1,968.416	903.814	1,966.026	891.958
Total Bathrooms	2.064	0.694	2.062	0.689
Single family	0.650	0.477	0.649	0.477
Townhome	0.265	0.441	0.265	0.441
Condo	0.085	0.280	0.085	0.279
Building Condition-Excellent	0.053	0.224	0.053	0.224
Building Condition-Good	0.171	0.377	0.171	0.376
Building Condition-Average	0.765	0.424	0.765	0.424
Building Condition-Fair	0.005	0.074	0.005	0.074
Building Condition-Poor	0.000	0.015	0.000	0.015
Home Age (Yr)	42.084	36.378	42.067	36.384
Redlined	0.133	0.339	0.133	0.339

Models

The model tends to include four regression models examining housing prices, columns 1 & 2 represents the regression model using data with outliers, whereas column 3 & 4 uses data without the outliers. For dependent variables, in models 1 & 3, we have direct sales prices; in models 2 & 4, we have a log of the sales prices. We used the BPG test to test the data for heteroskedasticity and models 2, 3 & 4 have been adjusted with the updated standard errors.

The log transformation of sales price helps account for the skewed distribution of housing prices and allows for interpreting coefficients as percentages change. Independent variables included in the model are property characteristics such as building square footage, a key determinant of property value as its usable living space. We have total bathrooms, an essential amenity influencing property utility and value. The property types indicated by single-family, which are standalone structures, typically offer more privacy and land. In contrast, townhomes with different ownership structures, like HOAs, are essential to consider in this model.

We also have excellent building conditions, good, average, and fair, because these variables capture the property's maintenance level and structural integrity, which directly affect both current value and future maintenance costs. Building conditions also represent generational wealth disparity, capturing the disability of homeowners from marginalized communities to maintain their properties, resulting in reducing presumed property values. Home age, a linear term capturing depreciation effects, is also an important aspect affecting sales price. We have included home-age, squared accounts for nonlinear age effects, such as historic properties that might have higher value due to historical significance, while middle-aged homes behave differently. Lastly, we included redlined, which captures historical discriminatory lending practices' long-term effects on property values.

Regression Models With and Without Outliers

	Dependent variable:			
	Sales Price	Log of Sales price	Sales Price	Log of Sales price
	With Outliers		Without Outliers	
	(1)	(2)	(3)	(4)
Redlined	-130,954.600**	-0.225***	-102,291.500***	-0.220***
	(64,756.970)	(0.021)	(13,879.030)	(0.021)
Building Square Footage	347.157***	0.0003***	358.432***	0.0003***
	(33.942)	(0.00002)	(22.081)	(0.00002)
Total Bathrooms - Count	78,004.980*	0.047**	28,458.110	0.045**
	(43,506.050)	(0.019)	(18,313.590)	(0.018)
Single Family	-317,692.500***	-0.131***	-320,694.400***	-0.134***
	(87,924.850)	(0.040)	(35,047.420)	(0.040)
Townhome	-231,917.600***	-0.195***	-249,563.800***	-0.195***
	(87,404.810)	(0.036)	(27,678.350)	(0.035)
Building Condition- Excellent	212,616.000	0.167	154,216.700***	0.150
	(323,551.100)	(0.109)	(59,832.150)	(0.106)
Building Condition-Good	-39,240.680	0.104	76,455.400	0.104
	(311,824.300)	(0.104)	(56,260.890)	(0.103)
Building Condition-Average	-26,115.570	0.037	47,037.270	0.038
	(307,154.200)	(0.102)	(55,908.250)	(0.101)
Building Condition-Fair	-58,170.780	0.039	50,951.030	0.043
	(418,870.000)	(0.126)	(73,794.500)	(0.126)
Home Age	1,438.724	-0.008***	-3,599.012***	-0.008***
	(2,697.622)	(0.001)	(750.397)	(0.001)
Home Age - squared	3.976	0.0001***	46.624***	0.0001***
	(23.736)	(0.00001)	(6.114)	(0.00001)
Constant	70,096.510	12.611***	116,715.300**	12.603***
	(307,260.600)	(0.102)	(53,950.800)	(0.101)
Observations	4,741	4,741	4,733	4,733
R^2	0.056	0.353	0.503	0.370
Adjusted R^2	0.054	0.351	0.502	0.368
Note:				*p<0.1; **p<0.05; ***p<0.01
	Standard errors have been corrected for heteroskedasticity for Model (2), (3) and (4)			

Table 1: Regression Summary

Results

Our analysis focuses on the data set after excluding outliers. The regression analysis reveals significant disparities in housing prices across historically redlined areas and property types. Controlling for building square footage, number of bathrooms, home age, type of property and building conditions, properties in redlined areas demonstrate a negative statistically significant ($p < 0.01$) association with sales price in all four of our models. Property in a redlined neighborhood sells for $102,291 less than that in a non-redlined neighborhood and approximately 19.75% less in the logarithmic model without outliers. Abstracting from all other property characteristics including building square footage and building conditions single family homes exhibit a negative significant association ($p < 0.01$) with sales price, with single-family homes selling for approximately $320,694 less than condominiums. This negative association persists in the logarithmic specification as well exhibiting a 12.5% decrease in sales price of single-family homes as compared to a condominium.

Home age exhibits the expected negative significant ($p < 0.01$) association with sales price, whereas the home age squared captures the significant ($p < 0.01$) non-linear relationship where the sales price initially decreases at an increasing rate and increases at a decreasing rate after reaching minimum critical value. The resulting U-shaped relationship indicates that newer properties command premium prices, followed by a period of accelerating depreciation that eventually transitions into appreciation for historic properties.

The removal of outliers substantially improves the model's explanatory power, as evidenced by the R^2 values in columns 3 and 4. The linear specification without outliers explains 50.3% of the variation in sales prices, while its logarithmic accounts for 37% of the variation. These values represent a considerable improvement over the models with outliers, which only explain 5.6% and 35.3% of the variation respectively. This marked increase in explanatory power suggests that extreme observations were potentially distorting the relationship between housing characteristics and sales prices, and their removal leads to more reliable estimates of these relationships.

Implications

Our analysis shows that while factors such as building square footage, property and home age do exhibit the expected strong association with sales price, homes sold in Seattle in 2014 also showed a statistically significant negative association with areas previously redlined, emphasizing the lingering effects of historical racial discrimination. Although the sales price exhibits the conventional positive association with excellent building conditions, it is also important to note that generational wealth disparity induced by the unfair redlining practice also limits the resources available to black homeowners in maintaining and upgrading their properties, which further negatively impacts the presumed value of their properties. The correlation of wealth disparity with building conditions further iterates its causal relationship with redlining.

Limitations

One limitation of this study and the analysis, in general, is the sample size and time frame since the initial dataset of 10,508 homes was significantly reduced to 4,733 homes, which represents a loss of over half the original data points. This substantial reduction could introduce bias into the analysis, as the homes excluded might have specific characteristics that are now underrepresented in the final sample. Also, by concentrating solely on data from 2014, the study provides a snapshot of the Seattle housing market at a specific point in time, which has several drawbacks such as real estate markets often experience seasonal fluctuations within a year, which may not be fully captured in this analysis. Also, 2014 may have been an atypical year for the Seattle housing market due to specific economic conditions, potentially skewing the results. Lastly, the study cannot account for multi-year trends in housing prices, neighborhood development, or demographic shifts that might influence property values over time.

Technical Appendix

The Breusch-Pagan-Godfrey (BPG) test is being used to check for heteroskedasticity in the regression models. For example, for the Model A (Sales Price with outliers), the BPG test result shows a p-value > 0.05, indicating that we cannot reject the null hypothesis of homoskedasticity. For Models B, C, and D, the BPG test results show p-values < 0.05, leading to the rejection of the null hypothesis of homoskedasticity. In cases where heteroskedasticity is detected (Models B, C, and D), the models are adjusted using robust standard errors to account for this issue. The BPG test is crucial in this analysis as it helps identify whether the variance of the residuals is constant across all levels of the independent variables.

For Model A (Dependent Variable -Sales Price, data set with outliers)

The p-value > 0.05, hence we cannot reject the null hypothesis of homoskedasticity.

```
	studentized Breusch-Pagan test

data:  Model_A
BP = 7.7561, df = 11, p-value = 0.735
```

Figure 1: BPG Test for Model A (1)

For Model B (Dependent Variable -Log of Sales Price, data set with outliers)

The p-value < 0.05, hence we reject the null hypothesis of homoskedasticity and adjust the model with the robust standard errors.

```
	studentized Breusch-Pagan test

data:  Model_B
BP = 23.849, df = 11, p-value = 0.01339
```

Figure 2: BPG Test for Model B (2)

(Intercept)	redlined	buildingareasqft	TotalCalculatedBathCount	Single_family
1.021544e-01	2.098902e-02	1.784364e-05	1.879372e-02	4.043841e-02
Townhome	BC_Excellent	BC_Good	BC_Average	BC_Fair
3.624798e-02	1.087462e-01	1.036537e-01	1.015519e-01	1.262613e-01
Home_Age	Home_Age_squared			
9.944128e-04	8.610918e-06			

Figure 3 : Model B - Matrix of heteroskedasticity-robust standard errors.

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	1.2611e+01	1.0215e-01	123.4533	< 2.2e-16	***
redlined	-2.2521e-01	2.0989e-02	-10.7299	< 2.2e-16	***
buildingareasqft	3.4270e-04	1.7844e-05	19.2055	< 2.2e-16	***
TotalCalculatedBathCount	4.7015e-02	1.8794e-02	2.5016	0.012396	*
Single_family	-1.3114e-01	4.0438e-02	-3.2431	0.001191	**
Townhome	-1.9532e-01	3.6248e-02	-5.3886	7.448e-08	***
BC_Excellent	1.6722e-01	1.0875e-01	1.5377	0.124185	
BC_Good	1.0350e-01	1.0365e-01	0.9985	0.318071	
BC_Average	3.6786e-02	1.0155e-01	0.3622	0.717188	
BC_Fair	3.8969e-02	1.2626e-01	0.3086	0.757609	
Home_Age	-7.5584e-03	9.9441e-04	-7.6009	3.526e-14	***
Home_Age_squared	8.0929e-05	8.6109e-06	9.3984	< 2.2e-16	***

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

Figure 4 - Model B - Updated model with robust standard errors

For Model C (Dependent Variable -Sales Price, data set without outliers)

The p-value < 0.05, hence we reject the null hypothesis of homoskedasticity and adjust the model with the robust standard errors.

```
        studentized Breusch-Pagan test

data:  Model_C
BP = 267.8, df = 11, p-value < 2.2e-16
```

Figure 5: BPG Test for Model C (3)

```
       (Intercept)             redlined      buildingareasqft TotalCalculatedBathCount         Single_family
      53950.801543         13879.034456            22.080917          18313.594407          35047.415633
          Townhome         BC_Excellent              BC_Good            BC_Average               BC_Fair
      27678.344578         59832.153739          56260.888236          55908.244883          73794.500066
          Home_Age     Home_Age_squared
        750.397290             6.114426
```

Figure 6: Model C - Matrix of heteroskedasticity-robust standard errors.

```
t test of coefficients:

                           Estimate  Std. Error t value  Pr(>|t|)    
(Intercept)              1.1672e+05  5.3951e+04  2.1634  0.030563 *  
redlined                -1.0229e+05  1.3879e+04 -7.3702 2.001e-13 ***
buildingareasqft         3.5843e+02  2.2081e+01 16.2327 < 2.2e-16 ***
TotalCalculatedBathCount 2.8458e+04  1.8314e+04  1.5539  0.120267    
Single_family           -3.2069e+05  3.5047e+04 -9.1503 < 2.2e-16 ***
Townhome                -2.4956e+05  2.7678e+04 -9.0166 < 2.2e-16 ***
BC_Excellent             1.5422e+05  5.9832e+04  2.5775  0.009982 ** 
BC_Good                  7.6455e+04  5.6261e+04  1.3589  0.174229    
BC_Average               4.7037e+04  5.5908e+04  0.8413  0.400206    
BC_Fair                  5.0951e+04  7.3795e+04  0.6904  0.489948    
Home_Age                -3.5990e+03  7.5040e+02 -4.7961 1.667e-06 ***
Home_Age_squared         4.6624e+01  6.1144e+00  7.6253 2.926e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 7: Model C - Updated model with robust standard errors.

For Model D (Dependent Variable – Log of Sales Price, data set without outliers)

The p-value < 0.05, hence we reject the null hypothesis of homoskedasticity and adjust the model with the robust standard errors.

```
	studentized Breusch-Pagan test

data:  Model_D
BP = 28.374, df = 11, p-value = 0.002836
```

Figure 8: BPG test for Model D (4)

```
      (Intercept)          redlined    buildingareasqft TotalCalculatedBathCount       single_family
     1.012687e-01      2.088650e-02        1.776817e-05         1.840460e-02        3.953192e-02
         Townhome      BC_Excellent             BC_Good           BC_Average             BC_Fair
     3.544985e-02      1.057834e-01        1.026963e-01         1.008225e-01        1.257673e-01
         Home_Age  Home_Age_squared
     9.591394e-04      8.341983e-06
```

Figure 9: Model D - Matrix of heteroskedasticity-robust standard errors

```
t test of coefficients:

                          Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)             1.2603e+01  1.0127e-01 124.4512 < 2.2e-16 ***
redlined               -2.1984e-01  2.0887e-02 -10.5255 < 2.2e-16 ***
buildingareasqft        3.4815e-04  1.7768e-05  19.5940 < 2.2e-16 ***
TotalCalculatedBathCount 4.5125e-02 1.8405e-02   2.4518 0.0142491 *
single_family          -1.3450e-01  3.9532e-02  -3.4022 0.0006739 ***
Townhome               -1.9454e-01  3.5450e-02  -5.4877 4.284e-08 ***
BC_Excellent            1.4958e-01  1.0578e-01   1.4140 0.1574289
BC_Good                 1.0393e-01  1.0270e-01   1.0120 0.3115688
BC_Average              3.8394e-02  1.0082e-01   0.3808 0.7033635
BC_Fair                 4.3254e-02  1.2577e-01   0.3439 0.7309207
Home_Age               -7.9155e-03  9.5914e-04  -8.2527 < 2.2e-16 ***
Home_Age_squared        8.4915e-05  8.3420e-06  10.1793 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 10: Model D - Updated Model with robust standard errors

Sources

[1]Domonoske, Camila. (2016, October 19). *Interactive Redlinging Map Zooms In On America's History Of Discrimination.* NPR. Retrieved November 27, 2024, from https://www.npr.org/sections/thetwo-way/2016/10/19/498536077/interactive-redlining-map-zooms-in-on-americas-history-of-discrimination

[2]Lathan, N. (2023, September 20). *50 Years after being outlawed, redlining still drives neighborhood health inequities.* UC Berkely Public Health. Retrieved November 27, 2024, from https://publichealth.berkeley.edu/news-media/research-highlights/50-years-after-being-outlawed-redlining-still-drives-neighborhood-health-inequities

[3]Nelson, R. K., Winling, L, et al. (2023). *Mapping Inequality: Redlining in New Deal America*. Digital Scholarship Lab. https://dsl.richmond.edu/panorama/redlining

[4]Sean Keeley, (2014). *The Most Expensive Seattle Properties Sold in 2014*, Mapped! https://seattle.curbed.com/maps/the-most-expensive-seattle-properties-sold-in-2014-mapped

Word Count

Word Count – 1533 words excluding sources, titles and captions.