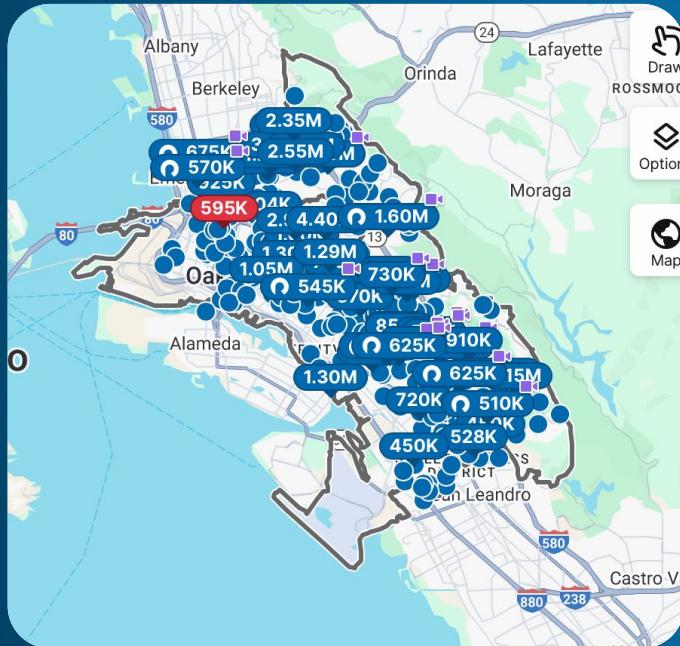
The background image is a wide-angle aerial photograph of a city, likely Oakland, California. It shows a dense urban area with numerous buildings, a river or bay to the right, and hills covered in trees in the distance under a sky filled with dramatic, white and grey clouds.

# Oakland Real Estate Assessment

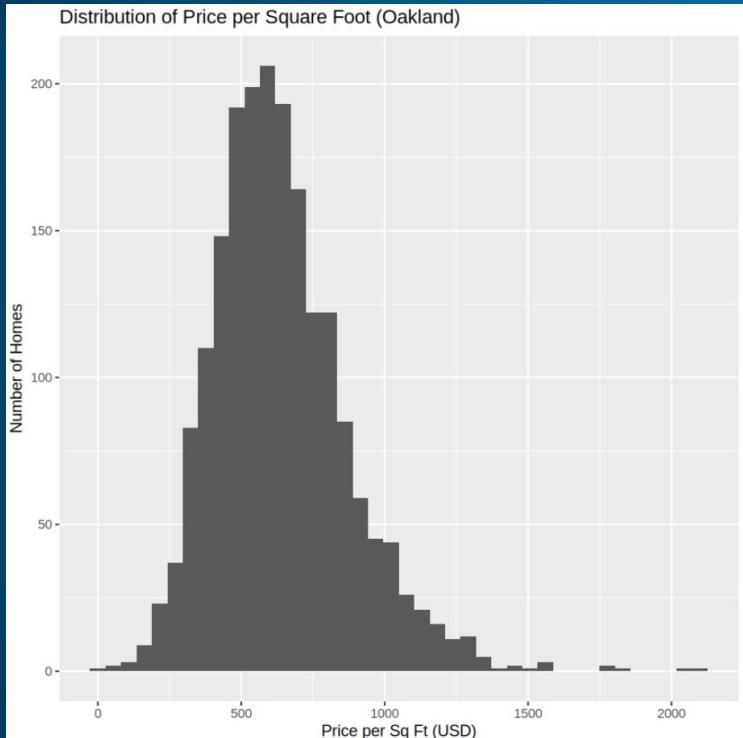
Persimmon Inc (Group 11)

# Oakland Home Sales Data Set



- After data cleaning, we focused on 1,956 listings located within the official Oakland city limits.
- Models 1–2 are estimated on homes with non-missing price, square footage, beds, and baths ( $N \approx 1,938$  homes)
- Model 3 uses a smaller subset ( $N = 1,621$ ) with additional required data on HOLC grade, amenities, ZIP, and sale year

# Price per Square Foot in Oakland



The average price per square foot among sold homes over the past 5 years is \$637, with most homes clustered between \$400–\$800 per square foot.

# Model 1 - Price vs. Square Feet

The intercept of \$242,100 represents the predicted price of a home with zero square feet

Call:

```
lm(formula = PRICE ~ SQUARE.FEET, data = oakland_homes)
```

Residuals:

Min	1Q	Median	3Q	Max
-3801447	-246827	-76955	187079	3800916

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )		
(Intercept)	2.421e+05	2.053e+04	11.79	<2e-16 ***		
SQUARE.FEET	4.820e+02	9.718e+00	49.60	<2e-16 ***		
---						
Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 ' '	1

Residual standard error: 493900 on 1948 degrees of freedom  
Multiple R-squared: 0.5581, Adjusted R-squared: 0.5578  
F-statistic: 2460 on 1 and 1948 DF, p-value: < 2.2e-16

# Model 1 - Price vs. Square Feet

For each additional square foot, the model predicts an average price increase of approximately \$482

Call:

```
lm(formula = PRICE ~ SQUARE.FEET, data = oakland_homes)
```

Residuals:

Min	1Q	Median	3Q	Max
-3801447	-246827	-76955	187079	3800916

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	2.421e+05	2.053e+04	11.79	<2e-16 ***
SQUARE.FEET	4.820e+02	9.718e+00	49.60	<2e-16 ***

---

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

Residual standard error: 493900 on 1948 degrees of freedom

Multiple R-squared: 0.5581, Adjusted R-squared: 0.5578

F-statistic: 2460 on 1 and 1948 DF, p-value: < 2.2e-16

# Model 1 - Price vs. Square Feet

Call:

```
lm(formula = PRICE ~ SQUARE.FEET, data = oakland_homes)
```

Residuals:

Min	1Q	Median	3Q	Max
-3801447	-246827	-76955	187079	3800916

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )		
(Intercept)	2.421e+05	2.053e+04	11.79	<2e-16 ***		
SQUARE.FEET	4.820e+02	9.718e+00	49.60	<2e-16 ***		
---						
Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 ' '	1



Residual standard error: 493900 on 1948 degrees of freedom  
Multiple R-squared: 0.5581, Adjusted R-squared: 0.5578  
F-statistic: 2460 on 1 and 1948 DF, p-value: < 2.2e-16

The extremely small p-value (< 2.2e-16) confirms this effect is highly statistically significant.

# Model 2 - Price vs. Square Feet + Bed + Baths

Comparison of Simple and Extended Price Models

Dependent variable: Sale Price (price)

	Size Only (1)	Size + Beds + Baths (2)
SQUARE.FEET	481.978*** (9.718)	570.244*** (20.409)
BEDS		-68,163.100*** (12,322.810)
BATHS		57,892.000*** (17,668.910)
Constant	242,149.700*** (20,534.490)	173,059.400*** (25,308.800)
Observations	1,950	1,938
R2	0.558	0.578
Adjusted R2	0.558	0.577

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

After controlling for beds and baths, the value of each extra square foot increases.

# Model 2 - Price vs. Square Feet + Bed + Baths

Comparison of Simple and Extended Price Models

-----  
Dependent variable: Sale Price (price)  
-----

	Size Only (1)	Size + Beds + Baths (2)
SQUARE.FEET	481.978*** (9.718)	570.244*** (20.409)
BEDS		-68,163.100*** (12,322.810)
BATHS		57,892.000*** (17,668.910)
Constant	242,149.700*** (20,534.490)	173,059.400*** (25,308.800)
Observations	1,950	1,938
R2	0.558	0.578
Adjusted R2	0.558	0.577

-----  
Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Holding size and baths constant, **more bedrooms are associated with a lower sale price**, suggesting buyers prefer fewer, larger rooms.

# Model 2 - Price vs. Square Feet + Bed + Baths

Comparison of Simple and Extended Price Models

-----  
Dependent variable: Sale Price (price)  
-----

	Size Only (1)	Size + Beds + Baths (2)
SQUARE.FEET	481.978*** (9.718)	570.244*** (20.409)
BEDS		-68,163.100*** (12,322.810)
BATHS		57,892.000*** (17,668.910)
Constant	242,149.700*** (20,534.490)	173,059.400*** (25,308.800)
Observations	1,950	1,938
R2	0.558	0.578
Adjusted R2	0.558	0.577

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Holding size and bedrooms constant, **more bathrooms are associated with a higher sale price**, suggesting buyers are willing to pay a premium for extra bathrooms.

# Model 2 - Price vs. Square Feet + Bed + Baths

Comparison of Simple and Extended Price Models

-----  
Dependent variable: Sale Price (price)  
-----

	Size Only (1)	Size + Beds + Baths (2)
SQUARE.FEET	481.978*** (9.718)	570.244*** (20.409)
BEDS		-68,163.100*** (12,322.810)
BATHS		57,892.000*** (17,668.910)
Constant	242,149.700*** (20,534.490)	173,059.400*** (25,308.800)
Observations	1,950	1,938
R2	0.558	0.578
Adjusted R2	0.558	0.577

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Overall, the extended model fits the data better (higher R<sup>2</sup>)

# Model 2 - Predicting Typical Home Price

We consider a 2-bedroom, 2-bath home with 1,500 square feet:

PRICE = 1.7305940 + 0.0057024 \*  
1500 -0.6816310 \* 2 + 0.5789200  
\*2=10.078772,  
meaning the predicted price is  
**\$1,007,884.**

The 95% confidence interval is  
[\$979,812 \$1,035,956]

Call:  
lm(formula = PRICE ~ SQUARE.FEET + BEDS + BATHS, data = homes\_df)

Residuals:

Min	1Q	Median	3Q	Max
-46.047	-2.414	-0.590	1.801	33.932

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.7305940	0.2530880	6.838	1.07e-11 ***
SQUARE.FEET	0.0057024	0.0002041	27.940	< 2e-16 ***
BEDS	-0.6816310	0.1232281	-5.531	3.61e-08 ***
BATHS	0.5789200	0.1766891	3.276	0.00107 **

---

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

Residual standard error: 4.716 on 1934 degrees of freedom

Multiple R-squared: 0.5779, Adjusted R-squared: 0.5772

F-statistic: 882.5 on 3 and 1934 DF, p-value: < 2.2e-16

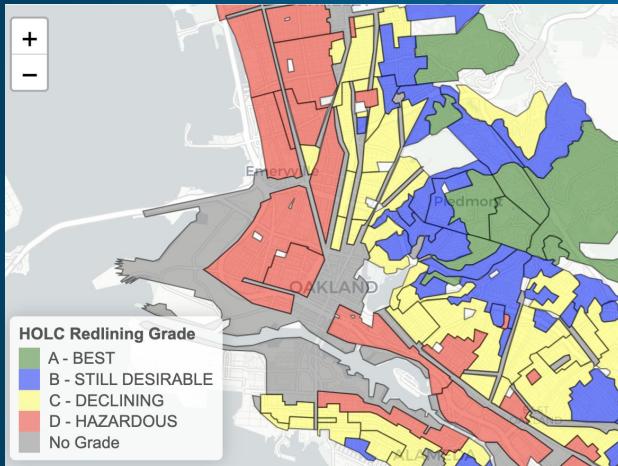
Regression equation:

PRICE = 1.731 + 0.006 \* SQUARE.FEET + -0.682 \* BEDS + 0.579 \* BATHS

PRICE = 1.7305940 + 0.0057024 \* square feet -0.6816310 \* bedrooms + 0.5789200 \* bathrooms

# Model 3 - Extended Model

+HOLC grade, ZIP, amenities, sale year



Historical redlining and neighborhoods



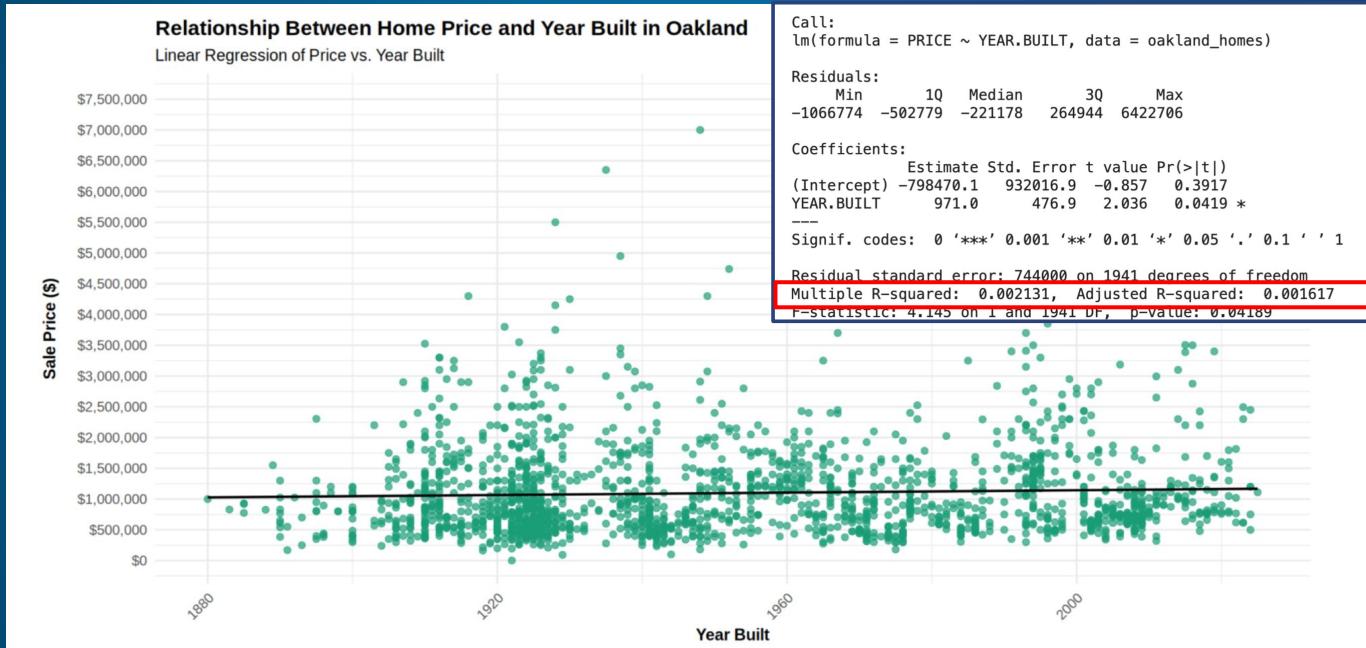
Distance to quality schools, job geography, amenities



Sale Year

# Model 3 - Extended Model

Age is Just a Number\*



\*Statistically significant, but explains almost 0% of price variation

# Model 3 - Extended Model

## Key Observations

### Historic HOLC grade

- HOLC **B** homes sell for about **\$80k more** than similar **HOLC C** homes (after controls).

### School quality access

- Within **0.5 miles** of a quality school → about **\$120k higher** sale price.

### Job geography

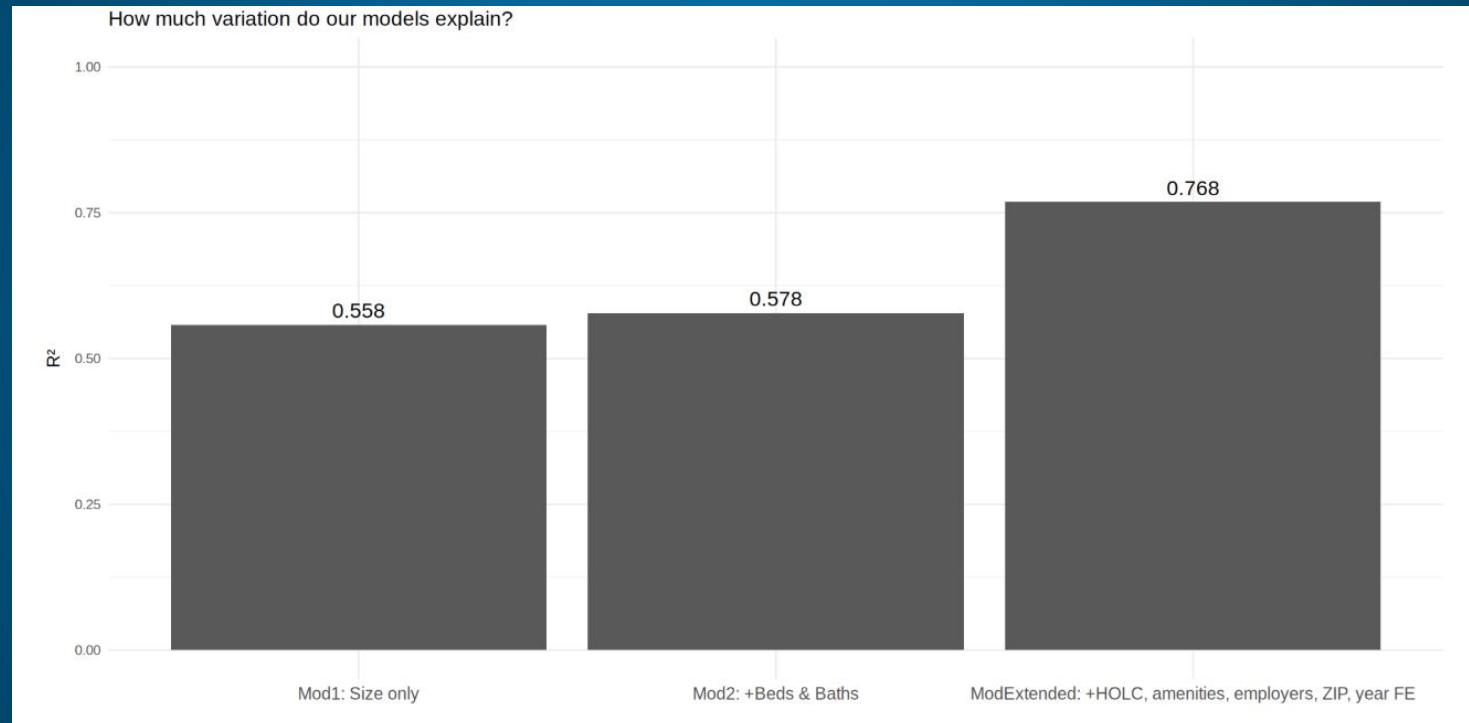
- **1 mile closer to Pixar/Emeryville** → roughly **+\$300k**
- Priciest homes are **farther from downtown and industrial areas** → **+\$264-584k per mile** further.

### Amenities & Waterfront

- Close to **grocery + retail/dining** → about **+\$180–200k**.
- **Waterfront** has the biggest premium: roughly **+\$380k per mile** closer.

# Model 3 - Extended Model

## Key Observations



# Model 3 - Extended Model

## Identifying Investment Opportunities

### Recommendations

- Target historically HOLC B neighborhoods (and select C/D blocks) that have strong present-day fundamentals (clean employers, good amenities)
- Prioritize neighborhoods within easy reach of quality schools if willing to pay for premium.
- **Avoid homes right next to heavy industrial/port areas** – buyers pay a clear premium to be farther away, so these locations sell at a discount.
- **Scan ZIP codes with premiums/discounts of about ±\$200k vs. 94609** – these “in-between” areas often look like emerging neighborhoods in transition or early gentrification.



## A Closer Look at Key Areas

- 94607 (West Oakland / Jack London): **prices are offset by strong location advantages.**
- 94608 (North Oakland / Emeryville): **market is driven by major attractors like Pixar.**

# How to Use our Models

The model can help identify high-ROI homes using 2 strategies:

## (1) Undervaluation / Mispricing with Model 3

- Perform a regression capturing recently-closed sales to estimate fair market value based on the number of beds, baths, and square feet
- Assess current listings to identify homes with a >10% delta between list price and the modeled fair value.
- Analyze each highlighted listing to ensure it isn't a false signal or outlier.

## (2) Value-Add / Forced Equity with Model 2

- Perform a regression to identify fair value based on bathrooms, bedrooms, or square footage
- Quantify the modeled price opportunity from adding a bathroom, bedroom, or expanding square footage
- Identify opportunities where fair value increase is significantly higher than cost to add beds/baths/sqft

• • •

**Let's Invest Together!**

• • •

# Appendix

# Effect of Lot Size on Non-SFR Homes

Single-family homes are about \$275K more expensive than the baseline (controlling for size and lot).

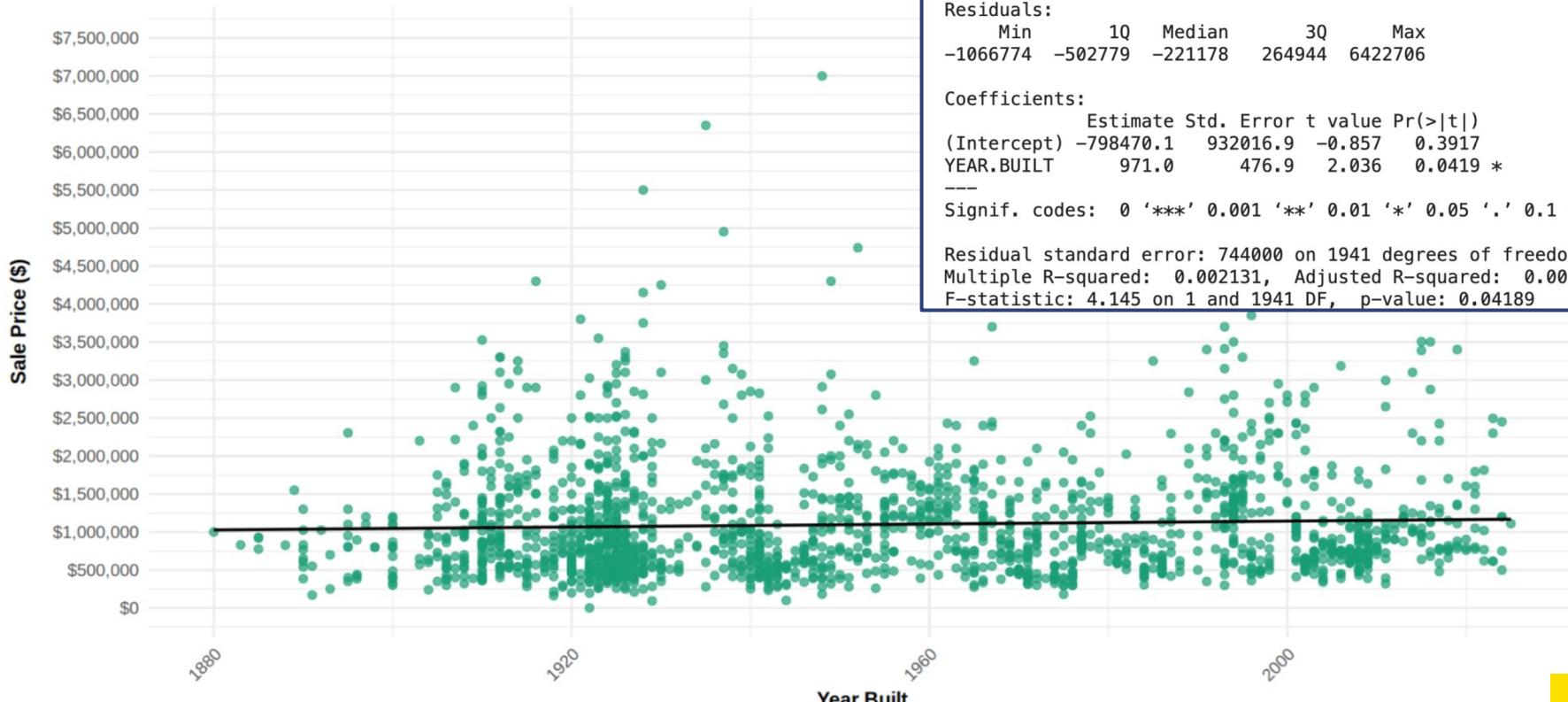
Townhouses are about \$121K more expensive than the baseline.

```
Call:  
lm(formula = PRICE ~ SQUARE.FEET + LOT.SIZE + PROPERTY.TYPE,  
    data = oakland_homes)  
  
Residuals:  
    Min      10     Median      30      Max  
-3944510 -267238   -60572   216120  3641437  
  
Coefficients:  
              Estimate Std. Error t value Pr(>|t|)  
(Intercept) 5.473e+03 5.032e+04  0.109  0.9134  
SQUARE.FEET  4.907e+02 1.060e+01 46.278 < 2e-16  
LOT.SIZE     1.715e+00 8.139e-01  2.107  0.0353  
PROPERTY.TYPEMulti-Family (2-4 Unit) -1.263e+05 1.337e+05 -0.944  0.3451  
PROPERTY.TYPEMulti-Family (5+ Unit) -2.191e+06 3.049e+05 -7.187 9.8e-13  
PROPERTY.TYPENon-Family Residential -8.950e+05 5.019e+05 -1.783  0.0747  
PROPERTY.TYPESingle Family Residential 2.749e+05 4.852e+04  5.667 1.7e-08  
PROPERTY.TYPETownhouse            1.211e+05 5.675e+04  2.133  0.0330  
  
(Intercept) ***  
SQUARE.FEET      *  
LOT.SIZE        .  
PROPERTY.TYPEMulti-Family (2-4 Unit) ***  
PROPERTY.TYPEMulti-Family (5+ Unit) ***  
PROPERTY.TYPENon-Family Residential ***  
PROPERTY.TYPETownhouse          *  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 5e+05 on 1730 degrees of freedom  
(212 observations deleted due to missingness)  
Multiple R-squared:  0.5764,    Adjusted R-squared:  0.5747  
F-statistic: 336.3 on 7 and 1730 DF,  p-value: < 2.2e-16
```

# Remember: Age is Just a Number

## Relationship Between Home Price and Year Built in Oakland

Linear Regression of Price vs. Year Built



Call:

```
lm(formula = PRICE ~ YEAR.BUILT, data = oakland_homes)
```

Residuals:

Min	1Q	Median	3Q	Max
-1066774	-502779	-221178	264944	6422706

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-798470.1	932016.9	-0.857	0.3917
YEAR.BUILT	971.0	476.9	2.036	0.0419 *

---

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

Residual standard error: 744000 on 1941 degrees of freedom

Multiple R-squared: 0.002131, Adjusted R-squared: 0.001617

F-statistic: 4.145 on 1 and 1941 DF, p-value: 0.04189

# Model 3 - Extended Model

## Key Observations

```

Call:
lm(formula = price_num ~ SQUARE.FEET + BEDS + BATHS + is HOLC_A +
  is_HOLC_B + is_HOLC_D + within_0_5_mile_quality_school +
  distance_to_bart_miles + dist_to_employer_pixar_miles + dist_to_employer_downtown_miles +
  dist_to_employer_industrial_miles + dist_to_employer_college_miles +
  dist_to_grocery_miles + dist_to_retail_dining_miles + dist_to_shopping_center_miles +
  dist_to_hiking_miles + dist_to_waterfront_miles + zip_factor +
  sale_year_fe, data = oakland_ext)

Residuals:
    Min      1Q  Median      3Q     Max 
-2989523 -182256   -9625   163769   3077364 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -532629.8  247789.9 -2.150 0.031744 *  
SQUARE.FEET      526.9    19.4 27.156 < 2e-16 *** 
BEDS             -26424.5  11744.2 -2.250 0.024585 *  
BATHS            15433.0   16104.7  0.958 0.338062    
is_HOLC_A        33710.3   40146.7  0.840 0.401215    
is_HOLC_B        82099.3   26351.0  3.116 0.001869 ** 
is_HOLC_D        -29918.8   35196.1 -0.850 0.395419    
within_0_5_mile_quality_school 124694.2  23305.7  5.350 1.01e-07 *** 
distance_to_bart_miles       10423.2   34825.1  0.299 0.764750    
dist_to_employer_pixar_miles -311526.0  72663.2 -4.287 1.92e-05 *** 
dist_to_employer_downtown_miles 587240.4  106906.7  5.493 4.60e-08 *** 
dist_to_employer_industrial_miles 264291.6  77384.2  3.415 0.000653 *** 
dist_to_employer_college_miles 199700.6   57627.3  3.465 0.000544 *** 
dist_to_grocery_miles         -199594.9  52139.8 -3.828 0.000134 *** 
dist_to_retail_dining_miles   -183961.5  46862.9 -3.926 9.03e-05 *** 
dist_to_shopping_center_miles -7877.0    40605.9 -0.194 0.846212    
dist_to_hiking_miles          57301.4   43794.8  1.308 0.190926    
dist_to_waterfront_miles      -381675.1  108424.8 -3.520 0.000443 *** 

```

zip_factor94501	-	-211650.1	375968.4	-0.563	0.573551
zip_factor94560	-	402506.5	377798.1	1.065	0.286857
zip_factor94601	-	58556.9	119494.3	0.490	0.624174
zip_factor94602	-	286778.3	98085.2	2.924	0.003507 **
zip_factor94603	-	151414.2	156983.7	0.965	0.334932
zip_factor94605	-	249015.9	135821.1	1.833	0.066929 .
zip_factor94606	-	-47549.1	103694.5	-0.459	0.646621
zip_factor94607	-	81473.3	119115.7	0.684	0.494085
zip_factor94608	-	-217837.7	118433.5	-1.839	0.06055 .
zip_factor94610	-	422247.8	93457.5	4.518	6.70e-06 ***
zip_factor94611	-	272840.6	75135.8	3.631	0.000291 ***
zip_factor94612	-	302547.2	114663.2	2.639	0.008407 **
zip_factor94618	-	537437.1	71285.3	7.539	7.92e-14 ***
zip_factor94619	-	273123.0	122465.6	2.230	0.025874 *
zip_factor94621	-	249501.6	155373.9	1.600	0.108514
sale_year_fe2021	-	150076.5	43421.1	3.456	0.000562 ***
sale_year_fe2022	-	203580.0	44945.6	4.529	6.36e-06 ***
sale_year_fe2023	-	114232.4	46976.6	2.432	0.015139 *
sale_year_fe2024	-	96792.0	47237.2	2.049	0.040621 *
sale_year_fe2025	-	-15757.2	47359.0	-0.333	0.739391

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

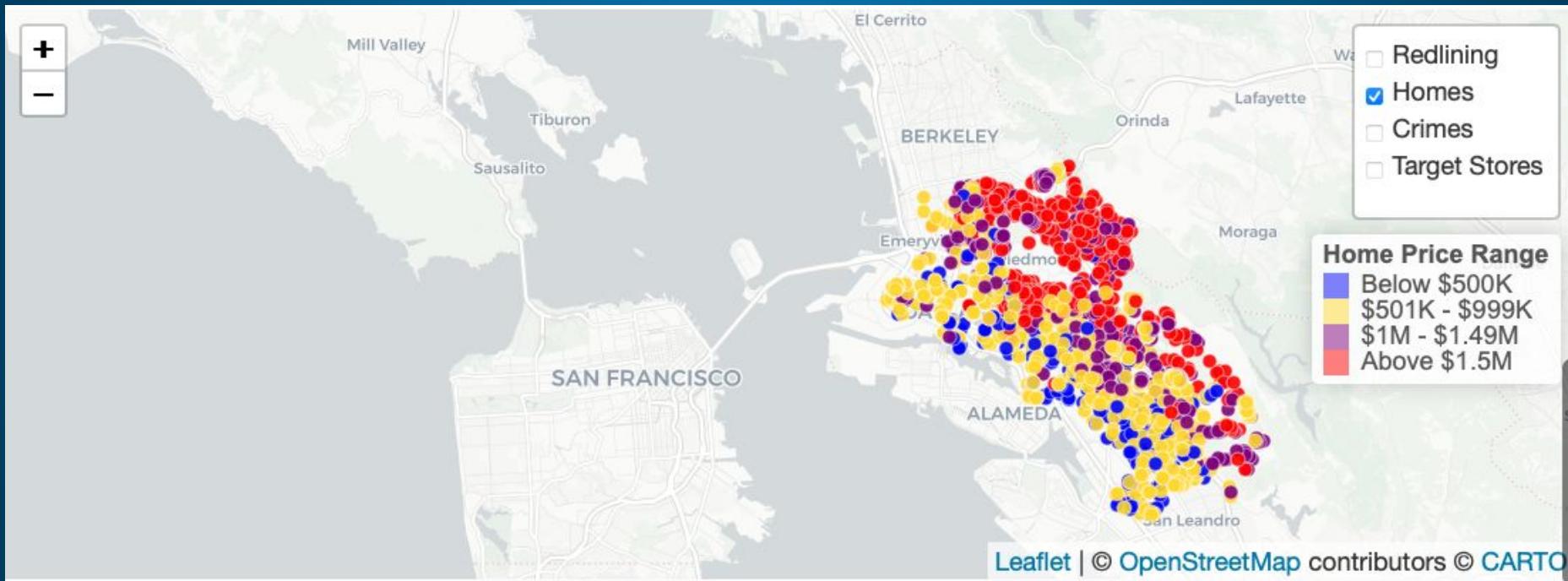
Residual standard error: 351500 on 1583 degrees of freedom

(317 observations deleted due to missingness)

Multiple R-squared: 0.7676, Adjusted R-squared: 0.7622

F-statistic: 141.3 on 37 and 1583 DF, p-value: < 2.2e-16

# Homes sold in the last 5 years



## Question 6

PRICE = 1.7305940 + 0.0057024 \* square feet - 0.6816310 \* bedrooms + 0.5789200 \* bathrooms

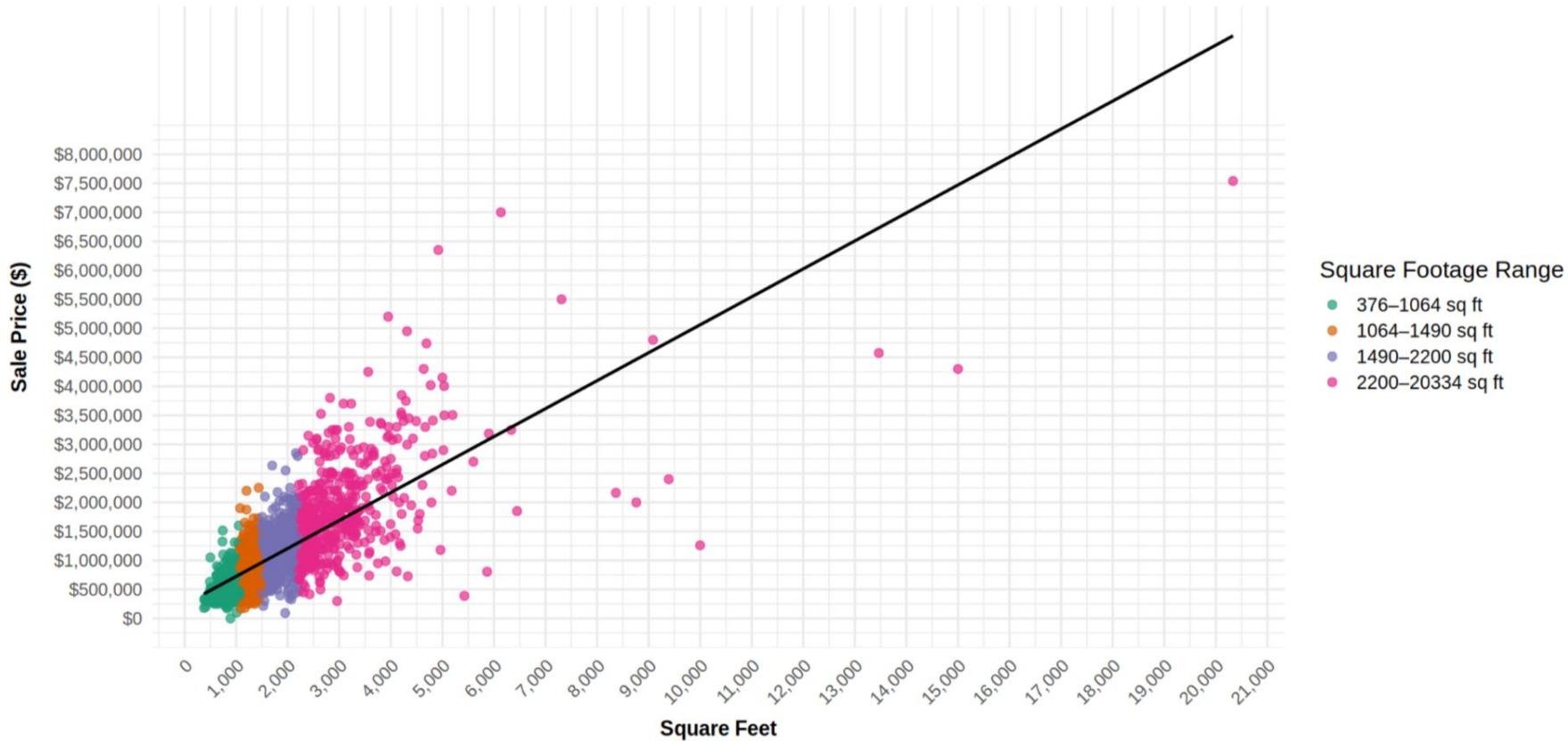
The 95% confidence interval is  
(9.79812,10.35956)

A matrix: 1 × 3 of type dbl

	<b>fit</b>	<b>lwr</b>	<b>upr</b>
<b>1</b>	10.07884	9.79812	10.35956

# Oakland Home Prices vs. Square Feet

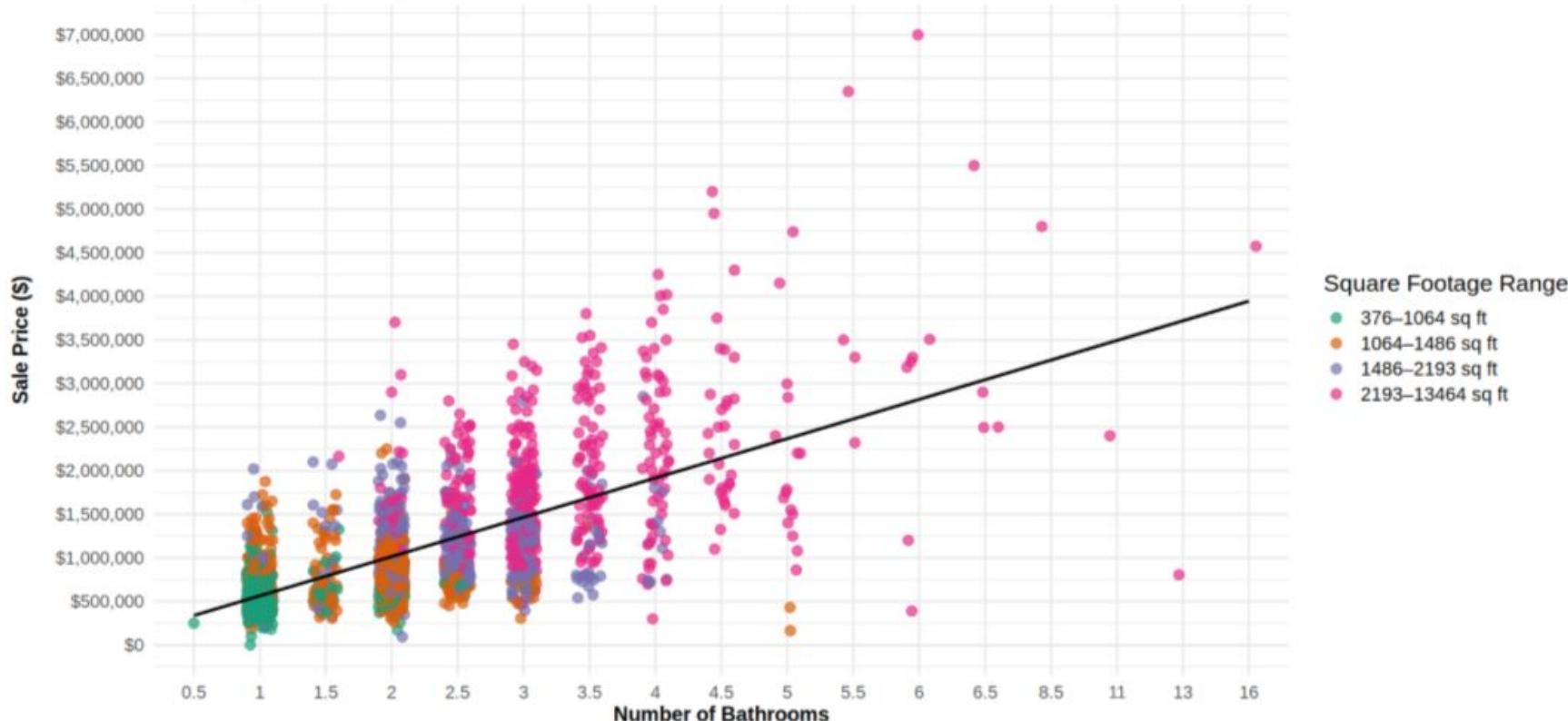
Linear Regression with Price and Square Footage Increments



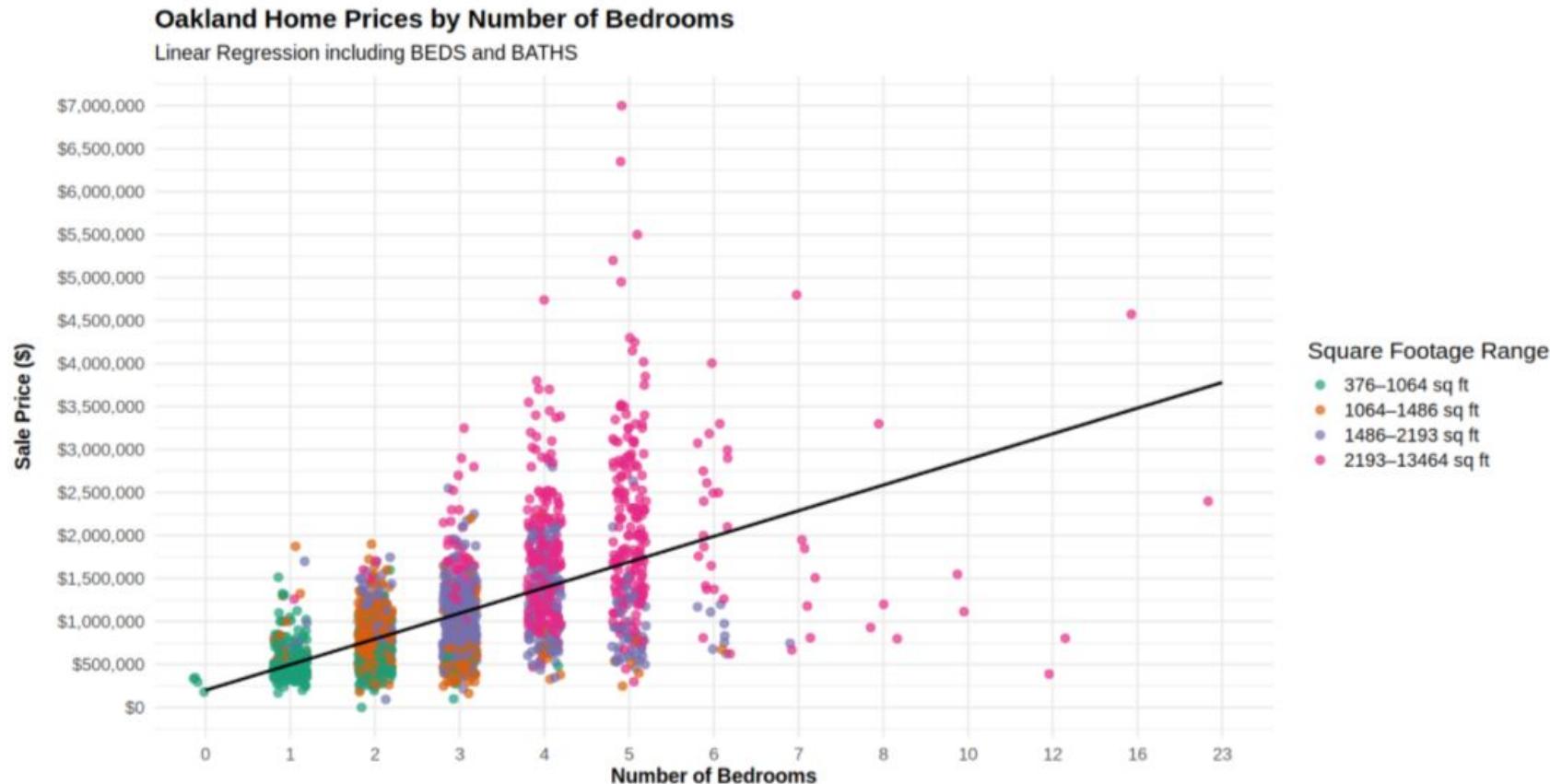
$$\text{PRICE} = 173,059.40 + 570.24 \times \text{SQUARE.FEET} - 68,163.10 \times \text{BEDS} + 57,892 \times \text{BATHS}.$$

### Oakland Home Prices by Number of Bathrooms

Linear Regression including BEDS and BATHS



$$\text{PRICE} = 173,059.40 + 570.24 \times \text{SQUARE.FEET} - 68,163.10 \times \text{BEDS} + 57,892 \times \text{BATHS}.$$

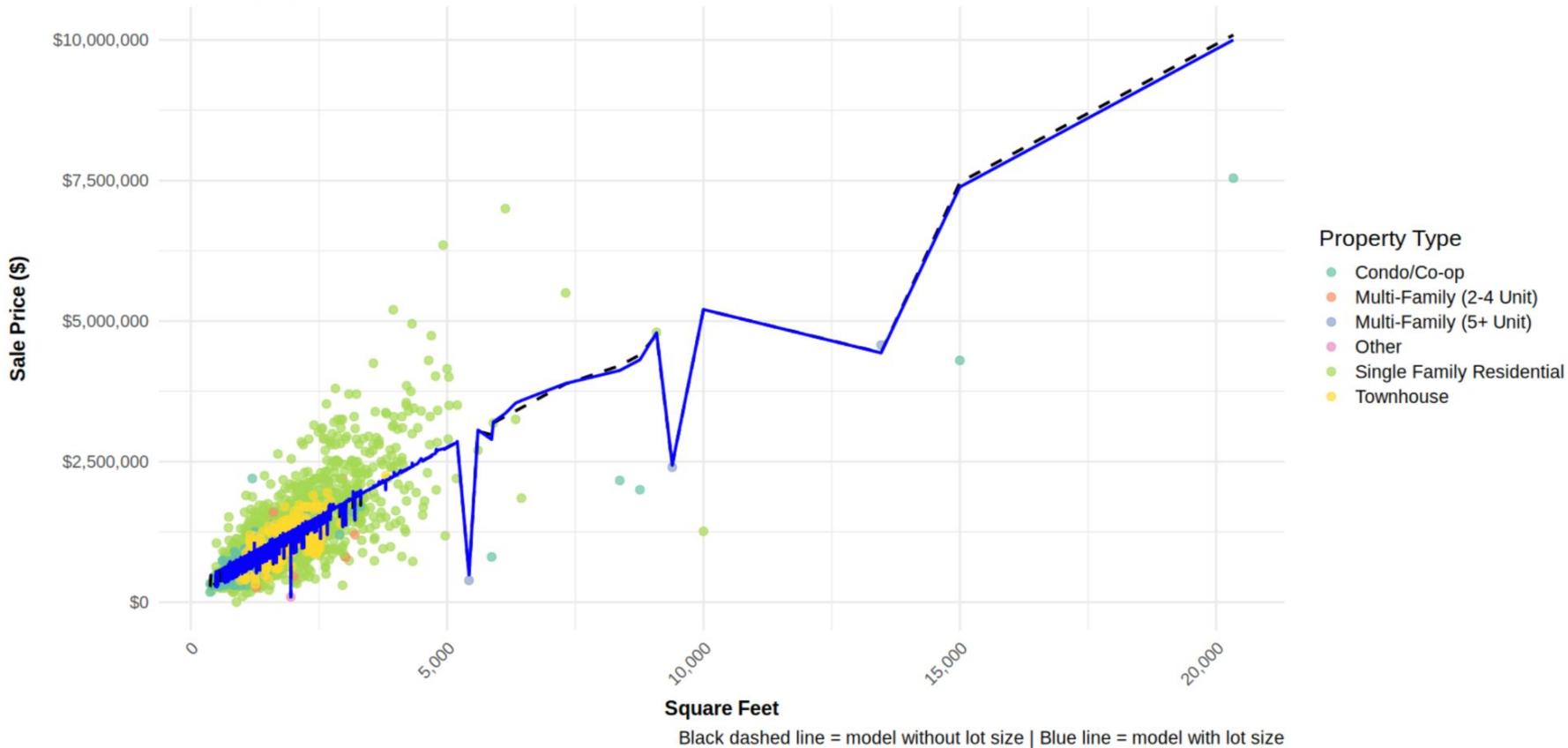


## Question 4

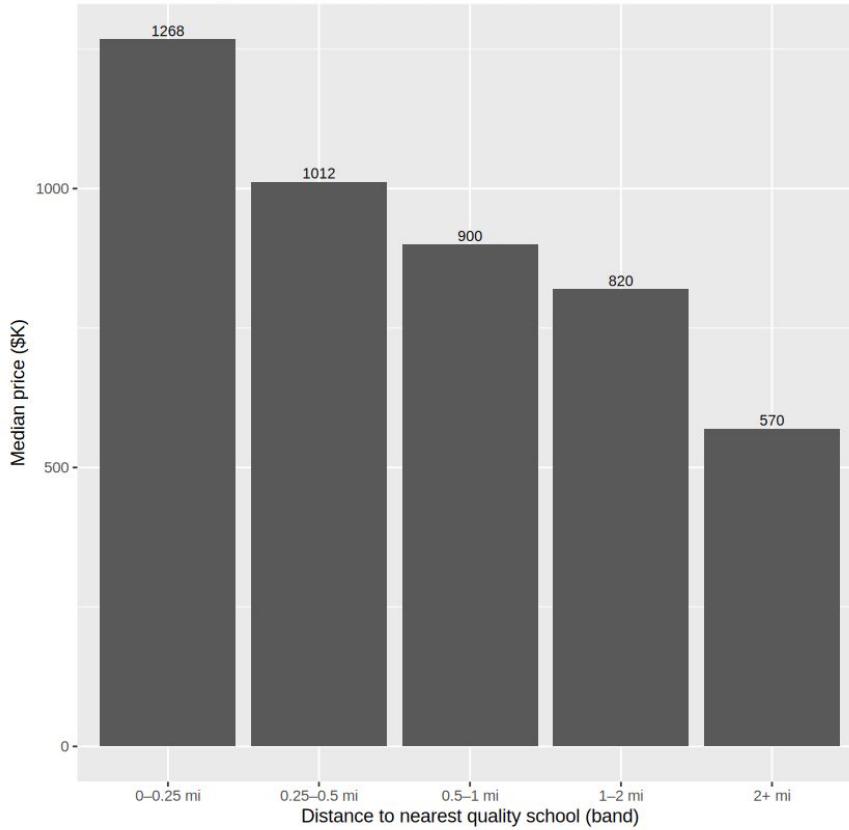
• • •

### Lot Size and Property Type on Price Prediction

All Property Types in Oakland



Median home price by distance to nearest quality school



# Redlining HOLC Grade



# Crimes in the last 14 days

