Pittsburgh Housing Price Analysis

By: Maxwell Snodgrass

The Research Question and General Approach

What is the most expensive neighborhood to buy a house in Pittsburgh?

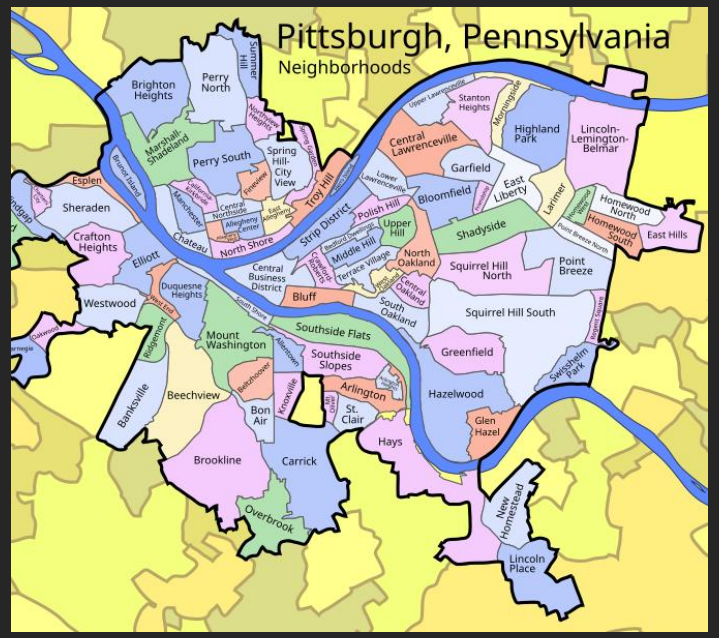
- Challenges
  - Rising house prices from 2012-2025
  - Types of sales (inheritance, corporation transfer, etc…)
  - Different housing characteristics

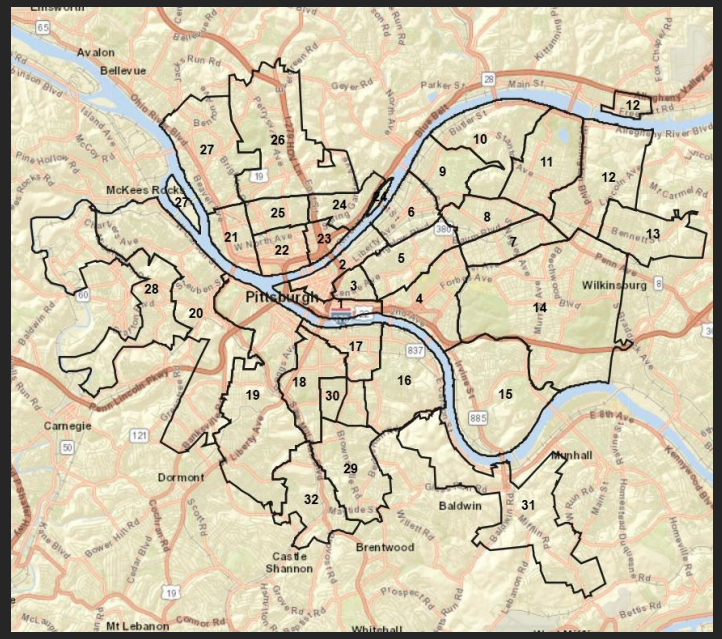
The Approach

1. Filter and clean data to include valid sales in Pittsburgh city limits
2. Visualize data and deal with problematic observations
3. De-trend and standardize prices based off the relationship between year and house prices
4. Merge sales data to housing characteristic data
5. Regress standardized house prices on housing characteristics
6. Obtain residuals (price deviations unexplained by home attributes)
7. Group residuals by neighborhood and take the average to create the final ranking

Data and Variables

- Allegheny County Housing Transaction Data
  - Pricing, time, and neighborhoods
    - Price
    - Municipality
    - Year
- Assessors Data Assessor Records
  - Housing characteristics
    - Area
    - Number of Stories
    - Year Built
    - Exterior Finish
    - Basement
    - Condition
    - Total Rooms
- Each dataset contains ‘PARID’
  - Merging



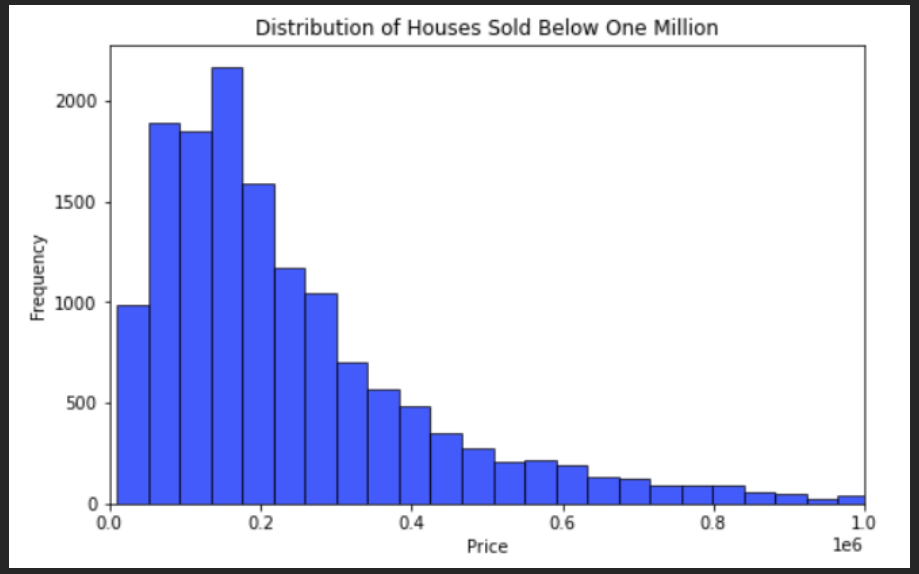


Municipality vs. Neighborhood

- We will use municipality as a proxy for neighborhood

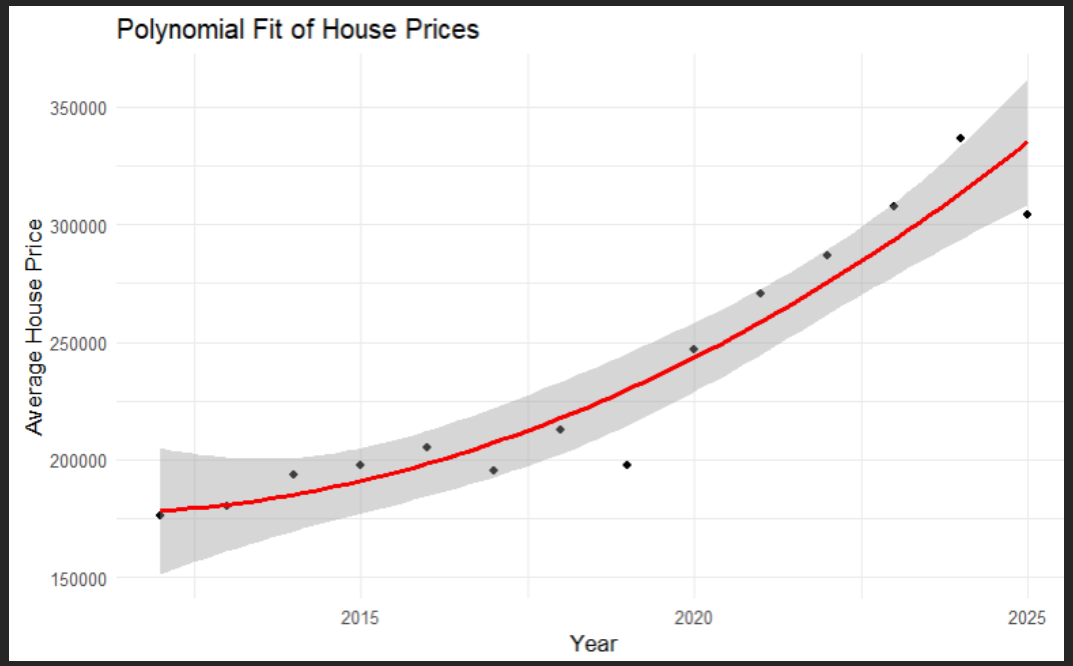
Filtering and Cleaning

1. Filter only “Valid Sales”
   - Avoid inheritance, corporate transfers, etc…
2. Filter only Pittsburgh city limit data
   - Only municipalities labelled as wards in Pittsburgh can stay
3. Filter out all house sales less than $10,000
   - This safely eliminates any possibility of potential data entry errors.
4. Account for missing values



Visualization and Outliers

- Two observations are over $60,000,000
  - Eliminate these two observations
  - They severely affect the data (mean changes by $13,109 dollars after removal)




```
Call:
lm(formula = PRICE ~ poly(YEAR, 2), data = sales)

Residuals:
     Min       1Q   Median       3Q      Max
 -321180  -130888   -67652    53073 17058145

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)      255764       2331 109.728  < 2e-16 ***
poly(YEAR, 2)1  6216803     281730  22.067  < 2e-16 ***
poly(YEAR, 2)2  1525128     281730   5.413 6.28e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 281700 on 14606 degrees of freedom
Multiple R-squared:  0.03414,   Adjusted R-squared:  0.03401
F-statistic: 258.1 on 2 and 14606 DF,  p-value: < 2.2e-16
```

Rising House Prices

- Time Trend Scatterplot
- Orthogonal Polynomial Regression Output
- Small R-Squared
  - Only about 3.4% of the price is explained by the year
- Nonetheless the trend shows significance

PREDICTED	DETRENDED	STANDARDIZED[, 1]
182854.7	-25854.683	-0.091777401
264687.8	55312.191	0.196343890
182663.1	82336.922	0.292274656
182663.1	659836.922	2.342249449
182663.1	-107663.078	-0.382175925
182663.1	-136663.078	-0.485118382
182663.1	-87663.078	-0.311181127
182663.1	-130663.078	-0.463819943
182663.1	-147663.078	-0.524165521
182663.1	262336.922	0.931227838
182663.1	-136163.078	-0.483343512
182663.1	1567336.922	5.563638409
182663.1	-102763.078	-0.364782200
182663.1	-102763.078	-0.364782200
182663.1	87336.922	0.310023356
182663.1	-162663.078	-0.577411620
182663.1	-132663.078	-0.470919423
182663.1	-32663.078	-0.115945433
182663.1	-32663.078	-0.115945433

Rising House Prices

- De-Trend
  - Create predicted price based off regression
  - Subtract predicted price from actual price
  - Standardize
- New column represents house prices controlled for time trend

```
Call:
lm(formula = STANDARDIZED ~ LOTAREA + STORIES + YEARBLT + EXTERIORFINISH +
    BASEMENT + CONDITION + TOTALROOMS, data = sales_standard3)

Residuals:
    Min      1Q  Median      3Q     Max
-2.7705 -0.3849 -0.1362  0.2258 10.8059

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    -1.063e+01  3.825e-01 -27.791  < 2e-16 ***
LOTAREA         2.656e-05  1.754e-06  15.141  < 2e-16 ***
STORIES         2.994e-01  1.432e-02  20.915  < 2e-16 ***
YEARBLT         5.101e-03  1.922e-04  26.542  < 2e-16 ***
EXTERIORFINISH  2.110e-02  3.890e-03   5.425 5.89e-08 ***
BASEMENT       -1.522e-01  5.696e-03 -26.722  < 2e-16 ***
CONDITION      -4.746e-02  6.537e-03  -7.261 4.06e-13 ***
TOTALROOMS      1.404e-01  3.576e-03  39.264  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.686 on 14098 degrees of freedom
Multiple R-squared:  0.2164,    Adjusted R-squared:  0.216
F-statistic: 556.3 on 7 and 14098 DF,  p-value: < 2.2e-16
```

Housing Characteristics

- Regress standardized price on housing characteristics in multiple linear regression
  - Statistical significance on all variables

The remaining residuals after this regression would represent price deviations unexplained by home attributes and unaffected by time, isolating the neighborhood effect.

Results

Top 5 Most Expensive Wards

MUNIDESC_x <chr>	mean_residual <dbl>
14th Ward - PITTSBURGH	0.6103213
7th Ward - PITTSBURGH	0.6050941
1st Ward - PITTSBURGH	0.5374408
9th Ward - PITTSBURGH	0.4326462
2nd Ward - PITTSBURGH	0.4224473
22nd Ward - PITTSBURGH	0.4119094

Homes in 14th Ward - PITTSBURGH **(Squirrel Hill and Point Breeze)** are priced on average about $137829.79 higher than the average house price in Pittsburgh when adjusting for rising house prices and controlling for housing characteristics.

Top 5 Least Expensive Wards

MUNIDESC_x <chr>	mean_residual <dbl>
20th Ward - PITTSBURGH	-0.3859240
26th Ward - PITTSBURGH	-0.3948611
29th Ward - PITTSBURGH	-0.4698300
30th Ward - PITTSBURGH	-0.5646511
12th Ward - PITTSBURGH	-0.5650727
13th Ward - PITTSBURGH	-0.6844009

6 rows

Homes in 13th Ward - PITTSBURGH **(Homewood and East Hills)** are priced on average about $154559.29 lower than the average house price in Pittsburgh when adjusting for rising house prices and controlling for housing characteristics.

The difference in average house price between 14th Ward – PITTSBURGH **(Squirrel Hill and Point Breeze)** and 13th Ward - PITTSBURGH **(Homewood and East Hills)** when controlling for rising prices and household characteristics is $292389.10.

Explanations of Differences

- When housing characteristics and time are controlled for, there is still unknown variation
- Potential sources of variation
  - Crime Rate
    - Perceived neighborhood safety
  - Demographics
    - Racial concentrations
    - Population density
  - Environmental
    - Proximity to green spaces
    - Flood risk
  - Other
    - Employment opportunities
    - School district prestige
    - Many others…