

A photograph of the Seattle skyline at sunset. The Space Needle is prominent on the left, with its observation deck lit up. Behind it is a dense cluster of modern skyscrapers of various heights and architectural styles. In the far distance, the majestic Mount Rainier rises against a hazy orange and yellow sky. The foreground shows some greenery and the tops of buildings.

Predicting King County House Prices

Exploratory Data Analysis

by Hannah Langen & Louisa Rudow

Objective

What were the most important factors in determining the house prices in King County between 2014 and 2015?

Method

1 Descriptive analysis
using Tableau

2 Linear regression model
using Pandas

Dataset



Data from houses sold King County, WA (USA)

Timeframe:	5/2014 - 5/2015
Columns:	21
Entries:	21.597
Categorical:	14
Numerical:	7

Descriptive Analysis

„Nowhere, though, has there been a more concerted effort to create a San Francisco-like tech scene with fewer downsides than in Seattle, the country’s second-biggest tech hub by some measures.“

Seattle Times 8.10.2015

„It’s getting harder to find a home for less than \$500,000 and easier to find one with a price tag of more than \$1 million.“

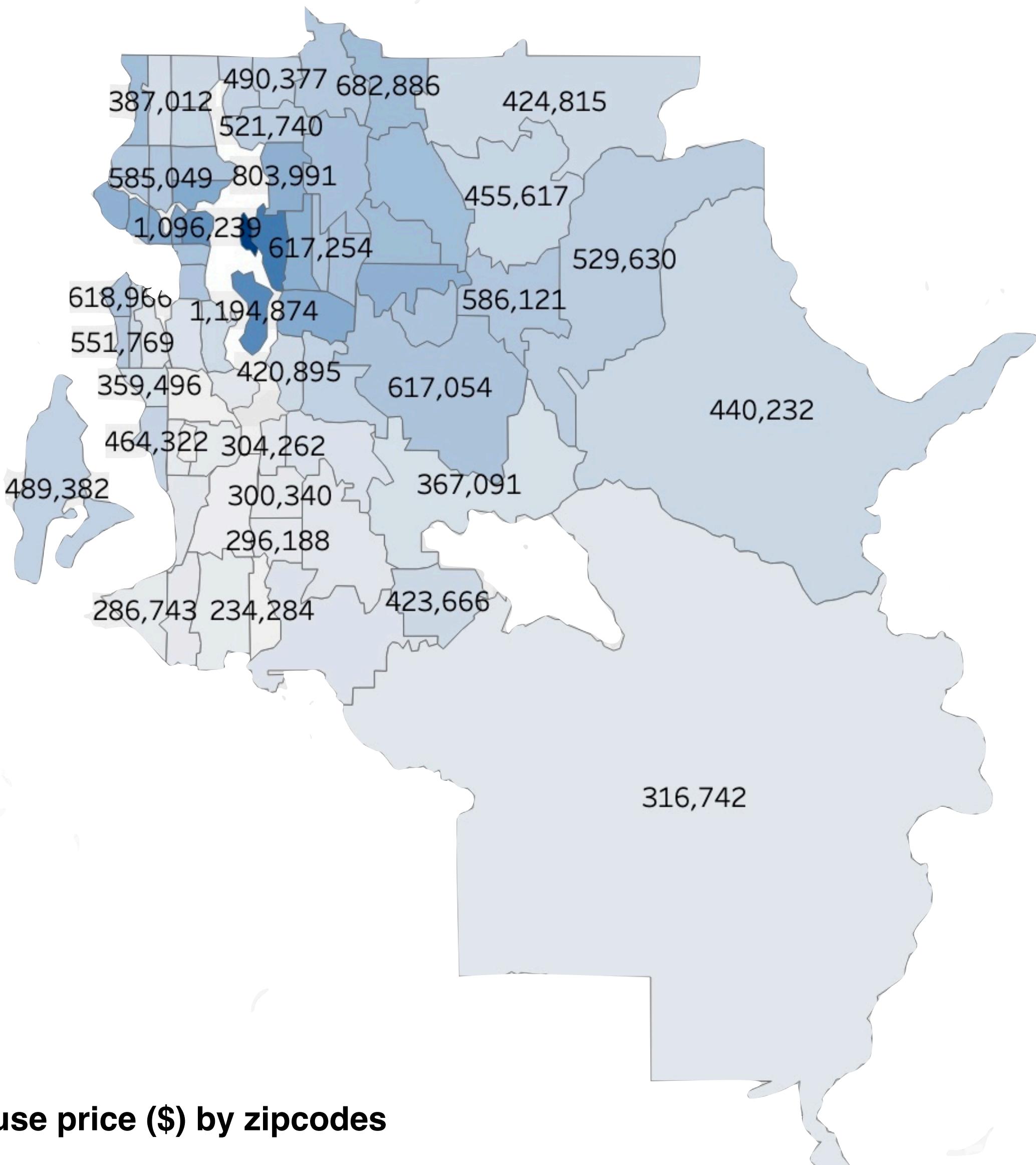
Seattle Times, 19.01.2022

„From 2011 to 2020, according to the data, the number of jobs in the Puget Sound area grew 30% while the number of housing units grew just 19%.“

The Business Journals, 24.01.2022

Some Assumptions:

1. Seattle is more expensive than the surrounding area
2. Location & house size are important features
3. House size becomes less relevant in higher price ranges



Average house price (\$) by zipcodes

AVG(Price)

234,284

2,161,300

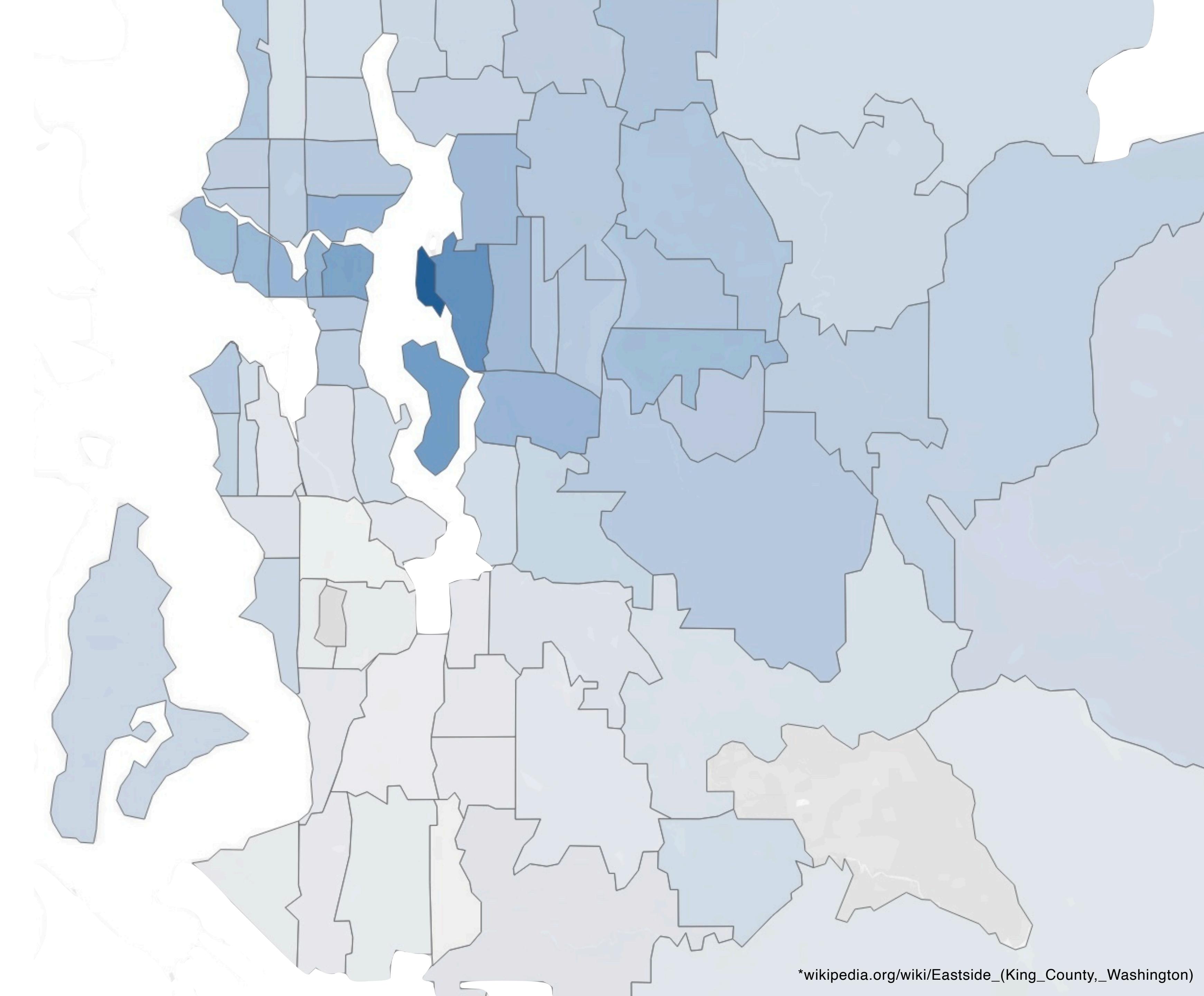
How are housing prices distributed throughout King County?

- **Clear pattern:** Concentration of higher housing values in the northwest corner of the county
- Most expensive area is not Seattle, but the **Eastside** area
- King County avg: ~ \$540.310
- National avg.: ~ \$349.375*

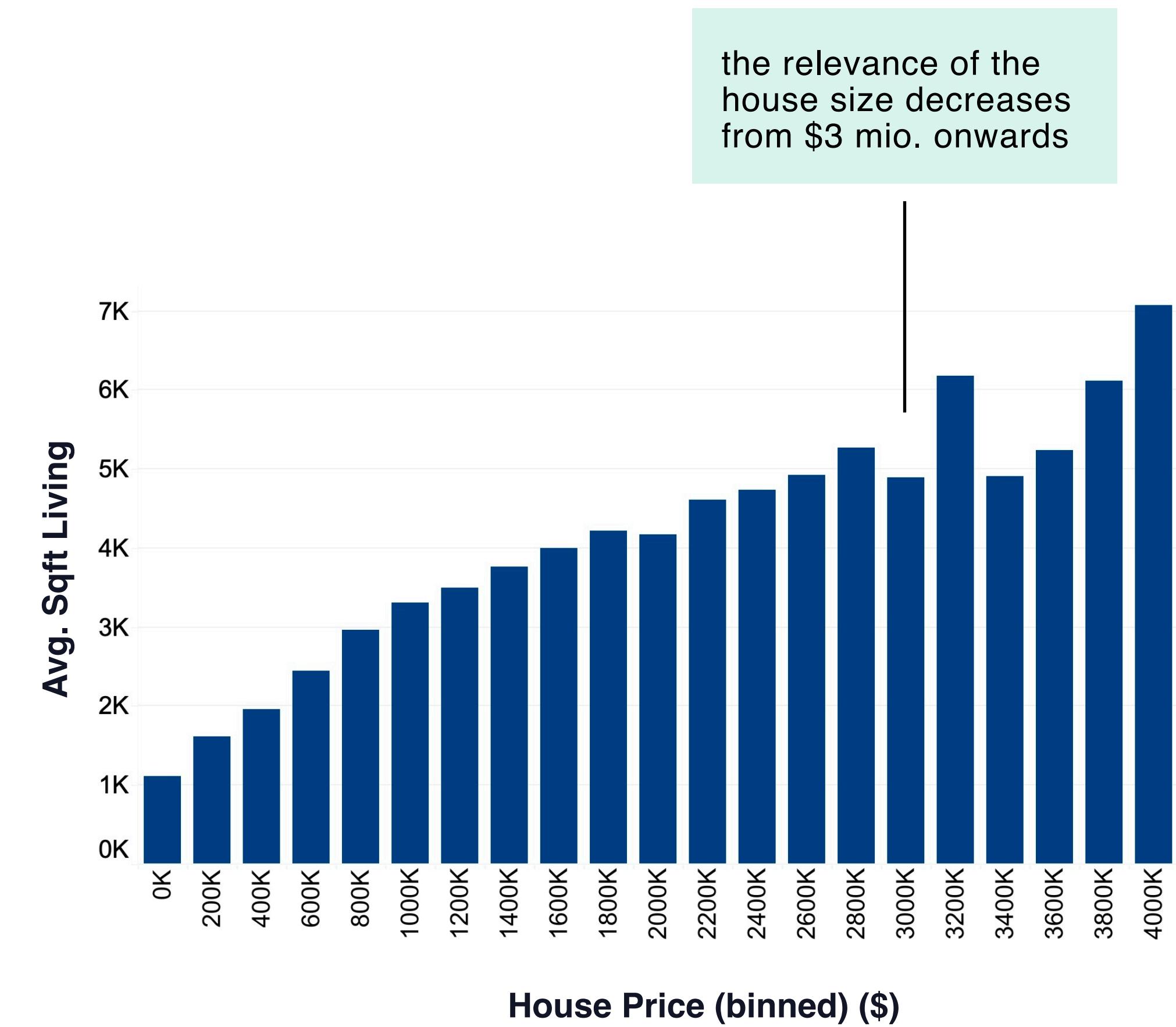
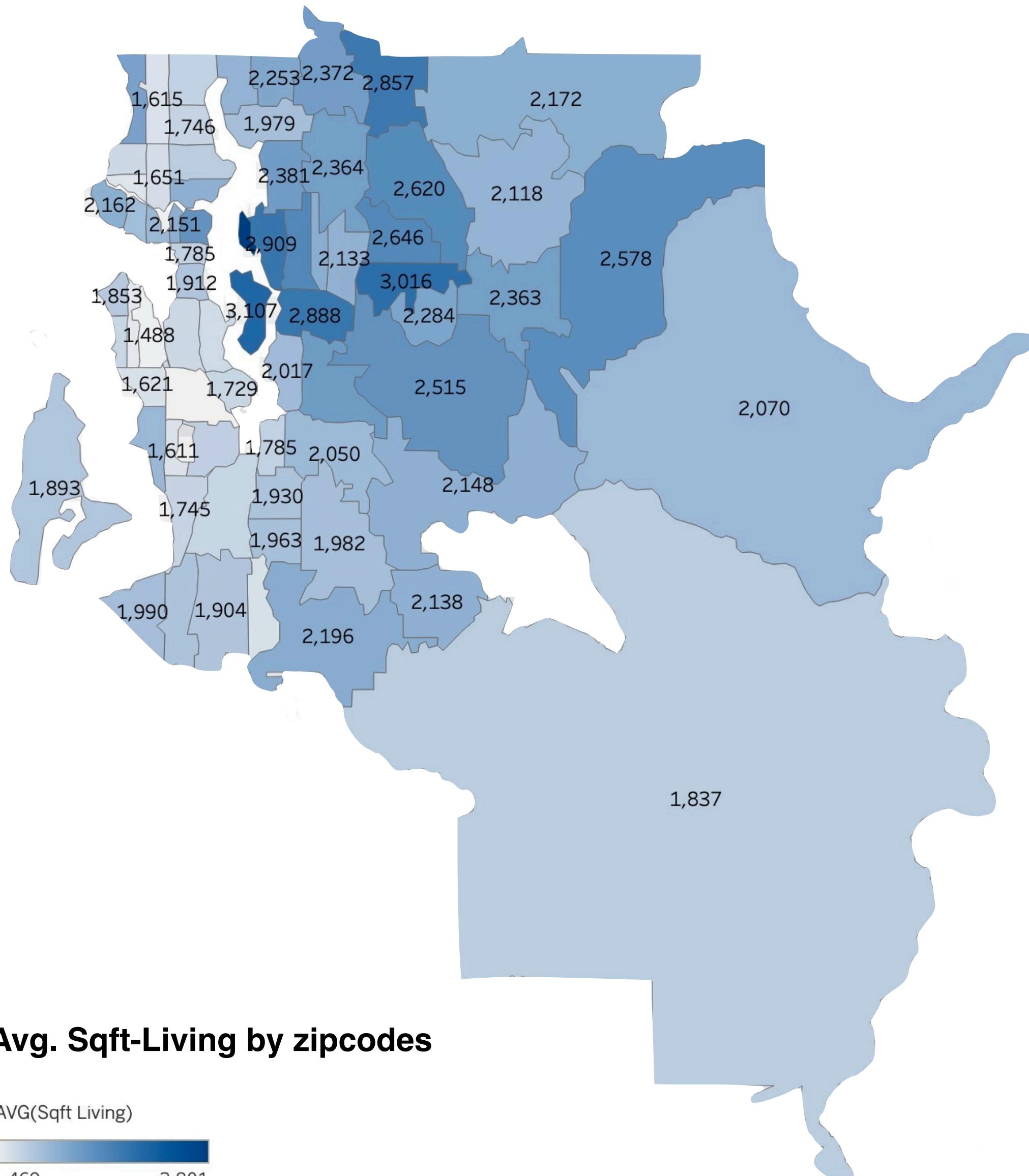
*<https://fred.stlouisfed.org/series/ASPUS>

Eastside

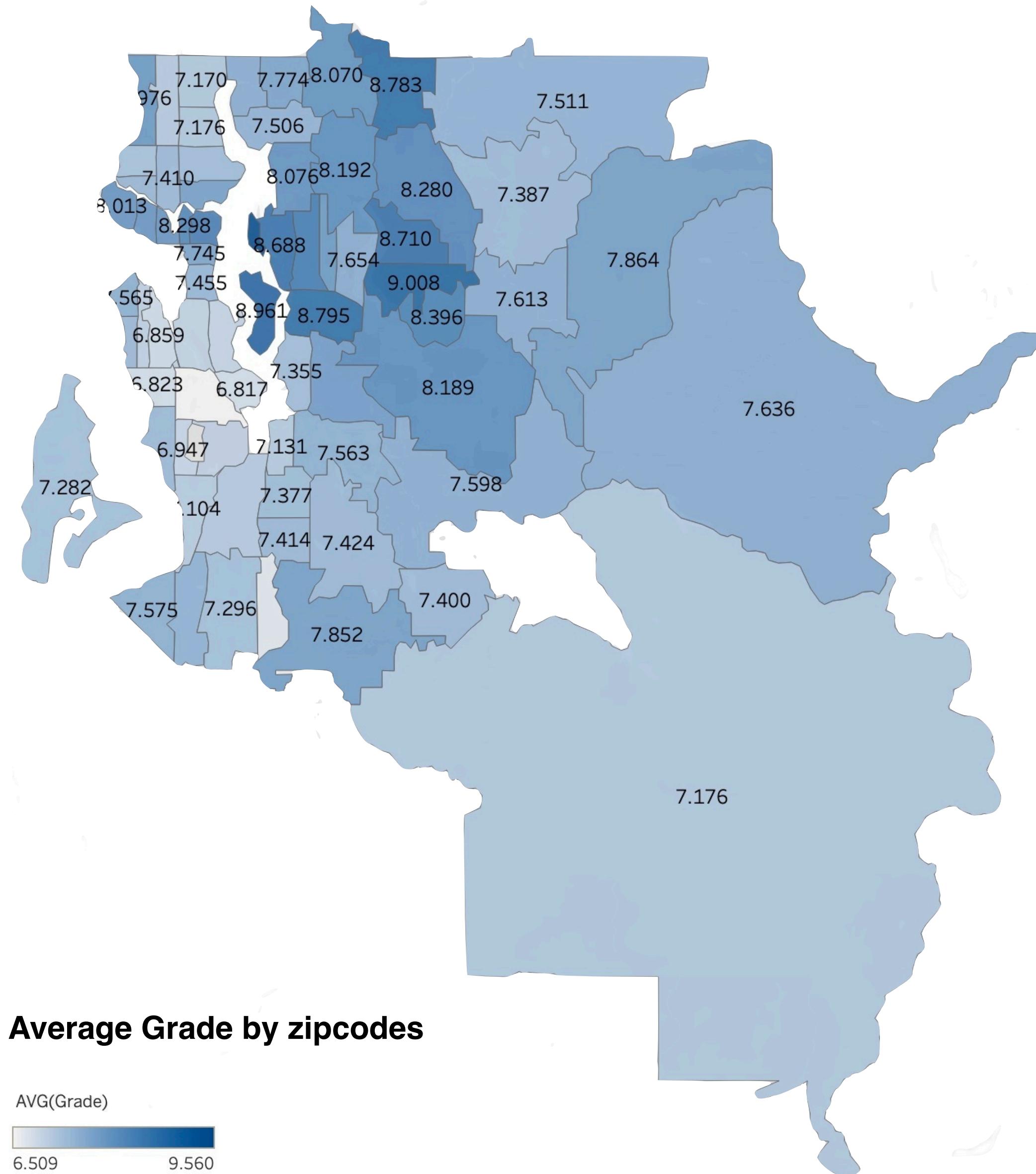
- major technology center: software, wireless, gaming industries, e.g. Microsoft, T-Mobile US, Nintendo US*
- hub for biotech and medical companies
- High profile personalities & company founders like Bill Gates, Jeff Bezos and Bill Boing live in this area



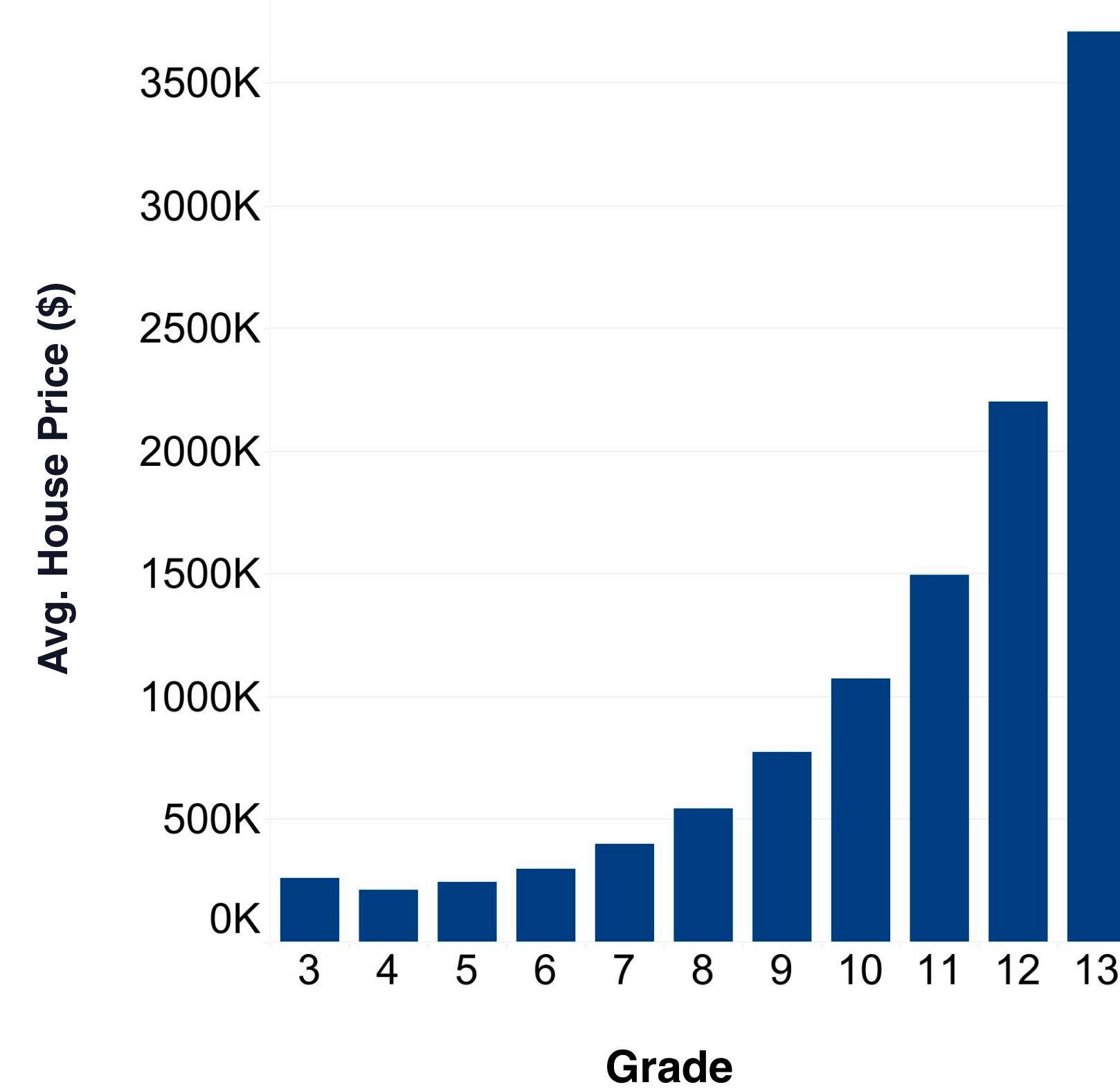
Can we see the same pattern in the avg. House Size?



How about the grading?



Grade: Classification by construction quality which refers to the types of materials used and the quality of workmanship. Buildings of better quality (higher grade) cost more to build per unit of measure and command higher value.*



Linear Regression

Creating the model

Categorical

id
date
bedrooms
bathrooms
floors
waterfront
view
condition
grade
yr_built
yr_renovated
zipcode
lat
long

Numerical

sqft_living
sqft_lot
sqft_above
sqft_basement
sqft_living15
sqft_lot15

house price

Dropping unimportant columns

Categorical

- id
- date
- bedrooms
- bathrooms
- floors
- waterfront
- view
- condition
- grade
- yr_built
- yr_renovated
- zipcode
- lat
- long



Numerical

- sqft_living
- sqft_lot
- sqft_above
- sqft_basement
- sqft_living15
- sqft_lot15

Creating/changing columns

Categorical

~~id~~
~~date~~
bedrooms
bathrooms
floors
waterfront
view
condition
grade
yr_built
~~yr_renovated~~
~~zipcode~~
lat
long
area

Numerical

sqft_living
~~sqft_lot~~
~~sqft_above~~
sqft_basement
sqft_living15
sqft_lot15

house price

area

Grouped zipcode

very_affordable_area

affordable_area

medium_priced_area

expensive_area

very_expensive_area

Average price

< \$250.000

$\geq \$250.000 < 500.000$

$\geq \$500.000 < \750.000

$\geq \$750.000 < \$1.000.000$

$\geq \$1.000.000$

Data Transformation

Categorical

bedrooms
bathrooms
floors
waterfront
view
condition
grade
yr_built
lat
long
area

Hot Encoding

house price

Quantile Transformer

Numerical

sqft_living
sqft_basement

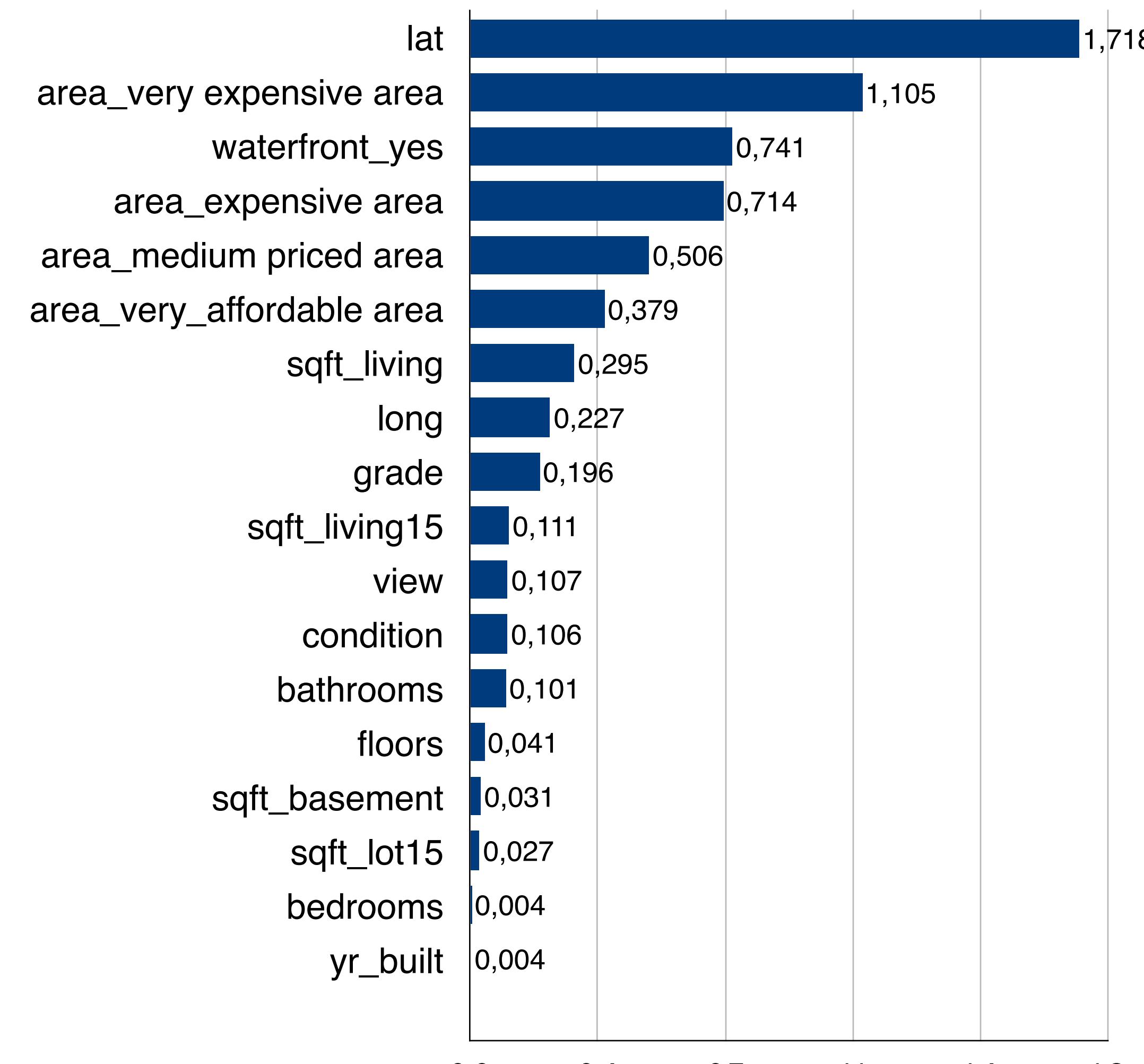
Standardize

sqft_living15
sqft_lot15

Standardize
Quantile Transformer

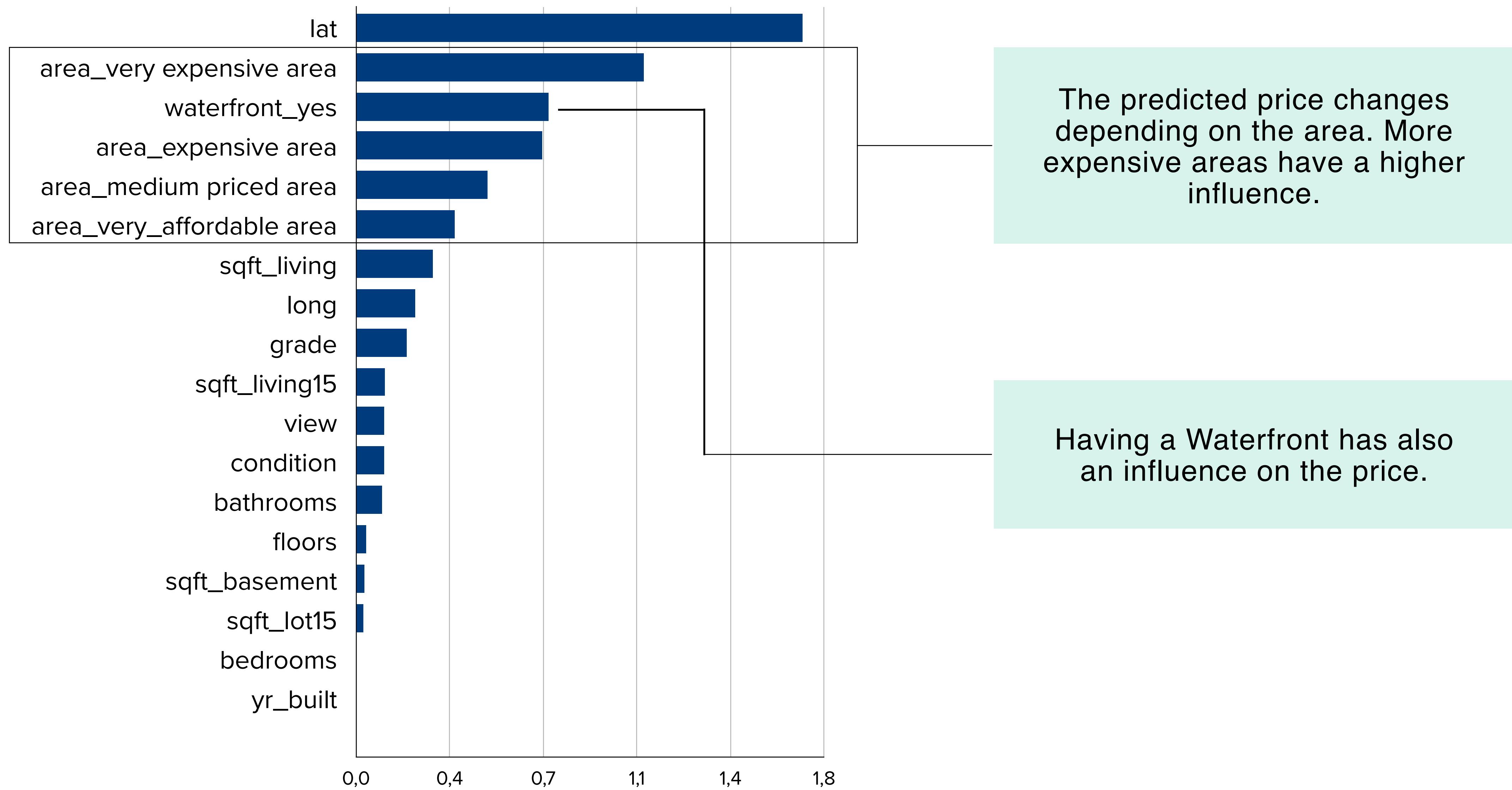
Results

82.6%
Adjusted R²



Feature Importance obtained from coefficients

Key Findings



Outlook

- Create coordinates using Lat/Long to have more accurate locations
- Use more data over a longer period of time (if available) to better assess the development
- Try different approaches (e.g. polynomial regression)

Thank you!