

What to Buy Where and When in King County?

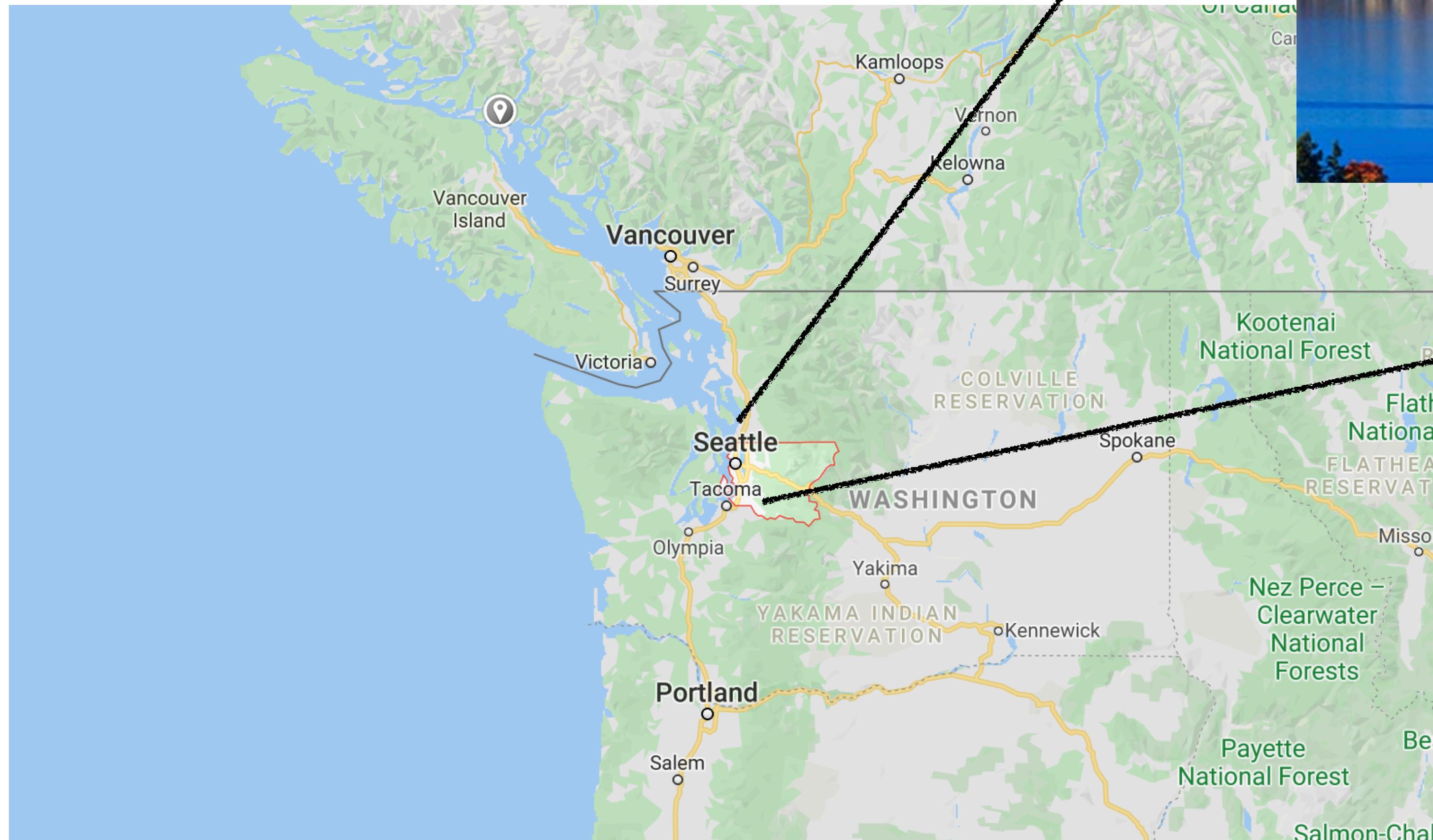
Marco Zagermann

Hamburg, June 10, 2020

Outline

1. King County
2. House prices: General parameters and data set
3. Five things you should know before you buy here!
4. A prediction model

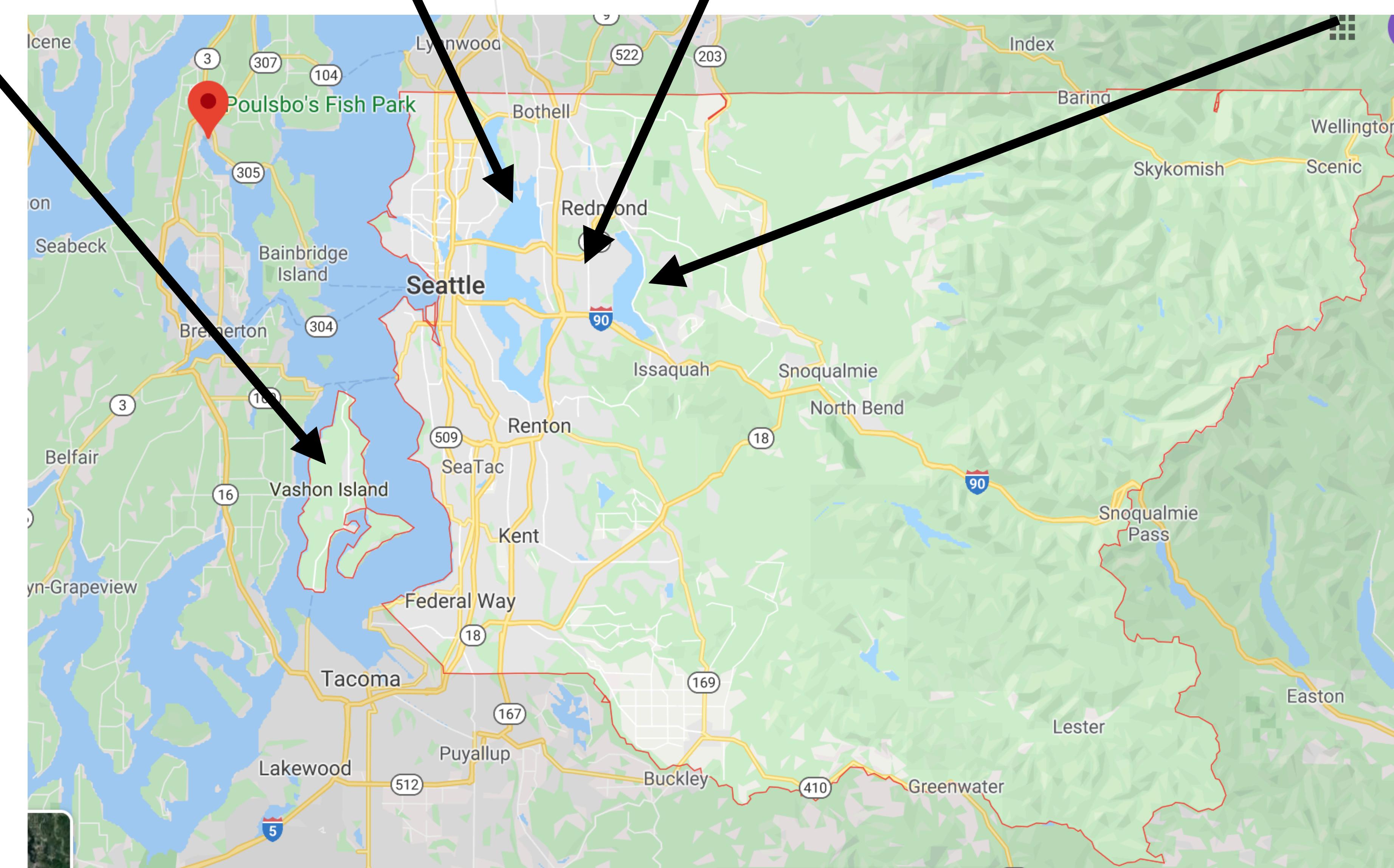
1. King County



Population: 2.2 M

Capital: Seattle

Vashon Island Lake Washington Bellevue Lake Sammamish



2. House prices: General parameters and data set

!



price	1.000000
sqrt_living	0.701511
grade	0.667093
sqft_above	0.603919
sqft_living15	0.586248
bathrooms	0.521651
view	0.405891
bedrooms	0.305679
lat	0.303124
waterfront	0.282324
floors	0.253959
yr_renovated	0.129272
sqrt_lot	0.083496
sqrt_lot15	0.081888
yr_built	0.051609
condition	0.041926
long	0.019851
id	-0.012410
zipcode	-0.056357
Name:	price, dtype: float64

Data from 2014/15

21597 Selling events

Direct importance of individual factors (correlations)

```
price           1.000000
sqft_living    0.701511
grade          0.667093
sqft_above     0.603919
sqft_living15  0.586248
bathrooms      0.521651
view            0.405891
bedrooms        0.305679
lat              0.303124
waterfront      0.282324
floors          0.253959
yr_renovated   0.129272
sqft_lot        0.083496
sqft_lot15     0.081888
yr_built        0.051609
condition       0.041926
long             0.019851
id               -0.012410
zipcode         -0.056357
Name: price, dtype: float64
```

Gives many important clues.

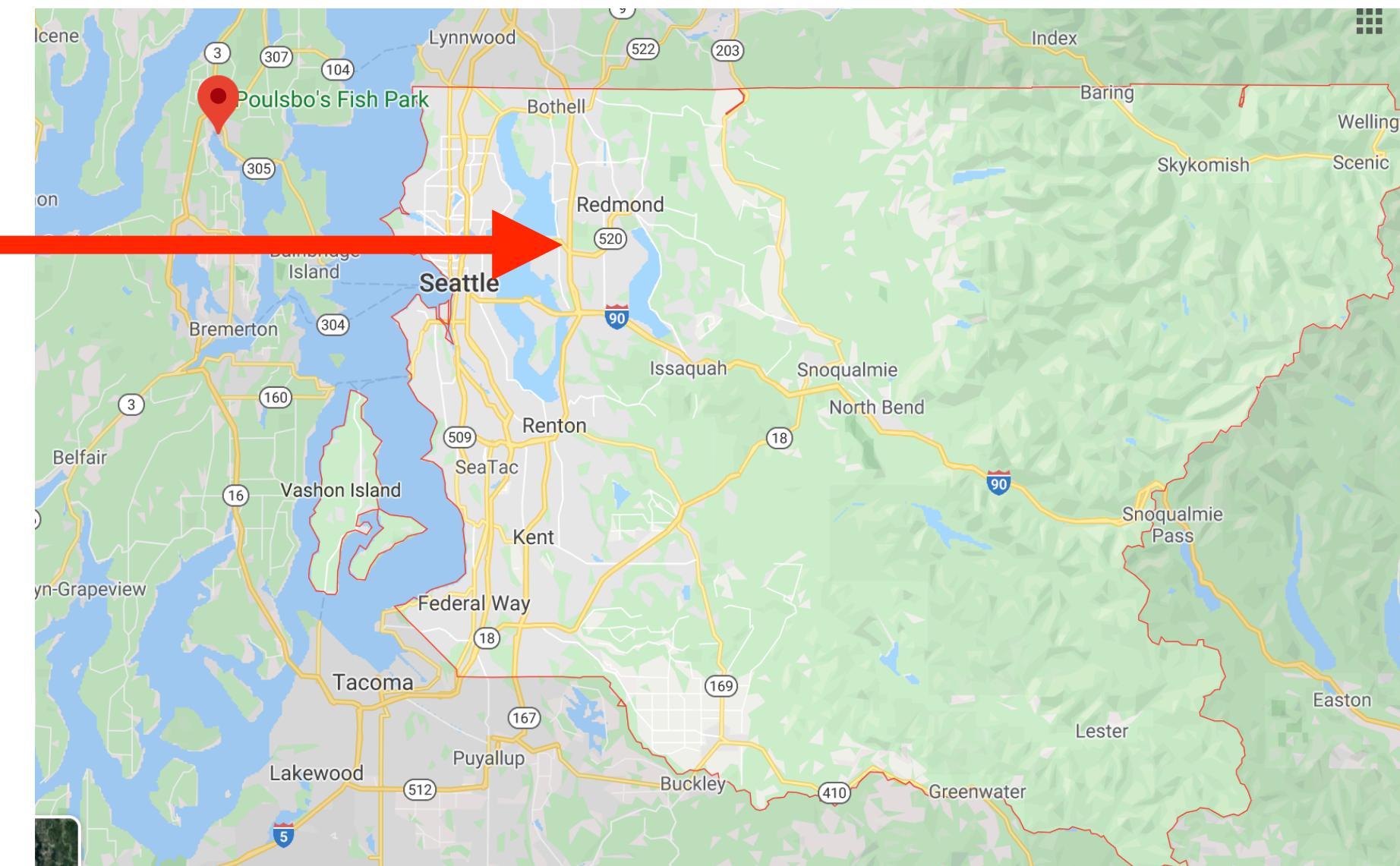
But: More useful information **hidden** in
combinations of various data !

3. Five things a house buyer should know here! Lake

(i) Tight budget?

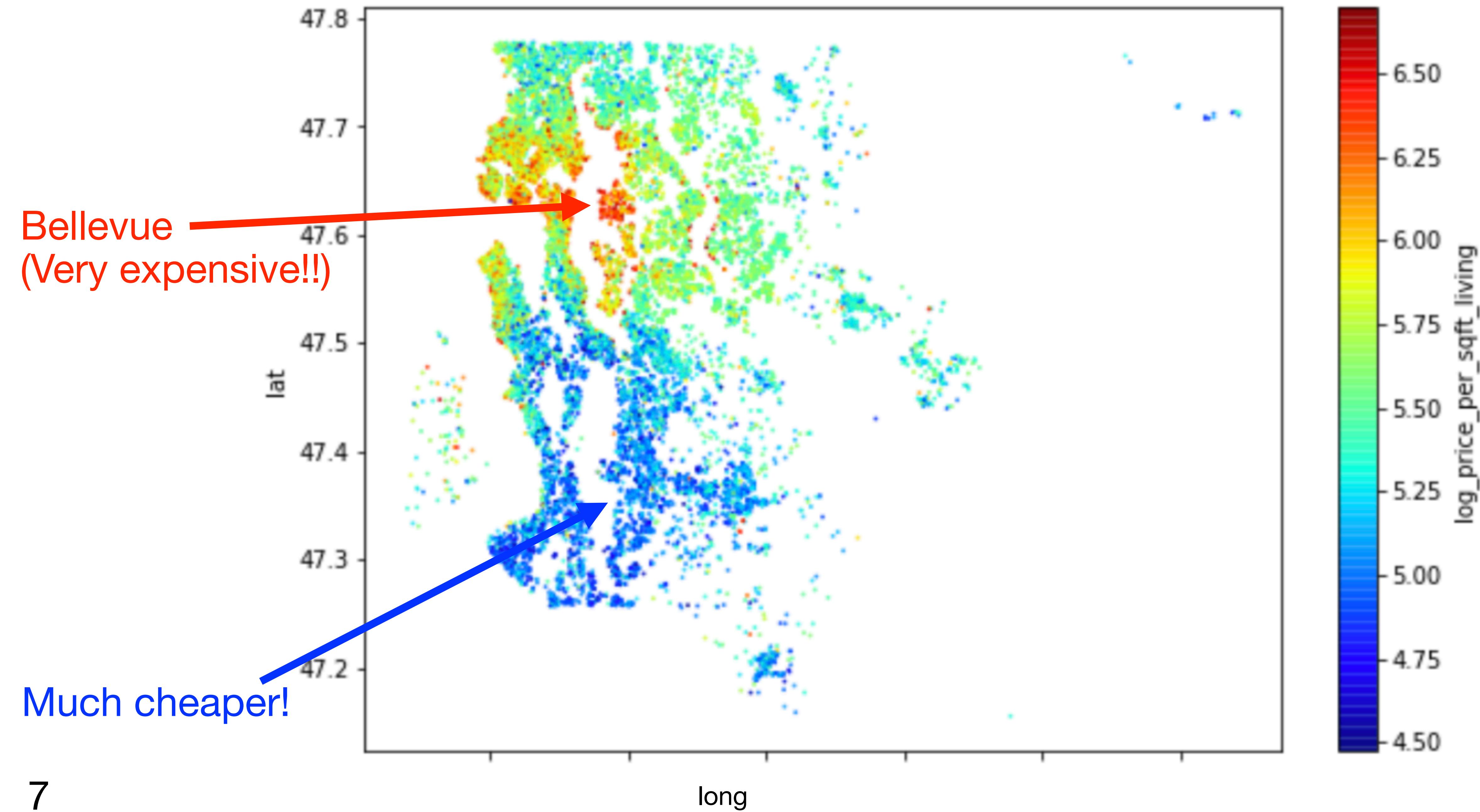
Then stay away from **Bellevue** as far as you can!

Bellevue —
Don't buy here!



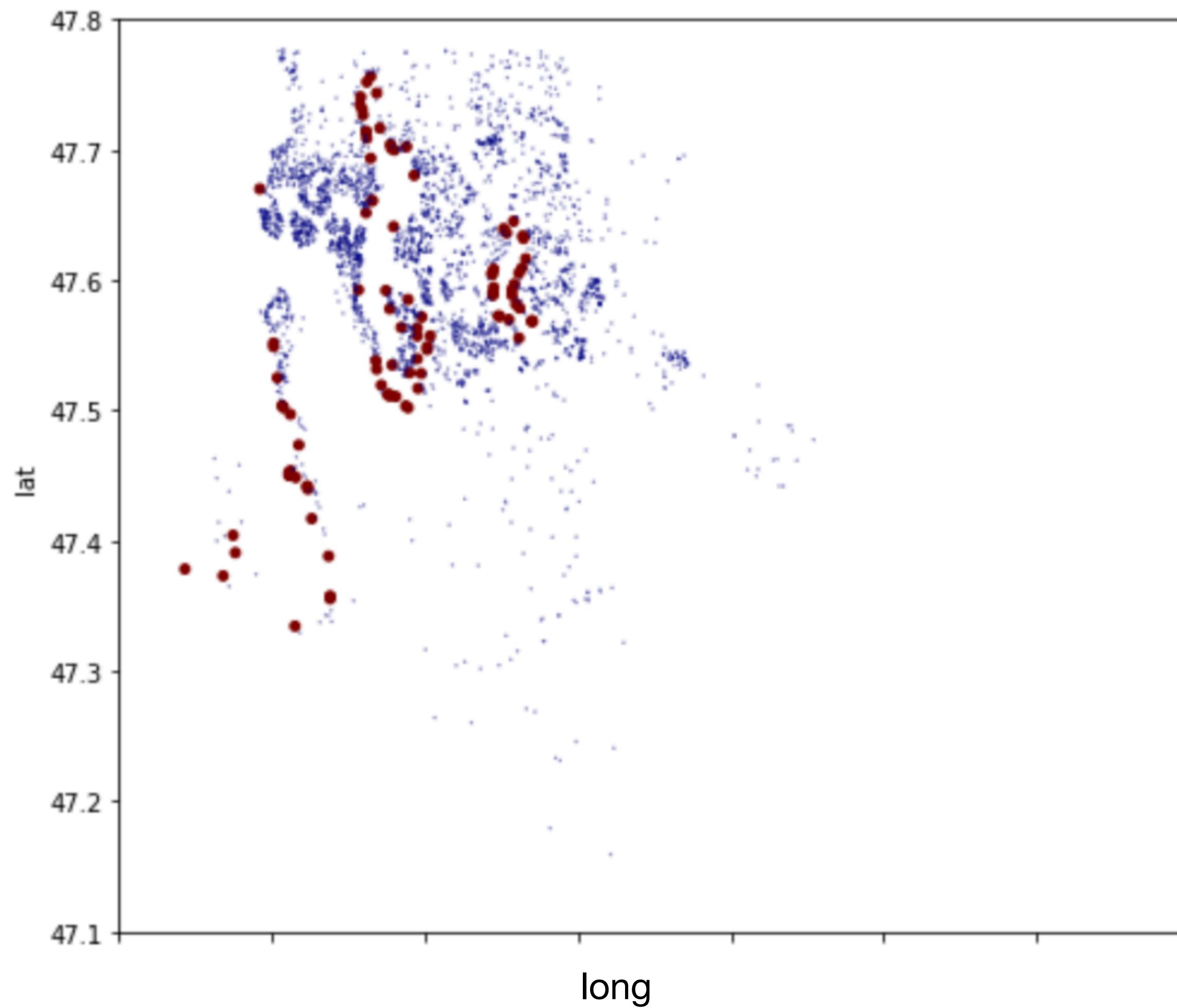
Source: Google Maps

Price per sqft_living

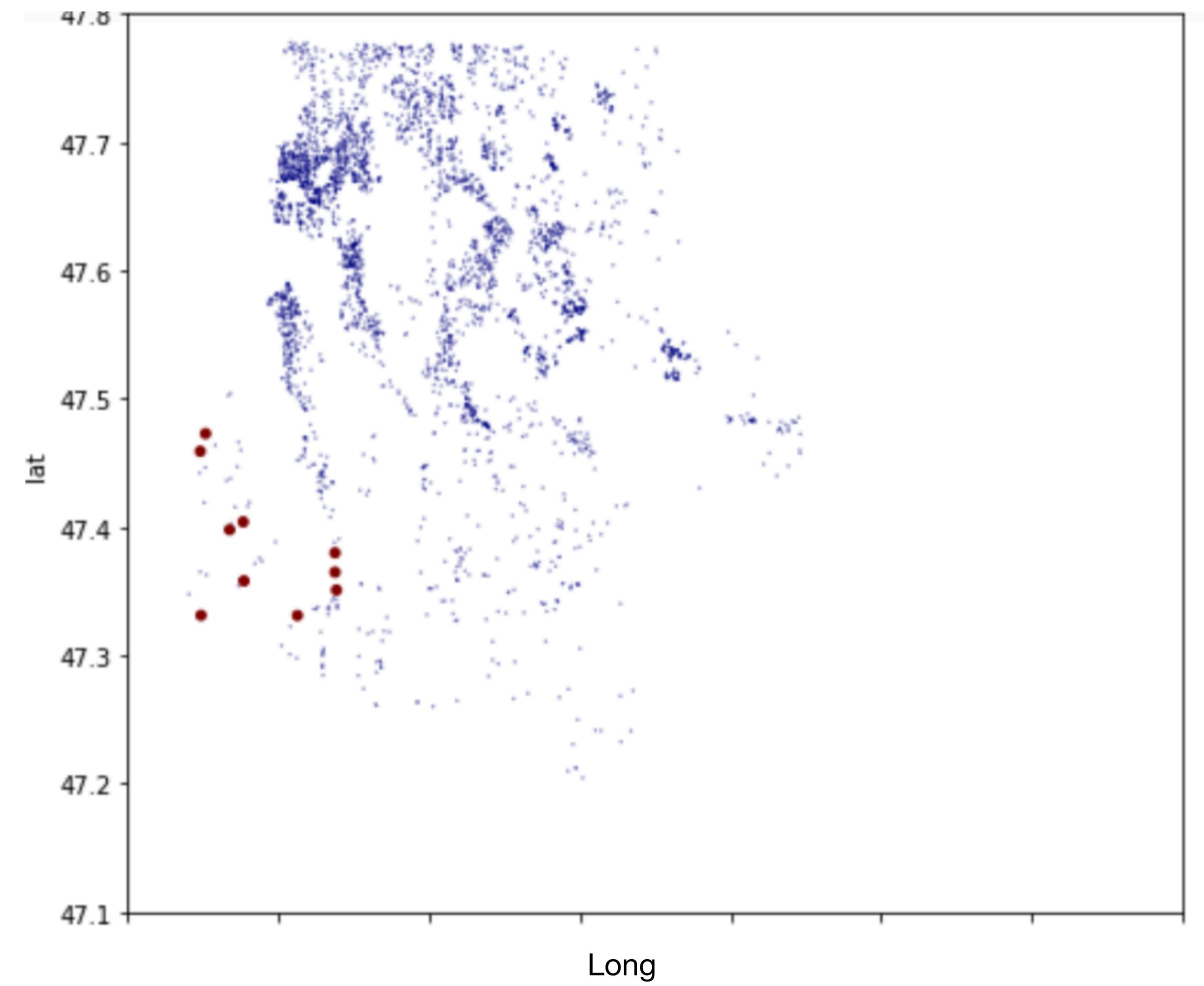


(ii) Small budget, but waterfront?

Very expensive (top 25 %)

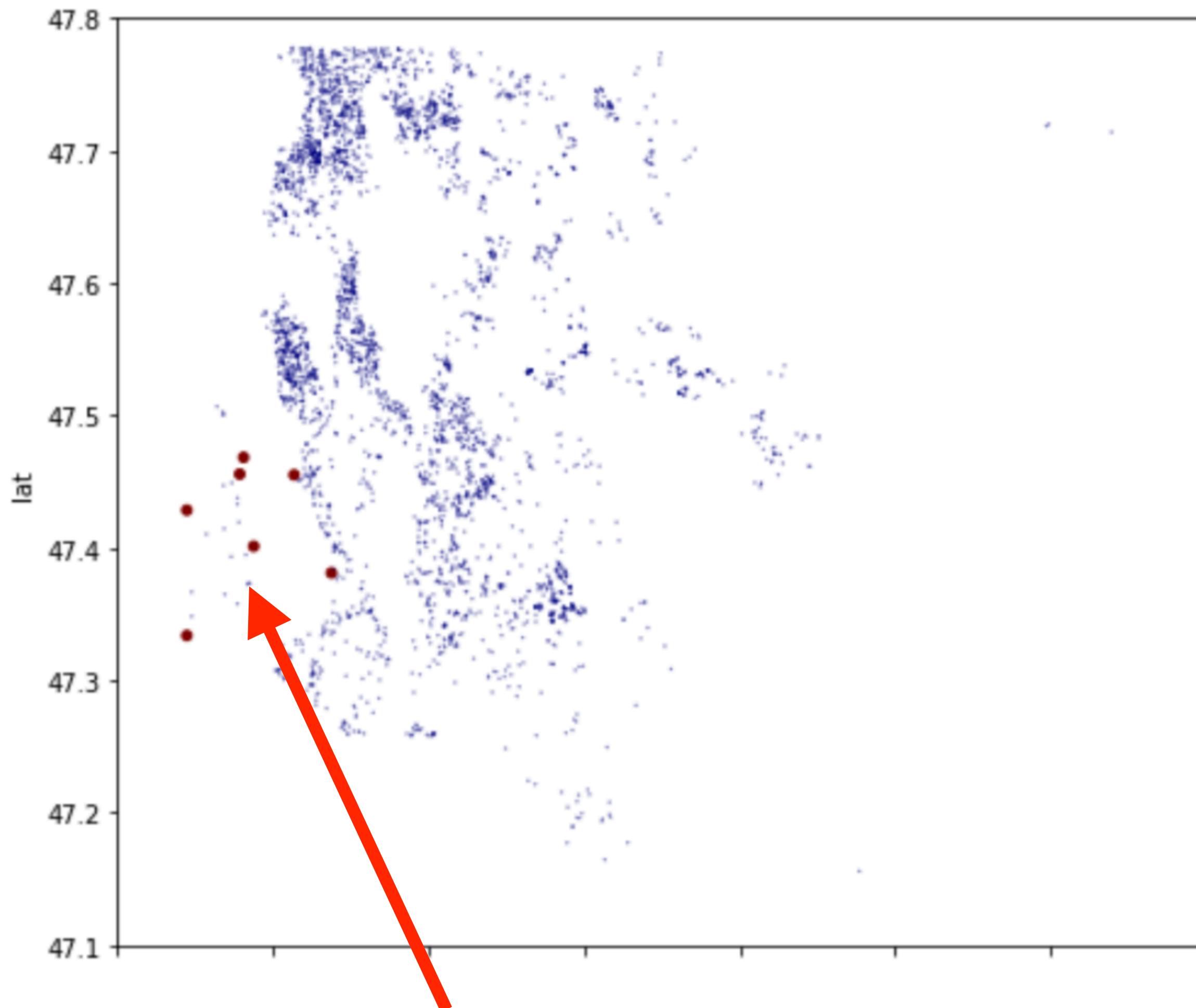


Expensive (top 50% - top 25 %)



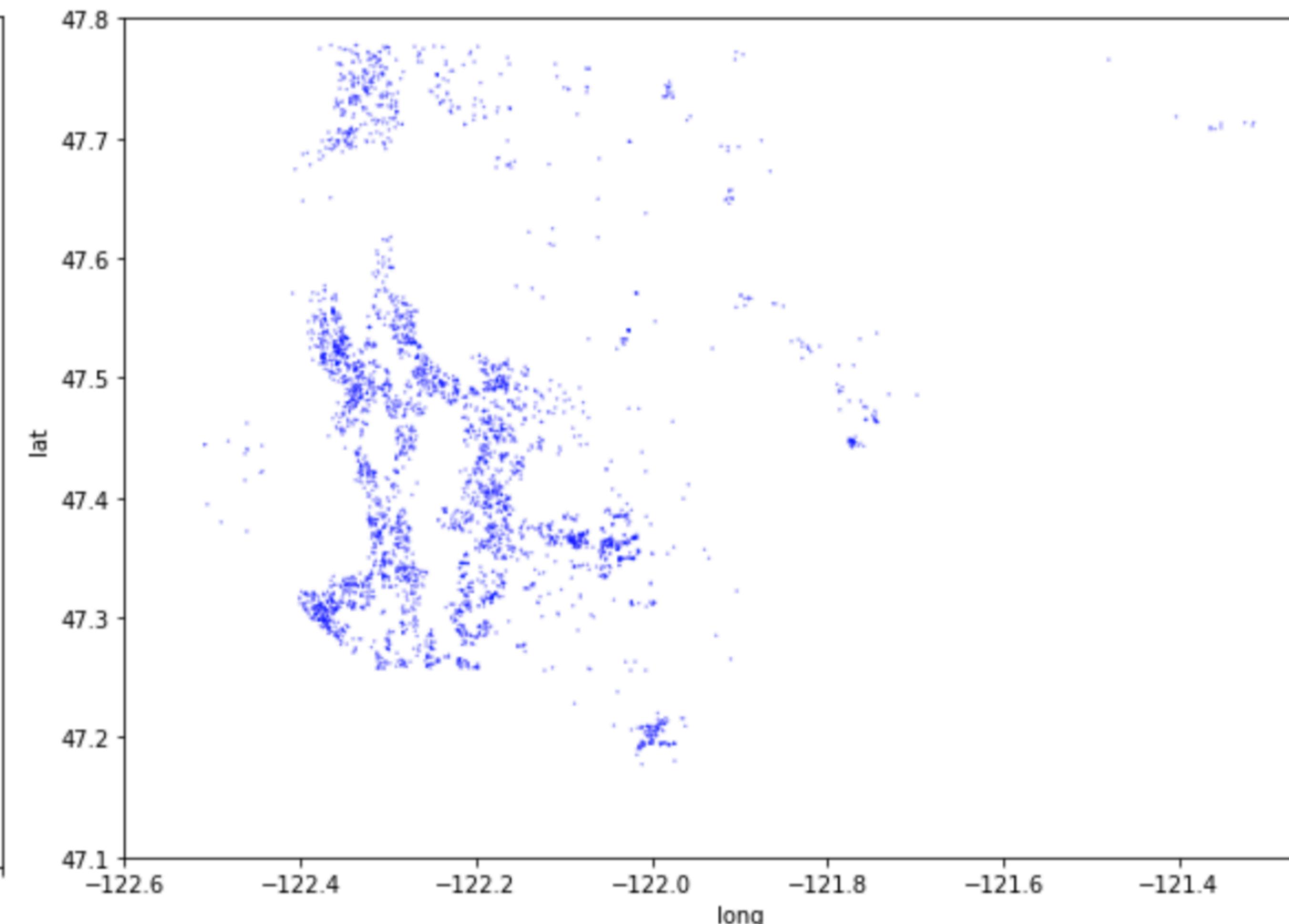
Red = waterfront

Cheap (top 75 % - top 50%)



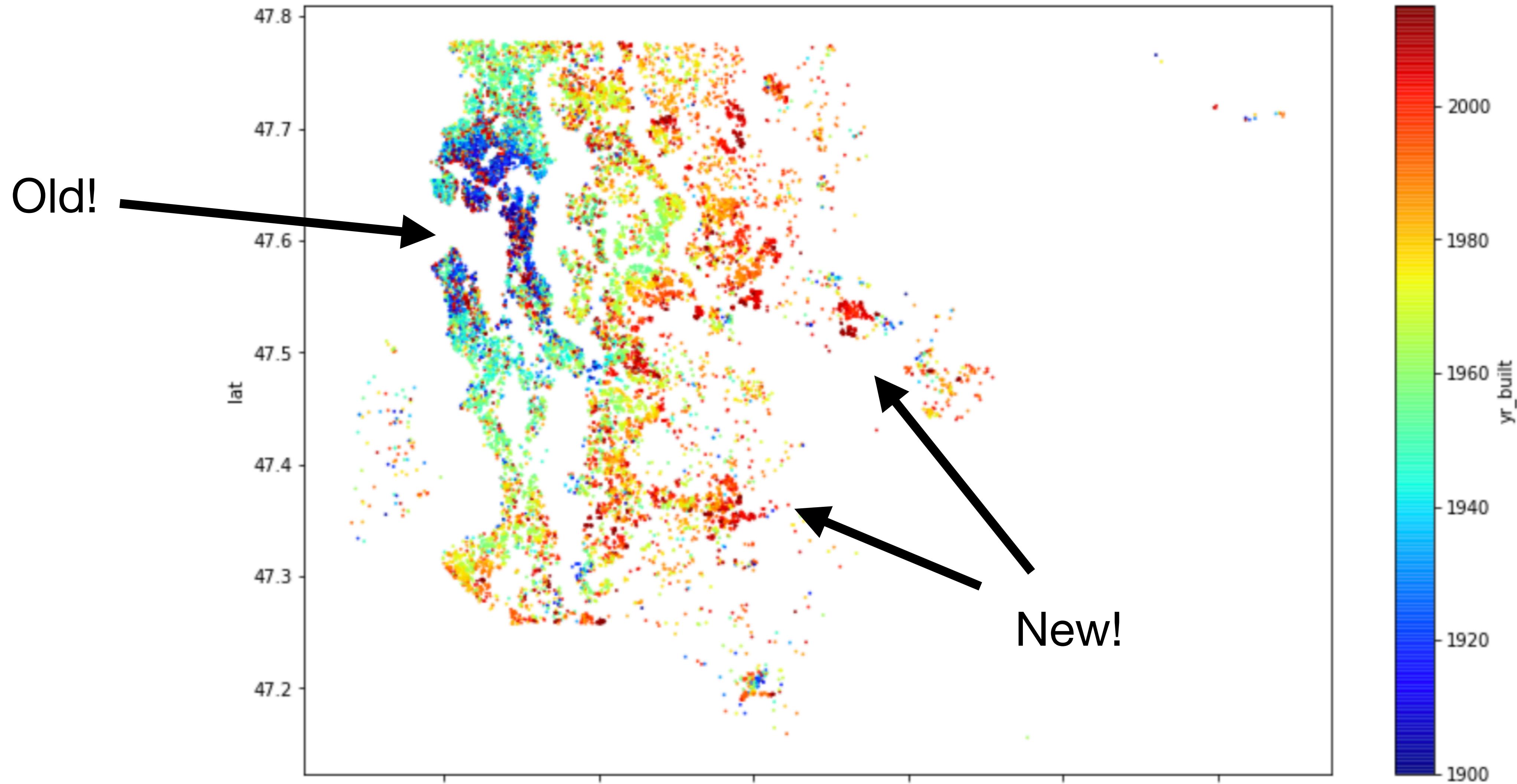
Vashon Island!

Very cheap (top 100 % - top 75 %)



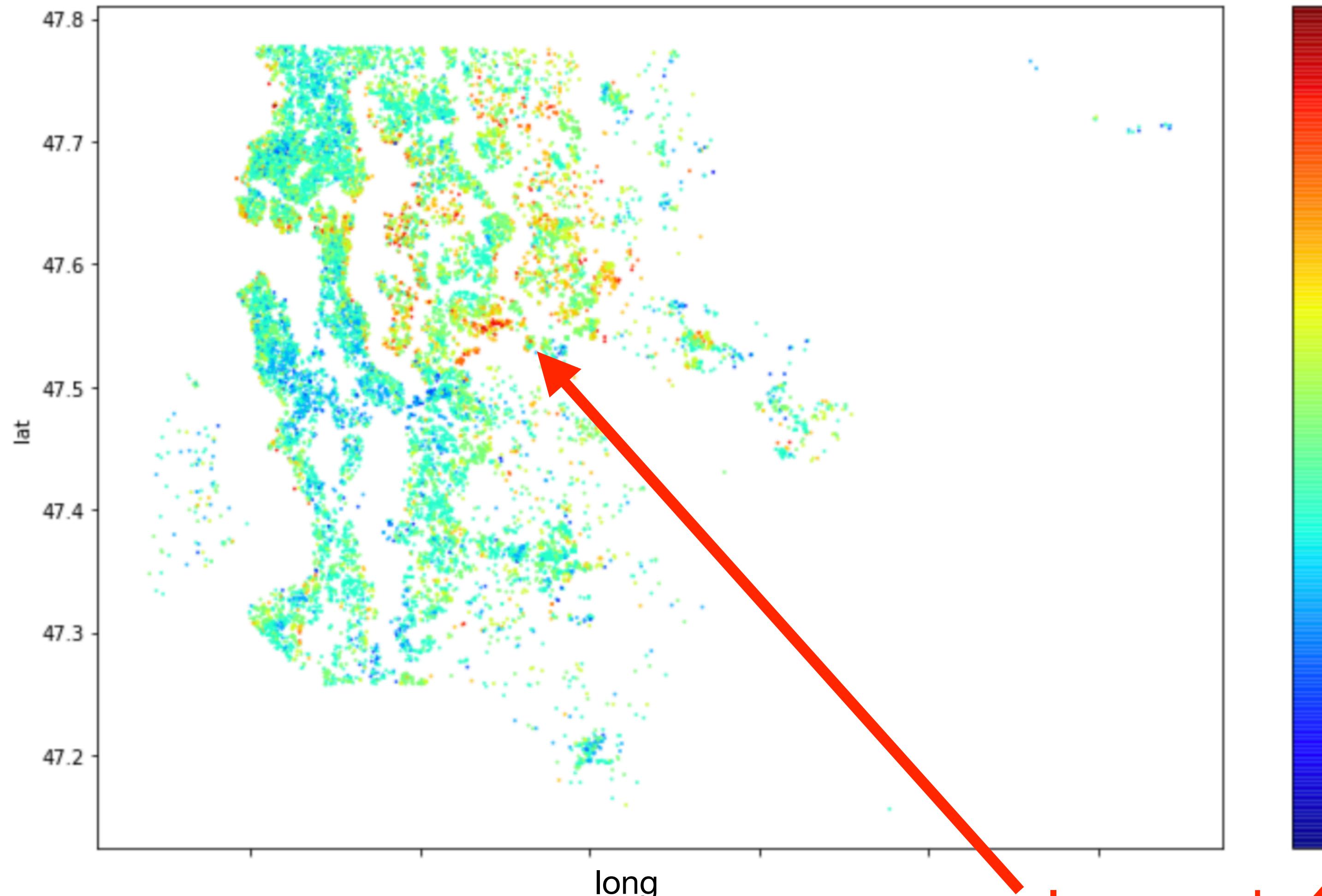
Red = waterfront

(iii) Big fan of old or new buildings?

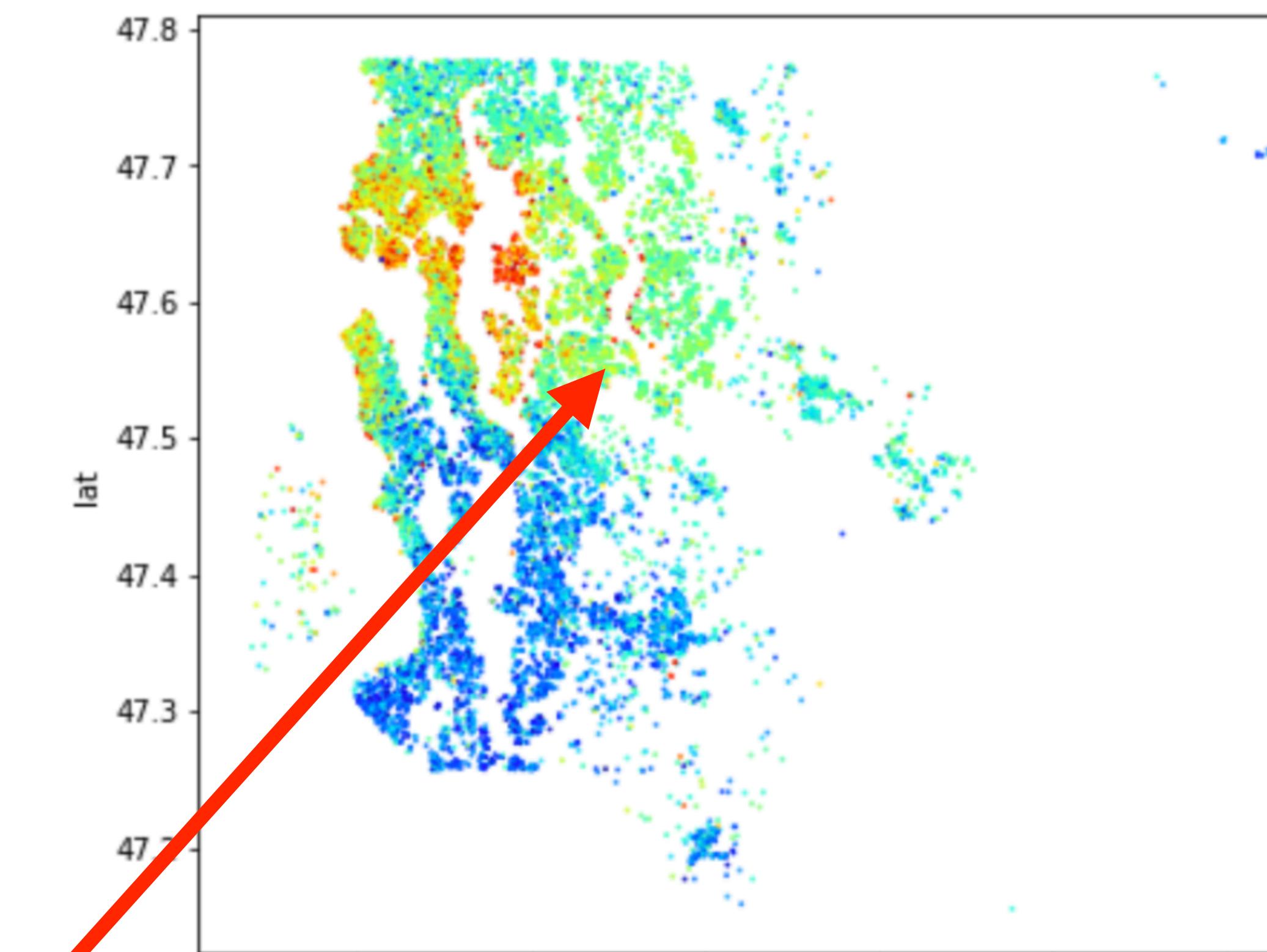


(iv) Highest grades but not highest price?

Grade



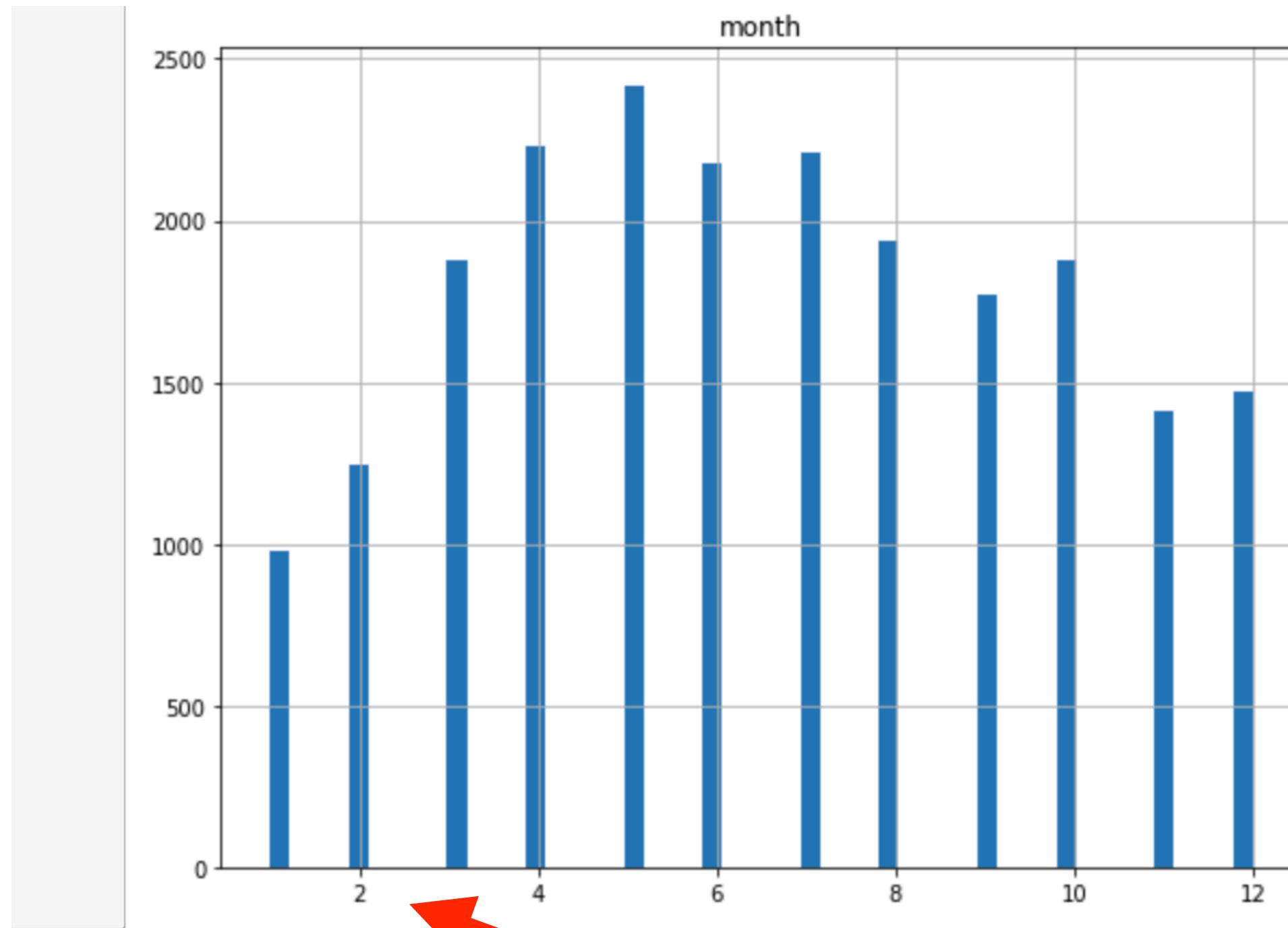
Price per living area



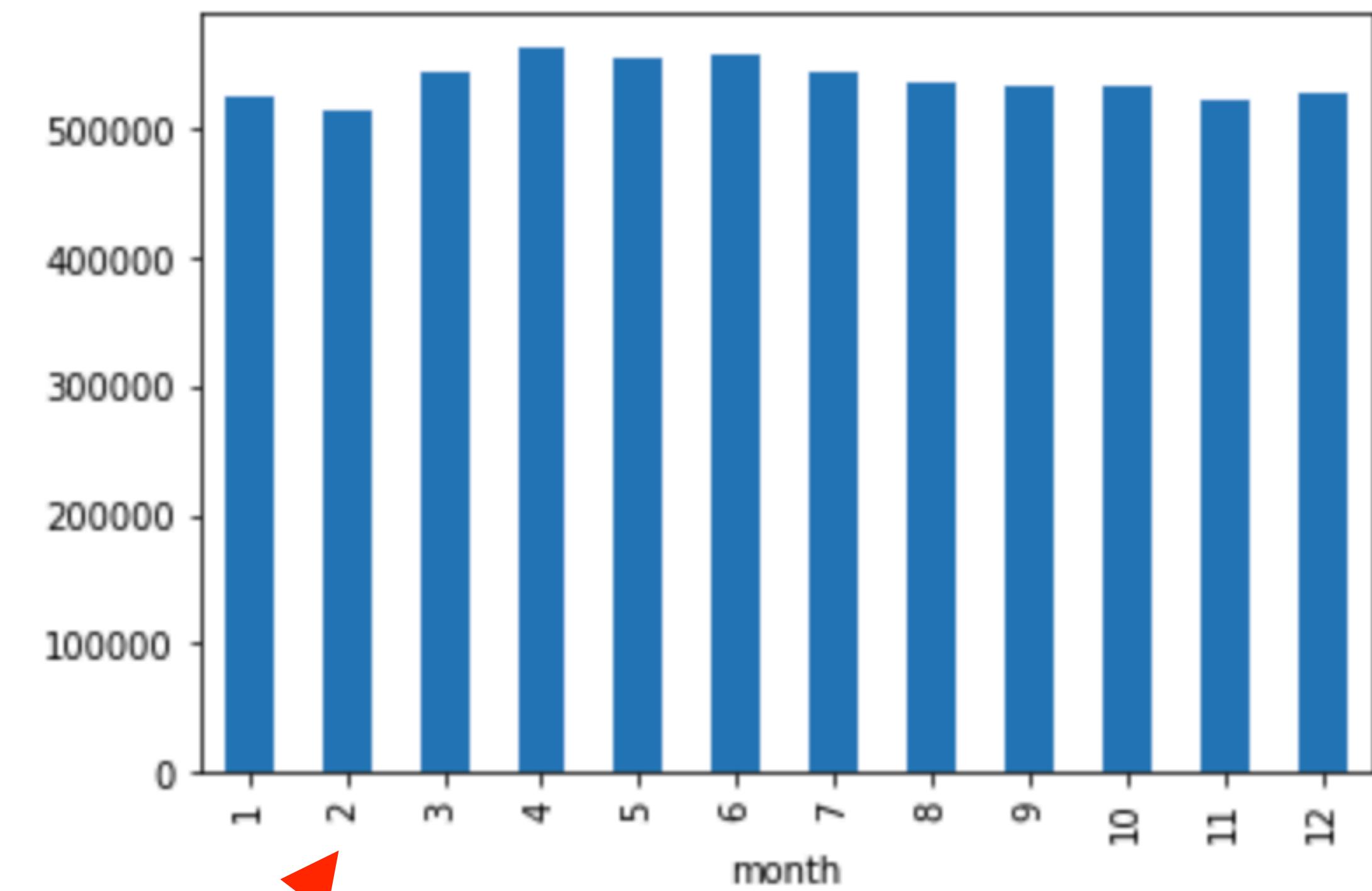
Issaquah
(Cheap and good !)

(v) Small budget or don't like crowds? Buy in winter!

Buying/Selling activity



Mean price



Buy in February!

4. A prediction model

Predict **log_price** from:

(works better than price)

grade, sqft_living, sqft_living15, sqft_above, bathrooms,
delta_lat, view, bedrooms, floors waterfront, yr_renovated,
sqft_lot, sqft_lot15, **water_distance**, condition, yr_built

Latitude difference
w.r.t. Bellevue

(works better than lat)



Minimal distance to a place
with waterfront = 1

(works better than waterfront)

$$R^2 = 0.81$$

What remains to be done:

- More careful statistical analysis of the model, take out some less significant variables
- Apply to test_set
- Make better use of some of the variables not studied here in detail

Thank you

The statistically most expensive spot in Bellevue according to this analysis:



