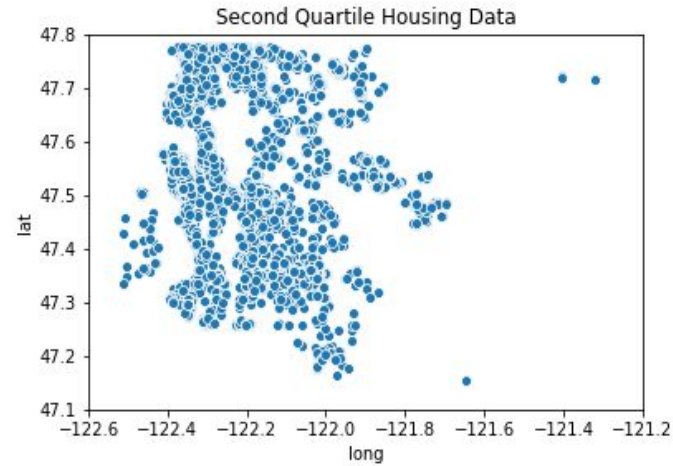
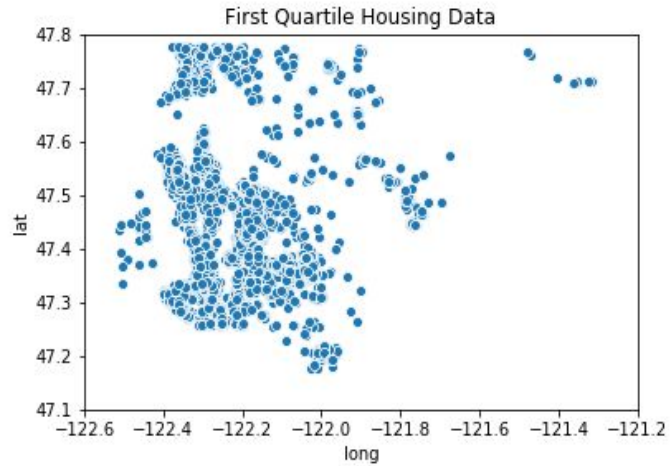


# Mod 1 Project

---

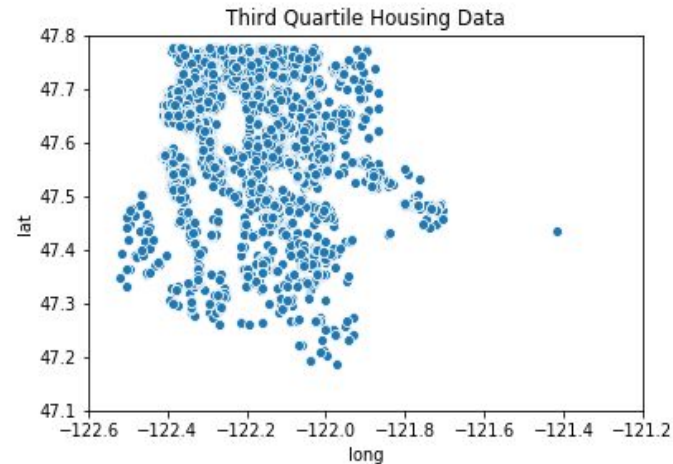
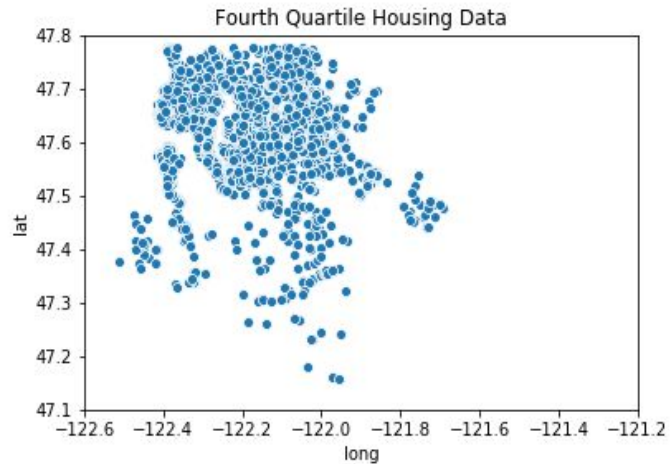
By Joe and Greg

# Does location affect the house price?



$Q1 < \$322,000$

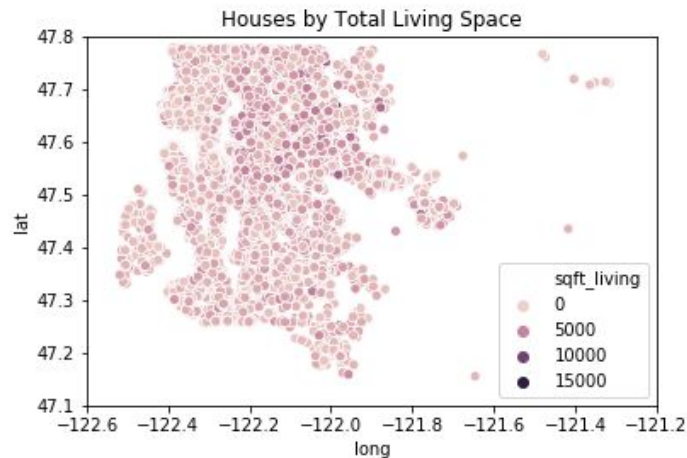
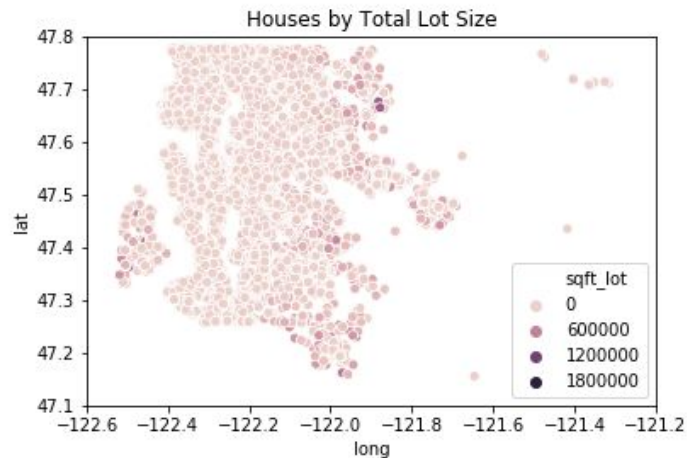
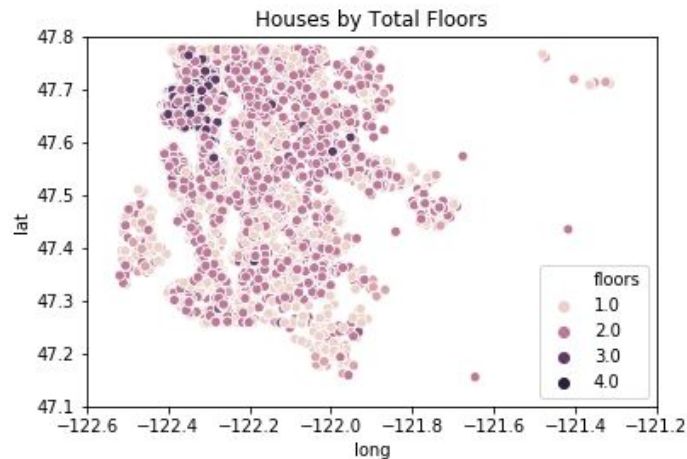
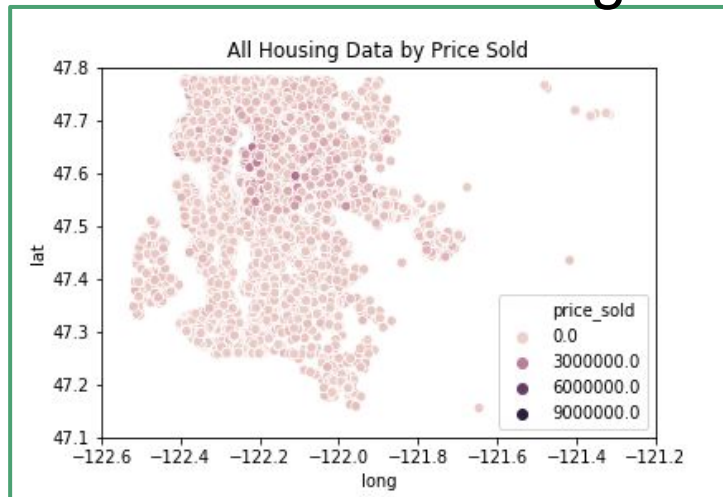
$Q1 < Q2 < \$450,000$



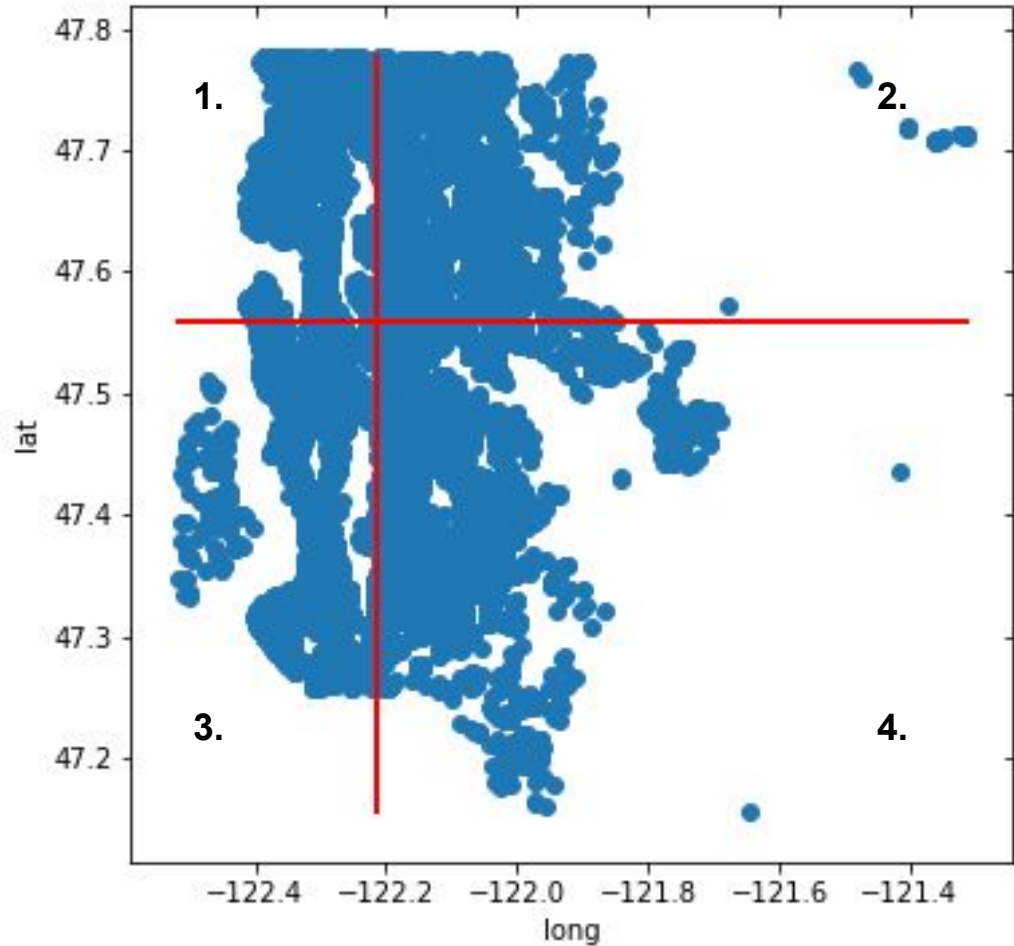
$Q2 < Q3 < \$645,000$

$Q3 < Q4$

# How else can we categorise the neighbourhoods?



Roughly split into  
regions with same  
number of properties.

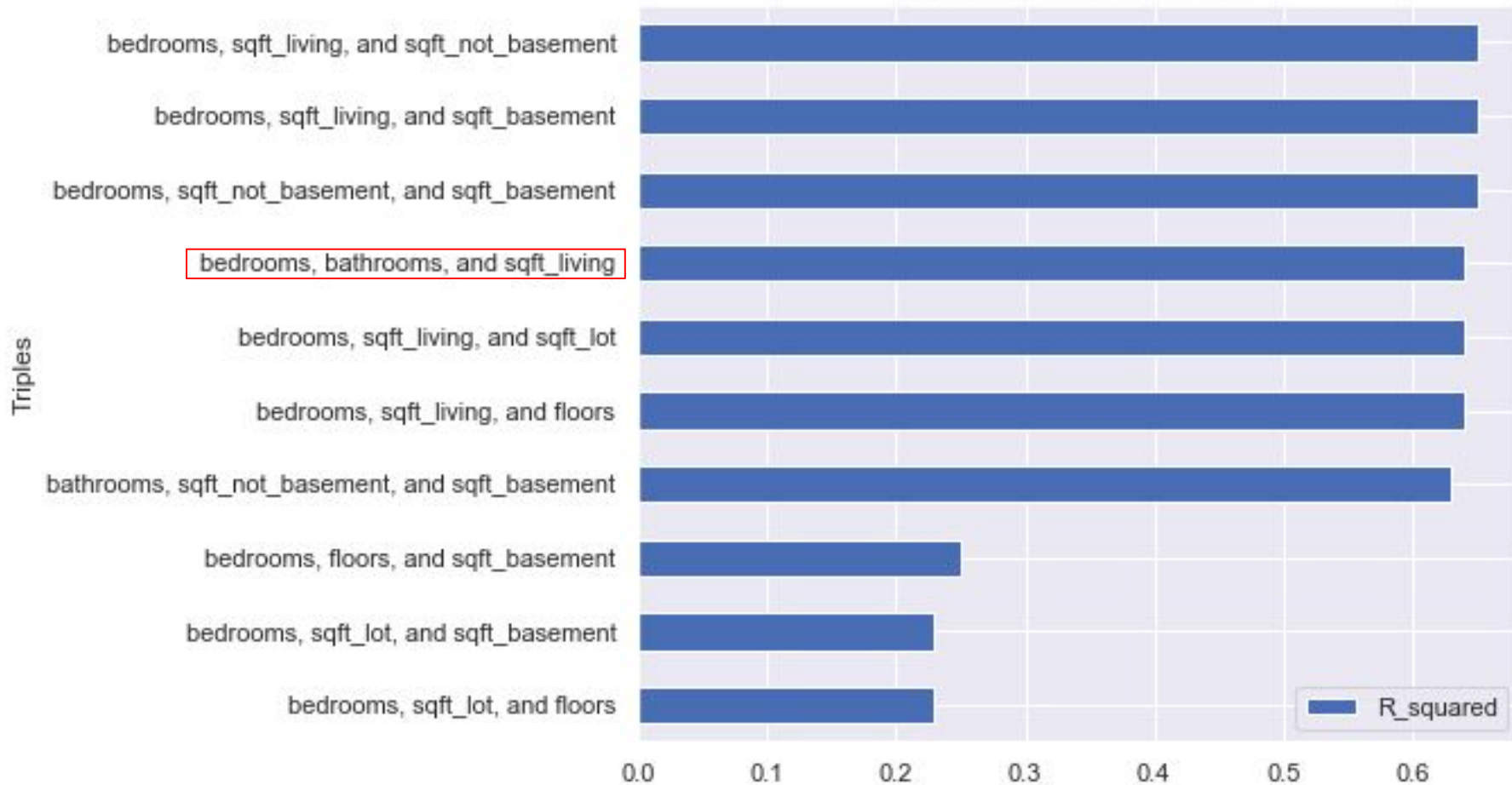


# What are the best three variables to choose?

We use the R squared value to determine which model is 'best'

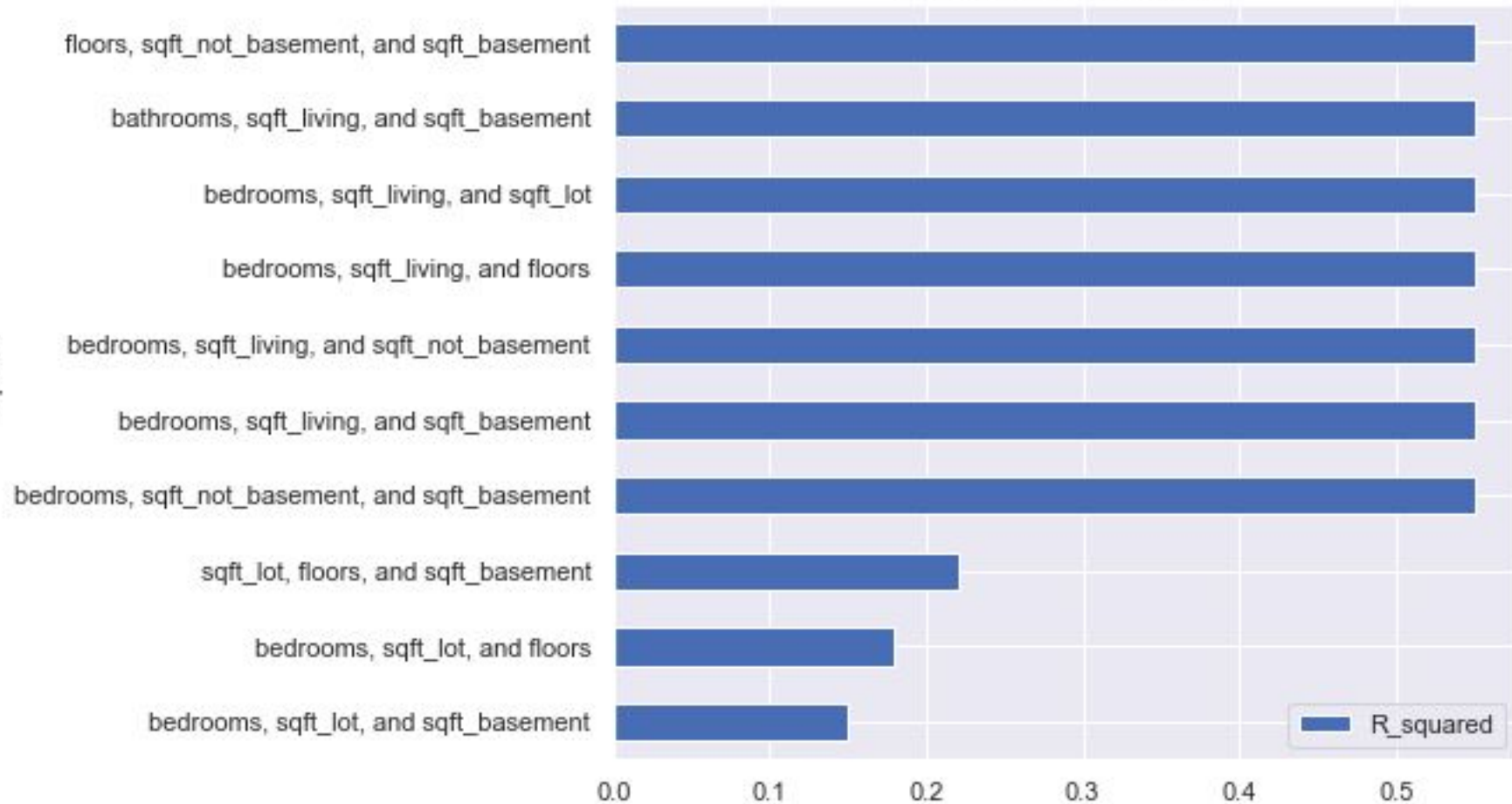
R squared is a measure of how good a models predictions are

Location 1 R\_squared trios

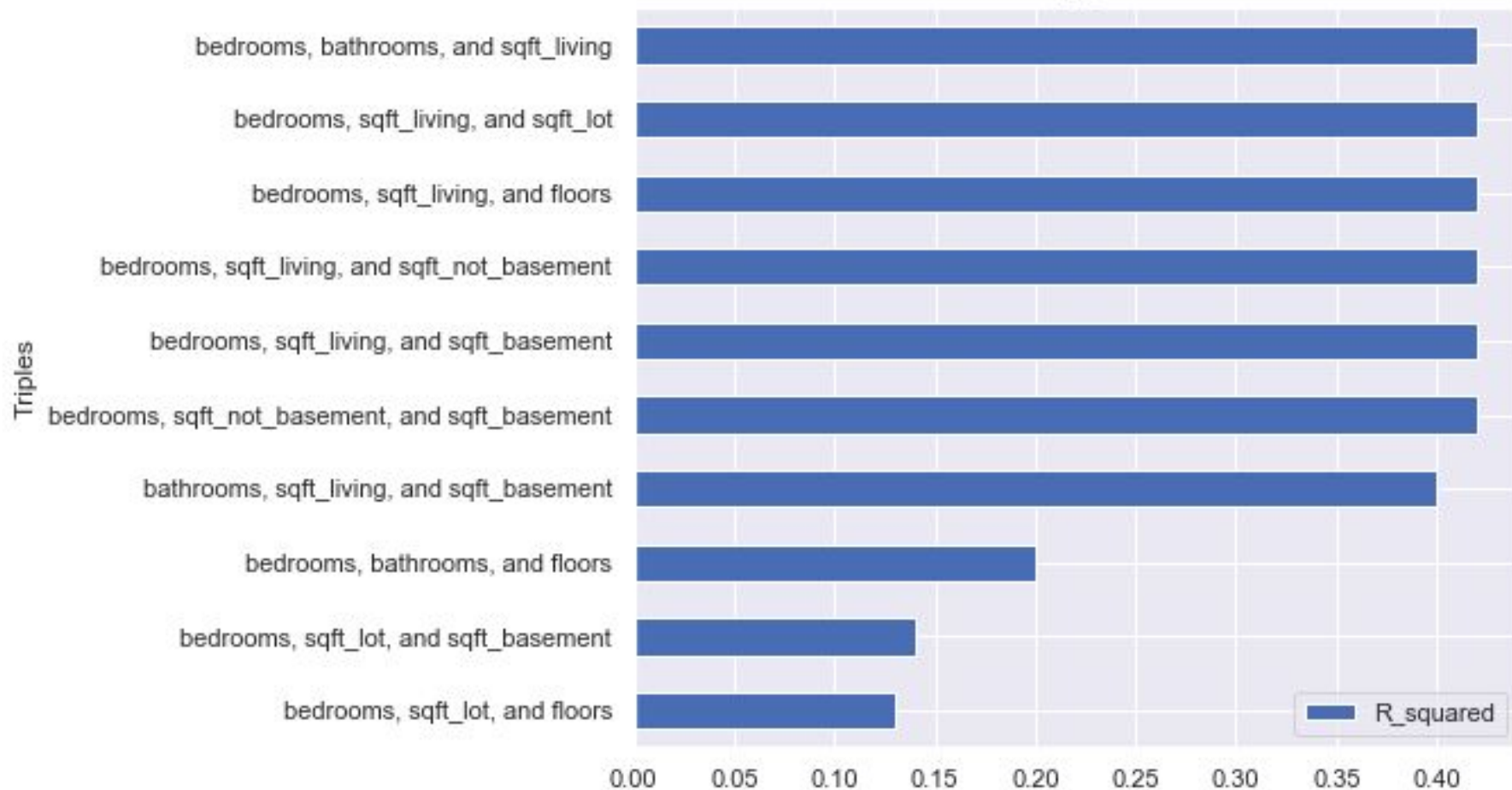


Location 2 R\_squared trios

Triples

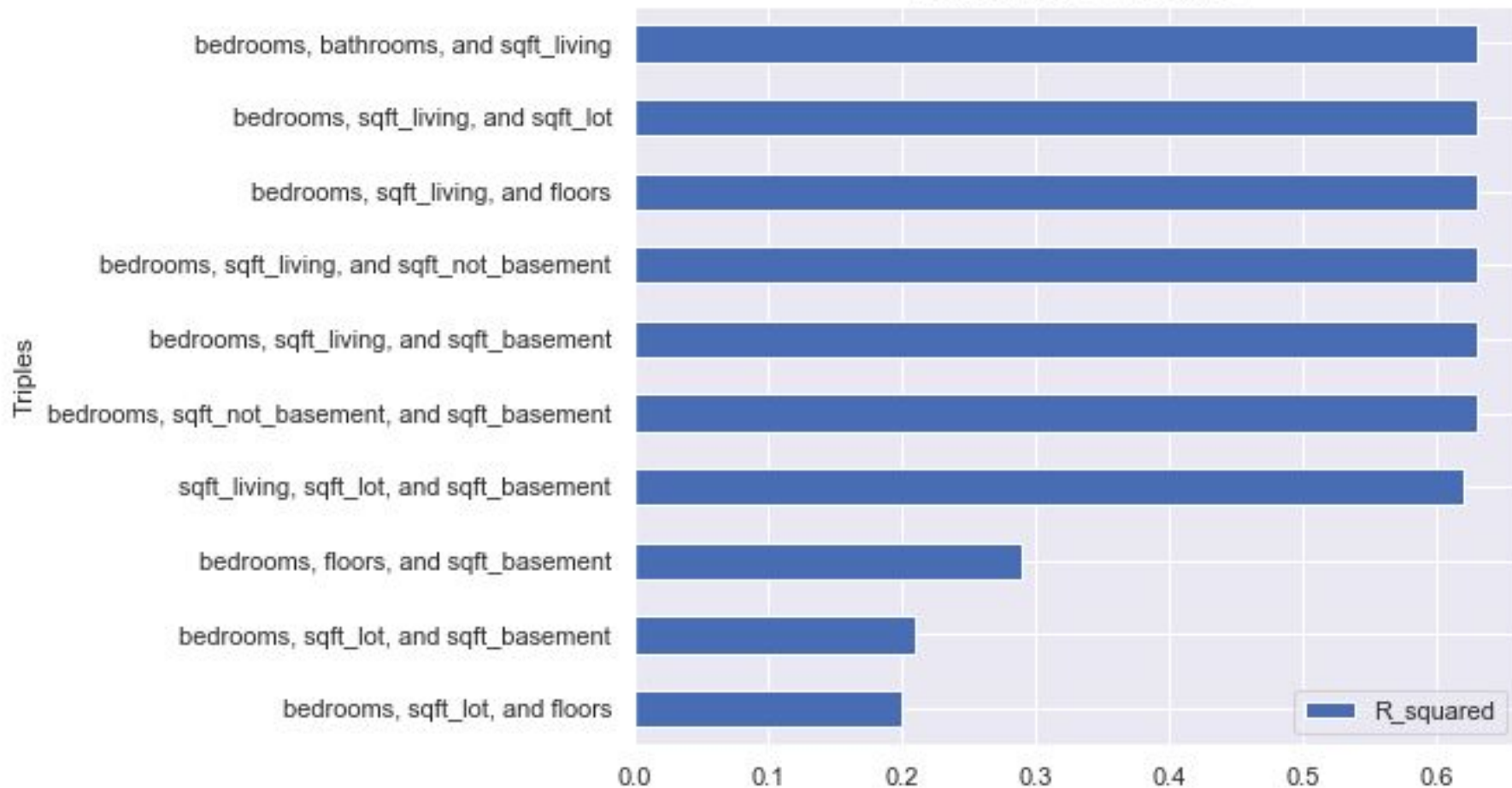


Location 3 R\_squared trios

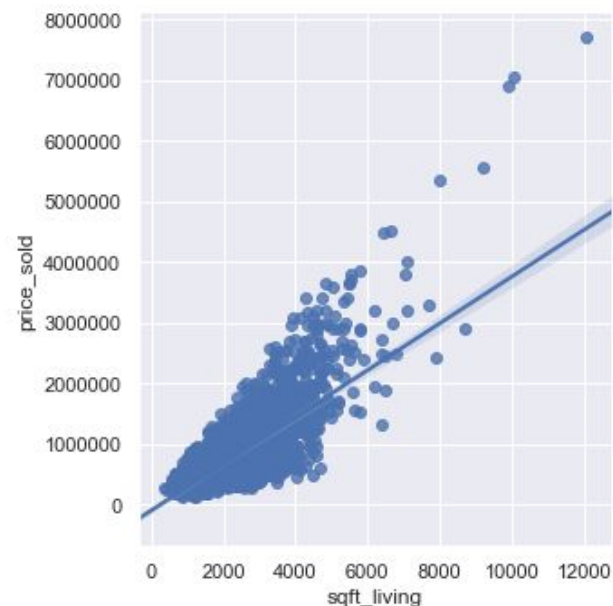
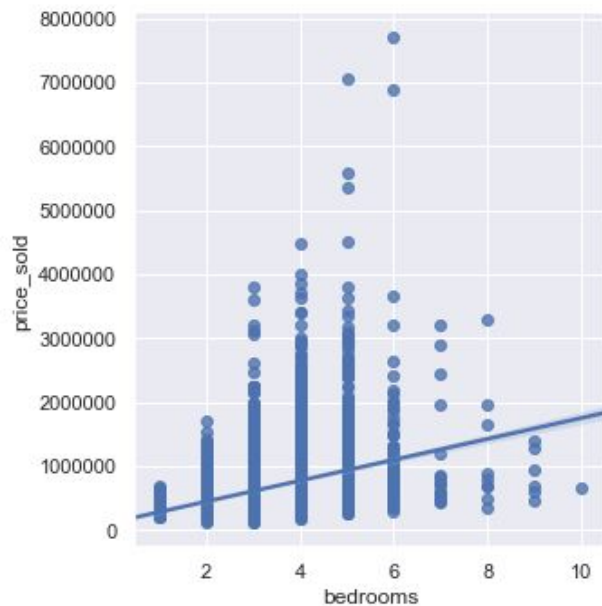
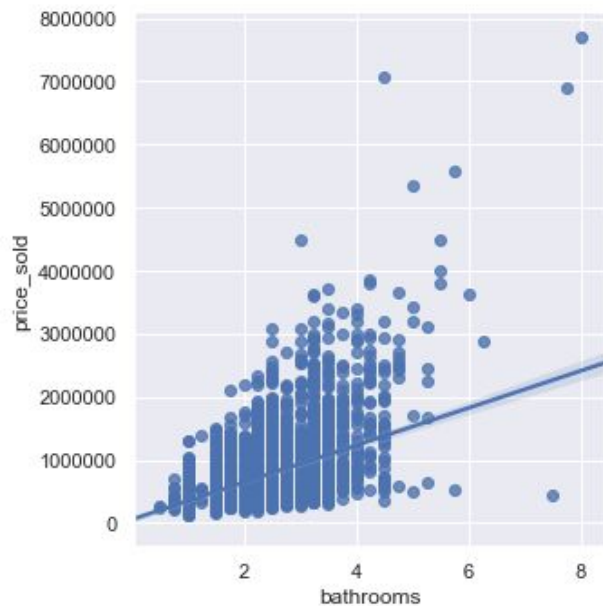




Location 4 R\_squared trios



# Seaborn linear regression for individual variables



## Linear regression for our 'best' variable trio

$R^2 = 0.64$

const            70400

bedrooms        -86400

bathrooms       -1370

sqft\_living      445

Not good!

We forgot to test for  
correlation between the  
variables

# What can we conclude?

- Clear split in property prices with location.
- Bigger lot size or more floors does NOT result in greater property value.
- Key factor is living space => Focus on this predictor.
- The accuracy of each predictor varies with location.
- Strongly correlated variables ruin our model meaning it is not an accurate predictor of price.

Thanks for listening!

Any  
questions?