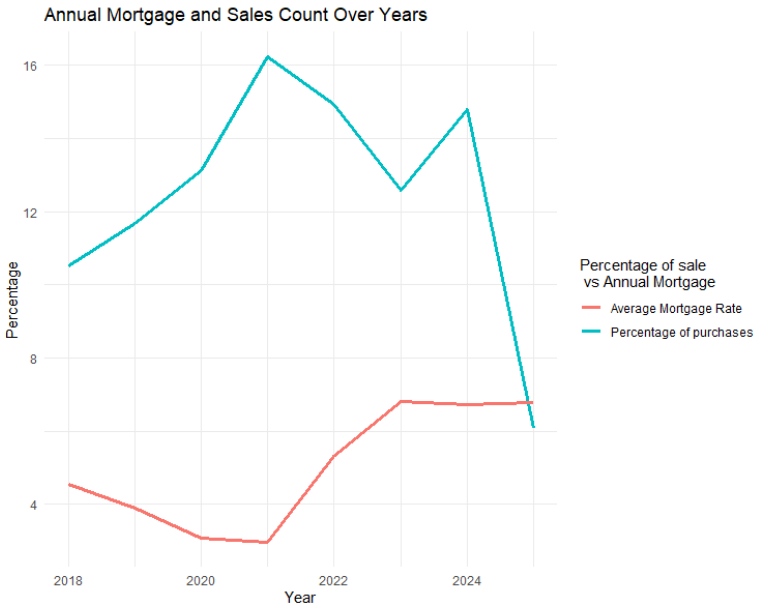
1 Introduction of the data set

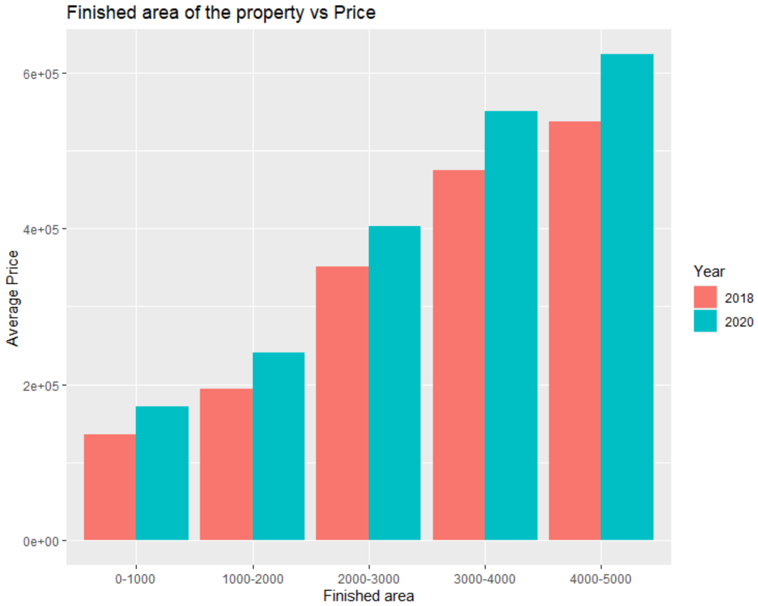
The original data set includes variables such as date of sale, sale price, finished square feet of the properties.
I added some additional variables such as annual mortgage rate, unemployment rates, total listing records according to the date of sale of the properties.

2 Visualizations

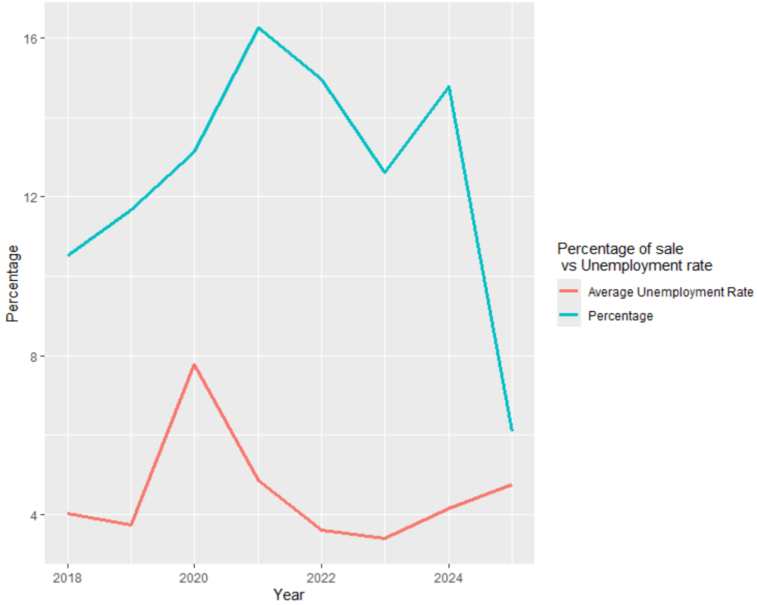
- For this visualization, I take the sum of sale each year, divide by the sum of all the purchases from 2018 to 2025 (total of 80297 purchase) and time 100 to have the percentage of the counts of sales compares to the whole. The blue line is the percentage of the sales of each year and the red line is the average annual rate of each year.



- The side-by-side bar chart illustrates the relationship between finished area and average sales price of 2018 and 2025. The data set was limited for the properties that have finished area less than 5000 squared feet, and the purchases are divided into groups based on the area of the property. Overall, the data reveals a clear positive correlation: larger finished areas generally correspond to higher sale prices.



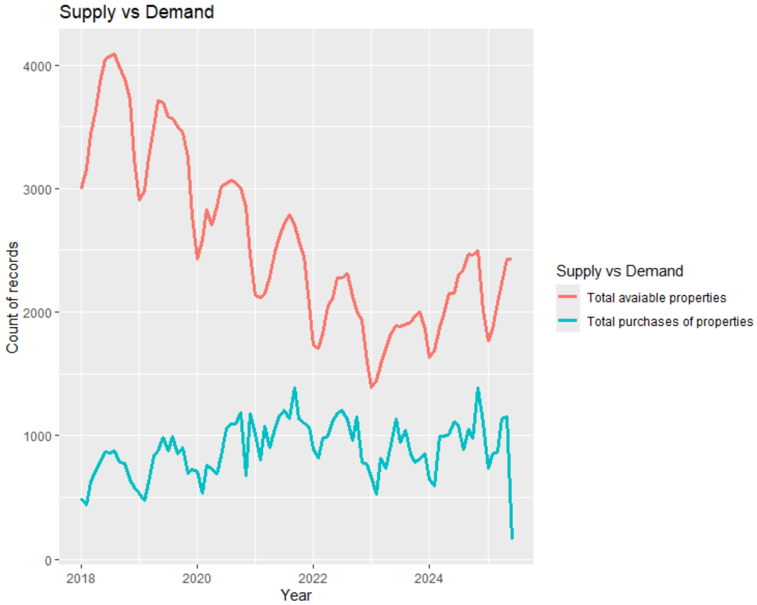
- The graph depicting percentage of sale counts in relation to unemployment rates does not reveal a clear inverse relationship between the two variables. Although in the period of time from 2020 to 2021, as the unemployment rate decreases, the rate of sales increases to the peak in 2021. However, that period of time is the only part that the two variables show a clear relationship.

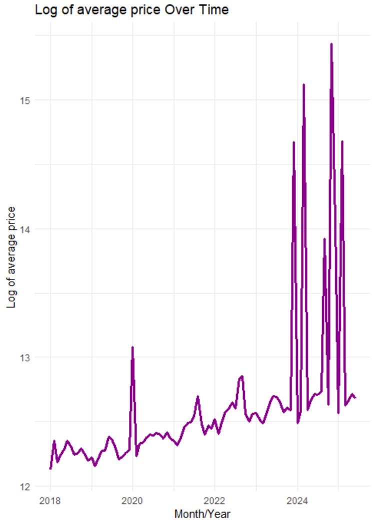


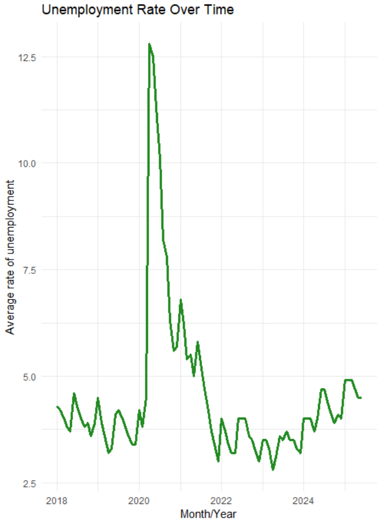
- The graph illustrates the relationship between records of total listing (available properties - supply) and the counts of sales(demand) over every month from 2018 to 2025. It appears that supply and demand have moved in parallel over time, both exhibiting a consistent upward (or downward) trend. — Issue — Explanation — — ——————————— — ———

——— — — **Doesn't capture unmet demand** — If prices are too high, many buyers may want to buy but **can't afford to** — that demand **won't show up** in transfer records. — — **Supply constraints** — In years with low inventory, fewer transactions may happen **not because demand is low**, but because **there's nothing to buy**. — — **Speculation investment** — Some transactions may be by **investors**, not end-users — so the data may mix **real housing demand** and **investment activity**. — — **Policy effects** — Tax incentives, interest rates, or regulations can **distort the timing** of transfers. —

- For thsi visualization, I want to depict the relationship between the unemployment rate and the average price of properties. However, the scale of two variables are different. Therefore, I made 2 separate visualizations and put them together. One is the unemployment rate,a dn one is the log of average price of properties over months from 2018-2025

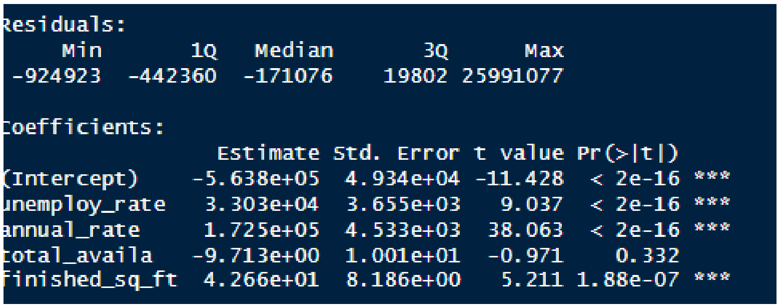

Supply vs Demand
Count of records
4000
3000
2000
1000
0
2018
2020
2022
2024
Year
Supply vs Demand
Total avaiable properties
Total purchases of properties


Log of average price Over Time
Log of average price
15
14
13
12
2018
2020
2022
2024
Month/Year


Unemployment Rate Over Time
Average rate of unemployment
12.5
10.0
7.5
5.0
2.5
2018
2020
2022
2024
Month/Year

3 Relationship between predictors vs response variable

To see if each predictors has statistically significant relationship with the response variable (sale price),I fit all the variables in linear regression. The multiple linear regression model reveals that annual mortgage rate, unemployment rate, and finished square footage are statistically significant predictors of sale price ($p < 0.001$) while total housing availability does not significantly affect sale price ($p = 0.332 > 0.05$).



```
Residuals:
     Min       1Q   Median       3Q      Max 
 -924923  -442360  -171076    19802 25991077 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)    -5.638e+05  4.934e+04 -11.428  < 2e-16 ***
unemploy_rate   3.303e+04  3.655e+03   9.037  < 2e-16 ***
annual_rate     1.725e+05  4.533e+03  38.063  < 2e-16 ***
total_availa   -9.713e+00  1.001e+01  -0.971    0.332    
finished_sq_ft  4.266e+01  8.186e+00   5.211 1.88e-07 ***
```