Final Project | STAT 419

Group 4

Abby Drongpa, Alex Sullivan, Chris Liu, Jett Palmer, Sumanth Thokala

2025-06-08

Introduction

Our data set is a subset of home in Cincinnati, Ohio, from the year 2002. It includes 552 observations and seven variables, each representing an individual house. This data set is meant to identify factors that influence and may predict the selling of homes.

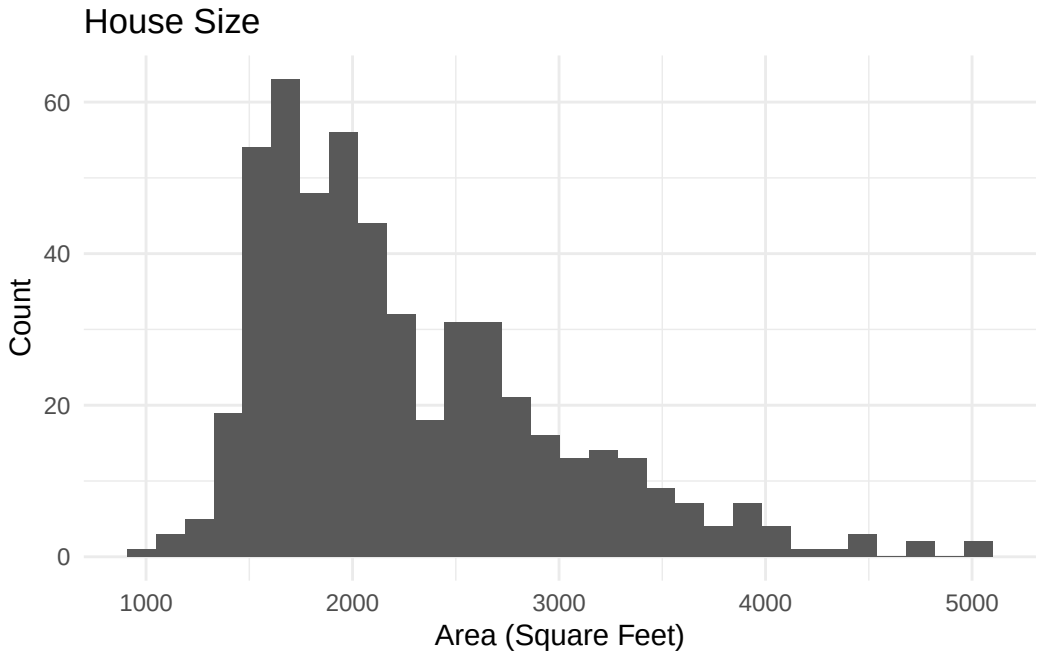
Our ordinal grouping variable, "Pricerank", categorizes home based on their selling prices. Houses sold for less than $190,000 are coded as 1, those sold between $190,000 and $285,000 are coded as 2, and houses sold for more than $285,000 are coded as 3.

The remaining six variables are explanatory:

- Area: The size of the house in square feet.
- BR: The number of bedrooms.
- BA: The number of bathrooms.
- Garage: The number of cars the garage(s) can accommodate.
- Quality: An index of construction quality, where 1 indicates high quality, 2 indicates medium quality, and 3 indicates low quality.
- Age: The age of the house as of 2002.

Graphs and Summary Statistics

House Area Histogram

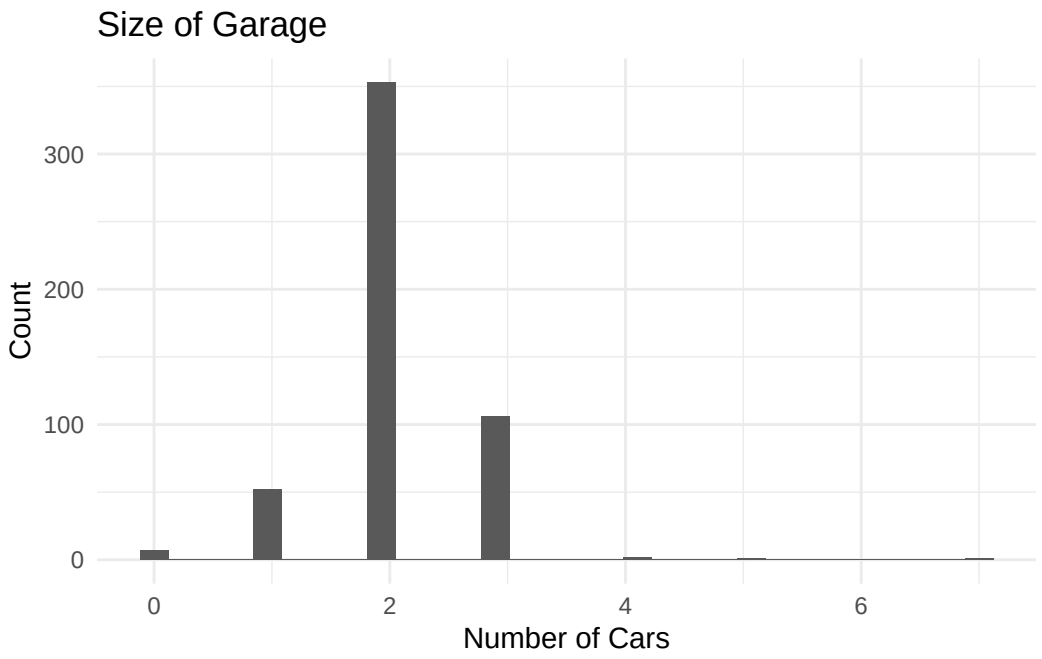



```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 980    1701    2061    2261    2636    5032
```

St Dev
711.066

The distribution is right skewed with a mean area of 2261 ft^2 and a median of 2061 ft^2. The standard deviation is 711.06 ft^2.

Car Storage Histogram

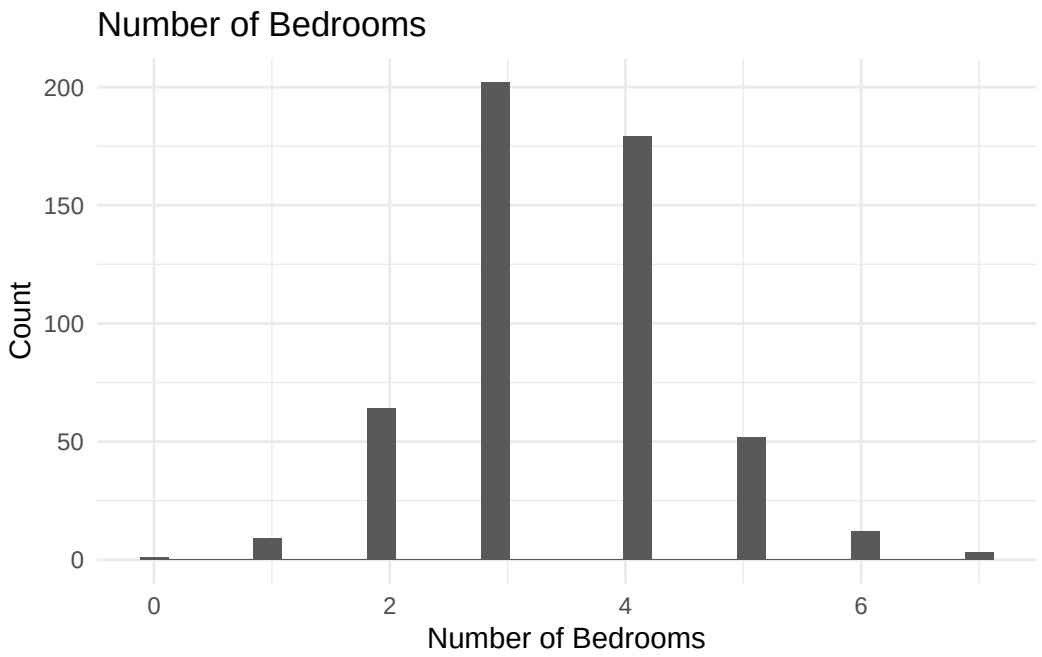


The mean number of cars able to be stored in the garage(s) is 2.1 with a median of 2 cars and a standard deviation of 0.65. The distribution reveals several high and low outliers at 0, 4 , 5, 6, and 7 cars. In other words, the distribution is highly concentrated at 2 car garages.

```
Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.0     2.0     2.0     2.1     2.0     7.0
```

St Dev
0.654

Bedroom Histogram

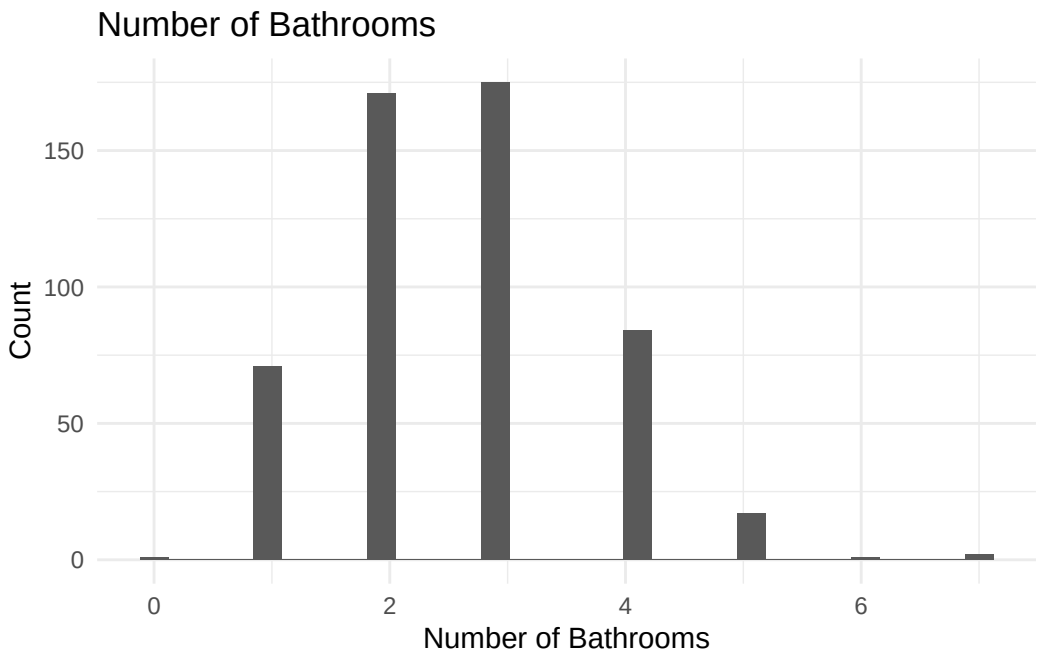


The distribution of number of bedrooms in each house is roughly normal, with a mean of 3.47 and median of 3 bedrooms. The standard deviation is 1.01 bedrooms.

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   3.000   3.000   3.471   4.000   7.000
```

St Dev
1.014

Bathrooms Histogram

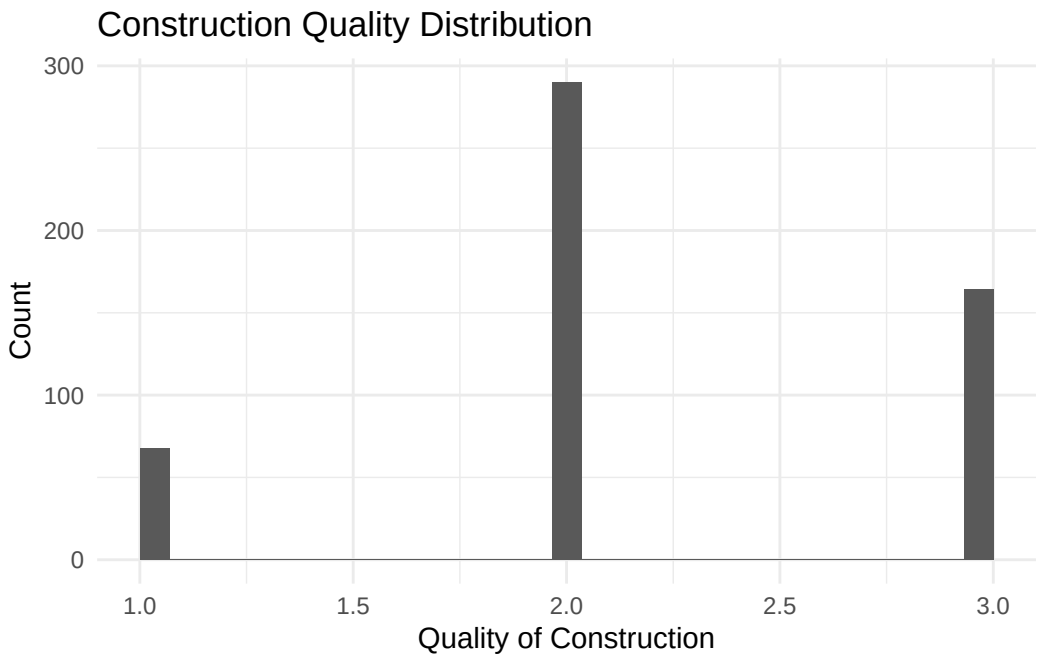


The distribution of number of bathrooms is skewed the right with a mean of 2.64 and median of 3 bedrooms. The standard deviation is 1.06.

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   2.000   3.000   2.642   3.000   7.000
```

St Dev
1.064

Construction Quality Bar Chart




```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.000   2.000   2.000   2.184   3.000   3.000
```

St Dev
0.641

Most houses were given a quality rating of 2 (medium), followed by 3 (low quality), and 1 (high quality).

Discriminant and Classification Analysis

Correlated Quantitative Variables

Discriminant Analysis

Classification Analysis

Summary

R Code