# Final Project

Shalom Mhanda, Sophia Perez

Introduction

Exploratory Data Analysis

The Saratoga, New York Housing Prices dataset is a random sample of 1057 houses taken from full Saratoga Housing Data (De Veaux). The data contains variables such as price (US Dollars), bedrooms (count), fireplaces (count),living area (square feet), age (years), bathrooms (count), lot size (acres), and fireplace (presence of fireplace, either true or false). The sample represents the houses in Saratoga, New York and their different characteristics. The data was collected as a random sample of the Saratoga houses. We are mainly focusing on checking if the number of bedrooms and number of fireplaces influences the house price. We first take a look at the variables bedrooms and fireplaces individually and how they relate to housing price and then look at both variables together.
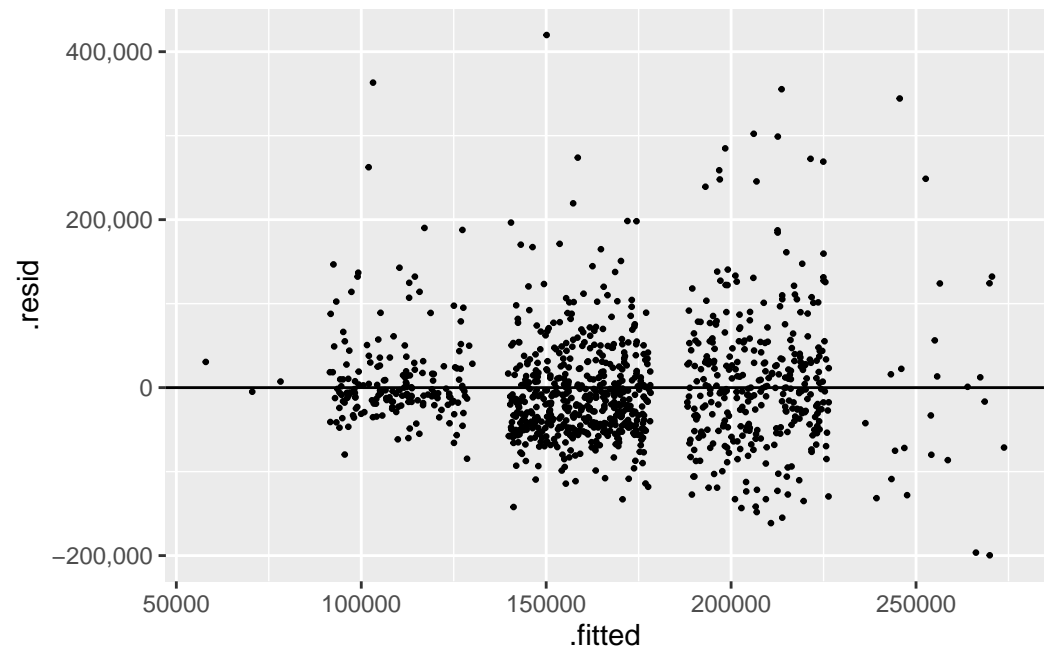
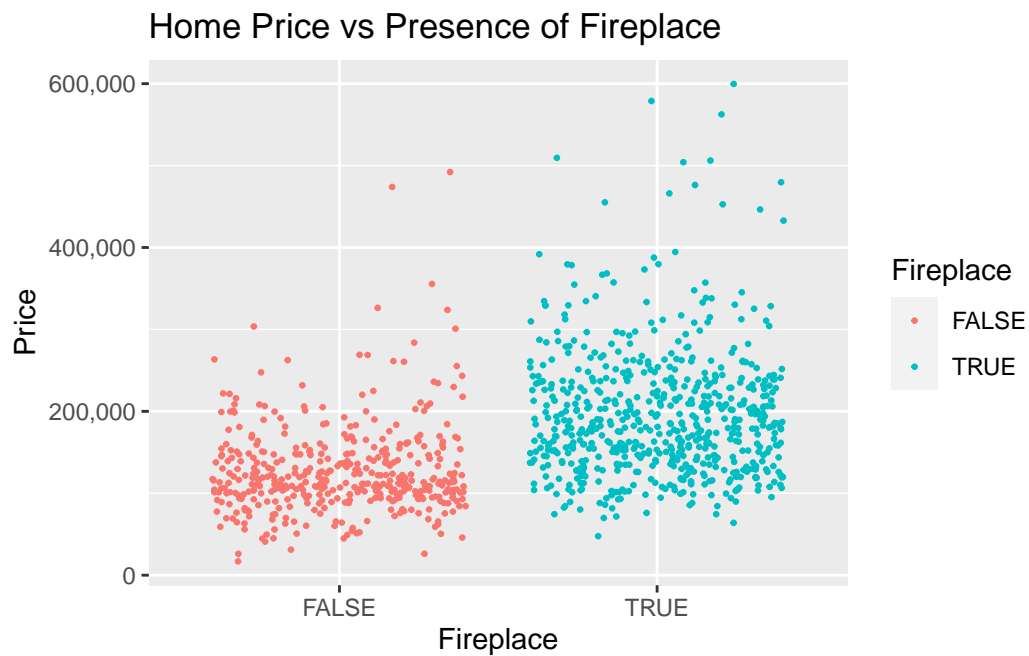**Is number of bedrooms a good predictor of house price?**
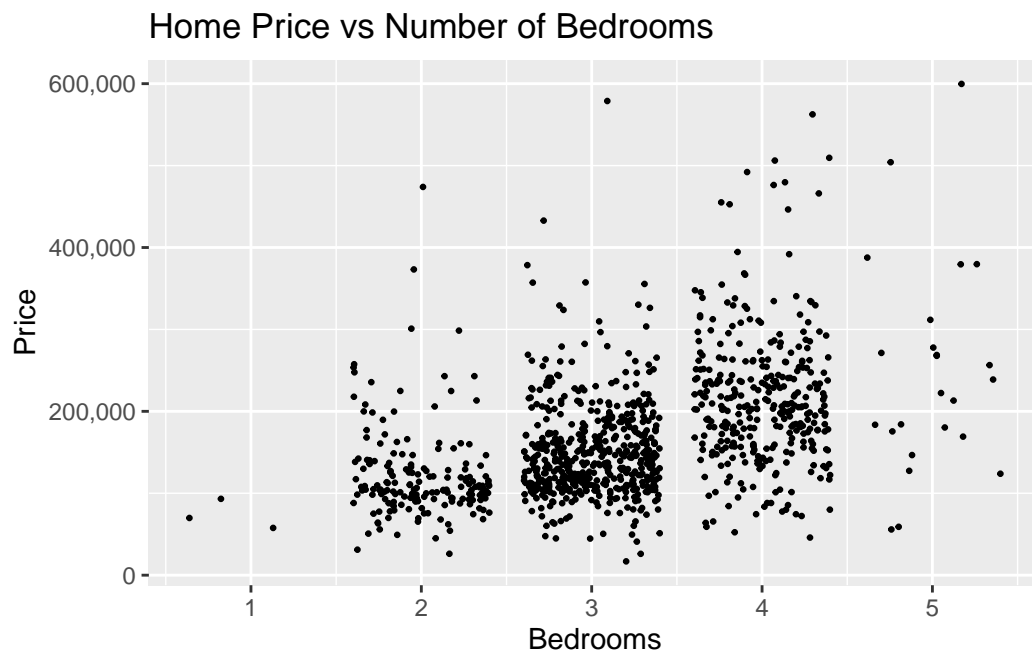
Home Price vs Number of Bedrooms

We draw a linear regression line to check for linearity between the house price and number of bedrooms.
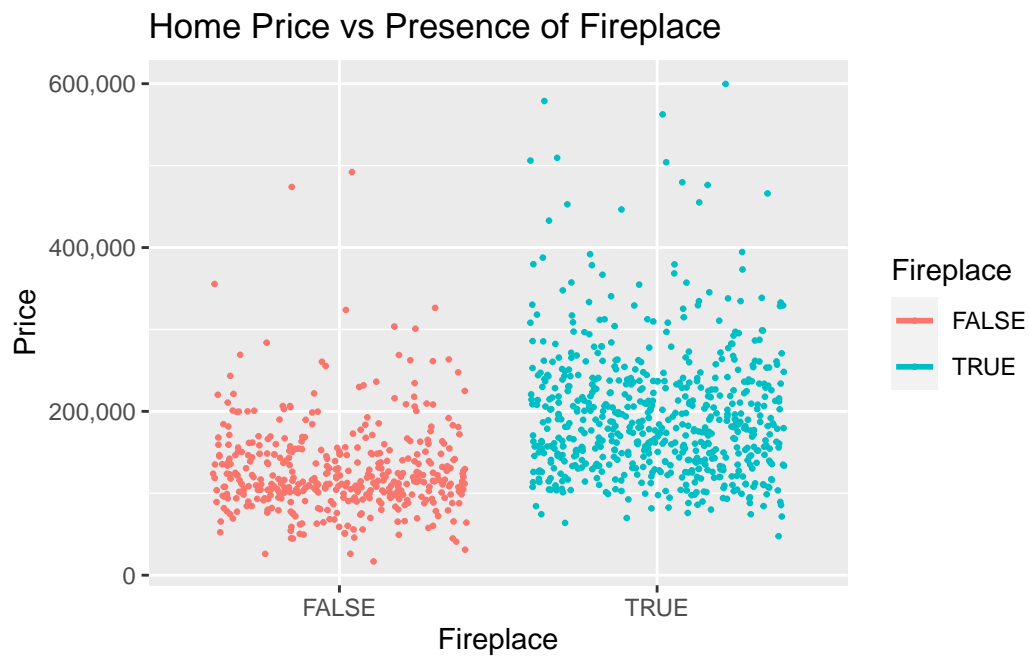
Home Price vs Number of Bedrooms
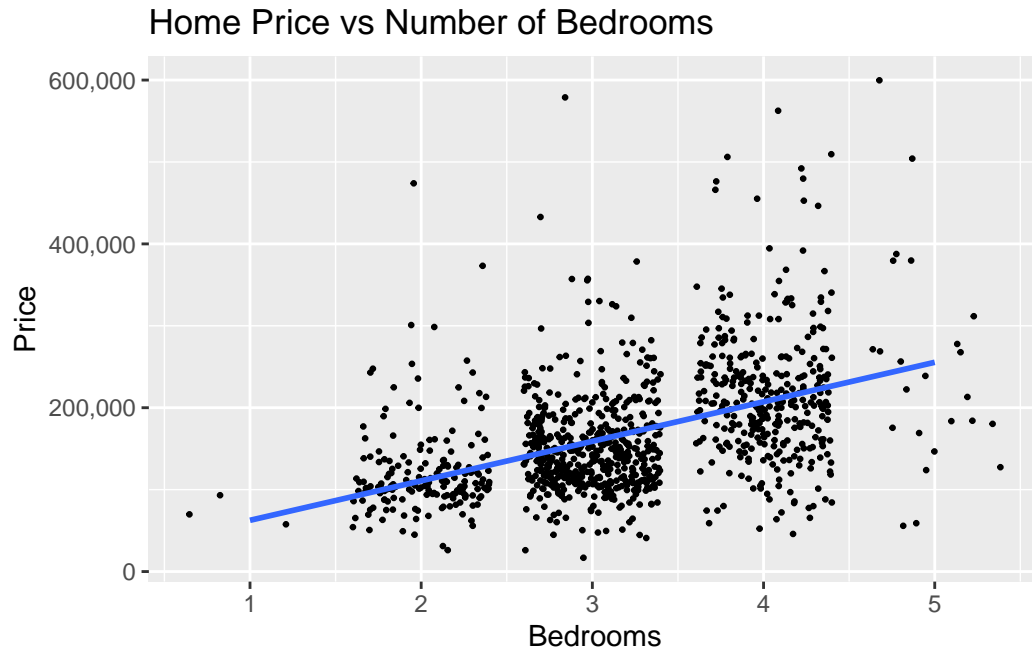
We plot a residual plot to check if the variables house price and number of bedrooms have normal residuals.

Is the presence of a fireplace a good predictor of house price?
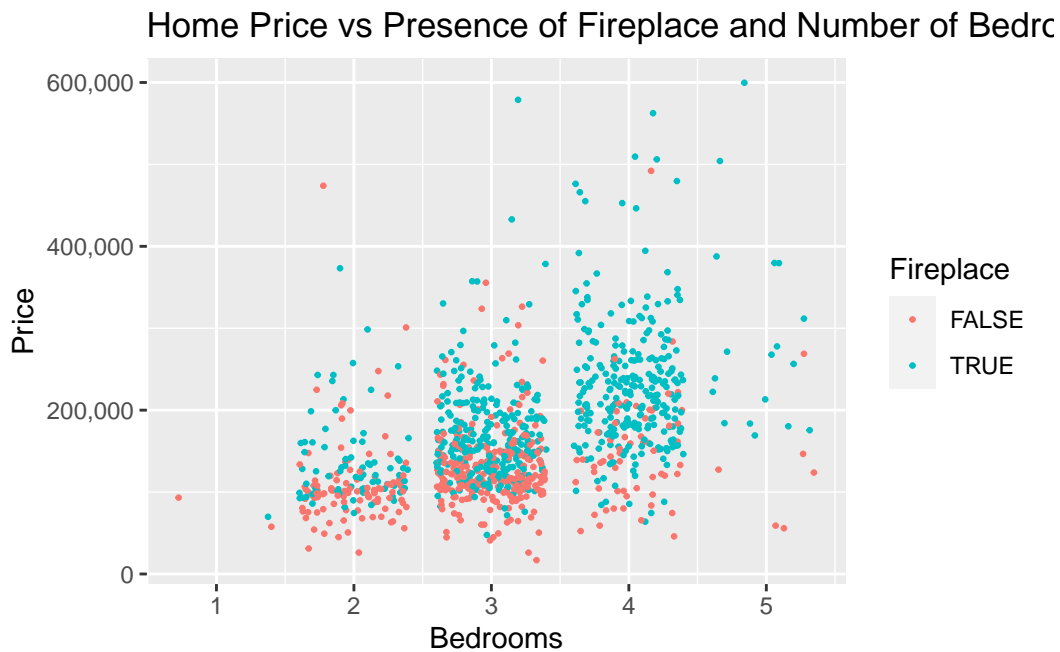
Home Price vs Presence of Fireplace

We try to draw a linear regression line to check for linearity between the house price and the presence of a fireplace. It seems there is no linearity as a linear line could not be drawn.

Home Price vs Presence of Fireplace

We plot a residual plot to check if the variables house price and presence of fireplace have normal residuals.
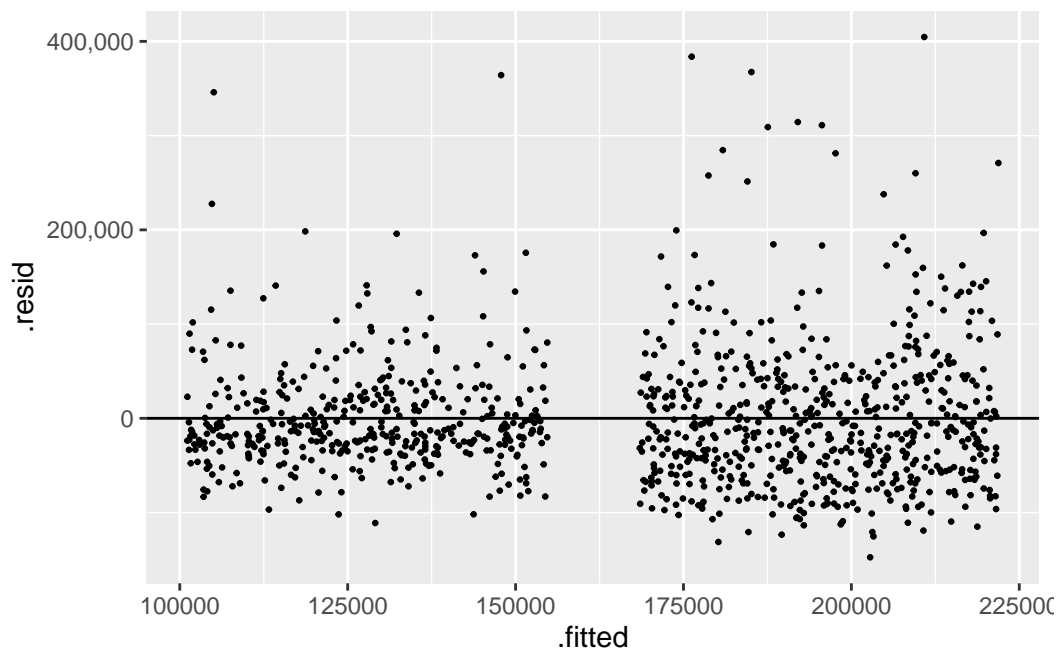
Multiple Predictors: Number of bedrooms and presence of fireplace as predictos of house price

Home Price vs Presence of Fireplace and Number of Bedrc

Results

To examine the relationship between number of bedrooms and presence of fireplace(s) on hose price, we do a hyphothesis test for the linear regression with multiple predictors. Our null hypothesis is that the slope of each predictor variable with respect to the response variable is 0. given other variables in the model. The alternative hupothesis is that the slope is not zero, given other variables. Our data comes from a random sample so it is independent.

```
# A tibble: 3 x 5
  term          estimate std.error statistic  p.value
  <chr>            <dbl>     <dbl>     <dbl>    <dbl>
1 (Intercept)     16201.     8745.      1.85 6.42e- 2
2 Bedrooms        38326.     2803.     13.7  2.65e-39
3 FireplaceTRUE   49830.     4226.     11.8  3.13e-30
```

Our p values for bedrooms and fireplace(True) are lvery small and less than $\alpha = 0.05$. Therefore, we colcude against the null hypothesis. The number of bedrooms and presence of fireplaces do not seem to be good predictors of house prices.

Conclusion