

ECON203 HW1

Charles Ancel

2024-01-27

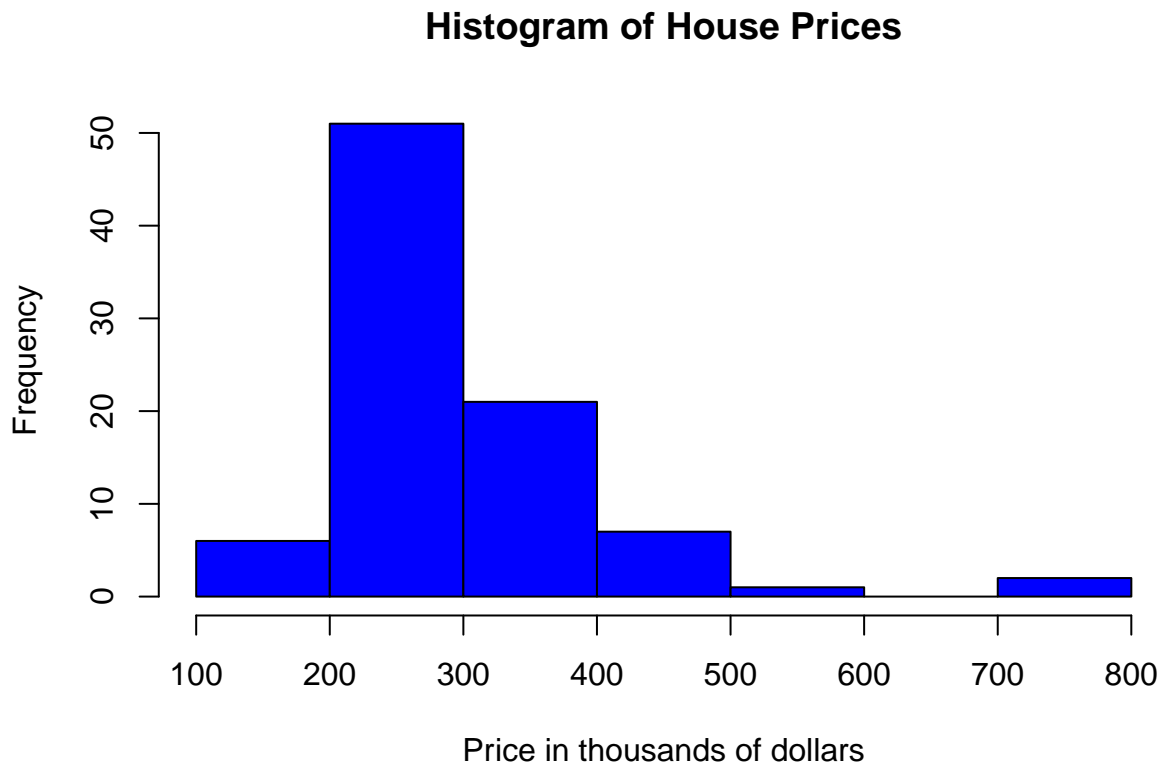
```
# Install the package (you only need to do this once)
install.packages("readxl")

##
## The downloaded binary packages are in
## /var/folders/d8/qgb_8zcs7vl6pjzppspg_0sw0000gn/T//RtmpDudf6t/downloaded_packages
# Load the package into R
library(readxl)

# Load the dataset
housing_data <- read_excel("housing.xls")
```

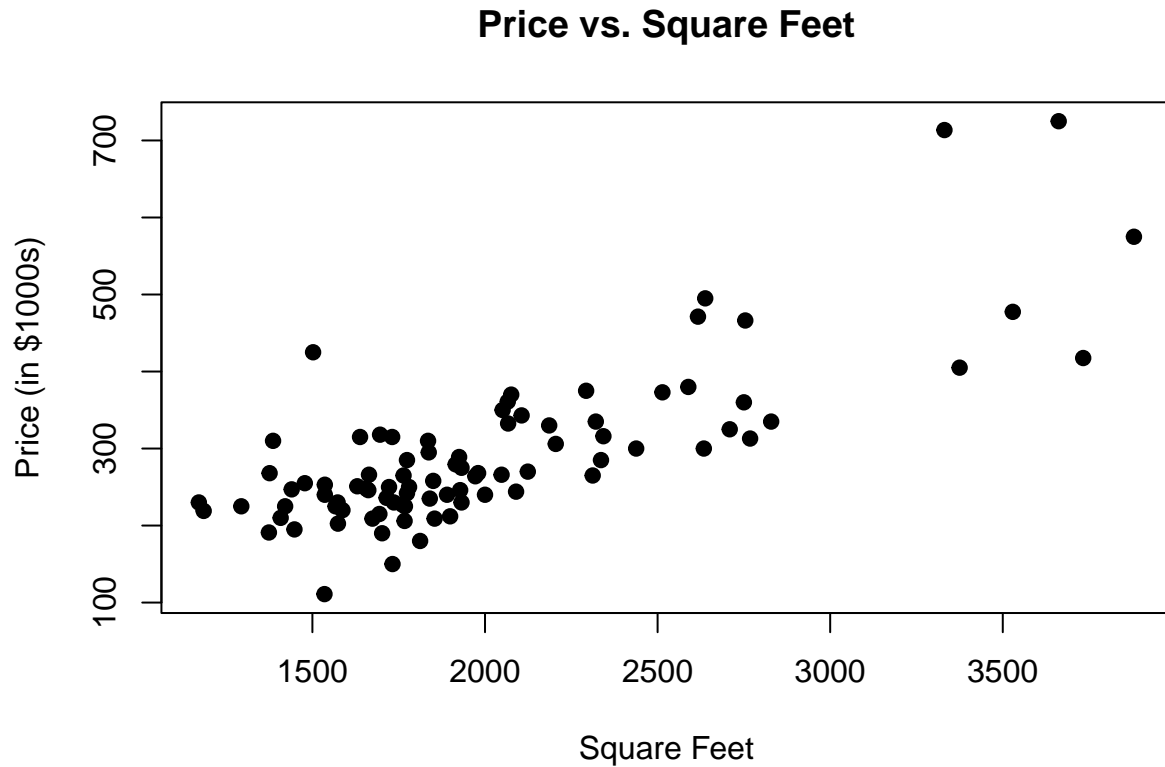
Question One:

```
# Plotting histogram of price
hist(housing_data$price, main="Histogram of House Prices", xlab="Price in thousands of dollars", col="b")
```



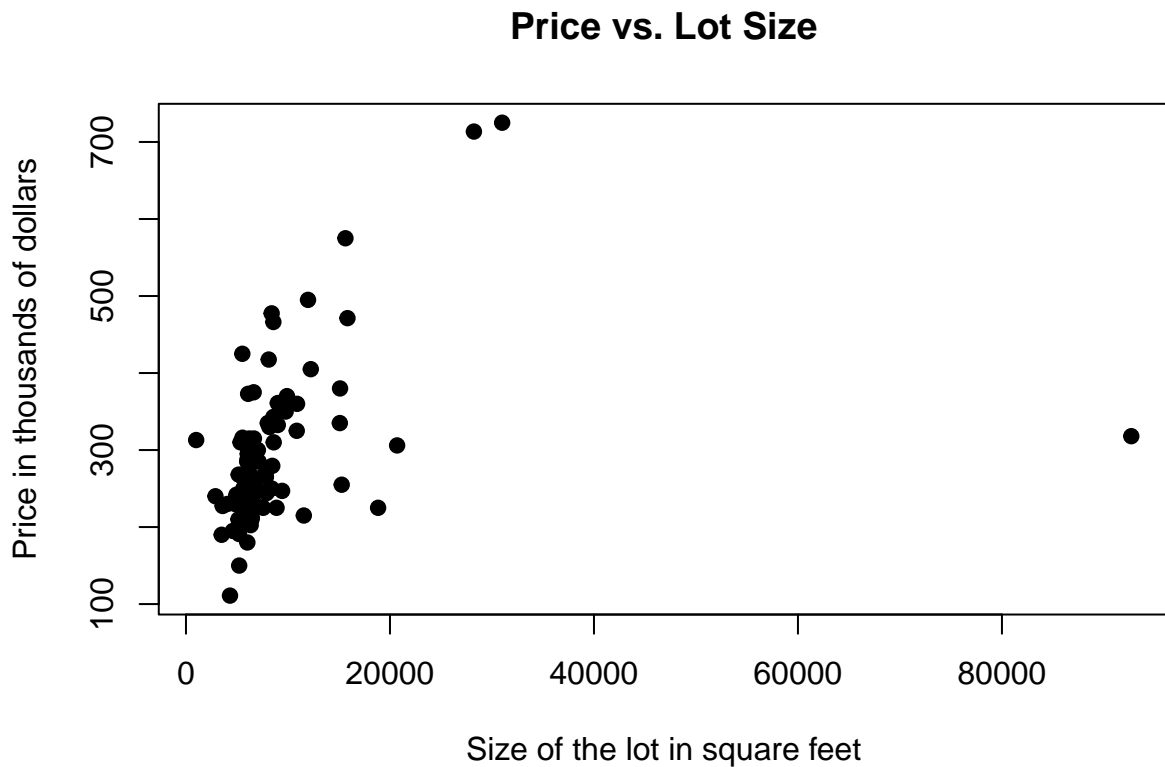
```
# Plotting price vs. square feet
```

```
plot(housing_data$sqrft, housing_data$price, main="Price vs. Square Feet", xlab="Square Feet", ylab="Price (in $1000s)", col="black", pch="n")
```



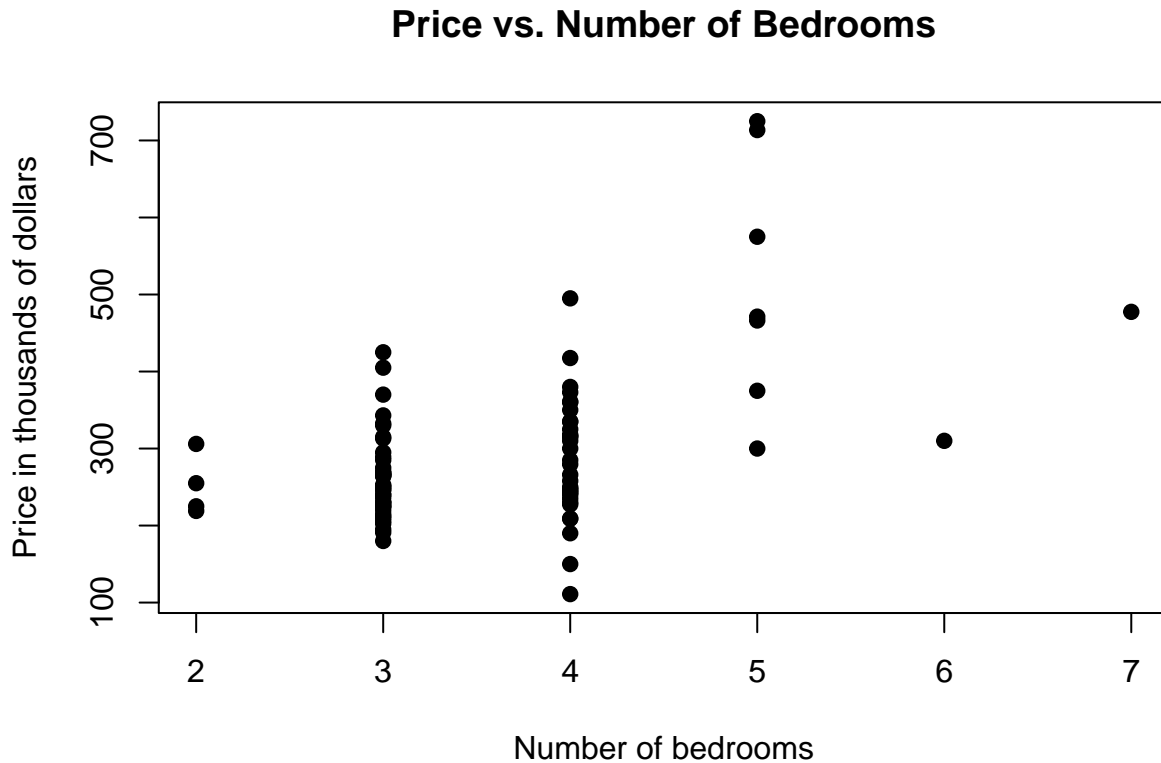
```
# Plotting price vs. lot size
```

```
plot(housing_data$lotsize, housing_data$price, main="Price vs. Lot Size", xlab="Size of the lot in square feet", ylab="Price in thousands of dollars", col="black", pch="n")
```



```
# Plotting price vs. number of bedrooms
```

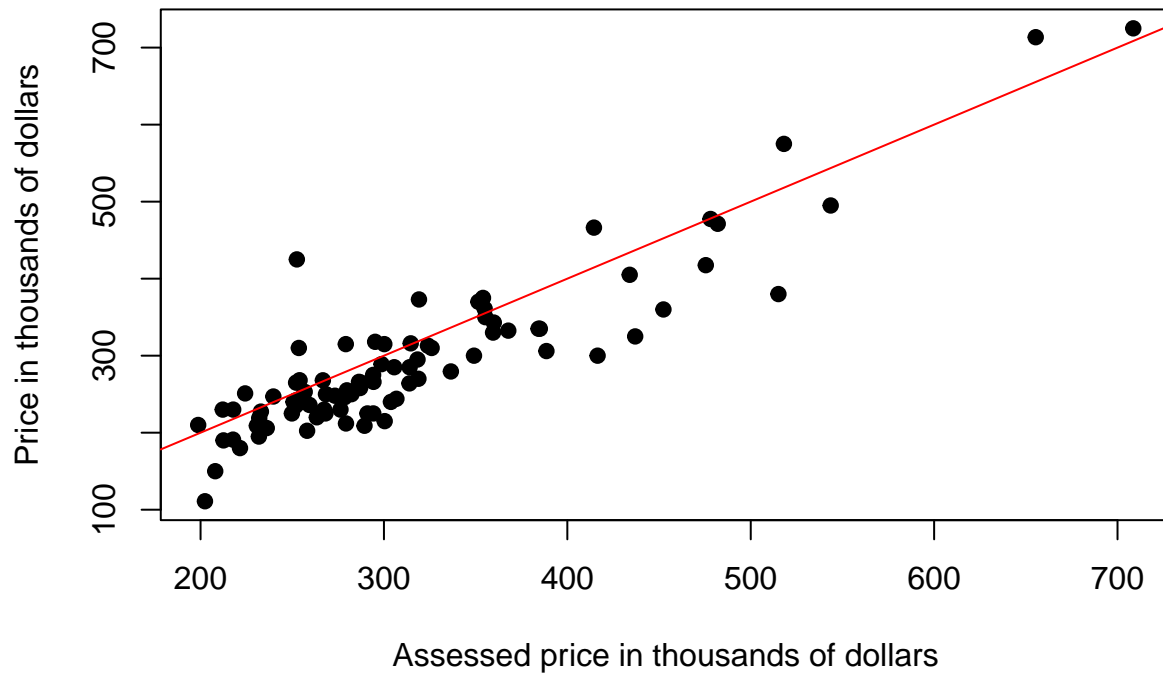
```
plot(housing_data$bdrms, housing_data$price, main="Price vs. Number of Bedrooms", xlab="Number of bedrooms", ylab="Price in thousands of dollars")
```



```
# Plotting price vs. assessed price
```

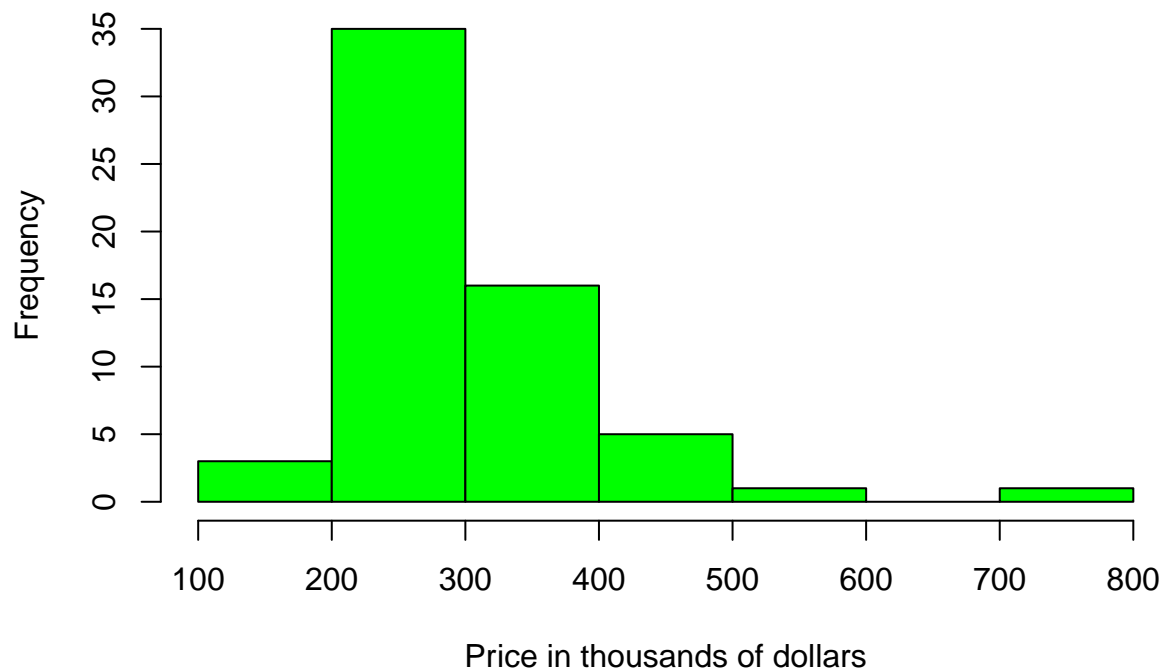
```
plot(housing_data$assess, housing_data$price, main="Price vs. Assessed Price", xlab="Assessed price in thousands of dollars", ylab="Price in thousands of dollars", col="red")  
abline(0, 1, col="red") # Adding a 45-degree line
```

Price vs. Assessed Price



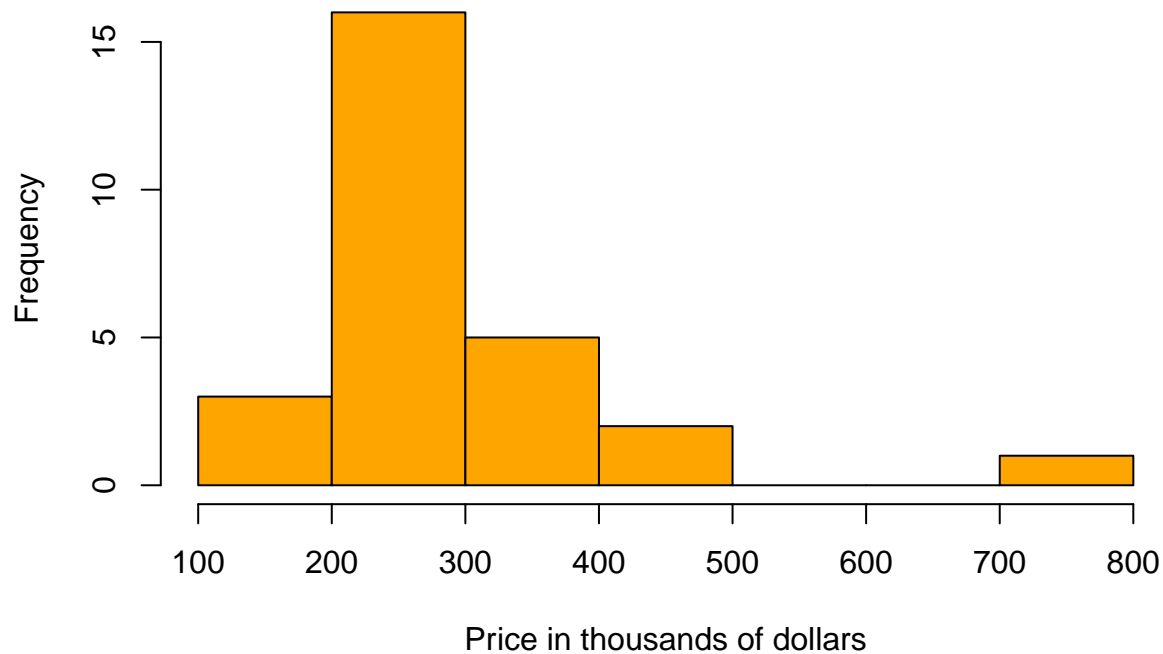
```
# Histogram for prices of colonial houses
hist(housing_data$price[housing_data$colonial == 1], main="Prices of Colonial Houses", xlab="Price in thousands of dollars", ylab="Frequency", col="red", border="black")
```

Prices of Colonial Houses



```
# Histogram for prices of non-colonial houses
hist(housing_data$price[housing_data$colonial == 0], main="Prices of Non-Colonial Houses", xlab="Price in thousands of dollars", ylab="Frequency", col="blue", border="black")
```

Prices of Non-Colonial Houses



```
head(housing_data)
```

```
## # A tibble: 6 x 6
##   price assess bdrms lotsize sqrft colonial
##   <dbl> <dbl> <dbl>   <dbl> <dbl>   <dbl>
## 1  300   349.     4   6126  2438     1
## 2  370   352.     3   9903  2076     1
## 3  191   218.     3   5200  1374     0
## 4  195   232.     3   4600  1448     1
## 5  373   319.     4   6095  2514     1
## 6  466.  414.     5   8566  2754     1
```

```
summary(housing_data)
```

```
##      price          assess          bdrms          lotsize          sqrft
##  Min.   :111.0   Min.   :198.7   Min.   :2.000   Min.   : 1000   Min.   :1171
## 1st Qu.:230.0   1st Qu.:253.9   1st Qu.:3.000   1st Qu.: 5733   1st Qu.:1660
## Median :265.5   Median :290.2   Median :3.000   Median : 6430   Median :1845
## Mean   :293.5   Mean   :315.7   Mean   :3.568   Mean   : 9020   Mean   :2014
## 3rd Qu.:326.2   3rd Qu.:352.1   3rd Qu.:4.000   3rd Qu.: 8583   3rd Qu.:2227
## Max.   :725.0   Max.   :708.6   Max.   :7.000   Max.   :92681   Max.   :3880
##      colonial
##  Min.   :0.0000
## 1st Qu.:0.0000
## Median :1.0000
## Mean   :0.6932
## 3rd Qu.:1.0000
## Max.   :1.0000
```

```
# Standard deviation for all variables
```

```
sapply(housing_data, sd, na.rm = TRUE)
```

```
##           price           assess           bdrms           lotsize           sqrft           colonial
## 1.027134e+02 9.531444e+01 8.413926e-01 1.017415e+04 5.771916e+02 4.638161e-01
```

Question Two:

```
library(e1071)
# Skewness of the price distribution
skewness_price <- skewness(housing_data$price)
print(paste("Skewness of Price: ", skewness_price))
```

```
## [1] "Skewness of Price: 1.96488249462324"
```

```
mean_price <- mean(housing_data$price)
median_price <- median(housing_data$price)
print(paste("Mean Price: ", mean_price))
```

```
## [1] "Mean Price: 293.546034090909"
```

```
print(paste("Median Price: ", median_price))
```

```
## [1] "Median Price: 265.5"
```

(b) The price average in the dataset is 293.55 dollars.

```
sd_price <- sd(housing_data$price)
print(paste("Standard Deviation of Price: ", sd_price))
```

(c) The price average in the dataset is smaller than the median price.

```
## [1] "Standard Deviation of Price: 102.713445172284"
```

(d) The integer part of the price standard deviation in the dataset is 102.

Question Three:

(a) The relation between prices and house size seems to be quadratic.

(b) There is one outlier observation in the plot of price against lotsize.

```
table(housing_data$bdrms)
```

(c) Most houses have 3 or 4 bedrooms.

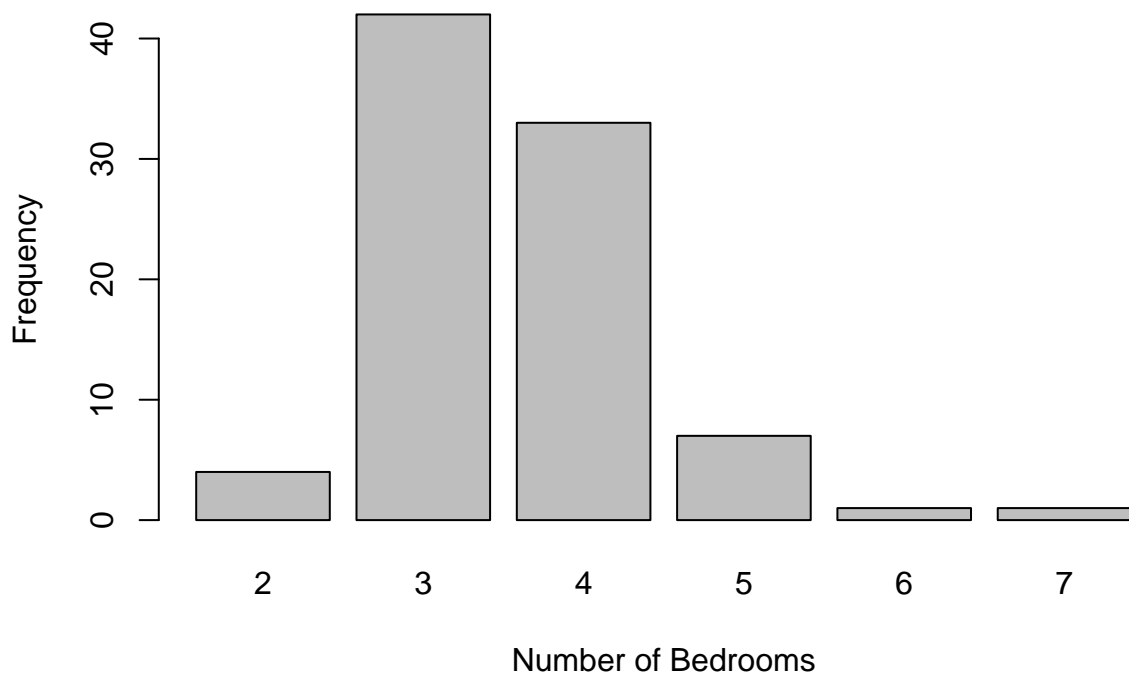
```
##
```

```
## 2 3 4 5 6 7
```

```
## 4 42 33 7 1 1
```

```
barplot(table(housing_data$bdrms), main="Distribution of Bedrooms", xlab="Number of Bedrooms", ylab="Fr
```

Distribution of Bedrooms



####

(d) Assessed prices are usually smaller than the selling prices.

Question Four:

```
mean_price_colonial <- mean(housing_data$price[housing_data$colonial == 1])
mean_price_noncolonial <- mean(housing_data$price[housing_data$colonial == 0])
print(paste("Mean Price - Colonial: ", mean_price_colonial))
```

(a) Colonial houses are, on average, more expensive than non-colonial houses.

```
## [1] "Mean Price - Colonial: 302.918868852459"
```

```
print(paste("Mean Price - Non-Colonial: ", mean_price_noncolonial))
```

```
## [1] "Mean Price - Non-Colonial: 272.37037037037"
```

```
var_price_colonial <- var(housing_data$price[housing_data$colonial == 1])
var_price_noncolonial <- var(housing_data$price[housing_data$colonial == 0])
print(paste("Variance of Price - Colonial: ", var_price_colonial))
```

(b) The price variance of colonial houses is smaller than the non-colonial ones.

```
## [1] "Variance of Price - Colonial: 9600.56373721585"
```

```
print(paste("Variance of Price - Non-Colonial: ", var_price_noncolonial))
```

```
## [1] "Variance of Price - Non-Colonial: 12475.1844729345"
```

(c) The price distribution of colonial houses is very different from that of non-colonial ones.

```
mean_bdrms_colonial <- mean(housing_data$bdrms[housing_data$colonial == 1])
mean_bdrms_noncolonial <- mean(housing_data$bdrms[housing_data$colonial == 0])
print(paste("Mean Bedrooms - Colonial: ", mean_bdrms_colonial))
```

(d) On average, colonial houses have more bedrooms than non-colonial ones.

```
## [1] "Mean Bedrooms - Colonial: 3.73770491803279"
```

```
print(paste("Mean Bedrooms - Non-Colonial: ", mean_bdrms_noncolonial))
```

```
## [1] "Mean Bedrooms - Non-Colonial: 3.18518518518519"
```