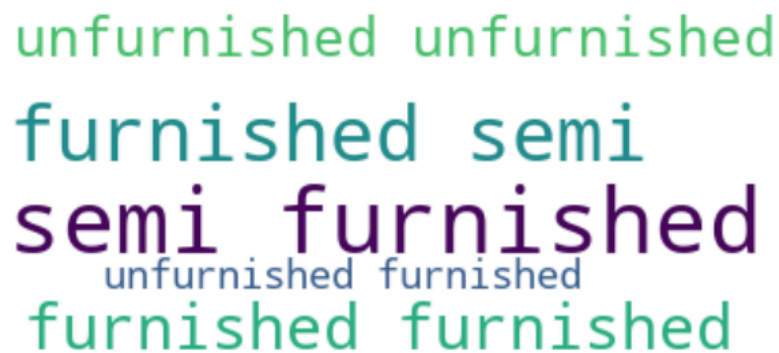| Name	Hrimkar Doshi
Uid	2021700021
Exp	5
Date	15/09/2024

AIM:

Create advanced charts using R programming language on the dataset - Housing data
Advanced - Word chart, Box and whisker plot, Violin plot, Regression plot (linear and
nonlinear), 3D chart, Jitter

RESULTS

Word Cloud

Observation

- This graph visualizes the frequency of different furnishing statuses ("furnished", "semi furnished", "unfurnished").
- "furnished" and "semi furnished" appear larger, suggesting they occur more frequently in the dataset, while "unfurnished" appears less frequently in comparison.

Code

```
library(wordcloud)
library(dplyr)
# Word chart for furnishingstatus
data <- read.csv("housing.csv")
furnish_status <- data %>% count(furnishingstatus)
wordcloud(words = furnish_status$furnishingstatus, freq = furnish_status$n, colors =
brewer.pal(8, "Dark2"))
```

**Boxplot of Price by Furnishing Status** :
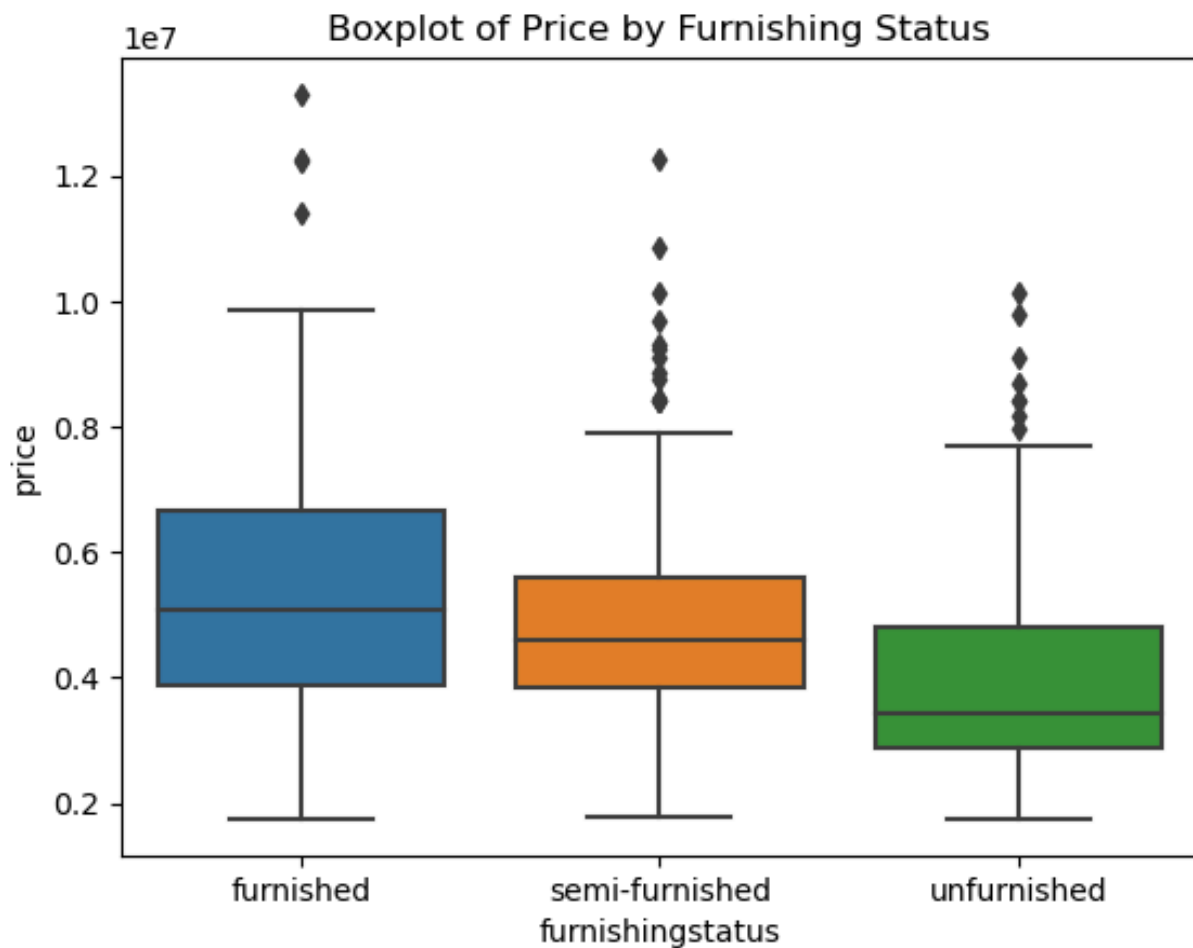
Observation
- The graph compares property prices across three categories: furnished, semi-furnished, and unfurnished.
- Furnished properties have a higher median price and a wider interquartile range (IQR), indicating more variability in their prices.
- Semi-furnished and unfurnished properties have lower median prices, with unfurnished properties showing the lowest overall price range.
- There are several outliers in each category, with the highest outliers found in the furnished category.

Code

```
boxplot(data$price ~ data$furnishingstatus, main="Boxplot of Price by Furnishing Status",
xlab="Furnishing Status", ylab="Price", col=c("lightblue", "lightgreen", "pink"))
```

3D Scatter Plot of Area vs. Bedrooms vs. Price :

3D Scatterplot
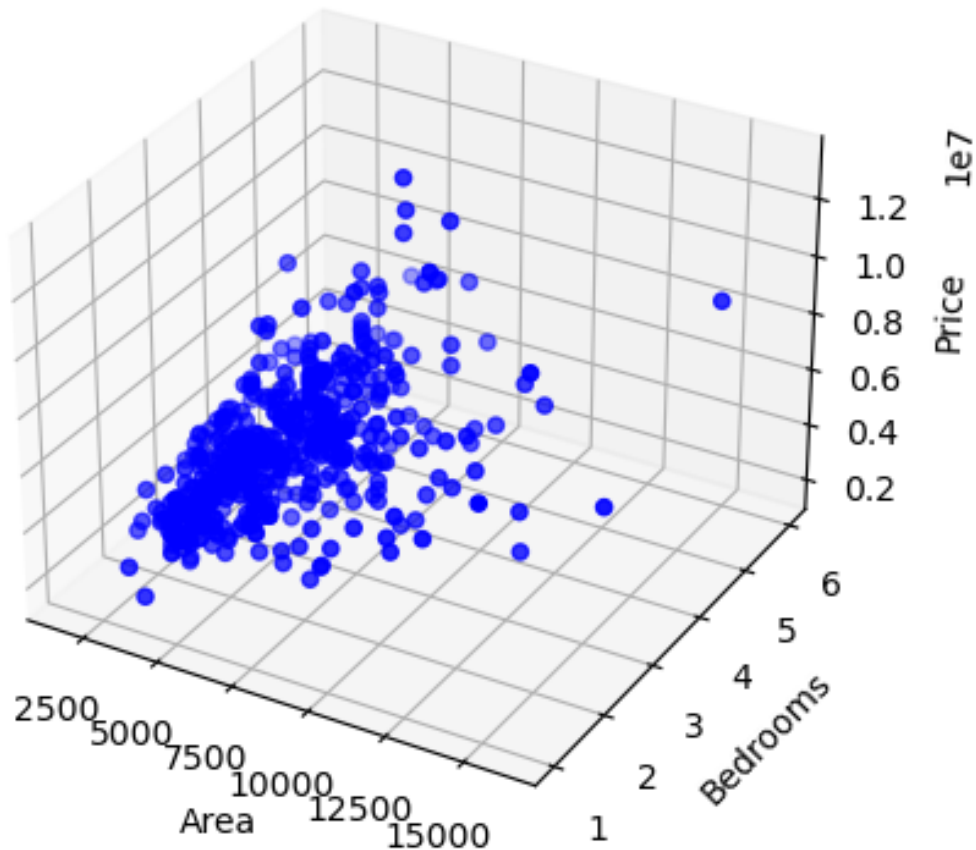
Observation
- This graph shows how area, number of bedrooms, and price relate to each other.
- Most data points cluster around properties with 2 to 4 bedrooms, an area of 2,500 to 12,500 sq. ft., and a price of 0.2 to 0.8 crores.
- Higher-priced properties (above 1 crore) seem to be associated with larger areas and more bedrooms, but the overall trend shows a concentration in mid-range areas and prices.

Code

```
library(scatterplot3d)

scatterplot3d(data$area, data$bedrooms, data$price, pch=16, color="blue", main="3D Scatterplot")
```

Jitter Plot of Price by Number of Bedrooms :


Jitter Plot of Price by Number of Bedrooms
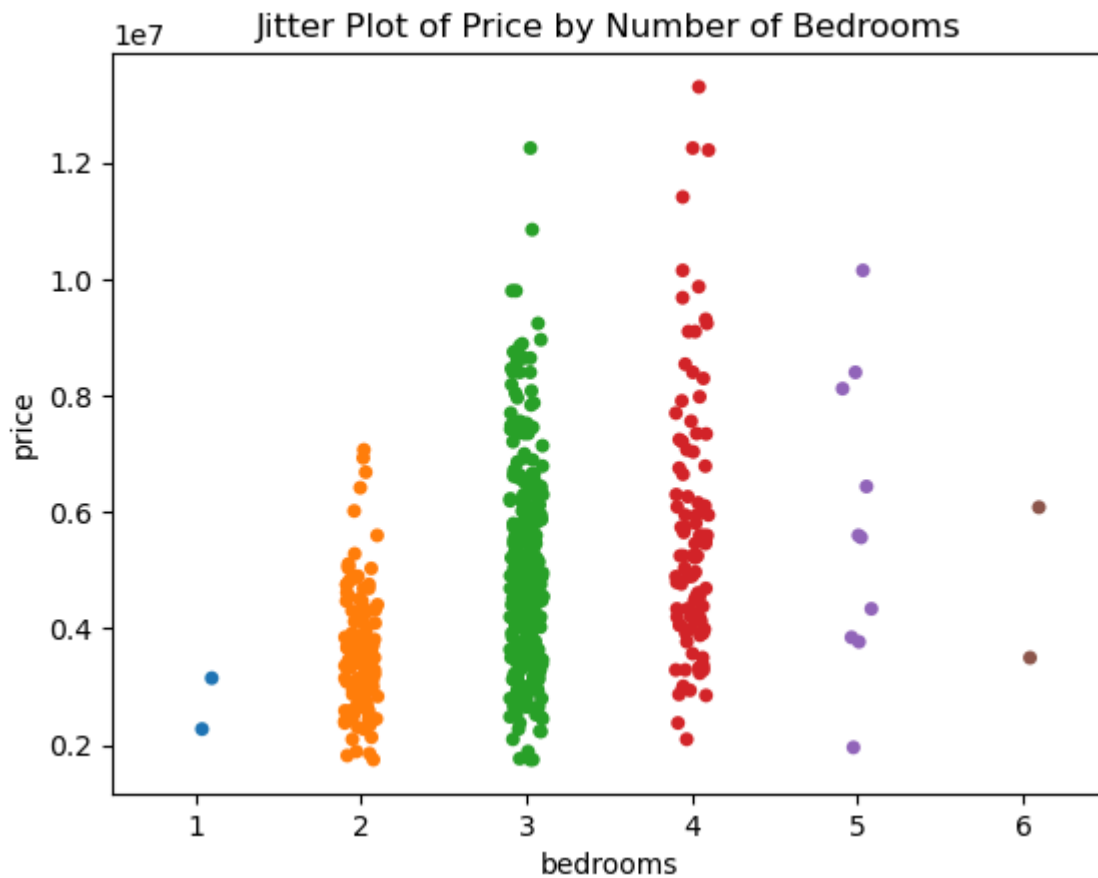
Observation

- This graph shows property prices distributed by the number of bedrooms.
- Properties with 3 and 4 bedrooms appear to be the most common.
- Prices generally increase as the number of bedrooms increases, with 1-2 bedroom properties priced on the lower end and 5-6 bedroom properties priced at the higher end.
- There is a significant variation in prices for properties with 3 and 4 bedrooms, suggesting that other factors besides the number of bedrooms (such as furnishing status or area) may influence price.

Code

```
ggplot(data, aes(x=factor(bedrooms), y=price)) +
  geom_jitter(width=0.2, height=0, color="darkblue") +
  theme_minimal() +
  labs(title="Jitter Plot of Price by Number of Bedrooms", x="Bedrooms")
```

Violin Plot of Price by Parking Spaces:
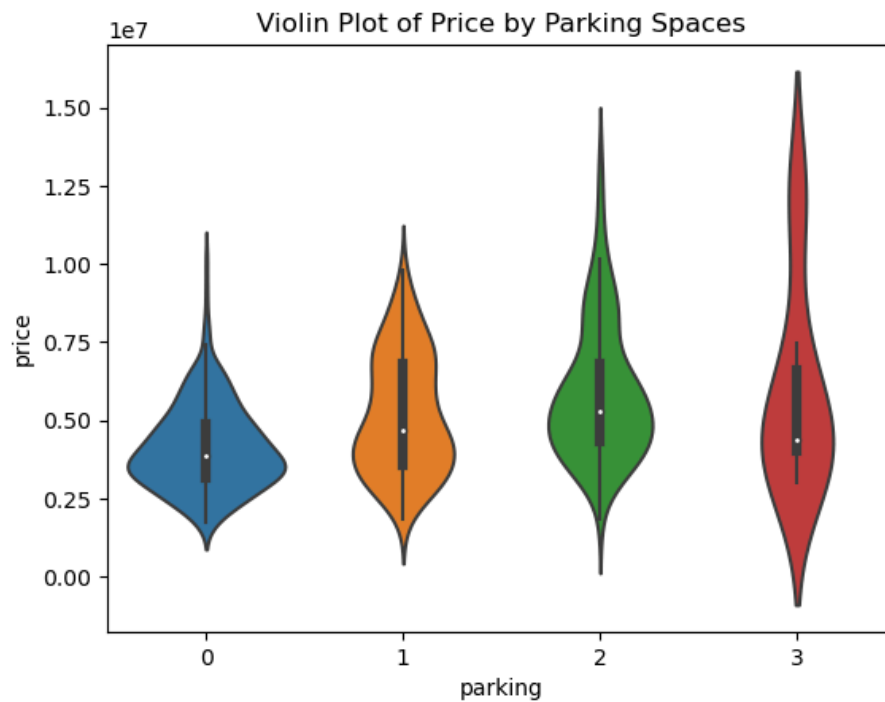
Observation

- This plot visualizes the distribution of house prices for different numbers of parking spaces.
- Properties with zero parking spaces have a wider range of prices in the lower spectrum, but fewer high-price outliers.
- As the number of parking spaces increases, the distribution becomes more stretched, especially for properties with 3 parking spaces, indicating a wider range of high prices.
- The median price (represented by the black bar inside the violin) increases with more parking spaces, suggesting that properties with more parking spaces tend to have higher prices.
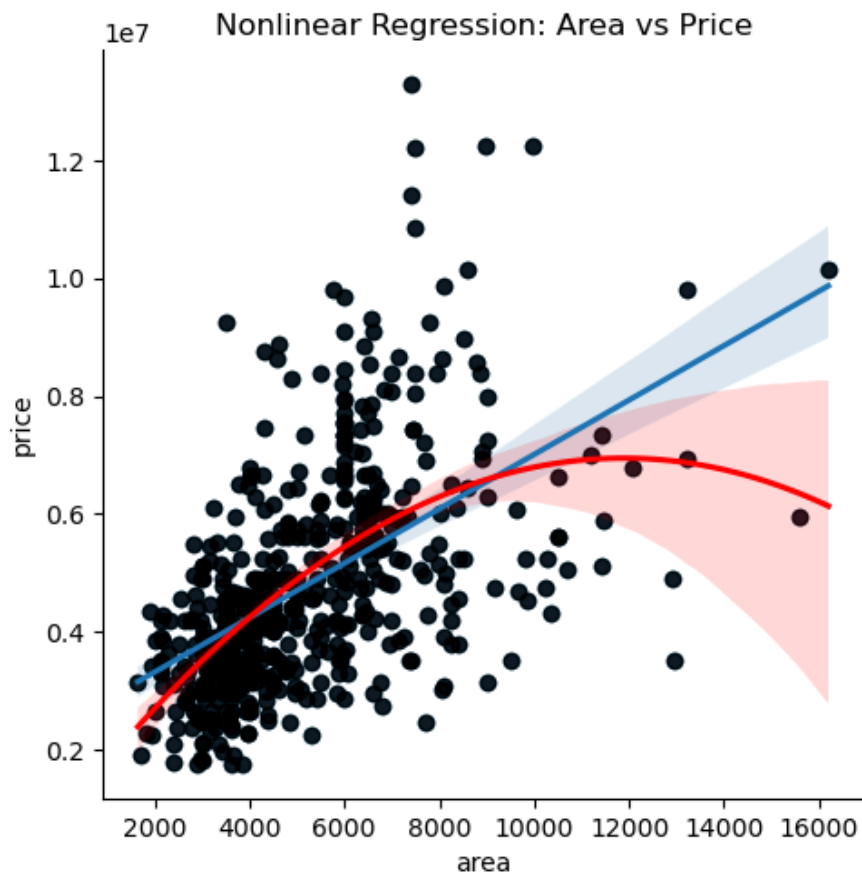
Code

```
sns.violinplot(x='parking', y='price', data=df)

plt.title('Violin Plot of Price by Parking Spaces')

plt.show()
```

Nonlinear Regression: Area vs Price:

Observation

- This scatter plot with a fitted nonlinear regression line shows the relationship between property area (x-axis) and price (y-axis).
- The red line suggests a nonlinear trend where price increases with area up to a certain point, but beyond approximately 10,000-12,000 sq ft, the price increase slows down.
- The blue line seems to indicate an alternative model where the price keeps increasing more rapidly, especially at higher area values.
- The shaded regions represent confidence intervals, showing more uncertainty in price predictions as the area increases.

Code

```
sns.regplot(x='bathrooms', y='price', data=df, scatter_kws={"color": "black"},
line_kws={"color": "red"}, order=2)

plt.title('Nonlinear Regression: Bathrooms vs Price')

plt.show()
```