Sardar Patel Institute of Technology
SEM VII:ADVANCE DATA VISUALIZATION.
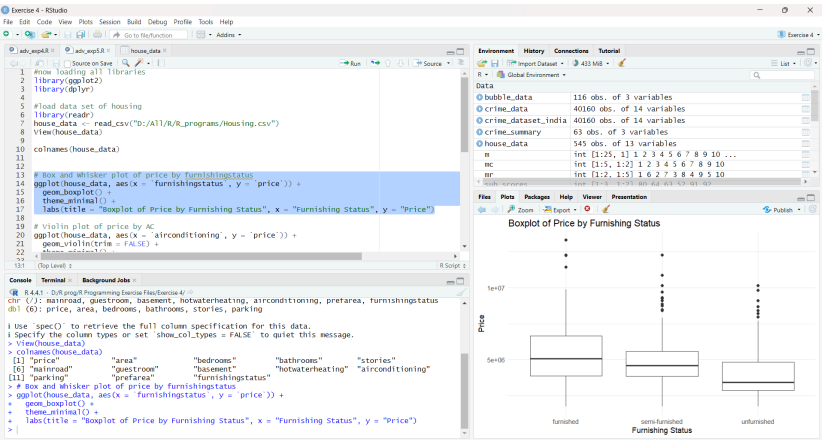
Name	Smruti Sonekar
UID no.	2021700064
Branch	BE CSE DS (BATCH B)
Experiment no.	5

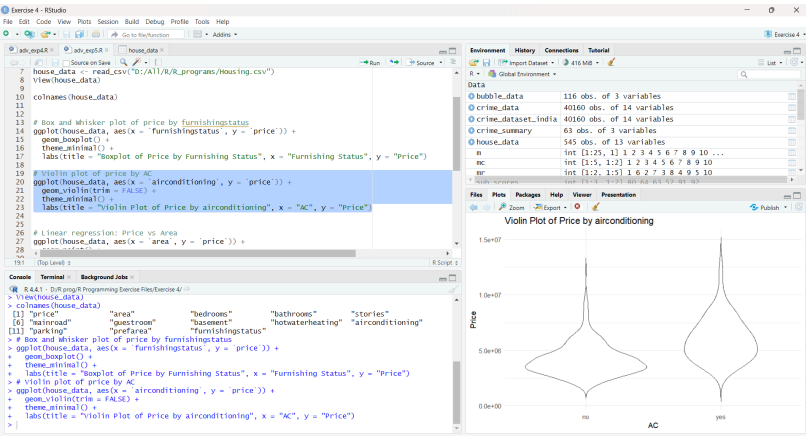| Topic:	Advance Data Visualization with R
Aim:	Create advanced charts using R programming language on the dataset - Housing data ● Advanced - Word chart, Box and whisker plot, Violin plot, Regression plot (linear and nonlinear), 3D chart, Jitter ● Write observations from each chart To explore and visualize housing data using advanced charts in R, including Word chart, Box and Whisker plot, Violin plot, Regression plot (linear and nonlinear), 3D chart, and Jitter plot, in order to uncover patterns and insights in the dataset. **Objectives:** 1. To visualize the distribution and relationship between various features in the housing dataset. 2. To identify potential outliers and understand the spread of the data. 3. To explore the relationship between independent variables and the target variable (e.g., house prices). 4. To create informative visualizations that can guide decision-making in the housing market.
Theory:	Link of DataSet: https://www.kaggle.com/datasets/yasserh/housing-prices-dataset This housing dataset contains information on various attributes of houses, including pricing and features. Here's a brief description of the columns: 1. **price**: The selling price of the house (numeric). 2. **area**: The total area of the house in square feet (numeric). 3. **bedrooms**: The number of bedrooms in the house (numeric). 4. **bathrooms**: The number of bathrooms in the house (numeric). 5. **stories**: The number of floors or stories in the house (numeric). 6. **mainroad**: Indicates if the house is located on the main road (binary: yes/no). 7. **guestroom**: Indicates if there is a guest room in the house (binary: yes/no). 8. **basement**: Indicates if the house has a basement (binary: yes/no). 9. **hotwaterheating**: Indicates if the house has hot water heating (binary: yes/no).

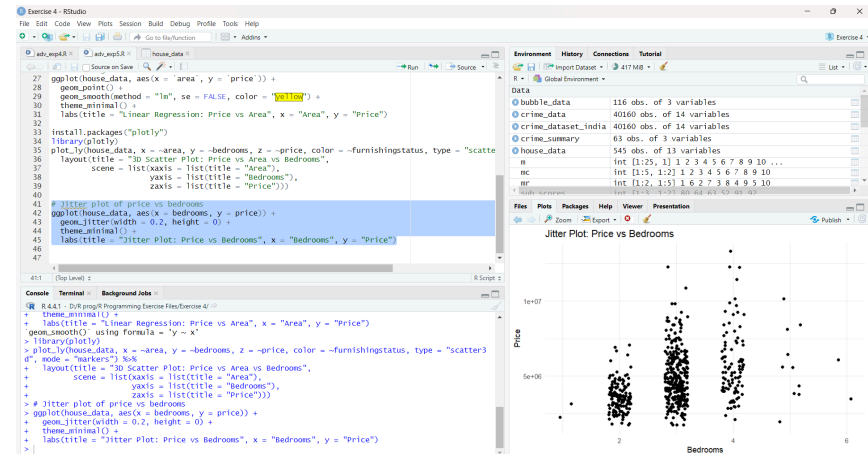|	
	10. **airconditioning**: Indicates if the house has air conditioning (binary: yes/no). 11. **parking**: The number of parking spaces available (numeric). 12. **prefarea**: Indicates if the house is in a preferred area (binary: yes/no). 13. **furnishingstatus**: The furnishing status of the house (categorical: furnished/semi-furnished/unfurnished).
Program:	1. **Box and Whisker Plot (Price by Furnishing Status)**: This plot shows the distribution of house prices based on whether the house is furnished, semi-furnished, or unfurnished. The box represents the interquartile range (IQR), and the line inside the box shows the median price. The "whiskers" extend to show the spread of the data, highlighting any outliers or extreme values. This plot helps compare price distributions for different furnishing categories. 2. **Violin Plot (Price by Air Conditioning)**: The violin plot provides a detailed view of the distribution of house prices, separated by whether the house has air conditioning or not. The shape of the plot shows the density of prices at different levels, similar to a box plot but with more information about the distribution's shape. Wider sections of the violin represent areas where price points are more concentrated. 3. **Linear Regression (Price vs Area)**: This scatter plot shows the relationship between the area of the house and its price. Each point represents a house, and the line represents the linear regression model, which estimates the relationship between area and price. The yellow regression line provides a clear visual of the trend: as the area of the house increases, the price tends to rise as well, indicating

a positive correlation.
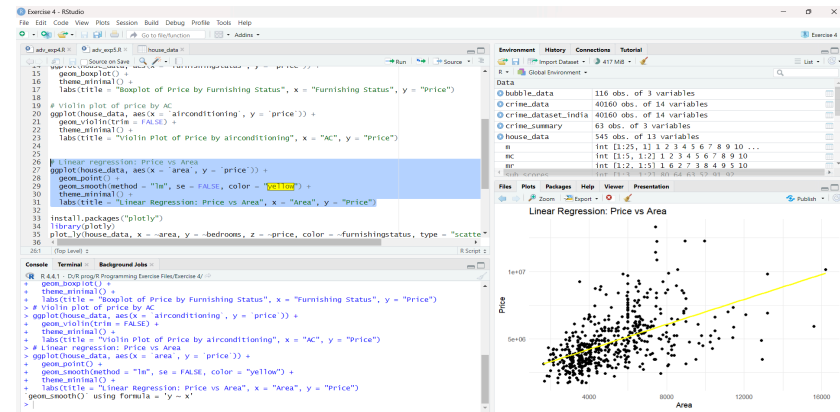
4. **3D Scatter Plot (Price vs Area vs Bedrooms)**: This 3D scatter plot visualizes the relationship between three variables: area, number of bedrooms, and price. Each point represents a house, with the color indicating the furnishing status. This plot helps in understanding how both area and the number of bedrooms together influence the house price, offering a more complex multi-dimensional analysis.
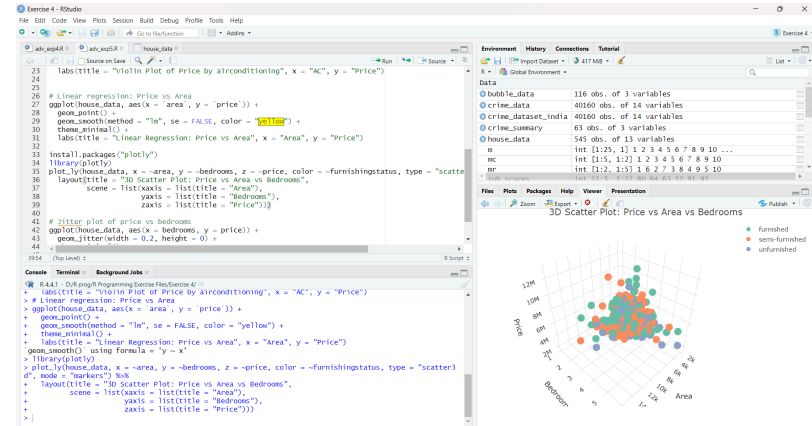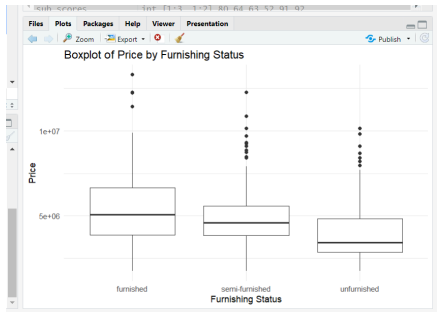

5. **Jitter Plot (Price vs Bedrooms)**: A jitter plot is used here to show the relationship between the number of bedrooms and the price of the house, with a slight random noise (jitter) added to avoid overlapping points. This visualization helps observe any patterns in pricing related to the number of bedrooms without data points being too cluttered together.
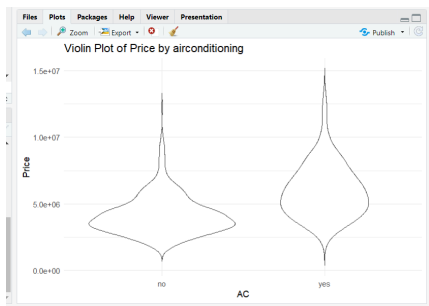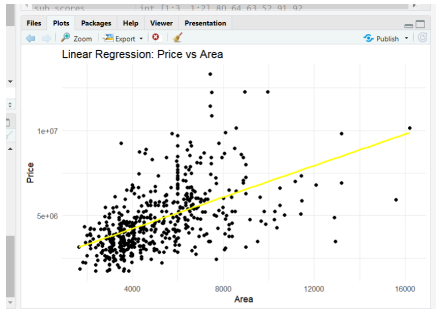
Result:	Box and Whisker Plot Violin Plot Regression Plot 3D Plot JitterPlot

Conclusion:	These visualizations highlight key trends in the housing dataset. Furnishing status and air conditioning influence house prices, with different price distributions for each category. A positive correlation is observed between house area and price, with larger houses generally costing more. The 3D plot reveals that both area and the number of bedrooms significantly affect price, while the jitter plot shows price variation across different bedroom counts. Overall, these charts provide valuable insights into how house features impact pricing.