



Name	Harditya Shah
UID No.	2021300114
Batch	L

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

Dataset used:

Crime stats data -

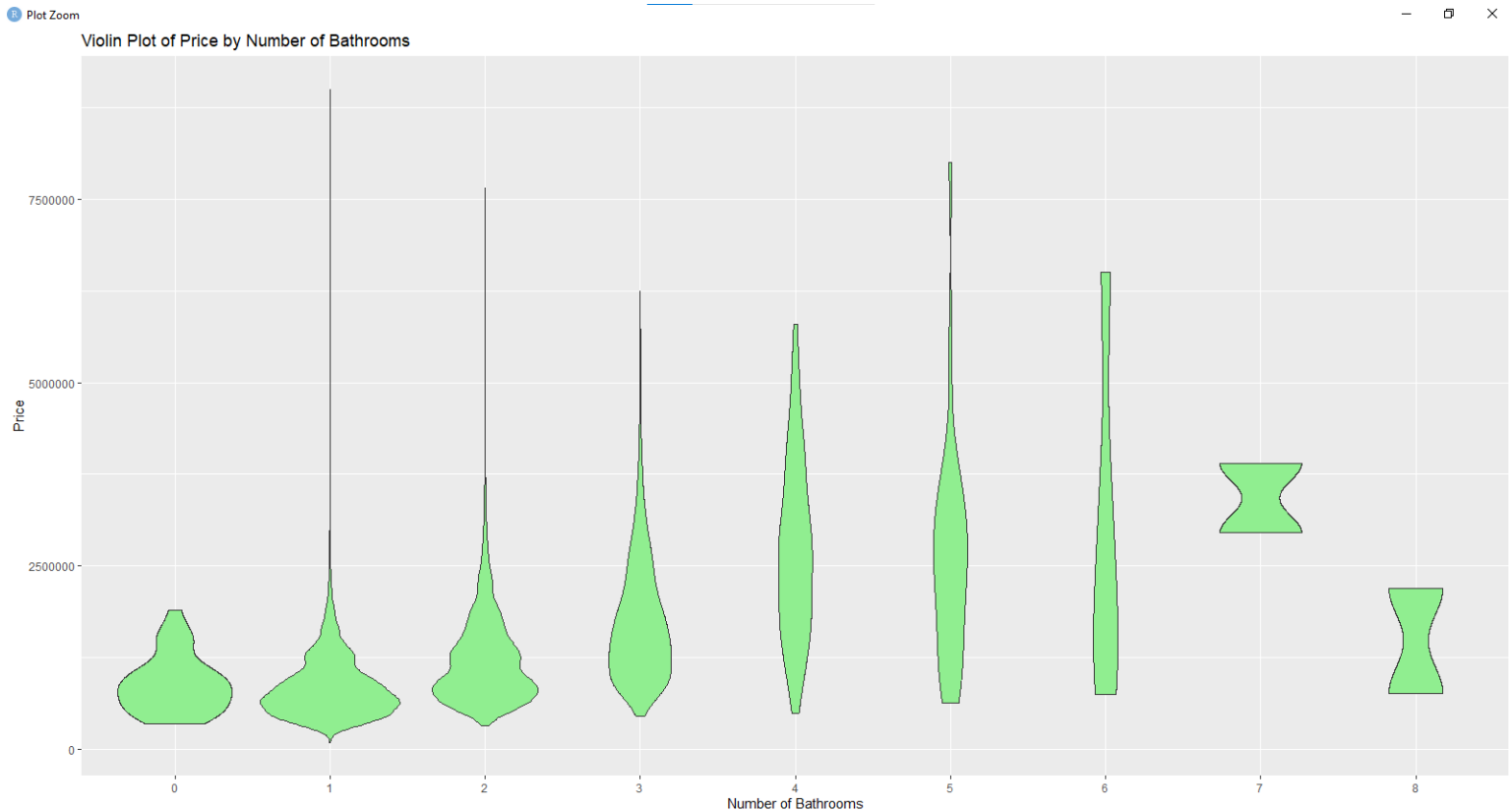
<https://www.kaggle.com/datasets/yasserh/housing-prices-dataset>

Objectives:

- To visualize the distribution and relationship between various features in the housing dataset.
- To identify potential outliers and understand the spread of the data.
- To explore the relationship between independent variables and the target variable (e.g., house prices).
- To create informative visualizations that can guide decision-making in the housing market.

1. Violin plot

```
ggplot(housing_data, aes(x=factor(Bathroom), y=Price)) +  
  geom_violin(fill="lightgreen") +  
  labs(title="Violin Plot of Price by Number of Bathrooms", x="Number of Bathrooms", y="Price")
```

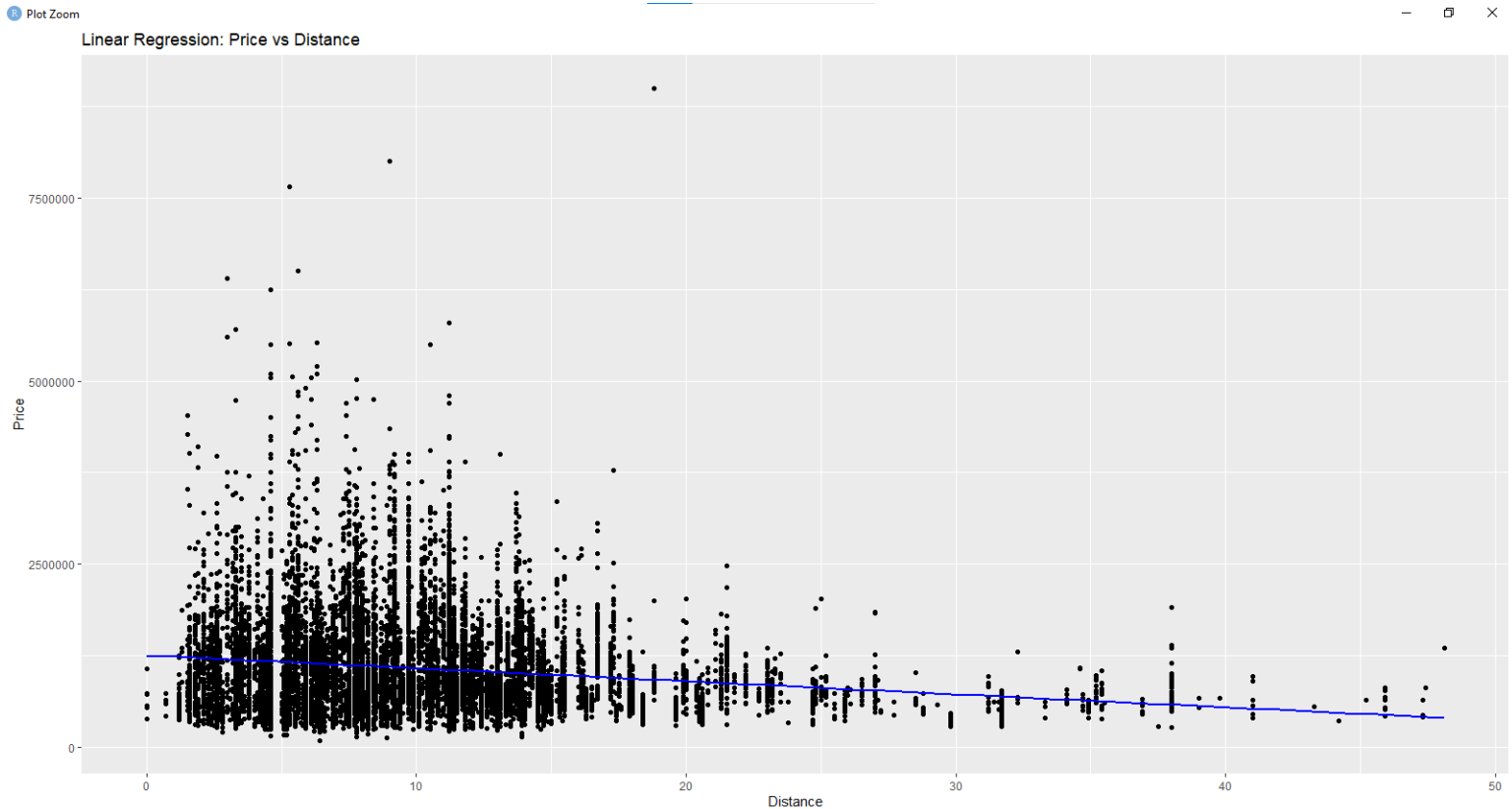


Observations:

- Homes with 1-2 bathrooms show a wide range of prices, while homes with more bathrooms (6-8) have more stable prices.
- Homes with 2-4 bathrooms tend to have prices within a narrower, more consistent range.
- Homes with fewer bathrooms (1-2) show significant price outliers compared to those with more bathrooms.

2. Linear Regression

```
ggplot(housing_data, aes(x=Distance, y=Price)) +  
  geom_point() +  
  geom_smooth(method='lm', se=FALSE, color="blue") +  
  labs(title="Linear Regression: Price vs Distance", x="Distance", y="Price")
```

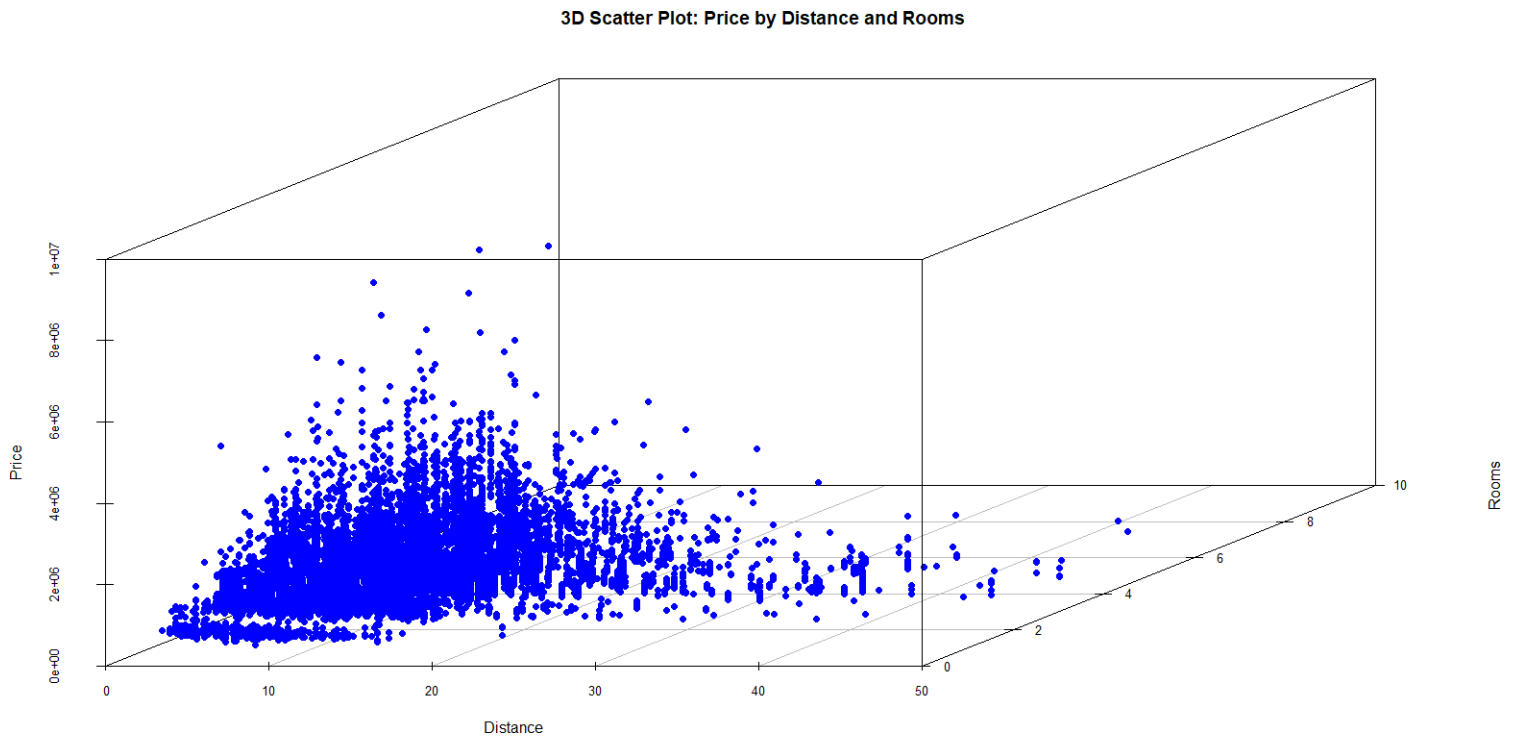


Observations:

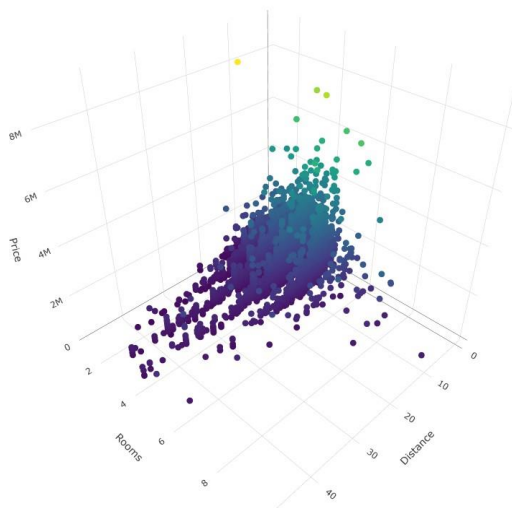
- As distance from the city increases, property prices tend to decrease, though the relationship is weak.
- Closer distances (below 10) show a wide range of prices, with some properties reaching very high prices. As distance increases, the price variability decreases, with fewer high-priced properties.
- There are a few properties with very high prices at both close and far distances, suggesting outliers.

3. 3D Scatter plot

```
scatterplot3d(housing_data$Distance, housing_data$Rooms, housing_data$Price,  
             xlab="Distance", ylab="Rooms", zlab="Price",  
             color="blue", pch=19, main="3D Scatter Plot: Price by Distance and Rooms")
```



Interactive 3D Scatter Plot: Price by Distance and Rooms

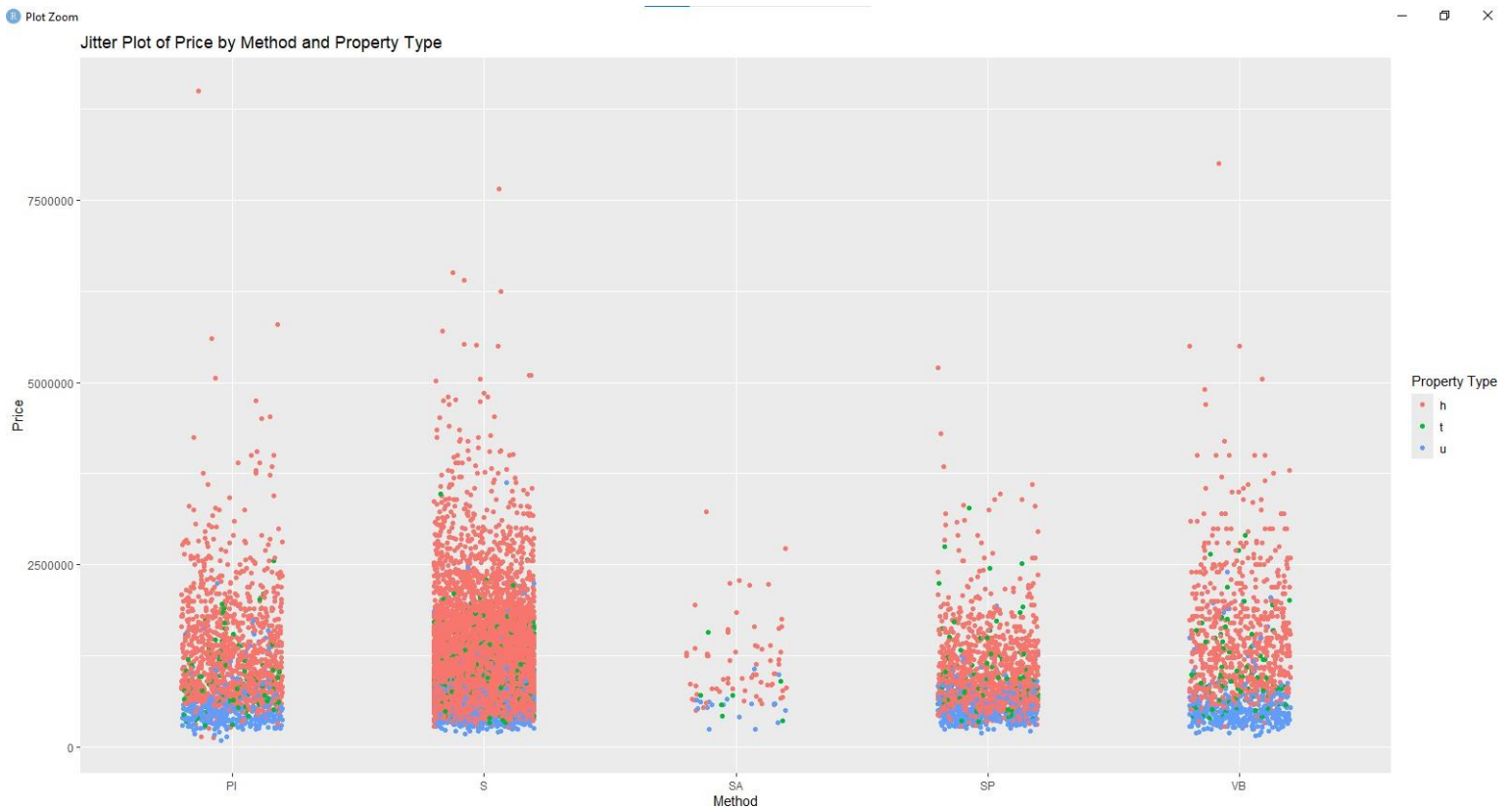


Observations:

- Price and Distance: Similar to the previous plot, prices generally decrease as the distance from the city center increases.
- Price and Rooms: Higher room counts (especially above 4) tend to correspond to higher prices, though there are still some lower-priced properties with many rooms.
- Most properties with moderate prices (under 2,000,000) and fewer rooms (up to 4) are clustered closer to the city (below 20 km).
- In short, distance negatively affects price, while a higher number of rooms positively correlates with price.

4. Jitter plot

```
ggplot(housing_data, aes(x=Method, y=Price, color=Type)) +  
  geom_jitter(width=0.2) +  
  labs(title="Jitter Plot of Price by Method and Property Type",  
        x="Method", y="Price", color="Property Type")
```



Observations:

- 'S' has the highest concentration of property sales with a wide spread in prices.
- 'SP' and 'SA' show a narrower price distribution.
- The red dots dominate the plot, indicating that this property type is the most common in the dataset.
- There's noticeable variation in prices depending on the method, with 'S' and 'VB' showing higher price points compared to 'SA'.