Name	Abhimanyu Kapoor
UID	2021300055
Experiment	5

Aim : Create advance charts using the R programming language on housing dataset.

- Advance- Word Cloud, Box plot, Violin plot, Jitter plot
- Write observations from each chart

Objectives:

- To understand and apply advance data visualization techniques in R.
- To create various types of charts using a housing-related dataset.
- To interpret and analyze the data through visual representations.

Theory:

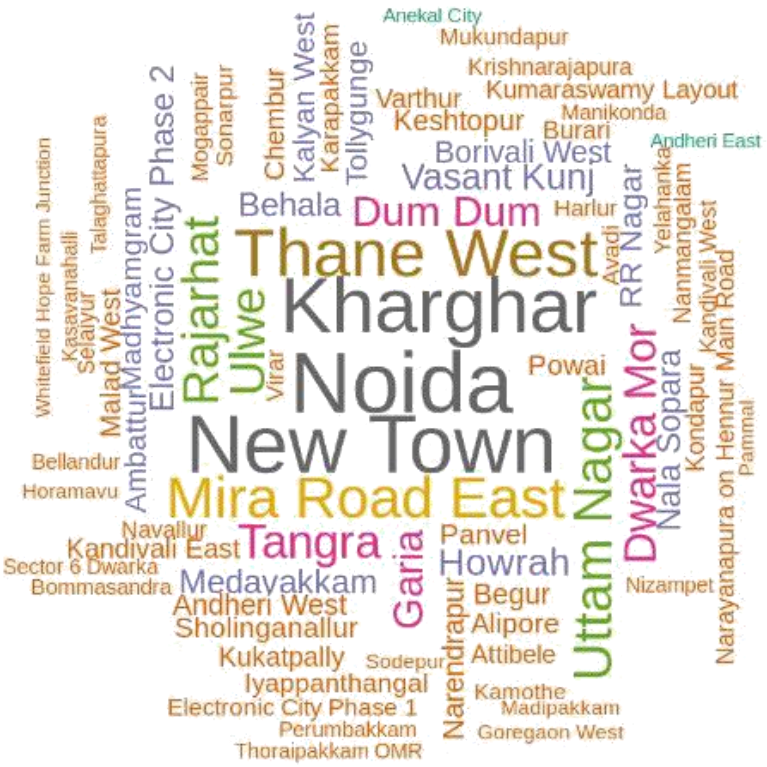
Data visualization is an essential skill in data analysis that helps in understanding trends, patterns, and relationships within a dataset. R, a powerful statistical programming language, provides a wide range of tools for creating visually appealing and informative charts. In this experiment, we will use advance chart types to analyze housing data and derive insights.

Dataset link :

https://www.kaggle.com/datasets/vijayjoshi17/housing-dataset-in-metropolitan-combined-dataset

1. Word Cloud :

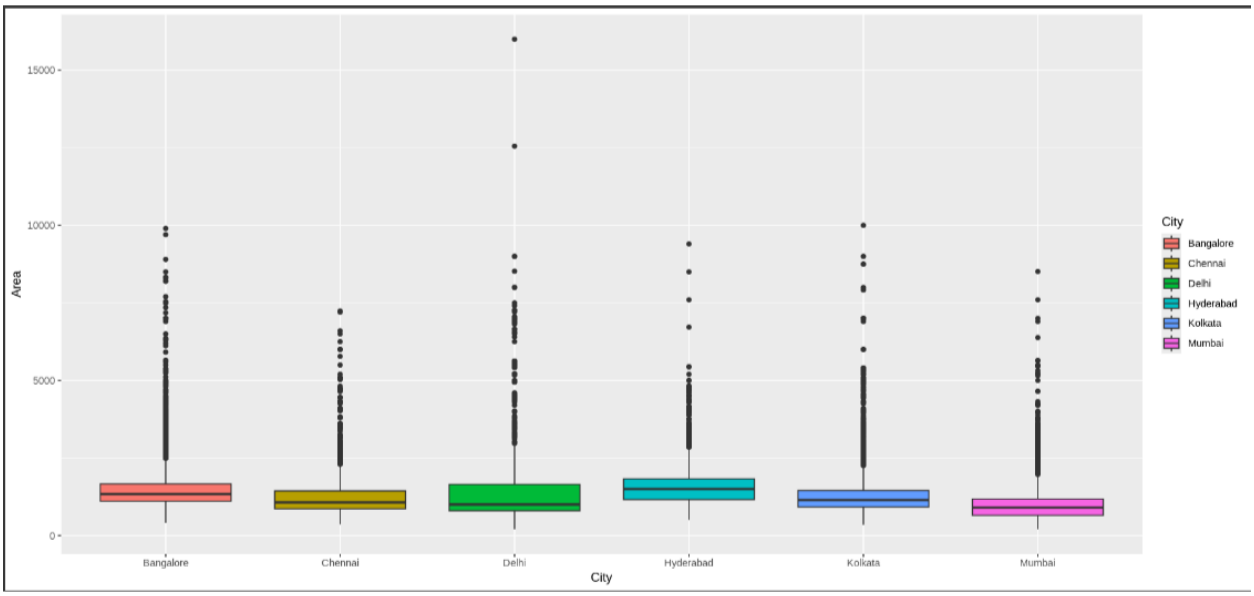
```
options(repr.plot.width = 15, repr.plot.height = 7)

location_count = housing_data %>% group_by(Location) %>% summarise(count = n(), .groups = 'drop')

wordcloud(words = location_count$Location, freq = location_count$count, min.freq = 5,
          max.words=75, random.order=FALSE,
          rot.per=0.35, colors=brewer.pal(8, "Dark2"))
```



Observation : Popular living areas are Noida, Kharghar, New Town and Thane West.

2. Box Plot:

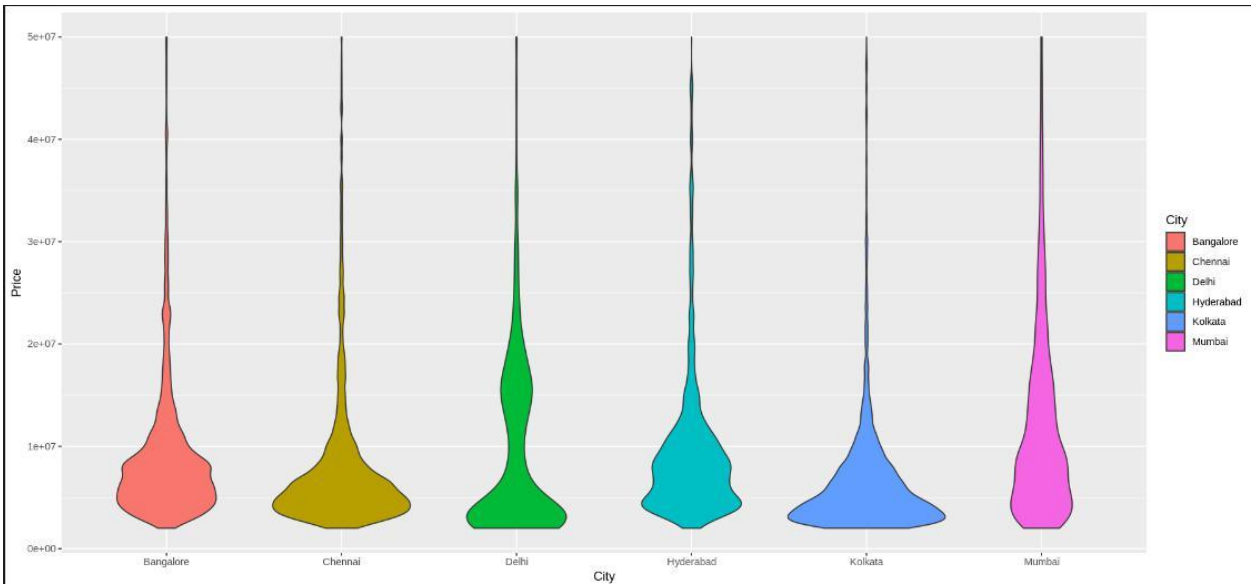
```
ggplot(housing_data, aes(x = City, y = Area, fill = City))
  + geom_boxplot()
```



Observation : Hyderabad and Bangalore have the highest price median while the median of Delhi is lowest and most of the prices are above median.

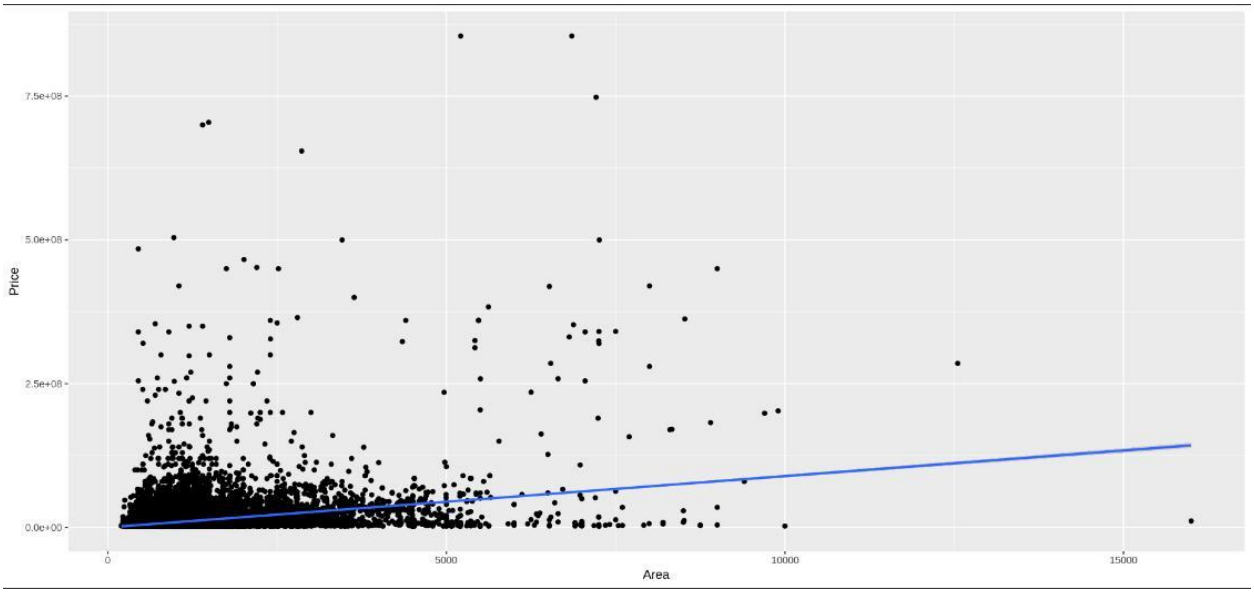
3. Violin Plot:

```
ggplot(housing_data, aes(x = City, y = Price, fill = City))
  + geom_violin(bounds = c(0, 5e7))
```



Observation : The plot for Bangalore and Hyderabad seems to be quite uniform indicating that prices may change depending on the area in both cities.

4. Regression Plot:



Conclusion :

From this experiment I was able to plot advance plots in R which can be used to provide more detailed insights about the data.