EXPT NO:6	OVER-PLOTTING REDUCTION TECHNIQUES
DATE: 09.02.2026	

PRE-LAB QUESTIONS

1. Why is over-plotting common in big data visualization?
   Over-plotting occurs when a large number of data points overlap in a limited visual space. In big data, millions of observations share similar or identical values, causing points to stack on top of each other and hide patterns.

2. How does data density affect perception?
   High data density overwhelms human visual perception. Dense regions appear as solid blocks, making it difficult to distinguish individual points, trends, clusters, or outliers.

3. What trade-offs exist between detail and clarity?

   - **More detail** → shows raw data but causes clutter

   - **More clarity** → reduces clutter but may hide individual variations

   Effective visualization balances accuracy with readability.

4. How do AI datasets increase visualization complexity?
   AI datasets are:

- High-dimensional
- Large-scale
- Continuous and dynamic

  This increases overlap, noise, and rendering cost, making traditional plots ineffective.

5. Why is over-plotting a serious analytical risk?
   Over-plotting can:

- Hide anomalies
- Mask trends
- Cause incorrect assumptions

  This leads to **wrong business or AI model decisions**.

OBJECTIVE : To apply techniques that reduce visual clutter in large-scale datasets.

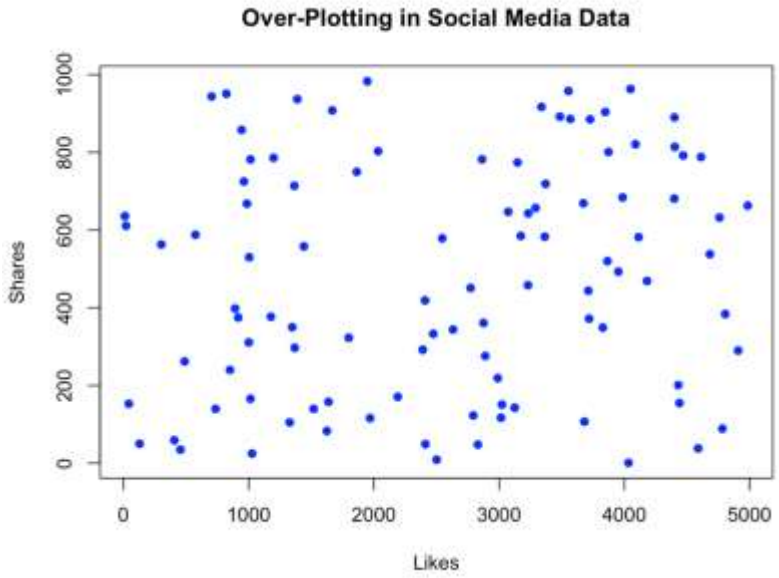
SCENARIO A social media analytics company visualizes millions of user interactions to study engagement patterns.

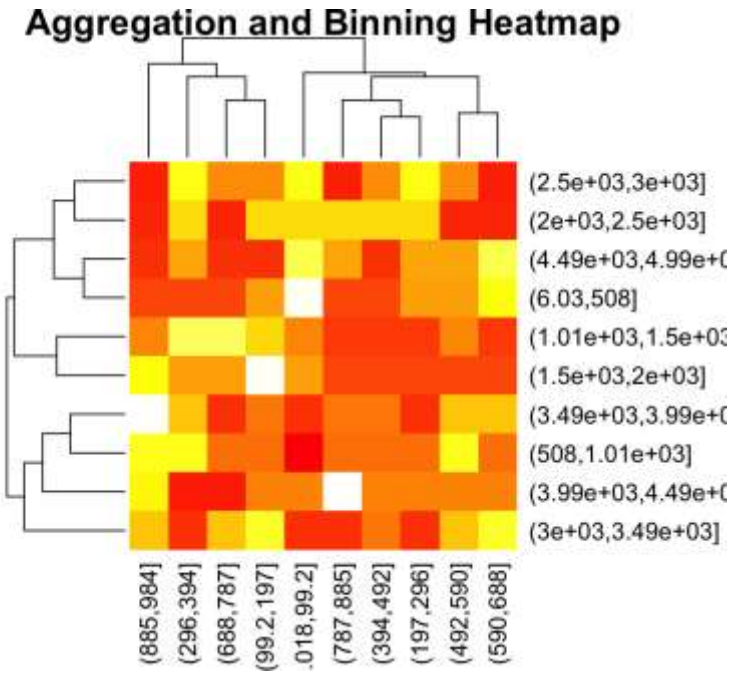
IN-LAB TASKS (Using R Language) • Apply alpha blending • Implement jittering techniques • Use aggregation and binning

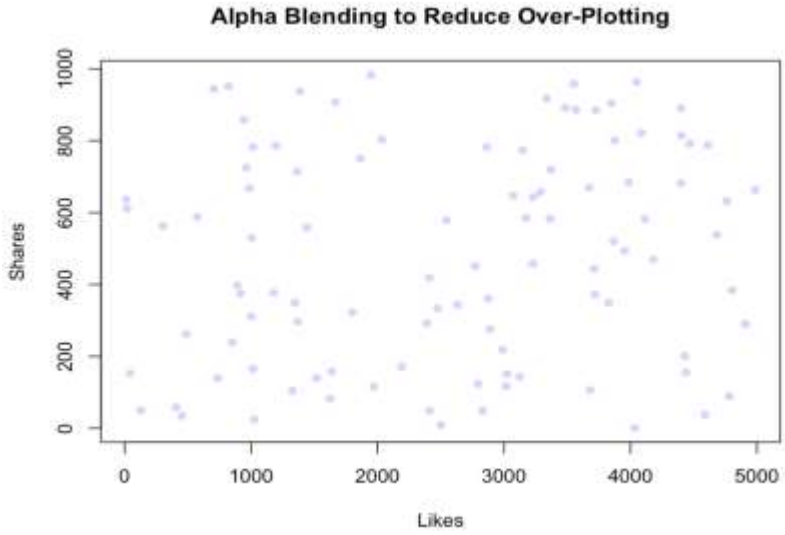
INPUT:

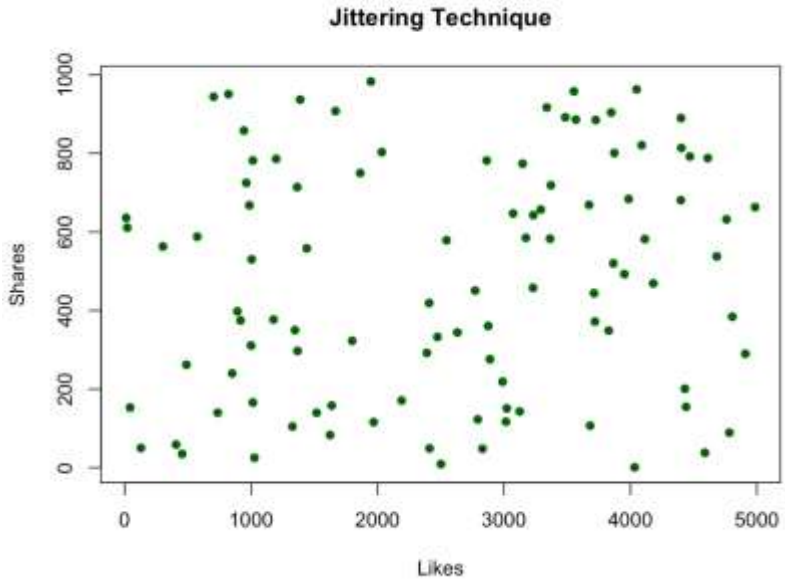
```r
# Convert to numeric (important)
data$Likes  <- as.numeric(data$Likes)
data$Shares <- as.numeric(data$Shares)

# Remove missing values
data <- na.omit(data)

# Basic scatter plot
plot(data$Likes, data$Shares,
     main = "Over-Plotting in Social Media Data",
     xlab = "Likes",
     ylab = "Shares",
     col = "blue",
     pch = 16)

# Alpha blending
plot(data$Likes, data$Shares,
     col = rgb(0, 0, 1, alpha = 0.2),
     pch = 16,
     main = "Alpha Blending to Reduce Over-Plotting",
     xlab = "Likes",
     ylab = "Shares")

# Jitter technique
plot(jitter(data$Likes),
     jitter(data$Shares),
     col = "darkgreen",
     pch = 16,
     main = "Jittering Technique",
     xlab = "Likes",
     ylab = "Shares")

# Aggregation / binning
likes_bins  <- cut(data$Likes, breaks = 10)
shares_bins <- cut(data$Shares, breaks = 10)

agg_data <- table(likes_bins, shares_bins)

heatmap(agg_data,
        col = heat.colors(256),
        main = "Aggregation and Binning Heatmap")
```

OUTPUT:


Over-Plotting in Social Media Data
Shares
Likes
0
200
400
600
800
1000
0
1000
2000
3000
4000
5000


Aggregation and Binning Heatmap
(2.5e+03,3e+03]
(2e+03,2.5e+03]
(4.49e+03,4.99e+0
(6.03,508]
(1.01e+03,1.5e+03
(1.5e+03,2e+03]
(3.49e+03,3.99e+0
(508,1.01e+03]
(3.99e+03,4.49e+0
(3e+03,3.49e+03]
(885,984]
(296,394]
(688,787]
(99.2,197]
.018,99.2]
(787,885]
(394,492]
(197,296]
(492,590]
(590,688]





POST-LAB QUESTIONS

1. Which technique provided the best clarity and why?
   Aggregation and binning provided the best clarity because it summarizes millions of points into meaningful density patterns without visual clutter.

2. How does over-plotting distort analytical conclusions?
   It hides true distributions, conceals outliers, and exaggerates or suppresses trends, leading to incorrect insights.
3. When should aggregation be preferred over raw plotting?
   Aggregation should be used when:

- Dataset size is very large
- Focus is on trends, not individual records
- Performance and scalability matter

4. How do these techniques support scalable AI analytics?

- Reduce rendering cost
- Improve interpretability
- Enable humans to validate AI outputs
- Support real-time visual monitoring

5. Explain real-world consequences of ignoring over-plotting.
   - Poor business decisions
   - Misleading AI model validation
   - Incorrect engagement analysis
   - Financial and reputational loss

LEARNING OUTCOME: Students master over-plotting reduction for big data visual analytics.

ASSESSMENT

Description	**Max Marks**	**Marks Awarded**
Pre Lab Exercise	**5**	
In Lab Exercise	**10**	
Post Lab Exercise	**5**	
Viva	**10**	
Total	**30**	
	Faculty Signature	