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 is common in big data visualization because large datasets contain thousands or millions of data points that overlap when plotted on the same axes. When many observations share similar or identical values, they are drawn on top of each other, making individual points indistinguishable.

2. How does data density affect perception?
High data density causes points to cluster tightly, making dense regions appear darker or cluttered. This hides patterns, trends, and outliers, leading to misinterpretation. Low-density areas may appear more important than they actually are, while dense areas may be underestimated.

3. What trade-offs exist between detail and clarity?
Showing all raw data points preserves detail but reduces clarity due to clutter. Reducing over-plotting improves clarity but may hide individual data variations. Therefore, a balance must be maintained between accurately representing data and ensuring visual readability.

4. How do AI datasets increase visualization complexity?
AI datasets are typically large, high-dimensional, and continuously generated. They include millions of interactions, features, and timestamps, which increases overlap and visual clutter, making traditional plotting methods ineffective without over-plotting reduction techniques.

5. Why is over-plotting a serious analytical risk?
Over-plotting can hide true patterns, exaggerate trends, or conceal outliers. This may lead to incorrect conclusions, flawed decisions, and biased models, especially in data-driven and AI-based systems.

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

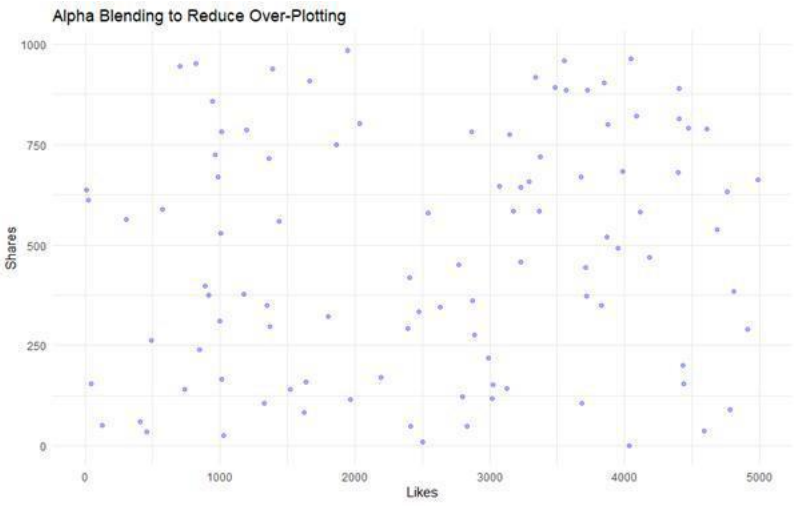
SCENARIO A social media analytics company visualizes millions of user interactions to study engagement patterns.

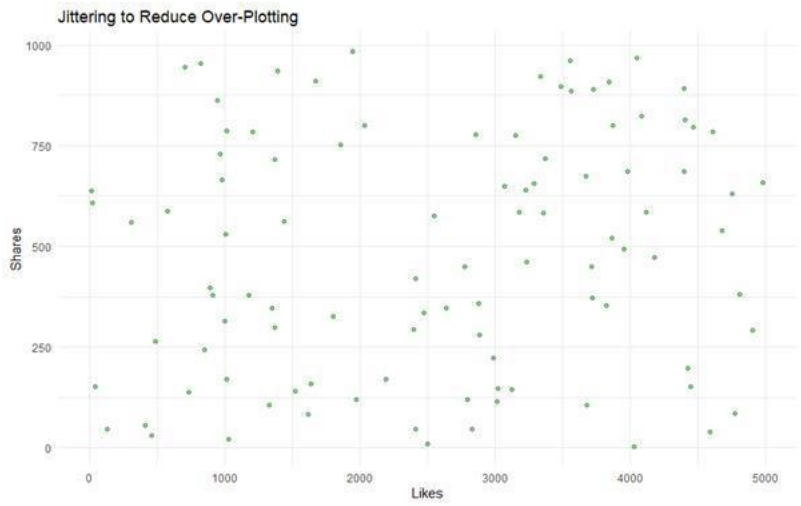
IN-LAB TASKS (Using R Language) • Apply alpha blending • Implement jittering techniques • Use aggregation and binning

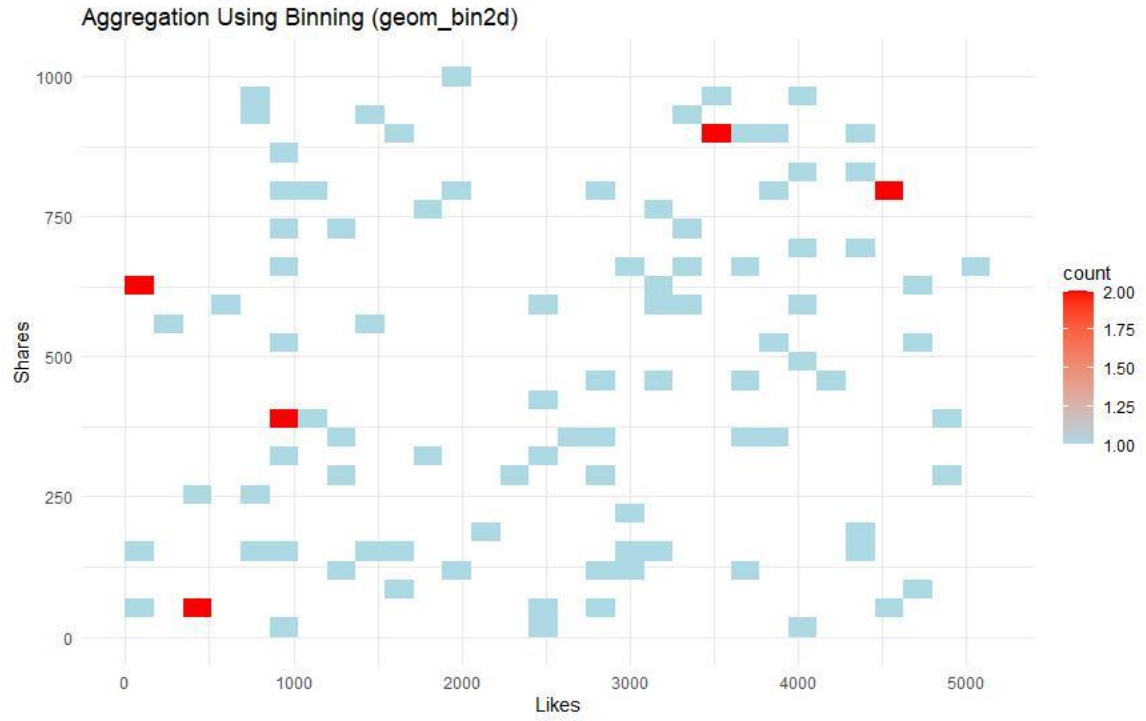
CODE:

```
# ----------------------------------------
# EXPERIMENT NO: 6
# Roll number: 23BAD095
# OVER-PLOTTING REDUCTION TECHNIQUES
# ----------------------------------------

library(ggplot2)
data <- read.csv("C:/Users/student.DESKTOP-D0TBOQD/Downloads/7.social_media_interactions.csv")
head(data)

# ----------------------------------------
# 1. NORMAL SCATTER PLOT (Over-plotting occurs)
# ----------------------------------------
ggplot(data, aes(x = Likes, y = Shares)) +
  geom_point() +
  ggtitle("Normal Scatter Plot (Over-Plotting)") +
  theme_minimal()

# ----------------------------------------
# 2. ALPHA BLENDING
# ----------------------------------------
ggplot(data, aes(x = Likes, y = Shares)) +
  geom_point(alpha = 0.3, color = "blue") +
  ggtitle("Alpha Blending to Reduce Over-Plotting") +
  theme_minimal()

# ----------------------------------------
# 3. JITTERING
# ----------------------------------------
ggplot(data, aes(x = Likes, y = Shares)) +
  geom_jitter(width = 5, height = 5, alpha = 0.4, color = "darkgreen") +
  ggtitle("Jittering to Reduce Over-Plotting") +
  theme_minimal()

# ----------------------------------------
# 4. AGGREGATION USING BINNING
# ----------------------------------------
ggplot(data, aes(x = Likes, y = Shares)) +
  geom_bin2d() +
  scale_fill_gradient(low = "lightblue", high = "red") +
  ggtitle("Aggregation Using Binning (geom_bin2d)") +
  theme_minimal()

# ----------------------------------------
# 5. HEXBIN AGGREGATION (Optional but Good)
# ----------------------------------------
ggplot(data, aes(x = Likes, y = Shares)) +
  geom_hex() +
  scale_fill_gradient(low = "yellow", high = "purple") +
  ggtitle("Hexagonal Binning for High Density Data") +
  theme_minimal()
```

OUTPUT:







POST-LAB QUESTIONS

1. Which technique provided the best clarity and why?
Aggregation and binning provided the best clarity because it grouped dense data points into bins, clearly showing overall patterns and distribution without clutter caused by individual overlapping points.

2. How does over-plotting distort analytical conclusions?
Over-plotting hides dense regions and trends, making sparse areas appear more significant. This can lead to incorrect assumptions about relationships, correlations, and data distribution.

3. When should aggregation be preferred over raw plotting?
Aggregation should be preferred when working with very large datasets where individual data points are less important than overall trends, patterns, or density distributions.

4. How do these techniques support scalable AI analytics?
These techniques enable clear visualization of large datasets, helping analysts validate patterns, detect anomalies, and interpret AI model behavior efficiently even as data volume scales.

5. Explain real-world consequences of ignoring over-plotting.
Ignoring over-plotting can result in incorrect business decisions, misleading analytics reports, biased AI models, and failure to detect critical trends such as declining user engagement or abnormal activity.

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	