EXPT NO:7	OVER-PLOTTING REDUCTION TECHNIQUES
DATE: 09.02.2026	

PRE-LAB QUESTIONS

1. **Why is over-plotting common in big data visualization?**
   Because millions of data points are plotted in limited screen space, causing points to overlap and hide patterns.
2. **How does data density affect perception?**
   High density creates dark clusters that dominate attention, while sparse areas may be overlooked, leading to biased interpretation.
3. **What trade-offs exist between detail and clarity?**
   Showing all data preserves detail but reduces readability; summarizing data improves clarity but may hide rare or important patterns.
4. **How do AI datasets increase visualization complexity?**
   AI datasets are large, high-dimensional, and continuous, making direct plotting cluttered and harder to interpret.
5. **Why is over-plotting a serious analytical risk?**
   It can conceal trends, outliers, and relationships, leading to incorrect conclusions and poor decision-making.

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

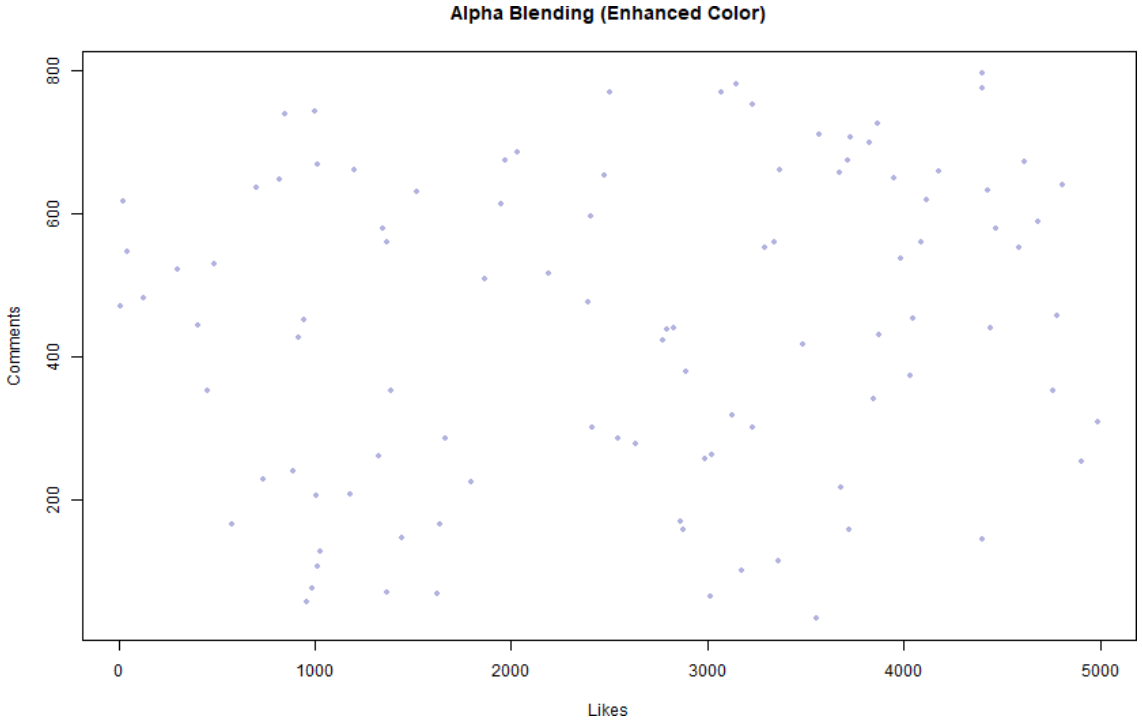
SCENARIO A social media analytics company visualizes millions of user interactions to study engagement patterns.

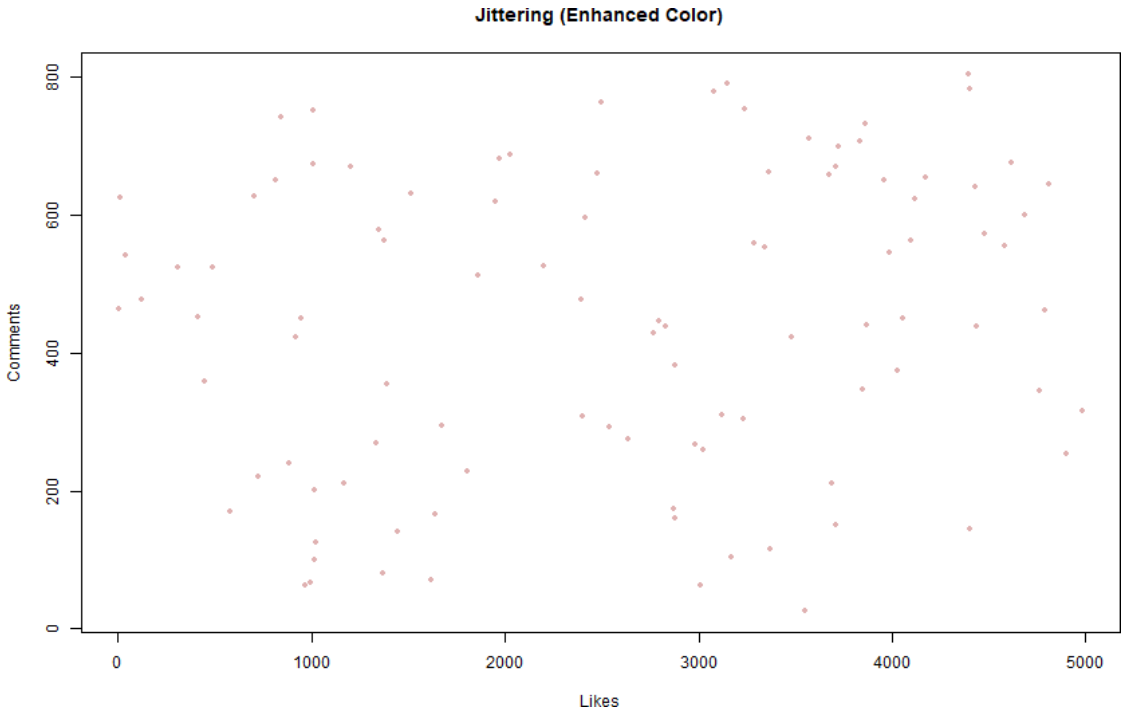
IN-LAB TASKS (Using R Language) • Apply alpha blending • Implement jittering techniques • Use aggregation and binning

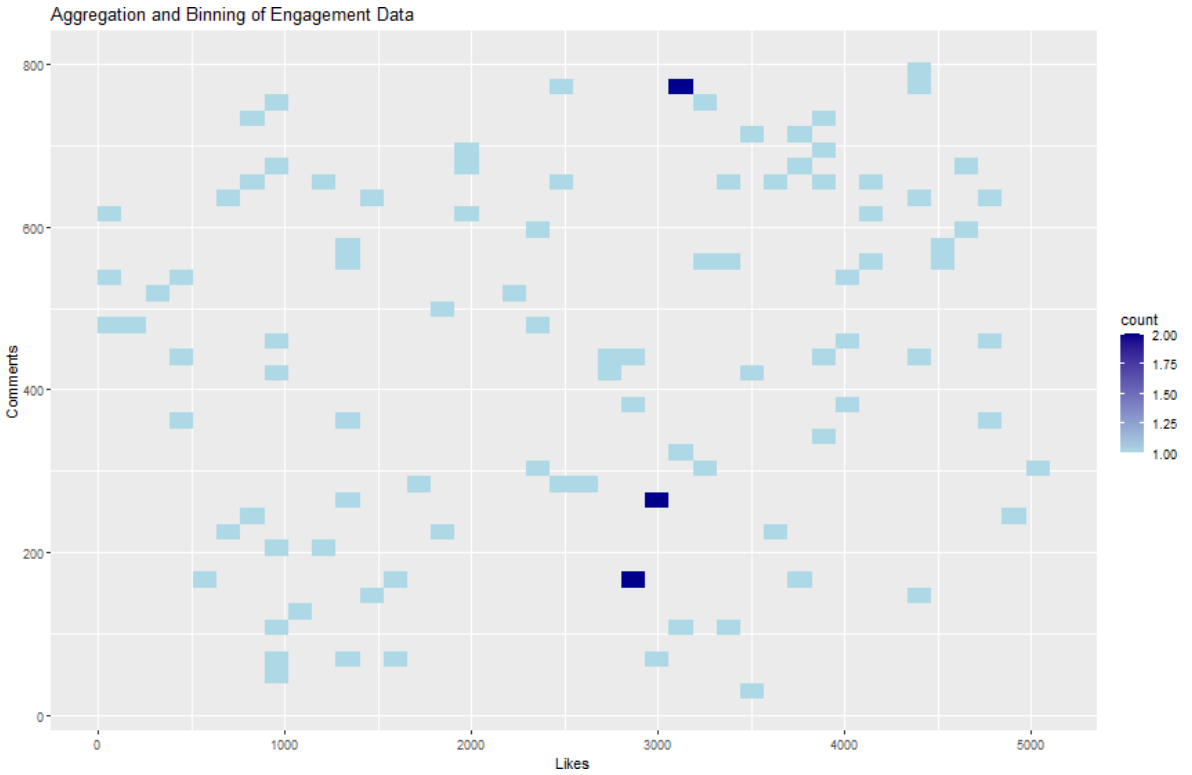
CODE:

```r
# ========================================
# Roll No: 23BAD101
# Experiment: Reducing Visual Clutter
# ========================================
# Load libraries
library(readxl)
library(ggplot2)
# ==============================
# 1. Import Dataset
# ==============================
df <- read_excel("C:\Users\student\Downloads\7.social_media_interactions.xlsm")
# ==============================
# 2. Basic Cleaning
# ==============================
df <- na.omit(df)
df <- unique(df)
# ==============================
# 3. Select Engagement Columns
# ==============================
x <- df$Likes
y <- df$Comments
# ==============================
# 4. Alpha Blending
# ==============================
plot(df$Likes, df$Comments,
     pch = 16,
     cex = 0.8,
     col = rgb(0, 0, 0.6, alpha = 0.3),    # darker blue
     xlab = "Likes",
     ylab = "Comments",
     main = "Alpha Blending (Enhanced Color)")
# ==============================
# 5. Jittering
# ==============================
plot(jitter(df$Likes, amount = 10),
     jitter(df$Comments, amount = 10),
     pch = 16,
     cex = 0.8,
     col = rgb(0.6, 0, 0, alpha = 0.3),    # darker red
     xlab = "Likes",
     ylab = "Comments",
     main = "Jittering (Enhanced Color)")
# ==============================
# 6. Aggregation & Binning
# ==============================
ggplot(df, aes(Likes, Comments)) +
  stat_bin2d(bins = 40) +
  scale_fill_gradient(low = "lightblue", high = "darkblue") +
  labs(title = "Aggregation and Binning of Engagement Data",
       x = "Likes",
       y = "Comments")

```


Alpha Blending (Enhanced Color)
Comments
Likes
0
1000
2000
3000
4000
5000
200
400
600
800


Jittering (Enhanced Color)
Comments
Likes
0
1000
2000
3000
4000
5000
0
200
400
600
800



POST-LAB QUESTIONS

1. **Which technique provided the best clarity and why?**
   Aggregation and binning gave the best clarity because they summarize millions of points into density regions, revealing overall engagement patterns clearly.
2. **How does over-plotting distort analytical conclusions?**
   It hides true data distribution, masks clusters and outliers, and can falsely suggest uniform or misleading trends.
3. **When should aggregation be preferred over raw plotting?**
   When datasets are very large or dense and individual points overlap excessively, making raw plots unreadable.
4. **How do these techniques support scalable AI analytics?**
   They enable efficient visualization of massive AI datasets by preserving patterns while reducing noise and computational load.
5. **Explain real-world consequences of ignoring over-plotting.**
   Misinterpreting user behavior, incorrect model insights, poor business decisions, and flawed AI-driven recommendations can result.

ASSESSMENT

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