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 because big data contains a very large number of data points. When many points share similar or identical values, they overlap on the graph, making individual points difficult to distinguish.

2. How does data density affect perception?
   High data density causes visual clutter, which makes it difficult to identify patterns, trends, or outliers. Dense regions may appear as solid blocks, hiding important information and misleading interpretation.

3. What trade-offs exist between detail and clarity?
   Showing every data point preserves detail but reduces clarity due to overlap. Reducing detail using techniques like aggregation or transparency improves clarity but may hide individual data points. A balance is required depending on the analysis goal.

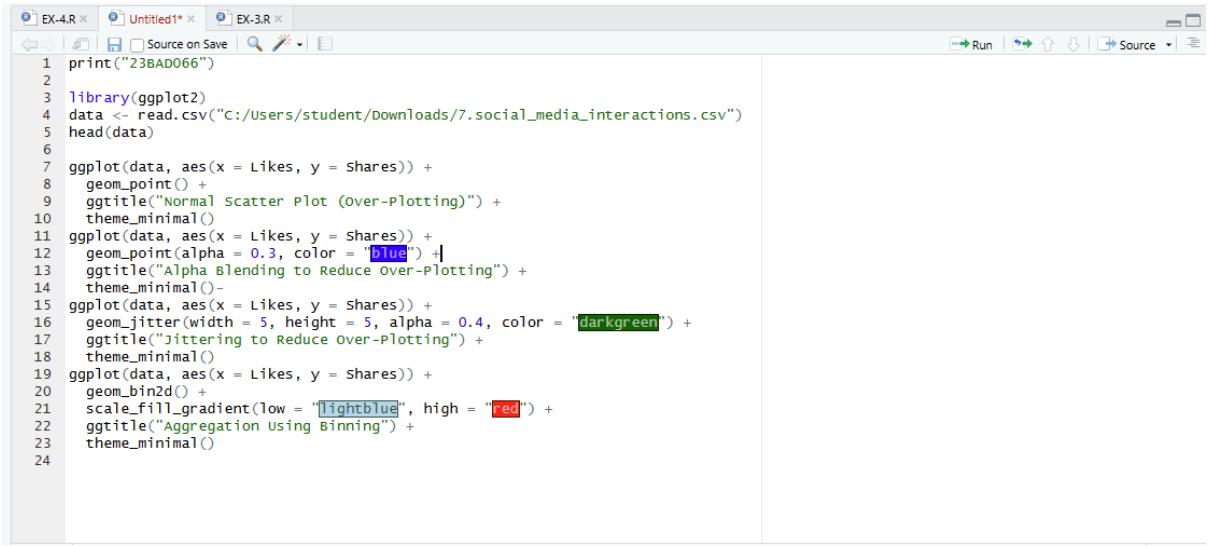
4. How do AI datasets increase visualization complexity?
   AI datasets are often large, high-dimensional, and continuously generated. They include multiple features and complex relationships, which increase overlap, clutter, and difficulty in visual interpretation.

5. Why is over-plotting a serious analytical risk?
   Over-plotting can hide trends, mask outliers, and create false impressions of data distribution. This can lead to incorrect conclusions, poor decisions, and biased model evaluation.

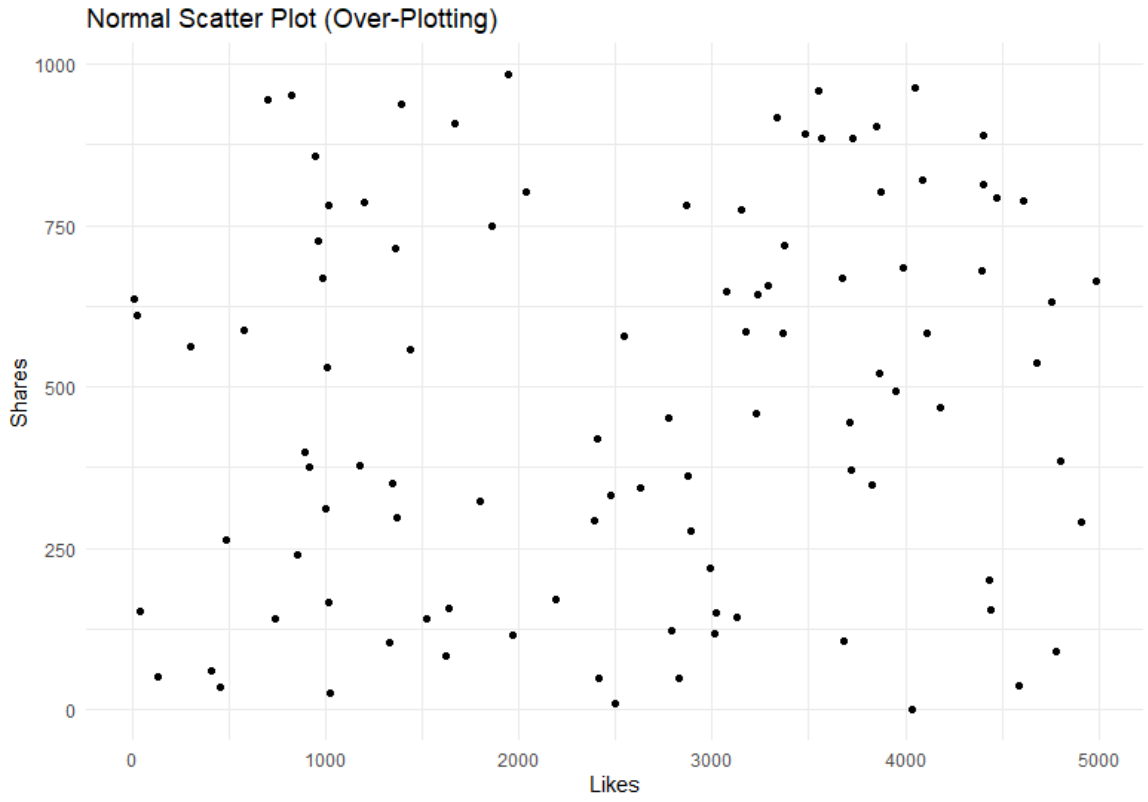
OBJECTIVE : To apply techniques that reduce visual clutter in large-scale datasets.

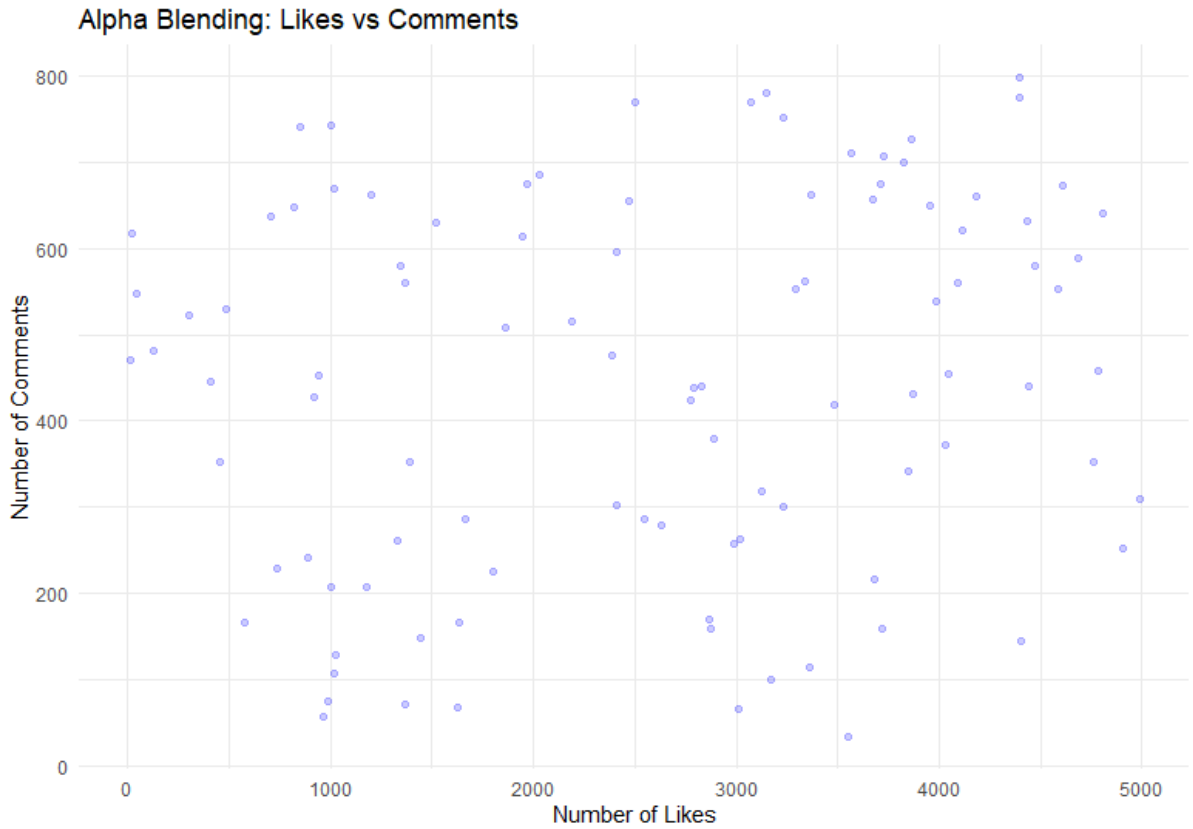
SCENARIO A social media analytics company visualizes millions of user interactions to study engagement patterns.

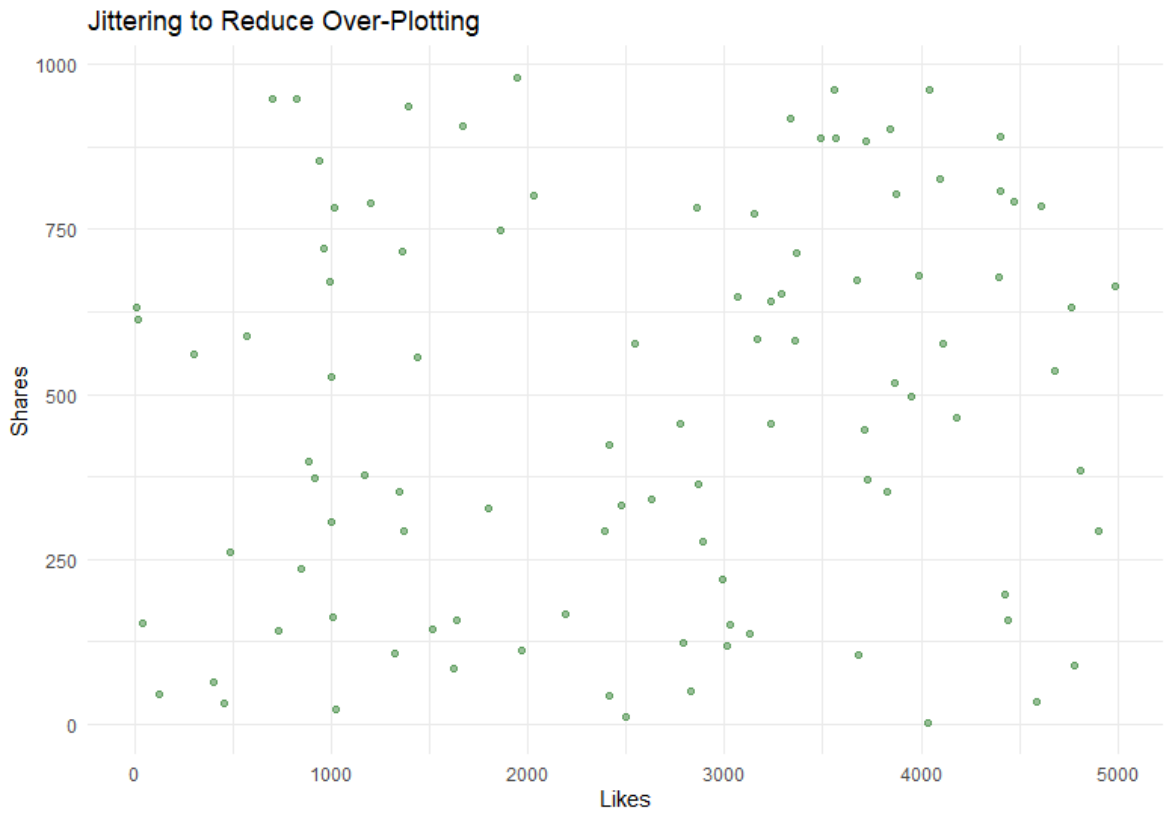
IN-LAB TASKS (Using R Language) • Apply alpha blending • Implement jittering techniques • Use aggregation and binning

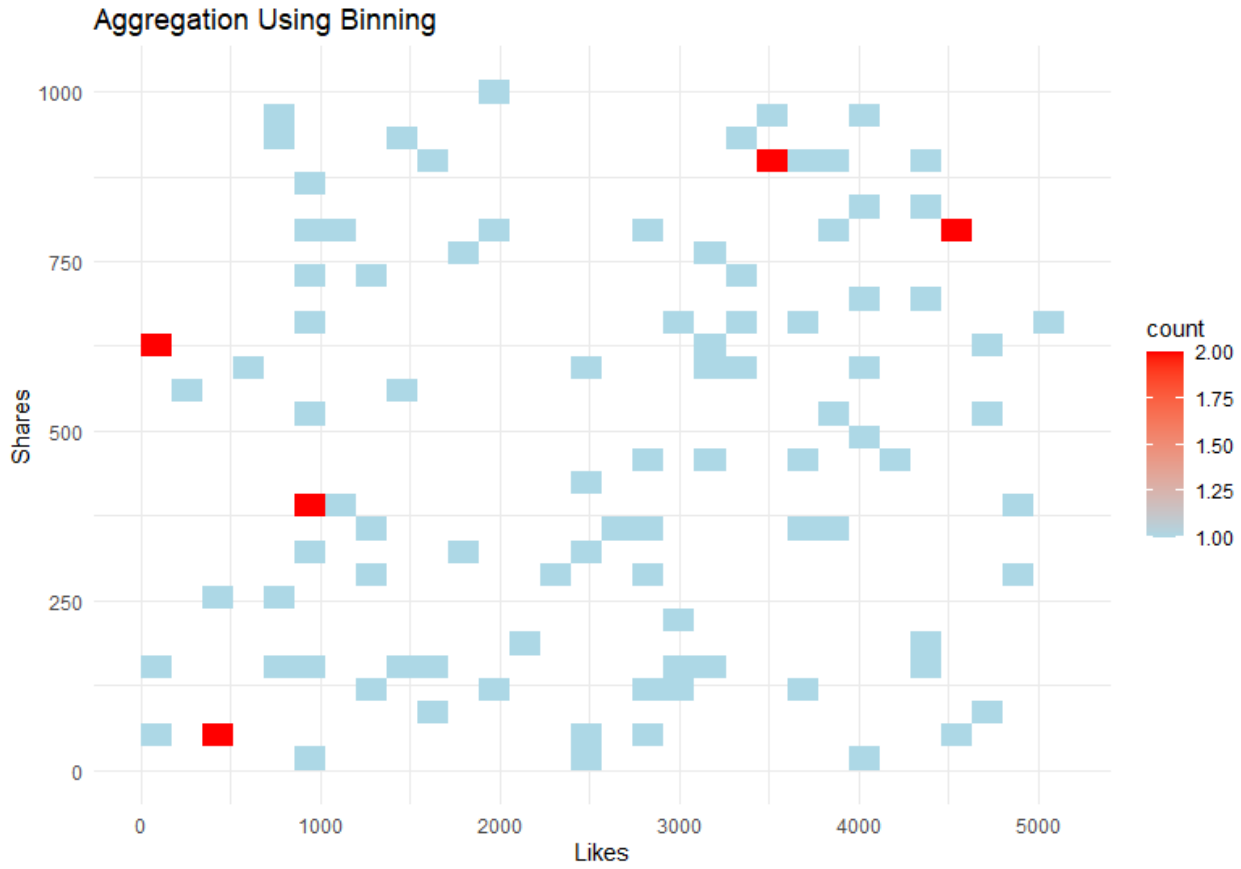


OUTPUT:


Normal Scatter Plot (Over-Plotting)
Shares
Likes
0
250
500
750
1000
0
1000
2000
3000
4000
5000


Alpha Blending: Likes vs Comments
Number of Comments
Number of Likes
0
200
400
600
800
0
1000
2000
3000
4000
5000


Jittering to Reduce Over-Plotting
Shares
Likes
0
250
500
750
1000
0
1000
2000
3000
4000
5000


Aggregation Using Binning
Shares
Likes
count
2.00
1.75
1.50
1.25
1.00
0
250
500
750
1000
0
1000
2000
3000
4000
5000

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	