EXPT NO:3	Designing Multivariate Patterns
DATE: 12.01.2026	

PRE-LAB QUESTIONS (PROVIDE BRIEF ANSWERS TO THE FOLLOWING QUESTIONS)

1. **Why is multivariate analysis essential in real-world AI problems?**

   Multivariate analysis helps AI systems understand relationships among multiple variables simultaneously, leading to more accurate predictions, better pattern recognition, and informed decision-making in complex real-world scenarios.

2. **What challenges arise when visualizing high-dimensional data?**

   High-dimensional data is difficult to visualize due to overlapping points, clutter, loss of interpretability, and limitations of representing more than three dimensions in 2D or 3D plots.

3. **How does correlation analysis support feature selection?**

   Correlation analysis identifies strongly related or redundant features, allowing selection of the most relevant variables and reducing dimensionality, overfitting, and computational cost.

4. **What are ethical concerns in healthcare data visualization?**

   Ethical concerns include patient privacy breaches, data misinterpretation, biased representations, and misleading visuals that may affect medical decisions.

5. **Give examples of multivariate data in AI systems.**

   - Patient records (age, BMI, blood pressure, glucose)
   - Financial datasets (income, spending, credit score)
   - Sensor data in IoT systems
   - Image data (pixel intensity, color channels)
   - Autonomous vehicle data (speed, distance, sensor readings)

IN-LAB EXERCISE:
OBJECTIVE:
To discover relationships among multiple variables using multivariate visualization.
SCENARIO:
A hospital analytics team studies patient health records to identify relationships between age, BMI, glucose levels, and blood pressure for early disease prediction.
IN-LAB TASKS (Using R Language)

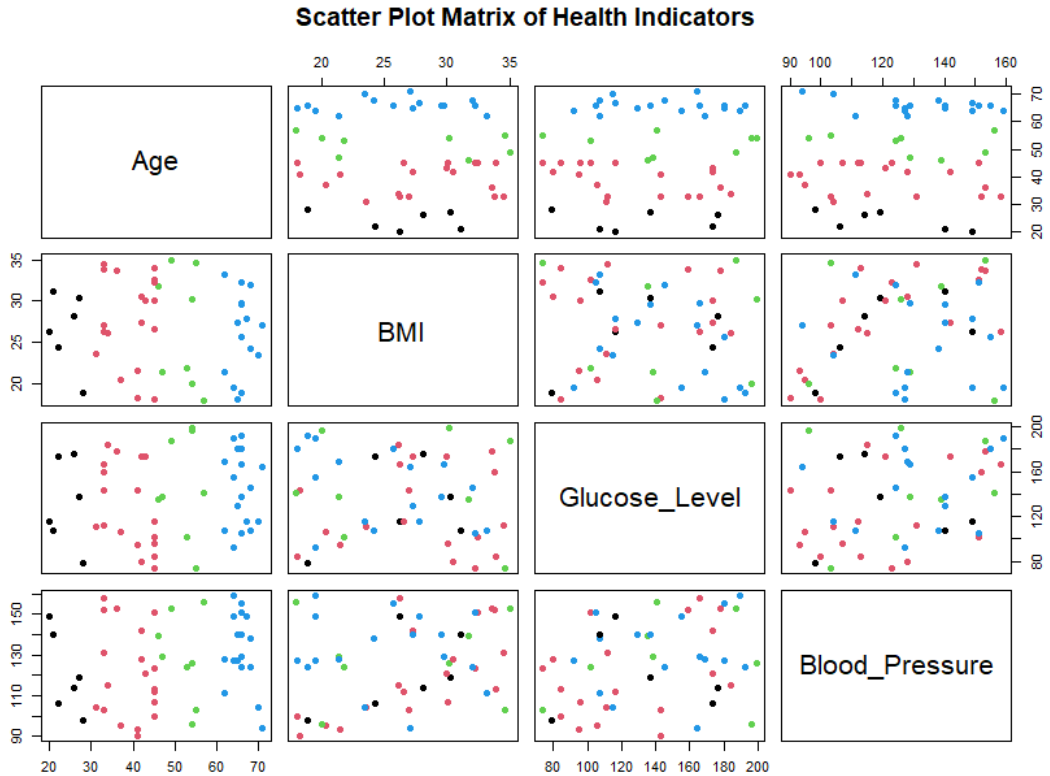
- Generate scatter plot matrix
- Apply color encoding for age groups
- Identify correlated health indicators

CODE:

```
# ==========================================
# EXPT NO: 3 - Designing Multivariate Patterns
# Roll No: 23BAD076
# ==========================================

# Load required library
library(ggplot2)

# Dataset already imported using Session -> Import Dataset
head(health_data)

# Create Age Groups
health_data$AgeGroup <- cut(
  health_data$Age,
  breaks = c(0, 30, 45, 60, 100),
  labels = c("Young", "Middle-aged", "Senior", "Elder")
)

# Scatter Plot Matrix
pairs(
  health_data[, c("Age", "BMI", "Glucose_Level", "Blood_Pressure")],
  col = as.numeric(health_data$AgeGroup),
  pch = 19,
  main = "Scatter Plot Matrix of Health Indicators"
)

# Correlation Analysis
cor_matrix <- cor(
  health_data[, c("Age", "BMI", "Glucose_Level", "Blood_Pressure")]
)

print(cor_matrix)
```

OUTPUT:

```
Console   Terminal ×   Background Jobs ×
R 4.4.1 · C:/Users/student/Downloads/exp3_dataset/
> # EXPT NO: 3 - Designing Multivariate Patterns
> # Roll No: 23BAD076
> # ========================================
>
> # Load required library
> library(ggplot2)
>
> # Dataset already imported using Session -> Import Dataset
> head(health_data)
  Patient_ID Age Gender  BMI Blood_Pressure Glucose_Level Cholesterol Disease_Risk    AgeGroup
1       3001  31   Male 23.6            104           111         255          Low Middle-aged
2       3002  45   Male 33.9            113            84         241         High Middle-aged
3       3003  65 Female 18.1            127           180         180         High       Elder
4       3004  53   Male 21.8            124           102         158         High      Senior
5       3005  68   Male 24.2            138           107         200         High       Elder
6       3006  33   Male 26.3            158           166         178       Medium Middle-aged
>
> # Create Age Groups
> health_data$AgeGroup <- cut(
+   health_data$Age,
+   breaks = c(0, 30, 45, 60, 100),
+   labels = c("Young", "Middle-aged", "Senior", "Elder")
+ )
>
> # Scatter Plot Matrix
> pairs(
+   health_data[, c("Age", "BMI", "Glucose_Level", "Blood_Pressure")],
+   col = as.numeric(health_data$AgeGroup),
+   pch = 19,
+   main = "Scatter Plot Matrix of Health Indicators"
+ )
>
> # Correlation Analysis
> cor_matrix <- cor(
+   health_data[, c("Age", "BMI", "Glucose_Level", "Blood_Pressure")]
+ )
>
> print(cor_matrix)
                     Age        BMI Glucose_Level Blood_Pressure
Age            1.0000000 -0.1538252     0.1146740      0.1672128
BMI           -0.1538252  1.0000000    -0.1210590      0.2485257
Glucose_Level  0.1146740 -0.1210590     1.0000000      0.2582461
Blood_Pressure 0.1672128  0.2485257     0.2582461      1.0000000
>
```



POST-LAB QUESTIONS (PROVIDE BRIEF ANSWERS TO THE FOLLOWING QUESTIONS)

1. **Which health parameters show strong correlation?**
   Health parameters such as Age and Blood Pressure and BMI and Glucose Level commonly show strong positive correlation in healthcare datasets.

2. **Why correlation does not imply causation in medical data?**
   Correlation only indicates a relationship between variables, not a cause-and-effect link, as confounding factors, lifestyle, or genetics may influence medical outcomes.

3. **How can these patterns assist predictive healthcare AI?**
   Identified patterns help AI models predict disease risks, detect early abnormalities, and improve treatment recommendations by learning relationships among health indicators.

4. **What visualization limitations exist for high-dimensional data?**
   High-dimensional data suffers from visual clutter, overlapping points, information loss, and difficulty in interpreting complex relationships in 2D or 3D plots.

5. **How can dimensionality reduction improve visualization?**
   Dimensionality reduction techniques (like PCA) compress data into fewer meaningful dimensions, making patterns easier to visualize while preserving important information.

ASSESSMENT

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