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?
2. What challenges arise when visualizing high-dimensional data?
3. How does correlation analysis support feature selection?
4. What are ethical concerns in healthcare data visualization?
5. Give examples of multivariate data in AI systems.

ANSWERS:

1. Multivariate analysis is essential in real-world AI problems because it analyzes multiple variables together, helping models understand complex relationships and improve prediction accuracy.
2. Visualizing high-dimensional data is challenging because it causes cluttered plots, overlapping information, and difficulty in interpreting patterns beyond three dimensions.
3. Correlation analysis supports feature selection by identifying redundant features, reducing dimensionality, and improving model efficiency and performance.
4. Ethical concerns in healthcare data visualization include patient privacy, data security, misrepresentation of data, bias, and potential misuse of sensitive information.
5. Examples of multivariate data in AI systems include healthcare records, customer behavior data, financial transaction data, and sensor data from autonomous systems.

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.

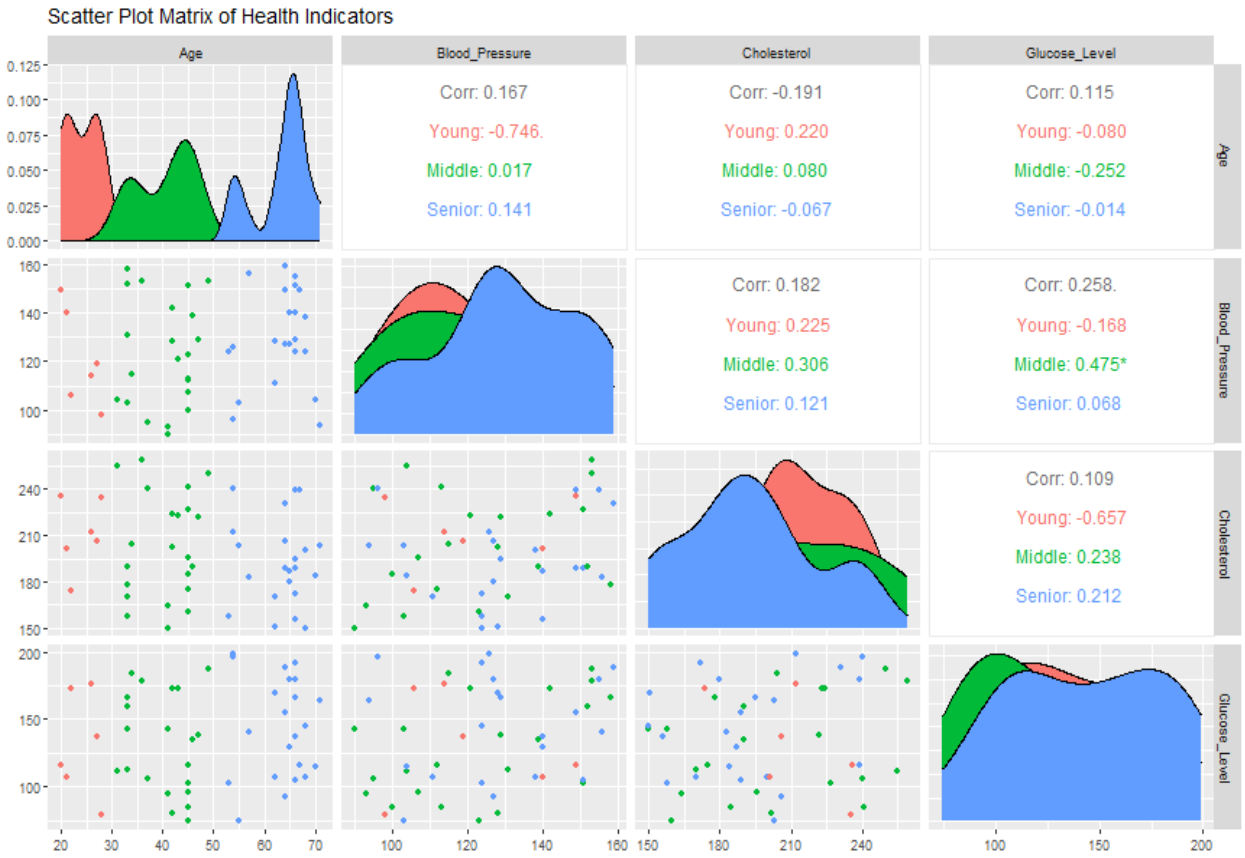
IN-LAB TASKS (Using R Language)

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

CODE:

```r
# ------------------------------------------------------
# Student Roll No: 23BAD095
# Subject: Data Analytics Lab
# Experiment: Scatter Plot Matrix & Correlation Analysis
# ------------------------------------------------------

# 1. Install required packages (run once)
install.packages(c("GGally", "dplyr"))

# 2. Load libraries
library(GGally)
library(dplyr)

# 3. Load healthcare dataset
health <- read.csv("C:/Users/student.DESKTOP-D0TBOQD/Downloads/3.healthcare_data.csv")

# 4. Check dataset structure
str(health)
head(health)

# 5. Create Age Groups for color encoding
health <- health %>%
  mutate(Age_Group = cut(
    Age,
    breaks = c(0, 30, 50, 100),
    labels = c("Young", "Middle", "Senior")
  ))

# 6. Scatter Plot Matrix with correct column names
ggpairs(
  health,
  columns = c("Age", "Blood_Pressure", "Cholesterol", "Glucose_Level"),
  aes(color = Age_Group),
  title = "Scatter Plot Matrix of Health Indicators"
)

# 7. Select numeric variables for correlation
health_numeric <- health %>%
  select(Age, Blood_Pressure, Cholesterol, Glucose_Level)

# 8. Correlation matrix
correlation_matrix <- cor(health_numeric)

# 9. Display correlation matrix
print(correlation_matrix)
```

OUTPUT:



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

1. Which health parameters show strong correlation?
2. Why correlation does not imply causation in medical data?
3. How can these patterns assist predictive healthcare AI?
4. What visualization limitations exist for high-dimensional data?
5. How can dimensionality reduction improve visualization?

ANSWERS:

1. No health parameters show a strong correlation. The highest observed correlation is between blood pressure and glucose level (≈ 0.26), which indicates only a weak to moderate positive relationship, while all other correlations are weak.
2. Correlation does not imply causation because health variables may be related due to confounding factors, coincidence, or indirect effects. Medical outcomes are influenced by multiple biological, environmental, and lifestyle factors, so a correlation alone cannot establish a cause–effect relationship.
3. These patterns assist predictive healthcare AI by helping identify relevant features, understand relationships among health indicators, and improve risk prediction models through better feature selection.
4. High-dimensional data visualization becomes complex and cluttered, making patterns difficult to interpret. Scatter plot matrices grow large, overlap increases, and important relationships may be overlooked.
5. Dimensionality reduction improves visualization by compressing multiple variables into fewer meaningful components, simplifying plots, revealing hidden patterns, and making high-dimensional data easier to interpret and analyze.

ASSESSMENT

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