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)

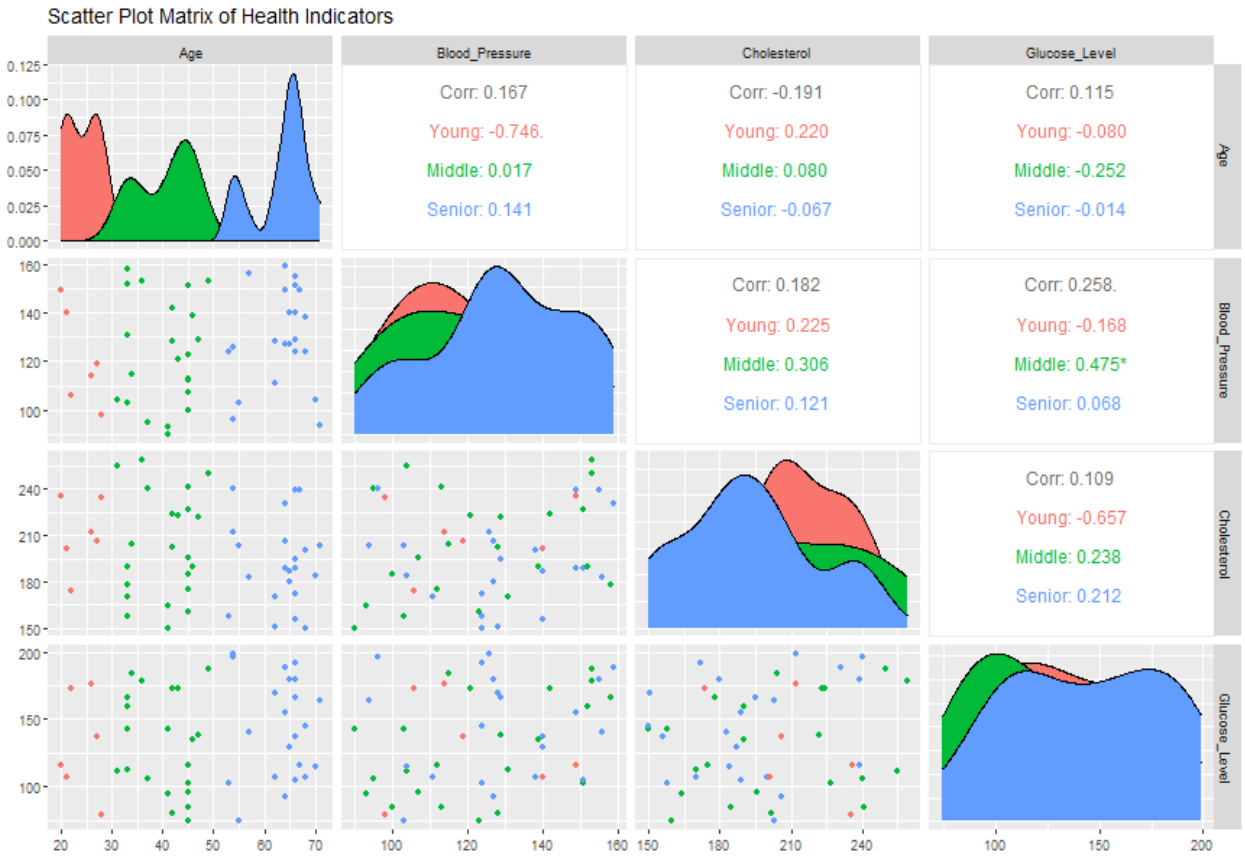
- Generate scatter plot matrix
- Apply color encoding for age groups
- Identify correlated health indicators

CODE:

```r

"Student name: Sai Vaishnavi R"
"roll no:23BAD094"
# 2. Load libraries
library(GGally)
library(dplyr)

# 3. Load healthcare dataset
health <- read.csv("C:/Users/student/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
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, use = "complete.obs")

# 9. Display correlation matrix
print(correlation_matrix)

```

OUTPUT:

```
> "Student name: Sai Vaishnavi R"
[1] "Student name: Sai Vaishnavi R"
> "roll no:23BAD094"
[1] "roll no:23BAD094"
> # 2. Load libraries
> library(GGally)
> library(dplyr)
>
> # 3. Load healthcare dataset
> health <- read.csv("C:/Users/student/Downloads/3.healthcare_data.csv")
>
> # 4. Check dataset structure
> str(health)
'data.frame':   50 obs. of  8 variables:
 $ Patient_ID    : int  3001 3002 3003 3004 3005 3006 3007 3008 3009 3010 ...
 $ Age           : int  31 45 65 53 68 33 45 64 46 28 ...
 $ Gender        : chr  "Male" "Male" "Female" "Male" ...
 $ BMI           : num  23.6 33.9 18.1 21.8 24.2 26.3 32.5 19.5 31.7 18.9 ...
 $ Blood_Pressure: int  104 113 127 124 138 158 151 149 139 98 ...
 $ Glucose_Level : int  111 84 180 102 107 166 102 155 135 79 ...
 $ Cholesterol   : int  255 241 180 158 200 178 227 189 190 235 ...
 $ Disease_Risk  : chr  "Low" "High" "High" "High" ...
> head(health)
  Patient_ID Age Gender  BMI Blood_Pressure Glucose_Level Cholesterol Disease_Risk
1       3001  31   Male 23.6            104           111         255          Low
2       3002  45   Male 33.9            113            84         241         High
3       3003  65 Female 18.1            127           180         180         High
4       3004  53   Male 21.8            124           102         158         High
5       3005  68   Male 24.2            138           107         200         High
6       3006  33   Male 26.3            158           166         178       Medium
>
> # 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
> 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, use = "complete.obs")
>
> # 9. Display correlation matrix
> print(correlation_matrix)
                    Age Blood_Pressure Cholesterol Glucose_Level
Age             1.0000000      0.1672128  -0.1911136     0.1146740
Blood_Pressure  0.1672128      1.0000000   0.1819998     0.2582461
Cholesterol    -0.1911136      0.1819998   1.0000000     0.1088433
Glucose_Level   0.1146740      0.2582461   0.1088433     1.0000000
>
```


Scatter Plot Matrix of Health Indicators
Age
Blood_Pressure
Cholesterol
Glucose_Level
Corr: 0.167
Young: -0.746.
Middle: 0.017
Senior: 0.141
Corr: -0.191
Young: 0.220
Middle: 0.080
Senior: -0.067
Corr: 0.115
Young: -0.080
Middle: -0.252
Senior: -0.014
Corr: 0.182
Young: 0.225
Middle: 0.306
Senior: 0.121
Corr: 0.258.
Young: -0.168
Middle: 0.475*
Senior: 0.068
Corr: 0.109
Young: -0.657
Middle: 0.238
Senior: 0.212
0.125
0.100
0.075
0.050
0.025
0.000
160
140
120
100
240
210
180
150
200
150
100
20
30
40
50
60
70
100
120
140
160
150
180
210
240
100
150
200

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	