****

**CHANDIGARH UNIVERSITY**

**DEPARTMENT: UIC**

Bachelors of Computer Application

Subject Name: Communication Skills

24CAP-161

**PROJECT**

**Analyzing and Visualizing Hospital Patient Data Using R**

Submitted by: **Harmanpreet Kaur**

**(24BCD10073)**

Submitted to: **Ambika Mam**

**Project Title:**

**Analyzing and Visualizing Hospital Patient Data Using R**

**Objective**

To analyze patient health and recovery data from a hospital using R. The aim is to understand patterns in vital signs, recovery duration, and demographic influences, thereby generating actionable healthcare insights through data visualization and statistics.

**Technologies Used**

* **Language**: R Programming
* **IDE**: RStudio
* **Libraries**:
  + ggplot2 – for data visualization
  + dplyr – for data manipulation
  + corrplot – for correlation analysis
* **Data Format**: CSV / In-memory data frame

**Methodology**

1. **Dataset Creation**

hospital\_data <- data.frame(

PatientID = 1:10,

Age = c(34, 57, 45, 60, 28, 70, 39, 52, 66, 48),

Gender = c("Female", "Male", "Female", "Male", "Female", "Male", "Female", "Male", "Female", "Male"),

BloodPressure = c(120, 145, 130, 155, 118, 160, 132, 140, 135, 150),

HeartRate = c(80, 95, 85, 102, 78, 98, 86, 90, 84, 100),

Temperature = c(98.6, 99.1, 98.4, 99.5, 98.3, 100.2, 98.9, 98.7, 99.0, 100.1),

Diagnosis = c("Fever", "Hypertension", "Infection", "Hypertension", "Fever", "Hypertension", "Infection", "Infection", "Fever", "Hypertension"),

AdmissionDate = as.Date(c("2024-01-02", "2024-01-05", "2024-01-07", "2024-01-10", "2024-01-12",

"2024-01-15", "2024-01-17", "2024-01-20", "2024-01-22", "2024-01-25")),

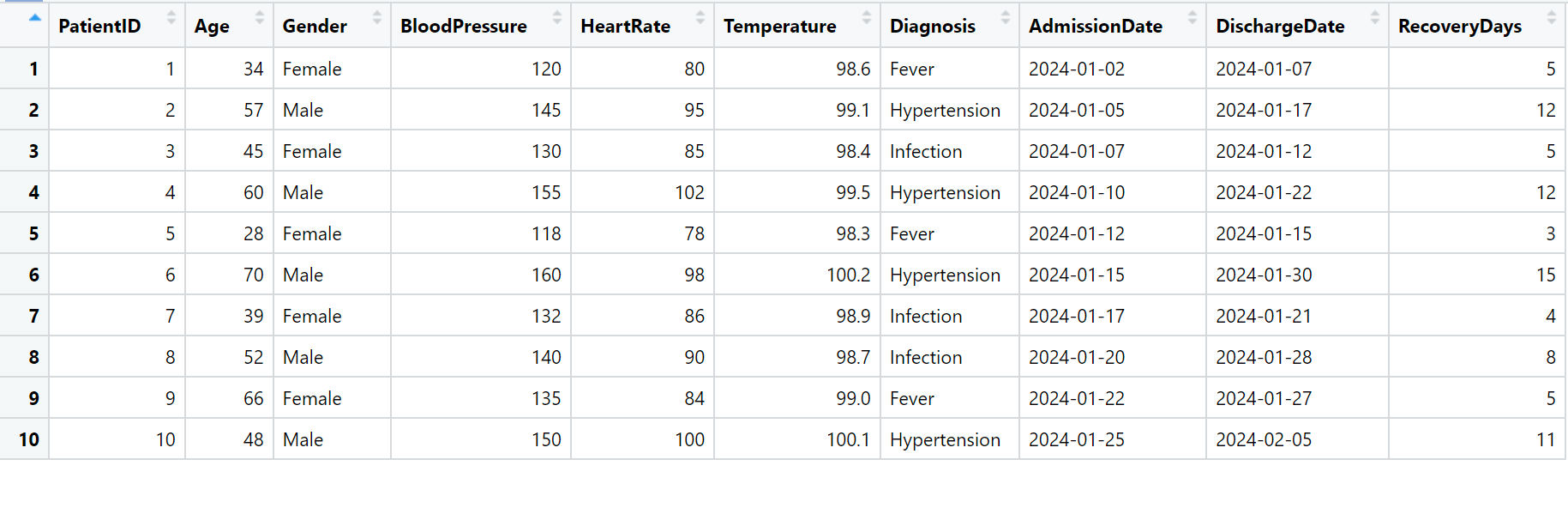
DischargeDate = as.Date(c("2024-01-07", "2024-01-17", "2024-01-12", "2024-01-22", "2024-01-15",

"2024-01-30", "2024-01-21", "2024-01-28", "2024-01-27", "2024-02-05"))

)

hospital\_data$RecoveryDays <- as.numeric(hospital\_data$DischargeDate - hospital\_data$AdmissionDate)

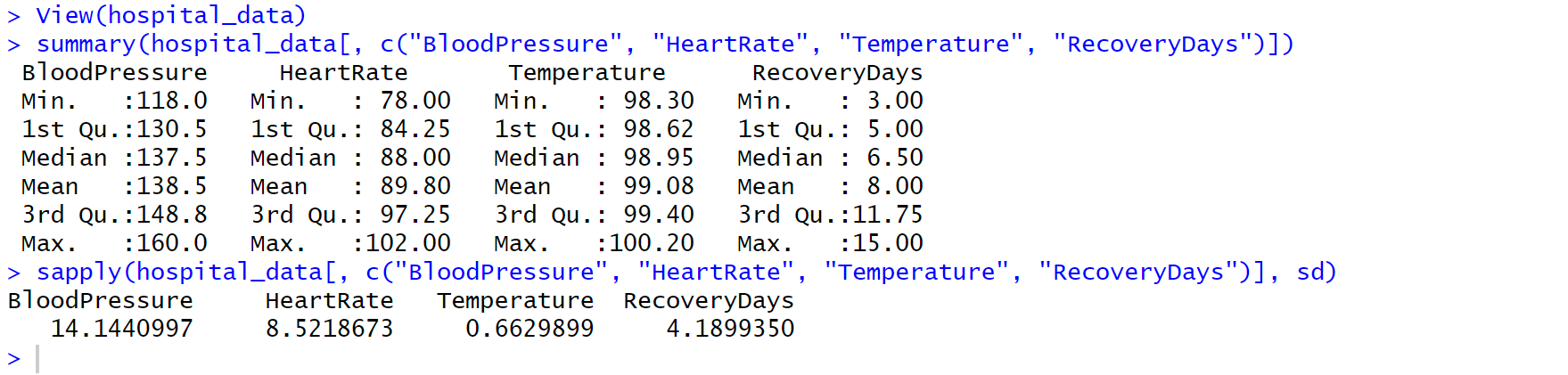
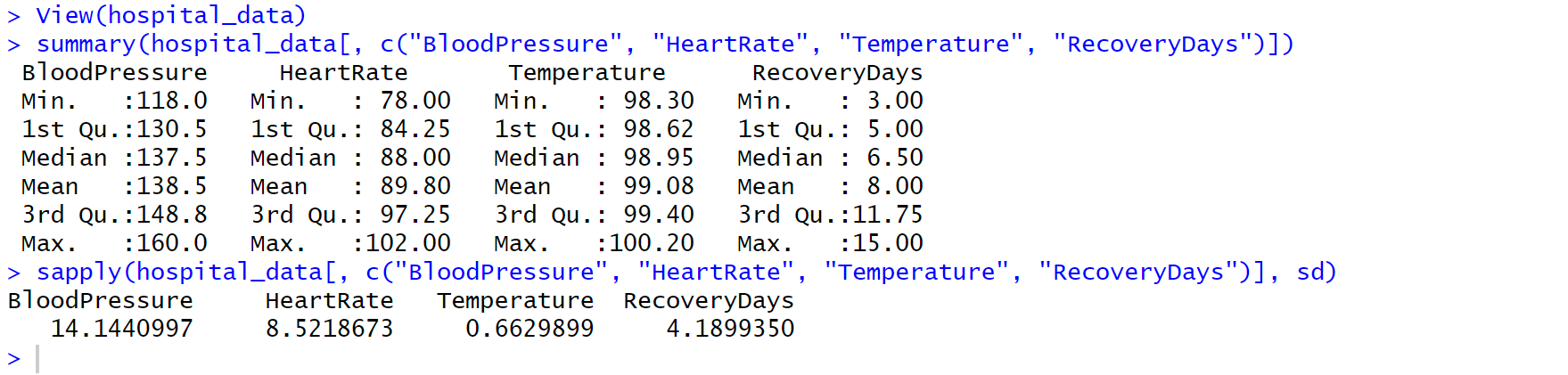
* **OUTPUT**



1. **Statistical Summary**

summary(hospital\_data[, c("BloodPressure", "HeartRate", "Temperature", "RecoveryDays")])

sapply(hospital\_data[, c("BloodPressure", "HeartRate", "Temperature", "RecoveryDays")], sd)

* **OUTPUT**

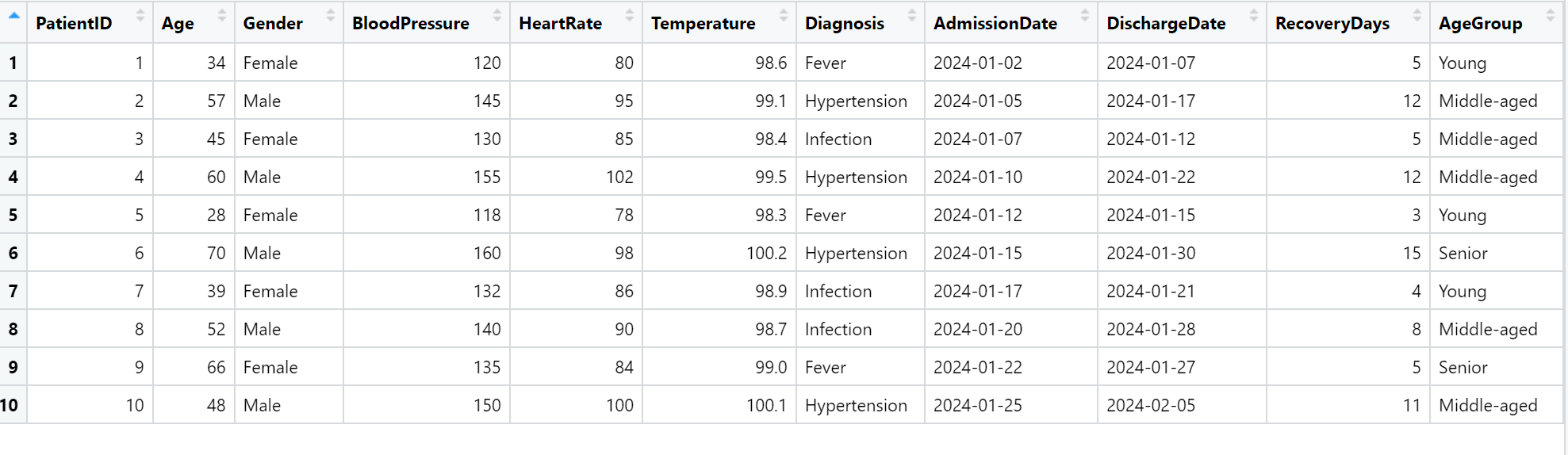
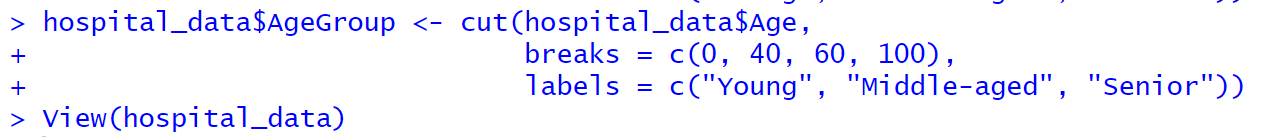
1. **Age Grouping**

hospital\_data$AgeGroup <- cut(hospital\_data$Age,

breaks = c(0, 40, 60, 100),

labels = c("Young", "Middle-aged", "Senior"))

* **OUTPUT**



1. **Correlation Analysis**

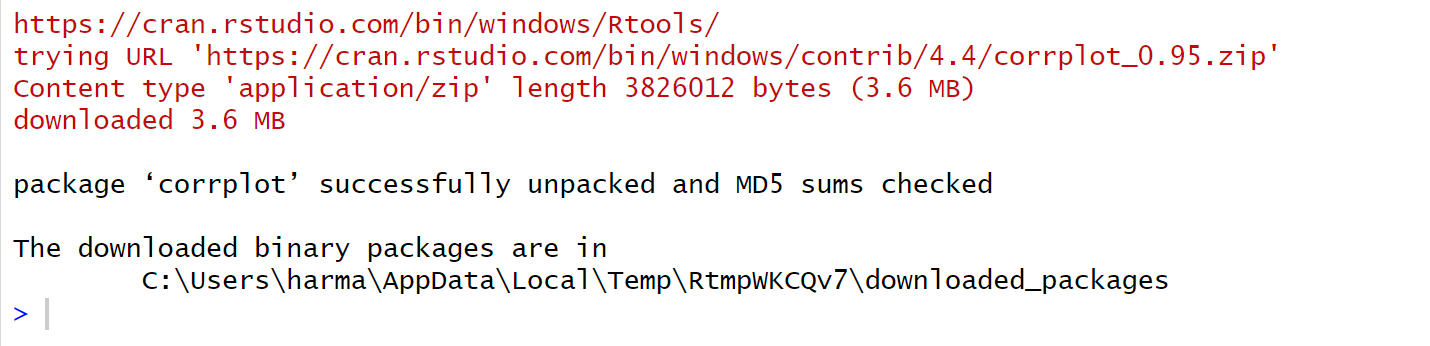
Install.packages(“corrplot”)

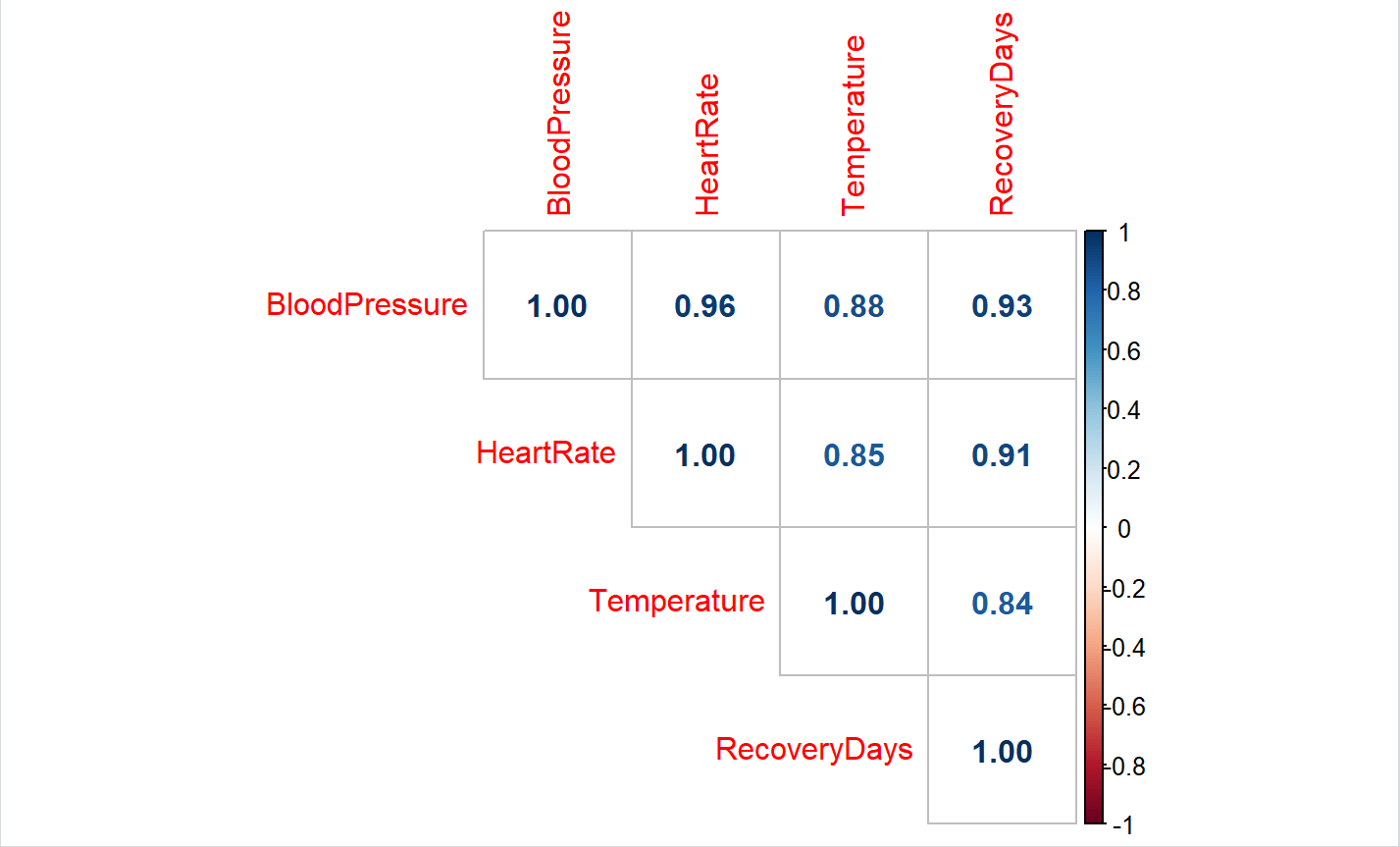
library(corrplot)

cor\_matrix <- cor(hospital\_data[, c("BloodPressure", "HeartRate", "Temperature", "RecoveryDays")])

corrplot(cor\_matrix, method = "number", type = "upper")

* **OUTPUT**





**Visualizations**

* **Recovery Days by Diagnosis**

install.packages("ggplot2")

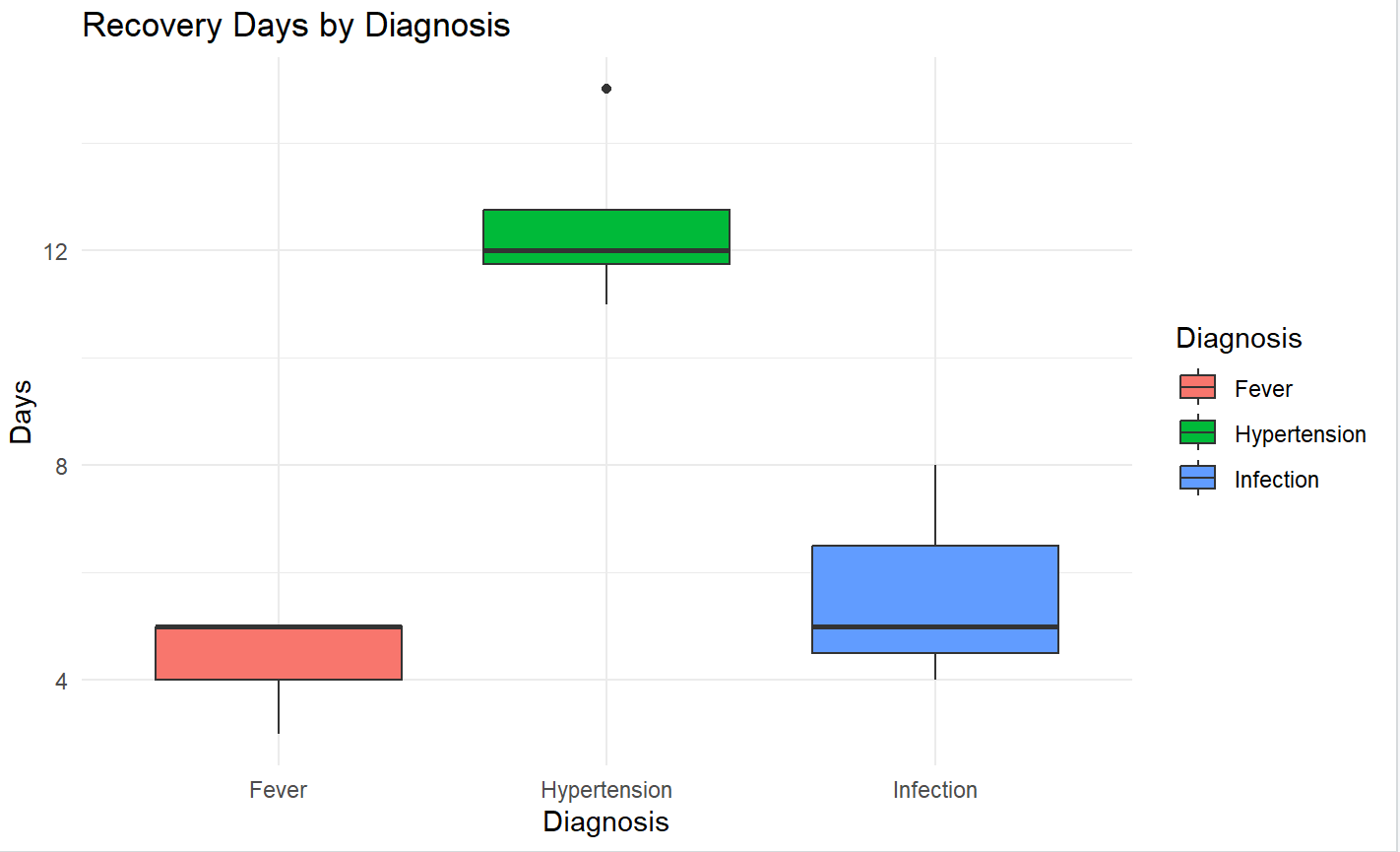
library(ggplot2)

ggplot(hospital\_data, aes(x = Diagnosis, y = RecoveryDays, fill = Diagnosis)) +

geom\_boxplot() +

theme\_minimal() +

labs(title = "Recovery Days by Diagnosis", x = "Diagnosis", y = "Days")

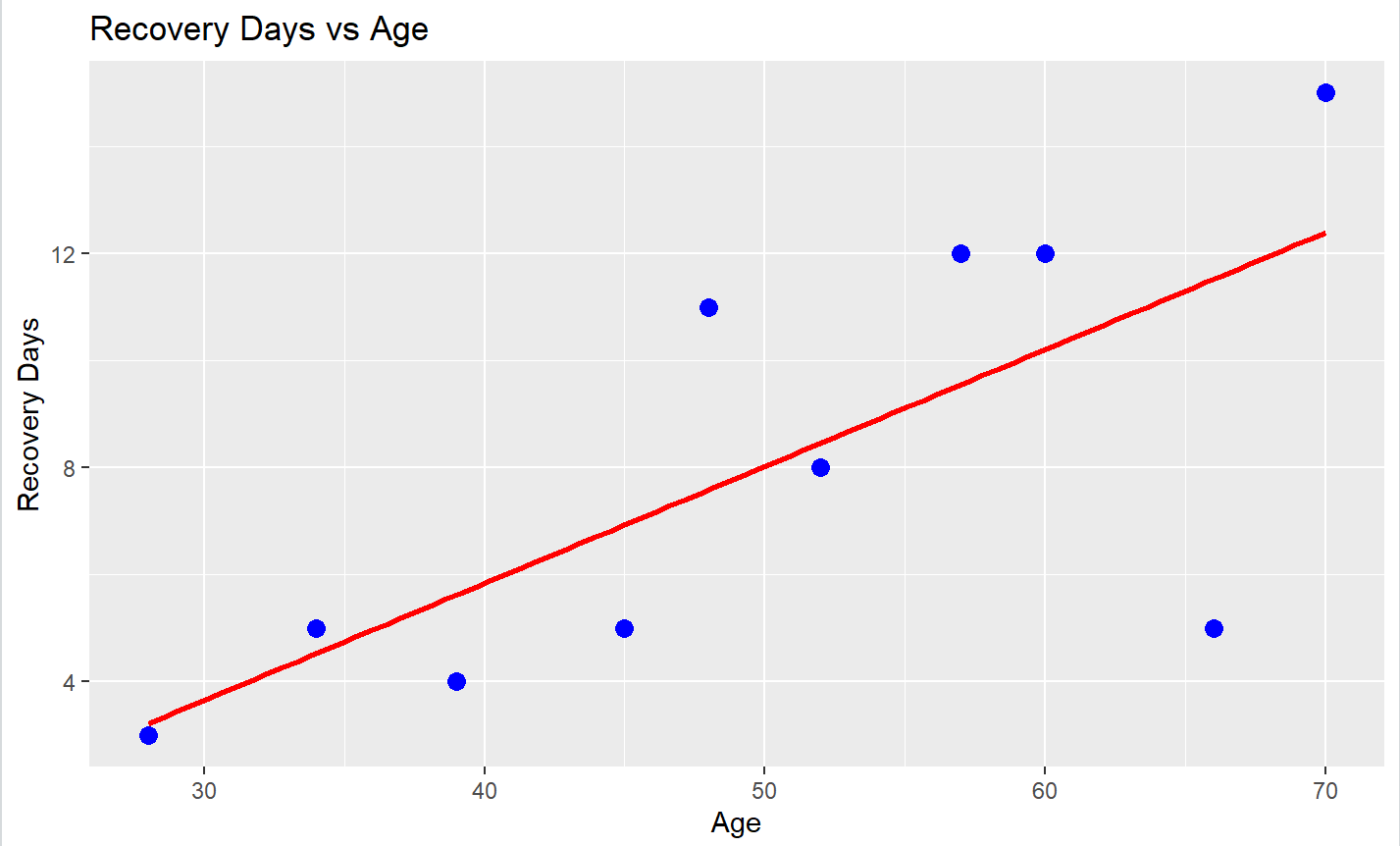
* **OUTPUT**
* **Age vs Recovery (Regression)**

ggplot(hospital\_data, aes(x = Age, y = RecoveryDays)) +

geom\_point(color = "blue", size = 3) +

geom\_smooth(method = "lm", se = FALSE, color = "red") +

labs(title = "Recovery Days vs Age", x = "Age", y = "Recovery Days")

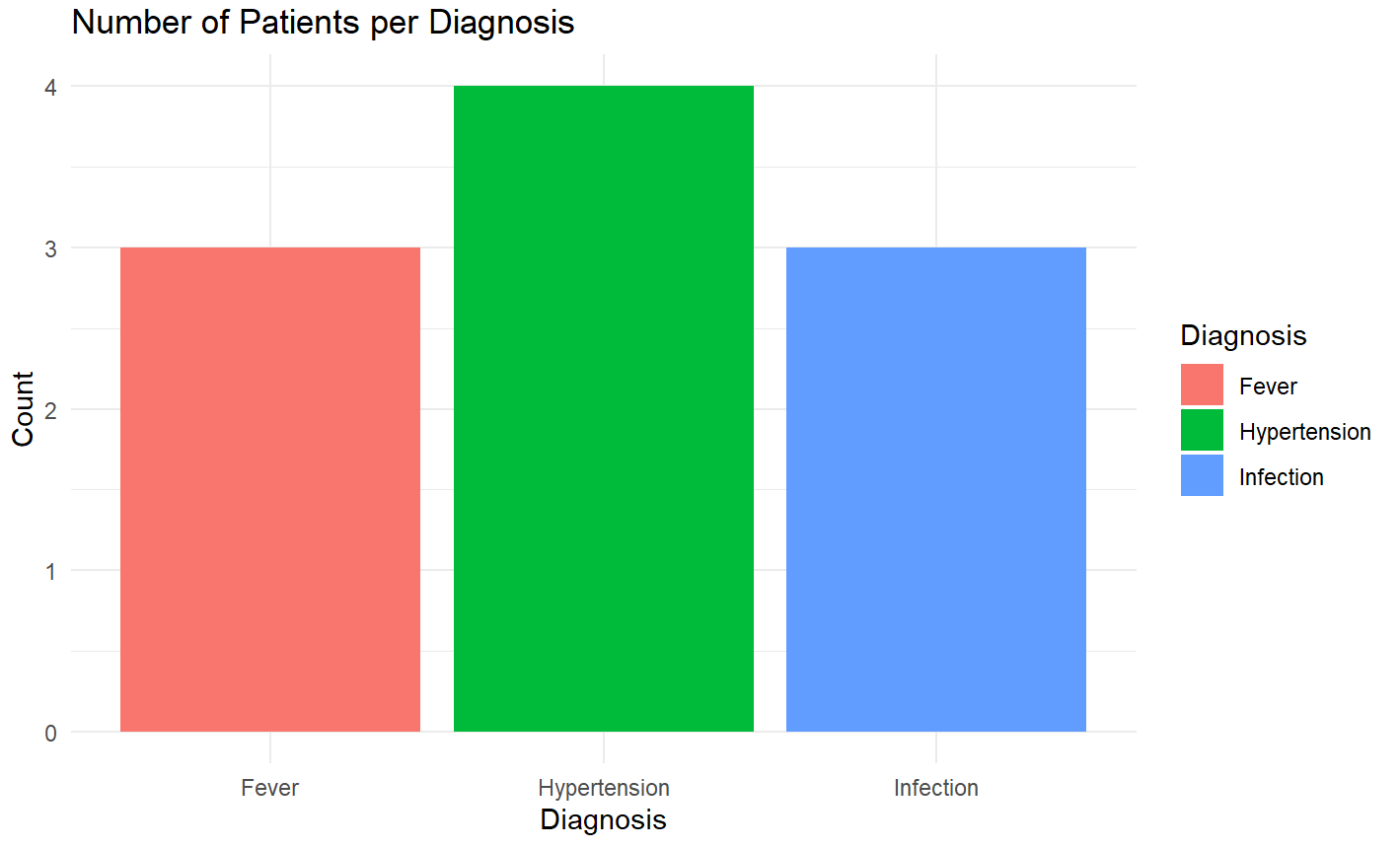
* **OUTPUT**
* **Number of Patients per Diagnosis**

ggplot(hospital\_data, aes(x = Diagnosis, fill = Diagnosis)) +

geom\_bar() +

theme\_minimal() +

labs(title = "Number of Patients per Diagnosis", x = "Diagnosis", y = "Count")

* **OUTPUT**

**Expected Output**

* **Summary Statistics** table
* **Correlation Matrix** between vitals and recovery
* **Boxplot** comparison by diagnosis
* **Scatterplot** showing regression between age and recovery
* **Categorized patients** by age group

**Insights**

* **Positive correlation** between age and recovery duration.
* **Hypertension patients** had longer recovery times than others.
* Higher **heart rate and blood pressure** were mildly associated with prolonged stays.
* Younger patients with **fever** showed fastest recovery.

**Learning Outcomes**

* Learned how to simulate and preprocess healthcare data in R.
* Performed descriptive and inferential statistical analysis.
* Used **data segmentation** and **grouped comparisons** to find patterns.
* Created **professional plots** for healthcare data analysis.
* Applied **correlation and regression** to interpret relationships in real-world datasets.

**References**

* R Documentation: <https://www.r-project.org/>
* Hadley Wickham’s *Tidyverse* package: <https://www.tidyverse.org/>
* ggplot2 Reference: https://ggplot2.tidyverse.org/
* Healthcare Dataset Inspiration: <https://www.kaggle.com/>