GITHUB PORTFOLIO

VITAL SIGNS DIAGNOSIS DATA

E REPORT

NAME OF PROJECT:
GITHUB PORTFOLIO

PRESENTED BY:
Morales, Enrico James E. &
Pinzon Jazreel S.

MAPUA UNIVERSITY
BI120L_CON29

Introduction

This report analyzes the Vital Signs Diagnosis Data - Group 007, which includes health and lifestyle variables such as BMI, blood pressure, glucose and cholesterol levels, age, sex, smoking status, physical activity, and sleep duration.

The dataset was cleaned by converting data types, removing outliers using the IQR method, and filtering implausible BMI values. Descriptive statistics and visualizations were used to explore distributions and trends across sex.

Statistical tests, including correlation analysis and a t-test on glucose levels by sex, were performed to identify significant relationships.
The goal is to understand how lifestyle factors relate to vital signs and overall health risks.

METHODS USED FOR ANALYSIS

Data analysis was conducted using R, a statistical programming language widely used for data science.
Several R packages were utilized to support the workflow:

- tidyverse – for data manipulation and visualization (read.csv(), filter(), select(), ggplot())
- psych – for descriptive statistics (describe())
- GGally – for advanced plotting (though not directly used in this script)
- corrplot – for visualizing correlation matrices (corrplot())

Key steps included:

- Data cleaning: Conversion of text to numeric (gsub(), as.numeric()), factor labeling (factor()), and removal of outliers via a custom remove_outliers() function using the IQR method.
- Descriptive statistics: Generated using summary() and describe() functions.
- Data visualization: Histograms, box plots, bar plots, and scatter plots created using ggplot2.
- Statistical testing: A t-test (t.test()) was performed to compare glucose levels by sex.
- Correlation analysis: A correlation matrix was computed using cor() and visualized with corrplot().

KEY RESULTS AND PLOTS

Key Results and Plots

Descriptive Statistics revealed:

- Average BMI: 26.82 (slightly overweight range)
- Average Glucose: 128.4 mg/dL
- Average Cholesterol: 189.84 mg/dL
- Participants were nearly evenly split by sex (Female: 488, Male: 502)
- Mean sleep duration was low (5.28 hours), with an average stress level of 5.32 (on a 10-point scale)

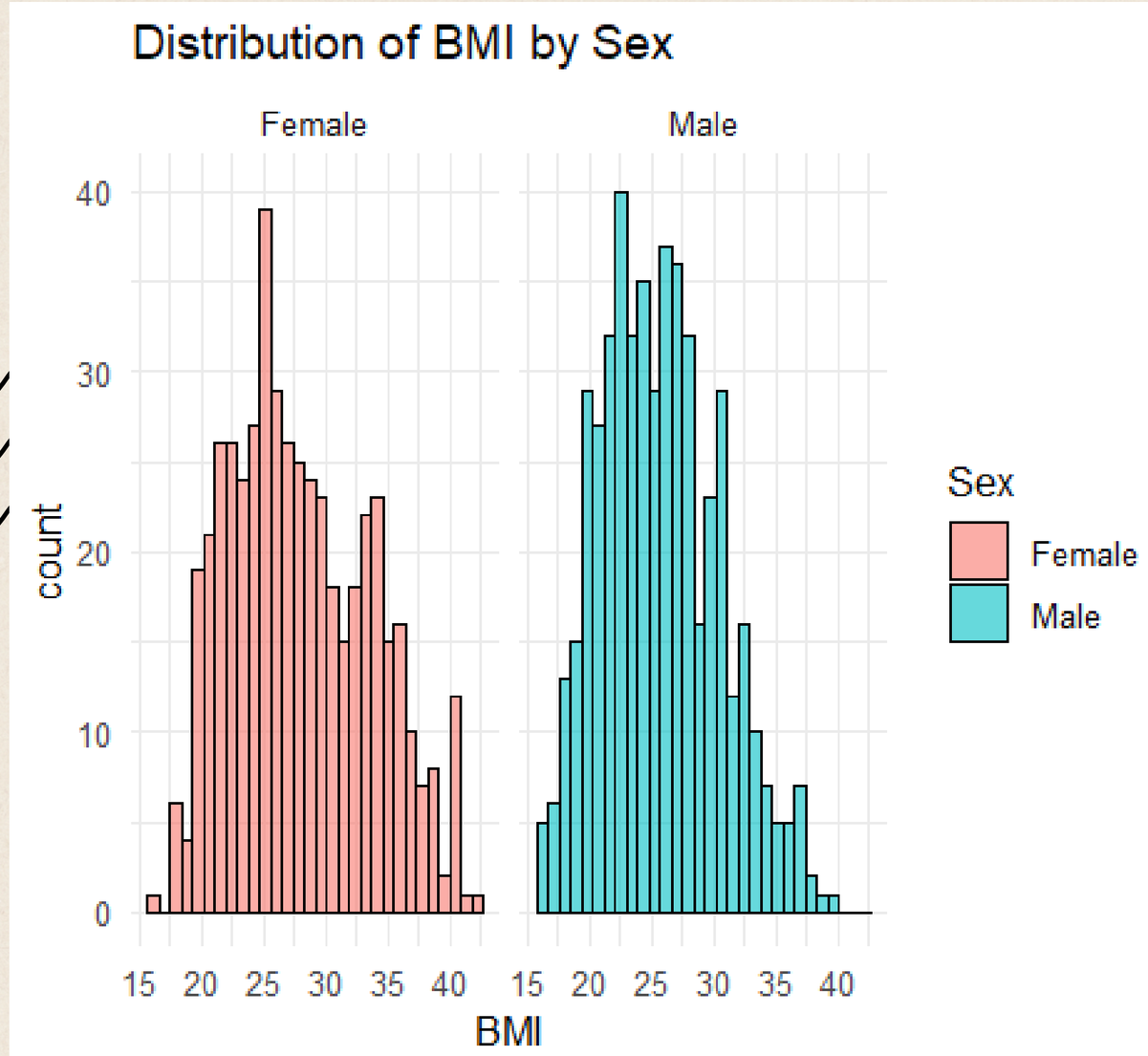
t-test showed a statistically significant difference in glucose levels between sexes:

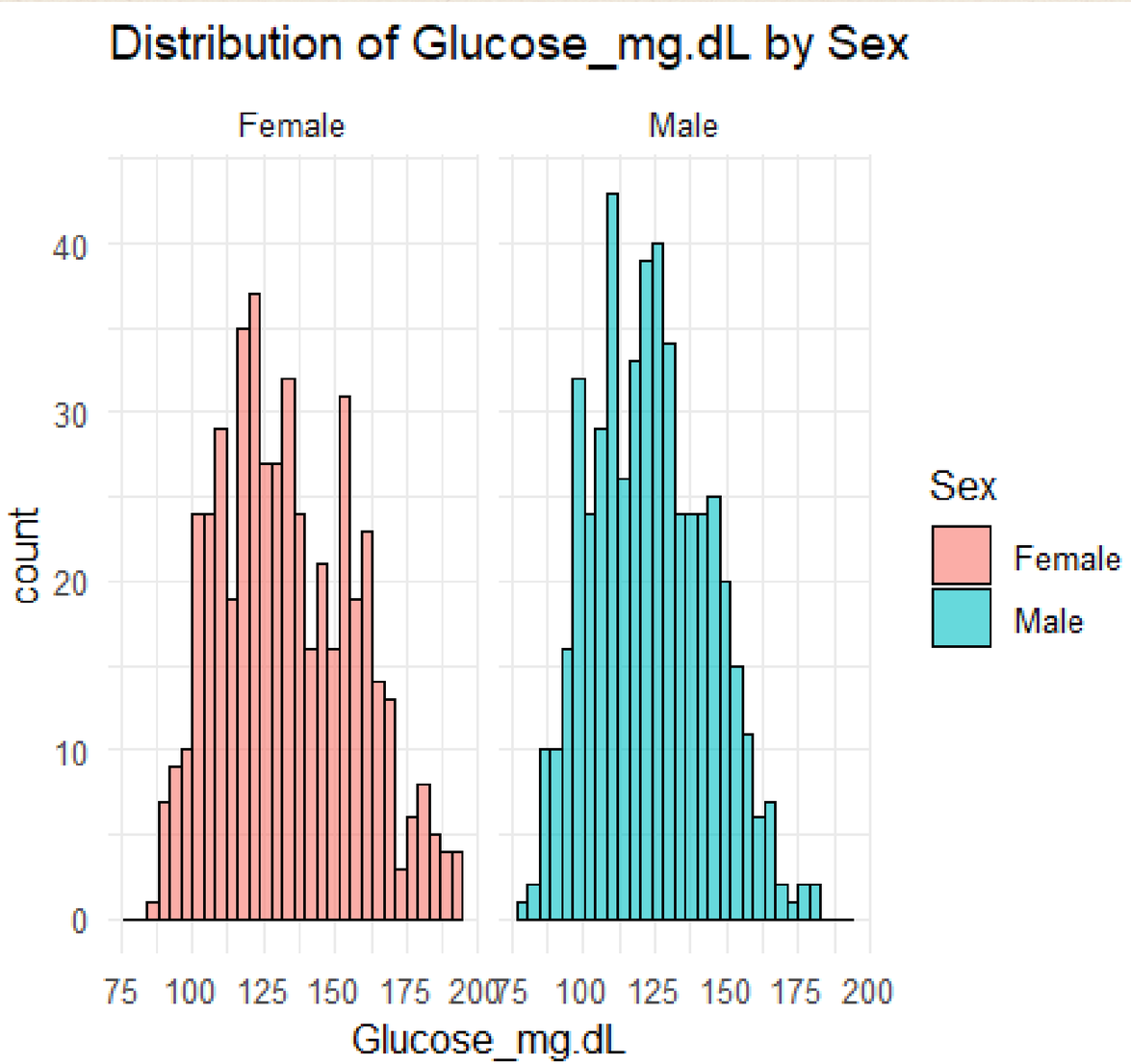
- Females had higher average glucose (133.61 mg/dL) than males (123.33 mg/dL),
- p-value = 3.921e-13, indicating a highly significant result.

Correlation analysis showed:

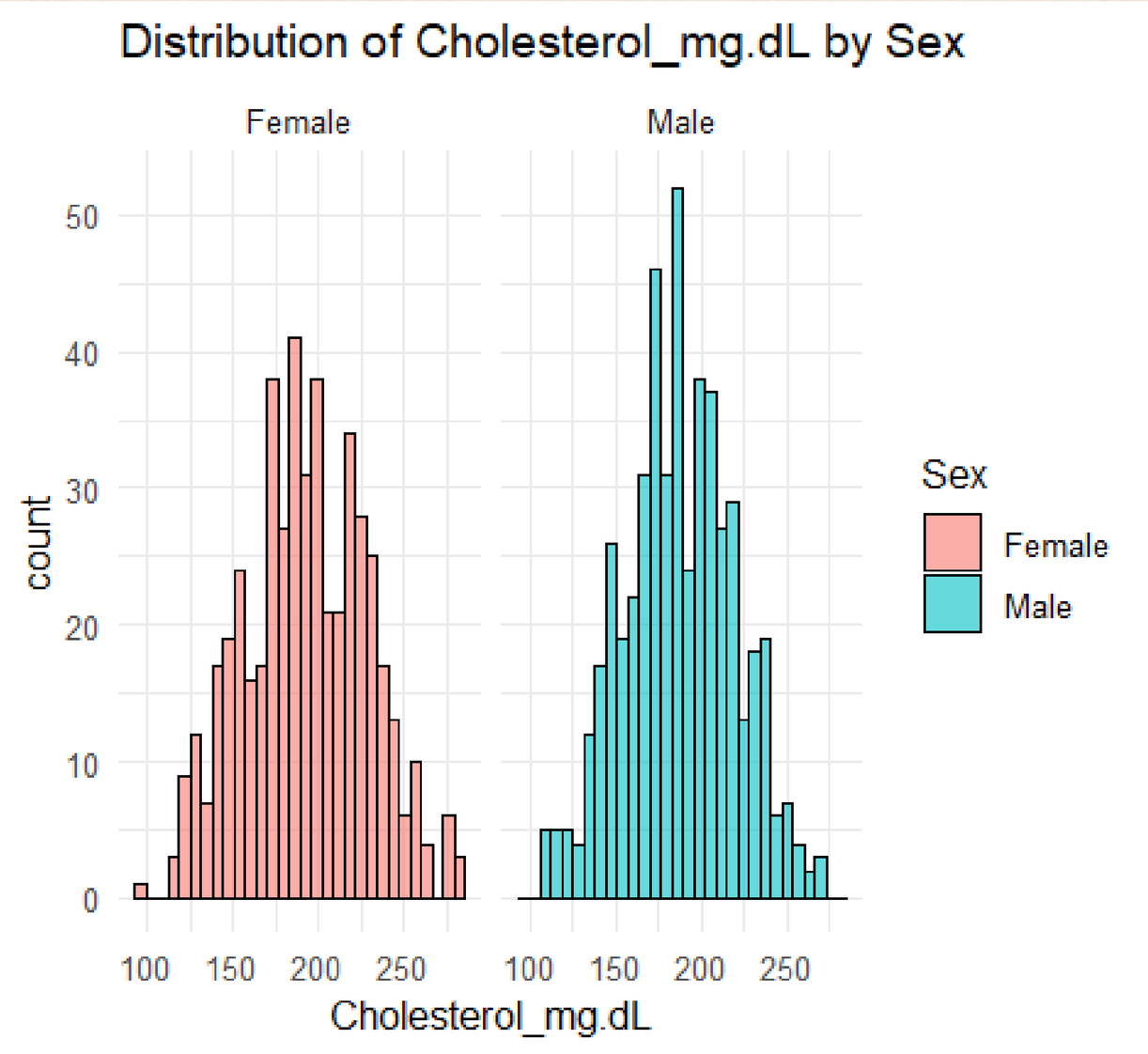
- BMI had a strong positive correlation with glucose (r = 0.97) and cholesterol (r = 0.69)
- Systolic and diastolic BP were also strongly correlated (r = 0.80)
- Heart rate was negatively correlated with physical activity (r = –0.59) and sleep duration (r = –0.47)

PLOTS


Distribution of BMI by Sex
Female
Male
count
BMI
Sex
Female
Male


Distribution of Glucose_mg.dL by Sex
Female
Male
count
Glucose_mg.dL
Sex
Female
Male

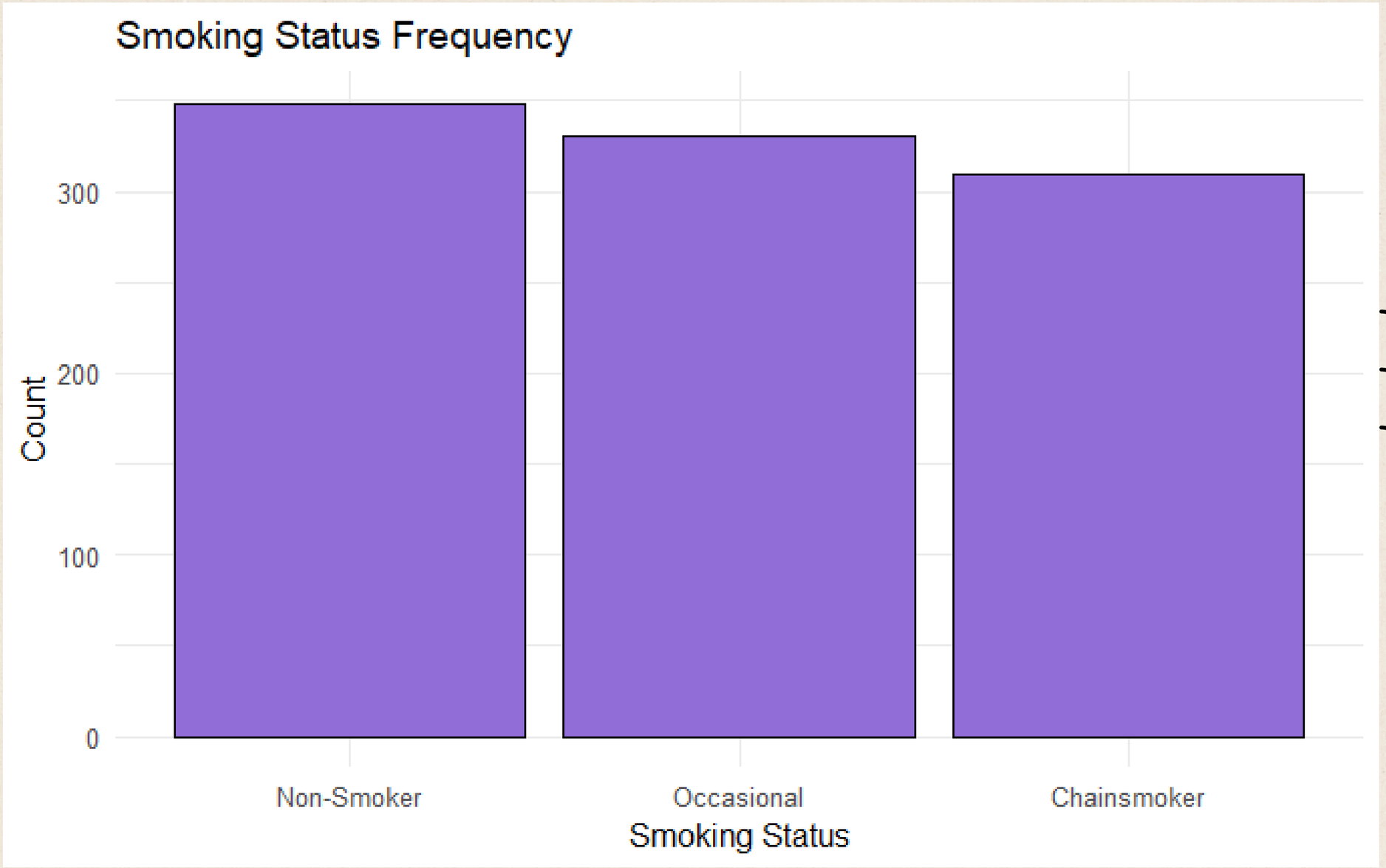
PLOTS



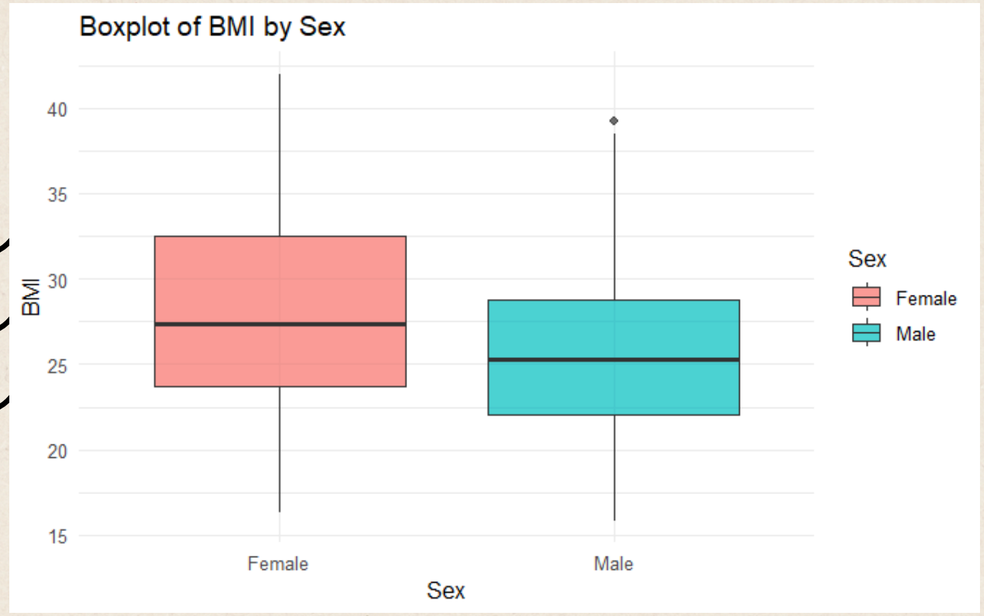
Histograms(Figure 1-3)
Glucose is in diabetic range (128.4), nomal range 70-99. BMI avg is in the overweight range. While the Cholesterol is in the Normal Range

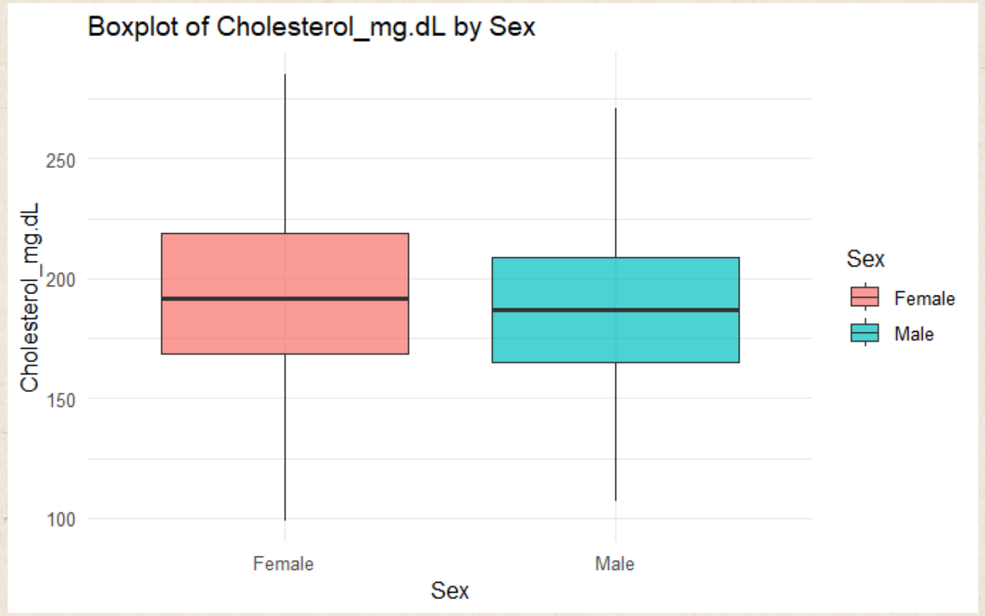
PLOTS

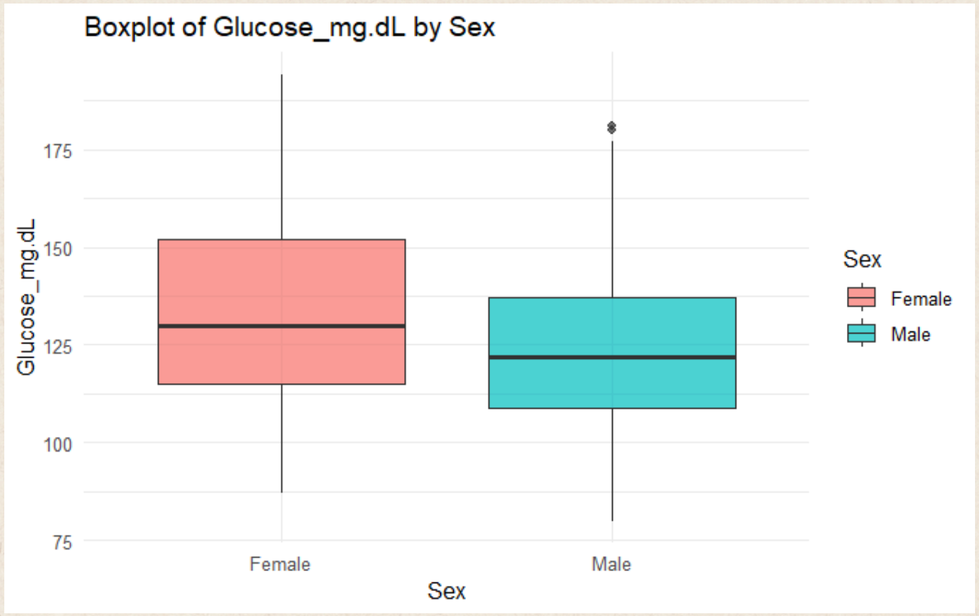
Bar graph(Figure 4) SMOKING STATUS



PLOTS



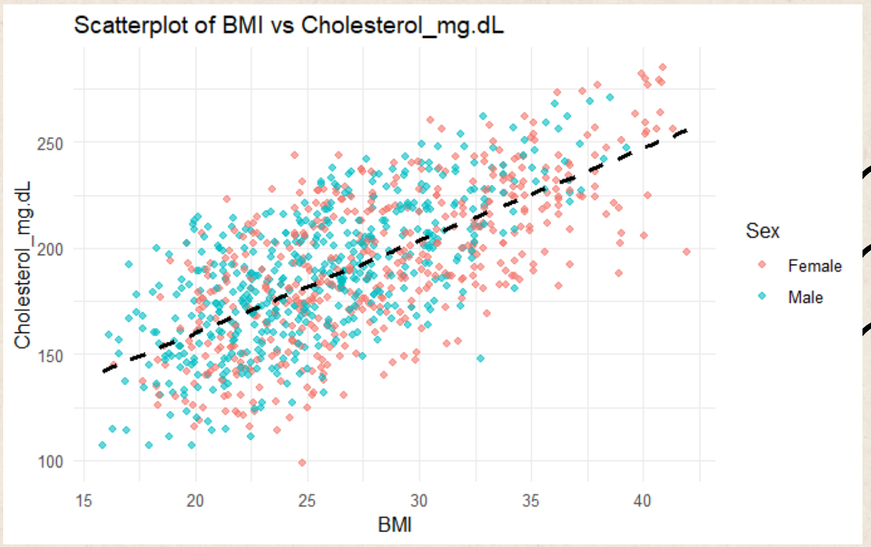


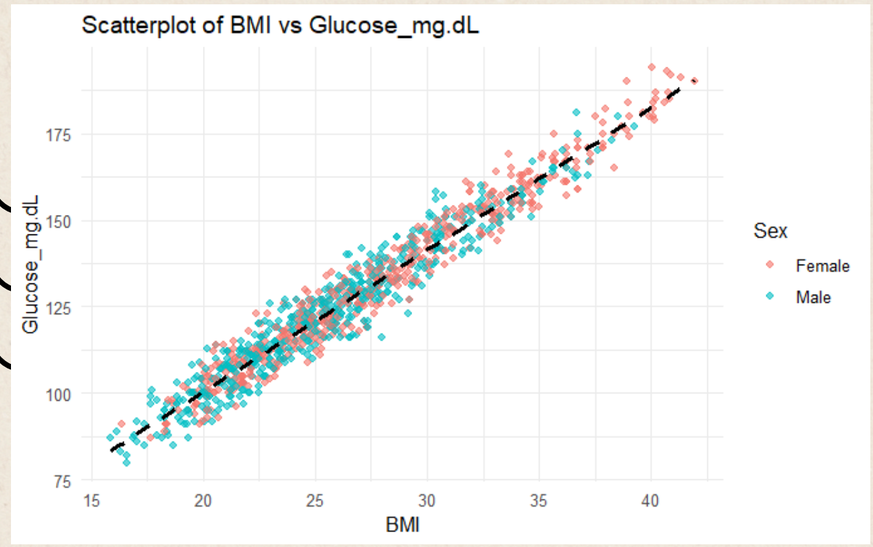


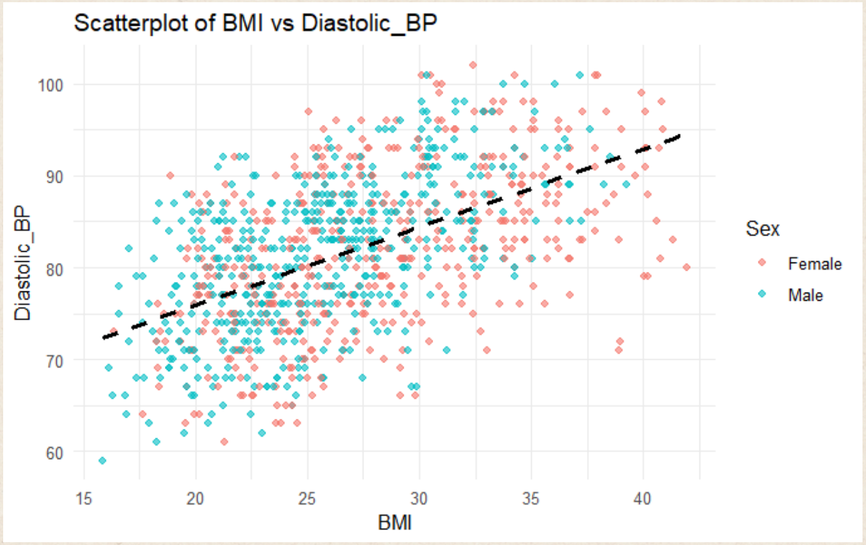
BOXPLOTS(Figure 5-7)

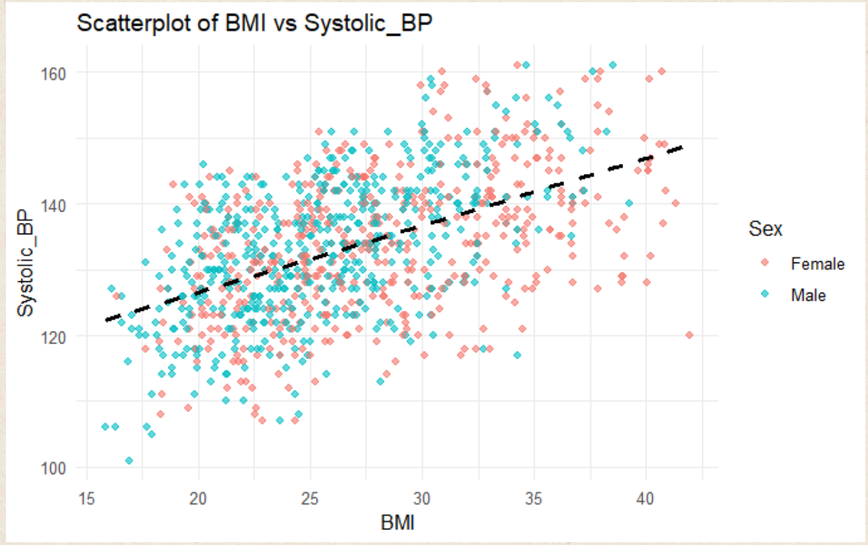
The boxplots showed that females had higher and more variable glucose and cholesterol levels compared to males, while BMI distributions were similar across sexes, with males showing slightly higher outliers.

PLOTS





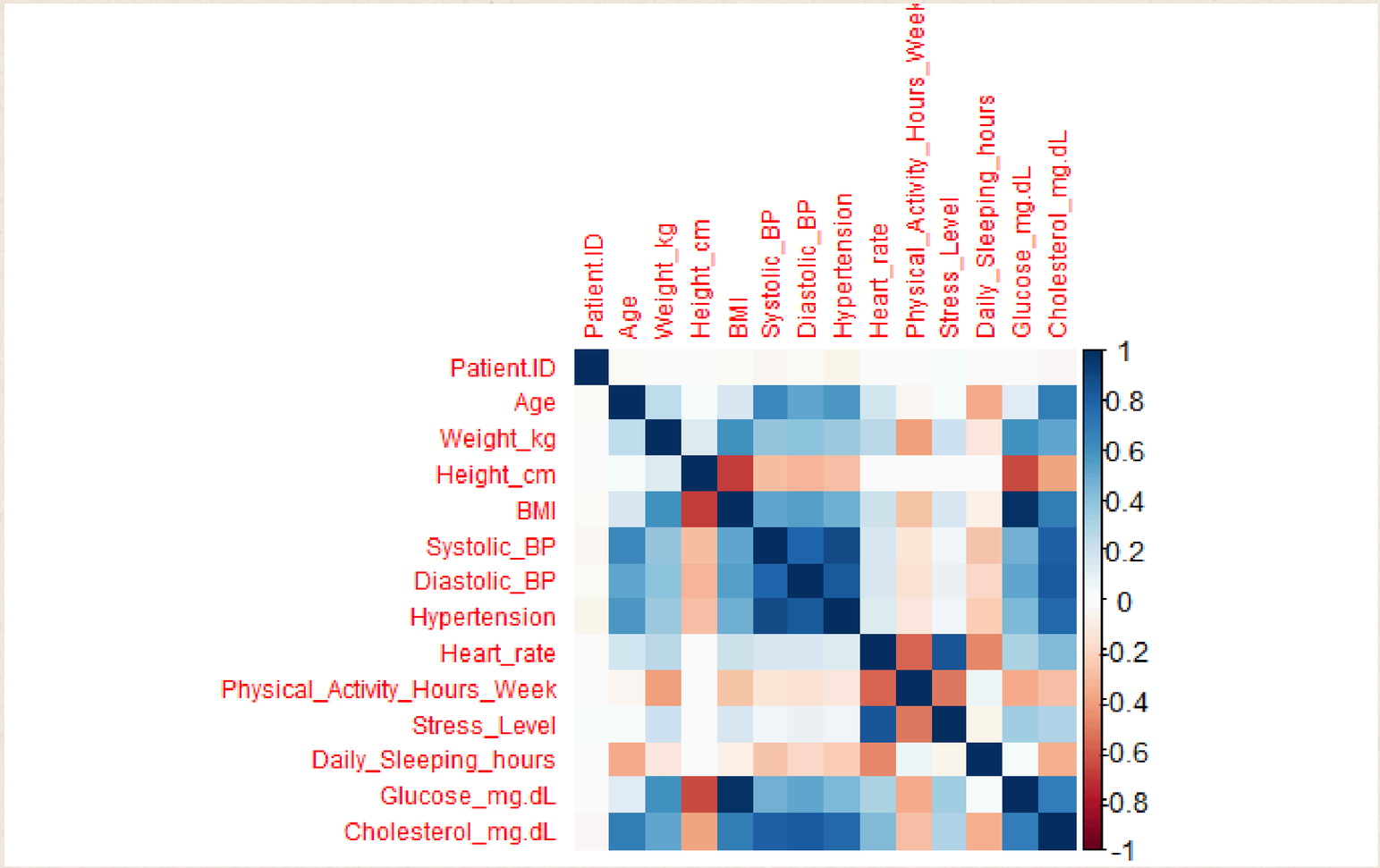




Scatter plots(Figure 8-11)

BMI IS POSITIVELY CORRELATED with diastolic_BP, Systolic_BP, Glucose, and Cholesterol

PLOTS



CORR PLOT(Figure 12)

The correlation plot visually highlighted strong positive relationships between BMI, glucose, and cholesterol levels. It also showed high correlations between systolic and diastolic blood pressure, as well as a negative correlation between physical activity and heart rate.

INTERPRETATION AND CONCLUSION

THE ANALYSIS REVEALED KEY HEALTH PATTERNS IN THE DATASET. HIGHER BMI WAS STRONGLY ASSOCIATED WITH ELEVATED GLUCOSE AND CHOLESTEROL LEVELS, INDICATING INCREASED METABOLIC RISK. FEMALES SHOWED SIGNIFICANTLY HIGHER GLUCOSE LEVELS THAN MALES, AS CONFIRMED BY STATISTICAL TESTING. ADDITIONALLY, LOWER PHYSICAL ACTIVITY AND SLEEP DURATION WERE LINKED TO HIGHER HEART RATES AND STRESS LEVELS.

OVERALL, THE FINDINGS SUGGEST THAT LIFESTYLE FACTORS SUCH AS PHYSICAL ACTIVITY, SLEEP, AND BMI PLAY A CRITICAL ROLE IN INFLUENCING VITAL HEALTH INDICATORS. PROMOTING HEALTHIER HABITS MAY HELP REDUCE RISKS RELATED TO BLOOD PRESSURE, GLUCOSE, AND CHOLESTEROL.