**NAME: SIDDHI SUBHEDAR**

**PRN: 1262240621**

**DESCRIPTIVE STATISTICS IN R**

install.packages("readr")

library(readr)

myData <- read\_csv("C:/Users/SoECE/Downloads/archive (16)/CardioGoodFitness.csv")

head(data)

> head(data)

# A tibble: 6 × 9

Product Age Gender Education MaritalStatus Usage Fitness Income Miles

*<chr>* *<dbl>* *<chr>* *<dbl>* *<chr>* *<dbl>* *<dbl>* *<dbl>* *<dbl>*

1 TM195 18 Male 14 Single 3 4 29562 112

2 TM195 19 Male 15 Single 2 3 31836 75

3 TM195 19 Female 14 Partnered 4 3 30699 66

4 TM195 19 Male 12 Single 3 3 32973 85

5 TM195 20 Male 13 Partnered 4 2 35247 47

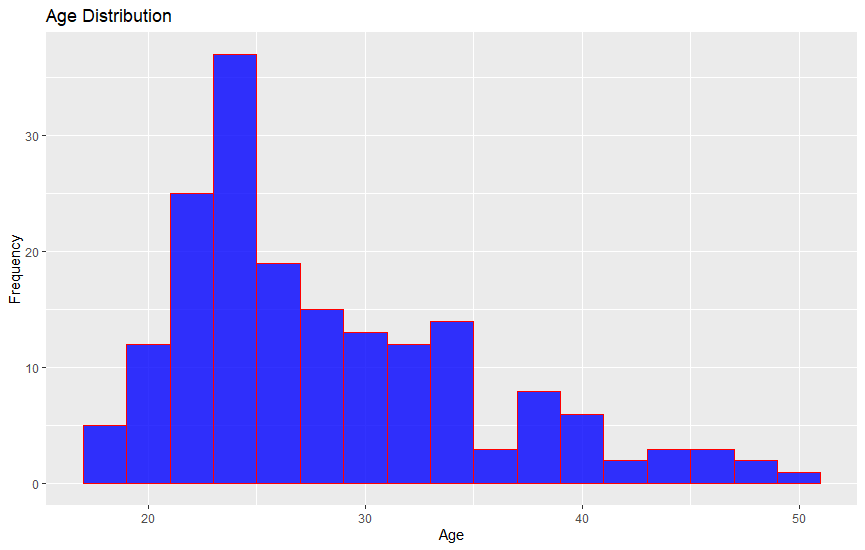
6 TM195 20 Female 14 Partnered 3 3 32973 66

library(ggplot2)

ggplot(myData, aes(x = Age)) +

geom\_histogram(binwidth = 2, fill = "blue", color = "red", alpha = 0.8) +

labs(title = "Age Distribution", x = "Age", y = "Frequency")

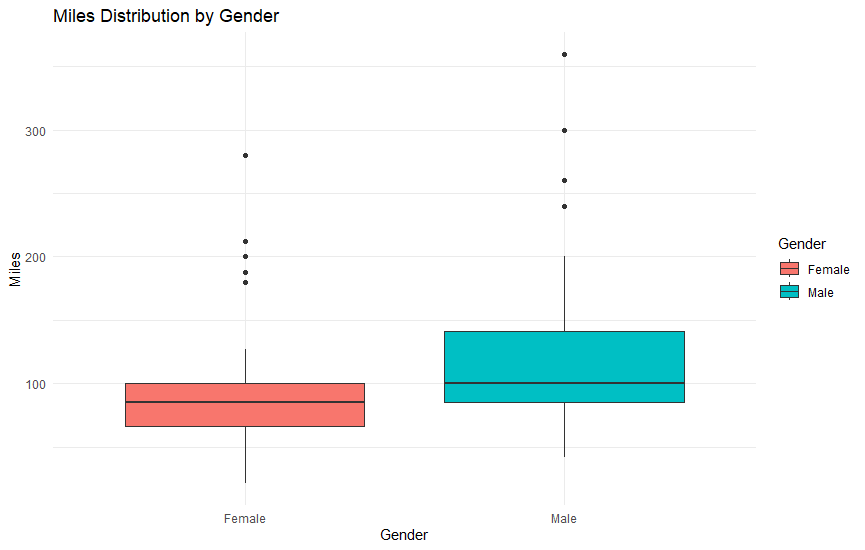


ggplot(myData, aes(x = Gender, y = Miles, fill = Gender)) +

geom\_boxplot() +

labs(title = "Miles Distribution by Gender", x = "Gender", y = "Miles") +

theme\_minimal()



mean = mean(myData$Age)

print(mean)

> print(mean)

[1] 28.78889

median = median(myData$Age)

print(median)

> print(median)

[1] 26

# Define a function to calculate mode

calculate\_mode <- function(x) {

# Remove NA values

x <- na.omit(x)

# Get the frequency of each value

freq\_table <- table(x)

# Get the mode (the value with the highest frequency)

mode\_value <- as.numeric(names(freq\_table[freq\_table == max(freq\_table)]))

return(mode\_value)

}

# Example: Assuming myData is your dataframe and Age is the column

myData <- data.frame(Age = c(21, 22, 23, 21, 22, 21)) # Example data

# Calculate the mode of the Age column

mode <- calculate\_mode(myData$Age)

# Print the mode

print(mode)

> print(mode)

[1] 21

max = max(myData$Age)

# Calculate the minimum

min = min(myData$Age)

# Calculate the range

range = max - min

cat("Range is:\n")

print(range)

Range is:

> print(range)

[1] 2

r = range(myData$Age)

print(r)

> print(r)

[1] 21 23

variance = var(myData$Age)

print(variance)

> print(variance)

[1] 0.6666667

std = sd(myData$Age)

print(std)

> print(std)

[1] 0.8164966

summary = summary(myData$Age)

print(summary)

> print(summary)

Min. 1st Qu. Median Mean 3rd Qu. Max.

21.00 21.00 21.50 21.67 22.00 23.00

summary = summary(myData)

print(summary)

> print(summary)

Age

Min. :21.00

1st Qu.:21.00

Median :21.50

Mean :21.67

3rd Qu.:22.00

Max. :23.00