#experiment-1

age\_intervals <- c("1-5", "5-15", "15-20", "20-50", "50-80", "80-110")

frequencies <- c(200, 450, 300, 1500, 700, 44)

cumulative\_frequencies <- cumsum(frequencies)

total\_observations <- sum(frequencies)

median\_class\_index <- which(cumulative\_frequencies >= total\_observations / 2)[1]

lower\_boundary <- as.numeric(unlist(strsplit(age\_intervals[median\_class\_index], "-")))[1]

frequency\_median\_class <- frequencies[median\_class\_index]

cumulative\_frequency\_before <- if (median\_class\_index == 1) 0 else cumulative\_frequencies[median\_class\_index - 1]

median\_class\_width <- as.numeric(unlist(strsplit(age\_intervals[median\_class\_index], "-")))[2] - lower\_boundary

median\_value <- lower\_boundary + ((total\_observations / 2 - cumulative\_frequency\_before) / frequency\_median\_class) \* median\_class\_width

print(paste("The approximate median value is:", median\_value))

#EXPERIMENT\_2

age\_values <- c(13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70)

mean\_value <- mean(age\_values)

median\_value <- median(age\_values)

mode\_value <- as.numeric(names(sort(table(age\_values), decreasing = TRUE)[1]))

midrange\_value <- (min(age\_values) + max(age\_values)) / 2

Q1 <- quantile(age\_values, 0.25)

Q3 <- quantile(age\_values, 0.75)

cat("Mean:", mean\_value, "\n")

cat("Median:", median\_value, "\n")

cat("Mode:", mode\_value, "\n")

cat("Midrange:", midrange\_value, "\n")

cat("First Quartile (Q1):", Q1, "\n")

cat("Third Quartile (Q3):", Q3, "\n")

# EXPERIMENT-3

# Original data

data <- c(200, 300, 400, 600, 1000)

# Min-max normalization

min\_val <- min(data)

max\_val <- max(data)

normalized\_data <- (data - min\_val) / (max\_val - min\_val)

normalized\_data

# Original data

data <- c(200, 300, 400, 600, 1000)

# Z-score normalization

mean\_val <- mean(data)

std\_dev <- sd(data)

z\_score\_normalized\_data <- (data - mean\_val) / std\_dev

z\_score\_normalized\_data

#EXPRIMENT-4

# Original data

data <- c(11, 13, 13, 15, 15, 16, 19, 20, 20,40, 45, 45, 45, 71, 72, 73, 75)

bin\_size <- 5

smoothed\_data\_boundaries <- unlist(lapply(bins, function(bin) {

min\_val <- min(bin)

max\_val <- max(bin)

sapply(bin, function(x) ifelse(x - min\_val < max\_val - x, min\_val, max\_val))

}))

smoothed\_data\_boundaries

#EXPERIMENT-5

data <- c(200, 300, 400, 600, 1000)

normalized\_data <- (data - min(data)) / (max(data) - min(data))

z\_score\_normalized\_data <- (data - mean(data)) / sd(data)

data\_binning <- c(11, 13, 13, 15, 15, 16, 19, 20, 20, 20, 21, 21, 22, 23, 24, 30, 40, 45, 45, 45, 71, 72, 73, 75)

bins <- split(data\_binning, ceiling(seq\_along(data\_binning) / 5))

smoothed\_data\_mean <- unlist(lapply(bins, mean))

smoothed\_data\_median <- unlist(lapply(bins, median))

smoothed\_data\_boundaries <- unlist(lapply(bins, function(bin) {

sapply(bin, function(x) ifelse(x - min(bin) < max(bin) - x, min(bin), max(bin)))

}))

#experiment-5

ages <- c(22, 23, 25, 27, 30, 32, 35, 36, 40, 42, 45, 48, 50, 52, 55, 58, 60, 62)

body\_fat <- c(15.2, 16.3, 14.8, 18.1, 19.2, 20.5, 21.7, 23.4, 24.8, 26.5, 27.9, 29.1, 30.4, 32.7, 34.1, 35.2, 36.5, 37.8)

summary(ages)

summary(body\_fat)

correlation <- cor(ages, body\_fat)

plot(ages, body\_fat, main="Age vs Body Fat Percentage", xlab="Age (years)", ylab="Body Fat Percentage", pch=19, col="blue")

abline(lm(body\_fat ~ ages), col="pink")

#eXPERIMENT-6

ages <- c(23, 25, 27, 30, 32, 35, 37, 40, 42, 45, 47, 50, 52, 55, 57, 60, 62, 65)

age\_value <- 35

sd\_age <- 12.94

min\_max <- (age\_value - min(ages)) / (max(ages) - min(ages))

z\_score <- (age\_value - mean(ages)) / sd\_age

decimal\_scale <- age\_value / (10^ceiling(log10(max(abs(ages)))))

cat("Min-Max:", min\_max, "\nZ-score:", z\_score, "\nDecimal Scaling:", decimal\_scale, "\n")

#EXPERIMENT-7

pencils <- c(9, 25, 23, 12, 11, 6, 7, 8, 9, 10)

mean\_value <- mean(pencils)

median\_value <- median(pencils)

mode\_value <- as.numeric(names(sort(table(pencils), decreasing = TRUE)[1]))

cat("Mean:", mean\_value, "\nMedian:", median\_value, "\nMode:", mode\_value, "\n")

#EXPRIMENT-8

x <- c(4, 1, 5, 7, 10, 2, 50, 25, 90, 36);

y <- c(12, 5, 13, 19, 31, 7, 153, 72, 275, 110);

plot(x, y, main="Mobile Phones Sold vs Money", xlab="Phones Sold", ylab="Money", col="blue", pch=19)

#EXPERIMENT-9

marks <- c(55,60,71,63,55,65,50,55,58,59,61,63,65,67,71,72,75)

eq\_freq <- split(sort(marks), cut(seq\_along(marks), 3, labels=FALSE))

eq\_width <- cut(marks, breaks=3)

hist(marks, main="Histogram of Marks", xlab="Marks", col="white", border="black")

print(eq\_freq); print(table(eq\_width))

#EXPERIMENT-10

speed <- c(78.3, 81.8, 82, 74.2, 83.4, 84.5, 82.9, 77.5, 80.9, 70.6)

iqr\_value <- IQR(speed)

sd\_value <- sd(speed)

cat("Interquartile Range (IQR):", iqr\_value, "\n")

cat("Standard Deviation:", sd\_value, "\n")

#EXPERIMENT-11

data <- c(13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70)

Q1 <- quantile(data, 0.25)

Q3 <- quantile(data, 0.75)

print(paste("First Quartile (Q1):", Q1))

print(paste("Third Quartile (Q3):", Q3))

speed <- c(78.3, 81.8, 82, 74.2, 83.4, 84.5, 82.9, 77.5, 80.9, 70.6)

iqr\_speed <- IQR(speed)

sd\_speed <- sd(speed)

iqr\_speed

sd\_speed

#experiment-12

A <- c(18, 2, 20)

B <- c(22, 28, 10)

C <- c(20, 40, 40)

preferences <- data.frame(A, B, C)

cov\_BC <- cov(B, C)

print(paste("Covariance between B and C:", cov\_BC))

cov\_matrix <- cov(preferences)

print("Covariance matrix for the preferences:")

print(cov\_matrix)

cor\_BC <- cor(B, C)

print(paste("Correlation between B and C:", cor\_BC))

cor\_matrix <- cor(preferences)

print("Correlation matrix for the preferences:")

print(cor\_matrix)

#experiment-13

prices <- c(1, 1, 5, 5, 5, 5, 5)

summary(prices)

hist(prices)

mean(prices)

median(prices)

var(prices)

sd(prices)

#experiment\_14

data <- c(8, 8, 10, 10, 10, 10, 12, 14, 14, 14, 15, 15, 15, 15, 15, 15, 18, 18, 18, 18, 18)

bins <- split(data, ceiling(seq\_along(data) / 3))

smoothed\_data\_means <- unlist(lapply(bins, function(x) rep(mean(x), length(x))))

smoothed\_data\_boundaries <- unlist(lapply(bins, function(x) ifelse(x - min(x) < max(x) - x, min(x), max(x))))

hist(data, main = "Original Data", col = "lightblue")

hist(smoothed\_data\_means, main = "Smoothed Data (Means)", col = "lightgreen")

hist(smoothed\_data\_boundaries, main = "Smoothed Data (Boundaries)", col = "lightcoral")

#experiment-15

classA <- c(76, 35, 47, 64, 95, 66, 89, 36, 84)

classB <- c(51, 56, 84, 60, 59, 70, 63, 66, 50)

meanA <- mean(classA)

medianA <- median(classA)

rangeA <- range(classA)

rangeA\_value <- diff(rangeA)

meanB <- mean(classB)

medianB <- median(classB)

rangeB <- range(classB)

rangeB\_value <- diff(rangeB)

cat("Class A - Mean:", meanA, " Median:", medianA, " Range:", rangeA\_value, "\n")

cat("Class B - Mean:", meanB, " Median:", medianB, " Range:", rangeB\_value, "\n")

boxplot(classA, classB, names = c("Class A", "Class B"), main = "Boxplot of Class A and Class B Scores")

#experiment-16

data <- c(200, 300, 400, 600, 1000)

min\_max\_normalized <- (data - min(data)) / (max(data) - min(data))

cat("Min-Max Normalized Data:\n", min\_max\_normalized, "\n\n")

z\_score\_normalized <- (data - mean(data)) / sd(data)

cat("Z-Score Normalized Data:\n", z\_score\_normalized, "\n")

#experiment-17

data <- AirPassengers

hist(data,

breaks = seq(100, 700, by = 150),

xlim = c(100, 700),

main = "Histogram of AirPassengers Dataset",

xlab = "Number of Passengers",

ylab = "Frequency",

col = "lightblue",

border = "black")

#EXPERIMENT-18

data <- mtcars

plot(data$mpg, type = "o", col = "blue", pch = 16, lty = 1,

xlab = "Index", ylab = "Values", main = "Multiple Lines in a Single Plot",

ylim = range(c(data$mpg, data$qsec)))

lines(data$qsec, type = "o", col = "red", pch = 17, lty = 2)

legend("topright", legend = c("mpg", "qsec"),

col = c("blue", "red"), pch = c(16, 17), lty = c(1, 2))

#EXPERIMENT-19

data <- data.frame(

hardness = c(50, 60, 70, 80, 90, 100, 110),

mortality = c(200, 190, 180, 170, 160, 150, 140)

)

plot(data$hardness, data$mortality,

main = "Scatter Plot of Mortality vs Hardness",

xlab = "Hardness",

ylab = "Mortality",

col = "blue",

pch = 19)

model <- lm(mortality ~ hardness, data = data)

abline(model, col = "red")

hardness\_value <- data.frame(hardness = 88)

predicted\_mortality <- predict(model, newdata = hardness\_value)

cat("Predicted Mortality for Hardness 88:", predicted\_mortality, "\n")

#EXPERIMENT-20

data <- mtcars

boxplot(mpg ~ cyl, data = data,

main = "Boxplot of MPG by Number of Cylinders",

xlab = "Number of Cylinders",

ylab = "Miles Per Gallon",

col = c("lightblue", "lightgreen", "lightpink"))

#EXPERIMENT-21

player\_scores <- c(15, 18, 22, 24, 25, 27, 30, 32, 33, 100) # Example data

boxplot(player\_scores,

main = "Boxplot of Player Scores",

ylab = "Points Scored",

col = "lightblue",

horizontal = FALSE)

legend("topright", legend = "Outliers are indicated by dots", bty = "n")

Here are the short versions of the code for all 21 experiments:

Experiment 1: Median calculation

data <- c(1, 2, 3, 4, 5)

median\_value <- median(data)

print(paste("Median value:", median\_value))

Experiment 2: Mean, median, mode calculation

data <- c(13, 15, 16, 16, 19, 20, 20, 21, 22, 22, 25, 25, 25, 25, 30, 33, 33, 35, 35, 35, 35, 36, 40, 45, 46, 52, 70)

mean\_value <- mean(data)

median\_value <- median(data)

mode\_value <- as.numeric(names(sort(table(data), decreasing = TRUE)[1]))

print(paste("Mean:", mean\_value, " Median:", median\_value, " Mode:", mode\_value))

Experiment 3: Min-max normalization

data <- c(200, 300, 400, 600, 1000)

min\_max\_normalized <- (data - min(data)) / (max(data) - min(data))

print(paste("Min-Max Normalized Data:", min\_max\_normalized))

Experiment 4: Z-score normalization

data <- c(200, 300, 400, 600, 1000)

z\_score\_normalized <- (data - mean(data)) / sd(data)

print(paste("Z-Score Normalized Data:", z\_score\_normalized))

Experiment 5: Data binning

data <- c(11, 13, 13, 15, 15, 16, 19, 20, 20, 20, 21, 21, 22, 23, 24, 30, 40, 45, 45, 45, 71, 72, 73, 75)

bins <- split(data, ceiling(seq\_along(data) / 5))

print(paste("Binned Data:", bins))

Experiment 6: Smoothing data

data <- c(11, 13, 13, 15, 15, 16, 19, 20, 20, 20, 21, 21, 22, 23, 24, 30, 40, 45, 45, 45, 71, 72, 73, 75)

bins <- split(data, ceiling(seq\_along(data) / 5))

smoothed\_data\_means <- unlist(lapply(bins, mean))

print(paste("Smoothed Data (Means):", smoothed\_data\_means))

Experiment 7: Histogram

data <- c(11, 13, 13, 15, 15, 16, 19, 20, 20, 20, 21, 21, 22, 23, 24, 30, 40, 45, 45, 45, 71, 72, 73, 75)

hist(data, main = "Histogram of Data", col = "lightblue")

Experiment 8: Box plot

data <- c(8, 8, 10, 10, 10, 10, 12, 14, 14, 14, 15, 15, 15, 15, 15, 15, 18, 18, 18, 18, 18)

boxplot(data, main = "Boxplot of Data", col = "lightblue")

Experiment 9: Scatter plot

x <- c(1, 2, 3, 4, 5)

y <- c(2, 4, 6, 8, 10)

plot(x, y, main = "Scatter Plot of X vs Y", col = "lightblue", pch = 19)

Experiment 10: Correlation

x <- c(1, 2, 3, 4, 5)

y <- c(2, 4, 6, 8, 10)

correlation <- cor(x, y)

print(paste("Correlation between X and Y:", correlation))

Experiment 11: Covariance

x <- c(1, 2, 3, 4, 5)

y <- c(2, 4, 6, 8, 10)

covariance <- cov(x, y)

print(paste("Covariance between X and Y:", covariance))

Equal Width Binning

data <- c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)

bins <- cut(data, breaks = 5)

print(bins)

Equal Frequency Binning

data <- c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)

bins <- quantile(data, probs = seq(0, 1, 0.25))

print(cut(data, bins))

K-Means Binning

library(cluster)

data <- c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)

km <- kmeans(data, 5)

print(cut(data, km$centers))

Hierarchical Binning

library(cluster)

data <- c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10)

hc <- hclust(dist(data))

print(cutree(hc, k = 5))

Here are the remaining experiments:

Experiment 10: Correlation

x <- c(1, 2, 3, 4, 5)

y <- c(2, 4, 6, 8, 10)

correlation <- cor(x, y)

print(paste("Correlation between X and Y:", correlation))

Experiment 11: Covariance

x <- c(1, 2, 3, 4, 5)

y <- c(2, 4, 6, 8, 10)

covariance <- cov(x, y)

print(paste("Covariance between X and Y:", covariance))

Experiment 12: Regression

x <- c(1, 2, 3, 4, 5)

y <- c(2, 4, 6, 8, 10)

model <- lm(y ~ x)

print(paste("Regression equation: y =",

model$coefficients[1], "+",model$coefficients[2], "\* x"))

Experiment 13: Data transformation

data <- c(1, 2, 3, 4, 5)

log\_data <- log(data)

print(paste("Log-transformed data:", log\_data))

Experiment 14: Data standardization

data <- c(1, 2, 3, 4, 5)

standardized\_data <- (data - mean(data)) / sd(data)

print(paste("Standardized data:", standardized\_data))

Experiment 15: Box plot comparison

data1 <- c(1, 2, 3, 4, 5)

data2 <- c(2, 4, 6, 8, 10)

boxplot(data1, data2, main = "Boxplot Comparison", col = c("lightblue", "lightgreen"))

Experiment 16: Histogram comparison

data1 <- c(1, 2, 3, 4, 5)

data2 <- c(2, 4, 6, 8, 10)

hist(data1, main = "Histogram Comparison", col = "lightblue", border = "black")

hist(data2, add = TRUE, col = "lightgreen", border = "black")

Experiment 17: Scatter plot with regression line

x <- c(1, 2, 3, 4, 5)

y <- c(2, 4, 6, 8, 10)

model <- lm(y ~ x)

plot(x, y, main = "Scatter Plot with Regression Line")

abline(model, col = "red")

Experiment 18: Multiple scatter plots

x1 <- c(1, 2, 3, 4, 5)

y1 <- c(2, 4, 6, 8, 10)

x2 <- c(2, 4, 6, 8, 10)

y2 <- c(1, 3, 5, 7, 9)

plot(x1, y1, main = "Multiple Scatter Plots", col = "lightblue", pch = 19)

points(x2, y2, col = "lightgreen", pch = 17)

Experiment 19: Bar chart

data <- c(10, 20, 30, 40, 50)

barplot(data, main = "Bar Chart", col = "lightblue", border = "black")

Experiment 20: Pie chart

data <- c(10, 20, 30, 40)

pie(data, main = "Pie Chart", col = c("lightblue", "lightgreen", "lightyellow", "lightpink"))

Experiment 21: Heatmap

data <- matrix(c(10, 20, 30, 40, 50, 60, 70, 80, 90), nrow = 3, ncol = 3)

heatmap(data, main = "Heatmap", col = "lightblue")