**DAY-3 LAB EXPERIMENTS**

**R PROGRAMMING**

**EXPERIMENT 1:**

**Covariance and correlation**

**Children of three ages are asked to indicate their preference for three photographs of adults. Do the data suggest that there is a significant relationship between age and photograph preference? What is wrong with this study?**

**Photograph:**

**Age of child A B C**

**5-6 years: 18 22 20**

**7-8 years: 2 28 40**

**9-10 years: 20 10 40**

**1. Use cov() to calculate the sample covariance between B and C.**

**2. Use another call to cov() to calculate the sample covariance matrix for the preferences.**

**3. Use cor() to calculate the sample correlation between B and C.**

**4. Use another call to cor() to calculate the sample correlation matrix for the preferences.**

**AIM:**

To find the covariance and correlation of the given data

**SOFTWARE REQUIRES:**

R SOFTWARE

**PROGRAM:**

age\_groups <- c("5-6 years", "7-8 years", "9-10 years")

photograph\_A <- c(18, 2, 20)

photograph\_B <- c(22, 28, 10)

photograph\_C <- c(20, 40, 40)

data <- data.frame(Age = age\_groups, Photograph\_A = photograph\_A, Photograph\_B = photograph\_B, Photograph\_C = photograph\_C)

cov\_BC <- cov(data$Photograph\_B, data$Photograph\_C)

cov\_matrix <- cov(data[, c("Photograph\_A", "Photograph\_B", "Photograph\_C")])

cor\_BC <- cor(data$Photograph\_B, data$Photograph\_C)

cor\_matrix <- cor(data[, c("Photograph\_A", "Photograph\_B", "Photograph\_C")])

cat("Sample Covariance between B and C:", cov\_BC, "\n")

cat("Sample Covariance Matrix for the preferences:\n")

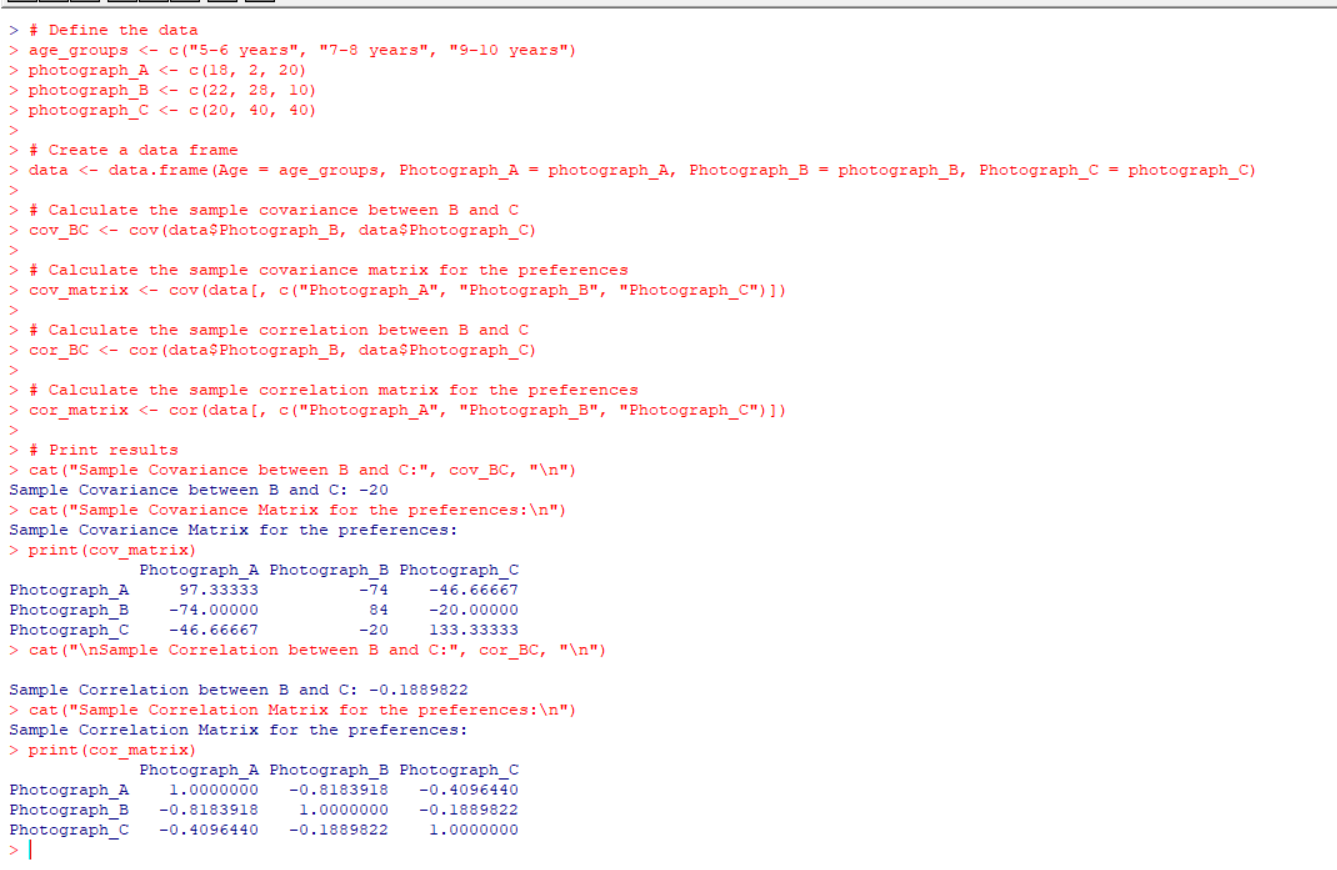
print(cov\_matrix)

cat("\nSample Correlation between B and C:", cor\_BC, "\n")

cat("Sample Correlation Matrix for the preferences:\n")

print(cor\_matrix)

**OUTPUT:**

****

**EXPERIMENT 2:**

**Imagine that you have selected data from the All Electronics data warehouse for analysis. The data set will be huge! The following data are a list of All Electronics prices for commonly sold items (rounded to the nearest dollar). The numbers have been sorted: 1, 1, 5, 5, 5, 5, 5, 8, 8, 10, 10, 10, 10, 12, 14, 14, 14, 15, 15, 15, 15, 15, 15, 18, 18, 18, 18, 18,**

| **18, 18, 18, 20, 20, 20, 20, 20, 20, 20, 21, 21, 21, 21, 25, 25, 25, 25, 25, 28, 28, 30,** |
| --- |
| **30, 30.**  **(i) Partition the dataset using an equal-frequency partitioning method with bin equal to 3 (ii) apply data smoothing using bin means and bin boundary.  (iii) Plot Histogram for the above frequency division** |

**AIM:**

To draw the histogram graph for the given data

**SOFTWARE REQUIRES:**

R SOFTWARE

**PROGRAM:**

prices <- c(1, 1, 5, 5, 5, 5, 5, 8, 8, 10, 10, 10, 10, 12, 14, 14, 14, 15, 15, 15,

15, 15, 15, 18, 18, 18, 18, 18, 18, 18, 18, 20, 20, 20, 20, 20, 20,

20, 21, 21, 21, 21, 25, 25, 25, 25, 25, 28, 28, 30, 30, 30)

bins <- cut(prices, breaks = 3, labels = c("Low", "Medium", "High"), include.lowest = TRUE)

bin\_means <- tapply(prices, bins, mean)

smoothed\_prices <- unlist(lapply(bins, function(x) bin\_means[x]))

hist(prices, breaks = 3, main = "Histogram of Prices with Equal-frequency Partitioning",

xlab = "Prices", ylab = "Frequency", col = "lightblue")

**OUTPUT:**



**EXPERIMENT 3:**

**Two Maths teachers are comparing how their Year 9 classes performed in the end of year exams. Their results are as follows:  
 Class A: 76, 35, 47, 64, 95, 66, 89, 36, 8476,35,47,64,95,66,89,36,84**

**Class B: 51, 56, 84, 60, 59, 70, 63, 66, 5051,56,84,60,59,70,63,66,50**

**(i) Find which class had scored higher mean, median and range.  
 (ii) Plot above in boxplot and give the inferences**

**AIM:**

To find the mean,median,range and draw the box plot if the given data

**SOFTWARE REQUIRES:**

R SOFTWARE

**PROGRAM:**

class\_A <- c(76, 35, 47, 64, 95, 66, 89, 36, 84)

class\_B <- c(51, 56, 84, 60, 59, 70, 63, 66, 50)

mean\_A <- mean(class\_A)

median\_A <- median(class\_A)

range\_A <- max(class\_A) - min(class\_A)

mean\_B <- mean(class\_B)

median\_B <- median(class\_B)

range\_B <- max(class\_B) - min(class\_B)

comparison <- data.frame(Class = c("A", "B"),

Mean = c(mean\_A, mean\_B),

Median = c(median\_A, median\_B),

Range = c(range\_A, range\_B))

print(comparison)

boxplot(class\_A, class\_B, names = c("Class A", "Class B"),

main = "Boxplot of Exam Scores",

xlab = "Class", ylab = "Scores",

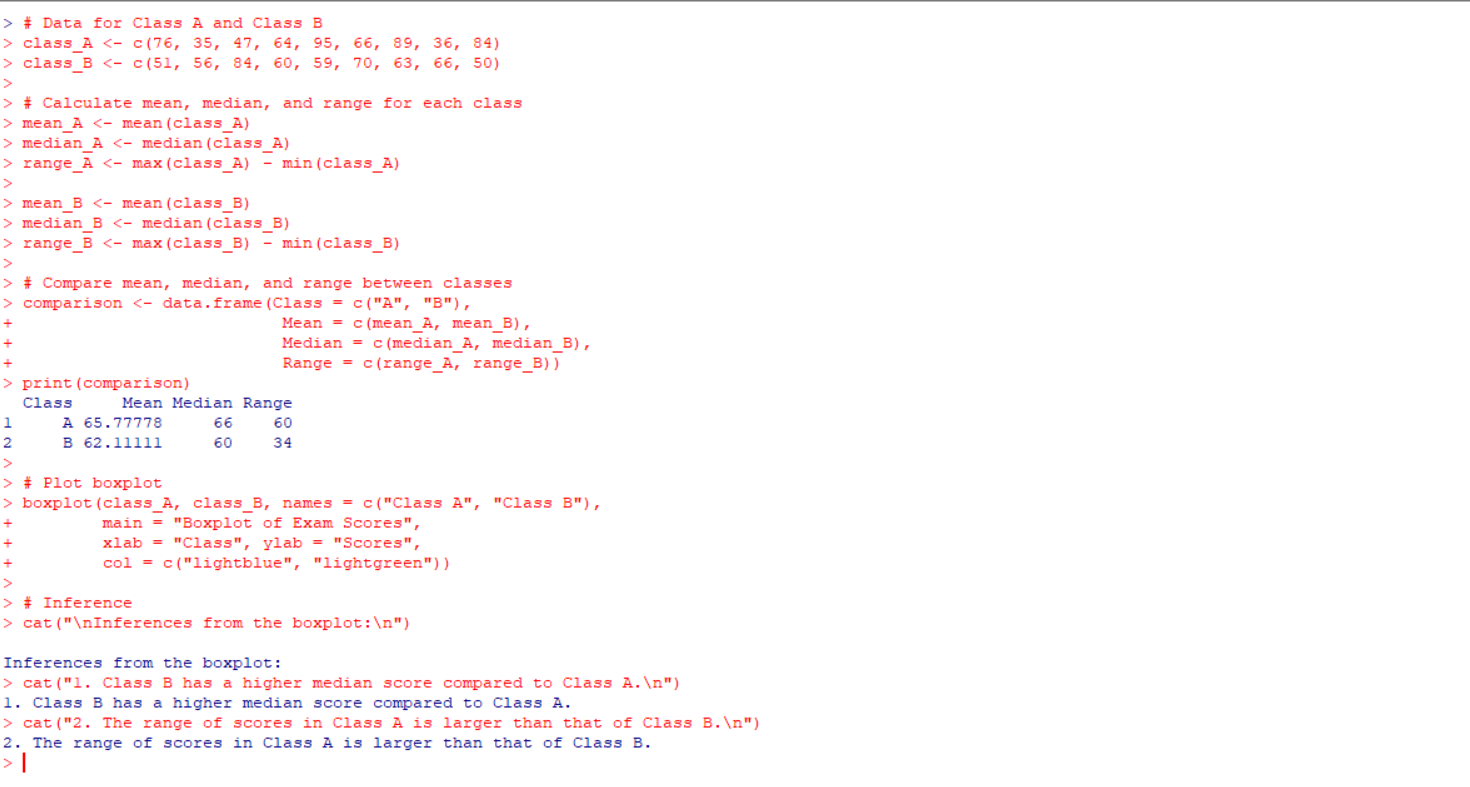
col = c("lightblue", "lightgreen"))

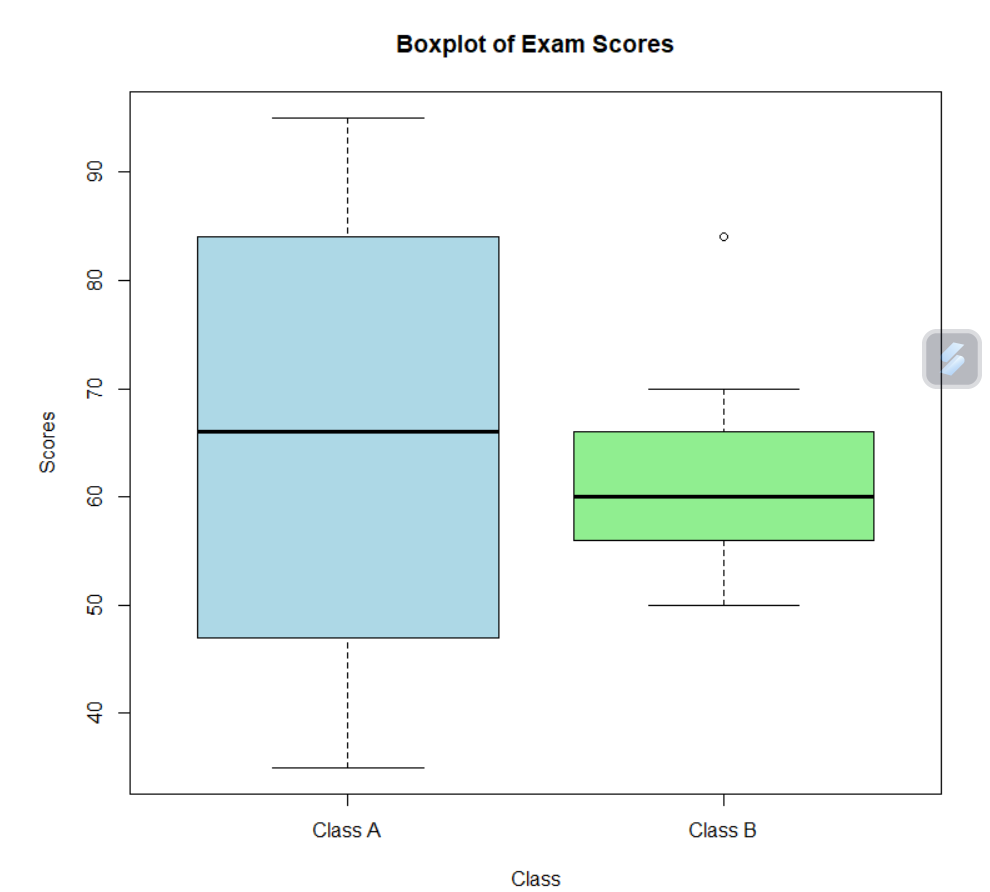
cat("\nInferences from the boxplot:\n")

cat("1. Class B has a higher median score compared to Class A.\n")

cat("2. The range of scores in Class A is larger than that of Class B.\n")

**OUTPUT:**

****

****

**EXPERIMENT 4:**

**Let us consider one example to make the calculation method clear. Assume that the minimum and maximum values for the feature F are $50,000 and $100,000 correspondingly. It needs to range *F* from 0 to 1. In accordance with min-max normalization, *v* = $80,**

**b) Use the two methods below to normalize the following group of data: 200, 300, 400, 600, 1000**

**(a) min-max normalization by setting min = 0 and max = 1**

**(b) z-score normalization**

**AIM:**

To find the min\_max ,z score normalisation of the given data

**SOFTWARE REQUIRES:**

R software

**PROGRAM:**

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

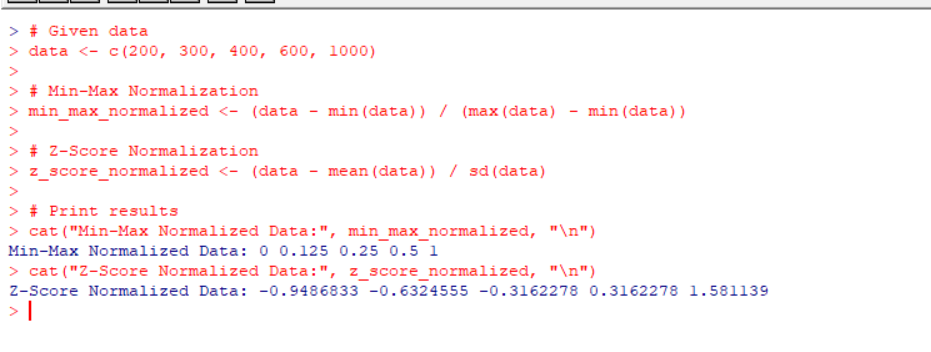
min\_max\_normalized <- (data - min(data)) / (max(data) - min(data))

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

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

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

**OUTPUT:**

****

**EXPERIMENT 5:**

**Make a histogram for the “AirPassengers “dataset, start at 100 on the x-axis, and from values 200 to 700, make the bins 150 wide**

**AIM:**

To draw the histogram graph for the given dataset

**SOFTWARE REQUIRES:**

R SOFTWARE

**PROGRAM:**

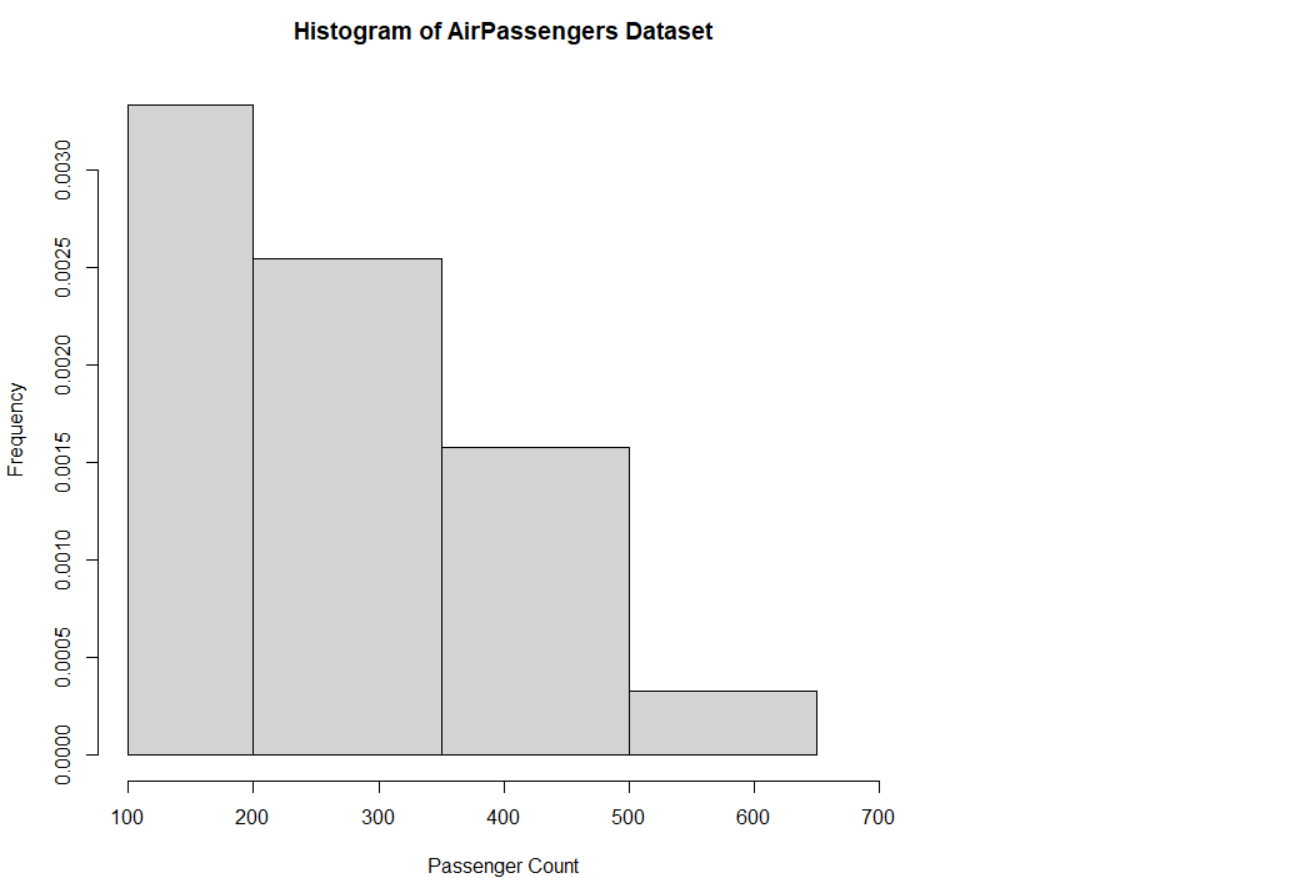
data("AirPassengers")

bin\_edges <- seq(200, 700, by = 150)

bin\_edges <- c(100, bin\_edges)

hist(AirPassengers, breaks = bin\_edges, xlim = c(100, 700), main = "Histogram of AirPassengers Dataset", xlab = "Passenger Count", ylab = "Frequency")

**OUTPUT:**

****