**DATA SCIENCE WITH R PROGRAMMING**

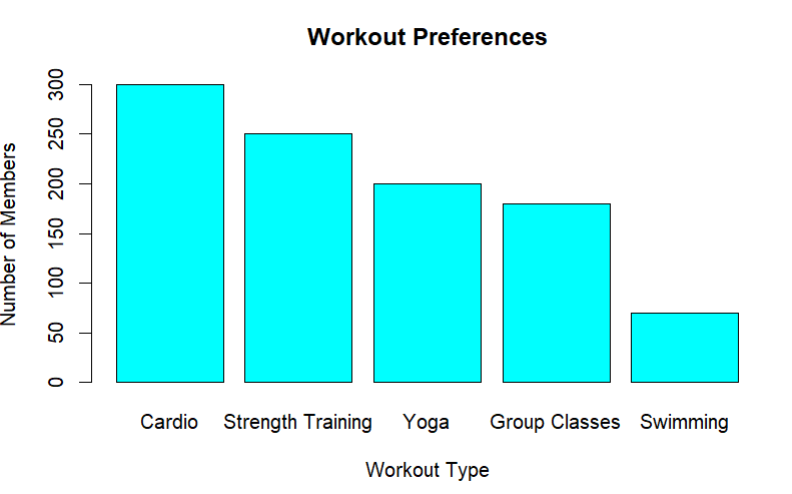
*1.Bar Chart:*

library(ggplot2)

workout <- c(300,250,200,180,70)

names(workout) <- c("Cardio","Strength Training","Yoga", "Group Classes","Swimming")

barplot(workout, main="Workout Preferences", xlab="Workout Type", ylab="Number of Members", col="cyan")



*2.Histogram:*

age\_gr <- c("18-25 years", "26-35 years", "36-45 years", "46-55 years", "56+ years")

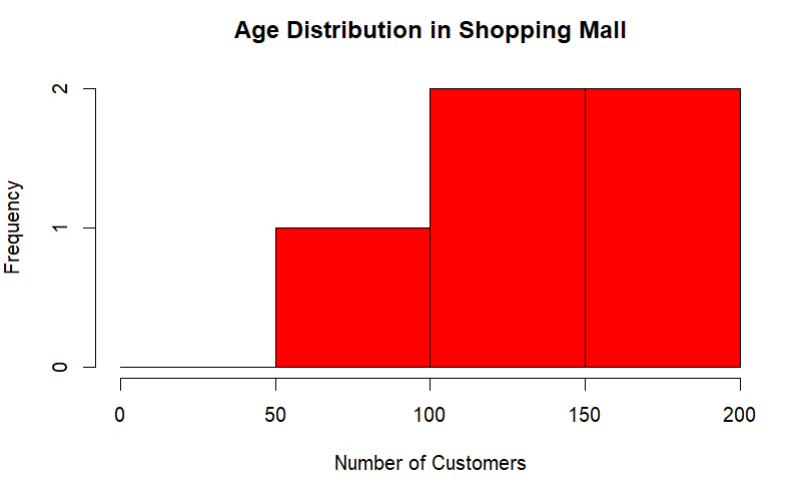
customer\_co <- c(200, 180, 150, 120, 70)

age\_data <- data.frame(Age\_Gr = age\_gr, Customers = customer\_co)

hist(age\_data$Customers, breaks = c(0, 50, 100, 150, 200), col = "Red",

xlab = "Number of Customers", ylab = "Frequency",

main = "Age Distribution in Shopping Mall")



*3.Boxplot:*

dep <- c("Marketing", "Finance", "IT", "Operations", "Other")

salary\_ranges <- list(

Marketing = c(50000, 80000),

Finance = c(60000, 90000),

IT = c(70000, 100000),

Operations = c(55000, 85000),

Sales = c(65000, 95000)

)

salary\_data <- data.frame(Department = rep(dep, each = 100),

Salary = unlist(lapply(salary\_ranges, function(range) {

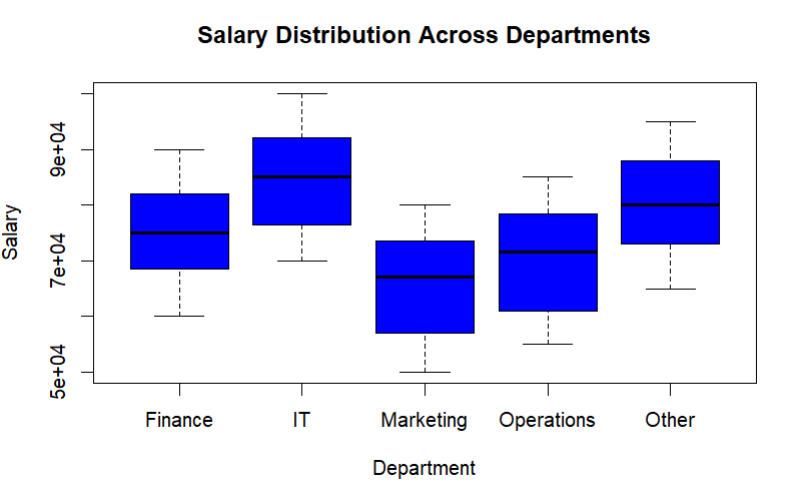
sample(seq(range[1], range[2], by = 1000), 100, replace = TRUE)

})))

boxplot(Salary ~ Department, data = salary\_data, col = "blue",

xlab = "Department", ylab = "Salary",

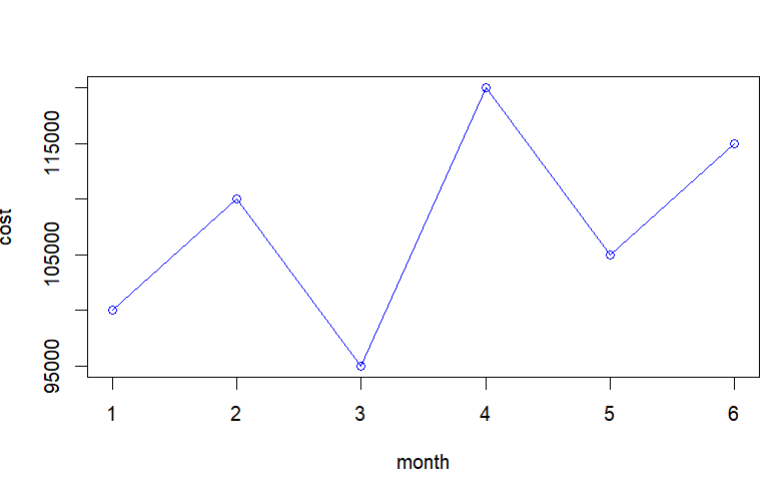
main = "Salary Distribution Across Departments")



*4.Line Chart:*

cost=c(100000,110000,95000,120000,105000,115000)

plot(cost,type='o',xlab='month',ylab='cost',col='blue')

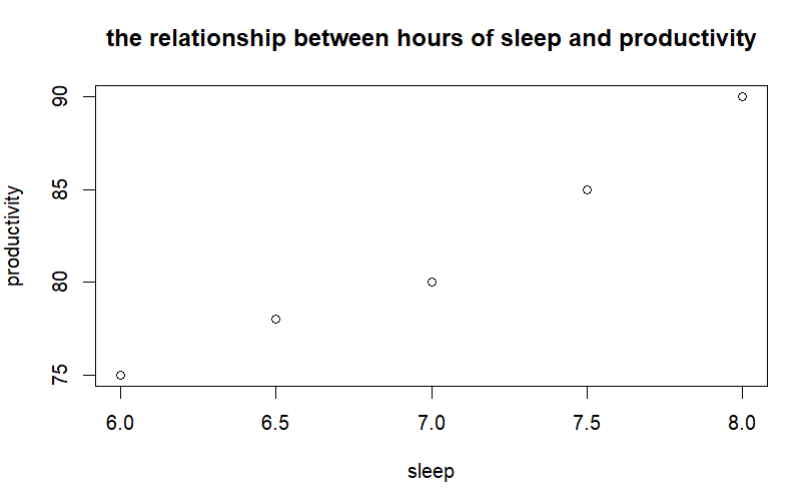


*5. Scatter Plot:*

sleep=c(7,6,8,6.5,7.5)

prod=c(80,75,90,78,85)

plot(sleep,prod,ylab='productivity',main='the relationship between hours of sleep and productivity')



*6.Bar Chart(mtcars database):*

data(mtcars)

average\_mpg <- tapply(mtcars$mpg, mtcars$cyl, mean)

barplot(average\_mpg,

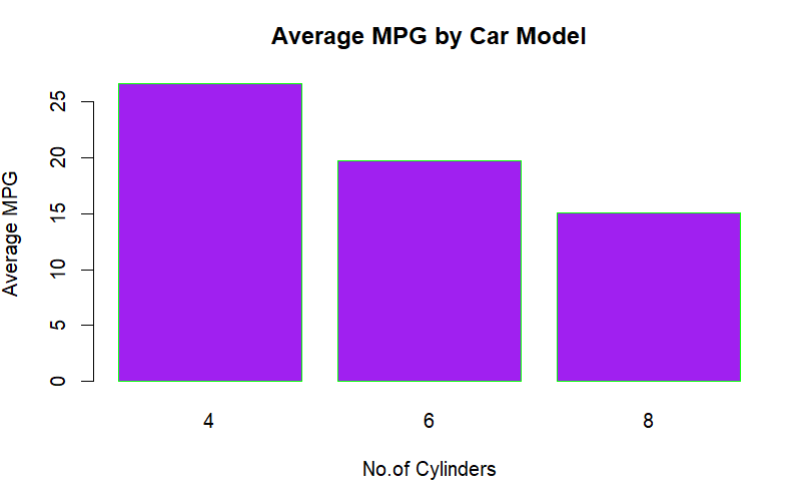
main = "Average MPG by Car Model",

xlab = "No.of Cylinders",

ylab = "Average MPG",

col = "purple",

border = "green")

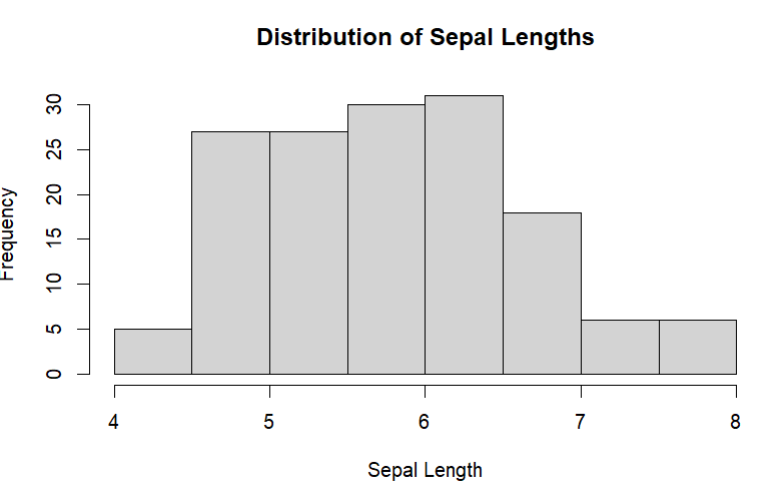


*7.Histogram (iris dataset):*

data(iris)

sepal\_lengths <- iris$Sepal.Length

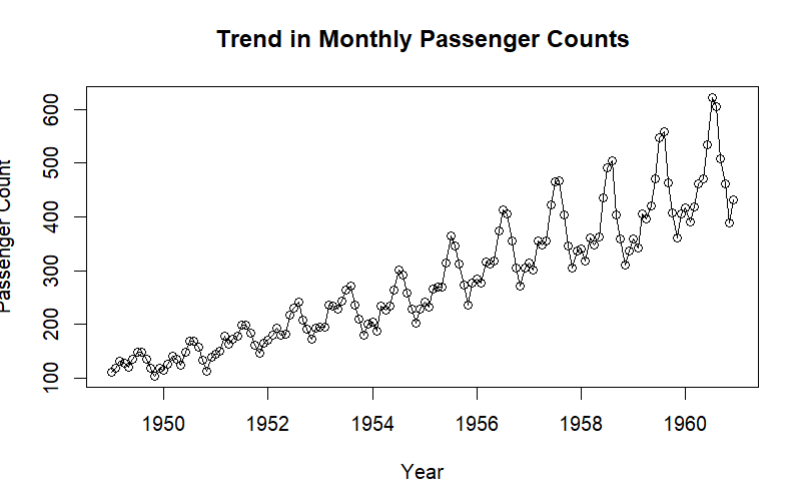
hist(sepal\_lengths, main = "Distribution of Sepal Lengths", xlab = "Sepal Length", ylab = "Frequency")



*8.Line Chart (AirPassenger dataset):*

data(AirPassengers)

plot(AirPassengers, type = "o", main = "Trend in Monthly Passenger Counts", xlab = "Year", ylab = "Passenger Count")



*9. Probability Distribution:*

lambda <- 27

x <- 20

probability <- 1 - ppois(x - 1, lambda)

cat("a) Probability of having 20 or more cars:", probability, "\n")

random\_numbers <- runif(19, min = 4, max = 5)

cat("b) 19 random numbers between 4 and 5:", random\_numbers, "\n")

rate <- 2

time <- 2

probability\_2\_wheeler <- ppois(1, rate \* time)

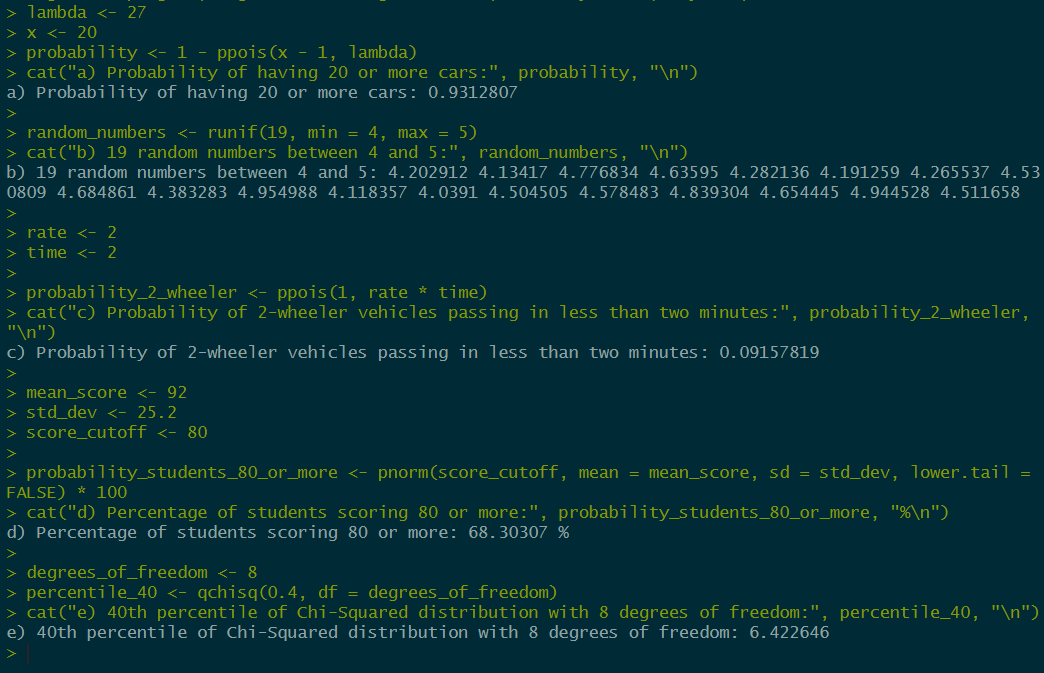
cat("c) Probability of 2-wheeler vehicles passing in less than two minutes:", probability\_2\_wheeler, "\n")

mean\_score <- 92

std\_dev <- 25.2

score\_cutoff <- 80

probability\_students\_80\_or\_more <- pnorm(score\_cutoff, mean = mean\_score, sd = std\_dev, lower.tail = FALSE) \* 100



*10. Measures of Tendency and Dispersion:*

x <- c(300,310,290,320,330,300,310,330,340,320)

range(x)

sd(x)

var(x)

install.packages("moments")

library(moments)

reaction\_times <- c(300, 310, 290, 320, 330, 300, 310, 330, 340, 320)

range\_reaction <- max(reaction\_times) - min(reaction\_times)

std\_dev\_reaction <- sd(reaction\_times)

var\_reaction <- var(reaction\_times)

print(paste("Reaction Times: Range =", range\_reaction,

"Standard Deviation =", std\_dev\_reaction,

"Variance =", var\_reaction))

completion\_times <- c(15, 18, 20, 25, 22, 17, 21, 19, 24, 23)

mean\_completion <- mean(completion\_times)

median\_completion <- median(completion\_times)

mode\_completion <- as.numeric(names(which.max(table(completion\_times))))

skewness\_completion <- skewness(completion\_times)

kurtosis\_completion <- kurtosis(completion\_times)

print(paste("Completion Times: Mean =", mean\_completion,

"Median =", median\_completion,

"Mode =", mode\_completion,

"Skewness =", skewness\_completion,

"Kurtosis =", kurtosis\_completion))

