|  |  |  |
| --- | --- | --- |
|  | **SAVEETHA SCHOOL OF ENGINEERING**  **SAVEETHA INSTITUTE OF MEDICAL AND TECHNICAL SCIENCES**  **INSTITUTE OF COMPUTER SCIENCE AND ENGINEERING**  **ITA 04 - STATISTICS WITH R PROGRAMMING MARCH 2023**  **ASSIGNMENT 5** |  |

1. i) Describe how histogram charts are created in R. Create a histogram chart for the below given age attribute.

Age : 5,45,23,30,33,32,34,35,42,41,28,29

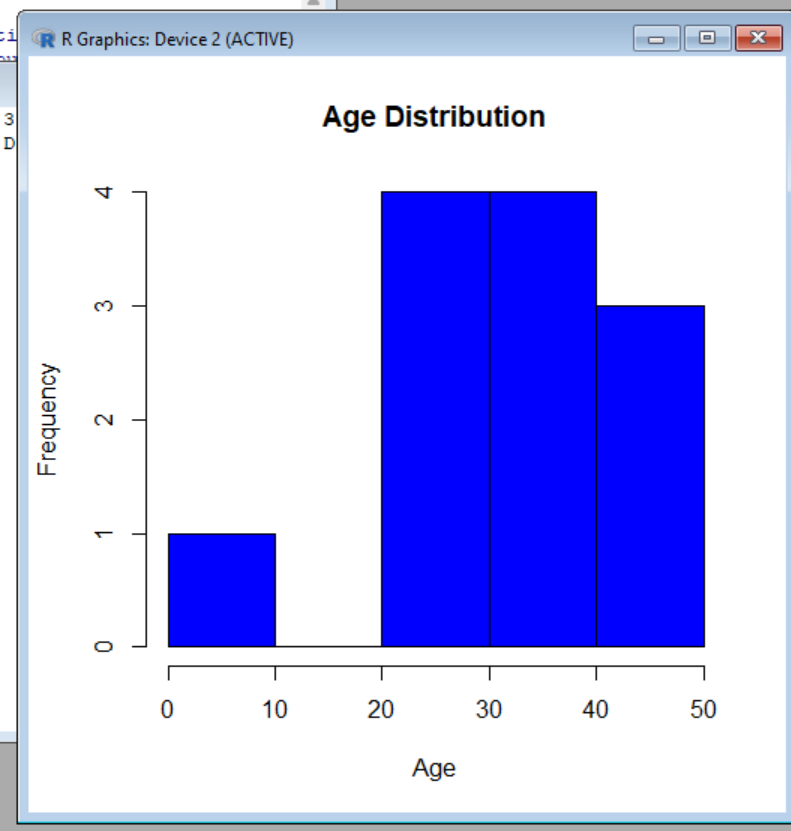
# A.

SOURCE CODE:

#

age <- c(5,45,23,30,33,32,34,35,42,41,28,29)

hist(age, main="Age Distribution", xlab="Age", ylab="Frequency", col="blue")



ii) Create a 3D Pie Chart for the dataset “political Knowledge” with suitable labels and colour.

SOURCE CODE:

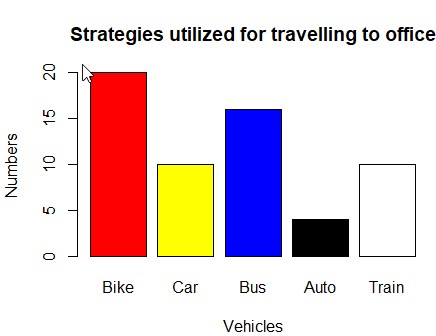
library(plotrix)

political\_knowledge <- c(25, 30, 10, 35)

# Create 3D pie chart

pie3D(political\_knowledge, labels=c("Very Low", "Low", "Medium", "High"), col=c("red", "yellow", "green", "blue"), explode=0.1)

1. Write R code for the below output Figure 1 shows Bike is assigned red , car is assigned yellow , bus is assigned blue , auto is assigned black , and train is assigned white. Mention the parameters used in the below barchart.



# A.

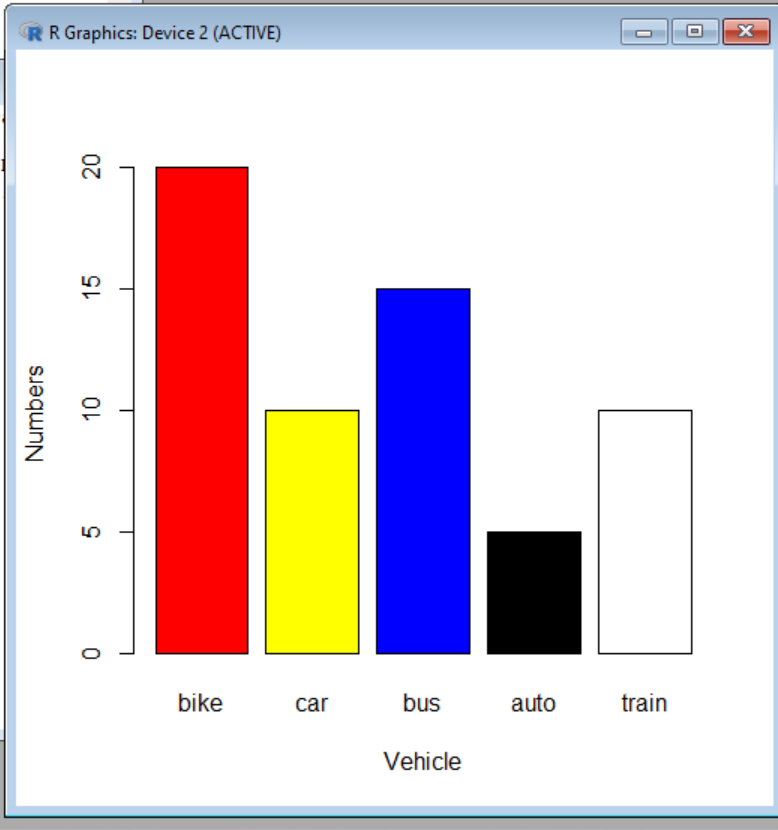
SOURCE CODE:

x <- c("bike","car","bus","auto","train")

> y <- c(20,10,15,5,10)

> colurs <- c("red","yellow","blue","black","white")

> barplot(y,names.arg=x,xlab="Vehicle",ylab="Numbers",col=colurs)



1. Create a Boxplot graph for the relation between "mpg"(miles per galloon) and "cyl(number of Cylinders) for the dataset "mtcars" available in R Environment. v)Write R program to find the sum of Natural Numbers using Recursion

# A.

SOURCE CODE:

# Load the mtcars dataset

data(mtcars)

# Create boxplot graph for mpg and cyl

boxplot(mpg ~ cyl, data = mtcars, main = "Miles per Gallon vs Number of Cylinders", xlab = "Number of Cylinders", ylab = "Miles per Gallon")

# Function to calculate the sum of natural numbers

sum\_natural\_numbers <- function(n) {

if(n == 1) {

return(1)

} else {

return(n + sum\_natural\_numbers(n-1))

}

}

# Test the function

sum\_natural\_numbers(5)

1. a. Melt ‘airquality’ data set which inbuild dataset in ‘R’ and display as a long – format data?

# A.

# Load the "reshape2" package

library(reshape2)

# Melt the "airquality" dataset

airquality\_melted <- melt(airquality)

# Print the melted dataset

print(airquality\_melted)

* 1. Melt air quality data and specify month and day to be “ID variables”?

# A.

# Melt the "airquality" dataset and specify "month" and "day" as ID variables

airquality\_melted <- melt(airquality, id.vars = c("Month", "Day"))

# Print the melted dataset

print(airquality\_melted)

* 1. Cast the molten ‘airquality’ data set.

# A.

# Cast the molten "airquality" dataset

airquality\_casted <- dcast(airquality\_melted, Month + Day ~ variable)

# Print the casted dataset

print(airquality\_casted)

* 1. Use cast function appropriately and compute the average of Ozone, Solar, Wind and temperature per month?

# A.

# Load the "reshape2" package

library(reshape2)

# Cast the "airquality" dataset

airquality\_casted <- cast(airquality, Month ~ variable, mean)

# Select only the relevant columns

airquality\_casted <- airquality\_casted[, c("Month", "Ozone", "Solar.R", "Wind", "Temp")]

# Print the casted dataset

print(airquality\_casted)

* 1. Create a boxplot for ozone reading of ‘airquality’ dataset. Add title, label and color.

# A.

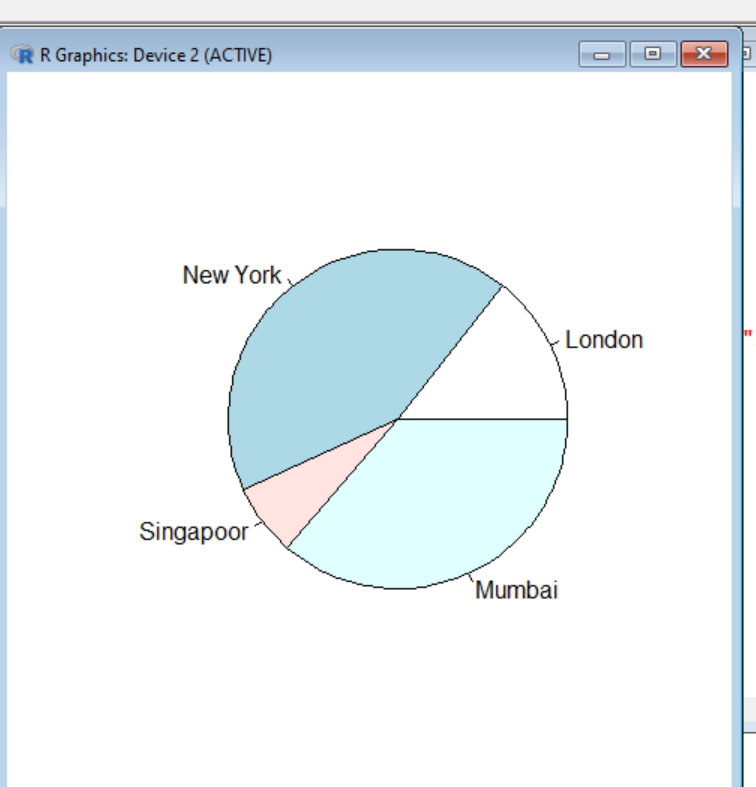
boxplot(airquality$Ozone, main = "Ozone Reading of Air Quality Dataset", xlab = "Ozone Reading", col = "blue")

1. a. Write a program for creating a pie-chart in R using the input vector (21,62,10,53). Provide labels for the chart as ‘London’, ‘New York’, ‘Singapore’, ‘Mumbai’. Add a title to the chart as ‘city pie-chart’ and add a legend at the top right corner of the chart.

R <- c(21,62,10,53)

la <- c("London","New York","Singapoor","Mumbai")

pie(R,la)



* 1. Using linear regression analysis establish a relationship between height and weight of a person using the input vector given below.

# Values of height

151, 174, 138, 186, 128, 136, 179, 163, 152, 131 # Values of weight.

63, 81, 56, 91, 47, 57, 76, 72, 62, 48 Predict the weight of a person with height 170.

x <- c(151, 174, 138, 186, 128, 136, 179, 163, 152, 131 )

y <- c(63, 81, 56, 91, 47, 57, 76, 72, 62,48)

r <- lm(y~x)

a <- data.frame(x = 170)

print(predict(r,a))

c. Visualize the regression graphically.

d. Call ‘mtcars’ which is built in dataset in ‘R’ and plot distribution of mpg feature.

Make x axis range from 10 to 35 and plot title as “More trends in 70’s Vehicles”. e. Find statistical summary of the ‘mtcars’ dataset.

1. There is a popular built-in data set in R called "mtcars" (Motor Trend Car Road Tests), which is retrieved from the 1974 Motor Trend US Magazine.

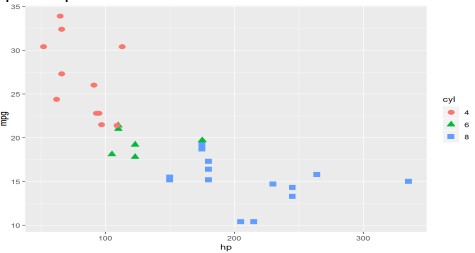
(i)Find the dimension of the data set & Give the statistical summary of the features. (ii)Create correlation matrix between mpg vs all other features and print the high 3 correlated Features(both +ve and -ve)

(iii)Plot the Box plot for “mpg” group by “cyl” feature.

(iv)Create a scatter plot graph for the relation between "mpg"(miles per gallon) and

"hp"(horse power) group by cyl(number ofcylinder)

**Sample Output:**



(v) Generate a multiple regression model to establish the relationship between "mpg" as a response variable with "disp","hp" and "wt" as predictor variables. Plot the regression line. Find the MSE of the model.

1. (i) Use melt and cast function to find mean of numeric data in dataset based on Species group.

A <- c(1,2,3,4,2,3,4,1)

B <- c(1,2,3,4,2,3,4,1)

a <- c(10,20,30,40,50,60,70,80)

b <- c(100,200,300,400,500,600,700,800)

data <- data.frame(A,B,a,b)

print("Original data frame:\n")

print(data)

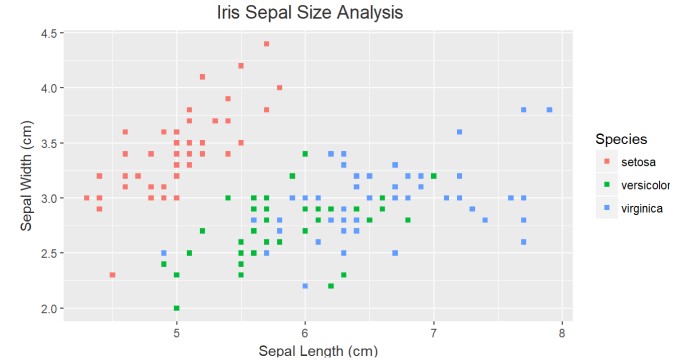
melt\_data <- melt(data, id = c("A","B"))

print("Reshaped data frame:\n")

print(melt\_data)

* 1. Generate a suitable plot which summaries statistical parameter of Sepal.Width based on Species group
  2. Generate scatter plot between Sepal.Length vs Sepal.Width grouped by Specias.

**Sample Output:**



1. A) Heights(in cm) of father and son are given as follows

Father(X): 150 152 155 157 160 161 164 165 Son (Y) : 154 156 158 159 160 162 161 164

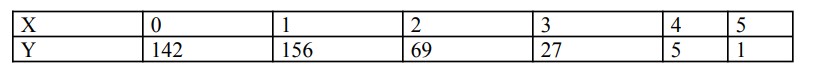
Fit a regression line parameters to predict the height of son given the height of father. Write R code for same.

x <- c(150 ,152 ,155 ,157 ,160 ,161, 164 ,165 )

y <- c(154 ,156, 158 ,159 ,160 ,162 ,161, 164 )

r <- lm(y~x)

B) Fit a regression line parameters distribution with the following data



x <- c(0,1,2,3,4,5)

y <- c(142,156,69,27,5,1)

l <- lm(y~x)