RWorksheet_Freires#4b

2024-10-29

Using Loop Function

for() loop

1. Using the for loop, create an R script that will display a 5x5 matrix as shown in Figure 1. It must contain vectorA = [1,2,3,4,5] and a 5 x 5 zero matrix. Hint Use abs() function to get the absolute value

```r
vectorA <- c(1, 2, 3, 4, 5)
zero_matrix <- matrix(0, nrow = 5, ncol = 5)

for (i in 1:5) {
  for (j in 1:5) {
  zero_matrix[i, j] <- abs(vectorA[i] - vectorA[j])
  }
}
print(zero_matrix)
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    0    1    2    3    4
## [2,]    1    0    1    2    3
## [3,]    2    1    0    1    2
## [4,]    3    2    1    0    1
## [5,]    4    3    2    1    0
```

2. Print the string "*" using for() function. The output should be the same as shown in Figure

```r
for (i in 1:5) {
  cat(rep('"*"', i), "\n")
}
```

```
## "*"
## "*" "*"
## "*" "*" "*"
## "*" "*" "*" "*"
## "*" "*" "*" "*" "*"
```

3. Get an input from the user to print the Fibonacci sequence starting from the 1st input up to 500. Use repeat and break statements. Write the R Scripts and its output.

```r
x <- 0
y <- 1

num <- readline(prompt = "Enter the starting number: ")
```

```
## Enter the starting number:
```

```r
3
```

```
## [1] 3
```

```
repeat {
  num <- x + y
  if (num > 500) break
  x <- y
  y <- num
  print(num)
}
```

```
## [1] 1
## [1] 2
## [1] 3
## [1] 5
## [1] 8
## [1] 13
## [1] 21
## [1] 34
## [1] 55
## [1] 89
## [1] 144
## [1] 233
## [1] 377
```

Using Basic Graphics (plot(),barplot(),pie(),hist())

4. Import the dataset as shown in Figure 1 you have created previously.

a. What is the R script for importing an excel or a csv file? Display the first 6 rows of the dataset? Show your codes and its result

```
library(readxl)
data_table <- read_excel("/cloud/project/Worksheet#4/data_table.xlsx")
print(head(data_table))
```

```
## # A tibble: 6 x 3
##   shoe_size height gender
##       <dbl>  <dbl> <chr>
## 1       6.5   66   F
## 2       9     68   F
## 3       8.5   64.5 F
## 4       8.5   65   F
## 5      10.5   70   M
## 6       7     64   F
```

b. Create a subset for gender(female and male). How many observations are there in Male? How about in Female? Write the R scripts and its output.

```
males <- subset(data_table)
females <- subset(data_table)

n_males <- nrow(males)
n_females <- nrow(females)

cat("Number of Male observations: ", n_males, "\n")
```

```
## Number of Male observations:  28
```

```
cat("Number of Female observations: ", n_females, "\n")
```

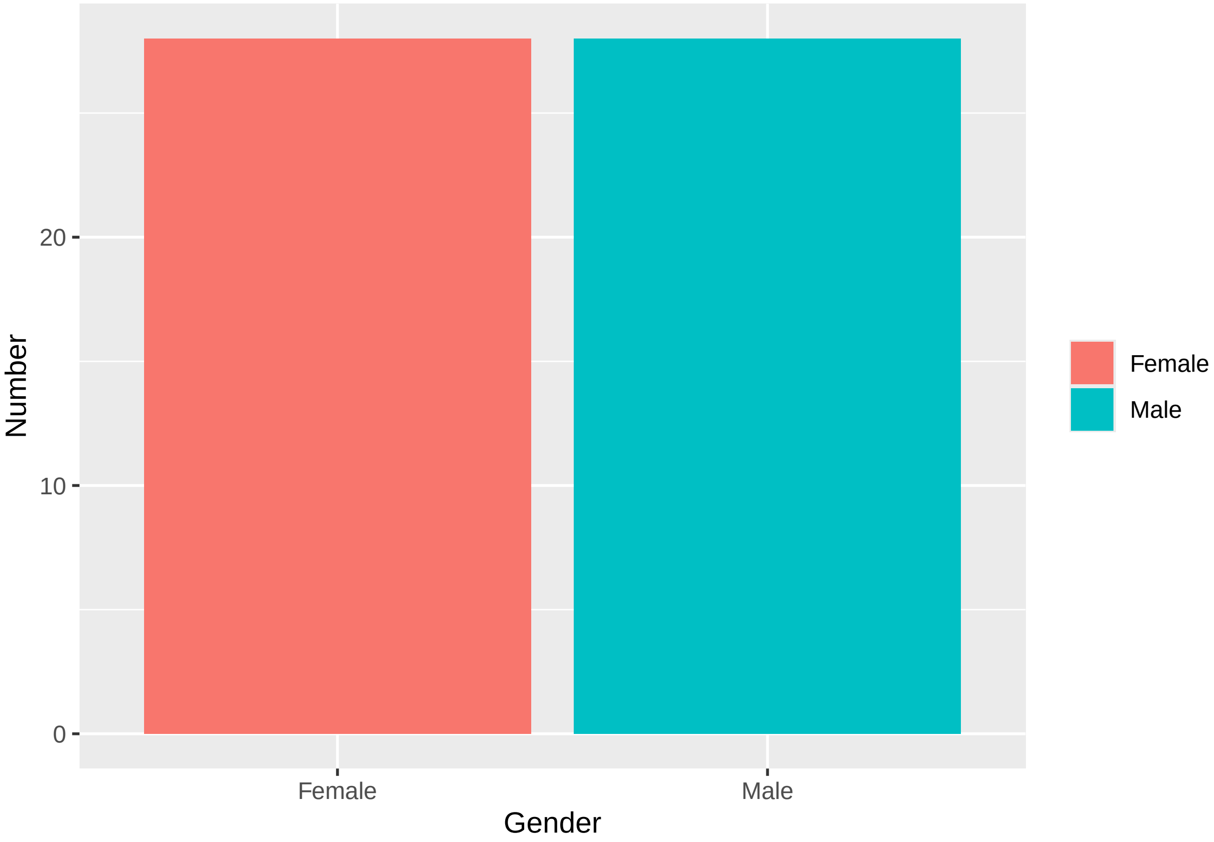
```
## Number of Female observations:  28
```

c. Create a graph for the number of males and females for Household Data. Use plot(), chart type = barplot. Make sure to place title, legends, and colors. Write the R scripts and its result.

```
library(ggplot2)


  Gender = c("Male", "Female")
  Number = c(28, 28)
  data_table <- data.frame(Gender, Number)

ggplot(data_table, aes(x = Gender, y = Number, fill = Gender)) +
  geom_bar(stat = "identity") +
  theme(legend.title = element_blank())
```



5. The monthly income of Dela Cruz family was spent on the following: Food Electricity Savings Miscellaneous 60 10 5 25

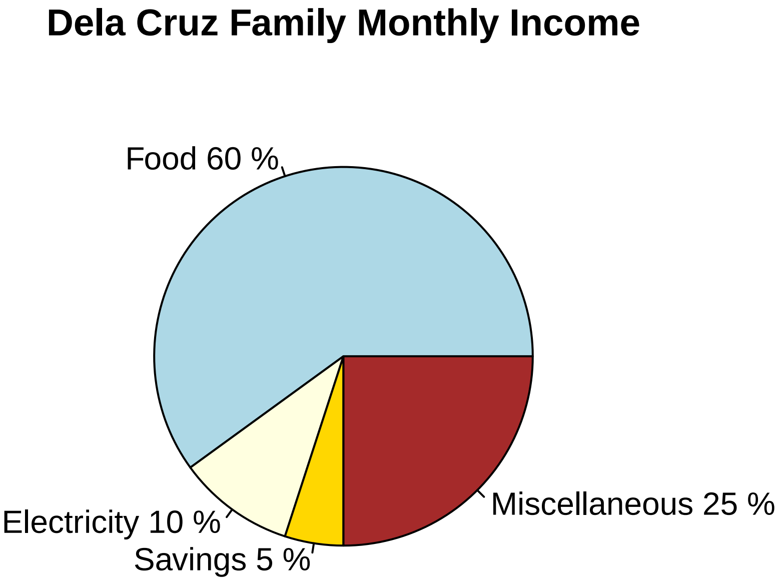
a. Create a piechart that will include labels in percentage.Add some colors and title of the chart. Write the R scripts and show its output.

```
library(ggplot2)

   bills <- c(60, 10, 5, 25)
   categories <- c("Food", "Electricity", "Savings", "Miscellaneous")

   percentages <- round(bills / sum(bills) * 100, 1)
   labels <- paste(categories, percentages, "%")
```

```r
pie(
  bills,
  labels = labels,
  col = c("lightblue", "lightyellow", "gold", "brown"),
  main = "Dela Cruz Family Monthly Income"
)
```



6. Use the iris dataset. data(iris)

a. Check for the structure of the dataset using the str() function.

- Describe what you have seen in the output.
- Based on my observations, the iris data set is a data frame that has 5 variables and 150 obs. The following variables are Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, and Species with 3 Factor Levels5

```r
data(iris)
str(iris)
```

```
## 'data.frame':    150 obs. of  5 variables:
##  $ Sepal.Length: num  5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...
##  $ Sepal.Width : num  3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...
##  $ Petal.Length: num  1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...
##  $ Petal.Width : num  0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...
##  $ Species     : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...
```

b. Create an R object that will contain the mean of the sepal.length, sepal.width,petal.length,and petal.width. What is the R script and its result?

```r
data(iris)
value <- colMeans(iris[, 1:4])
print(value)
```

```
## Sepal.Length  Sepal.Width Petal.Length  Petal.Width
##     5.843333     3.057333     3.758000     1.199333
```

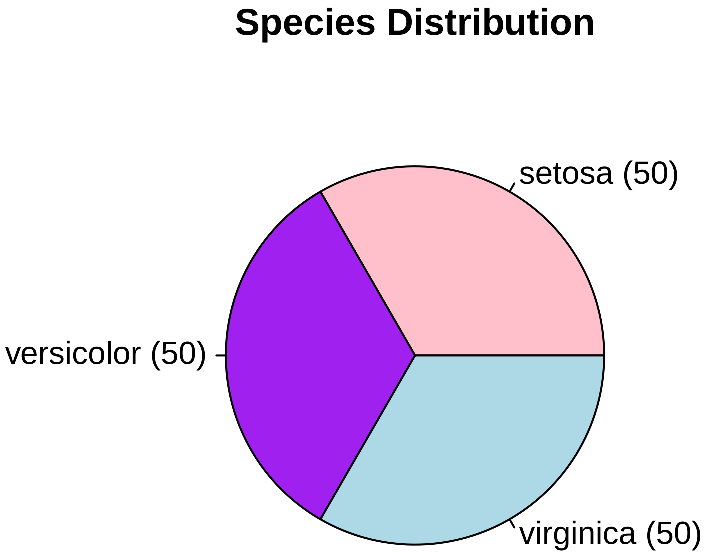
c. Create a pie chart for the Species distribution. Add title, legends, and colors. Write the R script and its result.

```r
data(iris)

species_data <- table(iris$Species)

labels <- paste(names(species_data), species_data, sep = " (")
labels <- paste(labels, ")", sep = "")

pie(
  species_data,
  labels = labels,
  col = c("pink", "purple", "lightblue"),
  main = "Species Distribution"
)
```



d. Subset the species into setosa, versicolor, and virginica. Write the R scripts and show the last six (6) rows of each species.

```r
setosa_sub <- subset(iris, Species == "setosa")
versicolor_sub <- subset(iris, Species == "versicolor")
virginica_sub <- subset(iris, Species == "virginica")

print(tail(setosa_sub))
```

```
##    Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 45          5.1         3.8          1.9         0.4  setosa
## 46          4.8         3.0          1.4         0.3  setosa
## 47          5.1         3.8          1.6         0.2  setosa
## 48          4.6         3.2          1.4         0.2  setosa
## 49          5.3         3.7          1.5         0.2  setosa
## 50          5.0         3.3          1.4         0.2  setosa
```

```r
print(tail(versicolor_sub))
```

```
##     Sepal.Length Sepal.Width Petal.Length Petal.Width    Species
## 95           5.6         2.7          4.2         1.3 versicolor
## 96           5.7         3.0          4.2         1.2 versicolor
## 97           5.7         2.9          4.2         1.3 versicolor
## 98            6.2         2.9          4.3         1.3 versicolor
## 99            5.1         2.5          3.0         1.1 versicolor
## 100           5.7         2.8          4.1         1.3 versicolor
```

```
print(tail(virginica_sub))
```

```
##     Sepal.Length Sepal.Width Petal.Length Petal.Width   Species
## 145          6.7         3.3          5.7         2.5 virginica
## 146          6.7         3.0          5.2         2.3 virginica
## 147          6.3         2.5          5.0         1.9 virginica
## 148          6.5         3.0          5.2         2.0 virginica
## 149          6.2         3.4          5.4         2.3 virginica
## 150          5.9         3.0          5.1         1.8 virginica
```

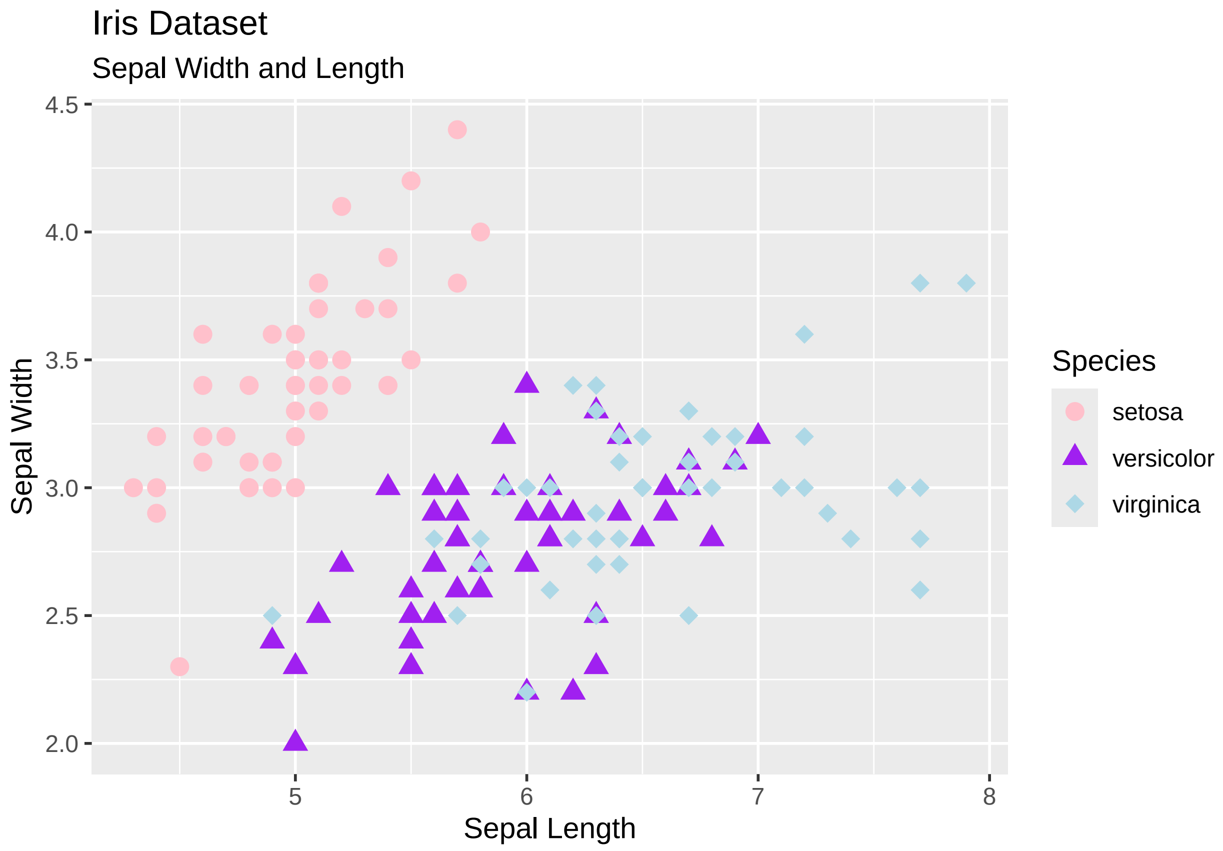
e. Create a scatterplot of the sepal.length and sepal.width using the different species(setosa,versicolor,virginica). Add a title = “Iris Dataset”, subtitle = “Sepal width and length, labels for the x and y axis, the pch symbol and colors should be based on the species.

```
library(ggplot2)
data(iris)

iris$Species <- as.factor(iris$Species)

scatter_plot <- ggplot(iris, aes(x = Sepal.Length, y = Sepal.Width, color = Species, shape = Species)) +
  ggtitle("Iris Dataset") +
  labs(subtitle = "Sepal Width and Length", x = "Sepal Length", y = "Sepal Width") +
  geom_point(size = 3) +
  scale_color_manual(values = c("setosa" = "pink", "versicolor" = "purple", "virginica" = "lightblue"))
  scale_shape_manual(values = c(16, 17, 18))


print(scatter_plot)
```



Hint: Need to convert to factors the species to store categorical variables.

f. Interpret the result.

- The results show the Sepal Width and Length of each species, The setosa has the most width than length , the versicolor has more length than width, and the virginica has the most length than width