RWorksheet_Magallanes#4b

Killy Magallanes

2024-10-28

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.

```
vectorA <- c(1, 2, 3, 4, 5)
matrix_result <- matrix(0, nrow = 5, ncol = 5)
for (i in 1:5) {
  for (j in 1:5) {
    matrix_result[i, j] <- abs(vectorA[i] - vectorA[j])
  }
}
print(matrix_result)
```

```
##      [,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

```
num_lines <- 5

for (i in 1:num_lines) {
  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.

```
fibonacci_sequence <- function(start) {
  a <- 0
  b <- 1

  repeat {
    fib <- a + b
    a <- b
    b <- fib
    if (fib > 500) {
      break
    }
    if (fib >= start) {
      cat(fib, "\n")
    }
  }
}
user_input <- as.numeric(readline(prompt = "Enter a starting number for Fibonacci sequence: "))
```

```
## Enter a starting number for Fibonacci sequence:
```

```
if (!is.na(user_input) && user_input >= 0) {
  cat("Fibonacci sequence starting from", user_input, "up to 500:\n")
  fibonacci_sequence(user_input)
} else {
  cat("Please enter a valid non-negative number.\n")
}
```

```
## Please enter a valid non-negative number.
```

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

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.

```
female_subset <- subset(ShoeSizes, Gender == "F")
male_subset <- subset(ShoeSizes, Gender == "M")
num_female <- nrow(female_subset)
num_male <- nrow(male_subset)
cat("Number of Female observations:", num_female, "\n")
```

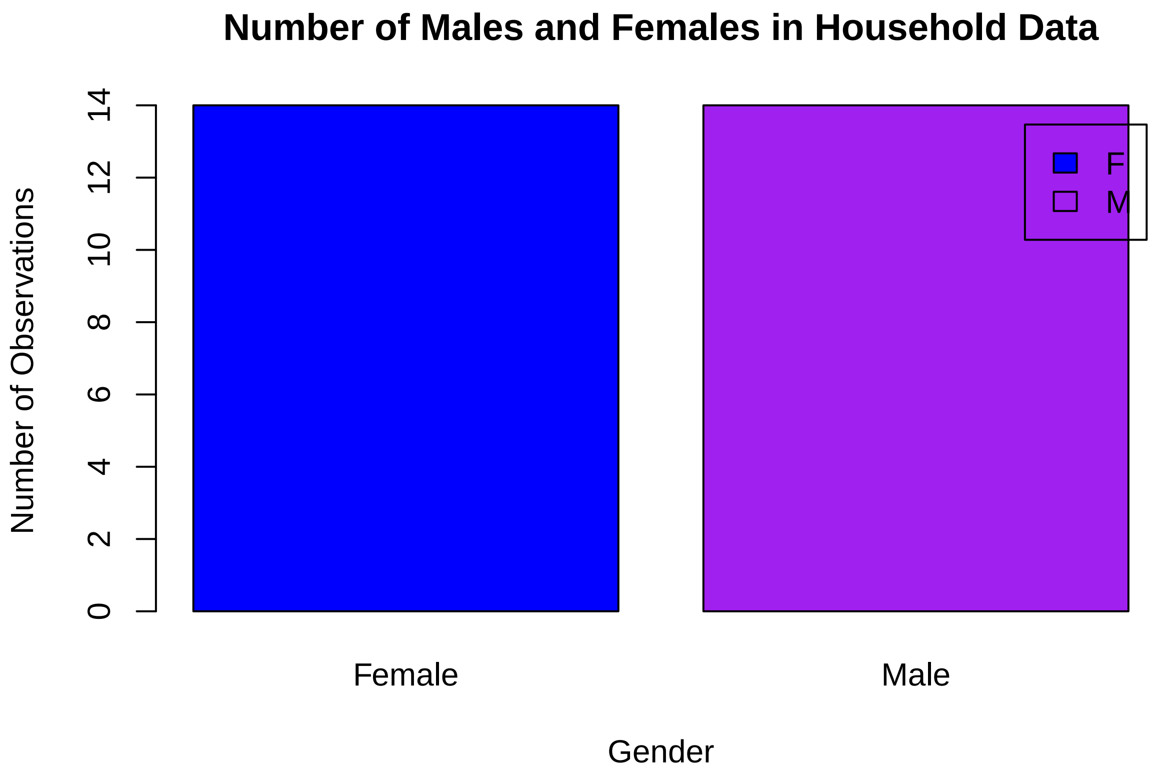
```
## Number of Female observations: 14
```

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

```
## Number of Male observations: 14
```

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.

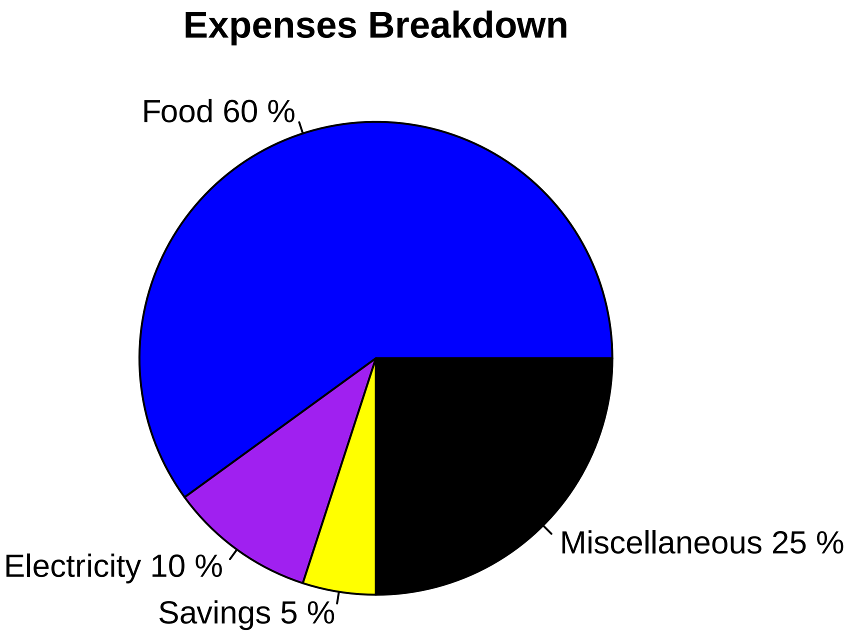
```
gender_count <- table(ShoeSizes$Gender)
bar_colors <- c("blue", "purple")
barplot(gender_count,
        main = "Number of Males and Females in Household Data",
        xlab = "Gender",
        ylab = "Number of Observations",
        col = bar_colors,
        names.arg = c("Female", "Male"),
        legend.text = TRUE)
```



5.

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.

```
categories <- c("Food", "Electricity", "Savings", "Miscellaneous")
values <- c(60, 10, 5, 25)
percentages <- round(values / sum(values) * 100)
pie_colors <- c("blue", "purple", "yellow", "black")
pie(percentages,
    labels = paste(categories, percentages, "%"),
    col = pie_colors,
    main = "Expenses Breakdown",
    radius = 1)
```



6. Use the iris dataset a. a. Check for the structure of the dataset using the str() function. Describe what you have seen in the output.

```
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 ...
```

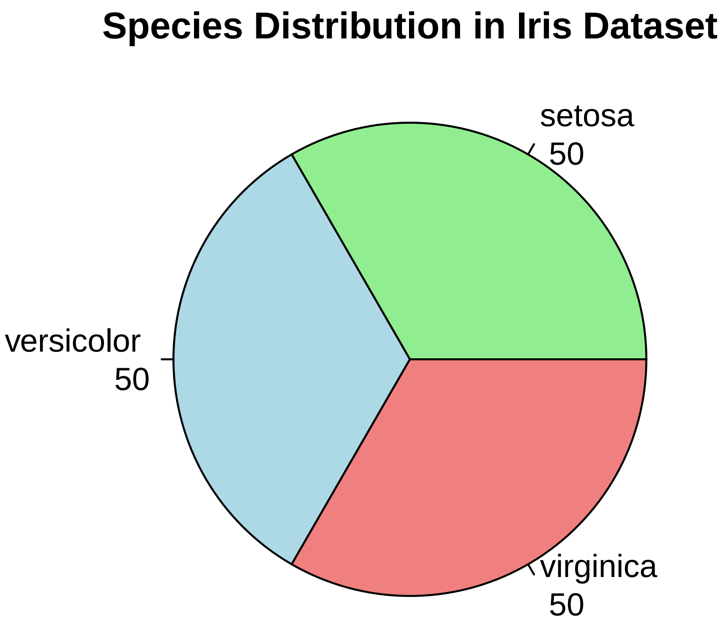
b. Create an R object that will contain the mean of the sepal.length, sepal.width,petal.length,and petal.width. What is the R script and its result?

```
data(iris)
means <- colMeans(iris[, 1:4])
means
```

```
## 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.

```
data(iris)
species_count <- table(iris$Species)
pie_colors <- c("lightgreen", "lightblue", "lightcoral")
pie(species_count,
    labels = paste(names(species_count), "\n", species_count),
    col = pie_colors,
    main = "Species Distribution in Iris Dataset",
    radius = 1)
```



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

```
data(iris)

setosa <- subset(iris, Species == "setosa")
versicolor <- subset(iris, Species == "versicolor")
virginica <- subset(iris, Species == "virginica")

last_six_setosa <- tail(setosa, 6)
last_six_versicolor <- tail(versicolor, 6)
last_six_virginica <- tail(virginica, 6)

cat("Last six rows of Setosa:\n")
```

```
## Last six rows of Setosa:
```

```r
print(last_six_setosa)
```

```
##    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
cat("\nLast six rows of Versicolor:\n")
```

```
##
## Last six rows of Versicolor:
```

```r
print(last_six_versicolor)
```

```
##     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
```

```r
cat("\nLast six rows of Virginica:\n")
```

```
##
## Last six rows of Virginica:
```

```r
print(last_six_virginica)
```

```
##     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.

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

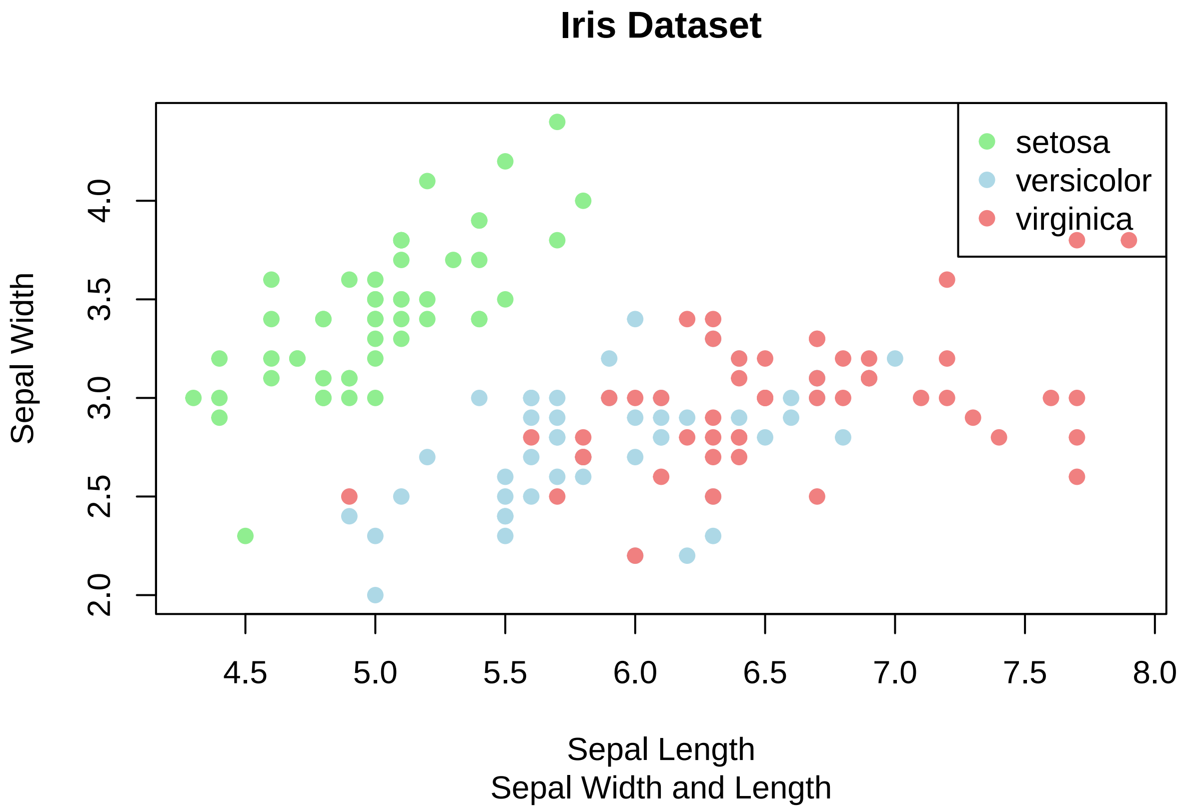
```
data(iris)

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

species_colors <- c("setosa" = "lightgreen", "versicolor" = "lightblue", "virginica" = "lightcoral")


plot(iris$Sepal.Length, iris$Sepal.Width,
     col = species_colors[iris$Species],
     pch = 19,
     main = "Iris Dataset",
     sub = "Sepal Width and Length",
     xlab = "Sepal Length",
     ylab = "Sepal Width")

legend("topright", legend = levels(iris$Species),
       col = species_colors, pch = 19)
```



f. Interpret the results

The scatterplot of Sepal.Length and Sepal.Width from the iris dataset reveals distinct differences among the three species. Setosa (light green) is clearly separated from Versicolor (light blue) and Virginica (light coral), exhibiting smaller sepal dimensions, which allows for easy identification. Versicolor shows a moderate range of sepal lengths and widths, while Virginica occupies the upper right area of the plot, indicating it has the largest sepals. There is some overlap between Versicolor and Virginica, suggesting that they may be more challenging to distinguish based solely on these measurements. Overall, the plot highlights a positive correlation between sepal length and width for Versicolor and Virginica, providing valuable insights for species classification in botanical research and data analysis.

7.

```
Alexafile <- read_excel("C:\\Users\\killy\\Documents\\pushed\\RWorksheet_Magallanes#4b\\alexa_file.xlsx
Alexafile
```

```
## # A tibble: 3,150 x 5
##    rating date                variation          verified_reviews      feedback
##     <dbl> <dttm>              <chr>              <chr>                    <dbl>
##  1      5 2018-07-31 00:00:00 Charcoal Fabric    Love my Echo!                1
##  2      5 2018-07-31 00:00:00 Charcoal Fabric    Loved it!                    1
##  3      4 2018-07-31 00:00:00 Walnut Finish      Sometimes while play~        1
##  4      5 2018-07-31 00:00:00 Charcoal Fabric    I have had a lot of ~        1
##  5      5 2018-07-31 00:00:00 Charcoal Fabric    Music                        1
##  6      5 2018-07-31 00:00:00 Heather Gray Fabric I received the echo ~       1
##  7      3 2018-07-31 00:00:00 Sandstone Fabric   Without having a cel~        1
##  8      5 2018-07-31 00:00:00 Charcoal Fabric    I think this is the ~        1
##  9      5 2018-07-30 00:00:00 Heather Gray Fabric looks great                 1
## 10      5 2018-07-30 00:00:00 Heather Gray Fabric Love it! I've listen~       1
## # i 3,140 more rows
```

a. Rename the white and black variants by using gsub() function.