RWorksheet_Nava#4b

Samantha Nicole Nava

2024-10-28

Using Loop Function 1.

```r
vectorA <- c(1, 2, 3, 4, 5)


matrixResult <- matrix(0, nrow = 5, ncol = 5)

for (i in 1:5) {
  for (j in 1:5) {

    matrixResult[i, j] <- abs(vectorA[i] - vectorA[j])
  }
}


print(matrixResult)
```

```
##      [,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.

```r
numRows <- 5

for (i in 1:numRows) {
  for (j in 1:i) {
    cat('"*"', " ")
  }
  cat("\n")
}
```

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

3.

```r
startNum <- 0
```

```r
if (is.na(startNum) || startNum < 0) {
  stop("Please enter a valid non-negative number.")
}


fib1 <- 0
fib2 <- 1


cat("Fibonacci sequence starting from", startNum, "up to 500:\n")
```

```
## Fibonacci sequence starting from 0 up to 500:
```

```r
repeat {
  if (fib1 >= startNum) {
    cat(fib1, "\n")
  }
  nextFib <- fib1 + fib2
  fib1 <- fib2
  fib2 <- nextFib


  if (fib1 > 500) {
    break
  }
}
```

```
## 0
## 1
## 1
## 2
## 3
## 5
## 8
## 13
## 21
## 34
## 55
## 89
## 144
## 233
## 377
```

Using Basic Graphics (plot(),barplot(),pie(),hist()) 4. a.

```r
library(openxlsx)


dataset <- read.xlsx("household_data2.xlsx", sheet = 1)


print(head(dataset))
```

```
##   Shoe.Size Height Gender
## 1       6.5   66.0      F
## 2       9.0   68.0      F
```

```
## 3        8.5    64.5        F
## 4        8.5    65.0        F
## 5       10.5    70.0        M
## 6        7.0    64.0        F
```

```r
print(str(dataset))
```

```
## 'data.frame':    28 obs. of  3 variables:
##  $ Shoe.Size: num  6.5 9 8.5 8.5 10.5 7 9.5 9 13 7.5 ...
##  $ Height   : num  66 68 64.5 65 70 64 70 71 72 64 ...
##  $ Gender   : chr  "F" "F" "F" "F" ...
## NULL
```

```r
dataset$Gender <- as.character(dataset$Gender)  # Convert to character if needed
```

b.

```r
library(openxlsx)

dataset <- read.xlsx("household_data2.xlsx", sheet = 1)

print(head(dataset))
```

```
##   Shoe.Size Height Gender
## 1       6.5   66.0      F
## 2       9.0   68.0      F
## 3       8.5   64.5      F
## 4       8.5   65.0      F
## 5      10.5   70.0      M
## 6       7.0   64.0      F
```

```r
uniqueValues <- unique(dataset$Gender)
print(uniqueValues)
```

```
## [1] "F" "M"
```

```r
male_subset <- subset(dataset, Gender == "M")
female_subset <- subset(dataset, Gender == "F")

numMales <- nrow(male_subset)
numFemales <- nrow(female_subset)

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

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

```r
cat("Number of observations in Female:", numFemales, "\n")
```

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

5.

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


percentages <- round((expenses / sum(expenses)) * 100)
```
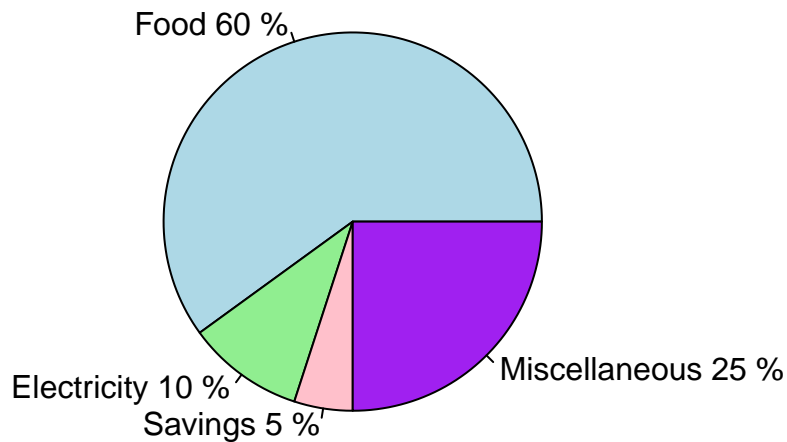
```r
labels <- paste(categories, percentages, "%", sep = " ")


colors <- c("lightblue", "lightgreen", "pink", "purple")

# Create the pie chart
pie(expenses, labels = labels, col = colors, main = "Monthly Income Expenses of Dela Cruz Family")
```

Monthly Income Expenses of Dela Cruz Family

6. a.

```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.

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

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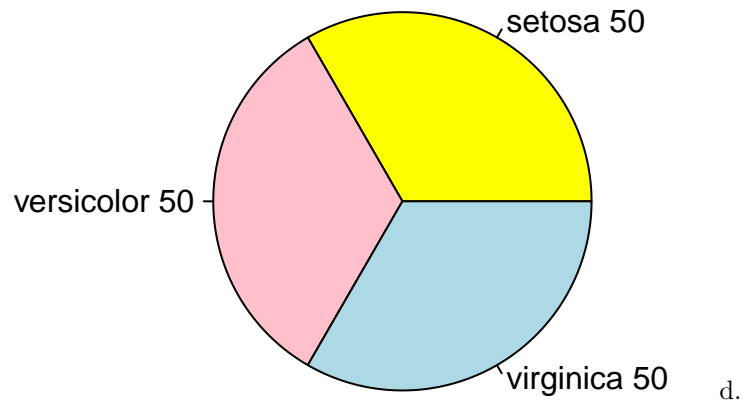
c.

```r
species_count <- table(iris$Species)


labels <- paste(names(species_count), species_count)


colors <- c("yellow", "pink", "lightblue")
```

```r
pie(species_count, labels = labels, col = colors, main = "Species Distribution in Iris Dataset")
```

Species Distribution in Iris Dataset


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


lastSixSetosa <- tail(setosa, 6)
lastSixVersicolor <- tail(versicolor, 6)
lastSixVirginica <- tail(virginica, 6)


lastSixSetosa
```

```
##    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
lastSixVersicolor
```

```
##     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
lastSixVirginica
```

```
##     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
```
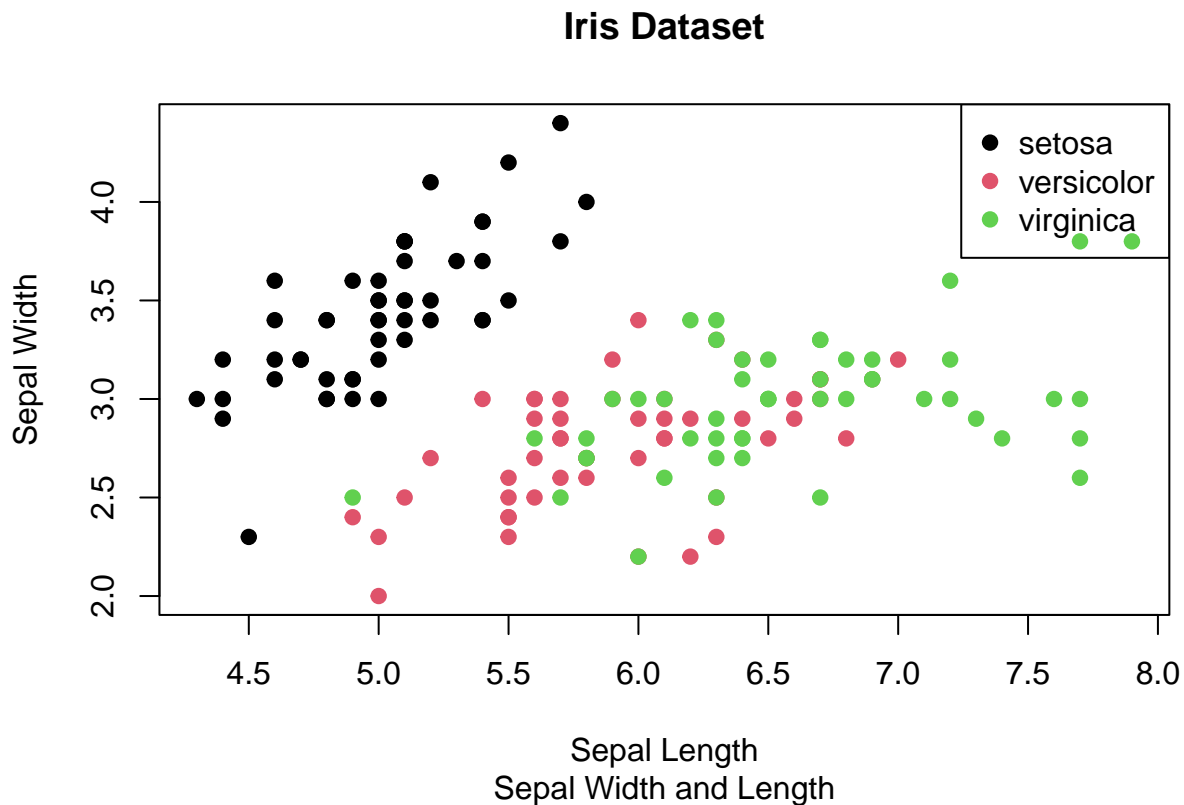
e.

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


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

legend("topright", legend = levels(iris$Species), col = 1:3, pch = 19)
```

Iris Dataset

f.

The scatter plot is a display of the variation within the iris species for changes in sepal length and width. You can see there is clustering - where some species are very close to another, hence similar in measurement. For example, setosa may typically be distinguished from the rest, say versicolor and virginica, with smallersepal length and width. This might also suggest that setosa may be somewhat different from the other two species, whose measures overlap.: The means computed in b give a summary measure for each measurement, providing the mean central tendency of each characteristic for all of the species. In this case, the higher mean values in the versicolor and virginica species for the characteristics Petal.Length and Petal.Width indicate that these species have larger petals compared to setosa.

7.

```
library(openxlsx)
```

```r
alexa_data <- read.xlsx("alexa_file.xlsx")


print("Original Dataset:")

## [1] "Original Dataset:"
print(head(alexa_data))  # Show the first 6 rows

##      Variants Count
## 1  Black Dot    10
## 2 Black Plus    15
## 3 Black Show     5
## 4 Black Spot     8
## 5  White Dot    12
## 6 White Plus    20
alexa_data$Variants <- gsub("^\\s+|\\s+$", "", alexa_data$Variants)  # Trim leading and trailing spaces
alexa_data$Variants <- gsub("\\s+", " ", alexa_data$Variants)  # Replace multiple spaces with a single


alexa_data$Variants <- gsub("Black Dot", "BlackDot", alexa_data$Variants)
alexa_data$Variants <- gsub("Black Plus", "BlackPlus", alexa_data$Variants)
alexa_data$Variants <- gsub("Black Show", "BlackShow", alexa_data$Variants)
alexa_data$Variants <- gsub("Black Spot", "BlackSpot", alexa_data$Variants)
alexa_data$Variants <- gsub("White Dot", "WhiteDot", alexa_data$Variants)
alexa_data$Variants <- gsub("White Plus", "WhitePlus", alexa_data$Variants)
alexa_data$Variants <- gsub("White Show", "WhiteShow", alexa_data$Variants)
alexa_data$Variants <- gsub("White Spot", "WhiteSpot", alexa_data$Variants)

print("Cleaned Dataset Snippet:")

## [1] "Cleaned Dataset Snippet:"
print(head(alexa_data))

##      Variants Count
## 1  BlackDot    10
## 2 BlackPlus    15
## 3 BlackShow     5
## 4 BlackSpot     8
## 5  WhiteDot    12
## 6 WhitePlus    20
write.xlsx(alexa_data, "cleaned_alexa_file.xlsx")
```

b.

```r
library(dplyr)

variationCounts <- alexa_data %>%
  count(Variants)


save(variationCounts, file = "variations.RData")
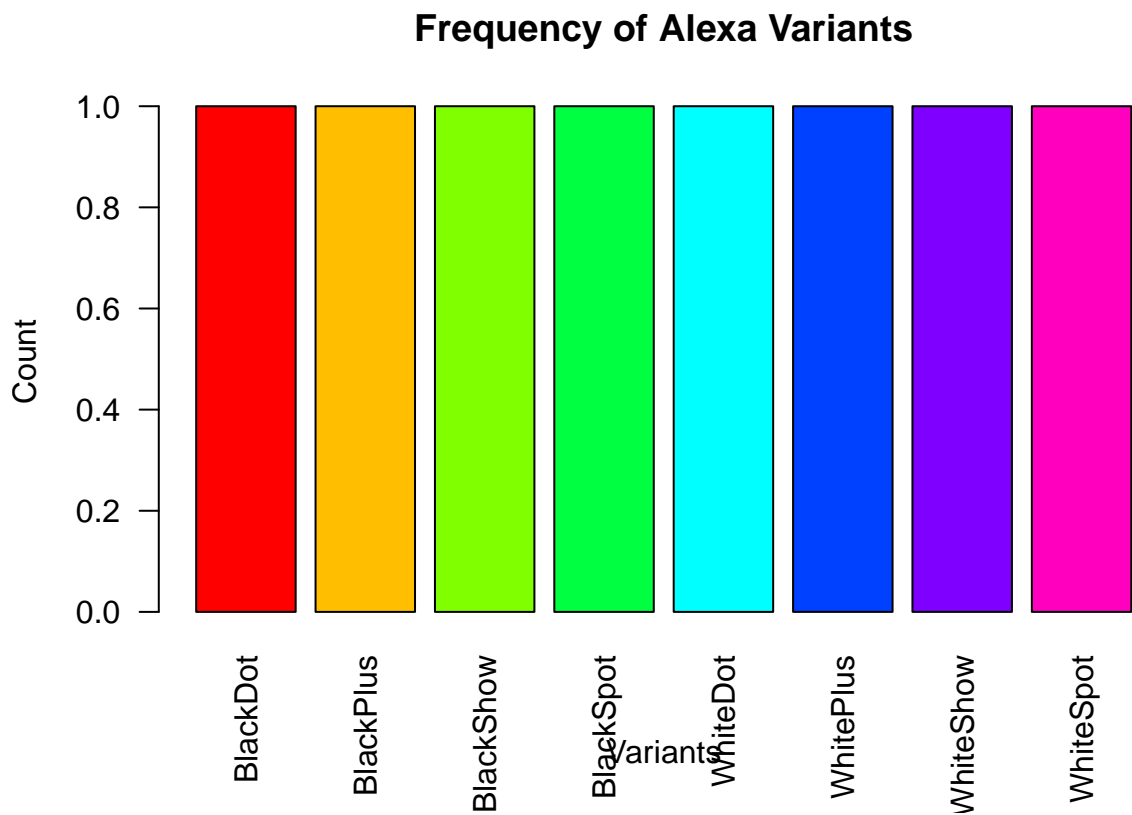```

```
print("Variation Counts:")
```

```
## [1] "Variation Counts:"
```
```
print(variationCounts)
```

```
##    Variants n
## 1  BlackDot 1
## 2 BlackPlus 1
## 3 BlackShow 1
## 4 BlackSpot 1
## 5  WhiteDot 1
## 6 WhitePlus 1
## 7 WhiteShow 1
## 8 WhiteSpot 1
```

c.

```
barplot(variationCounts$n,
        names.arg = variationCounts$Variants,
        col = rainbow(nrow(variationCounts)),
        main = "Frequency of Alexa Variants",
        xlab = "Variants",
        ylab = "Count",
        las = 2)
```
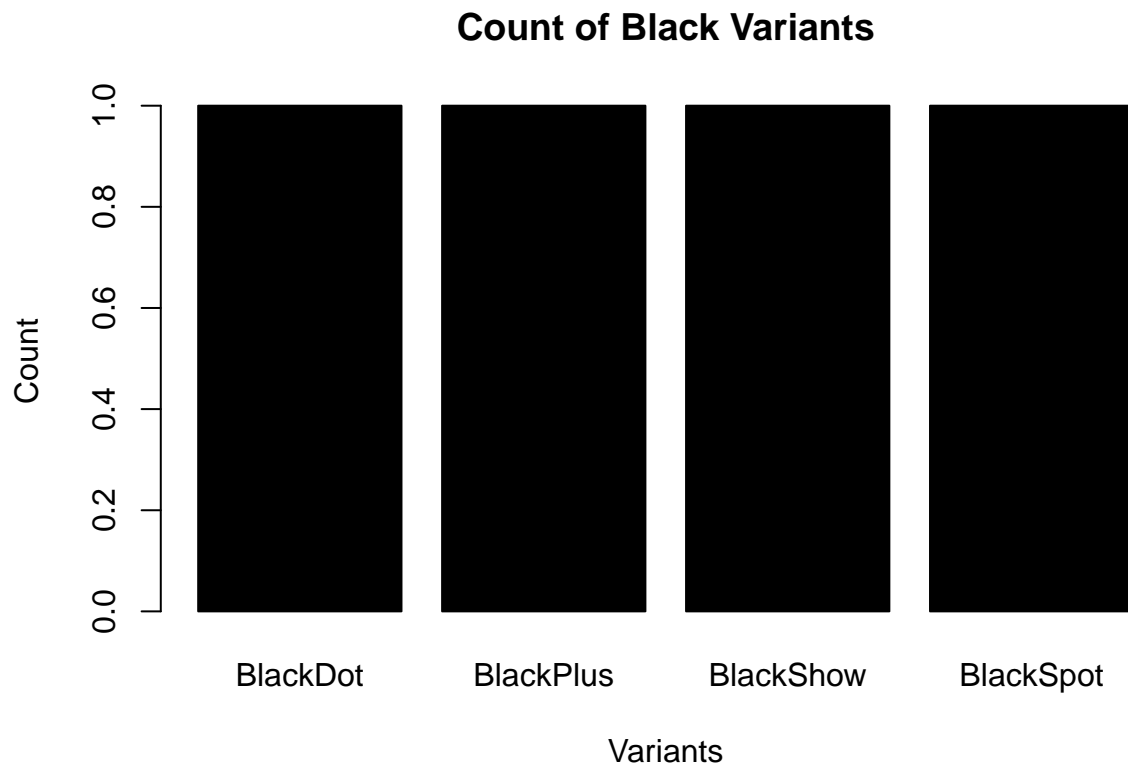

d.

```
# Step 5: Separate counts for black and white variations
blackVariations <- variationCounts[grep("Black", variationCounts$Variants), ]
whiteVariations <- variationCounts[grep("White", variationCounts$Variants), ]
```

```
# Barplot for Black Variants
barplot(blackVariations$n,
        names.arg = blackVariations$Variants,
        col = "black",
        main = "Count of Black Variants",
        xlab = "Variants",
        ylab = "Count")
```



```
# Barplot for White Variants
barplot(whiteVariations$n,
        names.arg = whiteVariations$Variants,
        col = "lightgrey",
        main = "Count of White Variants",
        xlab = "Variants",
        ylab = "Count")
```

Count of White Variants

Count

Variants

WhiteDot   WhitePlus   WhiteShow   WhiteSpot