

RWorksheet#4B_Calopez

Rey Angelo Calopez

2023-11-08

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)

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

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

matrixA

```
##      [,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 2

```
for (i in 1:5) {
  cat(paste0("\n", rep("*", i), "\n"), "\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.

```
# Get user input
n <- as.integer(readline(prompt = "Enter the number of terms: "))
```

```
## Enter the number of terms:
```

```
# Initialize variables
```

```
a <- 0
```

```
b <- 1
```

```
# Print Fibonacci sequence
```

```
cat("Fibonacci sequence:", a, b)
```

```
## Fibonacci sequence: 0 1
```

```
# Generate Fibonacci sequence
```

```
repeat {
  next_term <- a + b
  if (next_term > 500) {
    break
  }
  cat(next_term, " ")
  a <- b
  b <- next_term
}
```

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

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

```
accessData <- read.csv("df_shoes.csv")
head(accessData)
```

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

4b. 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.

```
male_subset <- subset(accessData, Gender == "M")
male_subset
```

```
##      X Shoe_Size Height Gender
## 5  5         10.5   70.0      M
## 9  9         13.0   72.0      M
## 11 11         10.5   74.5      M
## 13 13         12.0   71.0      M
## 14 14         10.5   71.0      M
## 15 15         13.0   77.0      M
## 16 16         11.5   72.0      M
## 19 19         10.0   72.0      M
## 22 22          8.5   67.0      M
## 23 23         10.5   73.0      M
## 25 25         10.5   72.0      M
## 26 26         11.0   70.0      M
## 27 27          9.0   69.0      M
## 28 28         13.0   70.0      M
```

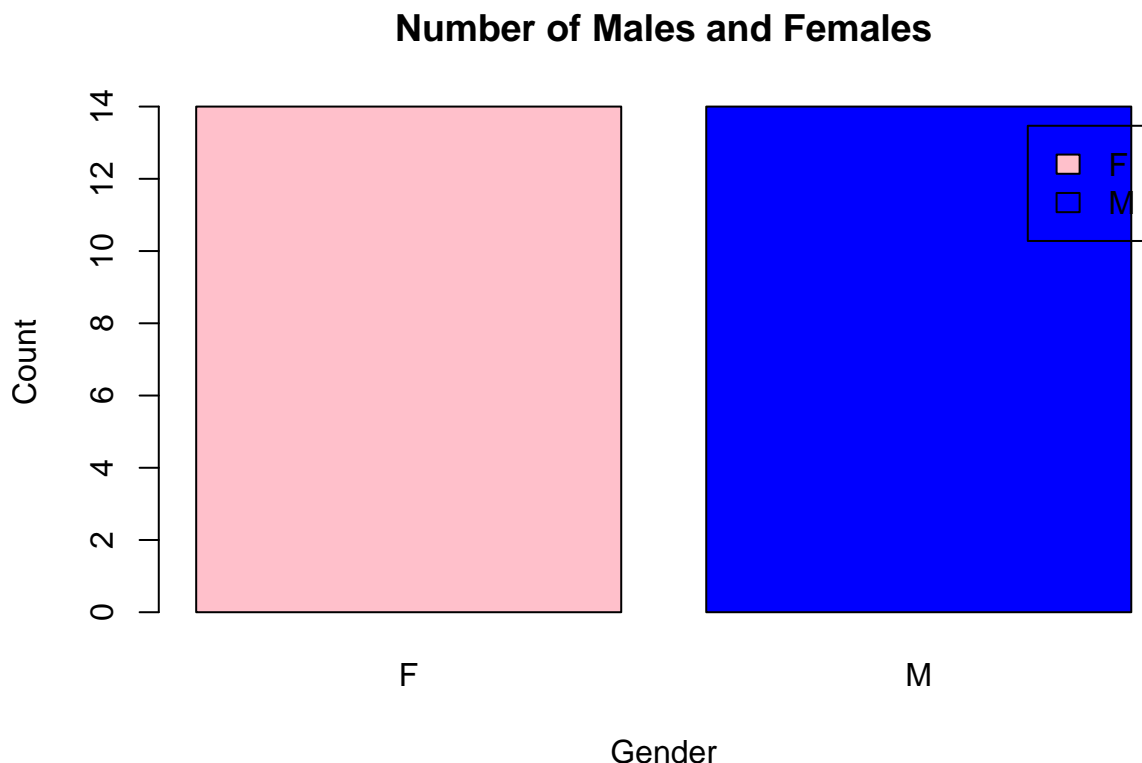
```
female_subset <- subset(accessData, Gender == "F")
female_subset
```

```
##      X Shoe_Size Height Gender
## 1  1         6.5   66.0      F
```

```
## 2 2 9.0 68.0 F
## 3 3 8.5 64.5 F
## 4 4 8.5 65.0 F
## 6 6 7.0 64.0 F
## 7 7 9.5 70.0 F
## 8 8 9.0 71.0 F
## 10 10 7.5 64.0 F
## 12 12 8.5 67.0 F
## 17 17 8.5 59.0 F
## 18 18 5.0 62.0 F
## 20 20 6.5 66.0 F
## 21 21 7.5 64.0 F
## 24 24 8.5 69.0 F
```

4c. 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

```
totalMaleFemale <- table(accessData$Gender)
barplot(totalMaleFemale,
        main = "Number of Males and Females",
        xlab = "Gender",
        ylab = "Count",
        col = c("pink", "blue"),
        legend.text = rownames(totalMaleFemale),
        beside = TRUE)
```



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

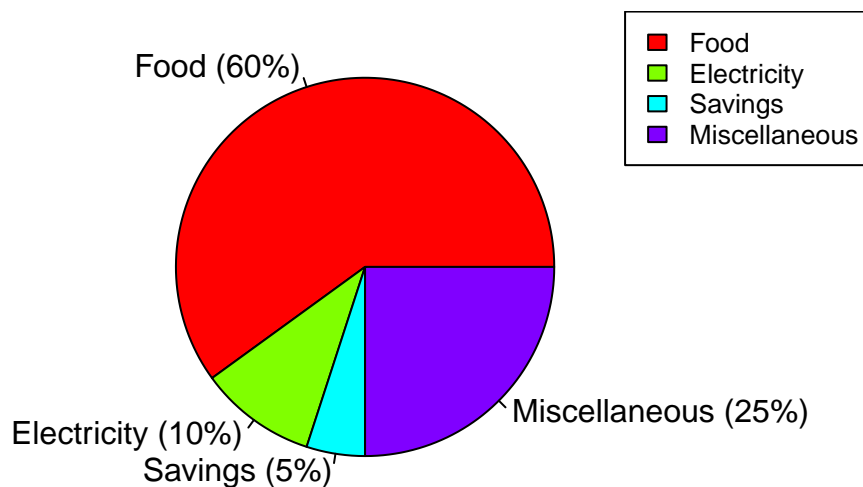
```
# Define the data
data <- c(Food = 60, Electricity = 10, Savings = 5, Miscellaneous = 25)
```

```
# Calculate percentages and format them as strings
percentages <- paste(round(100 * data / sum(data), 1), "%", sep = "")

# Create a pie chart
pie(data, labels = paste(names(data), " (", percentages, ")", sep = ""), col = rainbow(length(data)), margin = 10)

# Add a legend
legend("topright", names(data), cex = 0.8, fill = rainbow(length(data)))
```

Expense Distribution



6. Use the iris dataset

```
data(iris)
```

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

```
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 ...
```

The dataset contains information on iris blossoms. It contains information on the length and width of

6b. 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?

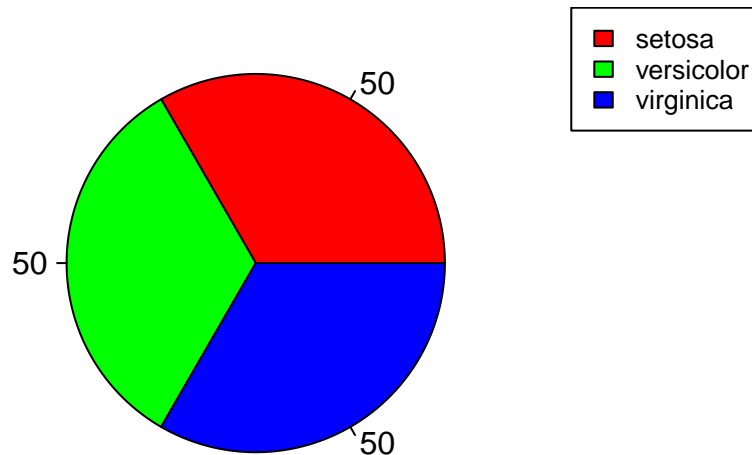
```
meanOfFlowers <- colMeans(iris[,1:4])
meanOfFlowers
```

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

6c. Create a pie chart for the Species distribution. Add title, legends, and colors. Write the R script and its result.

```
species_count <- table(iris$Species)
pie(species_count, labels = species_count, col = rainbow(length(species_count)), main = "Species Distribution",
legend("topright", names(species_count), cex = 0.8, fill = rainbow(length(species_count))))
```

Species Distribution



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

```
# Subset the iris data set into the three species.
setosa_subset <- subset(iris, Species == "setosa")
versicolor_subset <- subset(iris, Species == "versicolor")
virginica_subset <- subset(iris, Species == "virginica")

# Display the last six rows of each species.
tail(setosa_subset, 6)
```

```
##      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
```

```
tail(versicolor_subset, 6)
```

```
##      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
```

```
tail(virginica_subset, 6)
```

```
##      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
```

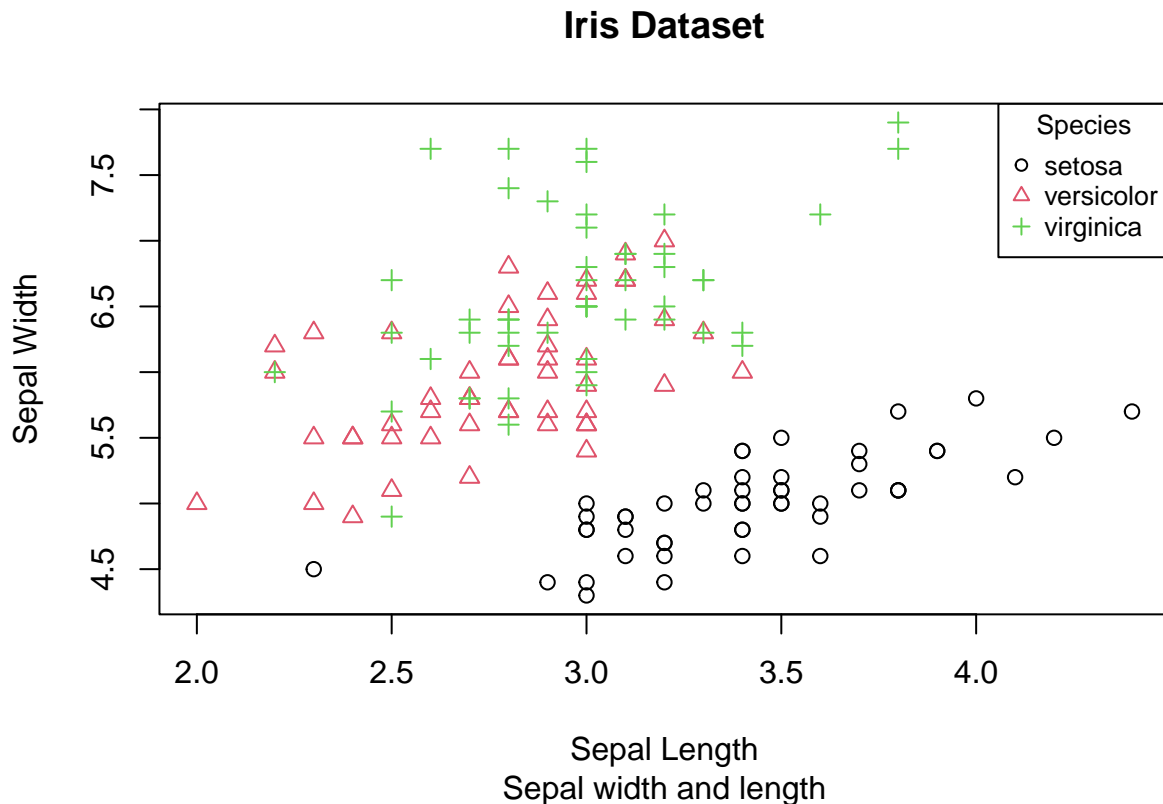
```
## 147      6.3      2.5      5.0      1.9 virginica
## 148      6.5      3.0      5.2      2.0 virginica
## 149      6.2      3.4      5.4      2.3 virginica
## 150      5.9      3.0      5.1      1.8 virginica
```

6e. 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.

```
# Convert the "Species" column to a factor
iris$Species <- as.factor(iris$Species)

# Create a scatterplot
plot(
  Sepal.Length ~ Sepal.Width,
  data = iris,
  pch = as.integer(iris$Species), # Use different pch symbols for each species
  col = as.integer(iris$Species), # Use different colors for each species
  xlab = "Sepal Length",
  ylab = "Sepal Width",
  main = "Iris Dataset",
  sub = "Sepal width and length"
)

# Add a legend
legend("topright", legend = levels(iris$Species), col = 1:3, pch = 1:3, cex = 0.8, title = "Species")
```



6f. Interpret the result.

```
# The dataset consists of five variables (columns) and 150 observations (rows) in a data frame format.
# Petal.Length, Petal.Width, Sepal.Length, and Sepal.Width are the names of the four numerical variables.
# The factor variable Species, which represents the species of iris flowers, is the sixth variable. The
```

7. Import the alexa-file.xlsx. Check on the variations. Notice that there are extra whitespaces among black variants (Black Dot, Black Plus, Black Show, Black Spot). Also on the white variants (White Dot, White Plus, White Show, White Spot).

```
library(readxl)
alexa_file <- read_excel("alexa_file.xlsx")
alexa_file

## # A tibble: 3,150 x 5
##   rating date      variation      verified_reviews      feedback
##   <dbl> <dtm>      <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
```

7a. Rename the white and black variants by using gsub() function.

```
alexa_file$variation <- gsub("Black Dot", "BlackDot", alexa_file$variation)
alexa_file$variation <- gsub("Black Plus", "BlackPlus", alexa_file$variation)
alexa_file$variation <- gsub("Black Show", "BlackShow", alexa_file$variation)
alexa_file$variation <- gsub("Black Spot", "BlackSpot", alexa_file$variation)

alexa_file$variation <- gsub("White Dot", "WhiteDot", alexa_file$variation)
alexa_file$variation <- gsub("White Plus", "WhitePlus", alexa_file$variation)
alexa_file$variation <- gsub("White Show", "WhiteShow", alexa_file$variation)
alexa_file$variation <- gsub("White Spot", "WhiteSpot", alexa_file$variation)

alexa_file

## # A tibble: 3,150 x 5
##   rating date      variation      verified_reviews      feedback
##   <dbl> <dtm>      <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
```

7b. Get the total number of each variations and save it into another object. Save the object as variations.RData. Write the R scripts. What is its result?

```
library("dplyr")

##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:stats':
##
##   filter, lag
##
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

var_total <- alexa_file %>%
  count(alexa_file$variation)

var_total

## # A tibble: 16 x 2
##   `alexa_file$variation`      n
##   <chr>                  <int>
## 1 Black                  261
## 2 BlackDot               516
## 3 BlackPlus              270
## 4 BlackShow              265
## 5 BlackSpot              241
## 6 Charcoal Fabric        430
## 7 Configuration: Fire TV Stick 350
## 8 Heather Gray Fabric    157
## 9 Oak Finish              14
## 10 Sandstone Fabric       90
## 11 Walnut Finish          9
## 12 White                  91
## 13 WhiteDot              184
## 14 WhitePlus              78
## 15 WhiteShow              85
## 16 WhiteSpot             109

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

7c. From the variations.RData, create a barplot(). Complete the details of the chart which include the title, color, labels of each bar.

```
load("variations.RData")
var_total

## # A tibble: 16 x 2
##   `alexa_file$variation`      n
##   <chr>                  <int>
## 1 Black                  261
## 2 BlackDot               516
## 3 BlackPlus              270
## 4 BlackShow              265
## 5 BlackSpot              241
## 6 Charcoal Fabric        430
```

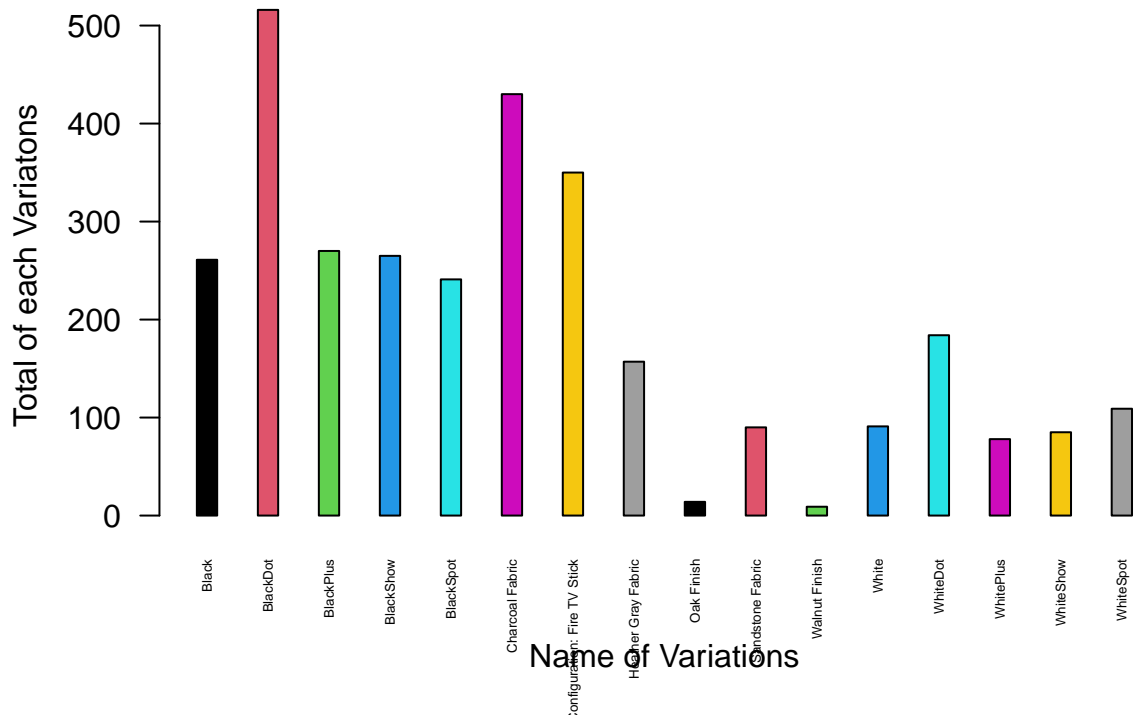


```
## 7 Configuration: Fire TV Stick    350
## 8 Heather Gray Fabric            157
## 9 Oak Finish                     14
## 10 Sandstone Fabric              90
## 11 Walnut Finish                 9
## 12 White                         91
## 13 WhiteDot                      184
## 14 WhitePlus                     78
## 15 WhiteShow                     85
## 16 WhiteSpot                     109
```

```
varNames <- var_total$`alexa_file$variation`

totalPlot <- barplot(var_total$n,
  names.arg = varNames,
  main = "Total number of each variations",
  xlab = "Name of Variations",
  ylab = "Total of each Variatons",
  col = 1:16,
  space = 2,
  cex.names = 0.4,
  las = 2 )
```

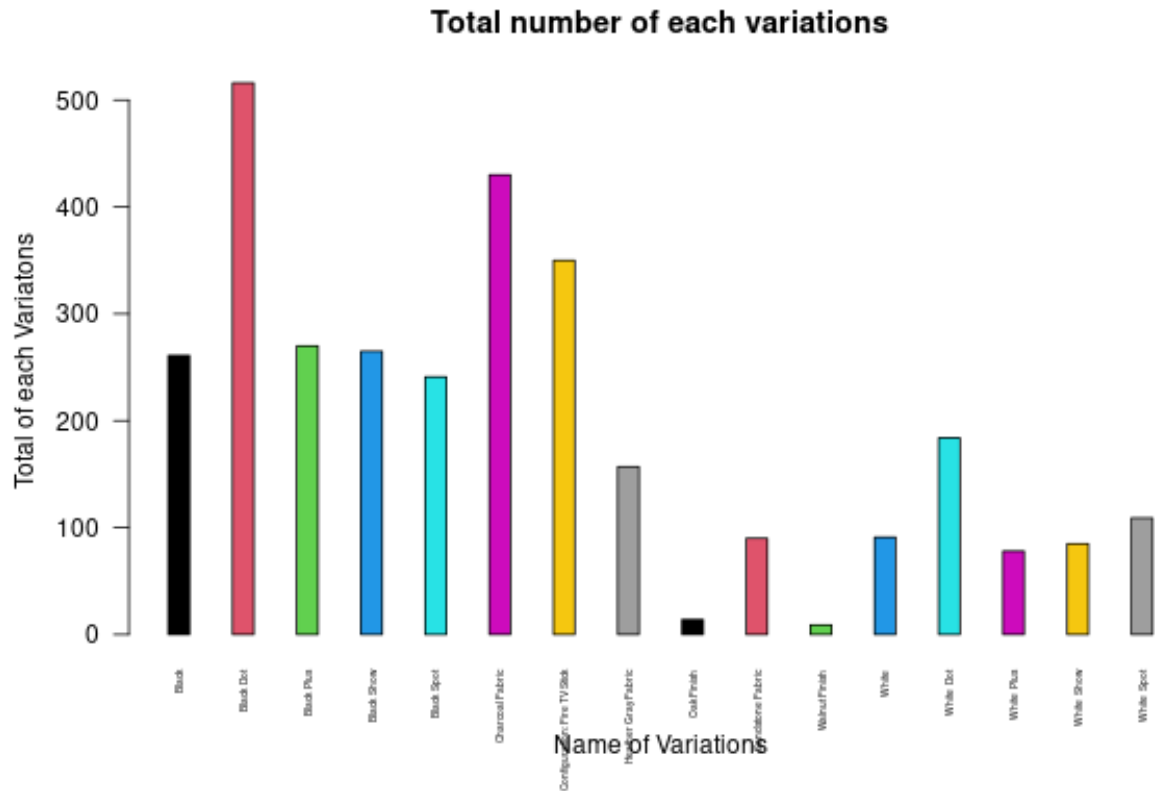
Total number of each variations



```
png("totalPlot.png")
dev.off()
```

```
## pdf
## 2
```

```
knitr::include_graphics("/cloud/project/RWorksheet4/RWorksheet#4b/totalPlot.png")
```



7d. Create a `barplot()` for the black and white variations. Plot it in 1 frame, side by side. Complete the details of the chart.

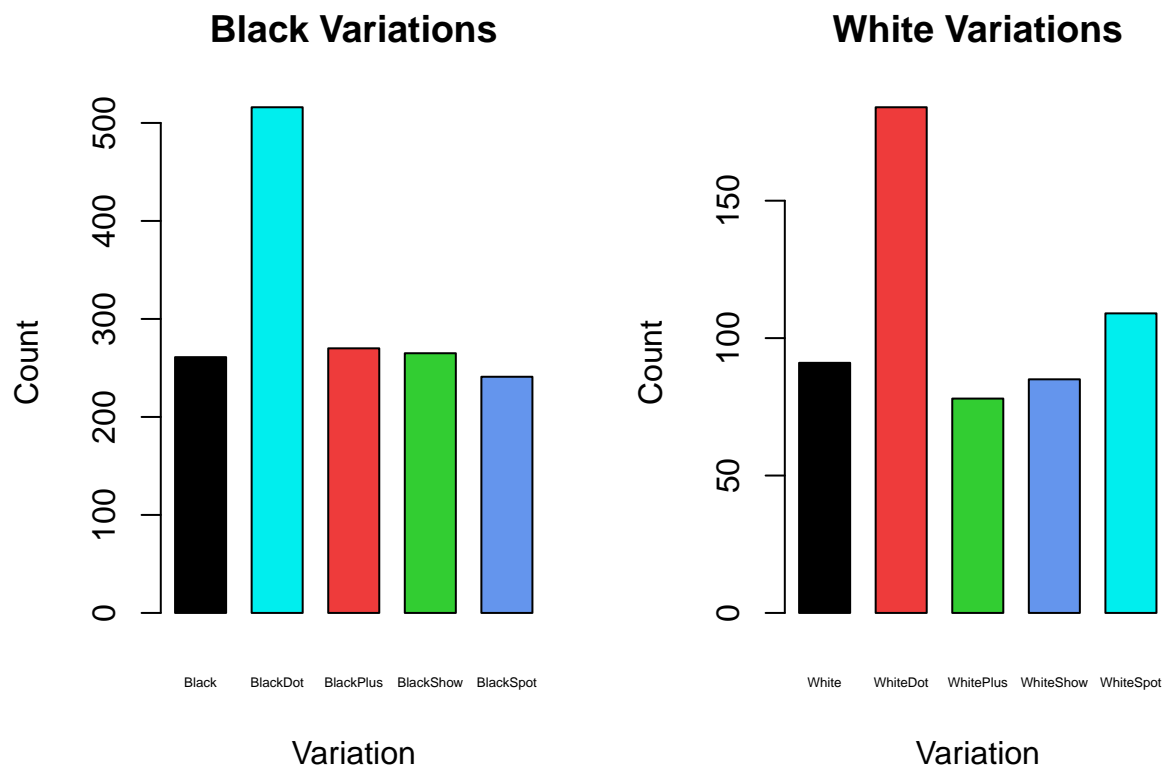
```
library(graphics)
blackVars <- var_total[var_total$`alexa_file$variation` %in% c("Black","BlackPlus","BlackShow","BlackSpot"),]
whiteVars <- var_total[var_total$`alexa_file$variation` %in% c("White","WhiteDot","WhitePlus","WhiteShow"),]

par(mfrow = c(1,2))

blackPlot <- barplot(height = blackVars$n,
  names.arg = blackVars$`alexa_file$variation`,
  col = c("black","cyan2","brown2","limegreen","cornflowerblue"),
  main = "Black Variations",
  xlab = "Variation",
  ylab = "Count",
  border = "black",
  space = 0.5,
  cex.names = 0.4)

whitePlot <- barplot(height = whiteVars$n,
  names.arg = whiteVars$`alexa_file$variation`,
  col = c("black","brown2","limegreen","cornflowerblue","cyan2"),
  main = "White Variations",
```

```
xlab = "Variation",
ylab = "Count",
border = "black",
space = 0.5,
cex.names = 0.4)
```



```
png("BlackN_whitePlot.png")
dev.off()
```

```
## pdf
## 2
```

```
knitr::include_graphics("/cloud/project/RWorksheet4/RWorksheet#4b/BlackN_whitePlot.png")
```

