

RWorksheet_Cautivar#4b.Rmd

James Clark Cautivar

2024-10-27

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)

mat <- matrix(nrow = 5, ncol = 5)

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

for (i in 1:5) {
  for (j in 1:5) {
    cat(mat[i, j], " ")
  }
  cat("\n")
}
```

```
## 1  2  3  4  5
## 2  1  2  3  4
## 3  2  1  2  3
## 4  3  2  1  2
## 5  4  3  2  1
```

```
cat("\n")
```

```
matrixz <- matrix(0, nrow = 5, ncol = 5)
```

```
for (i in 1:5) {
  for (j in 1:5) {
    cat(matrixz[i, j], " ")
  }
  cat("\n")
}
```

```
## 0  0  0  0  0
## 0  0  0  0  0
## 0  0  0  0  0
## 0  0  0  0  0
## 0  0  0  0  0
```

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

```
for (i in 1:5) {
  for (j in 1:i) {
    cat("*", " ")
  }
  cat("\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.

```
inp <- as.integer(readline(prompt = "Input a number: "))
```

```
## Input a number:
```

```
inp <- 0 #example value of input because i cant knit it if it has a value of NA
```

```
a <- 0
```

```
b <- 1
```

```
repeat {
```

```
  if (a >= inp) {
    cat(a, " ")
  }
```

```
  fib <- a + b
  a <- b
  b <- fib
```

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

```
}
```

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

```
cat("\n")
```

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

```
file <- read.table("RespondentsData.csv", header = TRUE, sep = ",")
file
```

```
## 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
## 7       9.5  70.0    F
## 8       9.0  71.0    F
## 9      13.0  72.0    M
## 10      7.5  64.0    F
## 11     10.5  74.5    M
## 12      8.5  67.0    F
## 13     12.0  71.0    M
## 14     10.5  71.0    M
## 15     13.0  77.0    M
## 16     11.5  72.0    M
## 17      8.5  59.0    F
## 18      5.0  62.0    F
## 19     10.0  72.0    M
## 20      6.5  66.0    F
## 21      7.5  64.0    F
## 22      8.5  67.0    M
## 23     10.5  73.0    M
## 24      8.5  69.0    F
## 25     10.5  72.0    M
## 26     11.0  70.0    M
## 27      9.0  69.0    M
## 28     13.0  70.0    M
```

```
head(file)
```

```
##   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
```

- 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.

```
females <- subset(file, Gender == "F")
males <- subset(file, Gender == "M")
females
```

```
##   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
## 6      7.0   64.0      F
## 7      9.5   70.0      F
## 8      9.0   71.0      F
## 10     7.5   64.0      F
## 12     8.5   67.0      F
## 17     8.5   59.0      F
## 18     5.0   62.0      F
## 20     6.5   66.0      F
## 21     7.5   64.0      F
```

```
## 24      8.5   69.0    F
```

```
males
```

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

```
#Number of Observations
```

```
femalesObs <- nrow(females)
```

```
malesObs <- nrow(males)
```

```
femalesObs
```

```
## [1] 14
```

```
malesObs
```

```
## [1] 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.

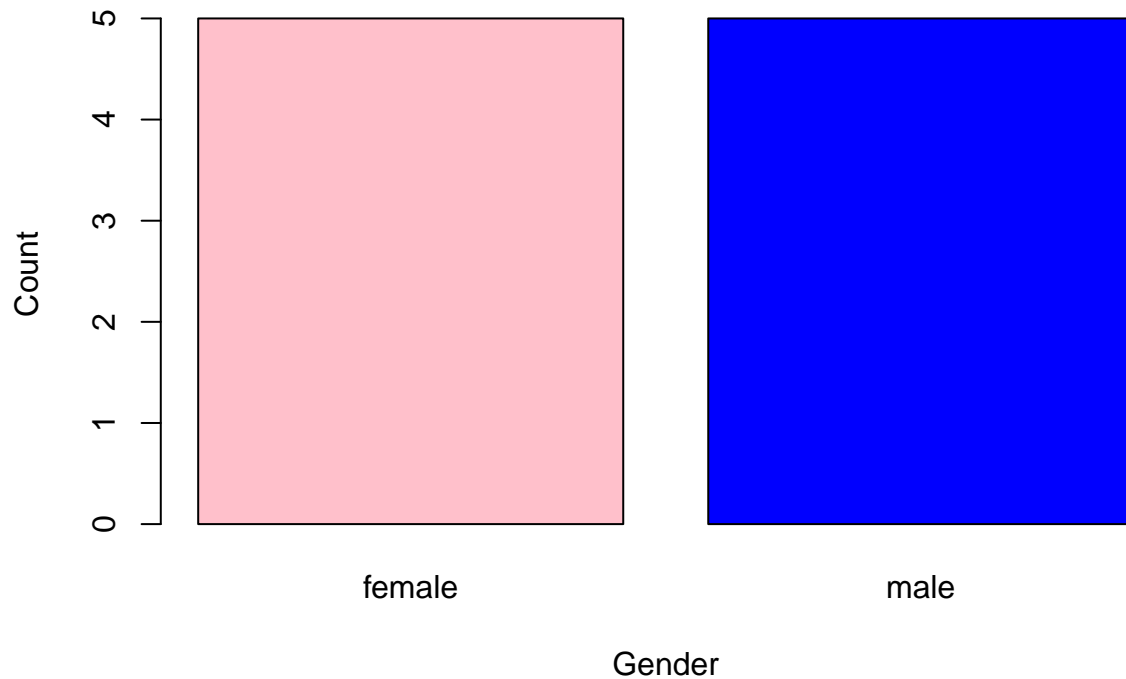
```
load("householdData.RData")
```

```
#i searched and used this function because the original householdData dataframe has extra spaces and it
householdData[["Sex"]] <- gsub(" ", "", householdData[["Sex"]])
```

```
#i also searched the table function to easily count the frequency of each gender. This is where i encour
genderCounts <- table(householdData$Sex)
```

```
barplot(
  genderCounts,
  col = c("pink", "blue"),
  main = "Number of Males and Females in Household",
  xlab = "Gender",
  ylab = "Count"
)
```

Number of Males and Females in Household

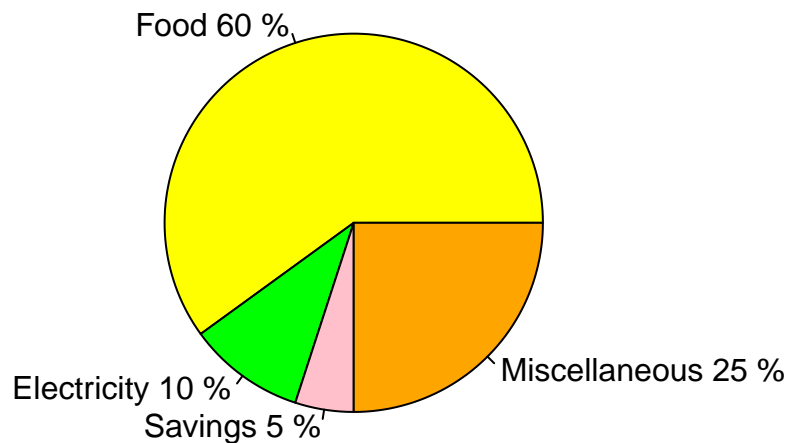


5. The monthly income of Dela Cruz family was spent on the following: a. Create a piechart that will include labels in percentage. Add some colors and title of the chart. Write the R scripts and show its output.

```
amounts <- c(60, 10, 5, 25)
categories <- c("Food", "Electricity", "Savings", "Miscellaneous")
labels <- paste(categories, amounts, "%")

pie(
  amounts,
  labels = labels,
  col = c("yellow", "green", "pink", "orange"),
  main = "Expenses of Dela Cruz Family"
)
```

Expenses of Dela Cruz Family



6. Use the iris dataset. 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 ...
```

The iris data is a dataframe and based on the `str` function, it has 150 observations/rows and 5 variables/columns. The variables are Sepal.Length, Sepal.Width, Petal.Length, Petal.Width, and Species.

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?

```
means <- c(
  Sepal.Length = mean(iris$Sepal.Length),
  Sepal.Width = mean(iris$Sepal.Width),
  Petal.Length = mean(iris$Petal.Length),
  Petal.Width = mean(iris$Petal.Width)
)
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.

```
species <- table(iris$Species)
colors <- c("blue", "green", "pink")

pie(
  species,
  main = "Species Distribution",
  col = colors,
```

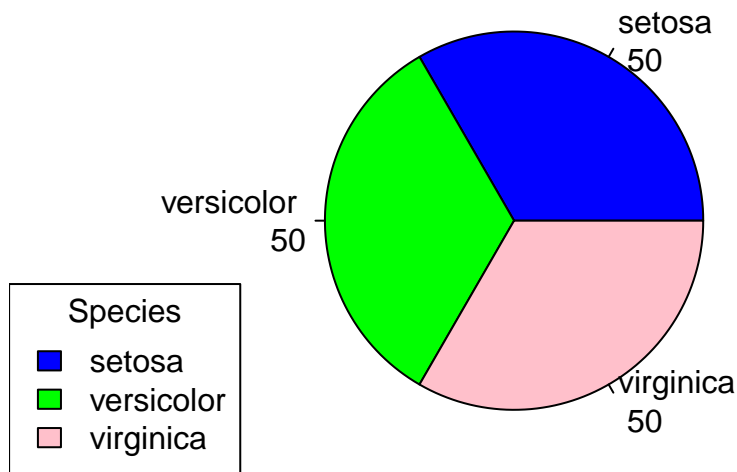
```

labels = paste(names(species), "\n", species),
)

legend(
  "bottomleft",
  legend = names(species),
  fill = colors,
  title = "Species"
)

```

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.

```

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

tail(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

```

```
tail(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

```

```
tail(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.

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

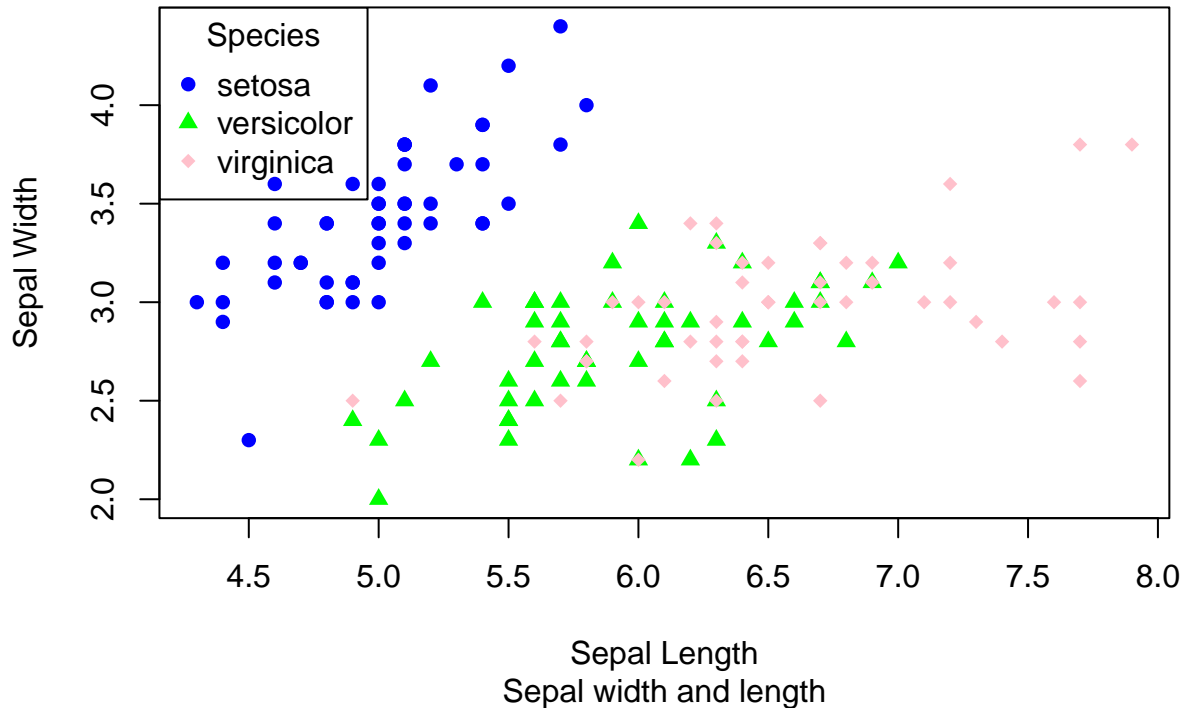
```
colors <- c("setosa" = "blue", "versicolor" = "green", "virginica" = "pink")
```

```
symbols <- c("setosa" = 16, "versicolor" = 17, "virginica" = 18)
```

```
plot(
  iris$Sepal.Length, iris$Sepal.Width,
  col = colors[iris$Species],
  pch = symbols[iris$Species],
  main = "Iris Dataset",
  sub = "Sepal width and length",
  xlab = "Sepal Length",
  ylab = "Sepal Width"
)
```

```
legend(
  "topleft",
  legend = levels(iris$Species),
  col = colors,
  pch = symbols,
  title = "Species"
)
```


Iris Dataset



f. Interpret the result. The scatterplot shows the relationship between Sepal Length and Sepal Width for the three species. Setosa has the widest sepal but also has the shortest sepal length among the species. Versicolor on the other hand has average sepal width and longer sepal. Lastly, the Virginica has the longest sepal and an average sepal width.

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 <- read_xlsx("alexa_file.xlsx")
alexa
```

```
## # 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
```

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

```

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

```

```
alexa$variation[1052:2000]
```

```

## [1] "WhiteSpot" "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot" "WhiteSpot"
## [7] "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot" "WhiteSpot"
## [13] "WhiteSpot" "BlackSpot" "WhiteSpot" "WhiteSpot" "WhiteSpot" "WhiteSpot"
## [19] "BlackSpot" "BlackSpot" "WhiteSpot" "BlackSpot" "BlackSpot" "BlackSpot"
## [25] "WhiteSpot" "BlackSpot" "BlackSpot" "WhiteSpot" "BlackSpot" "BlackSpot"
## [31] "BlackSpot" "BlackSpot" "WhiteSpot" "BlackSpot" "WhiteSpot" "BlackSpot"
## [37] "BlackSpot" "WhiteSpot" "WhiteSpot" "BlackSpot" "BlackSpot" "BlackSpot"
## [43] "BlackSpot" "WhiteSpot" "WhiteSpot" "BlackSpot" "BlackSpot" "BlackSpot"
## [49] "WhiteSpot" "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot"
## [55] "WhiteSpot" "BlackSpot" "WhiteSpot" "BlackSpot" "WhiteSpot" "BlackSpot"
## [61] "BlackSpot" "WhiteSpot" "WhiteSpot" "WhiteSpot" "BlackSpot" "BlackSpot"
## [67] "BlackSpot" "WhiteSpot" "WhiteSpot" "BlackSpot" "BlackSpot" "BlackSpot"
## [73] "WhiteSpot" "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot"
## [79] "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot"
## [85] "BlackSpot" "WhiteSpot" "WhiteSpot" "WhiteSpot" "BlackSpot" "BlackSpot"
## [91] "WhiteSpot" "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot" "WhiteSpot"
## [97] "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot" "WhiteSpot" "BlackSpot"
## [103] "BlackSpot" "WhiteSpot" "BlackSpot" "WhiteSpot" "BlackSpot" "BlackSpot"
## [109] "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot" "WhiteSpot"
## [115] "BlackSpot" "WhiteSpot" "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot"
## [121] "BlackSpot" "WhiteSpot" "BlackSpot" "WhiteSpot" "BlackSpot" "WhiteSpot"
## [127] "BlackSpot" "BlackSpot" "BlackSpot" "WhiteSpot" "WhiteSpot" "BlackSpot"
## [133] "WhiteSpot" "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot" "WhiteSpot"
## [139] "BlackSpot" "WhiteSpot" "BlackSpot" "BlackSpot" "BlackSpot" "WhiteSpot"
## [145] "BlackSpot" "BlackSpot" "BlackSpot" "WhiteSpot" "WhiteSpot" "BlackSpot"
## [151] "BlackSpot" "BlackSpot" "WhiteSpot" "WhiteSpot" "BlackSpot" "WhiteSpot"
## [157] "BlackSpot" "BlackSpot" "WhiteSpot" "BlackSpot" "BlackSpot" "WhiteSpot"
## [163] "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot"
## [169] "WhiteSpot" "WhiteSpot" "BlackSpot" "WhiteSpot" "BlackSpot" "WhiteSpot"
## [175] "BlackSpot" "BlackSpot" "BlackSpot" "WhiteSpot" "BlackSpot" "BlackSpot"
## [181] "WhiteSpot" "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot"
## [187] "BlackSpot" "BlackSpot" "WhiteSpot" "WhiteSpot" "BlackSpot" "WhiteSpot"
## [193] "WhiteSpot" "WhiteSpot" "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot"
## [199] "WhiteSpot" "BlackSpot" "WhiteSpot" "WhiteSpot" "WhiteSpot" "BlackSpot"
## [205] "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot"
## [211] "WhiteSpot" "WhiteSpot" "BlackSpot" "WhiteSpot" "WhiteSpot" "WhiteSpot"
## [217] "BlackSpot" "BlackSpot" "BlackSpot" "WhiteSpot" "WhiteSpot" "BlackSpot"
## [223] "WhiteSpot" "WhiteSpot" "WhiteSpot" "WhiteSpot" "BlackSpot" "WhiteSpot"
## [229] "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot" "WhiteSpot" "WhiteSpot"
## [235] "WhiteSpot" "BlackSpot" "BlackSpot" "BlackSpot" "WhiteSpot" "WhiteSpot"
## [241] "BlackSpot" "WhiteSpot" "BlackSpot" "BlackSpot" "WhiteSpot" "BlackSpot"
## [247] "WhiteSpot" "BlackSpot" "WhiteSpot" "BlackSpot" "BlackSpot" "BlackSpot"
## [253] "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot" "BlackSpot"

```

[illegible]

[illegible]

```
## [907] "BlackPlus" "BlackPlus" "BlackPlus" "BlackPlus" "BlackPlus" "BlackPlus"
## [913] "BlackPlus" "BlackPlus" "BlackPlus" "BlackPlus" "WhitePlus" "WhitePlus"
## [919] "BlackPlus" "BlackPlus" "WhitePlus" "WhitePlus" "BlackPlus" "BlackPlus"
## [925] "BlackPlus" "BlackPlus" "WhitePlus" "BlackPlus" "BlackPlus" "BlackPlus"
## [931] "BlackPlus" "WhitePlus" "BlackPlus" "BlackPlus" "BlackPlus" "BlackPlus"
## [937] "BlackPlus" "BlackPlus" "WhitePlus" "WhitePlus" "BlackPlus" "BlackPlus"
## [943] "WhitePlus" "BlackPlus" "WhitePlus" "BlackPlus" "BlackPlus" "BlackPlus"
## [949] "BlackPlus"
```

- b. 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

totalVar <- alexa %>%
count(alexas$variation)
totalVar

## # A tibble: 16 x 2
##   `alexas$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

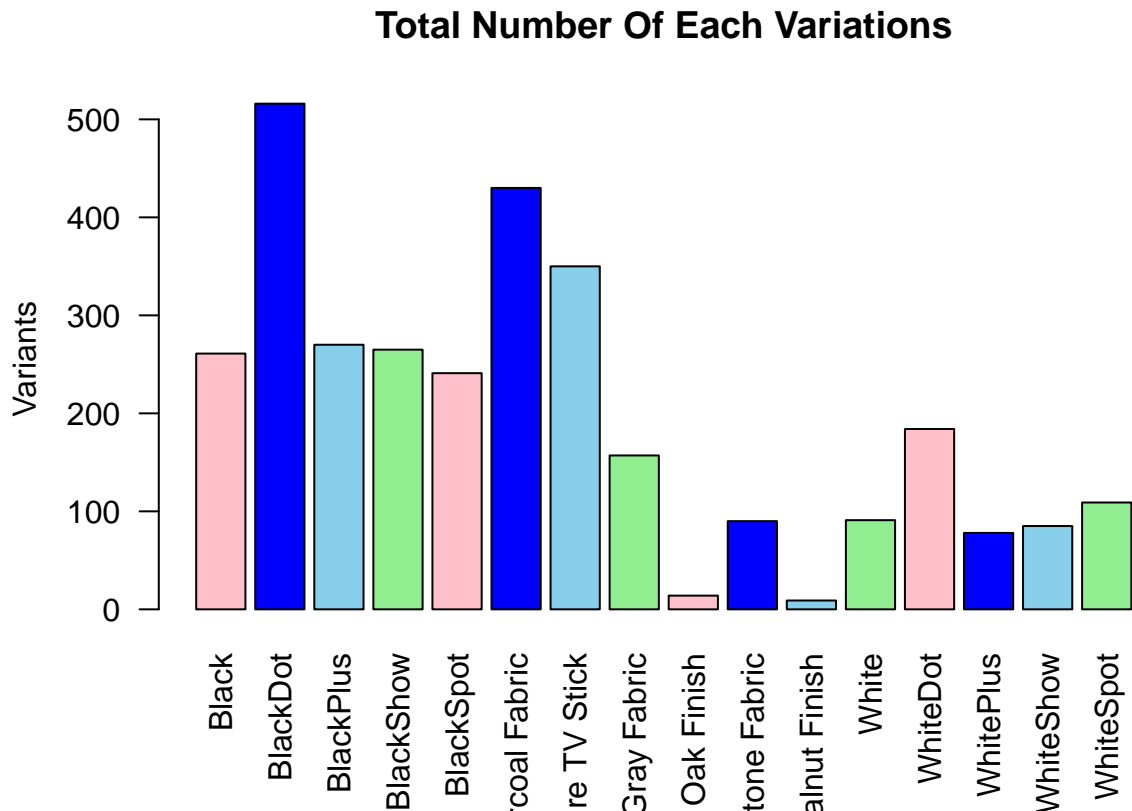
save(totalVar, file = "variations.RData")
```

- c. 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")

barplot(
  totalVar$n,
```

```
col = c("pink", "blue", "skyblue", "lightgreen"),
main = "Total Number Of Each Variations",
ylab = "Variants",
names.arg = totalVar$`alexa$variation`,
las = 2
)
```



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

```
library(dplyr)

load("variations.RData")

par(mfrow = c(1, 2))

blackV <- totalVar %>%
  filter(`alexa$variation` %in% c("Black", "BlackDot", "BlackPlus", "BlackShow", "BlackSpot"))

barplot(
  height = blackV$n,
  names.arg = blackV$`alexa$variation`,
  col = c("pink", "blue", "skyblue", "lightgreen", "cyan"),
  main = "Black Variants",
  xlab = "Total Numbers",
  ylab = "Variations",
  las = 2
)
```

```

WhiteV <- totalVar %>%
  filter(`alexa$variation` %in% c("White", "WhiteDot", "WhitePlus", "WhiteShow", "WhiteSpot"))

barplot(
  height = WhiteV$n,
  names.arg = WhiteV$`alexa$variation`,
  col = c("pink", "blue", "skyblue", "lightgreen", "cyan"),
  main = "White Variants",
  xlab = "Total Numbers",
  ylab = "Variations",
  las = 2
)

```

