ASSIGNMENT 2

Data Science with R

```
library(tidyverse)

library(rvest)



#ques-d

file <- read_html("https://stats.stackexchange.com/questions?tab=Votes")

question_title <- file %>% html_elements("#questions .s-link") %>% html_text()

views <- file %>% html_elements(".is-supernova .s-post-summary--stats-item-number") %>% html_text()

answers <- file %>% html_elements(".has-answers") %>% html_text()

votes <- file %>% html_elements(".s-post-summary--stats-item__emphasized .s-post-summary--stats-item-number") %>% html_text()

Dataset <- data.frame("The title of the question"= question_title, "The number of views"= views,"The number of answers"= answers, "The notes of votes"= votes)



#ques-a

data(iris)

windows(width = 10, height = 8)

#side-by-side boxplots of all continuous variables based on the column species

boxplot(Sepal.Length ~ Species, data = iris, col = c("red", "green", "yellow"), main = "Boxplot of Sepal Length by Species")

boxplot(Petal.Length ~ Species, data = iris, col = c("red", "green", "yellow"), main = "Boxplot of Petal Length by Species")

boxplot(Sepal.Width ~ Species, data = iris, col = c("red", "green", "yellow"), main = "Boxplot of Sepal Width by Species")

boxplot(Petal.Width ~ Species, data = iris, col = c("red", "green", "yellow"), main = "Boxplot of Petal Width by Species")

#scatterplot of Sepal.Length and Petal.Length
```

```
plot(Sepal.Length ~ Petal.Length, data = iris, col = c("red","green","yellow"), xlab =
"Petal.Length", ylab = "Sepal.Length", pch = 16)


#From the scatterplot we can observed the relation between Speal.Length and Petal.Length




#ques-b
library(imager)
dog <- load.image("dog.jpeg")
col.mat <- as.array(dog[,,1,])
dims <- dim(col.mat)
flip <- array(0,dim = c(dims[1],dims[2],dims[3]))

 for(i in 1:dims[1])
{
  for(j in 1:dims[2])
  {
    flip[i,j,] <- col.mat[dims[1]-i+1,j,]
  }
}
#plot size by side
par(mfrow = c(1,2))
plot(dog)
plot(as.cimg(flip))



#ques-c
library(MASS)

data(ships)
```

```
ships_dataset <- data.frame(ships)
#plot to prove the hypothesis
library(ggplot2)
ggplot(ships,aes(type,incidents))+
  geom_point()
#From the plot it is clear that for ship type "B" the number of damage incidents is maximum
#Hence Ship type B is the least trustworthy ship



#ques-e
days <- 1
probability <- 1/100
expected_days <- 0
while(probability > 0){
  expected_days <- expected_days + days*probability
  days <- days + 1
  probability <- probability*(99-days+1)/(100-days+1)

}
print(expected_days)

##########################################################
```