Assignment - 2

Data Science with R

Deadline : June 13, Tuesday, 23:59 IST

Instructions :

1. You need to go through the whole assignment and attempt all problems in a single R script file.
2. Make a Github Repository ,and upload your R Script file in that Repository. We will Circulate a Google form, in which all of you need to push the link of the repository, for the grading of your assignment.
3. We have **Strict Advisory** for **Cheating, Copying & Plagiarizing** from others assignments. Anyone should not share their solutions of assignment with others. If caught, both(or the group if more than 2 are involve) will be awarded with 0 marks in this assignments, no arguments will be entertain. This will effect your ratification chances badly.
4. You can discuss problems with your friends and welcome to discuss with mentors over discord.
5. This assignments contains 25% weightage of overall project.
6. Partial Grading will be done, if your solution doesn't give desired output.

Questions :

a. Load the dataset iris by using command data(iris). Draw side-by-side boxplots of all continuous variables based on the column Species. Make a scatterplot of Sepal.Length and Petal.Length coloring each plot by Species. What can you conclude from this scatterplot?

b. Write a function, `flip()`, that takes an `imager` image as an argument, and returns an `imager` image that is mirrored across the the vertical axis. For example, if input image is shown on the left below, and the output image should be the one on the right.


Note: The input argument for the function MUST be an image loaded with `imager` and NOT the location or file name of the image.

c. In library `MASS` is the dataset `ships` which contains the dataset of damage caused by waves on various cargo ships. There are 5 variables in the dataset (you can learn more about it using `?ships`). The `type` of ship is listed as a factor with levels "A", "B", ..., "E". Looking at the data I have the following hypothesis:

"*Ship type B is the least trustworthy ship as it had the most accidents*"

Make a plot or plots to either "prove" or "disprove" my hypothesis. Submit all codes required to make your plots

d. From the website: https://stats.stackexchange.com/questions?tab=Votes, scrape the following information and make a dataframe:

1. The title of the questions
2. The number of views
3. The number of answers
4. The notes of votes

e. You take half of a vitamin every morning. The vitamins are sold in a bottle of 100 (whole) tablets, so at first you have to cut the tablets in half. Every day you randomly pull one thing from the bottle. If it's a whole tablet, you cut it in half and put the leftover half back in the bottle. If it's a half-tablet, you take the vitamin. You just bought a fresh bottle. How many days, on average, will it be before you pull a half-tablet out of the bottle?