# Assignment - 3

Data Science with R

**Deadline : June 27, 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:**

1. Make a scatterplot of Sepal.Length and Petal.Length (from `iris` Dataset) coloring each plot by Species. What can you conclude? you must do this in ggplot2.
2. Load the dataset txhousing in package ggplot2. Explore the dataset with plots similar to the way we did in last problem set. There are some missing values in the dataset. You may use function complete.cases() if required. **Note:** We are not asking anything particular in this question, as of now, whatever you have learned about to extract and observe information from visualization, statistics and dataset manipulation etc, you should try to implement in this question and extract useful insights about data.

3. Titanic Dataset: Load the dataset `titanic.csv` provided to you. This dataset contains information regarding the passengers on the Titanic. The columns are:

   - `PassengerId` - unique ID for each passenger
   - `Survived` - whether the passenger survived or nor
   - `Pclass` - the class of their ticket: 1st class, 2nd class, 3rd class
   - `Sex` - male or female
   - `Age` - age
   - `SibSp` - number of siblings/spouse traveling with them
   - `Parch` - number of parents/children traveling with them
   - `Fare` - cost of their ticket
   - `Embarked` - the Port from which they boarded the ship. C = Cherbourg, Q = Queenstown, S = Southampton. Write `R` code to recreate the plot below, **exactly**. The final `ggplot` plot should be saved in the object `finalP`. That is, your last line of code must look like

```
final_Plot <- ggplot( ... ) ....
```