Practice Assignment 07

Download the titanic dataset from [kaggle](#). Unzip the folder and keep only the train.csv file. Rename the file as titanic. csv

Create a GitHub repository called "st2195_assignment_7" and include an R script able to perform the operations described below.

1. Generate a series of bar charts to describe the gender, ticket class and survival of the passengers onboard. [1 point]

2. Generate a histogram for the passengers' age. Furthermore, describe the passengers' age using the following two boxplots: age per ticket class and age based on survival. [1 point]

3. Generate a histogram for the travel fare and a table showing the number of people who did not pay – you may want to check on Google why a handful of people was on board for free! [1 points]

4. A chart of your choice to describe the family size per ticket class [1.5 points]

5. A series of stacked bar charts to show the how survival differs for different gender and ticket class [1.5 points]

6. A violin chart describing how survival related to age and gender [1.5 points]

7. A violin chart describing the survival rate related to age and ticket class [1.5 points]

From the graphs above, what kind of associations between the variables like age, gender, ticket class and the survival do you observe? [1 point]

Note: it would be best to use ggplot.

Additional Notes:

- Download the Titanic dataset from Kaggle.
  - You may need to log in to Kaggle. If you don't already have an account, create one.
  - Read through the description of the dataset to understand what each variable contains.
  -

Data Dictionary

Variable	Definition	Key
survival	Survival	0 = No, 1 = Yes
pclass	Ticket class	1 = 1st, 2 = 2nd, 3 = 3rd
sex	Sex	
Age	Age in years	
sibsp	# of siblings / spouses aboard the Titanic	
parch	# of parents / children aboard the Titanic	
ticket	Ticket number	
fare	Passenger fare	
cabin	Cabin number	
embarked	Port of Embarkation	C = Cherbourg, Q = Queenstown, S = Southampton

  -

Variable Notes

pclass: A proxy for socio-economic status (SES)
1st = Upper
2nd = Middle
3rd = Lower

age: Age is fractional if less than 1. If the age is estimated, is it in the form of xx.5

sibsp: The dataset defines family relations in this way...
Sibling = brother, sister, stepbrother, stepsister
Spouse = husband, wife (mistresses and fiancés were ignored)

parch: The dataset defines family relations in this way...
Parent = mother, father
Child = daughter, son, stepdaughter, stepson
Some children travelled only with a nanny, therefore parch=0 for them.

  -

- Hints
  - Computing family size (for task 4 of requirements) can be achieved by using the formula: sibsp + parch + 1
    - sibsp: # of siblings / spouses aboard the Titanic
    - parch: # of parents / children aboard the Titanic