CPTR420
October 19, 2023
Class Exercise

This assignment will be graded for attempt rather than accuracy. You are encouraged to collaborate with your peers. Do as much as you can. Add comments/questions to your script where necessary

The titanic_train dataset contains data about passengers on the Titanic ship
The data is available at CRAN and can be installed with the titanic package. The package contains several datasets.

titanic_train has data about survivors as well as passengers who did not survive. Some data are missing from the set.

The columns are: PassengerId , Survived (1= yes, 0= no), Pclass (1= first, 2= second, 3=third), Name, Sex, Age, SibSp, Parch, Ticket, Fare, Cabin, Embarked (i.e port of embarkation)

1. Install and load the titanic package and store the titanic_train dataset as a dataframe in a new object
2. Prepare the data for visualization:
    a. Remove all observations where the Survived variable is not known
    b. Assign the mean age to all ages that are unknown
    c. Assign the median fare to all unknown fares
    d. Assign S to all Embarked values that are unknown
    e. Make sure that the Pclass (i.e. Passenger class), and Survived variables are factors.
3. Create another dataframe called survivors_df that has data for the survivors only
4. Create plots for the following
    a. Show how the number of males and females on the ship compare by age (overlapping histogram with 10 age bins)
    b. Show number of survivors by class
    c. Prepare a plot that compares the number of passengers and survivors/non-survivors from each class. Use a bar with one column of survivors and the other of non-survivors
    d. Show how the number of male survivors compares with that of females survivors by age. (bar plot)
    e. Is survival rate influenced by the port of embarkation?
    Use a bar chart with Embarked on X axis. plot the bars side by side
    f. Is there any relationship between Fare and Age? Use points and smooth
    g. Is survival related to this relationship between Age and Fare?
    h. Is there any correlation between Age and Survived and Gender?
    i. Create a plot that compares survival by age.

Upload your completed script to LearningHub