

Practice assignment 7

Download the titanic dataset from [kaggle](https://www.kaggle.com/titanic). 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.