

# Data exploration with the Titanic dataset

## Exercise 1: Use data transformation and visualization to answer the following questions (50 min):

Question 1: According to Wikipedia, there was an estimated 2,224 passengers and crew onboard the Titanic when it sank. How many of them do we have information for in this dataset? Of the people we have data for, how many of them survived and how many did not? What is the overall survival rate?

Question 2. How many passengers on the Titanic were males and how many were females? What do you find when you break it down by ticket class?

Question 3. How many passengers of each sex survived and how many of them did not? What is the survival rate for passengers of each sex?

Question 4. For how many passengers do we have age information (including estimated age)? For how many is the age information missing? What is the age distribution for passengers whose age information is available

Question 5. Show the age distribution per ticket class, per sex. What do you find?

Question 6. How do the sex, ticket class, and age of a passenger affect their chance of survival? Try to use a single plot to answer this question.

Question 7. Show the distribution of the number of family members (including siblings, spouses, parents, and children) that each passenger was accompanied by. Were most passengers travelling solo or with family?

Question 8. Which ticket class did most of the largest families get? And which ticket class has the lowest proportion of female passengers who travelled solo out of all the female passengers in that class?

Question 9. In this dataset, the Fare variable does not represent the fare per person. Instead, each ticket number has a corresponding fare, and some passengers share one single ticket number. Therefore, the Fare variable is the total fare for a group of passengers sharing the same ticket number. Knowing this, calculate the average fare per person. You don't need to show a table or a figure for this question, just show the code for the calculation.

Question 10. What is the distribution of the per-ticket fare for each ticket class?