For the following exercise, please use the ‘Titanic’ dataset provided to you.

1. Read the Titanic dataset.

2. Make two new dataframes : a subset of male survivors, and a subset of female survivors. [Hint: Use command ‘subset’]

3. Based on question 2, what was the name of the oldest surviving individual? In what class was the youngest surviving male female? Hint: use which.max, which.min on the subsets you just created.

4. Take 10 random names of passengers from the Titanic, and sort them alphabetically in descending order. Hint: use command ‘sample’ and then ‘sort’.

5. Convert the ’Name’ (passenger name) variable to a ’character’ variable, and store it in the dataframe.

6. How many observations of ’Age’ are missing from the dataframe? Hint: use ‘summary’

7. Make a new variable called ’Status’, based on the ’Survived’ variable already in the dataset. For passengers that did not survive, Status should be ’dead’, for those who did, Status should be ’alive’. Make sure this new variable is a factor

8. Count the number of passengers in each class (1st, 2nd, 3rd)

9. Using grep, find all passengers with the last name ’Moran’. Make this subset into a new dataframe. Did they all survive? [note: grep is the R command for regular expression]

10. For what proportion of the passengers is the age unknown? Was this proportion higher for 2nd class than 1st and 3rd? Hint: First make a subset of the dataframe where age is missing