5. Problem Statement

1. Import the Titanic Dataset from the following link:

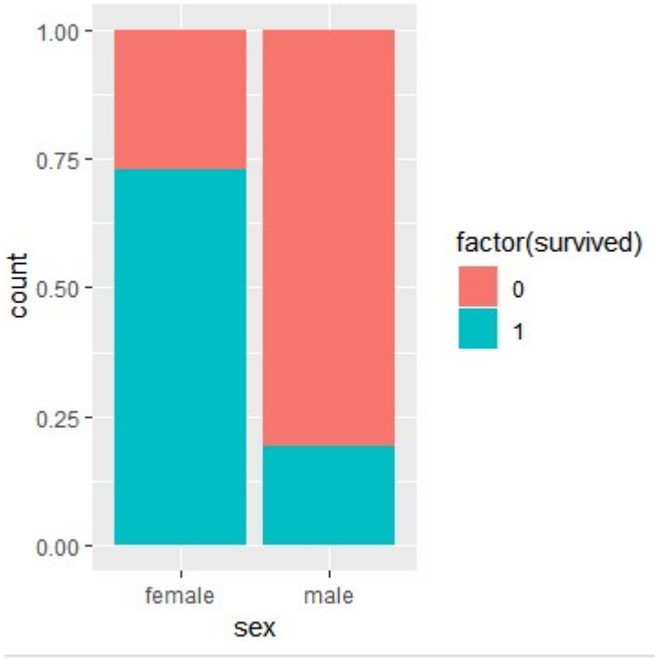
https://drive.google.com/file/d/1JTJCjdGuUxzKXYlwOavwovB01k6FWg3r/view?ts=5b42ea10

Perform the below operations:

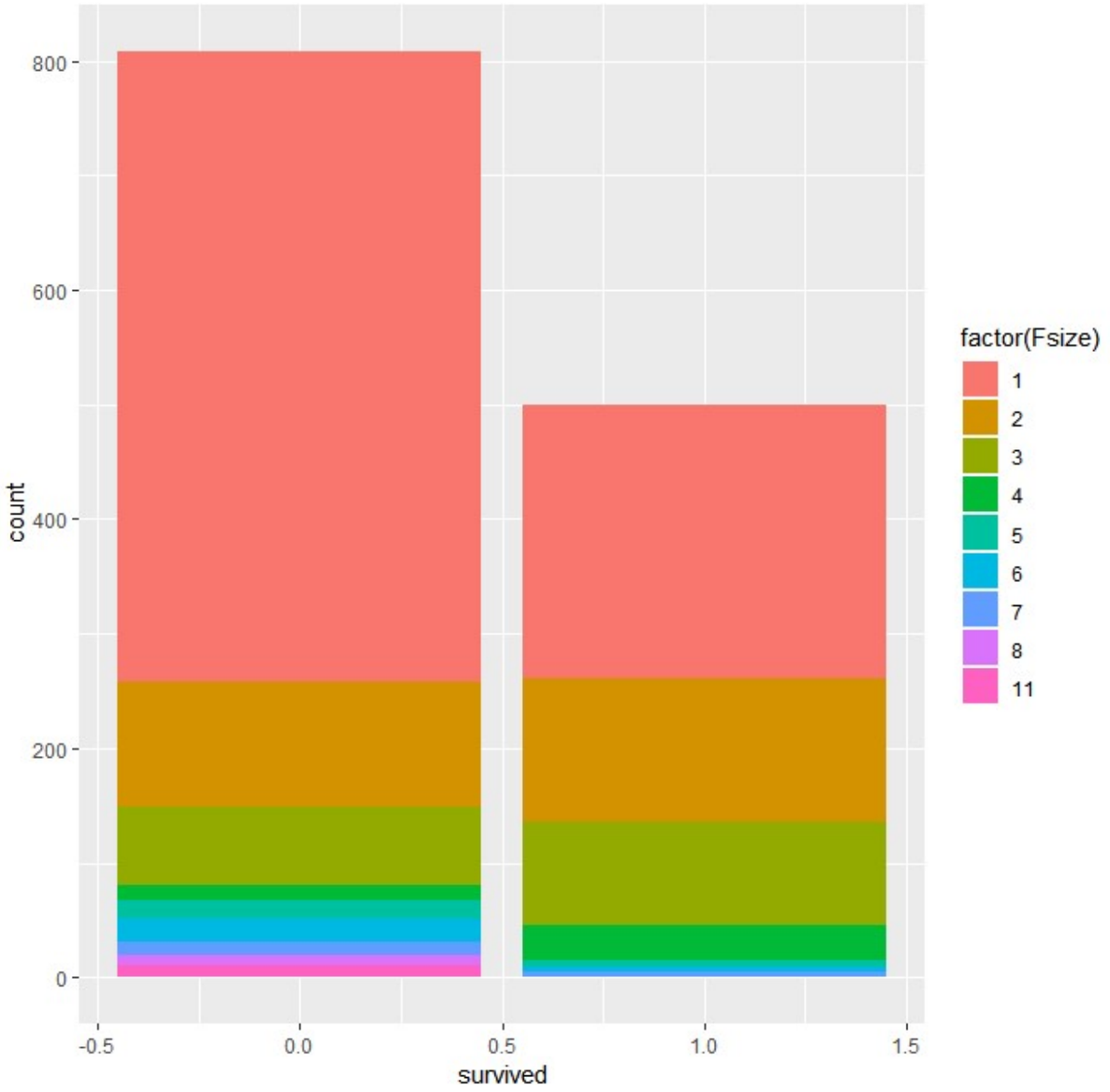
a. Pre-process the passenger names to come up with a list of titles that represent families and represent using appropriate visualization graph.

```
> titanic_nl=titanic3%>%separate(col=name, into=c("Last","First"),sep="\\. ")%>%separate(col=Last, into=c("Family","Title"),sep=", ")
Warning message:
Expected 2 pieces. Additional pieces discarded in 1 rows [248].
> View(titanic_nl)
```

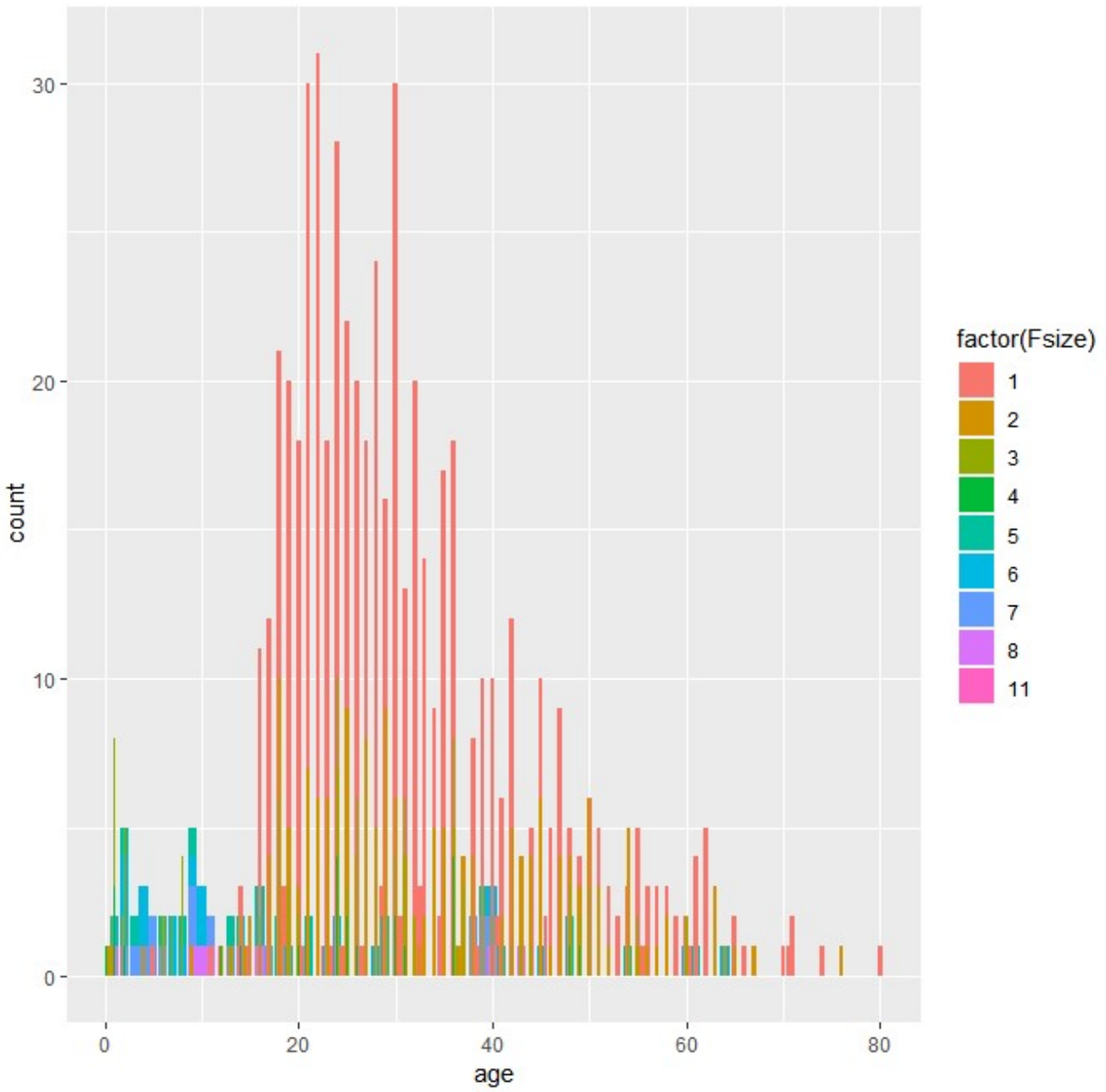
```
> titanic_nl%>%ggplot(aes(x=sex,fill=factor(survived) ))+geom_bar(stat = "count",position="fill")
```

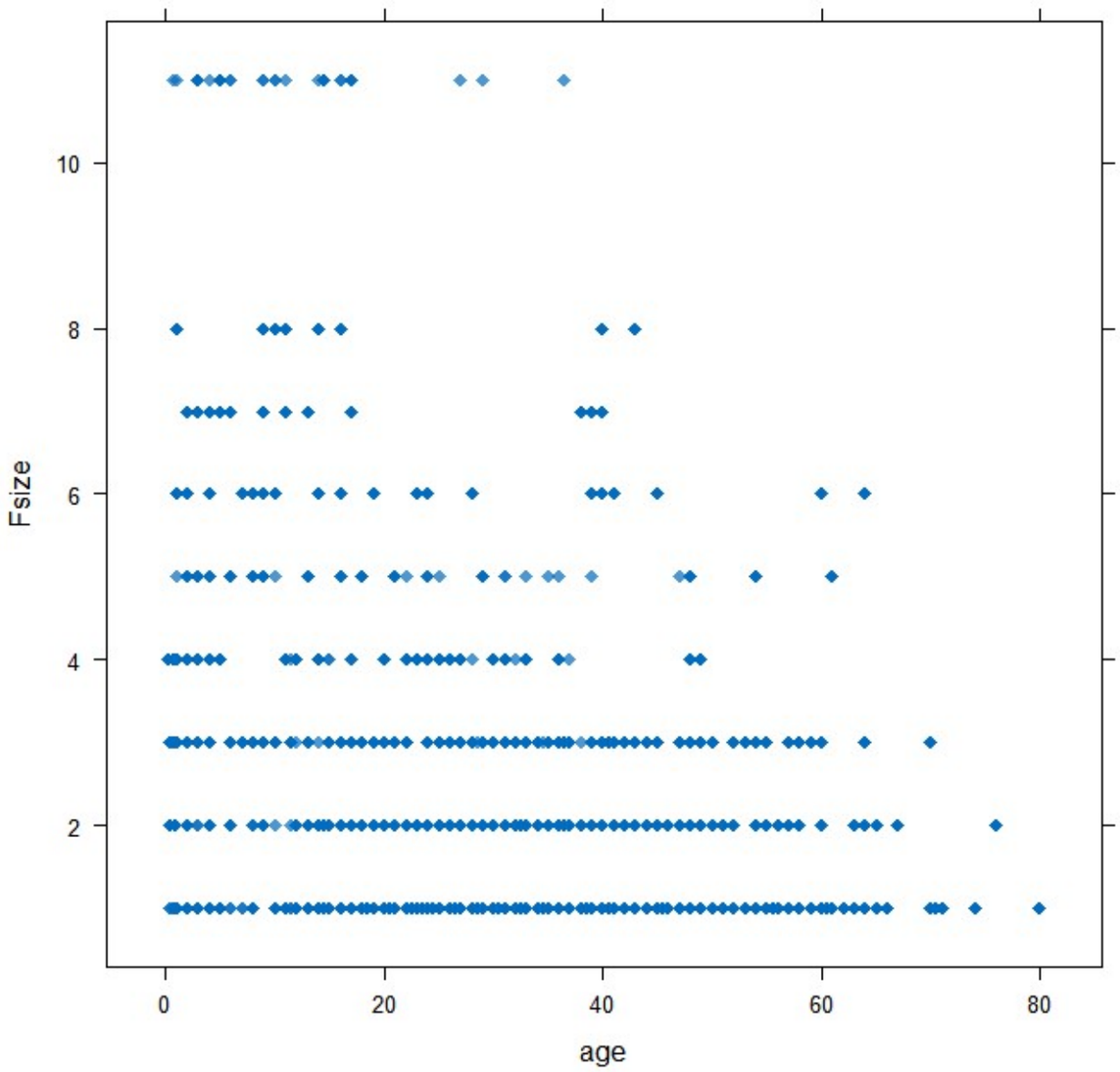


b. Represent the proportion of people survived by family size using a graph.



c. Impute the missing values in Age variable using Mice library, create two different graphs showing Age distribution before and after imputation




Fsize
10
8
6
4
2
0
20
40
60
80
age