ASSIGNMENT 6.1.a

Problem Statement

1. Import the Titanic Dataset from the following link:
https://drive.google.com/file/d/1JTJCjdGuUxzKXYlw0avwovB01k6FWg3r/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.

Answer:

Step1: Reading the ‘titanic.csv’ file into R

Step 2: Creating a function to find the “,” after each title and substr() function usage to pluck the titles.

```
getTitle<- function(x){
 if(str_detect(x, ",")== T){
  cPtr<- str_locate(x,",")
  titleName<- substr(x,1,cPtr-1)
  return(titleName)
 }
}
```

Step 3: Creating data frame and using the user defined function with only title names, and also counting the title names. All the below code are under the **mainFunc()**

```
mainFunc<- function(){
 titanicDf<- read.csv("C:/Users/arunabhl/Documents/MyRFiles/titanic3.csv")
 tDF<- data.frame(cbind(sapply(titanicDf$name, function(x) getTitle(x),
                       simplify =T )))
 colnames(tDF)<- "titleName"
 ttlCnt<- count(tDF, "titleName")
}
```

Step 4: To create visualization of the family title and the count. The below code is also under the **mainFunc()**

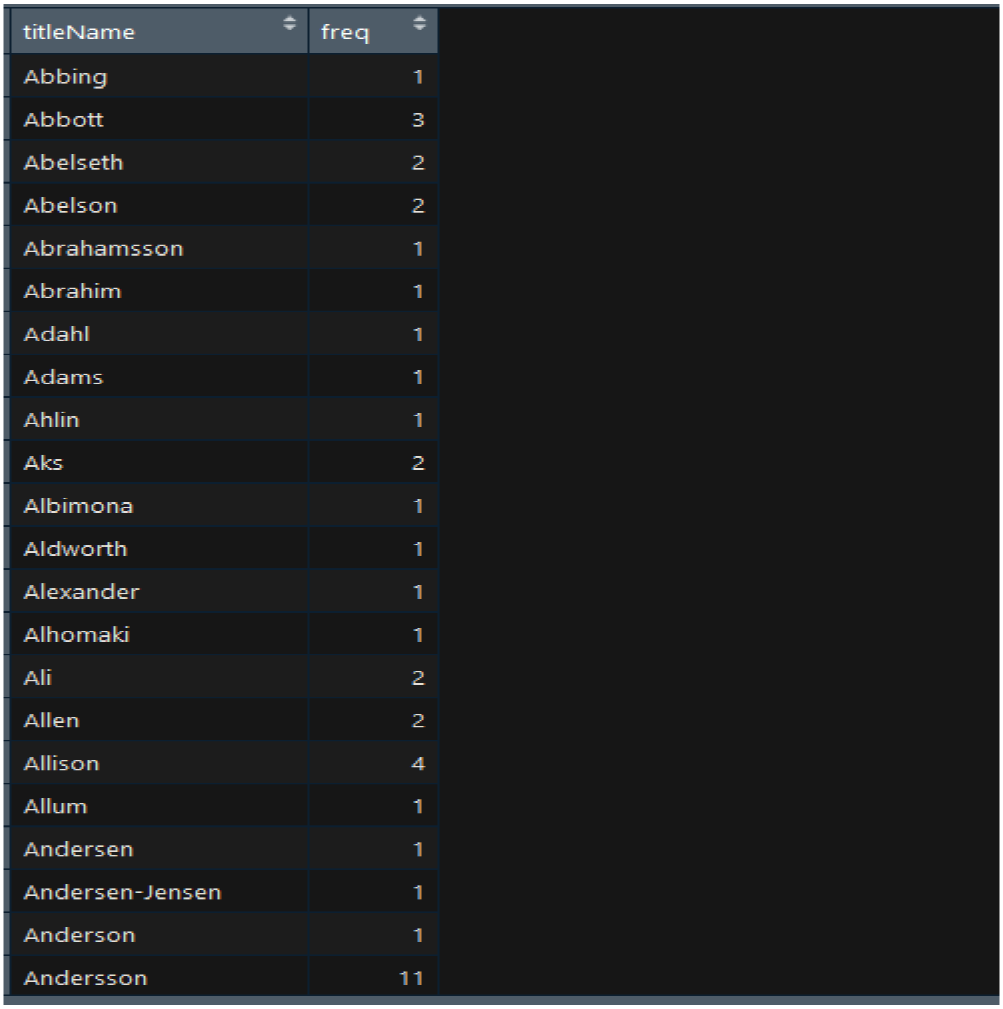
```
mCnt<- max(ttlCnt[,2])+1

  plot(ttlCnt,type="p",main="Family Title and Count Representation", ylab="No. of Family
members",

       xlab="Family Title", ylim=c(0,mCnt))
```

Step 5: To look for the family titles and the count, call the **ttlCnt variable**

Output:



titleName	freq
Abbing	1
Abbott	3
Abelseth	2
Abelson	2
Abrahamsson	1
Abrahim	1
Adahl	1
Adams	1
Ahlin	1
Aks	2
Albimona	1
Aldworth	1
Alexander	1
Alhomaki	1
Ali	2
Allen	2
Allison	4
Allum	1
Andersen	1
Andersen-Jensen	1
Anderson	1
Andersson	11

Note: R script is attached in the repository.