ACADGILD ASSIGNMENT 6.1

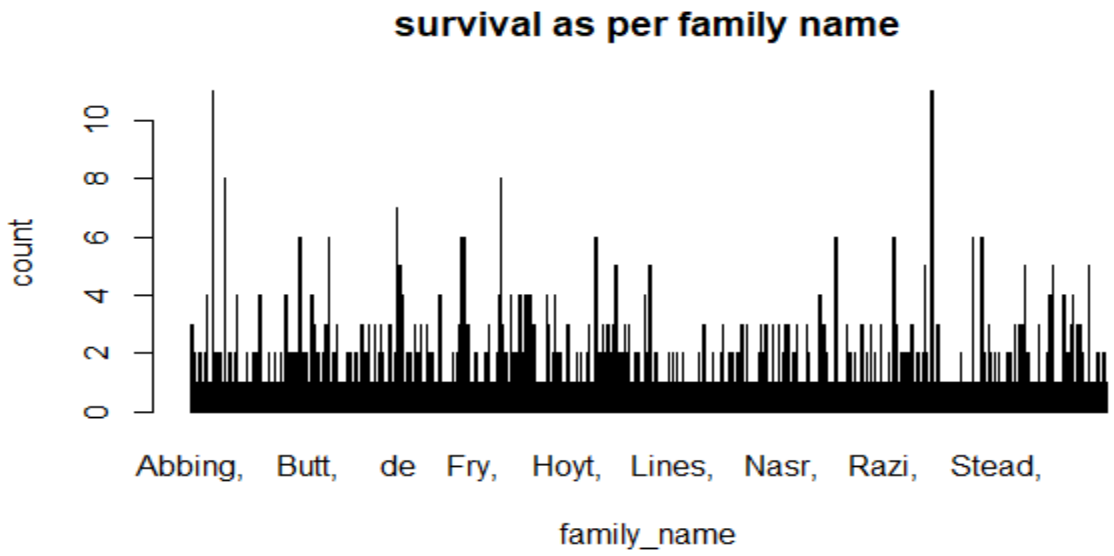
1. Titanic Dataset

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

Answer:

```
library(readr)

titanic3 <- read_csv("C:/Users/Rajesh
Chowdary/Downloads/titanic3.csv")

View(titanic3)

str(titanic3)

head(titanic3)

tail(titanic3)

str(titanic3$name)

titanic3$name<-as.character(titanic3$name)

namessplit<-do.call(rbind,strsplit(sub(" ",";",Titanic3$name),";"))

namessplit<-data.frame(namessplit)

names(namessplit)<-c("family_name", "name")

head(namessplit)

str(namessplit)

Title<-do.call(rbind,strsplit(sub(" ",";",namessplit$name),";"))
```

```
head(Title)
Title<-data.frame(Title)
names(Title)<-c("title", "first_name")
head(Title)
str(Title)
head(Title)
str(Titanic3)
TitanicData<-cbind(namessplit,Titanic3)
head(TitanicData)
View(TitanicData)
str(TitanicData)
TitanicData<-cbind(Title,TitanicData)
head(TitanicData)
View(TitanicData)
Title<-table(Title)
Title
View(Title)
familyname<-table(TitanicData$family_name)
View(familyname)
```

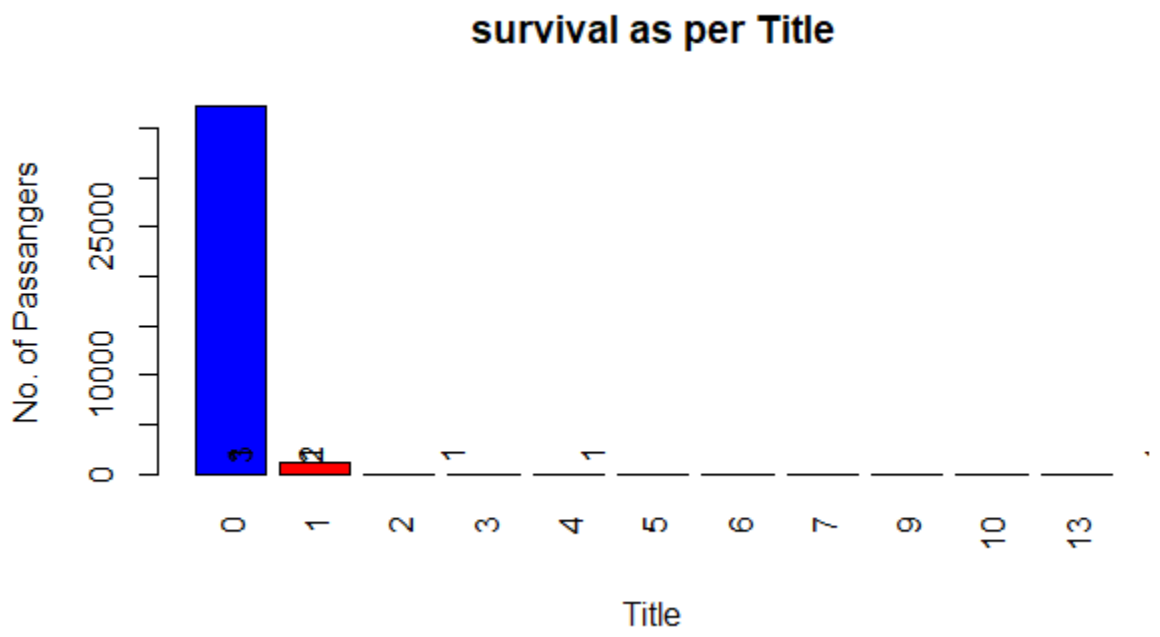


b. Represent the proportion of people survived by family size using a graph.

Answer:

```
View(Title)

barplot(Title,xlab = "Title", ylab = "No. of Passangers",

      main = "survival as per Title" , col = c("blue", "red"), las=3)

text(Title, 0,table(Title), pos = 3, srt = 90)

View(TitanicData)

SurvivedTitle<-table(TitanicData$Survived, TitanicData$title)

barplot(p,xlab = "Title", ylab = "survived",
```

main= "Survival as per title", col=rainbow(length(p



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

```
library(mice)

sum(is.na(TitanicData$age))

str(TitanicData)

mini_data <- TitanicData[-c(1,2,3,4,5,7,12,13,14,16,17,18)]

View(mini_data)

library(dplyr)

mini_data <- mini_data %>%

  mutate(
```

```
    survived = as.factor(survived),
    sex = as.factor(sex),
    age = as.numeric(age),
    sibsp = as.factor(sibsp),
    parch = as.factor(parch),
    embarked = as.factor(embarked)
  )
str(mini_data)
mice_data <- mice(mini_data, m=5, maxit=10,seed=500)
summary(mini_data)
```