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.**

```
library(readr)
library(readxl)
Titanic3 <- read_excel("titanic3.xls")

head(Titanic3)
tail(Titanic3)

str(Titanic3$name)
Titanic3$name <- as.character(Titanic3$name)
str(Titanic3$name)

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

head(namesplit)

namesplit <- data.frame(namesplit)
names(namesplit) <- c("family_name", "name")
head(namesplit)
str(namesplit)

Title <- do.call(rbind, strsplit(sub(" ",";",namesplit$name),";"))
head(Title)
Title <- data.frame(Title)
names(Title) <- c("title", "first_name")
head(Title)
str(Title)
head(Title)
```

```
> library(readr)
> library(readxl)
> Titanic3 <- read_excel("titanic3.xls")
Warning message:
In read_fun(path = path, sheet_i = sheet, limits = limits, shim = shim,  :
  Coercing text to numeric in M1306 / R1306C13: '328'
>
> head(Titanic3)
# A tibble: 6 x 14
  pclass survived name  sex      age sibsp parch ticket  fare cabin embark
ed boat
   <dbl>    <dbl> <chr> <chr>  <dbl> <dbl> <dbl> <chr>  <dbl> <chr> <chr>
<chr>
1      1        1 Alle~ fema~ 29         0     0 24160  211.  B5    S
2
2      1        1 Alli~ male   0.917     1     2 113781 152.  C22 ~ S
11
```

```
3      1        0 Alli~ fema~  2         1     2 113781 152.  C22 ~ S
NA
4      1        0 Alli~ male  30         1     2 113781 152.  C22 ~ S
NA
5      1        0 Alli~ fema~ 25         1     2 113781 152.  C22 ~ S
NA
6      1        1 Ande~ male  48         0     0 19952   26.6 E12   S
3
# ... with 2 more variables: body <dbl>, home.dest <chr>
> tail(Titanic3)
# A tibble: 6 x 14
  pclass survived name  sex     age sibsp parch ticket  fare cabin embarke
d boat
   <dbl>    <dbl> <chr> <chr> <dbl> <dbl> <dbl> <chr>  <dbl> <chr> <chr>
<chr>
1      3        0 Yous~ male   NA       0     0 2627   14.5  NA    C
NA
2      3        0 Zabo~ fema~  14.5     1     0 2665   14.5  NA    C
NA
3      3        0 Zabo~ fema~  NA       1     0 2665   14.5  NA    C
NA
4      3        0 Zaka~ male   26.5     0     0 2656    7.22 NA    C
NA
5      3        0 Zaka~ male   27       0     0 2670    7.22 NA    C
NA
6      3        0 Zimm~ male   29       0     0 315082  7.88 NA    S
NA
# ... with 2 more variables: body <dbl>, home.dest <chr>
>
> str(Titanic3$name)
 chr [1:1309] "Allen, Miss. Elisabeth Walton" "Allison, Master. Hudson Tre
vor" ...
> Titanic3$name <- as.character(Titanic3$name)
> str(Titanic3$name)
 chr [1:1309] "Allen, Miss. Elisabeth Walton" "Allison, Master. Hudson Tre
vor" ...
>
> namesplit <- do.call(rbind,strsplit(sub(" ",";",Titanic3$name),";"))
>
> head(namesplit)
     [,1]        [,2]
[1,] "Allen,"    "Miss. Elisabeth Walton"
[2,] "Allison,"  "Master. Hudson Trevor"
[3,] "Allison,"  "Miss. Helen Loraine"
[4,] "Allison,"  "Mr. Hudson Joshua Creighton"
[5,] "Allison,"  "Mrs. Hudson J C (Bessie Waldo Daniels)"
[6,] "Anderson," "Mr. Harry"
>
> namesplit <- data.frame(namesplit)
> names(namesplit) <- c("family_name", "name")
> head(namesplit)
  family_name                                 name
1      Allen,                Miss. Elisabeth Walton
2    Allison,                 Master. Hudson Trevor
3    Allison,                   Miss. Helen Loraine
4    Allison,           Mr. Hudson Joshua Creighton
5    Allison, Mrs. Hudson J C (Bessie Waldo Daniels)
6   Anderson,                             Mr. Harry
> str(namesplit)
'data.frame':  1309 obs. of  2 variables:
 $ family_name: Factor w/ 868 levels "Abbing,","Abbott,",..: 16 17 17 17 1
7 21 25 25 28 31 ...
```

```
 $ name       : Factor w/ 1144 levels "Billiard, Master. James William",..
: 159 64 194 575 1019 549 232 864 982 792 ...
>
> Title <- do.call(rbind, strsplit(sub(" ",";",namesplit$name),";"))
> head(Title)
     [,1]      [,2]
[1,] "Miss."   "Elisabeth Walton"
[2,] "Master." "Hudson Trevor"
[3,] "Miss."   "Helen Loraine"
[4,] "Mr."     "Hudson Joshua Creighton"
[5,] "Mrs."    "Hudson J C (Bessie Waldo Daniels)"
[6,] "Mr."     "Harry"
> Title <- data.frame(Title)
> names(Title) <- c("title", "first_name")
> head(Title)
    title                          first_name
1   Miss.                    Elisabeth Walton
2 Master.                       Hudson Trevor
3   Miss.                       Helen Loraine
4     Mr.             Hudson Joshua Creighton
5    Mrs. Hudson J C (Bessie Waldo Daniels)
6     Mr.                               Harry
> str(Title)
'data.frame':  1309 obs. of  2 variables:
 $ title     : Factor w/ 34 levels "Billiard,","Brito,",..: 18 15 18 21 22
21 18 21 22 21 ...
 $ first_name: Factor w/ 1127 levels "(Ada E Hall)",..: 298 508 465 507 50
6 456 672 1025 271 909 ...
> head(Title)
    title                          first_name
1   Miss.                    Elisabeth Walton
2 Master.                       Hudson Trevor
3   Miss.                       Helen Loraine
4     Mr.             Hudson Joshua Creighton
5    Mrs. Hudson J C (Bessie Waldo Daniels)
6     Mr.                               Harry
```

str(Titanic3)
TitanicData<-cbind(namesplit,Titanic3)
head(TitanicData)
View(TitanicData)
str(TitanicData)
TitanicData<-cbind(Title,TitanicData)
head(TitanicData)
View(TitanicData)

str(Titanic3)
TitanicData<-cbind(namessplit,Titanic3)
head(TitanicData)
View(TitanicData)
str(TitanicData)
TitanicData<-cbind(Title,TitanicData)
head(TitanicData)
View(TitanicData)

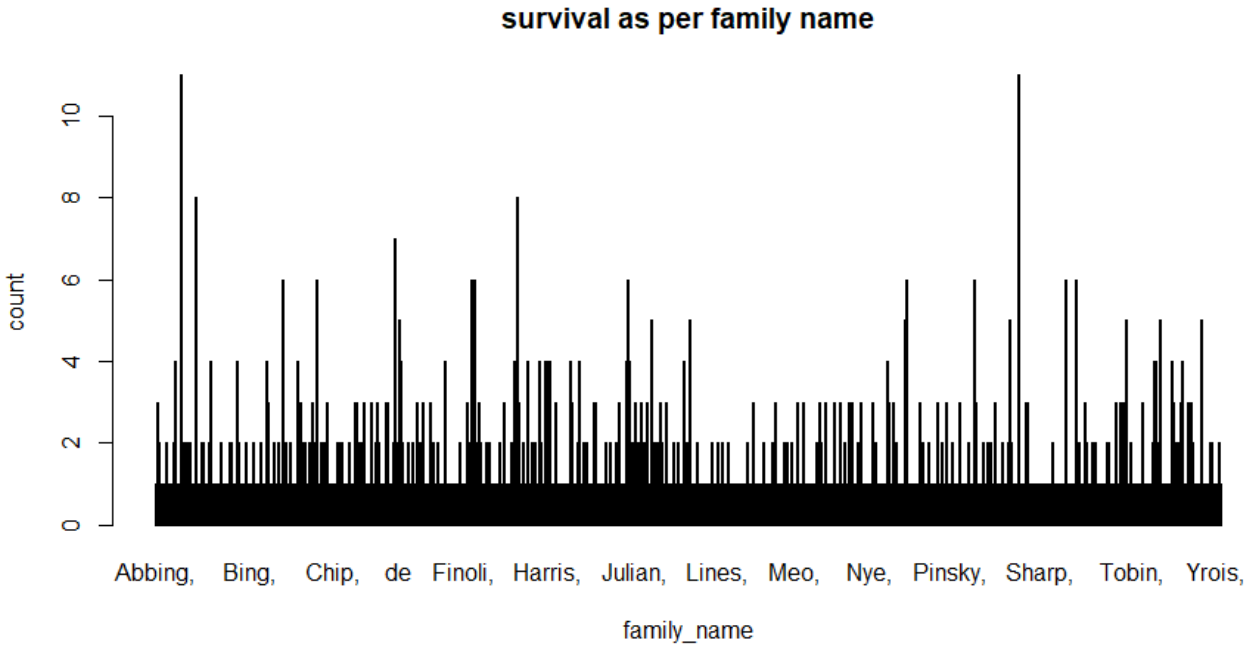
```
subtitles<-gsub("\\..*", " ", TitanicData$name)
head(subtitles)

Title<-gsub(".*\\ ", " ", subtitles)
Title
head(Title)
```

```
> str(Titanic3)
Classes ‘tbl_df’, ‘tbl’ and 'data.frame':    1309 obs. of  14 variables:
 $ pclass   : num  1 1 1 1 1 1 1 1 1 1 ...
 $ survived : num  1 1 0 0 0 1 1 0 1 0 ...
 $ name     : chr  "Allen, Miss. Elisabeth Walton" "Allison, Master. Hudso
n Trevor" "Allison, Miss. Helen Loraine" "Allison, Mr. Hudson Joshua Creig
hton" ...
 $ sex      : chr  "female" "male" "female" "male" ...
 $ age      : num  29 0.917 2 30 25 ...
 $ sibsp    : num  0 1 1 1 1 0 1 0 2 0 ...
 $ parch    : num  0 2 2 2 2 0 0 0 0 0 ...
 $ ticket   : chr  "24160" "113781" "113781" "113781" ...
 $ fare     : num  211 152 152 152 152 ...
 $ cabin    : chr  "B5" "C22 C26" "C22 C26" "C22 C26" ...
 $ embarked : chr  "S" "S" "S" "S" ...
 $ boat     : chr  "2" "11" NA NA ...
 $ body     : num  NA NA NA 135 NA NA NA NA NA 22 ...
 $ home.dest: chr  "St Louis, MO" "Montreal, PQ / Chesterville, ON" "Montr
eal, PQ / Chesterville, ON" "Montreal, PQ / Chesterville, ON" ...
> TitanicData<-cbind(namessplit,Titanic3)
Error in cbind(namessplit, Titanic3) : object 'namessplit' not found
> head(TitanicData)
    title                        first_name   title
1   Miss.                  Elisabeth Walton   Miss.
2 Master.                     Hudson Trevor Master.
3   Miss.                     Helen Loraine   Miss.
4     Mr.            Hudson Joshua Creighton     Mr.
5    Mrs. Hudson J C (Bessie Waldo Daniels)    Mrs.
6     Mr.                               Harry     Mr.
                      first_name   title                          first_n
ame
1                  Elisabeth Walton   Miss.                   Elisabeth Wal
ton
2                     Hudson Trevor Master.                      Hudson Tre
vor
3                     Helen Loraine   Miss.                      Helen Lora
ine
4           Hudson Joshua Creighton     Mr.           Hudson Joshua Creigh
ton
5 Hudson J C (Bessie Waldo Daniels)    Mrs. Hudson J C (Bessie Waldo Danie
ls)
6                             Harry     Mr.                              Ha
rry
  family_name                                    name pclass survived
1      Allen,                  Miss. Elisabeth Walton      1        1
2    Allison,                   Master. Hudson Trevor      1        1
3    Allison,                     Miss. Helen Loraine      1        0
4    Allison,             Mr. Hudson Joshua Creighton      1        0
5    Allison, Mrs. Hudson J C (Bessie Waldo Daniels)      1        0
6   Anderson,                                Mr. Harry      1        1
```

```
                                              name    sex     age sibsp par
ch
1                    Allen, Miss. Elisabeth Walton female 29.0000     0
0
2                   Allison, Master. Hudson Trevor   male  0.9167     1
2
3                     Allison, Miss. Helen Loraine female  2.0000     1
2
4             Allison, Mr. Hudson Joshua Creighton   male 30.0000     1
2
5 Allison, Mrs. Hudson J C (Bessie Waldo Daniels) female 25.0000     1
2
6                              Anderson, Mr. Harry   male 48.0000     0
0
  ticket     fare   cabin embarked boat body                       home.de
st
1  24160 211.3375      B5        S    2   NA                     St Louis,
MO
2 113781 151.5500 C22 C26        S   11   NA Montreal, PQ / Chesterville,
ON
3 113781 151.5500 C22 C26        S <NA>   NA Montreal, PQ / Chesterville,
ON
4 113781 151.5500 C22 C26        S <NA>  135 Montreal, PQ / Chesterville,
ON
5 113781 151.5500 C22 C26        S <NA>   NA Montreal, PQ / Chesterville,
ON
6  19952  26.5500     E12        S    3   NA                     New York,
NY
> View(TitanicData)
> str(TitanicData)
'data.frame':	1309 obs. of  22 variables:
 $ title      : Factor w/ 34 levels "Billiard,","Brito,",..: 18 15 18 21 2
2 21 18 21 22 21 ...
 $ first_name : Factor w/ 1127 levels "(Ada E Hall)",..: 298 508 465 507 5
06 456 672 1025 271 909 ...
 $ title      : Factor w/ 34 levels "Billiard,","Brito,",..: 18 15 18 21 2
2 21 18 21 22 21 ...
 $ first_name : Factor w/ 1127 levels "(Ada E Hall)",..: 298 508 465 507 5
06 456 672 1025 271 909 ...
 $ title      : Factor w/ 34 levels "Billiard,","Brito,",..: 18 15 18 21 2
2 21 18 21 22 21 ...
 $ first_name : Factor w/ 1127 levels "(Ada E Hall)",..: 298 508 465 507 5
06 456 672 1025 271 909 ...
 $ family_name: Factor w/ 868 levels "Abbing,","Abbott,",..: 16 17 17 17 1
7 21 25 25 28 31 ...
 $ name       : Factor w/ 1144 levels "Billiard, Master. James William",..
: 159 64 194 575 1019 549 232 864 982 792 ...
 $ pclass     : num  1 1 1 1 1 1 1 1 1 1 ...
 $ survived   : num  1 1 0 0 0 1 1 0 1 0 ...
 $ name       : chr  "Allen, Miss. Elisabeth Walton" "Allison, Master. Hud
son Trevor" "Allison, Miss. Helen Loraine" "Allison, Mr. Hudson Joshua Cre
ighton" ...
 $ sex        : chr  "female" "male" "female" "male" ...
 $ age        : num  29 0.917 2 30 25 ...
 $ sibsp      : num  0 1 1 1 1 0 1 0 2 0 ...
 $ parch      : num  0 2 2 2 2 0 0 0 0 0 ...
 $ ticket     : chr  "24160" "113781" "113781" "113781" ...
 $ fare       : num  211 152 152 152 152 ...
 $ cabin      : chr  "B5" "C22 C26" "C22 C26" "C22 C26" ...
 $ embarked   : chr  "S" "S" "S" "S" ...
 $ boat       : chr  "2" "11" NA NA ...
 $ body       : num  NA NA NA 135 NA NA NA NA NA 22 ...
```

```
 $ home.dest  : chr  "St Louis, MO" "Montreal, PQ / Chesterville, ON" "Mon
treal, PQ / Chesterville, ON" "Montreal, PQ / Chesterville, ON" ...
> TitanicData<-cbind(Title,TitanicData)
> head(TitanicData)
  Title   title                       first_name   title
1         Miss.                  Elisabeth Walton   Miss.
2       Master.                     Hudson Trevor Master.
3         Miss.                     Helen Loraine   Miss.
4           Mr.             Hudson Joshua Creighton     Mr.
5          Mrs. Hudson J C (Bessie Waldo Daniels)    Mrs.
6           Mr.                               Harry     Mr.
                         first_name   title                          first_n
ame
1                  Elisabeth Walton   Miss.                  Elisabeth Wal
ton
2                     Hudson Trevor Master.                     Hudson Tre
vor
3                     Helen Loraine   Miss.                     Helen Lora
ine
4            Hudson Joshua Creighton     Mr.           Hudson Joshua Creigh
ton
5 Hudson J C (Bessie Waldo Daniels)    Mrs. Hudson J C (Bessie Waldo Danie
ls)
6                              Harry     Mr.                              Ha
rry
  family_name                                    name pclass survived
1      Allen,                  Miss. Elisabeth Walton      1        1
2    Allison,                   Master. Hudson Trevor      1        1
3    Allison,                     Miss. Helen Loraine      1        0
4    Allison,             Mr. Hudson Joshua Creighton      1        0
5    Allison, Mrs. Hudson J C (Bessie Waldo Daniels)      1        0
6   Anderson,                                 Mr. Harry      1        1
                                        name    sex     age sibsp par
ch
1                  Allen, Miss. Elisabeth Walton female 29.0000     0
0
2                 Allison, Master. Hudson Trevor   male  0.9167     1
2
3                   Allison, Miss. Helen Loraine female  2.0000     1
2
4            Allison, Mr. Hudson Joshua Creighton   male 30.0000     1
2
5 Allison, Mrs. Hudson J C (Bessie Waldo Daniels) female 25.0000     1
2
6                              Anderson, Mr. Harry   male 48.0000     0
0
  ticket     fare   cabin embarked boat body                        home.de
st
1  24160 211.3375      B5        S    2   NA                     St Louis,
MO
2 113781 151.5500 C22 C26        S   11   NA Montreal, PQ / Chesterville,
ON
3 113781 151.5500 C22 C26        S <NA>   NA Montreal, PQ / Chesterville,
ON
4 113781 151.5500 C22 C26        S <NA>  135 Montreal, PQ / Chesterville,
ON
5 113781 151.5500 C22 C26        S <NA>   NA Montreal, PQ / Chesterville,
ON
6  19952  26.5500     E12        S    3   NA                     New York,
NY
> View(TitanicData)
```

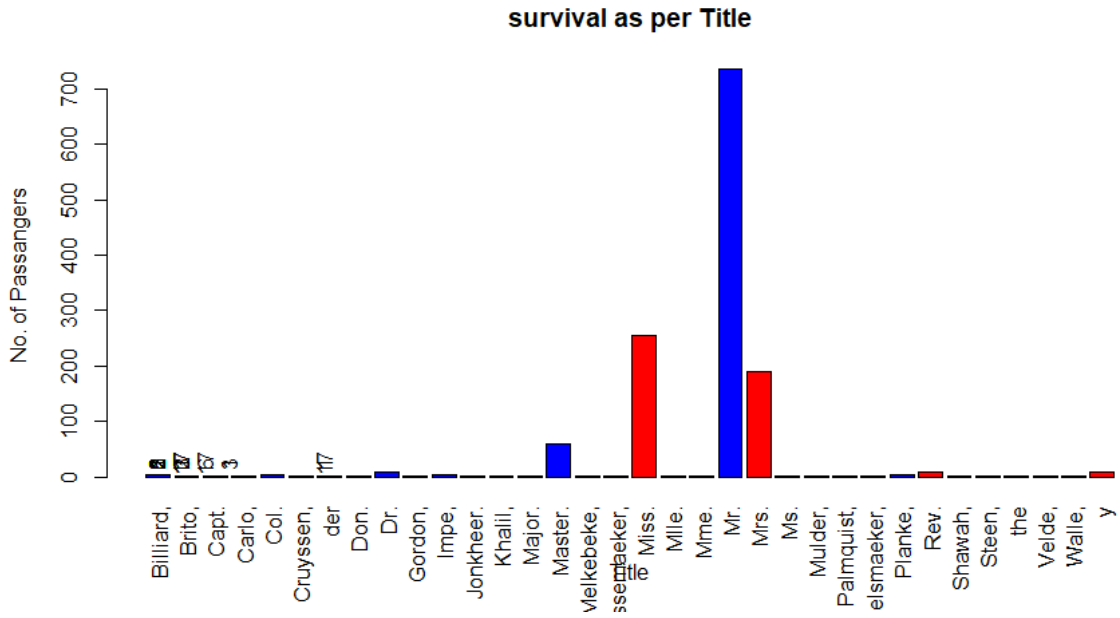
familyname<-table(TitanicData$family_name)
View(familyname)
barplot(familyname,main = "survival as per family name", xlab = "family_name", ylab = "count",col ="red")



Title <- table(TitanicData$title)
Title
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)

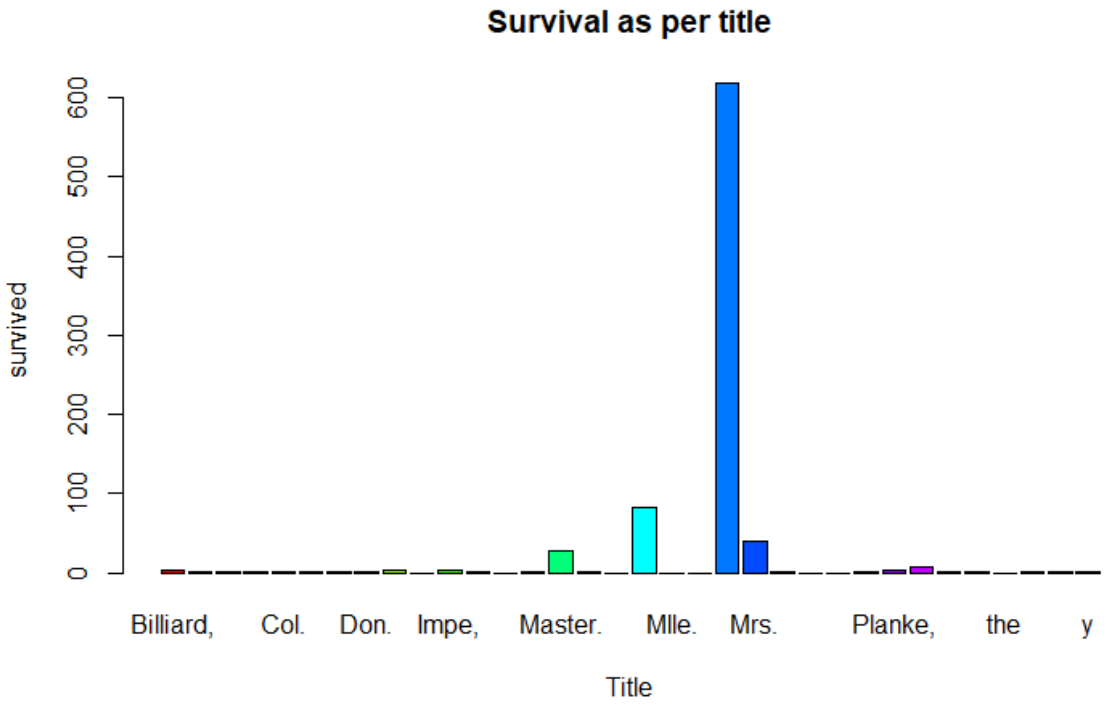
```
> Title <- table(TitanicData$title)
> Title

   Billiard,       Brito,        Capt.       Carlo,         Col.    Cruyss
en,
           3            1            1            2            4
1
         der         Don.          Dr.      Gordon,        Impe,    Jonkhe
er.
           1            1            8            2            3
1
     Khalil,       Major.      Master.   Melkebeke, Messemaeker,        Mi
ss.
           1            2           59            1            2
256
```

```
       Mlle.          Mme.           Mr.          Mrs.           Ms.       Muld
er,
          2             1           736           191             2
1
  Palmquist,   Pelsmaeker,       Planke,          Rev.       Shawah,        Ste
en,
          1             1            4             8             1
1
         the        Velde,        Walle,             y
          1             1            1             8
> View(Title)
> barplot(Title,xlab = "Title", ylab = "No. of Passangers", main = "surviv
al as per Title" , col = c("blue", "red"), las=3)
> text(Title, 0,table(Title), pos = 3, srt = 90)
```



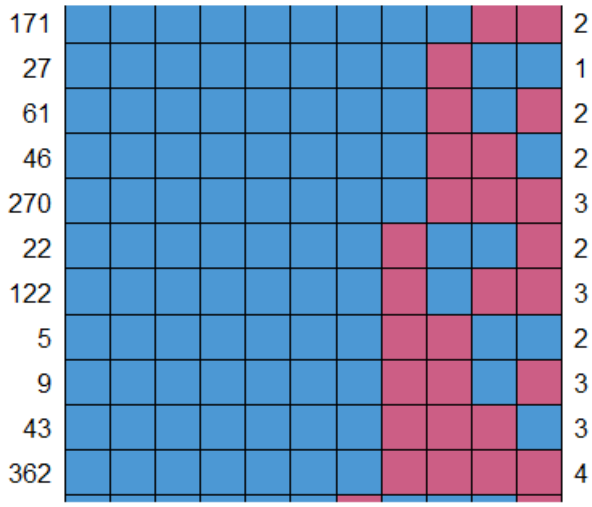
b. Represent the proportion of people survived from the family title using a graph.

```
View(TitanicData)
Survived_data <- table(TitanicData$survived, TitanicData$title)
psg <- Survived_data[1,]

barplot(psg,xlab = "Title", ylab = "survived",main= "Survival as per title",
      col=rainbow(length(psg)))
```



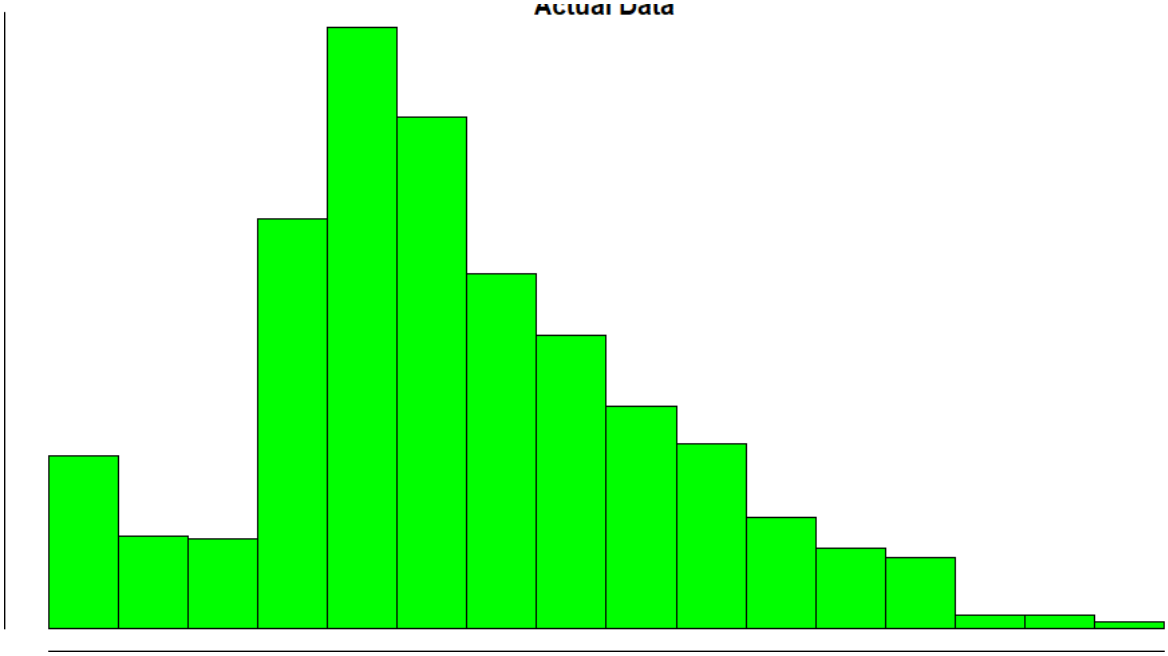
c. **c. Impute the missing values in Age variable using Mice library, create two different graphs showing Age distribution before and after imputation**

```
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)
md.pattern(mini_data)
```




```
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)
md.pattern(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)
Imputed=complete(mice_data,5)
```

```
hist(TitanicData$age, main='Actual
Data',col="green")
```




```
hist(Imputed$age, main='Imputed Data',col="black")
```