

Assignments for session on "VISUALIZATION AND PLOTTING"

1. Importing the Titanic Dataset
2. Preprocess to get the Family Names
3. Separate the Titles
4. Represent the Barplots by Family names
5. Proportion of the survived by graphical representation

```
titanic3 <- read.csv("C:/users/sarang_dani/Desktop/R/Acadgild's Assignments/titanic3.csv")
head(titanic3)
str(titanic3$name)
titanic3$name <- as.character(titanic3$name)
str(titanic3$name)
#To get the Family names, seprate the spaces and use rbind to consolidate the split string
names_split <- do.call(rbind,strsplit(sub(" ","",titanic3$name),";"))
head(names_split)
#converting to dataframe and give column names
names_split <- as.data.frame(names_split)
names(names_split) <- c("Family name","First Name")
head(names_split)

titanic3$Title<-(do.call(rbind,strsplit(titanic3$name," ")))[,2]
Title

#Barplot for family names
FamilyName <- table(names_split$`Family name`)
FamilyName
barplot(FamilyName,main = "Survival as per Family Name", xlab = "Family Name", ylab = "count",col

#Barplot for survivals as per Titles
Title<-table(Title)
Title
barplot(Title,xlab = "Title", ylab = "No. of Passangers",
        main = "survival as per Title" , col = c("blue", "red"), las=3)

# Represent the proportion of people survived from the family size using a graph.
survivedTitle<-table(titanic3$survived,titanic3$Title)
#Since the survived is 0, first row. we will take only that
p<-SurvivedTitle[1,]

#barplot of survived numbers per title
barplot(p,xlab = "Title", ylab = "survived",
        main= "Survival as per title", col=rainbow(length(p)))
#pie chaart showing proportion of survival title wise

pie_chart<-pie(p, main = "Pie-Chart of Titles survived", col = rainbow(length(p)) )
legend("topright", names(p), cex= 0.5, fill = 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.
text(Title, 0,table(Title), pos = 3, srt = 90)
```